

Proyecto de Bases de datos para el desarrollo de la plataforma Brujula Uni: Sistema de análisis comparativo de universidades en Bogotá para estudios en ciencias de la educación

Universidad Central
Maestría en analítica de datos
Curso de Bases de Datos
Bogotá, Colombia

{¹mfandinom@ucentral.edu.co, ²emarquezb@ucentral.edu.co, ³rpintoo@ucentral.edu.co}

November 17, 2023

Contents

| | | |
|----------|---|----------|
| 1 | Introducción (Max 250 Palabras) - (<i>Primera entrega</i>) | 3 |
| 2 | Características del proyecto de investigación que hace uso de Bases de Datos (Max 500 Palabras) - (<i>Primera entrega</i>) | 3 |
| 2.1 | Titulo del proyecto de investigación (Max 100 Palabras) - (<i>Primera entrega</i>) | 4 |
| 2.2 | Objetivo general (Max 100 Palabras) - (<i>Primera entrega</i>) | 4 |
| 2.2.1 | Objetivos especificos (Max 100 Palabras) - (<i>Primera entrega</i>) | 4 |
| 2.3 | Alcance (Max 200 Palabras) - (<i>Primera entrega</i>) | 4 |
| 2.4 | Pregunta de investigación (Max 100 Palabras) - (<i>Primera entrega</i>) . | 5 |
| 2.5 | Hipotesis (Max 100 Palabras) - (<i>Primera entrega</i>) | 5 |
| 3 | Reflexiones sobre el origen de datos e información (Max 400 Palabras) - (<i>Primera entrega</i>) | 6 |
| 3.1 | ¿Cual es el origen de los datos e información ? (Max 100 Palabras) - (<i>Primera entrega</i>) | 6 |
| 3.2 | ¿Cuales son las consideraciones legales o eticas del uso de la información? (Max 100 Palabras) - (<i>Primera entrega</i>) | 6 |
| 3.3 | ¿Cuales son los retos de la información y los datos que utilizara en la base de datos en terminos de la calidad y la consolidación? (Max 100 Palabras) - (<i>Primera entrega</i>) | 7 |
| 3.4 | ¿Que espera de la utilización de un sistema de Bases de Datos para su proyecto? (Max 100 Palabras) - (<i>Primera entrega</i>) | 7 |

| | | |
|----------|---|-----------|
| 4 | Diseño del Modelo de Datos del SMBD (Sistema Manejador de Bases de Datos)(Primera entrega) | 8 |
| 4.1 | Características del SMBD (Sistema Manejador de Bases de Datos) para el proyecto (Primera entrega) | 8 |
| 4.2 | Diagrama modelo de datos (Primera entrega) | 9 |
| 4.3 | Imágenes de la Base de Datos (Primera entrega) | 9 |
| 4.4 | Código SQL - lenguaje de definición de datos (DDL) (Primera entrega) | 11 |
| 4.5 | Código SQL - Manipulación de datos (DML) (Primera entrega) . . | 12 |
| 4.6 | Código SQL + Resultados: Vistas (Primera entrega) | 12 |
| 4.7 | Código SQL + Resultados: Triggers (Primera entrega) | 12 |
| 4.8 | Código SQL + Resultados: Funciones (Primera entrega) | 13 |
| 4.9 | Código SQL + Resultados: procedimientos almacenados (Primera entrega) | 13 |
| 5 | Bases de Datos No-SQL (Segunda entrega) | 15 |
| 5.1 | Diagrama Bases de Datos No-SQL (Segunda entrega) | 15 |
| 5.2 | SMBD utilizado para la Base de Datos No-SQL (Segunda entrega) | 17 |
| 6 | Aplicación de ETL (Extract, Transform, Load) y Bodega de Datos (Tercera entrega) | 18 |
| 6.1 | Ejemplo de aplicación de ETL y Bodega de Datos (Tercera entrega) | 18 |
| 6.2 | Automatización de Datos (Tercera entrega) | 18 |
| 6.3 | Integración de Datos (Tercera entrega) | 18 |
| 7 | Proximos pasos (Tercera entrega) | 19 |
| 8 | Lecciones aprendidas (Tercera entrega) | 20 |
| 9 | Bibliografía | 21 |

1 Introducción (Max 250 Palabras) - (Primera entrega)

En un mundo globalizado y cada vez más competitivo, la educación superior desempeña un papel esencial en la formación de profesionales calificados que se enfrentarán a los nuevos retos del mercado laboral. En palabras de Schultz (1960) en el desarrollo de la teoría del capital humano “al invertir en sí mismos, los seres humanos aumentan el campo de sus posibilidades. Es un camino por el cual los hombres pueden aumentar su bienestar”

Desde esta perspectiva, es necesario un enfoque que permita analizar cuál es la mejor decisión al momento de escoger en la oferta educativa que se encuentra en el país según el contexto de los interesados.

En Colombia, la prueba Saber Pro se ha convertido en una herramienta que nos da información para medir el rendimiento de los estudiantes que están por graduarse y la calidad de la educación superior. Tal como indica la página del icfes: “El examen de Estado de la Calidad de la Educación Superior, Saber Pro, es un instrumento de evaluación estandarizada para la medición externa de la calidad de la educación superior que evalúa las competencias de los estudiantes que están próximos a culminar los distintos programas profesionales.” (Acerca del examen Saber Pro, s/f)

En este proyecto, el objetivo es analizar y comparar el desempeño de las universidades que ofrecen las carreras de educación, inicialmente en Bogotá, utilizando como indicadores principales, los resultados de las pruebas Saber Pro entre 2018 a 2022, el número de inscripciones a primer semestre de las licenciaturas en el año 2022 y los aspectos relacionados con los precios para entender cómo estas variables influyen en la elección de los estudiantes.

También se busca proporcionar una herramienta valiosa para los futuros estudiantes de licenciatura, los cuales podrán tomar una decisión informada al evaluar tanto el rendimiento académico, como los aspectos prácticos al escoger universidad. A su vez, el proporcionar una evaluación comparativa, contribuirá a mejorar la calidad de la educación superior en Colombia e impulsará a las universidades para ofrecer programas académicos con un balance idóneo en relación precio-calidad.

2 Características del proyecto de investigación que hace uso de Bases de Datos (Max 500 Palabras) - (Primera entrega)

El presente proyecto de investigación, “Brújula Uni: Sistema comparativo de análisis de universidades en Bogotá para estudios en ciencias de la educación” se enfoca en el desarrollo de un comparador de universidades por carrera, para el alcance de este proyecto, se centrarán las bases en todas las carreras de ciencias de la educación, de la ciudad de Bogotá.

Su objetivo principal es proporcionar a los estudiantes una herramienta robusta y de fácil acceso que les permita evaluar de manera efectiva las opciones disponibles de carrera en el área de las ciencias de la educación en la ciudad de Bogotá. Esto se dará en función de una variedad de factores clave, que van desde

el desempeño académico de las universidades hasta detalles específicos de los programas de estudio y la popularidad de los programas medido en inscripciones.

2.1 Título del proyecto de investigación (Max 100 Palabras) - (Primera entrega)

Brujula Uni: Sistema para análisis comparativo de universidades en Bogotá para estudios en ciencias de educación

2.2 Objetivo general (Max 100 Palabras) - (Primera entrega)

Desarrollar un sistema de gestión de bases de datos (DBMS) que establezca la estructura de información necesaria para crear una herramienta que realice el análisis y la comparación del desempeño de las instituciones de educación superior que ofrecen carreras de licenciaturas.

2.2.1 Objetivos específicos (Max 100 Palabras) - (Primera entrega)

- Diseñar una base de datos que almacene y gestione la información relevante de los resultados de las pruebas saber pro, numero de inscripciones y costos de semestre sobre las instituciones de educación superior de bogotá enfocados en carreras de ciencias de la educación.
- Desarrollar un algoritmo de análisis de datos que proporcione a los usuarios recomendaciones personalizadas basadas en sus preferencias y necesidades académicas, ayudándoles a tomar decisiones informadas.
- Encontrar tendencias a través de la búsqueda, filtrado y procesamiento de los datos, permitiendo dar información relevante al análisis general por universidad y por licenciatura dentro de las ciencias de la educación

2.3 Alcance (Max 200 Palabras) - (Primera entrega)

El proyecto empezará con el diseño de un sistema de gestión de bases de datos, el cual almacenará información de las universidades que dictan las carreras de licenciatura en la ciudad de Bogotá, en este se tendrán en cuenta los resultados de las pruebas Saber Pro entre 2018 y 2022, el número de inscripciones a primer semestre en el año 2022 y los aspectos relacionados con los precios de matrícula.

A partir de allí, se realizarán diversas consultas que permitan generar un análisis de los datos macro de la información para encontrar tendencias entre las variables de las diferentes fuentes de información.

Luego se espera desarrollar una herramienta de consulta personalizada por usuario, el cual permitirá a los aspirantes, dependiendo de sus requerimientos, dar las opciones de universidad y carrera dentro de las licenciaturas que satisfaga sus necesidades a través de algoritmos de clasificación o recomendación.

2.4 Pregunta de investigación (Max 100 Palabras) - (*Primera entrega*)

¿Cuál es la estructura de información óptima que permite realizar análisis y comparaciones de desempeño y algoritmos de clasificación entre instituciones de educación superior?

2.5 Hipotesis (Max 100 Palabras) - (*Primera entrega*)

La categorización eficaz de la información en dimensiones críticas, como calidad académica, costos y accesibilidad, contribuirá a una estructura de información óptima.

Esta estructura, diseñada de manera comprensible para diversos usuarios, como estudiantes, docentes y responsables de políticas educativas, facilitará la realización de comparaciones entre instituciones de educación superior que impartan carreras de ciencias de la educación en bogotá y permitirá generar algoritmos de recomendación según las necesidades de los stakeholders.

3 Reflexiones sobre el origen de datos e información (Max 400 Palabras) - (*Primera entrega*)

Para desarrollar este proyecto, se recurrió a diversas fuentes de información de acceso público, proporcionadas por distintas entidades de gobierno.

Inicialmente, se empleó la base de datos con los resultados anonimizados de la prueba saber pro del periodo comprendido entre los años 2018 a 2022. También, se utilizó la base de datos que contiene a los estudiantes matriculados en educación superior para Colombia en 2022, esta base es suministrada por el ministerio de educación nacional del país. Finalmente, se incorporó la base de datos del SNIES que proporciona datos sobre los programas académicos ofrecidos en el área de ciencias de la educación en el país.

Sobre todas las bases de datos, se realizó un filtro a la información con los programas profesionales de ciencias de educación que se encuentren en la ciudad de Bogotá.

Es importante destacar que durante el proceso de análisis, se identificaron datos faltantes en algunos registros, particularmente en los resultados de las pruebas Saber Pro y en la información de los costos de matrícula de algunas universidades.

3.1 ¿Cual es el origen de los datos e información ? (Max 100 Palabras) - (*Primera entrega*)

Las bases de datos fueron recopiladas de las siguientes fuentes de información:

- Datos abiertos:

Según el Ministerio de Tecnologías de la Información y las Comunicaciones (MinTIC), es información pública dispuesta en formatos que permiten su uso y reutilización en formatos que permiten su uso y reutilización bajo licencia abierta y sin restricciones legales para su aprovechamiento. (MinTIC, [2021])

- SNIES:

Según el Ministerio de Educación, el Sistema Nacional de Información de la Educación Superior (SNIES), es un sistema de información que ha sido creado para responder a las necesidades de información de la educación superior en Colombia.

En este sistema se recopila y organiza la información relevante sobre la educación superior que permite hacer planeación, monitoreo, evaluación, asesoría, inspección y vigilancia del sector. (Ministerio de Educación Nacional, [2021])

3.2 ¿Cuales son las consideraciones legales o eticas del uso de la información? (Max 100 Palabras) - (*Primera entrega*)

Los datos abiertos son información pública que se puede reutilizar bajo licencia abierta y sin restricciones legales en Colombia. Esta regulación está respaldada

por la Ley 1712 de 2014, la cual establece la obligación de las entidades públicas de poner a disposición datos y define los datos abiertos en su numeral sexto como: "Todos aquellos datos primarios o sin procesar, que se encuentran en formatos estándar e interoperables que facilitan su acceso y reutilización, los cuales están bajo la custodia de las entidades públicas o privadas que cumplen con funciones públicas y que son puestos a disposición de cualquier ciudadano, de forma libre y sin restricciones, con el fin de que terceros puedan reutilizarlos y crear servicios derivados de los mismos". (Ley 1712 de 2014, Colombia)

3.3 ¿Cuales son los retos de la información y los datos que utilizara en la base de datos en terminos de la calidad y la consolidación? (Max 100 Palabras) - (Primera entrega)

Las 3 bases de datos fueron obtenidas en formato CSV "Comma Separated Values" para su reutilización, en la base de datos de resultados de la prueba saber pro, hay registros vacios de algunos estudiantes y en la base de datos del SNIES hay campos vacios con respecto al precio de matricula, por lo cual se requiere plantear un metodo para completar la información o manejar los resultados con los datos faltantes.

3.4 ¿Que espera de la utilización de un sistema de Bases de Datos para su proyecto? (Max 100 Palabras) - (Primera entrega)

Para este proyecto, se espera que el sistema de gestión de Bases de Datos nos de la información ordenada y facilite la gestión, análisis y presentación de datos de manera efectiva.

Permitirá también la integración de datos de las diferentes bases de datos, lo que es esencial para realizar análisis y comparaciones significativas entre los diferentes conjuntos de datos. Desde ahí, poder desarrollar algoritmos de recomendación o clasificación según las necesidades de los estudiantes.

Esta DBSM formará la estructura de información necesaria para crear una herramienta que cumpla con el análisis y la comparación del desempeño de las instituciones de educación superior y evaluará los factores más relevantes a la hora de la toma de decisiones para la escogencia de universidad.

4 Diseño del Modelo de Datos del SMBD (Sistema Manejador de Bases de Datos) *(Primera entrega)*

4.1 Características del SMBD (Sistema Manejador de Bases de Datos) para el proyecto *(Primera entrega)*

Para el desarrollo de este proyecto, se utilizó MySQL para el diseño de la SMBD, el almacenamiento y tratamiento de la información:

MySQL es un sistema de administración de bases de datos relacionales. Es un software de código abierto desarrollado por Oracle. Se considera como la base de datos de código abierto más utilizada en el mundo (Londoño, 2023)

Dentro de las ventajas competitivas que tiene este sistema de bases de datos tenemos:

- Código abierto: La flexibilidad que ofrece su carácter de código abierto es una gran ventaja de MySQL, además de ser gratuita y fácil de usar.
- Facilidad de uso: MySQL es fácil de configurar y requiere un ajuste mínimo para lograr excelentes niveles de rendimiento. Herramientas de interfaz gráfica de terceros como, MySQL WorkBench y dbForge Studio, hacen que MySQL sea aún más simple para empezar a usar la base de datos, lo que es un mano a mano para principiantes.
- Compatibilidad: Hoy, MySQL ofrece compatibilidad con la mayoría de las principales plataformas informáticas, como Linux, macOS, Microsoft Windows y Ubuntu. Además, proporciona un alto rendimiento para el almacenamiento de grandes volúmenes de datos o Business Intelligence. (MySQL: qué es y cuáles son sus beneficios, s/f)

4.2 Diagrama modelo de datos (Primera entrega)

Modelo lógico y relacional

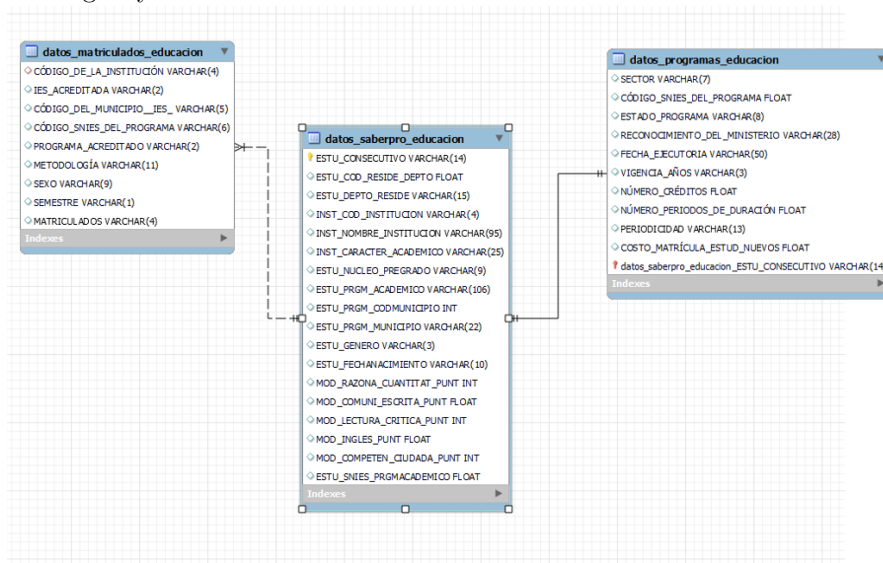


Figure 1: Datos Saber Pro

4.3 Imágenes de la Base de Datos (Primera entrega)

1 • SELECT * FROM DATOS_SABERPRO_EDUCACION;

| ESTU_CONSECUTIVO | ESTU_COD_RESIDE_DEPTO | ESTU_DEPTO_RESIDE | INST_COD_INSTITUCION | INST_NOMBRE_INSTITUCION | INST_CARACTER_ACADEMICO | ESTU_NUCLEO_PREGRADO |
|------------------|-----------------------|-------------------|----------------------|---|---------------------------|----------------------|
| EK201830001697 | 23 | CORDOBA | 1113 | UNIVERSIDAD DE CORDOBA-MONTERIA | UNIVERSIDAD | EDUCACIÓN |
| EK201830002147 | 18 | CAQUETA | 1115 | UNIVERSIDAD DE LA AMAZONIA-FLORENCIA | UNIVERSIDAD | EDUCACIÓN |
| EK201830002157 | 52 | NARIÑO | 1206 | UNIVERSIDAD DE NARIÑO-PASTO | UNIVERSIDAD | EDUCACIÓN |
| EK201830002160 | 18 | CAQUETA | 1115 | UNIVERSIDAD DE LA AMAZONIA-FLORENCIA | UNIVERSIDAD | EDUCACIÓN |
| EK201830002168 | 18 | CAQUETA | 1115 | UNIVERSIDAD DE LA AMAZONIA-FLORENCIA | UNIVERSIDAD | EDUCACIÓN |
| EK201830002174 | 18 | CAQUETA | 1115 | UNIVERSIDAD DE LA AMAZONIA-FLORENCIA | UNIVERSIDAD | EDUCACIÓN |
| EK201830002221 | 18 | CAQUETA | 1115 | UNIVERSIDAD DE LA AMAZONIA-FLORENCIA | UNIVERSIDAD | EDUCACIÓN |
| EK201830002233 | 18 | CAQUETA | 1115 | UNIVERSIDAD DE LA AMAZONIA-FLORENCIA | UNIVERSIDAD | EDUCACIÓN |
| EK201830002243 | 11 | BOGOTÁ | 1806 | UNIVERSIDAD LIBRE-BOGOTÁ D.C. | UNIVERSIDAD | EDUCACIÓN |
| EK201830002247 | 52 | NARIÑO | 1206 | UNIVERSIDAD DE NARIÑO-PASTO | UNIVERSIDAD | EDUCACIÓN |
| EK201830002253 | 50 | META | 1119 | UNIVERSIDAD DE LOS LLANOS-VILLAVICENCIO | UNIVERSIDAD | EDUCACIÓN |
| EK201830002261 | 11 | BOGOTÁ | 2710 | FUNDACION UNIVERSITARIA MONSERRATE... | INSTITUCIÓN UNIVERSITARIA | EDUCACIÓN |
| EK201830002295 | 50 | META | 1119 | UNIVERSIDAD DE LOS LLANOS-VILLAVICENCIO | UNIVERSIDAD | EDUCACIÓN |
| EK201830002315 | 11 | BOGOTÁ | 2710 | FUNDACION UNIVERSITARIA MONSERRATE... | INSTITUCIÓN UNIVERSITARIA | EDUCACIÓN |
| EK201830002317 | 18 | CAQUETA | 1115 | UNIVERSIDAD DE LA AMAZONIA-FLORENCIA | UNIVERSIDAD | EDUCACIÓN |
| EK201830002326 | 50 | META | 1119 | UNIVERSIDAD DE LOS LLANOS-VILLAVICENCIO | UNIVERSIDAD | EDUCACIÓN |
| EK201830002328 | 25 | CUKINAMARCA | 2710 | FUNDACION UNIVERSITARIA MONSERRATE... | INSTITUCIÓN UNIVERSITARIA | EDUCACIÓN |
| EK201830002335 | 18 | CAQUETA | 1115 | UNIVERSIDAD DE LA AMAZONIA-FLORENCIA | UNIVERSIDAD | EDUCACIÓN |
| EK201830002337 | 76 | VALLE | 1203 | UNIVERSIDAD DEL VALLE-CALI | UNIVERSIDAD | EDUCACIÓN |
| EK201830002346 | 50 | META | 1119 | UNIVERSIDAD DE LOS LLANOS-VILLAVICENCIO | UNIVERSIDAD | EDUCACIÓN |
| EK201830002357 | 18 | CAQUETA | 1115 | UNIVERSIDAD DE LA AMAZONIA-FLORENCIA | UNIVERSIDAD | EDUCACIÓN |
| EK201830002358 | 50 | META | 1119 | UNIVERSIDAD DE LOS LLANOS-VILLAVICENCIO | UNIVERSIDAD | EDUCACIÓN |

Figure 2: Datos Saberpro

CODIGO PROYECTO BASES. SQL File 3"

1 • SELECT * FROM DATOS_MATRICULADOS_EDUCACION;

Result Grid | Filter Rows | Export | Wrap Cell Contents | Fetch rows:

| CÓDIGO_DE_LA_INSTITUCIÓN | IES_AREDITADA | CÓDIGO_DEL_MUNICIPIO_IES | CÓDIGO_SNIES_DEL_PROGRAMA | PROGRAMA_AREDITADO | METODOLOGÍA | SEXO | SEMESTRE | MATRICULADOS |
|--------------------------|---------------|--------------------------|---------------------------|--------------------|-------------|-----------|----------|--------------|
| 1101 | SI | 11001 | 52746 | NO | Presencial | Masculino | 1 | 21 |
| 1101 | SI | 11001 | 52746 | NO | Presencial | Masculino | 2 | 33 |
| 1101 | SI | 11001 | 52746 | NO | Presencial | Femenino | 1 | 30 |
| 1101 | SI | 11001 | 52746 | NO | Presencial | Femenino | 2 | 32 |
| 1101 | SI | 11001 | 52978 | NO | Presencial | Masculino | 1 | 34 |
| 1101 | SI | 11001 | 52978 | NO | Presencial | Masculino | 2 | 29 |
| 1101 | SI | 11001 | 52978 | NO | Presencial | Femenino | 1 | 30 |
| 1101 | SI | 11001 | 52978 | NO | Presencial | Femenino | 2 | 30 |
| 1101 | SI | 11001 | 53264 | NO | Presencial | Masculino | 1 | 2 |
| 1101 | SI | 11001 | 53264 | NO | Presencial | Masculino | 2 | 6 |
| 1101 | SI | 11001 | 53264 | NO | Presencial | Femenino | 1 | 4 |
| 1101 | SI | 11001 | 53264 | NO | Presencial | Femenino | 2 | 5 |
| 1101 | SI | 11001 | 55136 | NO | Presencial | Masculino | 1 | 22 |
| 1101 | SI | 11001 | 55136 | NO | Presencial | Masculino | 2 | 23 |
| 1101 | SI | 11001 | 55136 | NO | Presencial | Femenino | 1 | 19 |
| 1101 | SI | 11001 | 55136 | NO | Presencial | Femenino | 2 | 17 |
| 1101 | SI | 11001 | 101795 | NO | Presencial | Masculino | 1 | 9 |
| 1101 | SI | 11001 | 101795 | NO | Presencial | Masculino | 2 | 10 |
| 1101 | SI | 11001 | 101795 | NO | Presencial | Femenino | 1 | 9 |
| 1101 | SI | 11001 | 101795 | NO | Presencial | Femenino | 2 | 8 |
| 1101 | SI | 11001 | 106672 | NO | Presencial | Masculino | 1 | 8 |
| 1101 | SI | 11001 | 106672 | NO | Presencial | Masculino | 2 | 14 |

Figure 3: Informacion Inscritos

CODIGO PROYECTO BASES. SQL File 3"

1 • SELECT * FROM DATOS_PROGRAMAS_EDUCACION;

Result Grid | Filter Rows | Export/Import | Wrap Cell Contents | Fetch rows:

| SECTOR | CÓDIGO_SNIES_DEL_PROGRAMA | ESTADO_PROGRAMA | RECONOCIMIENTO_DEL_PROGRAMA | FECHA_EXECUTORIA | VIGENCIA_AÑOS | NÚMERO_CREDITOS | NÚMERO_PERIODOS_DE_DURACIÓN | PERIODICIDAD | COSTO_MATRÍCULA_ESTUD_NUEVOS |
|---------|---------------------------|-----------------|------------------------------|---------------------|---------------|-----------------|-----------------------------|--------------|------------------------------|
| Oficial | 146 | Activo | Acreditación de alta calidad | 2023-03-03 00:00:00 | 6 | 170 | 10 | Semestral | 1284790 |
| Oficial | 147 | Activo | Acreditación de alta calidad | 2023-03-03 00:00:00 | 6 | 157 | 10 | Semestral | 1284790 |
| Oficial | 148 | Inactivo | 0 | 0 | 0 | 0 | 8 | Semestral | 0 |
| Oficial | 149 | Inactivo | 0 | 0 | 0 | 160 | 10 | Semestral | 0 |
| Oficial | 150 | Inactivo | Acreditación de alta calidad | 0 | 6 | 160 | 10 | Semestral | 0 |
| Oficial | 151 | Activo | Acreditación de alta calidad | 2023-03-03 00:00:00 | 6 | 168 | 10 | Semestral | 1284790 |
| Oficial | 152 | Inactivo | 0 | 0 | 0 | 0 | 10 | Semestral | 0 |
| Oficial | 153 | Inactivo | Acreditación de alta calidad | 0 | 6 | 160 | 10 | Semestral | 0 |
| Oficial | 154 | Inactivo | Acreditación de alta calidad | 0 | 6 | 160 | 10 | Semestral | 0 |
| Oficial | 155 | Activo | Acreditación de alta calidad | 2023-05-19 00:00:00 | 4 | 160 | 10 | Semestral | 1284790 |
| Oficial | 156 | Activo | Acreditación de alta calidad | 2022-05-24 00:00:00 | 6 | 160 | 10 | Semestral | 1284790 |
| Oficial | 157 | Activo | Acreditación de alta calidad | 2020-11-18 00:00:00 | 4 | 160 | 10 | Semestral | 1284790 |
| Oficial | 158 | Activo | Acreditación de alta calidad | 2018-04-25 00:00:00 | 6 | 160 | 10 | Semestral | 1284790 |
| Oficial | 159 | Activo | Acreditación de alta calidad | 2019-06-12 00:00:00 | 4 | 134 | 8 | Semestral | 1284790 |
| Oficial | 161 | Activo | Registro calificado | 2016-10-10 00:00:00 | 7 | 24 | 2 | Semestral | 0 |
| Oficial | 162 | Activo | Registro calificado | 2016-10-10 00:00:00 | 7 | 23 | 2 | Semestral | 5000000 |
| Oficial | 163 | Inactivo | Registro calificado | 2006-04-20 00:00:00 | 7 | 36 | 3 | Semestral | 0 |
| Oficial | 165 | Inactivo | 0 | 2016-10-10 00:00:00 | 0 | 24 | 2 | Semestral | 0 |
| Oficial | 166 | Inactivo | 0 | 0 | 0 | 0 | 4 | Semestral | 0 |
| Oficial | 167 | Inactivo | 0 | 0 | 0 | 0 | 4 | Semestral | 0 |
| Oficial | 169 | Activo | Acreditación de alta calidad | 2023-04-25 00:00:00 | 6 | 48 | 4 | Semestral | 6500000 |
| Oficial | 170 | Activo | Acreditación de alta calidad | 2023-03-03 00:00:00 | 6 | 50 | 4 | Semestral | 5000000 |
| Oficial | 171 | Inactivo | 0 | 0 | 0 | 0 | 4 | Semestral | 0 |
| Oficial | 172 | Inactivo | Acreditación de alta calidad | 2017-05-11 00:00:00 | 4 | 45 | 4 | Semestral | 4000000 |

Figure 4: Informacion Programas

4.4 Código SQL - lenguaje de definición de datos (DDL) (Primera entrega)

Listing 1: Código SQL

```
CREATE SCHEMA PROYECTO_BASE_DE_DATOS;

CREATE TABLE DATOS.SABERPRO.EDUCACION (
  ESTU_CONSECUTIVO VARCHAR(14) PRIMARY KEY,
  ESTU_COD_RESIDE_DEPTO FLOAT(7) NULL,
  ESTU_DEPTO_RESIDE VARCHAR(15) NULL,
  INST_COD_INSTITUCION INT NULL,
  INST_NOMBRE_INSTITUCION VARCHAR(95) NULL,
  INST_CARACTER_ACADEMICO VARCHAR(25) NULL,
  ESTU_NUCLEO_PREGRADO VARCHAR(9) NULL,
  ESTU_PRGM_LACADEMICO VARCHAR(106) NULL,
  ESTU_PRGM_CODMUNICIPIO INT NULL,
  ESTU_PRGM_MUNICIPIO VARCHAR(22) NULL,
  ESTU_GENERO VARCHAR(3) NULL,
  ESTU_FECHANACIMIENTO VARCHAR(10) NULL,
  MOD_RAZONA_CUANTITAT_PUNT INT NULL,
  MOD_COMUNLESCRITA_PUNT FLOAT(5) NULL,
  MOD_LLECTURA_CRITICA_PUNT INT NULL,
  MOD_INGLES_PUNT FLOAT(5) NULL,
  MOD_COMPETEN_CIUADADA_PUNT INT
);

ALTER TABLE DATOS.SABERPRO.EDUCACION
ADD ESTU_SNIES_PRGMACADEMICO FLOAT(8);

CREATE TABLE DATOS.MATRICULADOS.EDUCACION (
  CDIGO_DE_LA_INSTITUCION VARCHAR(4),
  IES_ACREDITADA VARCHAR(2),
  CDIGO_DEL_MUNICIPIO_IES VARCHAR(5),
  CDIGO_SNIES_DEL_PROGRAMA VARCHAR(6),
  PROGRAMA_ACREDITADO VARCHAR(2),
  METODOLOGIA VARCHAR(11),
  SEXO VARCHAR(9),
  SEMESTRE VARCHAR(1),
  MATRICULADOS VARCHAR(4)
);

ALTER TABLE DATOS.PROGRAMAS.EDUCACION
MODIFY COLUMN CDIGO_SNIES_DEL_PROGRAMA FLOAT(8) PRIMARY KEY;
```

```

CREATE TABLE DATOS.PROGRAMAS.EDUCACION (
  SECTOR VARCHAR(7) ,
  CDIGO_SNIES_DEL_PROGRAMA FLOAT(8) ,
  ESTADO_PROGRAMA VARCHAR(8) ,
  RECONOCIMIENTO_DEL_MINISTERIO VARCHAR(28) ,
  FECHA_EJECUTORIA VARCHAR(50) ,
  VIGENCIA_AOS VARCHAR(3) NULL,
  N_MERO_CR_DITOS FLOAT(5) NULL,
  N_MERO_PERIODOS_DE_DURACION FLOAT(4) NULL,
  PERIODICIDAD VARCHAR(13) ,
  COSTO_MATRÍCULA_ESTUD_NUEVOS FLOAT(11) NULL
);

```

4.5 Código SQL - Manipulación de datos (DML) (*Primera entrega*)

4.6 Código SQL + Resultados: Vistas (*Primera entrega*)

```

CREATE VIEW ESTUDIANTES_POR_GENERO AS
SELECT ESTU_GENERO, COUNT(*) AS CANTIDAD_ESTUDIANTES
FROM DATOS.SABERPRO.EDUCACION
GROUP BY ESTU_GENERO;

```

```

SELECT * FROM ESTUDIANTES_POR_GENERO;

```

```

CREATE VIEW ESTUDIANTES_POR_UNIVERSIDAD AS
SELECT INST_NOMBRE_INSTITUCION, COUNT(*) AS CANTIDAD_ESTUDIANTES
FROM DATOS.SABERPRO.EDUCACION
GROUP BY INST_NOMBRE_INSTITUCION;

```

```

SELECT * FROM ESTUDIANTES_POR_UNIVERSIDAD;

```

4.7 Código SQL + Resultados: Triggers (*Primera entrega*)

```

DELIMITER //

```

```

CREATE TRIGGER CALCULAR_PROMEDIO_COSTO_CARRERA
AFTER UPDATE ON DATOS.PROGRAMAS.EDUCACION
FOR EACH ROW
BEGIN
  DECLARE PROMEDIO_COSTO FLOAT;

  SELECT AVG(COSTO_MATRÍCULA_ESTUD_NUEVOS)

```

```

        INTO PROMEDIO_COSTO
        FROM DATOS.PROGRAMAS.EDUCACION
        WHERE CDIGO_SNIES_DEL_PROGRAMA = NEW.CDIGO_SNIES_DEL_PROGRAMA;

        UPDATE DATOS.PROGRAMAS.EDUCACION
        SET PROMEDIO_COSTO.CARRERA = PROMEDIO_COSTO
        WHERE CDIGO_SNIES_DEL_PROGRAMA = NEW.CDIGO_SNIES_DEL_PROGRAMA;
    END //

    DELIMITER ;

```

4.8 Código SQL + Resultados: Funciones (*Primera entrega*)

```

    DELIMITER ||

    CREATE FUNCTION CALCULAR_PROMEDIO_CALIFICACIONES (
        MOD_RAZONA_CUANTITAT_PUNT INT(3),
        MOD_COMUNLESCRITA_PUNT FLOAT(5),
        MOD_LLECTURA_CRITICA_PUNT INT(3),
        MOD_INGLES_PUNT FLOAT(5),
        MOD_COMPETEN_CIUADADA_PUNT INT(3)) RETURNS FLOAT
    BEGIN
        DECLARE PROMEDIO FLOAT;
        SET PROMEDIO = (MOD_COMPETEN_CIUADADA_PUNT +
        MOD_INGLES_PUNT +
        MOD_LLECTURA_CRITICA_PUNT +
        MOD_COMUNLESCRITA_PUNT +
        MOD_RAZONA_CUANTITAT_PUNT) / 5;
        RETURN PROMEDIO;
    END;

    SELECT CALCULAR_PROMEDIO_CALIFICACIONES(80, 85.5, 90, 88.5, 75);

```

4.9 Código SQL + Resultados: procedimientos almacenados (*Primera entrega*)

```

    DELIMITER //

    CREATE PROCEDURE PROMEDIO_GENERAL_DEPARTAMENTO()
    BEGIN
        SELECT
            ESTU_DEPTO_RESIDE AS Departamento,

```

```

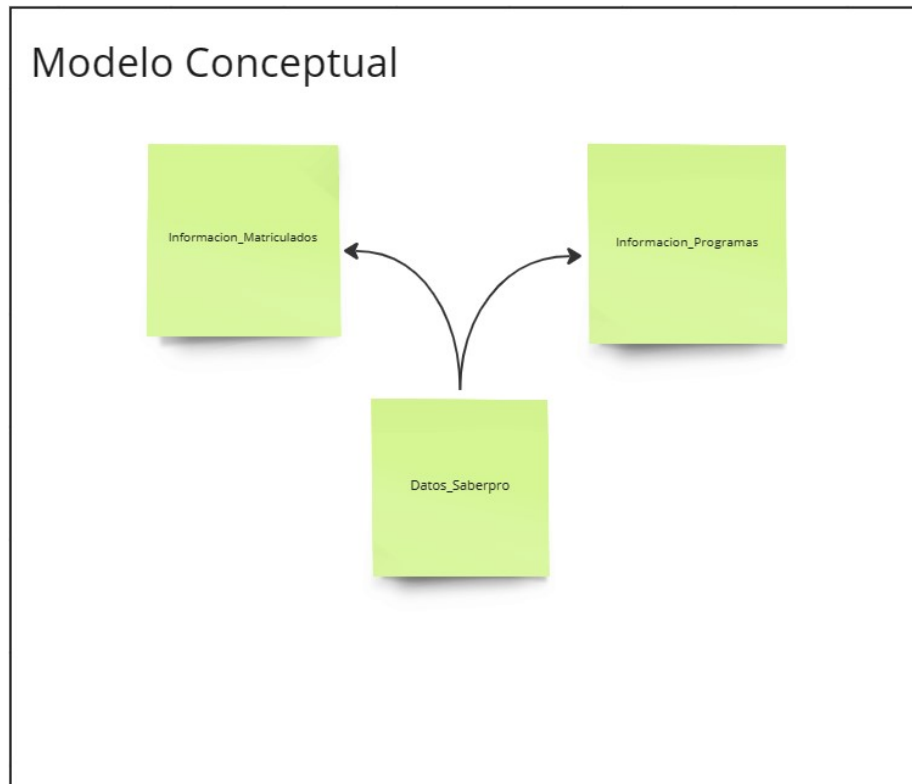
        AVG(MOD.RAZONA_CUANTITAT_PUNT) AS PROMEDIO_RAZONA_CUANTITAT,
        AVG(MOD.COMUNLESCRITA_PUNT) AS PROMEDIO_COMUNLESCRITA,
        AVG(MOD.LECTURA_CRITICA_PUNT) AS PROMEDIO_LECTURA_CRITICA,
        AVG(MOD.INGLES_PUNT) AS PROMEDIO_INGLES,
        AVG(MOD.COMPETEN_CIUADADA_PUNT) AS PROMEDIO_COMPETEN_CIUADADA
FROM
    DATOS.SABERPRO.EDUCACION
GROUP BY
    ESTU_DEPTO_RESIDE;
END //

DELIMITER ;

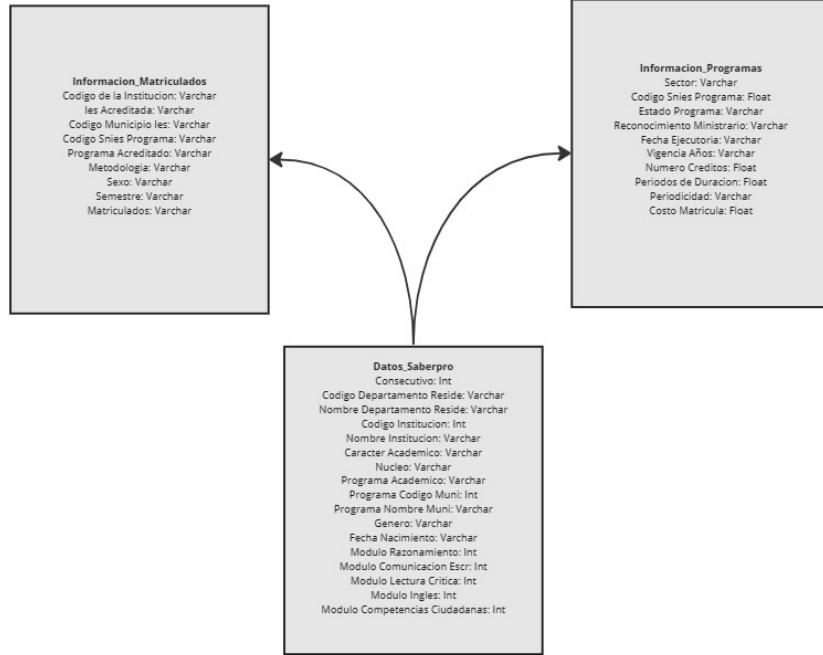
```

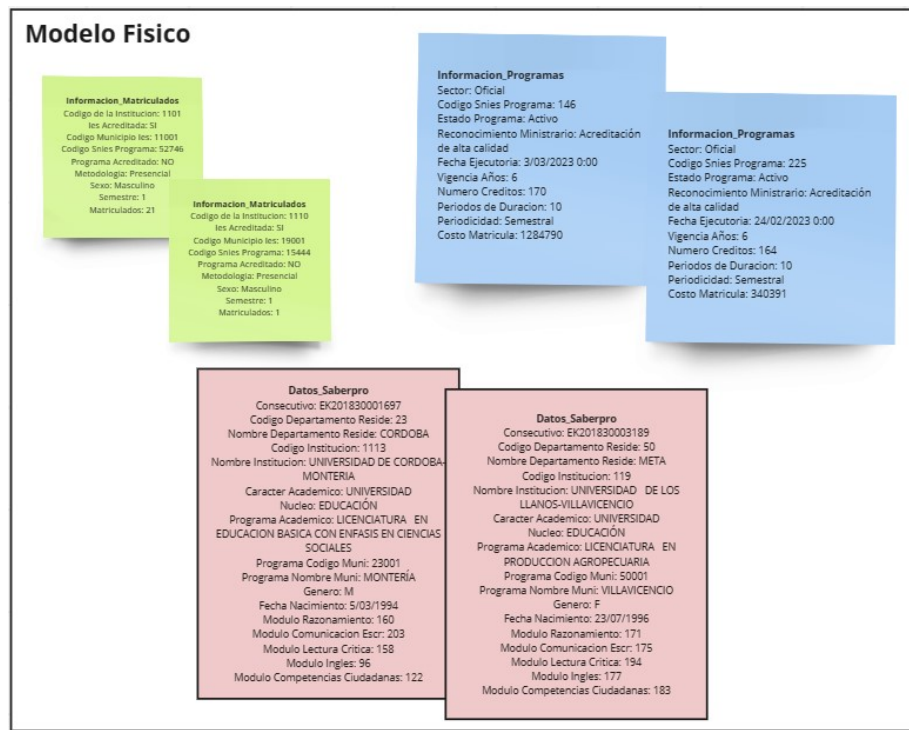
5 Bases de Datos No-SQL (*Segunda entrega*)

5.1 Diagrama Bases de Datos No-SQL (*Segunda entrega*)



Modelo Logico





5.2 SMBD utilizado para la Base de Datos No-SQL (*Segunda entrega*)

Para el proyecto de creación de un sistema de bases de datos para el desarrollo de Brújula Uni: Sistema comparativo de análisis de universidades en Bogotá para estudios en ciencias de la educación, se decidió utilizar Mongo db como SMBD para la base de datos No-SQL

MongoDB ofrece flexibilidad en el esquema de datos, lo cual es crucial dada la naturaleza dinámica y variada de la información relacionada al proyecto.

La capacidad de almacenar datos heterogéneos, facilita el diseño de una base de datos que se adapta a las necesidades cambiantes del proyecto sin comprometer la integridad de los datos, en este sentido, nos podría permitir almacenar los documentos que tienen la información correspondiente a pensum academico de los programas enlistados en este proyecto.

- 6 Aplicación de ETL (Extract, Transform, Load) y Bodega de Datos** (*Tercera entrega*)
 - 6.1 Ejemplo de aplicación de ETL y Bodega de Datos** (*Tercera entrega*)
 - 6.2 Automatización de Datos** (*Tercera entrega*)
 - 6.3 Integración de Datos** (*Tercera entrega*)

7 Proximos pasos (*Tercera entrega*)

8 Lecciones aprendidas (*Tercera entrega*)

9 Bibliografía

- Schultz, T. W. (1960). *American Economic Association*
- Acerca del examen Saber Pro. (s/f). Icfes. Recuperado el 6 de octubre de 2023, de <https://www.icfes.gov.co/acerca-del-examen-saber-pro>
- Ministerio de Tecnologías de la Información y las Comunicaciones. (2021). Datos abiertos. Recuperado de <https://gobiernodigital.mintic.gov.co/portal/Iniciativas/Datos-abiertos/>
- Ministerio de Educación de Colombia. (Año de publicación o actualización). EL SNIES - Qué es el SNIES. Recuperado de <https://snies.mineducacion.gov.co/portal/EL-SNIES/Que-es-el-SNIES/>
- Instituto Colombiano para la Evaluación de la Educación. (Año de publicación o actualización). EL Icfes - Qué es el Icfes. Recuperado de <https://www.datos.gov.co/Educacion/Resultados-nicos-Saber-Pro/u37r-hjmu>
- Congreso de la República de Colombia (Marzo 6, 2014). Ley 1712. Ley de Transparencia y del Derecho de Acceso a la Información Pública Nacional y se dictan otras disposiciones DO. 49218. <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=56882>
- Londoño, P. (2023, enero 17). *Qué es MySQL, para qué sirve y características principales*. Hubspot.es. <https://blog.hubspot.es/website/que-es-mysql>
- MySQL: qué es y cuáles son sus beneficios. (s/f). Hostgator.mx. Recuperado el 17 de noviembre de 2023, de <https://www.hostgator.mx/blog/mysql-conoce-que-es-y-que-ventajas-tiene/>