SRI AKILANDESWARI WOMEN'S COLLEGE, WANDIWASH

DATA WAREHOUSE

Class : III BCA

Mrs.M.SUMITHA

Assistant Professor Department of BCA

SWAMY ABEDHANADHA EDUCATIONAL TRUST, WANDIWASH

Data Warehouse

A read-only database for decision analysis

- Subject Oriented
- Integrated
- Time variant
- Nonvolatile

consisting of time stamped operational and external data.

Data Warehouse Purpose

- Identify problems in time to avoid them
- Locate opportunities you might otherwise miss

Data Warehouse: New Approach

An old idea with a new interest because of:

- Cheap Computing Power
- Special Purpose Hardware
- New Data Structures
- Intelligent Software

Warehousing Problems

Business Issues

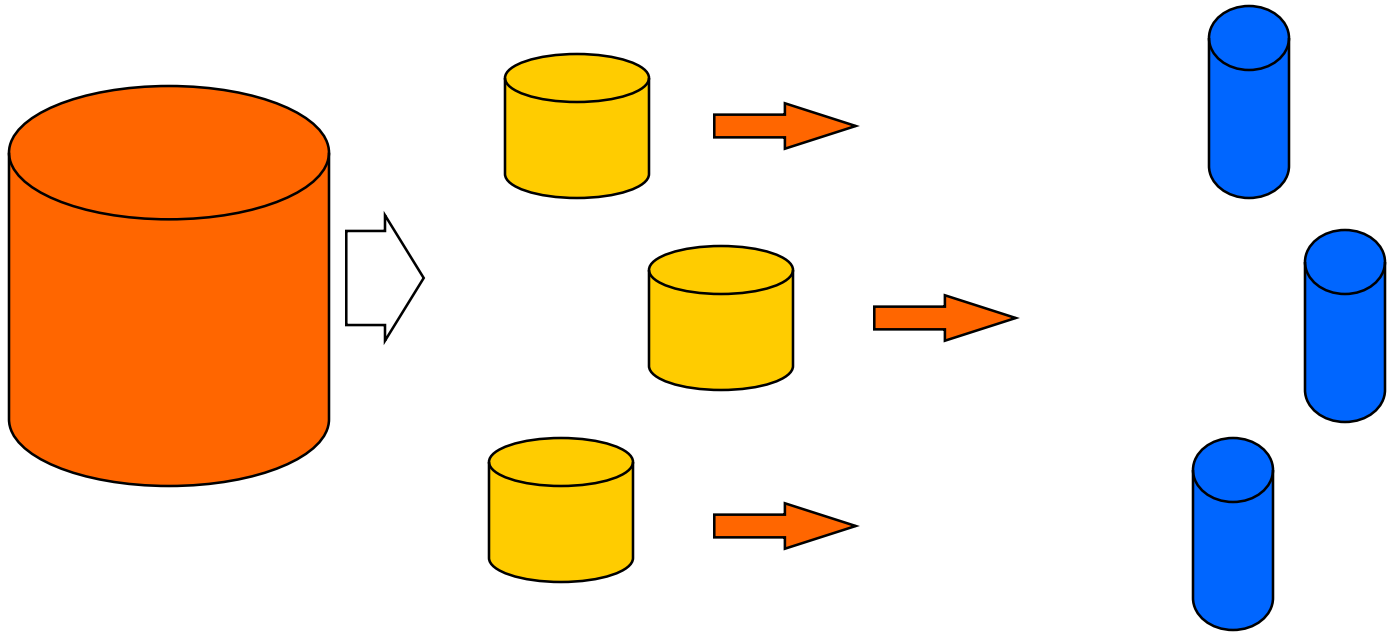
- Data Quantity
- Data Accuracy
- Maintenance
- Ownership
- Cost

Warehousing Problems

Analysis Issues

- User Interface
- Intelligent Processing

EUC Data Architecture



Enterprise Data Warehouse

Data Marts

Business Packages

Needed Environment

- Understand DSS technology
- Understand relational databases
- Know company data resources
- Transformation and integration tools
- Possess DBA skills
- Install network and telecommunication hardware and software
- Possess front end software
- Possess meta-data navigation tools

Two Approaches

- **Classical Enterprise Database**

  Typically contains operational data that integrates information from all areas of the organization.

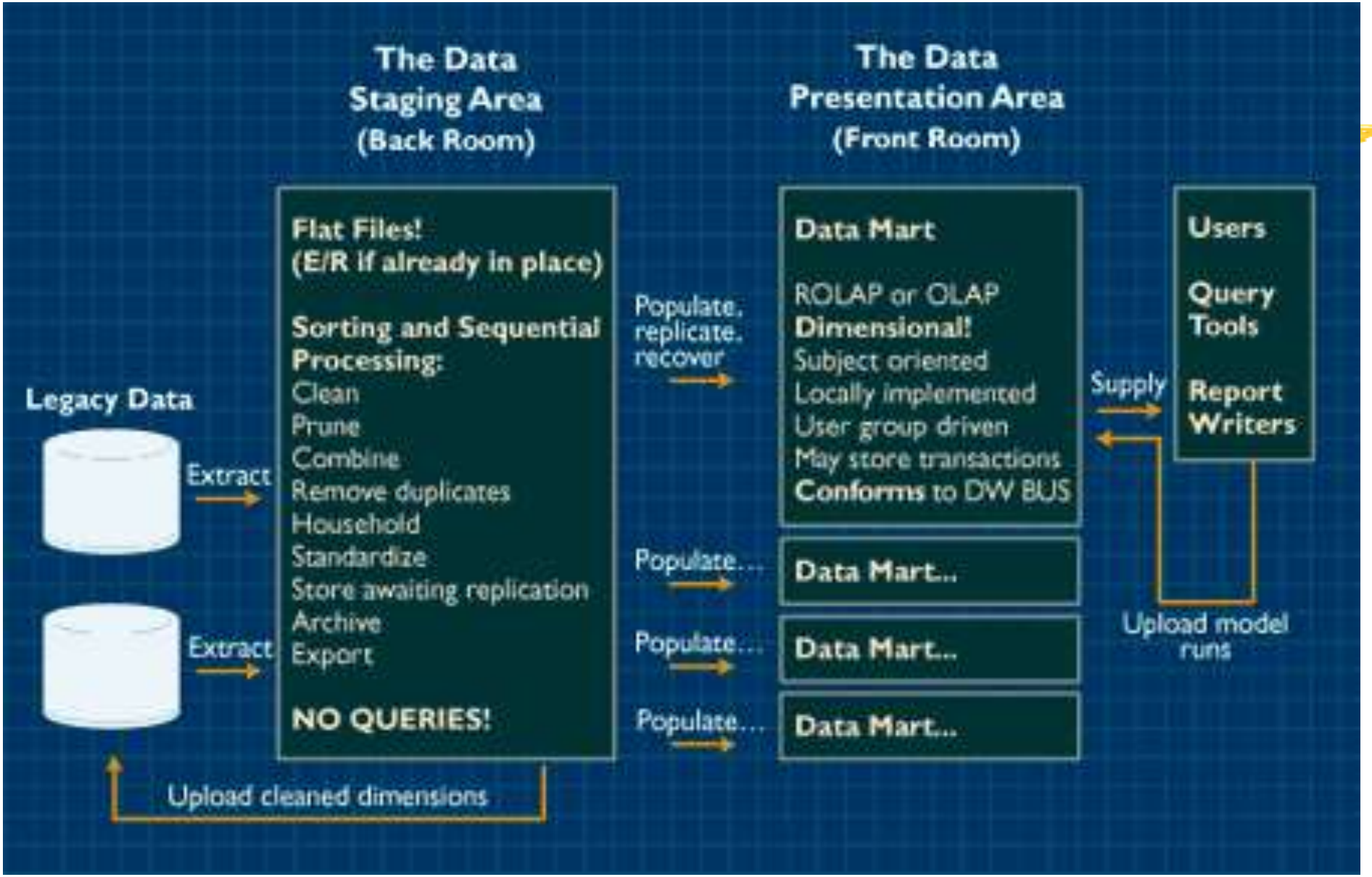
- **Data Mart**

  Extracted and managerial support data designed for departmental or EUC applications

Data Warehouse vs Operational Databases

- Highly tuned
- Real time Data
- Detailed records
- Current values
- Accesses small amounts of data in a predictable manner

- Flexible access
- Consistent timing
- Summarized as appropriate
- Historical
- Access large amounts of data in unexpected ways


The Data Staging Area (Back Room)
The Data Presentation Area (Front Room)
Flat Files! (E/R if already in place)
Sorting and Sequential Processing:
Clean
Prune
Combine
Remove duplicates
Household
Standardize
Store awaiting replication
Archive
Export
NO QUERIES!
Data Mart
ROLAP or OLAP
Dimensional!
Subject oriented
Locally implemented
User group driven
May store transactions
Conforms to DW BUS
Data Mart...
Data Mart...
Data Mart...
Users
Query Tools
Report Writers
Legacy Data
Extract
Extract
Populate, replicate, recover
Populate...
Populate...
Populate...
Supply
Upload model runs
Upload cleaned dimensions

Separate Read Only Database

Operational and decision support processing are fundamentally different.

- Flexible access
- Consistent timing
- Historical data
- Environmental references
- Common reference

Front End Access

Tailored access programs in user form, usually client-server

- General purpose GUI products (e.g. Access, PowerBuilder)
- Custom access routines

Development Cycles

Enterprise	Warehouse
Traditional SDLC	Iterative Development
Requirements Driven	Data Driven

Data Warehouse

A database for departmental decision support. Contains detail data brought in through a feed and clean process.

- Data extracted from source files
- Data is integrated and transformed
- Data resides in a read-only database
- User access via a front end tool or application

Data Warehouse Design Basics

Must provide flexible access, "what if" processing, and extensive reporting on data subsets.

- Can tables be scanned in a reasonable time
- Can indexes be added as needed.
- If not, consider partitioning or summary tables

Data Warehouse Design Basics

- Iterative Development with only part of the warehouse built at one time
- Used and modified with frequent user feedback and modification

1. Implement, test, bias & completeness
2. Create and execute applications
3. Identify Requirements
4. Iterate

Data Development Roles

- **Data Analyst:**

  Works with user groups to understand business and technical requirements.

- **Data Architect:**

  Manages company data. Must understand business needs as well as data implications.

Analyst-Architect Communication Requirements

- Formal lines of communication
- Data warehouse council that meets monthly
- Analyst and architect retreat
- Formally identify analysts and architects
- Ensure comparable levels of personnel
- Both understand business requirements
- Architect goals available to analysts
- Analyst strategy and issues available to architects

Data Warehouse & Design Issues

- Small systems are easier to manage than large ones.
- Adding attributes is easier than changing or deleting them.

Data Warehousedesign: Proposed Changes

- Will the change disrupt the current system?
- Will the change add significantly more detail?
- Will the change disregard existing data?
- Will the change add a new table or change an existing one?
- What granularity?
- Does it fit the current data model?
- How much programming is required?
- How many resources will it consume?

Data Warehouses Extracted Data

- *Clean and feed.*
- Internal data
  - Summarized
  - Historical
  - Accuracy
  - Integration
- External data
- Generated data

Data Warehouse Extraction Issues

Internal data often comes from separate operational databases.

- Reconcile formats
- Remove intelligent codes
- Verify accuracy
- Unify time stamps

OLAP and MDBMS

- Online Analytic Processing
  - Primarily relational models from which targeted extractions can be developed
  - Complex for users
- Multidimensional DBMS
  - Pre-modeled data cubes for efficient access
  - Complex for design