



UNIVERSIDAD CENTRAL DE VENEZUELA
FACULTAD DE CIENCIAS
ESCUELA DE COMPUTACIÓN
CENTRO DE INVESTIGACIÓN EN SISTEMAS DE INFORMACIÓN

**PROTOTIPO DE SOLUCIÓN ANALÍTICA PARA LA OBTENCIÓN DE INDICADORES DE
RENTABILIDAD EN INSTITUCIONES FINANCIERAS**

Trabajo Especial de Grado presentado ante la ilustre
Universidad Central de Venezuela por
Br. Francisco Chacoa
Para optar al título de Licenciado en Computación

Tutoras:

Profa. Concettina Di Vasta
Profa. Brenda Lopez

Enero, 2018

UNIVERSIDAD CENTRAL DE VENEZUELA
FACULTAD DE CIENCIAS
ESCUELA DE COMPUTACIÓN

ACTA

Quienes suscriben, miembros del jurado designado por el Consejo de la Escuela de Computación, para examinar el Trabajo Especial de Grado titulado “Prototipo de Solución de Inteligencia de Negocio para la obtención de indicadores de Rentabilidad en Instituciones Financieras” y presentado por el bachiller: Br. Chacoa Saputelli Francisco José titular de la Cédula de Identidad V-19.692.025, a los fines de optar al título de **Licenciado en Computación**, dejamos constancia de lo siguiente:


Leído como fue, dicho trabajo por cada uno de los miembros del jurado, se fijó el día 1 de Febrero de 2018, a las 9:30 am horas, para que el autor lo defendiera en forma pública, lo que este hizo en la Sala I 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 aprobar con la nota de 20 puntos.

Adicionalmente se hace constar que la profesora Brenda López también actuó como tutora del Trabajo Especial de Grado.

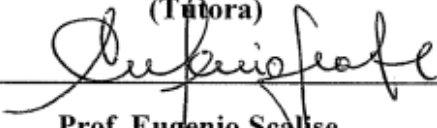
En fe de lo cual se levanta la presente Acta, en Caracas el día 1 de Febrero de 2018.



Prof. Concettina Di Vasta



Prof. Franky Uzcátegui
(Jurado)

(Tutora)


Prof. Eugenio Scalise
(Jurado)

AGRADECIMIENTOS

A Dios, mi familia y la Universidad Central de Venezuela.

Venciendo las sombras, siempre.

Francisco Chacoa

DEDICATORIA

A mi familia, a la de sangre y la de la vida, incondicionales siempre y un soporte de acero ante la dificultad. Son modelos a seguir invaluable, responsables del ser humano que soy e influencia para el futuro.

A mamá, papá, las tías todas, los abuelos, mis hermanos Ricardo, Diego, Luis y Mariano.

A mi madrina Paola, modelo y pilar en mi carrera.

A mi negra, Muguete, incondicional siempre, con paciencia y amor infinitos.

ÍNDICE

RESUMEN	8
INTRODUCCIÓN	9
CAPÍTULO I	11
1.1 Planteamiento del Problema	11
1.2 Objetivos.....	12
1.2.1 General	12
1.2.2 Específicos	12
1.3 Solución Propuesta	13
1.4 Justificación.....	14
CAPÍTULO II.....	16
MARCO CONCEPTUAL.....	16
2.1 Sistema Financiero.....	16
2.1.1 Funciones del Sistema Financiero	17
2.1.2 Importancia del Sistema Financiero	17
2.1.3 Componentes del Sistema Financiero	18
2.1.3.1 Activos financieros.....	18
2.1.3.2 Intermediarios financieros	19
2.1.3.3 Mercados Financieros.....	19
2.2 Rentabilidad.....	20
2.3 Indicadores.....	21
2.3.1 Tipos de indicadores	22
2.4 Sistema de Información.....	25
2.5 Inteligencia de Negocio.....	37
2.5.1 Arquitectura de una Solución analítica	38
2.5.2 Herramientas para la Construcción de una Solución Analítica	46
CAPÍTULO III	49
MARCO METODOLÓGICO	49
CAPÍTULO IV.....	57
MARCO APLICATIVO	57
4.1 Fases del Proyecto.....	58
4.1.1 Planificación del Proyecto.....	58
4.1.2 Definición de los Requisitos del Negocio.....	59
4.1.3 Diseño de la Arquitectura Técnica	65
4.1.4 Selección de Productos e Instalación.....	67
4.1.5 Diseño del Modelo Dimensional.....	69
4.1.5.1 Definir el Proceso de Negocio.....	69
4.1.5.2 Identificar el Nivel de Granularidad	69
4.1.5.3 Definir las Dimensiones	70
4.1.5.4 Identificación de los Hechos y las Tablas de Hechos	71
4.1.6 Diseño Físico.....	72
4.1.7 Diseño de procesos ETL	73
4.1.8 Desarrollo de las Aplicaciones Analíticas	75
4.1.9 Desarrollo	80
4.1.10 Mantenimiento y Escalabilidad	85
CONCLUSIONES	86

Universidad Central de Venezuela

Facultad de Ciencias

Escuela de Computación

Centro de Investigación en Sistemas de Información

**PROTOTIPO DE SOLUCIÓN ANALÍTICA PARA LA OBTENCIÓN DE
INDICADORES DE RENTABILIDAD EN INSTITUCIONES FINANCIERAS**

Autores: Chacoa Francisco

Tutoras: Profa. Concettina Di
Vasta, Profa. Brenda López

Año: 2017

RESUMEN

Hoy en día, más que en ningún momento de la historia, resulta muy importante la generación de información de manera eficiente y oportuna. Las ventajas y oportunidades de negocios duran apenas días, incluso horas. Conocer la rentabilidad de un activo o cliente oportunamente, en sus distintos puntos de vista, puede marcar la diferencia entre una inversión provechosa o el peor de los fracasos financieros.

Siendo así, el objetivo del presente Trabajo Especial de Grado es la implementación de un prototipo de solución analítica que permita generar indicadores de rentabilidad financiera, de manera eficaz y eficiente para asistir en la toma de decisiones en lo alto de la pirámide organizacional. Para esto, se apoya en las metodologías y conceptos presentados en el presente trabajo, relacionados con el área de inteligencia de negocio.

Haciendo uso de la metodología Kimball para el desarrollo de soluciones analíticas mediante el uso de almacenes de datos y con el apoyo de herramientas de software propietario (Oracle Database y Oracle Business Intelligence (OBI)) y software libre (Pentaho Data Integration o PDI) se definen luego una serie de indicadores de rentabilidad que darán soporte a las decisiones de la institución financiera que los utilice.

Palabras clave: Rentabilidad, Almacén de datos, Mercado financiero, Activos, Metodología Kimball, Inteligencia de negocio.

**Palabras clave: Prototipo , Inteligencia de negocio , Finanzas , Rentabilidad , Kimball
Solución analítica**

INTRODUCCIÓN

La economía es el motor que impulsa el desarrollo de cualquier país, de modo que una economía sana y en crecimiento es, sin duda, sinónimo de avance y superación, que se traduce, en la mayoría de los casos, en una mejor calidad de vida.

Tomando en cuenta lo anterior, se puede decir que la salud del sistema financiero depende, entre otras cosas, de tomar las decisiones correctas en el momento apropiado y de colocar el dinero de manera acertada en los actores correctos. En este aspecto, pocos indicadores proveen tanta información sobre el estado y la viabilidad de la colocación de capital como los indicadores de rentabilidad.

Dado que uno de los principales objetivos de las instituciones financieras es, servir como intermediarios entre actores con capital y otros que lo necesitan, es necesario que dichas instituciones estén en capacidad de evaluar el estado de las inversiones realizadas y prever el desempeño de inversiones por realizar.

Los indicadores de rentabilidad apoyan en gran medida la toma de decisiones relacionada a estos aspectos, toda vez que representan la relación entre el capital utilizado o invertido en un activo y el retorno que genera el mismo.

Pero se trata de indicadores complejos, que se alimentan de gran cantidad de datos y cuyas fuentes en muchos casos no se encuentran unificadas, y en muchos otros, los datos no son ni siquiera proporcionados por fuentes compatibles entre sí. Esto, en conjunto con la limitada capacidad y velocidad humana para procesar y unificar datos, hacen que la necesidad de generar una solución que realice estas tareas de manera eficiente sea crítica.

De la generación de información por parte de esta solución analítica depende, en gran medida, la toma de decisiones oportunas.

Por esta razón, el Trabajo Especial de Grado (TEG) presentado a continuación se centra en el desarrollo de un Prototipo de Solución Analítica para la obtención de indicadores de Rentabilidad Financiera; y posee la siguiente estructura:

El Capítulo I, contiene el planteamiento del problema que origina el trabajo en cuestión, así como el objetivo general, los objetivos específicos para luego proponer una solución y justificarla debidamente.

En el Capítulo II, se muestra el marco conceptual, en el cual se presentan los fundamentos teóricos investigados, para que sirvan de base al desarrollo de este trabajo.

El Capítulo III, describe la metodología propuesta por Ralph Kimball como la elegida para el desarrollo de una solución de inteligencia de negocio, compuesta por distintas fases que suponen una guía para la realización de la solución analítica.

En el Capítulo IV, se describen las actividades realizadas en cada una de las fases de la metodología de Ralph Kimball para el desarrollo particular del prototipo de solución analítica para la obtención de indicadores de rentabilidad en instituciones financieras.

Finalmente, se presentan las conclusiones, bibliografías y referencias digitales consultadas para la elaboración del trabajo.

CAPÍTULO I

PROBLEMA DE INVESTIGACIÓN

1.1 Planteamiento del Problema

Los indicadores de rentabilidad, en cualquiera de sus formas, proveen a las organizaciones financieras de información valiosa para los procesos de toma de decisiones. Pero, estos indicadores necesitan, para su cálculo, de muchos datos provenientes de fuentes diversas. Tradicionalmente este cálculo se realiza de manera manual o semi manual, haciendo uso de hojas de cálculo y unificando las fuentes de datos de manera poco confiable, es decir, que tanto la unificación de fuentes, como el cálculo de los indicadores y la identificación de los datos se realizan más lentamente, dada la limitada capacidad humana para dichas tareas. Derivando todo esto en la generación tardía de información crítica.

Es en este momento en el que destaca la importancia de poder calcular dichos indicadores de manera no solo confiable, sino también de manera eficaz y oportuna. Al tratarse de indicadores que pueden afectar la toma de decisiones en distintos procesos de negocio y en distintos niveles de la pirámide organizacional, la oportuna disponibilidad de los mismos es crítica para el aprovechamiento de ventajas de mercado, potenciación de fortalezas (como oficinas o clientes altamente rentables) o corrección de puntos débiles (llámese caída en la rentabilidad de un activo y los correctivos necesarios para detenerla o revertirla).

De modo que es evidente que los mecanismos tradicionales de procesamiento de datos y generación de indicadores son insuficientes hoy en día para satisfacer la demanda de información de cualquier institución financiera.

Resaltando así la necesidad de mecanismos eficaces y eficientes, que agilicen los procesos de unificación de fuentes de datos y estandarización de los mismos y además faciliten el acceso a través de herramientas sencillas de consulta.

1.2 Objetivos

1.2.1 General

Desarrollar un prototipo de solución de analítica para la obtención de indicadores que apoyen la toma de decisiones relacionadas a la rentabilidad en instituciones financieras.

1.2.2 Específicos

Entender e identificar las necesidades del negocio con respecto al acceso a la información y cálculo de indicadores relacionados a la rentabilidad financiera.

Definir los indicadores y reportes relacionados con la rentabilidad financiera.

Definir la arquitectura técnica del prototipo de la solución analítica.

Definir las políticas que se utilizarán para asegurar la calidad de los datos y garantizar una integración adecuada de los mismos.

Desarrollar un almacén de datos para que luego sea alimentado con datos a través de procesos de extracción, transformación y carga, unificando fuentes y tipos de datos para asegurar la calidad final de los datos, así como la integridad y veracidad de los reportes generados.

Conceptualizar y construir tanto los indicadores como los reportes de gestión propuestos.

Realizar pruebas a la solución analítica para evaluar su desempeño e identificar posibles errores.

1.3 Solución Propuesta

A continuación se plantea una solución a la problemática previamente identificada y detallada, dicha solución permite la unificación e integración de fuentes y tipos de datos, la alimentación del almacén de datos, elaboración de indicadores y el acceso a los mismos a través de herramientas de consulta y reportes.

Dicho almacén de datos contiene data histórica, de manera que permita la identificación de tendencias y asista la toma de decisiones de alto nivel gerencial, correspondientes a los indicadores de rentabilidad en sus diferentes presentaciones.

La herramienta de inteligencia de negocio, que da acceso fácil a los datos contenidos en el almacén, permite, entre otras cosas, la elaboración de consultas de manera sencilla y la construcción de reportes diversos, hechos de acuerdo con las necesidades de cada usuario.

En cuanto a la construcción de dicha solución, se utiliza como marco metodológico el método de Ralph Kimball, que permite el correcto desarrollo y mantenimiento de una solución analítica utilizando un almacén de datos como fuente unificada de datos, alimentada por procesos ETL que permitan la estandarización y unificación de datos procedentes de distintas fuentes.

De manera que se propone la siguiente arquitectura para la solución analítica (ver imagen 1):



Imagen 1: Arquitectura propuesta.

1.4 Justificación

Actualmente, se vive en una época en la cual la información es un factor clave para alcanzar la misión, visión y objetivos de una organización, en la que los ejecutivos requieren del acceso rápido y fácil de dicha información para la toma de decisiones, y lograr beneficios en un plazo de tiempo corto para mantener un desempeño funcional y óptimo dentro de la organización.

El escaso tiempo para el análisis de esta información, complica el hecho de tomar decisiones adecuadas en un entorno sujeto a constantes cambios, más aún cuando las organizaciones se ven enfrentadas a situaciones, en las que se deben tomar decisiones con acierto y celeridad, que determinan las futuras acciones a llevar a cabo. Contar con la información histórica y bien procesada de los procesos de negocio, permite comparar datos en varios períodos e identificar tendencias.

Por estas razones aunadas al análisis previo de la problemática planteada, la implementación de una solución de inteligencia de negocio, es una alternativa tecnológica que permite manejar la información para la toma de decisiones acertadas en todos los niveles de una organización, desde la extracción, depuración y transformación de datos, hasta la exploración y distribución de la información mediante herramientas de fácil uso para los usuarios, con el fin de satisfacer sus necesidades.

Además, hay que destacar que las soluciones de inteligencia de negocio les permiten a las personas encargadas de los negocios ser más productivas al:

- Minimizar el tiempo requerido para recolectar toda la información necesaria del negocio al evitar la ejecución de consultas directas sobre los sistemas transaccionales.
- Descubrir información no evidente a partir de los análisis realizados sobre los datos operativos.

- Evitar que se manipulen manualmente los datos, hecho que puede causar que el resultado de los análisis efectuados sea incorrecto.
- Permitir al usuario final realizar análisis de forma rápida, personalizada y fiable, en los que se obtenga la información que es requerida en un momento dado a través de cuadros de mando integral.

Asimismo, las instituciones financieras se ven beneficiadas de manera importante al dar el salto tecnológico que sugiere la utilización de la solución propuesta, identificando y aprovechando oportunidades y ventajas que sólo se hacen visibles luego del análisis efectivo de datos históricos.

CAPÍTULO II

MARCO CONCEPTUAL

2.1 Sistema Financiero

Para Díaz Mata & Hernández Almora (1999) el sistema financiero, es el conjunto de personas y organizaciones, tanto públicas como privadas, por medio de las cuales se captan, administran, regulan y dirigen los recursos financieros que se negocian entre los diversos agentes económicos, dentro del marco de la legislación correspondiente.

Dentro del sistema financiero coexisten varios tipos de instituciones y organizaciones. Instituciones regulatorias como el BCV en caso de Venezuela o la Reserva Federal en el caso de EE.UU, interactúan con otros tipos de organizaciones, como por ejemplo, instituciones bancarias, aseguradoras, casas de cambio y de bolsa, todas con actividades distintas y objetivos diversos dentro del sistema.

A pesar de su complejidad los sistemas financieros poseen funciones concretas, cuyo seguimiento sirve para evaluar e identificar sistemas financieros saludables.

2.1.1 Funciones del Sistema Financiero

Según Rodríguez & Capece (2001), los sistemas financieros tienen distintas funciones, todas importantes, pero según algunas de sus más básicas son:

- Obtener estabilidad monetaria, ya que el sistema financiero contribuye a salvaguardar el valor del dinero.
- Contribuir a las solvencias de las instituciones.
- Crear una variedad de activos financieros, con el fin de ofrecer un fácil acceso a las fuentes de financiación.
- Fomentar el excedente de los ingresos sobre los gastos de consumo, es decir, el ahorro, para capturarlo y convertirlo en recursos disponibles para la inversión.
- Lograr una eficaz asignación de los recursos. Es necesario que el sistema financiero seleccione las mejores oportunidades de inversión, garantizando que las inversiones potencialmente rentables no queden desasistidas.
- Conseguir un bajo costo de intermediación.

2.1.2 Importancia del Sistema Financiero

Incluso desde antes de que fuera definido como tal, los sistemas financieros han sido de vital importancia para el desarrollo y crecimiento de las poblaciones, dado que, como sugiere Dueñas Pietro (2008):

El Sistema Financiero opera como un intermediario entre las personas que disponen de suficiente capital y aquellas que necesitan y solicitan recursos monetarios para desarrollar proyectos de inversión e impulsar la actividad económica. (...) El Sistema Financiero facilita la circulación del dinero en la economía, permitiendo de esta manera la realización de numerosas transacciones diarias y fomentando de este modo el desarrollo de proyectos de inversión.

Sin embargo, la mejor manera de entender el funcionamiento de los sistemas financieros, es conocer sus componentes y sus funciones, para luego comprender sus interacciones.

2.1.3 Componentes del Sistema Financiero

Todo sistema financiero tiene tres componentes principales, cada uno con características propias (ver imagen 2), como son:

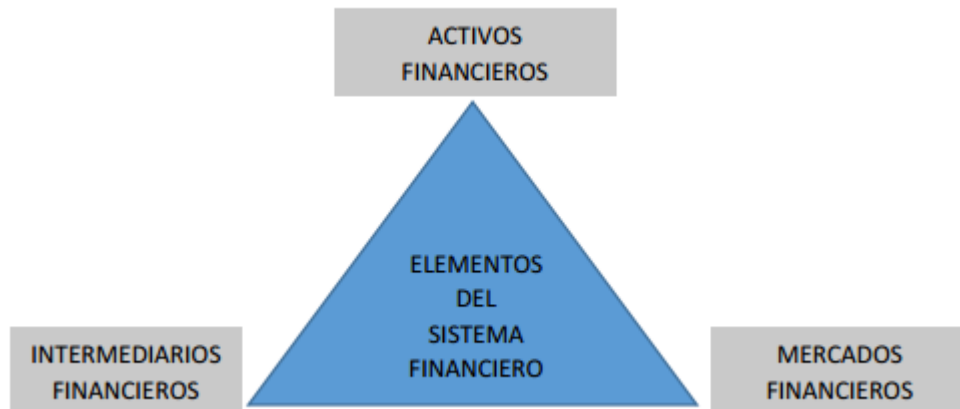


Imagen 2: Componentes del sistema financiero.

2.1.3.1 Activos financieros

Son títulos librados por las Unidades de Gastos Deficitarias (entidades que gastan más de lo que producen). Estos títulos son negociables y tienen tres características importantes (Dueñas Pietro, R., 2008):

- La liquidez, es decir, son de fácil conversión en efectivo.
- Bajo riesgo, es decir la posibilidad mínima de pérdida de valor.
- Rentabilidad, que es la ganancia sobre la inversión realizada.

2.1.3.2 Intermediarios financieros

Según Mochón y Beker (1998) los intermediarios financieros:

Son los entes que intervienen en el mercado atrayendo recursos monetarios, poniéndolos mediante operaciones crediticias y otras de tipo financiero. Comprende a personas o empresas que operan en forma pública o privada, en montos reducidos o considerables y que hacen de ello su actividad específica o no.

Los Intermediarios Financieros son los que como dice su nombre intermedian las operaciones entre los distintos participantes del mercado. Constituyen el conjunto de instituciones que intermedian entre los oferentes y demandantes de recursos financieros.

Los Intermediarios Financieros más importantes en todos los países son los bancos. Estos bancos y otras entidades financieras, son entidades que hacen de la intermediación pública de recursos financieros, su actividad específica y habitual, caracterizada por la captación de depósitos y el otorgamiento de créditos. Son ejemplos: Bancos Comerciales, Bancos de Inversión, Agentes de Bolsa, Compañías de Seguro entre otros.

2.1.3.3 Mercados Financieros

De acuerdo con Dueñas Pietro, R. (2008), un mercado financiero, es un sitio físico o virtual al cual concurren compradores y vendedores, donde se negocian los activos financieros. Las funciones de los mercados financieros son:

- Poner en contacto a los participantes
- Fijar los precios de los activos financieros
- Convertir en líquidos a los activos financieros.
- Reducir costos y plazos de intermediación.

2.2 Rentabilidad

Según Pablo García Estévez (2012):

La rentabilidad resulta de la relación entre el resultado contable (ingresos netos) y los recursos utilizados para obtenerlo, es decir que tan eficiente se es para producir ingresos teniendo una cantidad conocida de recursos (capital de inversionistas, activos, inversión externa, etc).

Dado que la rentabilidad puede ser calculada tomando distintos enfoques y cada uno aporta información sobre distintos aspectos del negocio, se utilizan estas diversas fuentes en conjunto para obtener una fotografía más amplia del desempeño de la organización.

Estas fuentes de información son, en sí, números, ya sean de obtención directa (capital invertido, volumen de dinero) o valores calculados (tasas, márgenes), y estos números nos ayudarán a medir, evaluar e impulsar el desarrollo de cada uno de los procesos de la organización. A dichos números nos referiremos de ahora en adelante como indicadores.

2.3 Indicadores

Una de las definiciones más utilizadas para indicadores pertenece al campo de las ciencias sociales, y aunque los contextos son distintos parece pertinente en este caso, ya que según (Bauer, 1966):

“Los indicadores sociales (...) son estadísticas, serie estadística o cualquier forma de indicación que nos facilita estudiar dónde estamos y hacia dónde nos dirigimos con respecto a determinados objetivos y metas, así como evaluar programas específicos y determinar su impacto”

También según (Marianela Armijo,2011):

“Un indicador es un instrumento que nos provee evidencia cuantitativa acerca de si una determinada condición existe o si ciertos resultados han sido logrados”

Es decir, que un indicador nos permite medir, evaluar y monitorear diferentes aspectos de un evento, situación o proceso, proporcionando información concreta sobre distintas áreas de lo antes mencionado.

Sin embargo no todos los indicadores tienen las mismas características o naturaleza y pueden ser clasificados por el tipo de información que brindan.

2.3.1 Tipos de indicadores

De acuerdo con Armijo (2011) los indicadores se dividen en:

- **Indicadores de eficacia:** Se refieren al Grado de cumplimiento de los objetivos de la organización, sin referirse al costo de los mismos.
- **Indicadores de eficiencia:** Proveen información sobre la relación ente dos magnitudes. Ejemplo: cantidad de recursos utilizados en la elaboración de determinada cantidad de productos. Los indicadores de eficiencia, en el contexto financiero, miden la relación de objetivos alcanzados entre recursos y costos generados, es decir, miden la habilidad de una persona o institución para obtener ingresos, generando costos bajos o aceptables.
- **Efectividad:** Dedicados a medir la capacidad de la organización para responder de manera correcta y rápida a las necesidades. Es decir, la relación entre eficiencia y eficacia

En el caso de la rentabilidad existen varios y muy diversos indicadores, en su mayoría de eficiencia, que nos permiten evaluar la viabilidad de, por ejemplo, la colocación de cajeros automáticos en distintos lugares, la eficiencia en la utilización de capital invertido o simplemente medir las ventajas económicas obtenidas a través de la utilización de determinados servicios por parte del cliente, como se muestra a continuación.

Rentabilidad sobre activos (ROA)

La rentabilidad sobre los activos o rentabilidad económica según Sánchez Ballesta (2002) es una medida de la capacidad de los activos de una empresa para generar valor con independencia de cómo han sido financiados, lo que permite la comparación de la rentabilidad entre empresas sin que la diferencia en las distintas estructuras financieras, puesta de manifiesto en el pago de intereses, afecte al valor de la rentabilidad.

Es decir, que el ROA es la relación entre lo invertido en un activo, y el retorno monetario que genera dicho activo, lo que da una idea de que tan eficiente es la organización en cuanto a generación de ingresos, y se calcula de la siguiente manera:

$$ROA = \frac{\text{Resultados netos sin intereses o impuestos} *}{\text{Valor medio del activo}}$$

*Definiendo resultados netos como: Ingresos – costos

Rentabilidad sobre capital propio (ROE)

Es una medida, referida a un determinado periodo de tiempo, del rendimiento obtenido por esos capitales propios, generalmente con independencia de la distribución del resultado. La rentabilidad financiera puede considerarse así una medida de rentabilidad más cercana a los accionistas o propietarios que la rentabilidad económica (Sánchez Ballesta, Juan Pedro, 2002).

En concreto, el ROE es tanto como la relación que existe entre el ingreso neto y el capital invertido en la organización por los accionistas. De manera que evidencia eficiencia al administrar y utilizar los recursos suministrados por los accionistas, y es uno de los indicadores que se busca maximizar en la mayoría de las organizaciones dado que está muy ligado al financiamiento.

En cuanto al cálculo del ROE, en forma resumida es como sigue:

$$ROE = \frac{\text{Resultados netos sin intereses o impuestos} *}{\text{Fondos propios en estado medio}}$$

*Definiendo resultados netos como: Ingresos - costos

Tanto éstos como el resto de los indicadores de rentabilidad, referenciados más adelante, se obtienen luego del correcto procesamiento de los datos a través de sistemas de información.

2.4 Sistema de Información

Un sistema de información (SI), según Laudon y Laudon (2012) es un conjunto de componentes interrelacionados que recolecta (o recupera), procesa, almacena y distribuye información para apoyar los procesos de toma de decisiones y de control en una organización.

Dichos sistemas contienen (y generan) información sobre personas, lugares o elementos importantes para la organización o su entorno, de manera que visto de manera general un sistema de información posee, en esencia, una entrada de datos, los elementos encargados de procesar y transformar dichos datos, y una salida de información pertinente a los diversos niveles organizacionales del negocio (Ver imagen 3).

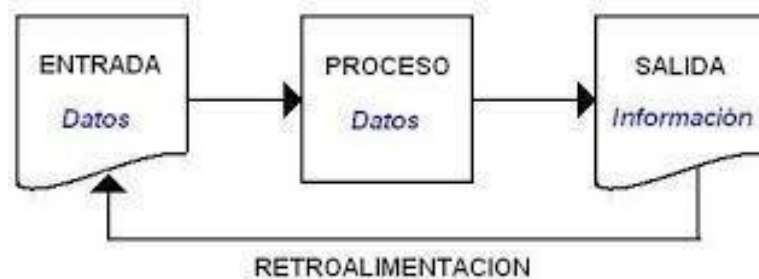


Imagen 3: Vista general de un sistema de información.

Al adentrarse en lo que se refiere a las partes de los sistemas de información, según Laudon, Laudon (2012) existen tres partes, o actividades principales que componen un sistema de información, la entrada, el procesamiento y la salida (ver imagen 4).

La **entrada** se refiere a la captura o recolección de los datos crudos provenientes de la organización misma o su entorno de negocios, dichos datos aún no aportan soporte significativo a la toma de decisiones en ninguno de los niveles.

La etapa de **procesamiento** es la que se encarga de transformar los datos obtenidos a través de la entrada y, a partir de estos, generar información a través de indicadores significativos para los usuarios finales de los resultados obtenidos. Dando un formato manejable a la información suministrada.

La tercera etapa corresponde a la **salida** o distribución de la información obtenida a través del procesamiento hacia las personas o actividades que harán uso de la misma. Procurando la distribución adecuada de dicha información se procura un análisis posterior más eficiente y una toma de decisiones más rápida y acertada.

Otro aspecto de la salida de datos es la **retroalimentación** del SI, que no es más que la salida de datos que llega a los actores correspondientes dentro de la organización de manera que les permita realizar una evaluación acertada sobre los procesos de entrada y procesamiento de dato.

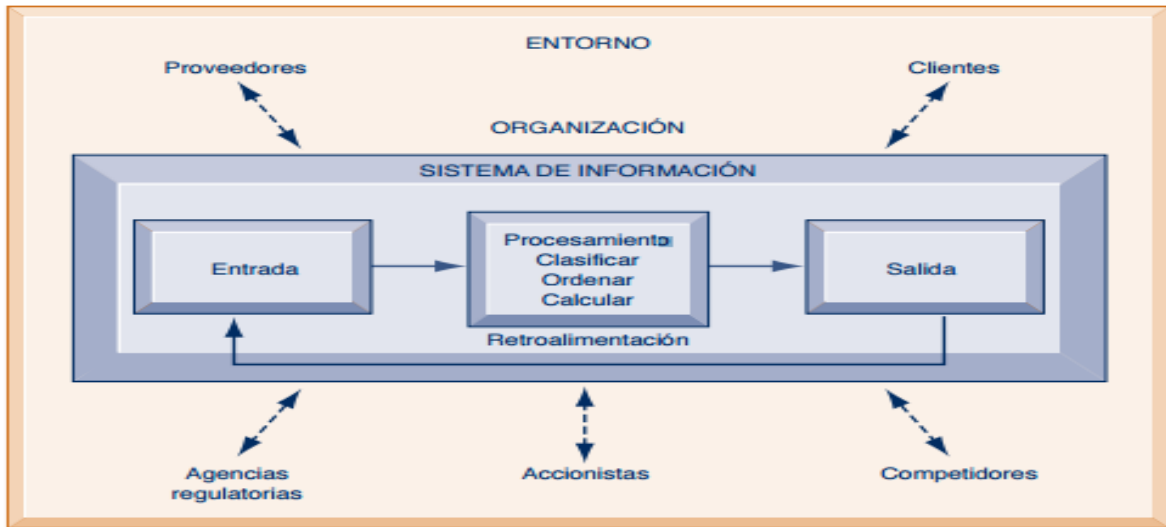


Imagen 4: Modelo de retroalimentación de un SI.

Tipos de sistemas de información

En lo que se refiere a sistemas de información existen varias clasificaciones, diferenciadas principalmente por el enfoque que se les dan a las actividades del negocio realizadas por estos.

A continuación se presentan dos clasificaciones de sistemas de información (SI), la primera (Imagen 5) responde a la pirámide organizacional y las áreas del negocio a las que brindan soporte, y la segunda los clasifica de acuerdo a la manera en la que procesan los datos.

La clasificación basada en la pirámide organizacional clasifica los SI de acuerdo al nivel de la organización en el que operan y que tipo de toma de decisiones apoyan.



Las organizaciones de negocios son jerarquías que consisten en tres niveles principales: gerencia de nivel superior, gerencia de nivel medio y gerencia operacional. Los sistemas de información dan servicio a cada uno de estos niveles. A menudo, los científicos y los trabajadores del conocimiento trabajan con la gerencia de nivel medio.

Imagen 5: La pirámide organizacional

Desde la base de la pirámide hacia arriba encontramos en primer lugar a la gerencia operacional, que hace uso de los TPS para dar soporte a sus actividades diarias, que contemplan gran número de acciones repetitivas.

TPS (Transaction processing system)

Para Laudon & Laudon (2012) un TPS es un sistema computarizado que efectúa y registra las transacciones diarias de rutina necesarias para realizar negocios, como introducir pedidos de ventas, reservaciones de hoteles, nómina, registro de empleados y envíos.

El principal propósito de los sistemas en este nivel es responder a las preguntas de rutina y rastrear el flujo de transacciones por toda la organización. ¿Cuántas piezas están en el inventario? ¿Qué ocurrió con el pago de ayer? Para responder dichas preguntas la información debe estar actualizada y ser de fácil acceso.

En el nivel operacional, las tareas, recursos y metas están predefinidos y muy estructurados, dado que cada uno de dichos aspectos se encuentra especificado en las reglas del negocio (Índice de precios, criterios de otorgamiento de créditos).

En la imagen 6 se observa un modelo de ejemplo de un TPS orientado al área del pago y consulta de nóminas:

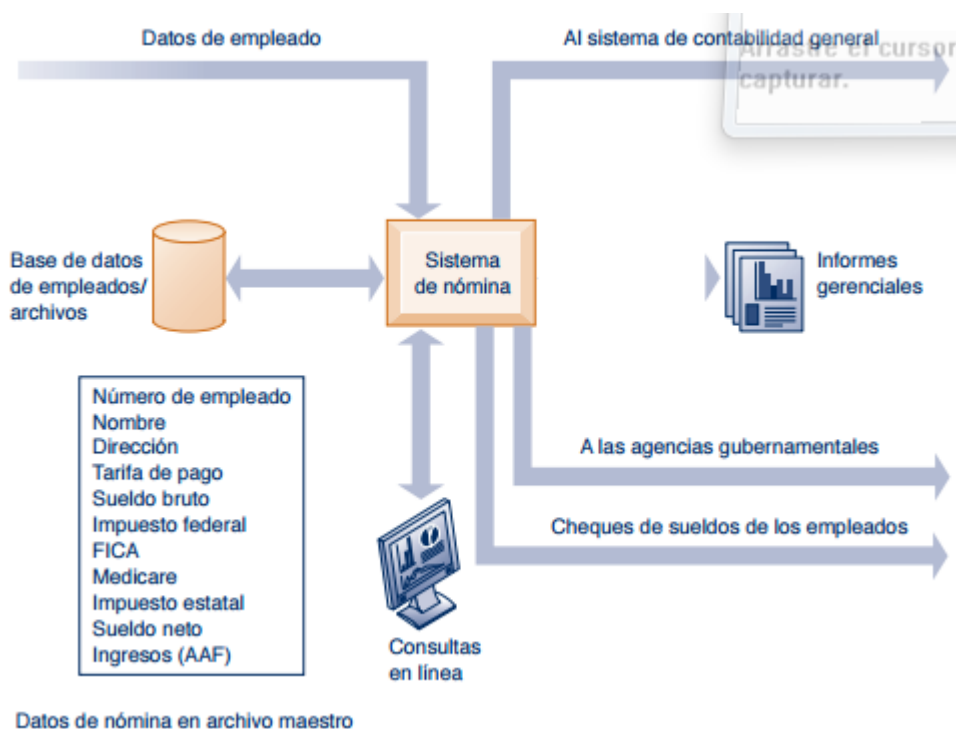


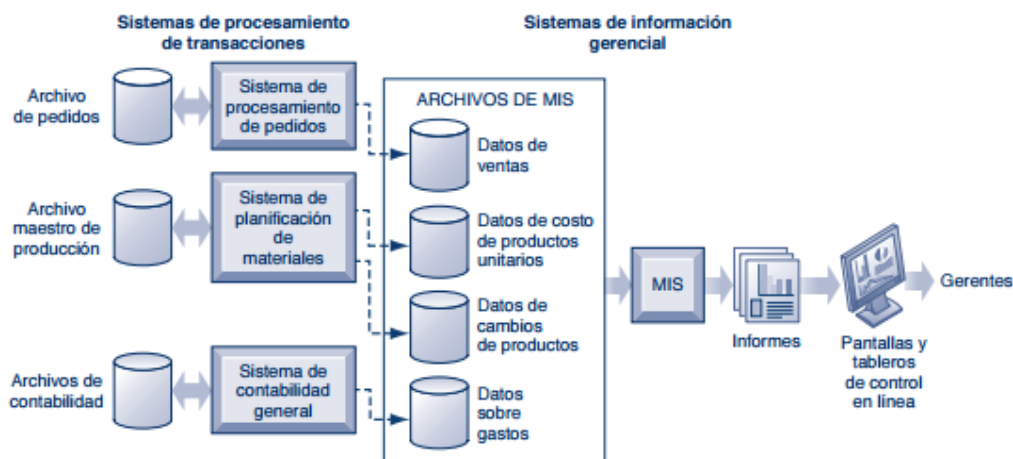
Imagen 6: TPS orientado a pagos y nómina.

En cuanto a la gerencia de nivel medio, las necesidades son distintas a las atendidas por los TPS, este nivel organizacional requiere herramientas que faciliten el monitoreo y control de los procesos de negocio de bajo nivel, así como la toma de decisiones oportuna.

MIS (Management Information System)

Los MIS son los encargados de proveer las herramientas necesarias a los gerentes de nivel medio de modo que, a través de reportes se pueda supervisar el desempeño de las actividades básicas de una organización, como ventas, transacciones bancarias, utilización de recursos, etc. Laudon & Laudon (2012).

La gráfica a continuación (Imagen 7) describe cómo un MIS toma datos de uno o varios TPS para elaborar reportes de desempeño:



En el sistema que se ilustra en este diagrama, tres TPS suministran datos de transacciones sintetizados al sistema de informes del MIS al final del periodo de tiempo. Los gerentes obtienen acceso a los datos de la organización por medio del MIS, el cual les provee los informes apropiados.

Imagen 7: Modelo de MIS

Dado que los MIS generan información con respecto a actividades básicas del negocio, éstos dan soporte a la toma de decisiones rutinarias, repetitivas y con poca variabilidad.

DSS (Decision Support System)

En contraste con los MIS y orientados a prestar apoyo a los niveles de gerencia media-alta se encuentran los DSS, dan apoyo directo a los directivos, durante el proceso de toma de decisiones y se utilizan para analizar posibles escenarios (Obrien, 2006).

Los DSS no dan soporte a decisiones rutinarias, sino a requerimientos a demanda o “Ad hoc” y consolidan información generada por los TPS y MIS, así como fuentes de datos externas, para que los usuarios puedan trabajar directamente sobre dicha información, previa aplicación de modelos analíticos y formulación de indicadores.

A continuación (Imagen 8) un modelo que representa un DSS utilizado para la estimación y proyección de viajes y transporte:

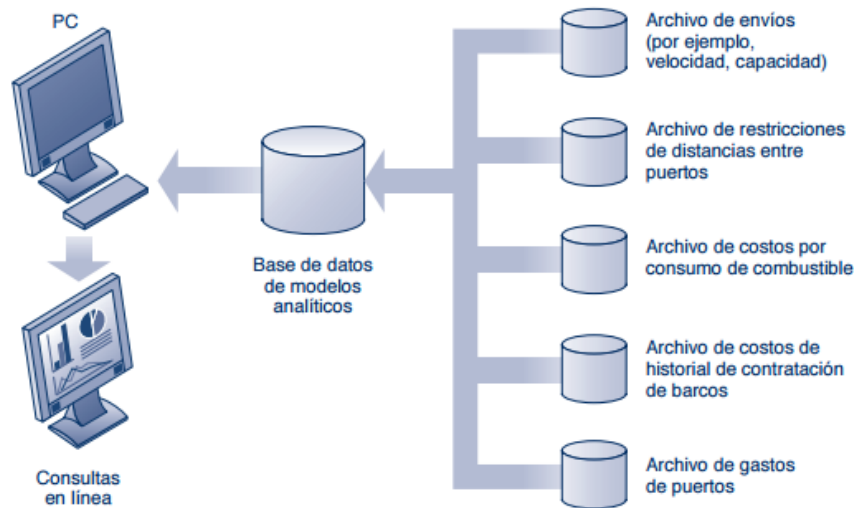


Imagen 8: Modelo de DSS

ESS o EIS (Executive Support Systems)

Pasando la gerencia de nivel medio y orientados hacia la alta gerencia se ubican los ESS también conocidos como EIS, que según Obrien (2006) dan soporte a las decisiones de la parte más alta de la pirámide organizacional y que aparte de no ser rutinarias requieren una perspectiva amplia, no solo de la organización sino de su entorno, así como la realización de proyecciones y estimaciones.

Dichos sistemas tienen muchas y diversas fuentes de información, desde bases de datos operacionales (TPS), pasando por sistemas analíticos de menor nivel (MIS, DSS) e incluso fuentes externas como noticias, indicadores financieros, leyes fiscales y redes sociales.

A menudo este gran volumen de información es resumido y presentado a los usuarios a través de un tablero de control o dashboard, para contribuir a la elaboración de la perspectiva general de la organización y su desempeño. A continuación (Imagen 9) un ejemplo de dicho tablero:

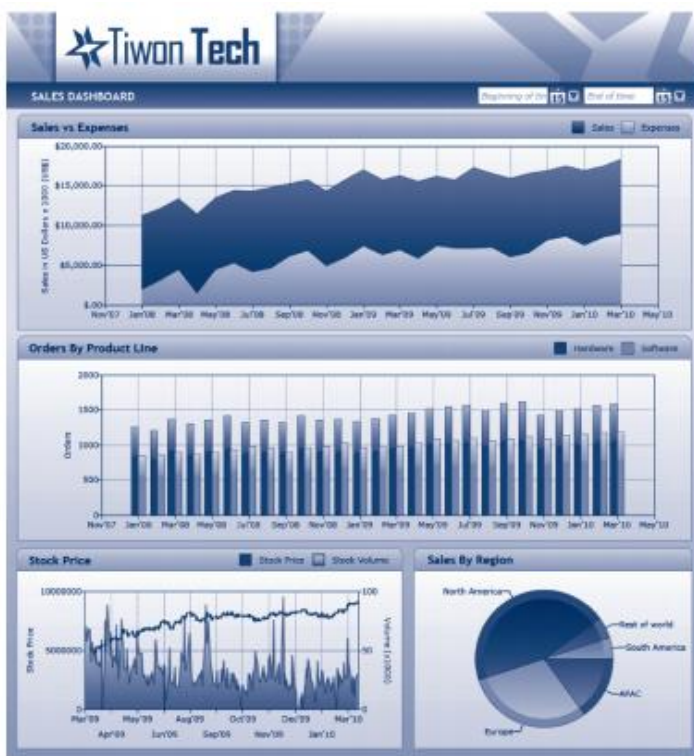


Imagen 9: Tablero de control común en los ESS/EIS

Pero como se mencionó anteriormente, la anterior no es la única manera de clasificar los sistemas de información, dado que pueden ser clasificados también de acuerdo a la manera en que se almacenan y procesan los datos.

De tal manera que obtenemos dos tipos de sistemas de información, los **OLTP**, dedicados a la actividad transaccional del negocio, y los **OLAP**, dedicados a asistir a la actividad gerencial y de toma de decisiones.

1. Sistemas de procesamiento de transacciones en línea (OLTP)

Los sistemas OLTP (On-Line Transaction Processing) son, según Obrien (2006), sistemas que brindan apoyo a aplicaciones orientadas a transacciones, típicamente son utilizados para el ingreso y almacenamiento de las transacciones asociadas a los procesos de negocios, i.e, ventas, liquidaciones de crédito, facturación, etc.

Suelen ser utilizados por organizaciones que poseen una red de comunicación distribuida, por lo que trabajan de modo cliente/servidor.

También son llamadas bases de datos dinámicas y numerosas, ya que la información se modifica en tiempo real, es decir, se insertan, se eliminan, se modifican y se consultan datos en línea durante el tiempo de operación del sistema.

Un ejemplo es el sistema transaccional bancario, en el que puede registrarse la liquidación de un crédito y alterar el saldo de la cuenta beneficiaria de manera inmediata.

Características de los sistemas OLTP

- Manejan poco volumen de información.
- Realiza transacciones rápidas, repetitivas.
- Realiza gran cantidad de transacciones.
- Debe cumplir las propiedades A.C.I.D:
 - Atomicidad: Las transacciones no pueden quedar a medias. En caso de aborto de una transacción debe revertirse todo cambio hecho por la misma.
 - Consistencia: Los datos siempre deben ser válidos, por lo que en caso de aborto, debe reestablecerse el último estado válido de la base de datos.
 - Independencia: Los efectos de una transacción no deben ser visibles para ninguna otra transacción hasta que la primera haya concluido de manera exitosa.
 - Durabilidad: Cada cambio realizado por una transacción que concluye de manera exitosa debe ser perdurable en el tiempo.

Aplicaciones de los sistemas OLTP

Los sistemas OLTP poseen numerosas y diversas aplicaciones, dado que registran transacciones de todo tipo, tienen cabida en la gran mayoría de las organizaciones existentes, sean instituciones bancarias o financieras, tiendas o redes de tiendas, registro de encuestas, mantenimiento de inventario, entre muchas otras aplicaciones.

A continuación (Imagen 10) un gráfico que ejemplifica un OLTP orientado al registro de pago por servicios televisivos (Pay per view):

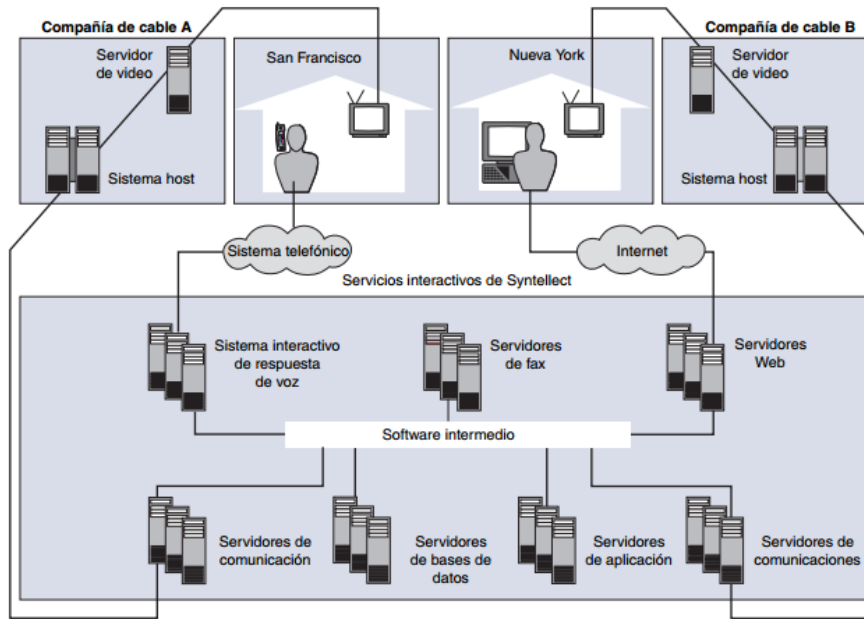


Imagen 10: Modelo de sistema OLTP

2. Sistemas de procesamiento analítico en línea (OLAP)

Los sistemas OLAP, como sugiere Obrien (2006), están orientados a brindar apoyo a las actividades estratégicas y gerenciales del negocio, de manera que no están ideados para registrar gran cantidad de transacciones, sino para almacenar información eficientemente y en cantidades suficientes para que se puedan generar reportes e indicadores que respondan a las preguntas generadas por el negocio y su entorno.

El procesamiento analítico en línea permite a los administradores y analistas, analizar y manipular en forma interactiva grandes cantidades de datos detallados y consolidados desde muchas perspectivas. El procesamiento analítico en línea implica el análisis de relaciones complejas entre miles o incluso millones de elementos de datos almacenados en mercados de datos, almacenes de datos y otras bases de datos multidimensionales para descubrir patrones, tendencias y condiciones de excepción. En la imagen 11 se describe ligeramente la arquitectura general de los sistemas OLAP.

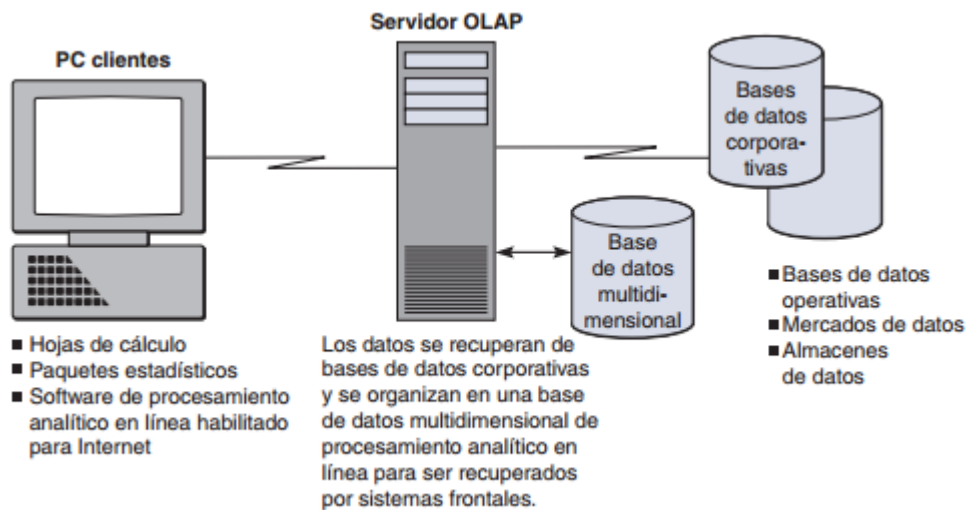


Imagen 11: Modelo de sistema OLAP

Características de los sistemas OLAP

- Manejan gran volumen de información.
- Están diseñados para optimizar el tiempo de acceso a los datos a pesar de su gran volumen.
- Realiza consultas formuladas sobre la marcha y con características distintas entre ellas.
- Realiza pocas consultas a la vez.
- Permite realizar operaciones de agregación o desglose de manera rápida y sencilla, permitiendo varias vistas distintas del mismo universo de datos.

Aplicaciones de los sistemas OLAP

Si existe un tema clave para el éxito de cualquier organización, es el de la planificación, poder proyectar el comportamiento del entorno del negocio y utilizar esta información para adelantarse a los hechos y estar prevenidos ante el cambio, es lo que diferencia a las organizaciones de las grandes organizaciones.

Los sistemas OLAP tienen aplicaciones en todos los niveles de gerencia y planificación estratégica, sea una organización de comercio electrónico o una compañía transnacional de seguros, todas pueden verse beneficiadas por los sistemas OLAP, tomando en cuenta que permiten el acceso oportuno y rápido a la información, además de proveer mecanismos para elaborar reportes desde diversos puntos de vista, partiendo desde un mismo conjunto de datos, lo que da a la gerencia de la organización una perspectiva muy amplia de las actividades del negocio, el comportamiento del entorno y finalmente les permite planificar, prever y corregir aspectos claves de desempeño de la organización.

Como ejemplo una vista al tablero de control de una organización con diversos puntos de vista del entorno y sus actores (Imagen 12):

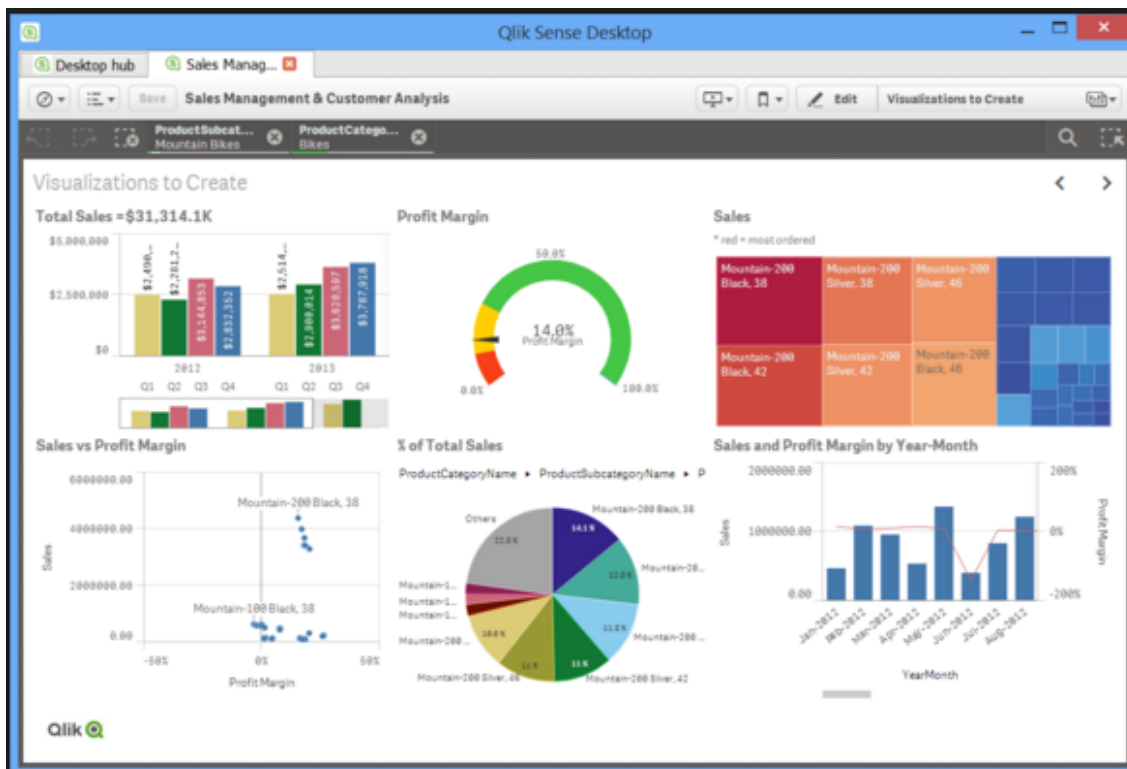


Imagen 12: Ejemplo de cuadro de mando (Dashboard)

Habiendo mencionado las características y ventajas de los sistemas OLAP, luce natural la relación entre estos y la toma de decisiones, tanto reactivas como preventivas, tomando en cuenta el entorno de la organización que los utiliza. Dicho proceso de análisis de información, generación de conclusiones y toma de decisiones se conoce como inteligencia de negocio.

2.5 Inteligencia de negocio

En la actualidad es evidente que la cantidad de datos en el entorno de negocio aumenta de manera acelerada, no así nuestra capacidad para analizarlos de manera eficiente y transformarlos en información. Nace así la necesidad de desarrollar métodos o procesos que agilicen y simplifiquen el proceso de transformación y análisis de dichos datos, sin importar la naturaleza del entorno de negocio, sea una fábrica, una organización política o una entidad financiera, la necesidad de información siempre estará presente y será cada vez mayor.

De lo antes mencionado surge la Inteligencia de negocio que, según Cano (2008), no es más que:

“(…) un proceso interactivo para explorar y analizar información estructurada sobre un área (normalmente almacenada en un datawarehouse), para descubrir tendencias o patrones, a partir de los cuales derivar ideas y extraer conclusiones.

El proceso de Business Intelligence incluye la comunicación de los descubrimientos y efectuar los cambios.

Las áreas incluyen clientes, proveedores, productos, servicios y competidores.”

Pero una solución de inteligencia de negocio comprende mucho más que el almacenamiento especialmente estructurado de los datos (datawarehouse), comprende una arquitectura entera que abarca desde las fuentes de datos hasta la manera en la que se muestran los mismos.

2.5.1 Arquitectura de una Solución analítica

La arquitectura de una solución analítica viene determinada por una situación central como fuente de información para las herramientas de análisis. La estructura básica (Ilustración 10) se esquematiza con los siguientes elementos:

- Fuentes de datos.
- Procesos de extracción, transformación y carga (ETL en inglés).
- Almacén de datos (datawarehouse)
- Herramientas de presentación de datos

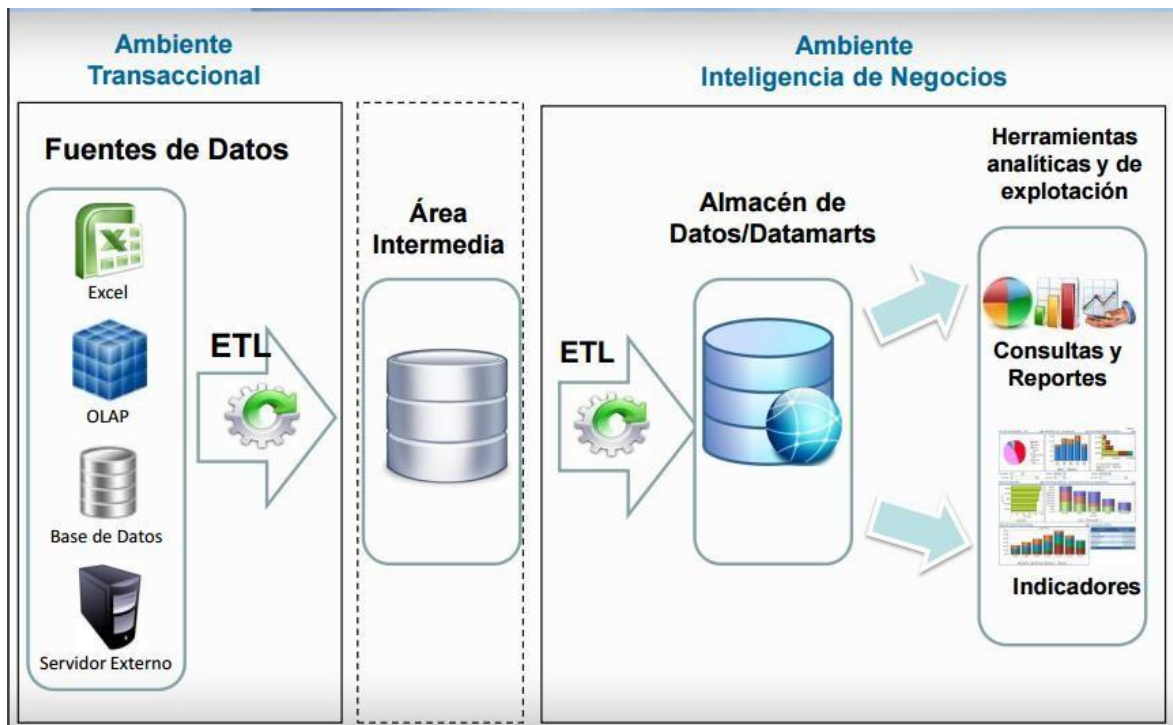


Imagen 13: Arquitectura de una solución de BI

- **Fuentes de datos**

Las fuentes de datos son de vital importancia para el correcto funcionamiento de una solución de inteligencia de negocio, según (Cano, 2008) entre los factores principales que aportan complejidad a la carga de datos se encuentra la diversidad en cuanto a la naturaleza de dichas fuentes, dado que en muchos lugares, o en la mayoría, se manejan varias bases de datos de distintos tipos, así como distintos sistemas que las alimentan. Sistemas de tipo ERP, CRM o de facturación pueden alimentar distintas bases de datos dentro de la misma organización.

Dicha convivencia de sistemas suma a la complejidad de tratamiento de los datos que generan, puesto que muchas veces no son sistemas integrados y la posterior integración y estandarización de los datos puede dificultarse en cuanto se utilicen distintos sistemas manejadores de bases de datos, hojas de cálculo o incluso texto plano para almacenar las transacciones realizadas.

Siendo así, se hace evidente que es necesario que existan procesos que faciliten la revisión, integración, transformación y carga de dichos datos en el datawarehouse.

- **Extracción, transformación y carga (ETL)**

Los procesos ETL son los responsables de la integración y almacenamiento exitosos de los datos provenientes de las fuentes. Dichos procesos son de gran importancia y por lo general ocupan entre el 60% y 80% del tiempo de un proyecto de inteligencia de negocio.

Por lo general se utiliza un área de preparación de datos (staging area), que no es más que una base de datos temporal, para realizar las alteraciones necesarias sobre los datos.

Los tres pasos que comprenden los procesos ETL según (Cano, 2008) son:

Extracción: se refiere a la obtención física de los datos almacenados en las distintas fuentes.

Transformación: es el proceso en el que se realizan todas las alteraciones necesarias sobre los datos crudos obtenidos durante la extracción. Comprende la unificación de fuentes y tipos de datos, eliminación de duplicados, sustitución de campos vacíos y resolución de problemas de consistencia. También dentro de este paso se realiza la agrupación de registros y el cálculo de indicadores.

Carga e integración: comprende los procesos de carga correcta de los datos limpios y correctamente transformados. Esta parte del proceso posee dos partes, la carga inicial, que es la primera carga masiva de datos al datawarehouse y la fase de actualizaciones, que permite añadir nuevos registros y estructuras al almacén.

- **Almacén de datos**

Según (Inmon, 2002) un datawarehouse es una colección de datos integrada, no volátil, orientada a un área o proceso específico, variable en el tiempo, que además, está destinada a dar soporte a procesos de toma de decisiones.

Integración: Siendo la más importante de las características de un datawarehouse la integración se refiere a consistencia en cuanto a tipo, valor y formato de los datos que son introducidos al almacén de datos, siendo que los mismos provienen de diversas fuentes, no siempre relacionadas y por tanto la posibilidad de la presencia de inconsistencias es alta.

No volatilidad: No se refiere en si a la permanencia de los datos en la memoria de, en este caso, el servidor, sino a que al momento de la carga de datos (que por lo general se hace en masa) se guarda una fotografía de un instante de tiempo determinado y a diferencia de los datos operacionales que son actualizados y editados con frecuencia, los datos contenidos en un almacén de datos sufre poco o ningún cambio, de manera que al realizarse nuevas cargas se obtiene un registro histórico de los hechos registrados por las fuentes de datos operacionales.

Orientación a un área o proceso: En general, los almacenes de datos se concentran en las divisiones o departamentos principales de una organización, así como también en los procesos de negocio principales de la misma.

De tal manera los almacenes de datos se concentran en captar los datos asociados a dichas divisiones o procesos desde las fuentes de datos operacionales.

Variabilidad temporal: Se refiere a la exactitud con la que se refleja un instante de tiempo a través de los registros almacenados en el datawarehouse, a través de la identificación temporal (marcas de tiempo, fechas de registro, etc).

Dichas marcas temporales permiten conocer el instante de tiempo al cual hacen referencia los datos contenidos en un registro y determinan la validez de los mismos con respecto al intervalo, o intervalos, de tiempo que se desean consultar.

Definido ya el almacén de datos es necesario mencionar las características internas del mismo, dado que es una base de datos particular, tiene una estructura muy específica que le permite realizar sus actividades y que se detalla a continuación.

Modelo Dimensional

Según (Kimball, 1996), el modelo dimensional de un almacén de datos, es una técnica de diseño lógico la cual tiene por fin presentar los datos de manera estándar e intuitiva permitiendo a su vez un acceso de alto rendimiento a esos datos; usada para la construcción de los almacenes de datos.

Para modelar un almacén de datos hay que tomar en cuenta la información que se desea mostrar en los reportes o indicadores que posteriormente serán analizados por los usuarios finales. Por lo tanto, para implementar un almacén de datos se debe comenzar por hacer un modelo dimensional, ya que a través de éste se identifican cuáles son los hechos que se desean medir y desde cuáles perspectivas se desean agrupar los mismos. En otras palabras, un modelo dimensional consta básicamente de una o varias tablas de hechos asociadas a varias dimensiones formando con ello un esquema que puede tener la siguiente estructura:

- ✓ **Esquema Estrella:** Es el esquema de modelado más simple. En este diseño la tabla de hechos está rodeada por dimensiones y juntos forman una estructura que permite implementar mecanismos básicos para poder utilizarla con una herramienta de consultas OLAP (Imagen 14). La tabla de hechos (fact table) representa la ocurrencia de los mismos y las dimensiones proveen de contexto a dichos hechos e indicadores de manera de que se puedan ubicar en el tiempo y espacio correctos.

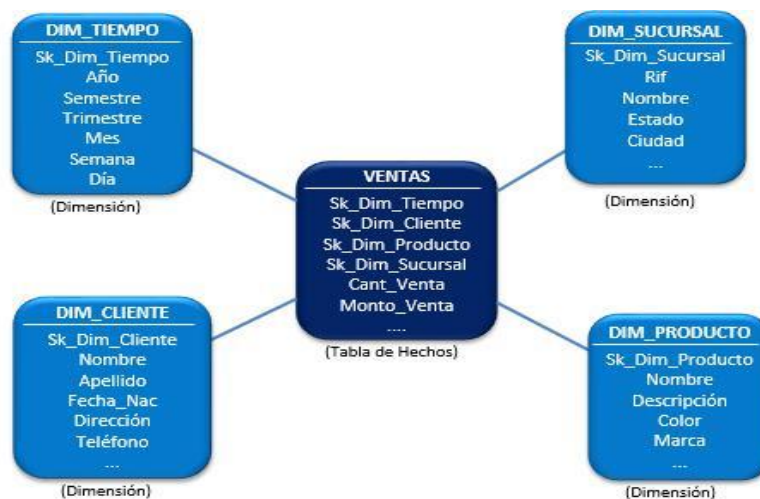


Imagen 14: Esquema estrella

- ✓ **Esquema Copo de Nieve:** Alternativa al esquema estrella con la notable diferencia de que las tablas dimensionales están normalizadas, por lo que se asemejan a las bases de datos relacionales (Imagen 15).

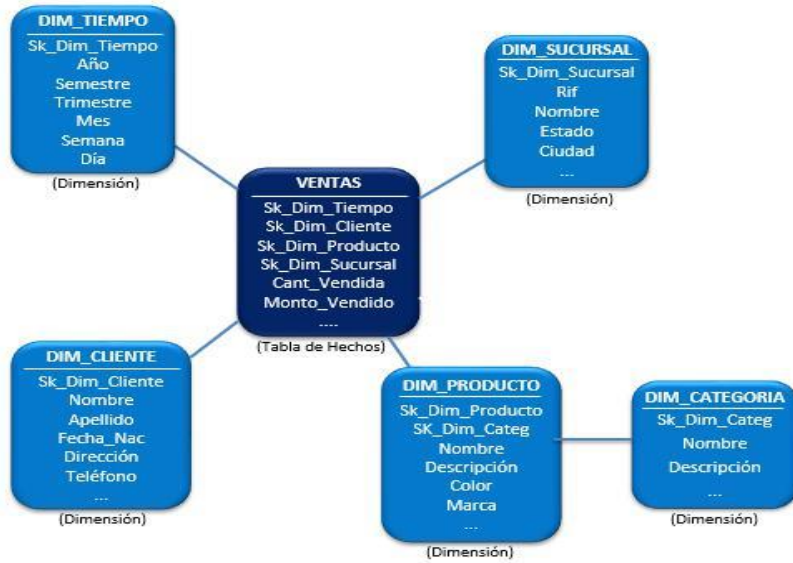


Imagen 15: Esquema de copo de nieve

- ✓ **Esquema de Constelación de Estrellas:** Es una combinación del esquema de estrella y un esquema de copo de nieve. Este esquema es más complejo que el estrella o copo de nieve debido a que contiene más de una tabla de hechos, de manera que existirán dimensiones que contextualicen varias tablas de hechos, aprovechando la representación normalizada de jerarquías cuando sea necesario (Imagen 16).

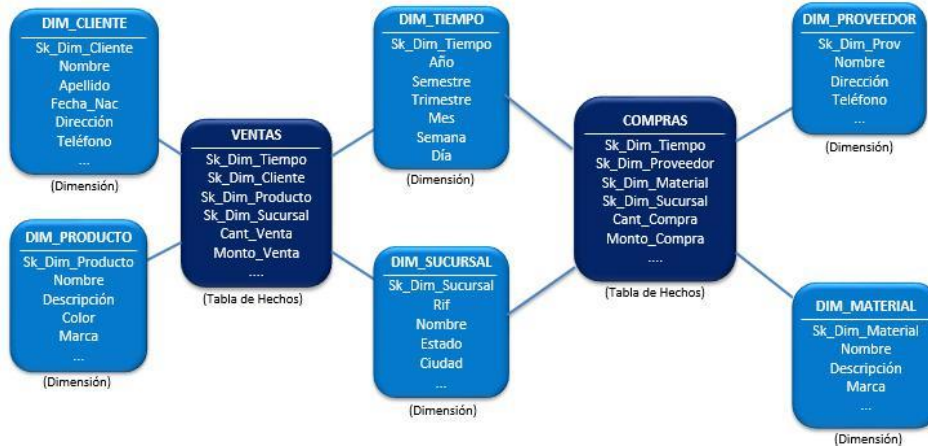


Imagen 16: Esquema de constelación de hechos

- **Herramientas de explotación y visualización de datos**

Son piezas de software o suites que permiten a los usuarios finales hacer uso de los datos ya contenidos en el almacén de datos, de manera que dichos datos puedan ser presentados de distintas formas, como:

- **Reportes estructurados:** reportes elaborados con una estructura fija, que responden a preguntas usuales y recurrentes del negocio, como por ejemplo, cierres de inventario diarios o reportes consolidados de ventas.
- **Consultas ad-hoc:** consultas inusuales que responden a preguntas no comunes del negocio. Por lo general, estas consultas deben ser realizadas con rapidez y eficacia, de manera que la solución analítica tiene formas de permitir que los usuarios finales las estructuren y ejecuten.
- **Cuadros de mando o dashboards:** Son grandes pantallas de resumen que presentan gran cantidad de indicadores con métodos que permiten su rápida interpretación como, por ejemplo, gráficos circulares, de barras, histogramas o números con escala de colores.

2.5.2 Herramientas para la Construcción de una Solución Analítica

Según (Martins, 2014), son un tipo de software diseñado para colaborar en los procesos de negocio de las organizaciones., específicamente se trata de herramientas que asisten en el análisis y la presentación de los datos. Este conjunto de herramientas tienen en común las siguientes características:

Accesibilidad a la Información: Los datos son la fuente principal de este concepto. Lo primero que debe garantizar este tipo de herramientas y técnicas es el acceso de los usuarios a los datos, con independencia de la procedencia de éstos.

Apoyo en la Toma de Decisiones: Se busca seleccionar y manipular sólo aquellos datos que interesen a los usuarios para que sirvan de base a las decisiones tomadas.

Orientación al Usuario Final: Se busca independencia entre los conocimientos técnicos de los usuarios y su capacidad para utilizar estas herramientas.

Como herramientas para el desarrollo de la solución analítica se eligen tres (3):

- Pentaho Data Integration:** Como herramienta esencial para el desarrollo de los procesos ETL que además de unificar las distintas fuentes de datos, permite la transformación de dichos datos, dando calidad a los mismos desde la estandarización y unificación de formatos, tipos de datos, etc. Cuenta con una interfaz gráfica sencilla y brinda buena interacción con muchas y diversas entradas o fuentes de datos, como hojas de cálculo, texto plano y diversos manejadores de bases de datos.

- Microsoft SQL Server:** Sistema manejador de bases de datos de Microsoft, elegido en gran medida por conveniencia, dado que la institución financiera en cuestión lo utiliza como fuente de datos para reportes de alta gerencia, dado que tiene buena integración con otros productos de Microsoft como Excel.

- Oracle Business Intelligence:** Herramienta orientada a la extracción de información por parte de los usuarios finales para realizar reportes que asistan a la toma de decisiones. Puede funcionar con diversos manejadores de bases de datos, los reportes proveen diferentes perspectivas y tienen características de personalización únicas. También permite la construcción de dashboards o cuadros de mando para integrar diferentes reportes y permite añadir filtros a los mismos.

CAPÍTULO III

MARCO METODOLÓGICO

El desarrollo de una solución analítica requiere de un método bien estructurado y de la correcta utilización de técnicas y tecnología, para así garantizar su correcto desempeño y la generación de resultados confiables y consistentes.

En el marco de lo dicho anteriormente, Ralph Kimball en 1998 define, a través del diagrama presentado a continuación, una secuencia de tareas de alto nivel requeridas para el diseño, desarrollo y despliegue efectivos del almacén de datos en el contexto de una solución analítica.

Dicho proceso toma en cuenta cada fase del desarrollo del almacén de datos, desde su conceptualización hasta su mantenimiento, pasando por su implantación

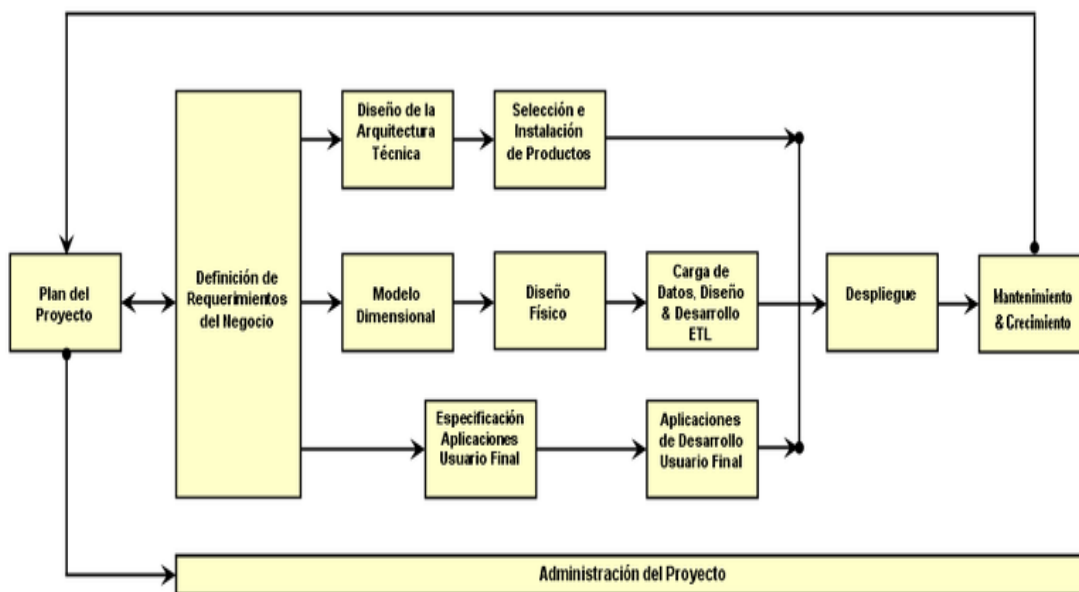


Imagen 17: Ciclo de vida dimensional de Kimball

3.1 Planificación del proyecto

Se refiere a la clara definición de varios temas importantes relacionados con el correcto desarrollo del proyecto, desde los requerimientos de personal (nivel de destreza, experiencia, cantidad), pasando por el financiamiento, su justificación y la asignación, duración y secuenciación de las tareas a realizar.



Definición del Proyecto: En el caso de la definición, existen factores identificados por el autor, que diferencian a los proyectos fallidos de los que han logrado seguir en pie. Estos factores son:

- Poseer un aliado empresarial comprometido con el proyecto. Su papel principal será de conciliador entre el proyecto y la organización.

- La existencia de una fuerte motivación organizacional por la construcción del almacén o bodega de datos.

- La factibilidad y disponibilidad los de datos. Uso de tecnologías de información.



Definición del Alcance: Una vez que se ha logrado un acuerdo con la organización, es el momento para establecer los límites del proyecto. La determinación del alcance requiere la participación conjunta tanto del departamento de tecnología de la información, como el de gestión empresarial.

- **Justificación:** Comprende la estimación de los beneficios y costos asociados al desarrollo e implementación de este tipo de solución. Es necesario considerar los costos aproximados para el hardware y software que serán usados. Debido a que los almacenes de datos tienden a expandirse rápidamente, es recomendable realizar estimaciones a corto plazo.

Es una de las fases más críticas del proyecto dado que, en la mayoría de los casos, la planificación impacta directamente en las posibilidades de éxito del proyecto y está íntimamente relacionada con la definición de los requerimientos del negocio.

3.2 Definición de los requerimientos del negocio

Los requerimientos son la pieza central del desarrollo exitoso de un datawarehouse, ya que estos junto con los usuarios del negocio afectan todos los aspectos de implementación de la solución analítica.

La definición de requerimientos contiene desde la disponibilidad de información, hasta frecuencia de actualización, pasando por la organización de la misma.

Los requerimientos de negocio no solo impactan en la arquitectura del datawarehouse, sino también en la definición de las interfaces de las aplicaciones de presentación de datos y en las fases de mantenimiento y crecimiento del almacén de datos.

3.3 Modelado dimensional

Para (Kimball & Ross, 2008) el modelo óptimo de bases de datos para de soporte a la toma de decisiones es el modelo dimensional, dado que combina simplicidad de utilización y alto rendimiento.

Al mismo tiempo definen un proceso de cuatro fases para la implementación eficaz de un modelo dimensional:

- **Escoger el área o proceso de negocios:** Se refiere a la elección del área del negocio a medir por el datawarehouse. Se investiga sobre el negocio y se establece comunicación con los encargados de los procesos del área para registrar sus necesidades y requerimientos.
- **Definir la granularidad:** Se entiende como definir el nivel de detalle con el que se almacenarán los datos en la tabla de hechos, una vez definido el proceso de negocios a trabajar.
- **Definición de las dimensiones:** Las dimensiones dan contexto a los registros presentes en la tabla de hechos y sus medidas. Adicionalmente proporcionan una variedad de maneras de observar la información.
- **Definición de los hechos:** Se determinan los aspectos del proceso de negocio que se pretenden medir, los indicadores utilizados para realizar dicha medición y el método de cálculo de los mismos, tomando la granularidad como referencia en todo momento.

3.4 Diseño físico

El modelo dimensional realizado anteriormente ha de ser trasladado a un diseño físico, es decir, que se procede a realizar la implementación física del almacén o bodega de datos que soportará la solución analítica.

En esta fase se deben seleccionar las estructuras necesarias que puedan dar soporte al diseño lógico, así como definir algunos aspectos importantes como son las convenciones establecidas para los nombres de los tributos, tipos de datos, declaraciones claves y permisividad de valores nulos. Además, también puede incluir aspectos referentes al ambiente de bases de datos, como la indexación, estrategias de particionamiento, ajustes de rendimiento y distribución de archivos.

3.5 Diseño de los procesos ETL

En esta fase se conceptualizan los procesos ETL que se utilizarán para la integración de las diferentes fuentes de datos que alimentarán el almacén de datos, de manera que puedan realizarse de manera automática y correcta. Implica la extracción de los datos generalmente de los sistemas operacionales. Las principales actividades de esta fase son:

- La extracción de datos desde las distintas fuentes.

- La transformación de los datos origen para que se convierta en la información estructurada que se ha definido en el modelo dimensional. Arreglando problemas de consistencia, estandarización, etc.

- La carga de datos en las estructuras definidas en la conceptualización de almacén de datos.

Esta fase ha sido muy subestimada, pero todas las actividades que la involucran son altamente críticas pues tienen que ver con la materia prima del almacén o bodega de datos: los datos. La desconfianza y pérdida de credibilidad en la información resultante del proyecto, provocará efectos inmediatos e inevitables si el usuario se encuentra con información inconsistente.

3.6 Especificación de aplicaciones de usuario final

Esta etapa se centra en la elección de las herramientas que utilizaremos para manipular y acceder a la información contenida en el almacén de datos. Cuadros de mando, reportes estáticos, herramientas de creación de reportes ad-hoc, son todas opciones válidas para presentar información al usuario.

3.7 Diseño de arquitectura técnica

Se trata de la esquematización y organización de todos los elementos que, dentro del ámbito de la tecnología, nos asistirán en el desarrollo de la solución de BI.

En general se seleccionan las herramientas que se utilizarán y se procede a implementar la solución de BI utilizando dichas herramientas de una manera organizada (implementación de modelos, ETL y realización de las primeras cargas masivas de datos).

El diseño técnico de la arquitectura sirve como un marco organizativo que soporta la integración de los elementos tecnológicos necesarios para el desarrollo de la solución de inteligencia de negocio. Este diseño permite identificar los componentes más importantes y minimizar imprevistos al momento de desarrollar e implementar la solución especificada, por medio del análisis de las problemáticas que podrían afectar el proceso de desarrollo.

3.8 Selección de Productos e Instalación

La selección de los productos está estrechamente relacionada al plan de arquitectura planteado en la fase anterior, e incluye una serie de actividades específicas para una buena elección de productos, los cuales vienen dados por la comprensión de los procesos de negocio, el establecimiento de un plan de evaluación del producto, donde se coloquen ponderaciones para indicar la importancia de los elementos y cuyos criterios comunes tomados en esta fase sean sobre la funcionalidad, la arquitectura técnica, las características de software, el impacto de las infraestructuras y la viabilidad de los proveedores.

3.9 Desarrollo de las Aplicaciones Analíticas

Luego de la especificación de las aplicaciones analíticas para usuarios finales, el desarrollo de esas aplicaciones involucra configuraciones de los metadatos y construcción de reportes específicos. Los usuarios accederán a los datos del almacén o bodega de datos por medio de una o varias herramientas de productividad, basadas en una interfaz gráfica de usuario. Estas herramientas pueden incluir software de consultas, generadores de reportes, procesamiento analítico en línea o herramientas de minería de datos, y se seleccionaran dependiendo de los tipos de usuarios y sus requerimientos particulares.

3.10 Despliegue

Por último, está la fase de despliegue. En ella convergen los flujos de actividades asociadas a las fases que involucran a la tecnología, datos y aplicaciones analíticas, donde se logra la unificación de todas ellas. Si la planificación se ha ejecutado correctamente y se ha asociado a las pautas indicadas, se puede asegurar que los resultados de los flujos de tecnología, datos y aplicaciones analíticas. La disponibilidad de la infraestructura de capacitación y apoyo.

3.11 Mantenimiento y crecimiento del almacén de datos

Luego de realizada la implementación se debe proceder al mantenimiento del almacén de datos, así como es necesario el adiestramiento de los usuarios que utilizarán las herramientas de presentación de datos.

Tras el despliegue, se inicia una fase de mantenimiento del sistema cuándo se encuentra en producción. Este mantenimiento incluye un conjunto de actividades técnicas operacionales que son necesarias para mantener en estado óptimo el flujo de trabajo habitual del sistema. En este punto, se deberán realizar actividades como:

- Monitorización del uso del sistema. Mantenimiento de la tabla de índices.

- Procedimiento de respaldo del sistema. Resolución de posibles incidencias.

Es necesario también que exista la capacidad de abordar nuevos requerimientos del negocio mediante la expansión del modelo dimensional, dado que los procesos de negocio rara vez son estáticos e impactan inevitablemente en los procesos analíticos relacionados a ellos.

CAPÍTULO IV

MARCO APLICATIVO

En este capítulo se presenta todo lo relacionado a la implementación realizada en este Trabajo Especial de Grado de acuerdo a la metodología definida por Ralph Kimball (ver Capítulo III), la cual permite el desarrollo de una solución analítica adaptable a cualquier institución financiera, basada en la obtención de indicadores relacionados a la rentabilidad.

De esta manera, se procede a describir las actividades realizadas a lo largo del desarrollo de la solución analítica, distribuidas en las tareas comprendidas en la metodología aplicada, abarcando desde la definición de requerimientos pasando por el diseño del almacén de datos y la elaboración de los procesos ETL, hasta la realización de consultas analíticas para el desarrollo y muestra de indicadores de gestión.

4.1 Tareas del Proyecto

4.1.1 Planificación del Proyecto

En esta tarea se realiza una planificación que permite desarrollar un conjunto de actividades que derivan en la implementación de la solución analítica planteada. Dentro de estas actividades resaltan: el levantamiento y análisis de información, las reuniones con los usuarios, la definición de la arquitectura de la solución, la instalación y configuración de los productos tecnológicos seleccionados, el diseño e implementación del almacén de datos, la elaboración y ejecución de los procesos ETL tanto de carga inicial como de carga incremental junto con la programación de las tareas automáticas que se encargan de ejecutar los *jobs* periódicamente, la construcción y publicación de indicadores y reportes de rentabilidad, y finalmente, la transferencia de conocimiento tecnológico a los usuarios finales.

Todas esas actividades se plantean siguiendo un orden de ejecución basado en el nivel de prioridad que poseen y en la dependencia que existe en el desarrollo de ellas.

Todas esas actividades se plantean siguiendo un orden de ejecución basado en el nivel de prioridad que poseen. No obstante, hay que destacar que se estima llevar a cabo esta planificación en un periodo de tiempo de 20 semanas aproximadamente, como se observa a continuación:

ACTIVIDADES	TIEMPO
Planificación del proyecto	2 Semana
Definición de requerimientos	4 Semanas
Especificación de aplicaciones	2 Semana
Diseño de arquitectura e instalación de productos	2 Semanas
Modelado dimensional	3 Semanas
Diseño físico	2 Semanas
Desarrollo de ETL y carga de datos	3 Semanas
Despliegue	2 Semanas
TOTAL	20 SEMANAS

4.1.2 Definición de los Requisitos del Negocio

La solución analítica a implementar, va a servir a las instituciones financieras, como herramienta para la obtención de datos para el análisis de rentabilidad en sus distintas formas. De esta manera se han identificado las necesidades del negocio, a través de la definición de los indicadores de rentabilidad necesarios para la obtención de la información deseada.

Cada uno de los indicadores que se muestran posteriormente sirve para brindar al usuario distintos puntos de vista en lo que a rentabilidad se refiere, haciendo uso de varios niveles de detalle (granularidad).

De esta manera el objetivo de la solución analítica es proveer a los usuarios de datos confiables, unificados y verificados de manera eficaz, eficiente oportuna.

Entre los indicadores necesarios para la correcta medición de la rentabilidad se pueden destacar:

- ROE: Relación entre el capital invertido en un activo y los ingresos que produce.
- ROA: Relación entre el valor de un activo y los ingresos que produce.
- Margen financiero bruto (MFB): Relación entre los ingresos y costos financieros, sin tomar en cuenta otros factores externos.

A continuación se especifican los indicadores a calcular y utilizar:

Nombre	Descripción	Método de cálculo
Saldo Actual	Saldo total.	$\Sigma(\text{SaldoFinal})$
Saldo Promedio	Saldo promedio.	$\Sigma (\text{MtoSaldoProm})$
Ingresos Financieros	Se refiere al Interés acumulado al mes proveniente de los rendimientos por Cartera de Créditos, Comisiones Flat (Cuenta Contable 513) y 1los Intereses Promocionales.	$\Sigma(\text{SaldoFinal})$ *De contratos asociados a la cuenta contable 513
Costos Financieros	Se refiere al Interés acumulado al mes proveniente de los costos por las captaciones del público (promocionales o convencionales). (Cuenta Contable 411)	$\Sigma(\text{SaldoFinal})$ *De contratos asociados a la cuenta contable 411
Ingresos Pool	Se refiere al interés obtenido al aplicar la Tasa Pool Activa (Referencia de la transferencia de recursos a un costo único aplicado a las operaciones activas) sobre los Saldos Promedios Mensuales.	$\Sigma (\text{MtoSaldoProm} * \text{tasaPoolActiva} / (360/30))$
Costo Pool	Se refiere al interés obtenido al aplicar la Tasa Pool Pasiva (Referencia de la transferencia de recursos a un costo único aplicado a las operaciones pasivas) sobre los Saldos Promedios Mensuales.	$\Sigma (\text{MtoSaldoProm} * \text{tasaPoolPasiva} / (360/30))$
Margen Financiero Bruto (MFB)	Una estimación que resulta del diferencial entre los ingresos financieros obtenidos por el rendimiento de los productos activos y el costo resultante de las captaciones del público.	(Ingresos financieros+Ingresos pool)- (Costos financieros+Costos Pool)
Margen Financiero Neto (MFN)	Consiste en una estimación resultante de tomar como insumo	MFB – MontoPrimaRiesgo + Recuperaciones de activos

	el resultado del Margen Financiero Bruto, así como las provisiones incobrables y los ingresos por recuperaciones de los activos financieros.	financieros
Prima de Riesgo	Se refiere a una provisión que consiste en la porción estimada de las cuentas incobrables sobre el saldo de la cartera vencida.	No calculado
Monto por prima de riesgo	Se refiere al monto estimado de las cuentas incobrables sobre el saldo de la cartera vencida.	$\Sigma (\text{MtoSaldoProm} * \text{MtoPorcentajePrima} / (360/30))$
Recuperaciones de activos financieros	Se refiere a los ingresos generados por la recuperación de créditos y otros activos financieros en castigos. (Cuenta Contable 520).	$\Sigma (\text{SaldoFinal})$ *Asociado a las cuentas 520
Margen Operativo Bruto	Constituye una estimación que toma en consideración el resultado obtenido del Margen Financiero Neto, así como el Módulo de Gastos definido por las áreas responsables, donde se tomarán en consideración los Gastos de Transformación asociados a la actividad bancaria.	MFN – Gastos generales y Administrativos
Gastos Generales y Administrativos	Se refiere a los gastos en los cuales incurre el Banco por concepto de gastos de personal, servicios externos, traslados y comunicaciones, seguros, mantenimiento y reparaciones. (Cuenta Contable 440)	$\Sigma (\Delta(\text{SaldoFinal}))$ *De contratos asociados a la cuenta contable 440
Margen Operativo Neto	Consiste en la estimación resultante de tomar como insumo el resultado obtenido en el Margen Operativo Bruto, así como los ingresos/gastos operativos y el impuesto sobre la renta.	MFN+Ingresos Operativos- Gastos operativos-Impuesto sobre la renta
Ingresos Operativos	Se refiere a los ingresos por Comisiones cobradas al cliente por Operaciones (tales como Fideicomisos, otras líneas de crédito, giros y transferencias, emisión y envío de tarjetas de crédito, comisiones netas POS) y por Otras Comisiones No Financieras (tales como Estados de cuenta, cheques devueltos,	$\Sigma (\text{MontoTransacción})$ *De contratos en las cuentas contables 5 correspondientes a rendimiento

	sobregiros, emisión/renovación y reposición de tarjetas).	
Gastos Operativos	Se refiere a los gastos por servicios (tales como Giros y transferencias, cámara de compensación, comisiones cedidas en POS).	$\Sigma(\text{MontoTransacción})$ *De contratos en las cuentas contables 4 correspondientes a costos
Rentabilidad sobre el Patrimonio (ROE)	El ROE, por sus siglas en inglés, es un indicador financiero que permite valorar la rentabilidad del capital empleado en una inversión, midiendo la capacidad de la empresa de remunerar a sus accionistas.	$\frac{\text{Margen Operativo Neto}}{\Sigma(\text{SaldoFinal})}$
Rentabilidad sobre los Activos (ROA)	El ROA (Return on Assets) consiste en un indicador financiero que permite valorar la rentabilidad de una empresa en función al total de sus Activos, midiendo la capacidad de la empresa de gestionar eficientemente sus activos para generar valor.	$\frac{\text{Margen Operativo Neto}}{\Sigma(\text{SaldoFinalActivo})}$

Reportes:

Cuadro integral de rentabilidad: Cuadro general de indicadores de rentabilidad que provee una fotografía global.

Indicadores:

- Ingresos Operativos
- Ingresos Pool
- Costos Operativos
- Costos Pool
- Ingresos Financieros
- Costos Financieros
- Margen Financiero Bruto
- Margen Financiero Neto
- ROE
- ROA

Frecuencia: Mensual

Unidad de medición: Bs

Nivel de detalle: Por región, zona, oficina.

Gráfico: Tabla.

Reporte Evolutivo: Describe mediante gráficos, la evolución mensual de los indicadores de rentabilidad especificados.

Indicadores:

- Margen Financiero Bruto
- Margen Financiero Neto
- ROE
- ROA
- Ingresos tipificados
- Costos tipificados

Frecuencia: Mensual

Unidad de medición: -

Nivel de detalle: Por región, zona, oficina.

Gráfico: Gráfica de líneas.

Cuadro de mando por oficina – Ingresos vs Costos: Describe mediante gráficos, la evolución mensual de varios de los indicadores utilizados para construir indicadores de más alto nivel, con el propósito de dar profundidad a la información suministrada.

En caso de querer visualizar alguno de los reportes

Indicadores:

- Ingresos Operativos
- Costos Operativos
- Ingresos Financieros
- Costos Financieros
- Ingresos Pool
- Costos Pool
- Saldos de cierre
- Saldos Promedio

Frecuencia: Mensual

Unidad de medición: Bs

Nivel de detalle: Por región, zona, oficina.

Gráfico: Gráfica de líneas.

4.1.3 Diseño de la Arquitectura Técnica

El diseño de la arquitectura de la solución analítica que se desarrolla se detalla a continuación (Imagen 18):

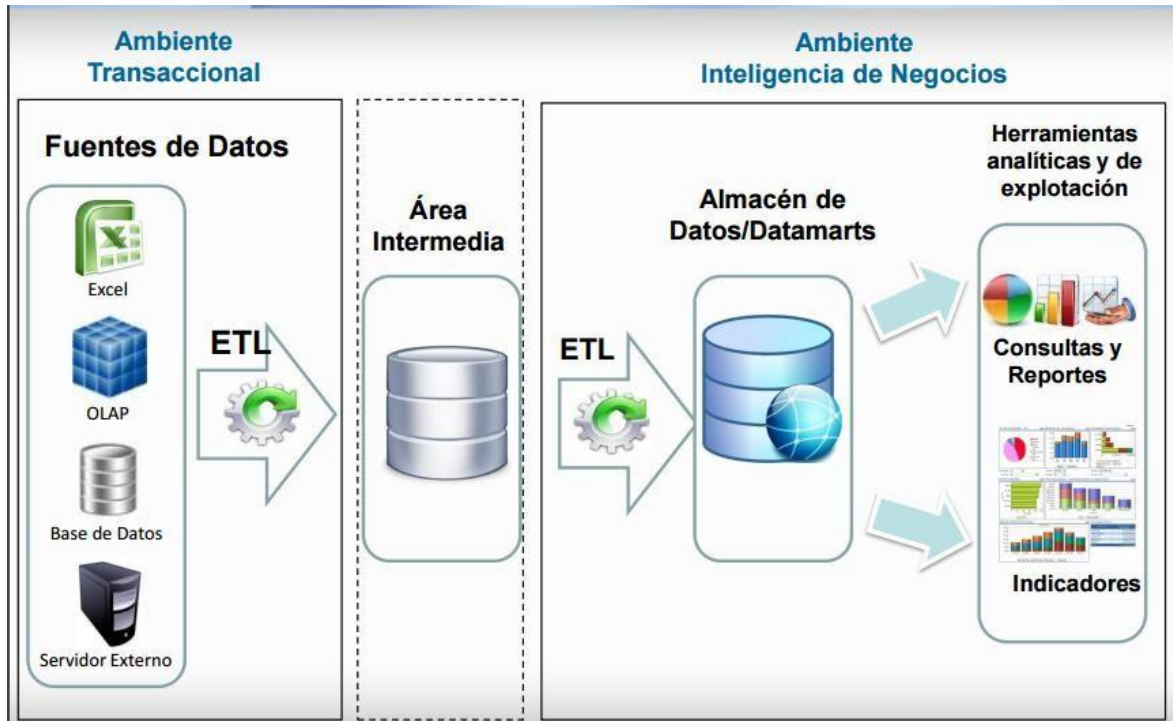


Imagen 18: Arquitectura estándar de una solución analítica

Teniendo como base dicha imagen se observa que la arquitectura está conformada por tres partes o componentes así como diversos procesos y elementos que permiten el funcionamiento de la misma.

El primer componente comprende las diversas fuentes de datos que alimentan la solución analítica. Estas fuentes pueden ser internas y/o externas, usualmente se utilizan aquellas provistas por los sistemas transaccionales relacionados con los procesos de negocio previamente estudiados. Para el caso de estudio del prototipo a construir, se utilizarán fuentes en Excel y CSV en su mayoría. Dado que los sistemas transaccionales de las instituciones financieras (IBS por ejemplo) solo permiten extraer datos en forma de hojas de Excel o formato CSV.

El segundo componente es el relacionado a los procesos de extracción, transformación y carga (ETL) que se ejecutarán periódicamente. Mediante estos procesos, se comienza a poblar el almacén con los datos, que desde las fuentes, son previamente transformados y unificados. En este ambiente se utiliza la herramienta de Pentaho Data Integration (PDI) para crear y ejecutar dichos procesos.

En cuanto al tercer componente, este se divide en dos componentes:

El primero es el almacén de datos, donde se localizan los datos históricos, que son el resultado de la integración de las diversas fuentes, de manera que las herramientas de consulta puedan acceder a estos de manera eficiente, dado que el almacén está estructurado en base a un modelo dimensional.

Este almacén será periódicamente alimentado por los procesos ETL. Para este componente se hace uso de la Base de Datos SQL Server.

El segundo componente se refiere a la capa de presentación, es decir, a las herramientas de acceso a los datos contenidos en el almacén para permitir a los usuarios finales organizar y visualizar los distintos indicadores y reportes definidos en la solución analítica. Para este componente se utiliza la herramienta Oracle Business Intelligence (OBI).

4.1.4 Selección de Productos e Instalación

La selección de las herramientas para el desarrollo e implementación de una solución analítica, puede resultar complejo, dada la diversidad de las mismas y a la variedad de sus características, siendo Pentaho Business Intelligence, Tableau Software, Oracle Business Intelligence, IBM Cognos y Microsoft Power Business Intelligence algunas de las más populares.

Sin importar la combinación de herramientas seleccionada, ésta debe permitir la implementación de un repositorio centralizado, la capacidad de extraer y unificar datos desde las diversas fuentes, visualización reportes y la posibilidad de construir consultas a demanda de acuerdo a los requerimientos de información.

Luego de tomar en cuenta diversos aspectos, se decidió por la instalación de SQL Server para la implementación y diseño de los diversos repositorios necesarios durante el desarrollo de la solución analítica. En lo referente a los procesos ETL y su implementación, se procedió a instalar Pentaho Data Integration dada su gran flexibilidad para el manejo y unificación de datos. Adicionalmente, Pentaho proporciona acceso a la gran mayoría de sus funcionalidades por ser una herramienta de código abierto.

Luego, en lo que respecta a las herramientas de extracción y visualización de datos, se utilizará Oracle BI 12c, siendo ésta una herramienta que proporciona todas las cualidades y funcionalidades mencionadas anteriormente.

Luego de estudiar las distintas herramientas presentes en el mercado, se optó utilizar para el desarrollo de la solución de inteligencia de negocio planteada para la obtención de indicadores de rentabilidad que apoyen la toma de decisiones, las herramientas de la suite Oracle Business Intelligence y una herramienta de la suite de Pentaho (Pentaho Data Integration). A continuación el esquema de la solución analítica con las herramientas ubicadas (Imagen 19).

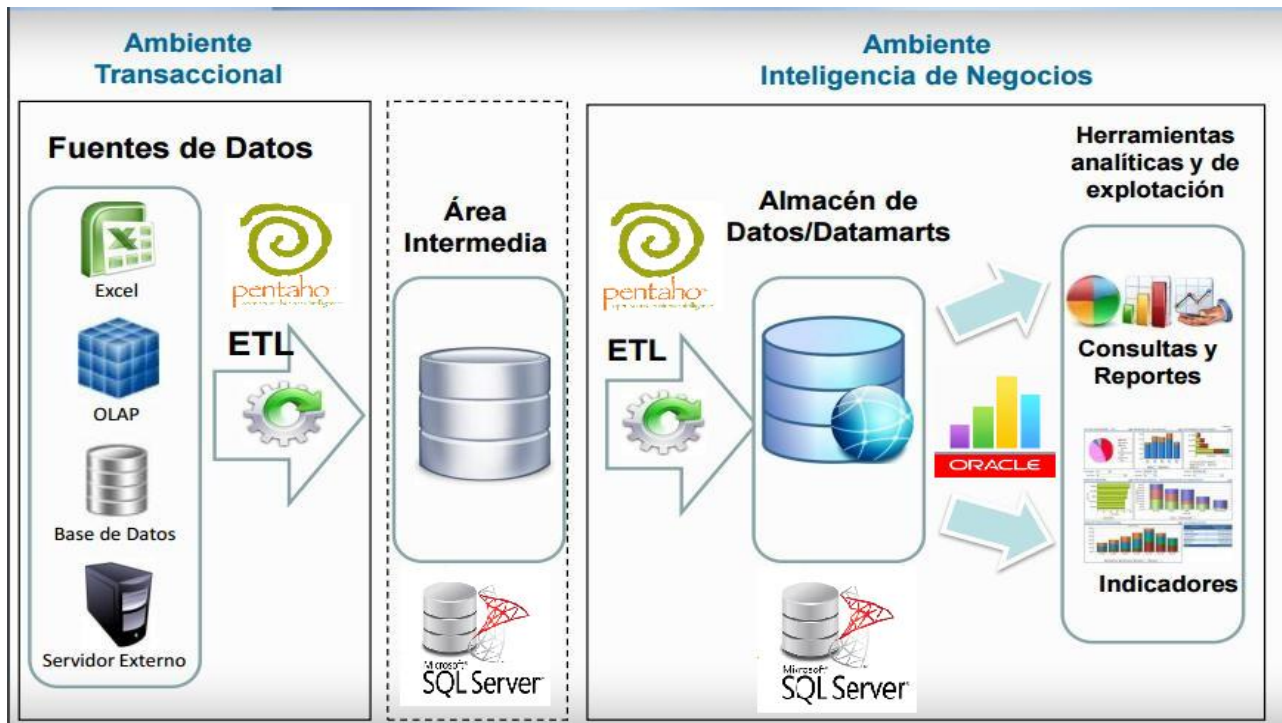


Imagen 19: Correspondencia entre herramientas y componentes de la solución

4.1.5 Diseño del Modelo Dimensional

En base a los requerimientos del negocio planteados en el punto 4.1.2 y la definición de la arquitectura técnica a emplear descrita en el punto 4.1.3, se evaluaron las características de los datos y las fuentes para identificar hechos medibles o dimensiones que pudieran formar parte de la solución analítica, para luego, dar inicio al diseño del modelo dimensional del almacén de datos, el cual comprende el proceso del cálculo de indicadores de rentabilidad al cierre de cada mes.

Sin embargo, de acuerdo con la metodología de Ralph Kimball, este paso comprende la realización de varias actividades, siendo éstas:

4.1.5.1 Definir el Proceso de Negocio

Se orienta el diseño al modelaje del proceso de cálculo de rentabilidad de una institución financiera.

4.1.5.2 Identificar el Nivel de Granularidad

Entendiendo como granularidad al nivel de detalle al que se almacenarán los datos en el *datawarehouse*, es útil utilizar los registros de la tabla de hechos como referencia. Siendo así, la granularidad del modelo dimensional se define como sigue:

Indicadores financieros (ROE, ROA, MFB, MFN), por oficina, almacenados de manera mensual (haciendo referencia a la fecha del último día del mes).

4.1.5.3 Definir las Dimensiones

De acuerdo con los requerimientos de información identificados en la fase de diseño, se identifican diversas dimensiones que contextualizan los hechos y permiten agrupar los datos siendo éstas:

DIMREGION: Guarda la información que describe a las regiones geográficas, primer nivel de la jerarquía.

DIMZONA: Almacena lo referente a las características de las zonas geográficas, que a su vez están dentro de las regiones.

DIMOFICINA: Posee la información sobre las oficinas o sucursales de la entidad financiera en cuestión y corresponde al nivel más bajo en la jerarquía geográfica.

DIMTEMPORAL: Almacena las fechas y es indispensable para el modelo, ya que ubica a los hechos modelados en el tiempo.

4.1.5.4 Identificación de los Hechos y las Tablas de Hechos

Luego de identificar las necesidades del negocio, así como los indicadores que sirven para medir la rentabilidad, se modela una tabla de hechos que abarca todos los indicadores y permite la manipulación efectiva de los datos para producir información.

Es importante destacar la naturaleza de la tabla de hechos definida a continuación como periódica o de tipo snapshot o fotografía, de temporalidad mensual, dado que este tipo de tablas de hechos modela de mejor manera los requerimientos del negocio, siendo que la evolución de los indicadores de rentabilidad no es significativa en espacios de tiempo menores y dichos indicadores no modelan hechos transaccionales como ventas o ingresos de pacientes.

FactRentabilidad:

- IFinancieros: Representa el monto en Bs de ingresos financieros (Cuentas 513).
- GFinancieros: Monto en Bs de costos financieros (Cuentas 411).
- IPool: Muestra el monto en Bs de los ingresos por tasas de intercambio interno (Saldos por tasa Pool Activa)
- GPool: Muestra el monto en Bs de los gastos por tasas de intercambio interno (Saldos por tasa Pool Pasiva)
- IOperativos: Monto en Bs de ingresos operativos (Montos de transacciones En cuentas 5)
- GOperativos: Monto en Bs de gastos operativos (Montos de transacciones En cuentas 4)
- Margen Financiero Bruto: Relación entre los ingresos (Financieros y Pool) y los gastos (Financieros y Pool)
- Margen Financiero Neto: Relación entre los ingresos (Financieros y Pool) y los gastos (Financieros y Pool), tomando en cuenta la prima de riesgo.
- Margen Operativo Bruto: MFN restando los costos operativos.
- Margen Operativo Neto: MFN sumando ingresos operativos y restando los costos operativos.
- ROE: Relación entre el MON y la totalidad de los saldos.
- ROA: Relación entre el MON y la totalidad de los saldos activos.

4.1.6 Diseño Físico

El diseño físico de la solución consiste en la creación de las estructuras que conforman el modelo dimensional del almacén de datos, construido a partir de lo definido en las fases anteriores (Imagen 20).

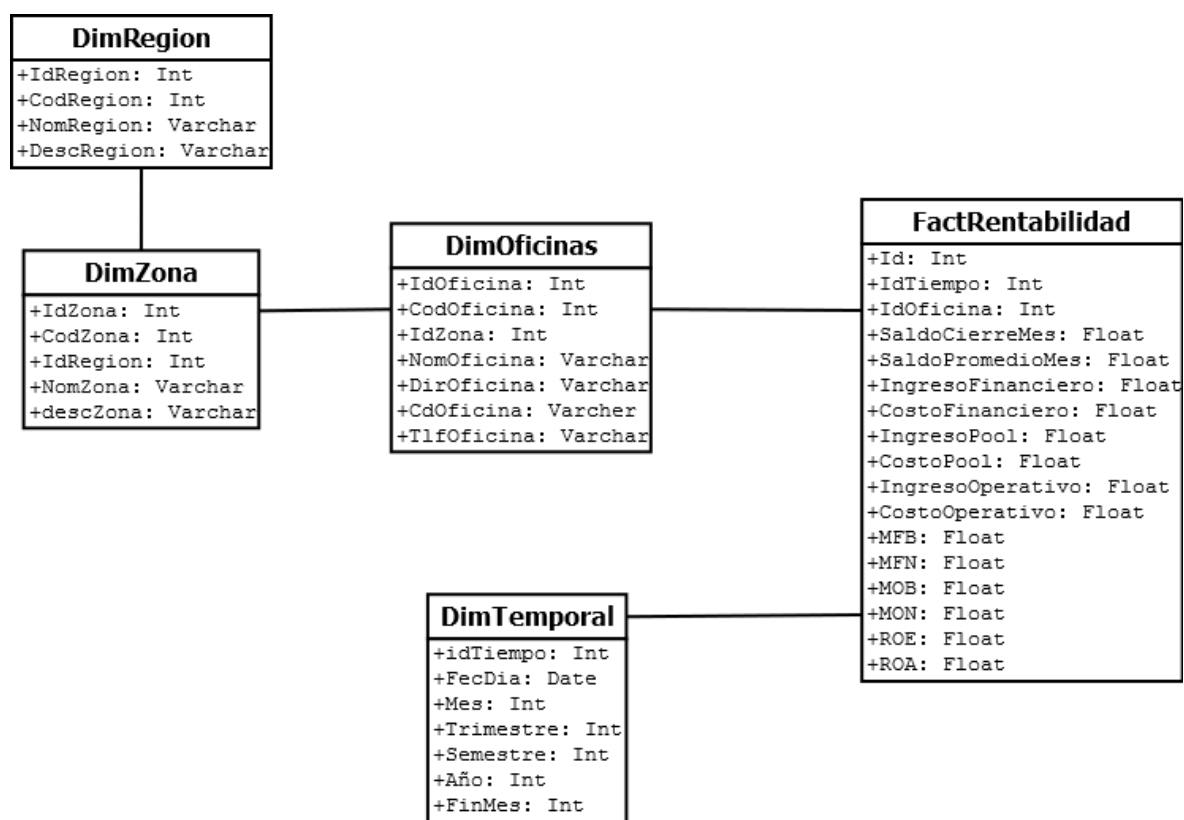


Imagen 20: Diseño físico del Datawarehouse

Una vez definido el modelo dimensional, se procede a implementar dicho modelo en una base de datos SQL Server, mediante el uso de la herramienta SQL Server Management Studio. Dado que el modelo es utilizado por una sola institución financiera en cualquier momento, solo se necesita un espacio de trabajo.

4.1.7 Diseño de procesos ETL

Después de crear en físico el almacén de datos en SQL Server, se comienzan a diseñar e implementar los procesos ETL para extraer, transformar y cargar los datos procedentes de las fuentes de datos (suministradas por la institución financiera en su mayoría en Excel y CSV) al almacén, de manera que luego de poblarlo correctamente con datos unificados y los indicadores necesarios, los usuarios puedan obtener fácilmente a través de herramientas de visualización, información de manera precisa y oportuna para la toma de decisiones estratégicas en cuanto a rentabilidad se refiere. Sin embargo, cabe señalar que los datos utilizados como fuentes de datos no son reales, es decir, que toda la información que arrojará la aplicación es simulada.

Para el desarrollo de los procesos ETL, se utiliza la herramienta Pentaho Data Integration (PDI - Kettle), como se mencionó anteriormente. A través de Pentaho Data Integration se crean las transformaciones (*transformations*) que luego de obtener los datos desde las fuentes, los unifican, seleccionan y estandarizan, para luego calcular los indicadores necesarios y así poblar las tablas del almacén de datos de manera correcta. Luego de la implementación de dichos ETL, se procede a crear los *Jobs* o *Trabajos*, también se crean *jobs* (trabajos), que permitirán ejecutar un conjunto de transformaciones en el orden correcto.

El primero de dichos trabajos es el que combina la extracción desde las fuentes de datos relacionadas a las dimensiones y la realización del backup de cada una de las mismas en la computadora (Imagen 21).

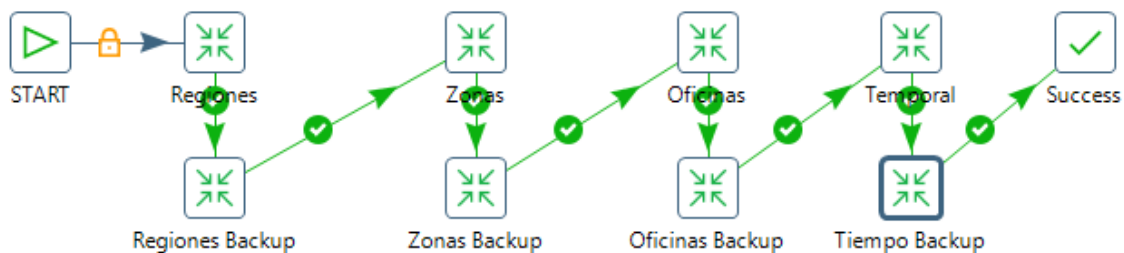


Imagen 21: Job encargado de cargar y respaldar las dimensiones.

Luego encontramos el Job encargado de la extracción y unificación de datos desde las fuentes contables, para luego cargar indicadores calculados y directos a la tabla de hechos. Dichas fuentes se procesan en dos partes, dadas las limitaciones físicas de la computadora que realiza las operaciones. Luego del procesamiento se realiza un backup local, como medida de seguridad. Dicho backup no sería viable si se tuviera una tabla de hechos con mayor detalle (granularidad), pero dada la naturaleza de la solución analítica, es completamente factible (Imagen 22).

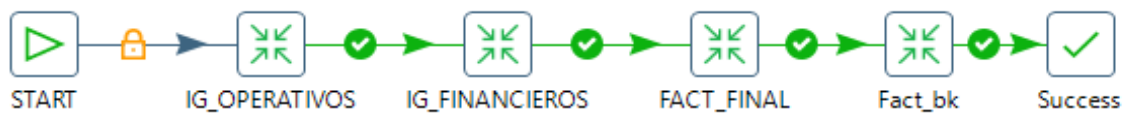


Imagen 22: Job encargado de cargar y respaldar la tabla de hechos.

Los Jobs mencionados deben ser ejecutados en un orden específico, dado que la estructura del almacén de datos lo requiere. Para cumplir este objetivo se pueden utilizar archivos .BAT organizados por el manejador de tareas de Windows. Siendo así, estos serán ejecutados los últimos días de cada mes, luego de que estén disponibles los cierres contables de la institución financiera en cuestión.

Para que dichos procesos no tengan incidencia en el desempeño de las actividades diarias de la institución, se realizarán en horas nocturnas, específicamente entre las 10 pm y las 12 pm.

4.1.8 Desarrollo de las Aplicaciones Analíticas

Por medio de la herramienta Oracle BI Server, específicamente Administration Tool, se da inicio al desarrollo de la aplicación de la solución analítica, en donde se despliega en cada una de las capas que componen a la herramienta, toda la estructura y los datos del almacén de datos implementado anteriormente.

Las tres (3) capas que integran a esta herramienta, las cuales se pueden apreciar en la imagen 23, son:

- Capa física.
- Capa de negocio y mapeo.
- Capa de presentación.

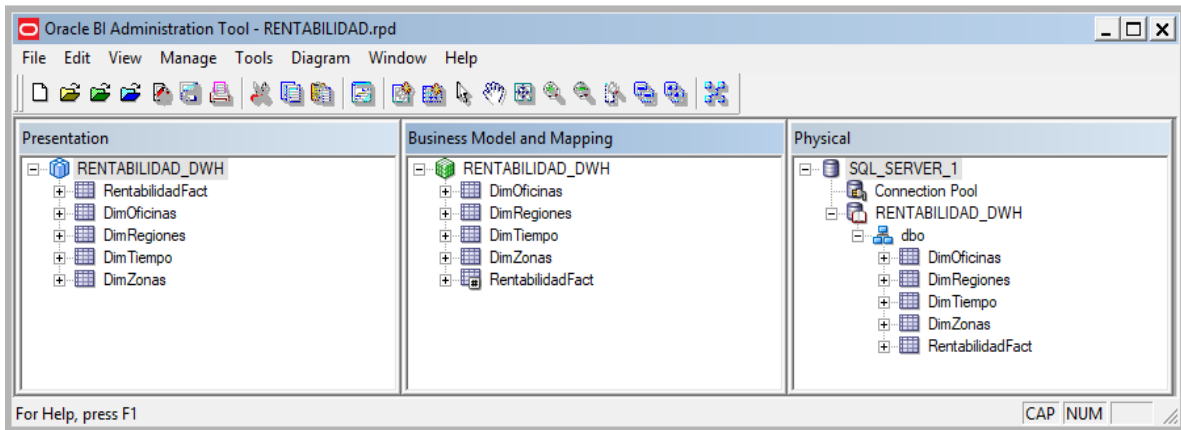


Imagen 23: Vista del repositorio en Oracle Admin Tool

Dentro de la capa física se tiene acceso a la estructura física del almacén de datos que se encuentra en la base de datos SQL Server, por tanto, a través de ella se importan los metadatos asociados al modelo dimensional que conforma el almacén que se implementó, mediante la creación de una conexión a base de datos . Este almacén se encuentra en el esquema SQL_SERVER_1. Una vez que se importan los metadatos, muchas de las propiedades de las fuentes se configuran automáticamente como lo son las relaciones de claves primarias y foráneas. Imagen 24 se muestra parte de la capa física con más detalle.

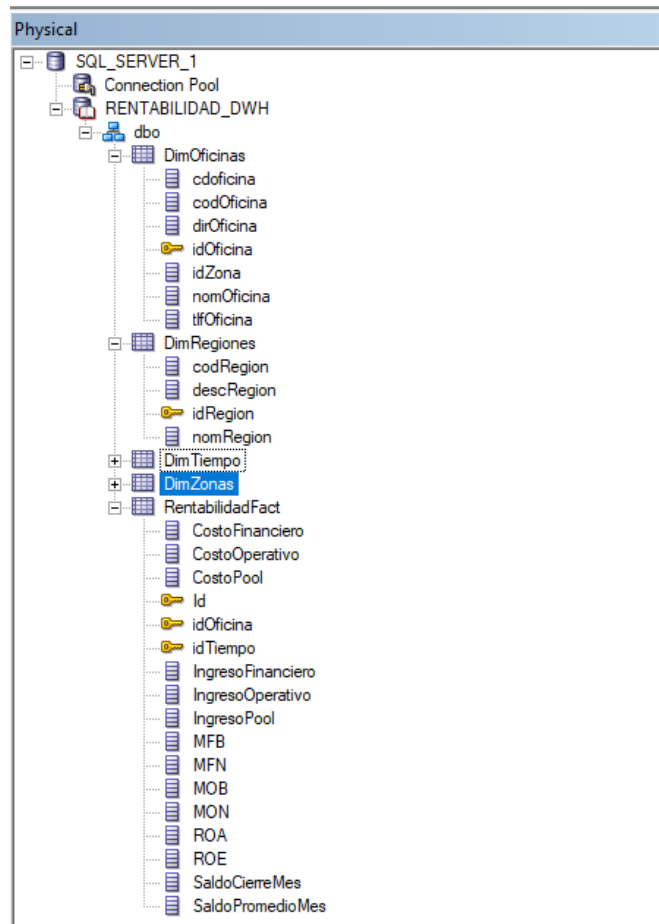


Imagen 24: Vista física del repositorio en Oracle Admin Tool

En la capa de negocio y mapeo, se determinan las distintas áreas temáticas que se identificaron en la definición de los requerimientos del negocio. Para esta solución el área a trabajar es la rentabilidad.

Sin embargo, también se crean las jerarquías lógicas para las dimensiones, permitiendo con ello realizar operaciones de *drill down* o *roll up*, se especifican las operaciones de agregación sobre las métricas de las tablas de hechos y se eliminan los atributos que no son significativos para los usuarios finales (Imagen 25).

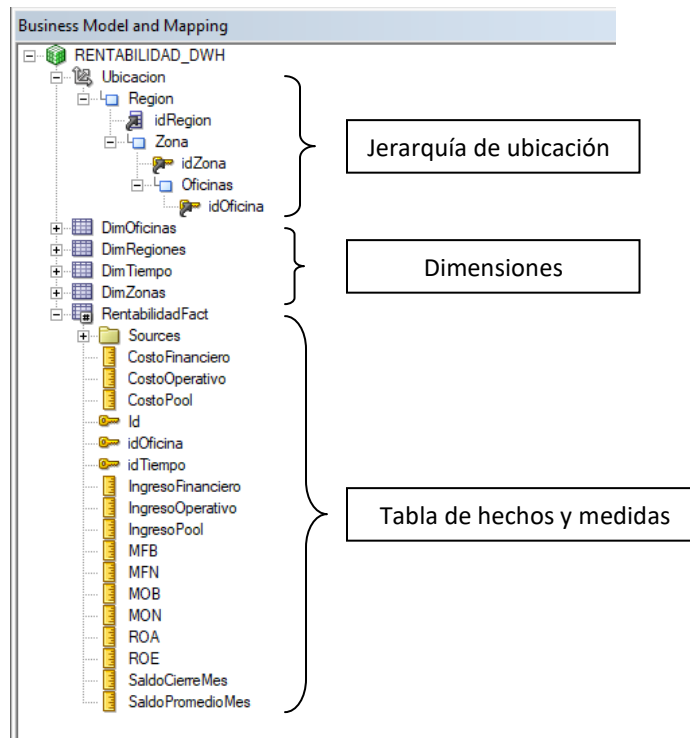


Imagen 25: Modelo de negocio del repositorio en Oracle Admin Tool

Al momento de definir las tablas de hechos dentro de la capa de negocio y mapeo, se especifican las operaciones de agregación que se le aplican a cada una de las métricas que las representan (Imagen 26).

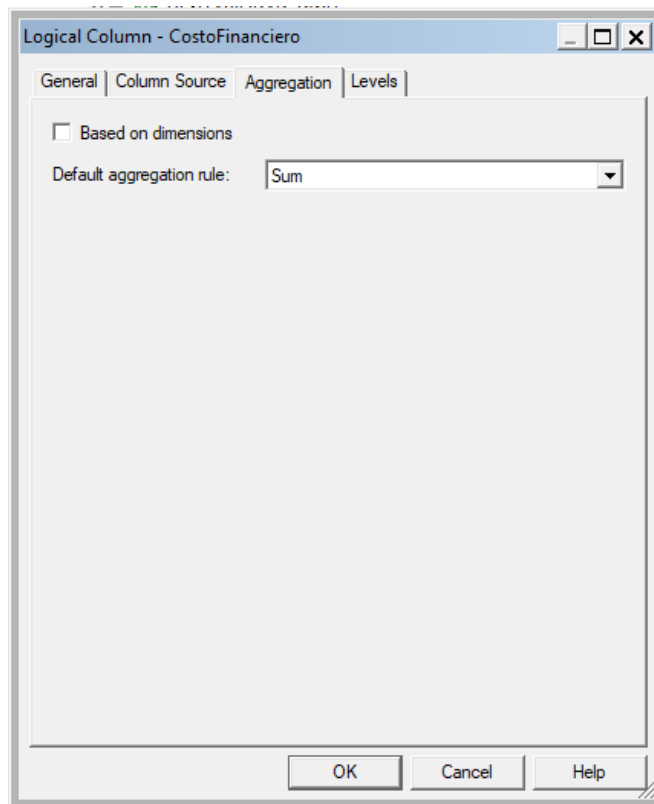


Imagen 26: Agregaciones para las medidas en Oracle Admin Tool

Por otro lado, en la capa de presentación, se encuentran los datos que ven los usuarios por medio de la aplicación analítica, después de haber sido tratados y especificados bajo la estructura lógica definida en el modelo de negocio. En esta capa se pueden reorganizar los datos y renombrarlos para hacerlos más entendibles para el usuario final (Imagen 27).

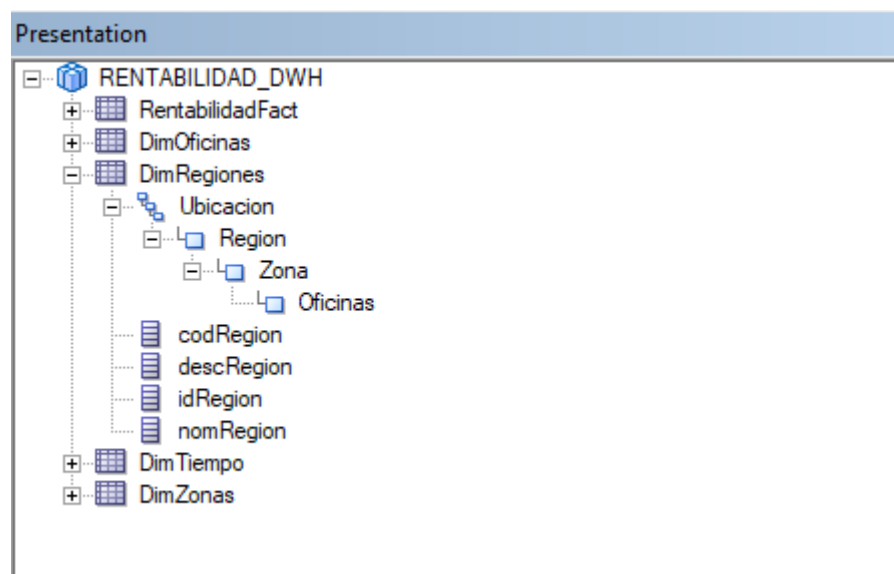


Imagen 27: Vista de presentación del repositorio en Oracle Admin Tool.

4.1.9 Desarrollo

Una vez definidos en la herramienta OBI Admin Tool, los datos que consumirá la aplicación para el usuario final, se hace uso de Oracle Business Intelligence, donde se realizan los reportes y análisis de los datos usando los indicadores definidos anteriormente.

Mediante Oracle Business Intelligence se visualiza el portal web, que constituye parte del producto final de esta solución de analítica de cara a la capa de extracción y visualización de datos.

Es aquí donde se muestran los indicadores de rentabilidad más relevantes, en granularidad baja para ofrecer una vista general del negocio.

Luego se encuentran los reportes que permiten obtener vistas comparativas de diversos indicadores, en temporalidades distintas, permitiendo identificar tendencias y relaciones poco obvias entre las dimensiones que contextualizan a los hechos.

La solución analítica, toma como principal elemento de muestra de datos al cuadro de mando o *dashboard*. Estos son reportes complejos, compuestos a su vez de reportes más pequeños, permitiendo que se muestren indicadores diversos.

El primero de estos cuadros de mando tiene como objetivo mostrar los indicadores de rentabilidad más relevantes (ROE, ROA, MFB, MFN, etc), a través del año, dividido por meses (Imagen 28).

En el caso de ROE y ROA, se muestra el valor correspondiente a toda la institución financiera. Mientras que para gastos e ingresos se muestran los valores por región, con distinción entre los tipos de los mismos (financieros, operativos y pool).

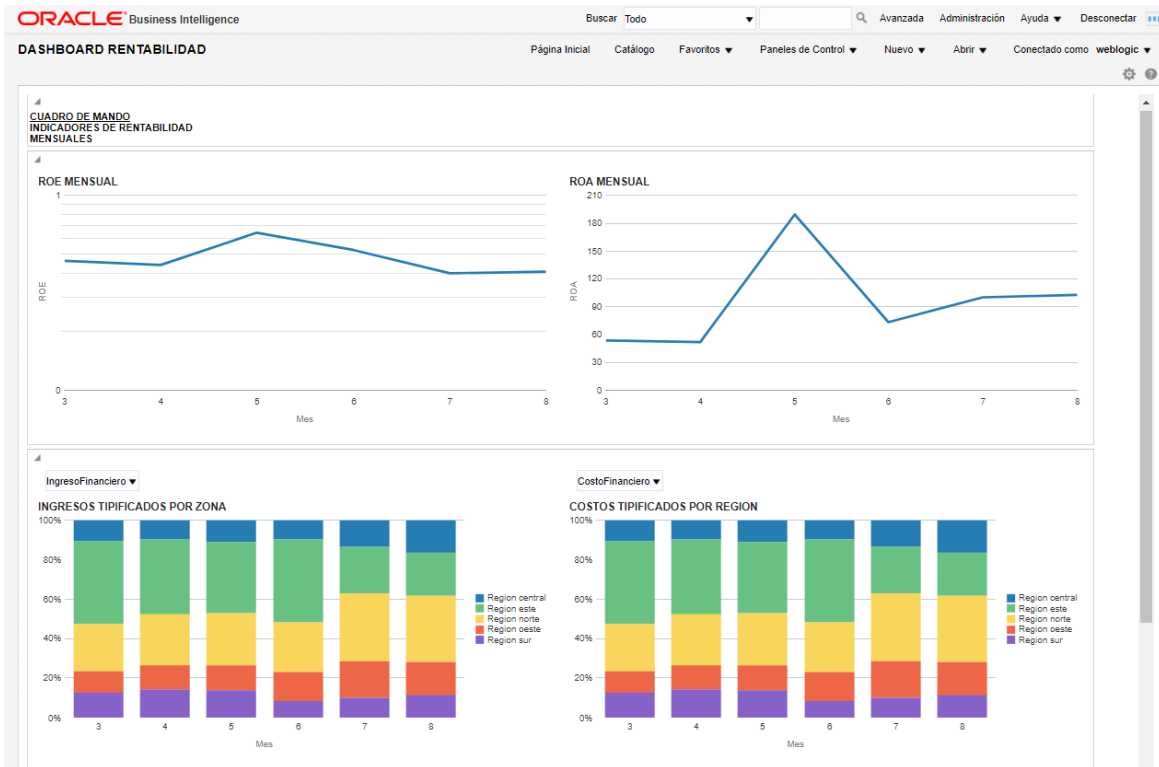


Imagen 28: Dashboard principal de rentabilidad, presentando indicadores generales.

El segundo cuadro de mando se centra en aportar información sobre las oficinas, haciendo referencia a indicadores de nivel medio (saldos, ingresos y costos). De esta manera se provee a los usuarios de información detallada, permitiendo la identificación de factores importantes para lograr cambios significativos en la gestión de la rentabilidad.

Este cuadro de mando está provisto de un filtro de oficinas que permite la visualización de los indicadores por una o varias oficinas (Imagen 29).

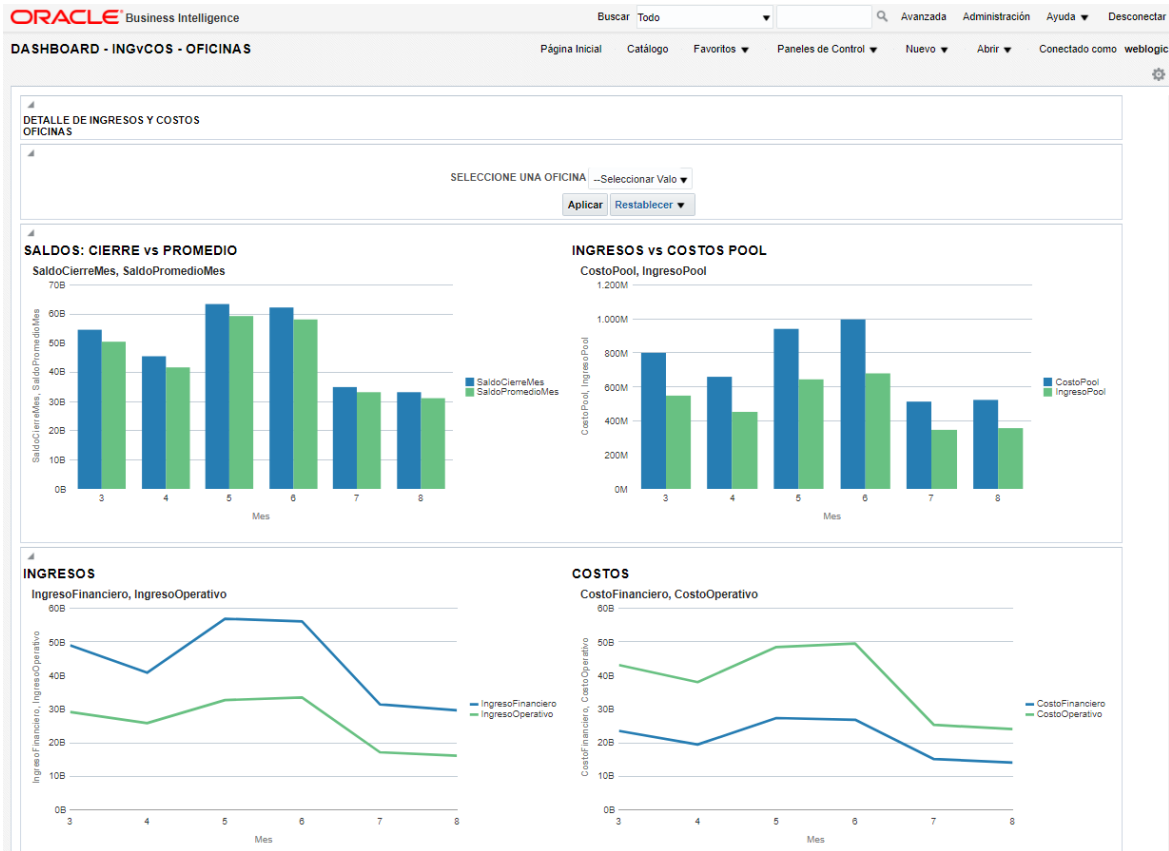


Imagen 29: Dashboard de costos vs ingresos por oficina.

Si bien los cuadros de mando desarrollados permiten visualizar información tanto de alto como de relativamente bajo nivel, las posibilidades de visualización e interrelación de datos es casi infinita, lo que hace inviable modelarlas todas.

De manera que, para proveer a los usuarios de una herramienta que les permita elaborar consultas a demanda (AD-HOC), OBI está provisto de la herramienta para la elaboración de análisis. Herramienta de sencilla utilización pero que a la vez permite infinidad de opciones a los usuarios para representar sus consultas, como tablas dinámicas y gráficos de distintas naturalezas (Imagen 30).

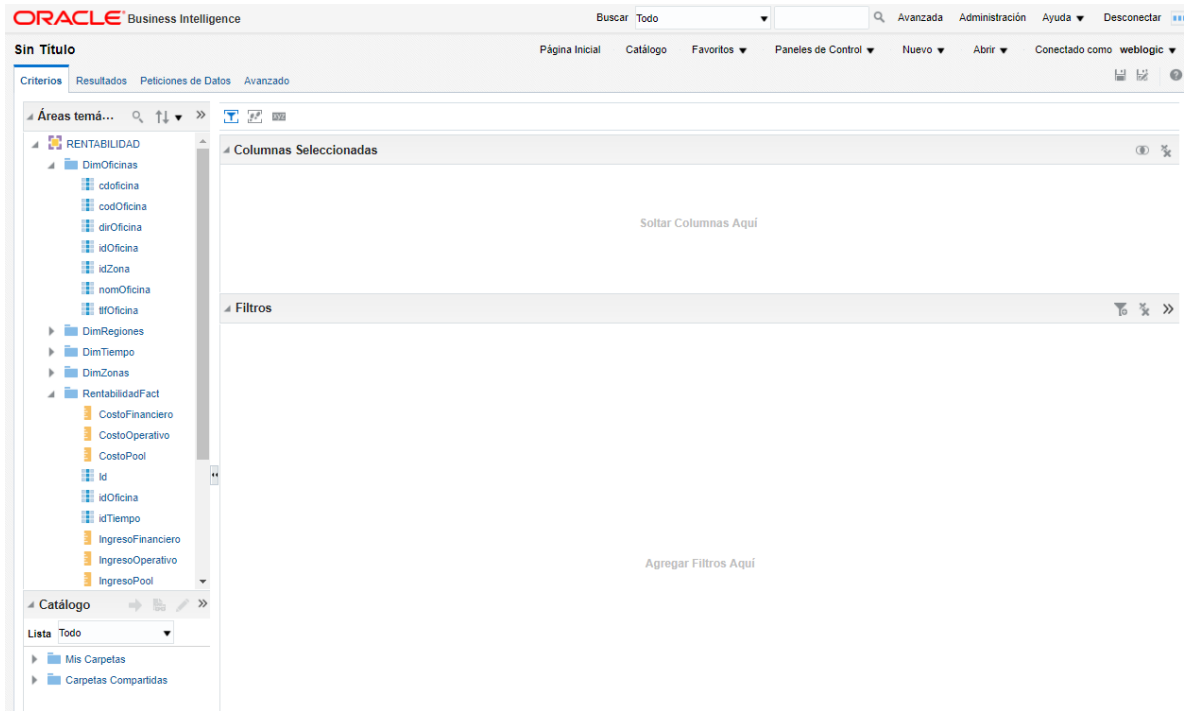


Imagen 30: Pantalla inicial de Oracle Análisis

Como se evidencia en la imagen anterior, a la izquierda se tienen los elementos de datos pertenecientes al modelo de datos y a la derecha los campos de modelado y filtros que permiten añadir o eliminar elementos con facilidad.

A continuación (Imagen 31) un reporte AD-HOC que refleja los saldos promedios por zona, para el mes de Mayo.

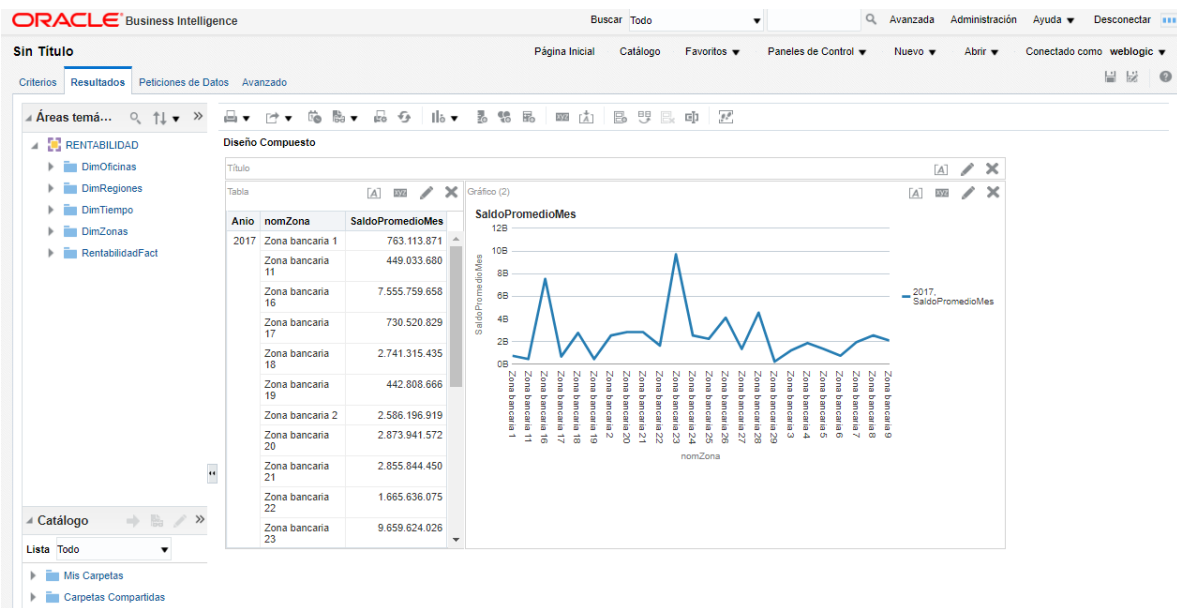


Imagen 31: Análisis completo con gráfico de seguimiento de tendencias.

De esta manera se puede observar que las consultas se realizan de manera rápida y flexible, siendo posible la inserción de gráficos, notas, títulos y tablas.

Verificación de datos e indicadores.

Para verificar la veracidad de los indicadores calculados se utilizó la oficina 103 como ejemplo, tomando las fuentes que alimentan los procesos ETL.

A continuación la secuencia de construcción del indicador de MFB.

(Ingresos pool+financieros)-(Costos Pool + Financieros)

Ingresos Pool (MtoSaldoProm * tasaPoolActiva / (360/30)):

$$((55797917,12)*(0,13))/(360/30) = \mathbf{604477,43}$$

Ingresos financieros (Saldo final de las cuentas 513):

50218125,41

Costos Pool (MtoSaldoProm * tasaPoolPasiva / (360/30)):

$$((55797917,12)*(0,19))/(360/30) = \mathbf{883467,02}$$

Costos financieros (Saldo final de las cuentas 411):

23993104,36

Al sustituir valores en la fórmula, se obtiene:

$$(\mathbf{604477,43+50218125,41})-(\mathbf{883467,02+23993104,36})$$

$$(\mathbf{50822602,84-24876571,38})= \mathbf{25946031,46 MFB}$$

Que al ser comparado con la aplicación resulta **correcto**.

4.1.10 Mantenimiento y Escalabilidad

Una vez completado el desarrollo de la solución, se procede a realizar un conjunto de pruebas para certificar que la información reflejada a través de los reportes e indicadores que se muestran en el portal web interactivo, es consistente con los valores que se encuentran en las fuentes de datos suministradas por la institución financiera.

Luego de esas pruebas, se mantiene el monitoreo de la solución con el fin de corregir los posibles errores que se puedan presentar una vez se encuentre la solución en ambiente de producción. Existe la posibilidad de mejorar el desempeño o la visualización de ciertos indicadores, así como la realización de reportes nuevos.

En conclusión:

- Se da soporte continuo a la solución desarrollada con el fin de corregir posibles incidencias.

- Se analizan, contemplan y desarrollan nuevos requerimientos para la solución analítica desarrollada

CONCLUSIONES

Se han logrado satisfactoriamente los objetivos del Trabajo Especial de Grado, que consisten en el desarrollo de una Solución Analítica para la obtención de indicadores que apoyen la toma de decisiones en el ámbito de la rentabilidad financiera, definiendo e identificando previamente las necesidades del negocio, los indicadores y reportes necesarios y la arquitectura técnica de la solución analítica. También, se definieron las políticas para asegurar la calidad de los datos, para luego poblar el almacén de datos correctamente realizado.

Luego se conceptualizaron y desarrollaron los reportes definidos, después de calcular correctamente los indicadores de rentabilidad pertinentes.

Finalmente se realizaron las pruebas pertinentes para garantizar el correcto funcionamiento de la solución, proporcionando información de calidad basada en los hechos reales del negocio y con ello evitar decisiones sin basamento real, lo cual puede resultar perjudicial para el desarrollo y crecimiento dentro del mundo financiera, gracias a la investigación previamente realizada sobre los diferentes aspectos de la rentabilidad.

La solución desarrollada proporciona una visión integral de la rentabilidad en cada uno de sus aspectos relevantes, permitiendo contrastar puntos de vista e información real y oportuna. Brindando ventajas competitivas sobre otras instituciones de la misma naturaleza.

También reduce significativamente el tiempo en el que se generan los indicadores de rentabilidad, pasando de los tradicionales 20 – 25 días, a solo 5-7 días.

Los indicadores de rentabilidad más relevantes como el MFB, ROE y ROA se reflejan en la solución, permitiendo el estudio exhaustivo de la rentabilidad a diferentes niveles de detalle, haciendo posible la identificación oportuna de eventualidades y su origen.

Es indispensable resaltar que para el desarrollo de la solución, se utilizó la metodología de Ralph Kimball, utilizando diversas herramientas especiales para el desarrollo de aplicaciones de este tipo, entre las que destacan Pentaho Data Integration y Oracle BI.

Finalmente, realizado el proceso dispuesto por la metodología Kimball, se desarrolló la solución analítica que provee a la parte alta de la pirámide organizacional de una herramienta de suma utilidad, tanto para la toma de decisiones como para la planificación a mediano y largo plazo de estrategias financieras que permitan aumentar el rendimiento de los activos de la institución. Además, esta solución permite a los usuarios extraer información de manera dinámica, a través de reportes o consultas a demanda (AD-HOC), generando así información virtualmente infinita.

BIBLIOGRAFÍA

- Laudon, K. and P, L. (2000). *Sistemas de información*. 1st ed. Pearson Custom Publishing.
- O'Brien, J. (2006). *Sistemas de información gerencial*. 1st ed. Bogotá: Irwin/McGraw-Hill.
- Kimball, R. (2008). *The data warehouse lifecycle toolkit*. 1st ed. Indianapolis, IN: Wiley Pub.
- Dueñas Prieto, R. (2008). *Introducción al sistema financiero*.
- Sánchez Ballesta, J. (2002) *Análisis de rentabilidad de la empresa*.
- Mochón y Beker (1998) *Economía, principios y aplicaciones*.
- Díaz Mata y Hernández Almora (1999) *Sistemas financieros mexicano e internacional en internet*.
- Rodríguez y Capece (2001) *El sistema financiero argentino*.
- Armijo, M. (2011) *Planificación estratégica e indicadores de desempeño en el sector público*
- GARCÍA ESTÉVEZ, P. (2008). *El método de valoración por múltiplos*. Colección Finanzas.
- Bauer, R. (1966). *Social indicators*. 1st ed. Cambridge, Mass.: M.I.T. Press.