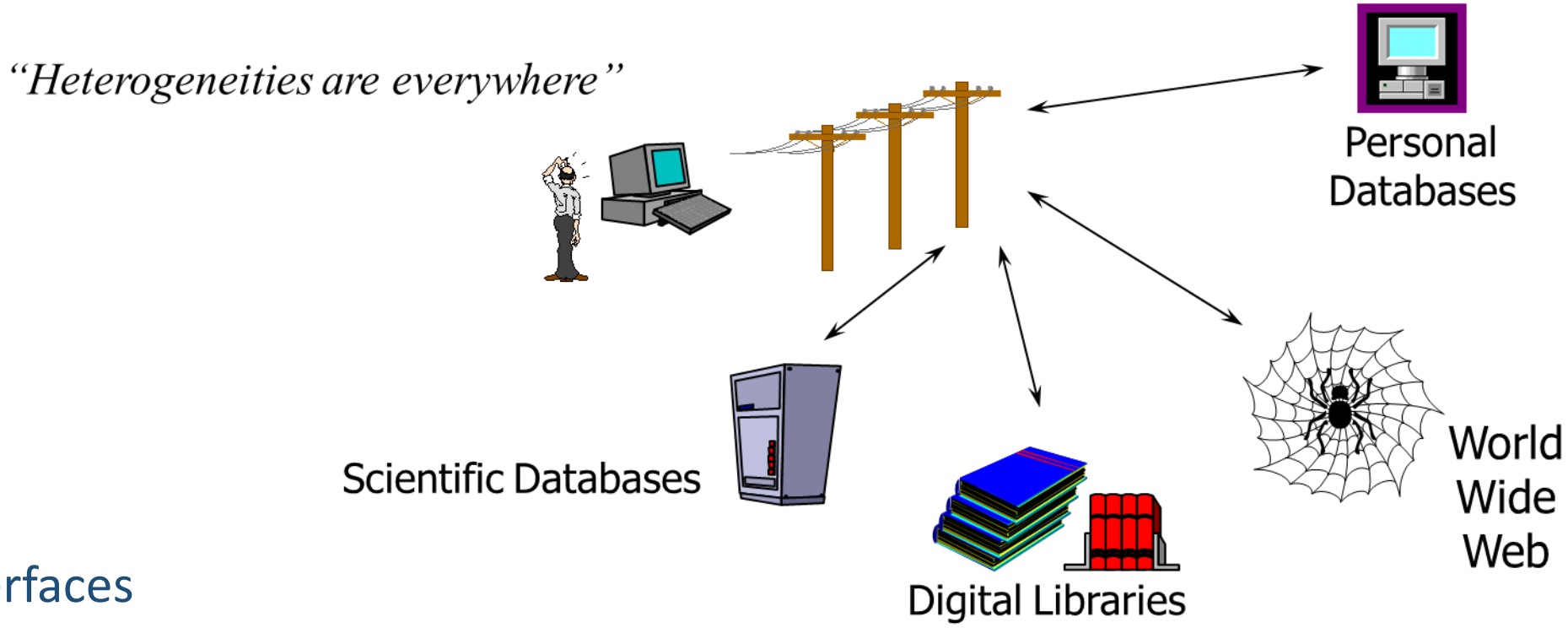
FOUNDATIONS OF BUSINESS INTELLIGENCE
Data Warehouses and Information Management

Lesson 08

Management and Information Systems

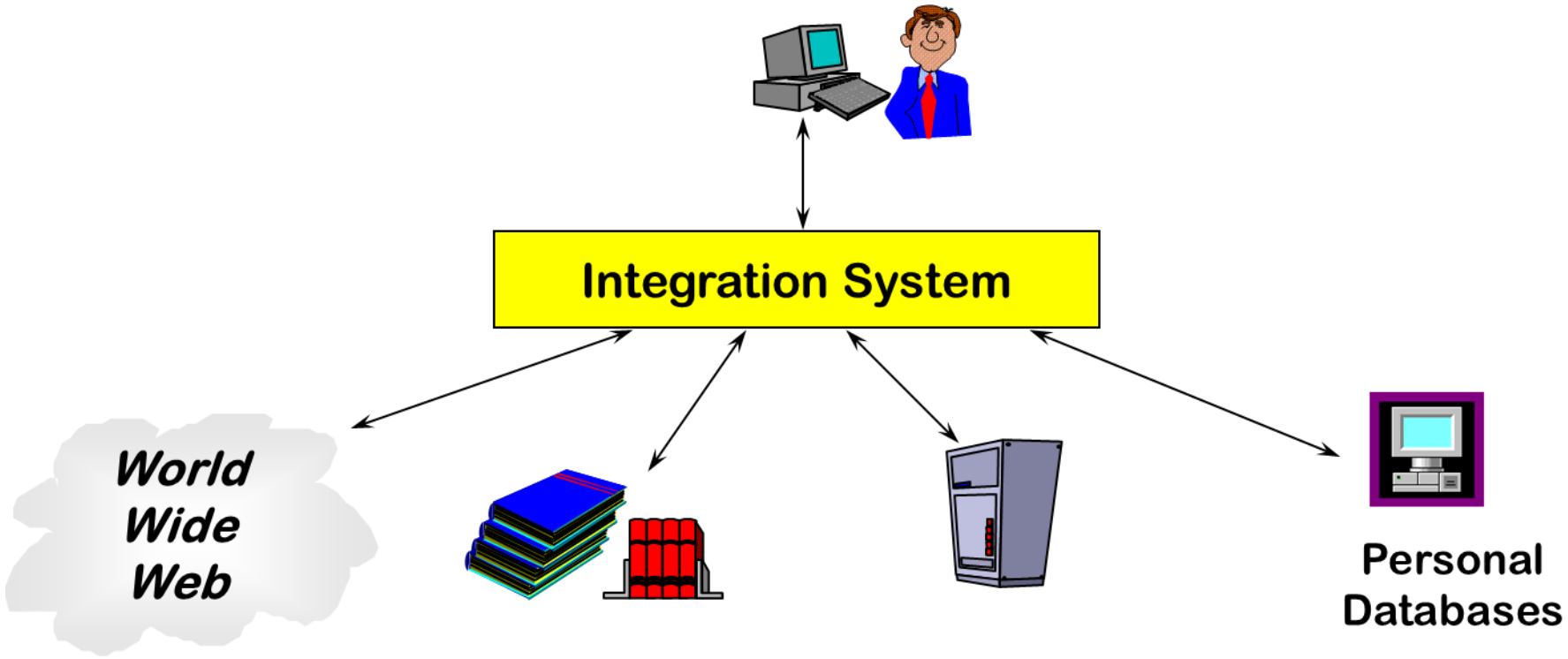
Department of Computing and information Systems

Problem: Heterogeneous Information Sources



- Different interfaces
- Different data representations
- Duplicate and inconsistent information

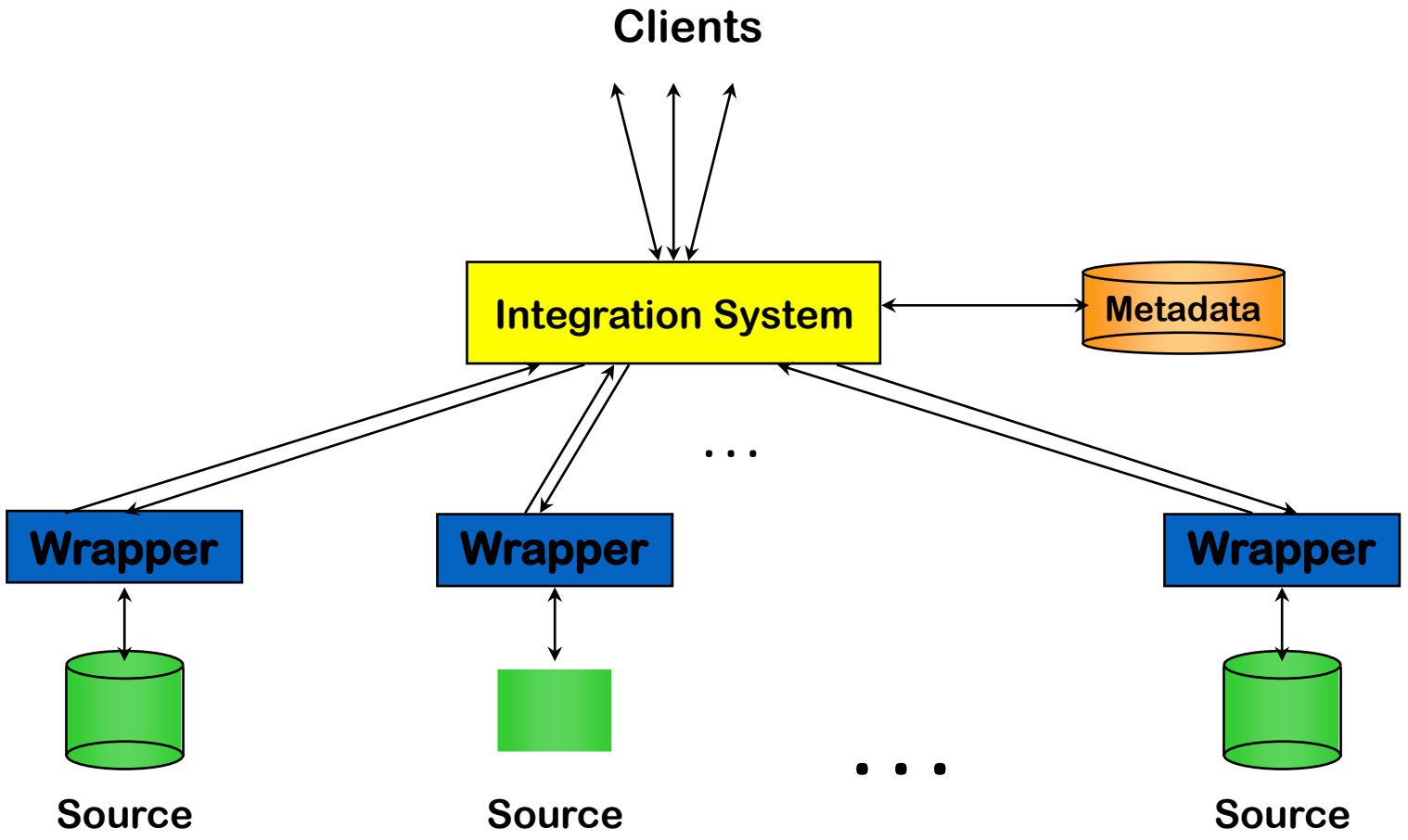
Goal: Unified Access to Data



- Collects and combines information
- Provides integrated view, uniform user interface
- Supports sharing

The Traditional Research Approach

Query-driven (lazy, on-demand)

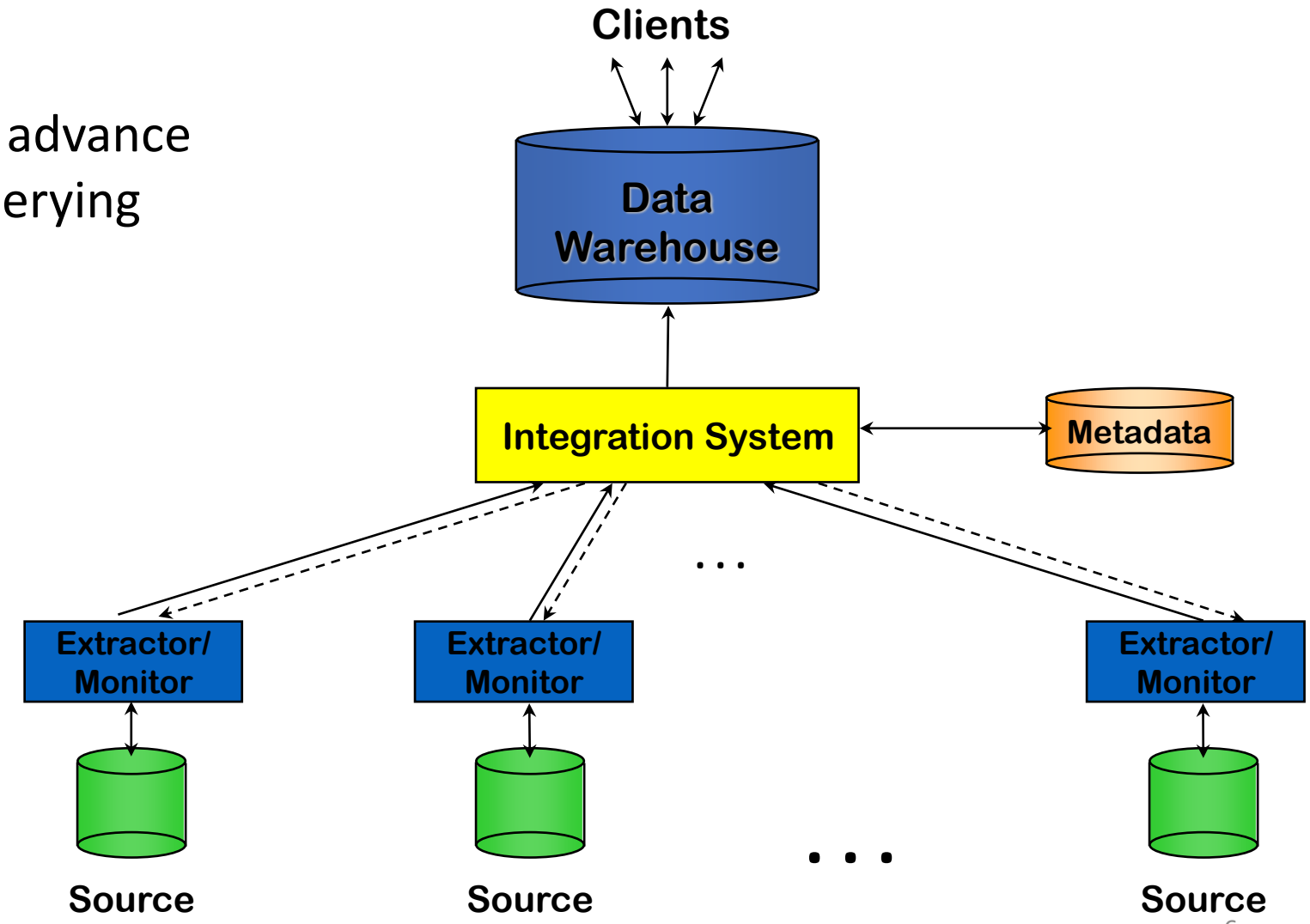


Disadvantages of Query-Driven Approach

- Delay in query processing
  - Slow or unavailable information sources
  - Complex filtering and integration
- Inefficient and potentially expensive for frequent queries
- Competes with local processing at sources
- *Hasn't* caught on in industry

The Warehousing Approach

- Information integrated in advance
- Stored in wh for direct querying and analysis



What is a Data Warehouse?

A Practitioners Viewpoint

"A data warehouse is simply a single, complete, and consistent store of data obtained from a variety of sources and made available to end users in a way they can understand and use it in a business context."

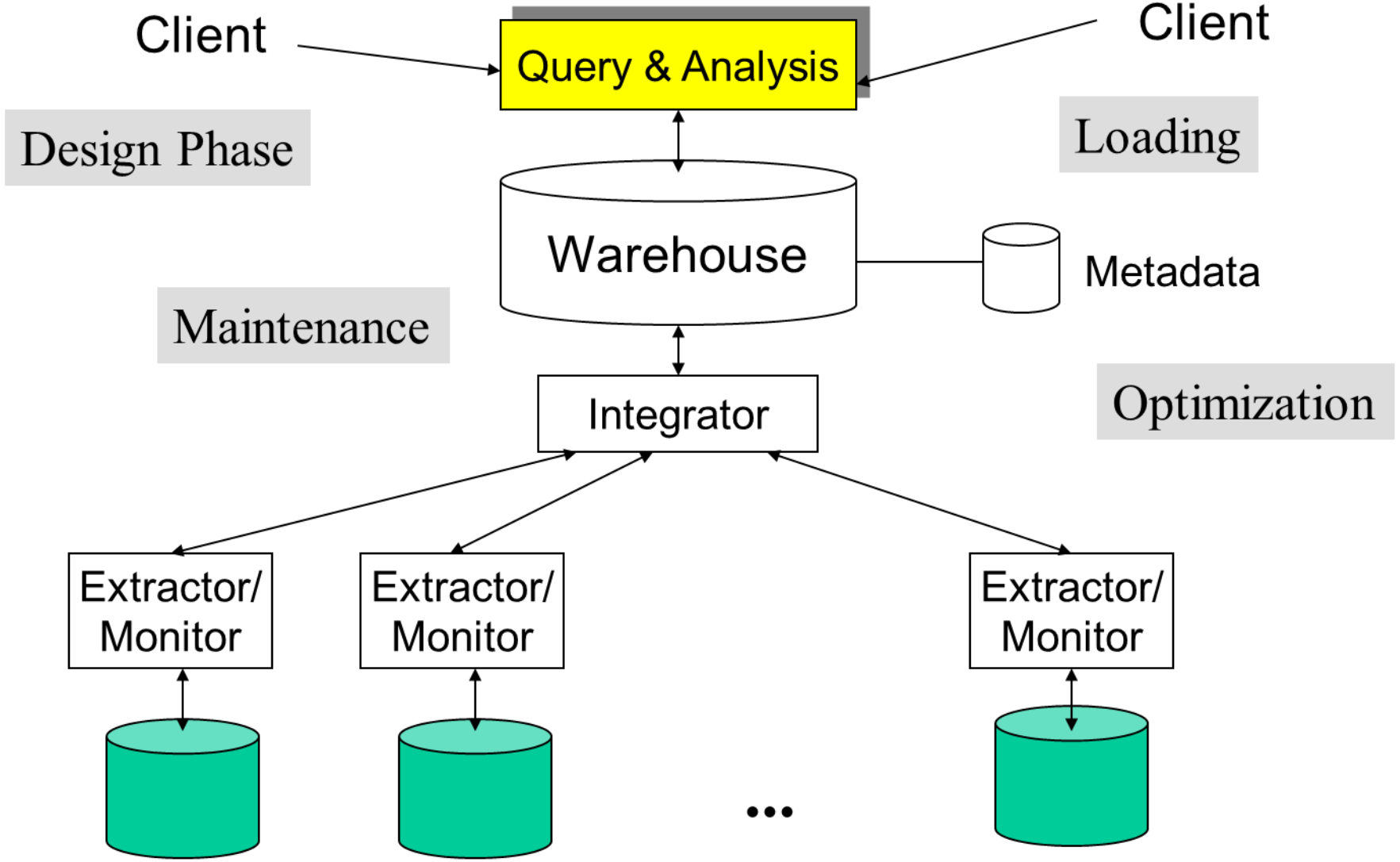
-- Barry Devlin, *IBM Consultant*

A Data Warehouse is...

- Stored collection of diverse data
  - A solution to data integration problem
  - Single repository of information
- Subject-oriented
  - Organized by subject, not by application
  - Used for analysis, data mining, etc.
- Optimized differently from transaction-oriented db
- User interface aimed at executive
- Large volume of data (Gb, Tb)
- Non-volatile
  Historical
  Time attributes are important
- Updates infrequent

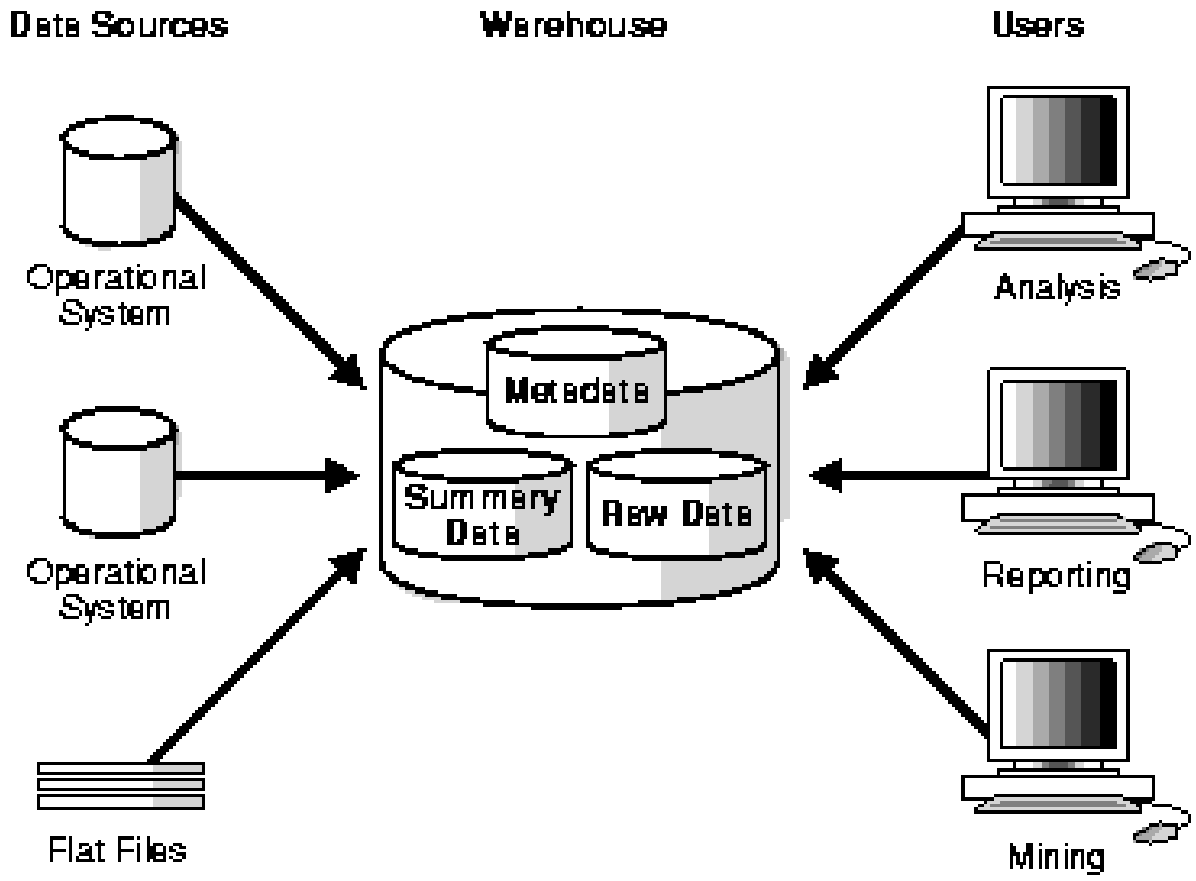
Ex:
Complete client histories at insurance firm

Generic Warehouse Architecture



Data Warehouse Concepts

- Basic Data Warehouse

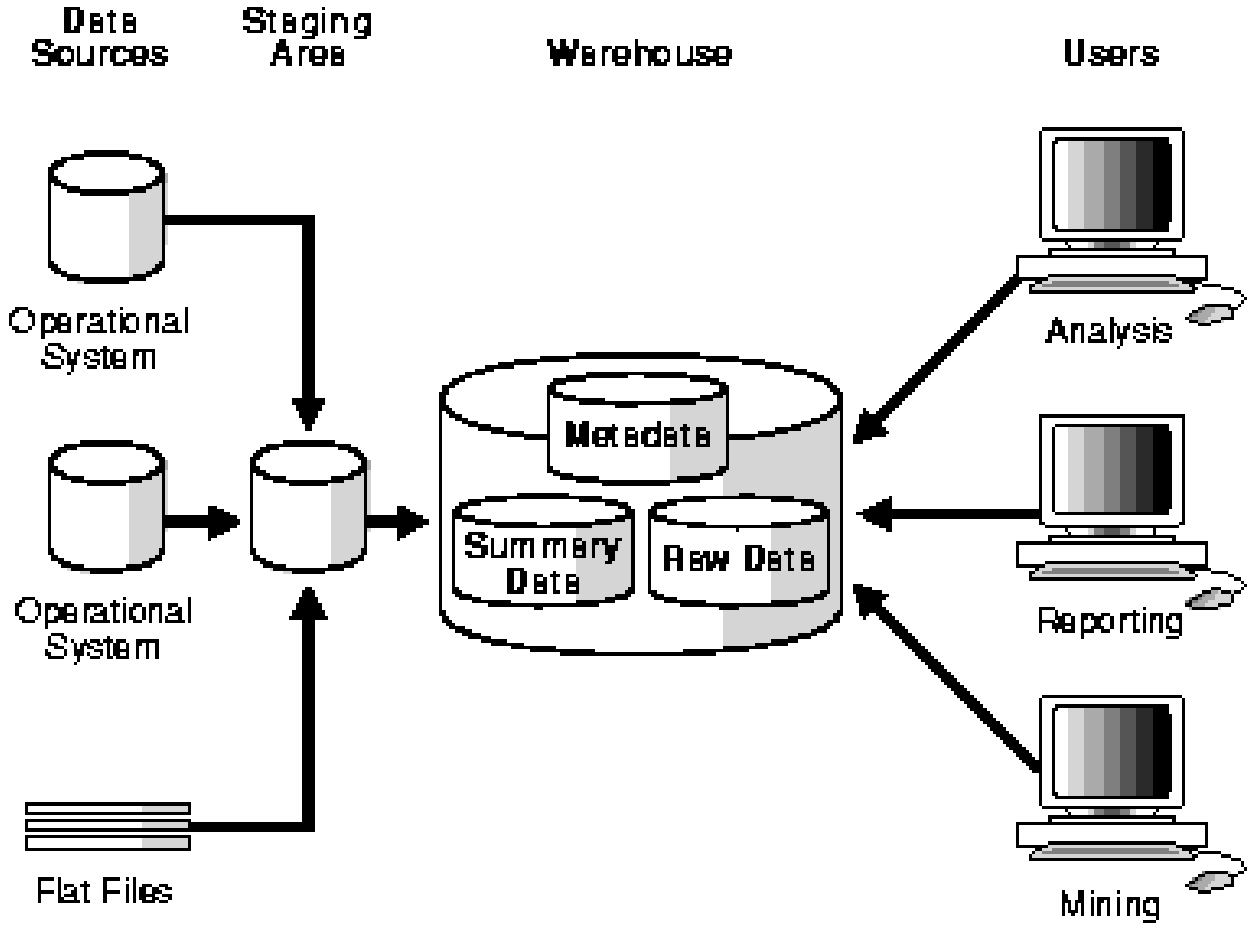


Data Warehouse Concepts

- Data Warehouse with Staging Area

Some data warehouses clean and process data before moving it into storage.

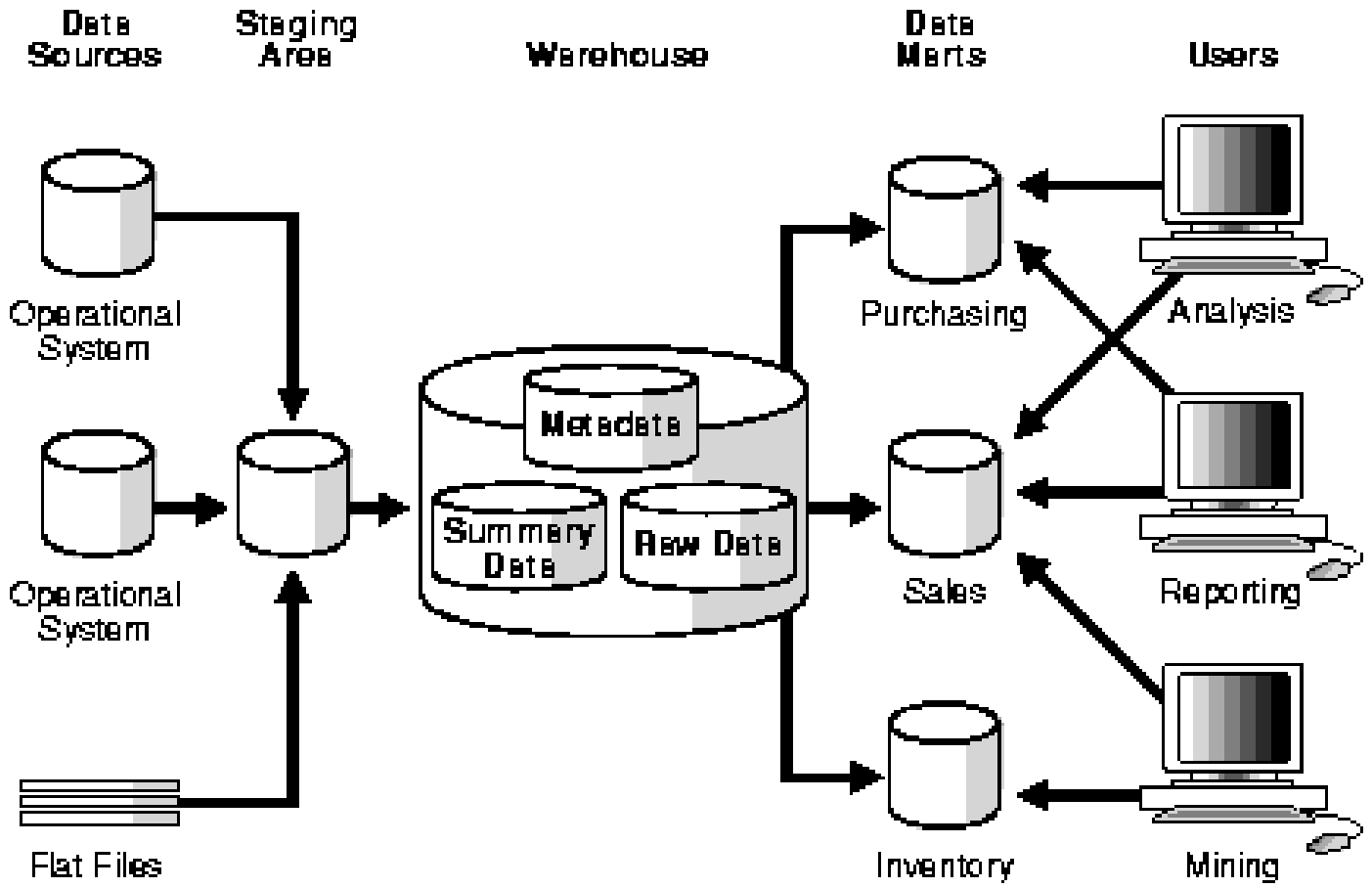
These systems have "staging areas" where information is reviewed, evaluated, then deleted or transferred into the warehouse.



Data Warehouse Concepts

- Data Warehouse With Data Marts

- Data marts add another level of customization to your data warehouse.
- Once data is processed and evaluated, data marts streamline information to teams and employees who need it most.



Advantages of Warehousing Approach

- High query performance
- Not necessarily most current information
- Complex queries at warehouse
- OLTP at information sources
- Information copied at warehouse
- Can modify, annotate, summarize, restructure, etc.
- Can store historical information
- Security, no auditing
- Has caught on in industry

Data Warehouse vs. Data Marts

- Enterprise warehouse: collects all information about subjects (customers, products, sales, assets, personnel) that span the entire organization
  - Requires extensive business modeling (may take years to design and build)

- Data Marts: Departmental subsets that focus on selected subjects
  - Marketing data mart: customer, product, sales
  - Faster roll out, but complex integration in the long run

OLAP for Decision Support

- OLAP = Online Analytical Processing
- Support (almost) ad-hoc querying for business analyst
- Think in terms of spreadsheets
  - View sales data by geography, time, or product
- Extend spreadsheet analysis model to work with warehouse data
  - Large data sets
  - Semantically enriched to understand business terms
  - Combine interactive queries with reporting functions
- Multidimensional view of data is the foundation of OLAP
  - Data model, operations, etc.

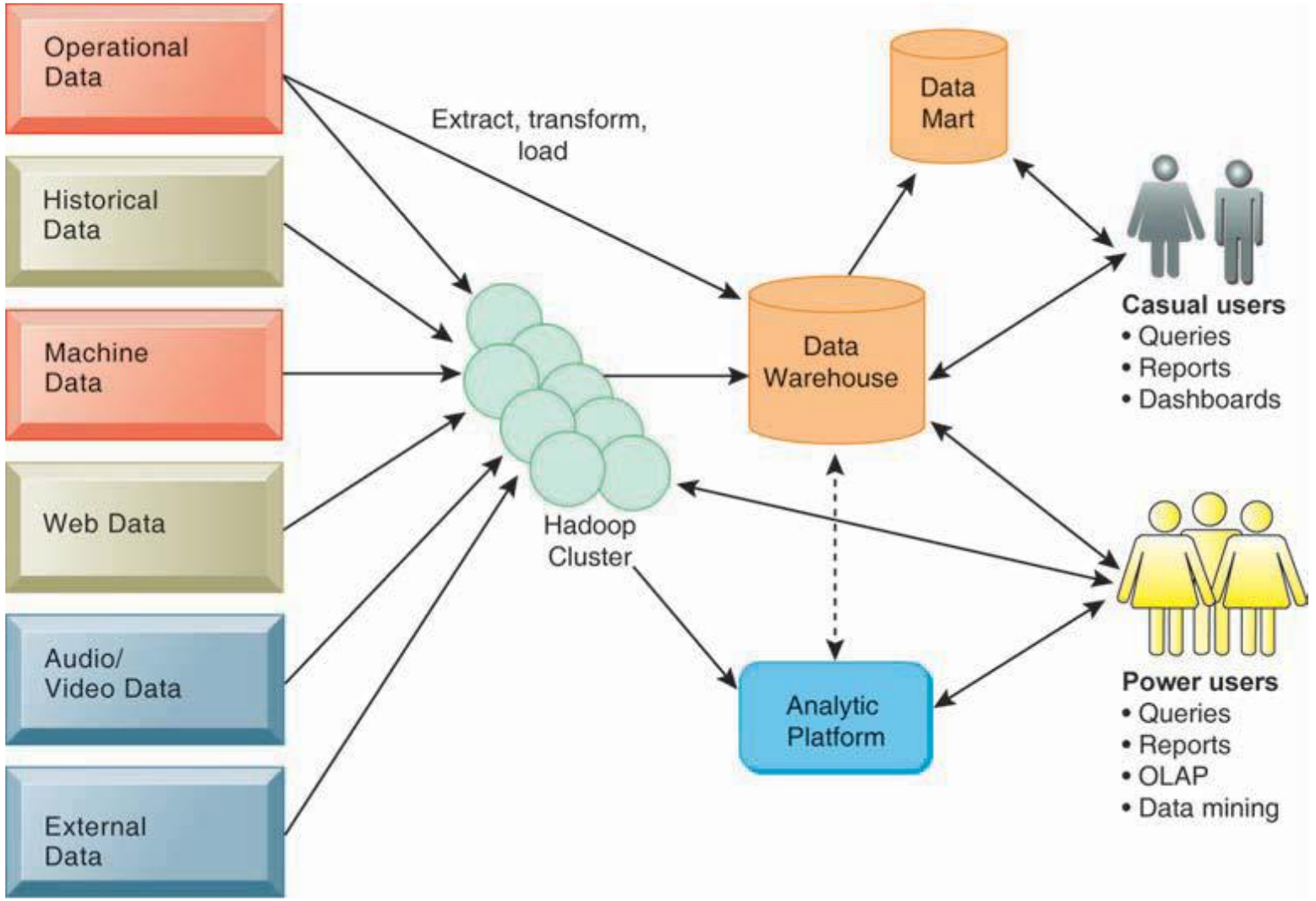
Architecture for BIS

Analytic Platforms

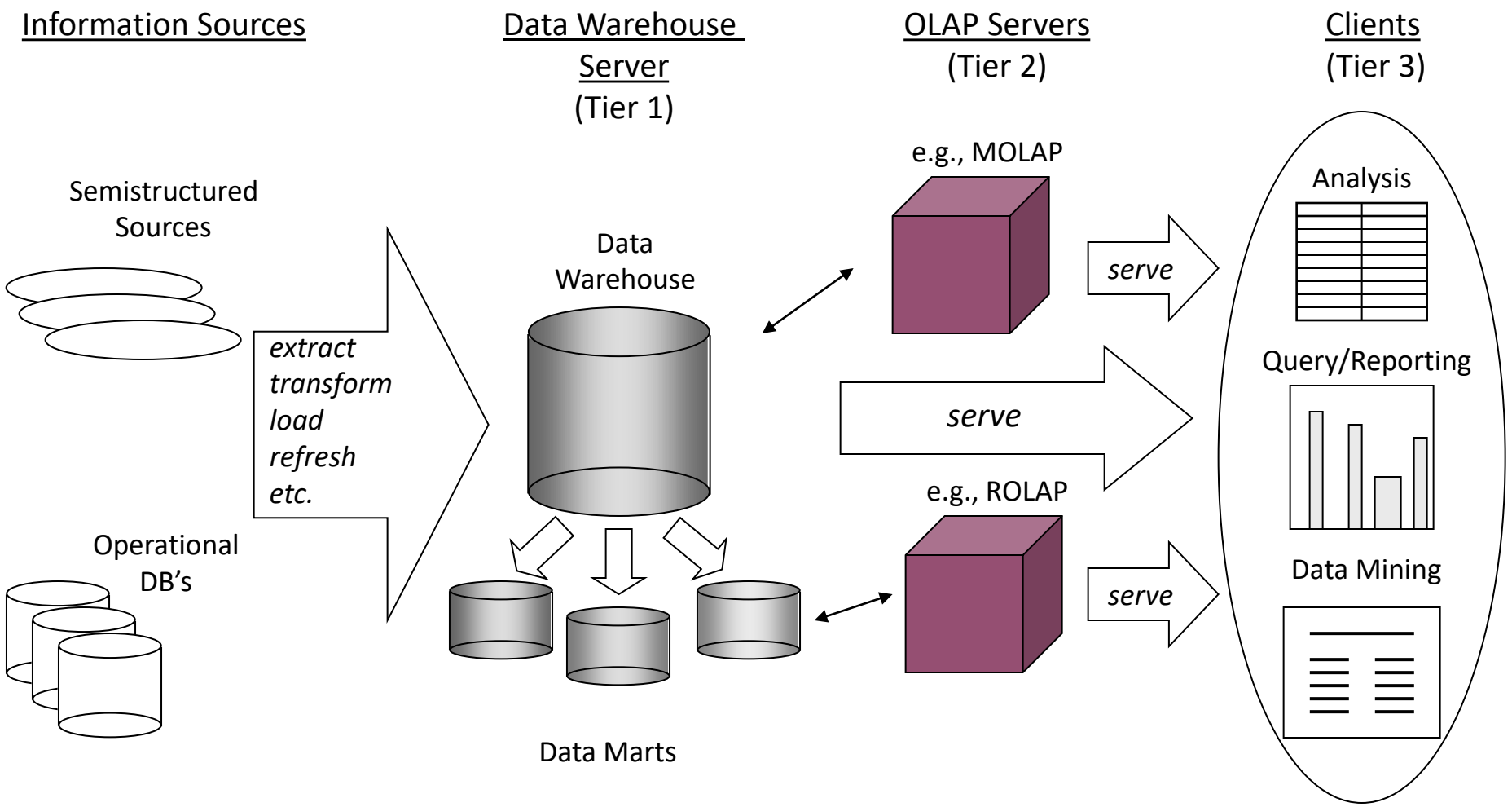
- Commercial database vendors have developed specialized high-speed analytic platforms optimized for analyzing large datasets.
- These analytic platforms, such as IBM Netezza and Oracle Exadata, feature preconfigured hardware-software systems that are specifically designed for query processing and analytics.

Hadoop

It breaks a big data problem down into sub-problems, distributes them among up to thousands of inexpensive computer processing nodes, and then combines the result into a smaller data set that is easier to analyze.



The Complete Decision Support System



Approaches of BIS

The following techniques and methodologies could be taken as approaches to BIS

- Improving reporting and analytical capabilities
- Using scorecards and dashboards
- Enterprise Reporting
- On-line Analytical Processing (OLAP) Analysis
- Advanced and Predictive Analysis
- Alerts and Proactive Notification
- Automated generation of reports with user subscriptions and "alerts" to problems and/or opportunities.

Capabilities of BIS

- Data Storage and Management
  - Data warehouse
  - Ad hoc analysis
  - Data quality
  - Data mining
- Information Delivery
  - Dashboard
  - Collaboration /search
  - Managed reporting
  - Visualization
  - Scorecard
- Query, Reporting and Analysis
  - Ad hoc Analysis
  - Production reporting
  - OLAP analysis

Benefits of BIS

- Improved Management Processes.
- Planning, controlling, measuring and/or applying changes that results in increased revenues and reduced costs.
- Improved business operations.
- Fraud detection, order processing, purchasing that results in increased revenues and reduced costs.
- Intelligent prediction of future.

Q/A