

UNIVERSIDAD POLITÉCNICA SALESIANA

SEDE QUITO

CARRERA:

INGENIERÍA DE SISTEMAS

**Trabajo de titulación previo a la obtención del título de:
INGENIEROS DE SISTEMAS**

TEMA:

**IMPLEMENTACIÓN DE UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS
PARA EL ANÁLISIS DE DATOS RELACIONADOS CON LOS PROYECTOS DE
SOFTWARE DEL SECTOR PÚBLICO EN EL ECUADOR EN LA ÚLTIMA
DÉCADA**

AUTORES:

**JOSUE FRANCISCO COBA MEDINA
ESTALIN ALBERTO REVELO PAZ**

TUTOR:

FRANKLIN EDMUNDO HURTADO LARREA

Quito, julio del 2021

CESIÓN DE DERECHOS DE AUTOR

Nosotros Coba Medina Josue Francisco y Revelo Paz Estalin Alberto, con documentos de identificación N° 1723980387 y N° 1720698867, manifestamos nuestra voluntad y cedemos a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que somos autores del trabajo de titulación intitulado: IMPLEMENTACIÓN DE UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS PARA EL ANÁLISIS DE DATOS RELACIONADOS CON LOS PROYECTOS DE SOFTWARE DEL SECTOR PÚBLICO EN EL ECUADOR EN LA ÚLTIMA DÉCADA, mismo que ha sido desarrollado para optar por el título de: ingenieros de sistemas, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En aplicación a lo determinado en la Ley de Propiedad Intelectual, en nuestra condición de autores nos reservamos los derechos morales de la obra antes citada. En concordancia, suscribimos este documento en el momento que hacemos entrega del trabajo final en formato digital a la Biblioteca de la Universidad Politécnica Salesiana.

Quito, julio del 2021



.....
Coba Medina Josue Francisco

CI: 1723980387



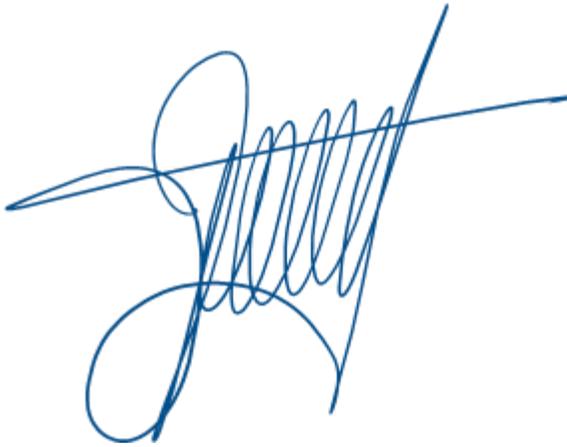
.....
Revelo Paz Estalin Alberto

CI: 1720698867

**DECLARATORIA DE COAUTORÍA DEL DOCENTE
TUTOR/A**

Yo declaro que bajo mi dirección y asesoría fue desarrollado el Proyecto Técnico, IMPLEMENTACIÓN DE UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS PARA EL ANÁLISIS DE DATOS RELACIONADOS CON LOS PROYECTOS DE SOFTWARE DEL SECTOR PÚBLICO EN EL ECUADOR EN LA ÚLTIMA DÉCADA, realizado por Coba Medina Josue Francisco y Revelo Paz Estalin Alberto, obteniendo un producto que cumple con todos los requisitos estipulados por la Universidad Politécnica Salesiana, para ser considerado como trabajo final de titulación.

Quito, julio del 2021.

A handwritten signature in blue ink, consisting of a series of loops and a long horizontal stroke extending to the right.

.....

Franklin Edmundo Hurtado Larrea

CI: 1713382016

DEDICATORIA

A Dios por ser mi guía espiritual, quien me ha enseñado ser mejor persona y sobre todo amar a mi prójimo.

A mis padres Francisco y Rosa que han sido fuente de inspiración en cada etapa de mi vida, quienes con sus enseñanzas y consejos me han inculcado valores para ser una persona honrada.

Finalmente, a mi hermana, quien ha sido una compañera incondicional, consejera y guía en mi vida.

Josue

DEDICATORIA

A mis padres por haberme brindado su cariño y apoyo incondicional, al resto de mi familia que ha compartido conmigo su paciencia y generosidad.

A mi novia por su apoyo en el camino de la vida.

A Dios por brindarme la fuerza y la esperanza de un nuevo comienzo en el diario por venir.

AGRADECIMIENTO

Agradecemos a la Universidad Politécnica Salesiana, en especial a la carrera de Ingeniería de Sistemas por formarnos profesionalmente, brindándonos las herramientas necesarias para conquistar metas laborales, sin perder el punto de vista humanitario.

Al Ing. Franklin Hurtado Larrea, por brindarnos su guía, conocimiento y tiempo durante el ciclo de vida del presente proyecto y en la carrera universitaria.

A Dios por brindarnos de su sabiduría y poder superar obstáculos tanto en el presente proyecto como en la vida universitaria.

A nuestros amigos y compañeros de clases quienes han sido fuente de apoyo en el transcurso de la carrera.

*Josue Coba Medina
Estalin Revelo Paz*

ÍNDICE GENERAL

INTRODUCCIÓN	1
ANTECEDENTES	2
PROBLEMA DE ESTUDIO	3
JUSTIFICACIÓN	4
OBJETIVOS	5
OBJETIVO GENERAL	5
OBJETIVOS ESPECÍFICOS	5
MARCO METODOLÓGICO.....	6
METODOLOGÍA KIMBALL PARA EL DESARROLLO DEL DATA	
WAREHOUSE	6
Planificación del proyecto:	6
Definición de los requerimientos del negocio:	6
Modelado dimensional:	6
Diseño e implementación de la técnica de ETL (Extract-Transform-Load):	6
Diseño de la arquitectura técnica.....	6
MODELO DE DESARROLLO INCREMENTAL (POR ETAPAS).....	7
Especificación conceptual	7
Análisis de requisitos.....	7

Diseño inicial.....	7
Codificación y Depuración.....	7
METODOLOGÍA DE PRUEBAS AL DATA WAREHOUSE	8
Calidad de los datos.....	9
Seguridad.....	9
Interfaz de Usuario	9
MARCO REFERENCIAL.....	10
SISTEMA DE CONTRATACIÓN PÚBLICA DEL ECUADOR.....	10
CAPÍTULO 1.....	11
1. MARCO TEÓRICO	11
1.1. INTELIGENCIA DE NEGOCIOS	11
1.1.1 Descripción	11
1.1.2 Beneficios	11
1.1.3 Proceso ETL (Extract, Transform, Load)	12
1.1.4 Reportes	13
1.2 METODOLOGÍA KIMBALL PARA EL DESARROLLO DEL DATA WAREHOUSE	14
1.2.1 Modelo de desarrollo Incremental aplicado a la metodología Kimball.	16
1.2.2 Modelo multidimensional de base de datos (MDB).	19
1.2.3 Modelo estrella	19

1.2.4	Tabla de hechos	19
1.2.5	Data Warehouse	19
1.2.6	Data Mart	20
1.2.7	Cubos OLAP (On-line Analytical Processing).....	20
1.3	METODOLOGÍA DE PRUEBAS AL DATA WAREHOUSE	20
1.3.1	Calidad	20
1.3.2	Seguridad	21
1.4	HERRAMIENTAS	21
1.4.1	PDFElement.....	21
1.4.2	XAMPP.....	22
1.4.4	Pentaho Community Edition.....	22
1.4.6	Schema Workbench	23
1.4.7	Saiku Analytics	23
CAPÍTULO 2.....		24
2.	DISEÑO Y CONSTRUCCIÓN.....	24
2.1	ESPECIFICACIONES DE REQUERIMIENTOS	24
2.2	DESCRIPCIÓN DE LA SOLUCIÓN	25
2.3	IMPLEMENTACIÓN DEL PROCESO ETL.....	26
2.3.1	Extracción.....	26
2.3.2	Transformación.....	27

2.3.3	Carga.....	30
2.4	ANÁLISIS Y DISEÑO DEL DATA WAREHOUSE.....	33
2.4.1	Dimensiones	34
2.4.2	Tabla de hechos	35
2.4.3	Schema Workbench.....	37
2.4.4	Versiones de depuración.....	39
2.4.5	Medidas.....	40
2.4.6	Análisis de los datos	40
2.5	CREACIÓN DE REPORTES.....	41
2.5.1	Reporte estático.....	41
2.5.2	Ad-hoc	42
2.6	DESCRIPCIÓN DEL PROCESO DE PRUEBAS AL DATA WAREHOUSE	43
2.6.1	Seguridad	44
2.6.2	Prueba de integración.....	46
2.6.3	Prueba de rendimiento	47
2.6.4	Prueba de disponibilidad.....	47
2.6.5	Interfaz de usuario	48
	CAPÍTULO 3.....	49
3.	EXPLOTACIÓN	49

3.1	DESCRIPCIÓN DEL PROCESO DE DATOS AL DATA WAREHOUSE	
		49
3.1.1	Datos almacenados en el Sistema de Contratación Pública del Ecuador	
		49
3.1.2	Depuración.....	51
3.1.3	Proceso ETL a la base de datos	53
3.1.4	Creación del Data Warehouse.....	54
3.2	DESCRIPCIÓN DEL ACCESO A REPORTES	56
3.3	LEVANTAMIENTO DE LA SOLUCIÓN EN EL SERVIDOR	58
3.3.1	Requerimiento.....	58
3.3.2	Instalación y configuración.....	60
3.4	HALLAZGOS DE INFORMACIÓN	61
	CONCLUSIONES	67
	RECOMENDACIONES.....	69
	REFERENCIAS BIBLIOGRÁFICAS	70
	ANEXOS	72

ÍNDICE DE TABLAS

Tabla 1: Descripción de estándar de datos.....	28
Tabla 2: Descripción base de datos bi_coba_revelo.....	30
Tabla 3: Matriz de pruebas	44
Tabla 4: Perfiles y roles en la inteligencia de negocios	57
Tabla 5: Descripción del servidor.....	58
Tabla 6: Requisitos mínimos Hardware Pentaho Server	59
Tabla 7: Requisitos mínimos Software Pentaho Server.....	59
Tabla 8: Requisitos mínimos XAMPP.....	60

ÍNDICE DE FIGURAS

Figura 1: Diagrama Modelo Incremental	8
Figura 2: Diagrama ciclo de vida dimensional del negocio	15
Figura 3: Descripción estructura base de datos.....	32
Figura 4: Diagrama físico de tabla de hechos.....	35
Figura 5: Diagrama físico de la estructura estrella	36
Figura 6: Esquema del Data Warehouse.....	38
Figura 7: Reportes Estáticos	42
Figura 8: Reportes Ad-hoc.....	43
Figura 9: Creación de perfiles, roles y permisos de analista	45
Figura 10: Creación de perfiles, roles y permisos de público.....	46
Figura 11: Sistema Oficial de Contratación Pública.....	49
Figura 12: PDF descargado del Sistema Oficial de Contratación Pública	51
Figura 13: Formulario de extracción mediante PDFElement	52
Figura 14: Visualización del tipo de dato en Pentaho Data Integration (PDI)	53
Figura 15: Proceso de carga en Pentaho Data Integration (PDI).....	54
Figura 16: Dimensiones y medidas del Data Warehouse	55
Figura 17: Publicación del esquema	56
Figura 18: Dimensiones y medidas para realizar el análisis del hallazgo A.....	61
Figura 19: Reporte Hallazgo A.....	62
Figura 20: Dimensiones y medidas para realizar el análisis del hallazgo B.....	63
Figura 21: Reporte hallazgo B	64
Figura 22: Dimensiones y medidas para realizar el análisis del hallazgo C.....	65
Figura 23: Reporte Hallazgo C	66

RESUMEN

El presente documento expone una solución de inteligencia de negocios para el análisis de datos relacionados con los proyectos de software del sector público del Ecuador en la última década (2010-2020), los datos se encuentran en un repositorio digital en formato PDF, este proceso (sin la solución) retrasa la recolección y el tratamiento de la información contenida en estos archivos, se aplicó el proceso ETL (Extract-Transform-Load), en apoyo a la extracción, depuración y estandarización de los datos, a su vez se cargó esta información en un Data Warehouse, se analizó las dimensiones y medidas que lo conforman por medio de la inteligencia de negocios en base a factores de los proyectos de software, a esta información se accede mediante la herramienta Pentaho Community por medio de este se consigue la creación y descarga de reportes con información relevante para los ciudadanos.

ABSTRACT

This document exposes a business intelligence solution for the analysis of data related to software projects in the public sector of Ecuador in the last decade (2010-2020), the data are in a digital repository in PDF format, this process (without the solution) delays the collection and processing of the information contained in these files, the ETL process (Extract-Transform-Load) was applied, The ETL (Extract-Transform-Load) process was applied to support the extraction, debugging and standardization of the data, in turn this information was loaded into a Data Warehouse, the dimensions and measures that make it up were analyzed by means of business intelligence based on factors of the software projects, this information is accessed through the Pentaho Community tool through which the creation and downloading of reports with relevant information for citizens is achieved.

INTRODUCCIÓN

El Sistema de Contratación Pública del Ecuador, es el ente responsable de establecer los principios y regulaciones de los procesos de contratación realizadas por entidades públicas, actualmente almacena su información en un repositorio, el cual dificulta el análisis de los datos de proyectos de software.

Al compilar la información de proyectos realizados por varias entidades públicas en distintos periodos de tiempo dificulta el análisis de los datos, es la razón de ser del presente proyecto de titulación, este hace énfasis en el análisis de los proyectos de software durante la última década (2010-2020) por medio de la inteligencia de negocios.

Una de las principales motivaciones de la realización del presente proyecto es la implementación de una herramienta que aporte a los ciudadanos analizar los factores de los proyectos de software sin importar el nivel de conocimiento de los mismos, así también el proporcionar la destreza de descubrir información por sí mismos.

En el capítulo uno se detalla el sustento teórico que se emplea como referente para el presente proyecto.

En el capítulo dos se presenta la información sobre el diseño y construcción tanto del proceso ETL como el Data Warehouse, a su vez se realiza la implementación de la herramienta Saiku Analytics para la generación de reportes a través de Pentaho Community. El capítulo tres contiene la explotación de los datos, así como la descripción de los procesos de datos al Data Warehouse, la ejecución de pruebas para validar la solución y los hallazgos que se lograron.

ANTECEDENTES

El Sistema de Contratación Pública del Ecuador es un repositorio digital donde las entidades públicas realizan la entrega de información acerca de diversos proyectos que se han ofertado en el transcurso de años, contiene información con cierto detalle acerca del ciclo de vida de los proyectos en forma de pliegos digitales, a su vez cuenta con información general y resumida de los parámetros más importantes en dicho proyecto.

Los datos generados del repositorio son de importancia para los ciudadanos que deseen o necesiten dicha información, con lo cual es prioritario una manera ágil y sencilla de manejar estos grandes volúmenes de datos sin perder la calidad de los mismos.

El manejo y análisis de estos datos no es el más adecuado, esto se debe a los escasos medios para la obtención de los mismos, sin mencionar que no se puede realizar un análisis que permita hallar resultados significativos.

La inteligencia de negocios solventa las necesidades planteadas, es por ello que hoy en día la implementación de la misma es un recurso que grandes empresas buscan, el principal motivo es el manejo e interpretación de los datos y el potencial que esto representa para los usuarios.

PROBLEMA DE ESTUDIO

El desarrollo, contratación y adquisición de software de TICS (Tecnologías de información y Comunicación) en el Ecuador ha tenido un crecimiento exponencial en el transcurso de la última década esto en referencia al sector público. (Andrade et al., 2012).

El Sistema de Contratación Pública del Ecuador es el ente encargado del proceso de proveer información para el desarrollo e implementación de proyectos de software, actualmente los documentos que se generan se encuentran almacenados en un repositorio digital el cual posee limitaciones para realizar un análisis de sus datos, este provee los datos en formatos PDF, a la cual su acceso, búsqueda y descarga se realiza uno a la vez, actualmente no dispone de un lugar en donde sea asequible almacenar datos, analizar e interpretar estos a través de la inteligencia de negocios que permita identificar factores con respecto al tiempo y costo que influyen en los proyectos para su desarrollo.

JUSTIFICACIÓN

El Sistema de Contratación Pública del Ecuador cada año propone varios proyectos de software por lo cual distintas entidades aplican a estos mediante un concurso de méritos, este se adjudica a la entidad que cumpla con los requisitos, cada uno de los factores que integran un proyecto de software pueden llegar a ser muy variables o cambiar su curso de desarrollo dependiendo de qué componentes lo conforman, tales como el tiempo, el costo, esto se da puesto que se encuentran en un entorno impredecible. Estos factores determinan el rumbo del proyecto y su posible éxito o fracaso. Hoy en día las empresas han transformado los procesos del negocio de una manera sencilla y direccionada a la gestión de estos con base en la información.

El organizar, depurar y almacenar los datos que provee el Sistema de Contratación Pública del Ecuador en un Data Warehouse posibilita la disposición de la información de manera precisa, confiable y rápida, a través de una herramienta de BI que realice una gestión ordenada de estos, con ello se analiza e identifica factores de importancia que se interpretan a través de la información previamente tratada, de manera que se tomen decisiones en base a los requerimientos de los proyectos con respecto al tiempo y el costo. Por otro lado, se obtendrá un repositorio de información del Estado en base a los proyectos realizados cada año, cada uno con requerimientos claves que se llevaron a cabo en cada proyecto, logrando así una mejor planificación, adquisición, gestión de recursos y la calidad del producto final, esto servirá de guía para empresas, consultores o entidades que se encuentren a la vanguardia en TICS.

OBJETIVOS

OBJETIVO GENERAL

Implementar una solución de inteligencia de negocios para el análisis de datos relacionados con los proyectos de software del sector público en el Ecuador de la última década.

OBJETIVOS ESPECÍFICOS

- Analizar e identificar los datos que provee el Sistema de Contratación Pública del Ecuador para una solución de inteligencia de negocios que apoye a la gestión de los proyectos de software.
- Aplicar el proceso de ETL para el manejo y organización de estos datos que se encuentran en el repositorio digital del Sistema de Contratación Pública del Estado.
- Implementar un repositorio multidimensional (Data Warehouse) para el manejo de los datos y manejo en la herramienta de BI.
- Generar reportes con la información obtenida del análisis de la herramienta de BI la cual sirva de guía o punto de partida de futuros análisis.

MARCO METODOLÓGICO

METODOLOGÍA KIMBALL PARA EL DESARROLLO DEL DATA WAREHOUSE

Por medio de la metodología de Kimball se desarrolló e implementó el Data Warehouse el cual aplica el ciclo de vida dimensional del negocio.

Planificación del proyecto:

El propósito del proyecto es establecer la identificación de los factores relevantes en el proceso de desarrollo, contratación y adquisición de software de TICS en el Ecuador en la última década (2010-2020).

Definición de los requerimientos del negocio:

Se identificó los factores como: tiempo, costo, estado de contratación y la forma de adquisición de software de TICS en el Ecuador última década (2010-2020).

Modelado dimensional:

Se estableció el nivel de granularidad de la información, las dimensiones y medidas del Data Warehouse.

Diseño e implementación de la técnica de ETL (Extract-Transform-Load):

El proceso de extracción se ejecutó por medio del portal web del Sistema de Contratación Pública del Ecuador, la transformación se realizó mediante macros de Excel y la herramienta Pentaho Data Integration (PDI) que contiene procesos de depuración y carga de los datos hacia el repositorio.

Diseño de la arquitectura técnica

Se estableció los requisitos tanto de hardware y software para determinar la arquitectura donde se implementará la solución de inteligencia de negocios, para el hardware se hizo uso del contenedor Tomcat en donde se aloja el módulo de ETL, para crear y

visualizar los reportes se utilizó la de la herramienta Saiku Analytics del software Pentaho Community.

MODELO DE DESARROLLO INCREMENTAL (POR ETAPAS)

Se opta por el modelo de desarrollo incremental el motivo de esta selección es la retroalimentación a través de reuniones programadas por parte del tutor del proyecto, con ello se busca obtener una mejor calidad del producto final, a su vez tener la facilidad de realizar cambios en la solución como se lo plantea en la Figura 1.

Especificación conceptual

Se especificaron los requerimientos esenciales para el proyecto, a la vez de recopilar fuentes de información necesaria, se determinó 3 etapas importantes: ETL, Data Warehouse e inteligencia de negocios.

Análisis de requisitos

Se realizaron proyecciones en cuanto tiempo y herramientas necesarias para cumplir con el proyecto, esta etapa se realizó mediante reuniones periódicas entre el tutor y el equipo de trabajo para definir las metas e iteraciones.

Diseño inicial

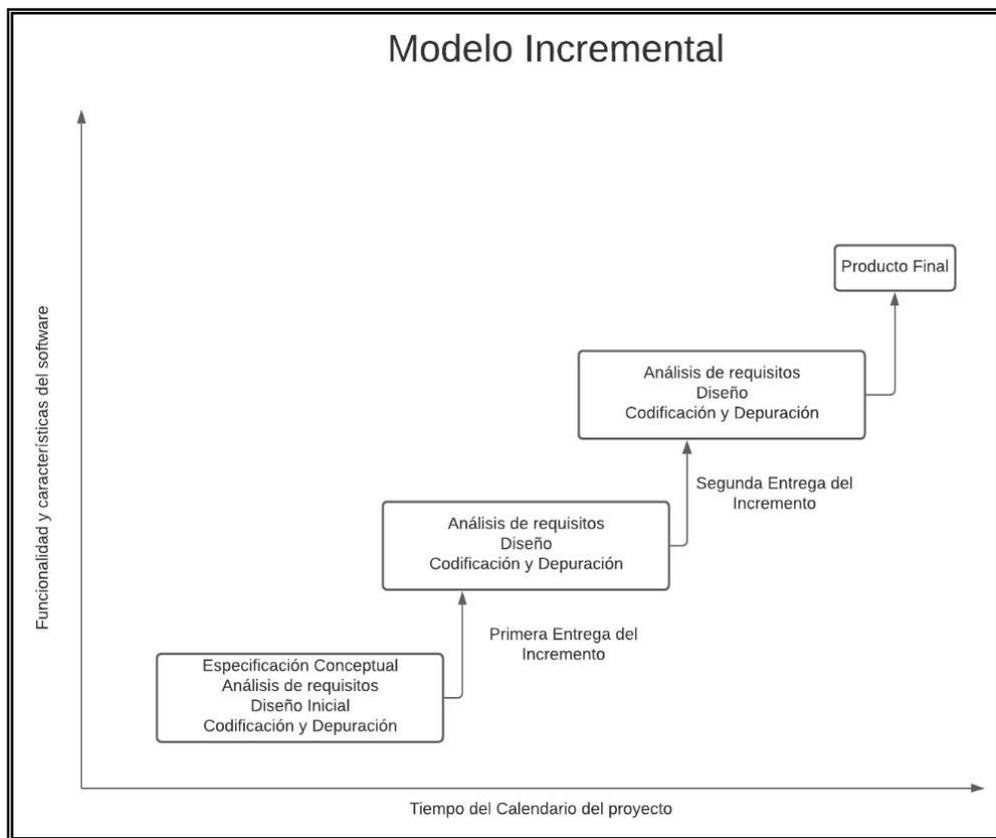
Se construyó la solución, componentes y funcionalidades que se desea obtener en las primeras etapas de desarrollo.

Codificación y Depuración

Se construyó la estructura funcional para la entrega de un primer incremento en este se realizó la limpieza y depuración de datos que es la principal fuente de información.

Figura 1

Diagrama Modelo Incremental.



Nota: Descripción de la ruta de desarrollo en base a un modelo incremental por etapas.

Elaborado por: Josue Coba y Estalin Revelo.

La Figura 1 realiza una referencia a las iteraciones del desarrollo del modelo incremental, así como a sus diferentes etapas que lo conforman.

METODOLOGÍA DE PRUEBAS AL DATA WAREHOUSE

La metodología que se implementó en las pruebas de un Data Warehouse descrita por Doug Vucevic y Wayne Yadow (2012) donde se centra en datos, a su vez hace énfasis en las diferentes etapas que deben cumplir estos como:

Calidad de los datos

Pruebas del sistema de integración. Se validó cada uno de los componentes de la solución de la inteligencia de negocios, el cual se enfoca en los procesos críticos que conforman el ETL.

Prueba de rendimiento. Se validó el rendimiento de la solución en condiciones de trabajo.

Seguridad

Prueba de validación. Se definen roles y perfiles de usuario para la manipulación de los datos de la herramienta Pentaho Community.

Interfaz de Usuario

Prueba de validación. Se define como se presenta la información, esto se realiza por medio de la herramienta Saiku Analytics de Pentaho Community.

MARCO REFERENCIAL

SISTEMA DE CONTRATACIÓN PÚBLICA DEL ECUADOR

El Sistema de Contratación Pública del Ecuador es el ente encargado de determinar los principios y normas para regular los procedimientos de contratación para la adquisición o arrendamiento de bienes, ejecución de obras y prestación de servicios, incluidos los de consultoría que realicen:

- Los organismos y dependencias de las Funciones del Estado.
- Los organismos Electorales.
- Los organismos de Control y Regulación.
- Las entidades que integran el Régimen Seccional Autónomo.
- Los organismos y entidades creados por las Constitución o Ley para el ejercicio de la potestad estatal, para la prestación de servicios públicos o para desarrollar actividades económicas asumidas por el Estado.
- Las personas jurídicas creadas por acto legislativo seccional para la prestación de servicios públicos.
- Las corporaciones, fundaciones o sociedades civiles.

CAPÍTULO 1

1. MARCO TEÓRICO

1.1. INTELIGENCIA DE NEGOCIOS

Según Díaz (2012) se entiende por “inteligencia de negocio al conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización” (p. 18), este modelo logra la integración de diferentes tecnologías y herramientas de análisis, a su vez se toma en cuenta los distintos modelos que poseen los negocios logrando así unificarlos con la información respectiva que cada uno de estos poseen.

1.1.1 Descripción

En la actualidad las empresas generan y manejan un gran volumen de datos éste representa un reto empresarial y surgen preguntas como ¿Qué hacer con estos datos?, ¿Cómo usar estos datos para tomar mejores decisiones?, la inteligencia de negocios satisface esta necesidad, ofrece manejar la información por procesos de extracción, transformación y almacenamiento de datos en un repositorio que integra diferentes tecnologías para el análisis de datos mediante la estadística descriptiva y datos de alta calidad, donde se visualicen estos mediante tableros e informes.

1.1.2 Beneficios

Implementar un sistema de inteligencia de negocios otorga los múltiples beneficios para el negocio, siendo así los principales:

- “Se establecen métricas e indicadores claves de rendimiento (KPI, Key Performance Indicador) e indicadores claves de meta (KGI, Key Goal Indicator) fundamentales para la empresa”. Díaz (2012) (pg. 20). Los KPIs (indicadores claves de rendimiento) ayudan al negocio a mejorar mediante la gestión de sus

procesos, productos y servicios en base a los indicadores los cuales permiten realizar un benchmarking que brinda la oportunidad de contrastar la información con empresas que se encuentren en el mismo segmento del negocio.

- Mejorar la competitividad de la organización como resultado y ser capaces de:
 - a) Diferenciar lo relevante sobre lo superfluo.
 - b) Acceder más rápido a la información.
 - c) Tener mayor agilidad en la toma de las decisiones. Díaz (2012) (pg. 20).

1.1.3 Proceso ETL (Extract, Transform, Load)

El Proceso de extracción, transformación y carga (Extract, Transform, Load) integra diferentes módulos que siguen una serie de procedimientos. Kimball define que la aplicación del proceso ETL es un reto, en este existen condicionantes externos que definen su desarrollo, diseño e implementación, estos dos últimos procesos consumen la mayor cantidad de tiempo y precisión para obtener información íntegra, depurada y confiable.

Extracción. Proceso de recolección de datos de los sistemas de origen o repositorios en los cuales estos se almacenan, este integra subprocesos como:

Integridad referencial de datos. Este subproceso almacena los datos conservando su formato original (cadena de caracteres, entero, fechas, etc.), simultáneamente segrega la información que se extrae.

Calidad de datos. Este subproceso verifica la idoneidad de los datos y valida que las operaciones entre los datos de origen y destino se encuentren relacionados de tal manera que sean idénticos.

Transformación. Proceso que establece un estándar de la información, aquí se identifican diferentes problemas con el formato de los datos, campos nulos e información

faltante. Este proceso agrega valor a los datos por medio de distintos subprocesos que lo conforman como son:

Corrección sintáctica. Subproceso que realiza un análisis lexicográfico en el que se identifica caracteres alfabéticos (letras), alfanuméricos (números) y caracteres especiales, por lo tanto, se depura cada uno de estos de tal manera que la información final sea íntegra y no se pierda coherencia en esta.

“Kimball señala como ejemplo la corrección en los caracteres de índole latina como la ‘ñ’” (Bermeo y Simbaña, 2020) (pg. 40).

Estandarización. Subproceso que transforma los caracteres alfabéticos a un formato establecido sea este de mayúsculas o minúsculas y los generaliza como una cadena de caracteres, al igual que las cadenas alfanuméricas en cuanto a caracteres numéricos como (Enteros, punto flotante, etc.) (Bermeo y Simbaña, 2020) (pg. 40).

Carga. Proceso que continua a partir de la estandarización de los datos donde su estructura se organiza, convirtiéndose en el más importante y crítico para el almacenamiento de información.

Validación de los datos. Subproceso que corrobora que los datos de la fuente sean semejantes tanto en sus filas y columnas del esquema estructural como en su destino.

Carga de dimensiones. Subproceso que almacena los datos en las distintas dimensiones que componen el Data Warehouse, este se basa en el anterior subproceso, a su vez se realiza la carga de la tabla de hechos que la conforma. (Bermeo y Simbaña, 2020) (pg. 41).

1.1.4 Reportes

Según Díaz (2012), “Un informe es un documento a través del cual se presentan los resultados de uno o varios procesos de negocios. Suele contener texto acompañado de

elementos como tablas o gráficos para agilizar la comprensión de la información presentada” (p. 137). Considerando toda la definición expuesta por el autor Díaz el análisis se debe respaldar en cuanto a los hallazgos que se obtiene al usar la herramienta de BI, por consiguiente, el reporte es la parte final y más importante de la misma, donde la información encontrada se presentada de manera clara y concisa a los usuarios con el fin de interpretar estas características en su beneficio.

Estático Según Díaz (2011), “Tienen un formato preestablecido inamovible” (p. 139). Este tipo de reporte contiene como característica principal que el usuario visualice los resultados del análisis de datos, se utiliza con mayor frecuencia en negocios que se encargan de enviar estadísticas a sus clientes que desconocen de ¿cómo realizar el análisis de los datos?, estos optan por la contratación de este servicio y únicamente reciben el producto final.

Ad-hoc Según Díaz (2012), “Son creados por el usuario final a partir de la capa de metadatos que permite usar el lenguaje de negocios propio” (p. 139). El reporte Ad-hoc tiene como objetivo la manipulación de datos para identificar nueva información relevante para el usuario.

1.2 METODOLOGÍA KIMBALL PARA EL DESARROLLO DEL DATA WAREHOUSE

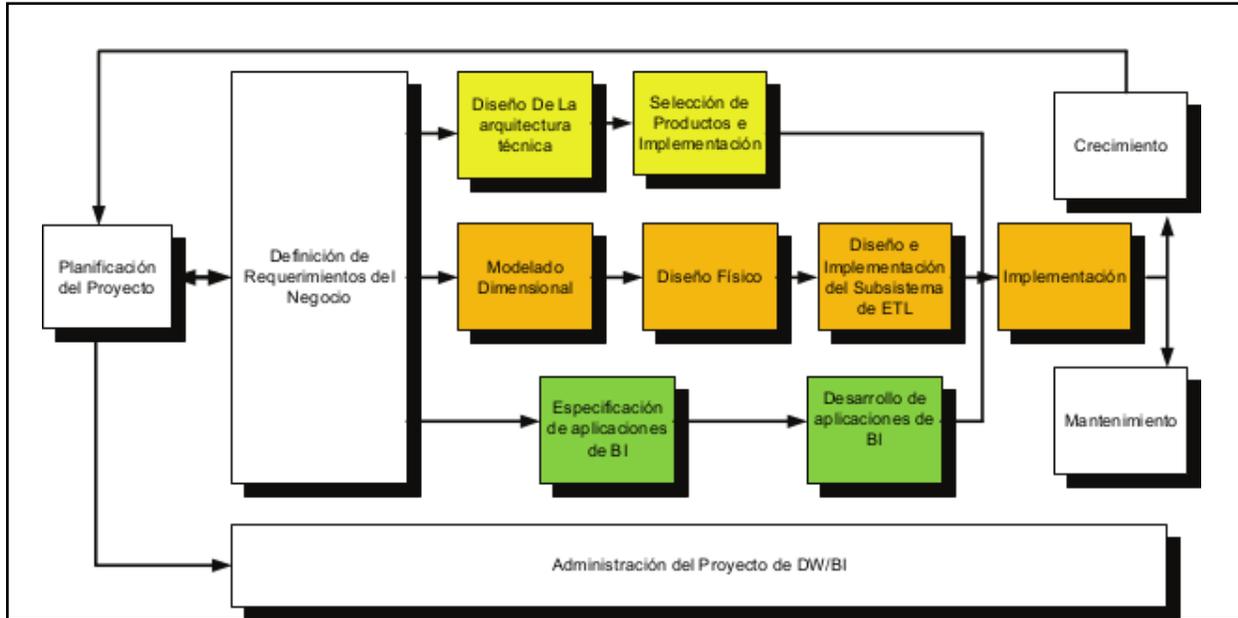
La metodología Kimball indica que el desarrollo de un Data Warehouse esté basado en el ciclo de vida dimensional del negocio, se representa en la Figura 2, este presenta cuatro principios fundamentales (Kimball & Ross, 2013).

- Agregar valor al negocio.
- Diseñar y construir una infraestructura única.
- Realizar entregas incrementales añadiendo valor al negocio.

- Presentar una solución sólida.

Figura 2

Diagrama ciclo de vida dimensional del negocio.



Nota: Descripción del ciclo de vida dimensional del negocio según Kimball. Fuente
:(Kimball & Ross, The Data Warehouse Lifecycle Toolkit, 2013).

Kimball define 6 etapas para el diseño y construcción de la inteligencia de negocio, estas son:

- Planificación del proyecto: determinar el alcance, propósito y solución del mismo.
- Definición de los requerimientos del negocio: establecer los factores clave del negocio.
- Modelado dimensional: define las dimensiones del Data Warehouse, se detalla la información relevante.
- Diseño Físico: identificar la mejor estructura a fin de implementar al modelo dimensional.

- Diseño e implementación de la presentación de datos: implementar el proceso de ETL (Extract, Transform, Load).
- Diseño de la arquitectura técnica: determinar la arquitectura de software y hardware que implementa la inteligencia de negocios.

1.2.1 Modelo de desarrollo Incremental aplicado a la metodología Kimball.

Descripción del Modelo. El modelo de desarrollo incremental iterativo o por etapas comenta sus ideales el profesor Harlan Mills en 1980, esta incorporación a los modelos de desarrollo establecidos demostró ser una revolución en el campo de la ingeniería de software y las bases para las metodologías ágiles modernas (Costanzo, 2020).

A medida que transcurre el tiempo existen cambios notables en el desarrollo tanto de aplicaciones web y móviles haciendo una comparativa de años posteriores con exactitud a mediados de los 2000, el modelo que regía como estándar para el desarrollo de cualquier proyecto de software era el modelo en cascada, no obstante al encontrarse en un entorno en constante evolución y cambios, más si se habla de tecnología, los modelos y estándares deben evolucionar también, es por ello que el modelo incremental es considerado un estándar para el desarrollo de aplicaciones web, móvil, servicios, entre otras (Costanzo, 2020).

El modelo incremental proporciona una retroalimentación por parte del usuario hacia el equipo de desarrollo al finalizar el proceso en cada una de sus etapas, agilizando y fortaleciendo la solución del proyecto, con lo cual se reduce el riesgo de cambios significativos y disminuye el tiempo de implementación.

Compatibilidad de la metodología Kimball aplicando el desarrollo incremental.

Según Kimball (2003), “El Data Warehouse no es nada más que la unión de todos los Data Marts que lo constituyen” (p. 45). Kimball detalla en su metodología que al tratarse de

la construcción del Data Warehouse se reduce gran cantidad de su complejidad al acoplar los distintos Data Marts por medio del desarrollo incremental, esto se debe que al tratarse de construcciones individuales logre compactar en una única unidad donde cada uno de los elementos es rigurosamente tratado, es decir las presentaciones periódicas por parte del equipo de desarrollo hacia el cliente suma un control en la construcción desde el inicio del proyecto, esta metodología reduce el impacto en los cambios en su estructura, diseño o construcción (Sevilla, 2003).

Objetivo. Se centra en una contribución de elementos, con exactitud de dos modelos semejantes, los cuales son el modelo lineal secuencial en conjunto con la construcción de prototipos, es decir la presentación de incrementos secuenciales a los interesados del proyecto con la finalidad de optimizar recursos prioritarios dentro del desarrollo del mismo como son recurso humano, tiempo y costos (Cahuana, 2015). Un punto importante que se debe recalcar es la presencia de secuencias lineales de forma escalonada en un periodo de tiempo establecido en el calendario de actividades o del ciclo de vida del proyecto. Como todo proyecto el primer incremento que se presenta establece las bases y requerimientos del mismo, denomina como núcleo (Cahuana 2015). Plantear el incremento es la parte esencial de este modelo, es su razón de ser, su distintivo de los demás modelos, se realiza pequeñas entregas de partes del software cada parte debe constar con una funcionalidad para ser utilizada, el primer entregable o núcleo es una presentación básica del producto final no obstante debe contar con sus funciones que permite hacer uso del mismo, tras la presentación de cada incremento se desarrolla otro de una forma secuencial por lo que se logra así un producto final robusto, aprobado por los interesados, con un control de calidad óptimo y completamente funcional.

Fases. Pérez (2016) plantea las fases del modelo incremental se considera un modelo flexible, esto quiere decir que puede variar u omitir ciertas etapas dependiendo del tipo de proyecto que se desee realizar. Las etapas que conforman este modelo son:

Requerimientos. En esta etapa se establecen los requerimientos mínimos que debe tener para su funcionalidad, a su vez establecer los objetivos a los cuales debe aspirar el proyecto (Cahuana, 2015).

Definición de las etapas y las iteraciones. Esta etapa determina las tareas que se van a cumplir para alcanzar dicho objetivo, es importante aclarar que el trabajo que se implementará no sigue un orden aleatorio, cada una de ellas tiene un grado de importancia y sincronización entre las demás (Cahuana, 2015).

Diseño y Desarrollo de los incrementos. Etapa que establece las iteraciones del proyecto, esta define las nuevas metas que se deben cumplir para alcanzar el objetivo general, esta es la esencia del modelo que se encuentra en constante cambio y mejora, en este proceso es indispensable para la retroalimentación que se obtiene del cliente, (Cahuana, 2015).

Validación de incrementos. En esta etapa se presentan las nuevas iteraciones con los cambios solicitados con la finalidad de que tanto el cliente como el responsable de la gestión del proyecto queden satisfechos con los resultados, caso contrario lo más recomendable es realizar una revisión de las tareas cumplidas en búsqueda de los factores que puedan representar conflictos o recibir algún tipo de cambio o mejora (Cahuana, 2015).

Integración de incrementos. En esta etapa como su nombre lo indica se integran los incrementos que se realizan, estos al ser ya aprobados por todas las partes interesadas contribuyen al resultado final deseado (Cahuana, 2015).

Entrega del producto. Es la etapa final del proyecto en donde al tener los incrementos validados e integrados se procede a la entrega final del producto (Cahuana, 2015).

1.2.2 Modelo multidimensional de base de datos (MDB).

El modelo multidimensional, es una técnica de diseño lógico este presenta la información de los procesos de forma estándar, cumple con el proceso de recuperación adecuada de los datos que se encuentran ligados al negocio. Los datos se almacenan como hechos y dimensiones en un modelo de datos relacional, siendo así el objetivo principal con el fin de establecer una interacción del usuario con los datos de manera sencilla (Domínguez, 2008).

1.2.3 Modelo estrella.

El modelo estrella de un Data Warehouse y Data Mart se conforma de una o más tablas de hechos, a su vez este modelo brinda la oportunidad de compartir dimensiones y minimiza el manejo de tablas, al mismo tiempo estas contienen una mayor cantidad de información.

1.2.4 Tabla de hechos.

Es una tabla central de un esquema dimensional, esta almacena los datos que da un valor agregado al negocio por lo cual se conservan características que facilita las consultas de la información histórica, su almacenamiento en las tablas llamadas “lock up”, contienen un conjunto de valores relacionados a una dimensión en particular.

1.2.5 Data Warehouse.

Según Kimball, un Data Warehouse es una base de datos corporativa, desnormalizada no modelada en base a los sistemas transaccionales, que es alimentada por todos los departamentos que conforman la organización y usualmente están modelados en base a esquemas multidimensionales como son los modelos de estrella o de copo de nieve. (Kimball, 2013)

1.2.6 Data Mart.

Según Kimball el Data Warehouse es una versión extendida de los Data Mart, en el cual el Data Warehouse es departamental y este se beneficia de un nivel de granularidad mayor este se centra en un proceso, departamento en la organización.

1.2.7 Cubos OLAP (On-line Analytical Processing).

OLAP es un proceso analítico en línea que se caracteriza por ser ágil y flexible al manejar los datos, contempla varios objetivos como el recuperar y manipular los mismos mediante consultas al modelo multidimensional, expone en forma de respuesta la información solicitada, por medio del análisis multidimensional se puede procesar grandes cantidades de datos.

Cubos MOLAP (Multidimensional OLAP). Es un modelo de datos multidimensional, este es muy compacto para conjuntos de datos de pocas dimensiones ya que su almacenamiento y extracción de los mismos se logra con una pre estructuración de los datos agregados por medio de una réplica de su fuente, esta lee datos pre compilados, a su vez esta requiere de cálculos intensivos de compilación.

Cubos ROLAP (Relational OLAP). Es una versión de OLAP que se centra en el trabajo con base de datos relacionales, optimizadas para extraer información mediante consultas, esta posee gran similitud con MOLAP, evita la redundancia de datos.

1.3 METODOLOGÍA DE PRUEBAS AL DATA WAREHOUSE

1.3.1 Calidad

Pruebas del sistema de integración. Validar los componentes que conforman la solución de la inteligencia de negocios, se lleva a cabo durante la etapa de desarrollo mediante el proceso de elementos más críticos de esta siendo así la lógica ETL, reglas del negocio, etc. (Toledo, 2017).

Asegura que no se produzcan inconvenientes al momento de ejecutar el trabajo, con ello garantiza la construcción y el éxito al desplegar la solución, con el fin de ejecutar los procesos para el uso de la herramienta. (Toledo, 2017)

Prueba de rendimiento. Validar el rendimiento de la solución en condiciones de trabajo, se consideran factores importantes como el hardware y software y la complejidad de las consultas. (Toledo, 2017)

Prueba de disponibilidad. Asegurar que los datos se encuentren disponibles en las plataformas establecidas en el hardware y software para los usuarios. (Toledo, 2017)

1.3.2 Seguridad

Segregación Funcional. La segregación funcional valida las funciones y responsabilidades para que el usuario participe en la interacción con la herramienta de BI. (Alfaro y Varga, 2016)

Interfaz de Usuario. La interfaz de usuario presenta la forma en cómo se va a consumir y visualizar la información en la herramienta, la cual provee de una forma de interacción con los datos. (Ramos, 2017)

1.4 HERRAMIENTAS

1.4.1 PDFElement

PDFElement es una herramienta de software que realiza todo tipo de tareas con archivos, siendo las principales: editar archivos, convertir archivos PDF a otros formatos, crear PDF a partir de otro tipo de archivos, combinar diferentes archivos, cuenta con un potente sistema OCR (Optical Character Recognition), su principal función es reconocer el texto de los archivos que editamos, por lo que posibilita el modificar o extraer cualquier tipo de texto y almacenarlo en archivos de formato CSV.

1.4.2 XAMPP

Es una herramienta de desarrollo que integra tecnologías de software libre, este provee de una configuración funcional, el cual maneja una distribución de Apache, las herramientas que lo componen son:

1.4.3 MySQL

Es un SGBD (Sistema Gestor de Base de Datos) adquirido con mayor frecuencia en el mercado bajo una licencia GPL (Licencia Pública General), o en resumen software libre, pese al pensamiento de muchas personas cuenta con grandes herramientas que llegan a competir con software con licencia. Un rasgo que resalta entre las demás herramientas es su compatibilidad, ya que puede ejecutarse en distintos sistemas operativos (Linux, Windows, entre otros) sin modificar sus características independientemente del sistema que se encuentre. Al definirse como un SGBD cuenta con una curva de aprendizaje buena, ya que su interfaz de fácil entendimiento le otorga la característica de ser un software intuitivo, es decir que puede ser usado sin complicaciones por usuarios no experimentados o expertos.

1.4.4 Pentaho Community Edition

Pentaho es una herramienta de software libre desarrollada por Hitachi, integra diferentes instrumentos para la inteligencia de negocios, en esta se realizan análisis de datos mediante conexiones con bases de datos no relacionales y relaciones, a su vez permite la generación de reportes, esta versión cuenta con diferentes herramientas las cuales son:

1.4.5 Pentaho Data Integration (PDI)

Herramienta basada en tecnología Rapid Application Development (RAD), ésta mediante su interfaz gráfica ofrece la oportunidad de diseñar flujos de trabajo para coordinar recursos y ejecución de actividades como extracción, transformación y carga (ETL).

1.4.6 Schema Workbench

Herramienta para el diseño visual, desarrollada en JAVA, ésta mediante su interfaz permite crear cubos OLAP, integra distintos gestores de base de datos, para implementar un Data Warehouse de la manera óptima y precisa.

1.4.7 Saiku Analytics

Herramienta para el análisis mediante la web, fácil de usar, crea y visualiza la información que contiene un Data Warehouse de manera estructural o gráfica.

CAPÍTULO 2

2. DISEÑO Y CONSTRUCCIÓN

2.1 ESPECIFICACIONES DE REQUERIMIENTOS

En la construcción de la solución de inteligencia de negocios se toman como pilares sustanciales los siguientes requerimientos, en donde se lleva un control de actividades o metas a cumplir por parte de la solución:

Para ello se estableció una lista de requerimientos que aporta una guía para el cumplimiento correcto de la solución:

- Permitir acceder a los datos de entidades públicas en cuanto a los tipos de proyectos de TI mediante la herramienta Saiku Analytics en los cuales se tendrán acceso tanto a sus dimensiones (Entidad, Fecha, Estado, etc.) como a sus medidas (Presupuesto y Plazo Entrega).
- Obtener una respuesta rápida de las consultas al Data Warehouse mediante la herramienta Saiku Analytics.
- Otorgar acceso a un respaldo de los datos obtenidos con los cuales está conformada la inteligencia de negocios.
- Permitir el acceso de usuarios a la información por medio de la herramienta Saiku Analytics.
- Incorporar un control sobre la manipulación de datos de la herramienta de BI, establecerá limitantes para salvaguardar la veracidad de estos.
- Generar reportes del análisis realizado por los usuarios mediante el uso de la herramienta Saiku Analytics.

Según Ramos (2017) la construcción de estos requerimientos es de suma importancia no solo en etapas tempranas del desarrollo de la solución, ya que pueden ser incorporados

otros conceptos o presentarse cambios. En la misma publicación el autor establece las categorías más relevantes para una solución de inteligencia de negocios tales como:

- Calidad de datos
- Seguridad
- Interfaz de usuario
- Conformidad legal
- Integración de Datos y Archivos
- Linaje de datos

2.2 DESCRIPCIÓN DE LA SOLUCIÓN

Actualmente los procesos de adquisición de bienes y servicios para entidades públicas del Ecuador respaldan su información a través de un repositorio digital, es el encargado de recopilar los datos que se abarcan durante el ciclo de vida del proyecto, tales como entidad a la que se encargó el proyecto, presupuesto, fecha, duración, entre otros. Estos datos se convergen en un formato único de PDF individual, por cada proyecto vigente o pasado, esto conlleva a una cierta dificultad para realizar un análisis de factores de interés para los usuarios. Al no contar con una herramienta que facilite dicho análisis y sumado al considerable volumen de proyectos que se han realizado en distintos periodos de tiempo es pertinente la implementación de una solución de inteligencia de negocios que permita realizar un análisis de los datos de una forma ágil y sencilla para el usuario.

El proceso que se lleva a cabo en el Sistema de Contratación Pública del Ecuador tiene como finalidad establecer políticas y condiciones en el transcurso de una contratación pública, esto se realiza con la finalidad de prevenir, detectar y detener intentos de corrupción en donde, de ser el caso, deberá afrontar todo el peso de la ley hasta las últimas consecuencias.

La información que se encuentra en el Sistema de Contratación Pública del Ecuador posee una opción para obtenerla, a través de un mecanismo manual y en periodos cortos de tiempo, seis meses como máximo, por este motivo se dificulta realizar un análisis de datos, sin mencionar que la interpretación de la información obtenida no del todo comprensible para todos los usuarios.

La inteligencia de negocios busca satisfacer estas necesidades y busca una mejor experiencia a los usuarios, desde un manejo más intuitivo para la información como son interfaces de usuarios hasta interpretación de datos con más detalle y de fácil comprensión por medio de gráficos estadísticos o tablas de datos expuestas.

La solución que se presenta posee componentes comerciales con licencia GNU (General Public License), tales como MySQL y el paquete de software que provee Pentaho Community el cual se hace uso de Pentaho Server, Pentaho Data Integration (PDI), Schema Workbench y Saiku Analytics las cuales trabajan en sincronía entre las herramientas y el modelo multidimensional (estrella).

2.3 IMPLEMENTACIÓN DEL PROCESO ETL

2.3.1 *Extracción.*

La etapa de extracción de datos del repositorio del Sistema de Contratación Pública del Ecuador se la ejecutó de forma manual, en primer lugar se realizó la búsqueda mediante el explorador de documentos las siguientes palabras clave: Software, Redes, Desarrollo, Telecomunicaciones, Web visualizando así la información para su descarga en formato PDF, se encontró alrededor de 11.318 archivos que se extrajeron de esta manera, posteriormente se trató la información que estos contenían a través de herramientas para realizar los procesos de transformación y almacenamiento de los datos en la base de datos multidimensional.

PDFElement. La herramienta PDFElement mediante su librería OCR extrae la información seleccionada de cada uno de los archivos en formato PDF, siendo así su punto de partida: Entidad, Objeto de Proceso, Código, Tipo Compra, Presupuesto Referencial Total (sin IVA), Tipo de Contratación, Forma de Pago, Total de Adjudicación, Plazo de Entrega, Vigencia de Oferta, Funcionario encargado del proceso, Fecha de publicación, esta se almacena en diferentes archivos en formato CSV, a estos datos se realiza una segregación mediante el tutor del proyecto de titulación evaluó cada uno de estos campos y brindó su apoyo en esta etapa, se obtuvo información relevante para su posterior tratamiento.

2.3.2 Transformación.

La etapa de transformación de los datos previamente obtenidos y almacenados en formatos CSV, se trató mediante macros en Excel que permite la estandarización, se transformó el formato original de letra de minúscula a mayúscula, se realiza la estandarización de los nombres a los campos de la hoja de Excel tomando como ejemplo el campo “año” reemplazándolo por “anio”, a su vez se depuran los signos de exclamación y de interrogación que generarían problemas al momento de cargar los datos, se realiza la separación de la información del CSV de fecha en día, mes y año (anio) para no generar conflictos al momento de realizar la carga y análisis de datos, se establece un límite de datos para el archivo CSV de “objeto” que contiene información acerca del objetivo de cada proyecto por ejemplo “DESARROLLO E IMPLEMENTACIÓN DE MÓDULOS DE REPORTES Y DE MANTENIMIENTO DE INFORMACIÓN DE LOS BIENES DE REGISTRO E INVENTARIO PERTENECIENTES A LOS FONDOS YACIMIENTOS ARQUEOLÓGICOS, COLECCIONES ARQUEOLÓGICAS, BIENES INMUEBLES, BIENES INMATERIALES, BIENES DOCUMENTALES DE ARCHIVO E HISTÓRICO”,

esto con la finalidad de obtener información depurada e íntegra para realizar el proceso de carga a la base de datos que alimentará el Data Warehouse.

Tabla 1

Descripción de estándar de datos

Archivo CSV	Datos del archivo sin aplicar el estándar	Datos del archivo aplicado el estándar	Estándar
Objeto	“Adquisición de materiales informáticos”	“ADQUISICIÓN DE MATERIALES INFORMÁTICOS”	Se realizó el cambio de minúsculas a mayúsculas
Objeto	“ADQUISICIÓN DE MATERIALES INFORMÁTICOS”	ADQUISICION DE MATERIALES INFORMATICOS	Se realizó la depuración de caracteres especiales como: tildes, punto y coma, signos de exclamación e interrogación y comillas.
Objeto	30 TUBOS PVC 4" SANITARIO DESAGÜE PLASTIGAMA PARA PASES DE AGUA EN EL SECTOR DE EL ROSAL.	Eliminado	Se elimina la información no relevante de acuerdo a los proyectos de software
Objeto	DESARROLLO E IMPLEMENTACION DE MODULOS DE REPORTES Y DE MANTENIMIENTO DE INFORMACION DE LOS BIENES DE REGISTRO E INVENTARIO PERTENECIENTES A LOS FONDOS YACIMIENTOS ARQUEOLOGICOS ...	DESARROLLO E IMPLEMENTACION DE MODULOS DE REPORTES Y DE MANTENIMIENTO DE INFORMACION DE LOS BIENES DE REGISTRO E INVENTARIO PERTENECIENTES A LOS FONDOS YACIMIENTOS ARQUEOLOGICOS	De acuerdo a los campos de las tablas dimensionales el máximo valor almacenable es 400 caracteres.

Fecha	20/05/2010 0:00:00	DIA:20 MES:5 ANIO:2010	Se extrae los campos: día, mes y año del formato fecha.
Fecha	AÑO	ANIO	Se realizó el cambio de AÑO por ANIO tanto en la base de datos como en el archivo CSV
Presupuesto	3.730,42	3730,42	Se retiró el carácter separador de unidades de miles (punto)
Todos los archivos CSV	No Disponible	Eliminado	Se eliminó los datos con información no disponible

Nota: Tabla que describe la etapa de Transformación de los datos. Elaborado por: Josue

Coba y Estalin Revelo.

En la Tabla 1 se puede evidenciar los cambios que se han realizado en los datos como son: mayúsculas, eliminación de tildes y sustitución de palabras con caracteres especiales, estos cambios se ejecutaron con el fin de establecer un estándar para los datos para su interpretación sin inconvenientes.

2.3.2.1 Depuración. La etapa de depuración contempla varios aspectos que al llegar al término de la misma brindan información con integridad y calidad, estos procesos se detallan a continuación:

- Se eliminó los caracteres especiales como caracteres como: comillas dobles (“”), comillas simples (‘’), puntos y coma (;), dos puntos (:), punto (.), coma (,), tildes (ó), barra inversa invertida (\), signos de interrogación, signos de exclamación,

espacios en blanco, para ello se crearon macros las cuales se encuentran detalladas en el anexo 5 en la versión 2 de depuración.

- Se eliminan los espacios en blanco tanto al inicio, en el intermedio y el final de cada uno de los campos obtenidos.
- Se realizó la revisión de duplicidad de información esto fue mediante la herramienta de Excel, por medio del cual se filtró columnas y filas que contengan información repetida en esta y se procedió a eliminarlas.
- Se realizó la revisión de campos nulos e información sin relación al desarrollo de software, en este proceso se obtiene como resultado 3500 datos los cuales serán posteriormente cargados a la base de datos.

2.3.3 Carga

La etapa del proceso de carga a la base de datos se realiza a través de la herramienta de Pentaho Data Integration (PDI) en conjunto con la herramienta MySQL.

MySQL. La herramienta MySQL permitió implementar el diagrama de la base de datos de las 9 tablas y se estableció los tipos de dato de cada una de ellas siendo así:

Tabla 2

Descripción base de datos bi_coba_revelo.

NOMBRE DE LA TABLA	TIPO DE DATO	DESCRIPCIÓN
Contratación	DIM_CONTRATACION: varchar (100) CONTRATACION: varchar (400)	Tabla dimensional que hace alusión al tipo de contratación que se llevó a cabo: - Concurso Público. - Subasta inversa electrónica. - Contratación directa. - Menor cuantía. - Cotización. - Licitación.

Plazo_entrega	DIM_PLAZO_ENTREGA: varchar (100) PLAZO_ENTREGA: int (10)	Tabla dimensional que hace alusión al plazo de entrega para el proyecto que se llevó a cabo en días.
Estado	DIM_ESTADO: varchar (100) ESTADO: varchar (400)	Tabla dimensional que hace alusión al estado que se encuentra el proyecto: - Finalizada. - Parcial. - Por adjudicar. - Desierta. - Adjudicada. - Cancelado. - En recepción. - Finalizado por mutuo acuerdo. - Ejecución de contrato.
Entidad	DIM_ENTIDAD: varchar (100) ENTIDAD: varchar (400)	Tabla dimensional que hace alusión a la entidad pública encargada el proyecto.
Presupuesto	DIM_PRESUPUESTO: varchar (100) PRESUPUESTO: decimal (40,2)	Tabla dimensional que hace alusión al presupuesto acordado para el desarrollo del proyecto en dólares.
Vigencia_oferta	DIM_VIGENCIA_OFERTA: varchar (100) VIGENCIA_OFERTA: int (10)	Tabla dimensional que hace alusión al tiempo que estuvo en vigencia el proyecto en días.
Objeto	DIM_OBJETO: varchar (100) OBJETO: varchar (400)	Tabla dimensional que hace alusión al objetivo del proyecto.
Tipo_compra	DIM_TIPO_COMPRA: varchar (100) TIPO_COMPRA: varchar (400)	Tabla dimensional que hace clasifica el proyecto en tres categorías: - Bien. - Servicio. - Consultoría.
Fecha	DIM_FECHA_PUBLICACION: FECHA_PUBLICACION: date DIA: int (2) MES: int (2) ANIO int (4)	Tabla dimensional que refleja la fecha en el que se ha publicado el proyecto, tomando en cuenta el año, mes y día.

Nota: Describe el tipo de dato de la base de datos. Elaborado por: Josue Coba y Estalin

Revelo.

En la Tabla 2 se establecen los tipos de valores con los cuales están construidas las tablas dimensionales en la base de datos, así como su nombre y la especificación del tamaño de datos.

A continuación, se presenta el esquema de la base de datos.

Figura 3

Descripción estructura base de datos.



Nota: Estructura de la base de datos de datos conforme a la tabla 2. Elaborado por: Josue Coba y Estalin Revelo mediante phpMyAdmin.

En la Figura 3 se puede mostrar cómo están construidas las tablas dimensionales en phpMyAdmin y sus tipos de datos, así como sus respectivos tamaños, cada uno de tablas permitió almacenar la información que se obtuvo previamente y la cual conforma el Data Warehouse.

Cada una de estas dimensiones representa un factor crucial para realizar el análisis de los datos. Los datos almacenados en dichas dimensiones se pueden evidenciar de mejor manera en la Tabla 2.

2.4 ANÁLISIS Y DISEÑO DEL DATA WAREHOUSE

Tomando en cuenta los requerimientos propuestos en el inicio del proyecto se diseñó la solución de inteligencia de negocios mediante una metodología Kimball, tiene como propósito el desarrollo de Data Mart individual para posteriormente unificarlos, de esta manera se busca la creación de un Data Warehouse fácil, en donde hace énfasis en la información obtenida del Sistema de Contratación Pública del Ecuador en la última década.

Previo al modelamiento de la base de datos, el proceso que se llevó a cabo fue el de evidenciar toda aquella información que se puede obtener de cada proyecto. Mediante este estudio de la información se afrontó un gran obstáculo, algunos datos que son prioritarios en muchos de los casos no se encontraban disponibles o carecían de información, tras una primera depuración se constató una cantidad importante de datos incompletos, lo que llevó a su eliminación para así no disminuir la calidad de la solución de inteligencia de negocios. Tras realizar este análisis y una primera depuración de los datos se establece la granularidad y la profundidad del análisis a llegar con la solución de inteligencia de negocios, ya que al ser información de varios periodos de tiempo es importante realizar un análisis medianamente profundo, con la finalidad de satisfacer las necesidades del usuario al momento que sea requerida la búsqueda de información.

El nivel de granularidad para el Data Mart es: el presupuesto que se invierte en proyectos de software en un periodo de tiempo determinado y el estado en que se terminó dicho proyecto.

Considerando lo anteriormente mencionado y teniendo en claro a dónde se desea llegar es pertinente establecer los siguientes parámetros a lo que se refiere a modelado y posterior construcción del Data Mart:

Los datos de fecha se segmentan en los componentes más relevantes: día, mes y año a su vez mantiene el formato conjunto en donde se combinan estos elementos en uno solo. Tras identificar la cantidad de caracteres que posee el objeto del proyecto de software se limitan sus caracteres para reducir tiempos de ejecución. Contemplar el tipo de dato que representa presupuesto al contar con una cantidad decimal, se debe tener un estándar para evitar cambios de valores en la implementación. La información que representa tiempo referente a proyecto se maneja en dos tablas para mejor segregación de información como: tiempo de oferta y plazo de entrega. Establecer una considerable cantidad de caracteres para las tablas de entidad y objeto, ya que en ciertos casos poseen cadenas de caracteres amplias. Al tratarse de cantidades considerables a lo que presupuesto se refiere es prescindible otorgar un tipo de dato que permita abarcar el valor total que en muchos de los casos puede ser por miles. Identificar la información que no sea de relevancia para el usuario como datos incompletos, datos no disponibles o sin relación para el objetivo de la solución de inteligencia de negocios, es decir, los proyectos de software.

2.4.1 Dimensiones

Las tablas de dimensiones o tablas dimensionales que se consideran en la base de datos son todas aquellas que representan o contienen datos descriptivos de los proyectos de software, con los cuales se obtiene un análisis de la información. Cabe aclarar que toda información que se ingresa en las tablas dimensionales es proporcionada por los datos obtenidos a través del Sistema de Contratación Pública del Ecuador, en donde en ciertos

casos los datos ingresados al portal carecen de un estándar o no disponen la información completa con lo cual fue indispensable realizar constantes depuraciones de los datos. Cada una de estas tablas dimensionales describen los parámetros de proyectos de software los cuales al hacer analizados se obtiene información. El modelo de tablas dimensionales se rige a la Figura 3.

2.4.2 Tabla de hechos

Tras dar por terminado el análisis de las dimensiones mencionadas y desarrolladas en puntos anteriores, se presenta una tabla de hechos que tiene como finalidad la recolección de todos los datos con la finalidad de realizar consultas que satisfagan las necesidades de los usuarios. La estructura cómo está construida la tabla de hechos está compuesta por los identificadores de cada una de las tablas dimensionales como se describe en la Figura 4.

Figura 4

Diagrama físico de tabla de hechos.



Column Name	Data Type
DIM_ENTIDAD	varchar(100)
DIM_OBJETO	varchar(100)
DIM_TIPO_COMPRA	varchar(100)
DIM_PRESUPUESTO	varchar(100)
DIM_CONTRATACION	varchar(100)
DIM_PLAZO_ENTREGA	varchar(100)
DIM_VIGENCIA_OFERTA	varchar(100)
DIM_ESTADO	varchar(100)
DIM_FECHA_PUBLICACION	varchar(100)

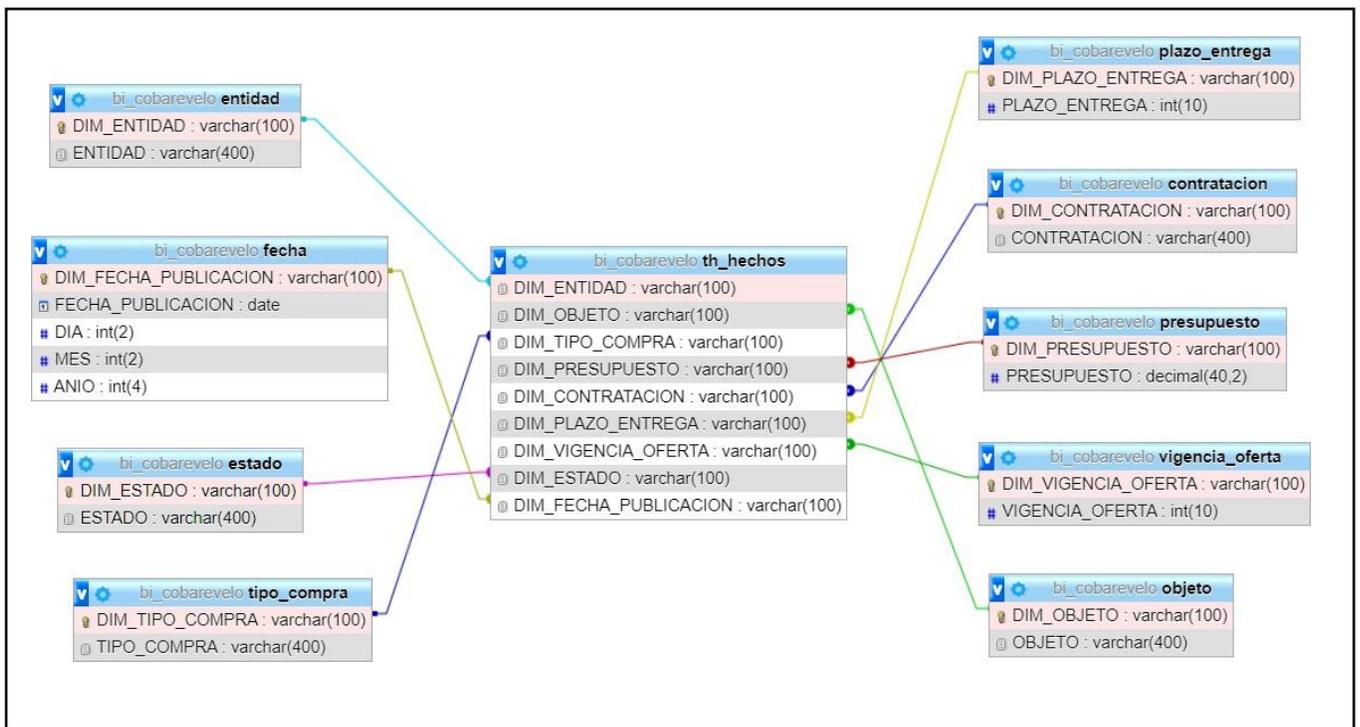
Nota. Diagrama de la tabla de hechos que usa la inteligencia de negocios. Elaborado por: Josue Coba y Estalin Revelo mediante phpMyAdmin.

La Figura 4 muestra la estructura de la tabla de hechos del Data Warehouse, tal como sucede en las dimensiones, se encuentran los tipos de datos con sus respectivos tamaños.

Tras modelar las tablas de hechos y dimensiones con sus respectivos campos se procede en la implementación de la metodología Kimball, es un proceso de construcción físico de la base de datos, para realizar esta tarea se utilizó phpMyAdmin en donde su herramienta de diseñador fue de gran utilidad, ya que al ser intuitiva y de fácil uso en especial al momento de crear relaciones entre la tabla de hechos y las tablas dimensiones se obtiene la estructura de la Figura 5:

Figura 5

Diagrama físico de la estructura estrella.



Nota. Diagrama físico con la estructura estrella que usa la inteligencia de negocios.

Elaborado por: Josue Coba y Estalin Revelo mediante phpMyAdmin.

El diagrama físico está basado en el modelo estrella, se encuentra en el SGBD (Sistema de Gestión de Base de Datos) es totalmente funcional en donde se puede realizar funciones SQL como insertar, actualizar y eliminar datos, no obstante existe una brecha que impide la comunicación con la herramienta de Pentaho Community, para ello debe ser implementado un Data Warehouse que pueda ser publicado en la herramienta de Pentaho Community, para su creación hace uso de la herramienta Schema Workbench de Pentaho Community permite realizar una conexión de la base de datos MySQL, para posterior a la creación del Data Warehouse y su configuración, pueda ser publicado satisfactoriamente en Pentaho Community.

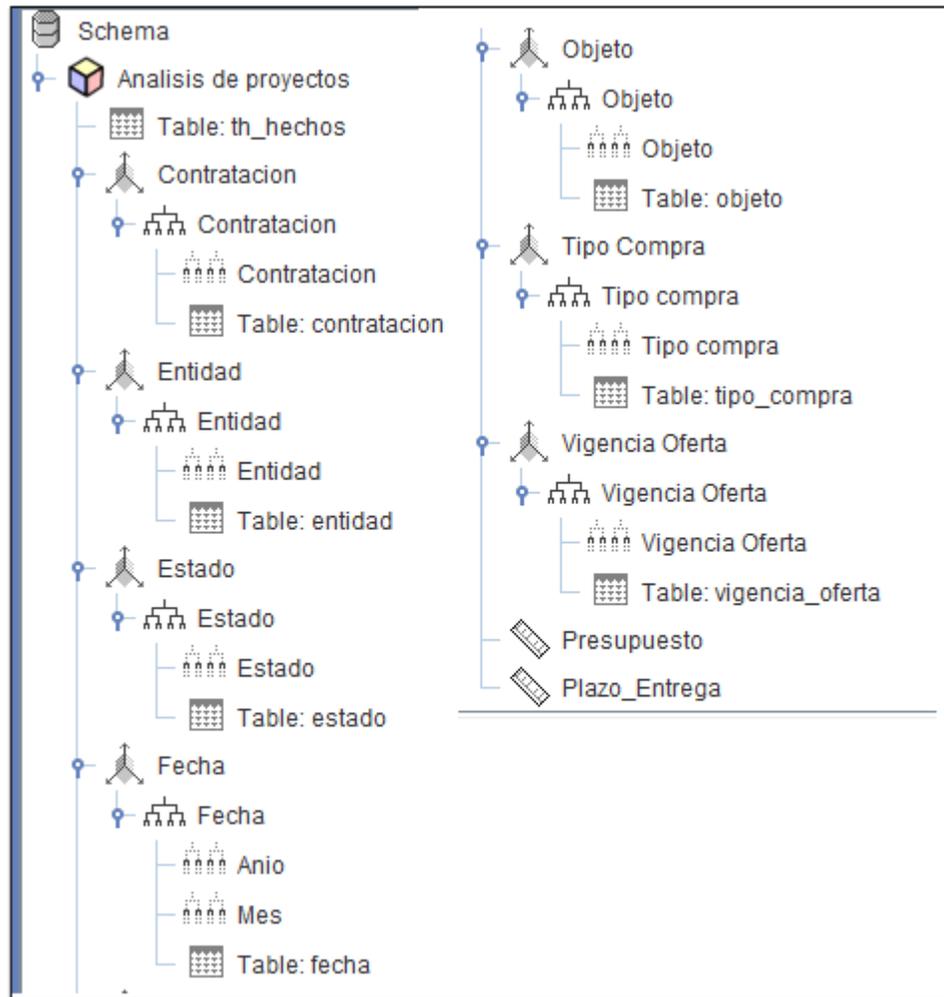
2.4.3 Schema Workbench

Para la creación del Data Warehouse, Pentaho Community ofrece Schema Workbench, cuenta con la posibilidad de realizar conexiones con distintas bases de datos y varios SGBD (Sistema de Gestión de Base de Datos), no obstante, la conectividad con MySQL presenta un inconveniente que se soluciona descargando conector (mysql-connector-java-8.0.23), esto por parte de la conexión con la base de datos.

Un punto a destacar es la declaración de medidas, dimensiones y la tabla de hechos, a su vez la granularidad en el campo de fecha, ya que al existir campos en los cuales se pueden subdividir como son el año y mes deben establecerse sus correspondientes niveles y jerarquías. Tras la creación del Data Warehouse dimensional con sus correspondientes componentes el proceso de publicación del mismo a Pentaho Community se presenta un inconveniente, al ser la primera vez que se publica en esta plataforma es necesario realizar un cambio de contraseña en la plataforma de Pentaho Community, una vez realizado los cambios, el Data Warehouse se encuentra conformado como se muestra en la Figura 6.

Figura 6

Esquema del Data Warehouse.



Nota. Esquema del Data Warehouse con sus dimensiones, niveles, jerarquías y tablas.

Elaborado por: Josue Coba y Estalin Revelo mediante Schema Workbench.

Como se muestra en la Figura 6 cada dimensión del Data Warehouse tiene componentes en común como son la jerarquía y la tabla de donde obtiene los datos, salvo en la dimensión fecha que al tener varios campos se debe establecer niveles, también se evidencia la creación de medidas como es presupuesto y plazo de entrega las cuales como las dimensiones obtienen la información de la base de datos creada.

Al realizar este proceso se debe configurar un último detalle, esta vez en Pentaho Community, en la sección de Manage Data Source al igual que Schema Workbench se establece recursos JDBC mediante el mismo controlador que se utilizó en el desarrollo del Data Warehouse.

2.4.4 Versiones de depuración

Alrededor del presente proyecto se realizaron diferentes versiones, las cuales fueron:

Versión 1: Descarga de datos del Sistemas de Contratación Pública del Ecuador. La presente versión tiene como referencia la extracción de los archivos en formato PDF, el almacenamiento de cada uno de estos y la resolución de los inconvenientes que presentaron al momento de descargar cada uno de ellos, en caso de necesitar más información acerca de este proceso consulte el Anexo 5.

Versión 2: Extracción y depuración de los datos mediante la herramienta PDFElement. Versión que tiene como objetivo la extracción de la información contenida en los archivos PDF y la depuración de los caracteres especiales mediante macros en la herramienta de Excel, en caso de necesitar más información acerca de este proceso consulte el Anexo 5.

Versión 3: Segunda depuración de los datos obtenidos mediante la herramienta PDFElement. Versión que tiene como objetivo la corrección del campo de Presupuesto en el cual el sistema de contratación pública maneja las unidades, decenas y centenas de manera distinta, este cambio se lo realizó en el estado del presupuesto con esta construye la medida de Presupuesto que conforma el Data Warehouse, en caso de necesitar más información acerca de este proceso consulte el Anexo 5.

Versión 4: Tercera depuración de los datos obtenidos mediante la herramienta PDfElement. Versión que tiene como objetivo la corrección del campo objeto en este se corrige el inconveniente de acuerdo a la información sobrecargada en este, para el despliegue de los reportes de la herramienta Saiku Analytics, en caso de necesitar más información acerca de este proceso consulte el Anexo 5.

Cada una de estas etapas conforman el proceso de ETL, éste brinda una retroalimentación en cada una de estas, llegando así a obtener un producto con calidad e integridad de los datos.

2.4.5 Medidas

En el análisis y la construcción del Data Warehouse se deben establecer medidas que permitan realizar consultas con la información obtenida de la base de datos como son: el presupuesto, tiempo, entidades y objeto de estudio. Para esta información se establece como medidas el presupuesto y plazo de entrega que representan puntos de interés para los usuarios como se evidencia en la Figura 6. Es importante aclarar que las medidas tienen una funcionalidad importante dentro del análisis, ya que serán ejes fundamentales para la obtención de información relevante, para ello Schema Workbench aporta con la función de agregación a las medidas las cuales fueron una suma de valores de sus respectivas tablas dimensionales.

2.4.6 Análisis de los datos

Una vez creado correctamente el Data Warehouse y publicado satisfactoriamente en la plataforma de Pentaho Community se hace uso de la herramienta Saiku Analytics, en donde es posible visualizar de mejor manera los datos obtenidos por consultas mediante

tablas o gráficos estadísticos, el análisis realizado da paso a la creación de reportes: Ad-hoc y estáticos considerado como punto clave de una inteligencia de negocios.

2.5 CREACIÓN DE REPORTE

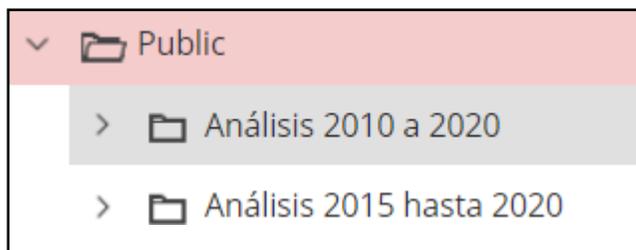
La creación de reportes de una solución de inteligencia de negocios es de suma importancia, ya que esta es la encargada de transmitir los hallazgos encontrados mediante los datos que se encuentran en el Data Warehouse, es necesaria la herramienta de Saiku Analytics, permite la manipulación de datos y generar reportes del análisis realizado en distintos formatos que sea de preferencia para el usuario como: PDF, Excel, CSV o PNG. La creación de reportes de una solución de inteligencia de negocios es de suma importancia, ya que esta es la encargada de transmitir los hallazgos encontrados mediante los datos que se encuentran en el Data Warehouse, es necesaria la herramienta de Saiku Analytics, permite la manipulación de datos y generar reportes del análisis realizado en distintos formatos que sea de preferencia para el usuario como: PDF, Excel, CSV o PNG.

2.5.1 *Reporte estático*

Las generaciones de reportes estáticos de la solución de inteligencia de negocios se encuentran ubicados en una sección de libre acceso en donde cualquier tipo de usuario (analista o público) puede acceder a ellos, como se muestra en el Anexo 2, para la realización de este tipo de reporte se realizó una encuesta a los diferentes tipos de usuarios como se evidencia en el Anexo 6. Teniendo los resultados de la encuesta, los criterios deseados por los usuarios se generaron reportes anuales que contengan cierta información específica cómo es el caso del estado, la entidad y el objeto de estudio de cada proyecto realizado por las entidades públicas, además de contar también con una segmentación de los últimos cinco años. Como se muestra en la Figura 7.

Figura 7

Reportes Estáticos.



Nota. Reportes anuales y de los últimos cinco años. Elaborado por: Josue Coba y Estalin Revelo mediante Pentaho Community.

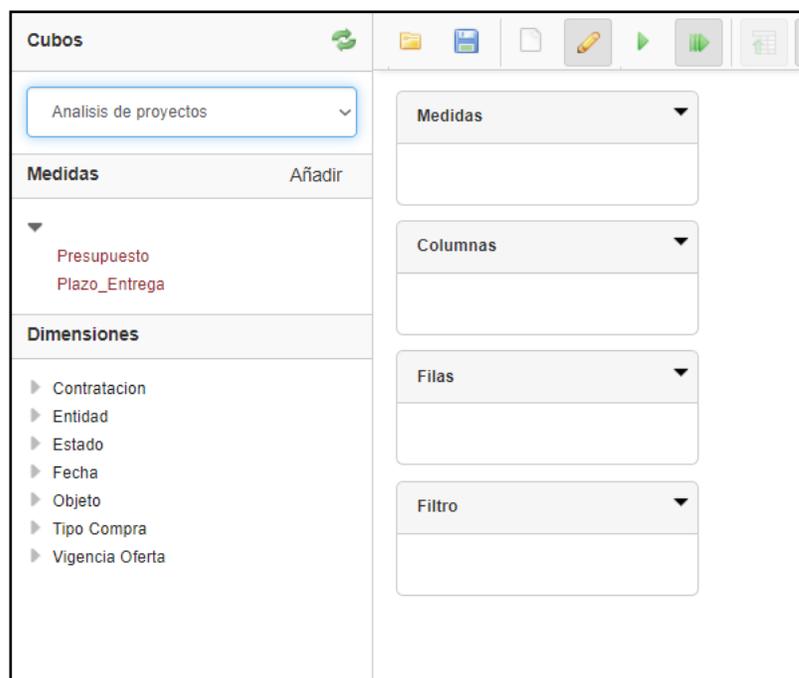
El reporte estático tiene como cualidad mostrar al usuario información previamente analizada y obtenida, como se puede evidenciar en el Anexo 6.

2.5.2 *Ad-hoc*

Con respecto a los reportes Ad-hoc también se hace uso de la herramienta Saiku Analytics, lo que diferencia este tipo de reportes es en la manipulación de datos, ya que se puede establecer una granularidad más específica, además de analizar a profundidad mediante el uso de filtros que Saiku Analytics otorga.

Figura 8

Reportes Ad-hoc.



Nota. Panel para manipular las dimensiones y medidas creadas en el Data Warehouse.

Elaborado por: Josue Coba y Estalin Revelo mediante Saiku Analytics.

El reporte Ad-hoc tiene como cualidad manipular información, como se puede evidenciar en el Anexo 1.

2.6 DESCRIPCIÓN DEL PROCESO DE PRUEBAS AL DATA WAREHOUSE

Una vez el Data Warehouse se implementa, se procede con la fase de pruebas la cual contempla varias categorías, las cuales se presentan en la siguiente Tabla 4:

La Tabla 4 representa una matriz de pruebas con tres categorías principales: calidad de datos, seguridad e interfaz de usuario, para mayor detalle diríjase al Anexo 3 y Anexo 1.

Tabla 3*Matriz de pruebas.*

Documento de Requerimientos				
Categoría	Tipo de requerimiento	N°	Requerimiento	Tipo de Prueba
Seguridad	Seguridad	1	Incorporar un control sobre la manipulación de datos de la herramienta Pentaho Server el cual establecerá limitantes para salvaguardar la veracidad de estos.	Pruebas de validación de datos
Calidad de datos	Compleitud	2	Permitir acceder a los datos de entidades públicas en cuanto a los tipos de proyectos de TI mediante la herramienta Saiku Analytics en los cuales se tendrán acceso tanto a sus dimensiones (Entidad, Fecha, Estado, etc.) como a sus medidas (Presupuesto, Plazo Entrega).	Prueba de integración.
	Eficiencia	3	Obtener una respuesta rápida de las consultas al Data Warehouse mediante la herramienta Saiku Analytics.	Prueba de rendimiento
	Disponibilidad	4	Permitir el acceso de usuarios a la información por medio de la herramienta Saiku Analytics	Pruebas de disponibilidad
Interfaz de Usuario	Reportería	5	Generar reportes de la interacción que los usuarios realicen mediante el uso de la herramienta Saiku Analytics	Pruebas de validación de datos

Nota: Descripción y resultados de la matriz de pruebas (Anexo 3). Elaborado por: Josue

Coba y Estalin Revelo.

2.6.1 Seguridad

Definir las funciones y responsabilidades de los usuarios y administrador con acceso a la información del Data warehouse.

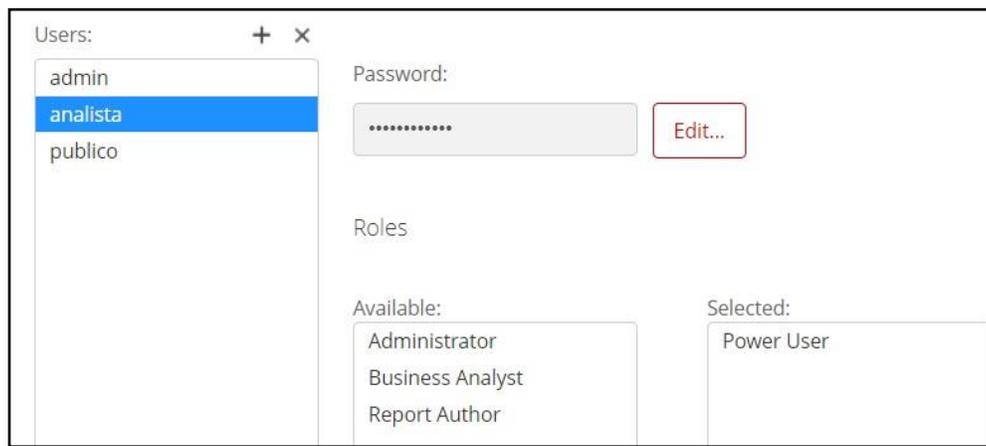
Descripción de la prueba. Establecer perfiles de acceso a los datos en cuanto a la visualización de los datos (administrador y usuario), establecer roles de acuerdo al tipo de usuarios que accedan a la herramienta (comunes y avanzados).

Resultados Esperados. Establecer control de acceso a la información mediante mecanismos reconocidos, con el fin de mantener la información con un grado de veracidad y evitar alteraciones innecesarias.

Resultados Obtenidos. Se implementa roles y perfiles, estos brindan la posibilidad de diferenciar a los usuarios comunes de los analistas de la información, también mantendrán acceso limitado a la información en cuanto a generación de reportes.

Figura 9

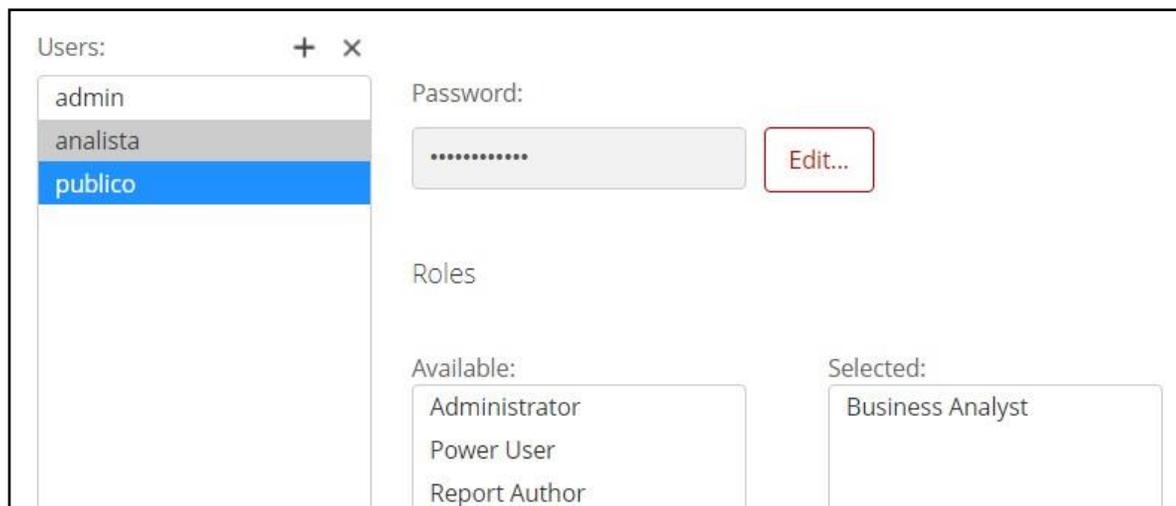
Creación de perfiles, roles y permisos de analista.



Nota: Creación de perfiles para analista con la finalidad de evitar un mal uso de la información. Elaborado por: Josue Coba y Estalin Revelo mediante la herramienta Pentaho Server.

Figura 10

Creación de perfiles, roles y permisos de público.



Nota: Creación de perfiles para público general con la finalidad de evitar un mal uso de la información. Elaborado por: Josue Coba y Estalin Revelo mediante la herramienta Pentaho Server.

Conclusiones. La creación de perfiles con sus respectivos roles y funcionales, así como sus limitantes para la manipulación y acceso a la información permite obtener seguridad en los datos, logrando un trato amigable con el usuario como salvaguardar la información.

2.6.2 Prueba de integración

Asegurar que los datos cumplan con los procesos de Extraer, Transformar y Cargar, por medio del proceso ETL donde se alimenta de información el Data Warehouse, interactuando con sus respectivas dimensiones y medidas.

Descripción de la prueba. Extraer la información del repositorio digital del Sistema de Contratación Pública del Ecuador, provee de archivos en formato PDF la extracción se realizará de forma manual mediante el uso de la herramienta PDFElement se obtiene la

información de estos documentos para posteriormente utilizar la herramienta Pentaho Data Integration (PDI) para transformarlos y cargarlos respectivamente al repositorio digital de datos que alimenta el Data Warehouse.

Resultados Obtenidos. Se Transforma los archivos por lotes de PDF obtenidos del Sistema de Contratación Pública del Ecuador por medio de la herramienta PDFElement estos datos son almacenados en archivo CSV. La herramienta de Pentaho Data Integration (PDI) carga los datos obtenidos de los archivos CSV, los cuales ingresaron de manera exitosa a la base de datos. La herramienta Schema Workbench ayudó en la creación del Data Warehouse, adicionalmente la integración de medidas y niveles para un próximo análisis de datos en la herramienta de Pentaho Server.

2.6.3 Prueba de rendimiento

Asegurar que las consultas al Data Warehouse se realicen dentro de los plazos previstos.

Descripción de la prueba. Realizar consultas de los datos previamente cargados en el Data Warehouse, mediante la interacción de sus dimensiones y medidas establecidas.

Resultados Obtenidos. Se realiza la consulta de cuantos proyectos de software se han finalizado en diciembre del año 2016 y su presupuesto y esto arroja un resultado exitoso en cuanto a la información solicitada en esta.

2.6.4 Prueba de disponibilidad

Asegurar de que la solución de almacenamiento de datos permite la interacción con los datos (dimensiones y medidas).

Descripción de la prueba. Solicitar información acerca del servicio de hospedaje del servidor en cuanto a rendimiento y posible mantenimiento del mismo, a su vez desarrollar un

diccionario de datos para que los usuarios interactúen con las dimensiones y medidas del Data Warehouse

Resultados Obtenidos. Se obtuvo la información solicitada al Data Center del Campus Sur de la Universidad politécnica Salesiana, en donde se obtuvo respuesta de una administradora Marcela Gallegos donde informó acerca del rendimiento y calendarización del mantenimiento del mismo, a su vez se creó un diccionario de datos para los usuarios que deseen información acerca de los datos de esta.

2.6.5 Interfaz de usuario

Prueba de validación. Los informes del usuario final son un componente importante de la herramienta, esté posibilita la interacción de datos.

Descripción de la prueba. Generar reportes en los cuales se visualice la información que el usuario requiera conforme a ello permitir la descarga en formato PDF

Resultados Obtenidos. Se presentan los reportes estáticos y Ad-hoc establecido en la herramienta Saiku Analytics mediante el cual interactúa el usuario con estos, minimizando así el tiempo de consulta al Data Warehouse, a su vez se encuentra disponible la descarga de los archivos en formato PDF.

CAPÍTULO 3

3. EXPLOTACIÓN

3.1 DESCRIPCIÓN DEL PROCESO DE DATOS AL DATA WAREHOUSE

3.1.1 Datos almacenados en el Sistema de Contratación Pública del Ecuador

El Sistema de Contratación Pública del Ecuador es la fuente de datos en donde se obtuvo la información acerca de los proyectos de software de las entidades públicas desde el 2010 hasta el 2020. Cabe aclarar que el sistema carece de la posibilidad de la descarga de archivos por bloques, lo que conlleva a la descarga manual de cada uno de los archivos durante los periodos de tiempo que el sistema soportaba, los cuales eran de 6 meses como máximo.

Figura 11

Sistema Oficial de Contratación Pública.



The screenshot shows the search interface of the Sistema Oficial de Contratación Pública. At the top, there is a logo for 'COMPRAS PÚBLICAS' and the title 'Sistema Oficial de Contratación Pública'. Below the title, the date and time are displayed as 'Jueves 20 de Mayo de 2021 | 23:42'. The main section is titled '»Búsqueda de Procesos de Contratación' and includes a note: 'Para Buscar Procesos: Debe Presionar el botón 'Buscar Entidad', e ingresar el nombre de la Entidad que desea consultar.' The search form contains several fields: 'Palabras claves' (text input), 'Entidad Contratante' (text input with a search icon), 'Tipo de Contratación' (dropdown menu set to 'TODOS'), 'Tipo de compra' (text input), 'Estado del Proceso' (text input), 'Código del Proceso' (text input), and 'Por Fechas de Publicación (*)' with 'Desde:' (2020-11-20) and 'Hasta:' (2021-05-20) date pickers. At the bottom, there is a link for 'Para una búsqueda más avanzada' pointing to 'Buscar en google'.

Nota. Sistema donde se almacenan los documentos requeridos para la extracción de información de proyectos de software de entidades públicas. Elaborado por: Josue Coba y Estalin Revelo.

La Figura 9 presenta un repositorio gracias al cual se ejecutó la extracción de datos, que se trataron posteriormente en las etapas de ETL.

La información plasmada en los archivos PDF se presenta en la Figura 10, en donde se identifican los datos de: entidad, objeto de proceso, código, tipo compra, presupuesto referencial total sin IVA, tipo de contratación, forma de pago, tipos de adjudicación, plazo de entrega, vigencia de oferta, funcionamiento encargado del proceso, estado del proceso, estado que finalizó el proceso, descripción y fechas, estos campos sirvieron para efectuar la extracción de la información de cada uno de los archivos PDF descargados.

Figura 12

PDF descargado del Sistema Oficial de Contratación Pública.

Sistema Oficial de Contratación Pública

Sábado 3 de Octubre de 2020 | 20:58 [Ingresar al Sistema]

» Información Proceso Contratación

Proceso de Contratación

Entidad:	Empresa Pública de hidrocarburos del Ecuador EP PETROECUADOR
Objeto de Proceso :	ACTUALIZACIÓN DEL SISTEMA DE GESTION DE LA RED WAN CISCO, INSTALACIÓN DE EQUIPOS MULTIMARCA DEL SCADA Y REPOTENCIACIÓN DEL FLUJO DE DATOS HACIA EL CMCH DE ARCH
Código:	CS-06-OSC-ZO-O-2014
Tipo Compra:	Servicio
Presupuesto Referencial Total (Sin Iva):	USD 180,000.00
Tipo de Contratación:	Cotización
Forma de Pago:	Anticipo 0% Saldo: Pago Otra - Revisar términos de referencia 100.00%
Tipo de Adjudicación:	Total
Plazo de Entrega:	90 días
Vigencia de Oferta:	90 días
Funcionario encargado del proceso:	rgalarza@ind.eppetroecuador.ec
Estado del Proceso:	Finalizada
Estado en el cual finalizó el Proceso	En Recepción
Descripción:	ACTUALIZACIÓN DEL SISTEMA DE GESTION DE LA RED WAN CISCO, INSTALACIÓN DE EQUIPOS MULTIMARCA DEL SCADA Y REPOTENCIACIÓN DEL FLUJO DE DATOS HACIA EL CMCH DE ARCH

Fechas de Control del Proceso

Fecha de Publicación	2014-09-16 09:00:00	Indicar la fecha real en la cual desea publicar el Proceso.
Fecha Limite de Preguntas	2014-09-18 09:00:00	Fecha máxima para solicitar aclaraciones respecto al Proceso de Contratación.
Fecha Limite de Respuestas	2014-09-23 22:00:00	Fecha máxima para solventar cualquier inquietud relacionada al Proceso de Contratación.
Fecha Limite de Propuestas	2014-09-25 10:00:00	Fecha máxima para la entrega de propuestas.
Fecha Apertura de Ofertas	2014-09-25 10:30:00	Fecha para la apertura de los sobres de las ofertas.
Fecha Estimada de Adjudicación	2014-10-10 11:00:00	Fecha estimada para la Adjudicación de la compra.

Nota. Información de un archivo descargado y almacenado en el sistema oficial de contratación pública. Fuente: Josue Coba y Estalin Revelo.

3.1.2 Depuración

La información obtenida del Sistema Oficial de Contratación Pública presentó ciertas características que no aportaron valor para un análisis, estos fueron: datos mal redactados, campos faltantes, texto mal escrito, entre otros. En la primera depuración de datos tras utilizar la herramienta PDFElement mediante su función de lectura OCR se convirtió los datos de

los archivos en formato PDF a texto plano almacenados en un documento CSV, en este proceso la herramienta confundió caracteres similares tales como el número 1 por la letra “l”, entre otras similitudes, por medio de la herramienta Microsoft Excel y las funciones de filtrado y macros se depuró y pulió la información de acuerdo a los estándares establecidos en la tabla 1, es por ello que la importancia de esta primera depuración fue de gran utilidad para la próxima iteración en el proceso ETL.

La Figura 11 especifica el proceso de extracción de información de archivos PDF a un archivo CSV mediante la herramienta de PDFElement se realizó por medio de campos hacia el documento.

Figura 13

Formulario de extracción mediante PDFElement.

The screenshot shows a web form titled "Sistema Oficial de Contratación Pública". At the top, it displays the date and time "Miércoles 26 de Agosto de 2020 | 13:42" and a link "[Ingresar al Sistema]". The main section is titled "»Información Proceso Contratación" and contains a "Proceso de Contratación" form. The form fields are as follows:

Entidad:	SUPERINTENDENCIA DE ECONOMIA POPULAR Y SOCIAL	Entidad
Objeto de Proceso :	CONTRATACIÓN DEL SERVICIO DE MANTENIMIENTO DE LICENCIA SOFTWARE DE CONSULTA DE LAVADO DE ACTIVOS	Objeto
Código:	SIE-SEPS-001R-2020	Codigo
Tipo Compra:	Servicio	Tipo_compra
Presupuesto Referencial Total (Sin Iva):	USD 19,642.86	Presupuesto
Tipo de Contratación:	Subasta Inversa Electrónica	Contratacion
Forma de Pago:	Anticipo 0% Saldo: Pago Pago contra entrega de bienes o servicio 100.00%	Pago
Tipo de Adjudicación:	Total	Total_Adjudicacion
Plazo de Entrega:	10 días	Plazo de Entrega
Vigencia de Oferta:	90 días	Vigencia_Oferta
Funcionario encargado del proceso:	karina.bedon@seps.gob.ec	Funcionario_encargado
Estado del Proceso:	Finalizada	
Estado en el cual finalizó el Proceso	En Recepción	
Descripción:	CONTRATACIÓN DEL SERVICIO DE MANTENIMIENTO DE LICENCIA SOFTWARE DE CONSULTA DE LAVADO DE ACTIVOS	
Variación mínima de la Oferta durante la Puja:	1.00% Tipo Variación: Precio total	

Below the main form is a section titled "Fechas de Control del Proceso" with a field for "Fecha de Publicación" and a label "Fecha_Publicacion" with the instruction "Indicar la fecha real en la cual desea publicar el Proceso."

Nota. Formularios de extracción de datos de interés para el análisis. Fuente: Josue Coba y Estalin Revelo mediante la herramienta PDFElement.

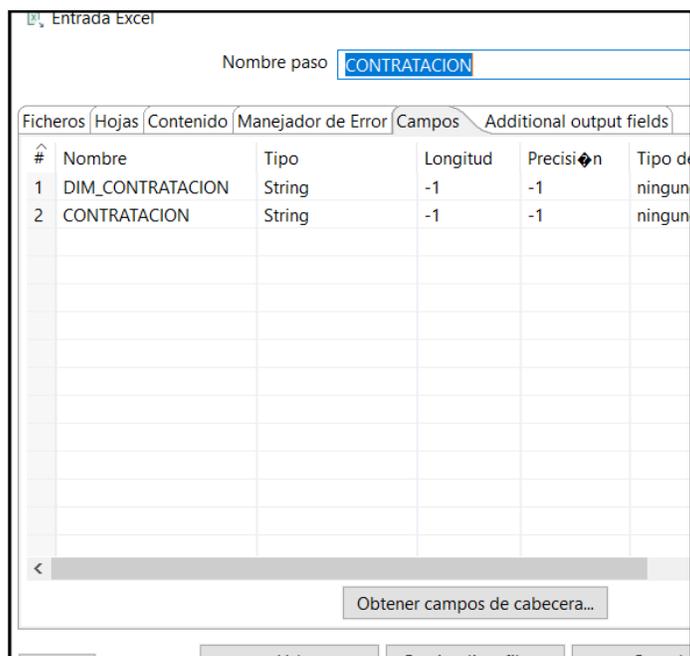
3.1.3 Proceso ETL a la base de datos

Mediante el uso de la herramienta Pentaho Data Integration (PDI) se obtuvo como resultado la extracción de datos ubicados en archivos Excel hacia tablas ubicadas en una base de datos en MySQL, para ello los datos deben estar depurados completamente, ya que este proceso es sensible a cambios como se muestra en la Figura 13 y Figura 14.

La Figura 13 presenta el tipo de información almacenada en los archivos CSV, se toma como ejemplo el documento de contratación este contiene 2 campos, el primero DIM_CONTRATACION de tipo cadena de carácter y clave primaria, el segundo es CONTRATACION de tipo cadena de carácter.

Figura 14

Visualización del tipo de dato en Pentaho Data Integration (PDI).



The screenshot shows the 'Entrada Excel' (Excel Input) tool configuration in Pentaho Data Integration. The 'Nombre paso' (Step Name) is 'CONTRATACION'. The 'Campos' (Fields) tab is selected, displaying a table with the following data:

#	Nombre	Tipo	Longitud	Precisión	Tipo de
1	DIM_CONTRATACION	String	-1	-1	ningun
2	CONTRATACION	String	-1	-1	ningun

At the bottom of the window, there is a button labeled 'Obtener campos de cabecera...' (Get header fields...).

Nota. Descripción del tipo de dato almacenado en el archivo CSV. Fuente: Josue Coba y Estalin Revelo mediante la herramienta Pentaho Data Integration (PDI).

La Figura 13 presenta en su columna izquierda los archivos CSV que contienen la información previamente depurada de cada uno (CONTRATACION, ENTIDAD, ESTADO, FECHA, OBJETO, PLAZO_ENTREGA), enlazados con las columnas de la parte derecha representando las tablas de cada una de las dimensiones de la base de datos (CONTRATACION 2, ENTIDAD 2, ESTADO 2, FECHA 2, OBJETO 2, PLAZO_ENTREGA 2) y su carga en estas al Data Warehouse.

Figura 15

Proceso de carga en Pentaho Data Integration (PDI).



Nota. Proceso de entrada Excel y salida a tablas en base de datos MySQL. Fuente: Josue Coba y Estalin Revelo mediante la herramienta Pentaho Data Integration (PDI).

3.1.4 Creación del Data Warehouse.

Mediante la herramienta Schema Workbench se crearon las dimensiones y medidas que serán utilizadas para el análisis, para ello se cuenta con jerarquías en donde se insertan

las tablas que se encuentran en la base de datos, así como su tipo. Otro elemento importante son las medidas que pertenecen al Data Warehouse, las cuales al igual que las dimensiones tiene conexión con las tablas en las bases de datos, además de establecer un tipo de dato y el agregado.

La Figura 14 se visualiza el esquema del Data Warehouse, donde se almacena las dimensiones de: Contratación, Entidad, Estado, Fecha, Objeto, Tipo compra y Vigencia Oferta, con sus 2 medidas respectivas de Presupuesto y Plazo_Entrega.

Figura 16

Dimensiones y medidas del Data Warehouse.

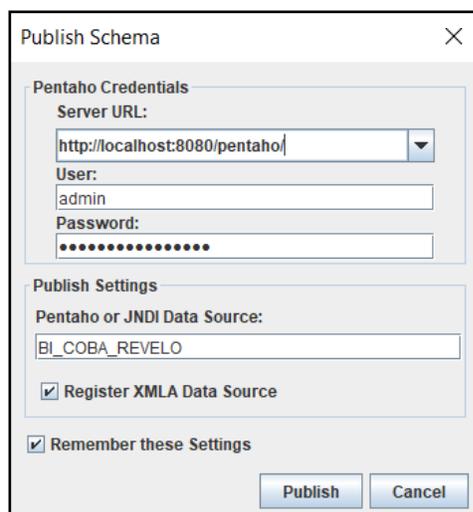


Nota. Establecer las medidas y dimensiones del Data Warehouse. Fuente: Josue Coba y Estalin Revelo mediante la herramienta Schema Workbench.

El siguiente proceso es la publicación del esquema o Data Warehouse hacia la plataforma de Pentaho Community como se muestra en la Figura 15.

Figura 17

Publicación del esquema.



Nota. Publicación del esquema hacia Pentaho Server mediante el administrador. Fuente: Josue Coba y Estalin Revelo mediante la herramienta Schema Workbench.

Como se muestra en la Figura 15 para la publicación es necesario contener las credenciales del administrador, ya que en su perfil se publicará el Data Warehouse, esto se realiza para evitar cualquier tipo de error por falta de permisos.

3.2 DESCRIPCIÓN DEL ACCESO A REPORTE

La solución de inteligencia de negocios cuenta con la generación de reportes, cabe aclarar que no todos los perfiles incluidos en el proyecto tienen libre acceso a la generación de reportes, por otro lado, todos los perfiles pueden visualizar los reportes estáticos ya creados.

Tabla 4*Perfiles y roles en la inteligencia de negocios.*

Usuario	Rol de Usuario	Funciones
Administrador	Administrador	Perfil con todos los permisos. Capacidad de creación de perfiles. Asignación de roles a perfiles. Generación de Reportes (estáticos y Ad-hoc). Acceso al navegador de archivos. Administrar la fuente de datos. Acceso a la documentación.
Público en General	Business Analyst	Acceso al navegador de archivos (reportes estáticos). Acceso a la documentación.
Analista	Power User	Generación de Reportes (estáticos y Ad-hoc). Acceso al navegador de archivos. Acceso a la documentación.

Nota. Roles de perfiles con sus funciones dentro de la inteligencia de negocios Fuente: Josue Caba y Estalin Revelo.

La tabla 3 realiza una referencia en cuanto a la segregación funcional de cada uno de los perfiles que los usuarios usen para el acceso a la herramienta Pentaho Server.

Como se menciona en el Anexo 2, el acceso a la solución está definida por roles, los cuales pueden ser modificados únicamente mediante el perfil del administrador, esto se implementó como medida de seguridad tanto para los usuarios como para los datos almacenados en inteligencia de negocios, una vez aclarado la generación de reportes, estos cumplen con dos características: la primera, guardar dentro de las carpetas de Pentaho Server para que sean visualizados por otros usuarios indistintamente de sus roles y la segunda Saiku Analytics permite la descarga de los reportes (Ad-hoc y estáticos) mediante los formatos

PDF, Excel, PNG y CSV, esta es una gran calidad, ya que se está trabajando con una herramienta gratuita.

3.3 LEVANTAMIENTO DE LA SOLUCIÓN EN EL SERVIDOR

Otorgar una guía de soporte para el manejo correcto del sistema con la finalidad de brindar navegabilidad continua en el sistema y explotar las herramientas que ofrece el sistema para el descubrimiento de información mediante la manipulación de los datos.

Tabla 5

Descripción del servidor.

DESCRIPCIÓN	NOMBRE DE LA MÁQUINA VIRTUAL	RAM	DISCO	CPU
WINDOWS 10	VM_TESIS_EREVELO_W10	6 GB	80 GB	4

Nota: Descripción del servidor implementado en el ICC de la Universidad

Politécnica Salesiana Campus Sur. Elaborado por: Coba J y Revelo E.

La Tabla 5 describe las características del servidor en cuanto a hardware y software implementado para el despliegue de la inteligencia de negocios.

3.3.1 *Requerimiento*

Para un correcto funcionamiento, Pentaho Server y XAMPP se recomienda

Tabla 6*Requisitos mínimos Hardware Pentaho Server*

Procesador	4 Cores
RAM	8 GB (4 GB para Pentaho Data Integration PDI Dedicado, 4 Servidor Apache Tomcat)
Espacio en disco	80 GB

Nota. Se presentan los requisitos mínimos de hardware para instalar Pentaho Server.

Fuente: Carlos A. Jaime M (2015).

La Tabla 6 describe las características en cuanto a hardware implementado en la solución de inteligencia de negocios.

Tabla 7*Requisitos mínimos Software Pentaho Server*

JRE	Java Runtime Environment Java 7
JDK	Java Developer Kit Java 7
Driver BD	Driver jdbc mysql-connector-java-8.0.23
Pentaho BI-Server	Pentaho BI Suite CE biserver-ce-5.4.0.1-130
Contenedor Web	Apache Tomcat 6.043
Procesamiento de datos (ETLs)	Pentaho Data Integration PDI CE pdi-ce-5.4.0.1
Front-End	Browsers Mozilla Firefox, Google Chrome o Internet Explorer

Nota. Se presentan los requisitos mínimos de software para instalar Pentaho Server. Fuente:

Carlos A. Jaime M (2015).

La Tabla 7 detalla la configuración de las librerías y herramientas que se instalaron para el despliegue de la inteligencia de negocios.

Tabla 8

Requisitos mínimos XAMPP

RAM	256 MB
Almacenaje	85 MB

Nota. Se presentan los requisitos mínimos para instalar XAMPP. Fuente: Jonex (2020).

La Tabla 8 detalla los requisitos de la herramienta XAMPP para el despliegue del proceso ETL en este.

3.3.2 Instalación y configuración

En este literal se contempla la instalación y configuración de los principales componentes del proyecto como es Pentaho Server y Saiku Analytics.

Instalación y configuración de Pentaho Server. El proyecto a presentarse utiliza una versión gratuita de Pentaho, denominada Pentaho Community el cual puede descargarse desde el repositorio oficial, para más detalle revisar el Anexo 2 en donde se puede encontrar las diferentes versiones que se han desarrollado de Pentaho Community, así como sus diferentes componentes: server, cliente, plugins entre otros. La inteligencia de negocios se encuentra desarrollada en Pentaho 9.1.

Descargado los archivos necesarios previamente requeridos, se procede con la configuración de las herramientas, para la iniciar el servicio de Pentaho Server es necesario contar con distribución de Apache y Tomcat, XAMPP es una herramienta que solventó dichas necesidades, ya que cuenta con los servicios integrados de MySQL, Apache y Tomcat.

Instalación y configuración de Saiku Analytics. La descarga de la herramienta Saiku Analytics se debe realizar desde el Marketplace de Pentaho Server. Una vez ingresado en el Marketplace se procede a la búsqueda y descarga de Saiku Analytics y su licencia free,

dispone un archivo que se encuentra alojado en un repositorio GitHub. Para más detalles de los pasos de instalación ver Anexo 2.

<http://192.168.1.118:8080/pentaho/>

3.4 HALLAZGOS DE INFORMACIÓN

Mediante la reunión mantenida con el tutor del proyecto de titulación, sugirió un análisis de estos datos en los cuales se realizaron los siguientes descubrimientos:

- A. Análisis de la dimensión del tiempo de entrega de los proyectos en base a la medida de presupuesto del año 2010.

Este hallazgo involucra las dimensiones de tipo compra, vigencia oferta, año y la medida de presupuesto, esta se detalló en la Figura 16.

Figura 18

Dimensiones y medidas para realizar el análisis del hallazgo A.

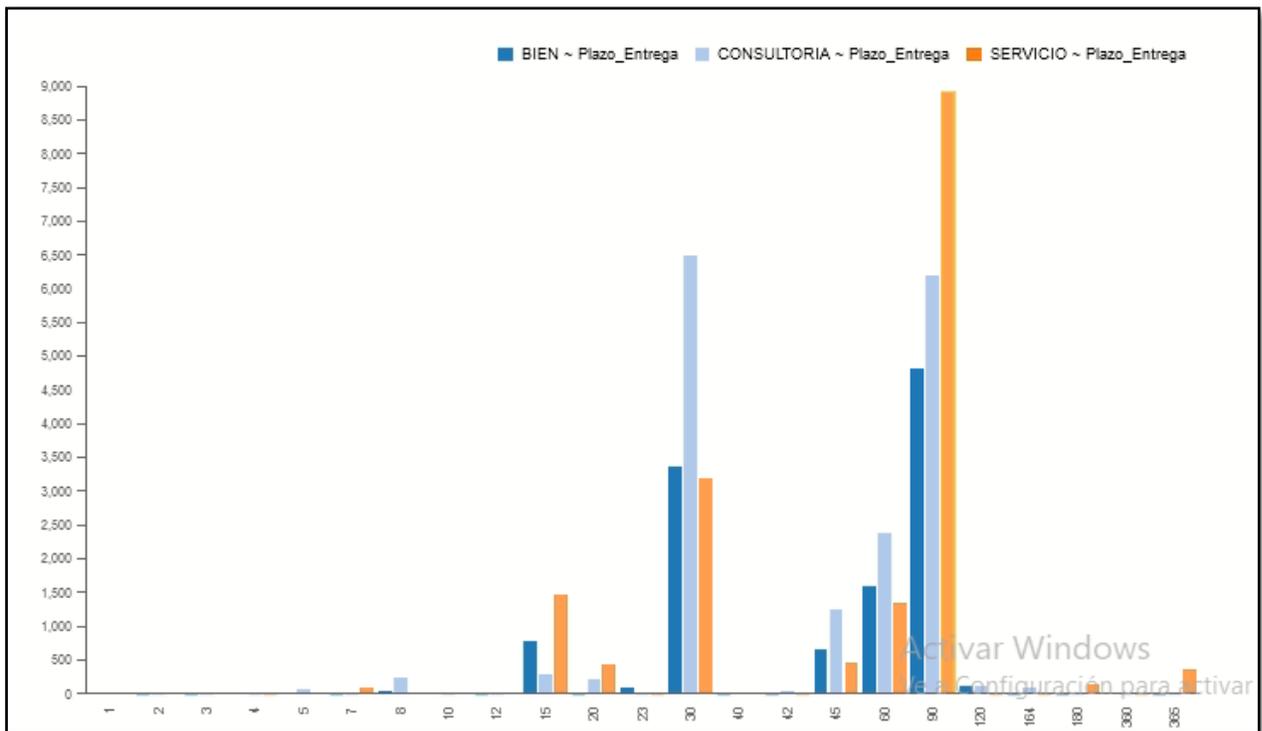


Nota. Las dimensiones que se usaron para realizar el hallazgo A fueron: “Presupuesto, tiempo de entrega de los proyectos y el año 2010”. Elaborado por: Josue Coba y Estalin Revelo mediante la herramienta Saiku Analytics.

Se observa, gráficamente el tiempo máximo que se puede cumplir con los diferentes proyectos de software y el costo de cada una de sus formas de implementación, estos segregados por el año 2010, representada en la Figura 17, se evidencia que el mayor costo de un proyecto es de 9000 dólares en servicios en un periodo de 90 a 120 días en el año 2010.

Figura 19

Reporte Hallazgo A.



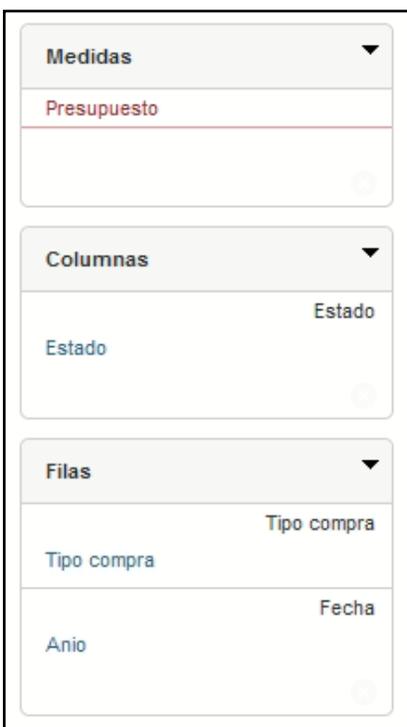
Nota. Reporte del hallazgo en base al análisis de las dimensiones “Presupuesto y tiempo de entrega de los proyectos”. Elaborado por: Josue Coba y Estalin Revelo mediante la herramienta Saiku Analytics.

B. Presupuesto por cada año de acuerdo al tipo de compra y el estado de los proyectos de software.

Este hallazgo involucra las dimensiones de tipo de compra, año y estado y la medida de presupuesto, está se detalla en la Figura 18.

Figura 20

Dimensiones y medidas para realizar el análisis del hallazgo B.

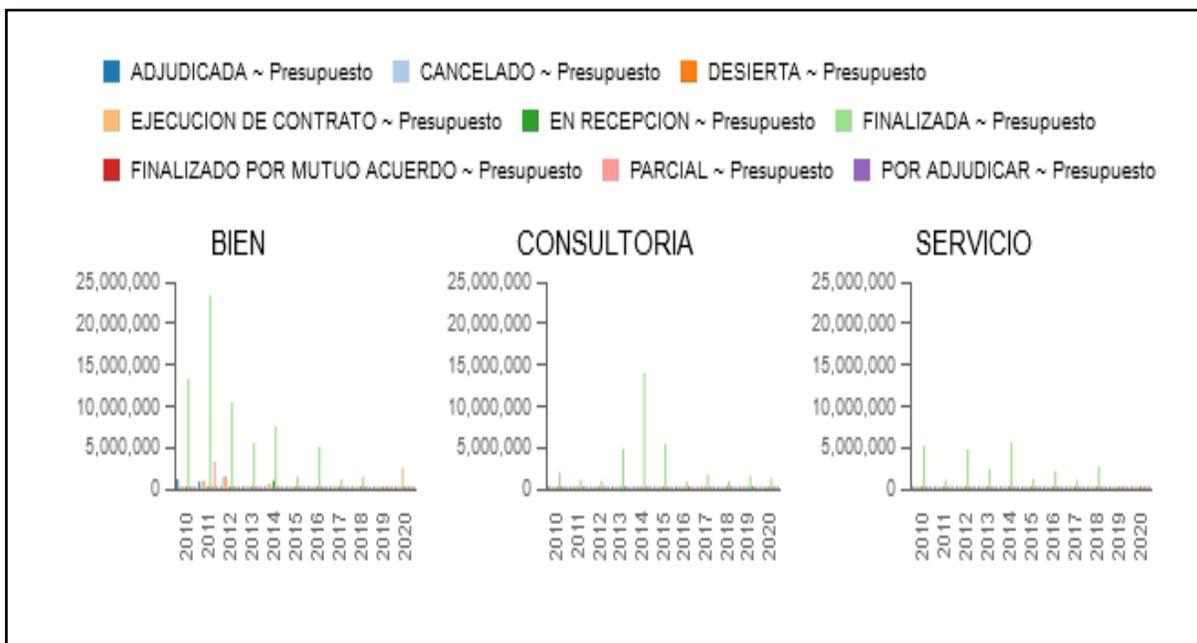


Nota. Descripción las dimensiones que se usaron para realizar el hallazgo B estas fueron: “Presupuesto, estado, tipo compra y año desde el 2010-2020”. Elaborado por: Josue Coba y Estalin Revelo mediante la herramienta Saiku Analytics.

Se visualiza gráficamente que el costo de la adquisición en cuanto a bienes del Estado tiene un rango entre 20 y 25 millones de dólares de los proyectos finalizados desde el año 2010 hasta el 2020 resaltando en este el año 2011, solicitar una consultoría por parte del estado se encuentra en un rango de 10 y 15 millones de dólares en el mismo periodo y estado de los proyectos, el Estado invierte en contratación de servicios un presupuesto máximo de 5 millones de dólares, esto se detalla en la Figura 19.

Figura 21

Reporte hallazgo B.



Nota. Reporte del Hallazgo B en base al análisis de las dimensiones “Tipo de compra, año y estado en base al presupuesto, segregado por cada año.” Elaborado por: Josue Coba y Estalin Revelo mediante la herramienta Saiku Analytics.

C. Asignación de presupuesto de acuerdo al año y de acuerdo al tipo de contratación y su estado.

Este hallazgo contempla las dimensiones de año, contratación, estado y la medida de presupuesto, que se detalla en la Figura 20.

Figura 22

Dimensiones y medidas para realizar el análisis del hallazgo C.



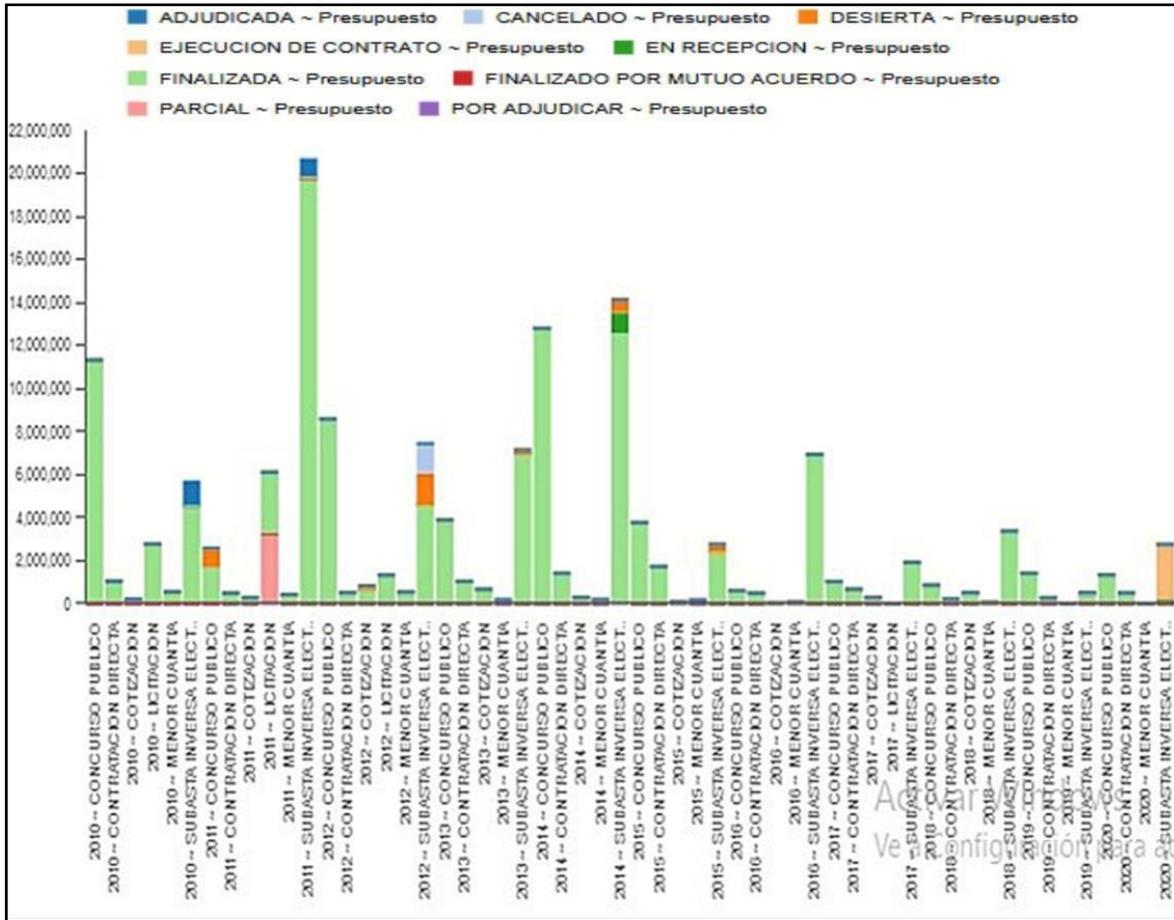
Nota. Reporte del Hallazgo C en base al análisis de las dimensiones “año, contratación, estado y la medida de presupuesto segregado por cada año.” Elaborado por: Josue Coba y Estalin Revelo mediante la herramienta Saiku Analytics.

Se visualiza gráficamente el costo en cuanto al estado de los proyectos esta toma valores como Adjudicó, cancelado, desierto, ejecución de contrato, en la recepción, finalizada, finalizado por mutuo acuerdo, parcial y por adjudicar este se encuentra segregado por año.

En el año 2011 se observa que el costo de inversión del Estado se encuentra entre 20 y 22 millones de dólares para los proyectos que se encuentren adjudicados, esto significa que cada uno de los participantes ha cumplido con los requerimientos solicitados en cada uno de los contratos, este proceso se realizó por medio de subasta inversa electrónica.

Figura 23

Reporte Hallazgo C.



Nota. Reporte del Hallazgo C, visualización de las dimensiones de acuerdo a “Asignación de presupuesto desde el año 2010 hasta el 2020, su tipo de contratación y estado del proyecto”

Elaborado por: Josue Coba y Estalin Revelo. Mediante la herramienta Saiku Analytics.

CONCLUSIONES

- Se realiza una encuesta a diferentes personas para identificar la importancia de los datos, que se extrajeron de los archivos PDF, a esta respondieron 32 personas siendo las dimensiones y medidas más importantes: Presupuesto, Tipo Contratación, Plazo_Entrega y Fecha para lo cual se tomó de referencia estas para la implementación del Data Warehouse, al mismo tiempo se identificó que la mayoría de personas prefiere visualizar los datos en barras y extraerlos en formatos PDF, para lo cual se ejecutó mediante la implementación de la herramienta Saiku Analytics.
- En el proceso ETL se debe ser muy riguroso en la minería y limpieza de datos ya que en cada una de estas etapas se encuentran con errores comunes como la confusión de caracteres, errores por parte de los sistemas OCR, errores de almacenamiento, información faltante, por medio de la metodología incremental se perfecciona el proceso ETL, evitando así la falta de veracidad en los datos para futuros análisis.
- El proceso de carga de la información a la base de datos se visualiza un error al conformar el Data Warehouse puesto que el análisis de datos en cuanto a la dimensión del campo objeto contaba con cadenas de caracteres extensas, esto provocó un retraso en la carga y análisis de los datos, esto se logró validar en el proceso ETL mediante versiones implementadas, posibilitando la integración de los datos hacia la herramienta Pentaho, desplegando así los datos con mejor tiempo.

- La metodología incremental implementada con la inteligencia de negocios obtiene una retroalimentación constante en el proceso ETL, alcanzando así un menor riesgo ante cambios significativos en etapas finales del proyecto, demostrando la compatibilidad con la metodología Kimball.
- La creación de reportes estáticos y Ad-hoc mediante la herramienta Saiku Analytics proporciona un análisis con mayor comprensión acerca de los factores como el costo y tiempo en el ciclo de vida de los proyectos de software por parte de entidades públicas.

RECOMENDACIONES

- Los datos como fuente principal del presente proyecto se debieron tener sumo cuidado especialmente en el proceso ETL, es por ello que antes de realizar dicho proceso es indispensable realizar depuraciones minuciosas, esto con la finalidad de evitar cualquier tipo de errores a futuro.
- Durante el levantamiento de la solución de negocios se presentaron ciertos inconvenientes, en especial en la creación de reportes por parte de la solución, ya que se trabajó con una licencia libre de Saiku Analytics, con lo cual en primeras instancias la herramienta no respondía y se dio solución a este inconveniente mediante la descarga de licencia gratuita en un repositorio externo.
- Para futuros proyectos que se realicen con la misma temática y en especial en el área de reportes, una buena práctica es adquirir la licencia de pago de la herramienta de Saiku Analytics, otorgando más funcionalidades que su versión gratuita como por ejemplo la posibilidad de crear cuadros de mando.
- Adquirir un servidor público en donde se aloje la solución, con la finalidad de conceder acceso y hacer uso de la misma a los ciudadanos en cualquier lugar, dicho servidor debe tener protocolos de seguridad y la capacidad de operar sin interrupciones para otorgar una experiencia satisfactoria a los ciudadanos.
- Implementar un proceso automatizado para la recolección y actualización de los datos que alimentan a la solución de inteligencia de negocios logrando así el uso continuo de la misma.

REFERENCIAS BIBLIOGRÁFICAS

- KIMBALL, R. (2013). The Data Warehouse Toolkit Classics: The Data Warehouse Toolkit, 3rd Edition.
- Vucevic, D., & Yaddow, W. (2012). Testing the data warehouse practicum: Assuring data content, data structures and quality. Trafford Publishing.
- Kimball, Ralph, et. al. The Data Warehouse Lifecycle Toolkit, John Wiley & Sons, 1998.
- Díaz, J. C. (2012). Introducción al business intelligence. Editorial UOC.
- Toledo, F. (2017, 10 julio). Testing en Business Intelligence. federico-toledo.com. <https://www.federico-toledo.com/testing-en-business-intelligence/>
- Ramos, S. (2019, 29 octubre). Toma de requisitos para la creación de sistemas de BI: Procesos de Negocios (05). SolidQ Blogs. <https://blogs.solidq.com/es/business-analytics/requisitos-procesos-de-negocios-business-intelligence-05/>
- Costanzo, M. (2020). Modelo de desarrollo incremental iterativo de software. mwebs. <https://mwebs.com.uy/blog/qu%C3%A9-es-el-desarrollo-incremental/23>
- Ortiz, M. (2012, septiembre). Modelo Incremental. Blogger. <http://isw-udistrital.blogspot.com/2012/09/ingenieria-de-software-i.html>
- Pérez, A. (2016, 16 agosto). Características y fases del modelo incremental. OBS Business School. <https://www.obsbusiness.school/blog/caracteristicas-y-fases-del-modelo-incremental>.
- Cahuana, M. (2015, 7 septiembre). Ingeniería de software modelo incremental. slideshare. <https://es.slideshare.net/mardluz/ingeniera-de-software-modelo-incremental>

Andrade, C., Proaño, J., Ricaurte, D., & Sáenz, J. (2012). Estudio de Mercado Servicio Desarrollo de Software en Ecuador 2012. docplayer. <https://docplayer.es/609232-Estudio-de-mercado-servicio-desarrollo-de-software-en-ecuador.html>.

Bermeo, L., & Simbaña, E. (2020). CREACIÓN DE UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS PARA LA GENERACIÓN DE REPORTES CON INFORMACIÓN RELEVANTE DEL PROCESO DE TITULACIÓN DE LOS ESTUDIANTES DE LA CARRERA DE INGENIERÍA DE SISTEMAS DE LA UNIVERSIDAD POLITÉCNICA SALESIANA SEDE QUITO. [Tesis de ingeniería]. Universidad Politécnica Salesiana

Sevilla, E. (2003). GUÍA METODOLÓGICA PARA LA DEFINICIÓN Y DESARROLLO DE UN DATA WAREHOUSE. [Tesis de ingeniería]. Universidad Americana Facultad de Ingeniería.

Alfaro Viana, I. A., & Vargas León, E. (2016). Diseño del plan de seguridad informática del sistema de información misional de la Procuraduría General de la Nación [Especialista en Seguridad Informática]. Universidad Piloto de Colombia.

Domínguez Martínez, J. A. (2008). Diseño de un modelo multidimensional de data mart del área de capacitación en el INEGI [Maestría en informática y tecnologías computacionales]. Universidad Autónoma de Aguascalientes.

Sánchez, J. (2018). Aplicación del Modelo Incremental para el desarrollo del sistema de Información Docente [Tesis de Ingeniería en Computación]. Universidad Autónoma del Estado de México.

ANEXOS

Para revisión de anexos del trabajo, solicite a la universidad

Anexo 1: Informe de Pruebas.

Anexo 2: Manual del Usuario.

Anexo 3: Plan de Pruebas.

Anexo 4: Diccionario de datos.

Anexo 5: Informe de Versiones.

Anexo 6: Encuesta a usuarios