Data Warehousing

Architecture

Introduction

- The structure that brings all the components of a data warehouse together is known as the architecture.
- Correct architecture is critical for the success of your data warehouse.
- Architecture includes a number of factors
- It includes the integrated data
- The architecture includes everything that is needed to prepare the data and store it.
- It also includes all the means for delivering information from your data warehouse
- The architecture is further composed of the rules, procedures, and functions that enable your data warehouse to work
- The architecture is made up of the technology that empowers your data warehouse.

What is a Data Warehouse Architecture

- Primarily based on the business processes of a business enterprise
- Conceptualization of how the data warehouse is built

Purpose of the Architecture

- The architecture provides the overall framework for developing and deploying your data warehouse;
- It is a comprehensive blueprint.
- The architecture defines the standards, measurements, general design, and support techniques.

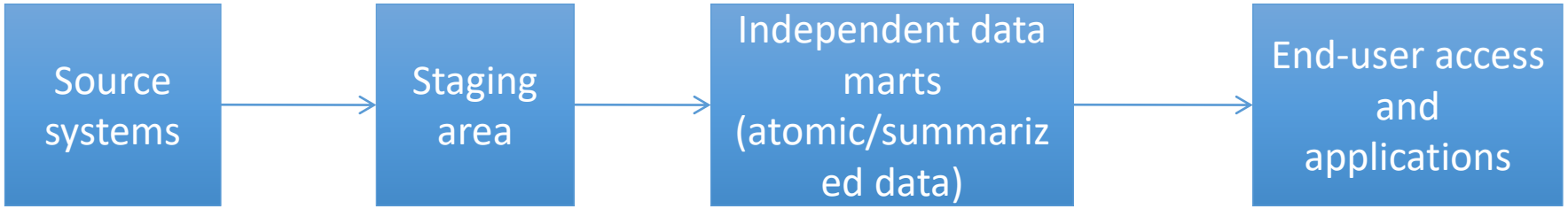
Five Main Data Warehouse Architectures

- Independent Data Marts
- Data Mart Bus Architecture
- Hub-and-Spoke
- Centralized Data Warehouse
- Federated Architecture

Independent Data Marts

- Data marts that are independent of each other
- Often created by organization units
- Inconsistent data definitions and different dimensions and measures

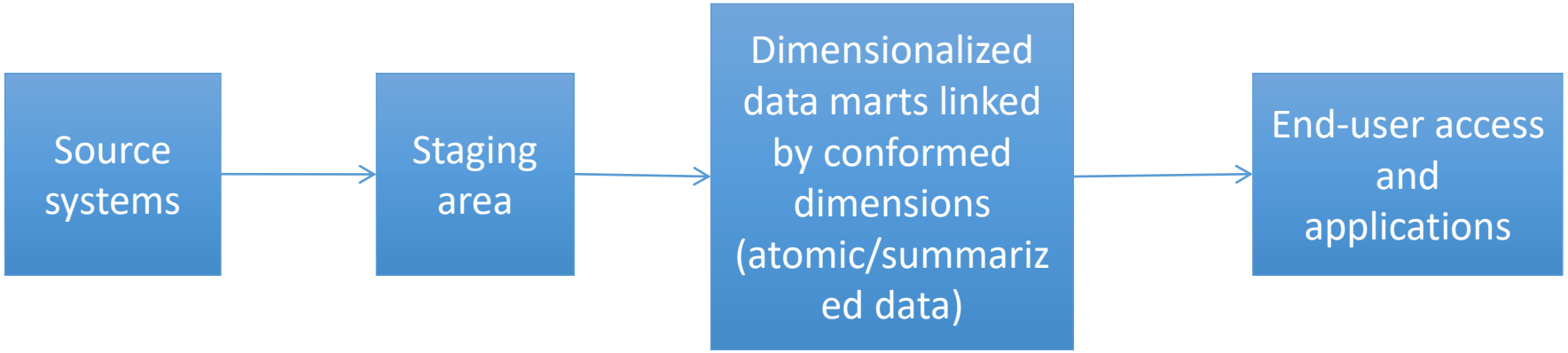
Independent Data Marts Diagram



Data Mart Bus Architecture

- Creation starts with a business requirements analysis for a specific process such as orders, deliveries, customer calls, or billing.
- One mart is created for a single business process
- Additional marts are developed using the conformed dimensions of the first mart

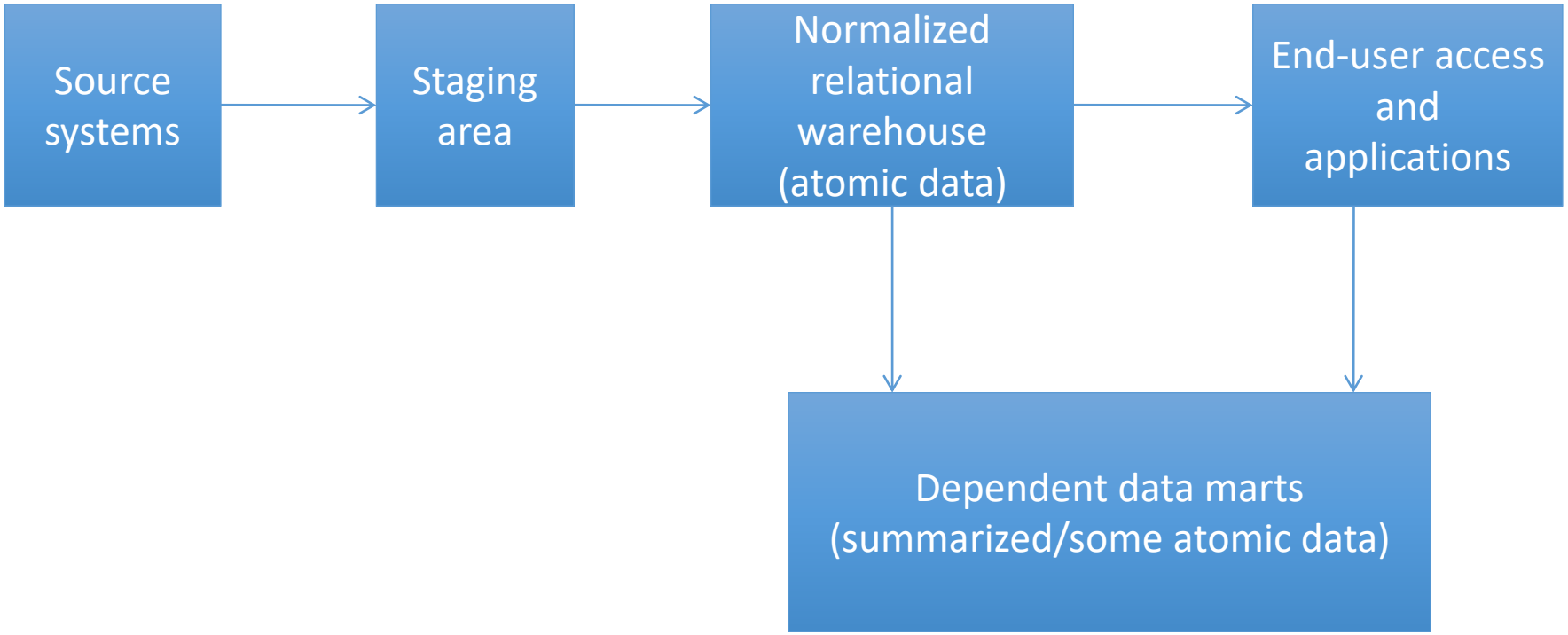
Data Mart Bus Architecture Diagram



Hub-and-Spoke

- Developed after an enterprise-level analysis of data requirements
- Focused on building a scalable and maintainable infrastructure
- Developed in an iterative manner
- Dependent data marts, obtain the data from the warehouse
- Consist of a centralized hub that accepts requests from multiple applications that are connected through spokes

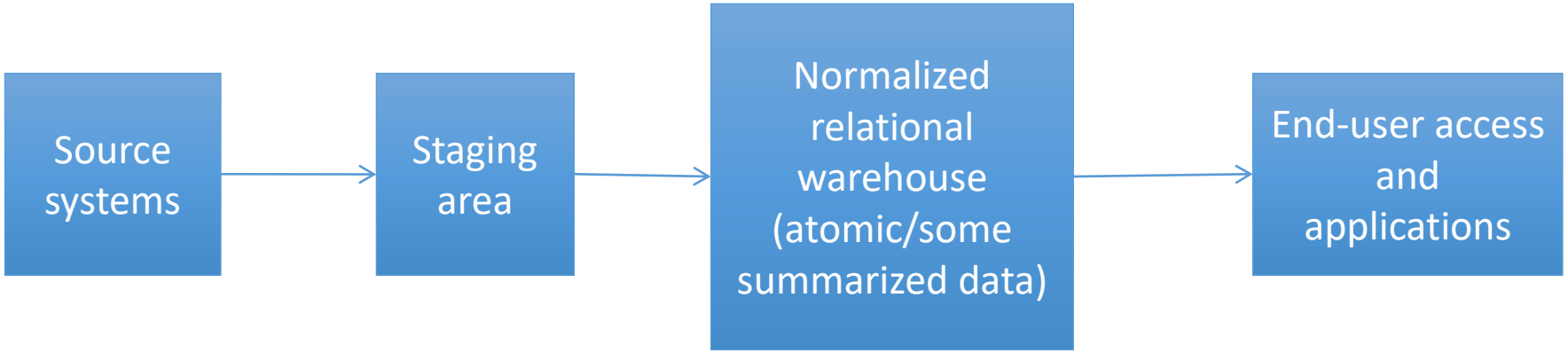
Hub-and-Spoke Diagram



Centralized Data Warehouse

- Similar to the hub-and-spoke architecture except there are no dependent data marts
- Contains atomic-level data, some summarized data, and logical dimensional view of the data
- Queries and applications access data

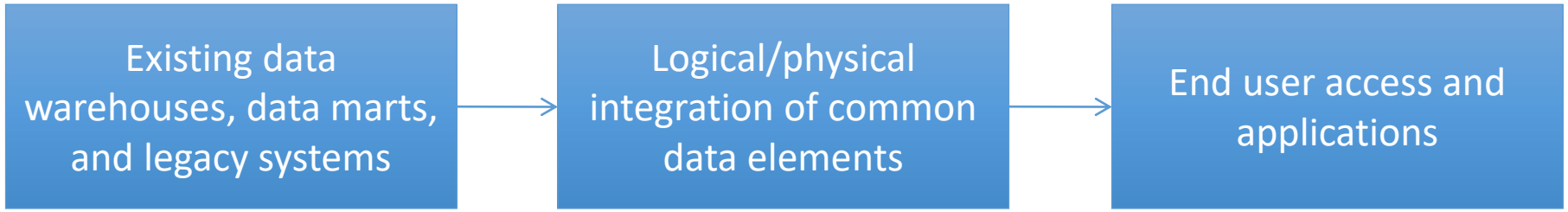
Centralized Data Warehouse Diagram



Federated Architecture

- Leaves existing decision-support structures in place
- Shares information among a number of different systems
- Data is either logically or physically integrated
  - Shared keys
  - Global metadata
  - Distributed queries

Federated Architecture Diagram



Factors That Affect Choosing A Data Warehouse Architecture

- Information Interdependence between Organizational Units
- Upper Management’s Information Needs
- Urgency of Need for a Data Warehouse
- Nature of End-User Tasks
- Constraints on Resources

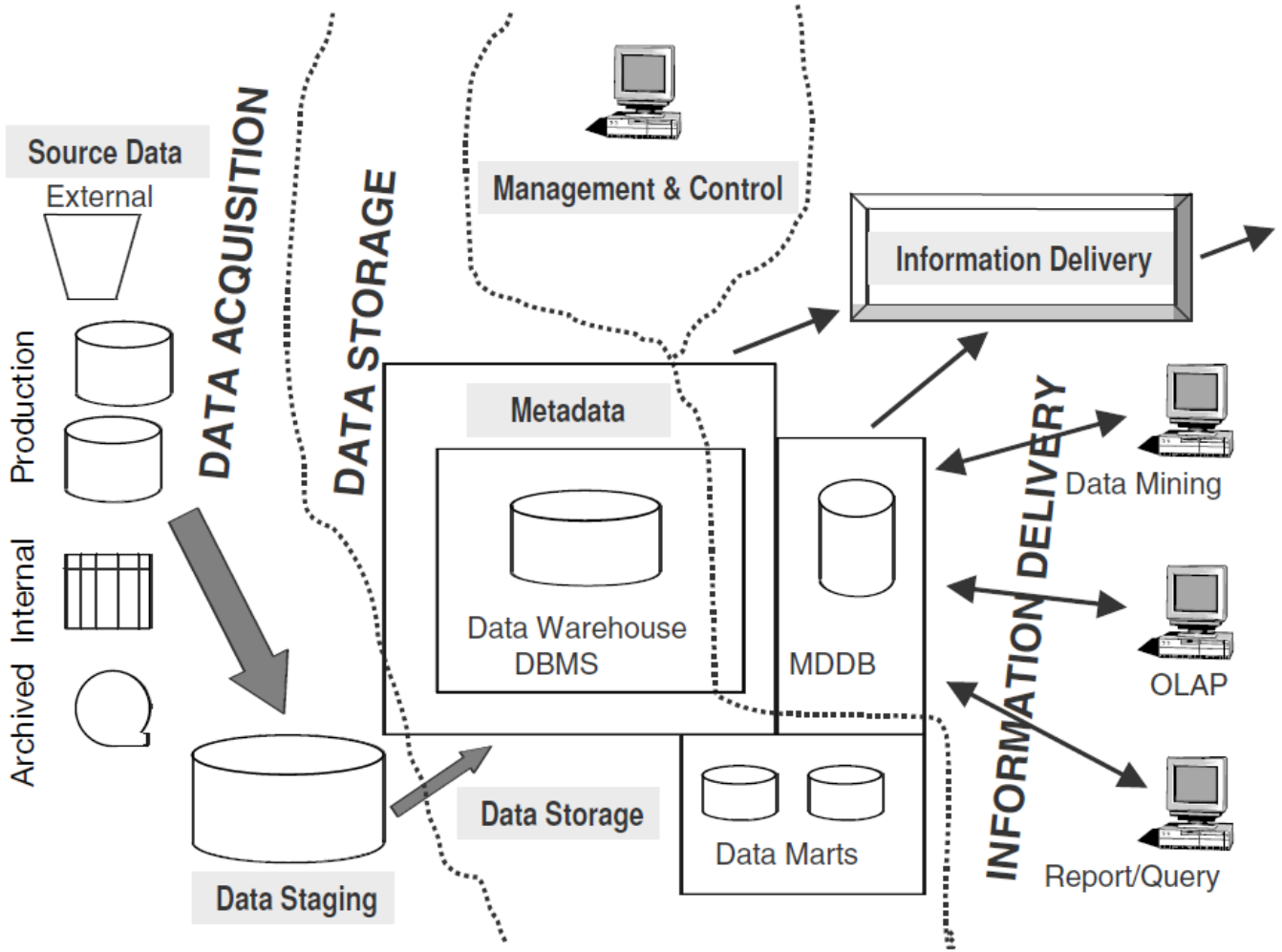
Factors That Affect Choosing A Data Warehouse Architecture

- Strategic View of the Data Warehouse Prior to Implementation
- Compatibility with Existing Systems
- Perceived Ability of the In-House IT Staff
- Technical Issues
- Social/Political Factors

Architecture in the three major areas

- Data acquisition
- Data storage
- Information delivery

Architectural Components in the three Major Areas



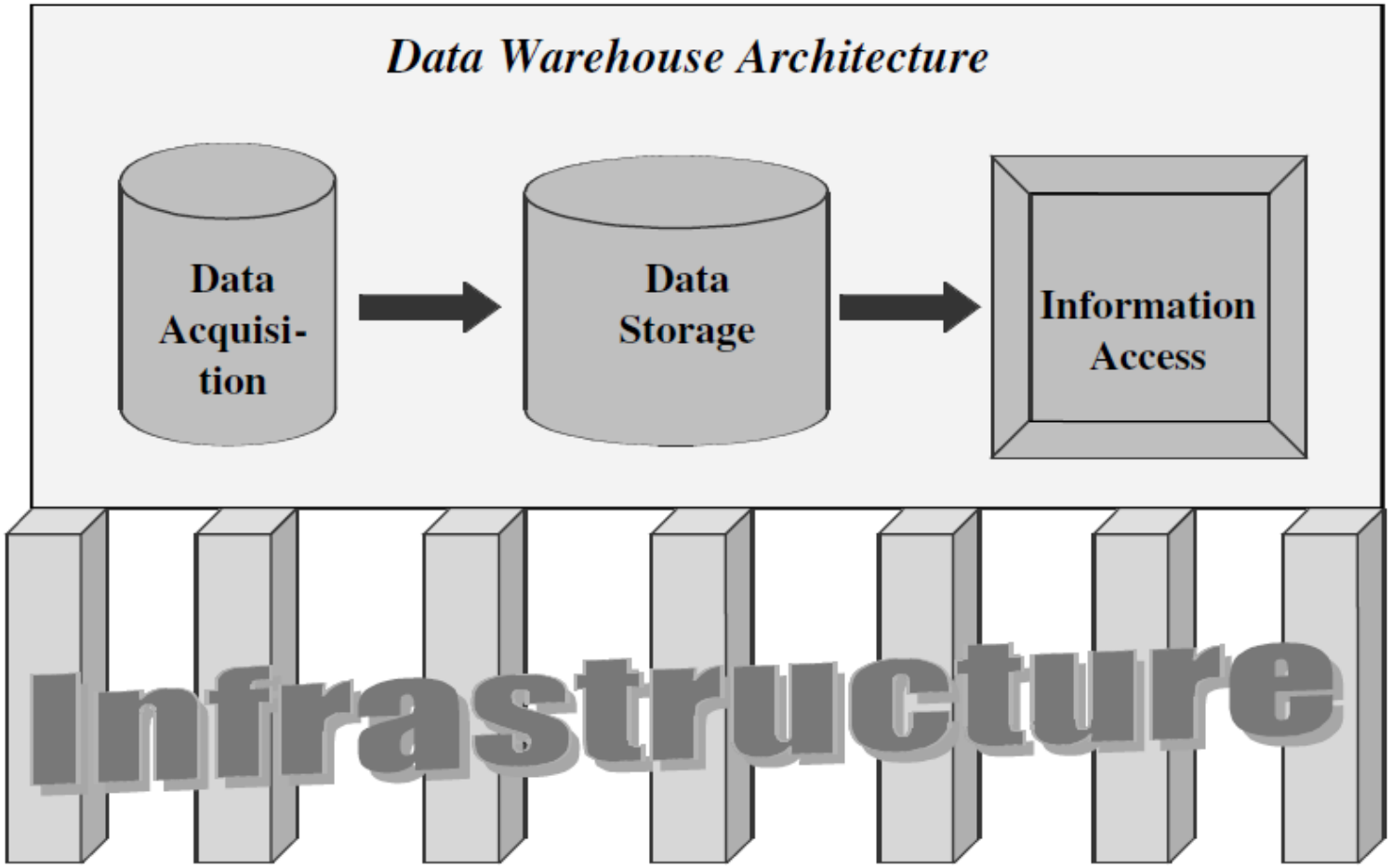
Technical Architecture

- The technical architecture of a data warehouse is the complete set of functions and services provided within its components.
- The technical architecture also includes the procedures and rules that are required to perform the functions and provide the services.
- The technical architecture also encompasses the data stores needed for each component to provide the services.
- Note:
  - *The architecture is not the set of tools needed to perform the functions and provide the services.*
  - *Tools are the means to implement the architecture.*

Data Warehouse Infrastructure

- Infrastructure is the foundation supporting the architecture.
- The foundational infrastructure includes many elements.
- The infrastructure includes:
  - Computing platform
    - Hardware, operating system, network software, LAN, WAN
  - Database management system
  - People
  - Procedure

Infrastructure supporting Architecture



Classification of Data Warehouse Infrastructure

- Operational Infrastructure
- Physical Infrastructure

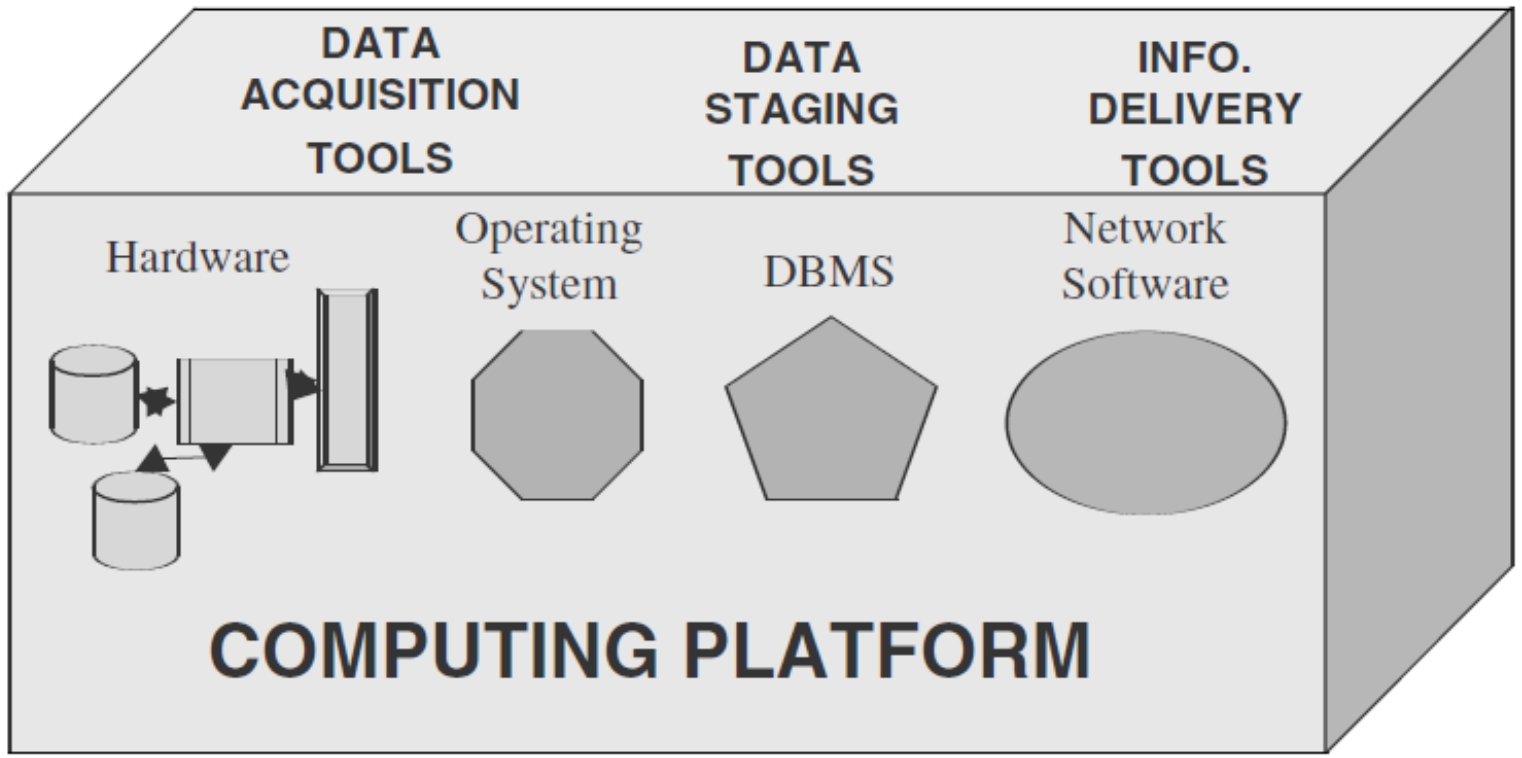
Operational Infrastructure

- People
- Procedures
- Training
- Management software
- These support the management of the data warehouse and maintain its efficiency.
- Even though you may have the right hardware and software, your data warehouse needs the operational infrastructure in place for proper functioning.

Physical Infrastructure

- Computing platform
  - Hardware
  - System software
  - Network
  - Network software
- Database Management System

`


DATA ACQUISITION TOOLS
DATA STAGING TOOLS
INFO. DELIVERY TOOLS
Hardware
Operating System
DBMS
Network Software
COMPUTING PLATFORM