**大数据（HadoopNoSQL等）的技术与实践 课程大纲（课时共2天）**

| **主题** | **授课内容** |
| --- | --- |
| 大数据时代关系数据库的挑战与应对 | 1. 现代数据管理技术综述  2. 关系数据库技术的核心特征  3.主流关系数据库的挑战  4. 改进型关系数据库 |
| 大数据技术综述 | 1. 大数据概念澄清  2. 大数据技术家族  3. NoSQL技术综述  3.1最早的NoSQL---BDB  3.2 Hadoop之Hbase与Facebook之Cassandra  3.3 MongoDB与CouchDB  3.4Memcached与Redis  3.5图形数据库Neo4j  4. MapReduce  5.关系数据库联邦  6.海量分布式文件系统     7大数据技术理论基础         CAP，BASE，ACID |
| Hadoop实用教程 | 1. Hadoop技术概论  1.1 Hadoop体系架构总论        1.2 HDFS-工作原理与架构        1.3 平民化的分布式计算MapReduce        1.4 MapReduce工作原理与架构        1.5 Hadoop数据仓库-Hive        1.6 Hadoop NoSQL数据库-HBase        1.7 工作流调度-Ooize        1.8 分布式协调系统--Zookeeper  2. Hadoop部署  2.1 Hadoop版本介绍与选择        2.2 Hadoop部署实践      2.3 Hadoop安装文件构成与配置体系      2.4 机器硬件建议配置      2.5 系统环境配置      2.6 基本参数配置与说明      2.7 进程分布规划与启动  3. 分布式文件系统HDFS实用教程  3.1 HDFS操作  3.2 HDFS编程—文件读写  3.3 HDFS数据压缩  3.4 HDFS技术要点  4.  MapReduce实用教程  4.1 MapReduce原理与架构  4.2 MapReduce编程方法  4.3 MapReduce实用技术要点  4.4 MapReduce排序与关联  4.5 MapReduce工作流  4.6 MapReduce调优  5. MapReduce2.0-YARN  5.1 YARN的原理  5.2 YARN设计架构  5.3 YARN工作流程  5.4 YARN与MapReduce1.0比较  6. MapReduce实例讲解  6.1普通实例  6.2 高级实例         6.3 MapReduce高级数据分析（时间允许时）  7. Hadoop数据仓库Hive  7.1 Hive编程        7.2 Hive环境部署与搭建        7.3 Hive工作机制  7.4   Hive语法与实践  8. 其它ZooKeeper，Sqoop，Chukwa，Avro…… |
| Hadoop技术分析 | 1.       Hadoop MapReduce技术解析  6.1关于效率  6.2关于扩展性  6.3关于可靠性与可用性  6.4关于与关系数据库  6.5关于适用的数据类型  6.6关于数据存储与管理  2.       Hadoop与关系数据库  2.1   MapReduce与关系数据库  2.2   Hive与MPP关系数据库 |
| NoSQL实用教程 | 1.NoSQL理论基础---CAP与BASE深入分析  2.NoSQL实用教程     2.1 HBase实用教程       2.1.1 HBase原理       2.1.2 HBase实用安装部署要点       2.1.3 HBase数据模型       2.1.4 HBase索引与关联的实现       2.1.5 HBase使用       2.1.6HBase性能调优       2.1.7 HBase高级设计教程---如何真正用好HBase       2.1.8 HBase与关系数据库结合  3. NoSQL设计实例     3.1 HBase实现全属性查询     3.2 HBase实现时间序列数据管理     3.3 HBase与MapReduce结合示例  4. Facebook Cassandra介绍  5.MongoDB介绍  6.图数据库Neo4J介绍 |
| NoSQL技术分析 | 1. NoSQL技术手段总结  1.1 水平分割  1.2 数据副本与读写一致性  1.3   In-Memory架构  1.4 MVCC  1.5列存储  1.6 COW  2. NoSQL技术解析  2.1   关于水平扩展性  2.2关于模式自由  3. NoSQL与关系数据库     3.1 理论原则分析     3.2 逻辑模型分析     3.3 物理模型分析     3.4 索引、事务与关联     3.5 使用场景定位     3.6 企业数据体系定位 |
| Spark教程 | 1. Spark组成与体系架构     2. Spark原理     3. Spark与Hadoop     4. Scala简介     5. Spark技术流程 |
| 超越Hadoop | 1.       Hadoop技术体系的不足与尴尬  2.       新技术介绍  3.       互联网技术体系介绍  4.       数据管理技术发展趋势分析 |
| 大数据技术实践分享 | 1.       海量数据处理架构设计  2.       大数据驱动与企业业务/运营  3.       实践中的企业大数据分析技术流程  3.1采集---各种方法的比较  3.2存储---原始数据与业务数据提取  3.3模型---Web分析指标体系  3.4分析---大数据分析方法  3.5 行动---个性化推荐  4. 大数据与企业交易  5. 大数据与企业交互  5. 自已设计大数据技术体系 |
| 大数据与企业新一代数据体系建设 | 1.传统的以关系数据库为主的企业数据架构  2.大数据时代的新一代企业数据逻辑架构  2.1   数据分类  2.2   数据分布  2.3   数据流转  2.4   数据集成  2.5   数据交换  2.6   数据分析  2.7   应用展示  3. 新一代企业数据技术架构  3.1逻辑架构  3.2技术方法  3.3物理平台  4.       新一代企业数据架构中Hadoop/NoSQL与关系数据库的相互配合  5.       典型场景示例 |