Bases de Datos Orientada a Grafos Neo4j

1ra Laura Camila Scarpetta Rodríguez
Universidad Distrital Francisco José de Caldas
Maestría en Ciencias de la Información y la Comunicación
Bogotá D.C., Colombia
lcscarpettar@correo.udistrital.edu.co

2do José Manuel Vargas Montero
Universidad Distrital Francisco José de Caldas
Maestría en Ciencias de la Información y la Comunicación
Bogotá D.C., Colombia
jomvargasm@correo.udistrital.edu.co

***Resumen*—En este documento se verá una introducción al concepto de bases de datos orientada a grafos. Luego se hablará sobre la base de datos Neo4j. Se explicará su proceso de instalación para diferentes sistemas operativos y sus configuraciones iniciales. Posterior a ello se mostrará un manejo básico con este gestor de base de datos.**

***Index Terms*—Neo4j, BDOG, base de datos, relaciones, consultas**

I. Introducción

Las bases de datos orientadas a grafos ofrecen una gran oportunidad en la actualidad en la que tenemos una creciente cantidad de información.
Este modelo ofrece tantas ventajas, como lo pueden resumir las famosas cuatro V: volumen, velocidad, veracidad y variedad.
Este modelo de bases de datos se puede implementar por medio de diferentes gestores de bases de datos. Uno de los que toma gran importancia es Neo4j, por ser este de los pocos de OpenSource existentes.
Neo4j inició en el año 2007. Opera sobre Java. Es un sistema multiplataforma, desarrollado por la empresa Neo Technology.
Como ejemplo de su gran importancia y rápido crecimiento se debe decir que Neo4j tiene como clientes a las empresas Hewlett-Packard, eBay y Cisco.
En la Figura 1 se observa el logo del software Neo4j.

Figura 1. Logo de Neo4j.

II. Bases de datos orientadas a grafos

Las bases de datos orientadas a grafos (DBOG) cambian el concepto completamente de las bases de datos relacionales (SQL) y hasta de las bases de datos no relaciones (NoSQL). La información, los objetos, de la base de datos se representan por nodos. Las relaciones entre los objetos se relacionan en los grafos con las aristas. Un ejemplo de un diagrama de grafos para una base de datos se muestra en la Figura 2.

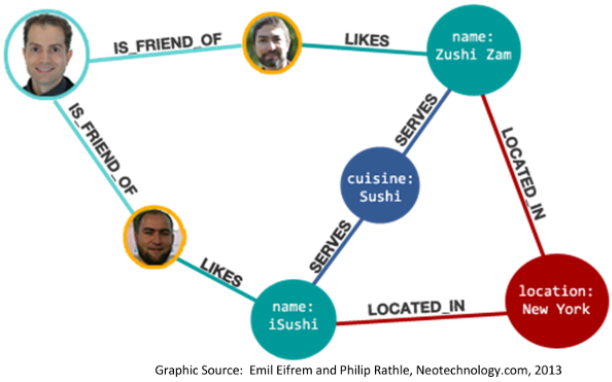

Figura 2. Example of a DBOG.

Como se observa, el diagrama es bastante intuitivo de lo que nosotros entendemos por una relación.
Neo4j usa grafos de propiedad. Son grafos con peso, con etiquetas y donde podemos asignar propiedades tanto a nodos como relaciones (por ejemplo, cuestiones como nombre, edad, país de residencia,nacimiento).

III. Instalación

El proceso de instalación del programa Neo4j es diferente para cada sistema operativo. En el presente documento se explicará el proceso para la instalación en Windows y en una distribución de Linux (Ubuntu).
Se recomienda visitar la página de Neo4j para mayor información:

- Requerimientos: https://neo4j.com/docs/operations-manual/current/installation/requirements/
- Instalación para distribuciones Linux de Debian (Ubuntu): https://neo4j.com/docs/operations-manual/current/installation/linux/debian/

- Instalación para sistemas Window: https://neo4j.com/docs/operations-manual/current/installation/windows/

III-A. Instalación en Windows

Para instalar Neo4j en un computador (o servidor) con Windows se debe descargar el instalador de la página: https://neo4j.com/download/
Una vez se haya descargado, se debe dar doble clic para ejecutarlo y se deben seguir las instrucciones.
Después de aceptar el acuerdo de licencia se mostrará la siguiente ventana, como se observa en la Figura 3.

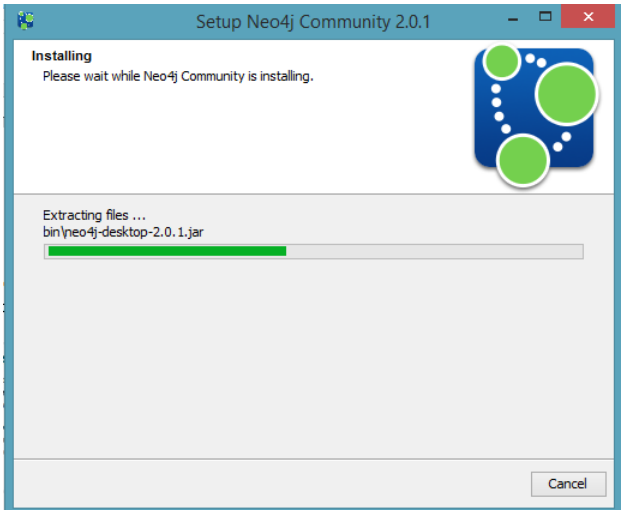

Figura 3. Instalación de Neo4j en Windows.

Una vez termine el proceso de instalación, se debe ejecutar Neo4j. Se abrirá la ventana que se muestra en la Figura 4.

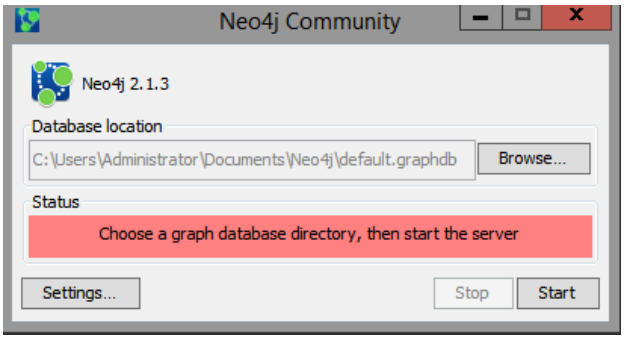

Figura 4. Ejecución de Neo4j en Windows.

Si se desea realizar configuraciones como la cantidad de memoria RAM asignad a los nodos y las aristas, la configuración del acceso desde el explorador o desde un sitio remoto, el puerto de la conexión a la base de datos, habilitar auto indexación, entre otras opciones, se debe dar clic en Settings... Se desplegará una ventana como la que se muestra en la Figura 5.

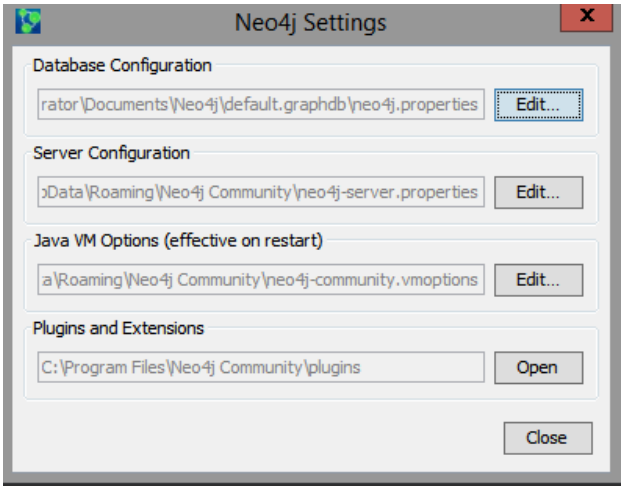

Figura 5. Configuraciones de Neo4j en Windows.

Finalmente, hechas las configuraciones deseadas, se debe dar clic en Start para hacer que el servidor arranque. La ventana de Neo4j debe quedar como en la Figura 6.

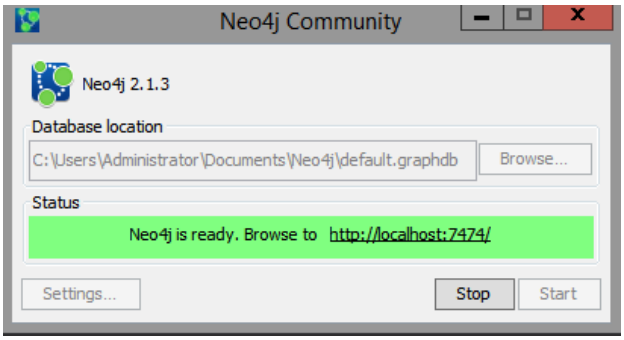

Figura 6. Arranque de Neo4j en Windows.

III-B. Instalación en Linux (Debian/Ubuntu)

Para las distribuciones de Linux el proceso es diferente. Es recomendable ir a la página https://neo4j.com/docs/operations-manual/current/installation/linux/debian/ para mayor información.
Para instalarlo se debe cumplir con el requisito de tener instalado Java 8 runtime. Para ello se debe ejecutar las siguientes sentencias, para tener acceso al repositorio e Java 8, como se muestran en las Figuras 7 y 8

```
pruebas@pruebas-VirtualBox:~$ sudo add-apt-repository ppa:webupd8team/java
```

Figura 7. Agregar repositorio de Java Runtime.

```
pruebas@pruebas-VirtualBox:~$ sudo apt-get update
```

Figura 8. Actualizar repositorios de Linux.

Se usa la siguiente sentencia para poder instalar como tal el runtime de Java, como se observa en la Figura 9

```
pruebas@pruebas-VirtualBox:~$ sudo apt-get install oracle-java8-installer
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following additional packages will be installed:
  gsfonts-x11 java-common oracle-java8-set-default
Suggested packages:
  binfmt-support visualvm ttf-baekmuk | ttf-unfonts | ttf-unfonts-core
  ttf-kochi-gothic | ttf-sazanami-gothic ttf-kochi-mincho
  | ttf-sazanami-mincho ttf-arphic-uming
The following NEW packages will be installed:
  gsfonts-x11 java-common oracle-java8-installer oracle-java8-set-default
0 upgraded, 4 newly installed, 0 to remove and 3 not upgraded.
Need to get 54.8 kB of archives.
After this operation, 272 kB of additional disk space will be used.
Do you want to continue? [Y/n]
```

Figura 9. Instalación del Runtime de Java en Linux.

Una vez hecho esto, se debe agregar el repositorio de Neo4j. Para ello se ejecuta la siguiente sentencia desde la terminal, como se muestra en la Figura 10.

```
pruebas@pruebas-VirtualBox:~$ wget -O - https://debian.neo4j.org/neotechnology.g
pg.key | sudo apt-key add -
--2017-10-27 15:39:18--  https://debian.neo4j.org/neotechnology.gpg.key
Resolving debian.neo4j.org (debian.neo4j.org)... 52.0.233.188
Connecting to debian.neo4j.org (debian.neo4j.org)|52.0.233.188|:443... connected
.
HTTP request sent, awaiting response... 200 OK
Length: 4791 (4.7K) [application/octet-stream]
Saving to: 'STDOUT'

-                   100%[===================>]   4.68K  --.-KB/s    in 0s

2017-10-27 15:39:37 (434 MB/s) - written to stdout [4791/4791]

OK
pruebas@pruebas-VirtualBox:~$ echo 'deb http://debian.neo4j.org/repo stable/' |
sudo tee -a /etc/apt/sources.list.d/neo4j.list
deb http://debian.neo4j.org/repo stable/
pruebas@pruebas-VirtualBox:~$ sudo apt-get update
```

Figura 10. Repositorios de Neo4j en Linux.

Finalmente, se instala Neo4j, haciendo lo que se muestra en la Figura 11.

```
pruebas@pruebas-VirtualBox:~$ sudo apt-get install neo4j
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following additional packages will be installed:
  cypher-shell daemon
The following NEW packages will be installed:
  cypher-shell daemon neo4j
0 upgraded, 3 newly installed, 0 to remove and 3 not upgraded.
Need to get 87.7 MB of archives.
After this operation, 102 MB of additional disk space will be used.
Do you want to continue? [Y/n] y
```

Figura 11. Repositorios de Neo4j en Linux.

Para iniciar el servidor y arrancar en gestor de base de datos, se debe ejecutar el comando que se muestra en al Figura 12.

```
pruebas@pruebas-VirtualBox:~$ sudo service neo4j start
```

Figura 12. Inicio del servidor de Neo4j en Linux.

III-C. Prueba de la instalación

Sin importar qué distribución o qué sistema operativo se haya utilizado, se puede probar la correcta instalación utilizando el explorador Web. Se dirige a la dirección $localhost : 7474/browser/$.
En esta dirección se tendrá todo el acceso y gestión de la base de datos: es un aplicativo corriendo sobre un explorador Web. Esto se muestra en la Figura 13.

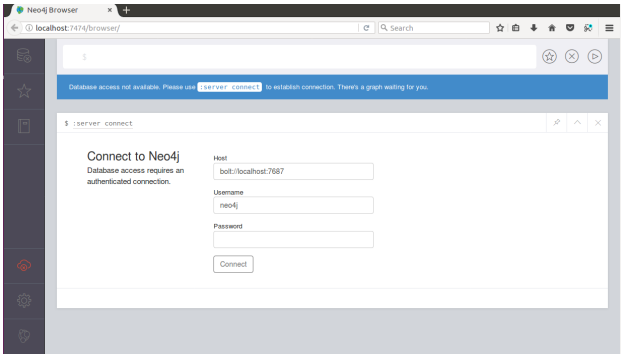

Figura 13. Neo4j corriendo desde el explorador Web.

La contraseña al iniciar Neo4j por primera vez por defecto es $neo4j$. Luego de ingresarla se pedirá se cambia la contraseña, como se muestra en la Figura 14.

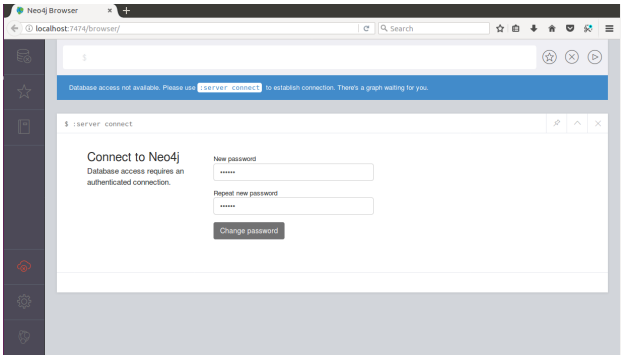

Figura 14. Cambio de contraseña en Neo4j por primera vez.

III-D. Equations

Number equations consecutively. To make your equations more compact, you may use the solidus (/), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in:

$$a + b = \gamma \tag{1}$$

Be sure that the symbols in your equation have been defined before or immediately following the equation. Use “(1)”, not “Eq. (1)” or “equation (1)”, except at the beginning of a sentence: “Equation (1) is . . .”

III-E. LaTeX-Specific Advice

Please use “soft” (e.g., `\eqref{Eq}`) cross references instead of “hard” references (e.g., `(1)`). That will make it possible to combine sections, add equations, or change the order of figures or citations without having to go through the file line by line.

Please don’t use the `{eqnarray}` equation environment. Use `{align}` or `{IEEEeqnarray}` instead. The

{eqnarray} environment leaves unsightly spaces around relation symbols.

Please note that the {subequations} environment in LaTeX will increment the main equation counter even when there are no equation numbers displayed. If you forget that, you might write an article in which the equation numbers skip from (17) to (20), causing the copy editors to wonder if you've discovered a new method of counting.

BibTeX does not work by magic. It doesn't get the bibliographic data from thin air but from .bib files. If you use BibTeX to produce a bibliography you must send the .bib files.

LaTeX can't read your mind. If you assign the same label to a subsubsection and a table, you might find that Table I has been cross referenced as Table IV-B3.

LaTeX does not have precognitive abilities. If you put a \label command before the command that updates the counter it's supposed to be using, the label will pick up the last counter to be cross referenced instead. In particular, a \label command should not go before the caption of a figure or a table.

Do not use \nonumber inside the {array} environment. It will not stop equation numbers inside {array} (there won't be any anyway) and it might stop a wanted equation number in the surrounding equation.

III-F. Some Common Mistakes

- The word "data" is plural, not singular.
- The subscript for the permeability of vacuum μ_0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter "o".
- In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
- A graph within a graph is an "inset", not an "insert". The word alternatively is preferred to the word "alternately" (unless you really mean something that alternates).
- Do not use the word "essentially" to mean "approximately" or "effectively".
- In your paper title, if the words "that uses" can accurately replace the word "using", capitalize the "u"; if not, keep using lower-cased.
- Be aware of the different meanings of the homophones "affect" and "effect", "complement" and "compliment", "discreet" and "discrete", "principal" and "principle".
- Do not confuse "imply" and "infer".
- The prefix "non" is not a word; it should be joined to the word it modifies, usually without a hyphen.
- There is no period after the "et" in the Latin abbreviation "et al.".
- The abbreviation "i.e." means "that is", and the abbreviation "e.g." means "for example".

An excellent style manual for science writers is [7].

III-G. Authors and Affiliations

The class file is designed for, but not limited to, six authors. A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

III-H. Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is "Heading 5". Use "figure caption" for your Figure captions, and "table head" for your table title. Run-in heads, such as "Abstract", will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced.

III-I. Figures and Tables

III-I0a. Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation "Fig. 15", even at the beginning of a sentence.

Cuadro I
TABLE TYPE STYLES

Table Head	**Table Column Head**		
	Table column subhead	***Subhead***	***Subhead***
copy	More table copy[a]		

[a]Sample of a Table footnote.

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity "Magnetization", or "Magnetization, M", not just "M". If including units in the label, present

Figura 15. Example of a figure caption.

them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

ACKNOWLEDGMENT

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks . . .”. Instead, try “R. B. G. thanks. . .”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

REFERENCES

Please number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first . . .”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

REFERENCIAS

[1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4] K. Elissa, “Title of paper if known,” unpublished.

[5] R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.