

Trabajo Práctico 2 - Bases de datos NoSql

Bases de datos

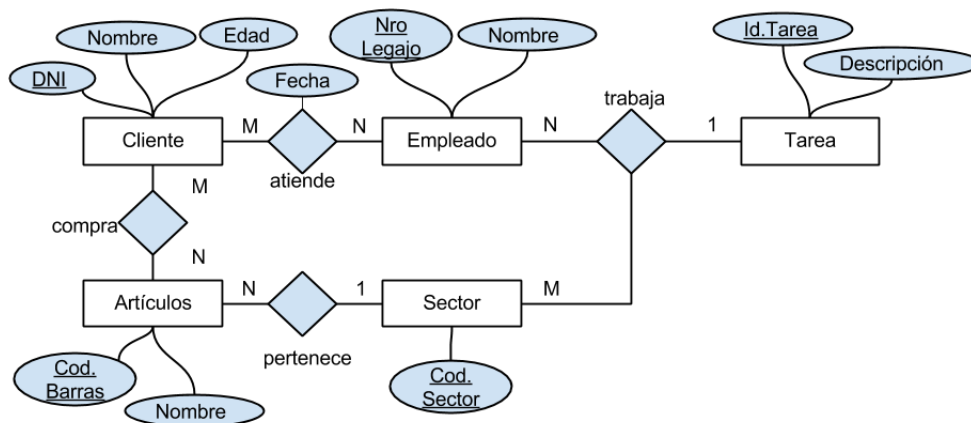
10 de octubre de 2015

Introducción

Este trabajo práctico consiste en realizar distintos ejercicios utilizando una base de datos no relacional. En esta oportunidad, elegimos trabajar con base de datos orientadas a documentos. La implementación que utilizaremos de éstas es *MongoDB*. Para todo el transcurso utilizaremos como referencia su manual oficial¹.

Parte 1 - Desnormalización

Analizar el siguiente DER y realizar un diseño orientado a documentos utilizando desnormalización.



El diseño debe cumplir contestando las siguientes consultas sin utilizar *Map Reduce*:

- Los empleados que atendieron clientes mayores de edad.
- Los artículos más vendidos.
- Los sectores donde trabaja exactamente 3 empleados. Puede haber un empleado que contabiliza para varios sectores.
- El empleado que trabaja en más sectores.
- Ranking de los clientes con mayor cantidad de compras.
- Cantidad de compras realizadas por clientes de misma edad.

¹Éste se encuentra en <http://docs.mongodb.org/manual/>

Parte 2 - Map Reduce

Utilizar los datasets provistos por la cátedra, llamados `disposiciones_201X.json` para los años 2012, 2013 y 2014, y realizar las siguientes consultas utilizando el esquema *Map Reduce*:

- Devolver la cantidad de disposiciones tipo resoluciones que se hayan realizado en Abril del 2013.
- Devolver la cantidad de disposiciones por cada tipo definido.
- Devolver la fecha más citada para todos los informes. Descartar las fechas embebidas en descripción.
- Devolver la mayor cantidad de páginas utilizadas por cada tipo de disposición.

Parte 3 - Sharding

- Dado el siguiente esquema:

```
"personas": {  
  "nombre"      : String,  
  "password"    : String,  
  "codigo_postal" : Integer,  
  "genero"      : String,  
  "edad"        : Integer,  
  "fecha_creacion" : Date  
}
```

Nota: `codigo_postal` es un código de 6 cifras y `fecha_creacion` posee rango a partir de primero de Enero de 2014 hasta la fecha actual.

- Crear una base de datos en MongoDB y configurarla tal que trabaje en un esquema de Sharding para el atributo `codigo_postal`. La cantidad de Shards es a elección. Para realizar este paso lea **tutorial_sharding.txt**. Al crear el índice, utilice uno **simple**.
 - Programar una rutina, a través de la consola de MongoDB, que ingrese datos para este esquema. Cada ítem debe crear un código postal de forma aleatoria. Utilizar `db.people.getShardDistribution()` y `sh.status()` para comprobarlo. Recomendamos ingresar al menos 500.000 registros en intervalos de 20.000.
 - Graficar cómo van cambiando los Shards a través del ingreso de los datos.
 - Realizar los pasos **a.** al **c.** pero utilizando como atributo a `_id` y definiendo éste como índice **hashed**.
 - Comparar ambas configuraciones y gráficos de ingreso de datos.
- Describa al menos tres escenarios posibles donde se defina un atributo para trabajar utilizando Sharding. Justifique.
 - Analice las características que debe poseer un atributo para poder trabajar como clave en un esquema de Sharding.

Parte 4 - Otras base de datos NoSql

Elegir otro tipo de bases de datos NoSql, o sea Key-value, Column o Graph database, y describir cuáles serían los cambios realizados para realizar cada uno de los puntos de este TP. No es necesario realizar cada punto sino describir cómo realizarlo y qué características cambian. Investiguen las diferencias que hay al trabajar con este tipo de base de datos elegida; y las ventajas y desventajas que surgen.

Condiciones de entrega:

La entrega deberá constar, como mínimo, de la siguiente documentación:

- Carátula con tabla de contenidos, fecha y nombre de los autores.
- Introducción y explicación del problema a resolver.
- Detalle de los supuestos asumidos para la resolución del problema.
- Diseño físico correspondiente a la solución y explicación del diseño
- Código y explicación de las consultas así como también de cada decisión tomada.
- Conclusiones.

Fecha de entrega: 11 de Noviembre de 2015