

**Memoria Trabajo Práctico 2 (TP2)**

Bases de Datos Orientadas a Grafos: Neo4j

***Alumno****: Francisco Javier Piqueras Martínez*

***Asignatura****: Gestión y Almacenamiento de Información no Estructurada*

***Fecha******de******entrega****: 13 de mayo de 2020*

Índice

[1. Descripción del documento 4](#_Toc40299775)

[2. Ejercicios 4](#_Toc40299776)

[2.1. Ejercicio 1: Creación de la base de datos. 4](#_Toc40299777)

[2.2. Ejercicio 2: Consultas sobre la base de datos. 6](#_Toc40299778)

[2.3. Ejercicio 3: Creación de una base de datos propia. 10](#_Toc40299779)

[3. Opinión del Trabajo Práctico 15](#_Toc40299780)

Ilustraciones

[Ilustración 2‑1. Acceso al contenedor Neo4j 5](#_Toc40299783)

[Ilustración 2‑2. Confirmación carga Base de Datos Neo4j 5](#_Toc40299784)

[Ilustración 2‑3. Información de la base de datos tras la creación. 6](#_Toc40299785)

[Ilustración 2‑4. Esquema de la base de datos. 6](#_Toc40299786)

[Ilustración 2‑5. Resultado consulta número 1. 7](#_Toc40299787)

[Ilustración 2‑6. Resultado consulta número 2. 7](#_Toc40299788)

[Ilustración 2‑7. Resultado consulta número 3. Test 1. 8](#_Toc40299789)

[Ilustración 2‑8. Eliminación nodo Demetrio 8](#_Toc40299790)

[Ilustración 2‑9. Eliminación nodo Genara 8](#_Toc40299791)

[Ilustración 2‑10. Resultado consulta número 3. Test 2. 9](#_Toc40299792)

[Ilustración 2‑11. Eliminación nodo Demetrio. Segunda vez. 9](#_Toc40299793)

[Ilustración 2‑12. Resultado consulta número 3. Test 3. 9](#_Toc40299794)

[Ilustración 2‑13. Base de datos ejercicio 3. 12](#_Toc40299795)

[Ilustración 2‑14. Resultado consulta ejercicio 3-1. 13](#_Toc40299796)

[Ilustración 2‑15. Resultado consulta ejercicio 3-2. 13](#_Toc40299797)

[Ilustración 2‑16. Resultado consulta ejercicio 3-3. 13](#_Toc40299798)

[Ilustración 2‑17. Resultado consulta ejercicio 3-4. 14](#_Toc40299799)

[Ilustración 2‑18. Resultado consulta ejercicio 3-5. 14](#_Toc40299800)

[Ilustración 2‑19. Resultado de la consulta ejercicio 3-6. 15](#_Toc40299801)

# Descripción del documento

Este documento consiste en la memoria del Trabajo Práctico 3, correspondiente al Módulo 5 de la asignatura de “Gestión y Almacenamiento de Información no Estructurada” del “Máster en Ingeniería y Ciencia de Datos” de la UNED.

El trabajo consiste en la realización de tres ejercicios con el objetivo de familiarizarse con todos los aspectos de la utilización de bases de datos orientadas a grafos, mediante el uso de la herramienta Neo4j.

La realización de estos ejercicios se entregará en unos ficheros de texto adjuntos en la entrega llamados ejercicio1.txt, ejercicio2.txt y ejercicio3.txt.

Para realizar esta práctica se va a utilizar un sistema basado en virtualización ligera con Docker. Para desplegar dicho sistema, deberemos tener Docker instalado en nuestro ordenador. En este caso, se ha realizado con el sistema operativo MAC OS.

Nota: Nótese que los enunciados de los ejercicios destacan por emplear letra en formato *cursiva*.

# Ejercicios

## Ejercicio 1: Creación de la base de datos.

*Crea una base de datos con los datos que se adjuntan en el archivo grafo.txt disponible en el curso virtual. El procedimiento de despliegue del sistema neo4j se detalla en el apartado 3 y la base de datos se explica en el apartado 4.*

En primer lugar, se ha realizado la ejecución del siguiente comando:

|  |
| --- |
| $ docker run –-publish=7474:7474 –-publish=7687:7687 –-volume=$HOME/neo4j/data:/data neo4j |

Tras esta ejecución, se ha descargado la imagen oficial de Neo4j, y se ha creado un contenedor accesible a través de la dirección http://localhost:7474. Además, crea un volumen para que la base de datos persista fuera del contenedor. A continuación, se muestran las trazas del sistema:

|  |
| --- |
| (base) MacBook-Pro-de-Javier:TP3 - Neo4j javierpiquerasmartinez$ **docker run --publish=7474:7474 --publish=7687:7687 --volume=$HOME/neo4j/data:/data neo4j**  Unable to find image 'neo4j:latest' locally  latest: Pulling from library/neo4j  54fec2fa59d0: Pull complete  b7dd01647a92: Pull complete  793cbc6f8a59: Pull complete  9a08b87df712: Pull complete  27fd00dc5a07: Pull complete  769f827a6cd8: Pull complete  1ea3d537b7db: Pull complete  199ed086dc4a: Pull complete  Digest: sha256:772647cac3fe0e4dbd92ac0d08f2c50dd3be617c04948428b6d828807847fc39  Status: Downloaded newer image for neo4j:latest  Warning: Folder mounted to "/data" is not writable from inside container. Changing folder owner to neo4j.  Directories in use:  home: /var/lib/neo4j  config: /var/lib/neo4j/conf  logs: /logs  plugins: /var/lib/neo4j/plugins  import: /var/lib/neo4j/import  data: /var/lib/neo4j/data  certificates: /var/lib/neo4j/certificates  run: /var/lib/neo4j/run  Starting Neo4j.  2020-05-12 17:36:04.010+0000 INFO ======== Neo4j 4.0.4 ========  2020-05-12 17:36:04.016+0000 INFO Starting...  2020-05-12 17:36:39.193+0000 INFO Bolt enabled on 0.0.0.0:7687.  2020-05-12 17:36:39.193+0000 INFO Started.  2020-05-12 17:36:40.123+0000 INFO Remote interface available at http://localhost:7474/ |

El acceso a dicho contenedor se realizará usando la clase neo4j y la contraseña neo4j como se puede observar en la siguiente Ilustración.

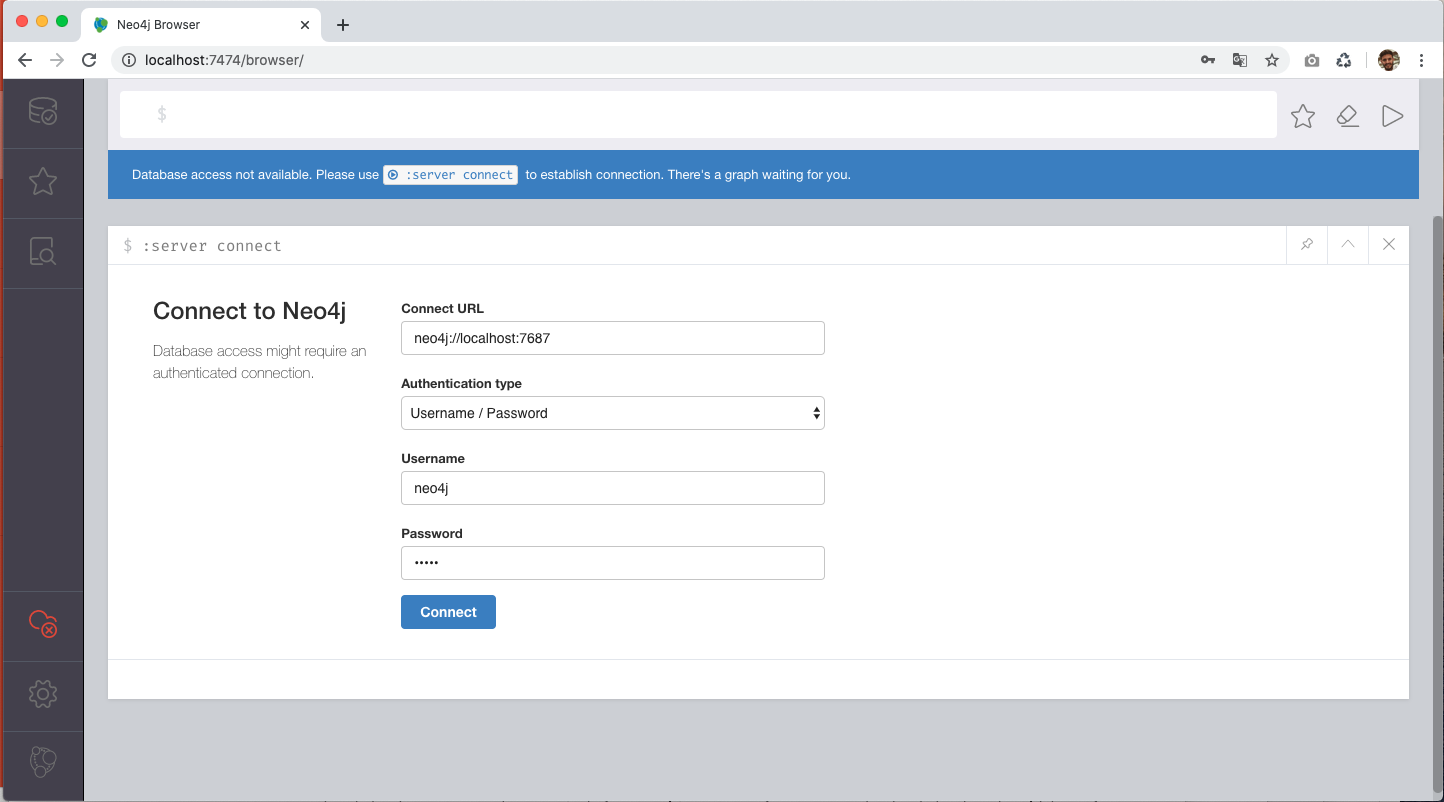


Ilustración 2‑1. Acceso al contenedor Neo4j

Para la creación de la base de datos con los datos del fichero “grafo.txt”, se ejecutan los comandos que este archivo contiene en Neo4j Browser. Se obtiene la confirmación de la ejecución como se muestra a continuación:

Captura de pantalla de un celular

Descripción generada automáticamente

Ilustración 2‑2. Confirmación carga Base de Datos Neo4j

Se comprueba que se ha creado correctamente y se observa la estructura de ésta en el panel izquierdo.

Captura de pantalla de un celular

Descripción generada automáticamente

Ilustración 2‑3. Información de la base de datos tras la creación.

Además, también se puede ver esta estructura con la ejecución del siguiente comando:

|  |
| --- |
| CALL db.schema.visualization() |

El cual proporciona el resultado mostrado en la siguiente Ilustración.

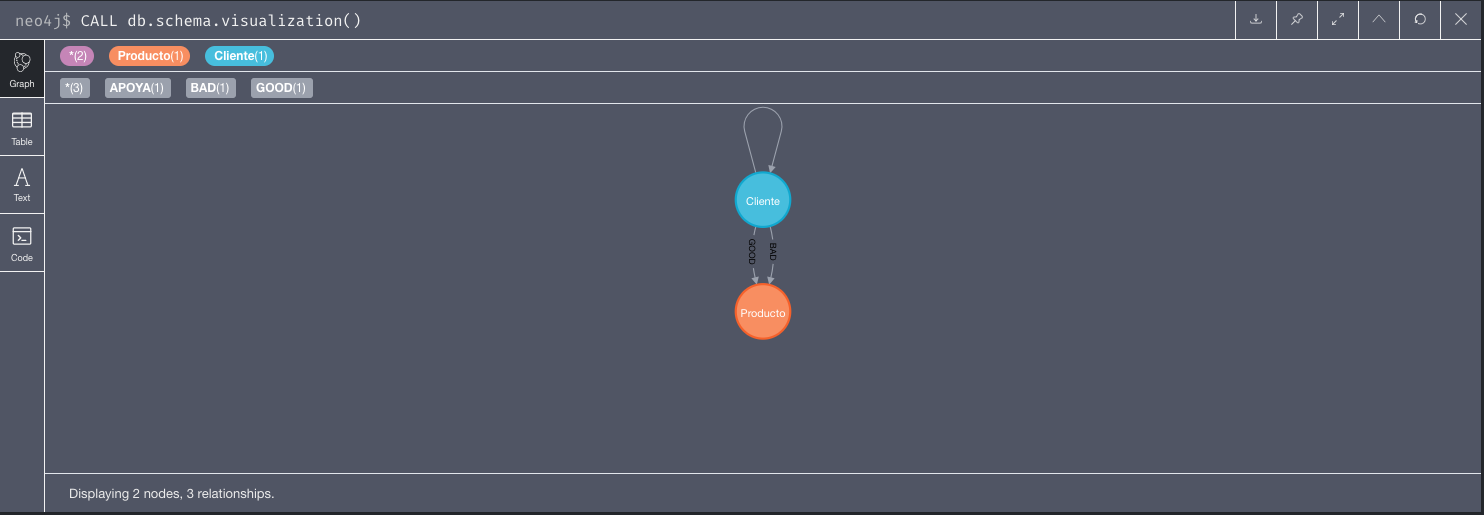


Ilustración 2‑4. Esquema de la base de datos.

## Ejercicio 2: Consultas sobre la base de datos.

*Sobre la base de datos, implementa consultas para conseguir la siguiente información:*

* *Queremos conocer nodos tipo Cliente que hayan emitido una opinión tipo BAD del producto con id=2 (hay 3 clientes con estas características).*

**MATCH (client:Cliente) -[:BAD]-> (:Producto {id:2})**

**RETURN client**

Captura de pantalla de un celular

Descripción generada automáticamente

Ilustración 2‑5. Resultado consulta número 1.

La consulta anterior también se puede hacer haciendo uso del “WHERE”:

**MATCH (client:Cliente) -[:BAD]-> (product:Producto)**

**WHERE product.id = 2**

**RETURN client**

* *Queremos buscar clientes que formen un “circuito” de relaciones de tipo “APOYA”, es decir una secuencia de relaciones, todas en la misma dirección, que acaba y empieza en el mismo Cliente.*

**MATCH circuito=(clientSource:Cliente) -[a:APOYA\*..]-> (clientSource)**

**RETURN circuito**

*Captura de pantalla de un celular

Descripción generada automáticamente*

Ilustración 2‑6. Resultado consulta número 2.

* *Demetrio ha decidido mostrar su apoyo a Genara. Añadir la relación APOYA entre los nodos correspondientes. No se debe suponer que el nodo existe (ni que no existe), solo se debe asegurar que tras la instrucción ambos existen (sin duplicarlos) y con la relación APOYA indicada.*

**MERGE (demetrio:Cliente {nombre: 'Demetrio'})**

**MERGE (genara:Cliente {nombre: 'Genara'})**

**CREATE (demetrio) -[a:APOYA]-> (genara)**

**RETURN demetrio, a, genara**

Captura de pantalla de un celular

Descripción generada automáticamente

Ilustración 2‑7. Resultado consulta número 3. Test 1.

Para probar todos los casos, se eliminan ambos nodos:

**MATCH (c:Cliente {nombre:'Demetrio'}) DETACH DELETE c**

**MATCH (c:Cliente {nombre:'Genara'}) DETACH DELETE c**

Captura de pantalla de un celular

Descripción generada automáticamente

Ilustración 2‑8. Eliminación nodo Demetrio

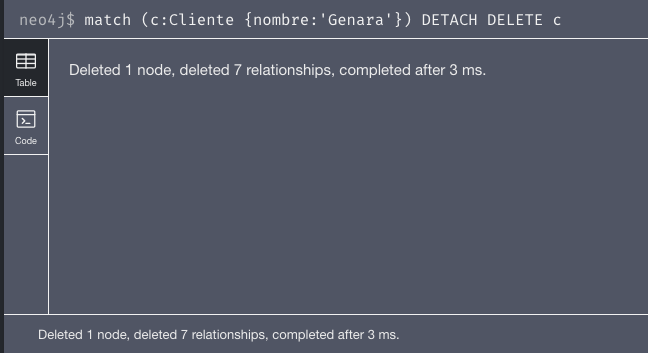


Ilustración 2‑9. Eliminación nodo Genara

Y se vuelve a ejecutar la consulta. Se obtiene el mismo resultado.

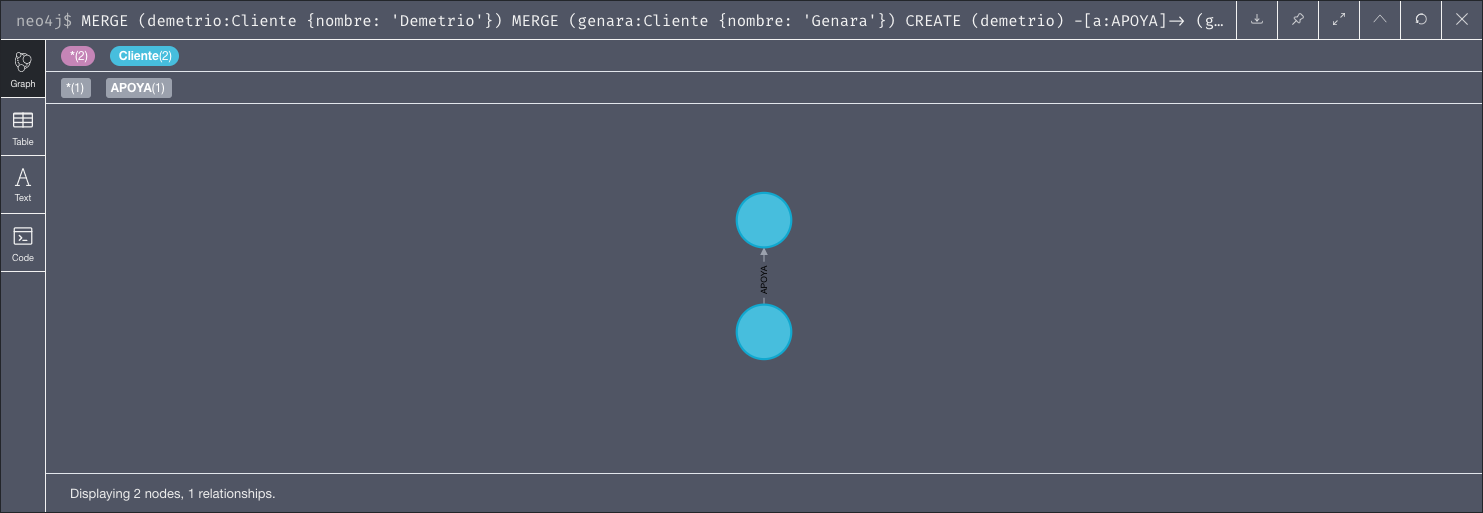


Ilustración 2‑10. Resultado consulta número 3. Test 2.

Finalmente se elimina un solo nodo de los dos:

**MATCH (c:Cliente {nombre:'Demetrio'}) DETACH DELETE c**

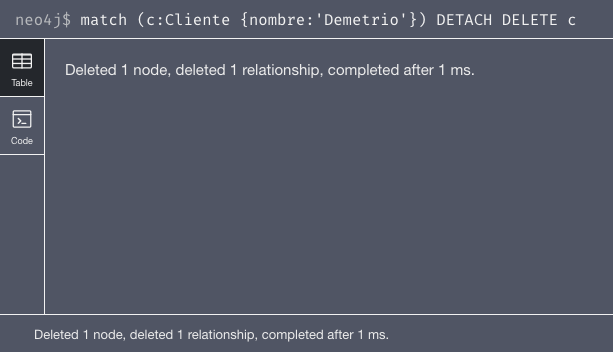


Ilustración 2‑11. Eliminación nodo Demetrio. Segunda vez.

Se ejecuta la consulta por última vez. Se obtiene el mismo resultado.

Captura de pantalla de un celular

Descripción generada automáticamente

Ilustración 2‑12. Resultado consulta número 3. Test 3.

## Ejercicio 3: Creación de una base de datos propia.

*Crea tu propia base de datos sobre Neo4j, e implementa al menos 2 consultas sobre dichos datos. Se valorará la originalidad y la complejidad de la base de datos y de las consultas creadas.*

Se crea una nueva base de datos:

|  |
| --- |
| (base) MacBook-Pro-de-Javier:TP3 - Neo4j javierpiquerasmartinez$ **docker run --publish=7474:7474 --publish=7687:7687 --volume=$HOME/companysystemuse/data:/data neo4j**  Unable to find image 'neo4j:latest' locally  latest: Pulling from library/neo4j  54fec2fa59d0: Pull complete  b7dd01647a92: Pull complete  793cbc6f8a59: Pull complete  9a08b87df712: Pull complete  27fd00dc5a07: Pull complete  769f827a6cd8: Pull complete  1ea3d537b7db: Pull complete  199ed086dc4a: Pull complete  Digest: sha256:772647cac3fe0e4dbd92ac0d08f2c50dd3be617c04948428b6d828807847fc39  Status: Downloaded newer image for neo4j:latest  Warning: Folder mounted to "/data" is not writable from inside container. Changing folder owner to neo4j.  Directories in use:  home: /var/lib/neo4j  config: /var/lib/neo4j/conf  logs: /logs  plugins: /var/lib/neo4j/plugins  import: /var/lib/neo4j/import  data: /var/lib/neo4j/data  certificates: /var/lib/neo4j/certificates  run: /var/lib/neo4j/run  Starting Neo4j.  2020-05-12 17:36:04.010+0000 INFO ======== Neo4j 4.0.4 ========  2020-05-12 17:36:04.016+0000 INFO Starting...  2020-05-12 17:36:39.193+0000 INFO Bolt enabled on 0.0.0.0:7687.  2020-05-12 17:36:39.193+0000 INFO Started.  2020-05-12 17:36:40.123+0000 INFO Remote interface available at http://localhost:7474/ |

Para insertar el contenido de la base de datos, se lanzan las siguientes sentencias:

|  |
| --- |
| match(n) detach delete(n);  CREATE (joseLuis:Trabajador {nombre:'Jose Luis García', edad:34})  CREATE (juan:Trabajador:Técnico {nombre:'Juan García', edad:33})  CREATE (adolfo:Trabajador:Técnico {nombre:'Adolfo Safalia', edad:34})  CREATE (fernando:Trabajador:Administrativo {nombre:'Fernando Calisto', edad:49})  CREATE (gabriel:Trabajador:Administrativo {nombre:'Gabriel Hernández', edad:19})  CREATE (cesar:Trabajador:Administraivo {nombre:'César Genero', edad:43})  CREATE (oscar:Trabajador:Administrativo {nombre:'Óscar Ortega', edad:46})  CREATE (amancio:Trabajador:Tecnico {nombre:'Amancio Retalde', edad:33})  CREATE (gustavo:Trabajador:Tecnico {nombre:'Gustavo Menier', edad:19})  CREATE (horacio:Trabajador:Ingeniero {nombre:'Horacio Pérez', edad:24})  CREATE (segismundo:Trabajador:Ingeniero {nombre:'Segismundo Martínez', edad:18})  CREATE (antonio:Trabajador:Director {nombre:'Antonio García', edad:51})  CREATE (impresoraSalaUno:Dispositivo:Impresora {id:1, descr:'Impresora de la Sala 1', tonerPagsRestantes:5000})  CREATE (impresoraSalaDos:Dispositivo:Impresora {id:2, descr:'Impresora de la Sala 2', tonerPagsRestantes:4000})  CREATE (impresoraPasilloUno:Dispositivo:Impresora {id:3, descr:'Impresora del Pasillo 1', tonerPagsRestantes:785})  CREATE (impresoraPasilloDos:Dispositivo:Impresora {id:4, descr:'Impresora del Pasillo 2', tonerPagsRestantes:87})  CREATE (impresoraDespacho:Dispositivo:Impresora {id:5, descr:'Impresora del despacho', tonerPagsRestantes:689})  CREATE (lectorHuella:Dispositivo:Lector {id:6, descr:'Lector huella fichaje'})  CREATE (cesar)-[:IMPRIME {id:1, fecha: datetime("2020-05-12T11:30:00.000+0100"), numeroDePaginas: 50}]->(impresoraSalaUno),  (oscar)-[:IMPRIME {id:2, fecha: datetime("2020-05-11T10:14:00.000+0100"), numeroDePaginas: 1}]->(impresoraSalaDos),  (amancio)-[:IMPRIME {id:3, fecha: datetime("2020-05-10T08:54:00.000+0100"), numeroDePaginas: 2}]->(impresoraSalaUno),  (fernando)-[:IMPRIME {id:4, fecha: datetime("2020-05-08T12:24:00.000+0100"), numeroDePaginas: 35}]->(impresoraPasilloDos),  (joseLuis)-[:IMPRIME {id:5, fecha: datetime("2020-05-07T15:56:00.000+0100"), numeroDePaginas: 15}]->(impresoraPasilloUno),  (juan)-[:IMPRIME {id:6, fecha: datetime("2020-05-05T17:41:00.000+0100"), numeroDePaginas: 24}]->(impresoraSalaUno),  (juan)-[:IMPRIME {id:7, fecha: datetime("2020-05-05T19:26:00.000+0100"), numeroDePaginas: 5}]->(impresoraSalaDos),  (segismundo)-[:IMPRIME {id:8, fecha: datetime("2020-05-11T20:14:00.000+0100"), numeroDePaginas: 2}]->(impresoraSalaDos),  (antonio)-[:IMPRIME {id:9, fecha: datetime("2020-05-12T13:16:00.000+0100"), numeroDePaginas: 352}]->(impresoraDespacho),  (fernando)-[:IMPRIME {id:10, fecha: datetime("2020-04-13T09:18:00.000+0100"), numeroDePaginas: 24}]->(impresoraPasilloUno),  (antonio)-[:IMPRIME {id:11, fecha: datetime("2020-02-10T10:26:00.000+0100"), numeroDePaginas: 12}]->(impresoraDespacho),  (joseLuis)-[:FICHA {id:12, fecha: datetime("2020-01-13T08:45:00.000+0100")}]->(lectorHuella),  (joseLuis)-[:FICHA {id:13, fecha: datetime("2020-05-11T17:49:00.000+0100")}]->(lectorHuella),  (juan)-[:FICHA {id:14, fecha: datetime("2020-05-13T07:50:00.000+0100")}]->(lectorHuella),  (juan)-[:FICHA {id:15, fecha: datetime("2020-05-13T18:46:00.000+0100")}]->(lectorHuella),  (adolfo)-[:FICHA {id:16, fecha: datetime("2020-05-13T08:01:00.000+0100")}]->(lectorHuella),  (adolfo)-[:FICHA {id:17, fecha: datetime("2020-05-13T17:01:00.000+0100")}]->(lectorHuella),  (fernando)-[:FICHA {id:18, fecha: datetime("2020-05-13T08:25:00.000+0100")}]->(lectorHuella),  (fernando)-[:FICHA {id:19, fecha: datetime("2020-05-13T18:24:00.000+0100")}]->(lectorHuella),  (gabriel)-[:FICHA {id:20, fecha: datetime("2020-05-13T08:03:00.000+0100")}]->(lectorHuella),  (gabriel)-[:FICHA {id:21, fecha: datetime("2020-05-13T17:47:00.000+0100")}]->(lectorHuella),  (cesar)-[:FICHA {id:22, fecha: datetime("2020-05-13T08:41:00.000+0100")}]->(lectorHuella),  (cesar)-[:FICHA {id:23, fecha: datetime("2020-05-13T14:52:00.000+0100")}]->(lectorHuella),  (oscar)-[:FICHA {id:24, fecha: datetime("2020-05-13T08:14:00.000+0100")}]->(lectorHuella),  (oscar)-[:FICHA {id:25, fecha: datetime("2020-05-13T17:30:00.000+0100")}]->(lectorHuella),  (amancio)-[:FICHA {id:26, fecha: datetime("2020-05-13T08:24:00.000+0100")}]->(lectorHuella),  (amancio)-[:FICHA {id:27, fecha: datetime("2020-05-13T18:34:00.000+0100")}]->(lectorHuella),  (gustavo)-[:FICHA {id:28, fecha: datetime("2020-05-13T08:30:00.000+0100")}]->(lectorHuella),  (gustavo)-[:FICHA {id:29, fecha: datetime("2020-05-13T17:10:00.000+0100")}]->(lectorHuella),  (horacio)-[:FICHA {id:30, fecha: datetime("2020-05-13T08:01:00.000+0100")}]->(lectorHuella),  (horacio)-[:FICHA {id:31, fecha: datetime("2020-05-13T18:24:00.000+0100")}]->(lectorHuella),  (segismundo)-[:FICHA {id:32, fecha: datetime("2020-05-13T09:10:00.000+0100")}]->(lectorHuella),  (segismundo)-[:FICHA {id:33, fecha: datetime("2020-05-13T17:02:00.000+0100")}]->(lectorHuella),  (antonio)-[:FICHA {id:31, fecha: datetime("2020-05-13T08:30:00.000+0100")}]->(lectorHuella),  (antonio)-[:FICHA {id:33, fecha: datetime("2020-05-13T20:30:00.000+0100")}]->(lectorHuella) |

Aquí se puede observar el resultado de la misma:

Imagen que contiene flor

Descripción generada automáticamente

Ilustración 2‑13. Base de datos ejercicio 3.

Esta es la base de datos de una empresa que registra la interacción de sus trabajadores (:Trabajador), que pueden ser (:Tecnico, :Administrativo, :Ingeniero o :Director) con los dispositivos (:Dispositivo) de la empresa (:Impresora, :Lector).

Cada vez que uno de los trabajadores imprime algo en un impresora, se crea una relación (:IMPRIME) entre el nodo del mismo, y la impresora, con las propiedades id, fecha, y numeroDePaginas. Cada vez que un usuario ficha, se crea una relación (:FICHA) con las propiedades id y fecha. Los trabajadores tienen id y nombre. Las impresoras tienen id, descripción y tonerPagsRestantes (número de páginas restantes en el tóner). El lector de huella dactilar tiene id y descripción.

Se realizan las siguientes consultas:

* Obtener los trabajadores con cargo ‘Administrativo’.

**MATCH (a:Administrativo) RETURN a**

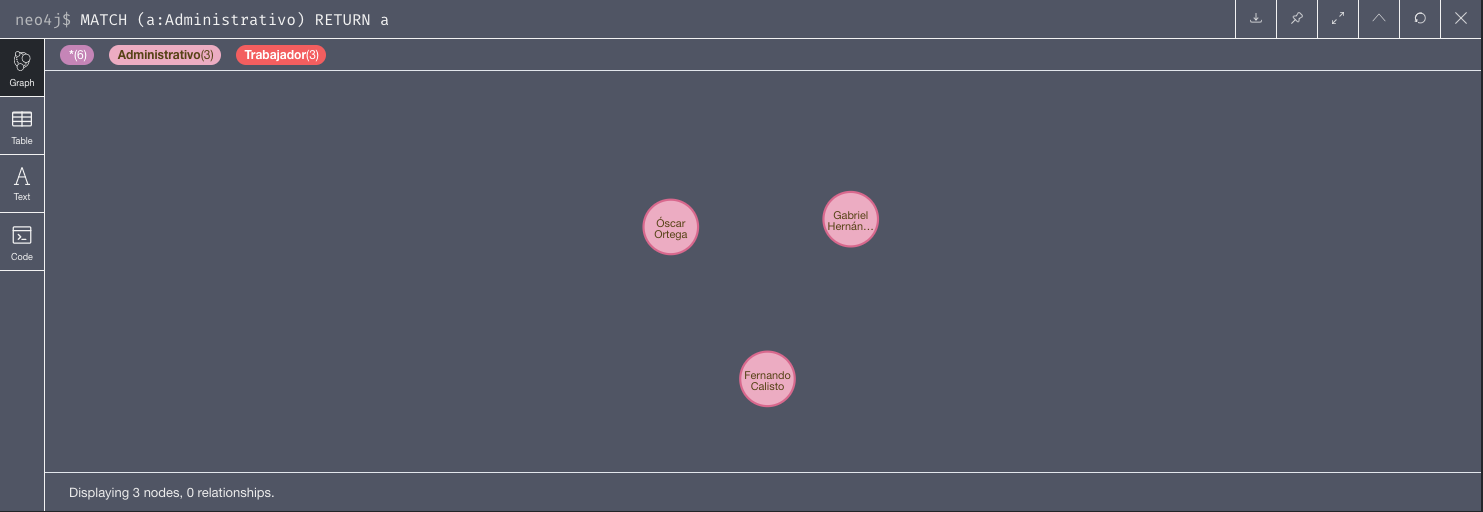


Ilustración 2‑14. Resultado consulta ejercicio 3-1.

* Obtener los trabajadores cuyo nombre es ‘Fernando’.

**MATCH (a:Trabajador)**

**WHERE a.nombre STARTS WITH 'Fernando'**

**RETURN a**

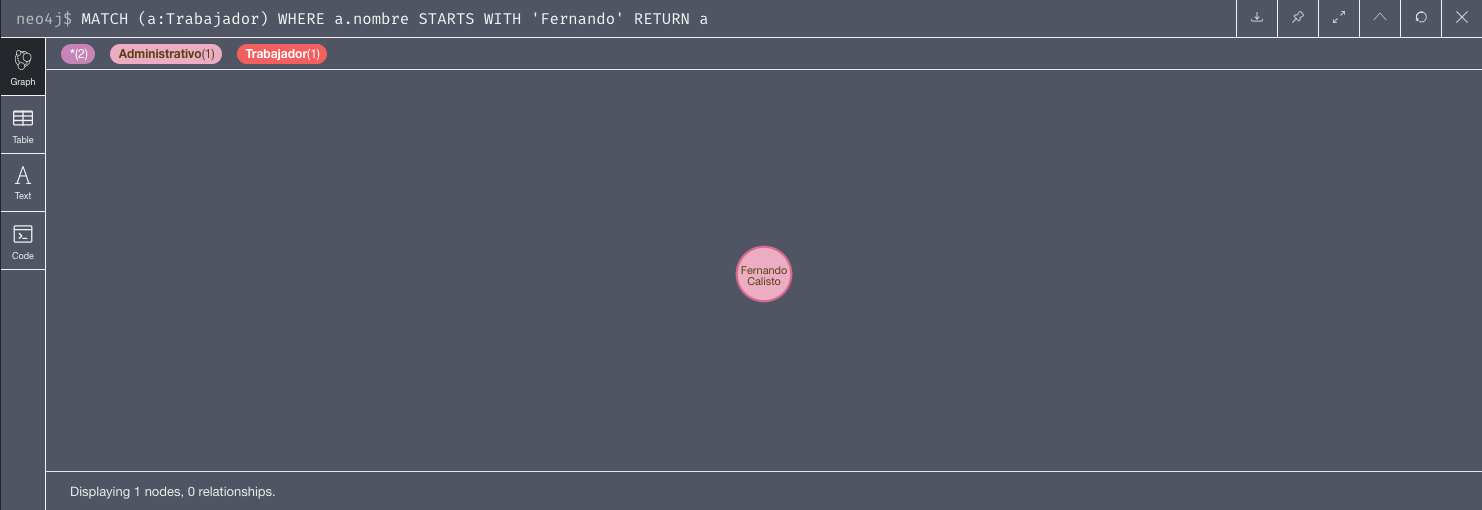


Ilustración 2‑15. Resultado consulta ejercicio 3-2.

* Obtener el total de impresiones.

**MATCH (:Trabajador)-[impresion:IMPRIME]->(:Impresora)**

**RETURN sum(impresion.numeroDePaginas)**

Imagen que contiene captura de pantalla

Descripción generada automáticamente

Ilustración 2‑16. Resultado consulta ejercicio 3-3.

* Obtener qué trabajadores han realizado más de 100 en un momento determinado, en qué impresora, cuántas páginas y cual fue el momento en el que eso ocurrió.

**MATCH (t:Trabajador)-[impresion:IMPRIME]->(impresora:Impresora)**

**WHERE impresion.numeroDePaginas > 100**

**RETURN t.nombre, impresora.descr, impresion.numeroDePaginas, impresion.fecha**

Captura de pantalla de un celular

Descripción generada automáticamente

Ilustración 2‑17. Resultado consulta ejercicio 3-4.

* Obtener la hora de entrada del usuario ‘Gustavo Menier’ el día 13 de mayo de 2020.

**MATCH (:Trabajador {nombre: 'Gustavo Menier'})-[fichaje:FICHA]->(:Lector)**

**WHERE fichaje.fecha.year=2020 and fichaje.fecha.month=05 and fichaje.fecha.day=13**

**RETURN min(fichaje.fecha)**

Captura de pantalla de un celular

Descripción generada automáticamente

Ilustración 2‑18. Resultado consulta ejercicio 3-5.

* Qué impresoras tienen poco tóner (menos de 100 páginas restantes).

**MATCH (impresora:Impresora)**

**WHERE impresora.tonerPagsRestantes < 100**

**RETURN impresora.descr, impresora.tonerPagsRestantes**

Captura de pantalla de un celular

Descripción generada automáticamente

Ilustración 2‑19. Resultado de la consulta ejercicio 3-6.

# Opinión del Trabajo Práctico

Este trabajo práctico me ha gustado mucho. Además de la buena documentación que se ha proporcionado directamente por parte del equipo docente, me ha gustado mucho también el hecho de haber podido realizar un ejercicio libre en la práctica, algo que en las anteriores no habíamos podido hacer. De esta forma, conseguimos enfrentarnos directamente al uso de este tipo de base de datos desde el planteamiento inicial del problema.