

Modelo de tripletas RDF contra el modelo llave-valor

*Comparación de características sobre estos dos modelos de base de datos no relacionales

1st Alejandro Hernández González
Ingeniería de sistemas y computación
Pontificia Universidad Javeriana Cali
Cali, Colombia
alejo9720@javerianacali.edu.co

Abstract—En este documento trataremos las diferentes características que tienen los modelos para las bases de datos, en este caso analizaremos el modelo de tripletas RDF orientado a grafos, contra el modelo llave-valor(Key-value), expondremos sus ventajas y desventajas, y por último el mejor modelo según los requerimientos que se pidan para la base de datos.

Index Terms—Modelo, Tripletas, Llave-valor, comparación

I. INTRODUCCIÓN

El presente ensayo contiene las características principales de dos modelos los cuales determinan la estructura lógica de base de datos de tal manera que formaliza el modo de almacenar, organizar y manipular los datos, además de la comparación entre ellos, exponiendo así sus ventajas y desventajas. Con el fin de poder escoger cual podrá realizar mejor una tarea a la hora de guardar información.

Para tanto el modelo de tripletas RDF(Triple Store), como el modelo llave-valor(Key-value) se definirá de manera puntual las características más importantes de cada modelo para poder así realizar una comparación más pertinente y objetiva.

II. MARCO TEÓRICO

A. Modelo de tripletas RDF(Triple Store)

El modelo de tripletas se basa en el modelo estándar para la representación de datos en la web el cual lo propuso la W3C(World Wide Web Consortium), que busca "Representar propiedades con nombre y sus valores. Estas propiedades sirven tanto para definir atributos de recursos, como para representar relaciones entre recursos" [1]. Esto nos permite mezclar, exponer y compartir datos estructurados y no estructurados.

Los elementos que guarda el modelo son únicos a nivel global, esto se debe a que cada elemento tiene un IRI(Internationalized Resource Identifier) el cual tiene de la tarea de identificar un objeto o un concepto. Dichos elementos se usan para construir sentencias de la forma: sujeto-predicado-objeto (Tripletas) como se puede ver en la figura 1 en la que el predicado tiene como función unir el sujeto y el objeto, estableciendo una relación, además define una propiedad del sujeto.

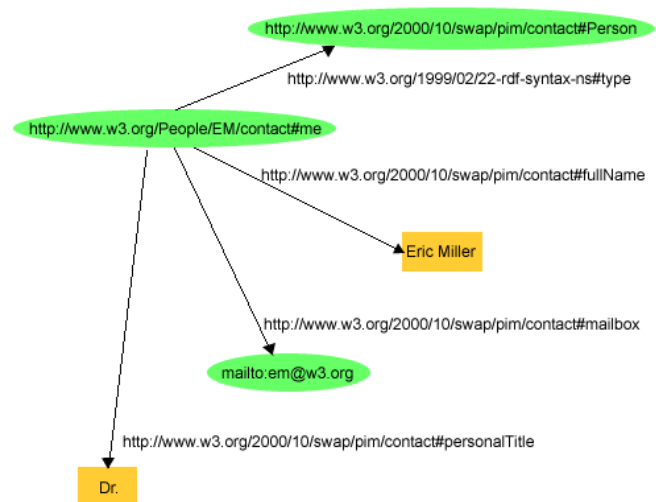


Fig. 1. Grafo que describe a Eric Miller [2]

A partir de esto, al tener un conjunto de tripletas se le podría denominar un grafo RDF, el cual no es un grafo en el sentido matemático, si no que es una forma de presentar gráficamente, donde sujetos y objetos son nodos y las propiedades los arcos. La figura 1 es un gran ejemplo de esto, ya que aquí podemos ver que consta de tripletas que se relacionan entre sí, podemos ver que los arcos son los predicados y los nodos son los sujetos y los objetos.

Este modelo cumple con las propiedades ACID: Atomicidad, la cual es que se deben de ejecutar una serie de pasos para completar la operación de lo contrario no se hace, Consistencia, que se encarga de velar por que se haga aquello que pueda acabar, en este modelo no soporta reglas de consistencia definida por el usuario sino que la encargada de velar por la consistencia es la aplicación, Aislamiento(Isolation), que busca que dos transacciones que se hagan al mismo tiempo no afecte la información a la vez implementando *Snapshot Isolation Model* en la que cada transacción pareciera que estuviera operando en una instantánea y consistente de la base de datos, los cambios que realice sobre la base de datos solo

podrá verlos el hasta el momento que se realice un commit. Y por ultimo, Durabilidad, la cual asegura que si se realiza una transacción esta debe de permanecer aun as el sistema falle.

B. Modelo llave-valor(Key-value)

Este modelo consta de una llave principal que se asocia con un valor. Estos valores pueden ser estructuras complejas como colecciones, diccionarios, arreglos asociados o caches además permiten textos largos, no solo enteros o cadenas cortas, lo cual lo hace ideal para guardar bastante contenido.

El objetivo principal de este modelo es "asegurar la disponibilidad de los datos frente a fallos del sistema o de la red, aprovechando la mas pequena posibilidad de proveer el servicio, aun cuando partes del conjunto de datos falten temporalmente" [3]. Esto nos muestra que el modelo tiene una consistencia débil pero una alta disponibilidad.

Key	Value
K1	AAA,BBB,CCC
K2	AAA,BBB
K3	AAA,DDD
K4	AAA,2,01/01/2015
K5	3,ZZZ,5623

Fig. 2. Ejemplo de modelo llave-valor

Los datos que usualmente son redundantes, se almacenan en una tabla hash simple, a partir de esto hace que todas las operaciones acceden usando la llave primaria. Por lo general existen tres tipos de operaciones: Put, que sirve para almacenar un valor para la llave. Get, que recupera el valor dada una llave. Y por ultimo, Delete que sirve para borrar el valor y la llave dada una llave.

En general no es posible recupera el valor si no se conoce la llave, ya que el valor no tiene ningn esquema definido. En el caso de que no se conozca la llave, es posible obtener una lista de las llaves o en algunos sistemas permite usar indices adicionales. Usualmente, la estructura de los valores son objetos como JSON o XML, la cual se asemeja a la aplicación, ya que esta maneja la integridad de los datos y tambien es la encargada de resolución de conflictos de lectura. Todo esto hace que sea vital la definición de las llaves, por lo que no es recomendable usar solamente un numero secuencial.

Para este modelo se utiliza una consistencia BASE: Basically Available, Soft state y Eventually consistent. La cual nos indica que la base de datos tendrá una buena disponibilidad aunque eventualmente la información sera consistente.

III. COMPARACIÓN

Para empezar es preciso aclaras que ambos modelos son no relacionales o también conocido como noSQL; pero se puede ver una gran diferencia debido a que el modelo de tripletas se basa principalmente en guardar la consistencia, ya que cuenta con las propiedades ACID el cual genera una alta consistencia

debido a sus características. Por otro lado, el modelo de llave-valor usa la consistencia BASE el cual es una mas débil.

A la hora de plantear el esquema de la base de datos el modelo de tripletas se hace mas intuitivo, debido a que representar la información en grafos utilizando RDF es mas sencillo, mientras que el modelo de llave-valor se necesita saber previamente las consultas para poder generar la llave apropiada que cumpla los criterios y así generar el esquema de la base de datos, permitiendo una mejor accesibilidad al momento de realizar las consultas.

A partir de esto sale otra gran diferencia a la hora de realizar las consultas, ya que el modelo de tripletas las consultas pueden ser diferentes y mas amplias. Tambin es preciso denotar que no hay un numero definido de atributos, lo cual lo hace muy dinámico. Además, de tener todas la teora de grafos a su disposición, permitiendo encontrar relaciones entre los datos de una manera mas rápida, haciendo que este modelo sea ideal para el Big Data, ya que descubre patrones en los datos gracias a como se estructuran las consultas para este modelo. Es preciso recalcar que el rendimiento es mucho mayor que el modelo llave-valor, ya que las consultas se realizan de manera que recorre los nodos entre las relaciones(aristas) que posea la base de datos, por lo que el tiempo de ejecución seria proporcional al tamao que deba recorrer para realizar la consulta.

Por otro lado, el modelo llave-valor, es un modelo el cual es ideal para manejar volúmenes de información muy grandes, pero esto hace que sea notoria su principal falencia la cual es la integridad, haciendo que la parte de la validacin sea por parte del usuario. Alguno de sus mas frecuente uso se ve en almacén de sesiones orientado a la web ya que cuando un usuario inicia y cierra sesión, la aplicación almacena todos los datos relacionados con la sesión en la base de datos, y otro uso es en el carro de compras por que permite guardar grandes cantidades de pedidos a la vez y el cambio de estados es muy altos ya que se atiende a una gran cantidad de personas.

IV. CONCLUSIONES

Por lo tanto, a pesar de ser modelos noSQL se puede notar una gran diferencia debido a que el modelo de tripletas se preocupa por guardar la consistencia mientras que el modelo llave valor es un es un poco mas débil.Tambin es preciso retomar que el modelo llave-valor se debe de generar a partir de las consultas que se pidan, mientras que el modelo de tripletas es un poco mas amplio frente a esto, haciendo que su uso dependa de la situacin en la que estemos y la forma en la que queramos guardar la información.

REFERENCES

- [1] Hipertexto, el nuevo concepto de documento en la cultura de la imagen "RDF", Maria Jesus Lemarca Puente.[En línea].Disponible en: <http://www.hipertexto.info/documentos/rdf.htm> [Accedido:12-nov-2018]
- [2] RDF Primer, Frank Manola y Eric Miller.[En línea]. Disponible en: <https://www.w3.org/TR/rdf-primer> [Accedido:12-nov-2018]
- [3] "Sistema de base de datos:Key-Value store", Prof Maria Constanza Pabón.[En línea]. [Accedido:12-nov-2018]
- [4] Washington A. Velsquez Vargas, "Bases de datos orientadas a grafos y su enfoque en el mundo real".[En línea]. Disponible en:<https://bit.ly/2FAnw4y> [Accedido:12-nov-2018]

[5] "Qué es una base de datos clave-valor?".[En línea]. Disponible en:<https://aws.amazon.com/es/nosql/key-value/> [Accedido:12-nov-2018]