

L1

- 1.什么是大数据
- 2.大数据的定义、基本属性、技术特征、算法特征

L2 大数据计算体系

- 1.系统总体架构
- 2.数据存储、处理、应用系统
- 3.统一访问接口
- 4.计算模式与平台
- 5.大数据计算架构与计算模型

L3 数据采集方法

- 1.系统日志数据采集 工具及过程
- 2.网络数据采集过程
- 3.网络爬虫 工作原理等
- 4.网页搜索策略
- 5.网页分析算法
- 6.数据采集接口 rest API

L4 数据预处理技术

- 1.数据预处理--数据噪声-- 分箱
- 2.数据噪声处理--平滑处理
- 3.数据噪声处理--聚类
- 4.数据变换--规范化
- 5.数据规约

L5 数据分析算法-数据关系

- 1.词袋模型
- 2.TF-IDF算法
- 3.余弦相似性**
- 4.Apriori算法**
- 5.pageRank算法

L6 数据分析算法-分类与聚类

1.朴素贝叶斯分类器

2.AdaBoost分类器

3.支持向量机SVM

4.k邻近算法KNN

5.k-Means聚类算法

6.最大期望算法

7.EM算法

L7 数据分析算法-数据决策

1.决策树

2.ID3算法

3.pseudo-code

4.c4.5

5.CART

5.1 GINI

6.ID3、C4.5、CART比较

L8文本读写技术

1.读取文本文件

2.写入文本文件

3.Apriori python实现

4.数据库连接

L9数据处理技术

1.合并数据集

2.数据转换

L10数据分析技术

1.Numpy

2.pandas

L11数据可视化

1.Matplotlib绘图

2.Pandas绘图工具

3.大屏可视化设计

L12Hadoop生态系统

1.Hadoop总体架构

集群配置、生态体系、各种系统

1.2HDFS文件系统

数据容错恢复、机架感知副本存放、文件错误检测和恢复、EC纠删码

L13分布式存储架构

0.三种分布式存储架构

1.Hbase存储模型与架构

2.二次索引表机制

3.分布式协同管理

4.Hadoop资源管理与作业调度

L14 MapReduce计算模型

1.分布式并行计算

2.计算架构

3.计算流程

4.实际算例

L15图并行计算框架

1.图计算问题

2.BSP图计算模型

3.图计算架构

L16流计算模型

1.流计算概念

2.流计算模式

3.流计算模型

L17 Storm计算架构

1.逻辑架构

2.系统架构

3.Storm实现机制

L18 内存计算模型

- 1.内存计算概念
- 2.分布式缓存系统
- 3.内存计算技术

L19 Spark内存计算

- 1.逻辑计算模型
- 2.物理计算架构
- 3.RDD数据模型
- 4.作业模型

L20 ElasticSearch 分布式搜索引擎

- 1.ES特点
- 2.ES计算模型
- 3.ES计算架构

L21 云计算基础

- 1.云计算概念
- 2.云计算架构
- 3.实现机制

L22 Google云计算平台

- 1 Google文件系统GFS
- 2 分布式数据处理MapReduce
- 3 分布式锁服务Chubby
- 4 分布式结构化数据表Bigtable

L23 Amazon 云

- 1 基础存储架构Dynamo
- 2 弹性计算云EC2
- 3 简单存储服务S3
- 4 非关系型数据库服务SimpleDB和DynamoDB
- 5 关系数据库服务RDS

L24 阿里云

1 阿里云计算体系架构

2 弹性计算服务 (ECS)

3 开放存储服务OSS

4 开放结构化数据服务OTS

5 开放数据处理服务ODPS

6 关系型数据库RDS

7.云创存储万物云服务