Hbase-2.0.5配置

一、HBase支持随机写

HBase的读写操作还是借助HDFS完成，要完成随机写，根本上还是需要复合HDFS的特性！

HDFS只支持追加写！

随机的操作： Update+Delete 借助 追加写+时间戳(版本号)

只允许客户端查询时返回时间戳最新的数据！

二、 HBase支持海量数据的实时读写

①分布式

②索引，LSM树

③kv

④吃内存

⑤列式存储

⑥布隆过滤器（查询）

三、HBase是什么

HBase是基于hadoop的数据库！

提供一个十亿级行\*百万级列级别的表存储，对表中的数据提供实时的随机读写操作！

四、HBase的安装和配置

前提：保证已经配置了JAVA\_HOME,HADOOP\_HOME环境变量

保证Hadoop，zookeeper正常部署

**1、配置**

(1)将hbase解压到指定目录

(2)编辑conf/hbase-env.sh

显示行号：set nu(在非编辑状态下)

128行： export HBASE\_MANAGES\_ZK=false（不用HBase自带的zookeeper就改为false）

(3)修改配置在:hbase/conf/hbase-site.xml

<property>

<name>hbase.rootdir</name>

<value>hdfs://hadoop1:9000/hbase</value>

</property>

<property>

<name>hbase.cluster.distributed</name>

<value>true</value>

</property>

<property>

<name>hbase.zookeeper.quorum</name>

<value>hadoop1,hadoop2,hadoop3</value>

</property>

<property>

<name>hbase.unsafe.stream.capability.enforce</name>

<value>false</value>

</property>

<property>

<name>hbase.wal.provider</name>

<value>filesystem</value>

</property>

(4)编辑regionservers

hadoop101

hadoop102

hadoop103

(5)将hbase远程发送给其它集群

xsync hbase/

**2、启动**

启动顺序

(1) hadoop

(2)zookeeper

(2)hbase

启动方式

(1) bin/hbase-daemon.sh start master

bin/hbase-daemon.sh start regionserver

(2) 群起：bin/start-hbase.sh

群停：bin/stop-hbase.sh

**3、查看：**

jps查看

访问web界面： 访问master进程所在机器:16010

端口说明： 16000是master进程的RPC端口！

16010是master进程的http端口！

16020是RegionServer进程的RPC端口！

16030是RegionServer进程的http端口！

**4、提示：可能会出现的问题**

(1)如果集群之间的节点时间不同步，会导致regionserver无法启动，抛出ClockOutOfSyncException异常。

修复提示：

a、同步时间服务

请参看帮助文档：《尚硅谷大数据技术之Hadoop入门》

b、属性：hbase.master.maxclockskew设置更大的值

<property>

<name>hbase.master.maxclockskew</name>

<value>180000</value>

<description>Time difference of regionserver from master</description>

</property>

**(2)如果hbase开启后HM过几秒就自动关闭，但是HRegionServer进程正常运行**

查看Hadoop的core-site.xml 文件的端口号是否与hbase的hbase-site.xml端口号是否一致。

修改hbase-site.xml的端口号与hadoop的端口号一致。



