Chapter - 6

DATA WAREHOUSING, DATA MINING AND
MANAGEMENT INFORMATION SYSTEM

6.1   Introduction

6.1.1   The term *Management Information System (MIS)* is not new to the banking sector. Since the early 80s, banks have been using this terminology to refer to the process of generating various reports and analyses at the Corporate/Head offices for their decision making for own use as well as for conveyance to authorities in charge of regulation. Often, these reports are generated through computers and can be generated at any point of time. However, the usage of the terms *data warehousing* and *data mining* are relatively new. These terms have gained significance with the growing sophistication of technology and the need for predictive analysis with *What if* simulations. *MIS* in the present context of high availability of voluminous data on electronic media at diverse locations and on diverse platforms, has become more pertinent to banks' decision-making process, thanks to the availability of new tools of technology such as data warehousing, data mining. Data warehousing which refers to collection of data from various sources (internal and external) and placing them in a form suitable for further processing which will gain critical importance in the presence of data mining which refers to the process of extracting hidden information and generating several types of analytical reports which are usually not available in the original transaction processing systems. A brief write-up on the concepts of data warehousing and data mining is placed at **Annexure 15**. Management Information System would thus be the end product of both the processes - data warehousing and data mining.

6.2   Relevance of Data Warehousing and Data Mining for banks in India

6.2.1   Banking being an information intensive industry, building a Management Information System within a bank or an industry is a gigantic task. It is more so for the public sector banks which have a wide network of bank branches spread all over the country. It becomes all the more difficult due to prevalence of varying degrees of computerisation. At present, banks generate MIS reports largely from periodic paper reports/ statements submitted by the branches and regional/zonal offices. Except for a few banks which have been using technology in a big way, MIS reports are available with a substantial time lag. Reports so generated have also a high margin of error due to data entry being done at various levels and the likelihood of varying interpretations at different levels.

6.2.2   Though computerisation of bank branches has been going on at a good pace, MIS requirements have not been fully addressed to. It is on account of the fact that most of the Total Branch Comuputerisation (TBC) software packages are transaction processing oriented. They have been designed primarily for day-to-day operations at the branch level and day-end balancing of books. There are only a few packages used by a limited number of branches which can easily be interfaced with the computer systems at Zonal/ Head

Offices and have the capability to generate MIS data. Banks have not implemented such packages partly because of the high costs and partly because of the absence of any strategic plan to collate information at the corporate level by using the TBC packages.

6.2.3   However, the need for building MIS at the corporate level has increased considerably during the last few years because of the following reasons :

- Regulatory requirements indicated by the RBI for preparation of Off-site Monitoring Surveillance (OSMOS) Reports on a regular basis in electronic format
- Regulatory requirement of filing of statutory returns such as the one under Section 42 of the Reserve Bank of India Act, 1934 for working out Cash Reserve Ratio (CRR) and Statutory Liquidity Ratio (SLR) obligations in electronic format
- Asset Liability Management (ALM) guidelines for banks being implemented by the RBI w.e.f. April 1, 1999 with the stipulation that the banks should capture 100 percent of their business through the ALM system by April 1, 2000.
- Need for timely submission of Balance Sheets and Profit & Loss Accounts
- Focus on transaction costing and a need for relating the service charges levied on the customers to be based on cost of servicing
- Need for Inter-Branch Reconciliation of Accounts within a definite time frame
- Need to meet the stipulations made by the Central Vigilance Commission (CVC) to computerise at least 70 percent of banking business by January 1, 2001.
- Need to undertake risk management strategies and for this purpose build up appropriate sets of data and market intelligence reports.

Building MIS to meet the above objectives would not be possible by replicating the existing TBC package. MIS capability needs to be incorporated in the packages so that the controlling offices/corporate offices can generate the required monitoring reports without having to make another round of data entry at their end.

6.2.4   In this connection, application of data warehousing and data mining techniques appears to be the appropriate solution. The implication of adopting such technology in a bank would be as under :
- All transactions captured at the branch level would get consolidated at a central location. Such a central location could be called the Data Warehouse of the concerned bank. For this to happen, one of the requirements would be to establish connectivity between the branches on the one hand and the Data Warehouse platform on the other.
- For banks with large number of branches, it may not be desirable to consolidate the transaction details at one place only. It can be decentralised by locating the services on regional basis. The regional Data marts as developed can provide mutual back-up and could be linked to the central Data Warehousing server so that for the purpose of MIS at the corporate level, data can be accessed from all the regional Data marts.
- By way of data mining techniques, data available at various computer systems can be accessed and by a combination of techniques like *classification, clustering, segmentation, association rules, sequencing, decision tree*

(described in detail at **Annexure-15**), various ALM reports such as Statement of Structural Liquidity, Statement of Interest Rate Sensitivity etc. or accounting reports like Balance Sheet and Profit & Loss Account can be generated instantaneously for any desired period/date.

6.2.5   Significant cost benefits, time savings, productivity gains and process re-engineering opportunities are associated with the use of data warehouse for information processing. Data can easily be accessed and analysed without time consuming manipulation and processing. Decisions can be made more quickly and with confidence that the data are both time-relevant and accurate. Integrated information can be also kept in categories that are meaningful to profitable operation.

6.2.6   Trends can be analysed and predicted with the availability of historical data and the data warehouse assures that everyone is using the same data at the same level of extraction, which eliminates conflicting analytical results and arguments over the source and quality of data used for analysis. In short, data warehouse enables information processing to be done in a credible, efficient manner.

6.3   **Recommendations**

6.3.1   The Committee recognises the need for data warehouses both at the individual bank level and at industry level. The argument that *it is too early for such technology in India* does not hold good for the banking industry which is primarily an industry dealing with facts and figures. For implementing various regulatory guidelines including the latest one on ALM, a robust MIS, founded on data warehousing and data mining, at individual bank level is essential. The structure, configuration and design of the data warehouse may, however, differ from bank to bank.

6.3.2   It is not necessary to wait for all bank branches to be computerised for setting up of data warehouse. Neither is it necessary for all branches to have the same TBC software package. Data warehouse can be established even across multiple computer platforms as long as the transaction details are made available to the data warehouses in standardised formats. Therefore, banks should standardise the data formats and start supplying the data on a continuous basis from the branches which have already been computerised. It is expected that the computerised branches themselves would provide the critical data for a data warehouse to go live. The Committee recommends that all banks should put in place their data warehouse strategy by January 1, 2001. The banks with a large number of computerised branches may start their pilot projects by warehousing certain categories of data (if not all the transactions) by April 1, 2001. Some illustrative application areas are :
  - Investment Analysis
  - Credit Analysis
  - Customer Base Analysis
  - Defaulters Analysis

6.3.3   For building data bases at the individual customer level within a bank or at the industry level, it may be advisable to follow a unique identification number for all bank

customers. A Task Force may be set up by IBA to explore feasible methodology for working out a unique identification system.

6.3.4 While building the industry level data warehouse, legal questions relating to confidentiality of information may arise. The Standing Committee on Legal Issues relating to Electronic Banking as recommended in Chapter 7 of the Report may examine this issue. However, for the data collected under the regulatory provisions, the Reserve Bank of India could establish a Data Warehouse on Banking and Finance. The Department of Supervision, the Department of Banking Operations & Development and the Exchange Control Department of the Reserve Bank of India have already been receiving large amount of data. The Department of Statistical Analysis & Computer Services and the Department of Economic Analysis & Policy have also been receiving various statistical returns. Data so collected do not have any legal sensitivity and can well be used for data warehousing and data mining.

6.3.5 The Indian Banks' Association may initiate the process of building another Industry Level Data Warehouse, based on agreements to be signed by the participating banks on sharing of data. This data warehouse may mask the customer information, but it should be based on individual customer information so that the participating institutions can derive the benefit of business segmentation analysis and trend forecasting on various banking operations.