



Universidad Cenfotec

Maestría Profesional en Bases de Datos

Documento final de Proyecto de Investigación Aplicada 2.

Evaluación de la factibilidad del uso exclusivo de herramientas de *software* de código abierto para la construcción de cubos de información que permitan realizar análisis de datos con el fin de abaratar costos en licenciamiento en un proyecto de inteligencia de negocios.

Hernández Alfaro, Kenyi

Julio, 2023

Se autoriza la reproducción parcial o total de este documento exclusivamente por motivos académicos.

Dedico esta investigación a mi esposa Diana y a mis dos hijos, Mateo y Andrés, quienes aportaron tiempo en familia para la realización de esta.

TRIBUNAL EXAMINADOR

Este proyecto fue aprobado por el Tribunal Examinador de la carrera: **Maestría Profesional en Bases de Datos**, requisito para optar por el título de grado de **Maestría**, para el estudiante: **Hernández Alfaro Kenyi**.



Digitally signed by JOSE
ALBERTO CABEZAS
JAIKEL (FIRMA)
Date: 2023.07.06
12:19:19 -06'00'

MBD José Cabezas Jaikel
Tutor

MARCO ANTONIO
HERNANDEZ
VASQUEZ (FIRMA)

Firmado digitalmente por
MARCO ANTONIO
HERNANDEZ VASQUEZ
(FIRMA)
Fecha: 2023.07.13 10:15:36
-06'00'

M.Sc. Marco Hernández Vásquez
Lector 1

IGNACIO
TREJOS
ZELAYA
(FIRMA)

Firmado digitalmente
por IGNACIO TREJOS
ZELAYA (FIRMA)
Fecha: 2023.07.13
21:06:59 -06'00'

M.Sc. Ignacio Trejos Zelaya
Lector 2



San José, Costa Rica, 5 de julio de 2023

Tabla de contenido

Tabla de contenido.....	i
Lista de tablas.....	vii
Lista de figuras.....	viii
Resumen	1
Capítulo 1. Introducción	2
1.1 Generalidades	2
1.2 Antecedentes del problema	2
1.3 Definición y descripción del problema.....	2
1.4 Justificación.....	3
1.5 Viabilidad.....	4
1.5.1 Punto de vista técnico	4
1.5.2 Punto de vista operativo	4
1.5.3 Punto de vista económico	5
1.6 Objetivos	6
1.6.1 Objetivo general	6
1.6.2 Objetivos específicos	6
1.7 Alcances y limitaciones.....	7
1.7.1 Alcances	7
1.7.2 Limitaciones	8
1.8 Marco de referencia organizacional y socioeconómico.....	8
1.8.1 Ubicación	9
1.8.2 Recurso humano.....	9
1.8.3 Asesoría externa	9

1.8.4 Productos de TTO Innova	9
1.8.5 Gestión comercial de la empresa TTO Innova.....	10
1.8.6 Historia	10
1.8.7 Tipo de negocio y mercado meta	12
1.8.8 Misión, visión y valores.....	12
1.8.9 Políticas institucionales	13
1.9 Estado de la cuestión	13
1.9.1 Planificación de la revisión	13
1.9.2 Ejecución de la revisión	20
Capítulo 2. Marco conceptual.....	25
2.1 Conceptos relacionados a la investigación	26
Capítulo 3. Marco metodológico.....	28
3.1 Tipo de investigación.....	28
3.2 Alcance investigativo	28
3.3 Enfoque.....	29
3.3.1 Enfoque cuantitativo.....	29
3.3.2 Enfoque cualitativo	29
3.3.3 Enfoque mixto	29
3.4 Diseño	30
3.4.1 No experimental	30
3.4.2 Transversal	31
3.5 Población y muestreo	31
3.6 Instrumentos de recolección de datos	32
3.6.1 Entrevistas	32

3.6.2 Análisis documental	32
3.7 Técnicas de análisis de información	33
3.7.1 Herramientas.....	33
Capítulo 4. Análisis del diagnóstico.....	37
Capítulo 5. Propuesta de solución.....	44
5.1 Detalle de los componentes de la solución	47
5.1.1 Capa de datos. Base de datos MySQL 8.0.....	47
5.1.2 Capa cliente. Pentaho Data Integration 9.3	49
5.1.3 Capa Cliente. Schema Workbench 9.3.....	52
5.1.4 Capa cliente. Metabase 0.45	53
Capítulo 6. Conclusiones y recomendaciones.....	56
6.1 Conclusiones.....	56
Glosario	57
Referencias.....	58

Lista de tablas

Tabla 1. Cálculo salario por hora.....	5
Tabla 2. Costo de horas investigación total.....	6
Tabla 3. Listado de palabras.....	15
Tabla 4. Criterio de inclusión y exclusión de estudios.....	18
Tabla 5. Estudios encontrados en Google Académico.....	21
Tabla 6. Extracción de información inteligencia de negocios en un ambiente de big data.....	22
Tabla 7. Extracción de datos análisis comparativo en inteligencia de negocios. ..	23
Tabla 8. Características de las herramientas encontradas en las búsquedas.	38
Tabla 9. Peso de cada característica que posee una herramienta en investigación.	41
Tabla 10. Peso asignado según sus características para cada herramienta seleccionada en la investigación.....	41
Tabla 11. Proveedores de herramientas.....	46

Lista de figuras

Figura 1. Procedimiento repetitivo para la selección de los estudios.....	20
Figura 2. Ejecución de la selección.	21
Figura 3. Webometrics ranking de la universidad.....	22
Figura 4. Nube de conceptos.	25
Figura 7. Principio de Pareto.....	36
Figura 8. Esquema de infraestructura de la solución.....	47
Figura 9. Modelo entidad relación de la base de datos transaccional.....	48
Figura 10. Modelo en estrella de la base de datos OLAP.....	49
Figura 11. ETL para llenado de la dimensión fecha.	50
Figura 12. ETL para migrar los datos de la base de datos OLTP a OLAP.....	51
Figura 13. ETL para carga de tabla de hechos.....	52
Figura 14. Schema Workbench y el cubo tesis.....	53
Figura 15. Vista de archivo XML en editor de texto.	53
Figura 16. Análisis de la distribución de la tabla de hechos.	54
Figura 17. Análisis de la tabla de hechos según la dimensión de clientes.	55

Figura 18. Análisis de distribución de la tabla de hechos según la dimensión de
empleados. 55

Resumen

En la actualidad, los datos son el activo más importante de las organizaciones. Transformar esos datos en conocimiento es el quehacer de la inteligencia de negocios. Las organizaciones deben invertir tiempo y recursos en proyectos de inteligencia de negocios para ser más competitivos y tomar decisiones de forma más certera. Implementar proyectos en inteligencia de negocios que sean rentables y, de paso, tratar de abaratar los costos por medio de la utilización de *software* libre es una tarea de los responsables de los proyectos en tecnología.

Con esta investigación, se pretende responder a la interrogante: ¿es posible construir cubos de información para el análisis de datos con el uso exclusivo de herramientas de licenciamiento open source? Para solventar la interrogante, se identificaron todas las posibles herramientas disponibles en el mercado; luego, se clasificaron según sus características y, posteriormente, se escogió una herramienta que solvente las necesidades específicas y que se plasmó en una guía la secuencia de cómo construir un cubo de información, haciendo uso de esta tecnología de *software*. Se ha considerado que la solución responda a la interrogante de si se pueda utilizar un grupo de herramientas que éstas sean complementarias.

Palabras clave: investigación, inteligencia de negocios, análisis de datos, cubos de información, código abierto, open source

Capítulo 1. Introducción

1.1 Generalidades

El documento guía resultante de la investigación es puntual en los pasos a seguir por cualquier profesional que desee construir cubos de información para el análisis de datos. La investigación es general y, por sus características propias, es aplicable a cualquier organización que se interese por el uso de herramientas de *software* con licenciamiento de código abierto.

1.2 Antecedentes del problema

El autor de la investigación cuenta con más de quince años de experiencia en el manejo técnico de proyectos relacionados a la temática de inteligencia de negocios. Se ha participado en varios proyectos de desarrollo de *software* para la construcción de cubos de información con el objetivo de realizar análisis de datos. En todos los proyectos en los que ha participado el autor se han utilizado herramientas de desarrollo de *software* que tienen licencias de pago. Estas licencias son emitidas por empresas privadas que las disponen en el mercado actual. Sin embargo, también existen en el mercado programas para el licenciamiento de herramientas de *software* de código abierto que podrían ayudar a alcanzar los mismos objetivos de desarrollo que las herramientas de pago. Realizar la investigación solventó la necesidad de conocimiento del autor y los eventuales lectores de esta investigación y de la guía resultante para la construcción de cubos de información con herramientas de código abierto.

1.3 Definición y descripción del problema

El problema general es la incertidumbre existente de poder utilizar en forma exclusiva herramientas de código abierto para la construcción de cubos de información, con el fin de realizar el análisis de datos. Antes de iniciar la investigación, se desconocía la existencia de herramientas de código abierto que permitan alcanzar los mismos objetivos de construcción de cubos de información que se logra con otras herramientas de licenciamiento de pago. La secuenciación de los objetivos específicos son el cómo, paso a paso, se aclaró la incertidumbre.

La investigación que se realizó fue evaluativa y, a partir de los resultados que se obtengan, se emitió criterio propio acerca de los hallazgos y características de las herramientas evaluadas.

1.4 Justificación

La empresa TTO Innova requiere definir un proyecto de inteligencia de negocios, con la finalidad de recopilar y analizar datos para mejorar los procesos y su relación con los clientes y proveedores. Esto como parte de hacer la empresa más competitiva y rentable en un entorno donde producto de la pandemia y la crisis de las cadenas de abastecimiento ha causado que sea de crucial importancia mejorar la logística y el manejo de datos en todos los niveles.

Por esta razón, se justifica la construcción de cubos de información que permitan realizar análisis de datos. Como se indica, los cubos de información son matrices multidimensionales de datos que permiten el análisis de información para la toma de decisiones. Como manifiesta el sitio Justexw (2021), los cubos “son matrices de datos entendidas donde los datos están estructurados multidimensionalmente” (párr. 2). Sin embargo, la compañía TTO Innova no posee los recursos para comprar un *software* que construya cubos de información. Por ello es que se valora en qué medida se pueden usar herramientas de código abierto de una forma eficiente y satisfactoria. Los resultados de la investigación no solo benefician a la empresa citada, sino que, además, puede servir de parámetros a otros negocios que tengan las mismas necesidades, lo que impactaría en el desempeño empresarial de muchas pequeñas y medianas empresas de Costa Rica que en este momento tienen recursos limitados.

Asimismo, la investigación permite aclarar la duda del autor de esta investigación respecto a la posibilidad de construir cubos de información para análisis por medio del uso exclusivo de herramientas de código abierto, con un menor costo, con respecto al licenciamiento total del proyecto final. También se deseó realizar la investigación, pensando en todas aquellas organizaciones que desean personalizar el *software* que se vaya a construir.

Además, el conocimiento generado es de gran valor para la experiencia del autor de la investigación y le permite abrir nuevos horizontes como consultor independiente, sobre todo, en un momento en el que básicamente las empresas buscan proyectos innovadores que les ayuden a magnificar su actividad empresarial.

1.5 Viabilidad

La empresa TTO Innova ha indicado que tiene una total disposición para brindar la información de su organización, así como la posibilidad de consultar a sus diferentes empleados, pues considera este proyecto importante, ya que, desde hace un tiempo, se ha tenido la preocupación por brindar una solución a este tipo de proyectos.

1.5.1 Punto de vista técnico

Es importante subrayar que se tiene experiencia en *software* de código abierto, ya que el autor cuenta con experiencia en el manejo de herramientas de inteligencia de negocios y conoce las características que debe solventar una herramienta de inteligencia de negocios para la confección de cubos de información.

Además, el investigador posee experiencia de más de quince años, lo que le permite evaluar herramientas de código abierto y compararlas con las características de las herramientas de paga, que proveen empresas privadas. Asimismo, se podrían evaluar las de código abierto que cuenten con versiones de prueba gratuitas para evaluadores.

1.5.2 Punto de vista operativo

El tiempo que se invirtió en la investigación era el necesario requerido por el investigador para solventar las inquietudes que surgieron durante la investigación, no se requirió de la disposición de un tercero para poder darle continuidad a la investigación o poder solventar dudas durante la investigación.

Además, la empresa se manifestó dispuesta a brindar los datos necesarios, ya la empresa está organizada y poseen una serie de directrices que se pueden compartir en forma electrónica. Asimismo, cualquier consulta a los funcionarios se realiza por videollamada o correo electrónico con peticiones que no abarcan mucho tiempo.

1.5.3 Punto de vista económico

La investigación requirió de muchas horas de investigación y evaluación de herramientas de *software* de código abierto. Los costos de esas horas hombre para realizar el trabajo se sufragaron por el mismo investigador, ya que se tenía la necesidad de cumplir con el trabajo final de graduación para la Maestría de Tecnología de Datos de la Universidad CENFOTEC. También el autor deseaba conocer las herramientas disponibles en el mercado de código abierto para la construcción de cubos de información. No se incurrió en costos por licenciamiento, ya que, para evaluar las herramientas, se utilizaron las versiones de prueba de los proveedores de *software* que estaban disponibles en sus páginas web.

El costo estimado de la investigación ascendió a los ocho mil novecientos dólares estadounidenses. Según el Manual de Puestos del Sistema Empresarial de Información Salarial de Price Warehouse Coopers, publicado para el primer semestre del 2021 en Costa Rica.

Tabla 1. Cálculo salario por hora.

Rubro	Monto
Salario anual	₡ 18 602 100
Aguinaldo	₡ 1 550 175
Total salario anual	₡ 20 152 275
Tipo cambio ago 2022	₡ 675,02
Dolarizado	\$ 29 855

Costo hora	\$ 14,35
------------	----------

Fuente: Elaboración propia.

Tabla 2. Costo de horas investigación total.

Consultor	Horas semanales	Duración meses	TFG	Duración TFG horas	Costo estimado de la investigación
Kenyi Hernández	10	12	480		US\$8890

Fuente: Elaboración propia.

1.6 Objetivos

Se utilizó la taxonomía de Benjamín Bloom de 1956. Esta es la taxonomía recomendada por la Universidad CENFOTEC para la definición de objetivos en los trabajos finales de graduación. Durante el periodo de preparación de la presente investigación, se estudió, con adecuada documentación, el motivo por el cual se utiliza esta guía para definir el escalamiento de los objetivos, ya que ha sido el estándar en la educación costarricense y es utilizada por el Consejo Nacional de Educación Superior (CONESUP). Adicionalmente, su estructura jerárquica permite la definición de los objetivos desde lo general hasta lo específico.

1.6.1 Objetivo general

Evaluar la factibilidad del uso exclusivo de herramientas de *software* de código abierto para la construcción de cubos de información que permitan realizar análisis de datos, con el fin de abaratar costos en licenciamiento en un proyecto de inteligencia de negocios.

1.6.2 Objetivos específicos

- Identificar las herramientas de código abierto disponibles en la actualidad que permitan realizar inteligencia de negocios, con el fin de contabilizarlas por medio de búsquedas en sitios especializados en línea.
- Comentar las características de las herramientas de código abierto que permiten la construcción de cubos de información, con el objetivo de construir una base de datos para realizar comparaciones consultando los sitios web publicados por cada proveedor.
- Escoger una herramienta o un grupo complementario de herramientas que permitan la construcción de un cubo de información aplicando el criterio experto de selección, con el objeto de concentrar el foco de la investigación.
- Experimentar la construcción de un cubo de información para comprobar las características de la solución seleccionada poniendo en práctica el funcionamiento de la herramienta o grupo complementario de herramientas escogidas.
- Crear un esquema de infraestructura para la construcción de cubos de información con la finalidad de dibujar un plano final de la solución utilizando una herramienta o un grupo de herramientas que sean de código abierto ilustrando la interacción de las herramientas en la solución.

1.7 Alcances y limitaciones

1.7.1 Alcances

El resultado final de la investigación es encontrar la respuesta a la interrogante de si es posible construir cubos de información utilizando exclusivamente herramientas de *software* de código abierto, además de averiguar qué herramienta se recomienda utilizar, cabe también la posibilidad de que la solución sea utilizar un grupo de herramientas que entre ellas se complementen para construir un cubo. Adicionalmente, se definió un esquema de la infraestructura requerida, en caso de que fuera posible, de cubos de información con herramientas de código abierto. Los lectores también pueden consultar en un documento las características más destacadas de las herramientas encontradas

de código abierto disponibles en el mercado para la construcción de cubos para el análisis de datos. Y, por último, los lectores pueden tener acceso a un cubo de información construido con la herramienta o grupo de herramientas que se seleccionen en la investigación.

Lo anterior implica brindar un reporte de los resultados del estudio a los miembros de la empresa, donde se identifique la viabilidad de la utilización de herramientas de código abierto para construir los cubos de información, así como los costos (no del código de abierto, sino de la contratación de personal profesional para implementarlo en la empresa), junto con un cronograma de los tiempos que se requiere para que la empresa puede aplicar estas herramientas, pues se requiere de un proceso de familiarización. A esto seguirá una breve capacitación al personal de dos horas, tanto con los directivos como con las personas que maneja la parte de las bases de datos.

1.7.2 Limitaciones

Una de las limitaciones es el desconocimiento inicial de la cantidad exacta de herramientas de código abierto disponibles en el mercado para la construcción de cubos de información, el grupo por definir es incierto. En caso de encontrar gran cantidad de ellas, los alcances de la investigación deben ser reevaluados. Además, no se realizará un estudio de costos por licenciamiento para todas las herramientas que se investiguen y la otra limitante es que la implementación de la propuesta depende de las decisiones de la gerencia, el trabajo será un insumo para la toma de decisiones.

1.8 Marco de referencia organizacional y socioeconómico

Según datos de la página web, TTO Innova S.A. fue fundada en el año 2005, por la empresaria Diana Montero Granados, que visualizó una oportunidad de distribuir e importar al país productos provenientes de otros países especialmente calzado para adultos.

El negocio se orientó a ofrecer productos de buena calidad a buenos precios y competir contra empresas que ofrecen productos similares en el

mercado. Inicialmente, se empezó con la venta de productos en línea con distribución a todo el país, debido al éxito de ventas y la aceptación de los productos se abrió la división de ventas al por mayor a detallistas.

1.8.1 Ubicación

La empresa, en sus inicios, se ubicó en Heredia, debido al éxito de la empresa y su crecimiento, se trasladaron a la Trinidad de Moravia, luego de adquirir allí un terreno y construir oficinas y una bodega de almacenamiento de producto.

1.8.2 Recurso humano

Actualmente, trabajan de forma directa con la empresa doce personas y tres asesores en forma indirecta que realizan labores administrativas, legales, contables y financieras, varios de los empleados de ventas tienen más de cinco años de laborar de forma ininterrumpida, lo que muestra la gran estabilidad laboral existente.

La empresa cuenta con gerente general, gerente operativo, jefatura de importaciones, jefatura de ventas y tres vendedores adicionalmente el resto del personal se distribuye ayudantes de seguridad y bodegueros (Empresa TTO Innova, 2019).

1.8.3 Asesoría externa

La empresa es asesorada por otras empresas como Imercka de San José S.A., quien le brinda asesoría en el área de mercadeo y ventas. La empresa Consultores Empresariales la asesora en el área de contabilidad y finanzas. Y Transportes Hidalgo es contratado para realizar la labor de distribución y logística de los productos de la compañía (Empresa TTO Innova, 2019).

1.8.4 Productos de TTO Innova

Dentro de los productos que se comercian se encuentran los zapatos ecológicos UIN, el negocio ha estado orientado a ofrecer productos de buena calidad a buenos precios y competir contra distribuidores de marcas tradicionales como Reebok, Nike o Adidas (Empresa TTO Innova, 2019).

1.8.5 Gestión comercial de la empresa TTO Innova

Distribución.

La empresa realiza la distribución de sus productos por medio de la tercerización del servicio de distribución con siete camiones de reparto para la entrega de sus productos punto por punto en todo el país, una vez realizada la venta, su compromiso de entrega es de 48 horas, esto permite realizar un análisis de crédito y cobro de cada cliente para ver si se puede o no despachar la mercadería (Empresa TTO Innova, 2019).

Ventas.

Las ventas operan bajo la modalidad de preventa, lo que significa que los vendedores visitan a cada cliente en su punto de ventas y, luego de realizar un análisis de inventarios y rotación de producto, le dan un sugerido al cliente y en un periodo de 48 horas se le entrega la mercadería. Se cuenta con una jefatura de ventas y tres vendedores en el terreno que atienden a todos los clientes del país (Empresa TTO Innova, 2019).

Compras a proveedores.

La empresa importa sus productos principales de China (Master Trade Corp.) México (Absormex) y Centroamérica (Alas Doradas del Salvador) y en el mercado local cuenta con varios socios comerciales quienes venden sus productos en la modalidad de distribuidores (Empresa TTO Innova, 2019).

1.8.6 Historia

De acuerdo con Marín (2017), la inteligencia de negocios ha recorrido un largo camino desde sus orígenes, muchas personas han aportado su trabajo a

hacer crecer esta rama de la tecnología y convertirla en lo que hoy conocemos como una herramienta para la gestión de negocios.

Marín (2017) indica que hace aproximadamente sesenta y tres años se conoció la primera persona en acuñar el término *business intelligence*, fue Hans Peter Luhn, un investigador de la empresa IBM, quien en su artículo “A Business Intelligence System” publicó la siguiente definición: “es la habilidad de aprender las relaciones de hechos presentados de forma que guíen las acciones hacia una meta deseada” abre las puertas a un campo de investigación a una de las tecnologías más importantes en la actualidad. En 1962 el canadiense Kenneth Iverson hizo un importante aporte para el futuro de la inteligencia de negocios. Inventa el primer lenguaje de programación multidimensional que se convierte en el primer paso para lo que hoy conocemos como el procesamiento analítico en línea (OLAP). Pero no fue hasta siete años más tarde que se creó el concepto de bases de datos. En la década siguiente, los años setenta se crearon grandes bases de datos y aplicaciones para empresas desarrollando el concepto de inteligencia de negocios.

En la década de los años ochenta, dio paso a la creación de concepto del *datawarehouse* y el lanzamiento de la primera hoja de cálculo. Para la década de los años noventa, se populariza el término *business intelligence* en manos de Howard Dresner y durante ese mismo periodo se crearon múltiples aplicaciones de *business intelligence* que conectaban bases de datos con la información estructurada que generaban las empresas. Con la llegada del milenio, se estrena el concepto *business intelligence 2.0* que, adicionalmente, a la información

estructurada se agrega todos los datos no estructurados y cualquier otro tipo de información. Y hoy en día, ya que las dimensiones de la información son muy extensas y la información crece de forma exponencial, se crea el concepto del *big data* nos relata (Marín, 2017).

1.8.7 Tipo de negocio y mercado meta

La investigación se realizará para la empresa de venta de mercancías minoristas TTO Innova. Pero, como se ha indicado, los resultados benefician a otras organizaciones que pueden mejorar su eficiencia, reducir costos y mejorar su competitividad, todo sin asumir el costo de pagar por licencias.

1.8.8 Misión, visión y valores

1.8.8.1 Misión.

Ser una empresa líder en venta de mercancías minoristas de artículos de ropa y artículos para el hogar en el cantón de San José.

1.8.8.2 Visión.

Llegar a competir en precio y calidad con otras empresas minoristas con una visión de servicio al cliente y uso intensivo de tecnología.

1.8.8.3. Valores.

- Honestidad.
- Búsqueda de la eficiencia.
- Buenas relaciones con los clientes.
- Excelente comunicación con proveedores.

1.8.9 Políticas institucionales

La empresa no ha desarrollado políticas institucionales como tal, sin embargo, puede decirse que su norte ha sido reducir costos, pero brindando productos que tengan buena aceptación y calidad. Además, se tiene un manejo corporativo que incluye una gerencia general, una de recursos y una financiera que tiene un departamento que maneja la información.

1.9 Estado de la cuestión

Investigadores de todo el mundo realizan aportes frecuentes al campo de la inteligencia de negocios. Estos aportes enriquecieron el concepto del *business intelligence* e hicieron que crezca y evolucione con mucha rapidez. Lo podemos notar al encontrar infinidad de estudios y aportes publicados en diferentes revistas de tecnología. Sin embargo, la documentación relacionada con inteligencia de negocios por medio del uso de licencias de código abierto es limitada. En la siguiente sección se desea hacer un compendio de documentos técnicos en relación con la investigación y se realiza su identificación, clasificación y análisis.

1.9.1 Planificación de la revisión

A continuación, se expone una interrogante clara y debidamente depurada del tema a investigación. Se hizo una búsqueda de documentación relacionada al tema con el objetivo de conocer el desarrollo investigativo existente sobre el tema, definiendo áreas claves que se pudieran encontrar y que no se encuentren estudios duplicados en el tema investigativo.

1.9.1.1 Formulación de la pregunta.

Al formular la pregunta, se está enmarcando las posibles búsquedas de información e investigación. El objetivo es encontrar nuevas respuestas que ayuden a contribuir al campo de conocimiento de los cubos de información, sin tener que realizar el pago de licencias.

1.9.1.1.1 Foco de la pregunta.

Se requirió determinar los estudios o publicaciones recientes, de no menor a cinco años, desde su fecha de publicación, que aborden el uso de tecnología de código abierto para el desarrollo de cubos de información importantes para la toma de decisiones.

1.9.1.1.2 Amplitud y calidad de la pregunta.

En esta sección se expone la pregunta a responder queriendo basar la investigación en un problema a resolver. Se enumeraron los términos clave de importancia para la búsqueda de información. Se definieron medidas a utilizar para determinar la respuesta a la pregunta.

1. Problema.

Crear proyectos de tecnología en el área de inteligencia de negocios conlleva la inversión de muchos recursos, en los cuales suele resaltar el tiempo para obtener resultados después de la inversión. Una buena parte del costo del proyecto está en la compra de *hardware* y licenciamiento de *software*. Es donde resulta una buena solución incorporar licenciamiento de *software open source*, para aminorar la inversión monetaria y así poder llevar a conclusión el proyecto.

2. Pregunta.

Dando continuidad a la problemática anteriormente descrita, nació la siguiente interrogante:

¿Qué investigaciones se han publicado recientemente que describieron la creación de cubos de información con herramientas de código abierto?

3. Palabras clave y sinónimos.

Seguidamente, se detalla un listado de palabras clave que se utilizan para realizar las búsquedas de trabajos de investigación o documentos que estén relacionados con la investigación. La mayoría de las palabras utilizadas estaban en el idioma inglés, pues la gran mayoría de publicaciones se encuentra en ese idioma. Estas palabras están descritas en la tabla 3.

Tabla 3. Listado de palabras.

Palabra	Equivalente en idioma inglés.
Código abierto	Open-Source
Cubo	Cube
Datos	Data
Herramientas	Tools
Inteligencia de negocios	Business Intelligence
Proceso Analítico en Línea (OLAP)	On-line Analytical Process (OLAP)

Fuente: Elaboración propia.

4. Intervención.

Recopilar los resultados de todos los estudios de construcción de cubos de información para la toma de decisiones con herramientas que poseen licenciamiento del tipo de código abierto.

Selección de los artículos o documentos más importantes para la investigación y realizar análisis de los resultados obtenidos.

5. Control.

Al iniciar con la investigación, se desconocía la cantidad exacta de herramientas que existen en el mercado para la construcción de cubos de información. Tampoco se conocía sobre algún nombre de herramientas que permitan la construcción de un cubo de información que sea de licenciamiento

open source. La búsqueda se inició a partir de cero, empleando las palabras clave anteriormente descritas.

6. Efectos.

Se tenía la esperanza de poder contar con la documentación suficiente para iniciar la investigación y tener claridad de la cantidad de herramientas disponibles en el mercado para la construcción de cubos de información y que estas herramientas tengan su licenciamiento del tipo de código abierto.

7. Medida de salida.

Se revisa la calidad de la información encontrada en los sitios web dedicados para tal fin.

8. Población.

La población de la investigación son todas las investigaciones a nivel internacional que han publicado sus resultados, luego de construir cubos de información, utilizando herramientas de código abierto.

9. Aplicación.

La presente investigación es de interés para todas aquellas personas que deseen desarrollar cubos de información con el uso exclusivo de herramientas de código abierto, los beneficios pueden ir desde la reducción de costos en licenciamiento a la mejora en los procesos de desarrollo y construcción de cubos de información, así como el descubrimiento de nuevas formas de análisis de información y procesamiento de datos.

10. Diseño experimental.

Cuando se ejecutó el diseño experimental, se realizó un análisis y clasificación de los estudios obtenidos tomando como base la calidad del contenido e importancia para la investigación realizada. Se cuenta con la cantidad suficiente de información para poder respaldar los hallazgos.

1.9.1.2 Selección de fuentes.

En esta sección se identifican las fuentes principales que se consideraron para la investigación.

1.9.1.2.1 Definición del criterio de selección de fuentes.

Es de suma importancia para la investigación contar con fuentes confiables y verídicas de información, por lo cual se verifica que cuenten con la credibilidad de los investigadores y el respaldo teórico y metodológico con que cuenta la investigación. También que sean vigentes y que sean publicados en revistas de prestigio. Si los artículos son muchos, se escogerán los más relevantes.

1.9.1.2.2 Lenguaje de estudio.

Las búsquedas de la investigación se realizaron en idioma español y en inglés para aumentar los posibles documentos resultantes.

1.9.1.2.3 Identificación de fuentes.

A continuación, se describe el proceso de selección de fuentes para clasificar los documentos como pertenecientes a la investigación. También se describe el proceso de ejecución de las búsquedas y la lista de fuentes:

1. Método de selección de fuentes:

En esta sección se comprueba el respaldo con que cuenta la fuente ante la comunidad de investigadores en tecnologías de información y la calidad de la publicación, con respecto a otros documentos de investigación.

2. Cadena de búsqueda:

Se realizan mediante la utilización de las palabras clave “AND” y “OR”, “business intelligence” AND "open source"

3. Lista de fuentes:

Para aumentar la calidad de los resultados obtenidos y seleccionar documentos con una mejor aceptación en la comunidad de investigadores en

tecnología de información, se consideró la utilización de la siguiente fuente de información: IEEE Digital Library.

1.9.1.2.4 Selección de fuentes después de la evaluación.

La fidelidad de las búsquedas resultantes se obtiene de refinar los criterios de búsqueda y de los documentos resultantes en la búsqueda. Es de importancia para la investigación contar con acceso a los datos resultantes después de los criterios de búsqueda.

1.9.1.2.5 Comprobación de las fuentes.

En el momento de iniciar con la investigación, no se contaba con criterio experto para la selección de las fuentes de información; sin embargo, se escogieron las más utilizadas para obtener documentación relacionada con tecnología de datos. Se espera obtener un criterio experto para refinar la lista o agregar más, según sea necesario.

1.9.1.3 Selección de estudios.

Una vez que se han definido las búsquedas, se selecciona los trabajos resultantes del buscador que van a ser parte de la investigación.

1.9.1.3.1 Definición del criterio de inclusión y exclusión de estudios.

Para la selección o exclusión de un estudio, se utilizaron los criterios detallados en la tabla 4. Si un artículo cumple con los requisitos, entonces es considerado en la selección.

Tabla 4. Criterio de inclusión y exclusión de estudios.

Pregunta en investigación	Término principal para criterio de investigación	Criterio de exclusión
---------------------------	--	-----------------------

¿Qué investigaciones se han publicado recientemente que describan la creación de cubos de información con herramientas de código abierto?	"Código Abierto", "Open-Source" "Cubo", "Cube" "Datos", "Data" "Herramientas", "Tools" "Inteligencia de Negocios", "Business Intelligence" "Proceso Analítico en Línea (OLAP)", "On-line Analytical Process (OLAP)"	Documentos sobre herramientas de creación de cubos que sean de licenciamiento de pago.
---	---	--

Fuente: Elaboración propia.

1.9.1.3.2 Procedimiento para la selección de los estudios.

Se aplicó el siguiente proceso iterativo por cada fuente obtenida para la selección de los estudios relevantes:

1. Utilizar la opción de búsqueda avanzada o búsqueda general disponible en las fuentes seleccionadas.
2. Basado en la cantidad de resultados obtenidos, emplear las respectivas cadenas de búsqueda aplicables para obtener resultados considerados de interés o que satisfagan los criterios de inclusión.
3. Si al ejecutar el paso dos, la lista de resultados es superior a los 50 resultados, aplicar filtros o cadenas adicionales para disminuir la lista de resultados, por ejemplo, aplicando rangos de fechas de los estudios.
4. Evaluar los resultados obtenidos y aplicar los criterios de exclusión basados en el *abstract* y palabras clave del artículo.
5. Seleccionar los resultados considerados relevantes para la fuente consultada y repetir el proceso con las demás fuentes disponibles.

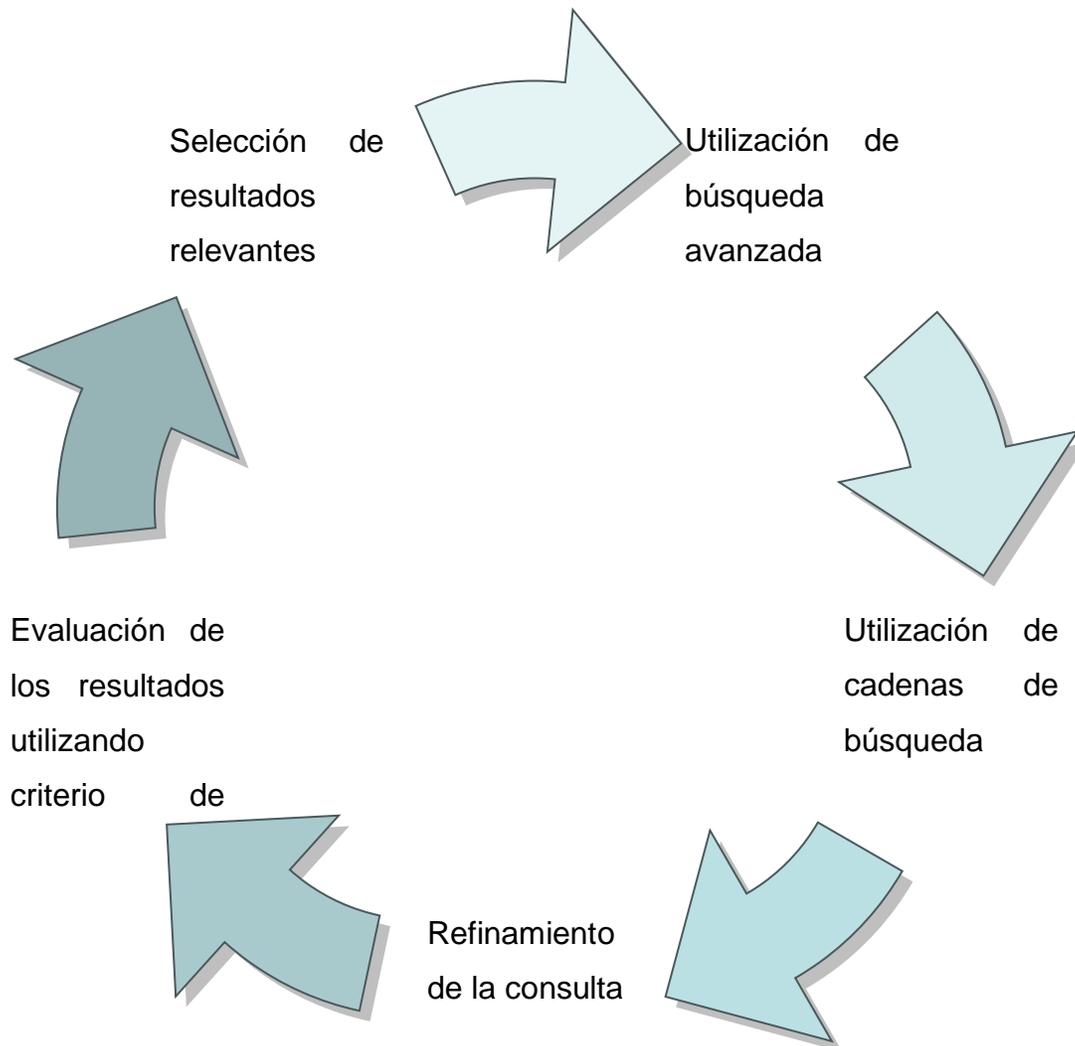


Figura 1. Procedimiento repetitivo para la selección de los estudios.

Fuente: Salazar y Tenorio (2020).

1.9.2 Ejecución de la revisión

En seguida, se muestra proceso de selección llevado a cabo para cada una de las fuentes.

1.9.2.1 Ejecución de la selección en la fuente Google Académico.

1.9.2.1.1 Selección de estudios iniciales.

Utilizando las recomendaciones brindadas por Blanco et al. (2007), se realizan las búsquedas como se muestra a continuación:

Parámetros de búsqueda:

intitle:"open source" intitle:"business intelligence" +"tools"

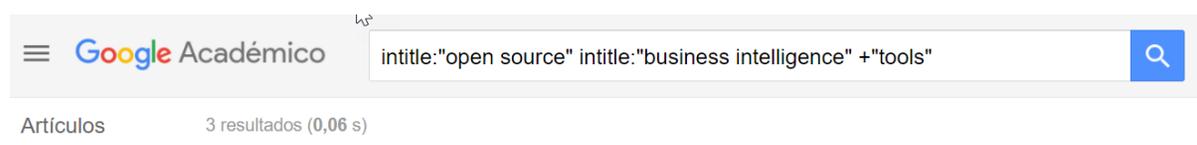


Figura 2. Ejecución de la selección.

Fuente: Google Académico.

Al realizar la consulta en la base de datos de Google Académico, se encontraron tres resultados, de los cuales se seleccionó solamente el primer documento, ya que se utilizó el método de exclusión propuesto.

Tabla 5. Estudios encontrados en Google Académico.

No	Título	Autores	Año	URL
1	Use of open-source tools for business intelligence in an environment of big data	Michal Stora	2017	https://is.muni.cz/th/mnpva/DP_Stora_Archive.pdf

2	A Comparative Analysis on Various Business Intelligence (BI), Data Science and Data Analytics Tools.	Koustubh Sharma, Aditya Shetty, Arnish Jain, Ritesh Kumar Dhanare	2021	https://ieeexplore.ieee.org/document/9402226
---	--	---	------	---

1.9.2.1.2 Evaluación de los estudios.

El primer documento es una tesis de maestría de la Universidad Masaryk, la cual está clasificada como número ocho en el ranking de universidades en Webometrics de Europa Central y del Este (Stora, 2017).

Resultados de la búsqueda

Central & Eastern Europe

Ranking	World Ranking	University	Det.	Country	Impact	Openness	Excellence
8	444	Masaryk University in Brno / Masarykova Univerzita v Brně MU			351	678	605

Figura 3. Webometrics ranking de la universidad.

Fuente:

El segundo documento es una publicación de la IEEE muy reciente sobre herramientas utilizadas en inteligencia de negocios. La reputación de la IEEE es alta entre los ingenieros de tecnología (Sharma, Shetty, Jain y Kumar, 2021).

1.9.2.1.3 Revisión de la selección.

Se seleccionan las publicaciones, luego de la revisión del resumen de cada una de ellas y la relevancia e importancia con respecto a la investigación en curso.

1.9.2.1.4 Extracción de la información.

Para cada documento encontrado que es relevante para la investigación, se extraen sus características más destacadas

Tabla 6. Extracción de información inteligencia de negocios en un ambiente de big data.

Repositorio	Google Académico
Título	Use of open-source tools for business intelligence in an environment of big data.
Publicación	Primavera 2017, Facultad de Informática de la Universidad Masaryk, República Checa, Europa del Este.
Autor	Michal Stora.
Referencia	Stora, B. M. Use of open-source tools for business intelligence in an environment of big data.
Descripción	
Área	Inteligencia de negocios con open-source.
Resumen	En el contenido la tesis de maestría de informática, se menciona los requisitos del sistema en estudio y se compara múltiples soluciones de inteligencia empresarial de código abierto para elegir la herramienta que más se ajusta a un problema en investigación.

Tabla 7. Extracción de datos análisis comparativo en inteligencia de negocios.

Repositorio	Institute of Electrical and Electronics Engineers (IEEE)
Título	A Comparative Analysis on Various Business Intelligence (BI), Data Science and Data Analytics Tools
Publicación	2021 International Conference on Computer Communication and Informatics (ICCCI)
Autores	Koustubh Sharma; Aditya Shetty; Arnish Jain; Ritesh Kumar Dhanare.
Referencia	K. Sharma, A. Shetty, A. Jain and R. K. Dhanare, "A Comparative Analysis on Various Business Intelligence (BI), Data Science and Data Analytics Tools," 2021 International Conference on Computer Communication and

	Informatics (ICCCI), 2021, pp. 1-11, doi: 10.1109/ICCCI50826.2021.9402226.
Descripción	
Área	Herramientas utilizadas en la analítica empresarial y análisis de datos.
Resumen	<p>Los datos juegan un papel muy importante en el crecimiento de cualquier tipo de negocio, ya que, mediante el análisis de los datos, se puede realizar la estructura de trabajo posterior y una adecuada planificación y ejecución. Las decisiones basadas en datos tienen una alta probabilidad de éxito y ayudarán a hacer crecer el negocio. La inteligencia empresarial, así como el análisis de datos, ayudan a analizar los datos para obtener información sobre la dilución. Tanto la inteligencia empresarial como la ciencia de datos giran en torno a los datos. Si bien, la ciencia de datos es el grupo más grande que contiene más información, se puede considerar como parte del panorama más amplio. Además, el alcance de la inteligencia empresarial es limitado. Hoy en día, los términos "inteligencia empresarial" y "análisis" se utilizan hasta cierto punto de manera intercambiable. Ambos términos describen ampliamente la práctica de tomar mejores decisiones comerciales mediante el uso de grandes datos. Pero ¿son las dos tecnologías iguales? La sincronización de BI y Big Data Analytics es necesaria para un mejor proceso de toma de decisiones. Ambos no son lo mismo, pero tienen muchos objetivos comunes. Hoy en día, las empresas de todo el mundo están cambiando rápidamente y las empresas se han dado cuenta de que no solo se ocupan de vender registros o datos estructurados, que de otro modo serían procesados, utilizados, analizados y logrados. BI & Big Data Analytics tiene como objetivo explorar por qué, qué y cómo los clientes, productos, empleados y empresas están involucrados en las ventas. Sin embargo, en este documento puede averiguar cómo las organizaciones cumplen con estas expectativas, cuál es el camino para seguir en BI, los desafíos futuros, las oportunidades y la necesidad de análisis.</p>

descrito, a continuación, se esbozará el marco conceptual del presente documento.

2.1 Conceptos relacionados a la investigación

Código abierto: es de suma importancia definir qué son herramientas de código abierto. Se encuentran definidas de la siguiente manera:

Las herramientas open-source o de código abierto y de *software* libre son softwares desarrollados y distribuidos con una licencia de uso que permite que cualquier persona pueda ver el código y utilizarlo libremente, sin restricciones.

Detrás de este concepto, hay dos movimientos que apoyan el intercambio de ideas y la colaboración entre los expertos en áreas de la evolución de la informática y la tecnología.

Software libre y software de código abierto son conceptos similares, pero no son exactamente lo mismo. En la mayoría de los casos, un software libre es, a su vez, de código abierto, y viceversa. Sin embargo, hay casos de open-source que no son software libre. Detrás de cada concepto hay un movimiento diferente. La diferencia entre ambos radica, sobre todo, en la filosofía de cada movimiento, con unas libertades específicas.

Inteligencia de negocios: el concepto de inteligencia de negocios lo define Rouse (2018) de la siguiente forma:

La inteligencia de negocios o inteligencia empresarial (Business Intelligence, BI) es un proceso impulsado por la tecnología para analizar datos y presentar información procesable para ayudar a los ejecutivos, gerentes y otros usuarios finales corporativos a tomar decisiones comerciales informadas (párr. 1):

Cubos OLAP: los cubos OLAP se pueden definir como, de acuerdo con Microsoft (2021), así:

Un cubo OLAP es una estructura de datos que supera las limitaciones de las bases de datos relacionales y proporciona un análisis rápido de datos.

Los cubos OLAP pueden mostrar y sumar grandes cantidades de datos, al tiempo que proporcionan a los usuarios acceso que permite búsquedas a cualquier punto de datos para que los datos se puedan sumar y segmentar según sea necesario para controlar la variedad más amplia de preguntas que son relevantes para el área de interés de un usuario (párr. 5).

Capítulo 3. Marco metodológico

3.1 Tipo de investigación

Se empleó el tipo de investigación aplicada. Como indica Rodríguez (2020), esta busca resolver una serie de problemas que se presentan en situaciones determinadas. Utiliza conocimientos en forma práctica para resolver necesidades concretas. Rodríguez (2020) añade que: “La investigación aplicada es el tipo de investigación en la cual el problema está establecido y es conocido por el investigador, por lo que utiliza la investigación para dar respuesta a preguntas específicas”.

Como se indica, se requiere adquirir conocimiento para la construcción de cubos de información por medio del uso de herramientas de código abierto realizando la evaluación, comparando las características que destacan los proveedores de *software* en los productos que promocionan y por medio de criterio propio se elaboraron conclusiones y recomendaciones. Así también se construye una guía para que cualquier desarrollador de *software* pueda realizar la construcción de una solución de inteligencia de negocios utilizando *software* libre disponible en el mercado.

3.2 Alcance investigativo

Es una investigación exploratoria, como expresan Hernández, Fernández y Baptista (2014), se trata de aquellas donde se busca dar respuesta a fenómenos novedosos. Como es una primera aproximación, se procura tener resultados no concluyentes, pero que pueden servir para resolver una situación concreta y para investigaciones posteriores.

La revisión de la literatura ha demostrado que hay poca documentación con respecto al tema de la realización de cubos de información con herramientas de *software* libre. Adicionalmente, previo a la realización de la investigación, se desconocía la cantidad exacta de herramientas disponibles en el mercado que sean de código libre. Por lo tanto, el alcance investigativo que se realizó se cataloga como exploratorio. El estudio conllevó muchas horas de investigación y

comparación de características del *software*, esto permitió llegar a conclusiones que luego requieren una comparación más exacta, aunque se trata de que las conclusiones sean lo más acertadas posible.

3.3 Enfoque

En esta sección se pretende dar a conocer el enfoque que se desea aplicar; en este caso, es la investigación mixta, que combina el enfoque cuantitativo y cualitativo. Para una mejor explicación se definen individualmente:

3.3.1 Enfoque cuantitativo

El enfoque cuantitativo se define como un conjunto de estrategias para recolectar y procesar información, realizando un análisis de métodos científicos y numéricos, así como empleando las herramientas del ámbito de las estadísticas, en relación con las causas y efectos de la investigación. Hernández y Mendoza (2018) señalan que este enfoque está asociado con explorar y definir fenómenos, variables o hechos, así como comparar desarrollos tecnológicos para resolver problemas.

3.3.2 Enfoque cualitativo

El enfoque cualitativo consiste en estudiar a las personas, en función de sus experiencias, de lo que dicen y hacen en contextos sociales y culturales. Asimismo, el enfoque cualitativo puede generar preguntas antes, durante y después de la recolección de los datos, por ser un proceso dinámico, ya que los diferentes estudios no siempre dan los mismos resultados. En otras palabras, según la teoría de Hernández, Fernández y Baptista (2014), este enfoque cualitativo “utiliza la recolección y análisis de los datos para afinar las preguntas de investigación o revelar nuevas interrogantes en el proceso de interpretación” (p. 7).

3.3.3 Enfoque mixto

Ahora bien, el enfoque de la investigación será mixto, dado que se utilizarán en conjunto técnicas cuantitativas y cualitativas para interpretar los resultados de la tesis, ya que el enfoque mixto se ajusta a las necesidades de la investigación,

puesto que se medirá cuantitativamente y analizará cualitativamente los resultados obtenidos.

En el presente caso, la parte cuantitativa está dada por un número determinado de los artículos y publicaciones válidos de los artículos disponibles sobre el tema. Estos deben ser comentados y analizados de acuerdo con una serie de escalas que partan de las variables definidas. Su número se obtiene de los criterios de búsqueda del investigador y los artículos que son seleccionados.

En el caso de los aspectos cualitativos, se consulta a una serie de profesionales con experiencia en código abierto o conocimiento en cubos de información, la factibilidad del uso exclusivo de herramientas de *software* de código abierto para la construcción de cubos de información que permitan realizar análisis de datos basados en razonamientos e inferencias.

3.4 Diseño

El diseño, según Hernández, Fernández y Baptista (2014), se define como un plan o estrategia que se desarrolla con el objetivo de obtener la información requerida en una investigación y responde al planteamiento. Con la finalidad de responder al planteamiento del problema, se programa el desarrollo de una investigación cualitativa, el análisis y la recolección de datos para enriquecerla.

Por medio de la confección del diseño, se quiere dar respuesta a la pregunta ¿cómo se procede a desarrollar la investigación? De acuerdo con el tipo de investigación y el enfoque que se se le dé a esta, se establecerá el diseño metodológico. Para la presente investigación, se explican los diseños no experimental, seccional y transversal que se explican a continuación.

3.4.1 No experimental

Un primer diseño identificado para este estudio es el de tipo no experimental, el cual, como expresan Hernández, Fernández y Baptista (2014), es el "estudio que se realiza sin la manipulación deliberada de variables y en los que solo se observan los fenómenos en su ambiente natural para después analizarlos" (p. 205). Es donde se observan situaciones ya existentes, no provocadas

intencionalmente por quien las realiza, además de que sus variables independientes ocurren, no se tiene control directo sobre ellas, ni se manipulan, porque ya habían sucedido.

El presente estudio, se dice que tiene un diseño no experimental, debido a que no se realiza la manipulación de ninguna variable propuesta, sino que la idea principal proviene de la recopilación de datos e información de los sujetos sin ninguna intención de manipular este tipo de respuestas, más bien el objetivo está en observar cómo estas respuestas son brindadas en ambientes más naturales y cotidianos para luego proceder con su exhaustivo análisis.

3.4.2 Transversal

Por último, se puede mencionar que el estudio posee un corte de tipo transversal, perteneciente a los seccionales acorde con Juárez (1993), es decir, aquel que pretende demostrar una situación presente o ausente en una población y periodo determinado.

En el caso del diseño transversal, se aplica para esta investigación en el sentido de que las variables principales por estudiar incluyen la recolección de la información en un momento dado del tercer cuatrimestre del año 2022.

3.5 Población y muestreo

En este caso, de acuerdo con las características de la investigación, se acude al muestreo no probabilístico, este se puede definir de la siguiente manera: “se caracteriza porque a los elementos de la muestra no se les ha definido la posibilidad de ser incluidos en la misma” (Méndez, 2014, p. 184). En este caso la posibilidad está dada porque se consideren representativos de la población.

En el presente trabajo, la muestra en todas las situaciones es no probabilística, es decir, se escoge a las personas con relación con la investigación. Por un lado, funcionarios de la empresa que manejan la parte de *software* y programación en la empresa de productos de la empresa, o sea, dos personas. Luego, cuatro personas que son expertas en cubos de información para el análisis de datos.

3.6 Instrumentos de recolección de datos

Este proyecto de investigación utilizará los siguientes instrumentos de recolección:

3.6.1 Entrevistas

En cuanto a la definición de entrevista, Hernández, Fernández y Baptista (2014) la definen como las “herramientas para recolectar datos cualitativos, se emplean cuando el problema de estudio no se puede observar o es muy difícil hacerlo por ética o complejidad” (p. 403). Es básicamente una conversación entre dos personas que parte de una serie de preguntas generadoras.

En esta investigación, se realiza un análisis por medio de entrevista a 12 colaboradores, esta puede ser tanto presencial como virtual, para que los expertos se expresen libremente y brinden opiniones sobre estos aspectos y la forma de mejorar las deficiencias que se valoren.

3.6.2 Análisis documental

De acuerdo con Hernández, Fernández y Baptista (2017), este tipo de investigación es la que se realiza, como su nombre lo indica, apoyándose en fuentes de carácter documental, esto es, en documentos de cualquier especie. Como subtipos de esta investigación, se encuentra la investigación bibliográfica, la hemerográfica y la archivística; la primera se basa en la consulta de libros, la segunda en artículos o ensayos de revistas y periódicos, y la tercera en documentos que se encuentran en los archivos, como cartas, oficios, circulares, expedientes, etcétera.

Se caracteriza porque busca la aplicación o utilización de los conocimientos que se adquieren. En una investigación empírica, lo que le interesa al investigador, primordialmente, son las consecuencias prácticas. En este caso, es la recolección y ordenamiento temporal de los diferentes datos sobre *software* de código abierto

y cubos de información. Todo esto agrupados en hojas de procesamiento de texto y estadísticas.

3.7 Técnicas de análisis de información

3.7.1 Herramientas

3.7.1.1 Diagrama de Ishikawa.

El diagrama Ishikawa, conocido también con el nombre de espina de pescado, es un medio de recolectar información sobre las características de calidad generadas en la prestación de un servicio y esquematizarlas ordenadamente en categorías (Acuña, 2015).

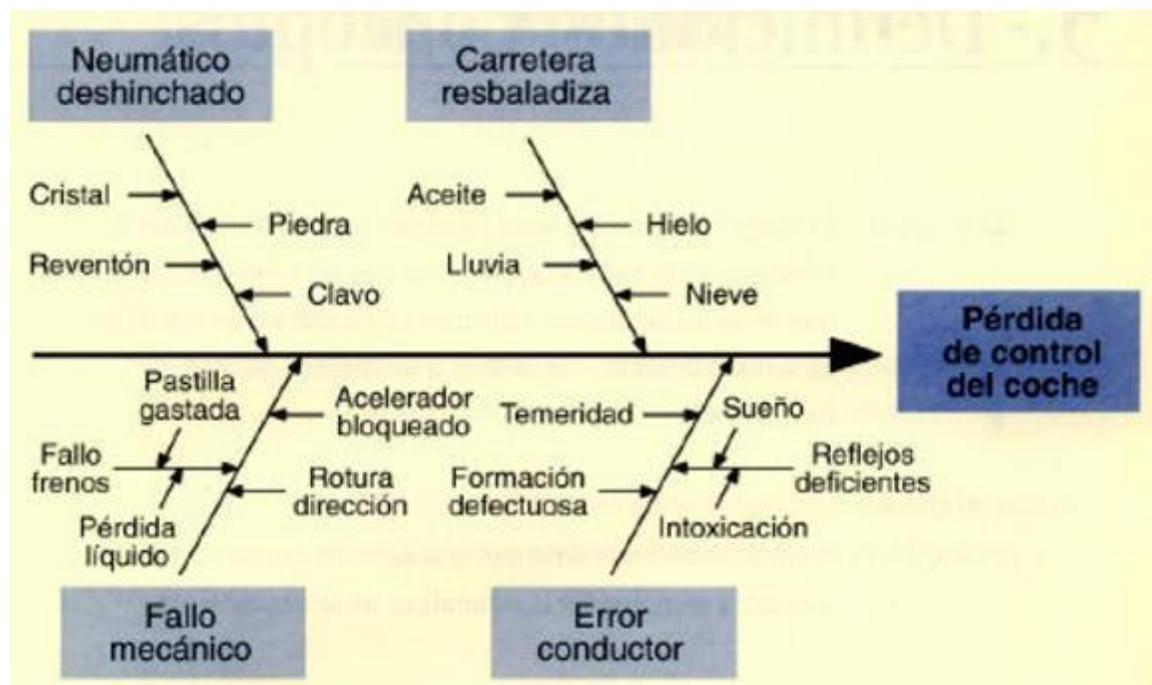


Figura 5. Diagrama de Ishikawa.

Fuente: www.fundibeq.com (2020).

El diagrama de Ishikawa se puede utilizar, de acuerdo con el sitio web Calidad PUCP (2014), para:

- Visualizar, en equipo, las causas principales y secundarias de un problema.

- Analizar procesos en búsqueda de mejoras.
- Mostrar el nivel de conocimientos técnicos que existe en la empresa sobre un determinado problema.
- Prever los problemas y ayudar a controlarlos, no sólo al final, sino durante cada etapa del proceso.

Según Acuña (2015), existen dos tipos de diagramas de Ishikawa; el diagrama de procesos, en el que se colocan los diversos procesos requeridos para la prestación del servicio en las ramas; y el diagrama general en el que se colocan todas las fuentes de causas directamente en las ramas, de acuerdo con su ocurrencia. El procedimiento para construir un Ishikawa de proceso es el siguiente:

1. Elegir el servicio o proceso que será objeto de estudio.
2. Colocar la frase procesos para el servicio de interés o el nombre del servicio en el extremo derecho de la flecha horizontal.
3. Hacer una lista de todas las características de calidad que se generan en cada etapa del proceso, o en forma general. Se debe tener en cuenta al cliente interno y auditores, de ninguna manera se puede realizar individualmente.
4. Ordenar la información de acuerdo con las etapas que conforman el proceso.
5. Dibujar las flechas diagonales (ramas principales) sobre las que se representarán los procesos o las características de calidad.
6. Dibujar subramas y anotar las características de calidad generadas en cada proceso anotado en las ramas principales.
7. Verificar que todas las características hayan sido anotadas.

3.7.1.2 La gráfica de Pareto.

Acuña (2015) explica que el gráfico de Pareto es muy útil para el análisis y categorización de problemas, utilizando como herramienta de ponderación el

Principio de Pareto, mediante el cual se seleccionan aquellos problemas cuya solución realmente favorezca el desempeño del sistema.

El diagrama de Pareto se puede considerar como una distribución de frecuencias o histograma de datos de atributos ordenados por categorías (Acuña, 2015). Acá se organizan diversas clasificaciones de datos por orden descendente, de izquierda a derecha por medio de barras sencillas para calificar las causas; de modo que se pueda asignar un orden de prioridades.

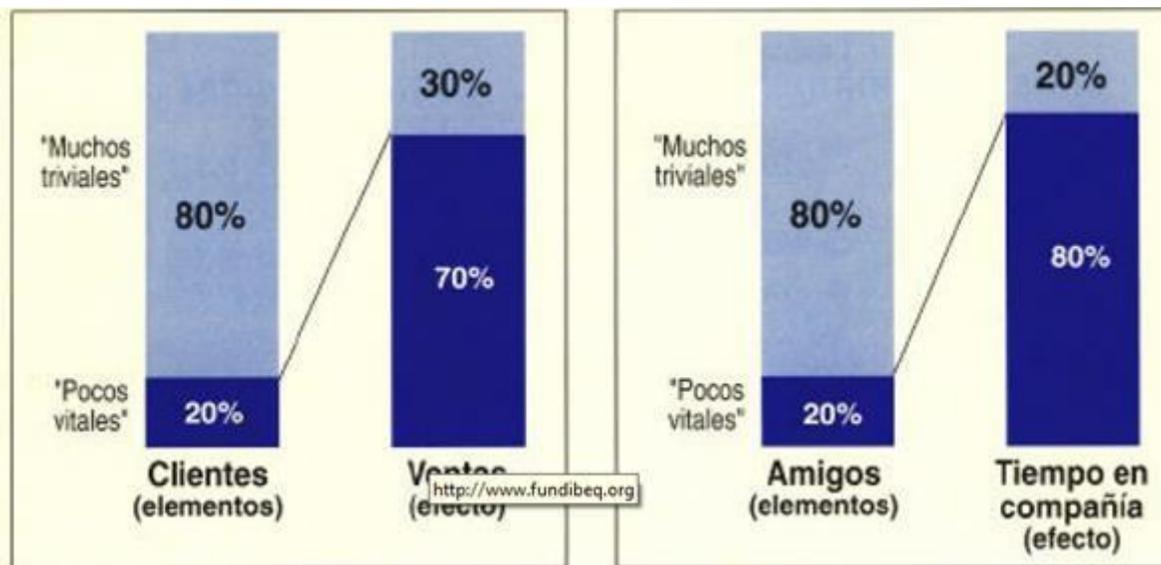


Figura 6. Diagrama de Pareto.

Fuente: www.fundibeq.com (2020).

Para Piñedo (2007), la gráfica de Pareto separa los “pocos vitales” de los “muchos triviales”. Es decir, muestra la importancia relativa de todos los problemas o condiciones con el fin de determinar el punto de inicio en la solución de los problemas de tal manera que un equipo sepa dónde dirigir sus esfuerzos para mejorar. Reducir los problemas de impacto (las barras más largas en una Gráfica Pareto) servirá más para una mejora general del proceso. Con frecuencia, el 20 % de las causas representarán el 80 % de los problemas.

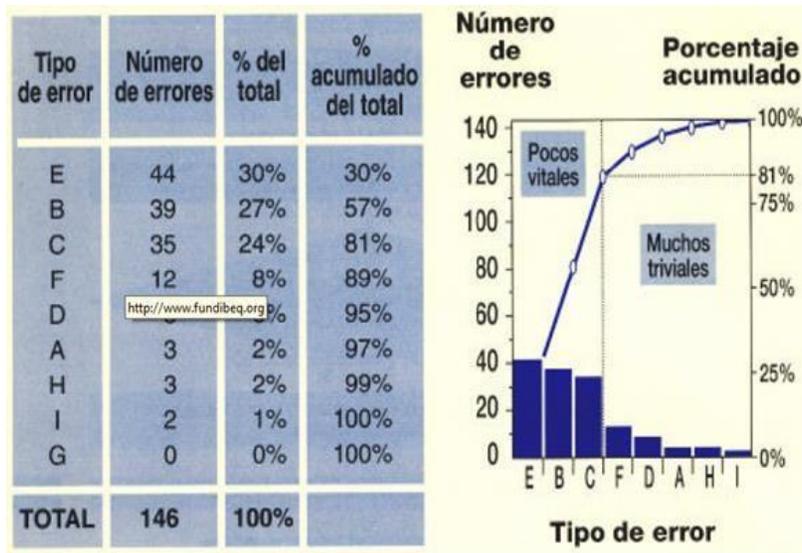


Figura 7. Principio de Pareto.

Fuente: www.fundibeq.com (2020).

Según Pinedo (2018), la gráfica de Pareto se utiliza principalmente cuando:

- Se necesita identificar un producto o servicio para el análisis con el fin de mejorar la calidad.
- Se requiere identificar oportunidades para mejorar.
- Se necesita analizar las diferentes agrupaciones de datos (ej.: por producto, por segmento, del mercado, área geográfica, etc.).
- Se desea buscar las causas principales de los problemas y establecer la prioridad de las soluciones.
- Se necesita evaluar los resultados de los cambios efectuados a un proceso.

Capítulo 4. Análisis del diagnóstico

A continuación, se comentan las características principales de las herramientas para realizar inteligencia de negocios y que su licenciamiento es de código abierto encontradas en las búsquedas definidas para este estudio.

Las características principales que las herramientas para realizar inteligencia de negocios deben cumplir son las siguientes:

1. Permite, de manera ágil, la extracción de datos de un origen de datos digital, con ello, se entiende que se puede seleccionar datos para ser consultados posteriormente por el usuario final del aplicativo. El principal objetivo de la extracción es la construcción de un repositorio histórico de datos en el cual se almacenan copias instantáneas de los datos a través del tiempo.
2. Permite realizar modificaciones a los datos previamente extraídos para ser modelados según los requerimientos de los usuarios finales. La información extraída luego es maleable para convertirse en los datos requeridos por los usuarios de la aplicación.
3. Se debe garantizar el acceso a los datos por parte de los usuarios con total independencia de la fuente de obtención de éstos.
4. Facilita la carga de datos digitales de forma ágil a una base de datos destino o repositorio integral de datos.
5. Permite la construcción de un repositorio operacional de los datos transaccionales para compartir características técnicas a través de todos los datos almacenados en el repositorio integral de datos.
6. Cuenta con herramientas que permiten la agrupación de los datos en dimensiones, según sus características principales para brindar apoyo en la toma de decisiones.
7. La herramienta permite la construcción de variables desde los datos por medio de la agrupación de datos, llamadas medidas.

8. Esta herramienta debe ser facilitadora de la construcción de reportes multidimensionales realizando cruces entre las dimensiones y las medidas, con el objetivo de consolidar los datos sin tener que contar con mucha experiencia técnica respecto a la construcción de repositorios de datos.
9. Permite la construcción de fotografías de los datos según la historia en el tiempo de los datos transaccionales.

Las herramientas analizadas en este estudio permiten la construcción de cubos de información, con el objetivo de definir una base de datos para realizar comparaciones, la siguiente es una tabla de los datos obtenidos al consultar los sitios web publicados por cada proveedor de herramienta para realizar inteligencia de negocios. Con el propósito de comentar las propiedades de las herramientas de inteligencia de negocios encontradas en la investigación y cumplir con el objetivo específico número dos de esta investigación se enumeran las características de cada herramienta encontrada en la tabla 8.

Tabla 8. Características de las herramientas encontradas en las búsquedas.

Nombre de la herramienta	Características destacadas
Pentaho Business Analytics	<ul style="list-style-type: none"> • Publicado por: Pentaho Corporation. • Licencia: Open source por medio de la versión comunitaria. Pero también posee una versión de pago por suscripción (Enterprise) que incluye todas las funcionalidades disponibles. • Integración de datos (ETL): Sí, Pentaho Data Integration (Kettle). • Provee Servicios OLAP (cubos): Sí. • Reportería de datos: Sí. • Dashboards: Sí. • Minería de datos: Sí. • Plataforma base: Java

Kibana	<ul style="list-style-type: none"> • Publicado por: Elastic. • Licencia: Open source, licenciamiento Apache. • Integración de datos (ETL): Por medio de la herramienta de datos que provee Elastic, llamada Logstash la cual es una herramienta de ETL. • Provee servicios OLAP (cubos): No. • Reportería de datos: Sí. • Dashboards: Sí. • Minería de datos: No. • Plataforma base: JavaScript • Nota: Los datos deben almacenarse únicamente en Elasticsearch.
Knime	<ul style="list-style-type: none"> • Publicado por: Science Team, University of Konstanz. • Licencia: Open source. • Integración de datos (ETL): Sí. • Provee servicios OLAP (cubos): No. • Reportería de datos: Sí. • Dashboards: Sí. • Minería de datos: No. • Plataforma base: Java
Jaspersoft Business Intelligence Suite.	<ul style="list-style-type: none"> • Publicado por: Tibco. • Licencia: Open source. • Integración de datos (ETL): Si, mediante la herramienta Jaspersoft Community ETL desarrollada en conjunto con Talend y su comunidad. • Provee servicios OLAP (cubos): Sí. • Reportería de datos: Sí. • Dashboards: Sí. • Minería de datos: No.

	<ul style="list-style-type: none"> • Plataforma base: Java
Apache Zeppelin	<ul style="list-style-type: none"> • Publicado por: Apache. • Licencia: Open source de Apache. Exploración de datos para el motor de datos Hadoop y Spark. • Integración de datos (ETL): Sí. • Provee servicios OLAP (cubos): No. • Reportería de datos: No. • Dashboards: No. • Minería de datos: No. • Plataforma base: Java
Talend Open Studio for Data Integration	<ul style="list-style-type: none"> • Publicado por: Talend. • Licencia: Open source. • Integración de datos (ETL): Sí. • Provee servicios OLAP (cubos): No. • Reportería de datos: No. • Dashboards: No. • Minería de datos: No. • Plataforma base: Java

Fuente: Elaboración propia.

Utilizando los datos mostrados en la tabla ocho y con la finalidad de continuar con la investigación y solventar el objetivo específico número tres, que consiste en escoger una herramienta o un grupo complementario de herramientas que permitan la construcción de un cubo de información para análisis de datos, se procederá a asignar un peso a cada una de las características de cada herramienta y así seleccionar la herramienta seleccionada para construir una solución de inteligencia de negocios.

Tabla 9. Peso de cada característica que posee una herramienta en investigación.

Característica de la herramienta a evaluar:	Puntaje asignado o Peso sugerido.
¿Es la licencia “Open source”?	10 puntos
¿Posee integración de datos (ETL)?	5 puntos
¿Provee servicios OLAP construcción de cubos?	10 puntos
¿Cuenta con reportería de datos?	7 puntos
¿Permite la construcción de <i>dashboards</i> ?	8 puntos
¿Es la minería de datos parte de la suite de herramientas?	6 puntos

Fuente: Elaboración propia

Se realiza la asignación de valores según propiedades de las herramientas, con el fin de determinar cuál de las herramientas será la seleccionada para construir un prototipo del proyecto para realizar análisis de datos. Este prototipo cumple con el objetivo número cuatro de la presente investigación, el cual requiere la experimentación de la construcción de un cubo de información para comprobar las características de la solución seleccionada, poniendo en práctica el funcionamiento de la herramienta o grupo complementario de herramientas escogidas.

Tabla 10. Peso asignado según sus características para cada herramienta seleccionada en la investigación.

Pentaho	Kbana	Knime	Jaspersoft	Apache Zeppelin	SpagoBI	Talend
---------	-------	-------	------------	-----------------	---------	--------

¿Es la licencia <i>open source</i> ?	10	10	10	10	10	10	10
¿Posee integración de datos (ETL)?	5	5	5	5	5	0	5
¿Provee servicios OLAP construcción de cubos?	10	0	0	10	0	10	0
¿Cuenta con reportería de datos?	7	7	7	7	0	7	0
¿Permite la construcción de <i>dashboards</i> ?	8	8	8	8	0	8	0
¿Es la minería de datos parte de la suite de herramientas?	6	0	0	0	0	0	0
Totales finales	46	30	30	40	15	35	15

Fuente: Elaboración propia.

L

a
tabl
a
10
nos
mu

estra en forma de resumen los totales finales obtenidos según el peso asignado a cada una de las características de las herramientas que se encuentran en investigación. La herramienta que obtiene el mayor puntaje es la seleccionada para cumplir con el objetivo específico de experimentar la construcción de cubo OLAP. Esta experimentación permitirá adquirir los conocimientos necesarios para brindar a la empresa asociada con la investigación de una solución en inteligencia de negocios, con el propósito de solventar sus requerimientos de información. La herramienta que obtuvo el mayor puntaje fue Pentaho con 46 puntos, seguida de Jaspersoft. Lo que quiere decir que se experimentará la construcción de un cubo

de información utilizando la suite de Pentaho, porque cumple la mayor cantidad de características para ser una herramienta de inteligencia de negocios.

Capítulo 5. Propuesta de solución

Para implementar la solución de inteligencia de negocios utilizando únicamente *software* de código abierto en la empresa, se propone una metodología basada en características propias de la inteligencia de negocio, así como hacer referencia a las buenas prácticas para la implantación de sistemas basados en *software* libre.

Los requisitos para la implementación de una herramienta de inteligencia de negocios deben considerar las necesidades actuales de la empresa u organización con respecto al proceso de toma de decisiones. Así se podrán definir los requisitos mínimos que debe cumplir la solución de *software* de inteligencia de negocios que se desea implementar.

El proceso parte de la definición de las debilidades y desventajas que el *software* puede ayudar a mejorar en la organización y del levantamiento de requerimientos de la solución a desarrollar. Luego de ello, se debe analizar el proceso de toma de decisiones y establecer las características que solventará la herramienta de *software*.

Posteriormente, se debe determinar cuál es la información más importante de la empresa y colocarla en la gestión de información de la herramienta. Así se podrá configurar el modelo de datos que se utilizará para solventar las necesidades inmediatas.

Se deberá definir para la empresa los siguientes puntos importantes:

1. Extraer características del *software* utilizado por la empresa actualmente.
2. Mapeo de datos.
3. Extracción de datos.
4. Desarrollo e implementación de una solución de inteligencia de negocios.
5. Consolidación de la información.

6. Construcción del cubo de inteligencia de negocios.

7. Elaboración de análisis de datos.

Según los resultados obtenidos en el capítulo anterior, en el cual se escogió una herramienta para la construcción de un cubo para análisis de datos; se procede ahora a mostrar los resultados obtenidos de la experimentación en la construcción de una solución de inteligencia de negocios utilizando las herramientas de Pentaho. En este mismo capítulo se muestra la creación de un esquema de infraestructura para combinar herramientas de inteligencia de negocio de código abierto y solventar la necesidad de información de la empresa. La experimentación con la construcción del cubo y el dibujo de un esquema de infraestructura, mostrados en este capítulo, resuelven las interrogantes planteadas en los objetivos específicos cuatro y cinco de la presente investigación. Pentaho software y su suite de herramientas para inteligencia de negocios fue el que obtuvo la mayor nota según el peso asignado y los criterios evaluados, las herramientas de Pentaho incorporan las siguientes características:

- Tipo de licenciamiento del tipo de código abierto.
- Provee integración de sistemas por medio de herramientas para la construcción de ETL (Extracción, Transformación y Carga de datos).
- Posee servicios de análisis en línea de procesos (OLAP) para la construcción de cubos para el análisis de datos.
- Cuenta con herramientas para la construcción de reportes transaccionales.
- Permite la construcción de paneles de indicadores gerenciales.
- Cuenta con una herramienta para la minería de datos.

La solución de inteligencia de negocios construida durante la experimentación utilizó como datos transaccionales la base de datos **classicmodel**, la cual es una base de datos pública en Internet que permite analizar la funcionalidad y características de MySQL 8.0. Su descarga es gratuita y

contiene información pública para el análisis. No se realizó la experimentación con datos de la empresa, ya que la información de su base de datos es confidencial y los representantes legales de la empresa no cedieron los permisos para utilizar sus datos en una investigación abierta como la actual.

Es importante señalar que las empresas u organizaciones que se inclinen por la implementación de soluciones de inteligencia de negocio de basadas en *software* de código abierto tendrán beneficios, tal y como menciona Flórez (2012), en reducción de costos, flexibilidad para adaptar sus soluciones de *software* los productos y la independencia de sus proveedores.

Para la construcción de un cubo para análisis de datos, se deben descargar e instalar las siguientes herramientas *open source*, en el caso de la base de datos se selecciona como opción utilizar un proveedor en línea, servicio pago contra demanda y utilización:

Tabla 11. Proveedores de herramientas.

Proveedor	Herramienta	Versión funcional utilizada
Oracle	Java Runtime	Versión 8 Update 351
Hitachi Vantara	Pentaho Data Integrator	Version 9.3.0.0-428
Hitachi Vantara	Pentaho Server	Version 9.3.0.0-428
Hitachi Vantara	Pentaho Schema Workbench	Version 9.2.0.0-290
Google Cloud	MySQL	Version 8.0
Metabase, Inc	Metabase	Version 0.45.1

Fuente: Elaboración propia.

La figura 8 muestra el esquema de infraestructura para la construcción de cubos de información, con la finalidad de dibujar un plano final de la solución utilizando una herramienta o un grupo de herramientas que sean de código abierto ilustrando la interacción de las herramientas en la solución.

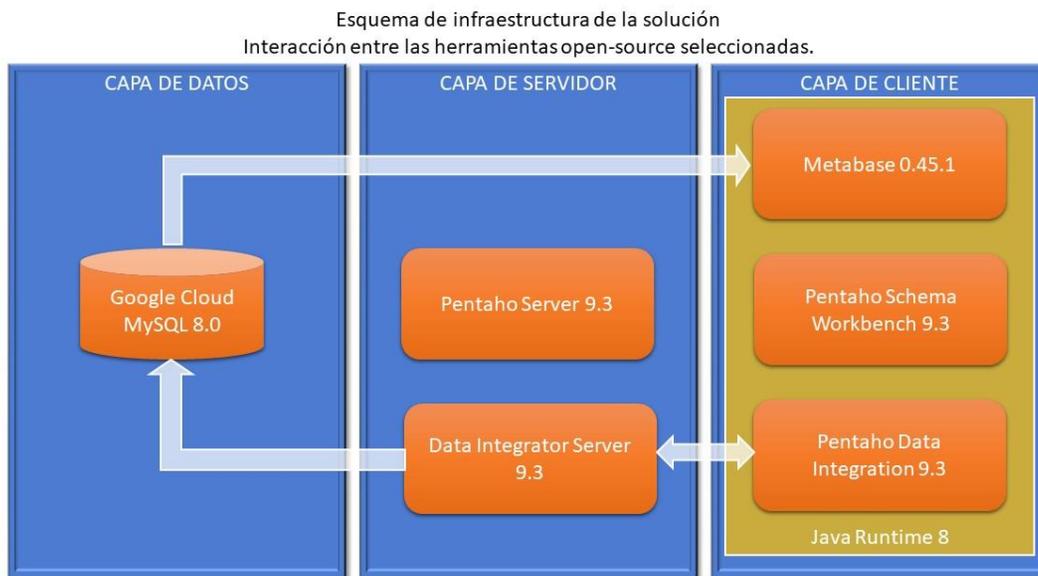


Figura 8. Esquema de infraestructura de la solución.

Fuente: Elaboración propia.

5.1 Detalle de los componentes de la solución

5.1.1 Capa de datos. Base de datos MySQL 8.0

Esta herramienta es un motor de base de datos que permite la creación de base de datos en forma de estrella para plasmar las bases de la construcción del cubo de información. También, aquí se creó la base de datos transaccional que contiene los datos OLTP o transaccionales. A partir de estos dos modelos de base de datos es que se nutre de información y datos a todos los elementos de la solución.

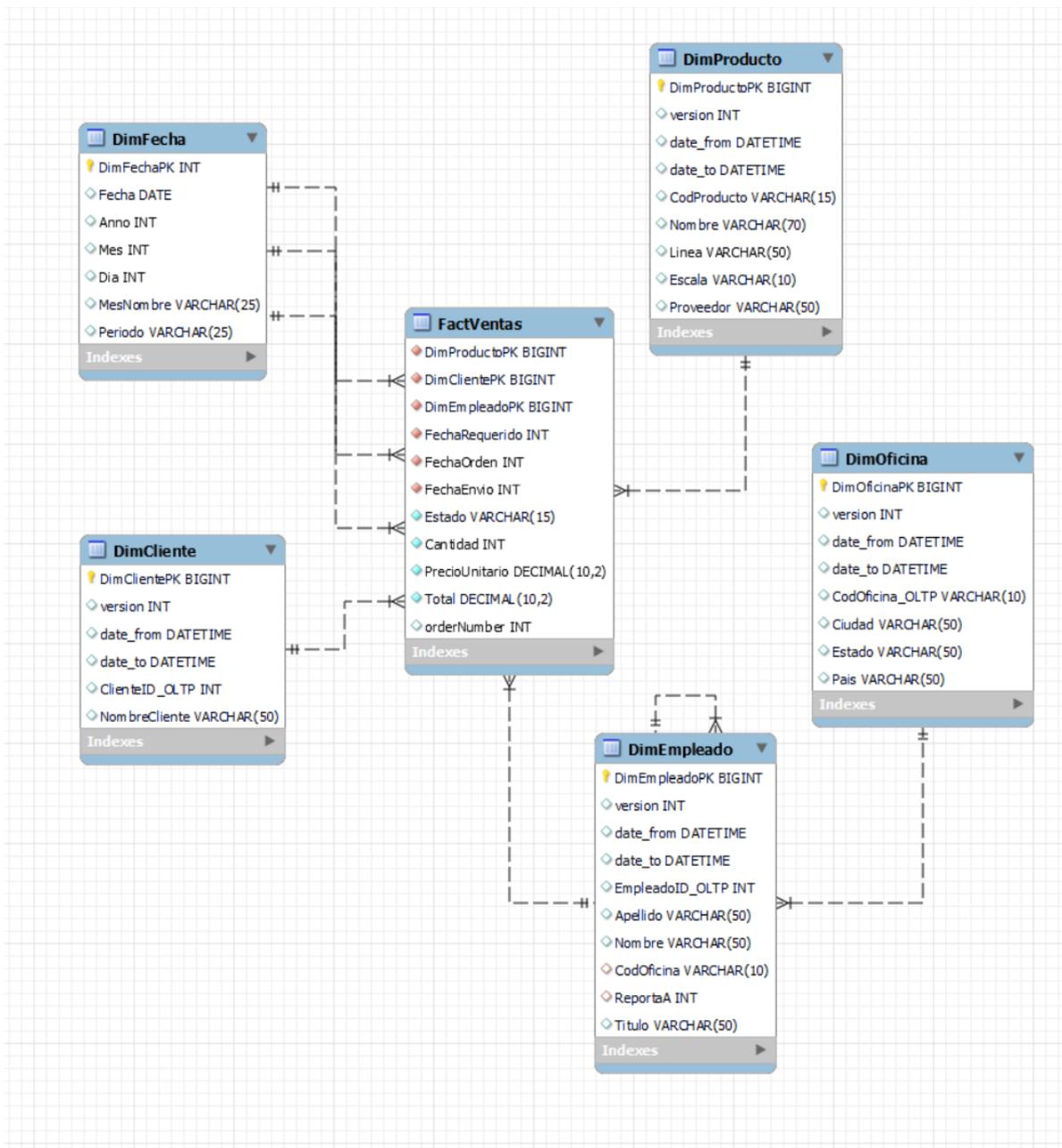


Figura 10. Modelo en estrella de la base de datos OLAP.

Fuente: Elaboración propia.

5.1.2 Capa cliente. Pentaho Data Integration 9.3

La herramienta Pentaho Data Integration permite la construcción de los ETL, los cuales son procesos que permiten la extracción, transformación y carga de datos desde diferentes fuentes y destinos de datos. En este caso se utilizó la

herramienta Pentaho Data Integrator para construir un ETL que migre los datos transaccionales de la base de datos OLTP (classicmodels) a la base de datos OLAP, la cual se llama tesis. Esta migración de datos permite que los datos puedan ser analizados utilizando como criterios de búsqueda sus dimensiones y realizar operaciones de agrupación por medio de funciones de agregación para las medidas definidas en el modelo de datos. Los ETL definidos y construidos son los siguientes:

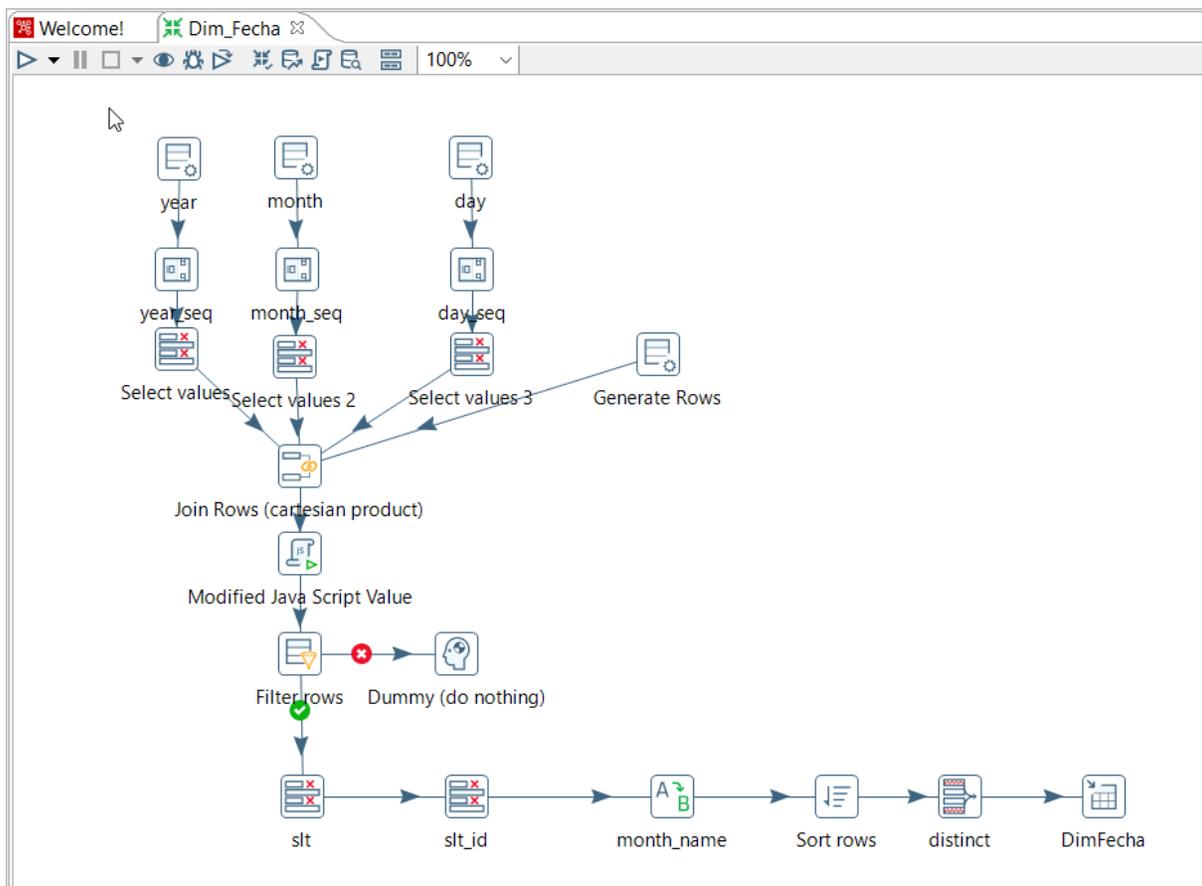


Figura 11. ETL para llenado de la dimensión fecha.

Fuente: ETL para la carga de Dim_Fecha.

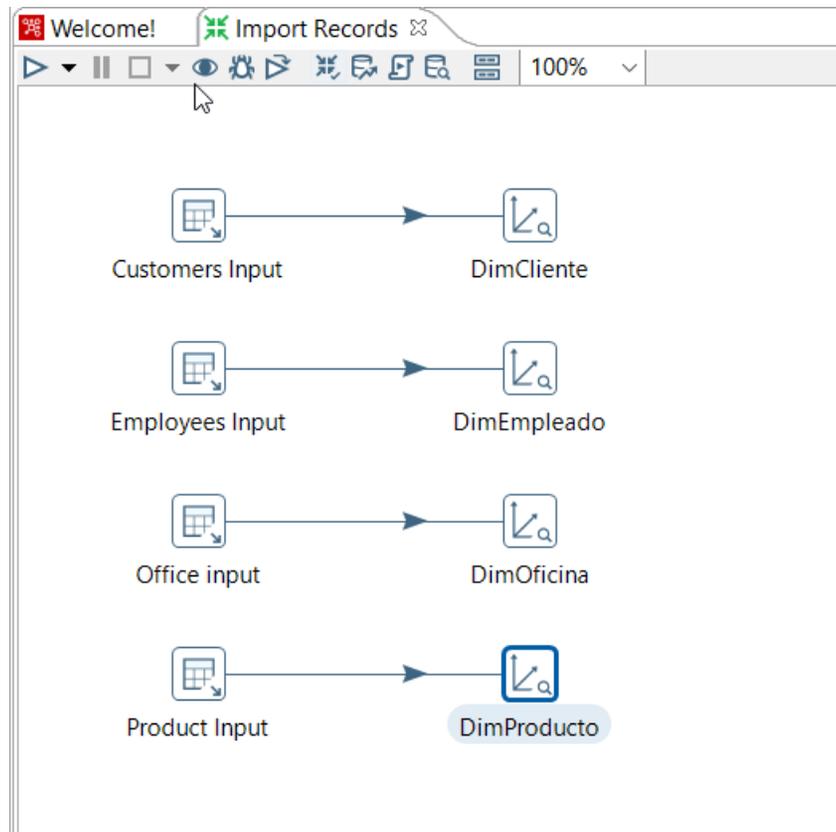


Figura 12. ETL para migrar los datos de la base de datos OLTP a OLAP.

Fuente: ETL para carga de datos a OLAP.

En la figura 12 se muestran los ETL con los que se experimentó la construcción. Ellos permiten la migración de datos desde la base de datos transaccional a la base de datos analítica.

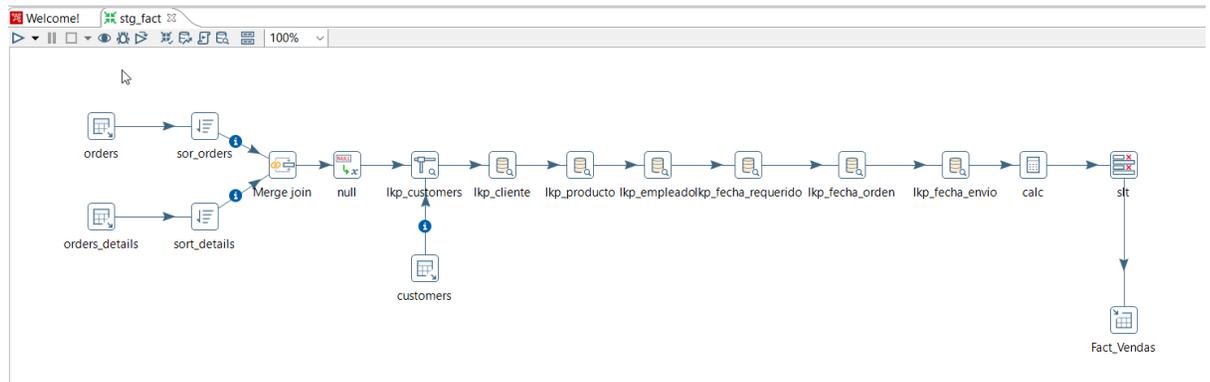


Figura 13. ETL para carga de tabla de hechos.

Fuente: ETL para importar tabla de hechos.

La tabla de hechos almacena todos los datos transaccionales que serán analizados según sus dimensiones. Por su naturaleza, esta tabla puede contener miles o millones de datos, dependiendo de qué tan grande puede ser la base de datos transaccional o cuantos datos copia se almacenen en ella.

5.1.3 Capa Cliente. Schema Workbench 9.3

Este elemento de la suite de herramientas de Pentaho que se seleccionó permite el diseño y la creación de cubos de información para el análisis de datos. La estructura del cubo se convierte en un archivo .xml que contiene todas las características definidas del cubo de información que se ha construido. Para la investigación, se creó el cubo tesis.xml, el cual posee la siguiente estructura:

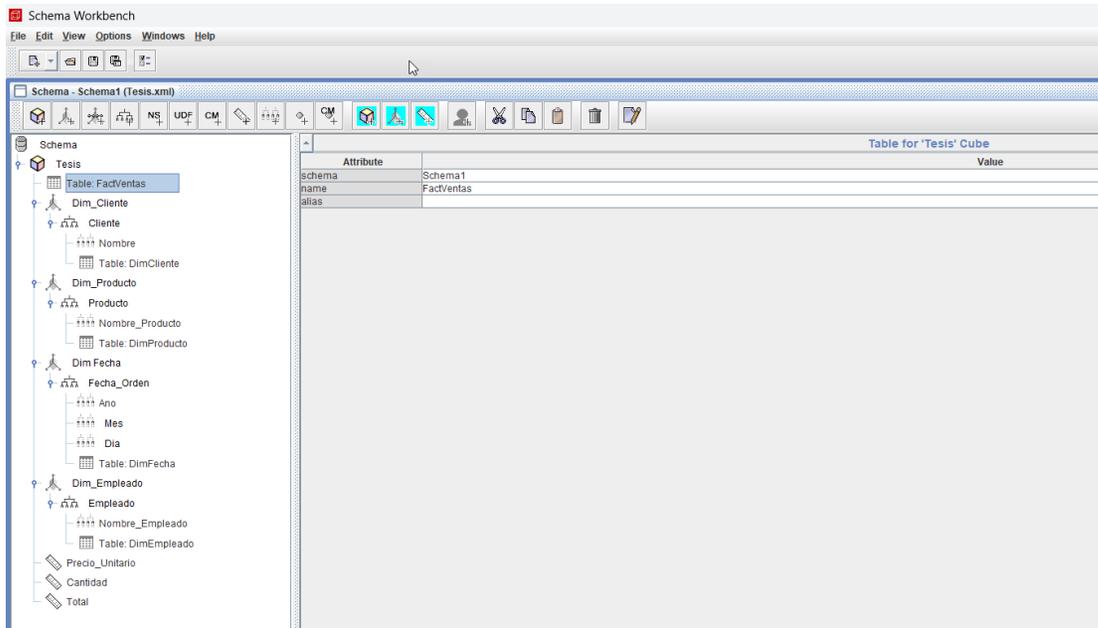


Figura 14. Schema Workbench y el cubo tesis.

Fuente: cubo tesis.xml.

Seguidamente, se muestra un vistazo de cómo se ve la estructura del archivo tesis.xml en un editor de texto.

```
<Schema name="Schema1">
  <Cube name="Tesis" visible="true" cache="true" enabled="true">
    <Table name="FactVentas" schema="Schema1">
    </Table>
    <Dimension type="StandardDimension" visible="true" foreignKey="DimClienteFK" highCardinality="false" name="Dim_Cliente">
      <Hierarchy name="Cliente" visible="true" hasAll="false" primaryKey="DimClienteFK">
        <Table name="DimCliente">
        </Table>
        <Level name="Nombre" visible="true" table="DimCliente" column="NombreCliente" nameColumn="NombreCliente" parentColumn="NombreCliente" type="String" uniqueMembers="false" levelType="Regular">
        </Level>
      </Hierarchy>
    </Dimension>
    <Dimension type="StandardDimension" visible="true" foreignKey="DimProductoFK" highCardinality="false" name="Dim_Producto">
      <Hierarchy name="Producto" visible="true" hasAll="true" primaryKey="DimProductoFK">
        <Table name="DimProducto">
        </Table>
        <Level name="Nombre_Producto" visible="true" table="DimProducto" column="Nombre" nameColumn="Nombre" parentColumn="Nombre" type="String" uniqueMembers="false" levelType="Regular">
        </Level>
      </Hierarchy>
    </Dimension>
    <Dimension type="TimeDimension" visible="true" foreignKey="FechaOrden" highCardinality="false" name="Dim Fecha">
      <Hierarchy name="Fecha_Orden" visible="true" hasAll="true" primaryKey="DimFechaFK">
        <Table name="DimFecha">
        </Table>
        <Level name="Ano" visible="true" table="DimFecha" column="Anno" nameColumn="Anno" ordinalColumn="Anno" parentColumn="Anno" type="Integer" internalType="int" uniqueMembers="false" levelType="Regular">
        </Level>
        <Level name="Mes" visible="true" table="DimFecha" column="MesNombre" nameColumn="MesNombre" ordinalColumn="MesNombre" parentColumn="MesNombre" type="String" internalType="String" uniqueMembers="false" levelType="Regular">
        </Level>
        <Level name="Dia" visible="true" table="DimFecha" column="Dia" nameColumn="Dia" parentColumn="Dia" type="Integer" internalType="int" uniqueMembers="false" levelType="Regular">
        </Level>
      </Hierarchy>
    </Dimension>
    <Dimension type="StandardDimension" visible="true" foreignKey="DimEmpleadoFK" highCardinality="false" name="Dim_Empleado">
      <Hierarchy name="Empleado" visible="true" hasAll="true" primaryKey="DimEmpleadoFK">
        <Table name="DimEmpleado">
        </Table>
        <Level name="Nombre_Empleado" visible="true" table="DimEmpleado" column="Nombre" nameColumn="Nombre" ordinalColumn="Nombre" parentColumn="Nombre" type="String" uniqueMembers="false" levelType="Regular">
        </Level>
      </Hierarchy>
    </Dimension>
    <Measure name="Precio_Unitario" column="PrecioUnitario" datatype="Numeric" aggregator="sum" visible="true">
    </Measure>
  </Cube>
</Schema>
```

Figura 15. Vista de archivo XML en editor de texto.

Fuente: cubo tesis.xml

5.1.4 Capa cliente. Metabase 0.45

Por medio la incorporación al proyecto del uso de la herramienta Metabase 0.45, se puede lograr la exploración y análisis de datos. Se sugiere la herramienta como un complemento al cubo, ya que facilita la explotación de los datos para el usuario final. Esta misma herramienta cuenta con análisis exploratorios, lo que permite diagnosticar la información con la que se cuenta. Algunos análisis con los que se puede contar de primera mano son los siguientes:

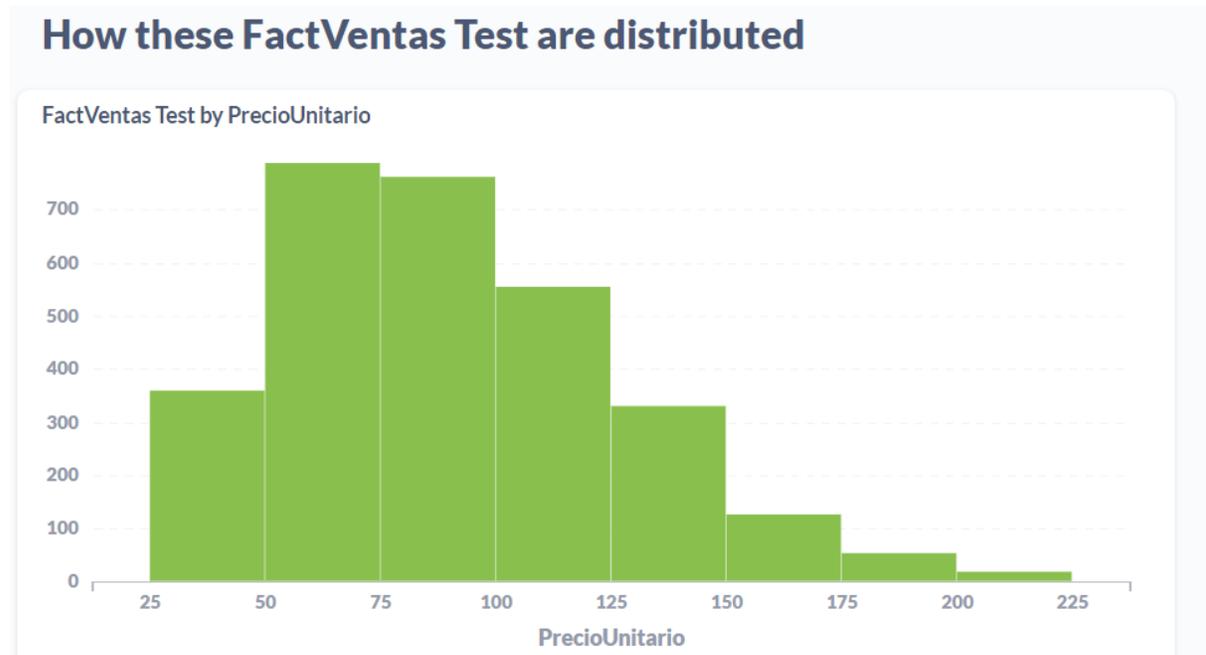


Figura 16. Análisis de la distribución de la tabla de hechos.

Fuente: Metabase.

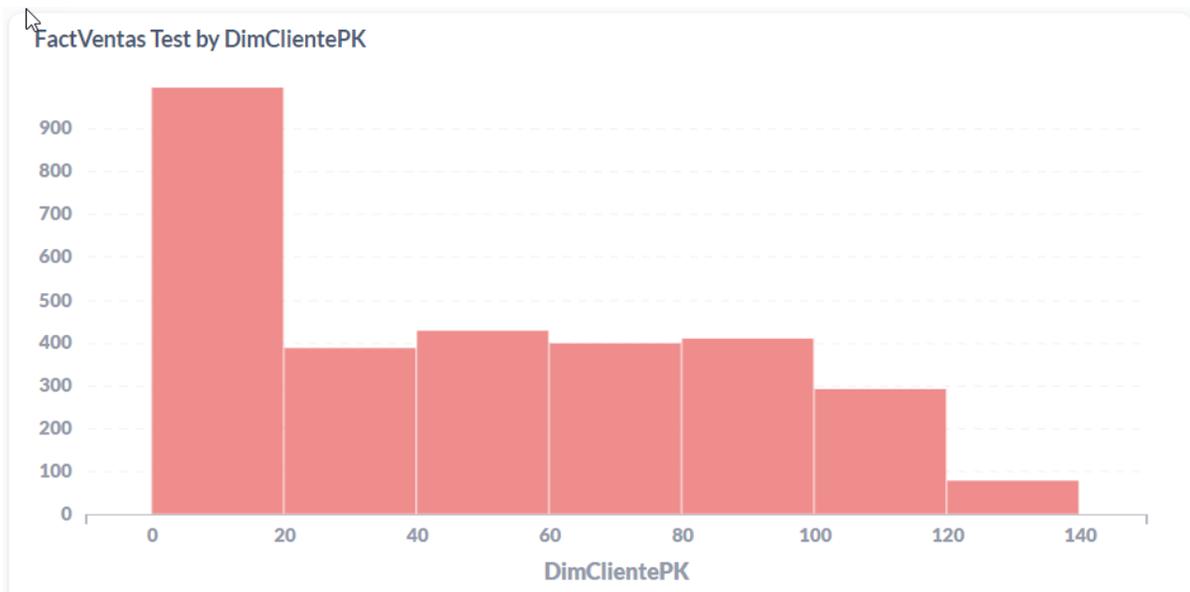


Figura 17. Análisis de la tabla de hechos según la dimensión de clientes.

Fuente: Metabase.

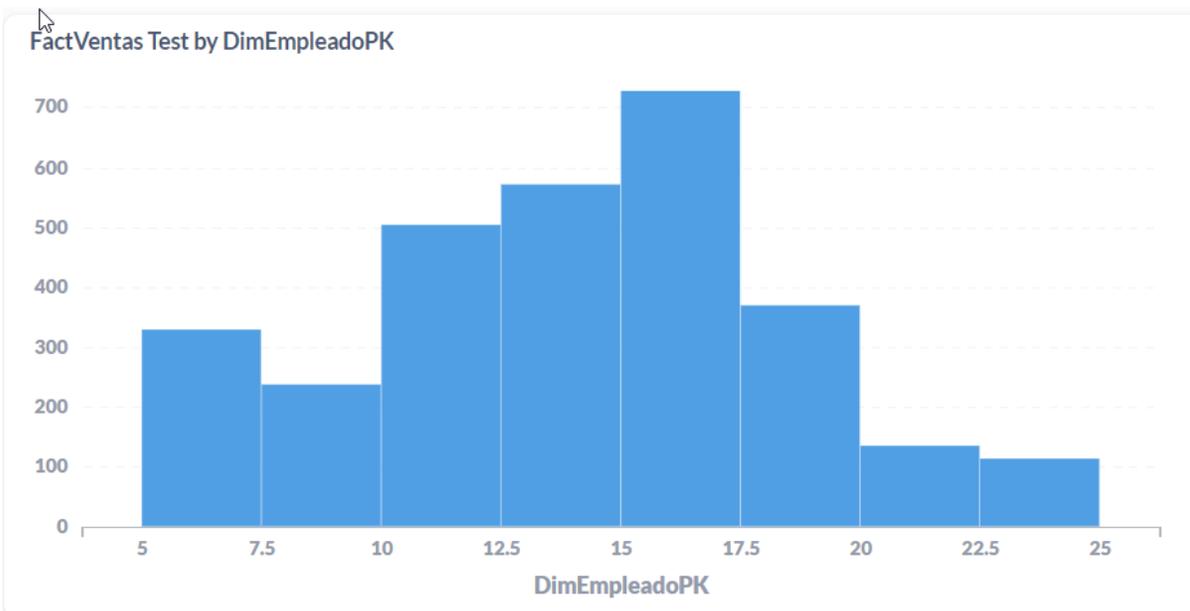


Figura 18. Análisis de distribución de la tabla de hechos según la dimensión de empleados.

Fuente: Metabase.

Capítulo 6. Conclusiones y recomendaciones

6.1 Conclusiones

- Se construyó un listado en el cual se pueden identificar siete herramientas de código abierto disponibles en la actualidad y que permiten realizar inteligencia de negocios; adicionalmente, estas herramientas mencionadas se han contabilizado por medio de búsquedas en sitios especializados en línea.
- Se comentaron las características de las herramientas de código abierto que permiten la construcción de cubos de información, con el objetivo de construir una base de datos para realizar comparaciones consultando los sitios web publicados por cada proveedor.
- Se escogió Pentaho como herramienta de experimentación, esta permite la construcción de un cubo de información por medio de la aplicación del criterio experto de selección con el objeto de concentrar el foco de la investigación.
- Se experimentó la construcción de un cubo de información para comprobar las características de la solución seleccionada, poniendo en práctica el funcionamiento de la herramienta escogida.
- Se creó un esquema de infraestructura para la construcción de cubos de información, con la finalidad de dibujar un plano final de la solución, utilizando un grupo de herramientas que son de código abierto ilustrando la interacción de las herramientas en la solución.
- Es factible el uso exclusivo de herramientas de *software* de código abierto para la construcción de cubos de información que permitan realizar análisis de datos, con el fin de abaratar costos en licenciamiento en un proyecto de inteligencia de negocios.

Glosario

OLTP	<i>Online Transactional Process</i> . Procesamiento transaccional de los datos. Se refiere a la sección de base de datos transaccional de un modelo de inteligencia de negocios.
OLAP	<i>Online Analytical Process</i> . Procesamiento analítico en línea. Se refiere a la sección analítica de la base de datos de un modelo de inteligencia de negocios.

Referencias

- Hernández, R. y Mendoza, C. (2018). *Metodología de la investigación. Las rutas cuantitativa, cualitativa y mixta*. (7a. ed.). México: Editorial McGraw Hill Education. <https://b-ok.lat/book/5375854/6f5ae2>
- Hernández, R., Fernández, C. y Baptista, P. (2014). *Metodología de Investigación*. México: Mc Graw Hill. Tercera Edición.
- Juárez, M. (1993). *Trabajo social e investigación: temas y perspectivas*. Madrid: Universidad Pontificia Comillas.
- Justexw. (2021). *Funciones de CUBO en Excel*. <https://es.justexw.com/funciones-de-excel/cubo>
- Marín, A. (2017, 4 abril). *Breve historia del Business Intelligence: Origen y evolución*. Time Manager. <https://www.timemanagerweb.com/breve-historia-del-business-intelligence/>
- Microsoft. (2021, 17 noviembre). *Uso de cubos OLAP para análisis avanzado*. Microsoft Docs. <https://docs.microsoft.com/es-es/system-center/scsm/olap-cubes-overview?view=sc-sm-2019>

- Redacción España. (2019). *Qué son las herramientas de código abierto y de software libre*. <https://agenciab12.com/noticia/que-son-herramientas-codigo-abierto-software-libre>
- Rodríguez, E. (2020). *Investigación aplicada: características, definición, ejemplos*. Lifereder. <https://www.lifereder.com/investigacion-aplicada/>
- Rouse, M. (2018, 27 diciembre). *Inteligencia de negocios (BI)*. ComputerWeekly.es. <https://www.computerweekly.com/es/definicion/Inteligencia-de-negocios-BI>
- Sharma, K., Shetty, A., Jain, A., Kumar, R. (2021). *A Comparative Analysis on Various Business Intelligence (BI), Data Science and Data Analytics Tools. International Conference on Computer Communication and Informatics (ICCCI)*. <https://ieeexplore.ieee.org/document/9402226>
- Stora, B. (2017). *Use of open-source tools for business intelligence in an environment of big data*. República Checa : Masaryk University.

