



Universidad Central de Venezuela
Facultad de Ciencias
Escuela de Computación

Desarrollo de un sistema de Inteligencia de Negocios para el apoyo en la toma de decisiones en una empresa de servicios de procesamiento, recogida y entrega de valores.

Trabajo Especial de Grado presentado ante la ilustre Universidad Central de Venezuela por las
bachilleres

Br. Cumana, Josuahna
Br. Requena, María Andreina
Para optar por el título de
Licenciado en Computación

Tutor: Prof. Rangel, Wilfredo
Ciudad Universitaria, Octubre 2017

ACTA

Quienes suscriben, miembros del Jurado designado por el Consejo de la Escuela de Computación, para examinar el Trabajo Especial de Grado presentado por el bachiller Josuahna Cumana CI 17.746.012, con el título **“Desarrollo de un sistema de Inteligencia de negocios para el apoyo en la toma de decisiones en una empresa de servicios de procesamiento, recogida y entrega de valores”** a los fines de optar por el título de Licenciado en Computación, deja constancia de lo siguiente:

Leído como fue dicho por cada uno de los miembros del jurado, se fijó el día 04 de octubre de 2017 a las 12:00 m, para que los autores lo defendieran en forma pública, lo que se hizo en el salón PB-3 de la Escuela de Computación mediante una presentación oral de su contenido, luego de lo cual respondió a las preguntas formuladas.

Finalizada la defensa pública del Trabajo Especial de Grado, el jurado decidió aprobarlo con la calificación de 18 puntos.

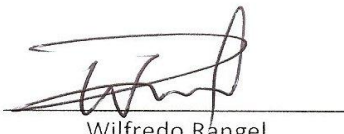
En fe de lo cual se levanta la presente Acta, en Caracas a los 04 días del mes de octubre del año 2017.



Antonio Silva
Jurado



Ana Morales
Jurado



Wilfredo Rangel
Tutor

ACTA

Quienes suscriben, miembros del Jurado designado por el Consejo de la Escuela de Computación, para examinar el Trabajo Especial de Grado presentado por el bachiller María Andreina Requena CI 13.477.729, con el título **“Desarrollo de un sistema de Inteligencia de negocios para el apoyo en la toma de decisiones en una empresa de servicios de procesamiento, recogida y entrega de valores”** a los fines de optar por el título de Licenciado en Computación, deja constancia de lo siguiente:

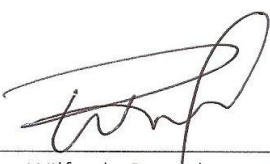
Leído como fue dicho por cada uno de los miembros del jurado, se fijó el día 04 de octubre de 2017 a las 12:00 m, para que los autores lo defendieran en forma pública, lo que se hizo en el salón PB-3 de la Escuela de Computación mediante una presentación oral de su contenido, luego de lo cual respondió a las preguntas formuladas.

Finalizada la defensa pública del Trabajo Especial de Grado, el jurado decidió aprobarlo con la calificación de 18 puntos.

En fe de lo cual se levanta la presente Acta, en Caracas a los 04 días del mes de octubre del año 2017.



Antonio Silva
Jurado

Ana Morales
Jurado

Wilfredo Rangel
Tutor

Universidad Central de Venezuela
Facultad de Ciencias
Escuela de Computación

Desarrollo sistema de Inteligencia de negocios para el apoyo en la toma de decisiones en una empresa de servicios de procesamiento, recogida y entrega de valores.

Autores: Br. Cumana Josuahna
Br. María Andreina Requena
Tutor: Prof. Wilfredo Rangel
Fecha: Octubre 2017

RESUMEN

Un elemento fundamental en el proceso de Atención al Cliente es la gestión de quejas y reclamos. Para que un modelo de atención sea exitoso debe asegurar dos premisas: asegurar la satisfacción del cliente en la gestión de sugerencias, quejas y reclamos y a partir de la información que se maneja generar informes y estadísticas que permitan tomar decisiones para el mejoramiento continuo en los servicios que se prestan y en caso de ser necesario hacer los correspondientes correctivos. La empresa de nuestro caso de estudio presta servicios de recogida, transporte y procesamiento de valores; actualmente dispone de un Sistema de Gestión que permite registrar y realizar seguimiento de las sugerencias, quejas y reclamos de los clientes, además de canalizarlos hacia las distintas áreas resolutorias internas involucradas en su resolución para que sean gestionados y procesados en forma oportuna y completa, sin embargo el área Atención al Cliente no cuenta con estrategias que permitan manipular estos datos y brindar información estadística para la toma de decisiones en la organización. Bajo este contexto, el presente Trabajo Especial de Grado (T.E.G) propone el análisis, diseño e implementación de una sistema que a partir de la información suministrada de las quejas y reclamos, permita generar indicadores que sirvan de apoyo a la toma de decisiones en la Gerencia de Atención al Cliente y a la organización como tal, basándonos en los conceptos de Inteligencia de Negocios (BI), haciendo uso de la metodología planteada por Kimball (BDL, por sus siglas en inglés) y las tecnologías de Pentaho BI Suite Community Edition, PostgreSQL y Community Dashboard Editor (CDE).

Palabras clave: Inteligencia de Negocios.

Índice

ACTAS	2
RESUMEN	4
Índice	5
Índice de ilustraciones	6
Índice de tablas	6
INTRODUCCIÓN	9
CAPITULO I- Problema de Investigación	12
PLANTEAMIENTO DEL PROBLEMA Y JUSTIFICACIÓN DE LA INVESTIGACIÓN	12
OBJETIVO DE LA INVESTIGACIÓN	14
OBJETIVOS ESPECÍFICOS	14
ALCANCE	15
CAPÍTULO II - Marco Conceptual	16
Conceptos básicos	16
Servicio al Cliente	17
Sistemas de Información	18
Funciones de un Sistema de Información	20
Tipos de Sistemas de Información	21
Inteligencia de negocios	23
Arquitectura y componentes de un BI	23
Beneficios de los BI	25
Competencias de los BI	26
Los 5 estilos de un BI en un escenario empresarial	28
Plataforma de Inteligencia de Negocios Pentaho BI Suite Community Edition	30
Bases de datos relacionales	32
Data Warehouse	32
Características de un Data Warehouse	33
Ventajas y desventajas de un Data Warehouse	34
Arquitectura de un Data Warehouse	34
Conceptos Asociados a un Data Warehouse	35
Formatos de almacenamiento y MDX	39
Dimensión, tablas y tablas de hechos	40
Sistemas OLTP	42
Diferencias entre OLTP y OLAP	42
Diferencias entre OLTP y Data Warehouse	43
Diseño de un data warehouse propuesto por Kimball	44
CAPÍTULO III – Marco Aplicativo	51
Planificación del Proyecto	51
Definición de los requerimientos del negocio	52
Selección de herramientas	55
Selección de la herramienta de Inteligencia de Negocios	56

Selección del Sistema Manejador de Base de Datos	59
Diseño técnico de la arquitectura	60
Modelo dimensional	60
Diseño físico	66
Diseño de procesos ETL.....	70
Especificación de la aplicación de usuario final	78
Implementación	79
Despliegue	93
Mantenimiento y crecimiento	93
Capítulo IV – Conclusiones.....	94
REFERENCIAS BIBLIOGRÁFICAS.....	95

Índice de ilustraciones

Ilustración 1: Datos e Información (Laudon&Laudon, 2012).....	19
Ilustración 2: Funciones de un sistema de información (Laudon&Laudon, 2012)	20
Ilustración 3: Tipos de sistemas de información (Laudon&Laudon, 2012)	21
Ilustración 4: Arquitectura conceptual de un BI	24
Ilustración 5: Arquitectura analítica de un BI	25
Ilustración 6: Estilos de BI y grupos de usuarios (MicroStrategy)	29
Ilustración 7: Arquitectura de un Data Warehouse.....	35
Ilustración 8: Diferencias entre OLTP y OLAP	43
Ilustración 9: Diagrama ciclo de vida dimensional del negocio.....	44
Ilustración 10: Relaciones de la tabla de hechos fact_incidente.....	70
Ilustración 11: Ejemplo de la hoja de cálculo	71
Ilustración 12: JobServPan	75
Ilustración 13: dim_cliente	75
Ilustración 14: dim_categoria	76
Ilustración 15: dim_oficina	76
Ilustración 16: dim_estado	77
Ilustración 17: dim_fecha	77
Ilustración 18: dim_traza	78
Ilustración 19: fact_incidente	78
Ilustración 20: Página de inicio de sesión en el sistema	80

Índice de tablas

Tabla 1: Planificación del proyecto	52
Tabla 2: Ficha técnica del requerimiento 1.....	53

Tabla 3: Ficha de especificación del requerimiento 1	53
Tabla 4: Ficha técnica del requerimiento 2.....	54
Tabla 5: Ficha de especificación del requerimiento 2	55
Tabla 6: Cuadro comparativo de productos BI Open Source	58
Tabla 7: Comparación entre SMBD's	59
Tabla 8: Dimensiones	60
Tabla 9: Jerarquía Dimensión Área	61
Tabla 10: Atributos Dimensión Área	61
Tabla 11: Jerarquía Dimensión Cliente	61
Tabla 12: Atributos Dimensión Cliente	61
Tabla 13: Jerarquía Dimensión Estado.....	62
Tabla 14: Atributos Dimensión Estado	62
Tabla 15: Jerarquía Dimensión Fecha	62
Tabla 16: Atributos Dimensión Fecha	62
Tabla 17: Jerarquía Dimensión Oficina	62
Tabla 18: Atributos Dimensión Oficina	63
Tabla 19: Jerarquía Dimensión Producto.....	63
Tabla 20: Atributos Dimensión Producto.....	63
Tabla 21: Jerarquía Dimensión Tipo de requerimiento	63
Tabla 22: Atributos Dimensión Tipo de requerimiento	63
Tabla 23: Jerarquía Dimensión Requerimiento	64
Tabla 24: Atributos Dimensión Requerimiento	64
Tabla 25: Jerarquía Dimensión Sucursal	64
Tabla 26: Atributos Dimensión Sucursal	64
Tabla 27: Jerarquía Dimensión Traza	64
Tabla 28: Atributos Dimensión Traza.....	65
Tabla 29: Tabla de hechos.....	65
Tabla 30: Dimensiones vs. Hechos.....	65
Tabla 31: Granularidad fact_incidente	66
Tabla 32: Medidas fact_incidente.....	66
Tabla 33: Diseño físico de la dimensión Área	67
Tabla 34: Diseño físico de la dimensión Cliente	67
Tabla 35: Diseño físico de la dimensión Estado	67
Tabla 36: Diseño físico de la dimensión Fecha	68
Tabla 37: Diseño físico de la dimensión Oficina	68
Tabla 38: Diseño físico de la dimensión Producto	68
Tabla 39: Diseño físico de la dimensión Tipo de requerimiento	68
Tabla 40: Diseño físico de la dimensión Requerimiento	69
Tabla 41: Diseño físico de la dimensión Sucursal	69
Tabla 42: Diseño físico de la dimensión Traza	69
Tabla 43: Diseño físico de la tabla de hechos Incidente	70
Tabla 44: Definición de tabla dimensión dim_cliente	71

Tabla 45: Definición de tabla dimensión dim_req_tipo	72
Tabla 46: Definición de tabla dimensión dim_area	72
Tabla 47: Definición de tabla dimensión dim_producto	72
Tabla 48: Definición de tabla dimensión dim_requerimiento	72
Tabla 49: Definición de tabla dimensión dim_oficina	72
Tabla 50: Definición de tabla dimensión dim_sucursal	72
Tabla 51: Definición de tabla dimensión dim_fecha	73
Tabla 52: Definición de tabla dimensión dim_estado	73
Tabla 53: Definición de tabla dimensión dim_traza	73
Tabla 54: Mapeo de tabla dimensión dim_cliente	73
Tabla 55: Mapeo de tabla dimensión dim_req_tipo	73
Tabla 56: Mapeo de tabla dimensión dim_area	73
Tabla 57: Mapeo de tabla dimensión dim_producto	73
Tabla 58: Mapeo de tabla dimensión dim_requerimiento	73
Tabla 59: Mapeo de tabla dimensión dim_oficina	74
Tabla 60: Mapeo de tabla dimensión dim_sucursal	74
Tabla 61: Mapeo de tabla dimensión dim_fecha	74
Tabla 62: Mapeo de tabla dimensión dim_estado	74
Tabla 63: Mapeo de tabla dimensión dim_traza	74
Tabla 64: Entregables ETL	75
Tabla 65: Roles de usuario para el Sistema Automatizado de Inteligencia de Negocios	79
Tabla 66: Reportes desarrollados	79
Tabla 67: Vistas Dinámicas desarrolladas	79
Tabla 68: Dashboard desarrollados	79
Tabla 69: Descripción del entregable Vista dinámica Casos creados por cliente	81
Tabla 70: Descripción del entregable Vista dinámica Casos creados por estado	82
Tabla 71: Descripción del entregable Vista dinámica Casos creados por fecha	83
Tabla 72: Descripción del entregable Vista dinámica Casos creados por oficina	84
Tabla 73: Descripción del entregable Vista dinámica Casos creados por oficina y requerimiento	85
Tabla 74: Descripción del entregable Vista dinámica Casos creados por tipo de requerimiento	85
Tabla 75: Descripción del entregable Vista dinámica Casos cerrados por fecha	86
Tabla 76: Descripción del entregable Vista dinámica Casos cerrados por oficina y estado	86
Tabla 77: Descripción del entregable Reporte Casos por cliente	88
Tabla 78: Descripción del entregable Reporte Casos por tipo de requerimiento	90
Tabla 79: Descripción del entregable Dashboard Casos anualmente creados	91
Tabla 80: Descripción del entregable Dashboard Casos anualmente cerrados	92
Tabla 81: Pruebas de aceptación de los usuarios	93

INTRODUCCIÓN

El desarrollo de soluciones de Inteligencia de Negocios permite a cualquier organización o empresa acceder a información cuyo fin último es facilitar y optimizar el proceso de toma de decisiones. Por lo general toda la información que debe manipularse proviene de fuentes de diversas, tales como archivos de Excel, bases de datos, servidores, etc.; y son estos sistemas los encargados de integrarla y transformarla para posteriormente almacenarla en estructuras multidimensionales. El análisis, diseño y modelado de estas estructuras son las que permitirán hacer las combinaciones necesarias para dar sentido y mostrar la información relevante y necesaria para el negocio. Finalmente, el resultado de la solución de inteligencia de negocios se verá plasmada a través de diversas herramientas que permitirán generar y obtener reportes o vistas dinámicas se generen con la información obtenida de la solución.

La organización de nuestro caso de estudio tiene como objetivo suministrar servicios y productos para el manejo integral de valores y riesgos asociados con seguridad, respaldo y tecnología. Esta organización cuenta con una trayectoria de más de 50 años en el mercado y está conformada por varias unidades de negocios a nivel nacional e internacional, manteniendo el liderazgo en su ramo dada la diversidad y el alcance de sus servicios orientados a clientes financieros y comerciales. Algunos de los servicios que presta son: procesamiento y transporte de valores terrestre, aéreo y marítimo con riesgos asociados, atención integral de cajeros automáticos, fabricación instalación y operación de cofres de seguridad, administración de bóvedas, entre otros.

En la actualidad, en concordancia con las múltiples tendencias de mejoramiento continuo, los cambios en los entornos organizacionales y las exigencias de los clientes y usuarios, toda organización que presta un servicio debe mantener como estrategia un enfoque orientado hacia el cliente, es por esto que las empresas buscan estar a la vanguardia y realizar cambios en operaciones, servicios y estrategias que les permitan mantener la armonía entre sus objetivos empresariales y las necesidades de cada uno de sus clientes.

El Centro de Gestión al Cliente (CGC) hace vida dentro de ésta organización; esta unidad representa el canal de atención, canalización y es responsable de dar respuesta a los requerimientos, quejas y reclamos de los clientes relacionados con los servicios contratados. Para realizar esto, cuenta con un sistema automatizado donde se registra toda interacción con el cliente. Es decir, por cada sugerencia, queja o reclamo que reporte un cliente, el Centro de Gestión al Cliente registra un ticket en este sistema y según sea el caso, este va asignando

actividades a las áreas resolutorias involucradas en su gestión, hasta su cierre. Sin embargo, contar con un sistema de atención donde solo se registren y canalicen los requerimientos de los clientes no es suficiente si no se maneja esta información, también es necesario suministrar una base de información para la continua revisión y mejora de los procesos internos a través del análisis de las quejas y reclamos como oportunidades de negocio y herramientas de gestión para la empresa.

El sistema de gestión de quejas y reclamos que maneja actualmente la empresa de nuestro caso de estudio dispone de mecanismos para el registro y canalización de las quejas y reclamos además de contabilizar los tiempos de respuesta de las áreas resolutorias. Sin embargo, por requerimientos de la región, se planteó la necesidad de ir a un esquema de indicadores gestión interna de procesos y de satisfacción del cliente basada en el manejo de esta información, de esta manera se podrá sustentar la toma de decisiones.

Bajo este escenario, el presente Trabajo Especial de Grado (T.E.G) propone el análisis, diseño e implementación de una Sistema de Inteligencia de Negocios (BI) para el apoyo en la toma de decisiones a la Gerencia Nacional de Post Venta; producto del análisis de la información suministrada por el Centro de Gestión al Cliente; haciendo uso de la metodología planteada por Kimball (BDL, por sus siglas en inglés) y las tecnologías de Pentaho BI Suite Community Edition, PostgreSQL y Community Dashboard Editor (CDE), tomando como base herramientas de Inteligencia de Negocio (BI) de Software Libre, ya que están enfocadas en la administración y creación de conocimiento mediante el análisis de datos existentes.

El presente Trabajo Especial de Grado (T.E.G) lo conforman cuatro (4) capítulos:

Capítulo I – Problema de Investigación: Se realiza una descripción del desarrollo de un sistema automatizado de inteligencia de negocios para la empresa de procesamiento y transporte de valores, justificación de la investigación y situación actual.

Capítulo II – Marco Conceptual: Se desarrollan todos los conceptos y bases teóricas que fundamentan el presente T.E.G. y que sustentan la solución planteada. Partiendo de las nociones básicas de Atención al Cliente, Gestión de Quejas y Reclamos, se describen Sistemas de Información, conceptos relacionados con Almacenes de Datos seguidos por los conceptos relacionados con Inteligencia de Negocios, particularmente la herramienta Pentaho BI Suite Community Edition finalizando con la Metodología de Desarrollo de Sistemas de Inteligencia de Negocios planteada por Kimball (2008).

Capítulo III – Marco Aplicativo: Se describen las actividades realizadas en las distintas fases de desarrollo de la solución: Planificación del Proyecto, Definición de Requerimientos del Negocio, Selección de Herramientas, Diseño de la Arquitectura, Modelo Dimensional, Diseño Físico, Diseño del proceso de Extracción, Transformación y Carga (ETL por sus siglas en ingles), Especificación de la Aplicación de Usuario Final, Implementación y Fase de pruebas.

Capítulo IV– Conclusiones y Recomendaciones: Se presentan las conclusiones y recomendaciones para próximos trabajos.

CAPITULO I- Problema de Investigación

PLANTEAMIENTO DEL PROBLEMA Y JUSTIFICACIÓN DE LA INVESTIGACIÓN

En la actualidad la mayoría de las organizaciones manejan altos flujos de información, gracias a esto uno de los aspectos fundamentales para el funcionamiento de cualquier empresa es justamente el manejo adecuado de ésta información. La manipulación de los datos se ha convertido para muchas empresas en una tarea titánica, es por esto que se han creado un conjunto de herramientas especializadas que pueden ayudar a las empresas a adquirir un mejor entendimiento del modelo de negocio con la intención de poder manipular la información de manera más sencilla y entender el porqué de su desempeño o mejor aún, plantear escenarios a futuro, lo que permite la toma de mejores decisiones.

A continuación se detalla la situación que justifica la realización del presente T.E.G. describiendo la situación actual que vive la organización de nuestro caso de estudio, líder en el procesamiento y transporte integral de valores y riesgos asociados, precisando la solución propuesta, objetivo general, objetivos específicos, alcance, importancia y justificación de la investigación.

SITUACIÓN ACTUAL

La Organización de nuestro caso de estudio tiene más de cinco (5) décadas de experiencia resguardando el producto del esfuerzo de todos los venezolanos con los mayores niveles de seguridad, profesionalismo, responsabilidad y eficiencia. Cuenta con oficinas a todo lo largo del territorio nacional garantizando cobertura total en todo el país, centros de logística de efectivo distribuidos en las regiones con mayor impacto económico permitiendo la mayor agilidad en el conteo, clasificación, depuración y preparación de remesas, así como su inmediato cargo en cuenta para el mayor rendimiento financiero del efectivo, centros integrales de procesamiento de monedas, además, cuenta con el soporte tecnológico más avanzado en procesos de seguridad, interconexión y encadenamiento nacional de bóvedas, circuitos de video digitales de grabación continua, para seguimiento y control de operaciones y verificación de clientes; posee la flota de camiones y aeronaves de mayor importancia en lo que se refiere a empresas transportistas de valores en Latinoamérica, y único con un Centro de Gestión al Cliente que garantiza atención oportuna y directa en la resolución de requerimientos, solicitudes, quejas y reclamos de sus clientes.

El modelo actual para la gestión de sugerencias, quejas y reclamos desarrollado en el Centro de Gestión al Cliente se basa en asegurar respuestas al cliente en la gestión sus incidencias, para

ello dispone de “Ejecutivos de Atención” quienes reciben el reporte del cliente vía telefónica o correo electrónico y a través de un sistema automatizado se encargan de registrar y realizar seguimiento a todos y cada uno de los casos que reporta el cliente. Cada requerimiento, queja o reclamo se registra en esta herramienta y genera un número de incidencia, adicionalmente esta herramienta es capaz de generar reportes en hojas de Excel donde muestra información sobre los casos registrados tal como: número de incidencia, cliente asociado, oficina donde se generó, fecha de creación y cierre del caso, si el cierre es procedente o no, etc.

Estos reportes contienen grandes volúmenes de información y manipularlos se hacía bastante complejo. Bajo este escenario, la Gerencia de Atención al Cliente no contaba con datos veraces y oportunos que sirvieran de soporte para la toma de decisiones en función de garantizar la satisfacción de sus clientes, optimización de sus procesos internos y mejora continua. Basándonos en el hecho de que la identificación de las necesidades del cliente marca la pauta para el cumplimiento de los objetivos institucionales para una organización y que la organización de nuestro caso de estudio se encontraba en la necesidad de ir a un esquema de indicadores gestión interna de procesos y de satisfacción del cliente; se planteó la implementación de un sistema que permitiese el manejo de la información proveniente de los requerimientos, quejas y reclamos de sus clientes proporcionada por el Centro de Gestión al Cliente que sirvieran de apoyo a nivel corporativo y gerencial para la toma de decisiones oportunas y orientadas a generar los informes y estadísticas para la gestión, a fin de que los reclamos, quejas y sugerencias sean conocidos por las áreas responsables de su origen y les sea posible identificar las acciones correctivas y/o preventivas a fin de garantizar la satisfacción del cliente y la mejora de sus procesos internos.

SOLUCIÓN PROPUESTA

El presente T.E.G. aborda la necesidad de desarrollar una solución de inteligencia de negocios que permita mejorar la manipulación de la información proveniente de las quejas y reclamos de los clientes a fin de facilitar la visualización de datos y generar indicadores de gestión que permitan a la Gerencia del Centro de Atención al Cliente y a la Gerencia Nacional de Servicio Post Venta la toma de decisiones oportunas para la mejora de los procesos internos de la organización.

Para ello partimos de modelado multidimensional orientado a responder cada uno de los requerimientos de las gerencias involucradas el cual se alimenta a partir de un proceso de ETL (Extracción, Transformación y Carga) de los datos provenientes de los archivos arrojados por el

sistema operacional existente hasta la base de datos multidimensional para posteriormente generar informes y lograr visualizarlos.

Partiendo de la necesidad de responder los requerimientos planteados mencionamos las herramientas necesarias para el desarrollo de la solución propuesta; destacando que son de Software Libre, aprovechando así las ventajas competitivas que nos ofrece, adicionalmente son requerimientos dados por la institución:

- Pentaho Data Integration (PDI)
- Pentaho Report Designer
- Pentaho Schema Workbench (PSW)
- Kettle
- Mondrian
- PostgreSQL
- PgAdmin

OBJETIVO DE LA INVESTIGACIÓN

El objetivo principal de esta investigación puede ser establecido como:

“Desarrollar e implementar una sistema de Inteligencia de Negocios para el apoyo a la toma de decisiones para facilitar la gestión de la información producto de los requerimientos, quejas y reclamos, manejada por el Centro de Gestión al Cliente en una organización que presta servicio de procesamiento, recogida y entrega de valores.”

OBJETIVOS ESPECÍFICOS

En función del objetivo principal se tienen los siguientes objetivos específicos:

- Analizar el modelo de negocios asociado.
- Definir los requerimientos de la solución planteada.
- Diseñar una arquitectura de software que responda a las necesidades planteadas y sean adaptables en la organización de nuestro caso de estudio.
- Diseñar el modelo multidimensional de la solución.
- Realizar el proceso de extracción, transformación y carga de datos (ETL) para la unificación de la data del proceso de gestión por resultados.
- Construir la solución de inteligencia de negocios para el estudio de indicadores y facilitación de creación de conocimiento a partir de del conocimiento de los mismos.
- Realizar las pruebas necesarias para validar el correcto funcionamiento de la gestión de resultados en la solución desarrollada.

ALCANCE

El sistema contempla:

- La definición los requerimientos, funcionales, como no funcionales del sistema, con la finalidad de comprender toda la información pertinente a cada uno de los procesos y requerimientos de la organización.
- Elaborar el modelo multidimensional, definir sus métricas y dimensiones asociadas.
- Realizar el proceso de ETL entre la base de datos del módulo operacional actual de la empresa y la base de datos de la solución propuesta de Inteligencia de Negocios.
- Diseñar e implementar la solución de Inteligencia de Negocios en Pentaho BI Suite, que permita la generación de indicadores y reportes necesarios.

El sistema no contempla:

- La modificación del código fuente de la herramienta Pentaho BI Suite.
- Funcionalidades que no estén contempladas en la fase de requerimientos del negocio.
- La modificación de la herramienta para el registro de los requerimientos, quejas y reclamos que se encuentra actualmente activa.

CAPÍTULO II - Marco Conceptual

Conceptos básicos

Queja: Manifestación de un usuario o grupo de usuarios con el fin de hacer constar una situación anómala derivada de un servicio o de un procedimiento inadecuados con el objetivo de que ésta sea subsanada por parte del servicio correspondiente.

Reclamo: Manifestación de un usuario o grupo de usuarios con el fin de hacer constar una situación irregular o ilegal derivada de un servicio o de un procedimiento inadecuado con el objetivo de que ésta sea subsanada legalmente. Es decir; un reclamo es igualmente una expresión de insatisfacción hecha a una organización con respecto a sus productos pero que pide o pretende algún tipo de compensación.

Sugerencia: Propuesta documentada de un usuario o grupo de usuarios cuyo objetivo es mejorar una situación, un servicio o un procedimiento.

El objetivo primordial, aunque no único de un Centro de Servicios es servir de punto de contacto entre los usuarios y la Gestión de Servicios. Un Centro de Servicios en su concepción más moderna, debe funcionar como centro neurálgico de todos los procesos de soporte al servicio que se está prestando:

- Registrando y monitorizando incidentes.
- Gestionando solicitudes, requerimientos o cambios solicitados por los clientes mediante peticiones de servicio

Adicionalmente, debe jugar un papel importante dando soporte al negocio identificando nuevas oportunidades en sus contactos con usuarios y clientes. Es esencial para el buen desarrollo del negocio que los clientes y usuarios perciban que están recibiendo una atención personalizada y ágil que les ayude a:

- Emitir peticiones de servicio.
- Resolver rápidamente las interrupciones del servicio.
- Informarse sobre el cumplimiento de los acuerdos de servicio y promesas de valor establecidos.
- Recibir información comercial en primera instancia.

Los principales beneficios de una correcta implementación del Centro de Servicios se resumen en:

- Reducción de costes mediante una eficiente asignación de recursos.
- Una mejor atención al cliente que repercute en un mayor grado de satisfacción y fidelización del mismo.
- Apertura de nuevas oportunidades de negocio.
- Centralización de procesos que mejoran la gestión de la información y la comunicación.
- Soporte al servicio proactivo.

En general, el concepto de “Atención al Cliente” está relacionado con el servicio que se presta y proporciona a las empresas de servicios comunicación directa y de primera línea con sus clientes en caso que estos necesiten manifestar reclamos, sugerencias, plantear inquietudes sobre el producto o servicio en cuestión, solicitar información adicional, solicitar servicio particular o simplemente plantear sus necesidades. Así, el éxito de una empresa estará estrechamente vinculado a la solución de demandas de sus clientes. Es por esto que, sin lugar a dudas, el éxito de una empresa dependerá fundamentalmente de que las demandas que provienen de sus clientes sean en principio recibidas, atendidas y satisfechas.

Servicio al Cliente

El Servicio al Cliente puede ser definido en un sentido amplio como la medida de actuación del sistema logístico para proporcionar en tiempo y lugar un producto o servicio. El concepto de servicio al cliente es a menudo confundido con el de satisfacción al cliente, que es un concepto más amplio ya que incluye elementos del marketing tales como: producto, precios, promoción y distribución.

Es importante incorporar también el concepto de Servicio Post Venta, este incluye todo el soporte que debe realizarse después de la venta. Tradicionalmente se ha tomado el Servicio Post Venta como un área poco importante, sin embargo en la actualidad juega un papel protagonista y crucial integrado a la cadena de suministros ya que, un servicio Post Venta de calidad y a tiempo atrae al cliente potencial y garantiza la fidelización de los clientes actuales; es por ello que no debe considerarse como un centro de costos sino como un generador de ingresos.

La satisfacción del cliente es el resultado de comparar su expectativa con la percepción del producto o servicio recibido, cuanto mejor sea la percepción del cliente y siempre y cuando esté acorde a sus expectativas, mayor será la satisfacción ante el servicio prestado. Esta situación pone de manifiesto la necesidad de mejora continua como única vía para satisfacer y superar las crecientes expectativas del cliente.

Sistemas de Información

Antes de entrar al estudio de los sistemas de información es necesario dejar claro lo que se entiende como información y la diferencia entre datos, información y conocimiento.

Aunque toda información procede de algún dato anterior no todos los datos son información. Es más, se puede tener muchos datos y mucha información y, sin embargo, no haber acumulado ningún conocimiento.

Por información nos referimos a los datos que se han modelado en una forma significativa y útil para los seres humanos. Por el contrario, los datos son flujos de elementos en bruto que representan los eventos que ocurren en las organizaciones o en el entorno físico antes de ordenarlos e interpretarlos en una forma que las personas puedan comprender y usar. (Laudon & Laudon, 2012)

Los datos no son más que elementos de tipo numérico, alfanumérico, imágenes o sonidos que permiten describir ciertos eventos, actividades o transacciones. Se tratan de elementos que pueden estar almacenados e incluso clasificados, pero no organizados para dar respuestas a cuestiones específicas o a significados concretos.

La información proviene de la ordenación de los datos de forma que éstos tengan un significado concreto e incluso un valor determinado. De esta forma, el destinatario puede analizar dicho significado y obtener conclusiones. Mientras que los datos en sí no son nada, sí lo es la información que se puede obtener de ellos. Las decisiones que se toman habitualmente en las empresas provienen de información recabada a partir de datos, previamente clasificados y presentados adecuadamente para facilitar su proceso. Un breve ejemplo en el que se comparen la información y los datos. Las cajas en los supermercados exploran millones de piezas de datos de los códigos de barras, que se encargan de describir cada uno de los productos disponibles. Se puede obtener un total de dichas piezas de datos y analizar para conseguir información relevante, como el número total de botellas de lavaplatos que se vendieron en una tienda específica, las marcas de lavaplatos que se venden con más rapidez en esa tienda o territorio de

ventas, o la cantidad total que se gastó en esa marca de lavaplatos en esa tienda o región de ventas. (Ilustración 1)

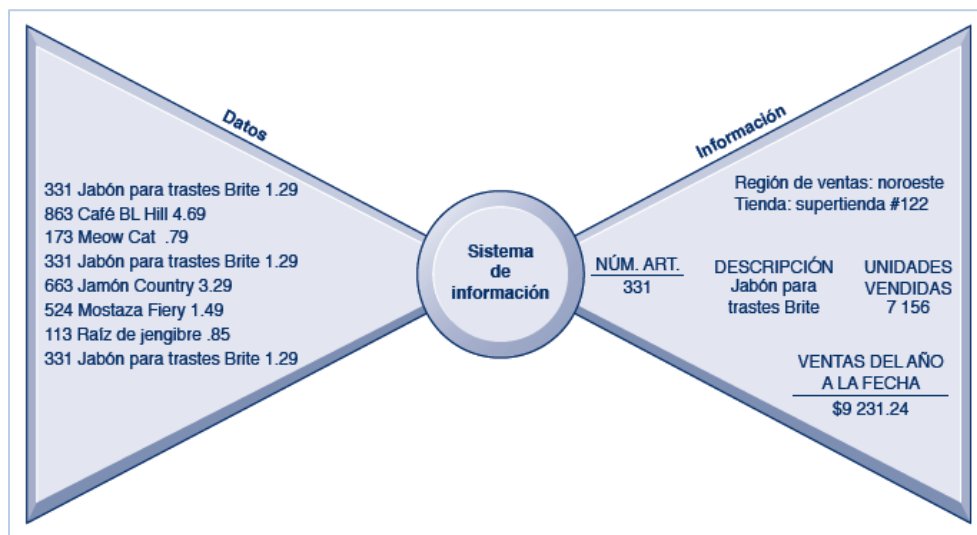


Ilustración1: Datos e Información (Laudon&Laudon, 2012)

Ahora se definirá lo que es un sistema; un Sistema posee un conjunto de componentes (elementos o subsistemas) que interaccionan (relaciones) para lograr un objetivo común dentro de un límite (frontera).

En conclusión, podemos plantear la definición técnica de un sistema de información como un conjunto de componentes interrelacionados que recolectan (o recuperan), procesan, almacenan y distribuyen información para apoyar los procesos de toma de decisiones y de control en una organización. Además de apoyar la toma de decisiones, la coordinación y el control, los sistemas de información también pueden ayudar a los gerentes y trabajadores del conocimiento a analizar problemas, visualizar temas complejos y crear nuevos productos. Los sistemas de información contienen información sobre personas, lugares y cosas importantes dentro de la organización, o en el entorno que la rodea. (Laudon & Laudon, 2012)

Sin embargo, estar informados no es suficiente para tomar decisiones acertadas. Cualquier decisión requiere también de la experiencia, historia profesional y personal de quien la toma, así como de sus propios elementos de juicio de evaluación de la información. La información evaluada por una persona dentro de un contexto se convierte en conocimiento. Los sistemas de información, vistos desde esta perspectiva, pueden mejorar el desarrollo de una base de conocimiento si se coloca, al alcance de los agentes que toman las decisiones, información relevante. De hecho, en entornos cambiantes una buena base de conocimientos constituye a

menudo una fuente fundamental de ventaja competitiva, y en estas situaciones los sistemas de información pueden resultar esenciales.

Funciones de un Sistema de Información

Hay tres actividades en un sistema de información que producen los datos necesarios para que las organizaciones tomen decisiones, controlen las operaciones, analicen problemas y creen nuevos productos o servicios. Estas actividades son: entrada, procesamiento y salida (*Ilustración 2*).

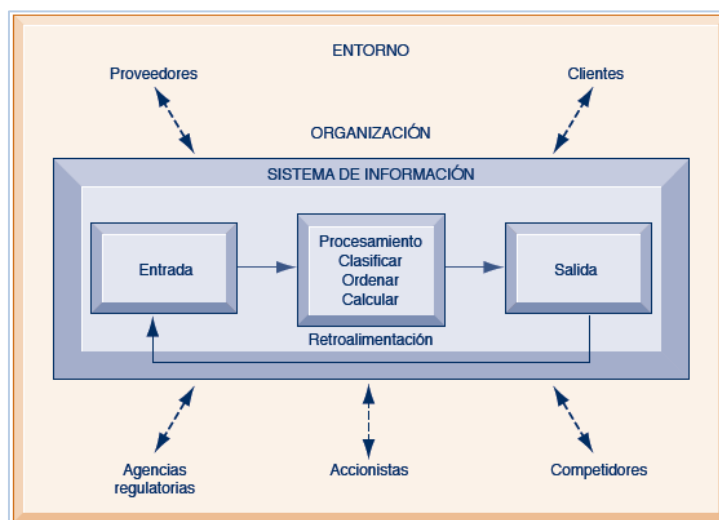


Ilustración 2: Funciones de un sistema de información (Laudon&Laudon, 2012)

- La entrada captura o recolecta los datos en crudo desde el interior de la organización o a través de su entorno externo.
- El procesamiento convierte esta entrada en bruto en un formato significativo.
- La salida transfiere la información procesada a las personas que harán uso de ella, o a las actividades para las que se utilizará.

Los sistemas de información también requieren retroalimentación: la salida que se devuelve a los miembros apropiados de la organización para ayudarles a evaluar o corregir la etapa de entrada.

Aunque los sistemas de información basados en computadora usan la tecnología computacional para procesar los datos en bruto y convertirlos en información significativa, hay

una clara distinción entre una computadora y un programa computacional por un lado, y un sistema de información por el otro. Las computadoras electrónicas y los programas de software relacionados son la base técnica, las herramientas y materiales, de los sistemas de información modernos. Las computadoras proveen el equipo para almacenar y procesar la información. Los programas de computadora, o software, son conjuntos de instrucciones de operación que dirigen y controlan el procesamiento de la máquina. Es importante saber cómo funcionan las computadoras y los programas computacionales para diseñar soluciones a los problemas organizacionales, sin embargo las computadoras sólo son parte de un sistema de información.

Tipos de Sistemas de Información

Las distintas clases de sistemas de información surgen de la satisfacción de diferentes necesidades. Si el sistema de información satisface los requerimientos del sistema-organización, las distintas clases de subsistemas de información habrán de responder a las necesidades de distintos subsistemas de la organización. (*Ilustración 3*)

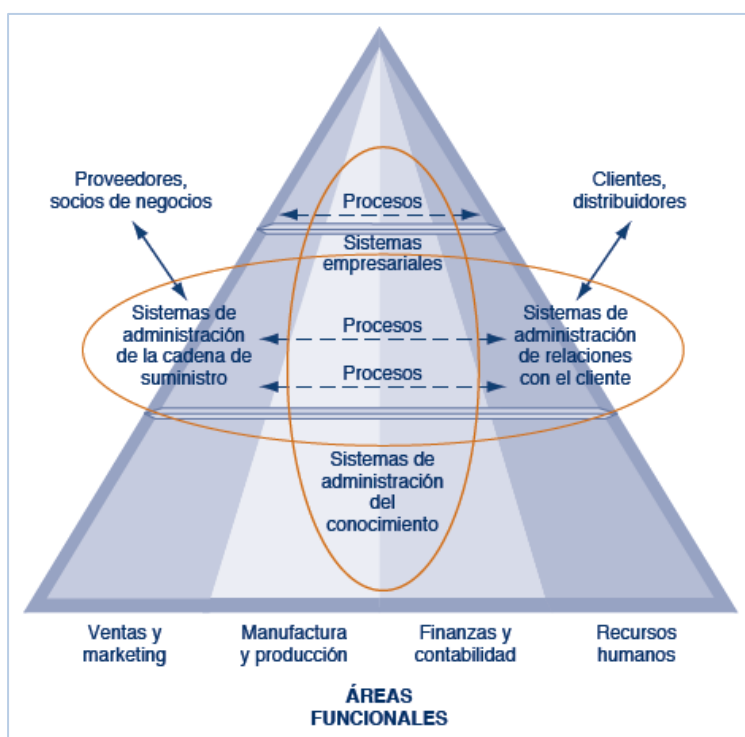


Ilustración 3: Tipos de sistemas de información (Laudon&Laudon, 2012)

A continuación se describen cada uno de los sistemas:

- **Sistemas empresariales:** Las empresas usan sistemas empresariales, también conocidos como sistemas de planificación de recursos empresariales (ERP), para integrar los procesos de negocios en manufactura y producción, finanzas y contabilidad, ventas y marketing, y recursos humanos en un solo sistema de software. La información que antes se fragmentaba en muchos sistemas distintos ahora se guarda en un solo almacén de datos exhaustivo, en donde se puede utilizar por muchas partes distintas de la empresa.
- **Sistemas de administración del conocimiento:** Algunas empresas funcionan mejor que otras debido a que tienen un mejor conocimiento en cuanto a cómo crear, producir y ofrecer productos y servicios. Este conocimiento empresarial es difícil de imitar, único y se puede aprovechar para obtener beneficios estratégicos a largo plazo. Los sistemas de administración del conocimiento (KMS) permiten a las organizaciones administrar mejor los procesos para capturar y aplicar el conocimiento y la experiencia. Estos sistemas recolectan todo el conocimiento y experiencia relevantes en la empresa, para hacerlos disponibles en cualquier parte y cada vez que se requieran para mejorar los procesos de negocios y las decisiones gerenciales. También enlazan a la empresa con fuentes externas de conocimiento.
- **Sistemas de administración de la cadena de suministro:** Las empresas usan sistemas de administración de la cadena de suministro (SCM) para ayudar a administrar las relaciones con sus proveedores. Estos sistemas ayudan a proveedores, empresas de compras, distribuidores y compañías de logística a compartir información sobre pedidos, producción, niveles de inventario, y entrega de productos y servicios, de modo que puedan surtir, producir y entregar bienes y servicios con eficiencia. El objetivo primordial es llevar la cantidad correcta de sus productos desde el origen hasta su punto de consumo en el menor tiempo posible y con el costo más bajo. Estos sistemas aumentan la rentabilidad de las empresas al reducir los costos de transportación y fabricación de los productos, y al permitir a los gerentes tomar mejores decisiones en cuanto a la forma de organizar y programar el suministro, la producción y la distribución. Los sistemas de administración de la cadena de suministro son un tipo de sistema inter-organizacional, debido a que automatizan el flujo de información a través de los límites organizacionales.
- **Sistemas de administración de relaciones con el cliente:** Las empresas usan sistemas de administración de relaciones con el cliente (CRM) para que les ayuden a administrar las relaciones con sus clientes. Los sistemas CRM proveen información para coordinar todos los procesos de negocios que tratan con los clientes en ventas, marketing y servicio para

optimizar los ingresos, la satisfacción de los clientes y la retención de éstos. Esta información ayuda a las empresas a identificar, atraer y retener los clientes más rentables; a proveer un mejor servicio a los consumidores existentes; y a incrementar las ventas.

Inteligencia de negocios

La inteligencia de negocios (BI por sus siglas en inglés – *Business Intelligence*), es un término genérico para describir el aprovechamiento de los recursos de información internos y externos de la organización para la toma de mejores decisiones de negocio. (Kimball R. y., 2002)

El concepto de BI incluye una amplia categoría de metodologías, aplicaciones y tecnologías que permiten reunir, acceder, transformar y analizar los datos, transacciones e información no estructurada (interna y externa), con el propósito de ayudar a los usuarios a tomar mejores decisiones de negocio. BI también puede intervenir en todos los procesos de una organización, actuando en las tareas y actividades de los empleados, creando nuevas actividades y nuevas habilidades, mejorando la comunicación entre departamentos e incrementando la capacidad de reacción de la compañía. (Boletín Digital, 2008)

De las definiciones anteriores se puede decir que BI no es un producto ni un sistema, *es una arquitectura y una colección de aplicaciones operacionales así como aplicaciones de soporte de decisiones y bases de datos que proveen a la comunidad de negocios fácil acceso a los datos de negocios*. También se puede decir que se entiende por BI al conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización.

Arquitectura y componentes de un BI

En la *Ilustración 4* se pueden apreciar los componentes de una arquitectura conceptual de un BI, la cual está formada por:

- **Sistemas transaccionales:** son las fuentes de datos, las cuales pueden ser de distintos tipos, tales como, archivos planos, hojas de cálculo, bases de datos, archivos XML, entre otros.

- Repositorios de información: en el cual se encuentran los datos transformados, representados en modelos multidimensionales, dimensiones y tablas de hechos.
- Herramientas de acceso a los datos: son las que tienen acceso a los datos y luego de manipularlos y ordenarlos proveen de información que soporte las necesidades gerenciales, entre ellas se encuentran los reportes, indicadores de gestión, análisis OLAP y minería de datos.

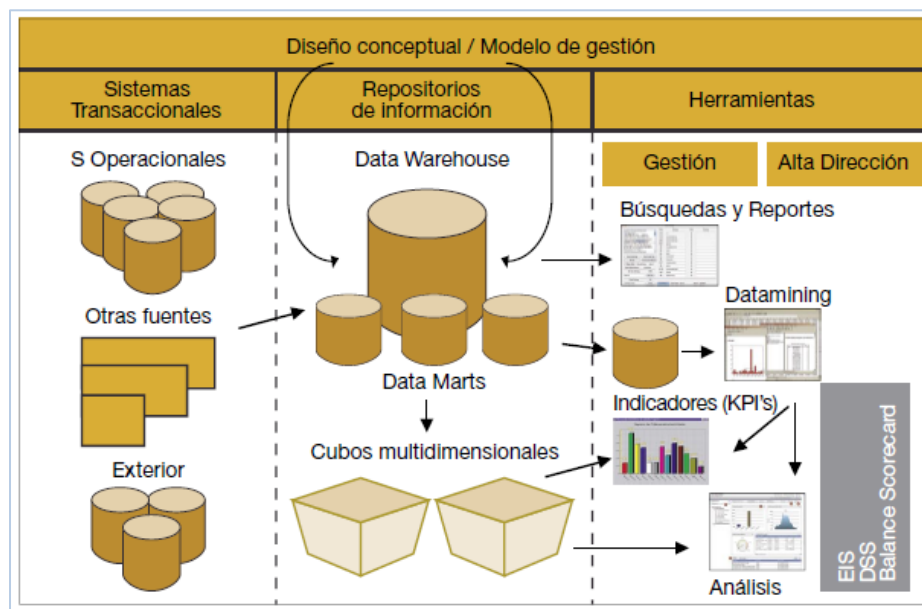


Ilustración 4: Arquitectura conceptual de un BI

A esta arquitectura se le suman unos componentes que permiten proveer una solución analítica (Ilustración 5).

- ETL, acrónimo de *Extraction, Transformation and Load*, proceso de extracción, transformación y carga, en el cual se toman los datos de las fuentes y se indica cuáles campos se van a utilizar, cuáles necesitan transformación y en dónde van a ser almacenados, esto viene apoyado de un proceso denominado mapeo o "*mapping*" el cual tiene toda la información de los datos que se encuentran en las fuentes.
- Base de datos analítica, RDBMS en la cual reside las estructuras de datos dimensionales, optimizada para acceso OLAP.
- Servidor OLAP, procesa consultas MDX, retorna resultados multidimensionales; puede emitir consultas SQL al RDBMS de base de datos analíticas; manejo de caché optimizado

y construcción de agregados para mejorar rendimiento; manejo de estados para sesiones de usuario final.

- Cliente OLAP, herramienta de usuario final que provee segmentación, navegación y pivoteo; genera MDX dinámicamente a partir de las peticiones de los usuarios; generalmente provee una metáfora de “hoja de cálculo” para manipulación de datos y tablas rudimentarias.

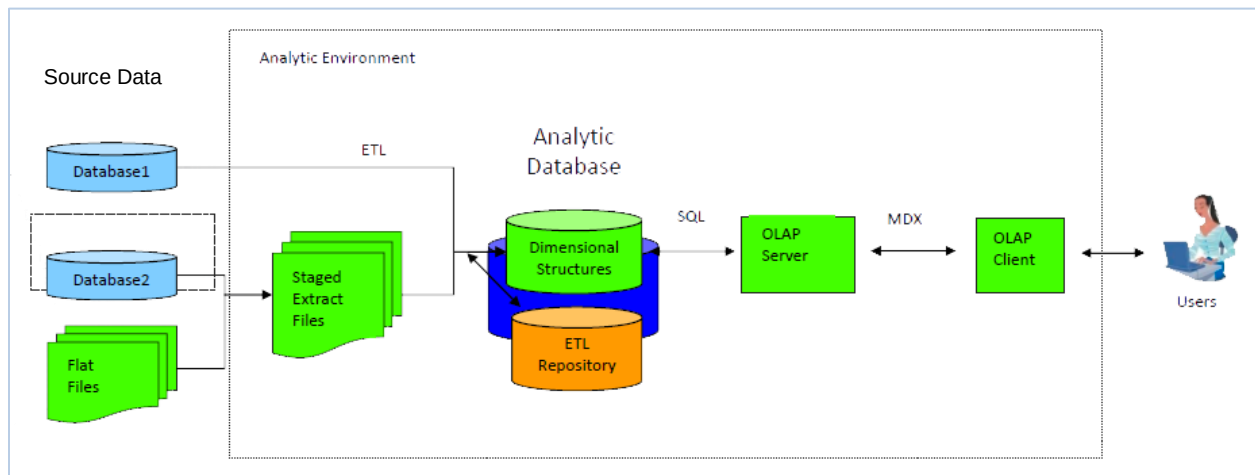


Ilustración 5: Arquitectura analítica de un BI

Beneficios de los BI

Según Espiñeira, Sheldon y Asociados – 2008, la implantación de estos sistemas de información proporciona diversos beneficios, entre los que se pueden destacar:

- Disposición de la información correcta en el momento adecuado para la toma de decisiones.
- Los datos se encuentran almacenados en un único lugar y se puede extraer de manera sencilla y en tiempo real.
- Permite agrupar información de distintas áreas en un solo cuadro.
- Permitir una visión única, conformada, histórica, persistente y de calidad de toda la información.
- Provee la capacidad de evaluar distintos escenarios al mismo tiempo.
- Crear, manejar y mantener métricas, indicadores claves de rendimiento e indicadores claves de metas fundamentales para la empresa.

- Aportar información actualizada tanto a nivel agregado como en detalle.

Además un BI adecuado puede llevar a:

- Incremento de la rentabilidad: BI puede ayudar a los clientes del negocio a evaluar el valor y la expectativa de rentabilidad a corto plazo de un cliente y usar este conocimiento para distinguir entre cliente rentable y no rentable.
- Disminuir costos: así sean logísticas de gerencia, costos operacionales o la disminución de la inversión solicitada para hacer ventas, BI puede ser utilizada para ayudar la evaluación de los costos organizacionales.
- Gestión mejorada de la relación con el cliente (CRM): esta es básicamente una aplicación BI que aplica el análisis de la información del cliente para proveer al mismo un mejor servicio, descubrir ventas cruzadas y aumentar las oportunidades de ventas y por sobre todo incrementar la lealtad del cliente.
- Disminución de riesgos: aplicar métodos BI a datos de créditos puede mejorar el análisis de riesgos de crédito, analizando las actividades y confiabilidad de proveedores y clientes puede proveer una visión en como agilizar la cadena de suministros.

Competencias de los BI

El MagicQuadrant de Gartner (2013) define la plataforma de BI y analítica, como una plataforma de software que ofrece 15 competencias a través de 3 categorías:

- Integración:
 - Infraestructura: todas las herramientas en la plataforma usan la misma seguridad, metadatos, administración, integración de portales, modelo de objetos y motor de consulta, y deben compartir el mismo look & feel.
 - Gestión de metadatos: las herramientas deben aprovechar los mismos metadatos, y las herramientas deben proveer una forma robusta de realizar las búsquedas, capturar, almacenar y publicar objetos de metadatos, tales como dimensiones, jerarquías, medidas, métricas de desempeño y objetos de diseño de informe.
 - Herramientas de desarrollo: la plataforma debe proveer un conjunto de herramientas programáticas y visuales, junto con un kit de desarrollo de software para crear aplicaciones analíticas, integrándolas al proceso de negocio, y/o incrustarlas en otra aplicación.

- Colaboración: permite a los usuarios compartir y discutir información y contenido analítico, y/o gestionar jerarquías y métricas a través de hilos de discusiones, chats y comentarios.
- Entrega de información
 - Informes: provee la capacidad de crear reportes con formatos e interactivos, con o sin parámetros, con capacidades de distribución y programación altamente escalables.
 - Dashboards: incluye la capacidad de publicar informes basados en web o móviles con pantallas intuitivas e interactivas que indican el estado de una métrica de desempeño comparada con una meta o un valor objetivo. Cada vez más los dashboards son utilizados para difundir datos de aplicaciones operacionales en tiempo real, o en conjunto con un motor de procesamiento complejo de eventos.
 - Consultas Ad-hoc: permite a los usuarios realizar sus propias preguntas a los datos, sin depender de un IT para crear un reporte.
 - Integración con Microsoft Office: algunas veces, Microsoft Office (particularmente Excel) actúa como el cliente de reportes o análisis. En estos casos, es vital que las herramientas ofrezcan esta integración, incluyendo soporte para formatos de documentos y presentaciones, fórmulas, refrescamiento de datos y tablas pivotes. Integraciones avanzadas incluyen bloqueo de celdas y write-back.
 - BI basado en búsqueda: aplica un índice de búsqueda a fuentes de datos estructuradas y no estructuradas y los mapea en una estructura de clasificación de dimensiones y medidas que los usuarios pueden navegar y explorar fácilmente utilizando una interfaz de búsqueda.
 - BI móvil: permite a la organización entregar contenido analítico a dispositivos móviles en un modo de publicación y/o interactivo, y se aprovecha del conocimiento de la ubicación del cliente móvil.
- Análisis
 - Procesamiento analítico en línea (OLAP): permite a los usuarios analizar los datos con consulta rápida y cálculo de rendimiento, lo que permite un estilo de análisis conocido como “slice and dice”. También tiene la capacidad de escribir de nuevo los valores en una base de datos propia para propósitos de planificación y “what if” (“qué pasaría si”).

- Visualización interactiva: ofrece a los usuarios la capacidad de mostrar numerosos aspectos de los datos de una manera más eficiente mediante el uso de imágenes y gráficos interactivos, en lugar de filas y columnas.
- Modelado predictivo y minería de datos: permite a la organización clasificar variables categóricas, y estimar variables continuas utilizando algoritmos matemáticos.
- Scorecards: estos toman las métricas que se muestran en un dashboard y aplican un mapa de estrategias que alinea los indicadores claves de rendimiento (KPIs) con un objetivo estratégico.
- Modelado prescriptivo, simulación y optimización: apoya la toma de decisiones al permitir a las organizaciones seleccionar el valor correcto de una variable basado en un conjunto de restricciones para procesos deterministas y modelando los resultados para procesos estocásticos.

Los 5 estilos de un BI en un escenario empresarial

De acuerdo a MicroStrategy existen 5 estilos de BI, los cuales se derivaron de la necesidad de las empresas de utilizar la información de nuevas maneras, como apoyo a la toma de decisiones, a los reportes de operación y a la optimización de procesos. Durante esta etapa de invención, los proveedores de tecnología de BI construyeron un software que satisficiera la evolución tecnológica del mercado y sus actuales necesidades. Estas aplicaciones se basan en los 5 estilos mencionados a continuación:

- Reportes corporativos o empresariales: este estilo se centra en la distribución de informes financieros u operacionales estandarizados; donde estos informes dan más información que un scorecard o un panel de control en un formato conciso que puede ser utilizado por todos los usuarios.
- Análisis de cubos: ideal para el análisis básico y predefinido por ejemplo análisis de ventas por región durante ciertos periodos de tiempo.
- Análisis y consultas Ad Hoc: se refiere a los análisis investigativos de los datos empresariales, incluso llegando al nivel de transacción.
- Análisis estadístico y minería de datos: es utilizado para descubrir relaciones sutiles, y/o generar proyecciones; tratamiento estadístico y otras funciones matemáticas avanzadas.

- Entrega de información y alertas: se basa en la distribución de reportes en forma cronológica o por eventos; es decir, la Información es distribuida a un largo número de usuarios internos o externos basado en calendarios, excepciones o por demanda.

Las organizaciones líderes han reconocido los beneficios de colocar información en manos de todos sus empleados sin importar el cargo o función. Por esto, es importante contar con una infraestructura de BI que sea capaz de proveer los cinco estilos a todos los usuarios en una organización, ofreciendo diferentes niveles de funcionalidad personalizados a las necesidades y características de cada usuario.

Una forma de ver a estos diferentes estilos de BI es colocarlos en un espacio bidimensional (*Ilustración 6*) donde el eje vertical representa la sofisticación y la interactividad del proceso analítico y el eje horizontal representa la escala o tamaño de la población de usuarios:

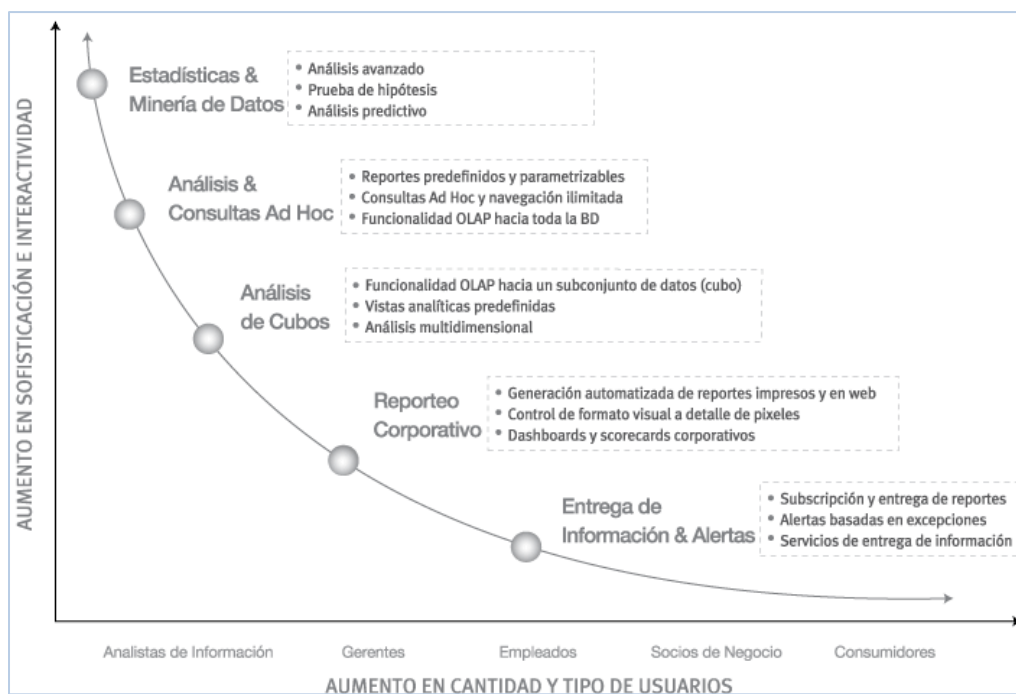


Ilustración 6: Estilos de BI y grupos de usuarios (MicroStrategy)

Como se puede observar el más interactivo y sofisticado estilo de BI es utilizado por un grupo de usuarios relativamente pequeño formados por analistas de información, para los cuales los datos y los análisis son sus trabajos principales. El estilo de BI menos interactivo entrega datos

básicos y resultados que son pertinentes a un gran grupo de usuarios, que abarcan todos los niveles de la organización.

Plataforma de Inteligencia de Negocios Pentaho BI Suite Community Edition

Pentaho se define a sí mismo en su página web como una plataforma de BI (Business Intelligence) “orientada a la solución” y “centrada en procesos” que incluye todos los principales componentes requeridos para implementar soluciones basadas en procesos. Esta plataforma provee el soporte y la infraestructura necesarios para crear soluciones de inteligencia empresarial, incluyendo un motor de flujo de trabajo integrado que se puede acoplar fácilmente a los procesos de negocios, asimismo, incluye un framework que proporciona los servicios básicos, incluidos los certificados, registro, auditoría, flujo de trabajo, servicios web y los motores de reglas.

La plataforma también incluye un motor de solución que integra reportes, análisis, tableros de mando (dashboards) y componentes de minería de datos para formar una plataforma de BI sofisticado y completo. Actualmente, existe una versión comercial y una versión de Software Libre desarrollada por la comunidad. 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 muy flexible al cubrir una alta gama de necesidades empresariales.

Características

- Lenguaje: Plataforma 100% J2EE, asegurando la escalabilidad, integración y portabilidad.
- Fuente de datos: Amplia conectividad a cualquier tipo de fuente de datos.
- Arquitectura: Abierta y basada en estándares, además todos los componentes están expuestos vía servicios web para facilitar la integración con Arquitecturas Orientadas a Servicios (SOA).
- Servidor: Puede correr en servidores compatibles con J2EE como JBOSS AS, WebSphere, Tomcat, WebLogic y Oracle AS.
- Base de datos: Es compatible con una gran variedad de fuentes de datos, a través de la conectividad a base de datos de java vía JDBC. Soporta base de datos propietarias y de Software Libre, tales como: IBM DB2, Microsoft SQL Server, MySQL, Oracle, PostgreSQL, NCR, Teradata, Firebird, EnterpriseDB, entre otras.
- Sistema operativo: Es multiplataforma y su lenguaje es interpretado.
- Lenguaje de programación: Java, Javascript, JSP, XSL (XSLT /XSL-FO).

- Interfaz de desarrollo: Java SWT, Eclipse, Web-based. Repositorio de datos basado en XML.
- Seguridad: Autenticación de usuario basada en roles y la autorización de la integración de sesión única y fácil con los sistemas de seguridad externos como
- LDAP, Active Directory y otros.
- Salida de datos: Salida en formatos populares como HTML, Excel, CSV, PDF y RTF.

Componentes

- Data Integration: Muchas organizaciones tienen información disponible en aplicaciones y base de datos separadas. Pentaho Data Integration extrae, limpia e integra esta valiosa información y la pone en manos del usuario, proveyendo consistencia, ya que centraliza una versión de todos los recursos de información, que es uno de los más grandes desafíos para las organizaciones de hoy en día. Pentaho Data Integration anteriormente conocida como Kettle incluye cinco herramientas las cuales son:
 - Spoon: Es una herramienta gráfica para diseñar ETL.
 - PAN: Ejecutar transformaciones diseñadas en el Spoon.
 - CHEF: Herramienta para ejecutar trabajos complejos que automatizan los procesos de actualización de la base datos.
 - Kitchen: Herramienta que ayuda a ejecutar los trabajos por lotes permitiendo iniciar y controlar fácilmente el proceso de ETL.
 - Carter: Es un servidor web que permite la supervisión remota el proceso de ETL.
- Analysis: Suministra a los usuarios un sistema avanzado de análisis de información, permitiéndole al mismo navegar por los datos, ajustando la visión de los datos, los filtros de visualización, añadiendo o quitando los campos de agregación. Permite el análisis a través de cubos OLAP y a través de lasherramientas Mondrian, Schema Workbench y Aggregation Designer.
- Reporting: Solución proporcionada por Pentaho e integrada en su suite para el desarrollo de informes. Dentro de dicha suite existen tres opciones con diferentes enfoques y dirigidos a diferentes tipos de usuarios.
- Pentaho Report Designer: Editor basado en eclipse con prestaciones profesionales de calidad, con capacidad de personalización de informes a las necesidades del negocio destinado a desarrolladores. Incluye Asistentes para facilitar la configuración de propiedades. Está estructurado de forma que los desarrolladores pueden acceder a sus

prestaciones de forma rápida: Incluye un editor de consultas para facilitar la confección de los datos.

- Pentaho Report Design Wizard: Herramienta de diseño de informes, que facilita el trabajo y permite a los usuarios obtener resultados de forma inmediata, está destinada a usuarios con un menor conocimiento técnico.
 - Web Ad-Hoc reporting: Es similar a la herramienta anterior pero vía web. Extiende la capacidad de los usuarios finales para la creación de informes a partir de plantillas reconfiguradas a través de un asistente de creación.
 - Dashboard (Tableros de mando): Incorporar una gran variedad de tipos de gráficos, tablas y velocímetros, para brindar a los usuarios la posibilidad de visualizar de forma atractiva información crítica para el proceso de negocio. También se puede integrar con los Portlets JSP para visualizar informes, gráficos y análisis OLAP. Para esto Pentaho se vale de la herramienta CDF (Community Dashboard Framework).
- Data Mining (Minería de datos): Es el proceso de correr datos en algoritmos sofisticados, para revelar significativos patrones y correlaciones que pueden estar escondidos. Esto puede ser usado para ayudar a entender lo mejor para el negocio y explotar el rendimiento de este en un futuro prediciendo completamente en el análisis. Se realiza con la herramienta Weka.

Bases de datos relacionales

Las bases de datos relacionales son una colección estructurada de datos relacionados entre sí con un propósito específico, su función es proveer una plataforma de almacenamiento eficiente y confiable. Es debido señalar que este tipo de bases de datos se encuentran optimizadas para manejar grandes volúmenes de transacciones; permitiendo a su vez realizar consultas tanto simples como estructuradas. El almacenamiento en estas, se lleva a cabo por medio de tablas y relaciones entre las mismas, donde por un lado las filas de la tabla representan registros; mientras que las columnas representan atributos o campos individuales.

Data Warehouse

Un almacén de datos es una colección de datos consistente semánticamente que sirve como implementación física de un modelo de datos de apoyo a la toma de decisiones y que almacena la información que la organización necesita para la toma de decisiones estratégicas. Además, en

muchas ocasiones, el almacén de datos es visto también como una arquitectura construida para integrar datos de múltiples fuentes heterogéneas y dar soporte a consultas estructuradas y/o ad hoc, informes analíticos y toma de decisiones. Un almacén de datos presenta una vista multidimensional de los datos y es por esto que se denominan bases de datos multidimensionales o cubos de datos. Un punto en un cubo de datos almacena una medida consolidada de los valores correspondientes a los valores de las dimensiones ya sean tablas o relaciones en un espacio multidimensional.

Un almacén de datos es útil, desde la perspectiva de un usuario por el hecho de que toda la información está en un lugar, y no es necesario andar por varias fuentes dispares de información o tratando de encontrar los archivos más antiguos en un desordenado sistema de correo electrónico o estructura de carpetas. Asimismo la información se encuentra más actualizada, los datos del almacén de datos son automáticamente cargados y actualizados en forma regular, lo que significa que nunca están fuera de la fecha o el estudio de la información antigua.

Se posee un acceso rápido al almacén de datos, encontrándose optimizado para la recuperación rápida de la información. El almacén de datos da respuesta a las preguntas mucho más rápido que tiendas locales archivo o archivos de correo electrónico sin límites de tamaño de cálculo sólo puede almacenar una cantidad limitada de datos y a menudo necesitan ser divididos en trozos para dar cabida a toda la información requerida. En cambio un almacén de datos puede almacenar una cantidad casi ilimitada de datos. En cuanto a la historia disponible; el almacén de datos no sólo contiene información actual, sino también los datos de última semana, el mes pasado, el año pasado, y varios años atrás también. Esto significa que cualquier análisis de tendencias o comparaciones con el tiempo se apoya en el almacén de datos. De hecho, si nunca son eliminados los datos del almacén de datos, a menudo estos contienen mucha más información histórica que los sistemas de origen.

Características de un Data Warehouse

De acuerdo a Inmon – 2002, las principales características de un data warehouse son las siguientes:

- Orientado hacia la información relevante de la organización: se diseña para consultar eficientemente la información relativa a las actividades básicas de la organización, como

lo son compras, producción e incluso ventas, y no para soportar los procesos que se realizan en la organización, como lo es la gestión de pedidos por ejemplo.

- Datos integrados: integra datos recolectados de diferentes sistemas operacionales de la organización y/o de fuentes externas, destacando que cuando se habla de fuentes externas se toman en cuenta diversos formatos desde txt hasta html por citar algunos.
- Variable en el tiempo: los datos son relativos a un periodo de tiempo y estos deben ser integrados periódicamente, es importante señalar que estos datos son almacenados como fotos que se corresponden a períodos de tiempo.
- No volátil: los datos que son almacenados no sufren de ninguna actualización solo son incrementados. Es importante destacar que el período de tiempo cubierto por un data warehouse va de 2 a 10 años aproximadamente.

Ventajas y desventajas de un Data Warehouse

Su principal ventaja es que presenta una rentabilidad de las inversiones realizadas para su creación, además de producir un aumento de la competitividad en el mercado y favorecer a la productividad de los técnicos de dirección.

Pero a pesar de sus considerables ventajas, también se encuentran sus puntos en contra como lo son, la infravaloración del esfuerzo necesario para su diseño y creación, además de los recursos necesarios para la captura, carga y almacenamiento de los datos, así mismo se debe tener en cuenta el aumento constante en los requerimientos de los usuarios y la privacidad de los datos.

Arquitectura de un Data Warehouse

Debido a que los almacenes de datos puede llegar a ser bastante grandes y complejos, es debido utilizar una arquitectura para alcanzar el éxito (*Ilustración 7*). En la cual se destacan las siguientes secciones:

- Uno o más sistemas fuente, como lo son archivos, DBMS, e incluso ERP.
- Un proceso de Extracción, Transformación y Carga de los datos. A menudo este proceso contiene un área de ensayo utilizada como área intermedia en donde se extraen los datos y se realiza la transformación de los datos iniciales y su limpieza. Los datos de

ensayo pueden ser usados tanto una base de datos como en ficheros planos. En muchos casos mediante archivos planos se permite un procesamiento más rápido.

- El almacén de datos, que consiste en la base de datos del almacén central y cero o más mercados de datos.
- La capa de usuario final identificada como EUL, con las diversas herramientas para trabajar con los datos, como lo son informes, cuadros de mando, hojas de cálculo y los documentos publicados. Generalmente, la combinación del almacén central y los mercados de datos es considerado como el almacén de datos, y el término de Data Warehouse es utilizado para denotar el proceso completo de construcción, carga y la gestión de los datos del almacén.

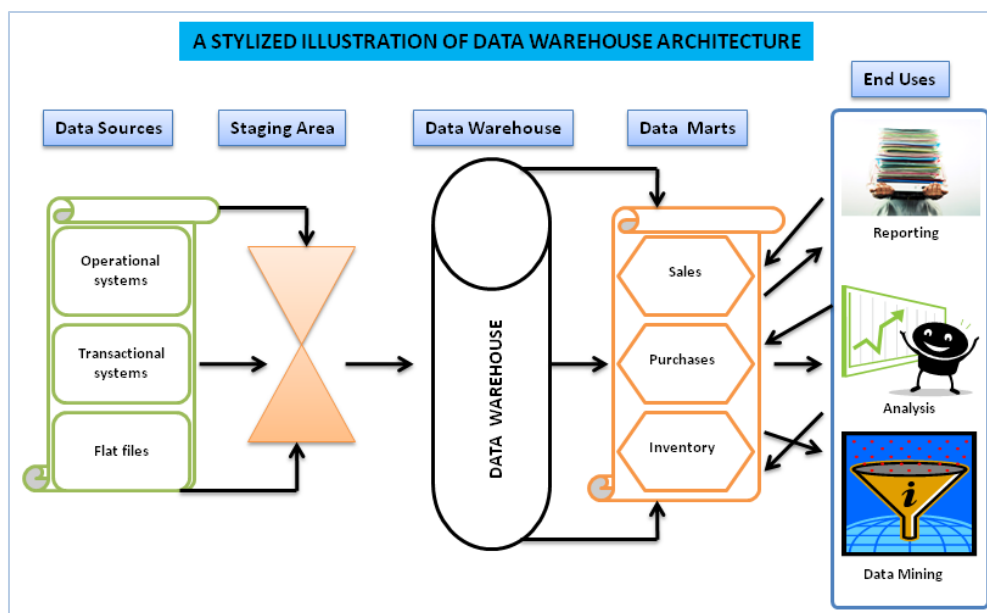


Ilustración 7: Arquitectura de un Data Warehouse

Conceptos Asociados a un Data Warehouse

Multidimensionalidad

Es convertir los datos de varias fuentes, tablas relacionales o archivos planos en una estructura donde los datos estén agrupados en dimensiones separadas y heterogéneas, que generalmente son llamadas cubos. Las dimensiones son perspectivas de alto nivel de los datos que representan la información más importante de un negocio. Destacando que en una aplicación OLAP estas dimensiones tienden a no cambiar en el tiempo, así como cada dimensión tiene

componentes los cuales son llamados miembros. Por otra parte el esquema multidimensional puede ser implementado por un esquema relacional; donde dos tipos de esquemas relacionales pueden implementar el esquema multidimensional, como lo son el esquema en estrella y el esquema copo de nieve.

Esquema en estrella

Un esquema estrella es un modo de representar datos multidimensionales en una base de datos relacional, donde las tablas de dimensión guardan información descriptiva acerca de sus miembros y sus relaciones, mientras que las tablas de hechos almacenan datos del negocio. Es debido destacar que los queries multidimensionales son hechos haciendo joins entre las tablas de hechos y las dimensiones, y que existen sistemas OLAP que utilizan este tipo de modelos de bases de datos y son llamados ROLAP.

Por otra parte una de las ventajas obvias de utilizar un esquema como este es su simplicidad y comprensión para los usuarios finales. Muy a menudo durante la fase de diseño de un almacén de datos, esquemas de estrella se utilizan para señalar la traducción inicial de las preguntas de negocio en diagramas de base de datos lógico.

Medidas

Una medida es un tipo de dato cuya información es usada por los analistas en sus queries para medir el rendimiento del comportamiento de un proceso o un objeto del negocio. Las medidas candidatas son los datos numéricos, pero no todo dato numérico es una medida candidata, de igual forma las medidas se encuentran involucradas con los cálculos de los resúmenes.

Dimensiones

Una dimensión es una entidad o una colección de entidades relacionadas, usadas por los analistas para identificar el contexto de las medidas con las que trabajan, estas determinan el contexto para las medidas. El número y tipo de dimensiones para cada medida del modelo debe ser determinada cuidadosamente, ya que el significado de una medida está influenciado por las definiciones de los tipos de medidas que tiene. Por otra parte cuando se trata de definir las dimensiones el añadir, eliminar o cambiar propiedades particulares de las dimensiones

candidatas cambia el contexto y en consecuencia el significado de la medida candidata. Por último es debido señalar que las dimensiones son referenciadas por las llamadas llaves de dimensión y que estas poseen entidades, atributos, jerarquías e incluso niveles de agregación.

Hechos

Un hecho es una colección de medidas relacionadas con sus dimensiones relacionadas, representadas por las llaves de dimensión. De igual forma un hecho puede representar un objeto de negocio, una transacción o un evento que es utilizado por el analista de información. Los hechos contienen un identificador de hecho, llaves de dimensión que lo enlazan con las demás dimensiones, medidas y varios tipos de atributos, que son los que usualmente se derivan de otros datos en el modelo. Por otra parte los hechos relacionados al negocio representan o bien objetos del negocio cuyo estado es de interés del analista de información, e incluso pueden representar transacciones o eventos del negocio.

Data Mart

Un Data Mart es una versión especial de un almacén de datos. Son subconjuntos de datos con el propósito de ayudar a que un área específica dentro del negocio pueda tomar mejores decisiones. Los datos existentes en este contexto pueden ser agrupados, explorados y propagados de múltiples formas para que diversos grupos de usuarios realicen la explotación de los mismos de la forma más conveniente según sus necesidades.

Esquema de copo de nieve

Un esquema en copo de nieve es una estructura algo más compleja que el esquema en estrella. Se da cuando alguna de las dimensiones se implementa con más de una tabla de datos. La finalidad es normalizar las tablas y así reducir el espacio de almacenamiento al eliminar la redundancia de datos; pero por otra parte este esquema genera peores rendimientos al tener que crear más tablas de dimensiones y más relaciones por medio de los joins entre las tablas lo que tiene un impacto directo sobre el rendimiento.

Cubo OLAP

Un cubo OLAP es 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.

OLAP

Codd describe el proceso OLAP, como el “proceso en que se muestran los datos empresariales en la perspectiva multidimensional, realizándose el análisis en línea de los mismos, utilizando fórmulas matemáticas y de análisis estadístico para consolidar y sumar los datos; es decir, OLAP es un enfoque para proveer rápidamente respuestas a consultas analíticas que son de naturaleza multidimensional” [Codd, E.F. et al (1993)].

Las herramientas OLAP permiten realizar análisis en línea de los datos desde una perspectiva multidimensional, donde las principales operaciones que pueden realizar sobre un cubo son:

- Drill-Down (profundización progresiva).
- Roll-Up (generalización progresiva).
- Filtering (filtro o selección)
- Pivoting (rotación).
- Slice (rebanar).
- Dice (cortar en cuadros).

Es importante señalar que existen características esenciales de OLAP descritas por Codd, las cuales han sido ampliadas obteniendo las siguientes:

- Vista conceptual multidimensional.
- Manipulación intuitiva de los datos.
- Accesibilidad.
- Extracción en lote Vs. Interpretativa.
- Modelo de análisis de OLAP.
- Arquitectura cliente/servidor.
- Transparencia.
- Soporte multiusuario.
- Tratamiento de datos no-normalizados.

- Almacenamiento de los resultados OLAP manteniéndolos separados de los datos originales.
- Extracción de los valores omitidos.
- Tratamiento de los valores omitidos.
- Flexibilidad de reportes.
- Desempeño uniforme de reportes.
- Ajuste automático del nivel físico.
- Dimensionalidad genérica.
- Niveles ilimitados de dimensiones y agregaciones.
- Operaciones a través de dimensiones sin restricciones.

En esencia, el procesamiento analítico en línea se puede ver como la síntesis, análisis y consolidación de grandes volúmenes de datos con un enfoque multidimensional.

Formatos de almacenamiento y MDX

Existen tres variaciones de bases de datos OLAP, cada una con un nombre basado en el formato de almacenamiento que se utiliza:

- MOLAP: (OLAP Multidimensional) Se refiere al formato original de OLAP en que los datos se almacenan en un formato multidimensional de propiedad. Todos los datos detallados y agregados se almacenan en el archivo de cubo.
- ROLAP: (Relational OLAP) En este caso, los datos y todos los agregados son almacenados en una base de datos relacional estándar. El motor ROLAP traduce consultas multidimensionales en SQL optimizados. Pentaho Mondrian es un ejemplo perfecto de un motor ROLAP.
- HOLAP: (Hybrid OLAP) En HOLAP, agregados y datos de navegación se almacenan en una estructura MOLAP pero los datos detallados se mantienen en base de datos relacional. Hasta la fecha, no hay solución de código abierto HOLAP disponibles.

Todas estas soluciones tienen las expresiones multidimensionales o consulta MDX, el idioma en común. MDX proporciona una sintaxis especializada para consultar datos que son almacenados en cubos OLAP, así como SQL hace para tablas relacionales. El lenguaje primero fue introducido por Microsoft en 1997 como parte de la OLE DB para la especificación OLAP, pero fue rápidamente adoptada por la mayoría de los vendedores de OLAP.

La razón para el desarrollo de un lenguaje especializado para el análisis de datos de almacenes de datos OLAP, es que SQL no se adapta muy bien al hacerlo. Aunque a través de los años, el lenguaje SQL se ha ampliado con características analíticas, aún MDX es el estándar de facto en el mundo OLAP. En el año 2000, el código XML for Analysis (XML / A) fue introducido, destacando que ahora es un estándar de facto para consultar bases de datos OLAP. XML / A se extiende el lenguaje MDX con la normalización de etiquetas XML para permitir ejecución de instrucciones de consulta a través de HTTP utilizando métodos SOAP.

El servidor OLAP Mondrian soporta MDX y XML / A, lo que hace una solución muy versátil para todo tipo de escenarios, pero ha de tenerse en cuenta que Mondrian no es ni un almacenamiento de datos, ni un instrumento de análisis, aunque de todos modos tendrá una interfaz para el análisis de datos

Dimensión, tablas y tablas de hechos

La distinción que hay entre la dimensión y tablas de hechos no siempre es clara, e incluso hay ocasiones en las que una tabla de dimensiones en el esquema de una estrella puede convertirse en una tabla de hechos de otro esquema estrella. Una explicación simple es que las tablas de dimensiones contienen información sobre las entidades de negocio como lo son clientes, productos, e incluso tiendas y las tablas de hechos sobre eventos de negocios como lo son ventas, envíos, y pedidos por citar algunos. La diferencia más notable se encuentra en las columnas medibles tales como ingresos, costos, y elementos, que forman parte de las tablas de hechos. Además, todos los diferentes ángulos y los atributos que se requieren para resumir estos hechos se almacenan en las tablas de dimensiones.

En realidad es bastante simple si se traduce una solicitud de informes típicos como "Muéstrame el total valor del pedido por mes por grupo de productos" en un modelo tridimensional, el elemento calculado de interés, es decir, la suma del orden de valor, es un hecho, el mes es un atributo tiempo, que pertenece a la dimensión tiempo, y el grupo de productos es un producto de atributo que pertenece a la dimensión del producto.

Una tabla de hechos por lo tanto sólo contiene las claves externas que apunta a las tablas de dimensión y atributos que se pueden agregar, los cuales son elementos cuantitativos. Por otra parte las tablas de dimensión contienen todos los atributos que describen una perspectiva de cierta organización, es decir, los elementos cualitativos.

Todas las filas de hecho se almacenan en el nivel más bajo de granularidad posible. Por granularidad, se entiende el grano de los datos, una tabla de hechos a nivel de la fecha tiene una granularidad más baja que una tabla de hechos a nivel de mes. La granularidad de la tabla de hechos está determinada por la combinación de las granularidades de las tablas de dimensiones. No todas las combinaciones tienen que estar presentes, por supuesto todas las medidas en la tabla de hechos se pueden agrupar por los elementos de las tablas de dimensiones.

Granularidad

Por granularidad, se enfoca la atención, en el nivel de detalle en el cual los datos se almacenan en los data warehouse. La regla de oro es almacenar los datos en el nivel más bajo de detalle posible. Uno de los conceptos erróneos acerca de los esquemas de estrella es que las tablas de hechos se debe siempre pre agregar. La agregación puede ser buena para aumentar el rendimiento de la consulta, pero sólo después de cargar los datos en el nivel más bajo de detalle.

Tipos de tablas de hechos

Las tablas de hechos no siempre se basan solo en transacciones. Las tablas de hecho transaccionales son el tipo más común, pero hay un par de variaciones que son necesarias tenerlas en cuenta.

El primero es la tabla de acumulación de instantáneas periódica; la única diferencia entre una tabla de hechos ordinarios y ésta son los campos de fecha atado a un proceso específico que necesitan ser actualizados cuando se producen nuevos eventos.

Un tipo muy diferente de la tabla de hechos, es la instantánea periódica, que son como fotos tomadas cada día, semana o mes. Instantáneas periódicas se utilizan cuando se necesita congelar los datos que sólo están indirectamente relacionados con las transacciones. Otra forma de ver las transacciones y las tablas de instantáneas es por el hecho de los tipos de medidas que contienen.

Las medidas en una tabla de transacciones regulares, como los ingresos por ventas, se pueden resumir en todas las dimensiones, incluyendo el tiempo, estos se conocen como aditiva de

hechos o medidas. Las medidas en las tablas de instantáneas periódicas no puede ser fácil de resumir, al menos no cuando los periodos de múltiples están involucrados. Existen medidas que se denominan semi-aditivo, lo que significa que pueden añadirlas en conjunto, pero es necesario el incluir un período de tiempo o de filtro para cortar los datos. Para finalizar el último tipo de medida es el hecho de no ser aditivos. Un buen ejemplo de este es la temperatura ambiente, no tiene sentido hacer un resumen de la temperatura de diferentes salas o períodos diferentes, aunque siempre se pueden calcular promedios.

Sistemas OLTP

Son los sistemas conocidos con el nombre de On-line Transactional Processing, estos procesan las transacciones de tiempo real de un negocio. Cabe destacar que poseen estructuras de datos optimizadas para la introducción y la adición de los datos.

Su principal desventaja es que proporciona capacidades muy limitadas para la toma de decisiones. Por otra parte es posible decir que estos sistemas definen el comportamiento operacional de un entorno operacional de gestión, como lo son: modificaciones, consultas, consultas rápidas y encuestas, así como poco volumen de información, las transacciones rápidas y los grandes volúmenes de concurrencias.

Diferencias entre OLTP y OLAP

Las siguientes diferencias es lo que provoca que sea imposible la convivencia en una misma base de datos de los entornos OLAP y OLTP:

- Pérdida del rendimiento del entorno OLTP.
- Falta de integración entre distintas aplicaciones OLTP.
- Tecnologías de bases de datos sin capacidad para soportar aplicaciones OLTP.
- Incorporación de datos externos difícilmente aplicables a las bases de datos OLTP.
- Distribución de los datos de manera no adecuada para OLTP.

Otras diferencias se pueden observar en la *Ilustración 8*.

	OLTP System Online Transaction Processing (Operational System)	OLAP System Online Analytical Processing (Data Warehouse)
Source of data	Operational data; OLTPs are the original source of the data.	Consolidation data; OLAP data comes from the various OLTP Databases
Purpose of data	To control and run fundamental business tasks	To help with planning, problem solving, and decision support
What the data	Reveals a snapshot of ongoing business processes	Multi-dimensional views of various kinds of business activities
Inserts and Updates	Short and fast inserts and updates initiated by end users	Periodic long-running batch jobs refresh the data
Queries	Relatively standardized and simple queries Returning relatively few records	Often complex queries involving aggregations
Processing Speed	Typically very fast	Depends on the amount of data involved; batch data refreshes and complex queries may take many hours; query speed can be improved by creating indexes
Space Requirements	Can be relatively small if historical data is archived	Larger due to the existence of aggregation structures and history data; requires more indexes than OLTP
Database Design	Highly normalized with many tables	Typically de-normalized with fewer tables; use of star and/or snowflake schemas
Backup and Recovery	Backup religiously; operational data is critical to run the business, data loss is likely to entail significant monetary loss and legal liability	Instead of regular backups, some environments may consider simply reloading the OLTP data as a recovery method

Ilustración 8: Diferencias entre OLTP y OLAP

Diferencias entre OLTP y Data Warehouse

Las principales diferencias existentes entre OLTP y un Data Warehouse son las siguientes:

- Las aplicaciones OLTP se encuentran organizadas para ejecutar las transacciones para las que fueron hechas; en cambio, el data warehouse se encuentra organizado en base a conceptos.
- El número de usuarios de un data warehouse en general es menor al número existente en un sistema OLTP.
- Los sistemas OLTP realizan cientos de transacciones por segundo, mientras que por el contrario una sola consulta en un data warehouse podría llevarse varios minutos.
- El diseño de un OLTP es considerablemente normalizado mientras que un data warehouse tiende a ser desnormalizado.
- Por lo general los OLTP son de menor tamaño que los data warehouse, debido a que los data warehouse están formados por la información de varios OLTP.
- El OLTP normalmente está formado por un número mayor de tablas, con pocas columnas; mientras que en un data warehouse el número de tablas es menor, pero cada una de estas tiende a ser mayor en número de columnas.

- Los OLTP son comúnmente actualizados por los sistemas operacionales del día, mientras que los data warehouse son actualizados en lote de manera periódica.
- Las estructuras de los OLTP son muy estables, rara vez cambian, mientras que por otra parte los data warehouse sufren cambios constantes derivados de su evolución.

Diseño de un data warehouse propuesto por Kimball

En la *Ilustración 9* se observan las actividades principales del ciclo de vida dimensional del negocio. En el diagrama se observa la secuencia de tareas, dependencias y concurrencias.

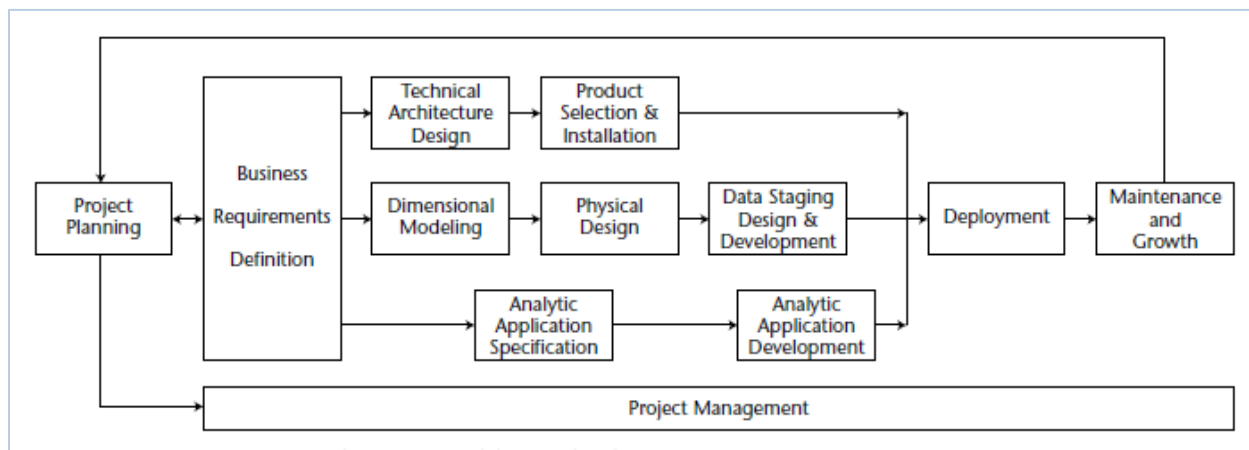


Ilustración 9: Diagrama ciclo de vida dimensional del negocio

Gestión y planificación del proyecto

Evaluar la preparación para un data warehouse

Antes de iniciar con el data warehouse, es prudente evaluar la disposición de la organización para proceder con el mismo. Existen 5 factores identificados que permiten diferenciar un proyecto de ejecución sin inconvenientes versus aquellos que son una lucha constante para lograr su ejecución.

El factor más crítico para lograr desarrollar un data warehouse exitosamente es tener un patrocinador fuerte. Los patrocinadores deben tener una visión del impacto potencial que tendrá un data warehouse en la organización. Deben ser apasionados y estar personalmente convencidos del valor del proyecto además de ser realistas.

El segundo factor es contar con una motivación convincente y fuerte por el negocio para construir un data warehouse. Este factor usualmente va de la mano con el patrocinio. Un proyecto de data warehouse no puede simplemente proveer una capacidad “buena de tener”, esta necesita resolver problemas críticos del negocio con el fin de reunir los recursos necesarios para un lanzamiento exitoso y una esperanza de vida sana.

El tercer factor al evaluar la preparación es la factibilidad. Hay varios aspectos de factibilidad, tales como factibilidad técnica o factibilidad de recursos, pero la factibilidad de datos es la más importante. La factibilidad de datos es la mayor preocupación debido a que no hay arreglos a corto plazo si no se recopilan los datos desde una fuente razonablemente limpia con la granularidad correcta.

El cuarto factor se enfoca en la relación entre el negocio y las organizaciones de TI. ¿La organización de TI entiende y respeta el negocio? ¿El negocio entiende y respeta la organización de TI? Si no se puede contestar afirmativamente a ambas preguntas no significa que no se pueda continuar sino que se debe velar porque tanto el negocio como la organización de TI estén alineados.

El quinto y último factor es la cultura analítica actual de la empresa. Si la empresa actualmente no valora la información y los análisis, su preparación para un warehouse es cuestionable. Los data warehouses más exitosos son los implementados en organizaciones en las cuales la toma de decisiones basadas en hechos se fomenta y se recompensa.

Definición del alcance inicial

Estos son algunos lineamientos generales para desarrollar el alcance general del data warehouse:

- La definición del alcance debe ser un esfuerzo conjunto de la organización de TI y de la gestión empresarial.
- El alcance inicial del proyecto del data warehouse debe ser significativo y manejable.
- Debe centrarse inicialmente en un solo requerimiento del negocio con el apoyo de una sola fuente de datos.
- Limitar el número de usuarios que inicialmente accederán al data warehouse.
- Comenzar a establecer criterios de éxito para el proyecto mientras el alcance se define.
- Una vez la organización de TI y la gerencia empresarial acuerden el alcance, este debe ser documentado para lograr una comprensión coherente del alcance preliminar proyecto.

Generar la justificación del negocio

Una vez establecido el alcance, se debe generar la justificación. Esto significa que se identifican costos anticipados y beneficios asociados al proyecto. Una forma de hacer esto es con el factor retorno de la inversión, que consiste en comparar el retorno financiero esperado, es decir, los beneficios del negocio contra la inversión esperada que representan los costos.

Establecer la identidad del proyecto

El proyecto del data warehouse debe tener un nombre. Algunas organizaciones crean logos los cuales incluyen en presentaciones, franelas, tazas, entre otros.

Dotar de personal al proyecto

El proyecto requiere de diferentes roles y habilidades de ambas partes (organización de TI y gestión empresarial), los cuales pueden ser cubiertos por una o varias personas dependiendo del tamaño, también puede darse el caso en el que cada miembro del equipo pueda desempeñar varios roles, esto viene a depender de los requerimientos y del tamaño del proyecto

Desarrollo del plan del proyecto

Esto incluye la identificación de las actividades necesarias para implementar el data warehouse, los recursos que las realizarán y el tiempo que tomará realizar cada una de ellas, las dependencias que existan entre actividades, de tal manera que se pueda dar seguimiento al progreso del proyecto.

Administración del proyecto

En este punto se discuten las técnicas a utilizar para darle seguimiento al proyecto, monitoreo de estatus, del alcance y estrategias de comunicación. Se establecen reuniones de personal en la cual se maneja una agenda y se abre el canal de comunicación entre todas las áreas que abarca el proyecto. El monitoreo se realiza periódicamente analizando el estatus del proyecto en diferentes estados de tiempo.

Definición de requerimientos

Para entender los requerimientos del negocio, se recomienda hablar con los usuarios del negocio ya que al final estos serán los que harán uso del data warehouse. Las preguntas deben estar orientadas hacia cómo es su trabajo, sus objetivos, sus retos e intentar descifrar cómo toman sus decisiones, tanto hoy como en el futuro. Se deben considerar las necesidades del negocio en conjunto con la disponibilidad de los datos que soportan el requerimiento. Se deben llevar a cabo reuniones de negocio y una vez que se capta un tema que sea constante en los usuarios se toma este punto y se le plantea a los responsables de los datos para conseguir los

detalles de los sistemas de origen, esto para entender si estos datos en realidad apoyan lo que los usuarios están solicitando y si la misma está completa y es confiable.

Estas son las actividades que se llevan a cabo para tener éxito en la definición de los requerimientos:

- Preparación para la entrevista: se debe definir el equipo que llevará a cabo las entrevistas y los roles y responsabilidades que tendrán.
- Llevar a cabo la investigación pre-entrevista: es importante conocer la empresa antes de empezar con los requerimientos, se recomienda leer el informe anual de la misma, también es importante conocer la estructura organizacional y la cadena de mando.
- Seleccionar los entrevistados: es ineficiente e imposible reunirse con todo el personal de la empresa, por lo cual se necesita realizar la selección de representantes de cada una de las áreas que serán afectadas por el desarrollo del data warehouse.
- Desarrollar cuestionarios: el preparador de las entrevistas debe desarrollar varios cuestionarios, ya que estos serán distintos dependiendo del entrevistado, en cuando a trabajo y nivel de funciones. El cuestionario debe ser de una sola página, para evitar exceso de tiempos en la entrevista.
- Agendar y preparar las entrevistas: es necesario contar con una herramienta que permita el seguimiento de las entrevistas en cuanto a fecha y hora de las mismas, también es importante realizar una reunión con los entrevistados seleccionados para presentarles el calendario de entrevistas.
- Terminar las entrevistas: luego de cada entrevista se debe estipular un tiempo para establecer la conclusión de la entrevista, en la cual se haga un resumen de los puntos que se han discutido en la misma.
- Preparar y publicar los requerimientos: es esencial tener documentada la información obtenida en las entrevistas, esto permite revisarlas y comentarlas, sin mencionar que toda documentación se vuelve material de referencia para el equipo de trabajo. Documentando cada una de las sesiones evitará que se pierda información crítica en caso de que un miembro del equipo decida irse, y ayudará a su reemplazo a ponerse al día con mayor facilidad.

Diseño de la arquitectura técnica

La arquitectura técnica es el plano de los servicios y elementos técnicos del data warehouse. Esta sirve como un marco organizativo para apoyar la integración de las tecnologías. La arquitectura técnica consiste en una serie de modelos que ahondan en mayores detalles con

respecto a cada uno de los principales componentes. La arquitectura identifica los componentes que son necesarios de inmediato y los que serán incorporados más adelante. La arquitectura sirve como una herramienta de comunicación con respecto a un conjunto de requerimientos técnicos entre el equipo, la gerencia y los proveedores.

Proceso de ocho (8) pasos para crear la arquitectura técnica:

1. Establecer el grupo de trabajo para la arquitectura.
2. Recoger los requerimientos relacionados con la arquitectura.
3. Documentar los requerimientos de la arquitectura.
4. Desarrollar un modelo de arquitectura de alto nivel.
5. Diseñar y especificar los subsistemas.
6. Determinar las fases de implementación de la arquitectura.
7. Documentar la arquitectura técnica.
8. Revisar y finalizar la arquitectura técnica.

Selección e instalación de productos

El esquema de la arquitectura es parecido a una lista de compras. Se seleccionan los productos que encajan con el esquema para proveer la funcionalidad necesaria.

Las tareas asociadas a la selección de productos incluyen las siguientes:

- Entender el proceso de compra corporativo.
- Desarrollar una matriz de evaluación de productos.
- Llevar a cabo investigaciones de mercado.
- Disminuir las opciones a una lista corta y ejecutar evaluaciones detalladas.
- Elaborar un prototipo, de ser necesario.
- Seleccionar el producto, instalar la versión de prueba y negociar.

Modelo dimensional

El modelado dimensional es una técnica de diseño lógico que busca presentar los datos en un marco estándar que sea intuitivo y permita acceso de alto rendimiento. Es inherentemente dimensional y se adhiere a una disciplina que utiliza el modelo relacional con importantes restricciones. Cada modelo dimensional es compuesto por una tabla con una clave multiparte, formada por claves foráneas, llamada tabla de hechos y un conjunto de tablas más pequeñas llamadas tablas dimensión. Cada tabla dimensión tiene una clave primaria que corresponde

exactamente con uno de los componentes (clave foránea) de la clave multipartes de la tabla de hechos. Esta estructura es denominada modelo estrella.

Para lograr el modelo dimensional se deben considerar cuatro (4) pasos en particular:

1. Seleccionar el proceso de negocio a modelar: un proceso es una actividad del negocio llevada a cabo en la organización que es típicamente apoyada por un sistema fuente de datos.
2. Establecer la granularidad del proceso de negocio: significa especificar exactamente que representa una fila individual de la tabla de hechos.
3. Escoger las dimensiones que aplican a cada línea de la tabla de hechos: las dimensiones contestan a la pregunta “¿Cómo las personas del negocio describen los datos resultantes del proceso de negocio?”.
4. Identificar los hechos numéricos que poblarán cada línea de la tabla de hechos: los hechos responden a la pregunta “¿qué se está midiendo?”.

Diseño físico

El diseño físico difiere del diseño lógico en término de detalles especificados para la base de datos física, incluyendo los nombres físicos de las columnas, tipos de datos, declaraciones de restricciones y permisibilidad de nulos. En este punto el diseño físico contiene actividades como ajustes de desempeño, particionamiento y diseño de archivo.

Diseño y desarrollo de datos o sistema ETL

Este proceso comprende varios aspectos determinantes para el data warehouse. Para garantizar un buen desarrollo se debe seguir un plan para poder explicar al resto del equipo la evolución de este proceso. La estructura para describir el plan de diseño de datos y su proceso de implementación consta de diez (10) pasos:

- Planificación
 1. Crear un plan de alto nivel.
 2. Probar, escoger e implementar una herramienta ETL.
 3. Profundizar en la tabla de destino, gráficamente dibujando cualquier reestructuración o transformación de datos complejos. Ilustrar gráficamente el proceso de generación de claves sustitutas. Desarrollar secuencia de trabajo preliminar.
- Carga de dimensiones

4. Construir y probar una carga de tabla dimensión simple. El objetivo principal de este paso es resolver cualquier detalle de infraestructura, incluyendo la conectividad, transferencia de archivos y problemas de seguridad.
5. Construir y probar el proceso de cambio de una dimensión.
6. Construir y probar las cargas de las demás dimensiones.
- Tabla de hechos y automatización
 7. Construir y probar las cargas de las tablas de hechos históricas, incluyendo búsqueda y sustitución de claves sustitutas.
 8. Construir y probar el proceso de carga incremental.
 9. Construir y probar las cargas de tablas agregadas y/o cargas MOLAP.
 10. Diseñar, construir y probar la automatización de la aplicación.

Especificación de la aplicación analítica

Utilizando como base la definición de requerimientos del negocio, se necesitan revisar los hallazgos e informes para identificar al menos un conjunto de 10 aplicaciones. Se requiere disminuir el foco inicial de las competencias críticas para manejar las expectativas y asegurar el tiempo de entrega. Se deben establecer los estándares para las aplicaciones (menús)

Desarrollo de la aplicación analítica

En este punto se abordan los estándares de nombramiento, cálculos, librerías y escritura de código, esto para evitar retrabajo en el futuro. Esta actividad inicia una vez el diseño de la base de datos está lista, las herramientas de acceso a los datos y los metadatos están instalados y una parte de los datos históricos han sido cargados.

Despliegue

Es la unión de tecnología, datos y aplicaciones en el escritorio del usuario, junto con la formación y estructura de soporte al usuario. Es ya la entrega del producto final.

Mantenimiento y crecimiento

En esta etapa del ciclo se continúa con la inversión de recursos en las áreas de: soporte, formación, soporte técnico y soporte de programas.

CAPÍTULO III – Marco Aplicativo

La metodología escogida para el desarrollo del Sistema de Inteligencia de Negocios expuesto en este T.E.G es “Ciclo de vida dimensional del negocio” de Ralph Kimball (2008), debido a que esta permite llevar el control de las actividades involucradas en el proceso de desarrollo de un proyecto. A continuación se muestra la aplicación de la metodología antes mencionada para el Sistema de Inteligencia de Negocios.

Planificación del Proyecto

Para llevar a cabo el Sistema Automatizado de Inteligencia de Negocios planteado, inicialmente se realizó la planificación de actividades, dada la naturaleza de la organización y sus niveles de seguridad se establecieron contactos presencial y también online con un personal del área de tecnología y se definió una persona representante de la Gerencia Nacional de Postventa.

En la Tabla 1 se muestra un resumen de cómo fueron llevadas a cabo las actividades para la ejecución del proyecto, especificándose las fechas planificadas y reales, al igual que el tiempo de duración de cada una de ellas indicado en días.

	Duración (días)	Fecha Inicio Planificada	Fecha Fin Planificada	Fecha Inicio Real	Fecha Fin Real	Duración Real
Definición del proyecto	5	01/04/15	07/04/15	01/04/15	03/04/15	3
Gestión y planificación del proyecto	20	08/04/15	05/05/15	06/04/15	24/04/15	15
Definición de requerimientos	20	06/05/15	02/06/15	27/04/15	22/05/15	20
Diseño de la arquitectura técnica	5	03/06/15	09/06/15	25/05/15	27/05/15	3
Aplicación de la arquitectura técnica	10	10/06/15	23/06/15	28/05/15	10/06/15	10
Selección de productos	3	24/06/15	26/06/15	11/06/15	15/06/15	3
Instalación de productos (por cada producto utilizado)	2	29/06/15	30/06/15	16/06/15	22/06/15	5
Diseño del modelo dimensional	5	01/07/15	07/07/15	23/06/15	29/06/15	5
Diseño físico	3	08/07/15	10/07/15	30/06/15	02/07/15	3
Implementación de Base de datos	5	13/07/15	17/07/15	03/07/15	09/07/15	5
Diseño de ETL	5	20/07/15	24/07/15	10/07/15	14/07/15	3
Desarrollo de ETL	25	27/07/15	28/08/15	15/07/15	18/08/15	25

Diseño de la aplicación de BI	5	31/08/15	04/09/15	19/08/15	25/08/15	5
Desarrollo de la aplicación de BI	30	07/09/15	16/10/15	26/08/15	07/10/15	31
Prueba de la aplicación de BI	20	19/10/15	13/11/15	08/10/15	04/11/15	20
Implementación del sistema	5	16/11/15	20/11/15	05/11/15	11/11/15	5
Despliegue frente al usuario	5	23/11/15	27/11/15	12/11/15	18/11/15	5

Tabla1: Planificación del proyecto

Definición de los requerimientos del negocio

En ésta etapa del proyecto se definieron entrevistas con la Gerencia de Atención al Cliente perteneciente a la Gerencia Nacional de Servicio Postventa. En ellas se identificaron los requerimientos reales del negocio y se ubicaron reportes producto de la herramienta para la Gestión de Quejas y Reclamos que manejan actualmente que servirían de insumo para el Sistema de Inteligencia de Negocios planteado. Se puntualizó la necesidad de contabilizar:

- Total de incidencias reportadas mensual / anual
- Cantidad de Incidencias reportadas por categoría

Ficha técnica del requerimiento 1:

Proceso	Incidencias		
Reporte Analítico	Reporte de Casos por Cliente		
Objetivo Estratégico	Obtener información de los casos creados por un cliente		
Observación			
Frecuencia de solicitud / análisis			
☐ Diario ☐ Semestral			
☐ Semanal ☒ Anual			
☒ Mensual ☐ Otro (especificación):			
☐ Trimestral			
Indicadores / Medidas a visualizar			
Indicador / Medida	Descripción	Unidad	
Total Casos	Total de los casos por cliente en un mes de un año	-	
Almacenamiento de la data	Mínimo un día de historia		
Usuarios del reporte	Analista del Departamento de Atención de Quejas y Reclamos		
Parámetros del reporte			
Nombre	Descripción	Filtro	Formato
Cliente	Nombre del cliente	Obligatorio	String

Año	-	Obligatorio	yyyy
Mes	-	Obligatorio	String

Tabla 2: Ficha técnica del requerimiento 1

Ficha de especificación del requerimiento 1:

Dato	Posición	Tamaño / Tipo de letra	Formato	Observación
Cabecera				
Logo de la empresa	Superior izquierda	-	-	-
Fecha	Superior derecha	Arial 9	daydd-mmmm-yyyy	Fecha de generación del reporte, Color: #000099
Barra roja	Centrado	-	-	Color: #FF0000
Título del reporte	Izquierda	Arial 18	-	Color: #000099
Cliente consultado	Izquierda	Arial 12	-	Color: #000000
Encabezado específico al reporte				
Tipo de requerimiento	Centrado	Arial 10	-	Color: #ffffff
Área	Centrado	Arial 10	-	Color: #ffffff
Producto	Centrado	Arial 10	-	Color: #ffffff
Requerimiento	Centrado	Arial 10	-	Color: #ffffff
Oficina	Centrado	Arial 10	-	Color: #ffffff
Sucursal	Centrado	Arial 10	-	Color: #ffffff
Estado	Centrado	Arial 10	-	Color: #ffffff
Fecha de apertura	Centrado	Arial 10	dd/mm/yyyy	Color: #ffffff
Pie de página del detalle				
Total casos creados	Alineado a la izquierda	Arial 12	-	Color: #ffffff
Pie de página del reporte				
Barra roja	Centrado	-	-	Color: #FF0000
Número de página	Inferior derecha	Arial 9	{0} / {1}	Color: #000099

Tabla 3: Ficha de especificación del requerimiento 1

Ficha técnica del requerimiento 2:

Proceso	Incidencias		
Reporte Analítico	Reporte de Casos por Tipo		
Objetivo Estratégico	Obtener información de los casos creados por tipo		
Observación			
Frecuencia de solicitud / análisis			
☐ Diario ☐ Semestral ☐ Semanal ☒ Anual ☒ Mensual ☐ Otro (especificación): ☐ Trimestral			
Indicadores / Medidas a visualizar			
Indicador / Medida	Descripción	Unidad	
Total Casos	Total de los casos por cliente en un mes de un año	-	
Almacenamiento de la data	Mínimo un día de historia		
Usuarios del reporte	Analista del Departamento de Atención de Quejas y Reclamos		
Parámetros del reporte			
Nombre	Descripción	Filtro	Formato
Tipo	Tipo de caso	Obligatorio	String
Año	-	Obligatorio	yyyy
Mes	-	Obligatorio	String

Tabla 4: Ficha técnica del requerimiento 2

Ficha de especificación del requerimiento 2:

Dato	Posición	Tamaño / Tipo de letra	Formato	Observación
Cabecera				
Logo de la empresa	Superior izquierda	-	-	-
Fecha	Superior derecha	Arial 9	daydd-mmmm-yyyy	Fecha de generación del reporte, Color: #000099
Barra roja	Centrado	-	-	Color: #FF0000
Título del reporte	Izquierda	Arial 18	-	Color: #000099
Tipo de caso consultado	Izquierda	Arial 12	-	Color: #000000
Encabezado específico al reporte				
Área	Centrado	Arial 10	-	Color: #ffffff

Producto	Centrado	Arial 10	-	Color: #ffffff
Requerimiento	Centrado	Arial 10	-	Color: #ffffff
Oficina	Centrado	Arial 10	-	Color: #ffffff
Sucursal	Centrado	Arial 10	-	Color: #ffffff
Estado	Centrado	Arial 10	-	Color: #ffffff
Fecha de apertura	Centrado	Arial 10	dd/mm/yyyy	Color: #ffffff
Pie de página del detalle				
Total casos	Alineado a la izquierda	Arial 12	-	Color: #ffffff
Pie de página del reporte				
Barra roja	Centrado	-	-	Color: #FF0000
Número de página	Inferior derecha	Arial 9	{0} / {1}	Color: #000099

Tabla 5: Ficha de especificación del requerimiento 2

Selección de herramientas

La organización de nuestro caso de estudio que requiere el Sistema Automatizado de Inteligencia de Negocios que aborda el presente T.E.G. decidió aprovechar las ventajas competitivas ofrece que el Software Libre, que el software propietario no puede ofrecer, entre las que están según PICÓ (2004):

- **Economía:** El bajo o nulo coste de los productos libres, permite proporcionar servicios a las organizaciones y ampliar sus infraestructuras sin que se vean mermados sus intentos de crecimiento, por no poder hacer frente al pago de grandes cantidades en licencias, pudiendo beneficiarse de aplicaciones de alta calidad a bajo coste.
- **Software adaptado:** El acceso al código fuente del programa proporciona la posibilidad de ajustar una aplicación a las necesidades concretas de cualquier institución.
- **Independencia del proveedor:** El Software Libre al no depender de un proveedor único, permite que cualquier organización pueda proporcionar servicios de soporte sobre una aplicación. De esta manera si un proveedor desaparece, siempre se podrá continuar mejorando dicho software.
- **Mejores prestaciones con el mismo hardware:** Por lo general los requisitos de procesamiento y memoria del Software Libre son menores que en las aplicaciones propietarias y optimizan los recursos del ordenador, lo que permite no tener que renovar el parque informático de una empresa cada pocos años o recuperar equipos informáticos obsoletos ya retirados, para realizar algunas acciones determinadas.
- **Aumento de la productividad:** El acceso al código fuente permite el desarrollo de nuevos productos sin la necesidad de desarrollar todo el proceso partiendo de cero.

Selección de la herramienta de Inteligencia de Negocios

En los últimos años el Open Source ha entrado en las herramientas de BI del mercado, por lo cual se evaluará el estado actual de las principales herramientas de BI Open Source del mercado con la intención de conseguir criterios de discriminación para el uso de las mismas, según Gartner las principales herramientas son JasperSoft y Pentaho. Los productos de información de JasperSoft se han convertido en la piedra angular de la misión de muchas soluciones críticas de aplicación en los principales segmentos de mercado tales como financieros, información al por menor y fabricación, proporcionando en tiempo real de entrega este servicio, para aplicaciones críticas tales como la auditoría y presentación de informes, los clientes de autoservicio, de gestión de conformidad y desempeño de los sistemas e incluso la gestión de sintonización. La misión y su plan de éxito se basan en hacer que el BI sea más accesible para los usuarios de negocios.

JasperSoft ofrece productos de código abierto que se ofrecen como aplicaciones independientes o combinados en una solución completa, de aquí se desprenden las siguientes herramientas:

- JasperReport, permite la presentación de informes, esta herramienta posee la habilidad de entregar contenido rico en la pantalla, impreso o en PDF, HTML, XLS, CSV y XML; pueden ser independientes o estar incorporados directamente en la aplicación de un usuario para darle capacidades avanzadas de informes.
- JasperServer, es un servidor independiente o de servicios web de información del motor; este reduce drásticamente el tiempo necesario para construir y desplegar aplicaciones de servidor que necesitan informes y análisis, que da acceso operativo a los tomadores de decisiones a los datos que anteriormente sólo tenían acceso a la alta gerencia.
- JasperAnalysis, proporciona análisis de datos; y se puede utilizar para explorar las tendencias, los patrones, las anomalías, y las correlaciones de los datos al permitir dinámicamente a los usuarios slice and dice, de pivote, el filtro, gráfico, drill-down, o un roll-up, cubo de datos en tiempo real.
- JasperETL, corresponde a la integración de herramientas para mover la información de múltiples bases de datos en un almacén de datos y el formato de los datos para que pueda ser analizada.

Por su parte Pentaho fabrica componentes para la comunidad Open Source y mejora los componentes desarrollados por otros. Pentaho integra componentes en bloques coherentes y flexibles que los desarrolladores de Java pueden utilizar para montar rápidamente soluciones

personalizadas, así como la completa plataforma de BI para usuarios finales. Además de prestar apoyo técnico integral, administración de versiones, control de calidad, y manejo de los servicios empresariales. En su concepción Pentaho gestiona, facilita, apoya y toma el papel principal en el desarrollo del proyecto. Pentaho BI es una iniciativa pionera por el desarrollo de la comunidad del código abierto al proporcionar a las organizaciones un conjunto completo de capacidades de BI que les permitan mejorar radicalmente el rendimiento del negocio, la eficiencia y eficacia. Pentaho ofrece una gama de productos que cubre las áreas de aplicación de Informes, Análisis, cuadros de mando, minería de datos e integración de dato entre los que se destacan:

- Pentaho Reporting: es la solución proporcionada por Pentaho e integrada en su suite para el desarrollo de informes. Originariamente el proyecto se llamaba JFreeReports hasta la adquisición por parte de Pentaho. Pentaho permite a las organizaciones acceder fácilmente, formato, y distribuir información a los empleados, clientes y socios. Ofrece un despliegue flexible de presentación de informes de escritorio independiente, para la presentación de informes interactivos basados en Web para la empresa inteligencia de negocios.
- Pentaho Analysis: es una herramienta poderosa que ayuda a los usuarios para operar con la máxima eficacia mediante la obtención de los conocimientos y la comprensión que necesitan para hacer un uso óptimo de decisiones, lo que les permite explorar la información de negocios por arrastrar, soltar, y perforar, así como los datos de tabulación cruzada. Ofrece un amplio análisis, capacidades que incluye una tabla dinámica de espectadores JPivot, posee avanzados gráficos utilizando SVG o Flash, de igual manera cuenta con reproductores integrados, tableros de instrumentos, la integración de minería de datos, integración de portales, y la integración del flujo de trabajo.
- Pentaho Dashboards: ofrece una perspectiva inmediata individual, departamental, o del rendimiento empresarial. Ofrece amplias capacidades de gestión de métricas que permiten la definición y seguimiento de indicadores críticos en cada nivel de organización, posee una adecuada visualización, la cual, ayuda a los usuarios ver de inmediato que métricas de negocios van por buen camino y los alerta que necesitan más atención. Es importante señalar la integración con Pentaho Reporting y Analysis lo que le permite a los usuarios ver los detalles subyacentes a informes y análisis para entender qué factores están contribuyendo al bien o al mal rendimiento.
- Pentaho Data Mining: incorpora Weka, una colección de algoritmos de aprendizaje automático, inclusión de la agrupación, la segmentación, árboles de decisión, los

bosques al azar, y el análisis de componentes principales. Estos algoritmos se combinan con tecnologías OLAP para proporcionar análisis de los datos. Las herramientas de Data Mining pueden analizar los datos históricos para crear modelos predictivos y luego usando Pentaho Informes y Análisis de las empresas, pueden distribuir esta información a las personas adecuadas.

- Pentaho Data Integration: los datos están disponibles en cualquier forma posible y en todas partes. El suministro de datos consistente a través de todas las fuentes de información es uno de los mayores desafíos que enfrentan las organizaciones de TI hoy en día. Para superar este problema Pentaho ha adoptado el proyecto de código abierto Kettle y hoy en día ofrece una poderosa Extracción, Transformación y Carga de los datos, así mismo Pentaho Data Integration, proporciona una forma fácil de usar gráficos, e incluso el medio ambiente de arrastrar y soltar.

La Plataforma Pentaho BI provee el soporte y la infraestructura necesarios para la construcción de soluciones a los problemas específicos de BI. La plataforma incluye un incrustado flujo de trabajo del motor y se puede integrar fácilmente en los procesos empresariales. Núcleo de servicios, incluyendo la autenticación, registro, auditoría, flujo de trabajo, servicios web, y motores de reglas son ofrecidas en dicha plataforma. El marco también incluye un motor de solución que integra los informes anteriores presentados, análisis, dashboards y componentes de minería de datos para formar una plataforma de BI sofisticada y completa.

Funcionalidades	Jaspersoft	Pentaho
Integración de la infraestructura	Si	Si
Reportes	Si	Si
Dashboard	No	Si
Reportes y consultas Ad-hoc	Limitado	Si
Servicios de hojas de cálculo	No	Si
Servidor OLAP	Si	Si
Interfaz OLAP	Si	Si
Data Mining	No	Si
ETL	Si	Si
Alertas	Si	Si
Repositorio	Si	Si
Seguridad	Si	Si
Programación	Si	Si

Tabla 6: Cuadro comparativo de productos BI Open Source

De acuerdo a lo expuesto anteriormente, se selecciona Pentaho, debido a que es una suite basada en elementos estándares que reúne todo los aspectos básicos que debe contar una herramienta de BI, se destaca por poseer una capacidad de personalización, pudiendo adaptarse mejor a los detalles solicitados en los formatos de los informes, los cubos OLAP y los dashboard.

Selección del Sistema Manejador de Base de Datos

SMDB	Características
PostgreSQL	• Licencia Open source completamente, lo que permite crear derivados comerciales propietarios sin pagar regalías. • Enfocado en la fiabilidad, integridad de datos y características integradas enfocadas al desarrollador. • Tiene un planificador de consultas capaz de unir cantidades relativamente grandes de tablas eficientemente. • Multiplataforma. • Copias de seguridad en caliente. • Acceso encriptado vía SSL. • Posibilidad de definir nuevos tipos de datos. • Ilimitado número de índices por tabla.
MYSQL	• Licencia Open source Copyleft, que obliga a los vendedores de software libre a liberar su código o adquirir una licencia propietaria de una sola entidad comercial y dueña de derechos de autor. • Enfocado en aplicaciones web de lectura mayormente. • Su principal preocupación es la optimización de consultas sencillas. • Multiplataforma. • Utiliza un diseño de servidor multicapas con módulos independientes. • Provee almacenamiento transaccional y no transaccional. • Soporta un máximo de 64 índices por tabla.

Tabla 7: Comparación entre SMDB's

Se selecciona como SMDB a PostgreSQL, ya que el mismo además de ser multiplataforma, está diseñado para ambientes de alto volumen tanto de transacciones como de usuarios, al tiempo que posee una alta escalabilidad y además es de licencia Software Libre.

Diseño técnico de la arquitectura

Modelo dimensional

La propuesta estuvo basada en una arquitectura de desarrollo basada en tres capas: cliente, servidor y base de datos, donde toda la lógica se encuentra en la aplicación servidor y los datos son almacenados en la base de datos.

Análisis de requerimientos dimensionales

Gerencia de Atención al Cliente

Se desea analizar las incidencias que cumplan con lo siguiente:

- Hayan sido creados en el tiempo seleccionado.
- Hayan sido creados por los clientes seleccionados.
- Los resultados sean medidos en: Año y Mes.
- Sean del tipo de requerimiento seleccionado.

Dimensiones

Las dimensiones que conforman el modelo de datos dimensional son las contenidas en la Tabla 8.

#	Dimensión
1	Área
2	Cliente
3	Estado
4	Fecha
5	Oficina
6	Producto
7	Tipos de requerimiento
8	Requerimiento
9	Sucursal
10	Traza

Tabla 8: Dimensiones

Dimensión Área

Descripción: Contiene la información de las áreas en las cuales se pueden reportar incidentes.

Jerarquías

Nivel	Atributos
Área	Nombre de área

Tabla 9: Jerarquía Dimensión Área

Atributos

Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por defecto
Id_area	Identificador del área		Ninguno
Nombre_area	Nombre del área		Ninguno

Tabla 10: Atributos Dimensión Área

Dimensión Cliente

Descripción: Contiene la información de los clientes que reportan incidentes.

Jerarquías

Nivel	Atributos
Cliente	Nombre del cliente

Tabla 11: Jerarquía Dimensión Cliente

Atributos

Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por defecto
Id_cliente	Identificador del cliente		Ninguno
Nombre_cliente	Nombre del cliente		Ninguno

Tabla 12: Atributos Dimensión Cliente

Dimensión Estado

Descripción: Contiene los estados en los cuales se puede encontrar un incidente.

Jerarquías

Nivel	Atributos
Estado	Nombre del estado

Tabla 13: Jerarquía Dimensión Estado

Atributos

Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por defecto
Id_estado	Identificador del estado		Ninguno
Nombre_estado	Nombre del estado		Ninguno

Tabla 14: Atributos Dimensión Estado

Dimensión Fecha

Descripción: Contiene las fechas en las cuales se aperturan y cierran los incidentes.

Jerarquías

Nivel	Atributos
Día	Día
Mes	Mes
Año	Año

Tabla 15: Jerarquía Dimensión Fecha

Atributos

Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por defecto
Día	Día del período		Ninguno
Mes	Mes del período		Ninguno
Año	Año del período		Ninguno

Tabla 16: Atributos Dimensión Fecha

Dimensión Oficina

Descripción: Contiene un listado de las oficinas de la empresa.

Jerarquías

Nivel	Atributos
Oficina	Nombre de la oficina

Tabla 17: Jerarquía Dimensión Oficina

Atributos

Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por defecto
Id_oficina	Identificador de la oficina		Ninguno
Nombre_oficina	Nombre de la oficina		Ninguno

Tabla 18: Atributos Dimensión Oficina

Dimensión Producto

Descripción: Contiene los productos sobre los cuales se pueden generar incidentes.

Jerarquías

Nivel	Atributos
Producto	Nombre del producto

Tabla 19: Jerarquía Dimensión Producto

Atributos

Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por defecto
Id_producto	Identificador del producto		Ninguno
Nombre_producto	Nombre del producto		Ninguno

Tabla 20: Atributos Dimensión Producto

Dimensión Tipo de requerimiento

Descripción: Contiene los tipos de requerimiento de los incidentes.

Jerarquías

Nivel	Atributos
Tipo de requerimiento	Nombre del tipo de requerimiento

Tabla 21: Jerarquía Dimensión Tipo de requerimiento

Atributos

Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por defecto
Id_req_tipo	Identificador del tipo de requerimiento		Ninguno
Nombre_req_tipo	Nombre del tipo de requerimiento		Ninguno

Tabla 22: Atributos Dimensión Tipo de requerimiento

Dimensión Requerimiento

Descripción: Contiene un listado de los requerimientos.

Jerarquías

Nivel	Atributos
Requerimiento	Nombre del requerimiento

Tabla 23: Jerarquía Dimensión Requerimiento

Atributos

Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por defecto
Id_req	Identificador del requerimiento		Ninguno
Nombre_req	Nombre del requerimiento		Ninguno

Tabla 24: Atributos Dimensión Requerimiento

Dimensión Sucursal

Descripción: Contiene un listado de las sucursales de la empresa.

Jerarquías

Nivel	Atributos
Sucursal	Nombre de la sucursal

Tabla 25: Jerarquía Dimensión Sucursal

Atributos

Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por defecto
Id_sucursal	Identificador de la sucursal		Ninguno
Nombre_sucursal	Nombre de la sucursal		Ninguno

Tabla 26: Atributos Dimensión Sucursal

Dimensión Traza

Descripción: Contiene el listado de las causas por las cuales el incidente ha sido creado o cerrado.

Jerarquías

Nivel	Atributos
Traza	Nombre de la traza

Tabla 27: Jerarquía Dimensión Traza

Atributos

Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por defecto
Id_traza	Identificador de la traza		Ninguno
Nombre_traza	Nombre de la traza		Ninguno

Tabla 28: Atributos Dimensión Traza

Tabla de hechos

La tabla de hechos que conforma el modelo dimensional es la siguiente (Tabla 29)

#	Tema	Hecho
1	Control de incidencias	fact_incidente

Tabla 29: Tabla de hechos

Dimensiones vs. Hechos

Dimensiones	Hechos
	Fact_incidente
Área	X
Cliente	X
Estado	X
Fecha	X
Oficina	X
Producto	X
Tipos de requerimiento	X
Requerimiento	X
Sucursal	X
Traza	x

Tabla 30: Dimensiones vs. Hechos

Fact incidente

Descripción: Esta tabla permite obtener el número de incidentes creados, de acuerdo a filtros que se definan en las consultas

Granularidad

N°	Nombre de la Dimensión	Descripción	Llave Primaria
1	Área	Es el listado de áreas sobre las cuales se crean los incidentes	Si

N°	Nombre de la Dimensión	Descripción	Llave Primaria
2	Cliente	Es el listado de clientes que han reportado los incidentes	Si
3	Estado	Es el listado estados por los que pasa un incidente, en los cuales se van a evaluar las cantidades	Si
4	Fecha	Período de evaluación de incidentes	Si
5	Oficina	Es el listado de las oficinas en las que se reportan incidentes	Si
6	Producto	Es el listado de productos sobre los cuales se reportan incidentes	Si
7	Tipos de requerimiento	Es el listado del tipo de requerimiento que puede tener un incidente	Si
8	Requerimiento	Es el listado del tipo de requerimiento que puede tener un incidente	Si
9	Sucursal	Es el listado de las oficinas en las que se reportan incidentes	Si
10	Traza	Es el listado de trazas que dan valor agregado a un incidente	Si

Tabla 31: Granularidad fact_incidente

Medidas

N°	Nombre	Descripción
1	Casos cliente	Cantidad de casos por cliente
2	Casos tipo req	Cantidad de casos por tipo de requerimiento
3	Casos área	Cantidad de casos por área
4	Casos producto	Cantidad de casos por producto
5	Casos requerimiento	Cantidad de casos por requerimiento
6	Casos oficina	Cantidad de casos por oficina
7	Casos sucursal	Cantidad de casos por sucursal
8	Casos creados	Cantidad de casos creados
9	Casos estado	Cantidad de casos por estado
10	Casos cerrados	Cantidad de casos cerrados

Tabla 32: Medidas fact_incidente

Diseño físico

Para seleccionar la estructura necesaria para dar soporte al diseño lógico de los datos se tomó en cuenta la definición de estándares del entorno de base de datos. A continuación se describe el diseño físico para las dimensiones y tablas de hechos en el modelado dimensional. Posteriormente se muestran las relaciones entre tablas de hechos y dimensiones.

Dimensión: Dim_Area

Nombre de la tabla		
dim_area		
Descripción		
Tabla que contiene todas las áreas		
Columna	Tipo	Descripción
area_id	integer	Identificador único de la dimensión área
area_nombre	charactervarying	Nombre del área

Tabla 33: Diseño físico de la dimensión Área

Dimensión: Dim_Cliente

Nombre de la tabla		
dim_cliente		
Descripción		
Tabla que contiene todos los clientes		
Columna	Tipo	Descripción
cliente_id	integer	Identificador único de la dimensión cliente
cliente_nombre	charactervarying	Nombre del cliente

Tabla 34: Diseño físico de la dimensión Cliente

Dimensión: Dim_Estado

Nombre de la tabla		
dim_estado		
Descripción		
Tabla que contiene todos los estados		
Columna	Tipo	Descripción
estado_id	integer	Identificador único de la dimensión estado
estado_nombre	charactervarying	Nombre del estado

Tabla 35: Diseño físico de la dimensión Estado

Dimensión: Dim_Fecha

Nombre de la tabla		
dim_fecha		
Descripción		
Tabla que contiene toda la información de la fecha		
Columna	Tipo	Descripción
fecha_id	integer	Identificador único de la dimensión

		fecha
fecha_jahr	integer	Número del año
fecha_dia	integer	Número del día
fecha_mes	integer	Número del mes
fecha_mes_nom	charactervarying	Nombre del mes
fecha_nombre	charactervarying	Fecha completa en formato dd/mm/yy

Tabla 36: Diseño físico de la dimensión Fecha

Dimensión: Dim_Oficina

Nombre de la tabla		
dim_oficina		
Descripción		
Tabla que contiene todos las oficinas		
Columna	Tipo	Descripción
oficina_id	integer	Identificador único de la dimensión oficina
oficina_nombre	charactervarying	Nombre de la oficina

Tabla 37: Diseño físico de la dimensión Oficina

Dimensión: Dim_Producto

Nombre de la tabla		
dim_producto		
Descripción		
Tabla que contiene todos los productos		
Columna	Tipo	Descripción
producto_id	integer	Identificador único de la dimensión producto
producto_nombre	charactervarying	Nombre del producto

Tabla 38: Diseño físico de la dimensión Producto

Dimensión: Dim_Req_Tipo

Nombre de la tabla		
dim_req_tipo		
Descripción		
Tabla que contiene todos los tipos de requerimiento		
Columna	Tipo	Descripción
req_tipo_id	integer	Identificador único de la dimensión tipo de requerimiento
req_tipo_nombre	charactervarying	Nombre del tipo de requerimiento

Tabla 39: Diseño físico de la dimensión Tipo de requerimiento

Dimensión: Dim_Requerimiento

Nombre de la tabla		
dim_requerimiento		
Descripción		
Tabla que contiene todos los requerimientos		
Columna	Tipo	Descripción
requerimiento_id	integer	Identificador único de la dimensión requerimiento
requerimiento_nombre	charactervarying	Nombre del requerimiento

Tabla 40: Diseño físico de la dimensión Requerimiento

Dimensión: Dim_Sucursal

Nombre de la tabla		
dim_sucursal		
Descripción		
Tabla que contiene todos las sucursales		
Columna	Tipo	Descripción
sucursal_id	integer	Identificador único de la dimensión sucursal
sucursal_nombre	charactervarying	Nombre de la sucursal

Tabla 41: Diseño físico de la dimensión Sucursal

Dimensión: Dim_Traza

Nombre de la tabla		
dim_traza		
Descripción		
Tabla que contiene todos las trazas		
Columna	Tipo	Descripción
traza_id	integer	Identificador único de la dimensión traza
traza_nombre	charactervarying	Nombre de la traza

Tabla 42: Diseño físico de la dimensión Traza

Fact_incidente

Nombre de la tabla		
fact_incidente		
Descripción		
Contiene la información de los casos creados		

Columna	Tipo	Descripción
incidente_id	Integer	Número de caso
id_cliente	Integer	Identificador de la dimensión cliente
id_req_tipo	Integer	Identificador de la dimensión tipo de requerimiento
id_area	Integer	Identificador de la dimensión área
id_producto	Integer	Identificador de la dimensión producto
id_requerimiento	Integer	Identificador de la dimensión requerimiento
id_oficina	Integer	Identificador de la dimensión oficina
id_sucursal	Integer	Identificador de la dimensión sucursal
id_estado	Integer	Identificador de la dimensión estado
id_fecha_apertura	Integer	Identificador de la dimensión fecha
id_fecha_cierre	Integer	Identificador de la dimensión fecha
id_traza	Integer	Identificador de la dimensión traza

Tabla 43: Diseño físico de la tabla de hechos Incidente

Relaciones de la tabla de hechos

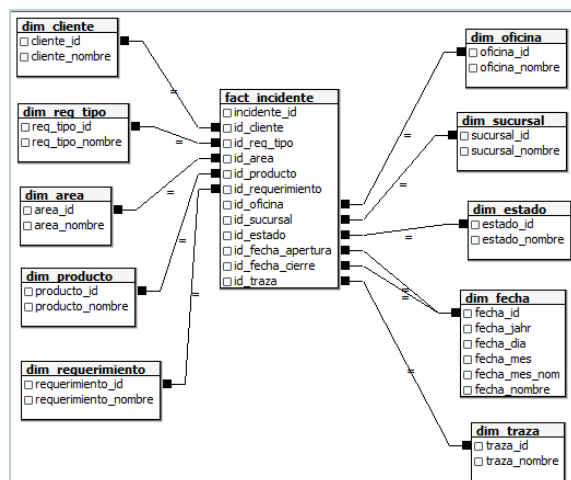


Ilustración10: Relaciones de la tabla de hechos fact_incidente

Diseño de procesos ETL

Para llevar a cabo el proceso de Extracción, Transformación y Carga (ETL por sus siglas en inglés) se realizaron una serie de transformaciones y trabajos (transformations, jobs) utilizando la herramienta *Spoon de Pentaho*. Esta herramienta permite almacenar transformaciones en archivos locales. A continuación se muestran las transformaciones y trabajos realizados para la implementación del Sistema de Inteligencia de Negocios propuesto.

Extracción:

Para el proceso de extracción tenemos que la fuente son las hojas de cálculo de Excel generadas por los usuarios, las cuales son alojadas en un servidor y se accede a las mismas vía FTP.

	A	B	C	D	E	F	G	H
1								
2		Resumen de Incidencias						
3								
4		Org Cliente	Categoria	Oficina	Fecha Apertura	Estado	Fecha Cierre	Traza Act
5		DIA DIA SUPERMERCADOS, C.A.	Reclamos. TV. Traslado. Ausencia en Entrega y Recogida	SERPAPRO - El Tuy	18/11/2013	Cerrado - Procedente	20/11/2013	Caso Cerrado
6		DIA DIA SUPERMERCADOS, C.A.	Reclamos. TV. Traslado. Ausencia en Entrega y Recogida	SERPAPRO - Los Ruices	18/11/2013	Cerrado - Procedente	20/11/2013	Caso Cerrado
7		DIA DIA SUPERMERCADOS, C.A.	Reclamos. TV. Traslado. Ausencia en Entrega y Recogida	BLINPASA - Maracay	18/11/2013	Cerrado - Procedente	20/11/2013	Caso Cerrado
8		DIA DIA SUPERMERCADOS, C.A.	Reclamos. TV. Traslado. Ausencia en Entrega y Recogida	SERPAPRO - El Tuy	18/11/2013	Cerrado - Procedente	20/11/2013	Caso Cerrado
9		DIA DIA SUPERMERCADOS, C.A.	Reclamos. TV. Traslado. Ausencia en Entrega y Recogida	SERPAPRO - Tranex	18/11/2013	Cerrado - Procedente	20/11/2013	Caso Cerrado
10		DIA DIA SUPERMERCADOS, C.A.	Reclamos. TV. Traslado. Ausencia en Entrega y Recogida	BLINPASA - Valledela P.	18/11/2013	Cerrado - Procedente	20/11/2013	Caso Cerrado

Ilustración11: Ejemplo de la hoja de cálculo

Transformaciones y trabajos implementados:

Para realizar la estandarización de datos y limpieza de los mismos se realiza la lectura de las columnas de la hoja de cálculo, se transforman los datos y se llenan las tablas correspondientes. A continuación se indica el mapeo entre las columnas del archivo y las tablas destino.

Columna Org Cliente

Nombre	Llave	Tipo	Formato	Limpieza	Valor por defecto
cliente_id	PK	integer	Numérico		000000
cliente_nombre		character	Texto (Mayúscula)	No debe ser nulo	No tiene

Tabla 44: Definición de tabla dimensión dim_cliente

Columna Categoria

En este caso se mapean cuatro (4) tablas dimensión.

Nombre	Llave	Tipo	Formato	Limpieza	Valor por defecto
req_tipo_id	PK	Integer	Numérico		000000
req_tipo_nombre		character	Texto	No debe ser nulo	No tiene

Tabla 45: Definición de tabla dimensión dim_req_tipo

Nombre	Llave	Tipo	Formato	Limpieza	Valor por defecto
area_id	PK	Integer	Numérico		000000
area_nombre		character	Texto	No debe ser nulo	No tiene

Tabla 46: Definición de tabla dimensión dim_area

Nombre	Llave	Tipo	Formato	Limpieza	Valor por defecto
producto_id	PK	Integer	Numérico		000000
producto_nombre		character	Texto	No debe ser nulo	No tiene

Tabla 47: Definición de tabla dimensión dim_producto

Nombre	Llave	Tipo	Formato	Limpieza	Valor por defecto
requerimiento_id	PK	Integer	Numérico		000000
requerimiento_nombre		character	Texto	No debe ser nulo	No tiene

Tabla 48: Definición de tabla dimensión dim_requerimiento

Columna Oficina

En este caso se mapean dos (2) tablas dimensión.

Nombre	Llave	Tipo	Formato	Limpieza	Valor por defecto
oficina_id	PK	Integer	Numérico		000000
oficina_nombre		character	Texto	No debe ser nulo	No tiene

Tabla 49: Definición de tabla dimensión dim_oficina

Nombre	Llave	Tipo	Formato	Limpieza	Valor por defecto
sucursal_id	PK	Integer	Numérico		000000
sucursal_nombre		character	Texto	No debe ser nulo	No tiene

Tabla 50: Definición de tabla dimensión dim_sucursal

Columnas Fecha Apertura y Fecha Cierre

Estas columnas son mapeadas a una sola dimensión.

Nombre	Llave	Tipo	Formato	Limpieza	Valor por defecto
fecha_id	PK	Integer	Numérico	No debe ser nulo	000000
fecha_jahr		Integer	Numérico		No tiene
fecha_dia		Integer	Numérico		No tiene
fecha_mes		Integer	Numérico		No tiene
fecha_mes_nom		character varying(11)	Texto		No tiene

Nombre	Llave	Tipo	Formato	Limpieza	Valor por defecto
fecha_nombre		character varying(10)	Texto		No tiene

Tabla 51: Definición de tabla dimensión dim_fecha

Columna Estado

Nombre	Llave	Tipo	Formato	Limpieza	Valor por defecto
estado_id	PK	Integer	Numérico		000000
estado_nombre		character	Texto	No debe ser nulo	No tiene

Tabla 52: Definición de tabla dimensión dim_estado

Columna Traza Act

Nombre	Llave	Tipo	Formato	Limpieza	Valor por defecto
traza_id	PK	Integer	Numérico		000000
traza_nombre		character	Texto	No debe ser nulo	No tiene

Tabla 53: Definición de tabla dimensión dim_traza

Tablas destino

Tabla:	dim_cliente		
Campo	Tipo	Mapeo	
cliente_id	Integer	nextval('todos_inc':regclass)	
cliente_nombre	Carácter	Mayuscula(Org Cliente)	

Tabla 54: Mapeo de tabla dimensión dim_cliente

Tabla:	dim_req_tipo		
Campo	Tipo	Mapeo	
req_tipo_id	Integer	nextval('todos_inc':regclass)	
req_tipo_nombre	Carácter	Calculo(Split(Categoría,','))	

Tabla 55: Mapeo de tabla dimensión dim_req_tipo

Tabla:	dim_area		
Campo	Tipo	Mapeo	
area_id	Integer	nextval('todos_inc':regclass)	
area_nombre	Carácter	Calculo(Split(Categoría,','))	

Tabla 56: Mapeo de tabla dimensión dim_area

Tabla:	dim_producto		
Campo	Tipo	Mapeo	
producto_id	Integer	nextval('todos_inc':regclass)	
producto_nombre	Carácter	Calculo(Split(Categoría,','))	

Tabla 57: Mapeo de tabla dimensión dim_producto

Tabla:	dim_requerimiento		
Campo	Tipo	Mapeo	
requerimiento_id	Integer	nextval('todos_inc':regclass)	
requerimiento_nombre	Carácter	Calculo(Split(Categoría,','))	

Tabla 58: Mapeo de tabla dimensión dim_requerimiento

Tabla:	dim_oficina		
Campo	Tipo	Mapeo	
oficina_id	Integer	nextval('todos_inc':regclass)	
oficina_nombre	Carácter	Calculo(Split(Oficina,'-'))	

Tabla 59: Mapeo de tabla dimensión dim_oficina

Tabla:	dim_sucursal		
Campo	Tipo	Mapeo	
sucursal_id	Integer	nextval('todos_inc':regclass)	
sucursal_nombre	Carácter	Calculo(Split(Oficina,'-'))	

Tabla 60: Mapeo de tabla dimensión dim_sucursal

Tabla:	dim_fecha		
Campo	Tipo	Mapeo	
fecha_id	Integer	Split(Fecha Apertura,'/') or Split(Fecha Cierre,'/')	
fecha_jahr	Integer	ExtraeFinal(Fecha Apertura,4) or ExtraeFinal(Fecha Cierre,4)	
fecha_dia	Integer	ExtraeInicio(Fecha Apertura,2) or ExtraeInicio(Fecha Cierre,2)	
fecha_mes	Integer	ExtraeMedio(Fecha Apertura,2) or ExtraeMedio(Fecha Cierre,2)	
fecha_mes_nom	character varying(11)	ConvierteANombre(fecha_mes)	
fecha_nombre	character varying(10)	EliminarCeros(Fecha Apertura) or EliminarCeros(Fecha Cierre)	

Tabla 61: Mapeo de tabla dimensión dim_fecha

Tabla:	dim_estado		
Campo	Tipo	Mapeo	
estado_id	Integer	nextval('todos_inc':regclass)	
estado_nombre	Carácter	EliminarEspacios(Estado)	

Tabla 62: Mapeo de tabla dimensión dim_estado

Tabla:	dim_traza		
Campo	Tipo	Mapeo	
traza_id	Integer	nextval('todos_inc':regclass)	
traza_nombre	Carácter	EliminarEspacios(Traza Act)	

Tabla 63: Mapeo de tabla dimensión dim_traza

Entregables generados para el proceso de ETL

Entregable	Descripción
JobServPan	Trabajo que se encarga de correr las transformaciones que llenan las dimensiones
dim_cliente	Transformación usada para la extracción, transformación y carga de los datos de la dimensión cliente

Entregable	Descripción
dim_categoria	Transformación usada para la extracción, transformación y carga de los datos de las dimensiones área, producto, requerimiento y tipo de requerimiento
dim_oficina	Transformación usada para la extracción, transformación y carga de los datos de las dimensiones oficina y sucursal
dim_estado	Transformación usada para la extracción, transformación y carga de los datos de la dimensión estado
dim_fecha	Transformación usada para la extracción, transformación y carga de los datos de la dimensión fecha
dim_traza	Transformación usada para la extracción, transformación y carga de los datos de la dimensión traza
fact_incidente	Transformación usada para la extracción, transformación y carga de los datos de la tabla de hechos fact_incidente

Tabla 64: Entregables ETL

Detalle de los entregables

Trabajo JobServPan

Trabajo que se encarga de correr las transformaciones que llenan las dimensiones

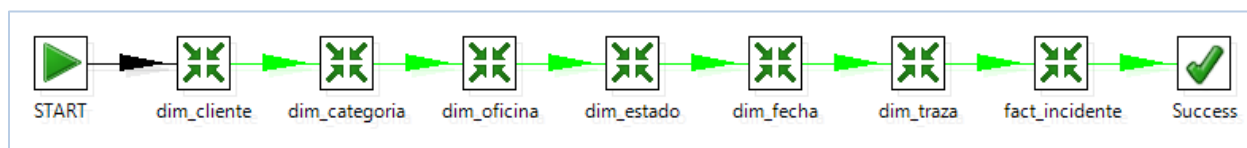


Ilustración 12: JobServPan

Transformación dim_cliente

Es una transformación que crea la dimensión que recopila toda la información de los clientes. Recibe datos de los reportes de incidencias, luego unifica estos datos y elimina los datos nulos y repetidos, para luego insertarlos en la dimensión cliente.

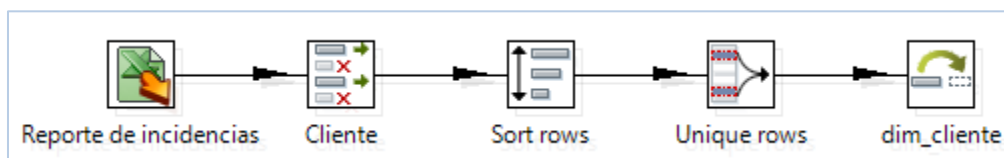


Ilustración 13: dim_cliente

Transformación dim_categoria

Es una transformación que crea la dimensión que recopila toda la información de los tipos de requerimientos, áreas, productos y requerimientos. Recibe datos de los reportes de incidencias, luego unifica estos datos y elimina los datos nulos y repetidos, para luego insertarlos en las dimensiones tipo de requerimiento, área, producto y requerimiento.

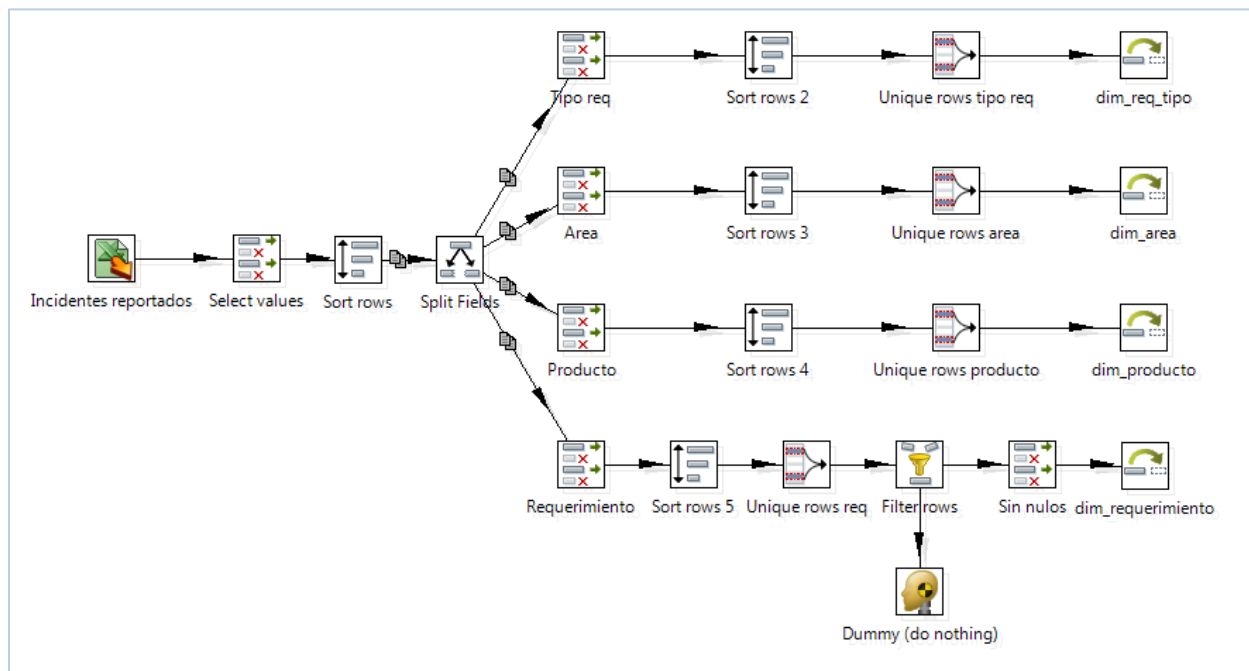


Ilustración 14: dim_categoria

Transformación dim_oficina

Es una transformación que crea la dimensión que recopila toda la información de las oficinas y sucursales. Recibe datos de los reportes de incidencias, luego unifica estos datos y elimina los datos nulos y repetidos, para luego insertarlos en las dimensiones oficina y sucursal.

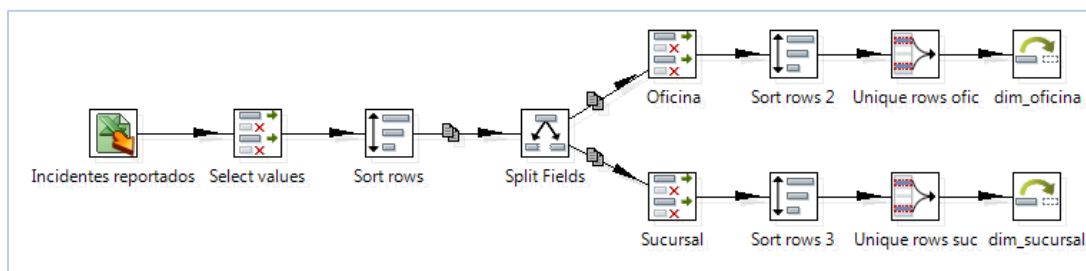


Ilustración 15: dim_oficina

Transformación dim_estado

Es una transformación que crea la dimensión que recopila toda la información de los estados. Recibe datos de los reportes de incidencias, luego unifica estos datos y elimina los datos nulos y repetidos, para luego insertarlos en la dimensión estado.

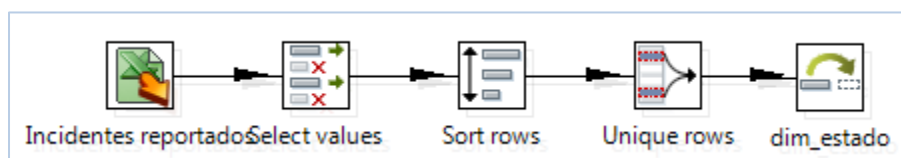


Ilustración 16: dim_estado

Transformación dim_fecha

Es una transformación que crea la dimensión que recopila toda la información de las fechas de apertura y cierre. Recibe datos de los reportes de incidencias, luego unifica estos datos y elimina los datos nulos y repetidos, para luego insertarlos en la dimensión fecha.

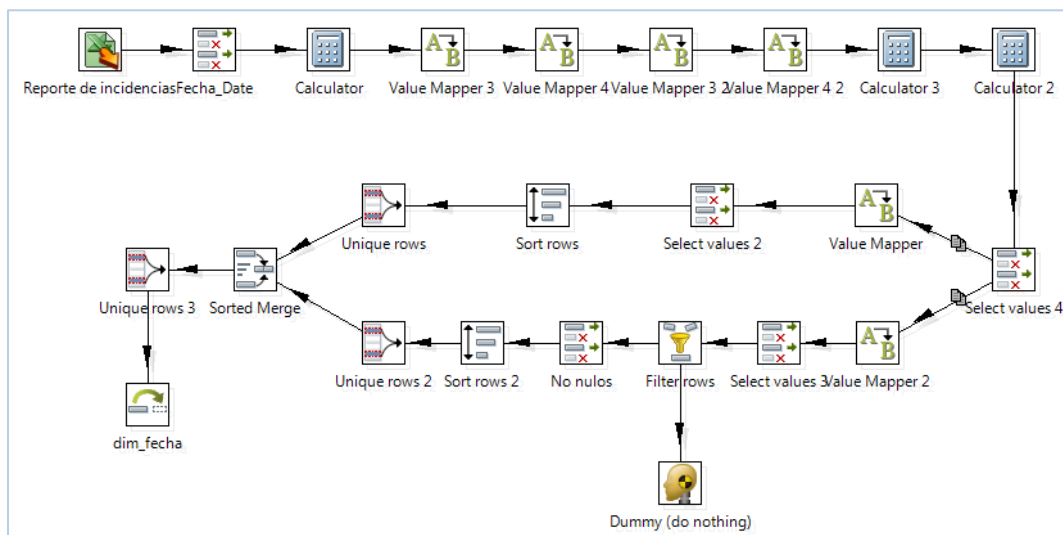


Ilustración 17: dim_fecha

Transformación dim_traza

Es una transformación que crea la dimensión que recopila toda la información de las trazas. Recibe datos de los reportes de incidencias, luego unifica estos datos y elimina los datos nulos y repetidos, para luego insertarlos en la dimensión traza.

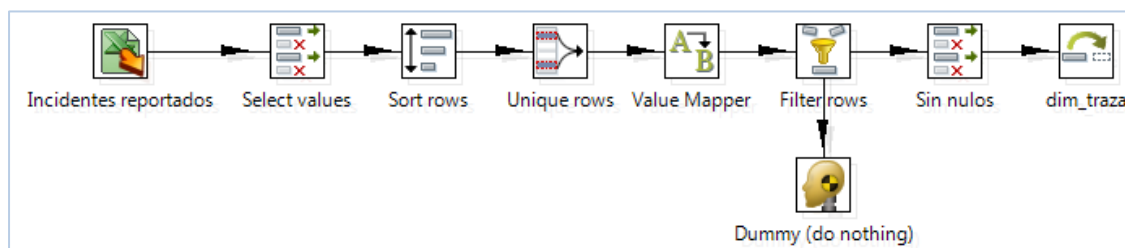


Ilustración 18: dim_traza

Transformación fact_incidente

Es una transformación que carga los datos de la tabla de hechos. Recibe datos de los reportes de incidencias, luego separa cada fila del archivo en los campos correspondientes a los atributos de la tabla de hechos, realiza la búsqueda de los identificadores de cada una de las dimensiones involucradas en la misma y realiza la inserción en la base de datos.

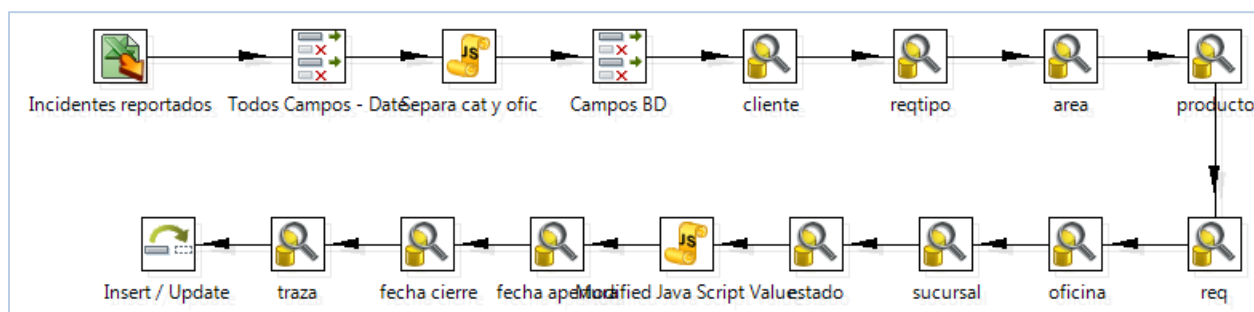


Ilustración 19: fact_incidente

Especificación de la aplicación de usuario final

Perfiles de Usuario:

Usuarios gerenciales: Son usuarios con un perfil estratégico y no predecible.

Usuarios operacionales: Son usuarios que consumen una serie de reportes estándares y pueden consultar información de los cubos.

Usuarios técnicos y de soporte: Son usuarios pertenecientes a la Gerencia de Tecnología, llevan el control de los servidores y el manejo técnico de toda la infraestructura.

Roles de Usuario

Usuario	ETL	Reportes	Cubos	Dashboard
Gerente	Sin acceso	Acceso total	Acceso total	Acceso total
Operacionales	Sin acceso	Acceso total	Acceso total	Acceso total
Técnicos	Acceso total	Acceso total	Acceso total	Acceso total

Usuario	ETL	Reportes	Cubos	Dashboard
Soporte	Con autorización	Con autorización	Con autorización	Con autorización

Tabla 65: Roles de usuario para el Sistema Automatizado de Inteligencia de Negocios

Desarrollo de la aplicación de usuario final

Siguiendo con la especificación de la metodología, el desarrollo de las aplicaciones de los usuarios finales involucra configuraciones de metadatos y la construcción de reportes específicos, vistas dinámicas de los cubos OLAP y dashboard:

Reportes

Reporte	Tiempo de respuesta del sistema
Casos por cliente	10 segundos por mes
Casos por tipo	10 segundos por mes

Tabla 66: Reportes desarrollados

Vistas Dinámicas de los cubos

Casos creados por cliente
Casos creados por estado
Casos creados por fecha
Casos creados por oficina
Casos creados por oficina y requerimiento
Casos creados por tipo de requerimiento
Casos cerrados por fecha
Casos cerrados por oficina
Casos cerrados por oficina y estado

Tabla 67: Vistas Dinámicas desarrolladas

Dashboard

Casos anualmente creados
Casos anualmente cerrados

Tabla 68: Dashboard desarrollados

Implementación

Para el desarrollo propuesto surgió la necesidad de establecer distintos niveles de acceso definiendo roles para los usuarios.

Adicionalmente, luego de realizar el proceso de ETL se realizó la creación de distintos cubos OLAP representando un conjunto de dimensiones, métricas y tablas de hechos, de manera tal de que los resultados de las consultas cuenten con filtros que permitan utilizar distintos criterios de búsqueda.

A continuación la vista inicial donde se muestra la pantalla de ingreso donde debe colocarse el usuario y contraseña asignados.



Ilustración 20: Página de inicio de sesión en el sistema

Vistas dinámicas

Gerencia	Gerencia de Atención al Cliente
Entregable	Casos creados por cliente
Descripción	Permite visualizar la cantidad de casos que han sido creados por un cliente

Captura de pantalla

Cliente.Cliente	Medidas
● Casos_Cliente	
All DCliente.HClientes	58.802
100 % BANCO, BANCO UNIVERSAL, C.A.	12
1932 PREBO, C.A.	2
3J INVERSIONES, C.A.	2
A. COMPANY, C.A.	2
A.C DE RESPONSABILIDAD LIMITADA PUNTO FIJO	2
A.C. CLINICA DISPENSARIO PADRE MACHADO	2
A.C. CLUB DE LEONES SAN CRISTOBAL MONARCA	4
A.C. DE RESPONSABILIDAD LIMITADA SAN JOSE OBRERO	6
A.C. HOSPITAL SAN JOSE DE LAS HERMANITAS DE LOS POBRES	2
A.E AEROEXPRESOS EJECUTIVOS CA	8
ACERO GANGA C A	10
ACERO OFERTA, C.A.	14
ACEROS LAMINADOS, C.A.	6
ACRILUM, C.A.	44
ACUMULADORES DUNCAN, C.A.	2
ADMIN PROV COMPLEJO TURISTICO RECREACIONAL VEGASOL	2
ADMINISTRADORA DANORAL, C.A.	2
ADMINISTRADORA DARMAY, C.A.	54
ADMINISTRADORA ESTACECETE, C.A.	14
ADMINISTRADORA SAN FRANCISCO, C.A.	4

Tabla 69: Descripción del entregable Vista dinámica Casos creados por cliente

Gerencia	Gerencia de Atención al Cliente
Entregable	Casos creados por estado
Descripción	Permite visualizar la cantidad de casos que han sido creados por el estado en que se encuentran los mismos

Captura de pantalla

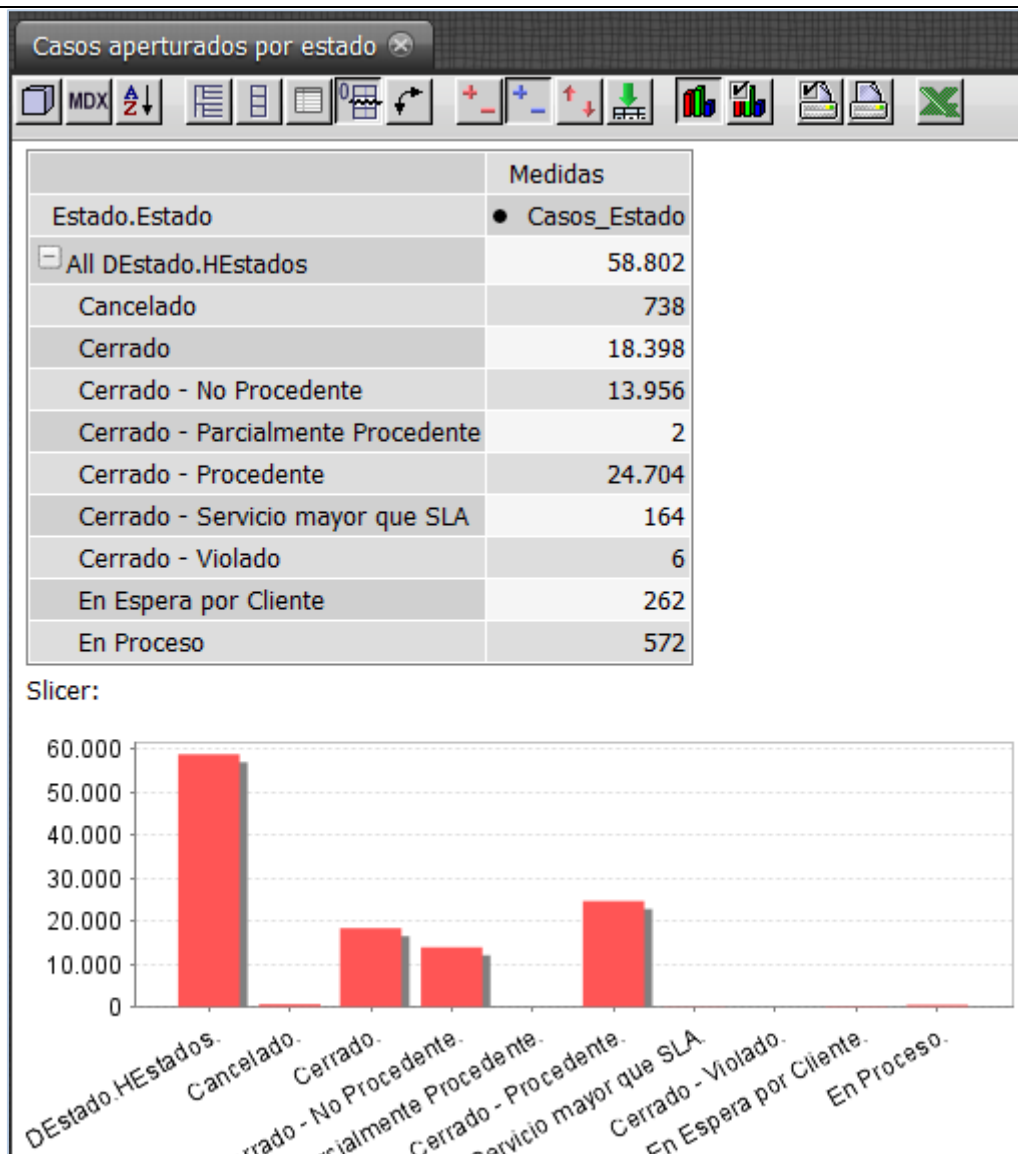


Tabla 70: Descripción del entregable Vista dinámica Casos creados por estado

Gerencia	Gerencia de Atención al Cliente
Entregable	Casos creados por fecha
Descripción	Permite visualizar la cantidad de casos que han sido creados por fecha

Captura de pantalla

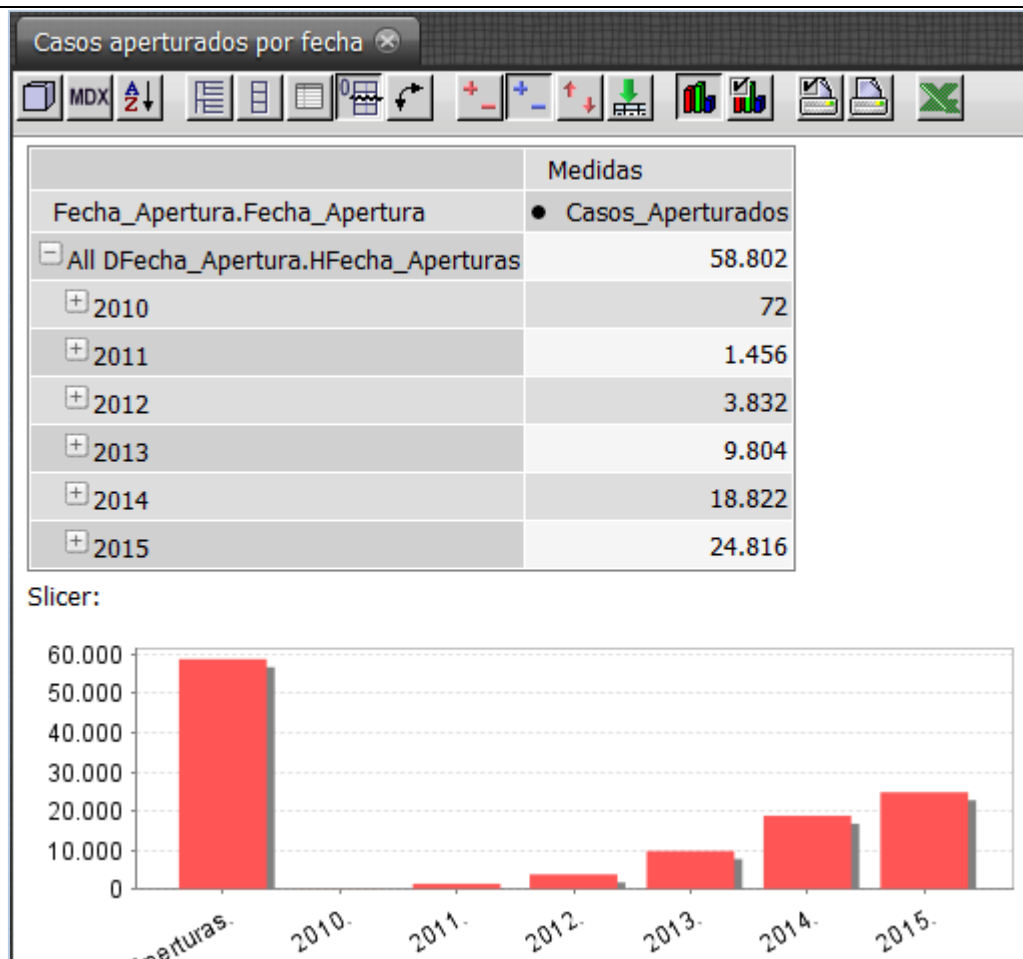


Tabla 71: Descripción del entregable Vista dinámica Casos creados por fecha

Gerencia	Gerencia de Atención al Cliente
Entregable	Casos creados por oficina
Descripción	Permite visualizar la cantidad de casos que han sido creados por oficina. Permite un Drill-Down de las sucursales que pertenecen a una oficina.

Captura de pantalla

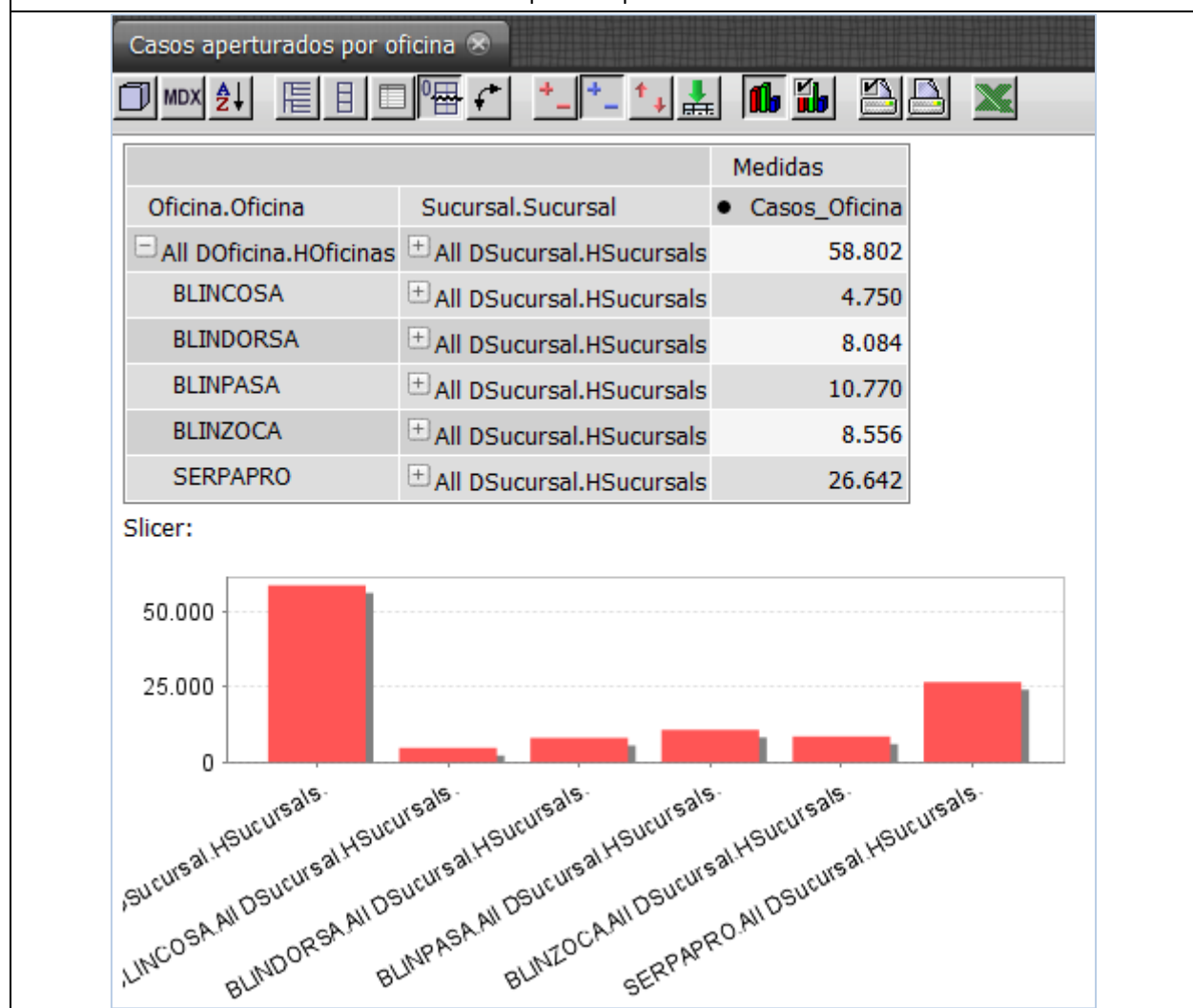


Tabla 72: Descripción del entregable Vista dinámica Casos creados por oficina

Gerencia	Gerencia de Atención al Cliente
Entregable	Casos creados por oficina y requerimiento
Descripción	Permite visualizar la cantidad de casos que han sido creados por oficina. Permite un Drill-Down de la oficina, visualizando la sucursal, el tipo de requerimiento, el área, el producto y el requerimiento del caso o casos creados, según la cantidad.

Captura de pantalla

Casos abiertos por oficina y requerimiento						
Oficina.Oficina	Sucursal.Sucursal	Tipo_Req.Tipo_Req	Area.Area	Producto.Producto	Requerimiento.Requerimiento	Medidas
All DOficina.HOficinas	All DSucursal.HSucursals	All DTipo_Req.HTipo_Reqs	All DArea.HAreas	All DProducto.HProductos	All DRequerimiento.HRequerimientos	● Casos_Apertura
BLINCOSA	All DSucursal.HSucursals	All DTipo_Req.HTipo_Reqs	All DArea.HAreas	All DProducto.HProductos	All DRequerimiento.HRequerimientos	58.802
	Acarigua	All DTipo_Req.HTipo_Reqs	All DArea.HAreas	All DProducto.HProductos	All DRequerimiento.HRequerimientos	4.750
	Barinas	All DTipo_Req.HTipo_Reqs	All DArea.HAreas	All DProducto.HProductos	All DRequerimiento.HRequerimientos	1.210
	Barquisimeto	All DTipo_Req.HTipo_Reqs	All DArea.HAreas	All DProducto.HProductos	All DRequerimiento.HRequerimientos	492
		All DTipo_Req.HTipo_Reqs	All DArea.HAreas	All DProducto.HProductos	All DRequerimiento.HRequerimientos	3.048
BLINDORSA	All DSucursal.HSucursals	All DTipo_Req.HTipo_Reqs	All DArea.HAreas	All DProducto.HProductos	All DRequerimiento.HRequerimientos	8.084
BLINPASA	All DSucursal.HSucursals	All DTipo_Req.HTipo_Reqs	All DArea.HAreas	All DProducto.HProductos	All DRequerimiento.HRequerimientos	10.770
BLINZOCA	All DSucursal.HSucursals	All DTipo_Req.HTipo_Reqs	All DArea.HAreas	All DProducto.HProductos	All DRequerimiento.HRequerimientos	8.556
SERPAPRO	All DSucursal.HSucursals	All DTipo_Req.HTipo_Reqs	All DArea.HAreas	All DProducto.HProductos	All DRequerimiento.HRequerimientos	26.642

Tabla 73: Descripción del entregable Vista dinámica Casos creados por oficina y requerimiento

Gerencia	Gerencia de Atención al Cliente
Entregable	Casos creados por tipo de requerimiento
Descripción	Permite visualizar la cantidad de casos que han sido creados por tipo de requerimiento

Captura de pantalla

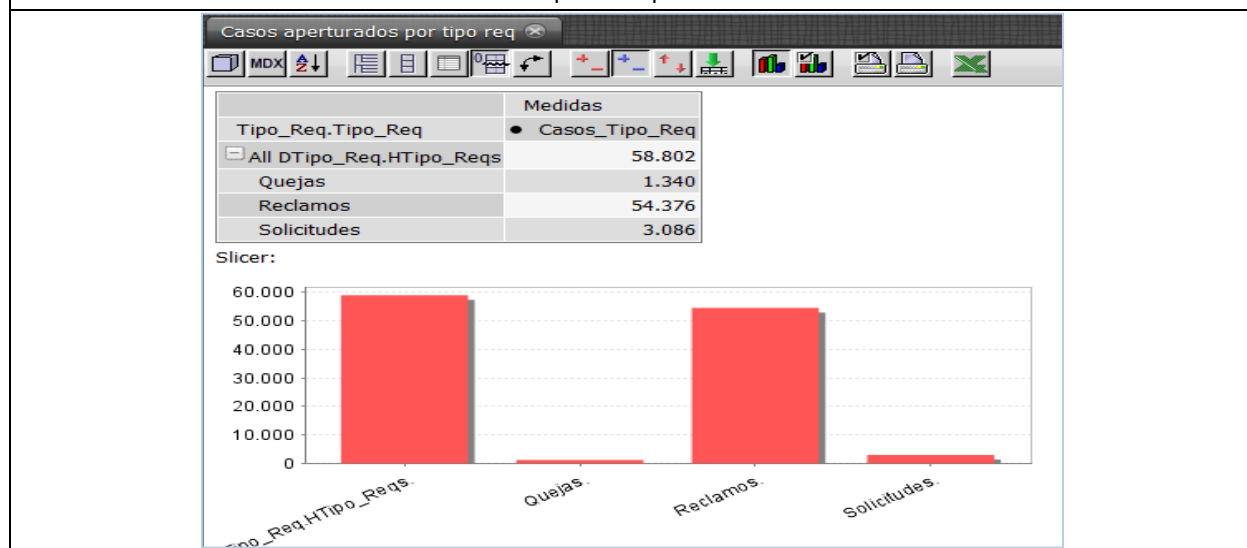


Tabla 74: Descripción del entregable Vista dinámica Casos creados por tipo de requerimiento

Gerencia	Gerencia de Atención al Cliente
Entregable	Casos cerrados por fecha
Descripción	Permite visualizar la cantidad de casos que han sido cerrados por fecha

Captura de pantalla

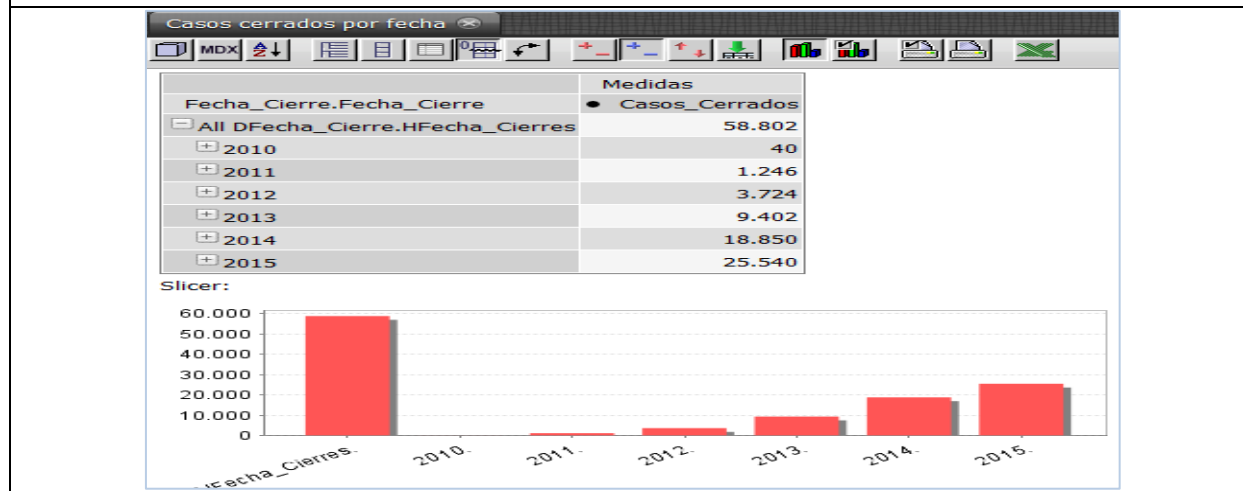


Tabla 75: Descripción del entregable Vista dinámica Casos cerrados por fecha

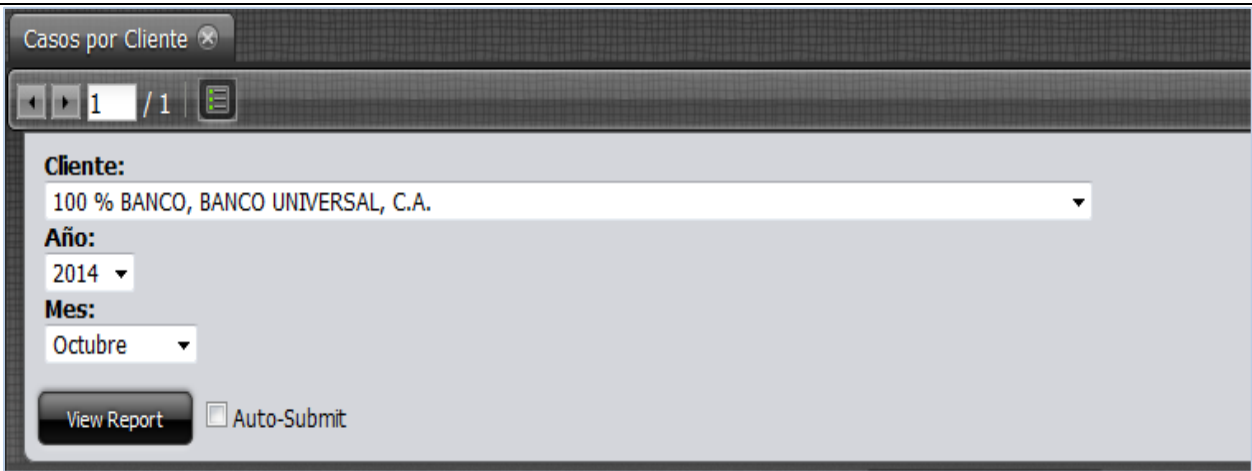
Gerencia	Gerencia de Atención al Cliente
Entregable	Casos cerrados por oficina y estado
Descripción	Permite visualizar la cantidad de casos que han sido cerrados por oficina. Permite un Drill-Down de la oficina, visualizando la sucursal, el estado, la traza, la fecha de apertura y la fecha de cierre del caso o casos, según la cantidad.

Captura de pantalla

Oficina.Oficina	Sucursal.Sucursal	Estado.Estado	Traza.Traza	Fecha_Apertura.Fecha_Apertura	Fecha_Cierre.Fecha_Cierre	Medidas
All DOficina.HOficinas	All DSucursal.HSucursals	All DEstado.HEstados	All DTraza.HTrazas	All DFecha_Apertura.HFecha_Aperturas	All DFecha_Cierre.HFecha_Cierres	● Casos_Cerrados
BLINCOSA	All DSucursal.HSucursals	All DEstado.HEstados	All DTraza.HTrazas	All DFecha_Apertura.HFecha_Aperturas	All DFecha_Cierre.HFecha_Cierres	58.802
BLINDORSA	All DSucursal.HSucursals	All DEstado.HEstados	All DTraza.HTrazas	All DFecha_Apertura.HFecha_Aperturas	All DFecha_Cierre.HFecha_Cierres	4.750
BLINPASA	All DSucursal.HSucursals	All DEstado.HEstados	All DTraza.HTrazas	All DFecha_Apertura.HFecha_Aperturas	All DFecha_Cierre.HFecha_Cierres	8.084
BLINZOCA	All DSucursal.HSucursals	All DEstado.HEstados	All DTraza.HTrazas	All DFecha_Apertura.HFecha_Aperturas	All DFecha_Cierre.HFecha_Cierres	10.770
SERPAPRO	All DSucursal.HSucursals	All DEstado.HEstados	All DTraza.HTrazas	All DFecha_Apertura.HFecha_Aperturas	All DFecha_Cierre.HFecha_Cierres	8.556
						26.642

Tabla 76: Descripción del entregable Vista dinámica Casos cerrados por oficina y estado

Reportes

Gerencia	Gerencia de Atención al Cliente
Entregable	Casos por cliente
Definición	Permite obtener información de los casos creados por un cliente
Campo / Descripción	
Parámetros: Cliente: Indica el cliente por el cual se filtrarán los casos. Año: Indica el año por el cual se filtrarán los casos, de acuerdo al cliente. Mes: Indica el mes por el cual se filtrarán los casos, de acuerdo al año. Reporte: Cliente: Muestra el nombre del cliente por el cual se han filtrado los casos. Tipo de requerimiento: Indica el tipo de requerimiento del caso. Área: Indica el área del caso. Producto: Indica el producto del caso. Requerimiento: Indica el requerimiento del caso. Oficina: Indica la oficina del caso. Sucursal: Indica la sucursal del caso. Estado: Indica el estado del caso. Fecha de apertura: Indica la fecha de apertura del caso. Total casos creados: Indica el total de casos creados por el cliente.	
Captura de pantalla	
	

Reporte 17 de febrero de 2016

Casos abiertos por Cliente

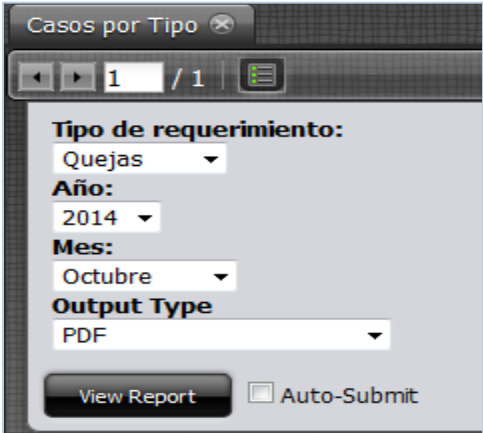
Cliente: 100 % BANCO, BANCO UNIVERSAL, C.A.

Tipo de Requerimiento	Año	Producto	Requerimiento	Oficina	Sucursal	Estado	Fecha de apertura
Reclamos	GEN	Información General	Información General	BUNPASA	Valencia	Cerrado	03/03/14
Reclamos	GEN	Información General	Información General	BUNPASA	Valencia	Cerrado	03/03/14

Total casos abiertos: 2

8/1

Tabla 77: Descripción del entregable Reporte Casos por cliente

Gerencia	Gerencia de Atención al Cliente
Entregable	Casos por tipo de requerimiento
Definición	Permite obtener información de los casos creados por un cliente
Campo / Descripción	
Parámetros: Tipo de requerimiento: Indica el tipo de requerimiento por el cual se filtrarán los casos. Año: Indica el año por el cual se filtrarán los casos, de acuerdo al tipo de requerimiento. Mes: Indica el mes por el cual se filtrarán los casos, de acuerdo al año. Output Type: Permite escoger el formato en que se mostrarán los resultados. Reporte: Tipo de requerimiento: Muestra el nombre del cliente por el cual se han filtrado los casos. Área: Indica el área del caso. Producto: Indica el producto del caso. Requerimiento: Indica el requerimiento del caso. Oficina: Indica la oficina del caso. Sucursal: Indica la sucursal del caso. Estado: Indica el estado del caso. Fecha de apertura: Indica la fecha de apertura del caso. Total casos creados: Indica el total de casos creados por el cliente.	
Captura de pantalla	
	

Página: 1 de 5 70%

W B Rincón 17 de febrero de 2018

Casos abiertos por Tipo

Tipo: Quejas

Área	Producto	Requerimiento	Oficina	Sucursal	Estado	Fecha de apertura
GEN	Servicio	Quejas por Servicio	BLINCOSA	Barquisimeto	Cerrado	02.02.14
GEN	Servicio	Quejas por Servicio	BLINCOSA	Barquisimeto	Cerrado	02.02.14
GEN	Servicio	Quejas por Servicio	BLINCOSA	Barquisimeto	Cerrado	21.02.14
GEN	Servicio	Quejas por Servicio	BLINCOSA	Barquisimeto	Cerrado	21.02.14
GEN	Servicio	Quejas por Servicio	BLINDORSA	Portafar	Cerrado	01.02.14
GEN	Servicio	Quejas por Servicio	BLINDORSA	Portafar	Cerrado	01.02.14
GEN	Servicio	Quejas por Servicio	BLINDORSA	Portafar	Cerrado	02.02.14
GEN	Servicio	Quejas por Servicio	BLINDORSA	Portafar	Cerrado	02.02.14
GEN	Servicio	Quejas por Servicio	BLINDORSA	Portafar	Cerrado	12.02.14
GEN	Servicio	Quejas por Servicio	BLINDORSA	Portafar	Cerrado	12.02.14
GEN	Servicio	Quejas por Servicio	BLINDORSA	Portafar	Cerrado	26.02.14
GEN	Servicio	Quejas por Servicio	BLINDORSA	Portafar	Cerrado	26.02.14
GEN	Servicio	Quejas por Servicio	BLINDORSA	Puerto Ordaz	Cerrado	08.02.14
GEN	Servicio	Quejas por Servicio	BLINDORSA	Puerto Ordaz	Cerrado	08.02.14
Total casos abiertos: 70						

1/1

Tabla 78: Descripción del entregable Reporte Casos por tipo de requerimiento

Dashboards

Gerencia	Gerencia de Atención al Cliente
Entregable	Casos anualmente creados
Definición	Permite visualizar información relevante acerca de los casos que se aperturan al año
Descripción Parámetro año: Indica el año por el cual se filtrará la información a mostrar. Dashboard: Consta de cinco (5) cuadros. 1. Casos por tipo: muestra un gráfico de barras verticales con el total de casos creados por tipo de requerimiento. 2. Casos por mes: muestra un gráfico de barras verticales con el total de casos creados por mes. 3. Top 10 Clientes: muestra un gráfico de torta con los 10 clientes que más casos han aperturado, y el porcentaje que representan. 4. Casos por oficina: muestra un gráfico de torta del total de casos abiertos por oficina, así como el porcentaje que estas representan. 5. Top 10 Sucursales: muestra un gráfico de torta con las 10 sucursales con más casos creados, y el porcentaje que representan.	
Captura de pantalla	
	

Tabla 79: Descripción del entregable Dashboard Casos anualmente creados

Gerencia	Gerencia de Atención al Cliente
Entregable	Casos anualmente cerrados
Definición	Permite visualizar información relevante acerca de los casos que se cierran al año

Descripción

Parámetro año: Indica el año por el cual se filtrará la información a mostrar.

Dashboard: Consta de cinco (5) cuadros.

1. Casos por tipo: muestra un gráfico de barras verticales con el total de casos cerrados por tipo de requerimiento.
2. Casos por mes: muestra un gráfico de barras verticales con el total de casos cerrados por mes.
3. Top 10 Clientes: muestra un gráfico de torta con los 10 clientes que más casos han cerrados, y el porcentaje que representan.
4. Casos por oficina: muestra un gráfico de torta del total de casos cerrados por oficina, así como el porcentaje que estas representan.
5. Top 10 Sucursales: muestra un gráfico de torta con las 10 sucursales con más casos cerrados, y el porcentaje que representan.

Captura de pantalla



Tabla 80: Descripción del entregable Dashboard Casos anualmente cerrados

Despliegue

En esta etapa se realizaron las pruebas de aceptación y se incluye el mantenimiento del sistema. A continuación se describen cada una:

Pruebas de aceptación

Se realizaron pruebas funcionales y no funcionales de la aplicación, dada la dinámica de la organización se mantuvo comunicación constante con la Gerencia de Atención al Cliente a fin de garantizar que el desarrollo propuesto cumpliera con sus expectativas.

Pruebas de aceptación	Fecha de aceptación
Pruebas Funcionales: Ejecución, revisión y retroalimentación de las funcionalidades previamente diseñadas para la Gerencia de Atención al Cliente	10 /10/2015 al 28/10/2015
Pruebas No-Funcionales: Retroalimentación de la facilidad con que los usuarios pueden acceder a la información de los reportes, cubos y dashboard.	29/10/2015 al 4/01/2016

Tabla 81: Pruebas de aceptación de los usuarios

Mantenimiento y crecimiento

En cuanto al mantenimiento y crecimiento, se realizó la documentación de todas las fases de desarrollo del software según la metodología empleada.

Capítulo IV – Conclusiones

Como resultado de éste T.E.G podemos decir que se logró implementar el sistema de para inteligencia de negocios para el apoyo a la toma de decisiones y facilitar la gestión de la información producto de los requerimientos, quejas y reclamos manejada por el Centro de Gestión al Cliente en una organización que presta servicio de procesamiento, recogida y entrega de valores. Este sistema se desarrolló siguiendo la metodología de Kimball, denominada Business Dimensional Lifecycle (BDL).

Trabajamos con la plataforma PentahoBI Suite Community Edition, ya que permitió realizar extracción, carga y transformación de la data que se obtuvo del sistema operacional. Adicionalmente se realizó el modelado multidimensional, se diseñaron los reportes solicitados así como la creación de diversas vistas dinámicas de los cubos OLAP y la visualización de los dashboards a través de su servidor de aplicaciones. Como sistema manejador de Base de Datos utilizamos PostgreSQL dado que nos permitía aplicar el diseño físico de la base de datos.

El sistema desarrollado fue sujeto a varias pruebas por parte de las áreas involucradas y se encuentra en producción actualmente.

REFERENCIAS BIBLIOGRÁFICAS

- 2ndQuadrant. (s.f.). *2ndQuadrant Professional PostgreSQL*. Recuperado el 23 de 02 de 2016, de PostgreSQL: <http://2ndquadrant.com/es/postgresql/postgresql-vs-mysql/>
- Boletín Digital. (2008). *Boletín de Asesoría Gerencial. La inteligencia de negocios (Business Intelligence)*, 13. Caracas, Distrito Capital, Venezuela: Espiñeira, Sheldon y Asociados.
- Bouman, R., & Dongen, J. (2009). *Pentaho Solutions Business Intelligence and Data*. Indianapolis: Wiley Publishing, Inc.
- Cannon, D., & Wheeldon, D. (2007). *Service Operation* (Primera ed.). Londres: The Stationery Office. Recuperado el 21 de 10 de 2014
- Codd, E., Codd, S., & Salley, C. (20 de 09 de 1998). *Providing OLAP to User-Analysts: An IT Mandate*. E.F. Codd Associates. California: Hyperion Solutions Corporations. Recuperado el 20 de Septiembre de 2014, de Universität Jena Web site: http://www.minet.uni-jena.de/dbis/lehre/ss2005/sem_dwh/lit/Cod93.pdf
- Gartner. (2013). *Magic Quadrant for Business Intelligence and Analytics Platforms*. Stamford: Gartner, Inc. Recuperado el 22 de Septiembre de 2014, de http://www.walmeric.com/pdf/2013_gartner_magic_quadrant_for_bi_and_analytics.pdf
- Inmon, W. (2002). *Building the Data Warehouse*. Estados Unidos: John Wiley & Sons, Inc.
- Kimball, R. y. (2002). *The data warehouse toolkit : the complete guide to dimensional modeling*. Estados Unidos: Wiley Computer Publishing.
- Kimball, R., Margy, R., Warren, T., Joy, M., & Bob, B. (2002). *The data warehouse lifecycle toolkit*. Estados Unidos: Wiley Computer Publishing.
- Langton, R. (s.f.). *Chief Executive SKF*. North America.
- Laudon, K. C., & Laudon, J. P. (2012). *Sistemas de información gerencial*. México: Pearson.
- Martínez, R. (02 de 10 de 2010). *PostgreSQL-es*. Recuperado el 23 de 02 de 2016, de Portal en español sobre PostgreSQL: http://www.postgresql.org.es/sobre_postgresql

MicroStrategy. (2002). *The 5 Styles of Business Intelligence: INDUSTRIAL-STRENGTH BUSINESS INTELLIGENCE*. Virginia: MicroStrategy.

MySQL. (22 de 02 de 2016). *MySQL*. Recuperado el 23 de 02 de 2016, de The world's most popular open source database: <http://dev.mysql.com/doc/refman/5.7/en/features.html>

Opensource. (s.f.). *Opensource.com*. Recuperado el 24 de Septiembre de 2014, de Discover an open source world: opensource.com/resources/what-open-source