在互联网、政治经济、社会、公共卫生、金融、医疗健康、环境、海洋等等各个领域，都存在着PB量级的数据有待研究，且大数据分析的结果已经产生实际效应。

本课程将从数据统计分析的数学基础、大数据智能分析与处理方法、大数据的分布式处理与并行计算以及大数据的多领域应用等不同层次，教授大数据分析与处理的基本方法、工具及应用。具体内容包括：

（1）大数据分析与处理基础：数据科学发展情况；大数据的时代背景及其重要性；大数据分析与处理的典型应用。

（2）数据统计分析的数学基础：数据统计知识回顾；多维数据分布；数据采样；主成份分析与因子分析；假设检验；线性相关与回归；非线性回归；方差分析；P-Value。

（3）大数据的智能处理：时间序列分析；动态贝叶斯网络；矩阵分解理论；机器学习；大规模数据上的模型应用。

（4）大数据的分布式处理与并行计算：HDFS与HBASE；Hadoop与Mapreduce，MapReduce并行计算；基于MapReduce的数据分析与智能处理方法的并行化。

（5）大数据分析与处理前沿：社交媒体大数据分析与处理；城市交通大数据分析与处理；金融大数据分析与处理；医疗大数据分析与处理等。