# hadoop

1、启动后，直接窒执行 hdfs dfs -ls / 的时候报错:fs.dfaultFS is not set类似的信息

问题分析，直接使用的命令，是/bin下的hdfs或者hadoop命令，这个命令是在使用yum安装impala的时候

安装的，并不是指向我所安装的hadoop的可执行文件，所以，我的配置都没有用到。

解决：删除/bin下的hadoop和hdfs，把我安装的hadoop的目录添加到环境变量，source /etc/profile一次即可

# hive

1、zookeerper问题

由于使用的是内置的，所以不需要单独安装了

2、启动错误，直接使用的是/bin下的hive脚本，但是这个脚本指向的并不是我安装的hive的可执行文件。

问题分析：在装impala的时候，依赖里有hive，导致又装了一次hive，并且自动设置了环境变量

解决，删除 /bin里的hive脚本，把我安装的hive目录添加到环境变量，source /etc/profile一次即可

3、对hive metastore 和 hiveserver2的理解

hive metastore统一管理与mysql的连接，多个hiveserver2直接与hive metastore连接，不用与mysql连接

好处是万一要换mysql，只需要更改hive metastore里的mysql配置就行，其他的hiveserver2都不需要动。

hive metastoer启动方式：nohup hive --service metastore &

hiveserver2启动方式 ：nohup hive --service hiveservice2 &

4、hue建表报错：

org.apache.hadoop.hive.ql.exec.DDLTask. MetaException(

message:Got exception: org.apache.hadoop.ipc.RemoteException User: hdfs is not allowed to impersonate hdfs)

问题分析：不清楚

解决：修改hadoop配置文件core-site.xml，标红的root不确定是否对，我的部署中写的是hadoop和root

<property>

      <name>hadoop.proxyuser.root.groups</name>

      <value>root</value>

      <description>Allow the superuser oozie to impersonate any members of the group group1 and group2</description>

 </property>

 <property>

      <name>hadoop.proxyuser.root.hosts</name>

      <value>\*</value>

      <description>The superuser can connect only from host1 and host2 to impersonate a user</description>

  </property>

# impala

注意，hbase不是给impala用的，装多了

1、因为impala是yum安装的，所以不存在hadoop和hive里第一条的问题，直接使用/bin下的脚本即可

2、impala各个角色

 1.impala-state-store   -- 协调服务

          2.impala-catalog --名称服务

          3.impala-server(impalad)  节点服务

         4.impala-shell 交互服务

    其中 state-store和catalog类似于 namenode一般部署在主节点，而impala-server类似于datenode部署在子节点上

    最后impala-shell可以部署在任意几点上。

3、节点服务启动失败，

报错：

Impalad services did not start correctly,exiting.Error:Couldn't open transport for impala2:24000 (connect() failed: Connection refused)

问题分析：

结合impala.2，知道“节点服务”肯定要连接到“协调服务”和“名称服务”，这里配置的impala2是本机，当然就错了

解决：配置/etc/default/impala，把里面的IMPALA\_CATALOG\_SERVICE\_HOST= 和 IMPALA\_STATE\_STORE\_HOST= 设置为“协调服务”和“名称服务”所在的机器

4、创建数据库报错create database impalaTest;

报错：

Query: create database impalaTest

ERROR: ImpalaRuntimeException: Error making 'createDatabase' RPC to Hive Metastore:

CAUSED BY: MetaException: Got exception: org.apache.hadoop.security.AccessControlException Permission denied: user=impala, access=WRITE, inode="/user/hive/warehouse":root:supergroup:drwxr-xr-x

问题分析：hadoop的hdfs里用户权限问题，如果想设置权限，要开启ACL功能：

在hdfs-site.xml里添加：

<property>

<name>dfs.namenode.acls.enabled</name>

<value>true</value>

</property>

然后进行授权：

切到 $hadoop/bin下执行sudo -u root ./hdfs dfs -setfacl -m user:impala:rwx /user/hive/\*

这里的root应替换为/user/hive/\*所属用户名

参考：<https://blog.csdn.net/fenglibing/article/details/82261340>

Hadoop权限管理参考：<http://hadoop.apache.org/docs/r1.0.4/cn/hdfs_permissions_guide.html>

5、impala和hive关系

impala只是使用到了hive的metadata(元数据)

impala的计算和hive无关，impala的计算利用impala-server

所以，impala里只要配置hive的metadata就好了

除此之外hive集群和impala集群没有关系

# hue

1、hue Could not connect to localhost:10000

所在机器需要启动hiveserver2

nohup huve --service hiveserver2

2、Failed to open new session: java.lang.RuntimeException: java.lang.RuntimeException: org.apache.hadoop.security.AccessControlException: Permission denied: user=hadoop, access=EXECUTE, inode="/tmp/hive/hdfs":hdfs:supergroup:drwx------

首次打开网页时设置的用户时hadoop，没有权限，参考impala.4授权

# kudu

1、/etc/default/kudu-master里没有添加其它master的地址，

导致“选举”失败，老是有master挂掉

2、kudu2服务器的hostname错误，

在更改hostname的时候，虽然/etc/sysconfig/network里和/etc/hosts里都写对了，但在执行 hostname kudu2的时候， 执行了hostname kude2 ,第二个u写成了e

导致 kudu1和kudu3上的tserver老是因为连不是kude2而挂掉

# streamsets

1、最大打开文件描述符数只有1024，导致启动失败。

解决办法：

ulimit -n 10240

# server 运行错误

问题记录：

1、

按照上午说的方式启动：

source /home/kela/env/bin/activate

cd /home/kela/server

/home/kela/env/bin/python /home/kela/env/bin/gunicorn --config deploy/gunicorn.conf run:app

会报错： File "/home/kela/env/bin/pip", line 7, in <module>

from pip.\_internal import main

File "/home/kela/env/lib/python2.7/site-packages/pip/\_internal/\_\_init\_\_.py", line 2, in <module>

from \_\_future\_\_ import absolute\_import

ImportError: No module named \_\_future\_\_

然后我安装future: pip install future

再次启动，依然报错（和上面一样）

--------------------------------------------------

不按照上午说的方式启动：

cd /home/kela/server

python /home/kela/env/bin/gunicorn --config deploy/gunicorn.conf run:app

(这里用的是系统自带的python，能够启动成功)

## order\_dict模块缺失

不用env里的任何东西，用系统的python，

会报错：order\_dict模块找不到，卸载urllib3，安装urllib3的1.2.1版本

解决：先卸载urllib3，安装时不指定版本报错的时候有提示，指定版本用 pip install urllib3==1.21.1）

## no module named pre.steps

但是在执行:python run.py nanmatou 20181122 20181122 all

的时候报错：Traceback (most recent call last):

File "run.py", line 6, in <module>

from pre.steps import stat\_mote\_hour, stat\_portal, logs, motes\_gws\_hour, stat\_gw\_hour, motes\_illegal, stat\_app\_hour

ImportError: No module named pre.steps

但是这个问题可以解决：

把/pre/run.py拷贝到上一级目录（改个名字，防止覆盖上级目录的run.py）

## connection reset by peer

时间跨度太大，预处理程序报错

2018-12-11 19:33:28,592 ERROR Failed to open transport (tries\_left=3)

。。。。。。

error: [Errno 104] Connection reset by peer

。。。。。。。

HiveServer2Error: Invalid query handle: 734c6d982980e84c:fe962f9e00000000

解决办法：每4天分一批，多批次次执行