



## **Descripción - Proyecto 1**

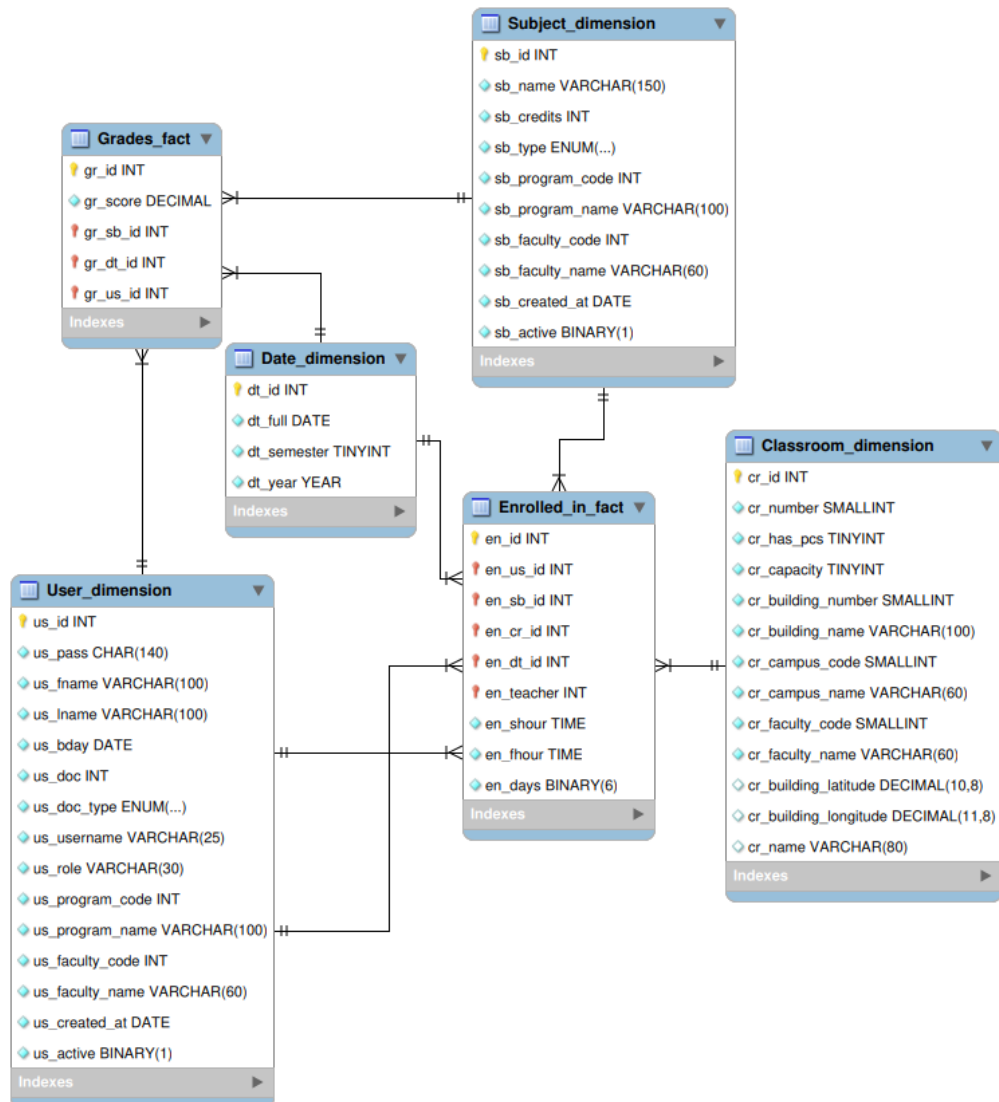
Camilo Arturo Echeverry Ayala  
Daniel Alejandro Melo Nuvar  
Jairo Daniel Salas Mosquera

Elizabeth León Guzmán

Universidad Nacional de Colombia  
Facultad de ingeniería  
Bases de datos avanzadas  
Bogotá, Colombia  
Marzo, 2023

## Modelo dimensional

El modelo que se genera es enfocado en el proceso de calificación e inscripción de materias de la Universidad Nacional. Dicho modelo consta de 2 tabla de hechos y 4 dimensiones, lo cual se puede considerar que tiene un esquema de constelación, el modelo es el siguiente:



## ETL

- Extracción:

Para la extracción de datos se usó la herramienta conocida como [Selenium](#), el cual consta en generar pruebas automatizadas en aplicaciones web. Esta herramienta junto al lenguaje de programación Python se procede a generar "Web Scraping", que en palabras simples es un programa que extrae información de una página web, esto junto al SIA se pudo extraer de manera óptima la información de las materias ofertadas por la universidad, para las facultades de, artes, ciencias e ingeniería. De igual forma para la información de facultades y carreras.

Para el registro de la información de usuarios (docentes y estudiantes) y de inscripción de asignaturas se opta por un script hecho en Python para el caso de los usuarios, genera información aleatoria, en este caso de personas; como nombres, fecha de nacimiento, número de documento, programa y facultad al que pertenece, etc. Por otra parte, el script tiene en cuenta las asignaturas ofertadas por cada programa, junto con el programa al que pertenece un estudiante para generar la información de la inscripción.

- Transformación y carga:

Se crea un programa con lenguaje de programación Python que utiliza las librerías `mysql.connect`, `pandas` y `tqdm` para consecutivamente: conectarse a la base de datos, cargar y operar sobre los datos de los archivos de excel y dar un indicador visual al usuario de este proceso, la ETL diseñada permite reiniciar de forma completa cada una de las tablas del data warehouse para una inserción completa o una modalidad de inserción sin modificar los datos actuales para agregar nuevos datos sobre los existentes, posteriormente se aplican técnicas de cifrado y reformato donde es requerido y mediante una conexión directa al motor de base de datos, los diferentes modelos creados y modificados son cargados a la estructura del data warehouse donde posteriormente serán procesados por el sistema OLAP y el sistema de BI

## **Dashboard**

El dashboard está hecho mediante la herramienta de Microsoft Power BI, el cual muestra información relacionada con:

- Cantidad de estudiantes inscritos por materia, programa y facultad en un semestre o año determinado.
- Cantidad de créditos inscritos por los estudiantes por materia, programa y facultad en un semestre o año determinado.
- Cantidad de clases que dicta un profesor por facultad y programa en un semestre.
- Promedio, máximo, mínimo y desviación estándar de notas de los estudiantes por materia, programa y facultad en un periodo determinado semestre o año determinado.
- Horas de uso de los espacios de clases.
- Número de estudiantes, profesores activos y materias que actualmente se encuentran ofertadas

Periodo - Inscripción

All

Subject\_dimension\_sb\_faculty\_name

All

Total de créditos de estudiante por periodo				
full_name	2021	2022	2023	Total
Aline Saiens	10	4		14
Allys Mackness	5	30	10	45
Andria Spurrer	4	15	9	28
Annetta Showl	9	3	14	26
Anvia Carryer	21	6		27
Augie Sharpley	26	3	2	31
August Mountlow	12	6	9	27
Total	1019	919	713	2651

Total clases por profesor				
full_name	2021	2022	2023	Total
Carmelina Cradoc	3	3	2	8
Catarina Dmytryk	6		2	8
Corbie Massingberd	1	4	1	6
Corby Trippitt	6	5	3	14
Damaris Dutnall	2		5	7
Dariusz Benane	1	4	4	9
Derek Boshere	2	1		3
Total	78	60	56	160

Cantidad de horas por aula	
Classroom_dimension.cr_building_name	2021
Centro de investigación	5,29H
Galería de aulas	12,72H
Laboratorios de desarrollo	8,03H
Total	26,04H



sb\_faculty\_name

All

832.00

Cantidad de materias activas

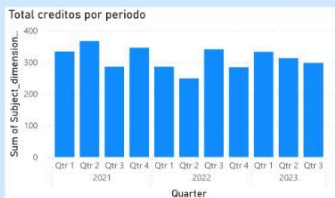
76.00

Cantidad de estudiantes activos

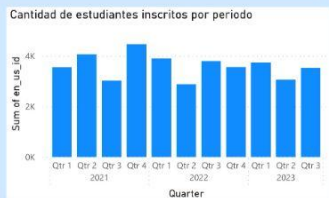
24.00

Cantidad de profesores activos

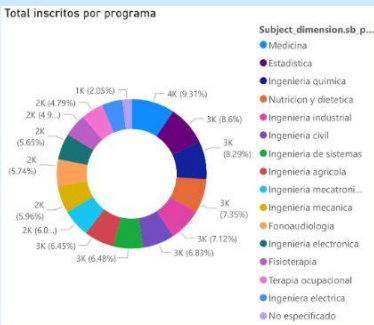
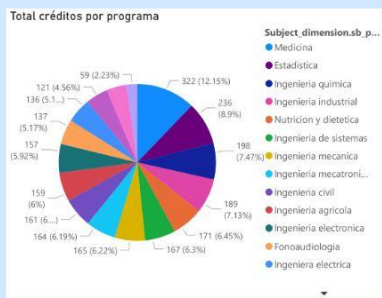
Total créditos por periodo				
Subject_dimension.sb_faculty_name	2021	2022	2023	Total
Ciencias naturales	89	86	61	236
Ingeniería	530	552	414	1496
Medicina	376	258	226	860
No especificado	24	23	12	59
Total	1019	919	713	2651



Cantidad de estudiantes inscritos				
Subject_dimension.sb_faculty_name	2021	2022	2023	Total
Ingeniería	80	85	73	100
Medicina	59	55	45	90
Ciencias naturales	23	22	17	52
No especificado	8	7	4	17
Total	92	94	89	100



Total horas de clase por profesor				
Subject_dimension.sb_faculty_name	2021	2022	2023	Total
Ciencias naturales	288.00	400.00	128.00	816.00
Ingeniería	3,856.00	2,480.00	3,104.00	9,440.00
Medicina	1,744.00	1,312.00	1,104.00	4,160.00
No especificado	128.00	120.00	64.00	312.00
Total	6,016.00	4,224.00	4,400.00	14,640.00



Periodo - Notas

Multiple selections

Subject\_dimensions\_sb\_faculty\_name

All

Total notas por periodo				
Subject_dimension.sb_faculty_name	2023	Total		
Ciencias naturales	3.20	3.20		
Ingeniería	3.06	3.06		
Medicina	2.90	2.90		
No especificado	2.50	2.50		
Total	3.01	3.01		

5.00

Nota Máxima

1.20

Desviación Estándar

1.00

Nota Mínima

3.04

Nota Promedio