## 企业应用

第一个方面：

数据分析

第二个方面：

数据检索

## 基本组件

\* Zookeeper

分布式协作框架

节点数目

测试集群：3 个

生产集群：

>> 小型集群, 3 个或者 5个

>> 中型机群，５　个　或者　７个

>> 大型机群, 奇数个

\* HDFS

存储海量数据

\* YARN

集群资源管理

资源调度

\* MapReduce

并行计算框架

思想：

分而治之

分：map

合：reduce

## 集群大小

>>> 测试集群

机器数量：

5 台 - 10 台

机器配置：

>> 内存：24 G / 32 以上

>> 硬盘：4 TB 以上

>> CPU 核：6 核CPU 以上

>> 网卡：万兆

>>> 生产集群

>> 小型集群

机器数量：

20 台以下

>> 中型机群

50 台以下

>> 大型机群

50 台以上

## Hadoop 发行版本

>> Apache

>> CDH

Cloudera公司

只支持64位操作系统

安装

>> tar 包

http://archive.cloudera.com/cdh5/cdh/5/

>> rpm 包

http://archive.cloudera.com/cdh5/redhat/6/x86\_64/cdh/

>> parcels 包

cdh 4.1.2 以后才有的

官方推荐的安装方式

Cloudera Manager进行安装

>> HDP

Hortonworks

## 集群环境准备

### 机器

#### 磁盘阵列

RAID 0 RAID1 JBOD

RAID 1

两块磁盘映射为一块磁盘使用。

CentOS 6.4版本，其中一块磁盘损坏，不会影响系统

JBOD

DataNode存储数据推荐的方式

/dfs/data01

/dfs/data02

/dfs/data03

/dfs/data04

#### 操作系统

redhat 6.x /CentOS 6.x

6.4 6.4

### 系统

#### IP 地址

尽可能在同一网段，

交换机\ 机架Rack

#### 主机名

bigdata-cdh01.ibeifeng.com

bigdata-cdh02.ibeifeng.com

bigdata-cdh03.ibeifeng.com

cmd：

# hostname bigdata-cdh01.ibeifeng.com

# vi /etc/sysconfig/network

#### Ip 与 主机名 映射

# vi /etc/hosts

注意：

如果是客户端的机器，也需要添加映射

>>>禁用IPv6

# echo "alias net-pf-10 off" >> /etc/modprobe.d/dist.conf

# echo "alias ipv6 off" >> /etc/modprobe.d/dist.conf

#### 普通用户

用于安装软件

# adduser beifeng

# passwd beifeng

#### 配置普通用户的sudo权限

/etc/sudoers

>>> 写权限

# chmod u+w /etc/sudoers

>>> 添加

# vi /etc/sudoers

内容：

beifeng ALL=(root)NOPASSWD:ALL

>>> 收回权限

# chmod u-w /etc/sudoers

#### 防火墙

关闭

$ sudo service iptables stop

$ sudo chkconfig iptables off

#### Selinux

禁用

$ sudo vi /etc/sysconfig/selinux

内容：

SELINUX=disabled

#### 卸载JDK

$ sudo rpm -qa|grep java

$ sudo rpm -e --nodeps xxx yyy zzz

#### 设置文件打开数量和用户最大进程数

>> 文件打开数量

$ ulimit -a

>> 用户最大进程数

$ ulimit -u

>> 设置

$ sudo vi /etc/security/limits.conf

内容：

\* soft nofile 65535

\* hard nofile 65535

\* soft nproc 32000

\* hard nproc 32000