

Data Warehousing for Tamil Nadu Government

GISTNIC DATA WAREHOUSE

- The 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.
- The GISTNIC data warehouse is a web-enabled SAS software solution.
- The data warehouse aims to provide 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 a valuable information for better decision making.
- The GISTNIC web site has online data warehouse which includes data marts on village amenities, rainfall, agricultural census data, essential commodity prices, malaria statistics etc.

Continue..

- Indian economy statistics and school health dataware house was developed during 1998-99.
- It is used to provide easy access with point and click interfaces to key decision makers across various level in the government.
- It endorses the beginning of a strategic business alliance between SAS Institute and the government sectors.
- Tamil Nadu Government will be India's first state Government which will ride on the technology wave to streamline processes and drive better decision making on the basis of communication of information in whatever form it may take, unconstrained by distance, time and volume.
- It indicates the commitment of SAS Institute in working together with the government to enable them to move towards being IT-savvy.

OBJECTIVES OF WEB ENABLED DATA WAREHOUSE

- To provide a 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 Analysis

VARIES

1. **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.

Application

- Village & range amenities analysis
- Irrigation & top/bottom analysis

2. **Rainfall statistics:** This data mart has information on daily levels of rainfall across various weather stations in Tamil Nadu. This will help them plan the water supply to various districts in Tamil Nadu and use various models to forecast rainfall levels.

Application:

- Time & Geography based rainfall analysis

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

Application

- Land-holding analysis-multidimensional
- Top/bottom & Medium-holding analysis

4.Essential Commodities: 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.

Application

- Time-wise commodity & commodity price analysis
- Forest-rice prices analysis

6.Indian economy: This data mart has information about statistics on the telecom sector, stock statistics on the telecom sector, stock exchange (NSE & BSE) and India foreign trade. This data is collected on monthly basis from CMIE, Mumbai.

Application

- Capital market analysis-MDDB report
- Basic telecom analysis-overview & stepwise
- External trade analysis(1997-1998)
- Combine report

7. School Health: This data mart 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.

Application

- Disease analysis- MDDB report & graph
- Immunization analysis-MDDB & graph

DATA WAREHOUSING FOR THE MINISTRY OF COMMERCE

The ministry of commerce has set up the following seven export processing zones at various locations.

- 1.Kandla free trade zone gandhidham
2. Santacruz electronics export processing zone bombay
3. Falta export processing zone , west bengal
4. Madras export processing zone, chennai
- 5.Cochin export processing zone ,cochin
6. Noida export processing zone noida
7. Vishakapatnam export processing zone visakhapatnam

This case study report presents how the data warehouse can be effectively built for the ministry of commerce.

✓The following are the objectives

1. Globalization of india's foreign trade
2. Attracting foregn investment
3. Scaling down tariff barriers.
4. Encouraging high and internationally acceptable standards of quality
5. Simplification and streamlining of procedures governing imports and exports.

DATA WAREHOUSE:

Data warehouse is subject oriented, integrated, time variant and non volatile collection of data in support of management decision making process

✓ OBJECTIVES OF DATA WAREHOUSE FOR THE MINISTRY OF COMMERCE

The ministry of commerce has been regularly reviewing the data warehouse in its board meetings. The ministry is equipped with all analysis variables a reporting form and the zone performance and the progress of exports are reviewed in each zone of the country.

The data warehouse includes all analysis variables into consideration from all the zones with an option to drill down to the daily data.

DATA WAREHOUSE IMPLEMENTATION STRATEGY FOR EXPORT PROCESSING ZONES

✓ INFRASTRUCTURE

The basic infrastructure required for building the warehouse for the ministry of commerce is based on the communication infrastructure, hardware/software/tools and manpower.

The common infrastructure

This includes all the tasks necessary to provide the technical basis for the warehouse environment. This includes the connectivity between the legacy environment and the new warehouse environment on a network as well as on database level.

Man power requirements.

The senior officials in the ministry of commerce sponsored the whole warehouse implementation and played active role as EXIM policy and business architect for data warehouse and also subject area specialists.

KEY AREAS FOR ANALYSIS

The data topic deals with the data access mapping, derivation , transformation and aggregation according to the requirement of ministry. The key areas are decision making , which will be used for analytical processing and for data mining at the ministry of commerce are listed as follows

1. Unitwise, sectorwise and countrywise imports and exports
2. Direction of imports and exports
3. Sectorwise , countrywise and zone wise trends imports and exports
4. Comparative country wise export and import for each sector
5. DTA sales
6. Claims of reimbursement of central sales tax of zones
7. Deemed export benefits
8. Employment generation
9. Investments in the zone
10. Deployment of infrastructure
11. Growth of industrial units
12. Occupancy details

IMPLEMENTATION OF DATA WAREHOUSE

Operational data systems

The operational data system keeps the organization running by supporting daily transactions such as import and export bills submitted to customs department in each zone, daily transactions of permissions allotments etc.

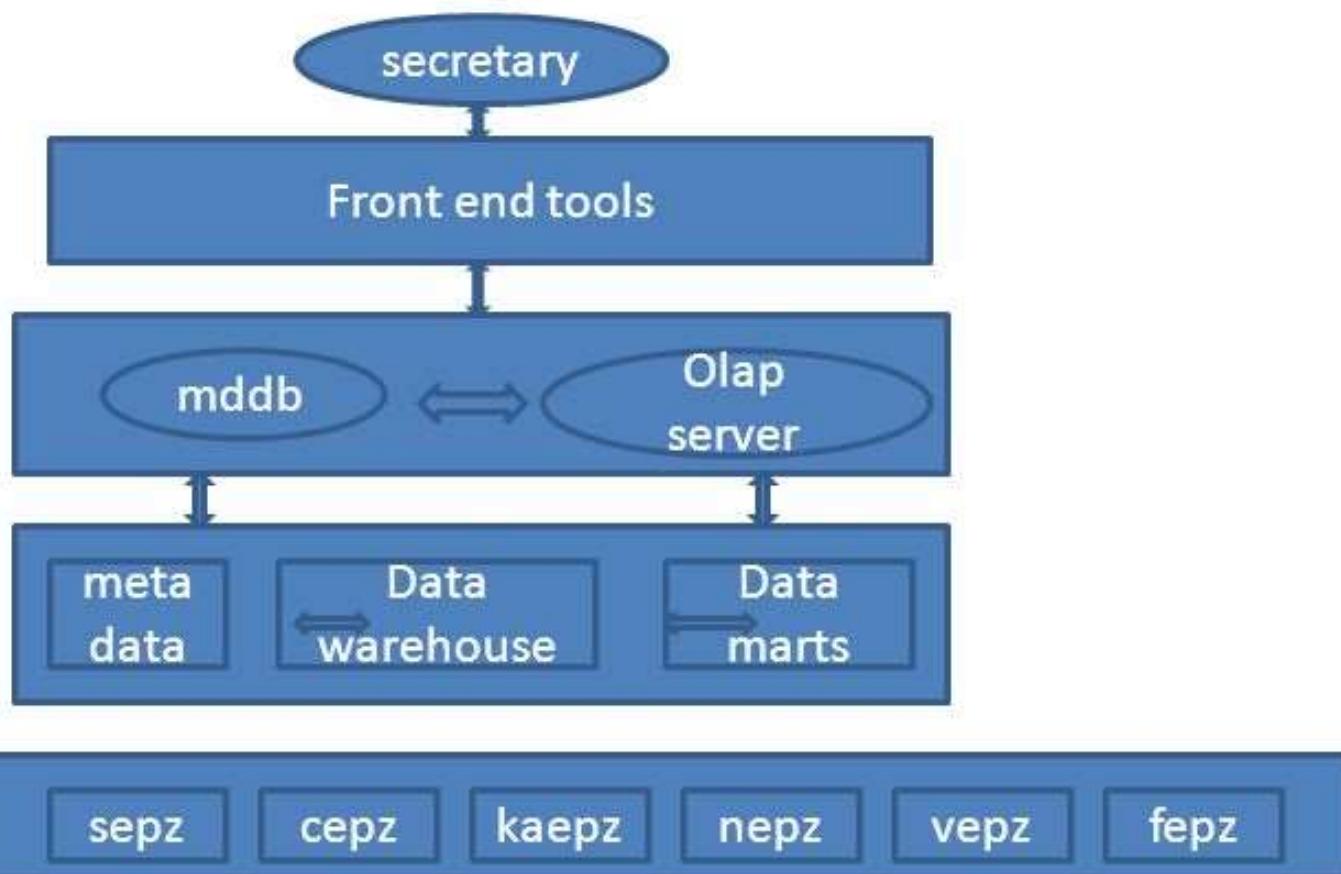
ARCHITECTURE

The architecture of the OLAP implementation consists of 5 layers.

All the 7 zones have DBMS/RDBMS data for internal management of zone activities and been forwarding the MIS reports to the MOC , new delhi.

The second layer is located at the MoC new delhi with large metadata repository and data warehouse

The metadata warehouse tools and OLAP server handling and maintaining same are focussed in level-3 of the architecture.



DESIGN OF ANALYSIS/CATEGORY VARIABLES

The data model is prepared by the entire data availability and data requirement are analyzed and the analysis variables for building the multidimensional cube are listed as follows

1. Employment generation managerial/skilled/unskilled classification with zone/unit/industry break-up
2. Investments in the zone
3. Performance of units and close monitoring during production
4. Deployment of infrastructure etc.

Related tables:

Ep_mast, ep_stmst ,dist_mast ,states , indu_mast; eo_mast, eo_stmst.

Analysis variables:

1. EPZ or EOU
2. Zone
3. Type of approval
4. Type of industry
5. State
6. District
7. Year of approval
8. Month of approval
9. Day of approval
10. Year of production commencement

11. Month of production commencement
12. Day of production commencement
13. Current status
14. Date of current status
15. Net foreign exchange percentage
16. Stipulated NFE
17. Number of approvals
18. Number of units

exports and imports values with zone sector port year month and day break-ups deal with the following performances

Zone-wise performance

Industry-wise performance

Country-wise performance

This will indicate the following direction of exports

Country-wise performance overall and with zone break-up

Port-wise performance which will decide the examination of infrastructure in these ports

Export performance during different time periods and the analysis of the same.

Trend over years/quarters/months for different country, sector, and zone

Comparative country-wise import/export for each sector

Related tables:

Shipbill, country, industry, currency, shipment_mode , export , bill entry

Export analysis variables:

Zone , export type, year of shipping bill, month of shipping bill, country, currency, mode of shipment, destination port, value of export etc.

Import analysis variables:

Auto/approval, zone, import type, import year, import month, import day, type of goods, import value, duty foregone, import country and mode of shipment

DEEMED EXPORT BENEFITS

Analysis variables:

Zone , sector, claims received amount disbursed year/quarter/month

EMPLOYMENT GENERATION

Analysis variables:

Zone , male/female, managerial/skilled/unskilled, number of employees

INVESTMENTS IN THE ZONE

Analysis variables:

Zone, unit, NRI foreign investment, Indian investment, remittances received, approved value.

CONCLUSION

The data warehouse for EPZ provides the ability to scale large volume of data seamless presentation of historical , projected and derived data.

It helps the planners in what-if analysis and planning without depending on the zones to supply the dat.

The time lag between the zone and the ministry is saved and then the analysis can be carried out with the speed of thought analysis.

The data warehouse for the ministry of commerce can add more dimensions to the propose warehouse with the data collected from other offices to evolve a data warehouse model for better analysis of promotion of imports and exports in the country.

This will provide an excellent executive information system to the secretary joint secretaries of the ministry.

DATA WAREHOUSING FOR THE GOVERNMENT OF ANDHRA PRADESH

DATA WAREHOUSE FOR FINANCE DEPARTMENT

Responsibilities of the finance department:

The finance department of the government of andhra pradesh has the following responsibilities.

1. Preparing a department-wise budget up to the sub-detail head and submission to the legislature for its approval.
2. Watching out the government expenditure and revenue department-wise
3. Looking after development activities under various plan schemes
4. Monitoring other administrative matters related to all heads of the department.

Treasuries in andhra pradesh:

Money standing in the government account are kept either in treasuries or in banks. Money deposited in the banks shall be considered as general fund held in the books of the banks on behalf of the state.

Director of treasuries:

Treasuries are under the general control of the director of treasuries and accounts.

Sub treasuries:

If the requirements of the public business make necessary the establishment of one or more sub-treasures under a district treasury.

The accounts of receipts and payments at a sub-treasury must be included monthly in the accounts of the district treasury.

Treasuries handle all the government receipts and payments.

Every transaction in the government is made through related departments.

DESCRIPTION OF THE ACCOUNT HEAD	LENGTH OF THE CODE
MAJOR	4
SUB-MAJOR	2
MINOR	3
GROUP SUB-HEAD	1
SUB-HEAD	2
DETAIL HEAD	3
SUB-DETAIL HEAD	3

LESS THAN 2000	RECEIPTS
2000<4000	SERVICE MAJOR HEADS
4000<6000	CAPITAL OUTLAY
6000<8000	LOANS
MORE THAN 8000	DEPOSITS

A project for building data warehouse for OLAP is implemented by NIC for the department of treasuries. The concept of building DW in the department of treasuries has been established for providing easy access to integrated up to date data related to various aspects of the department functions.

DW technology is used to develop analytical tools designed to provide support for decision making at all levels of the department.

Traditional information systems implemented in the department of treasuries are based on transactional databases

Which are not designed for providing fast and efficient access to information critical for decision making.

Data required for analysis are typically distributed among a number of isolated information systems meeting the needs of different sub-treasuries.

Data warehouse technology provided to the department of treasuries by national informatics centre eliminates these problems by storing current historical data from disparate information systems.

Data warehouse provide efficient analysis and monitoring of financial data of treasuries.

It also evaluates the internal and external business factors related to operational economic and financial conditions of treasuries budget utilization.

DIMENSIONS COVERED UNDER FINANCE DATA WAREHOUSE

The different dimensions taken for drill down approach against two measures (payments and receipts)

1. Department
2. District treasury office
3. Sub-treasury office
4. Drawing and disbursing officer
5. Time
6. Bank-wise
7. Based on different forms

COGNOS GRAPHIC USER INTERFACE FOR TREASURIES DATA WAREHOUSE

Impromptu:

It is used for generating various kind of reports like simple crosstab etc.

Transformer:

Transformer model objects contain definition of queries dimensions measures dimension views user classes and related authentication information as well as objects for one or more cubes that transformer creates for viewing and reporting in powerplay.

Powerplay:

COGNOS powerplay is used to populate reports with drill-down facility.

Scheduler:

Scheduler coordinates the execution of automated processes called task , on a set date and time, or at recurring intervals.

Authenticator:

Authenticator is a user class management system. It provides COGNOS client applications with the ability to create and show data based on user authenticated access.

DATA WAREHOUSING IN HEWLETT-PACKARD

HP can easily access and quickly analyze enormous volumes of self through data to help its reseller customers improve the efficiency and profitability of their businesses with an OLAP system based on Microsoft SQL server and knosys proclarity.

Hewlett-Packard is a worldwide market leader in the \$18 million inkjet industry. In the past, HP's brand recognition and reputation for reliability were enough to ensure that a reseller would carry HP products.

ACCESS TO INFORMATION NEEDED USING DATA WAREHOUSING TECHNOLOGY

HP has captured and stored the information both from primary research and third parties. The businesss analysis group decided they needed a system that would provide market metric data to help field sales force managers or account teams make brand and channel management decisions.

HP requires a system that requires low cost, low maintenance and as simple to administer as possible. So the group turned too Knosys, a Boise, Idaho based software company that has developed a business analysis/online analytical processing(OLAP) package called ProClarity.

The solution enables to move the data so quickly and at such a low cost of maintenance and ownership that it would solve the problems.

Knosys has helped the group build the data flow algorithms with microsoft visual basic development system and SQL server 7.0 data transformation services.

Due to HP's enormous sell through data volumes, it would take too long to build analytical models with a pure, multidimensional OLAP solution.

HP used SQL server 7.0 virtual cubes and cube partitioning capabilities.

Virtual cube capabilities allows decision makers to cross analyze data from all these OLAP sources.

Cube partitioning allows HP to more effectively manage large number of OLAP cubes .

Knosys Pro clarity provides HP decision makers with the key to analyzing masses of data.

Pro clarity is fully integrated with Microsoft products, and its PC-based client is modelled after Internet explorer 4.0

Pro clarity powerful analytical features takes full advantage of the robust capabilities found in SQL server 7.0 OLAP services

CONCLUSION

HP's pro clarity system and SQL server 7.0 provides the system with more accurate, detailed and timely data, that makes the business more efficient and helps its resellers make their businesses more efficient.

DATA WAREHOUSING IN LEVI STRAUSS

In 1998, ArsDigita corporation built a web service as a front-end to an experimental custom clothing factory operated by Levi strauss.

The whole purpose of the factory and web service was to test and analyze the consumer reaction to this method of buying clothes. Therefore , a data warehouse was built into their project almost from the beginning. The public website was supported by a mid-range HP Unix server that had ample leftover capacity to run the data warehouse.

The new 'dw' oracle user was created to support the data warehouse.

GRANTed SELECT on the OLTP tables to the 'dw' user and wrote procedures to copy all the data from the OLTP system into a star schema of tables owned by the 'dw' user

This kind of schema was proven to scale to the world's largest data warehouses.

In star join schema, there is one fact table that references a bunch of dimension tables.

The following dimension tables are designed after discussing with Levi's executives

- 1.TIME:** for queries comparing sales by season, quarter or holiday
- 2.PRODUCT:** for queries comparing sales by color or style
- 3.SHIP TO:** for queries comparing sales by region or state.
- 4.PROMOTION:** for queries aimed at determining the relationship between discounts and sales.
- 5.USER EXPERIENCE:** for queries looked at returned versus exchanged versus accepted items

DATA WAREHOUSING IN WORLD BANK

The world bank collects and maintains huge data of economic and developmental parameters for all the third world countries across the globe.

The bank performed analysis on this huge data manually and later with limited tools for analysis.

The bank collects and analyzes macro economic financial statistics and also information on parameters such as poverty, health, education, environment and public sector.

THE LIVE DATABASE DATA WAREHOUSE

The OLAP cubes were defined for this database using OLAP server module of SQL server 2000.

Universal access was provided for this data warehouse was called **live database**

BENEFITS OF THE SECOND-GENERATION LDB DATA WAREHOUSE

the first generation LDB data warehouse had certain limitations. Therefore the second generation ODB datawarehouse was built using SQL server 2000 analysis server along with proclarity of knosys corporation,

The package offered direct user functionality which was otherwise requires technical intervention by a programmer.

Proclarity also provides web enablement, thereby ensuring universal accessibility.

It results in significant cost savings by reducing the time and effort required to prepare a large variety of reports to suit varying needs of the economists and other governmental decision makers to aid effective and better economic planning.