|  |
| --- |
| 大数据培训指南昆山华信软件技术有限公司 |

**任务计划书**

* 了解Hadoop产生的历史背景、适用场景
* 理解HDFS、MapReduce、YARN、Zookeeper原理
* 掌握在Linux（CentOS 6.5）上安装Mysql、DB2、Oracle数据库
* 掌握Hive、Sqoop、Flume等组件的使用及基本操作
* 掌握TDH的安装和配置,理解各服务重要参数的配置和意义
* 熟悉将RDBMS中的数据导入HDFS和使用Flume将半结构化数据导入HDFS中，并通过数据库连接工具（Squirrel、Tableau）对数据进行查询和优化
* 理解Hive中的内表、外表、ORC格式表、分区表、分桶表、内存表的概念
* 熟悉大数据基准测试TPC-DS工具的使用
* 掌握基本故障的排错方法

1. **任务一**

目标：在Linux中安装JDK、Tomcat、Maven，并安装Mysql、DB2、Oracle数据库。以下所有实验都需要截图并配有相关说明

**\*实验时间：3天**

**\*考核标准**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 姓名 |  |  |  |  |
| 操作文档得分 |  |  |  |  |
| 在线考试得分 |  |  |  |  |
| 总分 |  |  |  |  |

1. **任务二**

目标：掌握安装TDH及其相关组件，安装前必须检查Checklist，并了解各个组件的作用

**\*实验时间：3天**

**\*考核标准**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 姓名 |  |  |  |  |
| 操作文档得分 |  |  |  |  |
| 在线考试得分 |  |  |  |  |
| 总分 |  |  |  |  |

1. **任务三**

目标：使用Sqoop工具实现将RDBMS（Mysql、DB2、Oracle）与HDFS之间的数据迁移

**\*实验时间：1天**

**\*考核标准**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 姓名 |  |  |  |  |
| 操作文档得分 |  |  |  |  |
| 总分 |  |  |  |  |

1. **任务四**

目标：使用Flume工具实现将半结构化数据批量导入HDFS

**\*实验时间：1天**

**\*考核标准**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 姓名 |  |  |  |  |
| 操作文档得分 |  |  |  |  |
| 总分 |  |  |  |  |

1. **任务五**

目标：掌握在Inceptor中建立内表、外表、ORC格式表、分区分桶表、内存表

**\*实验时间：2天**

**\*考核标准**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 姓名 |  |  |  |  |
| 操作文档得分 |  |  |  |  |
| 总分 |  |  |  |  |

1. **任务六**

目标：掌握TPC-DS的使用

**\*实验时间：1天**

**\*考核标准**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 姓名 |  |  |  |  |
| 操作文档得分 |  |  |  |  |
| 总分 |  |  |  |  |

1. **任务七**

目标：掌握通过日志的错误信息进行错误的排查

**\*实验时间：1天**

**\*考核标准**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 姓名 |  |  |  |  |
| 操作文档得分 |  |  |  |  |
| 总分 |  |  |  |  |