

ARQUITECTURA DEL DATA WAREHOUSE

Título del anteproyecto:

**Replicación incremental por medio de archivos log
en el Data Warehouse**

Versión del documento:

1.0

Fecha:

09.03.2020

Grupo:

G2006

Participantes:

Carlos Daiha Basna

Ciprian Ilut

Alejandro Nieto Jeux

Javier del Río García

Contenido

1 Resumen3

2 Introducción.....3

3 Objetivos3

 3.1 Objetivo general3

 3.2 Objetivos específicos.....3

4 Aspectos metodológicos4

5 Estimación temporal.....4

6 Tareas y asignación de roles4

7 Bibliografía.....4

1 Resumen

Ante las dificultades que supone realizar réplicas completas de tablas en un sistema relacional dentro del Data Warehouse debido a los volúmenes tan grandes de datos que se manejan, se propone realizar un estudio de las posibles soluciones de replicación existentes, centrándonos más en concreto en la técnica de replicación incremental basada en archivos log para un sistema relacional. Además proponemos realizar un caso práctico sobre este tema en el cual utilizaremos la técnica de replicación anteriormente mencionada para levantar una base de datos de prueba a través de los logs generados por otra.

2 Introducción

En el ODS a veces es necesario llevar a cabo un proceso de replicación de datos, ya sea para realizar un backup, distribución de datos o para realizar tarea de clustering. Si ponemos como ejemplo un gestor relacional, la forma más sencilla de realizar esta tarea es mediante la replicación completa de las tablas.

Sin embargo, si nos movemos al campo de los Data Warehouses, esta técnica supone un problema, pues el volumen de datos es notablemente superior y provocaría problemas de escalabilidad en términos de almacenamiento. Por este motivo se debe buscar una alternativa que reduzca el espacio ocupado por las réplicas.

Una de estas alternativas puede ser la replicación incremental basada en los archivos log. Estos archivos contienen las transacciones realizadas en la base de datos. Gracias a ellos, se puede realizar una réplica utilizando únicamente las transacciones y ahorrando de esta manera una gran cantidad de espacio.

3 Objetivos

3.1 *Objetivo general*

Estudiar las posibles soluciones para realizar réplicas dentro del Data Warehouse en contraposición a una técnica de replicación completa en un sistema gestor de bases de datos relacional MySQL.

3.2 *Objetivos específicos*

- Estudiar soluciones alternativas existentes para realizar réplicas en el Data Warehouse.
- Analizar los ficheros log generados por el SGBDR MySQL.
- Estudiar la técnica de replicación incremental basada en archivos log aplicada a MySQL.
- Caso práctico de creación de levantamiento de una base de datos a partir de los logs generados por otra.

4 Aspectos metodológicos

Distinguiremos entre tareas prácticas y teóricas.

- Dentro de las tareas teóricas se encuentran el estudio y el análisis de soluciones al problema de replicación dentro del Data Warehouse para un SGBDR como MySQL y más en concreto la técnica de replicación incremental basada en archivos log.
- Dentro de las tareas prácticas se encuentran el análisis de los ficheros log generados por el SGBDR MySQL y el caso práctico de levantamiento de una base de datos a partir de los logs generados por otra.

5 Estimación temporal

Dado que desconocemos el tema por completo, la estimación temporal será bastante imprecisa. Al contar con una tarea práctica y un estudio teórico creemos que la realización de este proyecto conllevará entre 20 y 30 días.

6 Tareas y asignación de roles

	Soluciones alternativas	Análisis de los ficheros log generados en MySQL	Replicación incremental basada en archivos log	Caso práctico
Carlos	x		x	x
Alejandro	x	x		x
Ciprian	x		x	x
Javier		x	x	x

Tabla 1: Reparto de las tareas

7 Bibliografía

[1] Stitch's Log-based Incremental Replication,

<https://www.stitchdata.com/docs/replication/replication-methods/log-based-incremental>

[2] Replicación de datos en MySQL,

<https://www.adictosaltrabajo.com/2009/12/08/mysql-replicacion/>