BUSINESS INTELLIGENCE SYSTEM

BUSINESS INTELLIGENCE (BI) SYSTEMS

- are information systems that process operational, social, and other data to identify patterns, relationships, and trends for use by business professionals and other knowledge workers.

BUSINESS INTELLIGENCE

- are patterns, relationships, trends, and predictions.

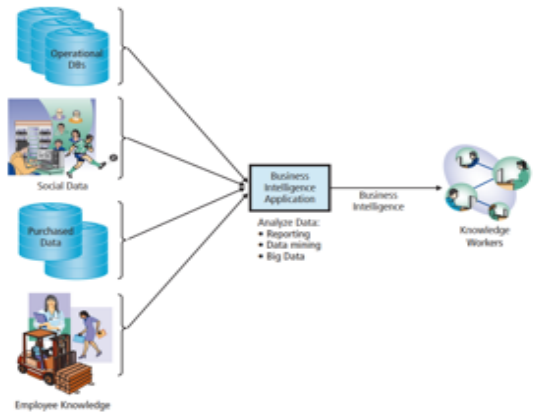
BI SYSTEMS FIVE STANDARD COMPONENT

- hardware, software, data, procedures, and people

BI APPLICATION

- is the software component of a BI system.

COMPONENTS OF BI SYSTEM



TYPES OF BUSINESS INTELLIGENCE SYSTEM

1. Informing

- Which products are selling quickly?

- Which products are most profitable?

2. Deciding

- Which customers shop at each location?

- Create custom marketing plans per store.

3. Problem solving

- How can we increase sales?

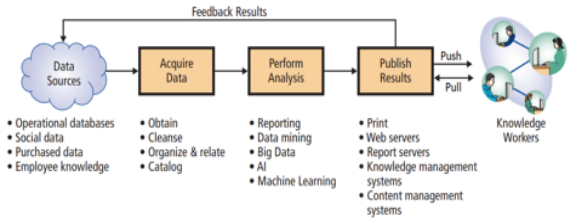
- How can we reduce food waste?

4. Project Management

- Build in-store cafés

- Expand to other locations

THREE PRIMARY ACTIVITIES IN THE BI PROCESS



THREE PRIMARY ACTIVITIES IN THE BI PROCESS CONT.

1. Data acquisition

- is the process of obtaining, cleaning, organizing, relating, and cataloging source data.

2. BI analysis

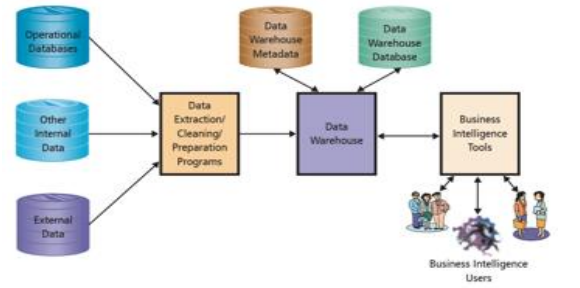
- is the process of creating business intelligence. (ex. Reporting, Data Mining, and Big Data)

3. Publish results

- is the process of delivering business intelligence to the knowledge workers who need it.

> **Push publishing** delivers business intelligence to users without any request from the users.
> **Pull publishing** requires the user to request BI results.

COMPONENTS OF A DATA WAREHOUSE



How Do Organizations Use Data Warehouses and Data Marts to Acquire Data?

Data Marts

- have a single-subject focus that filter and summarize information from data warehouse

Data warehouse

- which is a facility for managing an organization's BI data.

- **Functions**:

  Obtain data

  Cleanse data

  Organize and **relate** data

  Catalog data

PROBLEMS WITH OPERATIONAL DATA

- Dirty data (irrelevant data)

- Missing values

- Inconsistent data

- Data not integrated

- Wrong granularity (refers to the level of detail represented by the data)

  Too fine (the data can be made coarser by summing and combining)

  Not fine enough (there is no way to separate the data into constituent parts)

  Too much data

  Too many attributes, Too many data points

DATA WAREHOUSES VS. DATA MARTS

Data warehouse

- takes data from the data manufacturers (operational systems and other sources), cleans and processes the data, and locates the data on the shelves, so to speak, of the data warehouse.

Data mart

- is a data collection, smaller than the data warehouse, that addresses the needs of a particular department or functional area of the business.

THREE TECHNIQUES FOR PROCESSING BI DATA

BI Analysis Type	Goal	Characteristics
Reporting	Create information about past performance.	Process structured data by sorting, grouping, summing, filtering, and formatting.
Data mining	Classify and predict.	Use sophisticated statistical techniques to find patterns and relationships.
Big Data	Find patterns and relationships in Big Data.	Volume, velocity, and variety force use of MapReduce techniques. Some applications use reporting and data mining as well.

(1) REPORTING ANALYSIS

- is the process of sorting, grouping, summing, filtering, and formatting structured data.

- **Structured data** is data in the form of rows and columns. Most of the time structured data means tables in a relational database.

- **Reporting application** is a BI application that inputs data from one or more sources and applies reporting processes to that data to produce business intelligence.

- **Exception reports** are produced when something out of predefined bounds occurs.

RFM ANALYSIS

- a technique readily implemented with basic reporting operations, is used to analyze and rank customers according to their purchasing patterns.

RFM considers how or how much:

recently (R) a customer has ordered,

frequently (F) a customer ordered,

money (M) the customer has spent.

ONLINE ANALYTICAL PROCESSING (OLAP)

- provides the ability to sum, count, average, and perform other simple arithmetic operations on groups of data.

The **viewer** of the report can change the report's format—hence the term **ONLINE**.

An OLAP report has **measures** and **dimensions**.

A **measure** is the data item of interest. (Total sales, average sales, and average cost).

A **dimension** is a characteristic of a measure. (Purchase date, customer type, customer location, and sales region)

(2) DATA MINING ANALYSIS

DATA MINING

- is the application of statistical techniques to find patterns and relationships among data for classification and prediction.

- resulted from a convergence of disciplines, including **artificial intelligence** and **machine learning**.

DATA MINING TECHNIQUES CATEGORIES

1. Unsupervised data mining

- analysts do not create a model or hypothesis before running the analysis. (**Cluster analysis,** statistical techniques identify groups of entities that have similar characteristics).

2. Supervised data mining

- data miners develop a model prior to the analysis and apply statistical techniques to data to estimate parameters of the model. (**Regression analysis**, measures the effect of a set of variables on another variable)

(3) BIG DATA or BigData

- is a term used to describe data collections that are characterized by huge volume, rapid velocity, and great variety.

- **data sets** that are at least a **petabyte** in size, and usually larger.

A **data set** containing all Google searches in the United States on a given day is Big Data in size.

Additionally, Big Data has **high velocity**, meaning that it is generated rapidly.

Big Data may have structured data, but it also may have free-form text, dozens of different formats of Web server and database log files, streams of data about user responses to page content, and possibly graphics, audio, and video files.

BIG DATA TOOLS

MapReduce

- a technique for harnessing the power of thousands of computers working in parallel.

- the Big Data collection is broken into pieces, and hundreds or thousands of independent processors search these pieces for something of interest.

- that process is referred to as the **Map phase**.

HADOOP

- An open-source program supported by the **Apache Foundation10** that implements MapReduce

- drive the process of finding and counting the Google search terms, but Google uses its own proprietary version of MapReduce to do so instead.

- **Amazon** supports Hadoop as part of its EC3 cloud offering.

- **Microsoft** offers Hadoop on its Azure platform as a service named HDInsight.

Hadoop includes a query language titled **Pig**.

CHARACTERISTICS OF BI PUBLISHING ALTERNATIVES

Static reports

- are BI documents that are fixed at the time of creation and do not change.

Dynamic reports

- are BI documents that are updated at the time they are requested.

BI SERVER

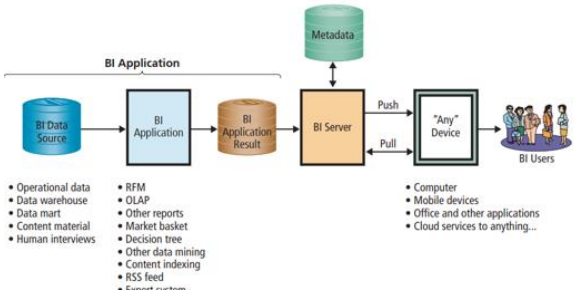
- is a Web server application that is purpose-built for the publishing of business intelligence.

- **use metadata to determine what results to send to which users and, possibly, on which schedule**.

- Functions: **management** and **delivery**.

The **management** function maintains metadata about the authorized allocation of BI results to users.

ELEMENTS OF A BI SYSTEM



ROLE OF KNOWLEDGE MANAGEMENT SYSTEMS

Knowledge management (KM)

- is the process of creating value from intellectual capital and sharing that knowledge with employees, managers, suppliers, customers, and others who need that capital.

- done before social media.

KM benefits organizations in two fundamental ways:

It improves **process quality**.

It increases **team strength**.

CONTENT MANAGEMENT SYSTEMS (CMS)

- are information systems that support the management and delivery of documents including reports, Web pages, and other expressions of employee knowledge.

Typical users:

- **companies** that sell complicated products that want to explain those to stakeholders

CHALLENGES IN CONTENT MANAGEMENT SYSTEM

- Content databases are huge
- CMS content is dynamic
- Documents do not exist in isolation from each other
- Contents are perishable
- Content is provided in many languages

WHY IS ARTIFICIAL INTELLIGENCE (AI) IMPORTANT?

Artificial intelligence (AI)

- is the ability of a machine to simulate human abilities such as vision, communication, recognition, learning, and decision making in order to achieve a goal.

- use AI to increase the **automation**, or the process of making systems operate without human intervention, of mundane tasks typically done by humans.

ADVANCES IN AI

Six main forces that have helped advance AI in recent years:

First, **computing power** has been increasing exponentially for several decades (Moore's Law), while earlier waves of AI lacked the necessary computing power.

Second, the **availability of large data sets** has advanced to the point that AI is viable technology.

Third, **cloud computing** has advanced AI development because it has made scalable resources available at very low costs.

Fourth, the rapid increase in **network connected smart devices** is producing vast amounts of data for AI applications.

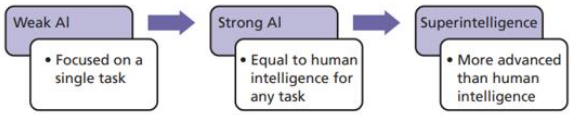
Fifth, **fundamental breakthroughs** in AI techniques have made AI useful for a variety of tasks. (**Neural Network/Deep learning**)

Sixth, recent advances in AI have been driven by the demand for applications that **solve practical problems**.

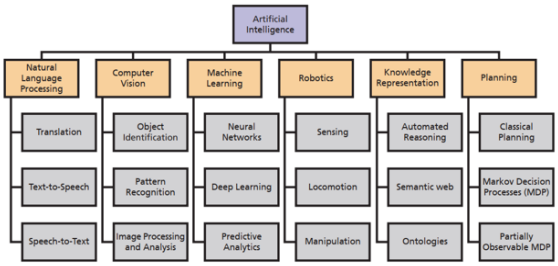
PRODUCTIVITY GAINS FROM AUTOMATION

Productivity Gains from Automation
Can work 24 hours, 365 days
Immediately trained, no "onboarding"
No breaks during work hours
No impaired workers
No time-wasting activities
No accidents or injuries
No arguments with other employees or managers
No scheduling issues
All holiday shifts covered
More accurate, precise, and consistent

EVOLUTION OF AI ABILITIES



MAJOR AI RESEARCH AREAS



MACHINE LEARNING

- A subset of AI or the extraction of knowledge from data based on algorithms created from training data.

- focused on **predicting outcomes based on previously known training data**.

- example, machine learning can be used to teach a system to recognize and classify dog breeds.

- EX. Microsoft **Fetch!** that does just that.

USING MACHINE LEARNING TO AUTOMATICALLY DETECT SPAM

Algorithm

- a set of procedures used to solve a mathematical problem, that best fits our situation.

- ex. **Naïve Bayes Classifier** that predicts the probability of a certain outcome based on prior occurrences of related events.

- predict whether a new email is spam or not based on attributes of previous spam messages.

IBMs Watson

- IBM's artificial intelligence named Watson is a question answering system that draws on several areas of AI.

- First, it uses **natural language processing (NLP)**, or the ability of a computer system to understand spoken human language, to answer questions.

- It was designed to play against world champions on the quiz show **Jeopardy**! and in 2011 it won.