Data Mining Interface, Security

Data Mining Interface

- Data Mining Interface (DMI) is a web-based interactive dynamic report building module
- DMI report immediately access current data and refresh automatically when new data becomes available.
- Trending and baseline data is also available for customized report
- Trending data is transparently used when necessary while baseline data is mixed with current data on the same screen
- DMI report have variable time range settings resolution setting and dynamic sorting and filter mechanism

- Report definitions are saved in database and reports are rerun when opened
- The DMI is equipped with an integrated persistent report cache that optimizes report re-run requests in context of real time data changes in database
- DMI uses product specific data views. Each data view supports its own set of dimensions and metrices.
- Trending and baseline data is also available for DMI reports. Trending data is transparently used when necessary while baseline data is mixed with current data on the same screen.
- use DMI to generate tabular reports and charts and mix multiple report section on same page. The reports can have a hierarchical structure with contextual drill down, sibling and parent report

Data Mining interface or Data warehouse interface and performance

- There are two kind of interfaces
- Programmatic Interface

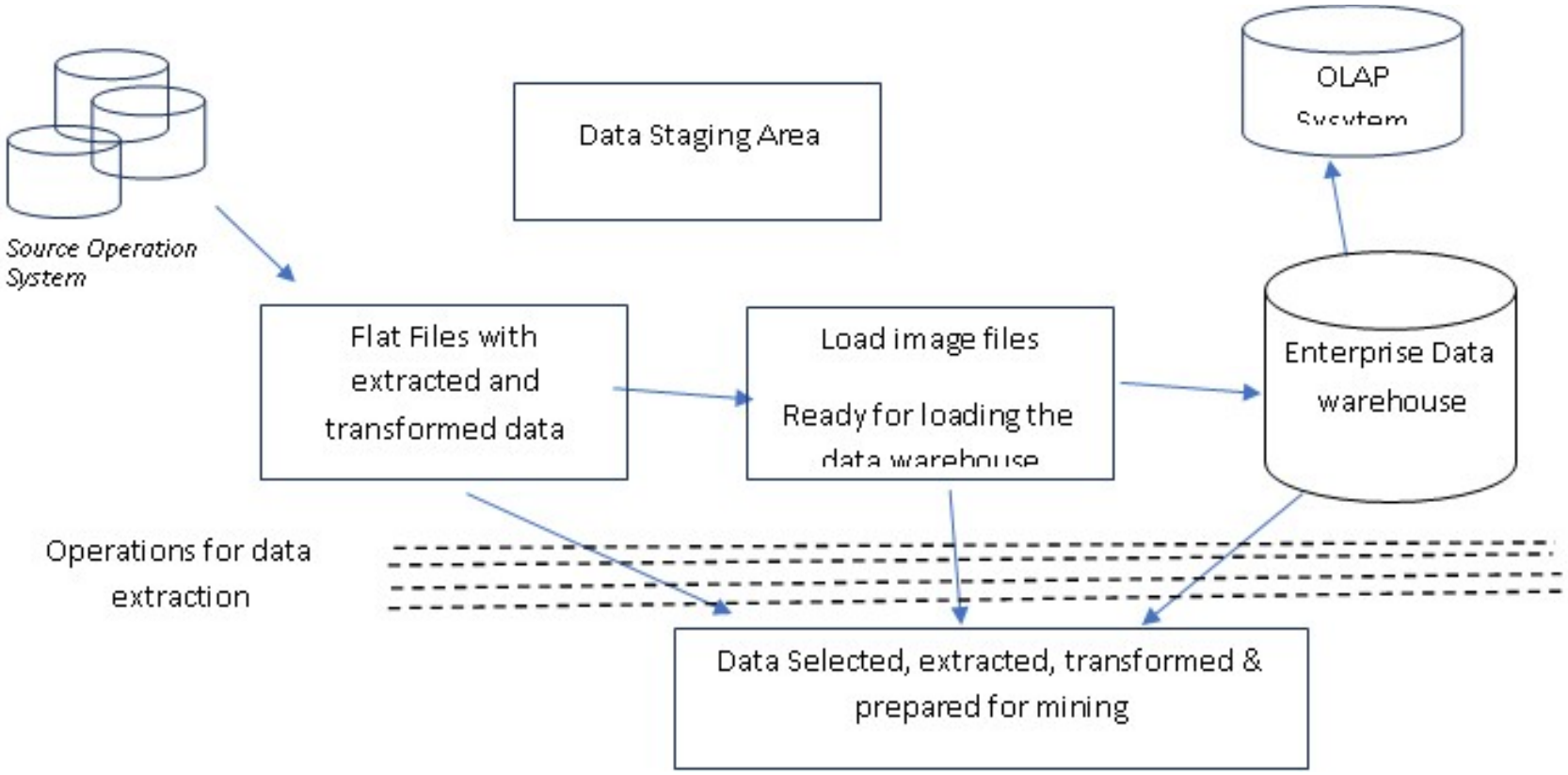
  For ex:

  - Java interface based on Java data Mining

  - PL/SQL Interface

- Graphical User Interface

Graphical User Interface



Data Warehousing - Security

- The objective of a data warehouse is to make large amounts of data easily accessible to the users, hence allowing the users to extract information about the business as a whole.
- But we know that there could be some security restrictions applied on the data that can be an obstacle for accessing the information.
- If the analyst has a restricted view of data, then it is impossible to capture a complete picture of the trends within the business.

Security Requirements

- We should consider the following possibilities during the design phase.
  - Whether the new data sources will require new security and/or audit restrictions to be implemented?
  - Whether the new users added who have restricted access to data that is already generally available?
- The following activities get affected by security measures −
  - User access
  - Data load
  - Data movement
  - Query generation

User Access

- We need to first classify the data and then classify the users on the basis of the data they can access.
- **Data Classification**
- The following two approaches can be used to classify the data
  - Data can be classified according to its sensitivity.
  - Data can also be classified according to the job function.
- **User classification**
  - The following approaches can be used to classify the users
  - Users can be classified as per the hierarchy of users in an organization, i.e., users can be classified by departments, sections, groups, and so on.
  - Users can also be classified according to their role, with people grouped across departments based on their role.

Audit Requirements

- Auditing is a subset of security, a costly activity. Auditing can cause heavy overheads on the system.
- To complete an audit in time, we require more hardware and therefore, it is recommended that wherever possible, auditing should be switched off.
- Audit requirements can be categorized as follows −
  - Connections
  - Disconnections
  - Data access
  - Data change

Network Requirements

- Network security is as important as other securities. We cannot ignore the network security requirement.
- We need to consider the following issues −
  - Is it necessary to encrypt data before transferring it to the data warehouse?
  - Are there restrictions on which network routes the data can take?

Data Movement

- Suppose we need to transfer some restricted data as a flat file to be loaded.
- When the data is loaded into the data warehouse, the following questions are raised −
  - Where is the flat file stored?
  - Who has access to that disk space?
  - Do you backup encrypted or decrypted versions?
  - Do these backups need to be made to special tapes that are stored separately?
  - Who has access to these tapes?
  - Where is that temporary table to be held?
  - How do you make such table visible?

Documentation

- The audit and security requirements need to be properly documented.
- This will be treated as a part of justification. This document can contain all the information gathered from −
  - Data classification
  - User classification
  - Network requirements
  - Data movement and storage requirements
  - All auditable actions

Impact of Security on Design

- Security affects the application code and the development timescales.
- Security affects the following area −
  - Application development
  - Database design
  - Testing