



Data Mining and Business Intelligence

Lecture 1/DMBI/IKI83403T/MTI/UI



Yudho Giri Sucahyo, Ph.D, CISA (yudho@cs.ui.ac.id)
Faculty of Computer Science, University of Indonesia

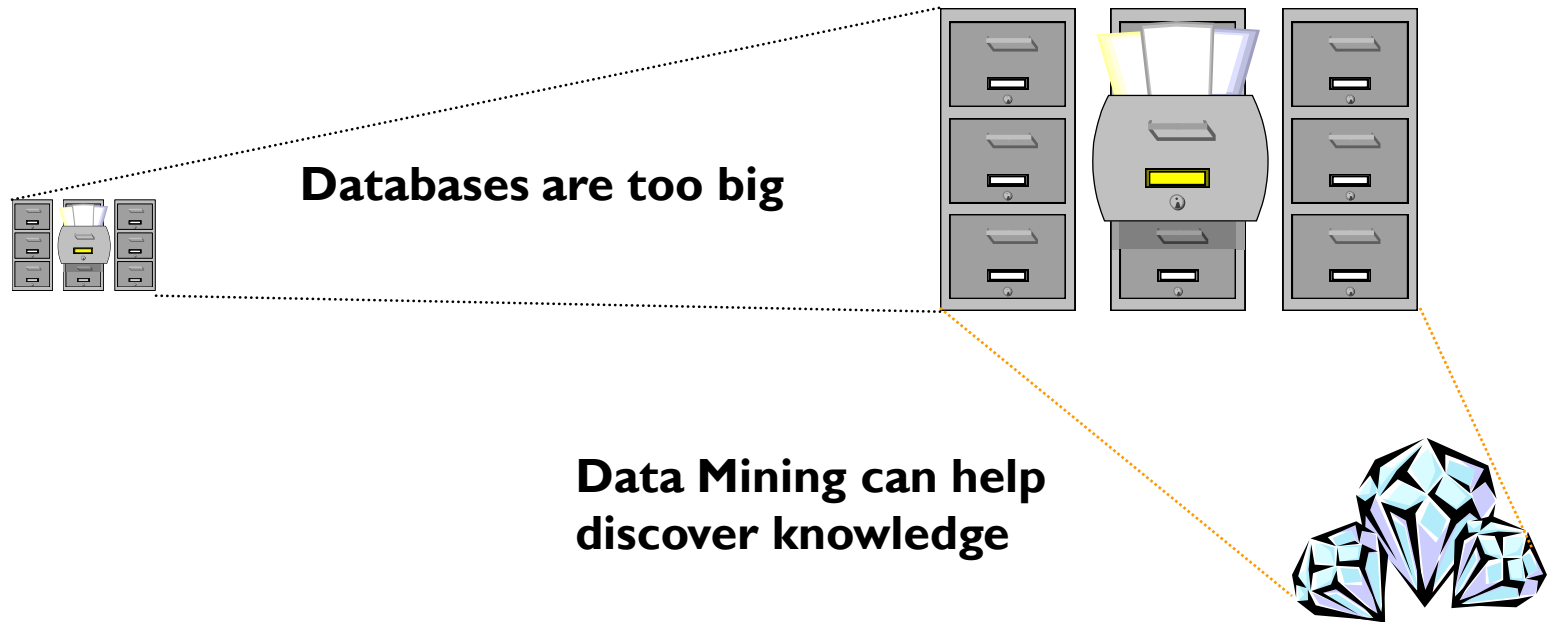
Objectives

- ▶ Motivation: Why data mining?
- ▶ What is data mining?
- ▶ Understand the drivers for BI initiatives in modern organizations
- ▶ Understand the structure, components, and process of BI

Motivation: Why data mining?

- ▶ **Data explosion problem:**
 - ▶ Automated data collection tools and mature database technology lead to tremendous amounts of data stored in databases, data warehouses and other information repositories.
- ▶ **We are drowning in data, but starving for knowledge!**
- ▶ **Data Mining:**
 - ▶ Extraction of interesting knowledge (rules, regularities, patterns, constraints) from data in **large** databases [JH].
 - ▶ Analysis of the **large** quantities of data that are stored in computers [DO],
- ▶ **Alternative names**
 - ▶ KDD, knowledge extraction, data archeology, information harvesting, **business intelligence**, etc.

Data Rich but Information Poor



Evolution of Database Technology

1960

- ▶ Data Collection, database creation, network DBMS

1970

- ▶ Relational data model, relational DBMS implementation

1980

- ▶ RDBMS, advanced data models (extended-relational, OO, etc.) and application-oriented DBMS (spatial, scientific, engineering, etc.)

1990

- ▶ Data mining and data warehousing, multimedia databases, and Web technology

Potential Applications

- ▶ See TSBD lecture notes – Data Mining
- ▶ See Chapter I of DO
 - ▶ Retailing
 - ▶ Banking
 - ▶ Credit Card Management
 - ▶ Insurance
 - ▶ Telecommunications
 - ▶ Telemarketing
 - ▶ Human Resource Management

Data Mining Should Not be Used Blindly

- ▶ Data mining find regularities from history, but history is not the same as the future.
- ▶ Association does not dictate trend nor causality.
- ▶ Some abnormal data could be caused by human.

Another view of BI

- ▶ BI is a broad field and it is viewed differently by different people.
- ▶ Common agreement on major components:
 - ▶ A centralized repository of data → *data warehouse*
 - ▶ An end-user set of tools to create reports and queries from data and information and to analyze the data, information, and reports → *business analytics*
 - ▶ To find non-obvious relationship among large amounts of data → *data mining*, for text → *text mining*, for web → *web mining*
 - ▶ Business Performance Management (BPM) to set goals as metrics and standards and monitoring and measuring performance by using the BI methodology.

Drivers of BI

- ▶ Organizations are being compelled to capture, understand, and harness their data to support decision making in order to improve business operations
- ▶ Business cycle times are now extremely compressed; faster, more informed, and better decision making is therefore a competitive imperative
- ▶ Managers need the *right information* at the *right time* and in the *right place*
- ▶ *Case Study 1: BI success story at Toyota Motor Company (Chapter 1 ET pg. 4-6).*

Business Value of BI

TABLE 11.1 Business Value of BI Analytical Applications

Analytical Application	Business Question	Business Value
Customer segmentation	What market segments do my customers fall into and what are their characteristics?	Personalize customer relationships for higher customer satisfaction and retention.
Propensity to buy	Which customers are most likely to respond to my promotion?	Target customers based on their need to increase their loyalty to your product line. Also, increase campaign profitability by focusing on the most likely to buy.
Customer profitability	What is the lifetime profitability of my customers?	Make business interaction decisions based on the overall profitability of customers or customer segments.
Fraud detection	How can I detect which transactions are likely to be fraudulent?	Quickly detect fraud and take immediate action to minimize cost.
Customer attrition	Which customers are at risk of leaving?	Prevent loss of high-value customers and let go of lower-value customers.
Channel optimization	What is the best channel to reach my customers in each segment?	Interact with customers based on their preference and your need to manage cost.

Source: Ziama and Kasher (2004). Courtesy of Teradata, division of NCR Corp.

Data Mining Functionality

▶ Association

- ▶ From association, correlation, to causality
- ▶ Finding rules like “A -> B”

▶ Classification and Prediction

- ▶ Classify data based on the values in a classifying attribute
- ▶ Predict some unknown or missing attribute values based on other information

▶ Cluster analysis

- ▶ Group data to form new classes, e.g., cluster houses to find distribution patterns

▶ Outlier and exception data analysis

▶ Time series analysis (trend and deviation)

- ▶ Trend and deviation analysis: regression, sequential pattern, similar sequences e.g. Stock analysis

Sarbanes-Oxley Act of 2002

(extracted from Gartner, Inc., 2004)

- ▶ **The Sarbanes-Oxley Act of 2002 mandates** drove one firm to implement a new financial performance management system, capable of meeting the new requirements to:
 - ▶ Perform flawless analysis and compilation of thousands of transactions and journal entries.
 - ▶ Balance more access to data with the need to control access to sensitive insider information.
 - ▶ Deliver reports to the SEC in less time.

Sarbanes-Oxley Act of 2002 (extracted from Gartner, Inc., 2004) ... continued

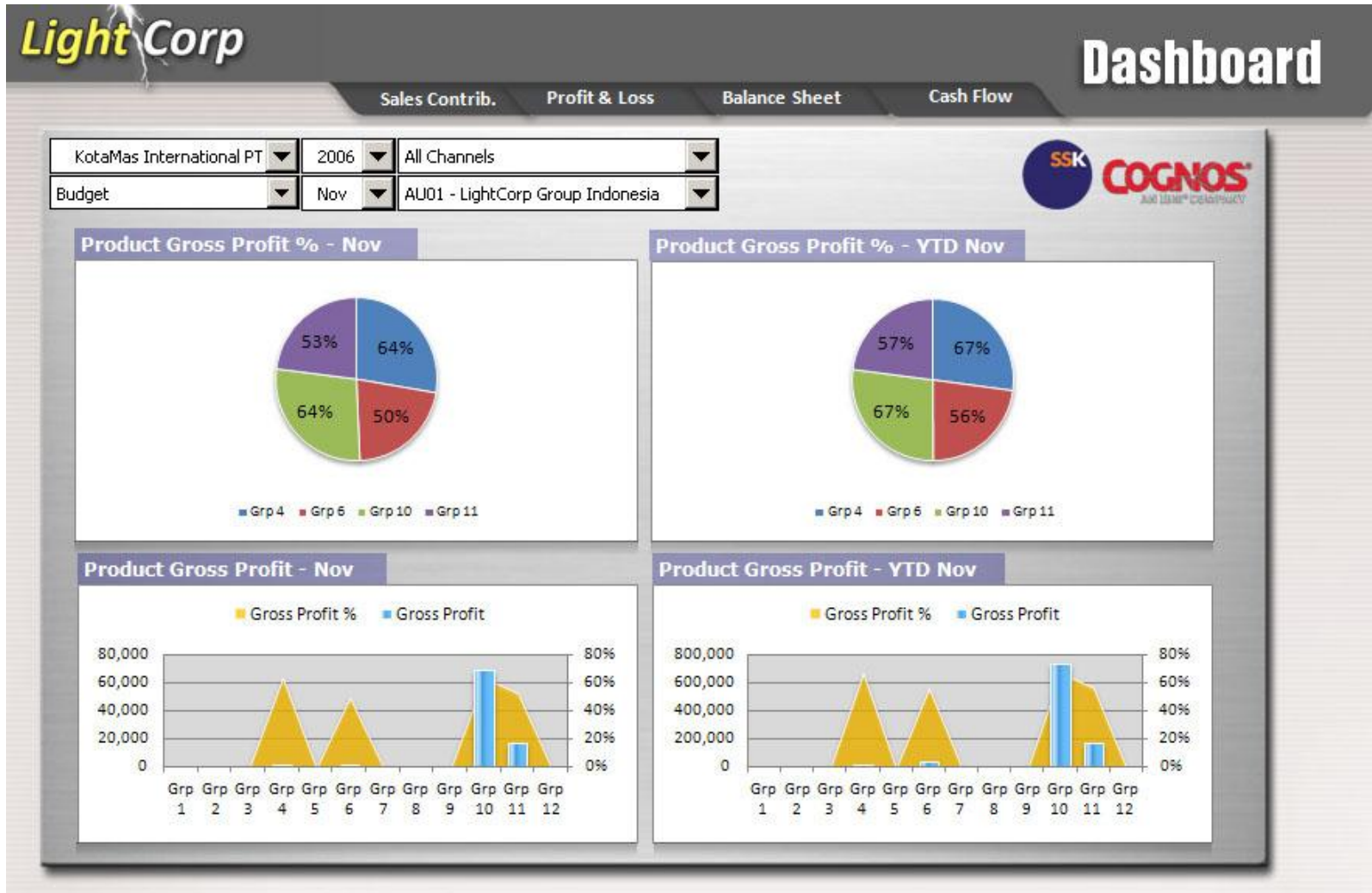
- ▶ Within the overarching goal of achieving financial-reporting compliance, these objectives included the following:
 - ▶ Get **more eyes on the data** and KPI and build in strict security controls
 - ▶ Provide live reports that allow people to drill down to the lowest level of transaction detail
 - ▶ Proactively scour the financial databases for anomalies, using variance triggers
 - ▶ Gather all financial data into a cohesive database
- ▶ Complement accounting and budgeting applications for flexible reporting, free-form investigation, and automated data analysis.
- ▶ BI can proactively alert specific individuals whenever an anomaly is detected.

Now let us see some screenshots.....

Dashboard



And another dashboard.....



And another dashboard....



Financial Reporting

Light Corp

Cash Flow

Sales Contrib.

Dashboard

Profit & Loss

Balance Sheet

Sentralasia PT

2006

Jun

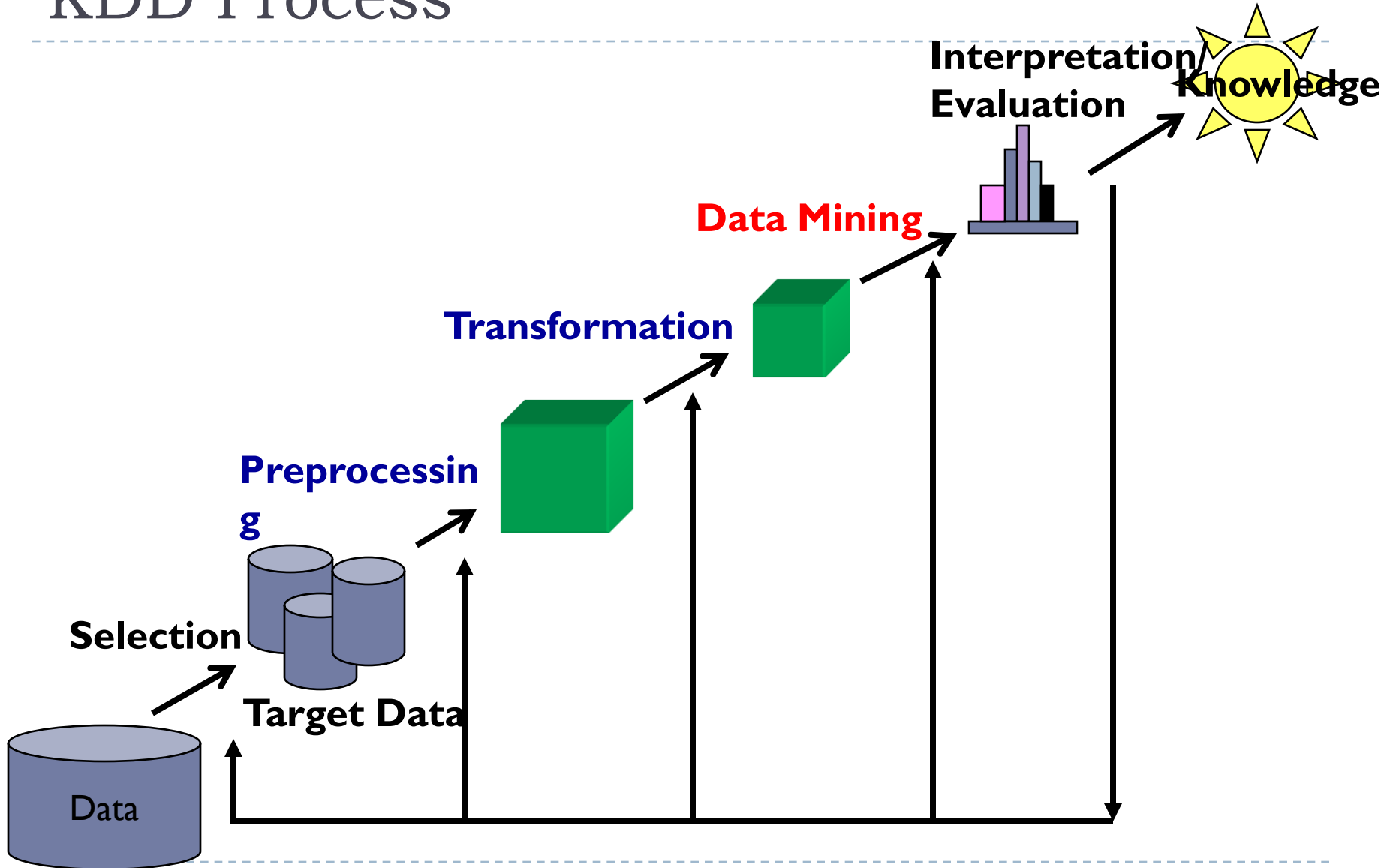
Jun 2006

YTD Jun 2006

	Jun 2006			YTD Jun 2006		
	Actual	Budget	Variance	Actual	Budget	Variance
Net cash from operating activities	12,138,530	13,559,897	(1,421,367)	62,652,755	81,381,376	(18,728,621)
EBIT	6,545,213	10,998,201	(4,452,988)	10,132,612	43,340,786	(33,208,174)
Add back:						
Amortisation (including impairment)	-	-	-	-	-	-
Depreciation (including impairment)	7,000,000	7,000,000	-	42,000,000	42,000,000	-
Depreciation (including impairment)	7,000,000	7,000,000	-	42,000,000	42,000,000	-
Gain / (Loss) on sale of fixed assets	-	-	-	12,782	11,376	1,406
Changes on Working Capital	(1,406,683)	(1,638,844)	232,161	10,520,143	6,042,826	4,477,317
Increase / (decrease) in receivables	1,658,813	1,890,973	(232,161)	(11,449,809)	(6,972,491)	(4,477,317)
Increase / (decrease) in other payables and accrued expenses	-	-	-	-	-	-
Increase / (decrease) in inventories	882,454	882,454	-	(3,253,829)	(3,253,829)	(0)
Increase / (decrease) in trade payables	1,134,584	1,134,584	-	(4,183,495)	(4,183,495)	(0)
Increase / (decrease) in other payables and accrued expenses	-	-	-	-	-	-
Increase / (decrease) in provisions	-	-	-	-	-	-
Cash flows from investing activities	-	-	-	-	-	-
Purchase of property, plant and equipment (excluding acquisitions)	-	-	-	-	-	-
Purchase of intangibles	-	-	-	-	-	-
Proceeds from sale of PPE	-	-	-	-	-	-
Proceeds from sale of intangible assets	-	-	-	-	-	-
Loan repayments received / (Loans granted)	-	-	-	-	-	-
Interest received	-	-	-	-	-	-
Cash flows from financing activities	(1,872,667)	(1,666,667)	(206,000)	88,764,000	90,000,000	(1,236,000)
Increase in minority interests	-	-	-	-	-	-
(Repayments) / Proceeds of/from borrowings	-	-	-	100,000,000	100,000,000	-
Interest paid	1,872,667	1,666,667	206,000	11,236,000	10,000,000	1,236,000
Dividends paid to shareholders IC	-	-	-	-	-	-
Opening balance - cash and cash equivalents	70,242,289	88,579,522	(18,337,254)	(70,908,623)	(70,908,623)	-
Increase / (decrease) in cash and cash equivalents	10,265,864	11,893,231	(1,627,367)	151,416,755	171,381,376	(19,964,621)
Closing balance - cash and cash equivalents	80,508,132	100,472,753	(19,964,621)	80,508,132	100,472,753	(19,964,621)

Back to theory..... 😊

KDD Process



Steps of a KDD Process

- ▶ Learning the application domain:
 - ▶ relevant prior knowledge and goals of application
- ▶ Creating a target data set: data selection
- ▶ Data cleaning and preprocessing: (may take 60% of effort!)
- ▶ Data reduction and projection:
 - ▶ Find useful features, dimensionality/variable reduction, invariant representation.
- ▶ Choosing functions of data mining
 - ▶ summarization, classification, regression, association, clustering.
- ▶ Choosing the mining algorithm(s)
- ▶ Data mining: search for patterns of interest
- ▶ Interpretation: analysis of results.
 - ▶ visualization, transformation, removing redundant patterns, etc.
- ▶ Use of discovered knowledge

Teradata Advanced Analytics Methodology (similar to CRISP-DM)

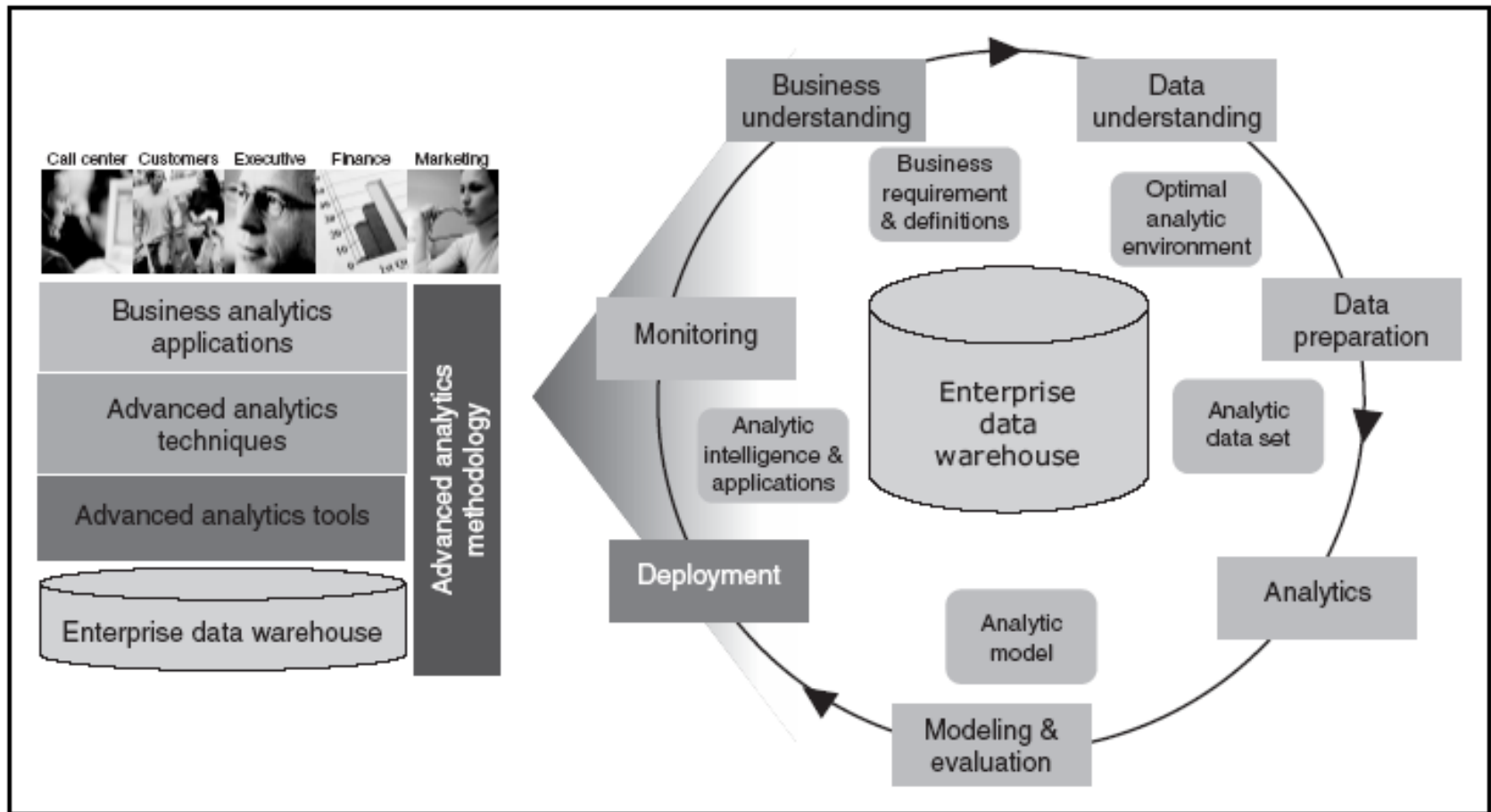
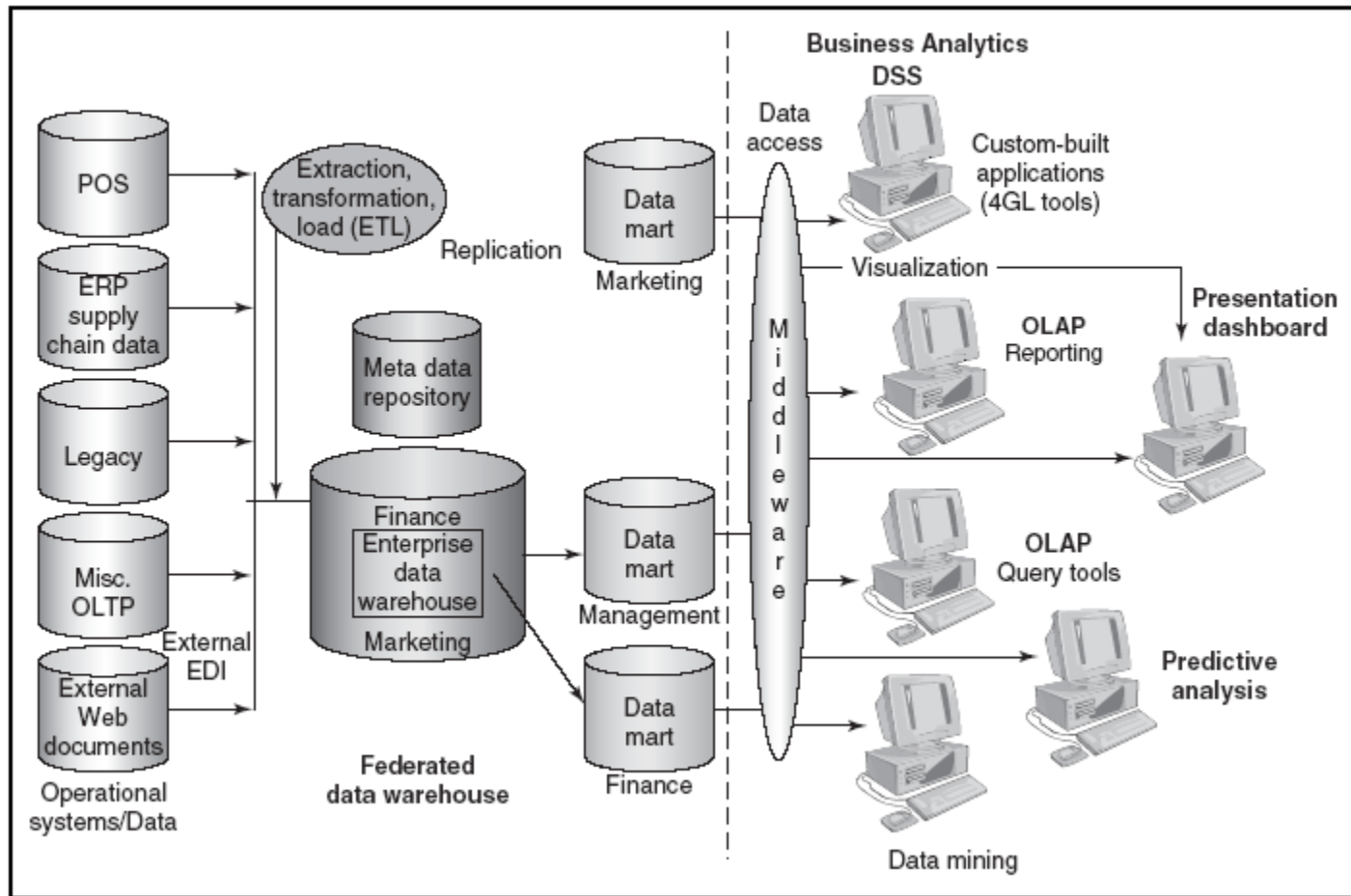


FIGURE S.3 Teradata Advanced Analytics Methodology

Structure and Components of BI

FIGURE 5.4 The Major Components of Business Intelligence



Structure and Components of BI... continued

▶ Data Warehouse

- ▶ Data flows from operational systems (e.g., CRM, ERP) to a DW, which is a special database or repository of data that has been prepared to support decision-making applications ranging from those for simple reporting and querying to complex optimization

▶ Business Analytics/OLAP

- ▶ Software tools that allow users to create on-demand reports and queries and to conduct analysis of data

Structure and Components of BI... continued

▶ Data Mining

- ▶ *Data mining* is a class of database information analysis that looks for hidden patterns in a group of data that can be used to predict future behavior
- ▶ Used to replace or enhance human intelligence by scanning through massive storehouses of data to discover meaningful new correlations, patterns, and trends, by using pattern recognition technologies and advanced statistics

Structure and Components of BI... continued

▶ **Business Performance Management (BPM)**

- ▶ Based on the *balanced scorecard* methodology—a framework for defining, implementing, and managing an enterprise's business strategy by linking objectives with factual measures

▶ **Dashboards**

A visual presentation of critical data for executives to view. It allows executives to see hot spots in seconds and explore the situation

BI: Today and Tomorrow

- ▶ Recent industry analyst reports show that in the coming years, millions of people will use BI visual tools and analytics every day
- ▶ BI takes advantage of already developed and installed components of IT technologies, helping companies leverage their current IT investments and use valuable data stored in legacy and transactional systems
- ▶ Some Issues:
 - ▶ Mining information from heterogeneous databases and global information systems
 - ▶ Handling relational and complex types of data
 - ▶ Efficiency and scalability of data mining algorithms