

Replicación de bases de datos NoSQL en dispositivos móviles

AUTOR: GABRIEL CAPDEVILA

DIRECTOR: PROF. MG. JAVIER BAZZOCCO

Motivaciones

En los últimos años aumenta considerablemente la cantidad de dispositivos capaces de conectarse a internet, la cantidad de usuarios y el flujo de información.

Las nuevas arquitecturas distribuidas generaban nuevos problemas en los sistemas RDBMS.

Surgen alternativas NoSQL. Nuevas tecnologías que facilitaban la escalabilidad y el uso de sistemas distribuidos.

Influencia en el desarrollo de aplicaciones móviles.

Objetivos

Realizar un análisis comparativo de la confiabilidad de una aplicación que utilice una base de datos NoSQL con copias locales y sincronización con un servidor remoto con la de una aplicación que utilice bases de datos enteramente locales en el dispositivo.

Los retos de los RDBMS

Estructura

- Estructura de datos debe ser bien definida de antemano

BigUsers

- Debe soportar un gran número de usuarios concurrentes

BigData

- Al manejar mucha información deben ser capaces de distribuirse eficientemente

Escalabilidad

- Debe ser capaz de incrementar el rendimiento agregando recursos

Introducción a NoSQL

Contexto e historia

- BigTable de Google. Lucene. Dynamo de Amazon. Hadoop de Yahoo.
- 2009 NoSQL Meetup organizada por Johan Oskarsson

¿Qué es NoSQL?

- No usan SQL
- Desarrolladas para correr en clusters
- Operan sin esquema

¿Por qué usar NoSQL?

- Modelos de datos más flexibles
- Escalabilidad
- Mayor performance de desarrollo

Tipos de bases de datos NoSQL

Clave-Valor

- Tabla Hashing con dos valores: ID y VALOR
- El valor es un BLOB y es responsabilidad de la aplicación interpretarlo

Column-Family

- Información guardada en columnas. No almacena una columna cuando un valor no existe

Basadas en grafos

- El grafo permite que la información sea almacenada una vez e interpretada según las relaciones
- Permite almacenar entidades (NODOS) y relaciones entre ellas (ARISTAS)

Basadas en Documentos

- Almacena y recupera documentos

NoSQL basados en documentos

La base de datos almacena documentos

- Pueden ser JSON, XML, BSON, etc.
- Estructuras auto descriptivas que pueden contener mapeos, colecciones, etc.

La estructura de los documentos puede variar

- El esquema de datos puede diferir entre documentos de la misma colección

No guarda información innecesaria

- No hay atributos vacíos. Si un atributo no se encuentra se lo considera no relevante.
- Se pueden agregar nuevos atributos sin setearlos previamente

Disponibilidad

- Mejoran la disponibilidad con replicación en sistemas distribuidos

Escalabilidad

- Escalar para lograr más carga de lectura y escritura

Modelado de datos en agregados

Modelo a través del cual percibimos o manipulamos nuestros datos

El modelo relacional toma la información que queremos y la divide en tuplas

NoSQL opera sobre información organizada en estructuras más complejas. NoSQL tiende a formar agregados

Los límites son puestos donde sea necesario dependiendo de la forma en la que se accederá a la información

Los desafíos de desarrollar aplicaciones móviles

Los dispositivos móviles presentan nuevas restricciones a la hora de desarrollar aplicaciones:

Son relativamente pobres en recursos

- Comparados con elementos estáticos

La movilidad es arriesgada

- Los dispositivos portátiles son más vulnerables

La conectividad móvil es variable

- Las redes inalámbricas no siempre funcionan

Se basan en fuentes finitas de energía

- El consumo eficiente de batería es vital

MCC vs Funcionalidad Offline

Para superar la falta de recursos se utilizan soluciones como Mobile Cloud Computing que permite utilizar infraestructura, plataformas y software a bajo costo en la nube.

Tanto el almacenamiento de la información como el procesamiento ocurren fuera del dispositivo. La nube le quita el tiempo de procesamiento a los dispositivos para dárselo a plataformas más poderosas.

Así, los dispositivos móviles no necesitan tantos recursos.

MCC vs Funcionalidad Offline (2)

Las conexiones de datos o Wifi son fácilmente interrumpidas o simplemente no existen en ciertos lugares.

Es importantes diseñar una aplicación que tenga Funcionalidad Offline y permita funcionar cuando la conectividad es interrumpida.

Así, no depende del Wifi o las redes móviles para su funcionamiento.

No puede garantizar que la información esté completamente actualizada en todo momento.

MCC vs Funcionalidad Offline (3)

Por un lado, los servicios web llegaron a tener mucha información para que puedan ser manejados por una sola computadora y se alocan en la nube.

Por otro lado, hay que tener en cuenta la conectividad y el hecho de que no se pueda garantizar una conexión constante. Las soluciones offline a su vez pueden tener problemas de consistencia.

Por estos motivos, no es del todo conveniente manejar NoSQL en ambientes móviles con un enfoque completamente nativo u otro completamente orientado a Cloud Computing. Surge entonces como alternativa, la Replicación.

Replicación

Maneja una BBDD distribuida donde todos los usuarios pueden acceder a la información sin interferir con los demás

Consiste en una copia frecuente de datos desde una BBDD a otra para que todos los nodos compartan el mismo nivel de información.

Los cambios realizados en la información guardada en una locación cualquiera se reflejarán en las demás locaciones.

Porqué usar replicación

Conexiones poco confiables

- Se necesitan copias locales de la información importante en caso que no haya conexión

Disponibilidad y Performance

- Sirve para lograr una mayor disponibilidad usando múltiples copias del sistema servidor
- Mejora la performance al disminuir el tiempo de inactividad y el tiempo de espera

Especialmente cuando el link entre dos sistemas está disponible sólo intermitentemente

Replicación en NoSQL

Las bases de datos NoSQL están diseñadas para arquitecturas distribuidas. La escalabilidad y la consistencia eventual facilitan el uso de replicación.

Stores Clave-valor con replicación: Riak, Tokyo Cabinet, Redis, Scalaris.

Stores basados en documentos: Simple DB, CouchDB.

CouchDB

- Asíncrono
- maestro-a-maestro
- Tolerante a fallos.

Lite Movie Database

El trabajo tiene su eje central en el estudio de distintas técnicas de utilización de bases de datos NoSQL en ambientes móviles.

Se busca realizar un análisis comparativo de la confiabilidad de una aplicación que utilice una base de datos NoSQL con copias locales y sincronización con un servidor remoto con la de una aplicación que utilice bases de datos enteramente locales en el dispositivo

LMD: La Aplicación

Lite Movie Database maneja una base de datos con información de Películas, Series y Personas relacionadas a ellas (actores, directores, guionistas, etc.)

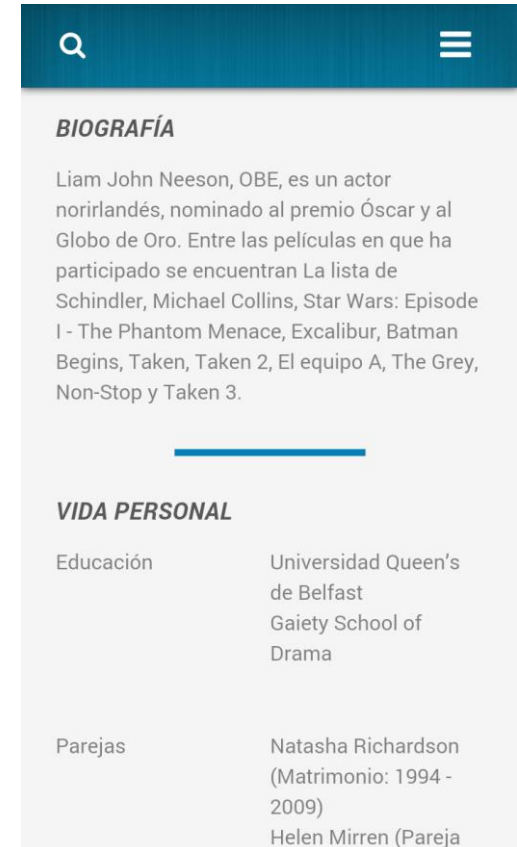
La idea principal es que el usuario pueda buscar dicha información de manera rápida y confiable desde el celular o la tablet.

Dos grandes diferencias con otros servicios similares: Lite y Offline.

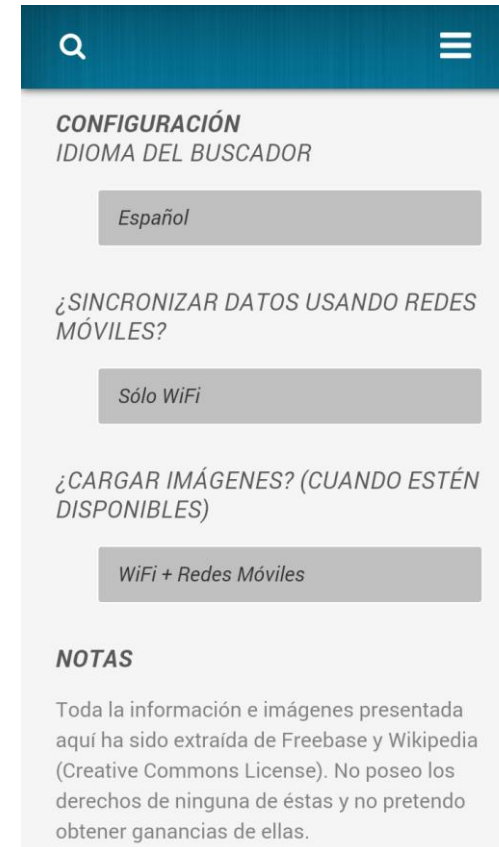
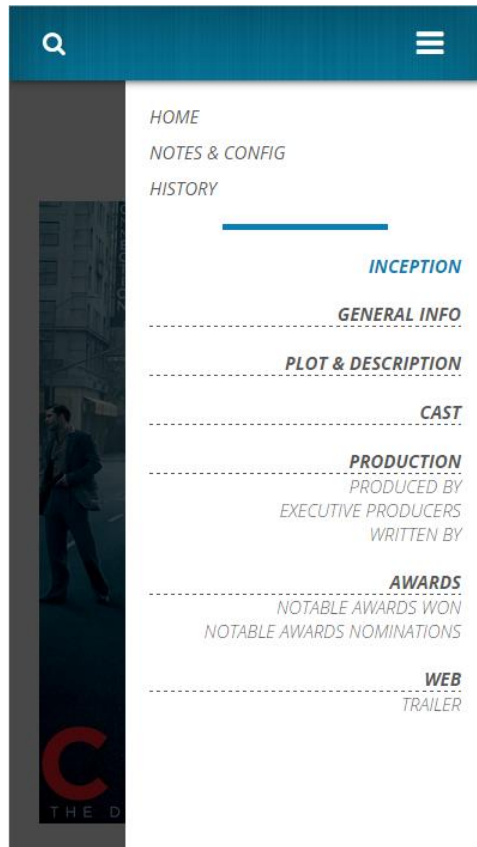
Es Lite porque trae la información absolutamente necesaria y evita el gasto extra de datos.

Es Offline porque guarda los documentos en el dispositivo y se puede acceder sin conexión.

LMD: Funcionamiento



LMD: Funcionamiento (2)



LMD: Tecnologías

PhoneGap

- Framework para el desarrollo de aplicaciones móviles con HTML, CSS y JS
- Renderiza las vistas HTML a una app ejecutable por dispositivos móviles

PouchDB

- Base de datos Javascript inspirada en CouchDB
- Diseñada para correr en el navegador
- Funciona tanto online como offline

CouchDB en couchappy.com

- Una instancia CouchDB en el servidor para replicar desde los dispositivos

LMD: Arquitectura y desarrollo

Desarrollada en HTML, JS y CSS para correr en Android sobre PhoneGap

El almacenamiento interno es enteramente NoSQL

- Contiene internamente dos instancias PouchDB
- La configuración y el historial se guardan como cualquier otro documento

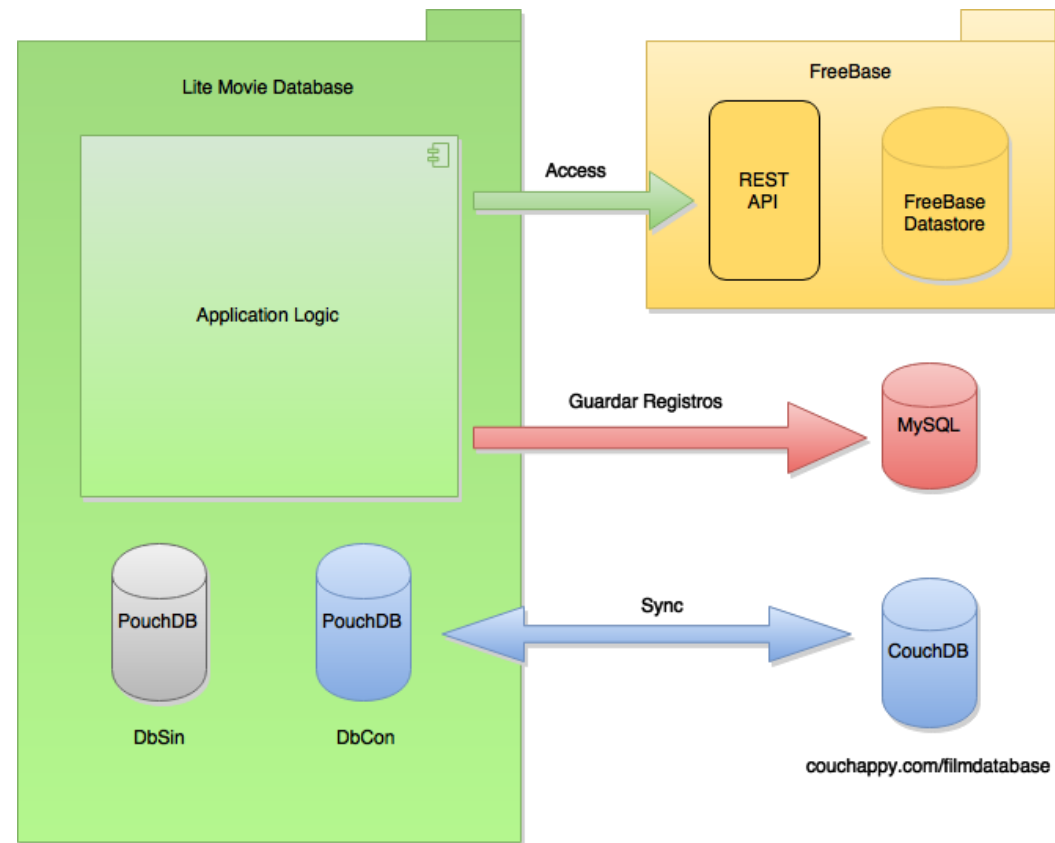
Se accede a los Doc a través del ID

- es-the_matrix-1999-movie
- en-liam_neeson-1952-actor

Solo una base de datos se replica

- Las consultas se hacen sobre ambas bases de datos
- Se busca simular el comportamiento de dos aplicaciones distintas donde una solo tenga los documentos buscador por el usuario y la otra se vaya completando automáticamente con replicación

LMD: Arquitectura y desarrollo (2)



LMD: Distribución

Se distribuyó la app en el playstore y fue descargada por 150 personas de las cuales 75 aportaron datos de sustento.

Según el uso de cada usuario se iba registrando el resultado de los accesos a las bases de datos, marcando cuando había acierto y cuando no.

Se corrieron los Tests sobre la herramienta RGA9 de ReliaSoft utilizando el modelo de confiabilidad NHPP.

LMD: Datos obtenidos

Del 11/03/15 al 22/05/15. 1752 hs (73 días)

Se obtuvieron 22.052 registros de los 75 usuarios.

Se obtuvieron los siguientes datos:

	WIFI	DATA	NO	TOTALES
DbSin aciertos	2453	394	289	3136
DbSin fallas	4509	433	2953	7895
DbSin total	6962	827	3242	11031
DbCon aciertos	3930	429	2658	7017
DbCon fallas	3031	393	580	4004
DbCon total	6961	822	3238	11021

LMD: Tests

	DMTBF 1752 Hr	CNOF 1752 HR	IMTBF 10000 Hr	CNOF 10000 Hr
DbSin	0,1833 Hr	7895	0,1270 Hr	65030
DbCon	0,5109 Hr	4004	0,6561 Hr	17797
DbSin WIFI	0,4813 Hr	4509	0,6734 Hr	18396
DbCon WIFI	0,7805 Hr	3031	1,2262 Hr	11011
DbSin DATA	5,0623 Hr	433	7,1810 Hr	1742
DbCon DATA	5,8018 Hr	393	8,6851 Hr	1498
DbSin NO	0,0808 Hr	2953	0,0000012776 Hr	1065400000
DbCon NO	0,3794 Hr	580	0,0000020517 Hr	612120000

DMTBF: Tiempo promedio entre fallas demostrado (Demonstrated Mean Time Between Failures)

IMTBF: Tiempo promedio entre fallas instantáneo según una variable de tiempo (Instantaneous Mean Time Between Failures)

CNOF: Número acumulado de fallas (Cumulative Number of Failures)

Conclusiones

La cantidad de veces que se necesitará una conexión a internet fiable será menor, ya que la replicación hace que la probabilidad de buscar y encontrar un documento en la base de datos sea mayor.

Si bien puede llegar a presentar desventajas en algunos otros aspectos, podemos decir que la incidencia de la replicación en la búsqueda es clara y presenta ventajas al momento de hacer consultas sobre diversos documentos.

Trabajos futuros

- Mayor difusión para lograr tests con más usuarios, más tiempo y más información
- Centrar el análisis en el uso de la aplicación con un solo tipo de conexión a la red
- Analizar el uso de la aplicación offline
- Extenderse más allá de las apps de consultas
- Estudiar los posibles conflictos entre los documentos replicados