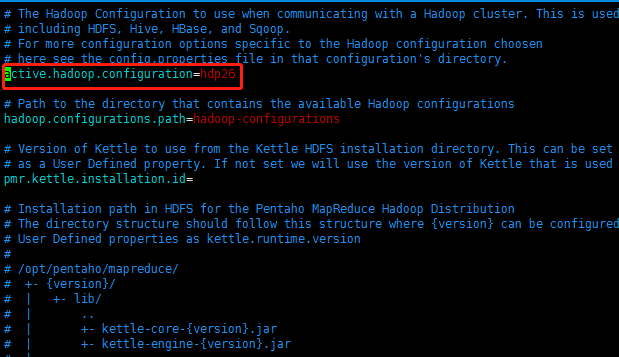
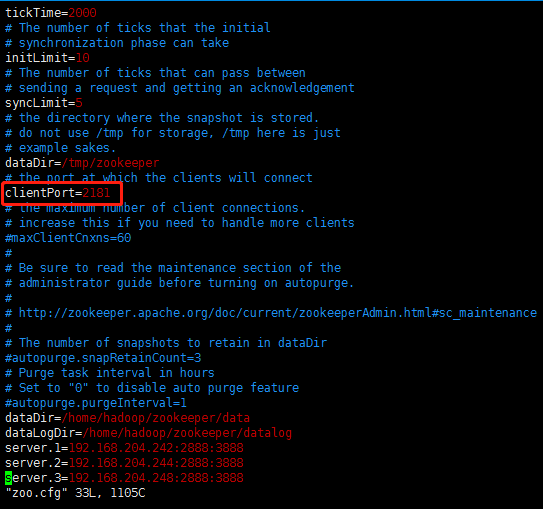
本次用的是hadoop2.9.2,kettle8.2在安装目录下/home/hadoop/kettle/data-integration/plugins/pentaho-big-data-plugin/hadoop-configurations可以看到该版本支持的是hdp26和hdp30（即支持2.6.x和3.0.x），在网上找不到hadoop2.9.2的shim包，但是返现hdp26可以支持hadoop2.9.2

在目录/home/hadoop/kettle/data-integration/plugins/pentaho-big-data-plugin下更改plugin.properties,将active.hadoop.configuration属性值设置为支持的hdp26



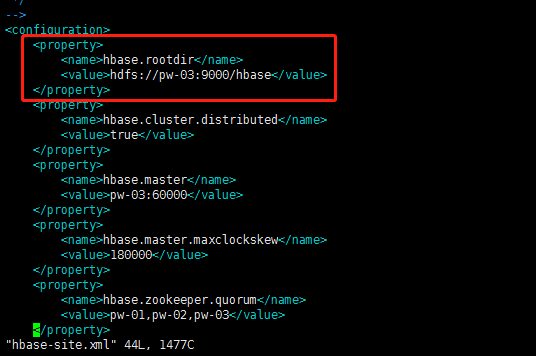
查看zookeeper的端口

在zookeeper的安装目录下/home/hadoop/zookeeper/zookeeper-3.4.10/conf查看zoo.cfg

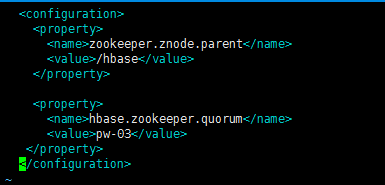


查看hbase的配置文件及端口

在hbase的安装目录/home/hadoop/hbase/conf查看hbase-site.xml，可以看到端口为9000，并且hbase数据的目录是在/hbase下



设置kettle的大数据配置，进入到kettle用到的shim包，进入/home/hadoop/kettle/data-integration/plugins/pentaho-big-data-plugin/hadoop-configurations/hdp26下，查看hbase-site.xml



<configuration>

<property>

<name>zookeeper.znode.parent</name>

<value>/hbase</value>

</property>

<property>

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

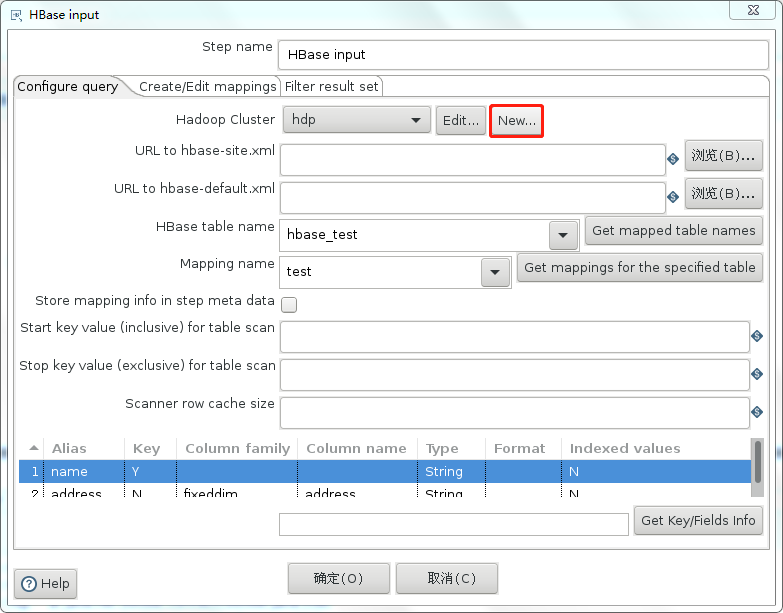
<value>pw-03</value>

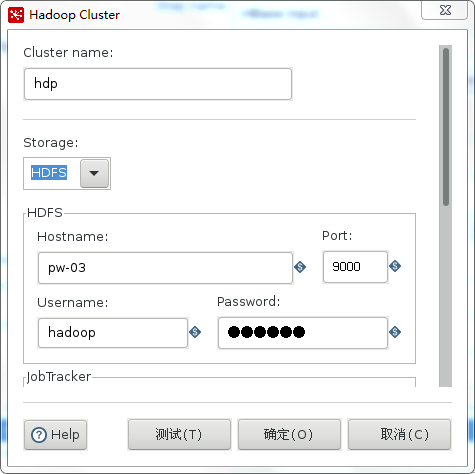
</property>

</configuration>

zookeeper.znode.parent设置为hbase数据的目录/habse,将hbase.zookeeper.quorum配置为zookeeper的主服务器(这里是pw-03)

新建转换，点击 核心对象-Big Data-HBase Input,新建一个Hadoop Cluster(hadoop集群)

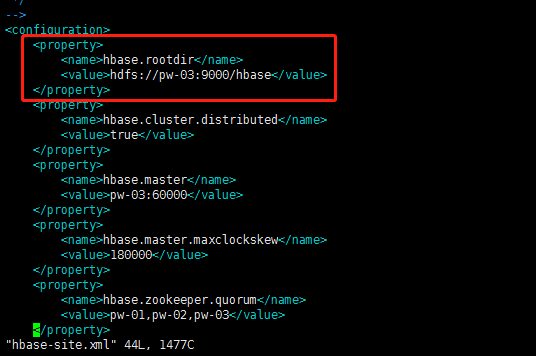




Cluster name是设置连接的名称(自己随便填)，Storage默认是选HDFS(Hadoop的类型),HDFS里面的内容是Hbase的配置

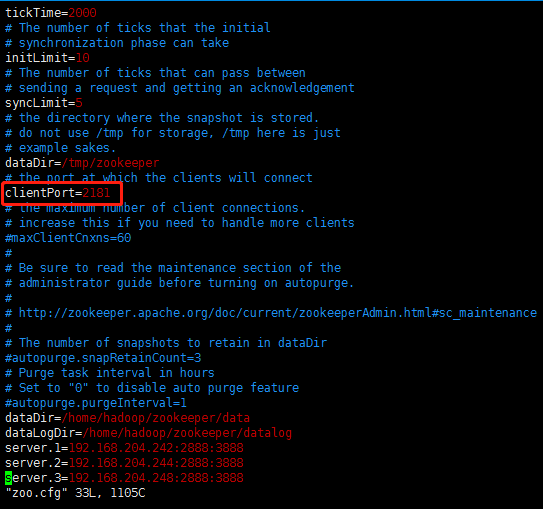
Hostname和port都是从hbase的配置文件来的，这里

Hostname是pw-03 port是9000 username和password是hbase安装所在的用户名密码

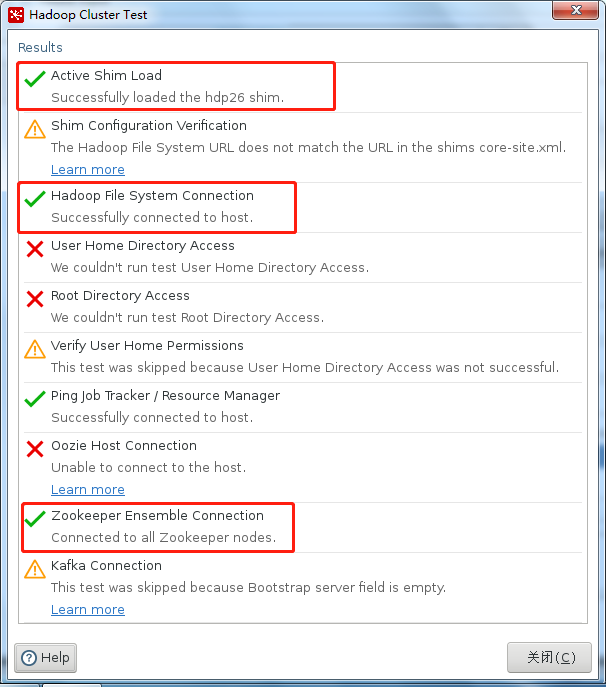


JobTracker中Hostname选择hadoop的用户pw-03，port选择默认的8032，

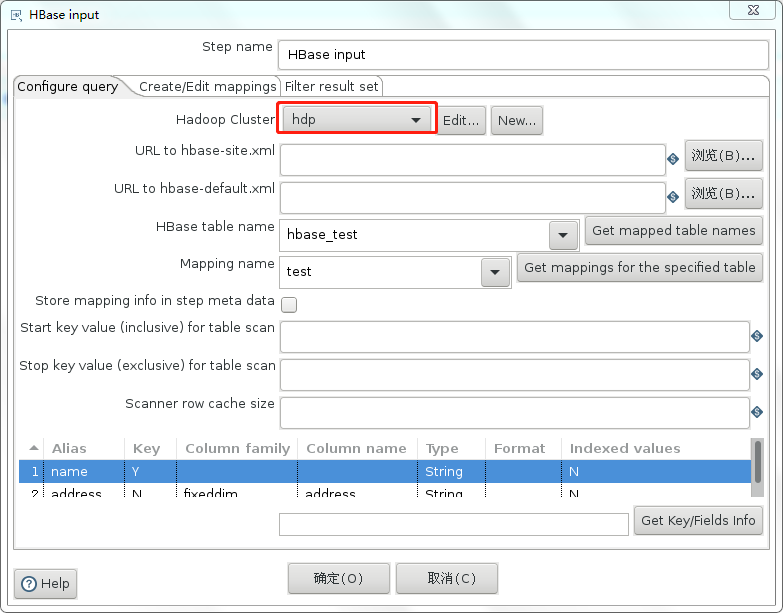
Zookeeper配置，Hostname选择主服务器pw-03,port根据配置文件选择，这里是2181



然后测试，可以看到shim包加载完成，zookeeper连接完成



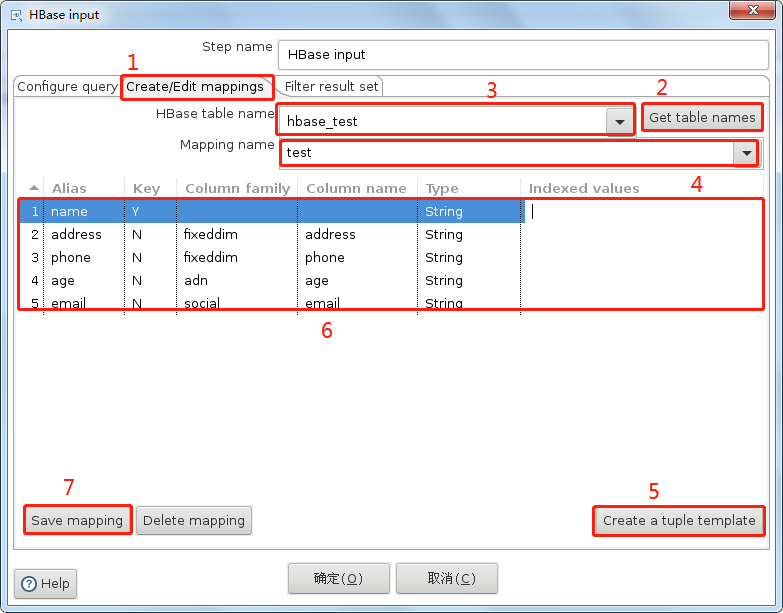
完成后再Hadoop Cluster里面添加进去



点击Create/Edit mappings

1. 点击Get table names,获取需要的表名
2. Mapping name自己取个名字
3. 点击Create a tuple template获取一个模板
4. 按顺序填写Column mappings(列簇)，Column name(列名),Type(字段类型)，Alias(对应列名

5.点击Save mapping



然后点击Configure query

1. 点击Get mapped table names获取表名
2. 点击 Get mappings for the specified table获取映射名
3. 点击Get Key/Fields info 获取字段对应表
4. 点击确定

