**如何配置UDH HBase Replication集群**

作者：费英林

目录

[1. HBase主从集群安装 2](#_Toc423332467)

[1.1. 主集群安装 2](#_Toc423332468)

[1.2. 从集群安装 2](#_Toc423332469)

[2. 定义表结构 3](#_Toc423332470)

[3. 定义Replication Peer 3](#_Toc423332471)

[4. 历史数据迁移 3](#_Toc423332472)

[5. 验证 4](#_Toc423332473)

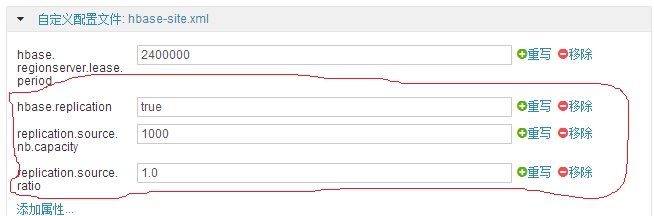
[附：常用的replication命令 4](#_Toc423332474)

HBase Replication是HBase集群间的一种数据备份机制，主集群可将数据复制到一个或多个从集群。

# HBase主从集群安装

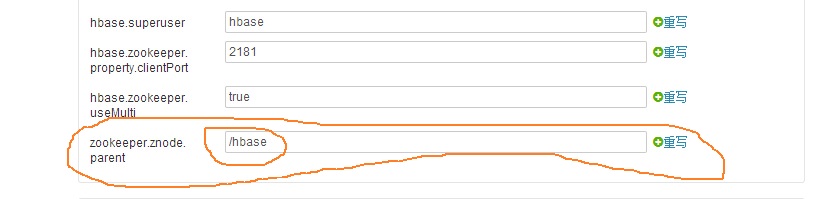
## 主集群安装

通过UDH安装HBase集群，打开HBase配置页面，点击“自定义配置文件: hbase-site.xml”，添加如下属性后重启服务：



## 从集群安装

通过UDH安装HBase集群，打开HBase配置页面，点击“高级”，修改如下属性后重启服务：



# 定义表结构

在HBase主集群和从集群上创建相同的表结构，打开Replication模式。Replication是基于列族的备份机制。例如：

create 't1', {NAME => 'colfam1', REPLICATION\_SCOPE => 1}

表t1的colfam1列族数据将从主集群复制到从集群。主从集群上的表结构定义要完全一致。

# 定义Replication Peer

在主集群上打开hbase shell，添加peer:

add\_peer '1', "search6.yonyou.com,search7.yonyou.com,search8.yonyou.com:2181:/hbase"

其中：

* + - ‘1’是peer ID，必须是一个short类型的整数；
    - "search6.yonyou.com,search7.yonyou.com,search8.yonyou.com:2181:/hbase"是从集群的连接串，定义为“从集群的ZK集群：从集群的ZK端口：从集群的HBase ZK节点（这个值要与前述画面中定义的值一致，必须是/hbase）”

# 历史数据迁移

如果在启用Replication之间HBase中的表中已经存在数据，这部分数据需要手动拷贝到从集群，可以使用HBase提供的CopyTable工具，下面是几个例子：

hbase org.apache.hadoop.hbase.mapreduce.CopyTable \

--peer.adr=search6.yonyou.com:2181:/hbase \

--families=colfam1 \

t1

以上命令是将表t1中列族colfam1的所有数据拷贝到从集群的t1表中。

hbase org.apache.hadoop.hbase.mapreduce.CopyTable \

--starttime=1 --endtime=1432037334102 \

--peer.adr=search6.yonyou.com:2181:/hbase \

--families= colfam1 \

t1

以上命令是将表t1中列族colfam1满足时间条件的所有数据拷贝到从集群的t1表中，其中starttime为1表示起始时间是1（最小值），结束时间是1432037334102。

hbase org.apache.hadoop.hbase.mapreduce.CopyTable \

--starttime= 1432038826918 --endtime= 1432522466099\

--peer.adr=search6.yonyou.com:2181:/hbase \

--families= colfam1 \

t1

以上命令是将表t1中列族colfam1满足时间条件的所有数据拷贝到从集群的t1表中，其中起始时间是1432038826918，结束时间是1432522466099。

# 验证

在主集群中打开HBase shell，新增一条记录：

put 't1', 'row1', 'colfam1', 'value1'

在从集群中打开HBase shell，查询复制过来的数据：

get 't1', 'row1'

# 附：常用的replication命令

add\_peer <ID> <CLUSTER\_KEY>

peer添加命令

list\_peers

列出本集群所有的peer

enable\_peer <ID>

启用之前禁用的peer

disable\_peer <ID>

禁用一个peer

remove\_peer <ID>

删除一个peer

enable\_table\_replication <TABLE\_NAME>

启用一个表中所有列族的replication

disable\_table\_replication <TABLE\_NAME>

禁用一个表中所有列族的replication