Testing MySQL Cluster with Star Schema Benchmark

Introduction

• Decision Support Systems (DSS) are increasingly used in industry, public organizations and the scientific community.
• Largest and most complex transactional systems;
• The amount of data businesses retain for analytic purposes is growing at a rate of 50% per year, according to Forrester Research;
• These events => need for database engines able to deal with large amounts of data in an acceptable time frame.

NEEDS:
• Continuous analytics;
• more complex exploration of business data;
• new business intelligence functionalities

=> SCALABILITY is an essential characteristic for successful DSS.

AIMS:
• Estimate the gains with the increase of resources allocated to the MySQLCluster.
• understanding the limitations and characteristics of MySQLCluster scalability for DSS.

A system whose performance increases by adding hardware, proportionally to the increased capacity is called “scalable system”.

MySQLCluster is an engine for scalable databases
We evaluate its scalability and performance
Evaluating over DSS => appropriate benchmark, the Star Schema Benchmark (SSB).

Index
The SSB was created to test a star schema of a data warehouse. This four testing groups simulates a DSS.

SSB queries are called query flights.

Each query flight answers a specific business question: 3 or 4 statements with increased selectivity.

The SSB is based on TPC-H benchmark. The queries are also based on TPC-H queries.

MySQL Cluster with Star Schema Benchmark (SSB)

SSB data model

Dimension tables:
- CUSTOMER
- SUPPLIER
- PART
- DATE

Fact table:
- LINEORDER

MySQL Cluster

Management node - Manage the other nodes of the MySQL Cluster, provide configuration data, starting and stopping nodes, perform backups.

Data Node - Stores the data.

SQL node - Accessing the data in the cluster

MySQL Cluster with 4 data nodes, 2 node groups, 2 replicas and 4 partitions

SSB (dbgen) data generator => 3 datasets, one of 1GB, one of 3GB and 6GB.

The fact tables had 6M, 18M and 36M records respectively.

Three MySQL cluster scenarios:
- 1 data node;
- 2 data nodes;
- 4 data nodes; they all had an additional node that housed the management node and sql node.

Experimental setup (i)

**Experimental setup (ii)**

- All data nodes had homogeneous features.
- A data node was composed of:
  - 2 virtual processors,
  - 8 GB RAM,
  - one 125 GB disk (disk 0),
  - one 15 GB disk (disk1),
  - 1 NIC

  the disk 0 was used for the operating system, MySQL binaries and data replicas; the disk 1 was used for swap file.

**LAB summary description of (iii)**

- We evaluated the MySQLCluster’s scalability, running the 4 SSB queries groups, for each cluster configuration and dataset.
- All the queries were performed 3 times, each execution time recorded, we calculated the average value.

**Results (i)**

Execution time of first queries of each group for the 6GB dataset, with different number of data nodes

**Results (ii)**

Execution time of first queries of each group for the 4 nodes scenario, with different dataset size

**Results (iii)**

Total average execution time of each query group for all dataset.

Conclusions and Future Work

- MySQL Cluster is scalable, although at a level lower than expected, far from a direct proportional relationship.
- Level of scalability hit is weak.
- Performance worsens significantly when the database engine starts using the swap file as memory.
- We will compare this engine with more traditional engines such as MySQL InnoDB and Microsoft SQL Server 2012.