Data Model: A Blueprint for Data Warehouse

Jasmeet Singh Birgi [*], Mahesh Khaire [**], Sahil Hira [**]

[*] Teradata Data Analyst
[**] BI Application Developer

Abstract- [1][9]Data modeling is the basic step of any database design, which is a powerful expression of any company business requirements. It is nothing but an act of exploring data oriented structures and the conceptual models in form of diagrams and flow charts on how data items relate to each other. Data modeling tools and techniques capture and translate complex system design into clearly understood representation of data flows and processes. The main focus of data modeler is on data and exploring data issues, therefore they tend to get much more accurate data than object modelers. They often use multiple models to view the same data and ensuring that all processes, entities, relationships and data flows have been processed.

Index Terms- Fundamental terms used in Data Modelling, Three intrinsic parts of the data modelling (top-down or bottom-up approach) like Conceptual, Logical and Physical Data Model.

I. FUNDAMENTAL TERMS USED IN DATA MODELLING

There are various fundamentals of data modeling that one should know about, and should perform various tasks:

A. Identify Entity Types: - Entity is nothing but the concept of class, which represents the collection of similar objects. It is slightly different from class, as class involves both data and behavior, whereas entity involves only data.

B. Identify Attributes: - There are one or more attributes of an entity, sometimes the specifying the attribute can have a significant impact on the development and maintenance efforts, but these can also overbuild the system and increase the cost for the same.

C. Apply Data Naming Conventions: -The basic idea for this is that all should agree to follow a common set of modeling standards on a project, this can help in better understanding of the code.

D. Identify Relationships:-Relationship between entities is similar to the relationship between objects. For example in an online ordering system, the Customer, Address and Order are the entities which are related to each other.

E. Apply Data Model Pattern: - There are many data modeling structures that can be applied in data models such as Data Flow Diagram, Information Model, Object Model, Object Role Model and Unified Modeling language Model.

F. Assign Keys: - The two main methods of assigning keys to the table is to either assigning a natural key or introducing a surrogate key, which can include all the columns from the table. It is observed that introducing a surrogate key is a much better option than natural key.

G. Normalize to Reduce Data Redundancy: - The main motive of data normalization is to reduce data redundancy, as it is extremely difficult to store same information at several places. We use 1NF, 2NF and 3NF.

H. De-normalize to Improve Performance: - As normalization focus on reducing redundancy, it can have a serious impact on the performance to access data. Therefore, sometimes one need to neglect redundancy to improve performance of the system.

[6]Data Model can be considered as a blueprint of a Data warehouse, providing conceptual representation of Database Objects and how they relate to one another in a Data warehouse. A good Data Model should represent the nature of data, how it will be organized and the business rules that govern them.

II. THREE INTRINSIC PARTS OF THE DATA MODELLING

There are three parts of a data model which can be created by either Top Down approach or Bottom Up approach:

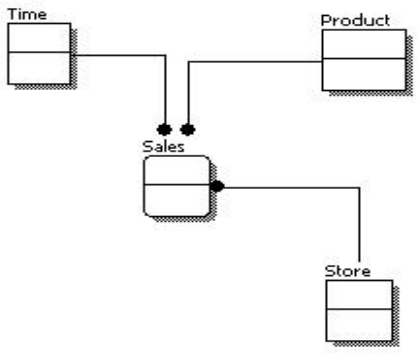
A. Conceptual Data Model:-

[8][11]Conceptual model is the first step in a top down approach and is a clear and accurate visual of business of the organization. These are created as a precursor to Logical data models or as an alternative to Logical data models. It determines the highest level of relationships between different entities. These can be called domain models and used to explore domain concepts of the project. It can be used to explore high level static business structures and concepts.

There are various features and characteristics of conceptual data models such as:

1. Including important entities and relationships among them
2. No attribute is specified, only entities
3. Entities will have definition
4. No primary key is specified

Conceptual data model is nothing but the summary level data model which describes the entire enterprise but in a much abstract way. Therefore, it provides very little information, for example in the below diagram, Product, Time, Sales and Stores are the entities and relationship between those entities is shown. For instance, diagram shows the relationship between Time and Sales, Product and Sales and Sales and Stores.



B. Logical data Model:-

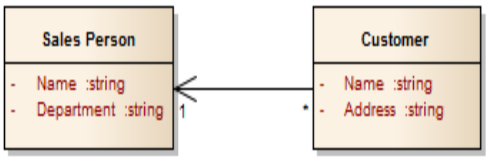
[2][4][7]Logical Data Model used as basis for Physical Data Model and depicts the detailed structure of the elements / Objects in a system and relationship between these elements. Once Conceptual data model is finalized, we can start with Logical Data model. A good logical model is created to represent the business requirements and should include all required entities, attributes, key groups and relationships that represent and define the business rules.

Typical Deliverables of Logical Data Mode include: -

1. Entity relationship diagrams
2. Objective is to give development team an idea about the different categories of data business will have and how it will be related.
3. Business process diagrams: - Briefs the development team on how data move within the organization.
4. User feedback documentation

A simple example of a Logical data model is provided below:

Note that the data elements Salesperson and Customer contain UML Attributes the Attribute types, however, remain platform-independent.



Platform-specific Attribute types and other meta-data that relate to a specific DBMS implementation are defined by the Physical data model [2].

Benefits for creating a LDM: -

1. The ability to access information across the enterprise increases with shared terminology.
2. Consistent reporting and analytics boost transparency internally as well as externally. This can help organizations make smarter strategic and tactical decisions, in addition to quickly demonstrating audit and regulatory requirements more easily.
3. Better data consistency and accuracy improve operational efficiency while reducing risk.
4. Eliminating the technical and human redundancy inherent in soloed systems as well as in data marts reduces expenses.

Benefits in financial risk management:

1. Improves the categorization of risk information and risk-measurement and -management capabilities.
2. Facilitates benchmarking the current state of an organization's risk information against best practices.
3. Allows organized growth of the data warehouse for more than risk management uses, over time, without having to re-architect.
4. Provides discipline and structure to the complexities inherent in risk management data.
5. Facilitates communication with consistent terminology among business units to ensure company leaders get the risk information they need in time to act.

"Creating a detailed logical model is an indispensable step for financial institutions wanting to maintain solid footing in shaky financial times and beyond." Sam Harris.

In short LDM is the very first step an organization takes to create an enterprise data warehouse (EDW) that can quickly and easily access and analyze risk data from disparate sources across the organization.

C. Physical Data Model:-

[3][10][12]Physical Data Model depicts how elements /database will actually be built from the business requirements that were established during the Logical data model. Physical modeling is the next stage, once business requirements have been gathered and logical model has been formed we have to convert this logical/ business model, into a relational database model. When physical modeling occurs, objects are being defined at the schema level. A schema is a group of related objects in a database. In short a physical database model will show all table structures, including column name, column data type, column constraints, primary key and relationships between tables. Additionally contrary to LDM Physical modeling is database software specific, meaning that the objects defined during physical modeling can vary depending on the relational database software being used.

For example, most RDBMS have variations with the way data types are represented and the way data is stored, although basic data types are conceptually the same among different implementations.

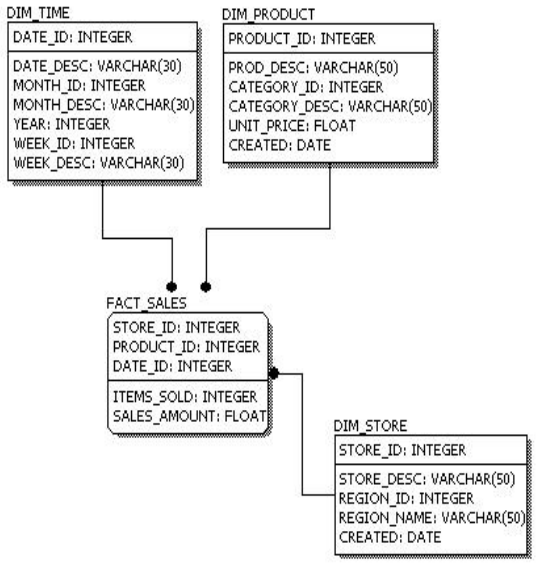
Physical data models are used to determine the internal schema of the database, determining data tables, data columns of those tables, and the relationship between those tables. It also includes column constraints, primary key, and foreign key can also be used for both agile and traditional projects. It is quite different from Logical Data Model as it represents the application and database specific implementation of logical data

modeling. Database performance, Indexing strategy, physical storage and de-normalization are important parameters of physical data model.

Typical Features of Physical Data Model include:-

1. Specifications of all tables and columns.
2. Foreign keys used to identify relationships between tables.
3. De-normalization can be occurred on basis on user requirements.
4. Physical data model will be different for different RDBMS,

For example, data type are different for different database.



Typical Deliverables of Physical Data Modeling include the following: -

1. Server Model Diagram: - shows tables, columns, relationship within a database.
2. User Feedback Documentation
3. Database Design Documentation

Steps for Physical Data Model Design are as follows: -

1. Converting entities into tables.
2. Converting relationships into foreign Keys.
3. Converting Attributes into Columns.
4. Modifying PDM based on physical constraints or requirements.

Main difference between Conceptual Data Model and Logical Data Model: -

1. In Conceptual data model primary keys are not present, whereas in logical data model, primary keys are present.
2. We cannot specify any attributes in a conceptual data model. However, in a logical data model, all attributes are specified within an entity.
3. Relationships are simply stated in a conceptual data model, not specified contrary to logical data model where relationships between entities are specified using primary keys and foreign keys.

Main difference between Logical Data Model and Physical Data Model: -

1. Entities will be replaced as table names.
2. Attributes will become column names.
3. Data types for each column will be specified. Data types can be different depending upon actual database being used.

III. CONCLUSION

Harnessing the quantity of data available with organization can derive huge business benefits. But to derive these benefits we must model our data warehouse correctly. We should create logical models for the many consumers of our data and physical models to ensure proper database performance; and most importantly, we must strive to meet the needs of the business. As the complexity increases from conceptual to logical to physical data model, therefore we always start with the conceptual data model. So that we can easily understand the high level, then we move on to logical data model, where we can understand details of our data. Finally we move to physical and implement it.

ACKNOWLEDGMENT

We would like to take this opportunity to thank & express our special gratitude to **Prof.Dr.Amresh Nikam**(Sinhgad Institutes,Pune) for their kind support.

REFERENCES

[1] The Data Model Resource Book: Universal Patterns for Data Modeling
[2] Data Model Patterns: Conventions of ThoughtData Modeling, A Beginner's Guide
[3] Data Modeling, A Beginner's Guide
[4] http://searchdatamanagement.techtarget.com/definition/data-modeling
[5] http://assets.teradata.com/resourceCenter/downloads/WhitePapers/EB6546.pdf?processed=1
[6] http://www.teradatamagazine.com/Article.aspx?id=11686
[7] http://learndatamodeling.com/blog/logical-data-modeling
[8] http://www.1keydata.com/datawarehousing/conceptual-data-model.html
[9] http://www.agiledata.org/essays/dataModeling101.html
[10] http://www.1keydata.com/datawarehousing/physical-data-model.html
[11] http://learndatamodeling.com/blog/conceptual-data-modeling-tutorial/
[12] http://learndatamodeling.com/blog/physical-data-modeling-tutorial/

AUTHORS

First Author – Jasmeet Singh Birgi, BTECH ,IBM India Pvt Ltd js.birgi@hotmail.com.
Second Author – Mahesh Khaire, MCA, Atos India Pvt. Ltd. maheshmkhaire@gmail.com
Third Author – Sahil Hira, BTECH,Atos India Pvt. Ltd. sahil.hira162@gmail.com

CorrespondenceAuthor–Dr.Amresh Nikam,
amreshn2000@gmail.com, +919921696327. (Sinhgad
Institutes,Pune