2 Introduction to Data Science

2.1 Data Science

- Data science is a field that uses statistics, scientific computing, scientific methods, processes, algorithms and systems to extract knowledge and insights from noisy, structured and unstructured data.

- It unifies domain knowledge with statistics and data analytics.

2.1.1 Challenges of Traditional Systems

1. Data set size is assumed to be limited and fixed. Huge data sets are slow to handle in traditional systems.

2. Designed to handle structured data only (containing numbers and categories). They cannot directly work on unstructured data like text, video and audio.

3. They are efficient in carrying out relatively easy-to-perform queries and analysis, but not otherwise.

4. They use conventional methods (like basic statistics) for analysis and need to support new methodologies (like machine learning).