



UNIVERSIDAD CÉSAR VALLEJO

FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

Tablero de mando basado en Linux para la Planta HFC del Grupo IMATRA SAC

TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:
Ingeniero de Sistemas

AUTORES:

Condori Juárez, Jonathan Roberto (ORCID: 0000-0002-0958-8125)

Muñoz Ronceros, Jim David (ORCID: 0000-0002-5947-9998)

ASESOR:

Dr. Hilario Falcon, Francisco Manuel (ORCID: 0000-0003-3153-9343)

LÍNEA DE INVESTIGACIÓN:

Sistema de Información y Comunicaciones

LIMA – PERÚ

2021

Dedicatoria

A Dios, por brindarnos la fuerza, salud necesaria para poder realizar y cumplir nuestros objetivos, derramándonos su bendición y llenarnos de un alma espiritual durante todo este proceso de vida.

A nuestros padres y todos nuestros seres queridos que nos ofrecieron su ayuda incondicional para seguir adelante en este proyecto de tesis y cumplir nuestros objetivos para alcanzar el título universitario.

Agradecimiento

A nuestro asesor de investigación, por la oportunidad de compartir sus conocimientos en este camino de la investigación y poder culminar nuestra Tesis.

A la empresa Grupo IMATRA, por brindarnos su apoyo en esta experiencia y poder utilizar su información en este gran proyecto de estudio.

Índice de contenidos

Carátula	i
Dedicatoria	ii
Agradecimiento	iii
Índice de Contenidos	iv
Índice de tablas	v
Índice de figuras	vi
RESUMEN	viii
ABSTRACT	ix
I. INTRODUCCIÓN	1
II. MARCO TEÓRICO	8
III. METODOLOGÍA	26
3.1 Tipo y diseño de investigación	27
3.2 Variables y operacionalización	27
3.3 Población, muestra y muestreo	28
3.4 Técnicas e instrumentos de recolección de datos	29
3.5 Procedimientos	30
3.6 Método de análisis de datos	31
3.7 Aspectos éticos	31
IV. RESULTADOS.....	33
V. DISCUSIÓN.....	46
VI. CONCLUSIONES.....	49
VII. RECOMENDACIONES	52
REFERENCIAS	54
ANEXOS.....	57

Índice de Tablas

Tabla 1 Tabla de variable	32
Tabla 2 Matriz correlacional de la hipótesis general	33
Tabla 3 Matriz correlacional específica 1	34
Tabla 4 Matriz correlacional específica 2	35
Tabla 5 Matriz correlacional específica 3	37
Tabla 6 Cuadro comparativo de la variable	38
Tabla 7 Mejora de los datos	39
Tabla 8 Transformación de los datos	40
Tabla 9 Almacenamiento de datos	41
Tabla 10 Tiempo de Respuesta	42
Tabla 11 Resumen de los resultados de las comprobaciones de las hipótesis	43
Tabla 12. Ficha de Observación	62
Tabla 13 Ficha de Observación	63
Tabla 14 Evaluación de dimensión e indicadores	64
Tabla 15 Validación de instrumento	66
Tabla 16 Cuestionario - Fuente: Elaboración propia	67
Tabla 17 Matriz de datos – Fuente: SPSS Statistics	69
Tabla 18 Esquema de trazabilidad: tabla: Extract_definition	77
Tabla 19 Esquema de trazabilidad: tabla: Standard_definition	77
Tabla 20 Esquema de trazabilidad tabla: structure_definition	78
Tabla 21 Esquema de trazabilidad tabla: source_definition	78
Tabla 22 esquema de trazabilidad tabla: source_field	79
Tabla 23 Esquema de trazabilidad tabla: target_definition	79
Tabla 24 esquema de trazabilidad tabla: target_field	80
Tabla 25 Esquema de trazabilidad tabla: transformation_definition	80
Tabla 26 Esquema de trazabilidad tabla: target_export	81
Tabla 27 esquema de trazabilidad tabla: export_definition	81
Tabla 28 Tabla de Credenciales	92
Tabla 29 Esquema de Tablas	95
Tabla 30 Entorno de software y herramientas	110
Tabla 31 configuración de fechas de extracción	112

Índice de figuras

Figura 1 Los sistemas de información en la organización (Laudón y Laudón 2012 p. 13).....	12
Figura 2 Sistemas de información más efectiva (Laudón y Laudón 2012 p. 18).....	13
Figura 3 Áreas del conocimiento en el sistema de información (O'Brien, Marakas 2006 p. 7)_	14
Figura 4 Almacén de datos. Los componentes de un sistema completo de almacén de datos. (O'Brien, Marakas 2006 p. 145).....	16
Figura 5 Almacén de datos y sistemas operacional. (Laudón y Laudón 2012 p. 223).....	16
Figura 6 Almacén de datos y sus subconjuntos. Un almacén de datos y sus subconjuntos de mercados (O'Brien, Marakas 2006 p. 146).....	17
Figura 7 Análisis de datos. (Laudón y Laudón 2012 p. 472).....	18
Figura 8 Modelado de la información. (Laudón y Laudón 2012 p. 16).....	19
Figura 9 Requerimientos de información por jerarquía. (Laudón y Laudon2012 p. 456).....	20
Figura 10 Resumen de los atributos de la calidad de información. (O'Brien, Marakas 2006 p. 323).....	22
Figura 11 Dimensiones de la calidad de información. (Ohlborst, Frank 2012).....	23
Figura 12 Esquema Hadoop Fuente: Universidad Nacional de Entre Ríos (2016), Sitio Web: http://sedici.unlp.edu.ar/bitstream/handle/10915/52766/Documento_completo.p df?sequence=1&isAllowed=y	23
Figura 13 Arquitectura de datos.....	71
Figura 14 Repositorio de datos.....	82
Figura 15 Flujo de datos del proceso NIFI.....	85
Figura 16 Esquema traceability.....	86
Figura 17 Esquema de control.....	87
Figura 18 Configuración de la tabla Export_definition.....	88
Figura 19 Configuración de datos de la tabla extract_definition.....	88
Figura 20 Configuración de los datos de la tabla source_definition.....	89
Figura 21 Configuración de los datos de la tabla source_field.....	89
Figura 22 Diagrama de flujo: Proceso de obtención de fuentes.....	90
Figura 23 Diagrama de flujo: Proceso de estandarización.....	91
Figura 24 Credenciales entorno Linux.....	92
Figura 25 Entorno Linux.....	93
Figura 26 Listado de scripts.....	94
Figura 27 Esquema dev_stage_gt.....	95
Figura 29 Tablas de configuración.....	96
Figura 30 Script create_table: ubb_docsis_prueba.....	97
Figura 31 Creación de una tabla en HIVE.....	98
Figura 32 Listado de planos del entorno HDFS – Hadoop.....	99
Figura 33 Carpetas de validación de del proceso de archivo plano.....	100
Figura 34 NIFI Entorno de los process group.....	101
Figura 35 Proceso de extracción de los planos ETL.....	101
Figura 36 proceso de extracción por FTP.....	102
Figura 37 Desarrollo en Phyton de la extracción de los datos.....	103
Figura 38 Desarrollo del proceso de ingesta.....	104
Figura 39 Flujo del proceso de desarrollo de la arquitectura completa.....	105
Figura 40 Flujo del proceso de estandarización hacia la ingesta de la data en las tablas.....	106
Figura 41 Proceso de exportación y obtención de los reportes finales.....	107
Figura 42 Resultado de los reportes finales obtenidos.....	108
Figura 43 Tabla de KPI de resultados por semana en un porcentaje de percentil.....	108
Figura 44 Estructura del reporte final de saturación de puertos.....	109
Figura 45 Process group de prueba.....	111
Figura 46 Conexión del Proceso del ListSFTP.....	112
Figura 47 Configuración del proceso ListSFTP.....	113

Figura 48 Proceso DescomprimirArchivoEdge.....	113
Figura 49 Proceso de configuración de variables para descomprimir plano de una ruta especifica.....	114
Figura 50 Proceso de Actualización de parámetros.....	114
Figura 51 Configuración de procesos de actualización de parámetros.....	115
Figura 52 Proceso de carga de los planos hacia la ruta HFS.....	115
Figura 53 Configuración del proceso para mover el plano hacia una ruta HFS.....	116
Figura 54 Proceso del plano HFS.....	116
Figura 55 Shells del proceso de almacenamiento.....	117
Figura 56 proceso de carga hacia una tabla en HIVE.....	117
Figura 57 configuración de proceso de ingesta a HIVE.....	117
Figura 58 validación de la cantidad de registros de la carga ingestada en HIVE.....	118
Figura 59 Muestra de resultados de una tabla injestada en HIVE.....	118
Figura 60 Muestra del proceso finalizado correctamente desde el LogFile en NIFI.....	118

Resumen

En la presente investigación se planteó un Tablero de mando basado en Linux para la toma de decisiones de la planta HFC en el área de operaciones del Grupo Imatra SAC.

En el cual la problemática que presenta en el área de operaciones es en los tiempos de espera para obtener los reportes finales para que puedan tomar decisiones con respecto al análisis de los datos transformados, para obtener estos reportes el área de operaciones lo extrae de una forma manual de otros pequeños sistemas de apoyo, lo cual les toma mucho tiempo procesarlos.

La población estuvo conformada por el área de operaciones con un total de 35 personas, el cual estuvo liderada por 1 gerente y 30 gestores. El trabajo de investigación fue realizado con una muestra de 30 personas.

La presente investigación fue de enfoque cuantitativa, tipo de investigación aplicada, y nivel de investigación Explicativa y el diseño de investigación No experimental – Transversal.

Como resultado se ha obtenido un Tablero de mando para la toma de decisiones de la planta HFC en el área de operaciones del Grupo Imatra SAC, cabe indicar que el sistema cumplió con los objetivos y se logró optimizar los tiempos de espera tanto en reproceso como en los reportes finales, se puede hacer análisis de alto nivel para predecir en el tiempo y tener una buena toma de decisión.

Palabras clave: Tablero de mando, Toma de decisión, área de operaciones, optimización, reportes.

Abstract

In the present investigation a Linux-based dashboard was proposed for the decision making of the HFC plant in the operations area of Grupo Imatra SAC.

In which the problem that presents in the operations area is the waiting time to obtain the final reports so that they can make decisions regarding the analysis of the transformed data, to obtain these reports the operations area extracts them manually from other small support systems, which takes them a long time to process them.

The population consisted of the operations area with a total of 35 people, which was led by 1 manager and 29 managers. The research work was carried out with a sample of 30 people.

The present research was of quantitative approach, type of applied research, and Explanatory research level and Non-experimental - Transversal research design.

As a result, a dashboard has been obtained for decision making in the HFC plant in the operations area of Grupo Imatra SAC, it should be noted that the system met the objectives and it was possible to optimize the waiting times both in reprocessing and in the final reports, it is possible to make high-level analysis to predict in time and have a good decision making.

Keywords: Dashboard, decision making, operations area, optimization, reports.

I. INTRODUCCIÓN

Actualmente en el mundo los SIG se han consolidado como un gran factor importante dentro de la organización de las empresas.

Su importancia en el crecimiento dentro de las finanzas, contabilidad, administración, etc. (2020); muestra un generoso desarrollo de sistemas de información, destinados para el manejo integral de los activos corporativos, son un conjunto de aplicaciones que se conectan entre sí, proporcionando datos para las necesidades operativas y administrativas de una organización; La meta primaria es mostrar una visión global de las circunstancias ordinarias de la organización para diseñar, controlar, coordinar y establecer las mejores opciones para el avance fructífero de una asociación.

En el caso principal, la innovación mantiene capacidades, por ejemplo, la correspondencia en medios generales (comunicación fija, portátil o VOIP), la correspondencia compuesta (correo electrónico, SMS, visita), la admisión de información debido a la digitalización, el acopio y la transmisión de registros de la organización y la indagación de datos en Internet.

Las obtenciones de mejores resultados para beneficio de la empresa van de la mano con el uso de la tecnología para certificar la calidad, inteligencia competitiva, automatización de procesos y sistemas para toma de decisiones de alto nivel. Monforte (1994) caracteriza al sistema estratégico de información como: “parte importante de la empresa, lo que conlleva una ventaja competitiva por sí mismo, debido a que forma parte esencial del negocio y contribuye como una cualidad única a los productos, operaciones o toma de decisiones”, Laudón(2012) define un sistema de información a un conjunto de partes interrelacionadas que reúnen (o recuperan), procesan, almacenan y distribuyen información para ayudar a las medidas dinámicas y de control en una asociación.

Actualmente a nivel Sudamérica se cuenta con sistemas de apoyo para la toma de decisión basado en sistemas integrales para obtener mejores resultados, optimización de procesos, automatización y mejora en los tiempos de respuesta de las organizaciones, Las futuras asociaciones deberían pensar en nuevas normas para las empresas y centrarse en las medidas como punto central de consideración, que no se tienen en cuenta en las asociaciones habituales. Sathiri

(2015) la finalidad del SIG es la de mejorar los activos y crear respuestas para los beneficiarios del sistema, así como los usuarios; esto es generalmente concebible con estructuras de TI más adaptables y maniobrables en sus productos, Los sistemas educativos están en evolución constante debido a la necesidad de un desarrollo informático a la par.

Actualmente en el Perú se cuenta con muchos sistemas de integración para poder tener un control de las operaciones de ciertas áreas establecidas, pero a que costo o tiempo de respuesta se demoran en obtener un resultado para poder tomar una decisión. El cambio y la implementación de nuevas herramientas para el apoyo de la gestión de trabajo hacen que no actualicen el modelo de trabajo ya que no están acostumbrados al cambio.

Actualmente el GRUPO IMATRA S.A.C no cuenta con un SIG para la planta HFC, sin embargo, cuenta con otras plataformas de monitoreo servicios de la planta DHFC, monitoreo de fuente de energía y otras plataformas que brindan información relevante para las operaciones de la red HFC de GRUPO IMATRA S.A.C.

Esta actividad hasta la fecha se extrae la información de varias plataformas, lo que existe una alta probabilidad en los tiempos de espera y pérdida de información, de los cuales se generen errores en las mediciones de la planta HFC y no sean reportes con alto valor.

Seguidamente mostramos las justificaciones de nuestra investigación:

La justificación teórica en este estudio de investigación se realiza con la finalidad del conocimiento de los SIG en la automatización de los procesos, cuyos resultados de esta investigación es mejorar la toma de decisiones en los reportes finales de los Dashboards de la planta HFC. Vargas (2018) la investigación permite realizar tecnologías innovadoras para el desarrollo de sistemas de información ejecutiva para el control integral.

Actualmente para obtener toda la información de la planta HFC se obtiene de varias aplicaciones, lo cual dificulta el tiempo de espera de los reportes para poder hacer un análisis correspondiente.

La solución planteada está basada en un ambiente Big Data; las de base de datos PostgreSQL y a la par la visualización de los reportes Excel.

La justificación tecnológica en este proyecto de investigación se realiza por qué se realizó un análisis de impacto de las áreas involucradas al obtener los reportes de sus sistemas, para ello se pidió nuevas tecnologías de desarrollo para un sistema que contenga todos los procesos de la planta HFC para lo cual afianzaremos en SIG para la toma de decisiones en las operaciones de redes y obtener reportes óptimos para que los involucrados puedan hacer su análisis y obtengan resultados óptimos. Santa Cruz(2018) alcanzar una alta calidad en el servicio a nuestros clientes, comparar resultados y mejorar las fallas en la organización son puntos de suma importancia dentro de un SIG.

Para la Justificación económica de esta investigación se tomará las jornadas necesarias para poder hacer el proyecto que beneficiará a los involucrados del negocio. GRUPO IMATRA SAC cuenta con los ingresos de los proyectos de sus clientes. Paulsen (2020) la implementación de un SIG para el área de Bienestar Laboral ocasionó un gran impacto económico dentro de la organización.

En la Justificación metodológica de esta investigación, la metodología a implementar en este proyecto plantea nuevas técnicas de desarrollo para medir los SIG para la toma de decisiones en la planta HFC en el área de operaciones de GRUPO IMATRA SAC, el instrumento usado fue un cuestionario, compuesto por 26 preguntas del sistema implementado. Vargas (2018) mediante el proyecto se logra indagar diversas metodologías validadas y probadas en diversas situaciones, en esta ocasión en el campo de la Salud.

En lo referente a la realidad problemática de nuestra investigación se planteó el problema general y específico. El problema general de la investigación

fue ¿De qué manera la implementación de un Tablero de mando basado en Linux va a mejorar la toma de decisiones para la Planta HFC en el área de operaciones del Grupo IMATRA SAC?

A continuación, se muestra los problemas específicos:

- **PE1:** ¿Cómo la implementación del Tablero de Mando basado en Linux reduce los riesgos en la toma de decisiones de la Planta HFC en el área de operaciones del Grupo IMATRA SAC?
- **PE2:** ¿Cómo la implementación del Tablero de Mando basado en Linux influye en la mejora de los procesos para la toma de decisiones de la Planta HFC en el área de operaciones del Grupo IMATRA SAC?
- **PE3:** ¿Cómo la implementación del Tablero de Mando basado en Linux influye en los resultados para la toma de decisiones de la Planta HFC en el área de operaciones del Grupo IMATRA SAC?

El objetivo general fue Implementar un Tablero de Mando para la mejora de la toma de decisiones de la Planta HFC en el área de operaciones del Grupo IMATRA SAC. Se muestra a continuación los objetivos específicos:

- **OE1:** Determinar cómo el Tablero de Mando basado en Linux reduce los riesgos en la toma de decisiones de la Planta HFC en el área de operaciones del Grupo IMATRA SAC.
- **OE2:** Identificar la influencia de un Tablero de Mando basado en Linux en la mejora de los procesos para la toma de decisiones de la Planta HFC en el área de operaciones del Grupo IMATRA SAC.
- **OE3:** Determinar la influencia de un Tablero de Mando basado en Linux en los resultados para la toma de decisiones de la Planta HFC en el área de operaciones del Grupo IMATRA SAC.

La hipótesis general de la investigación fue La Implementación de un Tablero de Mando mejoró significativamente la toma de decisiones de la Planta HFC en el área de operaciones del Grupo IMATRA SAC.

- **HE1:** La Implementación de un Tablero de Mando logró reducir los riesgos en la toma de decisiones de la Planta HFC en el área de operaciones del GRUPO IMATRA SAC, Santiago de Surco. Saldaña (2020) concluyó que el rendimiento de TI es óptimo, teniendo un alto rendimiento en las probabilidades estadísticas de estudio de los indicadores.
- **HE2:** El Tablero de Mando mejoró los procesos en la toma de decisiones de la Planta HFC en el área de operaciones del Grupo IMATRA SAC, Santiago de Surco. Domínguez (2012) los SI son considerados como una gran fuente de análisis y respuesta de alto impacto, obteniendo información para su explotación.
- **H3:** El Tablero de Mando influyó en los resultados para la toma de decisiones de la Planta HFC en el área de operaciones del Grupo IMATRA SAC. Hopkin (2017 pág. 73) la estrategia de peligro de la junta directiva establecerá las responsabilidades relativas al peligro, al igual que los planes de juego para ejecución del enfoque. Las convenciones del tablero de peligro se establecerán en una progresión de reglas y convenciones de riesgo para conducir la evaluación del procedimiento.

II. MARCO TEÓRICO

En este capítulo que se realiza tiene como finalidad obtener información óptima y concreta acerca de los sistemas de información gerencial y toma de decisión, se ubicaron los siguientes trabajos vinculados con la tecnología, procesos en big data utilizadas en la presente investigación.

Esta parte del capítulo presenta los antecedentes de los trabajos de investigación, donde se incluyeron estudios relacionados con sistemas de información gerencial y principal proceso dentro de la organización. Se tiene Gurmendi (2017), para lograr la mejora de procesos a través de la toma de decisiones realizó un sistema de información gerencial. También, Santa Cruz (2018) detallar la influencia existente ente la gestión de calidad del servicio de la empresa y el sistema de información gerencial es su principal factor en esta investigación.

Pezo (2017) tiene como principal objetivo implementar un sistema de información gerencial para que nos permita mejora en la toma de decisión, ayudando de esta forma a las áreas involucradas. Con esta herramienta nos permite mejorar la calidad de servicio de los asegurados del Hospital II-2 Tarapoto. Pezo (2017) concluyó con los resultados obtenidos fueron correctos, se utilizó un diseño pre – experimental, mejoraron los procesos rotundamente obteniendo indicadores favorables. Se obtuvo una muestra de 6 personas de estadísticas del Hospital II-2 Tarapoto.

De acuerdo con Santa Cruz (2018) la empresa Casconta EIRL Ventanilla - Callao y su principal factor del sistema de información de cómo afecta en la calidad de sus servicios. Santa Cruz (2018) el método utilizado fue deductivo, con una justificación descriptiva, de tipo aplicada, diseño no experimental a nivel transversal. Se tiene una muestra de 40 trabajadores de la organización Casconta E.I.R.L. Santa Cruz (2018) concluyó, que el SI de Gestión obtiene una variación del 43.6% de calidad de servicio ya que los resultados determinan que $R^2 = 0.436$. Mejorando de esta forma la calidad de la organización.

Vargas (2018) el despliegue de un sistema de información gerencial para mejorar el control de los procesos de hemodiálisis. Vargas (2018) para

este caso de estudio, se utilizó metodología ágil ICONIX, la programación se desarrolló en PHP y para el almacenamiento de la información se utilizó MySQL. Tiene una investigación aplicada, con una población obtenida de la sesión de hemodiálisis desarrollada en el mes de octubre del 2018. Vargas (2018) se concluyó que para los 3 indicadores los resultados obtenidos de las sesiones de hemodiálisis disminuyeron favorablemente en un tiempo de 78%, obteniendo un decremento del 27.39 minutos.

Acuña (2019) este estudio tiene una línea de exploración, Despliegue de las tecnologías de información y comunicación (TIC) para el progreso incesante de la calidad en asociaciones peruanas. Acuña (2019) el objetivo es ejecutar un sistema de información gerencial para una dinámica suficiente en la organización JM Comercial y Servicios Generales. La extensión prevista es ayudar a la organización en esa capacidad y en particular a la administración. Acuña (2019) la construcción de los datos fue en una muestra de 30 trabajadores, el instrumento de recolección fue el cuestionario para calcular el nivel de satisfacciones estudio se realizado fue en un enfoque cuantitativa de tipos descriptiva. Acuña (2019) según los datos obtenidos en el primer tamaño se obtuvo que el 86.67% de sondeo no están de acuerdo con la toma de decisión actual de la organización y en el segundo tamaño se obtuvo que el 100% del sondeo están de acuerdo con el despliegue de un SIG.

Padilla (2018) esta investigación se realizó en Cotopaxi – Ecuador, Requieren implementar un sistema de información gerencial para medir indicadores en sus transformaciones internas, para poder medir la eficiencia y las estrategias de la organización. Padilla (2018) la finalidad general es desplegar un sistema de información gerencial, cumpliendo los requisitos obtenidos de la investigación. Se tuvo que implementar una investigación de tipo descriptiva. Padilla (2018) concluyó que los resultados se muestran, las Cajas Solidarias han funcionado durante los últimos 7 años con cálculos de 556 asociados aproximadamente. Aplican el 0,50% en depósitos y en préstamos varía entre 1,0% y 1,5%. Una desigualdad de 1,0% a 1,5%, cumplen su misión solidaria.

Chamba (2017) el presente trabajo de investigación se realizó en Machala – Ecuador. En la actualidad toda la información de la Municipalidad de Santa Rosa se genera de forma manual, los reclamos o solicitudes que realizan los ciudadanos o entidades. Chamba (2017) concluyó que el despliegue de un sistema de informático administrativo cumplirá con los objetivos planteados que permitirá la solución de los problemas que presenta el gobierno de Santa Rosa.

Herrera (2016) el objetivo fundamental que se tiene es la problemática de la contratista, de la cual su principal servicio es de suministrar personal a sus clientes. Herrera (2016) posteriormente, con el despliegue de un sistema de información gerencial, ayudará en buscar herramientas administrativas para poder obtener una mejor gestión de la problemática expuesta. Herrera (2016) la consecución de las propuestas realizadas permite lograr la mejora en los tiempos de respuesta de sus trabajadores en relación con los clientes, consiguiendo la efectividad en la inversión de los presupuestos relacionados con cada sector de la organización, logrando la optimización de recursos y así consiguiendo la disminución del mal uso y derroche de recursos.

Ruiz (2016) esta investigación se realizó en Guayaquil- Ecuador, el estudio propuesto pretende en un sistema de información gerencial que permite alcances de una nueva solución a la informalidad de liquidación de fruta semanal, de tal forma ayude a la gerencia general. Solución integral organizacional, económica y tecnológica. Ruiz (2016) para el caso de estudio se propone el uso de metodologías estructuradas, que permita hacer estudios y modelar los procesos para la ejecución de un sistema de información. Ruiz (2016) concluyó que la compañía Bandecua S.A. tiene la capacidad de ejecución de un sistema de información gerencial que le permitirá manejar la información para un correcto análisis del desempeño de la empresa y así optimizar la transformación de toma de decisión al momento de analizar futuras compras de fruta.

Angarita (2015) esta investigación se realizó en Ocaña- Colombia, este proyecto se centra en una aplicación para la comprobación de registro

asistencial del colegio la sallé, el principal objetivo del software se basa en registrar las asistencias e inasistencias de los estudiantes y docentes. Angarita (2015) el docente al ingresar a clase debe registrar los datos y revisar el listado de los alumnos donde ingresaron o no a su respectiva clase. Al finalizar semestre académico el gestor del sistema podrá mirar los reportes de los estudiantes que inasistencias durante todo el ciclo.

Seguidamente, se describe las teorías relacionadas sobre los conceptos fundamentales que dieron las principales definiciones a la investigación. Los Sistemas de Información Gerencial son un apoyo fundamental para la organización ya que ayudan en los procesos de información para poder analizar y tomar buenas decisiones.

Sobre los SIG Laudón (2012 pág. 7), el constante cambio en la tecnología, uso administrativo e impacto en los logros empresariales son los temas más emocionantes en los SIG. La aparición de nuevos negocios e industrias, así como su desaparición y son las empresas que tienen éxito las que logran usar nuevas tecnologías. Los más importantes nuevos temas son sintetizados brindando importancia al uso comercial de los sistemas de información.

Según Laudón (2012 pág. 13), nos define que:

Los sistemas tendrán la capacidad de realizar., incrementar la colaboración en el mercado, volverse un productor de excelente calidad o costo muy accesible, fomentar nuevos artículos y aumentar el desempeño de los empleados son procesos que cada vez más necesitan de los tipos y la calidad de los sistemas de información en la empresa, como se logra ver en la Figura 1

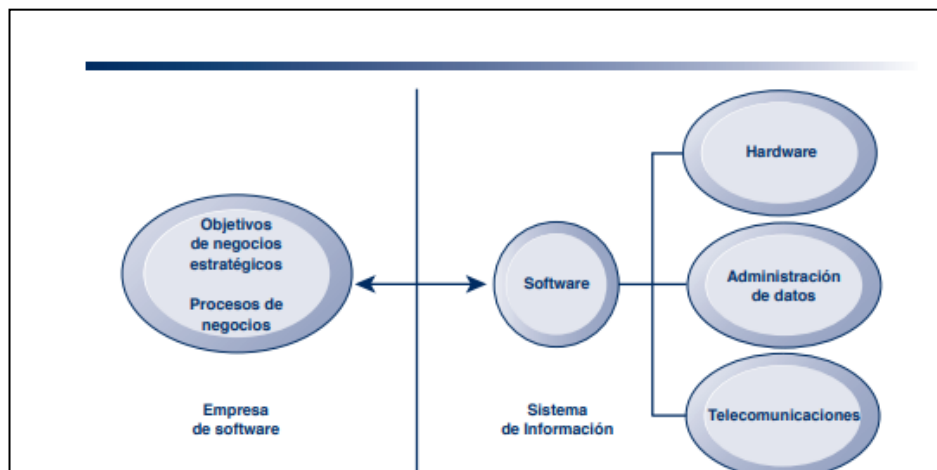


Figura 1 Los sistemas de información en la organización (Laudón y Laudón 2012 p. 13)

También Cohen (2000 pág. 1) indica que un sistema de información consta de un conjunto de partes que se interrelacionan con la finalidad de dar apoyo a las funciones dentro un negocio o empresa

Según Laudón (2012 pág. 18), plantean el SIG:

Una agrupación de piezas interrelacionadas que ensamblan (o recuperan), interactúan, almacenan y transmiten datos para ayudar a las medidas dinámicas y de control de la asociación. Brindar apoyo a la dinámica, manejo y la coordinación, los sistemas de información pueden igualmente apoyar a los administradores y a los usuarios de la información a diseccionar los problemas, a visualizar cuestiones complejas y a fomentar nuevos elementos.

También Fernandez (2006 pág. 11) define sistema de información:

Un SI, es un conjunto de elementos que interactúan entre sí para poder lograr un objetivo común. A pesar de que hay una extraordinaria variedad de SI, la mayor parte de ellos puede abordarse a través de un modelo formado por cinco casillas esenciales: componentes de entrada, componentes de rendimiento, segmento de cambio, instrumentos de control y destinos, los activos acceden al sistema de información a través de sus componentes para ser alterados en el segmento de cambio. Este ciclo está limitado por el componente de control para lograr el objetivo fijado.

Según Laudón (2012 pág. 18), aporta que:

Para comprender completamente los marcos de datos, debe conocer las medidas de autoridad más extensas, los ejecutivos, y la innovación de datos de los marcos, junto con su capacidad para brindar respuestas a los obstáculos y problemas en el clima de negocios, como se puede visualizar en la Figura 2.



Figura 2 Sistemas de información más efectiva (Laudón y Laudón 2012 p. 18)

Según Domínguez (2012 pág. 39) exponen que:

Un marco de datos de administración es un conjunto de marcos de datos que se comunican entre sí y, por tanto, proporcionan datos a los ejecutivos sobre las necesidades de las tareas.

El SIG se conoce como un conjunto de datos amplio y facilitado de subsistemas relacionados que convierten los datos en información en un surtido de enfoques para desarrollar aún más la eficiencia con el estilo y los atributos del administrador.

De la misma manera Laudón (2012 pág. 15) nos dice que los sistemas de información:

Son capaces de almacenar diversos tipos de información, como podrían ser de personas, localizaciones y objetos relevantes para la empresa. De

igual manera, nos hace la aclaración que, al hablar de información, hace referencia a datos relevantes y de gran utilidad para las personas que lo manejarán y estarán a cargo de su ordenamiento y apreciación, debiendo ser personas capaces para su comprensión y uso.

Según O'Brien (2006 pág. 6) los sistemas de información gerencial (SIG):

El sistema de información nos muestra cada una de las partes y activos importantes para dar su información y sus funciones a la organización. Por el contrario, el término tecnología de información alude a los componentes de hardware mixtos indispensables para que el sistema pueda operar satisfactoriamente. En principio, un sistema de información podría utilizar elementos básicos de hardware como un lápiz y papel o carpetas de archivo para obtener y almacenar sus datos.

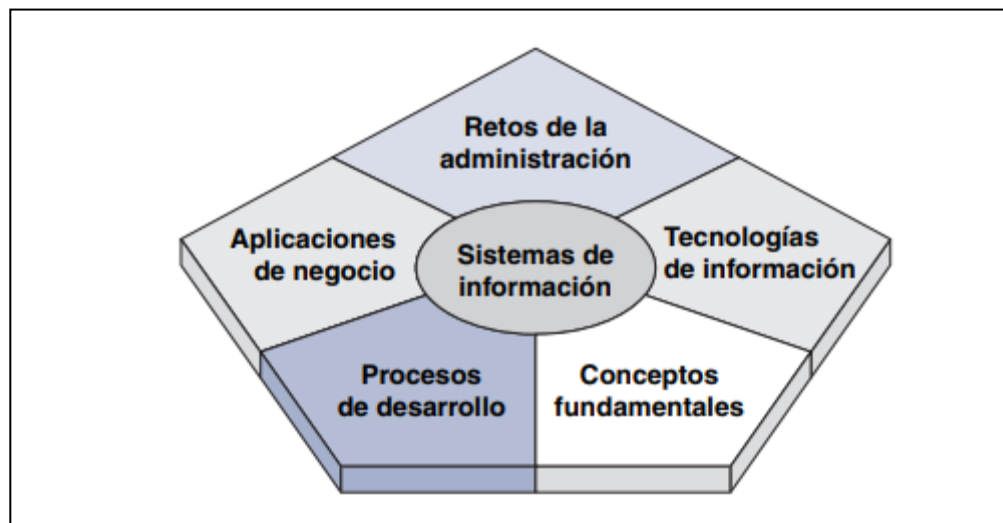


Figura 3 Áreas del conocimiento en el sistema de información (O'Brien, Marakas 2006 p. 7)

Por otro lado, Bocchino (1995 pág. 134) considera que Los marcos de datos son un conjunto de partes y ciclos que pueden manejar eficientemente la información para cumplir con las necesidades legítimas y funcionales, dar datos para planificar, controlar y producir informes en línea con las reuniones externas.

Los procesos de datos son un factor importante para manejar volúmenes de información complejas para poder explotar y tener buen cálculo de una solución y no se requieran estrategias complejas para generar alternativas

O'Brien (2006 pág. 30), procesamiento de datos:

La preparación de la información está sujeta en su mayor parte a ejercicios de manipulación, como el cálculo, el examen, la organización, la caracterización y la sinopsis. Estos ejercicios reúnen, examinan y manejan la información hacia los datos para los clientes finales. La naturaleza de cualquier información guardada en un marco de datos también puede mantenerse a través de un curso consistente de ejercicios de revisión y actualización.

Según Perissé (2001 pág. 5), nos indica que:

Son materiales para cuestiones organizadas en las que los objetivos están fundamentados, se puede indicar la información vital, existen técnicas estándar para registrar una respuesta y no se necesitan sistemas complejos para producir y evaluar las opciones. Las cuestiones completamente organizadas son procesables y es factible elegir si se defiende continuar con el cálculo, teniendo en cuenta la medida de tiempo y los activos necesarios.

La transformación en unos procesos de inicio, transformación y fin, para obtener unos datos procesados y claros, que contenga toda la información a trabajar y poder amoldar a las necesidades de la organización.

O'Brien (2006 pág. 144), un almacén de datos:

Almacena la información que se ha separado de los diferentes conjuntos de datos funcionales, externos y diferentes de una asociación. Se trata de una fuente de información purificada, modificada y clasificada para que los directores y otros expertos empresariales puedan utilizarla para la extracción de información, la gestión de información en línea y diferentes tipos de investigación empresarial, el estudio estadístico y la ayuda a la elección, tal y como se muestra en la figura 4.

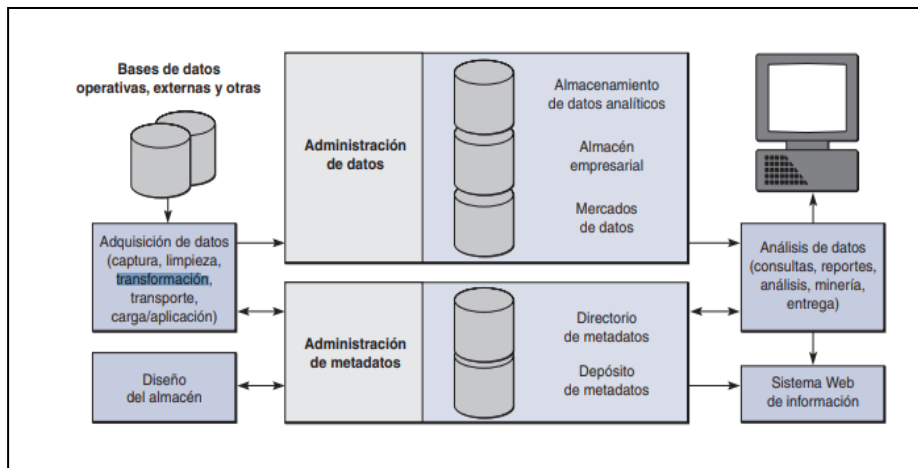


Figura 4 Almacén de datos. Los componentes de un sistema completo de almacén de datos. (O'Brien, Marakas 2006 p. 145)

Según Laudón (2012 pág. 254), un almacén de datos:

Es un conjunto de datos que almacena datos actuales y registrados de posible interés para los jefes de la organización. La información comienza a partir de muchos marcos de intercambio funcional del centro, como los marcos de ofertas, registros de clientes, fabricación, y puede incorporar información condicional de los locales de la Web. El centro de distribución de la información solidifica y normaliza los datos de varios conjuntos de datos funcionales, por lo que tiene a ser utilizado en todo el esfuerzo para el examen y la dinámica de los ejecutivos, como se aprecia en la Figura 5.

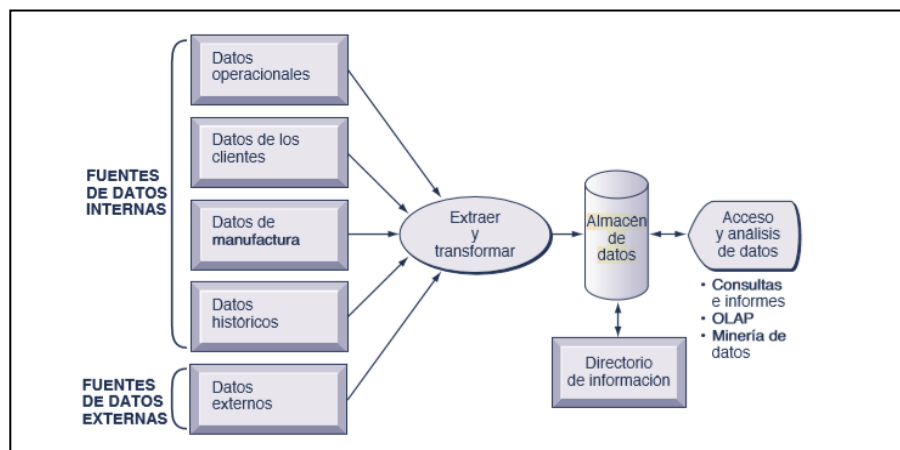


Figura 5 Almacén de datos y sistemas operacional. (Laudón y Laudón 2012 p. 223)

También O'Brien (2006 pág. 147), el almacén de datos en la empresa:

La información puede ser trasladada a mercados de datos o a un almacén analítico de datos, según los tipos de análisis para que puedan ser contenidos de la manera más eficiente. Los metadatos (datos acerca de los datos) se guardan en un archivo de metadatos y se registran mediante un índice de metadatos. Para finalizar, puede proporcionar una gran cantidad de herramientas analíticas de software para consultar, reportar, realizar minería y analizar los datos para que sean distribuidos a través de Internet y los sistemas de intranet en Web a los clientes finales de negocio, como podemos ver en la Figura 6.

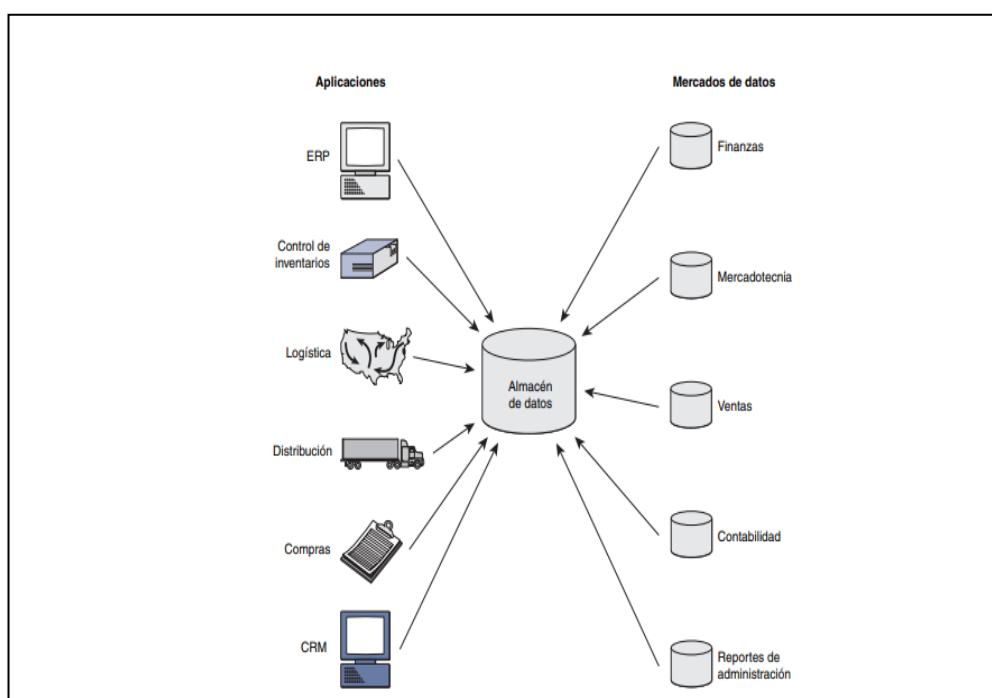


Figura 6 Almacén de datos y sus subconjuntos. Un almacén de datos y sus subconjuntos de mercados (O'Brien, Marakas 2006 p. 146)

Los Análisis de datos como una herramienta principal para la inteligencia de negocios. La minería de datos disecciona considerables porciones de datos, agregando lo incluido en los almacenes de datos, para descubrir patrones y reglas que puedan utilizarse para prever la conducta futura y orientar de forma correcta la toma de decisiones.

Laudón (2012 pág. 504), análisis de datos multidimensionales y OLAP:

Como una de las innovaciones esenciales de la visión empresarial. Las páginas de contabilidad cuentan con un elemento comparativo para la

investigación de diversas dimensiones conocido como tabla de giro, que los jefes e investigadores “superusuarios” utilizan para reconocer y comprender los diseños en los datos empresariales que pueden ser valiosos para la dinámica semiorganizada, como se encuentra en la Figura 7.

Costos fijos totales	19000					
Costo variable por unidad	3					
Precio promedio de venta	17					
Margen de contribución	14					
Punto muerto	1357					
		Costo variable por unidad				
Precio de venta	1357	2	3	4	5	6
	14	1583	1727	1900	2111	2375
	15	1462	1583	1727	1900	2111
	16	1357	1462	1583	1727	1900
	17	1267	1357	1462	1583	1727
	18	1188	1267	1357	1462	1583

Figura 7 Análisis de datos. (Laudón y Laudón 2012 p. 472)

También Cabero (2013 pág. 134), el concepto de análisis de datos:

La evaluación de un cuestionario antes de su aplicación, deber realizado por expertos en el rubro, dándonos como juicio u opinión las ventajas que nos pueda brindar su aplicación, de igual manera la probabilidad de conseguir una estandarización e información detallada del objeto de estudio.

La información es un conjunto de datos convertidos de manera que ayudan a disminuir la duda del futuro y, de esta manera, apoya a la toma de decisiones.

Laudón (2012 pág. 365) define información:

A los datos modelados significativamente que sean útiles para las personas, como vemos en la Figura 8.

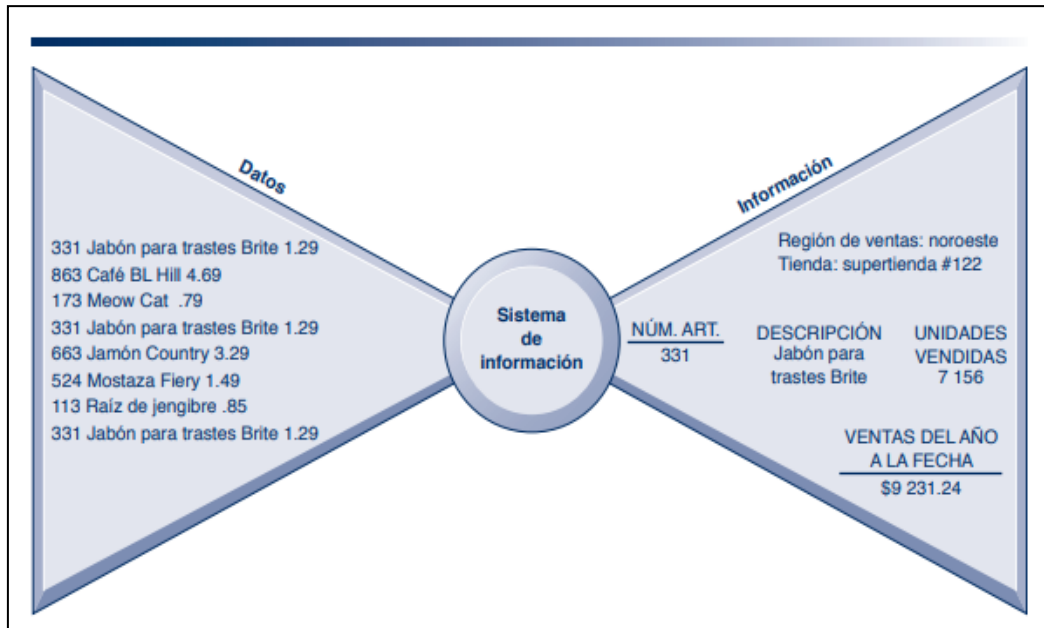


Figura 8 Modelado de la información. (Laudón y Laudón 2012 p. 16)

Según Chiavenato (2009 pág. 350), define información:

Los datos son un montón de información significativa en un entorno determinado que ayuda a disminuir el borde de la vulnerabilidad en cuanto a la dinámica. Los datos serán información significativa para alguien en un entorno determinado, y en los negocios la marca más aplicable es su calidad.

Urso (2014 pág. 118), caracteriza que estos ciclos incorporan ver cómo va el emprendimiento, cómo se está haciendo la ejecución y en un nivel muy básico, cómo se evalúa que la tarea se termine, el control debe tener una metodología indispensable, es decir, debe verificar el avance del emprendimiento, tanto lo que se ha hecho como lo que se evalúa que se termine. Para ello debemos pensar en varias perspectivas, a las que aludiremos aprobación y control del avance de la obra, del plan, de gastos, de riesgos, de calidad, de correspondencia, de adquisición, de intereses de los socios y control de cambios.

Laudón (2012 pág. 455) define tipo de decisiones:

“Cada nivel jerárquico contiene variados requerimientos de información para el soporte de decisiones y responsabilidad para los diferentes tipos de decisiones”, como se aprecia en la Figura 9.

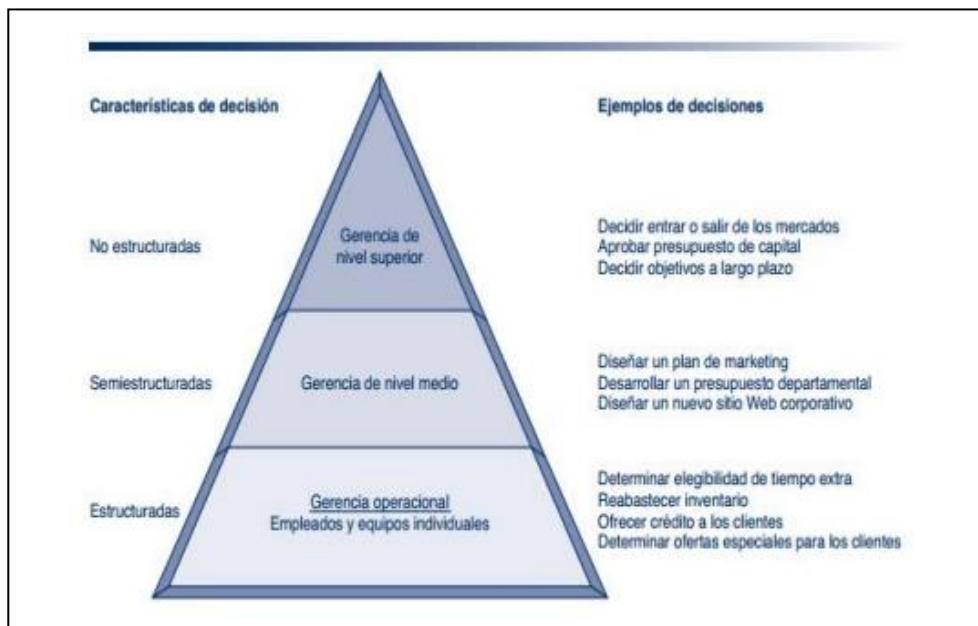


Figura 9 Requerimientos de información por jerarquía. (Laudón y Laudón 2012 p. 456)

El riesgo es el medio por donde se mide una toma de decisión correctamente factible dentro de una organización. Cuando se incrementa el nivel de riesgo baja significativamente el nivel de rendimiento esperado.

O'Brien (2006 pág. 51), define riesgo:

La introducción de mejoras revolucionarias en los ciclos de negocio para incrementar la competencia y la adecuación no es una tarea sencilla. En diversas organizaciones han utilizado programas de organización de activos empresariales (ERP) para rediseñar, mecanizar y coordinar sus medidas empresariales de ensamblaje, dispersión, dinero y recursos humanos.

Señala Knight (2012 pág. 2) nos menciona:

Que todas las organizaciones, desde las más grandes a hasta las más pequeñas, necesitan gestionar elementos exógenos y endógenos que incrementan la vulnerabilidad para lograr sus objetivos. Este impacto de

ausencia de seguridad es lo que se denomina riesgo y esencial para todas las actividades.

También Hopkin (2017) así mismo:

Dividen el riesgo en 3 categorías:

Riesgos peligrosos (o puros): Considerados como riesgos operacionales o asegurable debido a que la empresa puede estar protegida ante un suceso inesperado. Siempre conllevan a malos resultados para la empresa.

Riesgos de control (o incertidumbre): es usado comúnmente para comprobar que los resultados obtenidos de las actividades de la empresa se encuentren en un rango adecuado. Son relacionados con la gestión de proyectos y dan espacio a la incertidumbre referente al resultado más lógico frente a una situación.

Riesgos de oportunidad (o especulativos): son muy llamativos para las empresas debido a que el retorno de inversión es altamente rentable, mientras más riesgo se tome se obtendrán más beneficios. Son asumidos por la empresa para alcanzar una alta rentabilidad y rendimientos satisfactorios, específicamente en los riesgos comerciales o de mercado.

Un proceso es toda una lógica o un ciclo de vida que está compuesto por un conjunto de pasos u órdenes que deben de estar en ejecución.

O'Brien (2006 pág. 51), define proceso:

Los datos obsoletos, erróneos o difíciles de comprender no serán extremadamente significativos, útiles o importantes para usted o para otros expertos empresariales. Las personas necesitan datos excelentes, es decir, elementos de datos cuyos atributos, características o particularidades hacen que los datos sean más importantes para ellos. Es valioso considerar que los datos tienen tres medidas: tiempo, contenido y estructura, como se ve en la Figura 10.

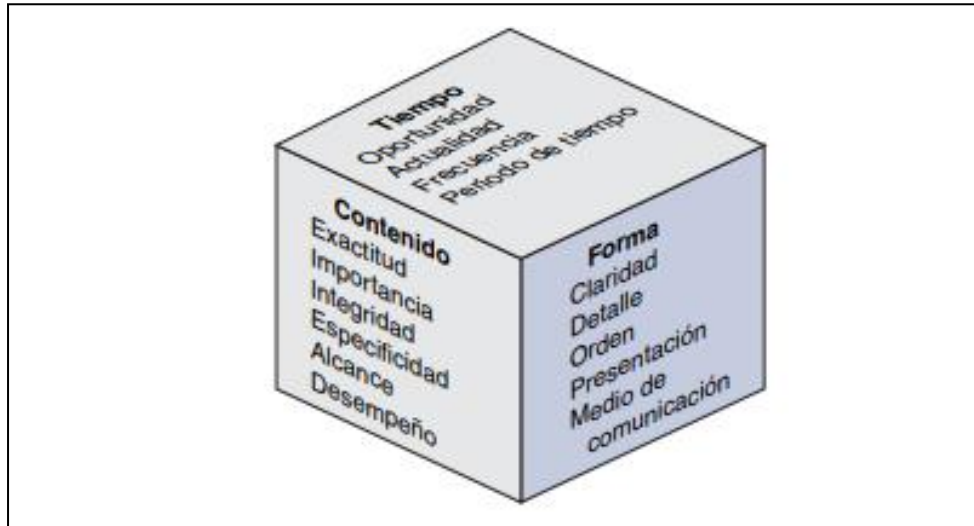


Figura 10 Resumen de los atributos de la calidad de información. (O'Brien, Marakas 2006 p. 323)

Según (Ramos, 2014) el autor nos dice “Es una red personal de bajo alcance, que nos permite relacionar los smartphones y ordenadores a través de un aparato central”.

Ohlhorst (2012 pág. 2), el término BIG DATA es definido:

Como una circunstancia en la que los índices informativos se han desarrollado hasta grandes dimensiones que las innovaciones de datos tradicionales. Es más importante tratar el volumen del conjunto de datos o la medida y el desarrollo del conjunto de datos. A fin de cuentas, el conjunto de datos se ha desarrollado tanto que es difícil de controlar y, sorprendentemente, más difícil de obtener estima de él. Las principales dificultades son la obtención de información, la capacidad, la búsqueda, el cambio, el examen y la percepción, como se visualiza en la figura 11.

DIMENSIÓN DE CALIDAD	DESCRIPCIÓN
Precisión	¿Los datos representan la realidad?
Integridad	¿Son consistentes la estructura de los datos y las relaciones entre las entidades y atributos?
Consistencia	¿Están definidos los elementos de datos de una manera consistente?
Compleción	¿Están presentes todos los datos necesarios?
Validez	¿Los valores de datos se encuentran dentro de rangos definidos?
Puntualidad	¿Están disponibles los datos cuando se necesitan?
Accesibilidad	¿Es posible acceder a los datos, comprenderlos y utilizarlos?

Figura 11 Dimensiones de la calidad de información. (Ohlborst, Frank 2012)

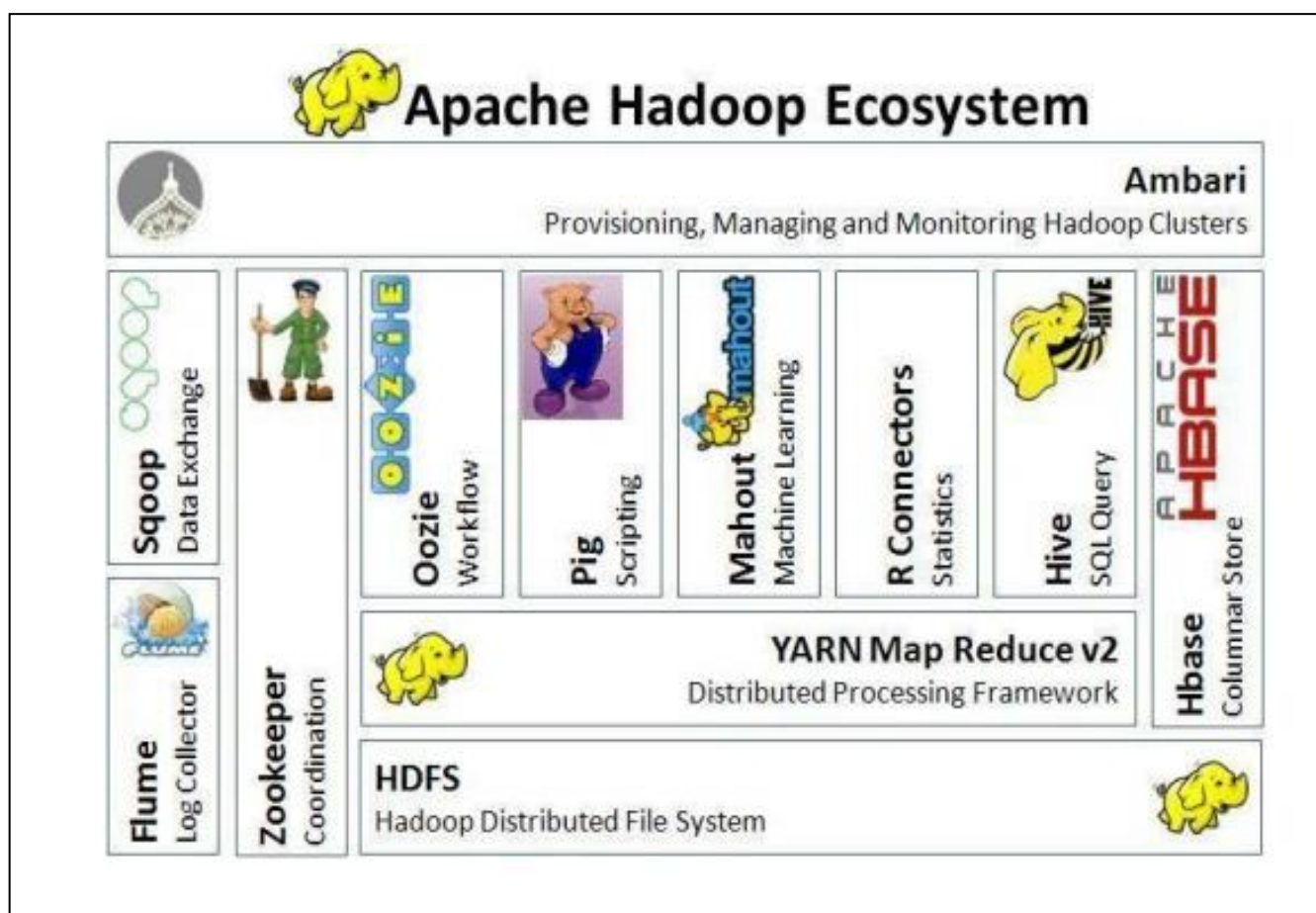


Figura 12 Esquema Hadoop Fuente: Universidad Nacional de Entre Ríos (2016), Sitio Web: http://sedici.unlp.edu.ar/bitstream/handle/10915/52766/Documento_completo.pdf?sequence=1&isAllowed=y

Ruiz (2016 pág. 4), define Hadoop:

Plataforma realizada para la preparación apropiada de información de enorme volumen transportada en correspondencia con la asistencia de partes de MapReduce y YARN. Fue creado para encontrar y supervisar las decepciones en la capa de uso.

Hadoop tiene 2 partes centrales, un servidor de archivos (HDFS) así como el algoritmo utilizado para particionar los datos (MapReduce), como se aprecia en la Figura 12.

III. METODOLOGÍA

3.1 Tipo y diseño de investigación

El tipo de investigación del presente estudio reúne las condiciones metodológicas de una investigación aplicada, porque nuestra investigación tiene como objetivo implementar un Tablero de Mando basado en Linux para la planta HFC en el área de operaciones del GRUPO IMATRA SAC. Según Arias (2012 pág. 22) el desarrollo de nuevas investigaciones ayuda en el proceso de incrementar los nuevos fundamentos teóricos, teniendo principios de una aplicación clara ante nuevas soluciones o problemas específicos.

La presente de investigación será Explicativa, presenta dicho nivel de investigación, ya que busca determinar cómo influye un Tablero de Mando basado en Linux para la planta HFC en el área de operaciones de GRUPO IMATRA SAC. Para Hernández (2003 pág. 95) las investigaciones que son de tipo explicativa intentan evaluar los datos transformados de las causas de los eventos y fenómenos físicos o sociales. Como su mismo nombre lo dice, su principal factor es buscar y explicar el por qué suceden dichos fenómenos y en qué aspectos se representan en una o más variables.

Esta investigación es de diseño no experimental de tipo transversal, ya que la recolección de los datos se hizo en un solo momento en la unidad de análisis.

Acorde a lo expuesto por Hernández (2003 pág. 149) los fenómenos que suscriben en el ambiente de un análisis de estudio que se despliega en un diseño no experimental que sucede sin la manipulación de las variables.

Se observan casuísticas que ya existen y no es posible manipular las variables, no se puede influenciar sobre dichas variables porque ya se dieron y se expusieron de esa forma, en un estudio no experimental.

3.2 Variables y operacionalización

Para los SIG, Según nos indica Mcleod (2006 pág. 13) define que un marco de datos de administración (MIS) como un marco basado en PC que proporciona datos a los clientes que tienen necesidades comparables. Los clientes suelen establecer un elemento formal de autoridad: una organización o una unidad

auxiliar. Los datos describen la organización o uno de sus marcos significativos en cuanto a lo que ha ocurrido antes, lo que está ocurriendo ahora y lo que probablemente va a ocurrir más adelante. En su mayor parte, se presentan en forma de informes intermitentes, informes excepcionales y rendimientos de recreación numérica. Los datos entregados son utilizados por los directores y no jefes de la misma para establecer las opciones que abordan los problemas de la organización.

Para complementar las definiciones de Toma de decisiones, Según Laudón (2012 pág. 14) nos indica que los focos dinámicos en la decisión se encuentran dentro de las opciones con respecto a una cuestión o caso a tratar. En el clima de negocios, la dinámica tendrá un nivel más prominente de responsabilidad, utilizando la capacidad de apoyar los datos y que esto impacta una dinámica decente, llenando como una técnica de negocio.

3.3 Población, muestra y muestreo

La población, en este presente proyecto se desarrolla considerando como población de 35 trabajadores en la empresa GRUPO IMATRA S.A.C

Un grupo de objetos, documentos o personas a ser estudiadas forman parte fundamental dentro de las Investigaciones comúnmente. A continuación, Ñaupas (2013 pág. 205) indica que, La población es un conjunto de personas o establecimientos que son objeto de examen.

La Muestra será de 30 personas, cuando se tiene identificada la población que se someterá a estudio y cuando esta, debido a su volumen no es posible tener en cuenta su totalidad en la aplicación de instrumentos de investigación; se genera la necesidad de establecer una muestra. Según Hernández (2014 pág. 174) nos indica que, se piensa en una parte de la población, que mediará para la obtención de los datos que necesitaremos.

Muestreo Probabilístico aleatorio simple, Partiendo de esta premisa, podemos decir que es la parte que nos podrá ayudar en la recolección de información que necesitamos para la obtención de la medición y observación de las variables

estudiadas, es por ello que el tipo de muestreo se determinó que será de 30 trabajadores. Según Hernández (2014 pág. 4) indica que el muestreo es un caso de estudio. Es una unidad de análisis de la muestra.

3.4 Técnicas e instrumentos de recolección de datos

Las técnicas e instrumentos que serán empleados en la investigación son de suma importancia y de mucho valor.

Según Hernández (2014 pág. 12) la recopilación de información incluye el fomento de una disposición definitiva de los sistemas que nos llevará a recopilar información por un motivo determinado en instrumentos normalizados.

Es uniforme para todos los casos. Los datos son obtenidos mediante observación, medición y documentación. Se utilizan instrumentos que han demostrado ser válidos y confiables en estudios previos o son generados nuevos basándose en la revisión de la literatura y se prueban y ajustan. Las preguntas, ítems o indicadores utilizados son específicos con posibilidades de respuesta o categorías predeterminadas”.

En este proyecto de investigación, la técnica utilizada es la encuesta, este método de investigación permite recabar datos de grupos de personas que se encuentran involucradas con el tema de estudio y nos permiten acceder a la información desde la fuente. En este sentido se aplicó una encuesta con 26 preguntas al personal de operaciones.

Según Arias (2012 pág. 111) los métodos de recopilación de información son las distintas estructuras o métodos de obtención de datos. Ejemplos de estrategias son: la percepción inmediata, la visión general en sus dos modalidades: oral o compuesta (sondeo), la charla, la investigación narrativa, el examen de contenidos, etc.

El presente proyecto de investigación está utilizando como instrumentos cuestionario

5 = Totalmente desacuerdo

4 = En Desacuerdo

3 = Ni de acuerdo ni en desacuerdo

2 = De Acuerdo

1 = Totalmente acuerdo

- Ficha de Observación

Conforme a lo expuesto por Hernández (2014 pág. 289) esta técnica de recopilación de información consiste en el registro metódico, sustancial y fiable de prácticas y circunstancias discernibles, a través de un conjunto de clases y subcategorías. Este procedimiento incluye la captación de la imagen o de la vista, de forma eficiente, de cualquier realidad, maravilla o circunstancia que se produzca en la naturaleza o en el ámbito público, según los objetivos de investigación preestablecidos.

- Encuesta

Conforme a lo expuesto por Arias (2012 pág. 33) la revisión oral o compuesta cuya intención es adquirir datos. Evidentemente, los datos adquiridos son legítimos sólo para el periodo en el que se han recogido, ya que los dos atributos y valoraciones pueden ser diferentes a largo plazo. La encuesta fue definida de diversas maneras: como un método, diseño o simplemente como una técnica.

3.5 Procedimientos

El GRUPO IMATRA, todo su análisis de información lo realiza a través de Excel o pequeños sistemas que realizan el procesamiento de la información. Teniendo en cuenta que al realizar todo este proceso toma mucho tiempo en obtener los resultados. Se requiere optimizar y obtener los reportes en un tiempo que sea óptimo y con datos claros y concisos.

Luego de realizar una investigación interna dentro de la organización, se llegó a la solución correcta y efectiva. De acuerdo con las necesidades de GRUPO IMATRA, se recomienda en la implementación del Tablero de mando para el

apoyo en la toma de decisiones, para obtener mejores resultados y sea más eficiente en las decisiones a optar.

3.6 Método de análisis de datos

En este caso los datos se analizaron, la información integrada será listada a partir de los datos obtenidos, cuando se trata del procesamiento de información, nos referimos a los métodos estadísticos utilizados y los procedimientos específicos a complementar para procesar los datos recopilados, en nuestro caso usaremos SPSS.

En este sentido, SPSS ha realizado aportes al campo de la metodología de la investigación científica cuantitativa y al desarrollo de toda la investigación, y ha jugado un papel importante en la academia y la sociedad civil. Además de las actividades habituales de investigación, docencia y producción de conocimiento.

SPSS le permite crear fácilmente archivos de datos de manera estructurada, consolidada, y también puede organizar bases de datos que se pueden analizar mediante diversas técnicas estadísticas. Aunque existen otros programas por ejemplo (Microsoft Excel) para organizar datos y crear archivos electrónicos. SPSS le permite capturar y analizar datos sin depender de otros programas.

Por otro lado, también es posible convertir una base de datos creada en Microsoft Excel a una base de datos SPSS

3.7 Aspectos éticos

La consideración de la información de esta tesis se refiere a la ética informática, cumpliendo con todos los aspectos establecido, se aplicó las citas de acuerdo con las normas ISO 690-2.

El software cumple con las normas de ética y seguridad de la información de la organización, para mantener todos los estándares profesionales, dando la confiabilidad de los involucrados del negocio, obteniendo las buenas prácticas de los instrumentos realizados en la encuesta.

El desarrollo del presente trabajo de investigación cuenta con la autorización del área de operaciones de GRUPO IMATRA, en el cual nos brinda la autorización del análisis respectivo de toda su información, teniendo en cuenta que sus datos se van a utilizar de forma íntegra y profesional.

IV. RESULTADOS

En este capítulo se elaboró una descripción al detalle de los resultados obtenidos en la investigación, basándonos en los indicadores de cantidad de incremento de datos, incremento del porcentaje de información, reducción de riesgo e incremento del tiempo de respuesta. Se logró determinar el efecto del uso del Tablero de mando basado en Linux para la mejora de la planta HFC. Posteriormente se realizó el procesamiento de cada uno de los indicadores presentados, se plantearon mediante encuesta. Los resultados obtenidos se realizaron a través del programa SPSS Statistics.

4.1. Normalización de la influencia de la variable

a) Ho: "La variable el efecto del uso del Tablero de mando basado en Linux en la mejora de la planta HFC se distribuye en forma normal"

H1: "La variable el efecto del uso del Tablero de mando basado en Linux en la mejora de la planta HFC no se distribuye en forma normal"

b) N. S= 0.05

Tabla 1 Tabla de variable

	Kolmogórov-Smirnov		
	Estadístico	gl	Sig.
Tablero de Mando	0,122	30	0,025
Toma de Decisiones	0,114	30	0,03

Fuente: Elaboración propia

c) Se observa en la columna sig. Kolmogorov-Smimov de todos son menores que 0.05, lo cual se rechaza la hipótesis Nula.

d) Concluimos que La variable el efecto del uso del Tablero de mando basado en Linux en la mejora de la planta HFC no se distribuye en forma normal. por tanto, aplicaremos la prueba estadística no paramétrica de escala ordinal de rho de Spearman.

a) El Planteo de las Hipótesis General

Ho: “La Implementación de un Tablero de Mando no mejorará significativamente la toma de decisiones de la Planta HFC en el área de operaciones del Grupo Imatra SAC.”

H₁: “La Implementación de un Tablero de Mando mejorará significativamente la toma de decisiones de la Planta HFC en el área de operaciones del Grupo Imatra SAC.”

a. N.S: 0.05

b. La Contrastación de la Hipótesis:

Pruebas estadísticas no paramétricas de escala Ordinal.

Utilizaremos la prueba de Rho de Spearman.

Tabla 2 Matriz correlacional de la hipótesis general

Matriz de Correlaciones

			Tablero de Mando	Toma de Decisiones
Rho de Spearman	Tablero de Mando	Coeficiente de correlación	1,000	0,896
		Sig. (bilateral)	.	0,014
		N	30	30
	Toma de Decisiones	Coeficiente de correlación	0,896	1,000
		Sig. (bilateral)	0,014	.
		N	30	30

Fuente: Elaboración propia en SPSS

Finalmente se observa que hay una marcada relación entre el tablero de mando y la toma de decisiones en un 89.6%

- c. Conclusión:
Se puede concluir que, La Implementación de un Tablero de Mando mejorará significativamente la toma de decisiones de la Planta HFC en el área de operaciones del Grupo Imatra SAC, a un nivel de significancia del 5% bilateral.

b) El Planteo de las Hipótesis Específica 1

Ho:” La Implementación de un Tablero de Mando no reducirá los riesgos en la toma de decisiones de la Planta HFC en el área de operaciones del Grupo Imatra SAC”

H1: “La Implementación de un Tablero de Mando reducirá los riesgos en la toma de decisiones de la Planta HFC en el área de operaciones del Grupo Imatra SAC.”

- a. N.S: 0.05
- b. La Contrastación de la Hipótesis:
- c. Pruebas estadísticas no paramétricas de escala Ordinal. Utilizaremos la prueba de Rho de Spearman.

Tabla 3 Matriz correlacional específica 1

Matriz de Correlaciones

			Riesgo	Tablero de Mando
Rho de Spearman	Riesgo	Coefficiente de correlación	1,000	0,923
		Sig. (bilateral)	.	0,017
		N	30	30
	Tablero de Mando	Coefficiente de correlación	0,923	1,000
		Sig. (bilateral)	0,017	.
		N	30	30

Fuente: Elaboración propia en SPSS

Finalmente se observa que hay una relación entre el riesgo y el Tablero de mando en un 92,30%.

d. La conclusión:

Se puede concluir, que La Implementación de un Tablero de Mando reducirá los riesgos en la toma de decisiones de la Planta HFC en el área de operaciones del Grupo Imatra SAC a un nivel de significancia del 5% bilateral.

c) El Planteo de las Hipótesis Específica 2

Ho: " El Tablero de Mando no mejorará los procesos en la toma de decisiones de la Planta HFC en el área de operaciones del Grupo Imatra SAC"

H1: "El Tablero de Mando mejorará los procesos en la toma de decisiones de la Planta HFC en el área de operaciones del Grupo Imatra SAC."

a. N.S: 0.05

b. La Contrastación de la Hipótesis:

c. Pruebas estadísticas no paramétricas de escala Ordinal. Utilizaremos la prueba de Rho de Spearman.

Tabla 4 Matriz correlacional específica 2

Matriz de Correlaciones

			Proceso	Tablero de Mando
Rho de Spearman	Proceso	Coeficiente de correlación	1,000	0,923
		Sig. (bilateral)	.	0,020
		N	30	30

	Tablero de Mando	Coefficiente de correlación	0,923	1,000
		Sig. (bilateral)	0,020	.
		N	30	30

Fuente: Elaboración propia en SPSS

Finalmente se observa que hay una marcada relación entre el proceso y el tablero de mando en un 92.3%

d. La conclusión:

Se puede concluir, que El Tablero de Mando mejorará los procesos en la toma de decisiones de la Planta HFC en el área de operaciones del Grupo Imatra SAC a un nivel de significancia del 5% bilateral.

d) El Planteo de las Hipótesis Específica 3

Ho: " El Tablero de Mando no influye en los resultados para la toma de decisiones de la Planta HFC en el área de operaciones del Grupo Imatra SAC."

H1: "El Tablero de Mando influye en los resultados para la toma de decisiones de la Planta HFC en el área de operaciones del Grupo Imatra SAC."

a. N.S: 0.05

b. La Contrastación de la Hipótesis:

c. Pruebas estadísticas no paramétricas de escala Ordinal. Utilizaremos la prueba de Rho de Spearman

Tabla 5 Matriz correlacional específica 3

Matriz de Correlaciones

			Resultados	Tablero de Mando
Rho de Spearman	Resultados	Coefficiente de correlación	1,000	0,905
		Sig. (bilateral)	.	0,024
		N	30	30
	Tablero de Mando	Coefficiente de correlación	0,905	1,000
		Sig. (bilateral)	0,024	.
		N	30	30

Fuente: Elaboración propia en SPSS

Finalmente se observa que hay una marcada relación entre los resultados y el tablero de mando en un 90.5%

d. La conclusión:

Se puede concluir, que El Tablero de Mando influye en los resultados para la toma de decisiones de la Planta HFC en el área de operaciones del Grupo Imatra SAC a un nivel de significancia del 5% bilateral.

4.2. La contrastación de hipótesis

La hipótesis general se contrastará mediante la prueba estadística no paramétrica de escala ordinal, por la prueba de rho de Spearman determinará que el efecto del tablero de mando para la planta HFC en el área de operaciones de grupo Imatra SAC.

Tabla 6 Cuadro comparativo de la variable

N° de Encuestado	Efecto del uso del Tablero de mando basado en Linux en la mejora de la planta HFC.																															
	DIMENSIÓN 1: Tiempo de procesamiento						DIMENSION 2: Porcentaje de información						DIMENSION 3: Riesgo				DIMENSION 4: Tiempo de Respuesta				Proceso						Resultados					
	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	p12	p13	p14	p15	p16	p17	p18	p19	p20	p21	p22	p23	p24	p25	p26	p27	p28	p29	p30	p31	p32
si	24	24	25	22	26	27	28	28	25	26	26	21	25	27	20	30	30	19	7	27	23	24	22	23	21	26	27	21	26	27	12	30
no	6	6	5	8	4	3	2	2	5	4	4	9	5	3	10	0	0	11	0	3	7	6	8	7	9	4	3	9	4	3	18	0
total	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	7	30	30	30	30	30	30	30	30	30	30	30	30	30

Fuente: Elaboración propia en SPSS

4.3. Aplicación de la estadística descriptiva de la variable

4.3.1. Variable: Efecto del uso del Tablero de mando basado en Linux en la mejora de la planta HFC.

Tabla 7 Mejora de los datos

		pregunta01			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	si	24	80,0	80,0	80,0
	no	6	20,0	20,0	100,0
	Total	30	100,0	100,0	

Fuente: Elaboración propia de autor

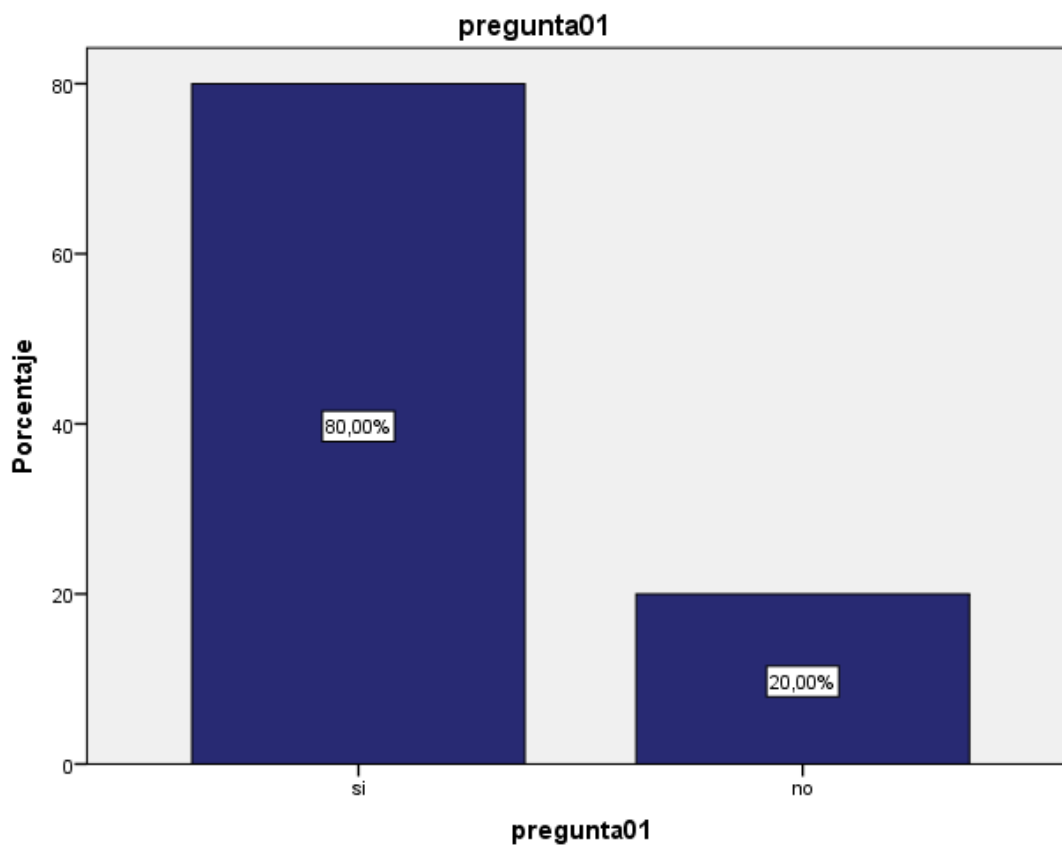


Gráfico 1: Mejora de los datos
Fuente: Elaboración propia

INTERPRETACIÓN:

De los 30 encuestados el 80% respondieron si sobre la pregunta: ¿El tablero de mando, ayuda en el cumplimiento y mejora de los datos? y el 20% respondieron no.

Tabla 8 Transformación de los datos

		pregunta02			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	si	24	80,0	80,0	80,0
	no	6	20,0	20,0	100,0
	Total	30	100,0	100,0	

Fuente: Elaboración propia

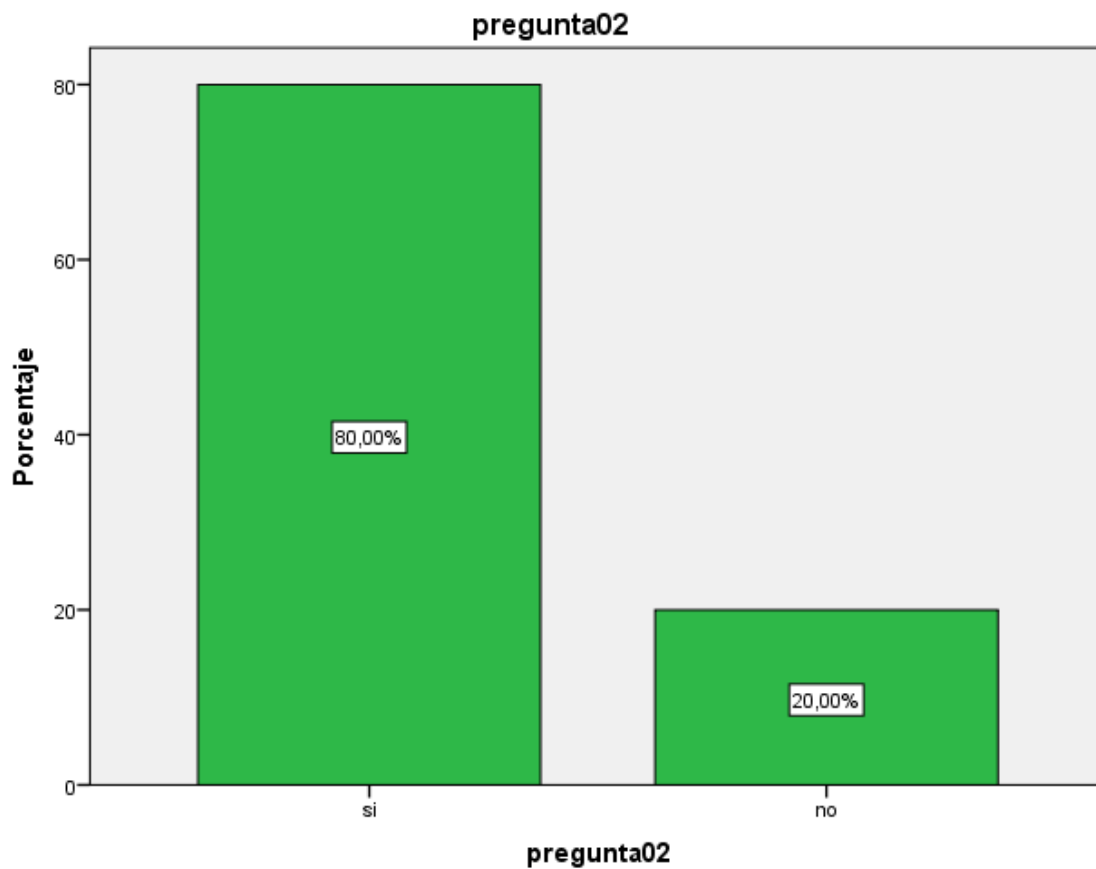


Gráfico 2: Transformación de los datos
Fuente: Elaboración propia

INTERPRETACIÓN:

De los 30 encuestados el 80% respondieron si a la pregunta: ¿Está de acuerdo que la transformación de la información mejora el procesamiento de los datos? y el 20% respondieron no.

Tabla 9 Almacenamiento de datos

		pregunta03			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	si	25	83,3	83,3	83,3
	no	5	16,7	16,7	100,0
	Total	30	100,0	100,0	

Fuente: Elaboración propia

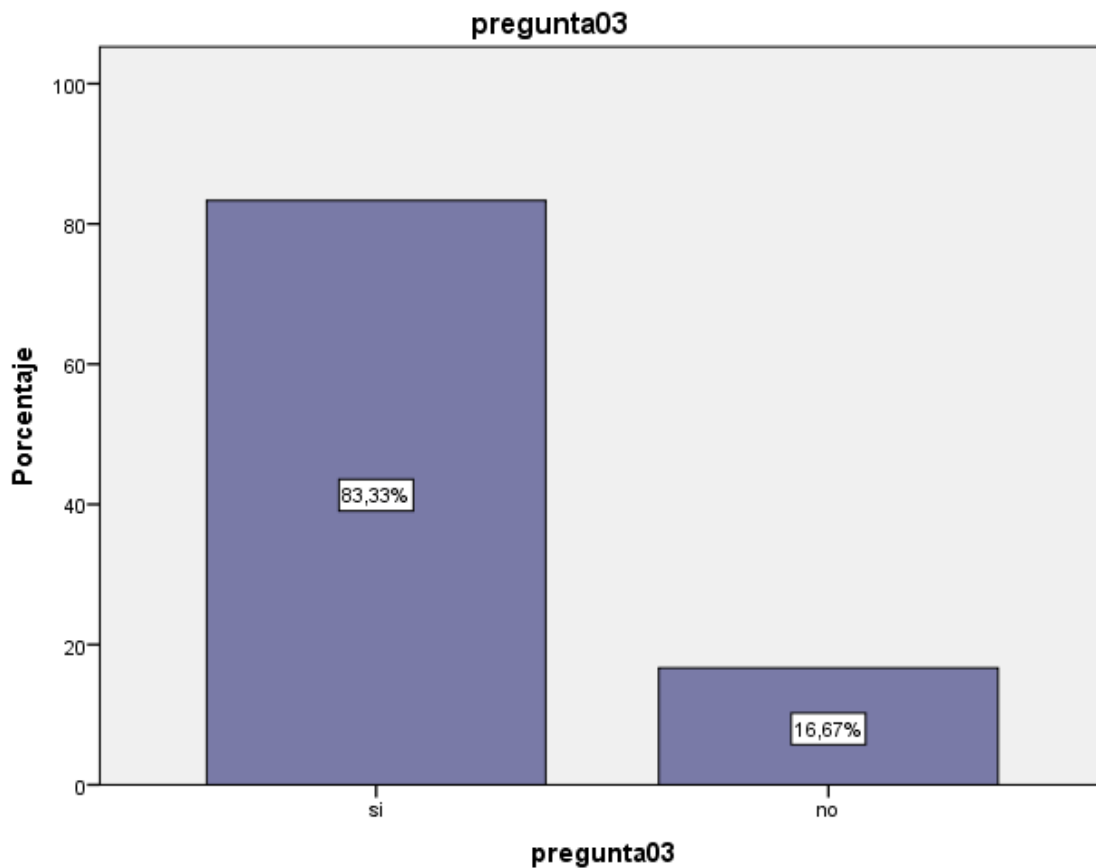


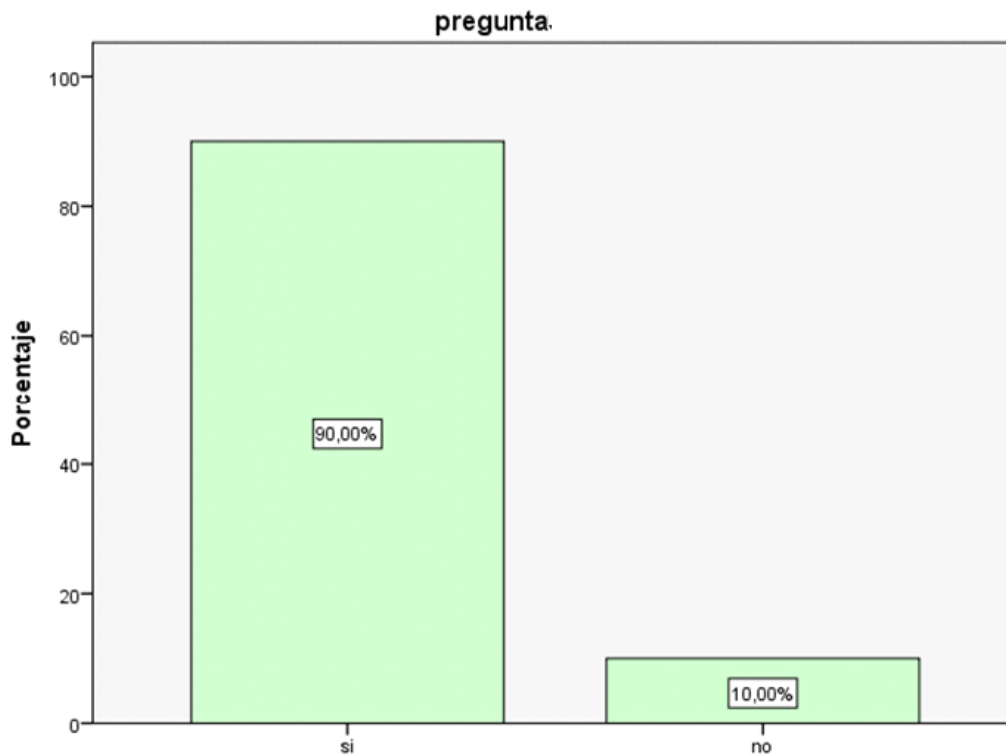
Gráfico 3: Almacenamiento de datos
Fuente: Elaboración propia de autor

INTERPRETACIÓN:

De los 30 encuestados se observa que el 83.33% respondieron si a la pregunta: ¿Considera usted que el almacenamiento de datos mejorará los resultados de los procesos? y el 16.67% respondieron no.

Tabla 10 Tiempo de Respuesta

		pregunta04			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	si	27	90,0	90,0	90,0
	no	3	10,0	10,0	100,0
	Total	30	100,0	100,0	



Fuente: Elaboración propia

Gráfico 4: Tablero de mando
Fuente: Elaboración propia de autor

INTERPRETACIÓN:

De los 30 encuestados se observa que el 90.00% respondieron si a la pregunta: ¿Cree usted que el tablero de mando ayuda en los tiempos de respuesta? y el 10.00% respondieron no.

4.4. Resumen

Tabla 11 Resumen de los resultados de las comprobaciones de las hipótesis

Cod.	Hipótesis	Resultado (Aceptada o Rechazada)
HE1	La Implementación de un Tablero de Mando reduce los riesgos en la toma de decisiones de la Planta HFC en el área de operaciones del Grupo Imatra SAC.	Aceptada
HE2	El Tablero de Mando mejora los procesos en la toma de decisiones de la Planta HFC en el área de operaciones del Grupo Imatra SAC.	Aceptada
HE3	El Tablero de Mando influye en los resultados para la toma de decisiones de la Planta HFC en el área de operaciones del Grupo Imatra SAC.	Aceptada
HG	La Implementación de un Tablero de Mando mejora significativamente la toma de decisiones de la Planta HFC en el área de operaciones del Grupo Imatra SAC.	Aceptada

Conforme se muestra en la tabla 13, basándonos en los resultados obtenidos se logró demostrar que las hipótesis alternas planteadas en un principio fueron aceptadas, pudiendo cumplir con el objetivo general y los objetivos específicos. Se generó un considerable incremento en la cantidad de almacenamiento de datos del 80%, incremento del porcentaje de información del 83% y un incremento positivo en lo que refiere al tiempo de respuesta del 90%.

V. DISCUSIÓN

La investigación realizada tiene como objetivo principal implementar un Tablero de mando basado en Linux para la Planta HFC del GRUPO IMATRA SAC, Santiago de Surco. Donde se concluyó que el tablero de mando es un apoyo fundamental para la organización en la mejora de la toma de decisiones en la planta HFC, haciendo que se obtenga mejores resultados con una alta calidad de reportes y buena toma de decisión de parte del área de operaciones.

Algunos estudios como Vargas (2018) en su investigación “SIG para mejorar el Control Integral de las Sesiones de Hemodiálisis de EsSalud” se realizó en La Libertad - Perú, los resultados obtenidos de esta investigación hicieron la medición en indicaron para hacer mejoras en el tiempo de las sesiones de hemodiálisis de los pacientes, logrando reducir el tiempo en un 78%, se obtuvo una disminución de 27.39 minutos.

Así mismo Tapia (2019) en su investigación “SIG para web y control de procesos de producción en la Granja Avícola REC S.A.C” se realizó en Ica – Perú, los resultados obtenidos, se alcanza optimizar el tiempo, esfuerzo y manejo correcto de la información, así como la mejora de la gestión de alimentos, disminución de la mortalidad de aves y la gestión en el control de calidad de la producción de huevos.

El tiempo de respuesta de los reportes disminuyó considerablemente con la implementación del Tablero de mando, el personal del área de operaciones confío en los reportes y calidad de los resultados obtenidos, con un nivel de confiabilidad del 30% al 90 % de efectividad.

Para la mejora de los procesos del Sistema de información se establecieron la Ingesta, Transformación y Delivery de la información, se logra la eficiencia de la entrega de los reportes con una productividad de 90% de los datos obtenidos de los archivos input, reduciendo los tiempos de espera en un 75%.

El resultado adquirido de esta investigación se comprueba que haber Implementado un Tablero de mando facilitó el conocimiento de los resultados para la toma de decisiones de la planta HFC en el área de operaciones, se puede extraer la información con las características establecidas por los usuarios, para la mejora continua de los procesos de Grupo Imatra.

VI. CONCLUSIONES

En relación con el presente estudio de un Tablero de mando basado en Linux para la Planta HFC de GRUPO IMATRA SAC, Santiago de Surco, podemos concluir lo siguiente.

Se presentan a continuación las conclusiones de la investigación:

1. El indicador de encuesta realizada al personal del área de operaciones, el 86.67% están de acuerdo que los análisis de datos ayudan en la inteligencia de negocios, por lo cual nos indica que es confiable las herramientas para la explotación de los datos.

Se concluye que la implementación del Tablero de Mando es un apoyo fundamental en la organización para optimizar tiempo de espera de los reportes diarios.

2. De los resultados obtenidos el 93.33% está de acuerdo que el Tablero de Mando brinda apoyo fundamental para la toma de decisión.

Se concluye que hay muy buena consistencia interna de los ítems del instrumento, por tanto, existe muy buena confiabilidad para el recojo de la información de la presente tesis, de la variable toma de decisiones al 92.4%.

3. De los resultados obtenidos el 80 % está de acuerdo con que el tablero de mando ayuda en el cumplimiento y mejora de los datos.

Se concluye que la implementación del tablero de mando brinda una mejora sustancial en lo referente al cumplimiento y mejora de procesamiento de los datos para empresa.

4. El indicador de encuesta realizado al personal del área de operaciones, el 63.3 % están de acuerdo en que una buena toma de decisión nos permite determinar el riesgo en el tiempo para obtener buenos resultados para la empresa.

Se concluye que la toma decisión de manera acertada y correcta en la empresa es muy importante para evitar la cantidad de riesgos en el tiempo y poder tener un adecuado procesamiento de datos y mejora en la empresa.

5. De los resultados obtenidos el 86.67% está de acuerdo en que los resultados de los procesos han mejorado dentro de la organización con un Tablero de mando.

Se concluye que la implementación del tablero de mando es fundamental en la mejora de los resultados de los procesos para la Empresa.

6. De los resultados obtenidos el 88.33% está de acuerdo en que la implementación de un Tablero de Mando ayuda a no tener pérdida de información.

Se concluye que el porcentaje de pérdida de información baja considerablemente debido a la implementación del Tablero de mando dentro de la Empresa.

VII. RECOMENDACIONES

Con los resultados obtenidos del presente estudio de investigación se consideran los objetivos a mejorar en los procesos internos de la organización, teniendo en cuenta las siguientes recomendaciones:

1. Se recomienda establecer un control interno del personal del área de operación del uso correcto de la aplicación y que la información rote al personal, para que cada uno esté capacitado para las ejecuciones y posibles errores de la información.
2. Se recomienda tener un monitoreo de las aplicaciones, que alimentan el Tablero de Mando, para poder así evitar la pérdida de información
3. Se debe realizar una evaluación constante de los planos fuentes que alimentan el Tablero de mando, para que así se pueda realizar mejoras de rendimiento y de productividad de la aplicación
4. Se debe tener en cuenta los datos huérfanos, para que así esa información no se pierda y se pueda guardar en una tabla y más adelante se pueda utilizar.
5. Se recomienda tener un mejor manejo de la información a través del uso del Tablero de Mando para así lograr mejoras sustanciales en los procesos de la empresa.

REFERENCIAS

- ACUÑA. Implementación de un Sistema de Información Gerencial en JM Comercial y Servicios Generales – Huarmey. Huarmey, Lima, Peru : s.n., 2019.
- AICTS. *El mundo en tus manos*. 2020. Disponible en:
http://www.aicts.info/calixam/index.php?option=com_content&view=article&id=7:sistemas-de-informacion-gerencial&catid=10&Itemid=142.
- ANGARITA. Desarrollo de un Sistema de Información Para el Control de Registro Asistencial del Colegio la Salle en Ocaña. 2015.
- ARIAS, F. *El Proyecto de Investigacion Introduccion a la metodologia cientifica*. 6a Edicion. Caracas : Suplidora Van, C.A, 2012. pág. 146. Disponible en:
<https://doi.org/Q180.55-M4A7>.
- BALLESTEROS, A.. movilonia. *movilonia*. [En línea] 25 de 1 de 2017. Disponible en:
https://www.movilonia.com/noticias/diferencias-banda-ancha-fija/#Fibra_optica_la_puerta_a_un_futuro_prometedor.
- BOCCHINO, W. *Sistemas de información para la administración: técnicas e instrumentos*. Segunda Edicion. s.l. : Trillas, 1995. pág. 403.
- BORJA, E. y FERRUZOLA, E. *Los Sistemas de información gerencial: aplicabilidad en procesos empresariales y de educación superior en el Ecuador*. s.l. : SATHIRI, Enero de 2015, LOS SISTEMAS DE INFORMACIÓN GERENCIAL:, pág. 140.
- CABERO, J y LLORENTE, M. *La aplicación del juicio de experto como técnica de evaluación de las tecnologías de la información y comunicación (TIC)*. 2013.
- CARDOZO, F. *Evaluación de nuevas tecnologías de última milla para acceso dedicado a internet: ADSL (Asymmetric Digital Subscriber Line) Y HFC (Hybrid Fiber Coax)*. 2002. 1, Zulia : s.n., 2002, Télématique, Vol. 1, págs. 10-20.
- CASTRO, C. redesimalambricasunidad56. *redesimalambricasunidad56*. [En línea] 10 de Diciembre de 2014. Disponible en:
<http://redesimalambricasunidad56.blogspot.com/2014/12/511-arquitectura-de-una-red-wimax.html>.
- CAZAREZ, G. *DISEÑO DE UN PROTOTIPO DIDÁCTICO PARA LA IMPLEMENTACIÓN DE REDES DE SENSORES INALÁMBRICOS BASADOS EN EL PROTOCOLO ZIGBEE*. 2011, Ra Ximhai, pág. 12.
- CHAMBA, L. Automatización del sistema manual de información en secretaria general del GAD Municipal de Santa Rosa. 2017.
- CHIAVENATO, I. *Gestión del Talento Humano*. Mc Graw Hill. Sede Académica La Paz : Mc Graw Hill, 2009. pág. 626.
- COHEN, D y ASIN, E. *Sistemas de información para los negocios*. Tercera Edicion. s.l. : McGraw-Hill, 2000. pág. 413.

- DAVIS, G. *Sistema de Informacion Gerencial*. s.l. : McGraw-Hill, 1987.
- DOMINGUEZ, L. *Analisis de Sistemas de Informacón*. Primera edicion. Mexico : RED TERCER MILENIO S.C, 2012. pág. 105.
- DURAN, F., MONDRAGON, N. y SANCHEZ, M. *Redes cableadas e inalámbricas para transmisión de datos*. 2008, Científica, págs. 113-118.
- ENRIQUEZ, A., HAMILTON, J. y TAHA, B. *Banda ancha inalámbtrica*. españa : close mobile, 2013. Vol. 3.
- FERNANDEZ, K. *Redes de HFC. Redes de HFC*. [En línea] 6 de 4 de 2012. Disponible en: <http://redes-hfc.blogspot.com/2012/04/redes-hfc-hybrid-fibre-coaxial.html>.
- FENANDEZ, V. *Desarrollo de sistemas de informacion. Desarrollo de sistemas de informacion*. [En línea] 2006. Disponible en: https://books.google.com.pe/books?id=Sqm7jNZS_L0C&printsec=frontcover&dq=que.
- GUERRERO. *Teorema de Decisión*. 2008. pág. 120.
- GURMENDI, P. *Gestión de los sistemas de información en la toma de decisiones de la Municipalidad Distrital de El Tambo* 2018. Huancayo, Peru : s.n., 2019.
- HALLBERG, B. *Fundamentos de redes*. mexico : McGRAW-HILL/INTERAMERICANA EDITORES, 2007.
- HAYDEN, M. *EcuRed. EcuRed*. [En línea] 6 de 11 de 2001. Disponible en: https://www.ecured.cu/Redes_punto_a_punto.
- HERNANDEZ, R., FERNANDEZ, C. y BAPTISTA, L. *Metodología de la investigación*. Mexico : McGraw-Hill, 2014. pág. 634.
- HERNANDEZ, R, FERNANDEZ, C y Lucio, BAPTISTA. *Metodologia de la investigacion*. Quinta edicion. Mexcio : McGraw-Hil, 2003.
- HERNANDEZ, R., FERNANDEZ, C. y BAPTISTA, P. *Metodologia de la investigacion*. sexta edicion. México D.F. : McGRAW-HILL / INTERAMERICANA EDITORES, S.A, 2014.
- HERNANDO, J., MENDO, L. y RIERA, J. *TRANSMISION POR RADIO*. España : EDITORIAL UNIVERSITARIA RAMON ARECES, 2013.
- HERRERA. *Sistemas de información gerencial en el suministro de personal de Martínez Orozco E.U*. Disponible en: <https://repositorio.unicartagena.edu.co/bitstream/handle/11227/5480/Pasantias%20Final%20-%20LILIA%20HERRERA.pdf?sequence=1&isAllowed=y>. 2016.
- HOPKIN, P. *Fundamentals of risk management: understanding, evaluating and implementing effective risk management*. s.l. : Kogan Page Publishers, 2017. pág. 385.
- HUIDOBRO, J.. *Telecomunicaciones, Tecnologías, Redes y servicios*. madrid : RA-MA, 2014.

- KAST, F. *clasiytipodecisiones.blogspot. clasiytipodecisiones.blogspot.* [En línea] 1979. Disponible en: <http://clasiytipodecisiones.blogspot.com/2016/12/teoria-de-la-toma-de-decisiones.html>.
- KNIGHT, k. Future ISO 31000 standard on risk management, ISO focus. *Future ISO 31000 standard on risk management, ISO focus.* [En línea] 2012.
- LAPIEDRA, R. y DEVECE, C. *Introducción a la gestión de sistemas de información en la empresa.* Primera Edición. Madrid : Campus del Riu Sec, 2011.
- LAUDON, K y LAUDON, J. *Sistemas de información gerencia.* [ed.] Luis Miguel Cruz Castillo. Decimo Segunda Edición. Mexico : Pearson Educación, 2012. pág. 640.
- MARTINEZ, C y NESTOR, A. *Evaluación del desempeño de la tecnología adsl en la red de internet banda ancha.* Zulia : Télématique, 2005. págs. 141-163.
- MARTÍNEZ, J. Prored. *Prored.* [En línea] 1 de 5 de 2018. Disponible en: <https://www.prored.es/blog/fibra-optica/tecnologia-ftth-la-fibra-hasta-el-hogar/>.
- MCLEOD, J. *Sistemas de Información Gerencial.* Pearson Educación. Mexico : Pearson Educación, 2006.
- MONACHESI, E y otros , Editorial de la Universidad Tecnológica Nacional,2011, págs. 1-20.
- MONFORTE, M. *Sistemas de información para la dirección.* [ed.] Pirámide. s.l. : Pirámide, 1994. pág. 264.
- MUÑOZ, D. *SISTEMAS INALAMBRICOS DE COMUNICACION PERSONAL.* mexico : S.A. MARCOMBO, 2012.
- ÑAUPAS, H. *Metodología de la Investigación Científica y Elaboración de Tesis.* Tercera Edición. Lima : CEPREDIM, 2013.
- O'BRIEN, J y MARAKAS, G. *Sistemas de Informacion Gerencial.* Septima Edición. Mexico D.F. : McGraw-Hill, 2006. pág. 625.
- OCANDO, A. y UGAS, L. *Tecnologías para redes inalámbricas en las organizaciones del estado Zulia.* 2005, Télématique, págs. 70-86.
- OHLHORST, F. Turning Big Data into Big Money. 2018, de SAS. [En línea] 2012. Disponible en: https://www.sas.com/storefront/aux/en/spbdabm/65113_excerpt.pdf.
- OLAVSRUD. Big Data Causes Concern and Big Confusion.2018. [En línea] 2012. Disponible en: http://www.cio.com/article/700804/Big_Data_Causes_Concern_and_Big_Confusion?page=2&taxonomyId=3002.
- PADILLA. Sistema de información gerencial para las cajas solidarias de Ecuador. 2018.
- PERISSE, M. *Proyecto informático: una metodología simplificada.* 2001. Se presenta una metodología estructurada simplificada, para la automatización de

procedimientos administrativos. Esta metodología puede ser llevada adelante en proyectos de pequeño y mediano porte, por un grupo de dos a cuatro personas..

- PEZO, V. Implementación de un sistema de información gerencial para la gestión de indicadores hospitalarios del POI y PPR en la unidad de estadística del Hospital II-2 Tarapoto MINSa. Tarapoto, San Martín : s.n., 2017.
- RABANOS, J. *COMUNICACIONES MOVILES*. España : EDITORIAL UNIVERSITARIA RAMON ARECES, 2015.
- RAMOS, W. *Redes y conectividad*. Lima : RITISA GRAFF S.R.L, 2014.
- PEREZ, C., DE JESUS, H. y SALAZAR, G. *Redes Inalámbricas 802.11n el Nuevo Estándar*. 2006, Conciencia Tecnológica, pág. 3.
- RUIZ, J. *Clasificación de tráfico de red usando Hadoop y GPUs*. Universidad Nacional de Entre Ríos. Buenos Aires : <http://sedici.unlp.edu.ar/handle/10915/52630>, 2016.
- SANTA CRUZ, K. Sistema de Información Gerencial y su Influencia en la Calidad de Servicio en la Empresa Casconta E.I.R.L Ventanilla - Callao. *Sistema de Información Gerencial y su Influencia en la Calidad de Servicio en la Empresa Casconta E.I.R.L Ventanilla - Callao*. Ventanilla, Callao, Perú : s.n., 2018.
- STALLINGS, W. *COMUNICACIONES Y REDES DE COMPUTADORES*. MADRID : Pearson educación, 2004.
- TANENBAUM, A. *Redes de computadoras*. México : Pearson educación, 2003. Vol. 4.
- TAPIA, J. Sistema de información gerencial para web y control de procesos de producción en la Granja Avícola REC S.A.C. Ica, Ica, Perú : s.n., 2019.
- URSO, v. *Dirección de Proyectos exitosos*. Granica. Buenos Aires : Granica, 2014. pág. 118.
- VARGAS, L. Sistema de Información Gerencial para mejorar el Control Integral de las Sesiones de Hemodiálisis de EsSalud La Libertad. Trujillo, La Libertad, Perú : s.n., 2018.
- VERA, C., BARBOSA, J. y PABON, D. *La Tecnología ZigBee estudio de las características de la capa física*. 2017. Disponible en: https://www.iso.org/files/live/sites/iso.org/files/archive/pdf/en/future_31000_junefocus2009.pdf. 2017, Scientia Et Technica, págs. 238-245.
- VILLAVICENCIO, A. Diseño de un sistema de información gerencial para la compañía comercializadora de banano BANDECUA S.A. Guayaquil, Ecuador : s.n., 2016.

MATRIZ DE CONSISTENCIA – PROYECTO DE INVESTIGACIÓN
TABlero DE MANDO BASADO EN LINUX PARA LA PLANTA HFC DEL GRUPO IMATRA SAC

PROBLEMA GENERAL	OBJETIVO GENERAL	HIPÓTESIS PRINCIPAL	VARIABLES	DIMENSIONES	INDICADORES
¿De qué manera la implementación de un Tablero de mando basado en Linux va a mejorar la toma de decisiones para la Planta HFC en el área de operaciones del Grupo Imatra SAC?	Implementar un Tablero de Mando para la mejora de la toma de decisiones de la Planta HFC en el área de operaciones del Grupo Imatra SAC.	La Implementación de un Tablero de Mando mejoró significativamente la toma de decisiones de la Planta HFC en el área de operaciones del Grupo Imatra SAC.	Efecto del uso del Tablero de mando basado en Linux en la mejora de la planta HFC. Acuña (2019 pág. 5)		
PROBLEMA ESPECÍFICO	OBJETIVOS ESPECÍFICOS	HIPÓTESIS ESPECÍFICA			
¿Cómo la implementación del Tablero de Mando basado en Linux reduce los riesgos en la toma de decisiones de la Planta HFC en el área de operaciones del Grupo Imatra SAC?	Determinar como el Tablero de mando reduce los riesgos en la Toma de decisiones de la Planta HFC en el área de operaciones del Grupo Imatra SAC.	La Implementación de un Tablero de Mando redujo los riesgos en la toma de decisiones de la Planta HFC en el área de operaciones del Grupo Imatra SAC. Saldaña (2020 pág. 59)		Tiempo de procesamiento El ahorro de tiempo y costos es importante en la automatización del procesamiento de la información. O'Brien (2006 pág. 55)	Cantidad de almacenamiento de datos El almacenamiento de datos cobra gran relevancia en la empresa debido al volumen de información que permite procesar. Laudón (2012 pág. 158)
¿Cómo la implementación del Tablero de Mando basado en Linux influye en la mejora de los procesos para la toma de decisiones de la Planta HFC en el área de operaciones del grupo Imatra SAC?	Identificar la influencia de un Tablero de Mando en la mejora de los procesos para la toma de decisiones de la Planta HFC en el área de operaciones del Grupo Imatra SAC.	El Tablero de Mando mejoró los procesos en la toma de decisiones de la Planta HFC en el área de operaciones del Grupo Imatra SAC. Domínguez (2012 pág. 37)		Porcentaje de información Se logra consolidar los procesos y la retroalimentación de los datos de manera concisa. Herrera (2016 pág. 39)	Incremento del porcentaje de información La cantidad de información es muy importante para la organización, destacando su calidad para facilitar su análisis y depuración. Hualpa (2020, p.14)
¿Cómo la implementación del Tablero de Mando influye en	Determinar la influencia de un Tablero de Mando en los resultados	El Tablero de Mando influyó en los resultados para la toma de		Riesgo Existen desventajas dentro de la organización como el riesgo de la pérdida de información por desconocimiento del manejo de procesos de los empleados. Cortez(2020, p.27)	Reducción de riesgo Con la finalidad de mantener una adecuada evaluación de riesgos es necesaria una revisión constante de los procesos e identificar las amenazas para así minimizar los riesgos al mínimo. Rojas (2020, p.45)

los resultados para la toma de decisiones de la Planta HFC en el área de operaciones de grupo IMATRA SAC?	para la toma de decisiones de la Planta HFC en el área de operaciones de grupo Imatra SAC.	decisiones de la Planta HFC en el área de operaciones del Grupo Imatra SAC. Hopkin (2017 pág. 73)			
				Tiempo de Respuesta Es de mucha importancia definir la efectividad y el tiempo de respuesta de cada uno de los procesos de trabajo para facilitar una mejor toma de decisiones. Pizarro (2020, p.12)	Incremento de tiempo de respuesta Proceso donde la información transcurre inmediatamente. Los sistemas tienen transacciones de realizar procesamiento en tiempo real Olavsrud (2012 pág. 254)

Anexo 4 Matriz de operacionalización de variable

MATRIZ DE OPERACIONALIZACION DE VARIABLES
TABLERO DE MANDO BASADO EN LINUX PARA LA PLANTA HFC DEL GRUPO IMATRA SAC

VARIABLE	DEFINICIÓN CONCEPTUAL	DEFINICIÓN OPERACIONAL	DIMENSIÓN	INDICADORES	INSTRUMENTO	ESCALA DE MEDICIÓN
Efecto del uso del Tablero de mando basado en linux en la mejora de la planta HFC. Acuña (2019 pág. 5)	El tablero de mando logró optimizar procesos para un mejor resultado para la organización. Permitted incrementar la satisfacción del usuario en el desempeño a través de su implementación. Además, facilitando comprender problemas para la mejora de toma de decisiones de las personas involucradas. Laudón (2016)	La idea principal del proyecto fue obtener como resultado un Tablero de mando para la mejora de toma de decisiones en la Planta HFC, se logró optimizar los tiempos de espera de los reportes finales, ayudando al área de operaciones a realizar análisis de alto nivel. los datos para la muestra fueron obtenidos a través de encuesta.	Tiempo de procesamiento El ahorro de tiempo y costos es importante en la automatización del procesamiento de la información. O'Brien (2006 pág. 55)	Cantidad de almacenamiento de datos El almacenamiento de datos cobra gran relevancia en la empresa debido al volumen de información que permite procesar. Laudón (2012 pág. 158)	- Encuesta - Ficha de Observación	E. Likert
			Porcentaje de información Se logra consolidar los procesos y la retroalimentación de los datos de manera concisa. Herrera (2016 pág. 39)	Incremento del porcentaje de información La cantidad de información es muy importante para la organización, destacando su calidad para facilitar su análisis y depuración. Hualpa (2020, p.14)	- Encuesta - Ficha de Observación	E. Likert

			<p>Riesgo Existen desventajas dentro de la organización como el riesgo de la pérdida de información por desconocimiento del manejo de procesos de los empleados. Cortez (2020, p.27)</p>	<p>Reducción de riesgo Con la finalidad de mantener una adecuada evaluación de riesgos es necesaria una revisión constante de los procesos e identificar las amenazas para así minimizar los riesgos al mínimo. Rojas (2020, p.45)</p>	<p>- Encuesta - Ficha de Observación</p>	<p>E. Likert</p>
			<p>Tiempo de Respuesta Es de mucha importancia definir la efectividad y el tiempo de respuesta de cada uno de los procesos de trabajo para facilitar una mejor toma de decisiones. Pizarro (2020, p.12)</p>	<p>Incremento de tiempo de respuesta Proceso donde la información transcurre inmediatamente. los sistemas tienen transacciones de realizar procesamiento en tiempo real Olavsrud (2012 pág. 254)</p>	<p>- Encuesta - Ficha de Observación</p>	<p>E. Likert</p>

Anexo 5 Instrumento de recolección de datos

FICHA DE OBSERVACIÓN.

Días procesamiento Tiempo de Respuesta	Tiempo de Procesamiento		
	Almacenamiento de Datos	Transformación de Datos	Análisis de Datos
1	60 min	45min	30 min
2	40 min	28min	30 min
3	30 min	28min	30 min
4	25 min	20min	30 min
5	25min	18min	30 min
6	20min	18min	15min
7	16min	18min	15min
8	13min	10min	15min
9	10min	10min	10 min
10	8min	10min	10 min

Tabla 12. Ficha de Observación del Tiempo de Procesamiento

Fuente: Elaboración propia

FICHA DE OBSERVACIÓN

Días procesamiento Tiempo de Diagnostico	Porcentaje de Información		
	Nivel de manejo de toda la información	Nivel de automatización de los procesos	Nivel de apoyo para la toma de decisión
1	25%	40%	45%
2	25%	40%	45%
3	25%	40%	45%
4	25%	40%	45%
5	50%	40%	45%
6	50%	40%	55%
7	50%	80%	85%
8	85%	80%	85%
9	85%	80%	85%
10	85%	80%	85%

Tabla 13 Ficha de Observación del porcentaje de información

Fuente: Elaboración propia

N.º	DIMENSIONES / ítems	Pertinencia		Relevancia		Claridad		Observaciones
		SI	NO	SI	NO	SI	NO	
	DIMENSIÓN 1: TIEMPO DE PROCESAMIENTO							
	Indicador: Cantidad de Almacenamiento de datos	x		x		x		
	DIMENSIÓN 2: PORCENTAJE DE INFORMACION							
	Indicador: Incremento del porcentaje de Información	x		x		x		
	DIMENSIÓN 3: RIESGO							
	Indicador: Reducción de Riesgo	x		x		x		
	DIMENSIÓN 4: TIEMPO DE RESPUESTA							
	Indicador: Incremento de tiempo de respuesta	x		x		x		

Tabla 14 Evaluación de dimensión e indicadores

Fuente: Elaboración propia

Anexo 6 Validación de Instrumento

Nº	Dimensiones / ítems	Pertinencia ¹		Relevancia ²		Claridad ³		Sugerencias
		Si	No	Si	No	Si	No	
	I. Tiempo de procesamiento							
1	¿El tablero de mando, ayuda en el cumplimiento y mejora del tiempo de procesamiento?	x		x		x		
2	¿Está de acuerdo que la transformación de la información mejora el tiempo de procesamiento?	x		x		x		
3	¿Considera usted que el almacenamiento de datos mejorará los resultados de los procesos?	x		x		x		
4	¿Cree usted que el tablero de mando ayuda en el tiempo de procesamiento?	x		x		x		
5	¿Está de acuerdo que los análisis de datos ayudan en la inteligencia de negocios?	x		x		x		
6	¿El tiempo de procesamiento es fundamental para el almacenamiento de un tablero de mando?	x		x		x		
	II. Porcentaje de información	Si	No	Si	No	Si	No	
7	¿Considera usted que un tablero de mando brinda apoyo para la toma de decisión?	x		x		x		
8	¿Está de acuerdo que la inyección de datos mejorará en la gestión de los procesos?	x		x		x		
9	¿Está de acuerdo de que un tablero de mando ayuda a no tener pérdida de información?	x		x		x		
10	¿Cree usted que la información almacenada en la base de datos ayuda en los resultados?	x		x		x		
11	¿Considera usted que analizar la integración de los procesos mejorará con el uso del tablero de mando?	x				x		
12	¿Piensa usted que la integridad de los procesos ayuda en la automatización de la información?	x		x		x		

Nº	Dimensiones / ítems	Pertinencia ¹		Relevancia ²		Claridad ³		Sugerencias
		Si	No	Si	No	Si	No	
	III. Riesgo							
13	¿Considera usted que un tablero de mando reduce el riesgo en la toma de decisión dentro de la organización?	x		x		x		
14	¿Piensa usted que una buena toma de decisión nos permite determinar el riesgo en el tiempo para obtener buenos resultados?	x		x		x		
15	¿El cumplimiento de la seguridad, es un factor impórtate al tomar una buena decisión?	x		x		x		
16	¿Está de acuerdo que en la organización la seguridad de riesgo es un factor importante?	x		x		x		
	IV. Resultados							
17	¿Piensa usted que el tiempo de espera ha mejorado en obtener los resultados en los Dashboards?	x		x		x		
18	¿Al termino de las actividades considera usted que el tablero de mando ha mejorado en los tiempos de esperas dentro de la organización?	x		x		x		
19	¿Considera usted que los resultados de los procesos han mejorado dentro de la organización con un tablero de mando?	x		x		x		
20	¿Cree usted que la retroalimentación nos ayuda a mejorar los resultados?	x		x		x		
21	¿La gerencia está de acuerdo que una buena planificación ayuda en los problemas de la organización?	x		x		x		
22	¿Está de acuerdo que una toma de decisión efectiva nos ayuda a obtener buenos resultados?	x		x		x		

Tabla 15 Validación de instrumento

Fuente: Elaboración propia

Anexo 7 Cuestionario

Tablero de mando basado en Linux para la planta HFC del Grupo Imatra SAC

I. DATOS GENERALES

Valora de acuerdo con la siguiente escala: marca con una "X" el casillero de su preferencia.

- (1) Totalmente de acuerdo
- (2) De acuerdo
- (3) Ni de acuerdo ni en desacuerdo
- (4) En Desacuerdo
- (5) Totalmente en desacuerdo

Tabla 16 Cuestionario - Fuente: Elaboración propia

Efecto del uso del Tablero de mando basado en Linux en la mejora de la planta HFC.		1	2	3	4	5
DIMENSIÓN 1: Tiempo de procesamiento						
1	¿El tablero de mando, ayuda en el cumplimiento y mejora del tiempo de procesamiento de los datos?	x				
2	¿Está de acuerdo que la transformación de la información mejora el tiempo de procesamiento de los datos?	x				
3	¿Considera usted que el almacenamiento de datos mejorará los resultados del tiempo de procesamiento?	x				
4	¿Cree usted que el tablero de mando ayuda en la mejora de tiempo de procesamiento datos?	x				
5	¿Está de acuerdo que los análisis de datos ayudan en la inteligencia de negocios?	x				
6	¿Cómo califica que el procesamiento de datos es fundamental para el almacenamiento del tablero de mando?	x				
DIMENSIÓN 2: Porcentaje de información						
7	¿Los datos extraídos ayudan en la transformación de los procesos?	x				
8	¿Cómo califica el control y la mejora del porcentaje de información analizada?	x				
9	¿Está de acuerdo de que un tablero de mando ayuda a no tener pérdida de información?	x				
10	¿Cree usted que la información almacenada en la base de datos ayuda en los resultados?	x				
11	¿Considera usted que analizar la información explotada mejorara con el uso del tablero de mando?	x				
12	¿Piensa usted que la integridad de los procesos ayuda en la automatización de la información?	x				

13	¿El manejo del porcentaje de información ayuda a medir los resultados administrativos?	x				
14	¿Está de acuerdo que la analítica de datos ayuda a manejar grandes volúmenes de información para el correcto almacenamiento en un ambiente Big Data?	x				
15	¿Cómo califica usted que la automatización de los procesos ayuda en la eficacia del tablero de mando?	x				
16	¿Considera usted que la automatización otorga mejor control de la información?	x				

		1	2	3	4	5
Efecto del uso del Tablero de mando basado en Linux en la mejora de la planta HFC.						
DIMENSIÓN 3: Riesgo						
17	¿Considera usted que un tablero de mando reduce el riesgo en la toma de decisión dentro de la organización?	x				
18	¿Piensa usted que una buena toma de decisión nos permite determinar el riesgo en el tiempo para obtener buenos resultados?	x				
19	¿El cumplimiento de la seguridad, es un factor impórtate al tomar una buena decisión?	x				
20	¿Está de acuerdo que en la organización la seguridad de riesgo es un factor importante?	x				
DIMENSIÓN 4: Tiempo de respuesta						
21	¿Piensa usted que el tiempo de espera ha mejorado en obtener los resultados en los DASHBOARD?	x				
22	¿Al termino de las actividades considera usted que el tablero de mando ha mejorado en los tiempos de esperas dentro de la organización?	x				
23	¿Considera usted que los tiempos de respuesta de los procesos han mejorado dentro la organización con un tablero de mando?	x				
24	¿Cree usted que la retroalimentación nos ayuda a mejorar los resultados?	x				
25	¿La gerencia está de acuerdo que una buena planificación ayuda en los problemas de la organización?	x				
26	¿Está de acuerdo una toma de decisión efectiva nos ayuda a obtener buenos tiempos de respuesta?	x				

Anexo 8 Matriz de Datos

N° de Encuestado	EFECTO DEL USO DEL TABLERO DE MANDO BASADO EN LINUX EN LA MEJORA DE LA PLANTA HFC.																																				
	DIMENSIÓN 1: Tiempo de procesamiento						DIMENSION 2 Porcentaje de información										DIMENSION 3 Riesgo						DIMENSION 4: Tiempo de respuesta						Resultados								
	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	p12	p13	p14	p15	p16	p17	p18	p19	p20	p21	p22	p23	p24	p25	p26	p27	p28	p29	p30	p31	p32					
1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	2	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1		
2	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	2	2	1	2	1	1	1	1	1	1	2	1	1	1	1	1	1	1		
3	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	2	2	1	2	1	1	1	1	1	1	2	1	2	1	2	1	1	1		
4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	1	1	1	1	2	1	1	1	2	1	1	1	1	2	1	1		
5	1	1	1	1	2	1	1	1	2	1	1	1	1	1	1	1	1	2	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	2	1	
6	1	1	2	1	2	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	2	1	1	1	1	1	1	1	1	2	2	1	1	
7	1	1	1	1	2	1	1	1	1	1	1	2	1	1	1	1	1	2	1	1	1	1	1	2	1	1	1	1	1	1	1	1	2	2	1	1	
8	1	1	1	1	2	1	1	1	1	2	1	2	1	1	1	1	1	2	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	
9	1	1	1	1	1	1	1	1	2	2	2	2	1	1	1	1	1	2	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
10	1	1	2	1	1	1	1	1	2	2	2	2	1	1	1	1	1	2	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
11	1	2	2	1	1	1	1	1	2	1	1	2	1	1	1	1	1	2	1	1	2	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	
12	1	1	1	1	1	1	1	1	1	1	1	2	1	2	1	1	1	2	1	1	1	2	2	2	2	1	2	2	1	1	1	1	1	1	1	1	
13	1	1	1	2	1	1	1	1	1	1	1	2	1	2	1	1	1	1	1	1	2	2	2	2	1	1	2	1	1	1	1	1	1	1	1	1	
14	2	1	1	2	1	1	1	1	1	2	1	1	1	2	1	1	1	1	1	1	2	2	2	2	1	2	2	1	1	1	1	1	1	1	1	1	
15	2	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	1	2	1	2	2	1	1	1	1	1	1	1	1	1	
16	1	1	1	1	1	1	1	2	1	1	1	1	1	1	2	1	1	1	2	1	1	2	2	1	1	1	1	2	1	1	1	1	1	1	1	1	
17	1	1	1	1	1	1	1	2	1	1	1	1	1	1	2	1	1	1	2	1	1	2	2	1	1	2	1	2	1	2	1	1	1	1	1	1	
18	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	2	1	1	1	2	1	1	2	1	1	1	1	1	1	1	1	2	1	
19	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	2	1	1	1	1	1	1	1	1	2	1	
20	1	2	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	2	1	1	1	1	1	2	1	1	1	1	1	1	1	1	2	1	
21	1	2	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1
22	1	2	1	2	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1
23	2	2	1	2	1	1	1	1	1	1	1	1	2	1	2	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	2	1
24	2	2	1	2	1	1	1	1	1	1	1	1	2	1	2	1	1	1	1	1	1	1	1	1	2	1	1	1	2	1	1	1	2	1	2	1	
25	2	1	1	2	1	1	1	1	1	1	2	2	1	1	2	1	1	1	1	1	1	1	1	1	2	1	1	1	2	1	1	1	2	1	2	1	
26	1	1	1	2	1	1	1	1	1	1	2	2	1	1	2	1	1	1	1	1	1	1	1	1	2	1	1	1	2	1	1	1	2	1	2	1	
27	1	1	2	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	2	1	1	1	2	1	1	1	2	1	2	1	
28	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	
29	1	1	2	1	1	2	2	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1
30	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	

Tabla 17 Matriz de datos – Fuente: SPSS Statistics

N° de Encuestado	VARIABLE INDEPENDIENTE:SISTEMA DE INFORMACIÓN GERENCIAL																VARIABLE DEPENDIENTE:TOMA DE DECISIONES																
	DIMENSIÓN 1: Tiempo de procesamiento						DIMENSION 2: Integración de la Información										DIMENSION 3 Riesgo				DIMENSION 4: Tiempo de respuesta						Resultados						
	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	p12	p13	p14	p15	p16	p17	p18	p19	p20	p21	p22	p23	p24	p25	p26	p27	p28	p29	p30	p31	p32	
si	24	24	25	22	26	27	28	28	25	26	26	21	25	27	20	30	30	19	7	27	23	24	22	23	21	26	27	21	26	27	12	30	
no	6	6	5	8	4	3	2	2	5	4	4	9	5	3	10	0	0	11	0	3	7	6	8	7	9	4	3	9	4	3	18	0	
total	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	7	30	30	30	30	30	30	30	30	30	30	30	30	30	30

Anexo 9 Propuesta de valor

1. Sistema de Información de la Metodología Aplicada

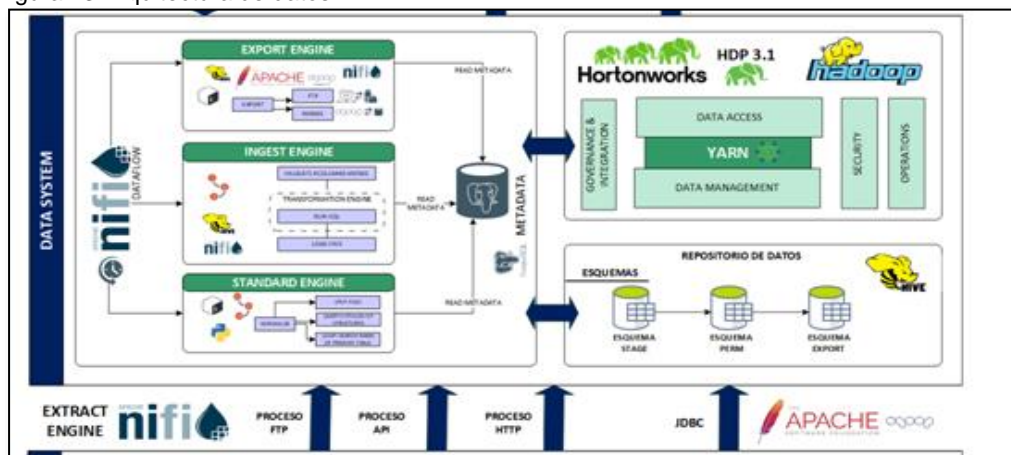
1.1 Nombre y descripción del Sistema de Información

El desarrollo del sistema está enfocado en un ambiente big data para la explotación de la información y la obtención de los resultados a través de reportes en Excel o Dashboards.

Teniendo en cuenta las siguientes características:

- Para la ingesta de los datos y almacenamiento de la información, se está implementando en base de datos NoSQL Hive y para la distribución de los datos Hadoop para almacenar gran cantidad de información.
- Se utiliza PostgreSQL, para almacenar la metadata de los planos a ingestar y de los reportes para exportar.
- Nifi para la extracción, estandarización, ingesta, transformación y exportación de los datos analizados
- Lenguaje de programación Python, para la lógica del desarrollo de los flujos que contiene toda la lógica del negocio de los procesos de ejecución.
- Todo este desarrollo se va ejecutar en un servidor Linux, ya que se va ejecutar todo por consola, y los resultados de las ejecuciones se va poder visualizar en las tablas log que va dejando el proceso Nifi.

Figura 13 Arquitectura de datos



1.2 Objetivo del Sistema de Información

El objetivo principal del sistema de información es, procesar toda la información en el servidor Linux, para poder incluir el nuevo proceso de Monitoreo; el cual está conformado por la Ingesta, Transformación y Delivery de la información.

Los cambios necesarios para la atención abarcan los siguientes sistemas integrados:

- Servidor LINUX
- Hive
- Postgress
- NiFi
- Clúster Hadoop (rutas HDFS)
- Reportes Excel

1.3 Alcance del Sistema de Información

El sistema de información gerencial permite reportes y Dashboards definidos por las áreas usuarias.

El sistema de información gerencial permite la determinación proactiva, predictiva y reactiva de eventos, anomalías que se presenten en el funcionamiento de las redes HFC, con alta disponibilidad y acceso a la información, manteniendo la información actualizada

- Alta tasa de correlacionamiento de datos que permita alertar, diagnosticar sobre cualquier tipo de incidencia en la red, así como emitir reportes a demanda, de forma casi inmediata.

1.4 Restricciones del Sistema de Información

- No se emita reportes a demanda
- No permita conexión con otras bases de datos relacionales
- No guardara data más de 1 año.

- Solo se podrá tener la foto del momento, en caso se caiga el archivo fuente y no se pueda extraer. Esa información se pierde y no es recuperable.

1.5 Componentes del Sistema de Información

1.5.1 Apache NIFI

El sistema de información gerencial se desarrolla con Apache Nifi 1.8

- Data Ingestion Processors: Se encargan de obtener datos input desde orígenes. El sistema de información gerencial que será desplegado en Linux, considera el uso de los siguientes procesadores: GetFTP, GetHTTP
- Routing and Mediation Processors: Se encargan de la evaluación y ruteo en base a un criterio específico del DataFlow. El sistema de información gerencial que será desplegado por linux considera el uso de los siguientes procesadores: RouteOnAttribute, ControlRate
- Database Access Processors: Se encargan de realizar tareas que incluyan el acceso a orígenes del tipo Base de datos a través de la especificación JDBC. El sistema de información gerencial que será desplegado en Linux, considera el uso de los siguientes procesadores: ExecuteScript, ExecuteSQL, SelectHiveQL
- Attribute Extraction Processors: Se encargan de la extracción de atributos de determinado input y convertirlos a determinado formato. El sistema de información sera desplegado en Linux, considera el uso de los siguientes procesadores: AttributesToJSON, AttributesToCSV, ExtractAvroMETADATA,
- System Interaction Processors: Se encargan de la interacción entre Apache Nifi y el clúster Big Data. El sistema de información

gerencial que será desplegado en Linux, considera el uso de los siguientes procesadores: ExecuteStreamCommand, ExecuteProcess, LogAttribute, UnpackContent, LogMessage, LookupAttribute

- Export Processors: Se encargan de la exportación de datos o archivos. El sistema de información gerencial que será desplegado en Linux, considera el uso de los siguientes procesadores: PutEmail, PutFile, PutSFTP, PutFTP, PutHDFS, PutParquet

- Data Transformation Processors: Se encargan de la transformación de datos El sistema de información gerencial que será desplegado en Linux, considera el uso de los siguientes procesadores: ConvertRecord.

- Splitting and Aggregation Processors: Se encargan de la división y agregación de archivos. El gestor de trazabilidad que será desarrollado para QHAPAX considera el uso de los siguientes procesadores: SplitAvro, SplitContent, SplitJson, SplitRecord, SplitXml.

- HTTP Processors: Se encargan de la conexión HTTP entre el clúster y entornos HTTP a través del uso de estos procesadores. El sistema de información gerencial para su desarrollo, considera el uso de los siguientes procesadores para la extracción de datos desde Web los siguientes procesadores: GetHTTP, HandleHttpRequest

- Format Conversion Processors: Se encargan de la conversión de formato de archivos dentro del flujo. El sistema de información gerencial para su desarrollo, considera el uso de los siguientes procesadores: ConvertAvroToJSON, ConvertAvroToORC, ConvertCSVToAvro, ConvertExcelToCSVProcessor, ConvertJSONToAvro.

1.5.1.1 Procesos con Apache NIFI

Obtención de archivos

HTTP: Lee la metadata y envía una solicitud del servidor web destino de esta forma obtiene los archivos que se enviarán al directorio input del HDFS.

API: Lee la metadata y envía una solicitud a una URL que devuelve un archivo de respuesta que son XML, JSON y los archivos recibidos se envían a un directorio input del HDFS.

Ingesta

XML: Ejecuta un query para mapear los campos y valores a cargar, valida si existe la tabla en el STAGE, si existe lo borra para crear la tabla e insertar los datos de los archivos con formato XML al esquema PERM.

TXT: Se realiza la validación de existencia de la tabla en el STAGE, si existe lo borra para crear la tabla e insertar los datos de los archivos con formato TXT al esquema PERM

CSV: Se realiza la validación de existencia de la tabla en el STAGE, si existe lo borra para crear la tabla e insertar los datos de los archivos con formato CSV al esquema PERM

JSON: Obtiene los valores del campo a través del uso de JSON PATH, valida si existe la tabla en el STAGE, si existe lo borra para crear la tabla e insertar los datos de los archivos con formato JSON al esquema PERM.

Exportación

FTP: Lee la metada de PostgreSQL para saber los datos del servidor destino y el archivo a exportar, luego obtiene el archivo del HDFS en el directorio Export, finalmente realiza el FTP al directorio destino y registra el log en la tabla event_output_detail.

SQOOP: Lee la metada de PostgreSQL para saber los datos de la base de datos destino y el archivo a exportar, luego obtiene el archivo del HDFS en el directorio Export, finalmente ejecuta el comando de exportación en SQOOP y registra el log en la tabla event_output_detail.

Depuración de archivos y BD:

Lee la metada de PostgreSQL, ejecuta el comando para depuración de archivos con latencia mayor a la configurada en la tabla de PostgreSQL y en la base de datos ejecuta una sentencia en HIVE para obtener los datos que tengan una mayor latencia a la configurada en la tabla de PostgreSQL.

1.5.2 Metadata (PostgreSQL)

El esquema traceability considera las tablas relacionadas a la configuración de los procesos de ingesta, proceso de transformación y proceso de exportación:

- **extract_definition**: Contiene los datos básicos para la extracción de fuentes.

Esquema: traceability

Nombre de la tabla: Extract_definition

Esquema	Tabla	Campo	Tipo de datos
traceability	extract_definition	extract_process	VARCHAR(50)

traceability	extract_definition	extract_source_hostname	VARCHAR(50)
traceability	extract_definition	extract_source_path	VARCHAR(50)
traceability	extract_definition	extract_source_user	VARCHAR(50)
traceability	extract_definition	extract_source_password	VARCHAR(50)
traceability	extract_definition	extract_target_path	VARCHAR(50)
traceability	extract_definition	extract_filename	VARCHAR(50)
traceability	extract_definition	extract_filetype	VARCHAR(50)
traceability	extract_definition	Versión	VARCHAR(50)

Tabla 18 Esquema de trazabilidad: tabla: Extract_definition

- **standard_definition:** Contiene los datos básicos de los distintos tipos de estandarización soportados por sistema de información gerencial.

Esquema: traceability

Nombre de la tabla: Standard_definition

Esquema	Tabla	Campo	Tipo de datos
traceability	standard_definition	standard_definition	VARCHAR(50)
traceability	standard_definition	type_standarization	VARCHAR(100)
traceability	standard_definition	source_name	VARCHAR(100)
traceability	standard_definition	Versión	VARCHAR(50)

Tabla 19 Esquema de trazabilidad: tabla: Standard_definition

- **structure_definition:** Contiene la METADATA de configuración de estructura diferentes, esta tabla aplica en el tipo de estandarización de estructuras variables, si un archivo fuente (source) pertenece a esta casuística entonces se consultará esta tabla como un catálogo de estructuras, además de presentar.

Esquema: traceability

Nombre de la tabla: structure_definition

Esquema	Tabla	Campo	Tipo de datos
traceability	structure_definition	structure_code	VARCHAR(50)
traceability	structure_definition	source_name	VARCHAR(100)

traceability	structure_definition	standard_definition	VARCHAR(50)
traceability	structure_definition	Versión	VARCHAR(50)

Tabla 20 Esquema de trazabilidad tabla: structure_definition

- **source_definition:** Contiene la configuración de la definición de orígenes de datos de los procesos de carga configurados en el gestor de trazabilidad.

Esquema: traceability

Nombre de la tabla: source_definition

Esquema	Tabla	Campo	Tipo de datos
traceability	source_definition	source_name	VARCHAR(50)
traceability	source_definition	Origin	VARCHAR(50)
traceability	source_definition	hdfs_path	VARCHAR(250)
traceability	source_definition	Delimiter	VARCHAR(50)
traceability	source_definition	load_type	VARCHAR(50)
traceability	source_definition	control_file	VARCHAR(50)
traceability	source_definition	purge_retention	VARCHAR(50)
traceability	source_definition	ingestion_frequency	VARCHAR(50)
traceability	source_definition	Priority	VARCHAR(50)
traceability	source_definition	last_update_ts	VARCHAR(50)
traceability	source_definition	project_name	VARCHAR(50)
traceability	source_definition	Versión	VARCHAR(50)

Tabla 21 Esquema de trazabilidad tabla: source_definition

- **source_field:** Contiene la configuración de campos que contendrán los archivos fuente.

Esquema: traceability

Nombre de la tabla: source_field

Esquema	Tabla	Campo	Tipo de datos
traceability	source_field	source_name	VARCHAR(50)
traceability	source_field	field_name	VARCHAR(50)
traceability	source_field	Datatype	VARCHAR(50)

traceability	source_field	Length	VARCHAR(50)
traceability	source_field	default_value	VARCHAR(50)
traceability	source_field	Format	VARCHAR(50)
traceability	source_field	Nullable	VARCHAR(50)
traceability	source_field	field_order	VARCHAR(50)
traceability	source_field	Versión	VARCHAR(50)

Tabla 22 esquema de trazabilidad tabla: source_field

- **target_definition:** Contiene la configuración de tablas destino en Hive, donde se debe almacenar la información proveniente de la Ingesta de los archivos fuente.

Esquema: traceability

Nombre de la tabla: target_definition

Esquema	Tabla	Campo	Tipo de datos
traceability	target_definition	target_name	VARCHAR(50)
traceability	target_definition	target_layout_name	VARCHAR(50)
traceability	target_definition	source_name	VARCHAR(50)
traceability	target_definition	target_table	VARCHAR(50)
traceability	target_definition	orphan_target_table	VARCHAR(50)
traceability	target_definition	Type	VARCHAR(50)
traceability	target_definition	Subtype	VARCHAR(50)
traceability	target_definition	Delimiter	VARCHAR(50)
traceability	target_definition	transformation_name	VARCHAR(50)
traceability	target_definition	project_name	VARCHAR(50)
traceability	target_definition	Versión	VARCHAR(50)

Tabla 23 Esquema de trazabilidad tabla: target_definition

- **target_field:** Contiene la configuración de los campos de las tablas destino en Hive, para los procesos de carga registrados en el Gestor de Trazabilidad.

Esquema: traceability

Nombre de la tabla: target_field

Esquema	Tabla	Campo	Tipo de datos
---------	-------	-------	---------------

traceability	target_field	target_layout_name	VARCHAR(50)
traceability	target_field	field_name	VARCHAR(50)
traceability	target_field	field_name_filter	VARCHAR(50)
traceability	target_field	Datatype	VARCHAR(50)
traceability	target_field	Length	VARCHAR(50)
traceability	target_field	default_value	VARCHAR(50)
traceability	target_field	Format	VARCHAR(50)
traceability	target_field	Nullable	VARCHAR(50)
traceability	target_field	field_order	VARCHAR(50)
traceability	target_field	Versión	VARCHAR(50)

Tabla 24 esquema de trazabilidad tabla: target_field

- **transformation_definition:** Contiene la configuración de los procesos de transformación configurados en el Gestor de trazabilidad.

Esquema: traceability

Nombre de la tabla: transformation_definition

Esquema	Tabla	Campo	Tipo de datos
traceability	transformation_definition	transformation_name	VARCHAR(50)
traceability	transformation_definition	transform_script	VARCHAR(50)
traceability	transformation_definition	orphan_indicator_column	VARCHAR(50)
traceability	transformation_definition	orphan_purge_retention	VARCHAR(50)
traceability	transformation_definition	Versión	VARCHAR(50)

Tabla 25 Esquema de trazabilidad tabla: transformation_definition

- **target_export:** Contiene la definición de los procesos de exportación.

Esquema: traceability

Nombre de la tabla: target_export

Esquema	Tabla	Campo	Tipo de datos
traceability	target_export	target_name	VARCHAR(50)
traceability	target_export	export_name	VARCHAR(50)
traceability	target_export	file_prefix	VARCHAR(50)

traceability	target_export	file_extension	VARCHAR(50)
traceability	target_export	post_export_script	VARCHAR(50)
traceability	target_export	orphan_post_export_script	VARCHAR(50)
traceability	target_export	script_type	VARCHAR(50)
traceability	target_export	flag_active	VARCHAR(50)
traceability	target_export	start_dt	VARCHAR(50)
traceability	target_export	end_dt	VARCHAR(50)
traceability	target_export	filter	VARCHAR(1000)
traceability	target_export	version	VARCHAR(50)

Tabla 26 Esquema de trazabilidad tabla: target_export

- **export_definition:** Contiene la configuración de los procesos de exportación de archivos especificando: servidor de destino, credenciales de acceso, ruta, tipo de archivo y nombre del archivo de exportación.

Esquema: traceability

Nombre de la tabla: export_definition

Esquema	Tabla	Campo	Tipo de datos
traceability	export_definition	export_name	VARCHAR(50)
traceability	export_definition	Type	VARCHAR(50)
traceability	export_definition	Hostname	VARCHAR(50)
traceability	export_definition	Username	VARCHAR(50)
traceability	export_definition	Password	VARCHAR(50)
traceability	export_definition	Filepath	VARCHAR(50)
traceability	export_definition	Versión	VARCHAR(50)

Tabla 27 esquema de trazabilidad tabla: export_definition

1.5.3 Hive

Hive almacena los datos resultantes del proceso de Ingesta de archivos fuentes, donde se pueden considerar procesos de transformación representados por la ejecución de scripts con extensión .hql que afectarán los datos almacenados en el esquema PERM. Por ello en Hive se cuenta con los siguientes esquemas:

- **STAGE:** Contiene las tablas que almacenan la información proveniente de los archivos fuente, considera únicamente la información del archivo fuente utilizado por el proceso de carga actual.
- **PERM:** Contiene las tablas pertenecientes a los procesos configurados en el Gestor de Trazabilidad, en este esquema se almacenará la información histórica de los procesos de carga.
- **Export:** Contiene las tablas generadas por el proceso de exportación (Export engine) del Gestor de Trazabilidad.

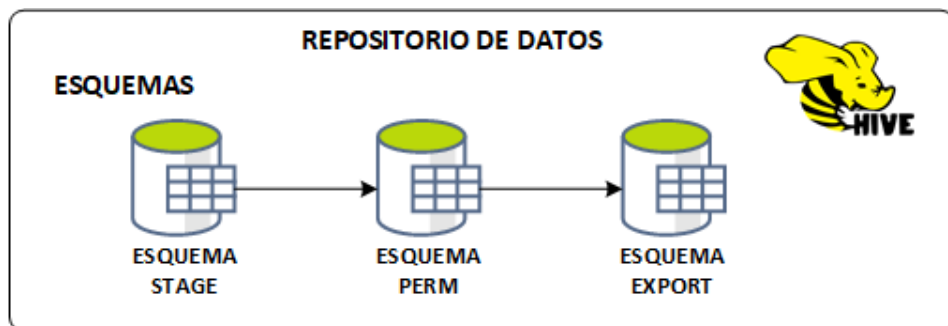


Figura 14 Repositorio de datos

1.6 Funcionalidades del Sistema de Información

Obtención de archivos fuente a través de los siguientes procesos:

- **Proceso FTP:** Se obtiene los archivos fuentes provenientes de otros servidores filesystem a través de FTP, estos archivos serán enviados al clúster al directorio ftp (`/var/opt/bd/input/ftp`). Se utilizarán credenciales configuradas en la tabla `origin_extrac_definition` de PostgreSQL. Se registrará los detalles del envío FTP en la tabla `event` del esquema control de PostgreSQL, además en el archivo log se almacenará detalles de validación de envío FTP.
- **Proceso API:** Se consume información proveniente de Web services, para ello se utilizará NIFI para la conexión al endpoint al web service respectivo, se obtendrá un response en formato (json/xml) el cual será enviado a la carpeta input. Se utilizarán

credenciales configuradas en la tabla `origin_extract_definition` de PostgreSQL. Se registrará los detalles del envío FTP en la tabla `event` del esquema `control` de PostgreSQL,

- **Proceso RDBMS:** Se obtiene archivos resultantes de exportación de consultas realizadas directamente a base de datos, el resultado del proceso de obtención RDBMS serán enviados a la carpeta `input` del clúster. Se utiliza credenciales configuradas en la tabla `origin_extract_definition` de PostgreSQL. Se registrará los detalles del envío FTP en la tabla `event` del esquema `control` de PostgreSQL,

Realizar la ingesta de archivos fuente de los tipos: Fuentes regulares, Fuentes de estructura con estandarización.

En los casos de tipos de fuentes pertenecientes a estructura con estandarización el sistema permite generar uno o más archivos de tipo regular que serán las entradas del proceso de ingesta de archivos.

Realiza procesos de transformación, el cual ejecutará en Hive los scripts configurados en la tabla de PostgreSQL: `transformation_definition`, estos procesos consideran la lógica de campos, agregaciones, filtros que sean necesarios para los procesos configurados.

Realiza la exportación de información según la definición configurada en las tablas `target_export` y `export_definition` de PostgreSQL, se considera la generación de archivos de exportación (formato: archivo plano) y envío a otros servidores filesystem, adicionalmente permite la exportación de información directa a través del uso de Apache Sqoop 1.4.7 a otras base de datos (RDBMS) considerando en el alcance actual: Oracle y Netezza.

1.7 Tipos de archivos soportador

Se consideran dos tipos de archivos soportados: Fuentes regulares, estructuras con estandarización.

- Fuentes regulares: Dentro de este grupo se consideran archivos estructurados y semi estructurados. Los formatos de archivos soportados por:
 - Archivos planos (csv, text file, .dat)
 - Archivos formato XML
 - Archivos formato JSON
 - Archivos formato AVRO

- Estructuras con estandarización: Se considera las siguientes casuísticas soportadas.
 - Archivos con estructura variable a través del tiempo.
 - Archivos que contienen más de una estructura y serán divididos por NIFI en distintos archivos por cada estructura.
 - Archivos que serán divididos en más de un archivo de acuerdo a las tablas que correspondan.
 - Archivos que deben ser extraídos de acuerdo a coincidencia con un patrón configurado.

Flujo de datos del proceso NIFI

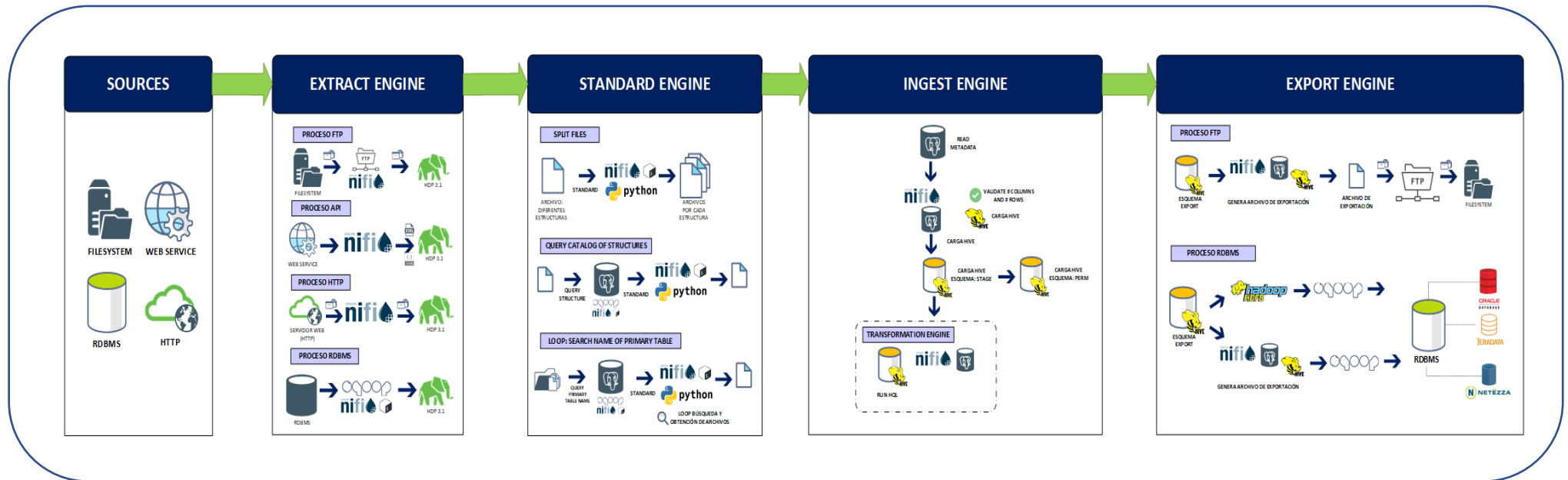


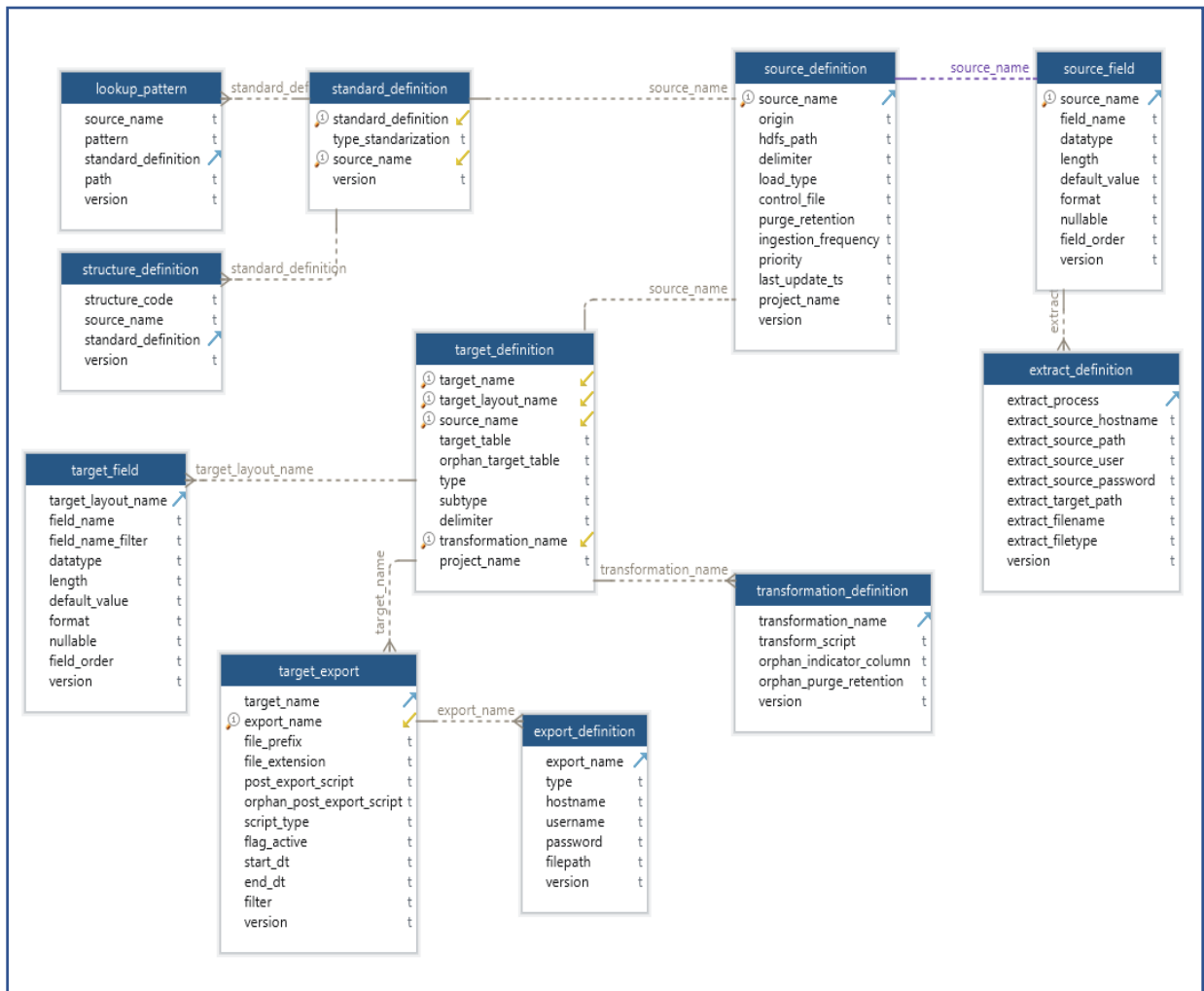
Figura 15 Flujo de datos del proceso NIFI

1.8 Modelo de datos

Esquema traceability

Esquema de todas las tablas que están involucradas en la inserción de la metadata, para poder extraer la información en los reportes.

Figura 16 Esquema traceability



Esquema control

Control de los jobs y de los logs de los procesos a ejecutar.

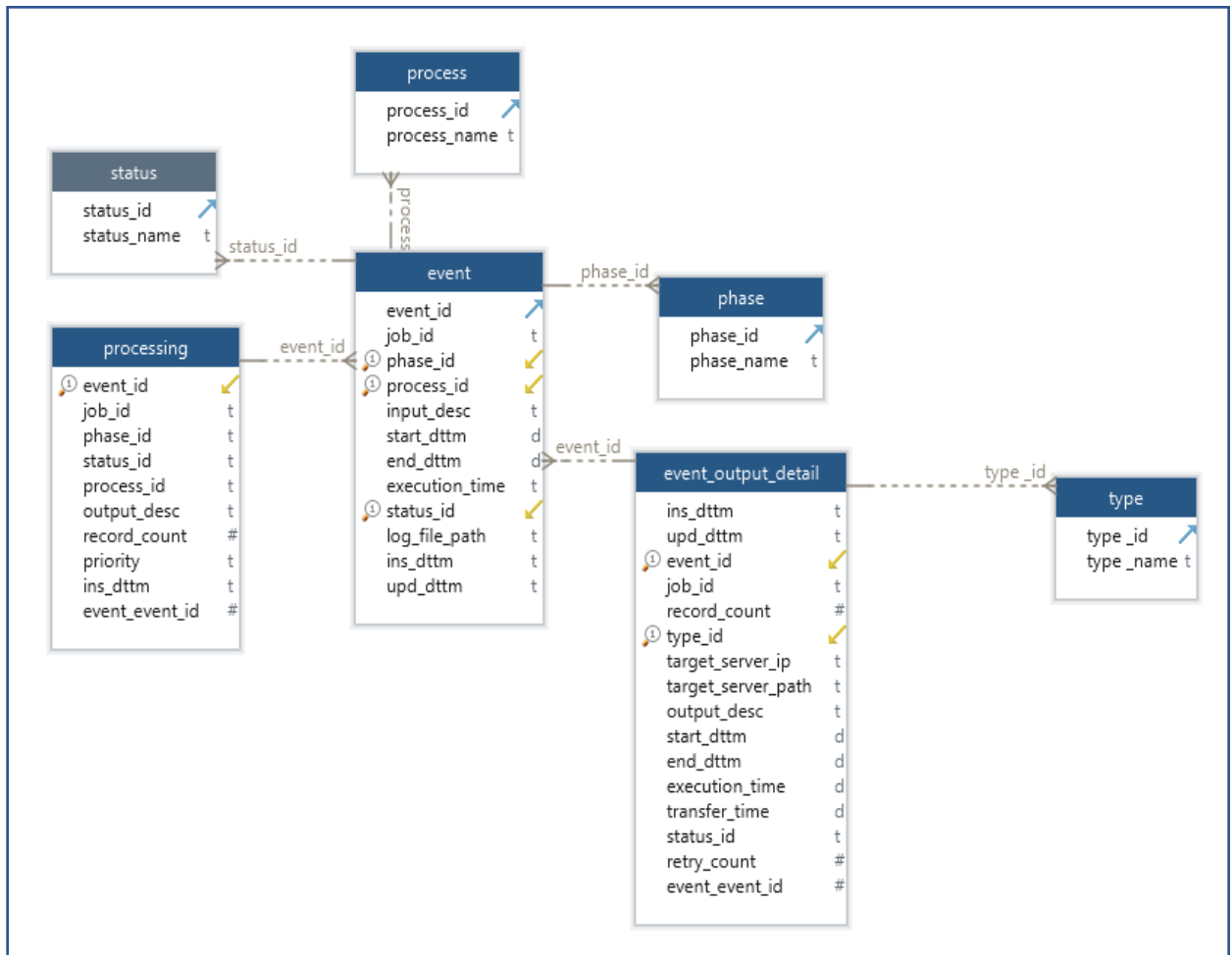


Figura 17 Esquema de control

Table export_definition Contiene la configuración de los procesos de exportación de archivos especificando: servidor de destino, credenciales de acceso, ruta, tipo de archivo y nombre del archivo de exportación.			
export_name	varchar(50)		
type	varchar(50)		
hostname	varchar(50)		
username	varchar(50)		
password	varchar(50)		
filepath	varchar(50)		
version	varchar(50)		
Foreign Keys			
fk_export_definition_target_export (export_name)	ref target_export (export_name)		

Figura 18 Configuración de la tabla Export_definition

Table extract_definition Contiene los datos básicos para la extracción de fuentes			
extract_process	varchar(50)		
extract_source_hostname	varchar(50)		
extract_source_path	varchar(50)		
extract_source_user	varchar(50)		
extract_source_password	varchar(50)		
extract_target_path	varchar(50)		
extract_filename	varchar(50)		
extract_filetype	varchar(50)		
version	varchar(50)		
Foreign Keys			
fk_extract_definition_source_field (extract_process)	ref source_field (source_name)		

Figura 19 Configuración de datos de la tabla extract_definition

Table source_definition Contiene la configuración de la definición de orígenes de datos de los procesos de carga configurados en el gestor de trazabilidad.			
	load_type	varchar(50)	
	control_file	varchar(50)	
	purge_retention	varchar(50)	
	ingestion_frequency	varchar(50)	
	priority	varchar(50)	
	last_update_ts	varchar(50)	
	project_name	varchar(50)	
	version	varchar(50)	
Indexes			
Unq	uniq_source_definition_source_name	source_name	
Foreign Keys			
	fk_source_definition_target_definition (source_name) ref target_definition (source_name)		
	fk_source_definition_standard_definition (source_name) ref standard_definition (source_name)		

Figura 20 Configuración de los datos de la tabla source_definition

Table source_field Contiene la configuración de campos que contendrán los archivos fuente			
Unq	source_name	varchar(50)	
	field_name	varchar(50)	
	datatype	varchar(50)	
	length	varchar(50)	
	default_value	varchar(50)	
	format	varchar(50)	
	nullable	varchar(50)	
	field_order	varchar(50)	
	version	varchar(50)	
Indexes			
Unq	uniq_source_field_source_name	source_name	
Foreign Keys			
	fk_source_field_source_definition (source_name) ref source_definition (source_name)		

Figura 21 Configuración de los datos de la tabla source_field

1.9 Diagrama de secuencias o de flujo

Diagrama de flujo: Proceso de Obtención de fuentes

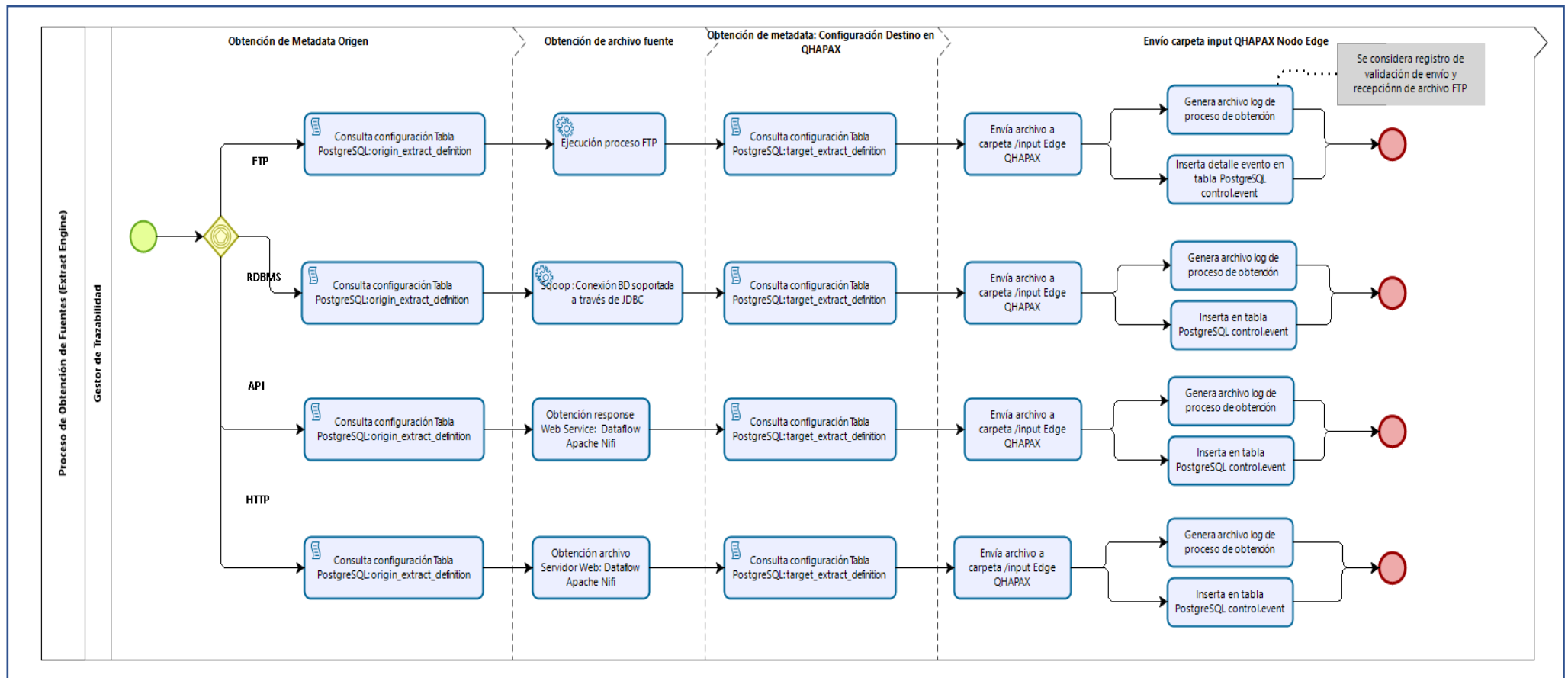


Figura 22 Diagrama de flujo: Proceso de obtención de fuentes

Diagrama de flujo: Proceso de Estandarización

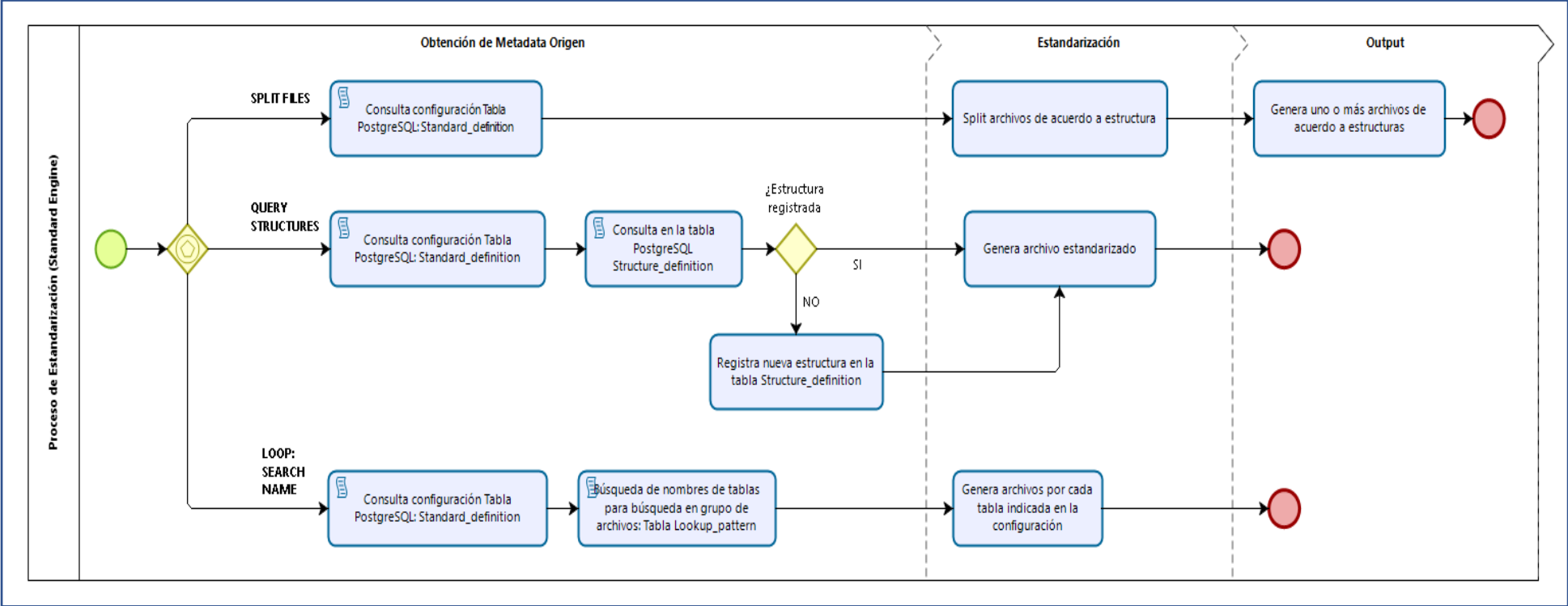


Figura 23 Diagrama de flujo: Proceso de estandarización
Fuente: elaboración propia

1.10 Funcionalidad

1.10.1 Inicio:

Para poder ingresar al servidor de Linux, se pedirá usuario y contraseña para poder hacer ejecuciones de funcionalidad.

Ip:	10.4.88.36
Puerto:	22
Usuario:	dev_nifi
Password:	xxxxxx

Tabla 28 Tabla de Credenciales

Se ingresa las credenciales

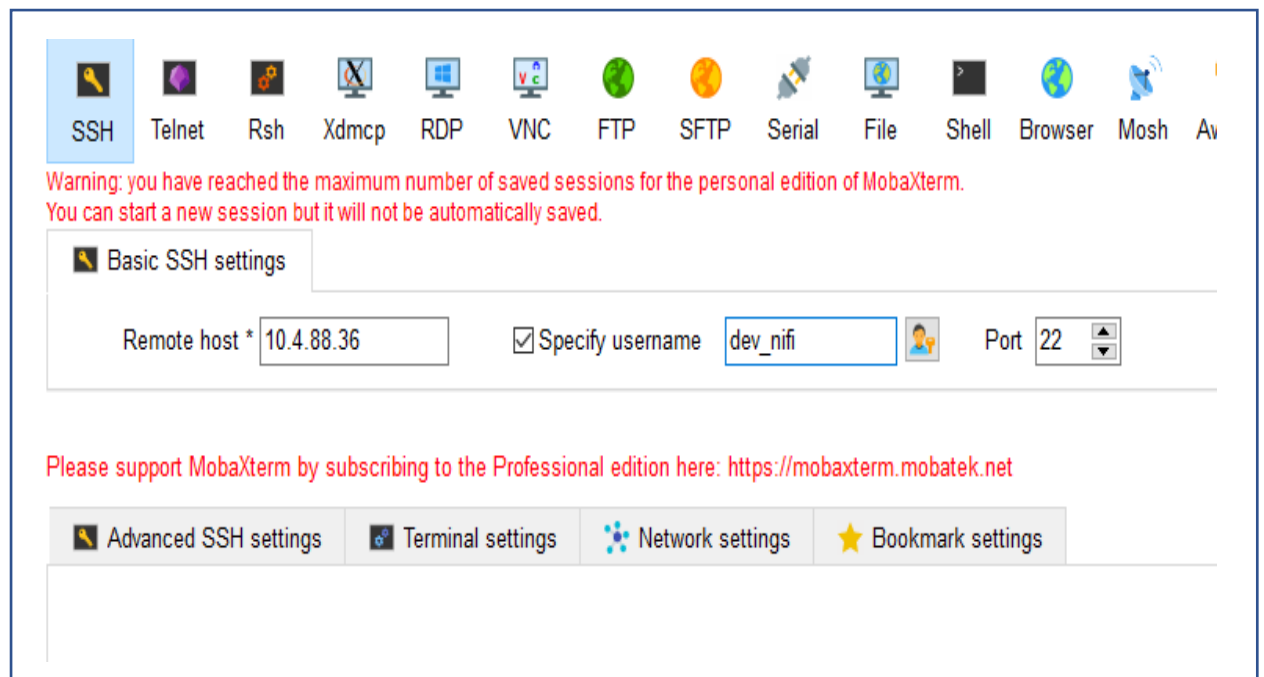
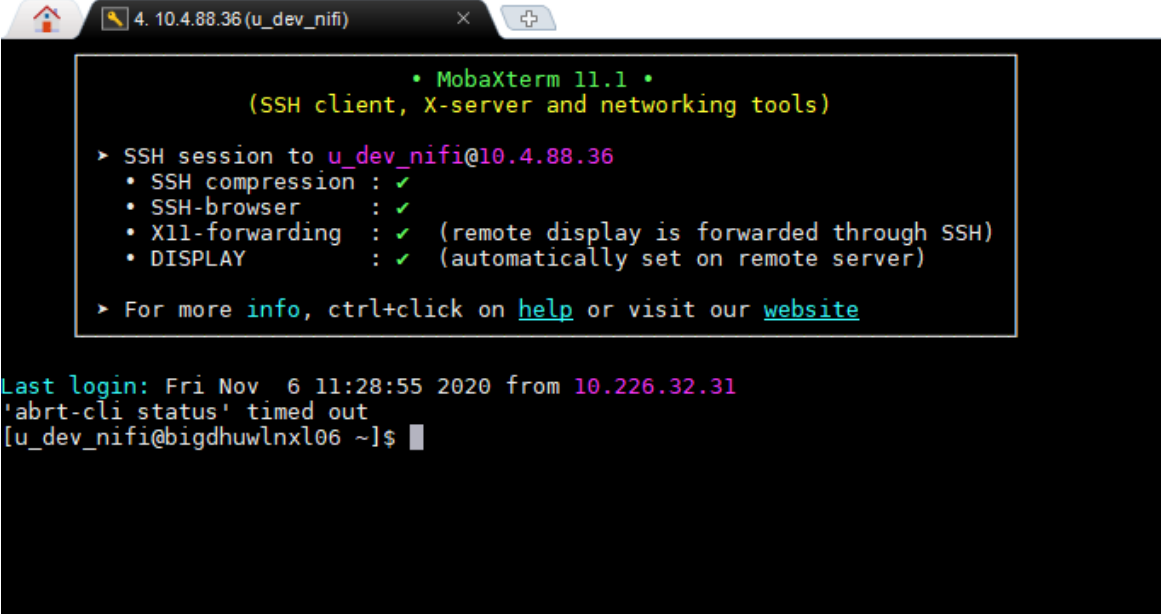


Figura 24 Credenciales entorno Linux

Ingresamos la contraseña, y nos muestra la siguiente pantalla



```
• MobaXterm 11.1 •
(SSSH client, X-server and networking tools)

▶ SSH session to u_dev_nifi@10.4.88.36
  • SSH compression : ✓
  • SSH-browser      : ✓
  • X11-forwarding  : ✓ (remote display is forwarded through SSH)
  • DISPLAY         : ✓ (automatically set on remote server)

▶ For more info, ctrl+click on help or visit our website

Last login: Fri Nov 6 11:28:55 2020 from 10.226.32.31
'abrt-cli status' timed out
[u_dev_nifi@bigdhuwlnxl06 ~]$
```

Figura 25 Entorno Linux

1.10.2 Muestra de componentes.

Se puede apreciar cómo se desplegó la información en un servidor Linux, aquí se encuentra toda la información de los procesos para la extracción, ingesta y transformación de los datos.

```

[lu_dev_nifi@bigdhuwlnx106 ~]$ ls -ltr
total 11023676
-rwxrwxrwx 1 u_dev_nifi g_dev_it      0 Jan  1  2012 newfile
-rwxrwxrwx 1 u_dev_nifi g_dev_it      0 Jan  1  2012 goldenfile
-rwxrwxrwx 1 u_dev_nifi g_dev_it 26473 Jul 19  2019 daas_prod_only_schemas_20190719_user_eaulloa.sql
-rwxrwxrwx 1 u_dev_nifi g_dev_it 164419 Aug 13  2019 hs_err_pid9005.log
drwxrwxrwx 2 u_dev_nifi g_dev_it      66 Aug 16  2019 FastData
drwxrwxrwx 2 u_dev_nifi g_dev_it      6 Aug 27  2019 jars
drwxrwxrwx 3 u_dev_nifi g_dev_it      32 Sep 14  2019 hdfs:
-rwxrwxrwx 1 u_dev_nifi g_dev_it 166310 Sep 15  2019 hs_err_pid22605.log
drwxrwxrwx 2 u_dev_nifi g_dev_it      6 Sep 21  2019 authorized keys
drwxrwxrwx 2 u_dev_nifi g_dev_it     125 Oct  1  2019 docs
-rwxrwxrwx 1 u_dev_nifi g_dev_it      0 Oct 14  2019 return
-rwxrwxrwx 1 u_dev_nifi g_dev_it      0 Oct 14  2019 select
-rwxrwxrwx 1 u_dev_nifi g_dev_it    2910 Oct 22  2019 tabla_general_configuration_20191022.sql
-rwxrwxrwx 1 u_dev_nifi g_dev_it   78089 Oct 22  2019 tablas_postgresql_20191022.sql
-rwxrwxrwx 1 u_dev_nifi g_dev_it      0 Nov  8  2019 to
drwxrwxrwx 2 u_dev_nifi g_dev_it     122 Nov 20  2019 .
-rwxrwxrwx 1 u_dev_nifi g_dev_it     194 Nov 21  2019 solution_productoffer.zip
drwxrwxrwx 7 u_dev_nifi g_dev_it     154 Nov 22  2019 solution_productoffe
-rwxrwxrwx 1 u_dev_nifi g_dev_it   23337 Nov 28  2019 execute_extract_ftp.py
-rwxrwxrwx 1 u_dev_nifi g_dev_it   39389 Dec 23  2019 execute_ingest_engine_1.py
-rwxrwxrwx 1 u_dev_nifi g_dev_it   19778 Jan 20  2020 standard_txt_parse_ubb_cbr.py
-rwxrwxrwx 1 u_dev_nifi g_dev_it 1280307 Jan 23  2020 Planta_consolidada_20200127.csv
-rw-r--r-- 1 u_dev_nifi g_dev_it      0 Jan 27  2020 org.apache.hadoop.hive.ql.io.orc.OrcSerde
-rw-r--r-- 1 u_dev_nifi g_dev_it     44 Jan 28  2020 script_hive.hql
-rwxrwxr-x 1 u_dev_nifi g_dev_it   661670 Feb  3  2020 Planta_consolidada_20200130.csv
-rw-r--r-- 1 u_dev_nifi g_dev_it     42 Feb  4  2020 script_jsp.hql

```

Figura 26 Listado de scripts

Hive: se encuentra toda la información de la ingesta de los archivos fuentes, se puede encontrar las tablas en fuentes, tablas temporales y tablas finales.

Se tienen 3 esquemas:

- Stage : se encuentra toda la data cruda
- Perm : se encuentra la data procesada y comprimida
- Delivery: tablas temporales y
- tablas finales

Schema	Tablas	Reportes
dev_stage_gt	ubb_show_docsis	Reporte de Saturación Reporte de Portadoras
dev_stage_gt	ubb_scmload_cisco_cbr	
dev_stage_gt	ubb_scmload_cisco_ubr	
dev_stage_gt	ubb_sum_casa	
dev_stage_gt	ubb_sum_cisco_cbr	
dev_stage_gt	ubb_scm_sum_olt_huawei	
dev_stage_gt	ubb_puertos_updown_casa	
dev_stage_gt	ubb_puerto_premium	
dev_stage_gt	ubb_puerto_cisco	
dev_stage_gt	ubb_intraway	
dev_stage_gt	ubb_cms_provincia	
dev_stage_gt	ubb_cms_lima	
dev_stage_gt	ubb_trafico_cablemodem	

Tabla 29 Esquema de Tablas

Ingresamos al esquema stage:

```

Beeline version 3.1.0.3.1.0.0-78 by Apache Hive
0: jdbc:hive2://bigdhwlnx106:2181,bigdhwlnx> show dev_stage_gt;
Error: Error while compiling statement: FAILED: ParseException line 1:5 cannot recognize input near 'show' 'dev_stage_gt' '<EO
>' in ddl statement (state=42000,code=40000)
0: jdbc:hive2://bigdhwlnx106:2181,bigdhwlnx> use dev_stage_gt;
INFO : Compiling command(queryId=hive_20201029233133_57e3498b-02f2-4366-8657-438a4ee21bdd): use dev_stage_gt
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO : Completed compiling command(queryId=hive_20201029233133_57e3498b-02f2-4366-8657-438a4ee21bdd); Time taken: 0.028 second
INFO : Executing command(queryId=hive_20201029233133_57e3498b-02f2-4366-8657-438a4ee21bdd): use dev_stage_gt
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20201029233133_57e3498b-02f2-4366-8657-438a4ee21bdd); Time taken: 0.015 seconds
INFO : OK

```

Figura 27 Esquema dev_stage_gt

Visualizamos las tablas que se encuentra dentro del esquema.

```
No rows affected (0.131 seconds)
0: jdbc:hive2://bigdhuwlnxl06:2181,bigdhuwlnx> showtables;
Error: Error while compiling statement: FAILED: ParseException line 1:0 cannot recognize input near
' (state=42000,code=40000)
0: jdbc:hive2://bigdhuwlnxl06:2181,bigdhuwlnx> show tables;
INFO : Compiling command(queryId=hive_20201029233225_8ff42623-cf93-4c4a-a6f2-c5ee96f7e1a3): show
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:tab_name, type:string, comment:
ties:null)
INFO : Completed compiling command(queryId=hive_20201029233225_8ff42623-cf93-4c4a-a6f2-c5ee96f7e1
s
INFO : Executing command(queryId=hive_20201029233225_8ff42623-cf93-4c4a-a6f2-c5ee96f7e1a3): show
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20201029233225_8ff42623-cf93-4c4a-a6f2-c5ee96f7e1
s
INFO : OK
+-----+-----+
|                tab_name                |
+-----+-----+
| cdr_cursado                             |
| cmactivosdiario                         |
| cmts_ip                                  |
| etiqueta_puertos                         |
| f00073_d_pmresult                       |
| f00073_pmresult                         |
| f00074_pmresult                         |
| f00086_pmresult                         |
| f00090_pmresult                         |
| f00091_pmresult                         |
| f00092_pmresult                         |
| glcifac041                               |
| kpipts_estaciones_bases                  |
| kpipts_transporte                       |
| kpipts_transporte_1                     |
| kpipts_transporte_2                     |
| marca_modelo_oui                        |
| movistar_total                           |
| movistar_total1                          |
+-----+-----+
```

Figura 28 Tablas de configuración dentro del esquema dev_stage_gt

Tablas de los archivos fuentes.

```
| reporte_firmware                        |
| scm_phy_cmts_casa                       |
| scm_phy_cmts_cisco                      |
| scm_phy_olt_huawei                       |
| scm_sum_cmts_casa                       |
| scm_sum_cmts_cisco                      |
| scm_sum_olt_huawei                       |
| scmload_load_cisco                      |
| scmoffline_cmts_casa                    |
| scmoffline_cmts_cisco                   |
| scmoffline_olt_huawei                    |
| subscriber                               |
| subscriberegionproduct                  |
```

Figura 28 Tablas de configuración

Lógica del negocio creación e inserción del cruce de tablas del Schema en dev_perm_gt.

```
create table ubb_docsis_prueba as (  
select *  
from ubb_show_docsis_casa_delivery);  
  
select  
nombre_cmts,  
count(nombre_cmts)  
from dev_delivery_gt.ubb_docsis_prueba  
group by nombre_cmts  
  
#####  
#####ubb_docsis_prueba copia de show docsis casa  
#####  
  
INSERT INTO TABLE dev_delivery_gt.ubb_show_docsis_casa_delivery PARTITION(fecha_carga)  
select  
SUBSTR(docsis.downstream,1,3) as troba,  
SUBSTR(docsis.puerto,1,3) as frecuencia,  
docsis.utilization as utilization,  
cata.nombre as nombre_cmts,  
cata.tipo_cmts as tipo_cmts,  
docsis.id as id_cmts,  
concat(cata.nombre,'-',SUBSTR(docsis.downstream,1,3)) as puerto,  
docsis.hora as hora,  
docsis.semana as semana,  
case when split(docsis.downstream,'\\\.')[1] is NULL then 'down' else 'up' end as estado,  
docsis.fecha as fecha_carga  
from dev_stage_gt.ubb_show_docsis_casa docsis  
inner join dev_perm_gt.ubb_catalogo_cmts cata on cata.id =docsis.id  
;
```

```
select  
t.nombre_cmts,  
t.tipo_cmts,  
t.id_cmts,t.troba,  
t.total,  
concat(t.nombre_cmts,'-',t.troba)as puerto,  
t.semana,t.fecha_carga  
from(  
select  
cata.nombre as nombre_cmts,  
cata.tipo_cmts as tipo_cmts,  
docsis.id as id_cmts,  
SUBSTR(docsis.downstream,1,3) as troba,  
count(SUBSTR(docsis.downstream,1,3)) as total,  
docsis.semana as semana,  
docsis.fecha as fecha_carga  
from dev_stage_gt.ubb_show_docsis_casa docsis  
inner join dev_perm_gt.ubb_catalogo_cmts cata  
on cata.id =docsis.id  
group by cata.nombre,cata.tipo_cmts,docsis.id  
order by SUBSTR(docsis.downstream,1,3)  
)t
```

Figura 29 Script create_table: ubb_docsis_prueba

Creación de una tabla en Hive, apuntando a una ruta de hdfs para que se almacene toda la información

```
CREATE TABLE `ubb_sum_casa_delivery` (  
  `puerto_up` string,  
  `puerto_down` string,  
  `trobas` string,  
  `consumido` string,  
  `nombre_cmts` string,  
  `id_cmts` string,  
  `hora` string,  
  `semana` string  
)  
PARTITIONED BY (  
  `fecha_carga` string)  
ROW FORMAT SERDE  
  'org.apache.hadoop.hive.ql.io.orc.OrcSerde'  
STORED AS INPUTFORMAT  
  'org.apache.hadoop.hive.ql.io.orc.OrcInputFormat'  
OUTPUTFORMAT  
  'org.apache.hadoop.hive.ql.io.orc.OrcOutputFormat'  
LOCATION  
  'hdfs://bigdhuwlnx102/dev/ubb_sum_casa_delivery/delivery'  
TBLPROPERTIES (  
  'bucketing_version'='2',  
  'discover.partitions'='true',  
  'external.table.purge'='true',  
  'transactional'='true',  
  'transactional_properties'='default',  
  'transient_lastDdlTime'='1578420146')
```

Figura 30 Creación de una tabla en HIVE

Hadoop:

Contiene todo el repositorio de la información, aquí se almacena toda la data que se registra por Hive.

Se puede apreciar todas las ingestas que se realizó de los planos transformados.

```
[u_dev_nifi@bigdhwlnx106 ~]$ hdfs dfs -ls /TEF/dev/UBB/
Found 97 items
drwxr-xr-x  - u_dev_nifi hdfs      0 2020-07-17 17:58 /TEF/dev/UBB/F00019_prueba_txt
drwxr-xr-x  - u_dev_nifi hdfs      0 2020-07-17 17:58 /TEF/dev/UBB/Inputs
drwxr-xr-x  - u_dev_nifi hdfs      0 2020-07-17 17:58 /TEF/dev/UBB/caidas_trobas
drwxr-xr-x  - u_dev_nifi hdfs      0 2020-07-17 18:20 /TEF/dev/UBB/cmactivosdiario
drwxr-xr-x  - u_dev_nifi hdfs      0 2020-07-17 18:21 /TEF/dev/UBB/cmts_cisco_ubr
drwxr-xr-x  - u_dev_nifi hdfs      0 2020-07-17 18:21 /TEF/dev/UBB/cmts_ip
drwxr-xr-x  - u_dev_nifi hdfs      0 2020-07-17 18:21 /TEF/dev/UBB/etiqueta_puertos
drwxr-xr-x  - u_dev_nifi hdfs      0 2020-07-17 18:21 /TEF/dev/UBB/glcifac041
drwxr-xr-x  - u_dev_nifi hdfs      0 2020-07-17 18:21 /TEF/dev/UBB/inputs
drwxr-xr-x  - u_dev_nifi hdfs      0 2020-07-17 18:21 /TEF/dev/UBB/marca_modelo_docsis
drwxr-xr-x  - u_dev_nifi hdfs      0 2020-07-17 18:21 /TEF/dev/UBB/marca_modelo_oui
drwxr-xr-x  - u_dev_nifi hdfs      0 2020-07-17 18:21 /TEF/dev/UBB/movistar_total
drwxr-xr-x  - u_dev_nifi hdfs      0 2020-07-17 18:21 /TEF/dev/UBB/prueba_avro
drwxr-xr-x  - u_dev_nifi hdfs      0 2020-07-17 18:21 /TEF/dev/UBB/prueba_json
drwxr-xr-x  - u_dev_nifi hdfs      0 2020-07-17 18:27 /TEF/dev/UBB/reporte_endpoints
drwxr-xr-x  - u_dev_nifi hdfs      0 2020-07-17 18:27 /TEF/dev/UBB/reporte_firmware
drwxr-xr-x  - u_dev_nifi hdfs      0 2020-07-17 18:27 /TEF/dev/UBB/scm_phy_cmts_casa
drwxr-xr-x  - u_dev_nifi hdfs      0 2020-07-17 18:27 /TEF/dev/UBB/scm_phy_cmts_cisco
drwxr-xr-x  - u_dev_nifi hdfs      0 2020-07-17 18:27 /TEF/dev/UBB/scm_phy_hist
drwxr-xr-x  - u_dev_nifi hdfs      0 2020-07-17 18:27 /TEF/dev/UBB/scm_phy_olt_huawei
drwxr-xr-x  - u_dev_nifi hdfs      0 2020-07-17 18:27 /TEF/dev/UBB/scm_sum_cmts
drwxr-xr-x  - u_dev_nifi hdfs      0 2020-07-17 18:27 /TEF/dev/UBB/scm_sum_cmts_casa
drwxr-xr-x  - u_dev_nifi hdfs      0 2020-07-17 18:27 /TEF/dev/UBB/scm_sum_cmts_cisco
drwxr-xr-x  - u_dev_nifi hdfs      0 2020-07-17 18:27 /TEF/dev/UBB/scm_sum_olt_huawei
drwxr-xr-x  - u_dev_nifi hdfs      0 2020-07-17 18:28 /TEF/dev/UBB/scmload_load_cisco
drwxr-xr-x  - u_dev_nifi hdfs      0 2020-07-17 18:28 /TEF/dev/UBB/scmoffline_cmts_casa
drwxr-xr-x  - u_dev_nifi hdfs      0 2020-07-17 18:28 /TEF/dev/UBB/scmoffline_cmts_cisco
drwxr-xr-x  - u_dev_nifi hdfs      0 2020-07-17 18:28 /TEF/dev/UBB/scmoffline_olt_huawei
drwxr-xr-x  - u_dev_nifi hdfs      0 2020-07-17 18:28 /TEF/dev/UBB/trabajos_programados
drwxr-xr-x  - u_dev_nifi hdfs      0 2020-07-17 18:28 /TEF/dev/UBB/ubb_ancho_banda_hfc_delivery
```

Figura 31 Listado de planos del entorno HDFS - Hadoop

Carpetas de validación de del proceso de archivo plano.

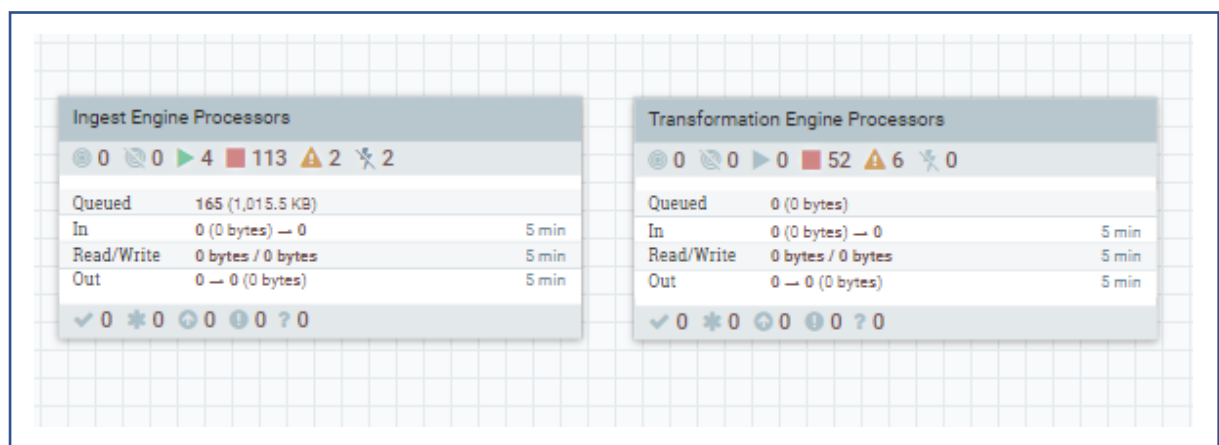
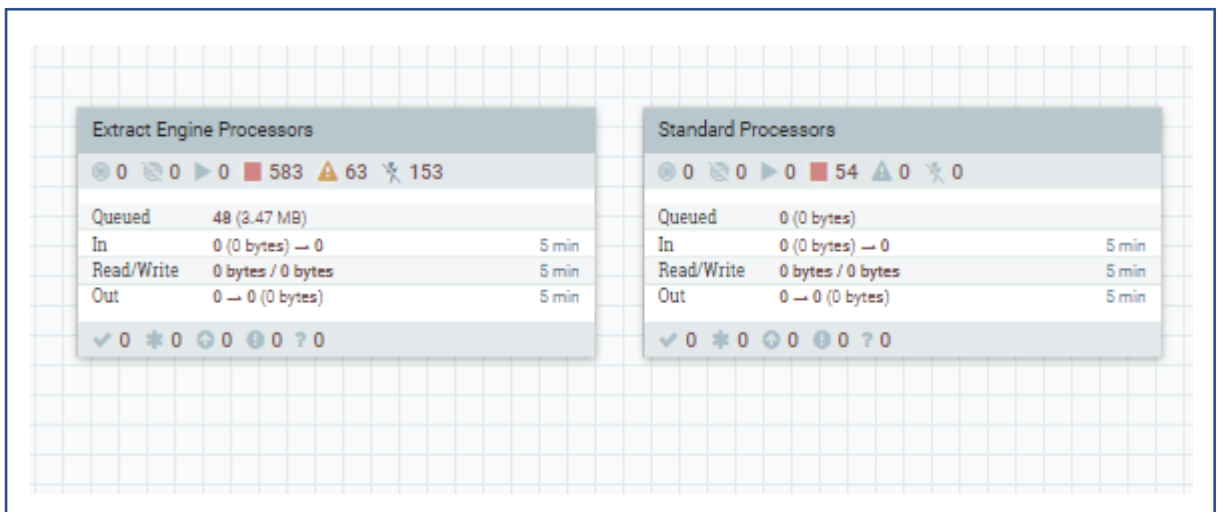
```
[u_dev_nifi@bigdhuwlnxl06 ~]$ hdfs dfs -ls /TEF/dev/UBB/ubb_show_docsis_casa
Found 7 items
drwxr-xr-x - u_dev_nifi hdfs      0 2020-07-17 18:29 /TEF/dev/UBB/ubb_show_docsis_casa/export
drwxr-xr-x - u_dev_nifi hdfs      0 2020-07-17 18:29 /TEF/dev/UBB/ubb_show_docsis_casa/invalid
drwxr-xr-x - u_dev_nifi hdfs      0 2020-07-17 18:29 /TEF/dev/UBB/ubb_show_docsis_casa/original
drwxr-xr-x - u_dev_nifi hdfs      0 2020-07-17 18:29 /TEF/dev/UBB/ubb_show_docsis_casa/perm
drwxr-xr-x - u_dev_nifi hdfs      0 2020-07-17 18:29 /TEF/dev/UBB/ubb_show_docsis_casa/processed
drwxr-xr-x - u_dev_nifi hdfs      0 2020-07-17 18:29 /TEF/dev/UBB/ubb_show_docsis_casa/processing
drwxr-xr-x - u_dev_nifi hdfs      0 2020-07-17 18:29 /TEF/dev/UBB/ubb_show_docsis_casa/stage
```

Figura 32 Carpetas de validación de del proceso de archivo plano

NIFI

Apache NIFI se encarga de ejecutar todos los procesos desarrollados en Python.

En NIFI se tiene configurado los procesos ETL.



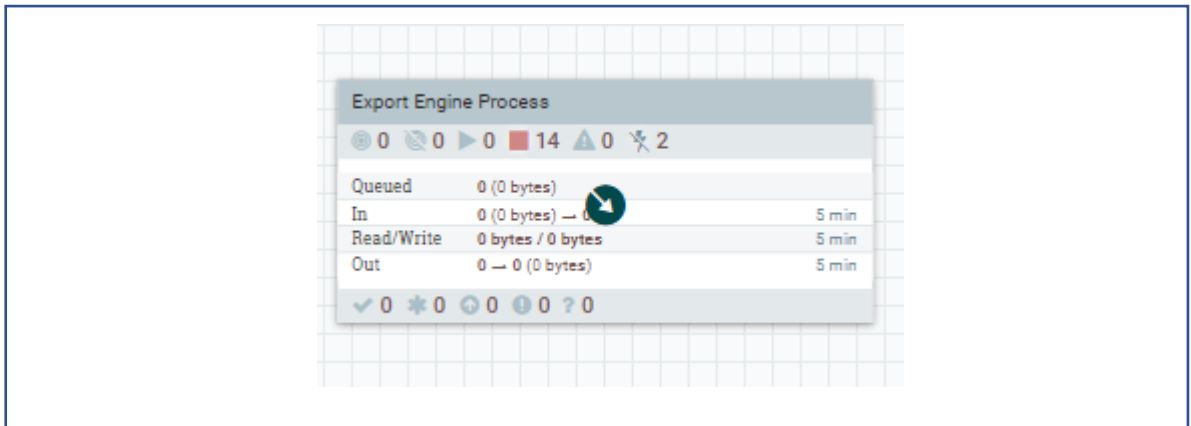


Figura 33 NIFI Entorno de los process group

Para la extracción de los planos se tiene configurado los siguientes process group

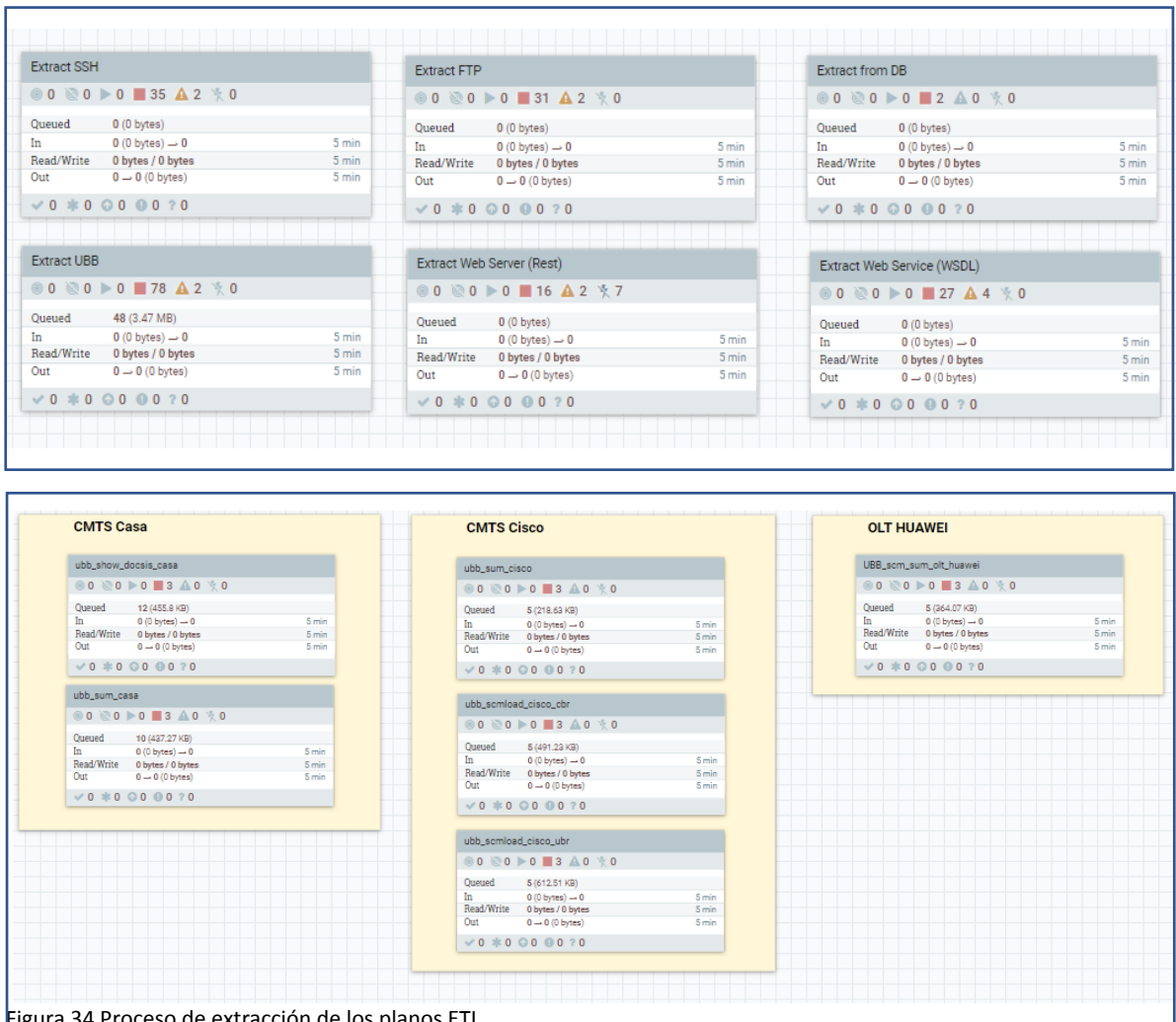


Figura 34 Proceso de extracción de los planos ETL

Ejemplo de proceso de extracción por FTP:

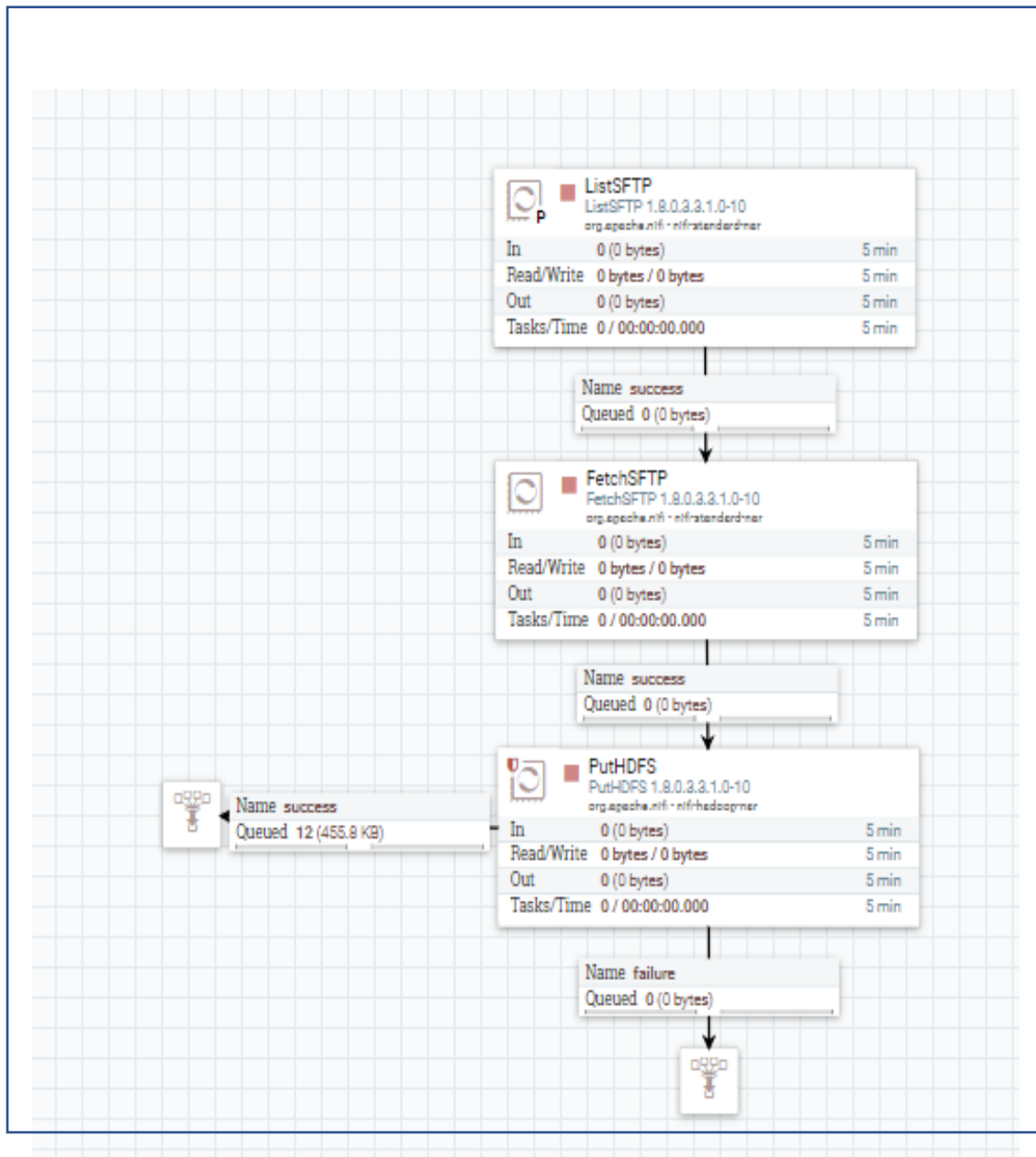


Figura 35 proceso de extracción por FTP

Python

Se encuentra todo el desarrollo, toda la lógica de la programación de la conexión de los servicios, extracción de los componentes y de otros sistemas integrados.

Extracción:

```
# fetch_data: Function to retrieve data after executing a query in PostgreSQL
def fetch_data(query):
    conn, cur = connect()
    try:
        cur.execute(query)
        conn.commit()
    except Exception as e:
        logging.error('-----')
        logging.error('No se logró ejecutar el siguiente query: ')
        logging.error(query)
        logging.error('Detalle: ')
        logging.error(e)
        logging.error('-----')
    data = cur.fetchall()
    return data

# convert fotmat date extracted from metadata to valid value for python (used for ftp file name)
def file_format_time(format_original):
    if format_original == "":
        format_out = ""
    elif format_original == "YYYY":
        format_out = "-" + datetime.datetime.now().strftime("%Y")
    elif format_original == "YYYYMM":
        format_out = datetime.datetime.now().strftime("%Y%m")
    elif format_original == "YYYYMMDD":
        format_out = "-" + datetime.datetime.now().strftime("%Y%m%d")
    elif format_original == "YYYYMMDDhh":
        format_out = "-" + datetime.datetime.now().strftime("%Y%m%d%H")
    elif format_original == "YYYYMMDDhhmm":
        format_out = "-" + datetime.datetime.now().strftime("%Y%m%d%H%M")
    elif format_original == "YYYYMMDDhhmss":
        format_out = "-" + datetime.datetime.now().strftime("%Y%m%d%H%M%S")
    elif format_original == "YYYYMMDDhhmssff":
        format_out = "-" + datetime.datetime.now().strftime("%Y%m%d%H%M%S%f")
    else:
        process_time = datetime.datetime.now().strftime("%Y%m%d%H%M")

    ### DEFINE LOG CONFIGURATION ###
    # Creating directory log for process
    path_log_process = BASE_PATH_LOGS + p_source_name
    print(path_log_process)
    try:
        # Si no existe la carpeta, debemos crearla y asignarle permisos de lectura/escritura
        if not os.path.exists(path_log_process):
            os.makedirs(path_log_process, 0o775)
            #run_cmd(["chmod -775", path_log_process])
            print("Se creó el directorio: " + path_log_process)
    except OSError as error:
        print("No se logró crear el directorio : " + path_log_process)

    # Init log file
    engine = 'engineExtract_'
    file_log = path_log_current_date + engine + p_source_name + '_' + process_time + '.log'
    print(file_log)

    # Resetear los handlers del logging
    for handler in logging.root.handlers[:]:
        logging.root.removeHandler(handler)
    logging.basicConfig(filename=file_log, filemode="w", level=logging.DEBUG, format='%(asctime)s %(message)s',
                        datefmt='%d/%m/%Y %H:%M:%S')
    logger = logging.getLogger()
```

Figura 36 Desarrollo en Python de la extracción de los datos

Ingesta: Proceso obtiene el plano

```
#connect: Function to connect to PostgreSQL DB and read metadata of Ingest Engine
def connect():
    try:
        conn = pg8000.connect(database=db,
                              user=user_db,
                              password=pass_db)
        cur = conn.cursor()
    except (Exception, pg8000.DatabaseError) as error:
        print_with_logging("Error al intentar conectar con la base de datos PostgreSQL"+ error, 'error')
    return conn, cur

#fetch_data: Function to retrieve data after executing a query in PostgreSQL
def fetch_data(query):
    conn, cur = connect()
    try:
        cur.execute(query)
        conn.commit()
    except Exception as e:
        print_with_logging(repeat('-', 150), 'error')
        print_with_logging('No se logró ejecutar el siguiente query: ', 'error')
        print_with_logging(query, 'error')
        print_with_logging('Detalle: ', 'error')
        print_with_logging(e, 'error')
        print_with_logging(repeat_character('-', 100), 'error')
    data = cur.fetchall()
    return data

#hdfs_read: Function to read a file from HDFS into memory
def hdfs_read(fpath, text=False):
    cmd = '-text' if text else '-cat'
    proc = subprocess.Popen(['hadoop', 'fs', cmd, fpath],
                              stdout=subprocess.PIPE,
                              stderr=subprocess.PIPE,
                              universal_newlines=False)
    output, error = proc.communicate()
    if proc.returncode != 0:
        # Get parameter configured in table traceability.source_definition for process
        query="SELECT source_name, origin, hdfs_path, delimiter, File_type, load_type, control_file, header, partition_c
        print_with_logging(query, 'info')
    try:
        data_source_definition=fetch_data(query)
        print_data(data_source_definition)
        for row in data_source_definition:
            source_name=row[0]
            hdfs_path=row[2]
            delimiter=row[3]
            file_type=row[4]
            load_type=row[5]
            control_file=row[6]
            header=row[7]
            partition_column=row[8]
            file_name_stored=row[9]
            purge_retention=row[10]
            ingestion_frequency=row[11]
            priority=row[12]
            project_name=row[14]
        print_with_logging('Se obtuvo correctamente la información de traceability.source_definition para el source_
    except Exception as e:
        print_with_logging('No se logró obtener los datos de la tabla traceability.source_definition para el proceso
        print_with_logging('Detalle: '+e)
        sys.exit(1)

### GET AND READ FILE FROM HDFS ###
try:
    input_hdfs_path = BASE_PATH_INPUT_HDFS+p_source_name #GT_INPUT_HDFS_PATH+p_source_name
    filename=GetFirstElement_HDFSDirectory(input_hdfs_path)
except Exception as e:
    print_with_logging('No se logró obtener el archivo a procesar en el directorio HDFS: '+hdfs_path, 'error')
    print_with_logging('El directorio HDFS configurado en la tabla traceability.source_definition no existe, por
    print_with_logging("Detalle: "+e)
    sys.exit(1)
```

Figura 37 Desarrollo del proceso de ingesta

1.10.3 Esquema del desarrollo.

Se puede apreciar de una forma funcional todo el proceso de desarrollo de los componentes.

Flujo del proceso completo.

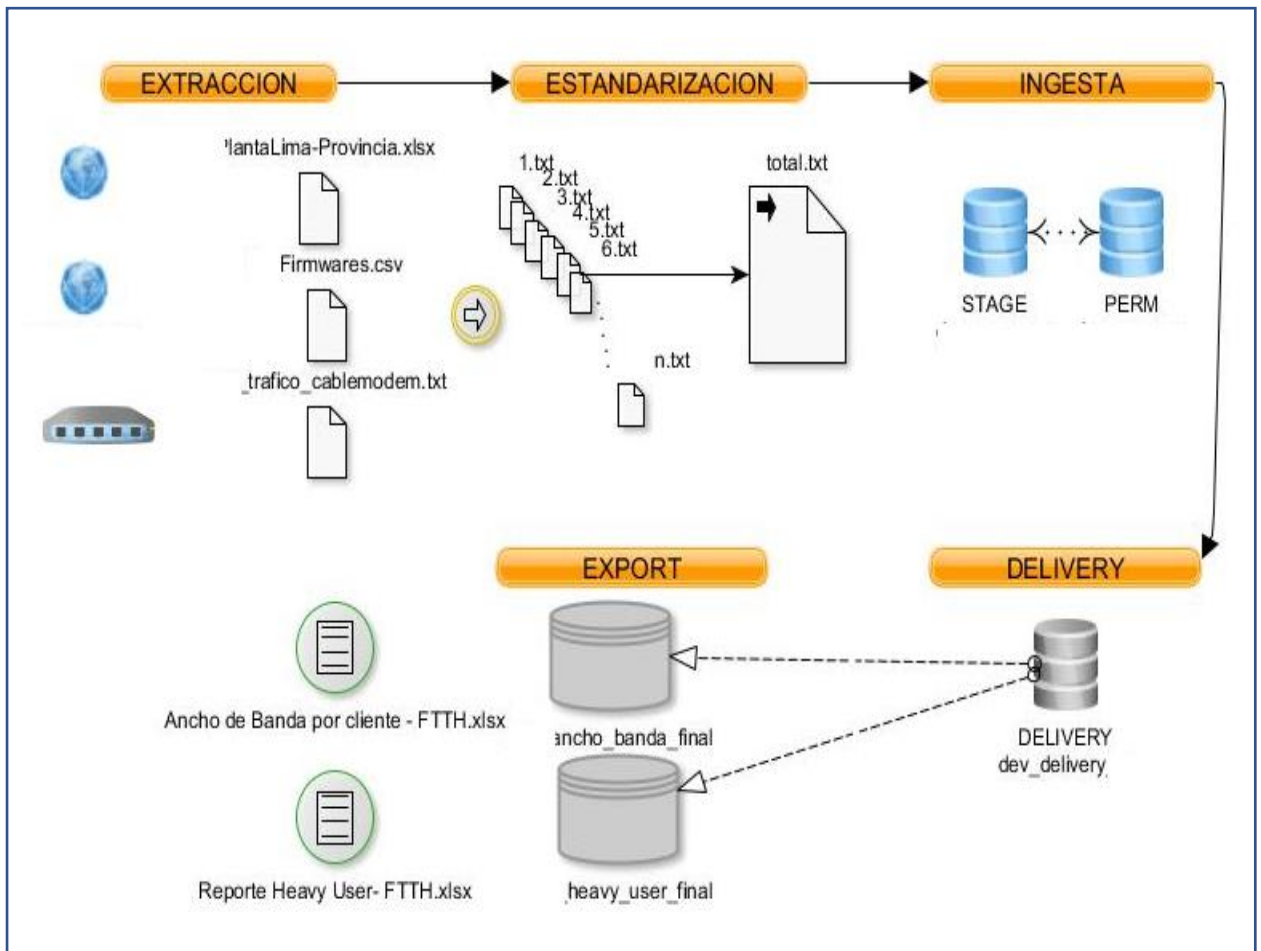


Figura 38 Flujo del proceso de desarrollo de la arquitectura completa.

Flujo del proceso de estandarización hacia la ingesta de la data en las tablas

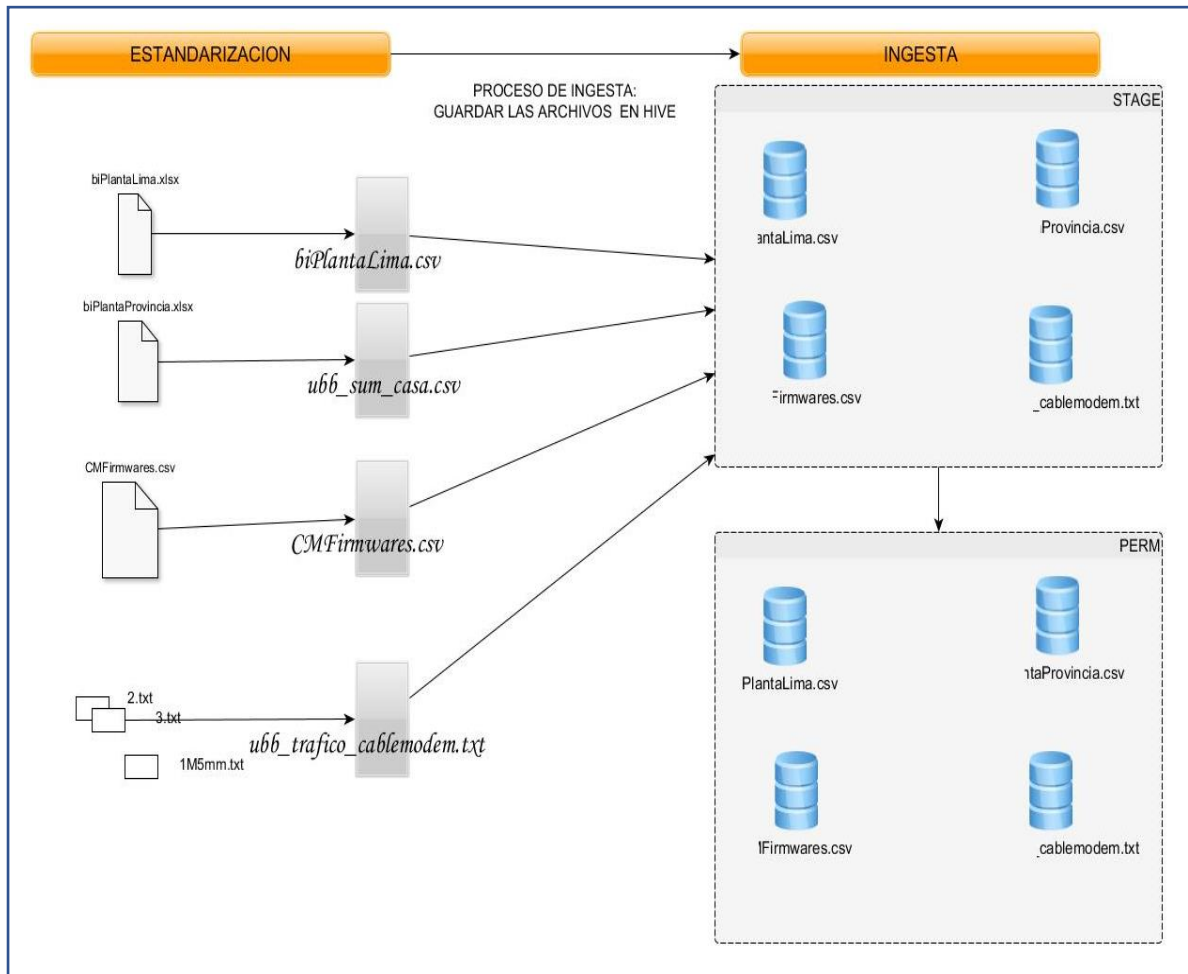


Figura 39 Flujo del proceso de estandarización hacia la ingesta de la data en las tablas

Proceso de exportación y obtención de los reportes finales.

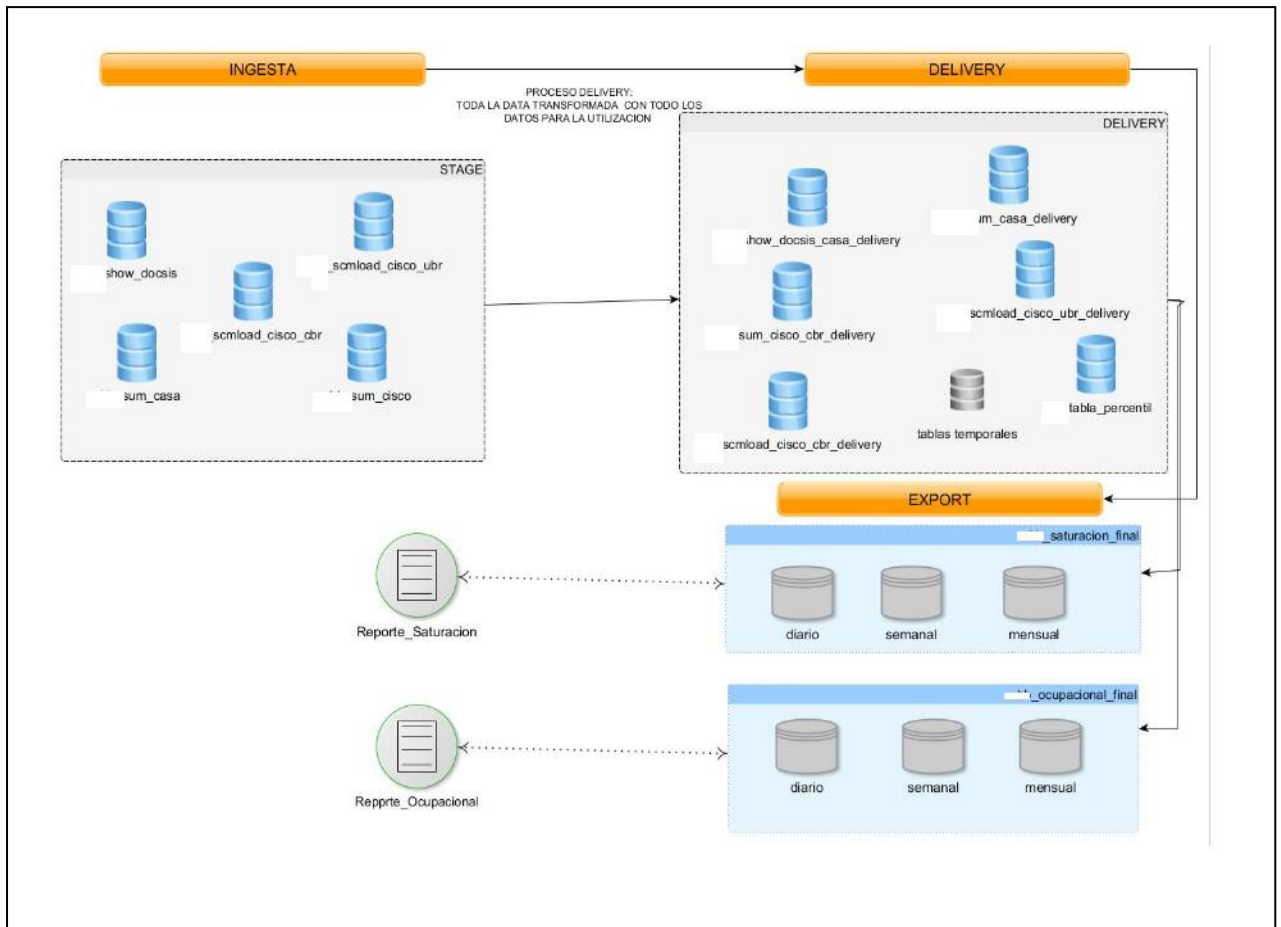


Figura 40 Proceso de exportación y obtención de los reportes finales.

1.10.4 Muestra de los reportes finales

Resultado de los reportes obtenidos.

REPORTE DE SATURACION				
ESTADO	NACIONAL	Premium	Alto Valor	Masivo
Sat >90	18	1	1	16
Sat 81-90	96	1	10	85
Sat 71-80	334	17	43	274
Sat <70	3765	420	523	2822
TOTAL	4213	439	577	3197

CLIENTES				
ESTADO	NACIONAL	Premium	Alto Valor	Masivo
Sat >90	11260	429	1315	9516
Centralizado	3384	429	1315	1640
Distribuido	7876	0	0	7876

Cientes	1391636
Centralizado	842552
Distribuido	549084

TROBAS CENTRALIZADO				
ESTADO	NACIONAL	Premium	Alto Valor	Masivo
Sat >90	4	1	1	2
Sat 81-90	59	1	8	50
Sat 71-80	211	14	38	159
Sat 61-70	373	48	66	259
Sat 51-60	422	106	78	238
Sat <50	801	163	124	514
TOTAL	1870	333	315	1222

Figura 41 Resultado de los reportes finales obtenidos

Sem35	Sem36	Sem37	Sem38	Sem39	Sem40	Sem41	Sem42	Sem43	Sem44	Sem45	Sem46	Sem47	Sem48	Sem49	Sem50	Sem51	Sem52	Sem01	Sem02
4.84%	4.37%	4.00%	3.77%	3.35%	3.35%	5.98%	3.79%	2.88%	4.11%	2.53%	1.77%	1.94%	1.72%	1.55%	1.55%	1.55%	1.55%	1.55%	1.55%
1.69%	1.19%	1.07%	1.04%	1.03%	1.03%	1.62%	0.98%	0.71%	1.11%	0.57%	0.52%	0.66%	0.49%	0.47%	0.47%	0.47%	0.47%	0.47%	0.47%

Sem35	Sem36	Sem37	Sem38	Sem39	Sem40	Sem41	Sem42	Sem43	Sem44	Sem45	Sem46	Sem47	Sem48	Sem49	Sem50	Sem51	Sem52	Sem01	Sem02
1304050	1304050	1304050	1304050	1304050	1304050	1304050	1304050	1304050	1304050	1304050	1331959	1388048	1400390	1406756	1415335	1423726	1432810	1386681	1391636
46089	32383	29977	25978	24427	23929	43649	25734	18908	28546	15405	13839	18514	14442	13469	7794	2842	2955	15489	11260
3.53%	2.48%	2.30%	1.99%	1.87%	1.83%	3.35%	1.97%	1.45%	2.19%	1.18%	1.04%	1.33%	1.03%	0.96%	0.55%	0.20%	0.21%	1.12%	0.81%

Figura 42 Tabla de KPI de resultados por semana en un porcentaje de percentil

	Puerto	Vendor	Tipo1	Tipo2	Portador	clientes	Anterior	Actual
2	arequipa4-1/5	casa	Masivo	No	36	239	Sat_<50	Sat_<50
3	arequipa4-1/4	casa	Alto Valor	Sí	48	292	Sat_51-60	Sat_51-60
4	arequipa4-1/3	casa	Masivo	No	32	292	Sat_<50	Sat_<50
5	arequipa4-1/2	casa	Alto Valor	Sí	32	479	Sat_61-70	Sat_61-70
6	arequipa4-1/1	casa	Masivo	No	32	194	Sat_<50	Sat_<50
7	arequipa4-1/0	casa	Masivo	No	36	372	Sat_61-70	Sat_61-70
8	arequipa4-1/6	casa	Masivo	No	32	466	Sat_71-80	Sat_71-80
9	arequipa4-1/7	casa	Alto Valor	Sí	32	280	Sat_<50	Sat_<50
0	arequipa4-2/0	casa	Alto Valor	Sí	32	327	Sat_51-60	Sat_61-70
1	arequipa4-2/1	casa	Alto Valor	Sí	32	384	Sat_<50	Sat_<50
2	arequipa4-2/2	casa	Alto Valor	Sí	32	161	Sat_<50	Sat_<50
3	arequipa4-2/3	casa	Premium	Sí	32	409	Sat_51-60	Sat_51-60
4	arequipa4-2/4	casa	Alto Valor	Sí	32	303	Sat_<50	Sat_<50
5	arequipa4-2/5	casa	Premium	Sí	32	304	Sat_<50	Sat_<50
6	arequipa4-2/6	casa	Premium	Sí	32	394	Sat_51-60	Sat_51-60
7	arequipa4-2/7	casa	Alto Valor	Sí	32	253	Sat_51-60	Sat_51-60
8	arequipa4-3/0	casa	Alto Valor	Sí	36	217	Sat_71-80	Sat_71-80
9	arequipa4-3/1	casa	Masivo	No	32	357	Sat_51-60	Sat_61-70
0	arequipa4-3/2	casa	Masivo	No	48	516	Sat_<50	Sat_<50
1	arequipa4-3/3	casa	Masivo	No	48	471	Sat_61-70	Sat_61-70
2	arequipa4-3/4	casa	Masivo	No	36	279	Sat_51-60	Sat_51-60
3	arequipa4-3/5	casa	Masivo	No	36	1	Sat_<50	Sat_<50
4	arequipa4-3/6	casa	Alto Valor	Sí	32	365	Sat_51-60	Sat_51-60
5	arequipa4-3/7	casa	Alto Valor	Sí	32	345	Sat_51-60	Sat_51-60

Figura 43 Estructura del reporte final de saturación de puerto

Anexo 10 Tablero de mando basado en Linux

1. INTRODUCCIÓN

a. OBJETIVO

Es demostrar la conformidad del software cumpliendo todos los requerimientos que se indicó en el desplégue inicial, cumpliendo estándares de rendimientos y la satisfacción de los usuarios al obtener los reportes finales. Ayudándolos en la optimización de tiempos y el control para tomar buenas decisiones.

b. PRODUCTO PROBADO

El entorno del software se utiliza apache NIFI, para el control y seguimientos de los procesos de cada flowFile, se debe tener en cuenta para el control del despliegue se desarrolla en cada processGroup para poder tener un mejor seguimiento de los proceso desarrollados

c. Entorno de Prueba

Para este proceso se muestra a continuación el despliegue y funcionamiento de cada proceso que se ejecuta dentro del desarrollo. Desde la parte inicial y el proceso final de la obtención del plano para el área correspondiente

Servidor: <http://10.3.88.47:8080/nifi>

Para las presentes pruebas fue requerido el siguiente software y herramientas:

#	Software o Herramienta	Descripción
1	Sistema Operativo	Windows 10
2	Servidor Apache NIFI	NIFI

Tabla 30 Entorno de software y herramientas

PROCESO DE CARGA FA_TRAFICO:

FLUJO NIFI:

Se encuentran todos los procesos de extracción, estandarización, ingesta y exportación de los procesos que cuentan nuestro tablero de mando.

<http://9090/nifi/?processGroupId=bd72621c-1181-35d4-aac5-d5c2f3a04d1f&componentIds=>

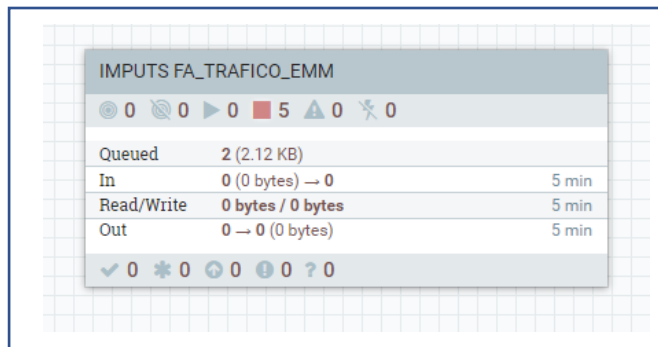


Figura 44 Process group de prueba

EXTRACION DEL PLANO:

Proceso de extracción del plano para que nuestro tablero de mando lo pueda consumir, se debe conectar con las siguientes credenciales al servidor linux y extraer los planos de las fechas específicas para que nuestro proceso NIFI lo pueda ingerir y exportar.

ip: 10.6.77.123

user: user20

passs: xxxxxx

Ruta: /sw/PG_SW_DESA/ShellsTemp

Proceso Fechas para extraer la información:

FECHA
2014
2015
2016
2017
2018

Tabla 31 configuración de fechas de extracción

Se deja un archivo plano programado de las fechas a extraer:

```
sh: Spool_Proceso_ I4AUXBAN.sh  
Param1: I4AUXBAN  
Param 2: D  
Param 3: /sw/desa_data/regulatorio/inputs /entrada
```

El flujo inicia de la siguiente manera:

Se activa el siguiente procesador, Se genera el nombre del archivo para extraer de la ruta específica.

Se crea la variable para poner el nombre del plano a buscar.

Procesador ListSFTP (Conectarse a una ruita ftp para extraer el plano para el proceso):

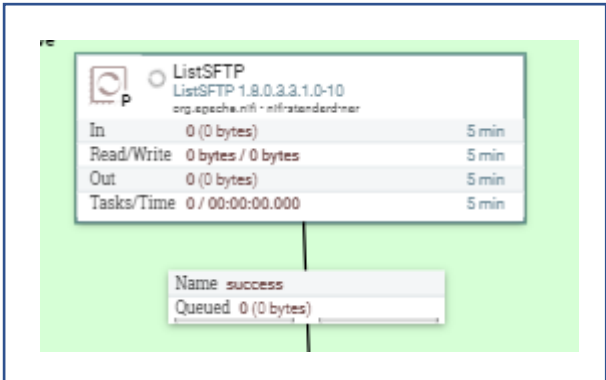
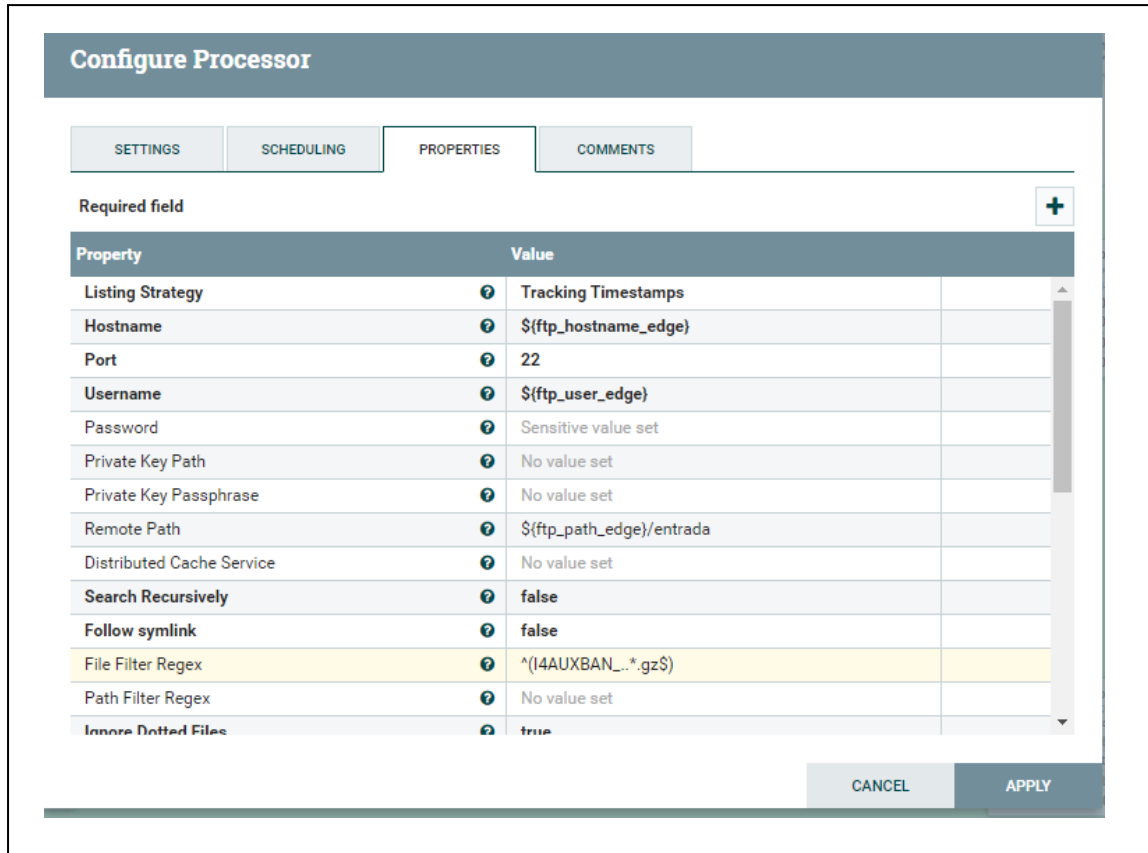


Figura 45 Conexión del Proceso del ListSFTP

Propiedades Procesador ListSFTP:



Configure Processor

SETTINGS SCHEDULING **PROPERTIES** COMMENTS

Required field +

Property	Value
Listing Strategy	Tracking Timestamps
Hostname	\${ftp_hostname_edge}
Port	22
Username	\${ftp_user_edge}
Password	Sensitive value set
Private Key Path	No value set
Private Key Passphrase	No value set
Remote Path	\${ftp_path_edge}/entrada
Distributed Cache Service	No value set
Search Recursively	false
Follow symlink	false
File Filter Regex	^(14AUXBAN_*.gz\$)
Path Filter Regex	No value set
Ignore Dotted Files	true

CANCEL APPLY

Figura 46 Configuración del proceso ListSFTP

Ejecutamos el siguiente proceso para poder descomprimir el archivo plano:
DescomprimirArchivoEdge

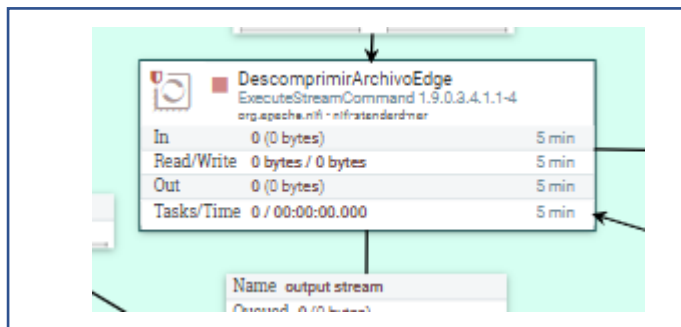


Figura 47 Proceso DescomprimirArchivoEdge

Descomprime un archivo plano .gz, para que luego se pueda mover a una ruta HDFS

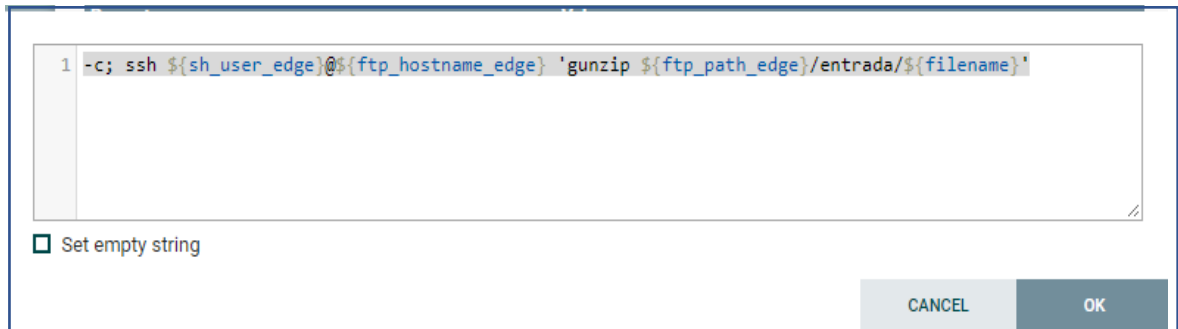


Figura 48 Proceso de configuración de variables para descomprimir plano de una ruta específica

Posteriormente mediante el siguiente procesador, se procede a actualizar la extensión del archivo plano a procesar.

Procesador Update Attribute(Actualizar parámetros):

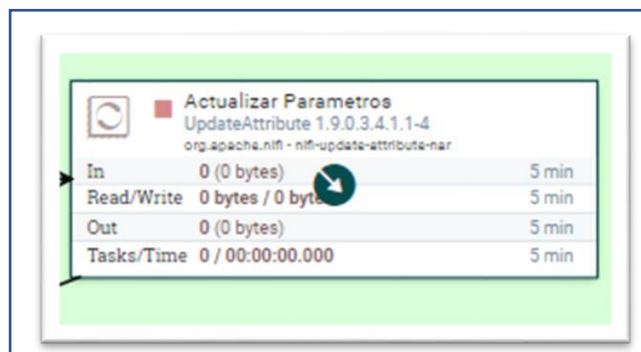


Figura 49 Proceso de Actualización de parámetros

Se verifica que se aplica la extensión el proceso

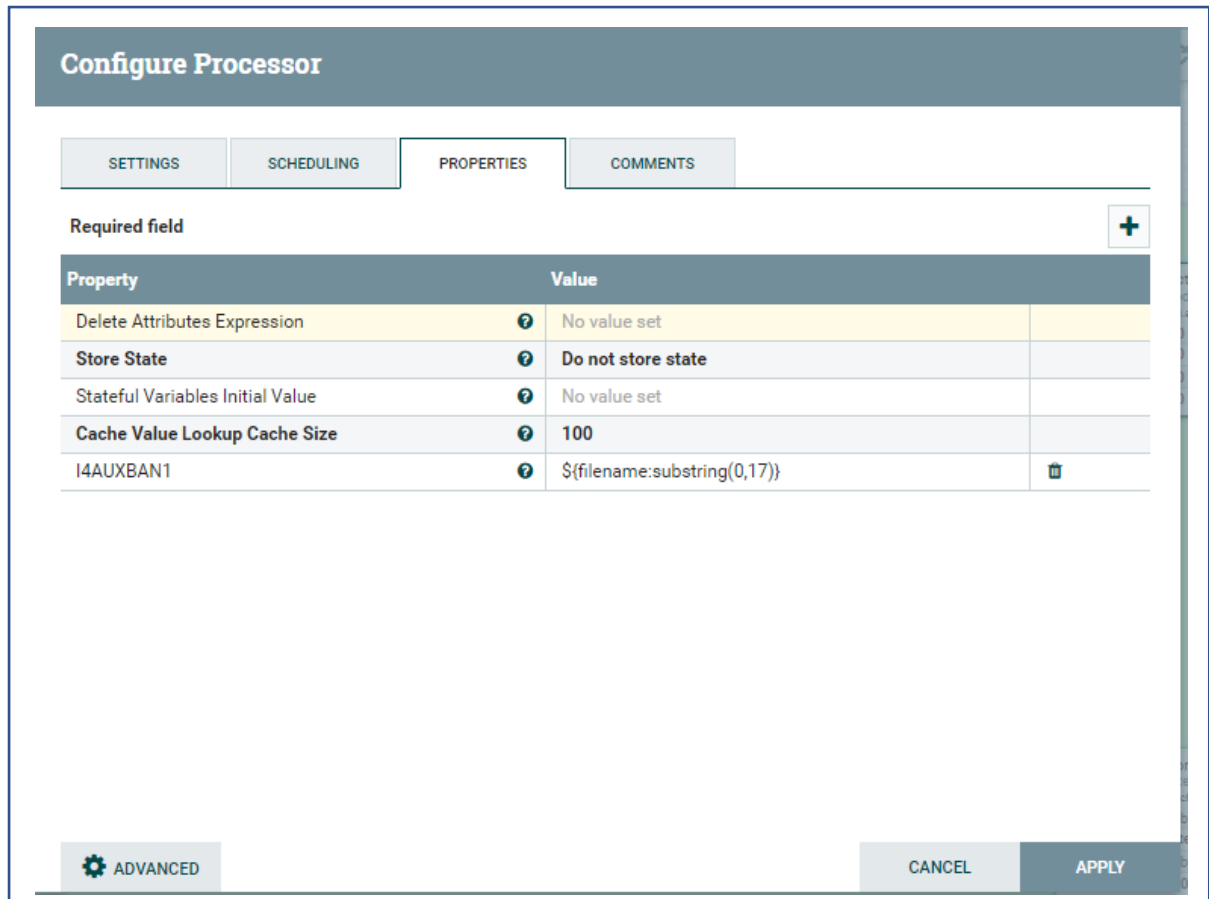


Figura 50 Configuración de procesos de actualización de parámetros

Como resultado de este paso se tiene el archivo para este caso I4AUXBAN_20140904.TXT en el path indicado en la imagen superior.

El próximo procesador estará enviando los archivos obtenidos a una ruta de HDFS.

Detalles Procesador ExecuteStreamCommand(Cargar Archivo_hdfs):

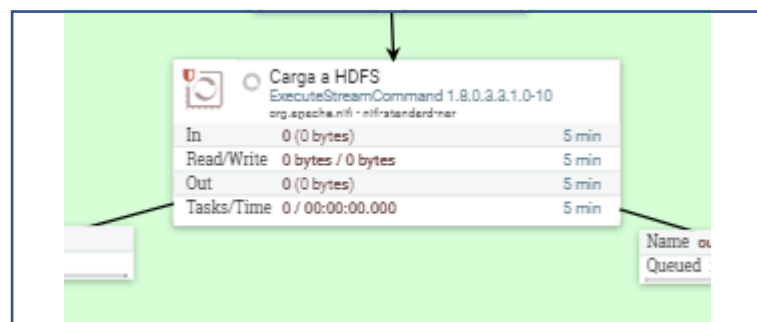


Figura 51 Proceso de carga de los planos hacia la ruta HFS

Configure Processor

SETTINGS
SCHEDULING
PROPERTIES
COMMENTS

Required field +

Property	Value
Command Arguments	-c; ssh \${sh_user_edge}@\${ftp_hostname_edge} \${path_...
Command Path	sh
Ignore STDIN	false
Working Directory	No value set
Argument Delimiter	;
Output Destination Attribute	No value set
Max Attribute Length	256

CANCEL
APPLY

Figura 52 Configuración del proceso para mover el plano hacia una ruta HFS

Como resultado de este paso se tiene el archivo para este caso I4AUXBAN_20140904.TXT en el path indicado en la imagen superior.

Archivo en HDFS:

```

-rw-r--r-- 1 u_dev_nifi_g_dev_it 1548760 Aug 11 15:19 I4AUXBAN_20140904.TXT

```

Figura 53 Proceso del plano HFS

Por último, se ejecutan en paralelo 2 procesos:

1. Inserción a Hive
2. Eliminar el archivo en el filesystem

1. Insercion a Hive:

En el procesador ExecuteStreamCommand se lanza la Shell de nombre “sh_carga_I4AUXBAN.sh”, esta tiene como objetivo sobrescribir la tabla final en hive “. I4AUXBAN” con nueva información de los archivos planos recolectados.

Rutas a utilizar:

Aquí se almacena la Shell: "sh_carga_I4AUXBAN.sh".

```
-rwxr-xr-x 1 u_dev_nifi g_dev_it 2984 Aug 11 17:07 sh_concat_regulator10_RLPR0.sh
-rwxrwxrwx 1 u_dev_nifi g_dev_it 3630 Aug 11 19:32 sh_hdfs_I4AUXBAN.sh
-rwxrwxrwx 1 u_dev_nifi g_dev_it 4981 Aug 12 17:08 sh_carga_I4AUXBAN.sh
```

Figura 54 Shells del proceso de almacenamiento

Procesador ExecuteStreamCommand(Poblado en Hive):

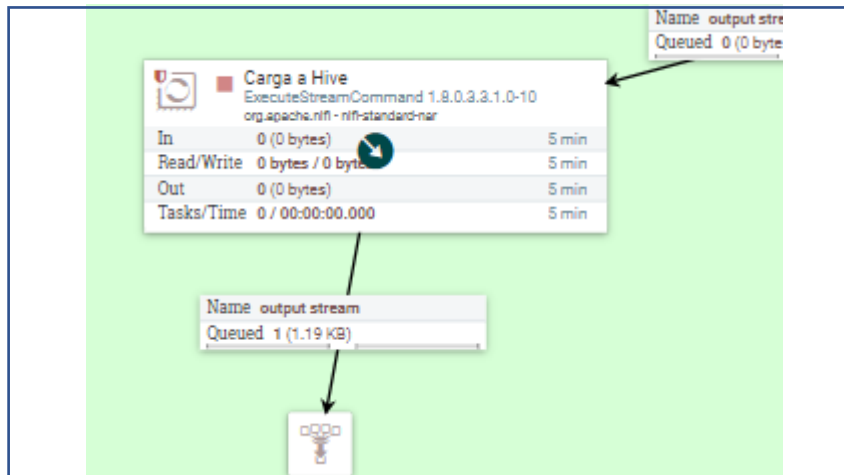


Figura 55 proceso de carga hacia una tabla en HIVE

- Nombre del Shell: file_script_sh_carga → sh_carga_I4AUXBAN.sh

```
1 -C; ssh ${ sh_user_edge } @ ${ ftp_hostname_edge } ${ path_script_sh } / ${ file_script_sh_carga } ${ filename }
```

Establecer cadena vacía

CANCELAR OK

Figura 56 configuración de proceso de ingesta a HIVE

Anexo 11 Constancia de la empresa donde se realizó la investigación



"Año del Bicentenario del Perú: 200 años de Independencia"

Presente.-

Por medio del presente hago constatar que los tesisistas Jonathan Roberto Condori Juarez y Jim David Muñoz Ronceros de la carrera de Ingeniería de Sistemas de la Universidad Cesar Vallejo, con nombre de tesis:

TABLERO DE MANDO BASADO EN LINUX PARA LA PLANTA HFC DEL GRUPO IMATRA SAC

Donde se autoriza la utilización de los datos e información de la empresa para que pueda aplicar en su proyecto de investigación, teniendo en cuenta que el sistema de información gerencial tiene un impacto fundamental dentro de la organización.

Sin más que agregar, quedo atento a cualquier duda o comentario que pudiera surgir. Reciba un cordial saludo.

Atentamente



GRUPO IMATRA S.A.C.
JORGE A. VARGAS LOPEZ
GERENTE GENERAL

Lima, Santiago de Surco, 04 de Enero de 2021