**TRABAJO PRÁCTICO Nº 3 – TIPOS DE BASES DE DATOS**

**Consigna**

El objetivo del presente trabajo consiste en aprender sobre bases de datos NoSQL. Para ello, partiendo de la información desarrollada en clase y los contenidos básicos desarrollados en este documento, los estudiantes organizados en grupo deberán profundizar los conocimientos sobre estas bases de datos.

La entrega consiste en un documento que debe contener información genérica sobre las bases de datos NoSQL, características, diferencia con las bases de datos relacionales, tipos, etc.

Una vez desarrollada la información general, deberán describir los distintos tipos de bases de datos especificando cómo funcionan, los SGBD con las cuales crearlas, para que son más usadas y dentro de lo posible pequeños ejemplos en cada caso.

Finalmente, teniendo en cuenta el TP2, deberán indicar con qué tipo de base de datos llevarían a cabo el almacenamiento en la propuesta BIG DATA.

**Sugerencias para el análisis del tema**

* Concepto. Entender el funcionamiento
* Describir sus ventajas y desventajas
* Describir cómo implementarlas
* Describir casos de aplicación
* Requerimientos técnicoss

**Contenido**

**Tipos de Bases de datos.**

**Clave – Valor**

Las bases de datos clave-valor es un tipo de base de datos no relacional (NoSQL) que emplea el paradigma clave-valor, una forma de almacenar datos pensados para editar, recuperar y eliminar todo tipo de información.

Este manera de definir datos es usado en múltiples lenguajes de programación. En Javascript tenemos el formato JSON, en Python tenemos los diccionarios y en Java o C++ tenemos los Mapas.

La información se guarda en lo que se conoce como tablas hash, donde se almacenan las claves con los correspondientes punteros que apuntan a la parte de memoria donde se encuentran los valores.

**Documentales**

Una base de datos documental es una de las principales variantes de las bases de datos no relacionales o NoSQL. Se caracterizan por almacenar la información en registros, cada uno de los cuáles funciona como una unidad autónoma de información.

Como su propio nombre indica, las bases de datos orientadas a documentos utilizan documentos para el almacenamiento de todos los registros y los datos asociados a ellos. Cada uno de estos registros puede almacenar distintos tipos de datos. A su vez, los documentos que contienen los registros pueden tener diferentes formatos, desde archivos JSON o XML hasta documentos de texto.

Al ser un tipo de bases de datos no relacionales, otra de sus principales características es que la información no está contenida en tablas. Por el contrario, están pensadas para el almacenamiento de datos semiestructurados, los cuáles se organizan en documentos con valores asignados, los cuáles pueden ser consultados por los usuarios gracias a las diversas herramientas de consulta existentes.

**Orientadas a grafos**

Hay dos modelos comunes de bases de datos orientadas a grafos: grafos de propiedades y grafos RDF. Los grafos de propiedades se centran en el análisis y las consultas, y los RDF se centran en la integración de datos. Ambos tipos de grafos consisten en un conjunto de puntos (vértices) y de las conexiones entre esos puntos (bordes). Sin embargo, también tienen diferencias.

Los grafos de propiedades se utilizan para modelar las relaciones entre los datos y permiten hacer análisis y consultas sobre los datos en función de estas relaciones. Un grafo de propiedades tiene vértices que pueden contener información detallada sobre un tema, así como bordes para resaltar la relación entre esos vértices. Los vértices y los bordes pueden tener atributos (llamados propiedades) a los que están asociados.

Debido a su versatilidad, los grafos de propiedades se utilizan en una amplia gama de industrias y sectores, como finanzas, fabricación, seguridad pública, venta minorista y muchos más.

Los grafos RDF (Resource Description Framework) cumplen con un conjunto de estándares W3C (Worldwide Web Consortium) diseñados para representar declaraciones; son la mejor opción a la hora de representar metadatos complejos y datos principales. Se suelen utilizar para datos vinculados, integración de datos y grafos de conocimiento. Pueden representar conceptos complejos en un dominio, así como ofrecer una semántica rica e inferencias sobre datos.

En el modelo RDF, una sentencia está representada por tres elementos: dos vértices conectados por un borde que representa el sujeto, el predicado y el objeto de una oración; esto se conoce como tripleta RDF. Cada vértice y borde se identifica con un URI exclusivo (identificador único de recurso). El modelo RDF aporta un método para publicar datos en un formato estándar y con una semántica bien definida, lo que facilita el intercambio de información. Las entidades estatales de estadísticas, las empresas farmacéuticas y las organizaciones sanitarias utilizan los grafos RDF de manera generalizada.

Los grafos y las bases de datos orientadas a grafos proporcionan modelos de grafos para representar las relaciones en los datos. Gracias a estas, los usuarios pueden hacer consultas transversales basadas en conexiones y aplicar algoritmos de grafos a fin de encontrar patrones, rutas, comunidades, factores influyentes, puntos únicos de fallo y otras relaciones. Esto permite un análisis más eficiente a gran escala con cantidades masivas de datos. El poder de los grafos reside en la analítica, los conocimientos que aportan y su capacidad para vincular fuentes de datos dispares.

En el análisis de grafos, los algoritmos exploran las rutas y la distancia entre los vértices, la importancia de estos y su agrupación. Por ejemplo, para determinar la importancia, los algoritmos suelen fijarse en los bordes de entrada, la importancia de los vértices circundantes y otros indicadores.

Las bases de datos orientadas a grafos se pueden usar en muchos escenarios diferentes, pero normalmente se utilizan para analizar redes sociales. De hecho, las redes sociales son el caso de uso ideal, ya que en ellas intervienen un gran volumen de nodos (cuentas de usuario) y conexiones multidimensionales (interacciones en muchas direcciones diferentes). En un análisis de grafos de una red social se puede determinar:

* ¿Qué actividad tienen los usuarios? (número de nodos)
* ¿Qué usuarios tienen mayor influencia? (densidad de conexiones)
* ¿Quién tiene el compromiso más bidireccional? (dirección y densidad de conexiones)

Sin embargo, esta información no resulta útil si los bots la sesgan de forma no natural. Afortunadamente, la analítica de grafos puede ser un método excelente para identificar y filtrar bots.

Estas funcionan muy bien para operar con grandes cantidades de datos con complejas relaciones entre ellos.

**Orientadas a Objetos.**

Una base de datos orientada a objetos almacena datos complejos y relaciones entre datos directamente, sin asignar filas y columnas, y esto hace que sean más adecuadas para aplicaciones que tratan con datos muy complejos. Los objetos tienen relaciones “muchos a muchos” y son accesibles mediante el uso de punteros. Estos punteros se vinculan a los objetos para establecer relaciones. Otro beneficio adicional de una base de datos orientada a objetos es que puede ser programada con pequeñas diferencias de procedimientos sin afectar a todo el sistema.