**UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS**

**FACULTAD DE INGENIERIA DE SISTEMAS E INFORMATICA ESCUELA ACADEMICO PROFESIONAL DE INGENIERIA DE SISTEMAS E INFORMATICA**



**Título de la tesis**

**DESARROLLO DE UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS EN TIEMPO REAL USANDO VIRTUALIZACIÓN DE DATOS PARA LA TOMA DE DECISIONES ORIENTADO AL ANÁLISIS Y MONITOREO DEL TRAFICO POSTAL DE LA EMPRESA SERVICIOS POSTALES DEL PERÚ**

**Trabajo para requerimiento parcial del curso de proyecto de tesis**

**Autor:**

QUIROZ ÑATO, Jorge Enrique

**Asesor**

Piedra Isusqui, José César

Lima – Perú 2016

FICHA CATALOGRÁFICA

QUIROZ ÑATO, Jorge Enrique

DESARROLLO DE UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS EN TIEMPO REAL USANDO VIRTUALIZACIÓN DE DATOS PARA LA TOMA DE DECISIONES ORIENTADO AL ANÁLISIS Y MONITOREO DEL TRAFICO POSTAL DE LA EMPRESA SERVICIOS POSTALES DEL PERÚ

Base de Datos

(Lima, Perú 2016)

Tesis, Facultad de Ingeniería de Sistemas, Pregrado, Universidad

Nacional Mayor De San Marcos

Formato 28 x 20 cm Paginas 145

**UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS**

Universidad del Perú, DECANA DE AMÉRICA

**FACULTAD DE INGENIERÍA DE SISTEMAS E INFORMÁTICA**

**ESCUELA ACADÉMICO PROFESIONAL DE INGENIERÍA DE SISTEMAS**

**DESARROLLO DE UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS EN TIEMPO REAL USANDO VIRTUALIZACIÓN DE DATOS PARA LA TOMA DE DECISIONES ORIENTADO AL ANÁLISIS Y MONITOREO DEL TRAFICO POSTAL DE LA EMPRESA SERVICIOS POSTALES DEL PERÚ**

Autor: Quiroz Ñato, Jorge Enrique

Asesor: Piedra Isusqui, José César

Título: Trabajo para requerimiento parcial del curso de proyecto de tesis

Fecha: Julio del 2016

**RESUMEN**

El presente trabajo de investigación tiene como objetivo desarrollar una solución de inteligencia de negocio (BI) que resuelva los problemas de análisis y monitoreo del trafico postal, en el Área de Postal de la empresa de Servicios Postales del Perú, para llevar a cabo un adecuado análisis de información, entregando una herramienta análisis y monitoreo de la información en tiempo real para que las áreas de Gerencia Postal en la toma de decisiones. Para seleccionar la solución de BI más opta a las necesidades de la empresa, se realizó un análisis comparativo con otras investigaciones similares, y un análisis de las arquitecturas que son usadas para las soluciones de inteligencia de negocio, obteniéndose como resultado el desarrollo de una solución usando virtualización de datos.

**Palabras claves:** Inteligencia de Negocios, Virtualización de datos

**MAJOR NATIONAL UNIVERSITY OF SAN MARCOS**

Universidad del Perú, DECANA DE AMÉRICA

**FACULTY OF SYSTEMS ENGINEERING AND INFORMATICS**

**Academic Professional School of Systems Engineering**

**DEVELOPMENT OF A SOLUTION OF BUSINESS INTELLIGENCE IN REAL-TIME USING DATA VIRTUALIZATION FOR DECISION MAKING ORIENTED ANALYSIS AND MONITORING OF TRAFFIC POSTAL OF THE ENTERPRISE POSTAL SERVICES OF PERU**

Author: Quiroz Ñato, Jorge Enrique

Advisor: Piedra Isusqui, José César

Títle: Trabajo para requerimiento parcial del curso de proyecto de tesis

Date: July 2016

**Abstract**

This research aims to develop solution business intelligence (BI) that solves the problems of analysis and monitoring of postal traffic in the area of Postal Company Postal Services of Peru, to carry out a proper information analysis, a tool providing analysis and monitoring of real-time information for management areas Post in decision-making. To select the BI solution best suited to the needs of the company, a comparative analysis with other similar research, and analysis of architectures that are used for intelligence solutions business, resulting in the development of a solution made using data virtualization.

**Palabras claves:** Business Intelligence, Data Virtualization

Índice

[Capítulo I Planteamiento metodológico 10](#_Toc455744470)

[1.1. Antecedentes 10](#_Toc455744471)

[1.2. Definición del Problema 10](#_Toc455744472)

[1.3. Formulación del problema 11](#_Toc455744473)

[1.4. Objetivos 11](#_Toc455744474)

[1.5. Justificación de la Investigación 12](#_Toc455744475)

[1.6. Justificación de la Técnica 12](#_Toc455744476)

[1.7. Alcance 12](#_Toc455744477)

[1.8. Limitaciones 12](#_Toc455744478)

[1.9. Diferencias con otras investigaciones 12](#_Toc455744479)

[Capítulo II Marco Teórico 13](#_Toc455744480)

[2.1. Inteligencia de Negocios 13](#_Toc455744481)

[2.2. Herramientas de Inteligencias de Negocios 13](#_Toc455744482)

[2.3. Niveles de gestión y toma de decisiones 13](#_Toc455744483)

[2.3.1. Nivel Estratégica 13](#_Toc455744485)

[2.3.2. Nivel táctica 13](#_Toc455744487)

[2.3.3. Nivel Operativo 14](#_Toc455744489)

[2.4. Los almacenes de datos de un sistema de inteligencia de negocio 14](#_Toc455744491)

[2.4.1. Data Warehouse 14](#_Toc455744493)

[2.4.2. Datamart 14](#_Toc455744494)

[2.4.3. Data staging area 14](#_Toc455744495)

[2.4.4. The Operational Data Store 14](#_Toc455744496)

[2.4.5. The Personal Data Stores 14](#_Toc455744497)

[2.5. Modelo Dimensional 15](#_Toc455744498)

[2.6. Virtualización de datos 15](#_Toc455744499)

[Capitulo III Estado del Arte 17](#_Toc455744500)

[3.1. Investigaciones Relacionadas 17](#_Toc455744501)

[3.2. Análisis y Resultados de las Investigaciones 19](#_Toc455744502)

[3.3. Arquitecturas de inteligencias de negocios de las soluciones 23](#_Toc455744503)

[3.4. Análisis y Resultados de las arquitecturas 30](#_Toc455744504)

[Capitulo IV Aporte Teórico 34](#_Toc455744505)

[4.1. Arquitectura propuesta para la solución 34](#_Toc455744506)

[4.2. Metodología usada para la gestión del proyecto de la presente investigación 37](#_Toc455744507)

[4.3. Descripción del proyecto en la empresa 39](#_Toc455744508)

[Referencias: 40](#_Toc455744509)

Índice de Figuras

[Figura 1 - Pirámide de Gestión. Fuente [Rick+12] 13](#_Toc455744255)

[Figura 2 - Arquitectura Tradicional de Inteligencia de Negocios. Fuente [Sahay+08] 23](#_Toc455744256)

[Figura 3 - Arquitectura de inteligencia de negocios en tiempo real. Fuente [Manh+05] 25](#_Toc455744257)

[Figura 4 - Arquitectura de la solución usada en la empresa SERPOST [SERPOST] 26](#_Toc455744258)

[Figura 5 - Arquitectura de un servidor de virtualización de datos. Fuente[Rick+12] 28](#_Toc455744259)

[Figura 6 - Módulos de una capa de virtualización. Fuente [Rick+12] 29](#_Toc455744260)

[Figura 7 - Arquitectura de virtualización de datos propuesta. Fuente[Propia] 34](#_Toc455744261)

[Figura 8 - Base de datos de la empresa y archivos Excel con información de los envíos 35](#_Toc455744262)

[Figura 9 - Envolturas de las tablas seleccionada y de los archivos Excel, que contiene los metadatos y especificaciones de los datos. Fuente[Propia] 35](#_Toc455744263)

[Figura 10 - Mapeo de las envolturas y generación de la tabla virtual. Fuente [Propia] 36](#_Toc455744264)

[Figura 11 - Datamart Virtual con su respectivo esquema de diseño. Fuente [Propia] 36](#_Toc455744265)

[Figura 12 - Conexión entre las aplicaciones de usuarios y el esquema del Datamart. Fuente [Propia] 37](#_Toc455744266)

Índice de Tablas

[Tabla 1 - Matriz de comparación de investigaciones. 20](#_Toc455582328)

[Tabla 2 - Semejanza entre las investigaciones mencionadas y la presente investigación.. 22](#_Toc455582329)

[Tabla 3 - Comparación entre los factores de las arquitecturas. 31](#_Toc455582330)

INTRODUCCIÓN

En los últimos años la información se ha convertido en uno de los activos más importantes, por no decir el más importante, de las organizaciones. Cada día se publican más estudios dirigidos analizar la información con el objetivo de presentar nuevas herramientas para la toma de decisiones que pudieran tener las organizaciones como, por ejemplo, cuando tienen que ordenar más materias primas y que cantidad, la forma de agilizar el proceso de compra y venta, que descuento ofrecerles a los clientes, y si externalizar o internalizar algunos de sus procesos, de esta necesidad de tener, usar y presentar mejor la información que poseen, se creó la inteligencia de negocios.

La razón principal por la cual las empresas desarrollan sistemas de inteligencia de negocio es para apoyar y mejorar sus procesos de toma de decisiones. Esto significa que los principales usuarios de estos sistemas son los que toman las decisiones, en otras palabras, los gerentes de las diferentes áreas de la empresa.

Actualmente existen una infinidad de herramientas y programas que ayudan en el proceso de analizar la información para la toma de decisiones. Sin embargo, este mundo de la toma de decisiones está cambiando, y el cambio más importante es que las organizaciones tienen que reaccionar más rápido. Hay cada vez menos tiempo disponible para tomar decisiones cruciales. El resultado es que tiene que ser posible los cambios de los informes existentes y el desarrollo de nuevos informes con mucha mayor rapidez.

Pero este no es el único cambio. Nuevas fuentes de datos se están haciendo disponibles para el análisis y presentación de informes. Hoy en día, hay muchos más sistemas que ofrecen datos valiosos, tales como weblogs, servidores de correo electrónico, y los sistemas de gestión de documentos. El análisis de estos datos puede conducir a una mejor comprensión de lo que piensan los clientes acerca de los servicios de una empresa y sus productos.

Al igual que en cualquier campo de investigación las nuevas tecnologías se han convertido en puentes hacia nuevas oportunidades y una forma de ampliar estos campos. En la inteligencia de negocios, existen muchas tecnologías disponibles como, por ejemplo, servidores de análisis de bases de datos, herramientas de presentación de información para móviles, análisis dentro de la base de datos, memoria masiva interna, plataformas de hardware altamente paralelizadas y computación en la nube. Todas estas nuevas tecnologías pueden ampliar drásticamente las características analíticas de información de las organizaciones permitiendo analizar los datos en sólo unos pocos minutos que de otro modo tomaría mucho tiempo con las tecnologías más antiguas o tradicionales.

Otro cambio evidente es acerca de los nuevos grupos de usuarios interesados en las herramientas de inteligencia negocio. Actualmente, la mayoría de los usuarios de los sistemas de inteligencia de negocio son los gerentes generales, directivos y gerentes de áreas, en otras palabras, los que se encuentran en los niveles de gestión estratégicos y tácticos. En la mayoría de los casos, estos usuarios pueden trabajar perfectamente con información que no está complemente actualizada hasta la fecha. Los datos que son de un día, una semana, o tal vez de hace un mes son más que suficiente. Lo nuevo es que, actualmente, los responsables de la toma de decisiones del nivel operativo se sienten atraídos por estos sistemas de inteligencia de negocio. Ellos entienden su valor, y por lo tanto quieren explotar el poder de la información y las herramientas analíticas. Sin embargo, ellos quieren analizar la información que esté completamente actualizada o casi completamente actualizada.

Todos estos cambios, especialmente en la toma de decisiones de una forma más rápida, son muy difíciles de implementar en los sistemas de inteligencia de negocio actuales, ya que requiere un rediseño bastante amplio. Esto se debe a que la mayoría de los sistemas de inteligencia de negocio que se han desarrollado en los últimos años se basan en una cadena de bases de datos. Los datos se extraen de bases de datos, transforman, se limpia y se cargar a otras bases de datos, replicando así los datos hasta llegar a un nivel de calidad adecuado para ser presentada en informes y usadas en herramientas analísticas.

Esta cadena de bases de datos y los procesos que extraen, transforman y cargan los datos son largos, complejos y altamente dependientes de las tecnologías con la que se implementaron inicialmente. Cada pequeño cambio que se desea realizar en un informe puede dar lugar a un gran cambio en las cadenas de bases de datos y en los procesos necesarios para generar estos informes. Podría tomar días, semanas o incluso meses antes de que un cambio aparentemente sencillo en los informes, sea implementado en todos los procesos para generar el informe. Además, debido a que muchos procesos de transformación son necesarios y cada uno de ellos lleva tiempo, es difícil entregar datos operacionales a un punto final de la cadena, tales como un datamart o un Data Warehouse. El resultado es que el departamento de inteligencia de negocio no puede mantenerse al día con la creciente velocidad de los cambios requeridos por el negocio. Esto conduce a un impacto negativo sobre la velocidad de la toma de decisiones y por ende la calidad de la organización.

Lo que se necesita es una arquitectura ágil que sea fácil de cambiar, y la mejor manera de hacerlo es crear una arquitectura que posea en un menor número de componentes, lo que significa un menor número de bases de datos y un menor número de procesos de transformación. Cuando se tiene un pequeño número de componentes, hay menos cosas que requieren cambios y es más fácil de gestionar, además de aumentar el nivel de agilidad.

Es en esta situación que la virtualización de datos entra en juego. En resumen, la virtualización es una tecnología alternativa para transformar los datos de distintas fuentes y hacerlas disponibles para presentaciones de informes o análisis. Con la virtualización de datos se necesita un menor número de bases de datos o repositorios para replicación y una menor cantidad de procesos de trasformación para las soluciones de inteligencia de negocio. En conclusión, con la virtualización de datos se simplifican las arquitecturas de inteligencia de negocios y los procesos de transformación dando como resultado una arquitectura ágil y sencilla que pueda responder con más eficiencia los continuos requerimientos de los usuarios.

# Capítulo I Planteamiento metodológico

* 1. **Antecedentes**

Servicios Postales del Perú – SERPOST S.A. es una empresa cuya actividad principal es la prestación de los servicios postales en toda su modalidad, así como, servicios y actividades conexas y complementarias al servicio postal, que opera en el ámbito de acción nacional, como también, internacional.

El objetivo Social de la Empresa es:

1. Prestar el servicio postal en todas sus modalidades en el ámbito nacional e internacional, así como servicios y actividades conexas y complementarias al servicio postal.
2. Prestar servicios de asesoría y consultoría nacional e internacional en materia postal.
3. Realizar actos y contratos relacionados con operaciones comerciales, industriales, de servicios o cualquier otra actividad, conexas con su objeto social, sin más limitaciones que las emanadas del ordenamiento del país.
4. En general toda clase de operaciones afines y/o complementarias permitidas por Ley.

Debido al avance tecnológico ha derivado en un menor uso de la correspondencia tradicional. El uso de la comunicación en tiempo real como la telefonía móvil y correos electrónicos ha reducido la utilización de cartas como medios de comunicación. Sin embargo, es de reconocerse el gran impacto que las tecnologías han tenido en el desarrollo del sector postal. Ya que, gracias al comercio electrónico, este sector ha evolucionado y se ha convertido en una herramienta de integración y acceso a este mercado, que cada día se expande más.

El estado de los procesos para las prestaciones de los servicios postales ha mejorado con la evolución de las tecnologías. Sin embargo, la gestión actual para el análisis del trafico postal y su monitoreo es deficiente y no ha variado en muchos años.

Actualmente la información de los envíos realizados en cualquier oficina postal, diferente a la central, son parcialmente registrados de forma manual y enviados a la central en archivos para su registro en la empresa, mientras que la información de los envíos que llegan desde el extranjero son enviados y registrados de forma completamente manual. Debido a esto, no se tiene un control de los indicadores del tráfico postal.

El establecimiento de las tarifas para los envíos, el control de flujo postal entre las oficinas, la asignación de presupuestos a las áreas, la creación de nuevas formas de envíos, son unas de las muchas tareas importantes que tiene la empresa y que se realizan mediante el análisis y estudios del tráfico postal. Pero debido al estado deficientes de los procesos para los registros y almacenamiento de estos, dificulta la realización de los análisis correspondientes para la toma de decisiones de muchos departamentos y áreas.

* 1. **Definición del Problema**

En este presente caso de estudio, la toma de decisiones de SERPOST es tomada por el área de “Gerencia Postal”, y lo hacen mediante informes de las diferentes áreas y departamentos ligados a tráficos postal, las cuales se encarga de realizar estudios de análisis para brindar información al área de gerencia para la toma de decisiones. Las áreas y departamentos tienen dos tipos de fuentes de información para poder generar sus informes, uno de ellos es la base de datos empresarial y la otra fuente son los registros entregados por las diferentes oficinas postales situadas a lo largo del país.

Tanto la base de datos de la organización como los registros de las oficinas postales poseen diversos problemas que dificultan la realización de los informes.

**Base de datos empresarial:**

* Posee mucha redundancia.
* No está completamente actualizada.
* No poseen parte de la información de los envíos de las oficinas postales.
* Posee información duplicada (mucha redundancia)

**Información de las oficinas postales:**

* La mayoría está en formato Excel.
* Esta parcialmente integrada con la base de datos empresarial.
* No tienen un formato definido

Ocasionando que se tenga que realizar un análisis previo para poder generar los informes para que estos sean entregados al área de gerencia postal. Los departamentos encargados de realizar estos informes, los realizan de manera improvisada y solo sirven para el mismo momento en que se crearon, por lo tanto, no se disponen de información fácil de analizar, oportuna, integral y veras, para generar informes y/o reportes que ayuden los departamentos correspondientes y para que apoye al área de gerencia postal en la toma de decisiones y poder tener un control del estado del tráfico postal.

En vista de los problemas identificados se presenta el siguiente trabajo, destinado al análisis estadístico del tráfico postal y apoyo a la toma de decisiones.

* 1. **Formulación del problema**
     1. **Problema General**

¿De qué manera se puede analizar y monitorear el tráfico postal en tiempo real de los diferentes tipos de envíos para la toma de decisiones de la de la empresa Servicios Postales del Perú?

* + 1. **Problemas Específicos**
* ¿De qué manera se puede integrar las fuentes de información del tráfico postal?
* ¿De qué manera gestionar la información para el análisis y monitoreo en tiempo real del tráfico postal?
* ¿De qué manera la extracción de conocimiento del tráfico postal puede mejora la toma de decisiones?
* ¿De qué manera una herramienta de análisis de información puede mejorar el proceso de análisis del tráfico postal para la toma de decisiones?
  1. **Objetivos**
     1. **Objetivo General:**

Desarrollar una solución de inteligencia de negocios en tiempo real para poder analizar y monitorear del tráfico nacional e internacional de la empresa Servicios Postales del Perú.

* + 1. **Objetivo Específico:**
* Identificar las diferentes fuentes de información que alimentara la solución.
* Seleccionar la tecnología para implementar la solución.
* Diseñar una arquitectura personalizada para la solución.
* Diseñar los reportes de fácil visualización conforme a los requerimientos de las áreas y/o departamentos interesados.
  1. **Justificación de la Investigación**

Debido a que el estado actual de la empresa cuenta con muchas dificultades para poder analizar y monitorear el estado de los envíos, esta investigación podrá dar una solución de inteligencia de negocios al problema para que se realice un mejor análisis y control de los envíos de una manera más integral y oportuna que la gestión actual.

* 1. **Justificación de la Técnica**

Con el uso de una solución de inteligencia de negocios en tiempo real usando virtualización de datos, le permitirá a la empresa tener las condiciones para disponer de información apropiada para la toma de decisiones de una manera mucho más eficiente que la gestión actual, facilitando el trabajo de las áreas y/o departamentos interesados y evitando la demora de generación de informes de análisis, además de poder monitorear el estado de los envíos en tiempo real a nivel nacional e internacional.

* 1. **Alcance**

La presente tesis se centrará en el análisis estadístico y predictivo del tráfico postal. La solución cubrirá la toma de decisiones del área de “Gerencia Postal”, más concretamente en sus subgerencias que son: **Sub-Gerencia de Administraciones Postales**, **Sub-gerencia de Operaciones Postales** y la **Sub-gerencia de Paquetería y Operaciones Aduaneras**, debido a son las más se relacionan al tráfico postal.

El analisis del trafico postal se centrara en los siquietnes puntos:

* Flujo de envios a nivel nacional.
* Flujos de envios a nivel internacional.
* Analizar la distribución por tipos de envios.
* Analizar la demanda por tipos de envios por periodos de tiempo.
* Analizar la demanda por tipos de envios por zonas.
* Analizar la demanda de envios a nivel local, nacional e internacional.
  1. **Limitaciones**

La presente solución contemplara solo los puntos que se relaciones al tráfico de envíos. La solución no contemplara el análisis de los recursos humanos ni contemplara temas financieros excepto los precios de las tarifas de envíos.

* 1. **Diferencias con otras investigaciones**

Los principales puntos que diferencian la presente investigación con otras de inteligencia de negocio o de tráfico postal son:

* Existen varias investigaciones sobre soluciones de inteligencia de negocio aplicadas a empresas en general y varias aplicadas a empresas Courier; sin embargo, estas son aplicadas por lo general a las áreas de finanzas, ventas, compras o de recursos humanos pero muy pocas orientadas hacia el análisis del trafico postal.
* La mayoría de empresas de servicios postales tiene un ámbito regional, siendo la mayoría de casos en Lima, la presente investigación tiene un ámbito más amplio, teniendo sucursales a nivel nacional y contando con convenios a nivel internacional.
* La mayoría de soluciones de inteligencia de negocios son con modelos tradicionales que tienen demora a la hora de la carga de datos para su análisis, la presente investigación posee una solución con una arquitectura ligera y ágil que presente los datos a los usuarios en tiempo real.

# Capítulo II Marco Teórico

En esta sección se definirán los conceptos que relacionados a inteligencia de negocios y virtualización de datos.

* 1. **Inteligencia de Negocios**

La Inteligencia de Negocios es un enfoque y un conjunto de herramientas para la gestión empresarial que permiten a una organización determinar qué información es relevante y útil para la toma de decisiones corporativas. [KIMBALL+02]

* 1. **Herramientas de Inteligencias de Negocios**

Es la gama de capacidades proporcionadas a los usuarios de negocio para aprovechar una estructura donde se organizan y almacenan datos para la toma de decisiones analítica. Una aplicación de BI puede ser tan simple como una herramienta de consulta ad hoc o tan compleja como una aplicación de modelado. [KIMBALL+02]

* 1. **Niveles de gestión y toma de decisiones**

La toma de decisiones está directamente relacionada con los niveles de gestión presente en una organización, la forma más popular y común de identificar estos niveles es con la pirámide de gestión, la cual se muestra en la Figura 1.

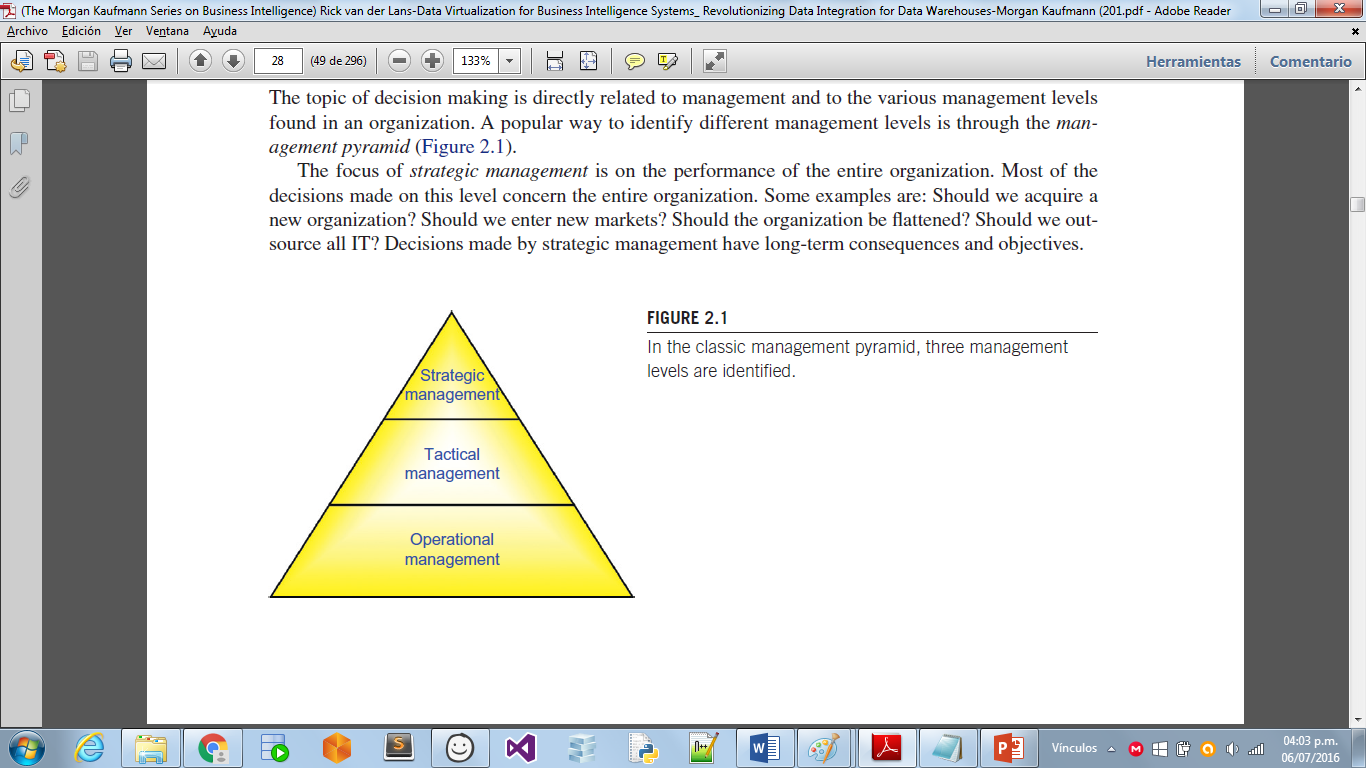


Figura 1 - Pirámide de Gestión. Fuente [Rick+12]

* + 1. **Nivel Estratégica**

Este nivel está enfocado en el rendimiento de toda la organización. La mayor parte de las decisiones que se toman en este nivel están orientadas a preocupaciones de toda la organización. Algunos ejemplos son: ¿Hay que adquirir un nuevo local? ¿Hay que entrar en nuevos mercados? ¿Hay que externalizar todas las áreas de tecnologías de información? Las decisiones tomadas por la dirección estratégica tienen consecuencias y objetivos a largo plazo. [Rick+12]

* + 1. **Nivel táctica**

Este nivel se ocupa de la planificación y el control de las unidades o áreas en una organización, tales como el marketing, las ventas y la producción. Normalmente su toma de decisiones tiene un impacto en el corto o medio plazo.

* + 1. **Nivel Operativo**

Este nivel se ocupa de las operaciones del día a día de la organización. Los ejemplos de las decisiones tomadas en este nivel de gestión son: ¿Cuántas paletas de refresco deberían entregarse a una tienda en particular al día siguiente? ¿Se han previstos suficientes conductores para entregar todos los paquetes? ¿Qué medidas se deben tomar ahora que un avión a llegado al aeropuerto con muchos pasajeros que tienen vuelo por tomar que se han visto seriamente retrasados?

* 1. **Los almacenes de datos de un sistema de inteligencia de negocio**

Los siguientes tipos de almacenes de datos se pueden encontrar en los sistemas de inteligencia de negocio:

* + 1. **Data Warehouse**

El conglomerado de componentes y procesos, donde los datos se estructuran específicamente para el rendimiento de consulta, análisis y facilidad de uso. [KIMBALL 2002]

* + 1. **Datamart**

Es un subconjunto lógico y físico del área de presentación de un Data Warehouse. Originalmente, los Data Mart’s eran definidos como un subconjunto altamente organizado de información, usualmente escogido para responder cuestiones específicas en el negocio. Esta definición no era exacta debido a que afirmaba que los Data Mart’s eran inflexibles y no podían combinarse unos a otros. La primera definición ha sido remplazada y ahora el Data Mart es definido como un juego flexible de datos, basado en la forma de dato más atómica posible para que se extraiga de una fuente operacional, y se presenta en un modelo simétrico que es más resistente cuando se tienen consultas inesperadas de los usuarios. Decimos que los Data Mart’s están conectados al bus del Data Warehouse. Y en su definición más simple un Data Mart representa información de un solo proceso del negocio. [KIMBALL 2002]

* + 1. **Data staging area**

Es un almacén de datos temporal e intermediario en donde se cargan los datos de los sistemas de producción. [Rick+12]

* + 1. **The Operational Data Store**

Es una construcción arquitectónica, donde se almacenan una colección integrada de datos operacionales. Se localiza entre los sistemas de producción o un stagin area si existiera y una Data Warehouse, son usados cuando existen sistemas TI que necesiten de data integrada y transformada. [Rick+12]

* + 1. **The Personal Data Stores**

Es un almacén de datos que está diseñado específicamente para las necesidades de un usuario individual. Un Personal Data Store puede ser una simple hoja de cálculo desarrollado por un usuario para almacenar descripciones de los códigos utilizados en algunas mesas o para almacenar valores de las ventas pronosticadas no disponibles en el almacén de datos. Un Personal Data Store también puede ser un simple archivo que contiene los datos de una organización externa. En la mayoría de los casos, los usuarios desean analizar estos datos, o quieren combinarlo con los datos internos. Algunos informes y herramientas de análisis crean una pequeña base de datos en la máquina del usuario. Esto también puede ser visto como un Personal Data Store. [Rick+12]

* 1. **Modelo Dimensional**

Una metodología para el modelado de datos lógicamente para el rendimiento de consulta y facilidad de uso, que tienen como objetivo presentar los datos en una arquitectura estándar para que los usuarios finales puedan tener un alto rendimiento al momento de acceder a la información. En el entorno de RDBMS, generalmente se implementa con una estructura de una tabla de hechos que se construye con un registro para cada medición discreta. Esta tabla de hechos es entonces rodeada por un conjunto de tablas de dimensiones que describen con precisión lo que se conoce en el contexto de cada informe de la medición. Esta estructura es llamada modelos tridimensionales o esquema estrella. Modelos tridimensionales son también la base lógica para todos los sistemas OLAP. [KIMBALL 2002]

* 1. **Virtualización de datos**
     1. **¿Qué es la virtualización?**

Es la creación, a través de software, de una versión virtual de algún recurso tecnológico, como una plataforma de hardware, memoria ram, un sistema operativo, un dispositivo de almacenamiento, recursos de red entre otros.

Dicho en otras palabras, la virtualización significa que una aplicación puede usar un recurso sin tener que preocuparse por detalles internos como, donde reside, como esta implementada, la plataforma que utiliza.

* + 1. **¿Qué es la virtualización de datos?**

La virtualización de datos es una de esas formas de virtualización. Como el término indica, el recurso que se virtualiza son los datos. En pocas palabras, la virtualización de datos significa que las aplicaciones pueden utilizar los datos sin tener que preocuparse de los detalles técnicos como: dónde reside, como es la interfaz técnica, cómo se encuentra implementado, que plataforma utiliza, y cómo está disponible. Por lo tanto, una solución de virtualización de datos encapsula las diferentes fuentes de datos, de modo que los detalles técnicos estén ocultos y la aplicación puede trabajar con una interfaz sencilla.

**Definición:** es una tecnología que ofrece a los usuarios o consumidores de datos una visión unificada de datos, abstraída, y encapsulada para consultar y manipular los datos almacenados en un conjunto heterogéneo de repositorios de datos. [Rick+12]

Virtualización de datos es un enfoque de integración de datos ágil que utilizan las organizaciones para obtener una visión más clara de sus datos, responder más rápidamente a las siempre cambiantes necesidades de inteligencia de negocios y ahorrar 50-75% durante la duplicación de datos y consolidación.

La virtualización de datos integra datos de fuentes dispersas, en distintas localizaciones y formatos, sin replicar los datos, para construir una capa de datos virtual que facilita la provisión de servicios de datos unificados para dar soporte a múltiples aplicaciones y usuarios. El resultado es un acceso más rápido a todos los datos, una menor replicación y menores costes asociados y una mayor agilidad frente al cambio.

Realiza muchas de las transformaciones y funciones de calidad de datos que utilizan las tecnologías más tradicionales de integración de datos (extracción – transformación – carga, ETL), replicación de datos, federación de datos, enterprise service bus (ESB), etc., pero utilizando tecnología moderna para permitir una integración de datos en tiempo real a un coste más bajo, más rápido y con una mayor agilidad. Puede reemplazar las tecnologías tradicionales de integración de datos y reducir la necesidad de datamarts replicados y data warehouses en muchos casos, pero no completamente.

* + 1. **Ventajas de la Virtualización de datos**
* Gana más conocimiento del negocio mediante el aprovechamiento de todos sus datos.
* Responde más rápidamente a las necesidades de análisis e inteligencia de negocios que una solución de integración tradicional.
* Evita duplicación de datos entre un 50% - 75%, reduciendo la complejidad y ahorran recursos.
* El tiempo de solución es acelerado, los proyectos de virtualización de datos se completan en menor tiempo de que los modelos de integración tradicional.

# Capitulo III Estado del Arte

En esta sección primero se mostrará una comparación entre las investigaciones que resuelven problemas similares a la presente investigación, luego se mostrarán las arquitecturas de las soluciones de inteligencia de negocios que pretenden resolver el problema de la presente empresa y la arquitectura de la solución que se plantea en esta investigación.

* 1. **Investigaciones Relacionadas**

A continuación, se mencionarán investigación que han generado soluciones a problemas similares.

* + 1. **ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS PARA EL ÁREA DE IMPORTACIONES EN UNA EMPRESA COMERCIALIZADORA/IMPORTADORA**

**PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ – Lima – 2013**

**Tesis realizada para optar por el Título de Ingeniero Informático**

El autor de esta investigación José Eduardo Córdova Yupanqui señala que las empresas comercializadoras al tener información oportuna y útil, les permitirán cumplir con sus objetivos específicos a mediano y largo plazo.

Su investigación aborda los problemas de la falta de un correcto análisis de los datos de las importaciones. La diversificación de los sistemas y la falta de comunicación e integración de ellos además de la falta de herramientas que unifiquen los datos, ocasionan que los análisis que se realizan pierdan valor por el tiempo que toman las áreas de compra, venta y almacén para organizar e integrar la información de manera manual, además que se añaden errores humanos por la manipulación de estos. Se tiene información desactualizada por la falta de integración que impide un correcto análisis para la toma de decisiones que permita a la empresa cumplir con sus objetivos.

Por los problemas vistos se presenta una solución de inteligencia de negocios que se ajuste a la realidad de la empresa y a sus requerimientos para el área de importación; de modo, que se pueda integrar y analizar la información de una manera más eficiente y se pueda cumplir con las metas de la empresa.

**Análisis:**

En esta investigación muestra como una solución de inteligencia de negocios con organizar e integrar las diferentes fuentes de datos a través de un proceso ETL, para posteriormente generar los datamart necesarios para la posterior explotación de los datos y creación de los reportes necesarios para la solución. La solución permitió un sistema de explotación, análisis de los datos y generación de reportes que apoyen a las diferentes áreas para poder obtener información actualizada y útil para la toma de decisiones. La solución cumplió con los objetivos de integración de la información dispersa y análisis de la información de las importaciones, además de crear una arquitectura de fácil mantenimiento.

* + 1. **PROPUESTA DE UN SERVICIO DE MENSAJERÍA EXTERNA PARA LA EMPRESA CP**

**INSTITUTO POLITÉCNICO NACIONAL - México. DF – 2010**

**Tesina para optar por los títulos de Ingeniero en Transporte, Licenciado en Administración Industrial, Ingeniero en Informática, Ingeniero Industrial**

Los autores de esta investigación Lisbeth García Morales, Luis Miguel Guerrero Durán Miriam Molina Velasco, Juan Gutiérrez Gutiérrez y Ricardo Daniel Pérez Navarro señalan que las empresas requieren la información de las diversas áreas que la componen para logran sus objetivos y el departamento de logísticas que se componen de diversas áreas como servicio al cliente, inventario y transporte de mensajería, no son la excepción.

La presente investigación, constituye un estudio que describir una estrategia de mejora para el servicio de mensajería interna con el que cuenta la empresa CP, en virtud de que, pueda realizar la implantación del mismo como un servicio externo al fin de ofrecer un nuevo servicio a sus clientes.

En esta investigación se abordan los problemas de deficiencia en los procesos de distribución de los envíos que evita que se aproveche completamente este proceso para brindar un mejor servicio al público en general y tener información de los envíos de manera más oportuna, para formular una estrategia de implementación de un servicio de mensajería externa.

Debido a esto se propone una solución de inteligencia de negocios basado en la herramienta Microsoft Suite BI (SQL Server Integration Services, SQL Server Analysis Services, SQL Server Reporting Services, SQL Server Management Studio) para resolver los inconvenientes encontrados en la operación actual del área de mensajería de paquetes y tener información oportuna del estado de los paquetes que fueron distribuidos además de tener un control y monitoreo de los envíos que fueron entregados.

**Análisis:**

Esta investigación muestra a Microsoft Suite BI como una herramienta capas de general reportes dinámicos que permite consultar el número de envíos o paquetes que deben de llegar a la central de la empresa o al destino estipulado, también se usaría para generar reportes o vistas de los paquetes, su ubicación, su estado, destino, origen, por fecha, por tienda, etc. También sirve como una herramienta para análisis de información a través de la tecnología OLAP mediante la implementación de un datamart. La solución permite tener un análisis y monitoreo de los envíos o paquetes para poder obtener información real y precisa para resolver los problemas existentes en la empresa. La investigación cumplió con el objetivo principal de Formular las estrategias para implementar el servicio de mensajería externa en la empresa CP aprovechando la capacidad instalada.

* + 1. **SISTEMAS DE INFORMACIÓN PARA EL ÁREA DE LOGÍSTICA, BASADO EN UN DATA WARAHOUSE APLICADO AL SEGUIMIENTO DE LAS SOLICITUDES DE REEMBOLSO POSTAL**

**Universidad Federal de Santa Catarina – Florianópolis – Brasil – 2004**

**Tesis para optar por el grado de Maestría en Ingeniería de Producción**

El autor Ricardo Guilherme Radünz Filho señala en su investigación que las empresas textiles necesitan más que nunca tener un mejor control del flujo de sus ventas para que los ejecutivos puedan tomar mejores y más rápidas decisiones para cumplir con los objetivos de la organización. Señala también que hay un problema en el área de logística para el análisis del flujo de las ventas, debido a que las herramientas actuales en la empresa generan informes con formato muy complicado y la falta de información actualizada e integrada genera problemas a la hora de tomar decisiones.

Frente a esto se propone una solución de inteligencia de negocio usando un datamart, para la empresa ViaBlu, para el procesamiento de las solicitudes de reembolso postal, permitiendo una reducción de costes y un mayor seguimiento de las solicitudes de los envíos y recepción de los clientes. La información que se analizara serán los contactos de todos los clientes potenciales de la empresa, las solicitudes de los pedidos realizados por estos, las facturas generadas por las ventas de las mercancías que se distribuirán y los plazos de entrada de los envíos hacia los clientes.

**Análisis:**

En esta investigación se presenta un datamart como una herramienta para el análisis de información de los envíos para la toma de decisiones de la parte de gerencia, se logró un mejor control de las entregas de los envíos a los clientes y una reducción del tiempo de entrega de estos. Se pudo generar reportes sobre el flujo de las ventas realizadas y un análisis de los envíos distribuidos y recibidos. Se dejó como trabajo futuro el poder analizar el flujo de los envíos para el seguimiento de estos a través de la web para que muchos usuarios puedan consultar esta información.

* 1. **Análisis y Resultados de las Investigaciones**
     1. **Matriz de Comparación de Investigaciones**

En la Tabla 1 se mostrarán las comparaciones de las investigaciones mencionadas anteriormente, donde las columnas serán las investigaciones y las filas los criterios de comparación. A continuación, se enumerarán los criterios de comparación.

1. **Proceso de negocio**

Se refiere al proceso o actividad de negocio en la cual se ha identificado un problema en donde se desea dar solución.

1. **Sistemas fuentes involucrados**

Se refiere a todos los sistemas de información propios de la organización que están involucrados en la investigación.

1. **Problemas identificados**

Se refiere a la cuestión que se plantea en la investigación, la cual se espera resolver.

1. **Objetivo a alcanzar**

Es la meta definida que cubrirá con las necesidades del negocio y dará solución al problema identificado.

1. **Solución creada**

Es la solución de inteligencia de negocio que resuelve los problemas identificados en la investigación y satisface los quebramientos del negocio.

1. **Esquema de Diseño usado en la solución**

Esta es el esquema de diseño usado en la solución. Por ejemplo: Estrella, copo de nieve, constelación, etc.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Empresa comercializadora | Empresa CP | Empresa ViaBlu |
| Proceso de Negocio | Integración y consolidación de los datos de las importaciones | Análisis y monitoreo de los estados de los envíos para formular una estrategia de implementación de un servicio de mensajería externa. | Análisis del flujo de las ventas del área de logística |
| Sistema fuentes involucrados | Sistema transaccional de información de los productos, los clientes y los envíos solicitados | Sistemas transaccionales de información de las diversas tiendas que posee la empresa, información de clientes y del departamento de mensajería. | Sistemas transaccionales de información de clientes, ventas, facturas por envíos, catálogo de productos, lista de solicitudes y costo por transporte usado. |
| Problema identificado | * Tarea manual en la extracción e integración de los datos. * Excesivo tiempo en la generación de reportes. * Diversificación y falta de comunicación de sistemas fuentes. | * Falta de información consolidada del volumen de envíos que llegan a la bodega y salen de ella. | * Generación de informes complejos y de difícil entendimiento. * Falta de información actualizada para el análisis. |
| Objetivo a alcanzar | Realizar el análisis, diseño e implementación de una solución de inteligencia de negocios para el área de importaciones. | Desarrollar una solución de inteligencia de negocio que ayude a determinar la viabilidad de la implementación de un servicio de mensajería externa. | Implementación de un nuevo sistema de información para el control de pedidos por Rembolso Postal a partir de la reestructura actual e implementación en paralelo del sistema de control existente. |
| Solución creada | Varios datamarts para el análisis de datos de las importaciones. | Un datamart del flujo de los envíos que llegan para el departamento de mensajería. | Un datamart para el análisis y control de las solicitudes de reembolso Postal. |
| Esquema de Diseño usado en la solución | Se utilizó el esquema estrella para cada datamart, en donde cada datamart estaba constituido por su tabla de hechos y sus respectivas dimensiones. | Usa un esquema de estrella para aumentar el rendimiento de consultas y tener las tablas del datamart los más des-normalizadas posibles. | Usa un esquema mixto, dependiendo del nivel de granularidad de los datos va normalizando o des-normalizando el esquema para mejorar las consultas. |

Tabla 1 - Matriz de comparación de investigaciones. Fuente [elaboración propia]

* + 1. **Análisis y Resultado de los Criterio de Comparación**

**Para el Análisis:**

En esta sección se describe los criterios que servirán para comparar las investigaciones ya nombradas con la presente investigación:

* **Criterio 1:** El proceso o actividad de negocio investigado está relacionado al análisis y monitoreo del tráfico postal.
* **Criterio 2:** Los sistemas de información involucrados en la investigación se relacionan al módulo de mensajería o son de trafico
* **Criterio 3:** Las cuestión o problemas que sean identificado en la investigación se relaciona con.
  + Diversificación y falta de comunicación entre la información del trafico postal.
  + Falta de información consolidada para la toma de decisiones.
  + Falta de sistema que extraigan información y la convierta en conocimiento.
  + Falta de sistemas que generen reporte de manera rápida, correcta y oportuna.
* **Criterio 4:** Si la solución creada o desarrollada en la investigación se esté relacionada al análisis y monitoreo del trafico postal.

**Criterio 5 y 6** no se pueden comparar puesto que aún no sea implementado la solución de la presente investigación.

Tomando todo lo visto en la tabla anterior se procederá a realizar el análisis de semana de las investigaciones con la presente investigación, definiendo los niveles de semejanza siguiente.

* **Bajo**
* **Medio**
* **Alto**

**Resultado del Análisis:**

En la Tabla 2 se observa los niveles de semejanza entre las investigaciones mencionadas y la presente investigación

|  |  |  |  |
| --- | --- | --- | --- |
| Criterios de comparación | Empresa comercializadora | Empresa CP | Empresa ViaBlu |
| Proceso de Negocio | **Medio** | **Alto** | **Bajo** |
| Sistema fuentes involucrados | **Medio** | **Alto** | **Medio** |
| Problema identificado | **Alto** | **Bajo** | **Medio** |
| Objetivo a alcanzar | **Medio** | **Bajo** | **Medio** |

Tabla 2 - Semejanza entre las investigaciones mencionadas y la presente investigación. Fuente Propia.

De la tabla 2 tenemos lo siguiente:

* Criterio 1: En la investigación realizada en la empresa CP obtuvo una calificación alta debido a que el proceso de negocio en que se enfoque es el análisis y monitoreo de los estados de los envíos que guarda mucha relación con la presente investigación. Por otro lado, las investigaciones de la empresa comercializadora y la empresa Viablu no guardada mucha relación, enfocándose en procesos que tiene poca relación o casi ninguna.
* Criterio 2: Los sistemas de información de la empresa CP obtuvo un puntaje alto debido a que los sistemas investigados tienen meran del mulo de envió principalmente, en donde se guardaban la información de los destinos de los envíos, los clientes de estos, las tiendas en donde serían distribuidos entre otros.
* Criterio 3: Los problemas encontrados en la empresa comercializadora obtuvo un puntaje alto debido que estos tenían mucha relación con los problemas de la presente investigación en los siguientes:
  + Tarea manual en la extracción e integración de los datos.
  + Excesivo tiempo en la generación de reportes.
  + Diversificación y falta de comunicación de sistemas fuentes
* Criterio 4: En ninguno de los objetivos de las investigaciones mencionadas obtuvieron puntajes altos debidos a que no apuntaban a el objetivo principal de la presente investigación, todos se relacionaba en cierta medida como los de los casos la empresa comercializadora y de la empresa Viablu, y otro como el caso de la empresa CP tenían muy poca relación.

Por lo analizado se considera que la investigación de la empresa comercializadora es más semejante a la presente investigación que los otros dos casos analizados.

Por lo tanto, se considerará los criterios 5 y 6 de las investigaciones en la presente instigación, que son la solución propuesta y el esquema de diseño.

* + 1. **Aspectos recogidos para la presente investigación**

A continuación, se describen los aspectos recogidos de las investigaciones analistas:

* Usar la técnica de datamart para la explotación de datos de las diferentes fuentes de información.
* Utilizar el esquema estrella en la solución planteada para poder analizar los datos de una manera eficiente.
  1. **Arquitecturas de inteligencias de negocios de las soluciones**
     1. **Solución Tradicional de Inteligencia de Negocios**

En la Figura 2 se muestra la solución más implementada en la mayoría de proyectos de inteligencia de negocios; si bien cada uno tiene ligeras diferencias, todas siguen el esquema de la arquitectura que se mostrara a continuación, esta solución presenta 4 capas como la que vemos en la siguiente imagen:

**Fuente:** Real time business intelligence in supply chain analytics

**Autor:** B.S. Sahay and Jayanthi Ranjan (2008)

**Repositorio:** Librería digital de ACM

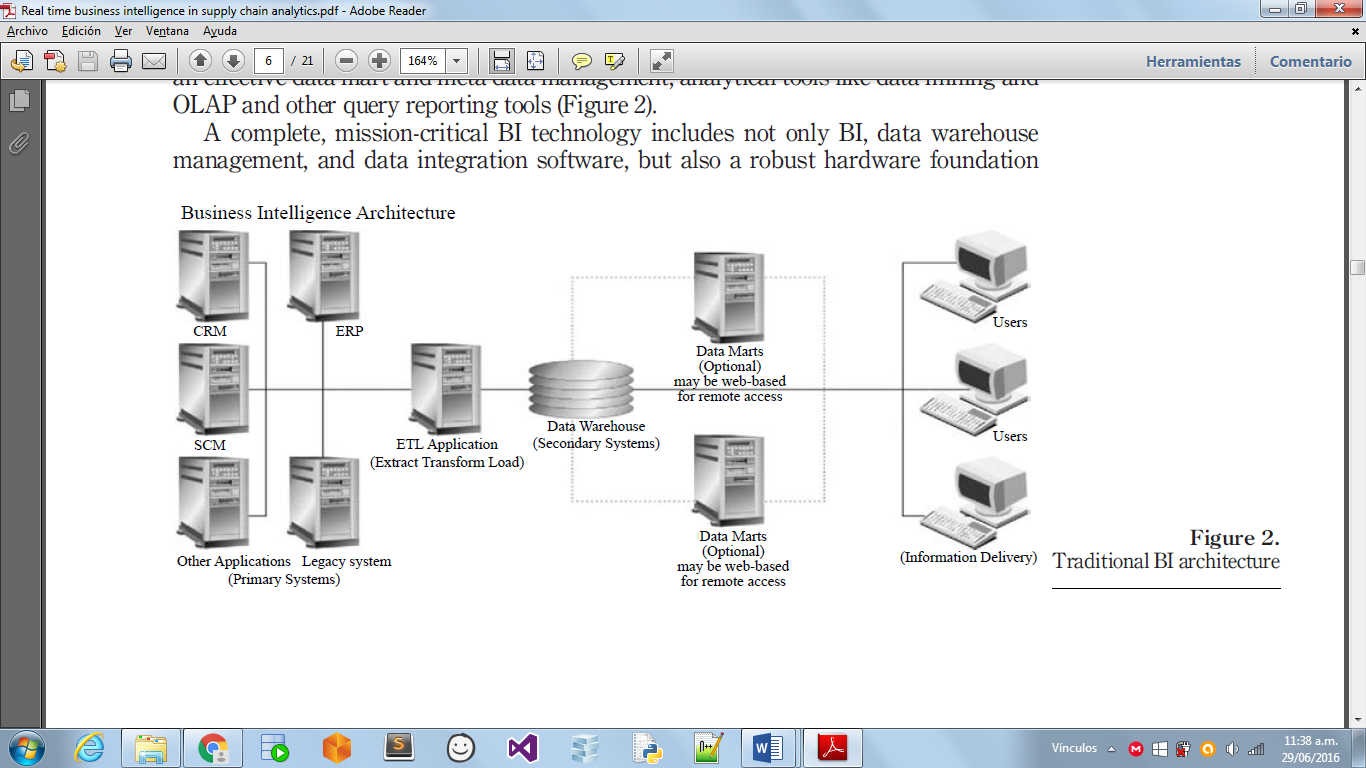


Figura 2 - Arquitectura Tradicional de Inteligencia de Negocios. Fuente [Sahay+08]

En esta investigación se realiza un estudio de cómo está compuesta una arquitectura clásica de inteligencia de negocios, enumerando sus características, ventaja y desventajas.

**Arquitectura:**

**Capa de Sistemas operacionales:** Son los sistemas operacionales de almacenamiento que captan los datos de las transacciones de la empresa. Consta de datos brutos y archivos.

**Capa de ETL:** Es la capa en donde se procesa información, en donde se realiza la operación de ETL (extracción transformación y carga). En esta capa se extrae información de los diferentes sistemas transaccionales y se realiza un proceso de integración y limpieza y organización de los datos en un repositorio intermedio llamado “Staging area” antes de enviarlo a la capa siguiente.

**Capa de Presentación:** Es la capa donde los datos se organiza, se almacenan, y se ponen a disposición para consulta directa por los usuarios, los generadores de informes, y otras aplicaciones de inteligencia de negocios analíticas. Aquí es donde se encuentran los datamart y Data Warehouse. Luego de cargar los datos en los sata wahouse y datamart existentes, se van cargando periódicamente los datos en las estructuras OLAP para el análisis de los datos y posteriormente donde se tomarán los datos para la presentación.

**Capa de aplicaciones de inteligencia de negocios:** Es la última capa y uno de los principales componentes de la arquitectura. Es la capa donde se encuentras las aplicaciones de inteligencia (BI) desde aplicaciones simples como la creación de informes ad hoc como aplicaciones complejas de minería de datos.

**Análisis:**

La arquitectura tradicional de inteligencia de negocios consta de muchos componentes, todos necesario para el procesamiento y presentación de información a los clientes, lo cual le confiere una naturaleza poco flexible y con la característica de ser muy dependiente de cómo está estructurada su arquitectura. Si se quiere realizar un cambio en la forma del análisis o el formato de los reportes para la toma de decisiones, se tiene que reestructurar la arquitectura en muchos de sus componentes. Esta arquitectura está dirigida para el análisis y la presentación de información para los niveles organizacionales estratégicos y tácticos, no soporta el nivel operativo debido que para este nivel se necesitan tener datos frescos y en tiempo real. Las actualizaciones de datos se realizan de forma periódica requiriendo la intervención humana.

* + 1. **Solución de Inteligencia de negocios en tiempo real**

En la Figura 3 se muestra la arquitectura de inteligencia de negocios presentada por Manh Nguyen, Josef Schiefer y Min Tjoa en 2005.

**Fuente:** Sense & Response Service Architecture (SARESA): An Approach towards a Real-time Business Intelligence Solution and its use for a Fraud Detection Application

**Autor:** Tho Manh Nguyen, Josef Schiefer and Min Tjoa (2005)

**Repositorio:** Librería digital de ACM

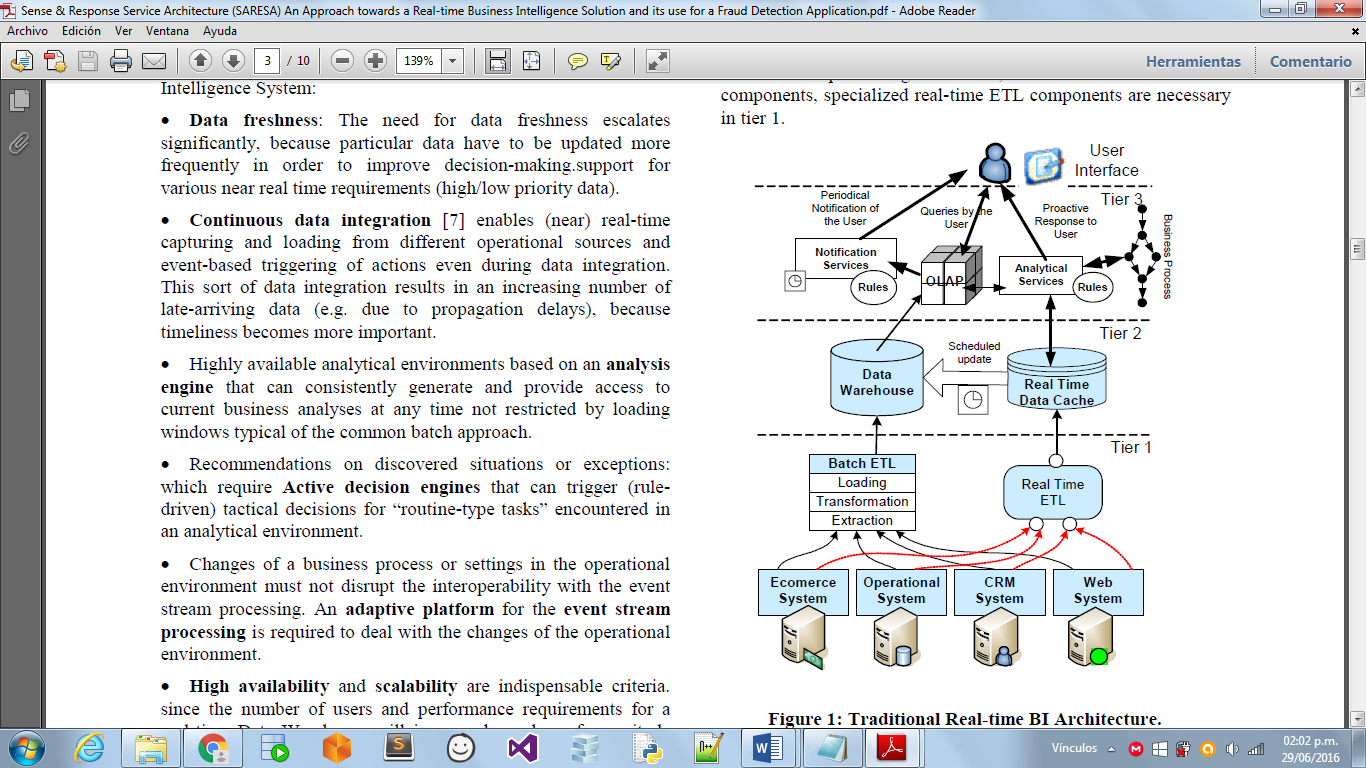


Figura 3 - Arquitectura de inteligencia de negocios en tiempo real. Fuente [Manh+05]

En esta investigación muestra como la arquitectura de inteligencia de negocios en tiempo real se puede adaptar a para una solución para detección de fraude, en ella se analiza las características de la arquitectura de inteligencia de negocios en tiempo real.

La idea de la inteligencia de negocios en tiempo real surgió de la necesidad de las empresas te poder tener información con bajo nivel de latencia, en otra palabra que se actualizara con mínimo retraso de tiempo para que las empresas pudieran reaccionar en base a la información actualizada.

La idea básica detrás de una estrategia inteligencia de negocios en tiempo real es acelerar el flujo de información y procesos de negocios para lograr una ventaja competitiva. Esto implica que todas las áreas de la empresa pueden responder a eventos tan pronto como se conozcan en cualquiera de ellas.

Esta modelo de solución de inteligencia de negocios trata de solucionar el problema de tener información en tiempo real. En otras palabras, la información de la empresa o área de la empresa, está en constante cambio. La información que se ingresa en los sistemas fuentes o sistemas transaccionales de la empresa son automáticamente integradas, transformadas y cargadas hacia los Data Warehouse’s y datamarts por un proceso ETL continuo, estos a su vez actualizan información periódicamente hacia las estructuras OLAP para que finalmente ellos brinden de información para las aplicaciones de inteligencia de negocios a los usuarios.

**Arquitectura:**

**Capa 1:** En esta capa los datos de las fuentes de datos se extraen, transforman y se cargan en el almacén de datos a través de los componentes de ETL. Esta capa presenta una descripción similar a la arquitectura tradicional, pero posee un componente adicional de ETL que procesa información continuamente y carga datos a un repositorio llamado “cache de datos en tiempo real”.

**Capa 2:** En esta capa los datos son almacenados en el Data Warehouse y el cache de datos en tiempo real. La cache de datos en tiempo real se encarga de consolidar los datos del proceso ETL, actualizar el Data Warehousey actualizar las estructuras OLAP en forma continua y con mínimo de latencia.

**Capa 3:** Esta capa es similar a la capa de presentación de la arquitectura tradicional de inteligencia de negocios, sirve como almacén para proporcionar un entorno para el análisis de datos y como fuente de información que alimenta las aplicaciones de los usuarios.

La arquitectura presentada no muestra la capa de las aplicaciones de usuarios o aplicaciones de inteligencia de negocio, esto lo hace porque esta capa no se diferencia con la presentada en la arquitectura tradicional.

**Análisis:**

En la arquitectura de inteligencia de negocios en tiempo real trata el principal problema que sufre la arquitectura de tradicional de inteligencia de negocios. El problema principal es de contar con información de las áreas de la empresa de manera rápida sin requerir una carga de datos periódica; en otras palabras, los procesos de ETL y de carga de los datos desde el Data Warehouse hacia las estructuras necesarias para el análisis, se realizan automática sin querer ningún proceso manual. Esta arquitectura necesita de componentes extra para poder realizar los análisis y actualizaciones de datos en tiempo real, lo cual genera que esta arquitectura sea mucho más compleja que la tradicional.

* + 1. **Solución usada por SERPOST**

En la Figura 4 se muestra la solución usada actualmente en la empresa SERPOST.

**Fuente:** SERPOST

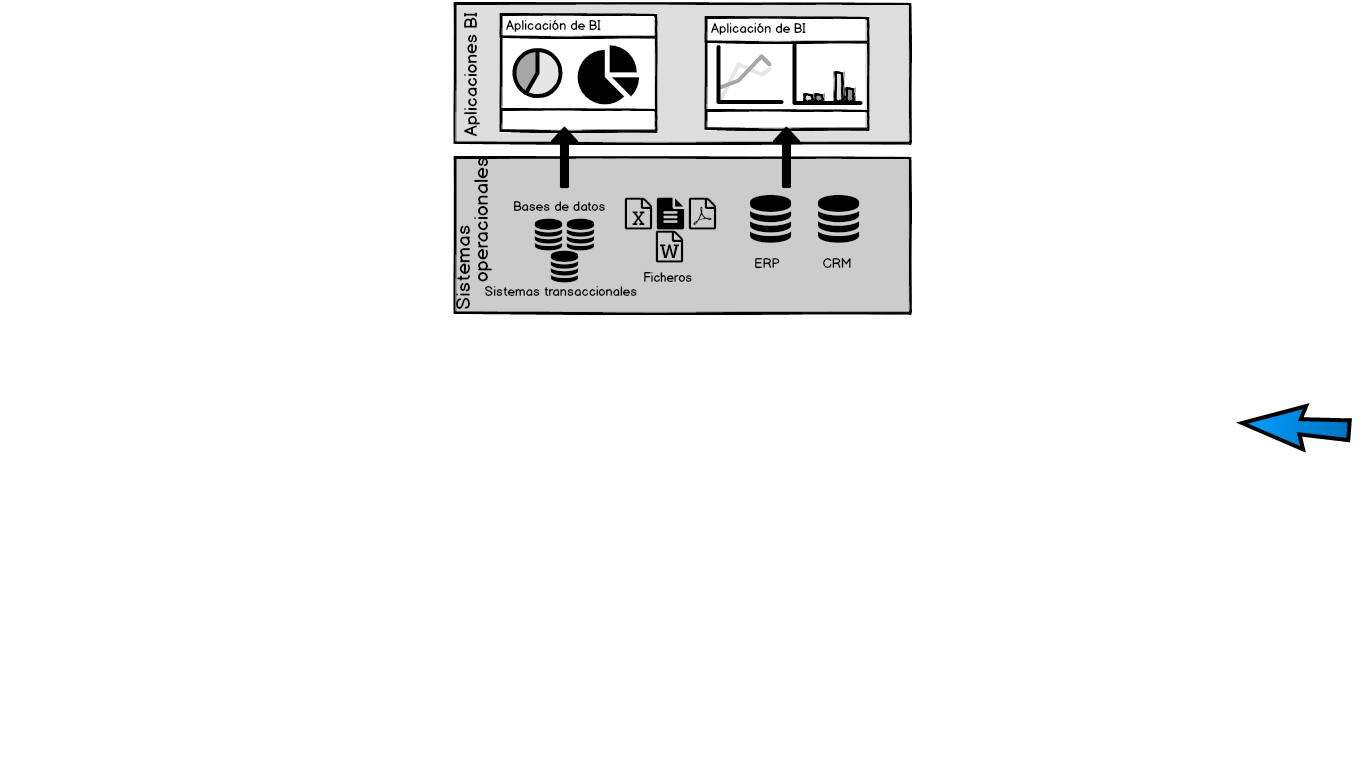


Figura 4 - Arquitectura de la solución usada en la empresa SERPOST [SERPOST]

La solución usada por SERPOST consta de 2 capas:

**Capa de aplicaciones de inteligencia de negocio:** En esta capa se encuentran todos los sistemas de información que dispone la empresa para sus tareas operativas

**Capa de sistemas transaccionales u operacionales:** En esta capa están las aplicaciones de inteligencia de negocio que usa SERPOST.

El principal objetivo que busca este tiempo de arquitectura es generar reportes simples conectándose directamente a la base de datos, debido a que carece de otro componente más, no puede realizar análisis complejos acerca de los datos presentes en los sistemas transaccionales.

**Análisis:**

La solución usada por SERPOST no dispone de repositorio de datos, como data warehouse y datamart, que optimicen las consultas para grandes cantidades de datos, además que tampoco posee estructura que permitan un análisis de por áreas de empresas o por temáticas como cubos OLAP, todas estas características presentan los siguientes problemas:

* Bajo rendimiento de consultas.
* Demora en la generación reportes para presentación de información.
* Imposibilidad de análisis de datos históricos para generar conocimiento.
* Trabajar con datos inconsistentes o erróneos.
* Al grado de redundancia de datos.
* Al grado de dependencia con las tecnologías de los sistemas transaccionales u operativos.

Este tipo de arquitectura suele usarle cuando no hay un área de inteligencia de negocios definida en la empresa.

* + 1. **Solución con Virtualización de Datos**

En la Figura 5 se muestra el diagrama de una arquitectura de virtualización de datos.

**Fuente:** Data Virtualization for Business Intelligence Systems Revolutionizing Data Integration for Data Warehouses

**Autor:** Rick F. van der Lans (2012)

**Repositorio:** Librería digital ScienceDirect

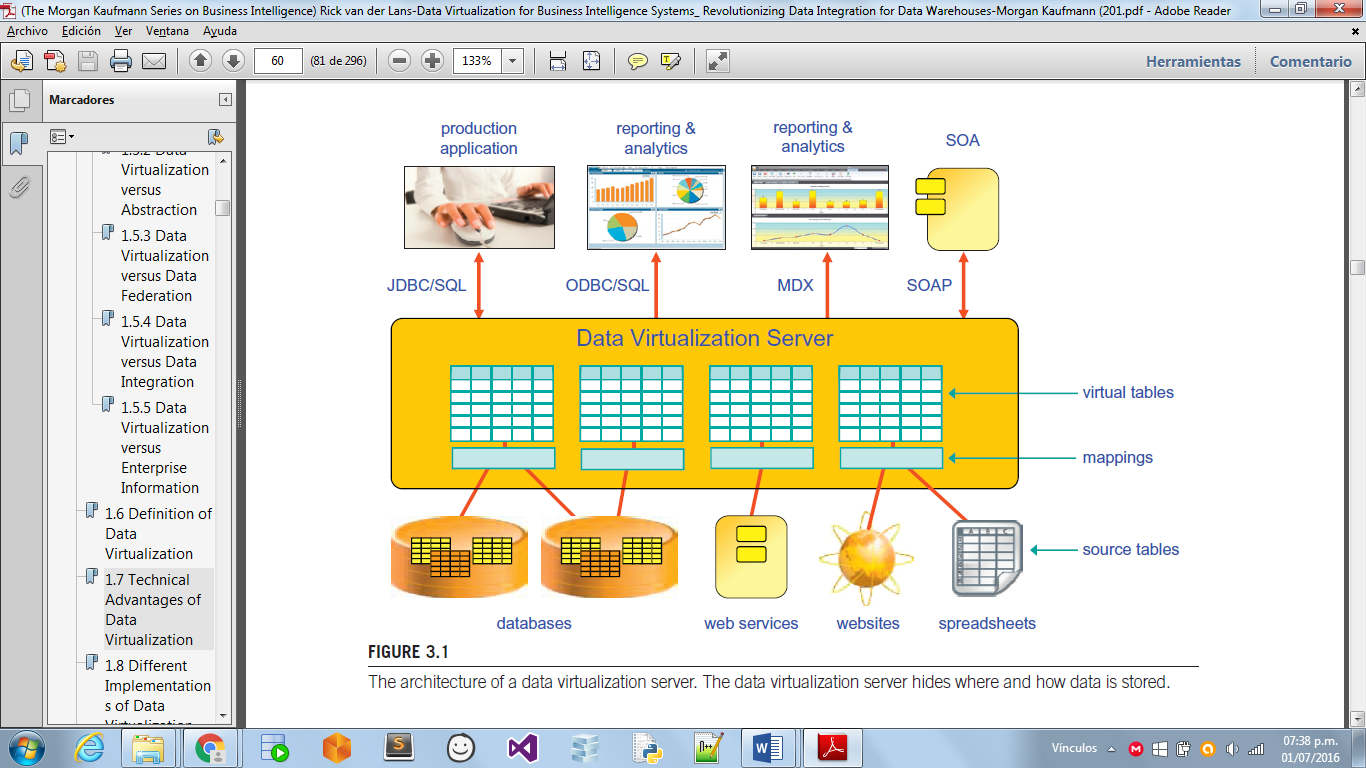


Figura 5 - Arquitectura de un servidor de virtualización de datos. Fuente[Rick+12]

Este libro especifica todos los conceptos relacionados a la virtualización de datos desde los tipos de virtualización de datos, formas de implementación, comparación con sistemas clásicos o tradicionales de inteligencia de negocio, ventajas y desventajas.

La arquitectura de virtualización de datos de 3 capas, la primera están todas las aplicaciones de los usuarios tanto las aplicaciones de análisis como de generación de reportes, la segunda que esta la capa de virtualización y la tercera que es donde se encuentran todos los sistemas transaccionales, documentos planos, datos albergados en servicios web, hojas de cálculo y cualquier sistema que la empresa usa para almacenar sus datos operativos.

Internamente, cada capa de virtualización de datos tiene su propia arquitectura a medida, pero en un nivel conceptual, todos ellos constan de al menos dos módulos principales: un **módulo de diseño** y un **módulo de tiempo de ejecución**. Además, hay un diccionario para almacenar las especificaciones de los metadatos. Estos tres se especificarán a continuación:

* **El módulo de diseño:** es donde los analistas, diseñadores, y posiblemente a algunos los usuarios pueden introducir especificaciones de virtualización. Aquí es donde las especificaciones tales como las definiciones de conceptos; modelos y esquemas de datos, los formatos de integración; y las especificaciones para la transformación, limpieza, e integración son introducidos. Todas estas especificaciones se almacenan en un diccionario.
* **El módulo de tiempo de ejecución:** es el módulo que es utilizado cuando los consumidores de datos o usuarios que necesiten información acceden a la capa de virtualización. En este módulo que manejan todas las peticiones entrantes. Se determina la mejor estrategia para el funcionamiento de las consultas, este módulo sabe cómo acceder a diferentes tecnologías de almacenamiento de datos, sabe cómo integrar los datos y como transformarlos. Se compone de un programador, un motor de almacenamiento, un gestor de memoria caché y un optimizador de consultas. Este módulo se encarga de que cualquier cambio, los sistemas de información transaccionales o bases de datos de la empresa, sea actualizado en la capa de virtualización en un tiempo mínimo de latencia, lo que mantiene los datos frescos y actualizados para el análisis y la creación de reportes.
* **Diccionario de Metadatos:** es donde se almacenaran todas las especificaciones de los metadatos de los datos que son mapeados de los sistemas transaccionales.

En la Figura 6 se muestran los módulos de la capa de virtualización:

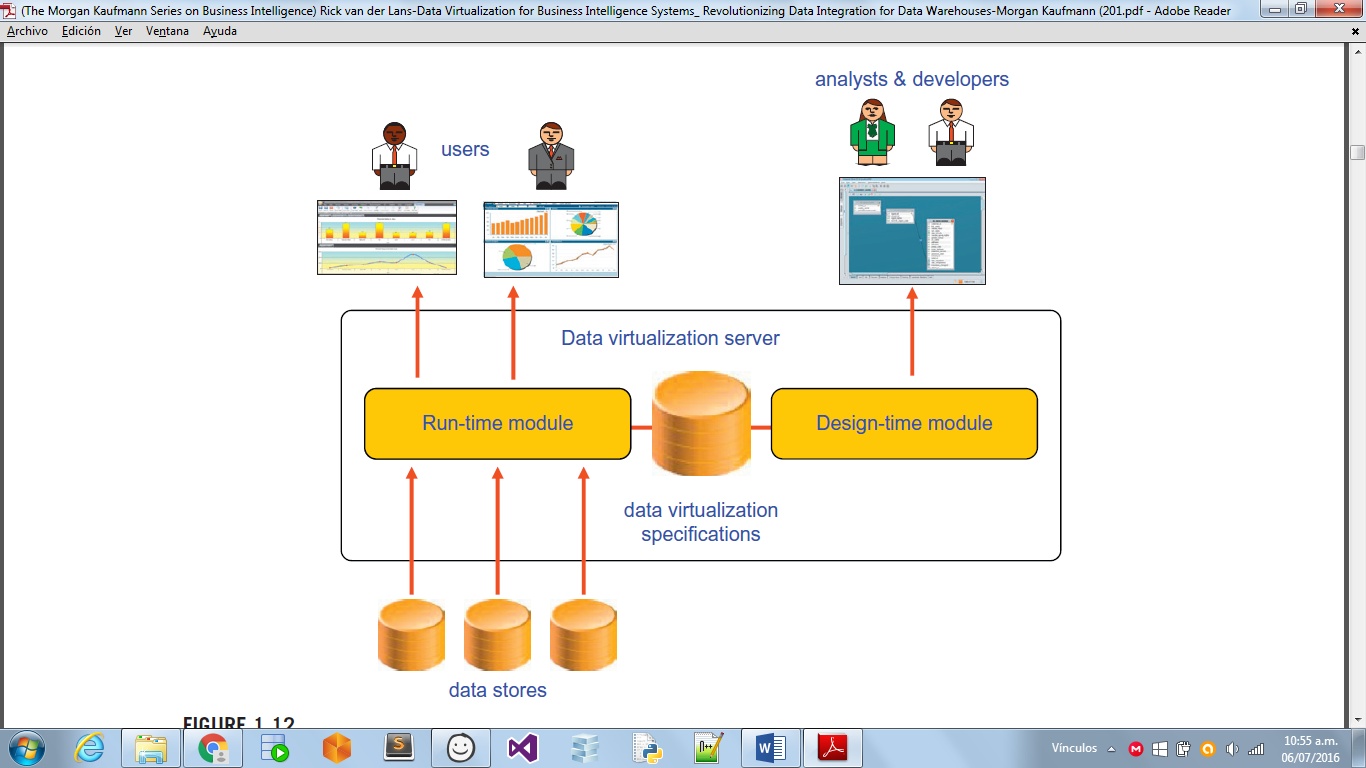


Figura 6 - Módulos de una capa de virtualización. Fuente [Rick+12]

**Análisis:**

La arquitectura de virtualización de datos consta de 3 capas, la capa de virtualización es la encargada de las operaciones de replicación, extracción, transformación y de carga. La capa de virtualización se compone de 2 módulos principales y un diccionario para almacenar las especificaciones de los metadatos. La arquitectura es bastante simple lo que le confiere un carácter de fácil mantenimiento, no presenta cadenas de base de datos en donde se dupliquen datos, lo que evita que use demasiados recursos y demás con el módulo de tiempo de ejecución le otorga una cualidad de actualización de datos con un grado bajo de latencia lo que mantiene los datos lo más actualizados posible para el análisis.

* 1. **Análisis y Resultados de las arquitecturas**

A continuación, se realizará el análisis comparativo entre las diferentes arquitecturas mostradas que nos permitirán de dar una solución de inteligencia de negocio en tiempo real y que demás estén alineados con los problemas de la presente investigación.

* + 1. **Definición de la necesidad del uso de una arquitectura de inteligencia de negocio**

La arquitectura que se definirá para la solución de la presente investigación deben estar alineado al procesamiento de información, de fuentes diversas, para la toma de decisiones en tiempo real.

* + 1. **Definición de los requerimientos**

Las arquitecturas analizadas deben soportar las actividades de análisis y monitoreo del tráfico postal además de estar alineados a los problemas de la presente investigación.

* + 1. **Alternativas a evaluar**

Se han considerado las siguientes arquitecturas a evaluar, a fin de tomarlos como referencia para la solución de inteligencia de negocios planteada en esta investigación.

* Arquitectura tradicional de inteligencia de negocios.
* Arquitectura de inteligencia de negocios en tiempo real.
* Arquitectura de inteligencia usada por SERPOST.
* Arquitectura de inteligencia usando virtualización de datos.
  + 1. **Definición de los factores de evaluación**

**Los factores a evaluar son los siguientes:**

* **Popularidad:** es un factor muy importante para que las organizaciones deseen optar por la arquitectura para su solución de inteligencia de negocios.
* **Orientación:** se refiere a los usuarios a los que va dirigido en los niveles de jerarquía de una organización.
* **Escalabilidad:** se refiere al grado de facilidad en que se pueden añadir más sistemas de que ayuden al rendimiento de las consultas y análisis de datos alterando mínimamente la arquitectura.
* **Flexibilidad al cambio:** Se refiere al grado de flexibilidad que tiene la arquitectura de poder adaptarse cuando se quiera un cambio en la presentación y análisis de información.
* **Adaptabilidad de implementación:** indica el grado de flexibilidad para personalizare a una implementación de solución específica.
* **Independencia de la tecnología BI:** se refiere al grado de no dependencia de funcionamiento de una solución o instancia de la arquitectura con respecto a las tecnologías de inteligencia de negocios usadas.
* **Independencia en los sistemas fuente:** se refiere al grado de no dependencia de funcionamiento de una solución o instancia de la arquitectura con respecto a las tecnologías de sistemas de información transaccionales.
* **Instalación:** se refiere al tiempo que demora instalar una solución de inteligencia de negocios usado una de las arquitecturas.
  + 1. **Definición del ponderado para los factores**

El ponderado para los factores a analizar será el siguiente:

* Pésimo:1
* Bajo:2
* Medio:3
* Bueno:4
* Excelente:5
  + 1. **Matriz de evaluación**

La valorización para cada factor se mostrará en la Tabla 3:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Factores | Arquitectura tradicional | Arquitectura en tiempo real | Arquitectura usada por SERPOST | Arquitectura usando virtualización de datos |
| Popularidad | **5** | **2** | **3** | **3** |
| Orientación | **3** | **4** | **2** | **5** |
| Escalabilidad | **3** | **3** | **2** | **4** |
| Flexibilidad al cambio | **3** | **3** | **1** | **5** |
| Adaptabilidad de implementación | **4** | **3** | **3** | **4** |
| Independencia de la tecnología BI | **3** | **3** | **1** | **5** |
| Independencia de los sistemas fuentes | **4** | **4** | **1** | **5** |
| Instalación | **3** | **2** | **5** | **4** |
| Total | **28** | **24** | **18** | **35** |

Tabla 3 - Comparación entre los factores de las arquitecturas. Fuente propia

* + 1. **Conclusión de la evaluación**

Como se muestra, la arquitectura que se usará en la solución en ña presente investigación será la de virtualización de datos. La selección se sustenta a continuación:

* El factor popularidad de las soluciones usando la arquitectura tradicional se le puso 5 debido a que es el más conocido, la mayoría de soluciones de inteligencia de negocio siguen esta arquitectura para su implementación, mientras que para la arquitectura en tiempo real es mucho menos conocido y menos implementado debido a sus mayores requerimientos tanto en las herramientas especiales de ETL para transformar los datos en un tiempo menor, como el la cantidad de componentes extra que tiene esta arquitectura, esto hace que sea muy poco atractiva para las empresas. La arquitectura usada para en SERPOST, es una que se suele utilizar cuando no hay un área de inteligencia de negocios en la empresa, suele usarse para esos casos, pero generalmente se prefiere por la tradicional y en cuando a la arquitectura de virtualización se encuentra en un auge de popularidad, aunque todavía no es muy difundida.
* El factor de orientación de las soluciones usando la arquitectura tradicional tiene un 3 debido a que está dirigido solo a las áreas de gerencia y a los niveles organizacionales más altos, por otra para la arquitectura usada por SERPOST obtuvo un 2 debido a que solo está orientada a las áreas de TI debido a que son conectada directamente con los sistemas de información empresariales y requieren un conocimiento técnico mayor para su uso. Las arquitecturas en tiempo real y virtualización de datos son las que está más orientadas a más personas debido que los datos se suelen actualizar con mucha frecuencia y están disponibles para muchas áreas; sin embargo, la arquitectura de virtualización de datos tiene una gama más amplia de orientación de debido a su cualidad de esconder los detalles técnicos para los usuarios.
* El factor de escalabilidad de las soluciones usando la arquitectura tradicional y a tiempo real obtuvieron un 3 debido a que cada vez que se intenta insertar componentes para la mejora de rendimiento, como servidores caches, se tiene que reestructura cierta parte de la arquitectura, mientras que en la de virtualización de datos cualquier inserción de nuevos componentes solo tiene como efecto una modificación en capa de virtualización. En cuanto a la arquitectura usada por SERPOST, si se desea ingresar un nuevo componente suele dar como efecto la actualización de la conexión para cada nuevo componente lo que lo hace poco escalable, razón por la cual obtuvo un 2.
* El factor de flexibilidad las soluciones usando la arquitectura tradicional y la de tiempo real obtuvieron un 3 debido que, si quiere cambiar algún aspecto del análisis o de presentación de información esto supone una restructuración de la arquitectura, como por ejemplo para las estructuras OLAP se tendría a que rediseñar su esquema y volver a replicar la data. Para la arquitectura de virtualización de datos los cambios tienen un efecto solo en la capa de virtualización; por ejemplo, si se desea cambiar la forma del análisis de datos en las estructuras OLAP solo se tiene que añadir una nueva dimensión o quitar algún sin necesidad de replicar ningún dato.
* El factor de adaptabilidad de implementación de las soluciones usando la arquitectura tradicional y la de virtualización de datos obtuvieron un 4 debido que son muy flexibles para la hora de implementarse para cualquier proyecto de cualquier tamaño, mientras que para la de tiempo real dependen más de las herramientas especiales que garanticen su tiempo de procesamiento con mínima latencia y la para la solución usando la arquitectura usada por SERPOST también depende de las tecnologías usada y de su compatibilidad con las herramientas de BI empleadas.
* El factor independencia de tecnologías BI de las soluciones de usando la arquitectura tradicional y en tiempo real obtuvieron 3 debido a que tiene la naturaleza de ser fieles a la tecnología de BI usada para su implementación, si se desea cambar se tendrá que reestructurar varias partes de la arquitectura. Pata la solución usando la arquitectura usada por SERPOST tiene un carácter de dependencia muy alto debido que se conectan directamente con los sistemas transaccionales y en cuando a la solución usando la virtualización de datos tiene solo es necesario instalar una nueva API para que se pueda conectar otra herramienta de BI a la solución.
* El factor de independencia de sistemas fuentes para las soluciones usando las arquitecturas tradicionales y de tiempo real suelen ser bastante independientes debido a que esto se encarga el proceso ETL de uniformizar la data. En cuando a la arquitectura usada por SERPOST tiene un alto grado de dependencia debido a que si se cambia o se añade un sistema transaccional se tiene que modificar muchas conexiones. Para la solución usando virtualización de datos, todos los sistemas transaccionales de integran en la capa de virtualización y esto encapsula todas las especificaciones técnicas asiendo que sea independiente a la tecnología de los sistemas transaccionales.
* El factor instalación de la arquitectura usada por SERPOST obtuvo un 5 debido que su instalación solo demora unas cuantas horas, mientras que las solucione con la arquitectura tradicional y con la de tiempo real suelen demorarse entre 3 a 8 meses. Las soluciones con una arquitectura de virtualización de datos suelen instalarse en un tiempo menor a 3 meses.

En conclusión, la arquitectura es la más apropiada para una solución de inteligencia de negocio en tiempo real es usando una arquitectura de virtualización de datos, teniendo una característica orientada a todo tipo de usuarios, es una solución mucho más escalable que las otras, que puede aplicarse y adaptarse a cualquier proyecto de cualquier tamaño, que posee mucha flexibilidad para cambiar y adaptarse a los requerimientos de los usuarios, con un grado de independencia alto con respecto a tecnologías inteligencia de negocios y transaccionales, y con un tiempo de instalación como solución muy corto con respecto a las de las alternativas. Es de aclarar que no es ideal para tareas de extracción de conocimiento a través de datos históricos como usando minería de datos; sin embargo, no se tomó en cuenta porque el problema a resolver de la presente investigación no está centrado en el análisis de información en tiempo real.

# Capitulo IV Aporte Teórico

La presente investigación se trata de una implementación de una solución de inteligencia de negocios en tiempo real para la empresa Servicios Postales del Perú.

En base al análisis realizado en la sección anterior se propone realizar una solución de inteligencia de negocios en tiempo real utilizando la arquitectura de virtualización de datos para poder resolver los problemas presentes en esta investigación.

La arquitectura de virtualización de datos se adaptará a los sistemas existentes en la empresa y a sus necesidades. Por otro lado, para el desarrollo de la solución se usará la metodología de gestión de proyectos SCRUM.

* 1. **Arquitectura propuesta para la solución**

La arquitectura de virtualización de datos que se va a mostrar a continuación es una adaptación de la arquitectura mostrada en el capítulo III, esta se encuentra alineada con los sistemas existente para la empresa la cual se destina la solución.

En la Figura 7 se muestra la arquitectura de virtualización de datos que se usara en la solución:

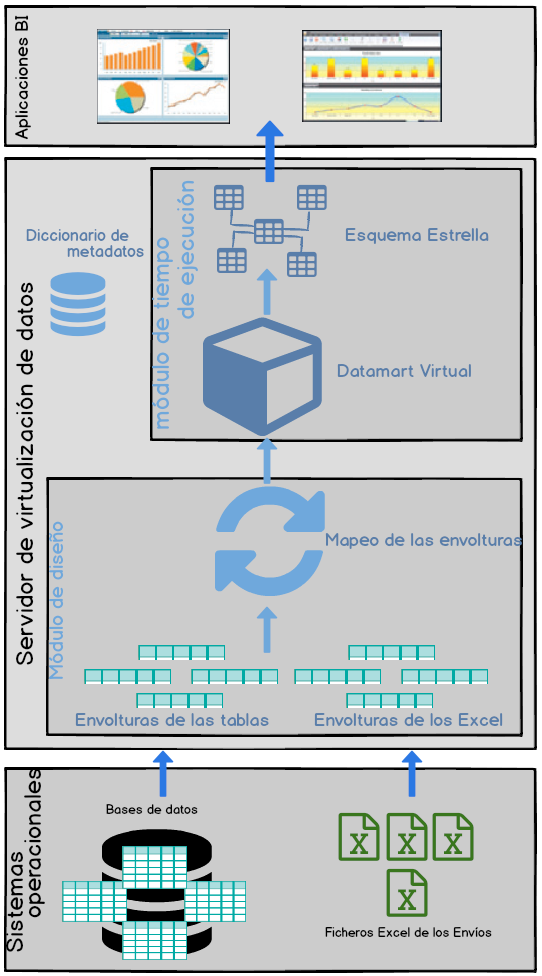


Figura 7 - Arquitectura de virtualización de datos propuesta. Fuente[Propia]

Para desarrollar la solución tenemos que identificar los sistemas fuentes a los cuales vamos a extraer la información, estos son la base de datos de la empresa y los archivos Excel que contiene la información de los envíos de las diferentes oficinas postales de SERPOST, estos se muestran en la Figura 8.

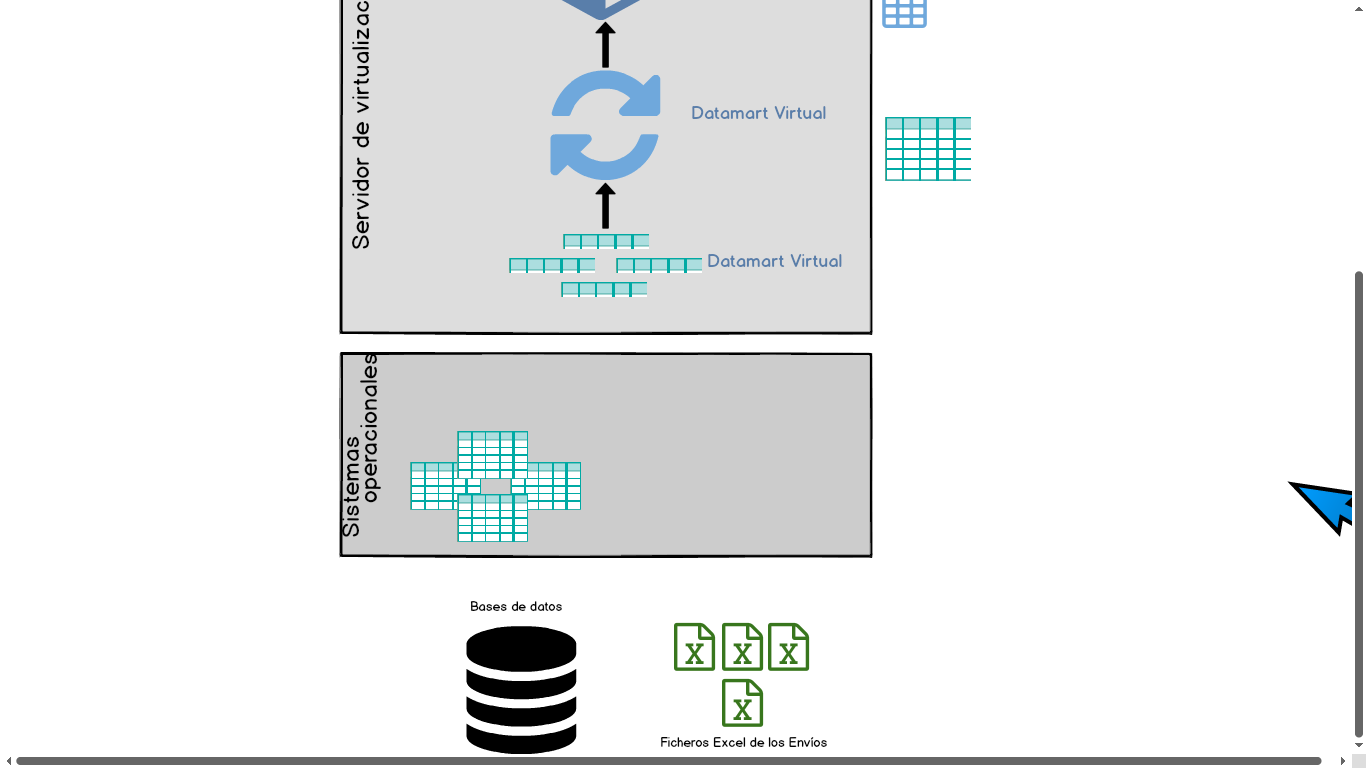


Figura 8 - Base de datos de la empresa y archivos Excel con información de los envíos

Lo siguiente es la configuración del servidor de virtualización de datos. Los primeros es especificar qué tipo de base de datos y que versión es, luego tenemos que importar las tablas de la base de datos que tenga la información de los diferentes tipos de envíos. Para los archivos Excel, tenemos que definir el formato en que se estructurara los datos del archivo de modo que se extraiga la información de los archivos Excel y se organice en el servidor de virtualización. Los formatos dependerán del tipo de análisis que necesite la empresa los cuales estarán especificados por los requerimientos de los usuarios. Estas especificaciones harán referencia a los datos contenidos en las tablas o ficheros importados. Este tipo de especificaciones se llama envoltura, y son en esencia los metadatos de la fuente donde se extrae, en la Figura 9 se muestran las envolturas.

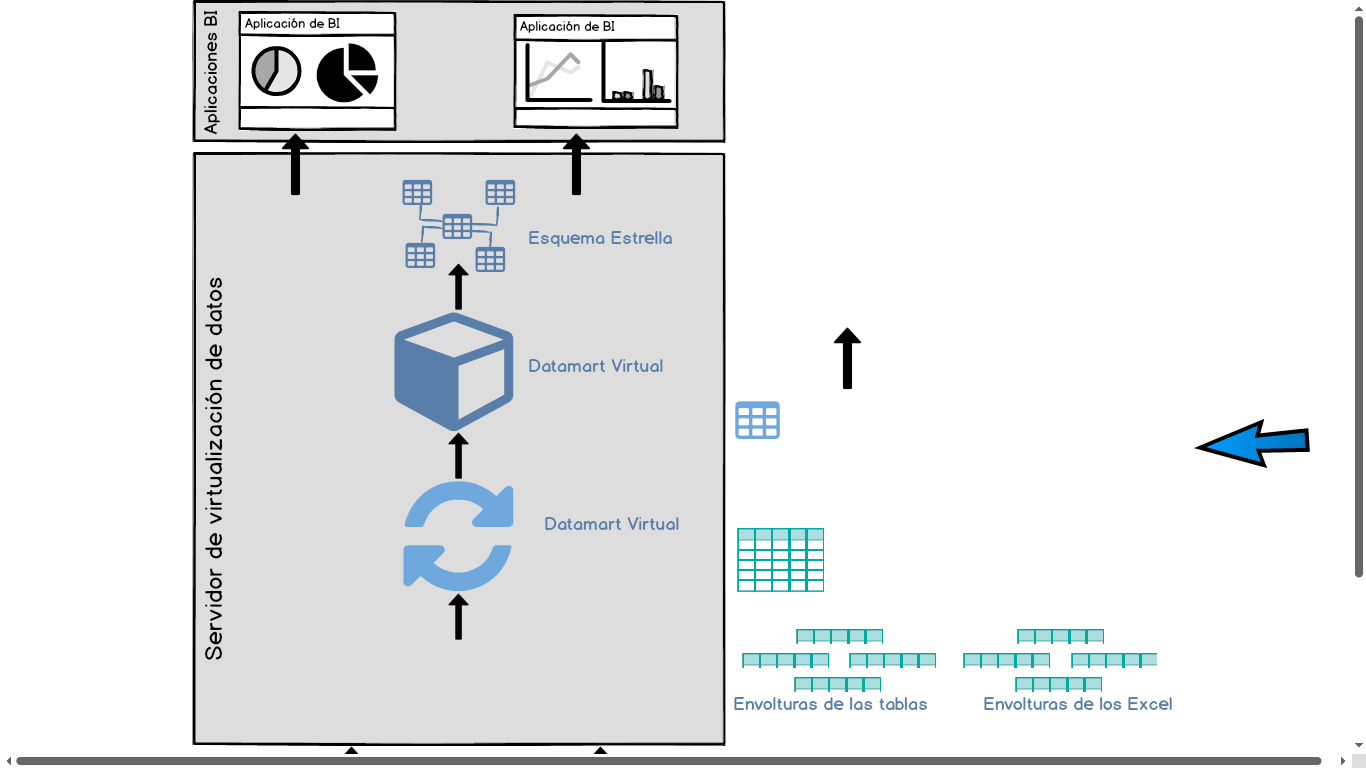


Figura 9 - Envolturas de las tablas seleccionada y de los archivos Excel, que contiene los metadatos y especificaciones de los datos. Fuente[Propia]

Lo siguiente es el mapeo de los atributos de las envolturas que hemos generado, es definir de lo que hemos importado tenemos que especificar qué datos nos servirán, que columnas serán usadas, y cuáles no, especificar un solo tipo de datos para realizar la integración y transformación de los datos de las diferentes fuentes. Después que se realice el mapeo de las envolturas se podrá generar las tablas virtuales, que será las tablas creadas de solo atributos necesarios e importantes para el análisis que nosotros queramos hacer. Estas tablas virtuales serán las tablas que compongan el Datamart Virtual que nos servirá para hacer los análisis de los datos. En la Figura 10 se muestra la imagen de del proceso de mapeo de las envolturas y la tabla virtual generada.

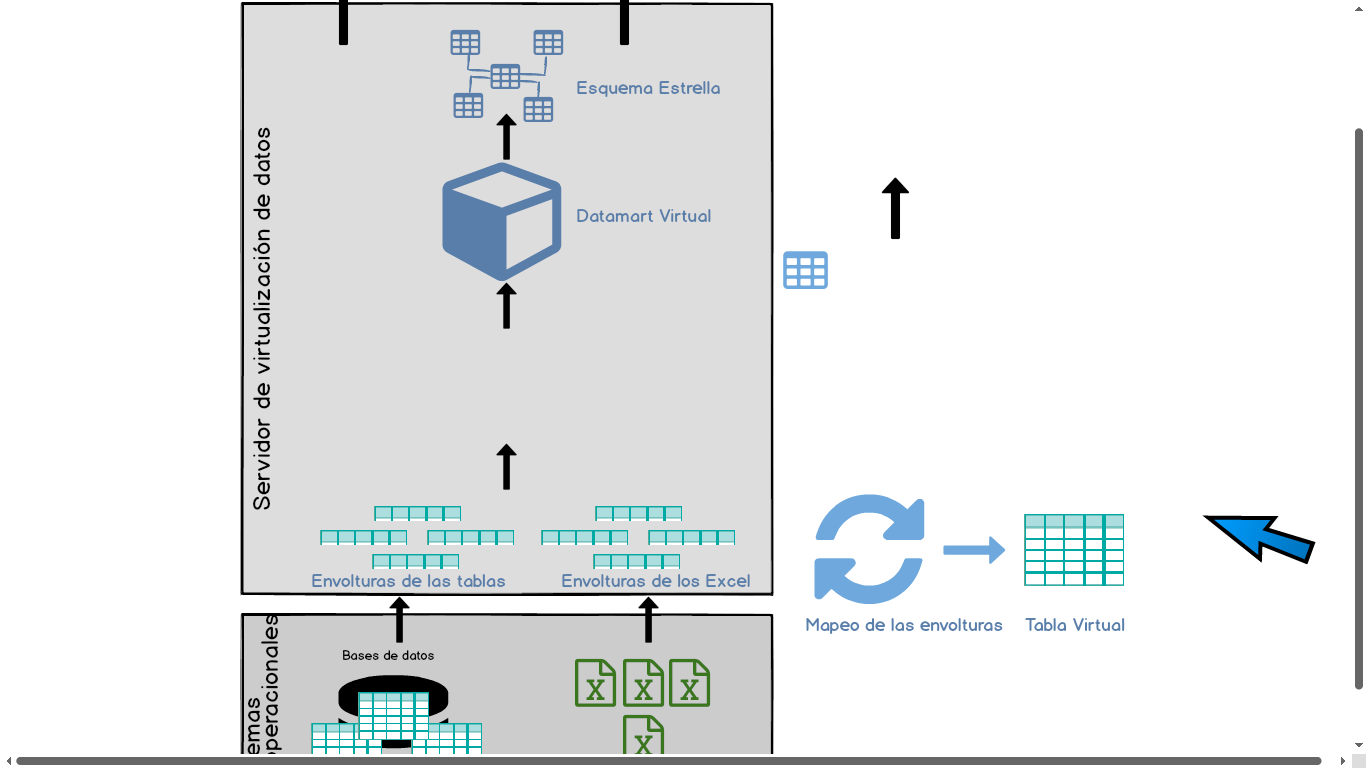


Figura 10 - Mapeo de las envolturas y generación de la tabla virtual. Fuente [Propia]

Luego de que generemos las tablas virtuales acorde con las especificaciones, que este caso serán las tablas que conformen el datamart, se procederá a crear el Datamart vinculando las tablas virtuales generadas para crear el esquema de diseño del Datamart, en la presente investigación se eligió el esquema estrella. En la Figura 11 se muestra el Datamart virtual con su respectivo esquema de diseño.

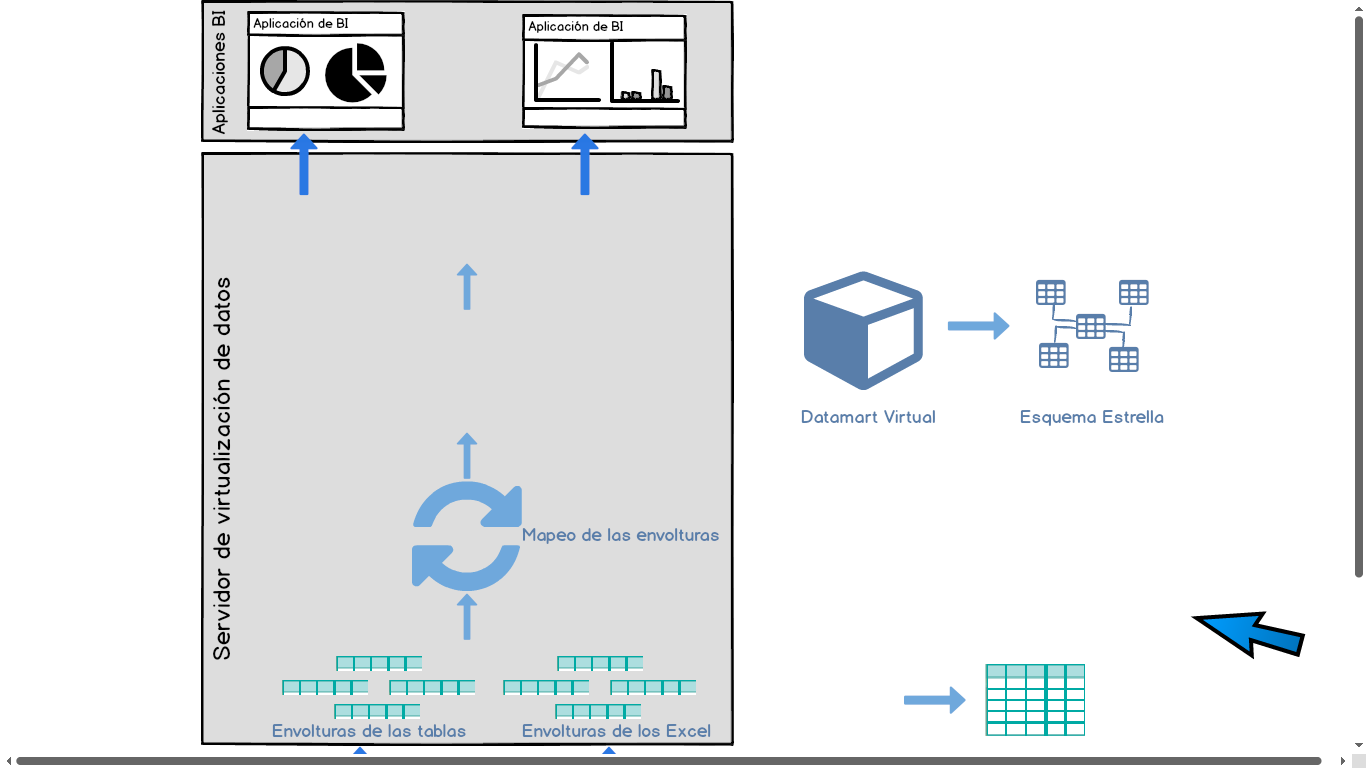


Figura 11 - Datamart Virtual con su respectivo esquema de diseño. Fuente [Propia]

Luego que este diseñado el Datamart se podrá realizar la conexión de con las aplicaciones de usuarios que se desee, también se puede especificar que herramientas de análisis se desea usar, las herramientas de virtualización de datos tienen una amplia gama de compatibilidad con diferentes herramientas de inteligencia de negocios, si se desea ingresar una se tiene que especificar qué tipo de tecnología se usa en el módulo de diseño. En la Figura 12 se muestra el proceso de conexión entre el Datamart y las aplicaciones de usuario.

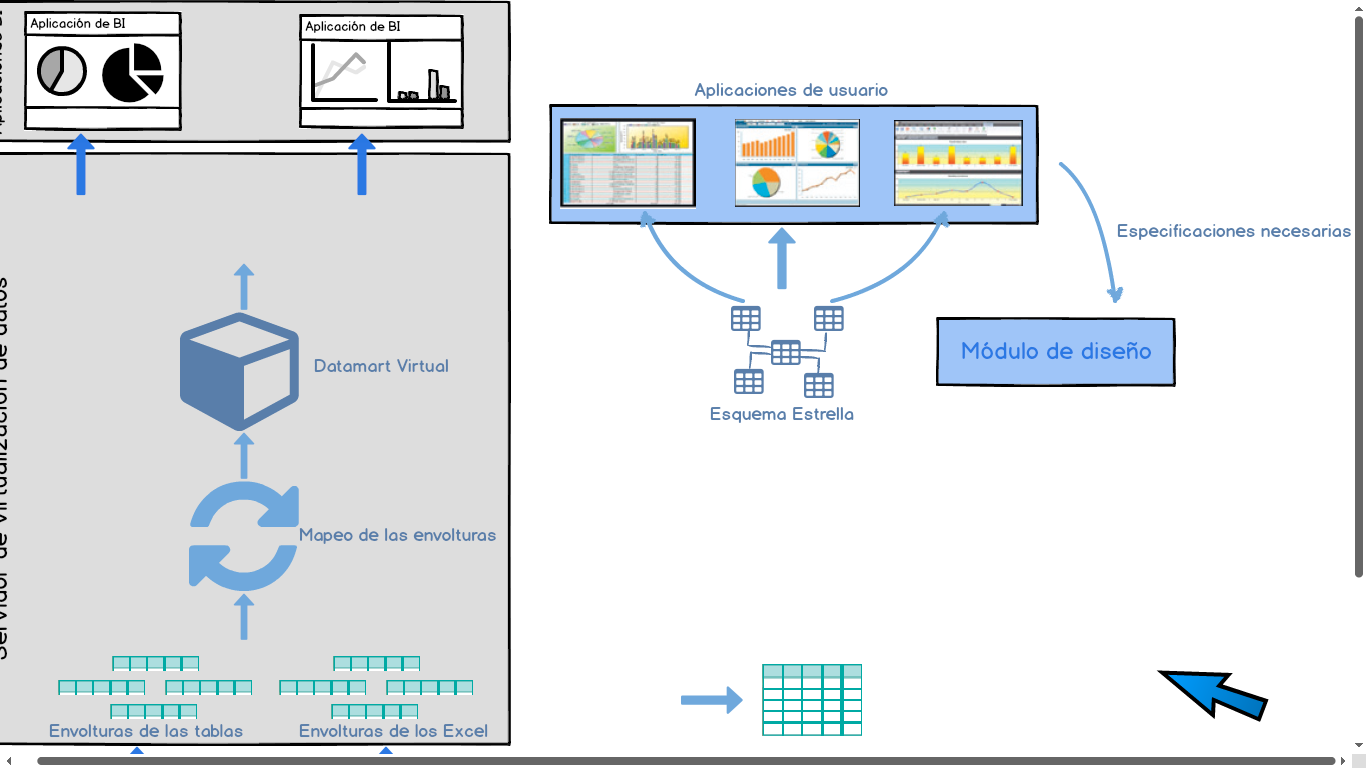


Figura 12 - Conexión entre las aplicaciones de usuarios y el esquema del Datamart. Fuente [Propia]

* 1. **Metodología usada para la gestión del proyecto de la presente investigación**

Para la gestión del proyecto de la presente investigación se optará por la metodología ágil SCRUM. En esta sección se definirá que es SCRUM, la justificación de porque se ha elegido esta metodología y se mostrara las especificaciones principales del proyecto de la investigación.

El objetivo en esta sección no es explicar a detalle cómo funciona o las características de SCRUM, sino solo mostrar los aspectos principales para definir como estará organizado el desarrollo de la solución.

* + 1. **Metodología a usar**

Como ya se mencionó, la metodología que se usara para la gestión del proyecto será SCRUM, la cual se alineara a los requerimientos del presente proyecto.

**Definición de SCRUM:**

Es una metodología ágil para la gestión de proyectos que fue desarrollada por Mike Beedle, Jeff Sutherland y Ken Schwaber que posee como principales características el adaptarse a cualquier cambio, ya sea de negocio o técnicos, para asegurar que la calidad del producto que se desarrolle sea alta. SCRUM consta de una serie de iteraciones o “sprints” en donde se van cubriendo los requerimientos especificados por los clientes mientras se realizan pruebas para comprobar y verificar que los requerimientos se están desarrollando según lo especificado, al final de cada iteración se realizan dos reuniones, una de ella es entre el equipo de desarrollo y los clientes llamado “review” en donde el equipo muestra lo avanzado al cliente y la retrospectiva (retrospective) en donde solo se reúne el equipo y se analiza lo que se avanzó, lo que no se avanzó y los obstáculos o inconvenientes que sufrió el equipo, además que se realizan reuniones diarias para que el equipo se apoye mutuamente.

* + 1. **Justificación**

Se ha escogido SCRUM para la gestión del presente proyecto porque es una metodología de gestión ágil y eficiente en la administración proyectos de cualquier tamaño o cualquier complejidad, que facilita el flujo de información y comunicación con el equipo del proyecto y el cliente que solicita dicho proyecto además de manejar de manera eficiente cualquier cambio y poder adaptarse a los nuevos requerimientos que es cliente desee solicitar,

En el caso del presente proyecto, de una solución de inteligencia de negocio, el uso de SCRUM es una gran ventaja para manejar los requerimientos de las áreas involucradas en el trafico postal, pudiendo adaptarse a cualquier nuevo requerimiento o cambio de este, como en la forma de presentación de la información de los reportes o en la forma de análisis de los datos para la toma de decisiones.

Bajo una arquitectura liviana como es la virtualización de datos, la metodología SCRUM permitirá desarrollar el proyecto con mayor eficiencia para generar entregable o vistas de los informes a los usuarios de las áreas interesadas en periodos cortos de tiempo, brindándoles a los clientes una vista del avance del proyecto generando mejor comunicación y encaminando la solución de la manera correcta.

* + 1. **Especificaciones del proyecto**

En esta sección se mostrarán todas las especificaciones del proyecto bajo la metodología SCRUM, demás los términos mostrados fueron adaptados para el presente proyecto. Se da por sentado que ya se conocen los términos asociados a SCRUM.

* + - 1. **Roles del proyecto**
         1. **Product Owner/Dueño del producto**

Ing. José Prieto Yalán – Jefe del Departamento de Sistemas

* + - * 1. **Desarrollador**

Jorge Enrique Quiroz Ñato – Autor de la investigación e implementación de la solución

* + - 1. **Eventos del proyecto**

Solo se definirán 2 eventos en esta sección, los demás se definirán en reunión entre los usuarios de las áreas interesadas.

* + - * 1. **Planeamiento del Sprint:** seta el evento que genere el plan para el proyecto presente, aquí se coordinara con las áreas de sub-gerencia relacionas al trafico postal y se definirá el cronograma, las historias de usuario y el tipo de pruebas y revisiones que sean necesarios.
        2. **Revisión del Sprint:** este evento será el que se realice al término de cada sprint, en donde se revisará los avances del desarrollo de los requerimientos planteados por los usuarios de las áreas interesadas en el trafico postal.
      1. **Herramientas SCRUM para el proyecto**

Solo se definirán 2 herramientas debido a que estas son esenciales para el proyecto, las demás serán definidas acorde con los requerimientos de los usuarios.

* + - * 1. **Backlog de Producto:** serán la lista de historias de usuario que soliciten los usuarios de las áreas interesadas en la información del tráfico postal.
        2. **Historias de Usuario:** Serán los tipos de reportes, tipos de análisis y funcionalidad de configuración que soliciten las áreas relacionadas al tráfico postal.
      1. **Iteraciones/sprints del Proyecto**

Los sprint del presente proyecto se definirán acorde con la reunión entre los usuarios interesados de las áreas relacionadas al trafico postal.

* 1. **Descripción del proyecto en la empresa**

El presente proyecto está enfocado en la empresa Servicios Postales del Perú y tiene como objetivo desarrollar una solución de inteligencia a de negocios en tiempo real usando una arquitectura de virtualización de datos para integrar las diferentes fuentes de datos de la empresa para posteriormente analizar mediante los datos para generar conocimiento útil para la toma de decisiones para las sub gerencias de **Administraciones Postales**, **Operaciones Postales** y la Sub-gerencia de **Paquetería y Operaciones Aduaneras**.

Esta solución proveerá de sistema de generación de reportes de análisis que simplificará la tarea de las diferentes áreas y departamentos involucrados en tráficos postal, servirá para realizar la toma de decisiones de las áreas de gerentica con la información correcta y necesaria, servirá de columna para el posterior desarrollo de un área de inteligencia de negocios y proveerá de una arquitectura de fácil mantenimiento.

**Referencias:**

[Rick+12] Data Virtualization for Business Intelligence Systems Revolutionizing Data Integration for Data Warehouses- Rick F. van der Lans (2012)

[Inmon+05] Building the Data Warehouse, Fourth Edition - Fourth Edition - W. H. Inmon (2005)

[Kimball+02] The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling, 2nd Edition - Ralph Kimball (2002)

[Kimball+13] The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling - Third Edition - Ralph Kimball (2013)

[Sahay+08] Real time business intelligence in supply chain analytics - B.S. Sahay and Jayanthi Ranjan (2008)

[Manh+05] Sense & Response Service Architecture (SARESA): An Approach towards a Real-time Business Intelligence Solution and its use for a Fraud Detection Application - Tho Manh Nguyen, Josef Schiefer and A Min Tjoa (2005)

[Hais+02] HAISTEN , M, Real-time Data Warehousing Defined, Library article from BetterManagement.com, 2002

[Terr+04] TERR, S, Real-time Data Warehousing 101, Article published at DataWarehouse.com, March 29,2004

[Tho+03] THO, N.; TJOA A., Zero-Latency Data Warehousing: Continuous Data Integration and Assembling Active Rules, 5th Intl. Conf. on Information Integration and Web-based Applications and Services(2003)

[Ricardo+04] Universidade Federal de Santa Catarina Programa de Pós-Graduação em Engenharia da Produção - Ricardo Guilherme Radünz Filho (2004)

[Jose+13] Análisis, diseño e implementación de una solución de inteligencia de negocios para el área de importaciones en una empresa comercializadora/importadora - José Eduardo Córdova Yupanqui (2013)

[Lisbeth+10] Propuesta de un Servicio de Mensajería Externa para la Empresa CP - Lisbeth García Morales, Luis Miguel Guerrero Durán Miriam Molina Velasco, Juan Gutiérrez Gutiérrez y Ricardo Daniel Pérez Navarro (2010)

[Azvine+06] Real Time Business Intelligence for the Adaptive Enterprise - B Azvine, Z Cui, D D Nauck and B Majeed (2006)

[Jin+08] A Business Activity Monitoring System Supporting Real-Time Business Performance Management - Jin Gu Kang and Kwan Hee Han (2008)

[Deanne+16] A review and future direction of agile, business intelligence, analyticsand data science - Deanne Larsona, Victor Chang (2016)

[Markus+10] Event-Driven Business Intelligence Architecture for Real-Time Process Execution in Supply Chains - Markus Linden, Sebastian Neuhaus, Daniel Kilimann, Tanja Bley, and Peter Chamoni (2010)

[Miguel+13] The Solid architecture for real-time management of big semantic data - Miguel A. Martínez-Prieto a, Carlos E. Cuesta b, Mario Arias c, Javier D. Fernández (2013)

[Shao+12] Research of Real-time Data Warehouse Storage Strategy Based on Multi-level Caches - Shao YiChuan, Xingjia Yao (2012)