

LABORATORIO II - Trabajo Práctico 1 Bases de Datos NoSQL

# Base de datos orientada a **Grafos**

Profesor: Juan Lopreiato

Integrantes:

- Bautista Bullejos
- Julián Martínez Fontañá
- Carla Montani
- Axel Speroni

## Índice

o1. Introducción .....	3
o2. Definición conceptual - Noción general del tema .....	3
o3. Ventajas - Desventajas .....	5
o4. Para que se utiliza .....	6
o5. Ejemplos de uso reales .....	7
o6. Motores de ejemplo .....	9
o7. Diferencias con otras DDBB .....	11
o8. Escalabilidad .....	12
Bibliografía .....	14

## 01. Introducción.

En los últimos años la necesidad de procesar grandes volúmenes de información en poco tiempo para el uso cotidiano en las redes sociales y dispositivos móviles llevo el desarrollo de nuevas soluciones tecnológicas, que sin desafiar la importancia de las bases de datos relacionales, buscan estructurar una forma diferente de almacenar datos. Las Bases de Datos NoSQL, nombre en el cual podemos incluir las bases de datos clave-valor, las documentales, las basadas en columnas y las basadas en grafos, surgieron como una solución para los problemas no resueltos eficientemente por la bases de datos relacionales, que no logran procesar satisfactoriamente conjuntos de datos grandes y complejos, las NoSQL (Not only SQL) abarcan diferentes tipos de bases de datos, cada una con sus características intrínsecas que le confieren un comportamiento más o menos apropiados para diferentes escenarios de aplicación.

En este caso vamos a profundizar en el estudio de las características, usos, ventajas y desventajas de las Bases de datos orientadas a Grafos.

## 02. Definición conceptual - Noción general del tema.

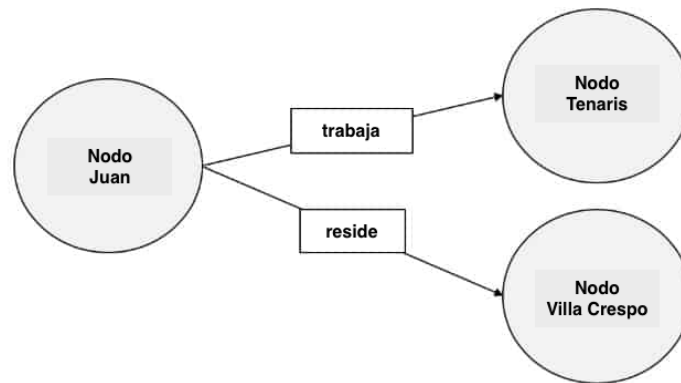
Las **Bases de datos orientadas a Grafos** (BDOG) son bases de datos que se basan en la teoría de grafos, una rama de la matemática y las ciencias de la computación que estudia las propiedades de los grafos, un **grafo** es un conjunto de objetos llamados **vértices** o **nodos** unidos por **enlaces** llamados aristas o arcos, que permiten representar relaciones binarias entre elementos de un conjunto. Los grafos, dentro del contexto de software, representan las entidades como nodos y las aristas por la forma como estos nodos se pueden encontrar relacionados, ya sean entre sí o no.

En este tipo de base de datos, los datos se muestran como entidades o nodos, los nodos pueden tener propiedades y se conectan con otros nodos por medio de relaciones o enlaces (edges) la conexión entre dos nodos son propiedades.

Los grafos son un conjunto de objetos (vértices y aristas) que permite representar datos interconectados, así como las relaciones entre ellas, de forma comprensible y como un único y más amplio conjunto de datos. Los grafos están formados por nodos o vértices, que son propiedades de datos u objetos claramente señalizados e identificables, y **aristas** o **arcos**, que representan las relaciones entre los objetos. Gráficamente, estos dos componentes tienen forma de puntos y líneas, respectivamente. Las aristas tienen un extremo inicial y uno final, mientras que cada nodo siempre contiene un número concreto de relaciones a otros nodos, ya sean de entrada o de salida.

Los nodos representan una entidad en la que almacenaremos piezas de datos o atributos de tipo clave-valor, mientras que las **aristas** representan cómo los nodos se conectan y asocian entre sí. Las relaciones también tienen nombre y dirección, de tal modo que cada relación siempre tiene un nodo de inicio y un nodo de fin.

Por ejemplo: Juan(nodo) trabaja en(enlace) Tenaris(nodo) y vive en (enlace) Villa Crespo



-

Grafos: relaciones nodos y aristas

Las relaciones también tienen nombres en función de los nodos que conectan. Esta variación en cómo funcionan las relaciones también causa que haya dos subtipos de bases de datos orientadas a grafos.

### **Labeled-Property Graph** (grafo de propiedades etiquetadas).

Se asignan propiedades (properties) concretas tanto a los nodos como a las aristas, cada nodo contiene todas sus propiedades contenidas dentro de sí mismo y las relaciones solo sirven para conectar distintos tipos de entidades. Un ejemplo simple de esto es un Nodo Empleado que tiene los atributos de un empleado como su legajo, nombre y dirección conectado al nodo Departamento, que tiene la descripción y dirección del departamento, donde la relación se llamaría “Trabaja En” y partiría desde el nodo empleado hasta el nodo Departamento, de tal modo que se lea Empleado X trabaja en Departamento Y.

### **Resource Description Framework** (grafo de descripción de recursos, RDF)

La estructura del grafo se regula mediante tripels y quads: los tripels se componen de tres elementos, siguiendo el esquema nodo-borde-nodo. Los quads complementan a los tripels con información de contexto adicional, facilitando así su separación en grupos. Los nodos optan un modelo más dinámico, donde tienen 1 o unos pocos atributos dentro de sí y cualquier otro atributo que se quiera añadir deberá ser agregado como un nodo con una relación específica para el tipo de atributo que es. Por ejemplo, si decimos que existe el nodo entidad Empleado que solo tiene el legajo del empleado, si después queremos añadir su nombre tendremos que crear el nodo de atributo Nombre y su relación sería Es Nombre, donde el nodo empleado apunta al nodo nombre.

### 03. Ventajas y Desventajas

Ventajas	Desventajas
<b>Rendimiento</b> pueden manejarse grandes cantidades de datos relacionados de forma rápida y efectiva. Las relaciones están guardadas en la propia base de datos, por lo que no son calculadas a partir de cada solicitud de búsqueda. Gracias a ello, la base de datos opera a gran velocidad incluso en búsquedas complicadas.	<b>Inexistencia de lenguaje estandarizado</b> cada plataforma de BDOG tiene un lenguaje de peticiones diferente por lo que no hay un lenguaje estandarizado para realizar peticiones a diferencia de las bases de datos relacionales con el SQL ( Structured Query Language ) Los más utilizados son PGQL, Gremlin, SPARQL, AQL, etc.
<b>Prevención de fraudes</b> gracias al análisis de las relaciones en grafos se pueden detectar de forma efectiva patrones de blanqueo y otros fraudes.	<b>Baja comunidad</b> al ser un tipo de BBDD bastante reciente, la comunidad aún no es muy grande y, por lo tanto, cuesta encontrar soporte a los distintos problemas que se pueden dar.
<b>Flexibilidad</b> La teoría de grafos en que se basan las BDOG permite solucionar múltiples problemas encontrando la solución más óptima. Facilidad para representar relaciones, dinamismo y flexibilidad para agregar nuevos datos.	<b>Ineficiencia para peticiones transaccionales</b> las bases de datos de grafos no funcionan bien para peticiones transaccionales a diferencia de las BBDD relacionales.
<b>Escalabilidad</b> las bases de datos orientadas a grafos permiten una buena escalabilidad gracias a que se pueden añadir nuevos nodos y nuevas relaciones entre ellos. Estructuras flexibles y ágiles	<b>Escalabilidad</b> puesto que están diseñadas especialmente para arquitecturas de un solo servidor, el crecimiento supone un desafío (matemático).
<b>Velocidad</b> la velocidad de búsqueda depende únicamente del número de relaciones concretas, no del conjunto de datos. Resultados en tiempo real.	<b>Altos volúmenes de datos</b> Ineficiente al manejar altos volúmenes de datos, para estos casos se debería usar con una combinación de DB Relaciones.
<b>Presentación</b> intuitiva y resumida de las relaciones.	<b>Búsqueda</b> La velocidad de búsqueda de nodos físicamente distribuidos es bastante lenta.

#### **04. Para que se utiliza**

El uso de una BDOG es una respuesta a la dificultad de representar los diversos sistemas complejos que caracterizan al mundo actual, y a la necesidad de tener un alto rendimiento de un sistema que se encuentre involucrado en contextos con altos volúmenes y concurrencia de datos.

##### **Redes sociales**

Las personas o grupos corresponden a nodos en una base de datos orientada a grafos, y las formas cómo interactúan dichos individuos generan las distintas relaciones de los mismos permitiendo así predecir sus comportamientos.

##### **Ecommerce**

Las redes neuronales de grafos, pueden ayudar también a nivel de las propuestas de ecommerce. Como ya comentamos, las recomendaciones suman un valor importante al tener contexto sobre el usuario, pero, con las redes neuronales se pueden generar predicciones en tiempo real tomando en cuenta los hábitos de compra del usuario para así poder impulsar diferentes tipo de acciones de conversión. Estos modelos de recomendación aprovechan el aprendizaje automatizado para proveer soluciones de valor. Diversos retails del mundo aprovechan estos modelos de soluciones basados en grafos para proponer de forma eficiente, nuevos productos a los usuarios.

##### **Recomendaciones**

Los algoritmos de recomendación establecen las relaciones entre individuos y los servicios a los que pueden estar sometidas las personas. Ya sea al momento de realizar una lectura de interés, la visualización de algún vídeo, las compras que realice una persona o las diversas variedades de consumo de los individuos tienden a establecer un interés en algún tema en particular generando una conducta de la cual se pueden abstraer y almacenar múltiples relaciones para su posterior recomendación.

##### **Geolocalización**

Las distintas operaciones geoespaciales dependen de estructuras de datos específicos, las cuales se pueden representar de una forma jerárquica; dicha representación facilita los cálculos de rutas o cualquier obtención de información entre las ubicaciones en alguna red específica tales como la de carreteras, ferroviaria o espacio aéreo. Las aplicaciones geoespaciales de las bases de datos orientadas a grafos son especialmente relevantes en las áreas de telecomunicaciones, logística, viajes, horarios y planificación de rutas. Las empresas que gestionan envíos les interesa calcular la ruta más rápida y eficiente

para aumentar el número de entregas por hora. Calcular la ruta más rápida puede hacerse usando el algoritmo de Dijkstra el cual se aplica en grafos.

### Controles de acceso

La autorización del uso de recursos o aplicaciones por parte de los diferentes tipos de usuarios basados en sus roles en un sistema permite que dicho flujo de información pueda ser representada mediante la utilización de grafos. La tecnología de grafos nos permite capturar en tiempo real, conexiones entre entidades de datos. También ayudan a detectar anomalías y comportamientos erróneos en los conjuntos de datos que se conectan en una red de TI para evitar de forma eficiente amenazas que pueden perjudicar la integridad de los datos. Numerosas empresas en la actualidad soportan algunos procesos de seguridad en este tipo de bases de datos, para contar con detección en tiempo real y visualización de datos para fortalecer la seguridad de sus datos.

### Política de privacidad de datos

La hiper conectividad de los últimos años, ha representado también un problema en cuanto a la privacidad de los datos. Algunas empresas deben limitar el acceso a datos claves, ya que no siempre dentro de una estructura, todos los usuarios tienen la capacidad o necesidad de acceder a todos los archivos. Los derechos de acceso suelen ser complejos de gestionar, por lo que una solución en grafos representa una estructura más fluida, escalable y confiable para lograr resultados efectivos. Un grafo puede ayudarnos a construir estructuras jerárquicas que sean dinámicas y que los diferentes usuarios puedan consultar dentro del grafo la información a la que tienen acceso, sin perder tiempo en diferentes autorizaciones.

## 05. Ejemplos de uso reales

### AirBnb

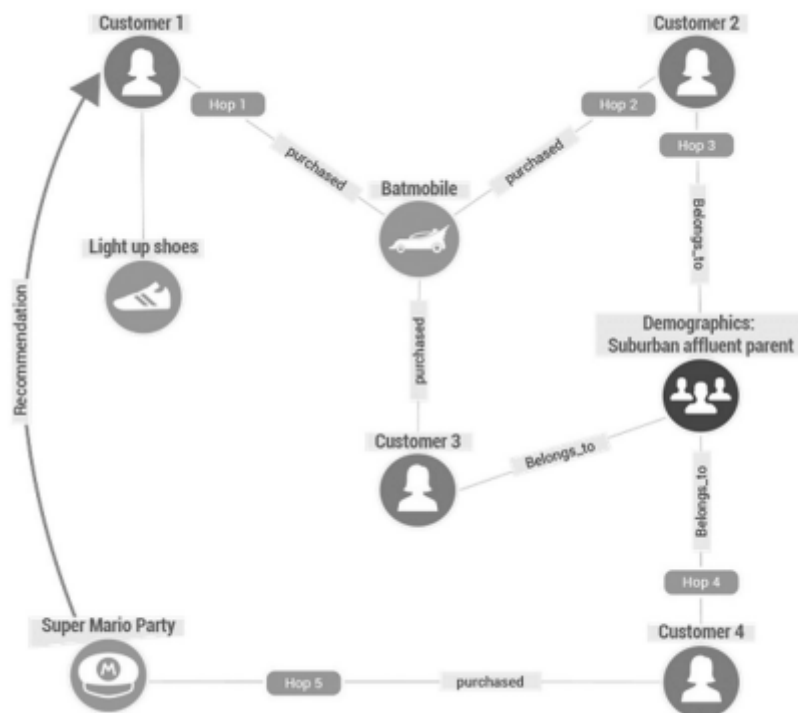
Gracias a la innovación, flexibilidad y prestación eficiente de servicios, Airbnb creció de forma muy rápida, los recursos de datos internos y externos también se incrementaron de forma exponencial y se volvieron un aspecto restrictivo e inmanejable, se determinó que los datos de Airbnb estaban siendo almacenados en ubicaciones restringidas y que carecían de un contexto adecuado para su análisis. Los datos no estaban democratizados y se hacía casi imposible contar con resultados totalmente confiables y fiables. Para eso, nació el proyecto de creación de **dataportal**.

Dataportal fue pensado para ser un motor de búsqueda de recursos de datos, donde las interacciones rápidas, detalladas y precisas contribuyen a fomentar la

exploración. Neo4j, ayuda en este caso a buscar relaciones y conexiones en cuestiones de segundos dentro de los millones de datos que integran las bases de datos de Airbnb.

### Wish

La plataforma tuvo un alto crecimiento en ecommerce a nivel mundial y quería mostrar a sus usuarios recomendaciones personalizadas, realizando consultas a la base de datos en tiempo real mostrando las correctas opciones que se adapten a sus usuarios. Mediante **TigerGraph** lograron mejorar la experiencia del usuario con recomendaciones personalizadas en menos tiempo y mejorar las ventas.



-  
Wish database TigerGraph

### Amazon

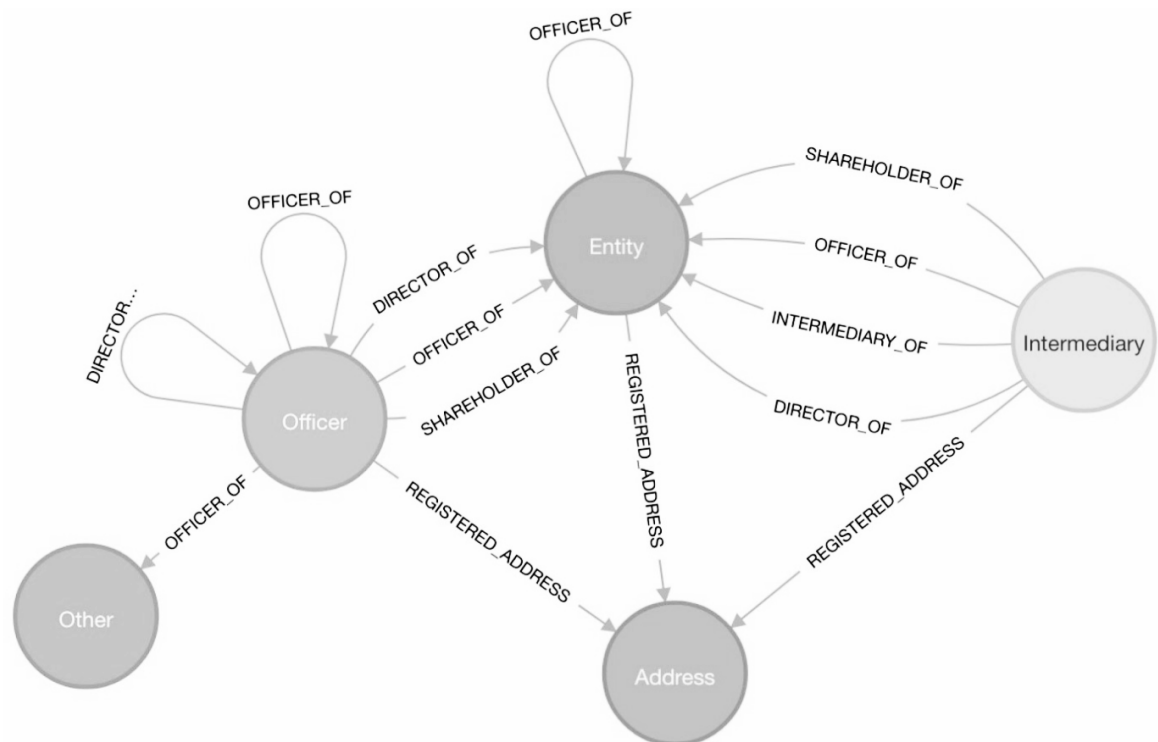
Amazon Neptune es un servicio de base de datos orientada a grafos completamente administrado que permite fácilmente integrar aplicaciones que funcionan con conjuntos de datos altamente conectados. Amazon Neptune es compatible con los lenguajes de consultas Apache TinkerPop Gremlin y SPARQL, los cuales permiten al usuario crear consultas para navegar por conjuntos de datos. Amazon Neptune es la solución ideal para casos de uso tales como crear motores de recomendaciones, ya que permite almacenar las relaciones que existen entre clientes y su historial de compras y así poder hacer consultas que



permitan emitir recomendaciones personalizadas. Con Amazon Neptune también se pueden crear consultas que permitan identificar patrones de fraude en transacciones financieras que contengan elementos en común con transacciones fraudulentas conocidas como correo electrónico, nombre de la persona o dirección IP. Otros casos de uso de Amazon Neptune incluyen su aplicación en redes sociales, operación de redes/T.I. y ciencias de la salud.

### Panama Papers

Un ejemplo (Detección de fraude) lo tenemos en los “Papeles de Panamá”, en el que tenemos más de 2.6 Terabytes de información, siendo complejo y costoso analizar toda esta información. Con un modelo basado en grafos, nos facilita el trabajo pudiendo seguir la pista, sobre todo si hay una representación visual de los datos, ejecutando consultas y búsquedas de los elementos, llegando a establecer las relaciones por donde haya podido circular el dinero.



-

Panama Papers database Neo4j

### o6. Motores de ejemplo (open source y/o pagos)

Existen varias plataformas, tanto de pago como gratuitas que permiten incorporar este tipo de base de datos.

### Neo4j

Esta BBDD orientada a grafos es muy popular en el mundo Big Data utilizada por empresas como **Ebay, Walmart** o **IBM**. Está pensada para manipular datos conectados de forma rápida y sencilla ayudando al análisis de la información. Tiene una gran comunidad por lo que es ideal para empezar a usar este tipo de bases de datos y está concebida como modelo de código abierto.

### ArangoDB

ArangoDB es una de las BDOG multi modelo **open source** más usadas del momento. Combina el modelo de grafos con el de clave-valor además de tener un completo motor de búsqueda de texto dando al usuario la posibilidad de combinar diferentes modelos de datos y realizar consultas a través de un lenguaje declarativo parecido a SQL.

### Amazon Neptune

Amazon Neptune es una base de datos de grafos de alto rendimiento con una baja latencia de realización de consultas. Permite almacenar información con millones de relaciones además de poder realizar predicciones de datos usando el módulo de aprendizaje autónomo Amazon Neptune ML. Puede usarse a través de la nube pública de Amazon Web Services y se abrió al público en 2018 como base de datos de alto rendimiento.

### FlockDB

FlockDB es una base de datos **open-source** orientada a grafos pensada para escalar de forma horizontal y no tanto de forma vertical. Al tener una menor capacidad que otro tipo de BBDD es usada en proyectos que requieren operaciones con una alta rapidez y baja latencia. Fue usada en sus inicios por Twitter para establecer relaciones entre los usuarios de su plataforma.

### SAP Hana Graph

SAP ha creado con SAP Hana una plataforma basada en un sistema de gestión de bases de datos relacional que se completa con el modelo integrado SAP Hana Graph, orientado a grafos.

### OrientDB

Esta graph database está considerada como uno de los modelos más rápidos disponibles actualmente.

### TigerGraph

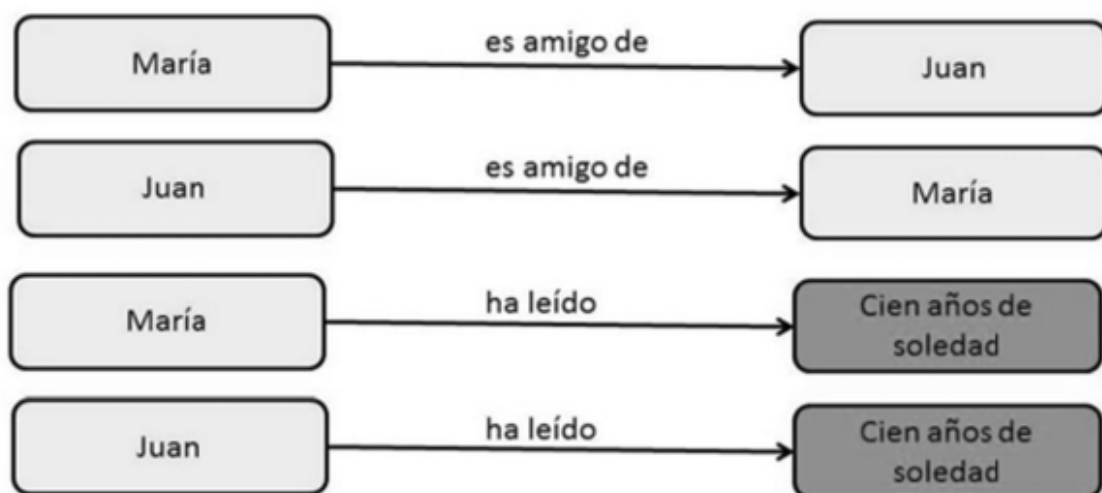
TigerGraph ofrece una amplia plataforma de gestión de bases de datos orientadas a grafos. Especialmente pensada para ser un acompañante en las soluciones

empresariales, este motor de grafos nos permite realizar análisis de conexiones entre datos a profundidad en tiempo real. Esto permite establecer mejores y más fiables modelos de toma de decisiones basadas en datos.

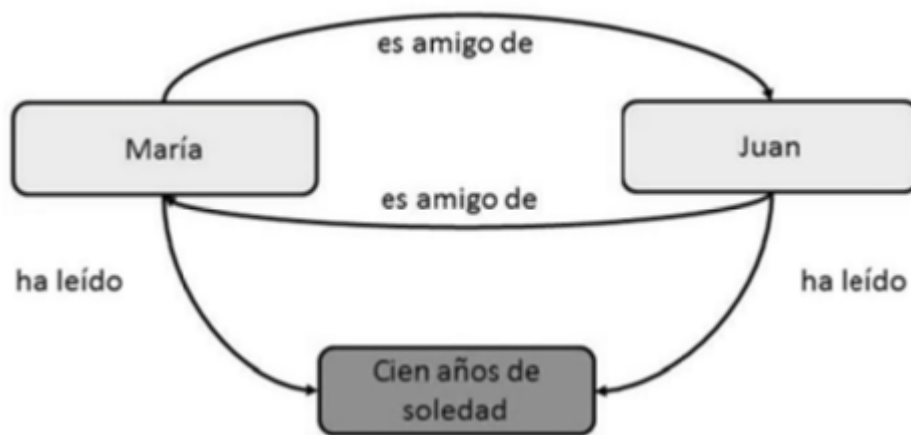
## 07. Diferencias con otras DDBB

Las bases de datos relacionales han sido las bases de datos estándar desde su nacimiento en 1970. A diferencia de las de grafos, las bases de datos relacionales tienen una estructura más compleja y orden más rígido, se basan en tablas que gestionan las relaciones de los registros de datos en cada línea. Las columnas, por su parte, contienen propiedades con diferentes valores de atributo. Además de por su estructura, las bases de datos relacionales también se diferencian esencialmente de las de grafos por su modo de funcionamiento. Para poder presentar y guardar las relaciones entre datos con interconexiones complejas, es necesario hacer cálculos con varias tablas superpuestas, lo cual suele suponer mucho tiempo y esfuerzo si se trata de un conjunto grande.

Las consultas en bases de datos relacionales normalizadas utilizando SQL (lenguaje de consulta estructurado) se realizan empleando la sentencia JOIN entre tablas, la cual une las diferentes tablas normalizadas para extraer los datos que se requieren. La dificultad se presenta si se desea obtener información cuando las entidades son altamente relacionadas con otras, pues implica construir consultas complejas difíciles de mantener y entender, que resultan muchas veces en la sobreutilización de recursos en los motores de bases de datos relacionales cuando se ejecutan dichas consultas. Para solucionar esto, los diseñadores de bases de datos deben encargarse de optimizar sus modelos y se ven obligados a adaptarlos a esta limitante, por eso los modelos de bases de datos muchas veces no representan claramente y de manera natural la lógica de negocio.



Mientras que las bases de datos basadas en tablas solo usan el lenguaje de consulta SQL (Structured Query Language), las bases de datos modernas del tipo NoSQL cada vez se alejan más de este lenguaje y del concepto relacional que va unido a él. En este enfoque también se incluyen las bases de datos de grafos como parte de la familia NoSQL. Además de las graph databases, muchos otros modelos también pertenecen a este grupo: las bases de datos key-value, las bases de datos orientadas a columnas y las bases de datos orientadas a documentos, por ejemplo. Se trata de repositorios de datos que gestionan y almacenan sobre todo conjuntos de datos estructurados, si bien menos interconectados.



-

Relaciones con Modelo de datos mediante grafos

En las BDOG los datos no se adhieren estrictamente a un tipo de dato, es decir nodos y relaciones similares y que pueden ser identificados como un tipo de datos por que comparten un conjunto de atributos, también pueden tener atributos propios y únicos respecto a otros nodos similares. La libertad en los esquemas permite de asociar la información en las diferentes formas de representarlos en nodos, propiedades o relaciones, permitiendo realizar cambios durante la evolución del modelo de la lógica de negocio, por ejemplo, una entidad asociado con un dato que puede constituir una propiedad en una entidad, también se puede modelar como una entidad completamente independiente.

## o8. Escalabilidad

La escalabilidad en bases de datos es un factor clave para quienes se desempeñan en el campo tecnológico. Para las bases de datos es uno de los procesos decisivos que inclina a

las organizaciones a elegir modelos, gestores e inclusive, sirve para determinar la cantidad de inversión que se debe realizar en equipamiento.

El concepto de escalabilidad se refiere a la capacidad que posee un software para adaptarse a las necesidades o las demandas de rendimiento a medida que el número de usuarios y operaciones crece. Generalmente al incrementarse ampliamente las transacciones en un sistema, el funcionamiento de las bases de datos empieza a verse afectado.

Sin embargo, cuando nos adentramos en el ámbito de las bases de datos orientadas a grafos, podemos encontrar un problema en la escalabilidad. Y es que su principal problema para escalar reside en su propio diseño: al estar diseñadas específicamente para un solo servidor, el crecimiento del volumen de datos acarrea un gran desafío matemático tanto para la base de datos como para quien la implementa.

### **Preguntas**

- **¿Qué es un nodo para las Bases de datos orientadas a Grafos (BDOG) ?**

1. Representan una entidad en la se almacenan piezas de datos.
2. Son una serie de puntos que podemos encontrar en una línea.
3. Son los vértices o elementos del Árbol.
4. Permiten conectar los diferentes equipos instalados en campo.

- **¿En qué caso aplicarías BDOG?**

1. Ecommerce
2. No code
3. Base de datos de un banco
4. Microemprendimiento

- **¿Cuáles son los beneficios de usar BDOG?**

1. Rendimiento manejan grandes cantidades de datos forma rápida
2. Gran soporte de de la comunidad para resolver problemas
3. Utilizan el lenguaje estandarizado SQL
4. Eficiencia para las peticiones transaccionales

- **¿Cuál empresa - institución usa BDOG?**

1. Wish
2. McDonald's
3. Anses
4. Mercado Libre

## **Bibliografía**

- <https://www.ionos.es/digitalguide/hosting/cuestiones-tecnicas/graph-database/>
- <https://icc.fcen.uba.ar/bases-de-datos-en-grafos-cuando-las-apariencias-no-enganan/>
- Bases de datos orientadas a grafos. Pinilla, C.; Bello, M.; Peña, C. (2017).
- <https://abdatum.com/informatica/base-datos-orientada-grafos>
- <https://www.grapheverywhere.com/comparativa-bases-de-datos-de-grafos-cual-es-mejor-cual-elegir/>
- <https://www.grapheverywhere.com/10-casos-de-uso-reales-basados-en-tecnologia-de-grafos/>