UNIT-V MINING ENVIRONMENT

Data Mining Environment: Case studies in building business environment, Application of data ware housing and Data mining in Government, National Data ware houses and case studies.

- A1) Define: Data mining
- A2) Explain about the applications of data mining in government
- A3) List some examples of use of data mining for application in science and business
- B1) Write a short notes on "Data mining environment".
- B2) Give examples of commercial Data Mining systems
- B3) Discuss about different case studies in building business environment
- B4) What is OLAP. In what it is different from OLTP. Name any ten differences between OLTP and OLAP

Data Mining Environment

What is Data Mining

Extracting or "mining" knowledge from large amounts of data.

Similar terms:

- knowledge mining from data,
- knowledge extraction,
- data/pattern analysis,
- data archaeology,
- data dredging.
- Knowledge Discovery from Data, or KDD.

Data mining as a step in the process of knowledge discovery

Knowledge discovery process consists of an iterative sequence of the following steps:

1. **Data cleaning** (to remove noise and inconsistent data)
2. **Data integration** (where multiple data sources may be combined)
3. **Data selection** (where data relevant to the analysis task are retrieved from the database)
4. **Data transformation** (where data are transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations)
5. **Data mining** (an essential process where intelligent methods are applied in order to extract data patterns)
6. **Pattern evaluation** (to identify the truly interesting patterns representing knowledge based on some interestingness measures)
7. **Knowledge presentation** (where visualization and knowledge representation techniques are used to present the mined knowledge to the user)
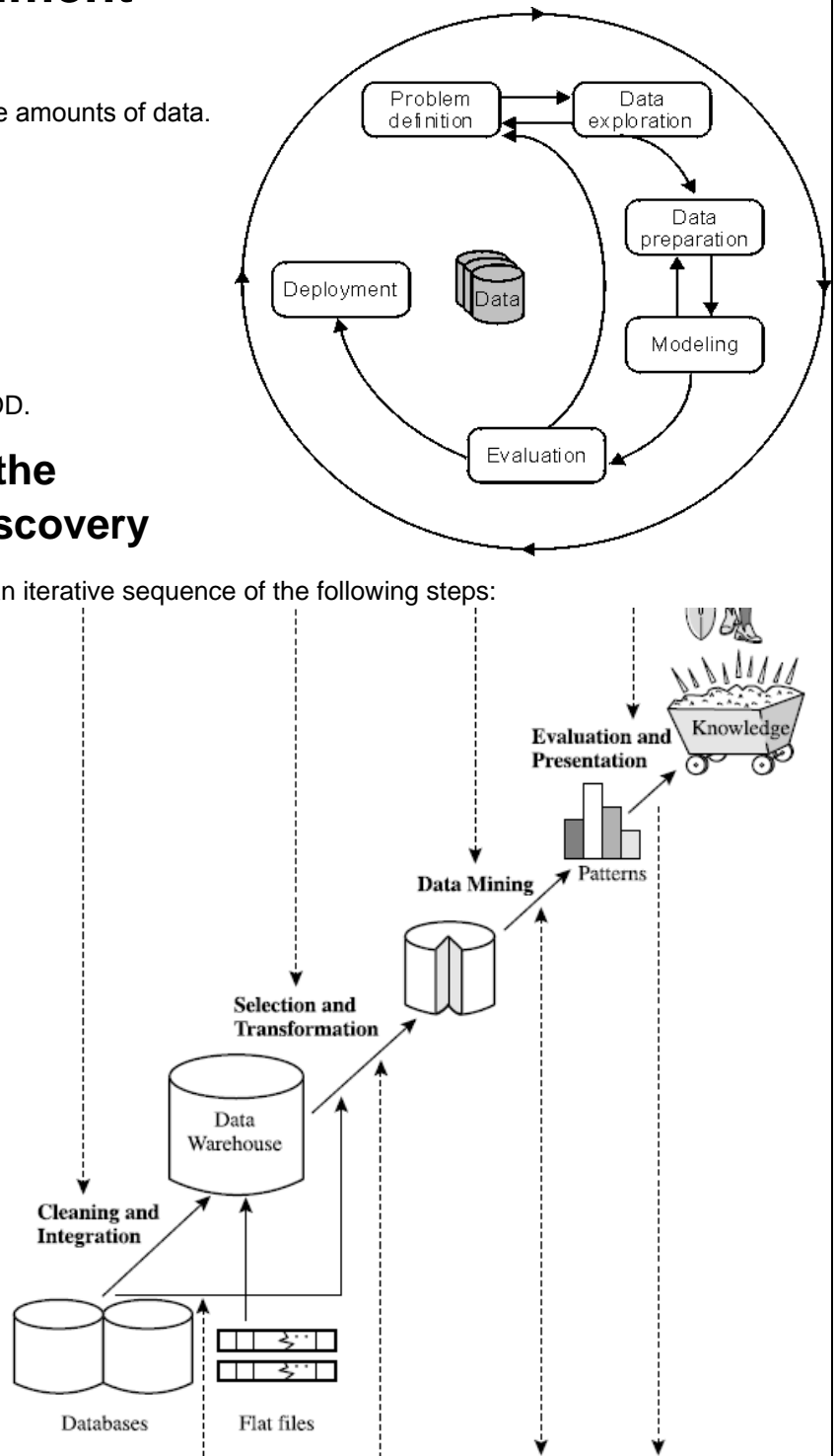
Architecture of a typical data mining system
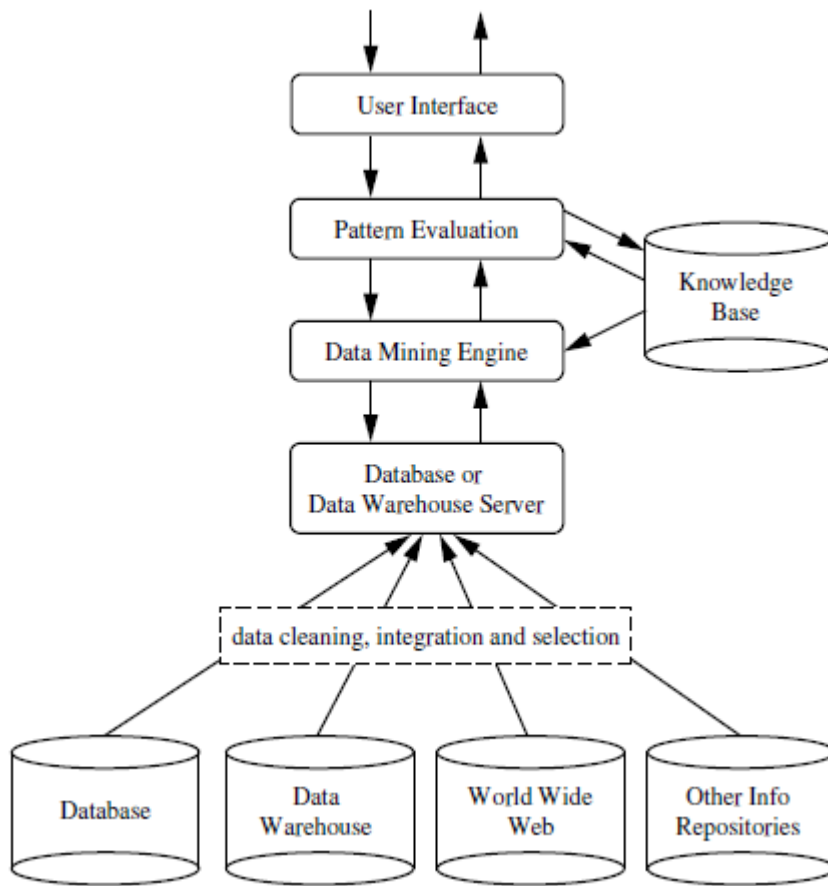
Database, data warehouse,WorldWideWeb, or other information repository:

- This is one or a set of databases, data warehouses, spreadsheets, or other kinds of information repositories
- Data cleaning and data integration techniques may be performed on the data.

Database or data warehouse server:

The database or data warehouse server is responsible for fetching the relevant data, based on the user's data mining request

Knowledge base

- This is the domain knowledge that is used to guide the search or evaluate the result patterns
- can include concept hierarchies, user beliefs, used to organize attributes or attribute values into different levels of abstraction.
- Other examples of domain knowledge are additional interestingness constraints or thresholds, and metadata (e.g., describing data from multiple heterogeneous sources).

Data mining engine

This is essential to the data mining system and consists of a set of functional modules for tasks such as

- characterization,
- association and correlation analysis,
- classification,
- prediction,
- cluster analysis,
- outlier analysis,
- evolution analysis.

Pattern evaluation module:

This component typically employs interestingness measures and interacts with the data mining modules so as to focus the search toward interesting patterns.

User interface:

- This module communicates between users and the data mining system,
- allowing the user to interact with the system by specifying a data mining query or task,
- providing information to help focus the search,
- performing exploratory data mining based on the intermediate data mining results.
- also allows the user to browse database and data warehouse schemas or data structures, evaluate mined patterns, and visualize the patterns in different forms.

Classification of Data Mining Systems
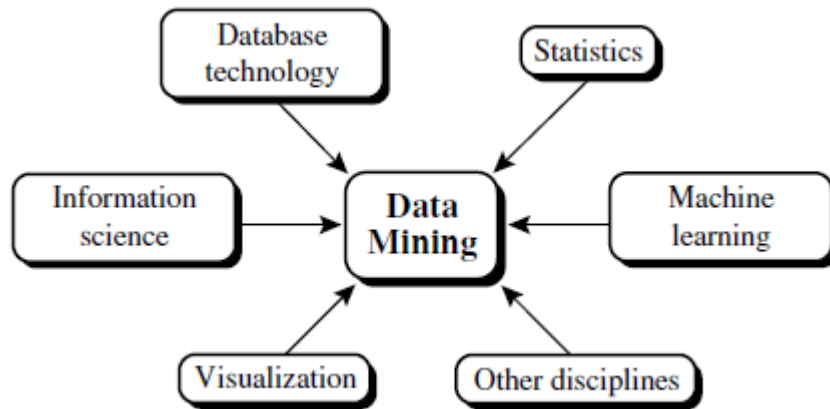
Kinds of Database technology

- Database systems can be classified according to different criteria (such as data models, or the types of data or applications involved)
- according to data models, relational, transactional, object-relational, or data warehouse mining system
- types of data handled: spatial, time-series, text, stream data, multimedia data mining system, or a World Wide Web mining system

Kinds of knowledge

- granularity or levels of abstraction of the knowledge mined
- generalized knowledge, primitive-level knowledge, knowledge at multiple levels

kinds of techniques

- degree of user interaction involved: autonomous systems, interactive exploratory systems, query-driven systems
- methods of data analysis employed: machine learning, statistics, visualization, pattern recognition, neural networks,
- applications adapted: finance, telecommunications, DNA, stock markets, e-mail

Applications of Data Mining (O)

Data Mining for Financial Data Analysis

- banks and financial institutions
- variety of banking services - checking and savings accounts for business or individual customers
- credit - business, mortgage, and automobile loans
- investment services – mutual funds
- Financial data collected in the banking and financial industry are often relatively complete, reliable, and of high quality, which facilitates systematic data analysis and data mining

Various cases are,

Design and construction of data warehouses for multi dimensional data analysis and data mining

one may like to view the debt and revenue changes by month, by region, by sector, and by other factors, along with maximum, minimum, total, average, trend, and other statistical information

Loan payment prediction and customer credit policy analysis

- Data mining methods, such as attribute selection and attribute relevance ranking, may help identify important factors
- example, factors related to the risk of loan payments include loan-to-value ratio, term of the loan, debt ratio (total amount of monthly debt versus the total monthly income), payment to income ratio, customer income level, education level, residence region, and credit history

Classification and clustering of customers for targeted marketing

Customers with similar behaviors regarding loan payments may be identified by multidimensional clustering techniques

Detection of money laundering and other financial crimes

- integrate information from multiple databases
- Multiple data analysis tools can then be used to detect unusual patterns, such as large amounts of cash flow at certain periods, by certain groups of customers etc

Data Mining for the Retail Industry

- collects huge amounts of data on sales, customer shopping history, transportation, consumption, service.
- many stores also have websites where customers can make purchases on-line
- Businesses such as Amazon exist solely on-line, without any brick-and-mortar (i.e., physical) store locations

Advantages

- identify customer buying behaviors,
- discover customer shopping patterns and trends,
- improve the quality of customer service,
- achieve better customer retention and satisfaction,
- enhance goods consumption ratios,
- design more effective goods transportation and distribution policies
- reduce the cost of business

Examples of data mining in the retail industry are

- Design and construction of data warehouses based on the benefits of data mining
- Multidimensional analysis of sales, customers, products, time, and region
- Analysis of the effectiveness of sales campaigns
- Customer retention—analysis of customer loyalty
- Product recommendation and cross-referencing of items

Data Mining for the Telecommunication Industry

- Multidimensional analysis of telecommunication data
  - used to identify and compare the data traffic, system workload, resource usage, group behavior, profit
- Fraudulent pattern analysis and the identification of unusual patterns
- Multidimensional association and sequential pattern analysis
- Mobile telecommunication services
- Use of visualization tools in telecommunication data analysis

Data Mining for Biological Data Analysis

- bioinformatics.The identification of DNA or amino acid sequence patterns

Data mining may contribute to biological data analysis in the following aspects

- Semantic integration of heterogeneous, distributed genomic and proteomic databases
- Alignment, indexing, similarity search, and comparative analysis of multiple nucleotide / protein sequences
- Discovery of structural patterns and analysis of genetic networks and protein pathways
- Association and path analysis: identifying co-occurring gene sequences and linking genes to different stages of disease development
- Visualization tools in genetic data analysis

Data Mining in Other Scientific Applications

- geosciences, astronomy, and meteorology
- sophisticated telescopes, multispectral high-resolution remote satellite sensors, global positioning systems
- fast numerical simulations in various fields, such as climate and ecosystem modeling, chemical engineering, fluid dynamics, and structural mechanics

challenges brought about by emerging scientific applications

- Data warehouses and data preprocessing
  - o geospatial data -spatial and temporal
- Mining complex data types
- Graph-based mining
- Visualization tools and domain-specific knowledge

Data Mining for Intrusion Detection

- misuse detection strategy – signature
- anomaly detection – profiles

Areas in which data mining technology may be applied or further developed for intrusion detection,

- Development of data mining algorithms for intrusion detection
- Association and correlation analysis, and aggregation to help select and build discriminating attributes
- Analysis of stream data
- Distributed data mining
- Visualization and querying tools

Examples of Commercial Data Mining Systems

Commercial data mining systems are organized into three groups:

- data mining products offered by large database or hardware vendors
- those offered by vendors of statistical analysis software
- those originating from the machine learning community

Offered By Large Database Or Hardware Vendors

Intelligent Miner by IBM

data mining functions

- association mining, classification, regression, predictive modeling, deviation detection, clustering, sequential pattern analysis

also provides an application toolkit containing

- neural network algorithms, statistical methods, data preparation tools, data visualization tools

Advantages

- scalability of its mining algorithms
- tight integration with IBM's DB2 relational database system

Microsoft SQL Server 2005

data mining functions

- association mining, classification (using decision tree, naïve Bayes, and neural network algorithms), regression trees, sequence clustering, time-series analysis
- integration of algorithms developed by third-party vendors and application users

MineSet from Purple Insight

data mining functions

- association mining, classification, advanced statistics and visualization tools

Uniqueness: has a set of robust graphics tools, including

- rule visualizer, tree visualizer, map visualizer,
- multidimensional data scatter visualizer for the visualization of data and data mining results

Oracle Data Mining (ODM)

- option to Oracle Database 10g Enterprise Edition

data mining functions

- association mining, classification, prediction, regression, clustering, sequence similarity search / analysis.

Offered By Vendors Of Statistical Analysis Software

Clementine by SPSS

- provides an integrated data mining development environment for end users and developers.

data mining functions

- association mining, classification, prediction, clustering, visualization tools

Uniqueness

object oriented, extended module interface, which allows users' algorithms and utilities to be added to Clementine's visual programming environment.

Enterprise Miner by SAS Institute, Inc

data mining functions

- association mining, classification, regression, clustering, time series analysis, statistical analysis packages

Uniqueness

- variety of statistical analysis tools, based on their long history

Insightful Miner, from Insightful Inc.

data mining functions,

data cleaning, classification, prediction, clustering, statistical analysis packages, visualization tools

distinguishing feature

its visual interface, which allows users to wire components together to create self-documenting programs

Originating From The Machine Learning Community

CART (Classification and Regression Trees), from Salford Systems

- It creates decision trees for classification and regression trees for prediction.
- CART employs **boosting** to improve accuracy.
- Several attribute selection measures are available.

See5 and C5.0, from RuleQuest

- See5 is the Windows version of C4.5, while C5.0 is its UNIX counterpart
- Both incorporate boosting.
- The source code is also provided.

Weka, developed at the University of Waikato in New Zealand

- open-source data mining software in Java
- It contains a collection of algorithms for data mining tasks, including data preprocessing, association mining, classification, regression, clustering, and visualization.

Case studies in building business environment

TBD

OLAP (On-line Analytical Processing)

- is characterized by relatively low volume of transactions.
- Queries are often very complex and involve aggregations.
- For OLAP systems a response time is an effectiveness measure.
- OLAP applications are widely used by Data Mining techniques.
- In OLAP database there is aggregated, historical data, stored in multi-dimensional schemas (star schema).
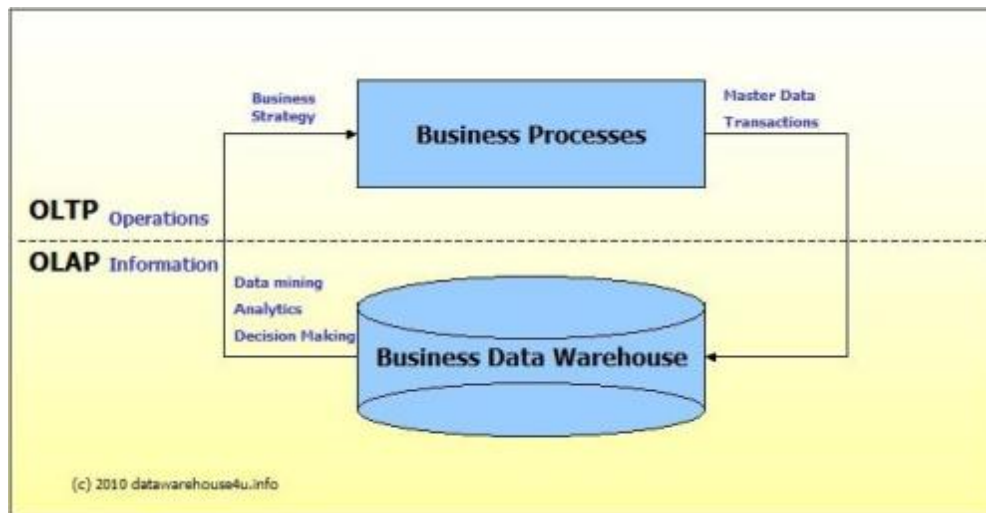
OLTP (On-line Transaction Processing)

- is characterized by a large number of short on-line transactions (INSERT, UPDATE, DELETE).
- The main emphasis for OLTP systems is put on very fast query processing, maintaining data integrity in multi-access environments and an effectiveness measured by number of transactions per second.
- In OLTP database there is detailed and current data, and schema used to store transactional databases is the entity model (usually 3NF).

Online Transaction Processing (OLTP)	Online Analytical Processing (OLAP)
Application Oriented	Used to analyze and forecast business needs
Up to date and consistent at all times	Data is consistent only up to the last update
Detailed data	Summarized data
Isolated data	Integrated data
Queries touch small amount of data	Queries touch large amounts of data
Fast response time	Slow response time
Updates are frequent	Updates are less frequent
Concurrency is the biggest performance concern	Each report or query requires lot of resources
Clerical Users	Managerial/Business Users
OLTP targets specific process like ordering from an online store	OLAP integrates data from different processes like (Ordering, processing, inventory, sales etc.,)
Performance sensitive	Performance relaxed
Few accessed records per time	Large volumes accessed at a time

Read/Update access	Mostly read and occasional update
No redundancy	Redundancy cannot be avoided
Databases size is usually around 100 MB to 100 GB	Databases size is usually around 100 GB to a few TB
Only current data available (old data is replaced by current data by updating)	Both current and historic data available (current is appended to historic data)
Short transactions (single granularity or more)	Long database transactions
Online update/insert/delete transactions	Batch update/insert/ delete transactions
High volume of transactions in a given period	Low volume transactions, periodic refreshing
Concurrency control and transaction recovery	No concurrent transactions and therefore no recovery upon failures required
Largely online ad hoc queries, requiring low level of indexing	Largely pre-determined queries requiring high level of indexing

System design	OLTP System Online Transaction Processing (Operational System)	OLAP System Online Analytical Processing (Data Warehouse)
Source of data	Operational data; OLTPs are the original source of the data.	Consolidation data; OLAP data comes from the various OLTP Databases
Purpose of data	To control and run fundamental business tasks	To help with planning, problem solving, and decision support
What the data	Reveals a snapshot of ongoing business processes	Multi-dimensional views of various kinds of business activities
Inserts and Updates	Short and fast inserts and updates initiated by end users	Periodic long-running batch jobs refresh the data
Queries	Relatively standardized and simple queries Returning relatively few records	Often complex queries involving aggregations
Processing Speed	Typically very fast	Depends on the amount of data involved; batch data refreshes and complex queries may take many hours; query speed can be improved by creating indexes
Space Requirements	Can be relatively small if historical data is archived	Larger due to the existence of aggregation structures and history data; requires more indexes than OLTP
Database Design	Highly normalized with many tables	Typically de-normalized with fewer tables; use of star and/or snowflake schemas
Backup and Recovery	Backup religiously; operational data is critical to run the business, data loss is likely to entail significant monetary loss and legal liability	Instead of regular backups, some environments may consider simply reloading the OLTP data as a recovery method

Application of data ware housing and Data mining in Government

- Introduction
- National Data Warehouses
  - ○ Census Data
  - ○ Prices of essential commodities
- Other Areas
  - ○ Agriculture
  - ○ Rural Development
  - ○ Health
  - ○ Planning
  - ○ Education
  - ○ Commerce and trade
  - ○ Other sectors
    - ▪ Tourism
    - ▪ Programme implementation
    - ▪ Revenue
    - ▪ Economic affairs
    - ▪ Audit and accounts
- Conclusion

Introduction

- Data warehousing and data mining are the important means of preparing the government to face the challenges of the new millennium.
- These technologies have extensive potential applications in the government: in various Central Government sectors such as Agriculture, Rural Development, Health and Energy.
- These technologies can and should therefore be implemented
- Similarly, in State Government activities also, large opportunities exist for applying these techniques.
- Almost all these opportunities have not yet been exploited.

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

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

General Information Services Terminal of National Informatics Centre (GISTNIC)

- 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 10 years, the data is quasistatic 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.

Prices of essential commodities

- The Ministry of Food and Civil Supplies, Government of India compiles
  - daily data (on weekly basis) for about
  - 300 observation centers 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 using SAS tools

Other Areas

Other possible areas for data warehousing and data mining in Central Government sectors are discussed in detail in the following sections

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
- Data on agricultural inputs such as **seeds and fertilizers** can also be effectively analysed
- Data from **livestock census** can be turned into a data warehouse
- **Land-use pattern** statistics can also be analysed 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
- substantial scope for application of data warehousing and data mining techniques in agricultural sector

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)
- Monitoring and analysis of progress made on implementation of rural development programmes
  - can be effectively utilized by OLAP and data mining technologies

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.

Planning

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

Education

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

Commerce and trade

- Data bank on **trade (imports and exports)** can be analysed 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

Other sectors

Number of other potential application areas for data warehousing and data mining, as follows

Tourism

Tourist arrival behavior 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

Paradigm Shift

- All government departments are deeply involved in generating and processing a large amount of data
- 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 were conventional statistical techniques on largely batch-mode processing
- the advent and prominence of the data warehousing and data mining technology, there is a paradigm shift
- may finally result in improved governance and better planning by better utilization of data
- can rely on data warehousing and data mining technologies for their day-to-day decision-making
- Different data marts for separate departments, can be integrated into one data warehouse for the 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 warehouse is required to be maintained by the ministry concerned for the concerned sector.
- A generic inter-sectoral data warehouse is maintained by a central body (as Planning Commission).
- 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

National Data ware houses and case studies.

- B1) Write a notes on national data warehouse
- B2) Discuss about national data warehouse in detail

CS1: Data Warehousing in the TN Government

- GISTNIC Data Warehouse
- Objectives of the web-enabled data warehouse
- Data Marts of various sectors and their applications

GISTNIC Data Warehouse

- General Information Service Terminal of National Informatics Centre (GISTNIC) Data Warehouse• is an initiative taken by National Informatics Centre (NIC)
- to provide a comprehensive information database by the government on national issues
- ranging across diverse subjects like food and agriculture to trends in the economy and latest updates on science and technology.
- This information base was collated to fulfil the information needs of bureaucrats, politicians, economists and, most important of all, the citizens.
- Web-enabled SAS software solution
- This data warehouse was developed during 1998-99
- aims at providing online information to key decision-makers in the government sector enabling them to make better strategic decisions with regard to administrative policies and investments.
- The Government of Tamil Nadu is the first one to perceive the need and importance of converting data into valuable information for better decision-making

The GISTNIC Web site currently has an online data warehouse which includes data marts on

- Village Amenities
- Rainfall
- Agricultural Census Data
- Essential Commodity Prices
- Malaria Statistics
- Indian Economy Statics
- School Health

Objectives of the web-enabled data warehouse

- To provide powerful decision-making tools in the hands of the end-users in order to facilitate prompt decision-making
- To reduce the amount of resources-time and manpower-spent on managing the volumes and variety of database handled by NIC.

Data Marts of various sectors and their applications

Village Amenities

- This data mart contains the 1991 census data of village amenities in all the villages in Tamil Nadu.
- It contains information on availability for amenities like education, health, drinking water, transportation, communication and irrigation

Applications

- Village amenities analysis
- Irrigation analysis
- Top/bottom analysis
- Range analysis- amenities
- More amenities-analysis

SAS/multidimensional report on the village amenities data.

This report is used to analyse villages in Tamil Nadu based on availability of various amenities


Rainfall

- This data mart has information on daily levels of rainfall across various weather stations in Tamil Nadu.
- This will help them to plan the water supply to various districts in Tamil Nadu and using various models to forecast rainfall levels

Applications

- Time-based rainfall analysis
- Geography/time-based rainfall analysis

A line plot screen for tracking rainfall levels at various weather stations in Tamil Nadu.

- On selecting one of the weather stations in the list box on the screen, the line plot changes to reflect the rainfall level for the selected weather station.
- On clicking any point on the line plot the graph displays the data and rainfall level for the data point
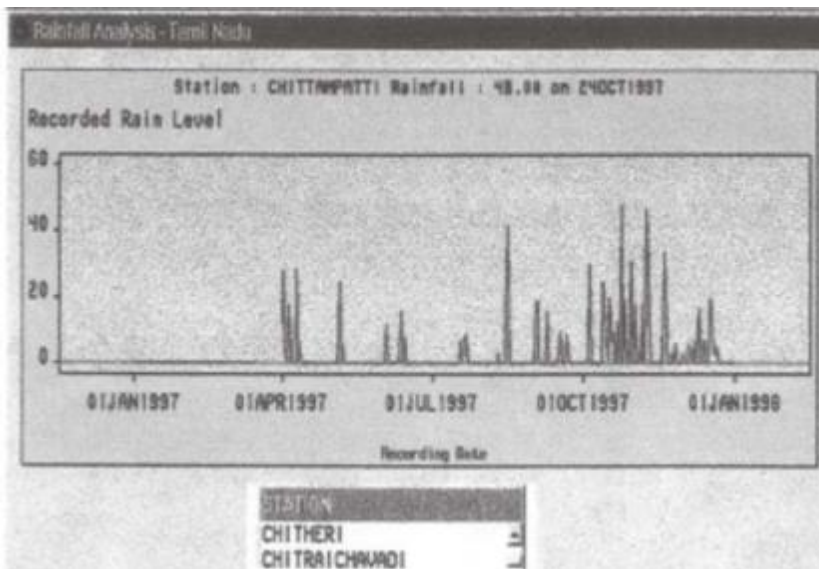
Agricultural Census Data

- This data mart has information on land-holding patterns across the villages in Tamil Nadu.
- It can be used to analyse information about land-holding amongst individuals, institutions, males, females, scheduled castes and scheduled tribes, etc

Applications

- Land-holding analysis
- Land-holding analysis - multi dimensional
- Top/bottom analysis
- Medium-holding analysis

A SAS/EIS multidimensional report on the agricultural census data

The report is displaying area under holding/number of holdings based on size, type of holder, sex of the holder, name of the district, etc

District Name			COIMBATORE		CUDDALORE		DHARMAPURI	
			Area - Total	Holdings - Total	Area - Total	Holdings - Total	Area - Total	H:
Size Class Type	Type Of Holder	Land Holder	SUM	SUM	SUM	SUM	SUM	St
LARGE	Individual	Male	36,297	2,539	9,132	715	11,408	78
		Female	4,558	306	388	27	806	68
	Institution	Institution	22,896	190	.	.	357	11
MARGINAL	Individual	Male	45,243	84,091	68,363	180,249	113,011	29
		Female	9,369	17,285	20,630	36,498	19,836	48
	Institution	Institution	93	171	.	.	58	11

Essential Commodity Prices

- To provide updated information on various essential commodity prices, NIC collects the retail/wholesale prices of various essential commodities like vegetables, sugar, rice, oil, cereals, etc.
- Using the GISTNIC data warehouse, end-users now have the updated information about trends in price change and will thus be able to closely monitor the prices more effectively

Applications

- Commodity price analysis
- Time-wise commodity
- Qtr3-4 analysis
- Rice analysis
- Forecast- rice prices

Malaria Statistics

- This data mart has information on various health camps conducted across Tamil Nadu to detect and cure malaria patients.
- This has vital information like number of people suffering from malaria, deaths caused due to malaria, source of malaria infection, demographic information of malaria patients, etc.
- Using the data warehouse, the end-users will be able to plan various precautionary measures to reduce the number of people suffering from malaria in Tamil Nadu

Applications

- MDR census on samples collected and tested
- MDR source of malarial parasites
- MDR age- and sex-wise malarial census
- Graph- and sex-wise malarial census

Indian Economy Statics

- This data mart has information about statistics on the telecom sector, stock exchange (NSE and BSE) and India's foreign trade.
- This data is collected on monthly basis from CMIE, Mumbai

Applications

- Capital market analysis
- Capital market analysis-MODS report
- Basic telecom analysis- overview
- Basic telecom analysis--state-wise
- External trade analysis (1997- 1998), Combine report

School Health

- has information about various health check-up camps conducted in various schools across Tamil Nadu.
- It has information about students suffering from various diseases, defects, immunization programmes, etc

Applications

- Disease analysis- MODS report, Disease analysis-graph, Immunization analysis- MODS
- Immunization analysis- graph

Other Case Studies (Skip)

- Data Warehouse for the Ministry of Commerce
- Data Warehouse for the Government of Andhra Pradesh: A Data Warehouse for Finance Department
- Data Warehousing in Hewlett-Packard
- Data Warehousing in Levi Strauss
- Data Warehousing in the World Bank
- HARBOR (Highly Available Replication-Based Online Recovery) Data Warehouse
- A Typical Business Data Warehouse for a Trading Company
- Customer Data Warehouse of the World's First and Largest Online Bank in the United Kingdom
- A German Supermarket EDEKA's Data Warehouse

Disclaimer

- Intended for educational purposes only. Not intended for any sort of commercial use
- Purely created to help students with limited preparation time. No guarantee on the adequacy or accuracy
- Text and picture used were taken from the reference items

Reference

- DATA WAREHOUSING: Concepts, Techniques, Products and Applications 3rd Ed, CSR Prabhu
- Data Mining: Concepts and Techniques 2nd Ed, Jiawei Han, Micheline Kamber

Credits

- Thanks to my family members who supported me, while I spent considerable amount of time to prepare these notes. Feedback is always welcome at GHCRajan@gmail.com