

Solución ágil de inteligencia de negocios utilizando un gestor de base de datos columnar y software open source – Caso de estudio: Dirección del Registro Interconectado de Programas Sociales de la Secretaría Nacional de Planificación y Desarrollo

Villagómez Tusa, Ana Gabriela

Vicerrectorado De Investigación, Innovación y Transferencia De Tecnología

Centro De Posgrados

Maestría En Gestión De Sistemas de Información e Inteligencia de Negocios

Trabajo de titulación, previo a la obtención del título de Magíster en Gestión De Sistemas de Información e Inteligencia de Negocios

Ph.D. Yoo, Sang Guun

19 de junio de 2020



Document Information

Analyzed document	ProyectoTitulacionMaestria_GabrielaVillagomez2019(1).docx (D75319032)
Submitted	6/19/2020 8:15:00 PM
Submitted by	
Submitter email	mgutierrez@difusion.com.mx
Similarity	4%
Analysis address	mgutierrez1.GDC@analysis.arkund.com

Firma:

.....

Yoo, Sang Guun

DIRECTOR



**VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y
TRANSFERENCIA DE TECNOLOGÍA
CENTRO DE POSGRADOS**

CERTIFICACIÓN

Certifico que el trabajo de titulación, **“Solución ágil de inteligencia de negocios utilizando un gestor de base de datos columnar y software open source – Caso de Estudio: Dirección del Registro Interconectado de Programas Sociales de la Secretaría Nacional de Planificación y Desarrollo”** fue realizado por la señora **Villagómez Tusa, Ana Gabriela** el mismo que ha sido revisado y analizado en su totalidad, por la herramienta de verificación de similitud de contenido; por lo tanto cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, razón por la cual me permito acreditar y autorizar para que lo sustente públicamente.

Sangolquí, 19 de junio de 2020

Firma:

Yoo, Sang Guun

Director

C.C.: 1306853720



**VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y
TRANSFERENCIA DE TECNOLOGÍA**

CENTRO DE POSGRADOS

RESPONSABILIDAD DE AUTORÍA

Yo **Villagómez Tusa, Ana Gabriela**, con cédula de ciudadanía n° 1720073509, declaro que el contenido, ideas y criterios del trabajo de titulación: **Solución ágil de inteligencia de negocios utilizando un gestor de base de datos columnar y software open source – Caso de Estudio: Dirección del Registro Interconectado de Programas Sociales de la Secretaría Nacional de Planificación y Desarrollo** es de mi autoría y responsabilidad, cumpliendo con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Sangolquí, 19 de junio de 2020.

Firma

Villagómez Tusa, Ana Gabriela

C.C.: 1720073509



**VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y
TRANSFERENCIA DE TECNOLOGÍA**

CENTRO DE POSGRADOS

AUTORIZACIÓN DE PUBLICACIÓN

Yo **Villagómez Tusa, Ana Gabriela** autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación: **Solución ágil de inteligencia de negocios utilizando un gestor de base de datos columnar y software open source – Caso de Estudio: Dirección del Registro Interconectado de Programas Sociales de la Secretaría Nacional de Planificación y Desarrollo** en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.

Sangolquí, 19 de junio de 2020

Firma

Villagómez Tusa, Ana Gabriela

C.C.: 1720073509

Dedicatoria

A Dios, por ser el motor de mi vida, guiar cada uno de mis pasos y sobre todo por su inmenso amor, por regalarme salud y otro amanecer.

Salmo 16:2. Tú eres mi Señor; todo lo bueno que tengo, lo he recibido de ti. Sin ti no tengo nada.

A mi esposo, por su amor, apoyo y comprensión que siempre me demuestra para verme bien y feliz; por ser el compañero de mi vida que el Señor puso en mi camino, por todo lo que hemos compartido y por luchar juntos por nuestro hogar.

A mis padres, por estar siempre a mi lado, por su amor incondicional, guía y ejemplo que siempre me han entregado, para hacer de mí una persona de bien.

A mis hermanos, por su amor, apoyo y por desear de corazón siempre lo mejor para mí.

A mi sobrinita, por ser tan especial para mí, por todas las manifestaciones de cariño y amor.

Agradecimiento

Agradezco infinitamente a Dios, porque sé, qué ninguna hoja de un árbol cae sin su consentimiento, porque en su palabra he encontrado la motivación necesaria para no desmayar, y sobre todo por llenar mi corazón de paz y amor.

Un agradecimiento especial a mi esposo, por todo su amor, apoyo incondicional demostrado en el logro de esta meta, y por cada cosa pequeña o grande que hace en beneficio de nuestro bienestar y felicidad.

A las autoridades de la Unidad del Registro Social, por permitirme desarrollar este trabajo de investigación, y a cada uno de mis compañeros que conforman la parte técnica de la Dirección del Registro Interconectado de Programas Sociales, gracias por la predisposición y colaboración brindada.

A mi Director de tesis, Ing. Sang Guun Yoo Ph.D. quién me ha guiado con su conocimiento, para la consecución de este proyecto.

A la Coordinadora de esta Maestría Ing. Tatiana Gualotuña, Ph.D. por toda la ayuda proporcionada para poder culminar este trabajo.

Índice de Contenidos

Hoja de resultados de la herramienta Urkund.....	2
Certificado del Director	3
Responsabilidad de autoría	4
Autorización de publicación	5
Capítulo I.....	16
Introducción.....	16
Antecedentes	16
Justificación e Importancia	18
Planteamiento del Problema.....	18
Objetivos	19
<i>Objetivo general</i>	19
<i>Objetivos específicos</i>	19
Hipótesis	20
Alcance	20
Capítulo II.....	21
Marco Teórico	21
Inteligencia de Negocio	21
Inteligencia de Negocio Ágil	24
Data Warehouse (DW)	27
Gestor de Base de Datos Columnar	29
Metodologías para Soluciones de Inteligencia de Negocio	33

<i>Metodología Hefesto</i>	33
<i>Metodología de Kimball</i>	36
<i>Propuesta metodológica para la integración de datos</i>	38
Revisión de Literatura.....	44
Capítulo III.....	50
Problemática Actual y Propuesta de la Solución de Inteligencia de Negocio	50
Análisis de Problemas	50
Arquitectura General de la Solución	52
Análisis de la Propuesta de la Solución de Inteligencia de Negocio	54
Capítulo IV	56
Implementación	56
Fase de Análisis	56
<i>Requisitos de datos</i>	56
<i>Análisis de requerimientos</i>	57
<i>Selección de herramientas de tecnológicas</i>	58
Fase Diseño	68
<i>Análisis y determinación de esquemas</i>	68
<i>Análisis y determinación de componentes del diseño de datos dimensional</i>	68
<i>Diseño del modelo dimensional</i>	70
Fase Construcción y Procesamiento	71
<i>Construcción de procesos y pruebas ETL</i>	71

	10
<i>Calendarización de procesos</i>	80
Fase de Explotación y Producción.....	81
<i>Construcción de reportes de datos</i>	81
Fase de Mantenimiento y Soporte	82
<i>Mantenimiento de base de datos y herramientas de software</i>	82
<i>Generación de respaldos</i>	82
Capítulo V	83
Evaluación de Resultados	83
Análisis de Resultados	83
Capítulo VI	92
Conclusiones y Recomendaciones.....	92
Conclusiones.....	92
Recomendaciones.....	93
Bibliografía	94

Índice de Tablas

Tabla 1	Análisis FODA para soluciones de BI en memoria.....	26
Tabla 2	Participantes del Grupo Focal.....	59
Tabla 3	Características de herramientas de gestión de datos columnar	60
Tabla 4	Características de herramientas de integración de datos open source	65
Tabla 5	Prueba de ejecución proceso de validación	74
Tabla 6	Prueba de ejecución proceso de consolidación	75
Tabla 7	Prueba de ejecución proceso de carga de dimensiones	77
Tabla 8	Prueba proceso de carga DWH	79
Tabla 9	Calendarización procesos.....	80

Índice de Figuras

Figura 1 Arquitectura de inteligencia de negocios	23
Figura 2 Componentes de un sistema de negocios ágil.....	24
Figura 3 Modelo de base de datos columnar.....	30
Figura 4 Teorema Brewer.....	32
Figura 5 Proceso de inteligencia de negocios	33
Figura 6 Enfoque Kimball solución de inteligencia de negocios.....	37
Figura 7 Fases de la propuesta metodológica	38
Figura 8 Artículos académicos encontrados	46
Figura 9 Arquitectura actual de inteligencia de negocios	52
Figura 10 Propuesta de la arquitectura de inteligencia de negocios	53
Figura 11 Puntuación gestores de base de datos Cassandra, HBase, MonetDB.....	58
Figura 12 Resultado de la evaluación a gestores de base de datos columnare open source	62
Figura 13 Cuadrante de Gartner para herramientas de integración de datos	65
Figura 14 Resultados de la evaluación de las herramientas de integración de datos open source.....	67
Figura 15 Esquemas para la implementación.....	68
Figura 16 Modelo dimensional.....	71
Figura 17 Preparación del ambiente de trabajo	72
Figura 18 Ejecución proceso validación	73
Figura 19 Ejecución proceso consolidación.....	75
Figura 20 Proceso carga dimensiones	77
Figura 21 Proceso carga DWH.....	78
Figura 22 Implementación modelo dimensional.....	80

Figura 23 Reporte de datos	81
Figura 24 Descripción problema I	83
Figura 25 Descripción solución problema I	84
Figura 26 Tiempo de respuesta prueba I	84
Figura 27 Resultados gestores base de datos prueba I	85
Figura 28 Tiempo de respuesta prueba II	86
Figura 29 Resultados gestores base de datos prueba II	86
Figura 30 Tiempo de respuesta prueba III	87
Figura 31 Resultados gestores base de datos prueba III	88
Figura 32 Creación de extractos para reportes de información en Tableau Software ..	89
Figura 33 Explotación de información	90

Resumen

Las entidades públicas buscan nuevas alternativas tecnológicas que permitan agilizar sus procesos de gestión de datos, mediante la implementación de soluciones de bajo costo y eficientes. La Unidad del Registro Social específicamente la Dirección del Registro Interconectado de Programas Sociales, cuya función es *consolidar y mantener interconectadas las bases de datos de beneficiarios de programas sociales a nivel interinstitucional e institucional para direccionar los proyectos y programas sociales*, cuenta con un data warehouse que contiene información histórica de carácter social, sin embargo, cada vez es más difícil manejar el crecimiento de los datos de manera óptima. Actualmente, se ha detectado problemas de rendimiento en el procesamiento de integración de datos afectando sus tiempos de respuesta. El presente proyecto tiene como propósito mejorar la solución de inteligencia de negocio de la institución, mediante un modelo de datos multidimensional, utilizando herramientas de gestión de datos columnar e integración de datos open source, para lo cual, se realizó un estudio que permita conocer las herramientas más eficientes; el desarrollo del proyecto está basado en una metodología para inteligencia de negocio. Como resultado, para el caso de estudio se disminuyó el tiempo de respuesta de acceso a los datos, utilizando herramientas open source, lo cual permitirá acelerar los procesos de integración de datos en la institución.

PALABRAS CLAVE:

- **ENTIDAD PÚBLICA**
- **DATA WAREHOUSE ÁGIL**
- **BASE DE DATOS COLUMNAR**

Abstract

Public entities are looking for new technological alternatives that make it possible to streamline their data management processes, through the implementation of low-cost and efficient solutions. The Social Registry Unit specifically the Directorate of the Interconnected Registry of Social Programs, whose function is to consolidate and maintain interconnected databases of beneficiaries of social programs at the inter-institutional and institutional level to direct social projects and programs, has a data warehouse that It contains historical information of a social nature, however, it is increasingly difficult to manage data growth optimally. Currently, performance problems have been detected in data integration processing affecting their response times. The purpose of this project is to improve the institution's business intelligence solution, using a multidimensional data model, columnar data management tools and open source data integration, for which, a study was conducted that allows knowing the more efficient tools; project development is based on a methodology for business intelligence. As a result, in the case of the study, the response time of data access was reduced, using open source tools, which will accelerate the data integration processes in the institution.

KEYWORDS:

- **PUBLIC ENTITY**
- **DATA WAREHOUSE ÁGIL**
- **COLUMN DATABASE**

Capítulo I

Introducción

Antecedentes

Mediante Decreto Ejecutivo Nro.712, de 11 de abril de 2019, publicado en el Registro Oficial, establece en el artículo 5: *“Créase la Unidad del Registro Social, como un organismo de derecho público, adscrito a la entidad rectora de la administración pública, con personalidad jurídica propia, dotado de autonomía administrativa, operativa y financiera, con sede en la ciudad de Quito, con facultades de coordinación, gestión, seguimiento y evaluación. Será la entidad encargada de la administración y el mantenimiento de los sistemas que permiten la gestión del Registro Social, y de la administración, el mantenimiento, la actualización y difusión de la información de la base de datos del mismo”*.

En su Disposición General, Primera señala *“Todas las representaciones, delegaciones, responsabilidades, partidas presupuestarias, bienes muebles, activos y pasivos; así como, los derechos y obligaciones constantes en convenios, contratos u otros instrumentos jurídicos nacionales e internacionales, relacionados con la administración, mantenimiento y actualización del Registro Social, a cargo de la Secretaria Nacional de Planificación y Desarrollo, serán asumidos y pasarán a formar parte del patrimonio de la Unidad del Registro Social, de acuerdo con las disposiciones establecidas en el proceso de levantamiento de información.”*

Mediante Resolución Nro. 003-2019-URS, de 01 de julio de 2019, se expide el Estatuto Orgánico de Gestión Organizacional por Procesos de la Unidad del Registro Social, en el cual establece en el artículo 10, la misión de la Dirección del Registro Interconectado de Programas Sociales es: *“Administrar la información de la base de datos del Registro Social; y, consolidar y mantener interconectadas las bases de datos de beneficiarios de programas sociales a nivel interinstitucional e institucional para direccionar los proyectos y programas sociales, facilitando la identificación y registro de beneficiarios”*.

Un estudio de *The Data Warehouse Institute* (TDWI) (Russom, Data Warehouse Modernization, 2016), indicó en su encuesta que los principales elementos de la modernización de un data warehouse incluyen realinear con objetivos comerciales recientes, aumentar la escala para grandes volúmenes de datos; habilitar nuevas herramientas de análisis o tipos de datos y sus prácticas correspondientes.

Según, la TDWI la modernización del data warehouse asume muchas formas, desde actualizaciones de servidor, ajustes para modelos de datos, agregar una nueva estructura al entorno de almacenamiento de datos extendido, hasta reemplazar la plataforma primaria del data warehouse (Russom, Data Integration for Real-Time Data Warehousing and Data Virtualization, 2010).

Anteriormente, la Dirección del Registro Interconectado de Programas Sociales, cumplió con ciertas funciones específicas, por lo cual fue implementado un data warehouse con la utilización de herramientas propietarias adecuadas para ese tiempo, sin embargo, los requisitos han ido cambiando; por tal motivo se considera importante realinear el entorno del data warehouse con los nuevos necesidades institucionales y desafíos tecnológicos.

Justificación e Importancia

La base de datos del Registro Interconectado de Programas Sociales ha sido integrada durante algunos años, con la finalidad de almacenar información relevante de programas sociales, siendo útil para la toma de decisiones institucional y, en general, para difundir información a la ciudadanía.

Actualmente, se tiene almacenada una gran cantidad de registros, en varios modelos de datos, por lo cual, se genera un tiempo considerable cuando se realiza algún tipo de consulta de datos; este proceso es realizado mediante la utilización de software propietario. Por tal motivo, se necesita desarrollar una solución de inteligencia de negocios que permita optimizar los tiempos de consulta, mediante una adecuada gestión e integración de datos, utilizando herramientas open source.

Planteamiento del Problema

La Dirección del Registro Interconectado de Programas Sociales (DRIPS) integra de forma periódica diferentes fuentes de datos, la información es almacenada en varios modelos de datos dimensionales según cada programa social, el acceso a los datos para realizar algún tipo de análisis toma un tiempo considerable al realizar junturas de datos para las consultas. Además, la falta de soporte técnico de las herramientas que se utiliza para llevar a cabo el proceso de gestión e integración de datos no permite tener las nuevas mejoras y actualizaciones del software, limitando el uso de funcionalidades, y en ocasiones produciendo un deterioro en el rendimiento; y finalmente, la falta de una metodología establecida, dificulta alcanzar objetivos de forma eficiente.

Por consiguiente, la DRIPS busca nuevas herramientas tecnológicas que permitan mejorar el procesamiento de gestión e integración datos, bajo el contexto de software open source y desarrollados mediante una metodología de inteligencia de negocio.

Objetivos

Objetivo general

Implementar un modelo multidimensional para la Dirección del Registro Interconectado de Programas Sociales, mediante el uso de un gestor de base de datos columnar y la aplicación de una metodología adecuada en base a ciertas prácticas, y técnicas para proyectos de inteligencia de negocio, con la finalidad de actualizar la solución existente y mejorar el rendimiento de la gestión de datos.

Objetivos específicos

- Realizar una revisión de literatura para determinar las características de herramientas de gestión de datos columnar y prácticas más adecuadas, para la implementación de proyectos de inteligencia de negocios, que contribuirá en el planteamiento de la propuesta de mejora.
- Implementar una nueva solución de inteligencia de negocios enfocada a la información de carácter social en el sector gubernamental, con el fin de modernizar la gestión de datos existente en la DRIPS.
- Evaluar los resultados mediante el uso de herramientas de integración de datos y gestión de datos, con el propósito de comprobar el tiempo de respuesta de acceso a los datos aplicado al modelo multidimensional en un gestor de datos relacional y columnar.

Hipótesis

La implementación de un modelo dimensional en un gestor de base de datos columnar, mediante el seguimiento de una metodología de integración de datos para proyectos de inteligencia de negocios enfocada al caso de estudio, permitirá mejorar el tiempo de acceso a los datos y actualizar la solución de inteligencia de negocios existente en la DRIPS.

Para la demostración de la hipótesis planteada se considera la utilización de herramientas de integración y gestión de datos demostrando el tiempo de respuesta de acceso a los datos, mediante la implementación de un modelo dimensional en un gestor de base de datos relacional y en un gestor de base de datos columnar.

Alcance

El alcance de la presente investigación es realizar un estudio que permita generar una propuesta de solución de inteligencia de negocio, con el apoyo de plataformas open source, para lo cual, se analizará las características de herramientas informáticas para gestores de base de datos columnares e integración de datos distribuidos actualmente, a fin de escoger las más apropiadas.

Además, se pretende plantear una metodología para la integración de datos; por medio de un análisis, realizar el diseño del modelo multidimensional e implementarlo con la utilización del software seleccionado y finalmente presentar una evaluación de los resultados obtenidos.

Capítulo II

Marco Teórico

En este capítulo, se presenta algunas definiciones y conceptos necesarios para el desarrollo e interpretación del proyecto y una revisión sistemática de literatura para evaluar y sustentar la investigación, con la finalidad de generar una propuesta adecuada para el área de interés.

Inteligencia de Negocio

La Inteligencia de Negocio o Business Intelligence (BI) ha sido definida como “las herramientas, tecnologías y procesos, permiten convertir los datos en información y la información en conocimientos y planes, con el fin de optimizar las acciones empresariales”. (Muntean & T.Surcel, 2013).

La Inteligencia de Negocios es la integración por un lado del almacenamiento y por el otro el procesamiento de grandes cantidades de datos, con el principal objetivo de transformarlos en conocimiento y en decisiones en tiempo real, a través de un sencillo análisis y exploración. Donde, “*Datos + Análisis = Conocimiento*”. Este conocimiento debe ser oportuno, relevante, útil y debe estar adaptado al contexto de la organización.

Es decir, la frase “Inteligencia de Negocios es el proceso de convertir datos en conocimiento y el conocimiento en acción, para la toma de decisiones”. (Bernabeu, 2010)

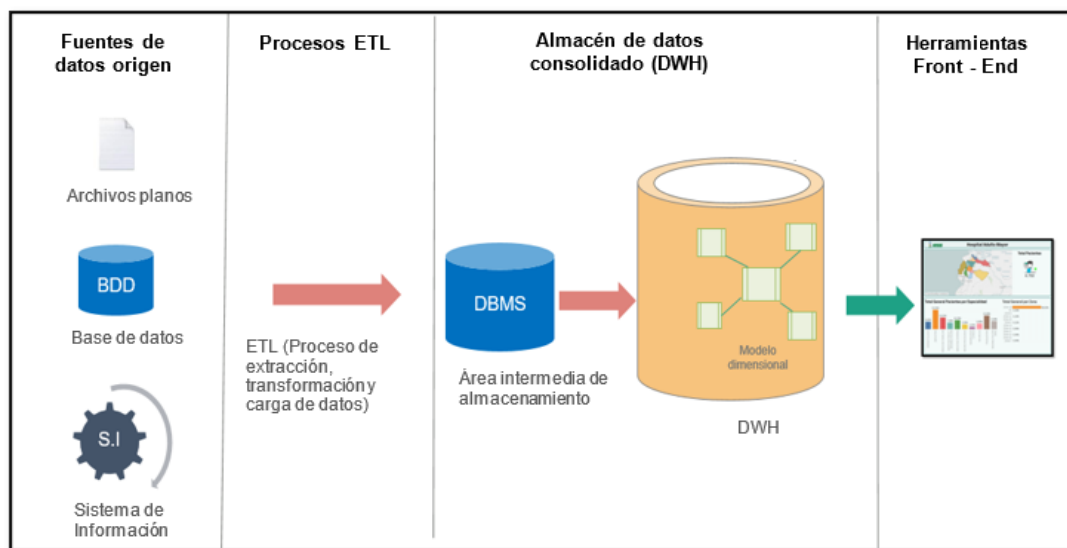
Según (Díaz, 2010), se entiende por Business Intelligence al conjunto de metodologías, aplicaciones, prácticas y la administración de la información que permite tomar mejores decisiones a los usuarios de una organización.

La inteligencia de negocios, hoy en día se basa en la entrega, análisis e integración de información de forma ágil. La entrega de información se enfoca en informes, consultas ad hoc, la integración de datos se basa en la consolidación de diferentes fuentes, el análisis se centra en capacidades como, el descubrimiento de datos basados en búsquedas, el análisis geoespacial y el procesamiento analítico en línea (OLAP).

Se apoya en la combinación de tecnologías que permitan acelerar los procesos de negocio, para tomar decisiones sensibles al tiempo en relación con las operaciones diarias que se llevan a cabo en un organismo. (Smith & Rege, 2017).

En el año 2017, Gartner, Inc. reconoce que el sector de la analítica de datos y BI estima un volumen de negocio mundial cercano a los 18.300 millones de dólares y un incremento del 25% que el negocio aumente en los próximos 3 años, debido a que estos sistemas darán accesibilidad, agilidad y conocimiento para dar continuidad a los negocios. Además, señala siete factores responsables de la evolución del mercado de BI y analítica: i) las herramientas modernas de BI y analítica; ii) la emergencia de nuevos desarrolladores de empresas establecidas; iii) la necesidad de dataset más complejos; iv) la habilidad para incrustar y extender contenidos; v) el soporte en tiempo real y los datos procedentes de streaming; vi) el crecimiento del despliegue de cloud de analítica y BI; vii) y los marketplaces. (Telefónica S.A., 2017).

La arquitectura de inteligencia de negocios nos permite visualizar el flujo que tendrán los datos hasta convertirse en información útil y generar conocimiento en una organización, en la Figura 1. se indica cada uno de los componentes necesarios.

Figura 1*Arquitectura de la Solución de Inteligencia de Negocio*

Nota: Recuperado de (Curto & Conesa, 2010)

Fuentes de datos de origen, insumo principal para el proceso de integración de datos, lo componen archivos planos que contienen información transaccional, base de datos estructurados, semiestructurados o no estructurados, sistemas de información operacionales, los cuales pueden ser internos o externos a la organización.

Proceso ETL, es el mecanismo mediante el cual varias fuentes de datos autónomas, distribuidas y heterogéneas, se consolidan y se asocian en un esquema global. Este proceso, facilita la gestión de los datos mediante una lógica y secuencia de pasos que permite la extracción, transformación y carga. (ETL). (Smith & Rege, 2017).

En el proceso los datos son extraídos, para ser transformados y pasar por varios procesos como son la discretización, limpieza, reemplazar valores faltantes, eliminación de ruido, entre otros, de modo de ser procesados de forma correcta y obtener datos confiables.

Almacén de datos consolidado, es el repositorio destino donde se almacenan todos los datos ya transformados. (Lara & Olivares, 2016).

Herramientas Front-End, está relacionado con el software que permite realizar reportes, cubos OLAP, las aplicaciones de análisis, la minería de datos y los KPIs. (Posada, Vega, Ampuero, & Eljaiek, 2015)

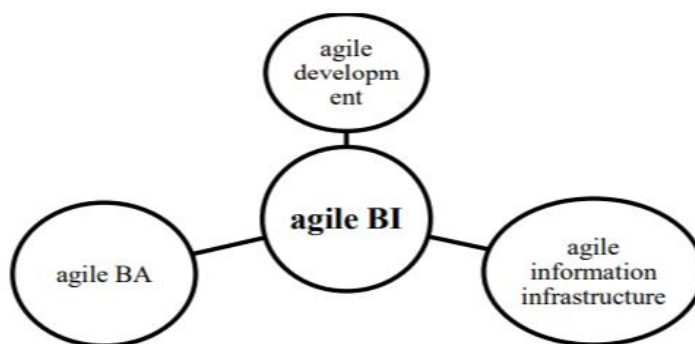
Inteligencia de Negocio Ágil

Según Forrester Research (Muntean & T.Surcel, 2013), define a los sistemas ágiles de BI como "un enfoque que combina procesos, metodologías, estructura organizativa, herramientas y tecnologías que posibiliten estrategias, decisión táctica y operativa; para ser más flexible y sensible al cambio rápido de los negocios y los requisitos reglamentarios".

A continuación, en la Figura 2. se presenta los componentes de un sistema de inteligencia ágil.

Figura 2

Componentes de un sistema de negocios ágil.



Nota: Recuperado de (Muntean & T.Surcel, 2013).

Ágil Desarrollo, una solución ágil de BI debe ser implementada de forma dinámica, el propósito es lograr reaccionar más rápidamente a los requisitos comerciales cambiantes. En la BI, existen dos enfoques distintos que son relevantes en el contexto del desarrollo de soluciones de BI: desarrollo en cascada y desarrollo ágil.

Los métodos ágiles ayudan a la organización o un equipo a adoptar el subconjunto de principios y prácticas basados en la cultura, los valores y el tipo de sistemas. Cada organización tiene sus propios métodos ágiles hechos a la medida de sus necesidades.

Las metodologías de desarrollo ágil de una solución de BI más populares son: Scrum, Scope Extremo y Almacenamiento ágil de datos. Estos proyectos, por lo general cuentan con diferentes características como: i) colaboración entre equipos funcionales, ii) desarrollos iterativos iii) tolerancia a los cambios. Las características pueden ser aplicadas a proyectos de BI.

Este enfoque aparece para solventar los principales problemas de proyectos tradicionales como son, los largos tiempos entre la solicitud del sistema y la entrega de la solución de BI, falta de flexibilidad en los cambios de los requisitos analíticos, pruebas al final de la vida del ciclo de desarrollo. (Devarapalli, 2013).

Ágil Analítica de Negocio (BA) para esta perspectiva, es necesario contar tanto con la tecnología apropiada y una metodología de diseño ágil. Hoy en día existen herramientas de BI con tecnología en memoria, cuyo principal objetivo es reemplazar las soluciones de BI basadas en almacenamiento en disco tradicionales cuyos modelos de datos son relacionales o basados en OLAP. La BI en memoria puede ahorrar tiempo de desarrollo significativo, eliminando la necesidad de almacenar datos precalculados o datos agregados relacionales. Las características más comunes son fáciles de usar, tienen un interfaz visual, velocidad de respuesta en el procesamiento y son rápidos para realizar un despliegue.

Según los análisis realizados la tecnología de BI en memoria tiene el potencial de ayudar a los sistemas a ser más ágiles. (Muntean & T.Surcel, 2013).

A continuación, en la Tabla 1. se muestra un análisis FODA para soluciones de BI en memoria.

Tabla 1

Análisis FODA para soluciones de BI en memoria

Fortalezas	Debilidades
<ul style="list-style-type: none"> • Mayor velocidad de respuesta, rápido acceso a informes, análisis y métricas de negocios. • Mejorar el autoservicio a través de la flexibilidad analítica. • Permite a las empresas integrar datos de sistemas transaccionales, fuentes de datos externas, hojas de cálculo o almacenes de datos. • Despliegue rápido. • Elimina la necesidad de almacenar datos precalculados en cubos OLAP o tablas relacionales agregadas. • Facilidad de uso para los usuarios finales. • Bajos Costos. 	<ul style="list-style-type: none"> • Gestión de metadatos limitado. • Limitado para procesos ETL. • Calidad de datos. • Limitado por la memoria física.
<ul style="list-style-type: none"> • Permite inteligencia de negocios en tiempo real sin un DW • Elimina la necesidad de un cubo OLAP precargado o un data mart • Computación en la nube 	<ul style="list-style-type: none"> • No es un análisis en tiempo real porque los datos se analizan en memoria, no en el almacén de datos.

Nota: Recuperado de (Muntean & T.Surcel, 2013)

Infraestructura de Información Ágil aborda la arquitectura e integración de datos, para asegurar agilidad y reaccionar a los requisitos de los negocios cambiantes. Debe ser capaz de extraer y combinar datos de cualquier fuente tanto internas y como externas, incluyendo XML (Lenguaje de Marcado Extensible), relacional, semiestructurado, multidimensional, entre otros.

Los sistemas de BI utilizan herramientas ETL para la extracción de datos de múltiples fuentes y almacenar temporalmente esos conjuntos de datos en un repositorio intermedio. A diferencia de las herramientas ETL, en algunas infraestructuras de información ágil se utiliza un servidor de virtualización, porque permite que las fuentes de datos permanezcan en sus sitios de origen, elimina el almacén de datos intermedio; admite tener datos resumidos en el inicio resolviendo la parte estructural y semántica. (Muntean & T.Surcel, 2013).

Data Warehouse (DW)

Definido por Bill Inmon (Inmon, 2002), como “Una colección de datos orientados a un tema específico, integrado, no volátil e histórico, organizados para dar soporte a la toma de decisiones de la gerencia”. El objetivo principal del almacenamiento de datos es reunir información de diferentes fuentes y colocar en un formato que sea propicio y sirva de insumo para el análisis de la información.

Según la TDWI (Russom, Data Integration for Real-Time Data Warehousing and Data Virtualization, 2010), habla de un almacén de datos en tiempo real (DW TR), como la composición de tecnologías que permiten la transferencia de datos en tiempo real o la obtención de datos desde aplicaciones operativas, para ser reportados en informes o paneles de información relevante, fundamental para lograr la toma de decisiones a tiempo. Un data warehouse está compuesto por una parte lógica denominado modelo multidimensional, que es un diseño conceptual que permite tener ideas más cercanas sobre las formas en que un usuario puede percibir un dominio de aplicación para su almacén de datos. Realmente, se considera un paso clave que garantiza el éxito de los proyectos de almacenamiento de datos, ya que define la expresividad.

El modelo multidimensional consiste en un hecho central que representa el tema de análisis rodeado de dimensiones para percibir la información desde diferentes perspectivas. (Thenmozhi & Vivekanandan, 2013).

Los elementos de diseño de un esquema multidimensional son:

- **Hechos:** son datos instantáneos en el tiempo, que son filtrados, agrupados y explorados a través de condiciones definidas en las tablas de dimensiones. Son almacenados en una estructura de datos denominado tabla de hechos, constituyen el volumen de la bodega, y pueden estar compuestos por millones de registros dependiendo de su granularidad y antigüedad de la organización. (Bernabeu, 2010).

La tabla de hechos está relacionada con un conjunto de dimensiones que representa las diferentes perspectivas mediante las cuales se analiza el hecho.

- **Dimensiones:** recogen los puntos de análisis de un hecho. Estos atributos se encuentran categorizados en una jerarquía o simplemente son descriptivos. (Curto & Conesa, 2010).
- **Medidas:** son las métricas de la tabla de hechos a las que se aplica una función de agregación como *count*, *distinct count*, *sum*, *max*, *avg*, etc, almacenados por su rápido acceso. (Curto & Conesa, 2010)

Los enfoques de diseño conceptual por lo general son basados según: i) Análisis exhaustivo de la fuente de datos, aquí, los requisitos se utilizan al final del diseño para filtrar los resultados. ii) Análisis del diseño en función de los requisitos, para generar el modelo multidimensional, aquí la fuente de datos se considera solo cuando se llena el almacén de datos. (Thenmozhi & Vivekanandan, 2013).

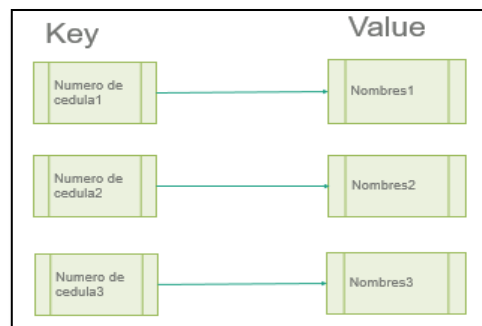
Gestor de Base de Datos Columnar

Las Bases de Datos Columnares se introdujeron por primera vez en 1970 en productos como Model 204 y ABABAS, este enfoque ha resurgido recientemente en consecuencia de las limitaciones que presentan los gestores de bases de datos relacionales y la necesidad de un sistema de almacenamiento diferente, por lo cual su utilización empieza a tomar fuerza en los últimos tiempos.

Como su nombre lo indica, las bases de datos están organizados en columnas en lugar de la fila, es decir se referencia como el valor en el par clave-valor, donde cada clave está asociada a uno o más atributos. Las claves son utilizadas como índices para realizar búsquedas y ser accedidos de manera rápida, los datos son almacenados contiguamente en un lugar separado en disco, por lo cual tienen un buen rendimiento para procesar grandes volúmenes de datos, simples y sin funcionalidades (Jara, 2014), como se visualiza en Figura 3. un ejemplo de acceso a los datos por medio de key-value.

Figura 3

Modelo de Base de Datos Columnar



Nota: Recuperado de (Jara, 2014)

Dentro de las principales características se destacan:

- **Compresión:** La compresión permite gastar menos tiempo en procesos de E/S mientras se leen los datos del disco, además reduce considerablemente el espacio ocupado. En consecuencia, la compresión tiene un gran impacto en el rendimiento de las consultas de datos si se accede a un alto porcentaje de columnas mediante una consulta. (S.G.Yaman, 2012).

- **Materialización:** La operación que transforma las columnas en filas se llama materialización. Una estrategia de materialización temprana, es cuando tan pronto empieza el procesamiento de una consulta empieza a formar tuplas formateadas en fila. Por otro lado, una estrategia de materialización tardía, no comienza a formar tuplas mientras una parte del plan de consulta se haya procesado. (Liu, Chen, M., & Otros, 2015).
- **Iteración por bloque:** para el procesamiento de una serie de tuplas, primero las tiendas de filas iteran a través de cada tupla, y luego requieren extraer los atributos necesarios mediante una interfaz de representación, los valores de los bloques de la misma columna se envían a un operador en una sola llamada a una función. Además, no se necesita la extracción de atributos, y si la columna tiene un ancho fijo, estos valores se pueden iterar directamente como una matriz. (J.Abadi, Madden, & Hachem, 2008)
- **Eficiencia en los operadores join:** con la reaparición de las tiendas de columnas y los trabajos realizados de joins de MonetDB y C-store ha promovido una serie de investigaciones para contar con joins eficientes en memoria principal, lo que representa grandes oportunidades para mejorar el rendimiento en tiendas de columnas. (D.Abadi, Boncz, & Harizopoulos, 2009).

Es importante mencionar que el término NoSQL describe las bases de datos no relacionales de alto desempeño, que utilizan varios modelos de datos incluidos los de tipo documentos, gráficos, claves-valores y columnas. (Stonebraker, 2010).

Con el crecimiento del volumen de los datos, fueron investigadas más soluciones escalables para las bases de datos ACID existentes, por tanto, Brewer y su grupo de trabajo (Brewer, 2012), crean un nuevo paradigma BASE (Basic, Availability, Soft-state, Eventual Consistency), los diseñadores e investigadores han utilizado el teorema de Brewer o CAP (Consistency, Availability, Partition Tolerance) como una razón para explorar una amplia variedad de novedosos sistemas distribuidos.

Según el teorema cuando se trabaja con sistemas distribuidos es imposible contar con las tres características simultáneamente, por tanto, es necesario renunciar parcialmente a una para obtener las otras dos (Brewer, 2012).

A continuación, en la Figura 4. se visualiza el Teorema Brewer.

Figura 4

Teorema Brewer



Nota: Recuperado de (Brewer, 2012)

Metodologías para Soluciones de Inteligencia de Negocio

Son creadas con el fin de entender cómo una organización puede crear inteligencia de sus datos, y así proveer a los usuarios finales información oportuna y acertada.

Metodología Hefesto

Según (Bernabeu, 2010), es una metodología que permite la construcción de un Data Warehouse, de forma sencilla, ordenada e intuitiva. Está fundamentada en una amplia investigación, comparación de metodologías existentes, experiencias propias del autor y feedbacks realizados en procesos de confección de almacén de datos.

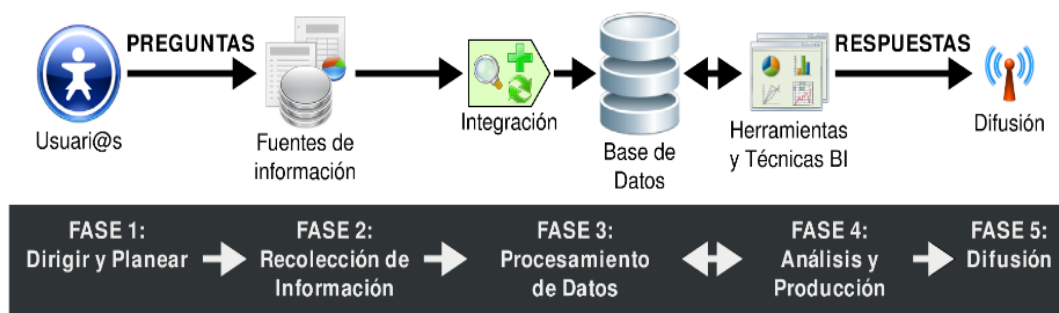
Su objetivo principal, es la comprensión de cada paso que se realiza, para no seguir un método al pie de la letra sin saber exactamente qué se está haciendo, ni por qué.

(Bernabeu, 2010)

A continuación, en la Figura 5. se presenta los procesos de la Inteligencia de Negocios definido según la metodología de HEFESTO y una breve descripción.

Figura 5

Arquitectura de Inteligencia de Negocios



Nota: Recuperado de (Bernabeu, 2010)

- **FASE 1: Dirigir y Planear** en esta fase inicial es donde se deberán recolectar los requerimientos de información específicos de los diferentes usuarios, así como entender sus necesidades, para que luego en conjunto con ellas se generen las preguntas que ayudarán a alcanzar los objetivos planteados.
- **FASE 2: Recolección de Información** es aquí, en donde se realiza el proceso de extraer desde diferentes fuentes de información de la empresa, tanto internas como externas, los datos que serán necesarios para encontrar las respuestas a las preguntas planteadas en el paso anterior.
- **FASE 3: Procesamiento de Datos** en esta fase, es donde se integran y cargan los datos en crudo en un formato utilizable para el análisis. Esta actividad puede realizarse mediante la creación de una nueva base de datos, agregando datos a una base de datos ya existente o bien consolidando la información en un nuevo repositorio.
- **FASE 4: Análisis y Producción** con los datos extraídos e integrados, se crea reportes de información utilizando herramientas de inteligencia de negocios. Como resultado final de esta fase se obtiene las respuestas de las preguntas planteadas.
- **FASE 5: Difusión**, esta fase se les entregará a los usuarios que lo requieran las herramientas necesarias, que les permitirán explorar los datos de manera sencilla e intuitiva.

Según (Bernabeu, 2010), entre los beneficios más importantes que BI proporciona a las organizaciones, vale la pena destacar lo siguiente:

- Reduce el tiempo mínimo que se requiere para recoger toda la información relevante de un tema en particular, ya que la misma se encontrará integrada en una fuente única de fácil acceso.
- Automatiza la asimilación de la información, debido a que la extracción y carga de los datos necesarios se realizará a través de procesos predefinidos.
- Proporciona herramientas de análisis para establecer comparaciones y tomar decisiones.
- Cierra el círculo que hace pasar de la decisión a la acción.
- Permite a los usuarios no depender de reportes o informes programados, porque los mismos serán generados de manera dinámica.
- Posibilita la formulación y respuesta de preguntas que son claves para el desempeño de la organización.
- Permite acceder y analizar directamente los indicadores empresariales.
- Se pueden identificar cuáles son los factores que inciden en el buen o mal funcionamiento de la organización.
- Se podrán detectar situaciones fuera de lo normal.
- Permitirá predecir el comportamiento futuro con un alto porcentaje de certeza, basado en el entendimiento del pasado. Los usuarios podrán consultar y analizar los datos de manera sencilla e intuitiva.

Metodología de Kimball

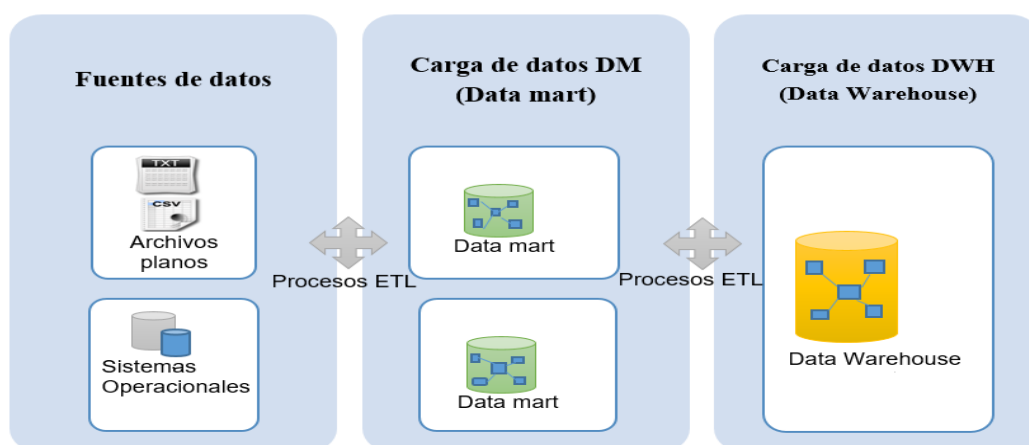
La metodología se basa en lo que Kimball denomina ciclo de vida dimensional del negocio, este ciclo de vida del proyecto de DW, según (Rivadera, 2010), está basado en cuatro principios básicos descritos de forma breve a continuación:

- **Centrarse en el negocio:** permite la identificación de los requerimientos y su valor asociado con el negocio.
- **Construir una infraestructura de información adecuada:** enfocado en el diseño de una base de información única, integrada, fácil de usar, de alto rendimiento que permita reflejar todos los requerimientos de negocio identificados en la empresa.
- **Realizar entregas en incrementos significativos:** permite la definición del plazo de los entregables en plazos de 6 a 12 meses, mediante la aplicación del concepto de incremento.
- **Ofrecer una solución completa:** contemplar todo el ciclo del proceso de inteligencia de negocios. Además, entregar herramientas de consulta ad hoc, aplicaciones para informes y análisis avanzado, capacitación, soporte, sitio web y documentación. Para contar con un almacén de datos sólido, bien diseñado, con calidad probada, y accesible.

El enfoque según Kimball, se enfoca en almacenar los datos en una estructura de forma dimensional porque permite optimizar las consultas a la base de datos y obtener mejores tiempos de respuesta para el acceso a la información de los diferentes procesos de negocio. Siendo en si una copia de datos transaccionales para ser analizados. En la Figura 6. se presenta el enfoque de Kimball y una breve descripción.

Figura 6

Enfoque Kimball solución Inteligencia de Negocios



Nota: Recuperado de (Rivadera, 2010)

El enfoque de Kimball, proporciona un enfoque de menor a mayor, muy versátil, por que indica que un data warehouse de una organización lo conforman varios data mart es decir bases de datos de menor tamaño de un departamento, mediante la ejecución de procesos ETL. Esto permite flexibilidad en el desarrollo, puesto que se incrementa componentes al análisis según las necesidades que tiene el área de interés.

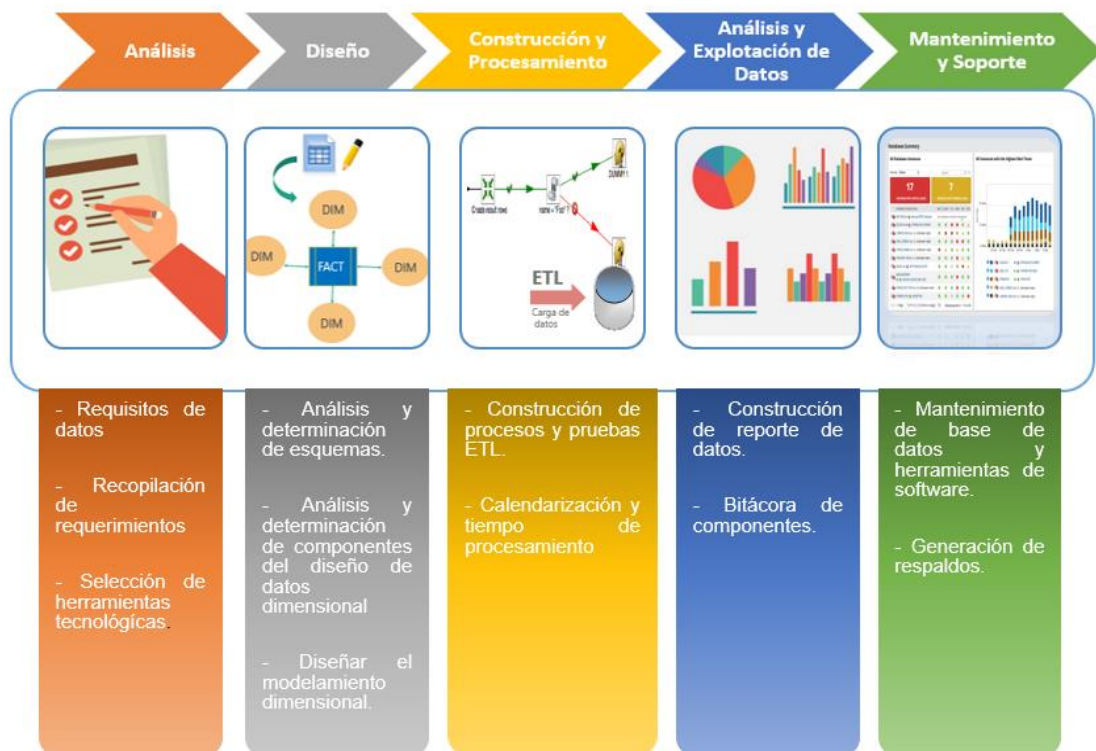
Propuesta metodológica para la integración de datos

Según (B.Husemann, Lechtenborger, & Vossen., 2000), las dificultades más recurrentes al momento de construir una data warehouse se encuentran en la recolección de requerimientos, el análisis, el diseño y selección de herramientas, debido a la falta de una metodología estándar para su desarrollo, la motivación principal es servir de instrumento para el ara técnica encargada de futuras implementaciones y mantenimiento.

La propuesta metodológica de la solución, presentada en la Figura 7. se basa en una adaptación de la metodología de Hefesto, descrito en la sección anterior, porque permite cubrir en general, todo el ciclo de implementación de la solución de inteligencia de negocio, para el caso de estudio.

Figura 7

Fases de la propuesta Metodológica



En base a lo expuesto, se presenta una breve descripción donde se define cada fase de la metodología de la solución propuesta.

Análisis. La primera fase de la metodología propuesta, permite conocer los requisitos, requerimientos que tiene el área de interés y delimitar las funcionalidades que tendrá su implementación.

Esta fase es importante, porque es aquí donde se establece la entrada de datos al data warehouse, por tanto se define cuáles son las fuentes de datos de origen, el tipo, procedencia, periodicidad, formatos, y periodicidad. Además, el análisis para escoger las herramientas tecnológicas apropiadas para el desarrollo de la solución.

A continuación, se presenta una breve descripción de cada ítem de la primera fase – *Análisis*, para el desarrollo de la implementación.

La **definición de los requisitos**, permite conocer las condiciones necesarias sobre la entrada de datos al DW, esto abarca desde conocer las cuáles serán las fuentes de datos de origen, el tipo, formatos, periodicidad de carga, y año mes al cual corresponde dicha información.

Este paso es importante porque determina y estandariza el proceso de integración de datos desde su punto de entrada, de tal forma de que pueda fluir los datos hasta lograr consolidarse en el DW y lograr su objetivo.

La **definición de los requerimientos**, permite identificar los requerimientos de los usuarios a través de diferente tipo de preguntas, de tal manera de conocer los objetivos de la organización.

Las preguntas planteadas serán analizadas a fin de identificar cuáles serán los indicadores y perspectivas que serán tomadas en cuenta para la construcción del DW. La idea central es, la formulación de preguntas complejas sobre el negocio, que incluyen variables de análisis que se consideren relevantes, ya que son estas las que permitirán estudiar la información desde diferentes perspectivas. (Bernabeu, 2010).

En consecuencia, en la fase de diseño con todos los requerimientos es posible levantar un modelo conceptual, donde se podrá visualizar el resultado obtenido.

El **análisis de las herramientas tecnológicas**, permite realizar una investigación cada cierto tiempo, para conocer las nuevas características, funcionalidades de las herramientas informáticas disponible en el campo de la inteligencia de negocios y la factibilidad de ser implementadas. Por tanto, la institución podrá estar acorde como avanza la tecnología y mantener el dinamismo que se requiere tener hoy en día en las aplicaciones y sistemas de información.

Diseño. La segunda fase de la metodología propuesta, permite recoger el análisis de requerimientos mediante la determinación de hechos, dimensiones y finalmente la creación del diseño dimensional para almacenar los datos.

A continuación, se presenta una breve descripción de cada ítem de la segunda fase – *Diseño*, para el desarrollo de la implementación.

El **análisis y determinación de esquemas**, permite definir en qué estructura de datos se va a guardar la información, con el fin de organizar y conocer el flujo de datos del proceso de integración. Se considera importante determinar cada repositorio de datos en función al contexto y utilidad que se le dará, de tal forma de organizar los datos esquemas, base de datos, vistas, entre otras; según lo establezca el equipo técnico que lleva a cabo este proceso.

Los **análisis y determinación de componentes del diseño de datos dimensional**, es determinar según los requerimientos establecidos cuáles serán los hechos (tabla de hechos), las perspectivas (tablas dimensionales) y medidas; de donde se extrae la información para ser analizada.

El **diseño del modelo dimensional**, permite conceptualizar todos los componentes analizados en el punto anterior, de manera que, se cree una estructura de datos adecuada que soporte la carga de datos al DW de forma completa.

Construcción y procesamiento. La tercera fase de la metodología propuesta, abarca la creación de procesos de extracción, transformación y carga (ETL), que permiten la carga de datos de un repositorio a otro, mediante la depuración, limpieza y conversión de datos de forma automática.

Además, otro punto es determinar la carga y procesos que se ejecutarán en el servidor establecido para la integración de datos, a fin de garantizar un óptimo rendimiento y calendarizar adecuadamente cada proceso.

A continuación, se presenta una descripción breve de cada ítem del ítem de la tercera fase –*Construcción y Procesamiento*, para el desarrollo de la implementación.

La **construcción de procesos y pruebas ETL**, al diseñar y crear los procesos ETLs es fundamental considerar que sean lo más eficientes posible, ya que una vez que se tenga un gran volumen de datos, el espacio en disco, el tiempo utilizado en procesamiento y acceso a la información serán esenciales para la carga de datos. (Bernabeu, 2010).

Para mejorar el rendimiento de un DW, se pueden llevar a cabo ciertas acciones como utilizar claves subrogadas, técnicas de indexación, técnicas de particionamiento, crear diferentes niveles de sumarización, crear vistas materializadas, técnicas de administración de datos en memoria caché, utilizar técnicas de multiprocesamiento de forma concurrente, con el objetivo de agilizar la obtención de resultados. (Bernabeu, 2010).

Además, es fundamental comprobar que los valores de datos luego de la transformación son los esperados mediante pruebas, y así asegurar la confiabilidad de la construcción de los procesos ETLs creados.

La **calendarización y tiempo de procesamiento**, permite conocer la carga de trabajo que tendrá un servidor y el tiempo estimado que demorará cada proceso en ser ejecutado, por lo tanto, es importante llevar una bitácora de procesos ETLs para estar informados y evitar cualquier contra tiempo en el caso de existir algún incidente.

Análisis y Explotación de Datos. La cuarta fase de la metodología propuesta, permite aprovechar todos los datos almacenada y extraerlos mediante la utilización de herramientas de inteligencia de negocios apropiadas para analizar la información; realizar comprobaciones permitirá verificar su funcionamiento y mostrar la información según los requerimientos del usuario. Finalmente, documentar el proceso mediante una bitácora de componentes servirá para conocer cualquier aspecto específico de una aplicación o reporte de datos planteado.

A continuación, se presenta una breve descripción de cada ítem de la cuarta fase – ***Análisis y Explotación de Datos***, para el desarrollo de la implementación.

La **construcción de reportes de datos**, dar un contexto a los datos es fundamental para convertirla en información, por tanto, es importante la utilización de herramientas de inteligencia de negocio u otra aplicación que permita extraer los datos almacenados en el data warehouse. Por lo cual, en este punto se debe tener claro los requerimientos planteados en la primera fase de la metodología, con el fin de responder a las preguntas planteadas en los reportes de datos, y ayudar al usuario final a analizar los datos y tener acceso a la información disponible tantas veces sea necesario.

Bitácora de componentes, permite llevar un registro de cada componente solicitado, de tal manera de documentar detalles específicos como por ejemplo un campo que se obtenga de la generación de una fórmula planteada por el usuario final.

Mantenimiento y soporte. La quinta fase de la metodología propuesta, permite tener el control de los cambios realizados en cualquier componente de la solución de inteligencia de negocios, porque es fundamental realizar el mantenimiento de tablas muy extensas o tuning de la base de datos para mejorar su rendimiento. Además, se requiere solventar cualquier incidente ocasionado en las herramientas de inteligencia de negocios y manejar una bitácora de errores y soluciones para evitar futuros inconvenientes.

A continuación, se presenta una descripción breve de cada ítem de la quinta fase - *Mantenimiento y Soporte*, en el desarrollo de la implementación:

El **mantenimiento de base de datos y herramientas de software**, este punto es importante porque permite abarcar los trabajos de mantenimiento para mejorar el rendimiento de las bases de datos o cualquier herramienta de software que intervenga en el proceso de gestión e integración de datos.

La **generación de respaldos** es relevante por temas de contingencia a fin de garantizar la operación del data warehouse, procesos ETLs, reportes de datos de forma permanente, y salvaguardar todos los componentes necesarios para el desarrollo de la solución de inteligencia de negocios.

En general, la metodología propuesta para el caso de estudio planteada se basa en Hefesto, sin embargo, se incluyó otros puntos como la i) calendarización y tiempo de procesamiento, ii) bitácora de componentes, iii) mantenimiento de base de datos y herramientas de software, y iv) generación de respaldos; con el fin de garantizar la disponibilidad y acceso a la información, y realizar acciones de contingencia en el caso de sufrir cualquier incidente que afecte la entrega de datos al usuario final.

Revisión de Literatura

Se realizó una revisión sistemática de literatura como un medio para que la investigación, sirva de apoyo para responder las preguntas planteadas.

Los criterios de inclusión definidos para la presente revisión sistemática de literatura son los siguientes:

- Con la finalidad de analizar ciertas técnicas y tecnologías utilizados en proyectos ágiles de data warehouse, fueron tomados en cuenta artículos a partir del año 2010.
- Los artículos contengan información sobre el objetivo general a ser analizado.
- Fueron considerados solamente artículos científicos y documentos de conferencias publicados en idioma inglés.

Los criterios de exclusión definidos para la presente revisión sistemática de literatura son los siguientes:

- Artículos que tengan temas de construcción de data warehouse con principios o técnicas tradicionales.
- Artículos que no estén en idioma de inglés.

Al final de la revisión de literatura se ha seleccionado una cadena de búsqueda.

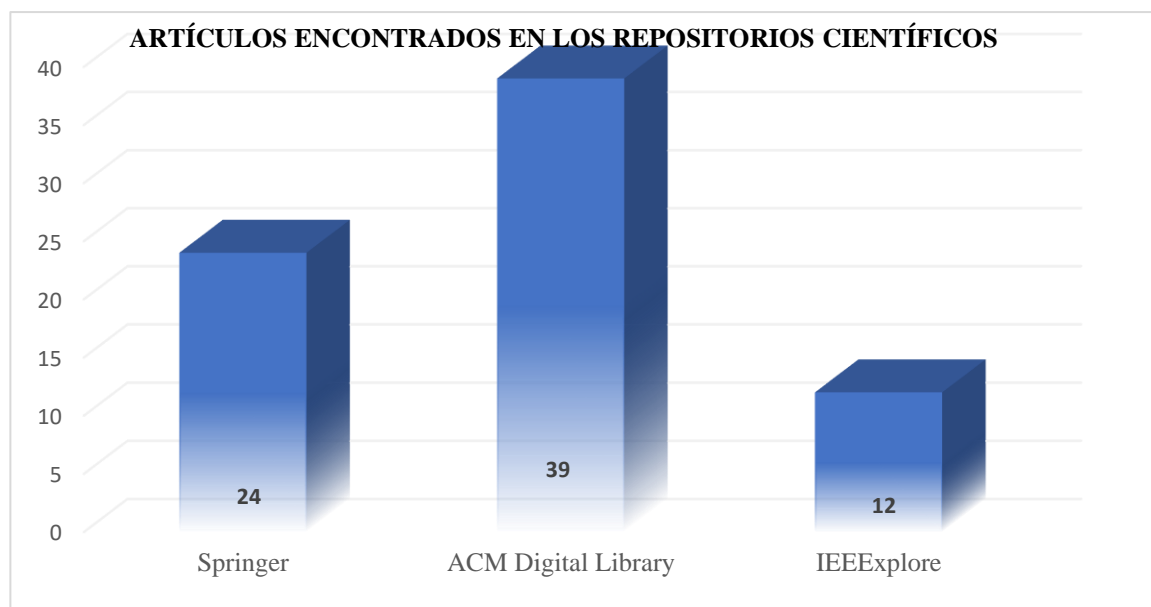
Para la construcción de la cadena de búsqueda se usan palabras que más se repiten en cada contexto definidos a partir de los estudios del grupo de control, que están direccionados al objetivo de la investigación, para ello se formaron los siguientes contextos: agile data warehouse, agile buisness intelligence, agile techniques for data warehouse, columnar database in memory.

**((agile Data Warehouse) OR (big data warehousing)) AND
(columnar in memory databases)**

A continuación, se presenta en la Figura 8. el número de artículos académicos encontrados en los repositorios científicos.

Figura 8

Artículos académicos encontrados



Al aplicar la cadena de búsqueda en los repositorios académicos seleccionados para la investigación, se filtra los documentos cuya fecha de publicación sea mayor al año 2010, sean artículos científicos o documentos de conferencias y estén en idioma inglés, por lo tanto, se obtienen los siguientes resultados presentados en la Figura 8.

Análisis de estudios encontrados para obtener información relevante sobre el objeto de la investigación.

Los estudios revisados se detallan a continuación:

Agile Data Warehouse – The Final Frontier (Bunio, 2012), el autor establece que los proyectos de data warehouse son diferentes a los proyectos de software por lo cual proporciona una guía de prácticas ágiles para el proceso de implementación de una data warehouse para ser evaluados y utilizados por otros proyectos es decir permite conocer ciertos principios y técnicas con un enfoque ágil.

Este estudio concluye la importancia de los modelos de datos ágiles, la utilización de subversión en la integración del modelamiento, añadir entidades o elementos de manera oportuna para alcanzar un equilibrio e integrar y automatizar el refactoring de la base de datos como piza clave para el éxito de un proyecto.

Agile BI – The Future of BI (Mihaela & Surcel, 2013), el estudio proporciona un análisis sobre los elementos claves que promueven una solución ágil los cuales son: una metodología de desarrollo ágil, ágil Business Analytics y una infraestructura de información ágil; además un breve análisis de las tecnologías que se pueden utilizar para crear una solución de BI ágil. El estudio revela que el objetivo de la solución debe implementarse de forma rápida y reaccionar más rápido a los requisitos.

El estudio concluye, las principales razones para implementar un BI ágil son: requisitos constantes cambiantes, incapacidad de TI para satisfacer las demandas de los usuarios, acceso lento a la información; las soluciones de BI ágil permiten a las organizaciones adaptarse a las cambiantes condiciones del negocio.

Big Data and New Data Warehousing Approaches (Ptiček & Vrdoljak, 2017), los autores indican que los sistemas tradicionales no soportan la analítica avanzada y la extracción del conocimiento no es capaz de hacer frente a grandes cantidades de datos diversos y no estructurados, por lo que estamos enfrentando un cambio de paradigmas en términos, tecnologías, metodologías utilizados, este tema se ha convertido en un área de investigación activa en los últimos años, por lo tanto el documento proporciona una visión general de las tendencias de estudio más importantes para big data warehouse, conceptos y tecnologías utilizados para el almacenamiento, ETL y los enfoques de investigación realizados en intentos de potenciar la guerra de datos tradicional.

The Data Warehousing (R) Evolution: Where's it headed next? (Smith & Rege, 2017), los autores indican que el documento de estudio permite conocer las principales tendencias en big data, lógica data warehouse, transacción híbrida, procesamiento analítico y computación en memoria. Además, señalan que las prácticas de data warehouse se han mantenido en los últimos treinta años con la tasa actual de cambio tecnológico y crecientes demandas analíticas, Gartner predice que para el 2019 las prácticas tradicionales de almacenamiento de datos ya no serán relevantes. Las nuevas tecnologías y arquitecturas de datos tendrán un impacto similar en el mundo de la gestión de información como la revolución industrial tuvo en el ámbito de la fabricación.

The Impact of Columnar In Memory Databases on Enterprise Systems (Plattner, 2014), el autor indica que los primeros análisis de aplicaciones productivas que han adoptado este concepto confirman que las arquitecturas en columnas habilitados en memoria son conceptualmente superiores para el procesamiento de transacciones en comparación con los enfoques basados en filas. Además, señala que es ampliamente adoptado en la industria y está comprobado que es factible ejecutar consultas analíticas en grandes conjuntos de datos directamente en un esquema libre de redundancia, eliminando la necesidad de mantener tablas agregadas preconstruidas.

El autor concluye, que en los últimos cinco años han demostrado la viabilidad de este enfoque y muchas empresas ya tienen esta arquitectura de base de datos en uso de forma productiva. Las experiencias adquiridas han confirmado que los sistemas basados en columnas se mantienen sin ninguna transacción y establece que todas las aplicaciones empresariales se construirán de forma agregada y sin redundancia en el futuro.

Los documentos escogidos en la revisión de literatura, mencionan que las consultas analíticas de grandes conjuntos de datos, gestionadas mediante tecnologías con una arquitectura datos columnar, elimina la necesidad de mantener tablas agregadas preconstruidas para la operación empresarial. Los primeros análisis de aplicaciones productivas para almacenes de columnas en memoria son conceptualmente superiores para el procesamiento de transacciones comerciales, en comparación con los enfoques basados en filas.

Las técnicas de integración de datos van desde el procesamiento de datos por lotes, eventos hasta mensajería. Se recomienda, aprovechar el desarrollo de las transformaciones de datos complejas en otros proyectos empresariales, las herramientas de integración de datos posean conectores y adaptadores que utilicen APIs nativas tradicionales ODBC / JDBC, para un rendimiento óptimo de datos.

Finalmente, los proyectos de data warehouse son diferentes a los de software por lo que, necesitan un conjunto de prácticas propias y el uso de nuevas herramientas de gestión de datos versátiles que permitan una gestión de datos eficaz. Los documentos revisados proponen adoptar las nuevas tendencias existentes en proyectos de éxito, a fin de aumentar la escalabilidad de procesamiento de datos y mejorar la solución de inteligencia de negocios empresarial.

Capítulo III

Problemática Actual y Propuesta de la Solución de Inteligencia de Negocio

En este capítulo, se presenta un análisis de los problemas existentes en el caso de estudio, en el ámbito de la gestión de datos, y en base a este análisis se genera una propuesta de la solución de inteligencia de negocio, con la finalidad de determinar los mecanismos y tecnologías más adecuadas que permitan contrarrestar los problemas que dificultan la consecución del proceso de integración de datos.

Análisis de Problemas

Problema I. El acceso a los datos al repositorio central (DW) cada vez tarda más tiempo, la DRIPS cuenta con una base de datos multidimensional almacenado en varios data mart, bajo este contexto, es importante analizar un nuevo modelo de datos dimensional y otros motores de gestión de datos que permita acceder de forma más rápida y así evitar futuros incidentes.

Problema II. Falta de soporte técnico y actualización de herramientas de software para la gestión e integración de datos, lo cual ocasiona que el personal técnico encargado designe un mayor tiempo en la solución de estos inconvenientes.

El factor económico para acceder al soporte y nuevas actualizaciones de las herramientas propietarias sin duda es un factor relevante, por tanto, en función de las políticas de gobierno es necesario utilizar herramientas open source para efectuar cualquier proceso.

Problema III. En general, la solución de inteligencia de negocios no se ha actualizado en la institución en los últimos años, no cuenta con una metodología para la integración de datos que permita llevar a cabo el proceso de forma sistemática, por lo cual, es necesario analizar e integrar nuevas tecnologías que permitan la consecución de los procesos de gestión de datos forma eficaz.

En base a los problemas identificados, se presenta respectivamente las acciones o alternativas para solventarlos.

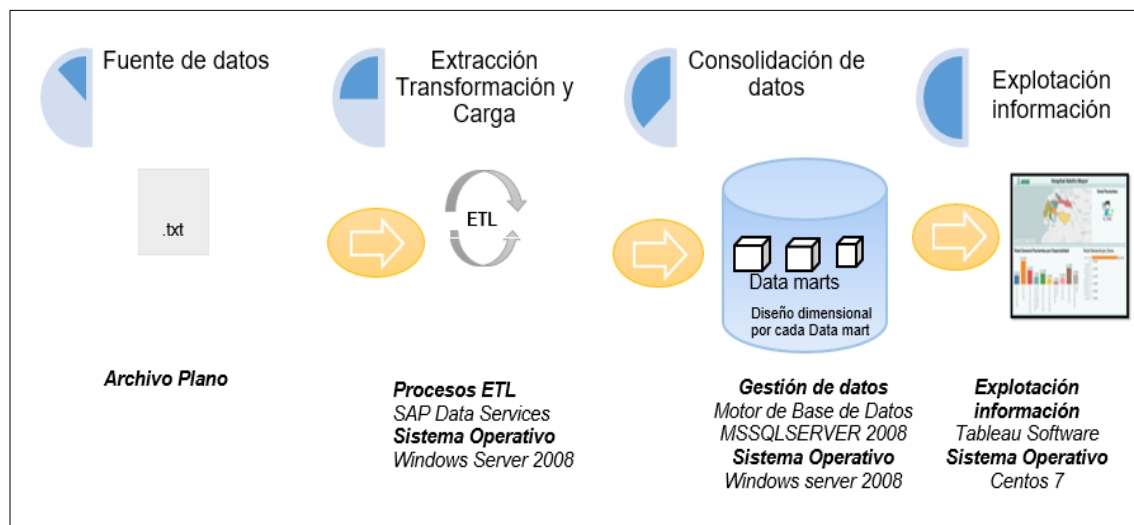
Solución I. Es fundamental analizar y generar un modelo dimensional que capture todos los campos necesarios en base a las necesidades actuales del caso de estudio, donde el diseño permita un acceso rápido a la información.

Solución II. Investigar, analizar y seleccionar herramientas open source para la gestión e integración de datos que permitan realizar las mismas tareas de forma eficiente.

Solución III. Proponer una metodología para la integración e implementar el modelo dimensional propuesto, a fin de evaluar los resultados obtenidos y la efectividad de la solución propuesta.

Con las alternativas mencionadas, se pretende aplacar los problemas existentes y realizar una propuesta de inteligencia de negocio escalable, eficaz y asequible en un entorno donde existe rotación de personal, en ocasiones un presupuesto limitado para acceder a herramientas tecnológicas y la necesidad de tener internamente un dinamismo en los procesos y así responder rápidamente a los requerimientos solicitados.

A continuación, se presenta en la Figura 9. una representación de la arquitectura actual de inteligencia de negocios del caso de estudio.

Figura 9*Arquitectura Actual de Inteligencia de Negocios***Arquitectura General de la Solución**

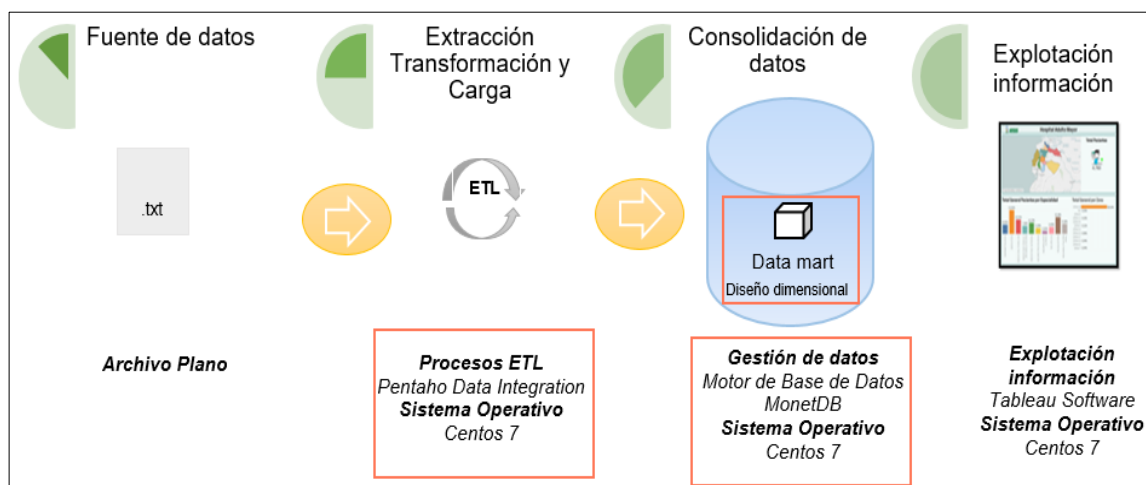
Reestructurar la arquitectura de la solución de inteligencia de negocio permite visualizar nuevas formas de realizar el proceso de gestión de la información, los componentes empleados y la interrelación entre cada uno.

La arquitectura general de la solución, se definió según lo expuesto según (Bernabeau D. , 2010).

A continuación, en la Figura 10. se presenta una propuesta de la arquitectura de inteligencia de negocio y en el capítulo 4, se describe el desarrollo del estudio investigativo específicamente el análisis de escoger dichas herramientas, la implementación y los resultados obtenidos.

Figura 10

Propuesta de la Arquitectura de Inteligencia de Negocios



Fuentes de datos aquí se obtiene del insumo principal para el proceso lo conforman los archivos planos en formatos .txt y csv.

Procesos ETL siendo de vital importancia en el proceso de integración de datos porque es aquí donde se construyen tareas automáticas que permiten extraer, transformar la data para mejorar su calidad y cargarlos al repositorio final para ser consumidos por otras aplicaciones. Este proceso se lo realiza mediante la utilización de herramientas de integración de datos, para el caso de estudio se seleccionó Pentaho Data Integration, según el análisis realizado en el capítulo 4. Además, permiten almacenar depurar los datos para guardar información valida a la base de datos del data warehouse.

Consolidación DWH donde mediante procesos ETL se carga la información en un modelo de datos dimensional que facilita el acceso a los datos, para la solución de inteligencia de negocios se escogió la herramienta MonetDB la cual permite un almacenamiento columnar.

Visualización de la Información, finalmente se tiene como último proceso la presentación de la información mediante la utilización de una herramienta de inteligencia de negocios, para nuestro caso de estudio se escogió Tableau Software debido a que la institución tiene licencias para su utilización, en esta fase se construyen tableros de información para usuarios internos y externos, con el fin de servir de insumo para el análisis de los resultados y así apoyar la toma de decisiones efectivas.

Análisis de la Propuesta de la Solución de Inteligencia de Negocio

Es fundamental analizar nuevos componentes y tecnologías para el proceso de gestión e integración de datos a fin de tener una solución de inteligencia de negocios mejorada.

Modelo dimensional: los requerimientos han ido cambiando con el tiempo y se pretende almacenar otro tipo de datos, por lo cual es importante la generación de un nuevo modelo dimensional para captar todos los requisitos actuales y enfocarse en crear una estructura de datos que permite optimizar los tiempos de respuesta en las consultas de datos siendo este un factor clave en la gestión de datos.

Herramientas de integración de datos: un limitante es la falta de soporte y desactualización de las herramientas de gestión e integración de datos impidiendo acceder a tecnologías innovadoras con las que se cuenta hoy en día para los procesos ETL. Fue revisada la herramienta de integración de datos Pentaho Data Integration

Edition utilizada en otra área en la institución, porque posee nuevas características que serán descritas en el capítulo 4, lo cual fue un buen referente para proponer una herramienta.

Nuevo gestor de datos columnar: el gestor de base de datos utilizada en la institución es MSSQLSERVER, esta herramienta es robusta y útil para realizar diferentes operaciones como INSERT, SELECT, UPDATE, DELETE, sin embargo, cuando se requiere acceder a los datos de diferentes data mart toma un tiempo considerable, por tal motivo al utilizar una herramienta de gestión de datos columnar en un ambiente Linux se pretende aprovechar las características de estos gestores de base de datos y ayudar en los procesos de integración de datos. Además, el correr sobre servidores en ambientes Linux permite tener mayor estabilidad, rendimiento, rapidez y acceder a software open source robustos

Capítulo IV

Implementación

En este capítulo, se presenta el desarrollo del proyecto según la propuesta metodológica descrita, de tal manera que permita cubrir todas las fases de la solución de inteligencia de negocios y apoye en la construcción de futuras implementaciones en el área de interés de la institución.

Fase de Análisis

Requisitos de datos

Los requisitos de datos permiten conocer las condiciones necesarias y establecer mecanismos y lineamientos con el fin de determinar los procesos de integración de datos.

Para empezar con la definición de los requerimientos se establecieron primero los requisitos de datos, los cuales son:

- Identificar las distintas fuentes de datos, para lo cual se generó tres archivos planos de tipo csv y .txt, donde el delimitador de columnas es el (;); con datos de prueba creados por el autor de forma automática.
- El tipo de información que va a ser integrada al almacén de datos corresponde a información estructurada.
- Al ser una propuesta, no aplica definir la periodicidad de integración de datos, porque únicamente se cargará una vez, sin embargo, es fundamental documentar para conocer cada qué tiempo deben ser cargados los datos.
- Es importante, establecer a que año y mes corresponde dicha información.

Análisis de requerimientos

La definición de los requerimientos se levantó a través de una entrevista al personal técnico involucrado en el proceso de integración de datos, que conoce las necesidades actuales del caso de estudio y la explotación de la información que se debe realizar.

Cabe recalcar la comprensión de los requerimientos por parte del investigador que forma parte del equipo de trabajo en la institución.

Las especificaciones de requerimientos se basan en construir un reporte de datos que responda ciertas preguntas relevantes y permita el análisis de la información.

Los requerimientos son:

- ¿Cuántas personas son beneficiarias de algún programa determinado?
- ¿Cuántas personas son beneficiarias según una ubicación geográfica específica?
- ¿Cuántas personas son beneficiarias de algún programa según la variable sexo?
- ¿Cuántas personas son beneficiarias de algún programa y cuál es su nivel de bienestar?
- Permitir un análisis de información en un determinado tiempo.
- Permitir un análisis de información específica según el programa, la ubicación geográfica y el tiempo.

Selección de herramientas de tecnológicas

En base a la investigación realizada sobre **herramientas de gestión de datos con un almacenamiento columnar open source** se escogió a las herramientas Cassandra, HBase por ser consideradas dentro del ranking de la Base de Conocimiento de Sistemas de Gestión de Base de Datos Relacionales y NoSQL de DB-Engines (SOLID IT, 2019), y ocupar la primera, segunda puntuación; finalmente se seleccionó a la herramienta MonetDB por estar calificada como uno de los primeros gestores de base de datos multimodelo que permite gestionar como una base de datos relacional y almacenar datos en columnas. Además, MonetDB está considerada en el forecast para el 2026 (Markets Gazette, 2019), como uno de los principales gestores de base de datos clave en las tendencias actuales y la demanda futura.

Las herramientas Cassandra, Hbase, y MonetDB se destacan por aportar en proyectos de éxito en temas de educación, investigación y a nivel empresarial.

A continuación, se presenta en la Figura 11. la puntuación obtenida de las herramientas Cassandra, Hbase y MonetDB según DB-Engines (SOLID IT, 2019).

Figura 11

Puntuación Gestores de Base de Datos Cassandra, HBase, MonetDB

Rank			DBMS	Database Model	Score		
Oct 2019	Sep 2019	Oct 2018			Oct 2019	Sep 2019	Oct 2018
1.	1.	1.	Cassandra +	Wide column	123.22	-0.18	-0.17
2.	2.	2.	HBase	Wide column	54.84	-0.88	-5.84
54.	↑ 57.	↓ 53.	MonetDB +	Relational, Multi-model i	1.88	+0.22	+0.26

Nota: Recuperado de DB-Engines (SOLID IT, 2019)

Para la selección de las herramientas se formó un grupo focal conformado por 5 profesionales de inteligencia de negocios e integración de datos incluido el investigador del estudio, que forman parte del equipo de trabajo en la institución. Los participantes fueron escogidos debido al conocimiento sobre el entorno y necesidades requeridas, por no ser partner de ninguna herramienta que dé lugar algún sesgo, y por haber realizado proyectos importantes sobre el tema tratado, como se visualiza en la Tabla 2.

Tabla 2

Participantes del Grupo Focal

Integrantes	Rol	Perfil Profesional	Proyectos Realizados
Gabriela Villagómez	Moderador Investigador del presente trabajo.	Ing. Sistemas Informáticos y Computación. Experiencia en el tema de estudio: 5 años	Proyecto de Inteligencia de Negocios e Integración de Datos para el Registro Interconectado de Programas Sociales.
Carlos Encalada	Participante	Ing. Informático Experiencia en el tema de estudio: 11 años	Proyecto de Inteligencia de Negocios e Integración de Datos para el Registro Social.
Mavelin Ati	Participante	Ing. Sistemas Informáticos y Computación e investigador del presente trabajo. Experiencia en el tema de estudio: 2 años	Proyecto de Inteligencia de Negocios para la Secretaría Técnica Plan Toda Una Vida. Proyecto de Integración de Datos para el Registro de Datos Crediticios.

Integrantes	Rol	Perfil Profesional	Proyectos Realizados
Rosley Naranjo	Participante	Ing. Sistemas Experiencia en Inteligencia de Negocios e Integración de Datos: 5 años	Proyecto de Inteligencia de Negocios e Integración de Datos para el Registro e Interconectado de Programas Sociales.
Alejandra Cuadros	Participante	Mcs. Gestión de Sistemas de Información e Inteligencia de Negocios Experiencia en el tema de estudio: 6 meses	Proyectos: Mapeo sistemático de literatura asequibles y reproducibles aplicando gestión del conocimiento y minería de procesos

A continuación, se describe en la Tabla 3. las características más importantes de cada herramienta según DB-Engines (SOLID IT, 2019).

Tabla 3

Características de herramientas de gestión de datos columnar

Características	Cassandra	Hbase	MonetDB
Almacenamiento columnar	Si Gestor de base de datos orientada en columnas basados en ideas de BigTable and DynamoDB.	Si Gestor de base de datos orientado en columnas basados en ideas Apache Hadoop y conceptos de BigTable.	Si Gestor de base de datos multimodelo: i) Relacional, y ii) Orientado en columnas.

Características	Cassandra	Hbase	MonetDB
Licencia Open Source	Si	Si	Si
Multiplataforma	Si BSD, Linux, OS Windows	Si Linux Unix Windows (Usando Cygwin)	Si FreeBSD, Linux, OS X, Solaris, Windows
Soporte SQL	Si SQL-like SELECT, DML and DDL statements (CQL)	No	Si
Foreign Keys	No	No	Si
Procedimientos almacenados	No	Si	Si (En Sql, C,R)
APIS y otros métodos de acceso	Si Protocolo propietario Thrift	Java API RESTful HTTP API Thrift	Si Native C Library JDBC ODBC
Acceso a claves foráneas	No	No	Si
Triggers	Si	Si	Si
Concurrencia	Si	Si	Si
Durabilidad	Si	Si	Si
Concepto de usuarios	Si Acceso para los usuarios definido por objeto	Si Listas de control de acceso (ACL) para RBAC, integración con Apache Ranger para RBAC y ABAC	Si Acceso específico según el estándar SQL
Soporta conceptos de transacciones	No	Si ACID de una fila (en millones de columnas)	Si ACID

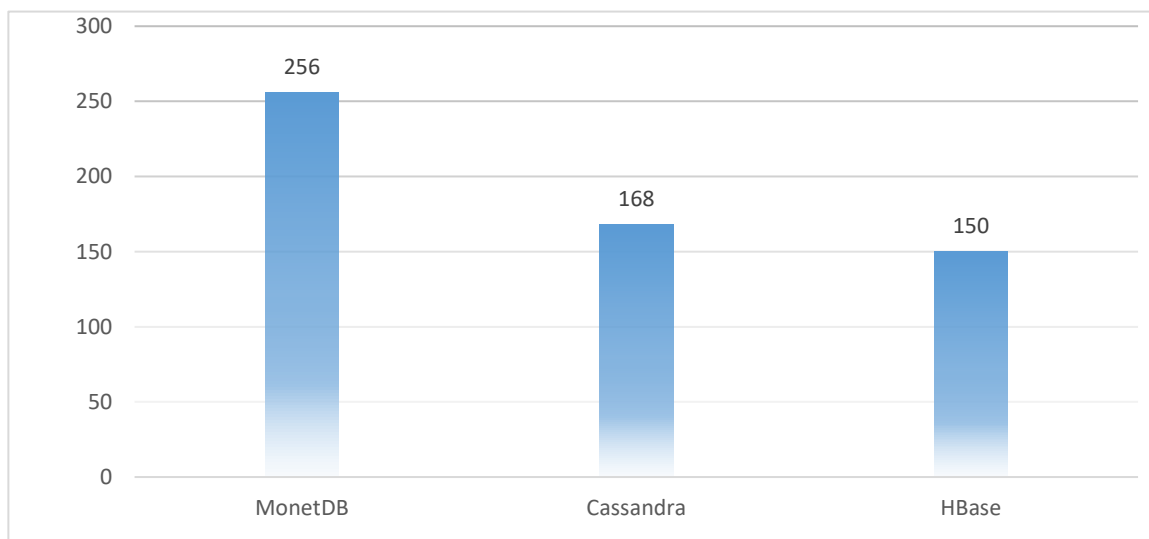
Características	Casandra	Hbase	MonetDB
Aplicación	Internet de las cosas (IOT), aplicaciones de detección de fraude, motores de recomendación, catálogos de productos y listas de reproducción y aplicaciones de mensajería.	Proyectos de gran procesamiento, almacenamiento y análisis de datos (Big data), informes de analítica de datos.	Base de datos analítica, Data warehouse, OLAP Data mining Scientific database

Nota: Recuperado de (SOLID IT, 2019)

Los integrantes del grupo focal evaluaron las características más importantes de cada herramienta, el resultado de la evaluación se visualiza en la Figura 12.

Figura 12

Resultado de la evaluación a gestores de base de datos columnares open source



Adicionalmente, para el análisis e investigación fueron probadas las herramientas de gestión de datos en un mismo ambiente en Linux comprobando así las características de cada herramienta. En consecuencia, se concluye la ventaja de **CASSANDRA** por su alta disponibilidad a través de la configuración entre nodos y semillas, lo cual permite garantizar su operación de forma continua. Sin embargo, la conexión para acceder a los datos es limitada, únicamente es posible establecer una conexión por medio del protocolo propietario thrift dificultando la comunicación con otras herramientas de software.

Por otra parte, sobre **HBASE** se puede señalar que es un gestor de base de datos que permite mantener los datos de forma redundante en varios nodos lo que permite una buena tolerancia a fallos, las consultas con funciones de agregación son relativamente rápidas, no obstante, debido a la falta de soporte de joins y acceso a claves foráneas utiliza mayor tiempo en acceder a datos de otras tablas.

Finalmente, **MONETDB** maneja transacciones apoyándose en sentencias SQL con un rápido tiempo de respuesta a los datos y la conectividad para acceder a los datos lo realiza a través de JDB, ODBC que son las implementaciones universales más utilizados permitiendo una mayor conexión con otras herramientas. Sin embargo, al trabajar con una arquitectura cliente servidor, no cuenta con otros nodos que permitan una alta disponibilidad y tolerancia a fallos.

Con el resultado de la evaluación sobre las características de las herramientas expuestas, y con la encuesta realizada al equipo de técnico de la Unidad del Registro Social, se decidió escoger la herramienta de gestión de datos columnar open source **MonetDB** como la más apropiada, porque posee características relevantes para desarrollar proyectos de éxito inteligencia de negocios con data warehouse de forma consistente y rápida, siendo una plataforma que maneja conceptos de relación y almacenamiento columnar lo que permite un acceso ágil cuando se realizan consultas en la misma tabla o tablas externas mediante claves foráneas, además tiene soporte con SQL lo cual permite un tiempo de aprendizaje corto para la implementación.

El proceso de integración se lo realiza por medio de **herramientas ETL Open Source** porque permiten consolidar la información de forma rápida, y por el procesamiento complejo que se realiza para procesar grandes cantidades de datos en tiempos adecuados, por lo cual se escogió las herramientas open source i) **Open Studio Data Integration** de Talend, ii) **Jaspersoft ETL Community** de TIBCO Software ii) **Pentaho Data Integration** de Hitachi Vantara para ser analizadas.

Estas herramientas son consideradas dentro del cuadrante de Gartner (*Gartner Inc, 2019*), en la parte de *Lideres* por haber obtenido mayor puntuación por la capacidad de visión de mercado y en la sección *Jugadores de Nicho* por concentrarse en segmentos del mercado, respectivamente, como se visualiza en la Figura 13.

Figura 13

Cuadrante de Gartner para Herramientas de Integración de Datos



Nota: Recuperado de (Gartner Inc, 2019)

A continuación, en la Tabla 4. se describe las características más importantes de cada herramienta.

Tabla 4

Características de herramientas de integración de datos Open Source

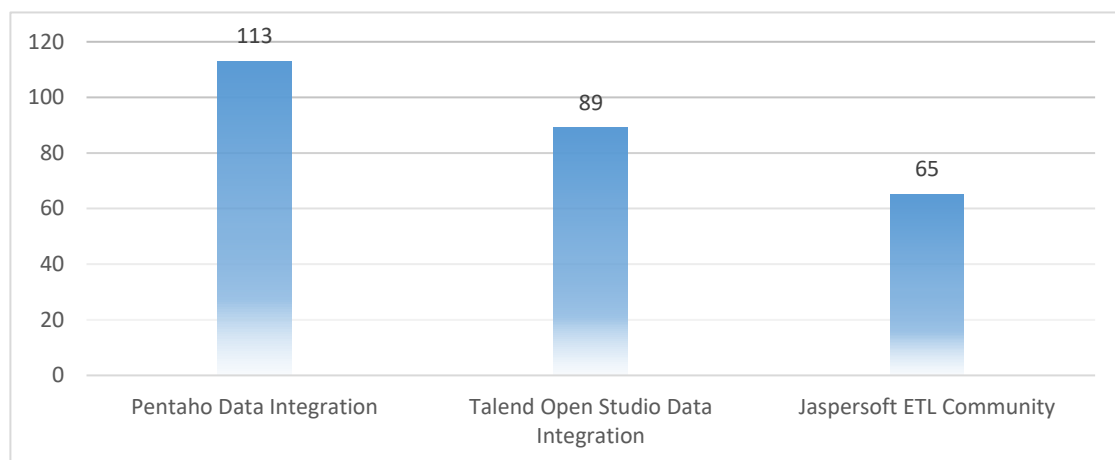
Características	Talend Open Studio Data Integration	Jaspersoft ETL Community	Pentaho Data Integration
Licencia Open Source	Si	Si	Si
Soporte Comunitario	Si	Si Soporte limitado para carga de trabajo en lotes. (Gartner Inc, 2019)	Si

Características	Talend Open Studio Data Integration	Jaspersoft ETL Community	Pentaho Data Integration
Versión Estable	Versión 7.3.1M2 Septiembre 2019	Versión 6.0.1 - released Marzo 2016	Versión 8.2 Octubre 2018
Multiplataforma	Si Windows Mac	Si Linux - versión 1.1.2 Windows Mac	Si Linux Windows Mac
Cargas de datos en lotes	Si	Si Por sí solo no puede abordar el alcance completo de las cargas de trabajo masivas / relacionadas con lotes. (Gartner Inc, 2019)	Si
Alto Rendimiento	Si	Si	Si
Conexión a múltiples fuentes de datos	Si Limitada	Si Limitada	Si Amplia

Con el mismo grupo focal conformado y detallado en la sección anterior, se evalúan las características de las herramientas de integración de datos analizados como se visualiza en la Figura 14.

Figura 14

Resultados de la evaluación de las herramientas de integración de datos open source



El resultado de la evaluación fue presentada al equipo de la institución, luego de un consenso se decidió utilizar la herramienta open source ***Pentaho Data Integration***, porque puede ser empleado en ambientes con sistemas operativos Linux permitiendo un procesamiento rápido y robusto; posee una amplia gama de componentes para procesar grandes cantidades de datos sin la necesidad de instalar otros productos, y principalmente porque parte del equipo de trabajo pertenecientes al grupo focal, adoptaron esta herramienta para la integración de datos en proyectos internos y según su experiencia el nivel de aprendizaje es rápido, fácil de instalar, configurar.

Integrantes del equipo de trabajo ha manifestado la utilización de la herramienta durante aproximadamente 6 años teniendo buenos resultados, dicha experiencia podrá ayudar en la construcción de futuros procesos ETL complejos; finalmente la herramienta cubre las necesidades el caso de estudio a ser desarrollado.

Fase Diseño

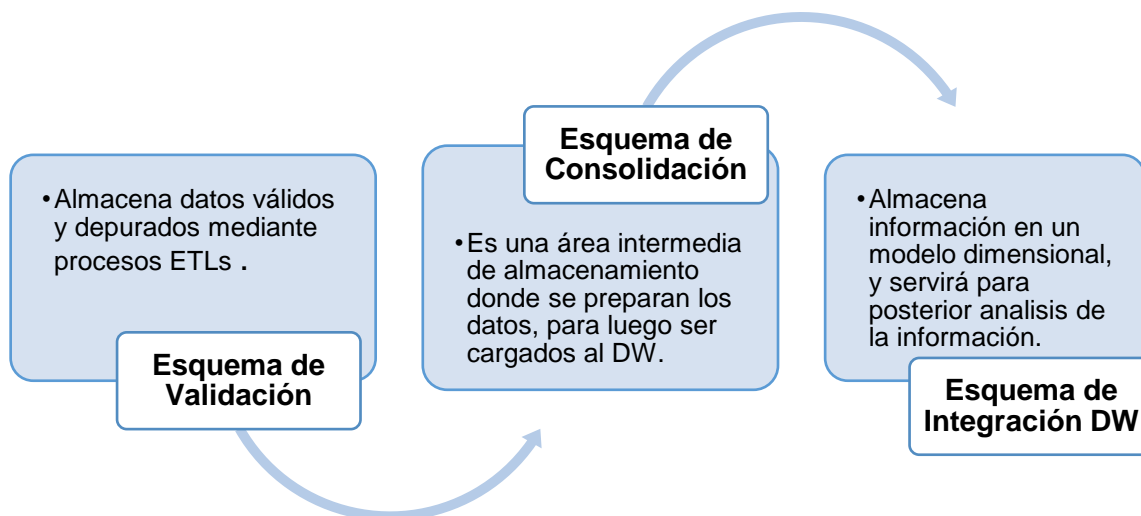
Análisis y determinación de esquemas

Para realizar la integración de datos al modelo dimensional se estableció tres esquemas con el fin de separar el proceso de validación, consolidación y la integración de datos al data warehouse.

A continuación, se presenta en la Figura 15. el proceso que se realiza en cada esquema de datos.

Figura 15

Esquemas para la implementación



Análisis y determinación de componentes del diseño de datos dimensional

En esta parte, se indica el análisis realizado para identificar las dimensiones, los diferentes hechos que serán almacenados para realizar el análisis de la información.

Para identificar los componentes del modelo dimensional, es importante responder una serie de preguntas que permitirán enfocarse en el objeto de análisis del caso de estudio:

Las dimensiones identificadas a través de ciertas preguntas, permiten conocer los indicadores de negocio, a continuación, se detallan cada uno:

- ¿Quién? ¿Quién generó el hecho? En el caso de estudio hace referencia a los beneficiarios de los distintos programas o servicios sociales, y personas en alguna condición de pobreza. Por lo tanto, se determina la dimensión DIM_BENEFICIARIO.
- ¿Que? ¿Qué deseo medir del hecho? El número de programas o beneficios que proporciona el estado por cada institución, establecimiento, su nivel de pobreza y la encuesta del Registro Social registrada. Por lo tanto, se determina la dimensión DIM_PROGRAMA, DIM_INSTITUCION, DIM_ESTABLECIMIENTO, DIM_NIVEL_BIENESTAR, DIM_CATASTRO y DIM_ENCUESTA.
- ¿Cuándo? ¿Cuándo ocurrió el hecho? Hace referencia a la fecha en la que ocurrió el hecho. Por lo tanto, se determina la dimensión DIM_FECHA.
- ¿Donde? ¿Donde ocurrió el hecho?, Hace referencia a la ubicación geográfica donde ocurrió el hecho. Por lo tanto, se determina la dimensión DIM_UBICACIÓN_GEOGRAFICA, cabe recalcar que el nivel de detalle que se aplicará es PROVINCIA, CANTÓN, PARROQUIA.

La tabla de hechos será definida por la cardinalidad de las perspectivas o dimensiones, es decir:

$$\text{TH} = \text{D1} \times \text{D2} \times \text{D2} \times \text{D3} \times \text{D4} \times \text{D5}$$

Donde:

TH: Tabla de Hechos BENEFICIOS

D1: Dimensión PROGRAMA

D2: Dimensión INSTITUCION

D3: Dimensión ESTABLECIMIENTO

D4: Dimensión NIVEL BIENESTAR

D5: Dimensión CATASTRO

D6: Dimensión ENCUESTA

D7: Dimensión FECHA

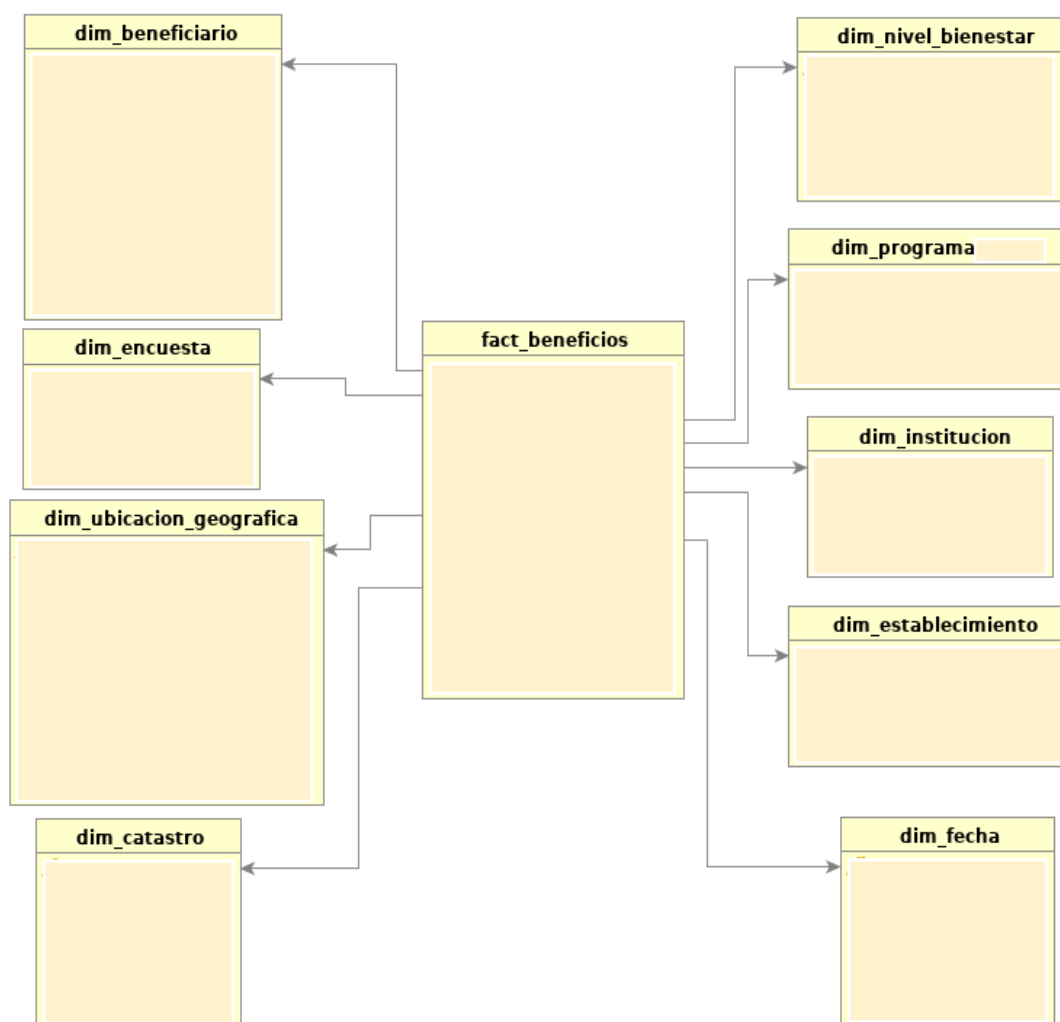
D8: Dimensión UBICACIÓN GEOGRÁFICA

Como se mencionó anteriormente, esta solución se enfoca en consolidar toda la información en una sola estructura de datos que permita analizar al beneficiario entorno a la información de los beneficios de programas sociales que recibe y a su condición de nivel de pobreza, para tener un mejor tiempo de respuesta en el acceso a los datos.

Diseño del modelo dimensional

En base al análisis de dimensiones y los hechos a continuación se presenta en la Figura 16. el modelamiento del diseño dimensional planteado.

Figura 16

Modelo Dimensional

Fase Construcción y Procesamiento

Construcción de procesos y pruebas ETL

En esta parte se construye los diferentes procesos de transformación de datos con la ayuda de los componentes de la herramienta de integración de datos seleccionada.

Las pruebas se realizan para verificar la completitud de los datos cargados desde la fuente al destino establecido en cada proceso luego de ser transformados.

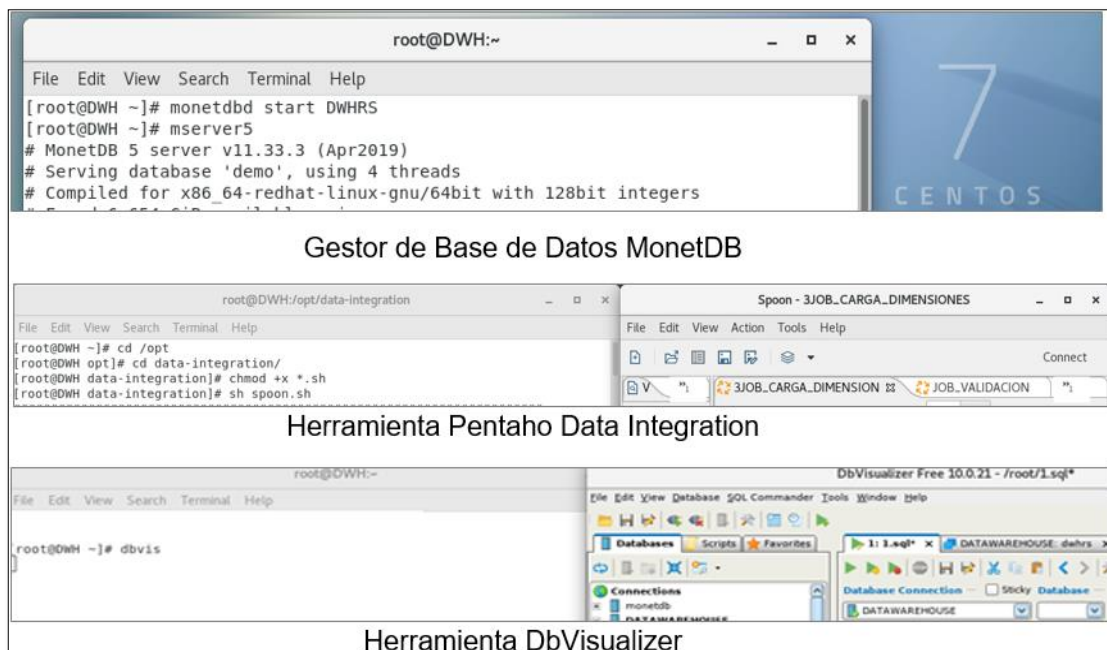
Para empezar con la construcción de los procesos ETL, se prepara el ambiente de trabajo, para lo cual se realiza la instalación de las herramientas seleccionadas:

- Gestor de base de datos: MonetDB-11.31.13
- Herramienta ETL: Pentaho Data Integration 7.1.0.12
- Herramienta de Administración Base de Datos: DbVisualizer Free

A continuación, se presenta en la Figura 17. la preparación del ambiente de implementación:

Figura 17

Preparación del Ambiente de Trabajo

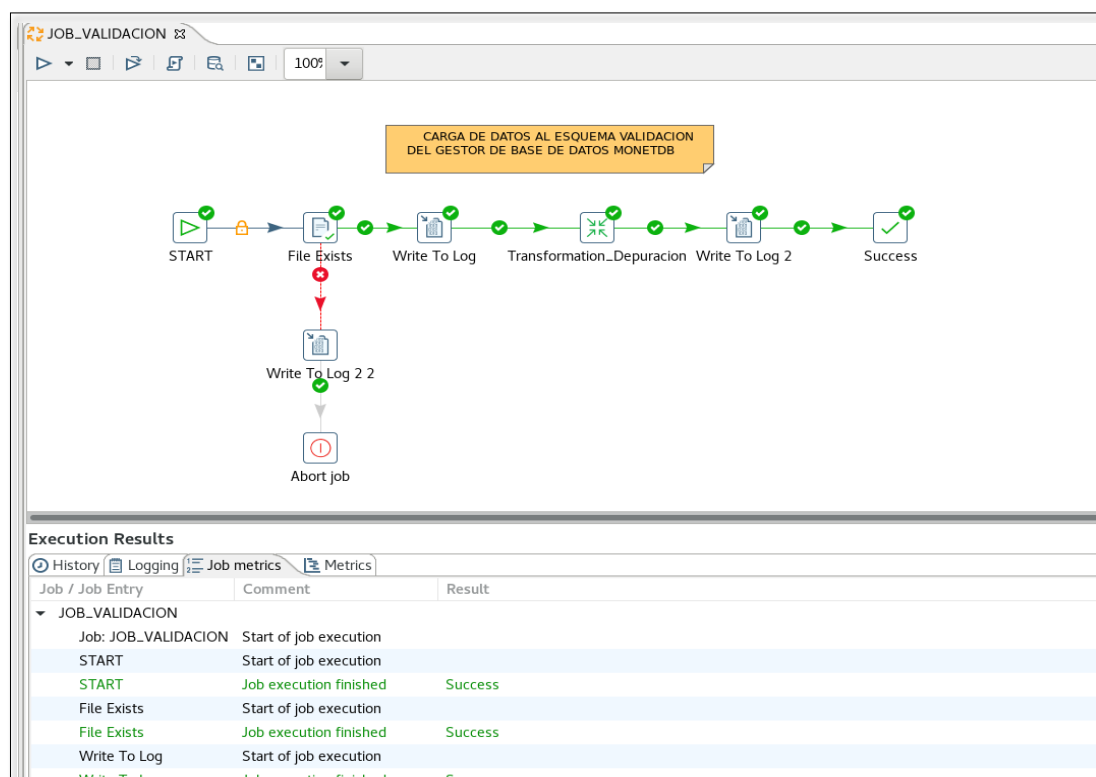


Primer proceso de validación. Este proceso realiza una carga de datos en el esquema establecido para la validación de datos para comprobar si existe el archivo plano y empezar el proceso, adicional se realiza una limpieza de datos para eliminar caracteres especiales, espacios en blanco innecesarios, conversiones, entre otros. Finalmente, se realiza la carga de datos a la tabla de datos destino.

A continuación, se presenta en la Figura 18. la ejecución del proceso JOB_VALIDACION.

Figura 18

Ejecución Proceso Validación.



A continuación, en la Tabla 5. se presenta el tiempo de procesamiento que toma ejecutar el proceso JOB_VALIDACION.

Tabla 5

Prueba de ejecución proceso de validación

Proceso	JOB_VALIDACION
Fuente Origen	Archivo .txt
Número total de registros 440.736	Tiempo ejecución: 9 s

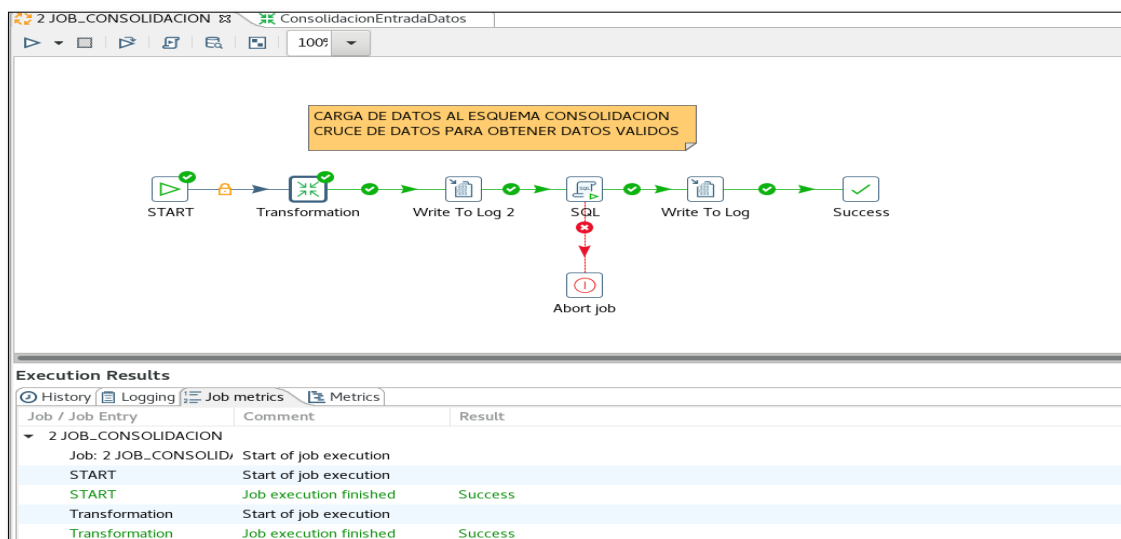
Como se puede observar en la Tabla 5. el tiempo de ejecución para el proceso JOB_VALIDACION es relativamente corto, esto puede deberse el haber utilizado un componente propio de la herramienta Pentaho Data Integration compatible con MonetDB denominado MonetDBBulkLoad que permite la carga de datos por lotes.

Segundo proceso de consolidación. Este proceso realiza una carga de datos a una tabla destino en el esquema definido para la consolidación, incluye sentencias SQL que permite un cruce de datos con otra fuente para validar la información. Al realizar este proceso se observó que es posible eliminar la interpretación de un identificador en MonetDB como palabra clave cuando se coloca comillas dobles en las consultas SQL, es así que un update denota una palabra clave, mientras que "update" expresa un nombre definido por el usuario.

A continuación, se presenta en la Figura 19. la ejecución del proceso JOB_CONSOLIDACION.

Figura 19

Ejecución proceso consolidación



A continuación, se presenta en la Tabla 6. la ejecución del proceso JOB_CONSOLIDACION.

Tabla 6.

Prueba de ejecución proceso de consolidación

PROCESO	JOB_CONSOLIDACION
Fuente Origen	Tabla de datos
Número total de registros 440.736	Tiempo ejecución: 50 min

Como se puede observar, el tiempo de ejecución para el proceso JOB_CONSOLIDACION, es mayor que en el proceso JOB_VALIDACION, teniendo la misma cantidad de registros sin embargo hay que considerar que la fuente de datos es una tabla de datos y se tienen sentencias SQL adicionales para realizar cruces de datos y cargar la data.

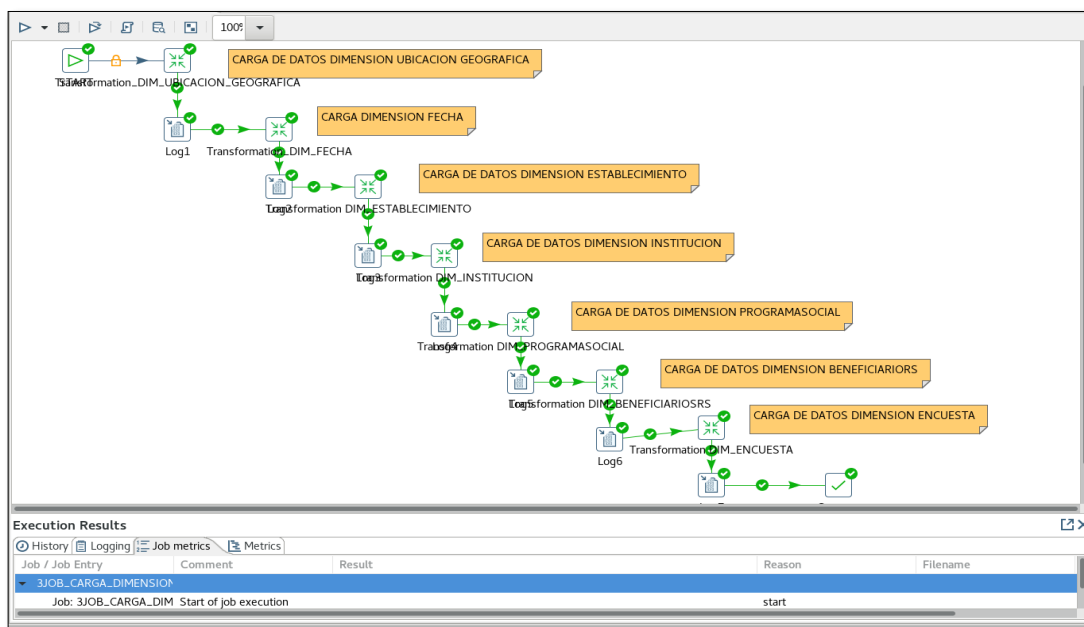
Tercer proceso de integración de datos al Data Warehouse. Este proceso realiza la carga de datos al esquema definido para la integración al data warehouse, se divide en dos procesos: i) Carga de datos a las dimensiones y ii) Carga de datos a la tabla de hechos.

La carga de datos a las dimensiones se realiza desde los catálogos proporcionados, tomados como fuente de origen a las distintas tablas *dimensiones destino* que son DIM_UBICACION_GEOGRAFICA, DIM_FECHA, DIM_INSTITUCION, DIM_ESTABLECIMIENTO, DIM_PROGRAMA_SOCIAL, DIM_BENEFICIARIO, DIM_NIVEL_BIENESTAR, DIM_CATASTO, con la ayuda de los diferentes componentes de la herramienta se prepara la data para cargar de forma consistente a las diferentes dimensiones.

A continuación, se presenta en la Figura 20. la ejecución del proceso JOB_CARGA_DIMENSIONES.

Figura 20

Proceso Carga Dimensiones



A continuación, se presenta en la Tabla 7. la ejecución del proceso JOB_CARGA_DIMENSIONES.

Tabla 7

Prueba de ejecución proceso de carga de dimensiones

PROCESO	JOB_CARGADWH
Fuente Origen	Tablas Catálogos
Proceso: Transformación DIM_UBICACION_GEOGRAFICA	Tiempo ejecución: 5 m 5 s
Proceso: Transformación DIM_FECHA	Tiempo ejecución: 5 m s
Proceso: Transformación DIM_ESTABLECIMIENTO	Tiempo ejecución: 3 s
Proceso: Transformación DIM_INSTITUCION	Tiempo ejecución: 1 s

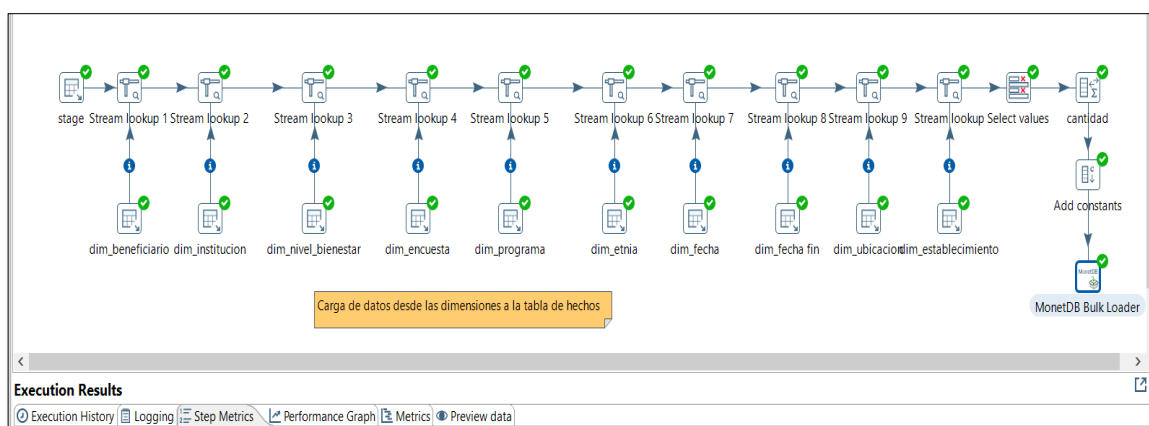
PROCESO	JOB_CARGADWH
Proceso: Transformación DIM_BENEFICIARIO	Tiempo ejecución: 15 m s
Proceso: Transformación DIM_ENCUESTA	Tiempo ejecución: 1 s
Proceso: Transformación DIM_CATASTRO	Tiempo ejecución: 15 m 35s
Número total de registros 440.736	

La carga de datos a la tabla de hechos se realiza desde las diferentes dimensiones DIM_UBICACION_GEOGRAFICA, DIM_FECHA, DIM_INSTITUCION, DIM_ESTABLECIMIENTO, DIM_PROGRAMA_SOCIAL, DIM_BENEFICIARIO, DIM_NIVEL_BIENESTAR, DIM_CATASTO tomadas como origen a la tabla de hechos destino FACT_BENEFICIOS, para lo cual se utilizó diferentes componentes que permiten guardar información de forma adecuada.

A continuación, se presenta en la Figura 21. la ejecución del proceso JOB_CARGA_DWH.

Figura 21

Proceso Carga DWH



A continuación, se presenta en la en la Tabla 8. la ejecución del proceso JOB_CARGA_DWH.

Tabla 8

Prueba de ejecución proceso de carga DWH

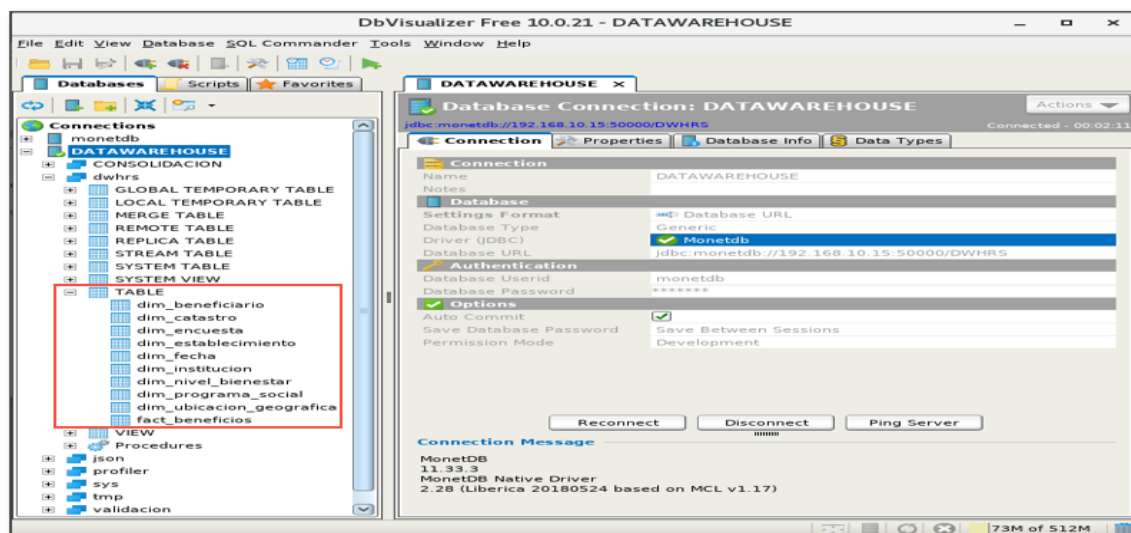
PROCESO	JOB_CARGA_DWH
Fuente Origen	DIM_UBICACION_GEOGRAFICA, DIM_FECHA, DIM_INSTITUCION, DIM_ESTABLECIMIENTO, DIM_PROGRAMA_SOCIAL, DIM_BENEFICIARIO, DIM_NIVEL_BIENESTAR, DIM_CATASTO
Proceso	Tiempo ejecución: 30 m 15 s
Número total de registros	440.736

Como se puede observar, el tiempo de ejecución para el proceso JOB_CARGA_DWH es relevante esto es debido a las diferentes transformaciones que se realiza y las condiciones necesarias para cargar los datos a la tabla de hechos de forma íntegra.

A continuación, en la Figura 22. se presenta la implementación de la base de datos multidimensional.

Figura 22

Implementación Modelo Dimensional



Calendarización de procesos

La calendarización de procesos permite conocer la carga que tendrá un servidor y planificar la ejecución de otros procesos. Algunas herramientas de integración de datos disponen una administración en la cual se puede observar la calendarización de procesos, sin embargo, Pentaho Data Integration no dispone de esta opción por lo cual se recomienda llevar una bitácora que permita llevar un control del procesamiento.

A continuación, se presenta en la Tabla 9. un ejemplo de cómo se podría llevarse un control de la calendarización de los procesos.

Tabla 8

Calendarización Procesos

Código Proceso	Proceso	Tiempo	Ejecución	Servidor	Nota
001	JOB_VALIDACION	9 s	15-05-2019	DWH	Éxito
002	JOB_CONSOLIDACION	50 m	15-05-2019	DWH	Éxito
003	JOB_DIMENSIONES	45 m 46 s	15-05-2019	DWH	Éxito
004	JOB_DWH	30m 15 s	15-05-2019	DWH	Éxito

Fase de Explotación y Producción

Construcción de reportes de datos

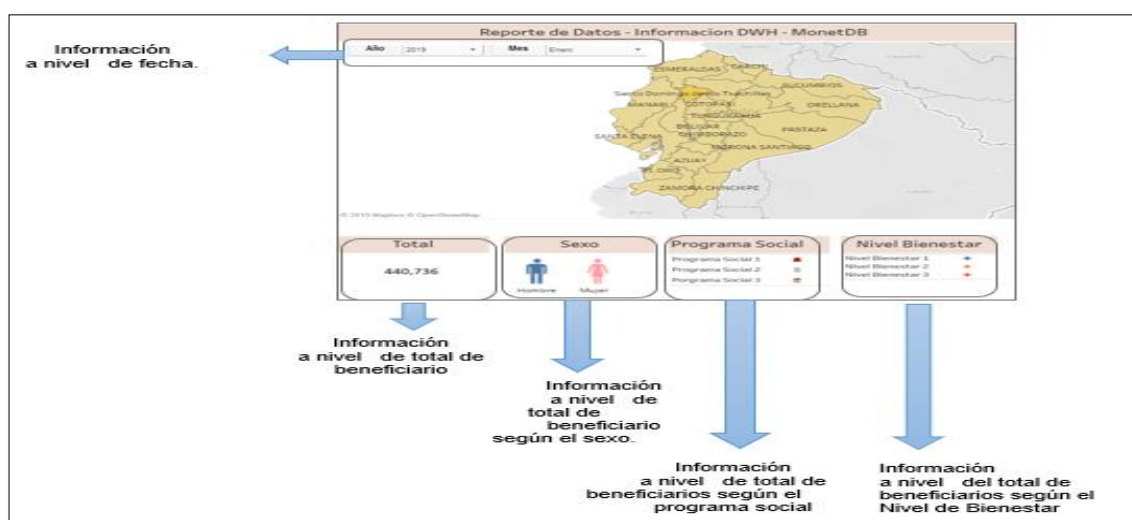
Para el proceso de explotación de los datos es importante el apoyo de una **herramienta de inteligencia de negocio**, por lo cual se utilizó la herramienta Tableau Software para realizar las pruebas, porque esta plataforma es utilizada por la institución actualmente, es fácil de implementar los reportes, y posee conectores para diversos motores de base de datos dentro de los cuales se encuentra para MonetDB.

Para desarrollar los reportes, es importante responder las preguntas realizadas en el análisis de requerimientos a fin de presentar información útil para el usuario final; para lo cual se utiliza componentes de mapa para mostrar la parte de ubicación geográfica, a adicional se incluyen elementos propios de la herramienta para presentar información de los beneficiarios a nivel de programa, nivel de bienestar, sexo y fecha.

A continuación, se presenta la Figura 23. que permite mostrar el reporte de datos.

Figura 23

Reporte de datos



Fase de Mantenimiento y Soporte

Mantenimiento de base de datos y herramientas de software

Las tareas de mantenimiento de base de datos y herramientas de software deben realizarse luego de ser implantada la solución, sin embargo, se recomienda realizar periódicamente un afinamiento de la base de datos a fin de garantizar un buen rendimiento de los procesos que se llevan a cabo y contar con información disponible al usuario final.

Generación de respaldos

La generación de respaldos se debe realizar luego de la implantación, sin embargo, se debe definir el tipo de backup que se debe obtener, estos pueden ser de tipo completo, incremental o diferencial. Además, es importante determinar un sitio externo al servidor de base de datos para ser guardados.

Capítulo V

Evaluación de Resultados

En este capítulo, se presenta la evaluación de los resultados obtenidos para la validación de la solución en función de la problemática descrita anteriormente.

Análisis de Resultados

Se presenta una evaluación de la solución propuesta para el caso de estudio, tomando como referencia los problemas y soluciones establecidos en el capítulo 3.

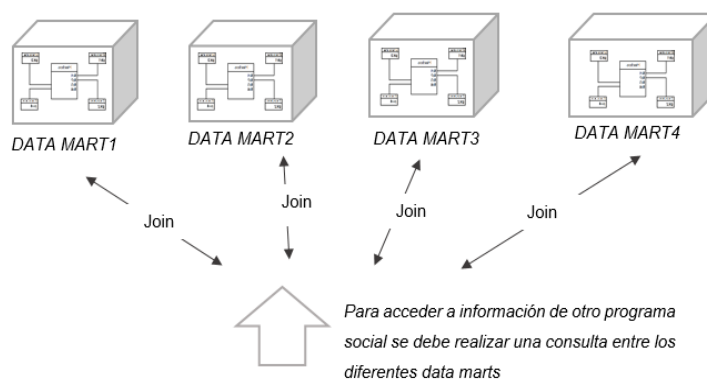
Análisis problema I

La estructura de datos existente en el caso de estudio se basa en la formación de varios data mart, que fueron construidos para tener información más detallada de cada programa social, por lo cual cuando se requiere acceder a información diversa se realizan consultas mediante uniones lo que significa un costo de rendimiento mayor por el tipo de consulta y la cantidad de datos que se tiene.

A continuación, se presenta en la Figura 24. una representación del problema mencionado.

Figura 24

Descripción problema I



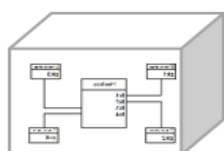
Solución I

Se implementa un modelo dimensional que permita consolidar información mínima de los diferentes programas sociales.

A continuación, se presenta en la Figura 25. una descripción de la solución del problema mencionado.

Figura 25

Descripción solución problema I



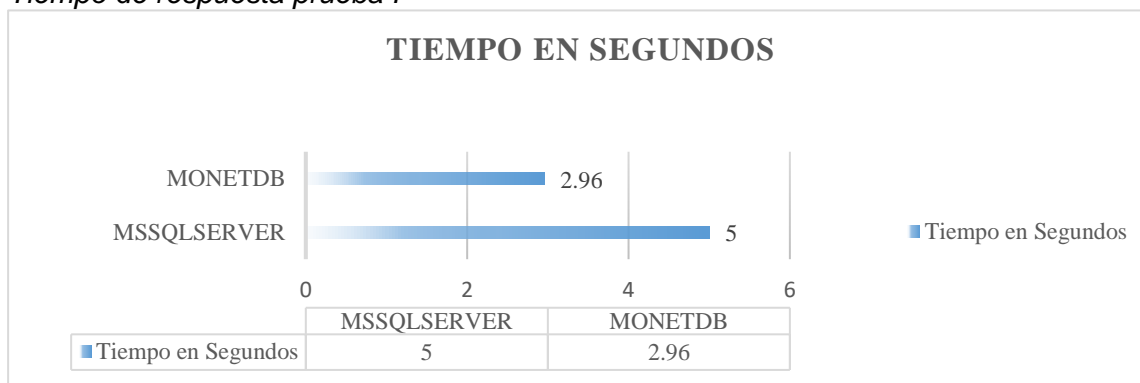
DATA MART

Para acceder a información únicamente se debe realizar una consulta en el data mart sin realizar ningún join

A continuación, se presenta la Figura 26. los tiempos de respuesta de acceso a los datos en el gestor de base de datos transaccional MSSQLSERVER y en el gestor de base de datos con almacenamiento columnar MonetDB al realizar una consulta de tipo Select se obtiene como resultado 440.736 registros.

Figura 26

Tiempo de respuesta prueba I



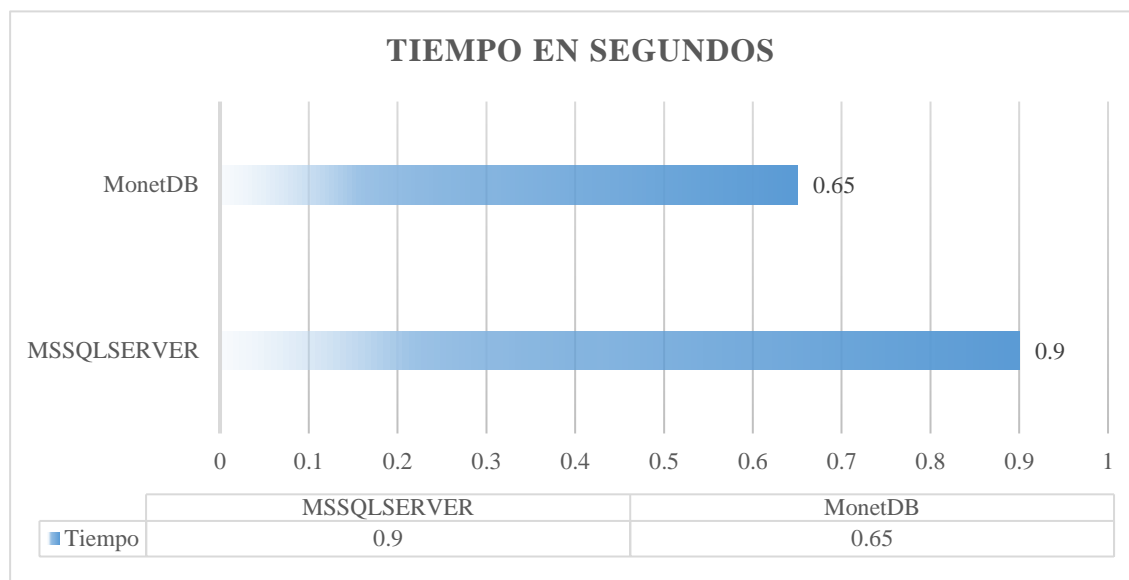
A continuación, se presenta la Figura 27. los tiempos de respuesta desde los motores de base de datos i) MSSQLSERVER y ii) MonetDB según la primera prueba realizada.

Figura 27

Resultados Gestores Base de Datos prueba I



A continuación, se presenta la Figura 28. los tiempos de respuesta de acceso a los datos en el gestor de base de datos transaccional MSSQLSERVER y en el gestor de base de datos con almacenamiento columnar MonetDB, cuando se realiza una consulta con un *INNER JOIN* se obtiene como resultado 440.736 registros.

Figura 28*Tiempo de respuesta prueba II*

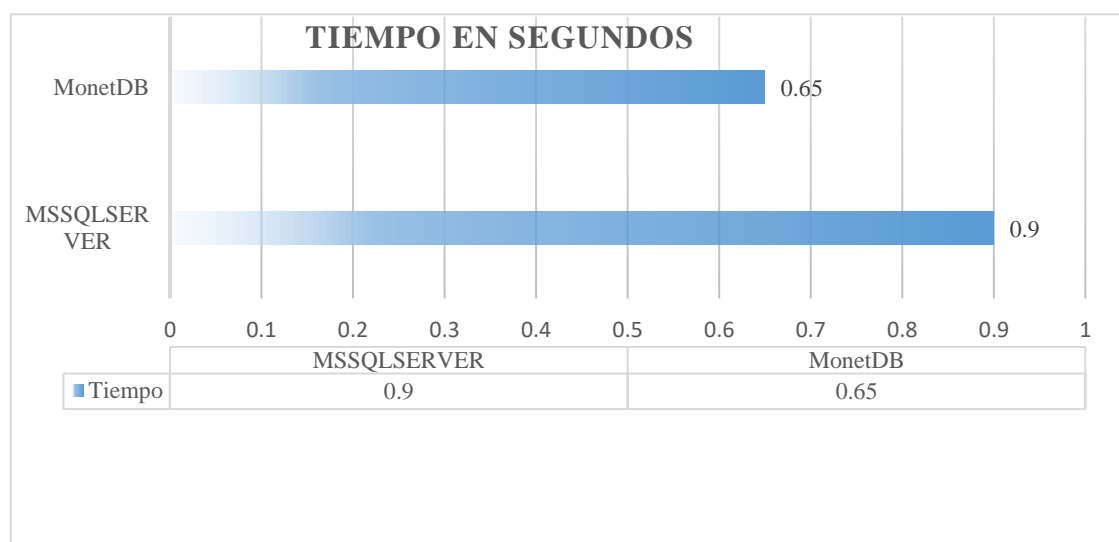
A continuación, se presenta la Figura 29. los tiempos de respuesta desde los motores de base de datos i) MSSQLSERVER y ii) MonetDB según la segunda prueba realizada.

Figura 29*Resultados Gestores Base de Datos prueba II*

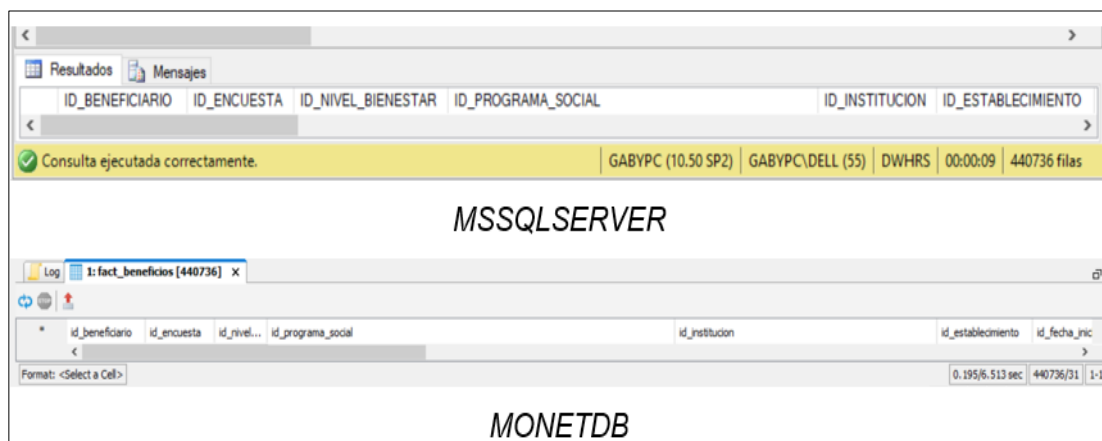
A continuación, se presenta la Figura 30. donde se refleja los tiempos de respuesta de acceso a los datos en el gestor de base de datos transaccional MSSQLSERVER y en el gestor de base de datos con almacenamiento columnar MonetDB, cuando se realiza una consulta con tres *INNER JOIN* con 440.736 registros.

Figura 30

Tiempo de respuesta prueba III



A continuación, se presenta la Figura 31. los tiempos de respuesta desde los motores de base de datos i) MSSQLSERVER y ii) MonetDB según la tercera prueba realizada

Figura 31*Resultados Gestores Base de Datos prueba III*

ID_BENEFICIARIO	ID_ENCUESTA	ID_NIVEL_BIENESTAR	ID_PROGRAMA_SOCIAL	ID_INSTITUCION	ID_ESTABLECIMIENTO
MSSQLSERVER					
Consulta ejecutada correctamente. GABYPC (10.50 SP2) GABYPC\DELL (55) DWHRS 00:00:09 440736 filas					
MONETDB					

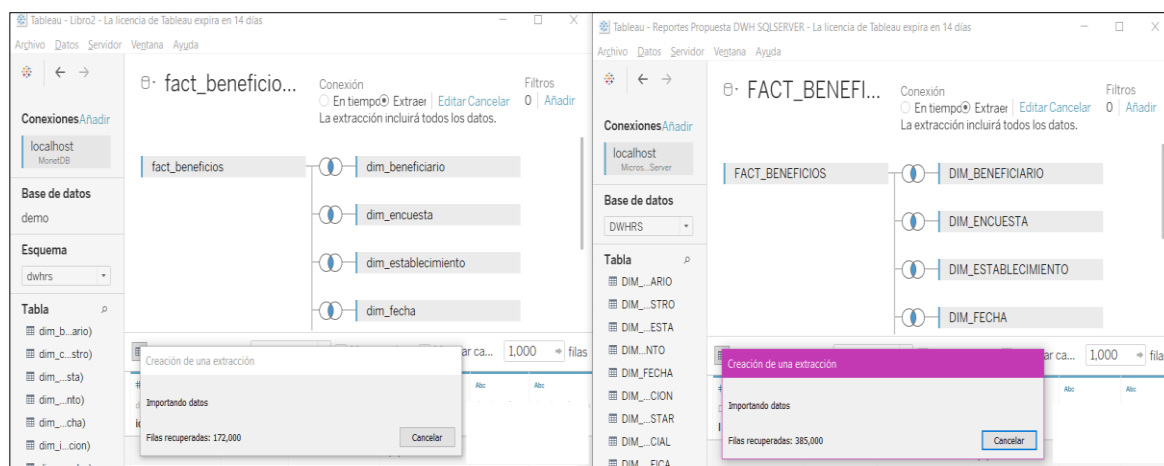
Como conclusión podemos sacar que el gestor de base de datos MonetDB permite un mejor almacenamiento por su arquitectura columnar, además mejoró la respuesta de acceso a los datos, sin embargo, en la última prueba cuando se realiza una consulta con más de un JOIN alcanzando tiempos de respuesta similar que el gestor de base de datos MSSQL SERVER.

Mediante la utilización de la herramienta Tableau Software, se crean los extractos de datos para MSSQL SERVER y MonetDB a fin de comprobar los objetivos del proyecto.

A continuación, se presenta la Figura 32. la creación de los extractos en un gestor de base de datos MSSQL SERVER y MonetDB.

Figura 32

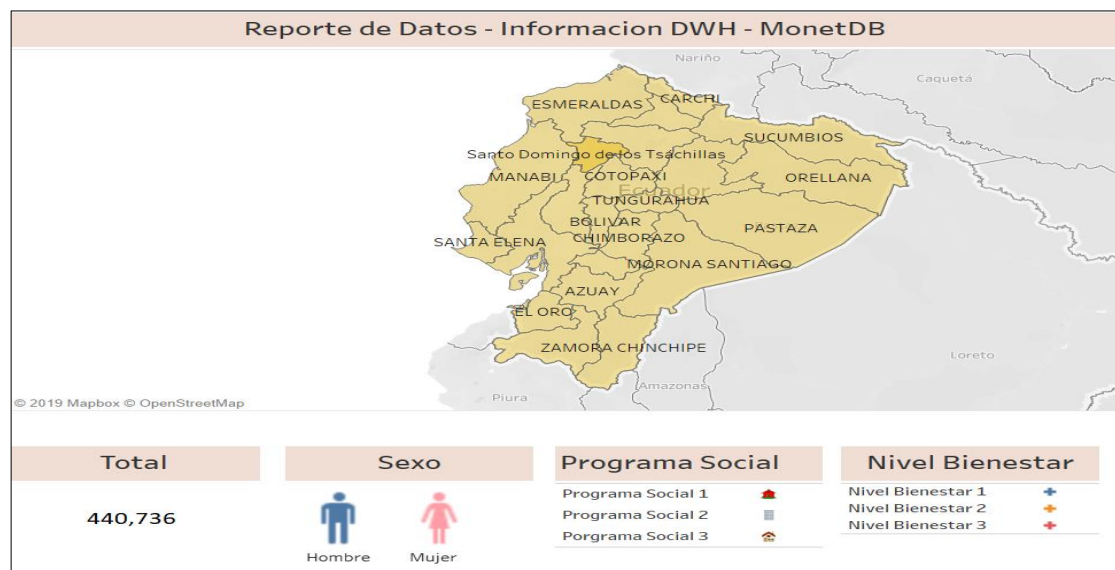
Creación Extractos Reportes de Información Tableau Software



Los tiempos de respuesta para la creación de los extractos de datos son de 2 a 1 considerando un total de registro de 440.736, por lo tanto para el estudio el motor MonetDB es relativamente más rápido con respecto a MSSQL SERVER.

Finalmente, se presenta en la Figura 33. un reporte de la información obtenida del proceso de integración de datos.

Figura 33

Explotación de Información**Análisis problema II**

Investigar, analizar y seleccionar nuevas herramientas open source eficientes para la integración de datos compatibles con un sistema operativo Linux.

Análisis solución II

Según el análisis realizado en la sección de herramientas tecnológicas se definió las características más relevantes, además se logró probarlas cuando se utilizó la herramienta Pentaho Data Integration descubriendo nuevos componentes para realizar cargas de datos masivas permitiendo la conexión con múltiples gestores de base de datos entre los cuales se encuentra MonetDB.

Se escogió la herramienta Tableau Software para el desarrollo del proyecto de inteligencia de negocios, porque es fácil de manipular, posee múltiples conectores para los gestores de bases de datos entre los cuales está MonetDB y actualmente es utilizada en la institución.

Análisis problema III

Desactualización de la solución de inteligencia de negocio para el caso de estudio, actualmente la institución no tiene una metodología para llevar a cabo el proceso de integración de datos.

Análisis solución III

En el capítulo 4 se propone una metodología para llevar a cabo el proceso de integración de datos paso a paso.

Cambios sugeridos para una Implantación

- Contar con un servidor robusto en memoria y disco, ya que los procesos que se realizan de almacenamiento columnar consumen una gran cantidad de recursos.
- Nuevo modelo dimensional para consolidar la información en una base de datos, basados en tres esquemas para su procesamiento validación, consolidación e integración al data warehouse descritos en el capítulo 4.
- Utilizar un servidor con sistema operativo Linux porque permite un procesamiento más rápido y seguro.
- La adopción de nuevas herramientas open source para la integración de datos como son MonetDB y Pentaho Data Integration debido a sus ventajas descritas en el capítulo 3.
- Tomar como base la metodología propuesta para el proceso de integración de datos para realizar futuras implementaciones.

Capítulo VI

Conclusiones y Recomendaciones

En este capítulo, se presentan las conclusiones obtenidas durante el desarrollo de este proyecto, así como las recomendaciones para dar continuidad a futuros estudios.

Conclusiones

Se realizó una revisión de literatura para determinar las características de herramientas de gestión de datos columnar y prácticas más adecuadas, para la implementación de proyectos de inteligencia de negocios que contribuyó en el planteamiento de la propuesta de mejora.

Se implementó, una nueva solución de inteligencia de negocios enfocada a la información de carácter social en el sector gubernamental, con el fin de modernizar la gestión de datos existente en la DRIPS.

Se evaluó, los resultados mediante el uso de herramientas de integración de datos y gestión de datos, con el propósito de comprobar el tiempo de respuesta de acceso a los datos aplicado al modelo multidimensional en un gestor de datos relacional y columnar.

Como resultado del proyecto, se demostró que la implementación de un modelo dimensional en un gestor de base de datos columnar mejoró el tiempo de acceso a los datos en un 28% respecto al gestor de datos transaccional actual, mediante la aplicación de una metodología de inteligencia de negocio.

Con la solución propuesta se intentó disminuir todos los data marts que actualmente conforma el data warehouse del caso de estudio, de tal manera de contar con solo modelo de datos multidimensional que almacene toda la información, y aprovechar las funcionalidades de las herramientas de integración y gestión de datos columnar para mejorar el rendimiento de acceso a los datos.

Recomendaciones

Como trabajo futuro, se podría realizar el mismo estudio, pero utilizando otros gestores de bases de datos columnar open source, para analizar los tiempos de respuesta y comparar los resultados obtenidos en este estudio.

Se recomienda, contar con una guía metodológica para realizar el proceso de integración de datos que se adapte al entorno y reglas de negocio de cada empresa.

Bibliografía

- Husemann, B., Lechtenbörger, J., & Vossen, G. (2000). Conceptual Data Warehouse Design. *DMDW*, 1-11.
- Bernabeau, D. (2009, 7 mayo). *DATAPRIX*. Obtenido de <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/-metodologia-hefesto/51-introduccion>
- Bernabeau, D. (2010). *HEFESTO: Metodología para la Construcción de un Data Warehouse*. Recuperado de https://www.academia.edu/23199027/HEFESTO_DATA_WAREHOUSING_Investigación_y_Sistematización_de_Conceptos_HEFESTO_Metodología_para_la_Construcción_de_un_Data_Warehouse
- Brewer, E. (2012). *How the "Rules" Have Changed*. *IEEE Computer Society*, 23-29.
- D. Abadi, Boncz, P., & Harizopoulos, S. (2009). Column-oriented database systems. *Proceedings of the VLDB Endowment*.
- Devarapalli, S. (2013). *Agile Business Intelligence Development Core Practices*.
- Díaz, J. C. (2010). *Introducción al Business Intelligence*. Barcelona, España: UOC.
- Díaz, J. C. (2014, 4 diciembre). Recuperado de <http://reader.digitalbooks.pro/content/preview/books/43005/book/OEBPS/chapter02.xhtml>
- Gartner Inc. (2019, 1 agosto). *Gartner*. Recuperado de <https://www.gartner.com/en/documents/3955823>
- J. Abadi, Madden, R., & Hachem, N. (Eds.). (2008). *Column-Stores vs. Row-Stores: How Different Are They Really*. ACM SIGMOD. <https://doi.org/10.1145/1376616.1376712>
- Jara, J. (2014). Big Data & Web Intelligence. *Universidad Católica "Nuestra Señora de la Asunción"*. <http://jeuazarru.com/wp-content/uploads/2014/10/BigData.pdf>.

- Lara, J., & Olivares, B. (2016). Propuesta de arquitectura para un módulo de inteligencia de negocios basado en minería de datos. *Research Computing Science*, 47-56.
- Liu, Y., Chen, M., M., & Otros. (2015). A distributed column-oriented database engine for big data analytics. *In Information and Communication Technology*, 289-299.
- Markets Gazette. (2019, 15 octubre). *SWOT Analysis of RDBMS SoftwMarket By Top Players like Cubrid, TmaxSoft, CA Technologies, ThinkSQL, FrontBase, MonetDB, VoltDB and Forecast To 2026*.
- Muntean, M., & Surcel, T. (2013). Agile BI - The future of BI. *Informatica Economica*, 17, 114-124. <https://doi.org/10.12948/issn14531305/17.3.2013.10>
- Muntean, M., & Surcel, T. (2013). Las principales desventajas de los sistemas de BI tradicionales, Septiembre 2013 [Tabla 1]. Recuperado de: *Informatica Economica*.
- Ptiček, M., & Vrdoljak, B. (2017). *Big Data and New Data Warehousing Approaches*, 1-5. United Kingdom, London: Association for Computing Machinery.
- Rivadera, G. (2010). La metodología de Kimball para el diseño de almacenes de almacenes de datos (Data Warehouse). *Universidad Catolica de Salta*.
- Russom, P. (2010). Data Integration for Real-Time Data Warehousing and Data Virtualization. *TDWI*, 1-9.
- Russom, P. (2016). *Data Warehouse Modernization*. TDWI.
- S.G.Yaman. (2012). Introduction to Column-Oriented Databases System. *Columnar Databases*, 6.
- Smith, J., & Rege, M. (2017). *The Data Warehousing (R) Evolution: Where's it headed*, 5. St. Paul, USA: ACM.
- Smith, J., & Rege, M. (2017). *The Data Warehousing (R) Evolution: Where's it headed next?*, 1-5. Lakeland, FL, USA.

- SOLID IT. (2019). *DB-ENGINES RANKING*. Recuperado de <https://db-engines.com/en/system/Cassandra%3BMonetDB>
- SOLID IT. (2019). Comparación Cassandra and MonetDB, Mayo 2019 [Tabla 3].
Recuperado de: <https://db-engines.com/en/system/Cassandra%3BMonetDB>
- Stonebraker, M. (2010). SQL databases vs. NoSQL databases. *Communications of the ACM*, 1-11.
- Telefónica S.A. (2017, 27 de Marzo). *El BI Será Una Tecnología Fundamental en El Crecimiento De Los Mercados*. Blogthinkbig.com. Recuperado de <https://blogthinkbig.com/el-business-intelligence-sera-una-tecnologia-fundamental-en-el-crecimiento-de-los-mercados>
- Thenmozhi, M., & Vivekanandan, K. (2013). A Tool for Data Warehouse Multidimensional Schema Design using Ontology. *International Journal of Computer Science Issue*, 1-8.