

DATA WAREHOUSE PRELUDE

- ✓ Data warehouse is a repository of an organization's electronically stored data and are designed to facilitate reporting and analysis
- ✓ Also emphasizes on the means to retrieve and analyze data, to extract, transform and load data, and to manage the data dictionary.
- ✓ An expanded definition for data warehousing includes business intelligence tools, tools to extract, transform, and load data into the repository, and tools to manage and retrieve metadata (data about data).
- ✓ In contrast to data, warehouses are operational systems that perform day-to-day transaction processing.
- ✓ A data warehouse is a collection of computer-based information that is critical to successful execution of enterprise initiatives
- ✓ It provides a tool to satisfy the information needs of the employee's at all organizational levels-not just for complex data queries but as a general facility for getting quick, accurate and often insightful information.
- ✓ It is designed so that its users can recognize the information they want and access that information using simple tools.
- ✓ One of the principal reasons for developing a Data Warehouse is to integrate operational data from various sources into a single and consistent architecture that supports analysis and decision making with the enterprise.
- ✓ Some of the applications data warehousing can be used for are:
 - Credit card churn analysis
 - Insurance fraud analysis
 - Call record analysis
 - Logistics management (part of Supply Chain Management that plans, implements, and controls the efficient, effective, forward, and reverse flow and

storage of goods, services, and related information between the point of origin and the point of consumption in order to meet customers' requirements)

ADVANTAGES

- **More cost effective decision making:** A data warehouse allows reduction of staff and computer resources required to support queries and reports against operational and production database. This typically offers significant savings.
- **Better enterprise intelligence:** Increased quality and flexibility of enterprise analysis arises from the multi-level data structure which guarantees data accuracy and reliability ensuring that a Data Warehouse contains only "trusted" data.
- **Enhanced customer service:** An enterprise can maintain better customer relationships by correlating all customer data via a single Data Warehouse Architecture.
- **Business reengineering:** Allowing unlimited analysis of enterprise information often provides insights to enterprise processes that may yield breakthrough ideas for engineering those processes. Knowing what information is important to an enterprise will provide direction and priority for reengineering efforts.

DATA MINING PRELUDE

- ✓ Data mining is the process of extracting hidden predictive information from a large database. As more data are gathered, with the amount of data doubling every year, data mining is becoming an increasingly important tool to transform this data into information.
- ✓ It is commonly used in a wide range of profiling practices, such as marketing, fraud detection and scientific discovery. Data mining can be applied to data sets of any size.

- ✓ Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge driven decisions.
- ✓ Data mining sometimes called data or knowledge discovery is the process of analyzing data from different perspectives and summarizing it into useful information.
- ✓ Data mining software is an analytical tool for analyzing data. It allows users to analyze data from many different dimensions, categorize it, and summarize the relationships identified.
- ✓ Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.
 - E.g. – analysis of retail sales data to identify apparently unrelated products that are often purchased together.
- ✓ **Database can be larger in both depth and breadth:**
 - The databases can have more columns and rows. High performance data mining allows users to explore full depth of a database, without pre-selecting a subset of variables. The data mining database contain larger samples (more rows) as they yield lower estimation errors and variance, and allow users to make conclusion about small but important segments of a population.

Data mining consists of five major elements:

- ✓ Extract, transform, and load transaction data onto the data warehouse system.
- ✓ Store and manage the data in a multidimensional database system.
- ✓ Provide data access to business analysts and information technology professionals.
- ✓ Analyze the data by application software.
- ✓ Present the data in a useful format, such as a graph or table.

ADVANTAGES OF DATA MINING

- ✓ **Automated prediction of trends and behaviors :**
 - Data mining automates the process of finding predictive information in large databases. Questions that traditionally required extensive hands-on analysis can now be answered directly from the data, quickly.
- ✓ **Automated Discovery of previously unknown patterns:**
 - Data mining tools sweep through databases and identify previously hidden patterns in one step.

CHAPTER 4

Applications of Data Warehousing and Data Mining in Government

4.1 INTRODUCTION

Data warehousing and data mining are the important means of preparing the government to face the challenges of the new millennium.

Data warehousing and data mining technologies have extensive potential applications in the government—in various Central Government sectors such as Agriculture, Rural Development, Health and Energy and also in State Government activities. These technologies can and should therefore be implemented.

In this chapter, we shall examine their potential applications in the State and Central Government.

4.2 NATIONAL DATA WAREHOUSES

A large number of national data warehouses can be identified from the existing data resources within the Central Government Ministries. Let us examine these potential subject areas on which data warehouses may be developed at present and also in future.

4.2.1 Census Data

The Registrar General and Census Commissioner of India decennially compiles information of all individuals, villages, population groups, etc. This information is wide ranging such as the individual-slip, a compilation of information of individual households, of which a database of 5% sample is maintained for analysis. A data warehouse can be built from this database upon which OLAP techniques can be applied. Data mining also can be performed for analysis and knowledge discovery.

A village-level database was originally developed by National Informatics Centre at Hyderabad under General Information Services Terminal of National Informatics Centre (GISTNIC) for the 1991 Census. This consists of two parts: primary census abstract and village amenities. Subsequently, a data warehouse was also developed for village amenities for Tamil Nadu. This enables multidimensional analysis of the village level data in such sectors as Education, Health and Infrastructure. The fact data pertains to the individual village data compiled under 1991 Census.

As the census compilation is performed once in ten years, the data is quasi-static and, therefore, no refreshing of the warehouse needs to be done on a periodic basis. Only the new data needs to be either appended to the data warehouse or alternatively a new data warehouse can be built.

There exist many other subject areas (e.g. migration tables) within the census purview which may be amenable and appropriate for data warehouse development, OLAP and data mining applications on which work can be taken up in future.

4.2.2 Prices of Essential Commodities

The Ministry of Food and Civil Supplies, Government of India, compiles daily data (on weekly basis) for about 300 observation centres in the entire country on the prices of essential commodities such as rice, edible oils, etc. This data is compiled at the district level by the respective State Government agencies and transmitted online to Delhi for aggregation and storage. A data warehouse can be built for this data, and OLAP techniques can be applied for its analysis. A data mining and forecasting technique can be applied for advance forecasting of the actual prices of these essential commodities. The forecasting model can be strengthened for more accurate forecasting by taking into account the external factors such as rainfall, growth rate of population and inflation.

A limited exercise in this direction was already executed at a State level (in Tamil Nadu).

4.3 OTHER AREAS FOR DATA WAREHOUSING AND DATA MINING

Other possible areas for data warehousing and data mining in Central Government sectors are discussed in detail as under.

4.3.1 Agriculture

The Agricultural Census performed by the Ministry of Agriculture, Government of India, compiles a large number of agricultural parameters at the national level. District-wise agricultural production, area and yield of crops is compiled; this can be built into a data warehouse for analysis, mining and forecasting. Statistics on consumption of fertilizers also can be turned into a data mart.

40 Applications of Data Warehousing and Data Mining in Government

Data on agricultural inputs such as seeds and fertilizers can also be effectively analyzed in a data warehouse. Data from livestock census can be turned into a data warehouse. Land-use pattern statistics can also be analyzed in a warehousing environment. Other data such as watershed details and also agricultural credit data can be effectively used for analysis by applying the technologies of OLAP and data mining.

Thus there is substantial scope for application of data warehousing and data mining techniques in Agricultural sector.

4.3.2 Rural Development

Data on individuals below poverty line (BPL survey) can be built into a data warehouse. Drinking water census data (from Drinking Water Mission) can be effectively utilized by OLAP and data mining technologies. Monitoring and analysis of progress made on implementation of rural development programmes can also be made using OLAP and data mining techniques.

4.3.3 Health

Community needs assessment data, immunization data, data from national programmes on controlling blindness, leprosy, malaria can all be used for data warehousing implementation, OLAP and data mining applications.

4.3.4 Planning

At the Planning Commission, data warehouses can be built for state plan data on all sectors: labour, energy, education, trade and industry, five year plan, etc.

4.3.5 Education

The Sixth All India Educational Survey data has been converted into a data warehouse (with about 3 GB of data). Various types of analytical queries and reports can be answered.

4.3.6 Commerce and Trade

Data bank on trade (imports and exports) can be analyzed and converted into a data warehouse.* World price monitoring system can be made to perform better by using data warehousing and data mining technologies. Provisional estimates of import and export also be made more accurate using forecasting techniques.

*This data is available with the Director General of Foreign Trade, Ministry of Commerce.

4.3.7 Other Sectors

In addition to the above mentioned important applications, there exist a number of other potential application areas for data warehousing and data mining, as follows:

Tourism. Tourist arrival behaviour and preferences; tourism products data; foreign exchange earnings data; and Hotels, Travel and Transportation data.

Programme Implementation. Central projects data (for monitoring).

Revenue. Customs data, central excise data, and commercial taxes data (state government).

Economic affairs. Budget and expenditure data; and annual economic survey.

Audit and accounts. Government accounts data.

All government departments or organizations are deeply involved in generating and processing a large amount of data. Conventionally, the government departments have largely been satisfied with developing single management information systems (MIS), or in limited cases, a few databases which were used online for limited reporting purposes. Much of the analysis work was done manually by the Department of Statistics in the Central Government or in any State Government. The techniques used for analysis were conventional statistical techniques on largely batch-mode processing. Prior to the advent of data warehousing and data mining technologies nobody was aware of any better techniques for this activity. In fact, data warehousing and data mining technologies could lead to the most significant advancements in the government functioning, if properly applied and used in the government activities. With their advent and prominence, there is a paradigm shift which may finally result in improved governance and better planning by better utilization of data. Instead of the officials wasting their time in processing data, they can rely on data warehousing and data mining technologies for their day-to-day decision making and concentrate more on the practical implementation of the decisions so taken for better performance of developmental activities.

Further, even though various departments in the government (State or Central) are functionally interlinked, the data is presently generated, maintained and used independently in each department. This leads to poor (independent) decision making and isolated planning. Herein lies the importance of data warehousing technology. Different data marts for separate departments, if built, can be integrated into one data warehouse for the government. This is true for State Government and Central Government. Thus data warehouses can be built at Central level, State level and also at District level.

Conclusion

In the government, the individual data marts are required to be maintained by the individual departments (or public sector organizations) and a central data

42 *Applications of Data Warehousing and Data Mining in Government*

warehouse is required to be maintained by the ministry concerned for the concerned sector. A generic inter-sectoral data warehouse is required to be maintained by a central body (as Planning Commission). Similarly, at the State level, a generic inter-departmental data warehouse can be built and maintained by a nodal agency, and detailed data warehouses can also be built and maintained at the district level by an appropriate agency. National Informatics Centre may possibly play the role of the nodal agency at Central, State and District levels for developing and maintaining data warehouses in various sectors.