Volume 5 ssue 2 (2017) 195-196 ISSN 2347 - 3258
International Journal of Advance Research and Innovation

Data Modeling Techniques for Data Warehouse

Sona Kumra
Department of Computer Application, Bharatiyavidhyapeeth,Pachim Vihar,New-Delhi, India

Article Info
Article history:
Received 25 January 2017
Received in revised form
20 February 2017
Accepted 28 February 2017
Available online 15 March 2017

Keywords
Data Modeling, Data Warehouse, Multidimensional Modeling

Abstract
The conceptual Entity-Relationship (ER) diagrams are extensively used for database design in relational database environment, which emphasized on day-to-day operations. Multidimensional (MD) data modelling, on the other hand, is crucial in data warehouse design, which targeted for managerial decision support. It supports decision making by allowing users to drill-down for a more detailed information, roll-up to view summarized information, slice and dice a dimension for a selection of a specific item of interest and pivot to re-orientate the view of MD data. When designing a MD model regardless whether it is a star or snowflake schema, it involves the identification of a fact, dimensions and measure attributes. This paper will explore on how the Multidimensional model can be used as the solution of data warehouse design instead of ER Model.

1. Data Modelling Techniques

Two data modelling techniques that are part of data warehousing are ER (Entity relationship modeling) and Multidimensional modeling.
Difference between ER and Multidimensional modeling
Dimensional modelling is very flexible for the user perspective. Dimensional data model is mapped for creating schemas whereas ER Model is not mapped for creating shemas and does not use in conversion of normalization of data into denormalized form.
ER Model is utilized for OLTP databases that uses any of the 1st or 2nd or 3rd normal forms, where as dimensional data model is used for data warehousing and uses 3rd normal form.
ER model contains normalized data where as Dimensional model contains denormalized data.

1.1 ER Modeling

ER diagrams are represented using conceptualdata: Entity, Relationships,Attributes.

1.1.1 Entity

In a logical sense, entities are the equivalent of grammatical nouns, such as employees, departments, products, or networks.

1.1.2 Relationships

Relationships are the equivalent of verbs or associations, such as the act of purchasing, the act of repairing, being a member of a group, or being a supervisor of a department. A relationship can be defined according to the number of entities associated with it, known as the degree.

1.1.3 Attributes

An entity can be defined by means of its properties, called attributes.

1.2 Multidimensional Modeling

The conceptual representation generated by multidimensional modeling consists of a set of fact schemata.
Fact schemata basically model facts, measures, dimensions and hierarchies.

1.2.1 Facts

Multidimensional Modeling is concept relevant to decision making process.Multidimensional Modeling typically models a set of events taking place within a company.

1.2.2 Dimension

Multidimensional Modeling is a fact with a finite domain and describes an analysis coordinate of the fact

1.2.3 Measure

Multidimensional Modeling is a numerical property of a fact and describes a quantitative fact aspect that is relevant to analysis.
Considering Relational context, there are two basic models that are used in dimensional modeling: (i) star model and (ii) snowflake model.

Corresponding Author,
E-mail address: : sonakumra1992@gmail.com,
Phone No.: +91 8285159898

2. Star Model vs Snowflake model.

The star model is the basic structure for a dimensional model. It has one large central table (fact table) and a set of smaller tables (dimensions) arranged in a radial pattern around the central table. (Figure 1). The snowflake model is the result of decomposing one or more of the dimensions. The many-to-one relationships among sets of attributes of a dimension can separate new dimension tables, forming a hierarchy. (Figure 2). The decomposed snowflake structure visualizes the hierarchical structure of dimensions very well.

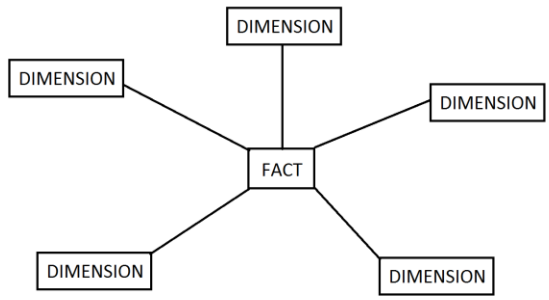

Fig. 1 Star Model

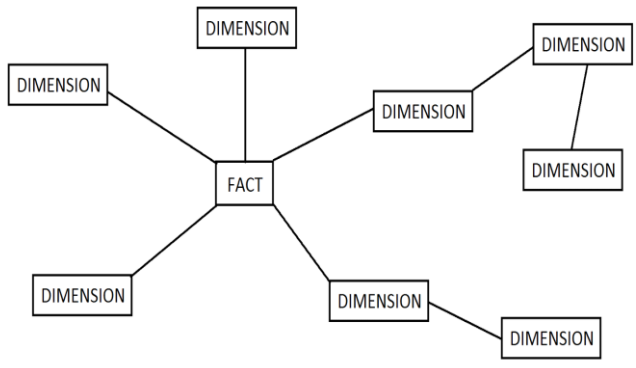

Fig. 2 Snowflake Model

3. Why multidimensional modeling more suitable than ER model for data warehouse?

1. End user prospective : end user cannot navigate complete ER model as there is no graphical user interface that makes it usable by end users.
2. ER modelling does not allow the high performance retrievel of data because it requires highly normalized relational tables.

4. Conclusions

This paper examines that ER modeling for data warehouse is not viable concept keeping in mind the recent developments in data warehousing. Various supporting arguments:

1. We cannot represent all the ER model into star or snowflake schema.
2. Poor query performance in large databases while using ER models.
3. Many ER diagrams don't illustrates many-many relations.
4. OLAP operations can be performed easily in data modeling schema .

References

[1] JA Blakeley, N Coburn, PA Larson. Updating derived relations: Detecting irrelevant and autonomously computable updates. ACM Transactions on Database Systems, September 1989 .

[2] D Lomet, J Widom. Special Issue on Materialized Views and Data Warehousing, IEEE Data Eng in eering Bulletin 18(2), 1995.

[3] Kimball, R. The Data Warehouse Toolkit. John Wiley, 1996.

[4] Barclay TR, Barnes J, Gray P Sundaresan. Loading Databases using Dataflow Parallelism. SIGMOD Record,.23(4), 1994.

[5] Blakeley JA, N Coburn, P Larson. Updating Derived Relations: Detecting Irrelevant and Autonomously Computable Updates. ACM TODS, 4(3), 1989.