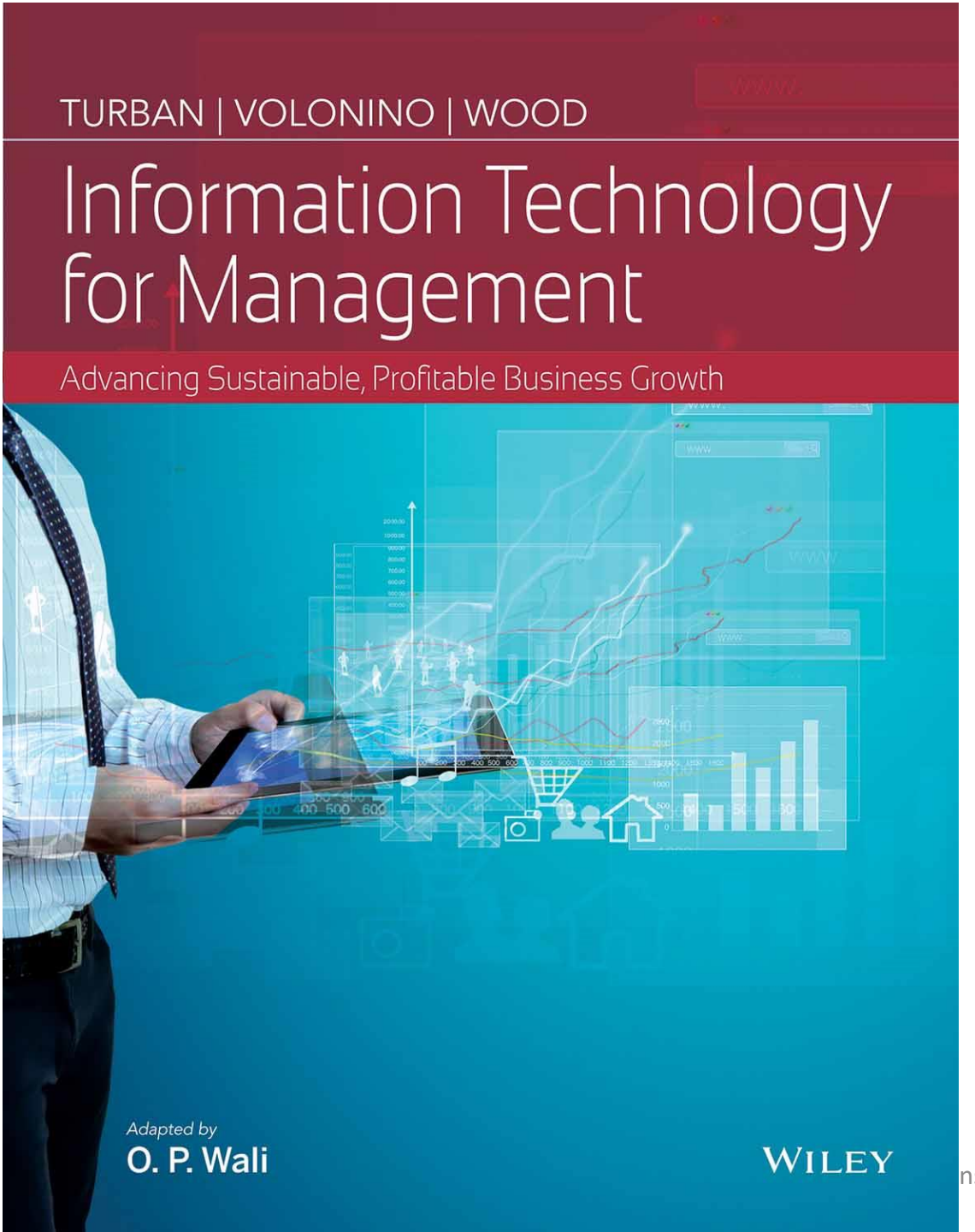


Chapter 3

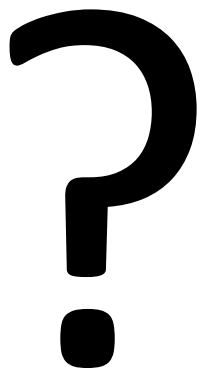
Managing Data to Improve Business Performance

Chapter Outline

1. Database Management Systems
2. Big Data and Analytics
3. Data and Text Mining
4. Business Intelligence

1. Database Management Systems

Examples of Data Usage



Examples of Data Usage

- Identify dissatisfied customers likely to defect
- Best retention incentive for them
- Detect fraudulent transactions and control fraud behavior
- Make automated recommendations for products for a customer based on the customer's profile
- Make split-second advertising decisions for online customers to avoid abandoned shopping carts.

Database Management Systems

- Databases
  - Collections of data sets or records stored in a systematic way.
  - Stores data generated by business apps, sensors, operations, and transaction-processing systems (TPS).
  - The data in databases are extremely **volatile**.
  - Medium and large enterprises typically have many databases of various types.

Volatile data changes frequently

Database Management Systems

- Data Warehouses
  - Integrate data from multiple *databases* and data silos, and organize them for complex analysis, knowledge discovery, and to support decision making.
  - May require formatting processing and/or standardization.
  - Loaded at specific times making them non-volatile and ready for analysis.

Database Management Systems

- Data Marts
  - Small-scale *data warehouses* that support a single function or one department.
  - Enterprises that cannot afford to invest in data warehousing may start with one or more data marts.

Database Management Systems

- Business Intelligence (BI)
  - Tools and techniques that process data and conduct statistical analysis for insight and discovery.
  - Used to discover meaningful relationships in the data, keep informed of real time, detect trends, and identify opportunities and risks.

Database Management Systems

- Database Management System (DBMS)
  - Integrate with data collection systems such as TPS and business applications.
  - Stores data in an organized way.
  - Provides facilities for accessing and managing data.
  - Standard database model adopted by most enterprises.
  - Store data in tables consisting of columns and rows, similar to the format of a spreadsheet.

Database Management Systems

- Relational Management System (DBMS)
  - Provides access to data using a declarative language.
- Declarative Language
  - Simplifies data access by requiring that users only specify what data they want to access without defining how they will be achieved.
  - Structured Query Language (SQL) is an example of a declarative language:

```
SELECT column_name(s)
FROM table_name
WHERE condition
```

Database Management Systems

- DBMS Functions (Tech Note 3.1)
  - Data filtering and profiling
  - Data integrity and maintenance
  - Data synchronization
  - Data security
  - Data access

Performance of DBMS – Imp Factors

- Data latency – gap between creation and availability of data
- Ability to handle the volatility of the data
- Query response time
- Data consistency
- Query predictability

Database Management Systems

Online Transaction Processing and Online Analytics Processing

- Online Transaction Processing (OLTP)
  - Designed to manage transaction data, which are volatile & break down complex information into simpler data tables to strike a balance between transaction-processing efficiency and query efficiency.
  - Cannot be optimized for data mining

Database Management Systems

Online Transaction Processing and Online Analytics Processing

- Online Analytics Processing (OLAP)
  - A means of organizing large business databases.
  - Divided into one or more cubes that fit the way business is conducted.

Database Management Systems

- DBMSs (mid-2014)
  - Oracle's MySQL
  - Microsoft's SQL Server
  - PostgreSQL
  - IBM's DB2
  - Teradata Database.

Database Management Systems

- Trend Toward NoSQL Systems (Not only)
  - Higher performance
  - Easy distribution of data on different nodes
    - enables scalability and fault tolerance
  - Greater flexibility
  - Simpler administration

Database Management Systems

Centralized and Distributed Database Architecture

- Centralized Database Architecture
  - Better control of data quality.
  - Better IT security.
- Distributed Database Architecture
  - Allow both local and remote access.
  - Use client/server architecture to process requests.

Database Management Systems

Garbage In, Garbage Out

- Dirty Data
  - Lacks integrity/validation and reduces user trust.
  - Incomplete, out of context, outdated, inaccurate, inaccessible, or overwhelming.

Cost of Poor Quality Data ≠ Lost Business ≠ Cost to Prevent Errors ≠ Cost to Correct Errors

Database Management Systems

- Principle of Diminishing Data Value
  - The value of data diminishes as they age.
  - Blind spots (lack of data availability) of 30 days or longer inhibit peak performance.
  - Global financial services institutions rely on near-real-time data for peak performance.
- Principle of 90/90 Data Use
  - As high as 90 percent, is seldom accessed after 90 days (except for auditing purposes).
  - Roughly 90 percent of data lose most of their value after 3 months.

Database Management Systems

- Principle of data in context
  - The capability to capture, process, format, and distribute data in near real time or faster requires a huge investment in data architecture.
  - The investment can be justified on the principle that data must be integrated, processed, analyzed, and formatted into “actionable information.”

Database Management Systems

Data Life Cycle

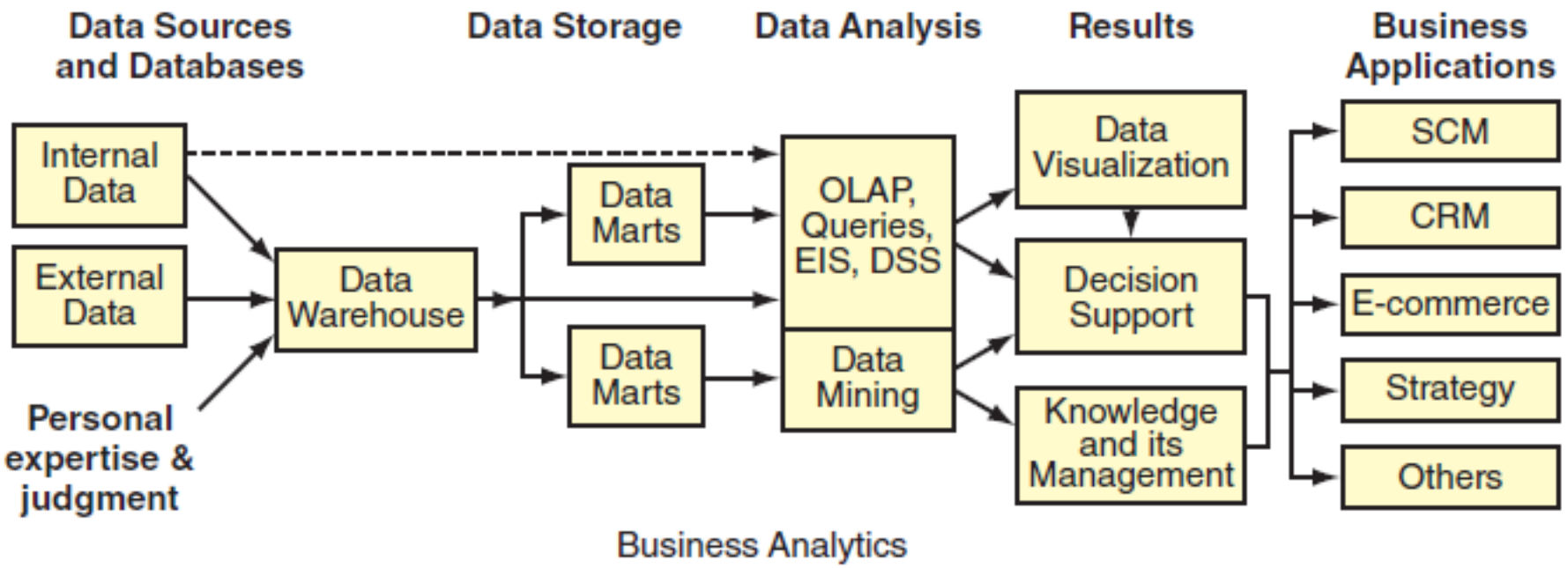


Figure 3.3 Data life cycle.

Database Management Systems

Master Reference File and Data Entities

- Data entity - Object, event, or phenomenon about which data is stored in a database.
- **Master data** represents the business objects which are agreed on and shared across the enterprise. Commonly confused with Reference Data – information that is key to the operation of a business – it also can cover transactional, unstructured, analytical, hierarchical and meta data.

Database Management Systems

- While master data is often non-transactional in nature, it is not limited to non-transactional data, and often *supports* transactional processes and operations.
- Master data may be about: customers, products, employees, materials, suppliers, and vendors, and it may also cover: sales, documents and aggregated sales.

Database Management Systems

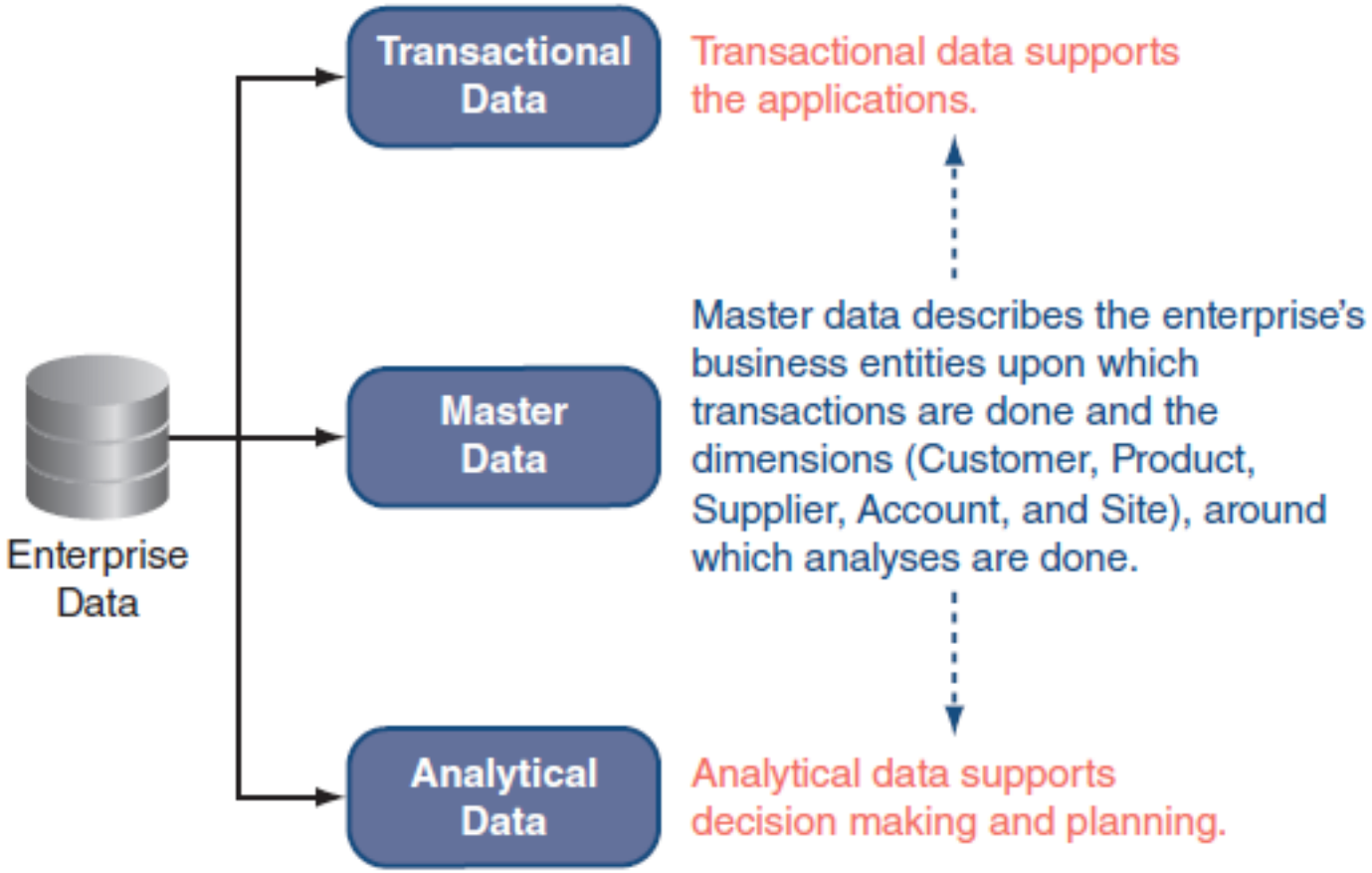


Figure 3.12 An enterprise has transactional, master, and analytical data.

2. Big Data and Analytics

Big Data and Analytics

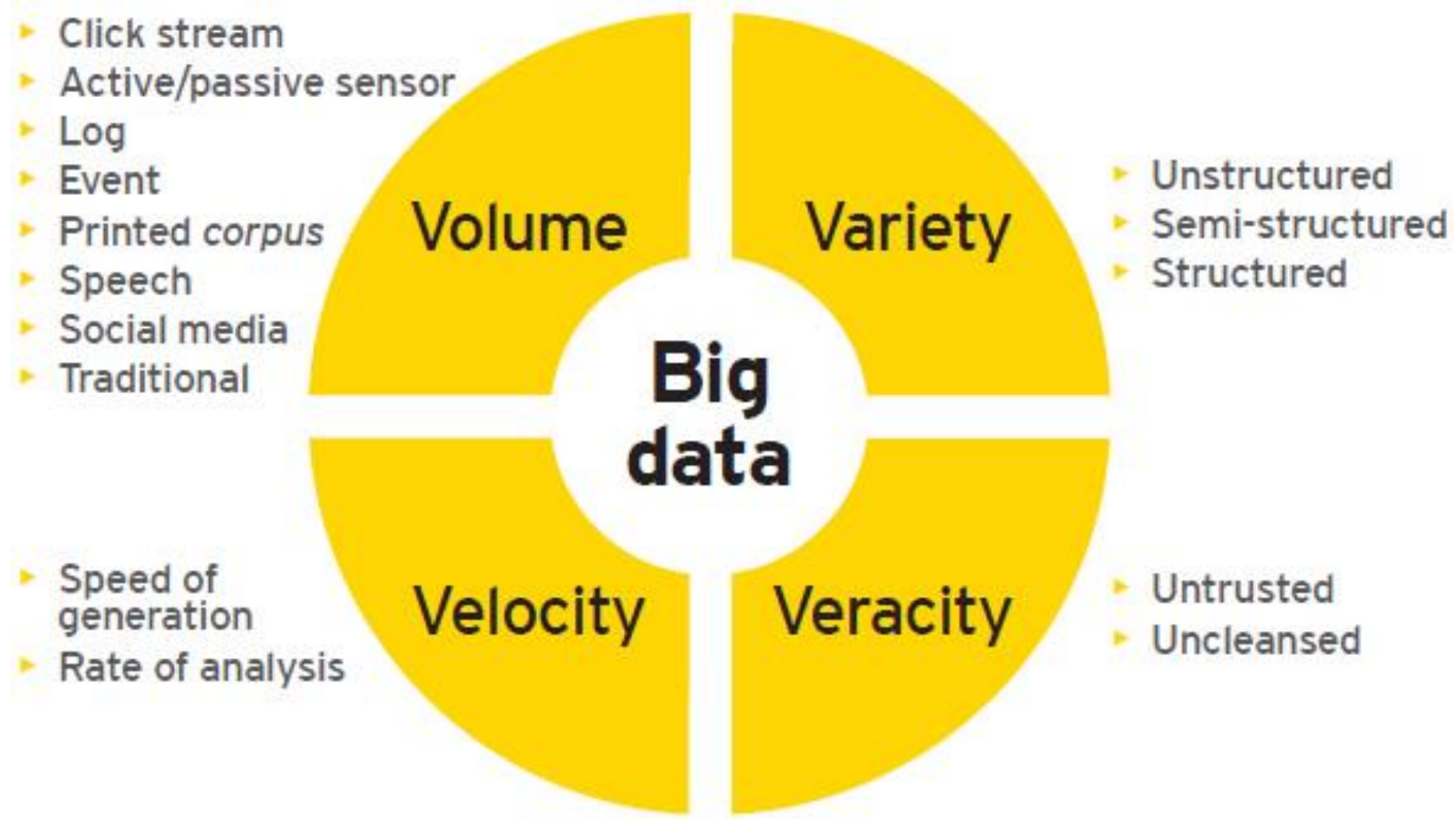
Major retailor, losing market share, learned that competitor had done:

- Invested heavily in IT to collect, integrate, and analyze data
- Linked data with suppliers:
  - To changes prices in real time
  - To reorder
  - To shift items from one store to another
- Was constantly using info from shop floor to CEO

Big Data and Analytics

- Market share
  - Percentage of total sales in a market captured by a brand, product, or company.
- Operating Margin
  - A measure of the percent of a company's revenue left over after paying variable costs: wages, raw materials, etc.
  - Increased margins mean earning more per dollar of sales.
  - The higher the operating margin, the better.

Four V's of Data Analytics



Data Warehouse and Big Data Analytics

TORTURE DATA LONG ENOUGH AND IT WILL CONFESS . . .

BUT MAY NOT TELL THE TRUTH

Case Study

Big Data and Analytics

- Human Expertise and Judgment Required
  - Data are worthless if you cannot analyze, interpret, understand, and apply the results in context.
  - Data need to be prepared for analysis.
  - Dirty data degrade the value of analytics.
  - Data must be put into meaningful context.

Big Data and Analytics

- Enterprise data warehouses (EDW)
  - Data warehouses that pull together data from disparate sources and databases across an entire.
  - Warehouses are the primary source of cleansed data for analysis, reporting, and Business Intelligence (BI).
  - Their high costs can be subsidized by using Data marts.

Big Data and Analytics

- Procedures to Prepare EDW Data for Analytics - ETL
  - Extract from designated databases.
  - Transform by standardizing formats, cleaning the data, integration.
  - Loading into a data warehouse.

Big Data and Analytics

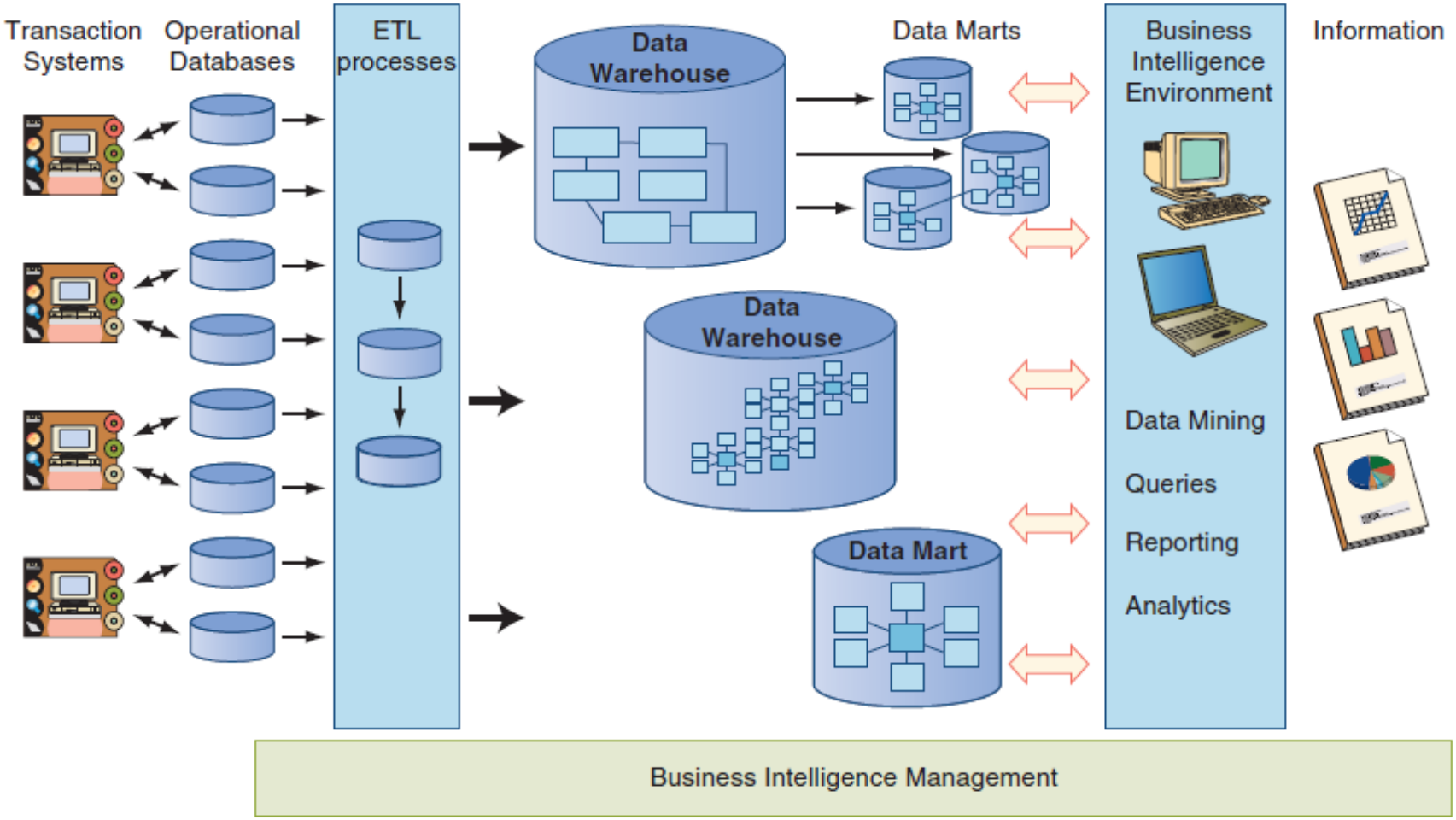


Figure 3.15 Database, data warehouse and marts, and BI architecture.

Big Data and Analytics

- Active Data Warehouse (ADW)
  - Real-time data warehousing and analytics.
  - Transform by standardizing formats, cleaning the data, integration.
- They Provide
  - Interaction with a customer to provide superior customer service.
  - Respond to business events in near real time.
  - Share up-to-date status data among merchants, vendors, customers, and associates.

Big Data and Analytics

- Supporting Actions as well as Decisions
  - Marketing and Sales
  - Pricing and Contracts
  - Forecasting
  - Sales
  - Financial

Industry Applications

TABLE 3.2 Data Warehouse Applications

Industry	Applications
Airline	Crew assignment, aircraft deployment, analysis of route profitability, customer loyalty promotions
Banking and financial services	Customer service, trend analysis, product and service promotions, reduction of IS expenses
Credit card	Customer service, new information service for a fee, fraud detection
Defense contracts	Technology transfer, production of military applications
E-Business	Data warehouses with personalization capabilities, marketing/ shopping preferences allowing for up-selling and cross-selling
Government	Reporting on crime areas, homeland security
Health care	Reduction of operational expenses
Investment and insurance	Risk management, market movements analysis, customer tendencies analysis, portfolio management
Retail chain	Trend analysis, buying pattern analysis, pricing policy, inventory control, sales promotions, optimal distribution channel decisions

Big Data and Analytics

- Really Big Data
  - Low-cost sensors collect data in real time in all types of physical things (machine-generated sensor data):
    - Regulate temperature and climate
    - Detect air particles for contamination
    - Machinery conditions/failures
    - Engine wear/maintenance

Big Data and Analytics

- Hadoop and MapReduce
  - Hadoop is an Apache processing platform that places no conditions on the processed data structure.
  - MapReduce provides a reliable, fault-tolerant software framework to write applications easily that process vast amounts of data (multi-terabyte data-sets) in-parallel on large clusters (thousands of nodes) of commodity hardware.
    - Map stage: breaks up huge data into subsets
    - Reduce stage: recombines partial results

Big Data and Analytics

1. Why are human expertise and judgment important to data analytics? Give an example.
2. What is the relationship between data quality and the value of analytics?
3. Why do data need to be put into a meaningful context?
4. What are the differences between databases and data warehouses?
5. Explain ETL and CDC.
6. What is an advantage of an active data warehouse (ADW)?
7. Why might a company invest in a data mart?
8. How can manufacturers and health care benefit from data analytics?
9. Explain how Hadoop implements MapReduce in two stages.

3. Data and Text Mining

Data and Text Mining

DBMSs support queries

vs.

Data Mining and Text Mining?

Data and Text Mining

- DBMS is a full-fledged system for housing and managing a set of digital databases.
- Data Mining is a technique or a concept of extracting useful and previously unknown information from raw data.

Data and Text Mining

- Creating Business Value
  - **Business Analytics**: the entire function of applying technologies, algorithms, human expertise, and judgment.
  - **Data Mining**: software that enables users to analyze data from various dimensions or angles, categorize them, and find correlative patterns among fields in the data warehouse.
  - **Text Mining:** broad category involving interpreted words and concepts in context.
  - **Sentimental Analysis**: trying to understand consumer intent.

Data and Text Mining

Business Value Falls in Three Buckets

- Making more informed decisions
- Discovering unknown insights, patterns, or relationships
- Automating and streamlining business process

Data and Text Mining

- Text Analytics (Mining) Procedure
  - Exploration
    - Simple word counts
    - Topics consolidation
  - Preprocessing
    - Standardization
    - May be 80% of processing time
    - Grammar and spell checking
  - Categorizing and Modelling
    - Create business rules and train models for accuracy and precision

Data and Text Mining

1. Describe data mining.
2. How does data mining generate or provide value? Give an example.
3. What is text mining?
4. Explain the text mining procedure.

4. Business Intelligence

Business Intelligence

- Key to competitive advantage
  - Across industries in all size enterprises
  - Used in operational management, business process, and decision making
  - Provides *moment of value* to decision makers
  - Unites data, technology, analytics, & human knowledge to optimize decisions

Business Intelligence

- Challenges
  - Data Selection & Quality
  - Alignment with Business Strategy and BI Strategy
- Alignment
  - Clearly articulates business strategy
  - Deconstructs business strategy into targets
  - Identifies KPIs
  - Prioritizes KPIs
  - Creates a plan based on priorities
  - Transform based on strategic results and changes

Factors for Increased use of BI

Smart devices everywhere

- Have created demand for effortless 24/7 access to insights

Data is big business

- When they provide insight that supports decisions and action

Advanced BI and analytics

- Help to ask questions that were previously unknown and unanswerable.

Cloud enabled BI and analytics

- Are providing low-cost and flexible solutions.

Business Intelligence

- BI Architecture and Analytics
  - Advances in response to big data and end-user performance demands.
  - Hosted on public or private clouds.
  - Limits IT staff and controls costs
  - May slow response time, add security and backup risks

Business Intelligence

1. How has BI improved performance management at Quicken Loans?
2. What are the business benefits of BI?
3. What are two data-related challenges that must be resolved for BI to produce meaningful insight?
4. What are the steps in a BI governance program?
5. What is a business-driven development approach?
6. What does it mean to drill down, and why is it important?
7. What four factors are contributing to increased use of BI?
8. How did BI help CarMax achieve record-setting revenue growth?

Adaptability: A Missing Ingredient

Mrs. Fields Cookies, a national chain of cookie stores, grew remarkably fast and successfully during the early 1980s. A key aspect of the company's strategy was to provide expertise directly from the headquarters to every store. As the number of stores increased, the only feasible way to achieve such direct control was through the use of information systems that were designed to mimic the decision making of Mrs. Fields herself. Decision-making systems were placed in each store.

Adaptability: A Missing Ingredient

The system would take input (such as the temperature, the day of the week, the date, and so forth), process the data, and provide, as output, information to each store manager about how many cookies of each type to bake each hour. In essence, the software provided each store manager with explicit directions for planning each day's production, sales, and labor scheduling, along with inventory control and ordering.

Adaptability: A Missing Ingredient

Because of the well-functioning computer systems, which in principle were systems designed to make Mrs. Fields' tacit knowledge available to all stores, the company was able to successfully function with few managerial levels.

However, as the market began to change and consumers became more health conscious, Mrs. Fields was very slow to respond.

Adaptability: A Missing Ingredient

In a sense, by embedding so much knowledge into systems that were incapable of adaptation, the organization tied itself to a certain way of doing things and failed to engage in knowledge creation. That is, it failed to pick up the signals in the environment, which might have suggested a change in strategy or product focus. By the early 1990s, the company had fallen into bankruptcy and sold to an investment firm.