



UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL

**SISTEMA DE POSGRADO
MAESTRÍA EN TELECOMUNICACIONES**

TEMA:

**Gestión y análisis de datos de estudiantes del Instituto Superior
Pedagógico Intercultural Bilingüe “Canelos” con el uso de la
herramienta Pentaho (Big Data)**

AUTOR:

David Hernán Paucar Naranjo

**Trabajo de titulación previo a la obtención del grado de
Magister en Telecomunicaciones**

TUTOR:

Msc. Manuel Romero Paz

Guayaquil, a los 11 días del mes de marzo del año 2022



UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL

**SISTEMA DE POSGRADO
MAESTRÍA EN TELECOMUNICACIONES**

CERTIFICACIÓN

Certificamos que el presente trabajo fue realizado en su totalidad por David Hernán Paucar Naranjo como requerimiento parcial para la obtención del Título de Magíster en Telecomunicaciones.

TUTOR

MSc. Manuel Romero Paz

DIRECTOR DEL PROGRAMA

MSc. Manuel Romero Paz

Guayaquil, a los 11 días del mes de marzo del año 2022



UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL

**SISTEMA DE POSGRADO
MAESTRÍA EN TELECOMUNICACIONES**

DECLARACIÓN DE RESPONSABILIDAD

YO, David Hernán Paucar Naranjo

DECLARO QUE:

El trabajo de Titulación, **Gestión y análisis de datos de estudiantes del Instituto Superior Pedagógico Intercultural Bilingüe “Canelos” con el uso de la herramienta Pentaho (Big Data)**, previa a la obtención del Título de **Magíster en Telecomunicaciones**, ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido.

Guayaquil, a los 11 días del mes de marzo del año 2022

EL AUTOR

David Hernán Paucar Naranjo



UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL

**SISTEMA DE POSGRADO
MAESTRÍA EN TELECOMUNICACIONES**

AUTORIZACIÓN

YO, David Hernán Paucar Naranjo

Autorizo a la Universidad Católica de Santiago de Guayaquil a la **publicación**, en la biblioteca de la institución del Trabajo de Titulación, **Gestión y análisis de datos de estudiantes del Instituto Superior Pedagógico Intercultural Bilingüe “Canelos” con el uso de la herramienta Pentaho (Big Data)**, cuyo contenido, ideas y criterios son de mi exclusiva responsabilidad y total autoría.

Guayaquil, a los 11 días del mes de marzo del año 2022

EL AUTOR

David Hernán Paucar Naranjo

REPORTE URKUND



The screenshot shows the URKUND interface with the following metadata:

Documento	TT David Paucar.docx (D129287305)
Presentado	2022-03-02 20:16 (-05:00)
Presentado por	Luis Córdova Rivadeneira (lcordova@yahoo.com)
Recibido	luis.cordova.ucsg@analysis.orkund.com
Mensaje	TT Ing. David Paucar Mostrar el mensaje completo

39% de estas 32 páginas, se componen de texto presente en 5 fuentes.

The toolbar below the metadata includes icons for document analysis, zoom, quote, and navigation (up, left, right arrows).

SISTEMA DE POSGRADO

MAESTRÍA EN TELECOMUNICACIONES

TEMA: Gestión y análisis de datos de estudiantes Instituto Superior Pedagógico Intercultural Bilingüe "Canelos" con el uso de la herramienta Pentaho (Big Data)

AUTOR: DAVID HERNÁN PAUCAR NARANJO

Trabajo de titulación previo a la obtención del grado de Magister en Telecomunicaciones

TUTOR: Msc. Manuel Romero Paz

Guayaquil, a los 27 días del mes Febrero del año 2022

SISTEMA DE POSGRADO MAESTRÍA EN TELECOMUNICACIONES

CERTIFICACIÓN Certificamos que el presente trabajo fue realizado en su totalidad por David Hernán Paucar Naranjo como requerimiento parcial para la obtención del Título de Magister en Telecomunicaciones.

TUTOR

DEDICATORIA

A toda mi familia, en especial a mi abuelito, por haberme impulsado a ser la persona que soy en la actualidad. Todos mis logros se los debo a todos ustedes, me inculcaron reglas que al seguirlas me ayudaron a ser una mejor persona y a cumplir todas mis metas propuestas

Gracias

AGRADECIMIENTOS

Agradezco a Dios por permitirme poder culminar una meta más en mi vida, sin él, nada de esto sería posible, tengo que agradecer a muchas personas ya que sin ellos no hubiera podido llegar a cumplir esta meta, ha sido un camino difícil, pero sin embargo aquí está el fruto del esfuerzo realizado. Agradezco a mi abuelito, a mis padres, a mis tías, primas, hermanas que estuvieron en todo el trayecto del camino y desearme lo mejor para mi vida. Gracias a mi novia que ha sido una persona que me ha brindado su apoyo y su amor incondicional, gracias a la Universidad Santiago de Guayaquil y a toda su planta docente por haberme brindado la oportunidad de formarme profesionalmente en sus aulas y permitir cumplir esta meta

Muchas Gracias



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL**

**SISTEMA DE POSGRADO
MAESTRÍA EN TELECOMUNICACIONES**

TRIBUNAL DE SUSTENTACIÓN

f.

MSc. Manuel Romero Paz

TUTOR

f.

MSc. Manuel Romero Paz

DIRECTOR DEL PROGRAMA

f.

MSc. Luis Córdova Rivadeneira

REVISOR

MSc. Edgar Quezada Calle

REVISOR

RESUMEN

En la actualidad los grandes volúmenes de datos generados por las instituciones han creado nuevos desafíos que son apoyados por las TICs para el manejo y análisis de la información. La presente investigación tuvo como objetivo realizar la gestión y análisis de datos de estudiantes en el Instituto Superior Pedagógico Intercultural Bilingüe “Canelos”, con el uso de la herramienta Pentaho (Big Data). Se aplicaron métodos inductivo, deductivo e hipotético – deductivo con investigaciones documental y de campo. La caracterización del software permitió determinar los componentes, productos, arquitectura, ventajas y funcionalidad. El diagnóstico situacional reflejó que actualmente se realiza de forma manual el manejo de información sobre estudiantes. Posteriormente, en función a la caracterización y el diagnóstico se desarrolló la propuesta para la gestión y análisis de información con el uso de Pentaho, la cual permitió crear información (tablas y gráficos) sobre estudiantes, pagos, matriculas, otros. Finalmente, la validación de la propuesta reveló que los administradores consideran muy importante el uso de los nuevos reportes por agilizar procesos y rendimiento, mejorar productividad y facilidad de uso. En conclusión, la propuesta diseñada permitirá tener información que puede ser utilizada por los gestores educativos como base para la toma de decisiones de una manera analítica y progresiva en el entorno académico.

Palabras clave: Gestión de datos, Análisis de datos, Pentaho, Big Data, ISPIB Canelos.

ABSTRACT

Currently, the large volumes of data generated by institutions have created new challenges that are supported by ICTs for the management and analysis of information. The objective of this research was to carry out the management and analysis of student data at the "Canelos" Bilingual Intercultural Higher Pedagogical Institute, with the use of the Pentaho (Big Data) tool. Inductive, deductive and hypothetical-deductive methods were applied with documentary and field research. The characterization of the software allowed to determine the components, products, architecture, advantages and functionality. The reflex situational diagnosis that is currently carried out manually in the management of information about students. Subsequently, based on the characterization and diagnosis, the proposal for the management and analysis of information was developed with the use of Pentaho, which allowed the creation of information (tables and graphs) about students, payments, enrollment, and others. Finally, the validation of the proposal revealed that administrators consider the use of the new reports very important to streamline processes and performance, improve productivity and ease of use. In conclusion, the designed proposal will allow information that can be used by educational managers as a basis for decision-making in an analytical and progressive manner in the academic environment.

Keywords: Data management, Analysis of data, Pentaho, Big Data, ISPIB Canelos

INDICE GENERAL

DEDICATORIA	VI
AGRADECIMIENTOS.....	VII
RESUMEN.....	IX
ABSTRACT	X
INDICE GENERAL	XI
ÍNDICE DE FIGURAS.....	XIII
ÍNDICE DE TABLAS.....	XIV
CAPITULO 1 . MARCO REFERENCIAL	15
1.1 INTRODUCCIÓN.....	15
1.2 ANTECEDENTES	15
1.3 PLANTEAMIENTO DEL PROBLEMA	16
1.4 DEFINICIÓN DEL PROBLEMA	17
1.5 JUSTIFICACIÓN DEL PROBLEMA	17
1.6 OBJETIVOS	17
1.6.1 OBJETIVO GENERAL	17
1.6.2 OBJETIVOS ESPECÍFICOS	18
1.7 HIPÓTESIS	18
1.8 METODOLOGIA DE LA INVESTIGACION	18
1.8.1 TIPOS DE INVESTIGACIÓN.....	18
1.8.2 MODALIDAD DE LA INVESTIGACIÓN	19
1.8.3 POBLACIÓN DE ESTUDIO.....	19
1.8.4 PROCEDIMIENTO	20
CAPITULO 2 . MARCO TEÓRICO	22
2.1 DATOS, INFORMACIÓN Y CONOCIMIENTO	22
2.2 SISTEMAS DE INFORMACIÓN	23
2.3 TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES.....	24
2.4 BUSINESS INTELLIGENCE	25
2.4.1 COMPONENTES	27
2.4.2 ARQUITECTURA.....	29
2.4.3 BENEFICIOS	31

2.5	ANALÍTICA DE DATOS	31
2.6	DATA WAREHOUSE	32
2.7	BIG DATA.....	34
2.7.1	BENEFICIOS	36
2.8	ADMINISTRADORES DE BASES DE DATOS	36
2.8.1	PENTAHO.....	37
CAPITULO 3 . CARACTERIZACIÓN Y DIAGNÓSTICO SITUACIONAL		43
3.1	CARACTERIZACIÓN DE LA HERRAMIENTA PENTAHO (BIG DATA).....	43
3.1.1	COMPONENTES	44
3.1.2	PRODUCTOS	44
3.1.3	ARQUITECTURA.....	45
3.1.4	VENTAJAS.....	46
3.1.5	FUNCIONALIDAD	46
3.2	DIAGNÓSTICO SITUACIONAL DEL MANEJO DE INFORMACIÓN SOBRE ESTUDIANTES EN EL ISPIB CANELOS.....	47
CAPITULO 4 . PROPUESTA PARA LA GESTIÓN Y ANÁLISIS DE DATOS DE ESTUDIANTES CON EL USO DE LA HERRAMIENTA PENTAHO.....		50
4.1	PROPUESTA	50
4.2	VALIDACIÓN	61
CONCLUSIONES		68
RECOMENDACIONES.....		69
REFERENCIAS BIBLIOGRÁFICAS.....		70
GLOSARIO DE TÉRMINOS		78
ANEXOS		79

ÍNDICE DE FIGURAS

Figura 2.1 Datos, información y conocimiento	22
Figura 2.2 Componentes de un Business Intelligence	27
Figura 2.3 Arquitectura básica en el desarrollo del BI	30
Figura 2.4 Arquitectura de un Data Warehouse	34
Figura 2.5 Tipos de Big Data	35
Figura 3.1 Logo herramienta Pentaho.....	43
Figura 3.2 Estructura del sistema Pentaho	45
Figura 3.3 Ficha de matrícula	48
Figura 3.4 Ficha socioeconómica	48
Figura 4.1 Herramienta Pentaho Data Integration Version 9.0.....	50
Figura 4.2 Pantalla inicial de la herramienta Pentaho Data Integration Version 9.0	51
Figura 4.3 Carga de información.....	52
Figura 4.4 Jobs de carga ETL.....	53
Figura 4.5 Schema Workbench.....	54
Figura 4.6 Inicio del sistema	54
Figura 4.7 Cubos principales del sistema	55
Figura 4.8 Información de los cubos	56
Figura 4.9 Ejecución de los cubos	57
Figura 4.10 Filtrado de información	57
Figura 4.11 Gráficas de información generada	59
Figura 4.12 Proceso propuesta establecida	60
Figura 4.13 Parámetro de evaluación beneficios de reportes	62
Figura 4.14 Parámetro de evaluación comprensión de reportes	63
Figura 4.15 Parámetro de evaluación eficacia y rendimiento de reportes	64
Figura 4.16 Parámetro de evaluación manejo de reportes.....	65
Figura 4.17 Parámetro de evaluación mejora de la productividad con los reportes	66
Figura 4.18 Parámetro de evaluación conocimiento de la información con los reportes.....	67

ÍNDICE DE TABLAS

Tabla 4.1 Beneficios de los reportes	62
Tabla 4.2 Datos de los reportes	63
Tabla 4.3 Eficacia y rendimiento de los reportes.....	63
Tabla 4.4 Manejo de los reportes.....	64
Tabla 4.5 Mejora de la productividad con los reportes	65
Tabla 4.6 Conocimiento de la información con los reportes	66

CAPITULO 1 . MARCO REFERENCIAL

1.1 INTRODUCCIÓN

Las Tecnologías de la Información y las Comunicaciones (TICs) han evolucionado a la sociedad de manera muy espontanea, ofreciendo nuevas formas de indagación de la información (Billnitzer & Jankovic, 2020). Estas han sido de impacto para el desarrollo de las empresas, optimizando el tiempo y agilizando los procesos generados dentro de las mismas (Castel, 2018). Son cada vez más utilizadas en el apoyo y automatización de la información generada dentro de una institución, pues se convierten en una herramienta con capacidad de adquirir, gestionar y transmitir el conocimiento estructural de la empresa (Buenrostro et al., 2019).

Las TICs dentro de las organizaciones se convierten en un conjunto de productos y procesos procedentes de nuevas herramientas (software y hardware), soportes y canales de comunicación, referentes con el almacenamiento, procesamiento y la transmisión digitalizada de los datos, mejorar la productividad, al ser aplicadas de manera inteligente (Pita, 2018). De esta forma, actualmente las empresas se encuentran en una etapa de transformación de una administración establecida en los documentos en papel a una manejada y producida en formatos electrónicos (Spremolla, 2017).

Por ende, las organizaciones de vanguardia son las que emplean tecnologías actuales, mismas que tienen la convicción de contar con datos reales, precisos el momento necesario de las instituciones para tomar decisiones (Roig, 2018), de forma efectiva para obtener mayor competitividad y supervivencia en el mercado actual (Martinez, 2019).

1.2 ANTECEDENTES

Con la utilización de la herramienta Pentaho se empodera a la institución en el flujo de información, con un sistema de análisis, modelos predictivos y seguridad en los datos (Sagubay et al., 2018). La plataforma cloud

permite la preparación y modificación de datos, estableciendo un ambiente inteligente y accesible. De esta manera, el almacenamiento de la información puede ser visualizado por medio de la web (Posada & Ampuero, 2021).

El manejo de la información implica las proyecciones administrativas que se enfocan en las necesidades del usuario, que se tiene en cuenta al momento de emplear una herramienta que pueda medir los datos para su utilidad y beneficio (Mendoza, 2018). Una de las instituciones que presentó la necesidad de gestionar de mejor manera la información de sus estudiantes es el Instituto Superior Pedagógico Intercultural Bilingüe Canelos (ISPIB Canelos). Mismo que se encuentra localizado en la parroquia Canelos, al sureste de la capital de Pastaza (Puyo). La parroquia de Canelos tiene una extensión de 440km². Limita al norte con la parroquia El Triunfo, al sur con la parroquia Simón Bolívar, al este con la parroquia Sarayacu y al oeste con la parroquia Veracruz. Su población es de 2.173 habitantes, de los cuales 1.774 (82%) se reconocen como indígenas (INEC, 2010).

Ante lo expuesto, la presente investigación estuvo encaminada a realizar la gestión y análisis de datos de estudiantes en el ISPIB Canelos, con el uso de la herramienta Pentaho (Big Data). A través, de la utilización de esta herramienta tecnológica se generan diferentes reportes en tiempo real necesarios para el análisis y presentación de información estadística, basada en procesos y actividades concretas.

1.3 PLANTEAMIENTO DEL PROBLEMA

La repentina invasión de la pandemia COVID-19 ha aplacado cambios sociales y económicos (García et al., 2021). Por ello, las organizaciones han visto la necesidad de cambiar sus modelos de negocio para asegurar su continuidad (Díaz et al., 2020). si bien tendencias como el teletrabajo, y el uso generalizado del internet parecen haber llegado para quedarse (Jácome, 2020).

Es así que, la gestión de los datos referente a estudiantes en el ISPIB Canelos, principalmente se basa en procesos sistemáticos que se realizan actualmente de forma manual. En este sentido, entre las causas del problema descrito se encuentran: el desconocimiento al realizar reportes, bajo nivel de preparación del personal y herramientas complejas de manejar. Esto conlleva consecuencias como: el mal manejo de la información, mala utilización de los recursos y demora en la presentación de reportes dentro de la organización.

Por lo tanto, resulta importante la elaboración de una propuesta para la gestión y análisis de datos de estudiantes con el uso de la herramienta Pentaho (Big Data) para agilizar los procesos y la generación de indicadores en cuanto a: número de estudiantes, pagos, materias, matriculas, entre otros.

1.4 DEFINICIÓN DEL PROBLEMA

¿Cómo incide el uso de la herramienta Pentaho en la gestión y análisis de datos de los estudiantes del Instituto Superior Pedagógico Intercultural Bilingüe “Canelos”?

1.5 JUSTIFICACIÓN DEL PROBLEMA

Al elaborar una propuesta para la gestión y análisis de datos de estudiantes con el uso de la herramienta Pentaho (Big Data) se contribuirá a la agilización de los procesos y la generación de indicadores en cuanto a: número de estudiantes, pagos, materias, matriculas, entre otros.

1.6 OBJETIVOS

1.6.1 OBJETIVO GENERAL

Realizar la gestión y análisis de datos de estudiantes en el Instituto Superior Pedagógico Intercultural Bilingüe “Canelos”, con el uso de la herramienta Pentaho (Big Data).

1.6.2 OBJETIVOS ESPECÍFICOS

- Caracterizar la herramienta Pentaho (Big Data).
- Realizar un diagnóstico situacional del manejo de información sobre estudiantes en el ISPIB Canelos.
- Elaborar una propuesta para la gestión y análisis de datos de estudiantes con el uso de la herramienta Pentaho.

1.7 HIPÓTESIS

Con el uso de la herramienta Pentaho (Big Data) se gestiona y analiza de mejor manera los datos de estudiantes en el Instituto Superior Pedagógico Intercultural Bilingüe “Canelos”.

1.8 METODOLOGIA DE LA INVESTIGACION

El presente estudio se diseñó de manera sistemática con la finalidad de garantizar el logro de resultados fiables y válidos, mismos que responden a los objetivos específicos planteados. De esta forma, los métodos aplicados fueron los siguientes:

- **Inductivo:** se aplicó este método para el logro de conclusiones generales a partir de premisas particulares. Por ello, se efectuó una observación general de los hechos para su posterior registro.
- **Deductivo:** se aplicó este método de razonamiento mediante el cual las consecuencias concretas se dedujeron en principios generales.
- **Hipotético – deductivo:** se aplicó este método ya que permitió obtener una mayor aproximación y comprensión del tema de investigación para llegar a la comprobación de la hipótesis planteada.

1.8.1 TIPOS DE INVESTIGACIÓN

Los tipos de investigación empleados para el logro de objetivos fueron:

- **Investigación cualitativa:** puesto que se realizó una investigación bibliográfica con el fin de identificar y ampliar los conocimientos indispensables para este trabajo para caracterizar a la herramienta Pentaho.
- **Investigación descriptiva:** puesto que se realizó un análisis de la realidad a través de una serie de parámetros planteados, para la realización del diagnóstico situacional del manejo de información sobre estudiantes en el ISPIB Canelos.
- **Investigación explicativa:** puesto que se buscó la relación existente entre variables para analizar y explicar su relación, para explicar la propuesta sobre la gestión y análisis de datos de estudiantes con el uso de la herramienta Pentaho.

1.8.2 MODALIDAD DE LA INVESTIGACIÓN

Las modalidades de investigación fueron:

- **Modalidad documental:** se realizó a partir de la revisión de diferentes fuentes bibliográficas para caracterizar a la herramienta Pentaho y para el diagnóstico situacional del manejo de información en la institución.
- **Modalidad de campo:** el conjunto de datos utilizados para la gestión y análisis de datos de estudiantes fue proporcionado por el ISPIB Canelos. Además, se realizó un análisis sistemático de información relevante para elaborar la propuesta con el uso de la herramienta Pentaho. Para la recolección de datos de la validación de la propuesta se utilizó como instrumento una encuesta a los responsables en el manejo de la información para la validación de la propuesta.

1.8.3 POBLACIÓN DE ESTUDIO

La población de la presente investigación estuvo constituida por 8 personas (Rector, Docente TICs, Coordinadora de Bienestar Institucional, Coordinadora de Carrera de TSDII, Coordinación de Vinculación, Coordinación de Investigación, Tesorería y Coordinación de Evaluación Interna) que administran la información de estudiantes dentro del ISPIB

Canelos. Ellos permitieron tener un diagnóstico inicial de cómo se maneja la información y también en conjunto se validó la propuesta de gestión y análisis de datos de estudiantes con el uso de la herramienta Pentaho.

1.8.4 PROCEDIMIENTO

- Para la caracterización de la herramienta Pentaho (Big Data), se realizó una revisión documental de: componentes, productos, arquitectura, ventajas y funcionalidad de la herramienta. El instrumento que se utilizó para la caracterización fue la guía de análisis de documentos, lo que proporcionó un análisis a fondo sobre la herramienta Pentaho.
- Para el diagnóstico situacional del manejo de información sobre estudiantes en el ISPIB Canelos, se aplicó la observación directa y la consulta a los administradores de la información. El instrumento que se utilizó fue la guía de observación.
- Para la elaboración de la propuesta para la gestión y análisis de datos de estudiantes con el uso de la herramienta Pentaho, se utilizó un servidor con requerimientos mínimos como: CPU básico, Core i3, Memoria 100gb y con instalación jdk 8 java. Posteriormente, en la herramienta Pentaho vinieron configurados todos los servicios con componentes principales como: **i) Pentaho report designer:** el cual permitió crear informes predefinidos, a través de la consulta de múltiples orígenes. **ii) Pentaho data integration o kettle:** el cual permitió crear procesos ETL, que son procesos de extracción, transformación y carga, con los que podemos extraer la información de diferentes orígenes, transformarlo en información útil y cargarlos en nuestro DWH, que es el almacén de los datos, **iii) Pentaho schema workbench:** el cual permitió crear el cubo OLAP a través del motor denominado Mondrian (potente en consultas OLAP), y también es Open Source y **iv) Pentaho server:** es el núcleo de la herramienta. De esta manera, mediante los ETL se cargaron las matrices a las bases de datos, dando origen a que los procesos en adelante sean automatizados.

Finalmente, para la validación de la propuesta se utilizó la técnica de la encuesta dirigida a los 8 administradores de la información en la institución. La misma se realizó con la finalidad de obtener una evaluación de los beneficios, comprensión de datos, mejora de eficiencia y rendimiento, dificultad de manejo, mejora de productividad, conocimiento de información en los nuevos reportes generados con la gestión y análisis de datos de estudiantes. El instrumento que se utilizó para determinar la validación fue el cuestionario diseñado con 6 preguntas cerradas (Anexo 1). El procesamiento de la información obtenida se realizó en el programa Microsoft Excel y el análisis se efectuó en base a la estadística descriptiva mediante la generación de la frecuencia absoluta y relativa.

CAPITULO 2 . MARCO TEÓRICO

El presente capítulo presenta los conceptos básicos para la comprensión de la gestión y análisis de datos:

2.1 DATOS, INFORMACIÓN Y CONOCIMIENTO

La forma apropiada para diferenciar los términos; los datos están localizados en el mundo y el conocimiento se encuentra ubicado en agentes de cualquier tipo, mientras que la información adopta un papel mediador entre ambos (Peñafiel et al., 2019).

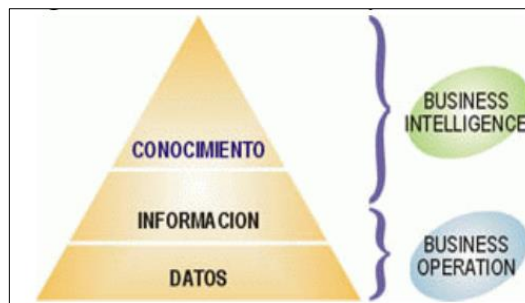


Figura 2.1 Datos, información y conocimiento

Fuente: (Peñafiel et al., 2019)

Los datos se consideran elementos primarios de la información cuyo valor por si solos no son de relevancia para la toma de decisiones, son la mínima unidad semántica; los datos pueden ser una colección de hechos almacenados en algún dispositivo; pueden ser internos o externos a la organización, objetivos o subjetivos o de tipo cualitativos y cuantitativos.

Según, Peñafiel et al. (2019) la información es un conjunto de datos procesados con un significado, un propósito y/o contexto; cuya utilidad reside en la toma de decisiones debido a la minimización de la incertidumbre, para la transformación en información de los datos se les añade valor:

- **Contextualizar:** contexto y propósito de su generación
- **Categorizar:** se conocen las unidades de medida que ayudan a interpretarlos.

- **Calcular:** procesamiento matemático o estadístico
- **Corregir:** depuración de los datos; errores e inconsistencias
- **Condensar:** resumen de datos

El conocimiento es la combinación de experiencias, valores, información y “saber hacer” que sirve como marco para la incorporación de nuevas experiencias e información, y es útil para la acción; para que la información se transforme en conocimiento es necesario realizar (Peñafiel et al., 2019):

- Comparación con otros elementos
- Predicción de consecuencias
- Búsqueda de conexiones
- Conversación con otros portadores de conocimiento

2.2 SISTEMAS DE INFORMACIÓN

Un sistema de información (SI) son los sistemas de software y hardware que soportan aplicaciones de uso intensivo de datos (Lapiedra et al., 2021). Con ello, es un conjunto ordenado de mecanismos que tienen como fin la administración de datos y la información, de manera que puedan ser recuperados y procesados fácil y rápidamente (Martínez, 2018). Permite además que la información esté disponible para satisfacer las necesidades en una organización.

Presenta como finalidad la gestión y administración de la información que la conforma. Además, es entender y analizar cómo ocurre el impacto de la adaptación de las tecnologías de la información en los procesos de decisiones gerenciales y administrativas de las empresas (de Pablos et al., 2019).

Los recursos de un sistema de información según Alvarado et al. (2018) son:

- **Datos:** cualquier tipo de información masiva que puede organizarse.
- **Actividades:** procedimientos, pasos a seguir, estaciones de trabajo, etc.

- **Recursos de informáticos:** aquellos determinados por la tecnología.

Asimismo, las actividades básicas de un sistema de información según Cobeña et al. (2018) son:

- **Entrada de información:** proceso en el cual el sistema toma los datos que se requieren.
- **Almacenamiento de información:** puede hacerse por computadora o por archivos físicos para conservar la información.
- **Procesamiento de la información:** ayuda a la transformación de la información, misma que puede ser útil en la toma de decisiones.
- **Salida de información:** en la cual el sistema produce la información que fue procesada.

2.3 TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES

Las TICs son el resultado de poner en interacción la informática y las telecomunicaciones. Todo con el fin de mejorar el procesamiento, almacenamiento y la transmisión de la información (Córdova et al., 2019). El elemento más representativo de las nuevas tecnologías es sin duda el ordenador y más específicamente el internet (Costa et al., 2018). El sector de las TICs se muestra actualmente como líder en la manipulación de datos provenientes de diferentes medios como redes sociales, plataformas tecnológicas y dispositivos electrónicos de comunicación que requieren emplear la información capturada o adquirida para ser transformada y suministrar informes estadísticos descriptivos o tal vez informes sobre modelos predictivos aplicables a diversas áreas del saber (Bravo et al., 2021).

La participación de las personas en el desarrollo de las TIC's contribuye en el crecimiento social, cultural, económico y político de un país. Pues la difusión de información a través de diversos medios permitirá que las personas se vayan forjando una idea de cómo se puede contar con la mayor información para la necesidad que se presente, pero no solo basta con estar informados, se debe ser objetivo, más realista y al mismo tiempo

visionario (Pita, 2018). De esta manera, el objetivo principal de las TICs es mejorar la calidad de vida y reducir los esfuerzos y el tiempo necesario para realizar procesos de trabajo y comunicación (Chuya et al., 2021).

Entre los elementos que componen las TIC se encuentran según (Muñoz et al., 2019):

- **Informática:** los recursos informáticos permiten procesar la información y los recursos a una velocidad muy alta, permitiendo resolver tareas muy complejas en muy poco tiempo.
- **Redes:** la evolución de las redes permite la comunicación de los recursos informáticos entre sí utilizando diferentes formas.
- **Software:** los programas y aplicaciones son importantes al momento de emplearlos recursos informáticos y las redes de comunicaciones para la creación de sistemas, plataformas y entornos, donde se puede almacenar, interactuar compartir y gestionar la información.

2.4 BUSINESS INTELLIGENCE

Business Intelligence (BI) es un término general para las estrategias, tecnologías y sistemas de información utilizados por las empresas para las herramientas, tecnologías, aplicaciones y prácticas utilizadas para recopilar, integrar, analizar y presentar los datos sin procesar de una organización para crear información de negocios (Mora, 2018), de forma que se puede optimizar el proceso de toma de decisiones en los negocios (Liang & Liu, 2018). Por otra parte, según de Sá et al. (2018) los sistemas de BI se pueden definir como herramientas para ayudar y ampliar los procesos de toma de decisiones y hacerlos más precisos y confiables, basados en el conocimiento generado por los datos de la empresa en valores intuitivos y experiencias personales. Desde un punto de vista directo con las TICs se constituye en el conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información desestructurada en información estructurada (Cheng et al., 2020).

Las empresas que utilizan plataformas de BI cuentan con herramientas analíticas que brindan información y datos esenciales para su gestión. En plena era digital, tomar decisiones bien informadas es uno de los principales factores de diferenciación de las empresas (Leite et al., 2019). Por ello, el BI combina análisis de negocios, minería de datos, visualización de datos, herramientas de infraestructura de datos y mejores prácticas para ayudar a las organizaciones a tomar decisiones basadas en datos (Chen & Lin, 2021). Las soluciones del BI modernas priorizan el análisis de autoservicio flexible, los datos gobernados en plataformas confiables, los usuarios comerciales empoderados y la velocidad para obtener información (Božič & Dimovski, 2019).

De esta manera, Cordero et al. (2020) manifiestan que las categorías en las que se distribuyen las soluciones de inteligencia de negocios son:

- **Data management tools:** ayuda a la depuración y estandarización de la información de diversa procedencia hasta la extracción, transformación y transferencia a un sistema.
- **Data discovery applications:** permiten recopilar y evaluar nueva información (data mining o minería de datos), y aplicar sobre esa información nueva o sobre la ya disponible técnica de análisis predictivo para realizar proyecciones de futuro.
- **Herramientas de reporting:** una vez recopilada y trasladada toda esa información preexistente o nueva, ayudan a las empresas a visualizarla de manera gráfica e intuitiva.

Sin duda, el uso de plataformas de BI es una opción eficaz para apoyar la toma de decisiones. Las plataformas de BI son un activo importante para cualquier empresa, ya que tienen múltiples beneficios que se convierten en una ventaja competitiva. Se sabe que los costos de implementación de BI son uno de los principales factores para la distribución de BI entre las pequeñas y medianas empresas (PYME) (Leite et al., 2019).

2.4.1 COMPONENTES

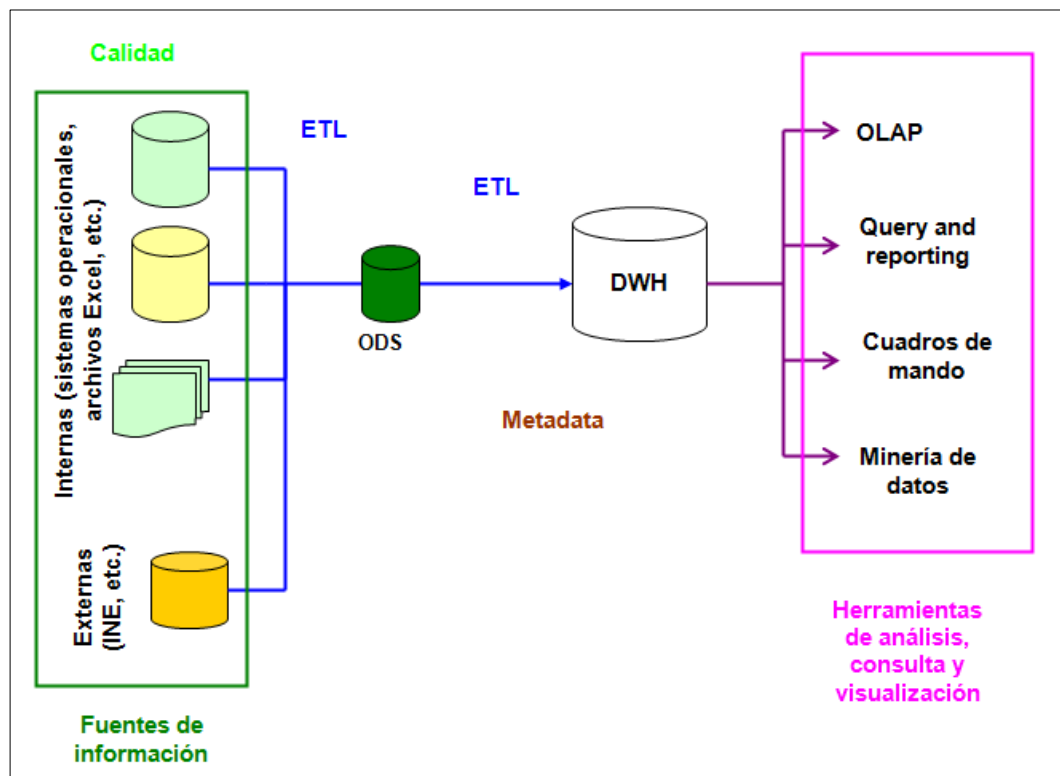


Figura 2.2 Componentes de un Business Intelligence

Fuente: (Pereáñez & Arango, 2020)

En este sentido, los componentes fundamentales del BI según se describen a continuación:

- La fuente que proporciona la información al almacén de datos de la organización o departamento.
- El Data Warehouse (DWH) constituye el almacén central de la información dentro de un sistema BI.
- El software utilizado para la extracción, transformación y carga de datos al DWH.
- El almacén de datos en conjunto con el diccionario de datos que provee acceso y administración de la información.
- El motor de online analytical processing (OLAP) para realizar cálculos, consultas, funciones, pronósticos y análisis de la información.
- El software de minería de datos para elaborar predicciones y tendencias sobre los procesos dentro de una organización.

- El software de reportes para presentar los datos especificados y mostrar mediante un navegador web o archivo de la información resultante del análisis de datos.
- Los datos almacenados dentro del DWH puede provenir de diversas fuentes en torno a la institución (excel, ficheros, CSV, otros), como también fuentes externas (geomarketing, sociodemográficas, información financiera, otras).

Sin embargo, Machado et al. (2019) menciona que los procesos de extract, transform and load (ETL) se convierten en la clave para el desarrollo de una solución de BI y se encuentra compuesto por las siguientes fases:

- **Extracción de datos:** puede realizarse mediante un desarrollo propio (manual), o empleando herramientas ETL existentes en el mercado (facilitan el mantenimiento, es la evolución del sistema).
- **Limpieza de datos:** para identificar valores contradictorios y cuando un mismo campo se está empleando con fines distintos, para establecer valores por defecto o ausencia de datos, errores debido a migraciones desde versiones anteriores, entre otros aspectos.
- **Transformación de datos:** una vez limpios, los datos son transformados. Lo que incluye sustitución de códigos, cambio de formato y creación de valores agregados y derivados.
- **Integración de datos:** carga de datos en el DWH.

Según Peñafiel et al. (2019) las funciones de los componentes de un BI son las siguientes:

- **Base de datos relacional:** depuración y homogeneización almacenamiento de datos.
- **Base de datos multidimensional:** velocidad de acceso y consulta. capacidad de análisis desde varios puntos de vista (dimensiones de análisis).
- **Visualizador:** análisis de la información (drill-down, drill-throughy rotación de filas y columnas.

2.4.1.1 HERRAMIENTAS DE APROVECHAMIENTO DE INFORMACIÓN

Entre las herramientas para la gestión de la información se encuentran las siguientes:

- **Cuadro de mando analítico:** son las herramientas de gestión definitiva. Los cuadros de mando brindan una visión general del rendimiento, gracias a una selección específica de los KPI, los cuales permiten medir el éxito de las actividades (Bermeo et al., 2020).
- **Cuadro de mando integral o estratégico:** en un método de gestión enfocado a la creación de valor, ayuda a las organizaciones a transformar sus estrategias en objetivos operativos medibles y relacionados entre sí (Meijide, 2019).
- **Data mining:** consiste en extraer información de un conjunto de datos y transformarla en una estructura comprensible, esto es, en información útil y accesible para que pueda ser usada posteriormente (Rangriz & Bayrami, 2019).
- **Online analytical processing:** es una tecnología que utiliza estructuras multidimensionales para proporcionar un acceso rápido a los datos para su análisis. OLAP permite a los analistas, gerentes y ejecutivos obtener información sobre los datos a través de un acceso rápido, consciente e interactivo a una amplia variedad de vistas posibles de la información. OLAP transforma los datos sin procesar para que reflejen la dimensión real de la empresa tal como la entiende el usuario (Tseng et al., 2020).
- **Query and reporting:** las consultas o informes libres trabajan tanto sobre el detalle como sobre las agregaciones de la información. El ejecutar este tipo de explotación de información en una base de datos implica tener una optimización del entorno mediante informes (reporting), puesto que el DWH conserva una tecnología y estructura muy apropiada para este tipo de requerimientos (Tangari et al., 2020).

2.4.2 ARQUITECTURA

El BI comienza con el modelo de origen de datos de una organización o empresa. Estos datos se extraen, se transforman y cargan mediante un

proceso intermediario entre la fuente de datos y el destino de la información. Los datos finales se almacenan en un DWH para la construcción de datamarts. su característica principal es optimizar el análisis de datos dentro de una organización en el DWH y los datamarts pueden ser utilizados mediante herramientas de análisis y reportes de datos. La información almacenada en el DWH se explota a través de la utilización de herramientas de reporting, análisis, alertas, otras.

En estas herramientas se basa también la construcción de productos BI más completos, como los sistemas de soporte a la decisión (DSS), los sistemas de información ejecutiva (EIS) y los cuadros de mando (CMI) o balancead scorecard (BSC). De esta forma, a continuación, se puede apreciar la arquitectura básica en el desarrollo del BI:

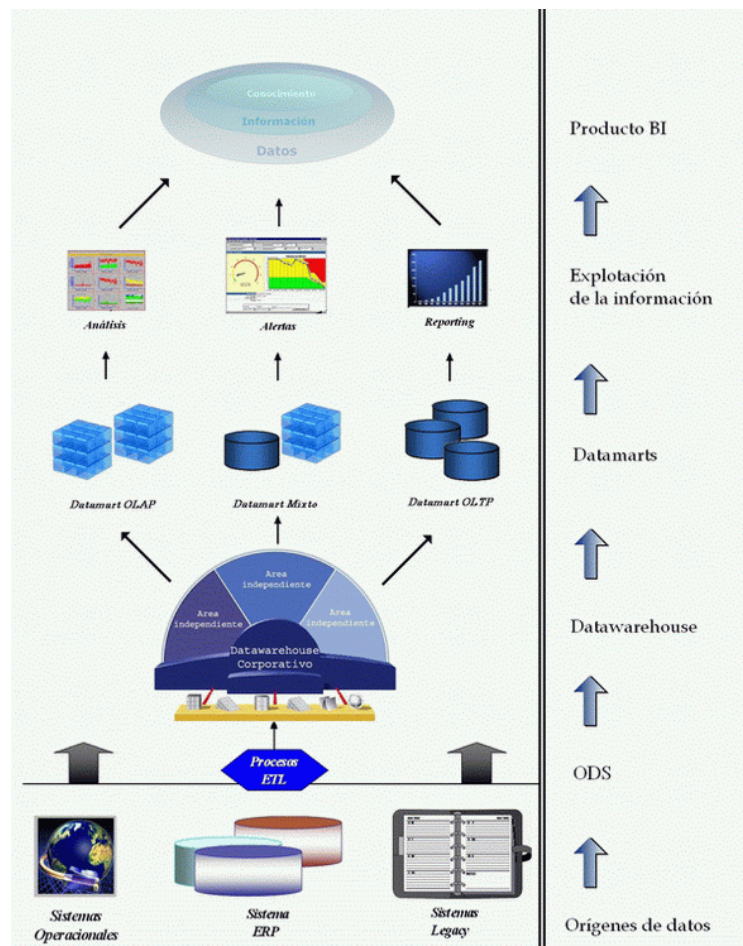


Figura 2.3 Arquitectura básica en el desarrollo del BI

Fuente: (Sorour et al., 2020)

2.4.3 BENEFICIOS

La principal ventaja del BI es que permite a las organizaciones entender las fuerzas que moldean los mercados y los negocios, favoreciendo así que puedan adelantarse a la competencia al tiempo que satisfacen las necesidades de los consumidores (Lopes et al., 2020).

Según Borse et al. (2019) los beneficios del BI son los siguientes:

- Aumenta la capacidad de planificar y realizar pronósticos con precisión gracias a la información.
- Aumenta la satisfacción de los clientes y los empleados.
- Mejora la toma de decisiones al estar fundadas en evidencias y datos concretos.
- Mejorar la toma de decisiones de negocio y la eficiencia operativa.
- Obtener datos de mayor calidad y alcanzar mayores ventajas competitivas.
- Permite la creación de informes de manera integra al emplear grandes cantidades de información.
- Realizar informes, planificar o analizar con bastante precisión y rapidez.
- Recopila y analiza información interna y externa, independientemente de su origen y fuente
- Reducir los costes y aumentar los ingresos.

2.5 ANALÍTICA DE DATOS

La analítica de datos es el proceso de examinar conjuntos de datos para encontrar tendencias y sacar conclusiones sobre la información que contienen. la analítica de datos se utiliza cada vez más con la ayuda de sistemas y software especializado (Treviño et al., 2020). Las tecnologías y técnicas de análisis de datos se utilizan ampliamente en las industrias comerciales para permitir a las organizaciones tomar decisiones empresariales más informadas. También se utiliza por científicos e investigadores para verificar o refutar modelos, teorías e hipótesis científicas (Aguilar, 2020).

Permite a las organizaciones analizar todos sus datos (en tiempo real, históricos, no estructurados, estructurados cualitativos) para identificar patrones y generar conocimientos para informar y como en algunos casos, automatizar decisiones, conectando la inteligencia y la acción (Bravo et al., 2021). El proceso analítico de datos conlleva i) comprender el problema empresarial recopilar e identificar datos relevantes para el problema, ii) preparar los datos para el análisis, iii) analizar los datos para generar conocimientos, iv) poner en funcionamiento los análisis y modelos y v) supervisar y optimizar el rendimiento (Tejada et al., 2018).

2.6 DATA WAREHOUSE

Un DWH es un almacén único, completo y consistente de datos obtenidos de una variedad de fuentes y puestos a disposición de los usuarios finales de una manera que puedan entender y utilizar en un contexto empresarial. Se trata de una base de datos corporativos que se caracteriza por integrar y depurar la información para luego procesarla, permitiendo el análisis de la información desde distintos puntos de vista (Peñafiel et al., 2019). La creación de un DWH representa en la mayoría de las ocasiones el primer paso, desde el punto de vista técnico, para implementar una solución completa y fiable del BI (Garcelon et al., 2018).

La ventaja principal de estos sistemas se basa en la estructura de la información, que supone el almacenamiento de los datos de forma homogénea y fiable, en una estructura basada en la consulta y el tratamiento jerarquizado de la misma, y en un entorno diferenciado de los sistemas operacionales (Wu et al., 2020).

Según Inmon (2000), quien acuñó el término DWH se caracteriza por ser:

- **Histórico:** permite almacenar datos en momentos concretos como para realizar análisis de tendencias en esos momentos o periodos de tiempo y poder realizar comparaciones entre esos valores tomados.

- **Integrado:** los datos almacenados tienen que ser integrados en una estructura consistente, lo que conlleva a la eliminación de las inconsistencias existentes entre sistemas de operaciones.
- **No volátil:** esto implica que la información almacenada en el data DWH es para ser leída pero no modificada. Por lo que la información es permanente.
- **Temático:** para facilitar el acceso a los datos y el entendimiento de ellos se suele organizar por temas.

Según Bouaziz et al. (2019) los beneficios de DWH son:

- Facilita la aplicación de técnicas estadísticas de análisis y modelización para encontrar relaciones ocultas entre los datos de almacén.
- Proporciona la capacidad de aprender de los datos del pasado y de presidir situaciones futuras en diversos escenarios.
- Proporciona una herramienta para la toma de decisiones en cualquier área funcional, basándose en información integrada y global del negocio.
- Simplifica dentro de la empresa la implantación de sistemas de gestión integral de la relación con el cliente.
- Supone una optimización tecnológica y económica en entornos de Centro de Información como estadística o de generación de informes.

Las etapas generales de almacén de los datos según Gunadi & Widiyanto (2020) son:

- Determinación de los objetivos comerciales
- Recopilación y análisis de la información
- Identificación de los procesos comerciales centrales
- Construcción de un modelo de datos conceptual
- Ubicación de las fuentes de datos y planificación de las transformaciones de datos
- Establecimiento de la duración del seguimiento
- Implementación del plan

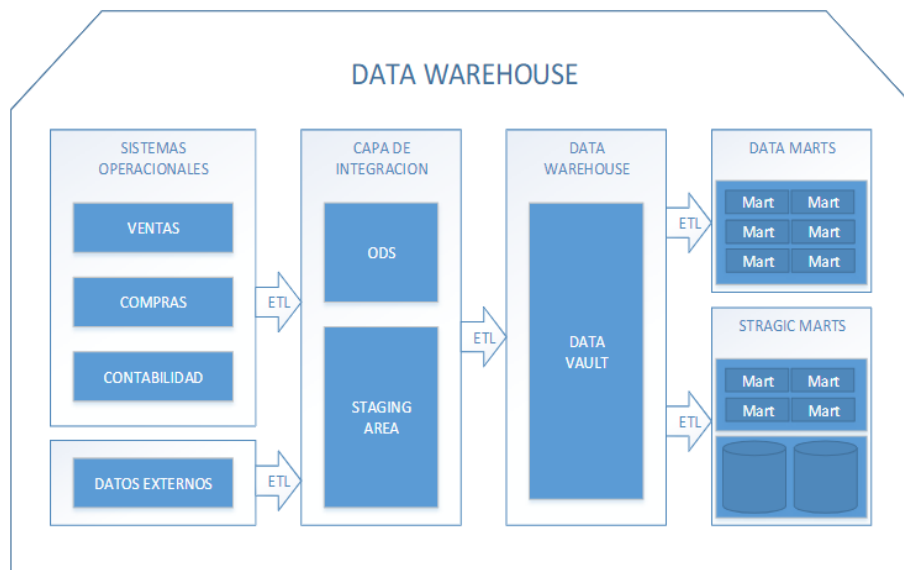


Figura 2.4 Arquitectura de un Data Warehouse

Fuente:(Bouaziz et al., 2019)

2.7 BIG DATA

La tecnología evoluciona progresivamente y en estos cambios constantes surge la Big Data o analítica de datos a gran escala, que apoyados en las nuevas tendencias tecnológicas como la computación en la nube permiten su almacenamiento y su explotación puede provocar un giro rápido a determinado sector (Tejada et al., 2018). En este sentido, Big Data es el conjunto de datos o combinaciones de conjuntos de datos cuyo tamaño, complejidad y velocidad de crecimiento dificultan su captura, gestión, procesamiento o análisis mediante tecnologías y herramientas convencionales (Márquez, 2020). Por otra parte, se refiere a los datos que son tan grandes como más rápidos o complejos que es difícil o imposible procesarlos con los métodos tradicionales (Wang et al., 2020).

El análisis de Big Data ayuda a las organizaciones a aprovechar sus datos y utilizarlos para identificar nuevas oportunidades. Y su importancia no gira en torno a la cantidad de datos que se tiene sino en lo que se puede hacer con dicha información (Lv & Qiao, 2020). Por ello, se puede tomar datos de cualquier fuente y analizarlos para encontrar respuestas que permitan: reducir los costos, reducir el tiempo, desarrollar nuevos productos y optimizar las ofertas, y tomar decisiones inteligentes (Wang et al., 2020).

Las nuevas tendencias tecnológicas en referencia a los datos a gran escala o también denominada Big Data en la actualidad, no fuese posible sin el uso de tecnologías como internet o la computación en la nube, que según cifras del portal de estadísticas internacionales “Internet World Stats”, el uso y población de internet en el mundo, hasta diciembre de 2017, el continente asiático domina con el 48.7%, seguido de África con 16.9%, mientras el continente con menos uso de internet está en Oceanía/Australia con el 0.6%. Latinoamérica y Caribe ocupan el cuarto lugar con el 8.5% referente al uso de internet en esta región (Internet Word Stats, 2018).

Según Ghani et al., (2019) los tipos de Big Data son los siguientes:

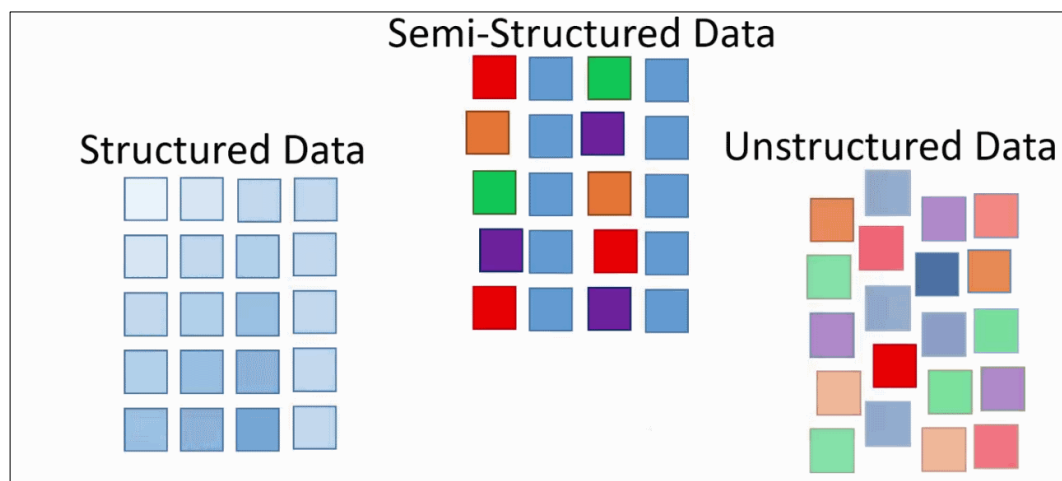


Figura 2.5 Tipos de Big Data

Fuente: (Ghani et al., 2019)

- **Datos estructurados:** tienen un formato fijo y a menudo son numéricos. Este tipo de datos es de información que ya está ordenada en bases de datos y hojas de cálculo, están almacenadas en bases de datos SQL, lagos de datos y almacenes de datos.
- **Datos no estructurados:** es información que carece de organización y no está en un formato predeterminado porque puede ser muy diversa. Se pueden convertir en archivos de documentos de texto almacenados en Hadoop, como clústeres o sistemas NoSQL.
- **Datos semiestructurados:** pueden contener ambas formas de datos, como por ejemplo los registros de servidores web o los datos de

sensores que se hayan configurado. Para ser precisos, son datos que, a pesar de no estar clasificados en un repositorio concreto (una base de datos), contienen información vital o etiquetas que segregan elementos individuales dentro de los datos.

2.7.1 BENEFICIOS

Según, Borja et al. (2020) los beneficios de un Big Data son los siguientes:

- Análisis de seguridad que permite la reducción de riesgos y pérdidas frente a fraudes (reducción de costos).
- Desarrollo de nuevos productos
- Filtros inteligentes de seguridad en el negocio electrónico
- Ofertas optimizadas y personalizadas
- Permite detectar patrones complejos de fraude en tiempo real analizando los datos históricos, el patrón de uso de información de geolocalización, análisis de transacciones y operaciones sospechosas (reducción de costos).
- Permite la simplificación de procesos actuales y control del negocio (reducción de costos)
- Reducción de costos
- Reducción de tiempos
- Soporte a la toma de decisiones a través de algoritmos automáticos.
- Toma de decisiones más inteligentes
- Una analítica sofisticada que analice todos los informes y datos que ayuda a la toma de decisiones, reduciendo los riesgos y descubre información que antes podría estar oculta, pero a la vez importante (ayuda a la toma de decisiones).

2.8 ADMINISTRADORES DE BASES DE DATOS

Las empresas utilizan bases de datos para mantener y actualizar grandes cantidades de información de forma estructurada para poder usarla más eficazmente (Ruiz et al., 2018). En este sentido, los administradores de bases de datos gestionan y mantienen las bases de datos informatizadas y

se aseveran de que éstas sean seguras y estén actualizadas. Son responsables de las bases de datos informatizadas de las empresas (Ballesteros et al., 2018).

2.8.1 PENTAHO

Es el líder mundial de sistemas de inteligencia a nivel empresarial, además es de código abierto, dispone de una suite llamada Pentaho suite donde ofrece todas sus herramientas enfocadas a la integración de datos y análisis inteligente los datos de la organización (de Sá et al., 2018). Se ha convertido en una gran alternativa a la hora de trabajar con sistemas BI. Incluye todo tipo de herramientas que están dirigidas principalmente a mejorar el proceso de toma de decisiones empresariales: software de gestión de almacenes de datos, herramientas de integración y análisis de datos, software para gestores y herramientas de minería de datos (Leite et al., 2019).

Entre los diversos softwares disponibles, Pentaho es un software libre que ofrece buenas posibilidades de exploración de recursos de BI y con posibilidad de personalización (Sá et al., 2018). Se creó en 2004 y comprende las herramientas Pentaho Reporting, Pentaho Reporting Server, Mondrian OLAP Server y Pentaho Data Integration. Estas herramientas componían la Pentaho Open BI Suite. En 2006, Pentaho abarcó los proyectos Kettle y Weka. En 2015, Hitachi Data Systems adquirió Pentaho. En los últimos años se ha lanzado una nueva edición por año, encontrándose actualmente en la versión 8.2. La plataforma Pentaho BI está disponible en dos versiones, ambas desarrolladas en Java. La Enterprise Edition, siendo esta la comercial y la Community Edition, la versión de código abierto (Pentaho, 2018).

Este software de gestión avanzada de datos utiliza técnicas ETL para analizar y extraer un gran volumen de datos con el objetivo de transformar números difíciles para el ser humano en informes gráficos más sencillos de entender, y que bien interpretados, se convierten en un conocimiento muy útil y valioso para las empresas, todo ello sin ingresar ni una sola línea de

código (Saraswati & Martarini, 2020). Los requisitos mínimos para su instalación son los siguientes: Windows, procesador de doble núcleo, RAM 4 Gb y HD 100 Gb (Leite et al., 2019).

2.8.1.1 CASOS DE ESTUDIO

A continuación, se describen varios estudios de caso que hacen referencia a la aplicación de la herramienta Pentaho, específicamente en instituciones de educación superior:

Análisis de información de aspirantes a ingresar a instituciones de educación superior utilizando Pentaho

Vázquez, J., Frías, R., González, L., & Farías, J. (2018)

Cuando concluye el proceso de admisión de una institución de educación superior (IES), en los registros institucionales como resultado del mismo, se cuenta con información recabada en base de datos mediante el registro y validación de aspirantes. Por la naturaleza del proceso, la recopilación y manejo de datos, dicha base de datos es consistente en datos de rendimiento académico, entre las que se mencionan: resultado de evaluaciones, becas, entre otros, sin perder de vista y tomando en consideración las políticas institucionales de manejo de información y datos personales, se identifica el procesamiento como un área de oportunidad para las entidades de desarrollo académico y seguimiento curricular de dichas instituciones. El presente trabajo consideró una propuesta de análisis utilizando las tecnologías de información y comunicación aplicada en la educación, en el ámbito de inteligencia de negocios utilizando Pentaho como herramienta de procesamiento y obtención de datos estadísticos, basados en procesos y actividades concretos, de donde proviene la propuesta de implementar una alternativa de apoyo a la toma de decisiones enfocada a los departamentos o áreas de las IES.

Business intelligence model using Pentaho for academic management at UNAMBA

Ponce, H. S., Ascue, M. P., & Aroni, J. L. M. (2019).

El estudio describe el proceso de desarrollo de un modelo de Business Intelligence utilizando Pentaho, para ayudar a mejorar el análisis de la información en la Gestión Académica de la Universidad Nacional Micaela Bastidas de Apurímac. Para ello, se implementó un Data Mart para obtener una serie de indicadores académicos y visualizar la información en informes analíticos y gráficos estadísticos. Para el diseño de la base de datos multidimensional se siguieron los lineamientos de la metodología Hephæstus; como solución analítica acorde a las necesidades de información se utilizaron herramientas de procesamiento analítico en línea. Para lograr la implementación se utilizaron herramientas de Pentaho Community como Pentaho Data Integration, Mondrian Server, Schema Workbench y Pentaho Business Analytics 7.0. La implementación del Modelo de Business Intelligence utilizando Pentaho, permitió a los usuarios inmersos en la gestión académica visualizar y analizar el comportamiento de los datos académicos, para que puedan tomar decisiones en base al conocimiento extraído de la información histórica almacenada en la facultad.

Importancia de las herramientas de Business Intelligence, para el análisis de Indicadores de Acreditación

Caicedo, R., Plua, C., & González, A. (2015).

El estudio trata sobre la importancia de utilizar herramientas de Business Intelligence, para el análisis de los indicadores de acreditación, eficiencia terminal, tasa de retención y nivelación del modelo genérico para la evaluación de las carreras presenciales y semipresenciales del Ecuador, para ello se plantea realizar un análisis de la base de datos de la carrera de Ingeniería en Computación y Redes. La investigación se llevó a cabo en la carrera de Ingeniería en Computación y Redes de la Universidad Estatal

del Sur de Manabí, en la cual se tomó como objetivo conocer la importancia de utilizar las herramientas de Business Intelligence Pentaho y RKward para la generación de nueva información, como estrategia para la construcción de indicadores de acreditación. Para el desarrollo de esta memoria se utilizó como metodología instrumentos que se basaron en documentos referativos, se utilizaron técnicas de análisis de bases de datos, observación, encuesta y entrevistas aplicadas a estudiantes y profesores, que ayudaron a justificar el problema de investigación, el cual consistía fundamentalmente en la obtención de nueva información a partir de los datos de la base de datos de la carrera por medio de las herramientas de BI, Pentaho y RKward, para la construcción de indicadores de acreditación.

Analysis and implementation of Pentaho indata control

Sagubay, P., Acosta, C., Wagnio, M., & Guaraca, T. (2018)

El manejo de los datos en su proporción abarca de manera desmedida las estrategias de negocios dentro del entorno actual, las proyecciones administrativas que se enfocan son las necesidades del cliente que se tiene en cuenta al momento de emplear una herramienta que pueda medir los datos para su utilidad y beneficio. La administración estudia las organizaciones y las técnicas que se van a ejecutar en el planteamiento del flujo de información, recursos y usos, según los objetivos propuestos en una estructura que obtenga máxima productividad haciendo de este, un proceso sistemático. Pentaho ofrece un rendimiento eficaz dentro de un mercado competitivo, que ofrece en su funcionalidad la implementación en una organización, lo cual permite interactuar a través de los datos la toma de decisiones que ayuden en el crecimiento, que establezca una evaluación en la ponderación de un resultado. La optimización en la funcionalidad de una herramienta facilita los procesos como un factor estratégico para identificar oportunidades para generar estrategias e implementar la información en la toma de decisiones, para dar soluciones en el mercado de negocios, para definir el control de gastos de recursos y el monitoreo, análisis y administración de la información. En conclusión se debe

establecer un seguimiento del mercado, en las medidas de las necesidades que se desea cubrir en los clientes o los requerimientos que exige el mercado para tener la capacidad de dirigir una empresa y se encamine en la dirección indicada.

Pentaho and Jaspersoft: a comparative study of business intelligence open-source tools processing big data to evaluate performances

Vargas, V., Syed, A., Mohammad, A., & Halgamuge, M. N. (2016)

Independientemente del crecimiento reciente en el uso de herramientas de "Big Data" y "Business Intelligence" (BI), se han realizado pocas investigaciones sobre las implicaciones involucradas. Las herramientas analíticas afectan el desarrollo y la sostenibilidad de una empresa, ya que evaluar las necesidades de la clientela para avanzar en el mercado competitivo es fundamental. Con el avance de la población, procesar grandes cantidades de datos se ha vuelto demasiado engorroso para las empresas. En algún momento del ciclo de vida de una empresa, todas las empresas necesitan crear nuevos y mejores sistemas de procesamiento de datos que mejoren sus procesos de toma de decisiones. Las empresas utilizan los resultados de BI para recopilar datos extraídos de interpretaciones agrupadas a partir de señales en el sistema de información de BI del conjunto de datos que ayuda a las organizaciones con actividades que les otorgan una ventaja en un mercado competitivo. Sin embargo, muchas organizaciones establecen este tipo de sistemas, sin realizar un análisis preliminar de las necesidades y deseos de una empresa, o sin determinar los beneficios y objetivos que pretenden alcanzar con la implementación. Rara vez miden los grandes costos asociados con la implementación de tales aplicaciones, lo que da como resultado estas soluciones impulsivas que no están terminadas o son demasiado complejas e inviables, en otras palabras, insostenibles incluso si se implementan. Las herramientas de código abierto de BI son herramientas específicas que resuelven este problema para las organizaciones que lo necesitan, con el almacenamiento y la gestión de datos. Este artículo compara dos de las herramientas de código abierto de BI mejor posicionadas en el mercado:

Pentaho y Jaspersoft, procesando Big data a través de seis bases de datos de diferentes tamaños, centrándose especialmente en sus procesos de extracción, transformación y carga (ETL) y elaboración de informes mediante la medición de su rendimiento utilizando sistemas de álgebra informática. Los resultados del análisis experimental de ETL muestran claramente que Jaspersoft BI tiene un incremento del tiempo de CPU en el proceso de datos sobre Pentaho BI, que está representado por un promedio de 42,28 % en las métricas de rendimiento sobre las seis bases de datos. Mientras tanto, Pentaho BI tuvo un marcado incremento del tiempo de CPU en el proceso de datos sobre Jaspersoft evidenciado por los resultados del análisis de informes con un promedio de 43.12% sobre seis bases de datos que prueban el punto de este estudio. Este estudio es una guía de referencia para muchos investigadores y profesionales de TI que respaldan las conveniencias del procesamiento de Big Data y la implementación de la herramienta de código abierto de BI en función de sus necesidades.

Business intelligence in education: an application of Pentaho software

de Sá, M., de Souza, S., da Silva, E., Cordeiro, R., & da Hora, H. (2018)

Los sistemas de BI se pueden definir como herramientas para ayudar y extender los procesos de toma de decisiones y hacerlos más precisos y confiables, basados en el conocimiento generado por los datos de la empresa que en valores intuitivos y experiencias personales. Entre los diversos softwares disponibles, Pentaho es un software libre que ofrece buenas posibilidades de exploración de recursos de BI y con posibilidad de personalización. Este trabajo tiene como objetivo presentar una aplicación del software Pentaho demostrando su uso en el contexto de una institución educativa. A través de la creación de un modelo multidimensional y la inserción de datos académicos de una institución educativa en un almacén de datos, la suite Pentaho se configuró para mostrar datos en forma multidimensional.

CAPITULO 3 . CARACTERIZACIÓN Y DIAGNÓSTICO SITUACIONAL

A continuación se realiza la caracterización de la herramienta Pentaho: componentes, productos, arquitectura, ventajas y funcionalidades. Además, se presenta la descripción de la realidad actual en la que se encuentra el ISPIB Canelos en el manejo de información de estudiantes.

3.1 CARACTERIZACIÓN DE LA HERRAMIENTA PENTAHO (BIG DATA)



Figura 3.6 Logo herramienta Pentaho

Parra & Halgamuge (2018) manifiestan que entre las características que presenta Pentaho están las siguientes:

- Compatible con la base de datos JDBC, IBM, DB2, SQL, Server, MySQL, Oracle, PostgreSQL, NCR Teradata, Firebird.
- De código abierto
- Diseño liviano
- Fácil de utilizar
- Flexible
- Instalación y configuración muy sencilla
- Potente
- Presenta una plataforma Java Enterprise Edition asegurando la escalabilidad, integración y portabilidad.
- Provee una interfaz de desarrollo Java SWT, Eclipse.
- Puede correr en distintos servidores compatibles con JEE como JBOSS AS Tomcat, WebLogic, WebSphere y Oracle AS.
- Repositorio de datos basado en XML.

- Sistema operativo de lenguaje interpretado
- Software multiplataforma
- Todos los componentes se comunican mediante Web Services para facilitar la integración con arquitecturas orientadas a servicios.
- Usa lenguaje de programación Java, JavaScript, JSP, XSL.

3.1.1 COMPONENTES

La plataforma integra los siguientes módulos según Leite et al. (2019):

- **Business Analytics Platform:** el servidor que proporciona varios servicios a los usuarios como informes e integración instrumentos;
- **Data Integration:** el ETL de la plataforma (Extract, Transform and Load), también conocido como Kettle, que permite la extracción, transformación y acciones de carga;
- **Report Designer:** la herramienta gráfica que permite diseñar informes y exportarlos directamente a Pentaho;
- **Aggregation Designer:** permite crear y mantener tablas agregadas;
- **Schema Workbench:** la interfaz visual para crear y prueba OLAP (procesamiento analítico en línea) cubos en Mondrian. Es capaz de exportar el cubo directamente en Pentaho;
- **Editor de metadatos:** la herramienta que simplifica la elaboración de informes con experiencia al permitir construir dominios de metadatos y modelos de datos relacionales.

3.1.2 PRODUCTOS

Según Barahama & Wardani (2021) los productos de Pentaho que pueden usarse de manera conjunta o separada según las necesidades de la organización son los siguientes:

- **Pentaho analysis:** permite consultar y analizar la información de un negocio de forma interactiva como seleccionando las diferentes perspectivas de la información en base a los criterios.

- **Pentaho Big Data:** provee las herramientas necesarias con el objetivo de poder analizar grandes volúmenes de datos.
- **Pentaho Data Integration (PDI):** transforma e integra los datos de entre sistemas de información y los datamarts que componen dicho sistema BI.
- **Pentaho Data Mining:** permite descubrir patrones de comportamientos o indicadores ocultos de la información de la empresa.
- **Pentaho Reporting (PRD):** recoge y muestra informes de los indicadores de la empresa, dichos informes pueden ser visualizados en formatos estándar por ejemplo: HTML, PDF y Excel.
- **SparkI:** permite la creación de plugins personalizados enfocados a determinadas funcionalidades y objetivos

3.1.3 ARQUITECTURA

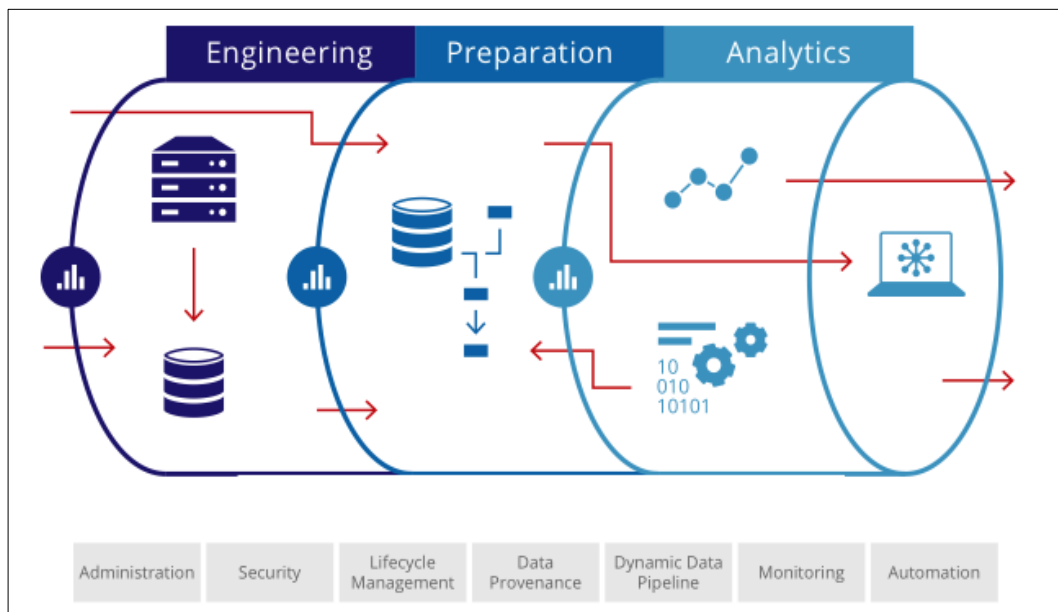


Figura 3.7 Estructura del sistema Pentaho

Fuente: (Sagubay et al., 2018)

La estructura del sistema Pentaho respalda el desempeño de las actividades y funciones que realiza el personal perteneciente a diversas áreas y niveles jerárquicos. Además, se encarga de brindar las facilidades del proceso que son necesarias para contribuir a la carga, operación y explotación de la aplicación desde su particular interés (Ponce et al., 2019).

El ambiente de implementación de Pentaho está basado en Java. Esto hace que Pentaho sea una solución muy flexible para cubrir una amplia gama de necesidades empresariales, los módulos de la plataforma Pentaho BI incluyen: Reporting, Business Analytics, Dashboards, Cloud Business Analytics, Data Integration (de Sá et al., 2018).

3.1.4 VENTAJAS

Entre las ventajas que presenta Pentaho según Parra & Halgamuge (2018) se encuentran las siguientes:

- Arquitectura Open Source que hace fácil su integración con todo tipo de infraestructuras.
- Es compatible con diferentes plataformas.
- Fácil acceso a los datos desde varias fuentes como base de datos o archivos planos.
- fácil integración de datos de diversas fuentes y formatos permitiendo su transformación y limpieza.
- Generación de informes y cuadros de mando mediante herramientas gráficas.
- La generación de reportes es rápida ya que utiliza técnicas para almacenar los datos de memoria cache.
- Los usuarios pueden acceder a visualizar los datos fácilmente.
- Provee herramientas para extraer datos de Hadoop y motores NoSQL y transformar los datos de información relevante.

3.1.5 FUNCIONALIDAD

La funcionalidad de Pentaho se puede resumir en las siguientes herramientas según Leite et al. (2019):

- **Análisis de datos:** mediante base de datos multidimensional.
- **Datos en tablas:** datos generados desde los requerimientos iniciales.
- **Gráficos interactivos:** gráficos predeterminados de la herramienta.
- **Instalación:** se debe configurar algunas variables en el sistema.

- **Reportes:** debe ser instalado de manera individual para su uso.
- **Usabilidad:** se debe tener un conocimiento previo de la herramienta.

3.2 DIAGNÓSTICO SITUACIONAL DEL MANEJO DE INFORMACIÓN SOBRE ESTUDIANTES EN EL ISPIB CANELOS

En función de la normativa que se encuentra vigente sobre el Sistema Nacional de Nivelación y Admisión (SNNA) el listado de estudiantes admitidos después del examen de Ingreso a la Educación Superior se lo recibe en formato de hoja de cálculo (Excel) por parte del departamento encargado de la Secretaría Nacional de Educación Superior (SENESCYT), esta matriz se denomina MTN (Matriz de Tercer Nivel).

La matriz MTN es el insumo fundamental para la gestión de información de los estudiantes ya que con los datos que se encuentra ahí se toma contacto con ellos para que procedan a legalizar su matrícula, este proceso se lo realiza de manera presencial, el estudiante debe llevar para este efecto copia de los documentos personales, copia notariada del título, resultado del examen de tipificación de sangre, Certificado del CONADIS (en caso requiera) entre los más importantes. Estos documentos, constituyen la esencia de la carpeta estudiantil que irá admitiendo los documentos de certificados de matrícula y promociones conforme avance el tiempo.

El estudiante al momento de matricularse llena de forma manual una ficha de matrícula (Figura 3.3) diseñada por el departamento pertinente de la SENESCYT, la misma que contiene información personal del estudiante, con esta ficha el encargado de reportar las matrículas del estudiante tiene que transcribir las respuestas para cumplimentar los requerimientos de registro de estudiantes matriculados.

Además, el estudiante debe llenar una ficha socioeconómica (Figura 3.4) de forma física, esa información se la maneja de manera interna a través de la Coordinación de Bienestar Institucional para gestionar ayudas económicas y/o adaptaciones curriculares si fuera del caso.

1. Tipo de documento de identificación: <input type="checkbox"/> Cédula <input type="checkbox"/> Pasaporte	Formulario Nro. _____
2. Número de documento de identificación: _____	3. Sexo: <input type="checkbox"/> Hombre <input type="checkbox"/> Mujer
4. Género: <input type="checkbox"/> Masculino <input type="checkbox"/> Femenino	
5. Apellidos y nombres del estudiante: _____	
6. Correo electrónico: _____	7. Nro. Celular: _____
8. Nro. Convencional: _____	9. Dirección: _____
10. Código postal: _____	11. En caso de emergencia contactar a: (apellidos y nombres)
12. Parentesco: _____	13. Nro. de contacto: _____
14. Etnia: <input type="checkbox"/> Indígena <input type="checkbox"/> Afroecuatoriana <input type="checkbox"/> Negra <input type="checkbox"/> Mulato <input type="checkbox"/> Mortuño <input type="checkbox"/> Mestizo <input type="checkbox"/> Blanco <input type="checkbox"/> Otro <input type="checkbox"/> No registra	
15. En caso de escoger etnia indígena u otro especifique su nacionalidad: _____	
16. ¿Habla algún idioma ancestral? <input type="checkbox"/> Sí <input type="checkbox"/> No	Especifique: _____
17. Fecha de nacimiento: _____	17.1. Edad: _____
18. Tipo de sangre: _____	
19. País de nacionalidad: _____	20. Provincia de nacimiento: _____
21. Cantón de nacimiento: _____	
22. Categoría migratoria: (en caso de ser extranjero) <input type="checkbox"/> Residente permanente <input type="checkbox"/> Residente transitorio o no residente <input type="checkbox"/> Residente temporal <input type="checkbox"/> Refugiado	
23. País de residencia: _____	24. Provincia de residencia: _____
25. Cantón de residencia: _____	
26. Estado civil: <input type="checkbox"/> Soltero <input type="checkbox"/> Casado <input type="checkbox"/> Unión libre <input type="checkbox"/> Divorciado <input type="checkbox"/> Viudo	
27. ¿Tiene alguna discapacidad?: <input type="checkbox"/> Sí <input type="checkbox"/> No	28. Nro. Carné de Conadis/Ministerio de Salud Pública: _____
29. Porcentaje de discapacidad: _____ %	30. Tipo: <input type="checkbox"/> Auditiva <input type="checkbox"/> Intelectual <input type="checkbox"/> Física <input type="checkbox"/> Mental <input type="checkbox"/> Visual <input type="checkbox"/> NA <input type="checkbox"/> Otra
31. Tipo de colegio: <input type="checkbox"/> Fiscal <input type="checkbox"/> Fisco-misional <input type="checkbox"/> Pariparic <input type="checkbox"/> Municipal <input type="checkbox"/> Extranjero <input type="checkbox"/> No registra	
32. Tipo de Bachillerato: <input type="checkbox"/> Técnico <input type="checkbox"/> Técnico productivo <input type="checkbox"/> BGU <input type="checkbox"/> BI <input type="checkbox"/> Otro	33. Año de graduación: _____
34. ¿Posee algún título de educación superior?: <input type="checkbox"/> Sí <input type="checkbox"/> No	Especifique el título: _____
35. Fecha en la que inició el estudiante la carrera: _____	36. Fecha de matrícula: _____
37. Tipo de matrícula: <input type="checkbox"/> Ordinaria <input type="checkbox"/> Extraordinaria <input type="checkbox"/> Especial	
38. Período / Ciclo académico: _____	39. Año del período académico: _____
40. Nivel académico: <input type="checkbox"/> 1º <input type="checkbox"/> 2º <input type="checkbox"/> 3º <input type="checkbox"/> 4º <input type="checkbox"/> 5º <input type="checkbox"/> 6º	41. Paralelo (A-T): _____
42. Nombre de la carrera: _____	43. Título que otorga la carrera: _____
44. Tipo de carrera: <input type="checkbox"/> Tecnológica <input type="checkbox"/> Tecnología <input type="checkbox"/> Presencial <input type="checkbox"/> Dual	45. Modalidad de la carrera: _____
46. Jornada en la que estudia: <input type="checkbox"/> Matutina <input type="checkbox"/> Vespertina <input type="checkbox"/> Nocturna <input type="checkbox"/> Intensiva	

Dirección: Albalana 67-183 entre Av. Diego de Almagro y Whymper.
 Código postal: 1701818 / Quito, Ecuador
 Teléfono: 593-2-3934-300 / www.educacionsuperior.gob.ec



47. ¿Ha repetido alguna materia? <input type="checkbox"/> No <input type="checkbox"/> Sí	48. ¿Ha repetido la graduación? <input type="checkbox"/> No <input type="checkbox"/> Sí
49. ¿Ha realizado prácticas preprofesionales?: <input type="checkbox"/> Sí <input type="checkbox"/> No	50. Nro. horas de prácticas preprofesionales realizadas: _____
51. Tipo de institución en el que realiza prácticas preprofesionales: <input type="checkbox"/> Pública <input type="checkbox"/> Privada <input type="checkbox"/> ONG <input type="checkbox"/> Otros <input type="checkbox"/> NA	
52. Sector económico en el que realiza prácticas preprofesionales: _____ (escribir el número de acuerdo al detalle)	
1. Agricultura, ganadería, caza y silvicultura 2. Minería y canchales 3. Industrias extractivas 4. Industrias de transformación 5. Comercio por mayor y comercio al por menor de mercancías de todo tipo 6. Comercio por menor de mercancías de todo tipo 7. Comercio por mayor y comercio al por menor de mercancías de todo tipo 8. Comercio por menor de mercancías de todo tipo 9. Actividades de transporte y almacenamiento 10. Actividades de información y comunicación 11. Actividades de servicios financieros e inmobiliarias 12. Actividades de servicios 13. Actividades de servicios de salud 14. Actividades de servicios de alojamiento y alimentación 15. Actividades de servicios de ocio, cultura, recreación y deportes 16. Actividades de servicios de otros tipos 17. Intermedios 18. Actividades de servicios de salud humana y de animales 19. Actividades de servicios de alojamiento y alimentación 20. Actividades de servicios de ocio, cultura, recreación y deportes 21. Actividades de servicios de otros tipos 22. NA	
53. ¿Ha participado en algún proyecto de vinculación con la sociedad en el instituto? <input type="checkbox"/> Sí <input type="checkbox"/> No	
54. ¿Cuál es el alcance del proyecto de vinculación con la sociedad? <input type="checkbox"/> Nacional <input type="checkbox"/> Provincial <input type="checkbox"/> Cantonal <input type="checkbox"/> Parroquial <input type="checkbox"/> I+D+D	
55. El estudiante se encuentra dedicado a: <input type="checkbox"/> Solamente al estudio <input type="checkbox"/> Trabajo y estudio	
56. ¿Cuál es el nombre de la empresa donde labora? _____	
57. Indique cuál es el sector económico de la empresa: _____ (escribir el número de acuerdo al detalle)	
1. Agricultura, ganadería, caza y silvicultura 2. Minería y canchales 3. Industrias extractivas 4. Industrias de transformación 5. Comercio por mayor y comercio al por menor de mercancías de todo tipo 6. Comercio por menor de mercancías de todo tipo 7. Comercio por mayor y comercio al por menor de mercancías de todo tipo 8. Comercio por menor de mercancías de todo tipo 9. Actividades de transporte y almacenamiento 10. Actividades de información y comunicación 11. Actividades de servicios financieros e inmobiliarias 12. Actividades de servicios 13. Actividades de servicios de salud 14. Actividades de servicios de alojamiento y alimentación 15. Actividades de servicios de ocio, cultura, recreación y deportes 16. Actividades de servicios de otros tipos 17. Intermedios 18. Actividades de servicios de salud humana y de animales 19. Actividades de servicios de alojamiento y alimentación 20. Actividades de servicios de ocio, cultura, recreación y deportes 21. Actividades de servicios de otros tipos 22. NA	
58. ¿Para qué emplea sus ingresos económicos el estudiante? <input type="checkbox"/> Sostener sus estudios <input type="checkbox"/> Para mantener a su familia <input type="checkbox"/> Gastos personales <input type="checkbox"/> NA	
59. ¿Usted o algún miembro de la familia recibe el bono de desarrollo humano? <input type="checkbox"/> Sí <input type="checkbox"/> No	
60. Nivel de formación del padre: <input type="checkbox"/> Centro de alfabetización <input type="checkbox"/> Jardín de infantes <input type="checkbox"/> Educación media <input type="checkbox"/> Primaria <input type="checkbox"/> Educación básica <input type="checkbox"/> Superior no universitario <input type="checkbox"/> Superior universitario <input type="checkbox"/> Postgrado <input type="checkbox"/> NA	
61. Nivel de formación de la madre: <input type="checkbox"/> Centro de alfabetización <input type="checkbox"/> Jardín de infantes <input type="checkbox"/> Primaria <input type="checkbox"/> Educación básica <input type="checkbox"/> Superior no universitario <input type="checkbox"/> Superior universitario <input type="checkbox"/> Postgrado <input type="checkbox"/> NA	
62. Ingresos del hogar: \$ _____	
63. Número de miembros del hogar: _____	

Consideración: Con base a lo dispuesto por la SENESCYT para el periodo académico 2023-I, se realizará una matrícula provisional hasta que el estudiante genere el certificado de No adeudar o la Orden de pago correspondiente en el sistema SIAU ONLINE una vez que se encuentre habilitado.

FIRMA ESTUDIANTE: _____

Dirección: Albalana 67-183 entre Av. Diego de Almagro y Whymper.
 Código postal: 1701818 / Quito, Ecuador
 Teléfono: 593-2-3934-300 / www.educacionsuperior.gob.ec



Figura 3.8 Ficha de matrícula

FICHA SOCIO – ECONÓMICA		FOTO
Carrera: Tecnología Superior en Dirección Integral Integral		
Ciclo: _____		
Per. Acad.: _____		
Datos de identificación del Estudiante		
Cédula de identidad: _____	Apellidos: _____	Nombres: _____
Lugar de nacimiento		
País: _____	Provincia: _____	Ciudad: _____
Nacionalidad: _____		
Fecha de nacimiento (d/m/a)		
Se: <input type="checkbox"/> Mestizo <input type="checkbox"/> Blanco <input type="checkbox"/> Indígena <input type="checkbox"/> Montuño <input type="checkbox"/> Negro <input type="checkbox"/> Mulato <input type="checkbox"/> Otro	Estado civil: <input type="checkbox"/> Sol. <input type="checkbox"/> Cas. <input type="checkbox"/> Vdo. <input type="checkbox"/> Div. <input type="checkbox"/> Otro	UI: _____
¿considera usted una persona: _____		
En caso de escoger etnia indígena u otro especifique su nacionalidad: _____		
Datos de identificación del conyuge (liene si es casado o en Unión Libre)		
Cédula de identidad: _____	Apellidos: _____	Nombres: _____
Dirección actual del Estudiante		
Sector de residencia: <input type="checkbox"/> Urbano <input type="checkbox"/> Rural	Provincia: _____	Cantón: _____
Parroquia: _____	Barrio: _____	Tlf. casa: _____
Tlf. cel.: _____		
Si posee estudios anteriores de nivel superior		
Institución: _____	Carrera o Especialidad: _____	Año o Semestre: _____
¿Quién financia los estudios del Estudiante?		
Padres: <input type="checkbox"/>	Conyuge: <input type="checkbox"/>	Estudiante: <input type="checkbox"/>
Familiares: <input type="checkbox"/>	Credito IFCE: <input type="checkbox"/>	Beca: <input type="checkbox"/>
Otros: <input type="checkbox"/>		
Aspectos laborales del Estudiante (liene en el caso de trabajar)		
Actividad laboral: _____	Lugar de trabajo: _____	Sueldo mensual: _____
Dirección del trabajo: _____		Tlf. del trabajo: _____

Parroquia Cañelos – Cantón y Provincia de Pastaza
 Teléfono: 03 2 787046 Email: bienestar@institucaneolos.edu.ec

Datos Económicos de la Familia del Estudiante					
Ingresos mensuales USD					
Sueldo del Padre del estudiante					
Sueldo de la Madre del estudiante					
Pensiones (alimentos, etc.) (yuldares)					
Becas					
Rentas (arriendos)					
Honorarios (de agencias)					
Migración (Herramientas del exterior)					
Otros					
El Estudiante vive con					
Padres	Esposa e hijos	Familiares	Amigo o compañero	Solo	
Estructura del hogar					
Núcleo Funcional (Padre y conyuge)	Núcleo Funcional (Conyuge e hijos)	Miembro Funcional (Padre o conyuge)	Potroto Funcional (Padre o conyuge)	Inngadato Funcional (Padre o conyuge)	
Datos relativos al Núcleo Familiar del Estudiante					
Nombre y Apellido	Parentesco	Edad (Años)	Estado civil	Ocupación (Profesión)	Teléfono
Número total de miembros de la familia que viven actualmente incluido el Estudiante: _____					
Aspectos de la vivienda del Estudiante					
Tenencia de la vivienda	Propia	Arrendada	Prestada		
Tipo de vivienda	Casa	Departamento	Cuartos		
La vivienda dispone de	Luz	Agua	Alcantarillado		
Nº de cuartos	Sala	Cocina	Baño		
Información de Estudiante con discapacidad:					
Tiene alguna discapacidad:	<input type="checkbox"/> Sí <input type="checkbox"/> No				
Tipo de discapacidad:	Física	Auditiva	Visual	Intelectual	
Parentesco:	%	%	%	%	
Número del carné de CONADIS					

Parroquia Cañelos – Cantón y Provincia de Pastaza
 Teléfono: 03 2 787046 Email: bienestar@institucaneolos.edu.ec

Figura 3.9 Ficha socioeconómica

De esta manera, se puede indicar que no se dispone de un sistema de manejo de información automatizado, se utilizan herramientas provenientes de los paquetes tradicionales de ofimática (hoja de cálculo y procesador de textos), cuando se requiere de alguna información específica de algún estudiante se lo hace a través del filtrado de la información en las hojas de cálculo, y esto se lo hace de manera manual.

Los responsables del manejo de la información son: el Rector (representante legal del Instituto), Docente TICs (persona responsable del ensamble de la matriz de matriculados), Coordinador/a de Bienestar Institucional (responsable del manejo de la información de la ficha socioeconómica) y Coordinador/a de Carrera (Manejo de información de los estudiantes matriculados)

Ante lo expuesto, actualmente la responsabilidad en el manejo de información se encuentra en las siguientes personas:

- **Rector:** Ing. Hernán Paucar Estrada Mgs.
- **Docente TICs:** Tlgo. Jaime Murchincho Buri.
- **Coordinadora de Bienestar Institucional:** Lic. Karina Ordoñez Reino MsC.
- **Coordinadora de Carrera de TSDII:** Lic. Cecilia López Huatatoca Mgs.
- **Coordinación de Vinculación:** Lic. Susana Morales Mgs.
- **Coordinación de Investigación:** Lic. Semu Saant Juank.
- **Tesorería:** Lic. Julián Tapuy.
- **Coordinación de Evaluación Interna:** Lic. Bertha Fonseca.

CAPITULO 4 . PROPUESTA PARA LA GESTIÓN Y ANÁLISIS DE DATOS DE ESTUDIANTES CON EL USO DE LA HERRAMIENTA PENTAHO

A partir de la caracterización de la herramienta Pentaho, el diagnóstico situacional del manejo de información de estudiantes en el ISPIB Canelos y el entendimiento que la gestión y análisis de datos constituye una estrategia, ya que las organizaciones confían cada vez más en activos intangibles para crear valor, se presenta la propuesta desarrollada:

4.1 PROPUESTA

La herramienta utilizada fue Pentaho Data Integration Version 9.0, misma que se caracteriza por ser una herramienta JAVA para crear procesos ETL de manera rápida y sencilla, en cualquier entorno. Presenta una interfaz Drag & Drop (arrastrar y soltar) mediante la cual se define el flujo de datos a través de conexión de pasos o steps con flechas. Es compatible con multitud de orígenes de datos y aplicaciones de terceros aplicadas al Big Data (Figura 4.1).

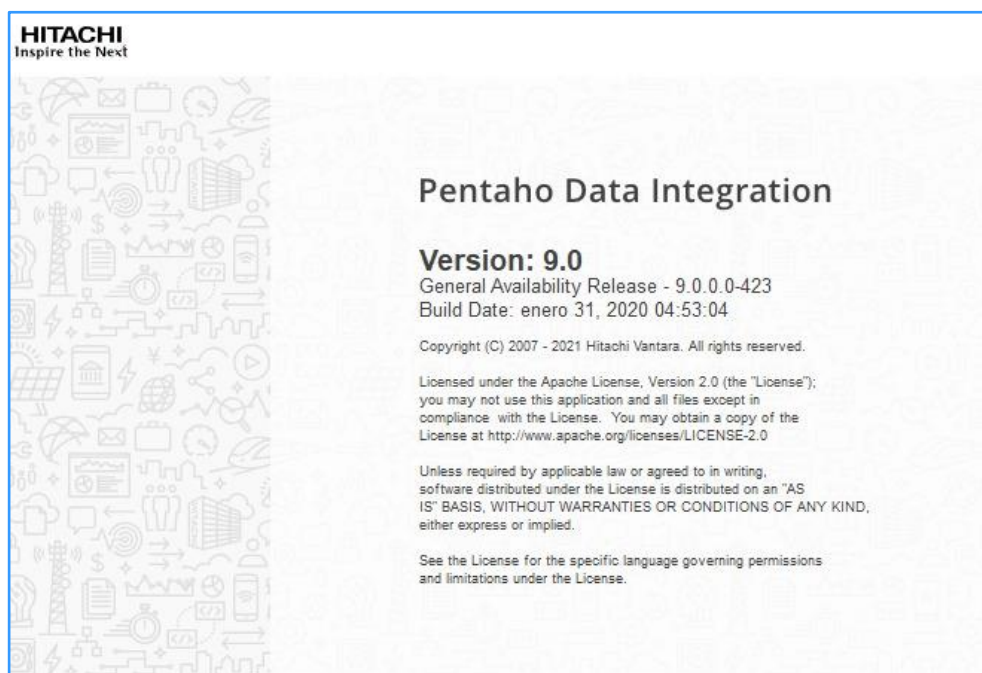


Figura 4.10 Herramienta Pentaho Data Integration Version 9.0

Una vez concluida la instalación y configuración del servicio, se identificaron las áreas de oportunidad que pueden ser impactadas por los resultados de la propuesta, y con el propósito de integrar un resultado tangible y visual, el formato de los reportes, tanto como su contenido, datos y presentación, de esta manera, se considera estar en condiciones para definir con las entidades cliente, en este caso las áreas académicas cuál será, la clasificación y confidencialidad de los reportes (Figura 4.2).

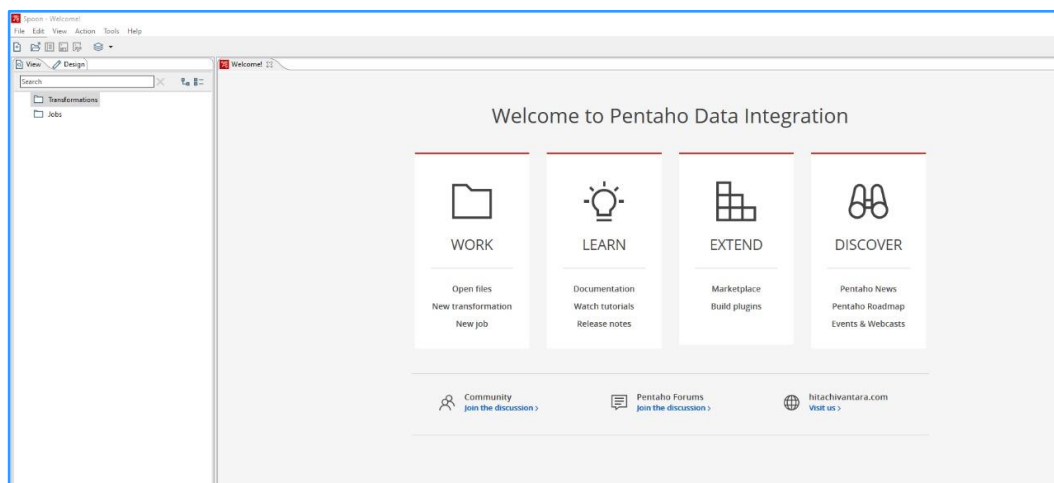


Figura 4.11 Pantalla inicial de la herramienta Pentaho Data Integration Version 9.0

Para la creación de las bases de datos se utilizaron técnicas de limpieza y calidad de datos. No obstante, se definieron políticas y estrategias para la carga inicial del DWH y su respectiva actualización. Asimismo, para crear las dimensiones que sirvieron para crear el DWH, se realizó la carga de información sobre los cubos de (Figura 4.3):

- Estudiantes de centros de apoyo (periodos y matriculas).
- Pagos realizados por estudiante que arrastran materias o cursos realizados dentro del instituto.
- Estudiantes con todas las matrículas en facultades, escuelas, carreras, paralelos, sexo, etc.
- Datos de todas las fuentes
- Periodos y matriculas

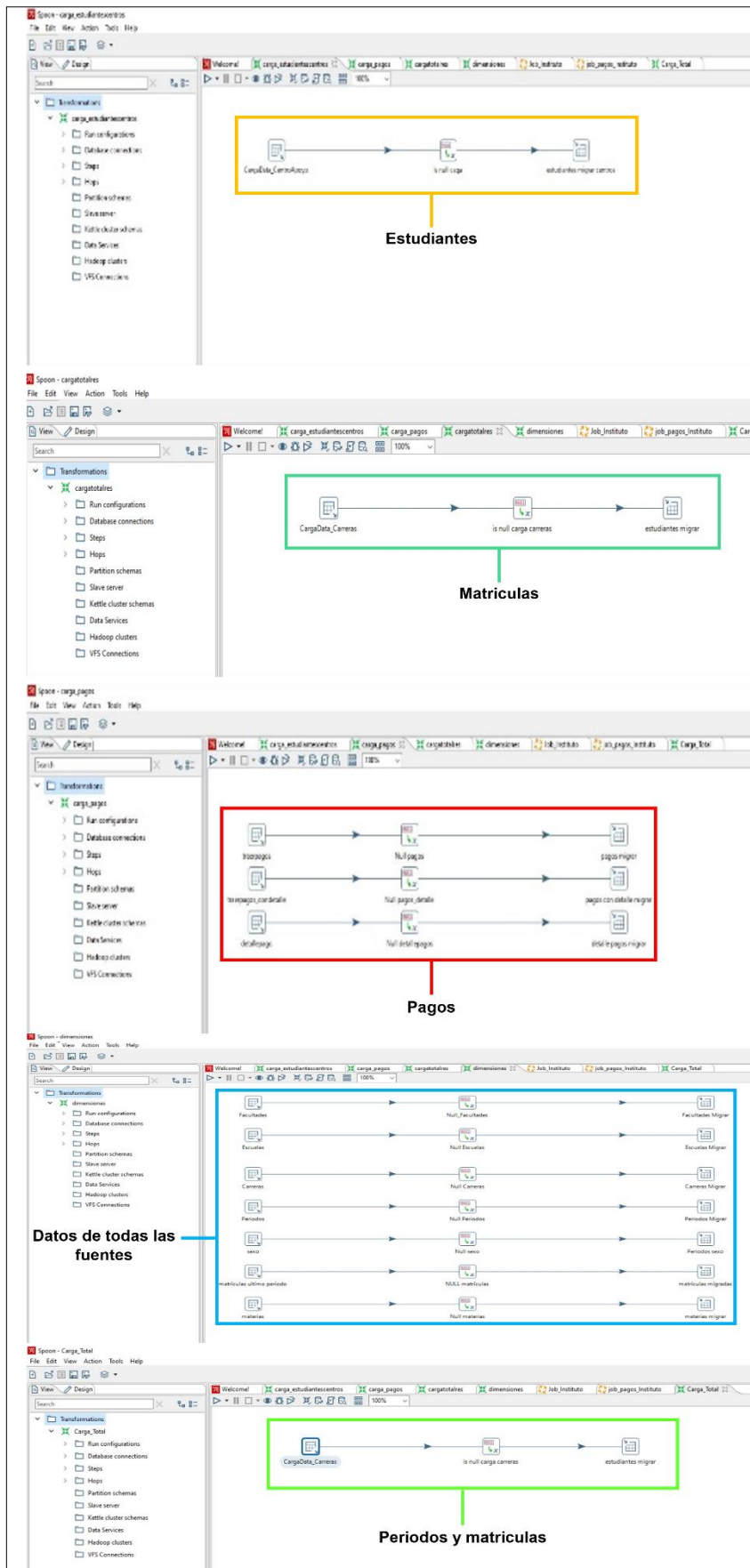


Figura 4.12 Carga de información

Los Jobs son las tareas iniciales que definen el flujo de ejecución de los diversos ETLs. De esta manera, se ejecutarán todos los ETLs creados que componen la migración de datos de los estudiantes del ISPIB Canelos, esto ejecutará cada uno de los ETLs de forma secuencial, por 2 minutos aproximadamente, durante todos los días. En el caso del Job de pagos, existe una actualización frecuente en su ejecución, la cual será 3 veces al día, durante horarios establecidos de 6:00 am, 12:00 pm y 20:00 pm, con el objetivo de obtener información actualizada de los mismo. El proceso de Extracción, Transformación y Carga se realizó en una de las herramientas de la suite Pentaho, Data Integration o Kettle. La herramienta brinda el soporte necesario para realizar todas las tareas de este proceso. En la etapa de extracción, se seleccionaron los datos que se utilizarían en el caso de estudio de la base de datos que se puso a disposición. Se han extraído varios campos que se pueden usar más adelante, pero solo los campos definidos previamente se usarán en el caso de estudio (Figura 4.4).

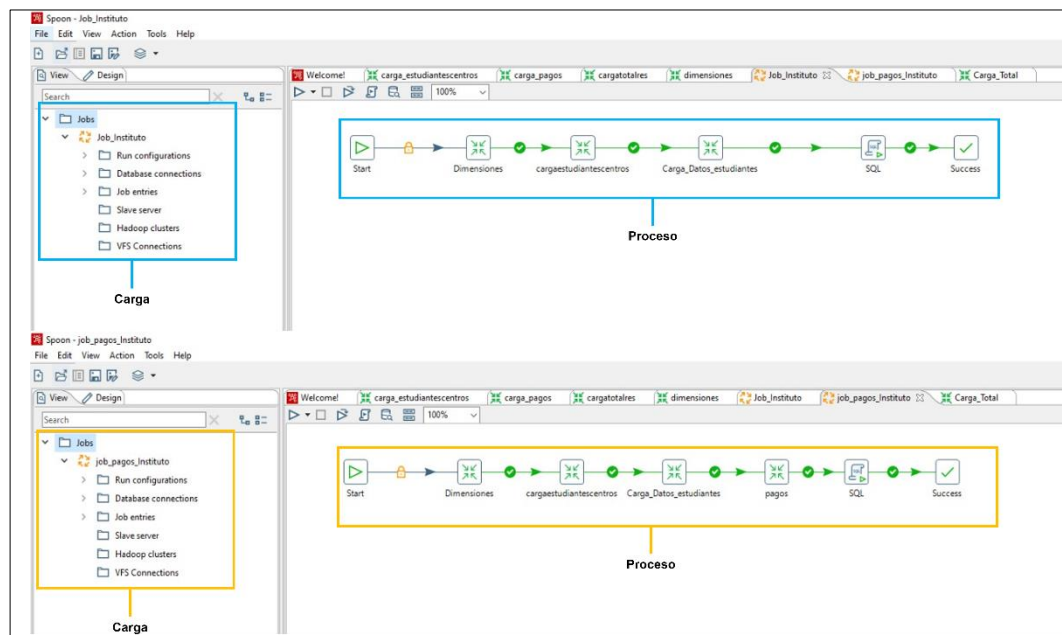


Figura 4.13 Jobs de carga ETL

En el Schema Workbench se crearon los cubos de los datos que se manejan en interfaz para interactuar con los análisis de datos. Estos archivos se generaron en extensión .XML. Es importante manifestar que en

aquí se manejan todos los cubos de datos y dimensiones de los que están compuestas (Figura 4.5).

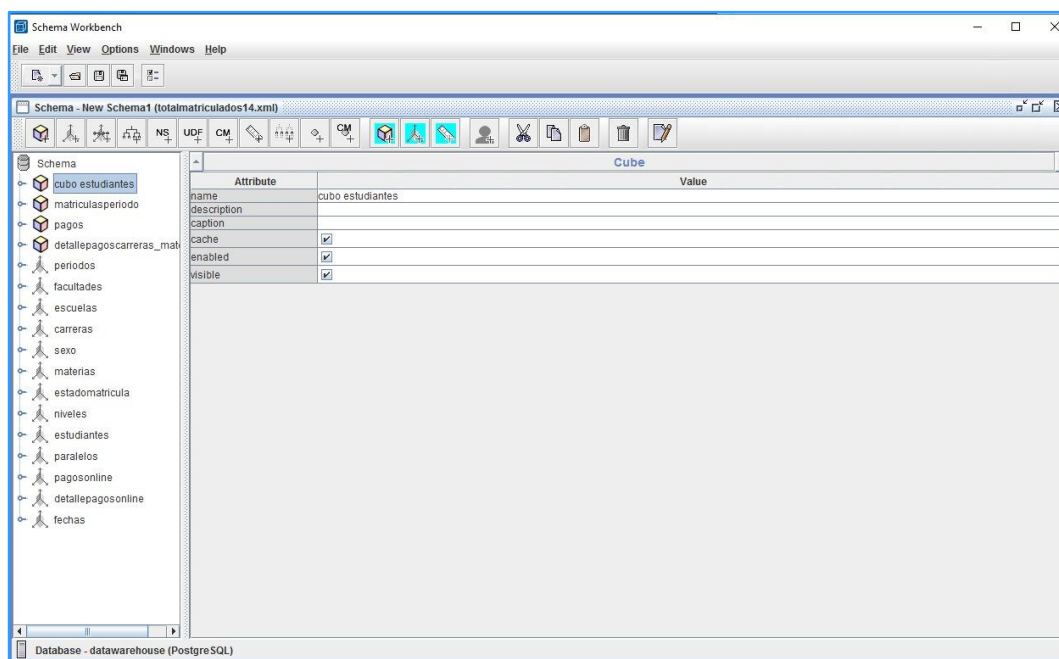


Figura 4.14 Schema Workbench

El sistema de información de estudiantes estará a cargo del administrador que puede acceder a todo el contenido y a los indicadores necesarios para cada uno de los procesos. A continuación, se expone la pantalla de inicio del sistema (ambiente de trabajo para los administradores) (Figura 4.6):

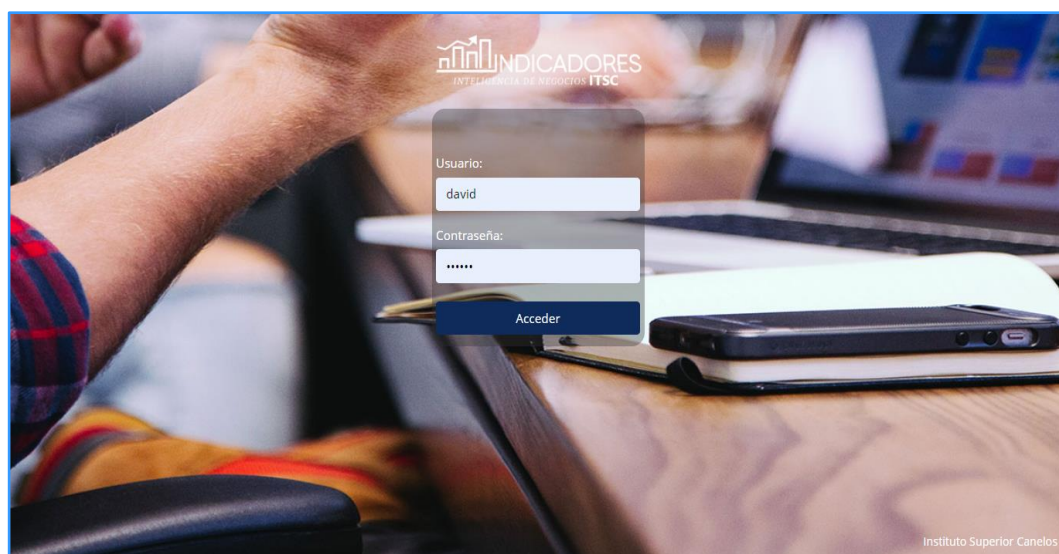


Figura 4.15 Inicio del sistema

Al ingresar al sistema se pueden observar los cubos principales entre ellos: estudiantes, detalle de pagos de carreras o materias, matriculas existentes en el período y pagos (Figura 4.7).

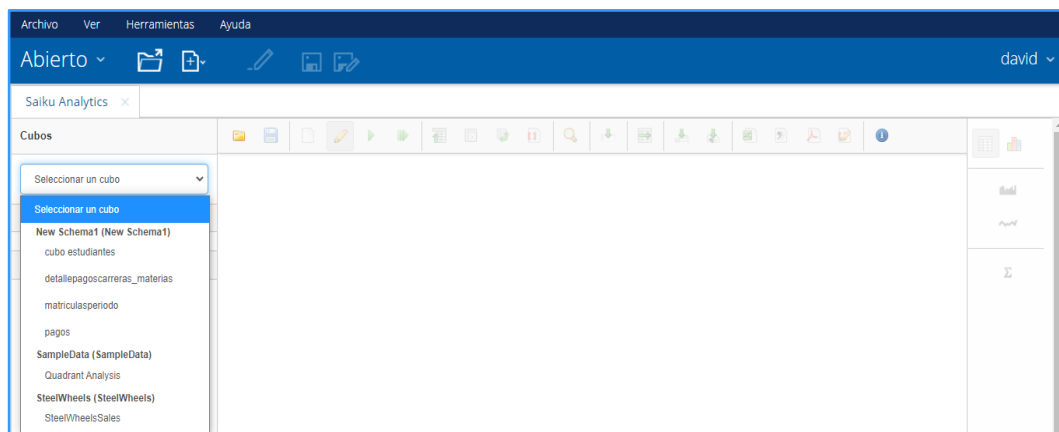


Figura 4.16 Cubos principales del sistema

Dentro de los diferentes cubos se pueden acceder a información como (Figura 4.8):

- **Estudiantes:** carreras, escuelas, facultades, períodos y sexo.
- **Pagos por carreras/ materias:** carrera, fechas, materias, pagos online, períodos.
- **Matriculas:** carreras, escuelas, estado matrícula, facultades, fechas, materias, niveles y paralelos.
- **Pagos:** pagos online, fechas, pagos online, períodos.

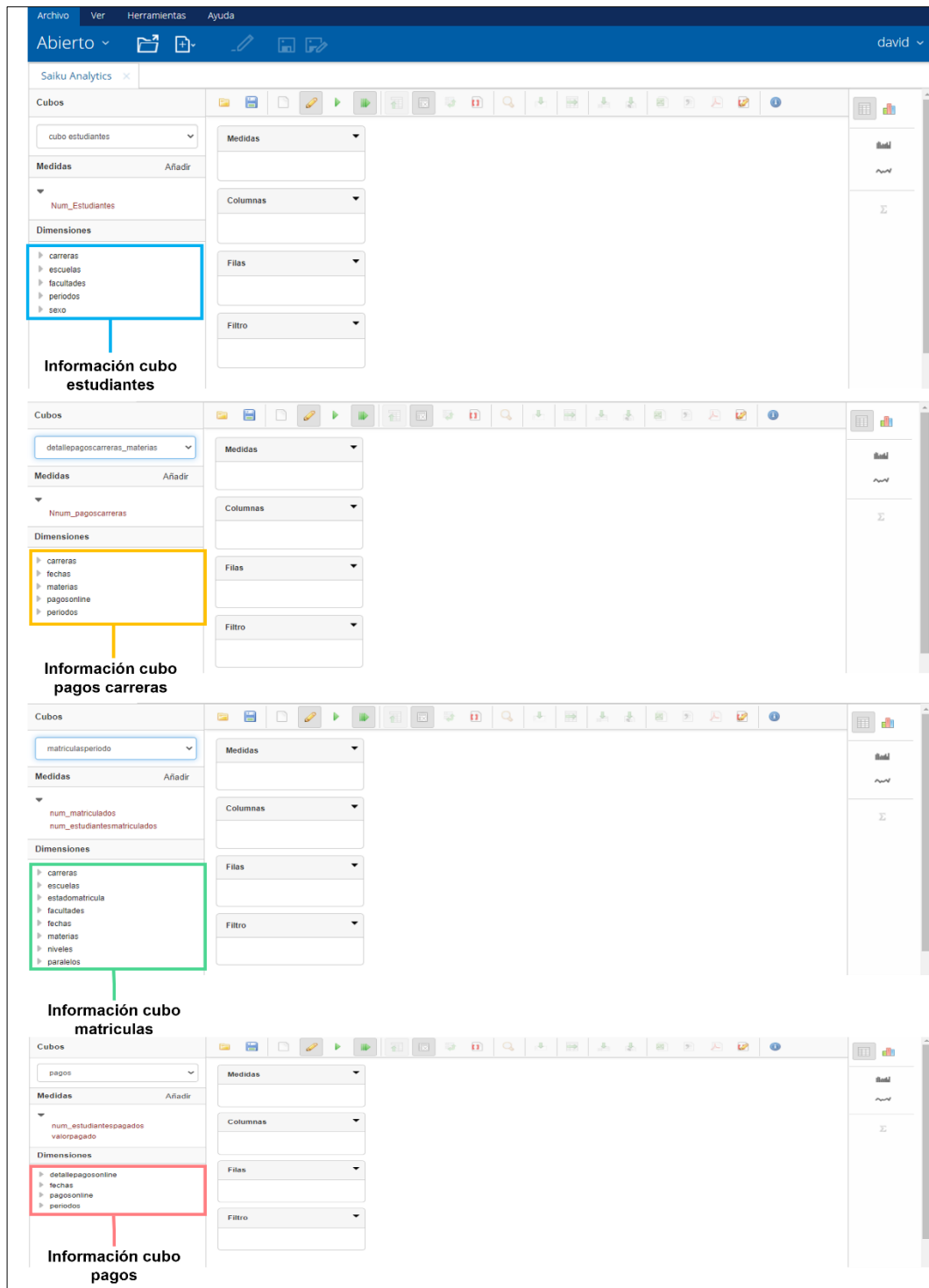


Figura 4.17 Información de los cubos

Los cubos de datos permiten que los datos se modelen y visualicen en múltiples dimensiones donde cada dimensión se implementa mediante la adopción de un modelo de esquema de estrella o copo de nieve. La ejecución de los cubos para realizar los diferentes análisis de información se encuentra distribuido por filas y columnas para poder obtener el

resultado deseado, el análisis depende del identificador, el cual realiza una búsqueda entre toda la Big Data y optimiza el tiempo de respuesta (Figura 4.9).

The screenshot shows the Saiku Analytics interface with a pivot table. The table has columns for 'carreras', 'FEMENINO', and 'MASCULINO'. The rows list various careers such as 'AJUSTE BASICO - FRN', 'AGRONOMIA', 'AGROINDUSTRIA', etc.

carreras	FEMENINO	MASCULINO
AJUSTE BASICO - FRN	440	539
AGRONOMIA	1.111	1.076
AGROINDUSTRIA	1.543	1.217
BIOTECNOLOGIA AMBIENTAL (SEDE MORONA SANTIAGO)	265	177
BIOQUIMICA Y FARMACIA	2.556	1.157
BIOFISICA	1.528	1.803
CICLO FORMATIVO	7.886	5.084
DISEÑO GRÁFICO	679	1.323
BIOQUIMICA Y FARMACIA	14.556	5.050
LICENCIATURA EN DISEÑO GRAFICO	2.311	2.892
EDUCACIÓN Y PROMOCIÓN DE LA SALUD	754	191
GASTRONOMÍA	5.524	5.522
INGENIERIA AUTOMOTRIZ	404	14.235
INGENIERIA AGRONOMICA	3.082	4.074
INGENIERIA EN CONTABILIDAD Y AUDITORIA	17.036	5.144
INGENIERIA DE EMPRESAS	8.200	0.275
CICLO DE FORMACION BASICA	370	1.088
INGENIERIA EN ECOTURISMO	5.938	4.170
INGENIERIA EN ELECTRONICA CONTROL Y REDES INDUSTRIALES	2.912	9.750
ADMINISTRACIÓN DE EMPRESAS	296	300
INGENIERIA ELECTRONICA Y COMPUTACION	5.914	8.862

Figura 4.18 Ejecución de los cubos

Es importante manifestar que se puede filtrar información desde cualquier variable que este dentro de los rangos de la data para poder facilitar el uso del misma (Figura 4.10).

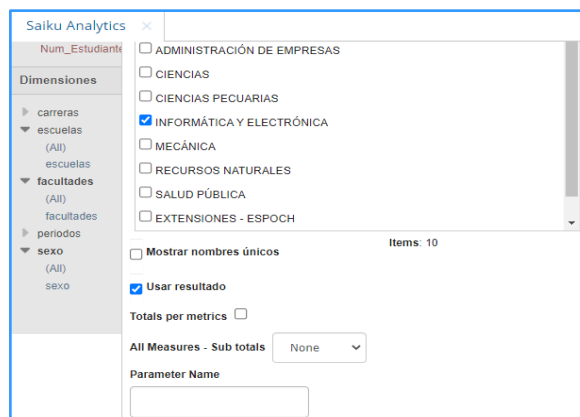
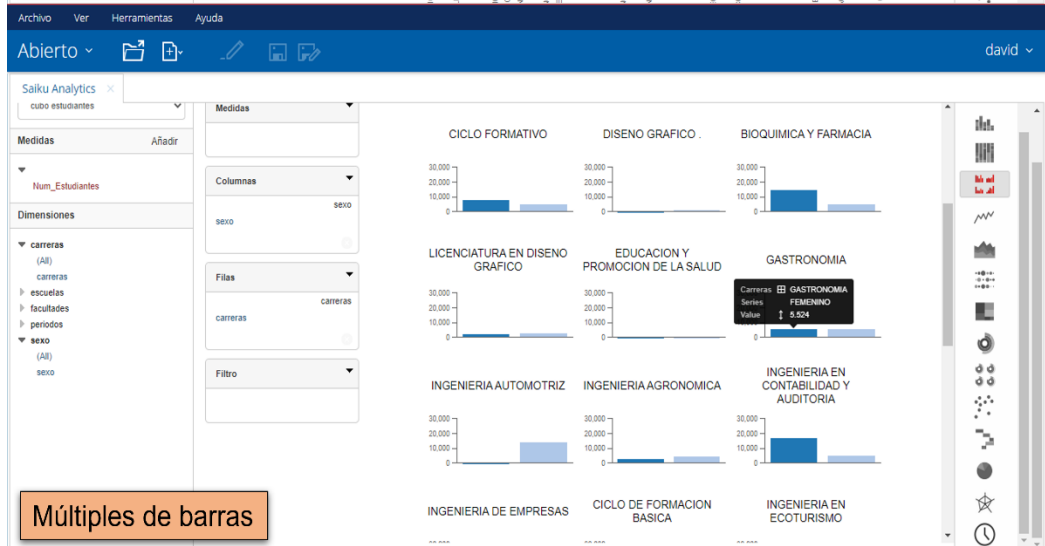
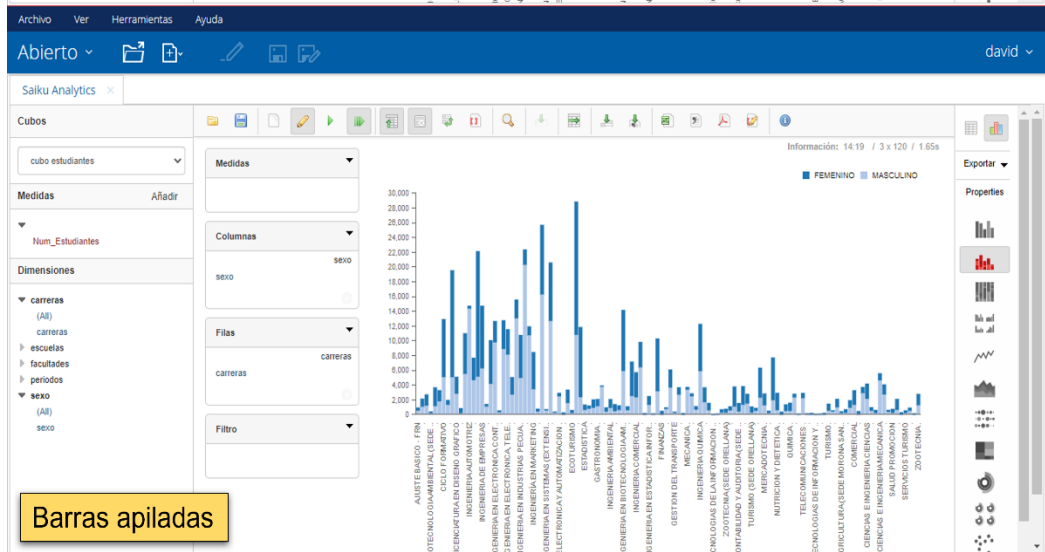
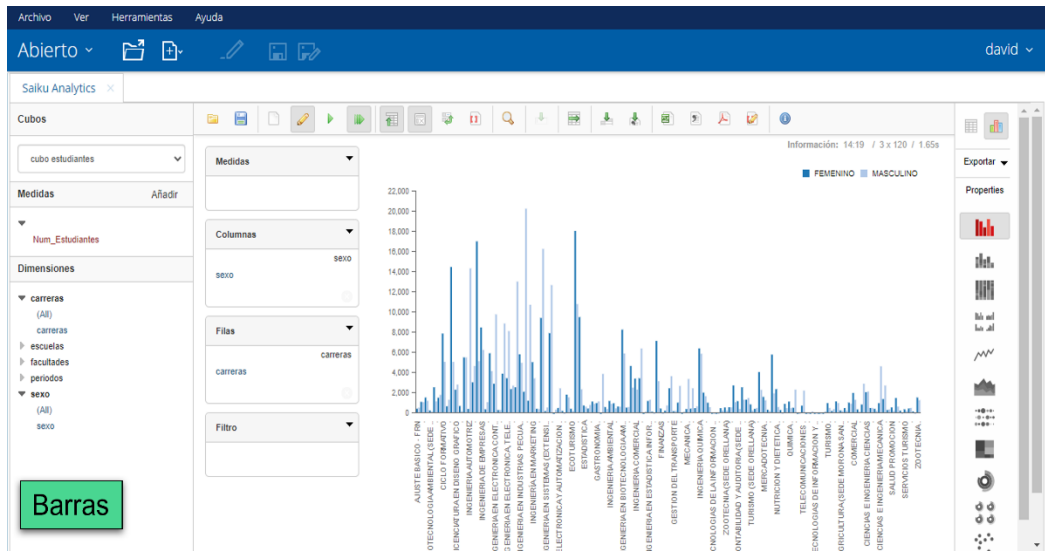


Figura 4.19 Filtrado de información

El resultado de la información se puede representar por medio de diferentes gráficas (barras, barras apiladas, barras 100%, múltiples de barras, áreas, cuadrículas, tree map, líneas, sunburst, multi sunburst, punto, cascada, torta y radar) que se adaptan de manera óptima a los datos analizados. Las gráficas generadas se pueden exportar en archivos PNG y PDF, según la necesidad. En cuanto a la información referente a las gráficas se puede exportar en archivos de Excel (.XLS) (Figura 4.11).



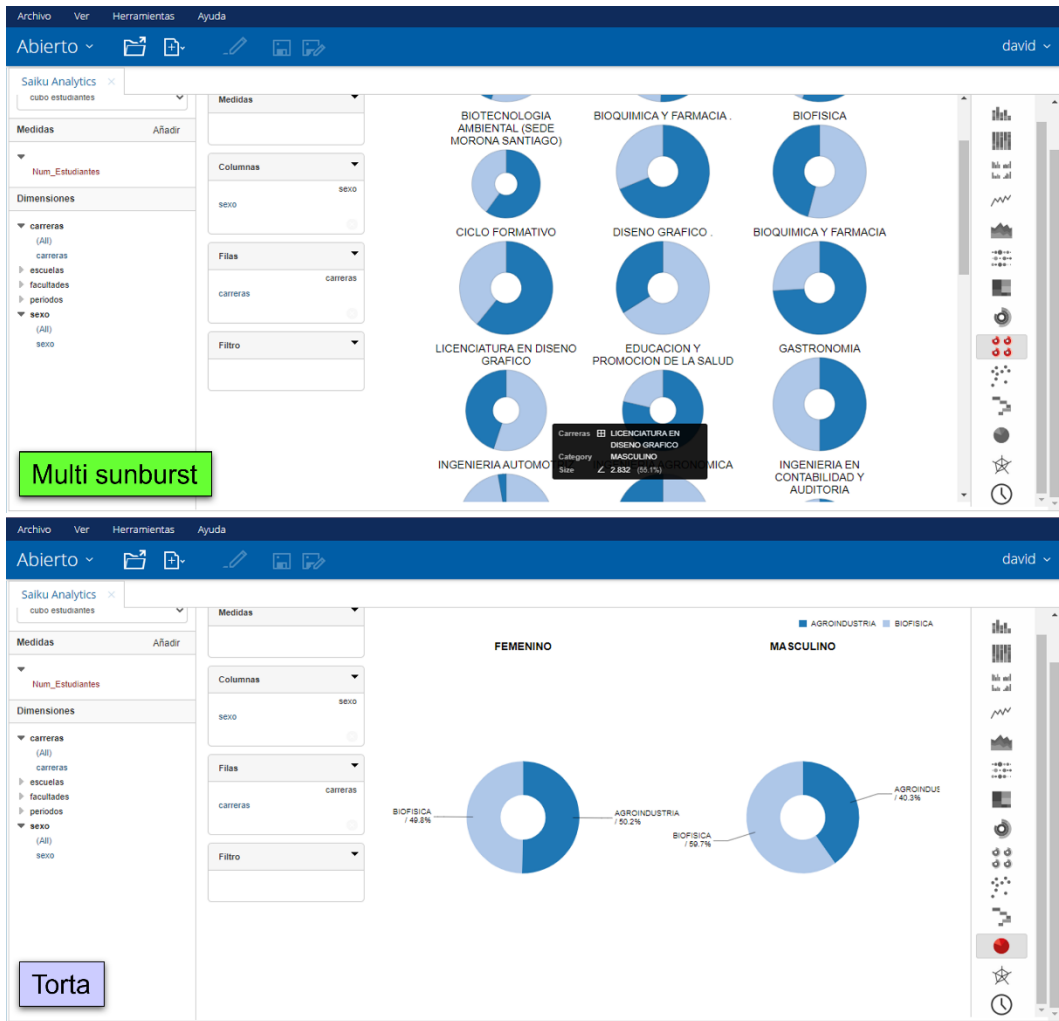


Figura 4.20 Gráficas de información generada

El proceso de la propuesta realizada se resume en la siguiente figura 4.12:

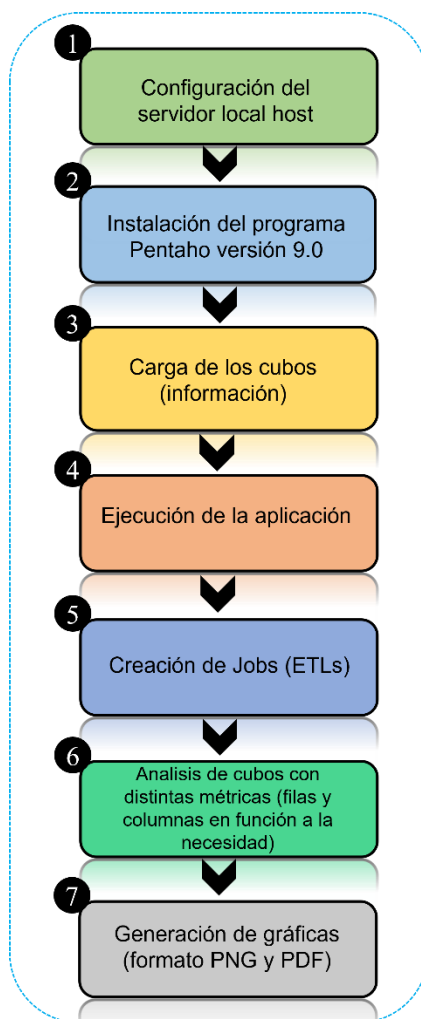


Figura 4.21 Proceso propuesta establecida

Una vez finalizada la propuesta para la gestión y análisis de datos de estudiantes con el uso de la herramienta Pentaho para el ISPIB Canelos se han obtenido los siguientes resultados:

- Generación de reportes personalizados con filtros y gráficos para una mejor visualización de los datos.
- Mejoramiento del acceso a la información sobre estudiantes que posee el instituto.
- Mejoramiento de la organización de la información y ahorro de tiempo al automatizar procesos en el análisis de datos.
- Selección y manipulación de la información de interés para los usuarios y así obtener una mejor toma de decisiones.

- La facilidad en la aprovechamiento de esta información es un factor decisivo para la toma de decisiones de los gestores en el tiempo y momento adecuado, haciendo un diferencial importante para los gestores del área educativa.
- Mejorar la capacidad de compartir, centralizar y agrupar la información de estudiantes.
- Analizar mediante gráficos dinámicos (barras, barras apiladas, barras 100%, múltiples de barras, áreas, cuadrículas, tree map, líneas, sunburst, multi sunburst, punto, cascada, torta y radar, etc.) el comportamiento de los principales indicadores de la institución de manera gráfica e intuitiva.
- Ampliar la visión estratégica, teniendo como base información histórica que permita la generación de conocimiento y la toma de decisiones.
- Tabular y diagramar la información de manera oportuna para los distintos entes que vigilan la entidad de educación superior y no manejar dicha información en archivos planos en Excel que pueden prestarse para manipulación de la información.

4.2 VALIDACIÓN

Posteriormente, a la realización de la propuesta de cómo gestionar y analizar la información de estudiantes, se procedió a realizar una validación dentro del ISPIB Canelos, con la finalidad de determinar la importancia de uso de los nuevos reportes realizados para el control de información de estudiantes, buscando conocer su funcionalidad y beneficios que aportan al personal de la institución encargada de manejar la información. Se obtuvieron los siguientes resultados:

Beneficios que proporcionan los nuevos reportes para el manejo de información de estudiantes

Tabla 4.1 Beneficios de los reportes

Parámetro	Frecuencia absoluta	Frecuencia absoluta agregada	Frecuencia relativa	Frecuencia relativa agregada
Agilitar los procesos	5	5	62,5%	62,5%
Generación de reportes	1	6	12,5%	75%
Consulta de datos	2	8	25%	100%
Total	8		100%	

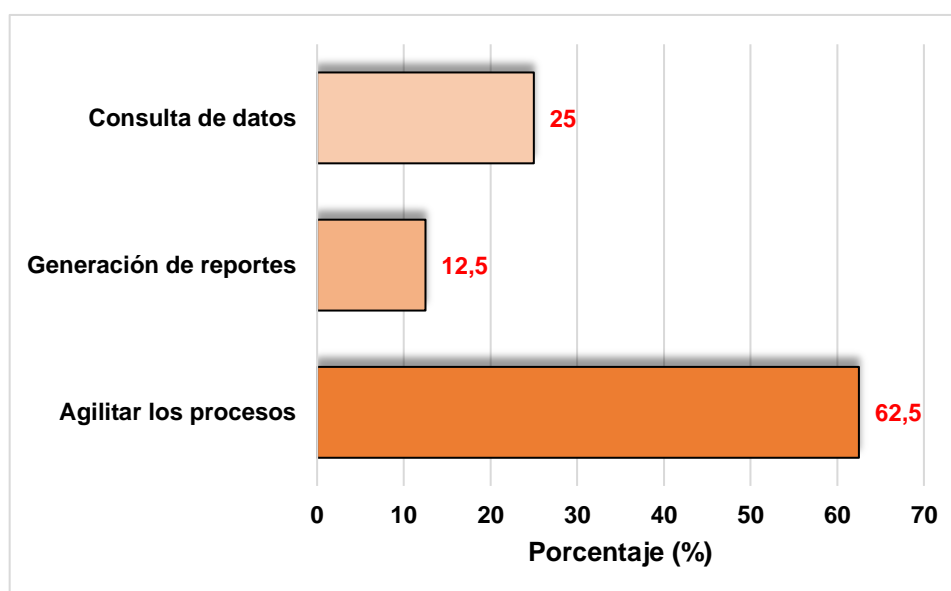


Figura 4.22 Parámetro de evaluación beneficios de reportes

La tabla 4.1 muestra el parámetro de evaluación sobre los beneficios que proporcionan los nuevos reportes para el manejo de información de estudiantes. El 62,5% de los administradores indicaron que la nueva gestión y análisis de datos ayuda agilizar los procesos, el 25% indicaron que ayuda a la consulta de datos y el 12,5% indicaron que pueden generar los reportes (Figura 4.13).

Comprensión de los datos de los reportes

Tabla 4.2 Datos de los reportes

Parámetro	Frecuencia absoluta	Frecuencia absoluta agregada	Frecuencia relativa	Frecuencia relativa agregada
Si	8	8	100%	100%
No	0		0%	
Total	8		100%	

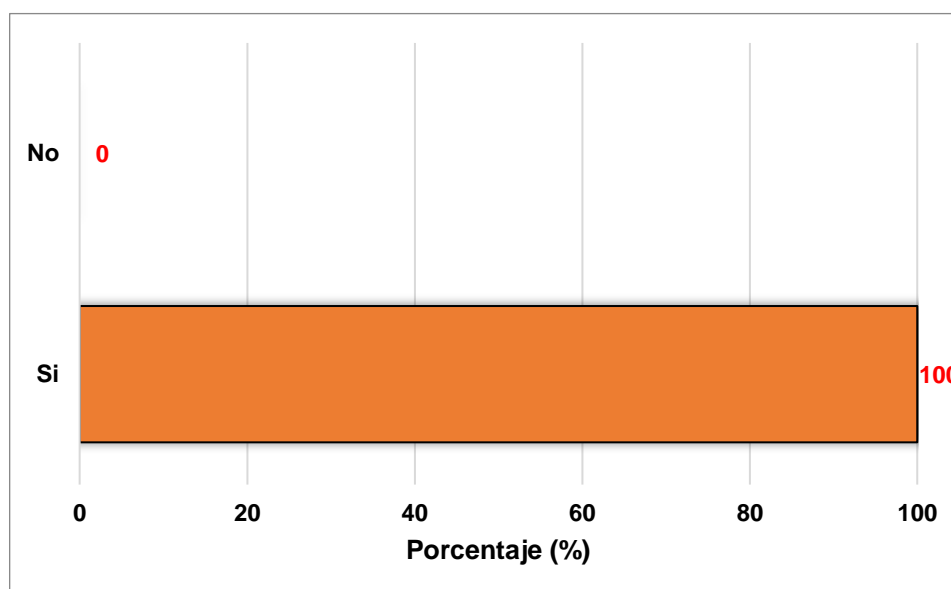


Figura 4.23 Parámetro de evaluación comprensión de reportes

La tabla 4.2 muestra el parámetro de evaluación sobre la comprensión de los datos de los reportes. El 100% de los administradores indicaron que la nueva gestión y análisis de datos si permite comprender los reportes generados (Figura 4.14).

Ayudan de los reportes a la eficiencia y rendimiento de las tareas

Tabla 4.3 Eficacia y rendimiento de los reportes

Parámetro	Frecuencia absoluta	Frecuencia absoluta agregada	Frecuencia relativa	Frecuencia relativa agregada
Si	8	8	100%	100%
No	0		0%	
Total	8		100%	

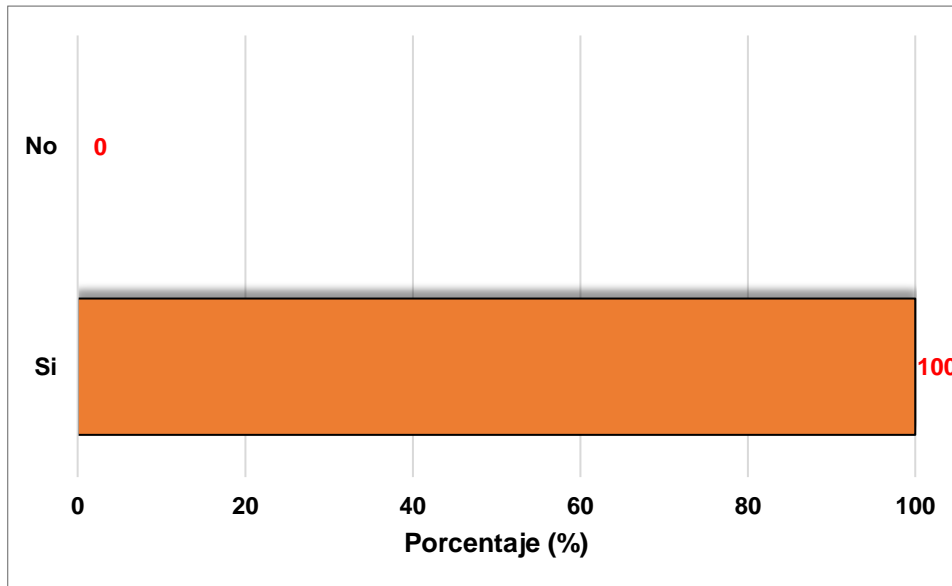


Figura 4.24 Parámetro de evaluación eficacia y rendimiento de reportes

La tabla 4.3 muestra el parámetro de evaluación sobre la ayuda de los reportes a la eficiencia y rendimiento de las tareas. El 100% de los administradores indicaron que la nueva gestión y análisis de datos si permite mejorar la eficacia y rendimiento en el manejo de la información de los estudiantes (Figura 4.15).

Dificultad para manejar los reportes

Tabla 4.4 Manejo de los reportes

Parámetro	Frecuencia absoluta	Frecuencia absoluta agregada	Frecuencia relativa	Frecuencia relativa agregada
Fácil	5	5	62,5%	62,5%
Medio	3	8	37,5%	100%
Difícil	0		100%	
Total	8		100%	

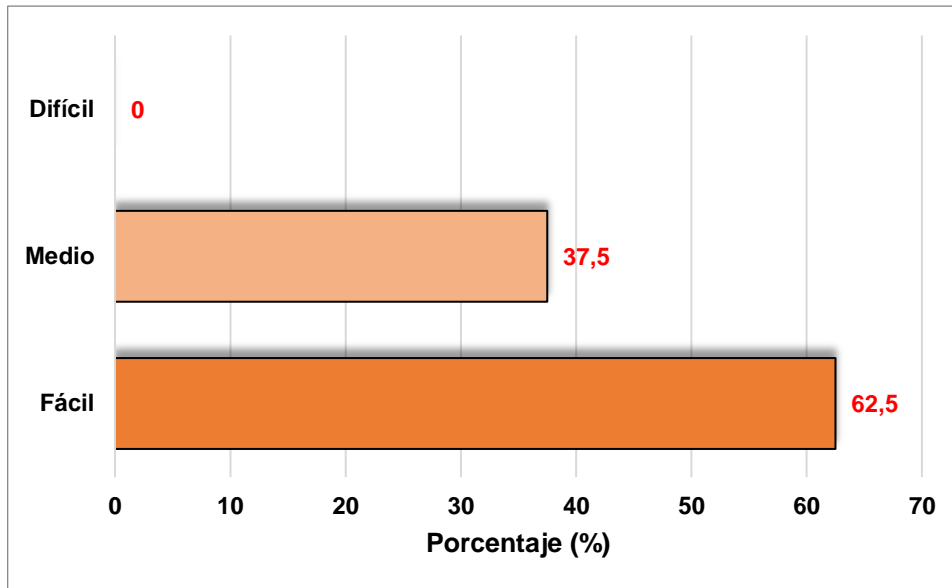


Figura 4.25 Parámetro de evaluación manejo de reportes

La tabla 4.4 muestra el parámetro de evaluación sobre el nivel de dificultad para manejar los reportes en el manejo de información de estudiantes. El 62,5% de los administradores indicaron que la nueva gestión y análisis de datos resulta fácil de manejar y el 37,5% mencionaron que el nivel de dificultad es medio (Figura 4.16).

Ayudan los reportes a mejorar la productividad

Tabla 4.5 Mejora de la productividad con los reportes

Parámetro	Frecuencia absoluta	Frecuencia absoluta agregada	Frecuencia relativa	Frecuencia relativa agregada
Si	8	8	100%	100%
No	0		0%	
Total	8		100%	

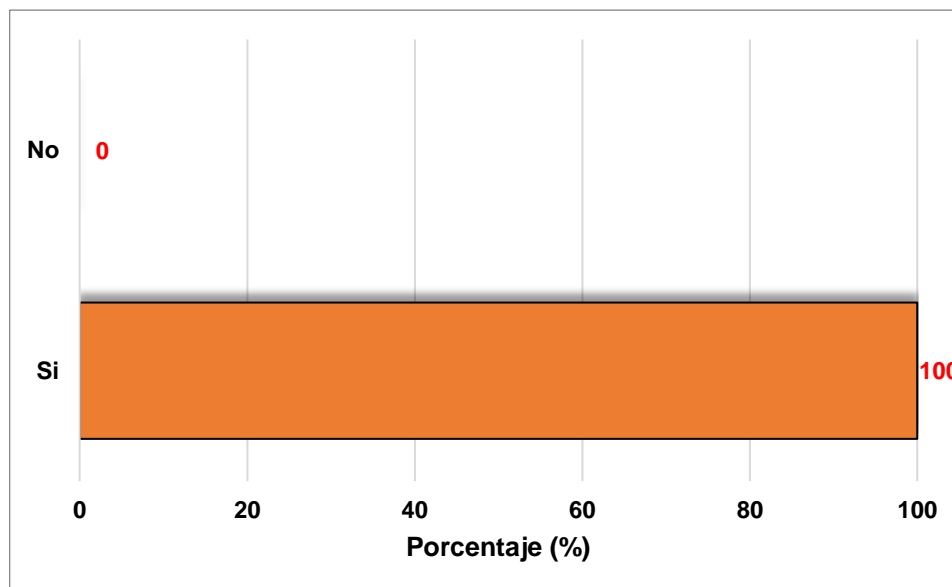


Figura 4.26 Parámetro de evaluación mejora de la productividad con los reportes

La tabla 4.5 muestra el parámetro de evaluación sobre la ayuda de los reportes a la mejora de la productividad. El 100% de los administradores indicaron que la nueva gestión y análisis de datos si permite mejorar la productividad en el manejo de la información de los estudiantes (Figura 4.17).

Ayudan los reportes a conocer información de los estudiantes de la institución

Tabla 4.6 Conocimiento de la información con los reportes

Parámetro	Frecuencia absoluta	Frecuencia absoluta agregada	Frecuencia relativa	Frecuencia relativa agregada
Si	8	8	100%	100%
No	0		0%	
Total	8		100%	

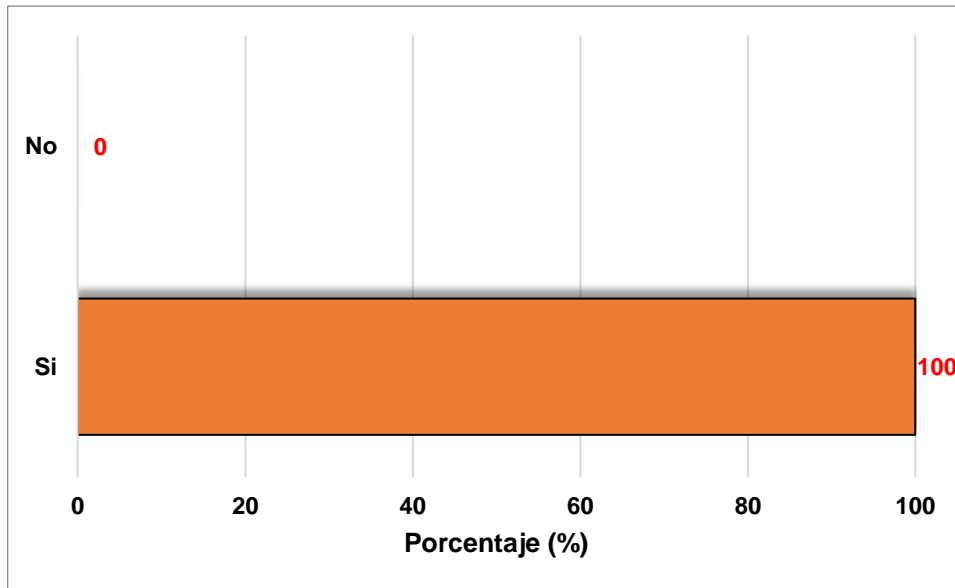


Figura 4.27 Parámetro de evaluación conocimiento de la información con los reportes

La tabla 4.6 muestra el parámetro de evaluación sobre la ayuda de los reportes a conocer información de los estudiantes de la institución. El 100% de los administradores indicaron que la nueva gestión y análisis de datos si permite conocer toda la información necesaria sobre los estudiantes pertenecientes a la Institución (Figura 4.18).

CONCLUSIONES

- En base a la caracterización de la herramienta Pentaho se determinaron los componentes, productos, arquitectura, ventajas y funcionalidad. Lo cual, permite indicar que el software se adapta a las necesidades de la institución, considerando aspectos como costos, rendimiento y usabilidad para el usuario.
- El diagnóstico situacional del manejo de información sobre estudiantes en el ISPIB Canelos permitió establecer que actualmente se lo realiza de forma manual, a través de llenado de las fichas de matrículas y la socioeconómica que se alojan en archivadores para luego ser transcritas en hojas de Excel.
- La propuesta desarrollada para la gestión y análisis de datos de estudiantes con el uso de la herramienta Pentaho, permite generar información (tablas y gráficos) sobre estudiantes, pagos, matriculas, otros. Estos datos permiten profundizar y mejorar la toma de decisiones de una manera analítica y progresiva en el ISPIB Canelos.
- Se identificó que la generación de los análisis con los datos de los estudiantes se presenta como una alternativa interesante para brindar información necesaria e importante, que permite realizar análisis específicos por parte de los gestores educativos.
- La validación de la propuesta sobre gestión y análisis de datos de los estudiantes reveló que los administradores consideran muy importante el uso de los nuevos reportes generados a través de la herramienta Pentaho para agilizar los procesos dentro de la Institución, mejorar la productividad y agilizar el rendimiento de las tareas. Además, los datos indican que los reportes son fáciles de usar y permiten conocer toda la información necesaria referente a los estudiantes de ISPIB Canelos.

RECOMENDACIONES

- Acoger y aplicar la presente propuesta para la gestión y análisis de datos de estudiantes con el uso de la herramienta Pentaho, con la finalidad de formular estrategias y reforzar la capacidad de competir en un entorno de cambios constantes dentro de las instituciones superiores que deben manejar y analizar gran cantidad de información.
- Pentaho se puede utilizar en cualquier entorno en el que sea necesaria una toma de decisiones informada, siempre que se disponga de tecnología y datos relevantes, organizados y fiables. Por tanto, es posible desarrollar nuevos trabajos que investiguen la eficacia del uso de herramientas de BI como ayuda a la gestión en otros entornos no empresariales, para procesamiento y obtención de datos estadísticos, basados en procesos y actividades concretas.
- Diagnosticar permanentemente las necesidades que presenta el ISPIB Canelos en la gestión y análisis de datos para ser apoyados las TICs que permiten la manipulación y análisis que de los mismos.

REFERENCIAS

- Aguilar, L. (2020). *Inteligencia de negocios y analítica de datos*. Marcombo.
- Alvarado, R., Acosta, K., & Buonaffina, Y. (2018). Necesidad de los sistemas de información gerencial para la toma de decisiones en las organizaciones. *InterSedes*, 19(39), 17-31.
- Ballesteros, H., Iñiguez, E., & Velasco, S. (2018). Minería de Datos. *RECIMUNDO: Revista Científica de la Investigación y el Conocimiento*, 2(1), 339-349.
- Barahama, A., & Wardani, R. (2021). Data analysis and data warehouse design based on Pentaho data integration (kettle) to support the determination of student learning achievement. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1098, No. 5, p. 052089). IOP Publishing.
- Bermeo, M., Montoya, L., Valencia A., & Cardona, M. (2020). Incursión de las TIC en la gestión de la información financiera en las empresas pyme comerciales: estudio de caso. *NOVUM*, 1(10), 25-41.
- Billnitzer, A., & Jankovic, J. (2020). Current management of tics and tourette syndrome: behavioral, pharmacologic, and surgical treatments. *Neurotherapeutics*, 17(4), 1681-1693.
- Borja, M., Pérez, M., & Luna, R. (2020). Beneficios ofrecidos por la gestión del Big Data en las instituciones gubernamentales en la era de la digitalización. *Revista La Propiedad Inmaterial*, 30, 93.
- Borse, A., Verma, S., Babu, S., & Kumar, G. (2019). Service oriented architecture paradigm for business intelligence: A survey. *International Journal of Advance Research, Ideas and Innovations in Technology*, 5(1), 196-199.

- Bouaziz, S., Nabli, A., & Gargouri, F. (2019). Design a data warehouse schema from document-oriented database. *Procedia Computer Science*, 159, 221-230.
- Božič, K., & Dimovski, V. (2019). Business intelligence and analytics for value creation: The role of absorptive capacity. *International journal of information management*, 46, 93-103.
- Bravo, L., López, H., & Guerrero, K. (2020). Transformación de la educación frente a la pandemia y la analítica de datos. *Revista Boletín Redipe*, 9(7), 91-99.
- Bravo, L., Molano, J., & López, H. (2021). Analítica académica: nuevas herramientas aplicadas a la educación. *Boletín Redipe*, 10(3), 137-158.
- Buenrostro, H., & Hernández, M. (2019). La incorporación de las TIC en las empresas. Factores de la brecha digital en las Mipymes de Aguascalientes. *Economía: teoría y práctica*, (50), 101-124.
- Caicedo, R., Plua, C., & González, A. (2015). Importancia de las herramientas de Business Intelligence, para el análisis de Indicadores de acreditación. *Revista Universidad de Guayaquil*, 121(3), 41-46.
- Castel, A. (2018). La integración de las TIC en los procesos educativos y organizativos. *Educar em Revista*, 34, 325-339.
- Chen, Y., & Lin, Z. (2021). Business intelligence capabilities and firm performance: A study in China. *International Journal of Information Management*, 57, 102232.
- Cheng, C., Zhong, H., & Cao, L. (2020). Facilitating speed of internationalization: The roles of business intelligence and organizational agility. *Journal of Business Research*, 110, 95-103.
- Chuya, J., Herrera, K., Aguirre, P., & Serrano, L. (2021). Economía digital, herramienta para mejorar la competitividad y productividad en las

- PYMES caso: Machala-Ecuador. *593 Digital Publisher CEIT*, 6(3), 76-86.
- Cobeña, C., Guerra, & Pozo, R. (2018). Los sistemas de información en la toma de decisiones gerenciales en las empresas comerciales de Portoviejo. *ECA Sinergia*, 9(2), 45-54.
- Cordero, E., Erazo, J., Narváez, C., & Cordero, D. (2020). Soluciones corporativas de inteligencia de negocios en las pequeñas y medianas empresas. *Revista Arbitrada Interdisciplinaria Koinonía*, 5(10), 483-514.
- Córdova, A., Taopanta, G., & Rojas, L. (2019). Tecnologías de información y comunicación (tics) aplicadas a las organizaciones empresariales. *Contribuciones a la economía*, (enero).
- Costa, M., Armijos, V., Loaiza, F., & Aguirre, G. (2018). Inversión en TICS en las empresas del Ecuador para el fortalecimiento de la gestión empresarial Periodo de análisis 2012-2015. *Revista Espacios*, 39(47).
- de Pablos, C., Agius, J., Romero, S., & Salgado, S. (2019). *Organización y transformación de los sistemas de información en la empresa*. esic.
- de Sá Mussa, M., de Souza, S., da Silva Freire, E., Cordeiro, R., & da Hora, H. (2018). Business intelligence in education: An application of Pentaho software. *Revista Produção e Desenvolvimento*, 4(3), 29-41.
- Díaz, H., García, V., & Hernández, I. (2020). La vulnerabilidad estructural de la economía mexicana ante la crisis derivada de la pandemia COVID-19. *Contaduría y administración*, 65(5), 2.
- Garcelon, N., Neuraz, A., Salomon, R., Faour, H., Benoit, V., Delapalme, A., & Rance, B. (2018). A clinician friendly data warehouse oriented toward narrative reports: Dr. Warehouse. *Journal of biomedical informatics*, 80, 52-63.

- García, M., Grilló, A., & Morte, T. (2021). La adaptación de las empresas a la realidad COVID: una revisión sistemática. *RETOS. Revista de Ciencias de la Administración y Economía*, 11(21), 55-70.
- Ghani, N., Hamid, S., Hashem, I., & Ahmed, E. (2019). Social media big data analytics: A survey. *Computers in Human Behavior*, 101, 417-428.
- Gunadi, F., & Widiyanto, S. (2020). Perbandingan Data Warehouse Cloud Computing Menggunakan Konvensional Berbasis Kriptografi. In *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)* (Vol. 1, No. 1, pp. 69-73).
- Inmon, B. (2002) *Building the Data Warehouse*. 3rd Ed. Toronto: John Wiley & Sons, Inc.
- Jácome, I. (2020). Psicología organizacional en tiempos de la pandemia COVID-19. *Dominio de las Ciencias*, 6(3), 26-34.
- Lapedra, R., Forés, B., Puig, A., & Martínez, L. (2021). *Introducción a la gestión de sistemas de información en las empresas*. Publicacions de la Universitat Jaume I.
- Leite, N., Pedrosa, I., & Bernardino, J. (2019). Open Source Business Intelligence on a SME: A Case Study using Pentaho. In *2019 14th Iberian Conference on Information Systems and Technologies (CISTI)* (pp. 1-7). IEEE.
- Liang, T., & Liu, Y. (2018). Research landscape of business intelligence and big data analytics: A bibliometrics study. *Expert Systems with Applications*, 111, 2-10.
- Lopes, J., Guimarães, T., & Santos, M. (2020). Adaptive business intelligence: A new architectural approach. *Procedia Computer Science*, 177, 540-545.

- Lv, Z., & Qiao, L. (2020). Analysis of healthcare big data. *Future Generation Computer Systems*, 109, 103-110.
- Machado, G., Cunha, Í., Pereira, A., & Oliveira, L. (2019). DOD-ETL: distributed on-demand ETL for near real-time business intelligence. *Journal of Internet Services and Applications*, 10(1), 1-15.
- Márquez, J. (2020). Inteligencia artificial y Big Data como soluciones frente a la COVID-19. *Revista de bioética y derecho*, (50), 315-331.
- Martinez, A. (2019). La Inteligencia Artificial, el Big Data y la Era Digital: Una Amenaza para los Datos Personales. *Rev. Prop. Inmaterial*, 27, 5.
- Martínez, J. (2018). *Sistema de información de mercados*. Ediciones Paraninfo, SA.
- Meijide, D. (2019). Los ODS en la estrategia empresarial. El caso SUEZ. *Icade: Revista de la Facultad de Derecho*, (108), 14.
- Mendoza, O. (2018). Marco jurídico de la protección de datos personales en las empresas de servicios establecidas en México: desafíos y cumplimiento. *Revista IUS*, 12(41), 267-291.
- Mora, G. (2018). Siglo XXI economía de la información: gestión del conocimiento y Business Intelligence, el camino a seguir hacia la competitividad. *Signos: Investigación en sistemas de gestión*, 10(2), 161-174.
- Muñoz, G., Inda, A., González, M., & Alvarez, C. (2019). Las micro, pequeñas y medianas empresas, una estrategia de aplicación de tecnología para aumentar su competitividad. *Revista Espacios*, 40(20).
- Parra, V., & Halgamuge, M. (2018). Performance Evaluation of Big Data and Business Intelligence Open Source Tools: Pentaho and

- Jaspersoft. In *Internet of Things and Big Data Analytics Toward Next-Generation Intelligence* (pp. 147-176). Springer, Cham.
- Pentaho, "Pentaho Web Page," 2018. [Online]. Available: <http://www.pentaho.com/>. [Accessed: 01-Jan-2022].
- Peñafiel, G., Yáñez, V., Guamán, K., & Padilla, L. (2019). Análisis de metodologías para desarrollar Data Warehouse aplicado a la toma de decisiones. *Ciencia digital*, 3(34), 397-418.
- Pereáñez, J., & Arango, D. (2020). Sistema de gestión y análisis de variables climáticas mediante la utilización de Data Warehouse y Business Intelligence. *Revista Ibérica de Sistemas e Tecnologías de Informação*, (E35), 28-37.
- Pita, G. (2018). Las TICs en las empresas: evolución de la tecnología y cambio estructural en las organizaciones. *Dominio de las Ciencias*, 4(1), 499-510.
- Ponce, H., Ascue, M., & Aroni, J. (2019). Business Intelligence Model using Pentaho for Academic Management at UNAMBA. *C&T Riqchary Revista de investigación en ciencia y tecnología*, 1(1), 20-24.
- Posada, A., & Ampuero, M. (2021). Solución de inteligencia de negocio para la gestión de proyectos utilizando Microsoft y Pentaho. *Revista Cubana de Transformación Digital*, 2(2), 54-70.
- Rangriz, H., & Bayrami, Z. (2019). The impact of E-CRM on customer loyalty using data mining techniques. *BI Management Studies*, 7(27), 175-205.
- Roig, M. (2018). Logística y cadena de suministro en la nueva era digital. *Revista de economía, empresa y sociedad*.
- Ruiz, A., Maya, M., & Franco, A. (2018). Análisis de la formación de administradores desde un enfoque prospectivo: Caso Universidad del Valle–Bugá. *Actualidad contable faces*, 21(36), 151-172.

- Sagubay, P., Acosta, C., Wagnio, ., & Guaraca, T. (2018). Análisis e implementación de Pentaho en el control de datos. *Magazine de las Ciencias: Revista de Investigación e Innovación*, 3(1), 47-58.
- Saraswati, N., & Martarini, N. (2020). Extract Transform Loading Data Absensi STMIK STIKOM Indonesia Menggunakan Pentaho. *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, 19(2), 273-281.
- Sorour, A., Atkins, A., Stanier, C., & Alharbi, F. (2020, February). The role of business intelligence and analytics in higher education quality: A proposed architecture. In *2019 International Conference on Advances in the Emerging Computing Technologies (AECT)* (pp. 1-6). IEEE.
- Spremolla, G. (2017). El trabajo en la era digital. *Revista de derecho*, 16(31), 103-123.
- Tangari, G., Charalambides, M., Pavlou, G., Grazian, C., & Tuncer, D. (2020, June). Classification-assisted Query Processing for Network Telemetry. In *Proceedings of Network Traffic Measurement and Analysis Conference (TMA)*.
- Tejada, F., Murrieta, R., Villao, F., & Garzón, J. (2018). Big Data en la Educación: Beneficios e Impacto de la Analítica de Datos. *Revista Científica y Tecnológica UPSE*, 5(2), 80-88.
- Treviño, R., Rivera, F., & Garza, J. (2020). La analítica de datos como ventaja competitiva en las organizaciones. *Vinculategica Efan*, 6(2), 1063-1074
- Tseng, F., Yeh, C., & Chou, A. (2022). A Collaborative Framework for Customized E-Learning Services by Analytic Hierarchy Processing. *Applied Sciences*, 12(3), 1377.
- Vargas, V., Syed, A., Mohammad, A., & Halgamuge, M. (2016). *International Journal of Advanced Computer Science and Applications*, 7(10), 20-29.

- Vázquez, J., Frías, R., González, L., & Farías, J. (2018). Análisis de información de aspirantes a ingresar a instituciones de educación superior utilizando pentaho. *Pistas Educativas*, 39(127).
- Wang, J., Yang, Y., Wang, T., Sherratt, R., & Zhang, J. (2020). Big data service architecture: a survey. *Journal of Internet Technology*, 21(2), 393-405.
- Wu, Z., Zhou, Y., Wang, H., & Jiang, Z. (2020). Depth prediction of urban flood under different rainfall return periods based on deep learning and data warehouse. *Science of The Total Environment*, 716, 137077.

GLOSARIO DE TÉRMINOS

BI: Business Intelligence

CPU: Central Processing Unit

DWH: Data Warehouse

ETL: Extract, Transform and Load

IES: Institución de Educación Superior

INEC: Instituto Nacional de Estadística y Censos

ISPIB: Instituto Superior Pedagógico Intercultural Bilingüe

JSP: Java Server Pages

KPI: Key Performance Indicator

MTN: Matriz de Tercer Nivel

OLAP: Online Analytical Processing

PDI: Pentaho Data Integration

PRD: Pentaho Reporting

SENESCYT: Secretaría Nacional de Educación Superior

SI: Sistema de información

SNNA: Sistema Nacional de Nivelación y Admisión

TICs: Tecnologías de la Información y las Comunicaciones

XML: Extensible Markup Language

XSL: Extensible Stylesheet Language

ANEXOS

Anexo 1: Encuesta aplicada a administradores



UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL

UNIVERSIDAD CATÓLICA DE SANTIAGO DE GUAYAQUIL

ENCUESTA

Estimado administrador, la presente encuesta tiene como finalidad determinar la importancia de uso de los nuevos reportes realizados para el control de información de estudiantes, buscando conocer su funcionalidad y beneficios en el ISPIB Canelos. Es así que, se solicita comedidamente responderla con la mayor sinceridad.

Responda las siguientes preguntas:

- ¿Cuáles son los beneficios que proporcionan los nuevos reportes para el manejo de información de estudiantes?

Agilizar los procesos	
Generación de reportes	
Consulta de datos	

- ¿Resulta fácil la comprensión de los datos de los reportes?

Si	
No	

- ¿Ayudan de los reportes a la eficiencia y rendimiento de las tareas?

Si	
No	

- ¿Cuál es la dificultad para manejar los reportes?

Fácil	
Medio	
Difícil	

- ¿Ayudan los reportes a mejorar la productividad?

Si	
No	

- ¿Ayudan los reportes a conocer información de los estudiantes de la institución?

Si	
No	

¡Gracias por su colaboración!



Presidencia
de la República
del Ecuador



Plan Nacional
de Ciencia, Tecnología,
Innovación y Saberes



SENESCYT
Secretaría Nacional de Educación Superior,
Ciencia, Tecnología e Innovación

DECLARACIÓN Y AUTORIZACIÓN

Yo, David Hernán Paucar Naranjo, con C.C: # 0604235846 autor del trabajo de titulación: **Gestión y análisis de datos de estudiantes del Instituto Superior Pedagógico Intercultural Bilingüe “Canelos” con el uso de la herramienta Pentaho (Big Data)**, previo a la obtención del título de **Magíster en Telecomunicaciones** en la Universidad Católica de Santiago de Guayaquil.

1.- Declaro tener pleno conocimiento de la obligación que tienen las instituciones de educación superior, de conformidad con el Artículo 144 de la Ley Orgánica de Educación Superior, de entregar a la SENESCYT en formato digital una copia del referido trabajo de titulación para que sea integrado al Sistema Nacional de Información de la Educación Superior del Ecuador para su difusión pública respetando los derechos de autor.

2.- Autorizo a la SENESCYT a tener una copia del referido trabajo de titulación, con el propósito de generar un repositorio que democratice la información, respetando las políticas de propiedad intelectual vigentes.

Guayaquil, a los 11 días del mes de marzo del año 2022

f. _____

Nombre: David Hernán Paucar Naranjo

C.C: 0604235846



Presidencia
de la República
del Ecuador



Plan Nacional
de Ciencia, Tecnología,
Innovación y Saberes



SENESCYT

Secretaría Nacional de Educación Superior,
Ciencia, Tecnología e Innovación

REPOSITORIO NACIONAL EN CIENCIA Y TECNOLOGÍA

FICHA DE REGISTRO DE TESIS/TRABAJO DE TITULACIÓN

TÍTULO Y SUBTÍTULO:	Gestión y análisis de datos de estudiantes del Instituto Superior Pedagógico Intercultural Bilingüe “Canelos” con el uso de la herramienta Pentaho (Big Data)	
AUTOR(ES)	David Hernán Paucar Naranjo	
REVISOR(ES)/TUTOR	MSc. Luis Córdova Rivadeneira; MSc. Edgar Quezada Calle / MSc. Manuel Romero Paz	
INSTITUCIÓN:	Universidad Católica Santiago de Guayaquil	
FACULTAD:	Sistema de Posgrado	
PROGRAMA:	Maestría en Telecomunicaciones	
TÍTULO OBTENIDO:	Magister en Telecomunicaciones	
FECHA DE PUBLICACIÓN:	11 de marzo del 2022	No. DE PÁGINAS: 80
ÁREAS TEMÁTICAS:	TIC, Business Intelligence, Arquitectura, Analítica de Datos, Data Warehouse, Big Data, Pentaho	
PALABRAS CLAVES/ KEYWORDS:	Gestión de datos, Análisis de datos, Pentaho, Big Data, ISPIB Canelos	
RESUMEN/ABSTRACT:	<p>En la actualidad los grandes volúmenes de datos generados por las instituciones han creado nuevos desafíos que son apoyados por las TICs para el manejo y análisis de la información. La presente investigación tuvo como objetivo realizar la gestión y análisis de datos de estudiantes en el Instituto Superior Pedagógico Intercultural Bilingüe “Canelos”, con el uso de la herramienta Pentaho (Big Data). Se aplicaron métodos inductivo, deductivo e hipotético – deductivo con investigaciones documental y de campo. La caracterización del software permitió determinar los componentes, productos, arquitectura, ventajas y funcionalidad. El diagnóstico situacional reflejo que actualmente se realiza de forma manual el manejo de información sobre estudiantes. Posteriormente, en función a la caracterización y el diagnóstico se desarrolló la propuesta para la gestión y análisis de información con el uso de Pentaho, la cual permitió crear información (tablas y gráficos) sobre estudiantes, pagos, matriculas, otros. Finalmente, la validación de la propuesta reveló que los administradores consideran muy importante el uso de los nuevos reportes por agilizar procesos y rendimiento, mejorar productividad y facilidad de uso. En conclusión, la propuesta diseñada permitirá tener información que puede ser utilizada por los gestores educativos como base para la toma de decisiones de una manera analítica y progresiva en el entorno académico.</p>	
ADJUNTO PDF:	<input checked="" type="checkbox"/> SI	<input type="checkbox"/> NO
CONTACTO CON AUTOR/ES:	Teléfono: +593-960802491	E-mail: dhpaucarn@hotmail.com
CONTACTO CON LA INSTITUCIÓN (COORDINADOR DEL PROCESO UTE):	Nombre: Romero Paz Manuel de Jesús	
	Teléfono: +593-994606932	
	E-mail: manuel.romero@cu.ucsg.edu.ec	
SECCIÓN PARA USO DE BIBLIOTECA		
Nº. DE REGISTRO (en base a datos):		
Nº. DE CLASIFICACIÓN:		
DIRECCIÓN URL (tesis en la web):		