

大数据是指规模庞大到无法用传统工具在合理时间内处理的数据集，其核心特征可概括为“4V”：Volume（海量）、Velocity（高速）、Variety（多样）、Value（低价值密度）。

关键技术涵盖数据采集、存储、处理与分析四大环节。采集阶段通过传感器、日志爬虫等工具获取结构化（数据库表）、半结构化（XML）、非结构化（视频、文本）数据；存储依赖分布式文件系统（如 HDFS）和 NoSQL 数据库（如 MongoDB），解决海量数据的横向扩展问题；处理环节运用 MapReduce、Spark 等分布式计算框架，实现并行运算提升效率；分析则借助机器学习算法挖掘隐藏价值，例如电商平台通过用户行为数据预测消费偏好，推送个性化商品。

大数据技术已渗透至金融风控、智慧城市、精准医疗等领域，其核心价值在于从碎片化信息中提炼规律，辅助决策优化。