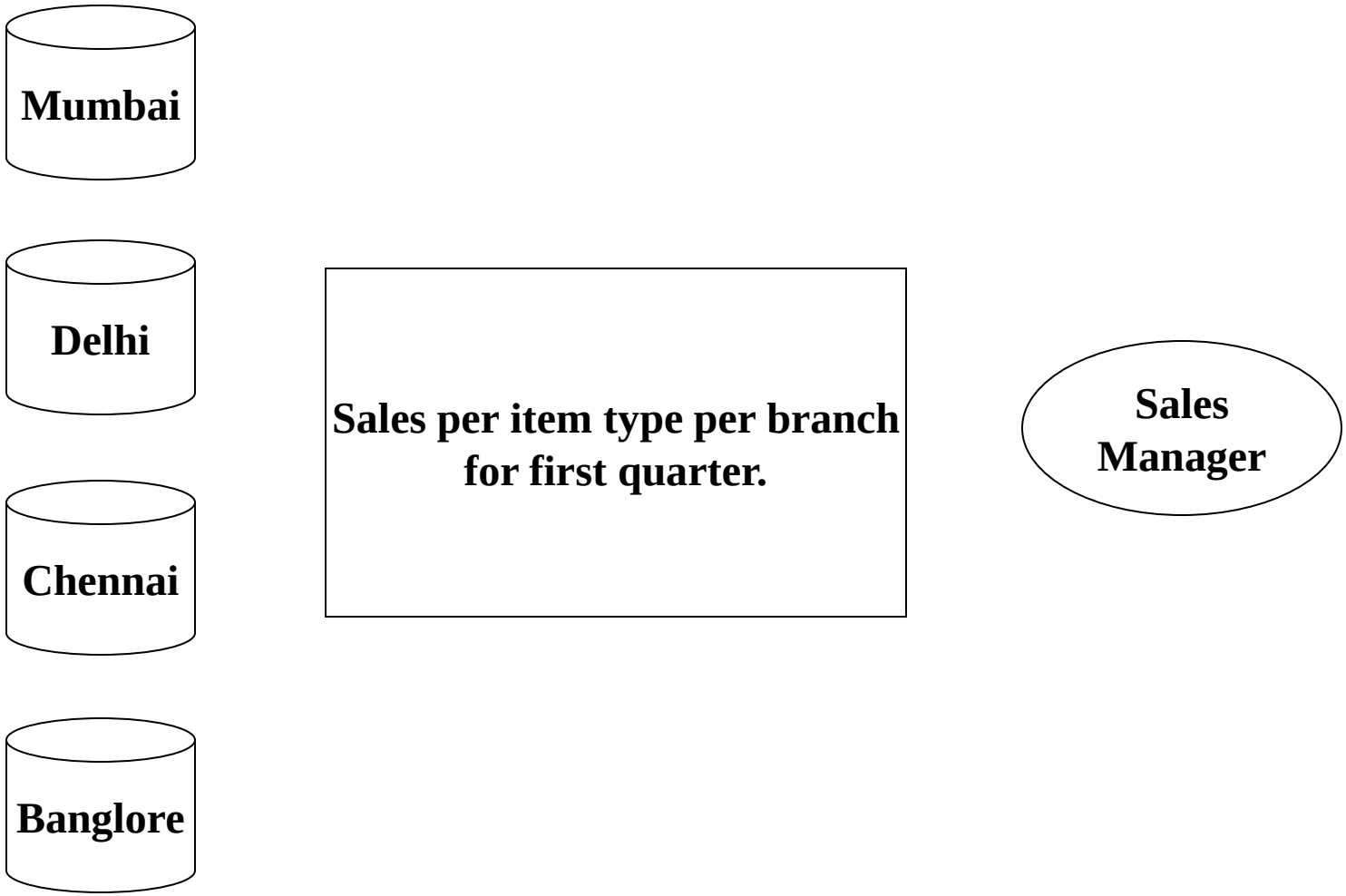
Introduction to Data Warehousing

Scenario 1

ABC Pvt Ltd is a company with branches at Mumbai, Delhi, Chennai and Banglore. The Sales Manager wants quarterly sales report.

Each branch has a separate operational system.

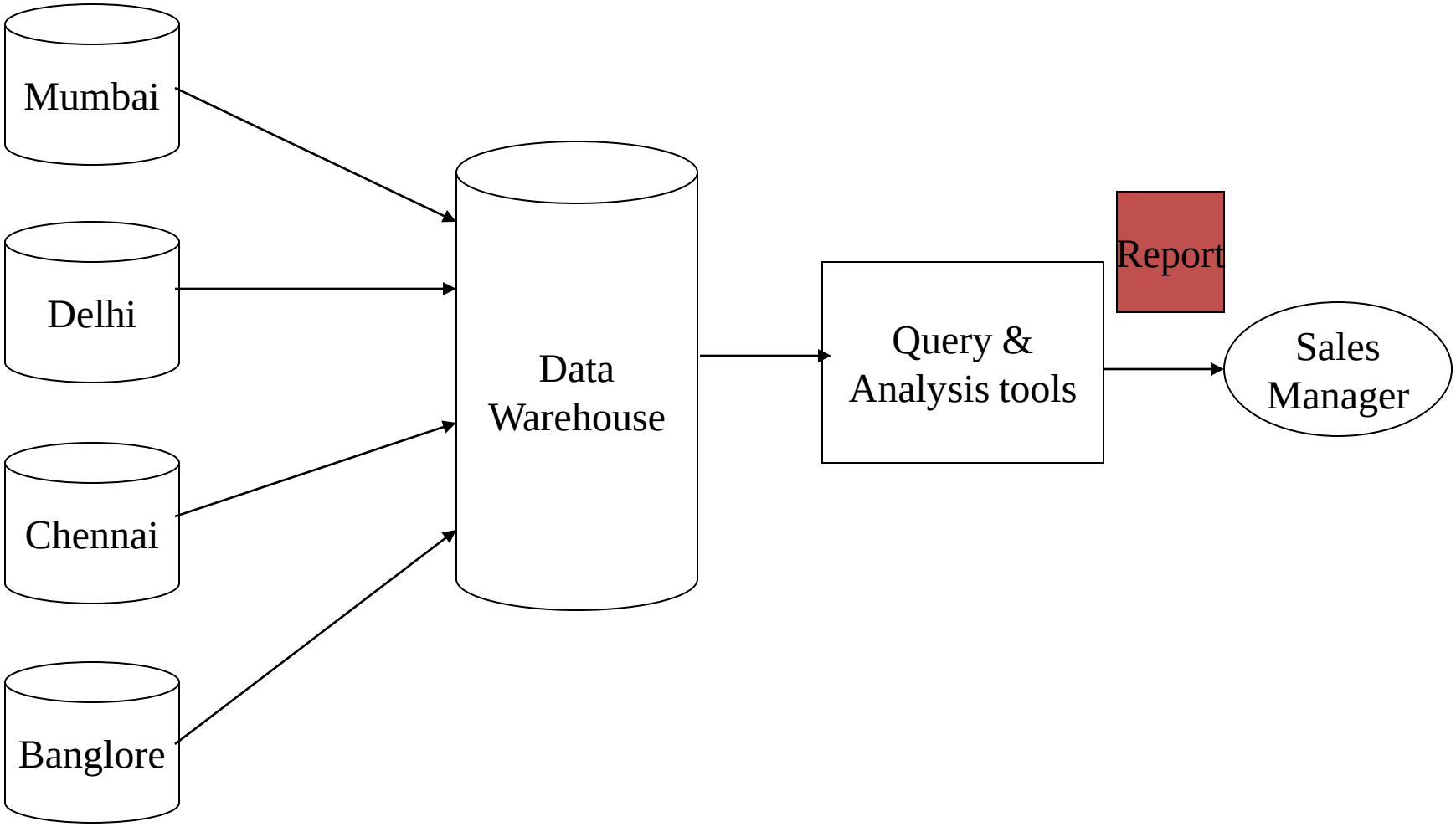
Scenario 1 : ABC Pvt Ltd.



Solution 1:ABC Pvt Ltd.

- Extract sales information from each database.
- Store the information in a common repository at a single site.

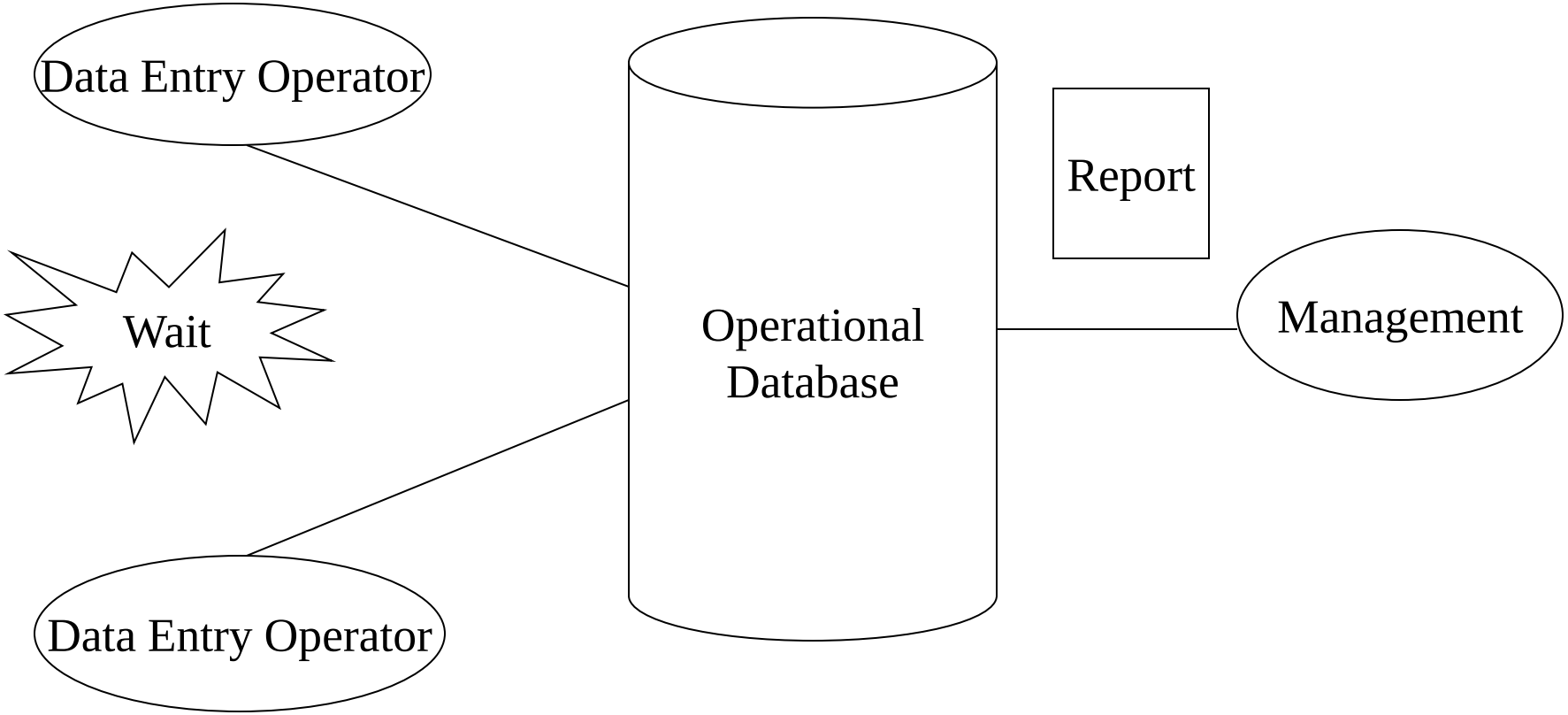
Solution 1:ABC Pvt Ltd.



Scenario 2

One Stop Shopping Super Market has huge operational database.Whenever Executives wants some report the OLTP system becomes
slow and data entry operators have to wait for some time.

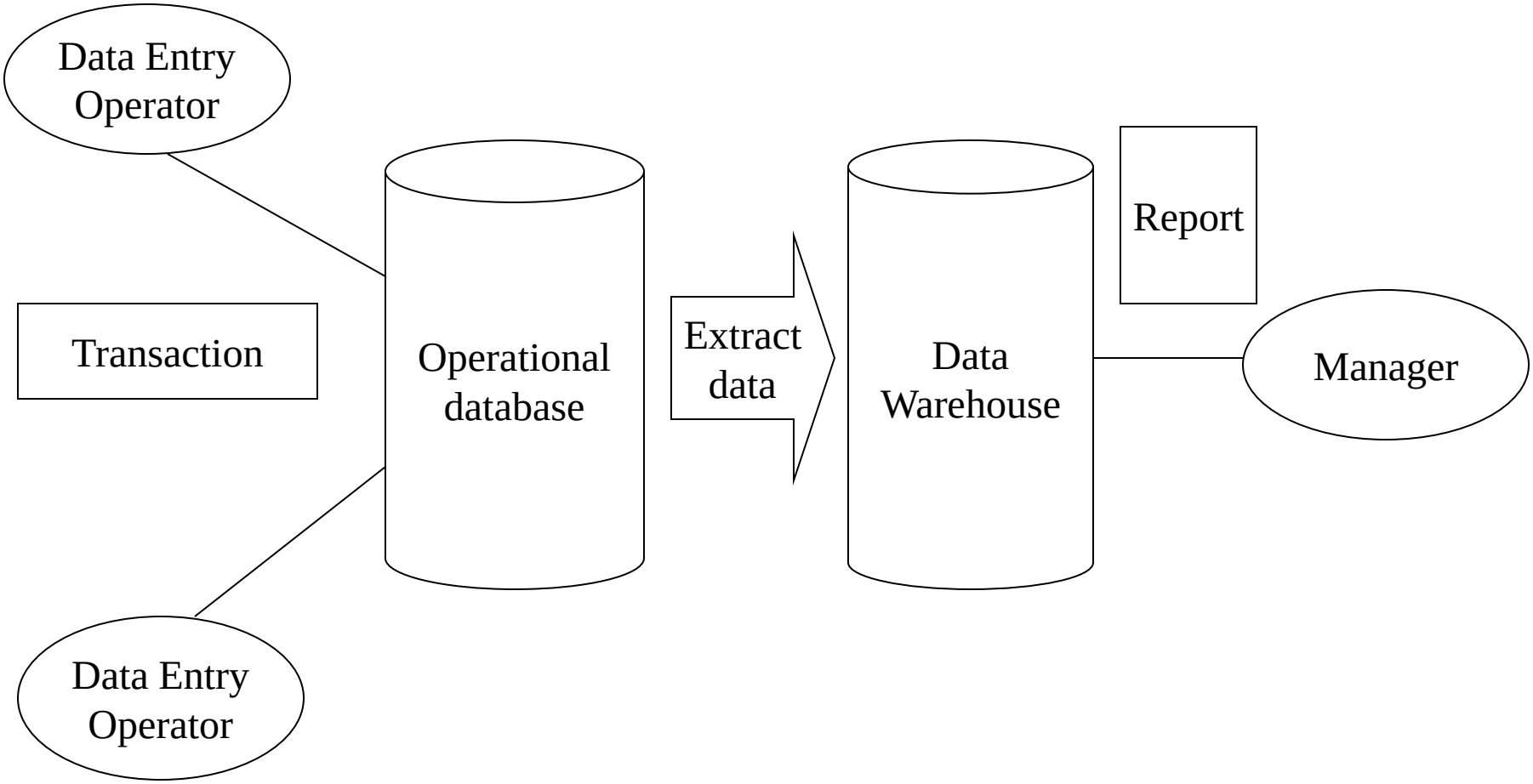
Scenario 2 : One Stop Shopping



Solution 2

- Extract data needed for analysis from operational database.
- Store it in warehouse.
- Refresh warehouse at regular interval so that it contains up to date information for analysis.
- Warehouse will contain data with historical perspective.

Solution 2


Data Entry Operator
Transaction
Data Entry Operator
Operational database
Extract data
Data Warehouse
Report
Manager

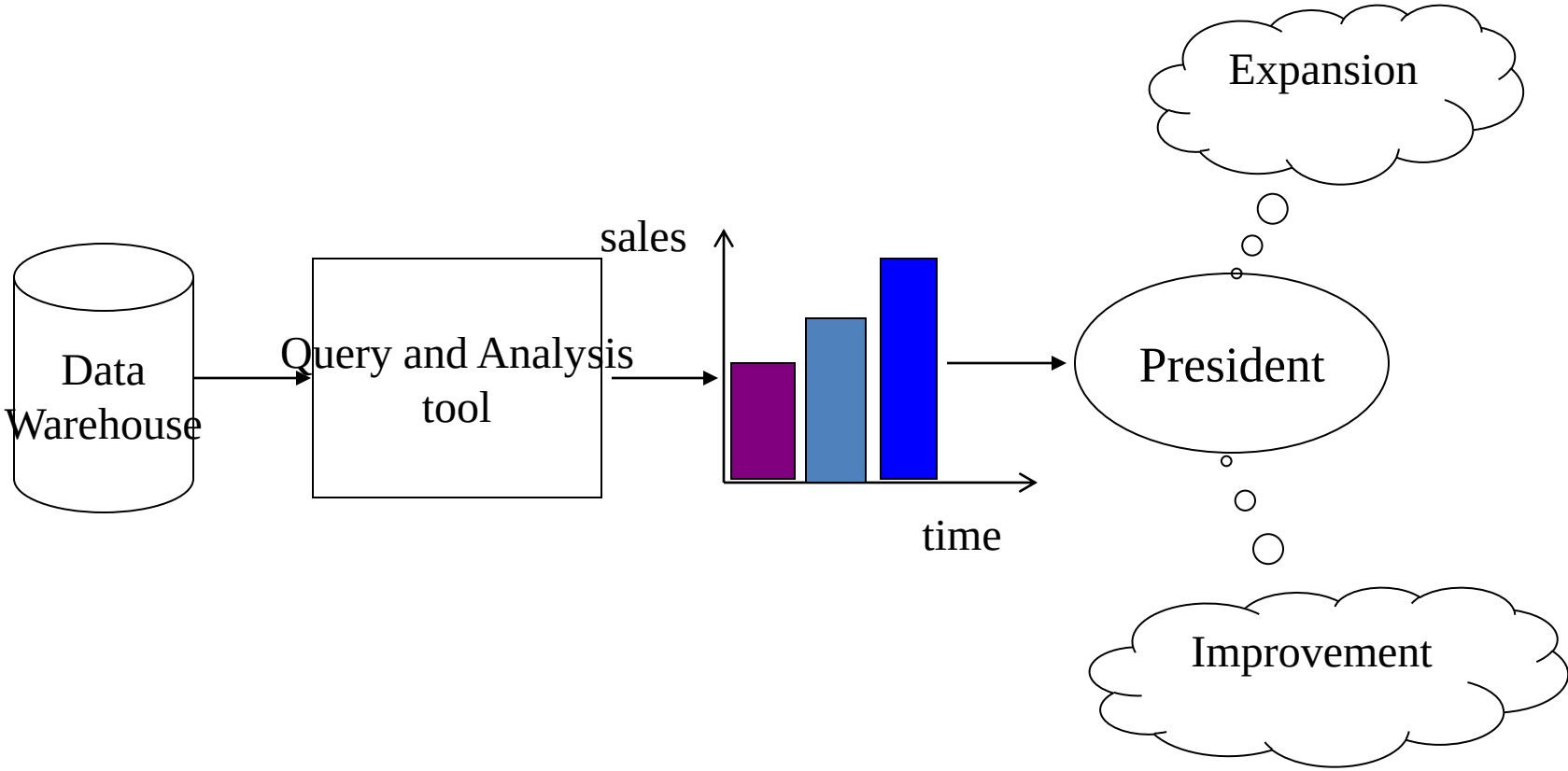
Scenario 3

Cakes & Cookies is a small,new company.President of the company wants his company should grow.He needs information so that he can make correct decisions.

Solution 3

- Improve the quality of data before loading it into the warehouse.
- Perform data cleaning and transformation before loading the data.
- Use query analysis tools to support adhoc queries.

Solution 3


Expansion
sales
Data Warehouse
Query and Analysis tool
President
time
Improvement

What is Data Warehouse??

Inmons's definition

A data warehouse is

-subject-oriented,

-integrated,

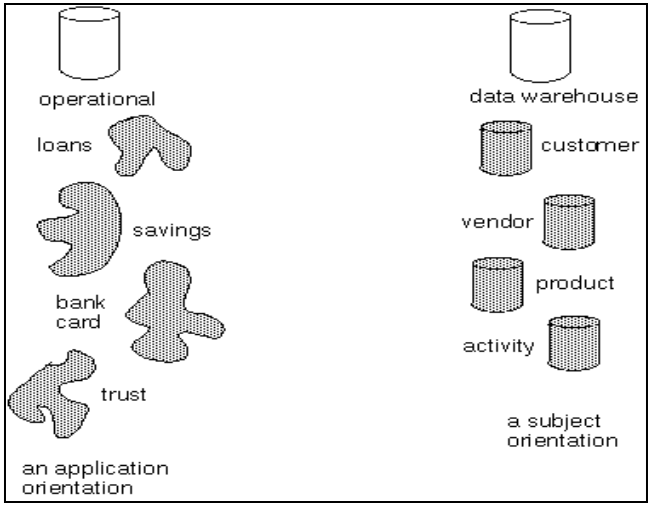
-time-variant,

-nonvolatile

collection of data in support of management's decision making process.

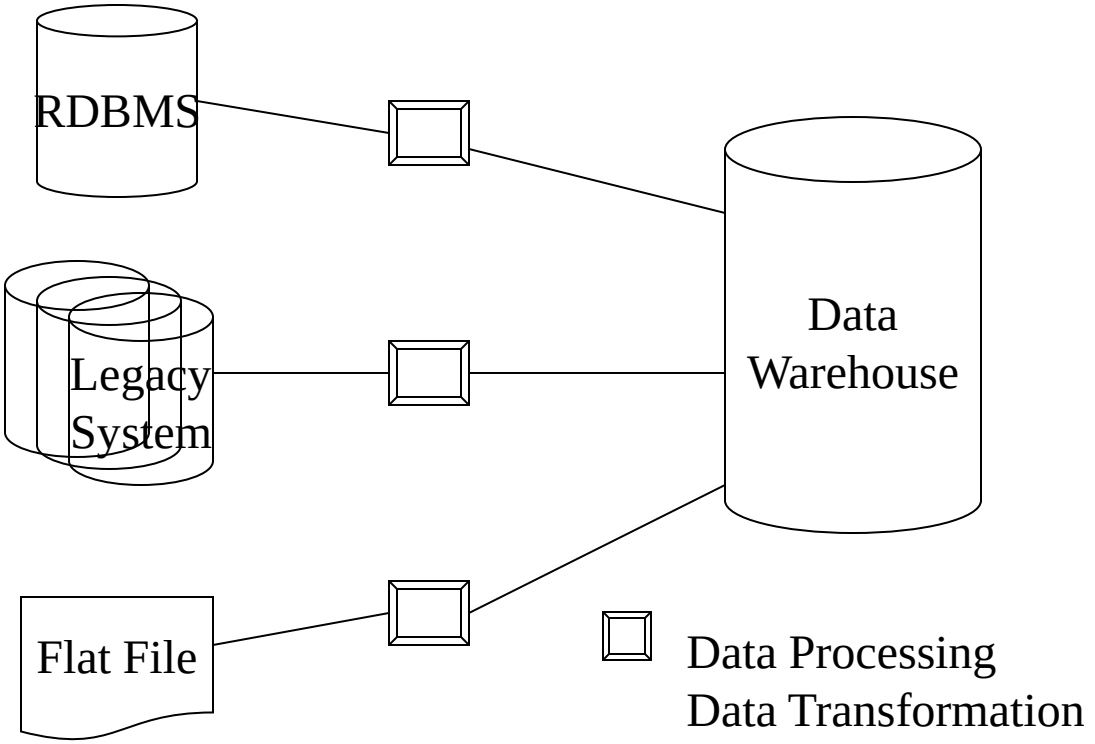
Subject-oriented

- Data warehouse is organized around subjects such as sales,product,customer.
- It focuses on modeling and analysis of data for decision makers.
- Excludes data not useful in decision support process.



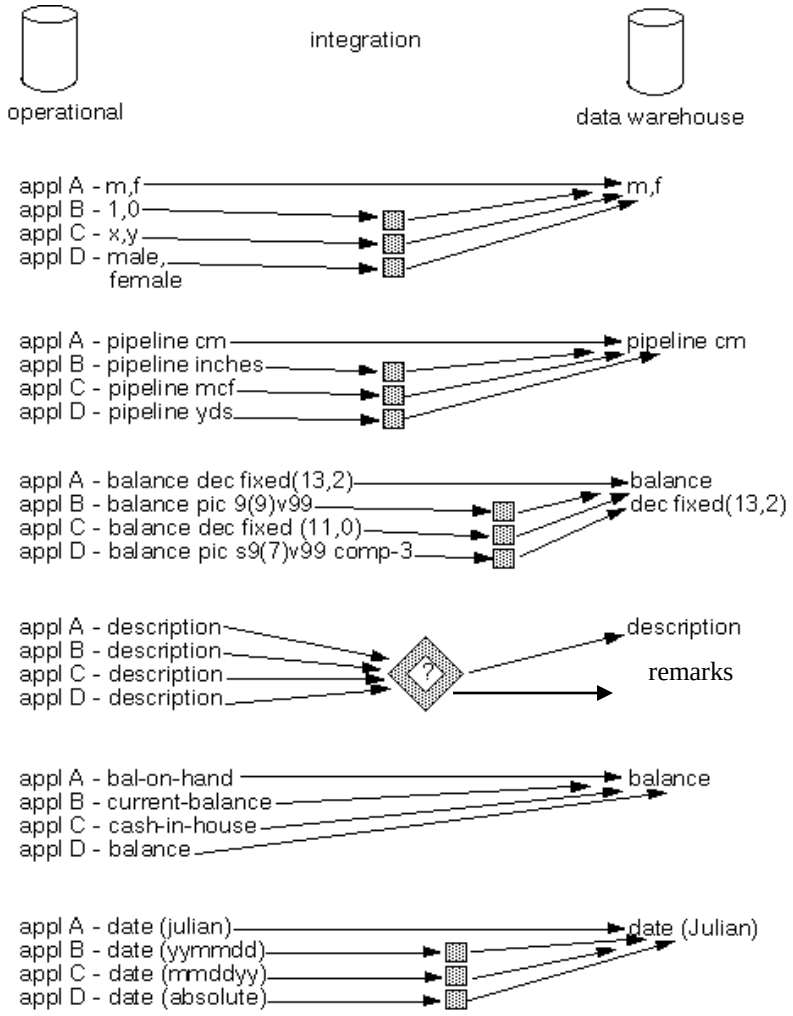
Integration

- Data Warehouse is constructed by integrating multiple heterogeneous sources.
- Data Preprocessing are applied to ensure consistency.



Integration

- In terms of data.
  - encoding structures.
  - Measurement of attributes.
  - physical attribute. of data
  - naming conventions.
  - Data type format

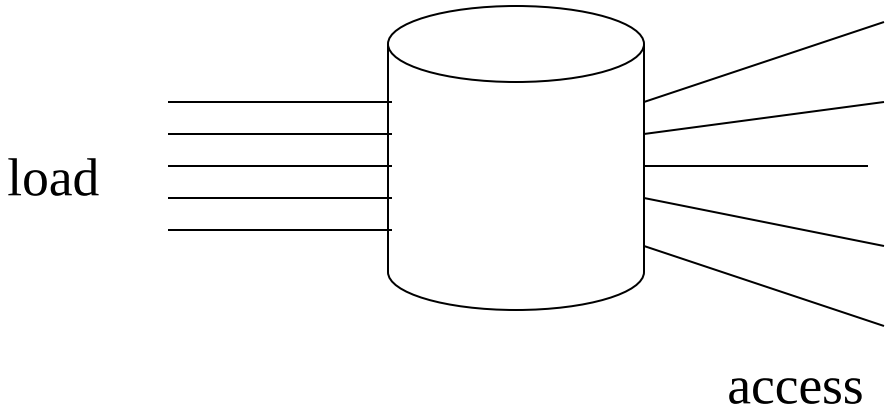


Time-variant

- Provides information from historical perspective e.g. past 5-10 years
- Every key structure contains either implicitly or explicitly an element of time

Nonvolatile

- Data once recorded cannot be updated.
- Data warehouse requires two operations in data accessing
  - Initial loading of data
  - Access of data



Operational v/s Information System

Features	Operational	Information
Characteristics	Operational processing	Informational processing
Orientation	Transaction	Analysis
User	Clerk,DBA,database professional	Knowledge workers
Function	Day to day operation	Decision support
Data	Current	Historical
View	Detailed,flat relational	Summarized, multidimensional
DB design	Application oriented	Subject oriented
Unit of work	Short ,simple transaction	Complex query
Access	Read/write	Mostly read

Operational v/s Information System

Features	Operational	Information
Focus	Data in	Information out
Number of records accessed	tens	millions
Number of users	thousands	hundreds
DB size	100MB to GB	100 GB to TB
Priority	High performance,high availability	High flexibility,end-user autonomy
Metric	Transaction throughput	Query througput

References

- Data Mining: Concepts and Techniques by Han,Kamber.
- www.dwinfocenter.org
- www.datawarehousingonline.com

Thank You