



UNIVERSIDAD CENTRAL DE VENEZUELA

FACULTAD DE CIENCIAS

ESCUELA DE COMPUTACIÓN

CENTRO DE INFORMACIÓN DE SISTEMAS DE INFORMACIÓN

**Sistema de Inteligencia de Negocios para Apoyar la Toma de Decisiones del
Proceso Encuesta Simple**

Trabajo de Especial de Grado presentado ante la ilustre

Universidad Central de Venezuela por el

Br. Simón Tribiño

Tutor: Prof. Franky Uzcátegui

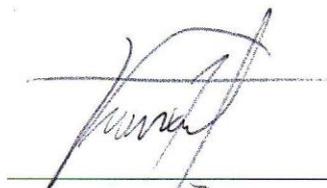
Caracas, Octubre 2014

ACTA

Quienes suscriben, miembros del Jurado designado por el Consejo de Escuela de Computación, para examinar el Trabajo Especial de Grado presentado por el bachiller Simón Alfredo, Tribiño Betancourt portador de la Cédula de Identidad V-19.379.291, con el título: **“Sistema de Inteligencia de Negocios para Apoyar la Toma de Decisiones del Proceso Encuesta Simple”**, a los fines de optar al título de Licenciado en Computación, dejan constancia de lo siguiente:

Leído como fue, dicho trabajo por cada uno de los miembros del jurado, se fijó el día 29 de Octubre de 2014 a la hora 11:00 am, para que su autor lo defienda en forma pública, lo que hizo en la Sala III de Postgrado, de la Escuela de Computación, mediante una presentación oral de su contenido, luego de lo cual respondió las preguntas formuladas. Finalizada la defensa pública del Trabajo Especial de Grado, el jurado decidió aprobarlo con la nota de 20 puntos.

En fe de lo cual se levanta la presente Acta, en Caracas a los veintinueve días del mes de Octubre del año dos mil catorce (2014), dejando constancia de que actuó como Coordinador del Jurado el Profesor tutor Franky Uzcátegui.



Prof. Franky Uzcátegui
Tutor



Prof. Franklin Sandoval
Jurado



Prof. Eric Gamess
Jurado

DEDICATORIA

A mis padres “Sheyla Betancourt” y “Pedro Rangel”,
a quienes le debo todo lo que soy.

A mi abuelo
el mejor ejemplo a seguir.

A toda mi familia,
por su cariño, afecto y por contar siempre con su apoyo.

Simón A. Tribiño B.

AGRADECIMIENTOS

Este trabajo especial de grado ha requerido de mucho esfuerzo y dedicación, su culminación involucra la ayuda desinteresada de todas las personas que agradeceré a continuación.

A mi mamá, Sheyla, por amarme y apoyarme incondicionalmente. A ella agradezco todo lo que soy y le debo el logro de esta meta. Es la base de mi vida, y todo lo que tengo no lo hubiera podido conseguir sin su apoyo. Gracias por todo tu amor mamá, eres la madre que todos desearían tener y solo yo tengo, te amo.

A mi papá, Pedro, por ser más que un padre ser un amigo, por escucharme y guiarme, por ayudarme en todo esos momentos difíciles en los que lo he necesitado y por apoyarme en todos mis sueños y creer en mí siempre.

A mi Abuelo, Luis José, por ser un padre, un amigo, un apoyo y el mejor ejemplo a seguir. Me llena de felicidad saber que si hoy estuvieras aquí conmigo, estarías muy orgulloso de mí. Te amo y te extraño.

A mi familia, le agradezco hoy y siempre el apoyo que me brindan no solo en el desarrollo de este trabajo sino en todos los aspectos de mi vida, por criarme de la manera correcta y encaminarme.

A mi amigo Víctor Vargas, el que me acompañó durante toda la carrera, en casi todas las materias, estudiando y haciendo proyectos, logrando alcanzar esta meta que hace cinco años comenzamos juntos y hoy terminamos juntos. Gracias por guiarme, aconsejarme, escucharme, soportarme y por todo lo que de ti he aprendido, de todo corazón te deseo lo mejor hermano.

A mi excelente tutor Franky Uzcátegui por enseñarme todo lo necesario para desarrollar esta tesis, por brindarme todo su apoyo y por ayudar en mi formación como profesional, gracias por creer en mí.

A mi casa de estudio la Ilustre “Universidad Central de Venezuela”, la casa que vence las sombras, por permitirme ser parte de ella y ser testigo de un crecimiento profesional.

En general quiero agradecer a todas y cada una de las personas que han vivido conmigo el desarrollo de esta tesis. No hace falta nombrarlos porque tanto ellos como yo sabemos que desde los más profundo de mi corazón les agradezco el brindarme todo el apoyo, la colaboración, el ánimo y sobre todo el cariño y la amistad.

A todos ustedes...

¡GRACIAS!

Simón A. Tribiño B.

RESUMEN

En el presente Trabajo Especial de Grado se exhiben un conjunto de conceptos, metodologías y herramientas que permiten llevar a cabo el desarrollo de una solución de inteligencia de negocios a partir del modelo de datos transaccional del proceso de negocio Encuesta Simple. La solución implementada se aplica en aquellas organizaciones que requieren conocer con qué información cuentan a partir de sus sistemas transaccionales, antes de definir sus requerimientos de información en forma de indicadores y reportes.

La utilización de una solución de inteligencia de negocios permitirá mejorar la toma de decisiones a nivel gerencial, lo cual contribuirá a incrementar los ingresos y las oportunidades de negocios, así como también ayudará a disminuir los riesgos y los costos de la organización.

Como método de desarrollo de la solución, se utilizó el enfoque de la empresa Tian Consultores, la cual es una adaptación basada en experiencias propias y en los enfoques ascendente (Kimball, R.) y descendente (Inmon, B.), que sintetiza y procedimenta sus mejores prácticas. La plataforma tecnológica utilizada como respaldo para el desarrollo de la solución, es la aplicación de Oracle Business Intelligence Standard Edition One.

ÍNDICE

ACTA.....	¡Error! Marcador no definido.
DEDICATORIA	3
AGRADECIMIENTOS	4
RESUMEN	6
ÍNDICE.....	7
ÍNDICE DE TABLAS	11
ÍNDICE DE ILUSTRACIONES	12
INTRODUCCIÓN	15
CAPÍTULO 1. PROBLEMA DE INVESTIGACIÓN.....	17
1.1 Problema	17
1.2 Objetivo General	17
1.3 Objetivos Específicos	17
1.4 Solución.....	18
1.5 Alcance.....	19
CAPÍTULO 2: MARCO CONCEPTUAL	20
2.1 Conceptos Básicos.....	20
2.1.1 Dato	20
2.1.2 Información	20
2.1.3 Conocimiento	20
2.1.4 Inteligencia	21
2.1.5 Sabiduría.....	21
2.2 Sistema	22
2.3 Sistemas de Información	22
2.3.1 Definición.....	22
2.3.2 Actividades de los Sistemas de Información.....	22
2.3.3 Clasificación de los Sistemas de Información.....	23

2.4	Sistemas OLTP y OLAP.....	26
2.4.1	Sistemas OLTP.....	26
2.4.2	Sistemas OLAP.....	27
2.4.3	Diferencias entre OLTP y OLAP.....	28
2.5	Base de Datos.....	29
2.5.1	Definición.....	29
2.5.2	Características de una Base de Datos.....	30
2.5.3	Sistema Manejador de Base de Datos.....	30
2.5.4	Modelos de Bases de Datos.....	31
2.6	Almacén de Datos.....	40
2.6.1	Introducción.....	40
2.6.2	Antecedentes Históricos.....	40
2.6.3	Definición.....	41
2.6.4	Objetivos de un Almacén de Datos.....	43
2.6.5	Dimensiones.....	43
2.6.6	Tabla de Hechos.....	44
2.6.7	Claves en las Tablas de Hechos.....	45
2.6.8	Cubos de Información.....	45
2.6.9	Variables.....	46
2.6.10	Granularidad.....	47
2.6.11	Dimensión Tiempo.....	48
2.6.12	Agregación.....	49
2.6.13	Jerarquía.....	49
2.6.14	Operaciones de Exploración Ascendente y Exploración Descendente.....	50
2.6.15	Metadatos.....	51
2.6.16	Ventajas de los Almacenes de Datos.....	51
2.6.17	Desventajas de los Almacenes de Datos.....	52

2.6.18	Tipos de Implementación de un Almacén de Datos	52
2.7	Inteligencia de Negocios	58
2.7.1	Definición.....	58
2.7.2	Usuarios de las Soluciones de Inteligencia de Negocios.....	59
2.7.3	Historia de la Inteligencia de Negocios	59
2.7.4	Características de la Inteligencia de Negocios	60
2.7.5	Arquitectura de una Solución de Inteligencia de Negocios.....	60
2.7.6	Procesos Básicos del Almacén de Datos (ETC).....	62
2.7.7	Inteligencia de Negocios en los Distintos Niveles Jerárquicos de la Organización.....	63
2.7.8	Inteligencia de Negocios y Almacén de Datos.....	64
2.7.9	Ventajas de la Inteligencia de Negocios	65
2.7.10	Principales Software para Inteligencia de Negocio.....	66
2.8	Tecnologías y Herramientas de Inteligencia de Negocios a Utilizar	68
2.8.1	Oracle BI (En español, Inteligencia de Negocio de Oracle).....	68
2.8.2	Oracle Warehouse Builder (En español, Desarrollador de Almacenes de Datos de Oracle) 70	
2.8.3	Oracle Administration Tool (En español, Herramienta de Administración de Oracle) .72	
2.9	Encuesta	74
2.9.1	Instrumento de Recolección.....	74
2.9.2	Definición.....	74
2.9.3	Definición del Universo de Estudio y la Elección de la Muestra	75
2.9.4	El Universo de Estudio	76
2.9.5	Modalidades de Encuesta.....	77
2.9.6	El Cuestionario	82
2.9.7	Tratamiento y Análisis de la Información.....	90
2.9.8	Software y Herramientas para Procesar Encuestas	92
2.9.9	Encuestas e Inteligencia de Negocio.....	92

2.9.10	Modelización de la Encuesta	92
2.9.11	Sistema de Información de Encuesta Simple	94
CAPÍTULO 3: MARCO METODOLÓGICO		97
3.1	Método Ascendente (Ralph Kimball)	97
3.1.1	Ciclo de Vida Dimensional del Negocio	99
3.2	Método Descendente (Bill Inmon).....	102
3.3	Metodología Tian Consultores.....	103
CAPÍTULO 4: MARCO APLICATIVO		107
4.1	Contexto del Desarrollo.....	107
4.2	Proyecto.....	107
4.3	Fases del Proyecto	107
4.3.1	Análisis del Negocio.....	107
4.3.2	Desarrollo del Almacén de Datos	119
4.3.3	Desarrollo de Requerimientos de Información	131
CONCLUSIONES		150
REFERENCIAS BIBLIOGRÁFICAS Y DIGITALES.....		152

ÍNDICE DE TABLAS

Tabla 1: Diferencias entre OLTP y OLAP	28
Tabla 2: Tabla Empleado	32
Tabla 3: Tablas Empleado y Departamento.....	33
Tabla 4: Comparación entre ROLAP y MOLAP.....	56
Tabla 5: Cuadro Comparativo de las Herramientas de BI Analizadas	68
Tabla 6: Ventajas y Desventajas de la Encuesta Capturada por medio del Papel	78
Tabla 7: Ventajas y Desventajas de las Encuestas Capturadas vía Telefónica.....	79
Tabla 8: Ventajas y Desventajas de las Encuestas en Línea	81
Tabla 9: Dimensión Encuestado	113
Tabla 10: Dimensión Ruta.....	114
Tabla 11: Dimensión Dispositivo	114
Tabla 12: Dimensión Encuestador.....	115
Tabla 13: Dimensión Medición	116
Tabla 14: Dimensión Tiempo	117

ÍNDICE DE ILUSTRACIONES

Ilustración 1: Arquitectura propuesta para la Solución de Inteligencia de Negocio.....	18
Ilustración 2: Relación Datos, Información, Conocimiento y Sabiduría.....	21
Ilustración 3: Tipos de Sistemas de Información.....	23
Ilustración 4: Modelo Multidimensional	34
Ilustración 5: Esquema Estrella	36
Ilustración 6: Esquema Copo de Nieve	37
Ilustración 7: Esquema Constelación.....	39
Ilustración 8: Características de un Almacén de Datos.....	42
Ilustración 9: Cubo Multidimensional	46
Ilustración 10: Jerarquía	50
Ilustración 11: Arquitectura ROLAP.....	53
Ilustración 12: Arquitectura MOLAP	55
Ilustración 13: Arquitectura HOLAP.....	57
Ilustración 14: Arquitectura del Almacén de Datos	62
Ilustración 15: Niveles Jerárquicos de la Organización	64
Ilustración 16: Oracle Business Intelligence.....	70
Ilustración 17: Oracle Warehouse Builder.....	72
Ilustración 18: Oracle Administration Tool.....	73
Ilustración 19: Relación entre Población y Muestra	75
Ilustración 20: Flujo de Actividades del Proceso de Encuesta Simple.....	94
Ilustración 21: Sistema de Información Encuesta Simple.....	95
Ilustración 22: Flujo de Actividades del Proceso de Encuesta Simple.....	96
Ilustración 23: Método Ascendente	99
Ilustración 24: Ciclo de Vida Dimensional del Negocio.....	100
Ilustración 25: Método Descendente	102

Ilustración 26: Modelo de Desarrollo de Sistemas de Información de Inteligencia de Negocios de Tian Consultores	106
Ilustración 27. Modelo Relacional del Ambiente Transaccional (Encuesta Simple)	109
Ilustración 28: Granularidad.....	110
Ilustración 29: Modelo Dimensional	111
Ilustración 30: Definición de Dimensiones	112
Ilustración 31: Centro de Diseño, Oracle Warehouse Builder	120
Ilustración 32: Asistente para la Creación de Dimensiones	121
Ilustración 33: Asistente para la Creación de Dimensiones: Paso 2.....	121
Ilustración 34: Resumen de la Creación de una Dimensión.....	122
Ilustración 35: Diseño Lógico de las Dimensiones y el Cubo con Oracle Warehouse Builder	123
Ilustración 36: Centro de Diseño OWB.....	124
Ilustración 37: Paleta de Funcionalidades de OWB.....	125
Ilustración 38: Ejemplo Disposición de Tablas para Proceso ETC.....	125
Ilustración 39: Proceso ETC Final para Dimensión Ruta	126
Ilustración 40: Consulta Tabla Encuestado, Dimensión Encuestado	128
Ilustración 41: Consulta Conteo de Tabla Encuestado, Dimensión Encuestado	128
Ilustración 42: Consulta Tabla Encuestador, Dimensión Encuestador.....	129
Ilustración 43: Consulta tabla Dispositivo, Dimensión Dispositivo	130
Ilustración 44: Consulta tablas Ruta y Región, Dimensión Ruta	131
Ilustración 45: Oracle Administration Tool.....	132
Ilustración 46: Capa Física	132
Ilustración 47: Capa de Modelo de Negocio y Asignación.....	133
Ilustración 48: Capa de Presentación.....	133
Ilustración 49: Cuadro de Mando: Inicio	134
Ilustración 50: Cuadro de Mando: Encuestas	135
Ilustración 51: Cantidad de Respuestas por Formulario	135

Ilustración 52: Porcentaje de Cantidad de Respuestas por Pregunta y Formulario	136
Ilustración 53: Cantidad de Encuestados por Formulario, Año y Mes.....	136
Ilustración 54: Cantidad de Encuestados por Año y Mes	137
Ilustración 55: Cuadro de Mando: Encuestador.....	137
Ilustración 56: Cantidad de Encuestados Total y por Región	138
Ilustración 57: Gráfico Cantidad de Encuestados por Región	138
Ilustración 58: Cantidad de Encuestados en Total por Encuestador	139
Ilustración 59: Cantidad de Encuestados por Encuestador y por Año	139
Ilustración 60: Cantidad de Encuestados Total por Encuestador y Año	140
Ilustración 61: Cuadro de Mando: Encuestado.....	140
Ilustración 62: Listado de Encuestados	141
Ilustración 63: Encuestados por Región y Ruta	142
Ilustración 64: Encuestados por Encuestador	143
Ilustración 65: Encuestados por Formulario	144
Ilustración 66: Cuadro de Mando: Ruta.....	145
Ilustración 67: Cantidad de Encuestados por Región y Ruta	145
Ilustración 68: Gráfico Cantidad de Encuestados por Región y Ruta	146
Ilustración 69: Cantidad de Encuestados por Región Ruta y Año	146
Ilustración 70: Gráfico Cantidad de Encuestados por Región Ruta y Año	147
Ilustración 71: Cantidad de Encuestadores por Región y Ruta	147
Ilustración 72: Gráfico Cantidad de Encuestadores por Región Ruta	148
Ilustración 73: Dispositivos y su Encuestador Asociado	148
Ilustración 74: Comparación del Porcentaje del Número de Dispositivos por Marca.	149

INTRODUCCIÓN

Tomar mejores decisiones de negocios es la clave para el éxito en el mercado competitivo que se vive actualmente. Este hecho en la actualidad es conocido por las organizaciones, por lo que están buscando mejorar sus sistemas de toma de decisiones al verse rebasados por el volumen y complejidad de los datos provenientes de sus sistemas transaccionales.

Las empresas que desean expandir su mercado o mejorar su rendimiento en el sector en el que se desempeñan deben ser competitivas y obtener ventajas de todos sus recursos para asegurar la efectividad de sus estrategias y su correspondiente remuneración. Uno de los recursos más importantes con los que cuenta cualquier organización es la información. El análisis detallado y oportuno de la información de la empresa habilita a los ejecutivos tomar decisiones más adecuadas y efectivas, fundamentalmente reconocer oportunidades y posibilidades de crecimiento.

Una buena parte de los datos que disponen las empresas provienen de aplicaciones de procesamiento transaccional, que procesan de manera automática, grandes volúmenes de datos referentes a las actividades rutinarias de la organización y son almacenadas en bases de datos transaccionales. De ellas se puede extraer información, fundamentalmente válidas para las transacciones del día a día, que sirven para apoyar y ejecutar las decisiones operativas que conducen las actividades básicas, pero no sirven para realizar análisis más avanzados, incluso de tipo estratégico ya que no están diseñadas para apoyar este tipo de tareas. Es aquí donde un sistema de información de inteligencia de negocio viene a apoyar a las organizaciones.

Un proceso importante dentro de muchas organizaciones que expone las características descritas en los párrafos anteriores, es la obtención de datos del entorno a través de encuestas.

Las encuestas son una parte importante en los estudios de mercado pero con la llegada de nuevas tecnologías y nuevos conceptos, están cambiando mucho en cuanto al formato y su uso, surgiendo con esto distintos tipos de encuesta como lo es la encuesta en línea. Cada día se está extendiendo más la realización de este tipo de encuestas debido a que es una manera más rápida y económica de realizarlas además de evitar tener que diseñarlas en papel y tener que repartirlas después entre todos los encuestados.

El usuario puede crear la encuesta desde cero o utilizando ejemplos de encuestas existentes y enviarla mediante correo electrónico, colgarla en su Web, etc. dejando así la encuesta disponible para los usuarios que la vayan a rellenar. Los resultados también son más fáciles de interpretar y permiten reducir considerablemente los costes ya que es suficiente con realizar la encuesta en línea y enviarla mediante correo electrónico, Facebook, Twitter o incluso enlazar la encuesta desde su página Web.

Además, en todas aquellas operaciones estadísticas que implican el uso de cuestionarios es de gran utilidad tener una visión global de los resultados, lo que permite detectar posibles errores o incoherencias de un solo vistazo.

En el presente trabajo de investigación se realizó un estudio de los conceptos, metodologías y herramientas tecnológicas que permitieron desarrollar de una solución de Inteligencia de Negocios a partir de la base de datos transaccional del sistema de información de encuestas simples. Este tipo de solución, se aplica en aquellas organizaciones que requieren conocer con qué información cuentan antes de definir los requerimientos de información para la toma de decisiones.

CAPÍTULO 1. PROBLEMA DE INVESTIGACIÓN

1.1 Problema

Una parte importante de los datos que disponen las empresas provienen de aplicaciones de procesamiento transaccional, que procesan de manera automática, grandes volúmenes de datos referentes a las actividades rutinarias de la organización y son almacenadas en bases de datos transaccionales. Estas bases de datos contienen información útil para las transacciones del día a día, que sirven para apoyar y ejecutar las decisiones operativas que conducen las actividades básicas, pero no sirven para realizar análisis más avanzados que faciliten la toma de decisiones, ya que no están diseñadas para apoyar este tipo de análisis.

Para esta investigación se toma como fuentes de datos transaccionales el Sistema de Información de Encuesta Simple. Este sistema realiza encuestas móviles que permiten a las organizaciones medir la satisfacción de sus clientes con respecto a un producto o servicio. Los datos se almacenan en una base de datos transaccional, estos datos son difíciles de interpretar y no es posible tener una visión global de los resultados, dificultando con esto la toma de decisiones y la detección posibles errores o incoherencias en los datos o en la estructura del cuestionario.

De lo descrito anteriormente puede resumirse el problema con la siguiente pregunta:

¿Cómo obtener información relevante que apoye la toma de decisiones del proceso de levantamiento de encuestas simples a partir de su esquema transaccional?

1.2 Objetivo General

Desarrollar una solución de inteligencia de negocio que apoye la toma de decisiones del proceso de levantamiento de Encuestas Simples, a partir de los datos de su esquema transaccional.

1.3 Objetivos Específicos

- Identificar Modelo de Datos del Negocio.
- Diseñar el Modelo Dimensional y los requerimientos de información.
- Construir el almacén de datos, los procesos de extracción, transformación y carga, y los reportes e indicadores de control.
- Ejecutar procesos de verificación de calidad de datos.

1.4 Solución

Para resolver los problemas expuestos en el punto anterior se desarrollo una solución, basado en las mejores prácticas de la metodología Ascendente de Ralph Kimball y las mejores prácticas incorporadas por Tian Consultores.

En primera instancia la se realizó el análisis del Sistema de Información: Encuesta Simple, identificando todas las actividades que se llevan a cabo para tener conocimiento total de la lógica de negocios del sistema. A partir del Modelo Transaccional se procedió a realizar el modelado dimensional, siguiendo los pasos de la metodología seleccionada, cuya estructura estará diseñada para satisfacer las necesidades del negocio. A partir de este diseño lógico resultante se procedió al desarrollo de la presentación de datos que tiene como principales actividades la extracción, transformación y carga (ETC). Estas actividades son fundamentales ya que trabajan con los datos, que son la materia prima del Almacén de Datos. Este desarrollo se realizó mediante la herramienta Oracle Warehouse Builder.

A continuación se propone una arquitectura que tendrá un proceso de Extracción, Transformación y Carga importante: el primero se encargará de la extracción, transformación y carga de los datos fuentes al almacén de datos de modo tal que los datos estén preparados para el análisis.

Por ultimo para la explotación y publicación del Almacén de Datos se desarrollaron e implementaron de aplicaciones de usuario final, que involucra configuraciones de los metadatos y construcción de reportes específicos. Para esta fase se utilizó la herramienta Oracle Bussiness Intelligence.

La Ilustración 1, muestra gráficamente la arquitectura descrita anteriormente:

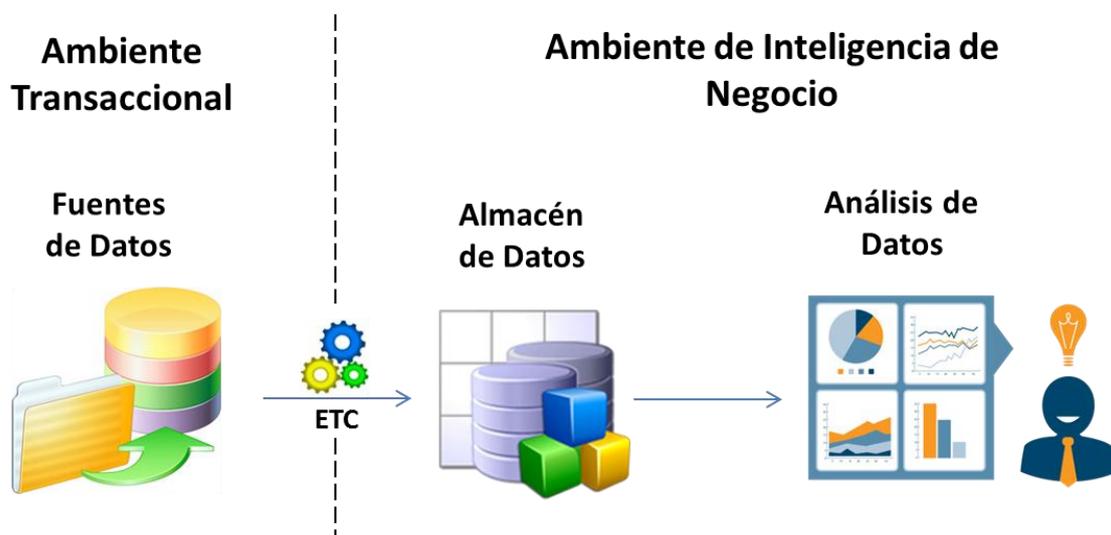


Ilustración 1: Arquitectura Propuesta para la Solución de Inteligencia de Negocio

Fuente: Manual de Desarrollo de Tian Consultores

1.5 Alcance

La solución de Inteligencia de Negocios desarrollada tiene como propósito que los usuarios asociados a las distintas organizaciones, puedan tener acceso rápido y estructurado a los datos relevantes de los resultados de las Encuestas Simples para que sirvan como soporte a la hora de tomar decisiones.

Por lo tanto el desarrollo de la solución de inteligencia de negocios se limitó al diseño y despliegue de los requerimientos asociados al proceso de levantamiento de Encuestas Simples, utilizando los datos del sistema transaccional como base para el llenado del almacén de datos.

CAPÍTULO 2: MARCO CONCEPTUAL

2.1 Conceptos Básicos

2.1.1 Dato

Los datos son un conjunto de hechos discretos y objetivos sobre acontecimientos, a los que describe, aunque no dicen nada sobre el porqué de las cosas y por sí mismos tienen poca o ninguna relevancia o propósito (Davenport & Prusak, 2000).

Según el DRAE, el concepto de dato en informática es “Información dispuesta de manera adecuada para su tratamiento por un ordenador”.

Los datos describen únicamente una parte de lo que pasa en la realidad y no proporcionan juicios de valor o interpretaciones, y por lo tanto, no son orientativos para la acción. La toma de decisiones se basa en datos, pero hace falta un juicio de valor para saber cuál es la alternativa más interesante (Alegre Vidal, 2004).

2.1.2 Información

La información es un conjunto organizado de datos procesados, que constituyen un mensaje que cambia el estado de conocimiento de la persona o sistema que recibe dicho mensaje.

Para Gilles Deleuze (1978), la información es un sistema de control, en tanto que es la propagación de consignas que deberíamos de creer. En tal sentido la información es un conjunto organizado de datos capaz de cambiar el estado de conocimiento.

Desde el punto de vista de la ciencia de la computación, la información es un conocimiento explícito extraído por seres vivos o sistemas expertos como resultado de interacción con el entorno o percepciones sensibles del mismo entorno. En principio la información, a diferencia de los datos o las percepciones sensibles, tienen estructura útil que modificará las interacciones del ente que posee dicha información con su entorno.

2.1.3 Conocimiento

El conocimiento es una mezcla de experiencias, información, habilidades y aptitudes que sirven como marco para la incorporación de nuevas experiencias e información que son útiles para llevar a cabo una acción (Davenport & Prusak, 2000).

En las organizaciones, el conocimiento se encuentra en rutinas organizativas, procesos, prácticas y normas institucionales (Nelson y Winter, 1982), además de encontrarse en documentos o almacenes de datos. El conocimiento relevante a las empresas incluye hechos, opiniones, ideas, teorías, principios, modelos, experiencias, valores, información contextual, percepciones de expertos e intuición.

Información se convierte en conocimiento una vez que se ha procesado en la mente de un individuo; y el conocimiento se vuelve a convertir en información cuando se articula o comunica a los demás por medio de un texto escrito, en formato electrónico, de forma oral o por otros medios (Alavi y Leidner, 1999).

2.1.4 Inteligencia

La inteligencia (del latín *intellegentia*) es la capacidad de pensar, entender, asimilar, elaborar información y utilizarla para resolver problemas. Es la habilidad que se tiene para analizar abstractamente un problema y la capacidad de organizar la información para un posterior razonamiento que permita elegir la mejor opción entre las que se nos brinda para resolver un problema.

2.1.5 Sabiduría

Es el nivel último del entendimiento. Cuando entendemos un abanico suficientemente amplio de patrones y de forma que los podemos utilizar y combinar de formas y situaciones nuevas totalmente diferentes. La sabiduría es, como el conocimiento, algo personal que se elabora íntimamente y que va con las personas y se pierde con ellas, a diferencia de los datos y la información. Es la utilización del conocimiento direccionado al mejoramiento de la calidad de decisiones.

En la Ilustración 2 se puede apreciar la relación entre las tres definiciones anteriores: Dato, Información, Conocimiento y Sabiduría:

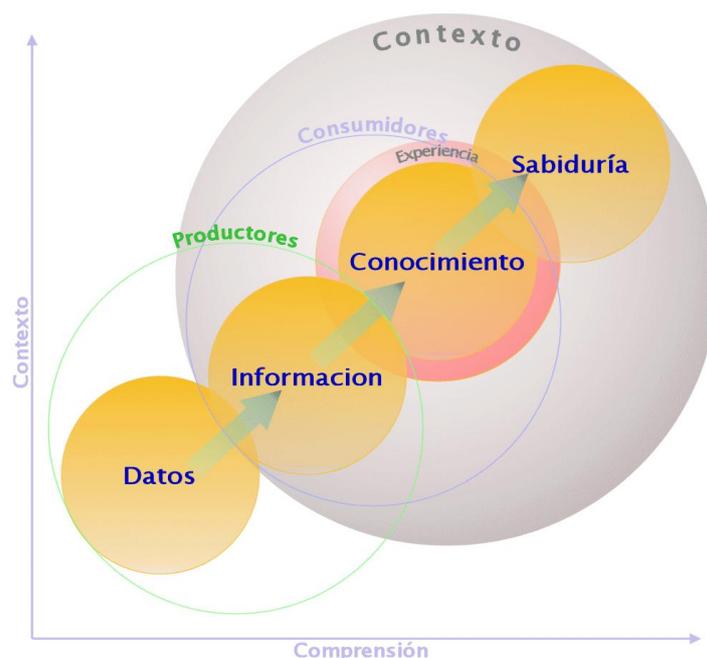


Ilustración 2: Relación Datos, Información, Conocimiento y Sabiduría

Fuente: <http://www.infovis.net/printMag.php?num=186&lang=1Sistema>

2.2 Sistema

Un sistema es un conjunto de partes o elementos organizados y relacionados que interactúan entre sí para lograr un objetivo.

Según Radatz (1997), un sistema es un todo integrado, aunque compuesto de diversas estructuras interactuantes y especializadas con un número de objetivos que pueden variar ampliamente de un sistema a otro.

Cada sistema existe dentro de otro más grande, por lo tanto un sistema puede estar formado por subsistemas y partes, y a la vez puede ser parte de un supersistema.

Los sistemas tienen límites o fronteras, lo que los diferencian del ambiente. El ambiente es el medio en externo que envuelve física o conceptualmente a un sistema. El sistema tiene interacción con el ambiente, del cual recibe entradas y al cual se le devuelve salidas.

Un grupo de elementos no constituye un sistema si no hay una relación e interacción, con un propósito definido.

2.3 Sistemas de Información

2.3.1 Definición

Según Peralta (2008), un sistema de información es un conjunto de elementos que interactúan entre sí con un único fin, el de apoyar todas las actividades de una empresa o negocio. Teniendo en cuenta el equipo computacional que se necesita para que el sistema de información pueda operar y un grupo de personas que es el recurso humano que interactúa con el sistema de información.

Un sistema de información es un conjunto de elementos que interactúan entre sí con el fin de apoyar las actividades de una empresa o negocio. También se puede definir como un conjunto de funciones o componentes interrelacionados que forman un todo, es decir, obtiene, procesa, almacena y distribuye información para apoyar la toma de decisiones y el control en una organización. Igualmente apoya la coordinación, análisis de problemas, visualización de aspectos complejos entre otros.

2.3.2 Actividades de los Sistemas de Información

Según Peralta (2008) los sistemas de información realizan cuatro actividades básicas: entrada, almacenamiento, procesamiento y salida de la información.

- **Entrada de Información:** proceso mediante el cual se toma los datos que requiere el sistema para procesar la información.
- **Almacenamiento de información:** es una de las actividades más importantes, ya que a través de esta propiedad el sistema puede recordar la información guardada en procesos anteriores.

- **Procesamiento de Información:** es la capacidad que poseen los sistemas de información para efectuar cálculos de acuerdo con una secuencia de operaciones preestablecida. Estos cálculos pueden efectuarse introduciendo datos al sistema o bien utilizando datos previamente almacenados. Esta característica permite la transformación de datos fuentes en información que puede ser utilizada para la toma de decisiones.
- **Salida de Información:** es la capacidad que poseen los sistemas de información para mostrar o presentar la información procesada al exterior. Es importante aclarar que la salida de un sistema de información puede constituir la entrada a otro sistema de información.

2.3.3 Clasificación de los Sistemas de Información

En cuanto a su naturaleza, pueden cerrados o abiertos:

- **Sistemas cerrados:** no presentan intercambio con el medio ambiente que los rodea. No reciben ningún recurso externo y nada producen que sea enviado hacia fuera. En rigor, no existen sistemas cerrados. Se da el nombre de sistema cerrado a aquellos sistemas cuyo comportamiento es determinístico y programado y que opera con muy pequeño intercambio de energía y materia con el ambiente.
- **Sistemas abiertos:** presentan intercambio con el ambiente, a través de entradas y salidas. Intercambian energía y materia con el ambiente. Son adaptativos para sobrevivir. Su estructura es óptima cuando el conjunto de elementos del sistema se organiza.

Según Peralta (2008), los sistemas de información, de manera general, se pueden clasificar de tres formas según sus propósitos generales: sistemas transaccionales, sistemas de soporte a la toma de decisiones y sistemas estratégicos. (Ver Ilustración 3)



Ilustración 3: Tipos de Sistemas de Información

Fuente: <http://www.udec.cl/dti/node/108>

2.3.3.1 Sistemas Transaccionales u Operacionales

Los Sistemas transaccionales son los encargados de recolectar, almacenar, modificar y recuperar toda información generada por las transacciones producidas en una organización, automatizando con esto los procesos operativos dentro de una organización, tales como pagos, cobros, pólizas, entradas, salidas, etc.

Se encuentran en el nivel más bajo de la pirámide, y no por ser menos importante, ya que representan la base de los siguientes niveles. Sin un sistema transaccional, los otros sistemas pierden su sentido.

Las características de los Sistemas Transaccionales u Operacionales son:

- Suelen lograrse ahorros de mano de obra, debido a que automatizan tareas operativas de la organización.
- Son el primer sistema de información que se implementa en la empresa.
- Son intensivos en entrada y salida de información, y sus cálculos y procesos suelen ser simples y poco sofisticados.
- Estos sistemas requieren de mucho manejo de datos para poder realizar sus operaciones y como resultado general también grandes volúmenes de información.
- Tienen la propiedad de ser recolectores de información, es decir, a través de estos sistemas se cargan las grandes bases de información para su explotación posterior.
- Son fáciles de ajustar ante la dirección general, ya que sus beneficios son visibles y palpables. En corto tiempo se puede evaluar los resultados y las ventajas que se tienen al implementarlo.

2.3.3.2 Sistemas de Soporte a la Toma de decisiones o Tácticos

Los Sistemas de soporte a la toma de decisión son herramientas de Inteligencia de Negocios enfocadas al análisis de los datos de una organización para tomar decisiones esporádicas situadas en el nivel estratégico de la organización.

Son sistemas de información utilizados normalmente por usuarios de rango organizacional más alto que los anteriores. Estos están diseñados como un conjunto de programas y herramientas que permiten obtener la información que se obtuvo de los sistemas transaccionales para ayudar al proceso de toma de decisiones.

Estos sistemas se incluyen con frecuencia luego de haber implantado los sistemas transaccionales, ya que estos últimos son su principal fuente de información.

Las características de los Sistemas de soporte a la toma de decisiones o tácticos son:

- Suelen introducirse después de haber implantado los Sistemas Transaccionales más relevantes de la empresa, ya que estos últimos constituyen su plataforma de información.
- La información que generan sirve de apoyo a los mandos intermedios y a la alta administración en el proceso de toma de decisiones.
- Suelen ser intensivos en cálculos y escasos en entradas y salidas de información.
- No suelen ahorrar mano de obra. Debido a ello, la justificación económica para el desarrollo de estos sistemas es difícil, ya que no se conocen los ingresos del proyecto de inversión.
- Suelen ser Sistemas de Información interactivos y amigables, con altos estándares de diseño gráfico y visual, ya que están dirigidos al usuario final.
- Apoyan la toma de decisiones que, por su misma naturaleza son repetitivos y de decisiones no estructuradas que no suelen repetirse.
- Estos sistemas pueden ser desarrollados directamente por el usuario final sin la participación operativa de los analistas y programadores del área de informática.

2.3.3.3 Sistemas Estratégicos

Son los sistemas de Información ubicados en el tope de la pirámide, estos sistemas se desarrollan en las organizaciones con el fin de lograr ventajas competitivas, a través del uso de la tecnología de información.

Ayudan en el proceso de mejora e innovación de productos y procesos dentro de las organizaciones buscando así ventajas sobre los principales competidores.

Las características de los Sistemas Estratégicos son:

Su función primordial no es apoyar la automatización de los procesos operativos ni proporcionar información para apoyar a la toma de decisiones. Sin embargo, este tipo de sistemas puede llevar a cabo dichas funciones.

Son utilizados solo por los altos ejecutivos de las organizaciones y suelen ser creados dentro de las organizaciones.

Cambian significativamente el desempeño de un negocio al medirse por uno o más indicadores clave, entre ellos, la magnitud del impacto.

Generan cambios fundamentales en la forma de dirigir una compañía, la forma en que compite o en la que interactúa con clientes y proveedores.

Su función es lograr ventajas que los competidores no posean, tales como ventajas en costos y servicios diferenciados con clientes y proveedores.

La Ilustración 3 presenta los distintos tipos de sistemas de información en la pirámide organizacional de una empresa:

2.4 Sistemas OLTP y OLAP

2.4.1 Sistemas OLTP

El nombre OLTP viene del acrónimo en inglés de Procesamiento de Transacciones en Línea. Es un tipo de sistema diseñado para admitir aplicaciones orientadas a transacciones. Los sistemas de OLTP se diseñan para responder inmediatamente a las solicitudes de usuario y están basados en bases de datos relacionales orientadas al procesamiento de transacciones.

Una transacción es una secuencia de operaciones realizadas como una sola unidad lógica de trabajo. Una unidad lógica de trabajo debe exhibir cuatro propiedades, conocidas como propiedades de atomicidad, coherencia, aislamiento y durabilidad (ACID), para ser calificada como transacción.

- **Atomicidad:** asegura que la operación se ha realizado correcta y completamente, de forma que no pueda quedar a medias en caso de que surja cualquier error.
- **Consistencia:** asegura la integridad de la base de datos, es decir, que solo se ejecutan aquellas operaciones que no van a romper las reglas, restricciones y claves foráneas. La propiedad de consistencia sostiene que cualquier transacción llevará a la base de datos desde un estado válido a otro también válido.
- **Aislamiento:** asegura que dos transacciones que afectan a la misma información (tabla, fila o celda) son independientes y no generan errores ni bloqueo mutuo.
- **Durabilidad:** asegura que una vez realizada la operación, ésta persistirá y no se podrá deshacer aunque falle el sistema. Será almacenada en disco, no solo en memoria.

Las características de los Sistemas OLTP son:

- **Son la versión tradicional de una base de datos:** se diseñan utilizando un modelo entidad-relación, se implementan en los motores típicos de base de datos (Oracle, SQLServer, MySQL, etc.) y dan soporte a la mayor parte del software del mercado.
- **Optimizadas para lecturas y escrituras concurrentes:** gracias a las propiedades ACID, el acceso a los datos está adaptado para tareas frecuentes de lectura y escritura.
- **Organizadas según la capa de aplicación:** las tablas y los datos se estructuran según el software que los maneja: programa de gestión a medida, ERP, CRM, BPM, etc.

- **Adaptadas a cada empresa o departamento:** dado que en muchas ocasiones se utiliza software no integrado, los formatos de los datos no suelen ser uniformes en los diferentes departamentos. Es común la falta de compatibilidad y la existencia de islas de datos.
- **Consultas realizadas en SQL, modificaciones en DML:** el SQL es el lenguaje de consulta universal para leer bases de datos relacionales, mientras que DML es el estándar para realizar modificaciones.
- **Gestión de datos históricos inexistente:** el historial de cambios suele limitarse a datos actuales o recientes. Salvo sistemas de respaldo, o que el software tenga una funcionalidad específica para ello, no se suelen manejar valores históricos para cada campo.

2.4.2 Sistemas OLAP

El nombre OLAP viene del acrónimo en inglés de Procesamiento Analítico en Línea. Es una solución utilizada en el campo de inteligencia de negocio cuyo objetivo es agilizar la consulta de grandes cantidades de datos con la finalidad de extraer algún tipo de información útil: tendencias de ventas, patrones de comportamiento de los consumidores, elaboración de informes complejos, etc.

Para representar la información, los sistemas OLAP utilizan los cubos OLAP, que no es más que un vector de varias dimensiones. Desde un punto de vista relacional, puede verse como una tabla de hechos que tiene dos tipos de columnas:

- **Indicadores:** también denominados métricas o ratios, son los valores numéricos con los que se opera. Por ejemplo: n° de clientes, n° de proveedores, importe de las ventas, n° de ventas, importe de las compras, n° de compras... etc.
- **Dimensiones:** son las características por las que se pueden filtrar y cruzar los indicadores. Por ejemplo: tiempo (fijando un determinado día, mes o año), geografía (fijando un determinado país, región o ciudad), proveedor, cliente, modo de pago... etc.

Las características de los Sistemas OLAP son:

- **Optimizadas para operaciones de lectura:** dado que la acción más común es la consulta, estas bases de datos disponen de valores agregados y resultados precalculados que les permiten responder en tiempo récord. Evitar las restricciones ACID les da agilidad.
- **Organizadas según las necesidades analíticas:** los datos están estructurados según las áreas de negocio, y los formatos de los datos están integrados de manera uniforme en toda la organización. Se busca evitar islas de datos.

- **Asíncronas:** no siempre se actualizan en tiempo real, sino que se suelen alimentar con información procedente de las bases de datos relacionales mediante un proceso de extracción, transformación y carga (ETC).
- **Gestión de datos históricos a largo plazo:** una de las exigencias analíticas consiste en realizar estudios de evolución a lo largo del tiempo, esto requiere que estas bases de datos mantengan un histórico a largo plazo, normalmente no inferior a cinco años.

2.4.3 Diferencias entre OLTP y OLAP

Existen diferencias notables entre los sistemas transaccionales (OLTP) y los sistemas analíticos (OLAP) que se presentan en la Tabla 1:

Características	OLTP	OLAP
Objetivos Principales	Asistir a aplicaciones específicas y garantizar la integridad y consistencia de los datos.	Consolidar los datos ya validados y adecuados a las necesidades para la toma de decisiones.
Orientación	Orientado a la aplicación, hace cumplir las reglas del negocio.	Orientado al sujeto, se define en base a lo que el analista necesita ver.
Vigencia de los datos	Los datos se usan a medida que se van produciendo y dejan de ser importantes a corto plazo.	Se guardan los datos actuales y los históricos para poder realizar análisis comparativos.
Granularidad de datos	Es dada por controles definidos por la organización, así como por normas legales importantes.	Estará dado por el tipo de análisis que se requiera analizar.
Organización	Es normalizado.	Se basa en estructuras jerárquicas desnormalizadas modeladas de acuerdo a como se analizaran los datos.
Cambios en los datos	Modifica sus datos en forma constante porque maneja las transacciones de la empresa.	No tiene como objetivo la presentación de los datos en línea ni tampoco modificar datos originales, solo consultarlos.
Acceso y manipulación de los datos	Realizan una manipulación de datos registro por registro con grandes cantidades de inserts, updates y deletes. Además necesitan de rutinas de validación y transacciones a nivel de registro.	Los DWS tiene una carga y acceso masivo de datos, no se realizan inserts, updates o deletes. La validación de datos o después de la carga.
Integración de datos	Los datos se encuentran típicamente no integrados, son calificados como datos primitivos o datos operacionales	Los datos deben estar integrados. Son conocidos como datos derivados o datos DSS dado que provienen de sistemas transaccionales o sistemas de archivos maestros preexistentes en las mismas organizaciones.

Tabla 1: Diferencias entre OLTP y OLAP

Fuente: <http://bddatawarehouseyessy.wordpress.com/2012/02/14/olap-y-oltp/>

Ambos sistemas se complementan, y aunque pareciera que una empresa puede sobrevivir únicamente con sistemas transaccionales, si no cuenta con un correcto sistema analítico estará basando la mayor parte de sus decisiones en la experiencia y en suposiciones que no siempre son válidas, desperdiciando el que es, muy probablemente, el recurso más importante de una organización: su información.

2.5 Base de Datos

2.5.1 Definición

Según Di vasta & Díaz (2001), una base de datos es un repositorio centralizado de datos lógicamente relacionados, que permite almacenar y organizar hechos o eventos y restituirlos a demanda de él, o los usuarios para producir información.

Una Base de Datos consolida muchos registros almacenados previamente en archivos independientes, de modo que un cúmulo (pool) común de registros sirvan como una sola central para muchas aplicaciones de procesos que necesitan este tipo de datos.

Normalmente el número de campos (columnas) que se pueden tener en una base varía según las necesidades en cuanto a gestión de datos, de forma que después se pueda explotar la información de forma ordenada y separada, aunque el resto de la información sigue almacenada en la base de datos.

Aparte de los datos que son almacenados en el archivo, también hay una serie de datos, en los que se informa del tipo de campo, los campos, la longitud de cada campo, entre otros. A estos datos que describen otros datos se le conoce como metadato.

En una base de datos podemos introducir, entre otros, los siguientes tipos de datos:

- **Vídeo:** secuencias de imágenes.
- **Gráficos:** archivos de imagen que pueden modificar la base de datos.
- **Sonido:** archivos de audio.
- **Lógicos:** tratan de establecer relaciones entre los datos.
- **Numéricos.**
- **Alfanuméricos.**

Los sistemas de información deben contar con la información que represente valor para las empresas, para esto debe tener acceso a las herramientas y mecanismos claves para poder obtener la información rápida y eficientemente. Por consiguiente, las bases de datos representan un componente fundamental para los sistemas de información que requieren manejar y almacenar pequeños o grandes volúmenes de información.

2.5.2 Características de una Base de Datos

Según Moreno A. (2000) podemos decir que las principales características de una Base de Datos son las siguientes:

- **Independencia de datos:** es el concepto que más ha ayudado a la rápida proliferación del desarrollo de Sistemas de Bases de Datos. La independencia de datos implica un divorcio entre programas y datos.
- **Integridad en la información:** la integridad de la base de datos se refiere a la validez y la consistencia de los datos almacenados. Es una de sus principales cualidades y tiene por objetivo que sólo se almacena la información correcta.
- **Redundancia mínima:** se trata de usar la base de datos como repositorio común de datos para distintas aplicaciones, evitando almacenar los mismos datos varias veces.
- **Globalización de la información:** permite a los diferentes usuarios considerar la información como un recurso corporativo que carece de dueños específicos.
- **Acceso concurrente por parte de múltiples usuarios:** la base de datos pertenece a la empresa y puede ser compartida por todos los usuarios que estén autorizados.
- **Consistencia de datos:** si existen dos o más archivos con la misma información, los cambios que se hagan a éstos deberán hacerse a todas las copias del archivo.
- **Consultas complejas optimizadas:** la optimización de consultas permite la rápida ejecución de las mismas.
- **Seguridad de acceso y auditoria:** se refiere al derecho de acceso a los datos contenidos en la base de datos por parte de personas y organismos. El sistema mantiene el control de acceso a la base de datos, con el objeto de saber qué o quién realizó una determinada modificación y en qué momento.
- **Acceso a través de lenguajes de programación estándar:** se refiere a la posibilidad ya mencionada de acceder a los datos de una base de datos mediante lenguajes de programación ajenos al sistema de base de datos propiamente dicho.

2.5.3 Sistema Manejador de Base de Datos

El sistema manejador de bases de datos se encarga de manejar la creación y todos los accesos a las bases de datos.

Estos sistemas tienen la capacidad de crear y gestionar las bases de datos, manejar el control de accesos a la información y brindar herramientas para la manipulación de los datos de acuerdo a las necesidades de los usuarios.

Según Silberschatz (2002) es un conjunto de programas, procedimientos, lenguajes, entre otros. Que suministra a los usuarios no informáticos, analistas, programadores, administradores de base de datos, los medios necesarios para describir, recuperar y manipular los datos almacenados en la base de datos, permitiendo mantener su integridad, confidencialidad y seguridad.

Entre las principales funciones de los SDBD se tienen:

- Almacenar, recuperar y actualizar los datos en el disco.
- Manejar la integridad y consistencia de los datos cuando estos son actualizados.
- Coordinar la concurrencia de acceso de los diferentes usuarios a la base de datos, garantizando que los datos no se dañen.

Un sistema manejador de base de datos es esencial para poder manipular y darle un adecuado funcionamiento a una base de datos.

2.5.4 Modelos de Bases de Datos

Un modelo de Base de Datos es un tipo de modelo que determina la estructura lógica de una base de datos, determinando con esto el modo de almacenar, organizar y manipular los datos. Los modelos de datos no son estructuras físicas: son abstracciones que permiten la implementación de un sistema eficiente de base de datos.

Algunos modelos con frecuencia utilizados en las bases de datos:

- Modelo jerárquico
- Modelo en red
- Modelo relacional
- Modelo orientado a objetos
- Modelo multidimensional

A continuación se describen los modelos de bases de datos Relacional y Multidimensional que son los que se adaptan para el contenido de este trabajo.

2.5.4.1 Modelo Relacional

El modelo relacional constituye una alternativa para la organización y representación de la información que se pretende almacenar en una base de datos. Se trata de un modelo teórico

matemático que, además de proporcionarnos los elementos básicos de modelado (las relaciones), incluye un conjunto de operadores (definidos en forma de un álgebra relacional) para su manipulación, sin ambigüedad posible (Ramez Elmasri, 2002).

En el modelo relacional se basa en el concepto matemático de relación. En este modelo, la información se representa en forma de “tablas” o relaciones, donde cada fila de la tabla se interpreta como una relación ordenada de valores (un conjunto de valores relacionados entre sí).

Características de las base de datos relacionales:

- Está compuesta por varias relaciones y varias tablas.
- No existen dos tablas con el mismo nombre.
- Cada tabla está compuesta por un conjunto de filas y columnas.
- La relación entre una tabla padre y una tabla hijo se lleva a cabo por medio de una clave primaria y una clave foránea.
- La clave primaria es la clave principal de un registro dentro de una tabla cumpliendo con la integridad de datos.
- La clave foránea se coloca en las tablas hijas, contiene el mismo valor que la clave primaria de la tabla padre, así es que se hacen las relaciones entre tablas.
- La estructura que tiene una base de dato relacional está basada en el esquema y los datos.

Para entender un poco más este concepto analicemos el siguiente ejemplo donde tenemos una tabla con los empleados de una organización. Esta tabla de empleados contiene el código del departamento y el nombre del departamento comenzará a repetir la información para los empleados que pertenezcan al mismo departamento (Ver Tabla 2: Tabla Empleado).

Nombre	Apellido	ID_Depto	Departamento
Simón	Tribiño	5	Informática
Víctor	Vargas	5	Informática
María	Gómez	5	Informática
Carlos	Pérez	2	Estadística
David	García	2	Estadística

Tabla 2: Tabla Empleado

Fuente: Elaboración Propia

Si el nombre del departamento cambiara de Informática a Tecnología, tendríamos que ir registro por registro modificando el nombre correspondiente y podríamos dejar alguna incongruencia en los datos. Una mejor solución es tener una tabla exclusiva de departamentos y solamente incluir la clave del departamento en la tabla de empleados (Ver Tabla 3).

De esta manera dejamos de repetir el nombre del departamento en la tabla de empleados y ahorramos espacios de almacenamiento. Y en caso de un cambio de nombre de departamento solamente debemos realizar la actualización en un solo lugar.

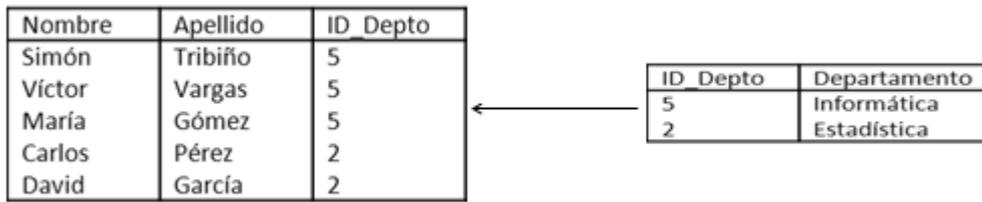


Tabla 3: Tablas Empleado y Departamento

Fuente: Elaboración Propia

2.5.4.2 Modelo Multidimensional

Según el Ralph Kimball (1997), es una técnica de diseño de bases de datos destinadas a apoyar a las consultas de los usuarios finales en un almacén de datos. Está orientado a la comprensibilidad y rendimiento. Según él, a pesar de que el modelo Entidad-Relación es de gran utilidad para la captura de transacciones, se debe evitar presentar al usuario final.

Por lo tanto el Modelo Multidimensional busca presentar la información de una manera estándar, sencilla y sobre todo intuitiva para los usuarios finales, además de permitir acceso rápido a la información por parte de los manejadores de bases de datos.

Cada Modelo Multidimensional está compuesto por una tabla llamada "tabla de hechos" y por un conjunto de tablas llamadas "dimensiones" (Ver Ilustración 4). Cada dimensión contiene una clave primaria que se relaciona con la tabla de hechos manteniendo una relación de 1 a muchos (1 a N).

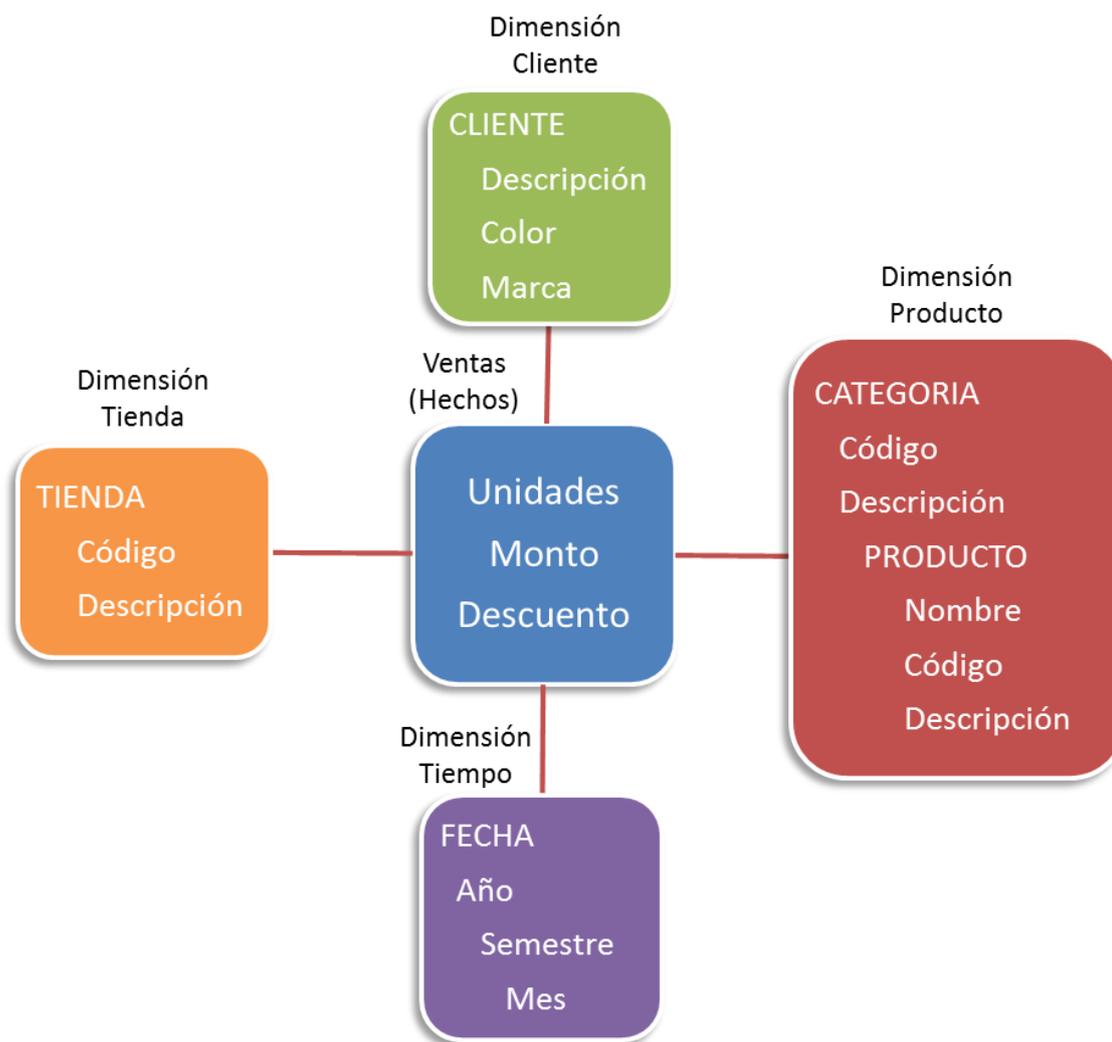


Ilustración 4: Modelo Multidimensional

Fuente: Elaboración Propia

La tabla de hechos contiene un conjunto de claves foráneas, cada una de ellas relacionada a una dimensión específica, es por eso que contiene una clave primaria compuesta, es decir, que se compone de varias columnas. Además de las claves foráneas de las dimensiones las tablas de hecho también incluyen otras columnas numéricas llamadas "métricas" o "hechos", por ejemplo en la ilustración anterior tenemos dos métricas: "Ventas en Pesos" y "Ventas en Cantidad".

La información contenida en una tabla de hechos por sí sola no nos dice nada, la mayoría de las veces contiene únicamente números, en cambio las dimensiones si contienen información descriptiva. Los atributos de las dimensiones muchas veces serán los encabezados de las columnas en los reportes.

Gracias a este modelo los usuarios finales pueden intuir fácilmente el contenido de la información y a su vez permite a las aplicaciones de inteligencia de negocios extraer la información rápida y eficientemente.

Entonces las bases de datos multidimensionales son una variación del modelo relacional que permite organizar los datos y expresar las relaciones entre ellos. Las principales ventajas de este tipo de bases de datos son la versatilidad para cruzar información y la alta velocidad de respuesta. Esto las convierte en herramientas básicas para soluciones de Inteligencia de Negocio, donde el análisis de los datos resulta crucial.

Los conceptos del modelado multidimensional se aplican a los modelos de datos lógicos y físicos. Un esquema multidimensional físico generalmente se representa en forma de esquema de estrella o de copo de nieve, en el que los objetos que contiene son en realidad tablas de base de datos. El esquema multidimensional puede incluso adoptar la forma de una sola tabla o vista, en la que todos los hechos y dimensiones están en columnas distintas de dicha tabla o vista. En un esquema dimensional lógico, los hechos, las medidas y las dimensiones se representan como entidades y atributos independientes a un proveedor de base de datos y, por lo tanto, se pueden transformar en un esquema dimensional físico para cualquier proveedor de base de datos.

Existen tres esquemas principales para llevar a cabo el diseño multidimensional de la base de datos del Almacén de Datos: Esquema de Estrella, Esquema de Copo de Nieve y Esquema de Constelación. Estos esquemas son usados para soportar operaciones de datos multidimensionales

2.5.4.3 Esquemas Estrella

Un esquema de estrella, se le conoce como el esquema tradicional y de este se derivan los demás esquemas conocidos. Su nombre se debe a que el modelado es representado como una estrella, en el centro de la estrella encontramos una tabla de hechos y una o varias tablas de dimensión.

En el esquema en estrella la tabla de hechos es la única tabla del esquema que tiene múltiples uniones que la conectan con otras tablas. El resto de tablas del esquema únicamente hacen uniones con esta tabla de hechos. Las tablas de dimensión se encuentran además totalmente denormalizadas, es decir, toda la información referente a una dimensión se almacena en la misma tabla. (Ver Ilustración 5)

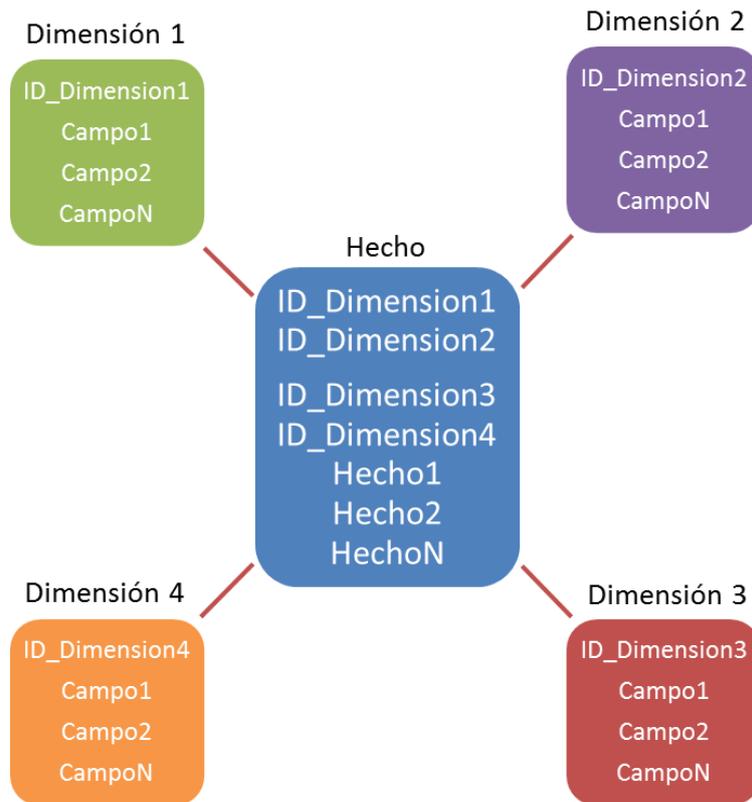


Ilustración 5: Esquema Estrella

Fuente: Elaboración Propia

Cuando se normaliza, se pretende eliminar la redundancia, la repetición de datos y que las claves sean independientes de las columnas, pero en este tipo de modelos se requiere no evitar precisamente esto.

Las ventajas que trae aparejada la desnormalización, son las de obviar uniones entre las tablas cuando se realizan consultas, procurando así un mejor tiempo de respuesta y una mayor sencillez con respecto a su utilización. El punto en contra, es que se genera un cierto grado de redundancia, pero el ahorro de espacio no es significativo.

El esquema en estrella es el más simple de interpretar y optimiza los tiempos de respuesta ante las consultas de los usuarios. Este modelo es soportado por casi todas las herramientas de consulta y análisis, y los metadatos son fáciles de documentar y mantener, sin embargo es el menos robusto para la carga y es el más lento de construir.

A continuación se destacarán algunas características de este modelo, que ayudarán a comprender mejor el porqué de sus ventajas:

- Posee los mejores tiempos de respuesta.
- Su diseño es fácilmente modificable.

- Existe paralelismo entre su diseño y la forma en que los usuarios visualizan y manipulan los datos.
- Simplifica el análisis.
- Facilita la interacción con herramientas de consulta y análisis.

2.5.4.4 Esquemas de Copo de Nieve

El esquema en copo de nieve es un esquema de representación derivado o extensión del esquema en estrella, en el que las tablas de dimensión se normalizan en múltiples tablas. Por esta razón, la tabla de hechos deja de ser la única tabla del esquema que se relaciona con otras tablas, y aparecen nuevas uniones gracias a que las dimensiones de análisis se representan ahora en tablas de dimensión normalizadas. En la estructura dimensional normalizada, la tabla que representa el nivel base de la dimensión es la que se une directamente con la tabla de hechos. (Ver Ilustración 6)

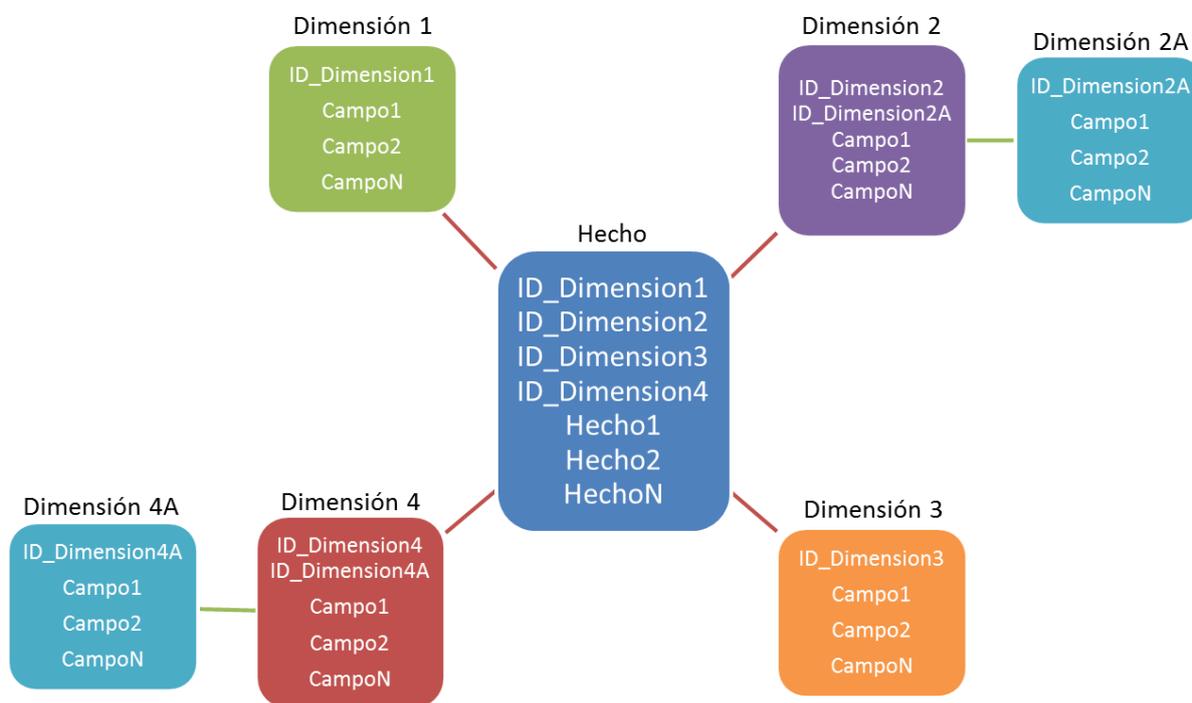


Ilustración 6: Esquema Copo de Nieve

Fuente: Elaboración Propia

La diferencia entre ambos esquemas reside entonces en la estructura de las tablas de dimensión. Para conseguir un esquema en copo de nieve se ha de tomar un esquema en estrella y conservar la tabla de hechos, centrándose únicamente en el modelado de las tablas de dimensión, que si bien en el esquema en estrella se encontraban totalmente denormalizadas, ahora se dividen en subtablas tras un proceso de normalización.

Este modelo es más cercano a un modelo de entidad relación, que al modelo en estrella, debido a que sus tablas de dimensiones están normalizadas.

Una de los motivos principales de utilizar este tipo de modelo, es la posibilidad de segregar los datos de las tablas de dimensiones y proveer un esquema que sustente los requerimientos de diseño. Otra razón es que es muy flexible y puede implementarse después de que se haya desarrollado un esquema en estrella.

Se pueden definir las siguientes características de este tipo de modelo:

- Posee mayor complejidad en su estructura.
- Hace una mejor utilización del espacio.
- Es muy útil en tablas de dimensiones de muchas tuplas.
- Las tablas de dimensiones están normalizadas, por lo que requiere menos esfuerzo de diseño.
- Puede desarrollar clases de jerarquías fuera de las tablas de dimensiones, que permiten realizar análisis de lo general a lo detallado y viceversa.

A pesar de todas las características y ventajas que trae aparejada la implementación del esquema copo de nieve, existen dos grandes inconvenientes de ello:

- Si se poseen múltiples tablas de dimensiones, cada una de ellas con varias jerarquías, se creará un número de tablas bastante considerable, que pueden llegar al punto de ser inmanejables.
- Al existir muchas uniones y relaciones entre tablas, el desempeño puede verse reducido.

La existencia de las diferentes jerarquías de dimensiones debe estar bien fundamentada, ya que de otro modo las consultas demorarán más tiempo en devolver los resultados, debido a que se deben realizar las uniones entre las tablas.

2.5.4.5 Esquemas de Constelación

Un esquema de constelación es una combinación de un esquema de estrella y un esquema de copo de nieve. Los esquemas de constelación son esquemas de copo de nieve en los que sólo algunas de las tablas de dimensiones se han desnormalizado.

Este modelo está compuesto por una serie de esquemas en estrella, y tal como se puede apreciar en la Ilustración 7, está formado por una tabla de hechos principal (“HECHOS_A”) y por una o más tablas de hechos auxiliares (“HECHOS_B”). Dichas tablas están ubicadas en el centro del modelo y están relacionadas con sus respectivas tablas de dimensiones.

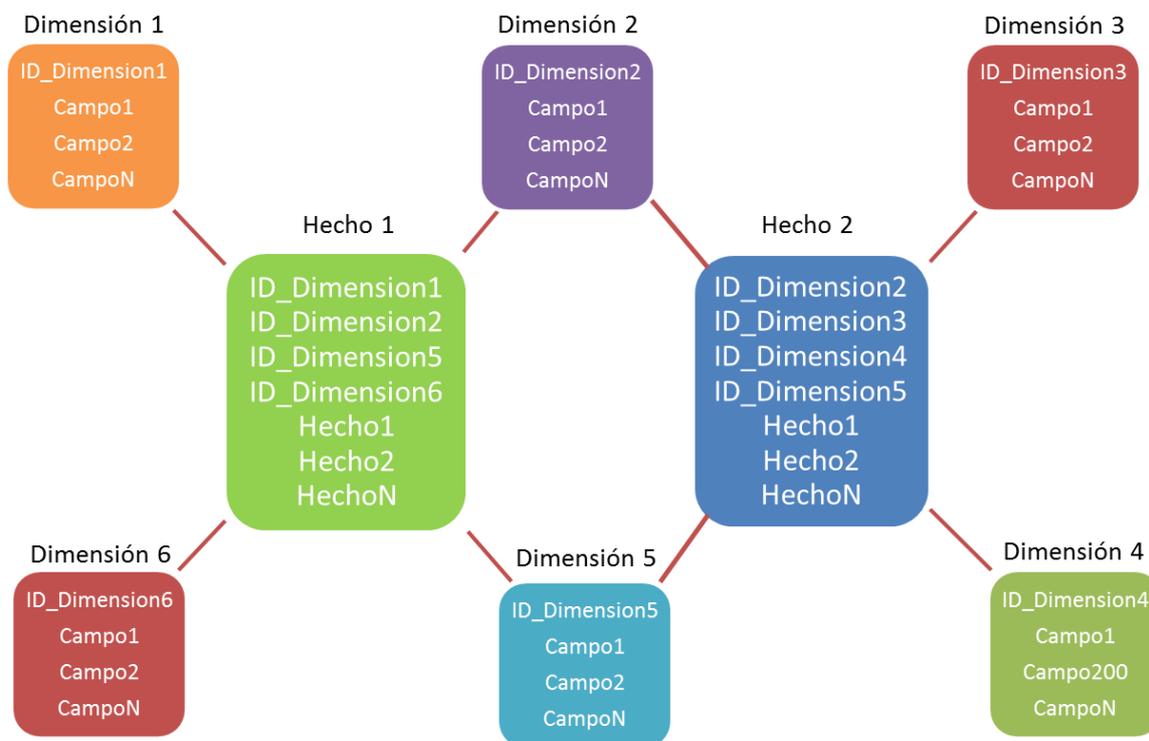


Ilustración 7: Esquema Constelación

Fuente: <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/arquitectura-del-data-warehouse/34-Almacén de Datos-manager>

No es necesario que las diferentes tablas de hechos compartan las mismas tablas de dimensiones, ya que, las tablas de hechos auxiliares pueden vincularse con solo algunas de las tablas de dimensiones asignadas a la tabla de hechos principal, y también pueden hacerlo con nuevas tablas de dimensiones.

Su diseño y cualidades son muy similares a las del esquema en estrella, pero posee una serie de diferencias con el mismo, que son precisamente las que lo destacan y caracterizan. Entre ellas se pueden mencionar:

- Permite tener más de una tabla de hechos, por lo cual se podrán analizar más aspectos claves del negocio con un mínimo esfuerzo adicional de diseño.
- Contribuye a la reutilización de las tablas de dimensiones, ya que una misma tabla de dimensión puede utilizarse para varias tablas de hechos.
- No es soportado por todas las herramientas de consulta y análisis.

2.6 Almacén de Datos

2.6.1 Introducción

Mejores decisiones de negocios es la clave para el éxito en el mercado empresarial hoy en día. Las organizaciones buscan que sus tomas de decisiones sean entendibles tomando en cuenta el volumen y complejidad de los datos disponibles en los sistemas operacionales y de producción. Facilitar estos datos a distintos usuarios con distintas necesidades de datos, es uno de los cambios más significativos para los profesionales de la información.

Muchas organizaciones escogen construir un Almacén de Datos para liberar la información en sus sistemas operativos y entender los verdaderos problemas del mundo de los negocios. El Almacén de Datos es un almacenamiento integrado de información obtenido de otros sistemas, que provee soporte de decisiones, disponibilidad y manipulación de información para el análisis de datos.

El Almacén de Datos representa el proceso de reunir la información histórica de una organización en un depósito central, y se ha convertido en una tecnología común y fundamental. Es la tendencia más grande dentro de la administración de información en la actualidad.

El proceso Almacén de Datos es complejo, costoso y que consume tiempo. Sin embargo, con el paso de los años se ha venido trabajando para crear una plataforma consistente, que permita reducir los costos e impulsar a las empresas a crear, administrar y utilizar un Almacén de Datos.

2.6.2 Antecedentes Históricos

El concepto de Almacén de Datos proviene de la combinación de dos necesidades que generalmente no están asociadas, sin embargo son tomadas en conjunto para obtener un mejor entendimiento del problema y presentar una posible solución. Estas necesidades son:

- Requerimiento del negocio, de una perspectiva amplia de la información de la compañía.
- La necesidad del manejo de la información por parte del departamento de sistemas.

Por separado, la demanda de los negocios requiere una vista amplia de la información que pueda liderar soluciones basadas en permitir a cualquier usuario acceder cualquier información sin importar donde esté ubicado. Sin embargo, estas soluciones son simples, porque ignoran fundamentalmente la distinción entre datos e información. De hecho, lo que los usuarios del negocio requieren es información.

Sin embargo, combinando las necesidades anteriormente mencionadas, se da una nueva perspectiva. Si la necesidad del sistema de información fuera tomada en cuenta, la necesidad de los negocios de una consulta amplia de los datos de los negocios es más fácil de reunir.

De igual manera la necesidad de obtener una consulta amplia de los datos y los obvios beneficios del negocio que estos tienen, son la justificación requerida para resolver el problema de administración de datos.

A mitad de los años noventa el Almacén de Datos se inició como una moda en la industria de la computación, se ha anunciado el valor del concepto por más de una década, tal popularidad trae sus propios problemas. Uno de estos es que en sus inicios, lleva a los vendedores a adaptar sus conceptos particulares para hacer una unión entre éste y los productos que ellos venden. Debe recordarse que el Almacén de Datos no puede ser manejado por sí sólo. Este se inventó por las compañías para satisfacer sus propias necesidades, éstas continúan existiendo y siguen creciendo de acuerdo al ambiente tecnológico en que los negocios se operen y sean más complejos.

2.6.3 Definición

Un almacén de datos es un repositorio de datos históricos integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones. Es una estructura de datos donde la información contenida está diseñada para favorecer el análisis y la divulgación eficiente de datos. Los almacenes de datos contienen a menudo grandes cantidades de información, que generalmente se subdividen en unidades lógicas más pequeñas, que dependen de la entidad de la que provengan. Dichas unidades se denominan Bodegas de Datos.

Es un repositorio de datos de muy fácil acceso, alimentado de numerosas fuentes, transformadas en grupos de información sobre temas específicos de negocios, para permitir nuevas consultas, análisis, reportes y decisiones.

Existen dos investigadores cuyos trabajos son muy difundidos con respecto al tema Almacén de Datos: Bill Inmon y Ralph Kimball.

Para Ralph Kimball (1997) el Almacén de Datos es una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis; es la unión de todos las Bodegas de Datos de una entidad".

Para Bill Inmon (1992), quien acuñó el término por primera vez, el Almacén de Datos es una colección de datos orientados al tema, integrados, no volátiles e históricos, organizados para el apoyo de un proceso de ayuda a la decisión. No obstante, y como cabe suponer, es mucho más que eso. (Ver Ilustración 8)



Ilustración 8: Características de un Almacén de Datos

Fuente: 1992. Building the Data Warehouse. 1st Edition. Wiley and Sons (1992)

Para comprender mejor esta definición, debemos estudiar con más detalle cada una de las características que caracterizan a un Almacén de Datos según Inmon:

- **Integrado:** los datos almacenados en el Almacén de Datos deben integrarse en una estructura consistente, por lo que las inconsistencias existentes entre los diversos sistemas operacionales deben ser eliminadas. La información suele estructurarse también en distintos niveles de detalle para adecuarse a las distintas necesidades de los usuarios.
- **Temático:** sólo los datos necesarios para el proceso de generación del conocimiento del negocio se integran desde el entorno operacional. Los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales. Por ejemplo, todos los datos sobre clientes pueden ser consolidados en una única tabla del Almacén de Datos. De esta forma, las peticiones de información sobre clientes serán más fáciles de responder dado que toda la información reside en el mismo lugar.
- **Histórico:** el tiempo es parte implícita de la información contenida en un Almacén de Datos. En los sistemas operacionales, los datos siempre reflejan el estado de la actividad del negocio en el momento presente. Por el contrario, la información almacenada en el Almacén de Datos sirve, entre otras cosas, para realizar análisis de tendencias. Por lo tanto, el Almacén de Datos se carga con los distintos valores que toma una variable en el tiempo para permitir comparaciones.

- **No volátil:** el almacén de información de un Almacén de Datos existe para ser leído, pero no modificado. La información es por tanto permanente, significando la actualización del Almacén de Datos la incorporación de los últimos valores que tomaron las distintas variables contenidas en él sin ningún tipo de acción sobre lo que ya existía.

Otra característica del Almacén de Datos es que contiene metadatos, es decir, datos sobre los datos. Los metadatos permiten saber la procedencia de la información, su periodicidad de refresco, su fiabilidad, forma de cálculo, etc.

2.6.4 Objetivos de un Almacén de Datos

Los principales objetivos de un Almacén de Datos son:

- Hacer que la información de la organización sea accesible, donde los contenidos del Almacén de Datos sean entendibles, navegables y el acceso a ellos este caracterizado por el rápido desempeño.
- Hacer que la información de la organización sea consistente, es decir, que la información de una parte de la organización coincida con la información de la cualquier otra parte de la organización.
- Proporcionar información adaptable y flexible. El Almacén de Datos está diseñado para cambios continuos. Al realizarse nuevas consultas o insertar datos nuevos al Almacén de Datos, los datos existentes y las tecnologías no cambian ni se corrompen.
- Proteger la información. El Almacén de Datos controla el acceso efectivo a los datos, como así también da a los dueños de la información gran visibilidad en el uso y abusos de los datos.
- Los datos deben ser integrados, no volátiles e históricos, organizados para el apoyo de un proceso toma de decisiones.

2.6.5 Dimensiones

Las dimensiones de un cubo son atributos relativos a las variables, son catálogos de información complementaria necesaria para la presentación de los datos a los usuarios, como por ejemplo: descripciones, nombres, zonas, rangos de tiempo, etc. Es decir, la información general complementaria a cada uno de los registros de la tabla de hechos.

Son la compañía integral de una tabla de hechos, contienen las descripciones textuales del negocio. En un modelo dimensional bien diseñado, las tablas dimensionales contienen muchas columnas o atributos que describen las filas. Se debe incluir tantas descripciones textuales como sea posible. Es común que una tabla dimensional tenga entre 50 y 100 atributos. Tienden a ser pequeñas en cuanto al número de filas (mucho menos de 1 millón) pero son muy extensas horizontalmente, con muchas

columnas. Cada dimensión tiene una clave primaria única, que se enlaza a través de integridad referencial a cualquier tabla de hechos a la cual esté junta.

Los atributos de las dimensiones son utilizados en condiciones de consultas, agrupamientos y reportes etiquetados, y son claves para hacer al Almacén de Datos usable y entendible. El Almacén de Datos es tan bueno como son los atributos dimensionales, depende de la calidad y profundidad de sus atributos dimensionales. Mientras más elementos y terminología del negocio se incorporen en los atributos dimensionales, mejor es el Almacén de Datos. Las tablas dimensionales son los puntos de entrada a las tablas de hechos. Atributos dimensionales robustos llevan a una capacidad de análisis robusto.

Los mejores atributos son textuales y discretos, deben ser palabras reales y completas en vez de abreviaciones.

Algunas veces en el diseño no está claro si un dato numérico que es extraído de un sistema fuente pertenece a una dimensión o a una tabla de hechos. Para tomar la decisión debe preguntarse si ese atributo tiene muchos valores y éstos servirán para hacer cálculos, si es así, pertenece a una tabla de hechos, o, si es una descripción discreta que es más o menos constante y participa en las condiciones de búsqueda, es un atributo dimensional.

2.6.6 Tabla de Hechos

Una tabla de hechos es la tabla central de un esquema multidimensional y almacena los valores de las medidas de negocio, como las ventas, el coste de las mercancías o las ganancias. Cada medida se toma mediante la intersección de las dimensiones que la definen, dichas dimensiones estarán reflejadas en sus correspondientes tablas de dimensiones que rodearán la tabla de hechos y estarán relacionadas con ella.

Una característica esencial de las tablas de hechos es que contienen datos numéricos (hechos) que se pueden resumir para proporcionar información sobre el historial de las operaciones de la organización. Cada tabla de hechos también incluye un índice de varias partes que contiene, como claves externas, las claves primarias de las tablas de dimensiones relacionadas así como los atributos de los registros de hechos. Las tablas de hechos no deben contener información descriptiva ni datos que no procedan de los campos de medición numéricos y los campos de índice que relacionan los hechos con las correspondientes entradas en las tablas de dimensiones. Para medir los datos de una tabla hechos, todas las medidas de una tabla de hechos debe corresponder al mismo grano.

La utilización de estas medidas garantiza que los datos se puedan recuperar y agregar de manera que el negocio pueda hacer uso de la riqueza de datos de negocio de la base de datos.

2.6.7 Claves en las Tablas de Hechos

Todas las tablas de hechos tienen dos a más claves foráneas que conectan a las claves primarias de las tablas dimensionales. Por ejemplo, la clave de producto en la tabla de hechos empatará con la clave de un producto de la tabla dimensional Producto. La tabla de hechos tiene su propia clave primaria conformada por un grupo de claves foráneas que es llamada clave compuesta o concatenada. Cada tabla que expresa una relación con cardinalidad muchos a muchos es una tabla de hechos, las demás son tablas de dimensiones.

Solamente un subconjunto de las claves foráneas que tiene una tabla de hechos basta para identificar a cada fila. De una docena o más de claves, pueden ser 2 o 3. Por ejemplo, código de venta y código de producto. En la mayoría de casos no hay ninguna ventaja en crear una clave primaria con un código en la tabla de hechos, pues la tabla se hace más grande, y un índice sobre este atributo no tiene ningún valor. Sin embargo, puede ser útil para facilitar la administración del sistema y cuando puede haber varias filas con todos los valores iguales.

2.6.8 Cubos de Información

Un Cubo de Información o Cubo OLAP, es un término acuñado por Edgar Frank Codd, quien lo defino como una base de datos multidimensional, en la cual el almacenamiento físico de los datos se realiza en un vector multidimensional.

Los cubos OLAP se pueden considerar como una ampliación de las dos dimensiones de una hoja de cálculo que representa o convierte los datos planos que se encuentran en filas y columnas, en una matriz de N dimensiones.

Los objetos más importantes que se pueden incluir en un cubo multidimensional, son los siguientes:

- **Indicadores:** sumalizaciones que se efectúan sobre algún hecho o expresiones basadas en sumalizaciones, pertenecientes a una tabla de hechos.
- **Atributos:** campos o criterios de análisis, pertenecientes a tablas de dimensiones.
- **Jerarquías:** representa una relación lógica entre dos o más atributos.

De esta manera en un cubo multidimensional, los atributos existen a lo largo de varios ejes o dimensiones, y la intersección de las mismas representa el valor que tomará el indicador que se está evaluando. En la Ilustración 9 se puede observar más claramente lo que se acaba de explicar:

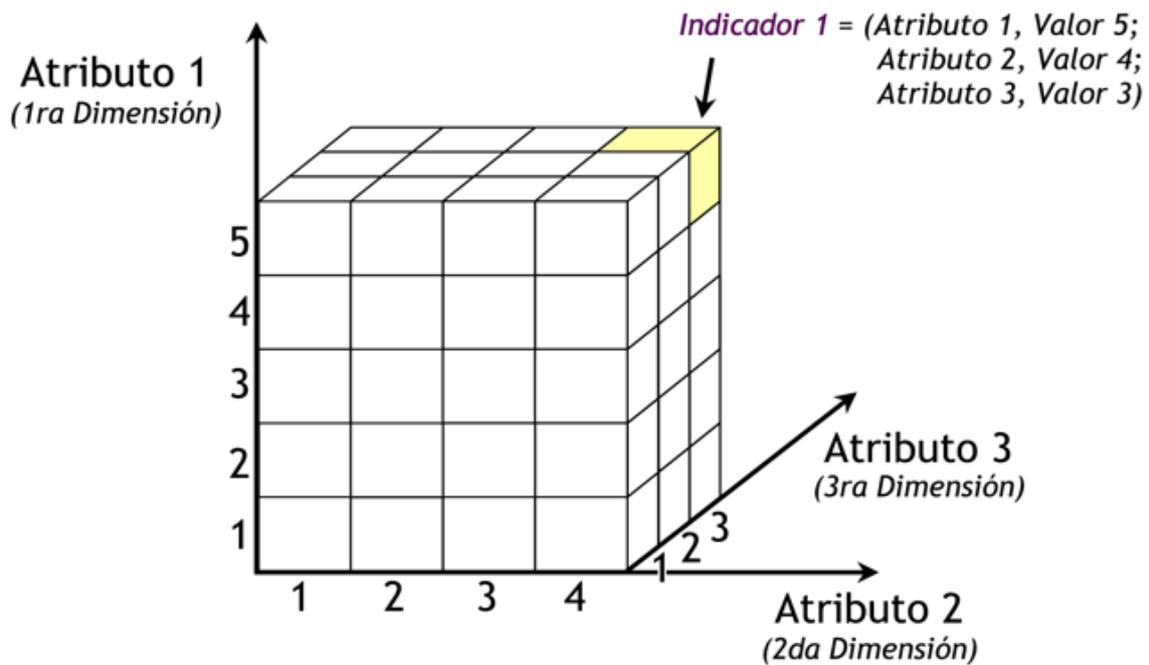


Ilustración 9: Cubo Multidimensional

Fuente: <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/arquitectura-del-data-warehouse/34-Almacén de Datos-manager>

Para la creación del cubo de la ilustración anterior, se definieron tres Atributos (“Atributo 1”, “Atributo 2” y “Atributo 3”) y se definió un Indicador (“Indicador 1”). Entonces el cubo quedó compuesto por 3 dimensiones una por cada Atributo, cada una con sus respectivos valores asociados. También, se ha seleccionado una intersección al azar para demostrar la correspondencia con los valores de las Atributos. En este caso, el indicador “Indicador 1”, representa el cruce del Valor “5” de “Atributo 1”, con el Valor “4” de “Atributo 2” y con el Valor “3” de “Atributo 3”.

Se puede observar, que el resultado del análisis está dado por los cruces matriciales de acuerdo a los valores de las dimensiones seleccionadas.

Más específicamente, para acceder a los datos del Almacén de Datos, se pueden ejecutar consultas sobre algún cubo multidimensional previamente definido. Dicho cubo debe incluir entre otros objetos: indicadores, atributos, jerarquías, etc. Basados en los campos de las tablas de dimensiones y de hechos, que se deseen analizar. De esta manera, las consultas son respondidas con un excelente rendimiento, minimizando al máximo el tiempo que se hubiese incurrido en realizar dicha consulta sobre una base de datos transaccional.

2.6.9 Variables

También llamadas “indicadores de gestión”, son los datos que están siendo analizados. Forman parte de la tabla de hechos. Más formalmente, las variables representan algún aspecto cuantificable o

medible de los objetos o eventos a analizar. Normalmente, las variables son representadas por valores detallados y numéricos para cada instancia del objeto o evento medido. En forma contraria, las dimensiones son atributos relativos a las variables, y son utilizadas para indexar, ordenar, agrupar o abreviar los valores de las mismas. Las dimensiones poseen una granularidad menor, tomando como valores un conjunto de elementos menor que el de las variables.

Una medida es tomada en la intersección de todas las dimensiones (día, producto y tienda). Esta lista de dimensiones define la granularidad de la tabla de hechos y nos dice cuál es el alcance de las mediciones. Una fila en una tabla de hechos corresponde a una medida. Todas las medidas en la tabla de hechos deben tener la misma granularidad.

Los hechos más usados son numéricos y sumarizables, tales como la cantidad en bolívares de cada venta. El que sean sumarizables es muy importante porque en las aplicaciones de Almacén de Datos casi nunca se recupera una sola fila de la tabla de hechos, más bien se recuperan cientos, miles o millones de filas de hechos a la vez, y lo mejor que se puede hacer con tantas filas es sumarlas o resumirlas.

Es posible que la medida de un hecho sea textual, pero es muy raro. Cuando esto sucede, el texto es una medida de algo, y son textos que provienen de una lista predefinida. Es preferible poner medidas textuales en las dimensiones porque pueden ser correlacionados con otros atributos textuales dimensionales y ahorran mucho espacio. Un hecho con datos en texto libre no es posible en un Almacén de Datos porque es imposible de analizar.

Si no hay ventas un determinado día, en una determinada tienda para un determinado producto, esa fila no se registra en la tabla de hechos. No se debe llenar la tabla de hechos con ceros, pues no sirven para el análisis. A pesar de esto, la tabla de hechos consume normalmente más del 90% del espacio de la BD dimensional. Las tablas de hechos tienden a ser profundas en términos del número de filas, pero estrechas en el número de columnas.

2.6.10 Granularidad

La granularidad es uno de los conceptos más importantes en el diseño de un Almacén de Datos porque es directamente proporcional al volumen de datos que almacenará el Almacén de Datos. La granularidad se refiere al nivel de detalle o resumen de los datos contenidos en el Almacén de Datos. A mayor detalle mayor nivel de granularidad, a menor detalle menor nivel de granularidad.

Podemos tomar por ejemplo un registro de datos correspondiente a una venta, el registro individualmente se correspondería con el mayor nivel de granularidad, mientras que una consulta acerca del total de ventas del mes se correspondería con un nivel de granularidad menor.

Evidentemente a mayor nivel de granularidad mayor será el volumen de información, mientras que a menor granularidad menor volumen de datos.

Los datos de origen en los sistemas operacionales se encuentran en su mayor grado de granularidad, es por tanto durante la fase de extracción de los datos de los sistemas operacionales hasta el Almacén de Datos cuando se produce la transformación de los mismos para adecuar su granularidad a la definida en el diseño del Almacén de Datos de destino. A través del nivel de granularidad se puede predecir el crecimiento en número de registros del Almacén de Datos, así como los requerimientos de espacio para el mismo.

La granularidad también afectará como es evidente al tiempo de respuesta de las consultas al Almacén de Datos. A mayor nivel de granularidad menor rendimiento y mayor tiempo de respuesta. A menor nivel de granularidad del sistema, menores recursos son utilizados y aumenta el rendimiento mejorando los tiempos de respuesta.

El nivel de granularidad de un Almacén de Datos es la clave de que esos datos puedan ser usados por diferentes tipos de usuarios en diferentes ocasiones y de diferente manera a lo largo de la compañía.

Por ejemplo los datos de facturación de una compañía pueden ser usados por usuarios de Marketing, Ventas, Operaciones y Financiero. Todos estos departamentos hacen uso de los mismos datos. Para el departamento de Marketing será interesante conocer las ventas por áreas geográficas por meses, para el departamento de ventas será necesario ver los datos desde la perspectiva de las ventas por vendedor o delegación comercial semanalmente. Para el departamento de operaciones será conveniente ver los datos desde la perspectiva de ventas por línea de producto con carácter mensual para de esa manera poder prever el nivel óptimo de stock en los almacenes. Finalmente para el departamento financiero lo importante será ver los datos desde la perspectiva de la rentabilidad de las ventas por línea de producto.

2.6.11 Dimensión Tiempo

En un Almacén de Datos, la creación y el mantenimiento de una tabla de dimensión Tiempo es obligatoria, y la definición de granularidad y estructuración de la misma depende de la dinámica del negocio que se esté analizando. Toda la información dentro del depósito, como ya se ha explicado, posee su propio sello de tiempo que determina la ocurrencia de un hecho específico, representando de esta manera diferentes versiones de una misma situación.

Es importante tener en cuenta que la dimensión tiempo no es sola una secuencia cronológica representada de forma numérica, sino que mantiene niveles jerárquicos especiales que inciden notablemente en las actividades de la organización. Esto se debe a que los usuarios podrán por

ejemplo analizar las ventas realizadas teniendo en cuenta el día de la semana en que se produjeron, quincena, mes, trimestre, semestre, año, estación, etc.

Existen muchas maneras de diseñar esta tabla, y en adición a ello no es una tarea sencilla de llevar a cabo. Por estas razones se considera una buena práctica evaluar con cuidado la temporalidad de los datos, la forma en que trabaja la organización, los resultados que se esperan obtener del almacén de datos relacionados con una unidad de tiempo y la flexibilidad que se desea obtener de dicha tabla.

2.6.12 Agregación

Las Agregaciones no son más que los datos precalculados que permiten una rápida respuesta en el momento de la consulta. Almacenando datos precalculados el servidor se recupera la información para estos cálculos ni hacer cualquier cálculo numérico. El uso de agregaciones es la base para una respuesta rápida en los sistemas OLAP.

Los cubos son la forma de que se almacenan las agregaciones en los sistemas OLAP. Las dimensiones significan consultas al cubo. Las agregaciones están almacenadas en la intersección de las dimensiones. Cada intersección (llamada celda) almacena un simple valor.

Para realizar una agregación es indispensable se considera tanto el espacio en disco como la explosión de los datos.

2.6.13 Jerarquía

Una jerarquía representa una relación lógica entre dos o más atributos pertenecientes a un cubo multidimensional, siempre y cuando posean su correspondiente relación de uno a muchos (1 a N).

Las jerarquías poseen las siguientes características:

- Pueden existir varias en un mismo cubo.
- Están compuestas por dos o más niveles.
- Se tiene una relación de uno a muchos (1 a N) entre atributos consecutivos de un nivel superior y uno inferior.

Por lo general, las jerarquías pueden identificarse fácilmente, debido a que existen relaciones de uno a muchos (1 a N) entre los propios atributos de un cubo.

La principal ventaja de manejar jerarquías, reside en poder analizar los datos desde su nivel más general al más detallado y viceversa, al desplazarse por los diferentes niveles. Un pequeño ejemplo de lo recién expuesto se presente en la Ilustración 10:

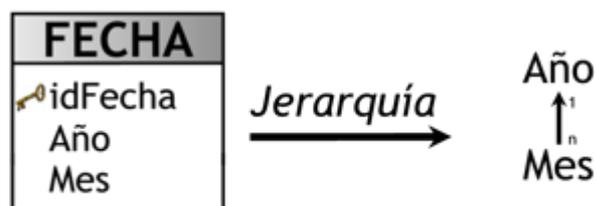


Ilustración 10: Jerarquía

Fuente: <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/arquitectura-del-data-warehouse/34-datawarehouse-manager>

Se crearon dos atributos: “Año” y “Mes”, pertenecientes a la dimensión “FECHA”, los cuales están constituidos de la siguiente manera:

”Año” = FECHA.Año

”Mes” = FECHA.Mes

A la izquierda de la igualdad se encuentra el nombre del atributo; a la derecha el nombre del campo de la tabla de dimensión “FECHA”.

Se creó la jerarquía llamada: “Jerarquía Fechas”, en la cual se colocó el atributo más general en la cabecera y se comenzó a disgregar los niveles hacia abajo. En este caso, la ilustración se explica cómo un mes del año pertenece solo a un año. Un año puede poseer uno o más meses del año.

2.6.14 Operaciones de Exploración Ascendente y Exploración Descendente

La funcionalidad de los sistemas OLAP se caracteriza por ser un análisis multidimensional de datos corporativos, que soportan los análisis del usuario y unas posibilidades de navegación, seleccionando la información a obtener.

Normalmente este tipo de selecciones se ve reflejada en la visualización de la estructura multidimensional, en unos campos de selección que permitan elegir el nivel de agregación (jerarquía) de la dimensión, y/o la elección de un dato en concreto, la visualización de los atributos del sujeto, frente a una(s) dimensiones en modo tabla, pudiendo con ello realizar, entre otras las siguientes acciones:

- **Exploración Ascendente (En inglés, Drill-Up):** Significa subir en la jerarquía, esto es aumentar el tamaño del grano al que están definidos los hechos. Al aplicar esta operación se necesita resumir información para adaptar el nivel de detalle de los hechos, en este proceso de resumen se utilizarán operadores de agregación. Por ejemplo: resumiendo los datos semanales en trimestrales o en anuales.

- **Exploración Descendente (En inglés, Drill-Down):** Esta operación es lo contrario de la exploración ascendente, ahora lo que se pretende es reducir el nivel de grano obteniendo un mayor nivel de detalle. Esto se traduce en cambiar el nivel de definición de los hechos a niveles inferiores de las jerarquías. Por ejemplo, en un informe de ventas en una compañía, se debe poder profundizar en los datos de cada región mundial para obtener los datos por país, y en el total de un país para obtener la información de las ciudades dentro del país.

2.6.15 Metadatos

Los metadatos son datos altamente estructurados que describen información, describen el contenido, la calidad, la condición y otras características de los datos. En pocas palabras son "datos sobre datos", como los define la Universidad del Norte de Arizona. Los metadatos proporcionan la estructura y la descripción de las cosas almacenadas y pueden y son aplicados en distintas áreas, desde catálogos de libros hasta las memorias de computadoras. Ofrece a los usuarios una manera de preservar y acceder a los recursos en largo plazo, protegiendo los datos para que no se pierdan.

Suministran contexto al dato y para efectos del Almacén de Datos es fundamental, ya que proporciona información referente acerca de la fuente de datos, especificaciones de las transformaciones, cargas, elementos eliminados y resumidos después de considerar aspectos como por ejemplo su antigüedad.

Los objetivos que deben cumplir los metadatos, según el colectivo al que va dirigido, son:

- Dar soporte al usuario final, ayudándole a acceder al Almacén de Datos con su propio lenguaje de negocio, indicando qué información hay y qué significado tiene. Ayudar a construir consultas, informes y análisis, mediante herramientas de Inteligencia de Negocio como DSS, EIS o CMI.
- Dar soporte a los responsables técnicos del Almacén de Datos en aspectos de auditoría, gestión de la información histórica, administración del Almacén de Datos, elaboración de programas de extracción de la información, especificación de las interfaces para la realimentación a los sistemas operacionales de los resultados obtenidos, etc.

2.6.16 Ventajas de los Almacenes de Datos

Hay muchas ventajas por las que es recomendable usar un almacén de datos. Algunas de ellas son:

- Proporciona información clave para la toma de decisiones empresariales. Mejora la calidad de las decisiones tomadas.
- Permite transformar información bruta en información útil.
- Especialmente útil para el medio y largo plazo.

- Son sistemas relativamente sencillos de instalar si las fuentes de datos y los objetivos están claros.
- Muy útiles para el almacenamiento de análisis y consultas de históricos.
- Los almacenes de datos hacen más fácil el acceso a una gran variedad de datos a los usuarios finales.
- Los almacenes de datos pueden trabajar en conjunto y, por lo tanto, aumentar el valor operacional de las aplicaciones empresariales, en especial la gestión de relaciones con clientes.

2.6.17 Desventajas de los Almacenes de Datos

Utilizar almacenes de datos también plantea algunos inconvenientes, algunos de ellos son:

- A lo largo de su vida los almacenes de datos pueden suponer altos costos. El almacén de datos no suele ser estático. Los costos de mantenimiento son elevados.
- Los almacenes de datos se pueden quedar obsoletos relativamente pronto.
- A menudo existe una delgada línea entre los almacenes de datos y los sistemas operacionales. Hay que determinar qué funcionalidades de estos se pueden aprovechar y cuáles se deben implementar en el Almacén de Datos, resultaría costoso implementar operaciones no necesarias o dejar de implementar alguna que sí vaya a necesitarse.
- No es muy útil para la toma de decisiones en tiempo real debido al largo tiempo de procesamiento que puede requerir. En cualquier caso la tendencia de los productos actuales (junto con los avances del hardware) es la de solventar este problema convirtiendo la desventaja en una ventaja.
- Requiere de continua limpieza, transformación e integración de datos. Mantenimiento.
- En un proceso de implantación puede encontrarse dificultades ante los diferentes objetivos que pretende una organización.
- Una vez implementado puede ser complicado añadir nuevas fuentes de datos.

2.6.18 Tipos de Implementación de un Almacén de Datos

Una base de datos con modelo dimensional, es una base de datos que tiene una estructura adecuada para resolver consultas analíticas. Se trata de modelos sencillos que aseguran unos buenos tiempos de respuesta, y que se corresponden bastante con el lenguaje de negocio de los usuarios. Las herramientas de Inteligencia de Negocio se conectan al modelo dimensional del Almacén de Datos.

Este modelo dimensional es el modo óptimo de organizar los datos en los sistemas de Inteligencia de Negocio, y puede hacerse mediante bases de datos relacionales (ROLAP), o utilizando bases de datos multidimensionales (MOLAP). Ambos tipos de herramientas, tanto ROLAP como MOLAP, están diseñadas para realizar análisis de datos a través del uso de modelos de datos multidimensionales, aunque en el caso de ROLAP estos modelos no se implementan sobre un sistema multidimensional, sino sobre un sistema relacional clásico.

2.6.18.1 ROLAP

El término ROLAP viene del acrónimo en inglés de Procesamiento Analítico Relacional en Línea, se trata de sistemas y herramientas OLAP construidos sobre una base de datos relacional. Es una alternativa a la tecnología MOLAP pero cuenta con todos los beneficios de un Sistema de Gestión de Base de Datos Relacional a los cuales se les provee extensiones y herramientas para poder utilizarlo como un Sistema Gestor del Almacén de Datos.

En la Ilustración 11 se muestra un ejemplo de arquitectura de una implementación de Almacén de Datos utilizando ROLAP:

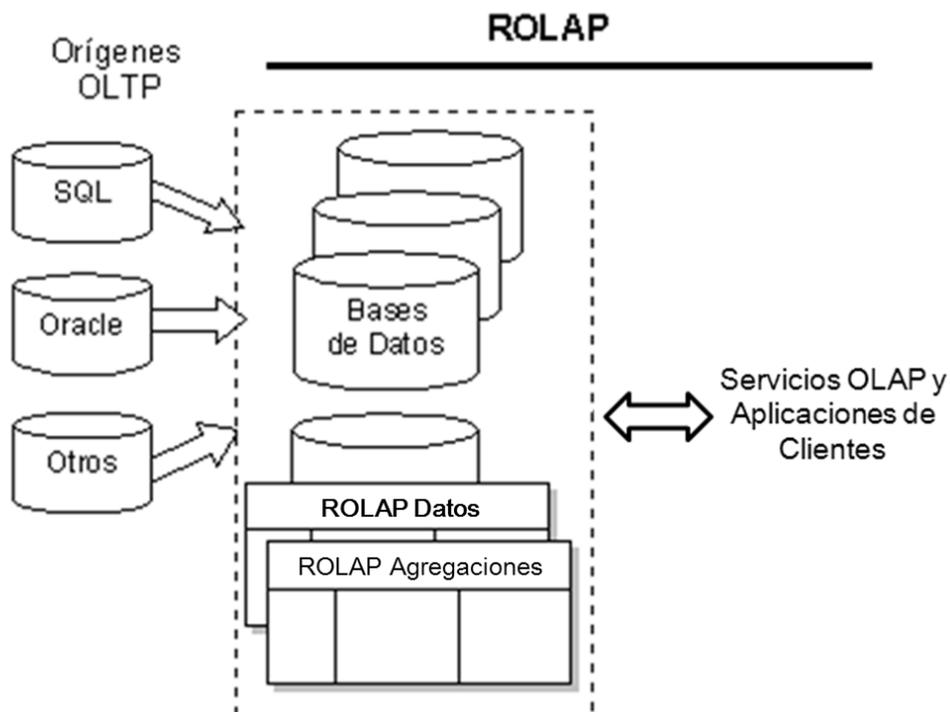


Ilustración 11: Arquitectura ROLAP

Fuente: http://www.tesis.ufm.edu.gt/67332/tesis.htm#_Toc519847876

En los sistemas ROLAP, los cubos multidimensionales se generan dinámicamente al instante de realizar las diferentes consultas, haciendo de esta manera el manejo de cubos transparente los usuarios. Este proceso se puede resumir a través de los siguientes pasos:

- Se seleccionan los indicadores, atributos, jerarquías, entre otros, que compondrán el cubo multidimensional.
- Se ejecutan las consultas sobre los atributos, indicadores, entre otros, seleccionados en el paso anterior. Entonces, de manera transparente a los usuarios se crea y calcula dinámicamente el cubo correspondiente, el cual dará respuesta a las consultas que se ejecuten.

Al no tener que intervenir los usuarios en la creación y el mantenimiento explícito de los cubos, ROLAP brinda mucha flexibilidad, ya que dichos cubos son generados dinámicamente al momento de ejecutar las consultas.

La principal desventaja de los sistemas ROLAP, es que los datos de los cubos se deben calcular cada vez que se ejecuta una consulta sobre ellos. Esto provoca que ROLAP no sea muy eficiente en cuanto a la rapidez de respuesta ante las consultas de los usuarios.

Para incrementar la velocidad de respuesta, en algunos casos se puede optar por almacenar los resultados obtenidos de ciertas consultas en la memoria caché, para que en un futuro, cuando se desee volver a ejecutar dicha consulta, los valores sean obtenidos más velozmente.

Cabe aclarar que si los datos del Almacén de Datos son almacenados y gestionados a través de un SGBD Relacional, no se requiere de otro software que administre y gestione los datos de manera Multidimensional.

Entre las características más importantes de ROLAP, se encuentran las siguientes:

- Almacena la información en una base de datos relacional.
- Utiliza índices de mapas de bits.
- Utiliza índices de Uniones.
- Posee optimizadores de consultas.
- Cuenta con extensiones de SQL.

Como se aclaró anteriormente, el almacén de datos se organiza a través de una base de datos multidimensional, sin embargo, puede ser soportado por un SGBD Relacional. Para lograr esto se utilizan los diferentes esquemas, en estrella, copo de nieve y constelación, los cuales transformarán el modelo multidimensional y permitirán que pueda ser gestionado por un SGBD Relacional, ya que solo se almacenarán tablas.

2.6.18.2 MOLAP

El término MOLAP viene del acrónimo en inglés de Procesamiento Analítico Multidimensional en Línea. El objetivo de los sistemas MOLAP es almacenar físicamente los datos en estructuras multidimensionales de manera que la representación externa y la interna coincidan.

Para ello, se dispone de estructuras de almacenamiento específicas y técnicas de compactación de datos que favorecen el rendimiento del Almacén de Datos.

En la Ilustración 12 se muestra un ejemplo de arquitectura de una implementación de Almacén de Datos utilizando MOLAP:

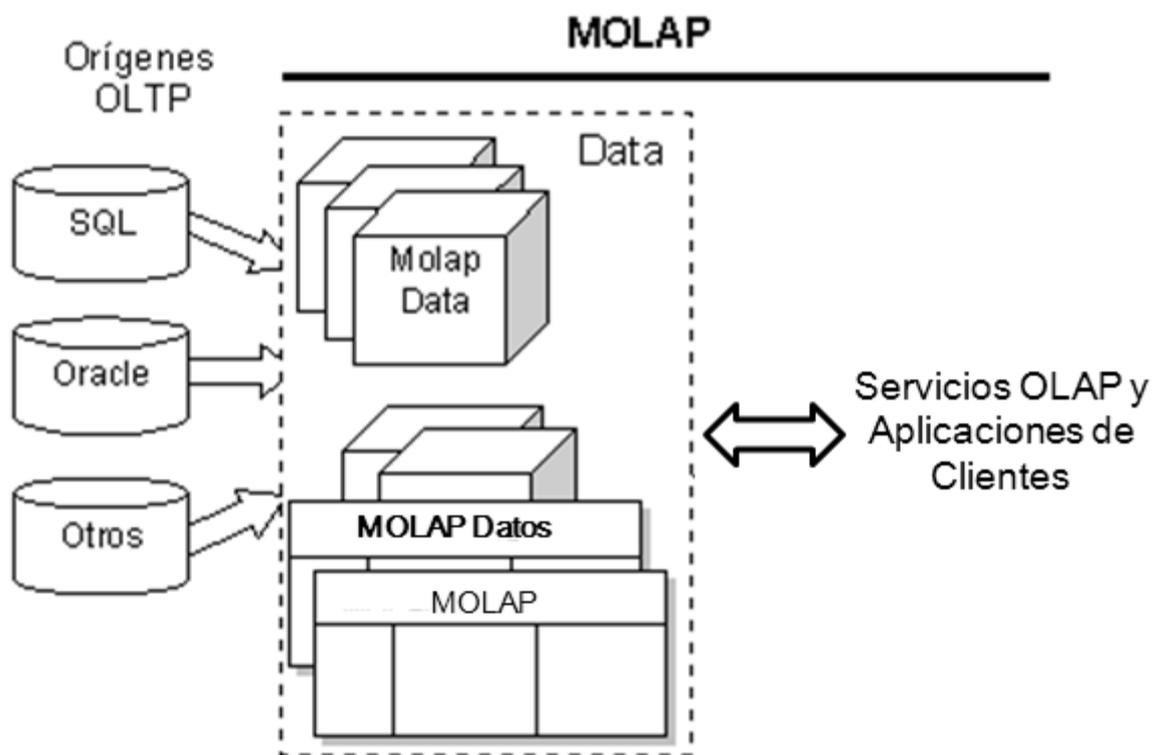


Ilustración 12: Arquitectura MOLAP

Fuente: http://www.tesis.ufm.edu.gt/67332/tesis.htm#_Toc519847876

MOLAP requiere que en una instancia previa se generen y calculen los cubos multidimensionales, para que luego puedan ser consultados.

El principal motivo de precalcular los datos de los cubos, es que posibilita que las consultas sean respondidas con mucha rapidez, ya que los mismos no deben ser calculados en tiempo de ejecución, obteniendo de esta manera una muy buen rendimiento.

Existen una serie de desventajas que están directamente relacionadas con la ventaja de precalcular los datos de los cubos multidimensionales, ellas son:

- Cada vez que se requiere o es necesario realizar cambios sobre algún cubo, se debe tener que recalcularlo totalmente, para que se reflejen las modificaciones llevadas a cabo. Provocando de esta manera una disminución importante en cuanto a flexibilidad.
- Se precisa más espacio físico para almacenar dichos datos (esta desventaja no es tan significativa).

Habitualmente, los datos del Almacén de Datos son almacenados y gestionados a través de SGBD Relacionales, ya que estos tienen la ventaja de poder realizar consultas directamente a través del lenguaje SQL. En estos casos, para la generación de los cubos multidimensionales se requiere de otro software que administre y gestione los datos de manera Multidimensional.

2.6.18.3 ROLAP vs MOLAP

En la Tabla 4 se pueden apreciar las principales diferencias entre estos dos tipos de implementación:

Rolap	Molap
Brinda mucha flexibilidad, ya que los cubos son generados dinámicamente al momento de hacer las consultas.	Cada vez que se requiere hacer un cambio en el cubo, se debe tener que recalcularlo todo, para que se reflejen las modificaciones. Provocando de esta manera una disminución importante en la flexibilidad.
Los datos de los cubos se deben calcular cada vez que se ejecuta una consulta. Esto provoca que Rolap no sea muy eficiente en cuanto a la rapidez de las consultas.	Las consultas son respondidas con mucha rapidez, ya que los datos no se deben ser calculados en tiempo de ejecución, obteniendo con esto un buen rendimiento.

Tabla 4: Comparación entre ROLAP y MOLAP

<http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/arquitectura-del-data-warehouse/34-Almacén de Datos-manager>

2.6.18.4 HOLAP

El término HOLAP viene del acrónimo en inglés de Procesamiento Analítico Híbrido en Línea. Constituye un sistema híbrido entre MOLAP y ROLAP, que combina estas dos implementaciones para almacenar algunos datos en un motor relacional y otros en una base de datos multidimensional. (Ver Ilustración 13)

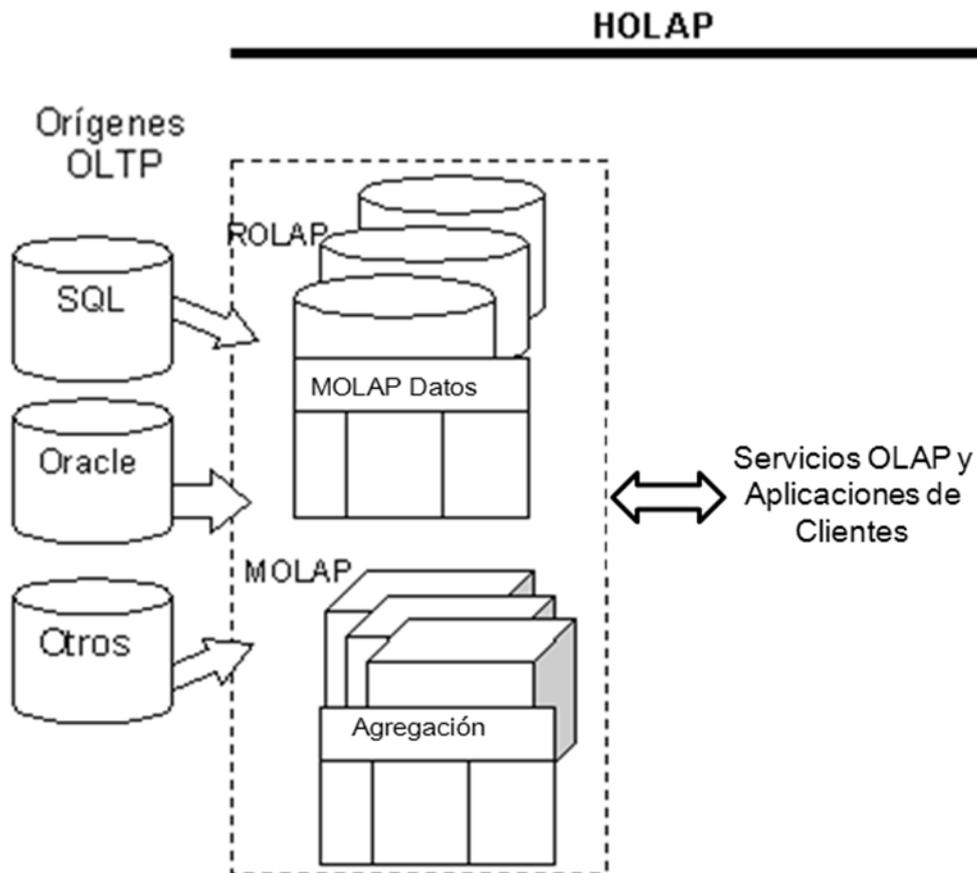


Ilustración 13: Arquitectura HOLAP

Fuente: http://www.tesis.ufm.edu.gt/67332/tesis.htm#_Toc519847876

Los datos agregados y precalculados se almacenan en estructuras multidimensionales y los de menor nivel de detalle en estructuras relacionales. Es decir, se utilizará ROLAP para navegar y explorar los datos, y se empleará MOLAP para la realización de tableros.

Como contrapartida, hay que realizar un buen análisis para identificar los diferentes tipos de datos.

2.7 Inteligencia de Negocios

2.7.1 Definición

La Inteligencia de Negocios cuenta con una diversidad de interpretaciones como muchos otros términos o conceptos. Su uso es justificado a todo lo que sea considerado como tecnología de información, pero no hay un consenso en lo respecta a su definición.

Inteligencia de Negocios es el conjunto de tecnologías que permiten a las empresas utilizar la información disponible en cualquier parte de la organización para hacer mejores análisis, descubrir nuevas oportunidades y tomar decisiones más informadas. Esta definición fue creada por Howard Dresner en 1989, siendo él el acreedor del término, a pesar de que H.P. Luhn en el año 1.958 escribió un artículo llamado "A Inteligencia de Negocio System" donde describe las características que debe tener un sistema de este tipo. De igual forma no fue hasta la década de los 90 que proliferó la multiplicación de aplicaciones de Inteligencia de Negocios, logrando facilitar el acceso a la información.

El término hizo su reaparición en 1996 cuando un reporte de Gartner Group dijo textualmente lo siguiente:

Para el año 2000, la Democracia de la Información emergerá en las empresas de vanguardia, con las aplicaciones de Inteligencia de Negocios ampliamente disponibles a nivel de empleados, consultores, clientes, proveedores y el público en general. La clave para surgir en un mercado competitivo es mantenerse delante de sus competidores. Se requiere más que intuición para tomar decisiones correctas basadas en información exacta y actualizada. Las herramientas de reporte, consulta y análisis de datos pueden ayudar a los usuarios de negocios a navegar a través de un mar de información para sintetizar la información valiosa que en él se encuentra - hoy en día esta categoría de herramientas se les llama "Inteligencia de Negocios"

Examinando las distintas definiciones queda claro que la Inteligencia de Negocios no consiste en una metodología, sistema, software o herramienta en particular, si no es un conjunto de tecnologías que tienen un fin común con el principal objetivo que es la de servir como soporte para la toma de decisiones.

La inteligencia de negocio actúa como un factor estratégico para una empresa u organización, generando una ventaja competitiva, que no es otra que proporcionar información privilegiada para responder a problemas de negocio como la entrada a nuevos mercados, promociones u ofertas de productos, control financiero, optimización de costes, planificación de la producción, análisis de perfiles de clientes, rentabilidad de un producto concreto, etc.

2.7.2 Usuarios de las Soluciones de Inteligencia de Negocios

A continuación se describen los dos tipos de usuarios que intervienen en una solución de inteligencia de negocios, según CANO (2007):

- **Productores de información:** Habitualmente son el 20% de los usuarios, que crean informes o modelos utilizando herramientas de escritorio. Donde predominan estadísticos que se valen de herramientas para minería de datos o son creadores de informes que utilizan herramientas para el diseño y/o programación de informes específicos. Regularmente son profesionales del área de sistema de información o usuarios muy avanzados con capacidades de comprender la información y la informática.
- **Los consumidores de información:** Son usuarios no habituales que regularmente consultan informes para la toma de decisiones, pero no acceden a los números o hacen análisis detallados. Los usuarios no habituales son directivos, gestores, responsables, colaboradores y usuarios externos. Este numeroso grupo está bien servido con cuadros de mando con análisis guiados, informes interactivos e informes de gestión estandarizados. La mayoría de estas herramientas proveen ahora acceso vía Web para promover el acceso desde cualquier lugar y facilitar el uso y minimizar los costes de administración y mantenimiento.

2.7.3 Historia de la Inteligencia de Negocios

En octubre de 1.958 H.P. Luhn de IBM, escribió un artículo llamado Inteligencia de Negocio System donde describe las características que debe tener un sistema de este tipo, en muchos aspectos de lo que escribió Luhn tiempo atrás, si se realiza un paralelismo de lo que se entiende hoy en día por inteligencia de negocios no varió mucho.

Edgar Frank Codd presenta el concepto de las bases de datos en el año 1.969, un año más tarde se desarrollaron las primeras base de datos con sus interfaz empresariales, estas aplicaciones, facilitan la entrada de datos en los sistemas, haciendo que aumente la información disponible, pero como el acceso a la información era de alta complejidad y difícil de acceder a las mismas no fue una solución completa.

Otro de los avances llegaron en el año 1.980 con la creación del concepto del Almacén de Datos por Ralph Kimball y Bill Inmon, con ellas aparecieron los primeros sistemas de reportes, con todo esto la solución seguía siendo compleja y funcionalmente pobre, se contaba con potentes sistemas de bases de datos pero no existían aplicaciones que facilitaran su explotación.

En 1.989 Howard Dresner difundió el término de Inteligencia de Negocio escrito por H.P. Luhn en el año 1.958, en la década de los 90 llegaba la Inteligencia de Negocio 1.0 y con ello la multiplicación de aplicaciones de Inteligencia de Negocios, logrando facilitar el acceso a la información

considerablemente pero empeoraron el problema que se quería resolver ya que seguían apareciendo múltiples soluciones sin poder consolidarse. Con la llegada de Inteligencia de Negocio 2.0 en el año 2.000 se logró consolidar las aplicaciones en una pocas plataformas a partir de ahí las herramientas empezaron a dar soluciones reales a las empresas u organizaciones.

2.7.4 Características de la Inteligencia de Negocios

- **Información:** el propósito de la Inteligencia de Negocios es proveer de información al usuario final para dar soporte a la toma de decisiones, por ende la información es la esencia de la Inteligencia de Negocios, estas pueden originarse desde las bases de datos operacionales, como también de arquitecturas de Bodega de Datos y Almacén de Datos diseñadas específicamente para el análisis.

El usuario requiere de información para apoyarse en el momento de tomar decisiones, pasando en segundo plano de dónde provenga esta, pudiendo ser de una fuente primitiva o derivada, para lo cual la inteligencia de negocios utiliza o crea fuentes de datos interna o externa con el fin de utilizarla como materia prima para lograr su objetivo.

- **Apoyo a la toma de decisiones:** básicamente consiste en organizar y presentar los datos relevantes para que sirvan como soporte a la hora de tomar decisiones. Esto implica la utilización de tecnologías, técnicas de análisis y todo lo que sea necesario con el fin de obtener solamente aquella información relevante y útil.
- **Orientación al usuario final:** un factor fundamental que tuvo su incidencia en la tecnología de la Inteligencia de Negocios para explotar información, fue que el usuario final no contaba con conocimientos y técnicas que le permita acceder de una manera sencilla y directa a los datos almacenados en los sistemas operacionales, ya que casi siempre necesitaba de ayuda de informáticos para acceder a la información, con la Inteligencia de Negocios, se elimina la dependencia de terceras personas para el acceso a los datos e información, siendo esta una herramienta sencilla y preparada para que interactúe directamente con el usuario final sin intermediarios.

2.7.5 Arquitectura de una Solución de Inteligencia de Negocios

La Arquitectura de una Solución de Inteligencia de Negocios viene determinada por su situación central como fuente de información para las herramientas de análisis. La estructura básica de la arquitectura:

- **Fuentes de datos:** Generalmente son los datos generados por sistemas operacionales, los cuales se utilizan para alimentar de información el Almacén de Datos. Las fuentes de información a las que se puede acceder son:

- Los sistemas operacionales, que contienen las aplicaciones desarrolladas a medida.
- Sistemas de información por sector: presupuestos, hojas de cálculo, etcétera o fuentes de información externa, compradas a terceros como por ejemplo el estudio del mercado.
- **Proceso ETC:** es el proceso de extracción, transformación y carga de los datos en el Almacén de Datos. Los datos antes de almacenarse en un Almacén de Datos, pasan por procesos de filtrado, limpieza, transformación y redefinición, habitualmente consume entre el 60% y el 80% de recursos en un proyecto de Inteligencia de Negocios tal como lo indica Ralph Kimball (2004), por lo que es un proceso clave en la vida de todo proyecto de esta naturaleza.
- **Área de Integración de Datos:** esta área se utiliza como repositorio para estandarizar, integrar, consolidar y extraer sólo la información necesaria para la construcción de análisis y reportes del negocio.
- **Almacén de datos:** en él se almacenan los datos de una manera que optimice su flexibilidad, facilidad de acceso y administración, en donde los datos están estructurados para generar informes que ayuden a la toma de decisiones.

El Almacén de Datos o almacén de datos, es la herramienta que surgió como respuesta a las necesidades de los usuarios de los niveles gerenciales, que necesitan información consistente, integrada, histórica y preparada para ser analizada y utilizada para mejorar la toma de decisiones.

El hecho de recuperar la información desde los distintos sistemas que posee la empresa, sean estos transaccionales o externos, para luego almacenarlos en un entorno cohesionado de información, como es un Almacén de Datos, permitirá analizar la información contextualmente y relacionada dentro de la organización.

- **Herramientas de Consulta:** representan los procedimientos y herramientas de trabajo que utiliza el usuario para consultar al modelo de datos preparado, por medio del planteamiento de preguntas, la búsqueda de información, la identificación de tendencias y todos aquellos elementos que contribuyen a la evaluación y toma de decisiones.

Son aplicaciones de generación de informes, que permiten generar vistas de datos agregadas para mantener a la gerencia informada sobre el estado de su negocio.

En la Ilustración 14, podemos apreciar los componentes anteriormente descritos y como se relacionan:

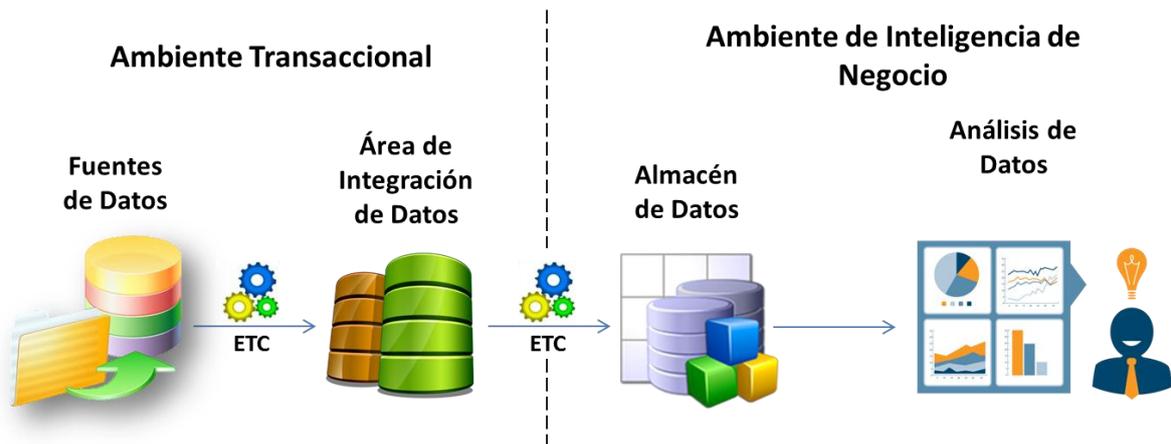


Ilustración 14: Arquitectura del Almacén de Datos

Fuente:

http://www.uaeh.edu.mx/docencia/P_Presentaciones/huejutla/sistemas/mineria_datos/almacen_datos.pdf

2.7.6 Procesos Básicos del Almacén de Datos (ETC)

El término ETC viene del acrónimo en inglés de Extraer, Transformar y Cargar y se refiere a los datos en una empresa. ETC es el proceso que organiza el flujo de los datos entre diferentes sistemas en una organización y aporta los métodos y herramientas necesarias para mover datos desde múltiples fuentes a un Almacén de Datos, reformatearlos, limpiarlos y cargarlos en otra base de datos o Bodega de Datos. ETC forma parte de la Inteligencia de Negocio.

La idea es que una aplicación ETC lea los datos primarios de unas bases de datos de sistemas principales, realice transformación, validación, el proceso cualitativo, filtración y al final escriba datos en el almacén y en este momento los datos son disponibles para analizar por los usuarios Ralph Kimball (2004).

- **Extracción:** Es el primer paso de obtener información hacia el ambiente del Almacén de Datos. Consiste en extraer los datos desde los sistemas de origen. La mayoría de los proyectos de Almacén de Datos fusionan datos provenientes de diferentes sistemas de origen. Cada sistema separado puede usar una organización diferente de los datos o formatos distintos. La extracción convierte los datos a un formato preparado para iniciar el proceso de transformación.

Una parte esencial del proceso de extracción es la de analizar los datos extraídos, luego se realiza un chequeo que verifica si los datos cumplen la pauta o estructura que se esperaba. Si no es así son rechazados.

Un requerimiento de importancia en proceso de extracción es que esta cause el menor impacto posible en el sistema origen. Esto se debe a que frecuentemente la gran cantidad de datos que

se deben extraer, este manejo masivo de información puede ralentizar e incluso colapsar el sistema origen. Por esta razón todas las operaciones de extracción suele realizarse en horarios o días donde este impacto sea mínimo.

La extracción incluye la limpieza de los datos. Las operaciones de limpieza básicas incluyen llenado de valores faltantes, corrección de errores tipográficos y otros errores de captura de datos, se procede a hacer reemplazo de sinónimos por identificadores que respeten un estándar y así unificar los criterios de los datos.

- **Transformación:** Luego de que la información fue extraída, se procede con la fase de extracción, esta aplica una serie de reglas de negocio sobre los datos extraídos para convertirlos en los datos que serán cargados.

Se deben codificar los datos que se han decodificado en forma inconsistente en distintas fuentes, limpieza de la información, eliminar lo que consideramos innecesario, seleccionar campos específicos que consideremos necesarios para el Almacén de Datos, también se deben tomar en cuenta algunas consideraciones como por ejemplo: si aceptara valores nulos, traducir códigos, codificar valores libres, unir datos de múltiples fuentes, abreviaturas de estados, códigos de productos, lugares de negocios, entre otros.

- **Carga:** Al final del proceso de transformación, los datos están preparados para ser cargados en el sistema destino. Dependiendo de los requerimientos de la organización, este proceso puede abarcar una amplia variedad de acciones diferentes. En algunas bases de datos se sobrescribe la información antigua con la información nueva. En cambio, en el Almacén de Datos se mantiene un historial de los registros de manera que se pueda hacer una auditoria de los mismos, teniendo almacenada toda la historia de un valor a lo largo del tiempo.

Las formas más básicas para desarrollar el proceso de carga son dos:

- **Acumulación simple:** consiste en realiza un resumen de todas las transacciones comprendida en el periodo de tiempo seleccionado y transportar el resultado como una única transacción hacia el Almacén de Datos para su almacenamiento.
- **Rolling:** almacena información resumida a distintos niveles, correspondientes a distintas agrupaciones de la unidad de tiempo o diferentes niveles jerárquicos. (Casos de varios niveles de granularidad)

2.7.7 Inteligencia de Negocios en los Distintos Niveles Jerárquicos de la Organización

La Inteligencia de Negocios es el conjunto de productos y servicios que permiten a los usuarios finales acceder y analizar de manera rápida y sencilla la información para la toma de decisiones de negocio a nivel operativo, táctico y estratégico (Ver Ilustración 15).



Ilustración 15: Niveles Jerárquicos de la Organización

Fuente: <http://www.udec.cl/dti/node/108>

A continuación se describen los tres niveles jerárquicos:

- **Inteligencia de Negocios a Nivel Operativo:** permite que los empleados que trabajan con información operativa puedan recibir dicha información de una manera oportuna, exacta y adecuada y se componen básicamente de herramientas de reportes u hojas de cálculo con un formato fijo cuya información se actualiza frecuentemente. Un ejemplo de esto podría ser un supervisor de ventas, quien es el encargado de monitorear el cumplimiento de las cuotas de ventas de los vendedores a su cargo.
- **Inteligencia de Negocios a Nivel Táctico:** permite que los analistas de datos y la gerencia media de la empresa utilicen herramientas de análisis y consulta con el propósito de tener acceso a la información sin intervención de terceros. Como ejemplo un gerente de ventas recibe un reporte en donde se indica que las ventas de una determinada categoría de productos o servicios, se incrementaron de manera inusual con relación al periodo anterior.
- **Inteligencia de Negocios a Nivel Estratégico:** permite que la alta dirección de las empresas pueda analizar y monitorear tendencias, patrones, metas y objetivos estratégicos de la organización. Un ejemplo de ello, lo constituye el cuadro de mando integral.

2.7.8 Inteligencia de Negocios y Almacén de Datos

La relación de soporte directo entre los sistemas de Inteligencia de Negocio y el Almacén de Datos se puede definir haciendo un especial énfasis en el modelado de los datos y el análisis de los mismos.

Inteligencia de Negocio es la habilidad de tomar mejores decisiones de manera más rápida. Un entorno de Inteligencia de Negocio enfocado al cliente provee de la infraestructura necesaria para proveer la información necesaria para maximizar la base de clientes. Esta infraestructura combina datos, canales y técnicas analíticas para mejorar la satisfacción del cliente y el beneficio a través de mayores puntos de contacto.

Los sistemas de ingreso de transacciones se ven afectados por las consultas a sus bases de datos. Generalmente poseen informes predefinidos, y cualquier modificación a los mismos debe ser solicitado al departamento de sistemas, que será en definitiva quien evaluará si tiene los recursos humanos disponibles como para realizarlos. El Almacén de Datos está separado de los sistemas transaccionales, por lo tanto las consultas no afectan la velocidad de registro de las operaciones. Se actualiza periódicamente, capturando datos de los distintos sistemas transaccionales. Una vez implementado, es importante elegir las herramientas de consulta al Almacén de Datos, que permitan un alto grado de autonomía a los usuarios.

2.7.9 Ventajas de la Inteligencia de Negocios

- Disponer de la capacidad de extraer, depurar, consolidar, sintetizar y presentar datos-información-conocimiento de forma automatizada.
- Poner a disposición de los usuarios que la necesiten la información necesaria para el análisis y la toma de decisiones: ágil, flexible y fiable, en el formato adecuado a cada uno. La información se obtiene sin dependencias de otros departamentos, con posibilidad de navegación OLAP por los propios usuarios, que permite profundizar en el análisis de forma interactiva en base a cualquiera de las dimensiones disponibles.
- Homogeneidad en la utilización de la información (interna y externa): utilización de la misma información al medir las mismas cosas.
- Sistema soportado sobre plataformas tecnológicas sólidas y escalables.
- Controlar los costos, ya que ofrece una solución que permite manejar fácilmente los distintos programas que se encuentran en los diferentes departamentos de la compañía.
- Mejorar la colaboración y la calidad de las decisiones, facilitando el acceso a la información en todos los niveles de la organización.
- Orientar las soluciones tecnológicas hacia el usuario, ya que reduce los tiempos de aprendizaje mediante el uso de herramientas.

- Proporcionar una profunda visión del negocio a través de un sistema integrado de usos: cuadros de mando, tableros de instrumentos, informes, minería de datos y almacenamiento analítico.
- Proveer asistencia a los ejecutivos para planear y pronosticar el trabajo, presentando una descripción común de los procesos del negocio de una compañía.

En conclusión, Inteligencia de Negocio es una herramienta moderna y de nueva generación, disponible a los gestores y directores del negocio quienes tienen la necesidad de analizar el pasado, usar herramientas estadísticas de predicción, y con ello estar un paso de los competidores y mejorar los resultados empresariales. Al fin y al cabo ese es el fin último de la tecnológica, mejorar el rendimiento y productividad de la organización.

Los sistemas de inteligencia de negocios han evolucionado a la fecha, incorporando nuevas y poderosas funcionalidades de análisis y presentación de información para distintos niveles de usuarios y áreas en una empresa.

2.7.10 Principales Software para Inteligencia de Negocio

Debido al acelerado desarrollo tecnológico en los últimos años es fácil imaginar que las empresas dedicadas a crear sistemas de información han tenido cambios considerables. A pesar de todas las adquisiciones en la industria, de la aparición de nuevas compañías y de la disolución de otras, el mercado está compuesto principalmente por las siguientes empresas: SAP, IBM Cognos, SAS Institute, Oracle y Microsoft. (O'Donnell, Sipsma, & Watt, 2012).

- **SAP:** Inició sus actividades en el año 1994 y ahora es la empresa con mayor participación en el mercado de los proveedores de plataformas de inteligencia de negocios controlando el 23,6% del mercado (Gartner, 2012). De acuerdo a los mismos clientes, SAP puede soportar la mayor cantidad de información y es considerada como una plataforma estándar de inteligencia de negocios. Sus fortalezas más reconocidas son las relacionadas con la tecnología OLAP.

Sin embargo, SAP no es bien catalogado en el área de soporte al cliente. Junto con eso se presentan problemas con la migración de plataformas y con los productos forzados.

- **Business Objects:** Ofrece a los usuarios el acceso constante y de manera sencilla y clara a los datos relevantes, para facilitar el análisis de la información que ha sido almacenada, y el posterior desarrollo de distintos informes para mejorar los procesos globales de la empresa.
- **Cognos:** IBM es una empresa reconocida en todo lo que es tecnología. Ofrecen tanto productos computacionales como software y servicios de Inteligencia de Negocios. Hace unos años, IBM adquirió Cognos, otra empresa importante en la industria de Inteligencia de Negocios. Con esto, IBM Cognos unificó todas sus tecnologías de Inteligencia de Negocios,

análisis y administración en una sola plataforma (Business Analytics), enviando un mensaje al mercado de que es necesario que las plataformas de Inteligencia de Negocios cuenten con todas esas tecnologías para ofrecer mejores soluciones a los negocios (Sallam, Richardson, Hagerty, & Hostmann, 2011).

Este brinda un sistema efectivo para llevar a cabo evaluaciones correctas de la información y una posterior toma de decisiones adecuada. Cabe destacar que Cognos cuenta además con una herramienta especial, que ha sido desarrollada para permitir la realización de modelos, pronósticos y simulaciones de negocios, con el fin de ofrecer la posibilidad de planificar el futuro de las operaciones.

- **Microstrategy:** Es otra de las herramientas más difundidas, cuya característica fundamental reside en que provee de las soluciones necesarias a los clientes de cualquier tipo de empresa o sector funcional de las mismas, con el objetivo de brindar un marco de ayuda que les permitirá acceder a un mayor conocimiento de la información que está siendo manejada por la empresa.

La herramienta entre sus características principales dispone de BBDD Analíticas Avanzadas y predictivas similares a OLAP y BBDD dimensionales como los almacenes de datos. Sin olvidar que ofrece una gran escalabilidad, posibilita la generación de cuadros de mando y la creación de informes empresariales.

- **Oracle BI** (En español, Inteligencia de Negocio de Oracle): Oracle es la segunda empresa que mayor dominio de mercado, con un 15,6% (Gartner, 2012). Lo que más destaca en Oracle es que cuenta con una gran variedad de aplicaciones para distintas funciones que puedan requerir los clientes. Además, cuenta con un amplio set de canales de distribución, especialmente para su producto OBBIE.

Al igual que las diversas aplicaciones de su tipo, ofrece la posibilidad de tener acceso a la información, compartir esos datos con los diversos sectores de la compañía, y permitir el análisis de dicha información para llevar a cabo la toma de decisiones correctas, que estarán basadas en datos obtenidos de manera sencilla y rápida.

- **WorkMeter:** Recopila información sobre el uso de las aplicaciones informáticas y de otros dispositivos (PBX, móviles,...) para determinar el nivel de carga y el esfuerzo de las personas. Ésta información es agrupada y presentada tanto al gestor como al mismo empleado, generando un ambiente de objetividad y transparencia. Workmeter ofrece la base objetiva necesaria para facilitar procesos de evaluación y mejora, permitiendo asimismo medir el impacto positivo de los cambios.

- **Pentaho BI** (En español, Inteligencia de Negocio de Pentaho): Es un conjunto de herramientas de inteligencia de negocios desarrollada bajo la filosofía del software libre para la gestión y toma de decisiones empresariales. Es una plataforma compuesta de diferentes programas que satisfacen los requisitos de la organización, ofreciendo soluciones para la gestión y análisis de la información, incluyendo el análisis multidimensional OLAP, presentación de informes, minería de datos y creación de cuadros de mando para el usuario. La plataforma ha sido desarrollada bajo el lenguaje de programación Java y tiene un ambiente de implementación también basado en Java, haciendo así que Pentaho sea una solución flexible.

A continuación se presenta el cuadro comparativo de algunas de las herramientas analizadas:

Herramienta BI	¿Es Software Libre o Propietario?	¿Es una Suite con Productos Integrados?	¿Ofrece servicio de Soporte de la herramienta?	¿Ofrece Reporte Empresarial?	¿Ofrece Cubos de Análisis?
SAP	Propietario	No	Sólo si dispone licencia	Si	Si
Cognos	Propietario	Si	Sólo si dispone licencia	Si	No
Microstrategy	Propietario	Si	Sólo si dispone licencia	Si	Si
OBI	Propietario	Si	Sólo si dispone licencia	Si	Si
Pentaho	Libre	Si	Sólo Versión Enterprise	Si	Si

Tabla 5: Cuadro Comparativo de las Herramientas de BI Analizadas

Fuente: <http://journal.tourgune.org/wp-content/uploads/2013/12/Issue-1.5-Estudio-comparativo-de-las-herramientas-de-Business-Intelligence.pdf>

2.8 Tecnologías y Herramientas de Inteligencia de Negocios a Utilizar

2.8.1 Oracle BI (En español, Inteligencia de Negocio de Oracle)

Oracle BI es un producto basado en Inteligencia de Negocios creado por la empresa Oracle, que brinda soluciones completas de negocio, arquitectura para empresas con capacidades de informes, consultas y análisis, sistema OLAP, cuadros de mando y cuadros de instrucciones. También ofrece todas las

fuentes de datos empresariales como mediciones, cálculos, definiciones, entre otros, y las jerarquías se crean de un modelo común de la información de la empresa, proporcionando a los usuarios una visión precisa y consistente, independientemente del lugar donde se encuentren. Los usuarios pueden acceder e interactuar con la información de múltiples maneras, incluyendo cuadros de mando interactivos basados en la Web, barras de búsqueda, aplicaciones ERP y CRM, dispositivos móviles y aplicaciones de Microsoft Office.

- **Oracle BI Dashboards** (En español, Cuadro de Mando de Inteligencia de Negocio de Oracle): Le ofrece impecables cuadros de mando, que le permiten conocer de un solo vistazo los factores clave en la gestión de la empresa. Incluye todo tipo de elementos como tablas dinámicas o gráficos interactivos, que le permitirán detectar los puntos clave en el menor tiempo posible.
- **Oracle BI Reporting** (En español, Reporteador de Inteligencia de Negocio de Oracle): Esta aplicación permite disponer de una alta capacidad a la hora de obtener información consistente para la toma de decisiones, sin necesidad de programación o de personal especializado. Es una aplicación Web orientada al usuario final, que permite crear análisis personalizados desde cero o modificar los existentes, ofreciendo unas completas opciones de visualización.
- **Oracle BI Scorecarding & Strategy Management** (En español, Desarrollo de Cuadros de Mando Integrales y Gestión de Estrategias de Inteligencia de Negocio de Oracle): Integra la gestión estratégica de la empresa, realizando un seguimiento automatizado de los objetivos empresariales. También se puede saber en todo momento cual es el grado de cumplimiento de los planes de la empresa y en qué puntos se está fallando.
- **Oracle BI Alerts** (En español, Alertas de Inteligencia de Negocio de Oracle): El sistema trabaja detectando anomalías que excedan los niveles de tolerancia definidos. Cuando esto sucede emite un mensaje de aviso con un informe vinculado sobre la anomalía.
- **Oracle BI Publisher** (En español, Reporteador de Inteligencia de Negocio de Oracle): ofrece un potente servidor de publicación que permite generar informes en base a múltiples fuentes de datos, con muchos tipos de documentos y distribuirlos por múltiples canales.
- **Oracle BI Server** (En español, Servidor de Inteligencia de Negocio de Oracle): es un servidor de consultas y análisis altamente escalable y eficiente que integra datos mediante sofisticadas capacidades de consulta federada a múltiples consultas relacionales, sistemas OLAP y aplicaciones pre-empaquetadas.

- **Oracle BI Answer** (En español, Respuestas de Inteligencia de Negocio de Oracle): una poderosa herramienta de consultas y análisis, que trabaja a través de vistas lógicas de la información proveniente de múltiples fuentes en un ambiente Web.

En la Ilustración 16 se muestra un ejemplo de un Panel de Control con distintos análisis en Oracle Business Intelligence:

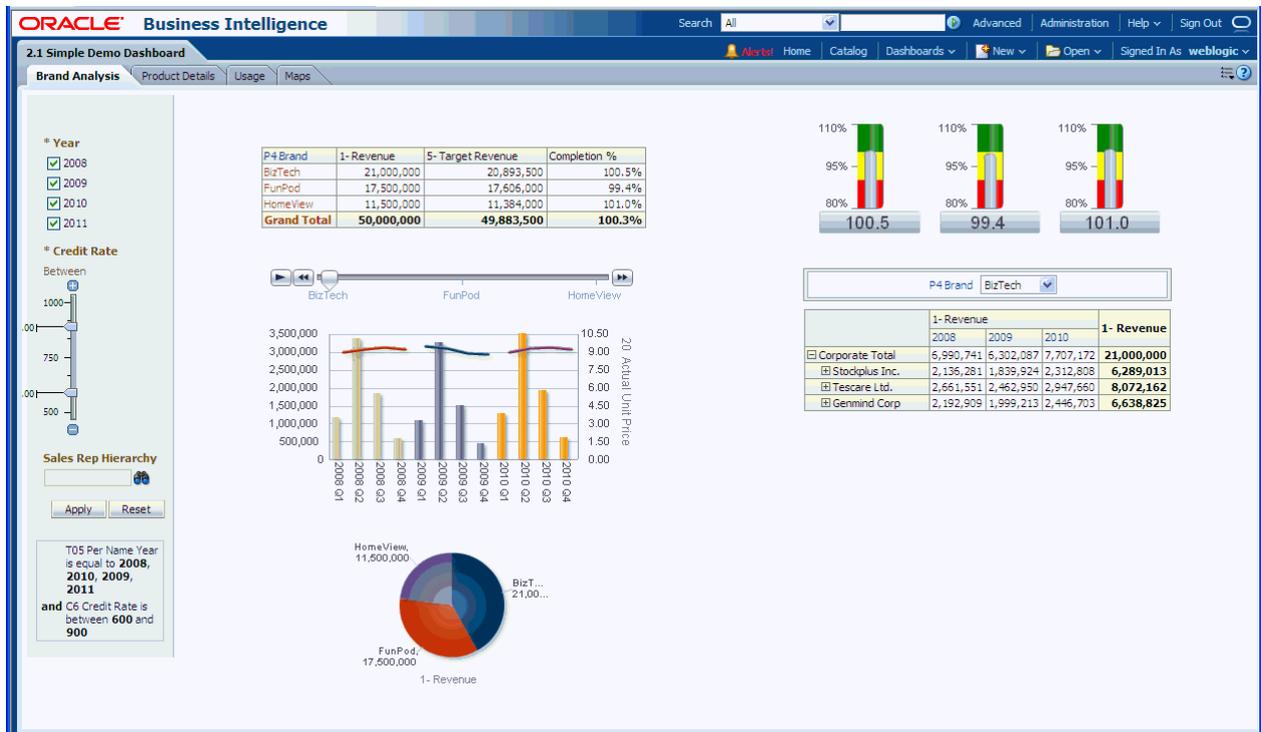


Ilustración 16: Oracle Business Intelligence

Fuente: http://docs.oracle.com/cd/E25178_01/fusionapps.1111/e20835/img/dashboardex.gif

2.8.2 Oracle Warehouse Builder (En español, Desarrollador de Almacenes de Datos de Oracle)

Es una herramienta para la creación de ETC producida por Oracle, ofrece un ambiente grafico para construir, manejar y mantener el proceso de integración de datos en los sistemas de inteligencia de negocios. Es una herramienta completa para todos los aspectos de administración de datos y metadatos. Brinda funcionalidades destinadas a la calidad y la integración de los datos, el modelado relacional y multidimensional, facilitando la administración de todo el ciclo de vida de datos y metadatos.

Warehouse Builder permite transformar los datos una base de datos transaccional información de calidad que facilite la toma de decisiones. Ofrece herramientas para la calidad de datos, auditoria de datos, interacción relacional completa y modelado dimensional, además del manejo completo de la metadata. Warehouse Builder permite crear Almacenes de datos, migrar datos de distintos sistemas,

integrar datos de distintas fuentes de datos, limpiar y transformar los datos para proveer información de calidad y el manejo de metada corporativa.

Entre las funcionalidades que ofrece Oracle Warehouse Builder encontramos las siguientes:

- **Opciones Avanzadas para Carga de Datos:** en grandes implementaciones, los procesos se ejecutan en paralelo, agotando los recursos debido a las grandes cantidades de datos que participan en la carga. Warehouse Builder permite ejecutar cargas de datos usando métodos rápidos y eficientes. Esta es una aproximación completamente diferente al movimiento de datos que se realiza por medio de conjuntos de datos controlados por el motor de base de datos. Los incrementos en velocidad de carga se deben entonces a la omisión de las capas SQL tradicionales.
- **Productividad para el Desarrollador:** Warehouse Builder posee una que permite que de manera gráfica, se creen piezas de transformación y se compartan a través de múltiples desarrolladores.
- **Soporte a Dimensiones Lentamente Cambiantes Integrado:** en Warehouse Builder la lógica de dimensiones lentamente cambiantes es diseñada en los metadatos de la dimensión. La dimensión captura toda la lógica que será aplicada a los datos que se insertan dentro de la misma.

Una vez, los usuarios de negocio han decidido que atributos son importantes, el modelador de datos diseña la dimensión. Con los pasos de ETC el desarrollador ve esta dimensión como cualquier otra. El siguiente paso es definir cómo efectuar los cambios y actualizaciones en la dimensión. Warehouse Builder automatiza esto basado en la definición de la dimensión. La combinación entre estos pasos, el diseño y la estandarización hacen que los procesos relacionados con dimensiones lentamente cambiantes presente un alto desempeño en su ejecución.

- **Traza Lineal de Principio a Fin:** los servicios de administración de dependencias le permiten prever el efecto que puedan tener los cambios que se hagan en cualquier lugar de los metadatos del sistema ETC antes de que estos puedan afectarlo y deteriorarlo. Por primera vez, se cuenta con la posibilidad de administrar los cambios de su sistema de manera proactiva y propagar los cambios a través de su sistema haciendo uso el Mapping Editor.
- **Administración de Configuraciones Avanzadas:** tener que mover sistemas o aplicaciones de un ambiente a otro (por ejemplo de desarrollo a producción) es un problema frecuente. Con la Enterprise ETL Option de Warehouse Builder, es posible generar un modelo para configurar los ambientes de desarrollo, pruebas y producción a niveles separados. Para mover los

cambios desde el ambiente de desarrollo a producción, el código generado en desarrollo es modificado en producción solo en áreas específicas. Por ejemplo, la información sobre las conexiones es substituida por la definida en desarrollo.

En la Ilustración 17 se presenta la pantalla principal del Oracle Warehouse Builder:

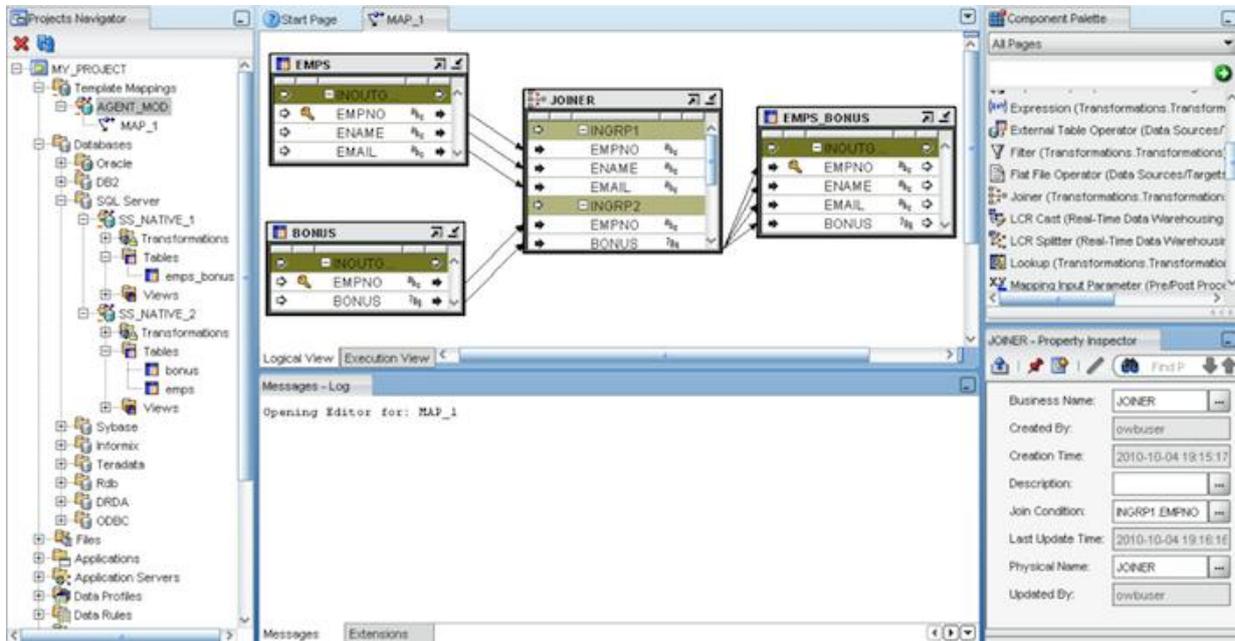


Ilustración 17: Oracle Warehouse Builder

Fuente: <http://www.oracle.com/ocom/groups/public/@otn/documents/webcontent/183940.png>

2.8.3 Oracle Administration Tool (En español, Herramienta de Administración de Oracle)

Es una herramienta de administración de Oracle BI que presenta una representación gráfica de las tres capas de un repositorio, permitiendo crear, editar, gestionar y administrar los repositorios de Oracle BI. Un repositorio de Oracle BI está compuesto por tres capas:

- **Capa Física:** Es la capa que contiene información acerca de los orígenes de datos físicos a los que Oracle BI envía las consultas. La forma más común de llenar la capa física es importa los metadatos de las bases de datos y otros orígenes de datos de los que necesitamos extraer los datos.

Al importar la metadata, muchas de las propiedades de los orígenes de los datos se configuran automáticamente según la información recopilada durante el proceso de importación. También se pueden definir otros atributos del origen de datos físicos que podrían no existir en los metadatos del origen de datos.

- **Capa de Negocio:** define los modelos lógicos o de negocio, especifica la asignación entre los modelos de negocio y los esquemas de la capa física. Es en esta capa donde se simplifican los esquemas físicos para formar la base de la vista de los datos por parte de los usuarios.

Para esta capa la herramienta contiene uno o más objetos de modelo de negocio. Un objeto de modelo de negocio contiene las definiciones y asignaciones de las tablas lógicas a las físicas para el modelo de negocio.

- **Capa de Presentación:** por último la capa de presentación es un medio para simplificar o personalizar aún más el modelo de negocio para los usuarios finales. Al simplificar la vista de los datos para los usuarios, les resulta más fácil elaborar consultas basadas en sus necesidades de negocio. Por ejemplo puede organizar las columnas de forma distinta a la estructura de la tabla que se encuentra en la capa de negocio, muestras menos columnas y cambiar el nombre de las mismas para que sean más comprensibles.

En la Ilustración 18 se presenta la pantalla principal del Oracle Administration Tool. En esta pantalla podemos apreciar las tres capas descritas anterior mente:

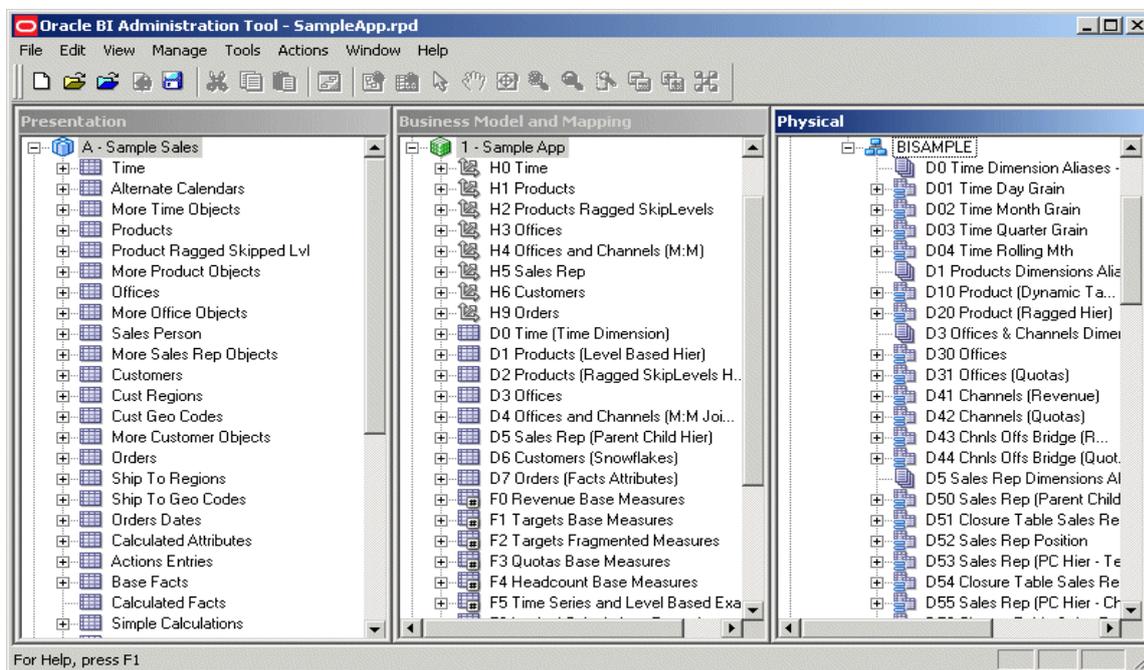


Ilustración 18: Oracle Administration Tool

Fuente: http://docs.oracle.com/cd/E21764_01/bi.1111/e10540/img/admin_tool_main_window.gif

2.9 Encuesta

2.9.1 Instrumento de Recolección

Es la herramienta que permite una adecuada explotación de la fuente de información. Cuando hablamos de recolección de datos nos estamos refiriendo a la información empírica abstraída en conceptos.

La recolección de datos está relacionada con el concepto de medición, la recolección es el proceso mediante el cual se obtiene el dato, valor o respuesta para la variable que se investiga.

La medición, etimológicamente viene del verbo medir y significa comparar una cantidad con su respectiva unidad con el fin de averiguar cuantas veces la segunda está contenida en la primera (Diccionario de la Real Academia Española).

El instrumento de recolección de datos está orientado a crear las condiciones para la medición. Todo lo empírico es medible y todo lo que es medible es cuantificable. No existe ningún aspecto de la realidad que escape a esta posibilidad.

2.9.2 Definición

La encuesta es quizás el instrumento más conocido y utilizado por los investigadores sociales cuando se quiere lograr precisión y representatividad partiendo directamente de consideraciones individuales y no estructurales, para acceder a la conclusión sobre la existencia de regularidades de estructuras sociales y sobre los sujetos insertos en ellas. Aunque los antecedentes de la encuesta nos remiten a los estudios de John Sinclair (Informe Estadístico de Escocia, 1791-1825), James Kay Shuttleworth (Las condiciones morales y físicas de la vida de los obreros de la industria textil en Manchester, 1832) y Charles Booth (Vida y trabajo de los habitantes de Londres, 1889-1891) no será hasta después de la Segunda Guerra Mundial cuando la encuesta empiece a ser la técnica predominante gracias a estudios del mercado estadounidense de carácter privado como los de Gallup, Crossley y Ropper.

Conceptualmente la encuesta puede considerarse como una técnica cuantitativa de investigación sociológica para la obtención de datos por observación directa, que consiste en una investigación realizada sobre una muestra de sujetos, representativa de un colectivo más amplio que se lleva a cabo en el contexto de la vida cotidiana, utilizando procedimientos estandarizados de interrogación con el fin de conseguir mediciones cuantitativas sobre una gran cantidad de características objetivas y subjetivas de la población o grupo que se pretende estudiar.

Estas encuestas pueden ser **Encuestas por Muestreo** en donde se elige una parte de la población que se estima representativa de la población total o podrá ser **Encuesta de la Población** (Ver Ilustración 19).

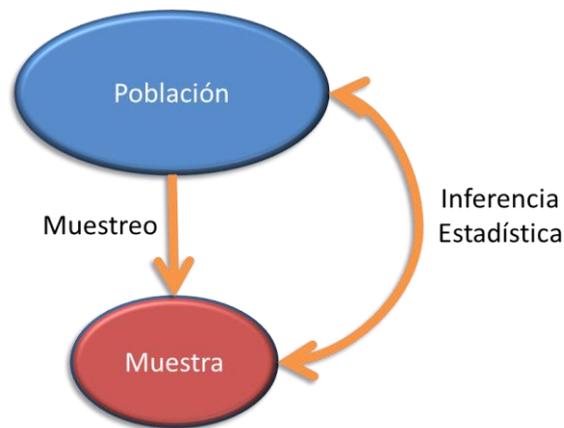


Ilustración 19: Relación entre Población y Muestra

Fuente: <http://www.estrucplan.com.ar/Producciones/entrega.asp?IdEntrega=438>

Una forma reducida de una encuesta por muestreo es un sondeo de opinión, esta forma de encuesta es similar a un muestreo, pero se caracteriza porque la muestra de la población elegida no es suficiente para que los resultados puedan aportar un informe confiable. Se utiliza solo para recolectar algunos datos sobre lo que piensa un número de individuos de un determinado grupo sobre un determinado tema.

La encuesta se ha convertido en una herramienta fundamental para el estudio de las relaciones sociales. Las distintas organizaciones contemporáneas, políticas, económicas o sociales utilizan esta técnica como instrumento indispensable para conocer el comportamiento de sus grupos de interés y tomar decisiones sobre ellos.

Por esta razón la encuesta adquiere en las diversas investigaciones una relevancia significativa pues al recoger información de la realidad se enfrenta tanto el campo teórico como el empírico, y se vale para esto de tipos de cuestionarios que se aplican a distintas personas ya sea en lugares de trabajo o en sus residencias. El investigador decide qué criterio de selección utilizará para la muestra, éste puede ser: sexo, edad o nivel socioeconómico, la muestra tiene la opción de ser estratificada y distribuida en diferentes sectores geográficos de una localidad o de una región determinada.

La encuesta es una fotografía que se le hace a un determinado sector de la sociedad con el objeto de extraer información que permita la constatación empírica de la investigación que se está efectuando.

Las diversas investigaciones en ciencias sociales en la actualidad recurren a la herramienta de la encuesta debido a que se hace necesario producir un nexo entre teorías y realidad observada y medida.

2.9.3 Definición del Universo de Estudio y la Elección de la Muestra

La definición del universo corresponde a todos los elementos o a todas las características de la población de estudio a las cuales se investigará desde la muestra. Para poder hacer esto es necesario

saber qué tipo de población es, si es homogénea o heterogénea, el número que está involucrado y sus características geográficas. Al tener esto claro se puede calcular la muestra y el tipo de muestra que corresponde de acuerdo a la población objeto de estudio.

2.9.4 El Universo de Estudio

Para Ferrando (1986), el universo o población es el conjunto de elementos objeto del estudio. En la realidad existen multitud de universos constituidos por elementos específicos que los diferencian. Cada uno de ellos, puede ser objeto de investigación mediante la técnica de la encuesta y de ahí que para realizar encuestas de opinión, controles de calidad, estudios de coyuntura, de mercado, etc., se utilicen muestras específicas adaptadas a los universos diferenciados de cada estudio.

Para que realmente se pueda iniciar con garantía el proceso de muestreo es necesario acotar el universo y conocer las unidades que lo componen. Acotar el universo significa concretar perfectamente la población que va a ser objeto de estudio. Las unidades del universo acotado constituyen el marco del que se va a sacar la muestra. En la práctica en gran parte de los estudios de ciencias sociales, el marco suele ser la población en general o un segmento de la misma. En estos la información sobre el universo se deduce de los censos generales de población que permiten conocer el marco de estudio.

Es necesario ahora, en consecuencia, establecer el tamaño de la muestra de acuerdo a la variabilidad. Dicho de otro modo, las estimaciones realizadas a partir de muestras muy pequeñas se desvían considerablemente de la realidad, produciéndose un ajuste cada vez mayor, a medida que se añaden nuevos elementos a la muestra.

Siendo así el problema radica en determinar el tamaño de la muestra de tal forma que una parte ofrezca la suficiente fiabilidad para los objetivos de la investigación, y por otra no suponga un aumento innecesario de los costos y del tiempo de realización.

Tratándose de muestras probabilísticas en cuyo proceso siempre intervine la selección aleatoria de los elementos, el tamaño de la muestra depende de:

- La varianza poblacional
- El nivel de confianza elegido
- El máximo error permitido en las estimaciones

Para la selección de la muestra se pueden utilizar distintos métodos o distintas combinaciones de métodos. Las más usuales son las siguientes:

- Muestreo aleatorio simple
- Muestreo aleatorio sistemático

- Muestreo aleatorio estratificado
- Muestreo por conglomerados
- Muestreo no aleatorio.

2.9.5 Modalidades de Encuesta

Las encuestas pueden ser clasificadas de distintas maneras:

Según sus objetivos:

- **Encuestas descriptivas:** Buscan reflejar o documentar las actitudes o condiciones presentes. Esto significa intentar describir en qué situación se encuentra una determinada población en momento en que se realiza la encuesta.
- **Encuestas analíticas:** Buscan en cambio, además de describir, explicar los porqués de una determinada situación. En este tipo de encuestas las hipótesis que las respaldan suelen contrastarse por medio de la examinación de por lo menos dos variables, de las que se observan interrelaciones y luego se formulan inferencias explicativas.

Según las preguntas:

- **De respuesta abierta:** En estas encuestas se le pide al interrogado que responda el mismo a la pregunta formulada. Esto le otorga mayor libertad al entrevistado y al mismo tiempo posibilitan adquirir respuestas más profundas así como también preguntar sobre el porqué y cómo de las respuestas realizadas. Por otro lado, permite adquirir respuestas que no habían sido tenidas en cuenta a la hora de hacer los formularios y pueden crear así relaciones nuevas con otras variables y respuestas.
- **De respuesta cerrada:** En éstas, los encuestados deben elegir para responder una de las opciones que se presentan en un listado que formularon los investigadores. Esta manera de encuestar da como resultado respuestas más fáciles de cuantificar y de carácter uniforme. El problema que pueden presentar estas encuestas es que no se tenga en el listado una opción que coincida con la respuesta que se quiera dar, por esto lo ideal es siempre agregar la opción “otros”.

Según la forma su alcance:

- **Encuestas simples** son encuestas de tipo descriptiva, de respuesta cerrada que están enfocadas a un producto o servicio específico, poseen una cantidad pequeña de preguntas por lo cual son diseñadas de manera rápida y normalmente son usadas cuando se requiere obtener resultados rápidos acerca del producto o servicio en el cual están enfocadas.

- **Encuestas complejas:** son encuestas que requieren de una estructura operativa para su ejecución, es decir, se debe realizar un análisis completo de las preguntas que se van a hacer y del conjunto de la población al cual irá dirigida dicha encuesta.

Según el medio de captura:

- **El papel:** Consisten en entrevistas directas o personales con cada encuestado. Tienen la ventaja de ser controladas y guiadas por el encuestador, además, se suele obtener más información que con otros medios. Sus principales desventajas son el tiempo que se tarda para la recolección de datos, su costo que es más elevado que las encuestas telefónicas o por Internet. El papel sigue siendo el medio más usado a pesar de los avances tecnológicos de las últimas décadas por su bajo costo, versatilidad y seguridad.

A continuación se muestra las ventajas y desventajas de las encuestas capturadas por medio del papel:

Ventajas	Desventajas
Elevado índice de respuestas.	Es caro y lento
Facilidad de cooperación de las personas entrevistadas debido a la presencia del entrevistador, a la vez que permite resolver dudas.	El entrevistador puede ejercer influencia y debe estar entrenado a la vez que controlado
Permite evitar influencia de otras personas.	Difícil acceso a ciertas poblaciones.
Posibilita la realización de entrevistas largas	
Pueden mostrarse materiales.	
Pueden obtenerse datos secundarios (presencia, ambiente, entre otros.)	

Tabla 6: Ventajas y Desventajas de la Encuesta Capturada por Medio del Papel

Fuente: Elaboración Propia

- **Entrevistas telefónicas:** Este tipo de encuesta consiste en una entrevista vía telefónica realizada por un equipo de personas entrenadas que serán las encargadas de verbalizar las preguntas y apuntar las respuestas. Sus principales ventajas son que se puede abarcar un gran número de personas en menos tiempo que la entrevista personal, sus costos suelen ser bajos y es de fácil administración (hoy en día, existen aplicaciones especializadas para la gestión de encuestas telefónicas). Las encuestas telefónicas se emplean cuando se desea aplicar un

cuestionario corto, de no más de diez preguntas, y se desean obtener resultados inmediatos; se utilizan en encuestas de coyuntura política, sondeos de opinión, recordación publicitaria y posicionamiento de marcas. La tasa de rechazo de este tipo de encuesta es comparativamente baja, siempre que se cuenten con encuestadores bien entrenados que logren persuadir al entrevistado para lograr la entrevista y mantener su atención.

A continuación se muestra en la Tabla 7, se presentan las ventajas y desventajas de las encuestas capturadas vía telefónica:

Ventajas	Desventajas
Económico con reservas	Presenta problemas de muestreo pues excluye a las personas que no tienen teléfono.
Técnica muy rápida tanto contactando como recogiendo la información	Existe un nivel de no respuesta no asociado al muestreo por el hecho de marcar números telefónicos al azar.
Puede utilizarse como medio único, como auxiliar o combinado con otras técnicas o modalidades de encuesta	No pueden exhibirse materiales
Permite entrevistar a determinadas poblaciones (personas importantes, muy ocupadas, de zonas mal comunicadas, entre otros.).	Brevidad de la entrevista.
Mejor inclinación a responder por teléfono.	No es apropiado para tratar temas delicados ni preguntas complejas.
Se necesitan menos entrevistadores y la supervisión es adecuada	

Tabla 7: Ventajas y Desventajas de las Encuestas Capturadas vía Telefónica

Fuente: Elaboración Propia

- **Encuestas en Línea:** se subdividen en dos categorías:
 1. **La Web:** Este tipo de encuesta consiste en un cuestionario en una página Web o en enviarlo a los correos electrónicos de un panel predefinido. Sus principales ventajas son la amplia cobertura a la que se puede llegar (incluso a miles de encuestados en varios países y al mismo tiempo), el ahorro de tiempo (se puede obtener miles de encuestas respondidas

en cuestión de horas), los bajos costos (que son menores a las encuestas cara a cara, por teléfono y postales) y la utilización de medios audiovisuales durante la encuesta. Sus desventajas son que no siempre se puede verificar la identidad del encuestado y la interrogante que deja la muestra en cuanto a su representatividad del universo. También tienen una baja tasa de respuestas, dado que habitualmente el sujeto no se motiva a responder, a no ser que tenga algún interés en los resultados del estudio o porque está recibiendo un pago. Una variante de la encuesta en la Web es la encuesta enviada por correo electrónico.

2. **Dispositivos móviles:** Las encuestas mediante dispositivos móviles permiten su aplicación con encuestador, grabando los datos directamente en algún dispositivo tipo teléfono celular o tableta, con o sin conexión a la Internet. El principal problema de usar este medio, más allá del costo del dispositivo, es que no pueden ser usados en sitios con alta tasa de delincuencia o pobreza, pues se corre el riesgo de perder tanto el equipo como los datos.

A continuación se muestran las ventajas y desventajas de las encuestas en línea:

Ventajas	Desventajas
Se precisa escaso personal para realizar la encuesta	Bajos índices de respuesta.
Accesibilidad	Inconvenientes formales: pueden darse errores en la cumplimentación, no se controla el orden de las preguntas y el cuestionario ha de ser corto
Flexibilidad de tiempo para el entrevistado que ofrece mayor calidad a la información	Impersonal
Inmediatez en el envío de las encuestas vía correo electrónico.	Pueden influir otras persona
La amplia cobertura a la que se puede llegar (incluso a miles de encuestados en varios países y al mismo tiempo).	Poco útil para encuestas de respuesta inmediata.
Reducción drástica de costes de envío: manipulación, sobres y franqueo.	Ciertos grupos quedan excluidos (analfabetos absolutos y funcionales).
Anonimato en las respuestas.	
Menor tiempo de respuesta (y por lo tanto mayor número de respuestas)..	
Inmediatez de resultados.	
Anulación de costes de recepción y manipulación de respuestas: se ven los resultados en línea	

Tabla 8: Ventajas y Desventajas de las Encuestas en Línea

Fuente: Elaboración Propia

Cada día se está extendiendo más la realización de encuestas en línea, debido a que es una forma de realizar encuestas mucho más rápida y barata y evita tener que diseñarlas en papel y tener que repartirlas después entre todos los encuestados. El usuario puede crear la encuesta desde cero o utilizando ejemplos de encuestas existentes y enviarla mediante correo electrónico, colgarla en su Web, etc. dejando así la encuesta disponible para los usuarios que la vayan a rellenar.

Los resultados también son más fáciles de interpretar, ya que el computador los muestra de una forma clara y ordenada (mediante gráficas, comparativas, etc.), para que se puedan analizar, compartir, redactar informes, etc.

Permiten reducir considerablemente los costes ya que es suficiente con realizar la encuesta en línea y enviarla mediante correo electrónico, Facebook, Twitter o incluso enlazar la encuesta desde su página Web.

2.9.6 El Cuestionario

Sir Francis Galton, un antropólogo británico explorador y estadista, inventó los cuestionarios a finales de 1800, según Galton el cuestionario es un instrumento de investigación que consiste en una serie de preguntas y otras indicaciones con el propósito de obtener información de los consultados. Aunque a menudo están diseñados para poder realizar un análisis estadístico de las respuestas, no es siempre así.

El cuestionario es un documento formado por un conjunto de preguntas que deben estar redactadas de forma coherente, y organizadas, secuenciadas y estructuradas de acuerdo con una determinada planificación, con el fin de que sus respuestas nos puedan ofrecer toda la información que se precisa.

Los cuestionarios es el documento básico para obtener la información en la gran mayoría de las investigaciones y estudios de mercado.

Algunas características de los cuestionarios son su bajo costo, no requieren de mucho esfuerzo por parte del consultado como por ejemplo las encuestas orales o telefónicas, y a menudo tienen respuestas estandarizadas que hacen más simple la tabulación de los datos. Sin embargo, estas respuestas estandarizadas pueden frustrar a los usuarios.

2.9.6.1 Tipos de Preguntas:

En los cuestionarios pueden aparecer diferentes tipos de preguntas según las características del tema a investigar. En general, pueden aparecer en los cuestionarios los tipos de preguntas siguientes:

- **Abiertas:** (preguntas que sólo formulan la pregunta, sin establecer categorías de respuesta) Se deben utilizar muy poco en las encuestas porque después de aplicar la encuesta hay que cerrarlas y luego estandarizarlas.
- **Cerradas:** Dicotómicas (establecen sólo 2 alternativas de respuesta, "Si o No" y a veces Ns/Nc) Se deben utilizar sólo para temas muy bien definidos que admiten estas 2 alternativas como respuesta.
- **Semicerradas o semiabiertas:** Contienen una serie de posibles respuestas previamente establecidas y una respuesta abierta tipo "otros" para incorporar respuestas de opinión.

Además de la pregunta, también se establecen las categorías de respuesta y a su vez se subdividen en:

- **De respuesta espontánea:** el encuestador no debe leerle la respuesta al encuestado.
- **De respuesta sugerida:** el entrevistador lee las preguntas al encuestado.
- **De valoración:** el entrevistador lee una escala de intensidad creciente o decreciente de categorías de respuesta.
- Según su función en el cuestionario, los siguientes tipos de preguntas responden a un objetivo concreto en la encuesta:
- **Introductivas:** Se realizan al principio del cuestionario para despertar el interés de la persona entrevistada y crear confianza. Normalmente no suelen tenerse en cuenta en la tabulación y normalmente son de tipo abierto.
- **Relajación:** Para establecer una pausa en la entrevista cuando el cuestionario es muy largo y pueda aburrir al entrevistado. Al igual que la anterior suelen ser preguntas de formato abierto que no suelen tenerse en cuenta en la tabulación.
- **Batería:** Son preguntas de una misma área temática realizadas a continuación de otras como consecuencia de la respuesta dada por el entrevistado.
- **Filtro:** Su objetivo es seleccionar una serie de personas que reúnen unas determinadas características. Si cumplen estas características se les somete a otra serie de preguntas. En el caso contrario se excluye al encuestado de la siguiente pregunta o de la propia encuesta.
- **Evaluación:** Tratan de obtener del entrevistado una respuesta jerarquizada subjetiva sobre las distintas escalas numéricas o verbales que se le indican en la pregunta.
- **Recuerdo:** Tiene por objeto obtener del entrevistado una información basada en el recuerdo de algún hecho pasado.
- **Tarjetas:** Consisten en mostrar al entrevistado una tarjeta en la que aparecen las diferentes opciones que tiene la pregunta formulada. Suelen utilizarse en las encuestas personales y cuando el número de alternativas es muy grande. Los entrevistadores deberán leer las respuestas en voz alta aunque los entrevistados las estén viendo.
- **Control:** Se realizan en ocasiones para comprobar respuestas veraces y coherentes. Puede ser una pregunta específica (por ejemplo, preguntar en qué revista vio anunciado un producto cuando en realidad ese producto no se publicitó en tal revista) o comparar varias preguntas cuyas respuestas podrían ser incompatibles.

- **Amortiguadoras:** se refieren a que cuando estamos preguntando temas escabrosos o pensamos que serán reticentes a contestar, hay que preguntar suavizando la pregunta y no preguntar de modo brusco y directo.

Por ultimo podemos categorizar las preguntas según su contenido:

- Identificación sitúan las condiciones en la estructura social. Ej. Edad, sexo, profesión.
- Acción tratan sobre las acciones de los entrevistados. Ej. ¿Va al cine? ¿fuma?
- Intención indagan sobre las intenciones de los encuestados. Ej. ¿Va a votar?
- Opinión tratan sobre la opinión encuestados sobre determinados temas. Ej. ¿Qué piensa sobre...?
- Información analizan el grado de conocimiento de los encuestados sobre determinados temas.
- Motivos tratan de saber el porqué de determinadas opiniones o actos.

Algunas de las reglas que se toman en cuenta para la formulación de las preguntas son:

- No deben ser excesivamente largos, porque en cuestionarios largos (+100 preguntas) disminuye el porcentaje de respuestas.
- Tiene que ser sencillas y redactadas de tal forma que puedan comprenderse con facilidad (no utilizar términos técnicos).
- No deben incorporar términos morales (juicios de valor).
- Nunca sugerir la respuesta, incitando a contestar más en un sentido que en otra.
- Todas deben referirse a 1 sola idea.
- Todas las que estén dentro de un mismo tema deben ir juntas en el cuestionario en forma de batería.
- No juntar preguntas cuya contestación a 1 de ellas influya sobre la contestación del otro, denominado efecto "halo".

2.9.6.2 Doble Propósito del Cuestionario

El cuestionario, o formulario de la entrevista, tiene dos propósitos principales. Primero, debe traducir los objetivos de la investigación en preguntas específicas cuyas respuestas proporcionarán los datos necesarios para comprobar las hipótesis o indagar el área determinada por los objetivos de la investigación. Para ello, cada pregunta debe comunicar al sujeto la idea o grupo de ideas requerida por los objetivos de la investigación para dar lugar a una respuesta que pueda analizarse en forma de que los resultados satisfagan la meta de estudio. Por otra parte, la pregunta debe cumplir estas dos

funciones con una distorsión mínima de la respuesta que provoca. Es decir, al hacer una pregunta al sujeto, suponemos que éste posee una actitud, una opinión o un conocimiento. Por lo tanto, debe formularse cada pregunta de modo que provoque una respuesta que refleje precisa y completamente la posición de cada sujeto.

La segunda función del cuestionario es ayudar al entrevistador en la tarea de motivar al sujeto para que comunique la información requerida. Como ya vimos, hay muchos factores que determinan la disposición del sujeto para participar en una entrevista. Por supuesto, las capacidades del entrevistador tienen gran importancia para la tarea de motivar al sujeto, pero el mismo cuestionario determina en gran medida el carácter de la relación entrevistador-sujeto y, en consecuencia, la cantidad y la calidad de datos reunidos.

Puesto que el cuestionario se construye sobre la base de los objetivos de investigación, es evidente que su elaboración no puede ser el primer paso de la realización de un proyecto de investigación. La formulación de los objetivos de investigación y la especificidad de los datos requeridos para satisfacerlos deben preceder a la elaboración del cuestionario.

Analizaremos los problemas que parecen más importantes en términos de los criterios citados y, por lo tanto, más significativos para la creación de un instrumento adecuado para reunir datos sociales mediante entrevistas.

2.9.6.3 Lenguaje

Al elaborar un cuestionario, el principal criterio para la elección del lenguaje es que el vocabulario y la sintaxis proporcionen el máximo de oportunidades para una comunicación completa y precisa de las ideas entre el entrevistador y el sujeto. No sólo deben elegirse palabras que se hallen dentro del vocabulario del sujeto, sino que es también preciso, en lo que concierne a sus expresiones coloquiales y clichés, o bien conocerlos y usarlos significativamente, o bien omitirlos por completo. Para decirlo en una forma más simple, el lenguaje del cuestionario debe aproximarse al del sujeto. En la mayoría de las encuestas de secciones transversales de la población se busca la simplicidad, y con razón. No obstante, para un estudio de médicos o abogados serían más apropiados vocabularios diferentes y especializados. Siempre que la población de sujetos es heterogénea, resulta inevitable recurrir a una solución de compromiso.

2.9.6.4 Marco de Referencia

Decir que un cuestionario debe formularse con el lenguaje del entrevistado es algo inequívoco y directo. Sin embargo, igualmente importante, y mucho más difícil, es formular preguntas que tengan en cuenta el marco de referencia en el cual el sujeto considera el tema tratado. No obstante, el cuestionario debe presentar cada tema en forma de vincularlo con las percepciones del sujeto y de que

sea coherente con las nociones de éste acerca de lo que es significativo y lo que no lo es para el problema en cuestión. El desarrollo de un tema de una a otra pregunta no sólo debe satisfacer los criterios de razonamiento y lógica del investigador, también debe satisfacer los del entrevistado. De este modo, el marco de referencia se convierte en otra dimensión en la cual el investigador debe partir del punto donde está el sujeto, debe orientarse hacia el sujeto.

El marco de referencia del sujeto también puede ser importante para determinar si estará dispuesto a comunicar un tipo dado de información. El sujeto puede resistirse a comunicar una información si no ve la relación entre una pregunta y su percepción de los objetivos de la investigación. De este modo, el sujeto de una encuesta que habló libremente sobre política exterior puede interrumpirse de pronto cuando se le pregunta su edad o educación. Aunque es improbable que estas preguntas parezcan amenazadoras, no se ajustan a su percepción de las necesidades de la investigación.

2.9.6.5 Nivel de Información

Las preguntas deben formularse en forma de que se vinculen significativamente con el nivel de información del sujeto en ese momento. No debe partirse de supuestos no realistas respecto de la capacidad del sujeto o de la información que posee. La importancia de esta norma para la elaboración del cuestionario reside en que cuando el entrevistador, con la autoridad de su rol, plantea una pregunta al sujeto se supone que éste posee una respuesta adecuada y que si no puede responder queda de alguna manera desacreditado. Por ejemplo, si en un cuestionario que se ocupa de las actitudes públicas hacia problemas de la energía atómica, un entrevistador pregunta a un sujeto: “¿Qué precauciones son adecuadas para que un técnico maneje isótopos radiactivos?” La reacción inmediata y muy común del sujeto será la incomodidad y el resentimiento porque se le hace una pregunta a la que no puede responder. El investigador no sólo habrá perdido la respuesta a una pregunta, sino que deberá pagar un alto precio en términos de una menor motivación para comunicarse con el sujeto. Por supuesto, existe la posibilidad de que el entrevistado, al sentirse obligado a demostrar su conocimiento, presuma conocimientos que no posee.

A veces se llama a este problema error de pericia; es decir, el error de adscribir al sujeto un grado de pericia en un campo particular que en realidad no posee. Estas preguntas “para peritos” pueden exigir al sujeto que se dedique a un incómodo autoanálisis, que verbalice material no analizado ni verbalizado y que por lo tanto, no es consciente.

2.9.6.6 Aceptación Social

Otra característica del cuestionario centrado en el sujeto consiste en subrayar la aceptabilidad de una amplia gama de respuestas. Ninguna pregunta debe enfrentar al sujeto con la necesidad de dar una respuesta socialmente inaceptable. Si esperamos que el sujeto conteste libre y espontáneamente,

debemos ayudarlo a sentir que toda la gama de posibles respuestas es aceptable; no sólo aceptable para el entrevistador, sino para él mismo y de acuerdo con sus propias normas.

Para proporcionar una gama de respuestas que satisfaga los criterios de aceptabilidad social del sujeto es necesario formular bien las preguntas. En un sentido más amplio podría decirse que las preguntas nunca deben constituir una amenaza para el yo del sujeto. Tal amenaza puede presentarse si se le pide que dé una respuesta que siente socialmente inaceptable, o si se lo coloca en una posición en que se siente menos informado de lo que debería estar.

2.9.6.7 Preguntas que Sugieren la Respuesta

Las preguntas deben formularse en forma de que no contengan ninguna sugerencia respecto de la forma más apropiada de contestarlas. Por ejemplo, una pregunta destinada a revelar actitudes generales hacia el control de rentas. Una formulación de la misma pregunta evidentemente tendenciosa sería: ¿Diría usted que está a favor del control de renta, no es así? Este tipo de parcialidad es tan fácil de reconocer que la evitamos sin inconveniente.

Una pregunta puede estimular una respuesta particular mediante la asociación de una de las posibles contestaciones con una meta tan deseable que difícilmente pueda negársela. De este modo, la pregunta: “¿Está Ud. a favor o en contra de los aumentos de impuestos destinados a prepararse para los peligros de la guerra?” Asocia a los aumentos de los impuestos con la defensa contra un posible ataque y supone que una respuesta negativa refleja una indiferencia ante la amenaza de ataque. Aun cuando el sujeto se permita una respuesta no estructurada, la pregunta influye en gran medida sobre su respuesta. Si sólo se le da la alternativa de la aceptación o el rechazo, como ocurre en la mayoría de las compulsas de la opinión pública, el efecto de la pregunta es todavía más grave.

2.9.6.8 La Idea Única

Las preguntas deben limitarse a una única idea o a una única referencia. Los problemas encontrados en esta área están ejemplificados en la siguiente pregunta: “¿Está Ud. a favor o en contra del seguro para desocupados y los planes de pensiones?” Muchas respuestas no permitirían al investigador determinar si el entrevistado responde a uno o ambos ítem mencionados en la pregunta. La formulación más aceptable de una pregunta en este sentido dependería de la medida de especificidad de los objetivos de la investigación. Si el propósito es determinar la actitud del sujeto hacia las pensiones y el seguro por desocupación en forma específica, sería necesario plantear dos preguntas, cada una relacionada con una de las dos proposiciones.

2.9.6.9 Secuencia de las Preguntas

Independientemente de la formulación de las preguntas en particular, el investigador debe considerar su distribución en el cuestionario. La secuencia de las preguntas debe determinarse principalmente de

acuerdo con el proceso de la entrevista y no por el proceso de la investigación. Un cuestionario bien diseñado facilita el pasaje fluido del sujeto de uno a otro ítem y a menudo lo lleva a anticipar la pregunta siguiente, guiándose por lo que le parece la lógica intrínseca del tema.

La secuencia de las preguntas también puede determinarse por lo que se llamó el Enfoque de embudo. Ello se refiere al procedimiento de plantear primero las preguntas más generales o menos restringidas y seguir luego con ítem más limitados. De este modo, en una secuencia de preguntas cada vez más específicas. El propósito de la secuencia en embudo es evitar que las primeras preguntas condicionen las posteriores y desde las primeras preguntas abiertas determinen en alguna medida el marco de referencia del sujeto.

En muchos casos, las primeras dos o tres preguntas de un cuestionario tienen una doble función. Por una parte, se las incluye para obtener información respecto de objetivos específicos de la investigación, pero también ayudan a preparar y motivar al sujeto. Durante las primeras dos o tres preguntas, el entrevistador, mediante sus sondeos, sus reacciones ante las respuestas y su conducta general ante el sujeto, lo prepara para el papel que se espera que desempeñe en la entrevista.

Además de su propósito de orientación preparación, las primeras preguntas sirven también para motivar al sujeto a participar más completamente, comprometiéndose en el tema tratado. De hecho, las primeras preguntas pueden establecer el tono de toda la entrevista.

2.9.6.10 La Forma de la Pregunta

Hasta ahora consideramos el problema de la formulación de las preguntas sin considerar el problema de la forma de la respuesta; es decir, si el sujeto debe contestar con sus propias palabras o si debe seleccionar entre una serie de categorías preestablecidas la respuesta que más se acerca a su opinión. Las preguntas del primer tipo se llaman “abierta” o “sin limitaciones”; el segundo tipo de pregunta es el “limitado” o “cerrado”. La pregunta abierta es aquella en que se estructura el tema para el sujeto, pero se deja a su cargo la tarea de responder con sus propias palabras, con libertad para estructurar la respuesta según le parezca conveniente y hablar todo lo que desee. Un ejemplo de pregunta abierta es el siguiente: “¿Cuál es su opinión ante la posibilidad de que en esta fábrica trabajen juntos los negros y los blanco?” La pregunta cerrada contiene en ella las posibles respuestas en forma de que el sujeto debe limitarse a seleccionar la categoría que más se aproxime a su posición. Un ejemplo de pregunta cerrada es el siguiente: “¿Piensa usted que su ingreso de este año será mayor, menor o aproximadamente igual al del año pasado?”

En términos generales, la pregunta cerrada se adapta bien a situaciones en que: 1) hay sólo un marco de referencia a partir del cual el sujeto puede contestar a la pregunta; 2) dentro de este único marco de referencia hay una gama conocida de posibles respuestas, y 3) dentro de esta gama hay puntos de elección claramente definidos que representan con precisión la posición de cada sujeto.

Un ejemplo del tipo cerrado es la pregunta: “¿Podría decir Ud. si su ingreso actual fue mayor, menor o aproximadamente igual al del año pasado?” En esta pregunta se pide al sujeto que compare dos hechos que conoce. El marco de referencia está limitado a la comparación de los ingresos de dos años, y las opciones son claras. Las preguntas abiertas tienen muchas ventajas, que resultan del hecho de estimular al sujeto para que estructure su respuesta como desee.

El intercambio relativamente libre entre el entrevistador y el sujeto, que es característico de la pregunta abierta, permite al investigador descubrir si el sujeto comprende con claridad la pregunta planteada. En cambio, una vez que el sujeto seleccionó una de las alternativas presentadas en una pregunta cerrada, el entrevistador sólo puede suponer que el sujeto comprendió la pregunta y eligió la alternativa que más se aproximaba a su posición.

2.9.6.11 La Prueba Previa

Independientemente del ingenio del investigador para formular sus preguntas y diseñar su cuestionario, necesita probarlos con sujetos antes de iniciar los trabajos de campo concretos. En un sentido, la prueba previa es en sí misma un estudio en miniatura. Su primera función es probar el cuestionario desde la perspectiva de la investigación. Deben analizarse las entrevistas para ver si las respuestas satisfacen los objetivos de la investigación. A menudo algunas de las “mejores preguntas” del investigador no logran provocar el tipo de respuesta que satisface los objetivos del estudio. Habrá mayores probabilidades de alcanzar las metas de la investigación cuando se analizan estas entrevistas de ensayo de acuerdo con ellos. A menudo la prueba previa requiere importantes revisiones de las preguntas, y es necesario efectuar varias pruebas de este tipo antes de hallar un cuestionario aplicable.

Un segundo objetivo de la prueba previa es determinar la medida en que el cuestionario satisface el criterio de orientación hacia el sujeto en todos sus aspectos. ¿Favorece el cuestionario la relación apropiada con los sujetos? ¿Comprenden éstos las preguntas? ¿Pueden plantearse las preguntas sin necesidad de explicarlas o volver a formularlas? No hay pruebas exactas para estas características. En este punto, es sumamente útil la ayuda de entrevistadores experimentados para obtener evaluaciones subjetivas del cuestionario.

2.9.6.12 Problemas más Usuales en los Cuestionarios

- Preguntas con alto grado de ambigüedad.
- Mala estructuración de las preguntas.
- Mala secuenciación de las preguntas.
- Inadecuada elección de muestras de población.
- Cuestionarios con excesivo número de ítems.

- Cuestionarios elaborados sin pruebas de validación- piloto
- Utilización errónea de escalas de medida.
- Errores en la codificación de las preguntas.

2.9.7 Tratamiento y Análisis de la Información

Se han descrito las distintas técnicas de obtención de información. Ahora bien, es necesario analizar esta información, generalmente abundante, y convertirla en conclusiones. Presentamos, pues, una serie de etapas a seguir a la hora de codificar, analizar, tratar e interpretar la información recogida a través de los cuestionarios:

- **Revisión de los cuestionarios.** En esta fase se trata de identificar y corregir las posibles fuentes de error. Para ello se revisan los cuestionarios buscando ambigüedades, como respuestas no legibles o en las que no se sabe qué cuadro se ha marcado, omisiones o incoherencias.
- **Codificación, clasificación, recopilación de cuestionarios y preparación de la base de datos.** Una vez depurados los cuestionarios se procede a la codificación de las preguntas para posibilitar el tratamiento informático. La codificación tiene por objeto sistematizar y simplificar la información procedente de los cuestionarios. En otras palabras, consiste en el establecimiento de grupos que permitan clasificar las respuestas. Conviene recordar que la codificación ha de ser lo más simple e intuitiva posible, siempre claramente unívoca, preferentemente numérica y lo más estandarizada posible. Con los códigos de respuesta asignados a cada variable, construiremos la base de datos que reflejará toda la información obtenida.
- **Análisis de las preguntas.** Una vez organizados los datos, comienza el análisis propiamente dicho. El primer paso es estudiar cada pregunta aislada, luego las preguntas por subgrupos y las relaciones entre pares de preguntas, y por último, se estudiarán las relaciones entre todas las preguntas.
- **Conocer la evolución del estudio.** Como la percepción de los problemas cambia pareja a cómo evoluciona la satisfacción de los usuarios de un servicio de información-documentación, hay que establecer un sistema de vigilancia de los principales parámetros de percepción, que nos avise cuando haya un nuevo problema sentido, cuando las cosas no van bien desde el punto de vista de los usuarios, etc. En consecuencia, es necesario desarrollar sondeos y estimaciones durante distintos momentos del análisis.

- **Anticipación del estudio de predicción.** No es suficiente con conocer la evolución del estudio o análisis, hemos de saber además hacia dónde nos lleva. Para ello, hemos de hacer proyecciones de futuro sobre los problemas, las necesidades, las demandas y la satisfacción de los usuarios objeto del estudio. El análisis ha de permitir asociar probabilidades y formular predicciones a corto plazo. En definitiva, ha de permitir anticiparnos y modificar nuestras actuaciones.
- **Presentación final de la información.** La información que se presente ha de ser muy clara, incluyéndose únicamente aquella que sea relevante para las decisiones que hay que tomar. A continuación, exponemos algunas recomendaciones útiles que pueden ayudar en presentación final:
 - Proporcionar la menor cantidad posible de datos. Eso sí, los que exponamos en el análisis deben de ser los más relevantes.
 - Hacer los comentarios e interpretaciones al final de cada grupo de resultados, diferenciando muy claramente lo que es un resultado de lo que es una interpretación personal.
 - Utilizar un lenguaje sencillo, sin posibilidad de mal interpretaciones.
 - La información sobre la metodología y los criterios de análisis para el estudio han de estar bien diferenciados del resto de la información.
 - En general, no se deben desagregar resultados, excepto que se haya demostrado una relación significativa con la variable de estratificación.
 - En estudios efectuados sobre muestras representativas hay que proporcionar las estimaciones por intervalo.
 - No se han de proporcionar fracciones decimales irrelevantes. En la mayoría de los casos con un decimal es más que suficiente.
 - No debemos dar información redundante.
 - Siempre que sea posible, se han de proporcionar los datos en forma gráfica. Además, en todas las tablas y gráficos presentados han de incluirse los nombres de las variables y las unidades de medida.
 - Al final del estudio se ha de incluir un breve resumen de los resultados y de las conclusiones. En algunos casos es conveniente plantear recomendaciones operativas.
 - Si existen niveles diferentes de usuarios, se debe seleccionar y segmentar la información.

2.9.8 Software y Herramientas para Procesar Encuestas

Existen en el mercado un sinnúmero de herramientas de software profesionales para realizar el procesamiento de la encuesta de forma eficiente y productiva. El tipo de software a utilizar dependerá en gran medida de la metodología de aplicación del instrumento en campo, es decir, si la encuesta será autoadministrada o administrada mediante entrevista personal; igualmente, el tipo de software dependerá del medio de captura, bien sea papel (PAPI), Web (CAWI), entrevista telefónica (CATI) o dispositivos móviles (CAPI).

Algunas herramientas ofrecen una facilidad llamada OLAP, lo cual permite almacenar los datos en un formato cúbico y así poder rotar las dimensiones de análisis para obtener múltiples vistas de la información y poder analizar fácilmente cada pregunta del cuestionario por sus variables de análisis (sexo, ciudad, edad, estrato social, entre otros).

2.9.9 Encuestas e Inteligencia de Negocio

En todas aquellas operaciones estadísticas que implican el uso de cuestionarios es de gran utilidad tener una visión global de los resultados, lo que permite detectar posibles errores o incoherencias de un solo vistazo, por ello la creación de una solución de inteligencia de negocio, es una efectiva solución. Se implementa con una arquitectura basa en un almacén de datos con una dimensión un tanto especial, a la que hemos llamado dimensión encuesta. Esta dimensión engloba todas las medidas del cuestionario, así como posibles miembros calculados derivados de las mismas. Planteamos las ventajas e inconvenientes del uso de este tipo de dimensión y su aplicación en encuestas económicas y demográficas.

2.9.10 Modelización de la Encuesta

Para modelizar la Encuesta partimos de los datos recogidos en el cuestionario y analizando los bloques en los que se divide podemos diferenciar claramente distintos tipos de datos:

- Datos de identificación y perfil de la empresa: NIDENT, provincia, municipio, razón social, naturaleza jurídica, titularidad, tamaño de la empresa, actividad económica, titularidad, pertenece a grupo de sociedades, tiene participación de capital extranjero, control sobre alguna empresa extranjera, entre otros.
- Datos de personal y horas trabajadas: nº medio de ocupados (remunerados y no remunerados), horas trabajadas (por personal remunerado y no remunerado).
- Ingresos de explotación: cifra de negocios, trabajos realizados para el inmovilizado, subvenciones de explotación, otros ingresos de explotación, entre otros.

- Existencias: existencias finales e iniciales de materias primas, otros aprovisionamientos, entre otros.
- Gastos de explotación: consumos, gastos de personal, servicios exteriores, dotaciones para amortización del inmovilizado, entre otros.
- Inversión: inversión en activos materiales e inversión en activos inmateriales.
- Impuestos sobre la actividad productiva: IVA, IAE, impuestos especiales, entre otros.
- Resultado del ejercicio
- Información complementaria relativa al empleo: asalariado a tiempo parcial y horas trabajadas por asalariados a tiempo parcial, trabajadores a domicilio, entre otros.
- Estructura de la cifra de negocios: actividad industrial principal, otras actividades industriales,...
- Subcontratación de servicios industriales: ingresos, pagos.
- Datos de localización: establecimiento industrial único o empresa multilocalizada

Como hemos comentado los hechos vendrían dados por medidas, es decir, por valores numéricos, así que directamente todo lo que son datos de identificación y otras características de la empresa o del establecimiento industrial no numéricas conformarían una o varias dimensiones.

Respecto a las medidas, en la mayoría de los casos se ven claramente definidos por distintas dimensiones. Por ejemplo: los datos de personal y horas trabajadas por la dimensión tipo de personal (remunerado y no remunerado), la inversión por el tipo de activo (materiales/inmateriales y dentro de cada tipo su desglose), los ingresos de explotación por el tipo de ingreso, etc.

Cada una de las métricas asociadas a distintas dimensiones dan lugar a tablas de hechos distintas bien diferenciadas.

Este análisis daría lugar a un modelo complejo en el que tendríamos distintas tablas de hechos con sus dimensiones asociadas.

Y este modelo habría de analizarse en distintos cubos OLAP, normalmente tantos como tablas de hechos o bien podríamos unir varias tablas de hechos en un mismo cubo, dando lugar a multitud de celdas vacías, tantas como resultaran de los cruces de dimensiones imposibles. Para poder extraer información deberemos conocer que cruces de dimensiones tienen sentido para cada medida.

Si quisiéramos realizar el cálculo de ratios, coeficientes o cualquier otra medida que combine datos de distintas tablas de hechos, por ejemplo comparar variables económicas con variables de personal:

salario medio, coste salarial medio, productividad, etc. Tendríamos que crear una nueva tabla que las uniera o un nuevo cubo en el que utilizaríamos ambas tablas de hechos.

2.9.11 Sistema de Información de Encuesta Simple

El proceso de Encuesta Simple, es muy utilizado por organizaciones en su Unidad de Mercadeo, con el fin de determinar una tendencia de gusto con respecto a un producto o servicio. Este proceso, es soportado por el Sistema de Información de Encuestas Simples. Este Sistema fue desarrollado por Tian Consultores para uno de sus clientes.

Es un Sistema de Información, que permite a diseñar, aplicar y administrar encuestas simples que permiten medir la satisfacción de sus clientes con respecto a un producto o servicio. Además, este sistema cuenta con una aplicación móvil que puede ser ejecutada sobre distintos sistemas operativos móviles.

La siguiente ilustración presenta el flujo de actividades del proceso del Sistema de Encuestas Simple descrito:

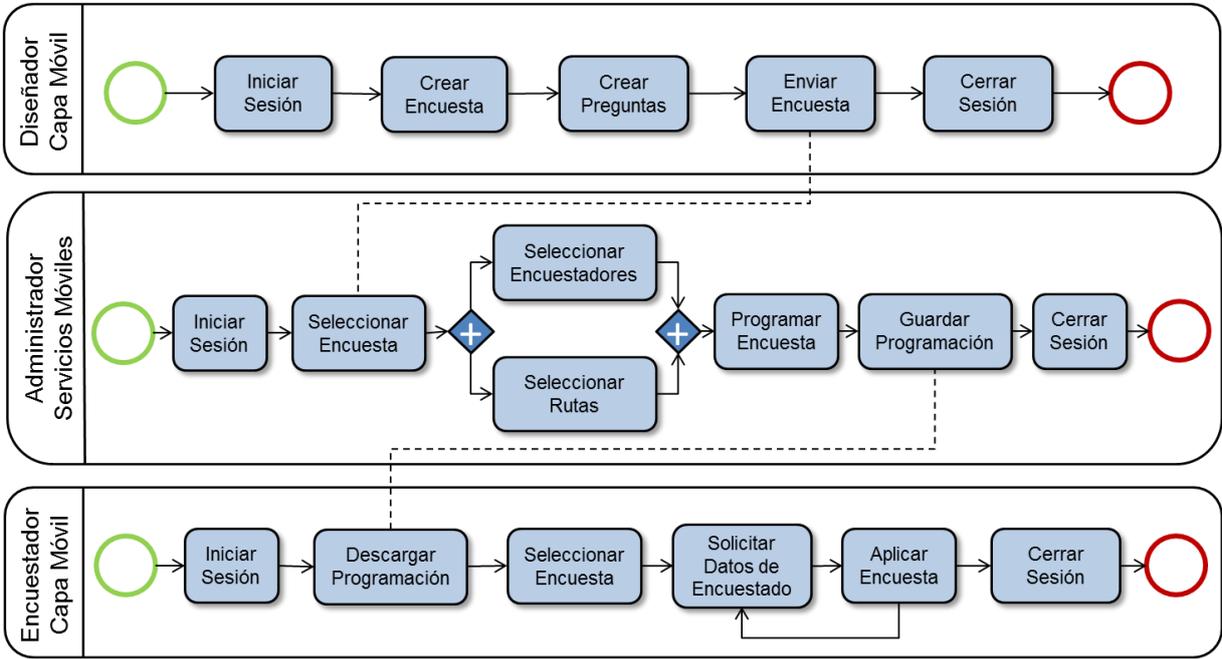


Ilustración 20: Flujo de Actividades del Proceso de Encuesta Simple

Fuente: Documentación del sistema de información de Encuestas Simples

El sistema se ha desarrollado utilizando tecnologías Web como HTML, JavaScript, CSS y PHP; el sistema manejador de base de datos Oracle, el sistema manejador de bases de datos móviles SQLite, y el framework PhoneGap.

Está basado en una arquitectura que consta de dos capas: la Capa Móvil y la Capa de Servicios Móviles. La primera de estas capas, permite a un usuario, según su rol de trabajo, crear encuestas o aplicar las encuestas existentes. La Capa de Servicios Móviles, mantiene el control de la información de los resultados obtenidos a través de las encuestas que han sido aplicadas, permite asignar a los encuestadores las encuestas que deben aplicar y lleva el control general del Sistema de Información (Ver Ilustración 21).

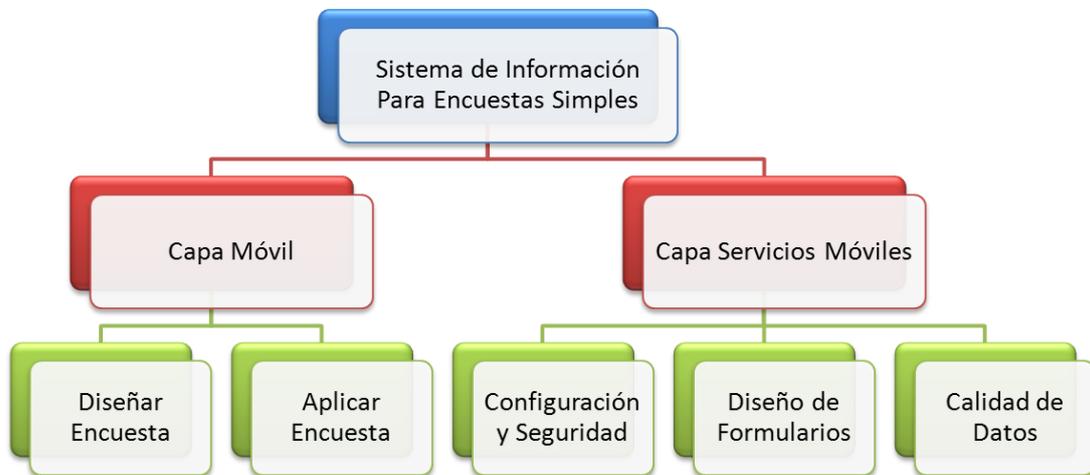


Ilustración 21: Sistema de Información Encuesta Simple

Fuente: Documentación del sistema de información de Encuestas Simples

La base de datos de la Capa de Servicios Móviles, permite manejar la información referente a las distintas encuestas, la configuración del sistema y la visualización de los resultados obtenidos.

La base de datos de la Capa Móvil, solo contiene la información necesaria para el encuestador que ha iniciado sesión, con el fin de no sobrecargar al dispositivo móvil con información, tomando en cuenta las capacidades de almacenamiento que éste puede tener.

La siguiente ilustración presenta la arquitectura descrita anteriormente:

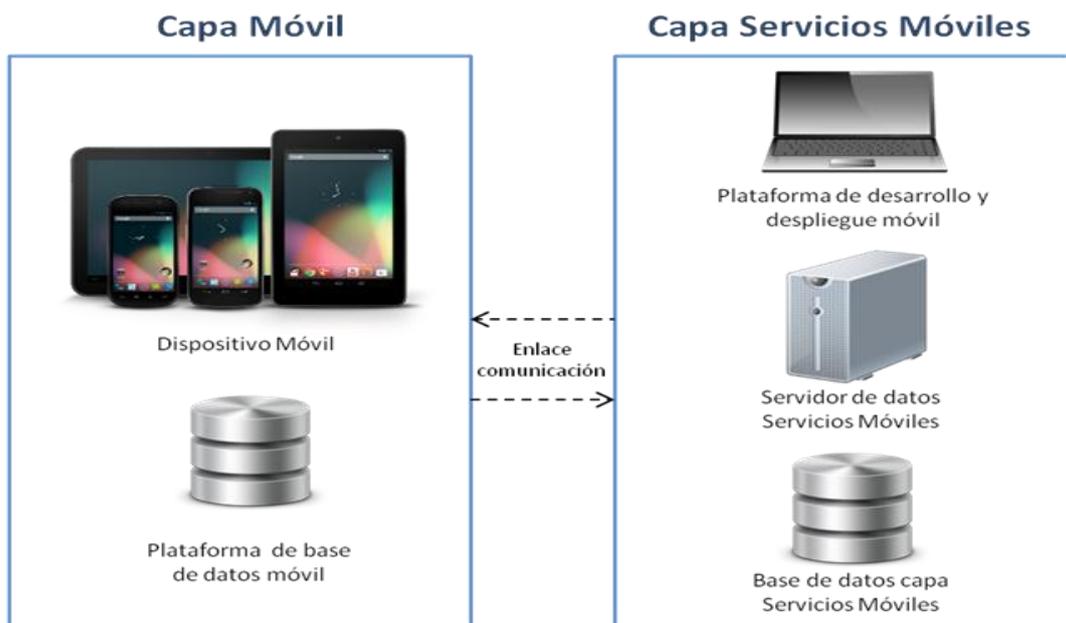


Ilustración 22: Flujo de Actividades del Proceso de Encuesta Simple

Fuente: Documentación del sistema de información de Encuestas Simples

Una organización que haga uso de este Sistema de Información, puede disminuir el tiempo que lleva todo el proceso creación y de almacenamiento de los resultados de las encuestas. De la base de datos resultante se puede extraer información que sirven para apoyar y ejecutar las decisiones operativas que conducen las actividades básicas, pero no sirven para realizar análisis más avanzados, incluso de tipo estratégico ya que no están diseñadas para apoyar este tipo de tareas. Además los resultados son difíciles de interpretar, y los tiempos de respuesta y la disposición de los datos, no permite tener una visión global de los resultados, dificultando con esto la toma de decisiones.

CAPÍTULO 3: MARCO METODOLÓGICO

Como se expuso anteriormente, el concepto de Almacén de Datos llegó de la mano de Bill Inmon y Ralph Kimball. Ambos pensaron en un único repositorio de información para poder integrar y explotar información de diversos sistemas fuentes. Pero, más allá de esta generalización conceptual, cada uno decidió hacerlo a su manera.

Para entender las diferencias entre ambos enfoques, es necesario en primer lugar tener claro la diferencia entre Almacén de Datos y Bodega de Datos:

- **Definición de Almacén de Datos:** Un Almacén de Datos proporciona una visión global, común e integrada de los datos de la organización, independiente de cómo se vayan a utilizar posteriormente por los consumidores o usuarios. Normalmente en el almacén de datos habrá que guardar información histórica que cubra un amplio período de tiempo. Pero hay ocasiones en las que no se necesita la historia de los datos, sino sólo sus últimos valores, siendo además admisible generalmente un pequeño desfase o retraso sobre los datos operacionales. En estos casos el almacén se llama almacén operacional.
- **Definición de Bodega de Datos:** Podemos entender una Bodega de Datos como un subconjunto de los datos del Almacén de Datos con el objetivo de responder a un determinado análisis, función o necesidad y con una población de usuarios específica. Al igual que en un Almacén de Datos, los datos están estructurados en modelos de estrella o copo de nieve y una Bodega de Datos puede ser dependiente o independiente de un Almacén de Datos.

¿Qué diferencia existe entonces entre una Bodega de Datos y un Almacén de Datos? Su alcance. La Bodega de Datos está pensada para cubrir las necesidades de un grupo de trabajo o de un determinado departamento dentro de la organización. Es el almacén natural para los datos departamentales. En cambio, el ámbito del Almacén de Datos es la organización en su conjunto. Es el almacén natural para los datos corporativos comunes.

Teniendo en cuenta esto, vamos a analizar qué es lo que propone cada uno de estos enfoques.

3.1 Método Ascendente (Ralph Kimball)

En el Modelo Dimensional se constituyen modelos de tablas y relaciones con el propósito de optimizar la toma de decisiones, con base en las consultas hechas en una base de datos relacional que están ligadas con la medición o un conjunto de mediciones de los resultados de los procesos de negocio.

El Modelo Dimensional es una técnica de diseño lógico que tiene como objetivo presentar los datos dentro de un marco de trabajo estándar e intuitivo, para permitir su acceso con un alto rendimiento. Cada Modelo Dimensional está compuesta por una tabla con una llave combinada, llamada tabla de

hechos, y con un conjunto de tablas más pequeñas llamadas tablas de dimensiones. Los elementos de estas tablas se pueden definir de la siguiente manera:

- **Hechos:** es una colección de piezas de datos y datos de contexto. Cada hecho representa una parte del negocio, una transacción o un evento.
- **Dimensiones:** es una colección de miembros, unidades o individuos del mismo tipo.
- **Medidas:** son atributos numéricos de un hecho que representan el comportamiento del negocio relativo a una dimensión.

Para Kimball el Almacén de Datos es un conglomerado de todas las Bodegas de Datos dentro de una empresa, siendo una copia de los datos transaccionales estructurados de una forma especial para el análisis, de acuerdo al Modelo Dimensional (no normalizado) que incluye, como ya vimos, las dimensiones de análisis y sus atributos, su organización jerárquica, así como los diferentes hechos de negocio que se quieren analizar. Por un lado tenemos tablas para las representar las dimensiones y por otro lado tablas para los hechos.

Las diferentes Bodegas de Datos están conectadas entre sí por la llamada bus estructura, que contiene los elementos anteriormente citados a través de las dimensiones conformadas (que permiten que los usuarios puedan realizar consultas conjuntas sobre las diferentes Bodegas de Datos, pues este bus contiene los elementos en común que las comunican). Una dimensión conformada puede ser, por ejemplo, la dimensión cliente, que incluye todos los atributos o elementos de análisis referentes a los clientes y que puede ser compartida por diferentes Bodegas de Datos (ventas, pedidos, gestión de cobros, etc.).

Al final el Almacén de Datos Corporativo no es más que la unión de las diferentes Bodegas de Datos. Esta característica le hace más flexible y sencillo de implementar, pues podemos construir una Bodega de Datos como primer elemento del sistema de análisis, y luego ir añadiendo otros que comparten las dimensiones ya definidas o incluyen otras nuevas. En este sistema, los procesos ETC extraen la información de los sistemas operacionales y los procesan igualmente en el área intermedia, realizando posteriormente el llenado de cada uno de la Bodega de Datos de una forma individual, aunque siempre respetando la estandarización de las dimensiones (dimensiones conformadas).

Kimball sugiere utilizar una metodología ascendente, donde la información se extrae de los sistemas transaccionales para ser cargada en diferentes Bodegas de Datos cada uno de los cuales son independientes, están modelados en forma dimensional y tienen foco departamental.

Kimball defiende su metodología mencionando que es mejor usar un método ascendente a la hora de diseñar un Almacén de Datos, es decir primero crear las Bodegas de Datos y después el Almacén de Datos (Ver Ilustración 23).

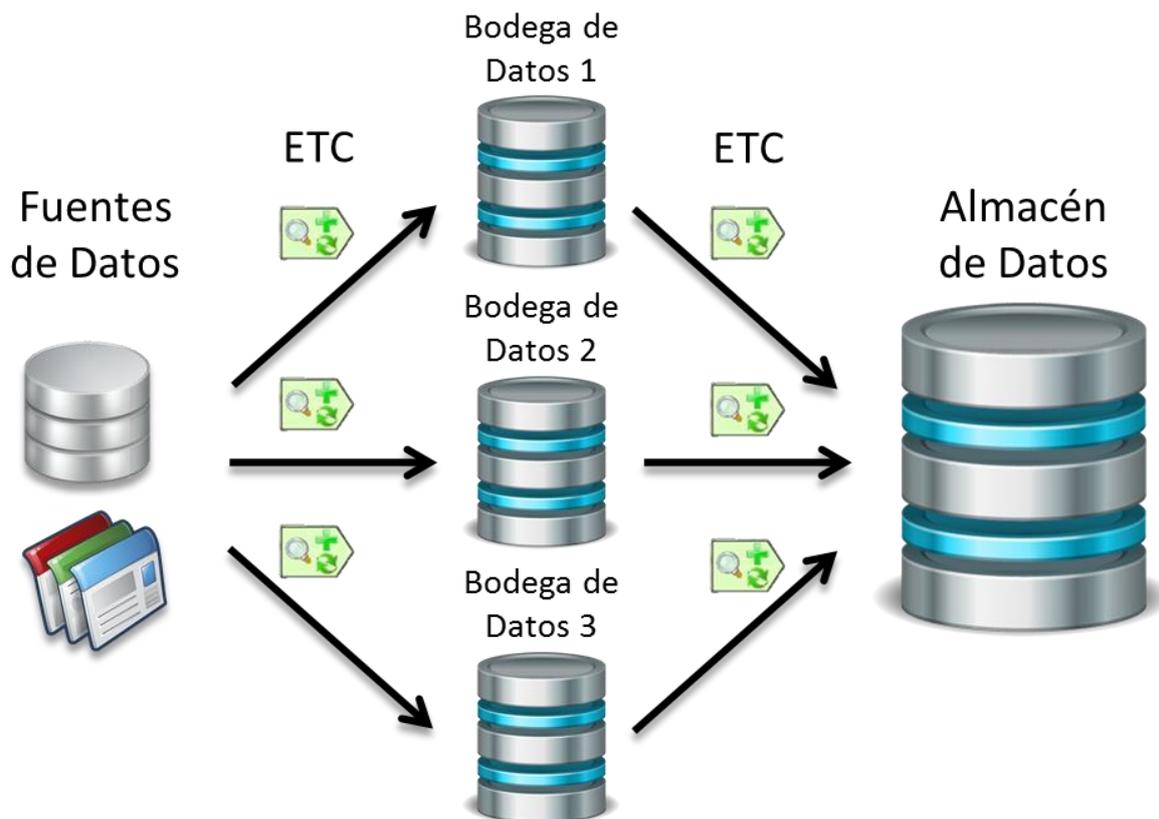


Ilustración 23: Método Ascendente

Fuente: <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/i-data-warehousing-investigacion-y-sistematizacion-concepto-13>

3.1.1 Ciclo de Vida Dimensional del Negocio

La metodología de Kimball, se basa en lo que se denomina Ciclo de Vida Dimensional del Negocio (Ver Ilustración 24). Esta metodología es considerada una de las técnicas favoritas a la hora de construir un Almacén de Datos.

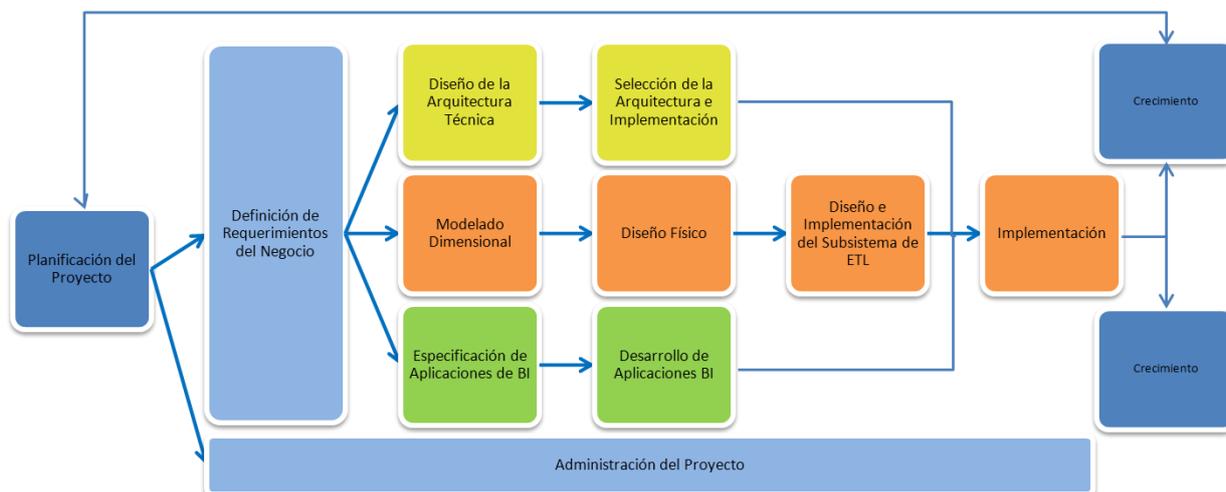


Ilustración 24: Ciclo de Vida Dimensional del Negocio

Fuente: Kimball et al., The Data Warehouse Lifecycle Toolkit. New York, Wiley, 1998

A continuación se hace una descripción de tallada de las fases que fueron expuestas en la Ilustración 24:

- **Planificación del Proyecto:** busca identificar la definición y el alcance que tiene el proyecto del Almacén de Datos. Esta etapa se concentra sobre la definición del proyecto, donde, a nivel de planificación, se establece la identidad del mismo, el personal, desarrollo del plan de proyecto, el seguimiento y la monitorización.
- **Definición de los Requerimientos del Negocio:** es un factor determinante en el éxito de un proceso del Almacén de Datos. Los diseñadores de los Almacenes de Datos deben tener en claro cuáles son los factores claves que guían el negocio para determinar efectivamente los requerimientos y traducirlos en consideraciones de diseño apropiadas.
- **Modelado Dimensional:** se comienza con una matriz donde se determina la dimensionalidad de cada indicador para luego especificar los diferentes grados de detalle dentro de cada concepto del negocio.
- **Diseño Físico:** se centra en la selección de las estructuras necesarias para soportar el diseño lógico. Un elemento principal de este proceso es la definición de estándares del entorno de la base de datos. La indexación y las estrategias de particionamiento se determinan en esta etapa.
- **Diseño y Desarrollo de la presentación de datos:** tiene como principales actividades la extracción, transformación y carga (ETC). Estas actividades son altamente críticas ya que tienen que ver con la materia prima del Almacén de Datos que son los datos.

- **Diseño de la arquitectura técnica:** en esta fase se deben tener en cuenta tres factores: los requerimientos de negocio, los actuales entornos técnicos, y las directrices técnicas y estratégicas futuras planificadas por la compañía, lo que permitirá establecer el diseño de la arquitectura técnica del entorno del Almacén de Datos. El proceso de diseño de la arquitectura técnica está compuesto de 8 pasos:
 1. Establecer un grupo de trabajo de arquitectura
 2. Requisitos relacionados con la arquitectura
 3. Documento de requisitos arquitectónicos
 4. Desarrollo de un modelo arquitectónico de alto nivel
 5. Diseño y especificación de los subsistemas
 6. Determinar las fases de aplicación de la arquitectura
 7. Documento de la arquitectura técnica
 8. Revisar y finalizar la arquitectura técnica

- **Selección de productos e instalación:** se evalúa y selecciona cuales son los componentes necesarios específicos de la arquitectura (plataforma de hardware, motor de la BD, herramienta de ETC, etc.). Luego de realizar la instalación de los componentes previamente evaluados y seleccionados, se recomienda una serie de premisas:
 - Comprender el proceso de compras corporativas
 - Elaborar una matriz de evaluación del producto
 - Realizar la investigación de mercados
 - Filtrar opciones y realizar evaluaciones más detalladas
 - Manejo de un prototipo
 - Selección del producto, instalación y negociación

- **Especificación de Aplicaciones para usuario finales:** se identifican los roles o perfiles de usuarios para los diferentes tipos de aplicaciones necesarias en base al alcance de los perfiles detectados.

- **Desarrollo de aplicaciones para usuario finales:** involucra configuraciones de los metadatos y construcción de reportes específicos.

- **Implementación:** representa el correcto funcionamiento de la tecnología, los datos y las aplicaciones de usuarios finales accesibles para el usuario del negocio.

- **Mantenimiento y crecimiento:** se basa en la necesidad de continuar con las actualizaciones de forma constante para así lograr la evolución de las metas por conseguir.
- **Gestión del proyecto:** asegura que todas las actividades del ciclo de vida se lleven a cabo de manera sincronizada.

3.2 Método Descendente (Bill Inmon)

Según Inmon (1990) el Almacén de Datos se nutre de un solo ETC, pero en este caso el Almacén de Datos no está modelado dimensionalmente, sino que está en tercera forma normal (3NF). Así, el creador de este modelo entiende que esta forma es mucho más rica y adaptable que el modelo de Kimball. Una vez que tenemos el Almacén de Datos generado de esta manera, se pueden crear las Bodegas de Datos para las áreas de negocio que necesitemos, y además lo podríamos utilizar para cualquier otro tipo de sistema de toma de decisiones como por ejemplo sistemas expertos, o minería de datos (Ver Ilustración 25).

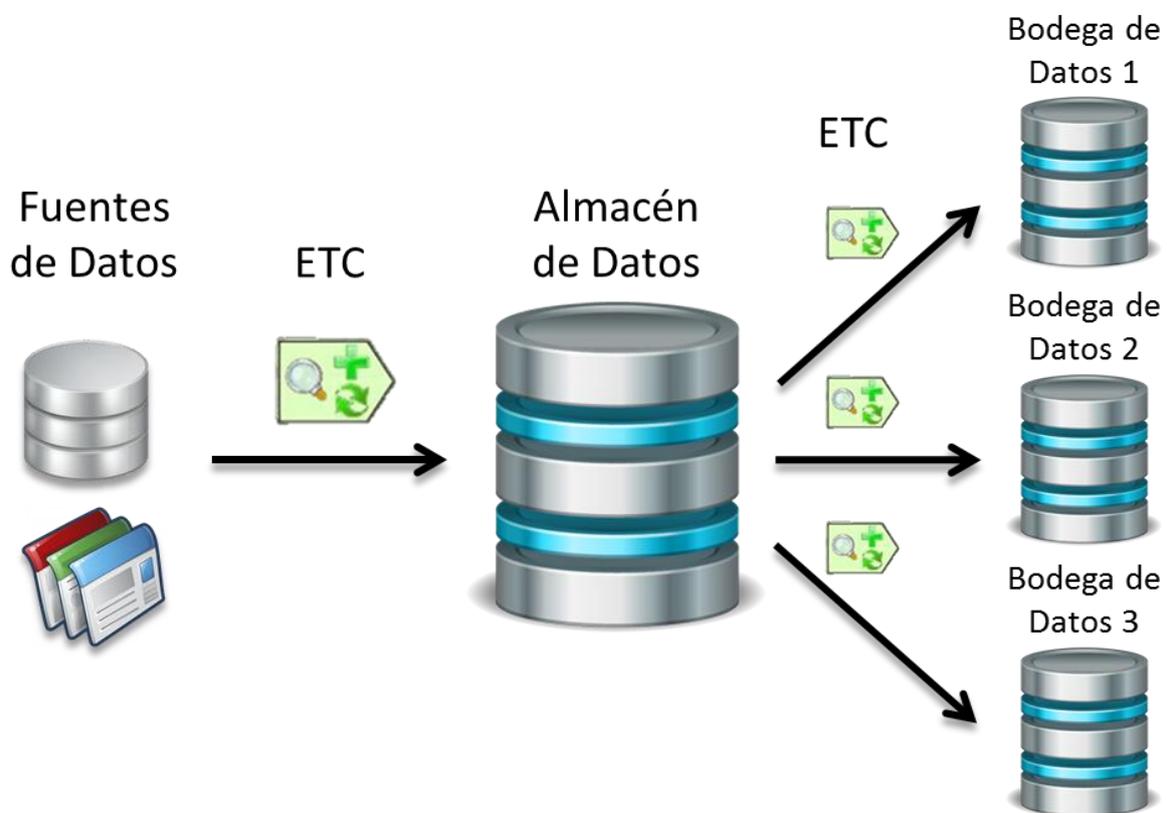


Ilustración 25: Método Descendente

Fuente: http://www.dataprix.com/files/uploads/250image/HEFESTO%20v2_0/data%20mart%20-%20top%20down.png

Bill Inmon ve la necesidad de transferir la información de los diferentes OLTP (Sistemas Transaccionales) de las organizaciones a un lugar centralizado donde los datos puedan ser utilizados para el análisis. Insiste además en que ha de tener las siguientes características:

- **Orientado a temas:** los datos en la base de datos están organizados de manera que todos los elementos de datos relativos al mismo evento u objeto del mundo real queden unidos entre sí.
- **Integrado:** la base de datos contiene los datos de todos los sistemas operacionales de la organización, y dichos datos deben ser consistentes.
- **No volátil:** la información no se modifica ni se elimina, una vez almacenado un dato, éste se convierte en información de sólo lectura, y se mantiene para futuras consultas.
- **Variante en el tiempo:** los cambios producidos en los datos a lo largo del tiempo quedan registrados para que los informes que se puedan generar reflejen esas variaciones.

La información ha de estar a los máximos niveles de detalle. Los Almacenes de Datos departamentales o Bodegas de Datos son tratados como subconjuntos de este Almacén de Datos corporativo, que son construidos para cubrir las necesidades individuales de análisis de cada departamento, y siempre a partir de este Almacén de Datos Central.

Inmon menciona en su metodología que es mejor usar el método descendente a la hora de diseñar el Almacén de Datos, es decir, las Bodegas de Datos se crean después de haber terminado de construir el Almacén de Datos completo, considerando así mejor la obtención y manipulación de los datos corporativos.

Los datos son extraídos de los sistemas operacionales por los procesos ETC y cargados en las áreas de almacenamiento intermedias, donde son validados y consolidados en el Almacén de Datos corporativo, donde además existen los llamados metadatos que documentan de una forma clara y precisa el contenido del Almacén de Datos. Una vez realizado este proceso, los procesos de refresco de la Bodega de Datos departamentales obtienen la información de él, y con las consiguientes transformaciones, organizan los datos en las estructuras particulares requeridas por cada uno de ellos, refrescando su contenido.

Al tener este enfoque global, es más difícil de desarrollar en un proyecto sencillo (pues estamos intentando abordar el “todo”, a partir del cual luego iremos al “detalle”).

3.3 Metodología Tian Consultores

El Manual de Desarrollo de Sistemas de Información de Inteligencia de Negocios de Tian Consultores presenta una metodología basada en las mejores prácticas, aprovechando la velocidad y la orientación de método ascendente de Kimball y la integración del método descendente de Inmon.

La metodología describe cuatro fases principales:

- **Análisis del Negocio:** busca identificar la definición y el alcance que tiene el proyecto. Esta fase comienza la identificación del Modelo de Análisis del Negocio. Según Tian Consultores el Modelo de Análisis de Negocio determina como se van a definir los requerimientos de información del negocio. También definen los 3 Modelos de Análisis de Negocio más importantes:
 - **Listado de Requerimientos:** Se modela en base a requerimientos de información predefinidos y suministrados por el responsable del proceso. Se presenta una lista de requerimientos bien definidos por este responsable de proceso y se analizan para ubicar y extraer los datos necesarios para el cálculo de los indicadores asociados, de las fuentes de datos transaccionales de la organización.
 - **Análisis del Proceso:** Se modela en base al proceso que necesita desarrollar el Sistema de Información Analítico. Los diseñadores de los Almacenes de Datos deben tener en claro cuáles son los factores claves que guían el negocio para determinar efectivamente los requerimientos y traducirlos en consideraciones de diseño apropiadas. Para esto se debe analizar a fondo el proceso de negocio. Una vez determinados los requerimientos se procede a analizar las fuentes de datos transaccionales con la finalidad de detectar los indicadores que satisfagan dichas necesidades.
 - **Análisis del Modelo de Datos Transaccional:** Consiste en obtener el modelo dimensional que sirve de base para el almacén de datos directamente del modelo físico de la base de datos del sistema transaccional, antes de definir los requerimientos que pueden ser satisfechos. Se genera el modelo de tal forma que a partir de él se puedan obtener los indicadores necesarios para satisfacer cualquier requerimiento de negocio. En algunos casos pueden presentarse requerimientos cuyos indicadores no se puedan obtener del modelo generado, en ese caso habría que hacer modificaciones al modelo transaccional y por ende en el modelo dimensional.

Luego de seleccionar el Modelo de Análisis de Negocio que mejor se adapta a nuestro caso, se procede a identificar las fuentes de datos para pasar a la fase de Modelamiento Dimensional que consiste en identificar la jerarquía de los datos involucrados, identificar los hechos a medir e identificar las dimensiones o criterios de clasificación de estos hechos.

En la fase de detallar requerimientos según el modelo de análisis se procede a detallar los requerimientos del negocio en base al modelo dimensional realizado y luego se documentan. En el caso de que en la etapa de identificación de modelo de análisis de requerimientos

hayamos escogido el Listado de Requerimientos, en esta fase solo se documenta ya tengo el listado de los requerimientos.

- **Desarrollo del Almacén de Datos:** Esta fase se basa en las mejores prácticas de las metodologías de Inmon y Kimball. Consiste en la creación del Almacén de Datos de acuerdo con el modelo dimensional obtenido. En esta fase se realiza la carga de los datos involucrados desde las fuentes de datos al Almacén de Datos. La arquitectura propuesta por la solución presenta dos procesos de Extracción, Transformación y Carga importantes: el primero se encargará de la extracción, transformación y carga de los datos fuentes al área de intermedia, esta área intermedia se utilizará como repositorio para estandarizar, integrar, consolidar y extraer sólo la información necesaria para la construcción de análisis y reportes del negocio. El segundo proceso será el encargado de la extracción, transformación y carga de la base de datos intermedia al almacén de datos de modo tal que los datos estén preparados para el análisis. Esta fase finaliza con el análisis de la consistencia de datos.
- **Desarrollo de Requerimientos de Información:** Involucra configuraciones de los metadatos. Esta fase inicia con la configuración del Administrador, este Administrador es una herramienta que me permite hacer la conexión entre el almacén de datos y la herramienta de Inteligencia de Negocios. Luego se procede a la creación de las consultas que satisfacen los requerimientos de información, y son estas consultas las que van a poblar los cuadros de mando de la herramienta. Por último se realiza el análisis de la consistencia de datos que se muestran en los cuadros de mando, esto implica verificar que los datos que se muestran en el cuadro de mando tengan una lógica, y que pueda brindar la información necesaria para el análisis. También se verifica si los datos son correctos con respecto a las distintas fuentes de datos.
- **Implementación de la Solución:** representa el correcto funcionamiento de la tecnología, los datos y las aplicaciones de usuarios finales accesibles para el usuario del negocio, quien debe ser adiestrado para su utilización. Esta etapa implica el ajuste de las consultas, este ajuste tiene la finalidad de mejorar el desempeño de la solución implementada. En esta etapa también se configuran los esquemas de seguridad necesarios, definiendo los roles y las permisologías necesarias, se automatizan los procesos manuales en un flujo de proceso y por último se adiestra el usuario para que explote su solución de inteligencia de negocio.

La siguiente ilustración presenta gráficamente la Metodología de Desarrollo de Sistemas de Información de Inteligencia de Negocios analizada anteriormente:

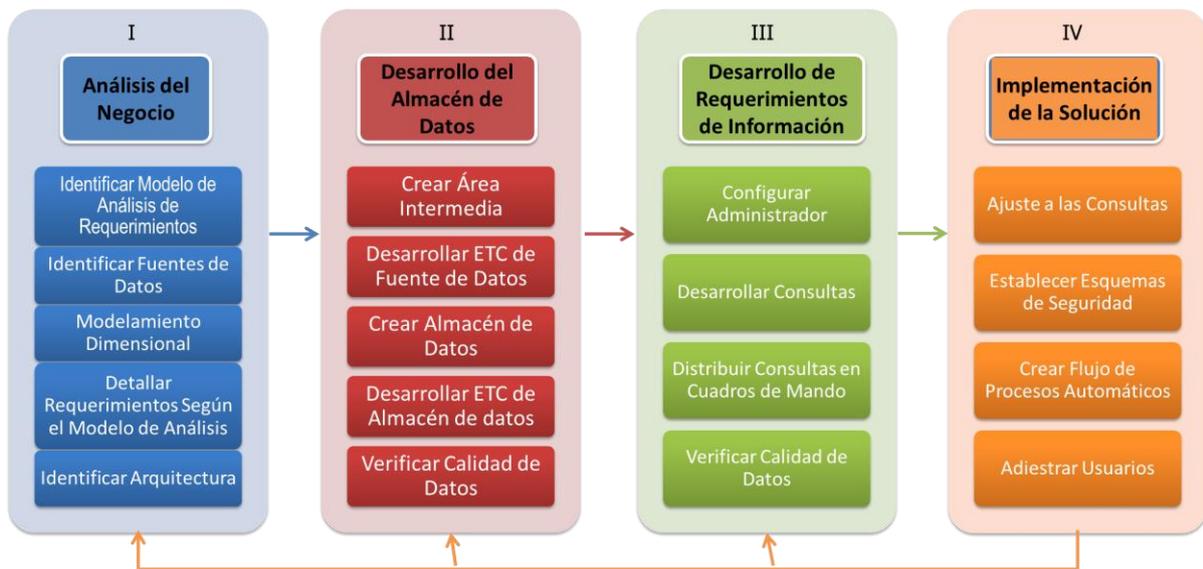


Ilustración 26: Modelo de Desarrollo de Sistemas de Información de Inteligencia de Negocios de Tian Consultores

Fuente: Manual de Desarrollo de Sistemas de Información de Inteligencia de Negocios de Tian Consultores

CAPÍTULO 4: MARCO APLICATIVO

En este capítulo se presenta la implementación del enfoque utilizado en este Trabajo Especial de Grado. Este enfoque es el utilizado por la empresa Tian Consultores, en el proceso de desarrollo de un Sistema de Información de Inteligencia de Negocios basado en experiencias propias, el método Ascendente de Kimball y el método Descendente de Inmon, gracias a esto podemos aprovechar la velocidad y la orientación de método ascendente y la integración del método descendente. De esta manera, se presenta el contexto de desarrollo y cada una de las fases establecidas en dicho método utilizado para el desarrollo de este Trabajo Especial de Grado.

4.1 Contexto del Desarrollo

Este proyecto ha sido desarrollado en el contexto del Trabajo Especial de Grado presentando ante la ilustre Universidad Central de Venezuela, Facultad de Ciencias, Escuela de Computación, bajo la tutoría del profesor Franky Uzcátegui, para optar al Título de Licenciado en Computación, por parte del bachiller Simón Tribiño como único integrante del equipo desarrollador.

4.2 Proyecto

El presente Trabajo Especial de Grado está basado en el desarrollo de una solución de Inteligencia de Negocios para el Sistema de Información Encuesta Simple, que permita a la gerencia media y alta, elaborar y obtener reportes y cuadros de mando que faciliten el análisis de los resultados de las encuestas simples aplicadas en sus organizaciones.

La solución se ha desarrollado utilizando la suite de Oracle, Oracle Business Intelligence Standard Edition One, la cual provee un conjunto de herramientas y componentes que facilitan las etapas de desarrollo, prueba e implementación.

Para la implementación de la solución de Inteligencia de Negocios se ha empleado el enfoque que utiliza la empresa Tian Consultores, el cual se basa en experiencias propias y en los métodos de desarrollo Ascendente (Kimball, R.) y Descendente (Inmon, B.), aprovechando las mejores prácticas de cada una.

4.3 Fases del Proyecto

A continuación se describe con detalle los procesos que se realizaron en cada una de las cuatro fases que se llevaron a cabo para la implementación de la solución:

4.3.1 Análisis del Negocio

Esta fase está constituida por la identificación y manejo del problema actual, en este caso se desarrolló una solución de Inteligencia de Negocios útil para todas aquellas organizaciones que deseen obtener

una herramienta que les permita facilitar el análisis de los resultados de sus encuestas simples a través de reportes y cuadros de mando.

Esta fase se subdivide en seis otras subfases, a continuación se describe en detalle los procesos realizados en cada una:

4.3.1.1 Identificación del Modelo de Análisis del Negocio

Esta fase comienza la identificación del Modelo de Análisis del Negocio. Según Tian Consultores el Modelo de Análisis de Negocio determina como se van a definir los requerimientos de información del negocio. Los tres Modelos de Análisis de Negocio más importantes son:

- **Listado de Requerimientos:** Se modela en base a requerimientos de información predefinidos y suministrados por el responsable del proceso. Se presenta una lista de requerimientos bien definidos por este responsable de proceso y se analizan para ubicar y extraer los datos necesarios para el cálculo de los indicadores asociados, de las fuentes de datos transaccionales de la organización.
- **Análisis del Proceso:** Se modela en base al proceso que necesita desarrollar el Sistema de Información Analítico. Los diseñadores de los Almacenes de Datos deben tener en claro cuáles son los factores claves que guían el negocio para determinar efectivamente los requerimientos y traducirlos en consideraciones de diseño apropiadas. Para esto se debe analizar a fondo el proceso de negocio. Una vez determinados los requerimientos se procede a analizar las fuentes de datos transaccionales con la finalidad de detectar los indicadores que satisfagan dichas necesidades.
- **Análisis del Modelo de Datos Transaccional:** Consiste en obtener el modelo dimensional que sirve de base para el almacén de datos directamente del modelo físico de la base de datos del sistema transaccional, antes de definir los requerimientos que pueden ser satisfechos. Se genera el modelo de tal forma que a partir de él se puedan obtener los indicadores necesarios para satisfacer cualquier requerimiento de negocio.

Para el presente Trabajo Especial de Grado se ha elegido este último, el Análisis del Modelo de Datos Transaccional.

4.3.1.2 Identificación de las Fuentes de Datos

Luego de seleccionar el Modelo de Análisis de Datos Transaccional, se procedió a identificar el área o ambiente transaccional desde donde se van a extraer los datos necesarios para representarse en los cuadros de mando.

De esta manera, el modelo relacional presentado en la Ilustración 27 representa un ambiente donde un usuario con roles específicos, que determinan sus acciones y opciones en el sistema, aplica un encuesta simple a un encuestado en un lugar y una fecha programada. A su vez el formulario está conformado por una serie de medidas o preguntas y dichas preguntas pueden tener ciertos valores definidos.

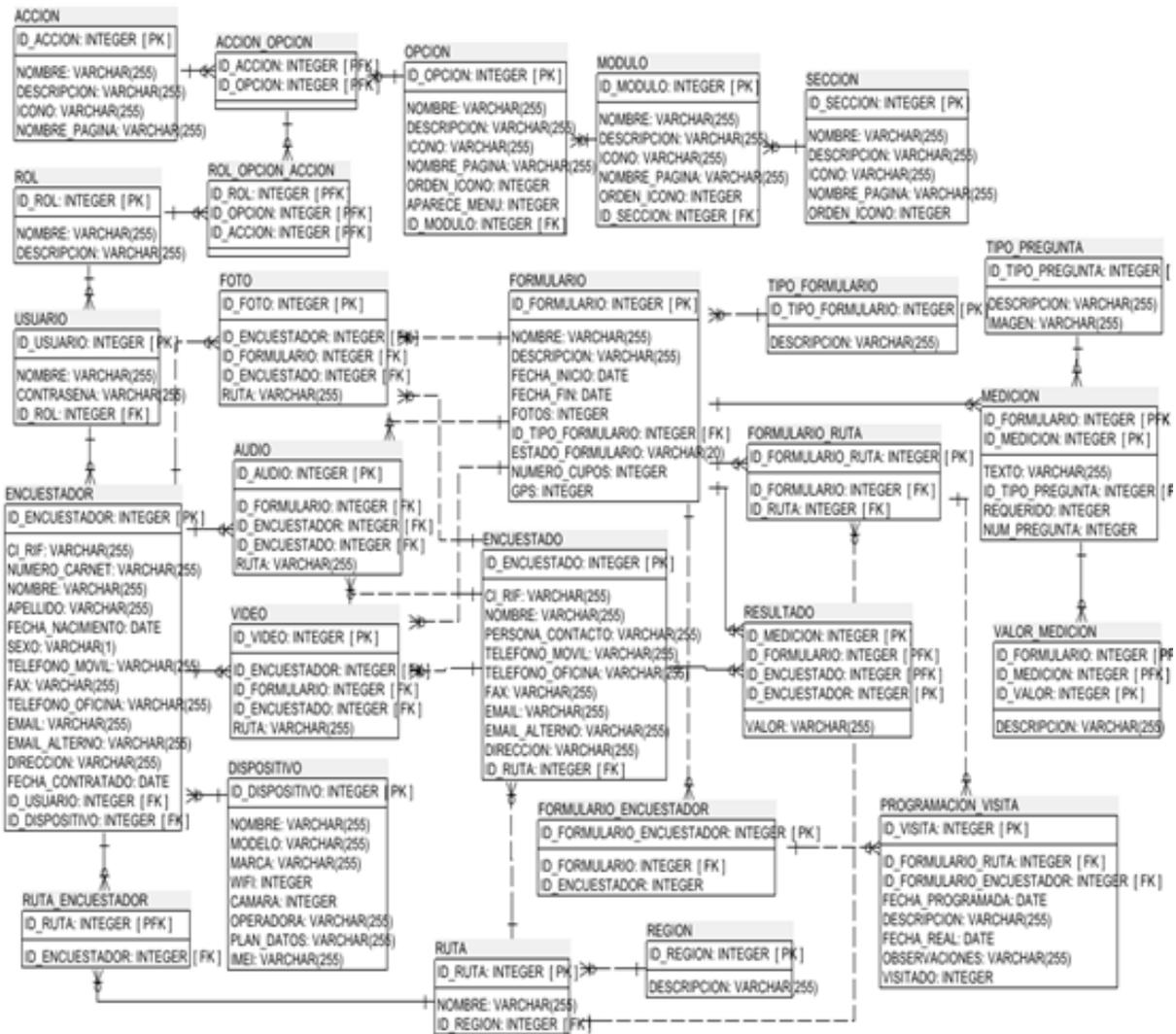


Ilustración 27. Modelo Relacional del Ambiente Transaccional (Encuesta Simple)

Fuente: Manual del Sistema Encuesta Simple de Tian Consultores

Luego de identificar las fuentes de datos se procede a la fase de Modelamiento Dimensional que consiste en identificar la jerarquía de los datos involucrados, identificar los hechos a medir e identificar las dimensiones o criterios de clasificación de estos hechos.

4.3.1.3 Modelamiento Dimensional

Antes de definir los requerimientos del negocio, de acuerdo con el alcance planteado, se determinan los datos necesarios para cumplir con cualquier criterio analítico que pueda presentarse en la organización. Para ello, se jerarquizan los distintos niveles de información que se identifican y a partir de esta jerarquización, se identifica el grado de granularidad requerida.

Esta jerarquía se basa en un encuestador que aplica una encuesta simple a un conjunto de encuestados, en una fecha y lugar planificado. Esta encuesta se aplica mediante un dispositivo móvil, donde se carga el formulario asociado a dicha encuesta, este formulario está conformado por un conjunto de medidas que a su vez tienen un conjunto de respuestas asociado. Estos formularios son categorizados en base a la manera en que serán aplicados (ver Ilustración 28).



Ilustración 28: Granularidad

Fuente: Elaboración Propia

Una vez obtenida la granularidad, a partir del nivel más bajo de la jerarquía, se definen las dimensiones de la solución, luego se especifican los diferentes grados de detalle (atributos) dentro de cada dimensión es decir la granularidad de cada dimensión, que se define como el nivel de detalle de cada indicador (variable o métrica) y por último, se definen las jerarquías que es el nivel de asociación entre los datos de la dimensión, todo esto es lo que le da forma al modelo dimensional del sistema de información Encuesta Simple.

El diseño del modelo dimensional se realizó basándose en el modelo de datos transaccional del Sistema de Información Encuesta Simple, este modelo fue analizado para determinar qué elementos

podrían considerarse como hechos medibles o dimensiones que pudieran formar parte de la solución de inteligencia de negocios.

Para poder saber que hechos y dimensiones eran necesarios para implementar la solución planteada, se desglosó una a una las tablas y los atributos del sistema transaccional, haciendo énfasis en determinar lo que se podía medir o contabilizar y bajo qué perspectivas se podían hacer.

El análisis indicó las diversas perspectivas con las cuales se desearía o se pudieran agrupar los datos. Es así como se identificaron las distintas dimensiones que se contemplan en la solución y que se mencionan a continuación:

- Encuestado
- Encuestador
- Dispositivo
- Tiempo
- Ruta
- Medición

Así como con las dimensiones se determinaron los hechos medibles, que para nuestra solución son los:

- Resultados.

Con los elementos expuestos anteriormente se pudo esbozar un primer modelo dimensional, el cual puede apreciarse en la Ilustración 29:



Ilustración 29: Modelo Dimensional

Fuente: Elaboración Propia

Una vez definidas las perspectivas a utilizar y los hechos a medir en la solución propuesta, se procedió a identificar por cada dimensión los atributos y las jerarquías que se consideraron para poder cumplir con los objetivos planteados.

La definición de los distintos niveles jerárquicos se puede ver en la Ilustración 30:

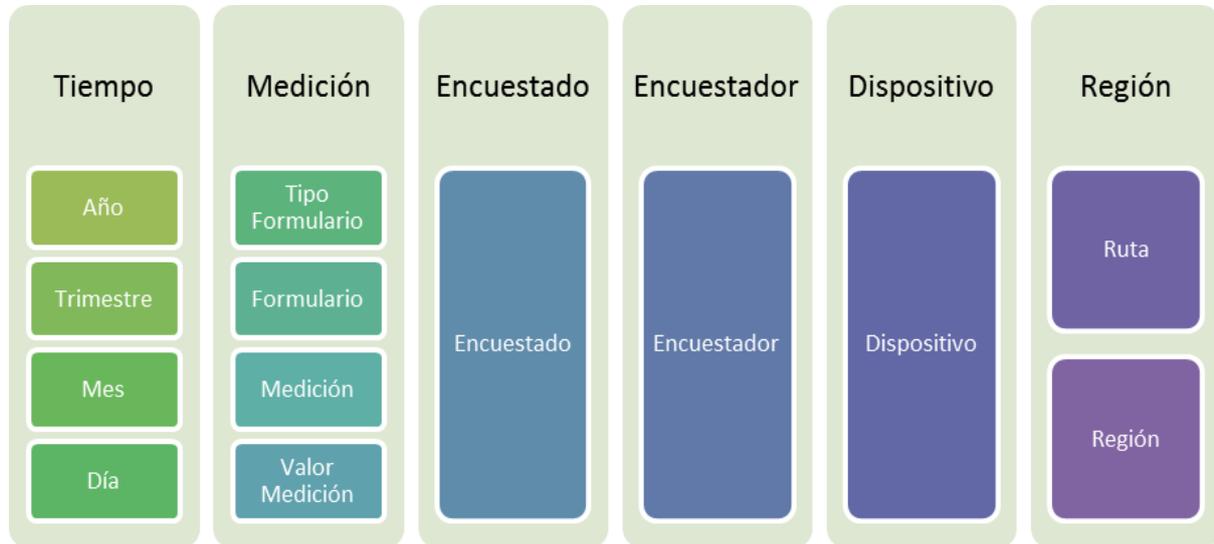


Ilustración 30: Definición de Dimensiones

Fuente: Elaboración Propia

La jerarquía en cada dimensión se define dependiendo de la lógica del negocio, en este caso, según lo definido en el Modelo de Datos Transaccional del Sistema de Información Encuesta Simple.

En el caso de la dimensión Medición, se definió que todas las mediciones o preguntas están agrupadas en formularios, que a su vez se clasifican según la forma en que dicho formulario será aplicado por los distintos encuestadores. A su vez estas mediciones tienen dos o más posibles respuestas definidas. Es así como surgieron los cuatro niveles jerárquicos contemplados para la dimensión medición (Ver Ilustración 30).

Para explicar esta jerarquización a continuación se presenta un ejemplo donde un formulario tiene dos preguntas y cada pregunta a su vez tiene dos posibles respuestas:

- Formulario 1 → Pregunta 1 → Respuesta 1
- Formulario 1 → Pregunta 1 → Respuesta 2
- Formulario 1 → Pregunta 2 → Respuesta 1
- Formulario 1 → Pregunta 2 → Respuesta 2

Para la dimensión tiempo los niveles jerárquicos surgen naturalmente, la única variante está en la granularidad o el nivel de detalle deseado para la dimensión. En este caso, el nivel contemplado fue el

que se definió en la Ilustración 30, sin embargo, dependiendo de las necesidades de la organización o de las consideraciones tomadas para la solución, se pudo haber hecho una jerarquización más detallada (contemplando semestres, quincenas y/o semanas, entre otros) o más simple (contemplando sólo año y mes).

Por último la dimensión Ruta los niveles jerárquicos también surgen naturalmente, dado que las rutas se establecen dentro de regiones. Un ejemplo donde se aplique esta jerarquización se muestra a continuación:

Región 1: Centro → Ruta 1: Av. Fuerzas Armadas.

Región 1: Centro → Ruta 2: Av. Libertador.

Región 2: Oeste → Ruta 1: Av. Lecuna.

Una vez definidos los niveles jerárquicos para cada dimensión contemplada para la solución de inteligencia de negocios propuesta, se procedió a definir los atributos necesarios para cumplir con los requerimientos planteados.

A continuación se describirán todas las dimensiones contempladas para la solución de inteligencia de negocios para el Sistema de Información Encuesta Simple:

Dimensión: Encuestado		
Descripción:	Contiene los distintas personas encargadas de aplicar los Formulario a los Encuestados en un lugar y una fecha determinada.	
Atributo	Tipo de Dato	Descripción
DIMENSION_KEY	NUMBER	Identificador único y técnico del registro en la tabla dimensión Encuestado.
EMAIL_ALTERNO	VARCHAR2(255 BYTE)	Correo electrónico alternativo del Encuestado
CI_RIF	VARCHAR2(255 BYTE)	Cédula de identidad del Encuestado
TELEFONO_MOVIL	VARCHAR2(255 BYTE)	Teléfono Móvil del Encuestado
FAX	VARCHAR2(255 BYTE)	Fax del Encuestado
ENCUESTADO_NOMBRE	VARCHAR2(255 BYTE)	Nombre del Encuestado
DIRECCION	VARCHAR2(255 BYTE)	Dirección del Encuestado
TELEFONO_OFICINA	VARCHAR2(255 BYTE)	Teléfono de la Oficina del Encuestado
PERSONA_CONTACTO	VARCHAR2(255 BYTE)	Persona contacto del Encuestado.
ENCUESTADO_CODIGO	VARCHAR2(255 BYTE)	Código del Encuestado.
EMAIL	VARCHAR2(255 BYTE)	Correo electrónico del Encuestado.
ENCUESTADO_ID	NUMBER	Identificador único y técnico del registro en la tabla Encuestado.

Tabla 9: Dimensión Encuestado

Fuente: Elaboración Propia

Dimensión: Ruta		
Descripción:	Contiene las regiones y las rutas donde los encuestadores van a aplicar los distintos formularios.	
Atributo	Tipo de Dato	Descripción
DIMENSION_KEY	NUMBER	Identificador único y técnico del registro en la tabla dimensión Ruta.
REGION_ID	NUMBER	Identificador único y técnico del registro en la tabla Región
REGION_DESCRIPCION	VARCHAR2(255 BYTE)	Descripción de la Región
REGION_CODIGO	VARCHAR2(255 BYTE)	Código de la Región.
RUTA_CODIGO	VARCHAR2(255 BYTE)	Código de la Ruta.
RUTA_DESCRIPCION	VARCHAR2(255 BYTE)	Descripción de la Ruta.
RUTA_ID	NUMBER	Identificador único y técnico del registro en la tabla Ruta.

Tabla 10: Dimensión Ruta

Fuente: Elaboración Propia

Dimensión: Dispositivo		
Descripción:	Contiene los distintos dispositivos utilizados por los encuestadores para aplicar las encuestas.	
Atributo	Tipo de Dato	Descripción
DIMENSION_KEY	NUMBER	Identificador único y técnico del registro en la tabla dimensión Dispositivo.
CAMARA	NUMBER	Atributo que determina si el dispositivo tiene cámara o no.
DISPOSITIVO_NOMBRE	VARCHAR2(255 BYTE)	Nombre del Dispositivo
DISPOSITIVO_ID	NUMBER	Identificador único y técnico del registro en la tabla
MARCA	VARCHAR2(255 BYTE)	Marca del Dispositivo.
PLAN_DATOS	VARCHAR2(255 BYTE)	Plan de datos que posee el Dispositivo.
WIFI	NUMBER	Atributo que determina si el dispositivo tiene wifi o no.
DISPOSITIVO_CODIGO	VARCHAR2(255 BYTE)	Identificador único y técnico del registro en la tabla Dispositivo.
MODELO	VARCHAR2(255 BYTE)	Modelo del Dispositivo.
OPERADORA	VARCHAR2(255 BYTE)	Operadora del Dispositivo.
IMEI	VARCHAR2(255 BYTE)	Código único que tiene el dispositivo físico.

Tabla 11: Dimensión Dispositivo

Fuente: Elaboración Propia

Dimensión: Encuestador		
Descripción:	Contiene los empleados encargados de aplicar los distintos formularios.	
Atributo	Tipo de Dato	Descripción
DIMENSION_KEY	NUMBER	Identificador único y técnico del registro en la tabla dimensión Encuestador
EMAIL	VARCHAR2(255 BYTE)	Correo electrónico del Encuestador
CI_RIF	VARCHAR2(255 BYTE)	Cédula de identidad del Encuestador
USUARIO_NOMBRE	VARCHAR2(255 BYTE)	Nombre del Usuario asociado al Encuestador
NOMBRE_USUARIO	VARCHAR2(255 BYTE)	Nombre del Encuestador
EMAIL_ALTERNATIVO	VARCHAR2(255 BYTE)	Correo electrónico alternativo del Encuestador
SEXO	VARCHAR2(1 BYTE)	Género del Encuestador
FECHA_NACIMIENTO	DATE	Fecha de Nacimiento del Encuestador
TELEFONO_OFICINA	VARCHAR2(255 BYTE)	Número de la oficina del Encuestador
APELLIDO	VARCHAR2(255 BYTE)	Apellido del Encuestador
USUARIO_CODIGO	VARCHAR2(255 BYTE)	Contraseña del Usuario asociado al Encuestador
TELEFONO_MOVIL	VARCHAR2(255 BYTE)	Teléfono Móvil del Encuestador
NUMERO_CARNET	VARCHAR2(255 BYTE)	Número de Carnet del Encuestador
DIRECCION	VARCHAR2(255 BYTE)	Dirección del Encuestador
FECHA_CONTRATADO	DATE	Fecha de contratado del Encuestador
FAX	VARCHAR2(255 BYTE)	Fax del Encuestador
USUARIO_ID	NUMBER	Identificador único y técnico del registro del Usuario asociado al Encuestador

Tabla 12: Dimensión Encuestador

Fuente: Elaboración Propia

Dimensión: Medición		
Descripción:	Contiene los distintos formularios, categorizados por tipos, con sus respectivas mediciones y los posibles valores de estas mediciones.	
Atributo	Tipo de Dato	Descripción
DIMENSION_KEY	NUMBER	Identificador único y técnico del registro en la tabla dimensión
TIPO_FORMULARIO_DESCRIPCION	VARCHAR2(255 BYTE)	Descripción del tipo de Formulario
TIPO_FORMULARIO_ID	NUMBER	Identificador único y técnico del registro en la tabla
TIPO_FORMULARIO_CODIGO	VARCHAR2(255 BYTE)	Código del tipo de Formulario
FECHA_FIN	DATE	Fecha final tope para aplicar el formulario
FORMULARIO_ID	NUMBER	Identificador único y técnico del registro en la tabla Formulario
ESTADO_FORMULARIO	VARCHAR2(255 BYTE)	Estado del formulario. (Aplicado / Sin aplicar)
NOMBRE_FORMULARIO	VARCHAR2(255 BYTE)	Nombre del Formulario
FECHA_INICIO	DATE	Fecha desde la que se puede empezar a aplicar le formulario
FORMULARIO_DESCRIPCION	VARCHAR2(255 BYTE)	Descripción del Formulario
FORMULARIO_CODIGO	VARCHAR2(255 BYTE)	Código del Formulario
NUMERO_CUPOS	NUMBER	Cantidad de personas a las cuales se les puede aplicar el formulario
GPS	NUMBER	Coordenadas GPS del lugar donde se aplico el formulario.
FOTOS	NUMBER	Atributo que indica si se tomaron o no fotos en la aplicación del formulario
MEDICION_DESCRIPCION	VARCHAR2(255 BYTE)	Descripción de la medición o pregunta.
REQUERIDO	NUMBER	Atributo que determina si la pregunta hay q contestarla obligatoriamente.
MEDICION_ID	NUMBER	Identificador único y técnico del registro en la tabla Medición
MEDICION_CODIGO	VARCHAR2(255 BYTE)	Código de medición.
NUMERO_PREGUNTA	NUMBER	Número de la pregunta dentro del formulario
VALOR_MEDICION_DESCRIPCION	VARCHAR2(255 BYTE)	Descripción de las posibles respuesta de una medición.
VALOR_MEDICION_CODIGO	VARCHAR2(255 BYTE)	Código de las posibles respuestas de una medición.
VALOR_MEDICION_ID	NUMBER	Identificador único y técnico del registro en la tabla

Tabla 13: Dimensión Medición

Fuente: Elaboración Propia

Dimensión: Tiempo		
Descripción:	Contiene las fechas en que se realizaron las distintas encuestas y sus distintas representaciones.	
Atributo	Tipo de Dato	Descripción
año	NUMBER	Número del año al que pertenece a la fecha.
trimestre	NUMBER	Número del trimestre al que pertenece a la fecha.
desc_trimestre	VARCHAR2(255 BYTE)	Descripción del trimestre al que pertenece la fecha.
abrev_trimestre	VARCHAR2(255 BYTE)	Abreviación de la descripción del trimestre.
mes	NUMBER	Número del mes al que pertenece la fecha.
desc_mes	VARCHAR2(255 BYTE)	Nombre del mes al que pertenece la fecha.
abrev_mes	VARCHAR2(255 BYTE)	Abreviación de la descripción del mes.
dia	NUMBER	Número del día al que pertenece la fecha.
desc_dia	VARCHAR2(255 BYTE)	Nombre del día al que pertenece la fecha.
abrev_dia	VARCHAR2(255 BYTE)	Abreviación de la descripción del día.

Tabla 14: Dimensión Tiempo

Fuente: Elaboración Propia

En resumen, se puede afirmar que en esta fase es donde se diseña cómo se va a agrupar la información transaccional, de tal forma que pueda ser representada y almacenada dentro del ambiente de Inteligencia de Negocios. Así mismo, se detallan cada una de las dimensiones y los hechos a medir, modelando el modelo transaccional de datos del Sistema de Información Encuesta Simple.

4.3.1.4 Detallar requerimientos según el modelo de análisis

Como se ha mencionado, la solución de Inteligencia de Negocios pretende ser útil a todas aquellas organizaciones que deseen obtener una herramienta que les permita facilitar el análisis de los resultados de sus encuestas simples través de reportes y cuadros de mando.

Por esta razón, la empresa Tian Consultores ha realizado una serie de investigaciones en varias organizaciones, con el fin de conocer cuáles son los problemas que tienen en común a la hora de realizar análisis de los resultados de sus encuestas simples. De esta manera, se contemplaron una serie de indicadores de gestión y métricas que permitirán a la organización poder mejorar su gestión, disminuir costos y evaluar resultados.

Éstos indicadores se enumeran a continuación:

1. Cantidad de encuestados en el tiempo.
2. Cantidad de encuestados por formulario.
3. Porcentaje de distribución de la cantidad de respuestas por pregunta.
4. Cantidad de encuestas aplicadas por encuestador.
5. Cantidad de encuestas aplicadas por encuestador y por región.
6. Cantidad de encuestas aplicadas por encuestador y por año.
7. Porcentaje de distribución de las encuestas aplicadas por encuestador y por año
8. Porcentaje de distribución de los encuestados por región y ruta.
9. Porcentaje de distribución de los encuestados por encuestador.
10. Porcentaje de distribución de los encuestados por formulario.
11. Cantidad de encuestados por ruta y región.
12. Porcentaje de distribución de la Cantidad de encuestados por ruta y región.
13. Cantidad de encuestados por ruta, región y año.
14. Porcentaje de distribución de la Cantidad de encuestados por ruta, región y año.
15. Cantidad de encuestadores por ruta y región.
16. Porcentaje de distribución de la Cantidad de encuestadores por ruta y región.
17. Porcentaje de distribución de los dispositivos por encuestador.
18. Porcentaje de distribución de las marcas de dispositivos más utilizadas.

Con la especificación de los indicadores a implementar en la solución planteada para cumplir con los requerimientos establecidos, se procedió a realizar el desarrollo de las aplicaciones analíticas. Dicho proceso se describirá más adelante.

4.3.1.5 Identificación de la Arquitectura

La arquitectura de la solución está conformada por tres componentes principales, que son la Fuente de Datos, el Almacén de Datos y los Análisis de Datos como se puede observar en la Ilustración 1:

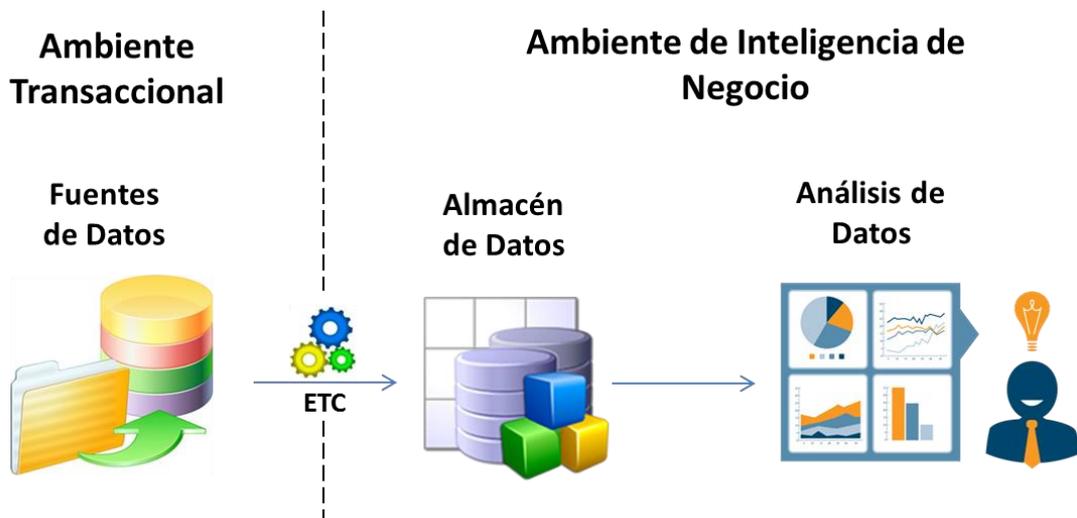


Ilustración 1: Arquitectura Propuesta para la Solución de Inteligencia de Negocio

Fuente: Elaboración Propia

El primer componente involucra a las diversas fuentes de datos que pudieran considerarse para la solución de inteligencia de negocios. Estos datos pueden provenir de fuentes internas y/o externas, usualmente se emplean aquellos almacenados en los sistemas transaccionales. En nuestro caso las fuentes de datos provienen de una sola Base de Datos, creada mediante el Sistema Manejador de Base de Datos de Oracle. Esta base de datos es la base de datos transaccional del Sistema de Información Encuesta Simple. Debido a que el sistema maneja sus procesos de forma paralela a la solución implementada y que los datos no provienen de más de una fuente de datos, se decidió no implementar un Área de Integración.

Luego se diseña y desarrolla un Almacén de Datos, basado en el modelo dimensional que permite obtener indicadores y reportes necesarios para el análisis de los resultados de Encuesta Simple. Este Almacén de Datos soporta los datos provenientes de la fuente de datos.

La arquitectura contiene un proceso ETC que está encargado de llevar los datos desde la Fuente de Datos al Almacén de Datos, este proceso es particular para cada organización debido a sus distintas fuentes de datos.

Una vez cargados los datos en el almacén de datos, están disponibles para el uso de herramientas de explotación y publicación que construyen las consultas y cuadros de mando que permiten visualizar los indicadores.

4.3.2 Desarrollo del Almacén de Datos

En el desarrollo de esta solución de Inteligencia de Negocios se utilizará la herramienta de Oracle Warehouse Builder para desarrollar el Almacén de Datos. Esta herramienta facilita la creación de las

jerarquías y la granularidad de cada dimensión, como también realiza el proceso de extracción, transformación y carga.

A continuación se muestra el desarrollo del Almacén de Datos con su respectivo proceso de extracción, transformación y carga.

4.3.2.1 Creación del Almacén de Datos

El almacén de datos se desplegó físicamente usando la herramienta Oracle Warehouse Builder, producto de la suite de Oracle Business Intelligence Standard Edition One, y basándose en el modelo dimensional realizado y descrito con anterioridad.

En una primera instancia, usando un asistente del Centro de Diseño de la herramienta seleccionada, se definieron cada una de las dimensiones contempladas en el modelo dimensional, estableciendo inicialmente el nombre de la dimensión, luego sus atributos y por último sus niveles jerárquicos.

A continuación se explicara el desarrollo de una dimensión ejemplo para ilustrar su uso.

Se creó un proyecto denominado TESIS, entro de dicho proyecto se crearon dos módulos, el primero llamado Origen, que es donde están los metadatos de la base de datos fuente y el segundo llamado Destino, que es donde desarrollamos nuestro Almacén de Datos. Ver Ilustración 31: Centro de Diseño, Oracle Warehouse Builder

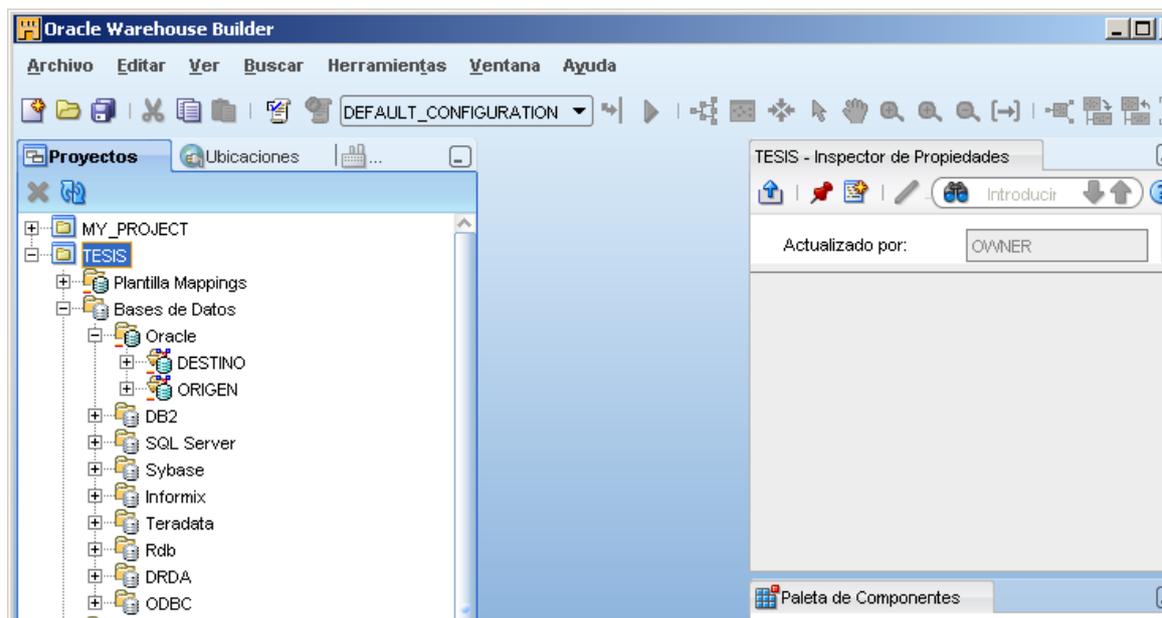


Ilustración 31: Centro de Diseño, Oracle Warehouse Builder

Fuente: Elaboración Propia

Dentro de nuestro módulo Destino nos ubicamos en la carpeta DIMENSIONES, presionamos el botón derecho del ratón, seleccionamos NUEVO ->NUEVA DIMENSIÓN. Ver Ilustración 32.

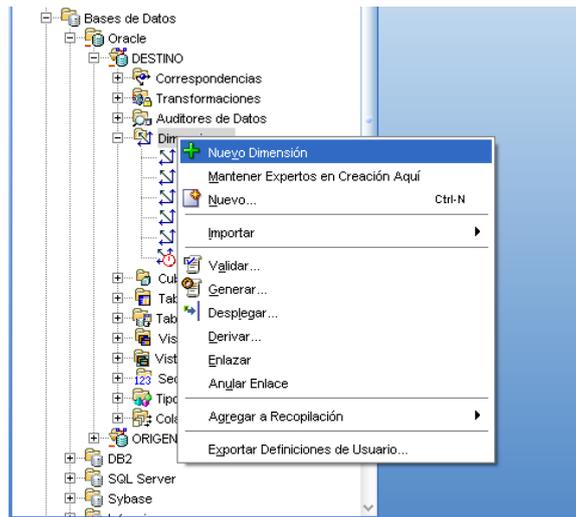


Ilustración 32: Asistente para la Creación de Dimensiones

Fuente: Elaboración Propia

Una vez seleccionada esa opción se mostrará una pantalla de bienvenida al asistente de la creación de dimensiones, en el cual nos explica los pasos que realizaremos. Luego se define el nombre de la dimensión, una descripción y el tipo de almacenamiento (Ver Ilustración 33). Para nuestra solución el tipo de almacenamiento es ROLAP.



Ilustración 33: Asistente para la Creación de Dimensiones: Paso 2

Fuente: Elaboración Propia

En el siguiente paso se definen los atributos que conforman la dimensión, su nombre, descripción, tipo de dato, longitud, precisión. Luego se definen los Niveles de jerarquía que conforman la dimensión y

se asocian los atributos a los distintos niveles de jerarquía al que pertenezcan. Por último se define la forma de almacenamiento de los cambios históricos y antes de terminar se muestra un resumen de todos los valores antes insertados, si es necesario hacer algún cambio se puede presionar el botón “ATRÁS”. (Ver Ilustración 34)



Ilustración 34: Resumen de la Creación de una Dimensión

Fuente: Elaboración Propia

De la misma forma fueron desarrolladas las dimensiones: Dispositivo, Encuestador, Encuestado, Ruta y Medición que pertenecen al modelo dimensional diseñado.

Para crear la dimensión de Tiempo se utilizó un asistente de la herramienta que permite generar específicamente este tipo de dimensiones. En este caso, se consideró (año, trimestre, mes y día), de esta forma, automáticamente la herramienta generó los niveles jerárquicos necesarios para soportar la selección realizada.

Una vez definidas todas las dimensiones se realizó el modelado del cubo, el cual incluye el hecho medible contemplado en la tabla de hecho. Luego de definir un nombre para el cubo, se seleccionaron las dimensiones involucradas en la solución y se especificaron los hechos medibles contemplados con anterioridad.

Si bien ya están todos los elementos del almacén de datos definidos en la herramienta, aún no se han creado ni desplegado físicamente, es decir, hasta ahora sólo se tiene el diseño lógico (ver Ilustración 35).



Ilustración 35: Diseño Lógico de las Dimensiones y el Cubo con Oracle Warehouse Builder

Fuente: Elaboración Propia

Una vez completado dicho diseño se realizó el despliegue del mismo usando el Administrador del Centro de Control de la herramienta seleccionada, de esta forma se crearon todas tablas y elementos que componen el almacén de datos propuesto para la solución de inteligencia de negocios.

Cabe destacar que al crear las tablas para el almacén de datos, se agregaron automáticamente una serie de campos o columnas necesarias para el correcto funcionamiento de las distintas aplicaciones que componen a la suite de Oracle Business Intelligence Standard Edition One.

Estos campos adicionales no interfieren de ninguna manera con el modelo propuesto para la solución de inteligencia de negocios, tanto los requerimientos como los objetivos establecidos pueden cumplirse aún con las modificaciones sufridas por las dimensiones.

Las tablas resultantes se almacenaron en una base de datos Oracle 11G optimizada para su funcionamiento como un almacén de datos.

4.3.2.2 Desarrollo del proceso de ETC del Almacén de Datos

El proceso de ETC para el almacén de datos está compuesto por siete procesos ETC en realidad, uno por cada dimensión definida y otro para el hecho. Estos procesos se realizaron con la ayuda de la herramienta Oracle Warehouse Builder la cual presenta una interfaz gráfica para ayudar a realizar los procesos de ETC. Ver Ilustración 36.

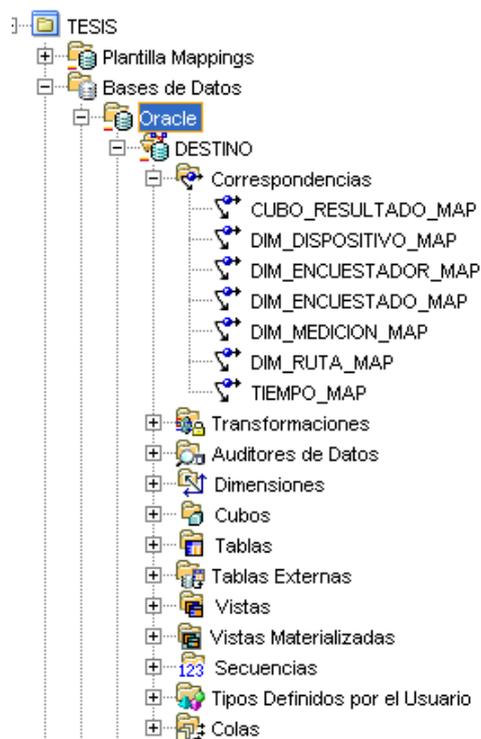


Ilustración 36: Centro de Diseño OWB

Fuente: Elaboración Propia

De esta forma se procedió a diseñar los procesos de ETC utilizando un editor del Oracle Warehouse Builder que permite, haciendo uso del *drag-and-drop* (en español Agarrar y Soltar), disponer de los elementos requeridos para realizar el proceso deseado.

Los elementos que pueden emplearse en el editor para realizar las tareas definidas van desde una paleta con diversas funcionalidades para transformar los datos (ver Ilustración 37), como evidentemente las tablas que han de usarse para extraer o cargar registros.



Ilustración 37: Paleta de Funcionalidades de OWB

Fuente: Elaboración Propia

En una primera instancia se realizaron los procesos de ETC para las dimensiones y de último se realizó lo propio para la tabla de hechos. Para ejemplificar lo realizado con las dimensiones, se tomará como ejemplo la dimensión Ruta. Usando el editor del Oracle Warehouse Builder, se dispusieron las tablas correspondientes de la base de datos origen y la dimensión seleccionada como destino (ver Ilustración 38).

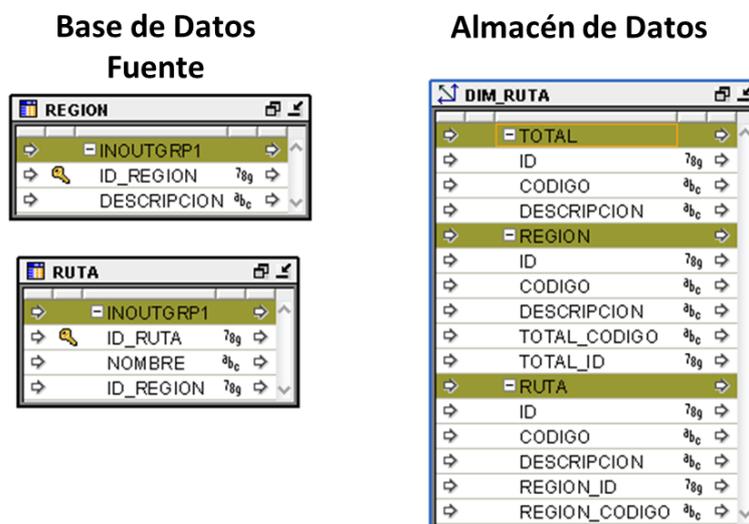


Ilustración 38: Ejemplo Disposición de Tablas para Proceso ETC

Fuente: Elaboración Propia

En la dimensión pueden apreciarse los distintos niveles jerárquicos que posee, yendo de arriba hacia abajo de lo más general a lo más específico. En todas las dimensiones se agregó una jerarquía denominada Total para que de esta forma se pueda ampliar la gama de posibilidades a la hora de realizar los reportes o indicadores, dado a que se pueden realizar mediciones tomando en cuenta toda una dimensión.

Aparte de las agrupaciones que pueden apreciarse producto de las jerarquías, en cada grupo distinto del nivel Total se muestran dos campos extras que referencian a su jerarquía padre.

Dado a que esta jerarquía total no se contempla en los datos que se manejaron, debió llenarse usando un componente de la herramienta denominado *Constant* (en español Constante) que permite colocar valores arbitrariamente como constantes. Dicha asociación se muestra en la Ilustración 39.

Por último, para completar el nivel de granularidad más bajo definido para la dimensión seleccionada, se realizó una combinación de los datos de las clasificaciones específicas con los productos. Esto se realizó utilizando el componente *Joiner* (en español Unión), de esta forma se definieron los campos involucrados en la combinación y se definió bajo qué criterio se iban a combinar los mismos.

En este caso, la combinación se realizó igualando el identificador de la clasificación a la que pertenecía la ruta y el identificador de las regiones existentes. El proceso resultante puede observarse en la Ilustración 39.

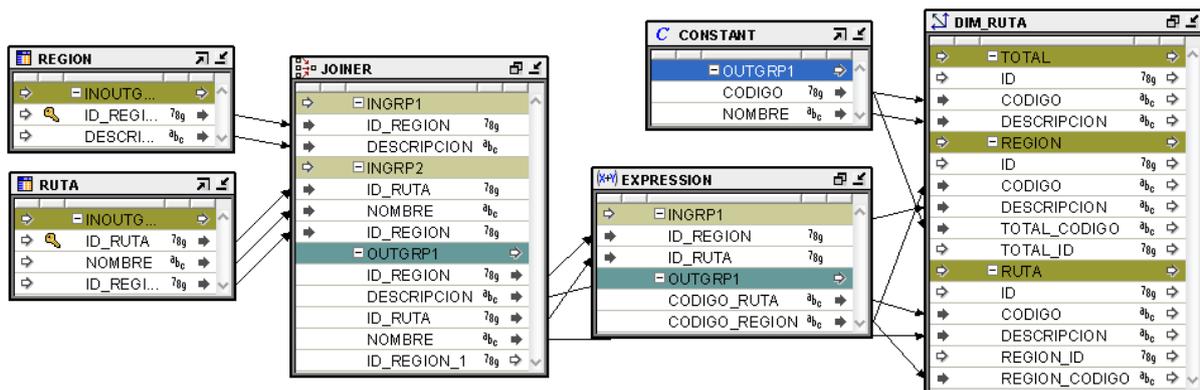


Ilustración 39: Proceso ETC Final para Dimensión Ruta

Fuente: Elaboración Propia

De manera similar se realizaron los demás procesos de extracción, transformación y carga para el resto de las dimensiones y la tabla de hechos contemplada en la solución. Dependiendo de la complejidad y de las necesidades en el diseño de los procesos, se utilizaron diversos componentes que ofrece la herramienta seleccionada.

Una vez diseñados todos los procesos necesarios para cargar los datos de la base de datos intermedia al almacén de datos, se procedió a crear los mismos físicamente dado que hasta ahora sólo se realizó el diseño lógico.

Al crear los procesos físicamente, Oracle Warehouse Builder crea por cada ETC un paquete, éstos son un conjunto de procedimientos programados en PL-SQL que se encuentran almacenados en la base de datos.

Luego de crear los procesos, se ejecutaron los mismos desde la herramienta utilizada para cargar de esta forma los datos correspondientes al almacén de datos.

4.3.2.3 Verificación de Calidad de Datos

Una vez terminado de construir el Almacén de Datos, se realiza el proceso de verificación de los datos, con la finalidad de asegurar su consistencia e integridad, así como también si son correctos con respecto a los datos transaccionales.

Esto se lleva a cabo realizando consultas a los sistemas transaccionales y al Almacén de Datos. Luego, se verifica que los datos obtenidos de las consultas de los sistemas transaccionales se correspondan con las consultas realizadas a la bodega de datos. De esta manera, se garantiza que la carga de datos se ha realizado de manera exitosa.

Ahora bien, se realizará la verificación a cada una de las dimensiones:

- **Dimensión Encuestado:** Para realizar la comparación entre las tablas, se seleccionaron el código y el nombre de cada Encuestado y se compararon los resultados tanto de la Dimensión del Almacén de Datos como la tabla de la Base de Datos Fuente. Ver Ilustración 40.

```

Select id_encuestado, nombre
from encuestado
ORDER by id_encuestado

```

ID_ENCUESTADO	NOMBRE
1	Abel Chen
2	Myles Benjamin
3	Jessica Garcia
4	Hyacinth Webster
5	Zeus Bean
6	Carson Wilkins
7	Quentin Craig
8	Allistair Pitts
9	Brendan Ellison
10	Adrienne Molina
11	Kelsey Bailey
12	Oleg Harper
13	Carter Roberson
14	Carol Strickland
15	Aimee Pennington
16	Zoe Middleton
17	Quail Zamora
18	Wade Mendoza
19	Quyn Rogers
20	George Henderson

```

Select encuestado_codigo, encuestado_nombre
from dim_encuestado_tab
where encuestado_id is not null
order by to_number(encuestado_codigo) ASC

```

ENCUESTADO_CODIGO	ENCUESTADO_NOMBRE
1 1	Abel Chen
2 2	Myles Benjamin
3 3	Jessica Garcia
4 4	Hyacinth Webster
5 5	Zeus Bean
6 6	Carson Wilkins
7 7	Quentin Craig
8 8	Allistair Pitts
9 9	Brendan Ellison
10 10	Adrienne Molina
11 11	Kelsey Bailey
12 12	Oleg Harper
13 13	Carter Roberson
14 14	Carol Strickland
15 15	Aimee Pennington
16 16	Zoe Middleton
17 17	Quail Zamora
18 18	Wade Mendoza
19 19	Quyn Rogers
20 20	George Henderson

Ilustración 40: Consulta Tabla Encuestado, Dimensión Encuestado

Fuente: Elaboración Propia

Adicionalmente, se realiza un conteo de la cantidad de registro que tiene cada tabla, de esta manera, si el primer par de consultas y el segundo par de consultas retornan la misma cantidad de registros con uno de diferencia, entonces se asegura que los datos cargados en la tabla de la dimensión Encuestado son consistentes, íntegros y correctos con respecto a los datos del sistema transaccional. Ver Ilustración 41.

```

Select COUNT(encuestado_codigo)
from dim_encuestado_tab

```

COUNT(ENCUESTADO_CODIGO)
833

```

Select COUNT(id_encuestado)
from encuestado

```

COUNT(ID_ENCUESTADO)
833

Ilustración 41: Consulta Conteo de Tabla Encuestado, Dimensión Encuestado

Fuente: Elaboración Propia

- Dimensión Encuestador:** Para realizar la comparación entre las tablas, se seleccionaron el código y el nombre de cada Encuestador. Adicionalmente, se realiza un conteo de la cantidad de registro que tiene cada tabla, de esta manera, si la primera consulta y la segunda consulta retornan la misma cantidad de registros y los datos coinciden, entonces se asegura que los datos cargados en la tabla de la dimensión Encuestador son consistentes, íntegros y correctos con respecto a los datos del sistema transaccional. Ver Ilustración 42

	USUARIO_CODIGO	USUARIO_NOMBRE
1	1	Simon
2	2	David
3	3	Leyton
4	4	Pedro
5	5	Alan
6	6	Jose
7	7	Carson
8	8	Giacomo
9	9	Elmo
10	10	Ricardo
11	11	Yvonne
12	12	Octavius
13	13	Sandra
14	14	Gisela
15	15	Drake
16	16	Sofia
17	17	Shafira
18	18	Pearl
19	19	Roberto
20	20	Raul

	ID_ENCUESTADOR	NOMBRE
1	1	Simon
2	2	David
3	3	Leyton
4	4	Pedro
5	5	Alan
6	6	Jose
7	7	Carson
8	8	Giacomo
9	9	Elmo
10	10	Ricardo
11	11	Yvonne
12	12	Octavius
13	13	Sandra
14	14	Gisela
15	15	Drake
16	16	Sofia
17	17	Shafira
18	18	Pearl
19	19	Roberto
20	20	Raul

Ilustración 42: Consulta Tabla Encuestador, Dimensión Encuestador

Fuente: Elaboración Propia

- Dimensión Dispositivo:** Para realizar la comparación entre las tablas, se seleccionaron el código y el nombre de cada Dispositivo. Adicionalmente, se realiza un conteo de la cantidad de registro que tiene cada tabla, de esta manera, si la primera consulta y la segunda consulta retornan la misma cantidad de registros y los datos coinciden, entonces se asegura que los datos cargados en la tabla de la dimensión Dispositivo son consistentes, íntegros y correctos con respecto a los datos del sistema transaccional. Ver Ilustración 43

Query 1 (Left):

```
Select dispositivo_codigo, dispositivo_nombre
from dim_dispositivo_tab
where dispositivo_codigo is not null
order by to_number(dispositivo_codigo) ASC
```

DISPOSITIVO_CODIGO	DISPOSITIVO_NOMBRE
1	Sony Xperia
2	Sony Xperia
3	Sony Xperia
4	Sony Xperia
5	Sony Xperia
6	Sony Xperia
7	Sony Xperia
8	Sony Xperia
9	Sony Xperia
10	Sony Xperia
11	Sony Xperia
12	Google Nexus One
13	Google Nexus One
14	Google Nexus One
15	Google Nexus One
16	Google Nexus One
17	Google Nexus One
18	Google Nexus One
19	Google Nexus One
20	Google Nexus One

Query 2 (Right):

```
Select id_dispositivo, nombre
from dispositivo
ORDER by id_dispositivo
```

ID_DISPOSITIVO	NOMBRE
1	1 Sony Xperia
2	2 Sony Xperia
3	3 Sony Xperia
4	4 Sony Xperia
5	5 Sony Xperia
6	6 Sony Xperia
7	7 Sony Xperia
8	8 Sony Xperia
9	9 Sony Xperia
10	10 Sony Xperia
11	11 Sony Xperia
12	12 Google Nexus One
13	13 Google Nexus One
14	14 Google Nexus One
15	15 Google Nexus One
16	16 Google Nexus One
17	17 Google Nexus One
18	18 Google Nexus One
19	19 Google Nexus One
20	20 Google Nexus One

Ilustración 43: Consulta tabla Dispositivo, Dimensión Dispositivo

Fuente: Elaboración Propia

- Dimensión Ruta:** Para realizar la comparación entre las tablas, se seleccionaron el código y el nombre de cada Ruta y el código de la Región asociada, se tomó el código de la región para verificar los datos teniendo en cuenta la jerarquía existente entre la región y su ruta. Entonces si la primera consulta y la segunda consulta retornan la misma cantidad de registros y los datos coinciden, se asegura que los datos cargados en la tabla de la dimensión Ruta son consistentes, íntegros y correctos con respecto a los datos del sistema transaccional. Ver Ilustración 44.

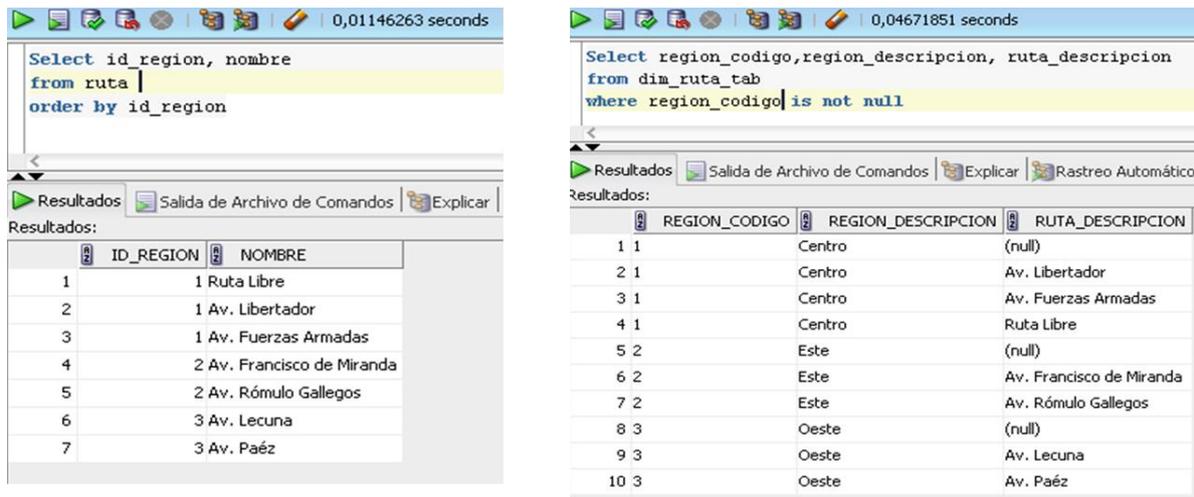


Ilustración 44: Consulta tablas Ruta y Región, Dimensión Ruta

Fuente: Elaboración Propia

En resumen, el proceso de verificación de los datos contenidos en las tablas de dimensiones es una fase importante durante el desarrollo de la solución, ya que ayuda a determinar posibles inconsistencias entre los datos, y de esta manera, realizar las correcciones necesarias antes de seguir avanzando en el desarrollo. Por lo tanto, se ahorra tiempo y eventualmente dinero.

4.3.3 Desarrollo de Requerimientos de Información

Esta fase del desarrollo involucra configuraciones de los metadatos. Esta fase inicia con la configuración del Administrador, que permite hacer la conexión entre el almacén de datos y la herramienta de Inteligencia de Negocios. Luego se procede a la creación de las consultas que satisfacen los requerimientos de información, y son estas consultas las que van a poblar los cuadros de mando de la herramienta. Por último se realiza el análisis de la consistencia de datos que se muestran en los cuadros de mando.

4.3.3.1 Configuración del Administrador

Para la integración y despliegue de la solución de inteligencia de negocio, se utilizó Oracle BI Administration Tool. Esta herramienta nos representa la capa física, la estructura lógica y la estructura de presentación en la cual se muestran los datos al usuario final. Ver Ilustración 45.

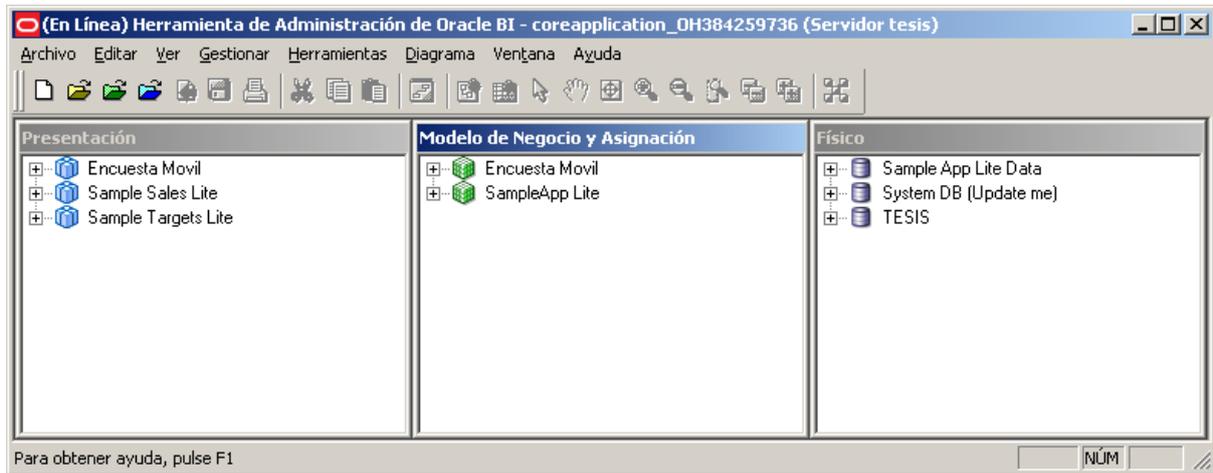


Ilustración 45: Oracle Administration Tool

Fuente: Elaboración Propia

Dentro de la capa física se especificaron las fuentes de datos físicas y las relaciones entre ellas, con las que Oracle BI Server trabaja para la elaboración de consultas. Se seleccionó la base de datos fuente a través del ODBC de TESIS especificando el esquema a usar, el cual fue OWNER.

Al importar metadatos, muchas de las propiedades de las fuentes se configuran automáticamente como lo son las relaciones de claves primarias y foráneas. En la Ilustración 46, se muestra la capa física con más detalle.

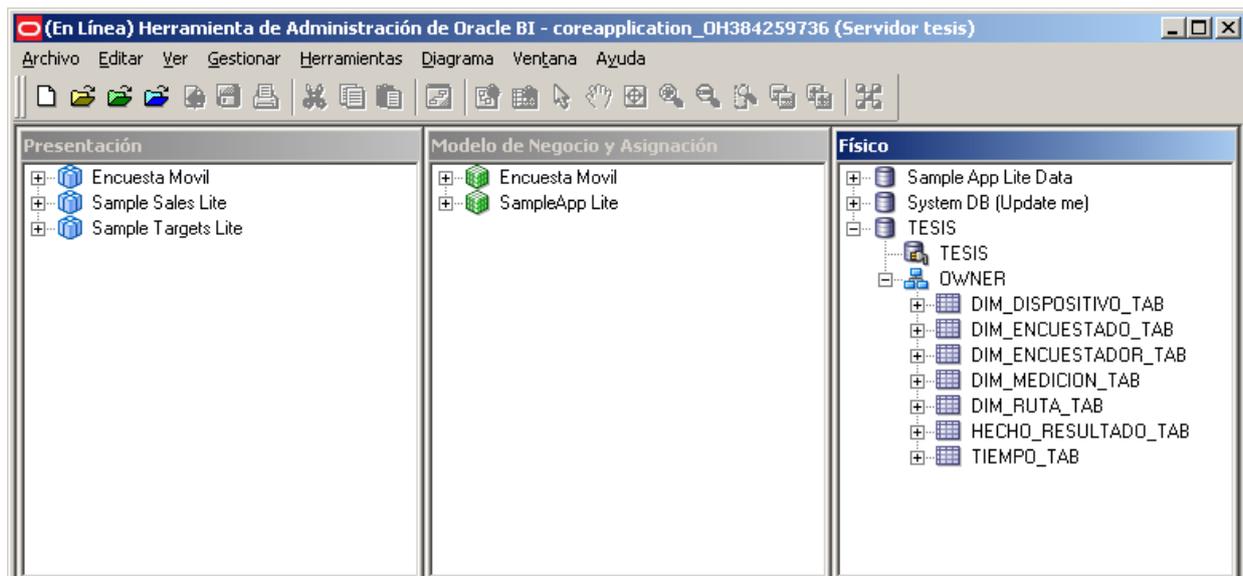


Ilustración 46: Capa Física

Fuente: Elaboración Propia

Dentro de la capa del Modelo de Negocio y Asignación, el objetivo fue captar cómo piensan los usuarios acerca del negocio utilizando su propio vocabulario, para esto se definió el modelo de negocio especificando las dimensiones, jerarquías y atributos de cada uno de los componentes y se renombraron alguno de los atributos, además de eliminar los que no eran significativos para los usuarios finales. En la Ilustración 47, se muestra la capa del Modelo de Negocio y Asignación.

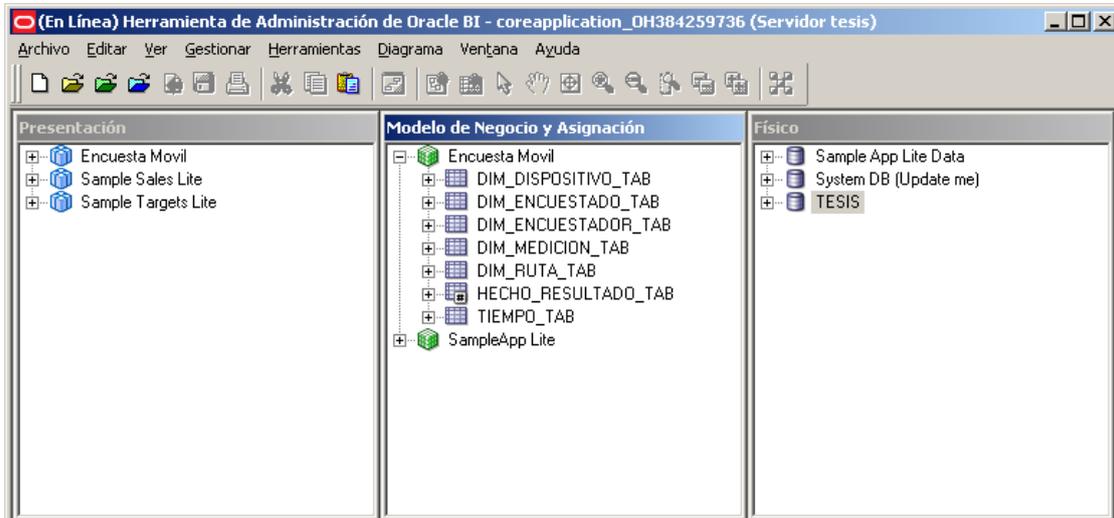


Ilustración 47: Capa de Modelo de Negocio y Asignación

Fuente: Elaboración Propia

En la capa de Presentación, se encuentran los datos que ven los usuarios con la aplicación analítica, el Oracle BI Answer. En esta capa se pueden reorganizar los datos como renombrarlos para hacerlos más entendibles para el usuario final. En este caso se dejaron los datos como los ubicados en la capa anterior, como se muestra en la Ilustración 48.

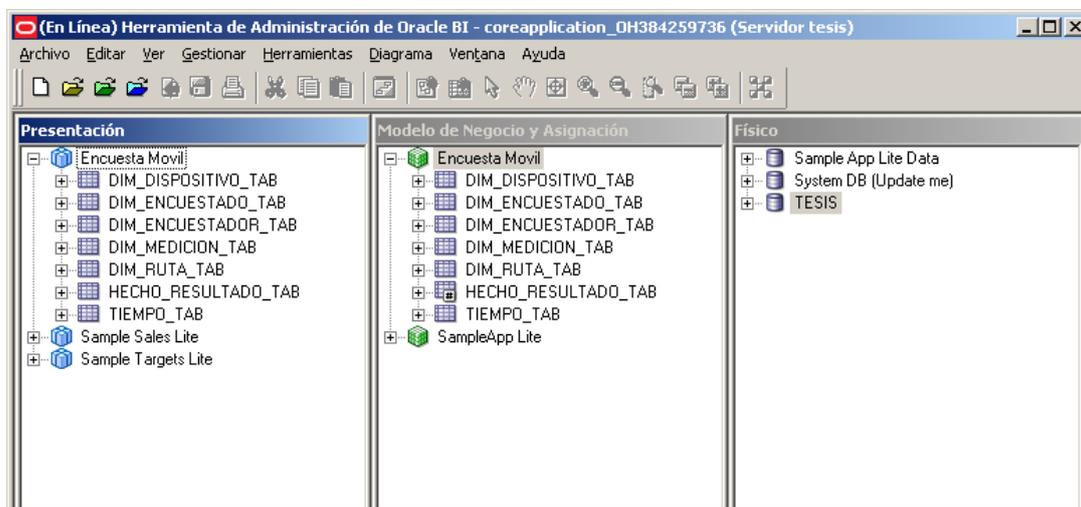


Ilustración 48: Capa de Presentación

Fuente: Elaboración Propia

Esta capa de presentación se muestra al usuario final utilizando la herramienta Oracle Business Intelligence Interactive Dashboard, en la cual se establece criterios de consultas para la creación de los indicadores y reportes mostrados a continuación.

4.3.3.2 Desarrollo de Consultas y Cuadros de Mando

Una vez realizada la integración y despliegue de los datos contenidos en la bodega, se crean las consultas necesarias para generar los reportes e indicadores descritos en la primera fase. Cada una de las consultas creadas se diseña de acuerdo a un criterio especificado que varía dependiendo del modelo de negocios. De esta manera a continuación, se describe como se diseñaron cada una de estas consultas.

En esta fase para la elaboración de los cuadros de analíticos se utilizó la herramienta Oracle Business Intelligence Answers, para luego ser mostrados en los cuadros de mando, a través de la herramienta Oracle Business Intelligence Interactive Dashboards. En las figuras siguientes, podemos observar algunos ejemplos de nuestros indicadores representados en el cuadro de mando.

El primer cuadro de mando representado en la Ilustración 49, podemos observar una pantalla inicial que contiene el título de la solución desarrollada, algunas imágenes, las áreas temáticas desarrolladas (Encuestas, Encuestador, Encuestados, Rutas y Dispositivos) y dos análisis, el primero representa la cantidad de encuestados por año y el segundo la cantidad de encuestados formulario.

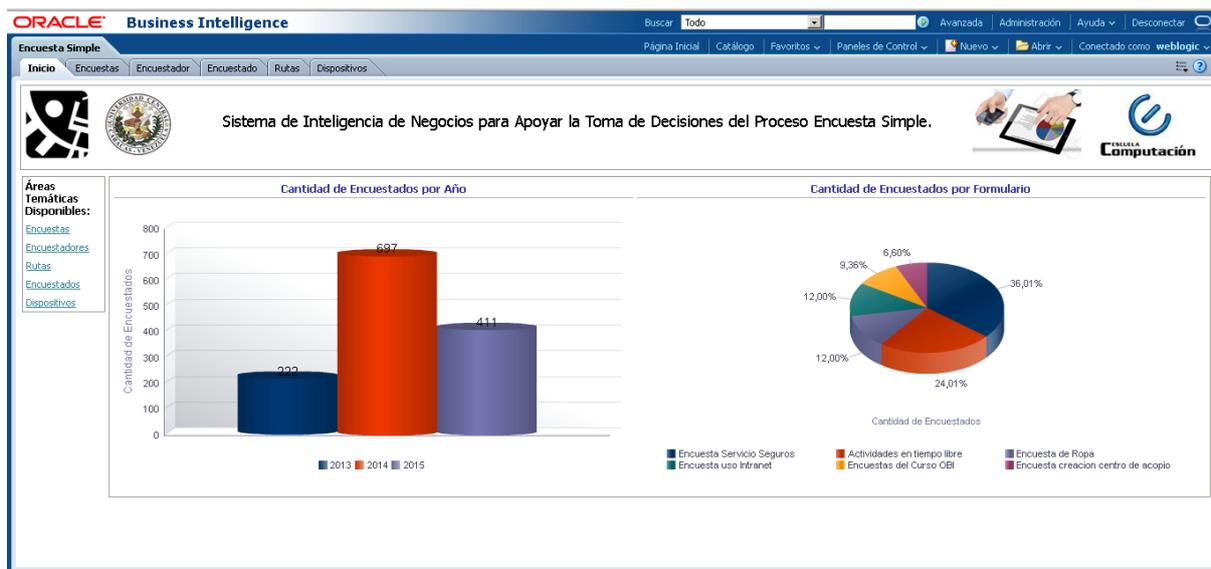


Ilustración 49: Cuadro de Mando: Inicio

Fuente: Elaboración Propia

En la Ilustración 50 podemos observar nuestro segundo cuadro de mando, llamado Encuestas. Este cuadro de mando está compuesto por cuatro análisis que serán descritos a continuación.

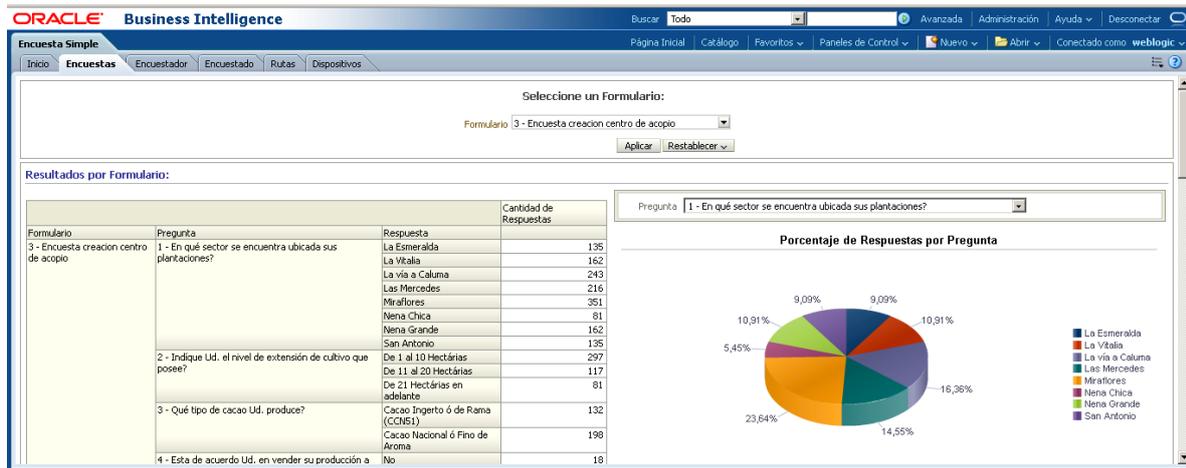


Ilustración 50: Cuadro de Mando: Encuestas

Fuente: Elaboración Propia

En la Ilustración 51, se muestra el primer análisis del cuadro de mando Encuestas, este análisis contiene la cantidad de respuestas que tuvo cada pregunta, categorizada por formulario.

Formulario	Pregunta	Respuesta	Cantidad de Respuestas
3 - Encuesta creación centro de acopio	1 - En qué sector se encuentra ubicada sus plantaciones?	La Esmeralda	135
		La Vitalia	162
		La vía a Caluma	243
		Las Mercedes	216
		Miraflores	351
		Nena Chica	81
		Nena Grande	162
		San Antonio	135
	2 - Indique Ud. el nivel de extensión de cultivo que posee?	De 1 al 10 Hectáreas	297
		De 11 al 20 Hectáreas	117
		De 21 Hectáreas en adelante	81
	3 - Qué tipo de cacao Ud. produce?	Cacao Ingerto ó de Rama (CCNS1)	132
		Cacao Nacional ó Fino de Aroma	198
	4 - Esta de acuerdo Ud. en vender su producción a un Centro de Acopio?	No	18
		Si	312
	5 - Por qué razón decidiría vender su producción a un Centro de Acopio	Calificación Transparente	90
		Facilidad de Transporte	135
		Por Pago Inmediato	105
		Precio Justo	285
		Venta Segura	210
	6 - Estaría dispuesto Ud. A seleccionar su producto, a cambio de un incentivo	No	30
Si		300	
7 - Estaría dispuesto a realizar Ud. Como productor un compromiso de venta	No	24	
	Si	306	

Ilustración 51: Cantidad de Respuestas por Formulario

Fuente: Elaboración Propia

El siguiente gráfico representado en la Ilustración 52, muestra una comparación por pregunta del porcentaje de la cantidad de respuestas que tuvo cada pregunta, por formulario.



Ilustración 52: Porcentaje de Cantidad de Respuestas por Pregunta y Formulario

Fuente: Elaboración Propia

Así mismo se realizó el siguiente análisis referente a la cantidad de encuestados de cada formulario por año y mes, como se muestra en la Ilustración 53.

FORMULARIO	AÑO		MES	CANTIDAD DE ENCUESTADOS
	2013	2014		
3 - Encuesta creacion centro de acopio	2013		09 - SEP	8
			12 - DIC	12
	2014		02 - FEB	14
			04 - ABR	11
			05 - MAY	10
			06 - JUN	32
			07 - JUL	25
			08 - AGO	12
			09 - SEP	10
			11 - NOV	12
	2015		01 - ENE	8
			07 - JUL	11
Cantidad Total de Encuestados:				55

Ilustración 53: Cantidad de Encuestados por Formulario, Año y Mes

Fuente: Elaboración Propia

Seguendo con el análisis de cantidad de encuestados por formulario, la Ilustración 54 muestra el un gráfico comparativo por mes y año, de la cantidad de encuestados por formulario.

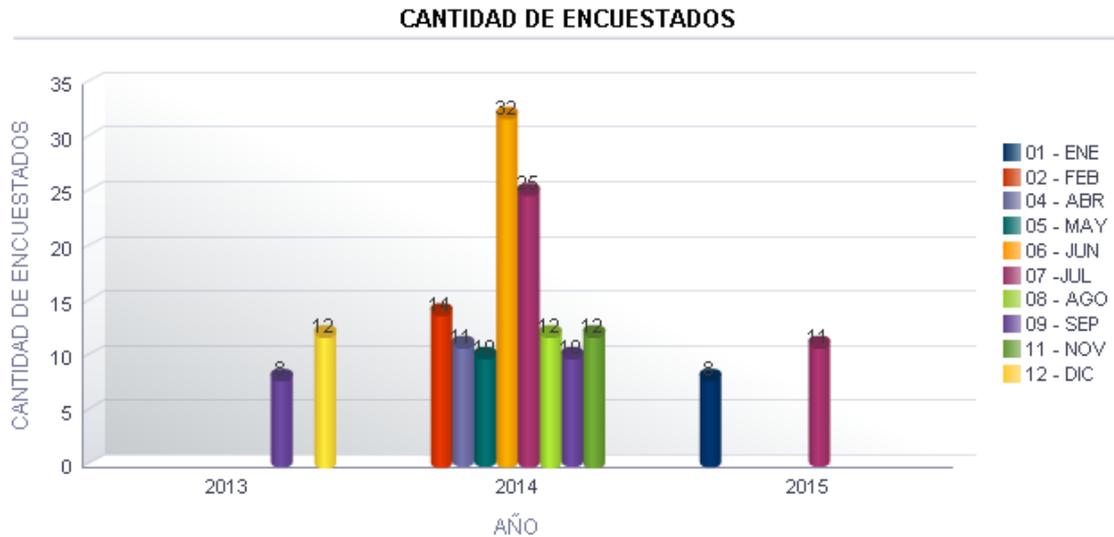


Ilustración 54: Cantidad de Encuestados por Año y Mes

Fuente: Elaboración Propia

En la Ilustración 55 podemos observar nuestro siguiente cuadro de mando, llamado Encuestador. Este cuadro de mando está compuesto por seis análisis que serán descritos a continuación.



Ilustración 55: Cuadro de Mando: Encuestador

Fuente: Elaboración Propia

En la Ilustración 56, se muestra los primeros dos análisis del cuadro de mando Encuestador, estos análisis contienen una tabla comparativa entre la cantidad de encuestados que entrevistó cada encuestador en total y por región.

		Cantidad de Encuestados					
CI_RIF	Nombre	Centro	Este	Oeste	CI_RIF	Nombre	Cantidad de Encuestados
V-10256751	Giacomo Slater	73		31	V-24280010	Tatum Cortez	108
V-12345896	Leyton Diaz	14	6		V-10256751	Giacomo Slater	104
V-12404414	Quon Robles	61		23	V-14620265	Elmo Holmes	102
V-14168950	Gisela Forbes	9		3	V-22403777	Carson Rowland	102
V-14620265	Elmo Holmes	65		37	V-12404414	Quon Robles	84
V-15451630	Yvonne Hansen	11		3	V-21392401	Alan Blackburn	47
V-18153124	Shafira Walters	5	11	5	V-21290773	David Hernandez	42
V-19650387	Pedro Perez	27	7		V-19650388	Jose Tribiño	35
V-19650388	Jose Tribiño	30	5		V-19650387	Pedro Perez	34
V-19800456	Drake Castro	5		3	V-28588464	Ifeoma Kim	23
V-20151388	Sandra Park	6		5	V-18153124	Shafira Walters	21
V-21290773	David Hernandez	33	9		V-12345896	Leyton Diaz	20
V-21392401	Alan Blackburn	34	13		V-6702078	Pearl Fry	20
V-22110717	Nigel Pitts	2	8	9	V-22110717	Nigel Pitts	19
V-22403777	Carson Rowland	71		31	V-25576614	Hop Osborne	17
V-24280010	Tatum Cortez	78		30	V-15451630	Yvonne Hansen	14
V-25354796	Octavius Cobb	6		4	V-14168950	Gisela Forbes	12
V-25576614	Hop Osborne	6	7	4	V-20151388	Sandra Park	11
V-28588464	Ifeoma Kim	4	8	11	V-25354796	Octavius Cobb	10
V-6702078	Pearl Fry	3	6	11	V-19800456	Drake Castro	8

Ilustración 56: Cantidad de Encuestados Total y por Región

Fuente: Elaboración Propia

De la misma manera se desarrolló el siguiente análisis que presenta un gráfico comparativo de la cantidad de encuestados por región de cada encuestador, como se muestra en la Ilustración 57.

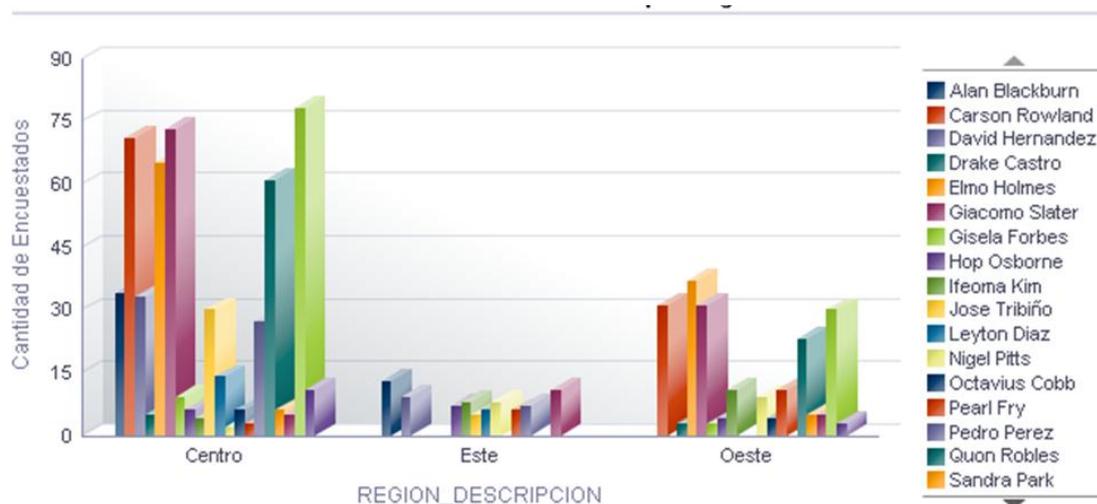


Ilustración 57: Gráfico Cantidad de Encuestados por Región

Fuente: Elaboración Propia

Siguiendo con el análisis de cantidad de encuestados por región, la Ilustración 58 muestra el un gráfico comparativo de la cantidad de encuestados en total por cada encuestador.

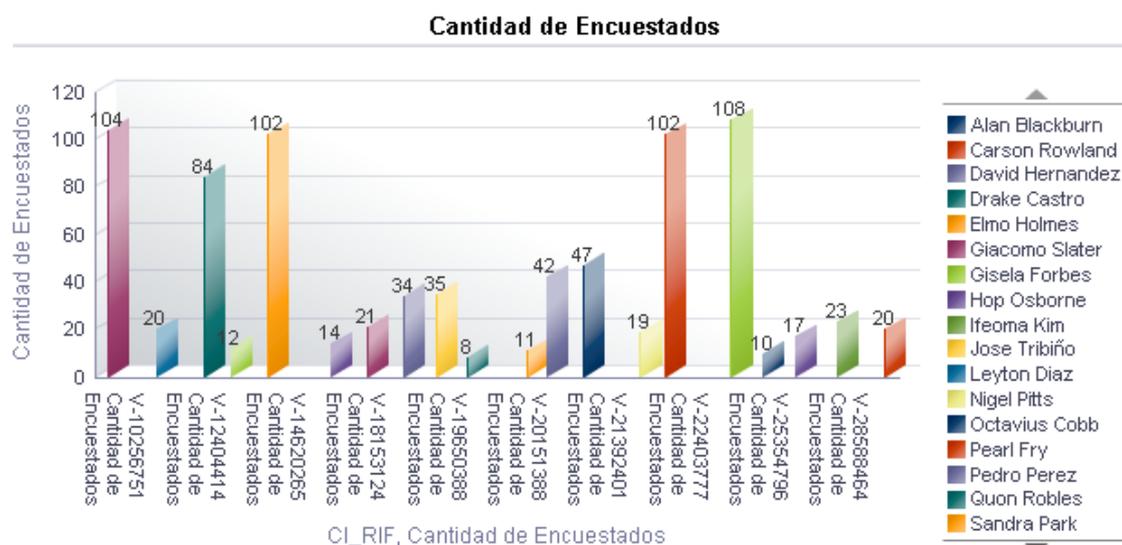


Ilustración 58: Cantidad de Encuestados en Total por Encuestador

Fuente: Elaboración Propia

En la Ilustración 59, se muestra un análisis que representa la cantidad de encuestados por encuestador y por año.

CI_RIF	Nombre	2013	2014	2015
		Cantidad de Encuestados	Cantidad de Encuestados	Cantidad de Encuestados
V-10256751	Giacomo Slater		104	60
V-12345896	Leyton Diaz		20	
V-12404414	Quon Robles	51	51	33
V-14168950	Gisela Forbes	12	12	
V-14620265	Elmo Holmes	62	102	
V-15451630	Yvonne Hansen		14	
V-18153124	Shafira Walters		21	21
V-19650387	Pedro Perez	13	13	21
V-19650388	Jose Tribiño		35	12
V-19800456	Drake Castro	8	8	8
V-20151388	Sandra Park		11	11
V-21290773	David Hernandez	20	22	22
V-21392401	Alan Blackburn		21	47
V-22110717	Nigel Pitts	19	19	19
V-22403777	Carson Rowland		66	36
V-24280010	Tatum Cortez		108	61
V-25354796	Octavius Cobb		10	
V-25576614	Hop Osborne	17	17	17
V-28588464	Ifeoma Kim		23	23
V-6702078	Pearl Fry	20	20	20

Ilustración 59: Cantidad de Encuestados por Encuestador y por Año

Fuente: Elaboración Propia

Siguiendo con el análisis de cantidad de encuestados por encuestador y año, la Ilustración 60 muestra el un gráfico comparativo de la cantidad de encuestados en total por cada encuestador categorizado por año.

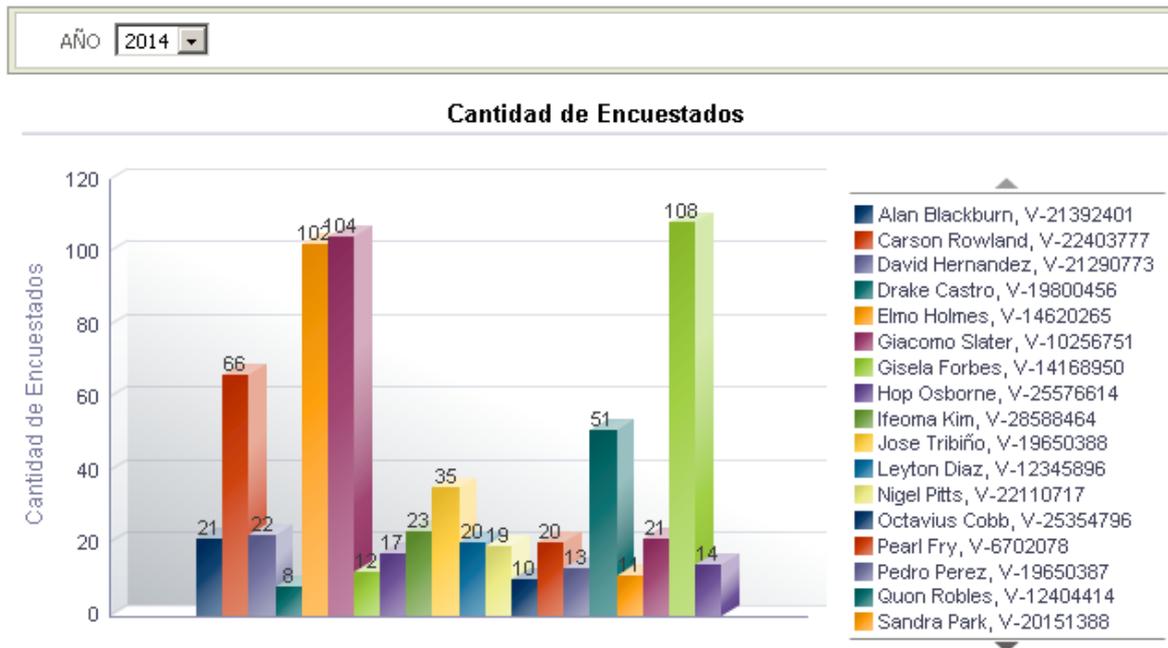


Ilustración 60: Cantidad de Encuestados Total por Encuestador y Año

Fuente: Elaboración Propia

En la Ilustración 61 podemos observar nuestro siguiente cuadro de mando, llamado Encuestado. Este cuadro de mando está compuesto por cuatro análisis que serán descritos a continuación.

CI_RIF	ENCUESTADO_NOMBRE	DIRECCION	EMAIL	TELEFONO_MOVIL	PERSONA_CONTACTO
V-10009147	Jensala Pennington	Apartado núm.: 959, 6649 Urrna,	Dus@Duisisus.co.uk	(420) 930-6947	Fuller Sykes
V-10012357	Ira Nicholson	Apdo.:850-5891 Egestas	consequat.nec.mollis@Aeneaneget.com	(861) 263-4246	MacKenzie Guthrie
V-10027049	Candice Case	Apartado núm.: 986, 6150 Elk C/	Aliquam.fringilla.cursus@Auncmauris.ca	(203) 443-6982	Bradley Hines
V-10078379	Uriel Mosley	Apartado núm.: 706, 9695 Diam Avenida	adpiscing.fringilla@semperretlacinia.net	(445) 597-1628	Aileen Briggs
V-10113475	Talon Booker	7526 Tempus Carretera	nunc.sed@insodales.edu	(325) 407-2182	Orlando Bartlett
V-10115953	Yetta Weiss	5640 Conditimentum C.	facilisis.Suspendisse@utpsum.com	(531) 118-0367	Whilmina Burke
V-10116194	Aline Dale	8789 Montes, Avda.	egestas.Aliquam.fringilla@Duisit.org	(787) 990-5888	Ruby Townsend

Ilustración 61: Cuadro de Mando: Encuestado

Fuente: Elaboración Propia

En la Ilustración 62 se muestra el primer análisis del cuadro de mando Encuestado, este análisis contienen una tabla con los distintos atributos relevantes de todos los encuestados, como lo son la cédula, el nombre, la dirección, el correo, el número telefónico y la persona de contacto.

CI_RIF	ENCUESTADO_NOMBRE	DIRECCION	EMAIL	TELEFONO_MOVIL	PERSONA_CONTACTO
V-10009147	Jamalia Pennington	Apartado núm.: 959, 6649 Urna,	Duis@Duisrisus.co.uk	(420) 930-6947	Fuller Sykes
V-10012357	Ira Nicholson	Apdo.:850-5891 Egestas	consequat.nec.mollis@Aeneaneget.com	(861) 263-4246	Mackensie Guthrie
V-10027049	Candice Case	Apartado núm.: 986, 6150 Elit C/	Aliquam.fringilla.cursus@Nuncmauris.ca	(283) 443-8882	Bradley Hines
V-10078379	Uriel Mosley	Apartado núm.: 706, 9895 Diam Avenida	adipiscing.fringilla@semperetlacinia.net	(445) 597-1628	Aileen Briggs
V-10113475	Talon Booker	7526 Tempus Carretera	nunc.sed@insodales.edu	(325) 407-2182	Orlando Bartlett
V-1011593	Yetta Weiss	5640 Condimentum C.	facilisis.Suspendisse@utipsum.com	(531) 118-0367	Whilemina Burke
V-10116194	Aline Dale	8789 Montes, Avda.	egestas.Aliquam.fringilla@Duisisit.org	(787) 990-5888	Ruby Townsend
V-10128332	Lance Meadows	172-6123 Eu, C.	risus@Nam.net	(447) 795-4953	Amena Carpenter
V-10147937	Isabelle Chavez	1626 Ornare, Calle	lorem.sit@a.edu	(737) 447-7836	Meghan Hewitt
V-10194133	Jocelyn Winters	7116 Eget, Avenida	Duis.mi.enim@accumsansed.org	(482) 318-0770	Rhea Daniel
V-10207474	Mari Espinoza	366-5547 Laoreet Avenida	posuere.at@eu.net	(233) 917-3066	Yardley McMahon
V-10250487	Renee Mckay	Apartado núm.: 426, 1442 Vitae	libero.et.tristique@nuncacmattis.net	(955) 546-0486	Malik Valencia
V-10316747	Justin Morin	9876 Sed C.	lobortis.tellus.justo@loremluctus.net	(415) 723-1128	Zelda Mcdaniel
V-10346520	Nita Patrick	8361 Quam	hymenaeos.Mauris.ut@Utsemperpretium.ca	(851) 176-9122	Delilah Baxter
V-10352753	Yen Riddle	Apdo.:841-898 Ac Calle	egestas.blandit.Nam@leoMorbineque.co.uk	(429) 813-1382	Fallon McGowan
V-10400100	Kyla Bender	Apartado núm.: 705, 2954 Metus Ctra.	eget@elit.edu	(776) 684-5066	Nolan Ford
V-10430462	Kareem Hurley	4295 Sed Av.	elementum@CuraePhasellusornare.com	(688) 199-9799	Scarlet Conway
V-10444315	Rogan Strong	6731 Diam	a@ullamcorper.net	(969) 765-9616	Buckminster Cantu
V-10554486	Donovan Workman	Apartado núm.: 134, 8151 Dignissim	ipsum.cursus.vestibulum@necimperdietnec.edu	(208) 944-5518	Carter Gallegos

Ilustración 62: Listado de Encuestados

Fuente: Elaboración Propia

El siguiente análisis presenta una tabla con el atributo cédula y nombre, de los encuestados, categorizados por región y ruta Ver Ilustración 63.

REGION_DESCRIPCION	RUTA_DESCRIPCION	CI_RIF	ENCUESTADO_NOMBRE
Centro	Av. Fuerzas Armadas	V-10430462	Kareem Hurley
		V-11655569	Boris Hopkins
		V-11918031	Cairo Avery
		V-13382458	Fleur Maddox
		V-1349493	David Mack
		V-15781527	Mariko Sheppard
		V-15979145	Noah MCGowan
		V-1784797	Abra Fuller
		V-17854691	Ignacia Sexton
		V-18256576	Frances Carr
		V-18426326	Nina Watson
		V-18488129	Samantha Horne
		V-26019263	MacKensie Alston
		V-26261811	Mona Johnston
		V-27333322	Gavin Shields
V-3282217	Rafael Lowe		
V-7058629	Emi Bolton		
V-8186343	Yvette Patton		
V-9302590	Naida Britt		

Ilustración 63: Encuestados por Región y Ruta

Fuente: Elaboración Propia

Así mismo se realizó el siguiente análisis que presenta una tabla con los encuestados categorizados por encuestador Ver Ilustración 64.

ENCUESTADOR: <input type="text" value="Alan Blackburn"/>				
CI_RIF	ENCUESTADOR	APELLIDO	CI_RIF	ENCUESTADO_NOMBRE
V-21392401	Alan	Blackburn	V-11045085	Oleg Harper
			V-11548530	Merrill Castaneda
			V-11663925	Emma Boyd
			V-11963740	Samuel Levine
			V-12079841	Veronica Merrill
			V-12234177	Keiko Rodriguez
			V-14144224	Leilani Clay
			V-14579619	Jordan Velez
			V-15214196	Rowan Johns
			V-15598557	Ursula Bradley
			V-16768675	Amaya Curry
			V-16846697	Winifred Solis
			V-17356038	Sonya Mcfadden
			V-17477585	Alden Lindsay
			V-17808743	Azalia Russell
			V-1805605	Tad Roberts
			V-18140065	Ariana Hernandez
			V-19906351	Haviva Colon
			V-20445270	Oscar Morgan
			V-20574733	Xyla Mann
			V-20659028	Brianna Bowman
			V-20960837	Quentin Craig
			V-21041194	Hyacinth Webster
			V-21892289	Jamal Odom
			V-22168871	Porter Chan

 Filas 1 - 25

Ilustración 64: Encuestados por Encuestador

Fuente: Elaboración Propia

Siguiendo con el análisis de los encuestados por encuestador, la muestra Ilustración 65 muestra una tabla comparativa entre los distintos encuestados categorizados por formulario.

NOMBRE_FORMULARIO Actividades en tiempo libre

NOMBRE_FORMULARIO	FORMULARIO_DESCRIPCION	CI_RIF	ENCUESTADO_NOMBRE
Actividades en tiempo libre	Encuesta sobre las actividades de los empleados en su tiempo libre	V-10009147	Jamalia Pennington
		V-10012357	Ira Nicholson
		V-10078379	Uriel Mosley
		V-10113475	Talon Booker
		V-10128332	Lance Meadows
		V-10207474	Mari Espinoza
		V-10554486	Donovan Workman
		V-10619434	Lois Larsen
		V-10843202	Moses Steele
		V-11172505	Mohammad Rocha
		V-11207106	Patrick Martin
		V-11492969	Cassady Carney
		V-11674384	Alden Patton
		V-11678854	Quyn Delacruz
		V-11718133	Uriel Caldwell
		V-12324833	Thor Tillman
		V-12496354	Gray Buckley
		V-12632504	Maris Norton
		V-12793342	Vielka Manning
		V-1304432	Dakota Cooke
		V-13078869	Jesse Koch
		V-1326204	Whilemina Dillon
		V-13330256	Jaime Hendrix
		V-13555060	Erich Ayers
		V-13883129	Adele Craig


 Filas 1 - 25

Ilustración 65: Encuestados por Formulario

Fuente: Elaboración Propia

En la Ilustración 66 podemos observar nuestro siguiente cuadro de mando, llamado Ruta. Este cuadro de mando está compuesto por seis análisis descritos a continuación.

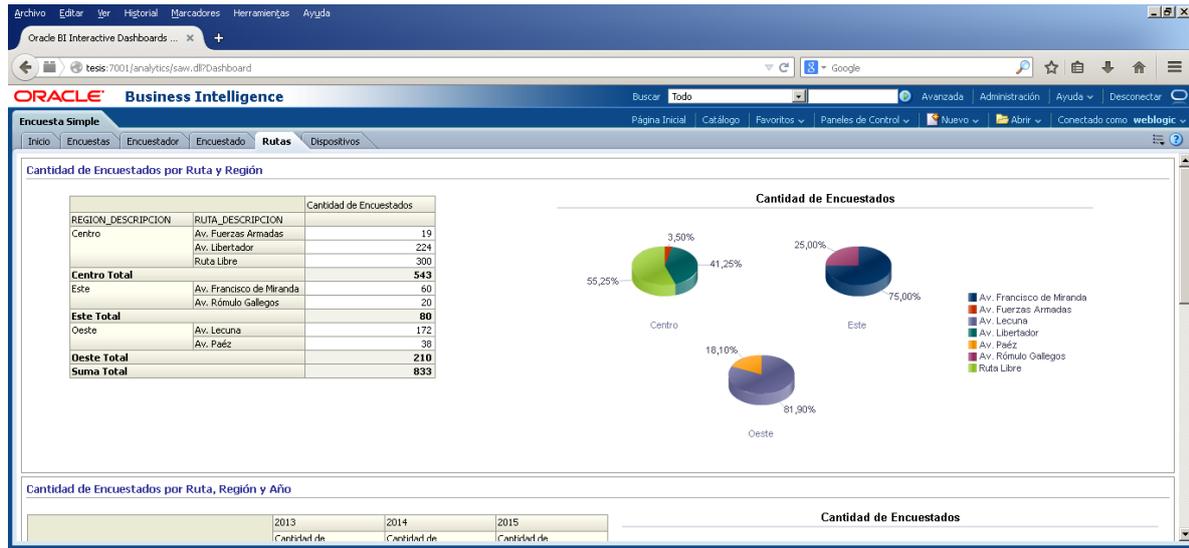


Ilustración 66: Cuadro de Mando: Ruta

Fuente: Elaboración Propia

En la Ilustración 67 se muestra el primer análisis del cuadro de mando Ruta, este análisis contiene la cantidad de encuestados agrupados por ruta y región.

REGION_DESCRIPCION	RUTA_DESCRIPCION	Cantidad de Encuestados
Centro	Av. Fuerzas Armadas	19
	Av. Libertador	224
	Ruta Libre	300
Centro Total		543
Este	Av. Francisco de Miranda	60
	Av. Rómulo Gallegos	20
Este Total		80
Oeste	Av. Lecuna	172
	Av. Paéz	38
Oeste Total		210
Suma Total		833

Ilustración 67: Cantidad de Encuestados por Región y Ruta

Fuente: Elaboración Propia

Siguiendo con el análisis anterior, la siguiente figura muestra un gráfico comparativo entre el porcentaje del número de encuestados por región y ruta.

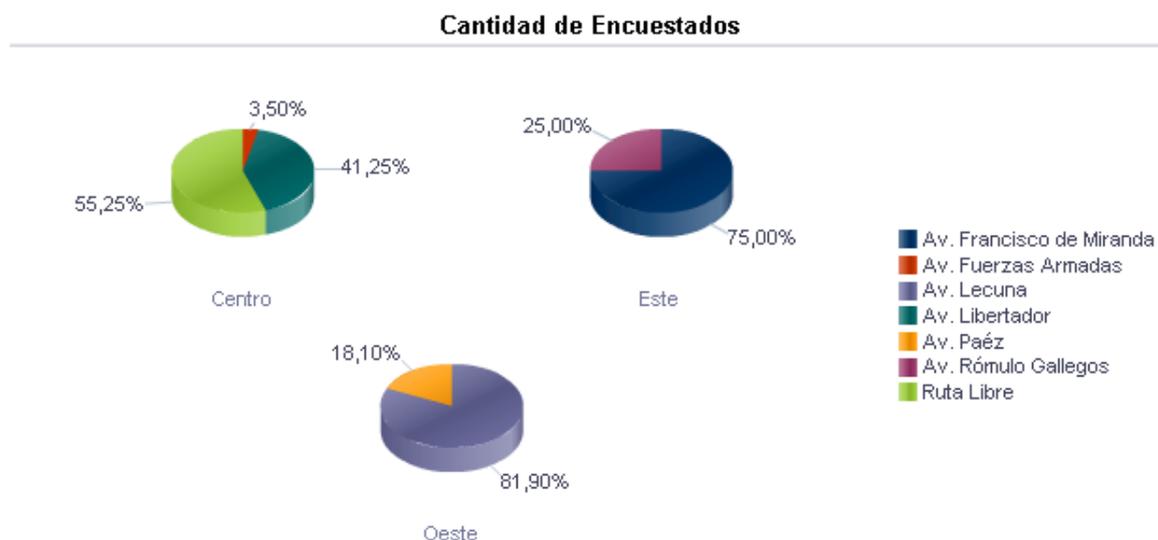


Ilustración 68: Gráfico Cantidad de Encuestados por Región y Ruta

Fuente: Elaboración Propia

De igual forma que el análisis anterior en la Ilustración 69 se presenta una tabla comparativa de la cantidad de encuestados, agrupados por ruta, región y años.

		2013	2014	2015
		Cantidad de Encuestados	Cantidad de Encuestados	Cantidad de Encuestados
REGION_DESCRIPCION	RUTA_DESCRIPCION			
Centro	Av. Fuerzas Armadas	4	19	7
	Av. Libertador	80	224	112
	Ruta Libre	20	164	116
Centro Total		104	407	235
Este	Av. Francisco de Miranda	20	60	47
	Av. Rómulo Gallegos	8	20	20
Este Total		28	80	67
Oeste	Av. Lecuna	73	172	81
	Av. Paéz	17	38	28
Oeste Total		90	210	109
Suma Total		222	697	411

Ilustración 69: Cantidad de Encuestados por Región Ruta y Año

Fuente: Elaboración Propia

Siguiendo con el análisis anterior, la siguiente figura muestra un gráfico comparativo entre el porcentaje del número de encuestados por región, ruta y año.

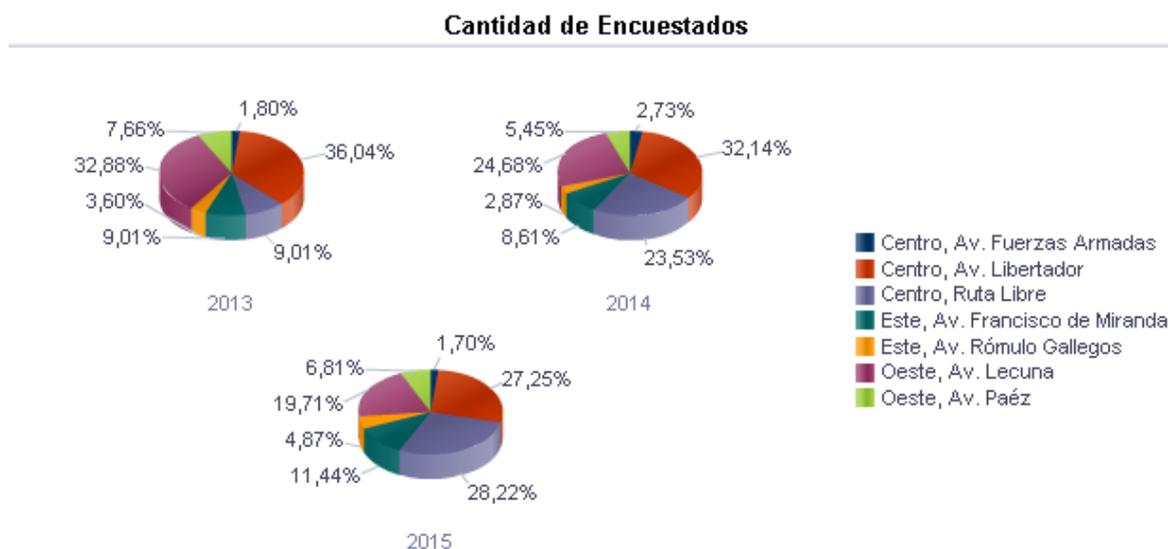


Ilustración 70: Gráfico Cantidad de Encuestados por Región Ruta y Año

Fuente: Elaboración Propia

De igual forma que el análisis anterior en la Ilustración 71 se presenta una tabla comparativa de la cantidad de encuestadores, agrupados por ruta y región.

REGION_DESCRIPCION		RUTA_DESCRIPCION	Cantidad de Encuestadores
Centro		Av. Fuerzas Armadas	5
		Av. Libertador	20
		Ruta Libre	10
Centro Total			20
Este		Av. Francisco de Miranda	10
		Av. Rómulo Gallegos	5
Este Total			10
Oeste		Av. Lecuna	10
		Av. Paéz	10
Oeste Total			15
Suma Total			20

Ilustración 71: Cantidad de Encuestadores por Región y Ruta

Fuente: Elaboración Propia

Siguiendo con el análisis anterior, la siguiente figura muestra un gráfico comparativo entre el porcentaje del número de encuestadores por región y ruta.

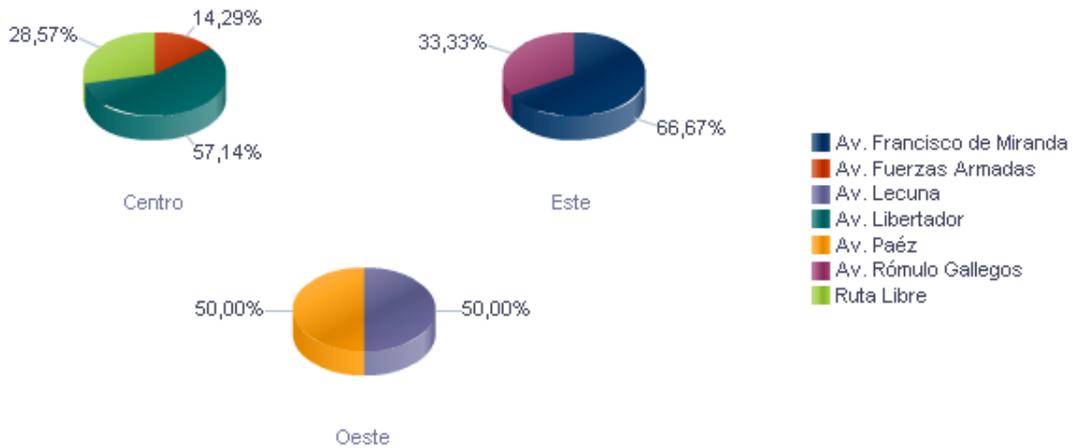


Ilustración 72: Gráfico Cantidad de Encuestadores por Región Ruta

Fuente: Elaboración Propia

Por último se presenta el cuadro de mando llamado Dispositivos, que está formado por dos análisis, el primero contiene una tabla con el listado de todos los dispositivos y sus atributos más relevantes y su encuestador asociado. Ver Ilustración 73.

ORACLE Business Intelligence

Encuesta Simple

Inicio Encuestas Encuestador Encuestado Rutas **Dispositivos**

Dispositivos:

MARCA	MODELO	OPERADORA	PLAN_DATOS	IMEI	CAMARA	WIFI	CÉDULA	ENCUESTADOR
Google	Nexus One	Digitel	1GB	BKG66VYV0HX		1	1 V-28588464	Sofia Kim
				HJQ85YDQ7EJ		1	1 V-25354796	Octavius Cobb
				KZI7905A4EK		1	1 V-14168950	Gisela Forbes
				QPG01CBD6HY		1	1 V-6702078	Pearl Fry
				TOQ05DJCIYE		1	1 V-22110717	Roberto Pitts
				XYO10BBF8WH		1	1 V-18153124	Shafira Walters
				YBV14JDM8XL		1	1 V-25576614	Raul Osborne
				YSO24RDL1RH		1	1 V-20151388	Sandra Park
				ZMO07BAM5YN		1	1 V-19800456	Drake Castro
				Sony	Xperia	Movistar	500MB	AWQ47XPB3LM
BSI82MU57BM		1	1 V-24280010					Ricardo Cortez
HCQ29WEB0GU		1	1 V-14620265					Elmo Holmes
KGL96QUL6EQ		1	1 V-15451630					Yvonne Hansen
MMK43RMM4MO		1	1 V-12404414					Jose Robles
MTX77OQJ2YM		1	1 V-21392401					Alan Blackburn
OCT52LJVM8DV		1	1 V-21290773					David Hernandez
Ovh92ADC2YG		1	1 V-19650388					Simon Tribiño
RRD48NSP7RH		1	1 V-19650387					Pedro Perez
SNS23ZKC3H5		1	1 V-12345896					Leyton Diaz
YHF73TNN0IN		1	1 V-10256751	Giacomo Slater				

Ilustración 73: Dispositivos y su Encuestador Asociado

Fuente: Elaboración Propia

Y el segundo y último análisis contiene un gráfico comparativo entre las distintas marcas de dispositivos utilizadas por los encuestadores Ver Ilustración 74.



Ilustración 74: Comparación del Porcentaje del Número de Dispositivos por Marca.

Fuente: Elaboración Propia

4.3.3.3 Verificar Calidad de Datos

Luego de crear las distintas consultas para distribuirlas en sus respectivos cuadros de mando, se realiza la última fase del método de desarrollo, el análisis de consistencia de los cuadros de mando. Esta fase se verificó la lógica de los datos contenidos en los cuadros de mando y su validez con respecto al almacén de datos, es decir, se analizó por medio del Oracle BI Interactive Dashboards que los distintos cuadros de mando tuvieran la información esperada y que los reportes tuvieran una lógica que apoye la toma de decisiones por parte de la alta gerencia.

También se verificó que la información de los cuadros de mando fuera consistente con la que se tiene tanto en la base de datos fuente como en el Almacén de Datos, para lograr esto, se elaboran consultas que reflejaron la misma información que los cuadros de mando, utilizando las herramientas de consulta asociadas a la fuente de datos y al almacén de datos.

CONCLUSIONES

El objetivo principal de este Trabajo Especial de Grado fue alcanzado exitosamente al lograr diseñar y desarrollar una solución de inteligencia de negocio del Sistema de Información Encuesta Simple a partir de su Modelo Transaccional de Datos, con la finalidad de apoyar la toma de decisiones precisas y a tiempo. En este sentido, se logró sustentar la toma de decisiones gracias a los indicadores y reportes que arrojan resultados reales de las encuestas simples aplicadas, permitiendo a los usuarios consultar en detalle y generar conocimiento sobre el negocio.

Se logró desarrollar una solución de Inteligencia de Negocios que, apoyado en un almacén de datos robusto y consistente, permite generar y representar el modelo de información del negocio, independientemente del modelo de información representado por los sistemas transaccionales. Por lo tanto, se obtiene una visión global de la información estratégica del negocio, la cual ayuda a definir un conjunto de reglas que permiten tener un control bien definido sobre dicha información.

También se diseñó con éxito un modelo dimensional gracias al uso de la metodología de Tian Consultores que como se mencionó anteriormente, es una adaptación de los enfoques ascendente (Kimball, R.) y descendente (Inmon, B.), facilitó su diseño y desarrollo, la cual nos sirvió de guía para la elaboración de cada uno de los componentes que conforman a la solución a través de una serie de pasos sencillos y fáciles de entender.

Adicionalmente, el seguimiento constante de la calidad de los datos en todas las fases del desarrollo, aseguró la confiabilidad de los resultados que se presentaron en las consultas y cuadros de mando. En consecuencia, el uso del método empleado cubrió las expectativas de los usuarios sobre sus requerimientos.

Se desarrollaron e implementaron los procesos encargados de extraer, transformar y cargar los datos desde las fuentes de datos, hacia el almacén de datos. Una vez construido y poblado el almacén de datos, se logró diseñar e implementar las consultas analíticas y reportes que dan respuesta a los indicadores de gestión que fueron definidos para esta solución de inteligencia de negocio, a través de una interfaz amigable y de fácil manejo.

Por último la metodología Tian Consultores en su cuarta fase expone un conjunto de pasos basados en el correcto funcionamiento de la tecnología, los datos y las aplicaciones de usuarios finales. Esta etapa implica una verificación del sistema por parte de los usuarios finales que conllevan el ajuste de las consultas, se configuran los esquemas de seguridad necesarios, se crean y programan los flujos de procesos necesarios para el poblamiento periódico y automático del almacén de datos y por último se procede a adiestrar a los usuarios finales. Estas actividades no están incluidas en el alcance del presente trabajo de investigación, por lo que no fueron desarrolladas.

En definitiva, la adquisición de sistemas o aplicaciones de Inteligencia de Negocios es determinante para las organizaciones que buscan apoyar y respaldar sus decisiones en información estratégica basada en eventos que vayan acorde con la realidad del negocio, y no basada en decisiones intuitivas tomadas a última hora. Incorporando este tipo de soluciones, facilitan a la gerencia media y alta el proceso de la toma de decisiones. Por lo tanto, el uso de este tipo de soluciones representa una ventaja competitiva de gran valor para las organizaciones que deseen mejorar su productividad

REFERENCIAS BIBLIOGRÁFICAS Y DIGITALES

Davenport, T. H. & Prusak, L. (2000). *Working knowledge: How organizations manage what they know*. Boston, MA: Harvard Business School Press.

Ennio Prada Madrid (2008) “Los insumos invisibles de decisión: datos, información y conocimiento”, <http://eprints.rclis.org/14256/1/ad11110.pdf>

Gerardo Arceo Moheno (2010). “El impacto de la gestión del conocimiento y las tecnologías de información en la innovación: un estudio en las pyme del sector agroalimentario de cataluña”, <http://www.eumed.net/tesis-doctorales/2010/gam/El%20conocimiento%20y%20su%20gestion.htm>

Alfonso Jimenez (2006). “Qué es la Inteligencia”, <http://www.xatakaciencia.com/otros/que-es-la-inteligencia>

Fuentes A. (2012). “Actividades de los sistemas de información”, http://sistemas-de-de-informacion.blogspot.com/p/actividades-basicas-de-un-sistema-de_06.html

Richardo R. (2008) “Principios de sistemas de información.”, <http://richardunefa.wordpress.com/principios-de-sistemas-de-informacion/>

Manuel Peralta (2006), “Sistemas de Información”, <http://www.monografias.com/trabajos7/sisinf/sisinf.shtml>

Ivan Turmero (2007) “Sistemas de Información Transaccionales”, <http://www.monografias.com/trabajos94/sistemas-informacion-transaccionales-bancos/sistemas-informacion-transaccionales-bancos.shtml#ixzz2pHaM3z2L>

R. Elmasri y S. Navathe (2002). *Fundamentos de los Sistemas de Bases de Datos* (3ª edición). Addison-Wesley.

A. Silberschatz, H. F. Korth y S. Sudarshan (2002). *Fundamentos de Bases de Datos* (4ª edición). McGraw Hill.

Laudon Keneth C. (2004), Laudon Jane P., *Sistemas de Información Gerencial* (8ª ed.)

María Isabel Guzmán (2001), “Implementación de un data warehouse para el soporte de toma de decisiones”, Guatemala.

Sabina Valeska Carrera Sepúlveda (2009), “Transformación de esquemas Multidimensionales difusos del Nivel conceptual al nivel Lógico”.

Jose Curto Díaz (2007), “Data Warehouse, Snow Flake Schema, Star Schema”, Venezuela.

Kimball (1998) et al., "The Data Warehouse Lifecycle Toolkit". New York, Wiley.
http://pic.dhe.ibm.com/infocenter/idm/docv3/index.jsp?topic=%2Fcom.ibm.datatools dimensional.ui.doc%2Ftopics%2Fc_dm_dimschemas.html

Cano, J. L. (2007). Business Inteligence. In J. L. Cano, Business Inteligence (p. 393).

Davenport, T. H., & Harris, J. G. (2009). Competing on Analytics. In T. H. Davenport, & J. G. Harris, Competing on Analytics (p. 276). México D.F.: Alfaomega.

Mary Breslin (2000), "Data Warehousing Battle of the Giants: Comparing the Basics of the Kimball and Inmon Models".

Bill Inmon (2002). "W.H. Building the Data Warehouse" (Third Edition), New York: John Wiley & Sons

Bill Inmon (2002). "Glossary of Data Warehouse", <http://www.inmoncif.com/library/glossary/#D>

Osmar Horacio Saldaño (2009), "Metodología de la Investigación",
<http://www.promonegocios.net/mercadotecnia/encuestas-tipos.html>

Ivan Thompson (2006), "Tipos de encuesta", <http://www.tiposde.org/escolares/123-tipos-de-encuestas/>

Stanton, Etzel y Walker, Mc Graw Hill, (2004). "Fundamentos de Marketing", 13a. Edición, España.

Etzel y Walker, Mc Graw Hill, (2004), "Diccionario de Marketing", de Cultural S.A., Edición 1999, España.

TiposDe.Org (2014). "Tipos de encuesta". <http://www.tiposde.org/escolares/123-tipos-de-encuestas/>

Eduardo Plasencia Alvarado (2013), "Como hacer una encuesta",
<http://blogparamaestros.blogspot.com/2013/03/como-hacer-una-encuesta-para-tesis.html>

Robert Johnson & Patricia Kuby (2005). "*Estadística elemental, lo esencial (3ª ed)*". Venezuela.

RRHH Portal (2007). "La encuesta y tipos de encuesta",
<http://www.gestiopolis.com/recursos/documentos/fulldocs/mar/tipencuch.htm>

Ruiz Avellan Joaquin (2007). "El cuestionario estructurado como herramienta básica para la evaluación de las instituciones documentales".

Leal, L. J. (2011). "Cómo diseñar una encuesta V",
<http://asesoriatesis1960.blogspot.com/2011/01/como-disenar-una-encuesta-iv.html>.

García Ferrando, M (1986). "El Análisis de la realidad social. Métodos y Técnicas de Investigación", Ed. Alianza, Madrid.

Mancilla Toledo, Nelson (1998). ¿Cómo Hacer un seminario de Título? Cuantitativa N°1. Ed. Mago Producciones, Santiago de Chile.

L. Festinger y D. Katz (1998). “Los Métodos de Investigación en Ciencias Sociales. Ed. Alianza”.

Uzcátegui, L. F. (2014). Manual “Metodología de Desarrollo de Sistemas de Información de Inteligencia de Negocio” de Tian Consultores. Caracas, Venezuela.

Lic. Franky Uzcátegui y Lic. César Bravo (2014). “Documentación del Sistema de Información Encuesta Simple”. Caracas, Venezuela.