Business Intelligence and Data Analytics (IS 350)

Lecture 2

An Overview of Business Intelligence and Analytics

Dr. Abdalla sayed

2023 - 2024

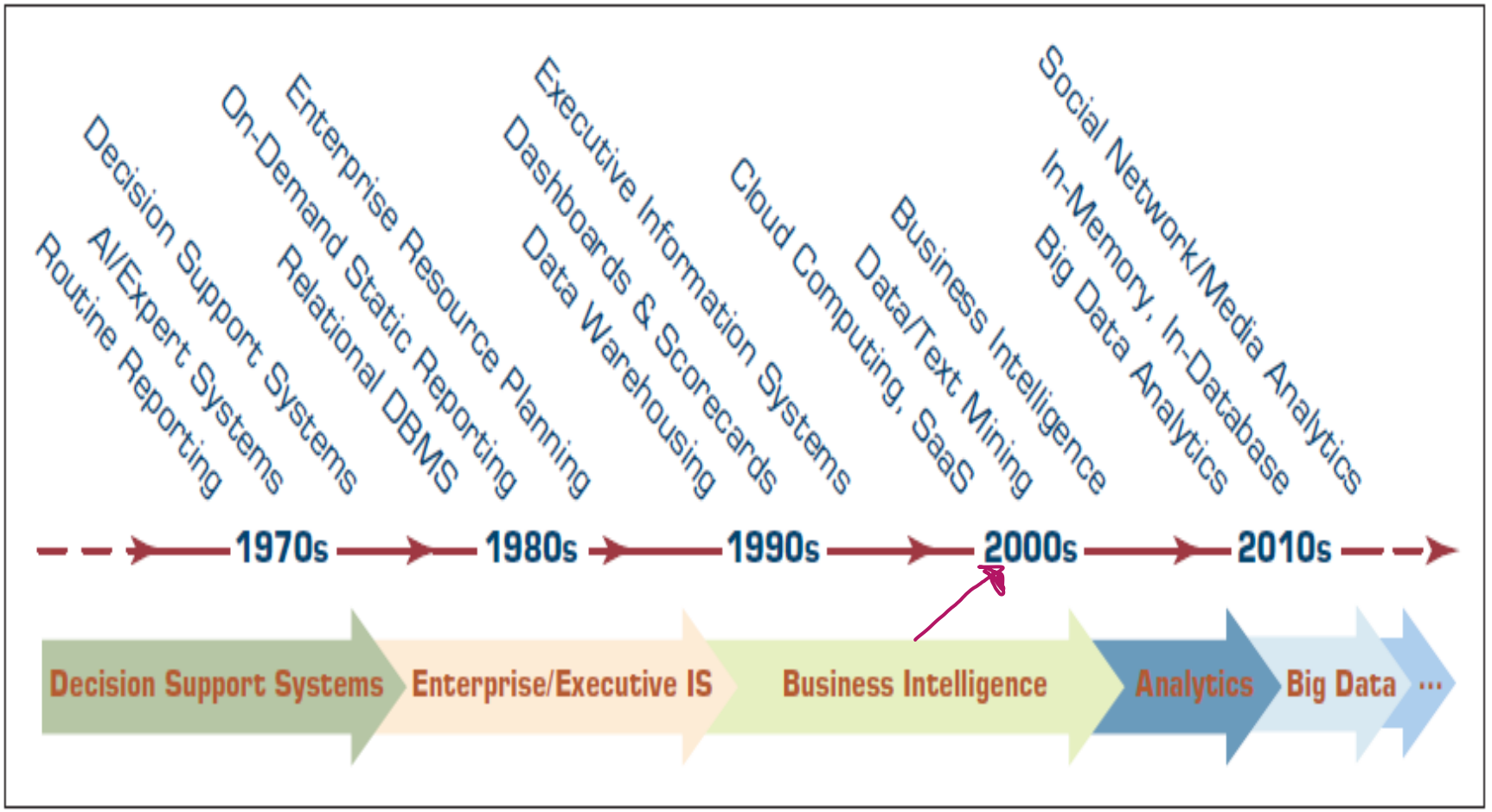
What is business intelligence (BI)

- **Business intelligence (BI)** is an umbrella term that combines architectures, tools, databases, analytical tools, applications, and methodologies.
- It is, like DSS, a content-free expression, so it means different things to different people.
- BI's major **objective** is to enable interactive access (sometimes in **real time**) to data, to enable manipulation of data, and to give business managers and analysts the ability to conduct appropriate analyses

What is business intelligence (BI)

- By analyzing historical and current data, situations, and performances, decision makers get valuable **insights** that enable them to make more informed and better **decisions**.
- The process of BI is based on the **transformation** of data to **information**, then to **decisions**, and finally to **actions**.

business intelligence history

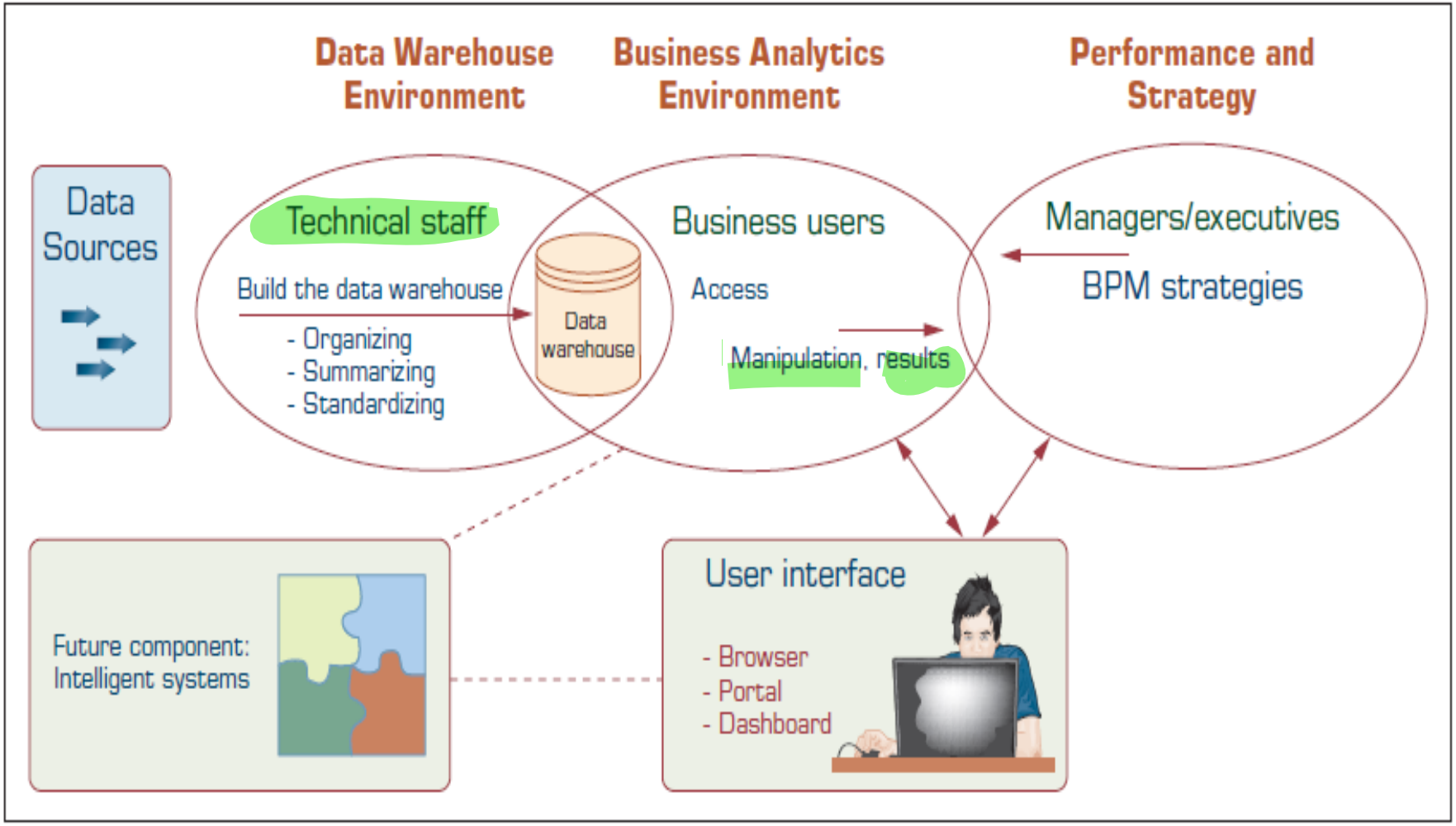


The Architecture of BI

A BI system has four major components

1. Data warehouse(DW), with its source data
2. Business Analytics, a collection of tools for manipulating, mining, and analyzing the data in the data warehouse;
3. Business Performance Management (BPM) for monitoring and analyzing performance
4. a user interface (e.g., dashboard)

The Architecture of BI



The Architecture of BI

- The data warehouse is a large repository of well-organized historical data.
- Business analytics are the tools that allow transformation of data into information and knowledge
- Business performance management (BPM) allows monitoring, measuring, and comparing **key performance indicators (KPI)**
- User interface (e.g., dashboards) allows access and easy manipulation of other BI components

Styles of BI

- MicroStrategy, Corp. distinguishes five styles of BI and offers tools for each
  1. report delivery and alerting
  2. enterprise reporting (using dashboards and scorecards)
  3. cube analysis (also known as slice-and-dice analysis)
  4. ad-hoc queries
  5. statistics and data mining

The benefits of BI

- The ability to provide accurate **information** when needed, including a real-time view of the corporate performance and its parts
- A survey by Thompson (2004)
  - **Faster**, more accurate reporting (81%)
  - **Improved decision making** (78%)
  - **Improved customer service** (56%)
  - **Increased revenue** (49%)

Transaction Processing vs Analytic Processing

- ***Transaction Processing* Systems** are constantly involved in handling updates to what we might call ***operational databases***.
- For example, in an ATM withdrawal transaction, we need to reduce our bank balance accordingly; a bank deposit adds to an account; and a grocery store purchase is likely reflected in the store's calculation of total sales for the day, and it should reflect an appropriate reduction in the store's inventory for the items we bought, and so on.

Transaction Processing vs Analytic Processing

- **Online Transaction Processing (OLTP)** systems handle a company's routine ongoing business.
- In contrast, a **DW** is typically a distinct system that provides storage for data that will be used for analysis. So , DWs are intended to work with informational data used for **online analytical processing (OLAP)** systems.
- OLTP is a term used for a transaction system that is primarily responsible for capturing and storing data related to **day-to-day** business functions such as ERP, CRM, SCM, POS, and so forth.

Transaction Processing vs Analytic Processing

- An **OLTP** system addresses a critical business need, automating daily business transactions, and running real-time reports and routine analysis. But these systems are not designed for ad hoc analysis and complex queries that deal with a number of data items.**OLAP**, on the other hand, is designed to address this need by providing ad hoc analysis of organizational data much more effectively and efficiently.
- **OLAP** and **OLTP** rely heavily on each other: OLAP uses the data captured by OLTP, and OLTP automates the business processes that are managed by decisions supported by OLAP

Normalization vs Denormalization

- **Normalization** is the method of arranging the data in the database efficiently. It involves constructing tables and setting up relationships between those tables according to some certain rules. The redundancy and inconsistent dependency can be removed using these rules in order to make it more flexible.
- There are 6 defined normal forms: 1NF, 2NF, 3NF, BCNF, 4NF and 5NF. Normalization should eliminate the redundancy but not at the cost of integrity

Normalization vs Denormalization

- **Denormalization** is the inverse process of normalization, where the normalized schema is converted into a schema which has redundant information.
- The performance is improved by using redundancy and keeping the redundant data consistent.
- **The reason for performing Denormalization** is the overheads produced in query processor by an over-normalized structure.

Key differences between Normalization vs Denormalization

- Normalization is the technique of dividing the data into multiple tables to reduce data redundancy and inconsistency and to achieve data integrity. On the other hand, Denormalization is the technique of combining the data into a single table to make data retrieval faster.
- Normalization is used in **OLTP** system, which emphasizes on making the insert, delete and update anomalies faster. As against, Denormalization is used in **OLAP** system, which emphasizes on making the search and analysis faster.

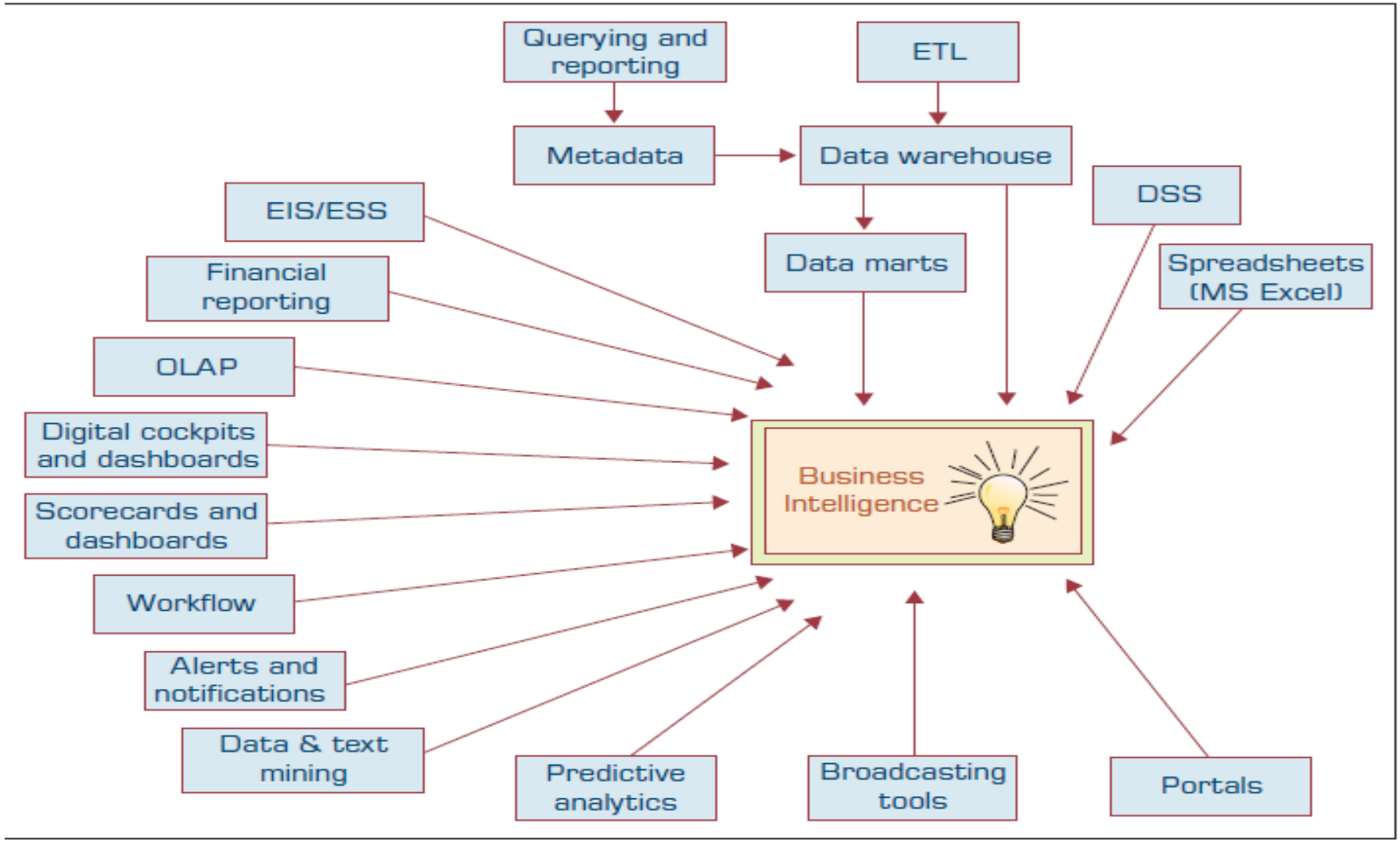
Key differences between Normalization vs Denormalization

- Data integrity is maintained in normalization process while in Denormalization data integrity harder to retain.
- Redundant data is eliminated when normalization is performed whereas Denormalization increases the redundant data.
- Normalization increases the number of tables and joins. In contrast, Denormalization reduces the number of tables and join.
- Disk space is wasted in Denormalization because same data is stored in different places. On the contrary, disk space is optimized in a normalized table.

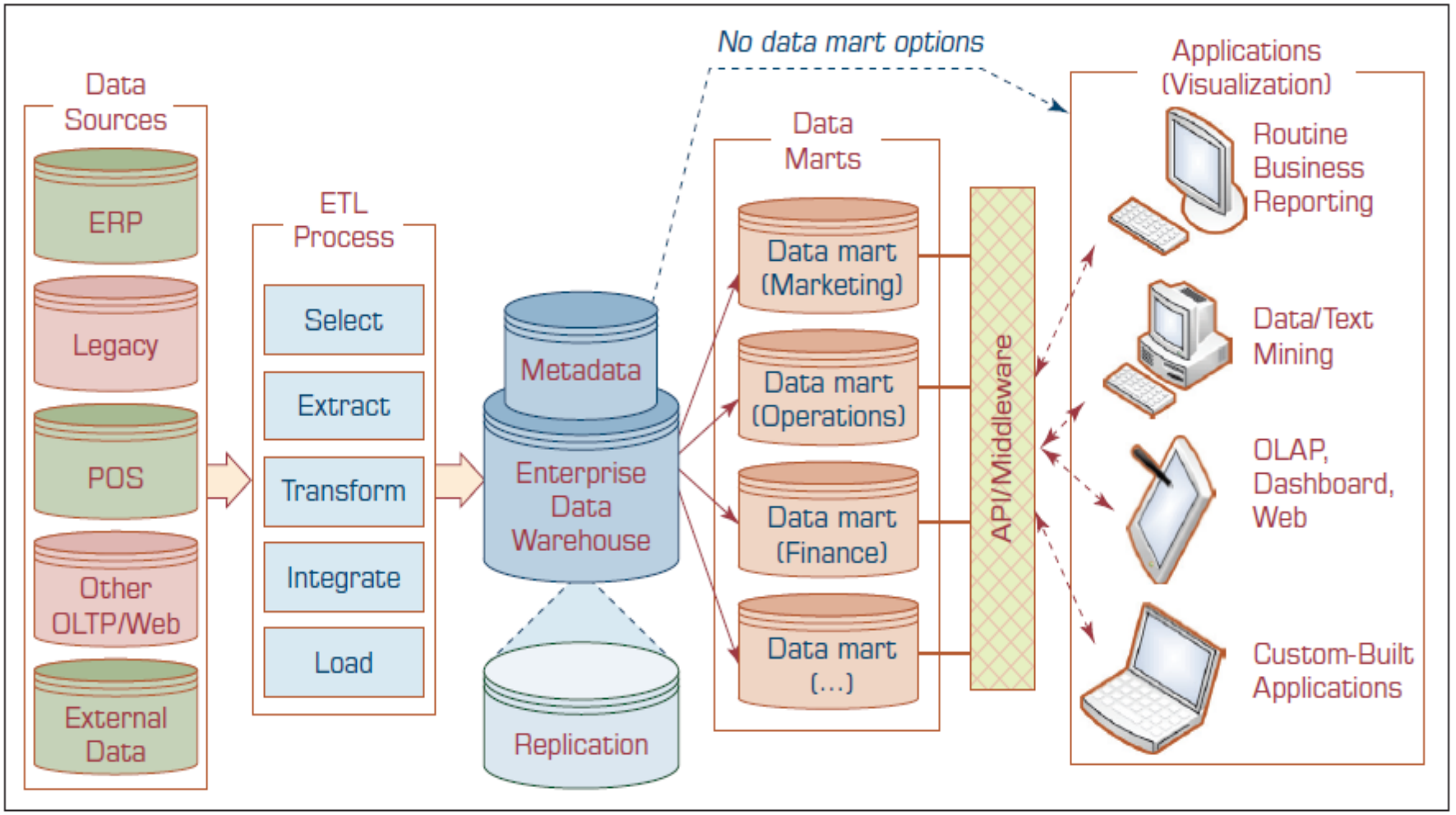
Transaction Processing vs Analytic Processing

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)

Evolution of BI



Data warehouse framework



Analytics overview

- **Analytics** is the process of developing actionable decisions or recommendations for actions based on insights generated from historical data.
- According to the Institute for Operations Research and Management Science (INFORMS), **analytics** represents the combination of computer technology, management science techniques, and statistics to solve real problems.

Analytics overview

- **Descriptive analytics** refers to knowing what is happening in the organization and understanding some underlying trends and causes of such occurrences.
- It involves the consolidation of data sources and availability of all relevant data in a form that enables appropriate reporting and analysis. Usually, the development of this data infrastructure is part of **DWs**.
- From this data infrastructure ,we can develop appropriate reports, queries, alerts, and trends using various reporting tools and techniques.

Analytics overview

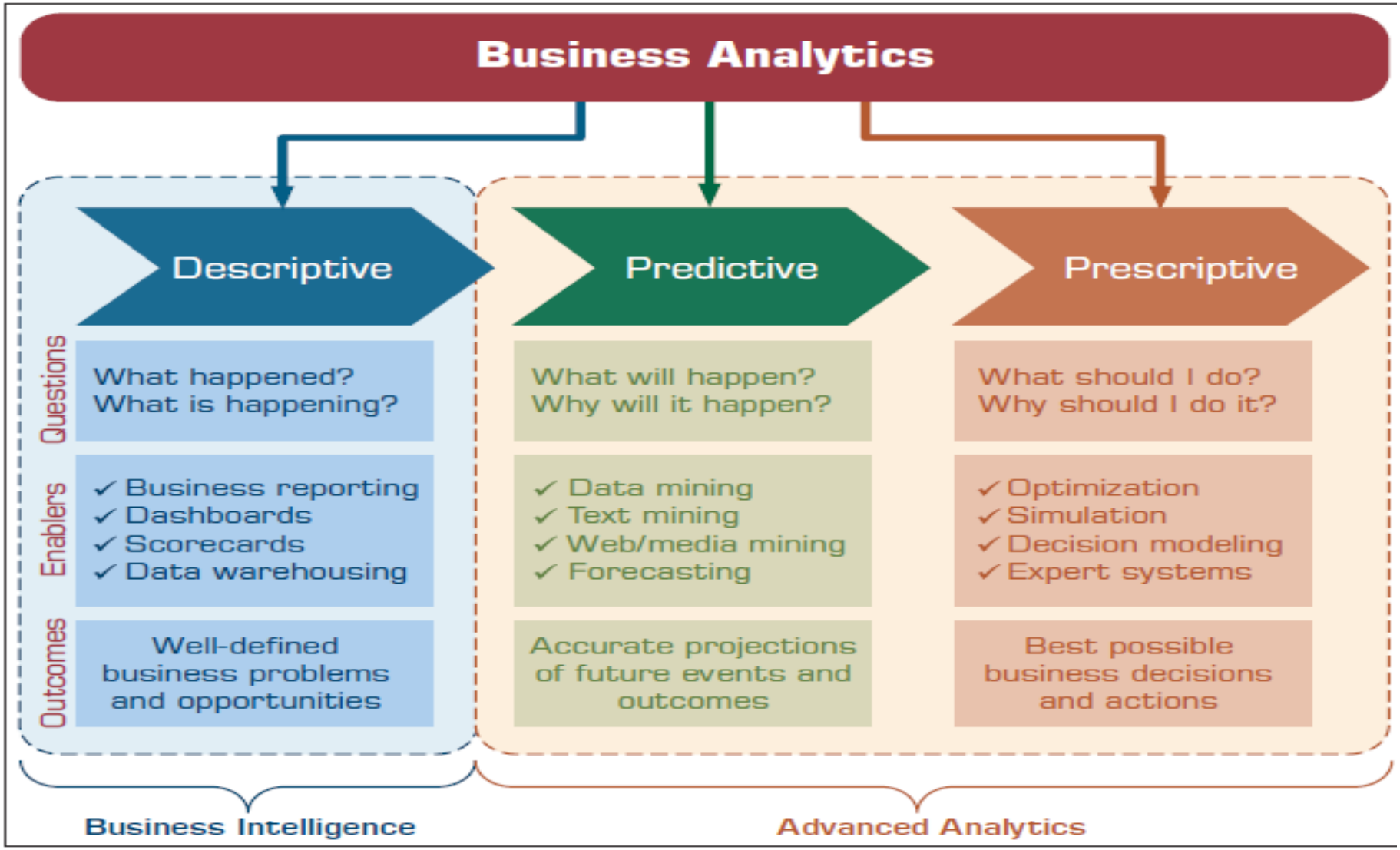
- **Predictive analytics** aims to determine what is likely to happen in the future. This analysis is based on statistical techniques as well as other more recently developed techniques that fall under the general category of data mining.
- The goal of these techniques is to be able to predict if the customer is likely to switch to a competitor (“churn”), what the customer would likely buy next and how much, what promotions a customer would respond to, whether this customer is a creditworthy risk, and so forth.
- A number of techniques are used in developing predictive analytical applications, including various classification algorithms

Analytics overview

Prescriptive Analytics aims to is to recognize what is going on as well as the likely forecast and make decisions to achieve the best performance possible. This group of techniques has historically been studied under the umbrella of **OR** or **management sciences** and are generally aimed at **optimizing** the performance of a system.

The goal here is to provide a decision or a recommendation for a specific action. These recommendations can be in the form of a specific yes/no decision for a problem, a specific amount (say, price for a specific item or airfare to charge), or a complete set of production plans. The decisions may be presented to a decision maker in a report or may be used directly in an automated decision rules system (e.g., in airline pricing systems). Thus, these types of analytics can also be termed decision or normative analytics.

Relationship between BA, BI , DW



Analytics overview

Data analyst vs Data scientist

- **Data Analyst** is just another term for professionals who were doing BI in the form of data compilation, cleaning, reporting, and perhaps some visualization. Their skill sets included Excel, some SQL knowledge, and reporting. You would recognize those capabilities as descriptive or reporting analytics.
- **Data scientist** is responsible for predictive analysis, statistical analysis, and more advanced analytical tools and algorithms. They may have a deeper knowledge of algorithms and may recognize them under various labels—data mining, knowledge discovery, or machine learning. Some of these professionals may also need deeper programming knowledge to be able to write code for data cleaning/analysis in current Web-oriented languages such as Java or Python and statistical languages such as R.

Data analyst vs Data scientist

	Data analyst	Data scientist
Mathematics	Foundational math, statistics	Advanced statistics, predictive analytics
Programming	Basic fluency in R, Python, SQL	Advanced object-oriented programming
Software and tools	SAS, Excel, business intelligence software	Hadoop, MySQL, TensorFlow, Spark
Other skills	Analytical thinking, data visualization	Machine learning, data modeling