

**UNIVERSIDAD NACIONAL AGRARIA LA
MOLINA**

FACULTAD DE ECONOMÍA Y PLANIFICACIÓN



**“DISEÑO E IMPLEMENTACIÓN DE UN DATAMART PARA EL
ÁREA DE ANÁLISIS DE UNA INSTANCIA TÉCNICA DEL
SECTOR EDUCATIVO”**

**TRABAJO DE SUFICIENCIA PROFESIONAL PARA OPTAR
TÍTULO DE INGENIERA ESTADÍSTICA E INFORMÁTICA**

KAREN ANGELA AYALA FLORES

LIMA – PERÚ

2021

**La UNALM es titular de los derechos patrimoniales de la presente investigación
(Artículo 24 – Reglamento de Propiedad Intelectual)**

**UNIVERSIDAD NACIONAL AGRARIA
LA MOLINA**

FACULTAD DE ECONOMÍA Y PLANIFICACIÓN

**“DISEÑO E IMPLEMENTACIÓN DE UN DATAMART PARA EL
ÁREA DE ANÁLISIS DE UNA INSTANCIA TÉCNICA DEL
SECTOR EDUCATIVO”**

**TRABAJO DE SUFICIENCIA PROFESIONAL PARA OPTAR
TÍTULO DE INGENIERA ESTADÍSTICA E INFORMÁTICA**

Presentado por:

KAREN ANGELA AYALA FLORES

Sustentado y aprobado ante el siguiente jurado:

**M.A. Fernando René Rosas Villena
PRESIDENTE**

**Mg. Iván Dennys Soto Rodríguez
ASESOR**

**Mg. Sc. Celso Gonzales Chavesta
MIEMBRO**

**Mg. Clodomiro Fernando Miranda Villagómez
MIEMBRO**

LIMA – PERÚ

2021

DEDICATORIA

Este trabajo se lo dedico en primer lugar a Dios, por fortalecerme en todo tiempo y por mostrarme cada día su misericordia y su gracia.

A mis padres, por su apoyo incondicional y sus sabios consejos.

A mis hermanas, por ser mi soporte.

A mi sobrino Eduardo, por ser el motor de mi vida.

AGRADECIMIENTO

En primer lugar, a Dios, porque todo se lo debo a él.

A mis padres, por ayudarme a llegar hasta aquí, y cumplir esta meta profesional.

A mis hermanas y mi sobrino Eduardo, porque día a día me motivan.

A mis familiares y seres queridos, por su amor incondicional.

A los profesores de mi alma máter, la UNALM, que contribuyeron compartiendo de sus conocimientos y en especial a mi asesor Iván Soto Rodríguez por su motivación y constante apoyo.

ÍNDICE GENERAL

I.	INTRODUCCIÓN	1
1.1.	Problemática	1
1.2.	Objetivos.....	2
II.	MARCO TEÓRICO.....	4
2.1.	Datos, información y conocimiento.....	4
2.2.	Inteligencia de Negocios	5
2.3.	Datawarehouse vs Datamart	6
2.4.	Metodologías	7
2.4.1.	Metodología Bill Inmon	8
2.4.2.	Metodología Ralph Kimball.....	8
2.4.3.	Comparación entre Bill Inmon y Ralph Kimball	9
2.5.	Proceso de Extracción, Transformación y Carga de datos (ETL)	10
2.5.1.	Herramientas de ETL	10
2.6.	Modelamiento de datos (Modelo dimensional)	11
2.6.1.	Modelo estrella	11
2.6.2.	Modelo copo de nieve	12
2.7.	Cuadro de mando (Dashboard).....	12
III.	MARCO METODOLÓGICO	15
3.1.	Planificación del proyecto	15
3.1.1.	El negocio.....	15
3.1.2.	Objetivos del proyecto	15
3.1.3.	Alcance del proyecto	16
3.1.4.	Justificación del proyecto.....	16
3.1.5.	Beneficios del proyecto	16
3.1.6.	Cronograma de actividades	18
3.1.7.	Equipo de trabajo	19
3.1.8.	Riesgos del proyecto	20
3.2.	Definición de requerimientos del negocio.....	21

3.2.1.	Reportes e informes.....	21
3.2.2.	Entrevistas	21
3.2.3.	Análisis de las fuentes de información.....	22
3.3.	Diseño de la arquitectura técnica.....	24
3.3.1.	Back Room.....	25
3.3.2.	Front Room	25
3.4.	Selección de productos e implementación.....	26
3.5.	Modelado dimensional	27
3.5.1.	Elegir el proceso de negocio	27
3.5.2.	Establecer el nivel de granularidad	27
3.5.3.	Elegir las dimensiones.....	27
3.5.4.	Identificar las tablas de hechos y medidas	37
3.6.	Diseño físico.....	39
3.6.1.	Sentencias SQL para crear las dimensiones.....	39
3.6.2.	Sentencias SQL para crear la tabla de hechos y medidas	50
3.7.	Diseño e implementación del subsistema de ETL.....	58
3.7.1.	ETL para las tablas dimensiones.....	58
3.7.2.	ETL para la tabla de hechos o tabla fact	65
3.8.	Especificación de aplicaciones de BI	67
3.9.	Desarrollo de aplicaciones de BI.....	68
3.10.	Implementación.....	71
3.10.1.	Plan de implementación	71
3.10.2.	Acceso a la herramienta de visualización Power BI	72
3.11.	Mantenimiento y crecimiento	72
3.11.1.	Mantenimiento	72
3.11.2.	Crecimiento	72
IV.	RESULTADOS Y DISCUSIÓN.....	74
V.	CONCLUSIONES Y RECOMENDACIONES	76
5.1.	Conclusiones.....	76
5.2.	Recomendaciones	76
VI.	REFERENCIAS BIBLIOGRÁFICAS	78

ÍNDICE DE TABLAS

Tabla 1: Comparación entre las metodologías Bill Inmon y Ralph Kimball	9
Tabla 2: Comparación herramientas ETL	10
Tabla 3: Cronograma de actividades para el desarrollo de las etapas de la metodología Ralph Kimball	18
Tabla 4: Equipo a cargo del desarrollo del proyecto.....	19
Tabla 5: Perfiles de usuarios del Datamart.....	20
Tabla 6: Características del hardware del servidor y de los usuarios del Datamart.....	26
Tabla 7: Software disponible para el desarrollo del proyecto	26
Tabla 8: Descripción de la estructura de la dim_area	27
Tabla 9: Descripción de la estructura de la dim_area_curricular	28
Tabla 10: Descripción de la estructura de la dim_caracteristica2	28
Tabla 11: Descripción de la estructura de la dim_caracteristica3	28
Tabla 12: Descripción de la estructura de la dim_dre.....	29
Tabla 13: Descripción de la estructura de la dim_estudiante.....	29
Tabla 14: Descripción de la estructura de la dim_grado.....	30
Tabla 15: Descripción de la estructura de la dim_incluir_resultados	30
Tabla 16: Descripción de la estructura de la dim_institucion	31
Tabla 17: Descripción de la estructura de la dim_lengua_eib	31
Tabla 18: Descripción de la estructura de la dim_lengua_materna	32
Tabla 19: Descripción de la estructura de la dim_n_ise.....	32
Tabla 20: Descripción de la estructura de la dim_nivel.....	32
Tabla 21: Descripción de la estructura de la dim_nivel_logro.....	33
Tabla 22: Descripción de la estructura de la dim_periodo.....	33
Tabla 23: Descripción de la estructura de la dim_seccion	33
Tabla 24: Descripción de la estructura de la dimsexo.....	34
Tabla 25: Descripción de la estructura de la dim_tipo_discapacidad_rnee	34
Tabla 26: Descripción de la estructura de la dim_subtipo_discapacidad_rnee.....	35
Tabla 27: Descripción de la estructura de la dim_tipo_evaluacion	35
Tabla 28: Descripción de la estructura de la dim_tipo_gestion2	36

Tabla 29: Descripción de la estructura de la dim_tipo_gestion3	36
Tabla 30: Descripción de la estructura de la dim_ubicacion	36
Tabla 31: Descripción de la estructura de la dim_ugel	37
Tabla 32: Descripción de la estructura de la fact_resultados_estudiantes	38

ÍNDICE DE FIGURAS

Figura 1. Pirámide de relación entre Datos, Información y Conocimiento.....	5
Figura 2. Arquitectura BI	6
Figura 3. Etapas de la metodología Ralph Kimball	9
Figura 4. Cuadrante Mágico de Gartner 2021 para plataformas de Análisis e Inteligencia de Negocios	14
Figura 5. Arquitectura del Datamart del área de Análisis de la Instancia Técnica	24
Figura 6. Diseño físico del Datamart del área de Análisis de la Instancia Técnica	57
Figura 7. Pasos realizados en el ETL de las tablas de dimensiones.....	58
Figura 8. Pasos realizados en el ETL de la tabla dim_area.....	59
Figura 9. Pasos realizados en el ETL de la tabla dim_area_curricular	59
Figura 10. Pasos realizados en el ETL de la tabla dim_caracteristica2	59
Figura 11. Pasos realizados en el ETL de la tabla dim_caracteristica3	60
Figura 12. Pasos realizados en el ETL de la tabla dim_dre	60
Figura 13. Pasos realizados en el ETL de la tabla dim_estudiante	60
Figura 14. Pasos realizados en el ETL de la tabla dim_grado	60
Figura 15. Pasos realizados en el ETL de la tabla dim_incluir_resultados.....	61
Figura 16. Pasos realizados en el ETL de la tabla dim_institucion.....	61
Figura 17. Pasos realizados en el ETL de la tabla dim_lengua_eib.....	61
Figura 18. Pasos realizados en el ETL de la tabla dim_lengua_materna.....	61
Figura 19. Pasos realizados en el ETL de la tabla dim_n_ise	62
Figura 20. Pasos realizados en el ETL de la tabla dim_nivel	62
Figura 21. Pasos realizados en el ETL de la tabla dim_nivel_logro	62
Figura 22. Pasos realizados en el ETL de la tabla dim_periodo	62
Figura 23. Pasos realizados en el ETL de la tabla dim_seccion	63
Figura 24. Pasos realizados en el ETL de la tabla dimsexo	63
Figura 25. Pasos realizados en el ETL de la tabla dim_tipo_discapacidad_rnee.....	63
Figura 26. Pasos realizados en el ETL de la tabla dim_subtipo_discapacidad_rnee	63
Figura 27. Pasos realizados en el ETL de la tabla dim_tipo_evaluacion.....	64
Figura 28. Pasos realizados en el ETL de la tabla dim_tipo_gestion2.....	64

Figura 29. Pasos realizados en el ETL de la tabla dim_tipo_gestion3.....	64
Figura 30. Pasos realizados en el ETL de la tabla dim_ubicacion.....	64
Figura 31. Pasos realizados en el ETL de la tabla dim_ugel.....	65
Figura 32. Pasos realizados en el ETL de la tabla de hechos o tabla fact.....	65
Figura 33. Pasos realizados en el ETL de la tabla fact_resultados_estudiantes.....	66
Figura 34. Reporte de resultados a nivel de estudiante innominado, evaluados en la ECE 2015 en 2.º grado de primaria.....	69
Figura 35. Reporte con la cantidad de IE y estudiantes evaluados en la ECE desde el 2007 al 2018, por lengua EIB.....	69
Figura 36. Reporte de resultados a nivel nacional por tipo de evaluación, grado y nivel, área curricular, periodo y los estratos sexo, gestión y área, según niveles de logro (porcentaje) y medida promedio.....	70
Figura 37. Reporte de resultados a nivel regional por tipo de evaluación, grado y nivel, área curricular, periodo y DRE, según niveles de logro (porcentaje) y medida promedio.....	71

RESUMEN

La Instancia Técnica del sector educativo al que hace referencia el presente proyecto, posee gran cantidad de datos históricos producto de las evaluaciones de logros de aprendizaje que realiza anualmente. Sin embargo, no cuenta con un almacén de bases de datos estructurado, integrado y ordenado que permita el uso de esta información de manera óptima. En consecuencia, demanda de mucho tiempo atender los pedidos de información y generar reportes, además de que por ser este un proceso manual se incurre muchas veces al error. Por todo ello, surgió la necesidad de incorporar soluciones tecnológicas e innovadoras, como es el proceso de Inteligencia de Negocios. Por consiguiente, el presente proyecto tiene como objetivo principal la implementación de un Datamart para el área de Análisis de la Instancia Técnica, con la finalidad de gestionar la información y apoyar en la toma de decisiones. Esta implementación se realizó empleando la metodología de Ralph Kimball describiendo detalladamente cada etapa. Además, cabe mencionar, que para la implementación del Datamart se utilizaron únicamente softwares libres, como el Pentaho Data Integration – Kettle (herramienta de extracción, transformación y carga) y el MySQL Workbench (herramienta de diseño y gestión de bases de datos). Adicionalmente, para elaborar los reportes e informes, se utilizaron las herramientas Microsoft Excel y Power BI. La implementación del Datamart permitió automatizar procesos, con la creación de consultas para la atención de los pedidos de información y generación de reportes, originando que los tiempos de respuesta disminuyeran. Además, toda la información ingresada al Datamart fue validada por lo que ahora es confiable y consistente. Se generaron también accesos para los usuarios en base a sus perfiles. Adicionalmente, se crearon Dashboards para el monitoreo, análisis y visualización de los principales indicadores y métricas para la toma de decisiones acertadas.

Palabras clave: Instancia Técnica, Inteligencia de Negocios, Datamart.

ABSTRACT

The Technical Instance of the education sector to which this project refers has a large amount of historical data from its annual learning achievement assessments. However, it does not have a structured, integrated and orderly database warehouse that allows for the optimal use of this information. As a result, it is very time-consuming to respond to requests for information and to generate reports, in addition to the fact that this is a manual process and often leads to errors. For all these reasons, the need arose to incorporate innovative technological solutions, such as the Business Intelligence process. Therefore, the main objective of this project is the implementation of a Datamart for the Analysis area of the Technical Instance, in order to manage information and support decision making. This implementation was carried out using Ralph Kimball's methodology, describing each stage in detail. Furthermore, it is worth mentioning that only free software was used for the implementation of the Datamart, such as Pentaho Data Integration - Kettle (extraction, transformation and loading tool) and MySQL Workbench (database design and management tool). In addition, Microsoft Excel and Power BI tools were used to produce reports. The implementation of Datamart made it possible to automate processes, with the creation of queries for information requests and the generation of reports, thus reducing response times. In addition, all information entered into the Datamart was validated and is now reliable and consistent. Access was also generated for users based on their profiles. Additionally, Dashboards were created for monitoring, analysis and visualization of the main indicators and metrics for accurate decisions making.

Keywords: Technical Instance, Business Intelligence, Datamart.

I. INTRODUCCIÓN

1.1. Problemática

En la actualidad, con los avances tecnológicos, las organizaciones se han convertido en grandes generadores de datos, sin embargo, no se trata de cuantos datos almacenes, sino de cómo usas esos datos para transformarlo en información y generar conocimiento que apoye en la toma de decisiones.

El sector educativo se caracteriza por ser uno de los sectores del cual se obtienen grandes volúmenes de datos, esto debido al tipo de actividades que realizan en este sector (evaluaciones), y al nivel de detalle con que se pueden obtener estos datos (resultados a nivel nacional, regional, local, de institución educativa y de estudiante).

La Instancia Técnica a la que hace referencia el presente proyecto no es ajena a esto, posee gran cantidad de datos históricos producto de las evaluaciones de logros de aprendizaje que realiza anualmente, a estudiantes de educación básica a nivel nacional, los cuales son: la Evaluación Censal de Estudiantes (ECE) y la Evaluación Muestral (EM). Sin embargo, no cuenta con un almacén de bases de datos estructurado, integrado y ordenado que permita el uso de esta información de manera óptima. Sus bases de datos, provenientes de múltiples fuentes de datos y en diferentes formatos, se encuentran dispersas (islas de información), almacenados en distintas carpetas y de fácil acceso a cualquier usuario, por lo que modificar o eliminar algún registro se hace muy frecuente. En consecuencia, demanda de mucho tiempo atender los pedidos de información y generar reportes, además de que por ser este un proceso manual se incurre muchas veces al error. Por todo ello, contar con información precisa, confiable y oportuna resulta una difícil tarea en el día a día. Adicionalmente, el tiempo que se invierte en el proceso

actual podría servir para hacer un análisis más exhaustivo de la información y para generar indicadores que apoyen en la toma de decisiones de la organización para la consecución de los objetivos.

Debido a la necesidad expuesta, se requiere incorporar soluciones tecnológicas e innovadoras, como es el proceso de Inteligencia de Negocios, para una mejor gestión de la información. Por ello, el presente proyecto tiene como finalidad la implementación de un Datamart que permita a los usuarios filtrar, clasificar, analizar y visualizar los datos de una manera rápida y sencilla, usando consultas para la atención de los pedidos de información y generación de reportes; y empleando Dashboards (cuadros de mando) para el monitoreo, análisis y visualización de los principales indicadores y métricas para la toma de decisiones acertadas. Adicional a ello, permitirá tener un control de acceso a los datos, en base a los perfiles de cada usuario. Es importante mencionar también, que el uso del Datamart no requerirá de personal especializado en TI, sino que será de fácil uso y comprensión para los usuarios, por lo que se convierte en una gran solución para la organización.

1.2. Objetivos

Objetivo principal

Diseñar e implementar un Datamart para el área de Análisis, con la finalidad de gestionar la información y apoyar en la toma de decisiones.

Objetivos específicos

- Mejorar los tiempos de respuesta en la atención de los pedidos de información y generación de reportes.
- Mejorar la consistencia de la información brindada.

- Crear informes empleando herramientas de visualización de datos para la toma de decisiones acertadas.

II. MARCO TEÓRICO

2.1. Datos, información y conocimiento

Según Oltra Badenes (2017), para una mejor comprensión de lo que es la Inteligencia de Negocios, resulta relevante conocer y diferenciar los conceptos de datos, información y conocimiento; ya que si bien estos conceptos se encuentran muy relacionados, no tienen el mismo significado: a los datos los define como un conjunto de caracteres (o símbolos) que por sí solos carecen de significado, también menciona que son la fuente de la que se extrae información; a la información la define como dato o conjunto de datos, que procesados o transformados, y puestos en un contexto determinado, adquieren significado; y por último define al conocimiento como información interpretada y relacionada, propiciador de la toma de decisiones. De manera similar Llombart (2003) define a los datos, información y conocimiento de la siguiente manera: a los datos como materia prima bruta, y una vez que a estos el usuario les atribuye algún significado en especial, se convierten en información, y a la interpretación de este último, con valor añadido, lo denomina conocimiento; también hace referencia a la relación que existe entre estos tres conceptos que son muy usados en la Inteligencia de Negocios, indica que en una pirámide, a menor jerarquía, mayor volumen de datos pero menor valor que los responsables de la toma de decisiones le dan a esa jerarquía, y de manera inversa, a mayor jerarquía, menor volumen de datos pero mayor valor que los responsables de la toma de decisiones le dan a esa jerarquía.



Figura 1. Pirámide de relación entre Datos, Información y Conocimiento

2.2. Inteligencia de Negocios

Según Joyanes Aguilar (2019), el concepto de Inteligencia de Negocios (Business Intelligence o BI) tiene su origen en los años 70, cuando se iniciaba con la automatización de las tareas en las empresas, pero que el término en sí fue adherido por Gartner a mitad de la década de los 90. La consultora Gartner, líder y referente mundial en TI (tecnologías de la información) y consultoría estratégica, en su reconocido glosario de términos, define la Inteligencia de Negocios como “un término general que incluye las aplicaciones, la infraestructura y las herramientas, y las mejores prácticas que permiten el acceso y el análisis de la información para mejorar y optimizar las decisiones y el rendimiento” (Gartner, s.f.). De manera similar, Medina La Plata (2013) define este concepto como “el conjunto de estrategias y herramientas enfocadas en la administración y creación de conocimiento mediante el análisis de datos existente en una organización”, y que por medio de estos sistemas de Inteligencia de Negocios se obtiene un mejor acceso a la información de la organización y ello permitirá realizar un mejor análisis que ayude en la toma de decisiones de todos sus usuarios. En Vásquez Castrillón & Sucerquia Osorio (2011) se menciona que la Inteligencia de Negocios es una herramienta del que se obtienen dos ventajas: una ventaja tecnológica, ya que permite la centralización, limpieza y afianzamiento de los datos, hallar información nueva que no era visible con las aplicaciones actuales y mejorar el rendimiento de los sistemas, y una ventaja competitiva, ya que permite dar acompañamiento al plan estratégico, mejorar las experiencias anteriores y optimizar la competitividad.

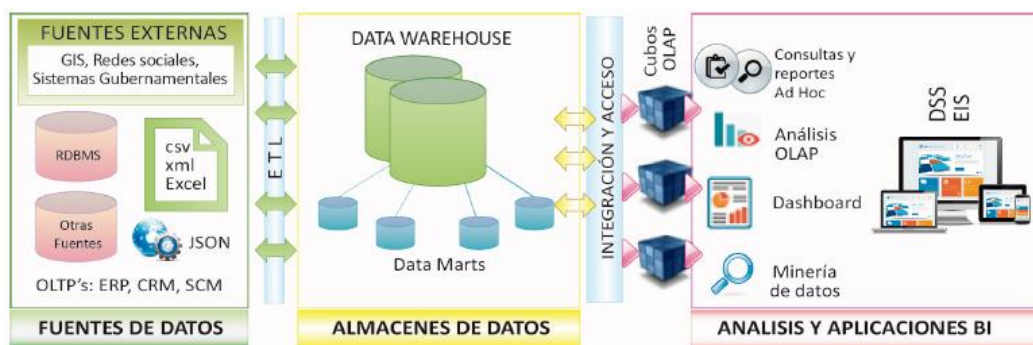


Figura 2. Arquitectura BI

Nota. Adaptado de “Dashboard para el soporte de decisiones en una empresa del sector minero” (p. 1220), por Mazón Olivo, y otros, 2017.

2.3. Datawarehouse vs Datamart

Es importante definir y diferenciar los conceptos de Datawarehouse y Datamart, ya que ambos hacen referencia a almacenes de bases de datos construidos con el objetivo de facilitar el análisis de los datos. En Joyanes Aguilar (2019) se hace la siguiente diferencia entre ambos conceptos: Mientras que el Datawarehouse se centra en una visión global e integra los datos de toda una organización, el Datamart se centra en las necesidades específicas de un departamento y su ventaja principal respecto al Datawarehouse es que por ser de menor tamaño es más fácil de construir. En Silva Solano (2017) se define al Datawarehouse como “una base de datos integral que consolida, estandariza y almacena datos actuales e históricos de potencial interés para los encargados de la toma de decisiones en la empresa”. Por su parte Mosquera & Hallo (2014) definen al Datamart como “un subconjunto de datos orientado al análisis de un área del negocio”. Adicionalmente Llombart (2003) define al Datamart como “un almacén de datos departamental, que reúne datos históricos pero tratados”. Por lo antes expuesto, la diferencia entre ambos almacenes de bases de datos radica en su alcance.

Existen casos de implementación del Datamart, en diferentes sectores, que resultaron ser exitosos, algunos de ellos son los siguientes:

- En Zambrano Tirado (2018), se implementó un Datamart para la gestión de incidencias en el centro de atención al usuario en la sede central del Ministerio de Educación, llegando a la conclusión de que el Datamart mejoró la gestión de incidentes, ya que “permitió incrementar el nivel de servicio del 61.01% al 75%, del mismo modo, se mejoró la atención del escalamiento jerárquico del 32.98% al 68.15%”.
- Asto Huamán & Arangüena Yllanes (2018), en su investigación para determinar el impacto que genera el uso del Datamart en la gestión académica en el sector de la educación superior, concluyeron que resultó ser eficiente frente a la toma de decisiones, ya que usando esta herramienta “se determinó tiempos de respuesta eficientes con un alto grado de confiabilidad debido a que se validan las reglas de negocio (Dimensiones establecidas por la metodología de Ralph Kimball) con la ISO/IEC 9126-4 (métricas de calidad en uso)”.
- Según Zamora Saldarriaga (2017), la implementación de un Datamart “influyó satisfactoriamente para la mejora en la toma de decisiones del control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional”; de igual manera, debido a la accesibilidad, fiabilidad, usabilidad y eficiencia que caracteriza al Datamart propuesto, este tuvo un impacto positivo para el cumplimiento de los procesos.

2.4. Metodologías

El uso de una metodología en la implementación de un Datamart resulta ser imprescindible, para garantizar que cada etapa está siendo ejecutada según lo previsto, con el fin de lograr el objetivo principal.

Hoy en día existen varias metodologías que se pueden emplear, sin embargo, “desde que se intentó por primera vez la construcción de almacenes de datos a principios de la década de 1990, han surgido dos modelos como dominantes: el de Inmon y el de Kimball” (Breslin, 2004).

2.4.1. Metodología Bill Inmon

Utiliza el enfoque “Top-Down” (descendente), en el propone que el proyecto parta de lo general a lo particular, es decir, que se inicie con la construcción del Datawarehouse y que de este se originen los Datamarts.

Inmon propone un modelo de datos de tres niveles:

- 1) Primer nivel: ERD (Entity Relationship Diagrams)
- 2) Segundo nivel: DIS (Data Item Set)
- 3) Tercer nivel: Modelo Físico (Physical Model)

Una vez implementado los tres niveles del modelo de datos utiliza la adaptación especial de la metodología de desarrollo en espiral. Adicionalmente, indica también que la implementación de los Datamarts debe realizarse con modelamiento dimensional.

2.4.2. Metodología Ralph Kimball

Utiliza el enfoque “Bottom-Up” (ascendente), en el propone que el proyecto parta de lo particular a lo general, es decir, que se inicie con la construcción de los Datamarts y que la unión de estos dé origen al Datawarehouse.

Esta metodología emplea el modelo dimensional y se basa en lo que Kimball denomina Ciclo de Vida Dimensional del Negocio (Business Dimensional Lifecycle), el cual tiene cuatro principios básicos: centrarse en el negocio, construir una infraestructura de información adecuada, realizar entregas en incrementos significativos y ofrecer la solución completa.

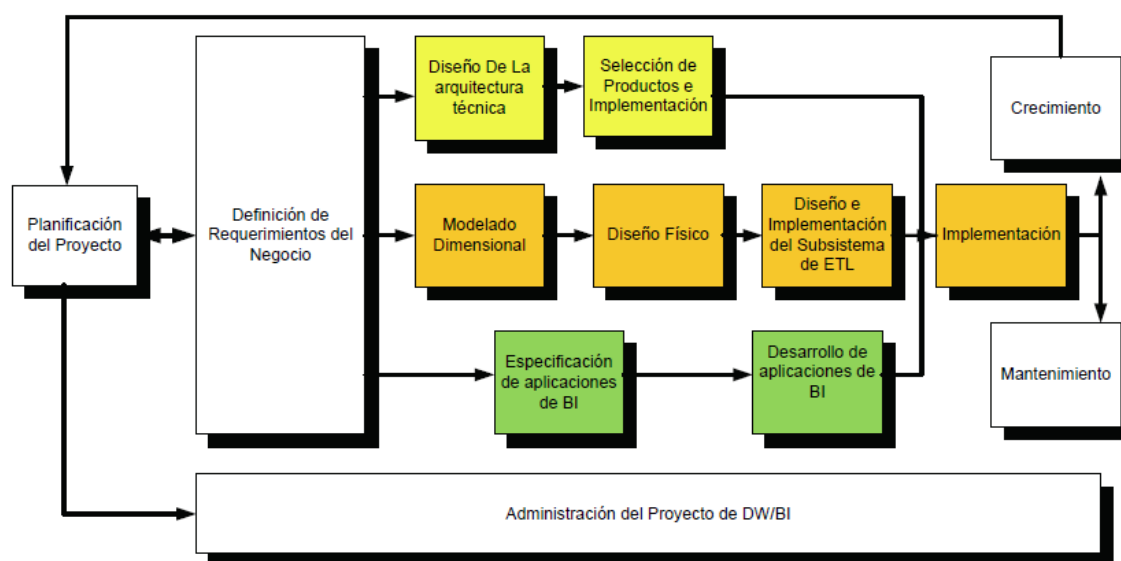


Figura 3. Etapas de la metodología Ralph Kimball

Nota. Adaptado de “La metodología de Kimball para el diseño de almacenes de datos (Data warehouses)” (p. 59), por Rivadera, 2010.

2.4.3. Comparación entre Bill Inmon y Ralph Kimball

Tabla 1: Comparación entre las metodologías Bill Inmon y Ralph Kimball

Factor	Metodología	
	Bill Inmon	Ralph Kimball
Presupuesto	El costo inicial es alto	El costo inicial es bajo
Plazos	Requiere de mayor tiempo para su desarrollo	Requiere de menor tiempo para su desarrollo
Especialización	Requiere de un equipo con especialización alta	Requiere de un equipo con especialización media
Alcance	Toda la organización	Áreas específicas de la organización
Mantenimiento	Es de fácil mantenimiento	El mantenimiento es más complejo

Para la implementación del presente proyecto, se optó por desarrollar la metodología Kimball creada por Ralph Kimball. Se escogió esta metodología, debido a su enfoque “Bottom-Up” (ascendente), ya que como punto de partida solo se requiere la construcción de un Datamart para el área de Análisis de la Instancia Técnica y esta metodología se ajusta perfectamente a las necesidades del proyecto. También se tomaron en cuenta dos factores muy importantes como el costo y el tiempo.

2.5. Proceso de Extracción, Transformación y Carga de datos (ETL)

El proceso ETL (Extract, Transform, Load) es uno de los procesos de mayor relevancia en la implementación de un almacén de datos. Este proceso se divide en tres etapas:

- **Extracción**

Consiste en extraer los datos desde las fuentes de origen, estos pueden ser desde simples hojas de cálculo hasta sistemas transaccionales u operacionales.

- **Transformación**

Consiste en transformar los datos extraídos previamente, estas modificaciones pueden ser simples conversiones, como cambiar el tipo de dato de entero a carácter hasta realizar cálculos generando algún valor con los datos extraídos, todo ello con el fin de generar bases de datos depuradas, consistentes y confiables.

- **Carga**

Consiste en trasladar los datos, procedentes de las etapas anteriores, al almacén de datos.

2.5.1. Herramientas de ETL

Existen muchas herramientas encargadas del proceso de ETL, sin embargo, para el presente proyecto se optará por utilizar una herramienta de software libre. A continuación, se compararán tres herramientas de gran utilidad, debido a su capacidad para extraer, transformar y cargar datos (Ruiz Borja, 2018):

Tabla 2: Comparación herramientas ETL

Herramienta	Tipo	Plataforma	Conexión a bases de datos
Pentaho Data Integration – Kettle (PDI)	Con licencia y versión gratuita	Windows, Linux y Mac	MySQL, Oracle, PostgreSQL, entre otros
Talend Data Integration (TDI)	Con licencia y versión gratuita	Windows y Mac	MySQL, Oracle, Microsoft SQL Server, entre otros
OpenRefine (OR)	Versión gratuita	Windows, Linux y Mac	MySQL, PostgreSQL y MariaDB

La herramienta que se seleccionó para el desarrollo del proceso de ETL del presente proyecto es el Pentaho Data Integration – Kettle (PDI), ya que si bien las tres herramientas descritas cumplen con los requisitos (software libre, plataforma Windows y la conexión a la base de datos MySQL), ya se contaba con conocimiento previo en la herramienta seleccionada.

2.6. Modelamiento de datos (Modelo dimensional)

Existen dos tipos de modelos que frecuentemente se utilizan al implementar un almacén de datos:

2.6.1. Modelo estrella

Es aquel modelo conformado por una tabla de hechos o tabla fact, y este a su vez se encuentra rodeado de las tablas de dimensiones. Todas las tablas de dimensiones tienen una columna que es la llave primaria (Primary Key o PK), el cual identifica de manera única a cada fila de la tabla. La tabla de hechos o tabla fact está conformada por las llaves primarias de las tablas de dimensiones, que vendrían a ser llaves foráneas en la tabla de hechos o tabla fact. Adicionalmente, esta última tabla contiene también las métricas o medidas.

Este modelo se caracteriza también por tener una tabla de dimensión por cada dimensión, lo que origina que tenga un bajo número de joins al realizar una consulta, esto hace que sus queries sean menos complejos y como consecuencia se realizan ejecuciones más rápidas. Adicionalmente se debe mencionar que, en este modelo las tablas de dimensiones y tabla de hechos o tabla fact se encuentran desnormalizadas.

2.6.2. Modelo copo de nieve

Este modelo presenta una estructura más compleja que el modelo estrella, debido a que algunas de las tablas de dimensiones no se encuentran relacionadas directamente con la tabla de hechos o tabla fact, sino que se encuentran relacionadas con ella por medio de otras tablas de dimensiones, lo que origina que tenga un alto número de joins al realizar una consulta, esto hace que sus queries sean más complejos y como consecuencia se toma más tiempo en realizar las ejecuciones debido a los cruces. Adicionalmente se debe mencionar que, en este modelo las tablas de dimensiones se encuentran normalizadas y la tabla de hechos o tabla fact desnormalizada.

Para el presente proyecto se optó por el modelo estrella, ya que se ajustaba más a las necesidades del proyecto y porque es el modelo más simple de interpretar, con mejores tiempos de respuesta y su diseño es sencillo de mantener y actualizar.

2.7. Cuadro de mando (Dashboard)

Otro concepto muy popular en el proceso de Inteligencia de Negocios son los denominados Dashboards, por lo que resulta de gran importancia conocer su definición. Según Reyes Dixson & Nuñez Maturel (2015), un cuadro de mando o Dashboard (en inglés) es una aplicación que permite la visualización de información, métricas e indicadores, cuyo uso tiene como objetivo mostrar el contenido de un sistema de soporte para la toma de decisiones de los usuarios finales. Este tipo de gráficos permite al usuario interactuar con la información contenida a un nivel de mayor detalle y de acuerdo con sus necesidades. De manera similar, Joyanes (2019) define un cuadro de mando (Dashboard) como “un panel de datos en el que las empresas visualizan la información más importante con el objetivo final de conseguir su mayor optimización”. En cambio, Flórez Fernández (2012) brinda información adicional acerca de los cuadros de mando indicando que “esta herramienta debe proveer interfaz personalizable y la capacidad de incluir datos de múltiples fuentes de información en tiempo real”.

Existen también casos de implementación del Dashboards, en diferentes sectores, que resultaron ser exitosos, algunos de ellos son los siguientes:

- En Pérez Paucar (2019) se implementó un Dashboard “para identificar las variables y dimensiones que tienen mayor impacto en el área educativa”, concluyendo que dicha herramienta permitió una mejor gestión y administración de la información, siendo esta ahora confiable y oportuna, logrando así una acertada toma de decisiones.
- En Mazón Olivo, y otros (2017) el objetivo era el de “implementar un Dashboard para la gestión de un sistema de soporte de decisiones (SSD) en las áreas de: contabilidad, producción y recursos humanos, de una empresa minera” concluyendo que dicha implementación logró que la empresa contara con información integrada y oportuna, permitiendo a sus usuarios interactuar con la información brindada, y realizar los análisis necesarios para tomar las mejores decisiones.

Para crear los Dashboards en el presente proyecto, se optó por utilizar la herramienta de visualización Power BI, ya que es una de las herramientas más completas, y porque según Gartner, uno de los máximos exponentes en lo que es Inteligencia de Negocios, en su Cuadrante Mágico 2021 para plataformas de Análisis e Inteligencia de Negocios, esta herramienta se posiciona como líder superando a las demás.



Figura 4. Cuadrante Mágico de Gartner 2021 para plataformas de Análisis e Inteligencia de Negocios
 Nota. Adaptado de <https://powerbi.microsoft.com/en-us/blog/microsoft-named-a-leader-in-2021-gartner-magic-quadrant-for-analytics-and-bi-platforms/>

III. MARCO METODOLÓGICO

Esta sección contiene el desarrollo de cada una de las etapas de la metodología Ralph Kimball, el cual se eligió por las razones expuestas en la sección anterior, para el diseño e implementación del Datamart del área de Análisis de la Instancia Técnica.

3.1. Planificación del proyecto

3.1.1. El negocio

La entidad a la que hace referencia el presente proyecto es una Instancia Técnica del sector educativo, responsable de diseñar e implementar evaluaciones de logros de aprendizaje a estudiantes de educación básica a nivel nacional. Tiene como objetivo producir y difundir información relevante y confiable sobre los resultados de estas evaluaciones y sus factores asociados, para contribuir a la toma de decisiones en las instancias de decisión política educativa, la comunidad educativa y la sociedad en general, con el fin de mejorar la calidad del sistema educativo del país.

3.1.2. Objetivos del proyecto

Objetivo general:

Diseñar e implementar un Datamart para el área de Análisis, con la finalidad de gestionar la información y apoyar en la toma de decisiones.

Objetivos específicos:

- 1) Mejorar los tiempos de respuesta en la atención de los pedidos de información y generación de reportes.
- 2) Mejorar la consistencia de la información brindada.
- 3) Crear informes empleando herramientas de visualización de datos para la toma de decisiones acertadas.

3.1.3. Alcance del proyecto

Permitirá al área de Análisis de la Instancia Técnica contar con datos centralizados para gestionarlos, entenderlos y confiar en ellos, mediante el uso de herramientas tecnológicas.

3.1.4. Justificación del proyecto

El área de Análisis de la Instancia Técnica posee gran cantidad de datos históricos producto de las evaluaciones de logros de aprendizaje que realiza anualmente, a estudiantes de educación básica a nivel nacional. Sin embargo, no cuenta con un almacén de bases de datos estructurado, integrado y ordenado que permita el uso de esta información de manera óptima. Es por ello que en el área surge la necesidad de implementar un Datamart que solucione la problemática descrita.

3.1.5. Beneficios del proyecto

- Datos integrados, estructurados, ordenados, actualizados, consistentes, de alta calidad y disponibles.
- Control de acceso a los datos según perfil de usuario.

- Reducción de los tiempos de respuesta en la atención de los pedidos de información y generación de reportes.
- Informes interactivos con información clara e intuitiva, que permite la monitorización y unificación de los datos relevantes.
- Apoyo en la toma de decisiones.

3.1.6. Cronograma de actividades

Tabla 3: Cronograma de actividades para el desarrollo de las etapas de la metodología Ralph Kimball

Actividad	Semana																							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Planificación del proyecto	X	X																						
Definición de requerimientos del negocio			X	X	X																			
Diseño de la arquitectura técnica						X	X																	
Selección de productos e implementación								X																
Modelado dimensional									X	X	X													
Diseño físico												X	X											
Diseño e implementación del subsistema de ETL														X	X	X								
Especificación de aplicaciones de BI																	X							
Desarrollo de aplicaciones de BI																		X	X	X				
Implementación																					X	X	X	
Mantenimiento y crecimiento																								X

3.1.7. Equipo de trabajo

La Instancia Técnica tiene a su cargo el área de Análisis conformado por cuatro equipos, entre ellos el equipo de Estadística y Gestión de datos. El área está liderada por un coordinador y cada equipo por un subcoordinador y a su vez este está conformado por especialistas con sólidos conocimientos en su campo. El presente proyecto se implementó para el área de Análisis de la Instancia Técnica, por lo que los usuarios son todos sus integrantes. El equipo a cargo y responsable de la implementación fue el equipo de Estadística y Gestión de datos.

a) Personas involucradas en el proyecto

El equipo a cargo del desarrollo del proyecto estuvo conformado por:

Tabla 4: Equipo a cargo del desarrollo del proyecto

Cargo	Función
Subcoordinador del equipo de Estadística	Supervisar que todos los procesos del proyecto se cumplan.
Especialista en Gestión de datos	Brindar soporte al proyecto.
Especialista en Estadística	Desarrollo del proyecto.

b) Usuarios del Datamart

El perfil de usuario que tiene cada integrante del área de Análisis de la Instancia Técnica es el siguiente:

Tabla 5: Perfiles de usuarios del Datamart

Cargo	Perfil de usuario
Coordinador del área de Análisis (1)	Consulta
Subcoordinador del equipo de Estadística (1)	Administrador
Otros subcoordinadores del área de Análisis (3)	Consulta
Especialista en Gestión de datos (1)	Administrador
Especialista en Estadística (1)	Administrador
Otros especialistas del área de Análisis (13)	Consulta

3.1.8. Riesgos del proyecto

Algunos de los riesgos que se podrían haber presentado durante la implementación del proyecto son los siguientes:

- Rotación de personal, lo que implicaría capacitar y dar a conocer el negocio al nuevo personal, generando retrasos en el tiempo establecido para la implementación del proyecto.
- Falta de compromiso o interés por parte del personal involucrado en el desarrollo del proyecto.
- Cambios en los requerimientos, lo que generaría retrasos en el desarrollo del proyecto.
- Tiempo mal planificado para la implementación del proyecto.

3.2. Definición de requerimientos del negocio

Para conocer los requerimientos del negocio y tener un mejor panorama acerca de las necesidades que tenía el área, se consideraron las siguientes fuentes de información:

3.2.1. Reportes e informes

Se realizó una investigación previa a la entrevista, para conocer los reportes e informes que generaba el área. Para ello se revisaron correos y documentaciones (oficios), de los pedidos de información más frecuentes tanto de los stakeholders internos (pertenecientes a la Instancia Técnica) como externos (público en general), también se accedió a las carpetas compartidas donde se almacenan los reportes e informes. Adicionalmente se hizo una revisión de las bases de datos con las que cuenta la Instancia Técnica y finalmente se consultó documentación que brinda mayor detalle de los procesos que realiza cada equipo del área de Análisis.

3.2.2. Entrevistas

Se entrevistó a los subcoordinadores y especialistas de los equipos que integran el área de Análisis (usuarios) y a los responsables en la toma de decisiones de la Instancia Técnica para conocer sus requerimientos. Los integrantes de cada equipo expusieron sus funciones, los análisis de información que realizan, las dificultades y limitaciones a las que se han tenido que enfrentar por no contar con un almacén de bases de datos integrado, y lo que esperan obtener con la implementación del Datamart en el área. Por su parte los responsables en la toma de decisiones manifestaron la necesidad de contar con información confiable y oportuna.

3.2.3. Análisis de las fuentes de información

En base al análisis de las fuentes de información se identificaron los siguientes requerimientos:

a) Requerimientos funcionales

- Reporte de resultados a nivel nacional por periodo, grado y área curricular, según medida promedio.
- Reporte de resultados a nivel nacional por periodo, grado y área curricular, según niveles de logro (cantidad y porcentaje).
- Reporte de resultados a nivel nacional por periodo, grado, área curricular y sexo (Mujer/Hombre), según medida promedio.
- Reporte de resultados a nivel nacional por periodo, grado, área curricular y sexo (Mujer/Hombre), según niveles de logro (cantidad y porcentaje).
- Reporte de resultados a nivel nacional por periodo, grado, área curricular y lengua materna, según medida promedio.
- Reporte de resultados a nivel nacional por periodo, grado, área curricular y lengua materna, según niveles de logro (cantidad y porcentaje).
- Reporte de resultados a nivel nacional por periodo, grado, área curricular y nivel socioeconómico, según medida promedio.
- Reporte de resultados a nivel nacional por periodo, grado, área curricular y nivel socioeconómico, según niveles de logro (cantidad y porcentaje).
- Reporte de resultados a nivel nacional por periodo, grado, área curricular y gestión (IE estatales/IE no estatales), según medida promedio.
- Reporte de resultados a nivel nacional por periodo, grado, área curricular y gestión (IE estatales/IE no estatales), según niveles de logro (cantidad y porcentaje).
- Reporte de resultados a nivel nacional por periodo, grado, área curricular y área (IE urbanas/IE rurales), según medida promedio.
- Reporte de resultados a nivel nacional por periodo, grado, área curricular y área (IE urbanas/IE rurales), según niveles de logro (cantidad y porcentaje).

- Reporte de resultados a nivel nacional por periodo, grado, área curricular y característica (IE polidocentes/IE multigrado), según medida promedio.
- Reporte de resultados a nivel nacional por periodo, grado, área curricular y característica (IE polidocentes/IE multigrado), según niveles de logro (cantidad y porcentaje).
- Reporte de resultados por periodo, grado, área curricular y DRE, según medida promedio.
- Reporte de resultados por periodo, grado, área curricular y DRE, según niveles de logro (cantidad y porcentaje).
- Reporte de resultados por periodo, grado, área curricular y UGEL, según medida promedio.
- Reporte de resultados por periodo, grado, área curricular y UGEL, según niveles de logro (cantidad y porcentaje).
- Reporte de resultados por periodo, grado, área curricular y distrito, según medida promedio.
- Reporte de resultados por periodo, grado, área curricular y distrito, según niveles de logro (cantidad y porcentaje).
- Reporte de resultados a nivel de IE, por grado, periodo y área curricular, según medida promedio y niveles de logro (cantidad y porcentaje).
- Reporte de resultados a nivel de estudiante nominado.
- Reporte de resultados a nivel de estudiante innominado.
- Reporte de resultados a nivel de estudiante, por tipo y subtipo de discapacidad del Registro de Necesidades Educativas Especiales (RNEE).
- Informe de IE y estudiantes participantes en las evaluaciones, por periodo, grado, área curricular y los estratos sexo, lengua materna, nivel socioeconómico, gestión, área, característica y lengua – Educación Intercultural Bilingüe (según corresponda).

b) Requerimientos no funcionales

- Accesibilidad al Datamart según los perfiles de usuarios establecidos.

- Disponibilidad del Datamart durante las 24 horas del día.
- Rendimiento óptimo del Datamart, disminuyendo los tiempos de respuesta en la generación de reportes.
- Interfaz de fácil uso y comprensión para los usuarios.
- Capacitación a usuarios no especializados en herramientas informáticas, sobre el correcto uso del Datamart.

3.3. Diseño de la arquitectura técnica

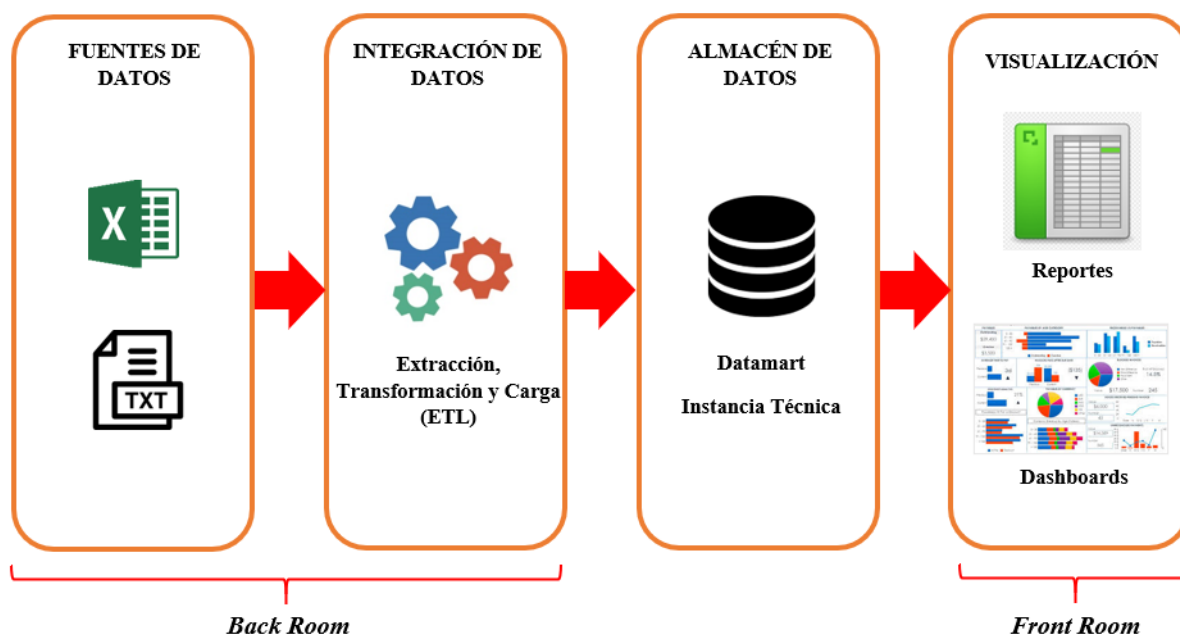


Figura 5. Arquitectura del Datamart del área de Análisis de la Instancia Técnica

En esta etapa se consolidaron tres factores: los requerimientos del negocio, el hardware y software de la Instancia Técnica, y los componentes requeridos en la implementación del Datamart. Para ello se tuvieron en cuenta dos escenarios: Back Room y Front Room. Asimismo, los componentes fundamentales que presenta la arquitectura son: los servicios, que son las funciones elementales para llevar a cabo las tareas que se requieren en el Datamart, y los repositorios de datos, que son lugares eventuales o permanentes donde se almacenan los datos.

A continuación, se describirá con mayor detalle los dos escenarios considerados en la arquitectura técnica.

3.3.1. Back Room

Abarca todos aquellos procedimientos y recursos necesarios para que se lleve a cabo el tratamiento de los datos, es decir el proceso de ETL.

Entre los servicios del Back Room están: la extracción, transformación y carga de datos y el control de tareas. De modo complementario, los repositorios de datos del Back Room son: los sistemas fuente, el área de ETL y los sistemas de administración de bases de datos (DBMS).

3.3.2. Front Room

Abarca todos aquellos procedimientos y recursos necesarios para mostrar los datos, previamente tratados, al usuario.

Entre los servicios del Front Room están: la navegación sobre el Datamart, el acceso y seguridad, la administración de consultas y los reportes estándar. De modo complementario, los repositorios de datos del Front Room son: las herramientas que dan acceso a los datos y las aplicaciones desarrolladas.

3.4. Selección de productos e implementación

a) Hardware

Tabla 6: Características del hardware del servidor y de los usuarios del Datamart

Características	Servidor	Usuarios
Marca	HP	Lenovo
Capacidad	Disco duro de 1 TB y 500GB	Disco duro de 1 TB
Procesador	Intel® Xeon® Gold 5118 CPU @ 2.30 GHz 2.29 GHz	Intel® Core™ i7-3770 CPU @ 3.40 GHz 3.40 GHz
Memoria RAM	32.0 GB	8.00 GB
Otros	Mouse y teclado	Mouse y teclado

b) Software

Tabla 7: Software disponible para el desarrollo del proyecto

Software	Descripción
MySQL	Motor de base de datos
Pentaho Data Integration (PDI) - Kettle	Herramienta ETL
MySQL Workbench 8.0.18	Herramienta de diseño y gestión de bases de datos
Microsoft Excel 2013	Herramienta de reportes
Power BI	Herramienta de visualización

Para la implementación del Datamart se usó únicamente productos de software libre, ya que no se contaba con un presupuesto asignado para dicho proyecto.

3.5. Modelado dimensional

3.5.1. Elegir el proceso de negocio

El proceso de negocio que se modelizó son los resultados obtenidos producto de las evaluaciones de logros de aprendizaje que la Instancia Técnica realiza anualmente, a estudiantes de educación básica a nivel nacional.

3.5.2. Establecer el nivel de granularidad

El nivel de granularidad definido, según los requerimientos realizados por parte de los usuarios, son los resultados a nivel de estudiante. En consecuencia, cada fila de la tabla de hechos corresponderá al resultado obtenido por parte de un estudiante por área curricular evaluada.

3.5.3. Elegir las dimensiones

Para la estructuración de las tablas dimensiones se tomaron en cuenta los requerimientos por parte de los usuarios, que se analizaron en etapas anteriores, estableciéndose 24 dimensiones en total, los cuales son:

- Dimensión área

Tabla 8: Descripción de la estructura de la dim_area

Campo	Tipo	Longitud	Descripción	Nulo	Llave
SK_AREA	int	1	Correlativo (Autoincrementable)	No	PK
ID_AREA	int	1	Identificador del área geográfica de la IE	No	
DESC_AREA	varchar	10	Área geográfica de la IE	No	

- Dimensión área curricular

Tabla 9: Descripción de la estructura de la dim_area_curricular

Campo	Tipo	Longitud	Descripción	Nulo	Llave
SK_AREA_CURRICULAR	int	2	Correlativo (Autoincrementable)	No	PK
ID_AREA_CURRICULAR	int	2	Identificador del área curricular evaluada	No	
ABREV_AREA_CURRICULAR	varchar	5	Abreviación del área curricular evaluada	No	
DESC_AREA_CURRICULAR	varchar	40	Área curricular evaluada	No	

- Dimensión característica (2 categorías)

Tabla 10: Descripción de la estructura de la dim_caracteristica2

Campo	Tipo	Longitud	Descripción	Nulo	Llave
SK_CARACTERISTICA2	int	1	Correlativo (Autoincrementable)	No	PK
ID_CARACTERISTICA2	int	1	Identificador de la característica de la IE (2 categorías)	No	
DESC_CARACTERISTICA2	varchar	40	Característica de la IE (2 categorías)	No	

- Dimensión característica (3 categorías)

Tabla 11: Descripción de la estructura de la dim_caracteristica3

Campo	Tipo	Longitud	Descripción	Nulo	Llave
SK_CARACTERISTICA3	int	1	Correlativo (Autoincrementable)	No	PK
ID_CARACTERISTICA3	int	1	Identificador de la característica de la IE (3 categorías)	No	
DESC_CARACTERISTICA3	varchar	40	Característica de la IE (3 categorías)	No	

- Dimensión Dirección Regional de Educación (DRE)

Tabla 12: Descripción de la estructura de la dim_dre

Campo	Tipo	Longitud	Descripción	Nulo	Llave
SK_DRE	int	2	Correlativo (Autoincrementable)	No	PK
ID_DRE	varchar	4	Identificador de la DRE (código DRE)	No	
DESC_DRE	varchar	30	Nombre de la DRE	No	

- Dimensión estudiante

Tabla 13: Descripción de la estructura de la dim_estudiante

Campo	Tipo	Longitud	Descripción	Nulo	Llave
SK_ESTUDIANTE	int	10	Correlativo (Autoincrementable)	No	PK
ID_ESTUDIANTE	varchar	19	Identificador del estudiante	No	
COD_ESTUDIANTE	varchar	2	Correlativo del estudiante en la sección	No	
APELLIDO_PATERO	varchar	40	Apellido paterno del estudiante	Sí	
APELLIDO_MATERNO	varchar	40	Apellido materno del estudiante	Sí	
NOMBRE_1	varchar	40	Primer nombre del estudiante	Sí	
NOMBRE_2	varchar	40	Segundo nombre del estudiante	Sí	
DNI	varchar	8	DNI del estudiante	Sí	
ID_SEXO	int	1	Identificador del sexo del estudiante	Sí	
ID LENGUA MATERNA	int	2	Identificador de la lengua materna del estudiante	Sí	

- Dimensión grado

Tabla 14: Descripción de la estructura de la dim_grado

Campo	Tipo	Longitud	Descripción	Nulo	Llave
SK_GRADO	int	1	Correlativo (Autoincrementable)	No	PK
ID_GRADO	int	1	Identificador del grado educativo evaluado	No	
ABREV_GRADO	varchar	3	Abreviación del grado educativo evaluado	No	
DESC_GRADO	varchar	10	Grado educativo evaluado	No	

- Dimensión incluir resultados

Tabla 15: Descripción de la estructura de la dim_incluir_resultados

Campo	Tipo	Longitud	Descripción	Nulo	Llave
SK_INCLUIR_RESULTADOS	int	1	Correlativo (Autoincrementable)	No	PK
ID_INCLUIR_RESULTADOS	int	1	Identificador de incluir resultados en los cálculos	No	
DESC_INCLUIR_RESULTADOS	varchar	20	Incluir resultados en los cálculos	No	

- Dimensión Institución Educativa (IE)

Tabla 16: Descripción de la estructura de la dim_institucion

Campo	Tipo	Longitud	Descripción	Nulo	Llave
SK_INSTITUCION	int	5	Correlativo (Autoincrementable)	No	PK
ID_INSTITUCION	varchar	8	Identificador de la IE	No	
COD_MODULAR	varchar	7	Código modular de la IE	No	
ANEXO	varchar	1	Anexo de la IE	No	
DESC_INSTITUCION	varchar	100	Nombre de la IE	No	
ID_UBICACION	varchar	6	Identificador de la ubicación de la IE	No	
ID_NIVEL	int	1	Identificador del nivel educativo de la IE	No	

- Dimensión lengua - Educación Intercultural Bilingüe (EIB)

Tabla 17: Descripción de la estructura de la dim_lengua_eib

Campo	Tipo	Longitud	Descripción	Nulo	Llave
SK LENGUA_EIB	int	2	Correlativo (Autoincrementable)	No	PK
ID LENGUA_EIB	int	2	Identificador de la lengua programada a evaluar	No	
DESC LENGUA_EIB	varchar	40	Lengua programada a evaluar	No	

- Dimensión lengua materna

Tabla 18: Descripción de la estructura de la dim_lengua_materna

Campo	Tipo	Longitud	Descripción	Nulo	Llave
SK LENGUA MATERNA	int	2	Correlativo (Autoincrementable)	No	PK
ID LENGUA MATERNA	int	2	Identificador de la lengua materna del estudiante	No	
DESC LENGUA MATERNA	varchar	40	Lengua materna del estudiante	No	

- Dimensión nivel del índice socioeconómico

Tabla 19: Descripción de la estructura de la dim_n_ise

Campo	Tipo	Longitud	Descripción	Nulo	Llave
SK_N_ISE	int	1	Correlativo (Autoincrementable)	No	PK
ID_N_ISE	int	1	Identificador del nivel del índice socioeconómico del estudiante	No	
DESC_N_ISE	varchar	10	Nivel del índice socioeconómico del estudiante	No	

- Dimensión nivel

Tabla 20: Descripción de la estructura de la dim_nivel

Campo	Tipo	Longitud	Descripción	Nulo	Llave
SK_NIVEL	int	1	Correlativo (Autoincrementable)	No	PK
ID_NIVEL	int	1	Identificador del nivel educativo de la IE	No	
DESC_NIVEL	varchar	20	Nivel educativo de la IE	No	

- Dimensión nivel de logro

Tabla 21: Descripción de la estructura de la dim_nivel_logro

Campo	Tipo	Longitud	Descripción	Nulo	Llave
SK_NIVEL_LOGRO	int	1	Correlativo (Autoincrementable)	No	PK
ID_NIVEL_LOGRO	int	1	Identificador del nivel de logro del estudiante	No	
DESC_NIVEL_LOGRO	varchar	20	Nivel de logro del estudiante	No	

- Dimensión periodo

Tabla 22: Descripción de la estructura de la dim_periodo

Campo	Tipo	Longitud	Descripción	Nulo	Llave
SK_PERIODO	int	2	Correlativo (Autoincrementable)	No	PK
ID_PERIODO	int	2	Identificador del año de la evaluación	No	
DESC_PERIODO	varchar	4	Año de la evaluación	No	
PERIODO_ANTERIOR	int	4	Año de la evaluación anterior	No	

- Dimensión sección

Tabla 23: Descripción de la estructura de la dim_seccion

Campo	Tipo	Longitud	Descripción	Nulo	Llave
SK_SECCION	int	3	Correlativo (Autoincrementable)	No	PK
ID_SECCION	varchar	2	Identificador de la sección evaluada	No	
DESC_SECCION	varchar	2	Sección evaluada	No	
ID_GRADO	int	1	Identificador del grado educativo evaluado	No	

- Dimensión sexo

Tabla 24: Descripción de la estructura de la dim_sexo

Campo	Tipo	Longitud	Descripción	Nulo	Llave
SK_SEXO	int	1	Correlativo (Autoincrementable)	No	PK
ID_SEXO	int	1	Identificador del sexo del estudiante	No	
DESC_SEXO	varchar	20	Sexo del estudiante	No	

- Dimensión tipo discapacidad - Registro de Necesidades Educativas Especiales (RNEE)

Tabla 25: Descripción de la estructura de la dim_tipo_discapacidad_rnee

Campo	Tipo	Longitud	Descripción	Nulo	Llave
SK_TIPO_DISCAPACIDAD_RNEE	int	2	Correlativo (Autoincrementable)	No	PK
ID_TIPO_DISCAPACIDAD_RNEE	int	2	Identificador del tipo de discapacidad R-NEE del estudiante	No	
DESC_TIPO_DISCAPACIDAD_RNEE	varchar	40	Tipo de discapacidad R-NEE del estudiante	No	

- Dimensión subtipo discapacidad - Registro de Necesidades Educativas Especiales (RNEE)

Tabla 26: Descripción de la estructura de la dim_subtipo_discapacidad_rnee

Campo	Tipo	Longitud	Descripción	Nulo	Llave
SK_SUBTIPO_DISCAPACIDAD_RNEE	int	2	Correlativo (Autoincrementable)	No	PK
ID_SUBTIPO_DISCAPACIDAD_RNEE	int	2	Identificador del subtipo de discapacidad RNEE del estudiante	No	
DESC_SUBTIPO_DISCAPACIDAD_RNEE	varchar	50	Subtipo de discapacidad R-NEE del estudiante	No	
ID_TIPO_DISCAPACIDAD_RNEE	int	2	Identificador del tipo de discapacidad R-NEE del estudiante	No	

- Dimensión tipo de evaluación

Tabla 27: Descripción de la estructura de la dim_tipo_evaluacion

Campo	Tipo	Longitud	Descripción	Nulo	Llave
SK_TIPO_EVALUACION	int	1	Correlativo (Autoincrementable)	No	PK
ID_TIPO_EVALUACION	int	1	Identificador del tipo de evaluación	No	
ABREV_TIPO_EVALUACION	varchar	4	Abreviación del tipo de evaluación	No	
DESC_TIPO_EVALUACION	varchar	100	Tipo de evaluación	No	

- Dimensión tipo de gestión (2 categorías)

Tabla 28: Descripción de la estructura de la dim_tipo_gestion2

Campo	Tipo	Longitud	Descripción	Nulo	Llave
SK_TIPO_GESTION2	int	1	Correlativo (Autoincrementable)	No	PK
ID_TIPO_GESTION2	int	1	Identificador del tipo de gestión de la IE (2 categorías)	No	
DESC_TIPO_GESTION2	varchar	20	Tipo de gestión de la IE (2 categorías)	No	

- Dimensión tipo de gestión (3 categorías)

Tabla 29: Descripción de la estructura de la dim_tipo_gestion3

Campo	Tipo	Longitud	Descripción	Nulo	Llave
SK_TIPO_GESTION3	int	1	Correlativo (Autoincrementable)	No	PK
ID_TIPO_GESTION3	int	1	Identificador del tipo de gestión de la IE (3 categorías)	No	
DESC_TIPO_GESTION3	varchar	30	Tipo de gestión de la IE (3 categorías)	No	

- Dimensión ubicación

Tabla 30: Descripción de la estructura de la dim_ubicacion

Campo	Tipo	Longitud	Descripción	Nulo	Llave
SK_UBICACION	int	4	Correlativo (Autoincrementable)	No	PK
ID_UBICACION	varchar	6	Identificador de la ubicación de la IE (código geográfico)	No	
DESC_DEPARTAMENTO	varchar	60	Departamento	No	
DESC_PROVINCIA	varchar	60	Provincia	No	
DESC_DISTRITO	varchar	60	Distrito	No	

- Dimensión Unidad de Gestión Educativa Local (UGEL)

Tabla 31: Descripción de la estructura de la dim_ugel

Campo	Tipo	Longitud	Descripción	Nulo	Llave
SK_UGEL	int	3	Correlativo (Autoincrementable)	No	PK
ID_UGEL	varchar	6	Identificador de la UGEL (código UGEL)	No	
DESC_UGEL	varchar	60	Nombre de la UGEL	No	
ID_DRE	varchar	4	Identificador de la DRE (código DRE)	No	

3.5.4. Identificar las tablas de hechos y medidas

Para la estructuración de la tabla de hechos y las medidas se tomaron en cuenta los requerimientos por parte de los usuarios, que se analizaron en etapas anteriores, estableciéndose 1 tabla de hechos, el cual es:

- Fact resultados estudiantes

Tabla 32: Descripción de la estructura de la fact_resultados_estudiantes

Campo	Tipo	Longitud	Descripción	Nulo	Llave
SK_PERIODO	int	2	PK dim_periodo	No	FK
SK_NIVEL	int	1	PK dim_nivel	No	FK
SK_GRADO	int	1	PK dim_grado	No	FK
SK_TIPO_EVALUACION	int	1	PK dim_tipo_evaluacion	No	FK
SK_DRE	int	2	PK dim_dre	No	FK
SK_UGEL	int	3	PK dim_ugel	No	FK
SK_UBICACION	int	4	PK dim_ubicacion	No	FK
SK_TIPO_GESTION3	int	1	PK dim_tipo_gestion3	Sí	FK
SK_TIPO_GESTION2	int	1	PK dim_tipo_gestion2	No	FK
SK_CARACTERISTICA3	int	1	PK dim_caracteristica3	Sí	FK
SK_CARACTERISTICA2	int	1	PK dim_caracteristica2	No	FK
SK_SECCION	int	3	PK dim_seccion	No	FK
SK_ESTUDIANTE	int	10	PK dim_estudiante	No	FK
SK_INSTITUCION	int	5	PK dim_institucion	No	FK
SK_SEXO	int	1	PK dimsexo	Sí	FK
SK LENGUA MATERNA	int	2	PK dim_lengua_materna	Sí	FK
SK TIPO DISCAPACIDAD RNEE	int	2	PK dim_tipo_discapacidad_rnee	Sí	FK
SK SUBTIPO DISCAPACIDAD RNEE	int	2	PK dim_subtipo_discapacidad_rnee	Sí	FK
SK INCLUIR RESULTADOS	int	1	PK dim_incluir_resultados	No	FK
SK_N_ISE	int	1	PK dim_n_ise	Sí	FK
SK AREA	int	1	PK dim_area	No	FK
SK LENGUA EIB	int	2	PK dim_lengua_eib	Sí	FK
SK AREA CURRICULAR	int	2	PK dim_area_curricular	No	FK
SK_NIVEL LOGRO	int	1	PK dim_nivel_logro	Sí	FK
ISE	decimal	(20,14)	Índice Socioeconómico	Sí	
MEDIDA	decimal	(20,14)	Medida Rash	Sí	
M500	decimal	(20,14)	Medida Rash 500	Sí	
ERROR	decimal	(20,14)	Error	Sí	
INFIT	decimal	(20,14)	Infit	Sí	
OUTFIT	decimal	(20,14)	Outfit	Sí	
PONDERADOR	decimal	(20,14)	Ponderador	Sí	
RESPUESTAS	varchar	150	Respuestas del estudiante	Sí	

Para la implementación del Datamart se optó por el modelo estrella, por los motivos expuestos en la sección anterior.

Para el Datamart de la Instancia Técnica, la tabla de hechos vendría a ser la tabla `fact_resultados_estudiantes`, rodeada por cada una de las 24 tablas de dimensiones previamente definidas. Las tablas de dimensiones tienen una columna que es la llave primaria (PK) el cual identifica de forma única a cada fila de la tabla. La tabla de hechos o tabla fact está conformada por las llaves primarias de las 24 tablas de dimensiones, que vendrían a ser llaves foráneas en la tabla de hechos. Adicionalmente la tabla de hechos contiene también las métricas o medidas y las claves de las respuestas de los estudiantes.

3.6. Diseño físico

La herramienta usada para el diseño y gestión de las bases de datos es el MySQL Workbench. En esta herramienta se crearon las tablas dimensiones y la tabla de hechos o tabla fact.

A continuación, se muestran las sentencias SQL generadas en la creación de cada una de las tablas:

3.6.1. Sentencias SQL para crear las dimensiones

- Dimensión área

```
CREATE TABLE `dim_area`  
(  
  `SK_AREA` int(1) NOT NULL AUTO_INCREMENT,
```

```

`ID_AREA` int(1) NOT NULL,
`DESC_AREA` varchar(10) NOT NULL DEFAULT "",
PRIMARY KEY (`SK_AREA`)
)
ENGINE=InnoDB AUTO_INCREMENT=1 DEFAULT CHARSET=utf8mb4

```

- Dimensión área curricular

```

CREATE TABLE `dim_area_curricular`
(
`SK_AREA_CURRICULAR` int(2) NOT NULL AUTO_INCREMENT,
`ID_AREA_CURRICULAR` int(2) NOT NULL,
`ABREV_AREA_CURRICULAR` varchar(5) NOT NULL DEFAULT "",
`DESC_AREA_CURRICULAR` varchar(40) NOT NULL DEFAULT "",
PRIMARY KEY (`SK_AREA_CURRICULAR`)
)
ENGINE=InnoDB AUTO_INCREMENT=1 DEFAULT CHARSET=utf8mb4

```

- Dimensión característica (2 categorías)

```

CREATE TABLE `dim_caracteristica2`
(
`SK_CARACTERISTICA2` int(1) NOT NULL AUTO_INCREMENT,
`ID_CARACTERISTICA2` int(1) NOT NULL,
`DESC_CARACTERISTICA2` varchar(40) NOT NULL DEFAULT "",
PRIMARY KEY (`SK_CARACTERISTICA2`)
)
ENGINE=InnoDB AUTO_INCREMENT=1 DEFAULT CHARSET=utf8mb4

```

- Dimensión característica (3 categorías)

```
CREATE TABLE `dim_caracteristica3`
(
  `SK_CHARACTERISTICA3` int(1) NOT NULL AUTO_INCREMENT,
  `ID_CHARACTERISTICA3` int(1) NOT NULL,
  `DESC_CHARACTERISTICA3` varchar(40) NOT NULL DEFAULT "",
  PRIMARY KEY (`SK_CHARACTERISTICA3`)
)
ENGINE=InnoDB AUTO_INCREMENT=1 DEFAULT CHARSET=utf8mb4
```

- Dimensión Dirección Regional de Educación (DRE)

```
CREATE TABLE `dim_dre`
(
  `SK_DRE` int(2) NOT NULL AUTO_INCREMENT,
  `ID_DRE` varchar(4) NOT NULL DEFAULT "",
  `DESC_DRE` varchar(30) NOT NULL DEFAULT "",
  PRIMARY KEY (`SK_DRE`)
)
ENGINE=InnoDB AUTO_INCREMENT=1 DEFAULT CHARSET=utf8mb4
```

- Dimensión estudiante

```
CREATE TABLE `dim_estudiante`
(
  `SK_ESTUDIANTE` int(10) NOT NULL AUTO_INCREMENT,
  `ID_ESTUDIANTE` varchar(19) NOT NULL DEFAULT "",
  `COD_ESTUDIANTE` varchar(2) NOT NULL DEFAULT "",
  `APELLIDO_PATERNO` varchar(40) DEFAULT NULL,
```

```

`APELLIDO_MATERNO` varchar(40) DEFAULT NULL,
`NOMBRE_1` varchar(40) DEFAULT NULL,
`NOMBRE_2` varchar(40) DEFAULT NULL,
`DNI` varchar(8) DEFAULT NULL,
`ID_SEXO` int(1) DEFAULT NULL,
`ID LENGUA MATERNA` int(2) DEFAULT NULL,
PRIMARY KEY (`SK_ESTUDIANTE`)
)
ENGINE=InnoDB AUTO_INCREMENT=1 DEFAULT CHARSET=utf8mb4

```

- Dimensión grado

```

CREATE TABLE `dim_grado`
(
`SK_GRADO` int(1) NOT NULL AUTO_INCREMENT,
`ID_GRADO` int(1) NOT NULL,
`ABREV_GRADO` varchar(3) NOT NULL DEFAULT "",
`DESC_GRADO` varchar(10) NOT NULL DEFAULT "",
PRIMARY KEY (`SK_GRADO`)
)
ENGINE=InnoDB AUTO_INCREMENT=1 DEFAULT CHARSET=utf8mb4

```

- Dimensión incluir resultados

```

CREATE TABLE `dim_incluir_resultados`
(
`SK_INCLUIR_RESULTADOS` int(1) NOT NULL AUTO_INCREMENT,
`ID_INCLUIR_RESULTADOS` int(1) NOT NULL,
`DESC_INCLUIR_RESULTADOS` varchar(20) NOT NULL DEFAULT "",
PRIMARY KEY (`SK_INCLUIR_RESULTADOS`)
)

```

```
)  
ENGINE=InnoDB AUTO_INCREMENT=1 DEFAULT CHARSET=utf8mb4
```

- Dimensión Institución Educativa (IE)

```
CREATE TABLE `dim_institucion`  
(  
  `SK_INSTITUCION` int(5) NOT NULL AUTO_INCREMENT,  
  `ID_INSTITUCION` varchar(8) NOT NULL DEFAULT "",  
  `COD_MODULAR` varchar(7) NOT NULL DEFAULT "",  
  `ANEXO` varchar(1) NOT NULL DEFAULT "",  
  `DESC_INSTITUCION` varchar(100) NOT NULL DEFAULT "",  
  `ID_UBICACION` varchar(6) NOT NULL DEFAULT "",  
  `ID_NIVEL` int(1) NOT NULL,  
  PRIMARY KEY (`SK_INSTITUCION`)  
)  
ENGINE=InnoDB AUTO_INCREMENT=1 DEFAULT CHARSET=utf8mb4
```

- Dimensión lengua - Educación Intercultural Bilingüe (EIB)

```
CREATE TABLE `dim_lengua_eib`  
(  
  `SK LENGUA_EIB` int(2) NOT NULL AUTO_INCREMENT,  
  `ID LENGUA_EIB` int(2) NOT NULL,  
  `DESC LENGUA_EIB` varchar(40) NOT NULL DEFAULT "",  
  PRIMARY KEY (`SK LENGUA_EIB`)  
)  
ENGINE=InnoDB AUTO_INCREMENT=1 DEFAULT CHARSET=utf8mb4
```

- Dimensión lengua materna

```
CREATE TABLE `dim_lengua_materna`  
(  
  `SK LENGUA MATERNA` int(2) NOT NULL AUTO_INCREMENT,  
  `ID LENGUA MATERNA` int(2) NOT NULL,  
  `DESC LENGUA MATERNA` varchar(40) NOT NULL DEFAULT "",  
  PRIMARY KEY (`SK LENGUA MATERNA`)  
)  
ENGINE=InnoDB AUTO_INCREMENT=1 DEFAULT CHARSET=utf8mb4
```

- Dimensión nivel del índice socioeconómico

```
CREATE TABLE `dim_n_ise`  
(  
  `SK_N_ISE` int(1) NOT NULL AUTO_INCREMENT,  
  `ID_N_ISE` int(1) NOT NULL,  
  `DESC_N_ISE` varchar(10) NOT NULL DEFAULT "",  
  PRIMARY KEY (`SK_N_ISE`)  
)  
ENGINE=InnoDB AUTO_INCREMENT=1 DEFAULT CHARSET=utf8mb4
```

- Dimensión nivel

```
CREATE TABLE `dim_nivel`  
(  
  `SK_NIVEL` int(1) NOT NULL AUTO_INCREMENT,  
  `ID_NIVEL` int(1) NOT NULL,  
  `DESC_NIVEL` varchar(20) NOT NULL DEFAULT "",  
  PRIMARY KEY (`SK_NIVEL`)
```



```
)  
ENGINE=InnoDB AUTO_INCREMENT=1 DEFAULT CHARSET=utf8mb4
```

- Dimensión nivel de logro

```
CREATE TABLE `dim_nivel_logro`  
(  
  `SK_NIVEL_LOGRO` int(1) NOT NULL AUTO_INCREMENT,  
  `ID_NIVEL_LOGRO` int(1) NOT NULL,  
  `DESC_NIVEL_LOGRO` varchar(20) NOT NULL DEFAULT "",  
  PRIMARY KEY (`SK_NIVEL_LOGRO`)  
)  
ENGINE=InnoDB AUTO_INCREMENT=1 DEFAULT CHARSET=utf8mb4
```

- Dimensión periodo

```
CREATE TABLE `dim_periodo`  
(  
  `SK_PERIODO` int(2) NOT NULL AUTO_INCREMENT,  
  `ID_PERIODO` int(2) NOT NULL,  
  `DESC_PERIODO` varchar(4) NOT NULL DEFAULT "",  
  `PERIODO_ANTERIOR` int(4) NOT NULL,  
  PRIMARY KEY (`SK_PERIODO`)  
)  
ENGINE=InnoDB AUTO_INCREMENT=1 DEFAULT CHARSET=utf8mb4
```

- Dimensión sección

```
CREATE TABLE `dim_seccion`
(
  `SK_SECCION` int(3) NOT NULL AUTO_INCREMENT,
  `ID_SECCION` varchar(2) NOT NULL DEFAULT "",
  `DESC_SECCION` varchar(2) NOT NULL DEFAULT "",
  `ID_GRADO` int(1) NOT NULL,
  PRIMARY KEY (`SK_SECCION`)
)
ENGINE=InnoDB AUTO_INCREMENT=1 DEFAULT CHARSET=utf8mb4
```

- Dimensión sexo

```
CREATE TABLE `dim_sexo`
(
  `SK_SEXO` int(1) NOT NULL AUTO_INCREMENT,
  `ID_SEXO` int(1) NOT NULL,
  `DESC_SEXO` varchar(20) NOT NULL DEFAULT "",
  PRIMARY KEY (`SK_SEXO`)
)
ENGINE=InnoDB AUTO_INCREMENT=1 DEFAULT CHARSET=utf8mb4
```

- Dimensión tipo discapacidad - Registro de Necesidades Educativas Especiales (RNEE)

```
CREATE TABLE `dim_tipo_discapacidad_ree`
(
  `SK_TIPO_DISCAPACIDAD_RNEE` int(2) NOT NULL
  AUTO_INCREMENT,
```

```

`ID_TIPO_DISCAPACIDAD_RNEE` int(2) NOT NULL,
`DESC_TIPO_DISCAPACIDAD_RNEE` varchar(40) NOT NULL
DEFAULT ",
PRIMARY KEY (`SK_TIPO_DISCAPACIDAD_RNEE`)
)
ENGINE=InnoDB AUTO_INCREMENT=1 DEFAULT CHARSET=utf8mb4

```

- Dimensión subtipo discapacidad - Registro de Necesidades Educativas Especiales (RNEE)

```

CREATE TABLE `dim_subtipo_discapacidad_rnee`
(
`SK_SUBTIPO_DISCAPACIDAD_RNEE` int(2) NOT NULL
AUTO_INCREMENT,
`ID_SUBTIPO_DISCAPACIDAD_RNEE` int(2) NOT NULL,
`DESC_SUBTIPO_DISCAPACIDAD_RNEE` varchar(50) NOT NULL
DEFAULT ",
`ID_TIPO_DISCAPACIDAD_RNEE` int(2) NOT NULL,
PRIMARY KEY (`SK_SUBTIPO_DISCAPACIDAD_RNEE`)
)
ENGINE=InnoDB AUTO_INCREMENT=1 DEFAULT CHARSET=utf8mb4

```

- Dimensión tipo de evaluación

```

CREATE TABLE `dim_tipo_evaluacion`
(
`SK_TIPO_EVALUACION` int(1) NOT NULL AUTO_INCREMENT,
`ID_TIPO_EVALUACION` int(1) NOT NULL,
`ABREV_TIPO_EVALUACION` varchar(4) NOT NULL DEFAULT ",
`DESC_TIPO_EVALUACION` varchar(100) NOT NULL DEFAULT ",

```

```

PRIMARY KEY (`SK_TIPO_EVALUACION`)
)
ENGINE=InnoDB AUTO_INCREMENT=1 DEFAULT CHARSET=utf8mb4

```

- Dimensión tipo de gestión (2 categorías)

```

CREATE TABLE `dim_tipo_gestion2`
(
`SK_TIPO_GESTION2` int(1) NOT NULL AUTO_INCREMENT,
`ID_TIPO_GESTION2` int(1) NOT NULL,
`DESC_TIPO_GESTION2` varchar(20) NOT NULL DEFAULT "",
PRIMARY KEY (`SK_TIPO_GESTION2`)
)
ENGINE=InnoDB AUTO_INCREMENT=1 DEFAULT CHARSET=utf8mb4

```

- Dimensión tipo de gestión (3 categorías)

```

CREATE TABLE `dim_tipo_gestion3`
(
`SK_TIPO_GESTION3` int(1) NOT NULL AUTO_INCREMENT,
`ID_TIPO_GESTION3` int(1) NOT NULL,
`DESC_TIPO_GESTION3` varchar(30) NOT NULL DEFAULT "",
PRIMARY KEY (`SK_TIPO_GESTION3`)
)
ENGINE=InnoDB AUTO_INCREMENT=1 DEFAULT CHARSET=utf8mb4

```

- Dimensión ubicación

```
CREATE TABLE `dim_ubicacion`  
(  
  `SK_UBICACION` int(4) NOT NULL AUTO_INCREMENT,  
  `ID_UBICACION` varchar(6) NOT NULL DEFAULT "",  
  `DESC_DEPARTAMENTO` varchar(60) NOT NULL DEFAULT "",  
  `DESC_PROVINCIA` varchar(60) NOT NULL DEFAULT "",  
  `DESC_DISTRITO` varchar(60) NOT NULL DEFAULT "",  
  PRIMARY KEY (`SK_UBICACION`)  
)  
ENGINE=InnoDB AUTO_INCREMENT=1 DEFAULT CHARSET=utf8mb4
```

- Dimensión Unidad de Gestión Educativa Local (UGEL)

```
CREATE TABLE `dim_ugel`  
(  
  `SK_UGEL` int(3) NOT NULL AUTO_INCREMENT,  
  `ID_UGEL` varchar(6) NOT NULL DEFAULT "",  
  `DESC_UGEL` varchar(60) NOT NULL DEFAULT "",  
  `ID_DRE` varchar(4) NOT NULL DEFAULT "",  
  PRIMARY KEY (`SK_UGEL`)  
)  
ENGINE=InnoDB AUTO_INCREMENT=1 DEFAULT CHARSET=utf8mb4
```

3.6.2. Sentencias SQL para crear la tabla de hechos y medidas

- Fact resultados estudiantes

```
CREATE TABLE `fact_resultados_estudiantes`  
(  
  `SK_PERIODO` int(2) NOT NULL,  
  `SK_NIVEL` int(1) NOT NULL,  
  `SK_GRADO` int(1) NOT NULL,  
  `SK_TIPO_EVALUACION` int(1) NOT NULL,  
  `SK_DRE` int(2) NOT NULL,  
  `SK_UGEL` int(3) NOT NULL,  
  `SK_UBICACION` int(4) NOT NULL,  
  `SK_TIPO_GESTION3` int(1) DEFAULT NULL,  
  `SK_TIPO_GESTION2` int(1) NOT NULL,  
  `SK_CARACTERISTICA3` int(1) DEFAULT NULL,  
  `SK_CARACTERISTICA2` int(1) NOT NULL,  
  `SK_SECCION` int(3) NOT NULL,  
  `SK_ESTUDIANTE` int(10) NOT NULL,  
  `SK_INSTITUCION` int(5) NOT NULL,  
  `SK_SEXO` int(1) DEFAULT NULL,  
  `SK LENGUA MATERNA` int(2) DEFAULT NULL,  
  `SK_TIPO_DISCAPACIDAD_RNEE` int(2) DEFAULT NULL,  
  `SK_SUBTIPO_DISCAPACIDAD_RNEE` int(2) DEFAULT NULL,  
  `SK_INCLUIR_RESULTADOS` int(1) NOT NULL,  
  `SK_N_ISE` int(1) DEFAULT NULL,  
  `SK_AREA` int(1) NOT NULL,  
  `SK LENGUA EIB` int(2) DEFAULT NULL,  
  `SK_AREA_CURRICULAR` int(2) NOT NULL,  
  `SK_NIVEL_LOGRO` int(1) DEFAULT NULL,  
  `ISE` decimal(20,14) DEFAULT NULL,  
  `MEDIDA` decimal(20,14) DEFAULT NULL,  
)
```

```

`M500` decimal(20,14) DEFAULT NULL,
`ERROR` decimal(20,14) DEFAULT NULL,
`INFIT` decimal(20,14) DEFAULT NULL,
`OUTFIT` decimal(20,14) DEFAULT NULL,
`PONDERADOR` decimal(20,14) DEFAULT NULL,
`RESPUESTAS` varchar(150) DEFAULT "",
KEY          `FK_FACT_RESULTADOS_ESTUDIANTES_DIM_AREA`
(`SK_AREA`),
KEY
`FK_FACT_RESULTADOS_ESTUDIANTES_DIM_AREA_CURRICULAR`
(`SK_AREA_CURRICULAR`),
KEY
`FK_FACT_RESULTADOS_ESTUDIANTES_DIM_CARACTERISTICA2`
(`SK_CARACTERISTICA2`),
KEY
`FK_FACT_RESULTADOS_ESTUDIANTES_DIM_CARACTERISTICA3`
(`SK_CARACTERISTICA3`),
KEY          `FK_FACT_RESULTADOS_ESTUDIANTES_DIM_DRE`
(`SK_DRE`),
KEY `FK_FACT_RESULTADOS_ESTUDIANTES_DIM_ESTUDIANTE`
(`SK_ESTUDIANTE`),
KEY          `FK_FACT_RESULTADOS_ESTUDIANTES_DIM_GRADO`
(`SK_GRADO`),
KEY
`FK_FACT_RESULTADOS_ESTUDIANTES_DIM_INCLUIR_RESULTADOS`
(`SK_INCLUIR_RESULTADOS`),
KEY `FK_FACT_RESULTADOS_ESTUDIANTES_DIM_INSTITUCION`
(`SK_INSTITUCION`),
KEY `FK_FACT_RESULTADOS_ESTUDIANTES_DIM LENGUA_EIB`
(`SK LENGUA_EIB`),

```

KEY
 `FK_FACT_RESULTADOS_ESTUDIANTES_DIM LENGUA MATERNA`
 (`SK_LENGUA_MATERNA`),
 KEY `FK_FACT_RESULTADOS_ESTUDIANTES_DIM NIVEL`
 (`SK_NIVEL`),
 KEY
 `FK_FACT_RESULTADOS_ESTUDIANTES_DIM NIVEL LOGRO`
 (`SK_NIVEL_LOGRO`),
 KEY `FK_FACT_RESULTADOS_ESTUDIANTES_DIM N_ISE`
 (`SK_N_ISE`),
 KEY `FK_FACT_RESULTADOS_ESTUDIANTES_DIM PERIODO`
 (`SK_PERIODO`),
 KEY `FK_FACT_RESULTADOS_ESTUDIANTES_DIM SECCION`
 (`SK_SECCION`),
 KEY `FK_FACT_RESULTADOS_ESTUDIANTES_DIM SEXO`
 (`SK_SEXO`),
 KEY
 `FK_FACT_RESULTADOS_ESTUDIANTES_DIM SUBTIPO DISCAPAC
 IDAD_RNEE` (`SK_SUBTIPO_DISCAPACIDAD_RNEE`),
 KEY
 `FK_FACT_RESULTADOS_ESTUDIANTES_DIM TIPO DISCAPACIDA
 D_RNEE` (`SK_TIPO_DISCAPACIDAD_RNEE`),
 KEY
 `FK_FACT_RESULTADOS_ESTUDIANTES_DIM TIPO EVALUACION`
 (`SK_TIPO_EVALUACION`),
 KEY
 `FK_FACT_RESULTADOS_ESTUDIANTES_DIM TIPO GESTION2`
 (`SK_TIPO_GESTION2`),
 KEY
 `FK_FACT_RESULTADOS_ESTUDIANTES_DIM TIPO GESTION3`
 (`SK_TIPO_GESTION3`),


```

KEY `FK_FACT_RESULTADOS_ESTUDIANTES_DIM_UBICACION`
(`SK_UBICACION`),
KEY `FK_FACT_RESULTADOS_ESTUDIANTES_DIM_UGEL`
(`SK_UGEL`),
CONSTRAINT
`FK_FACT_RESULTADOS_ESTUDIANTES_DIM_AREA` FOREIGN KEY
(`SK_AREA`) REFERENCES `dim_area` (`SK_AREA`) ON DELETE NO
ACTION ON UPDATE NO ACTION,
CONSTRAINT
`FK_FACT_RESULTADOS_ESTUDIANTES_DIM_AREA_CURRICULAR`
FOREIGN KEY (`SK_AREA_CURRICULAR`) REFERENCES
`dim_area_curricular` (`SK_AREA_CURRICULAR`) ON DELETE NO
ACTION ON UPDATE NO ACTION,
CONSTRAINT
`FK_FACT_RESULTADOS_ESTUDIANTES_DIM_CARACTERISTICA2`
FOREIGN KEY (`SK_CARACTERISTICA2`) REFERENCES
`dim_caracteristica2` (`SK_CARACTERISTICA2`) ON DELETE NO
ACTION ON UPDATE NO ACTION,
CONSTRAINT
`FK_FACT_RESULTADOS_ESTUDIANTES_DIM_CARACTERISTICA3`
FOREIGN KEY (`SK_CARACTERISTICA3`) REFERENCES
`dim_caracteristica3` (`SK_CARACTERISTICA3`) ON DELETE NO
ACTION ON UPDATE NO ACTION,
CONSTRAINT `FK_FACT_RESULTADOS_ESTUDIANTES_DIM_DRE`
FOREIGN KEY (`SK_DRE`) REFERENCES `dim_dre` (`SK_DRE`) ON
DELETE NO ACTION ON UPDATE NO ACTION,
CONSTRAINT
`FK_FACT_RESULTADOS_ESTUDIANTES_DIM_ESTUDIANTE`
FOREIGN KEY (`SK_ESTUDIANTE`) REFERENCES `dim_estudiante`
(`SK_ESTUDIANTE`) ON DELETE NO ACTION ON UPDATE NO
ACTION,

```

```

CONSTRAINT
`FK_FACT_RESULTADOS_ESTUDIANTES_DIM_GRADO` FOREIGN
KEY (`SK_GRADO`) REFERENCES `dim_grado` (`SK_GRADO`) ON
DELETE NO ACTION ON UPDATE NO ACTION,
CONSTRAINT
`FK_FACT_RESULTADOS_ESTUDIANTES_DIM_INCLUIR_RESULTAD
OS` FOREIGN KEY (`SK_INCLUIR_RESULTADOS`) REFERENCES
`dim_incluir_resultados` (`SK_INCLUIR_RESULTADOS`) ON DELETE NO
ACTION ON UPDATE NO ACTION,
CONSTRAINT
`FK_FACT_RESULTADOS_ESTUDIANTES_DIM_INSTITUCION`
FOREIGN KEY (`SK_INSTITUCION`) REFERENCES `dim_institucion`
(`SK_INSTITUCION`) ON DELETE NO ACTION ON UPDATE NO
ACTION,
CONSTRAINT
`FK_FACT_RESULTADOS_ESTUDIANTES_DIM LENGUA_EIB`
FOREIGN KEY (`SK LENGUA_EIB`) REFERENCES `dim_lengua_eib`
(`SK LENGUA_EIB`) ON DELETE NO ACTION ON UPDATE NO
ACTION,
CONSTRAINT
`FK_FACT_RESULTADOS_ESTUDIANTES_DIM LENGUA MATERNA`
FOREIGN KEY (`SK LENGUA MATERNA`) REFERENCES
`dim_lengua_materna` (`SK LENGUA MATERNA`) ON DELETE NO
ACTION ON UPDATE NO ACTION,
CONSTRAINT
`FK_FACT_RESULTADOS_ESTUDIANTES_DIM_NIVEL` FOREIGN
KEY (`SK_NIVEL`) REFERENCES `dim_nivel` (`SK_NIVEL`) ON DELETE
NO ACTION ON UPDATE NO ACTION,
CONSTRAINT
`FK_FACT_RESULTADOS_ESTUDIANTES_DIM_NIVEL_LOGRO`
FOREIGN KEY (`SK_NIVEL_LOGRO`) REFERENCES `dim_nivel_logro`

```

```

(`SK_NIVEL_LOGRO`) ON DELETE NO ACTION ON UPDATE NO
ACTION,
CONSTRAINT
`FK_FACT_RESULTADOS_ESTUDIANTES_DIM_N_ISE` FOREIGN KEY
(`SK_N_ISE`) REFERENCES `dim_n_ise` (`SK_N_ISE`) ON DELETE NO
ACTION ON UPDATE NO ACTION,
CONSTRAINT
`FK_FACT_RESULTADOS_ESTUDIANTES_DIM_PERIODO` FOREIGN
KEY (`SK_PERIODO`) REFERENCES `dim_periodo` (`SK_PERIODO`) ON
DELETE NO ACTION ON UPDATE NO ACTION,
CONSTRAINT
`FK_FACT_RESULTADOS_ESTUDIANTES_DIM_SECCION` FOREIGN
KEY (`SK_SECCION`) REFERENCES `dim_seccion` (`SK_SECCION`) ON
DELETE NO ACTION ON UPDATE NO ACTION,
CONSTRAINT
`FK_FACT_RESULTADOS_ESTUDIANTES_DIM_SEXO` FOREIGN KEY
(`SK_SEXO`) REFERENCES `dim_sexo` (`SK_SEXO`) ON DELETE NO
ACTION ON UPDATE NO ACTION,
CONSTRAINT
`FK_FACT_RESULTADOS_ESTUDIANTES_DIM_SUBTIPO_DISCAPAC
IDAD_RNEE` FOREIGN KEY (`SK_SUBTIPO_DISCAPACIDAD_RNEE`)
REFERENCES
`dim_subtipo_discapacidad_ree`
(`SK_SUBTIPO_DISCAPACIDAD_RNEE`) ON DELETE NO ACTION ON
UPDATE NO ACTION,
CONSTRAINT
`FK_FACT_RESULTADOS_ESTUDIANTES_DIM_TIPO_DISCAPACIDA
D_RNEE` FOREIGN KEY (`SK_TIPO_DISCAPACIDAD_RNEE`)
REFERENCES
`dim_tipo_discapacidad_ree`
(`SK_TIPO_DISCAPACIDAD_RNEE`) ON DELETE NO ACTION ON
UPDATE NO ACTION,
CONSTRAINT
`FK_FACT_RESULTADOS_ESTUDIANTES_DIM_TIPO_EVALUACION`

```

```

FOREIGN KEY (`SK_TIPO_EVALUACION`) REFERENCES
`dim_tipo_evaluacion` (`SK_TIPO_EVALUACION`) ON DELETE NO
ACTION ON UPDATE NO ACTION,
CONSTRAINT
`FK_FACT_RESULTADOS_ESTUDIANTES_DIM_TIPO_GESTION2`
FOREIGN KEY (`SK_TIPO_GESTION2`) REFERENCES
`dim_tipo_gestion2` (`SK_TIPO_GESTION2`) ON DELETE NO ACTION ON
UPDATE NO ACTION,
CONSTRAINT
`FK_FACT_RESULTADOS_ESTUDIANTES_DIM_TIPO_GESTION3`
FOREIGN KEY (`SK_TIPO_GESTION3`) REFERENCES
`dim_tipo_gestion3` (`SK_TIPO_GESTION3`) ON DELETE NO ACTION ON
UPDATE NO ACTION,
CONSTRAINT
`FK_FACT_RESULTADOS_ESTUDIANTES_DIM_UBICACION`
FOREIGN KEY (`SK_UBICACION`) REFERENCES `dim_ubicacion`
(`SK_UBICACION`) ON DELETE NO ACTION ON UPDATE NO ACTION,
CONSTRAINT
`FK_FACT_RESULTADOS_ESTUDIANTES_DIM_UGEL` FOREIGN KEY
(`SK_UGEL`) REFERENCES `dim_ugel` (`SK_UGEL`) ON DELETE NO
ACTION ON UPDATE NO ACTION
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4

```



Figura 6. Diseño físico del Datamart del área de Análisis de la Instancia Técnica

3.7. Diseño e implementación del subsistema de ETL

La herramienta que se empleó para la extracción, transformación y carga de las bases de datos es el Pentaho Data Integration (PDI) - Kettle. En esta herramienta se realizaron todas las transformaciones requeridas para el correcto ingreso de las bases de datos a las tablas de dimensiones y la tabla de hechos o tabla fact, que fueron creadas previamente en el MySQL Workbench.

3.7.1. ETL para las tablas dimensiones

Pasos realizados:

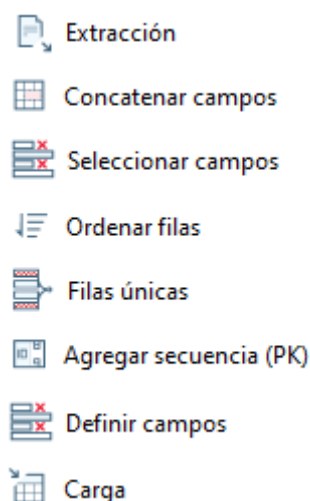


Figura 7. Pasos realizados en el ETL de las tablas de dimensiones

- 1) **Extracción:** Se extrajo la base de datos de un archivo en formato CSV.
- 2) **Concatenar campos:** Se concatenaron los campos para crear el ID necesario.
- 3) **Seleccionar campos:** Se seleccionaron solo los campos que conformarían la tabla dimensión.
- 4) **Ordenar filas:** Se ordenaron las filas por ID.
- 5) **Filas únicas:** Se eliminaron las filas duplicadas por ID.

- 6) **Agregar secuencia (PK):** Se agregó una columna de números correlativos autoincrementables, dando así origen a la llave primaria de la tabla dimensión.
- 7) **Definir campos:** Se ordenaron y definieron los campos de la tabla dimensión.
- 8) **Carga:** Previa transformación, se cargó la base de datos a una tabla ya creada en MySQL Workbench.

A continuación, se muestra el proceso de ETL que se realizó para poblar cada una de las tablas dimensiones:

- Dimensión área



Figura 8. Pasos realizados en el ETL de la tabla dim_area

- Dimensión área curricular



Figura 9. Pasos realizados en el ETL de la tabla dim_area_curricular

- Dimensión característica (2 categorías)



Figura 10. Pasos realizados en el ETL de la tabla dim_caracteristica2

- Dimensión característica (3 categorías)



Figura 11. Pasos realizados en el ETL de la tabla dim_caracteristica3

- Dimensión Dirección Regional de Educación (DRE)



Figura 12. Pasos realizados en el ETL de la tabla dim_dre

- Dimensión estudiante



Figura 13. Pasos realizados en el ETL de la tabla dim_estudiante

- Dimensión grado



Figura 14. Pasos realizados en el ETL de la tabla dim_grado

- Dimensión incluir resultados



Figura 15. Pasos realizados en el ETL de la tabla dim_incluir_resultados

- Dimensión Institución Educativa (IE)

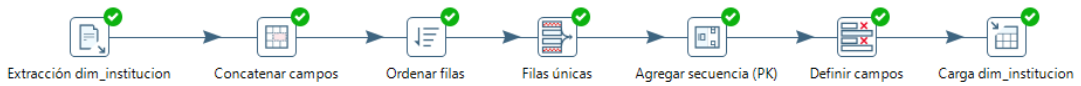


Figura 16. Pasos realizados en el ETL de la tabla dim_institucion

- Dimensión lengua - Educación Intercultural Bilingüe (EIB)



Figura 17. Pasos realizados en el ETL de la tabla dim_lengua_eib

- Dimensión lengua materna



Figura 18. Pasos realizados en el ETL de la tabla dim_lengua_materna

- Dimensión nivel del índice socioeconómico



Figura 19. Pasos realizados en el ETL de la tabla dim_n_ise

- Dimensión nivel



Figura 20. Pasos realizados en el ETL de la tabla dim_nivel

- Dimensión nivel de logro



Figura 21. Pasos realizados en el ETL de la tabla dim_nivel_logro

- Dimensión periodo



Figura 22. Pasos realizados en el ETL de la tabla dim_periodo

- Dimensión sección



Figura 23. Pasos realizados en el ETL de la tabla dim_seccion

- Dimensión sexo



Figura 24. Pasos realizados en el ETL de la tabla dim_sexo

- Dimensión tipo discapacidad - Registro de Necesidades Educativas Especiales (RNEE)



Figura 25. Pasos realizados en el ETL de la tabla dim_tipo_discapacidad_rnee

- Dimensión subtipo discapacidad - Registro de Necesidades Educativas Especiales (RNEE)



Figura 26. Pasos realizados en el ETL de la tabla dim_subtipo_discapacidad_rnee

- Dimensión tipo de evaluación



Figura 27. Pasos realizados en el ETL de la tabla dim_tipo_evaluacion

- Dimensión tipo de gestión (2 categorías)



Figura 28. Pasos realizados en el ETL de la tabla dim_tipo_gestion2

- Dimensión tipo de gestión (3 categorías)



Figura 29. Pasos realizados en el ETL de la tabla dim_tipo_gestion3

- Dimensión ubicación



Figura 30. Pasos realizados en el ETL de la tabla dim_ubicacion

- Dimensión Unidad de Gestión Educativa Local (UGEL)



Figura 31. Pasos realizados en el ETL de la tabla dim_ugel

3.7.2. ETL para la tabla de hechos o tabla fact

Pasos realizados:

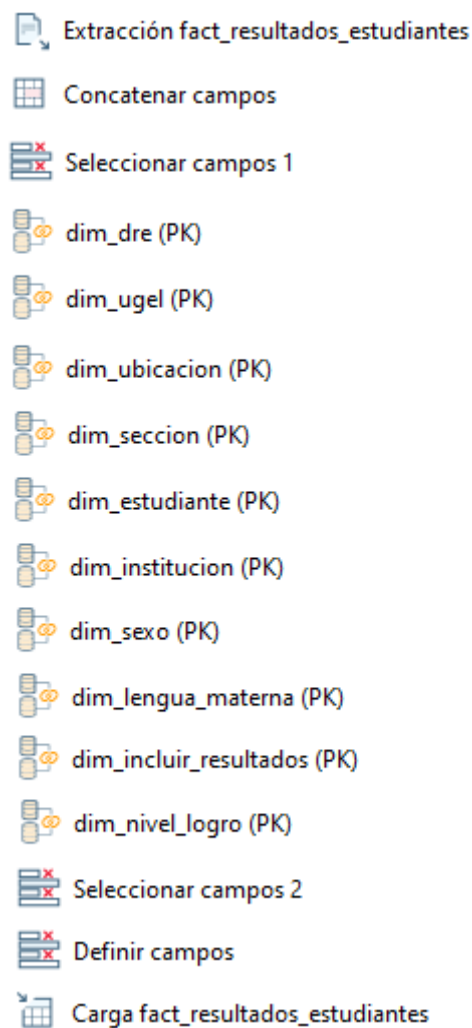


Figura 32. Pasos realizados en el ETL de la tabla de hechos o tabla fact

- 1) **Extracción:** Se extrajo la base de datos de un archivo en formato CSV.
- 2) **Concatenar campos:** Se concatenaron los campos para crear los ID necesarios.
- 3) **Seleccionar campos 1:** Se seleccionaron solo los campos que servirían para armar la tabla fact.
- 4) **dim_dre (PK), dim_ugel (PK), dim_ubicacion (PK), dim_seccion (PK), dim_estudiante (PK), dim_institucion (PK), dimsexo (PK), dim_lengua_materna (PK), dim_incluir_resultados (PK), dim_nivel_logro (PK):** Se agregó a la tabla fact las llaves primarias de las dimensiones que requerían dicha transformación, por medio de su ID.
- 5) **Seleccionar campos 2:** Una vez obtenido todos los PK de las tablas dimensiones, se eliminaron los ID que sirvieron para su armado.
- 6) **Definir campos:** Se ordenaron y definieron los campos de la tabla fact.
- 7) **Carga:** Previa transformación, se cargó la base de datos a una tabla ya creada en MySQL Workbench.

A continuación, se muestra el proceso de ETL que se realizó para poblar la tabla de hechos o tabla fact:

- Fact resultados estudiantes

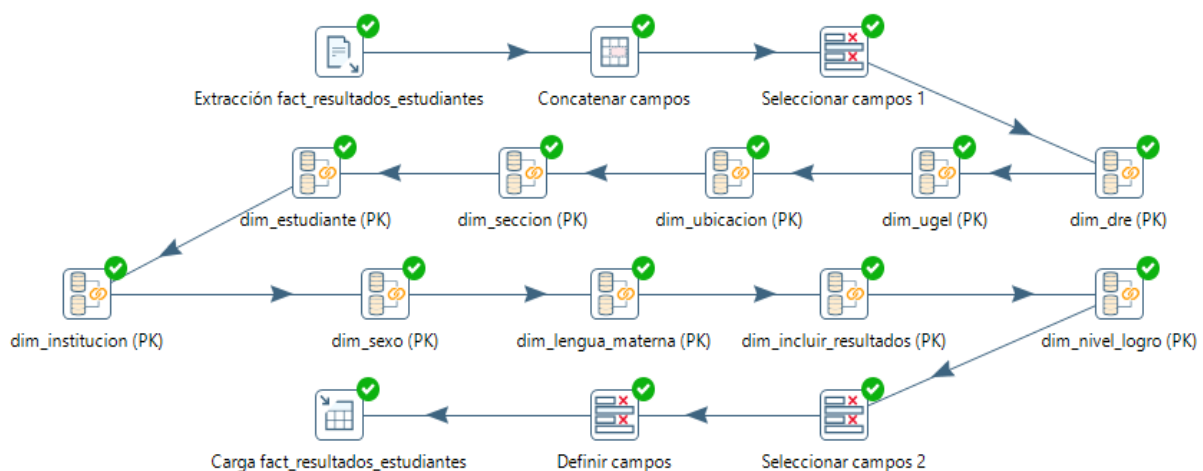


Figura 33. Pasos realizados en el ETL de la tabla fact_resultados_estudiantes

3.8. Especificación de aplicaciones de BI

A partir de los requerimientos realizados por parte de los usuarios en etapas anteriores, se establecieron 2 categorías para la elaboración y presentación de resultados:

a) Informes estándar

Son informes básicos, con un formato ya preestablecido y que son solicitados con mayor frecuencia. Tiene como finalidad la atención de los pedidos de información y la generación de reportes.

Este tipo de informe está dirigido al:

- Personal interno: integrantes de otras áreas que forman parte de la Instancia Técnica.
- Personal externo: aquellos que ingresan su pedido de información por medio de un oficio, formulario web o correo electrónico.

b) Aplicaciones analíticas

Son informes más elaborados, ya que implica un mayor análisis. Tiene como finalidad apoyar en la toma de decisiones acertada.

Este tipo de informe está dirigido a:

- La jefatura de la Instancia Técnica.
- Coordinadores y subcoordinadores de la Instancia Técnica.

3.9. Desarrollo de aplicaciones de BI

En base a las especificaciones anteriores, se procedió a desarrollar las dos categorías de presentación de resultados.

Para establecer la conexión del Datamart con las herramientas empleadas en ambas categorías, fue necesario instalar previamente el MySQL Connector/ODBC, cuyas siglas significan Open DataBase Connectivity. Este conector permite el acceso desde cualquier aplicación a una base de datos en MySQL. Generalmente es necesario instalarlo solo en equipos con Windows.

a) Informes estándar

- Para la atención de los pedidos de información, se crearon queries con el fin de automatizar aquellos que son más recurrentes. En consecuencia, para una próxima solicitud solo haría falta correr el query ya creado y exportar la base de datos. El formato de presentación para los pedidos de información es en un archivo Excel, debido a que se comparte tanto internamente como externamente y es un formato al que todos pueden tener acceso.
- Para la generación de reportes, en algunos casos se emplea el mismo procedimiento que para los pedidos de información. Existen otros reportes que requieren del análisis en los softwares estadísticos SPSS y R, por lo que es necesario su conexión con el Datamart. Para ello, se estableció la conexión de estos softwares con el Datamart mediante el conector ODBC y se accedió a las bases de datos mediante queries que se originan desde los mismos softwares. Posteriormente se realiza el análisis requerido y se comparte la información únicamente con el personal interno mediante un archivo en PowerPoint.

Algunos ejemplos de esta categoría son los siguientes:

Excel spreadsheet titled "ECE_2015_2P_alumnos - Excel". The interface shows the ribbon with options like ARCHIVO, INICIO, INSERTAR, DISEÑO DE PÁGINA, FÓRMULAS, DATOS, REVISAR, and VISTA. The active cell is A1, containing the formula "cod_mod7".

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	cod_mod7	anexo	ie	cod_DRE	nom_dre	cod_UGEL	nom_ugel	cdgeo	Region26	Provincia	Distrito	Gestion2	Area	Caracteristica2	seccion	cod_al	sexo_estu	grupo_3C	grupo_3M	M500_C_2015	M500_M_2015
2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
19	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
20	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
21	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
22	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
23	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
25	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
26	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Figura 34. Reporte de resultados a nivel de estudiante innominado, evaluados en la ECE 2015 en 2.º grado de primaria

Excel spreadsheet titled "CANT_IE_ALUM_EIB - Excel". The interface shows the ribbon with options like ARCHIVO, INICIO, INSERTAR, DISEÑO DE PÁGINA, FÓRMULAS, DATOS, REVISAR, and VISTA. The active cell is A1.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1		2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2018											
2	LENGUA EIB	IE	ALUMNOS	IE	ALUMNOS	IE	ALUMNOS	IE	ALUMNOS	IE	ALUMNOS	IE	ALUMNOS	IE	ALUMNOS	IE	ALUMNOS	IE	ALUMNOS	IE	ALUMNOS	IE	ALUMNOS
3	Aimara	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4	Asháninka	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5	Awajún	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6	Castellano como Segunda lengua	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	Quechua-Chanka	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8	Quechua-Cusco Collao	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9	Shipibo	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	TOTAL	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11																							
12																							
13																							
14																							
15																							
16																							
17																							
18																							
19																							
20																							
21																							
22																							
23																							
24																							

Figura 35. Reporte con la cantidad de IE y estudiantes evaluados en la ECE desde el 2007 al 2018, por lengua EIB

b) Aplicaciones analíticas

Para este tipo de informe se crearon Dashboards con el fin de tener una mejor visualización de los indicadores más relevantes y que esto conlleve a tomar mejores decisiones. La herramienta de visualización que se usó para la creación de estos Dashboards es el Power BI. Para ello, se estableció la conexión de esta herramienta con el Datamart mediante el conector ODBC y se accedió a las bases de datos mediante queries que se originan desde la misma herramienta. Posteriormente se crea el Dashboard y se comparte la información únicamente con el personal interno mediante el Power BI Service.

Algunos ejemplos de esta categoría son los siguientes:

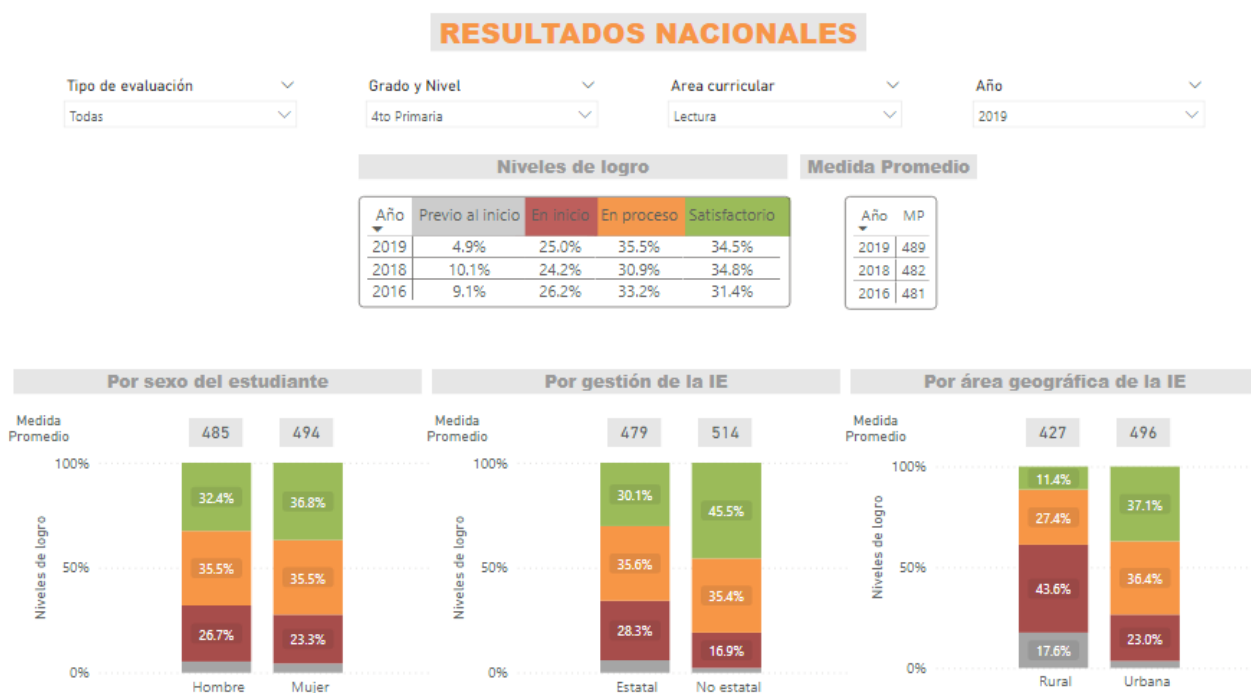


Figura 36. Reporte de resultados a nivel nacional por tipo de evaluación, grado y nivel, área curricular, periodo y los estratos sexo, gestión y área, según niveles de logro (porcentaje) y medida promedio

RESULTADOS REGIONALES

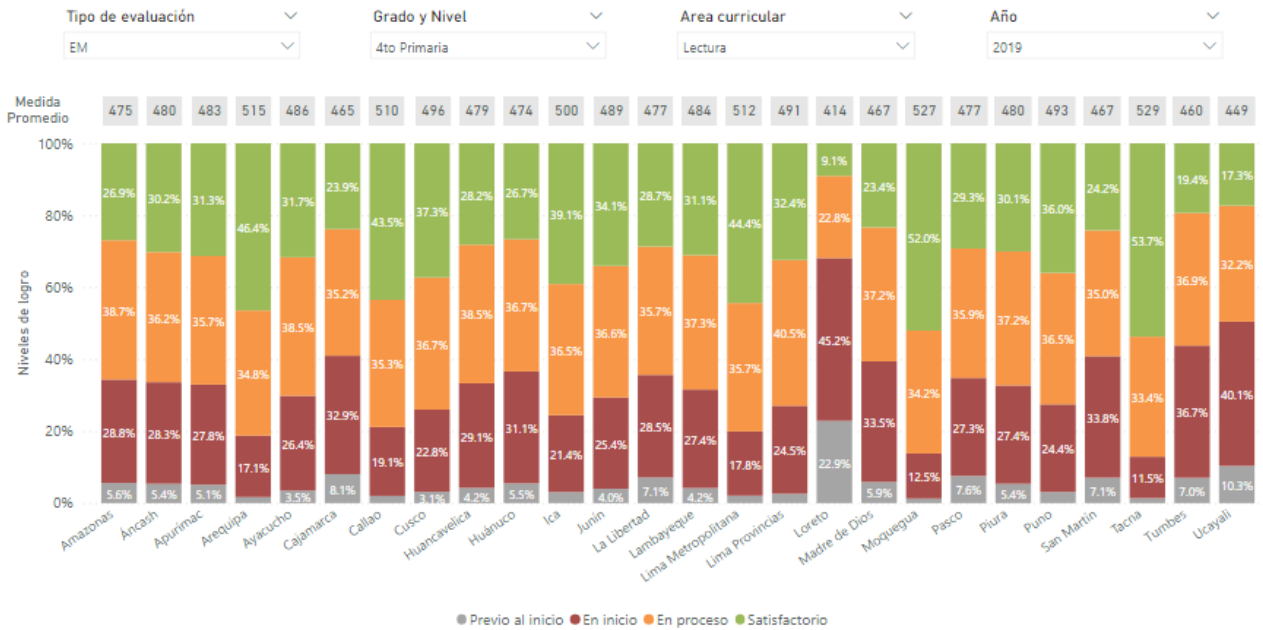


Figura 37. Reporte de resultados a nivel regional por tipo de evaluación, grado y nivel, área curricular, periodo y DRE, según niveles de logro (porcentaje) y medida promedio

3.10. Implementación

3.10.1. Plan de implementación

Para la implementación se desarrollaron las siguientes actividades:

- 1) **Validación de la información cargada en el Datamart:** Mediante la generación de reportes prueba se validó que la información cargada en el Datamart es correcta.
- 2) **Capacitación a los usuarios del Datamart:** Se realizaron reuniones para capacitar a los integrantes del área de Análisis.
- 3) **Capacitación a los usuarios tomadores de decisiones:** Se realizaron reuniones para capacitar a la coordinación y jefatura de la Instancia Técnica, respecto a la herramienta de visualización Power BI.

- 4) **Elaboración de insumos para el mejor entendimiento y uso correcto del Datamart:** Se elaboró el diccionario de datos del Datamart, un manual de uso y se crearon modelos de queries que apoye en la elaboración de los reportes.

3.10.2. Acceso a la herramienta de visualización Power BI

Para que la coordinación y la jefatura pueda tomar decisiones acertadas, se utilizó el Power BI Service, servicio en línea del Power BI, medio que permite visualizar los reportes generados con esta herramienta, desde cualquier dispositivo y con un usuario y contraseña establecidos.

3.11. Mantenimiento y crecimiento

3.11.1. Mantenimiento

Para el mantenimiento del Datamart, se ha establecido realizar un Backup una vez por semana, específicamente los días viernes al final del día.

3.11.2. Crecimiento

Respecto al crecimiento, se tiene previsto en un futuro realizar las siguientes actividades de mejora:

- Ingreso de nuevas bases de datos, producto de las evaluaciones que se realizarán en años posteriores.
- Actualización de las tablas que componen el Datamart y/o creación de nuevas, en base a nuevos requerimientos y necesidades del negocio.

- Implementación de nuevos análisis que permita sacarle el máximo provecho al Datarmart.
- Implementación de un aplicativo web, con el fin de que tanto personal interno como externo pueda tener acceso a las bases de datos del Datamart, según su perfil de usuario.
- Otras mejoras recomendadas por la coordinación y jefatura de la Instancia Técnica.

IV. RESULTADOS Y DISCUSIÓN

Los resultados se detallan en base a los siguientes indicadores:

- KPI 1: Tiempo empleado en la atención de los pedidos de información y generación de reportes (en minutos).

- Atención de los pedidos de información: Con el proceso anterior, atender un pedido de información tomaba aproximadamente entre 30 a 60 minutos dependiendo del requerimiento. Actualmente con la implementación del Datamart el tiempo que toma es de máximo 10 minutos.
- Generación de reportes: Con el proceso anterior, generar un reporte tomaba aproximadamente entre 120 a 180 minutos dependiendo del requerimiento. Actualmente con la implementación del Datamart el tiempo que toma es de máximo 45 minutos.

- KPI 2: Nivel de disponibilidad de la información

Actualmente el Datamart cuenta con información actualizada a la fecha, y esta se encuentra disponible las 24 horas del día para todos sus usuarios, por lo que el nivel de disponibilidad es en todo momento.

- KPI 3: Nivel de confiabilidad de la información.

Con los reportes prueba que se generaron para validar la información cargada en el Datamart, se obtuvo un nivel de confiabilidad muy alto ya que no se encontraron inconsistencias.

- KPI 4: Nivel de satisfacción de los usuarios respecto a los reportes generados.

Los usuarios manifestaron tener un nivel de satisfacción muy bueno, debido a dos factores principales: por la reducción en los tiempos de entrega de los reportes además de que la información contenida pasó a ser más confiable, y porque ahora la presentación de resultados es interactiva y permite visualizar la información en tiempo real lo que conlleva a tomar decisiones inmediatas y acertadas.

V. CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

- En base a los resultados obtenidos, la implementación del Datamart permitió mejorar los tiempos de respuesta en la atención de los pedidos de información y generación de reportes, dando un mayor tiempo para desarrollar análisis más exhaustivos con el fin de generar indicadores que apoyen en la toma de decisiones.
- La implementación de la Inteligencia de Negocios permite tener un control de accesos al Datamart y a los Dashboards según los perfiles de usuario.
- Se constató que la implementación del proceso de Inteligencia Negocios permite contar con información actualizada en tiempo real y disponible las 24 horas del día.
- Se verificó que con la implementación del Datamart se mejoró la consistencia de la información brindada.
- Se comprobó que la interfaz del Datamart y de los Dashboard generados, ayuda a que estos sean de fácil uso y comprensión para los usuarios.
- Se concluye que la implementación del proceso de Inteligencia de Negocios en la Instancia Técnica, aplicando la Metodología Kimball, resultó beneficioso, ya que permite contar con información integrada, confiable y oportuna. Adicionalmente, la presentación de resultados empleando Dashboards permite que los usuarios puedan identificar los indicadores más relevantes que apoyen en la toma de decisiones acertada.

5.2. Recomendaciones

- Se sugiere elaborar un diccionario de datos y un manual, con la finalidad de proporcionar estos insumos a los usuarios para que tengan una mejor y rápida

compresión de la funcionalidad del Datamart y del proceso de Inteligencia de Negocios en general.

- Se recomienda realizar capacitaciones a los usuarios con respecto al uso de las herramientas de Inteligencia de Negocios implementadas.
- Se propone seguir implementando la Inteligencia de Negocios en las demás áreas, con el fin de construir en un futuro el Datawarehouse de la Instancia Técnica.
- Se aconseja poner mayor énfasis en la etapa de definición de requerimientos del negocio, ya que de esa etapa depende la correcta implementación de la Inteligencia de Negocios.
- Se recomienda tener constante comunicación con los usuarios, con el fin de conocer mejor sus necesidades.
- Se sugiere realizar constantemente backups del Datamart, según la frecuencia con que se ingresa información nueva.

VI. REFERENCIAS BIBLIOGRÁFICAS

- Asto Huamán, L., & Arangüena Yllanes, M. R. (2018). *Inteligencia de negocios en la gestión académica de la educación superior universitaria*.
- Breslin, M. (2004). *Data Warehousing Battle of the Giants: Comparing the Basics of the Kimball and Inmon Models*.
- Flórez Fernández, H. A. (2012). Inteligencia de negocios como apoyo a la toma de decisiones en la gerencia. *Vínculos*, 11-23.
- Gartner. (s.f.). Obtenido de <https://www.gartner.com/en/information-technology/glossary/business-intelligence-bi>
- Joyanes Aguilar, L. (2019). *Inteligencia de negocios y analítica de datos. Una visión global de Business Intelligence & Analytics*. Bogotá: Alfaomega Grupo Editor, México.
- Llombart, O. A. (2003). *BI: Inteligencia aplicada al negocio*. DAA Contenidos Digitales.
- Mazón Olivo, B. E., Rivas Asanza, W. B., Pinta, M. A., Mosquera Franco, A. A., Astudillo Pizarro, L. A., Gallegos Macas, H. C., & Piedra Pineda, B. V. (2017). *Dashboard para el soporte de decisiones en una empresa del sector minero*. Machala.
- Medina La Plata, E. H. (2013). *Business Intelligence: la información como arma competitiva*.
- Mosquera, L., & Hallo, M. (2014). Data Mart Para El Sistema De Servicios Sociales Del Conadis. *EPN*.
- Oltra Badenes, R. F. (2017). *Business Intelligence. Definición*.
- Pérez Paucar, L. O. (2019). *Análisis de información utilizando un dashboard para la toma de decisiones en la coordinación zonal I - educación*. Ibarra.
- Reyes Dixson, Y., & Nuñez Maturel, L. (2015). La inteligencia de negocio como apoyo a la toma de decisiones en el ámbito académico. *GECONTEC: Revista Internacional de Gestión del Conocimiento y la Tecnología*, 63-73.

Rivadera, G. R. (2010). *La metodología de Kimball para el diseño de almacenes de datos (Data warehouses)*.

Ruiz Borja, J. E. (2018). *Comparación de herramientas ETL de código abierto*. Medellín.

Silva Solano, L. E. (2017). *Business Intelligence: un balance para su implementación*. InnovaG.

Vásquez Castrillón, J. B., & Sucerquia Osorio, A. (2011). *La Inteligencia de Negocios: Etapas del proceso*. Pereira.

Zambrano Tirado, L. J. (2018). *Datamart para la gestión de incidencias en el centro de atención al usuario en la sede central del Ministerio de Educación*. Lima.

Zamora Saldarriaga, J. E. (2017). *Implementación de un Datamart para la mejora en la toma de decisiones en el control de la demanda eléctrica del Comité de Operaciones Económicas del Sistema Interconectado Nacional*. Lima.