OPEN ACCESS

International
Journal
Of Modern Engineering Research (IJMER)

A Study on Data Warehousing with Metadata for Different OLAP Features Using SCD Types

K. Srikanth^1 , N.V.E.S. Murthy^2
*(Andhra university, Visakhapatnam, India
**(Andhra university, Visakhapatnam ,India

***ABSTRACT*:-** *A Study on Data warehouse with metadata for different online analytical processing (OLAP) and online transaction processing (OLTP) are basically main necessity of the large database collection in business, corporate fields and many areas using Research ideas and developing new ideas using SCD types. Nowadays many services, products and new techniques are offering possible many ideas in the Informatiaca Power center Designer and Database Management system . Data warehouses involve Cleaning of data, Integration of data, and Transformation of data. Moreover, data warehouses provide online analytical processing. Slowly Changing Dimensions are dimensions that have data that changes slowly, rather than changing on a time-based, regular schedule.*

***Keywords*:-** *Meta Data, Online Transaction Processing System, Online Analytical Processing, Data Warehouse, SCD.*

I. INTRODUCTION

On-line analytical processing (OLAP) is an aspect of decision support systems (DSS). OLAP server Relational OLAP (ROLAP) extended Data Base that maps operations on complex data to standard relational operators Multidimensional OLAP (MOLAP), specific purpose server that directly implements multidimensional data and operations. Different clients like Query and reporting tools, Analysis tools, Data mining tools For example, with OLAP solution, you can request information about company sales for business purpose and industrial development. OLAP systems are designed specifically for data analysis. All that is necessary for analysis is reading data. With this emphasis on reading only.

OLTP (On-line Transaction Processing) is represented by a large number of simple on-line transactions (INSERT, UPDATE, and DELETE)[1]. The main emphasis for OLTP systems is put on very immediate query processing, maintaining data integrity in multi-access environments and an powerful measured by number of transactions . In OLTP database there is detailed and current data, and schema used to store transactional databases is the entity model OLTP, OLAP, Metadata and Data warehouse are essential elements of supports system, which has developing become a center of the database industry. Many commercial products and services are now available, and all of principal database management system providers now offerings in these areas[2]. Informatica Power center receive different product functionality including ability to multiple register servers and metadata across the repository and partition data. You can join multiple sources use lookup You specify the target load order based on source qualifiers in a mapping. if u have the multiple source connected to the multiple destinations you can designate the order in which Informatics server loads data into the targets.

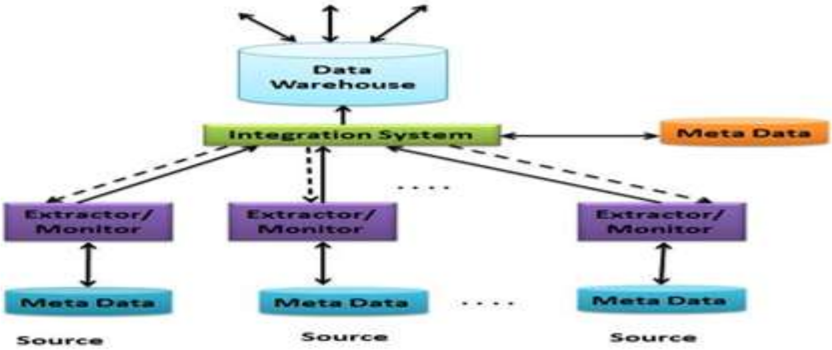

Fig 1: The basic requirements of Data warehouse, metadata, OLTP and OLAP.

1.1 Metadata

Meta data is “data about data” or Meta details “the data used to define other data”. It specifies source, values, usage and features of DWH data and defines how data can be changed and processed at every architecture layer. Meta data is stored in a Meta data repository which all the other architecture components can access[3]. The data that is used to represent other data is known as metadata. For example, the index of a book serves as a metadata for the contents in the book. In other words, we can say that metadata is the summarized data that leads us to detailed data. In terms of data warehouse.

1.2 Categories of Metadata

Metadata can be broadly categorized into three categories:

1.2.1 Business Metadata

It has the data ownership information, business definition, and changing policies.

1.2.2 Technical Metadata

It includes database system names, table and column names and sizes, data types and allowed values. Technical metadata also includes structural information such as primary and foreign key attributes and indices.

1.2.3 Operational Metadata

It includes currency of data and data lineage. Currency of data means whether the data is active, archived, or purged. Lineage of data means the history of data migrated and transformation applied on it.

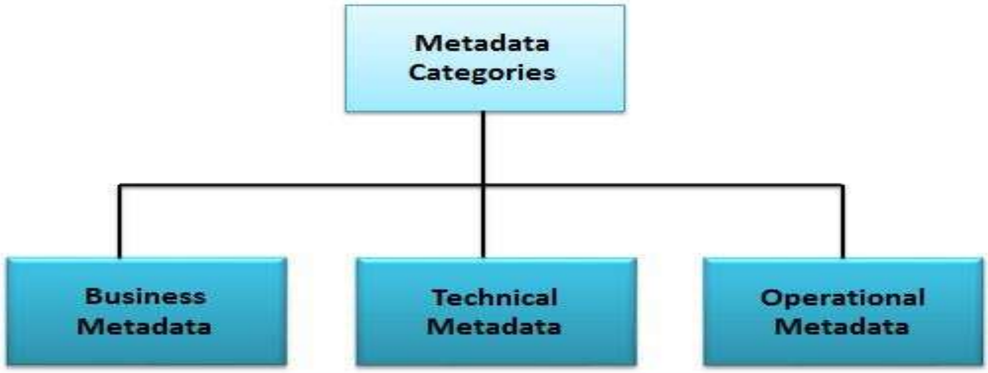

Fig 2: Diagram shows Metadata Categories

1.3 Role of Metadata

Metadata has a very important role in a data warehouse. The role of metadata in a warehouse is different from the warehouse data, yet it plays an important role. The various roles of metadata are explained below.

- ➢ Metadata helps in decision support system for mapping of data when data is transformed from operational environment to data warehouse environment.
- ➢ Metadata is used for query tools.
- ➢ Metadata is used in reporting tools.
- ➢ Metadata plays an important role in loading functions.

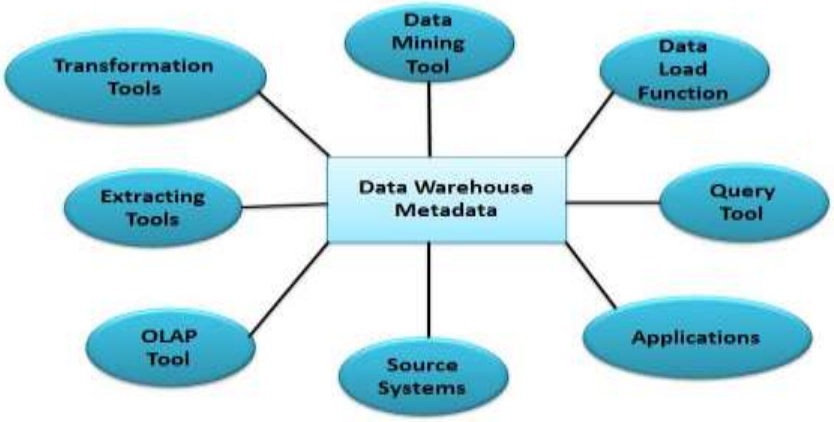

Fig 3: Diagram shows the roles of metadata.

1.4 Metadata Repository

Metadata repository is an integral part of a data warehouse system. It has the following metadata:

1.4.1 Definition of data warehouse - It includes the description of structure of data warehouse. The description is defined by schema, view, hierarchies, derived data definitions, and data mart locations and contents.

1.4.2 Business metadata - It contains has the data ownership information, business definition, and changing policies.

1.4.3 Operational Metadata - It includes currency of data and data lineage. Currency of data means whether the data is active, archived, or purged. Lineage of data means the history of data migrated and transformation applied on it.

1.4.4 Data for mapping from operational environment to data warehouse - It includes the source databases and their contents, data extraction, data partition cleaning, transformation rules, data refresh and purging rules.

1.4.5 Algorithms for summarization - It includes dimension algorithms, data on granularity, aggregation, summarizing, etc.

II. THE PROCESS OF DATA WAREHOUSE DESIGN

A DWH can be built using a Top –Down Approach, Bottom-Up Approach

2.1 Top-Down Approach

The data warehouse as a centralized repository for the entire enterprise. Data warehouse stores the 'atomic' data at the lowest level of detail[6]. Dimensional data marts are created only after the complete data warehouse has been created. Thus, data warehouse is at the center of the Corporate Information Factory (CIF), which provides a logical framework for delivering business intelligence.

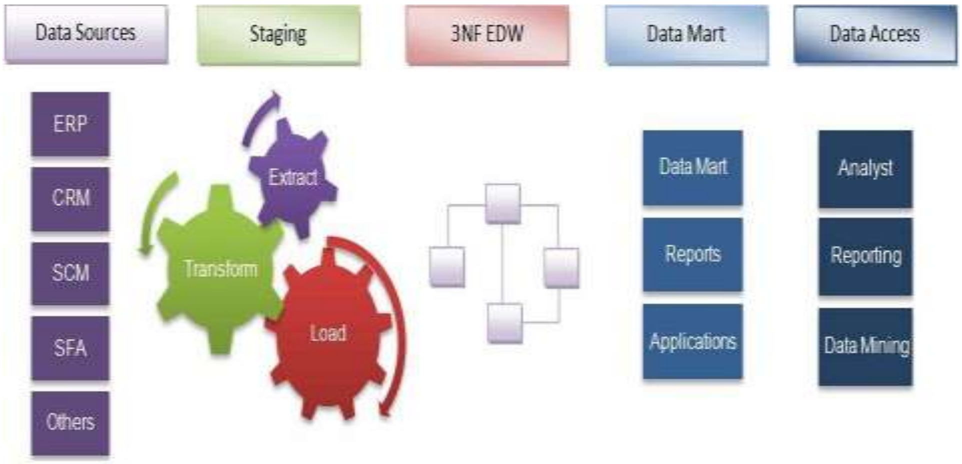

Fig 4: Top-Down Approach

The data warehouse in the following terms:

1. **Subject-oriented**

The data in the data warehouse is organized so that all the data elements relating to the same real-world event or object are linked together.

2. **Time-variant**

The changes to the data in the database are tracked and recorded so that reports can be produced showing changes over time.

3. **Non-volatile**

Data in the data warehouse is never over-written or deleted -- once committed, the data is static, read-only, and retained for future reporting.

4. **Integrated**

The database contains data from most or all of an organization's operational applications, and that this data is made consistent.

2.2 Bottom-Up Approach:

Keeping in mind the most important business aspects or departments, data marts are created first. These provide a thin view into the organizational data, and as and when required these can be combined into a larger data warehouse. Kimball defines data warehouse as “A copy of transaction data specifically structured for query and analysis”[7]. Kimball’s data warehousing architecture is also known as Data Warehouse Bus. Dimensional modeling focuses on ease of end user accessibility and provides a high level of performance to the data warehouse.

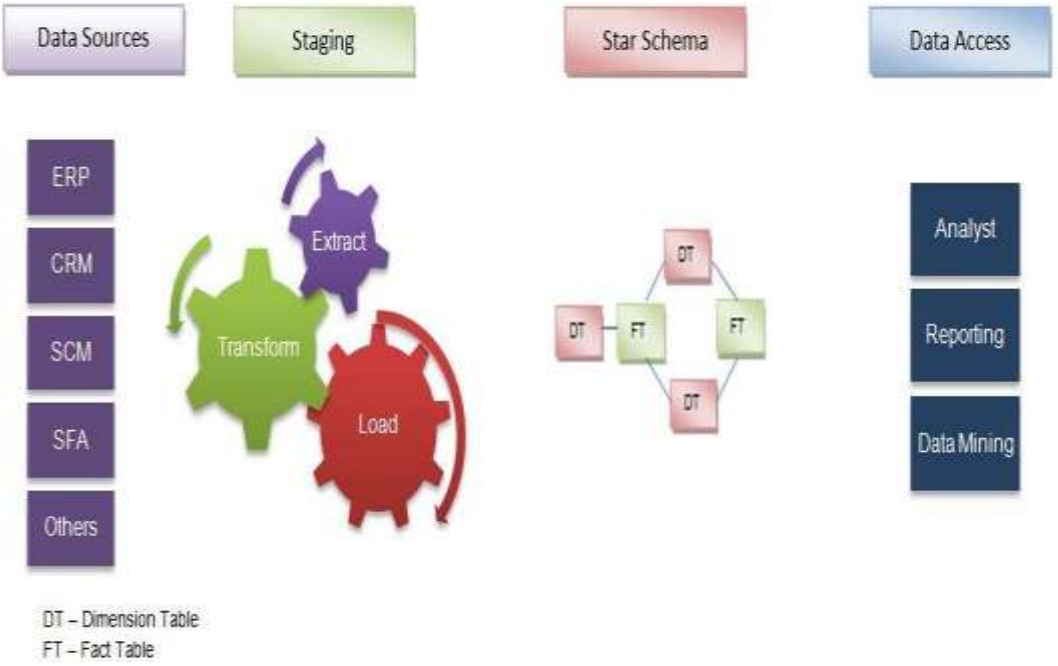

Fig 5: Bottom-Up Approach

3. OLAP Types

3.1. Relational OLAP (ROLAP) servers

3.1.1 ROLAP

This methodology relies on manipulating the data stored in the relational database to give the appearance of traditional OLAP's slicing and dicing functionality[4]. In essence, each action of slicing and dicing is equivalent to adding a "WHERE" clause in the SQL statement.

Advantages

➢ Can handle large amounts of data: The data size limitation of ROLAP technology is the limitation on data size of the underlying relational database. In other words, ROLAP itself places no limitation on data amount.

➢ Can leverage functionalities inherent in the relational database: Often, relational database already comes with a host of functionalities. ROLAP technologies, since they sit on top of the relational database, can therefore leverage these functionalities.

Disadvantages

➢ Performance can be slow Because each ROLAP report is essentially a SQL query (or multiple SQL queries) in the relational database, the query time can be long if the underlying data size is large[5].

➢ Limited by SQL functionalities: Because ROLAP technology mainly relies on generating SQL statements to query the relational database, and SQL statements do not fit all needs ROLAP technologies are therefore traditionally limited by what SQL can do

3.2 Typical data base schemes:

Star schema Star Schema consists of one or more fact table and one or more dimension tables that are related to foreign keys. Dimension tables are De-normalized, Fact table-normalized:

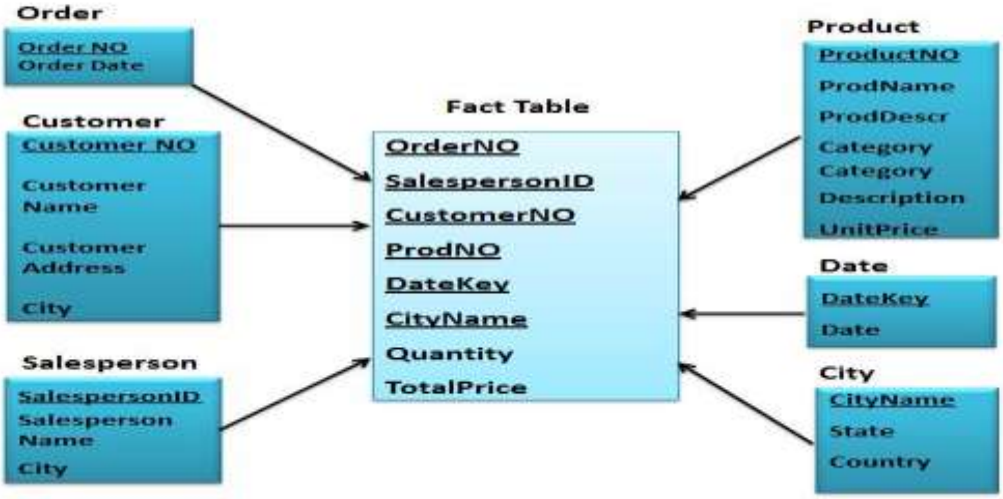

Fig 6: Star Schema

Snow flake schema is a normalize dimensions to eliminate the redundancy. The dimension data has been grouped into one large table. Both dimension and fact tables normalized.

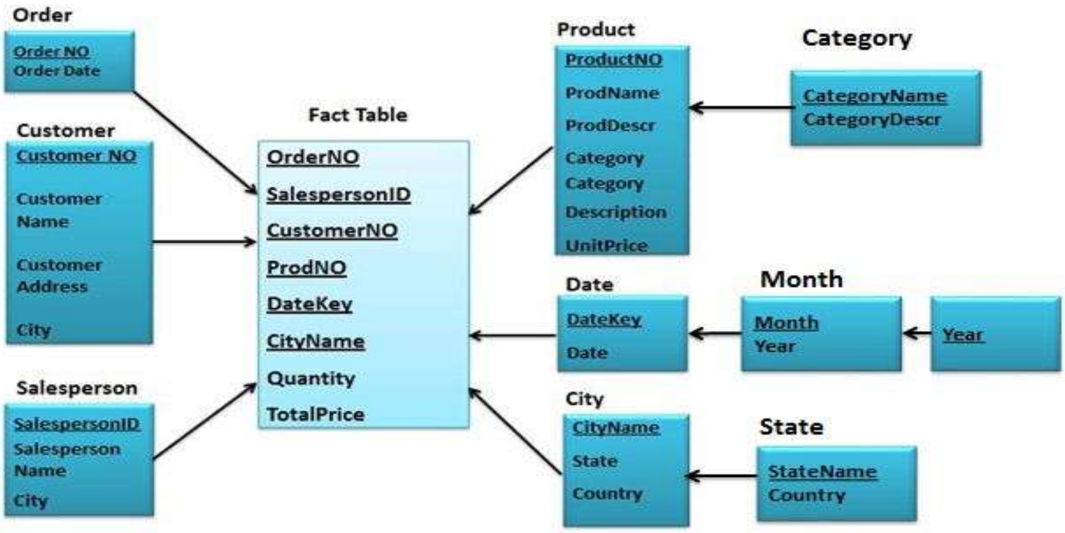

Fig 7: snowflake schema

3.3 Multi-dimensional OLAP (MOLAP) servers:

3.3.1 MOLAP

This is the more traditional way of OLAP analysis. In MOLAP, data is stored in a multidimensional cube. The storage is not in the relational database, but in proprietary formats.

Advantages

➢ Excellent performance: MOLAP cubes are built for fast data retrieval, and are optimal for slicing and dicing operations.

➢ Can perform complex calculations: All calculations have been pre-generated when the cube is created. Hence, complex calculations are not only doable, but they return quickly.

Disadvantages:

Limited in the amount of data it can handle Because all calculations are performed when the cube is built, it is not possible to include a large amount of data in the cube itself. This is not to say that the data in the cube cannot be derived from a large amount of data. Indeed, this is possible. But in this case, only summary-level information will be included in the cube itself.

3.4 Hybrid OLAP (HOLAP) servers

- The hybrid OLAP approach combines ROLAP and MOLAP technology, benefiting from the greater scalability of ROLAP and the faster computation of MOLAP.
- For example, a HOLAP server may allow large volumes of detailed data to be stored in a relational database, while aggregations are kept in a separate MOLAP store. The Microsoft SQL Server2000 supports a hybrid OLAP server[9].
- HOLAP technologies attempt to combine the advantages of MOLAP and ROLAP[11].
- When detailed information is needed, HOLAP can "drill through" from the cube into the underlying relational data

OLTP vs OLAP

(Below, there is no discussion of OLTP. How can it be OLTP versus OLAP? Change the side heading! What about, no side heading at all or Various OLAP.)

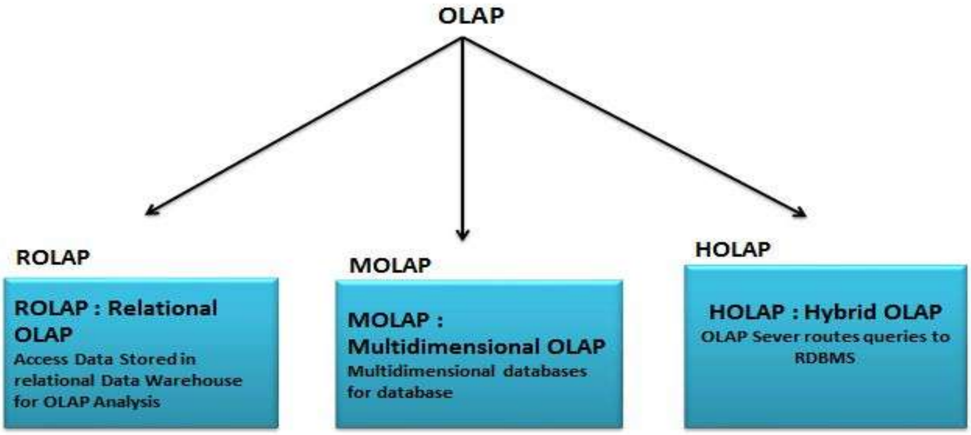
OLAP
ROLAP
MOLAP
HOLAP
ROLAP : Relational OLAP
Access Data Stored in relational Data Warehouse for OLAP Analysis
MOLAP : Multidimensional OLAP
Multidimensional databases for database
HOLAP : Hybrid OLAP
OLAP Sever routes queries to RDBMS

Fig 8: OLAP implementations technique

4. SCD Types:

In this case, an additional record is added into the customer dimension. SCD Type 1 The new incoming record (changed/modified data set) replaces the existing old record in target8]. SCD Type 3 In this approach, only the information about a previous value of a dimension is written into the database[10]. An 'old 'or 'previous' column is created which stores the immediate previous attribute. The beauty of this approach is it will maintain two versions, you will find two records the older version and the current version. In other words it maintains history. Again we can implement Type 2 in following methods.

1. Versioning
2. Effective Dates
3. Effective Dates & Current Record Indicators

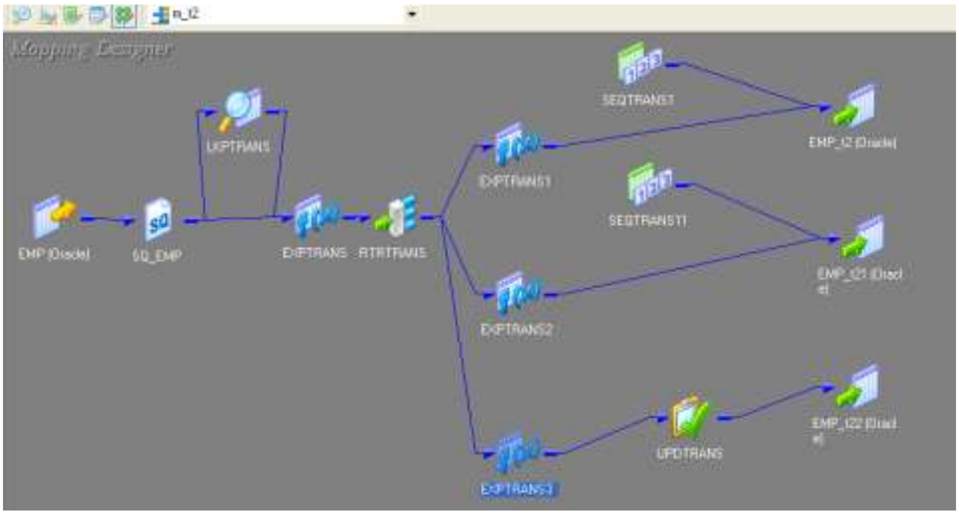

Fig 9: Slowly Changing Dimensions (SCDs) Flow

Customer Id	Customer First Name	Customer Last Name	Customer Country	Effective Year	Version
1	Kolli	Srikanth	India	2015	0
2	N.V.E.S	Murthy	U.S	2016	1

TABLE 1: Customer versioning Table

Customer Id	Customer First Name	Customer Last Name	Customer Country	Effective Start Date	Effective Start Date
1	Kolli	Srikanth	India	01/01/2015	12/31/2015
2	N.V.E.S	Murthy	U.S	01/01/2016	Till Date

TABLE 2: Customer Effective Dates Table

Customer Id	Customer First Name	Customer Last Name	Customer Country	Effective Start Date	Effective Start Date	Current Record
1	Kolli	Srikanth	India	01/01/2015	12/31/2015	No
2	N.V.E.S	Murthy	U.S	01/01/2016	Till Date	Yes

TABEL 3: Customer Effective Dates & Current Record Indicators Table

III. CONCLUSION

In Data base management system multiple requirements are available and include new business techniques. The above discussion we see that the Business Database management and Data warehousing depended on the performance of Meta data, OLTP and OLAP performance. In the data warehousing move toward information is requested, processed and merged continuously, so the information is readily available for direct querying OLAP and analysis at the warehouse. Slowly Changing Dimensions (SCDs) are dimensions that have data that changes slowly, rather than changing on a time-based, regular schedule. In SCD type 2 effective date, the dimension table will have Start Date and End Date as the fields. We can implementation on SCD TYPE-2 based on SCD TYPE-1 and new fields like Versioning, Effective Dates, and Effective Dates & Current Record Indicators.

REFERENCES

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

[2]. Zhuge, Y., H. Garcia-Molina, J. Hammer, J. Widom, “View Maintenance in a Warehousing Environment, Proc. Of SIGMOD Conf., 1995.

[3]. Metadata Standards and Metadata Registries: An Overview" (PDF). Retrieved 2011-12-23.

[4]. Data Warehousing, Data Mining and OLTP; Alex Berson, Stephen J. Smith 1997, McGraw Hill Page: 4, 14-16.

[5]. http: //searchbusinessintelligence.techtarget.in/tip/Inmon-vs-Kimball-Which- approach-is suitable-for-your-data-warehouse.

[6]. Informatica Power Center, Available at: www.informatica.com/ products/ data integration/ powercenter/ default.htm..

[7]. R. J. Davenport, September 2007. [Online] ETL vs. ELT: A Subjective View. In Source IT Consulting Ltd., U.K. Available at: http://www.insource.co.uk/pdf/ETL_ELT.pdf.

[8]. K.Srikanth, N.V.S.Murthy, J.Anitha : “Data Warehousing Concept Using ETL Process For SCD Type-1” Conf. on TIJCSA ,Volume 1, No. 10, December 2012 ISSN – 2278-1080.

[9]. T. Jun, C. Kai, Feng Yu, T. Gang, “The Research and Application of ETL Tools in Business Intelligence Project,” I n Proc. International Forum on Information Technology and Applications, 2009, IEEE, pp.620-623.

[10]. K.Srikanth, N.V.S.Murthy, J.Anitha : “Data Warehousing Concept Using ETL Process for SCD Type-2” Conf. on AJER ,Volume 2, , April 2013 Issue4, pp-86-91.

[11]. http://www.ijecs.in/issue/v2-i8/44%20ijecs.pdf. “Introduction on Data Warehouse with OLTP and OLAP, ISSN:2319-7242 Volume2 Issue 8 August, 2013 Page No. 2569-2573.