

UNIT 4

Applications of Data Warehousing and Data Mining in Government

(6 Hours) — E-Governance (CACS409)

⚠ Please Note — May Contain Errors

These notes were prepared with AI assistance and **may contain errors, omissions or outdated information** (names, dates, figures, acronyms and project details in particular). Please take care: cross-check everything against your prescribed textbook, class notes and official sources before relying on it for exams or assignments.

4.1 Introduction

Governments collect enormous volumes of data — censuses, taxes, land, health, education, prices and more. Most of it sits in separate departmental databases designed for day-to-day *transactions*, not for analysis. **Data warehousing** and **data mining** turn this scattered data into knowledge that supports planning, policy-making and monitoring.

Key Concepts

- **Data warehouse:** a subject-oriented, integrated, time-variant and non-volatile collection of data that supports management decision-making (Inmon's definition).
- **OLTP vs OLAP:** OLTP (online transaction processing) handles daily operations; OLAP (online analytical processing) supports multidimensional analysis, summaries and trends.
- **ETL:** Extract data from sources, Transform (clean, standardise, integrate), Load into the warehouse.
- **Data mart:** a smaller, subject-specific subset of a warehouse (e.g. agriculture mart).
- **Data mining:** discovering hidden patterns, relationships and trends in large datasets using techniques such as classification, clustering, association rules, regression and anomaly detection.

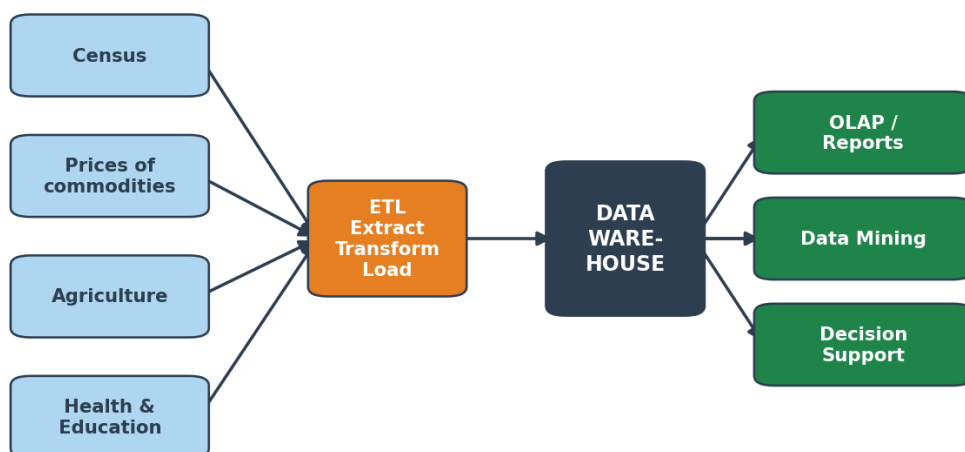


Figure 4.1 — Government data warehouse: sources → ETL → warehouse → analysis

Feature	Operational database (OLTP)	Data warehouse (OLAP)
Purpose	Day-to-day transactions	Analysis and decision support
Data	Current, detailed	Historical, summarised and integrated
Update	Frequent, real-time	Periodic loads; mostly read-only
Users	Clerks, operators	Planners, managers, analysts
Design	Application-oriented	Subject-oriented (e.g. population, prices)

Why are they useful in government?

- Integrate data from many departments into a single, consistent view.
- Track trends over time and across regions; evidence-based planning.
- Detect fraud, duplicate beneficiaries and tax evasion.
- Target subsidies and schemes to the right groups and areas.
- Forecast demand (health, education, food) and early-warn of problems (disease outbreaks, price spikes).

4.2 National Data Warehouses

1. Census Data

A national census collects data on population size, age, sex, households, literacy, occupation, housing, migration and more, usually every 10 years. A **census data warehouse** stores this data at various levels — nation, province, district, local level, ward — so that users can analyse it by many dimensions.

- **Uses:** allocating seats and budget, planning schools, hospitals and roads, estimating labour force and poverty, studying migration and urbanisation, preparing voters' lists and development indices.
- **OLAP examples:** literacy by sex × district × year; households without electricity by ward; age pyramid by province.
- **Data mining examples:** clustering districts with similar demographic profiles; finding factors associated with out-migration; predicting future population.
- **Benefits:** quick, flexible access to data that earlier required long printed reports; comparison across census years.

2. Prices of Essential Commodities

Government agencies monitor the prices of rice, wheat, pulses, edible oil, sugar, vegetables, fuel, etc. in different markets. A **price data warehouse** collects these daily/weekly prices from many markets and stores historical data for analysis.

- **Uses:** inflation (CPI) estimation, detecting price rises and hoarding, timing of imports/exports, setting support prices, targeting food subsidies, market information for farmers and consumers.
- **OLAP examples:** average price of rice by month × market × grade; price difference between hill and Terai markets.
- **Data mining examples:** forecasting seasonal price movements; detecting abnormal jumps; association between fuel price and vegetable price.
- **Benefits:** early warning, more informed regulation, transparency for consumers.

4.3 Other Areas for Data Warehousing and Data Mining

Sector	Typical data	Possible analysis / benefit
Agriculture	Crop area and yield, rainfall, soil, pest data, input use, market prices, subsidies	Yield prediction, crop planning, pest early warning, advisory services, price forecasting
Rural development	Poverty, infrastructure, employment schemes, beneficiaries, microfinance	Target the poorest villages; monitor scheme progress; detect leakage
Health	Disease reports, hospital records, immunisation, nutrition, health workers	Outbreak detection, resource planning, identifying high-risk groups, performance monitoring
Planning	Economic indicators, budgets, project data, national accounts	Scenario analysis, resource allocation, monitoring plan targets
Education	Enrolment, dropout, results, teachers, infrastructure	Identify low-performing areas, predict dropouts, allocate teachers and funds
Commerce and trade	Imports, exports, customs, licences, tariffs, company data	Trade trend analysis, detect under-invoicing, identify export potential
Other sectors	Tax, transport, energy, tourism, police, disaster, social security	Tax-evasion detection, traffic planning, crime hot-spot mapping, disaster risk mapping

Common Data Mining Techniques with Government Examples

Technique	Meaning	Example
Classification	Assign records to known classes	Classify taxpayers as low/high risk
Clustering	Group similar records without predefined classes	Group districts by development level
Association rules	Find items that occur together	Households lacking sanitation also lacking clean water
Regression / forecasting	Predict numeric values	Predict next year's crop yield or price
Anomaly detection	Find unusual records	Spot fraudulent claims or duplicate beneficiaries

Issues and Challenges

- Poor quality, inconsistent or incomplete data; legacy paper records.
- Lack of standards and sharing between departments (“data silos”).
- Privacy, confidentiality and security of citizens’ personal data.
- Shortage of skilled analysts and costly infrastructure.
- Risk of misinterpreting results; need for domain knowledge and ethical use.

Example 4.1 — Application question

A Ministry of Health wants to detect outbreaks of dengue early. Describe how a data warehouse and data mining could help.

Solution

Warehouse: use ETL to load weekly case reports from hospitals and health posts, weather data (temperature, rainfall) and population data into a health data warehouse, organised by disease, place and time.

OLAP: view cases by district × week and compare with previous years.

Data mining: association/regression to link rainfall and temperature with case counts; anomaly detection to flag districts with unusual rises; clustering to find similar high-risk areas.

Outcome: early alerts and better allocation of medicines, staff and awareness campaigns.

Key Takeaways — Unit 4

- ✓ Data warehouse = integrated, subject-oriented, time-variant, non-volatile store for decision support; data comes through ETL.
- ✓ OLAP provides multidimensional views; data mining discovers hidden patterns (classification, clustering, association, forecasting).
- ✓ National warehouses such as Census and Prices of Essential Commodities support planning, subsidy targeting and early warning.
- ✓ Other areas: agriculture, rural development, health, planning, education, commerce and trade, and further sectors.
- ✓ Challenges: data quality, silos, privacy and security, skills and cost.

— End of Unit 4 —

Reminder: these notes may contain errors — verify with your textbook and class notes.