# CDH集群组件kerberos认证方法

## 一、组件认证方法

### Spark

Spark任务kerberos认证有方式：命令行参数

* 命令行参数

--keytab KEYTAB：提交任务机器上keytab文件的绝对路径

--principal PRINCIPAL：用来登录KDC的Principal

如：

spark-submit

--master yarn

--deploy-mode cluster

--principal airflow@HADOOP.COM

--keytab /root/airflow.keytab

--class cn.tfinfo.Main kerberos-test.jar

在spark内使用jdbc或者其他方式连接集群组件，不用再进行重新认证

### **Livy**

Livy访问还是通过http正常请求即可，livy服务部署的时候内置执行用户为livy，所以执行代码时也会采用livy，这里就需要注意livy用户访问权限

### **Impala**

Impala JDBC URL与原URL有所变化，新增了一些参数，若部署了HAProxy连接上对应的host换成ha的节点即可

一般CDH开启生成服务的principal，如：impala/datapla2@HADOOP.COM

Impala JDBC URL如：

jdbc:impala://datapla2:21050/default;AuthMech=1;KrbRealm=HADOOP.COM;KrbHostFQDN=datapla2;KrbServiceName=impala

AuthMech：值可为1、2、3，1表示kerberos，2表示用户名密码，3不认证

KrbHostFQDN：Impala服务节点地址

KrbServiceName：Impala服务认证名，一般为impala

KrbRealm：域名

* Impala连接池方式

需要自建datasource，不再使用spring封装好的创建方式

详情见

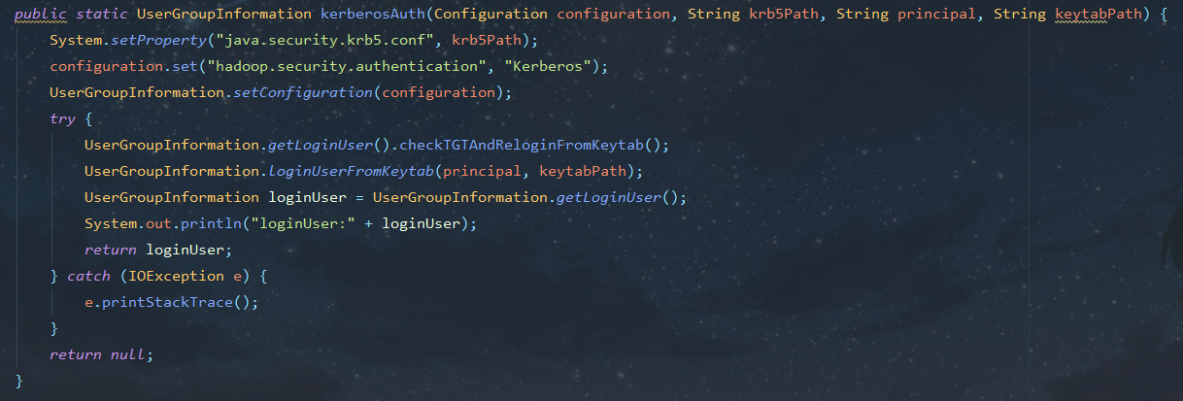
* Impala单连接方式

详情见

### **Hdfs**

Hdfs连接时可以不使用loginUser调用doAs方法

认证方法：



主方法：



## kerberos票据刷新机制说明

### kinit输密码方式

每次调用集群相关命令或者连接前执行下kinit user然后输入密码，一次输入在TGT有效期内就不用再执行kinit user了

### **Kinit用keytab方式**

每次调用集群相关命令或者连接前执行下kinit -kt /path/to/user.keytab user，一次输入在TGT有效期内就不用再执行上述命令了

### **TGT过期说明**

上述说的两种方式都会在登录的时候产生一个TGT，这个TGT有效期默认是24H，在命令行中执行kinit命令时，如果过了这个时间就需要再执行一次kinit命令，应用中会自动刷新。

记得调用下面这个代码!



### **Keytab过期**

Kerberos认证会定期使得keytab文件失效，当文件失效时就需要重新生成，一般默认时7天，我们可以采用定时任务定时拉取生成，对于长进程应用需要定期重新用keytab文件创建相应的连接池bean或者连接bean。