ENSAYO

DESAFÍOS, PROBLEMAS y DIFERENCIAS

PostgreSQL y MySQL

Una de las cosas más importantes a resolver cuando se va a implementar, escalar, desarrollar una aplicación o base de datos es decidir cuál es la que más conviene por su rendimiento o por su arquitectura que hace evidente una ventaja sobre los demás sistemas de Base de Datos, para eso vamos a analizar un poco sobre las características de estos 2 sistemas y el impacto que tuvo en un proyecto tan importante como UBER en por el mal dimensionamiento inicial.

Iniciemos revisando las características de PostgresSQL, entre las cuales se sabe que permite persistencia de datos, escalabilidad, adaptabilidad al hardware como CPY y memoria, es más eficiente con el uso de rollback’s, subconsultas y transacciones, comprueba la integridad referencial y permite almacenar procedimientos aproximándose mucho a gestores de Alto nivel como ORACLE; posiblemente características que tuvo en consideración en el principio UBER para usarlo como su sistema gestor de BD. Lo que no dimensionó fue el gran crecimiento casi exponencial, que impactaría factores como Rendimiento por consumo elevado de Recursos y limitaciones de arquitectura en limites por Fila de 8K, en corto plazo lo que forzaría la creatividad de su equipo de Ingeniería a migrar rápida y adecuadamente su plataforma a una con características más ajustadas a su cambio, será MySQL el mejor candidato?, veamos.

Por otro lado, tenemos a MySQL cuya arquitectura muestra que es muy rápido en la ejecución de operaciones con un consumo mínimo de recursos y con uno de los menores probabilidades de corrupción de datos, esto lo hace un gran candidato para ser escalable en la infraestructura necesaria. Sin embargo, que hacer frente a fallos sabiendo que no soporte Rollbacks?, además tampoco soporta subconsultas, transacciones ni maneja la integridad referencial haciendo de este gestor una solución pobre para muchas aplicaciones cuyas bases de datos son gigantes.

Teniendo claras las la limitaciones de arquitectura, analicemos cuales fueron los problemas adicionales que sufrieron en 2016 por el equipo de Ingeniería de UBER quienes inicialmente consideraron que PostgreSQL era su mejor opción y finalmente tuvieron que Migrar a MySQL.

Tuvieron problema para la escritura en disco SSD ya que la modificación de un dato en un campo significaba muchas más de 4 actualizaciones físicas, replicación y corrupción de datos haciéndolo muy irritante sobre todo por que impedía identificar en cuantas columnas se corrempen los datos cuando se promueven base de datos maestras para aumentar la capacidad, upgrades de versiones ya que no permite hacerlo sin antes cerrar la base de datos maestra, ejecutar el proceso de actualización, iniciar la maestra, copiarla, limpiar las copias anteriores, restaurar la nueva copia del maestro a la replica vuelva a colocar las replicas en el orden adecuado y rece lo que se sepa! Y por otro lado el soporte de Control de concurrencia multiversion (MVCC) ya que durante este proceso postgresql detienen el hilo del WAL hasta que finalice la transacción porque si una réplica de transmisión tiene una transacción abierta, las actualizaciones de la base de datos se bloquean si afectan a las filas que la transacción mantiene abiertas.

Ahora como evitaron todos estos problemas?, migrando a MySQL por supuesto ya que este permite que los índices secundarios apunten a los Primarios y estos últimos sean los que contienen las ubicaciones físicas en disco, admite múltiples modos de replicación diferente haciendo que sea trivial replicar entre diferentes versiones, manejando conexiones concurrentes con hilo por conexión.

En conclusión, se pudo prevenir todos los problemas presentados si se hubiera dimensionado desde el inicio incluso considerando una holgura importante ya que se trata de un servicio de consumo masivo en todo el mundo, ahora la pregunta es, será suficiente la plataforma actual para la creciente que pueda tener en los próximos años UBER conociendo que ha comenzado a incluir nuevos servicios?. Tener varios gestores de BD es una muy buena opción para la escabilidad.