



Data Warehouse

Ministry of

Commerce



Team Members:

Gaurang Ashava - RA2111027010007

Aiswaryaa Rajesh - RA2111027010012

Ritik Ranjan Sinha - RA2111027010013

Arunima Mishra - RA2111027010017



Index

S.No.	Topic	Page Num
1.	Introduction	4
2.	What is Datawarehouse ?	6
3.	Datawarehouse Architecture	10
4.	Objectives of Datawarehouse for MoC	12
5.	DESIGN OF ANALYSIS	14
6.	Implementation & Strategy	19
6.	ER Diagram (Schema)	23
7.	Executive Information System	28
8.	Benefits	36
9.	Conclusion	37

Introduction

- After the liberalization, foreign companies have shown interest in investing in India
- The investment boom has been injecting a high tide of competitiveness in Indian industries.
- In the wake of liberalization, many export-friendly policies have been drastically liberalised and the procedural bottlenecks have been removed a great deal over the years
- Various export promotions schemes like Export Oriented Unit (EOU) and Export Processing Zone (EPZ) schemes occupy an important place in the country export effort

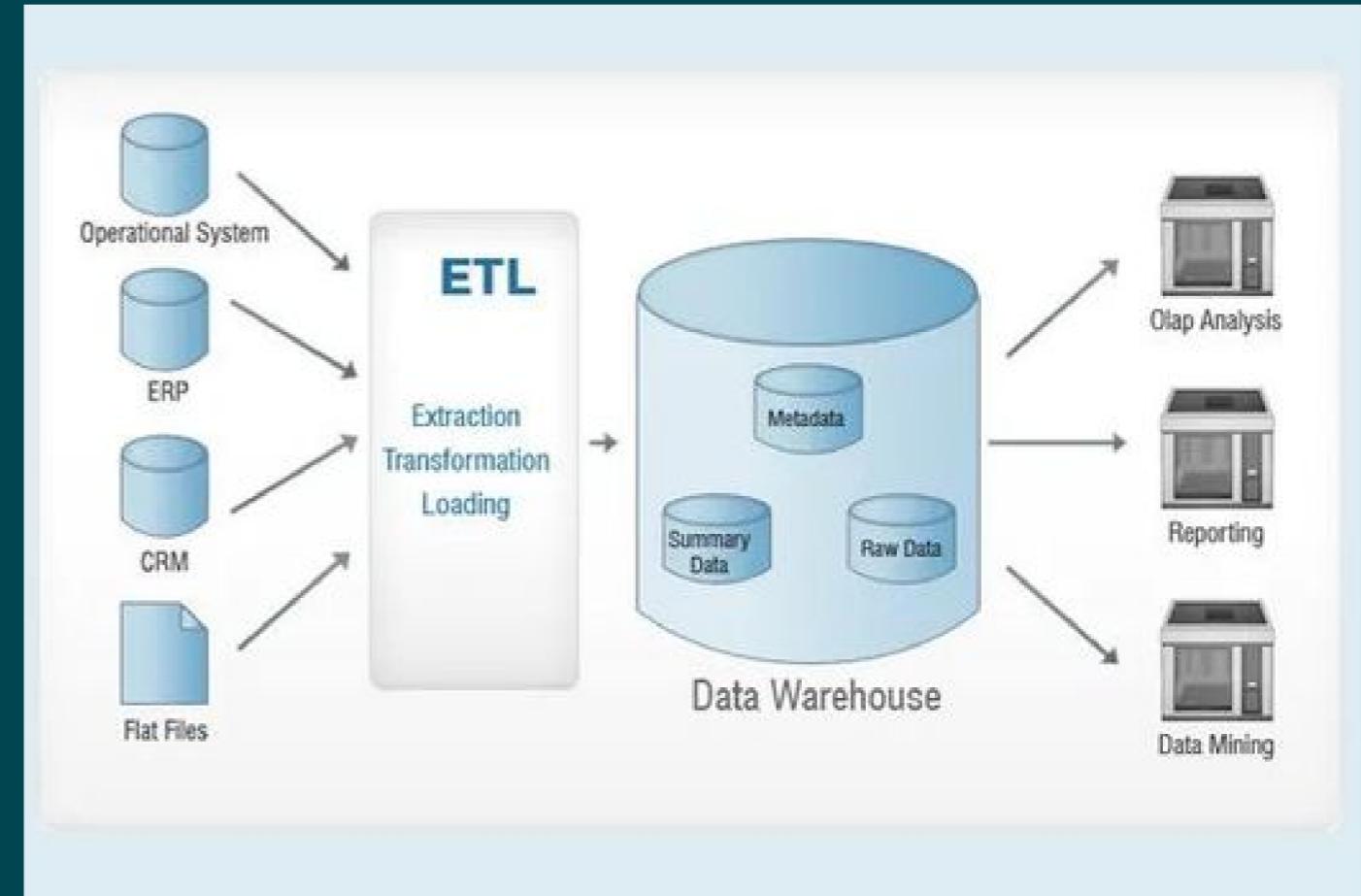
The EPZs are set up to provide an internationally competitive duty free environment for export production at low cost. They are responsible for administrating free trade zones in setting up 100% export-oriented multi-product industrial units

In regards to infrastructure, Data management and OLAP decision support tools are used to achieve :

- Globalization of Indian Foreign Trade
- Attracting foreign Investment
- Scaling down tariff barriers
- Encouraging high and internationally acceptable standards of quality
- Simplification and streamlining of procedures governing imports and exports

What is Data Warehouse

- Data warehouse is subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision-making progress.
- Data Warehouse focusses on data modelling and database design exclusively.



- Operational data in warehouse maintains an ongoing relationship between two or more tables based on policy that is in effect.
- Two types of operations can be performed ie loading of data and access of data .
- various rules are represented in data warehouse between two or more tables.
- Data in warehouse is integrated .
- Integration is stated as - in consistent naming conventions, in consistent measurement of variables, in consistent encoding structure, in consistent physical attributes of data.

- **Time-stamped Data:** In data warehouses, data is often time-stamped to indicate when it was collected or updated. This time-stamp allows users to track changes and analyze data at specific points in time.
- **Time-based Queries:** Data warehouses support time-based queries that enable users to retrieve and analyze data within specific time ranges.
- A data warehouse is suited for ad hoc analysis as well custom reporting

What is a data warehouse used for?

Making real-time decisions:

Analyze data in real time to proactively address challenges, identify opportunities, gain efficiency, reduce costs, and proactively respond to business events.

Enabling business reporting and ad hoc analysis:

Keep historical data on a separate server from operational data so that end users can access it and run their own queries and reports without impacting the performance of operational systems or waiting to get help from IT.

Implementing machine learning and AI:

Collect historical and real-time data to develop algorithms that can provide predictive insights, such as anticipating traffic spikes or suggesting relevant products to a customer browsing a website

Implementation Of Data Warehouse

A centralized database storing daily transactional data and management information from seven zones, facilitating internal management and reporting activities.

The architecture features three layers: zone-level databases, a central data warehouse with metadata at the Ministry of Commerce, and analytical tools.

Operational data systems

Architecture

Some analytical queries that are addressed directly by the ODS include:

- Determining the export volume of textile products specifically for the year 1997.
- Assessing the impact on export patterns resulting from adjustments to the overall transaction limit, particularly when capital equipment (including computer hardware for electronic units) is raised to 70%.
- Investigating the industry-wise export patterns within each zone.

- Understanding the sudden surge in exports of "computer software" within specific zones (MEPZ and SEEPZ) during March to June 1998, while other zones exhibit no such activity.
- Identifying changes in EXIM policies that could potentially boost exports by at least 100L in the upcoming financial year.



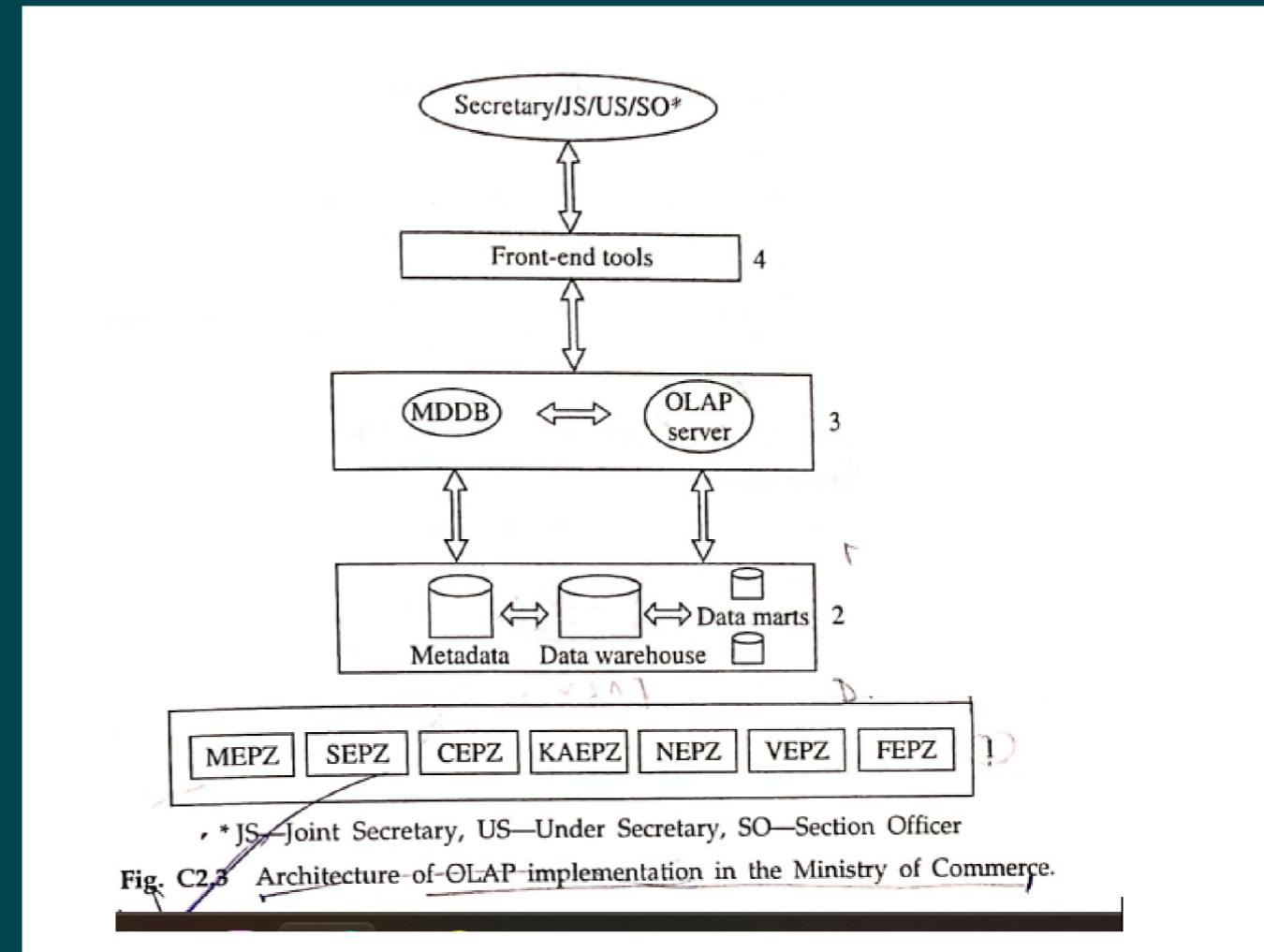
While the first three queries can be handled using SQL and reporting tools due to their straightforward nature, the last two queries require more intricate data processing methods. Analytical processing is essential for addressing these complex queries effectively.

Architecture

The OLAP implementation architecture consists of :

- Zone-Level DBMS/RDBMS Data
- MoC-Level Repository and Warehouse
- Metadata Warehouse Tools and OLAP Server

Metadata
and
Data



DESIGN OF ANALYSIS

- The data model creation process involves:
 - a.Analyzing available data and requirements.
 - b.Identifying analysis variables for multidimensional cubes.
 - c.Gathering information on approvals, unit status, production, and specifics such as employment generation and infrastructure deployment from relevant tables.

- Analysis variables encompass:
- EPZ/EOU status
- Zone
- Approval type
- Industry
- Location
- Production dates

Statistical functions generate additional variables, including:

- Number of approvals
- Number of units

- For export and import performance analysis, data is disaggregated by zone, sector, country, port, and time periods, enabling comprehensive assessments and strategic decision-making.
- Overall, the data model aims to provide insights into export/import performances, industry trends, and related variables to support strategic decision-making.

DESIGN OF ANALYSIS



DTA sales

Investments in the zone

**Claims of reimbursement of central
sales tax**

Deemed export benefits

Employment generation

OBJECTIVES OF DATA WAREHOUSE FOR THE MINISTRY OF COMMERCE

- 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, and so of the whole country .But any major setback in exports and defunct units, employees' problems etc, the analysis becomes difficult because of less analytical processing.
- The Data Warehouse includes all analysis variables into consideration from all the zones with an option to drill down to the daily data.

OBJECTIVES OF DATA WAREHOUSE FOR THE MINISTRY OF COMMERCE

- The aggregation by week, month, quarter ,year are possible in association with all other analysis variables. Implementing this takes planner to the problem area with few clicks.
- Data mining will help them further to discover many facts, which will help the planners in the way of revising the EXIM policy in favour of the country's progress.

Critical Decision Making

- Importance of data warehousing in facilitating informed decision-making.
- Illustrative examples demonstrating how real-time insights aid in critical decision-making processes.
- Highlighting the significance of accurate, reliable data for policy formulation and trade strategies.

Enhanced Forecasting

- Importance of data warehousing in improving forecasting accuracy.
- Utilizing predictive analytics for market trend analysis and forecasting.
- Enabling proactive decision-making based on predictive insights.
- Highlighting the impact of enhanced forecasting on the Ministry of Commerce's strategic planning and resource allocation.

Improved Data Governance

- Importance of data governance in ensuring data integrity, compliance, and security.
- Utilizing data warehousing for structured data governance practices.
- Ensuring compliance with regulations and policies to maintain trust and credibility.
- Highlighting the role of data security measures in safeguarding sensitive information and mitigating risks.

Implementation Strategy for Export Processing Zones

Infrastructure

Basic infrastructure is based on communication, hardware/software tools and manpower

Communication Infrastructure

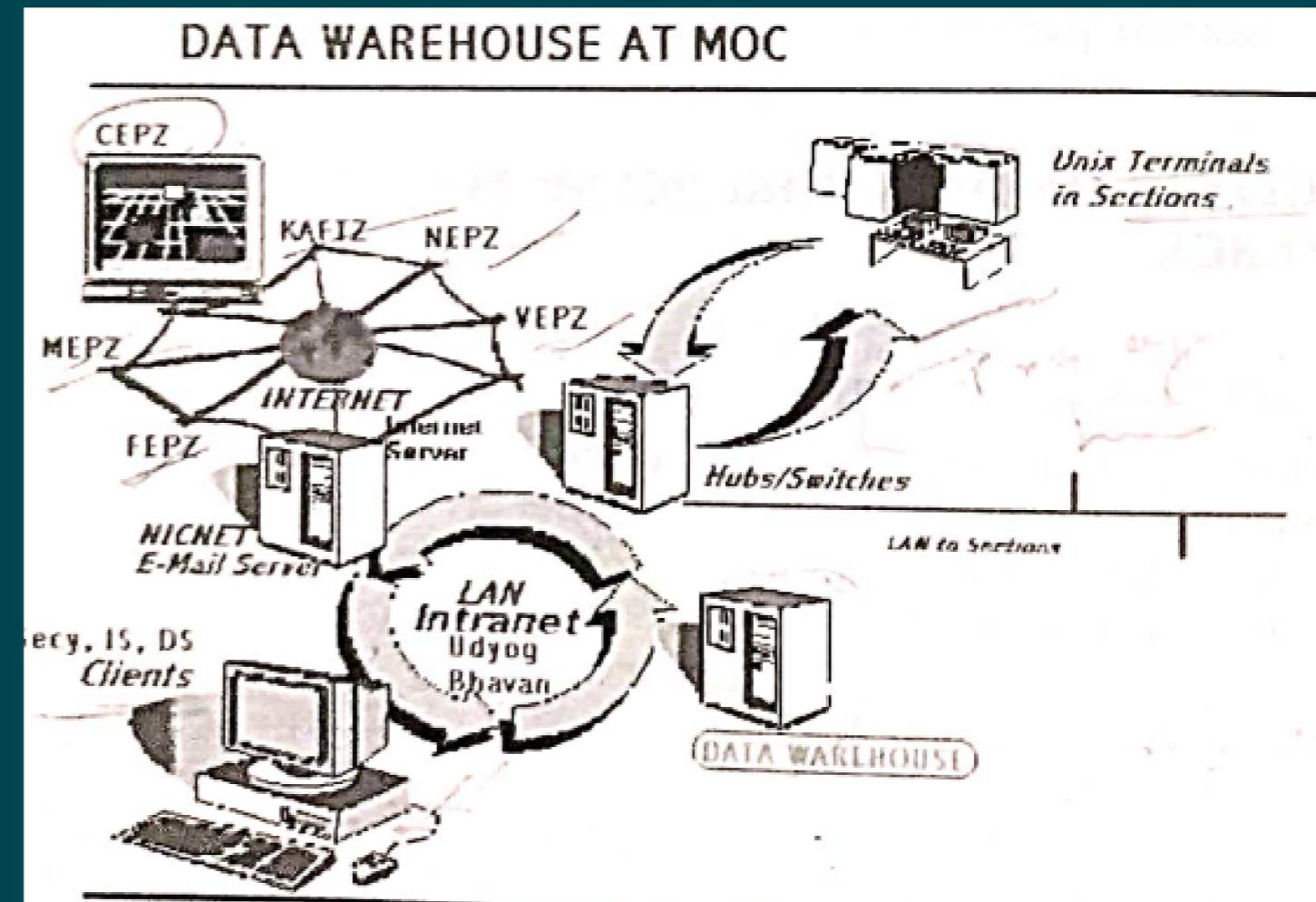
Connectivity between the legacy environment and the new warehouse environment on a network as well as on a database level and the implementation of a population subsystem, an administrative subsystem and a management subsystem



In Ministry of Commerce Delhi:

All export processing zones push data through VSAT into Datawarehouse service
This server is a part of Internet in Udyog Bhavan under the MOC

VSAT - A very small aperture terminal is a small-sized earth station used in the transmit/receive of data, voice and video signals over a satellite communication network, excluding broadcast television.



The Hardware and Software Tools

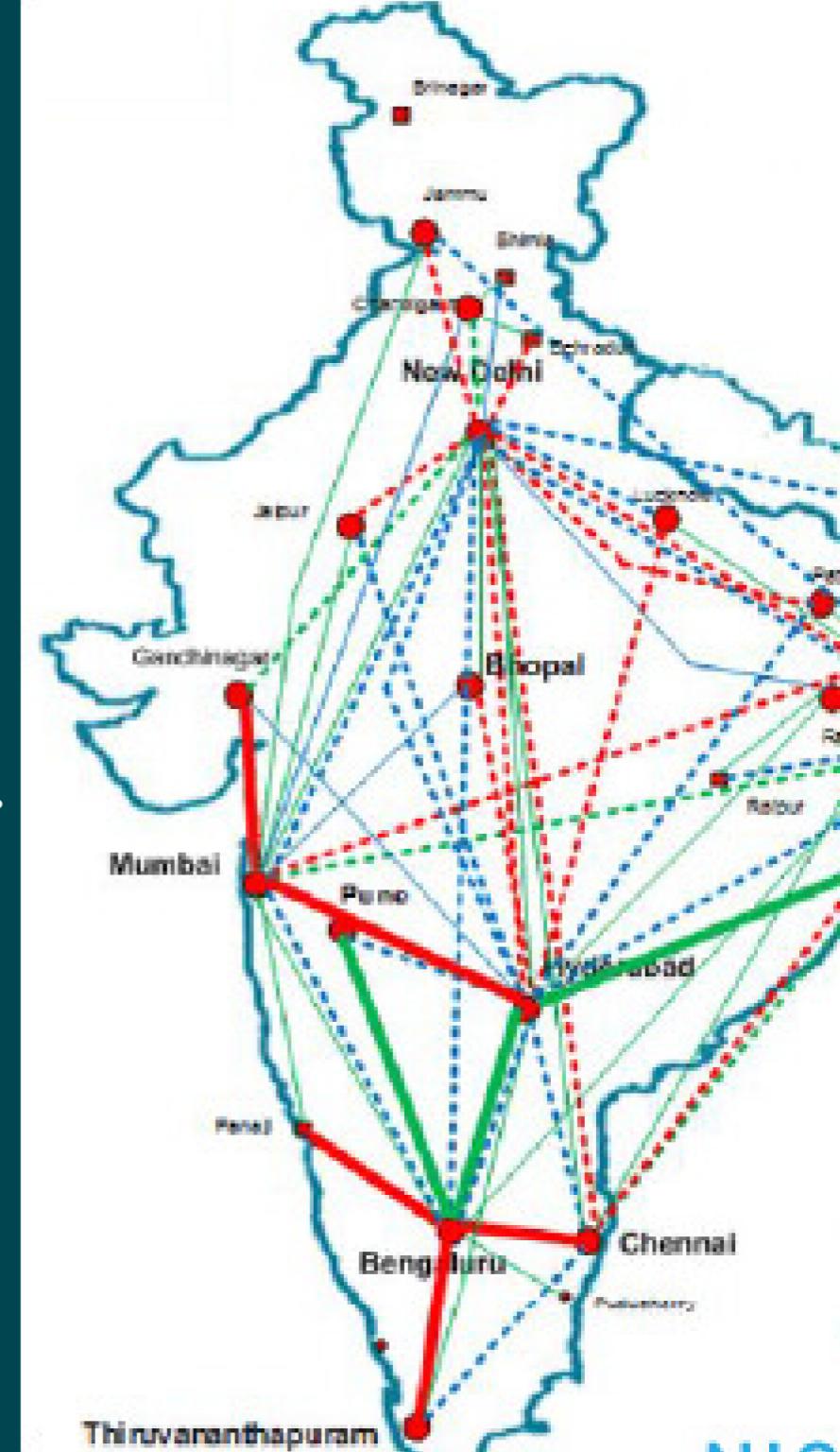
Hardware:

Necessary hardware for the warehouse server with backup server is placed in the existing WAN of NICNET. This connects all the existing zones and also clients in the MOC.

The MOC will be depending on the cost, performance and need and the tools for building the Datawarehouse.

NICNET:

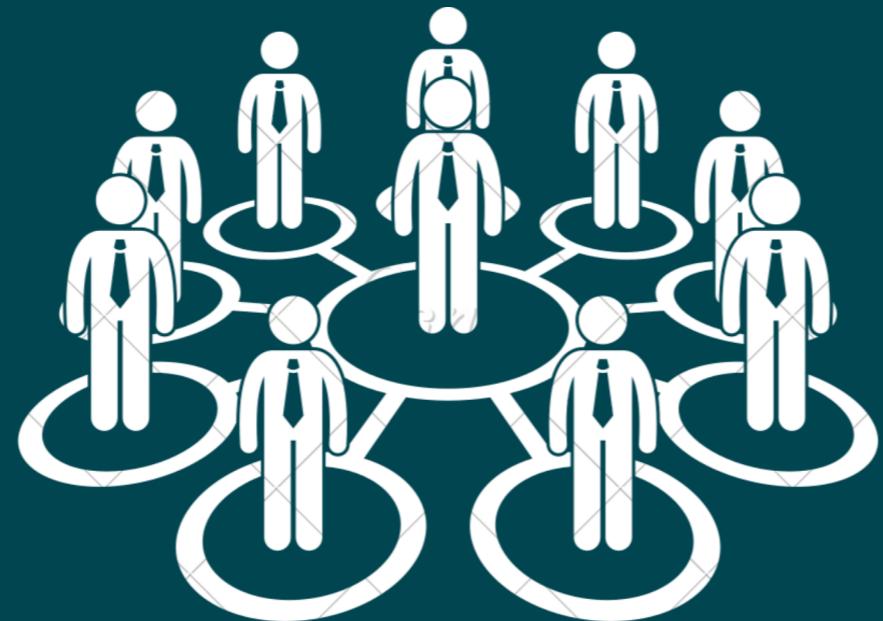
National Informatics Centre Network connecting all Districts and State Capitals with New Delhi. Pan India Communication Network for Government



Manpower Requirement

Senior officials sponsored the Datawarehouse implementation and played active role as EXIM policy, business architects and subject-area specialists.

NICNET provided the technical know-how for managers, administrators, platform specialists and tools specialists



EXIM Policy:

Export-Import Policy contains guidelines governing the imports and exports of products and services in and out of India

ER Diagram (Schema)

Sales			
sale_id (PK)			
date_key (FK)			
location_key (FK)			
product_key (FK)			
customer_key (FK)			
supplier_key (FK)			
quantity_sold			
unit_price			
total_price			
payment_method			
discount			

Time	Location	Product	Customer
date_key (FK)	location_key (PK)	product_key (PK)	customer_key (FK)
date	country	product_id	customer_id
day	region	product_name	customer_name
month	city	category	email
quarter	address	subcategory	phone
year		brand	address

Supplier	
supplier_key (PK)	
supplier_id	
supplier_name	
contact_person	
email	
phone	

The schema used for this diagram is a star schema as the fact table is centrally located and is surrounded by dimension tables.

The fact table stores quantitative information for analysis.

1. Sales
2. Supplier

A dimension table is a table or entity in a schema that stores details about the facts:

1. Time
2. Location
3. Product
4. Customer

Executive Information System



Data Analysis- SAS

Data Imported to SAS EIS Application and multi-dimensional analysis are carried out.

Analysis by Drilling Down

Helps us analyse the exports by rotating the cube, we can slice and dice the category decision variables

Multi Variance Analysis

Filters data for specific data requirement and the subset can be used for analysis.

Traffic Light Analysis

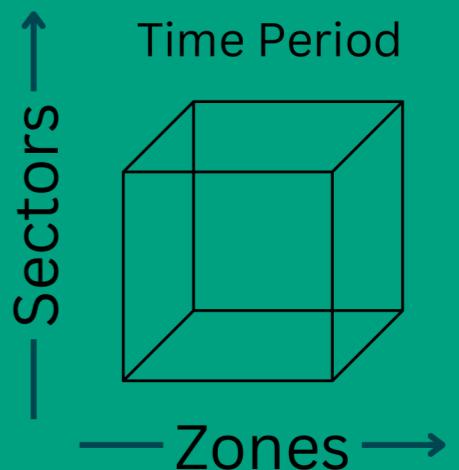
We define different colours for different themes or ranges.

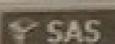


Data Analysis - SAS

A multi-dimensional database is built with various zones, sectors and the time periods in year as the 3 axes of a cube.

The data is imported into SAS EIS Application and the multi dimensional analysis are carried out.



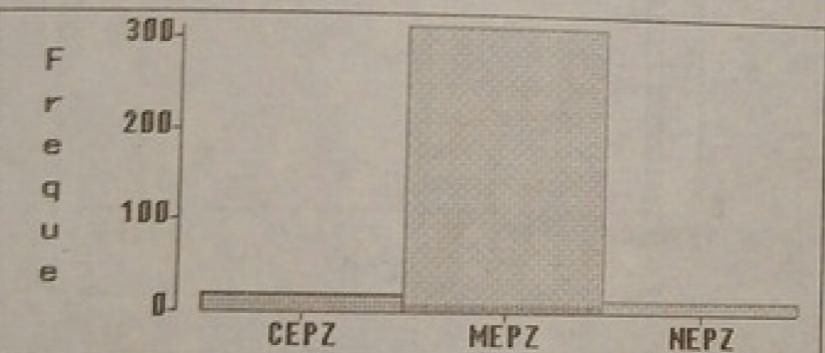


File Edit Analyze Tables Graphic Curves View Options Window Help

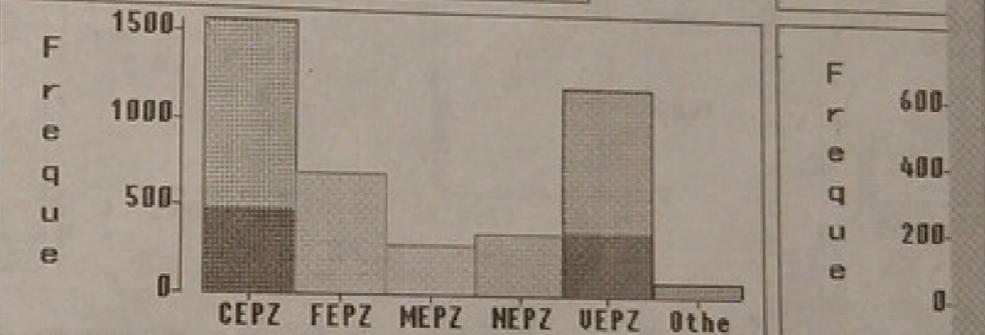


Bar Chart MEPZ.EXPORTS

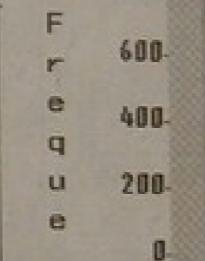
SECTOR = DRUGS



SECTOR = ELECTRONICS-HARDWARE

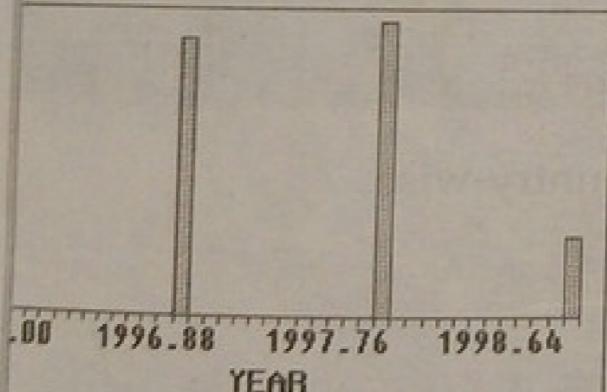


SECTOR

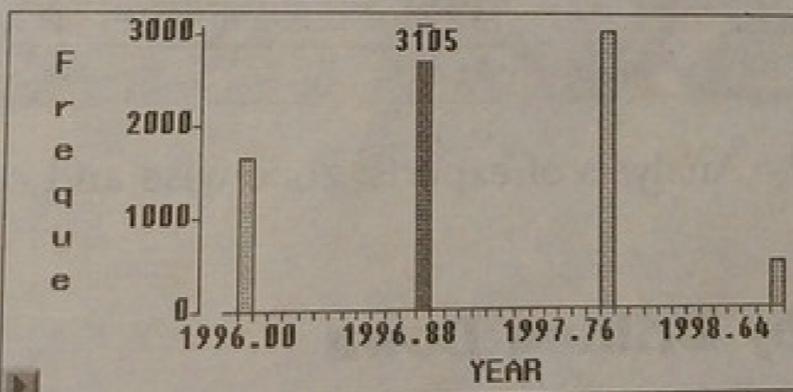


Histogram MEPZ.EXPORTS

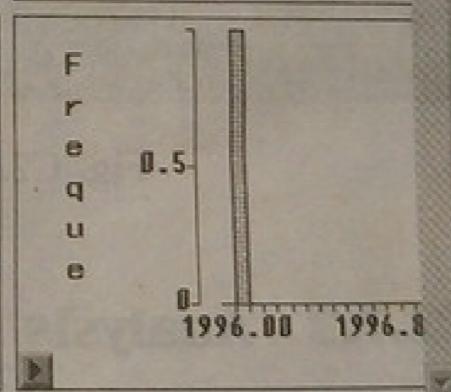
U A E



COUNTRY = U S A



COUNTRY = U S MISC F



File Edit View Insert Format Data Window Help

Start Microsoft Word - paper SAS

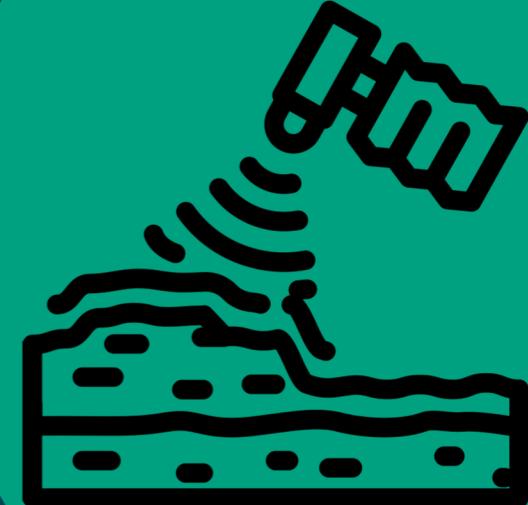
5:39 PM

Analysis by Drilling Down

We can rotate the cube, we slice and dice the category decision variables; any kind of hierarchy and drill down can be defined as per the hierarchy.

The hierarchy can be defined in the metadata repository or can be dynamically altered for any kind of analysis.

The analysis is very flexible and the performances can be reviewed anytime.

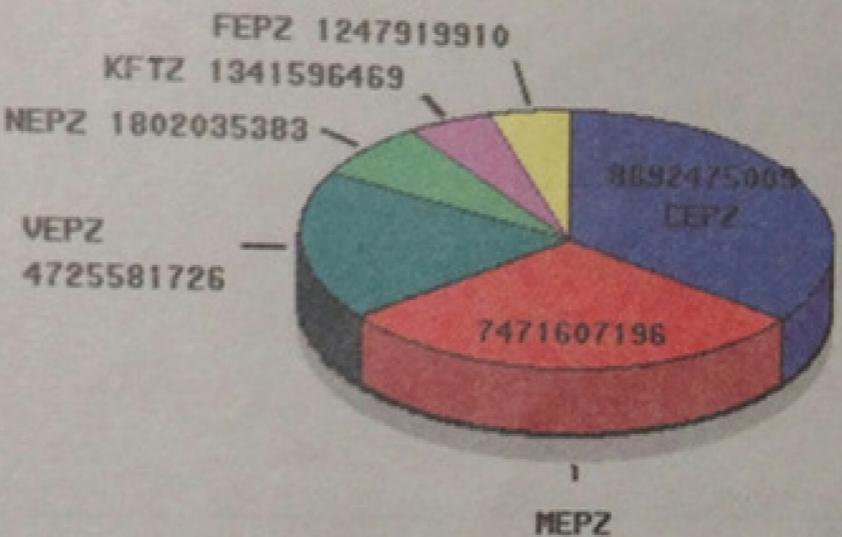




seminar



Zonewise Exports



Multi Variance Analysis

It filters data for specific data requirements and the subset can also be used for analysis.

The details of the actual data can be seen and this quickly narrows down our search to fish out the data for analysis and detail study.



SAS

File Edit View Globals Options Window Help

Zone

EXPORTS

SUM AVG

	ZONE	EXPORTS	
		SUM	AVG
MEPZ	TEXTILES	1292531503	705530.30
	RUBBER AND GLOVES	431613251.0	655947.19
	ENGINEERING	1612698947	2028552.13
	CHEMICAL & ALLIED	309036746.0	1197816.84
	ELECTRONICS-HARDWARE	1837555997	6292999.99
	LEATHER & PRODUCTS	997118463.0	1582727.72
	GEM AND JEWELLERY	USA	41018507.00
		SINGAPORE	4735650.00
		U A E	2997087.00
		HONG KONG	12942027.00
		THAILAND (BANGKOK)	2226050.00
	DRUGS	852104744.0	2713709.38
	GRANITE	75028224.00	609985.56
CEPZ		8892475009	4446237.50
VEPZ		4725581726	2362790.86
FEPZ		1247919910	623959.96
NEPZ		1802035383	901468.43

Page 25 Sec 3 25/26 At 5.5" Ln 1 Col 1 REC TRK EXT DRY WPH UX

Microsoft Word · paper31... 7:40 PM

Start

Traffic Light Analysis

We define different colours and different themes or ranges.

Traffic Light Assessment is a rating system for evaluating the performance in relation to predefined goals.



Benefits



As of November 2023, India's overall exports are estimated to be USD 62.58 billion, showing a positive growth of 1.23% compared to November 2022.

National Logistics Policy (NLP): NLP, launched in September 2022, aims to reduce logistics costs in India. Key initiatives include enhancing efficiency through GST implementation, promoting warehousing compliance, and encouraging geo-tagging of warehouses.

Unified Logistics Interface Platform (ULIP): The ULIP provides secure access to logistics-related information from various ministries. It integrates data from 30 systems across seven ministries, covering more than 1600 data fields for stakeholders' usage.

Conclusion

In conclusion, implementing a data warehouse for EPZ's provides the ability to scale large volume data with seamless presentation of historical, projected and derived data.

By centralizing and integrating trade data, industry trends, and business environment metrics, the MoC can make data-driven decisions to promote economic growth.

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

This translates to informed trade policies, targeted export strategies, improved business regulations, and ultimately, a stronger and more competitive Indian economy.



A close-up photograph of a man with long, dark hair and a full beard. He is looking directly at the camera with a neutral expression. The background is dark and out of focus, suggesting an indoor setting like a bar or restaurant.

**THANK YOU
FOR BEING HERE**