



Procesamiento de datos masivos

## Guía para la Tarea Final.

Recuerda que debes trabajar con el grafo de twitch. Es uno de los grafos disponibles para el sandbox de Neo4j, puedes acceder siguiendo las instrucciones del tutorial, y haciendo click según la imagen y buscando en “Pre-build Data”.

### Select a project

☐ For Developers (13)    ☒ For Data Scientists (5)

## Consultas en Neo4j.

- Para extraer las 10 personas con más cantidad de vistas totales:

```
MATCH (u:Stream)
```

```
WHERE u.total_view_count IS NOT NULL
```

```
RETURN u.name as streamer,
```

```
       u.total_view_count as total_view_count
```

```
ORDER BY total_view_count DESC
```

```
LIMIT 10;
```

- Para extraer las 10 personas con mayor cantidad de aristas entrantes:

```
MATCH (u:Stream)
WITH u, count{ (u)-[:CHATTER|VIP|MODERATOR]-() } as node_indegree
RETURN u.name as streamer,
       node_indegree
ORDER BY node_indegree DESC
LIMIT 10;
```

- Para extraer la cantidad de aristas entrantes de un nodo con nombre "string"

```
MATCH (u:Stream)
WITH u, count{ (u)-[:CHATTER|VIP|MODERATOR]-() } as node_indegree
WHERE u.name = "string"
RETURN u.name,node_indegree
LIMIT 1;
```

- Comando para poder ejecutar pagerank (ver página 9 de tutorial de twitch):

```
CALL gds.graph.project('twitch',
  ['User', 'Stream'],
  ['CHATTER', 'VIP', 'MODERATOR'])
```

- Para escribir en el grafo de twitch el pagerank de cada nodo. El pagerank se escribirá como un atributo adicional en cada nodo, llamado "pagerank". Este comando demora unos minutos.

```
CALL gds.pageRank.write('twitch',{writeProperty: "pagerank"})
```

- Para extraer el pagerank de un nodo con nombre "string", una vez ejecutados los comandos anteriores

```
MATCH (u)
```

```
WHERE u.name = "50988750"
```

```
RETURN u.name, u.pagerank
```

```
LIMIT 1;
```