

R语言操作hive, 并调度

2种方法连接hive

方法一 用jdbc连接

示例:

```
Sys.setenv(JAVA_HOME="C:/Program Files/Java/jdk1.8.0_121")
library('rJava')
library('DBI')
library('RJDBC')
setwd("D:\\work\\R\\RHive")
drv <- JDBC("org.apache.hive.jdbc.HiveDriver", "hive-jdbc-standalone.jar;hadoop-common.jar")
conn <-
dbConnect(drv, "jdbc:hive2://10.0.19.48:10000/default", "yarn", "***")
dbGetQuery(conn, "SELECT channel_id, create_time, deal_id FROM baofoo_cm.cm_deal limit 10")
```

方法二, 用odbc连接 (适用于windows)

先在cloudera官网下载odbc驱动, 并设置好账号和密码 如图:



示例：

```
library(RODBC)
ch <- odbcConnect("Sample Cloudera Hive DSN")
res <- sqlQuery(ch, paste("SELECT channel_id, create_time, deal_id
FROM baofoo_cm.cm_deal limit 10"))
```

shell 调用R脚本做定时计划

带参数脚本 myscript.R

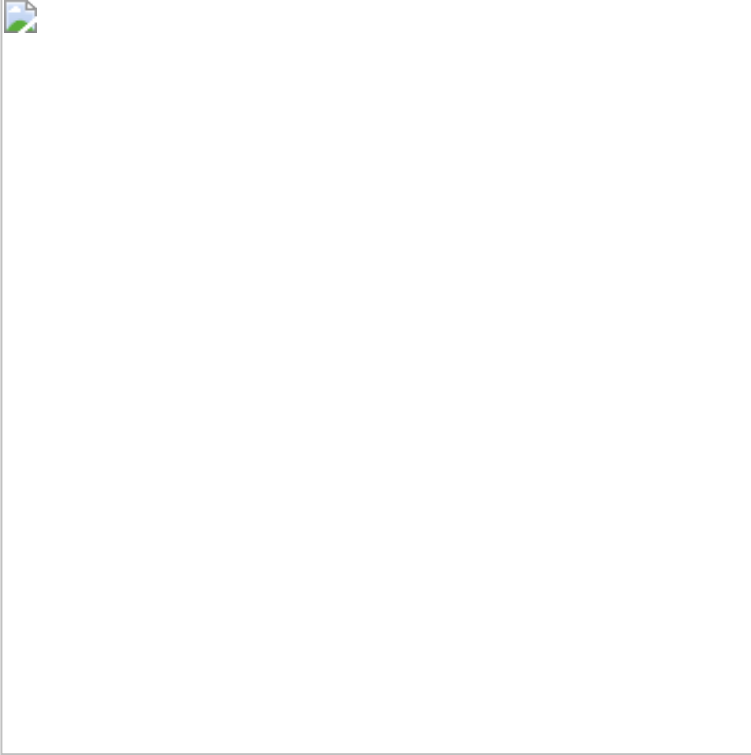
```
args <- commandArgs(TRUE)
```

```
args[1]
```

```
args[2]
```

```
$ Rscript myscript.R
```

结果:



使用crontab定时执行myscript.R

```
0 4 * * * Rscript /home/yz/myscript.R
```