

《大数据平台架构》——段方 北京大学博士后

1 背景

1.1 概述

1.1.1 大数据的概念和特点

1.1.2 大数据在不同领域中的应用

1.2 技术发展历史

1.2.1 数据统计系统

1.2.2 数据仓库系统

1.2.3 大数据系统

1.2.4 发展历程中的关键技术路线图

1.3

1.4 应用发展历史

1.4.1 报表和 KPI

1.4.2 数据挖掘/人工智能

1.4.3 对外服务

1.4.4 大数据运营

1.5 为什么是大数据

1.5.1 数据洞察一切

1.5.2 管理比拼的是技术能力

1.5.3 人工智能建立在大数据基础上

1.5.4 亟待新的管理理论突破

工业文明的管理理念

信息文明的管理理念呢？

1.6 【案例】大数据在一些行业的应用案例

2 系统架构实例

2.1 系统架构演进历史

2.1.1 数据库阶段

2.1.2 数据仓库阶段

数据标准

系统标准

技术选型

数据管理设计

2.1.3 大数据阶段

HADOOP 等技术引

入 中间的混搭架构

阶段 目标的大数据

架构

2.2 两级架构缘起

2.2.1 从两级法人说起

2.2.2 灵活性与标准性

2.2.3 不同阶段的不同理念

2.3 大数据的集中化

2.3.1 为什么要集中化？

2.3.2 大数据的审计职能

2.3.3 大数据的应用价值点

2.3.4 集中化的“痛”与“必然”

2.4 单点架构概述

2.4.1 整体架构

2.4.2 数据仓库平台

2.4.3 大数据平台

2.4.4 数据管理架构

2.4.5 数据应用架构

2.5 大数据架构详述

2.5.1 HADOOP 等技术族

2.5.2 HBASE/HIVE 设计

2.5.3 SPARK 设计

2.5.4 数据质量控制

2.5.5 数据源选择

2.6 架构设计的关键点

2.6.1 技术成熟度

2.6.2 架构设计的原则

2.6.3 架构设计的风险点

2.7 平衡的艺术

2.7.1 混搭的原因分析

2.7.2 硬件成本和人力成本

2.7.3 与业务发展的步调平衡

2.7.4 数据分析师的知识平衡

2.8 【案例】某企业大数据平台架构介绍

3 数据处理实例

3.1 概述

3.1.1 数据标准

3.1.2 业务标准

3.1.3 管理标准

3.2 数据内容

3.2.1 客户数据

3.2.2 产品数据

3.2.3 营销数据

3.2.4 员工数据

3.2.5 服务数据

3.3 数据标准的颗粒度

3.3.1 详细数据

3.3.2 汇总数据

3.3.3 衍生数据

3.4 数据如何收集

3.4.1 从内部各个 IT 系统获取

3.4.2 外部数据获取方法

3.4.3 批量获取/实时获取

3.5 数据清洗过程

3.5.1 数据抽取

3.5.2 数据转换

3.5.3 数据加载

3.6 跨平台数据接口标准

3.6.1 数据接口的定义

3.6.2 数据接口的管理职责

3.6.3 接口设计的原则

3.7 逻辑模型设计及实例

3.7.1 原因

3.7.2 逻辑模型设计的原则

3.7.3 设计实例

3.8 物理模型设计及实例

3.8.1 原因

3.8.2 物理模型设计原则

3.8.3 设计实例

3.9 数据字典及实例

3.9.1 概念

3.9.2 范围和内容

3.9.3 设计实例

3.10 【案例】某企业的数据资产治理案例

4 数据分析方法

4.1 综述

- 4.1.1 确定业务问题**
- 4.1.2 选择数据**
- 4.1.3 选择分析方法**
- 4.1.4 评估分析方法**
- 4.1.5 “最后一公里” 问题**
- 4.1.6 数据的可视化**
- 4.2 统计分析**
 - 4.2.1 统计分析的概念**
 - 4.2.2 统计分析的特点**
 - 4.2.3 统计分析的应用**
- 4.3 OLAP 分析**
 - 4.3.1 cube 的概念**
 - 4.3.2 上钻和下钻**
 - 4.3.3 OLAP 如何分析问题**
 - 4.3.4 OLAP 工具举例**
 - 4.3.5 【案例】 OLAP 分析实例**
- 4.4 数据挖掘分析**
 - 4.4.1 数据挖掘概念**
 - 4.4.2 数据挖掘算法介绍**

4.4.3 数据挖掘工具举例

4.4.4 【案例】数据挖掘实际案例

4.5 人工智能分析

4.5.1 人工智能的发展历史

4.5.2 人工智能的突破点

4.5.3 人工智能的算法

4.5.4 人工智能在金融行业应用案例

4.5.5 【案例】人工智能算法的应用案例（语音识别+图像识别）

4.6 【案例】人工智能在电信领域中的应用

5 大数据应用内容

5.1 从泰勒经济学说起

5.1.1 简述泰勒经济学

5.1.2 数据与企业管理的天然联系

5.1.3 让大数据渗透到企业的每个毛孔

5.2 客户分析

5.2.1 客户的数据维度

5.2.2 客户视图分析

5.2.3 客户离网分析等

5.2.4 客户细分分析

5.3 产品分析

5.3.1 产品数据维度

5.3.2 产品视图分析

5.3.3 产品的潜在客户分析

5.3.4 产品交叉分析

5.4 精准营销

5.4.1 营销的本质——配对

5.4.2 精准营销——数据的魅力

5.4.3 营销时机、渠道等选择

5.5 员工的管理

5.5.1 员工的量化分析

5.5.2 大数据下的“办公室分析”

5.5.3 员工绩效考评

5.5.4 如何避免“苍蝇犯大案”

5.6 外部生态圈的管理

5.6.1 生态圈的概念

5.6.2 金融企业的生态圈

5.6.3 上下游企业的管理

5.6.4 如何识别渠道欺诈？

6 某个大数据应用实例

6.1 【案例】外来工群体细分及营销

6.1.1 问题定义

哪个 **KPI**?

具体的口径

6.1.2 可选方法分析

数据分析

算法分析

6.1.3 潜在客户特征分析

客户样本选择

客户特征描述

6.1.4 客户细分实例

挖掘关键变量筛选

决策树分析实例

客户分群的营销建议

6.1.5 营销方案设计

营销渠道选择

营销时机选择

营销实例

6.1.6 营销过程

营销案实例

多波次设计

6.1.7 效果评估

LIFT 等

折算成经济效果

折算的社会效果

6.1.8 项目总结

7 大数据相关思考

7.1 大数据的建设

7.1.1 有哪些数据

7.1.2 收集哪些数据

7.1.3 采用什么技术

7.1.4 系统架构的设计

7.1.5 工程项目如何管理

7.2 大数据的管理

7.2.1 数据质量如何评估

7.2.2 数据质量如何控制

7.2.3 管理的作用甚于技术

7.2.4 数据资产的价值

7.3 大数据的应用

7.3.1 从业务需求出发

7.3.2 技术驱动和业务驱动结合

7.3.3 应用的突破点

7.3.4 应用的总体设计和规划

7.4 应用价值的评估

7.4.1 应用的评估方法

7.4.2 应用的时效性

7.5 还有哪些“痛”

7.5.1 如何容忍“互撕期”

7.5.2 数据质量的“苦”

7.5.3 技术选型的“惑”

7.5.4 项目管理的“难”

7.5.5 人才培养的“狂”

8 总结