

# 模块三：大数据技术及行业创新应用案例

## 一、大数据系统平台总体架构及部署

### 1、大数据系统平台总体架构描述

- (1) 大数据的业务目标
- (2) 大数据总体架构模型
- (3) 大数据总体架构特点
- (4) 大数据系统的物理架构
- (5) 大数据系统的逻辑架构
- (6) 大数据系统与现有系统的关系
- (7) 传统应用如何成为大数据系统的应用

### 2、大数据平台工作流程

- (1) 大数据采集系统架构及工作流程
- (2) 大数据存储架构及工作流程
- (3) 大数据处理架构及工作流程
- (4) 大数据挖掘分析架构及工作流程
- (5) 大数据展现平台及工作流程

### 3、大数据平台技术体系构成

- (1) 大数据采集技术（多样性：多源数据融合，数据抽取、转换和加载的过程优化)
- (2) 大数据存储技术（海量：结构化、半结构化和非结构化数据)

- (3) 大数据处理技术（快速：并行计算、流计算）
- (4) 大数据挖掘分析技术（价值：数据挖掘、商业智能）
- (5) 大数据可视化展示技术（价值：数据可视化）
- (6) 大数据隐私安全

#### 4、大数据相关概念与相关技术概览

- (1) 数据结构：结构化数据与非结构化数据
- (2) 数据库数据模型：关系型数据库与非关系型数据库
- (3) 数据处理特性：OLTP 与 OLAP
- (4) 数据一致性：强一致性与最终一致性
- (5) 数据存储方式：行式存储与列式存储
- (6) 数据库存储与处理架构：SMP 与 MPP
- (7) 数据存储架构：传统分布式文件与新型分布式文件
- (8) 数据处理架构：基于并行计算的分布式数据处理技术（MapReduce）

#### 5、大数据存储和处理技术

- (1) 分布式存储和计算平台- Hadoop
- (2) 分布式文件系统-HDFS
- (3) 分布式计算框架-YARN/MapReduce
- (4) 分布式数据库-NoSQL

#### 6、大数据查询和分析技术：SQL on Hadoop

- (1) Hive：基本的 Hadoop 查询和分析

- (2) Hive 2.0 : Hive 的优化和升级
- (3) 实时互动的 SQL : Impala 和 drill
- (4) 基于 PostgreSQL 的 SQL on Hadoop
- (5) 大数据高级分析和可视化技术
- (6) 传统数据仓库与联机分析处理技术
- (7) 大数据挖掘与高级分析
- (8) 大数据挖掘与高级分析库 : Mahout
- (9) 非结构化复杂数据分析
- (10) 实时预测分析
- (11) 开源可视化工具 : R 语言
- (12) 可视化技术

## 7、大数据 Hadoop2.0 平台架构及组件

- (1) Hadoop2.0 平台总体架构解析
  - Hadoop2.0 平台层次划分及构成
  - Hadoop2.0 平台软件系统构成
  - Hadoop2.0 平台各组件介绍
- (2) Hadoop2.0 平台各组件功能描述
  - 分布式文件系统-HDFS
  - 分布式批处理计算-YARN/MapReduce
  - 分布式数据库-HBase

- 分布式数据仓库-Hive
- 分布式协作服务-ZooKeeper
- 数据分析挖掘-Pig
- 数据采集系统-Chukwa
- 高吞吐量分布式消息系统-Kafka
- 系统监控- Ambari
- 跨语言服务间通信-Thrift
- 日志收集与转运工具-Flume
- 数据结果与线上决策的交互-Redis
- 流计算基础平台-Storm vs Spark
- 流式计算框架 Spark Streaming

## 二、大数据应用实践方法及行业应用案例分析

### 1、大数据应用实践方法

- (1) 业务需求定义
- (2) 大数据应用现状分析与标杆比较
- (3) 大数据系统平台规划和设计
- (4) 大数据技术切入与实施
- (5) 大数据试用和评估
- (6) 大数据应用推广

### 2、业务需求定义

- (1) 分析业务战略，了解战略层面的大数据需求
- (2) 调研业务经营模式、管理现状、进行大数据需求分析，发现应用机会和场景
- (3) 发现大数据应用对企业的机会和挑战
- (4) 标杆应用研究

### 3、各行业大数据应用的个性需求分析方法

- (1) “互联网与电子商务行业”大数据应用需求分析
- (2) “电信运营”大数据应用需求分析
- (3) “政府”大数据应用需求分析
- (4) “金融业”大数据应用需求分析
- (5) “零售业”大数据应用需求分析
- (6) “教育业”大数据应用需求分析
- (7) “医疗业”大数据应用需求分析
- (8) “能源业”大数据应用需求分析
- (9) “制造业”大数据应用需求分析
- (10) “交通物流业”大数据应用需求分析
- (11) 企业级大数据应用的共性需求分析
- (12) 客户分析
- (13) 绩效分析
- (14) 欺诈和风险评估

### 4、大数据应用现状分析与标杆比较

- (1) 分析数据应用现状和能力
- (2) 分析应用场景和数据容量、种类和速度
- (3) 确定大数据评价指标体系

(4) 应用场景的数据源和样本分析

(5) 标杆数据分析研究

## 5、大数据系统平台规划和设计

(1) 大数据战略目标定义

(2) 大数据系统平台架构设计

提出大数据架构

细化获取架构

处理和存储架构

分析架构

信息安全架构

大数据组织架构

## 6、关键技术系统设计

## 7、技术选型和采购建议

## 8、技术实施规划

(1) 大数据技术切入与实施

大数据技术试验

试点和数据采集、存储和分析

平台部署

(2) 大数据试用和评估

大数据试用

□大数据评估

## 9、大数据应用推广

- (1) 需求分析
- (2) 应用推广准备
- (3) 启动新的项目

## 10、以电信运营商客户分析为例的大数据实施案例分析

- (1) 大数据客户分析业务需求
- (2) 大数据客户分析现状与标杆比较
- (3) 大数据客户分析应用架构规划与设计
- (4) 大数据客户分析实施、试点和推广