

## **1 概述**

### **1.1 大数据的概念和特点**

#### **1.1.1 大数据的基础**

#### **1.1.2 大数据如何“与时俱进”？**

#### **1.1.3 大数据发展趋势**

人工智能

物联网结合

各个行业的深入

### **1.2 大数据的技术基础**

#### **1.2.1 从数据仓库开始**

#### **1.2.2 HADOOP 生态圈**

#### **1.2.3 与云计算的关系**

#### **1.2.4 数据运维能力提升**

### **1.3 大数据的应用举例**

#### **1.3.1 大数据提升客户分析能力**

**1.3.2** 大数据提升产品分析能力

**1.3.3** 大数据提升管理水平

**1.3.4** 大数据提升各行业“智慧”

**1.4** 大数据下的人工智能（AI）

**1.4.1** 什么是人工智能

**1.4.2** 人工智能改变哪些行业？

**1.4.3** 大数据下的人工智能有何不同？

**1.4.4** 人工智能的“颠覆”

**1.5** 大数据如何精细化管理

**1.5.1** 量化管理的引出

**1.5.2** 大数据如何提升“量化”的维度和深度

**1.5.3** 从艺术到技术

**1.5.4** 自动驾驶到自动管理？

**1.6** 电信企业的大数据“商机”

**1.6.1** 从网络运营到数据运营

**1.6.2** 提炼“内功”

**1.6.3 提升外部管理能力**

**1.6.4 扩展增值产品运营市场**

## **2 大数据的行业解决方案应用案例**

**2.1 基础应用范围**

**2.2 石油行业应用案例**

**2.3 交通行业应用案例**

**2.4 旅游行业应用案例**

**2.5 金融行业应用案例**

**2.6 电信行业应用案例**

**2.7 互联网行业应用案例等**

## **3 机器学习基础**

**3.1 多维分析方法**

**3.1.1 OLAP 分析**

**3.1.2 上钻和下钻**

**3.1.3 用 OLAP 分析问题**

## **3.2 分析算法**

### **3.2.1 回归算法**

线性回归

逻辑回归

### **3.2.2 决策树算法**

**C4.5** 算

法 **CART**

算法

### **3.2.3 贝叶斯算法**

朴素贝叶斯算法

**BBN (Bayesian Belief Network)** 算法

### **3.2.4 基于核的算法**

支持向量机 **SVM** 算法

线性判别分析(**Linear Discriminate Analysis** , **LDA**)

### **3.2.5 聚类算法**

**K-MEANS** 算法

期望最大化算法(**Expectation Maximization , ME**)

### **3.2.6 关联规则算法**

**Apriori** 算法

### **3.2.7 降低维度算法**

主成份分析(**Principle Component Analysis , PCA**)算法

偏最小二乘回归(**Partial Least Square Regression , PLS**)  
算法

### **3.2.8 集成算法**

随机森林算法

梯度推进机

## **3.3 机器学习方法在电信行业应用举例**

# **4 深度学习基础**

## **4.1 神经网络算法的演进历史**

### **4.1.1 传统的神经网络算法**

### **4.1.2 神经网络算法的特点**

### **4.1.3 神经网络算法的瓶颈点**

## **4.2 神经网络的算法原理**

### **4.2.1 神经元的原理**

### **4.2.2 神经网络的原理**

### **4.2.3 神经网络的 **BP** 算法**

## **4.3 多层神经网络算法**

### **4.3.1 单层神经网络**

### **4.3.2 两层神经网络**

### **4.3.3 多层神经网络——深度学习**

## **4.4 深度学习算法**

### **4.4.1 卷积神经网络 **CNN****

### **4.4.2 循环神经网络 **RNN****

### **4.4.3 深度神经网络 **DNN****

## **4.5 深度学习的应用案例**

### **4.5.1 深度学习用于图像识别**

### **4.5.2 深度学习用于语音识别**

### **4.5.3 深度学习用于语言翻译**

## **5 大数据技术基础**

### **5.1 从数据仓库开始**

#### **5.1.1 数据仓库的“集中 ”**

#### **5.1.2 数据仓库的模型标准化**

#### **5.1.3 大数据的演进**

### **5.2 HADOOP 生态圈**

#### **5.2.1 开源社区概述**

#### **5.2.2 开源改变了什么？**

#### **5.2.3 HADOOP 生态圈内容**

#### **5.2.4 HADOOP 的技术原则**

#### **5.2.5 HADOOP 的运维**

### **5.3 HADOOP 基础**

#### **5.3.1 HDFS 的原理**

#### **5.3.2 MAP/REDUCE 原理**

#### **5.3.3 YARN 原理**

## **5.4 HIVE/HBASE 技术**

### **5.4.1 HIVE 的原理**

### **5.4.2 HBASE 的原理**

### **5.4.3 两者的关系**

## **5.5 SPARK 技术**

### **5.5.1 基本原理**

### **5.5.2 应用案例**

## **5.6 KAFKA/FLUME**

### **5.6.1 基本原理**

### **5.6.2 应用案例**

## **5.7 各个技术的特点对比**

# **6 HADOOP 技术介绍**

## **6.1 发展历史**

### **6.1.1 google 的影响**

### **6.1.2 命名来源**

## **6.2 HDFS 原理**

### **6.2.1 适合做什么？**

### **6.2.2 不适合做什么？**

### **6.2.3 namenode 和 datanode**

## **6.3 HA 方法**

### **6.3.1 基本原理**

### **6.3.2 HADOOP 2.0 的 HA 实现方法**

## **6.4 MAP/REDUCE 原理**

## **6.5 YARN 原理**

### **6.5.1 2.0 引出的原因**

### **6.5.2 与容器 (docker) 的关系**

## **6.6 HIVE 基础**

### **6.6.1 基本原理**

### **6.6.2 应用案例**

## **6.7 HBASE 基础**

### **6.7.1 基本原理**

## **6.7.2 与 HIVE 的关系**

## **6.7.3 应用案例**

## **6.8 HADOOP 的难点**

### **6.8.1 安全性**

### **6.8.2 可操作性**

### **6.8.3 运维难题**

## **6.9 MR 的具体编程案例**

### **6.9.1 Wordcount 原理**

### **6.9.2 API 介绍及编程**

### **6.9.3 案例及高级编程**

# **7 SPARK**

## **7.1 背景**

### **7.1.1 对 HADOOP 的改良**

### **7.1.2 SPARK 的生态环境**

## **7.2 SPARK 的原理**

### **7.2.1 RDD 概念**

### **7.2.2 DAG 概念**

### **7.2.3 运行机制**

### **7.2.4 相关运算**

## **7.3 SPARK 的实施建议**

### **7.3.1 集群部署模式**

### **7.3.2 版本选择**

### **7.3.3 硬件配置建议**

### **7.3.4 混搭建议**

### **7.3.5 参数配置建议**

### **7.3.6 SPARK 做金融业的应用场景**

## **7.4 某公司 SPARK (500 节点) 部署实际案例**

### **7.4.1 建设历程**

### **7.4.2 技术要点**

计算引擎的高性能

如何实现多种异构环境透明访问

### **7.4.3 技术引入策略建议**

### **7.4.4 运维建设经验**

## **8 大数据下的企业运营**

### **8.1 从泰勒经济学说起**

#### **8.1.1 简述泰勒经济学**

#### **8.1.2 数据与企业管理的天然联系**

#### **8.1.3 让大数据渗透到企业的每个毛孔**

### **8.2 客户分析**

#### **8.2.1 客户的数据维度**

#### **8.2.2 客户视图分析**

#### **8.2.3 客户离网分析等**

### **8.3 产品分析**

#### **8.3.1 产品数据维度**

#### **8.3.2 产品视图分析**

#### **8.3.3 产品的潜在客户分析**

### **8.3.4 如何优化产品？**

## **8.4 精准营销**

### **8.4.1 营销的本质——配对**

### **8.4.2 精准营销——数据的魅力**

### **8.4.3 营销时机、渠道等选择**

## **8.5 员工的管理**

### **8.5.1 员工的量化分析**

### **8.5.2 大数据下的“办公室分析”**

### **8.5.3 员工绩效考评**

### **8.5.4 如何避免“苍蝇犯大案”**

## **8.6 外部生态圈的管理**

### **8.6.1 生态圈的概念**

### **8.6.2 电信企业的生态圈**

### **8.6.3 上下游企业的管理**

### **8.6.4 如何识别渠道欺诈？**

## **8.7 【例】某大数据系统应用案例**

## **9 \*大数据的互联网思维**

### **9.1 互联网思维概述**

#### **9.1.1 什么是互联网思维？**

#### **9.1.2 互联网思维在互联网企业中的落地**

#### **9.1.3 互联网思维的案例**

### **9.2 大数据的客户体验**

#### **9.2.1 什么是客户体验？**

#### **9.2.2 大数据的客户有哪些？**

#### **9.2.3 如何提升客户大数据使用体验？**

#### **9.2.4 【例】大数据的客户提升案例**

### **9.3 大数据的产品设计**

#### **9.3.1 大数据产品应该长什么样？**

#### **9.3.2 产品如何简约？**

#### **9.3.3 提供“千人千面”的产品**

#### **9.3.4 大数据的“爆品”在哪里？**

### **9.3.5 【例】大数据产品示例**

## **9.4 大数据的极致思维**

### **9.4.1 如何极致？**

### **9.4.2 数据的极致**

### **9.4.3 报告的极致**

### **9.4.4 讲故事的极致**

### **9.4.5 服务的极致**

## **9.5 大数据的平台思维**

### **9.5.1 什么是平台思维？**

### **9.5.2 大数据的平台化如何实现？**

### **9.5.3 从 APP STORE 到 BI STORE**

## **9.6 其它**

### **9.6.1 互联网思维如何用于企业运营？**

### **9.6.2 互联网思维堪比“文艺复兴”？**

## **9.7 【例】互联网思维在中国移动大数据系统中的落地实践**

# **10 大数据的收集及整理**

## **10.1 大数据的收集**

### **10.1.1 内部数据**

哪些 IT 系统？

财务系统

招聘系统等

### **10.1.2 外部数据**

爬虫获取数据

交换数据

购买数据等

## **10.2 大数据的整理**

### **10.2.1 如何表述数据间的关系？**

### **10.2.2 数据建模的概念**

### **10.2.3 数据建模的方法**

## **10.3 大数据的数据治理**

**10.3.1** 数据质量问题

**10.3.2** 数据质量如何管

控 **10.3.3** 数据字典的

定义

**10.4** 大数据的数据资

产 **10.4.1** 什么是数据

资产？

**10.4.2** 数据资产如何梳理？

**10.5** 【案例】某数据收集整理实际案例

## **11** 系统架构建设实施

**11.1** 项目的管理方法

**11.2** 大数据项目管理的特点

**11.3** 不同系统间数据接口控制

**11.4** 数据导入及管控

**11.5** 大数据系统建设过程中关键问题

**11.6** 大数据的数据质量保障

**11.7** 大数据的应用推广重点

**11.8** 大数据项目协调要点

**11.9** 【例】某大数据系统建设过程解剖

## **12** 电信行业大数据案例

**12.1** 某电信企业大数据建设案例分享

**12.2** 某电信企业大数据应用案例分享

**12.3** 某电信企业大数据管理案例分享

**12.4** 某电信企业大数据变现案例分享

**12.5** 其它行业大数据案例分享

## **13** 总结