DISEÑO E IMPLEMENTACIÓN DE UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS PARA LA GESTIÓN Y EL ANÁLISIS DE LA INFORMACIÓN A REPORTAR AL SNIES

David Vargas Ramos

Universidad Magdalena

Facultad, (Ingeniería)

Programa (Ingeniería de Sistemas)

Santa Marta, Colombia

2022

DISEÑO E IMPLEMENTACIÓN DE UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS PARA LA GESTIÓN Y EL ANÁLISIS DE LA INFORMACIÓN A REPORTAR AL SNIES

David Vargas Ramos

Trabajo presentado como requisito parcial para optar al título de:

Ingeniero de Sistemas

Director (a):

Doctor Alexander Bustamante

Línea de Investigación:

Ingeniería y Desarrollo de Software

Grupo de Investigación:

Desarrollo y Gestión de Tecnologías para las Organizaciones y la Sociedad - TecnOS

Universidad del Magdalena

Facultad, (Ingeniería)

Programa (Ingeniería de Sistemas)

Santa Marta, Colombia

2022

Nota de aceptación:

Aprobado por el Consejo de Programa en cumplimiento de los requisitos exigidos por el Acuerdo Superior N° 11 de 2017 y Acuerdo Académico N° 41 de 2017 para optar al título de Ingeniero de Sistemas.

Jurado

Jurado

Santa Marta,_____de____de _____

Resumen

La gestión eficiente de la información en las Instituciones de Educación Superior, se ha convertido en una condición indispensable para el correcto funcionamiento de las mismas y para soportar el proceso de toma de decisiones en que estas se ven involucradas diariamente. En este contexto, todas las universidades ubicadas dentro del territorio nacional deben compartir con el Ministerio de Educación Nacional información sobre su gestión. Para este fin, la Universidad del Magdalena comparte información a través del SNIES con el Ministerio, no obstante, a esta información enviada se le hace poco seguimiento de manera que ayude a posteriori a la Universidad a tomar decisiones que permitan mejoras en las gestiones realizadas.

Dado el contexto descrito, este proyecto se enfocó en diseñar e implementar una solución de inteligencia de negocios que reúna y centralice todos los datos que son compartidos a través del SNIES y que suelen estar dispersos en diferentes archivos bajo custodia del personal administrativo responsable de subir los datos dentro de la universidad. La solución construida se compone de una plataforma web que permite la administración de los datos y de los usuarios, una bodega de datos que consolida los datos y un componente de visualización que es usado en la generación de reportes gráficos que permiten ver detalladamente cada punto de gestión de la Universidad del Magdalena en sus diferentes dependencias.

Para la construcción de la solución se siguió un enfoque iterativo incremental que empezó con la identificación de las fuentes de datos que se debían utilizar en la plataforma, seguido del diseño de la bodega e implementación de todos los procesos de extracción, transformación y carga que permitiesen tener la información actualizada de manera oportuna y, finalmente, con la articulación de todos los componentes en la plataforma web.

Palabras claves: Inteligencia de negocios, Bodegas de datos, Cubos OLAP

ABSTRACT

The efficient management of information in higher education institutions has become and indispensable condition for their proper functioning and to support the decision-making process in which they are involved daily. In this context, all universities located within in the national territory must share information about their management with the Ministry of National Education. For this purpose, the University of Magdalena shares information through the SNIES with the Ministry, however, this information sent is little monitored so that it doesn't helps the University to make decisions that allow improvements in the procedures carried out.

Given the context described, this project focused on designing and implementing a business intelligence solution that gathers and centralizes all the data that is shared through the SNIES and that is usually dispersed in different files under the custody of the administrative staff responsible of upload the data within the university. The built solution consists of a web platform that allows the administration of data and users, a data warehouse that consolidates the data and a visualization component that is used in the generation of graphical reports that allow a detailed view of each point of information management of the University of Magdalena in its different dependencies.

For the construction of the solution, an incremental iterative approach was followed that began with the identification of the data sources that should be used in the platform, followed by the design of the warehouse and implementation of all the extraction, transformation and loading processes that allowed have the information updated in a timely manner and, finally, with the articulation of all the components in the web platform.

Keywords: Business Intelligence, Data Warehouse, OLAP Cube

Tabla de Contenido

Introducción 9
Justificación 10
Objetivos generales 12
Objetivos específicos 12
Fundamentación y marco teórico 13
  Inteligencia de negocio 13
  Cubos de Procesamiento Analítico en Línea 18
Estado del arte 23
Ejecución del proyecto 25
Identificación de las fuentes 25
  Construcción de modelo dimensional 30
    Dimensiones 30
    Tabla de Hechos por conjunto de datos 32
  Definición e implementación de estrategia para la carga de datos 34
Presupuesto 49
Cronograma 50
Conclusiones 51
Referencias 53

Listado de Ilustraciones

Ilustración 1. Modelado de esquema de estrella 17
Ilustración 2. Modelo conceptual cubo OLAP 19
Ilustración 3. Plantilla admitidos usadas por las oficinas de planeación de la Universidad del Magdalena 26
Ilustración 4. Registros Excel convertidos a un formato JSON. 35
Ilustración 5. Parámetros función dataProcesor del servicio insertion_manager... 37
Ilustración 6. Objeto para entidad dimensional programa presente en la función dataProcessor. 38
Ilustración 7. Objeto entidad de hecho docentes presente en la función dataProcessor. 38
Ilustración 8. Parámetros de función createOrUpdateRecord del servicio insertion_manager. 38
Ilustración 9. Cartas para la carga de datos de plantilla de población estudiantil. 40
Ilustración 10. Pasos a seguir para la carga de datos. 41
Ilustración 11. Arquitectura de manejo de diferentes herramientas para la visualización de los datos. 42
Ilustración 12. Modelo estudiantes en Power BI 43
Ilustración 13. Modelo de docentes en Power BI. 44
Ilustración 14. Modelo de información ejecución presupuestal en Power BI. 45
Ilustración 15. Modelo de balance generales en Power BI. 46
Ilustración 16. Modelo de información presupuestal Power BI. 46
Ilustración 17. Generador de código iframe 47
Ilustración 18. Visualización de gráficos de recursos humanos en plataforma web. 47
Ilustración 19. Visualización de gráficos de población estudiantil en plataforma web. 48
Ilustración 20. Programación de actualización de información en gráficos. 48

Listado de Tablas

Tabla 1. Definiciones de inteligencia de negocio. 13
Tabla 2. Variable tipo de identificación tomada del manual del SNIES 27
Tabla 3. Presupuesto 49
Tabla 4. Cronograma 50

Introducción

Las Instituciones de Educación Superior, en particular las universidades públicas, están obligadas a reportar información de su gestión a diferentes organismos como el Ministerio de Educación Nacional (MEN), la contraloría general de nación, etc. [1]. En particular el MEN hace uso del Sistema Nacional de Información de la Educación Superior (SNIES), que ha sido creado para responder a las necesidades de información de la educación superior en Colombia [2]. En este sistema las universidades están obligadas a reportar de manera periódica, información relacionada con su gestión financiera, académica, de extensión, de docencia, etc.

En el presente proyecto se construyó una solución de Inteligencia de Negocios que permite a las universidades públicas en general, y a la Universidad del Magdalena en particular, hacer uso de la información recopilada para reportar al SNIES, de manera que se pueda realizar una gestión cuantitativa de sus actividades en forma transparente y eficiente. Para esto se hizo uso de un grupo de tecnologías que se agrupan bajo el término de Inteligencia de Negocios: los procesos de Extracción, Transformación y Carga para la preparación y limpieza de los datos; la Bodega de Datos para la integración y consolidación de los datos provenientes de las diferentes fuentes; los cubos OLAP y las herramientas de visualización para la navegación de la información [3][4][5].

De esta manera, se aprovecharon las diferentes fuentes de información que tiene la Universidad del Magdalena en cada una de sus diferentes dependencias y se hizo uso de los diferentes tipos de variables presentes en las plantillas del SNIES [6] con el fin de generar reportes gráficos de gestión que permiten una mejor evaluación por parte de las universidades, de la información que están reportando.

Como síntesis, con el siguiente proyecto, (a) se hizo énfasis en la información que la Universidad del Magdalena recolecta con miras a reportar en el SNIES; (b) Se analizaron las diferentes fuentes de información y de qué manera la información es presentada al SNIES; (c) con base en la información presente en cada una de las fuentes obtenidas, se diseñó una estrategia para el almacenamiento de los datos (d) se construyó un sistema que permite hacer tanto seguimiento y gestión de los indicadores de los procesos como análisis de la información almacenada; (e) para lograr lo anteriormente descrito, se utilizaron las herramientas agrupadas bajo el término Inteligencia de Negocios para este fin.

Justificación

En Colombia las instituciones de educación superior deben realizar reportes de su gestión al MEN. Este hace uso del Sistema Nacional de Información de la Educación Superior (SNIES). En este sistema las universidades están obligadas a reportar de manera periódica información relacionada con su gestión financiera, académica, de extensión, de docencia, etc.

En este mismo orden de ideas, el SNIES, como se establece en la resolución 19591 del 2017 [6], define un conjunto de variables que las instituciones públicas habilitadas para prestar y ofrecer servicios públicos de educación superior, como la Universidad del Magdalena, deben reportar. Dicha tarea puede resultar poco eficiente, debido a que la información puede no estar centralizada, ni sistematizada, lo que haría que conseguir la información necesaria a reportar sea un proceso bastante engorroso y demorado.

A raíz de lo anteriormente dicho, fue uno de los puntos que este proyecto buscó solucionar; que los datos con los que la Universidad del Magdalena cuenta actualmente puedan convertirse en información valiosa para la misma, centralizar dicha información y luego, por consiguiente, poder generar y visualizar los reportes periódicos.

Otro de los aportes de este proyecto es permitir que exista un mayor aprovechamiento de la información en el proceso de toma de decisiones y que se pueda establecer una gestión cuantitativa de la universidad, es decir, que partiendo de la información que se recoleta para reportar al SNIES, se pueda monitorear y analizar el rendimiento de los procesos.

Para hacer posible lo anterior, en el proyecto se construyó una solución de Inteligencia de Negocios (BI, por el término en inglés, *Business Intelligence*), tecnología que permite a organizaciones tomar decisiones basadas en información precisa y oportuna [2], esta agrupa bajo su término otras tecnologías como: los procesos de Extracción Transformación y Carga para la preparación y limpieza de los datos; la Bodega de Datos para la integración y consolidación de los datos provenientes de las diferentes fuentes; los cubos OLAP y las herramientas de visualización para la navegación de la información.

Existen varias universidades en Colombia que han implementado de manera eficiente este tipo de tecnologías, como lo son, por ejemplo, la Universidad Francisco de Paula Santander, en la consolidación de indicadores institucionales [3] o la Universidad Distrital Francisco José de Caldas para la administración de indicadores [4].

Además de lo anterior dicho, una de las tareas dentro del plan de gobierno del actual rector Pablo Vera Salazar es el diseñar e implementar un sistema institucional de evaluación del desempeño y gestión de indicadores claves de todos los procesos de la institución [5], por lo que el desarrollo de este proyecto sería un paso en esta dirección.

Objetivos generales

Diseñar e implementar una solución de Inteligencia de Negocios que utilice las variables e indicadores definidos por el SNIES para hacer una gestión cuantitativa de los procesos de la Universidad de Magdalena.

Objetivos específicos

1. Identificar las fuentes de información que utiliza la Universidad del Magdalena para reportar al SNIES, y la calidad de estas, de manera que se pueda tener una relación de qué fuentes se utilizan en la universidad para cuáles variables del SNIES.
2. Construir un modelo dimensional genérico que estructure la información a reportar en el SNIES, modelo que debe ser extensible y pueda ser utilizado por universidades con la misma necesidad.
3. Definir una estrategia que se ajuste al estado actual de las fuentes de datos de la Universidad del Magdalena, para cargar los datos en la bodega de manera incremental.
4. Con base en la estrategia definida implementar cada uno de sus puntos.
5. Desarrollar una plataforma web que permita visualizar los datos almacenados en la bodega y los cubos OLAP de manera dinámica e interactiva.

Fundamentación y marco teórico

El proyecto se llevó a cabo gracias a varias tecnologías agrupadas bajo el término de inteligencia de negocio: bodega de datos, cubos OLAP, la visualización de datos, entre otras, por ello es importante entender, ¿Qué es la inteligencia de negocio y cuáles son las tecnologías agrupadas bajo este término?

Inteligencia de negocio

La inteligencia de negocio es una herramienta que permite a organizaciones tomar decisiones basados en información precisa y oportuna [2]. Varios autores expertos en la materia han definido en libros, artículos y comentarios la inteligencia de negocio, entre los cuales podemos encontrar varias definiciones en la Tabla 1 [1].

Tabla 1. Definiciones de inteligencia de negocio.

Publicación	Autor	Definición
Tutorial: Businnes Inteligence - Past, Present, and Future. Communications of the Association for Information System [7].	J. H. Watson	Es una categoría amplia de aplicaciones, tecnologías y procesos para la recolección, almacenamiento, el acceso y análisis de datos para ayudar a los usuarios a tomar mejores decisiones de negocio
Inteligencia de negocios – definición [8].	S. Españeida	Sirve para indicar el conjunto de conceptos y métodos para mejorar la toma de decisiones en los negocios, utilizando sistemas de apoyo basado en hechos
Business Intelligence Roadmap: The Complete Project Lifecycle for Decision Support Applications. Publisher: Addison Wesley Pub Date [9].	L. Moss	Es una arquitectura y una colección de aplicaciones que sirven de apoyo en la toma de decisiones operativas
Business intelligence: The savvy manager's guide [10].	D. Loshing	Los procesos, las tecnologías y las

		herramientas necesarias para convertir los datos en información, la información en conocimiento y el conocimiento en los planes que impulsan acción rentable empresarial
The profit Impact of Business Intelligence [11].	S. Willian	Información comercial y de negocios clave que conducen a decisiones y acciones y que se traducen en un mejor desempeño empresarial
Building a Data Warehouse: With Examples in SQL Server [12].	V. Rainardi	Conjunto de actividades para comprender las situaciones de negocios mediante la realización de diversos tipos de análisis sobre los datos de la compañía, así como en datos de terceros para ayudar a la toma de decisiones del negocio estratégicas, tácticas y operativas necesarias y tomar acciones para mejorar el desempeño del negocio

Tomando como base estas definiciones podemos concluir que la Inteligencia de Negocio son un conjunto de herramientas o aplicaciones, con las cuales se pueden obtener y guardar información, para posteriormente ser usada como apoyo en la toma de decisiones.

Como se puede observar en la conclusión, la inteligencia de negocio la conforma un conjunto de componentes o herramientas; para la obtención y carga de datos se tiene el proceso de ETL (*Extraction, Transformation, Load*), para el almacenamiento de dichos datos el área de almacenamiento: las Bodega de datos y para consultas de esos datos almacenados los Cubos de Procesamiento Analítico en Línea

(OLAP). En los siguientes apartados se describirán cada uno de los componentes anteriormente mencionados.

Bodega de datos

Una Bodega de datos es un repositorio o área de almacenamiento donde todos los datos de la empresa se guardan en un solo lugar. Esto incluye datos de diferentes fuentes, así como datos actuales e históricos [13]. Humphries define a la Bodega de datos como una colección de bases de datos integradas, de carácter temático diseñado para proporcionar la información necesaria para la toma de decisiones [14].

Una de las ventajas que tiene implementar una bodega de datos es que permite almacenar gran cantidad de datos, no obstante, esa gran cantidad de datos se almacena de una manera en específico que permite a los empresarios acceder a la información de una manera rápida, entonces, las bodegas de datos se convierten en el pilar necesario para la implementación de la inteligencia de negocios [13].

Las bodegas de datos requieren una combinación de metodologías, técnicas, *hardware* y los componentes de *software* que proporcionan en conjunto la infraestructura para soportar el proceso de información [15]. La estructura que se defina debe reflejar las necesidades y características del negocio, sus departamentos, equipos de trabajo y directivos, esto permitirá responder a interrogantes generados al tratar de tomar las decisiones [16] y con el tiempo se va convirtiendo en la memoria corporativa [17] describiendo el pasado y el presente de la empresa. Las bodegas de datos desglosan, resumen, ordenan y comparan, pero no descubren, ni predicen [18].

Mucha de la información acerca de una buena implementación de las bodegas de datos proviene de Ralph Kimball, sus técnicas para el modelado dimensional han sido reconocidas por ser rápidas y comprensibles, por ello muchas de las implementaciones de inteligencia de negocios que conllevan desarrollo de bodegas de datos utilizan las definiciones y técnicas dadas por Ralph Kimball en sus libros. Debido a esto para dar una mejor definición de las técnicas y conceptos usados dentro de la implementación correcta de las bodegas de datos, se mostrará las definiciones y ejemplos que ha dado Kimball a cada una de ellas [19].

Según Kimball, las bodegas de datos implementan un concepto llamado tablas de hechos y tablas de dimensiones. Para Kimball las tablas de hechos en un modelo dimensional almacenan las medidas de desempeño resultantes de los eventos de procesos de negocio de una organización. Las tablas de hecho según Kimball, tienen como atributos medidas contables empresariales tales como, sumas, promedios, etc. Por lo tanto, cada fila en esta tabla de hecho representa un evento de medición [19]. Por otra parte, una tabla de dimensión es un conjunto de datos

relacionados que brinda información acerca de un eje central llamado tabla de hecho [19]. En ese sentido las variables que estén dentro de una dimensión pertenecen al hecho y dan información acerca de este.

Dentro de los ejemplos que utiliza Kimball está el de un supermercado; por la caja registradora de dicho supermercado pasan productos y cada vez que algún producto pase por ella, la caja registradora cuenta una unidad en sus registros, en ese caso, producto sería una tabla de hecho. [19].

Las tablas de hechos almacenan por lo general valores numéricos. Kimball especifica que, si estos datos numéricos almacenados son aditivos, sería de más provecho dicha tabla de hecho, esto debido a que, una tabla de hecho puede devolver miles de filas, que representa miles de hechos a la vez, por ello, lo mejor y más útil que se puede hacer con esas cantidades de filas, es sumarlas [19].

Por otro lado, Kimball menciona que, las tablas de dimensiones son complementos integrales de una tabla de hechos. Las tablas de dimensiones contienen el contexto textual asociado con un evento de medición de proceso empresarial, por lo tanto, describen el quién, qué, dónde, cuándo, cómo y por qué asociado al evento [19].

Los datos en común de una organización deben agruparse en tablas de dimensiones. Kimball menciona que los datos que se deben almacenar deben contener el más mínimo nivel de granularidad o datos de bajo nivel, esto quiere decir que si tenemos por ejemplo un dato que especifique una fecha, se puede crear una tabla dimensional llamada “tiempo” que almacene datos como el año, mes, semana y día en diferentes campos de ese dato fecha suministrado [19].

Como se pudo ver en el ejemplo de tiempo anterior, las tablas de dimensión pueden contener muchos atributos. Kimball menciona que no es raro ver tablas de dimensiones que lleguen a contener de 50 a 100 atributos o columnas [19].

El ejemplo básico para representar un esquema relacional de (hecho – dimensión) es con un modelado de esquema de estrella. Si seguimos con el ejemplo del supermercado, un modelado de esquema de estrella tendría la forma que se observa en la Ilustración 1.

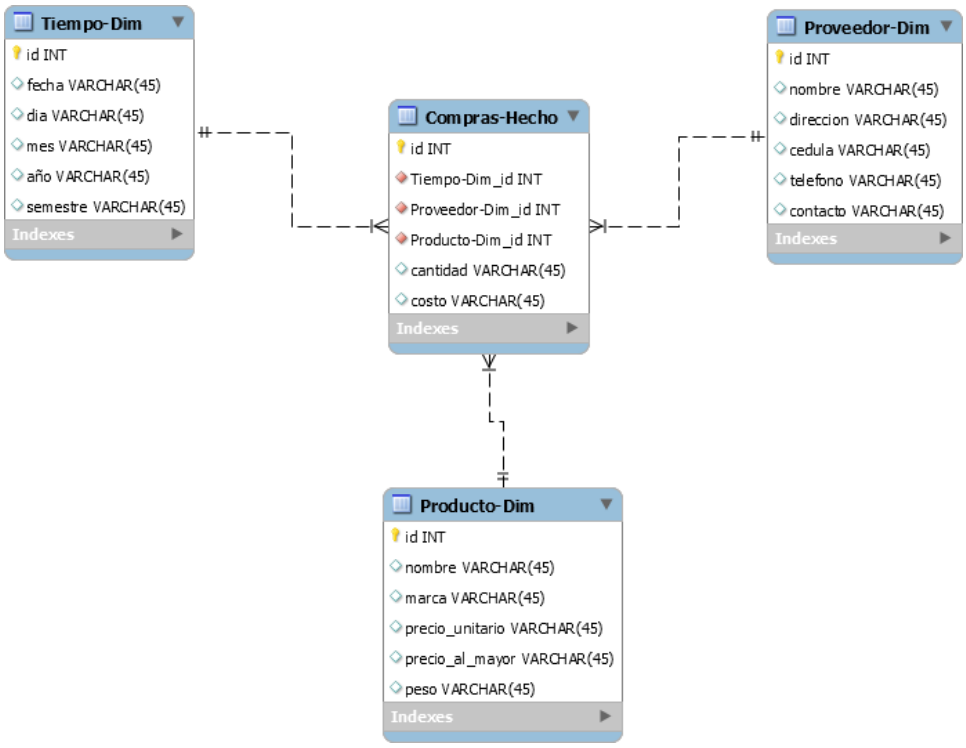


Ilustración 1. Modelado de esquema de estrella.

En donde en el centro se puede observar la tabla de hecho llamada “Compras-Hecho” y a sus costados las tablas dimensionales, “Tiempo-Dim”, “Producto-Dim” y “Proveedor-Dim”.

Extraer, transformar y cargar

El proceso de ETL (por sus siglas en inglés de *Extract, Transform and Load*) es el proceso en el cual se extrae, se transforma y se cargan los datos en la bodega de datos. Los datos vienen de diferentes fuentes internas recolectadas a través de sistemas de información, siendo posible extraer los datos e información histórica de archivos de texto plano u hojas de cálculo [20].

Kimball menciona que con la extracción lo que se busca es analizar los datos verificando si cumplen con la estructura necesaria [19]. Por ejemplo, la tabla dimensional de “Proveedor-Dim” de la Ilustración 1, almacena el nombre, dirección, cédula, teléfono y contacto de un proveedor; en el proceso de extracción validará que la fuente de datos que está suministrando la información tenga esos cinco datos que pide la tabla dimensional “Proveedor-Dim”.

Por otro lado, la transformación permite que esos datos que hayan sido extraídos se puedan modificar o adicionar información si se requiere, situaciones tales como

que existan datos o valores inadmisibles [21]. Por ejemplo, la tabla dimensional “Proveedor-Dim” espera en su campo “nombre” un dato de tipo texto, el proceso de transformación entonces valida que efectivamente el valor que está llegando como nombre sea un texto y no algo diferente como un número.

Por último, en la carga, los datos están debidamente estructurados, limpios y en algún tipo de esquema para luego ser cargados al sistema destino [21][22][23].

Cubos de Procesamiento Analítico en Línea

Los cubos de procesamiento analítico en línea (también conocido como cubos OLAP por sus siglas en inglés de *On-Line Analytical Processing*) son estructuras multidimensionales que permiten el análisis a bases de datos relacionales de gran tamaño. Una de sus características más importante es que las consultas que se realizan en ella se efectúan con gran rapidez [24]. A continuación, se encuentran otras de sus características:

- Estas herramientas soportan el análisis interactivo de la información de resumen, soportando muchas tareas de agrupación de datos que no pueden realizarse empleando las facilidades básicas de agregación y agrupamiento [25].
- A diferencia de las bases de datos dimensionales que almacenan su información en matrices unidimensionales, los cubos OLAP realizan un almacenamiento lógico de los datos en matrices multidimensionales [26].
- Para grandes cantidades de datos que se requiera una capacidad de respuesta rápida, los cubos OLAP presentan ventajas respecto a los clásicos modelados dimensionales, obviamente con ciertas desventajas como el costo en rendimiento que esto conlleva [19].
- Un concepto que se presenta también con los cubos OLAP es la medida, que es el valor que irá tomando las variables que se están analizando, normalmente las medidas son valores que se pueden contar [26].

Al hablar de matrices multidimensionales se debe tener en cuentan que cada dimensión de la matriz tiene cierto nivel de detalle y que estos niveles pueden formar una jerarquía [26].

Si representamos un cubo OLAP con el modelado relacional del ejemplo del supermercado, el cubo tendría la forma conceptual de la Ilustración 2.

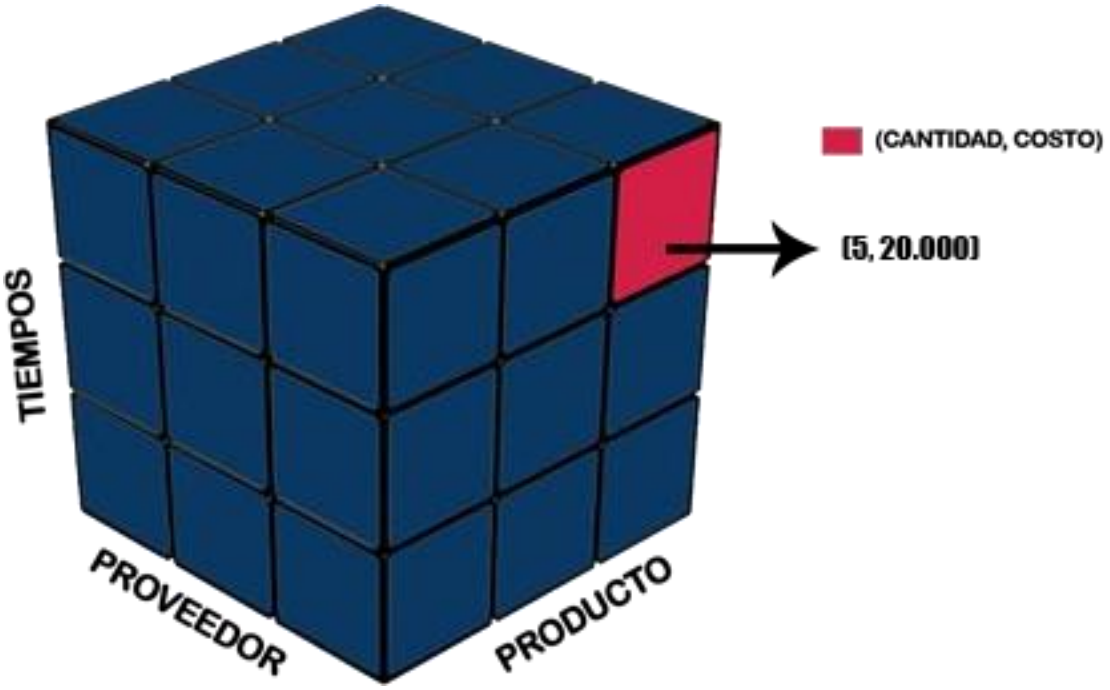


Ilustración 2. Modelo conceptual cubo OLAP.

Como se puede observar en el modelo conceptual de la Ilustración 2, en cada uno de sus ejes hay una dimensión de la tabla de hecho de compras y en rojo dos medidas a analizar: la cantidad y el costo, dos medidas almacenadas en la tabla de hecho.

En el eje de la parte derecha se tiene un producto del supermercado que alguna persona compró, en el eje de la parte inferior el proveedor de dicho producto y en el último eje el tiempo en el que la compra se realiza. La medida que se pretende analizar es la cantidad del producto que la persona compró y el costo que esta compra generó, una información que podría arrojar el análisis de dicha medida podría ser por ejemplo: “Compra de 5 Cheetos de Frito-Lay por un valor de 20.000, 17/03/2020”, donde el 5 sería la cantidad de producto y el 20.000 el valor que esa compra costó, Cheetos el nombre del producto en específico, Frito-Lay el proveedor y por último la fecha en que se realizó la compra.

Visualización de datos

Una vez se implementen herramientas de inteligencia de negocio, es necesario disponer de una plataforma de visualización de gráficos en donde se pueda observar los diferentes indicadores relevantes de análisis y reportes que la institución requiera.

Estas visualizaciones de gráficos les permiten a las compañías que implementen inteligencia de negocios entender con suma facilidad los datos que estos tienen almacenados en sus bases de datos. Estos reportes gráficos incluyen filtros que permiten los análisis personalizados de los datos que la situación requiera, esto quiere decir que, datos que antes estaban almacenados sin ninguna utilidad relevante, se convierten en información útil y valiosa para las compañías.

A continuación, se describirán las herramientas que fueron utilizadas para la realización de las visualizaciones de los datos en el proyecto, esto con el fin de que se conozca más a fondo el rol de cada una de ellas y lo importante que han sido en el mismo.

Power BI

Una potente herramienta que permite generar reportes gráficos a raíz de una base o bodega de datos es Power BI, una herramienta desarrollada por Microsoft que permite conectarse a los datos y posteriormente visualizarlos en gráficas unificadas [27].

En este orden de ideas, Power BI puede obtener los datos almacenados de una bodega de datos en MySQL, transformarlos y posteriormente realizar las gráficas que se crean necesarias. Además, se puede hacer uso de los datos obtenidos con Power BI para mejorar las operaciones y el rendimiento de las gestiones que debe realizar la Universidad del Magdalena, esto con el fin de que se puedan tomar las mejores decisiones basadas en los datos obtenidos al usar esta herramienta.

Power BI permite exportar las gráficas mediante un código *Iframe*, que puede ser incrustado en cualquier aplicativo.

Frontend

Un aplicativo de fácil acceso son las aplicaciones web, en donde los *Iframe* de gráficas generados por *Power BI*, pueden ser incrustados en estos.

El *Frontend* de un aplicativo, es la sección en donde los usuarios pueden acceder y ver la información, es esa parte en donde el usuario interactúa con el aplicativo.

Dentro de las tecnologías, herramientas y bibliotecas o librerías que se pueden encontrar cuando se habla del *Frontend* están HTML, CSS, JavaScript, React JS.

HTML

Según MDN Web Docs, HTML, es la pieza más básica para la construcción de la web y se usa para definir el sentido y estructura del contenido en una página web. (MDN Web Docs, 2020) [28].

CSS

CSS (Hojas de estilo en cascada) es un lenguaje que desarrolla la apariencia de un documento escrito con HTML. Este último define la estructura que debe tener la página y el otro define la apariencia de cada bloque de estructura, sección o elemento escrito en el documento.

JavaScript

JavaScript es un lenguaje de programación que permite insertar lógica programática en páginas web. MDN Web Docs habla acerca de la utilidad de JavaScript en las páginas web, menciona que cada vez que una página web hace algo más que sentarse allí y mostrar información estática para que la veas, muestra mapas interactivos, animaciones 2D y 3D, desplazamientos, reproductor de videos, tiene seguramente JavaScript implementando (MDN Web Docs, 2020) [28].

Javascript puede empaquetar múltiples variables en un tipo de datos conocido como objetos, que se comportan igual que un JSON.

Un JSON es un formato de texto que está formado por dos estructuras básicas:

1. Una colección de pares de key/valor (ítem) dentro de caracteres {}
2. Una lista ordenada de valores o colecciones (array) dentro de caracteres []

React JS

React JS, es una biblioteca de JavaScript que permite construir interfaces de usuarios [29], fue creada por Jordan Walke, ingeniero de Facebook.

Con React es posible desarrollar aplicativos webs de una sola página (SPA por sus siglas en *inglés de Single Page Aplication*), dando una experiencia de usuario agradable y rápida. Dentro de las grandes aplicaciones que han implementado esta biblioteca se pueden encontrar a las famosas redes sociales como Facebook e Instagram.

BACKEND

El *Backend* es el área lógica de un aplicativo. Implementa la arquitectura interna de cualquier aplicativo, los usuarios no pueden ver esta sección debido a que no tiene ningún elemento gráfico, dentro de sus funciones se pueden encontrar toda la lógica operacional a nivel de bases de datos como insertar, crear, actualizar, eliminar registros almacenados, también se encarga de la seguridad de dichos datos, encriptación de estos y creación de API REST.

La API REST, suministra una serie de *endpoint*, que, si luego se realiza una petición http a dichos *endpoint*, el *backend* realizará una tarea específica asociado a ellos.

Dentro de las herramientas o tecnologías que se pueden encontrar en un aplicativo web están NodeJS, ExpressJS, Sequelize.

Node JS

Node.js es un entorno de ejecución de JavaScript en tiempo real que permite ejecutar un programa desarrollado con JavaScript, en otras palabras, permite "sacar" a JavaScript del *frontend* y de su interpretación por el navegador, y lo compila fuera de este para lograr hacer aplicativos *backend*. Dentro de sus ventajas, como ya se mencionó, se compila a tiempo real, lo que trae consigo una mejor optimización de las peticiones que se le hagan a este, por lo tanto, brinda un alto rendimiento frente a otros lenguajes y tecnologías *backend*.

Grandes compañías como Netflix, Paypal y Uber implementan Node.js para sus diferentes aplicativos.

Express Js

Express.js es un *framework* de Node.js que permite crear aplicaciones web, permite gestionar sesión cookies, crear *endpoint*, administrar *cors*, implementar *middlewares*.

Sequelize

Sequelize es un ORM (*Object Relational Mapping,* modelo de programación que permite mapear las estructuras de una base de datos relacional) de Nodej.js basado en promesas para diferentes lenguajes de gestión de bases de datos tales como, MySql, SQLite, MariaDB entre otros, cuenta con un soporte sólido de transacciones, relaciones, carga y más [30].

Una de las muchas utilidades de utilizar un ORM es que permite efectuar operaciones sobre la base de datos utilizando un lenguaje de programación deseado, y tratar la información de esta como si fuesen objetos. A nivel de seguridad

también brinda ventajas debido a que protege la información de ataques como por ejemplo la inyección SQL.

Estado del arte

Existen múltiples proyectos que han usado las tecnologías y herramientas descritas en la fundamentación y marco teórico. Por ejemplo, la Universidad Distrital Francisco José de Caldas, usó elementos de inteligencia de negocios tales como las bodegas de datos, para administrar los indicadores del Sistema de Universidades Estatales (SUE) [31], la Universidad Distrital Francisco José de Caldas, implementó una bodega de datos para dar soporte a dichos reportes de indicadores, lo que permitió a la universidad almacenar, procesar y limpiar datos desde diferentes fuentes de la institución para luego poder generar análisis y reportes que dieran respuestas a los indicadores que pide el SUE.

Otra institución de educación superior que implementó inteligencia de negocios para la generación de indicadores utilizando bodega de datos y herramientas OLAP, fue la Universidad Francisco de Paula Santander [32], una vez implementado ayudó a la institución disminuir tiempos en la generación y entrega de los reportes exigidos por las entidades estatales de vigilancia y control, lo que ayudó a mejorar la credibilidad en la calidad de la información generada.

En la Universidad Distrital Francisco José de Caldas, dos estudiantes realizaron un prototipo de Bodega de datos para la toma de decisiones en el sector fiduciario utilizando la información histórica suministrada por el aplicativo SIFI [33].

El problema que se presentaba era que la información que manejaba las fiduciarias no estaba centralizada, sino que estaba almacenadas en diferentes fuentes y, por ende, dicha información no se estaba analizando y utilizando de manera adecuada, debido a que los datos únicamente estaban almacenados sin ningún tipo de propósito útil como, por ejemplo, generar reportes que ayudasen en la toma de decisiones.

Existen otros ejemplos de universidades de Colombia que han implementado las bodegas de datos como la Universidad Libre de Colombia, la cual implementó un diseño de bodega de datos para la empresa salón del automóvil [34], en esta empresa existía un problema con la información, debido a que la información que la empresa estaba almacenando tenía muchos datos redundantes, por lo tanto los reportes que se generaban no contenían información adecuada para analizar, por ello era necesario la implementación de una solución de inteligencia de negocios apoyada con bodegas de datos con el fin de almacenar grandes cantidades de datos

para luego poder acceder a ellos fácilmente y así generar reportes que ayuden en la toma de decisiones de dicha empresa.

Un grupo de estudiantes de Pontificia Universidad Javeriana, presentaron un proyecto que implementaban bodegas de datos para la utilización histórica para decisiones empresariales [35].

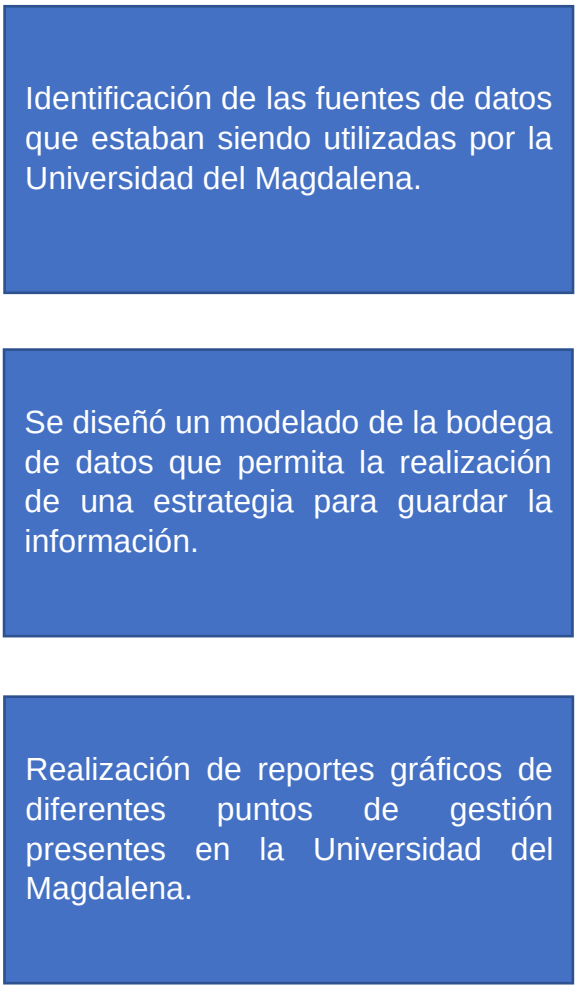
Se pueden encontrar variedades de proyectos realizados implementando cubos OLAP para resolver problemas para soluciones empresariales, uno de ellos fue el que realizaron dos estudiantes de la Universidad Estatal de Msilagro en donde diseñaron un cubo OLAP para la información contable generada por un *supermarket*, el problema que se presentaba en dicho *supermarket* era que los reportes que llegaban al área contable contenían muchas veces información no relevante y poco analítica, por lo tanto, fue necesario el desarrollo e implementación de un cubo OLAP que permitiese un análisis de la información gerencial del área contable [36].

Otro estudio que se realizó implementando cubos OLAP fue la tesis de Grace Isabel Núñez Soto en donde implementaron una solución de inteligencia de negocios para el área de finanzas de la municipalidad metropolitana de Lima, en donde el área de finanzas de dicha municipalidad presentaba problema para manejar la información de su gestión, por ello, se presentó una solución de inteligencia de negocios que implementó cubos OLAP que permitiese reducir el tiempo a los encargados al momento en que se necesitase tomar decisiones [37].

Por último, existen también proyectos que han implementado herramientas de visualización de datos, un claro ejemplo de este tipo de implementación de aplicativos webs para mostrar reportes gráficos es el de la Universidad de Eafit con su plataforma EAFIT en cifras, una plataforma que permite consultar información estadística de dicha Universidad tales como, éxito académico, indicadores históricos poblacionales, características de perfil de sus estudiantes, entre otros reportes gráficos [38].

Ejecución del proyecto

La ejecución del proyecto se desarrolló en varias etapas, las cuales son:



Se explicará a continuación a detalle, cada una de las etapas que permitieron el desarrollo oportuno del proyecto.

Identificación de las fuentes

Como se mencionó previamente, como base para la realización de este proyecto era necesario identificar qué fuentes utiliza la Universidad del Magdalena para reportar al SNIES, por ello se realizó una serie de reuniones con un representante

de las oficinas de planeación de la Universidad del Magdalena, debido a que una de las funciones de esta oficina, es la de presentar los informes a reportar a esta entidad de control.

En la primera reunión que se tuvo con el representante, este explicó en ella, por qué se reporta información de gestión de la Universidad al SNIES, qué fuentes de información usan, cómo presentan dichos reportes y qué guías utilizan para realizarlos. Como fuente de información, según explicó, están siendo usadas las diferentes dependencias que tiene la Universidad del Magdalena que son quienes tienen almacenado los datos pertinentes según el área al que pertenecen.

Teniendo la fuente de información, la Universidad del Magdalena sigue como guía la resolución 19591 de 2017 [6] en donde se especifican qué conjuntos de datos se debe reportar (llamados plantillas o módulos), así como también, los campos que tendrían dichos módulos, como, por ejemplo, las fechas que se deben tener en cuenta.

Entre estos módulos de datos se pueden encontrar, por ejemplo, módulos con nombre como: Información financiera, en donde se describe que dicho modulo hace referencia a un reporte trimestral de los saldos de PUC a nivel 6 y 8, de acuerdo con la resolución 643 de 2015. Otros módulos que se pueden encontrar allí son: matriculados, actividad cultural, cursos, etc.

A raíz de estos módulos, la oficina de planeación diseñó una serie de plantillas en Excel en donde cada una de estas tenía el nombre del módulo a la cual haría referencia y dentro de ellos en forma de columnas, los campos o variables pertenecientes a dichos módulos. Estas plantillas posteriormente se empaquetaron según su relación en carpetas a la cual llamamos conjuntos de datos.

Un ejemplo de ellos podría ser el conjunto de datos de población estudiantil, dentro del cual, están definidas las siguientes plantillas: Inscritos, Admitidos, Primer Curso, Matriculados y Graduados, en donde cada una de ellas contiene en forma de columnas sus respectivas variables a almacenar. (*Ilustración 3*)

	A	B	C	D	E	F	G	H	I	J	K
1	ID_IES	NOMBRE_IES	AÑO	SEMESTRE	COD_DANE	DEPARTAMENTO	MUNICIPIO	PRO_CONSECUTIVO	PROGRAMA	TIPO_DOCUMENTO	NUM_DOCUMENTO
2											
3											
4											
5											

Ilustración 3. Plantilla admitidos usadas por las oficinas de planeación de la Universidad del Magdalena.

Como resultado de esto, la oficina de planeación recolecta entonces la información necesaria de las diferentes dependencias y las usa para rellenar estas plantillas diseñadas, para presentar los informes de gestión.

Una guía extra que se tiene para que los datos que estén en cada una de las variables sea el correcto, es el manual de usuario del SNIES. En este manual el SNIES define el significado de todas las variables y la manera en las que estas variables deben ser presentadas. Se tiene por ejemplo la variable *"id_tipo_documento"* que está definida en el manual como: "código o sigla del tipo de documento principal de la persona que sirve como instrumento de reconocimiento de la identidad, nacionalidad y nombre", a raíz de esta definición presenta una tabla que indican los valores permitidos (ver Tabla 2).

Tabla 2. Variable tipo de identificación tomada del manual del SNIES.

ID_TIPO_DOCUMENTO	DESC_TIPO_DOCUMENTO	DEFINICIÓN
CC	Cédula de ciudadanía	Es el documento con el cual los colombianos mayores de edad pueden identificarse en todos los actos civiles, políticos, administrativos y judiciales.
DE	Documento de identidad extranjera	Cualquier documento expedido en el país de origen de las personas extranjeras con el cual se identifican.
CE	Cédula de extranjería	Es el documento expedido en Colombia por las personas extranjeras para identificarse en el territorio colombiano.
TI	Tarjeta de identidad	Documento oficial de identificación de los colombianos menores de edad entre los 7 y los 17 años de edad.
PS	Pasaporte	Documento que acredita la identidad y nacionalidad de las personas en el exterior, de acuerdo con la ley 12 de Versión: 2.4. Fecha de actualización: 10/Abril/2018 17 1947 mediante la cual Colombia se adhiere al Convenio sobre Aviación Civil Internacional de la Organización de Aviación Civil Internacional.
CA	Certificado cabildo	Documento que acredita una persona como un miembro de una comunidad indígena.

En total, los conjuntos de datos definidos por la oficina de planeación de la Universidad del Magdalena son de once conjuntos de datos, los cuales son Recursos Humanos, Población Estudiantil, Participante, Investigación, Internacionalización, Información Institucional, Información Financiera, Extensión, Derechos Pecuniarios y Bienestar.

Una vez se tuvo claro los tipos de fuentes usados por la Universidad del Magdalena, juntos con las guías y maneras en las que los reportes eran presentados, se tuvo una segunda reunión con el representante de la oficina de planeación, en la cual, mencionó que, de los once conjuntos de datos presentes, tenían más interés para una etapa inicial, únicamente trabajar sobre tres conjuntos de datos, los cuales son: Información Financiera, Población Estudiantil y Recursos Humanos. Este representante, además, dio acceso a plantillas de conjunto de datos de Población Estudiantil y Recursos Humanos con datos almacenados procedentes de los años 2018 y 2019.

Las plantillas presentes en cada uno de estos tres módulos están separadas por año y semestre, es decir, la plantilla de admitidos del conjunto de datos, población estudiantil, puede estar como, admitidos 2018-I y admitidos 2018-II, la información dentro de cada una de ellas, estará representando datos del año que especifique.

Mientras se iba avanzado en el proyecto, se notó que los campos presentes en las plantillas de Población Estudiantil no tenían datos suficientes para generar reportes gráficos puesto que la mayor cantidad de estos campos no eran datos cuantitativos, sino que, por lo contrario, era información cualitativa que hacía referencia a datos personales de los estudiantes como, por ejemplo, sus nombres y apellidos.

A raíz de esto, fue necesaria una tercera reunión con el mencionado representante, en la cual se acordó usar datos de estudiantes presentes en la plantilla “participante” del conjunto de dato con el mismo nombre.

Por ende, las plantillas que se trabajaron en el presente proyecto fueron.

- **Población estudiantil**
    - **Estudiantes inscritos:** Almacena información de los estudiantes inscritos a realizar el examen de admisión.
    - **Estudiantes admitidos**: Almacena información de los estudiantes admitidos en algún programa de la universidad.
    - **Estudiantes de primer curso**: Almacena información de los estudiantes nuevos en un año y semestre en específico.
    - **Estudiantes matriculados**: Almacena información de los estudiantes matriculados.
    - **Estudiantes graduados**: Almacena información de estudiantes egresados de la universidad.

- **Recursos Humanos**
    - **Docentes de universidad pública**: Almacena información acerca de cada uno de los docentes que hacen parte de la universidad.
    - **Contratos docentes**: Almacena información relevante de los contratos realizados a cada uno de los docentes de la universidad.
    - **Capacitación docente**: Almacena información de las capacitaciones que tiene cada uno de los docentes.
- **Información Financiera**
    - **Balance general y resultados**: Almacena información de relevante de resultados y balance de inversión realizada por la universidad en un semestre en específico.
    - **Ejecución Presupuestal**: Almacena información de ejecuciones realizadas por la universidad en base a su presupuesto financiero.
    - **Presupuesto Universidades Públicas**: Almacena información de los presupuestos manejados por la universidad.
- **Participantes**:
    - **Participantes**: Almacena información variada de los estudiantes.

Dentro del conjunto de datos de población estudiantil, los campos en general que pide cada plantilla de este, hace referencia a datos personales, académicos y financieros de los estudiantes, tales como, ciudad de residencia, número de documento, valores de la matrícula, sanciones académicas, materias pendientes entre otros datos.

Por otro lado, los campos que suele pedir el conjunto de datos de recursos humanos, además de los datos personales de quienes estén asociados contractualmente con la universidad, almacena dichos datos contractuales, datos como, cargo que ocupa, fecha de expiración del contrato, fecha de ingreso, periodo, títulos académicos, publicaciones (libros, artículos…) que hayan realizado los contratistas etc.

En el conjunto de datos de información financiera, los datos que allí se encuentra están orientados a datos de inversión de la universidad como pueden ser, recursos ejecutados, recursos aprobados, presupuestos aprobados, adiciones, reducciones, contra créditos, tipo de cuenta de ingresos de la universidad etc.

Por último, en la plantilla de participantes, se puede encontrar información variada de los diferentes participantes de la universidad; estudiantes, docentes, etc. Información tal como, sexo, país, novedad, entre otras. Como se mencionó, únicamente se tomaría la información de esta plantilla que haga referencia a estudiantes.

Construcción de modelo dimensional

A raíz de los conjuntos de datos anteriormente mencionados, se construye un modelo dimensional extensible para poder almacenar los campos o variables presentes en cada una de sus plantillas.

Los diferentes campos de las plantillas fueron agrupados según el tipo de información que suministran (conjunto de campos) para posteriormente ser modelados como tablas de dimensiones, por ende, las tablas de hechos serían la plantilla en sí misma, o como en el caso de población estudiantil, el conjunto de plantillas.

Dimensiones

En total la cantidad de dimensiones creadas una vez se analizó y se agruparon los diferentes campos fueron de veintidós dimensiones, los cuales son.

1. **Tiempos**, dos campos que se repetirían a lo largo de casi todas las plantillas de los diferentes conjuntos de datos es el campo AÑO que determina el año a reportar, y el campo SEMESTRE, que determina el semestre del año que se está reportando (1 o 2).

   Estas dos variables se agruparon, y se bajaron a su máximo nivel de granularidad, por lo tanto, esta dimensión tendría datos tales como, año, periodo, mes, semana del mes y día de la semana.

2. **Programas**, en esta dimensión se encuentran datos asociados al programa de estudio del estudiante o de asociación de contratistas, se pueden obtener datos como, nombre del programa, código SNIES del programa, municipio del programa, entre otras.

3. **Colegios**, En las diferentes plantillas del conjunto de datos de población estudiantil pide dos datos asociados con el colegio del estudiante: código de prueba saber de undécimo grado y título bachiller obtenido.

4. **Sanciones**, los campos contenidos en esta tabla de hecho hace referencia a datos acerca de sanciones recibidas por los estudiantes, tales como, por ejemplo, tipo de retiro, en caso de que aplique, el número de acto administrativo, la fecha del acto administrativo.

5. **Opcionesingresos**, se guardaron en esta tabla de hecho datos acerca de las opciones de ingreso de los estudiantes, orientados en su mayoría a el tipo de vinculación del estudiante con la que aplicó al ingresar, tales como, si el estudiante pertenece a un grupo étnico, comunidad negra, pueblo indígena, algún tipo de discapacidad, entre otras.

6. **Datosacademicos**, almacena datos referentes a la situación académica de un estudiante, promedio actual, promedio acumulado, números de materias inscritas, numero de materias aprobadas, semestre que cursa.

7. **Datosfinancieros**, tabla orientada a brindar información a la tabla de hecho del estudiante, en esta se almacena datos referentes a si un estudiante en específico ha recibido apoyos financieros o no por parte de la universidad.

8. **Datospersonales**, almacena información de datos personales acerca de alguna persona a almacenar, sexo biológico, municipio nacimiento, país de nacimiento, dirección, nombres, apellidos, fecha de nacimiento etc.

9. **Informaciongrados**, Se guardan datos acerca de información de grado del estudiante en caso de que aplique, datos como, por ejemplo, resultado de la prueba saber pro, observaciones de la prueba, número de folio, fecha del grado, entre otras.

10. **Estrategiacoberturas**, guarda información pertinente a tiempos y lugares de estudios del estudiante asociado, el municipio donde estudió, si estudió o no en la noche, si culminó los ciclos básicos.

11. **Junks**, las tablas de dimensiones que contienen *junks* en su nombre, son dimensiones que contienen campos que no tenían similitudes con otros campos y por ende no se podía agregar en ninguno de las otras dimensiones, por lo tanto, los datos que están en esta tabla son variados.

12. **Informacioncontractuales**, como su nombre lo indica, esta tabla de dimensión almacena datos pertinentes a información contractual de contratistas que hacen parte de la universidad, fecha de ingreso, cargo, fecha de expiración, etc.

13. **Informacionestudios**, almacena datos acerca de información de estudios asociado algún contratista de la universidad, por ejemplo, cuál es su título de pregrado, título de posgrado, qué título recibió, país donde estudió, nombre de la universidad donde obtuvo el título, etc.

14. **Junkdocentes**, dimensión *junks* orientado a la tabla de hecho docentes, sólo almacena un único dato que valida que el docente aplique al decreto 1279.

15. **Puntosdocentes**, almacena datos de validación de puntos obtenidos por los docentes en diferentes categorías, por ejemplo, si el docente tiene pregrado, maestrías, doctorados, si ha publicado artículos en revistas, si ha obtenido premios, si ha realizado traducciones, entre otras, se valida cada categoría y se almacena un valor 0 o 1 en cada una de ellas que representaría falso o verdadero respectivamente.

16. **Organogobiernos**, tabla de dimensión orientada a dar información a la tabla de hecho integrantegobiernos, almacena un único dato id órgano gobierno que almacena un valor entero según el área donde se ubique, estos valores son. Los valores asociados a cada uno de los valores están especificados en el manual guía del SNIES.

17. **Junkbalancegeneral**, dimensión *junks* orientado a la tabla de hecho balancegenerales.

18. **Junkinformacionpresupuestales**, dimensión *junks* orientado a la tabla de hecho información_ejecucion_presupuestales.

19. **Aportes**, almacena datos financieros de aportes de estampillas apropiadas y ejecutadas.

20. **Recursos**, datos de información financiera pertinentes a recursos propios o capitales apropiados y ejecutados.

21. **Transferencias**, almacena datos pertinentes a transferencias financieras apropiadas y ejecutadas por la universidad.

22. **Presupuestos**, datos presupuestales de la universidad, presupuestos aprobados, adicciones, créditos, reducciones, presupuestos ejecutados, etc.

Tabla de Hechos por conjunto de datos

1. Población estudiantil

A diferencia de los otros conjuntos de datos que serán posteriormente explicados, para el conjunto de datos de población estudiantil, al tener en sus diferentes plantillas casi los mismos campos o variables para almacenar, se agrupan y se modela una única tabla de hecho llamada estudiantes, la diferenciación entre el tipo de estudiante se realiza por medio de un campo

llamado estado, que, como su nombre indica, determina el estado actual del estudiante; inscrito, admitido, primer curso, matriculado o egresado.

Dentro de la tabla de hecho estudiantes, se agregan dos variables de referencia única del estudiante, su número de identificación y su código. Adicionalmente se agrega una variable llamada cantidad la cual todas las tablas de hechos en este proyecto la tendrán. Esta variable sirve para poder realizar conteos y operaciones estadísticas de una manera más fácil.

Las dimensiones, por ende, fueron el conjunto de todas las variables de las plantillas de este conjunto de datos, el total de dimensiones asociadas a la tabla de hecho de estudiantes son de 10 dimensiones que cubren en su totalidad todos los campos.

2. Recursos Humanos

Cada una de las plantillas presentes en este conjunto de datos, anteriormente descritas, representan una tabla de hecho.

2.1. **Docentes**, las plantillas de **docente de universidad pública**, **contratos docentes** y **capacitación del docente** se unen en una sola tabla de hecho llamada docentes y dentro de esta tabla de hecho se incluye un campo llamado tipo, que especifica a cuál de las tres plantillas mencionadas se está haciendo referencia.

Se almacena adicional a eso, la identificación del docente y seis dimensiones que brindan información de datos personales, tiempo reportado, información de estudio, información del contrato, y los puntos del docente.

2.2. **Autoridades**, incluye el campo identificación de la autoridad a la cual se hace referencia y tres dimensiones que brindan información acerca de los datos personales, el cargo que ocupa y el tiempo que se está reportando de la autoridad.

2.3. **Directivos**, incluye el campo identificación del directivo a la cual se hace referencia y cuatro dimensiones que brindan información de los datos personales, el cargo que ocupa, la información de estudio y el tiempo que se está reportando del directivo.

2.4. **Integrantesgobiernos**, hace referencia a la plantilla de integrantes máximo órgano de gobierno universitario, incluye el campo identificación del integrante a la cual se hace referencia y cuatro dimensiones que brindan información de los datos personales,

información contractual, el órgano de gobierno al que pertenece y el tiempo que se está reportando del integrante.

3. **Información Financiera**

   Cada una de las plantillas presentes en este conjunto de datos, anteriormente descritas, representan una tabla de hecho.

   3.1. **Balancegenerales**, hace referencia a la plantilla de balance generales y resultado, incluye dos dimensiones que hace referencia a información variada de balances generales y resultados de la universidad y los tiempos de reporte.

   3.2. **Información_presupuestal_universidad_publicas**, hace referencia a la plantilla de Presupuesto Universidades Públicas, incluye dos dimensiones que suministran información acerca de los presupuestos de la universidad como los tiempos de reportes de dichos presupuestos.

   3.3. **Información_ejecución_presupuestal**, hace referencia a la plantilla de Ejecución Presupuestal, incluye cinco dimensiones que hacen referencia a recursos, transferencias, aportes y tiempos de dichas ejecuciones presupuestales.

Definición e implementación de estrategia para la carga de datos

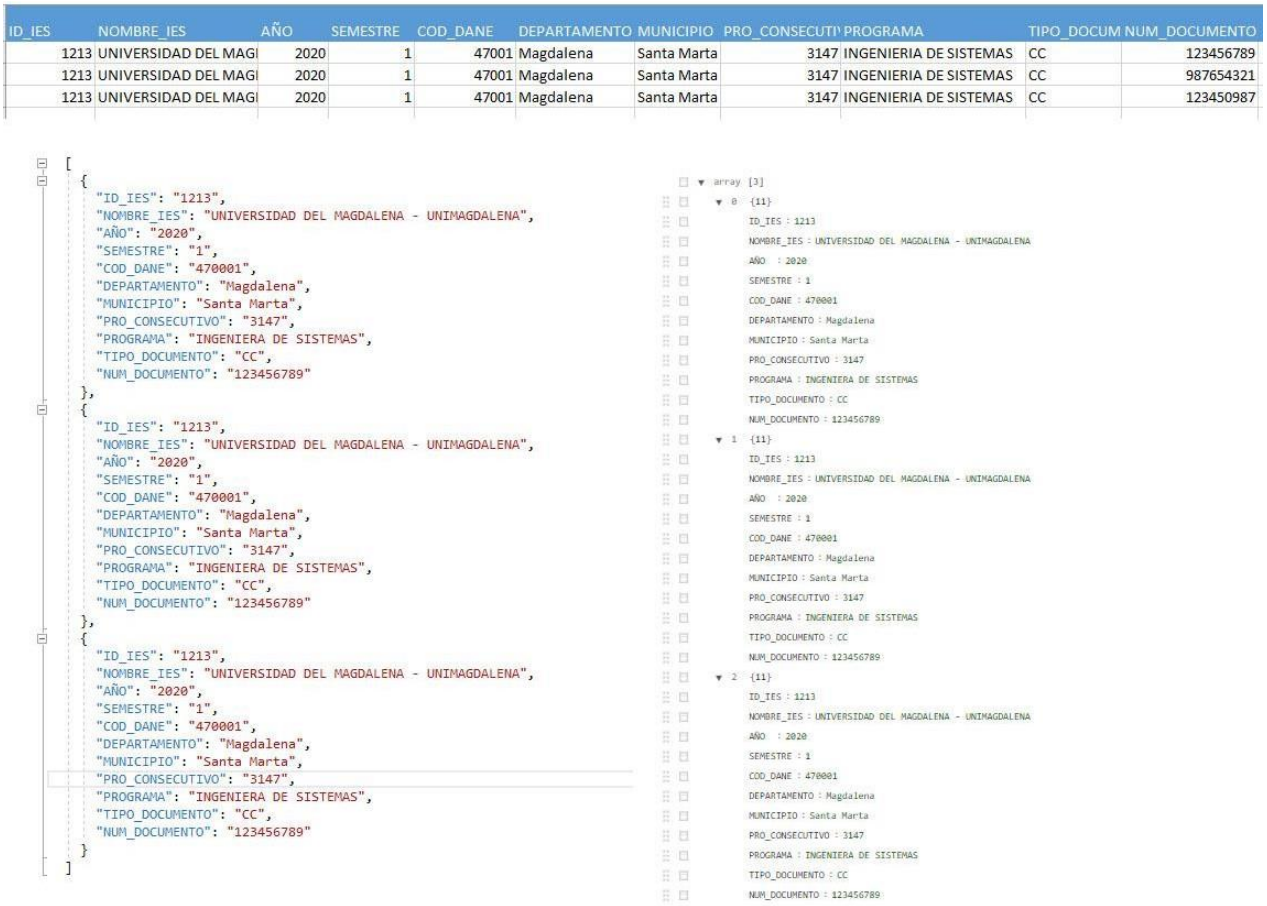
Al tener los datos empaquetados en plantillas y realizado el modelado de la bodega de datos dimensionales, con base en ellas fue posible reutilizar dichas plantillas para la carga de los datos.

Por ende, quienes quisieran cargas los datos a la bodega de datos, usarían una plataforma web para poder enviar las plantillas individuales llenas de datos a un servidor que procesaría cada fila de dicho registro y guardaría cada dato en su respectivo campo de la bodega de datos.

Implementación

Todas las plantillas enviadas al servidor siguen una serie de pasos intermedios, para luego, dependiendo del conjunto de datos al que pertenezca, un controlador define las condiciones que se deben tener para que los registros (filas independientes de las plantillas) puedan ser guardados en la bodega de datos.

Al llegar la plantilla al servidor, convertirá los registros almacenados en ella en un formato de texto JSON, que permite a *Javascript* un mejor manejo de dichos datos, de este modo los registros pasan de ser filas de una plantilla en específico a una colección de datos.



ID_IES	NOMBRE_IES	AÑO	SEMESTRE	COD_DANE	DEPARTAMENTO	MUNICIPIO	PRO_CONSECUTI	PROGRAMA	TIPO_DOCUM	NUM_DOCUMENTO
1213	UNIVERSIDAD DEL MAGI	2020	1	47001	Magdalena	Santa Marta	3147	INGENIERIA DE SISTEMAS	CC	123456789
1213	UNIVERSIDAD DEL MAGI	2020	1	47001	Magdalena	Santa Marta	3147	INGENIERIA DE SISTEMAS	CC	987654321
1213	UNIVERSIDAD DEL MAGI	2020	1	47001	Magdalena	Santa Marta	3147	INGENIERIA DE SISTEMAS	CC	123450987


```
[
  {
    "ID_IES": "1213",
    "NOMBRE_IES": "UNIVERSIDAD DEL MAGDALENA - UNIMAGDALENA",
    "AÑO": "2020",
    "SEMESTRE": "1",
    "COD_DANE": "470001",
    "DEPARTAMENTO": "Magdalena",
    "MUNICIPIO": "Santa Marta",
    "PRO_CONSECUTIVO": "3147",
    "PROGRAMA": "INGENIERA DE SISTEMAS",
    "TIPO_DOCUMENTO": "CC",
    "NUM_DOCUMENTO": "123456789"
  },
  {
    "ID_IES": "1213",
    "NOMBRE_IES": "UNIVERSIDAD DEL MAGDALENA - UNIMAGDALENA",
    "AÑO": "2020",
    "SEMESTRE": "1",
    "COD_DANE": "470001",
    "DEPARTAMENTO": "Magdalena",
    "MUNICIPIO": "Santa Marta",
    "PRO_CONSECUTIVO": "3147",
    "PROGRAMA": "INGENIERA DE SISTEMAS",
    "TIPO_DOCUMENTO": "CC",
    "NUM_DOCUMENTO": "123456789"
  },
  {
    "ID_IES": "1213",
    "NOMBRE_IES": "UNIVERSIDAD DEL MAGDALENA - UNIMAGDALENA",
    "AÑO": "2020",
    "SEMESTRE": "1",
    "COD_DANE": "470001",
    "DEPARTAMENTO": "Magdalena",
    "MUNICIPIO": "Santa Marta",
    "PRO_CONSECUTIVO": "3147",
    "PROGRAMA": "INGENIERA DE SISTEMAS",
    "TIPO_DOCUMENTO": "CC",
    "NUM_DOCUMENTO": "123456789"
  }
]
```



```
▼ array [3]
  ▼ 0 {11}
      ID_IES : 1213
      NOMBRE_IES : UNIVERSIDAD DEL MAGDALENA - UNIMAGDALENA
      AÑO : 2020
      SEMESTRE : 1
      COD_DANE : 470001
      DEPARTAMENTO : Magdalena
      MUNICIPIO : Santa Marta
      PRO_CONSECUTIVO : 3147
      PROGRAMA : INGENIERA DE SISTEMAS
      TIPO_DOCUMENTO : CC
      NUM_DOCUMENTO : 123456789
  ▼ 1 {11}
      ID_IES : 1213
      NOMBRE_IES : UNIVERSIDAD DEL MAGDALENA - UNIMAGDALENA
      AÑO : 2020
      SEMESTRE : 1
      COD_DANE : 470001
      DEPARTAMENTO : Magdalena
      MUNICIPIO : Santa Marta
      PRO_CONSECUTIVO : 3147
      PROGRAMA : INGENIERA DE SISTEMAS
      TIPO_DOCUMENTO : CC
      NUM_DOCUMENTO : 123456789
  ▼ 2 {11}
      ID_IES : 1213
      NOMBRE_IES : UNIVERSIDAD DEL MAGDALENA - UNIMAGDALENA
      AÑO : 2020
      SEMESTRE : 1
      COD_DANE : 470001
      DEPARTAMENTO : Magdalena
      MUNICIPIO : Santa Marta
      PRO_CONSECUTIVO : 3147
      PROGRAMA : INGENIERA DE SISTEMAS
      TIPO_DOCUMENTO : CC
      NUM_DOCUMENTO : 123456789
```


Ilustración 4. Registros Excel convertidos a un formato JSON.

Posterior a ello, los registros en formato *JSON* pasan por un filtro de validación, en donde se asegura que cada colección del *JSON* tenga los campos necesarios, así como que cada campo pertenezca a la plantilla a la cual se hace referencia.

Adicional a eso, valida que los valores de cada uno de los campos tengan el tipo correcto de datos, por ejemplo, que el campo “AÑO” sea un valor numérico y no de otro tipo.

En caso de que no ocurra ningún problema en las validaciones, se enviaría el *JSON* al controlador respectivo de cada tabla de hecho que maneja la lógica necesaria para la inserción de dichos datos en la bodega de datos.

Controladores

1. Validaciones.

La lógica para la inserción de los datos en la bodega de datos es la misma para todos los controladores, pero pueden presentar una variación en dicha lógica, según su tabla de hechos.

El principal objetivo de los controladores de estos conjuntos, es validar que cada registro no haya sido almacenado anteriormente en la bodega de datos, por ello, se obtiene la variable IDENTIFICACION que es una variable presente en todas las plantillas de población estudiantil y recursos humanos, así como también las variables, AÑO y PERIODO que están presentes en todas las plantillas.

Con estas variables, se realiza una consulta a la bodega de datos en busca de un registro que contenga una identificación dada en un año y periodo determinado. Se tienen en cuenta estas tres variables por las siguientes razones, la identificación es única por persona y esta se encuentra en la tabla de hecho, por lo tanto, en la bodega de datos no habría datos de una persona almacenada varias veces, no obstante, los datos almacenados pueden variar por año y periodo, por ejemplo, los datos académico de un estudiante, por ende, inicialmente un registro podría estar almacenado varias veces en la bodega de datos siempre y cuando no sean del mismo año y periodo.

Esto es así como lógica base de validación, sin embargo, como se mencionó anteriormente pueden existir variaciones a dicha lógica según la tabla de hecho.

Las plantillas del conjunto de datos de población estudiantil se unieron en una única tabla de hecho llamada estudiantes, por lo tanto, existen cinco plantillas que harían referencia a esta tabla de hecho, inscritos, admitidos, primer curso, matriculados y graduados, por ello, tienen una variación de la lógica base de validación del controlador; además de las tres variables ya mencionadas que esta valida, incluye la variable ESTADO en la validación.

Ocurre lo mismo con la tabla de hecho docentes, existen múltiples plantillas a la cuales se hace referencia, docente de universidad pública, contratos docentes y capacitación del docente, por ello, también implementa una variación en la lógica base, este incluye la variable TIPO en la validación.

Por otro lado, existen variaciones en la lógica base del controlador en todas las plantillas de información financiera, la validación de la variable identificación se elimina puesto que dicha variable no existe en sus tablas de hechos y únicamente se deja la validación de año y trimestre para la tabla de hecho de balances general y estado de resultado, validación de año en el hecho de información de ejecución presupuestal y las validaciones de año y periodo en el hecho de información presupuestal de universidad públicas.

2. Inserción de datos en bodega.

   Luego de las validaciones, se tiene en cuenta el resultado de esta para la inserción de los datos, si esta falla no se detiene el proceso, sino que no continúa al servicio de inserción de nuevos datos y por el contrario, continúa a un servicio de actualización de datos, esto es así porque es posible que exista algún error con un dato en concreto ya subido y se quiera corregir. Este servicio simplemente actualizaría el registro a nivel de la bodega de datos.

   Se debe tener en cuenta que los datos de un registro de las diferentes plantillas serán insertados en diferentes tablas de dimensiones y que posteriormente hará una inserción en la tabla de hecho relacionando los ids de las diferentes dimensiones previamente creadas.

   Para este fin se crea el servicio *insertion_manager* con dos funciones:

- *dataProcessor*, esta función crea el objeto a insertar de una determinada entidad, recibe tres parámetros: *data*, (registro de una determinada plantilla), *entityName* (nombre de la entidad en la cual se realizará la inserción o actualización) y *dataFact* (*ids* de las dimensiones a relacionar, solo para inserción o actualización de tablas de hechos)

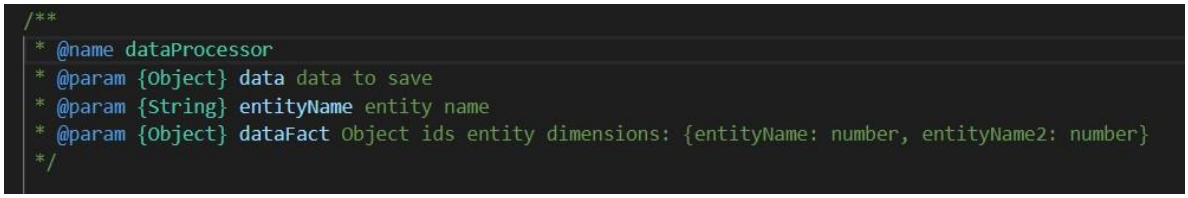


Ilustración 5. Parámetros función dataProcesor del servicio insertion_manager

Dentro de esta función se encuentran diferentes objetos que representan entidades de la bodega de datos con sus respectivos campos. Para los objetos que representan entidades dimensionales, sus campos serán rellenados con los valores presentes en el parámetro *data* en caso de que se encuentre el valor o *null* en caso contrario *(ver* Ilustración 6*).*

Para los objetos que representan entidad de tablas de hechos, sus campos serán rellenados con los *ids* presentes en el parámetro *dataFact (ver* Ilustración 7*).*

Posteriormente la función buscará y retornará el objeto que se desea insertar en la base de datos por medio de el parámetro *entityName*.

```
programas: {
    codigoDelProgramaSnies: String(data.PRO_CONSECUTIVO) || null,
    nombrePrograma: data.PROGRAMA || null,
    municipioPrograma: data.MUNICIPIO || null,
    idMunicipioPrograma: data.COD_DANE || null,
    departamentoPrograma: data.DEPARTAMENTO || null,
    id_ies: data.ID_IES || null,
    nombre_ies: data.NOMBRE_IES || null
},
```

Ilustración 6. Objeto para entidad dimensional programa presente en la función dataProcessor.

```
docentes: {
    cantidad: 1,
    tipo: data.tipo || null,
    identificacion: data.NUM_DOCUMENTO || null,
    tiemposId: dataFact.tiempos || null,
    datosPersonalesId: dataFact.datosPersonales || null,
    informacionEstudioId: dataFact.informacionEstudios || null,
    puntosDocenteId: dataFact.puntosdocentes || null,
    informacionContractualId: dataFact.informacionContractuales || null,
    junkDocenteId: dataFact.junkDocentes || null,
    capacitacionesId: dataFact.capacitaciones || null,
},
```

Ilustración 7. Objeto entidad de hecho docentes presente en la función dataProcessor.

- *createRecord*, realiza la inserción o actualización de los datos en una entidad determinada, recibe cinco parámetros; *entity* (Modelo sequelize de una entidad), *entityName*, *data*, *dataFact* y *idToUpdate* (id de entidad que se quiera actualizar)

  Dentro de esta función se realiza un llamado a la función *dataProcessor* para obtener el objeto y posterior a ello se realiza la creación o actualización del registro en la base de datos.

```
/**
 * @name createOrUpdateRecord
 * @param {Object} entity Entity model
 * @param {String} entityName entity name
 * @param {Object} data data to save
 * @param {Object} dataFact Object ids entity dimensions: {entityName: number, entityName2: number}
 * @param {Number} idToUpdate Id of entity to update
 */
```

Ilustración 8. Parámetros de función createOrUpdateRecord del servicio insertion_manager.

Endpoints

Carga de archivos

Para permitir el envío de las diferentes plantillas al servidor, se implementó una *API REST* con *endpoints* para cada una de las plantillas. Todos los *endpoints* generados siguen una convención dependiendo su uso.

Para el envío de plantillas es necesario especificar el conjunto de datos a la cual se hace referencia, seguido por la acción que se quiere realizar (para guardar siempre es "*upload*") y finalizando por el nombre de la plantilla a la cual se hace referencia, si queremos por ejemplo cargar datos de la plantilla de admitidos, su *endpoint* sería de la siguiente manera:

POST: */api/estudiantes/upload/admitidos*

La petición a estos *endpoints* puede devolver como resultados tres tipos de respuestas:

500: *Internal server error*: Un error inesperado ocurrido en el servidor.

400: *Bad request:* Puede suceder por dos razones, 1. La plantilla enviada no tiene una estructura válida o los datos contenidos no son válidos, 2. La plantilla no tiene datos.

200: *Success*: El servidor ha recibido las plantillas y ha enviado los datos al controlador.

Una vez el controlador finalice el proceso y los datos estén guardados en la bodega de datos, se envía un correo al usuario que subió los datos indicando que los datos han sido guardados.

Descarga de archivos

Cómo se mencionó anteriormente cuando una plantilla con datos es enviada al servidor se valida que tanto la estructura como los datos mismos sean acordes a lo que se pretende guardar, por ello también se generó una serie de *endpoints* para la descargar de estas plantillas.

Para descargar alguna plantilla se necesita especificar el conjunto de datos a la cual se hace referencia, seguido por la acción de descarga ("*download*") y finalizando por el nombre de la plantilla, siguiendo con el ejemplo del admitidos, para descargar su plantilla el *endpoint* sería:

POST: */api/estudiantes/download/admitidos*

Desarrollo de plataforma web

La plataforma web desarrollada tiene dos funciones en específico; permite la carga y envío de las diferentes plantillas al servidor que haría que la bodega de datos se vaya poblando y, por consiguiente, permite la visualización de dichos datos gráficamente.

Teniendo en cuenta los *endpoints* implementados, se desarrolló una plataforma web que hiciese uso de estos. Para la carga de plantillas cada conjunto de datos tiene su propia sección (llamada también página) en donde se mostrarán desplegados una serie de “cartas” especificando la plantilla a la cual hace referencia (Ilustración 9).

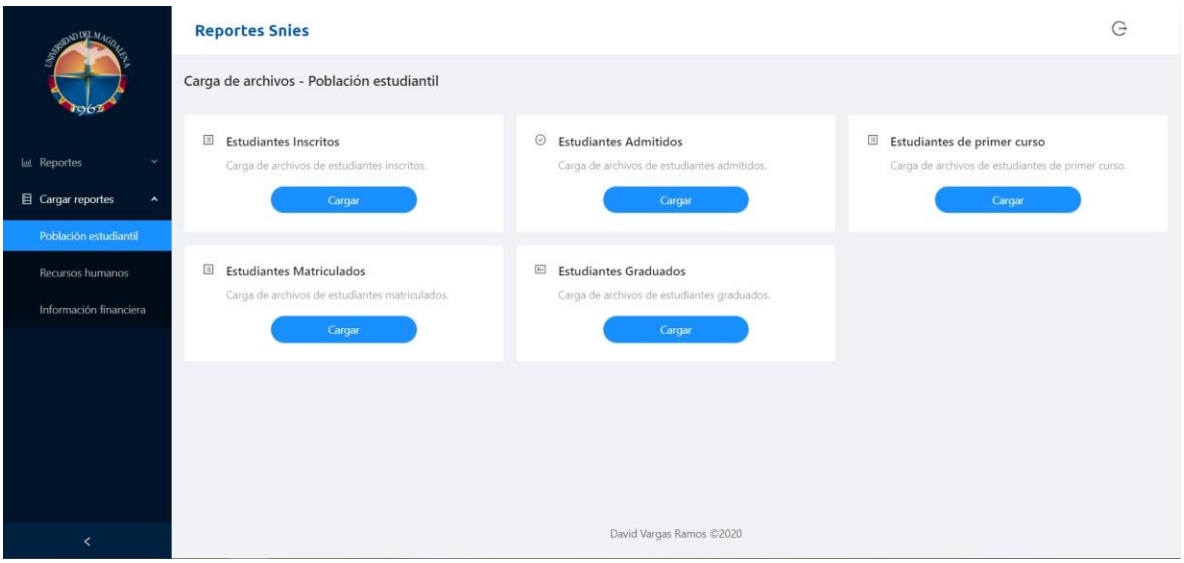


Ilustración 9. Cartas para la carga de datos de plantilla de población estudiantil.

En estas “cartas” también se incluyó un botón de carga que sirve para hacer desplegar un modal que explica la serie de pasos a seguir para poder subir correctamente las plantillas (Ilustración 10).

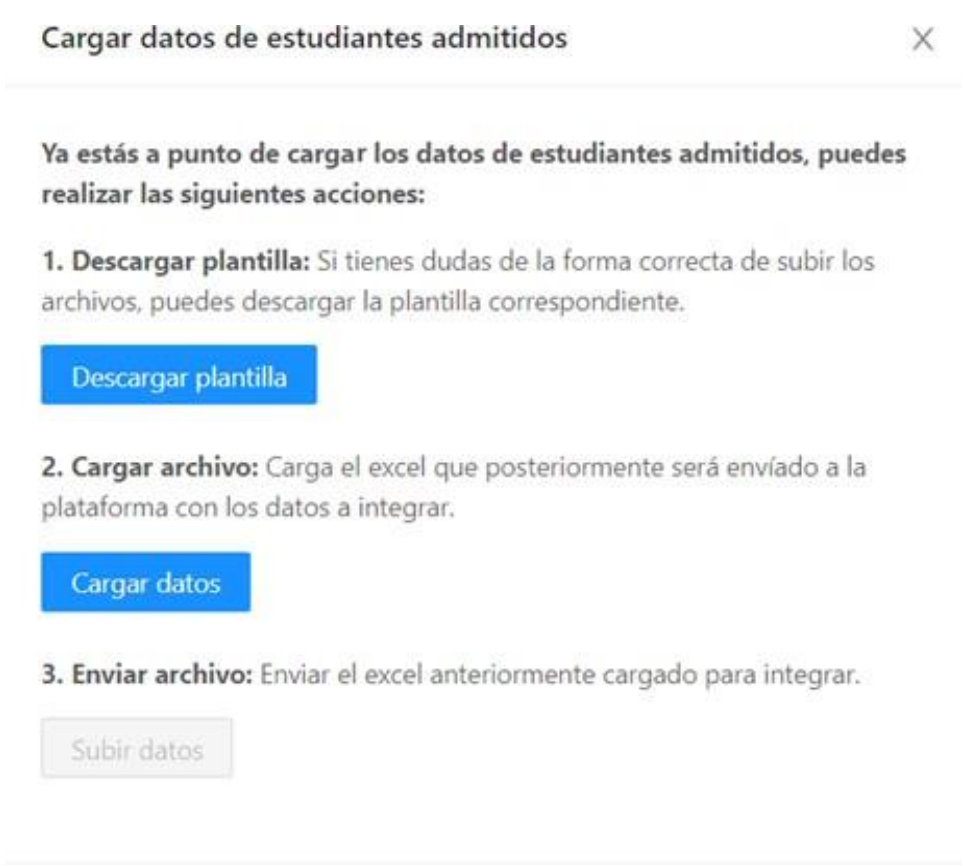


Ilustración 10. Pasos a seguir para la carga de datos.

Arquitectura de reportes gráficos

Para el buen funcionamiento de los reportes gráficos se implementó la arquitectura que se observa en la Ilustración 11.

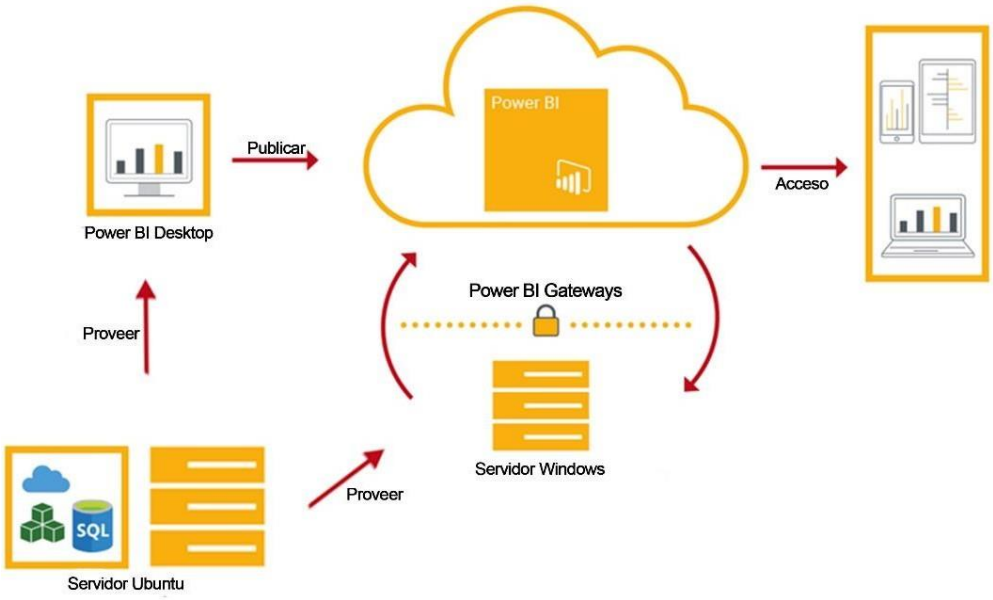


Ilustración 11. Arquitectura de manejo de diferentes herramientas para la visualización de los datos.

La bodega de datos está alojada en un servidor Ubuntu, esto hace posible que se pueda acceder a él (siempre que se tengan las credenciales) desde cualquier cliente SQL.

Power BI Desktop se conecta a la bodega y esta le provee todos los datos necesarios para la realización de los informes gráficos.

Una vez los informes gráficos estén realizados, es necesario publicar las gráficas en el servicio en la nube de *Power BI*, que permite generar una URL *Iframe* para posteriormente ser publicado en un aplicativo y así cualquier dispositivo tenga acceso a las gráficas.

Por último, era necesario que los nuevos registros que se ingresen en la bodega de datos fuesen obtenidos automáticamente por *Power BI* y así este pudiese actualizar automáticamente los reportes gráficos.

Para esto, se hizo uso de una herramienta llamada *Power BI Gateways* que crea un enlace entre la bodega de datos y *Power BI*, esta herramienta debe estar corriendo en un servidor, no obstante, únicamente está disponible para Windows, por ello, se obtuvo un servidor Windows únicamente para este cometido.

Cubos OLAP y reportes gráficos

Una vez enviadas las diferentes plantillas al servidor, y de que este hiciera todos los procedimientos para el envío de dichos datos a la bodega, se conectó la bodega de datos a *Power BI* para generar las diferentes gráficas según la información disponible en cada conjunto de datos.

Power BI por defecto convierte la estructura de datos a la cual se está accediendo en un cubo OLAP que permite posteriormente realizar reportes gráficos con él.

Al conectar la bodega de datos a *Power BI*, y crear éste los cubos, presenta el modelado de datos de la siguiente manera:

Modelo de estudiantes

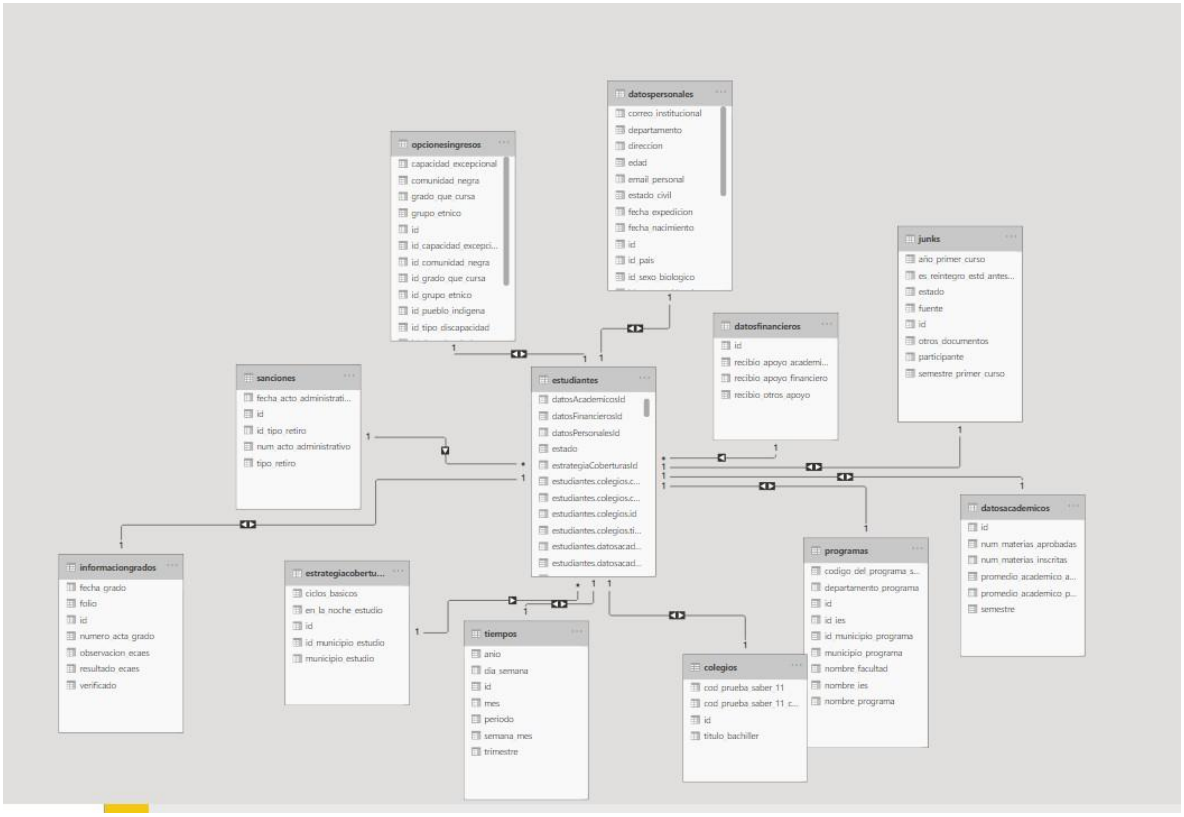


Ilustración 12. Modelo de estudiantes en Power BI.

Modelo de docentes

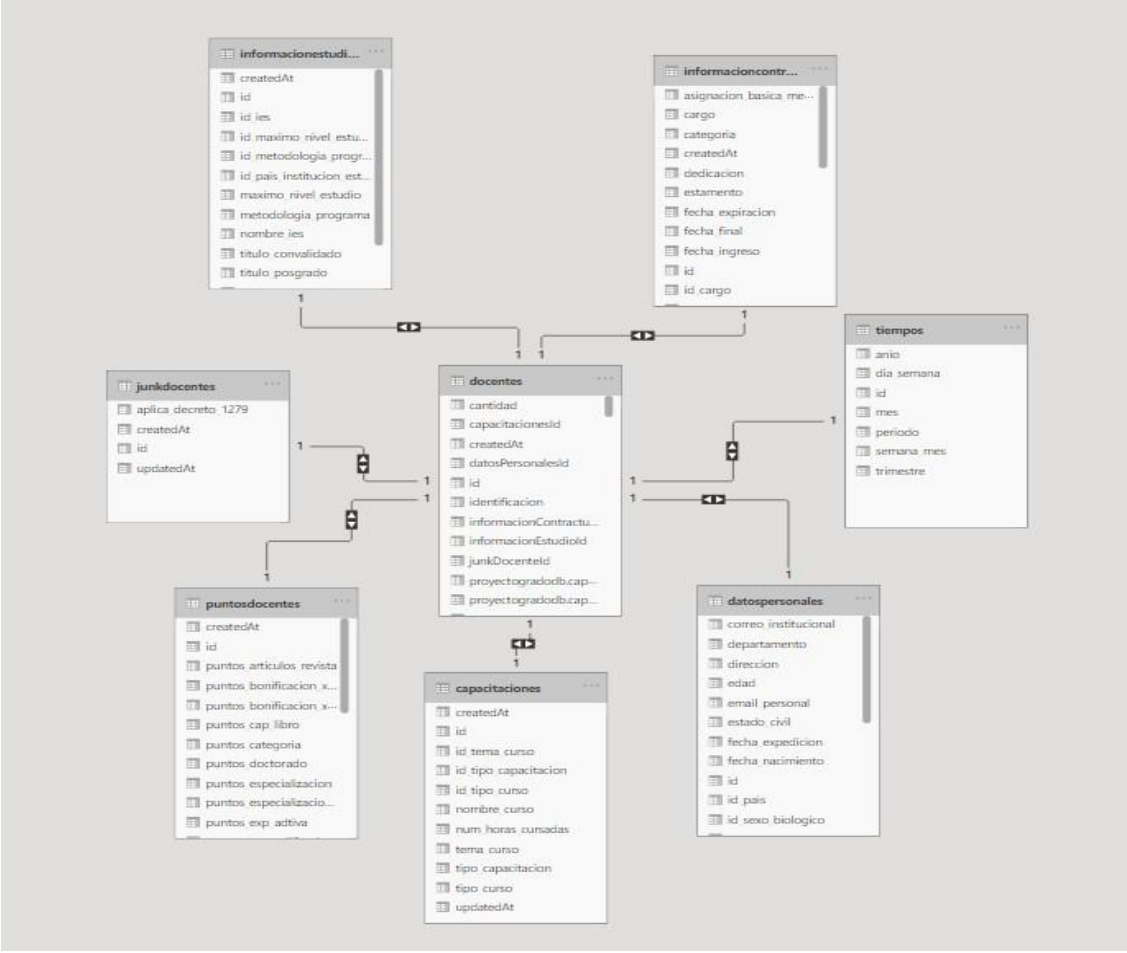


Ilustración 13. Modelo de docentes en Power BI.

Modelo de información ejecución presupuestal

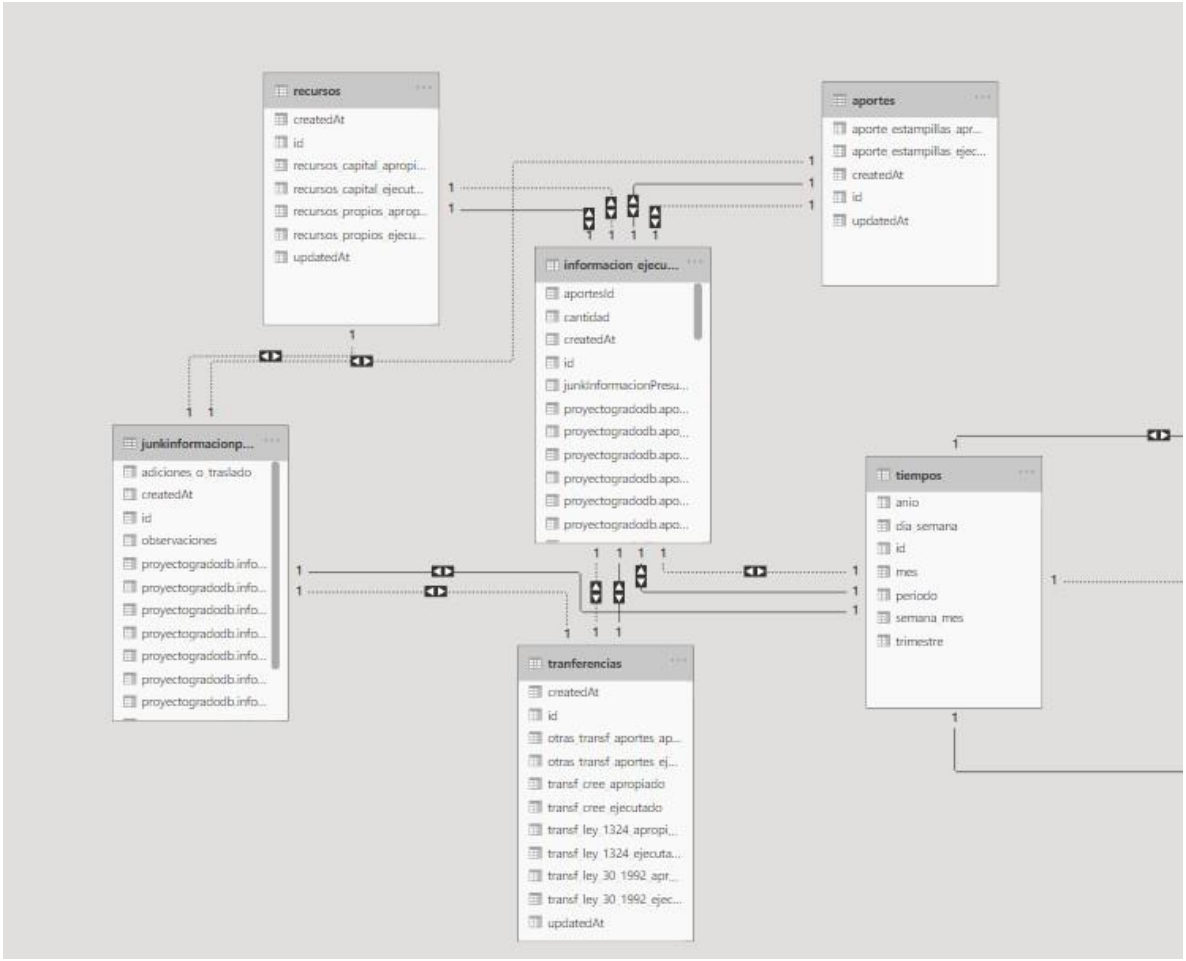


Ilustración 14. Modelo de información ejecución presupuestal en Power BI.

Modelo de balance generales

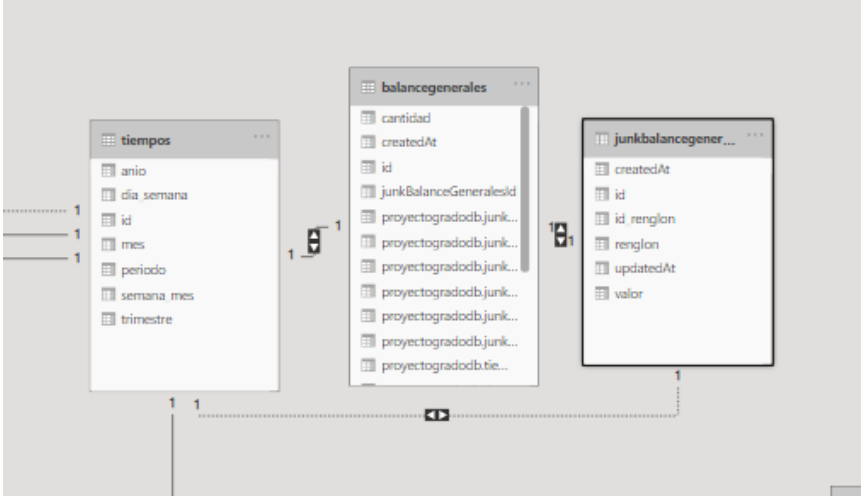


Ilustración 15. Modelo de balance generales en Power BI.

Modelo de información presupuestal

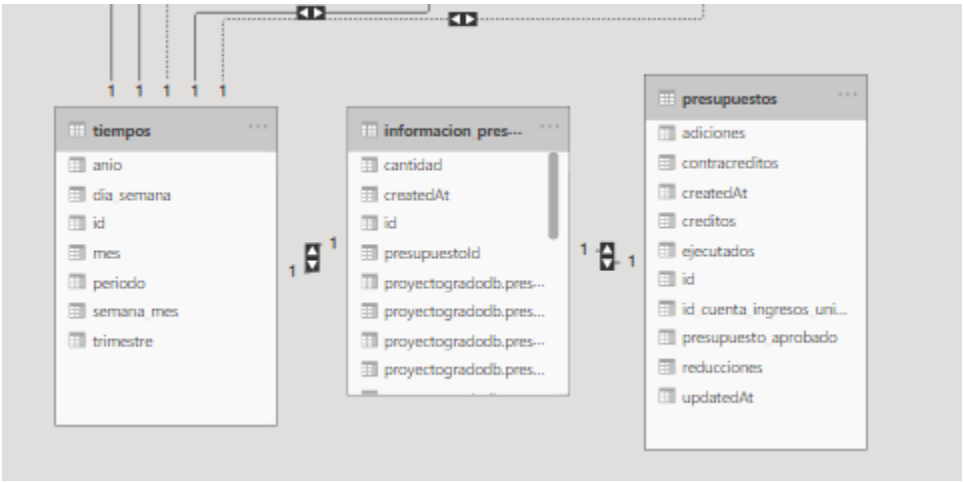


Ilustración 16. Modelo de información presupuestal Power BI.

Una vez creados los modelos, *Power BI* permite generar gráficas por medio de la información almacenada. Las gráficas generadas en *Power BI* pueden ser exportadas en un formato *Iframe*, formato que puede incluirse en cualquier documento HTML para su visualización (Ilustración 17).

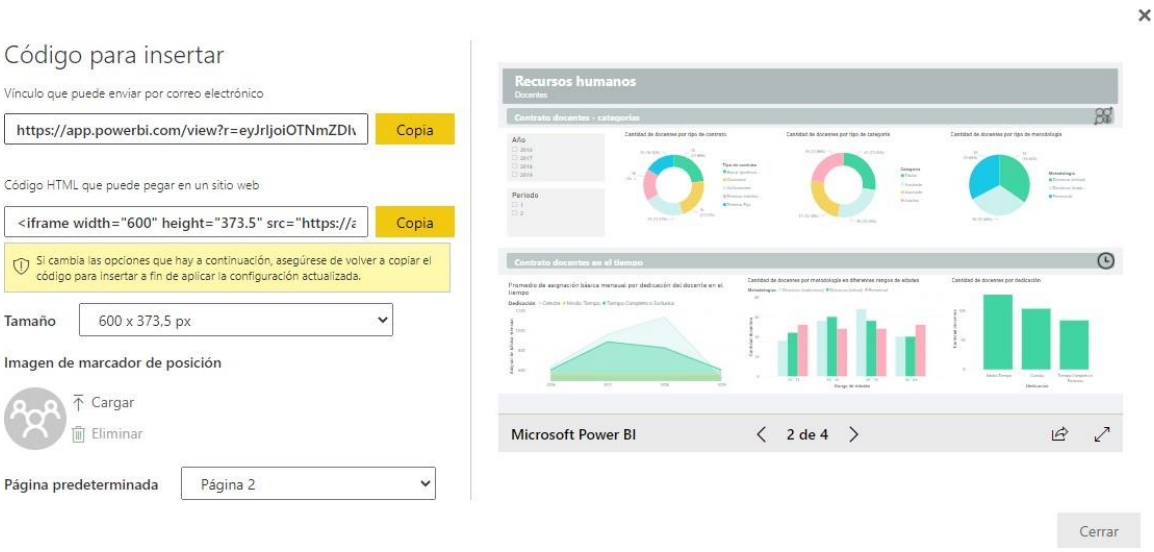


Ilustración 17. Generador de código iframe.

En ese orden de ideas, cada conjunto de datos tiene su propia sección gráfica en la cual contiene la etiqueta *Iframe* generada para cada uno de ellos (Ilustración 18 e Ilustración 19).

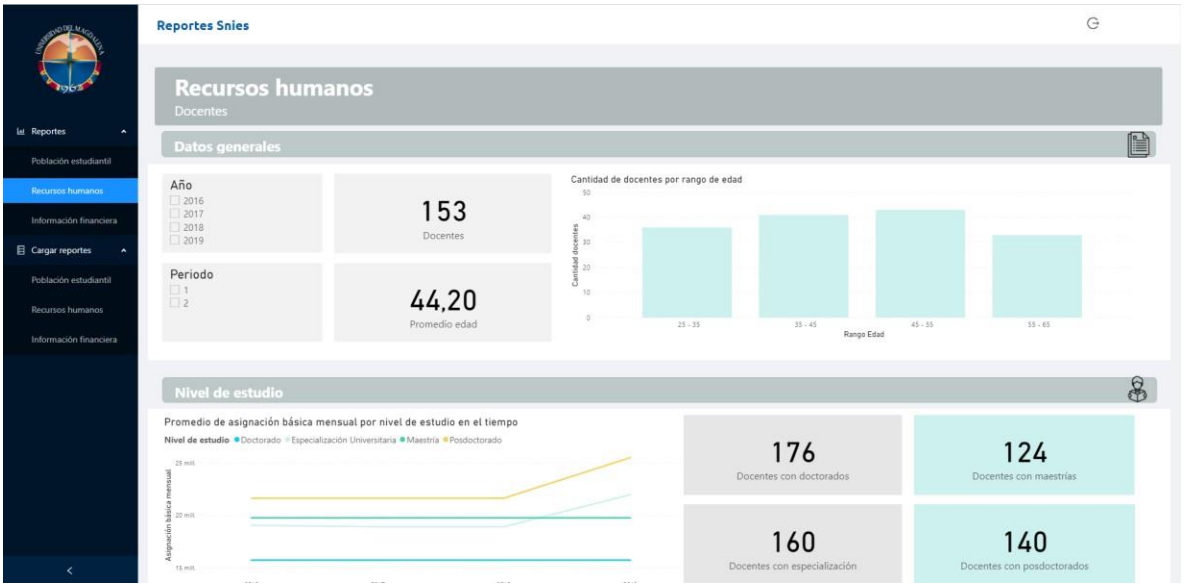


Ilustración 18. Visualización de gráficos de recursos humanos en plataforma web.

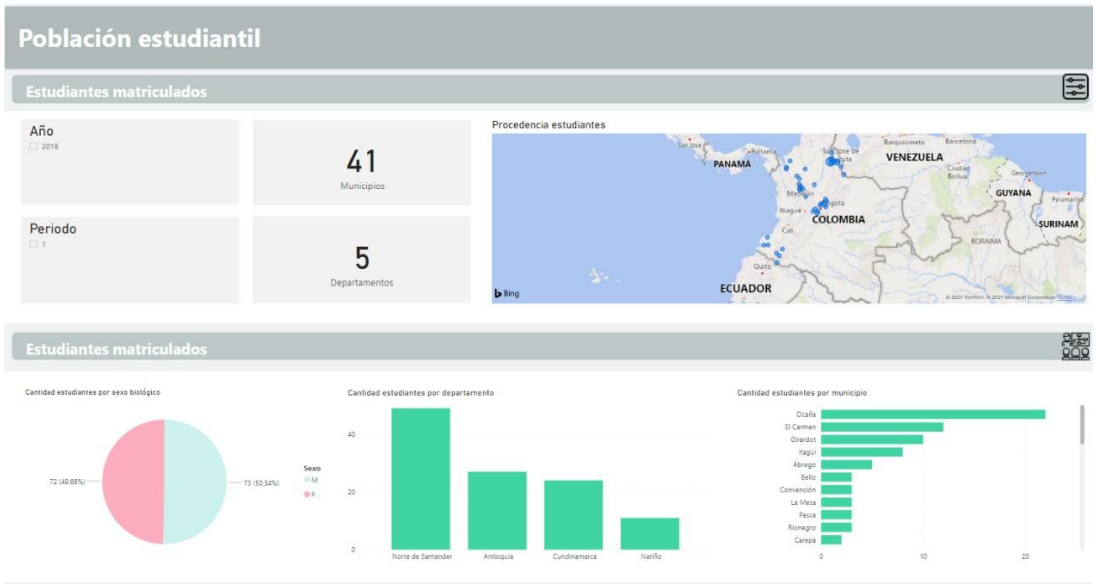


Ilustración 19. Visualización de gráficos de población estudiantil en plataforma web.

Finalmente, como se mencionó anteriormente, se implementó la herramienta *Gateways de Power BI* para la actualización automática de los reportes gráficos, no obstante, esta actualización no se realiza cada vez que se ingresan datos, sino que se deben programar. La programación para la actualización de los datos fue definida de la siguiente manera.

Ilustración 20. Programación de actualización de información en gráficos.

Presupuesto

Tabla 3. Presupuesto.

RUBROS GENERALES	UNIMAGDALENA		TOTAL RUBRO
	EFECTIVO	CAPACIDAD INSTALADA	
Equipos y Accesorios			
Materiales e Insumos			
Software	$ 1.000.000		$ 1.000.000
Servicios técnicos			
Bibliografía	$ 255.000		$ 255.000
Capacitaciones	$ 2.745.000		$ 2.745.000
Salidas de Campo			
TOTAL			$ 4.000.000

Cronograma

Tabla 4. Cronograma.

	2020					
Actividades	**Mar**	**Abr**	**May**	**Jun**	**Jul**	**Ago**
Diseño e implementación de bodega de datos						
Recreación de gráficas estadísticas presente en Unimagdalena en Cifras						
-Implementación Api Rest, para el consumo de datos presente en la bodega de datos						
Diseño plataforma web						
Implementación plataforma web						
Implementación de gráficas estadísticas en la plataforma web						
Construcción dashboard para la generación de reportes gráficos de indicadores de gestión						

Conclusiones

La culminación satisfactoria del proyecto permite tener la primera versión de un repositorio centralizado que utiliza la información compartida por la Universidad del Magdalena a través del SNIES al Ministerio de Educación Nacional de manera que pueda ser analizada desde diferentes perspectivas y llevar un registro histórico de toda esta información. Este repositorio permite integrar todas las fuentes, archivos en Excel, que eran utilizados tradicionalmente para llevar el registro de esta información.

La plataforma construida permite la gestión no solo de los datos de las diferentes fuentes para un solo análisis, sino que posibilita que se lleve un registro histórico de toda la información de manera que se puedan realizar análisis temporales. También, la plataforma posibilita la gestión de los usuarios que acceden a las diferentes funcionalidades de la plataforma de forma que se segregan las responsabilidades entre quienes son las personas responsables de subir la información y aquellas que pueden aprovecharla a través de la visualización.

Los procesos de extracción, trasformación y carga fueron construidos utilizando una estrategia de *lazy loading* donde los datos son cargados en segundo plano y a petición y disponibilidad de recursos de manera que no se bloquee la interacción de la plataforma cuando se están cargando los datos. De igual forma, cuando los datos son cargados se genera una actualización automática de todo el contenido que depende de los datos, como son los reportes.

El componente de visualización de construyó completamente en *Power BI*, una de las herramientas de mayor aceptación dentro del mercado de análisis de datos lo que permitió aprenderla y así expandir las posibilidades de entrada al mercado la laboral

La ejecución del proyecto de igual manera permitió la asimilación de nuevos conceptos relacionados con lo que es la inteligencia de negocios, la utilidad de las bodegas de datos y la diferencia con las bases de datos transaccionales. Al igual que conocer una nueva herramienta de amplia aceptación en el mercado como es *Power BI*, posibilitando así complementar mi visión más allá del desarrollo web que es campo donde tradicionalmente se tenía mayor fortaleza.

Más allá de la importancia asociada a este primer intento de incorporar la inteligencia de negocios dentro de los procesos de la Universidad del Magdalena, no se debe perder de mente que solo es un piolo que buscaba mostrar los beneficios que las mismas pueden acarrear, pero aún queda mucho camino por recorrer. A continuación, se mencionan solo algunos de estos trabajos futuros:

- Automatizar todo el proceso de carga de datos, no leyendo los mismos desde las fuentes secundarias como los archivos Excel que actualmente deben ser cargados, sino leyendo la información desde los mismos sistemas de información en el nivel de granularidad más bajo posible.
- Extender los datos que actualmente son cargados más allá de una necesidad de datos en particulares, llevarlos al nivel de procesos organizativos que poder ir siendo incorporados de manera incremental posibilitando así el cruce de información.

Para cerrar, este proyecto permitió organizar de una mejor manera los reportes enviados al SNIES, convertir esas plantillas usadas por la oficina de planeación en reportes gráficos útiles.

Referencias

[1] Decreto 1767 de junio 2 de 2006
[2] SNIES - Ministerio de Educación Nacional de Colombia (mineducacion.gov.co)
[3] R. Kimball, y M. Ross, The data warehouse toolkit: the complete guide to dimensional modeling, 2nd ed. United States of America, Wiley Computer Publishing, 2002.
[4] R. Kimball y J. Caserta,The data warehouse ETL toolkit: practical techniques for extracting, cleaning, conforming, and delivering data. United States of America, Wiley Publishing, 2004.
[5] F. Dávila, La inteligencia del negocio (business intelligence). Bogotá: Ediciones Politécnico Grancolombiano. 2006.
[6] Ministerio de educación nacional, Resolución 19591 del 2017, 27 septiembre del 2017.
[7] J. H. Watson, Tutorial: Businnes Inteligence -Past, Present, and Future. Communications of the Association for Information System.
[8] S. Españeida, Inteligencia de negocios – definición.
[9] Business Intelligence Roadmap: The Complete Project Lifecycle for Decision Support Applications. Publisher: Addison Wesley Pub Date.
[10] Business intelligence: The savvy manager's guide.
[11] The profit Impact of Business Intelligence.
[12] Building a Data Warehouse: With Examples in SQL Server.
[13] Spotless Plataform, Exploring Data Warehouses and Data Quality, 2019. <https://web.archive.org/web/20180726071809/https://spotlessdata.com/blog/exploring-data-warehousesand-data-quality>
[14] M. Tammeraja, What is OLAP – On-Line Analytical Processing? 1997
[15] R. Stackowiak, J. R. Rayman Greenewald. Oracle Data Warehousing and Business Intelligence Solutions, 2007.
[16] I. Witten, E. Frank, Data mining practical machine learning tools and techniques with java implementations, Academic Press, San Francisco, United States of America, 2000.
[17] J. Wang, Encyclopedia of Data Warehousing and Mining, 2009.
[18] A. Flores, Uso de triggers y snapshots como técnica incremental en el proceso de extracción, transformación y transferencia de datos (ETT) en un Data Warehouse, Junio 2003.
[19] R. Kimball, y M. Ross, The data warehouse toolkit: the complete guide to dimensional modeling, 2nd ed. United States of America, Wiley Computer Publishing, 2002. (3 in other)
[20] S. Williams y N. Williams, The Business Value of Business Intelligence. Business Intelligence Journal. 2003.

[21] R. Kimball y J. Caserta, The data warehouse ETL toolkit: practical techniques for extracting, cleaning, conforming, and delivering data. United States of America, Wiley Publishing, 2004. (4 in other)
[22] D. Darmawikarta, Dimensional Data Warehousing with MySQL: A Tutorial. Brainy Software Corp, 2007.
[23] R. Bouman y J. V Dongen, Pentaho® Solutions: Business Intelligence and Data Warehousing with Pentaho and MySQL. Wiley Publishing, Inc. Indianapolis, Indiana, 2009.
[24] E. F. Codd, S. B. Codd, C. T. Salley, Providing OLAP (On-Line Analytical Procesing) to User-Analysts: An IT Mandate, 1993.
[25] A.Silberschatz Korth H., S. Sudarshan, Fundamentos de Base de Datos, 2006
[26] J. Peña Y J. Suárez, (2005) “Utilización de información histórica para decisiones empresariales”
[27] Power BI <<https://powerbi.microsoft.com>>
[28] React JS <<https://developer.mozilla.org/es>>
[29] React JS <<https://es.reactjs.org>>
[30] Sequelize <<https://sequelize.org>>
[31] J. A. Hernández, A. F. Mora, Sistema de bodega de datos para la administración de indicadores del Sistema de Universidades Estatales (SUE) soportado por un sistema distribuido, para la Universidad Francisco José de Caldas. Vol. 8, 2011.
[32] A. A. Rosado Gómez, Consolidación de indicadores institucionales utilizando bodega de datos, vol. 11, no. 1, 2016.
[33] M. Purificadora, Y G. Gonzales, (2017) “Diseño de un prototipo de Data Warehouse para la toma de decisiones en el sector fiduciario utilizando la información histórica suministrada por el aplicativo SIFI”
[34] A. Pineda, Y W. Torres, (2011) “Prototipo de una bodega de datos para la empresa salón del automóvil”
[35] J. Peña Y J. Suárez, (2005) “Utilización de información histórica para decisiones empresariales”
[36] J. Barboza, (2020) “Diseño de un cubo OLAP para la información contable generada por el supermarket ‘La competencia’” del cantón Naranjito
[37] C. Basurto, (2010) “Análisis, diseño e implementación de una solución de inteligencia de negocios para el área de finanzas de la municipalidad metropolitana de lima”
[38] Eafit en cifras. <<http://entrenos.eafit.edu.co/noticias/2019/mayo/Paginas/eafit-en-cifras-la-nuevaplataforma-para-consultar-informacion-estadistica-de-la-universidad.aspx>>
[39] A. A. Rosado Gómez, D. W. Rico Bautista, Inteligencia de negocios: Estado del Arte, no. 44, abril 2010.