# AutoMDB: A Framework for Automated Multidimensional Database Design via Schema Transformation

Alfredo Cuzzocrea
ICAR-CNR, Univ. of Calabria
Italy
cuzzocrea@si.deis.unical.it

Rim Moussa
LaTICE Lab., Univ. of Tunis
Tunisia
rim.moussa@esti.rnu.tn

Hejer Akaichi
ESTI, Univ. of Carthage
Tunisia
hejer.akaichi@esti.rnu.tn

## Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: [Miscellaneous]; D.2.8 [**Software Engineering**]: Metrics—*Complexity measures*

## General Terms

Experimentation, Performance, Design

## Keywords

Multidimensional Databases, DataWarehousing, Schema Evolution, Logical OLAP Design.

## ABSTRACT

In this paper, we demonstrate *AutoMDB* -a framework for automated multidimensional database design. In order to prove the effectiveness and the reliability of our proposal, we tested it over the well-known TPC-H benchmark.

## 1. INTRODUCTION

One of the most powerful and prominent technologies for knowledge discovery in Decision Support Systems (DSS) environments are *Business Intelligence (BI) Suites* and particularly *On-line Analytical Processing (OLAP) technologies*. OLAP relies heavily upon a data model known as the *multidimensional databases* (MDB) [4]. Compared to relational databases, MDB enhance data presentation and increase performance by storing aggregated data. According to market watchers, BI platforms will remain one of the fastest growing software markets. Nevertheless, MDB design is often neglected and OLAP cubes are defined in a haphazard way, without worrying about performance and maintenance costs.

In the literature there are a number of papers that addressed MDB design. However, proposed methods were not generalized. Regarding benchmarks, TPC-H benchmark [3] is the most prominent DSS benchmark. However, TPC-H does not allow evaluation of OLAP technologies and does not implement the OLAP Mandate [1]. Therefore, in [2] we propose a framework for automated design of MDB. Our proposal was succesfully tested over TPC-H benchmark. In what follows, Section 2 presents the demonstration outline and Section 3 concludes the paper.


*The 19th International Conference on Management of Data (COMAD)*, 19th-21st Dec, 2013 at Ahmedabad, India.


## 2. DEMONSTRATION OUTLINE

The demonstration shows (i) *TPC-H*d* benchmark -a multidimensional TPC-H benchmark and (ii) *AutoMDB* -a recommender for MDB schema optimization.

### 2.1 Demonstration of TPC-H*$d$ Benchmark

Along the demonstration, we will present

- the process of generation of TPC-H*$d$ data set from the original TPC-H benchmark,
- the process of derivation of initial MDB schema [2] from TPC-H workload. Fig.1 illustrates the OLAP cube C8 proposed for SQL statement Q8,
- two implementations of TPC-H*$d$. The first is a JAVA client, which sends a stream of MDX statements to a RDBMS using the appropriate JDBC driver and OLAP4j. The second is a web-based implementation of TPC-H*$d$, which uses JPivot as an OLAP client, Apache Tomcat as a JSP container, Mondrian as a ROLAP server and MySQL as a relational DBMS,
- pivot tables and OLAP operations (such as slice, dice, ...) on OLAP cubes of TPC-H*$d$ benchmark. Fig.2 illustrates pivot tables obtained for C8 and Q8.

### 2.2 Demonstration of AutoMDB

Most BI solutions describe the MDB schema using XML. *First*, we implemented an XML parser (using SAX library) for loading the description of the cubes, namely measures' data, facts' data (i.e., physical tables or logical views), dimensions' data (i.e., hierarchies, levels and properties). *Second*, we implemented algorithms for cube merge and virtual cubes definition. Along the demo, we will,

- Load `tpch_multidim.xml` -the initial MDB schema of TPC-H*$d$ benchmark,
- Perform comparisons of OLAP cubes' characteristics. For each pair of OLAP cubes, we show whether they have same fact table or not and compute the number of shared | different | coalescable dimensions,
- Run merge of OLAP cubes using different similarity functions. For instance, Fig.3 shows a simple similarity function, which groups OLAP cubes having the same fact table.

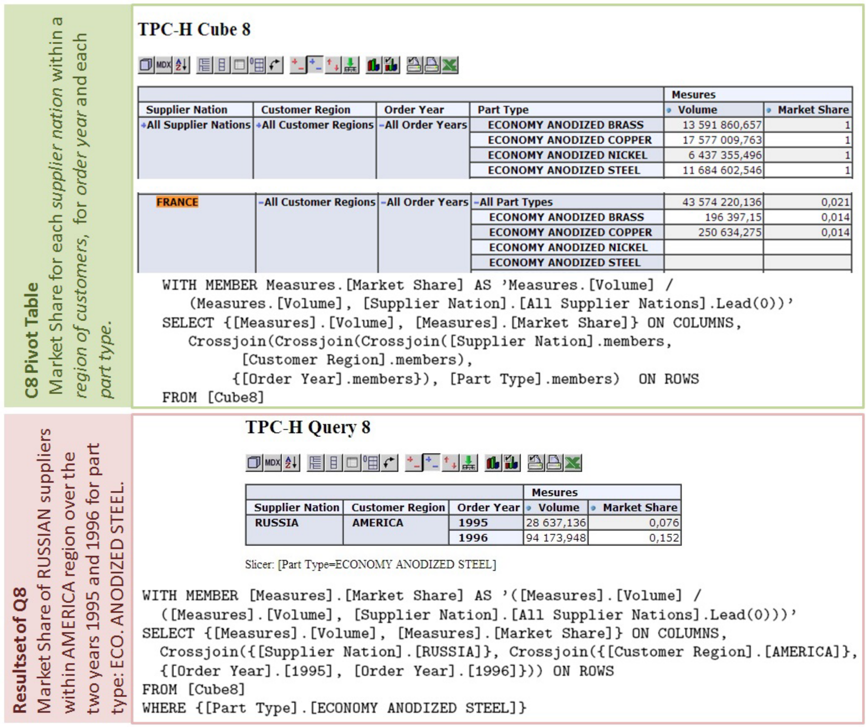


**Figure 2: Screenshots of C8 and Q8 Pivot Tables and corresponding MDX Statements.**

```
SELECT o_year, sum(case when nation = '[NATION]' then volume
                else 0 end) / sum(volume) as mkt_share
FROM (SELECT extract(year from o_orderdate) as o_year,
             l_extendedprice * (1-l_discount) as volume,
             n2.n_name as nation
      FROM part, supplier, lineitem, orders, customer, nation n1, nation n2, region
      WHERE p_partkey = l_partkey        AND s_suppkey = l_suppkey
        AND l_orderkey = o_orderkey      AND o_custkey = c_custkey
        AND c_nationkey = n1.n_nationkey AND n1.n_regionkey = r_regionkey
        AND r_name = '[REGION]'          AND s_nationkey = n2.n_nationkey
        AND o_orderdate between date '1995-01-01' and date '1996-12-31'
        AND p_type = '[TYPE]') as all_nations
GROUP BY o_year;
```

```
OLAP Cube 8
 |
 |-- Fact Table
 |    |-- LineItem Facts
 |-- Measures
 |    |-- M8.1: Σ(l_extendedprice×(1-l_discount))
 |-- Dimensions
 |    |-- D8.1: Part
 |    |    |-- L0: type
 |    |-- D8.2: Order Date
 |    |    |-- L0: year
 |    |-- D8.3: Customer Geography
 |    |    |-- L0: region
 |    |-- D8.4: Supplier Geography
 |    |    |-- L0: nation
```

**Figure 1: Turning Business Query Q8 of TPC-H benchmark into an OLAP Cube.**

# 3. CONCLUSIONS AND FUTURE WORK

In this demo, we demonstrate a framework for automating MDB schema design, which embeds several points of research innovation. Future work is mainly oriented towards the test of our framework using TPC-DS benchmark, and the emerging *Big Data*.

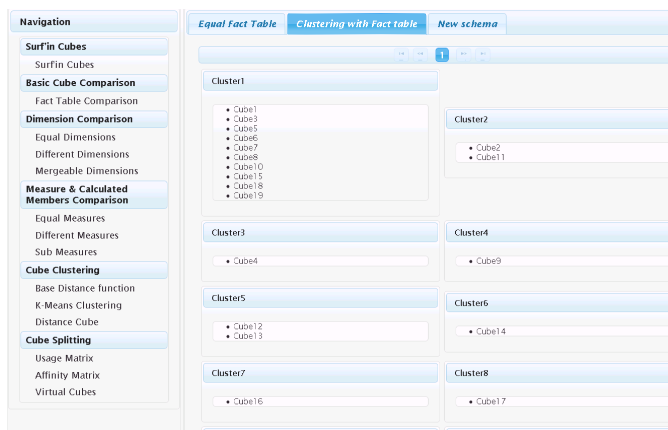


**Figure 3: Screenshot of AutoMDB -clustering OLAP cubes sharing same fact table.**

# 4. REFERENCES


[1] E. F. Codd, S. B. Codd, and C. T. Salley. Providing OLAP to user-analysts: An IT mandate. 1993.

[2] A. Cuzzocrea and R. Moussa. Multidimensional Database Design via Schema Transformation: Turning TPC-H into the TPC-H*d Multidimensional Benchmark. In *Proc. of 19th COMAD*, 2013.

[3] TPC-H benchmark. `http://www.tpc.org/tpch`.

[4] P. Vassiliadis. Modeling multidimensional databases, cubes and cube operations. In *Proc. of the 10th SSDBM*, pages 53–62, 1998.