

Big Data, NoSQL y Neo4j



Bayron Ortiz
Juan Cuero

Hoja de Ruta

1. Big Data.
2. Tipos de Datos en Big Data.
3. Características Big Data.
4. Bases de Datos NoSQL
5. Características BD. NoSQL
6. Tipos de BD NoSQL
7. Teorema CAP
8. Neo4j
9. Conceptos sobre Neo4j
10. Características Neo4j
11. Aplicaciones Neo4j.
12. Demo: GruposNeo



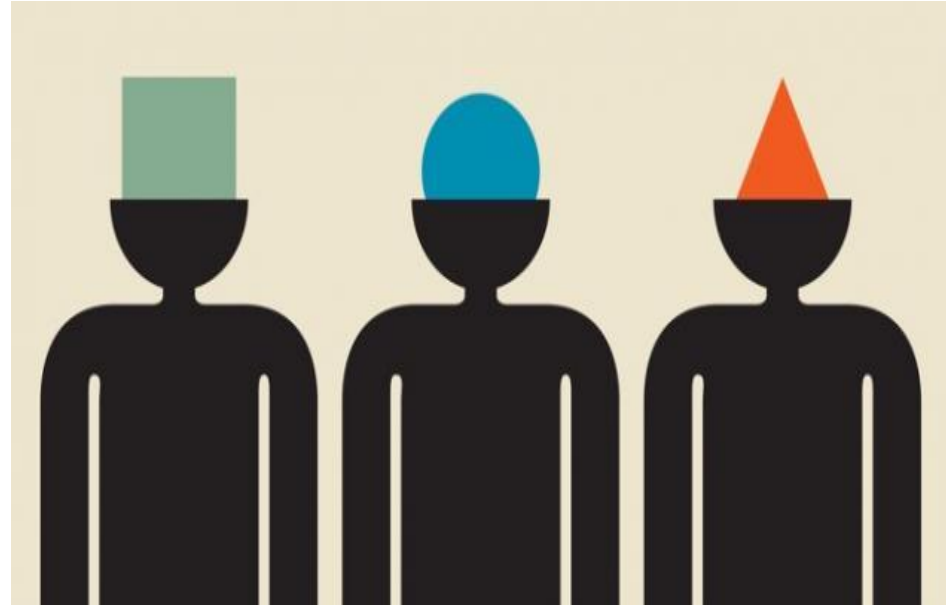
1. Big Data

Se define como un término que incluye diferentes tecnologías asociadas a la administración de grandes volúmenes de datos, provenientes de diferentes fuentes y que se generan con rapidez.[3]



2. Tipos de Datos en Big Data

- Datos Estructurados.
- Datos Semiestructurados.
- Datos No Estructurados.



2. Tipos de Datos en Big Data

Datos Estructurados

Datos con formato o esquema fijo que poseen campos fijos. Ej:

- Bases de datos relacionales.
- Hojas de cálculo.
- Fecha Nacimiento (D/M/Y).
- Cuentas Bancarias.
- Documentos de Identificación.



2. Tipos de Datos en Big Data

Datos Semiestructurados

Datos que no tienen un formato fijos, pero contiene etiquetas o marcadores que permiten separar los elementos dato. Ej:

- Archivos HTML.
- Archivos XML.
- Archivos de Log.
- Archivos CVS.



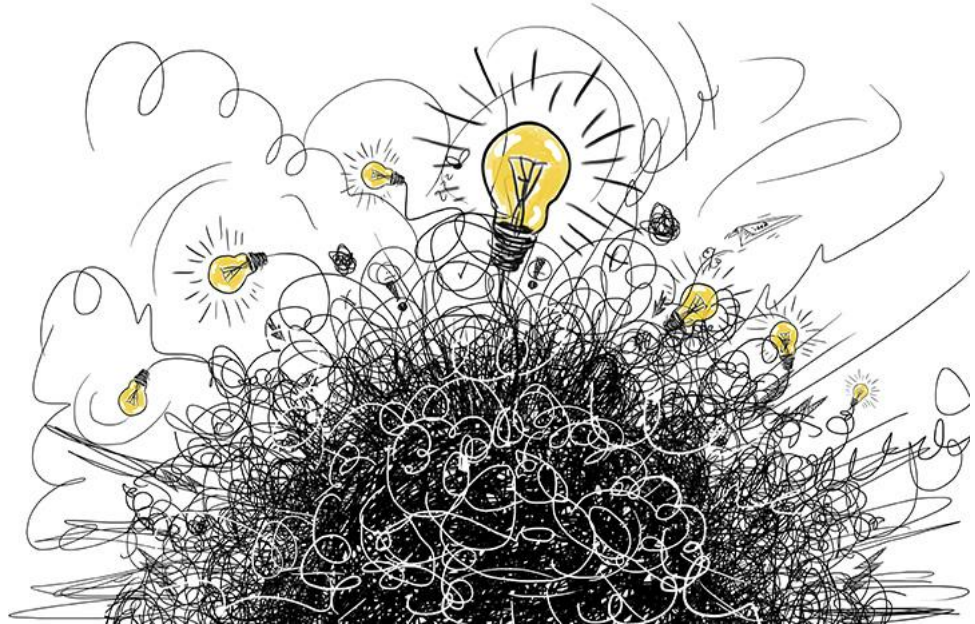
2. Tipos de Datos en Big Data

Datos No Estructurados

Aquellos que no poseen un formato ni campos fijos.

Ej:

- Imágenes.
- Audio.
- Video.
- SMS.
- Libros, artículos.



3. Características del Big Data

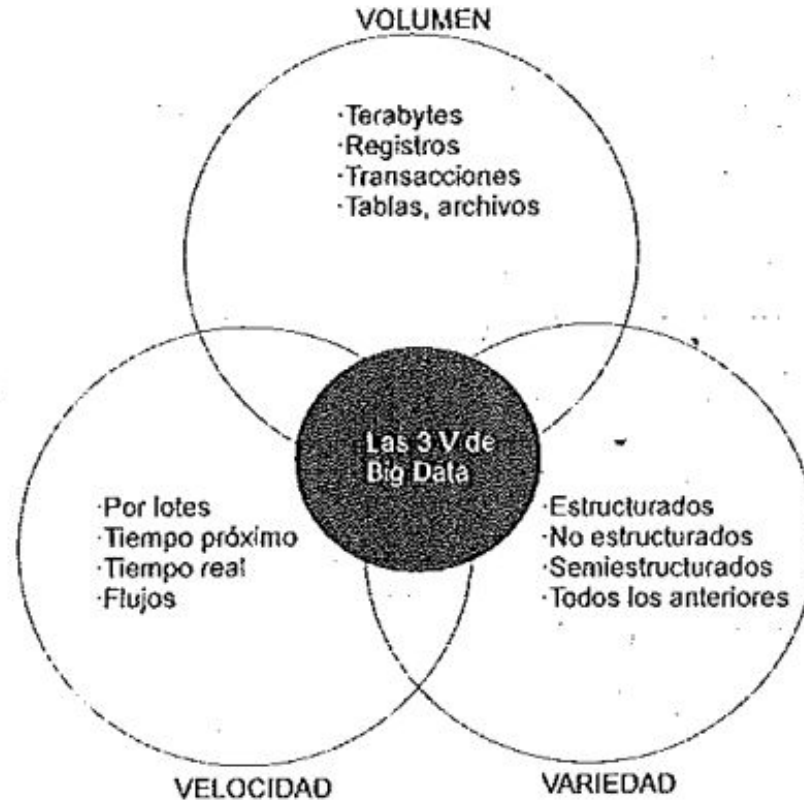


Figura 1. Las 3V del Big Data.

Tomado de [1]

4. Bases de Datos NoSQL

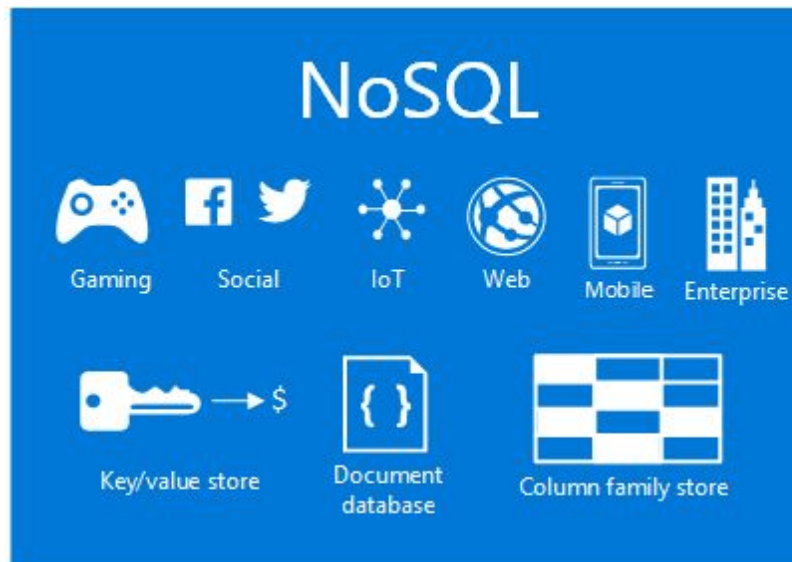
Las bases de datos NoSQL son sistemas de almacenamiento de información que no cumplen con el esquema entidad-relación. [7]











NoSQL

5. Características BD NoSQL

- Escalabilidad horizontal
- Habilidad de distribución
- Uso eficiente de recursos
- Libertad de esquema
- Modelo concurrencia débil
- Consultas simples



6. Tipos de BD NoSQL

Type	Example	
Key-Value Store	 redis	 riak
Wide Column Store	 HBASE	 cassandra
Document Store	 mongoDB	 CouchDB relax
Graph Store	 Neo4j	 The Distributed Graph Database

7. Teorema CAP



(Consistency **A**Availability **P**Partition tolerance):

En el 2000, Eric Brewer propuso la idea de que en un entorno distribuido un sistema no puede mantener continuamente consistencia perfecta, disponibilidad y tolerancia partición simultáneamente. [5]

Figura 2. Base de datos orientada a grafos. Tomado de [6]

8. Neo4J

Neo4j es una base de datos que almacena los datos en un modelo de grafo de propiedades.



9. Conceptos Sobre Neo4J

Elementos Principales

- Vértices (Nodos).
- Aristas (Relaciones)

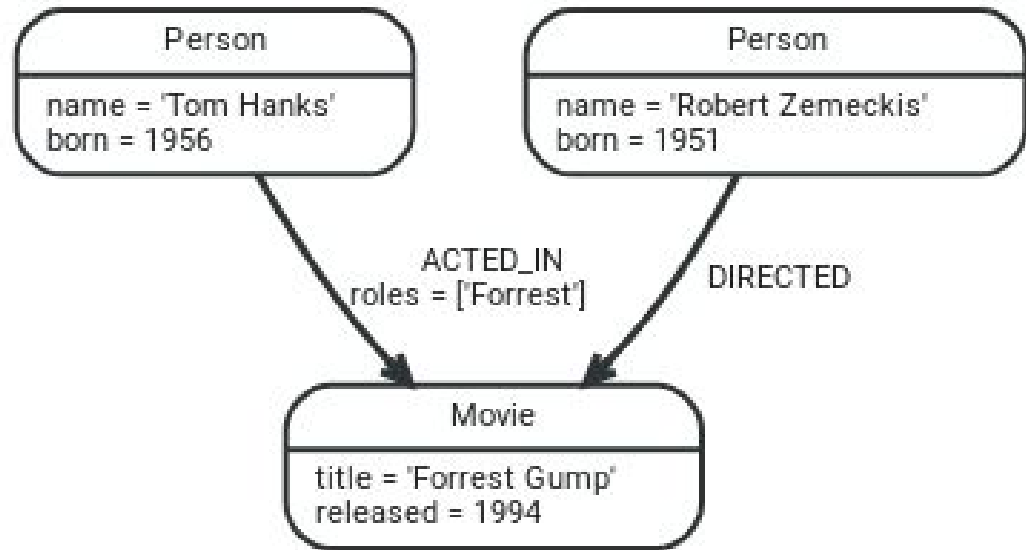


Figura 7. Ejemplo de Nodos con diferentes propiedades y relaciones. Tomado de [9]

9. Conceptos Sobre Neo4J

Tipos de Datos en Neo4J

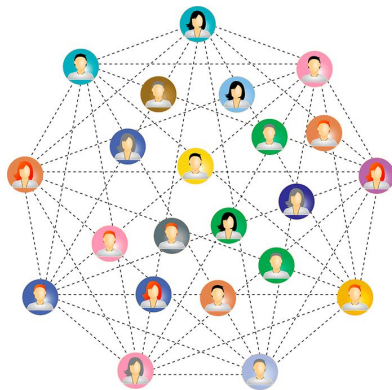
Valores de tipo Numéricos / Valores de tipo Booleanos / Valores de tipo String

Type	Description	Value range
<code>boolean</code>	binary logic value	true/false
<code>integer</code>	64-bit integer	-9223372036854775808 to 9223372036854775807, inclusive
<code>float</code>	64-bit IEEE 754 floating-point number	-
<code>string</code>	sequence of Unicode characters	infinite

Figura 8. Rango valores tipos de datos Neo4j. Tomado [9]

10. Características de Neo4J

- **Rendimiento:** Responden a las consultas actualizando el nodo y la relaciones de esa búsqueda y no todo el grafo completo.
- **Agilidad:**
 - +34 mil millones de nodos.
 - +34 mil millones de relaciones.
 - +68 mil millones de propiedades.
- **Flexibilidad y Escalabilidad:** Aumento de necesidades más nodos, más relaciones.



11. Aplicaciones

Detección Fraude

Las redes de fraude tienen mecanismos para delinquir que no son detectables con el análisis lineal de los datos. Pero con un análisis escalable de las múltiples relaciones entre los datos, esto es mucho más fácil.

Entre el 10% y el 20% de la deuda sin respaldo en los bancos líderes tanto en EEUU como en Europa se debe a este fraude.[9]



11. Aplicaciones

Recomendaciones y Redes Sociales

Conecta de forma eficaz a las personas con otros productos y servicios, en función de la información personal, sus perfiles en redes sociales y su actividad online reciente.



11. Aplicaciones

Gestión en Centros de Datos

Permite gestionar, monitorizar y optimizar todo tipo de redes físicas y virtuales pese a la gran cantidad de datos.



Demo



Referencias

[1] Joyanes, L., (2013), *Big Data Análisis De Grandes Volúmenes De Datos En Organizaciones*, México DF, México, Alfaomega Grupo Editor.

[2] Fernández, E. P. (2017). *Big Data: Eje Estratégico En La Industria Audiovisual*. Recuperado de <http://ebookcentral.proquest.com>

[3] Hernández, E., Duque, N., Moreno, J.,(15 de marzo de 2017). Big Data: una exploración de investigaciones, tecnologías y casos de aplicación, *TecnoLógica*. Recuperado de <http://www.scielo.org.co/pdf/teclo/v20n39/v20n39a02.pdf>

[4] Camargo, J., C. Ortega, J., Joyanes, L. (1 de diciembre de 2014). Conociendo Big Data. *Facultad de Ingeniería*. Recuperado de <http://www.scielo.org.co/pdf/rfing/v24n38/v24n38a06.pdf>

[5] Romero, A. C., Sanabria, J. S. G., & Cuervo, M. C. (2012). *Utilidad y funcionamiento de las bases de datos NoSQL*. *Facultad de Ingeniería*, 21(33), 21-32.

Referencias

- [6]** Pérez, M. A. C. (2017). NoSQL,¿ es necesario ahora?. Tecnología Investigación y Academia, 5(2), 174-179.
- [7]** Martín, A., Chávez, S. B., Rodríguez, N. R., Valenzuela, A., & Murazzo, M. A. (2013, June). Bases de datos NoSQL en cloud computing. In XV Workshop de Investigadores en Ciencias de la Computación.
- [8]** Antiñanco, M. J. (2014). Bases de Datos NoSQL: Escalabilidad y alta disponibilidad a través de patrones de diseño (Doctoral dissertation, Facultad de Informática).
- [9]** Neo4j, Inc. (2017). The Neo4j Developer Manual, (v3). Recuperado de <http://neo4j.com/docs/developer-manual/current/>