Kafka数据落地HDFS指导

# 落地前准备

## 检查要落地的Topic在Kafka中是否已经接入

可以联系运维确认是否数据正常接入。

## 确保落地flume使用的机机认证的账号有落地的HDFS路径的读写权限

# 部署Flume落地HDFS

## 部署Flume的机器

可选择机器必须满足：1.与kafka集群已经落地的FI集群网络互通。

## Flume部署

请参考 <<Flume安装和配置说明.docx>>



## Flume相关配置

（1）说明： $client\_dir 表示客户端安装路径

1. FI 管理台，新建机机账号（比如hdfstest），用户组包含hdfs
2. 下载hdfstest用户的kerberos证书文件，并上传至”flume安装目录/conf”下面；
3. 下载Kerberos认证

从集群内装有kerberos的节点下载kdc.conf  （/opt/huawei/Bigdata/FusionInsight/etc/\*\_\*\_KerberosClient/），并上传至”flume安装目录/conf”

1. 下载Zookeeper认证

如果集群中装有flume server，从/opt/huawei/Bigdata/FusionInsight/etc/\*\_\*\_Flume/目录下获取jaas.conf，将keytab修改为第2步获取到的keytab（使用绝对路径路径），principal修改为对应的用户名,并上传至”flume安装目录/conf”

1. 下载core-site.xml、hdfs-site.xml、hbase-site.xml

如果集群中装有flume server，从/opt/huawei/Bigdata/FusionInsight/etc/\*\_\*\_Flume/目录下获取这三个文件，并上传至客户端所在节点安装目录至”flume安装目录/conf

1. 客户端文件夹下修改flume-env.sh，JAVA\_OPTS添加（仅供参考，需要修改为文件实际路径）

-Djava.security.krb5.conf=$client\_dir/fusioninsight-flume-1.6.0/conf/kdc.conf -Djava.security.auth.login.config=$client\_dir/fusioninsight-flume-1.6.0/conf/jaas.conf -Dzookeeper.server.principal=zookeeper/hadoop.hadoop.com -Dzookeeper.request.timeout=120000

1. 修改hosts，将FI安装规划中所有host信息添加到客户端节点

vi /etc/hosts

1. properties.properties样例

请见跳板机 D:\liyue\properties.properties

1. 重启服务，测试是否可以连接集群HDFS