

# 《大数据基础及大数据安全》

——段方 北京大学博士后

## 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** 大数据业界趋势

**1.6.1** 数据更加丰富（跨行业）

**1.6.2** 计算能力更强（边缘计算）

### **1.6.3 结合业务应用场景（行业解决方案）**

### **1.6.4 开源改变了商务模式（生态圈）**

## **1.7 大数据商业战略**

### **1.7.1 数据变现能力**

### **1.7.2 分析变现能力**

### **1.7.3 服务变现能力**

### **1.7.4 产业链的服务模式**

## **1.8 智慧银行的概念**

### **1.8.1 广义概念**

企业角度的智慧

服务的智慧

产品的智慧

管理的“颠覆

”

### **1.8.2 狭义概念**

智慧营业厅

智慧化设备（排队机、柜台等）

## **1.9 大数据对传统银行的意义**

### **1.9.1 对客户的深度理解**

### **1.9.2 服务客户的每个金融需求**

### **1.9.3 重塑辉煌的唯一机遇**

### **1.9.4 逆转“中介”的宿命**

### **1.10 【思考】大数据为什么能够成为中国的国家战略？**

## **2 大数据技术基础（可选）**

### **2.1 从数据仓库开始**

#### **2.1.1 数据仓库的“集中”**

#### **2.1.2 数据仓库的模型标准化**

#### **2.1.3 大数据的演进**

### **2.2 HADOOP 生态圈**

#### **2.2.1 开源社区概述**

#### **2.2.2 开源改变了什么？**

#### **2.2.3 HADOOP 生态圈内容**

## **2.2.4 HADOOP 的技术原则**

## **2.2.5 HADOOP 的运维**

## **2.3 HADOOP 基础**

### **2.3.1 HDFS 的原理**

### **2.3.2 MAP/REDUCE 原理**

### **2.3.3 YARN 原理**

## **2.4 HIVE/HBASE 技术**

### **2.4.1 HIVE 的原理**

### **2.4.2 HBASE 的原理**

### **2.4.3 两者的关系**

## **2.5 SPARK 技术**

### **2.5.1 基本原理**

### **2.5.2 应用案例**

## **2.6 KAFKA/FLUME**

### **2.6.1 基本原理**

### **2.6.2 应用案例**

## **2.7 大数据如何助力智慧银行**

### **2.7.1 是所有新技术的基础**

**2.7.2 数据维度前所未有丰富**（物联网、**VR/AR**、区块链等收集的数据）

**2.7.3 分析能力前所未有**（人工智能等算法突破）

## **2.8 人工智能在智慧银行中的应用**

**2.8.1 客户的生物识别**

**2.8.2 算法精度改良**

**2.8.3 更强的风险监控**

## **2.9 【思考】各个技术的特点对比**

# **3 大数据安全——定义及范围**

**3.1 数据安全定义**

**3.2 与传统安全技术的关系**

**3.3 相关数据安全的法律规定**

**3.4 敏感数据的定义和范围**

**3.5 数据安全管控技术**

### **3.6 加密与解密的方法综述**

### **3.7 数据使用权限管理**

### **3.8 【案例】基于大数据进行群体欺诈分析**

## **4 大数据安全管理架构**

### **4.1 层层防御的技术理念**

#### **4.1.1 从敏感数据角度的层层防御**

#### **4.1.2 从数据出来环节的层层防御**

#### **4.1.3 从数据生命周期的层层防御**

### **4.2 数据访问策略**

#### **4.2.1 隐私数据策略管理**

#### **4.2.2 告警策略管理**

### **4.3 “三分技术、七分管理”**

#### **4.3.1 安全管理的重要性**

#### **4.3.2 安全管理的内容和范围**

#### **4.3.3 技术围绕管理需求**

## **4.4 与网络安全的关系**

### **4.4.1 层次不同**

### **4.4.2 解决网络突破后的数据安全问题**

## **4.5 数据导出管理与监控**

### **4.5.1 数据水印技术**

### **4.5.2 访问轨迹追踪**

## **4.6 安全管理的关键点**

### **4.6.1 明确的规章制度**

### **4.6.2 严格地执行规章制度**

## **4.7 访问终端的安全管控**

### **4.7.1 终端的安全管控原则**

### **4.7.2 应用的管控流程**

## **4.8 机房的安全管控要求**

### **4.8.1 设备隔离要求**

### **4.8.2 访问监控要求等**

## **4.9 【案例】某企业大数据安全架构示例**

## **5 数据加密方法（可选）**

### **5.1 加密算法概述**

#### **5.1.1 算法的分类**

#### **5.1.2 算法的特点评估**

### **5.2 加密算法的选择和评估**

#### **5.2.1 选择的原则**

#### **5.2.2 如何评估和对比**

### **5.3 加密算法的原理**

### **5.4 加密算法的对比**

#### **5.4.1 性能方面**

#### **5.4.2 可逆或不可逆**

#### **5.4.3 与数据特点的结合**

### **5.5 数据脱敏算法效率考量**

#### **5.5.1 数据量的评估**

#### **5.5.2 加密、解密效率评估**

### **5.5.3 安全等级选择**

## **6 数据金库等管理应用**

### **6.1 4A 认证模式**

#### **6.1.1 4A 概念**

#### **6.1.2 与大数据权限管理的关系**

### **6.2 金库管理模式**

#### **6.2.1 何为金库模式？**

#### **6.2.2 金库模式的要点**

#### **6.2.3 金库模式的意义**

### **6.3 管理规章制度设计**

#### **6.3.1 设计哪些规章制度**

#### **6.3.2 规章制度的设计原则**

#### **6.3.3 规章制度的设计分类**

### **6.4 访问日志管理**

#### **6.4.1 大数据的日志如何管理**

### **6.4.2 日志的价值分析**

### **6.4.3 日志的分析应用**

## **6.5 安全实时告警**

### **6.5.1 如何进行实时监控**

### **6.5.2 监控规则的实时计算**

### **6.5.3 告警信息的实时传送**

## **6.6 【案例】某企业数据金库案例**

# **7 安全的大数据分析（可选）**

## **7.1 安全态势感知概念和范围**

## **7.2 态势感知的大数据收集**

## **7.3 态势感知的检测**

## **7.4 态势感知的分析、响应**

## **7.5 态势感知的预测**

## **7.6 态势感知的防御**

## **7.7 还有哪些应用？**

## **7.8 【案例】安全态势感知技术介绍**

# **8 HADOOP 系统安全防护**

## **8.1 开源系统的安全防护问题**

## **8.2 与传统数据仓库安全的对比**

## **8.3 HADOOP 漏洞分析**

## **8.4 HADOOP 生态补丁策略**

## **8.5 加固 HADOOP 系统**

# **9 大数据安全事故案例分析**

## **9.1 内部人员的安全事故案例**

## **9.2 外部人员的安全事故案例**

## **9.3 黑客攻击的风险**

# **10 一些思考**

## **10.1 客户隐私与客户服务的矛盾**

## **10.2 客户隐私的规避思考**

## **10.3 客户隐私的防护成本**

## 11 总结