

大型架构及配置技术

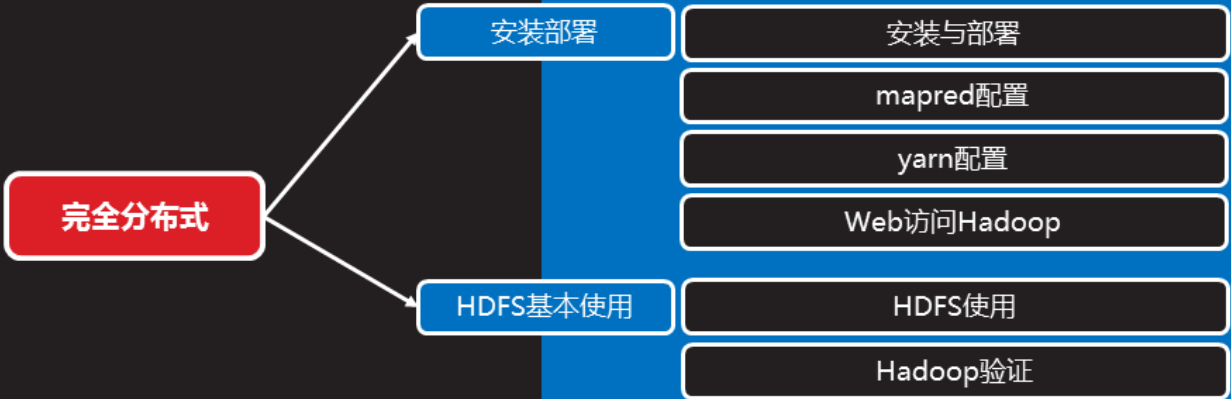
NSD ARCHITECTURE **DAY06**

内容

上午	09:00 ~ 09:30	作业讲解和回顾
	09:30 ~ 10:20	完全分布式
	10:30 ~ 11:20	
	11:30 ~ 12:00	
下午	14:00 ~ 14:50	节点管理
	15:00 ~ 15:50	NFS网关
	16:10 ~ 17:10	
	17:20 ~ 18:00	总结和答疑



完全分布式



安装部署

安装与部署

知识讲解

- Hadoop三大核心组件
 - 分布式文件系统
 - HDFS已经部署完毕
 - 分布式计算框架
 - MapReduce
 - 集群资源管理
 - yarn



安装与部署（续1）

- 系统规划

知识讲解

主机	角色	软件
192.168.1.21 master	NameNode SecondaryNameNode ResourceManager	HDFS YARN
192.168.1.22 node1	DataNode NodeManager	HDFS YARN
192.168.1.23 node2	DataNode NodeManager	HDFS YARN
192.168.1.24 node3	DataNode NodeManager	HDFS YARN



mapred部署

- 分布式计算框架mapred-site.xml

- 改名

FROM : mapred-site.xml.template

To : mapred-site.xml

- 资源管理类

mapreduce.framework.name

知识讲解



mapred部署 (续1)

知识讲解

- 分布式计算框架mapred-site.xml
 - 只支持local和yarn两种
 - 单机使用local
 - 集群使用yarn

```
<property>  
  <name>mapreduce.framework.name</name>  
  <value>yarn</value>  
</property>
```



yarn部署

知识讲解

- 资源管理yarn-site.xml
 - resourcemanager 地址

```
yarn.resourcemanager.hostname
```
 - nodemanager 使用哪个计算框架

```
yarn.nodemanager.aux-services
```
 - mapreduce_shuffle 计算框架的名称

```
mapreduce_shuffle
```



yarn部署 (续1)

- 资源管理yarn-site.xml
 - yarn-site.xml配置

知识讲解

```
<property>
  <name>yarn.resourcemanager.hostname</name>
  <value>nn01</value>
</property>
<property>
  <name>yarn.nodemanager.aux-services</name>
  <value>mapreduce_shuffle</value>
</property>
```



yarn部署 (续2)

- 启动服务
 - # /usr/local/hadoop/sbin/start-yarn.sh
- 验证服务
 - /usr/local/hadoop/bin/yarn node -list

知识讲解

```
[root@nn01 hadoop]# ./bin/yarn node -list
```

```
... ..
```

node2:33486	RUNNING	node2:8042	0
node1:35816	RUNNING	node1:8042	0
node3:40941	RUNNING	node3:8042	0



Web访问Hadoop

知识讲解

- 使用Web访问Hadoop
 - namenode web页面(nn01)
http://192.168.1.10:50070/
 - secondary namenode web 页面(nn01)
http://192.168.1.10:50090/
 - datanode web 页面(node1,node2,node3)
http://192.168.1.11:50075/



Web访问Hadoop (续1)

知识讲解

- 使用Web访问Hadoop
 - resourcemanager web页面(nn01)
http://192.168.1.10:8088/
 - nodemanager web页面(node1,node2,node3)
http://192.168.1.11:8042/



HDFS基本使用

HDFS使用

知识讲解

- HDFS基本命令

- # ./bin/hadoop fs -ls /

- 对应shell命令

- # ls /

- # ./bin/hadoop fs -mkdir /abc

- 对应shell命令

- # mkdir /abc

- # ./bin/hadoop fs -rmdir /abc

- 对应shell命令

- # rmdir /abc



HDFS使用（续1）

知识讲解

- HDFS基本命令

- # ./bin/hadoop fs -touchz /urfile

- 对应shell命令

- touch /urfile

- # ./bin/hadoop fs -cat /urfile

- 对应shell命令

- # cat /urfile

- # ./bin/hadoop fs -rm /urfile

- 对应shell命令

- # rm /urfile



HDFS使用（续2）

知识讲解

- HDFS基本命令

- 上传文件

- # ./bin/hadoop fs -put localfile /remotefile

- 下载文件

- # ./bin/hadoop fs -get /remotefile



案例2：Hadoop词频统计

课堂练习

1. 在集群文件系统里创建文件夹
2. 上传要分析的文件到目录中
3. 分析上传文件
4. 展示结果



Hadoop验证

知识讲解

- 创建文件夹

```
# ./bin/hadoop fs -mkdir /input
```
- 上传要分析的文件

```
# ./bin/hadoop fs -put *.txt /input
```



Hadoop验证（续1）

知识讲解

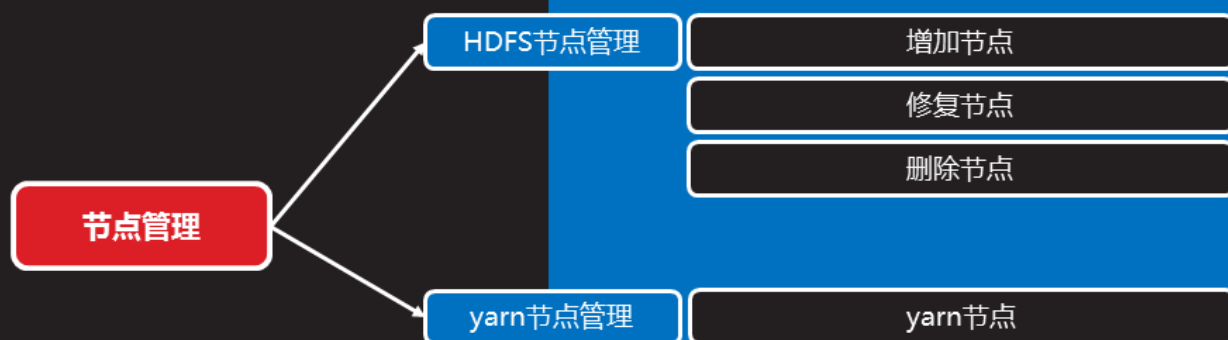
- 提交分析作业

```
# ./bin/Hadoop  
# jar ./share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.3.jar wordcount /input /output
```
- 查看结果

```
# ./bin/hadoop fs -cat output/*
```



节点管理



HDFS节点管理

增加节点

知识讲解

- HDFS增加结点
 - 启动一个新的系统，禁用Selinux、禁用firewalld
 - 设置SSH免密码登录
 - 在所有节点修改/etc/hosts，增加新节点的主机信息
 - 安装java运行环境（java-1.8.0-openjdk-devel）
 - 拷贝NameNode的/usr/local/hadoop到本机
 - 修改NameNode的slaves文件增加该节点
 - 在该节点启动DataNode
`./sbin/hadoop-daemon.sh start datanode`



增加节点（续1）

知识讲解

- HDFS节点管理
 - 设置同步带宽，并同步数据
`# ./bin/hdfs dfsadmin -setBalancerBandwidth 67108864`
`# ./sbin/start-balancer.sh`
 - 查看集群状态
`# ./bin/hdfs dfsadmin -report`



修复节点

知识讲解

- HDFS修复节点
 - 修复节点比较简单，与增加节点基本一致
 - 注意：新节点的ip和主机名要与损坏节点的一致
 - 启动服务

```
# ./sbin/hadoop-daemon.sh start datanode
```
 - 数据恢复是自动的
 - 上线以后会自动恢复数据，如果数据量非常巨大，可能需要一定的时间



删除节点

知识讲解

- HDFS删除节点
 - 配置NameNode的hdfs-site.xml
 - 增加dfs.hosts.exclude配置

```
<property>  
  <name>dfs.hosts.exclude</name>  
  <value>/usr/local/hadoop/etc/hadoop/exclude</value>  
</property>
```
 - 增加exclude配置文件，写入要删除的节点主机名
 - 更新数据

```
# ./bin/hdfs dfsadmin -refreshNodes
```



删除节点（续1）

知识讲解

- HDFS删除节点状态
 - 查看状态

```
# ./bin/hdfs dfsadmin -report
```
 - Normal : 正常状态
 - Decommissioned in Program : 数据正在迁移
 - Decommissioned : 数据迁移完成
 - 注意：仅当状态变成Decommissioned才能down机下线



案例3：节点管理

课堂练习

1. 增加一个新的节点
2. 查看状态
3. 删除节点

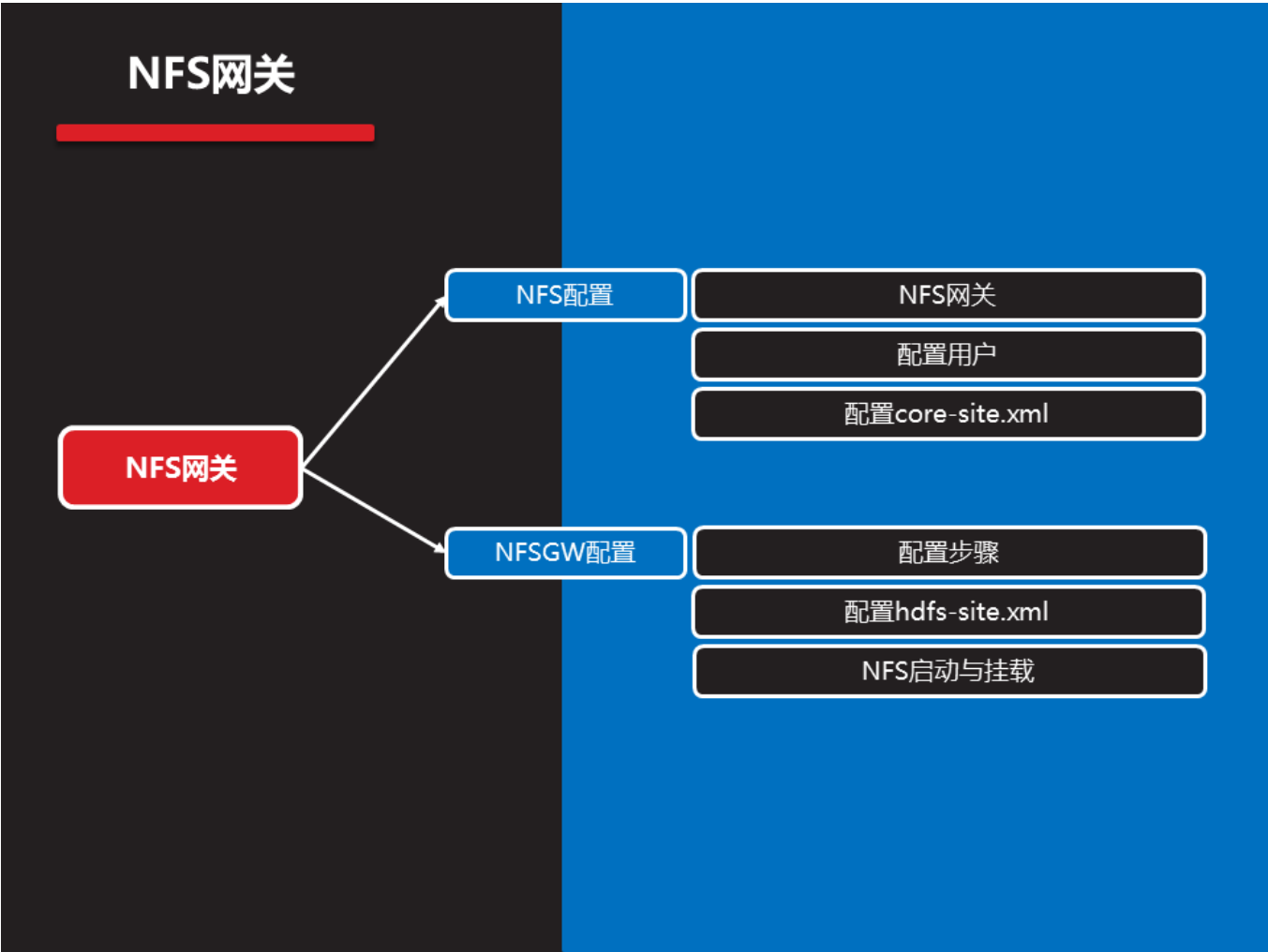


yarn节点管理

yarn节点

- yarn的相关操作
 - 由于Hadoop在2.x引入了yarn框架，对于计算节点的操作已经变得非常简单
 - 增加节点
 - # `sbin/yarn-daemon.sh start nodemanager`
 - 删除节点
 - # `sbin/yarn-daemon.sh stop nodemanager`
 - 查看节点 (ResourceManager)
 - # `./bin/yarn node -list`





NFS配置





