A Model&DBMS Independent Benchmark for Data Warehouses

Ibtisam Ferrahi*, Sandro Bimonte**, Kamel Boukhalfa***

* University Mhamed Bougara, Boumerdes, Algeria
ferrahi.ibtisam@umbb.dz,
** TSCF, Irstea. 9 Av. Blaise Pascal, Aubiere, France
Sandro.bimonte@irstea.f
***LSI Laboratory-USTHB, BP 32 ElAlia, Bab Ezzouare, Algiers
kboukhalfa@usthb.dz

Abstract. NoSQL systems support new data models, which propose alternative models to the well-known relational models, and query languages. Due to the lack of a well-accepted logical model for Data Warehouses (DWs), some preliminaries works proposes extensions/redefinition of relational star and snow-flake schemata for NoSQL families. However, many other modeling possibilities remain unexplored, and it is difficult to compare these proposals because of the lack of a well-recognized benchmarking framework for DWs. In this paper, we propose a generic extension of the relational Star Schema Benchmark, called GenSSB, to handle any kind of DBMS in terms of logical models. We validate our framework by instantiating GenSSB for some different logical models and DBMSs.

1 Introduction and motivation

Data Warehouses (DWs) and OLAP systems with their implementation in relational and multidimensional architectures have been widely studied in the last 30 years (Kimball and Ross, 2002). Nowadays, DWs and OLAP systems have reached a great maturity having different kind of applications in several domains such as marketing, health, agriculture, etc. Conceptual, logical and physical issues have been extensively investigated by academic and industrial communities. Several conceptual models based on ER, UML and other formalisms have been proposed, but no standard has been defined yet. Some logical models have been also proposed. Star and snowflake schemas are actually recognized as the de-facto standard logical models for DWs. The star schema denormalizes dimensional attributes to avoid expensive join operators. The snowflake schema is similar to the star schema except that dimensions are normalized into multiple related tables.

Based on these logical models, some optimization techniques (such as index, materialized views, fragmentation), and administration/tuning tools have been proposed. Therefore, specific benchmarks for relational DW (such as SSB, TPC-DS, etc.) (O'Neil et al., 2009) have been designed over these logical models to evaluate and compare performance of these optimization techniques.