

Neo4j

David Lampón
08 de mayo de 2013

¿Qué es Neo4j?

Sistema de gestión de bases de datos orientadas a grafos.

- ... de código abierto (open-source)
- ... de alto desempeño (high-performance)
- ... de escalamiento empresarial (enterprise-grade)

Además es ACID sobre las transacciones

A - Atomicidad	"todo o nada"
C - Consistencia	"todo sigue siendo"
I - aislamiento	"no afecto al origen"
D - Durabilidad	"larga vida"

... pero ¿Qué es una base de datos orientada a grafos?

Base de datos orientadas a grafos

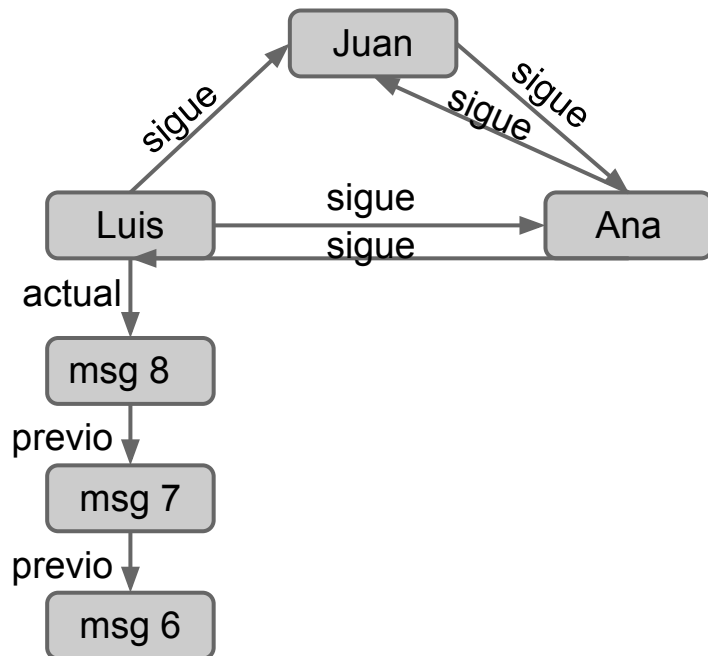
- Graph Databases:
- Sistema de gestión de bases de datos en línea con métodos de **CRUD** que se exponen en un modelo de datos en grafos.
- Generalmente se construyen para su uso con sistemas **OLTP**.
- La optimización de su desempeño depende de dos propiedades:
 - El almacenamiento subyacente (underlying storage)
 - El motor de procesamiento (processing engine).

CRUD: Creation, Read, Update, Delete

OLTP: On-Line Transaction Processing

Modelo de datos en grafos por propiedades

Twitter



1. Contiene nodos y aristas.
2. Nodos contienen propiedades (pares llave-valor).
3. Las aristas son nombradas y dirigida. Siempre tienen un nodo de origen y un nodo de destino.
4. Las relaciones también pueden tener propiedades.

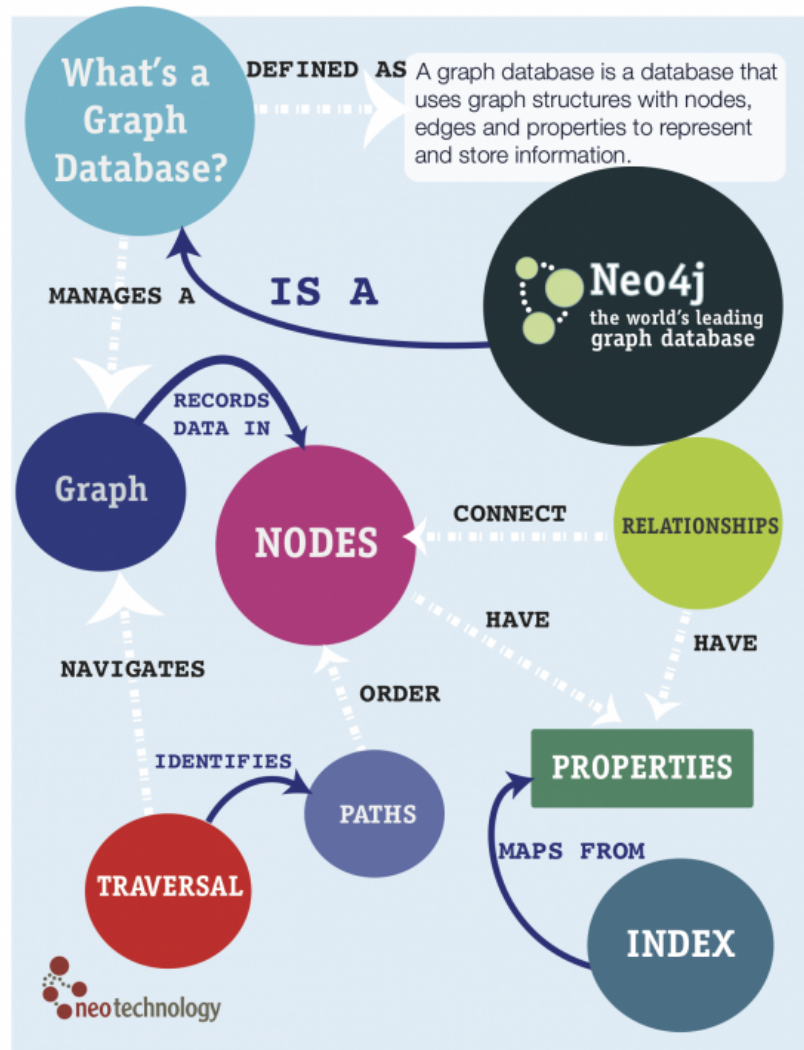
Libro descargable: <http://graphdatabases.com/>

Menos técnico:

- representan y almacenan información como nodos y aristas de un grafo,
- se puede usar teoría de grafos para recorrer la base de datos,
- debe estar normalizada, es decir, cada tabla tendría una sola columna y cada relación tan solo dos,
- cualquier cambio en la estructura de la información tenga un efecto tan solo local.

... regresando a Neo4j:

Dicho en un grafo...



Tecnología Neo4j

- Implementado en Java.
- Motor de persistencia embebido, i.e. Neo4j corre en el mismo proceso de la aplicación que requiere la base de datos.
- Almacenamiento subyacente y Motor de procesamiento nativos.
- Soporta aplicaciones en Java, JRuby, Ruby, Python, Scala, Clojure, C#, Erlang
- [Comparativo Hadoop\(HBase\) / Neo4j / PostgreSQL](#)

Ventajas

- **Intuitivo.** Usar un modelo de grafos para la representación de datos.
- **Confiable.** Con transacciones completamente ACID.
- **Durable y rápido.** Usando el disco para procesar datos.
- **Masivamente escalable.** Más de algunos billones de nodos/relaciones/propiedades.
- **Altamente-disponible.** Para distribuir el trabajo en múltiples máquinas.
- **NoSQL.** Para manejar grandes volúmenes de datos.
- **Expresivo.** Poderoso y leíble por un ser humano (graph query language)

Desventajas

- Tiene problemas representando estructuras de datos comunes como listas ordenadas, jerarquías, árboles o contenido de páginas web
- No es útil cuando el modelo de datos involucra tiempo
- Restricción semántica dada por almacenamiento de items como renglones en tablas

Consultando BDOG: Cypher

Cláusulas básica:

START + MATCH + RETURN

```
START a=node:user(name='Michael')
```

```
MATCH (a)-[:KNOWS]->(b)-[:KNOWS]->(c), (a)-[:KNOWS]->(c)
```

```
RETURN b, c
```

START. Especifica uno o más puntos de entrada (nodos o relaciones) en una gráfica (a través de índices o IDs).

MATCH. Se refiere a la especificación de lo que se está buscando .

RETURN. Especifica cuáles nodos, relaciones y propiedades en los datos encontrados por el MATCH deben ser regresados al cliente.

Además Neo4j soporta otros lenguajes de consulta como RDF, SPARQL y Gremlin.

Otras cláusulas de Cypher

- **WHERE:** Criterios para filtrar patrones que resultan del MATCH.
- **CREATE and CREATE UNIQUE:** Crea nodos y relaciones.
- **DELETE:** Borra nodos, relaciones y propiedades.
- **SET:** Asigna valores a las propiedades.
- **FOREACH:** Genera acciones para cada elemento de una lista.
- **UNION:** Une resultados de dos o más consultas (Funcionalidad que se introdujo hasta la versión 2.0 de Neo4j).
- **WITH:** Encadena consultas subsecuentes.

Instalación

1. Descargar <http://www.neo4j.org/download>

2. Descomprimir

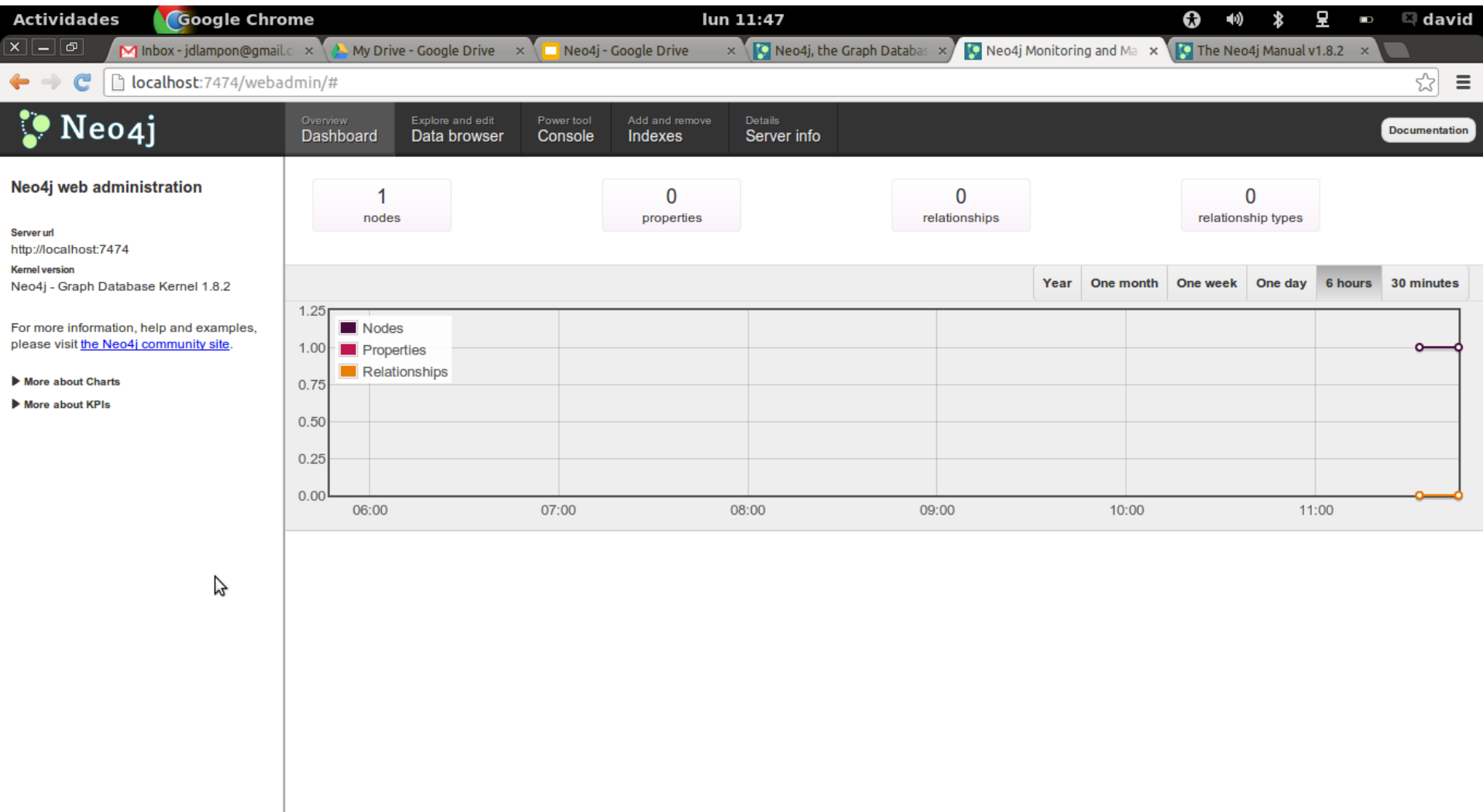
3. Iniciar:

```
david@lampon:~/granescala$ cd neo4j-community-1.8.2/
```

```
david@lampon:~/granescala/neo4j-community-1.8.2$ bin/neo4j start
```

4. Utilizar interfaz de administración web
<http://localhost:7474/webadmin/>

Interfaz - Dashboard



Interfaz - Data browser



Interfaz - Console

The screenshot shows the Neo4j web console interface. The browser's address bar displays `localhost:7474/webadmin/#/console/`. The interface has a dark theme with a top navigation bar containing links: Overview Dashboard, Explore and edit Data browser, Power tool Console (which is highlighted), Add and remove Indexes, and Details Server info. A 'Documentation' button is located on the right. On the left side, there is a sidebar with the Neo4j logo and text explaining the Neo4j Shell. The main area is a dark console with the prompt `neo4j-sh (0)$ |`. On the right side of the console, there are three tabs: 'Neo4j Shell' (selected), 'Gremlin', and 'HTTP'.

Actividades Google Chrome lun 11:55

Inbox - jdlampon@gmail.com x My Drive - Google Drive x Neo4j - Google Drive x Neo4j, the Graph Database x Neo4j Monitoring and Ma x The Neo4j Manual v1.8.2 x

localhost:7474/webadmin/#/console/

Neo4j

Overview Dashboard Explore and edit Data browser Power tool Console Add and remove Indexes Details Server info Documentation

The Neo4j Shell, allows you to use the Cypher query language for working with your graph, as well as other powerful shell features.

For command reference, see [the shell documentation](#).

For Cypher syntax help, see [the Cypher documentation](#). Remember to end your queries with a semicolon!

► Cheat sheet

► If the console hangs

neo4j-sh (0)\$ |

Neo4j Shell Gremlin HTTP

Interfaz - Indexes

The screenshot shows the Neo4j web interface in a Google Chrome browser. The browser's address bar displays `localhost:7474/webadmin/#/index/`. The interface has a dark header with the Neo4j logo and navigation tabs: Overview Dashboard, Explore and edit Data browser, Power tool Console, Add and remove Indexes, and Details Server info. A Documentation button is located on the right. The main content area is titled 'Index management' and contains two columns: 'Node indexes' and 'Relationship indexes'. Each column has a 'New [type] index' input field and a 'Create' button. A note on the left explains that index creation is for testing purposes and suggests using the Neo4j REST client for more complex configurations. Another note mentions querying indexes through the data browser. A 'More about indexes' link is at the bottom left.

Actividades Google Chrome lun 11:55

Inbox - jdlampon@gmail.com x My Drive - Google Drive x Neo4j - Google Drive x Neo4j, the Graph Database x Neo4j Monitoring and Ma x The Neo4j Manual v1.8.2 x

localhost:7474/webadmin/#/index/

Neo4j

Overview Dashboard Explore and edit Data browser Power tool Console Add and remove Indexes Details Server info Documentation

Index management

This interface lets you list, create and remove indexes from your database.

Note that index creation here is provided for testing purposes, and is only capable of creating default indexes. To create indexes with more complex configurations, please use your Neo4j REST client of choice.

Note: You can query your indexes through the data browser, see "syntax help" below the search bar there.

► More about indexes

Node indexes

New node index Create

Relationship indexes

New relationship index Create

Interfaz - Server Info

Actividades Google Chrome lun 11:55

Inbox - jdlampon@gmail.com x My Drive - Google Drive x Neo4j - Google Drive x Neo4j, the Graph Database x Neo4j Monitoring and Ma x The Neo4j Manual v1.8.2 x

localhost:7474/webadmin/#/info/

Neo4j

Overview Dashboard Explore and edit Data browser Power tool Console Add and remove Indexes Details Server info Documentation

org.neo4j

- Configuration
- Kernel
- Primitive count

com.sun.management

com.sun.management:type=HotSpotDiagnostic

java.lang

- Code Cache
- CodeCacheManager
- PS Eden Space
- PS MarkSweep
- PS Old Gen
- PS Perm Gen
- PS Scavenge
- PS Survivor Space
- java.lang:type=ClassLoading
- java.lang:type=Compilation
- java.lang:type=Memory
- java.lang:type=OperatingSystem
- java.lang:type=Runtime
- java.lang:type=Threading

java.util.logging

localhost:7474/webadmin/#/info/

Configuration

The configuration parameters used to configure Neo4j

forced_kernel_id An identifier that uniquely identifies this graph database instance within this JVM. Defaults to an auto-generated number depending on how many instance are started in this JVM.	
read_only Only allow read operations from this Neo4j instance.	false
neo4j.ext.udc.host Configuration attribute	udc.neo4j.org
logical_log The base name for the logical log files, either an absolute path or relative to the store_dir setting. This should generally not be changed.	nioneo_logical.log
node_auto_indexing Controls the auto indexing feature for nodes. Setting to false shuts it down, while true enables it by default for properties listed in the node_keys_indexable setting.	false
intercept_committing_transactions Determines whether any TransactionInterceptors loaded will intercept prepared transactions before they reach the logical log.	false
remote_shell_name Configuration attribute	shell
cache_type The type of cache to use for nodes and relationships. Note that the Neo4j Enterprise Edition has the additional 'gcr' cache type. See the chapter on caches in the manual for more information.. Valid options:[soft, weak, strong, none]	soft

Copyright (c) 2002-2012 Neo Technology. This is free software, available under the [GNU General Public License](#) version 3 or greater. Icons by [Glyphish](#) & [Tango](#).

Ejercicio de Carga

... **regresando a psql:** generar archivo de origen y destino sobre la tabla rita_1987

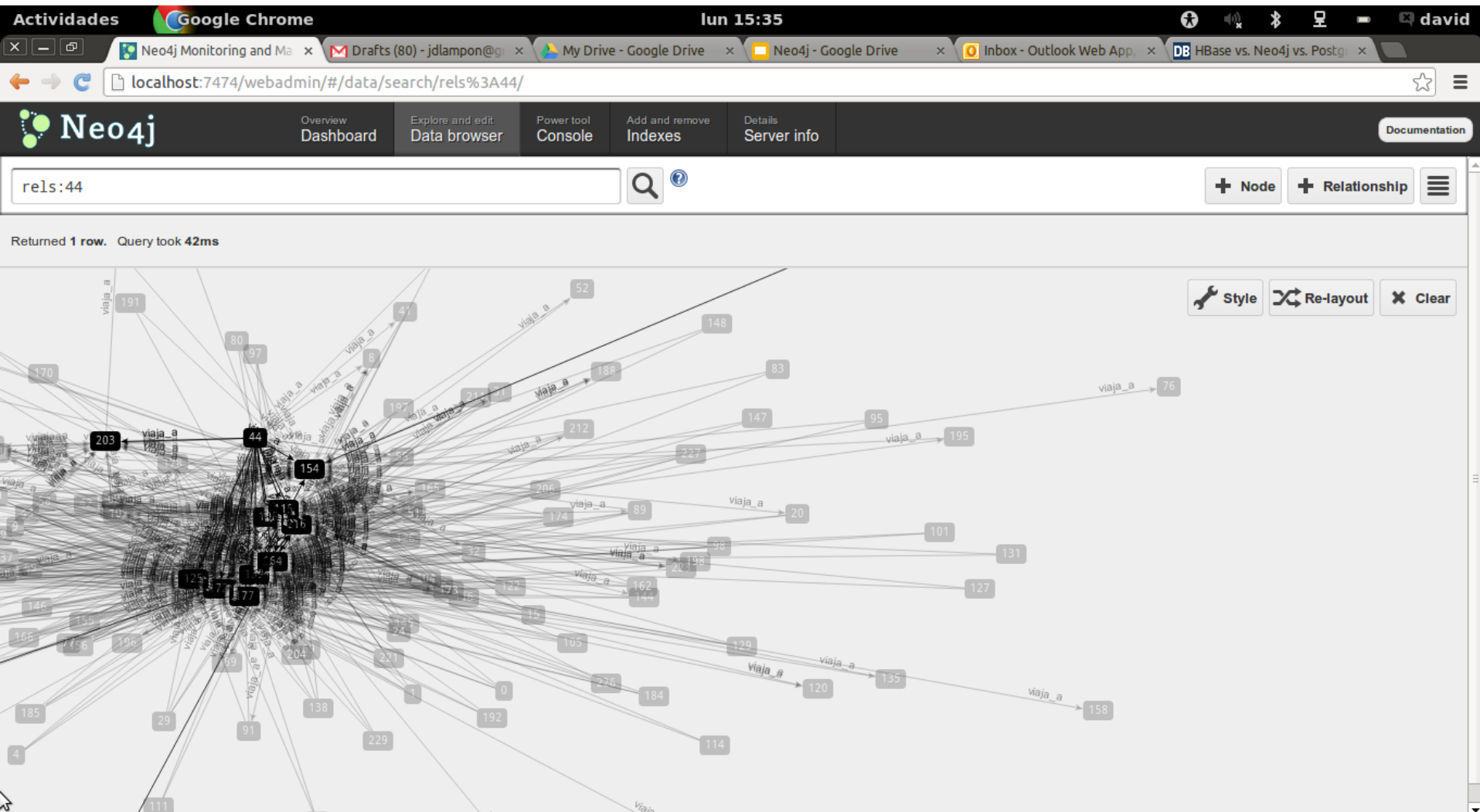
... **la magia del editor:** generar cypher queries para definir nodos y relaciones

```
create n={id:'1', name:'ABE', Type:'AA'};  
...  
start n1=node:node_auto_index(id='1'),n2=node:node_auto_index(id='9'  
create n1-[:viaja_a]->n2;  
...
```

... **y cargar en bash:** generando un archivo import.txt

```
cat rita/import.txt | bin/neo4j-shell -config conf/neo4j.properties -path  
data/graph.db
```

Ejercicio de Análisis



Referencias

- [1] Neo4j, *The World's Leading Graph Database*, <http://www.neo4j.org/>
- [2] Ian Robinson, Jim Webbwe & Emil Eifrem, *Graph Databases*, O'REILLY, 2013 NeoTechnology.
- [3] DB-ENGINES, *System Properties Comparison HBase vs Neo4j vs PostgreSQL*, <http://db-engines.com/en/system/HBase%3BNeo4j%3BPostgreSQL>
- [4] Michael Hunger, *Importing data into Neo4j - The spreadsheet way*, <http://blog.neo4j.org/2013/03/importing-data-into-neo4j-spreadsheet.html>, Marzo 06, 2013.
- [5] RITA: Research and Innovative Technology Administration, Bureau of Transportation Statistics, <http://www.transtats.bts.gov/>