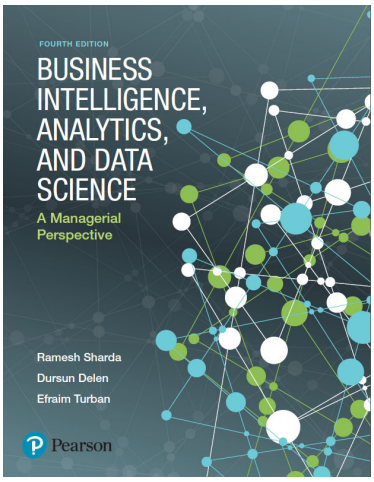
Business Intelligence, Analytics, and Data Science: A Managerial Perspective

Fourth Edition



Chapter 3

Descriptive Analytics II: Business Intelligence and Data Warehousing

Pearson

Learning Objectives (1 of 2)

3.1 Understand the basic definitions and concepts of data warehousing

3.2 Understand data warehousing architectures

3.3 Describe the processes used in developing and managing data warehouses

3.4 Explain data warehousing operations

3.5 Explain the role of data warehouses in decision support

Learning Objectives (2 of 2)

3.6 Explain data integration and the extraction, transformation, and load (ETL) processes

3.7 Understand the essence of business performance management (BPM)

3.8 Learn balanced scorecard and Six Sigma as performance measurement systems

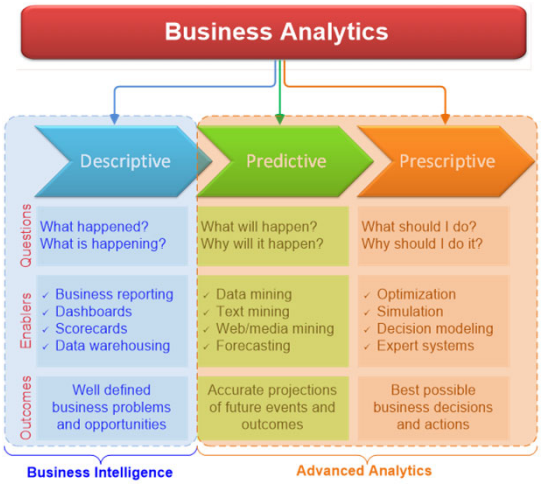
OPENING VIGNETTE

Targeting Tax Fraud with Business Intelligence and Data Warehousing

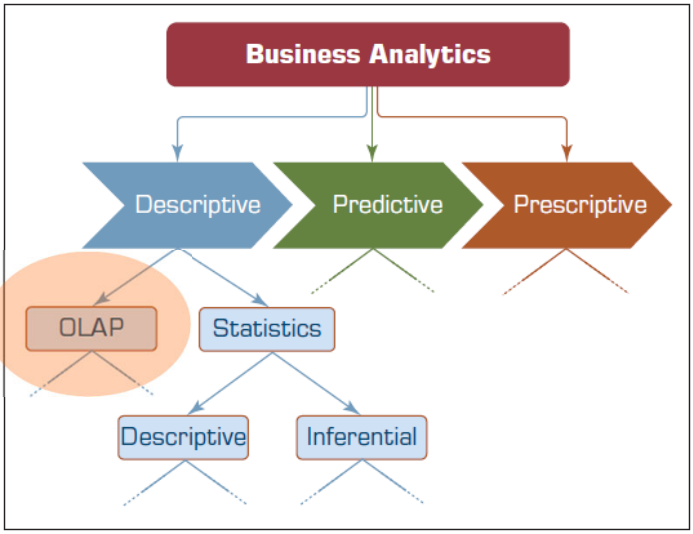
1. Why is it important for IRS and for U.S. state governments to use data warehousing and business intelligence (BI) tools in managing state revenues?
2. What were the challenges the state of Maryland was facing with regard to tax fraud?
3. What was the solution they adopted? Do you agree with their approach? Why?
4. What were the results that they obtained? Did the investment in BI and data warehousing pay off?
5. What other problems and challenges do you think federal and state governments are having that can benefit from BI and data warehousing?

Business Intelligence and Data Warehousing

- BI used to be everything related to use of data for managerial decision support
- Now, it is a part of Business Analytics
  - BI = Descriptive Analytics



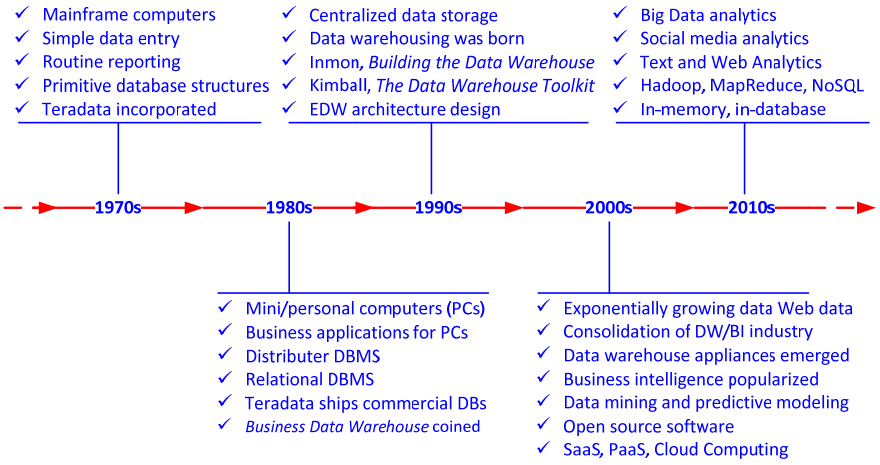
Statistical Modeling for Business Analytics



What is a Data Warehouse?

- A physical repository where relational data are specially organized to provide enterprise-wide, cleansed data in a standardized format
- A relational database? (so what is the difference?)
- "The data warehouse is a collection of integrated, subject-oriented databases designed to support DSS functions, where each unit of data is non-volatile and relevant to some moment in time"

A Historical Perspective to Data Warehousing



Characteristics of DWs

- Subject oriented
- Integrated
- Time-variant (time series)
- Nonvolatile
- Summarized
- Not normalized
- Metadata
- Web based, relational/multi-dimensional
- Client/server, real-time/right-time/active...

Data Mart

A departmental small-scale "DW" that stores only limited/relevant data

– Dependent data mart
  A subset that is created directly from a data warehouse

– Independent data mart
  A small data warehouse designed for a strategic business unit or a department

Other DW Components

- Operational data stores (ODS)
  - A type of database often used as an interim area for a data warehouse
- Oper marts
  - An operational data mart
- Enterprise data warehouse (EDW)
  - A data warehouse for the enterprise
- Metadata – "data about data"
  - In DW metadata describe the contents of a data warehouse and its acquisition and use

Application Case 3.1

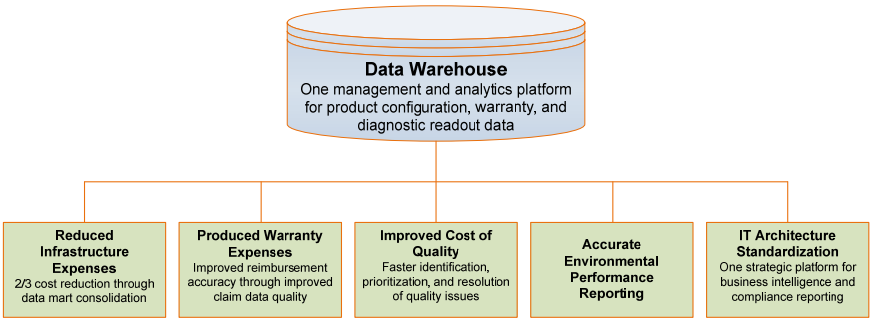
A Better Data Plan: Well-Established TELCOs Leverage Data Warehousing and Analytics to Stay on Top in a Competitive Industry

Questions for Discussion

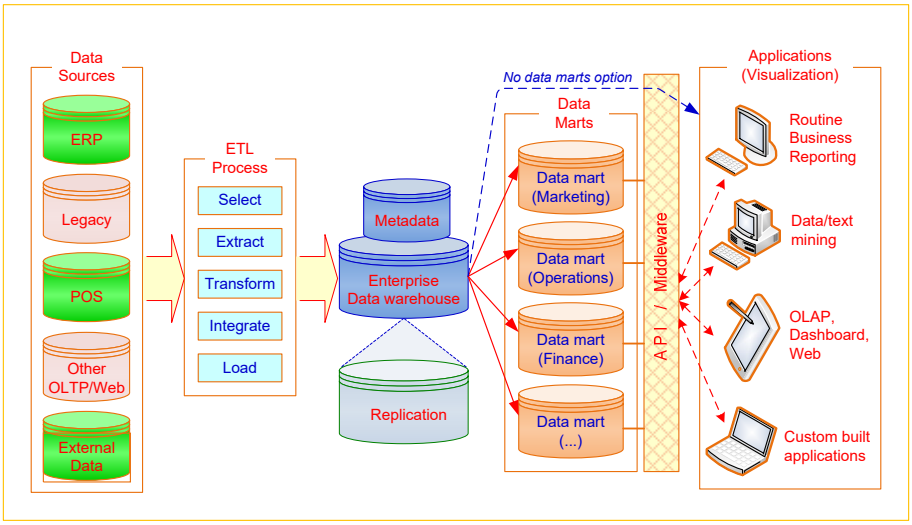
1. What are the main challenges for TELCOs?
2. How can data warehousing and data analytics help TELCOs in overcoming their challenges?
3. Why do you think TELCOs are well suited to take full advantage of data analytics?

DW for Data-Driven Decision Making

- An example of a DW supporting data-driven decision making in automotive industry



A Generic DW Framework



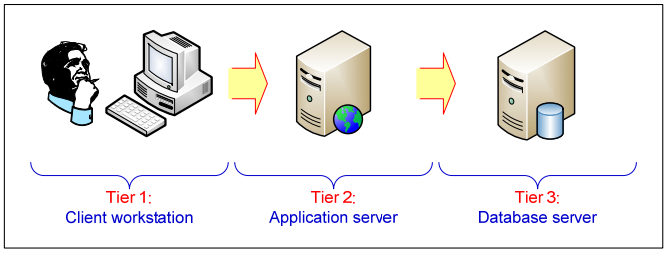
DW Architecture

- Three-tier architecture
  1. Data acquisition software (back-end)
  2. The data warehouse that contains the data & software
  3. Client (front-end) software that allows users to access and analyze data from the warehouse
- Two-tier architecture
  - First two tiers in three-tier architecture are combined into one

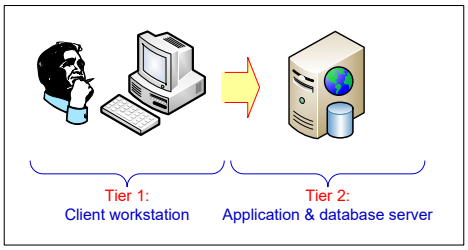
… sometimes there is only one tier?

DW Architectures

3-tier architecture



2-tier architecture

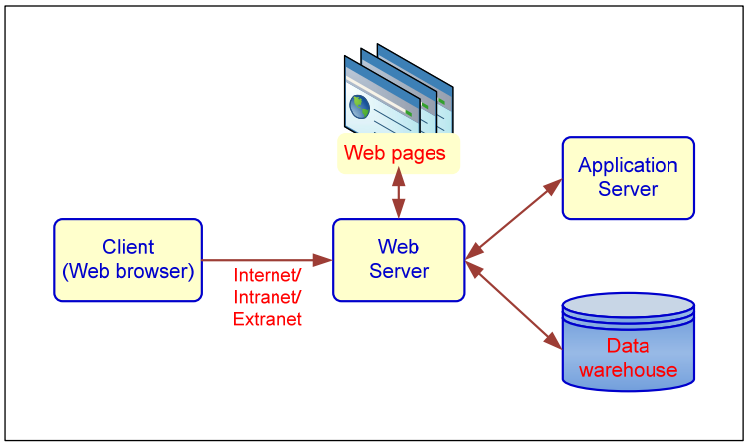


1-tier Architecture ?

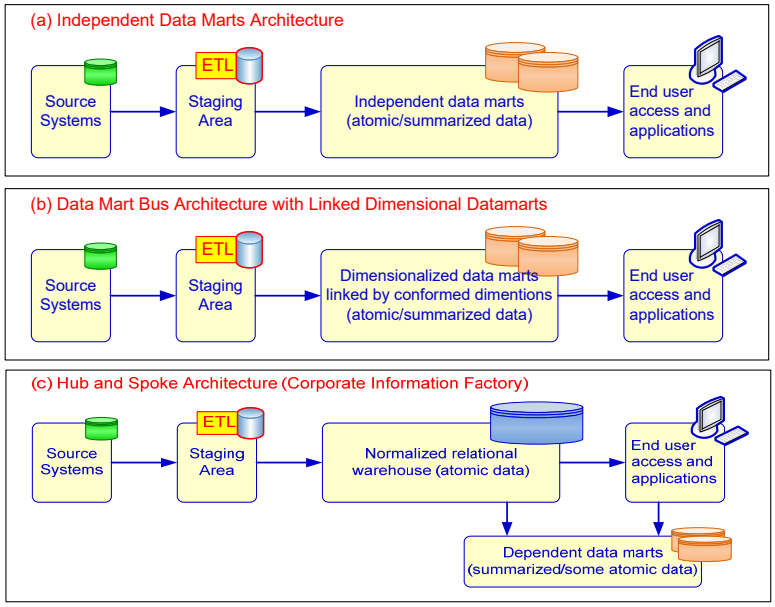
Data Warehousing Architectures

- Issues to consider when deciding which architecture to use:
  - Which database management system (DBMS) should be used?
  - Will parallel processing and/or partitioning be used?
  - Will data migration tools be used to load the data warehouse?
  - What tools will be used to support data retrieval and analysis?

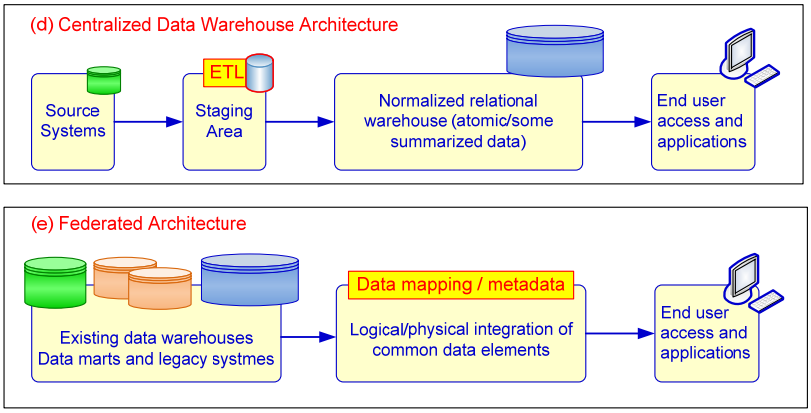
A Web-based DW Architecture



Alternative DW Architectures



Alternative DW Architectures



- Each architecture has advantages and disadvantages!
- Which architecture is the best?

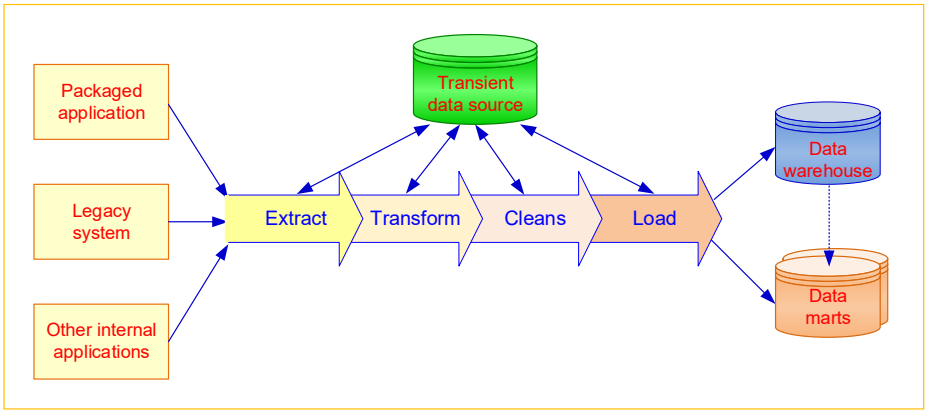
Ten Factors that Potentially Affect the Architecture Selection Decision

1. Information interdependence between organizational units
2. Upper management's information needs
3. Urgency of need for a data warehouse
4. Nature of end-user tasks
5. Constraints on resources
6. Strategic view of the data warehouse prior to implementation
7. Compatibility with existing systems
8. Perceived ability of the in-house IT staff
9. Technical issues
10. Social/political factors

Data Integration and the Extraction, Transformation, and Load Process

- ETL = Extract Transform Load
- Data integration
  - Integration that comprises three major processes: data access, data federation, and change capture.
- Enterprise application integration (EAI)
  - A technology that provides a vehicle for pushing data from source systems into a data warehouse
- Enterprise information integration (EII)
  - An evolving tool space that promises real-time data integration from a variety of sources, such as relational or multidimensional databases, Web services, etc.

Data Integration and the Extraction, Transformation, and Load Process



ETL (Extract, Transform, Load)

- Issues affecting the purchase of an ETL tool
  - Data transformation tools are expensive
  - Data transformation tools may have a long learning curve
- Important criteria in selecting an ETL tool
  - Ability to read from and write to an unlimited number of data sources/architectures
  - Automatic capturing and delivery of metadata
  - A history of conforming to open standards
  - An easy-to-use interface for the developer and the functional user

Application Case 3.2
BP Lubricants Achieves BIGS Success

Questions for Discussion

1. What is BIGS?
2. What were the challenges, the proposed solution, and the obtained results with BIGS?

Data Warehouse Development

- Data warehouse development approaches
  - Inmon Model: EDW approach (top-down)
  - Kimball Model: Data mart approach (bottom-up)
  - Which model is best?
- Table 3.3 provides a comparative analysis between EDW and Data Mart approach
- Another alternative is the hosted data warehouses

Comparing EDW and Data Mart

TABLE 3.3 Contrasts between the DM and EDW Development Approaches

Effort	DM Approach	EDW Approach
Scope	One subject area	Several subject areas
Development time	Months	Years
Development cost	$10,000 to $100,000+	$1,000,000+
Development difficulty	Low to medium	High
Data prerequisite for sharing	Common (within business area)	Common (across enterprise)
Sources	Only some operational and external systems	Many operational and external systems
Size	Megabytes to several gigabytes	Gigabytes to petabytes
Time horizon	Near-current and historical data	Historical data
Data transformations	Low to medium	High
Update frequency	Hourly, daily, weekly	Weekly, monthly
Technology		
Hardware	Workstations and departmental servers	Enterprise servers and mainframe computers
Operating system	Windows and Linux	Unix, Z/OS, OS/390
Databases	Workgroup or standard database servers	Enterprise database servers
Usage		
Number of simultaneous users	10s	100s to 1,000s
User types	Business area analysts and managers	Enterprise analysts and senior executives
Business spotlight	Optimizing activities within the business area	Cross-functional optimization and decision making

Application Case 3.3

Use of Teradata Analytics for SAP Solutions Accelerates Big Data Delivery

Questions for Discussion

1. What were the challenges faced by the large Dutch retailer?
2. What was the proposed multivendor solution? What were the implementation challenges?
3. What were the lessons learned?

Additional DW Considerations Hosted Data Warehouses

- Benefits:
  - Requires minimal investment in infrastructure
  - Frees up capacity on in-house systems
  - Frees up cash flow
  - Makes powerful solutions affordable
  - Enables solutions that provide for growth
  - Offers better quality equipment and software
  - Provides faster connections
  - … more in the book

Representation of Data in DW

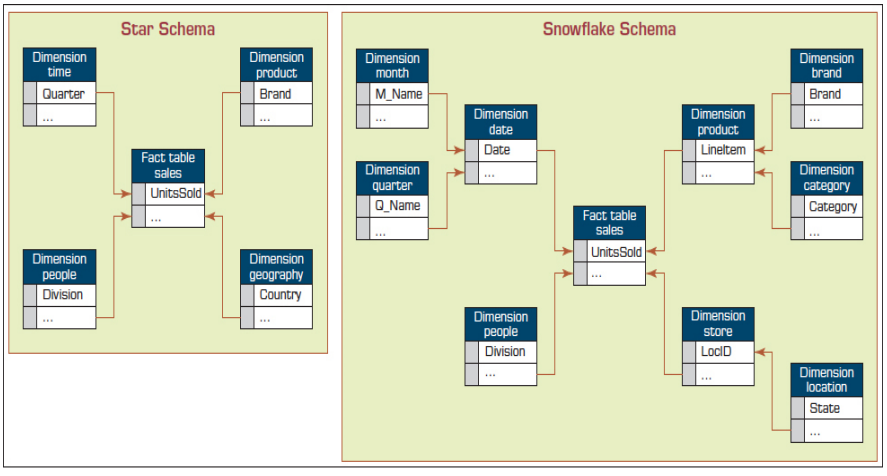
- Dimensional Modeling
  - A retrieval-based system that supports high-volume query access
- Star schema
  - The most commonly used and the simplest style of dimensional modeling
  - Contain a fact table surrounded by and connected to several dimension tables
- Snowflakes schema
  - An extension of star schema where the diagram resembles a snowflake in shape

Multidimensionality

The ability to organize, present, and analyze data by several dimensions, such as sales by region, by product, by salesperson, and by time (four dimensions)

- Multidimensional presentation
  - Dimensions: products, salespeople, market segments, business units, geographical locations, distribution channels, country, or industry
  - Measures: money, sales volume, head count, inventory profit, actual versus forecast
  - Time: daily, weekly, monthly, quarterly, or yearly

Star Schema versus Snowflake Schema



Analysis of Data in DW

- OLTP vs. OLAP…
- OLTP (Online Transaction Processing)
  - Capturing and storing data from ERP, CRM, POS, …
  - The main focus is on efficiency of routine tasks
- OLAP (Online Analytical Processing)
  - Converting data into information for decision support
  - Data cubes, drill-down / rollup, slice & dice, …
  - Requesting ad hoc reports
  - Conducting statistical and other analyses
  - Developing multimedia-based applications
  - …more in the book

OLAP vs. OLTP

TABLE 3.5 A Comparison between OLTP and OLAP

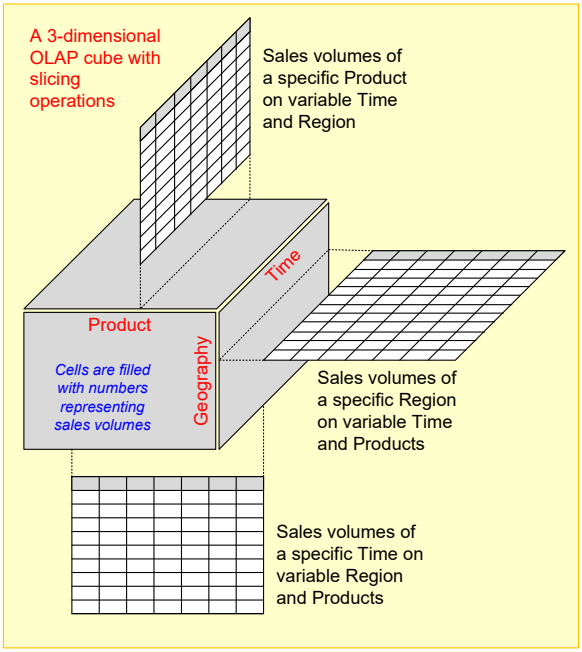
Criteria	OLTP	OLAP
Purpose	To carry out day-to-day business functions	To support decision making and provide answers to business and management queries
Data source	Transaction database (a normalized data repository primarily focused on efficiency and consistency)	Data warehouse or DM (a nonnormalized data repository primarily focused on accuracy and completeness)
Reporting	Routine, periodic, narrowly focused reports	Ad hoc, multidimensional, broadly focused reports and queries
Resource requirements	Ordinary relational databases	Multiprocessor, large-capacity, specialized databases
Execution speed	Fast (recording of business transactions and routine reports)	Slow (resource intensive, complex, large-scale queries)

OLAP Operations

- Slice - a subset of a multidimensional array
- Dice - a slice on more than two dimensions
- Drill Down/Up - navigating among levels of data ranging from the most summarized (up) to the most detailed (down)
- Roll Up - computing all of the data relationships for one or more dimensions
- Pivot - used to change the dimensional orientation of a report or an ad hoc query-page display

OLAP

Slicing Operations on a Simple Tree-Dimensional Data Cube



Successful DW Implementation Things to Avoid

- Starting with the wrong sponsorship chain
- Setting expectations that you cannot meet
- Engaging in politically naive behavior
- Loading the data warehouse with information just because it is available
- Believing that data warehousing database design is the same as transactional database design
- … more in the book

Massive DW and Scalability

- Scalability
  - The main issues pertaining to scalability:
    - The amount of data in the warehouse
    - How quickly the warehouse is expected to grow
    - The number of concurrent users
    - The complexity of user queries
  - Good scalability means that queries and other data-access functions will grow linearly with the size of the warehouse

Application Case 3.4
EDW Helps Connect State Agencies in Michigan

Questions for Discussion

1. Why would a state invest in a large and expensive IT infrastructure (such as an EDW)?
2. What is the size and complexity of the EDW used by state agencies in Michigan?
3. What were the challenges, the proposed solution, and the obtained results of the EDW?

DW Administration and Security

- Data warehouse administrator (DWA)
  - DWA should…
    - have the knowledge of high-performance software, hardware, and networking technologies
    - possess solid business knowledge and insight
    - be familiar with the decision-making processes so as to suitably design/maintain the data warehouse structure
    - possess excellent communications skills
- Security and privacy is a pressing issue in DW
  - Safeguarding the most valuable assets
  - Government regulations (HIPAA, etc.)
  - Must be explicitly planned and executed

The Future of DW

- Sourcing…
  - *Web, social media, and Big Data*
  - Open source software
  - SaaS (software as a service)
  - Cloud computing
  - Data lakes
- Infrastructure…
  - Columnar
  - Real-time DW
  - Data warehouse appliances
  - Data management practices/technologies
  - In-database & In-memory processing New DBMS
  - New DBMS, Advanced analytics, …

Data Lakes

- Unstructured data storage technology for Big Data
- Data Lake versus Data Warehouse

TABLE 3.6 A Simple Comparison between a Data Warehouse and a Data Lake

Dimension	Data Warehouse	Data Lake
The nature of data	Structured, processed	Any data in raw/native format
Processing	Schema-on-write (SQL)	Schema-on-read (NoSQL)
Retrieval speed	Very fast	Slow
Cost	Expensive for large data volumes	Designed for low-cost storage
Agility	Less agile, fixed configuration	Highly agile, flexible configuration
Novelty/newness	Not new/matured	Very new/maturing
Security	Well-secured	Not yet well-secured
Users	Business professionals	Data scientists

Business Performance Management

- Business Performance Management (BPM) is…

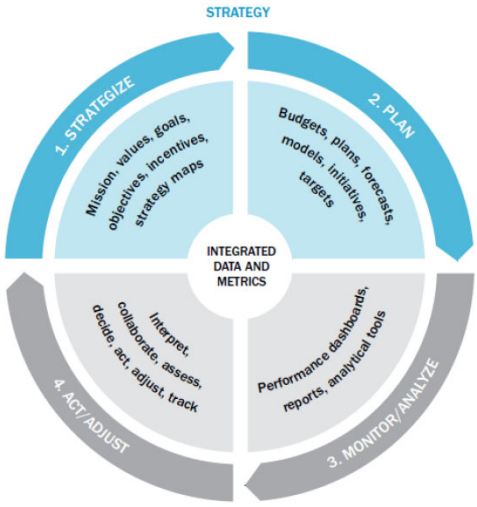
  A real-time system that alerts managers to potential opportunities, impending problems, and threats, and then empowers them to react through models and collaboration

- Also called corporate performance management (CPM by Gartner Group), enterprise performance management (EPM by Oracle), strategic enterprise management (SEM by SAP)

Business Performance Management

- BPM refers to the business processes, methodologies, metrics, and technologies used by enterprises to measure, monitor, and manage business performance.
- BPM encompasses three key components
  - A set of integrated, closed-loop management and analytic processes, supported by technology …
  - Tools for businesses to define strategic goals and then measure/manage performance against them
  - Methods and tools for monitoring key performance indicators (KPIs), linked to organizational strategy

A Closed-Loop Process to Optimize Business Performance



- Process Steps
  1. Strategize
  2. Plan
  3. Monitor/analyze
  4. Act/adjust

Each with its own sub-process steps

1 - Strategize: Where Do We Want to Go?

- Strategic planning
  - Common tasks for the strategic planning process:
    1. Conduct a current situation analysis
    2. Determine the planning horizon
    3. Conduct an environment scan
    4. Identify critical success factors
    5. Complete a gap analysis
    6. Create a strategic vision
    7. Develop a business strategy
    8. Identify strategic objectives and goals

2 - Plan: How Do We Get There?

- Operational planning
  - Operational plan: plan that translates an organization's strategic objectives and goals into a set of well-defined tactics and initiatives, resource requirements, and expected results for some future time period (usually a year).
- Operational planning can be
  - Tactic-centric (operationally focused)
  - Budget-centric plan (financially focused)

3 - Monitor/Analyze: How Are We Doing?

- A comprehensive framework for monitoring performance should address two key issues:
  - What to monitor?
    - Critical success factors
    - Strategic goals and targets
  - How to monitor?

4 - Act and Adjust: What Do We Need to Do Differently?

- Success (or mere survival) depends on new projects: creating new products, entering new markets, acquiring new customers (or businesses), or streamlining some process.
- Many new projects and ventures fail!
- What is the chance of failure?
  - 60% of Hollywood movies fail
  - 70% of large IT projects fail, …

Application Case 3.5

AARP Transforms Its BI Infrastructure and Achieves a 347% ROI in Three Years

Questions for Discussion

1. What were the challenges AARP was facing?
2. What was the approach for a potential solution?
3. What were the results obtained in the short term, and what were the future plans?

Performance Measurement

- **Performance measurement system**

  A system that assists managers in tracking the implementations of business strategy by comparing actual results against strategic goals and objectives
  - Comprises systematic comparative methods that indicate progress (or lack thereof) against goals

KPIs and Operational Metrics

- **Key performance indicator (KPI)**

  A KPI represents a strategic objective and metrics that measure performance against a goal
- Distinguishing features of KPIs
  - Strategy
  - Targets
  - Ranges
  - Encodings
  - Time frames
  - Benchmarks

Performance Measurement

- **Key performance indicator (KPI)**

  Outcome KPIs vs. Driver KPIs

  (lagging indicators (leading indicators

  e.g., revenues) e.g., sales leads)

- Operational areas covered by driver KPIs
  - Customer performance
  - Service performance
  - Sales operations
  - Sales plan/forecast

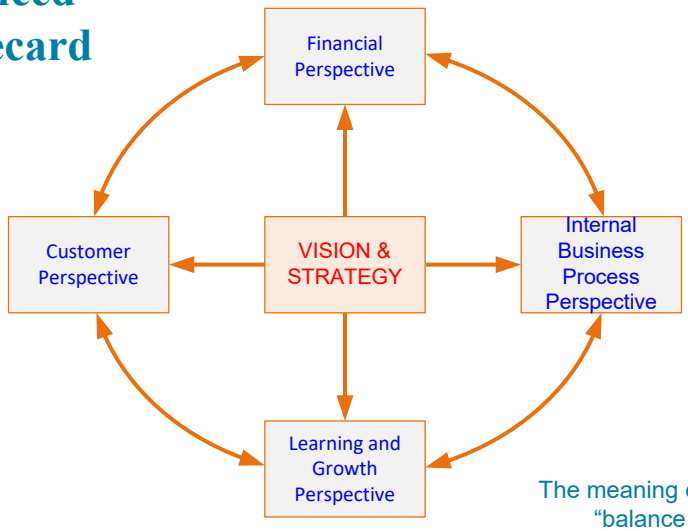
Performance Measurement System

- **Balanced Scorecard (BSC)**

  A performance measurement and management methodology that helps translate an organization's financial, customer, internal process, and learning and growth objectives and targets into a set of actionable initiatives

"The Balanced Scorecard: Measures That Drive Performance" *(HBR, 1992)*

Balanced Scorecard



The meaning of "balance" ?

Six Sigma as a Performance Measurement System

- **Six Sigma**

  A performance management methodology aimed at reducing the number of defects in a business process to as close to zero defects per million opportunities (DPMO) as possible

Six Sigma as a Performance Measurement System

- The DMAIC performance model

  A closed-loop business improvement model that encompasses the steps of defining, measuring, analyzing, improving, and controlling a process

- Lean Six Sigma
  - Lean manufacturing / lean production
  - Lean production versus six sigma?

Comparison of BSC and Six Sigma

TABLE 3.7 Comparison of the Balanced Scorecard and Six Sigma

Balanced Scorecard	Six Sigma
Strategic management system	Performance measurement system
Relates to the longer-term view of the business	Provides snapshot of business's performance and identifies measures that drive performance toward profitability
Designed to develop a balanced set of measures	Designed to identify a set of measurements that impact profitability
Identifies measurements around vision and values	Establishes accountability for leadership for wellness and profitability
Critical management processes are to clarify vision/strategy, communicate, plan, set targets, align strategic initiatives, and enhance feedback	Includes all business processes—management and operational
Balances customer and internal operations without a clearly defined leadership role	Balances management and employees' roles; balances costs and revenue of heavy processes
Emphasizes targets for each measurement	Emphasizes aggressive rate of improvement for each measurement, irrespective of target
Emphasizes learning of executives based on feedback	Emphasizes learning and innovation at all levels based on process feedback; enlists all employees' participation
Focuses on growth	Focuses on maximizing profitability
Heavy on strategic content	Heavy on execution for profitability
Management system consisting of measures	Measurement system based on process management

Effective Performance Measurement Should

- Measures should focus on key factors.
- Measures should be a mix of past, present, and future.
- Measures should balance the needs of shareholders, employees, partners, suppliers, and other stakeholders.
- Measures should start at the top and flow down to the bottom.
- Measures need to have targets that are based on research and reality rather than arbitrary.

Application Case 3.6

Expedia.com's Customer Satisfaction Scorecard

Questions for Discussion

1. Who are the customers for Expedia.com? Why is customer satisfaction a very important part of their business?
2. How did Expedia.com improve customer satisfaction with scorecards?
3. What were the challenges, the proposed solution, and the obtained results?

Plenty of Resources for DW @ TUN

- Teradata University Network (TUN)

End of Chapter 3

- Questions / Comments