MODULE 3

3.2 DATA WAREHOUSING

Real-Time Data Warehousing

- Real-time data warehousing integrates **real-time data** with **decision support systems (DSS)** and **business intelligence (BI)** tools to enhance business processes.
- It enables organizations to monitor the **current state** of business, quickly identify issues, and provide a foundation for solving them analytically.

Importance

- Combines data from multiple sources for analysis.
- Facilitates rapid decision-making by providing a **single version of truth** that is reliable, consistent, and timely.
- Supports innovative processes by ensuring decision-makers have access to updated and digestible information.

Data Warehouse (DW)

A **data warehouse** is a centralized repository of **current and historical data** structured to support decision-making.

Characteristics:

1. **Subject-Oriented**: Focused on specific business areas like sales, finance, or marketing.
2. **Integrated**: Data is unified from various sources, ensuring consistency.
3. **Time-Variant**: Tracks historical changes and trends over time.
4. **Non-Volatile**: Once data is entered, it remains unchanged for consistency.

Functions

- Enables **online analytical processing (OLAP)**, data mining, querying, and reporting.

- Provides essential tools for activities like **trend analysis, forecasting, and decision support.**

By centralizing fragmented data, data warehousing ensures that decision-makers access complete and dependable information necessary for both strategic and operational purposes.

Characteristics of Data Warehousing

The fundamental characteristics of data warehousing, as described by Inmon (2005), include:

1. Subject-Oriented

- Data is organized by subjects, such as sales, products, or customers, focusing on information relevant for decision support.
- Enables users to assess **how** and **why** their business is performing in a certain way.
- **Difference from Operational Databases**:
    - Operational databases are product-oriented and optimized for transaction updates.
    - Data warehouses provide a comprehensive view for analysis.

2. Integrated

- Combines data from multiple sources into a consistent format.
- Resolves naming conflicts and discrepancies (e.g., units of measurement).
- A data warehouse ensures **total integration** for uniformity.

3. Time-Variant (Time Series)

- Maintains **historical data** rather than just current status (except in real-time systems).
- Supports detection of trends, deviations, and long-term relationships for **forecasting** and **comparisons.**
- Includes multiple time points, such as daily, weekly, or monthly views.

4. Non-Volatile

- Once entered, data cannot be changed or updated.
- Updates are logged as **new data**; obsolete data is discarded.

- Enables optimization for data access rather than frequent modifications.

Additional Characteristics:

1. **Web-Based**: Designed for efficient computing environments for Web-based applications.
2. **Relational/Multidimensional**: Utilizes relational or multidimensional structures for organizing data.
3. **Client/Server**: Employs a client/server architecture to facilitate easy access for end users.
4. **Real-Time**: Supports active, real-time data access and analysis capabilities.
5. **Metadata-Inclusive**: Contains metadata (data about data), detailing how data is organized and how to use it effectively.

Types of Data Warehouses

1. **Data Marts**:
    - Focused on specific departments or business areas.
    - Smaller and more tailored than enterprise-wide data warehouses.
2. **Operational Data Stores (ODS)**:
    - Acts as an intermediary between operational databases and the data warehouse.
    - Stores short-term data for daily operations and reporting.
3. **Enterprise Data Warehouses (EDW)**:
    - Central repository for all data across an organization.
    - Provides a unified platform for decision support and analysis.

Data Warehousing Process

- **Definition**: Data warehousing refers to the entire process of storing, organizing, and managing data for analytical purposes.
- **Purpose**:
    - Enables decision support.
    - Provides easy access to business information.
    - Facilitates business insights.

This holistic approach ensures businesses have a robust foundation for analytical processing and informed decision-making.

Data Marts

- **Definition**: A **data mart** is a smaller, more focused subset of a data warehouse. It typically addresses a particular **subject area** (e.g., marketing, operations) or **department**.
- **Types**:
    1. **Dependent Data Marts**:
        - Created directly from a data warehouse.
        - Ensures consistent data models and high-quality data.
        - Promotes a single enterprise-wide data model, requiring the data warehouse to be built first.
    2. **Independent Data Marts**:
        - Stand-alone small data warehouses designed for specific business units or departments.
        - Does not rely on an enterprise data warehouse (EDW).
        - Cost-effective for organizations with limited budgets.

Operational Data Stores (ODS)

- **Definition**: A database that serves as an **interim staging area** for a data warehouse, providing a near-real-time view of current operational data.
- **Features**:
    - Continuously updated throughout business operations (unlike static data warehouses).
    - Supports **short-term decisions** for mission-critical applications.
    - Similar to **short-term memory** compared to a data warehouse's **long-term memory**.
    - Consolidates data from multiple sources for a near-real-time integrated view.
- **Use Cases**:
    - Facilitates operational decision-making for daily activities.
    - **Oper Marts**: Created when operational data from an ODS is analyzed multidimensionally.

Enterprise Data Warehouses (EDW)

- **Definition**: A large-scale data warehouse utilized across an entire organization for **decision support**.
- **Features**:
    - Integrates data from multiple sources into a standardized format.
    - Serves as a unified platform for a variety of decision support systems (DSS), such as:
        - **Customer Relationship Management (CRM)**
        - **Supply Chain Management (SCM)**
        - **Business Performance Management (BPM)**
        - **Product Lifecycle Management (PLM)**

        - **Revenue Management**
        - **Knowledge Management Systems (KMS)**
- **Examples**:
    - Used by Isle of Capri for enterprise-wide analytics and decision-making.

Metadata

- **Definition**: Metadata refers to "data about data," providing **context** and **structure** for effective data usage.
- **Types**:
    1. **Technical Metadata**: Details about the syntax and structure of the data.
    2. **Business Metadata**: Provides additional context for understanding structured data.
- **Categories**:
    - **Syntactic Metadata**: Describes data syntax.
    - **Structural Metadata**: Describes the organization and relationships of data.
    - **Semantic Metadata**: Explains the meaning of data in a specific context.

Key Insights into Metadata Management

1. **Purpose**:
    - Enhances data understanding and provides context for decision-making.
    - Facilitates the conversion of data into actionable knowledge.
2. **Implementation**:
    - Effective metadata management involves:
        - Ontologies and metadata registries.
        - Enterprise Information Integration (EII).

        - Extraction, Transformation, and Load (ETL) processes.
        - Service-Oriented Architectures (SOA).

3. **Requirements**:
    - Must be extensible, interoperable, efficient, and flexible.
    - Should support evolution, versioning, and low maintenance costs.
4. **Maturity Levels** (Zhao, 2005):
    - **Ad Hoc**: Initial, unstructured approach.
    - **Discovered**: Metadata begins to be recognized.
    - **Managed**: Metadata is structured and organized.
    - **Optimized**: Comprehensive management and usage.
    - **Automated**: Fully automated processes for metadata handling.

Ethical Considerations in Metadata

- Includes issues around **privacy** and **intellectual property** in the design, collection, and usage stages.
- Ethical metadata management ensures transparency and protects data ownership rights.

These components of data warehousing and metadata strategies collectively empower organizations to leverage data for informed decision-making while maintaining ethical standards.

3.3 DATA WAREHOUSING PROCESS OVERVIEW

Organizations continuously collect, store, and manage vast amounts of data, leading to challenges in maintaining, accessing, and utilizing this data effectively. The **data warehousing process** ensures data is structured and accessible for decision-making.

Key Components of the Data Warehousing Process

1. **Data Sources**:
    - Data is collected from multiple sources, including:
        - Operational legacy systems.
        - External providers (e.g., U.S. Census).

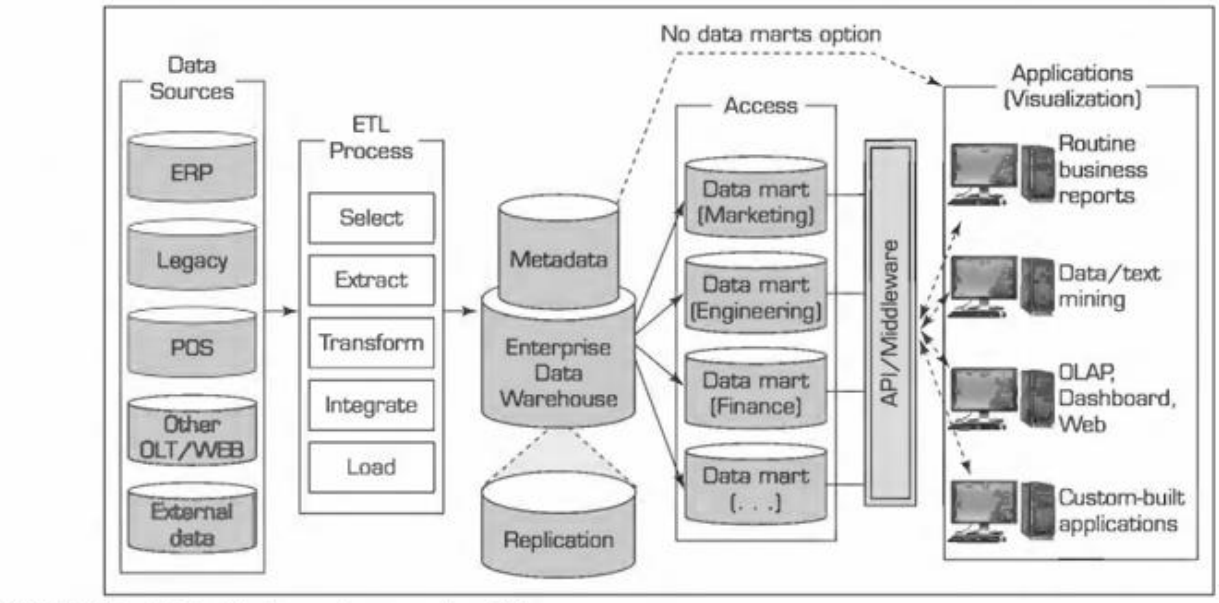


FIGURE 3.3 A Data Warehouse Framework and Views.

        - Online Transaction Processing (OLTP) or Enterprise Resource Planning (ERP) systems.
        - Web logs for online data.

---

2. **Data Extraction and Transformation (ETL)**:
    - **ETL Tools**: Data is extracted and transformed using either custom-built or commercial software.
    - **Purpose**: To standardize, cleanse, and reformat data for alignment with the organization's needs.

---

3. **Data Loading**:
    - Data is initially loaded into a **staging area** where it is cleansed and prepared for integration.
    - Post-transformation, data is loaded into the **data warehouse** or **data marts** for use.

---

4. **Comprehensive Database**:
    - The **Enterprise Data Warehouse (EDW)** acts as the centralized repository.
    - Supports decision-making by providing both summarized and detailed information from various sources.

---

5. **Metadata**:
    - **Definition**: Metadata refers to data about the structure, organization, and meaning of stored data.
    - **Uses**:
        - Assists IT personnel and end-users in understanding data.
        - Enables efficient indexing and searching, often integrated with web-based tools.

---

6. **Middleware Tools**:
    - Enable users to access and interact with the data warehouse through tools such as:
        - **SQL Queries**: For advanced users like analysts.
        - **Managed Query Environments**: Examples include Business Objects.
    - Support business users with applications for:
        - **Data Mining**

- **Online Analytical Processing (OLAP)**
- **Reporting**
- **Data Visualization**

Process Variants

- **Top-Down Approach**: Build the EDW first, then load data marts for specific departments.
- **Bottom-Up Approach**: Start with individual data marts that are later integrated into the EDW.
- **Alternative Approaches**: Some organizations skip creating data marts, opting to use BI tools for direct data manipulation.

This structured process ensures that raw data is transformed into actionable insights for strategic and operational decision-making.

3.4 DATA WAREHOUSING ARCHITECTURES

Data warehousing architectures refer to the structural frameworks used to organize, store, and process data for decision support. Different architectures cater to varying organizational needs, ranging from cost efficiency to scalability and holistic enterprise views.

---

1. Multi-Tier Architectures

- **Definition**: Multi-tier architectures are categorized as **client/server or n-tier architectures**. Among these, **two-tier** and **three-tier architectures** are the most commonly implemented for data warehousing.

Three-Tier Architecture:

- Divides the data warehouse into three parts:
    1. **Operational Systems**: Contain the data and backend software for data acquisition.

2. **Data Warehouse**: Acts as a central repository of data, storing both detailed and summarized information for analysis.
3. **DSS/BI/BA Engine**: Includes the application server and client software used for analysis and decision support.

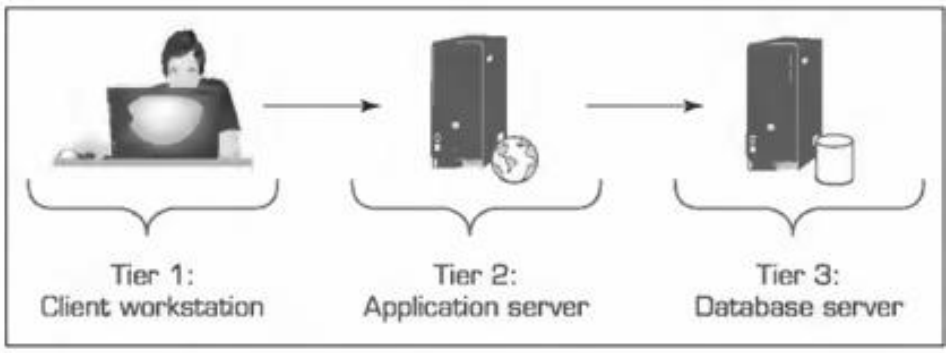


FIGURE 3.4 Architecture of a Three-Tier Data Warehouse.

- Data from the warehouse is often processed twice and organized in a multidimensional database for easy analysis or replicated in data marts.
- **Advantages**:
  - Separates functions for scalability.
  - Eliminates resource constraints.
  - Simplifies the creation of data marts.

Two-Tier Architecture:

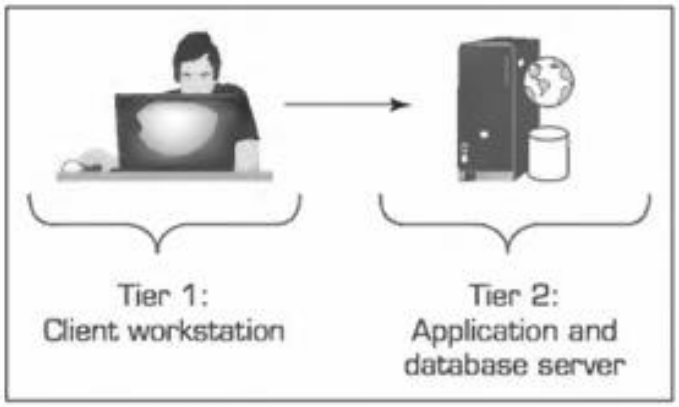


FIGURE 3.5 Architecture of a Two-Tier Data Warehouse.

- The DSS engine operates on the same hardware platform as the data warehouse.

- **Advantages**:
    - More economical than a three-tier system.
- **Disadvantages**:
    - May face performance issues when handling large-scale data warehouses or data-intensive applications.

---

2. Web-Based Data Warehousing

- **Definition**: Integrates data warehousing with Internet technologies, enabling users to access the warehouse via the web.
- **Structure**:
    1. **PC Client**: Requires a browser and an Internet connection.
    2. **Web Server**: Manages the flow of data between the client and the backend servers.
    3. **Application Server and Data Warehouse**: Store and process the data.
- **Communication Medium**: Typically uses the Internet, intranet, or extranet for connectivity between clients and servers.

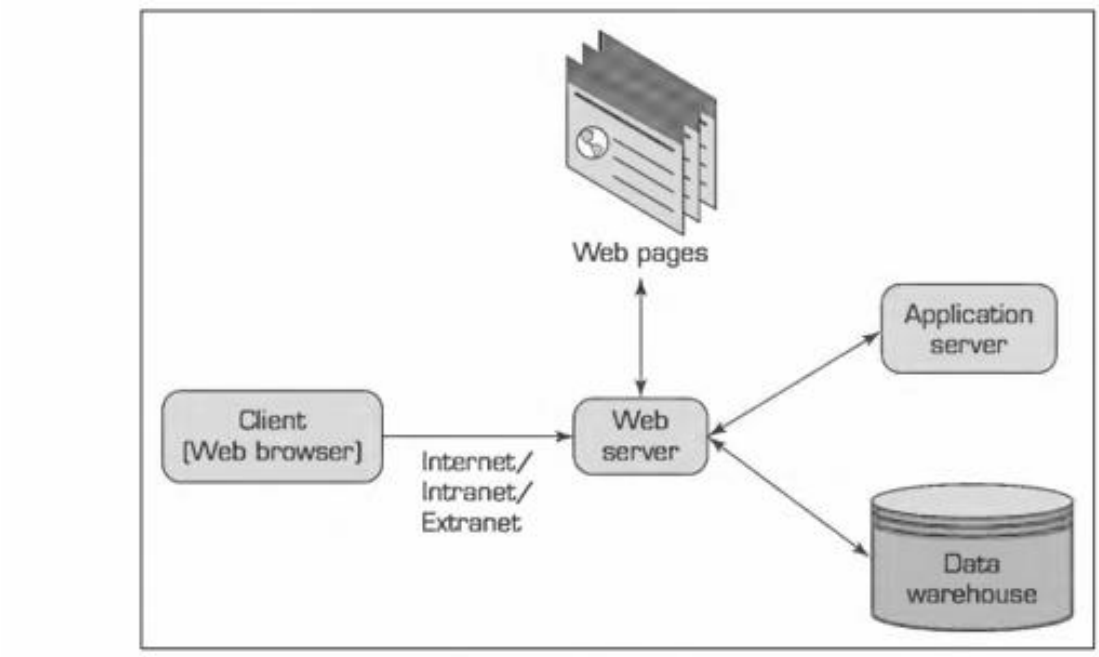


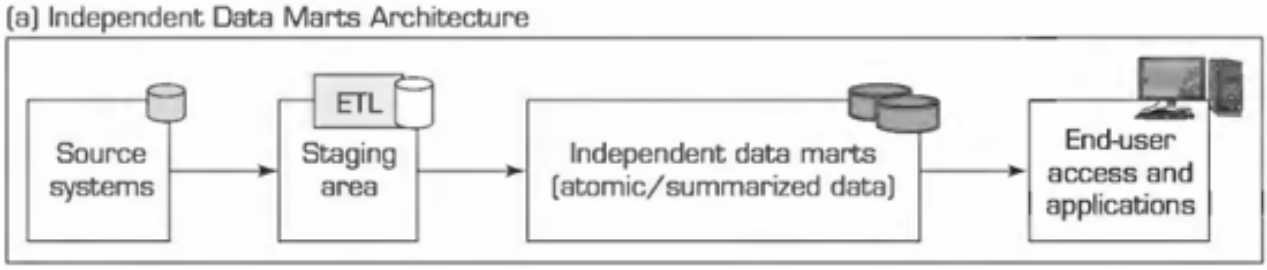
FIGURE 3.6 Architecture of Web-Based Data Warehousing.

- **Advantages**:
    - Simplified access using familiar web interfaces.
    - Platform independence across systems.
    - Lower operational and implementation costs.
- **Examples**:
    - The Vanguard Group moved to a three-tier web-based architecture to integrate data and provide customers and internal users with the same data views.
    - Hilton implemented a similar system, investing $3.8 million to integrate its data warehouse, which increased efficiency sixfold and provided annual savings of $4.5 to $5 million.

---

3. Alternative Data Warehousing Architectures

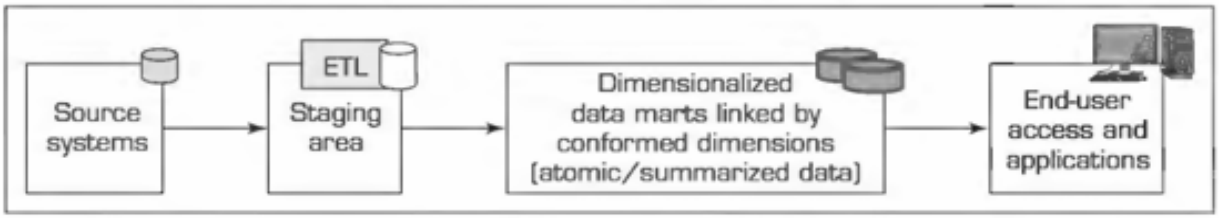
1. **Independent Data Marts**:
    - Stand-alone data marts are developed independently for specific organizational units or departments.



    - **Advantages**:
        - Simpler and less expensive compared to enterprise-wide solutions.
    - **Disadvantages**:
        - Inconsistent data definitions.
        - Different dimensions and measures across marts can make enterprise-wide analysis difficult or impossible.

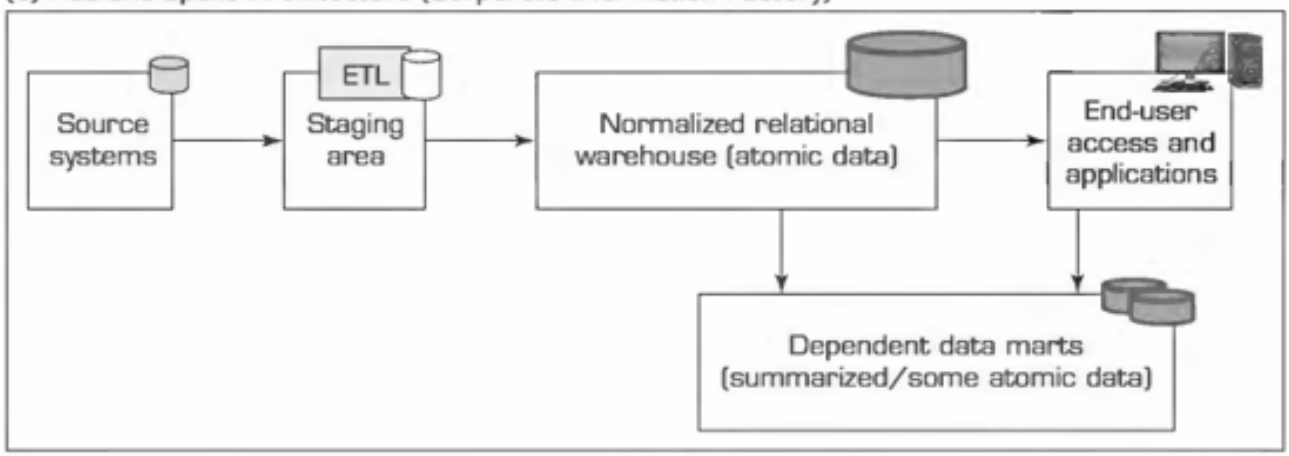
2. **Data Mart Bus Architecture**:
    - Links individual data marts using middleware to ensure better integration and consistency.

(b) Data Mart Bus Architecture with Linked Dimensional Data Marts



    - **Advantages**:
        - Provides metadata-level consistency across marts.
        - Allows for complex queries across data marts.
    - **Disadvantages**:
        - Query performance may not be optimal for large-scale or enterprise-wide data analysis.
3. **Hub-and-Spoke Architecture**:
    - Focuses on building a centralized data warehouse with dependent data marts for specific organizational units.

(c) Hub-and-Spoke Architecture (Corporate Information Factory)

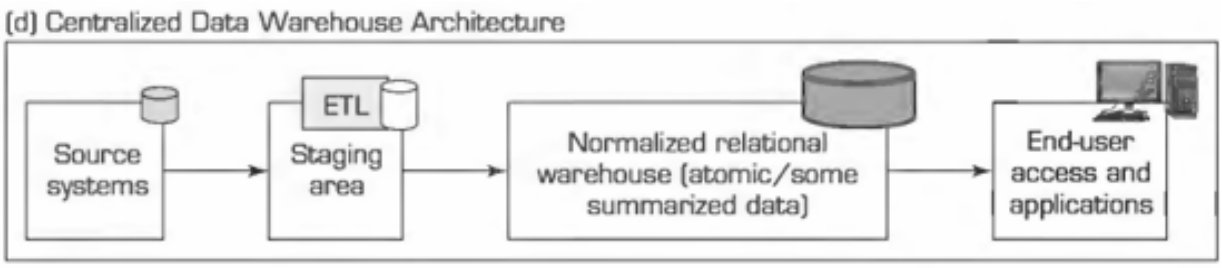


    - **Advantages**:
        - Scalable and maintainable infrastructure.

- Allows for easy customization of user interfaces and reports.

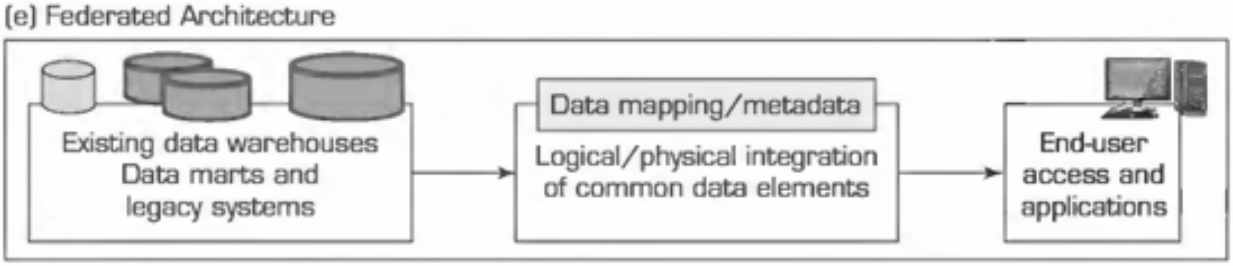
- **Disadvantages**:
  - May lead to data redundancy and latency.
  - Lacks a unified, enterprise-wide view.

4. **Centralized Data Warehouse**:
   - A single, large-scale enterprise data warehouse is used to serve the needs of all organizational units, eliminating the need for dependent data marts.



   - **Advantages**:
     - Provides a holistic view of the enterprise.
     - Simplifies data management and minimizes data redundancy.
   - **Disadvantages**:
     - High complexity and cost of design and implementation.
   - **Example**: Teradata advocates for this architecture to centralize all organizational data.

5. **Federated Data Warehouse**:
   - Integrates disparate systems and resources to meet changing business needs.



- **Advantages**:
  - Adaptable to evolving business conditions without replacing existing systems.
- **Disadvantages**:
  - Issues with data quality and query performance.
  - Best suited as a supplement to traditional data warehouses rather than a replacement.

Key Considerations for Choosing an Architecture

1. **Database Management System (DBMS)**:
   - Relational database systems like Oracle, SQL Server, and IBM DB2 are commonly used.
   - Must support both client/server and web-based architectures for flexibility.
2. **Parallel Processing and Partitioning**:
   - Parallel processing allows multiple CPUs to execute queries simultaneously, improving performance.
   - Partitioning divides large tables into smaller, more manageable pieces for efficient access.
3. **Data Migration Tools**:
   - Migration processes can range from simple to highly complex, depending on the diversity and location of existing data assets.
   - Tools may be developed in-house, purchased from third-party vendors, or included as part of the data warehouse system.
4. **Data Retrieval and Analysis Tools**:
   - Specialized tools are often required for data extraction, transformation, loading, and analysis.

- Real-time data migration or analysis may warrant the use of advanced third-party ETL tools.

Which Architecture Is the Best?

1. **Two Main Approaches**:
    - **Bill Inmon's Hub-and-Spoke Architecture**:
        - Centralized warehouse with dependent data marts.
        - Best for **enterprise-wide systems**.
        - Handles **large-scale data** and **complex requirements**.
        - **Expensive and time-consuming** to implement.
    - **Ralph Kimball's Data Mart Bus Architecture**:
        - Linked independent data marts with **shared dimensions**.
        - Focused on **specific departments**.
        - Simpler and less costly compared to hub-and-spoke.

---

2. **Survey Results (Ariyachandra and Watson, 2006b)**:
    - **Participants**: 454 companies (60% from the U.S.).
    - **Adoption Rates**:
        - **Hub-and-Spoke**: 39% (most widely used).
        - **Data Mart Bus**: 26%.
        - **Centralized Architecture**: 17%.
        - **Independent Data Marts**: 12%.
        - **Federated Architecture**: 4%.

---

3. **Success Scores** (7-point scale; higher is better):
    - **Independent Data Marts**: Scored the lowest across all metrics (not recommended).
    - **Federated Architecture**: Performed poorly; suitable only for short-term use.
    - **Hub-and-Spoke, Data Mart Bus, and Centralized Architectures**:
        - Scored **similarly high** on all success metrics:
            1. **Information Quality**.
            2. **System Quality**.
            3. **Individual Impacts**.
            4. **Organizational Impacts**.

---

4. **Key Characteristics**:
    - **Hub-and-Spoke**:
        - Used for **large, enterprise-wide implementations**.
        - Best for **scalability and integration**.
    - **Data Mart Bus**:
        - Focused on **department-specific needs**.
        - Easier to implement than hub-and-spoke.
    - **Centralized Architecture**:
        - Combines all data into one system.
        - Offers a **holistic enterprise-wide view**.
    - **Independent Data Marts**:
        - Simple and low-cost but lacks consistency and scalability.
    - **Federated Architecture**:
        - Integrates multiple systems temporarily; not ideal for long-term use.

5. **Conclusion**:
    - There is **no single best architecture**; the choice depends on the organization's needs:
        - Choose **hub-and-spoke** for **large-scale, enterprise-wide data systems**.
        - Use **data mart bus** for a **department-focused, cost-effective approach**.
        - Opt for **centralized architecture** if a **unified, organization-wide view** is the priority.

3.5 DATA INTEGRATION AND THE EXTRACTION, TRANSFORMATION, AND LOAD (Ell) PROCESSES

Data Integration

1. **Definition**:
    - Integrates data from multiple, disparate sources for unified analysis and storage.
    - Ensures seamless access to data for **ETL tools**, **data warehouses**, and **BI platforms**.
2. **Key Components of Data Integration**:
    - **Data Access**: Enables extraction from any data source.
    - **Data Federation**: Merges business views across multiple datasets into a unified structure.
    - **Change Capture**: Detects, captures, and delivers updates made to enterprise data sources.
3. **Technologies for Data Integration**:
    - **Enterprise Application Integration (EAI)**:

        - Focuses on reusing application functionality across systems using APIs or SOA.
        - Facilitates real-time or near-real-time data acquisition into data warehouses.
    - **Enterprise Information Integration (EII)**:
        - Pulls data from sources like relational databases, multidimensional databases, and web services.
        - Uses **metadata** to present a relational view of integrated data.
        - Leverages **XML tags** for data identification and adaptability.
    - **ETL (Extraction, Transformation, Load)**:
        - Core process for moving and cleaning data for use in warehouses or databases.
4. **Challenges in Data Integration**:
    - **Complexity**: Growing data warehouses increase integration complexity.
    - **Regulatory and Business Changes**: Evolving BI needs, mergers, and compliance introduce integration demands.
    - **Technical Issues**: Requires expertise to ensure compatibility and prevent system disruptions.

Extraction, Transformation, and Load (ETL)

1. **Definition**:
    - A process for moving data from source systems to a data warehouse by **extracting**, **transforming**, and **loading** it into the target system.
2. **Components of ETL**:
    - **Extraction**:
        - Reads data from various sources like databases, files, ERP systems, or CRM tools.

- Sources may include **OLTP databases**, flat files, Excel sheets, or message queues.

- **Transformation**:
  - Cleans, standardizes, and converts data into a format suitable for analysis.
  - Uses **business rules**, lookup tables, and data combinations.
  - Detects and removes anomalies in data to ensure consistency.
- **Load**:
  - Places transformed data into the target data warehouse or database.
  - Utilizes staging tables to facilitate and validate the loading process.

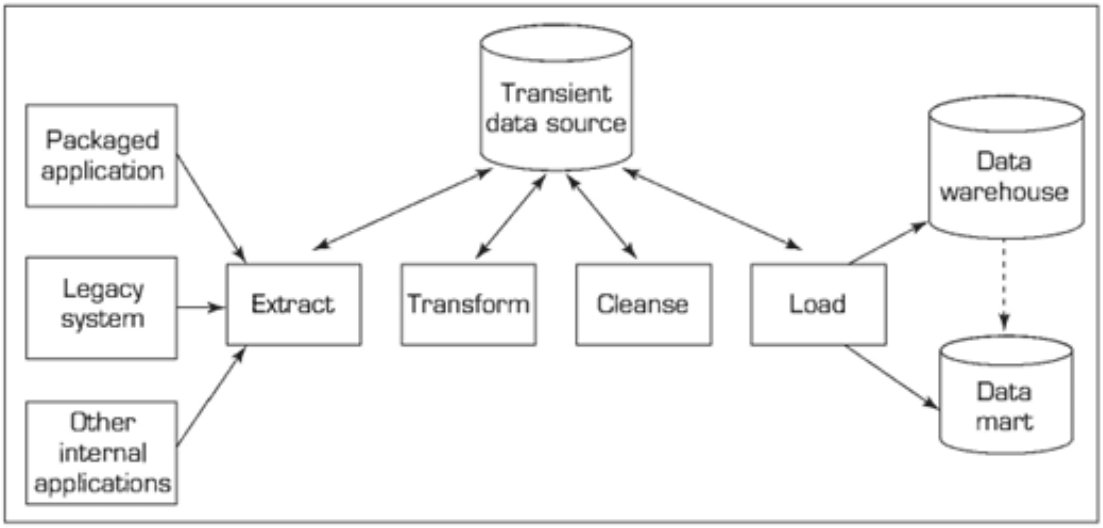


FIGURE 3.9 The ETL Process.

3. **Importance**:
   - **70% of time** in data-centric projects is spent on ETL processes.
   - Ensures that warehouses have **clean, reliable, and integrated data** for analysis.
4. **Functions of ETL Tools**:
   - Transport data between source and target systems.
   - Document and track metadata changes.

- Handle scheduling, error management, audit logs, and runtime statistics.
- Support data cleansing and scrubbing for accuracy and reliability.

5. **Advanced Features in ETL**:
   - **Metadata Integration**: Captures and uses metadata across systems.
   - **Standardization**: Conforms to open standards to ensure compatibility.
   - **Graphical User Interfaces (GUI)**: Simplifies rule development and management.
   - **Anomaly Detection**: Identifies redundant or inconsistent data.
6. **Criteria for Selecting ETL Tools**:
   - Ability to interact with diverse data source architectures.
   - Automatic metadata management and delivery.
   - Easy-to-use interface for developers and functional users.
   - Conformance to industry standards and flexibility for business needs.
7. **Examples of ETL in Action**:
   - **Motorola**:
     - Collects data from 30 procurement systems.
     - Uses ETL to load and analyze data in its global supply chain management (SCM) warehouse.
   - Vendors:
     - Notable ETL providers include **SAS, Microsoft, Oracle, IBM, Informatica, Tibco**, and Embarcadero.
8. **Challenges in ETL**:
   - **Costs**:
     - Tools are often expensive.
     - High initial investment and long learning curves.

- **Poor Data Management**:
  - Extensive ETL processes may indicate redundant or poorly managed data.
  - Overuse increases maintenance costs and reduces efficiency.
- **Tool Selection**:
  - Deciding between building custom ETL tools or purchasing off-the-shelf tools can be difficult.

9. **ETL as a Long-Term Investment**:
   - **Benefits**:
     - Simplifies maintenance.
     - Improves data quality for OLAP and data mining applications.
     - Reduces redundant data and streamlines processing.
   - **Strategic Impact**:
     - Properly designed ETL processes ensure enterprise-wide data consistency.
     - Well-managed ETL contributes to faster development, lower costs, and better data quality.
10. **Classification of ETL Tools (Solomon, 2005)**:
    - **Sophisticated**: Offers comprehensive documentation and accurate data management.
    - **Enabler**: Provides basic ETL functionality for mid-sized projects.
    - **Simple**: Handles minimal transformations and data movement.
    - **Rudimentary**: Limited to basic ETL tasks; often insufficient for large projects.
11. **Conclusion on ETL**:
    - A robust ETL strategy is essential for data warehousing success.

- It minimizes errors, ensures consistency, and prepares data for advanced BI tools.
- Poor ETL processes lead to higher costs, inefficiencies, and reduced data quality, making the choice of tools and methodologies critical.

Data Warehouse Development (Updated)

What is Data Warehouse Development?

- A major, complex project involving multiple departments, systems, and strategies.
- Influences organizational decision-making and CRM strategies.
- Helps store, manage, and analyze vast amounts of organizational data.

Benefits of Data Warehouses

1. **Direct Benefits**:
   - Users can perform **extensive analyses** in multiple ways.
   - Offers a **single version of truth** through a consolidated view of corporate data.
   - **Timely Information**:
     - Processing shifts from costly operational systems to low-cost servers.
     - Handles a higher volume of end-user requests quickly.
   - **Improved System Performance**:
     - Offloads reporting needs from operational systems to decision support systems (DSS).
   - **Simplified Access**: Data is easier to retrieve and use for decision-making.
2. **Indirect Benefits**:
   - Enhances **business knowledge** and provides a **competitive advantage**.

- Improves **customer service** and **satisfaction**.
- Facilitates better and faster decision-making.
- Helps reform business processes for efficiency and effectiveness.

Critical Success Factors

- Clearly define the **business objective** for the warehouse.
- Gather **support from management and end-users**.
- Set **reasonable timeframes and budgets**.
- **Manage expectations** to align with realistic goals and outcomes.

Return on Investment (ROI) for Data Warehouses

- Benefits categorized by Kelly (2001):
  - **Keepers** (20%): Cost savings from improved traditional decision-making functions.
  - **Gatherers** (30%): Savings from automated data collection and distribution.
  - **Users** (50%): Highest returns from decisions made using warehouse data.
- Costs include:
  - Hardware, software, network bandwidth, internal support, training, and external consulting.
- ROI is evaluated using **Net Present Value (NPV)** over the expected warehouse lifecycle.

Development Approaches

1. **Inmon's EDW Approach (Top-Down)**:
   - Focuses on building an **Enterprise Data Warehouse (EDW)** first.

    - Uses structured methodologies like **entity-relationship diagrams (ERD)**.
    - Allows for a **comprehensive and consistent view** of enterprise data.
    - Data marts may be added later as needed.
2. **Kimball's Data Mart Approach (Bottom-Up)**:
    - Builds **data marts** for specific departments (e.g., marketing, sales) first.
    - Uses **dimensional modeling** to structure data marts.
    - Gradually expands into a full data warehouse.
    - A **"plan big, build small"** strategy.
3. **Comparison**:
    - Inmon's approach is suitable for **large-scale enterprise-wide projects**.
    - Kimball's model is ideal for **smaller organizations** or when resources are limited.
    - Organizations often start with data marts and evolve toward an EDW.

Hosted Data Warehouses

- Organizations may choose to outsource data warehouse development.
- Hosted warehouses are managed by experienced third-party vendors.
- **Advantages**:
    - Eliminates the need for internal software and hardware management.
- **Concerns**:
    - **Privacy and security risks** associated with external hosting.

Dimensional Modeling in Data Warehouses

1. **Star Schema**:
    - Simplest and most common form of dimensional modeling.

- Features a central **fact table** connected to several **dimension tables**.
- **Relationship**:
  - The relationship between the **fact table and dimension tables is one-to-many**.
- Focuses on **fast query-response time**.
- Fact tables store metrics and performance data (e.g., sales, customer retention rates).
- Dimension tables provide details for slicing and dicing the data.

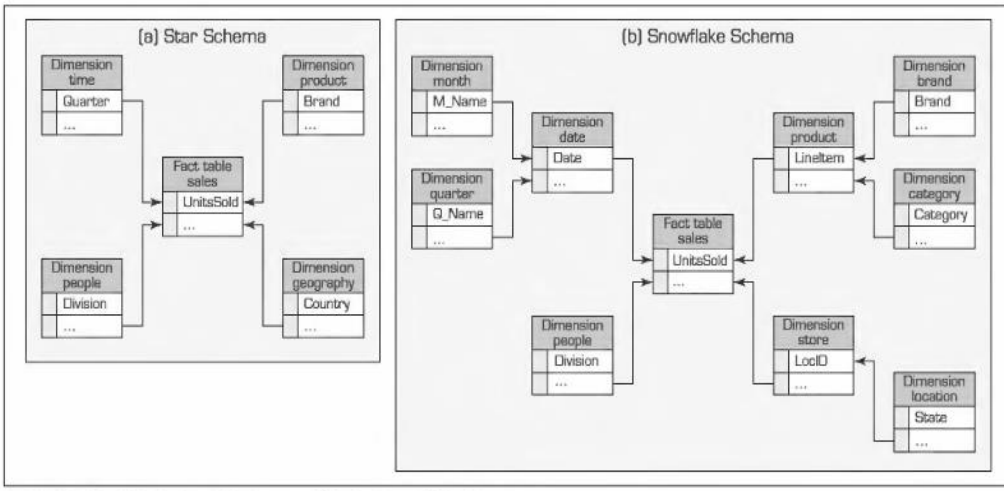


FIGURE 3.10 (a) The Star Schema, and (b) the Snowflake Schema.

2. **Snowflake Schema**:
   - A normalized version of the star schema.
   - Dimension tables are split into multiple related tables for additional structure.
   - Supports more complex data relationships but can slow query performance.
3. **Fact Table Normalization**:
   - Fact tables in star schemas are typically **completely de-normalized**.
   - This improves query performance and simplifies data analysis.

Analysis of Data in Data Warehouses

1. **OLAP (Online Analytical Processing)**:
    - A widely used method for analyzing data in warehouses.
    - Performs **multidimensional analytical queries** for decision support.
    - Enables fast answers to ad hoc questions.
2. **Querying**:
    - Combines data from multiple dimensions to provide insights.
    - Supports **ad hoc querying** for business users.

Why Develop a Data Warehouse?

- Organizes fragmented data into a unified platform.
- Simplifies **decision support** and enhances strategic planning.
- Provides clean, structured data for **advanced analytics**.

OLAP Versus OLTP

1. **OLTP (Online Transaction Processing)**:
    - Focuses on **day-to-day business transactions** (e.g., ERP, CRM, SCM, point-of-sale systems).
    - **Purpose**:
        - Captures and stores real-time data.
        - Runs routine reports and analyses.
    - **Limitations**:
        - Not designed for complex, ad hoc queries or multidimensional analysis.
2. **OLAP (Online Analytical Processing)**:
    - Designed for **ad hoc analysis** of organizational data.
    - **Purpose**:

        - Provides efficient multidimensional analysis.
        - Uses data captured by OLTP to support decision-making.
    - **Relationship**:
        - OLTP automates transactions; OLAP uses OLTP data for deeper insights.
3. **Comparison**:
    - OLTP handles **operational needs**, while OLAP supports **analytical requirements**.

TABLE 3.5 A Comparison Between OLTP and OLAP

Criteria	OLTP	OLAP
Purpose	To carry out day-to-day business functions	To support decision making and provide answers to business and management queries
Data source	Transaction database (a normalized data repository primarily focused on efficiency and consistency)	Data warehouse or data mart (a nonnormalized data repository primarily focused on accuracy and completeness)
Reporting	Routine, periodic, narrowly focused reports	Ad hoc, multidimensional, broadly focused reports and queries
Resource requirements	Ordinary relational databases	Multiprocessor, large-capacity, specialized databases
Execution speed	Fast (recording of business transactions and routine reports)	Slow (resource intensive, complex, large-scale queries)

OLAP Operations

1. **Cube**:
    - Core structure in OLAP.
    - A **multidimensional data structure** for analyzing data quickly and efficiently.
    - Overcomes relational database limitations in handling large, complex queries.
2. **OLAP Functions**:
    - **Slice**: Extracts a subset of data based on one dimension (e.g., sales in a specific region).

- **Dice**: Extracts data from multiple dimensions (e.g., sales by region and time).
- **Drill Down/Up**:
  - Drill Down: Navigates to detailed data.
  - Drill Up: Summarizes data at a higher level.
- **Roll-Up**: Computes data relationships across one or more dimensions.
- **Pivot**: Changes the dimensional orientation of reports or queries.

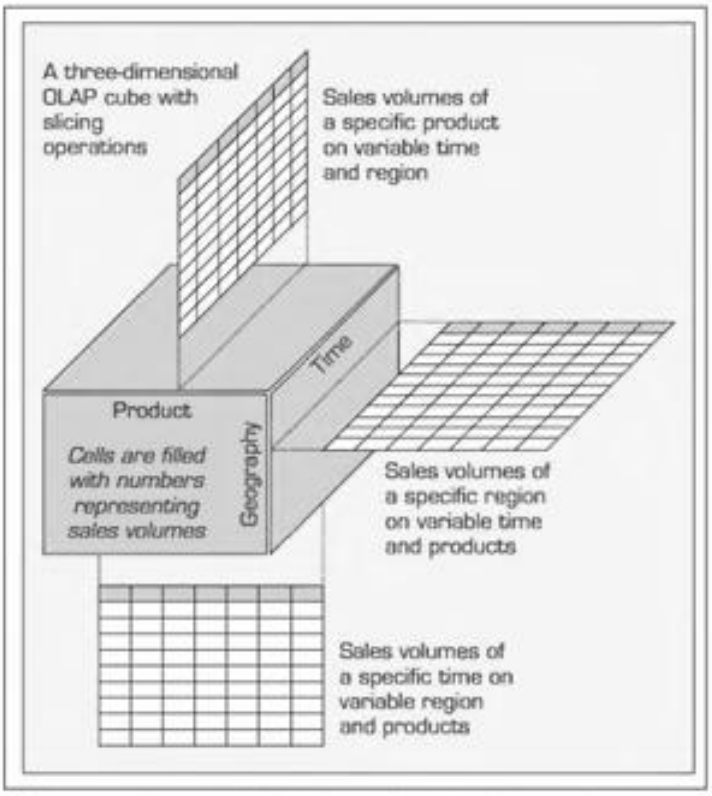


FIGURE 3.11 Slicing Operations on a Simple Three-Dimensional Data Cube.

3. **Variations of OLAP**:
   - **ROLAP (Relational OLAP)**:
     - Uses relational databases and SQL queries to generate results.
     - Requires carefully designed databases but avoids preprocessing.
   - **MOLAP (Multidimensional OLAP)**:
     - Precomputes and stores data in optimized multidimensional cubes.

        - Faster query performance but requires preprocessing.
    - **HOLAP (Hybrid OLAP)**:
        - Combines ROLAP and MOLAP.
        - Stores part of the data in MOLAP and the rest in ROLAP for flexibility.

4.4 DIFFERENT TYPES OF CHARTS AND GRAPHS

Basic Charts and Graphs

1. **Line Chart**:
    - **Structure**:
        - Data points connected by lines.
        - **X-axis**: Represents time or continuous variables.
        - **Y-axis**: Represents the measure being analyzed.
    - **Purpose**:
        - Best for tracking trends and changes over time.
    - **Examples**:
        - Tracking stock price changes over five years.
        - Monitoring daily customer service calls over a month.
2. **Bar Chart**:
    - **Structure**:
        - Bars represent data values; height/length corresponds to magnitude.
        - **X-axis**: Categories.
        - **Y-axis**: Numeric values or frequencies.
    - **Purpose**:
        - Compare data across categories.

- **Variants**:
  - Vertical, horizontal, and stacked bars.
- **Examples**:
  - Advertising spending by departments.
  - Sales by product categories.

3. **Pie Chart**:
   - **Structure**:
     - A circle divided into slices where slice size reflects proportions.
   - **Purpose**:
     - Show relative proportions of a single measure.
   - **Guidelines**:
     - Best for fewer than 4 categories; for more, use bar charts.
   - **Examples**:
     - Percentage of advertising budget spent on product lines.
     - Proportions of declared college majors.
4. **Scatter Plot**:
   - **Structure**:
     - Data points plotted on X and Y axes to represent two variables.
   - **Purpose**:
     - Explore relationships, trends, and outliers.
   - **Enhancements**:
     - Trend lines to indicate correlations.
   - **Examples**:
     - Correlation between age and weight of patients.

        - Relationship between customer service staff and open claims.

5. **Bubble Chart**:
    - **Structure**:
        - Similar to scatter plots but uses bubbles, with size and/or color adding more dimensions.
    - **Purpose**:
        - Enhance scatter plots by representing additional data dimensions.
    - **Examples**:
        - Profit margin by product type and geographic region.
        - Attendance by major and time of day.

Specialized Charts and Graphs

1. **Histogram**:
    - **Structure**:
        - Bars showing frequency distributions.
        - **X-axis**: Data ranges (bins).
        - **Y-axis**: Frequency or count.
    - **Purpose**:
        - Show the shape and distribution of data.
    - **Examples**:
        - Grade distribution in a class.
        - Age distribution of a customer base.
2. **Gantt Chart**:
    - **Structure**:
        - Horizontal bars represent tasks, with start/end dates and durations.

- **Purpose**:
  - Manage project timelines, milestones, and task dependencies.
- **Examples**:
  - Project schedules and task overlaps.
  - Resource allocation.

3. **PERT Chart**:
   - **Structure**:
     - Nodes (tasks/activities) connected by arrows (dependencies).
   - **Purpose**:
     - Simplify planning and scheduling of complex projects.
   - **Examples**:
     - Task sequences and dependencies in a project.
4. **Geographic Map**:
   - **Structure**:
     - Displays location-based data using markers, colors, or regions on a map.
   - **Purpose**:
     - Visualize data with geographic context.
   - **Examples**:
     - Mapping customer service requests by product type.
     - Deciding store locations based on population density.
5. **Bullet Graph**:
   - **Structure**:
     - Horizontal bar with markers to indicate progress toward goals.
   - **Purpose**:

        - Track performance metrics compared to benchmarks.
    - **Examples**:
        - Year-to-date revenue versus annual targets.
        - Sales representatives' progress toward quotas.
6. **Heat Map**:
    - **Structure**:
        - Grid format using color gradients to represent values.
        - **X-axis and Y-axis**: Represent dimensions or categories.
    - **Purpose**:
        - Compare values across two dimensions visually.
    - **Examples**:
        - Segmenting target markets by purchase amounts, age, and income.
        - Visualizing website traffic patterns by time and region.
7. **Highlight Table**:
    - **Structure**:
        - Similar to heat maps but includes numerical values overlaid on the color gradient.
    - **Purpose**:
        - Combine visual and numerical details for clearer analysis.
    - **Examples**:
        - Sales performance by product type and sales volume.
8. **Tree Map**:
    - **Structure**:
        - Nested rectangles representing hierarchical data.
        - Rectangle size reflects one dimension; color represents another.

- **Purpose**:
    - Visualize hierarchical data and identify patterns.
- **Examples**:
    - Performance of product categories.
    - Financial data by department.

Trends and Innovations

- **Hybrid Charts**:
    - Combine different chart types for more comprehensive analysis.
    - Example: Interactive, animated bubble charts showing trends over time.
- **Interactive Visualizations**:
    - Enable real-time exploration of complex data sets.
    - Example: Dynamic charts with animations for time-series data.

MODULE 3 –(2ND PART)

BUSINESS REPORTING, VISUAL ANALYTICS, AND BUSINESS PERFORMANCE MANAGEMENT

What Is a Business Report?

- A business report is a written document that contains information regarding business matters.
- Business reporting (also called enterprise reporting) is an essential part of the larger drive toward improved managerial decision making and organizational knowledge management.

Components of the Business Reporting System

Following are the most common components of a business reporting system.

• **OLTP** (online transaction processing). A system that measures some aspect of the real world as events (e.g., transactions) and records them into enterprise databases. Examples include ERP systems, POS systems, Web servers, RFID readers, handheld inventory readers, card readers, and so forth.

• **Data supply.** A system that takes recorded events/transactions and delivers them reliably to the reporting system. The data access can be push or pull, depending on whether or not it is responsible for initiating the delivery process. It can also be polled (or batched) if the data are transferred periodically, or triggered (or online) if data are transferred in case of a specific event.

• **ETL** (extract, transform, and load). This is the intermediate step where these recorded transactions/events are checked for quality, put into the appropriate format, and inse1ted into the desired data format.

• **Data storage.** This is the storage area for the data and metadata. It could be a flat file or a spreadsheet, but it is usually a relational database management system (RDBMS) set up as a data mart, data warehouse, or operational data store (ODS); it often employs online analytical processing (OLAP) functions like cubes.

• **Business logic**. The explicit steps for how the recorded transactions/events are to be converted into metrics, scorecards, and dashboards.

• **Publication.** The system that builds the various reports and hosts them (for users) or disseminates them (to users). These systems may also provide notification, annotation, collaboration, and other services.

• **Assurance.** A good business reporting system is expected to offer a quality service to its users. This includes determining if and when the right information is to be delivered to the right people in the right way/format.

4.6 PERFORMANCE DASHBOARDS

Performance dashboards are common components of most performance management systems, performance measurement systems, BPM software suites, and BI platforms.

- **Purpose:** Dashboards provide visual displays of important information, consolidated and arranged on a single screen so that it can be digested at a glance and easily drilled into for further exploration.
- **Example:** A typical executive dashboard (e.g., Figure 4.8) displays KPIs for a hypothetical software company, Sonatica (selling audio tools).
- **Features:**
    - Provides a high-level view of functional groups, including marketing, sales, finance, and support departments.
    - Allows executive decision-makers to quickly and accurately assess organizational performance.
- **Details on Display:**
    - **Left side:** Quarterly changes in revenues, expenses, and margins displayed in a time-series format, along with comparisons to previous years' monthly numbers.
    - **Upper-right side:** Two color-coded dials show monthly expenses for support services (left dial) and other expenses (right dial).

Dashboard Design

Dashboards have a long history, dating back to the EIS (Executive Information Systems) of the 1980s. Today, they are ubiquitous in organizations of all sizes and industries.

- **Prevalence:**
    - A few years ago, Forrester Research estimated that over 40% of the largest 2,000 companies used dashboards.
    - This number has likely increased significantly since then.
    - The Dashboard Spy website (dashboardspy.com/about) provides further evidence, showcasing thousands of dashboards used by businesses, nonprofits, and government agencies.
- **Distinctive Features (According to Eckerson, 2006):** Dashboards consist of three layers of information:

1. **Monitoring:** Graphical, abstracted data to track key performance metrics.
2. **Analysis:** Summarized dimensional data to identify the root cause of problems.
3. **Management:** Detailed operational data to determine actions for resolving issues.

- **Design Challenges (According to Few, 2005):**
    - Dashboards aim to display all required information on a single screen clearly and without distractions.
    - Information must be quickly assimilated by placing numbers in context using baseline or target comparisons, trend indicators, and specialized display components.
- **Common Comparisons in BI Dashboards:**
    - Comparisons against past values, forecasted values, targeted values, benchmark or average values, multiple instances of the same measure, and other related measures (e.g., revenues vs. costs).
- **Evaluative Context:**

- Important to clearly indicate whether a number is good or bad and if the trend is moving in the right direction.
- Common tools include:
  - **Specialized visual objects:** Traffic lights, gauges, etc.
  - **Visual attributes:** Color coding, arrows, and tones.

What to Look for in a Dashboard

Although performance dashboards and other information visualization frameworks differ in their purpose, they all share some common design characteristics. First, they all fit within the larger business intelligence and/or performance measurement system. This means that their underlying architecture is the BI or performance management architecture of the larger system. Second, all well-designed dashboard and other information visualizations possess the following characteristics (Novell, 2009):

• They use visual components (e.g., charts, performance bars, sparklines, gauges, meters, stoplights) to highlight, at a glance, the data and exceptions that require action.

• They are transparent to the user, meaning that they require minimal training and are extremely easy to use.

• They combine data from a variety of systems into a single, summarized, unified view of the business.

• They enable drill-down or drill-through to underlying data sources or reports, providing more detail about the underlying comparative and evaluative context.

• They present a dynamic, real-world view with timely data refreshes, enabling the end user to stay up to date with any recent changes in the business.

• They require little, if any, customized coding to implement, deploy, and maintain.

Best Practices in Dashboard Design

Below are detailed best practices in dashboard design:

- **Benchmark Key Performance Indicators with Industry Standards**

  Many customers eventually question whether the metrics they are tracking are appropriate. Often, they discover that some metrics are irrelevant or ineffective for monitoring business

performance. By conducting a **gap assessment with industry benchmarks**, organizations can align their metrics with industry best practices. This ensures the organization is tracking the right metrics for actionable insights.

- **Wrap the Dashboard Metrics with Contextual Metadata**

When presenting a report or dashboard to business users, several important questions often arise:

- Where is the data sourced from?
- What percentage of the data encountered quality problems during the ETL process?
- Is the information on the dashboard fresh or outdated?
- When was the data warehouse last refreshed, and when will it be refreshed again?
- Were any high-value transactions rejected during data loading that could affect the overall trends?

By addressing these questions, the dashboard provides more context and ensures users can interpret the data accurately. Metadata adds transparency and improves trust in the dashboard.

- **Validate the Dashboard Design by a Usability Specialist**

In many cases, dashboards are designed by tool specialists without considering usability principles. Even if the underlying data warehouse is well-engineered and performs efficiently, users may not adopt the dashboard if they perceive it as difficult to use. Poor usability leads to low adoption rates and resistance to change.
To prevent these issues, it is crucial to have a **usability specialist validate the dashboard design** during the development phase. This step ensures the dashboard is intuitive and user-friendly, increasing its adoption and effectiveness.

- **Prioritize and Rank Alerts/Exceptions Streamed to the Dashboard**

Given the vast amount of raw data available, it is important to proactively identify and highlight critical exceptions or behaviors for the user. This can be achieved by:

- Codifying business rules that detect patterns requiring immediate attention.

- Implementing programs (e.g., database-stored procedures) to scan fact tables for these patterns.
  This approach ensures that critical information reaches the user directly, eliminating the need for users to manually search for relevant trends in the data.

➢ **Enrich the Dashboard with Business Users' Comments**

Dashboards are often viewed by multiple users with different perspectives. By providing a **text box for users to add comments**, dashboards can capture valuable context from end-users. These comments can be tagged to the dashboard, adding a layer of qualitative insights to the structured KPIs. This enriches the overall utility of the dashboard and improves decision-making.

➢ **Present Information in Three Different Levels**

Dashboards should present information in layers based on the granularity of the data. The three levels are:

1. **Visual Dashboard Level:** Presents a concise set of 8 to 12 KPIs, offering a quick overview of what is performing well and what needs attention.
2. **Static Report Level:** Provides more detailed reports for deeper insights.
3. **Self-Service Cube Level:** Allows users to explore the data further, offering granularity for those who need it.

By layering the information, dashboards cater to diverse user needs, from high-level executives to detailed analysts.

➢ **Pick the Right Visual Construct Using Dashboard Design Principles**

Different types of data require different visual representations to be effective:

- Bar charts are suitable for comparing categories.
- Time-series line graphs work well for trends over time.
- Scatter plots are effective for showing correlations between variables.
- Simple tables are often the best choice for straightforward data presentation.

Documenting these design principles ensures consistency across all dashboards and allows developers to adhere to a unified standard while designing reports and dashboards.

- **Provide for Guided Analytics**

In any organization, users vary in their level of analytical maturity. Dashboards can guide less experienced users by offering intuitive navigation paths similar to those used by advanced analytical users. This ensures that all users, regardless of their expertise, can derive meaningful insights from the dashboard.

4.7 BUSINESS PERFORMANCE MANAGEMENT

Closed-Loop BPM Cycle

The most significant differentiator of BPM (Business Performance Management) from other BI tools and practices is its focus on strategy. BPM follows a **closed-loop set of processes** linking strategy to execution to optimize business performance. The cycle involves four key steps: **strategize**, **plan**, **monitor/analyze**, and **act and adjust**. The continuous nature of this cycle supports ongoing process improvement.

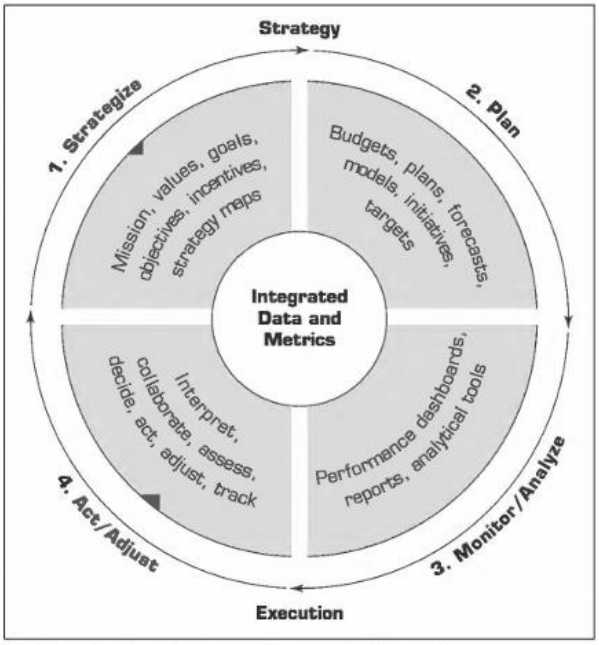


FIGURE 4.10 Closed-Loop BPM Cycle *Source: Business Intelligence, 2e.*

1. Strategize: Where do we want to go?

- **Definition:** Strategy is a high-level plan of action spanning a long time period (often several years) to achieve a specific goal.
- **Purpose:**
    - Identify and state the organization's mission, vision, and objectives.

    - Develop strategic, tactical, and operational plans to achieve these objectives.
- **Characteristics:**
    - Strategy is necessary when constraints like market conditions, resource availability, and legal/political factors must be addressed.
    - It provides overall direction to the enterprise and is the first process in BPM.
- **Responsibility:**
    - Corporate executives (led by the CEO) plan and create strategies.
    - Strategies are approved by the board of directors and implemented by the management team under senior executives.

2. Plan: How do we get there?

- **Definition:** Operational and financial plans translate strategic objectives into detailed tactics, resource requirements, and expected results for a specific period (usually a year).
- **Key Questions Addressed:**
    - What tactics and initiatives will meet the performance targets?
    - What are the expected financial results of executing these tactics?
- **Process:**
    - Strategies drive tactics, and tactics drive results.
    - Tactics must link directly to strategic objectives. If no linkage exists, management should question the necessity of the tactic.
- **Financial Planning and Budgeting:**
    - Revenue generation depends on producing or acquiring goods and services.
    - Associated costs include delivery, overhead, and capital requirements.
    - Collaborative input from departments is essential for clarity and communication of dependencies.

3. Monitor/Analyze: How are we doing?

- **Importance:** Monitoring ensures the organization tracks its progress effectively.
- **Key Issues:**
    1. **What to Monitor:** Organizations should focus on specific indicators or measures rather than attempting to track everything.

2. **How to Monitor:** Develop strategies for monitoring selected factors and responding effectively.

- **Key Performance Indicators (KPIs):**
  - Monitoring relies on KPIs to evaluate performance.
  - The process for determining KPIs and related topics like the balanced scorecard will be discussed further in this chapter.

4. Act and Adjust: What do we need to do differently?

- **New Projects:** Virtually all strategies depend on new initiatives, such as:
  - Creating new products.
  - Entering new markets.
  - Acquiring new customers or businesses.
  - Streamlining processes.
- **Challenges:**
  - Many projects fail due to lack of objectivity.
  - Failure rates vary by industry:
    - Hollywood movies: 60% failure rate.
    - Mergers and acquisitions: 60% failure rate.
    - Large IT projects: 70% failure rate.
    - New food products: 80% failure rate.
    - New pharmaceutical products: 90% failure rate.
  - Most new ventures fail between 60% and 80% of the time.
- **Key Question:** Given the high failure rates, the critical question becomes: "What do we need to do differently?

4.8 PERFORMANCE MEASUREMENT

Underlying BPM is a performance measurement system. According to Simons (2002), performance measurement systems: Assist managers in tracking the implementations of business strategy by comparing actual results against strategic goals and objectives. A performance measurement system typically comprises systematic methods of setting business goals together with periodic feedback reports that indicate progress against goals.

Key Performance Indicator (KPI)

There is a difference between a "run of the mill" metric and a "strategically aligned" metric. The term key performance indicator (KPI) is often used to denote the latter. A KPI represents a strategic objective and measures performance against a goal. According to Eckerson (2009), KPis are multidimensional. Loosely translated, this means that KPis have a variety of distinguishing **features**, including:

• **Strategy.** KPis embody a strategic objective.

• **Targets.** KPis measure performance against specific targets. Targets are defined in strategy, planning, or budgeting sessions and can take different forms (e.g., achievement targets, reduction targets, absolute targets).

• **Ranges**. Targets have performance ranges (e.g., above, on, or below target).

• **Encodings**. Ranges are encoded in software, enabling the visual display of performance (e.g., green, yellow, red). Encodings can be based on percentages or more complex rules.

• **Time frames**. Targets are assigned time frames by which they must be accomplished. A time frame is often divided into smaller intervals to provide performance mileposts.

• **Benchmarks.** Targets are measured against a baseline or benchmark. The previous year's results often serve as a benchmark, but arbitrary numbers or external benchmarks may also be used.

4.9 BALANCED SCORECARDS

The Four Perspectives

The balanced scorecard (BSC) suggests that an organization should be viewed from four key perspectives: customer, financial, internal business processes, and learning and growth. Objectives, measures, targets, and initiatives should be developed relative to each of these perspectives. Figure 4.11 illustrates these perspectives and their interrelationship with the organization's vision and strategy.

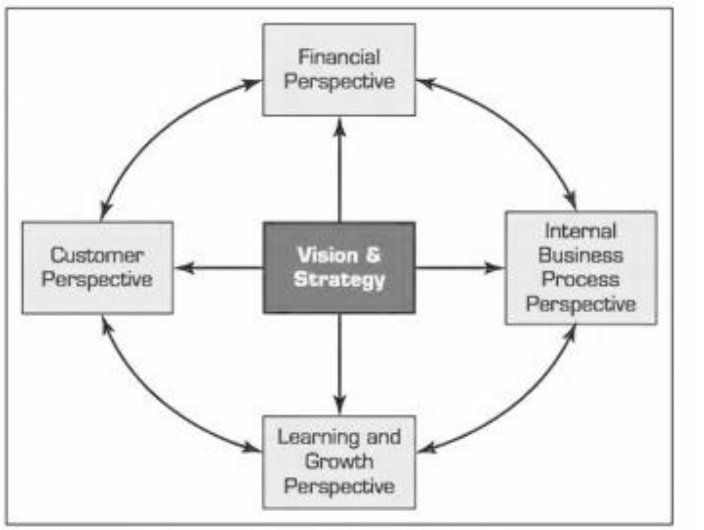


FIGURE 4.11 Four Perspectives in Balanced Scorecard Methodology.

THE CUSTOMER PERSPECTIVE

Recent management philosophies emphasize the importance of customer focus and satisfaction in business. These factors serve as leading indicators; if customers are not satisfied, they will eventually seek other suppliers. Poor performance in this area can predict future decline, even when financial performance seems strong. Metrics for customer satisfaction should consider different customer types and the specific processes for which products or services are provided to these groups.

THE FINANCIAL PERSPECTIVE

Kaplan and Norton acknowledge the traditional importance of financial data. Accurate financial information remains a priority, and corporate databases aim to centralize and automate much of the financial data processing. However, the current overemphasis on financials can lead to an imbalance with other perspectives. Additional financial data, such as risk assessments and cost-benefit analysis, could enhance this perspective.

THE LEARNING AND GROWTH PERSPECTIVE

This perspective answers the question, "To achieve our vision, how will we sustain our ability to change and improve?" It involves employee training, knowledge management, and corporate culture, focusing on both individual and organizational improvement. In a rapidly changing technological landscape, continuous learning and growth are essential for knowledge workers.

Metrics can guide managers in focusing training resources where they are most beneficial. "Learning" extends beyond training, encompassing mentoring, tutoring, and effective communication within the organization for problem-solving.

THE INTERNAL BUSINESS PROCESS PERSPECTIVE

This perspective emphasizes the importance of business processes. Metrics from this perspective help managers assess how well internal processes are performing and whether the outcomes meet customer requirements. These processes and outcomes directly impact the organization's mission and goals.

The Meaning of Balance in BSC

The BSC serves as both a performance measurement and management methodology. It helps translate an organization's financial, customer, internal process, and learning and growth objectives into actionable initiatives. As a performance measurement tool, the BSC overcomes the limitations of traditional financial systems by incorporating both financial and non-financial objectives. These non-financial objectives fall under three perspectives:

- **Customer:** This defines how the organization should appear to its customers to achieve its vision.
- **Internal business process:** Specifies the processes the organization must excel at to satisfy shareholders and customers.
- **Learning and growth:** Focuses on how an organization can improve its ability to change and grow to achieve its vision.

These non-financial objectives form a causal chain, where "learning and growth" drives changes in "internal business processes," which lead to improved "customer" outcomes, ultimately achieving "financial" objectives. This causal flow is demonstrated in Figure 4.12, which shows a strategy map and BSC for a fictitious company. The company’s financial goal, such as increasing net income, is driven by a customer objective like improving customer retention. In turn, customer retention is influenced by internal process objectives, such as enhancing call center performance. Learning objectives, such as reducing employee turnover, are at the base of this hierarchy.

The term "balance" in the BSC refers to the combination of different types of measures, ensuring that they are:

- Financial and nonfinancial
- Leading and lagging
- Internal and external
- Quantitative and qualitative
- Short-term and long-term

Dashboards Versus Scorecards

While dashboards and scorecards are often used interchangeably in trade journals, they serve different purposes. Executives, managers, and staff use scorecards to monitor strategic alignment and success in achieving strategic objectives and targets. The BSC is a well-known example. Dashboards, in contrast, are used at the operational and tactical levels. Operational dashboards track performance on a daily, weekly, or even hourly basis (e.g., production quality), while tactical dashboards monitor initiatives like marketing campaigns or sales performance.

4.10 SIX SIGMA AS A PERFORMANCE MEASUREMENT SYSTEM

 Introduction and Adoption:

- Six Sigma originated in the mid-1980s and has been widely adopted by companies globally.
- Initially used as a process improvement methodology rather than a performance measurement and management tool.

 Strategic Use:

- Some companies (e.g., Motorola) have recognized Six Sigma's value for strategic purposes.
- It helps measure and monitor key processes linked to profitability and accelerates business performance improvement.

 Focus on Business Processes:

- Six Sigma primarily addresses performance issues by focusing on improving business processes after problems are identified.

 Meaning of Sigma (σ):

- Sigma (σ) is a Greek letter used by statisticians to measure variability in a process.
- In quality management, variability equates to the number of defects in a process.

 Acceptable Variability in Business:

- Historically, companies accepted 6,200 to 67,000 defects per million opportunities (DPMO).
- Example: In an insurance company processing 1 million claims, 6,200 to 67,000 claims could be defective under normal conditions.

 Six Sigma Performance Level:

- Achieving Six Sigma means reducing defects to no more than 3.4 DPMO.

Six Sigma is a performance management methodology aimed at reducing the number of defects in a business process to as close to zero DPMO as possible.

The DMAIC Performance Model

Six Sigma rests on a simple performance improvement model known as DMAIC. Like BPM, DMAIC is a closed-loop business improvement model, and it encompasses the steps of defining, measuring, analyzing, improving, and controlling a process. The steps can be described as follows:

1. **Define**. Define the goals, objectives, and boundaries of the improvement activ ity. At the top level, the goals are the strategic objectives of the company. At lower levels- department or project levels- the goals are focused on specific operational processes.

2. **Measure.** Measure the existing system. Establish quantitative measures that will yield statistically valid data. The data can be used to monitor progress toward the goals defined in the previous step.

3. **Analyze.** Analyze the system to identify ways to eliminate the gap between the current performance of the system or process and the desired goal.

4. **Improve**. Initiate actions to eliminate the gap by finding ways to do things better, cheaper, or faster. Use project management and other planning tools to implement the new approach.

5**. Control.** Institutionalize the improved system by modifying compensation and incentive systems, policies, procedures, manufacturing resource planning, budgets, operation instructions, or other management systems.

TABLE 4.1 Comparison of Balanced Scorecard and Six Sigma

Balanced Scorecard	Six Sigma
Strategic management system	Performance measurement system
Relates to the longer-term view of the business	Provides snapshot of business's performance and identifies measures that drive performance toward profitability
Designed to develop balanced set of measures	Designed to identify a set of measurements that impact profitability
Identifies measurements around vision and values	Establishes accountability for leadership for wellness and profitability
Critical management processes are to clarify vision/strategy, communicate, plan, set targets, align strategic initiatives, and enhance feedback	Includes all business processes—management and operational
Balances customer and internal operations without a clearly defined leadership role	Balances management and employees' roles; balances costs and revenue of heavy processes
Emphasizes targets for each measurement	Emphasizes aggressive rate of improvement for each measurement, irrespective of target
Emphasizes learning of executives based on the feedback	Emphasizes learning and innovation at all levels based on the process feedback; enlists all employees' participation
Focuses on growth	Focuses on maximizing profitability
Heavy on strategic content	Heavy on execution for profitability
Management system consisting of measures	Measurement system based on process management