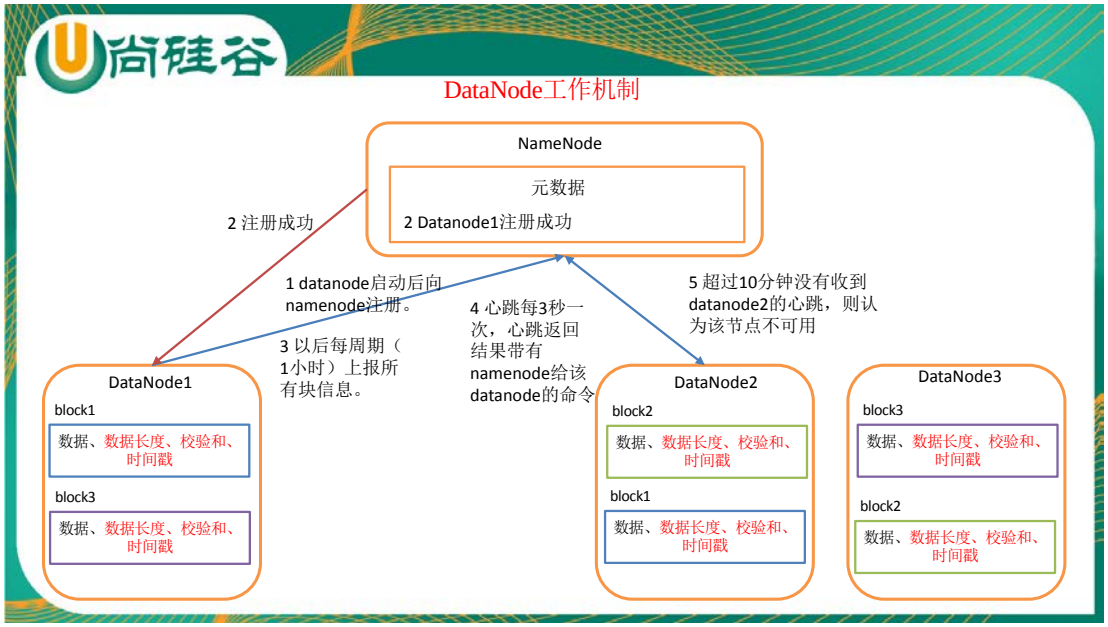
六 DataNode

6.1 DataNode 工作机制



1）一个数据块在 DataNode 上以文件形式存储在磁盘上，包括两个文件，一个是数据本身，一个是元数据包括数据块的长度，块数据的校验和，以及时间戳。

2）DataNode 启动后向 NameNode 注册，通过后，周期性（1 小时）的向 NameNode 上报所有的块信息。

3）心跳是每 3 秒一次，心跳返回结果带有 NameNode 给该 DataNode 的命令如复制块数据到另一台机器，或删除某个数据块。如果超过 10 分钟没有收到某个 DataNode 的心跳，则认为该节点不可用。

4）集群运行中可以安全加入和退出一些机器。

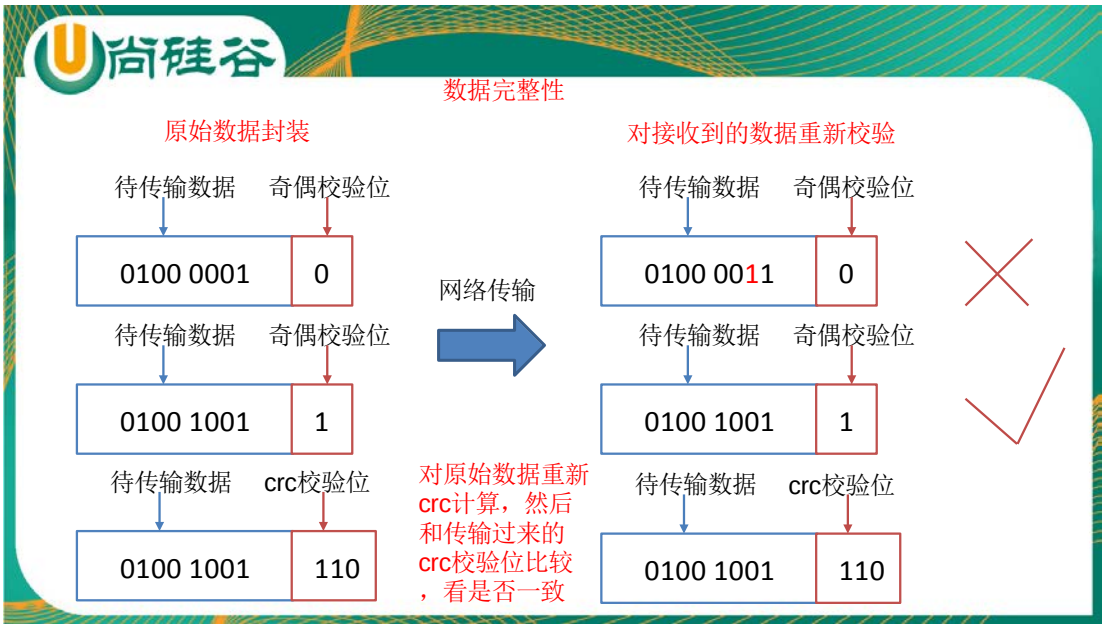
6.2 数据完整性

1）当 DataNode 读取 block 的时候，它会计算 checksum。

2）如果计算后的 checksum，与 block 创建时值不一样，说明 block 已经损坏。

3）client 读取其他 DataNode 上的 block。

4）datanode 在其文件创建后周期验证 checksum。



6.3 掉线时限参数设置

DataNode 进程死亡或者网络故障造成 DataNode 无法与 NameNode 通信，NameNode 不会立即把该节点判定为死亡，要经过一段时间，这段时间暂称作超时时长。HDFS 默认的超时时长为 10 分钟+30 秒。如果定义超时时间为 timeout，则超时时长的计算公式为：

timeout = 2 * dfs.namenode.heartbeat.recheck-interval + 10 * dfs.heartbeat.interval。

而默认的 dfs.namenode.heartbeat.recheck-interval 大小为 5 分钟，dfs.heartbeat.interval 默认为 3 秒。

需要注意的是 hdfs-site.xml 配置文件中的 heartbeat.recheck.interval 的单位为毫秒，dfs.heartbeat.interval 的单位为秒。

```
<property>
    <name>dfs.namenode.heartbeat.recheck-interval</name>
    <value>300000</value>
</property>
<property>
    <name> dfs.heartbeat.interval </name>
    <value>3</value>
</property>
```

6.4 服役新数据节点

0）需求：

随着公司业务的增长，数据量越来越大，原有的数据节点的容量已经不能满足存储数据的需求，需要在原有集群基础上动态添加新的数据节点。

1）环境准备

（1）克隆一台虚拟机

（2）修改 ip 地址和主机名称

（3）修改 xsync 文件，增加新增节点的 ssh 无密登录配置

（4）删除原来 HDFS 文件系统留存的文件

/opt/module/hadoop-2.7.2/data

2）服役新节点具体步骤

（1）在 namenode 的/opt/module/hadoop-2.7.2/etc/hadoop 目录下创建 dfs.hosts 文件

[atguigu@hadoop105 hadoop]$ pwd

/opt/module/hadoop-2.7.2/etc/hadoop

[atguigu@hadoop105 hadoop]$ touch dfs.hosts

[atguigu@hadoop105 hadoop]$ vi dfs.hosts

添加如下主机名称（包含新服役的节点）

hadoop102

hadoop103

hadoop104

hadoop105

（2）在 namenode 的 hdfs-site.xml 配置文件中增加 dfs.hosts 属性

```
<property>
        <name>dfs.hosts</name>
        <value>/opt/module/hadoop-2.7.2/etc/hadoop/dfs.hosts</value>
</property>
```

（3）刷新 namenode

[atguigu@hadoop102 hadoop-2.7.2]$ hdfs dfsadmin -refreshNodes

Refresh nodes successful

（4）更新 resourcemanager 节点

[atguigu@hadoop102 hadoop-2.7.2]$ yarn rmadmin -refreshNodes

17/06/24 14:17:11 INFO client.RMProxy: Connecting to ResourceManager at hadoop103/192.168.1.103:8033

（5）在 NameNode 的 slaves 文件中增加新主机名称

增加 105

hadoop102

hadoop103

hadoop104

hadoop105

（6）单独命令启动新的数据节点和节点管理器

[atguigu@hadoop105 hadoop-2.7.2]$ sbin/hadoop-daemon.sh start datanode

starting datanode, logging to /opt/module/hadoop-2.7.2/logs/hadoop-atguigu-datanode-hadoop105.out

[atguigu@hadoop105 hadoop-2.7.2]$ sbin/yarn-daemon.sh start nodemanager

starting nodemanager, logging to /opt/module/hadoop-2.7.2/logs/yarn-atguigu-nodemanager-hadoop105.out

（7）在 web 浏览器上检查是否 ok

3）如果数据不均衡，可以用命令实现集群的再平衡

[atguigu@hadoop102 sbin]$ ./start-balancer.sh

starting balancer, logging to /opt/module/hadoop-2.7.2/logs/hadoop-atguigu-balancer-hadoop102.out

Time Stamp Iteration# Bytes Already Moved Bytes Left To Move Bytes Being Moved

6.5 退役旧数据节点

1）在 namenode 的/opt/module/hadoop-2.7.2/etc/hadoop 目录下创建 dfs.hosts.exclude 文件

[atguigu@hadoop102 hadoop]$ pwd

/opt/module/hadoop-2.7.2/etc/hadoop

[atguigu@hadoop102 hadoop]$ touch dfs.hosts.exclude

[atguigu@hadoop102 hadoop]$ vi dfs.hosts.exclude

添加如下主机名称（要退役的节点）

hadoop105

2）在 namenode 的 hdfs-site.xml 配置文件中增加 dfs.hosts.exclude 属性

```
<property>
        <name>dfs.hosts.exclude</name>
        <value>/opt/module/hadoop-2.7.2/etc/hadoop/dfs.hosts.exclude</value>
```

```
</property>
```

3）刷新 namenode、刷新 resourcemanager

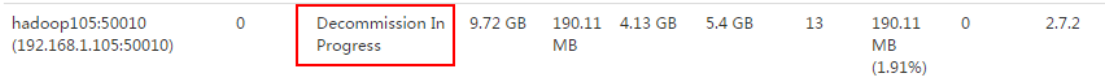
[atguigu@hadoop102 hadoop-2.7.2]$ hdfs dfsadmin -refreshNodes

Refresh nodes successful

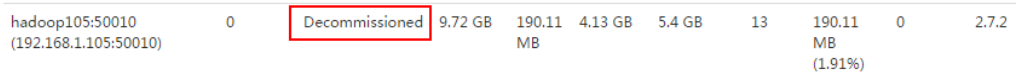
[atguigu@hadoop102 hadoop-2.7.2]$ yarn rmadmin -refreshNodes

17/06/24 14:55:56 INFO client.RMProxy: Connecting to ResourceManager at hadoop103/192.168.1.103:8033

4）检查 web 浏览器，退役节点的状态为 decommission in progress（退役中），说明数据节点正在复制块到其他节点。



5）等待退役节点状态为 decommissioned（所有块已经复制完成），停止该节点及节点资源管理器。注意：如果副本数是 3，服役的节点小于等于 3，是不能退役成功的，需要修改副本数后才能退役。



[atguigu@hadoop105 hadoop-2.7.2]$ sbin/hadoop-daemon.sh stop datanode

stopping datanode

[atguigu@hadoop105 hadoop-2.7.2]$ sbin/yarn-daemon.sh stop nodemanager

stopping nodemanager

6）从 include 文件中删除退役节点，再运行刷新节点的命令

（1）从 namenode 的 dfs.hosts 文件中删除退役节点 hadoop105

hadoop102

hadoop103

hadoop104

（2）刷新 namenode，刷新 resourcemanager

[atguigu@hadoop102 hadoop-2.7.2]$ hdfs dfsadmin -refreshNodes

Refresh nodes successful

[atguigu@hadoop102 hadoop-2.7.2]$ yarn rmadmin -refreshNodes

17/06/24 14:55:56 INFO client.RMProxy: Connecting to ResourceManager at

hadoop103/192.168.1.103:8033

7）从 namenode 的 slave 文件中删除退役节点 hadoop105

hadoop102

hadoop103

hadoop104

8）如果数据不均衡，可以用命令实现集群的再平衡

[atguigu@hadoop102 hadoop-2.7.2]$ sbin/start-balancer.sh

starting balancer, logging to /opt/module/hadoop-2.7.2/logs/hadoop-atguigu-balancer-hadoop102.out

Time Stamp Iteration# Bytes Already Moved Bytes Left To Move Bytes Being Moved

6.6 Datanode 多目录配置

1）datanode 也可以配置成多个目录，每个目录存储的数据不一样。即：数据不是副本。

2）具体配置如下：

hdfs-site.xml

```
<property>
        <name>dfs.datanode.data.dir</name>
<value>file:///${hadoop.tmp.dir}/dfs/data1,file:///${hadoop.tmp.dir}/dfs/data2</value>
</property>
```