Module 3: Big Data

- What is Big Data? How Big is Big Data?
- What are the advantages of Big Data?
- What are the challenges of Big Data?
- What are the widely used tools for Big Data?

Learning Objectives

- Finishing this module, you will be able to:
  - Explain what are big data using the eight V's
  - Explain big data applications
  - Understand and describe the challenges of big data
  - Explain the work flow of big data architecture
  - Identify the big data layers with example
  - Identify a few tools for big data

What Is Data?

- We learned what is data.
- We learned data will be useful if we can turn data into knowledge to help for decision making.
- We learned, in particular, structured data, that:
  - Organized in clearly defined tabular format
  - Queried with the Structured Query Language
  - Has been the most common practice in the field

Why Do We Need Data?

- Data is the future
  - Before 1600: Empirical Science
  - 1600-1950s: Theoretical science
  - 1950s-1990s: Computational science
  - 1990-now: Data science
    - 2012 Harvard Business Review: Data Scientists: The sexiest job of the 21st century.
    - 10 years after, it is still the case.

Big Data in Banking and Security

- Banking industry uses Big Data for risk analysis, fraud detection, and anti-money laundering, etc.
- Big Data can be used for security departments to catch illegal trading activities, such as using Network Analysis, Natural Language Processing, etc.

Big Data in Marketing

- Understanding customer behavior and preference is very important.
- Media content, online ads, content recommenders, are driven by the analyzing customer data at large scale.
- Youtube, Facebook, and TikTok!

Big Data in Healthcare

- Big Data is being used to deliver evidence-based diagnosis for reducing the cost and time in medical development.
- Big Data is also used in Machine Learning models to extract feature and patterns from x-ray images.

Big Data in Manufacturing

- Manufacturing in a Global setting is hard, and Big Data can provide quantitative support for decision making.
  - Supply chain management
  - Predictive maintenance
  - Quality control
  - Product forecasting
  - Risk management

Big Data in Retail

- Retail industry, especially, e-commerce industry, is probably the biggest application of big data.
- It combines many aspects that Big Data might support, from back-end supply chain management, to the front-end marketing strategies.