

 <p>UNIVERSIDAD CENTRAL</p>	<p><b>Maestría en Analítica de Datos</b> Curso de Big Data Semestre: 202402</p>	
--	---	---

## Proyecto Final – Buscador Relatoría

### Objetivos

- Desarrollar las habilidades básicas en la manipulación de datos semiestructurados.
- Realizar una práctica en el manejo de bases de datos NoSQL.

### Prerrequisitos

- Disponibilidad de python
- Disponibilidad de un dataset con audios.

### Herramientas

- Visual studio code
- *Datasets* de audios de relatoría  
(<https://drive.google.com/drive/folders/1kGixLfEiJ8F27PHU9WaLdsrgYVfDMbgM?usp=sharing>)
- *Dataset de similitud*
- *MongoDB*
- python

### Enunciado

#### 1) Entendimiento del proyecto y de los datos

- Revise el *dataset* provisto para el proyecto, comprenda la estructura de los datos y su contenido.

#### 2) El tipo de la providencia depende del nombre de la providencia:

- Si comienza con C el tipo es **Constitucionalidad**
- Si comienza con A el tipo es **Auto**
- Si comienza con T el tipo es **Tutela**.

d)

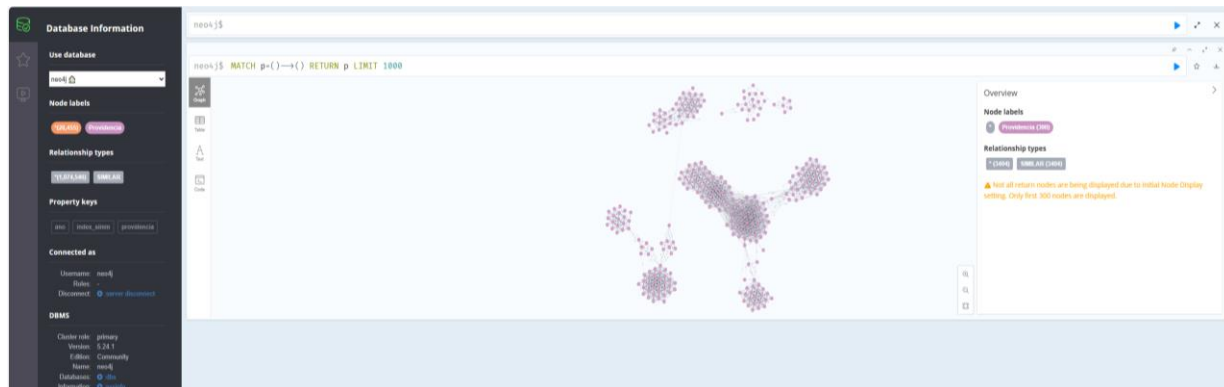
#### 3) Manipulando datos semiestructurados

- (20%) Cree una funcionalidad en python que permita extraer los textos de los archivos .wav entregados para el proyecto.
- (10%) Cree una base de datos en mongodb que permita almacenar un documento que contenga la siguiente información.

```
{
  "providencia": "C-411-22",
  "tipo": "Constitucionalidad",
  "anio": "2022",
  "texto": "Text extraído de los audios "
}
```

- c) (10%) Cree una funcionalidad que permita cargar el texto de las sentencias a la colección de mongo.
- d) (15%) Se entregara una colección de mongoDB que tiene la similitud entre las providencias suministradas en el dataset.
- i) Partiendo de esta colección se debe crear un grafo en neo4j. Con los nodos y relaciones. La estructura de la colección es la siguiente:

```
{
  "providencia1": "A-742-24",
  "providencia2": "C-440-22",
  "similitud": 29.613209919586332
}
```



- e) (55%) Cree una aplicación web que permita realizar consultas sobre la colección de mongo creada. Se evaluarán las siguientes búsquedas:
- Consultar por el nombre de una providencia.
  - Consultar las sentencias por Tipo.
  - Consultar por un año específico.
  - Consultar una providencia por cualquier texto (se debe crear un índice de texto).
  - Consultar las similitudes de una providencia específica
- f) **Bonus 1 pto:** Realizar un grafo que permita buscar por una sentencia y similitud específica. Debe realizar una búsqueda que visualice los nodos y las relaciones mayores al valor de una similitud específica.

## Metodología

- El proyecto se desarrolla en los grupos de estudiantes que están previstos para el desarrollo de trabajos prácticos.
- Se espera una participación equitativa de los integrantes del grupo

## Entrega de taller

Ultima sesión planeada a la clase.