

# Tópicos de NOSQL

## *Base de Datos de Grafos: NEO4J*

Santiago Camisay (47583) y Teresa Fontanella De Santis (52455)  
 Grupo 1 - Trabajo Práctico - Instituto Tecnológico de Buenos Aires (ITBA)

**Resumen**—En el presente informe se describe el diseño de una base de datos de grafos, dados un schema y una serie de consultas a dicha base. El motor de base de datos utilizado es neo4j en el lenguaje Java.

**Palabras claves**—NOSQL, neo4j, base de datos de grafos, TPC H benchmark.



## 1 INTRODUCCIÓN

El nodo LineItem

En la actualidad, cada vez más, se requieren consultas eficientes y rápidas a bases de datos que manejan grandes volúmenes de información. Es de esta necesidad que surgen las bases de datos NOSQL, que proponen flexibilizar la estructura de dichas bases según las reglas de negocio. En este caso, se utilizará una base de datos de grafos (en donde cada tupla de datos en el modelo relacional es un nodo) con el motor Neo4J. El problema a resolver es el "TPC- H benchmark" .

En la siguiente sección, se describen los tipos de nodos utilizados, así como las relaciones entre ellos y los índices correspondientes.

Además, para lograr que las consultas fueran más eficientes se agregaron índices en: L\_ShipDate en LineItem, O\_OrderDate en Order, C\_MktSegment en Customer, P\_Size y P\_Type en Part, y PS\_SupplyCost en PartSupplier. Para esto también se consideró que Neo4J sabe cuándo utilizar un índice u otro o ninguno.

## BIBLIOGRAFÍA

- [1] John Matthews and Kurtis Fink, *Métodos Numéricos con Matlab*, Fecha de publicacion: Mayo 2000 — ISBN-10: 8483221810 — ISBN-13: 978-8483221815 — Edición: 3
- [2] Apuntes de la Cátedra.
- [3] Consultas a la Cátedra, a quien agradecemos su gentil colaboración, por su experta ayuda en la preparación de este artículo.

## 2 ESTRUCTURA UTILIZADA

Siempre teniendo en cuenta las consultas que debían realizarse a la base de datos y el tamaño de las tablas correspondientes al modelo relacional, se consideró oportuno que los nodos principales fueran: Region, Nation, Part, Supplier, Customer, Order y LineItem; a cada uno de ellos, les corresponde una etiqueta (o *label*) del mismo nombre.

El nodo region tiene los campos R\_Name y R\_Comment (ambos textos variables), y está relacionado de manera unidireccional con el nodo nation por medio de HAS\_NATION.

El nodo Nation tiene los campos N\_Name y N\_Comment (ambos son de texto de longitud variable), y tiene las relaciones: HAS\_CUSTOMER (unidireccional, de nation a customer), y HAS\_SUPPLIER (unidireccional, de nation a customer).

El nodo Part tiene los campos: P\_PartKey, P\_Name, P\_Mfgr, P\_Brand, P\_Type, P\_Size, P\_Container, P\_RetailPrice y P\_Comment

El nodo Supplier

El nodo Customer

El nodo Order