

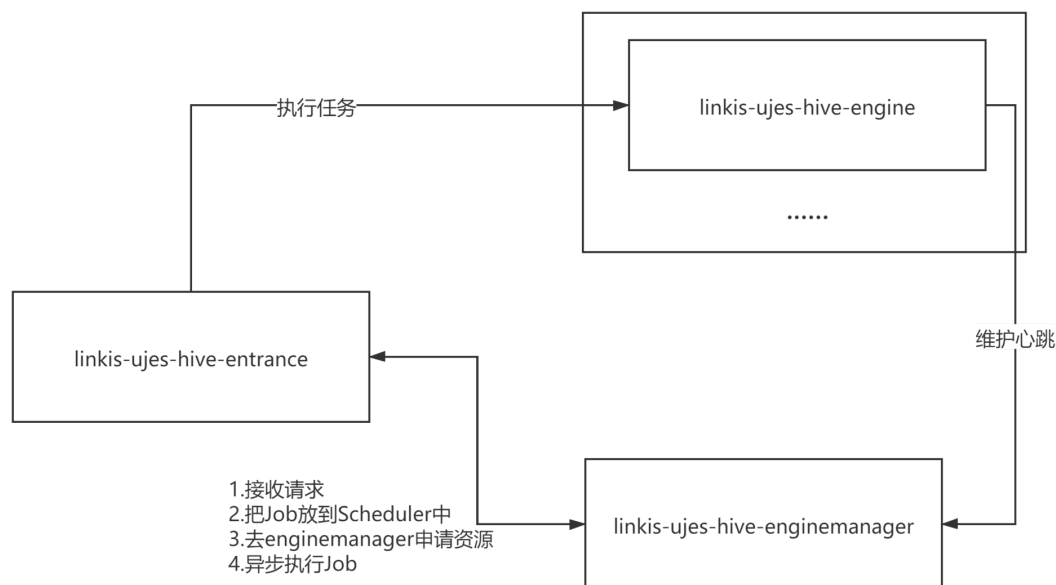
Linkis环境下Hive on Tez

一.简介

1.1 背景

Linkis是为了解决计算治理的中间件，用Hive常常会用到tez引擎，本文介绍如何基于Linkis打通Hive on Tez 。

1.2 架构



Linkis的Hive引擎执行任务流程

HiveEngine中用到org.apache.hadoop.hive.ql.Driver 去调用Hql，这里面涉及到Hive的执行流程。

- 由编译器解析查询语句并从AST 中生成一个Calcite逻辑计划。
- 优化逻辑计划。
- 优化逻辑计划转换物理计划。
- 向量化的执行计划。
- 生成具体的task，可以是mr或者Spark、Tez，并通过Driver提交任务到YARN。
- 执行结束后将结果返回给用户。

从执行逻辑角度看，基于Linkis去操作Hive on Tez不需要更改源码，之所以不能用Tez因为Hive引擎中缺失依赖，所以无法去执行。

二.部署

以CDH-5.16.2环境为例，假设大数据集群已经配置好Tez。

1.linkis-ujes-hive-enginemanager目录下，lib文件夹，添加T ez依赖。

把/opt/cloudera/parcels/CDH/lib/tez 目录下，tez-*前缀的jar都放到linkis-ujes-hive-enginemanager/lib目录下。

```
-rwxr-xr-x 1 hdfs hdfs 986163 7月 23 18:58 tez-api-0.8.5.jar
-rwxr-xr-x 1 hdfs hdfs 76816 7月 23 18:58 tez-common-0.8.5.jar
-rwxr-xr-x 1 hdfs hdfs 1324875 7月 23 18:58 tez-dag-0.8.5.jar
-rwxr-xr-x 1 hdfs hdfs 49583 7月 23 18:58 tez-examples-0.8.5.jar
-rwxr-xr-x 1 hdfs hdfs 77974 7月 23 18:58 tez-history-parser-0.8.5.jar
-rwxr-xr-x 1 hdfs hdfs 15261 7月 23 18:58 tez-javadoc-tools-0.8.5.jar
-rwxr-xr-x 1 hdfs hdfs 72845 7月 23 18:58 tez-job-analyzer-0.8.5.jar
-rwxr-xr-x 1 hdfs hdfs 290968 7月 23 18:58 tez-mapreduce-0.8.5.jar
-rwxr-xr-x 1 hdfs hdfs 196359 7月 23 18:58 tez-runtime-internals-0.8.5.jar
-rwxr-xr-x 1 hdfs hdfs 592962 7月 23 18:58 tez-runtime-library-0.8.5.jar
-rwxr-xr-x 1 hdfs hdfs 26332 7月 23 18:58 tez-yarn-timeline-history-0.8.5.jar
-rwxr-xr-x 1 hdfs hdfs 7742 7月 23 18:58 tez-yarn-timeline-history-with-acls-0.8.5.jar
```

2.CDH Hive配置tez依赖。

hive-site.xml 的 HiveServer2 高可用代码段 (安全码) 以 XML 格式查看

名称	tez.lib.uris
值	\$(fs.defaultFS)/apps/tez-0.8.5.tar.gz
说明	tez库路径
<input type="checkbox"/> 最终	
名称	hive.tez.container.size
值	3020
说明	
<input type="checkbox"/> 最终	

3.因为CDH配置的变量存储在元数据库中，所以把/etc/hive/conf配置独立出来，在linkis-ujes-hive-enginemanager conf文件夹下，把linkis-engine.properties和linkis.properties的hive.config.dir都配置独立出来的Hive相关的配置文件。

独立出来Hive配置文件，Hive-site.xml中追加属性（重点）：

```
1 <property>
2   <name>tez.lib.uris</name>
3   <value>hdfs:///apps/tez/tez-0.8.5.tar.gz</value>
4 </property>
5
6 <property>
7   <name>hive.tez.container.size</name>
8   <value>10240</value>
9 </property>
```

linkis-ujes-hive-enginemanager/conf/linkis.properties 内容：

```
1 wds.linkis.server.restful.scan.packages=com.webank.wedatasphere.linkis.entrance.restful
2 wds.linkis.engine.application.name=hiveEngine
3 wds.linkis.server.component.exclude.packages=com.webank.wedatasphere.linkis.engine.,com.webank.wedatasphere.linkis.udf.
4 wds.linkis.server.version=v1
5 #sudo script
6 wds.linkis.enginemanager.sudo.script=/home/hdfs/linkis/linkis-ujes-hive-enginemanager/bin/rootScript.sh
7 #hadoop config
8 hadoop.config.dir=/etc/hadoop/conf
9 #hive config
10 hive.config.dir=/home/hdfs/etc/hive
```

linkis-ujes-hive-enginemanager/conf/linkis-engine.properties 内容：

```
1 wds.linkis.server.restful.scan.packages=com.webank.wedatasphere.linkis.engine.restful
2 wds.linkis.engine.application.name=hiveEngine
3 wds.linkis.server.component.exclude.packages=com.webank.wedatasphere.linkis.enginemanager.,com.webank.wedatasphere.linkis.udf.
4 wds.linkis.server.component.exclude.classes=com.webank.wedatasphere.linkis.resourcemanager.service.annotation.RMAnnotationParser
5 wds.linkis.server.version=v1
6 #hadoop config
7 hadoop.config.dir=/etc/hadoop/conf
8 #hive config
10 hive.config.dir=/home/hdfs/etc/hive
```

/home/hdfs/etc/hive文件夹内容：

```
-rw-r--r-- 1 hdfs hdfs 20 7月 23 15:43 __cloudera_generation__
-rw-r--r-- 1 hdfs hdfs 66 7月 23 15:43 __cloudera_metadata__
-rw-r--r-- 1 hdfs hdfs 3852 7月 23 15:43 core-site.xml
-rw-r--r-- 1 hdfs hdfs 617 7月 23 15:43 hadoop-env.sh
-rw-r--r-- 1 hdfs hdfs 1721 7月 23 15:43 hdfs-site.xml
-rw-r--r-- 1 hdfs hdfs 3017 7月 23 15:43 hive-env.sh
-rw-r--r-- 1 hdfs hdfs 5617 7月 23 15:43 hive-site.xml
-rw-r--r-- 1 hdfs hdfs 310 7月 23 15:43 log4j.properties
-rw-r--r-- 1 hdfs hdfs 5087 7月 23 15:43 mapred-site.xml
-rw-r--r-- 1 hdfs hdfs 0 7月 23 15:43 redaction-rules.json
-rw-r--r-- 1 hdfs hdfs 315 7月 23 15:43 ssl-client.xml
-rw-r--r-- 1 hdfs hdfs 569 7月 23 15:43 topology.map
-rwxr-xr-x 1 hdfs hdfs 1594 7月 23 15:43 topology.py
-rw-r--r-- 1 hdfs hdfs 3553 7月 23 15:43 yarn-site.xml
```

三.作者介绍

李孟，目前就职于知因智慧数据科技有限公司，负责数据中台数据引擎基础架构设计和中间件开发，专注云计算大数据方向。