

Bases de Datos Distribuidas 2016 Facultad de Informática - Universidad Nacional de La Plata

Prof. Fernando G. Tinetti
Lic. Franco Agustin Terruzzi

Objetivo: conocer herramientas de desarrollo reales que se utilizan para gestionar sistemas distribuidos actuales y formas de medición para evaluar sus resultados. Desarrollar una bbdd utilizando varias estrategias, medir los resultados obtenidos y elaborar conclusiones.

Parte 1: Conceptos base

Desarrollar una REST API¹ en node.js² utilizando mongodb³ como motor de base de datos, tomando como guía el repositorio compartido⁴.

Definir el modelo de datos a utilizar evaluando distintas implementaciones orientadas a documentos (mongodb) a la siguiente base de datos relacional:

EMP			ASG			
ENO	ENAME	TITLE	ENO	PNO	RESP	DUR
E1	J. Doe	Elect. Eng.	E1	P1	Manager	12
E2	M. Smith	Syst. Anal.	E2	P1	Analyst	24
E3	A. Lee	Mech. Eng.	E2	P2	Analyst	6
E4	J. Miller	Programmer	E3	P3	Consultant	10
E5	B. Casey	Syst. Anal.	E3	P4	Engineer	48
E6	L. Chu	Elect. Eng.	E4	P2	Programmer	18
E7	R. Davis	Mech. Eng.	E5	P2	Manager	24
E8	J. Jones	Syst. Anal.	E6	P4	Manager	48
			E7	P3	Engineer	36
			E8	P3	Manager	40

PROJ			PAY	
PNO	PNAME	BUDGET	TITLE	SAL
P1	Instrumentation	150000	Elect. Eng.	40000
P2	Database Develop.	135000	Syst. Anal.	34000
P3	CAD/CAM	250000	Mech. Eng.	27000
P4	Maintenance	310000	Programmer	24000

Documentar en un informe las alternativas evaluadas y establecer conclusiones sobre la alternativa escogida.

Dar soporte (desarrollar los sustantivos necesarios) para todas las operaciones CRUD de los objetos resultantes del modelo de datos escogido, utilizando la API REST.

¹ <http://www.restapitutorial.com/lessons/whatisrest.html>

² <https://nodejs.org/en/>

³ <https://www.mongodb.com/es>

⁴ <https://bitbucket.org/t3rru/bbdd2016>

Utilizar postman para medir tiempos de respuesta sobre los distintos sustantivos gestionados, haciendo foco en distintas consultas sobre el modelo de datos, que involucren al menos entrecruzamiento de datos (joins relaciones), creación de documentos derivados (relaciones 1 a N y N a N) y eliminación de los mismos.

Establecer conclusiones en el informe sobre:

- Tolerancia a fallos (errores de sistema, conexión, fallos de base de datos)
- Tiempo de respuesta (ante operaciones con distinto grado de complejidad)
- Escalabilidad de la solución (mejora de tiempos de respuesta, reutilización de código, etc)

Herramientas a utilizar: mongodb, mongoose, nodejs, postman

Parte 2: Aplicación de conceptos de bbdd

Cambiar el modelo de datos de la API REST ya desarrollada para utilizar una base de datos mongodb *replicada* (utilizando replica Sets⁵) que posea al menos 3 esclavos y 1 master. (por motivos de infraestructura pueden utilizar 2 esclavos y 1 falso esclavo para desempatar las votaciones).

Desarrollar un helper (controlador interno en node.js) que encapsule los accesos al modelo de datos, con el objetivo de: tolerar fallos (consistencia, ACID), tracear operaciones (generar un log de todas las operaciones en el sistema), hacer profiling (realizar mediciones acerca de lo que tardan en tiempo distintas operaciones sobre la base de datos y almacenarlas en un modelo de datos).

Elaborar un informe sobre la infraestructura y tecnología utilizada y los cambios necesarios a implementar sobre la aplicación.

Investigar sobre los distintos modos de lecto-escritura⁶ que posee mongo utilizando los Replica Sets y realizar pruebas con postman sobre los distintos modos midiendo: tiempos de respuesta, tolerancia a fallos, y veracidad de los resultados. (Hacer énfasis en la capacidad y el tiempo de las modificaciones, inserciones y demás operaciones CRUD en las réplicas).

Experimentar y establecer conclusiones sobre los beneficios y los defectos que presenta el nuevo sistema replicado (tolerancia a fallos, tiempo de respuesta, dificultad de implementación).

Finalizar el informe con una conclusión sobre la posibilidad de implementar la alternativa de replica Sets en un entorno de producción actual, en la implementación de una API REST que necesite estar online (funcional) una gran parte del tiempo.

Herramientas a utilizar: virtualbox, replica Sets(mongodb)

Parte 3: Conceptos avanzados de bbdd

A partir de la conclusión anterior, establezca una serie de mejoras para potenciar el rendimiento general del sistema.

Investigar y construir un informe con los pasos a realizar y los cambios a efectuar sobre su sistema, para aplicar lo investigado. (no es necesario desarrollar).

⁵ <https://docs.mongodb.com/manual/administration/replica-sets/>

⁶ <https://docs.mongodb.com/manual/core/replica-set-write-concern/>

Alternativas de investigación:

- Mejora del tiempo de respuesta (optimización de consultas distribuidas con replicación, modelo lazy de operaciones CRUD, etc)
- Tolerancia a fallos (transacciones de 2 y 3 fases con replicación)
- Fragmentación (utilización de clusters en mongodb y sharding⁷)
- Medición y mejora sobre cantidad de réplicas (n réplicas, baja de esclavos, baja del master, etc)

⁷ <https://docs.mongodb.com/v3.0/sharding/>