CS660 – Database systems

Dr. John Conklin

UNIT 5

VARIATIONS IN DATABASE SYSTEMS

Advanced Database Architectures & Applications

UNIT 5 OVERVIEW

1. **Data Marts & Warehouses**
   Enterprise data storage and analytics infrastructure

2. **Decision Support Systems**
   OLAP, data mining, and analytical processing

3. **Distributed Databases**
   Multi-location data management and replication

4. **Web-Based Databases**
   Internet-connected database architectures

5. **Business Intelligence**
   Strategic insights from organizational data

DATA WAREHOUSES

What is a Data Warehouse?

A centralized repository that stores integrated data from multiple sources, optimized for query and analysis rather than transaction processing. Data warehouses support business intelligence, reporting, and data analysis.

Subject-Oriented

Organized around business subjects

Integrated

Combines data from multiple sources

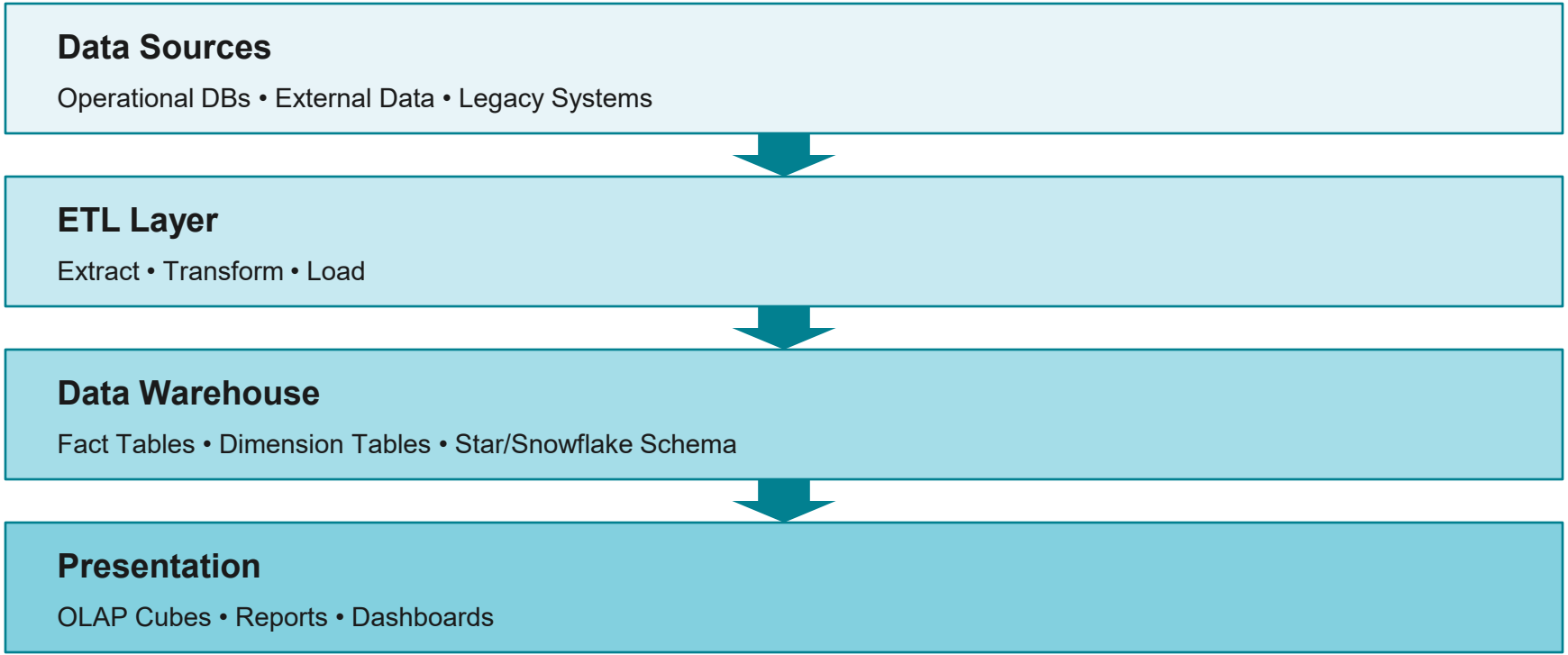
Time-Variant

Historical data with timestamps

Non-Volatile

Stable, not updated frequently

DATA WAREHOUSE ARCHITECTURE



DATA MARTS

What is a Data Mart?

A subset of a data warehouse focused on a specific business area (e.g., Sales, Finance, Marketing). Smaller, faster, and more focused than enterprise data warehouses.

Data Warehouse vs Data Mart

- DW: Enterprise-wide scope
- DM: Department-specific
- DW: Large (TB-PB)
- DM: Smaller (GB-TB)

Types of Data Marts

Dependent

Subset of enterprise DW

✓ Consistent data

⚠ Requires DW first

Independent

Standalone system

✓ Quick to deploy

⚠ Data silos risk

Hybrid

Combination approach

✓ Flexible

⚠ Complex integration

DECISION SUPPORT SYSTEMS

Decision Support Systems (DSS)

Interactive software systems that help decision-makers analyze data, identify patterns, and solve complex problems. DSS combines data, analytical models, and user-friendly interfaces to support strategic and operational decisions.

Data Management

Database, data warehouse, data marts

Model Management

Statistical, financial, optimization models

Knowledge Engine

Rules, patterns, expert knowledge

User Interface

Dashboards, reports, visualizations

OLAP: ONLINE ANALYTICAL PROCESSING

OLAP vs. OLTP

Purpose	Analysis & Reporting	Transaction Processing
Users	Analysts, Managers	Clerks, Customers
Data	Historical, aggregated	Current, detailed
Queries	Complex, read-heavy	Simple, read/write
Response	Seconds to minutes	Milliseconds
Size	Large (TB-PB)	Smaller (GB-TB)

OLAP Operations

Slice • Dice • Drill-Down • Drill-Up • Roll-Up • Pivot

DATA MINING

What is Data Mining?

The process of discovering patterns, correlations, and insights from large datasets using statistical and machine learning techniques. Data mining turns raw data into actionable knowledge.

Classification

Categorize data into classes • Example: Spam detection

Clustering

Group similar items • Example: Customer segmentation

Association

Find relationships • Example: Market basket analysis

Regression

Predict numeric values • Example: Sales forecasting

Anomaly Detection

Identify outliers • Example: Fraud detection

Sequential Patterns

Discover trends over time • Example: Click stream analysis

DISTRIBUTED DATABASES

What is a Distributed Database?

A database stored across multiple physical locations, connected by a network. Data can be distributed geographically while appearing as a single logical database to users.

Advantages

- Improved performance (data closer to users)
- Increased availability and reliability
- Scalability (add nodes as needed)
- Local autonomy for branches

Challenges

- Complexity in design and management
- Data consistency across locations
- Network dependency and latency
- Security across multiple sites

Distribution Strategies

Replication: Full copy at each site

Partitioning: Split data across sites

Hybrid: Combination approach

REPLICATION & FRAGMENTATION

Data Replication

Copies of data stored at multiple sites

- Full Replication: Complete copy everywhere
- Partial Replication: Selected data copied
- ✓ Faster reads, redundancy
- ⚠ Synchronization overhead

Data Fragmentation

Data split across multiple sites

- Horizontal: Rows split (e.g., by region)
- Vertical: Columns split (e.g., by dept)
- ✓ Reduced data transfer, local access
- ⚠ Complex joins across sites

CAP Theorem: Pick Two of Three

Consistency: All nodes see same data • Availability: System always responds • Partition Tolerance: Works despite network failures

Traditional RDBMS: CA (Consistency + Availability) • NoSQL: AP or CP

WEB-BASED DATABASES

What are Web-Based Databases?

Databases accessible through web browsers using internet protocols. Web applications interact with databases via middleware (application servers, APIs) to deliver dynamic content to users.

Three-Tier Architecture

Presentation Tier	*User Interface*	Technologies: HTML, CSS, JavaScript, React
Application Tier	*Business Logic*	Technologies: Node.js, Python, Java, PHP
Data Tier	*Data Storage*	Technologies: MySQL, PostgreSQL, MongoDB

WEB DATABASE TECHNOLOGIES

APIs	REST APIs • GraphQL • WebSockets
Middleware	Express.js • Django • Spring Boot
ORMs	Sequelize • SQLAlchemy • Hibernate
Cloud DBs	AWS RDS • Azure SQL • Google Cloud SQL

Security Considerations

SQL Injection Prevention • Authentication & Authorization • HTTPS/TLS Encryption • Input Validation • Prepared Statements

BUSINESS INTELLIGENCE

What is Business Intelligence (BI)?

Technologies, applications, and practices for collecting, integrating, analyzing, and presenting business data. BI transforms data into actionable insights that drive strategic and operational decisions.

Data Integration

ETL processes, data pipelines

Analytics

OLAP, data mining, predictive models

Reporting

Dashboards, reports, alerts

Visualization

Charts, graphs, interactive displays

BI TOOLS & PLATFORMS

Tableau
Strength: Interactive visualizations
Use Case: Enterprise dashboards

Power BI
Strength: Microsoft integration
Use Case: Business reporting

Qlik
Strength: Associative analytics
Use Case: Self-service BI

Looker
Strength: Cloud-native, Git-based
Use Case: Embedded analytics

SAP BusinessObjects
Strength: Enterprise scale
Use Case: Corporate reporting

IBM Cognos
Strength: AI-powered insights
Use Case: Financial analysis

BI PROCESS & LIFECYCLE

1. Requirements	Identify business questions, KPIs, stakeholders
2. Data Collection	Extract from sources, validate quality
3. Data Integration	Transform, cleanse, load into warehouse
4. Analysis	Apply OLAP, data mining, statistical models
5. Visualization	Create dashboards, reports, scorecards
6. Action	Make decisions, implement changes, monitor

KEY PERFORMANCE INDICATORS

KPIs are measurable values that demonstrate how effectively an organization is achieving key business objectives. BI systems track, analyze, and visualize KPIs to support decision-making.

Sales	Revenue Growth • Conversion Rate • Customer Acquisition Cost
Marketing	ROI • Lead Generation • Brand Awareness
Operations	Efficiency Rate • Cycle Time • Quality Score
Finance	Profit Margin • Cash Flow • Budget Variance

SMART KPI Criteria

Specific • Measurable • Achievable • Relevant • Time-bound

KEY TAKEAWAYS

- Data warehouses centralize enterprise data for analytics and reporting
- Decision support systems combine data, models, and interfaces for better decisions
- Distributed databases enable geographic scalability with replication and fragmentation
- Web-based databases power modern applications through three-tier architectures
- Business intelligence transforms data into strategic insights using OLAP, mining, and visualization

These advanced database systems power enterprise-scale applications! 🚀

INDIVIDUAL PROJECT

Individual Project

The case study retail store has expressed a desire to eventually be able to analyze the data that are collected from engaging in business, both in its brick-and-mortar store and in its online store. Because the system had to be redesigned from the ground up, the goal of taking the business online using a Web-based database was pushed back. Eventually, the company wants to be able to run statistical analyses against the data that it is collecting and to be able to drill down through the data to transform them into various desired formats. In addition, the company would like to acquire data sets from other providers to engage in decision-support initiatives. What recommendations can you propose to support these business intelligence goals?

Part 1

Future Database System Implementation Plan (4–5 pages)

- What fundamental differences exist between object-oriented and object-relational database systems and Web-based database systems?
  - Would these differences impact your retail store?
- Include details of what changes would need to be introduced to the database if it was used to build a data mart or a data warehouse.
- Include details of what considerations would need to be made if the database were to become a d  tabase.
- What specific types of business intelligence could be gathered from the database?
  - How would this information assist in the decision-making process for your retail store?
- How would your retail store benefit from data warehousing in the following areas?
  - Return on investment on business intelligence initiatives (Provide a 3-year estimate.)
  - Competitive advantage (based on local or target area)
  - Increased productivity of decision-makers (related to business process decision-making)
- How would you address the following data warehousing problems if they occurred in your retail store?
  - Required data were never captured.
  - There is a high demand for disk space and other resources.
  - There are hidden problems with source systems.
- Provide your analysis as to how this part of the project fulfills the mission and 1 or more goals of the case study organization.

All sources should be cited both in-text and in References using APA format.

Part 2

Argumentation PowerPoint (5-slide PowerPoint excluding title slide and References)

An *argument* in problem solving is used to permit the problem solver the opportunity to offer a set of reasons or evidence in support of his or her solution and related conclusions from the inquiry and research process. The argument is not an opinion and is not the conclusion restated, but rather a demonstration of the intellectual inquiry that you have made during the problem-solving process. Naturally, for the argument to gain credibility, it must be reinforced with scholarly references because the support of the conclusion comes from the quality of evidence that has been gathered.

You can skip this part of the assignment. ☺

Contact Information

Email:	Jconklin@coloradotech.edu
Phone:	602.796.5972
Website:	http://drjconklin.com
Office Hours:	Wednesdays: 6:00 PM – 7:00 PM (CST)
	Saturdays: 11:00 AM – 12:00 PM (CST)
Live Chats:	Wednesdays: 6:00 PM – 7:00 PM (CST)