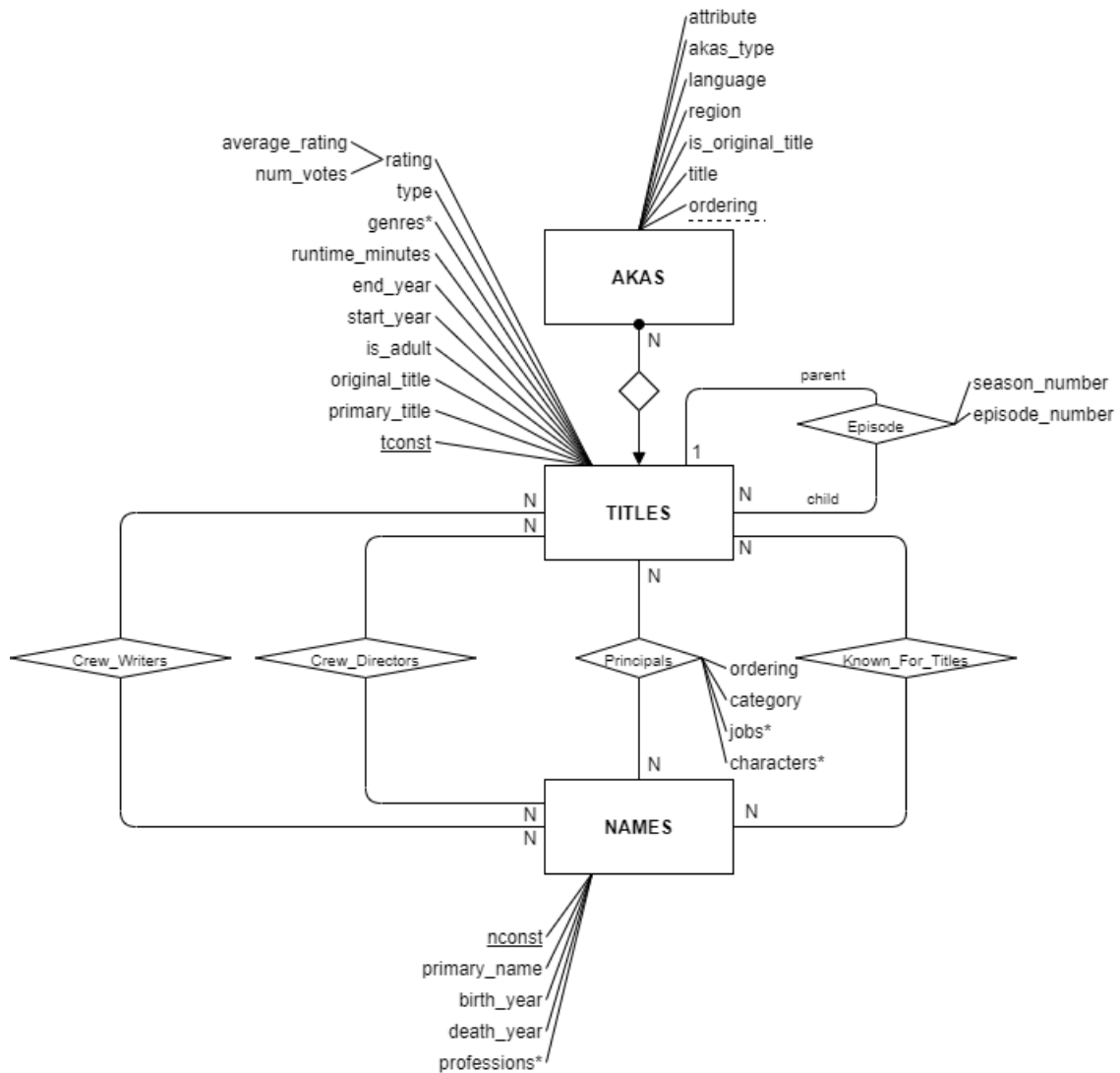


Tarea 3: Bases de datos orientadas a grafos

La IMDb (Internet Movie Database) es una base de datos en línea que almacena información relacionada con títulos (películas, series y contenidos digitales), y de sus correspondientes protagonistas, actores, directores, escritores y el resto del personal del equipo de producción.

Con el objetivo de almacenar en una base de datos la información que publica el IMDb¹ se generó el siguiente modelo conceptual (de entidad-relación):



Del análisis de los datos no se derivaron restricciones no estructurales que complementen el modelo anterior.

¹ <https://www.imdb.com/interfaces/>

Parte 1 - Diseño de base de datos

Partiendo del modelo de entidad relación anterior se diseñó una base de grafos en Neo4j. Para esta tarea se provee el siguiente archivo:

- `mpgvd_2022_neo4j_data.zip`: contiene los datos que deben cargarse de los distintos nodos y relaciones con sus respectivas propiedades

Describe el mapeo que permitió obtener como resultado la base de datos en Neo4j partiendo del modelo de entidad relación presentado en la sección anterior.

Parte 2 - Consultas

Diseñar las consultas que permitan obtener:

1. los títulos en los que 'Jim Carrey' tuvo un rol protagónico (principal) y los personajes que interpretó. El resultado debe ser una tabla que debe contener para cada título (`primaryTitle`, `characters`).
2. para cada título la información de sus episodios (`episodes`) relacionados directa o indirectamente (descendientes). El resultado debe contener para cada episodio (`parentTitle`, `depth`, `title`).

El valor de `depth` indica la cantidad de niveles de separación entre ambos títulos, iniciando en 0 si están directamente relacionados.

3. la cantidad de personajes principales y el título primario de cada película ('movie') dirigida por 'Alfred Hitchcock'. El resultado debe contener para cada título (`primaryTitle`, `charactersCount`).

Se debe entregar:

- código de cada una de las consultas resueltas utilizando Cypher
- si creó índices en la base de datos para resolver de forma más eficiente las consultas, incluir sus respectivos scripts de creación y justificar esta decisión

Parte 3 - Modelado

¿Identifica ventajas y desventajas al modelar esta realidad en una base de datos de grafos? Justifique su respuesta.

Entregable

Se deberá entregar un informe en formato PDF que contenga lo solicitado en las partes descritas en las secciones anteriores.

Anexo A: Instalación de ambiente

Los pasos a seguir para instalar el ambiente de trabajo y crear la base de datos son los siguientes:

1. Instalación JAVA Development Kit (JDK):

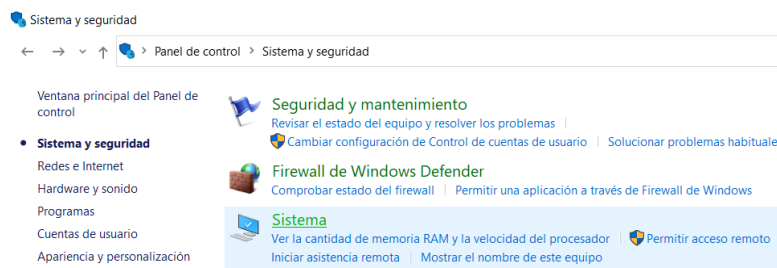
- i. Descargar **JDK 11.0.2 (build 11.0.2+9)** del siguiente link:

<https://jdk.java.net/archive/>

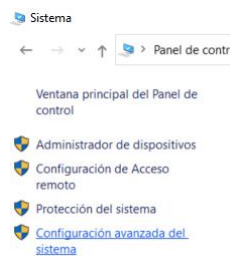
IMPORTANTE: no funciona con versiones de JDK superiores a la indicada

- ii. Descomprimir el archivo descargado en `C:\Program Files\Java`. Como resultado se creará un sub-directorio `C:\Program Files\Java\jdk-11.0.2`

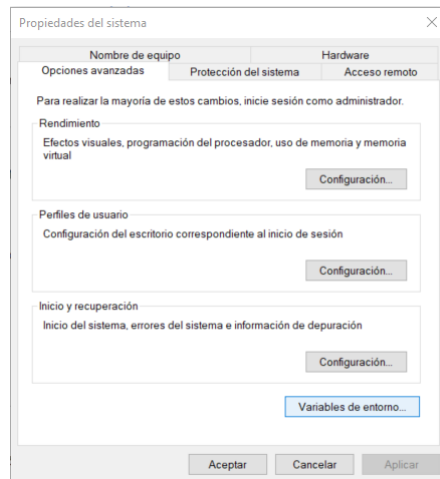
- iii. Ir a Panel de Control->Sistema y Seguridad->Sistema:



- iv. Ir a Configuración avanzada del sistema:



- v. Ir a Variables de Entorno:



- vi. Editar la variable de sistema PATH y agregar una línea con la siguiente ruta
C:\Program Files\Java\jdk-11.0.2\bin

2. Instalación Neo4j Community Edition:

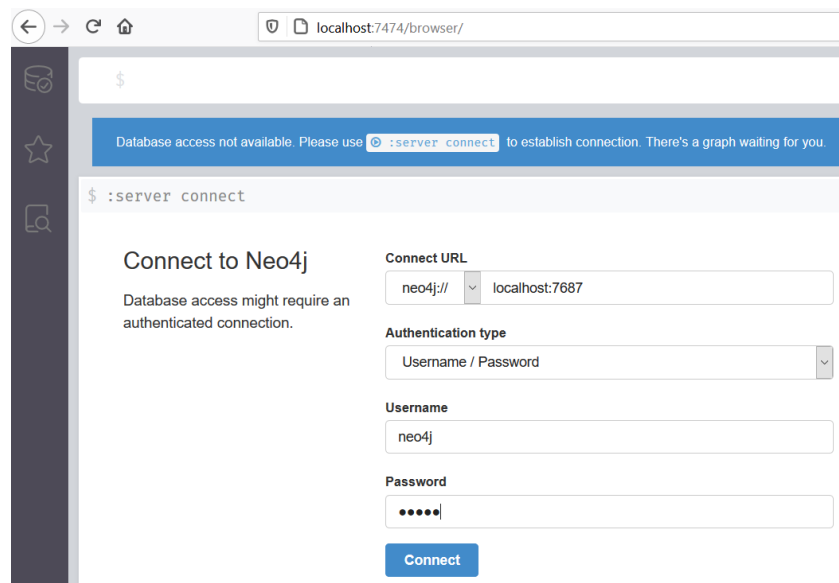
- i. Descargar Neo4j del siguiente link:
<https://neo4j.com/download-center/#community>
- ii. Descomprimir el archivo descargado en C:\ y renombrar el nuevo directorio creado de forma que quede C:\neo4j
- iii. Iniciar el servidor ejecutando C:\neo4j\bin\neo4j console:

```

C:\> Símbolo del sistema - neo4j.bat console
C:\neo4j\bin>neo4j.bat console
Directories in use:
home:          C:\neo4j
config:        C:\neo4j\conf
logs:          C:\neo4j\logs
plugins:       C:\neo4j\plugins
import:        C:\neo4j\import
data:          C:\neo4j\data
certificates:  C:\neo4j\certificates
licenses:      C:\neo4j\licenses
run:           C:\neo4j\run
Starting Neo4j.
2022-08-13 21:02:15.794+0000 INFO Starting...
2022-08-13 21:02:16.896+0000 INFO This instance is ServerId{          } (          )
2022-08-13 21:02:18.445+0000 INFO ===== Neo4j 4.4.10 =====
2022-08-13 21:02:20.059+0000 INFO Performing postInitialization step for component 'security-users' with version 3 and status CURRENT
2022-08-13 21:02:20.060+0000 INFO Updating the initial password in component 'security-users'
2022-08-13 21:02:20.770+0000 INFO Bolt enabled on 127.0.0.1:7687.
2022-08-13 21:02:21.724+0000 INFO Remote interface available at http://localhost:7474/
2022-08-13 21:02:21.734+0000 INFO id:
2022-08-13 21:02:21.747+0000 INFO name: system
2022-08-13 21:02:21.748+0000 INFO creationDate: 2022-08-13T20:37:39.777Z
2022-08-13 21:02:21.750+0000 INFO Started.

```

- iv. En cualquier navegador abrir la siguiente URL:
<http://localhost:7474>
- v. Acceder utilizando el usuario neo4j (la contraseña inicial es neo4j):



Al iniciar por primera vez se solicitará cambiar la contraseña.

3. Crear e importar BD en Neo4j:

- i. Descomprimir el archivo descargado de EVA con los datos para esta tarea en el directorio `import` dentro del directorio donde está Neo4j (`C:\neo4j\import`)
- ii. Para importar los datos en una nueva base de datos de nombre `mpgvd`, desde el directorio `C:\neo4j` ejecutar el siguiente comando:

```
bin\neo4j-admin import --database=mpgvd
--ignore-empty-strings=true --trim-strings=true
--nodes=import/node_titles.csv
--nodes=import/node_names.csv
--nodes=import/node_akas.csv
--relationships=import/relationship_episode_of.csv
--relationships=import/relationship_akas.csv
--relationships=import/relationship_known_for.csv
--relationships=import/relationship_principal_in.csv
--relationships=import/relationship_directed_by.csv
--relationships=import/relationship_written_by.csv
```

```

C:\neo4j>bin\neo4j-admin import --database=mpgvd --ignore-empty-strings=true --trim-strings=true --nodes=import/node_titles.csv --nodes=import/node_names.csv --nodes=import/node_akas.csv --relationships=import/relationship_episode_of.csv --relationships=import/relationship_akas.csv --relationships=import/relationship_known_for.csv --relationships=import/relationship_principal_in.csv --relationships=import/relationship_directed_by.csv --relationships=import/relationship_written_by.csv
Selecting JVM - Version:11.0.2+9, Name:OpenJDK 64-Bit Server VM, Vendor:Oracle Corporation
Neo4j version: 4.4.10
Importing the contents of these files into C:\neo4j\data\databases\mpgvd:
Nodes:
  C:\neo4j\import\node_titles.csv
  C:\neo4j\import\node_names.csv
  C:\neo4j\import\node_akas.csv

Relationships:
  C:\neo4j\import\relationship_episode_of.csv
  C:\neo4j\import\relationship_akas.csv
  C:\neo4j\import\relationship_known_for.csv
  C:\neo4j\import\relationship_principal_in.csv
  C:\neo4j\import\relationship_directed_by.csv
  C:\neo4j\import\relationship_written_by.csv

```

- iii. En el archivo `C:\neo4j\conf\neo4j.conf` editar el parámetro `dbms.default_database` para indicar la nueva base de datos creadas como base por defecto:

```
dbms.default_database=mpgvd
```

- iv. Iniciar el servidor ejecutando `C:\neo4j\bin\neo4j console`:

```

C:\neo4j\bin>neo4j.bat console
Directories in use:
home:          C:\neo4j
config:        C:\neo4j\conf
logs:          C:\neo4j\logs
plugins:       C:\neo4j\plugins
import:        C:\neo4j\import
data:          C:\neo4j\data
certificates:  C:\neo4j\certificates
licenses:      C:\neo4j\licenses
run:           C:\neo4j\run
Starting Neo4j.
2022-08-13 21:02:15.794+0000 INFO Starting...
2022-08-13 21:02:16.896+0000 INFO This instance is ServerId{          } (          )
2022-08-13 21:02:18.445+0000 INFO ===== Neo4j 4.4.10 =====
2022-08-13 21:02:20.059+0000 INFO Performing postInitialization step for component 'security-users' with version 3 and status CURRENT
2022-08-13 21:02:20.060+0000 INFO Updating the initial password in component 'security-users'
2022-08-13 21:02:20.770+0000 INFO Bolt enabled on 127.0.0.1:7687.
2022-08-13 21:02:21.724+0000 INFO Remote interface available at http://localhost:7474/
2022-08-13 21:02:21.734+0000 INFO id:
2022-08-13 21:02:21.747+0000 INFO name: system
2022-08-13 21:02:21.748+0000 INFO creationDate: 2022-08-13T20:37:39.777Z
2022-08-13 21:02:21.750+0000 INFO Started.

```

- v. En cualquier navegador abrir la siguiente URL:

<http://localhost:7474>

Anexo B: Referencias

A continuación, se incluyen algunos links de referencia que pueden ser de utilidad:

- Neo4j Cypher Refcard
<https://neo4j.com/docs/cypher-refcard/current/>
- Cypher Query Language
<https://neo4j.com/developer/cypher/>
- The Neo4j Cypher Manual v4.4
<https://neo4j.com/docs/cypher-manual/current/>
- Introducing the new Cypher Query Optimizer
<https://neo4j.com/blog/introducing-new-cypher-query-optimizer/>