

**DESARROLLO DE UNA GUÍA METODOLÓGICA COMO HERRAMIENTA DE APOYO
EN LA ELABORACIÓN DE UNA ARQUITECTURA DE INTELIGENCIA DE NEGOCIOS
EN RELACIÓN AL ÁREA DE SALUD DEL RÉGIMEN SUBSIDIADO Y CONTRIBUTIVO
EN LA ASOCIACIÓN INDÍGENA DEL CAUCA-EPS**



CIELO OTERO CAICEDO

CORPORACIÓN UNIVERSITARIA AUTÓNOMA DE CAUCA

PROGRAMA: INGENIERÍA DE SISTEMAS

POPAYÁN CAUCA, 2022

**DESARROLLO DE UNA GUÍA METODOLÓGICA COMO HERRAMIENTA DE APOYO
EN LA ELABORACIÓN DE UNA ARQUITECTURA DE INTELIGENCIA DE NEGOCIOS
EN RELACIÓN AL ÁREA DE SALUD DEL RÉGIMEN SUBSIDIADO Y CONTRIBUTIVO
EN LA ASOCIACIÓN INDÍGENA DEL CAUCA-EPS**



CIELO OTERO CAICEDO

**Trabajo de grado para optar al título de ingeniera de
Sistemas**

**Director: Carlos Antonio Flórez Arias
Ingeniero de Sistemas**

CORPORACIÓN UNIVERSITARIA AUTÓNOMA DE CAUCA

PROGRAMA: INGENIERÍA DE SISTEMAS

POPAYÁN CAUCA, 2022

Nota de aceptación

El Director y los Jurados han leído el presente documento, escucharon la sustentación del mismo realizada por el autor y lo encuentran satisfactorio.

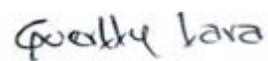


Ing. Carlos Antonio Flórez Arias

Director



Jurado Ing. Leonardo Javier Bastidas Moreno



Jurado Ing.

Popayán Cauca, 2022

DEDICATORIA

Quiero dedicar este trabajo de grado inicialmente a Dios que me acompaña e ilumina mis pasos, a él le doy gracias por darme la vida y proteger a mi familia, a mi madre por el esfuerzo y trabajo constante para ver mis sueños hechos de realidad, mi padre porque me ha inculcado valores especiales en mi formación como persona, a mis hermanos que siempre me han apoyado en mis proyectos de vida, a mi hijo David Santiago Velasco Otero que es la persona que me inspira a luchar y salir a adelante.

Gracias a mis amigas(o) y demás persona, la vida me ha premiado con personas que han sido un pilar fundamental para impulsarme a ser mejor cada día.

AGRADECIMIENTOS

Este trabajo de grado modalidad pasantía es fruto del aporte y apoyo de personas que desde sus conocimientos prácticos me aportaron experiencias, el cual me sirvieron para enriquecer el trabajo desarrollado y hacer esto posible.

Mis más sinceros agradecimientos al Ingeniero Carlos Antonio Flórez Arias, por su experiencia, paciencia, apoyo en la elaboración del trabajo de grado y formación académica.

Agradezco a la Corporación Universitaria Autónoma del Cauca y en su nombre los docentes que aportaron en el desarrollo y fortalecimiento de mis conocimientos. De igual manera a la empresa Asociación Indígena de Cauca EPS, especialmente al representante legal Oscar Elciario Bonilla Rivera y al grupo de trabajo proceso Gestión TIC por brindarme el espacio y acogimiento de tan nombrada y reconocida organización, para llevar a cabo el desarrollo del trabajo de grado.

Gracias a mi familia, compañeros por su cariño y confianza y apoyo incondicional. De igual manera agradezco a las personas que de una u otra forma aportaron para la culminación de este trabajo.

TABLA DE CONTENIDO

Contenido

RESUMEN	12
INTRODUCCIÓN	14
1.1.Planteamiento del Problema.....	1
1.2. Justificación.....	2
1.3Objetivo General.....	4
1.3.1 Objetivos Específicos	4
CAPITULO II. MARCO TEORICO O CONCEPTUAL	4
2.1.1 Sistemas de Información	4
2.1.2 Sistemas de Información	5
2.1.3 Tipos de sistemas de información.....	6
2.1.4 Fuentes de Información	8
2.1.5 Administración de Datos.....	8
2.1.6 Administración del conocimiento.....	9
2.1.7 ¿Que es un dato?	9
2.1.8 Historia sobre la Inteligencia de Negocio	9
2.1.9 Beneficios de la Inteligencia de Negocio.....	11
2.1.10 Importancia de la Inteligencia de Negocios.....	13
2.1.11 Componentes de una Arquitectura de Inteligencia de Negocios	13
2.1.12 ¿Qué es un Datawarehouse o Almacén de Datos?	16
2.1.13 Componentes de un Data Warehouse o Almacen de Datos	16
2.1.14 ¿Que es un DataMart?	17
2.1.15 Tipos de DataMart	17
2.1.16 Concepto Big Data.....	17
2.1.17 Minería de Datos	18
2.1.18 Concepto de Datos Multidimensionales (Cubos OLAP)	19
2.1.19 ¿Que es un sistema OLAP (On-Line Analytical Processing)?	20
2.1.20 Diferencia entre un sistema OLTP y OLAP	20
2.1.21 Herramientas Para El Desarrollo De Inteligencia de Negocios De Software Libre	21

2.1.22 CloverETL.....	21
2.1.23 CloverETL Engine.....	23
2.1.24 CloverETL Server	23
2.1.25 Pentaho Business Intelligence Suite	24
2.1.26 Pentaho Reporting.....	25
2.1.27 Pentaho Analysis	26
2.1.28 Pentaho Dashboards	27
2.1.29 Pentaho Data Integration	28
2.1.30 Pentaho Data Mining	30
2.1.31 Tableau	31
2.1.32 Herramientas Para El Desarrollo De Inteligencia de Negocios de Software Con Licencia	34
2.1.33 Oracle Business Intelligence Suite Enterprise Edition plus (EE)	34
2.1.34 SAP BI.....	35
2.1.36 TIBCO Spotfire	37
2.1.37 Sísense	38
2.1.38 SAS Business Inteligencia	39
2.1.39 Hyperion Interactive Reporting.....	40
2.1.40 Hyperion Financial Reporting.....	40
2.1.41 Metodologías Para La Construcción De Almacén De Datos	40
2.1.42 Metodología Ralph Kimball	40
2.1.43 Metodología Bill Inmon	47
2.1.44 Metodología Hefesto.....	51
2.2 ANTECEDENTES	53
2.2.1 Applications Guide Hdm-4 Volume Two (Guía De Aplicaciones)	53
2.2.2 Gmm-Bi: Una Guía Metodológica Para La Madurez Organizacional En Inteligencia De Negocios	54
2.2.3 Diseño De Modelo Tecnológico Para El Desarrollo Y El Uso De Big Data En El Análisis Y Visualización De Información Para Pequeñas Y Medianas Empresas.....	54
2.2.4 Una Metodológica De Toma De Decisiones Basadas En Inteligencia Empresarial Y Orientada A Objetivos.	55
2.2.5 Marco De Gestión Del Conocimiento Utilizando Arquitectura Empresarial E Inteligencia Empresarial.	55
2.2.6 Business Intelligence (Una Guía Práctica):.....	55
2.2.7 Diez Pasos Para Incrementar El Conocimiento Para Una Gestión Eficiente Del Capital Intelectual En La Inteligencia Empresarial.	56
2.2.8 Una Inteligencia De Negocios En Cinco Capas (Arquitectura).....	56

CAPITULO III. METODOLOGIA	57
3.3.1 Caracterización de la organización	57
3.3.2 Metodologías de Desarrollo de Almacén de Datos	60
3.3.3 Descripción de las etapas de la Guía Metodológica.....	62
3.3.3.1.1 Etapa 1: Planificación del Proyecto.....	64
3.3.3.1.1.1 Definición del Proyecto	64
3.3.3.1.1.2 Definición de un Plan de Trabajo	65
3.3.3.1.2 Etapa 2: Recolección de Requerimientos	65
3.3.3.1.2.1 Técnica de Entrevista Para la Recolección de Requerimientos.....	66
3.3.3.1.2.2 Identificación de las Fuentes de Información	68
3.3.3.1.2.3 Selección de la herramienta para proceso ETL.....	68
3.3.3.1.2.4 Construcción de Preguntas para Plantear el Modelo Dimensional.....	68
3.3.3.1.2.5 Matriz de Requerimientos	69
3.3.3.1.3 Etapa 3: Diseñar El Modelo Dimensional.....	69
3.3.3.1.3.1 Identificar el proceso de negocio.....	69
3.3.3.1.3.2 Definir las Tablas de Hechos	70
3.3.3.1.3.3 Definir las Tablas de Dimensiones.....	72
3.3.3.1.3.4 Diseñar el Modelo Lógico.....	73
3.3.3.1.3.5 Diseñar el Modelo Físico.....	73
3.3.3.1.4 Etapa 4: Diseñar el proceso E.T.L	74
3.3.3.1.4.1 Definición de Pruebas.....	76
3.3.3.1.5. Etapa 5: Implementar El Almacén De Datos	76
3.3.3.1.5.1 Documentación Final	77
3.2 Aplicación de la Guía en la empresa Asociación Indígena del Cauca - EPS.....	78
CAPITULO IV. RESULTADOS	113
CAPÍTULO V CONCLUSIONES Y RECOMENDACIONES	116
5.1 CONCLUSIONES.....	116
5.2 RECOMENDACIONES.....	116
6. GLOSARIO	117
BIBLIOGRAFIA	119
ANEXOS	125

LISTA DE TABLAS

Tabla 1: Costo de las Herramientas por Equipos y Organizaciones	33
Tabla 2: Tabla comparativa de metodologías para el desarrollo de un DW.....	61
Tabla 3: Presupuesto Del Proyecto Construcción Del Datamart.....	82
Tabla 4: Estructura del Archivo S1, MS Y NS	93
Tabla 5: Tabla de Indicadores y Perceptivas.....	95
Tabla 6: Tabla Definición de Requerimientos.....	96
Tabla 7: Tabla Análisis de Dimensiones	96
Tabla 8: Descripción de Campos y formato Dimensión Dinamizador	100
Tabla 9: Descripción de Campos y formato Dimensión Tiempo	102
Tabla 10 : Descripción de campo y atributos Dimensión Novedad.....	103

LISTA DE FIGURAS

Fig.1 Componentes de un sistema de información tomado de [15]	5
Fig.2. Niveles de los sistemas de información en la estructura organizacional empresarial. Tomado de[16].	6
Fig.3. Arquitectura de Inteligencia de Negocios. Tomado de [23]	14
Fig. 4. Modelo E.T.L. Tomado de [26]......	15
Fig.5. Elementos de Una Data Warehouse. Tomado de [27]......	16
Fig.6. Metodología KDD. Tomado de[30]	18
Fig.7. Cubo Multidimensional (OLAP). Tomado de [31]	19
Fig.8. Sistema OLTP Y OLAP. Tomado de[33].	21
Fig.9. Arquitectura de la Plataforma Pentaho Open BI Suite. Tomado de [34]	25
Fig.10. Interfaz Data Integración (Spoon). Tomado de [34]......	30
Fig.11: Elementos de la Plataforma Oracle BI Suite Enterprise Edition Plus. Tomado de [35]....	35
Fig.12: Método de Ralph Kimball. Tomado de [23]......	41
Fig.13. Gerenciamiento del Proyecto. Tomado de [27]	47
Fig.14. Fases de la Metodología Inmon. Tomado de[23]......	48
Fig.15. Componentes de la metodología Hefesto	52
Fig.16. Clasificación de las empresas de acuerdo con nivel del personal	58
Fig.17. Etapas de una Arquitectura de Inteligencia de Negocios	63
Fig. 18. Esquema en estrella. Tomado de [45].....	72
Fig.19 Esquema copo de nieve. Tomado de [45]......	72
Fig.20 Estructura Orgánica de la Asociación Indígena del Cauca – EPSI.....	80
Fig. 21 Flujograma Proceso novedad S1.....	84
Fig. 22 Flujograma Proceso novedad MS Y NS	85
Fig.23 Seguimiento de Actividades Proyecto Datamart	87
Fig.24 Modelo Dimensional esquema en estrella.....	98
Fig.25 ETL Carga de Archivo Novedad MS Reportado.....	104
Fig.26 ETL Carga de Archivo Novedad MS Validado.....	104
Fig.27 ETL Carga de Archivo Novedad S1 Reportado.....	105
Fig.28 ETL Carga de Archivo Novedad S1 Validado	105
Fig.39 ETL Carga de Archivo Novedad S1 Negado	105
Fig.30 ETL Carga de Archivo Novedad NS Reportado	106
Fig.31 ETL Carga de Archivo Novedad NS Validado	106
Fig.32 ETL Carga de Archivo Novedad NS Negado.....	106
Fig.33 ETL Cargue Dimensión Dinamizador.....	107
Fig.34 ETL Cargue de Dimensión Novedad.....	108
Fig.35 Códigos y Descripción de las Novedades	108

Fig.36 ETL Cargue Dimensión Tiempo.....	109
Fig.37 ETL Cargue Tabla de Hechos Reportes BDUA.....	110
Fig.38 Carga de Datos Archivo Novedad MS	110
Fig.39 Carga de Datos Archivo Novedad MS Validado.....	110
Fig.40 Carga de Datos Archivo Novedad MS Negado	111
Fig.41 Carga de Datos Archivo Novedad NS Reportado	111
Fig.42 Carga de Datos Archivo Novedad NS Validado	111
Fig.43 Carga de Datos Archivo Novedad NS Negado.....	111
Fig.44 Carga de Datos Archivo Novedad S1 Reportado.....	112
Fig.45 Carga de Datos Archivo Novedad S1 Validado.....	112
Fig.46 Carga de Datos Archivo Novedad S1 Negado	112
Fig.47 Carga de Datos Dimensión Dinamizador	112
Fig.48 Carga de Datos Dimensión Novedad.....	113
Fig.49 Carga de Datos Dimensión Tiempo	113
Fig.50 Carga de Datos Tabla de Hechos Novedades BDUA	113
Fig.51 Cantidad de Novedades Reportadas Semana dos enero 2021	114
Fig.52 Cantidad de Novedades Reportadas Mes de Enero 2021	114
Fig.53 Cantidad de Novedades Reportadas por Dinamizador mes Enero.....	115
Fig.54 Porcentaje de Cargue por Dinamizador.....	116

RESUMEN

El cumplimiento del presente trabajo se realizó bajo la modalidad de Pasantía en el área de Planeación subproceso Gestión Tic de la empresa Asociación Indígena del Cauca EPS, sede Popayán- Cauca, bajo el acompañamiento del personal encargados del proceso

La metodología guio el trabajo bajo un enfoque mixto, debido a que se realizó un proceso de recolección, análisis, agrupación de datos e información que contribuyó a conceptualizar detalladamente la información. Por tanto, se realizó bajo el cumplimiento de objetivos el cual se propuso el desarrollo de una guía metodológica como herramienta de apoyo en la elaboración de una arquitectura de Inteligencia de Negocios en relación al área de salud del régimen subsidiado y contributivo. La cual parte de los conceptos, técnicas, componentes, y métodos que se deben abordar para un proyecto de Inteligencia de Negocios. Con la guía se logra apoyar al subproceso de Gestión Tic, donde actualmente se realiza la propuesta para el desarrollo de un módulo BI.

Palabras Claves: Inteligencia Negocios, Parámetros, metodológica, subsidiado, contributivo, guía.

ABSTRACT

The current Project was developed as an internship for the company “Asociación Indígena del Cauca EPS” in the Popayán headquarters, working in the subprocess planning area of the Information Technology and Communications (ICT) management along with the staff that belongs to the company.

The approach for this project was divided in multiple tasks for a single process of recollection, analysis, and data clustering to conceptualize the information. Therefore, this process was implemented under a goal accomplishment focus to develop a methodology guide as a support tool in the design of a Business Intelligence Architecture for the subsidized and contributory system of the health area. Which uses the concepts, techniques, components, and methods that must be applied on a Business Intelligence project. With this guide is possible to support the subprocess planning area of ICT management where this proposal is developed.

Key words: Business Intelligence, Parameters, methodological, subsidized, contributory, guide.

INTRODUCCIÓN

En la actualidad, la forma cómo se gestiona la información es cada vez más importante y genera ventajas competitivas a nivel empresarial. De igual forma con la evolución de las tecnologías se han desarrollado sistemas que ayudan a gestionar la información. creando estrategias y herramientas dirigidas al tratamiento, análisis y administración de datos en las organizaciones dentro de estos sistemas se encuentra la Inteligencia de Negocios

Ahora bien, el objetivo de este documento es proporcionar a la empresa Asociación Indígena del Cauca-EPS una guía metodológica como herramienta de apoyo en la elaboración de una arquitectura BI en relación al área de salud del régimen subsidiado y contributivo en esta se establecen los principales lineamientos de una arquitectura como conceptos, métodos y técnicas. De acuerdo a esto, es importante mencionar que para el desarrollo de la guía se debe llevar a cabo las actividades necesarias para el cumplimiento de objetivos.

Para la organización es importante estar actualizada en cuanto a la implementación de nuevas tecnologías y herramientas que permitan descubrir nuevas tendencias en cuanto al manejo de la información. Por tal razón esta guía servirá como herramienta de apoyo al personal de Gestión TIC para abordar un proyecto de Inteligencia de Negocios

CAPITULO I. PROBLEMA

1.1. Planteamiento del Problema

Un problema muy común hoy en día en las organizaciones, es la falta de análisis de datos que se encuentran almacenados en los sistemas de información empresariales[1]. Esto se debe a que algunas entidades no implementan tecnologías y herramientas que permiten explorar y analizar estos datos[2].

Ahora bien, en el Municipio de Popayán departamento del Cauca se encuentra ubicada la AIC (Asociación Indígena del Cauca - EPS) entidad de carácter especial que administra recursos públicos provenientes del sector salud[3]; Por tanto, la empresa cuenta con un gran volumen de información como son: datos de usuarios, archivos de movilidad, novedades de afiliaciones, malla de validaciones y diferentes procesos que se realizan por cada persona que se encuentra vinculada al régimen subsidiado y contributivo[4]. Por tal razón la empresa desea explorar y analizar los datos que se encuentran almacenados, con el fin de encontrar nueva información relevante, con la que se pueda tomar decisiones estratégicas que le permita estar a la vanguardia del mercado cambiante[5].

Es importante mencionar que la empresa viene desarrollando un sistema de información que contiene varios módulos donde el subproceso de Gerencia de la información que pertenece a Gestión TIC tiene como objetivo el desarrollo del módulo BI (Inteligencia de Negocios)[6], por tanto, el área requiere fortalecer los procesos que se abordan en la construcción de este. Actualmente en la organización no se han establecido parámetros que definan los lineamientos específicos para la elaboración del módulo.

Por lo expuesto anteriormente se percibe la importancia de llevar a cabo en la AIC el desarrollo de una guía metodológica que sirva como herramienta de apoyo donde se establezcan los principales conceptos, procesos, métodos, técnicas y herramientas que se requieren para el desarrollo e implementación de una arquitectura de Inteligencia de Negocios.

De acuerdo con el planteamiento anterior se puede establecer como pregunta al problema lo siguiente:

¿Cómo desarrollar una guía metodológica que sirva como herramienta de apoyo en la elaboración e implementación de una Arquitectura de Inteligencia de Negocios en la Asociación Indígena del Cauca – EPS?

1.2. Justificación

A lo largo de los años la INFORMACIÓN ha sido vital para la superación de muchos obstáculos, y el crecimiento de las herramientas tecnológicas digitales han ayudado a que los procesos de la información sean cada vez más veloces y avanzados. ¿Pero cómo manejar esa información digital a nuestro beneficio?

La AIC EPS (Asociación Indígena del Cauca) acumula una gran cantidad de información, la cual debe ser llevada de una manera adecuada, para facilitar la toma de decisiones en la organización. Por consiguiente, en la organización se creó el subproceso de Gerencia de la Información que se encarga del hacer seguimiento de la calidad del dato, dado que la empresa comprende la importancia de gestionar la información[7], por tanto, la AIC mira el valor de implementar herramientas tecnológicas que permiten hacer análisis a los datos almacenados en los diferentes sistemas información de la organización, esto con

el fin encontrar información relevante para la buena toma de decisiones, dado que la información es un factor de valor incalculable para el éxito empresarial[8]. En este ámbito las tecnologías de información juegan un papel muy importante dado que existen herramientas que permiten diagnosticar escenarios en los que se pueda obtener información activa y relevante[9].

Por tanto, la AIC requiere implementar esas técnicas en las que pueda identificar información y aplicar estrategias que conlleven a cumplir con la visión de la organización de acuerdo a esto es importante que la organización cuente con información de calidad[10].

Dentro de este contexto, la intención de este proyecto es proporcionar una guía metodológica que sirva como referencia de apoyo para la elaboración e implementación de una Arquitectura de Inteligencia de Negocios en la Asociación Indígena del Cauca EPS, que va desde las técnicas necesarias para la extracción, transformación, carga de datos, hasta la selección de herramientas de análisis y presentación de la información que se deben aplicar para la implementación de un proyecto BI, teniendo en cuenta las políticas de datos que se maneja dentro de la organización. Cabe mencionar, que es importante guiarse en todos los procedimientos de cada caso. Por lo que es necesario tener un conocimiento fundamentado y descrito para la elaboración de proyectos[11].

Este proyecto beneficiará al área de Gestión de la Información TIC en la AIC, donde se llevará a cabo el desarrollo del módulo BI, por lo que la organización contará con una guía metodológica como herramienta de apoyo para el desarrollo del módulo.

La implementación de la Arquitectura de negocios inteligentes en la Asociación Indígena de Cauca EPS permitirá a los altos directivos de la organización tomar mejores decisiones y estrategias preventivas a su beneficio[12].

1.3 Objetivo General

Construir una guía metodológica como herramienta de apoyo en la elaboración de una arquitectura BI en relación al área de salud del régimen subsidiado y contributivo en la Asociación Indígena del Cauca- EPS.

1.3.1 Objetivos Específicos

- Analizar los componentes de una arquitectura de Inteligencia de negocios para la elaboración de una guía metodológica.
- Construir la guía basada en los componentes propuestos para la arquitectura de Inteligencia de Negocios.
- Implementar la guía metodológica en la construcción del módulo BI de la Asociación indígena del Cauca-EPS.

CAPITULO II. MARCO TEORICO O CONCEPTUAL

2.1.1 Sistemas de Información

Para [13]“la información es el significado que otorgan las personas a las cosas. Los datos se perciben mediante los sentidos, estos los integran y generan la información necesaria para el conocimiento quien permite tomar decisiones para realizar las acciones cotidianas [...]”. De lo expuesto se deduce que la información está compuesta por datos los cuales describen hechos, sucesos y entidades.

2.1.2 Sistemas de Información

El concepto de sistema de información según lo expresa[14] como “un conjunto de componentes interrelacionados que recaban procesa, almacenan, y distribuyen datos e información y proporcionan un mecanismo de retroalimentación para cumplir con un objetivo”. Según el planteamiento del autor en el párrafo anterior los SI (Sistemas de Información) contienen procesos que son ejecutados con el fin de dar respuesta a solicitudes, las cuales generan información útil para el usuario. En la figura 1 se ilustra los componentes de un sistema de información.



Fig.1 Componentes de un sistema de información tomado de [15]

Entrada: En la entrada es donde se recopilan la captura de datos que el sistema necesita para luego ser procesado, estos pueden ser manuales o automáticas.

Procesamiento: En esta etapa los datos son procesados es decir se le hace una transformación en el cual se crean los controles, almacenamientos y alternativas para la salida.

Salida: Es donde se muestra la información útil para el usuario. Dado que esta es visible en reportes, documentos entre otros.

Retroalimentación: En la parte de la retroalimentación se analiza y se toma en cuenta la información de acciones que se han ejecutado anteriormente. También se puede corregir los errores cometidos durante el proceso y la revisión del funcionamiento del sistema.

2.1.3 Tipos de sistemas de información

En la estructura organizacional como menciona [16]“los SIG se localizan en un nivel administrativo, en el que se relaciona con funciones de seguimiento, control, toma de decisiones y administración de recursos”. De acuerdo con el término que plantea el autor en la estructura se organizan los diferentes procesos de las áreas o departamentos a nivel empresarial como se observa en la siguiente figura.



Fig.2. Niveles de los sistemas de información en la estructura organizacional empresarial. Tomado de[16].

- **Sistemas de procesamiento de transacciones (TPS):** Un sistema de procesamiento de transacciones (TPS por sus siglas en inglés) se puede definir como un conjunto de actividades ejecutadas las cuales recolectan, almacenan, modifican y recuperan las acciones que son generadas por los procesos de la empresa.
- **Sistema de gestión del conocimiento (KWS):** la gestión del conocimiento se puede definir como el conjunto de actividades que se desarrollan al interior de una organización.
- **Sistemas de información de gestión (MIS):** la gestión de la información es ofrecer mecanismo que permitirán a la organización adquirir, producir y transmitir al menor tiempo posible datos e informaciones con una calidad y exactitud para cumplir con los objetivos empresariales.
- **Sistemas de soportes de decisiones (DSS):** Los sistemas de apoyo ayudan a la toma de decisiones gerenciales al combinar datos con sofisticados modelos analíticos en los procesos de toma de decisiones estructuradas y semiestructuradas.

Existen diferentes tipos de sistemas de soportes de decisiones como:

- Ciencias de la Administración (MS)
- Planificación de Recursos Empresariales (ERP)
- Análisis de Negocio
- Gestión de Recurso del Cliente (CRM)
- Sistemas Expertos (ES)
- Inteligencia de Negocios (BI)

- Gestión de recursos Empresariales (ERM)
- **Sistema de Información Ejecutiva (ESS):** Los sistemas ESS se basa en ciertos requerimientos estratégicos de la compañía, los altos directivos deben utilizar tecnologías apropiadas que brinden soluciones óptimas para la toma de decisiones, esto es con el fin de evitar alteraciones adversas a futuro.

2.1.4 Fuentes de Información

Una fuente de información es el medio por el cual procede la información. Esta puede provenir de diferentes sitios por ejemplo en las organizaciones se puede obtener de las bases de datos, archivos planos, archivos de texto o sistemas de información entre otros. Cabe mencionar que la información puede provenir de partes internas o externas de la organización.

- **Fuentes de información primarias:** Son aquellas que provienen de investigaciones. Esta información puede resultar de aplicar un método investigativo a personas o grupo de poblaciones. El cual puede ser por medio de entrevistas, libros, fotografías, discursos, noticias etc. Este tipo de información son primicias dado que no han sido expuestas, filtradas ni publicadas por primera vez.
- **Fuentes secundarias de información:** Estas fuentes describen la información primaria que se sintetiza para formar nuevamente un conjunto de información. Estas son diseñadas para restringir el acceso a fuentes primarias.

2.1.5 Administración de Datos

Desde mi punto de vista la administración de datos se encarga de recopilar, analizar y verificar el estado de estos. De igual forma se puede agregar que:

La administración de datos es la práctica de recopilar, mantener y utilizar datos de forma segura, eficiente y rentable. El objetivo de la administración de datos es ayudar a las personas, organizaciones y las cuestiones relacionadas a optimizar el uso de datos de datos dentro de los límites de las políticas y regulaciones para que puedan tomar decisiones y tomar medidas que maximicen los beneficios para la organización[17]

2.1.6 Administración del conocimiento

Este término lo define[18] como “el proceso sistemático de encontrar, seleccionar, organizar, destilar y preservar la información que mejore la comprensión de un empleado, de un grupo o de una organización completa en una específica de interés “. Según lo anterior la administración del conocimiento busca crear, analizar y compartir un conjunto de enfoques multidisciplinarios para el alcance de objetivos empresariales implementando las mejores técnicas.

2.1.7 ¿Que es un dato?

En mi criterio un dato puede representar un hecho, un suceso o una variable. Por tanto, si el dato se une con otros, estos generan información de algún evento.

2.1.8 Historia sobre la Inteligencia de Negocio

El termino BI (Business Intelligence - Inteligencia de Negocios) se dio a conocer por el investigador de IBM Hans Peter Luhn que en su artículo “A Business Intelligence System”

publicado en el año 1958 lo describe [19] “es la habilidad de aprender las relaciones de hechos presentados de forma que guíen las acciones hacia una meta deseada”.

Con este término se abre la puerta para las grandes investigaciones acerca del uso e implementación de una de las tecnologías más importantes para el apoyo en el proceso de toma de decisiones en grandes y pequeñas organizaciones.

Cuatro años más tarde, en 1962 el informático y matemático canadiense Kenneth Vieron aporta un avance importante acerca del futuro de la inteligencia de negocios inventando el primer lenguaje de programación multidimensional llamado OLAP que según [20] “se entiende por OLAP, o proceso analítico en línea, al método ágil y flexible para organizar datos, especialmente metadatos sobre un objetos o jerarquía de objetos como un sistema u organización multidimensional [...]” en otros términos este proceso pretende agilizar la consulta de grandes volúmenes de datos.

Pero fue hasta el año 1969, que un científico informático inglés Edgar Frank Codd crea el modelo relacional de datos, esto revolucionó la forma en que se almacena y se recuperan los datos.

Las primeras bases de datos creadas en los años 70'. Donde estas permitían el acceso a la información que se encontraba almacenada pero no contaban con buena eficacia y la manera en la que estaban organizadas no era tan útil, por tanto, se hacía difícil el acceso a la información específica.

En la siguiente década a mediados del año 1980 se crea el concepto Datawarehouse donde aparecen los sistemas de reporting los cuales no eran lo suficientemente funcionales, a pesar de que se contaba con buenas bases de datos, por tanto, no existía

aplicaciones que pudieran facilitar el uso y la explotación de estas mismas. Para entender el termino reporting el cual lo define según[21]“permiten obtener de los informes dinámicos los aspectos más críticos del negocio seleccionando la información que se necesita, quien la necesita y con qué periodicidad, permitiendo de esta forma poder tomar decisiones oportunas en cada momento”.

De esta manera iban surgiendo grandes investigaciones en las que se descubrían nuevas tecnologías y herramientas para el manejo de la información.

Es en la década de los 90' donde aparece el Business Intelligence 1.0. Y consigo mismo se desarrollan otras diversas aplicaciones BI, las cuales tenían algunos problemas en el momento de realizar el análisis al gran volumen de datos, ejecuciones en tiempo óptimo y la limitación de fuentes de datos. El uso de estas aplicaciones generaba precios inalcanzables. Por lo tanto, el BI no era accesible para todas las empresas.

Finalmente, en el año 2000 en la era de la interconexión y el surgimiento del manejo de grandes volúmenes de información como es el Big Data surge la idea “social BI” en la que se integraban datos externos a los sistemas operacionales. Así iba evolucionando y mejorando las aplicaciones del BI.

Actualmente estas herramientas de análisis permiten a los usuarios tomar decisiones en tiempos óptimos dado que se puede realizar un monitoreo constante a las bases de datos los cuales pueden brindar información útil para tomar decisiones.

2.1.9 Beneficios de la Inteligencia de Negocio

Los beneficios que puede tener una organización al implementar una inteligencia de negocios, los cuales se pueden clasificar en los siguientes:

1. Incremento de la eficiencia: Al ser los datos accesibles se puede generar información de gran valor, la cual podrá ser visualizada en plataformas especiales, lo que permite la visualización de datos los cuales generan informes, gráficos y resultados útiles para el uso y la toma de decisiones en tiempos óptimos.
2. Respuestas rápidas a situaciones de negocios: Para tomar una decisión es importante contar con buena información, de tal manera que sea sencilla en el momento de la búsqueda, donde el usuario pueda acceder a esta sin ninguna dificultad y de esta forma se pueda consolidar los datos. por consiguiente, el BI abarca un conjunto de herramientas que permiten dar respuesta rápida a la información.
3. Control de las áreas funcionales de la empresa: Cada organización maneja áreas en las cuales se generan a diario volúmenes de datos que brindan información. Esta puede ser aprovechada para conocer tendencias, proyectar datos y analizar escenarios.
4. Mejora el servicio al cliente: Al obtener la información más importante y en tiempo real, se puede brindar un mejor servicio a los clientes. Dado que se puede llegar a conocer los hábitos y necesidades más cotidianos de cliente.
5. Presenta la información por medio de tableros de indicadores con el fin de consolidar la comunicación en un tiempo óptimo y directo del panorama empresarial.
6. Los sistemas de inteligencia de negocios no solo brindan información analítica y fiable en tiempos oportunos si no, que también se puede visualizar la información

en formatos atractivos en los cuales el usuario pueda comprender el contenido de esta en una interfaz gráfica amigable.

2.1.10 Importancia de la Inteligencia de Negocios

La inteligencia de Negocios como se ha mencionado anteriormente es un conjunto de estrategias y herramientas que sirven para la toma de decisiones en grandes y pequeñas empresas, por tanto, el uso del BI en una organización puede contribuir al éxito empresarial

En [22] plantea que “[...] una solución BI completa permite a los ejecutivos y analistas del negocio observar que está ocurriendo, comprender por qué ocurre, predecir que va a ocurrir, colaborar con el equipo y decidir el camino que se debe seguir”.

Los directivos pueden obtener conocimientos de puntos clave como tendencias de mercado, competitividad entre otros. Es decir, con la información obtenida se puede planear y ejecutar acciones que generen impacto de crecimiento a nivel empresarial.

2.1.11 Componentes de una Arquitectura de Inteligencia de Negocios

Una Arquitectura de Inteligencia de Negocios está compuesta por diferentes módulos y procesos ver figura 3.



Fig.3. *Arquitectura de Inteligencia de Negocios. Tomado de [23]*

1. Fuente De Datos: Es el punto donde se encuentra la información útil para el análisis de negocios, por tanto, es importante conocer los datos internos y externos que se manejan en una organización. Estos datos pueden proporcionarse por medio de bases de datos, consultas, archivos de texto, sistemas ERP, sistemas contables, hojas de cálculo entre otros.

2. Proceso E.T.L: Según [24] plantea que “ Un proceso ETL es un programa que forma parte de un paquete ETL que recupera datos de una o varias fuentes y rellena una tabla de destino” dentro del mismo contexto se agrega que este proceso consiste en mover los datos los cuales pasan por un tratamiento y posteriormente son cargados a tablas de base de datos.

fases por los cuales debe pasar los datos.

- **Componentes de Extracción**

En[25] indica que “[...] esta fase convierte los datos a un formato homogéneo y consolidado para iniciar la fase siguiente. Uno de los requerimientos para esta secuencia, es que cause el menor impacto en el sistema origen. Si los datos a extraer son muchos, el sistema origen se podría ralentizar e incluso colapsar, provocando pérdida de información”.

- **Componentes de Transformación:** Según[25] la transformación de los datos consiste en “[...] reformatear y limpiar estos datos cuando sea necesario. En esta fase se aplica una serie de reglas de negocio o funciones sobre la información extraída para convertirlos en datos que después serán cargados”.
- **Componentes de Carga:** [...] la información que proviene de fase anterior es cargada en el sistema destino. Dependiendo de los requerimientos de la organización, este proceso puede abarcar una amplia variedad de acciones. Por ejemplo, en algunas bases de datos será necesario sobrescribir la información antigua con nuevos datos, mientras que, en otras, bastaría con resumir las transacciones y almacenar una magnitud considerada. En otras palabras, la fase corresponde a la carga de información en los repositorios de información destino[25].



Fig. 4. Modelo E.T.L. Tomado de [26].

2.1.12 ¿Qué es un Datawarehouse o Almacén de Datos?

Un Data Warehouse (ver figura 5) se puede definir como una bodega donde se integra los datos de diferentes fuentes los cuales son depurados. Se puede acceder a los datos que están depositados y tomar decisiones con respecto a la información que arrojan estos mismos. Antes de integrar los datos en el repositorio se debe realizar el proceso ETL mencionado anteriormente.

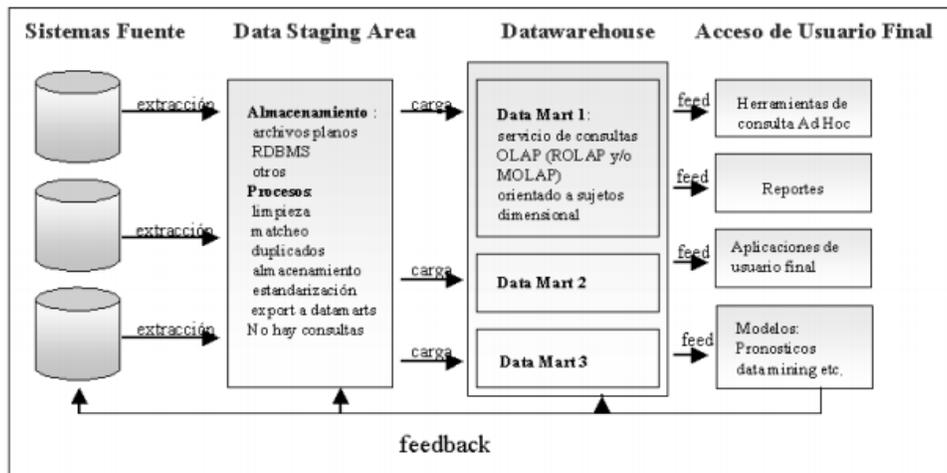


Fig.5. Elementos de Una Data Warehouse. Tomado de [27].

2.1.13 Componentes de un Data Warehouse o Almacén de Datos

1. Metadatos: son datos sobre los datos y describen cosas tales como:

- ✓ Cuando fue extraída la información del sistema de origen
- ✓ Cuando se cargaron los datos en el almacén de información
- ✓ De cual sistema de origen se generó un elemento
- ✓ Como fue calculado algo, por ejemplo, ganancia = (precio x cantidad vendida) – descuentos

2.1.14 ¿Que es un DataMart?

Según[28] “un **DataMart** es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio en específica [...]”. También se puede decir que el objetivo de este es satisfacer la necesidad de un proceso puntual, de acuerdo a esto[28] agrega que “un datamart puede ser alimentado desde los datos de un DataWarehouse, o integrar por sí mismo un compendio de distintas fuentes de información”.

2.1.15 Tipos de DataMart

Se observa la clasificación de los Data Mart según [28]

- **DataMart dependiente:** “los Datamart dependientes son aquellos que reciben los datos desde una Data Warehouse. En este tipo de Datamart la fuente de datos es única”.
- **DataMart independiente:** “son aquellos que toman sus datos directamente desde los sistemas transaccionales y no dependen de otros Datawarehouse. Este tipo de Datamart se alimenta generalmente de las organizaciones”.
- **DataMart híbrido:** “los Datamart híbridos permiten combinar las fuentes de datos un DataWarehouse corporativo con otras fuentes de datos tales como sistemas transaccionales y/o operacionales”.

2.1.16 Concepto Big Data

Para [29] “este concepto de Big Data se aplica para toda aquella información que no puede ser procesada o analizada utilizando procesos o herramientas tradicionales. Hay cuatro características claves que definen la información relativa al Big Data:”

- “Volumen: Los datos al Big Data se producen en cantidades muchos más grandes que los datos tradiciones”.
- “Velocidad: Los flujos de datos de medios sociales”.
- “Variedad: Los formatos de datos tradicionales tienden a ser relativamente bien definidos por un esquema de datos”.
- “valor: El valor económico de los diferentes datos varia significativamente”.

2.1.17 Minería de Datos

En[30] se describe el concepto de minería como “un proceso de descubrimiento de nuevas y significativas relaciones, patrones al examinar grandes cantidades de datos”. También se agrega que para el descubrimiento de los mismos. Se debe utilizar técnicas de minería de datos las cuales generan el descubrimiento de patrones y comportamientos que generan información relevante para la toma de decisiones.

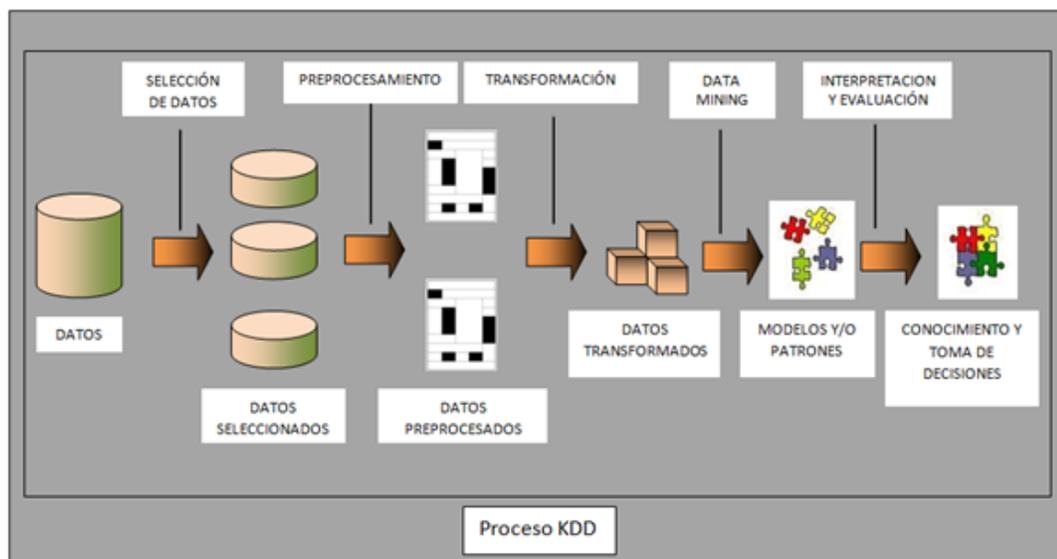


Fig.6. Metodología KDD. Tomado de[30]

2.1.18 Concepto de Datos Multidimensionales (Cubos OLAP)

El funcionamiento de estas aplicaciones posee un tipo de bases de datos con estructura multidimensional, esta es conocida como cubo OLAP. Es importante indicar estos cubos son vectores en los cuales la información está relacionada con otros ejes del cubo, por tanto, el ordenamiento que se lleva en esta jerarquía es posible que los datos sean analizados más rápidamente. También con la inclusión de estos vectores o cubos se ha ampliado el tratamiento masivo de las bases de datos relacionales, el cual permite un amplio procesamiento de grandes volúmenes de información.

Cada una de las dimensiones que alberga la base de datos incluye un campo explicito para un tipo de dato único, este posteriormente podrá ser contrastado con la información que posee las otras dimensiones que contiene esta base. Se realiza este procedimiento con el fin de obtener informes productivos y relevantes para la compañía.

De acuerdo a su nombre el cubo mencionado contiene la característica de ser una base de datos que posee varias dimensiones como se puede observar en la (figura 6).

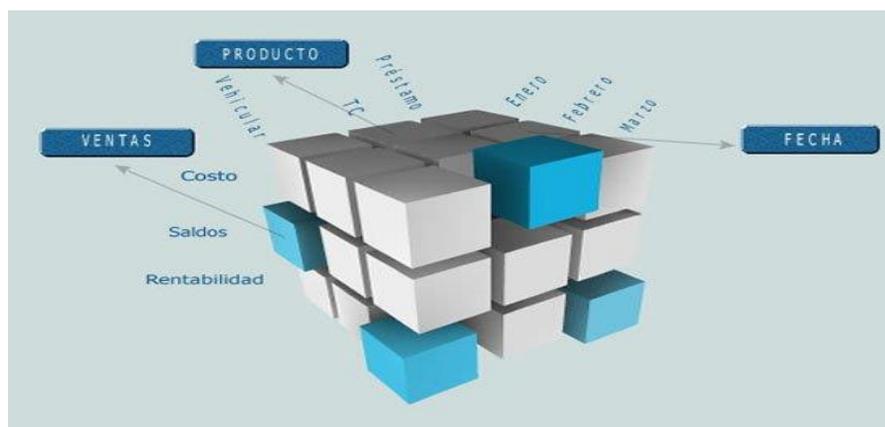


Fig.7. Cubo Multidimensional (OLAP). Tomado de [31]

2.1.19 ¿Que es un sistema OLAP (On-Line Analytical Processing)?

En [32] se indica que “Los sistemas de procesamiento analítico en línea se basan en la estructuración y presentación de la información por medio de ejes o dimensiones de interés para las personas que se responsabilizan del análisis de negocio de la organización”.

Es importante indicar que sistemas OLAP alimentan la información de los sistemas Data Warehouse mediante la profundización en los datos partiendo de lo siguientes:

- Consultas y elaboración de informes
- Sistemas de apoyo de decisión (DSS)
- visualización de datos

2.1.20 Diferencia entre un sistema OLTP y OLAP

La diferencia de estos sistemas es que el sistema OLTP son bases de datos orientados al procesamiento de transacciones en línea que contienen procesos que son ejecutados al definir una acción. Una transacción genera un proceso atómico este es validado por medio de parámetros los cuales definen unas acciones que involucra operaciones como inserción modificación y borrado. A diferencia de los sistemas OLAP las cuales son bases de datos orientadas al procesamiento analítico, usualmente son ejecutadas con el fin de realizar lectura y análisis a grandes cantidades de datos.

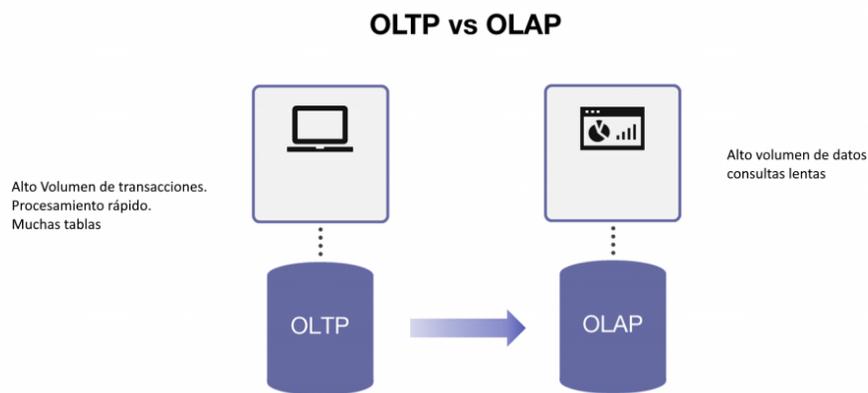


Fig.8. Sistema OLTP Y OLAP. Tomado de[33].

2.1.21 Herramientas Para El Desarrollo De Inteligencia de Negocios De Software Libre

Existen diferentes herramientas de inteligencia de negocios para diferentes sectores, es decir algunas están enfocadas explícitamente a pequeñas empresas y otras a multinacionales. Seleccionar la más conveniente dependerá la necesidad y presupuesto que posea la empresa.

En este punto se describe las funciones y características de cada herramienta

2.1.22 CloverETL

Es un software el cual está basado en la integración de datos estructurados, basada en la herramienta ETL es de código abierto. Esta plataforma se puede utilizar para migrar datos, limpieza y las operaciones de transformación de datos.

Esta herramienta elabora con datos estructurados. Estos pueden ser almacenados en archivos de texto, por ejemplo: registro de longitud fija, archivos CSV, XML o también los producidos por formatos MS Excel, dBase o FoxPro que estén en forma binaria. Por tanto,

CloverETL admite combinar, transformar, y circular datos provenientes de cualquier origen.

Aplicaciones para CloverETL

- Migración de datos
- Integración de datos
- Recopilación de datos
- ETL para almacenes de datos
- CloverETL se basa en el Grafico de Transformación
- Permite un diseño fácil e implementación de soluciones de migración de datos
- Cada componente realiza una acción u operación determinada (extracción de los datos almacenados, ordenamiento, filtrado y transparencia. etc.)

Especificaciones técnicas

Este software está escrito en lenguaje java y es compatible con las plataformas que se ejecutan en Java Virtual Machine (JVM) en la actualidad ha sido utilizado con gran comodidad en Linux, AIX, Solaris, HP-UX, Windows, AS/ 400 entre otros.

Requisitos hardware

- se debe contar con al menos 512 MB de RAM
- se debe contar con al menos 300 MB de espacio libre en el disco

Requisitos de software

- Sistemas operativos soportados: Microsoft Windows 32/64 bits, Linux de 32/64 bits, Mac OS x de 64 bits

- Oracle JDK 7

2.1.23 CloverETL Engine

Este es muy interactivo debido que se puede integrar con otras aplicaciones donde juega el rol de transformador de datos, es decir, toma los datos de la misma aplicación enviándolos y recibéndolos. Esta herramienta se basa en la plataforma extensible de Eclipse y hace parte de la familia CloverETL el cual es una aplicación basada en java para la extracción, transformación y carga de datos.

Características Principales

- Convierte los datos en diferentes tipos de caracteres (ASCII, UTF-8, ISO-8859-1, ISO-8859-2 etc.).
- Soporta las bases de datos comunes e importantes de la actualidad (Oracle, MS SQL, DB2, Infomix, Sybase) y software de código abierto (MYSQL, PostgreSQL) es compatible con otras bases de datos a través de la capa JDBC.
- se ejecuta en plataformas de 32 bits y 64 bits

2.1.24 CloverETL Server

CloverETL server es una herramienta de plataforma servidor que permite el monitoreo, implementación, automatización e integración de datos con proyectos de alto nivel de complejidad. De igual manera esta plataforma está diseñada para programar y ejecutar gráficos ETL.

CloverETL Server le permite lograr:

- Gestión centralizada de trabajos ETL
- Integración en los flujos de trabajos empresariales

- Entorno multiusuarios
- Ejecución de gráficos en paralelo
- Seguimiento de ejecuciones de grafico
- Programación de tareas
- Agrupación y ejecución distribuida de gráficos
- Lanzamiento de servicios

Requisitos hardware

- Memoria RAM 4 GB recomendada 16 GB
- Procesador de 4 núcleos
- Espacio en el Disco 1 GB

Requisitos software

- Microsoft Windows Server 2003/2008/2012 32/64 bit
- GNU/ Linux 32/ 64 bit
- Mac OS x
- Unix
- HP-UX
- AIX
- IBM System (AS/ 400)
- Oracle JDK 7/8 32/64 bit
- IBM SDK 7

2.1.25 Pentaho Business Intelligence Suite

Es líder en cuanto a soluciones de Business intelligence Open Source. De tal forma que ofrece un conjunto de soluciones propias con las cuales se puede mantener desarrollar y explotar proyectos BI. Estos pueden ser desde los procesos ETL, data integration y cuadros de mando. De acuerdo a lo descrito esta plataforma integra proyectos y por lo tanto esta herramienta es considerable flexible para el cubrimiento de las necesidades

de las organizaciones. En la (figura 9) se muestra la arquitectura de la Plataforma Open Source Pentaho

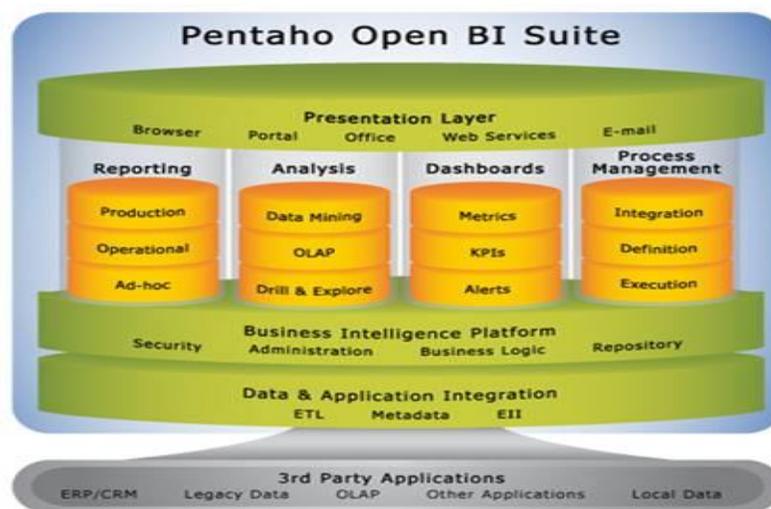


Fig.9. Arquitectura de la Plataforma Pentaho Open BI Suite. Tomado de [34]

2.1.26 Pentaho Reporting

Es un conjunto de herramientas de código abierto que permite la creación de informes analíticos y relacionales a partir de un gran volumen de datos provenientes de diferentes fuentes lo que generan información significativa. Las pantallas basadas en XML brindan un funcionamiento muy útil y por tanto es compatible con dispositivos pueda crear informes con excelentes pixeles en sus datos. Se conecta con PDF, Excel, HTML, impresoras, XHTML, archivos en formato XML y CSV.

Ademas esta es compatible con fuentes de datos como SQL, MDX, Pentaho Metadata entre otros.

Especificaciones

- Es una plataforma de código abierto, incluye Reporte Designer, Reporte Engine y Reporting SDK.
- Es un software que incluye un alto rendimiento y un bajo consumo de memoria
- El procesamiento de informes puede subir de nivel desde los más pequeños hasta informes empresariales con una gran envergadura
- Se integra perfectamente con el BI Server y por tanto permite compartir informes entre operadores.

2.1.27 Pentaho Analysis

Pentaho Analysis proporciona poder analítico a los usuarios permitiendo obtener conocimientos y comprensión necesaria para tomar decisiones en tiempos óptimos para beneficio del negocio. Esta herramienta se basa en el motor de procesamiento en línea (ROLAP).

El motor de Pentaho Analysis (Mondrian) se puede configurar por medio de un archivo de propiedades. Lo que genera un mejor rendimiento y funcionalidad del motor la fuente de datos en ciertas condiciones.

- Examina la información organizacional mediante la excavación y la tabulación cruzada de los datos.
- Experiencia y velocidad en tiempos de respuesta a las consultas analíticas.
- Observación de información multidimensional, elección de métricas y atributos específicos para analizar.
- Se implementa independiente o integrado con otros productos de la suite Pentaho BI.

Pentaho Analyzer: suministra el análisis de informes intuitivos e interactivos, es factible su comprensión haciendo que usuarios no expertos entiendan rápidamente la información de negocio.

Requerimientos Software

- Windows (XP SP2, 2008,7)
- Distribuciones modernas de Linux (SUSE Linux Enterprise Desktop y Server 10 y Red Hat Enterprise Linux)
- Mac OS X 10.5 (compatible con el más reciente)

Requerimientos Hardware

Servidor:

- RAM al menos 4GB
- Disco Duro al menos 2GB
- Procesador AMD 64 de doble núcleo o Intel EM64T

2.1.28 Pentaho Dashboards

Esta herramienta proporciona una perceptiva uniforme que permite crear secciones con o sin información. El tablero es la reunión de varios informes contenidos en una sola pantalla, si un usuario desea obtener varios informes al mismo tiempo esta permite realizar esas acciones. De este modo, si desea consultar un sitio web esta admite el acceso rápido a páginas gráficos y cuadros mientras se desarrolla otros informes. Es importante añadir que Pentaho Dashboard ofrece información compleja que se requiere para comprender y mejores procesos de la organización.

Visibilidades que ofrece:

- Incluye formatos con visualizaciones en Adobe Flash de modo que el usuario podrá realizar seguimiento y observar las métricas de negocio teniendo la preceptiva de lo que se desea.
- Permite crear con facilidad cuadros de mando
- Integración de herramienta Pentaho Reporting y Pentaho Analysis, da acceso a que los usuarios puedan obtener información de procedimientos que están generando buenas mejoras y cuáles no.
- Alerta integrada para inspeccionar continuamente las excepciones y notificar a los beneficiarios a tomar acciones

2.1.29 Pentaho Data Integration

Es una herramienta flexible que permite la recopilación de datos de fuentes plurales, migración de datos entre los repositorios, como son las bases de datos y archivos de aplicaciones. De la misma manera convierte los datos en un formato unificado para que este sea accesible y relevante para usuarios finales.

La integración proporciona el motor del proceso ETL por lo que facilita una correcta recopilación, limpieza y almacenamiento siempre utilizando un formato uniforme y consistente, por lo tanto, esta proporciona dos formas de almacenar sus transformaciones, trabajos y conexiones de bases de datos

- **Pentaho Enterprise Repositorio (repositorio empresarial Pentaho)**
- **File- Base** (archivo bases)

Componentes principales de Pentaho:

- Spoon: es una aplicación de escritorio que maneja una interfaz gráfica y un editor para transformaciones y trabajos ver **figura 6**. Este entorno integrado le permite, como desarrollador de BI trabajar en estrecha colaboración y cooperación con los usuarios empresariales para crear soluciones de inteligencia empresarial de forma más rápida y eficaz.
- Pan: es un proceso de línea de comando independiente que se puede utilizar para la ejecución de transformación y trabajo que se realizó en Spoon
- Kitchen: es el programa que ejecuta los trabajos que están diseñados en la gráfica Spoon, ya esté en XML o en repositorios.
- Carte: es un contener web el cual permite la configuración de un servidor ETL remoto.

Requerimientos

- Procesador Celeron. 2.0GHz
- Memoria RAM 128Mb o superior
- Espacio es el disco duro 200Mb
- Otros requerimientos Máquina virtual de Java (JRE) versión 1.5 o superior

Sistema operativo

- Windows, Linux, Unix, Macintosh
- Licencia GNU Lesser General Public License (versión 2.1)



Fig.10. Interfaz Data Integración (Spoon). Tomado de [34]

2.1.30 Pentaho Data Mining

Data Mining es el proceso de correr los datos a través de sofisticados algoritmos para descubrir patrones y correlaciones significativos que de otro modo pueden permanecer ocultas, estos pueden ser usados para ayudar a entender mejor el negocio, y también para explorar el futuro comportamiento del negocio a través de un análisis predictivo.

Pentaho Data Mining se diferencia por su carácter abierto, compatibles estándares naturales, uso de tecnología de minería de datos Weka, y una estrecha integración con un núcleo de las capacidades de inteligencia de negocios incluyendo informes, análisis y cuadros de mando.

Pentaho Data Mining se puede implementar como:

- En cuanto a los usuarios finales se refiere, la minería de datos opera enteramente en el fondo del sistema, los usuarios pueden ver los resultados a través del correo electrónico u otras páginas web, que puede incluir Pentaho dashboards.

- Un conjunto de componentes que permiten a los desarrolladores de Java crear rápidamente soluciones personalizadas de información utilizando objetos Java o Java Server Pages (JSP). Estos pueden ser estrechamente integrados con otras aplicaciones o portales.
- Funciona conjuntamente con otros componentes de la Suite Pentaho BI general.

Características y Beneficios

- Proporciona una mejor perspectiva en patrones ocultos y relaciones en los datos
- Permite aprovechar las correlaciones para mejorar el desempeño organizacional
- Provee indicadores de desempeño futuro
- Permite la incorporación de recomendaciones en sus aplicaciones
- Permite sacar el máximo provecho de una serie de algoritmos de minería de datos

Pentaho posee una Tecnología con un potente motor de minería de datos y herramientas de diseño gráfico.

2.1.31 Tableau

Es una herramienta de visualización de datos, es utilizada para realizar inteligencia de negocios la cual simplifica los datos generando fácil interpretación y manejo de la misma. Posee funciones de utilidad simple como, por ejemplo: el usuario puede realizar las acciones de arrastrar y soltar, esto permite que los datos sean accesibles, analizados, compartidos entre usuarios, además se pueden elaborar informes de diferentes procesos.

Tableau cuenta con una arquitectura de cliente- servidor de múltiples niveles lo que hace que sea escalable brindando los servicios a clientes móviles, clientes web y el software de escritorio.

- .

Características y funcionalidades de Tableau

- 1. Numerosas conexiones de datos:** se puede conectar a diferentes fuentes sin tener ningún grado de programación (Redshift, Cloudera, Hadoop, SQL Server, Google Analytics, entre otros.)
- 2. Tableros integrados:** se puede realizar integraciones de paneles en las aplicaciones existentes como por ejemplo Salesforce, SharePoint y Jive.
- 3. Modo “arrastrar y soltar”:** Este modo permite realizar una integración de forma sencilla con los datos y por lo tanto se puede crear elementos que generan fácil visualización.

Tableau Creator	70\$ por usuario y mes. Incluye la aplicación de Tableau Desktop, Tableau Pre y una licencia Creator en Tableau Server.
Tableau Explorer	35\$ por usuario y mes. Se exige un mínimo de 5 usuarios, el pack incluye una licencia de Explorer Tableau Server.

Tableau Viewer	12\$ por usuario y mes. Debe constar con 100 usuarios como mínimo. El pack incluye una licencia Viewer en Tableau Server
Nota: Si se busca una licencia de carácter individual, el valor es de 70\$ mensuales. Incluye Tableau Desktop, Tableau Prep y licencia Creator de Tableau Server o Tableau Online.	

Tabla 1: Costo de las Herramientas por Equipos y Organizaciones

Fuente: Elaboración Propia

Requisitos mínimos Tableau Desktop

- Microsoft Windows 7 o la más reciente
- Memoria 2GB
- 1,5 GB de espacio en el disco
- La CPU debe admitir conjunto de instrucciones SSE4.2 Y POPCNT

Requisitos mínimos del sistema Tableau Server

- Microsoft Windows server 2016,2012,2012 R2 2008 R2
- Amazon Linux 2, Red Hat Enterprise Linux (RHEL) 7.3+, CenOS 7.3, Oracle Linux 7.3+, Ubuntu 16.04 LTS en chipset *64
- 8 núcleos físicos, 16 CPU virtuales
- Memoria de RAM 32 GB
- Espacio mínimo en el disco de 50 GB

2.1.32 Herramientas Para El Desarrollo De Inteligencia de Negocios de Software Con Licencia

2.1.33 Oracle Business Intelligence Suite Enterprise Edition plus (EE)

Es una plataforma que permite al analista profesionales o desarrolladores a tomar decisiones rápidas bien fundamentadas. Debido a que su principal característica es el descubrimiento de patrones e interfaz de visualización de datos de manera sencilla e intuitiva, brinda una variedad de funciones como los cuadros de mando bajo el entorno OLAP, inteligencia artificial, reporting empresarial, detención de alertas, integración con Office, consultas ad-hoc, inteligencia predictiva en tiempo real, informes financieros minería de datos entre otras aplicaciones basadas en servicios web (J2EE y NET). Estas aplicaciones requieren un acceso a grandes volúmenes de datos. Los componentes de esta herramienta manejan una arquitectura en común, lo que permite al usuario trabajar sobre un ambiente amigable y sin problemas.

Principales beneficios

- 1. Visualización:** proporciona una apariencia muy amigable y de fácil interpretación para el usuario.
- 2. Usabilidad:** el usuario puede arrastrar y soltar columnas para elaborar informes a su modo, esto permite que el reporte sea fácil de usar.
- 3. Integración con las principales fuentes de datos:** Microsoft o proveedores de terceros se acoplan a las exigencias del usuario de la solución Oracle, el cual garantiza su integración para promover el reporting.

Oracle BI Suite Enterprise Edition Plus

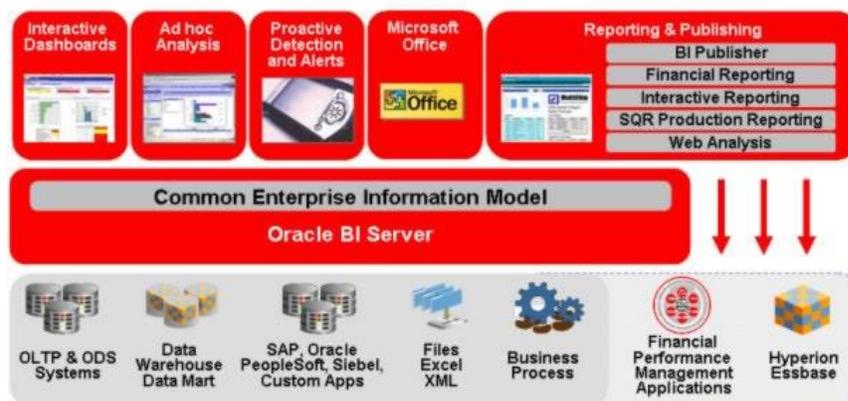


Fig.11: Elementos de la Plataforma Oracle BI Suite Enterprise Edition Plus. Tomado de [35].

2.1.34 SAP BI

Esta herramienta permite obtener datos de diferentes fuentes, se puede analizar realizar transformaciones, informes, cuadros de mando y realizar modelamientos de machine learnig como por ejemplo predicciones, proyecciones. Esta transforma los datos en información estratégica útil, disponible en cualquier momento lugar. De igual forma se dice que es un sistema de planificación de recursos empresariales en que todos los módulos trabajan en grupo para lograr un ordenamiento en cuanto al funcionamiento de la organización.

Posee ciertas características como:

- BI en tiempo real.
- Consumo de información simplificada, personalizada y dinámica.
- Está dirigido especialmente a la PYMES.
- Integra datos y procesamientos empresariales.

Requerimientos Sistemas operativos

- Windows server 2016(64 bits)
- Windows server 2012(64 bits)
- Windows server 2012 R2(64 bits)
- Windows server 2008 R2(64 bits)
- Windows server 2008 (64 bits)
- Memoria RAM: se recomienda 16GB
- Espacio en el disco duro: mínimo 100 GB
- CPU: 4 CPU con velocidad de 3 GHz

Nota: Para adquirir el software SAP BI es necesario como mínimo la compra de una licencia profesional, por lo general el valor aproximadamente es de \$ 2,700 USD; es importante comentar que no todas las empresas necesitan este tipo de herramientas.

2.1.35 IBM Cognos

Esta herramienta brinda a los usuarios un medio de trabajo muy interactivo debido que otorga facilidad para acceder a los datos a través de graficas atractivas, las cuales permiten identificar en tiempos oportunos la performance del negocio y así dar respuesta a las acciones correctivas necesarias. Además, suministra información confiable para que las decisiones se basen en datos correctos, certeros y seguros.

Características Principales:

- Los usuarios pueden realizar tareas administrativas.
- Los usuarios pueden acceder a la opción administración de contenido, donde se puede administrar las exportaciones, importaciones, comprobaciones de consistencia y actualizaciones de informes.

- Los usuarios pueden acceder al origen de los datos mediante la opción conexiones de datos. Se define el origen de los mismos y las conexiones e inicio de secciones
- los usuarios pueden asignar grupo y roles.

Requisitos

- Sistema operativo Windows
- Memoria RAM: 10 GB como mínimo
- Espacio en el Disco: mínimo 3,5 GB para instalar el software y 5 GB de espacio de espacio libre en la unidad que contiene el directorio temporal

Debe tener una de las bases de datos siguientes disponibles para almacenar los dtos

IBM Cognos:

- Oracle
- IBM Db2
- Microsoft SQL Server
- Informix
- Se necesita conectividad TCP/ IP para todos tipos de bases de datos.

2.1.36 TIBCO Spotfire

Es un software, en el cual se puede realizar reportes predictivos en base a **clustering y forecasting**. De igual manera incluye una potente plataforma para **geo codificación** y análisis de mapas geográficos. También permite a los usuarios visualizar y analizar sus datos sin asistencia técnica es compatible con una amplia gama de casos de uso como la creación de paneles hasta un caso de análisis predictivo en tiempo real. Esta posee un acceso rápido y conveniente desde archivos, bases de datos, servicios web, API o

servidor, la información puede combinarse desde múltiples fuentes para facilitar el descubrimiento profundo de datos

A continuación, se menciona las siguientes características:

- Almacenamiento seguro de datos.
- Análisis de rentabilidad
- Creación de informes ad hoc
- Análisis predictivo
- Desarrollo de aplicaciones
- La creación y el análisis no requieren descarga
- proporciona información fácil de interpretar que es vital para detectar tendencias y patrones valiosos.

Nota: el valor de licenciamiento es aproximadamente US\$ 650,00/ año, puede obtener la versión gratuita por un tiempo de 30 días

2.1.37 Sísense

Este software permite crear informes de inteligencia comercial, donde el usuario puede transformar de manera fácil los datos en informes muy completos, es decir que la información que poseen estos es deslumbrantes para la toma de decisiones. Con el uso de esta herramienta los departamentos o áreas comerciales podrán obtener la visibilidad completa de los datos y pueden generar conocimientos muy prácticos con gastos del área TI mínimo. Se puede conectar con las bases de datos, aplicaciones relevantes y mezclar diferentes fuentes datos para visualizar los datos en tableros o informes en formato PDF.

A continuación, se mencionan algunas características principales como:

- Análisis de rentabilidad, de tendencia, análisis predictivo, consulta ad hoc, creación de diagramas, importación y exportación de datos, métricas de rendimiento, arrastrar y soltar, alertas notificaciones y creación de informes en tiempo real, entre otras.

2.1.38 SAS Business Inteligencia

Esta herramienta posee las funciones visualización lo que permite a los productores de información explorar los datos, investigar patrones y descubrir información oculta de manera dinámica. Contiene una amplia gama de programas las cuales brindan opciones personalizadas como, por ejemplo: presentación grafica lo que permite al usuario tener el acceso a datos y perceptivas que nos visibles fácilmente en forma tabular. Entre las distintas funcionalidades está la posibilidad de crear análisis predictivos, visualizaciones, la obtención de indicadores de tendencia o rendimiento y elaborar estrategias empresariales.

Posee las siguientes características:

- Gestión de datos efectiva, garantiza que estos sean seguros y alineados con los objetivos de negocio y usados para dirigir y tomar mejores decisiones.
- Es una herramienta rápida y posee una interfaz simple comprensión.
- Brinda información correcta a los usuarios.
- Mediante la herramienta se puede tener conocimientos claro y visualizar los datos más importantes.

2.1.39 Hyperion Interactive Reporting

Esta herramienta brinda a los ejecutivos, usuarios empresariales y analistas, capacidad de análisis y consulta de usuario intuitivo y dirigido. Una interfaz intuitiva e interactiva, permite a los usuarios diseñar cuadros de mando, supervisar de forma rápida y navegar por información pertinente. Es fácil de implementar, proporciona acceso directo a los datos bloqueados en los sistemas transaccionales y de información y aprovecha los el almacén de datos actuales para entregar informes de inteligencia de negocio. Con seguridad incorporada, categorización de informes, versiones automáticas y archivo, el componente posee una completa solución basada en la web para distribución y administración de reportes listos para imprimir.

2.1.40 Hyperion Financial Reporting

Este software da una solución especial de reportes financieros, que genera un formato, libros financieros e informes administrativos que cumplen con los reglamentos y requisitos externos. Las características de este módulo integran información financiera que soporta cambios de moneda, GAAP, IFRS y otras normas financieras. Además, el módulo soporta XBRL para la transmisión electrónica y la presentación de información financiera.

2.1.41 Metodologías Para La Construcción De Almacén De Datos

2.1.42 Metodología Ralph Kimball

La metodología de Ralph Kimball tiene como base el ciclo de vida de los data warehouses, o almacén de datos llamado Ciclo de Vida Dimensional del Negocio o BDL (Business Dimensional Lifecycle), en el cual se representa la variedad de fases

por las que debe pasar un proyecto de Inteligencia de Negocios ver figura 14. Las etapas o fases desarrolladas por Kimball han sido diseñadas para trabajar en paralelo, es decir, se clasifican como una metodología “Bottom Up” debido a que la información es tomada de los sistemas transaccionales y es cargada directamente a algún tipo de datamart que están conectados por medio de las dimensiones que formaran el almacén de datos que será la parte central de la solución Business Intelligence.

A continuación, se presenta el esquema de la metodología de Ralph Kimball

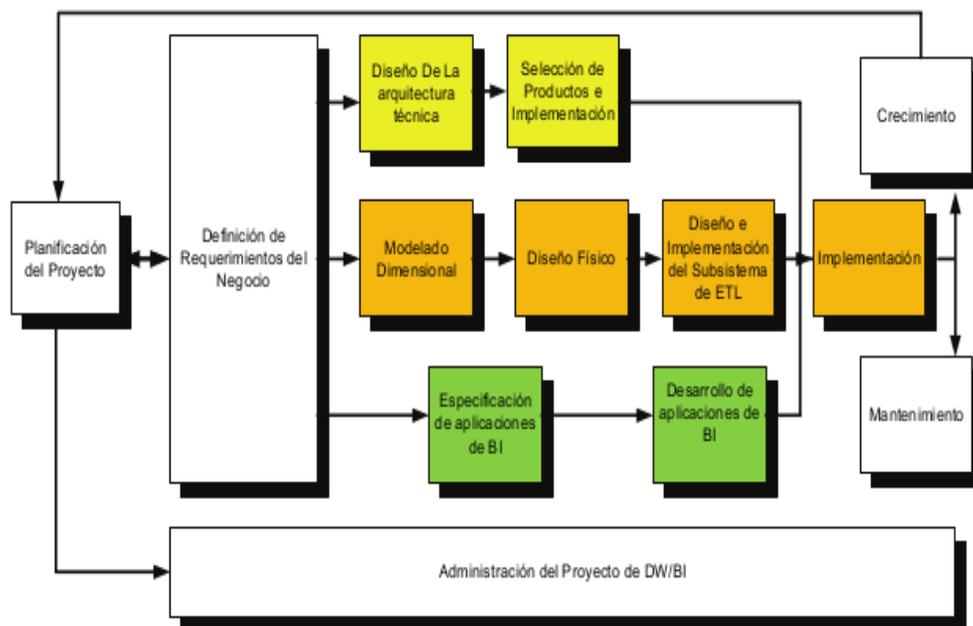


Fig.12: Método de Ralph Kimball. Tomado de [23]

Fase 1: Planificación De Proyecto

En este proceso se determina el propósito del proyecto de Business Intelligence, sus objetivos específicos, riesgos y el alcance del proyecto, basándose en la necesidad y requerimientos del negocio. Mas no en intervalos de tiempos establecidos, lo cual nos permitirá lograr los objetivos estratégicos empresariales.

Fase 2: Definición de los requerimientos del negocio

Para la definición de los requerimientos del negocio se ejecuta un proceso que consiste en realizar entrevistas al personal. Por tanto, se debe tener un buen conocimiento sobre la lógica del negocio, competidores, industria, los clientes y factores claves que guían al mismo. Lo anterior se realiza con el fin de obtener la información necesaria para el diseño del producto BI. También determinar los requerimientos y así establecer consideraciones de diseño propias.

La definición de los requerimientos es un poco compleja debido a que es la base para la elaboración las siguientes fases, en esta parte no puede cometer errores ya que podría crear una mala solución y sería un fracaso para lo toma de decisiones de la organización.

Fase 3: Diseño De La Arquitectura Técnica

La etapa de diseño se basa en conceptualizar los requerimientos que han recolectados para la elaboración del proyecto BI. Dentro del diseño se encuentra las siguientes actividades:

- 1. Diseñar la arquitectura técnica:** El diseño de la arquitectura técnica cubre el entorno general de los procesos y herramientas que se aplican a la información. Por tanto, se debe considerar los requerimientos del negocio, los sistemas actuales y la dirección técnica estratégica planeada. Esta arquitectura está compuesta por tres capas que son: la de integración, análisis y visualización.
 - **La capa de integración:** en esta parte se extraen los datos del origen de las fuentes y se enmarcan los campos explícitos conforme al modelo de datos, después pasan por el proceso E.L.T. donde se realiza el proceso de limpieza que consta en eliminar inconsistencias y errores. Por último, los datos son almacenados en tablas que

quedaran en un tipo de datamart.

- **La capa de análisis:** En esta capa se procede a aplicar las herramientas OLAP las cuales permitirán el análisis de la información.
- **Capa de visualización:** Muestra al usuario final los resultados, de las consultas obtenidas por las herramientas ilustrando datos comprensibles para el usuario.

En la arquitectura técnica existen dos conjuntos el back room y front room los cuales contienen diferentes requerimientos. Estos ofrecen servicios y componentes de almacenamiento de datos. Son independiente uno del otro.

- **Entorno Back Room:** Es la parte interna de la arquitectura técnica, aquí se detallan los procesos de E.T.L. que van desde los orígenes de datos hasta la carga de los nuevos datos en el almacén de datos.
- **Entorno Front Room:** Es el entorno donde se presentan los datos que son obtenidos del DW al usuario a través de servicios de navegación, seguridad, monitoreo, generación de reportes y manejo de consultas.

Dentro de la arquitectura técnica se debe identificar y analizar las fuentes de información. Las cuales darán el acceso a la extracción de los datos necesarios para enriquecer el nuevo sistema y de esta manera realizar el cumplimiento de los requerimientos

Fase 4: Selección de la Herramienta e Instalación

Este proceso se basa en el diseño de la arquitectura técnica es decir se realiza la selección e instalación del producto BI donde se evalúa y se tiene en cuenta ciertos criterios como:

- La plataforma hardware
- Las Bases de datos.
- Las Herramientas de Consulta.
- Las Herramientas de Reportes.
- La Instalación de productos, componentes y herramientas.

Una vez finalizado este proceso se debe realizar las pruebas de funcionamiento de los productos instalados. Esto con el fin de determinar y garantizar la integración con el entorno del almacén de datos.

Fase 5: Modelo Dimensional

El modelo dimensional es una técnica que permite mostrar la información de los procesos del negocio de una forma estandarizada y sencilla. Por lo cual se optimiza el acceso a los datos. En el modelo dimensional se crea el modelo lógico de la solución lo que permitirá adquirir la información necesaria para los procesos de gestión y toma de decisiones por parte de la organización.

Por otra parte, se dice que cada modelo dimensional está compuesto por una tabla que contiene las llaves de las otras tablas

Fase 6: Diseño y Desarrollo de Sistema ETL

El desarrollo del sistema ETL es muy importante, dado que es la etapa primordial para la elaboración del almacén de datos, y por ende se debe garantizar que el tratamiento de los datos sea consistente y que la calidad de los mismos sea adecuada para el proceso. Es importante mencionar que si los procesos E.T.L se diseñan adecuadamente se puede llevar a cabo los procesos de extracción, transformación y carga.

Se agrega que a un buen diseño de procesos E.T.L se le puede aplicar reglas para depuración de los datos, mejorar la calidad de los mismos y de esta manera integrar la información en un formato adecuado para su uso por parte de la variedad de herramientas de análisis.

Para el desarrollo del proceso ETL se debe diseñar el análisis a los usuarios finales. Esto hace referencia al diseño de las herramientas, las cuales deben de permitir a los usuarios realizar procedimientos de consulta y el análisis a los datos que están almacenados en la data Warehouse.

Fase 7: Diseño de Aplicaciones Business Intelligence

Esta fase sirve para identificar los diferentes perfiles de los usuarios lo cual nos permitirá establecer los diferentes tipos de aplicaciones que darán el acceso a la información de acuerdo con los perfiles asignados a cada usuario.

Kimball indica que las aplicaciones de Inteligencia de Negocios deben cumplir con un requisito esencial, estas aplicaciones deben tener la capacidad de realizar un análisis AD HOC, el análisis AD HOC permite a los usuarios cambiar los parámetros en un reporte para crear versiones personalizadas de dichos reportes.

Fase 8: Desarrollo de Aplicaciones Business Intelligence (BI)

Estas aplicaciones se desarrollan en base al diseño establecido en la fase o etapa anterior. Existen diferentes tipos de aplicaciones B.I. Algunas se basan en:

- **Basadas en Web:** Son aplicaciones para acceder mediante un navegador web. Los usuarios pueden observar los reportes a través de la intranet o internet.
- **Herramientas Independientes:** Con esta herramienta se diseñan plantillas de reportes que los usuarios van a poder acceder a través de una interfaz, no es una implementación muy flexible ya que los reportes son definidos con anterioridad y si se desea involucrar un nuevo reporte el nivel del costo sería muy alto
- **Herramienta de Interfaz Ejecutiva:** Provee una estructura de acceso a las plantillas de reportes a través de una serie de pantallas las cuales permiten una fácil navegación a través de estas.
- **Interfaz por código:** Proporcionan un API que permite diseñar una interfaz, se desarrollan reportes con una interfaz gráfica que admite un análisis más eficiente, estos reportes se pueden ajustar de acuerdo con las necesidades del usuario.
- **Documento Entregable:** Este resulta de la etapa de Diseño de la Arquitectura del proyecto.

Fase 9: Implementación

Esta fase es donde se unen las diferentes herramientas B.I., los datos y las aplicaciones B.I., para usuarios finales y factores externos tales como capacitación, soporte técnico, etc., que admitirá el despliegue del proyecto de inteligencia de negocios.

Fase 10: Gerenciamiento del Proyecto

En este paso se certifica que las actividades que comprenden el ciclo de vida se produzcan de manera adecuada, a tiempo y de forma sincronizada, el gerenciamiento acompaña todo el ciclo de vida, las operaciones principales son: la conducción del equipo de trabajo, el monitoreo del estado del proyecto, el mantenimiento del plan y documentación, el manejo del alcance, y la comunicación entre los requerimientos del negocio y las restricciones de los datos. De acuerdo a la descripción anterior sirve para poder manejar correctamente las expectativas dentro de la empresa, en la (figura 13) se ilustra el gerenciamiento de un proyecto.

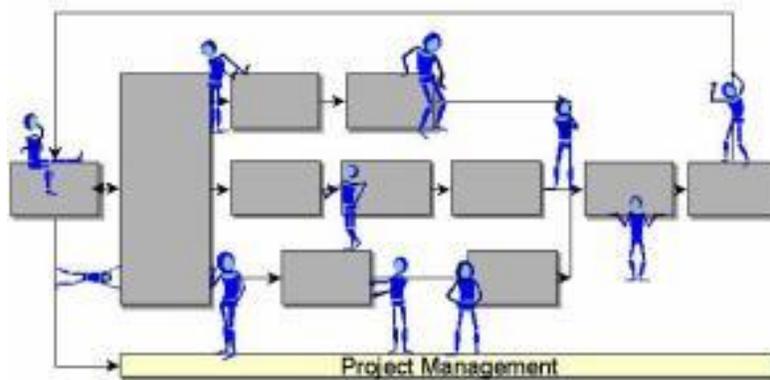


Fig.13. Gerenciamiento del Proyecto. Tomado de [27]

2.1.43 Metodología Bill Inmon

El enfoque Inmon suele denominarse como una metodología “Top Down” debido que se centra primeramente en una visión global de la organización, es decir, divide partes llamados sets de datos departamentales. Los DataMarts de la organización están conectados con el almacén de datos que son extraídos desde los sistemas operacionales, los cuales son cargados y consolidados en un data warehouse corporativo los cuales serán distribuidos a los diferentes datamarts de cada unidad del negocio para satisfacer los requerimientos de la organización. Para complementar este concepto se

dice que para Inmon la prioridad es que el modelo de datos este construido en tercera forma normal.



Fig.14. Fases de la Metodología Inmon. Tomado de[23]

La metodología de Bill Inmon se divide en tres componentes:

1) Desarrollo de los Sistemas Operacionales: Este componente comprende las siguientes etapas:

- **Actividades Iniciales del Proyecto:** Se obtiene los requerimientos del sistema a través de entrevistas, recopilación de datos, análisis del plan estratégico del negocio y de los requerimientos de los sistemas existentes para los nuevos sistemas BI
- **Uso de las Aplicaciones y Datos Existentes:** Se toma las aplicaciones y los datos existentes para formar los futuros componentes dentro del

proyecto que se lleva a cabo.

- **Determinación de Tamaño y Fases:** Una vez que se obtiene los requerimientos se debe determinar su tamaño y dividir el desarrollo del proyecto en fases funcionales y manejables.
- **Formalización de los Requerimientos:** Se asegura que todos los requerimientos sean completos, organizados, comprensibles y que tengan un nivel a mayor detalle que permita que estos sean efectivos. Para ello se debe tener en cuenta lo siguiente:

Modelado de Datos Comprende a:

- a) **Diagrama de Entidad Relación:** A partir de la especificación de los requerimientos, se debe identificar los procesos de mayor nivel que conformara el nuevo sistema con sus respectivas relaciones y cardinalidad
- b) **Conjunto de Elementos de Datos:** Cada proceso de negocio es dividido en conjunto de elementos de datos, es decir agrupar las tablas y los registros que representan el proceso de negocio. Un conjunto de elementos de datos contiene los atributos de los datos, agrupamiento de los atributos, claves, tipos de datos, entre otros.,
- c) **Diseño físico de la Base de Datos:** Se elaboran las tablas y bases de datos.

2) Desarrollo del DataWarehouse

El almacén de datos se encarga del desarrollo de sistemas y procesamiento de soporte a la toma de decisiones donde este comprende las siguientes etapas:

- **Análisis del Modelo de Datos:** Se verifica que el modelo de datos de la organización es consistente y que contiene los procesos de negocios prioritarios, cada proceso tiene su propia definición de datos tales como atributos, relaciones debidamente definidas, claves primarias, etc.
- **Análisis Bread box:** Este análisis consiste en realizar un tamaño estimado sobre el entorno del sistema de soporte de toma de decisiones, el cual proyecta la cantidad de datos que almacenara el DW, esto ayudara a definir si es necesario considerar múltiples niveles de granularidad. Es decir, el nivel de detalle que se almacenara la información del negocio que se está analizando.
- **Valoración Técnica:** Contiene definiciones técnicas que poseen la habilidad de manejar grandes cantidades de datos, permite organizar los datos de acuerdo al modelo de datos definido previamente en el diseño.
- **Preparación del Entorno Técnico:** En esta etapa se instala, ubica y desarrolla de los componentes técnicos que recibirán los datos: la red, el sistema operativo, la interfaz desde y hacia el data warehouse, el software que administrara el data warehouse.
- **Diseño del Datawarehouse:** Se diseña el data warehouse cumpliendo los requisitos de acoplamiento de los diferentes niveles de granularidad, orientación de los datos a los principales temas de la empresa, Variabilidad de tiempo en cada registro de datos y ausencia de datos que no apoyan

los sistemas de soporte de decisiones.

- **Análisis de los sistemas fuente:** Identificación de los sistemas de registro, se realizan los mapeos de los datos del ambiente operacional al ambiente analítico.
- **Especificaciones:** Descripción del entorno donde se permitirá la extracción e integración de los datos de forma eficiente y simple.
- **Programación:** Se fijan los programas que se utilizaran para realizar el proceso de extracción, transformación e integración de datos.
- **Población:** Consiste en la carga de datos para obtener el almacén de datos.
- **Repetición del Desarrollo Estándar:** Para conseguir los reportes estándares, el procesamiento analítico debe seguir las etapas de desarrollo de los sistemas operacionales tomando como fuente de datos el DW.
- **Determinación de los datos necesarios:** Se debe elegir los datos para los análisis necesarios que utilizarán para cubrir los requerimientos de los reportes.
- **Programas para extraer los datos:** Programas que se pueda acceder a los datos que se encuentran en el almacén de datos.
- **Análisis de datos:** Se examinan los resultados obtenidos para asegurar que satisfagan las necesidades de los usuarios y el reporte final

2.1.44 Metodología Hefesto

La metodología Hefesto según “se fundamenta en una amplia investigación, comparación de metodologías existentes y experiencias propias en procesos de confección de

almacenes de datos”. En la figura 17 se puede apreciar los pasos que se deben de abordar en cada fase de la metodología.

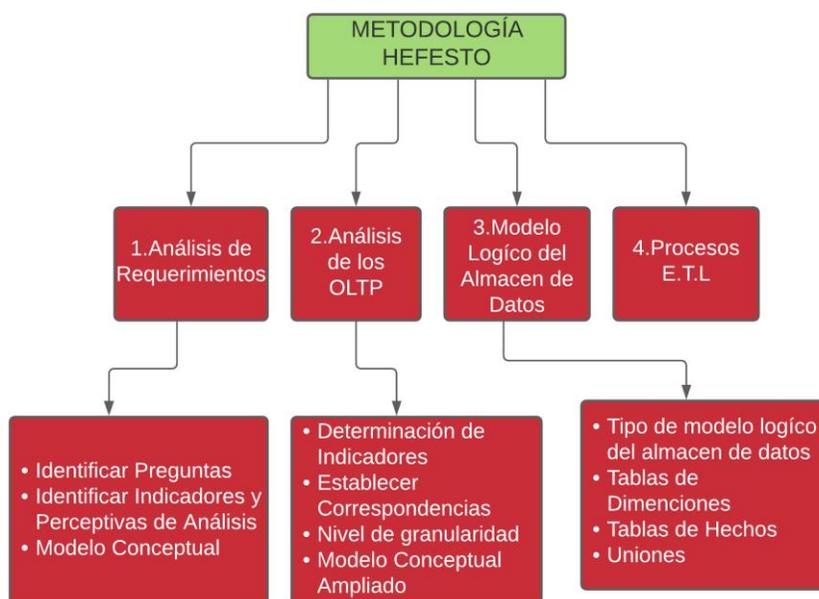


Fig.15. Componentes de la metodología Hefesto

Fuente: Elaboración propia

Esta metodología consta de cuatro fases:

- 1. Análisis de requerimientos:** en este paso se identifica los requerimientos de los usuarios a través de preguntas que explican los objetivos de la empresa, se realiza con el fin de identificar los indicadores y perspectivas que serán tomados en cuenta para construir el almacén de datos. Es de suma importancia indicar que esta metodología sirve para construir un almacén de datos o un DataMart, por ejemplo, si se construye dos DataMart se aplicaría la metodología dos veces.

2. Análisis de los OLTP: se analizarán las fuentes OLTP para determinar cómo serán calculados los indicadores y para establecer las respectivas correspondencias entre el modelo conceptual creado en el paso anterior y las fuentes de datos. Luego, se definirán qué campos se incluirán en cada perspectiva. Finalmente, se ampliará el modelo conceptual con la información obtenida en este paso.

3. Modelo lógico del almacén de datos:

Se confeccionará el modelo lógico de la estructura del DW, teniendo como base el modelo conceptual que ya ha sido creado. Para ello, primero se definirá el tipo de modelo que se utilizará y luego se llevarán a cabo las acciones propias al caso, para diseñar las tablas de dimensiones y de hechos. Finalmente, se realizarán las uniones pertinentes entre estas tablas.

4. Proceso ETL: Primero se cargarán los datos de las dimensiones y luego los de las tablas de hechos, teniendo en cuenta siempre, la correcta correspondencia entre cada elemento. En el caso en que se esté utilizando un esquema copo de nieve, cada vez que existan jerarquías de dimensiones, se comenzarán cargando las tablas de dimensiones del nivel más general al más detallado. Concretamente, en este paso se deberá registrar en detalle las acciones llevadas a cabo con los diferentes softwares.

2.2 ANTECEDENTES

2.2.1 Applications Guide Hdm-4 Volume Two (Guía De Aplicaciones)

Es una guía de tareas que describen ejemplos típicos de diferentes tipos de análisis, y esta será utilizada por personas que deseen tener conocimiento en el desarrollo de una tarea o un estudio[36].

Ventajas: Describe conceptos generales de ejemplos para diferentes tipos de análisis en la elaboración de un proyecto.

Desventajas: El documento no contiene información de las características que se requieren para la elaboración de proyectos asociados a la Inteligencia de Negocios.

2.2.2 Gmm-Bi: Una Guía Metodológica Para La Madurez Organizacional En Inteligencia De Negocios

Para el desarrollo de esta metodología propuesta, según el estudio realizado se tomó como referencia el modelo de madurez de capacidades de inteligencia empresarial, este modelo define un marco que se utiliza para medir, analizar e implementar mejoras de madurez en una organización. El modelo MEI se eligió porque es el único con una descripción explícita de los elementos del nivel KPA, que hace referencia a las actividades y prácticas las cuales permiten alcanzar las metas fundamentales en una organización[37].

Ventajas: una de las ventajas de este modelo es que no requiere de una estructura rígida, por tanto, es variante en las actividades, por lo que depende solo de la capacidad adaptiva de la organización.

Desventajas: el modelo solo se refiere a actividades, y no establece un procedimiento lineal en el cual una organización pueda guiarse para establecer el desarrollo de procesos.

2.2.3 Diseño De Modelo Tecnológico Para El Desarrollo Y El Uso De Big Data En El Análisis Y Visualización De Información Para Pequeñas Y Medianas Empresas.

El objetivo de esta investigación fue diseñar un modelo tecnológico que permitiera la incorporación sistemática de las PYMES al comercio electrónico, la propuesta fue diseñada por medio de la implementación de la tecnología Open Source por lo que se

desarrolló una metodología que facilita la implementación de las herramientas Big Data en las PYMES, en los cuales se incluye una serie de pasos que deben tomar en cuenta para la utilización de estas herramientas de análisis[38].

Ventajas: si se aplica el modelo propuesto, la empresa podrá tener conocimiento de patrones conductuales de los potenciales clientes. Es decir, la empresa podrá conocer un poco a sus clientes y tomar decisiones estratégicas para capturar la atención del cliente.

2.2.4 Una Metodológica De Toma De Decisiones Basadas En Inteligencia Empresarial Y Orientada A Objetivos.

Esta investigación presenta una metodología para la toma de decisiones basada en inteligencia empresarial y orientada a objetivos, estableciendo una metodología interactiva por lo que les permite a las empresas empezar con el descubrimiento de datos necesarios para la construcción de sus modelos, capturar objetivos de las partes interesadas la modulación de amenazas y oportunidades[27].

2.2.5 Marco De Gestión Del Conocimiento Utilizando Arquitectura Empresarial E Inteligencia Empresarial.

Este marco KM (knowledge Management) fue diseñado como una guía práctica para capturar utilizar y transferir conocimiento, por tanto, este marco es destinado a establecer un orden en el proceso de gestión del conocimiento. Con este marco práctico se puede ayudar a los altos directivos a la creación, captura, y digitación de conocimiento para la toma de decisiones[39].

Desventajas: el marco km no especifica las herramientas tecnológicas que se deben de utilizar para la elaboración de una arquitectura BI.

2.2.6 Business Intelligence (Una Guía Práctica):

Según lo expuesto por el autor del libro “describe la metodología, planificación de las iniciativas de un proyecto BI, los análisis de requerimientos, arquitectura y modelamiento

dimensional de datos y los procesos de extracción, transformación y carga (ETL) que se abordan en la implementación”[40].

Ventajas: este libro contiene conceptos de información útil para implementación de la arquitectura BI.

Desventajas: una de las desventajas es que el autor en el libro no hace comparaciones y análisis con datos de casos reales con los que se comparen las diferentes herramientas para el tratamiento de datos.

2.2.7 Diez Pasos Para Incrementar El Conocimiento Para Una Gestión Eficiente Del Capital Intelectual En La Inteligencia Empresarial.

Este trabajo propone una metodología, modelo basado en diferentes tipos de conocimiento que se utiliza durante el proceso de desarrollo. Como es el conocimiento táctico que es el que las personas poseen, pero no se describe en documentos, ni en ningún lugar lo que significa que solo reside en la cabeza. También se describe el conocimiento explícito que hace referencia al que se registra de alguna manera y por ende está disponible para las personas. El conocimiento de las organizaciones se compone por el conocimiento compartido por cada individuo, esto el con el fin de incrementar el conocimiento en las empresas para una gestión eficiente del conocimiento intelectual. Por lo que se define diez pasos como son: Obtener y utilizar, aprender, aportar, evaluar, sostener, apoyar e intercambiar[41].

2.2.8 Una Inteligencia De Negocios En Cinco Capas (Arquitectura)

Este estudio propone un marco de arquitectura BI que consta de 5 capas esenciales que son: Fuentes de datos, proceso ETL, transformación de datos, capas de almacén de datos y usuario final. Estas capas son fundamentales para garantizar el flujo de información, por tanto, estas fueron implementadas para realizar el marco propuesto.

Desventajas: una de las desventajas es que la propuesta realizada en este marco de arquitectura solo se describe como forma conceptual, por lo tanto, este marco debe ser validado aún, utilizando casos reales para afirmar su uso[42].

CAPITULO III. METODOLOGIA

Para el desarrollo de la propuesta de esta investigación será necesario utilizar un enfoque mixto, por lo que se llevará un proceso de recolección, análisis y la interpretación de datos.

Según [43] “en la ruta mixta se utiliza evidencia de datos numéricos, verbales, textuales, visuales, simbólicos y de otras clases para entender problemas en la ciencia”.

Según [44] “este tipo de estudios sirve para familiarizarnos con fenómenos relativamente desconocidos, obtener información sobre la posibilidad de llevar a cabo una investigación más completa respecto de un contexto particular, investigar nuevos problemas, identificar conceptos o variables promisoras establecer prioridades para investigaciones futuras o sugerir afirmaciones o postulados”.

Un proyecto basado en Inteligencia de Negocios suministra a cualquier organización el poder de organizar y administrar la información de una manera rápida y oportuna.

Este proyecto tiene como fin el desarrollo de una guía metodológica que sirva de apoyo para la elaboración de una arquitectura de inteligencia de negocios, en esta se detallara los procesos, métodos, componentes y etapas para implementar una solución de BI.

A continuación, se describe los componentes y etapas de la guía metodológica

- Caracterización de la Organización
- Metodologías de Desarrollo para la construcción de Almacén de Datos
- Descripción de Componentes de la Guía Metodológica

3.3.1 Caracterización de la organización

Esta parte del documento tiene finalidad presentar teóricamente el desarrollo de una guía metodológica basada en los componentes de una Arquitectura de Inteligencia de

Negocios para ello, se realizó una investigación en diferentes fuentes de información sobre la variedad de metodologías y modelos más comunes para la creación e implementación de una arquitectura BI, se toma como referencia estas metodologías y sus modelos con el fin de indicar y describir los componentes de la arquitectura.

De acuerdo, a los modelos que presenta cada metodología es importante clasificar el tamaño de la organización. Si esta se clasifica en microempresa, pequeña empresa, mediana empresa y gran empresa.

En la siguiente figura se ilustra el rango de clasificación de las empresas.

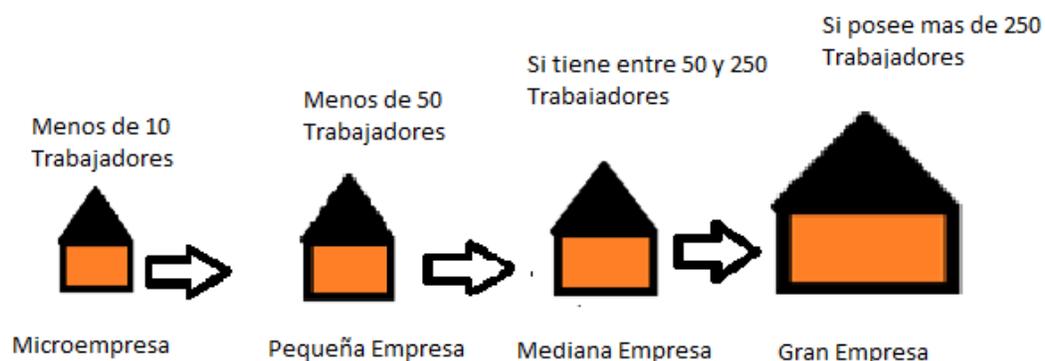


Fig.16. Clasificación de las empresas de acuerdo con nivel del personal

Fuente: Elaboración Propia

Ahora bien, teniendo en cuenta la clasificación de las empresas esta guía se desarrolla para ser implementada en organizaciones que manejan gran número de trabajadores y cuentan con múltiples procesos dentro del organigrama empresarial. De acuerdo con este contexto es importante indicar las ventajas y desventajas que poseen las grandes organizaciones para así mismo plantear las necesidades de esta.

Ventajas:

- Poseen gran capacidad de inversión
- Transformación de conocimientos y tecnología
- Capacidad para definir y cumplir objetivos a mediano y largo plazo
- Capacidad de negociación entre proveedores y financiadores
- Capacidad para afrontar situaciones de crisis
- Capacidad de mano de obra

Desventajas:

- Lentitud en la toma de decisiones (escasa flexibilidad)
- Estructuras más rígidas y complicadas
- Dificultad de coordinación e integración de todos los estamentos de la empresa
- Políticas más rígidas en cuanto al manejo de la información

A partir de las desventajas que poseen las grandes organizaciones se plantea el uso de técnicas y metodologías con el fin de contribuir a la mejora de los procesos estructurales por ejemplo estas permiten lo siguiente:

- Reducir actividades por procesos
- Facilitar el acceso y el intercambio de información
- Facilitar la toma de Decisiones
- Permite obtener información valiosa sobre el comportamiento del cliente, compras o ventas
- Permite ofrecer resultados concretos y precisos
- Permite realizar análisis en tiempo óptimos

3.3.2 Metodologías de Desarrollo de Almacén de Datos

En este punto se realizó una búsqueda de las metodologías de Kimball, Hefesto y Bill Inmon las cuales se tomaron como referencia para indicar los componentes de la guía metodológica, en cuanto a la revisión de estas se encontró que proponen un marco de trabajo a mayor detalle, es decir, se requiere diseñar cada artefacto siguiendo unos lineamientos para dar cumplimiento de los objetivos de cada fase.

A continuación, se describen las fases de cada una de las metodologías revisadas

Metodología	Fases
Kimball	<ul style="list-style-type: none">• Planificación del Proyecto• Definición de Requerimientos• Modelo Dimensional• Diseño Físico• Diseño e Implementación (Proceso E.T.L)• Mantenimiento y Crecimiento de DW• Especificaciones de Aplicaciones de BI• Diseño de la Arquitectura Técnica
Hefesto	<ul style="list-style-type: none">• Análisis de Requerimientos • Análisis de los OLTP• Modelo Lógico de DW• Integración de Datos
Bill Inmon	Esta metodología es similar a la planteada por Kimball, la diferencia está en esta se necesita realizar un proceso de normalización, es decir, en el modelo de datos se crea la normalización en tercera forma normal lo que se significa que se realiza un despliegue a las tablas de modelo.

Tabla 2. Tabla comparativa de metodologías para desarrollo de DW

Fuente: Elaboración Propia

Es importante indicar que al momento de elegir la metodología para desarrollar un almacén de datos se deben tener conocimiento sobre los aspectos que brinda cada metodología, es decir, si esta se acopla al entorno del equipo desarrollador y estructura organizacional. En la siguiente tabla se realiza la comparación de las metodologías mencionadas.

Metodologías			
Característica	Kimball	Hefesto	Bill Inmon
Presupuesto	Coste inicial bajo	Coste Inicial Bajo	Coste inicial alto
Plazos	Tiempo de desarrollo inferior	Se realizan en mediano y largo plazo	Requiere más tiempo de desarrollo
Experiencia	Equipo con especialización alta	Equipo con especialización media	Equipo con especificación media
Alcance	Departamentos individuales	Departamentos Individuales	Toda la compañía
Mantenimiento	Mantenimiento más complejo	Fácil Mantenimiento	Fácil mantenimiento

Tabla 2: Tabla comparativa de metodologías para el desarrollo de un DW

Fuente: Elaboración Propia

3.3.3 Descripción de las etapas de la Guía Metodológica

El desarrollo de guía parte del análisis realizado sobre los modelos de metodologías existentes mencionadas anteriormente, la cual se toman como referencia para describir cada componentes o etapa.

A continuación, se describe cada componente:

- **Planificación del Proyecto:** Es el primer paso para iniciar un proyecto de Inteligencia de Negocios, este paso consiste en comprender e identificar los escenarios del negocio en los cuales se ejecutará el proyecto.
- **Recolección de Requerimientos:** Este paso consiste en poder identificar las necesidades del negocio para poder definir y brindar una solución frente a la situación.
- **Diseño Modelo Dimensional:** El modelo dimensional es una técnica que permite mostrar la información de los procesos del negocio de una forma estandarizada y sencilla.
- **Diseño Proceso E.T.L:** El desarrollo del sistema ETL es muy importante, dado que es la etapa primordial para la elaboración del almacén de datos, donde se extraen los datos de las fuentes de origen, se realiza la transformación en los formatos deseados y el cargue de los mismos.
- **Implementación del Almacén de Datos:** Esta fase en donde se unen las diferentes las herramientas B.I., los datos y las aplicaciones para usuarios finales.

En la figura 17 se indica las etapas de la guía que representa la arquitectura de inteligencia de negocios.

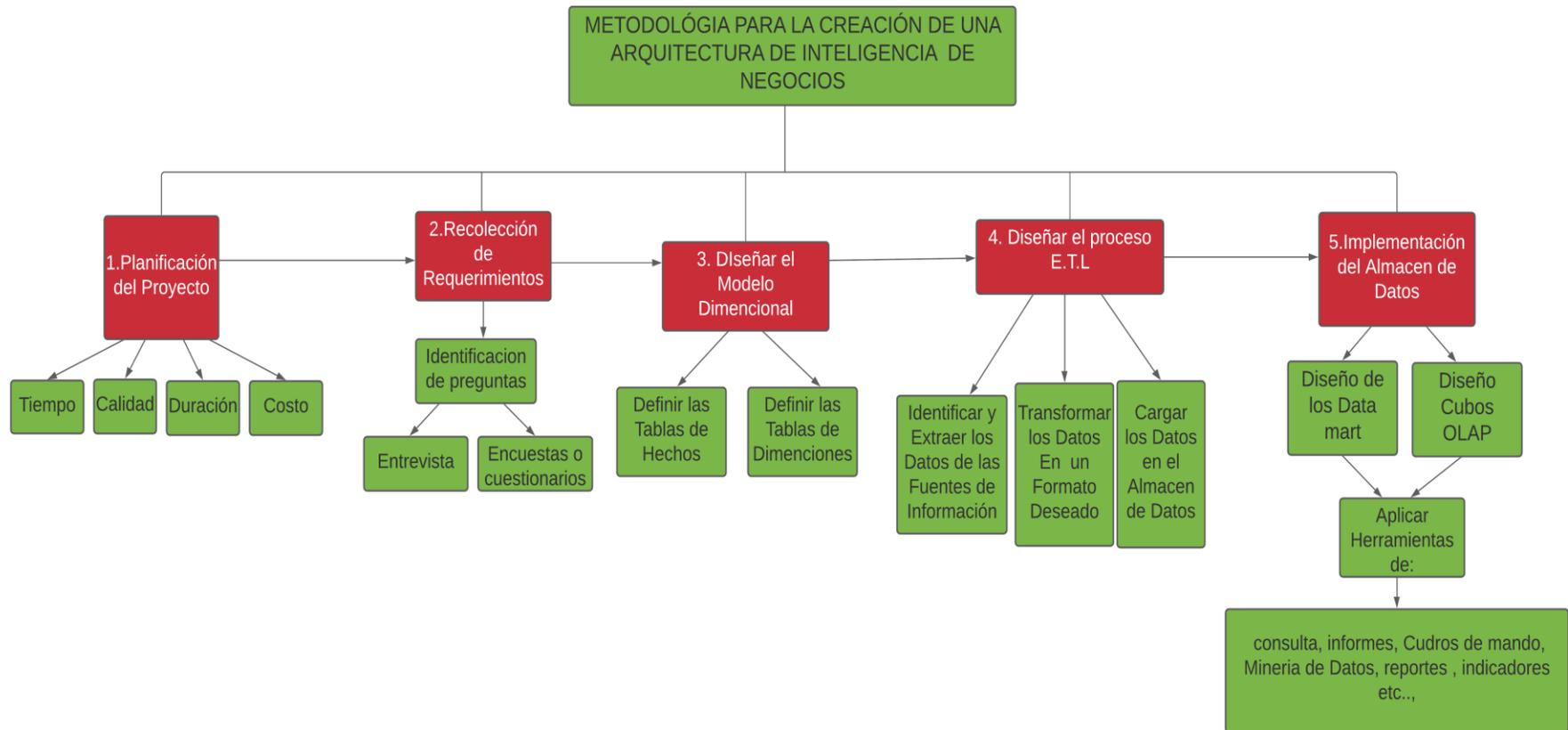


Fig.17. Etapas de una Arquitectura de Inteligencia de Negocios

Fuente: Elaboración Propia

En seguida se describe la forma en el cual se construyó cada etapa de la guía se tiene en cuenta que esta se compone de subfases descritas en la figura 19.

3.3.3.1.1 Etapa 1: Planificación del Proyecto

Es la etapa fundamental para el desarrollo del proyecto, es donde se realiza el acercamiento con el cliente, tiene como objetivo identificar los escenarios y el alcance a través de la justificación y las evaluaciones de factibilidad, en otras palabras, se puede decir que es la visión general del proyecto el cual provee un marco de trabajo en donde se desarrollara el proyecto. Para ello, se debe tener conocimiento sobre la estructura, recursos y procesos operacionales, esto permitirá hacer una estimación sobre los plazos y tiempos de ejecución de las actividades para el alcance del objetivo.

Dentro de la planificación del proyecto se encuentra lo siguiente:

1. Definición del Proyecto
2. Definición de un plan de Trabajo

3.3.3.1.1.1 Definición del Proyecto

En este punto se debe identificar los posibles escenarios para evaluar si es viable el desarrollo del proyecto, se determina la problemática y se establecen los objetivos los cuales deben de ser cuantificados y encaminados de acuerdo con la estrategia de la organización. El proyecto que se desarrolle dentro de la organización debe comprender a que procesos y subproceso de las áreas se realizara el mejoramiento

y, por tanto, se debe definir qué datos soportaran los nuevos modelos que se implementaran. En esta parte, se debe establecer la funcionalidad del producto.

3.3.3.1.1.2 Definición de un Plan de Trabajo

El coordinador del proyecto deberá proporcionar un plan de trabajo para realizar el seguimiento al progreso del proyecto, para la asignación de actividades tareas, recursos, costos y tiempos para el desarrollo del proyecto.

El plan de trabajo comprende lo siguiente:

- Definir las actividades, tareas, subtareas
- Establecer las secuencias de desarrollo de las actividades
- Realizar estimaciones de los horarios por cada actividad, tareas y subtareas
- Asignación de recursos

3.3.3.1.2 Etapa 2: Recolección de Requerimientos

En esta etapa de recolección de requerimientos es el punto de partida para la creación de una Arquitectura de Inteligencia de Negocios debido a que, es la etapa donde se establece la comunicación con el cliente quien tiene conocimiento sobre las reglas del negocio y plantea la necesidad. Se indica que esta etapa es un poco crítica debido que es la base para el desarrollo de las siguientes etapas, por tal motivo es importante definir con claridad los requerimientos del negocio. Durante este proceso no pueden existir errores ya que se podría construir una solución errónea que no aporte productividad a la organización.

Por ejemplo, en la metodología kimball en la fase de definición de los requerimientos del negocio se propone realizar entrevistas o encuestas al personal categorizado en cuatro grupos:

- Directivo quien es responsable de tomar las decisiones estratégicas
- Administradores intermedios responsables de explorar alternativas estratégicas para la toma de decisiones
- Personal tic, que son los encargados de conocer qué tipo de problemas informáticos y datos que existen dentro de la organización
- Personal que se entrevista por cuestiones políticas

Para llevar a cabo la entrevista primero se debe realizar los siguientes pasos: preparar la entrevista, extraer la información de los temas a analizar, agrupar la información para el análisis de los procesos de la organización. Una vez, realizada la entrevista se debe crear un documento donde se especifique claramente los requerimientos que quedaran como soporte sobre la necesidad expuesta por el cliente, partiendo de ello se prosigue a desarrollar las siguientes etapas.

3.3.3.1.2.1 Técnica de Entrevista Para la Recolección de Requerimientos

Esta técnica que permite establecer una comunicación entre el grupo de grupo técnico que lidera el proyecto en la organización y el personal que hacen parte del equipo de trabajo en los procesos y subprocesos de las diferentes áreas o departamentos de la empresa, estas preguntas se deben de realizar con el fin de obtener información relevante para la investigación, y de esta manera tener más recursos para el levantamiento de requerimientos y comprender cuál es la necesidad de la organización.

Tipos de preguntas que se pueden realizar en relación con la organización:

- ¿Cuál es la necesidad de la Organización?
- ¿Qué se resultados se esperan al implementar la arquitectura BI?
- ¿Cuál es estado actual de la organización en cuanto al manejo de la información?
- ¿Qué tipo de datos se manejan en la organización?
- ¿Cuáles son las fuentes de información que maneja la organización?
- ¿Cuál es la estructura que se maneja en las diferentes áreas de la organización?
- ¿Qué dificultad se ha presentado en cuanto al manejo de los sistemas de información?
- ¿Qué tendencias contribuyen hacer el trabajo más fácil o difícil?
- ¿Cuál es la estructura que se maneja en las diferentes áreas de la organización?
- ¿Qué pasos se han llevado a cabo actualmente para satisfacer la falta de información?
- ¿Cuál es el método de distribución de la información? ¿Es un informe de datos por escrito o por correo electrónico, o se trata de otro método?
- ¿Cómo se mide el rendimiento de las funciones?
- ¿Qué pasos se han llevado a cabo actualmente para satisfacer la falta de información?
- ¿Qué tipos de información faltan para realizar el análisis y la toma de decisiones?

3.3.3.1.2.2 Identificación de las Fuentes de Información

En esta etapa se identifica las diferentes fuentes de información con las que cuenta la organización esto se realiza con el fin de analizar los datos y que estos sean útiles para el modelo de negocio y para la toma de decisiones, estas fuentes se definen en:

- Sistemas ERP (sistema de planificación de recursos empresariales)
- Base de Datos
- Ficheros de tipo Excel
- Archivos planos
- Archivos CVS

3.3.3.1.2.3 Selección de la herramienta para proceso ETL

En esta parte se tiene en cuenta la selección de la herramienta con la cual se hará los procesos ETL, la cual servirá para realizar la extracción transformación y carga de datos, se debe tener conocimiento sobre el manejo y aplicaciones.

3.3.3.1.2.4 Construcción de Preguntas para Plantear el Modelo Dimensional

Una vez se realice el levantamiento de requerimientos y de acuerdo con estos se prosigue a realizar las preguntas con las cuales se definirá el modelo de datos, para ello se debe establecer los indicadores y perceptivas que se tendrán de la solución BI. Estos indicadores son los que permiten medir el valor determinado de un proceso, de igual manera las preceptiva es la noción consistente que se tiene para realizar el análisis de una solución. De acuerdo con estos dos conceptos se empieza a construir el modelo para dar respuesta a las preguntas planteadas

3.3.3.1.2.5 Matriz de Requerimientos

Esta parte se diseña con el fin de relacionar cada requisito con los objetivos de proyecto y objetivos estratégicos organizacionales que satisfacen y poder garantizar de esta forma que cada requisito esté agregando valor al presente y futuro del negocio.

3.3.3.1.3 Etapa 3: Diseñar El Modelo Dimensional

La etapa de diseño consiste en crear un modelo dimensional que consiga satisfacer los requerimientos capturados en la etapa anterior y definir los datos que se extraerán de las fuentes de información y la carga en las tablas de destino, este modelo también permite presentar información de los procesos de negocio de una manera más estandarizada optimizando el acceso a los datos.

La etapa del diseño comprende lo siguiente:

1. Identificar el proceso de negocio
2. Identificar las preguntas para construir el modelo
3. Definir las tablas de hechos
4. Definir las tablas de dimensiones
5. Diseñar el modelo lógico
6. Diseñar el modelo físico

3.3.3.1.3.1 Identificar el proceso de negocio

En esta actividad se debe identificar el proceso de la organización al cual se le asignará el modelo dimensional. Según sea la selección se reúnen los requisitos del

negocio, se tiene en cuenta que un proceso de negocio requiere más de un modelo debido a que contiene un conjunto de actividades relacionadas. Por esta razón se debe clasificar y priorizar los requisitos en determinados criterios:

- Calidad de los datos en los sistemas de origen
- Viabilidad
- Complejidad en los procesos de negocio

3.3.3.1.3.2 Definir las Tablas de Hechos

Se define las tablas de hechos que hacen parte del proceso de negocio. Cada tabla debe contener una estructura de claves llamadas foráneas que corresponde a la llave primaria de otras tablas llamadas dimensiones las cuales permiten establecer la relación entre las mismas. Generalmente estas tablas dirigen hechos que contienen casi siempre valores numéricos o indicadores. La segunda parte de la estructura son las medidas que almacena la información específica del negocio, estas recopilan datos cuantitativos como por ejemplo peso, altura y edad.

La tabla de hechos contiene lo siguiente:

- La tabla principal del modelo dimensional
- Contiene campos de la clave primaria de otras tablas llamadas dimensiones
- Contiene métricas sobre las medidas que queremos analizar (por ejemplo, valores numéricos)
- Evita redundancia de datos
- Contiene millones de registros

A continuación, se ilustra un ejemplo de lo que sería una tabla de hecho relacionada con las dimensiones sobre un proceso de negocio ver figura 18. Es importante indicar que para el diseño de estas tablas existen dos esquemas el de tipo estrella y copo de nieve.

- **Esquema estrella:** Este esquema consta de una tabla central llamada hechos, está relacionada con otras tablas llamadas dimensiones, las ramificaciones que se conectan con la tabla central indican una relación de muchos a uno. Ver figura 18.
- **Esquema copo de nieve:** consta de una o varias tablas de hechos conectada a tablas de dimensiones su relación es de muchos a uno, su principal característica es que está normalizada en tercera forma normal, es decir cada tabla está representada por un nivel de jerarquía, este esquema se recomienda cuando se requiere reducir el espacio de almacenamiento y eliminar renuncia de datos. Sin embargo, este puede ocasionar malos rendimientos debido al alto nivel de relación con la gran variedad de tablas dimensionales. En figura 19 se muestra la representación del esquema copo de nieve el cual las tablas de dimensiones están compuestas por varios niveles.

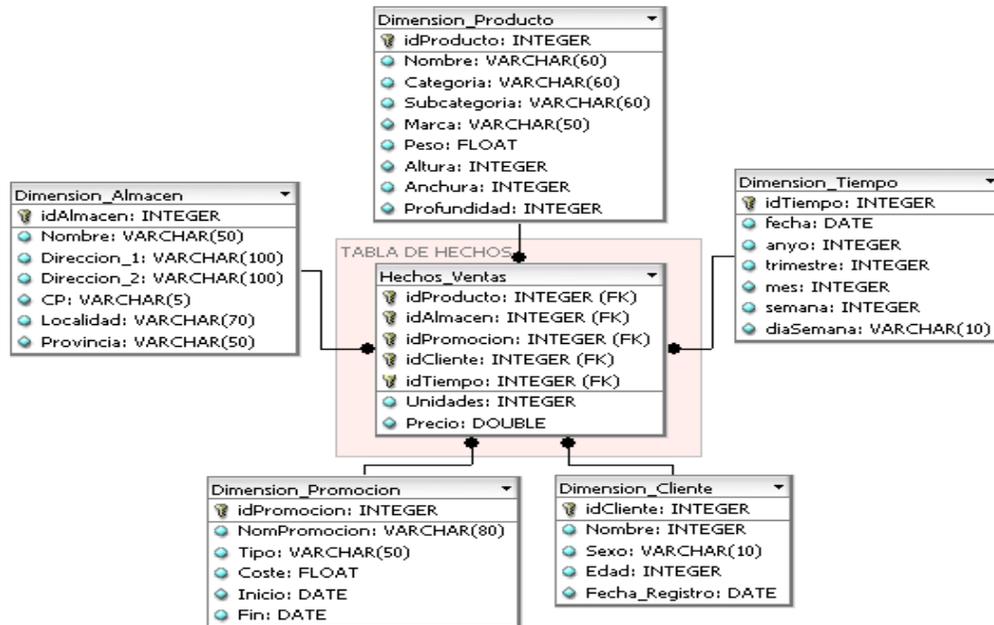


Fig. 18. Esquema en estrella. Tomado de [45]

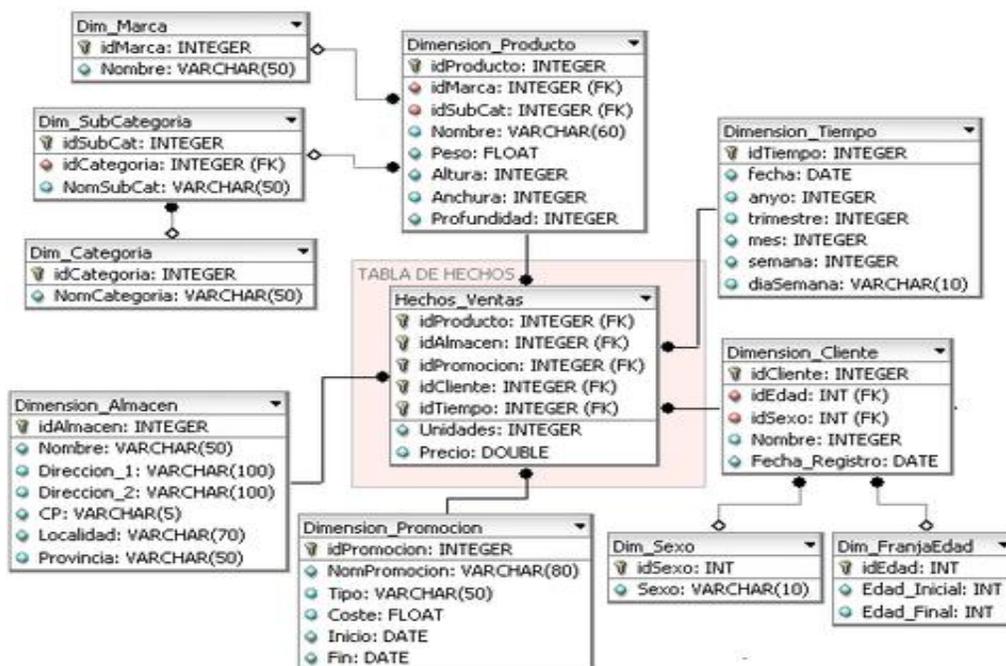


Fig.19 Esquema copo de nieve. Tomado de [45]

3.3.3.1.3.3 Definir las Tablas de Dimensiones

En este punto se define las tablas de dimensiones las cuales se van a relacionar con las tablas de hechos de acuerdo con las reglas del negocio, su estructura se puede organizar por medio de un nivel jerárquico en el cual los datos pueden verse en diferentes niveles, las tablas poseen un identificador único y unos atributos que describe criterios de análisis relevantes para la organización y por lo general estos siempre son de tipo texto sus características son las siguientes:

- Tablas en primera forma normal
- Contiene un identificador único (llave primaria) que se une a la tabla de hechos
- Los atributos de la tabla dimensión brinda información características de las tablas de hechos.
- Se puede definir más de una tabla de dimensión
- Las dimensiones pueden estar en un primer nivel o se puede desglosar a otros niveles

3.3.3.1.3.4 Diseñar el Modelo Lógico

Esta etapa se diseña con el objetivo de crear la estructura lógica del almacén de datos la cual soportara las consultas. Para elaborar este modelo se inicia con los indicadores de gestión que se han definido en las fases anteriores lo cual finaliza con la representación multidimensional de las variables que abarca cada indicador, para facilitar el proceso de modelado se puede hacer la representación gráfica en el esquema estrella o copo de nieve esto depende de los requisitos del negocio.

3.3.3.1.3.5 Diseñar el Modelo Físico

En el modelo físico se diseña con el fin de crear una estructura física que soporte el diseño lógico el cual será la base de implementación del proyecto, es aquí donde se define todos los componentes de las tablas como atributos, descripción de atributos, tipo de datos, claves primarias, foráneas, e índices y demás estructura física que representara la información, demás se establece las fuentes donde se va a extraer los datos para cada atributo. Es este proceso se conoce como mapeo de datos.

3.3.3.1.4 Etapa 4: Diseñar el proceso E.T.L

El desarrollo del sistema ETL es muy importante, dado que es la etapa primordial para la elaboración del almacén de datos, y por ende se debe garantizar que el tratamiento de los datos sea consistente y que la calidad de estos sea adecuada para el proceso. En el proceso E.T.L se definen que datos que tomaran de las diferentes fuentes de información, como se manipularan y en qué lugar se almacenaran estos procesos son los que soportaran el diseño físico.

Para realizar el proceso E.T.L se debe realizar lo siguiente:

- 1. Extracción de los Datos:** Este paso consiste en realizar una extracción de datos de las diferentes fuentes de información que maneja la organización, estos pueden ser proporcionados por medio de bases de datos, consultas, archivos de texto, sistemas ERP, sistemas contables, hojas de cálculo entre otros, fuentes de información externa, archivos planos etc...., los datos deben de pasar un proceso de preparación para llevar a cabo la integración de estos. Una vez se haya consolidado los datos críticos, se observará que la

estructura de los datos extraídos de las diferentes fuentes es heterogénea, por la misma razón estos datos ya compilados se deben organizar de acuerdo con la fecha, el tamaño y la fuente para adaptarse al proceso de transformación. La complejidad de este proceso puede variar según el tipo de dato, las fuentes de datos y el volumen de estos

2. Transformar los datos en formatos deseados: este punto se dice que es el más crítico debido a que se debe aplicar una serie de normas y funciones necesarias para convertirlos en un formato estandarizado el cual cumpla con los requisitos del esquema de BD de destino con los requisitos del negocio y posteriormente ser cargados. El proceso E.T.L en un almacén de datos implica lo siguiente:

- los datos se deben de convertir de acuerdo con los requerimientos del negocio.
- Realizar el proceso de formateo para establecer la compatibilidad

3. Cargar los datos en el Almacén de Datos: En este punto la información que proviene de fase anterior es cargada en el sistema destino dependiendo de los requerimientos de la organización, este proceso puede abarcar una amplia variedad de acciones. Por ejemplo, hay dos formas de realizar este paso la primera es por medio inserción SQL esto implica que el registro se hace manualmente en cada fila correspondiente a la tabla la otra forma es la carga masiva de datos con las herramientas que brinda el mercado actual para carga de datos.

En este paso se diseña las herramientas que permitirán al usuario final realizar consulta y análisis a los datos que se encuentran en el almacén de datos.

3.3.3.1.4.1 Definición de Pruebas

Se dice, que en cualquier proyecto se debe ejecutar pruebas al sistema. De acuerdo con esto se define los pasos para realizar este proceso:

- **Pruebas de procesos E.T.L:** estas se realizan para validar los procesos E.T.L los cuales deben cumplir con la funcionalidad de los requerimientos establecidos en el diseño.
- **Realizar pruebas al Data Warehouse:** estas pruebas consisten en validar el nuevo Almacén de datos si ha sido elaborado de acuerdo con los requerimientos y diseño establecido.
- **Pruebas de reporte:** el objetivo de estas es verificar si los reportes creados cumplen con los requerimientos definidos en las fases anteriores.

3.3.3.1.5. Etapa 5: Implementar El Almacén De Datos

Una vez se realice la etapa de pruebas se prosigue a efectuar la ejecución del almacén de datos, en el cual se implementará las herramientas B.I y las aplicaciones para usuarios finales y otros factores externos tales como la capacitación a usuarios, soporte técnico, etc., esto permitirá el despliegue del proyecto de inteligencia de negocios.

En seguida se formula una serie de instrucciones que garantizaran un proyecto exitoso y un producto de calidad.

- Se debe definir los planes y procedimientos de soporte para los

usuarios.

- Elaboración de entorno de prueba para el manejo de datos.
- Definición de niveles de soporte (a nivel de aplicación, modelo y calidad de datos).
- Proveer una documentación apropiada, para obtener un conocimiento adecuado sobre el proyecto
- Todos los componentes deben estar probados correctamente de principio a fin.

Es importante añadir que la lógica de negocio puede ser dirigido a todas las áreas que conforman la estructura organizacional o puede ser enfocada a ciertas áreas esto depende de los requerimientos que se definan. Se puede crear un sistema llamado DataMart el cual se orienta a un área en específico de la organización. De igual manera para el desarrollo de este sistema se debe ejecutar todas las fases mencionadas anteriormente.

3.3.3.1.5.1 Documentación Final

En esta parte hace referencia a la documentación de todos los procesos desarrollados por cada fase, se debe tener en cuenta las características y funciones de cada procedimiento realizado en cuanto a:

- Planificación del proyecto
- Levantamiento de requerimientos
- Diseño de modelo dimensional
- Diseño de proceso ETL

- E implementación del almacén de datos

3.2 Aplicación de la Guía en la empresa Asociación Indígena del Cauca - EPS

La Asociación Indígena del Cauca EPS-I es una entidad pública de carácter especial que tiene como objeto fortalecer la capacidad organizativa de los procesos en salud de los pueblos indígenas y demás población afiliada, a través de la administración de los recursos y el aseguramiento de los servicios dentro del Sistema General de Seguridad Social en Salud. En el transcurso de su existencia la organización acumula grandes volúmenes de datos los cuales incluyen información de procesos y subprocesos que se ejecutan en las diferentes áreas (ver figura 21).

Justificación Proyecto Datamart en la Asociación Indígena del Cauca

La Asociación Indígena del Cauca EPS-I mira el valor de implementar herramientas tecnológicas que permitan hacer análisis a los datos almacenados que se encuentran en sus sistemas de información, en los cuales manejan múltiples procesos.

Ahora bien, debido a que la empresa cuenta con un alto volumen de datos, la organización requiere diseñar sistemas que permitan automatizar los procesos.

En este punto, la intención del proyecto tiene como finalidad brindar a la organización una solución de inteligencia de negocios donde se pueda realizar el análisis y seguimiento sobre la cantidad de los procesos que reportan a la Base de Datos Única de Afiliados. El análisis parte donde los datos se encuentra en un nivel de jerarquía ya estructurados en los sistemas de origen, para realizar los procedimientos correspondientes para cada etapa de diseño de la solución.

Con la implementación de la solución el coordinador técnico podrá tomar decisiones en cuanto al resultado de la evolución del dinamizador encargado de reportar las diferentes novedades de los afiliados que se realizan en los diferentes municipios que tiene cobertura el AIC

Para el diseño e implementación del datamart o solución se tuvo en cuenta las etapas descritas en la guía metodológica.

A continuación, se describen las etapas que se desarrollaron:

- Planificación del Proyecto
- Recolección de Requerimientos
- Diseño Modelo Dimensional
- Diseño Proceso ETL
- Implementación del Datamart

PLANIFICACIÓN DEL PROYECTO

Para realizar la planificación del proyecto y determinar el alcance, justificación y objetivos fue necesario establecer reuniones con el equipo de trabajo encargado de gestionar los procesos BDUA del área de Aseguramiento y Gestión Tic que hacen parte de la estructura organizacional de la Asociación Indígena del Cauca ver figura 20.



Fig.20 Estructura Orgánica de la Asociación Indígena del Cauca – EPSI

Fuente: Asociación Indígena del Cauca

De este modo se tuvo en cuenta la información otorgada por equipo de trabajo para realizar la estimación del presupuesto del proyecto.

PRESUPUESTO				
ITEM	CANTIDAD	DESCRIPCION	VALOR UNITARIO	VALOR TOTAL
PAGOS DE SERVICIOS EXTERNOS				
1	6 meses	Internet	\$550000xmes	\$330.000
2	6 meses	Servicio de Agua	\$30000xmes	\$180.000
3	6 meses	Servicio de energía	\$30000xmes	\$180.000
SUBTOTAL				\$690.000
EQUIPO DE LABORATORIO Y CAMPO				
4	1 semana	Realización de reuniones y consultas con el personal del área aseguramiento y gestión tic	\$30.000Xdía viáticos y otros	\$200.000
SUBTOTAL				\$200.000
INSUMO DE LABORATORIO Y CAMPO				
5	1	Plan de datos y minutos	\$50.000	\$50.000
6	1	Motorola e ^{s6}	\$420.000	\$420.000
7	1	Computador Lenovo Core i7	\$180.0000	\$180.0000
SUBTOTAL				\$ 2.270.000
INSUMO DE OFICINA				
8	1	Bolígrafo	\$1.000	\$1.000
9	1	Lápiz	\$1.000	\$1.000

10	1	Cuaderno de apuntes	\$2.000	\$2.000
SUBTOTAL				\$4.000
PASAJES Y VIÁTICOS				
12	6 meses	Pasajes desde Morales Cauca hasta Popayán oficinas AIC	\$30.000xdia	\$3600.000
13	6	Insumos básicos	\$100.000	\$100.000
SUBTOTAL				\$370.0000
TOTAL, PROYECTO				\$6.864.000

Tabla 3: Presupuesto Del Proyecto Construcción Del Datamart

DEFINICION DEL PROYECTO

Para determinar el alcance del proyecto se realizó a cabo un estudio sobre el manejo de los procesos los cuales se definen de la siguiente manera, para el proceso de reportes que rige para los regímenes subsidiado y contributivo se genera por año un proceso que es la administración de la BDUA en la cual se encuentra los diferentes procesos como malla de validación, inconsistencias, archivos SAT, depurados y novedades BDUA entre otros. De acuerdo, con lo anterior se indica que el proyecto se enfoca en las novedades BDUA donde estos deben ser reportados semanalmente, mensual y anual.

A continuación, se describen las novedades de los afiliados que se deben reportar a la BDUA

- Archivo S1(Traslados – movilidad): Es el proceso de traslado que solicita el afiliado de una EPS a otra, y el proceso de movilidad hace referencia al cambio del régimen subsidiado al contributivo o viceversa dentro de la misma EPS.
- Archivo MS (Nuevas afiliaciones): las EPS deben reportar los datos del ingreso en la estructura del archivo maestro de afiliados para el régimen que aplique.
- Archivo NS (Novedades de las Afiliados): Las entidades deben mantener actualizada la información de sus afiliados mediante el envío de novedades.

Para entender el proceso de cargue de los reportes sobre las novedades de los afiliados se crea los siguientes flujogramas que corresponden a cada novedad.

En las figuras 21 se describe el flujograma del proceso S1 y en la figura 22 se describe el flujograma de los procesos MS Y NS.

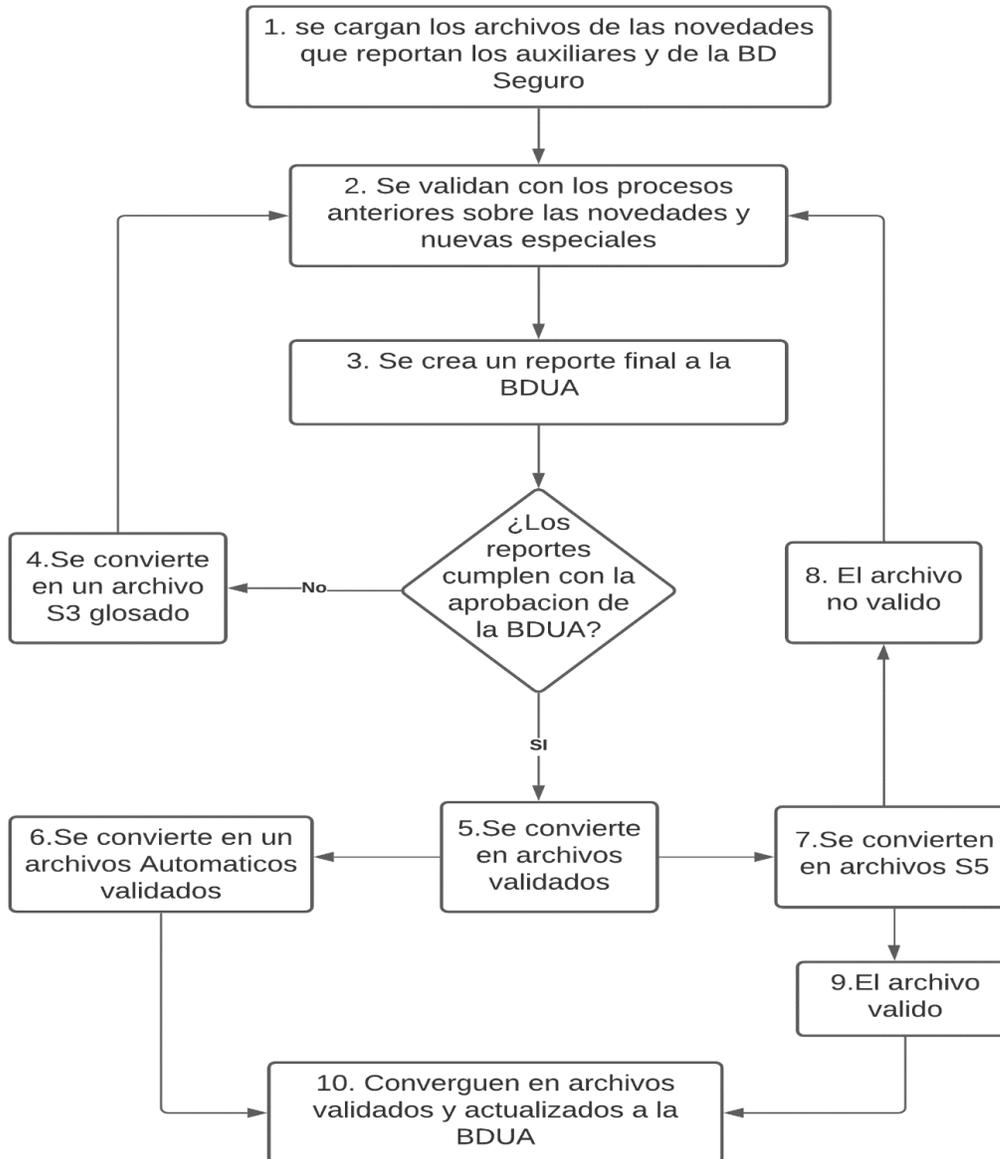


Fig. 21 Flujograma Proceso novedad S1

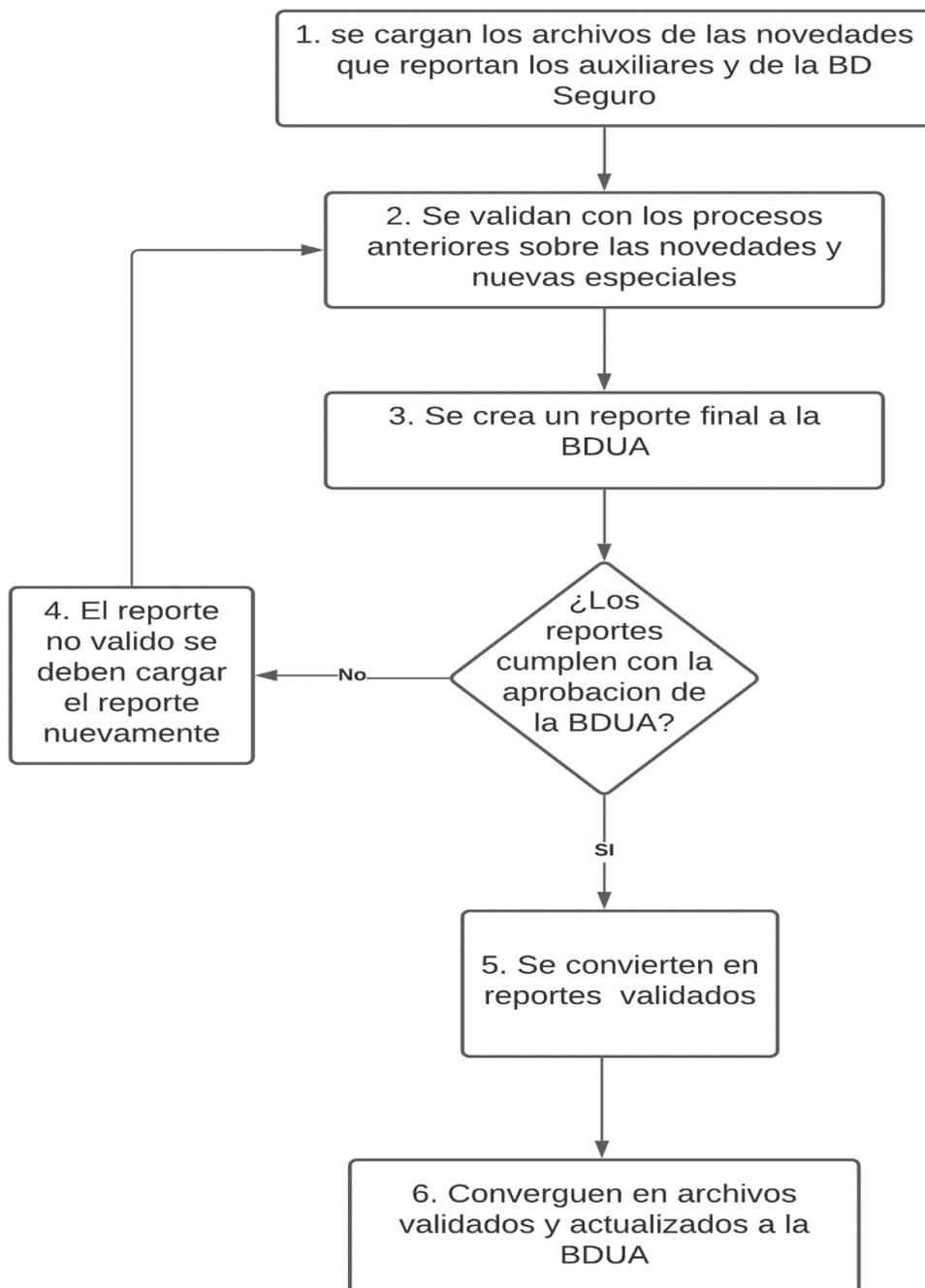


Fig. 22 Flujograma Proceso novedad MS Y NS

Una vez se realiza el reporte de las novedades, la BDUA en un tiempo determinado para devolver los resultados de los reportes, por ejemplo:

1. Archivo MS validado: hace referencia a todos los usuarios que fueron aprobados ya que no tuvieron inconsistencia.
2. Archivo MS negado: Todos los usuarios glosados que tuvieron inconsistencias.
3. Archivo NS validado: Todas las novedades que fueron avaladas y que no tuvieron inconsistencias.
4. Archivo NS negado: son las novedades que no pasaron y que tuvieron inconsistencias.
5. Archivo S1 validado: es el proceso de traslado de solicitud del afiliado y que fue avalado por la EPS
6. Archivo S1 negado: es el proceso de traslado de solicitud del afiliado y que fue negado por la otra EPS

Ahora bien, la intención del proyecto es brindarle a la organización una solución que permita generar indicadores para la automatización de los procesos donde se evalúa la cantidad de novedades reportadas por los dinamizadores que operan en diferentes departamentos y municipios. Para apoyar el desarrollo de la solución la organización cuenta con diferentes fuentes de información las cuales brindan los datos necesarios para realizar las transformaciones y modelos que se detallaran más adelante.

DEFINICIÓN DE UN PLAN DE TRABAJO

En esta etapa se define el plan de trabajo en el cual se realiza el seguimiento de las actividades del proyecto. Ver figura 23.

		Construcción de un DataMart en el Área de Aseguramiento Proceso y Gestion de Operaciones BDUA											
		Meses											
		Enero	Febreo	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
Actividad 1	inicio												
	Tarea 1	Desarrollo de la guía metodológica implementada en el anteproyecto											
Actividad 2	Definición del Proyecto												
	Tarea 1	Selección de la Organización											
	Tarea 2	Planteamiento de preguntas para la entrevista											
	Tarea 3	Realización de preguntas a los dinamizadores encargados de los procesos											
	Tarea 4	Definición del Área y proceso											
	Tarea 5	Definición del presupuesto											
Actividad 3	Definición de Requerimientos												
	Tarea 5	Definición de los requerimientos											
	Tarea 6	Definición de las fuentes de información											
	Tarea 7	Conocer las herramientas para los procesos ETL											
Actividad 3	Construcción del Modelo Dimensional												
	Tarea 1	Creación del modelo lógico											
	Tarea 2	Creación del modelo físico											
Actividad 4	Implementacion												
	Tarea 1	Realizar el proceso ETL											
	Tarea 2	Realizar Reportes											

Fig.23 Seguimiento de Actividades Proyecto Datamart

ETAPA 2: RECOLECCIÓN DE REQUERIMIENTOS

Para esta fase de identificación de los requerimientos se tuvo en cuenta la técnica de la entrevista, en la cual se plantearon las siguientes preguntas:

1. ¿cuál es la problemática de la organización?
2. ¿A qué área se dirige la problemática?
3. ¿Cuál es el proceso o subproceso donde se refleja la problemática?
4. ¿Cuál es el objetivo del proyecto?
5. ¿Qué información se manejan en el proceso?
6. ¿Qué proceso se pretende automatizar?
7. ¿Qué información se espera obtener con la implementación de la solución?
8. ¿Cuáles son las fuentes de información?

Una vez se planteó las preguntas se acordó una reunión en la cual participo el grupo de dinamizadores que son los encargados de administrar los diferentes procesos sobre la gestión de la información y se contó con personal de Gestión Tic quienes expusieron la necesidad.

Para dar respuesta en forma general sobre la definición de requerimientos, el equipo de trabajo expuso que la necesidad de la información se dirigía especialmente al área de Aseguramiento proceso gestión de operaciones BDUA. Este es uno de los procesos que hace parte de las diferentes operaciones que conforman la estructura organizacional. La problemática del proceso es, que las validaciones para el reporte final que se realiza a la BDUA son generadas manualmente lo cual hace un poco complejo la evaluación e identificación de dinamizadores que reportan las novedades de los diferentes municipios y departamentos.

IDENTIFICACIÓN DE LAS FUENTES DE INFORMACIÓN

Para el manejo de la información de los regímenes subsidiado y contributivo la Asociación Indígena del Cauca maneja gran cantidad de aplicaciones informáticas para cubrir las necesidades de los procesos y subprocesos de las diferentes áreas, la gran mayoría se encuentran alojadas en diferentes servidores con sistemas operativos Linux, Oracle, Windows Server 2008 y Windows 7, las bases de datos se encuentran creadas sobre motores de base de datos Firebird, Postgres, SqlServer entre otras.

Ahora bien, para identificar las fuentes de información que se requieren para realizar las validaciones sobre los procesos de las novedades que reportan los

dinamizadores y evaluar la efectividad de cargue, se requiere extraer la información de las siguientes fuentes de información.

- Archivos S1 (contiene archivos de procesos sobre las novedades que se reporta a la BDUA)
- Archivos S1 Validado (resultados de los procesos que reporta a la BDUA)
- Archivos S1 negado (resultados de los procesos que entrega la BDUA)
- Archivos NS (Reporte que se realiza a la BDUA de las novedades de los afiliados)
- Archivos NS Negado (Resultado del proceso que entrega la BDUA)
- Archivos NS Validado (Resultado del proceso que entrega la BDUA)
- Archivos MS (Reporte que se entrega a la BDUA de los nuevos afiliados)
- Archivos MS validado (Resultado del proceso que entrega la BDUA de las nuevas afiliaciones)
- Archivos MS negados (Resultado de los procesos que entrega la BDUA de los nuevos afiliados)
- Archivo Departamento1 (Contiene información sobre los departamentos y municipios)
- Archivo Municipio (Contiene información sobre los Municipios)
- Archivo ARC_DINA (Contiene información sobre los dinamizadores)
- Archivo COD_NOVEDAD (Contiene información sobre el código que se genera para los tipos de archivos de las novedades de la BDUA)

Nombre del Archivo (fuente de información)	Tipo _Formato
Archivos MS	TXT
Archivos MS Validado	TXT
Archivo MS Negado	TXT
Archivos NS	TXT
Archivo NS Negado	TXT
Archivo NS Validado	TXT
Archivo S1	TXT
Archivo S1 Automático	TXT
Archivo S3 del S1	TXT
Archivo Departamento1	CSV
Archivo Municipio	CSV
Arc_Dina	CVS
Cod_Novedad	CVS

Tabla 4. Descripción de Archivos de Novedades

A continuación, se describe la estructura y el tipo de datos de los archivos de las novedades que se deben reportar a la BDUA según la resolución 4622 del 2016 del Ministerio de Salud.

Nombre de la fuente de Información	Descripción	Campos
Archivos S1(Traslados, Movilidad)	Contiene la información de los traslados y movilidades de los afiliados. Es decir, si el	- Código de la entidad que solicita el traslado o movilidad

	<p>afiliado hace un cambio del régimen del subsidiado al contributivo o viceversa dentro de la misma EPS, los traslados se refieren a que el afiliado solicite el cambio de la EPS a otra ya sea por cualquier situación.</p>	<ul style="list-style-type: none"> - Tipo de documento del afiliado - Número de identificación del afiliado - Primer apellido del afiliado - Segundo apellido del afiliado - Primer nombre del afiliado - Segundo nombre del afiliado - Fecha de nacimiento del afiliado - Sexo del afiliado - Tipo de documento del afiliado - Número de identificación del afiliado - Primer apellido del afiliado - Segundo apellido del afiliado - Primer nombre del afiliado - Segundo nombre del afiliado - Código de departamento - Código del municipio - Zona - Fecha de afiliación y/o novedad - Tipo de población especial del régimen subsidiado - Nivel del SISBEN - Tipo de traslado - Número de identificación del cabeza de familia - Parentesco con la familia
--	---	--

<p>Archivo MS(Maestro Subsidiado)</p>	<p>Describe la información de las nuevas afiliaciones que aún no han sido ingresado por primera vez en la en la BDUA.</p>	<ul style="list-style-type: none"> - Código de la Entidad - Tipo de documento del afiliado - Número de identificación del afiliado - Primer apellido del afiliado - Segundo apellido del afiliado - Primer nombre del afiliado - Segundo nombre del afiliado - Fecha de nacimiento del afiliado - Sexo del afiliado - Código de departamento - Código municipio - Zona - Fecha de la afiliación y/o novedad (corresponde a la fecha de radicación del formulario único de afiliación. - Tipo de población especial del régimen subsidiado - Nivel SISBEN - Código IPS primaria - Condición de discapacidad - Tipo de documento de cabeza de familia - Parentesco con el cabeza de familia - Etnia
<p>Archivo NS(Novedades de Afiliaciones)</p>	<p>Describe la información sobre las novedades de las modificaciones de los afiliados o corrección de información (cambio de nombre, edad, cambio de documento, etc.)</p>	<ul style="list-style-type: none"> - Consecutivo de la novedad - Código de la entidad - Tipo de documento del afiliado

		<ul style="list-style-type: none"> - Número de identificación del afiliado - Primer apellido del afiliado - Segundo apellido del afiliado - Primer nombre del afiliado - Segundo nombre del afiliado - Fecha de nacimiento del afiliado - Código del departamento - Código del municipio - Código de la novedad - Fecha inicio de la novedad
--	--	--

Tabla 4: Estructura del Archivo S1, MS Y NS

SELECCIÓN DE HERRAMIENTA PARA PROCESO ETL

En la etapa de selección de herramientas para el proceso ETL se tiene en cuenta las herramientas Pentaho versión 8.3 y Talend Open Studio las cual poseen un conjunto de funciones para realizar los diferentes procesos. Además, se elige estas herramientas debido a que son softwares gratuitos y permiten conexión con diferentes administradores de base de datos como ORACLE, MYSQL, Para un completo funcionamiento se debe instalar jdk 8. Ahora bien, para la presentación de los informes se tuvo en cuenta la herramienta de Google Data Studio debido a que ofrece una variedad de presentación de informes y paneles e informáticos.

CONSTRUCCIÓN DE PREGUNTAS PARA PLANTEAR EL MODELO DIMENSIONAL

En esta parte de la etapa se creó las siguientes preguntas con el fin de determinar las perceptivas e indicadores que se utilizaran de análisis. Para ello, se debe tener en cuenta que estos sean realmente efectivos y deben mostrar lo que se quiere analizar. Con los indicadores se realiza el planteamiento del modelo dimensional el cual posee la tabla de hechos y dimensiones, estos indican el alcance del proyecto.

Ahora bien, de acuerdo con la información que se extrajo sobre las preguntas que se realizaron al equipo de trabajo las perceptivas que se tiene sobre los indicadores son las siguientes:

1. Conocer la cantidad de novedades reportadas, validadas y glosadas por semana, mes y año de acuerdo con el reporte de la BDUA.
2. Conocer la cantidad de novedades por dinamizador reportadas, glosadas, validadas por semana, mes, año por departamento y municipio
3. Conocer el porcentaje de efectividad de cargue que realiza cada dinamizador

En la siguiente tabla se visualizan los indicadores y las perspectivas obtenidas de acuerdo con los requerimientos establecidos.

Indicadores	Perspectivas
Total de novedades reportadas, validadas, glosadas sobre el reporte que devuelve la BDUA.	<ul style="list-style-type: none"> • Semana • Mes • Año
Total de novedades reportadas, glosadas, validadas que realiza un	<ul style="list-style-type: none"> • Semana • Mes • Año

dinamizador, por departamento y municipio.	
Total de porcentaje de cargue de efectividad de las novedades que realiza un dinamizador	<ul style="list-style-type: none"> • Dinamizador

Tabla 5: Tabla de Indicadores y Perceptivas

MATRIZ DE REQUERIMIENTOS

En esta parte se describen los requerimientos del negocio y tabla de hecho.

Definición de Requerimientos				
Nombre Requerimiento	Descripción	Agregación	Objetivo	Tiempo de Respuesta de la Información
CTNovedadesBDUA	Cantidad de novedades reportadas, validadas y glosadas que devuelve la BDU A	Count	Identificar la cantidad de novedades reportadas validadas y glosadas con base a los resultados de los procesos de la BDU A	2 días con tratamiento de la información manual/ tiempo de respuesta inmediato con la implementación del Datamart
CTNovedadesDinamizador	Cantidad de novedades reportadas, validadas y glosadas que realiza un dinamizador por departamento y municipio	Count	Identificar la cantidad de novedades que reporta un dinamizador por departamento y municipio	2 días con tratamiento de la información manual/ tiempo de respuesta inmediato con la implementación del Datamart

CPTEfectividadCargueDinamizador	porcentaje de cargue sobre las novedades que reporta un dinamizador	Porcentaje	Evaluar la efectividad de cargue de novedades que realiza un dinamizador mensualmente	2 días con tratamiento de la información manual/ tiempo de respuesta inmediato con la implementación del Datamart
---------------------------------	---	------------	---	---

Tabla 6: Tabla Definición de Requerimientos

A continuación, se especifican las dimensiones relacionadas con la tabla de hecho.

Nombre	Descripción	Tiempo	Novedades	Dinamizador
CTNovedadesBDUA	Cantidad de novedades reportadas, validadas y glosadas que devuelve la BDUA	X	X	
CTNovedadesDinamizador	Cantidad de novedades reportadas, validadas y glosadas que realiza un dinamizador por departamento y municipio	X	X	X
CPEfectividadCargueDinamizador	porcentaje de cargue sobre las novedades que reporta un dinamizador	X	X	X

Tabla 7: Tabla Análisis de Dimensiones

ETAPA 3: CONSTRUCCIÓN DEL MODELO DIMENSIONAL

Para el desarrollo del modelo dimensional lógico se tuvo en cuenta el esquema de diseño estrella debido al planteamiento de los requerimientos, en la figura 5 se observa el modelo que contiene las dimensiones y la tabla de hechos para realizar el análisis y dar respuesta al requerimiento del negocio.

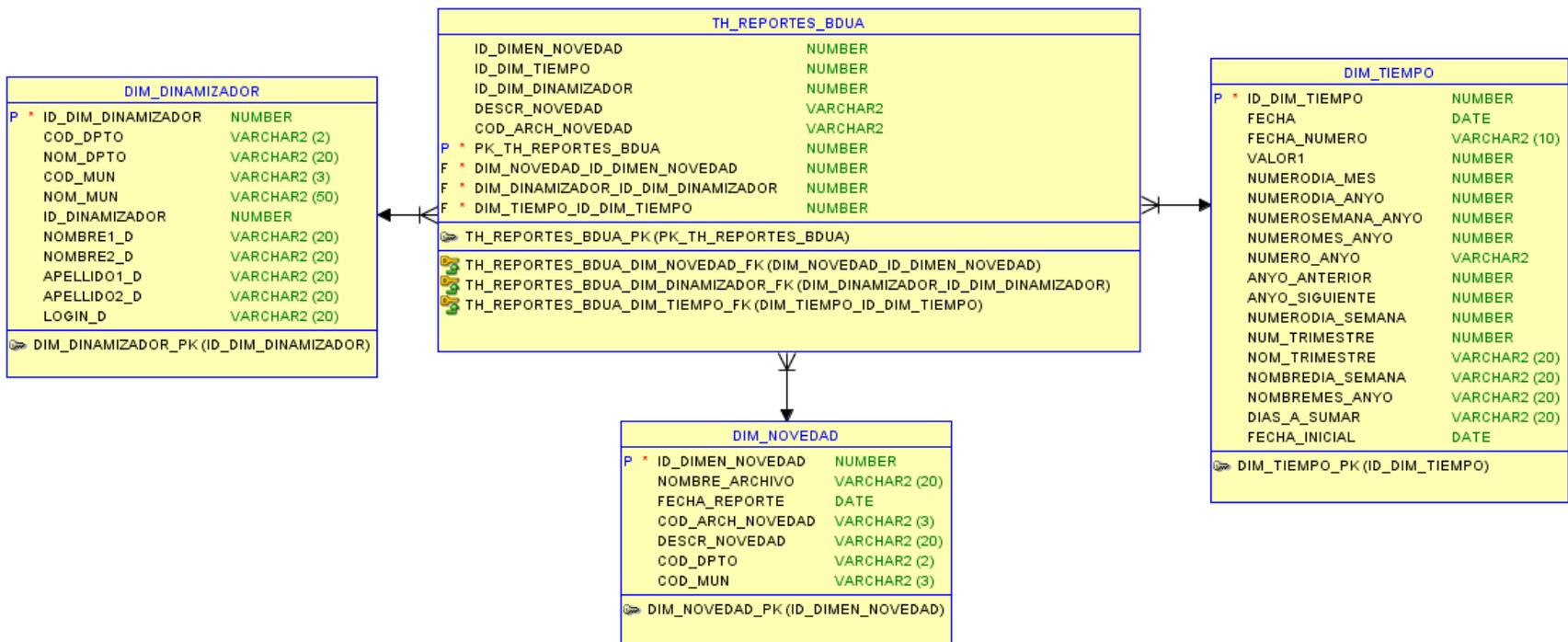


Fig.24 Modelo Dimensional esquema en estrella

A continuación, se describe las dimensiones del modelo con los respectivos campo y atributos.

DIM_DINAMIZADOR		
Nombre del Campo	Tipo de Dato	Detalle
Id_dim_dinamizador	Number	Hace referencia a la llave primaria de la identificación única de la dimensión
Cod_dpto	Varchar (2)	Contiene el código de departamento
Nom_dpto	Varchar (20)	Contiene el nombre de los departamentos
Cod_mun	Varchar (3)	Contiene el código del municipio
Nom_mun	Varchar(20)	Contiene el nombre de los municipios
ID_dinamizador	Number	Contiene un numero único de identificación para cada dinamizador
Nombre1_d	Varchar (20)	Contiene el primer nombre del dinamizador

Nombre2_d	Varchar (20)	Contiene el segundo nombre del dinamizador
Apellido1_d	Varchar (20)	Contiene el primer apellido del dinamizador
Apellido2_d	Varchar (20)	Contiene el segundo apellido del dinamizador
Login	Varchar(20)	Contiene el login del dinamizador

Tabla 8: Descripción de Campos y formato Dimensión Dinamizador

DIM_TIEMPO		
Nombre del Campo	Tipo de Dato	Detalle
Id_tiempo	Number	Hace referencia a la llave primaria de la identificación única de la dimensión
Fecha	Date	Contiene el formato fecha dd/mm/aa
Fecha_numero	Varchar (10)	Contiene la fecha en texto
Valor1	Number	Contiene un código único para las validaciones de fecha

Numerodia_mes	Number	Hace referencia al día del mes
Numerodia_anyo	Number	Contiene el número del día del año
Numerosemana_anyo	Number	Hace referencia al número de la semana en se encuentra
Numeromes_anyo	Number	Contiene el número del mes que pertenece
Numero_anyo	Number	Contiene el número del año
Año_anterior	Number	Hace referencia al año anterior
Año_siguiete	Number	Hace referencia al año siguiente
Numerodia_semana	Number	Hace referencia al valor de la semana
Num_trimestre	Number	Hace referencia al trimestre se encuentra
Nom_trimestre	Varchar (20)	Contiene el nombre del trimestre
Nombredia_semana	Varchar (20)	Contiene el nombre del día de la semana

Nombremes_año	Varchar (20)	Contiene el nombre del año
Días_a_sumar	Varchar (20)	Contiene los días que posee el mes
Fecha_inicial	Date	Contiene la fecha de inicio que se cargó en la dimensión

Tabla 9: Descripción de Campos y formato Dimensión Tiempo

DIM_NOVEDAD		
Nombre del Campo	Tipo de Dato	Detalle
Id_dimen_novedad	Number	Hace referencia a la llave primaria de la identificación única de la dimensión
Nombre_Archivo	Varchar (20)	Contiene el nombre del archivo de la novedad
Fecha_Reporte	Date	Contiene la fecha de reporte cuando se realizó a la BDU

Cod_Arch_Novedad	Varchar (3)	Contiene el código que se le asigno a la novedad, si es un S1, MS O NS
Descr_Novedad	Varchar (20)	Contiene el nombre que se le asigno a la novedad por ejemplo S1 reportado, S1 Validado S1 negado o glosado, etc.
Cod_Dpto	Varchar (2)	Contiene el código de departamento
Cod_Mun	Varchar (3)	Contiene el código del municipio

Tabla 10 : Descripción de campo y atributos Dimensión Novedad

ETAPA 5: DISEÑO DEL PROCESO ETL (EXTRACION TRANSFORMACION Y CARGA)

En la primera etapa del proceso ETL que consiste en la extracción de datos se tuvo en cuenta las diferentes fuentes de información que cuenta la organización para el alcance del proyecto, estas se describen en la etapa de levantamiento de requerimientos. Ahora bien, para la segunda etapa que consiste en la transformación se determinaron los archivos que se manejan en los procesos BDUA los cuales contienen una estructura única para cada proceso.

Descripción del Proceso ETL Carga de Datos del Plano Archivo MS Reportado

Este plano contiene la información sobre los nuevos afiliados a la BDUA

En la siguiente figura se muestra el proceso ETL para el cargue del archivo novedad MS



Fig.24 ETL Carga de Archivo Novedad MS Reportado

Descripción del Proceso ETL Carga de Datos del Plano Archivo MS Validado

En este plano se hace la transformación de cargue de la información de los nuevos afiliados que fueron aceptados y validados en la BDUA.

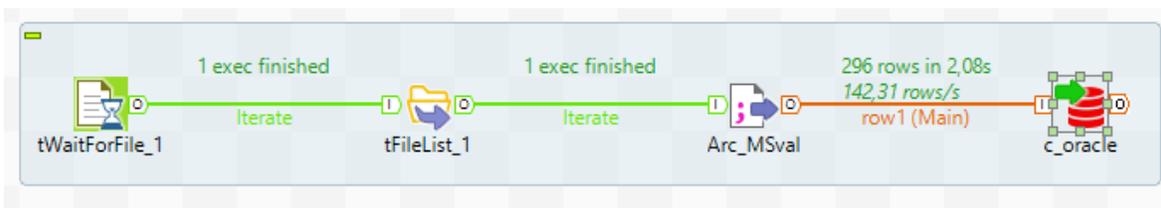


Fig.25 ETL Carga de Archivo Novedad MS Validado

Descripción del Proceso ETL Carga de Datos del Plano Archivo MS Negado

En este plano se hace la transformación de cargue de la información de los nuevos afiliados que no fueron aceptados y validados en la BDUA.

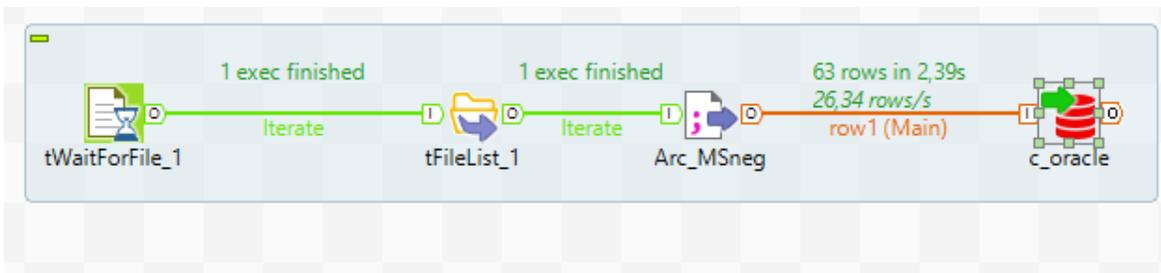


Fig.26 ETL Carga de Archivo Novedad MS Negado

Descripción del Proceso ETL Carga de Datos del Plano Archivo S1 Reportado

En este plano se realiza la transformación y cargue de la información que contiene los las novedades de movilidad y traslados de los afiliados de la EPS.

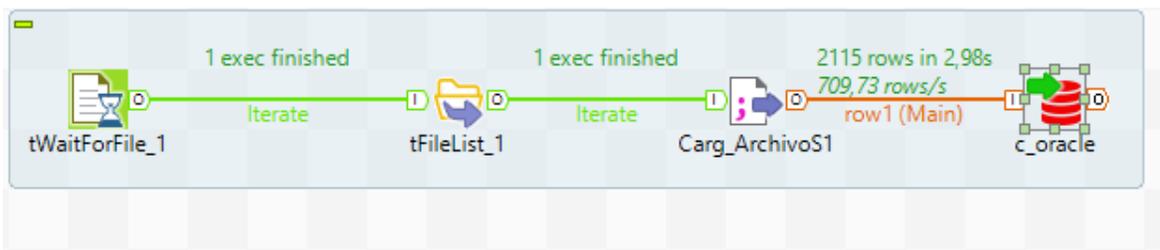


Fig.27 ETL Carga de Archivo Novedad S1 Reportado

Descripción del Proceso ETL Carga de Datos del Plano Archivo S1 Validado

En este plano se realiza la transformación y cargue de la información de novedades de movilidades y traslados que realizan a los afiliados en cuanto al cambio régimen y traslados a otras EPS.

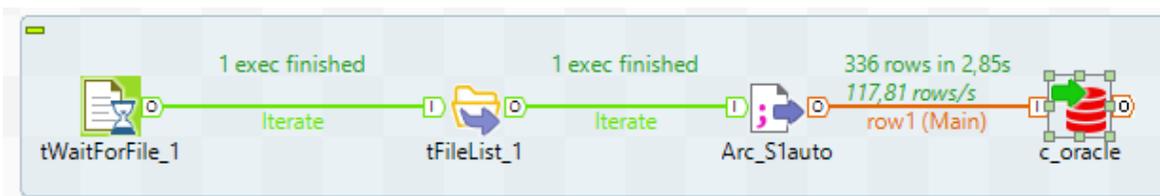


Fig.28 ETL Carga de Archivo Novedad S1 Validado

Descripción del Proceso ETL Carga de Datos del Plano Archivo S1 Negado

En este plano se realiza la transformación y cargue de la información de novedades de movilidades y traslados que realizan a los afiliados que no fueron validados en la BDU A



Fig.29 ETL Carga de Archivo Novedad S1 Negado

Descripción del Proceso ETL Carga de Datos del Plano Archivo NS Reportado

En este plano se realiza la transformación y cargue de la información sobre las novedades que se realizan en los afiliados por ejemplo cambio del nombre, apellidos, cambio de tipo documento entre otro tipo de novedades.



Fig.30 ETL Carga de Archivo Novedad NS Reportado

Descripción del Proceso ETL Carga de Datos del Plano Archivo NS Validado

En este plano se realiza la transformación y cargue de la información sobre las novedades que se realizan en los afiliados y que si fueron validados y aceptados en la BDU A.



Fig.31 ETL Carga de Archivo Novedad NS Validado

Descripción del Proceso ETL Carga de Datos del Plano Archivo NS Negado

En este plano se realiza la transformación y cargue de la información sobre las novedades que se realizan en los afiliados y que no fueron validados en la BDU A

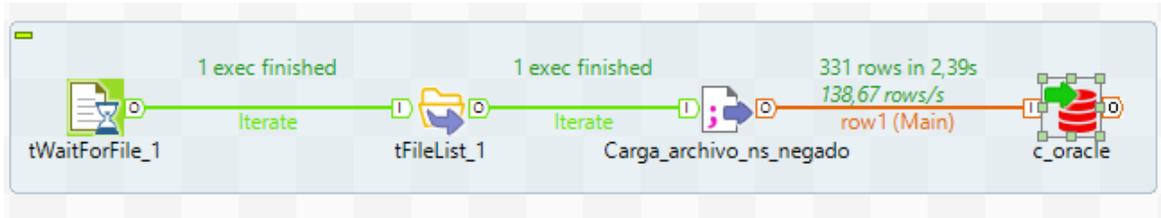


Fig.32 ETL Carga de Archivo Novedad NS Negado

Descripción del Proceso ETL Carga de Datos Dimensión Dinamizador

Para el cargue de este plano se analizó las diferentes fuentes de información donde estrago los datos de los dinamizadores, departamentos y municipio que tienen a cargo cada dinamizador se utilizó como herramienta para el proceso ETL Pentaho 8.3.

En la siguiente figura se ilustra el cargue de la dimensión dinamizador, para este proceso se extrae la información que se encuentra cargada en la tabla dinamizador en la base de

datos Registro luego se selecciona los campos y se carga en la tabla dimensión dinamizador que se encuentra creada en la base de datos AIC



Fig.33 ETL Cargue Dimensión Dinamizador

Descripción del Proceso ETL Carga de Datos Dimensión Novedad

Para el cargue de este plano se extrajeron de las fuentes de información los archivos de las novedades que se cargaron en la base de datos Registro este proceso de cargue de los archivos se detalló anteriormente.

En la figura 33 muestra el cargue de los datos de los archivos de las novedades que específicamente el campo **nombre archivo, fecha reporte, código del departamento y código del municipio**. Además, se agregan dos campos que contiene el código del archivo de la novedad y descripción del archivo de la novedad. Estos códigos se visualizan en la figura 35.

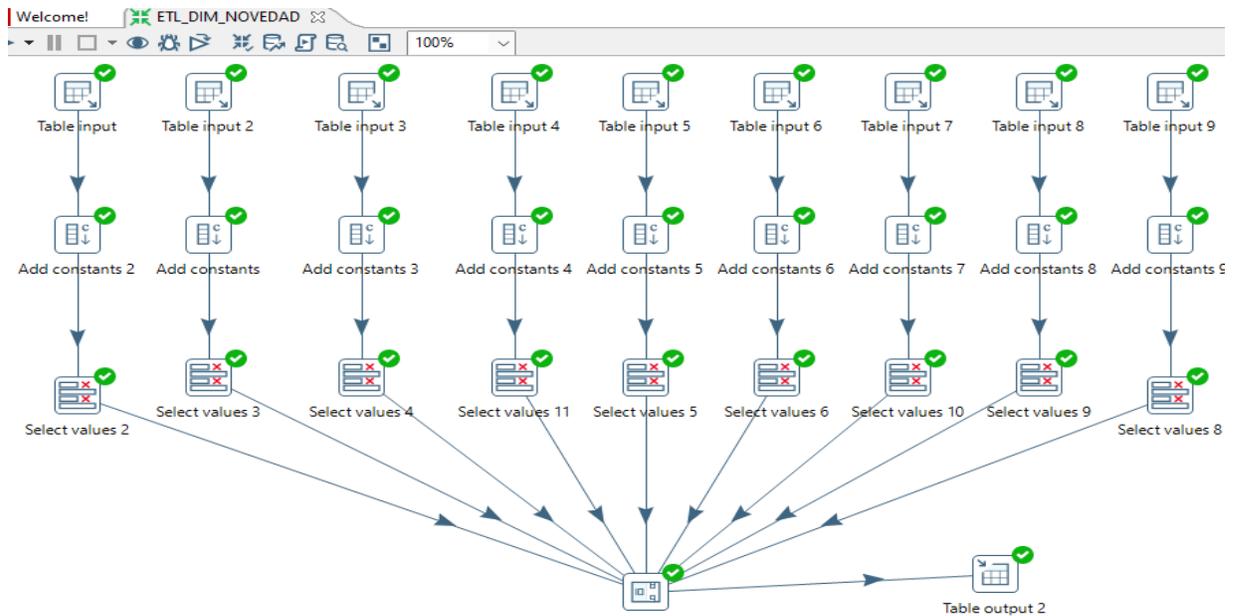


Fig.34 ETL Cargue de Dimensión Novedad

CODIGO_ARCHI_NOVEDAD	DESC_NOVEDAD
1	S1_VALIDADO
2	S3_NEGADO
11	S1_REPORTADO
9	MS_REPORTADO
5	MS_VALIDADO
6	MS_NEGADO
10	NS_REPORTADO
7	NS_VALIDADO
8	NS_NEGADO

Fig.35 Códigos y Descripción de las Novedades

Descripción del Proceso ETL Carga de Datos Dimensión Tiempo

En plano genera los datos vistos para cada fecha del desde el 01 de enero del 2015 hasta el año 01 de enero del 2030. Se realizo validaciones de fechas y fórmulas para mostrar los valores en la dimensión.

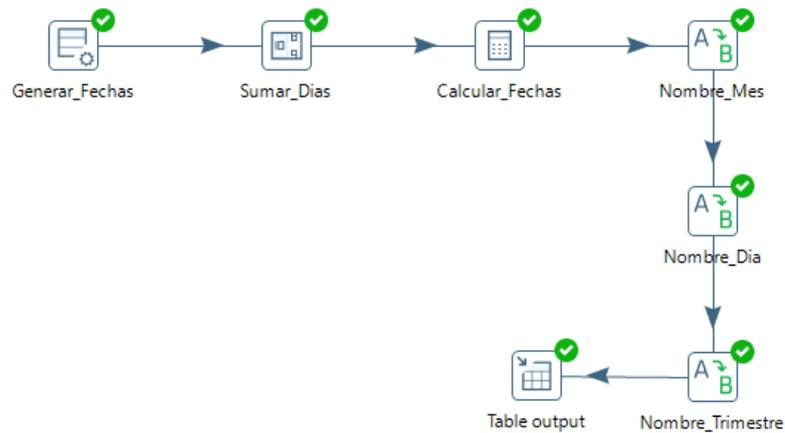


Fig.36 ETL Cargue Dimensión Tiempo

Descripción del Proceso ETL Carga de Datos Tabla de Hechos Reportes BDUA

En este plano se carga los ID (llaves primarias de las tablas de las dimensiones) en la primera entrada se selecciona la dimensión novedad en la cual se extrae los campos correspondientes, en la segundo item se concatena los campos en relación como el código de la archivo de la novedad, código del departamento y código municipio campos que se requieren para realizar las consultas, en el tercer item se selecciona nuevamente los campos de la dimensión mencionada, en el cuarto item Database Lookup se busca en la base de datos AIC la dimensión dinamizador y extrae el ID, código del departamento y municipio el cual tiene en relación con la dimensión anterior, en quinto item nuevamente se busca la dimensión tiempo en la base de datos y extrae el campo en relación en este caso la fecha de reporte que es contenida en la dimensión novedad. Para este caso se convierte la fecha de reporte en un formato donde se pueda igualar a la fecha numero de la dimensión tiempo esto es con el fin de poder identificar las fechas en la dimensión tiempo. De igual forma se añade una secuencia que corresponde a un identificador único

de la tabla de hechos (llave primaria) y por último se carga la tabla de hechos con los respectivos ID y campos para realizar las consultas.

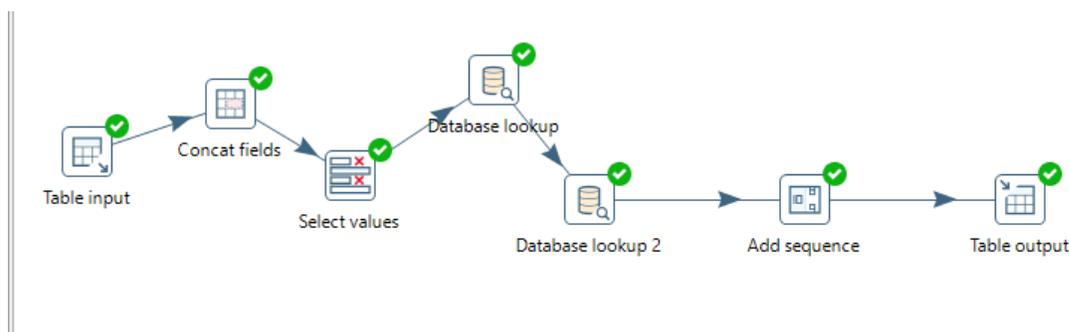


Fig.37 ETL Carga Tabla de Hechos Reportes BDUA

Pruebas de Carga de Datos (Proceso ETL)

Para realizar las pruebas de cargue se ejecutó cada uno de los ETL del plano correspondiente.

En la siguiente figura se muestra el llenado de la tabla del archivo novedad MS

IDTA...	NOMBRE_ARCHIVO	FECHA_REPORTES	COD_ENTIDAD	TIPO_DOC_AFILIADO	NUM_IDENTI_AFILIADO	APELLIDO1_AFILIADO	APELLIDO2_AFILIADO
58451	MSEPSI0315012021.TXT	15/01/21	EPSI03	CC	1002970270	QUILINDO	GUTIERREZ
58452	MSEPSI0315012021.TXT	15/01/21	EPSI03	CC	1004508905	RUALES	BUELVAS
58453	MSEPSI0315012021.TXT	15/01/21	EPSI03	CC	1006919755	EPIEYU	(null)
58454	MSEPSI0315012021.TXT	15/01/21	EPSI03	CC	1059609035	CHOCUE	COMETA
58455	MSEPSI0315012021.TXT	15/01/21	EPSI03	CC	1060109493	TOMBE	MEDINA
58456	MSEPSI0315012021.TXT	15/01/21	EPSI03	CC	1103118509	GONZALEZ	BASILIO

Fig.38 Carga de Datos Archivo Novedad MS

En la siguiente figura se muestra el llenado de la tabla del archivo novedad MS validado

ID_TA_MS...	NOMBRE_ARCHIVO	FECHA_REPORTES	COD_ENTIDAD	TIPO_DOC_AFILIADO	NUM_IDENTI_AFILIADO	APELLIDO1_AFILIADO	APELLIDO2
9762	MSEPSI0308012021.VAL	08/01/21	EPSI03	CC	1002840264	JIMENEZ	MELENJE
9763	MSEPSI0308012021.VAL	08/01/21	EPSI03	CC	10473340	CAMPO	ZAMBRANO
9764	MSEPSI0308012021.VAL	08/01/21	EPSI03	CC	1061815495	CHAVARRO	PEREZ
9765	MSEPSI0308012021.VAL	08/01/21	EPSI03	CC	1110118665	ESQUIVEL	ARIAS
9766	MSEPSI0308012021.VAL	08/01/21	EPSI03	CC	1124371839	BOURIYU	BOURIYU

Fig.39 Carga de Datos Archivo Novedad MS Validado

En la siguiente figura se muestra el llenado de la tabla archivo novedad MS Negado

ID_TAB_MS...	NOMBRE_ARCHIVO	FECHA_REPORTE	COD_ENTIDAD	TIPO_DOC_AFILIADO	NUM_IDENTI_AFILIADO	APELLIDO1_AFILIADO	APELLIDO2
7921	MSEPSI0308012021.NEG	08/01/21	EPSI03	CC	1060109493	TOMBE	MEDINA
7922	MSEPSI0308012021.NEG	08/01/21	EPSI03	CC	1061694857	RUIZ	RUIZ
7923	MSEPSI0308012021.NEG	08/01/21	EPSI03	CC	1067535109	CAMPO	MONTANO
7924	MSEPSI0308012021.NEG	08/01/21	EPSI03	CC	1131769404	BAILARIN	DOMICO
7925	MSEPSI0308012021.NEG	08/01/21	EPSI03	CC	625485	PEREZ	PADRON
7926	MSEPSI0308012021.NEG	08/01/21	EPSI03	CN	15630117	PESCADOR	GASPAR
7927	MSEPSI0308012021.NEG	08/01/21	EPSI03	CN	15631682	AYALA	ROJAS
7928	MSEPSI0308012021.NEG	08/01/21	EPSI03	CN	156334537	ARICAPA	TREJOS

Fig.40 Carga de Datos Archivo Novedad MS Negado

En la siguiente figura se muestra el llenado de la tabla archivo novedad NS Reportado

IDTAB_ARC...	NOMBRE_ARCHIVO	FECHA_REPORTE	CONSEC_NOVEDAD	COD_ENTIDAD	TIPO_DOC_AFILIADO	NUM_IDENTI_AFILIADO	APELLIDO1_AFILIADO	APELLIDC
29353	NSEPSI0308012021.txt	08/01/21	1077	EPSI03	CC	1002855263	NARVAEZ	
29354	NSEPSI0308012021.txt	08/01/21	213	EPSI03	CC	1002949180	QUIGUANAS	
29355	NSEPSI0308012021.txt	08/01/21	69	EPSI03	CC	1002953534	BOLADOS	
29356	NSEPSI0308012021.txt	08/01/21	160	EPSI03	CC	1002958416	SANCHEZ	
29357	NSEPSI0308012021.txt	08/01/21	109	EPSI03	CC	1002980087	CUETOCAMBO	
29358	NSEPSI0308012021.txt	08/01/21	1094	EPSI03	CC	1003084899	CRUZ	
29359	NSEPSI0308012021.txt	08/01/21	1150	EPSI03	CC	1003372561	VALENCIA	

Fig.41 Carga de Datos Archivo Novedad NS Reportado

En la siguiente figura se muestra el llenado de la tabla archivo novedad NS Validado

IDTAB_ARC...	NOMBRE_ARCHIVO	FECHA_REPORTE	CONSEC_NOVEDAD	COD_ENTIDAD	TIPO_DOC_AFILIADO	NUM_IDENTI_AFILIADO	APELLIDO1_AFILIADO	APELLIDC
1	NSEPSI0308012021.VAL	08/01/21	2	EPSI03	RC	1040522511	MENDEZ	ALVAREZ
2	NSEPSI0308012021.VAL	08/01/21	3	EPSI03	RC	1040522511	MENDEZ	ALVAREZ
3	NSEPSI0308012021.VAL	08/01/21	4	EPSI03	RC	1040800520	CARUPIA	DOMICO
4	NSEPSI0308012021.VAL	08/01/21	5	EPSI03	RC	1040799682	DOMICO	DOMICO
5	NSEPSI0308012021.VAL	08/01/21	6	EPSI03	RC	1040806369	BAILARIN	CARUPIA
6	NSEPSI0308012021.VAL	08/01/21	7	EPSI03	RC	1040806369	BAILARIN	CARUPIA

Fig.42 Carga de Datos Archivo Novedad NS Validado

En la siguiente figura se muestra el llenado de la tabla archivo novedad NS Negado

ID_TAB_NS...	NOMBRE_ARCHIVO	FECHA_REPORTE	CONSEC_NOVEDAD	COD_ENTIDAD	TIPO_DOC_AFILIADO	NUM_IDENTI_AFILIADO	APELLIDO1_AFILIADO	APELLIDC
4454	NSEPSI0308012021.NEG	08/01/21	1	EPSI03	CN	164172939	MENDEZ	
4455	NSEPSI0308012021.NEG	08/01/21	24	EPSI03	CN	162357177	CORDOBA	
4456	NSEPSI0308012021.NEG	08/01/21	25	EPSI03	RC	1038826338	CORDOBA	
4457	NSEPSI0308012021.NEG	08/01/21	26	EPSI03	RC	1038826338	CORDOBA	
4458	NSEPSI0308012021.NEG	08/01/21	34	EPSI03	CN	162141425	ARRIETA	
4459	NSEPSI0308012021.NEG	08/01/21	35	EPSI03	RC	1013375348	ARRIETA	
4460	NSEPSI0308012021.NEG	08/01/21	36	EPSI03	RC	1013375348	ARRIETA	
4461	NSEPSI0308012021.NEG	08/01/21	37	EPSI03	RC	1039653874	SAITAMO	

Fig.43 Carga de Datos Archivo Novedad NS Negado

En la siguiente figura se muestra el llenado de la tabla archivo novedad S1 Reportado

ID_TAB_AR...	NOMBRE_ARCHIVO	FECHA_REPORT	COD_ENTIDAD	TIPO_DOC_AFILIADO	NUM_IDENTI_AFILIADO	APELLIDO1_AFILIADO	APELLIDO
2827	S1EPSI0302032021.TXT	02/03/21	EPSI03	CC	100033433	MARTINEZ	YASNO
2828	S1EPSI0302032021.TXT	02/03/21	EPSI03	CC	1001243197	ROHATAN	HERNANDE
2829	S1EPSI0302032021.TXT	02/03/21	EPSI03	CC	1001499088	TASCON	GONZALEZ
2830	S1EPSI0302032021.TXT	02/03/21	EPSI03	CC	1001500245	BETANCUR	CHAVARRI
2831	S1EPSI0302032021.TXT	02/03/21	EPSI03	CC	1001849130	VILLADIEGO	ALVAREZ
2832	S1EPSI0302032021.TXT	02/03/21	EPSI03	CC	1002084572	HERNANDEZ	SOLANO
2833	S1EPSI0302032021.TXT	02/03/21	EPSI03	CC	1002755197	ARICAPA	BUENO
2834	S1EPSI0302032021.TXT	02/03/21	EPSI03	CC	1002777260	AGUILAR	MOMPOTE
2835	S1EPSI0302032021.TXT	02/03/21	EPSI03	CC	1002793761	BOMBA	CAMPO
2836	S1EPSI0302032021.TXT	02/03/21	EPSI03	CC	1002793913	CHEPE	POSCUE

Fig.44 Carga de Datos Archivo Novedad S1 Reportado

En la siguiente figura se muestra el llenado de la tabla archivo novedad S1 Validado

ID_S1...	NOMBRE_ARCHIVO	FECHA_REPORT	COD_ENTIDAD	TIPO_DOC_AFILIADO	NUM_IDENTI_AFILIADO	APELLIDO1_AFILIADO	APELLIDO2_
2518	S1EPSI0305012021.VAL	05/01/21	EPSI03	CC	1000325284	RODRIGUEZ	TORRES
2519	S1EPSI0305012021.VAL	05/01/21	EPSI03	CC	1000575451	JIPIZ	MEDINA
2520	S1EPSI0305012021.VAL	05/01/21	EPSI03	CC	1001024458	CORTES	TASCON
2521	S1EPSI0305012021.VAL	05/01/21	EPSI03	CC	1001534128	MORELO	MESTRA
2522	S1EPSI0305012021.VAL	05/01/21	EPSI03	CC	1001619217	DOMICO	MAJORE
2523	S1EPSI0305012021.VAL	05/01/21	EPSI03	CC	1001637511	DOMICO	UPUA
2524	S1EPSI0305012021.VAL	05/01/21	EPSI03	CC	1001772506	TASCON	TASCON

Fig.45 Carga de Datos Archivo Novedad S1 Validado

En la siguiente figura se muestra el llenado de la tabla archivo novedad S1 Negado

ID_TAB_ARC_S3...	NOMBRE_ARCHIVO	FECHA_REPORT	COD_ENTIDAD	TIPO_DOC_AFILIADO	NUM_IDENTI_AFILIADO	APELLIDO1_AFILIADO	APELLIDO2_
1	S3EPSI0305012021.TXT	05/01/21	EPSI03	AS	99524IA306	RAMIREZ	(null)
2	S3EPSI0305012021.TXT	05/01/21	EPSI03	CC	1000888104	ZAPATA	CATADO
3	S3EPSI0305012021.TXT	05/01/21	EPSI03	CC	1001499051	GONZALEZ	TASCON
4	S3EPSI0305012021.TXT	05/01/21	EPSI03	CC	1002755197	ARICAPA	BUENO
5	S3EPSI0305012021.TXT	05/01/21	EPSI03	CC	1002797469	MOLINA	CRUZ
6	S3EPSI0305012021.TXT	05/01/21	EPSI03	CC	1002853576	GUEVARA	HERNANDEZ
7	S3EPSI0305012021.TXT	05/01/21	EPSI03	CC	1002853847	DIAZ	GAVIRIA

Fig.46 Carga de Datos Archivo Novedad S1 Negado

En las siguientes figuras se muestra el cargue de las dimensiones del Datamart

ID_DIM...	ID_DINA...	NOMBRE1_D	NOMBRE2_D	APELLIDO...	APELLIDO2_D	LOGIN_D	COD_DPTO	COD_MUN	NOM_DPTO
1	1	MANUEL	FELIPE	MOLINA	CARDONA	MMOLINA	05	045	ANTIOQUIA
2	1	MANUEL	FELIPE	MOLINA	CARDONA	MMOLINA	05	051	ANTIOQUIA
3	1	MANUEL	FELIPE	MOLINA	CARDONA	MMOLINA	05	120	ANTIOQUIA
4	1	MANUEL	FELIPE	MOLINA	CARDONA	MMOLINA	05	172	ANTIOQUIA
5	1	MANUEL	FELIPE	MOLINA	CARDONA	MMOLINA	05	234	ANTIOQUIA
6	1	MANUEL	FELIPE	MOLINA	CARDONA	MMOLINA	05	250	ANTIOQUIA
7	1	MANUEL	FELIPE	MOLINA	CARDONA	MMOLINA	05	480	ANTIOQUIA
8	1	MANUEL	FELIPE	MOLINA	CARDONA	MMOLINA	05	659	ANTIOQUIA
9	1	MANUEL	FELIPE	MOLINA	CARDONA	MMOLINA	05	736	ANTIOQUIA
10	2	VIVIANA	ANDREA	GARCIA	TABORDA	VGARCIA	05	154	ANTIOQUIA
11	2	VIVIANA	ANDREA	GARCIA	TABORDA	VGARCIA	05	284	ANTIOQUIA
12	2	VIVIANA	ANDREA	GARCIA	TABORDA	VGARCIA	05	364	ANTIOQUIA

Fig.47 Carga de Datos Dimensión Dinamizador

ID_DI...	NOMBRE_ARCHIVO	FECHA_REPORT	COD_ARCH_NOVEDAD	DESCR_NOVEDAD	COD_DPTO	COD_MUN
1	MSEPSI0308012021.txt	08/01/21	9	MS_REPORTADO	19	392
2	MSEPSI0308012021.txt	08/01/21	9	MS_REPORTADO	19	392
3	MSEPSI0308012021.txt	08/01/21	9	MS_REPORTADO	19	100
4	MSEPSI0308012021.txt	08/01/21	9	MS_REPORTADO	19	473
5	MSEPSI0308012021.txt	08/01/21	9	MS_REPORTADO	19	473

Fig.48 Carga de Datos Dimensión Novedad

ID_DIM_TIEMPO	FECHA	FECHA_NUMERO	VALOR1	NUMERODIA_MES	NUMERODIA_ANYO	NUMEROSEMANA_ANYO	NUMEROMES_ANYO
1	01/01/15	20150101	1	1	1	1	1
2	02/01/15	20150102	1	2	2	1	1
3	03/01/15	20150103	1	3	3	1	1
4	04/01/15	20150104	1	4	4	1	1
5	05/01/15	20150105	1	5	5	2	1
6	06/01/15	20150106	1	6	6	2	1

Fig.49 Carga de Datos Dimensión Tiempo

ID_DIMEN_NOVEDAD	ID_DIM_DINAMIZADOR	ID_DIM_TIEMPO	PK_TH	DESCR_NOVEDAD	COD_ARCH_NOVEDAD
1	2958	22	2256	1 MS_VALIDADO	5
2	2959	22	2256	2 MS_VALIDADO	5
3	2960	22	2256	3 MS_VALIDADO	5
4	2961	33	2256	4 MS_VALIDADO	5
5	2962	49	2256	5 MS_VALIDADO	5

Fig.50 Carga de Datos Tabla de Hechos Novedades BDU A

CAPITULO IV. RESULTADOS

En esta fase se presenta la implementación del Datamart en el cual se realiza las consultas para generar los indicadores y dar respuesta a la necesidad del negocio.

Como resultado se obtiene la siguiente información.

1. Para el primer indicador que corresponde a dar respuesta sobre la cantidad de novedades reportadas validadas y negadas por semana, mes y año de acuerdo con el reporte de BDU A se expone lo siguiente:

Para la segunda semana de enero del año 2021 que corresponde al calendario hábil en la cual se deben de realizar los reportes se describe la siguiente información. Ver figura

18 en esta ilustra los resultados sobre la consulta que se realizó en el Datamart donde nos arroja información sobre de la cantidad de novedades reportadas para cada proceso BDUA

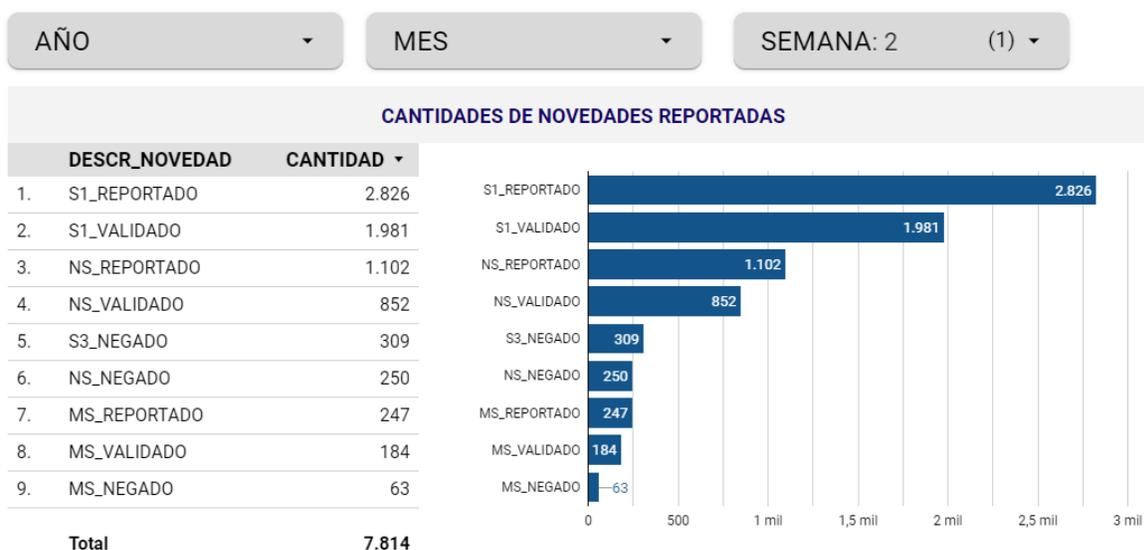


Fig.51 Cantidad de Novedades Reportadas Semana dos enero 2021

En la siguiente figura se ilustra la cantidad de novedades reportadas de los procesos BDUA.

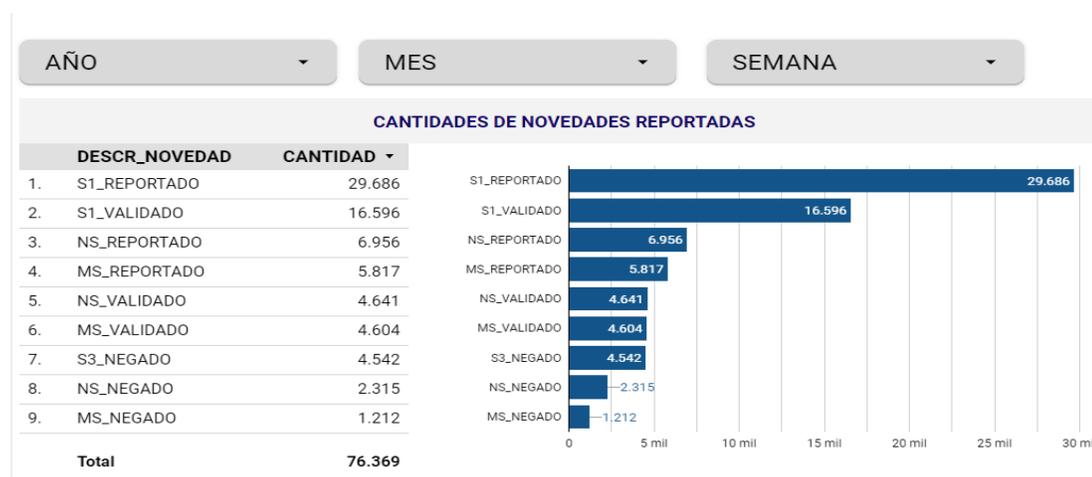


Fig.52 Cantidad de Novedades Reportadas Mes de Enero 2021

2. Para el segundo indicador que corresponde a dar respuesta sobre conocer la cantidad de novedades reportadas, validadas y glosadas por dinamizador semanal, mensual y anual se expone la siguiente información ver figura 19. En esta se ilustra el dinamizador y la cantidad de novedades reportadas en el mes de enero y el departamento y municipio que corresponde.

DEPARTAMENTO: LA ... (1) ▾		MUNICIPIO ▾			
NOMBRE: YULDOR HUE... (2) ▾		MES: Enero (1) ▾			
CANTIDADES DE NOVEDADES REPORTADAS					
	DEPARTAMENTO	MUNICIPIO	NOMBRE	DESCR_NOVEDAD	CANTIDAD ▾
1.	LA GUAJIRA	RIOHACHA	ADAN IGUARAN	S1_REPORTADO	903
2.	LA GUAJIRA	RIOHACHA	ADAN IGUARAN	S1_VALIDADO	99
3.	LA GUAJIRA	RIOHACHA	ADAN IGUARAN	MS_REPORTADO	64
4.	LA GUAJIRA	RIOHACHA	ADAN IGUARAN	NS_REPORTADO	61
5.	LA GUAJIRA	RIOHACHA	ADAN IGUARAN	NS_VALIDADO	52
6.	LA GUAJIRA	RIOHACHA	ADAN IGUARAN	MS_VALIDADO	44
7.	LA GUAJIRA	RIOHACHA	ADAN IGUARAN	S3_NEGADO	25
8.	LA GUAJIRA	RIOHACHA	ADAN IGUARAN	MS_NEGADO	20
9.	LA GUAJIRA	RIOHACHA	ADAN IGUARAN	NS_NEGADO	9
Total					1.277

Fig.53 Cantidad de Novedades Reportadas por Dinamizador mes Enero

3. Para dar respuesta a la pregunta conocer el porcentaje de cargue de efectividad de cargue que realiza cada dinamizador se creó el indicador que evalúa la cantidad de novedades reportadas versus la validadas.

Para este reporte se toma en cuenta dos dinamizadores como se puede evidenciar en la siguiente figura.

NOMBRE_COMPLETO: NELSON ... (2) ▾				
	NOMBRE_COMPLETO ▾	VALIDADOS	REPORTADOS	PORCENTAJE_...
1.	ADAN EBERTO SUAREZ IGUARAN	907	3681	24,6
2.	NELSON JULIAN RAMIREZ PEÑA	2221	3660	60,7

Fig.54 Porcentaje de Cargue por Dinamizador

CAPÍTULO V CONCLUSIONES Y RECOMENDACIONES

5.1 CONCLUSIONES

Cuando se va a seleccionar la herramienta BI de debe tener en cuenta si esta cumple con todos los requerimientos y requisitos que se han establecido, además se debe hacer un análisis en cuanto al tiempo de funcionamiento es decir si será utilizada para un corto o un largo plazo. Si la solución es utilizada a corto plazo no se recomienda usar la herramienta más costosa porque generaría más tiempo y costo en su implementación.

Al momento de implementar una solución de Inteligencia de Negocios se debe considerar las funcionalidades del sistema y que este se acople con las necesidades de usuarios en los diferentes departamentos de la organización.

Es importante conocer el organigrama, cultura de la organización para inculcar a los usuarios de la importancia de implementar este tipo de tecnologías, para ello se debe realizar campañas en las cuales los usuarios pueden captar el manejo o uso de la herramienta para su aceptación.

5.2 RECOMENDACIONES

La herramienta debe estar alineada con los objetivos de la organización.

Dedicar el tiempo necesario para todo el procesamiento de análisis de los datos de las fuentes de información. Para agilizar los procesos E.T.L y demás.

Elegir adecuadamente las herramientas, y que estas cumplan con las necesidades reglas y estándares de negocio, el proyecto debe cumplir con normas tecnológicas.

El proyecto se debe desarrollar en forma incremental para su mejora continua.

Se debe enfocar en las necesidades de información que posee la organización para así realizar la correcta implementación.

La herramienta debe estar desarrollada bajo un entorno amigable para el usuario, donde este pueda usarla con mayor facilidad.

Se recomienda avanzar en el desarrollo de aplicaciones para la unificación de las mismas.

Se recomienda realizar un análisis más profundo en los datos para detectar patrones de comportamiento en los afiliados.

6. GLOSARIO

- ❖ INTELIGENCIA DE NEGOCIOS: es un término que se conoce como un conjunto de aplicaciones, herramientas, estrategias y técnicas que se utilizan para la administración y obtención de conocimiento en los datos.
- ❖ GUÍA METODOLÓGICA: es la descripción de un conjunto de acciones, conceptos, metodologías y procesos que se abordan en la creación de algún proyecto.

- ❖ **MODELO DE MADUREZ:** este modelo es utilizado para la evaluación de procesos en una organización.
- ❖ **MODELO MEI:** este modelo define una línea bases para medir el valor de las iniciativas en un ámbito. Es decir, ayuda a las organizaciones a entender los aspectos que se deben mejorar y en qué situación está.
- ❖ **PYMES:** este término según la ley 590 del 2000 “se entiende por micro pequeña y mediana empresa toda unidad de explotación económica realizada por persona natural o jurídica en actividades empresariales, agropecuarias, industriales entre otras que cumplan con algunos requisitos”
- ❖ **ARQUITECTURA BI:** es el arte y técnica de diseñar, proyectar y aplicar una transformación estructural para optimizar un proceso analítico.
- ❖ **Big Data:** hace referencia al gran volumen de datos estructurados y no estructurados con los que cuenta una organización día a día.
- ❖ **ETL:** el proceso (Extraer, transformar y cargar) este permite a las empresas realizar limpieza, reestructuración y mover datos de diferentes fuentes para posteriormente cargarlos en una nueva base de datos.
- ❖ **KPA:** es el proceso en el cual se identifican actividades y prácticas las cuales permiten alcanzar las metas fundamentales.
- ❖ **Open Source:** es un software de código abierto en el cual las personas pueden modificarlo, ver y distribuirlo de tal manera en el que se considere conveniente. Por tanto, este permite la colaboración de trabajo grupal o empresarial.
- ❖ **MÓDULO BI:** Es un módulo que contiene la estructura de una arquitectura de Inteligencia de Negocios.

BIBLIOGRAFIA

- [1] J. E. Torres Romo, "Propuesta metodológica de una solución de inteligencia de negocios aplicada al sistema informático integrado de talento humano y sistema de registro de contratos y actas de finiquito," Universidad de las Américas, Quito, 2016.
- [2] A. A. Rosado Gómez and D. W. Rico Bautista, "Inteligencia de negocios: Estado del arte," *Scientia Et Technica*, 2010.
- [3] A. I. del Cauca, "Caracterización de la población Asociación Indígena del Cauca AIC EPS informe nacional." [Online]. Available: https://aicsalud.org.co/pdf/GDBV-F-8_Informe_capo_nacional_AIC_EPSI_2019.pdf.
- [4] Consejo Regional Indígena Del Cauca, "Resolución Noviembre 2017 (Por el se establecen las formas de participación social en la prestación de servicios de salud a los afiliados del régimen subsidiado y contributivo)," 2017. [Online]. Available: https://aicsalud.org.co/pdf/Resolucion_liga_de_usuarios.pdf.
- [5] Asociación Indígena del Cauca, "XI Asamblea General," 2018, [Online]. Available: http://aicsalud.org.co/pdf/informe_junta_directiva_23032018.pdf.
- [6] Asociación Indígena del Cauca AIC EPS I, "Asociación Indígena del Cauca," 2017. <https://aicsalud.org.co/>.

- [7] L. Fuentes Tapia and R. Valdivia Pinto, "Incorporación de elementos de inteligencia de negocios en el proceso de admisión y matrícula de una universidad chilena," Chile, 2010.
- [8] Y. Rodríguez Cruz and M. Pinto, "Modelo de uso de información para la toma de decisiones estratégicas en organizaciones de información," ; *Universidad de Granada, Facultad de Biblioteconomía y Documentación, Departamento de Información y Comunicación. Granada, España, 2018.*
- [9] E. Lago Vanegas and L. M. Guerra Cantero, "Sistema de inteligencia de negocios para el apoyo al proceso de toma de decisiones," Valencia Venezuela, 2013.
- [10] J. L. Cano, *Business Intelligence: Competir con informacion*. Banesto Fundacion Cultural, 2007.
- [11] R. Tippelt and H.-J. Lindemann, "El metodo de proyectos," *Ministerio de educacion - gobierno del Ecuador*, El Salvador, München Berlin, 2001.
- [12] H. Muñoz Hernández, R. C. Osorio Mass, and L. M. Zúñiga Pérez, "Inteligencia de los negocios. Clave del éxito en la era de la información," *Clío América*, 2016.
- [13] I. G. Camejo, "Algunas reflexiones sobre el concepto de información y sus implicaciones para el desarrollo de las ciencias de la información," 2000, [Online]. Available: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352000000300005.
- [14] D. Cohen, "CAPÍTULO Introducción a los sistemas de información 1 Principios objetivos de aprendizaje," pp. 2–33, 2009, [Online]. Available:

http://fcaenlinea1.unam.mx/anexos/1144/1144_u7_act1.

- [15] Ralph M Stair and George W Reynolds, *Principios de sistemas de información: enfoque administrativo*. México D.f: Cengage Learning Editores, SA de CV, 2010.
- [16] M. Proaño Castro, S. Orellana Contreras, and I. Martillo Pazmiño, “Los sistemas de información y su importancia en la transformación digital de la empresa actual,” *Espacios*, vol. 39, no. 45, pp. 1–4, 2018, [Online]. Available: <https://www.revistaespacios.com/a18v39n45/a18v39n45p03.pdf>.
- [17] U. Miranda and Z. Acosta, “Fuentes de información para la recolección de información cuantitativa y cualitativa,” *Univ. Nac. San Luis Gonzaga Ica*, pp. 1–20, 2008, [Online]. Available: <https://docs.bvsalud.org/biblioref/2018/06/885032/texto-no-2-fuentes-de-informacion.pdf>.
- [18] Oracle Corporation, “Administración de Datos,” *Copyright*, 2020. <https://www.oracle.com/co/database/what-is-data-management/>.
- [19] E. Prada Madrid, “Los insumos invisibles de decisión: datos, información y conocimiento,” *An. Doc.*, vol. 11, no. 11, pp. 183–196, 2008, doi: 10.6018/analesdoc.11.0.24881.
- [20] M. I. Carrillo Osorio and F. Hernandez Aguirre, “Inteligencia de Negocios. ¿ Qué es y su implementación?,” 2018. .
- [21] Camino a las TIC, “Perspectivas y aplicaciones del análisis de datos,” 2017. <http://anarociotic.blogspot.com/2017/08/perspectivas-y-aplicaciones-del.html> (accessed Sep. 03, 2021).

- [22] L. Muñiz, “El Reporting como sistema de Información para mejorar los resultados de su empresa,” [Online]. Available: https://accid.org/wp-content/uploads/2018/11/Sistema_de_reporting_SCG__E.pdf.
- [23] L. E. Silva Solano, “Business Intelligence: un balance para su implementación,” *InnovaG*, no. 3, pp. 27–36, 2017, [Online]. Available: <http://revistas.pucp.edu.pe/index.php/innovag/article/view/19742>.
- [24] I. Sánchez Sánchez, “Conceptos básicos de la metodología de la investigación,” p. 6, 2015, [Online]. Available: https://repository.uaeh.edu.mx/bitstream/bitstream/handle/123456789/16701/LEC_T133.pdf?sequence=1&isAllowed=y.
- [25] Conexiónsan, “¿Con cuál metodología de business intelligence debemos trabajar? | Gestión de Proyectos | Apuntes empresariales | ESAN,” 2016. <https://www.esan.edu.pe/apuntes-empresariales/2016/08/con-cual-metodologia-de-business-intelligence-debemos-trabajar/> (accessed Sep. 02, 2021).
- [26] D. Bernabeu R and M. García Mattío, “Hefesto | Proceso ETL |.” <https://troyanx.com/Hefesto/proceso-etl.html>.
- [27] J. L. Carhuallanqui Bastidas, “Diseño de una solución de inteligencia de negocios como herramienta de apoyo a la toma de decisiones en el área de ventas de la empresa farmacéutica Dispefarma,” 2017.
- [28] J. Holloran, “Specification or Prototype-Driven Design: Which One is Right for You?,” *LinkedIn*, 2019. <https://www.linkedin.com/pulse/specification-prototype-driven-design-which-one-right-holloran> (accessed Sep. 02, 2021).

- [29] E. Curry, "The Big Data Value Chain: Definitions, Concepts, and Theoretical Approaches," *New Horizons a Data-Driven Econ. A Roadmap Usage Exploit. Big Data Eur.*, pp. 29–37, Jan. 2016, doi: 10.1007/978-3-319-21569-3_3.
- [30] C. Pérez López and D. Santín González, *Minería de datos. Técnicas y herramientas*. Madrid España: Thomson Ediciones Paraninfo, S.A., 2007.
- [31] R. Espinosa, "Cubos OLAP (On-Line Analytic Processing)," *DataPrix*, 2010. <https://www.dataprix.com/es/blog-it/respinosamilla/cubos-olap-line-analytic-processing>.
- [32] E. Software, "Sistemas OLTP: procesamiento, administración y mantenimiento de transacciones - Evaluando Software," 2021. <https://www.evaluandosoftware.com/sistemas-oltp-procesamiento-administracion-mantenimiento-transacciones/> (accessed Sep. 03, 2021).
- [33] PC-Solucion, "Diferencias entre OLTP y OLAP," 2018. <https://pc-solucion.es/2018/04/23/diferencias-entre-oltp-y-olap/>.
- [34] Pentaho Corporation, "Pentaho Data Integration User Guide," 2012, [Online]. Available: https://docs.huihoo.com/pentaho/pentaho-business-analytics/4.8/pdi_user_guide.pdf.
- [35] A. Salinas, "Introducción a Pentaho (parte 1 de 2)," 2008. <https://gravitar.biz/bi/introduccion-pentaho-parte-1/>.
- [36] H. Kerali. G.R, D. McMullen, and J. . Odoki, "HDM - 4 Volume two Applications Guide."

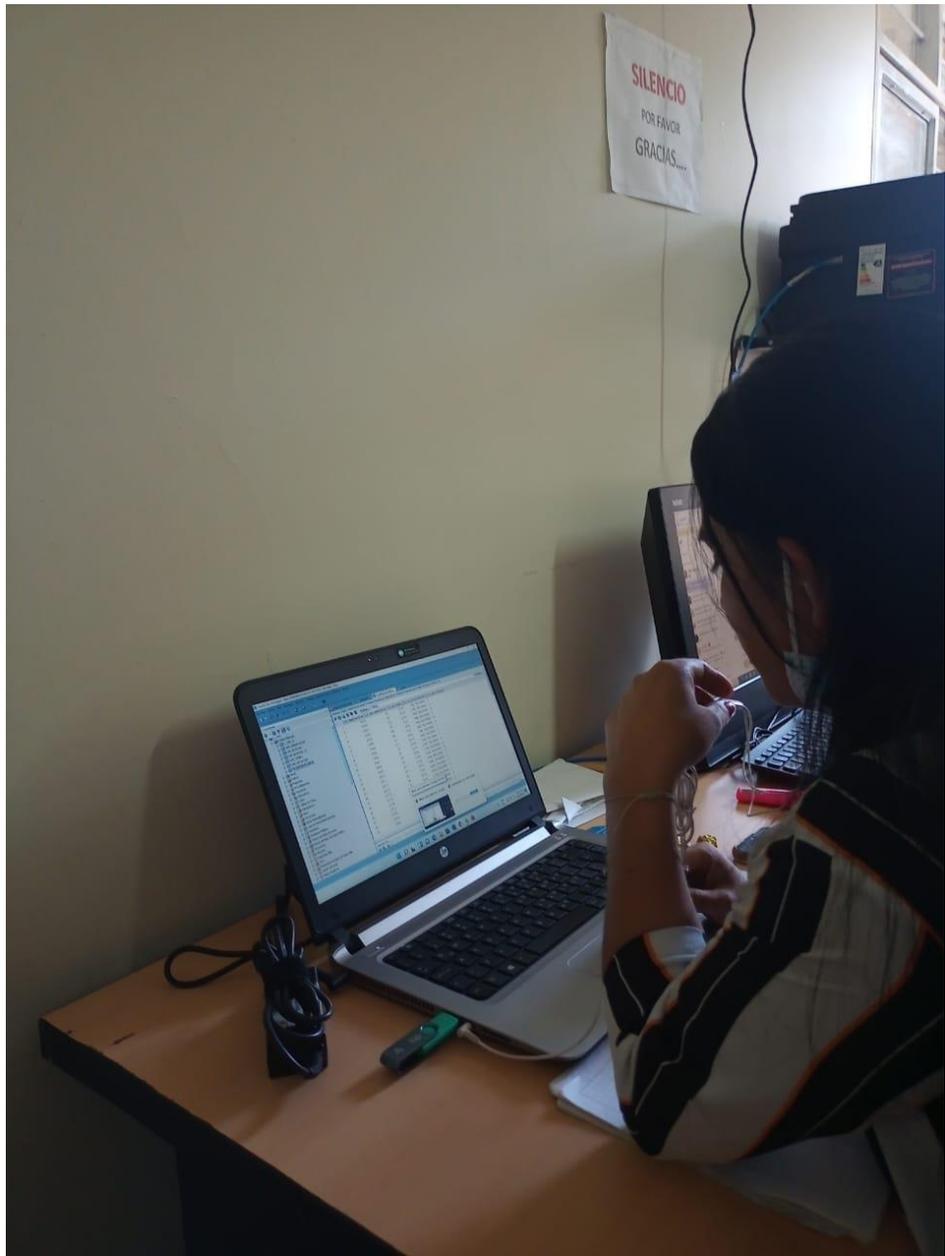
- [37] R. D. Prieto Morales, C. J. Meneses Villegas, and V. R. Vega Zepeda, "GMM-BI: Una guía metodológica para mejorar la madurez organizacional en inteligencia de negocios," Chile, 2014.
- [38] G. Rivera Pleitez, Y. S. Ventura Aguilar, and S. A. Cornejo Hernández, "Diseño de modelo tecnológico para el uso de big data en el análisis y visualización de información para la pequeña y mediana empresa," El Salvador, 2017.
- [39] C. Carreño González, "Gestión del conocimiento en el marco de la innovación en el programa de derecho de la facultad de ciencias jurídicas de la universidad de Manizales: Plan de acción," Universidad EAFIT, 2016.
- [40] E. Medina la Plata, *Business Intelligence - una guía practica*. Lima: Universidad Peruana de Ciencias Aplicadas S.A.C, 2012.
- [41] S. Marius Roşu and M. Guran, "Diez pasos para incrementar el conocimiento para una gestión eficiente del capital intelectual en la empresa Business Intelligence," Romania, 2008.
- [42] O. In Lih, S. Pei Hwa, and S. F. Wong, "A Five-Layered Business Intelligence Architecture," Salengor, Malasya, 2011.
- [43] R. Hernández Sampieri and C. P. Mendoza Torres, *Metodología de la investigación: las rutas cuantitativas cualitativas y mixtas*. Ciudad de Mexico: McGRAW-Hill Interamericana Editores, S.A. de C.V.
- [44] H. Fix Fierro, M. Padrón Innamorato, A. Pozas Loyo, C. Silva Forné, and A. A. Suárez Ávila, *Manual de sociología de derecho*. Ciudad de Mexico: D.R Fondo de

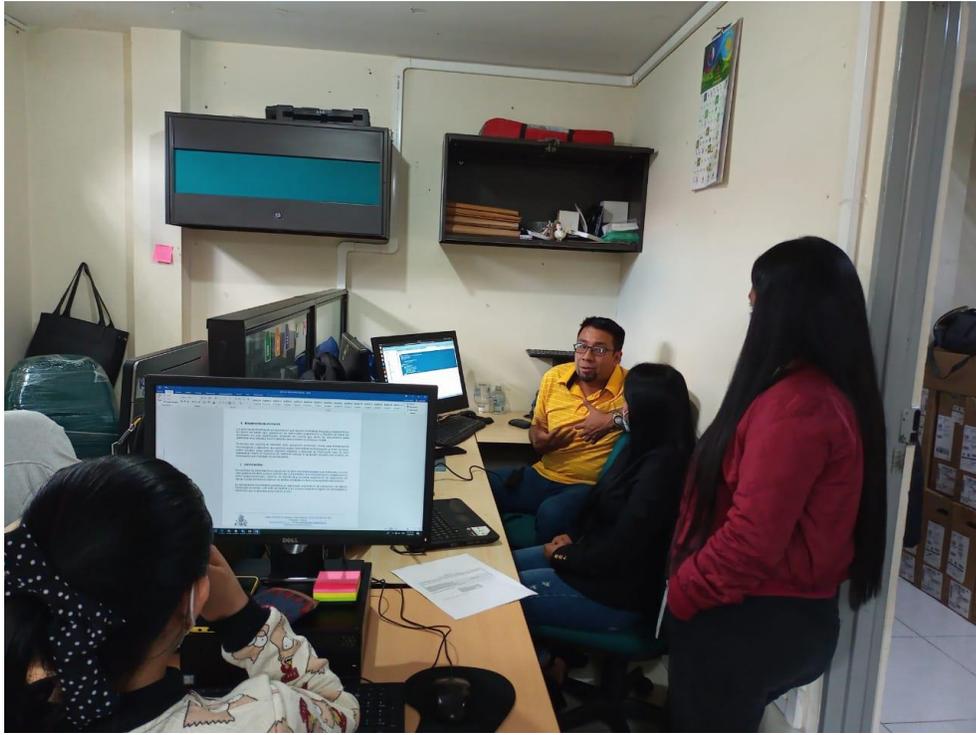
Cultura Economica, 2018.

- [45] C. Pesquera, "carlospesquera," 2018. <https://carlospesquera.com/que-son-las-tablas-de-hechos-y-de-dimension/>.

ANEXOS

Fotografía. Entrega final del producto





Fotografía. Certificación del cumplimiento del desarrollo del producto

	ASOCIACIÓN INDÍGENA DEL CAUCA Resolución 083 del 15 de diciembre de 1997 Dirección General de Asuntos Indígenas del Ministerio del Interior NIT: 817.061.723.3			
	CONSTANCIA LABORAL			
Código:	GDTH-F-19	Versión:	1.1	

AIC-TH-153

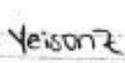
**EL SUSCRITO LÍDER DE TALENTO HUMANO
DE LA ASOCIACIÓN INDÍGENA DEL CAUCA AIC EPS-I**

CERTIFICA:

Que la estudiante **CIELO OTERO CAICEDO** identificada con Cédula de Ciudadanía N.º 109601217 expedida en Morales (Cauca), realizó satisfactoriamente la práctica empresarial con **LA ASOCIACION INDIGENA DEL CAUCA, AIC-EPS-I**, dando cumplimiento a los objetivos planteados en el proyecto titulado **DESARROLLO DE UNA GUIA METODOLOGICA COMO HERRAMIENTA DE APOYO EN LA ELABORACION DE UNA ARQUITECTURA DE INTELIGENCIA DE NEGOCIOS EN RELACION AL AREA DE SALUD DEL REGIMEN SUBSIDIADO Y CONTRIBUTIVO EN LA ASOCIACION INDIGENA DEL CAUCA AIC EPS-I**, de acuerdo al **CONVENIO N° 0039** suscrito entre la asociación indígena del cauca AIC EPS-I y la corporación universitaria autónoma del cauca

La presente certificación se firma en Popayán Cauca, a los 10 días del mes de marzo de 2022, a solicitud de la interesada


MANUEL LIBARDO CUETIA RAMOS
ASOCIACION INDIGENA DEL CAUCA
AIC EPS-I
LIDER DE TALENTO HUMANO

Elaboró:	Yelson Zemanate Burbano – Auxiliar de Talento Humano	
Revisó:	Manuel Libardo Cuetia Ramos - Líder de Talento Humano	

	Calle 1 # 4-66 Barrio Vásquez Cobo - Popayán Cauca Celular: 3102186472 - 3156032943 Email: talentohumano@aicsalud.org.co	
	Página 7 de 9	