

1. 下载

<https://www.knime.com/knime-extension-for-apache-spark>

选择对应的发行版本。

github地址：<https://github.com/spark-jobserver/spark-jobserver>

本文档按照spark-job-server-0.6.2.3-KNIME_cdh-5.7.tar.gz整理，操作系统RHEL6或者RHEL7.

2. 安装

```
root@host$ useradd -d /opt/${LINKNAME}/ -M -r -s /bin/false sparkjob-server
root@host$ su -l -c "hdfs dfs -mkdir -p /user/spark-job-server ; hdfs dfs -chown -R spark-job-server /user/spark-job-server" hdfs
root@host$ cp /path/to/spark-job-server-xxx.tar.gz /opt
root@host$ cd /opt
root@host$ tar xzf spark-job-server-xxx.tar.gz
root@host$ ln -s spark-job-server-xxx ${LINKNAME}
root@host$ chown -R spark-job-server:spark-job-server ${LINKNAME} spark-job-server-xxx/
```

设置开机启动：

(1)、RHEL6

```
root@host$ ln -s /opt/${LINKNAME}/spark-job-server-init.d \ /etc/init.d/${LINKNAME}
root@host$ chkconfig --levels 2345 ${LINKNAME} on
```

(2)、RHEL7

```
root@host$ ln -s /opt/${LINKNAME}/spark-job-server-init.d /etc/init.d/${LINKNAME}
root@host$ systemctl daemon-reload
root@host$ systemctl enable ${LINKNAME}
```

3. 修改配置文件

编辑 environment.conf，修改任务提交方式为yarn：

```
master = yarn-client
```

4. 修改Setting.sh配置文件

如果SPARK_HOME没有安装在给定的路径下，可以设置。

LOG_DIR设置日志路径，可以不设置，默认路径：/var/log/spark-job-server/

5. 启动服务

启动服务会有各种报错，查看日志即可。

下面的错误就是目录权限不对导致的。

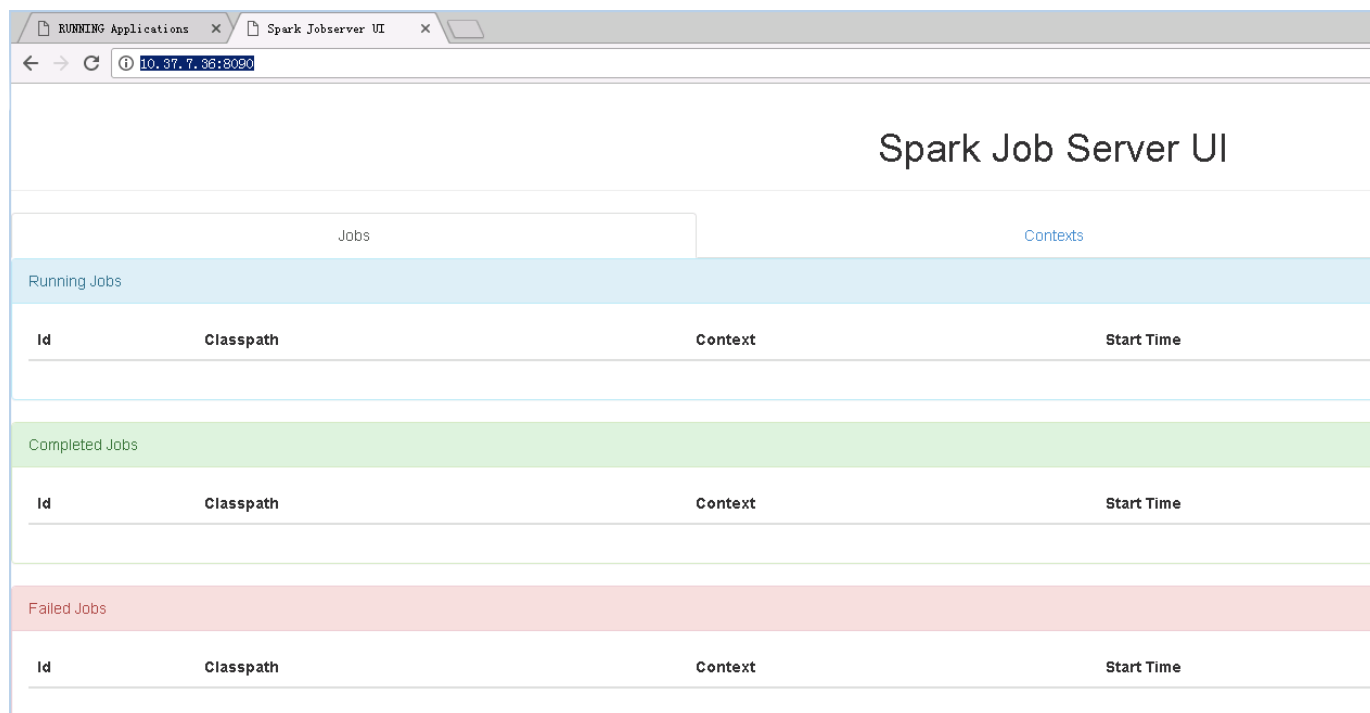
chmod 777

chown sparkjob-server: sparkjob-server ..

```
18-04-09 00:15:56,675] INFO InternalCommand.Migrate [] [] Schema Update is up to date, no migration necessary.
18-04-09 00:15:56,675] INFO spark.jobserver.io.JobSqlDAO [] [] - Running dummy query to initialize dbcp.
18-04-09 00:15:57,901] INFO k.jobserver.io.JobDAOActor [] [akka://JobServer/user/dao-manager] - Starting actor spark.jobserver
18-04-09 00:15:57,901] ERROR spark.jobserver.JobServer$ [] [] - Unable to start Spark JobServer:
java.io.FileNotFoundException: /tmp/spark-job-server/upload/files.data (权限不够)
    at java.io.FileOutputStream.open(Native Method)
    at java.io.FileOutputStream.<init>(FileOutputStream.java:212)
    at spark.jobserver.io.DataFileDAO.<init>(DataFileDAO.scala:55)
    at spark.jobserver.JobServer$.start(JobServer.scala:61)
    at spark.jobserver.JobServer$.main(JobServer.scala:130)
    at spark.jobserver.JobServer.main(JobServer.scala)
    at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
    at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:57)
    at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
    at java.lang.reflect.Method.invoke(Method.java:601)
    at org.apache.spark.deploy.SparkSubmit$.org$apache$spark$deploy$SparkSubmit$$runMain(SparkSubmit.scala:731)
    at org.apache.spark.deploy.SparkSubmit$.doRunMain$1(SparkSubmit.scala:181)
    at org.apache.spark.deploy.SparkSubmit$.submit(SparkSubmit.scala:206)
    at org.apache.spark.deploy.SparkSubmit$.main(SparkSubmit.scala:121)
    at org.apache.spark.deploy.SparkSubmit.main(SparkSubmit.scala)
18-04-09 00:15:57,978] ERROR ka.actor.OneForOneStrategy [] [akka://JobServer/user/dao-manager] - exception during creation
ka.actor.ActorInitializationException: exception during creation
    at akka.actor.ActorInitializationException$.apply(Actor.scala:218)
    at akka.actor.ActorCell.create(ActorCell.scala:578)
    at akka.actor.ActorCell.<init>(ActorCell.scala:495)
```

查看服务：

<http://localhost:8090/>



创建一个上下文：

```
curl -d "" 'localhost:8090/contexts/test?num-cpu-cores=6&mem-per-node=2G&spark.executor.instances=4'
```

上传jar包：

```
curl --data-binary @jobexample1.jar localhost:8090/jars/jobexample1
```

执行：

```
curl -d "input.strings=a b c a b see" "localhost:8090/jobs?appName=jobexample1&classPath=test.jobserver.WordCount&context=test"
```

	Spark job server	Zeppelin	Livy
是否支持 jar 包提交	支持	不支持	支持
是否支持代码段提交	不支持	支持	支持
是否支持 SparkContext 重用与管理	支持	支持重用，不能 stop SparkContext	支持
运行模式	Client	Client	Client/Cluster
多个 SparkContext 运行	在同一个 JVM，可能存在问题	运行在不同 JVM	运行在不同 JVM
接口	RestFul	RestFul/WebSocket	RestFul
SQL 是否支持	不支持	支持	不支持