



**Business School**  
UNIVERSITY OF COLORADO **DENVER**

**Information Systems Program**

# **Module 4**

## **Data Integration Concepts, Processes, and Techniques**

### **Lesson 1: Concepts of Data Integration Processes**



# Lesson Objectives

- Explain diagrams for refresh processing and typical tasks
- Discuss difficulties of initial population of a data warehouse
- Understand tradeoffs and constraints in managing refresh processing

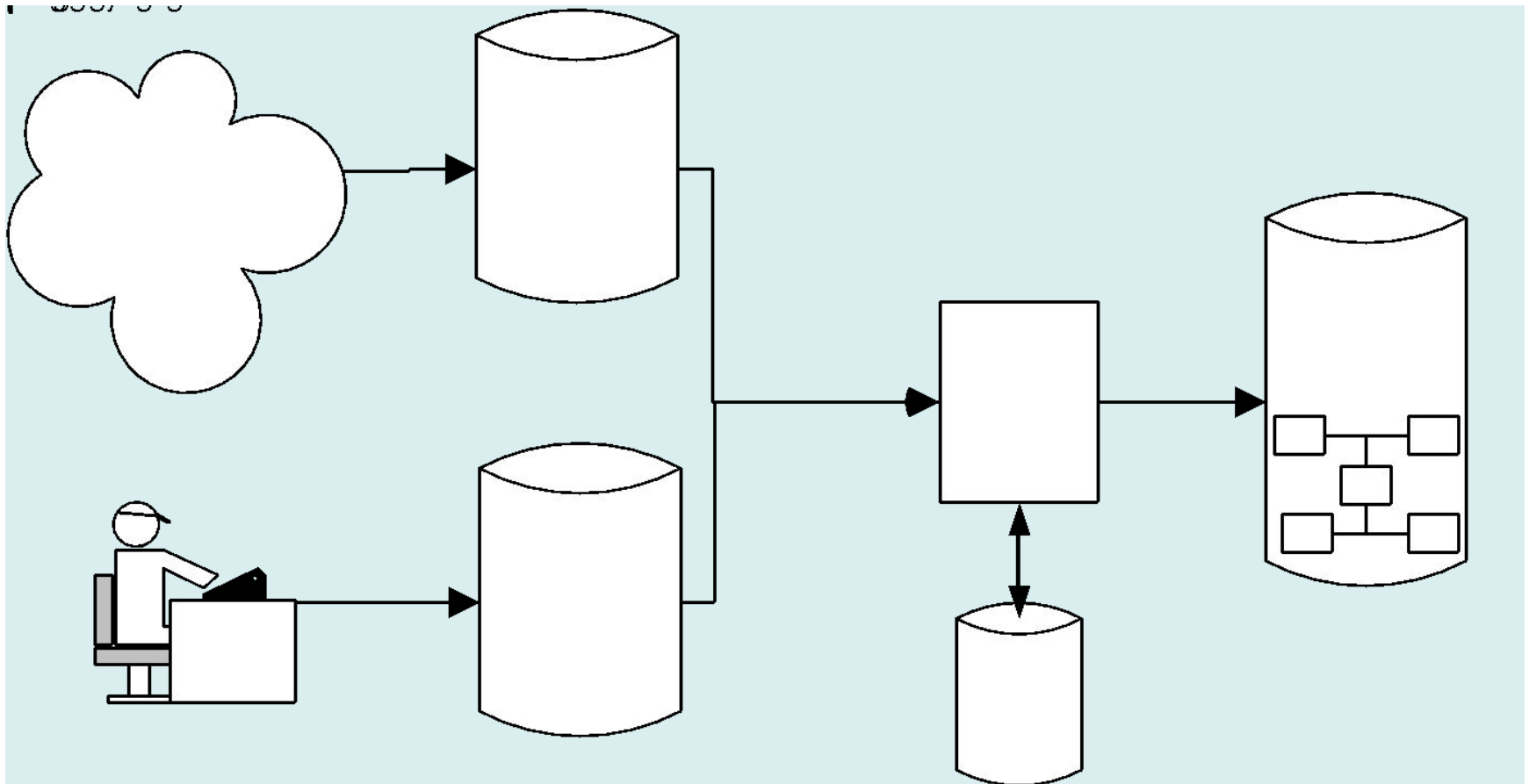


# Motivation for Data Integration

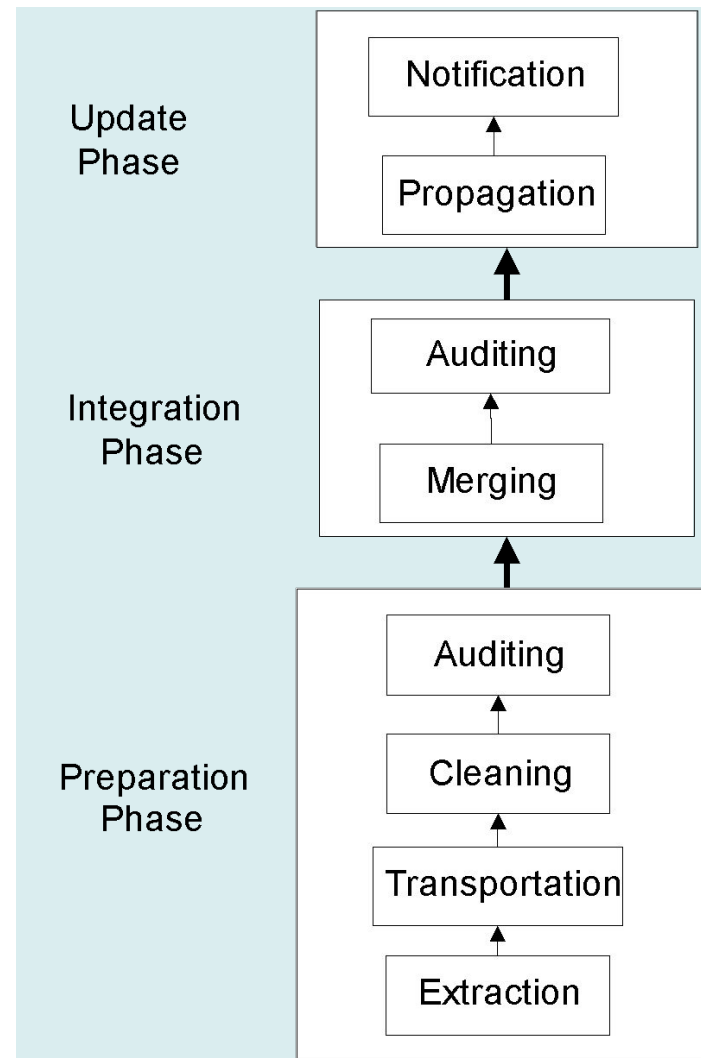
- Add value by data transformations
- Find single source of truth
- Overcome challenges
- Critical success factor for data warehouse projects
- Significant investments in effort, hardware, and software



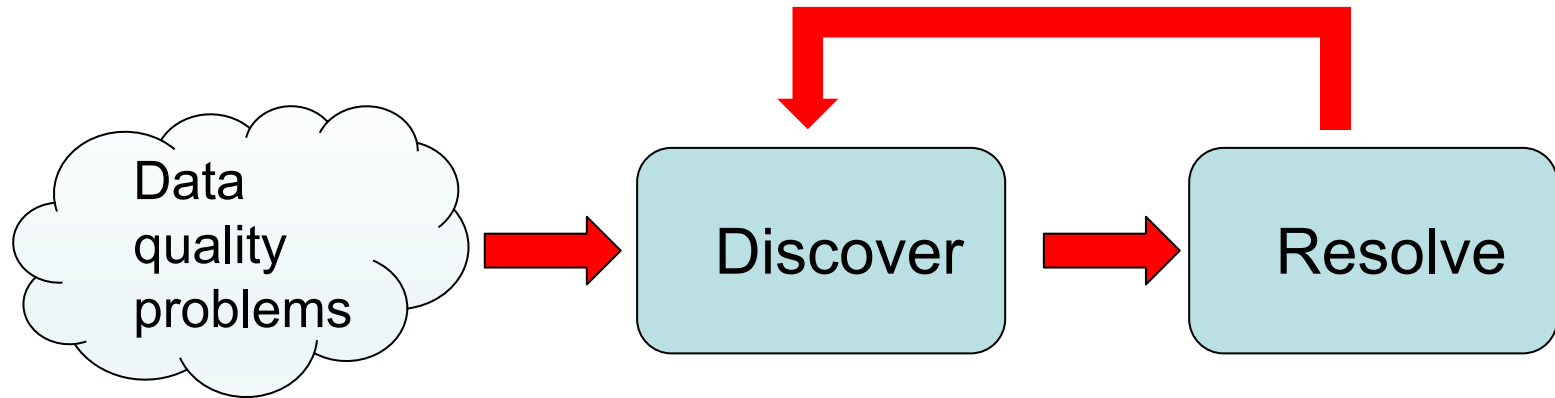
# Refresh Processing



# Refresh Workflow



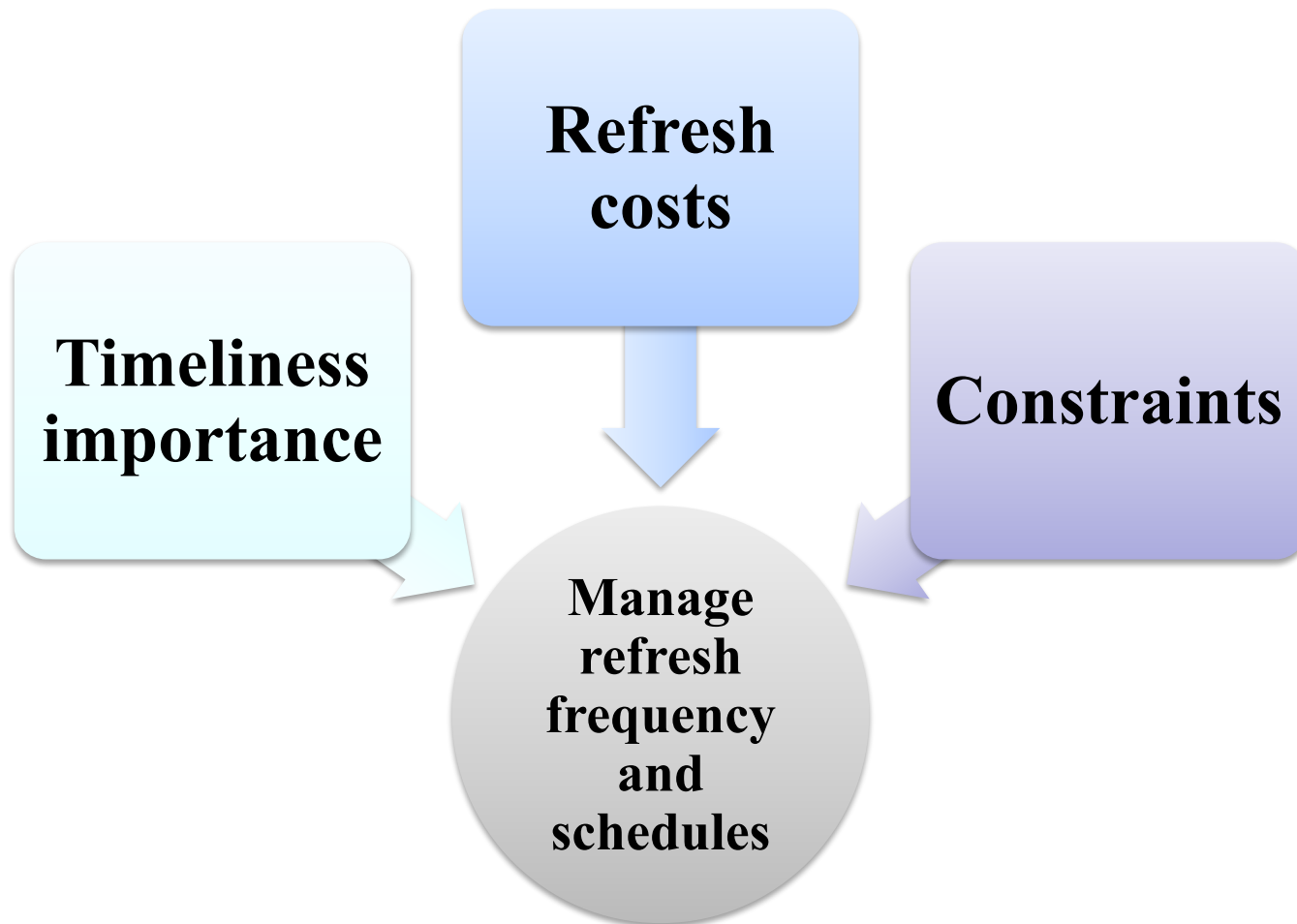
# Initial Data Warehouse Load



- Major development activity
- More open ended than refresh with difficult to estimate time requirements
- Use profiling tools to discover data quality problems
- Perform for major data warehouse extensions



# Refresh Processing Decision Making



# Refresh Constraints

- **Satisfy Constraints**

- **Source Access**

- **Integration**

- **Consistency**

- **Completeness**

- **Availability**





# Summary

- Add value by transformations
- Tedious but important processes for population and refreshment
- Substantial investments in technology and effort
- Data integration tools to improve productivity

