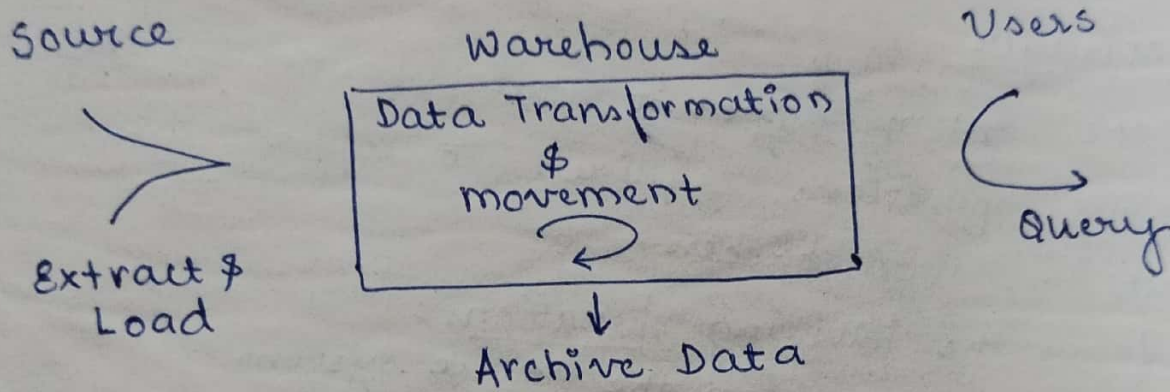


Data Warehouse Process

Process flow within a data warehouse

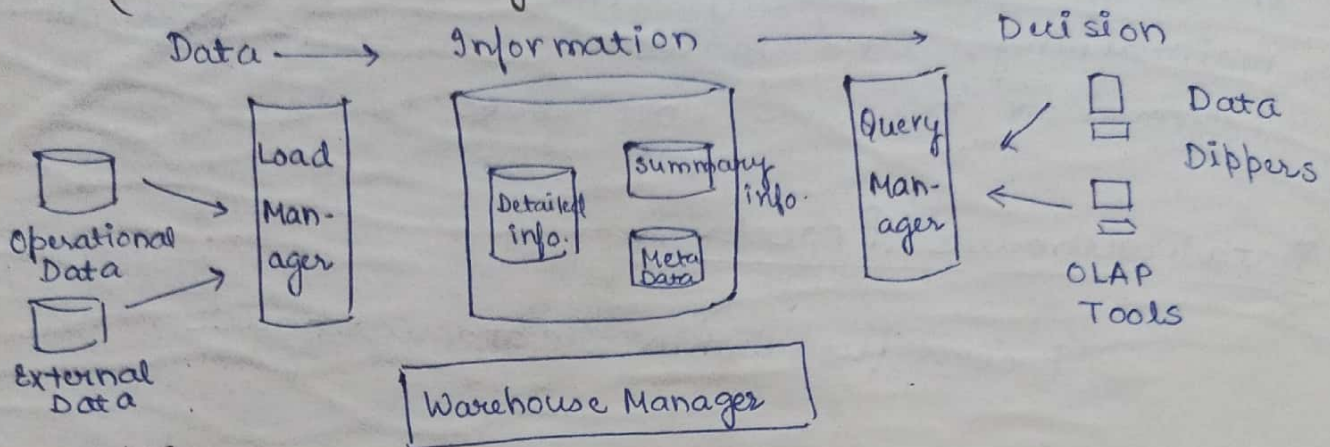


→ Before we create an architecture we first must understand the major processes that constitute the data warehouse. These processes are:

- 1) Extract and load the data
- 2) Clean and transform data into a form that can cope up with large volumes and provide good query performance.
- 3) Back up and archive data
- 4) Manage queries and direct them to the appropriate data sources.

Process Architecture

(Warehouse Management and support Processes)



6/3/19

Load Manager

It is the system component that performs all the necessary operations required to support the extract and load process.

It performs the following operations:

- 1) Extract the data from the source system.
- 2) Fast load the extracted data into a temporary data store.
- 3) Perform simple transformations into a structure similar to the one in the data warehouse.

Tools used:

- 1) Fast load - Database fast load facility
- 2) Simple transformation - copy management tool
- 3) Complex chunking - stored procedures, C, C++

Warehouse Manager

It is the system component that performs all the operations necessary to support warehouse mgmt. process.

It is typically constructed using combination of 3rd party system mgmt. sw, C program, shell scripts.

Its complexity vary blw specific solution.

Operations:

- 1) Analyze the data to perform consistency.
- 2) Create indexes, partition views, business views against the base data.
- 3) Generate denormalizations if appropriate.
- 4) Update all existing aggregations.
- 5) Back-up incrementally
- 6) Archive data that has reached the end of its capture life.

Query Manager

It performs all the operations necessary to support the query management process.

It is typically constructed using combination of access tools, monitoring tools, database facilities, C programs, shell script.

Operations:

- 1) Direct queries to the appropriate table.
- 2) Schedule the execution of user query.

