

Hadoop讲解

面向对象讲究低冗余高耦合

1) 完全分布式优势:

1、hadoop优势:

高可靠性: 底层维护多个数据副本, 某个计算元或存储出现故障不会导致数据丢失;

2、hadoop3x wordcount方法可直接进行词频统计

3、数据集成框架, 可整合推荐系统框架图

4、非关系型数据库: 堆, 栈,

2) 在linux中 配置网络主要有三种方法 :

(1) 桥接模式 (通过交换机连接外网)

(2) NAT模式VMnet8 (通过自己创建的虚拟路由<vm>ip去连接外网)

(3) 仅主机模式 (直接修改底层网络配置文件VMnet1<不能连接外网>)

以太网通过路由器 (pc) 连接主机

3) 子网掩码的类型:

A类网络缺省子网掩码: 255.0.0.0

B类网络缺省子网掩码: 255.255.0.0

C类网络缺省子网掩码: 255.255.255.0 (民用网)

实训课要求:

PPT,项目说明文档, 项目设计文档, 项目