# NN

## GC

## Core-site.xml

### 基本参数

#### 指定hdfs访问地址

<property>

<name>fs.defaultFS</name>

<value>hdfs://master:9000</value>

</property>

#### 数据临时目录

<property>

<name>hadoop.tmp.dir</name>

<value>/home/hadoopData</value>

</property>

### 优化

# DN

## HDFS-site.xml

### 优化

#### dfs.datanode.handler.count

– NameNode用来处理来自DataNode的RPC请求的线程数量

– 建议设置为DataNode数量的10%，一般在10~200个之间

默认值10。设置该值的一般原则是将其设置为集群大小的自然对数乘以20，即20logN，N为集群大小。

如果该值设的太小，会出现的情况是DataNode在连接NameNode的时候总是超时或者连接被拒绝

– 在NameNode上设定

– 默认值：10

#### **dfs.datanode.handler.count**

– DataNode用来连接NameNode的RPC请求的线程数量

– 取决于系统的繁忙程度

– 设置太小会导致性能下降甚至报错

– 在DataNode上设定

– 默认值：3

#### **dfs.datanode.max.xcievers**

– DataNode可以同时处理的数据传输连接数

– 默认值：256

– 建议值：4096

#### dfs.blockreport..intervalMsec

datanode向namenode报告块信息的时间间隔，默认6小时

#### dfs.datanode.directoryscan.interval

datanode进行内存和磁盘数据集块校验，更新内存中的信息和磁盘中信息的不一致情况，默认6小时.

### 基本参数

#### NameNode数据目录

<property>

<name>dfs.namenode.name.dir</name>

<value>/data/namenode</value>

</property>

#### DataNode数据目录

<property>

<name>dfs.datanode.data.dir</name>

<value>/data/datanode</value>

</property>

#### 副本数

<property>

<name>dfs.replication</name>

<value>3</value>

</property>

#### 是否启用webhdfs

<property>

<name>dfs.webhdfs.enabled</name>

<value>true</value>

</property>

#### 权限

<property>

<name>dfs.permissions.superusergroup</name>

<value>staff</value>

</property>

<property>

<name>dfs.permissions.enabled</name>

<value>false</value>

</property>

#### SecondaryNameNode配置

<property>

<name>dfs.namenode.secondary.http-address</name>

<value>linux123:50090</value>

</property>

#### 指定DataNode节点

在slaves中加入DataNode节点。