Hadoop+Spark+ iObjects for Spark

部署流程 v0.2

一、 准备

1.1 软件版本

- Ubuntu 14.04.4
- JDK 1.7.0
- Scala 2.10.6
- Hadoop 2.6.5
- Spark 1.6.3
- SuperMap iObjects Java 8.1.1
- SuperMap iObjects for Spark 8.1.1_60034

1.2 SSH 设置

修改/etc/ssh/sshd_config 文件，将以下三项开启 yes 状态

```
PermitRootLogin yes
PermitEmptyPasswords yes
PasswordAuthentication yes
```

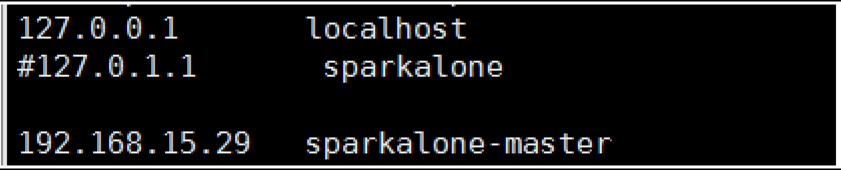
重启 ssh 服务

```
service ssh restart
```

这样 root 用户可直接登陆，以及为后续 ssh 无密码登录做准备。

1.3 绑定 IP 和修改计算机名

1.3.1 修改/etc/hosts,添加 IP 绑定注释 127.0.1.1 绑定（不注释会影响 hadoop 集群）



```
127.0.0.1       localhost
#127.0.1.1       sparkalone

192.168.15.29   sparkalone-master
```

1.3.2 修改/etc/hostname,为绑定计算机名。(计算机名和上面 hosts 绑定名必须一致)

1.4 SSH 无密码登陆

1、cd ~/.ssh #进到当前用户的隐藏目录（.ssh）

2、ssh-keygen -t rsa #用 rsa 生成密钥，一路回车

3、cp id_rsa.pub authorized_keys #把公钥复制一份，并改名为 authorized_keys，这步执行完后，在当前机器执行 ssh localhost 可以无密码登录本机了

*初次执行会提示确认，输入 yes 和登陆密码，之后就没提示了。

*后续如果做集群，可以 scp authorized_keys root@第二台机器名：/root/.ssh （把重命名后的公钥通过 ssh 提供的远程复制文件复制到从机）这样子节点启动也不需要再次输入密码。

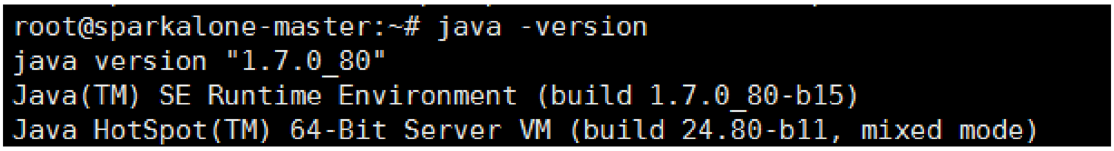
1.5 JDK 安装

下载：jdk-7u80-linux-x64.tar.gz 包，放到/opt 下解压

1.5.1 将 JDK 环境变量配置到/etc/profile 中

```
export JAVA_HOME=/opt/jdk1.7.0_80
export JRE_HOME=/opt/jdk1.7.0_80/jre
export CLASSPATH=$JAVA_HOME/lib:$JRE_HOME/lib
export PATH=$JAVA_HOME/bin: $PATH
```

1.5.2 检查 JDK 是否配置好



```
root@sparkalone-master:~# java -version
java version "1.7.0_80"
Java(TM) SE Runtime Environment (build 1.7.0_80-b15)
Java HotSpot(TM) 64-Bit Server VM (build 24.80-b11, mixed mode)
```

二、 Hadoop 部署

2.1 Hadoop 安装

1. 下载 hadoop2.6.5（hadoop-2.6.5.tar.gz）
2. 解压 tar -zxvf hadoop-2.6.5.tar.gz，并在主目录下创建 tmp、dfs、dfs/name、dfs/node、dfs/data

```
root@sparkalone-master:/opt/hadoop-2.6.5# mkdir tmp
root@sparkalone-master:/opt/hadoop-2.6.5# mkdir dfs
```

```
root@sparkalone-master:/opt/hadoop-2.6.5# mkdir dfs/name
root@sparkalone-master:/opt/hadoop-2.6.5# mkdir dfs/node
root@sparkalone-master:/opt/hadoop-2.6.5# mkdir dfs/data
```

2.2 Hadoop 配置

以下操作都在 hadoop-2.6.5/etc/hadoop 下进行

2.2.1 编辑 hadoop-env.sh 文件，修改 JAVA_HOME 配置项为 JDK 安装目录

```
export JAVA_HOME=/opt/jdk1.7.0_80
```

2.2.2 编辑 core-site.xml 文件，添加以下内容，

其中 **sparkalone-master** 为计算机名，**/opt/hadoop-2.6.5/tmp** 为手动创建的目录

```
<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://sparkalone-master:9000</value>
  </property>
  <property>
    <name>io.file.buffer.size</name>
    <value>131072</value>
  </property>
  <property>
    <name>hadoop.tmp.dir</name>
    <value>file:/opt/hadoop-2.6.5/tmp</value>
    <description>Abasefor other temporary directories.</description>
  </property>
  <property>
    <name>hadoop.proxyuser.spark.hosts</name>
    <value>*</value>
  </property>
<property>
    <name>hadoop.proxyuser.spark.groups</name>
    <value>*</value>
  </property>
</configuration>
```

2.2.3 编辑 hdfs-site.xml 文件，添加以下内容

其中 **sparkalone-master** 为计算机名，
file:/opt/hadoop-2.6.5/dfs/name 和 **file:/opt/hadoop-2.6.5/dfs/data** 为手动创建目录

```
<configuration>
<property>
  <name>dfs.namenode.secondary.http-address</name>
  <value>sparkalone-master:9001</value>
 </property>
  <property>
   <name>dfs.namenode.name.dir</name>
   <value>file:/opt/hadoop-2.6.5/dfs/name</value>
 </property>
 <property>
  <name>dfs.datanode.data.dir</name>
  <value>file:/opt/hadoop-2.6.5/dfs/data</value>
  </property>
 <property>
  <name>dfs.replication</name>
  <value>3</value>
 </property>
 <property>
  <name>dfs.webhdfs.enabled</name>
  <value>true</value>
 </property>
</configuration>
```

复制 mapred-site.xml.template 并重命名为 mapred-site.xml

```
cp mapred-site.xml.template mapred-site.xml
```

2.2.4 编辑 mapred-site.xml 文件，添加以下内容

其中 **sparkalone-master** 为计算机名

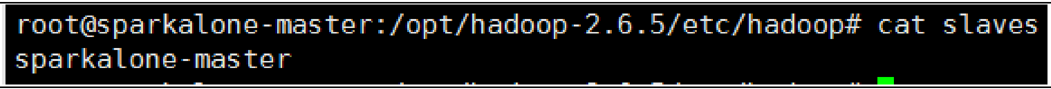
```
<configuration>
<property>
   <name>mapreduce.framework.name</name>
   <value>yarn</value>
 </property>
 <property>
  <name>mapreduce.jobhistory.address</name>
  <value>sparkalone-master:10020</value>
 </property>
 <property>
  <name>mapreduce.jobhistory.webapp.address</name>
  <value>sparkalone-master:19888</value>
 </property>
</configuration>
```

2.2.5 编辑 yarn-site.xml 文件，添加以下内容

其中 **sparkalone-master** 为计算机名

```
<configuration>
<!-- Site specific YARN configuration properties -->
    <property>
      <name>yarn.nodemanager.aux-services</name>
      <value>mapreduce_shuffle</value>
    </property>
    <property>
      <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
      <value>org.apache.hadoop.mapred.ShuffleHandler</value>
    </property>
    <property>
      <name>yarn.resourcemanager.address</name>
      <value>sparkalone-master:8032</value>
    </property>
    <property>
      <name>yarn.resourcemanager.scheduler.address</name>
      <value>sparkalone-master:8030</value>
    </property>
    <property>
      <name>yarn.resourcemanager.resource-tracker.address</name>
      <value>sparkalone-master:8035</value>
    </property>
    <property>
      <name>yarn.resourcemanager.admin.address</name>
      <value>sparkalone-master:8033</value>
    </property>
    <property>
      <name>yarn.resourcemanager.webapp.address</name>
      <value>sparkalone-master:8088</value>
    </property>
</configuration>
```

2.2.6 修改 slaves 文件，添加本机作为集群节点



```
root@sparkalone-master:/opt/hadoop-2.6.5/etc/hadoop# cat slaves
sparkalone-master
```

注意：

如果 hadoop 配置集群，可以将配置文件 etc/hadoop 下内容同步到其他机器上，并修改 slaves 文件

```
scp -r hadoop root@另一台机器名:/opt /hadoop-2.6.5/etc
```

2.3 Hadoop 启动

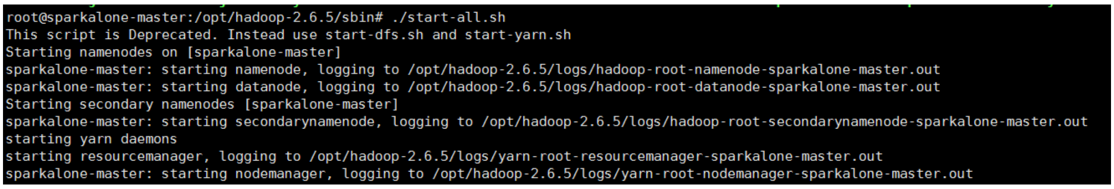
1. 格式化一个新的文件系统，进入到 hadoop-2.6.5/bin 下执行：

```
./hadoop namenode –format
```

2. **启动 hadoop**，进入到 hadoop-2.6.5/sbin 下执行：

```
./start-all.sh
```

看到如下内容说明启动成功

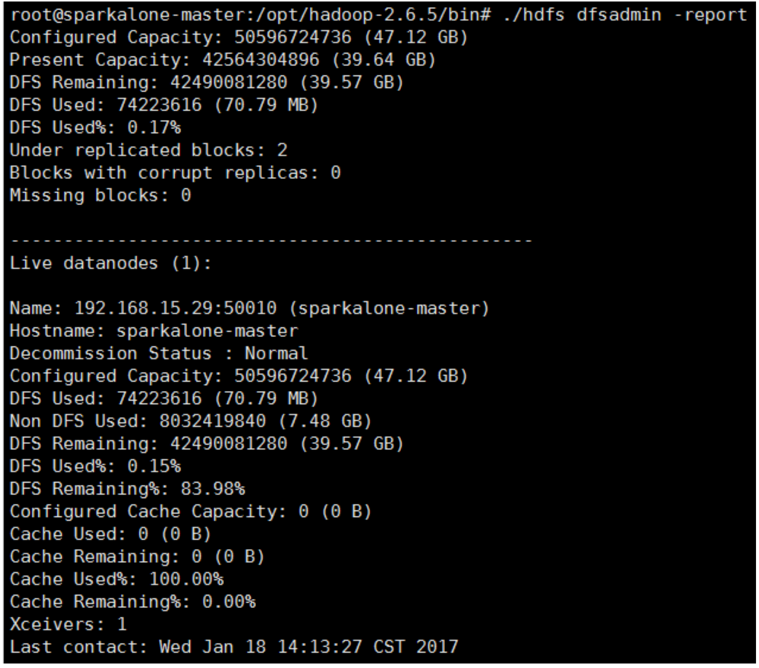


```
root@sparkalone-master:/opt/hadoop-2.6.5/sbin# ./start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
Starting namenodes on [sparkalone-master]
sparkalone-master: starting namenode, logging to /opt/hadoop-2.6.5/logs/hadoop-root-namenode-sparkalone-master.out
sparkalone-master: starting datanode, logging to /opt/hadoop-2.6.5/logs/hadoop-root-datanode-sparkalone-master.out
Starting secondary namenodes [sparkalone-master]
sparkalone-master: starting secondarynamenode, logging to /opt/hadoop-2.6.5/logs/hadoop-root-secondarynamenode-sparkalone-master.out
starting yarn daemons
starting resourcemanager, logging to /opt/hadoop-2.6.5/logs/yarn-root-resourcemanager-sparkalone-master.out
sparkalone-master: starting nodemanager, logging to /opt/hadoop-2.6.5/logs/yarn-root-nodemanager-sparkalone-master.out
```

2.4 Hadoop 集群检查

检查 hadoop 集群，进入 hadoop-2.6.5/bin 下执行

```
./hdfs dfsadmin -report
```

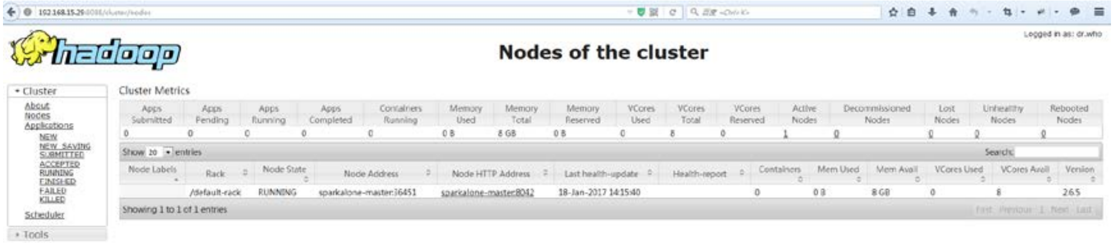


```
root@sparkalone-master:/opt/hadoop-2.6.5/bin# ./hdfs dfsadmin -report
Configured Capacity: 50596724736 (47.12 GB)
Present Capacity: 42564304896 (39.64 GB)
DFS Remaining: 42490081280 (39.57 GB)
DFS Used: 74223616 (70.79 MB)
DFS Used%: 0.17%
Under replicated blocks: 2
Blocks with corrupt replicas: 0
Missing blocks: 0

-------------------------------------------------
Live datanodes (1):

Name: 192.168.15.29:50010 (sparkalone-master)
Hostname: sparkalone-master
Decommission Status : Normal
Configured Capacity: 50596724736 (47.12 GB)
DFS Used: 74223616 (70.79 MB)
Non DFS Used: 8032419840 (7.48 GB)
DFS Remaining: 42490081280 (39.57 GB)
DFS Used%: 0.15%
DFS Remaining%: 83.98%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xceivers: 1
Last contact: Wed Jan 18 14:13:27 CST 2017
```

或者访问 http://192.168.15.29:8088/cluster/nodes



Nodes of the cluster

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved	Active Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes
0	0	0	0	0	0 B	8 GB	0 B	0	8	0	1	0	0	0	0

Show 20 entries Search:

Node Labels	Rack	Node State	Node Address	Node HTTP Address	Last health-update	Health-report	Containers	Mem Used	Mem Avail	VCores Used	VCores Avail	Version
	/default-rack	RUNNING	sparkalone-master:36451	sparkalone-master:8042	18-Jan-2017 14:15:40		0	0 B	8 GB	0	8	2.6.5

Showing 1 to 1 of 1 entries

三、 Spark 部署

3.1 Spark 安装

1. 下载 scala-2.10.6.tgz 和 spark-1.6.3-bin-hadoop2.6.tgz，放到 opt 下解压。
2. 将 scala 和 spark 环境变量配置到/etc/profile 中

```
export SCALA_HOME=/opt/scala-2.10.6
export SPARK_HOME=/opt/spark-1.6.3-bin-hadoop2.6
export PATH=$SCALA_HOME/bin:$JAVA_HOME/bin:$SPARK_HOME/bin:$PATH
```

3.2 Spark 配置

1. 进入 spark-1.6.3-bin-hadoop2.6/conf

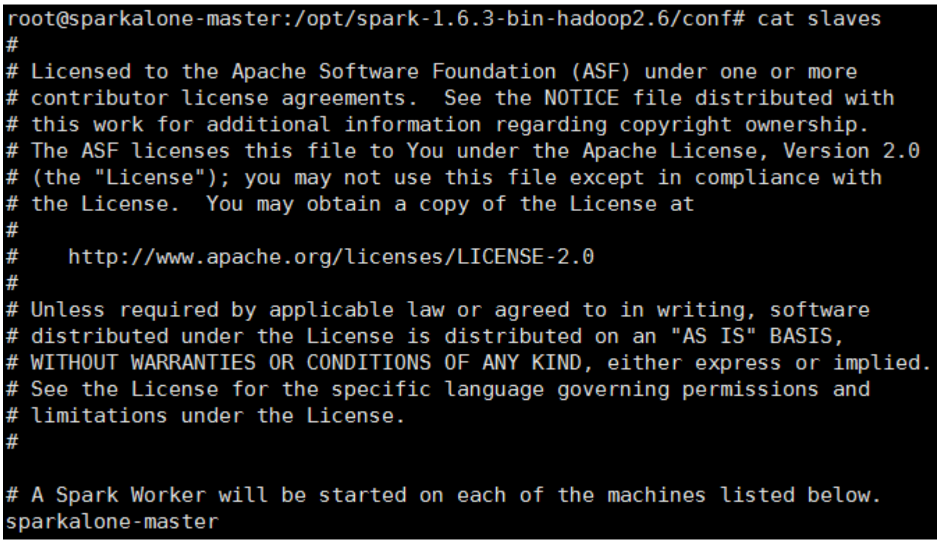
复制 spark-env.sh.template 并重命名为 spark-env.sh

```
cp spark-env.sh.template spark-env.sh
```

编辑 spark-env.sh 文件，添加以下内容

```
export JAVA_HOME=/opt/jdk1.7.0_80
export SPARK_MASTER_IP=192.168.15.29
export SPARK_WORKER_MEMORY=4g
export SCALA_HOME=/opt/scala-2.10.6
export HADOOP_HOME=/opt/hadoop-2.6.5/
export HADOOP_CONF_DIR=/opt/hadoop-2.6.5/etc/hadoop
export LD_LIBRARY_PATH=$LD_LIBRARY_PATH:$SUPERMAP_OBJ
export SPARK_CLASSPATH=$SPARK_CLASSPATH
```

2.编辑 slaves 文件，增加当前计算机名



```
root@sparkalone-master:/opt/spark-1.6.3-bin-hadoop2.6/conf# cat slaves
#
# Licensed to the Apache Software Foundation (ASF) under one or more
# contributor license agreements.  See the NOTICE file distributed with
# this work for additional information regarding copyright ownership.
# The ASF licenses this file to You under the Apache License, Version 2.0
# (the "License"); you may not use this file except in compliance with
# the License.  You may obtain a copy of the License at
#
#    http://www.apache.org/licenses/LICENSE-2.0
#
# Unless required by applicable law or agreed to in writing, software
# distributed under the License is distributed on an "AS IS" BASIS,
# WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
# See the License for the specific language governing permissions and
# limitations under the License.
#

# A Spark Worker will be started on each of the machines listed below.
sparkalone-master
```

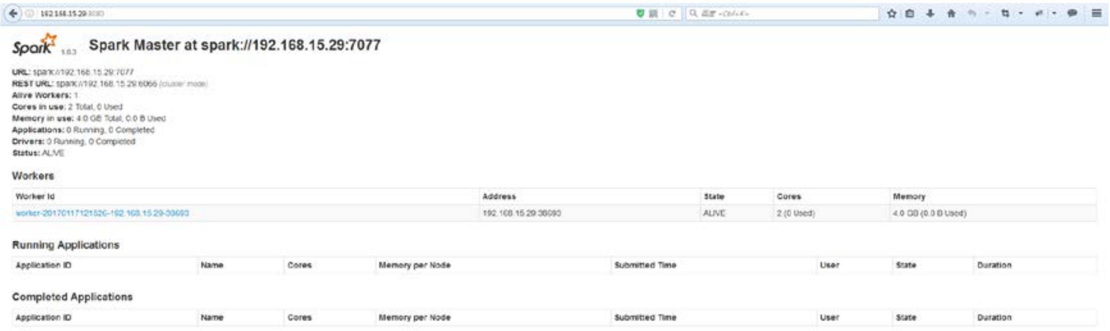
3.3　Spark 启动

启动 spark，进入 spark-1.6.3-bin-hadoop2.6/sbin 下执行

```
./start-all.sh
```

3.4　Spark 集群检查

访问 http://192.168.15.29:8080/



Spark Master at spark://192.168.15.29:7077

URL: spark://192.168.15.29:7077
REST URL: spark://192.168.15.29:6066 (cluster mode)
Alive Workers: 1
Cores in use: 2 Total, 0 Used
Memory in use: 4.0 GB Total, 0.0 B Used
Applications: 0 Running, 0 Completed
Drivers: 0 Running, 0 Completed
Status: ALIVE

Workers

Worker Id	Address	State	Cores	Memory
worker-20170117121526-192.168.15.29-38093	192.168.15.29:38093	ALIVE	2 (0 Used)	4.0 GB (0.0 B Used)

Running Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration

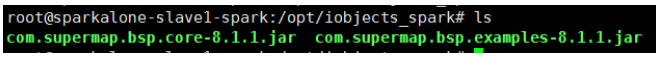
注意：如果将来配置 spark 集群，只需要保证子节点内容和主节点内容一致即可。

四、iObjects for Spark 部署

4.1 iObjects for Spark 安装

1. 下载 supermap_iobjects_for_spark_811_60034.tar.gz 和相应的 iObjects Java 811 组件
2. 放到 opt 下解压

（例如 for spark 中 lib 放到/opt/iobjects_spark，iObjects Java 放到/opt/iobjects_new）



```
root@sparkalone-slave1-spark:/opt/iobjects_spark# ls
com.supermap.bsp.core-8.1.1.jar  com.supermap.bsp.examples-8.1.1.jar
```

4.2 iObjects for Spark 配置

1. 将 for Spark 和 iObjects Java 环境变量配置到/etc/profile 中,结合之前 hadoop 和 spark 配置。总配置如下，其中红色为新增加配置。

```
export JAVA_HOME=/opt/jdk1.7.0_80
export SCALA_HOME=/opt/scala-2.10.6
export SPARK_HOME=/opt/spark-1.6.3-bin-hadoop2.6
export PATH=$SCALA_HOME/bin:$JAVA_HOME/bin:$SPARK_HOME/bin:$PATH
export SUPERMAP_OBJ=/opt/iobjects_new
export LD_LIBRARY_PATH=$LD_LIBRARY_PATH:$SUPERMAP_OBJ
export SPARK_CLASSPATH=$SPARK_CLASSPATH:/opt/iobjects_spark
export JRE_HOME=/opt/jdk1.7.0_80/jre
export CLASSPATH=$JAVA_HOME/lib:$JRE_HOME/lib
```

2. 进入 spark-1.6.3-bin-hadoop2.6/conf

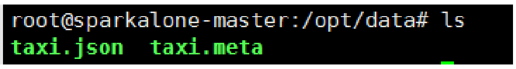
编辑 spark-env.sh 文件，总配置如下，其中红色为新增加配置。

```
export JAVA_HOME=/opt/jdk1.7.0_80
export SPARK_MASTER_IP=192.168.15.29
export SPARK_WORKER_MEMORY=4g
export SCALA_HOME=/opt/scala-2.10.6
export HADOOP_HOME=/opt/hadoop-2.6.5/
export HADOOP_CONF_DIR=/opt/hadoop-2.6.5/etc/hadoop
export SUPERMAP_OBJ=/opt/iobjects_new
export LD_LIBRARY_PATH=$LD_LIBRARY_PATH:$SUPERMAP_OBJ
export SPARK_CLASSPATH=$SPARK_CLASSPATH:/opt/iobjects_spark
```

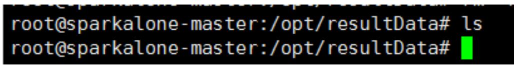
4.3 实例程序验证

我们使用 for Spark 示例来验证是否安装成功。

1. 将产品包中的示例数据放到中/opt/data 文件夹中，并创建/opt/resultData 空文件夹



```
root@sparkalone-master:/opt/data# ls
taxi.json  taxi.meta
```



```
root@sparkalone-master:/opt/resultData# ls
root@sparkalone-master:/opt/resultData# 
```

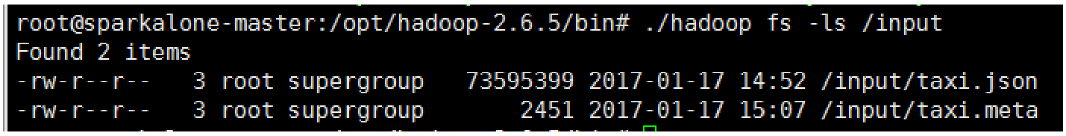
2. 将示例数据导入到 hdfs 中。

启动 hadoop，在 hadoop-2.6.5/bin 中执行

```
./hadoop fs -mkdir /input        #创建/input 目录
./hdfs dfs -put /opt/data/taxi.json /input/     #将 taxi.json 导入到/input 目录中
./hdfs dfs -put /opt/data/taxi.meta /input/   #将 taxi. meta 导入到/input 目录中
```

导入完成后，可以使用如下命名查看

```
./hadoop fs -ls /input
```



```
root@sparkalone-master:/opt/hadoop-2.6.5/bin# ./hadoop fs -ls /input
Found 2 items
-rw-r--r--   3 root supergroup   73595399 2017-01-17 14:52 /input/taxi.json
-rw-r--r--   3 root supergroup       2451 2017-01-17 15:07 /input/taxi.meta
```

3. 启动 Spark，切换到 spark-1.6.3-bin-hadoop2.6/bin 下执行（执行几何对象创建缓冲区）

```
./spark-submit --class com.supermap.bsp.examples.GeometryBufferTest --master
spark://192.168.15.29:7077 --conf spark.default.parallelism=3 --jars
/opt/iobjects_spark/com.supermap.bsp.core-8.1.1.jar
/opt/iobjects_spark/com.supermap.bsp.examples-8.1.1.jar --input hdfs://sparkalone-
master:9000/input/taxi.json --distance "10 meter" --output
/opt/resultData/BufferResult.json
```

简单解释下命令

--class　表示主类名称，含包名，本例子指的是需要执行的类

--master　Spark 集群总入口

--conf　spark.default.parallelism 值跟集群的总核数有关，建议是总核数的两到三倍

--jars　依赖的第三方 jar 包

--input　简单理解为操作数据来源

--distance　参数，这里只范例中缓冲距离

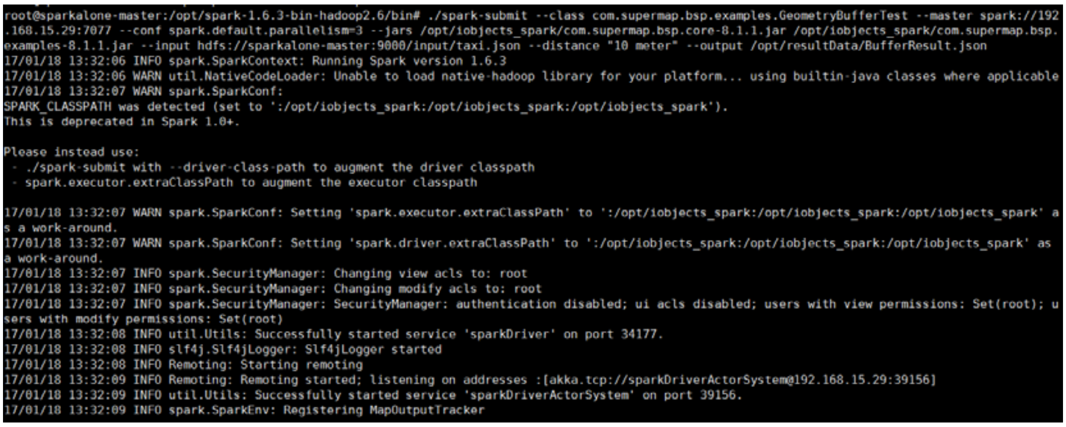
--output　结果输出到哪里

注意：如果集群模式运行不成功，

将-master spark://192.168.15.29:7077 修改为：-master local[4]本地运行

如果 local 模式运行还是报错，请参照常见问题 5.8

执行开始

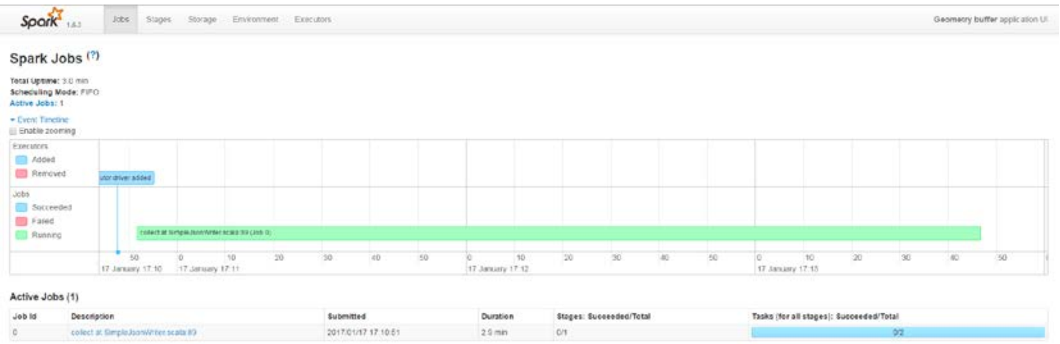


```
root@sparkalone-master:/opt/spark-1.6.3-bin-hadoop2.6/bin# ./spark-submit --class com.supermap.bsp.examples.GeometryBufferTest --master spark://192
.168.15.29:7077 --conf spark.default.parallelism=3 --jars /opt/iobjects_spark/com.supermap.bsp.core-8.1.1.jar /opt/iobjects_spark/com.supermap.bsp.
examples-8.1.1.jar --input hdfs://sparkalone-master:9000/input/taxi.json --distance "10 meter" --output /opt/resultData/BufferResult.json
17/01/18 13:32:06 INFO spark.SparkContext: Running Spark version 1.6.3
17/01/18 13:32:06 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
17/01/18 13:32:07 WARN spark.SparkConf:
SPARK_CLASSPATH was detected (set to ':/opt/iobjects_spark:/opt/iobjects_spark:/opt/iobjects_spark').
This is deprecated in Spark 1.0+.

Please instead use:
 - ./spark-submit with --driver-class-path to augment the driver classpath
 - spark.executor.extraClassPath to augment the executor classpath

17/01/18 13:32:07 WARN spark.SparkConf: Setting 'spark.executor.extraClassPath' to ':/opt/iobjects_spark:/opt/iobjects_spark:/opt/iobjects_spark' a
s a work-around.
17/01/18 13:32:07 WARN spark.SparkConf: Setting 'spark.driver.extraClassPath' to ':/opt/iobjects_spark:/opt/iobjects_spark:/opt/iobjects_spark' as
a work-around.
17/01/18 13:32:07 INFO spark.SecurityManager: Changing view acls to: root
17/01/18 13:32:07 INFO spark.SecurityManager: Changing modify acls to: root
17/01/18 13:32:07 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(root); u
sers with modify permissions: Set(root)
17/01/18 13:32:08 INFO util.Utils: Successfully started service 'sparkDriver' on port 34177.
17/01/18 13:32:08 INFO slf4j.Slf4jLogger: Slf4jLogger started
17/01/18 13:32:08 INFO Remoting: Starting remoting
17/01/18 13:32:09 INFO Remoting: Remoting started; listening on addresses :[akka.tcp://sparkDriverActorSystem@192.168.15.29:39156]
17/01/18 13:32:09 INFO util.Utils: Successfully started service 'sparkDriverActorSystem' on port 39156.
17/01/18 13:32:09 INFO spark.SparkEnv: Registering MapOutputTracker
```

执行过程中，可以访问 http://192.168.15.29:4040/jobs/ 查看执行情况



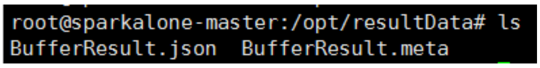
Spark Jobs (?)

Total Uptime: 3.0 min
Scheduling Mode: FIFO
Active Jobs: 1

Active Jobs (1)

Job Id	Description	Submitted	Duration	Stages: Succeeded/Total	Tasks (for all stages): Succeeded/Total
0	collect at SimpleJsonWriter.scala:89	2017/01/17 17:10:51	2.5 min	0/1	0/2

执行完成后，访问/opt/resultData 可以查看完成的内容

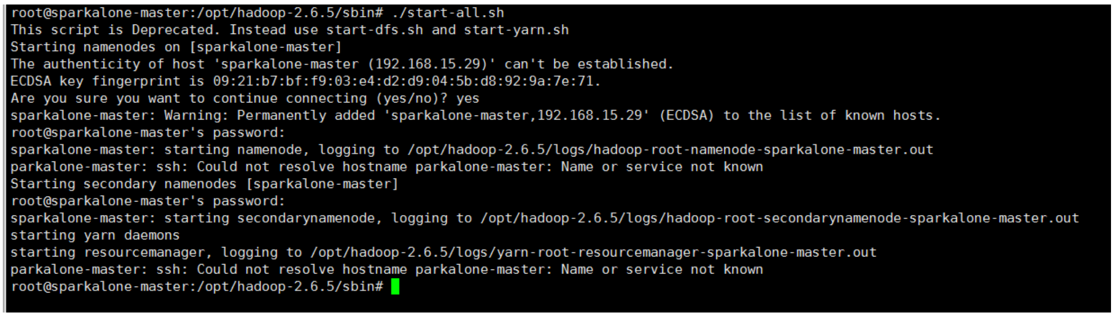


```
root@sparkalone-master:/opt/resultData# ls
BufferResult.json  BufferResult.meta
```

五、常见问题

5.1 启动 hadoop 来回让输入密码

答：没设置免 ssh 密码登陆， 截图如下，解决办法请参考 1.2 和 1.4



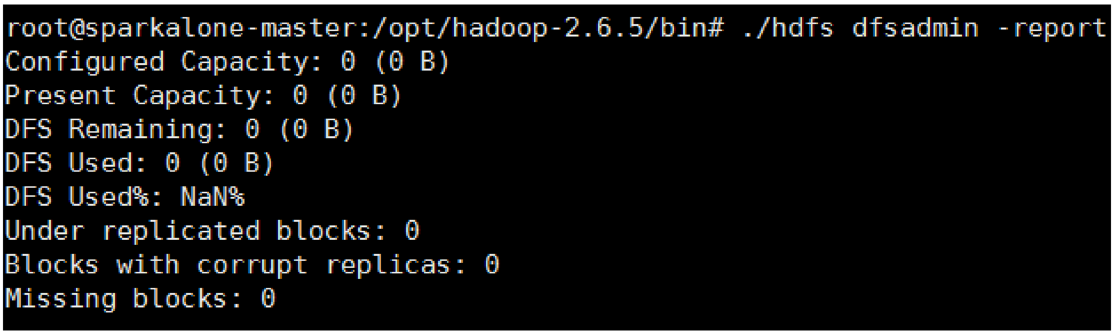
```
root@sparkalone-master:/opt/hadoop-2.6.5/sbin# ./start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
Starting namenodes on [sparkalone-master]
The authenticity of host 'sparkalone-master (192.168.15.29)' can't be established.
ECDSA key fingerprint is 09:21:b7:bf:f9:03:e4:d2:d9:04:5b:d8:92:9a:7e:71.
Are you sure you want to continue connecting (yes/no)? yes
sparkalone-master: Warning: Permanently added 'sparkalone-master,192.168.15.29' (ECDSA) to the list of known hosts.
root@sparkalone-master's password:
sparkalone-master: starting namenode, logging to /opt/hadoop-2.6.5/logs/hadoop-root-namenode-sparkalone-master.out
parkalone-master: ssh: Could not resolve hostname parkalone-master: Name or service not known
Starting secondary namenodes [sparkalone-master]
root@sparkalone-master's password:
sparkalone-master: starting secondarynamenode, logging to /opt/hadoop-2.6.5/logs/hadoop-root-secondarynamenode-sparkalone-master.out
starting yarn daemons
starting resourcemanager, logging to /opt/hadoop-2.6.5/logs/yarn-root-resourcemanager-sparkalone-master.out
parkalone-master: ssh: Could not resolve hostname parkalone-master: Name or service not known
root@sparkalone-master:/opt/hadoop-2.6.5/sbin# 
```

5.2 如何判断 hadoop 集群是否启动成功

答：可以，进入 hadoop-2.6.5/bin 下执行

```
./hdfs dfsadmin -report
```

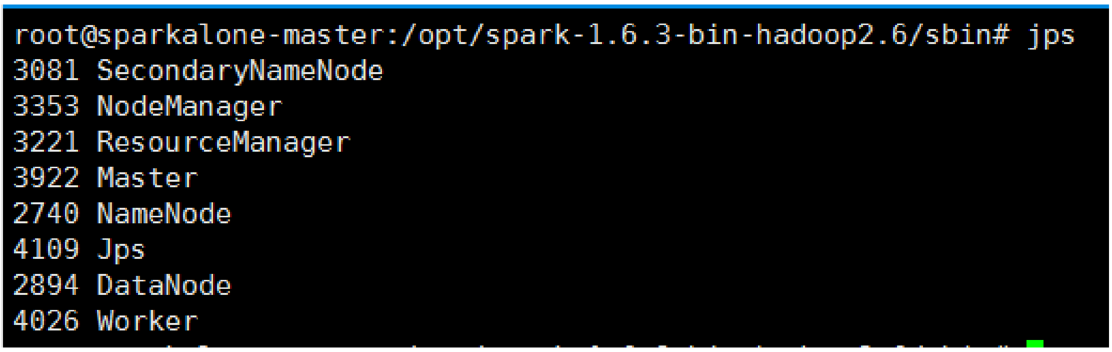
下图就是没启动成功，具体原因可以查看 hadoop-2.6.5/logs 日志



```
root@sparkalone-master:/opt/hadoop-2.6.5/bin# ./hdfs dfsadmin -report
Configured Capacity: 0 (0 B)
Present Capacity: 0 (0 B)
DFS Remaining: 0 (0 B)
DFS Used: 0 (0 B)
DFS Used%: NaN%
Under replicated blocks: 0
Blocks with corrupt replicas: 0
Missing blocks: 0
```

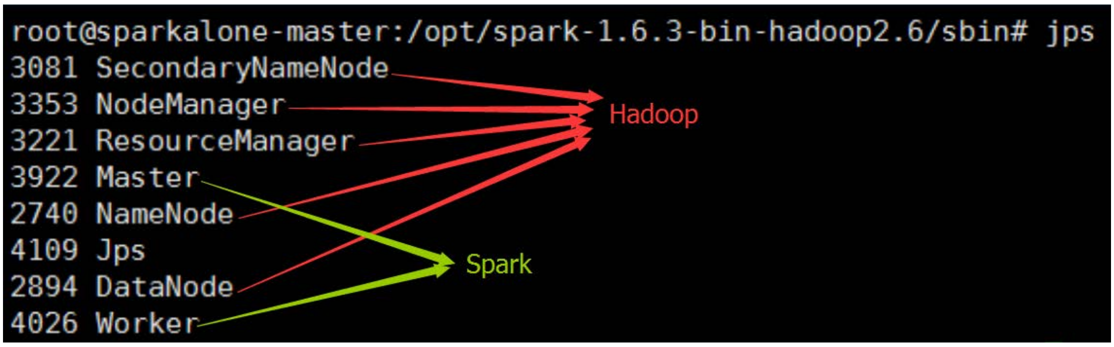
5.3 还有什么命令可以查看集群是否启动成功

答：输入 jps，根据启动的进程，查看是否启动成功



```
root@sparkalone-master:/opt/spark-1.6.3-bin-hadoop2.6/sbin# jps
3081 SecondaryNameNode
3353 NodeManager
3221 ResourceManager
3922 Master
2740 NameNode
4109 Jps
2894 DataNode
4026 Worker
```

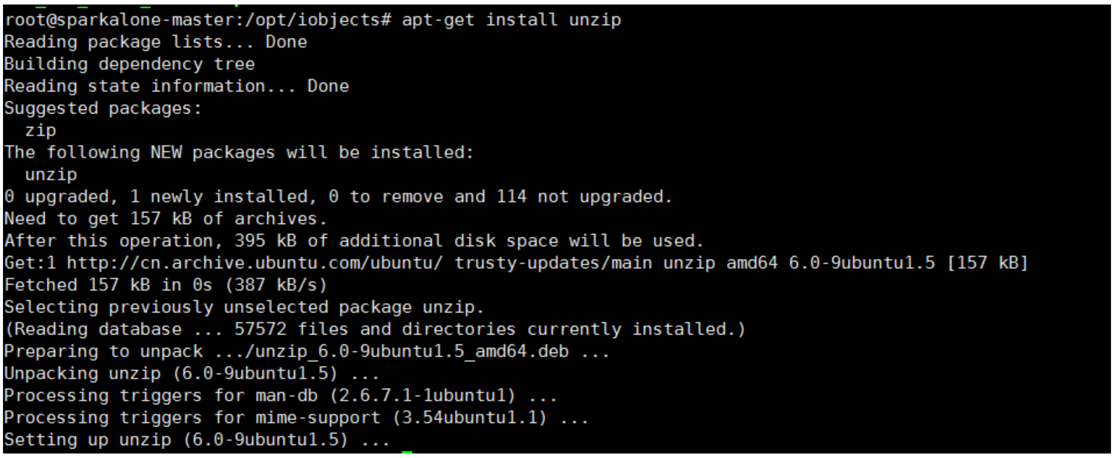
其中



```
root@sparkalone-master:/opt/spark-1.6.3-bin-hadoop2.6/sbin# jps
3081 SecondaryNameNode
3353 NodeManager
3221 ResourceManager
3922 Master
2740 NameNode
4109 Jps
2894 DataNode
4026 Worker
```

5.4 如何解压 zip 文件

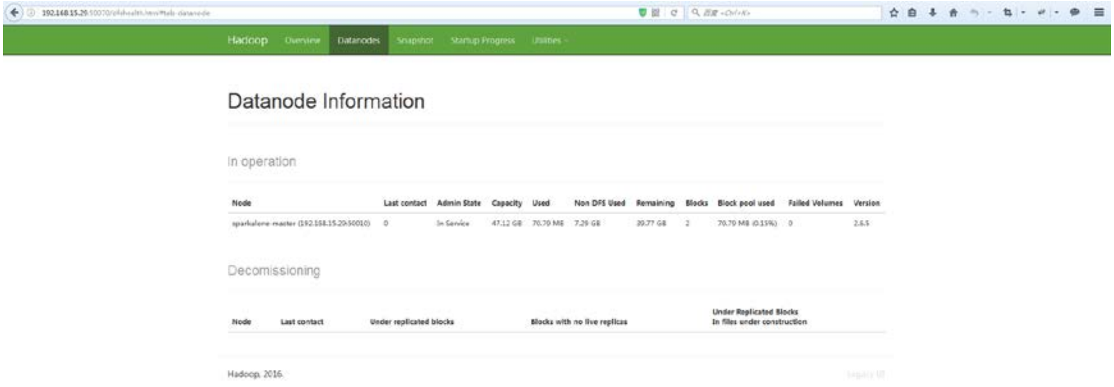
答：需要安装 unzip 命令，安装完成后使用 unzip+文件全名方式解压



```
root@sparkalone-master:/opt/iobjects# apt-get install unzip
Reading package lists... Done
Building dependency tree
Reading state information... Done
Suggested packages:
  zip
The following NEW packages will be installed:
  unzip
0 upgraded, 1 newly installed, 0 to remove and 114 not upgraded.
Need to get 157 kB of archives.
After this operation, 395 kB of additional disk space will be used.
Get:1 http://cn.archive.ubuntu.com/ubuntu/ trusty-updates/main unzip amd64 6.0-9ubuntu1.5 [157 kB]
Fetched 157 kB in 0s (387 kB/s)
Selecting previously unselected package unzip.
(Reading database ... 57572 files and directories currently installed.)
Preparing to unpack .../unzip_6.0-9ubuntu1.5_amd64.deb ...
Unpacking unzip (6.0-9ubuntu1.5) ...
Processing triggers for man-db (2.6.7.1-1ubuntu1) ...
Processing triggers for mime-support (3.54ubuntu1.1) ...
Setting up unzip (6.0-9ubuntu1.5) ...
```

5.5 如何有界面方式查看 hdfs 存储大小

答：访问 http://192.168.15.29:50070/dfshealth.html#tab-datanode



Datanode Information

In operation

Node	Last contact	Admin State	Capacity	Used	Non DFS Used	Remaining	Blocks	Block pool used	Failed Volumes	Version
sparkalone-master (192.168.15.29:50010)	0	In Service	47.12 GB	70.70 MB	7.29 GB	39.77 GB	2	70.70 MB (0.15%)	0	2.6.5

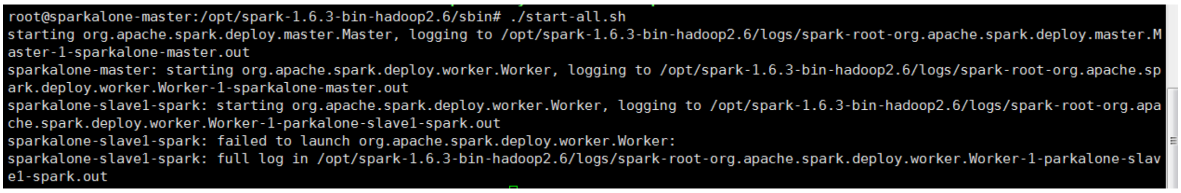
Decomissioning

Node	Last contact	Under replicated blocks	Blocks with no live replicas	Under Replicated Blocks In files under construction

Hadoop, 2016.

5.6 Spark 启动报错如何查看

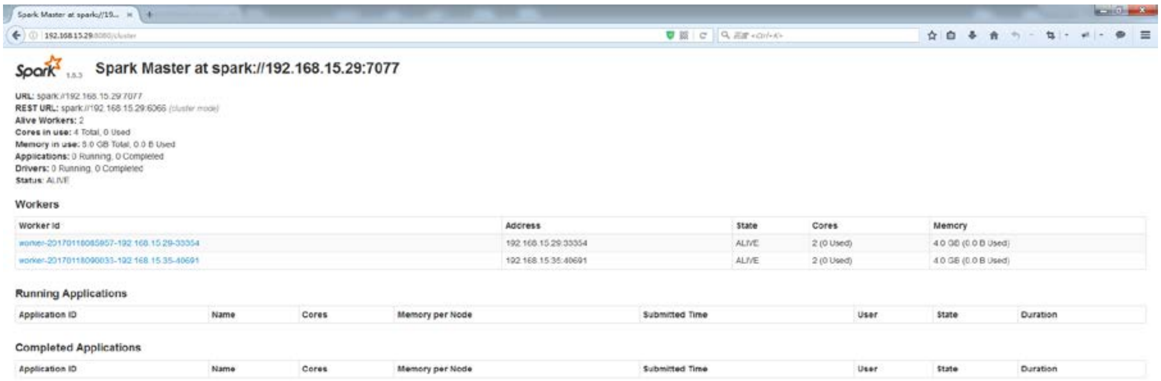
答：例如如下 Spark 启动报错，查看 sparkalone-slave1-spark 机器 Spark 下 logs 日志即可



```
root@sparkalone-master:/opt/spark-1.6.3-bin-hadoop2.6/sbin# ./start-all.sh
starting org.apache.spark.deploy.master.Master, logging to /opt/spark-1.6.3-bin-hadoop2.6/logs/spark-root-org.apache.spark.deploy.master.Master-1-sparkalone-master.out
sparkalone-master: starting org.apache.spark.deploy.worker.Worker, logging to /opt/spark-1.6.3-bin-hadoop2.6/logs/spark-root-org.apache.spark.deploy.worker.Worker-1-sparkalone-master.out
sparkalone-slave1-spark: starting org.apache.spark.deploy.worker.Worker, logging to /opt/spark-1.6.3-bin-hadoop2.6/logs/spark-root-org.apache.spark.deploy.worker.Worker-1-parkalone-slave1-spark.out
sparkalone-slave1-spark: failed to launch org.apache.spark.deploy.worker.Worker:
sparkalone-slave1-spark: full log in /opt/spark-1.6.3-bin-hadoop2.6/logs/spark-root-org.apache.spark.deploy.worker.Worker-1-parkalone-slave1-spark.out
```

5.7 Spark 集群启动成功和集群执行过程如何查看

答：访问 http://192.168.15.29:8080/



Spark Master at spark://192.168.15.29:7077

URL: spark://192.168.15.29:7077
REST URL: spark://192.168.15.29:6066 *(cluster mode)*
Alive Workers: 2
Cores in use: 4 Total, 0 Used
Memory in use: 8.0 GB Total, 0.0 B Used
Applications: 0 Running, 0 Completed
Drivers: 0 Running, 0 Completed
Status: ALIVE

Workers

Worker Id	Address	State	Cores	Memory
worker-20170118085957-192.168.15.29-33354	192.168.15.29:33354	ALIVE	2 (0 Used)	4.0 GB (0.0 B Used)
worker-20170118090033-192.168.15.35-40691	192.168.15.35:40691	ALIVE	2 (0 Used)	4.0 GB (0.0 B Used)

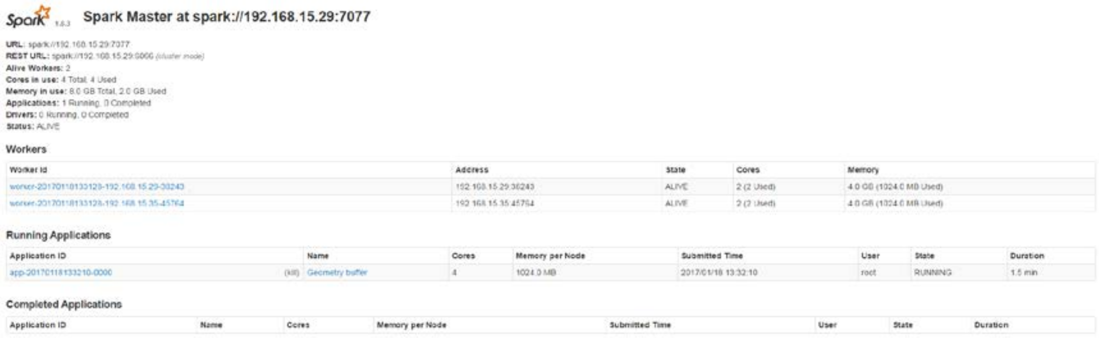
Running Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration

执行过程截图：



Spark 1.6.3 Spark Master at spark://192.168.15.29:7077

URL: spark://192.168.15.29:7077
REST URL: spark://192.168.15.29:6066 *(cluster mode)*
Alive Workers: 2
Cores in use: 4 Total, 4 Used
Memory in use: 8.0 GB Total, 2.0 GB Used
Applications: 1 Running, 0 Completed
Drivers: 0 Running, 0 Completed
Status: ALIVE

Workers

Worker Id	Address	State	Cores	Memory
worker-20170118133128-192.168.15.29-30243	192.168.15.29:30243	ALIVE	2 (2 Used)	4.0 GB (1024.0 MB Used)
worker-20170118133128-192.168.15.35-45764	192.168.15.35:45764	ALIVE	2 (2 Used)	4.0 GB (1024.0 MB Used)

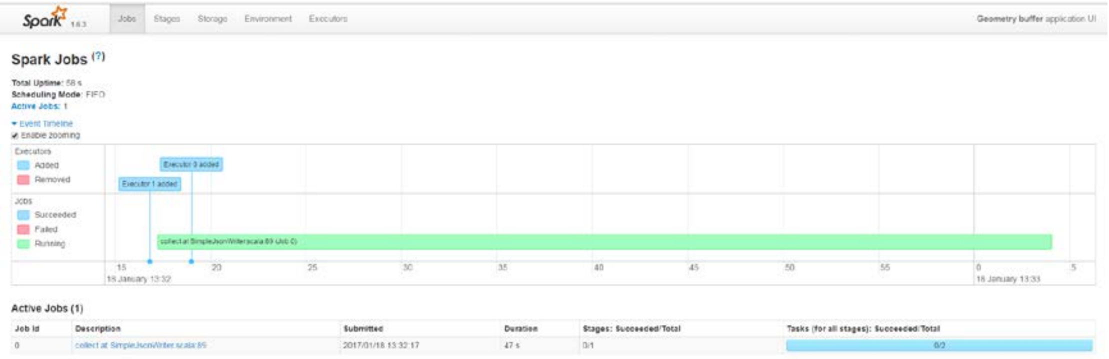
Running Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20170118133210-0000 (kill)	Geometry buffer	4	1024.0 MB	2017/01/18 13:32:10	root	RUNNING	1.5 min

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration

访问 http://192.168.15.29:4040/jobs/



Spark Jobs (?)

Total Uptime: 58 s
Scheduling Mode: FIFO
Active Jobs: 1

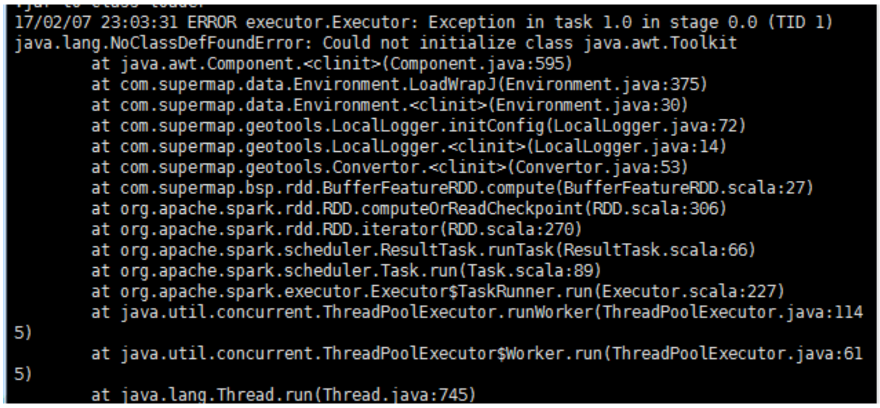
▾ Event Timeline
☑ Enable zooming

Active Jobs (1)

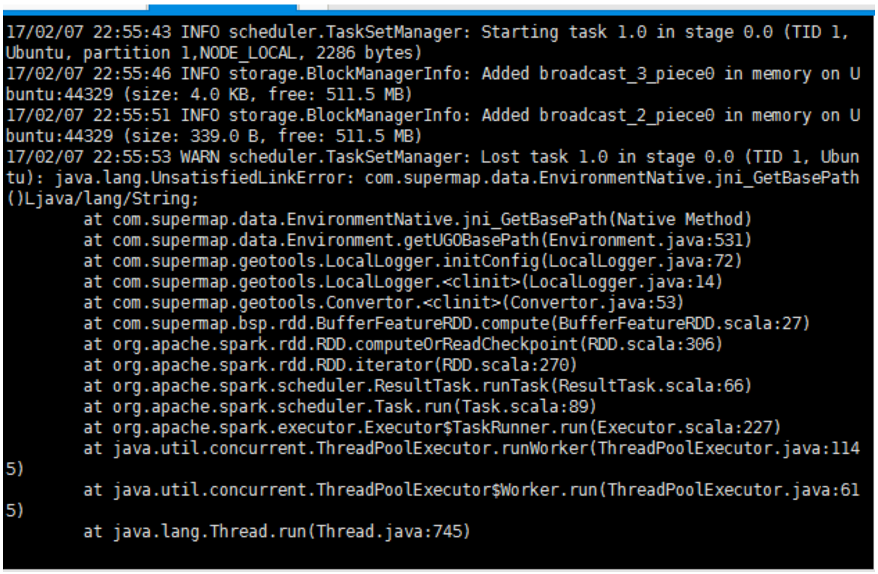
Job Id	Description	Submitted	Duration	Stages: Succeeded/Total	Tasks (for all stages): Succeeded/Total
0	collect at SimpleJsonWriter.scala:89	2017/01/18 13:32:17	47 s	0/1	0/2

5.8 iObjects for Spark 程序使用 local 模式运行不成功，报错

java.lang.NoClassDefFoundError: Could not initialize class java.awt.Toolkit
java.lang.NoClassDefFoundError: Could not initialize class com.supermap.geotools.Convertor



```
17/02/07 23:03:31 ERROR executor.Executor: Exception in task 1.0 in stage 0.0 (TID 1)
java.lang.NoClassDefFoundError: Could not initialize class java.awt.Toolkit
        at java.awt.Component.<clinit>(Component.java:595)
        at com.supermap.data.Environment.LoadWrapJ(Environment.java:375)
        at com.supermap.data.Environment.<clinit>(Environment.java:30)
        at com.supermap.geotools.LocalLogger.initConfig(LocalLogger.java:72)
        at com.supermap.geotools.LocalLogger.<clinit>(LocalLogger.java:14)
        at com.supermap.geotools.Convertor.<clinit>(Convertor.java:53)
        at com.supermap.bsp.rdd.BufferFeatureRDD.compute(BufferFeatureRDD.scala:27)
        at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:306)
        at org.apache.spark.rdd.RDD.iterator(RDD.scala:270)
        at org.apache.spark.scheduler.ResultTask.runTask(ResultTask.scala:66)
        at org.apache.spark.scheduler.Task.run(Task.scala:89)
        at org.apache.spark.executor.Executor$TaskRunner.run(Executor.scala:227)
        at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:114
5)
        at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:61
5)
        at java.lang.Thread.run(Thread.java:745)
```



```
17/02/07 22:55:43 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 0.0 (TID 1,
Ubuntu, partition 1,NODE_LOCAL, 2286 bytes)
17/02/07 22:55:46 INFO storage.BlockManagerInfo: Added broadcast_3_piece0 in memory on U
buntu:44329 (size: 4.0 KB, free: 511.5 MB)
17/02/07 22:55:51 INFO storage.BlockManagerInfo: Added broadcast_2_piece0 in memory on U
buntu:44329 (size: 339.0 B, free: 511.5 MB)
17/02/07 22:55:53 WARN scheduler.TaskSetManager: Lost task 1.0 in stage 0.0 (TID 1, Ubun
tu): java.lang.UnsatisfiedLinkError: com.supermap.data.EnvironmentNative.jni_GetBasePath
()Ljava/lang/String;
        at com.supermap.data.EnvironmentNative.jni_GetBasePath(Native Method)
        at com.supermap.data.Environment.getUGOBasePath(Environment.java:531)
        at com.supermap.geotools.LocalLogger.initConfig(LocalLogger.java:72)
        at com.supermap.geotools.LocalLogger.<clinit>(LocalLogger.java:14)
        at com.supermap.geotools.Convertor.<clinit>(Convertor.java:53)
        at com.supermap.bsp.rdd.BufferFeatureRDD.compute(BufferFeatureRDD.scala:27)
        at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:306)
        at org.apache.spark.rdd.RDD.iterator(RDD.scala:270)
        at org.apache.spark.scheduler.ResultTask.runTask(ResultTask.scala:66)
        at org.apache.spark.scheduler.Task.run(Task.scala:89)
        at org.apache.spark.executor.Executor$TaskRunner.run(Executor.scala:227)
        at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:114
5)
        at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:61
5)
        at java.lang.Thread.run(Thread.java:745)
```

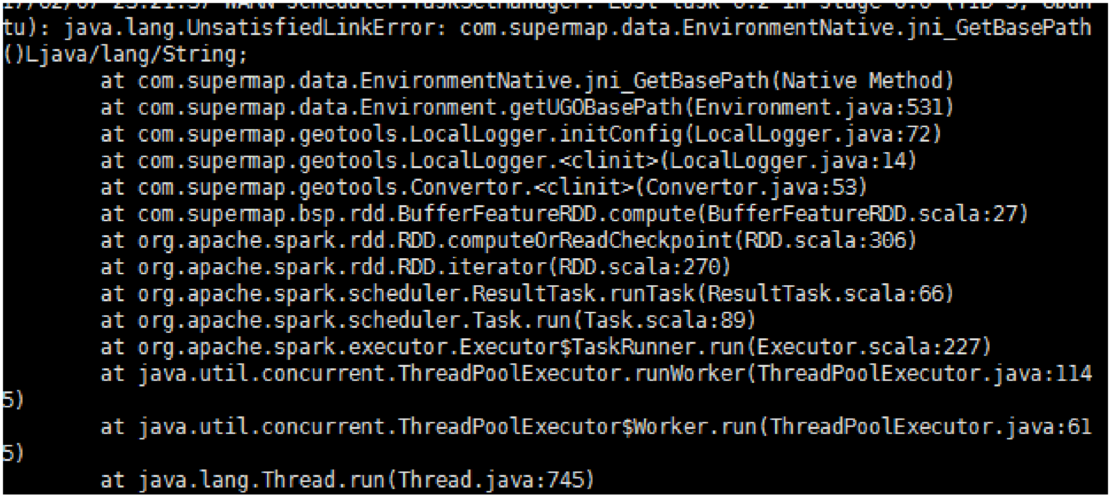
答：（原因是没装 X11 等依赖）

可以使用 iServer Support 中提供的一键化工具（810+版本），复制 support 目录到系统中。

```
./dependencies_check_and_install.sh install -y
```

5.9 iObjects for Spark 程序使用集群模式运行不成功，报错

```
java.lang.UnsatisfiedLinkError:
com.supermap.data.EnvironmentNative.jni_GetBasePath()Ljava/lang/String;
```



```
17/02/07 23:21:37 WARN scheduler.TaskSetManager: Lost task 0.2 in stage 0.0 (TID 3, Ubun
tu): java.lang.UnsatisfiedLinkError: com.supermap.data.EnvironmentNative.jni_GetBasePath
()Ljava/lang/String;
        at com.supermap.data.EnvironmentNative.jni_GetBasePath(Native Method)
        at com.supermap.data.Environment.getUGOBasePath(Environment.java:531)
        at com.supermap.geotools.LocalLogger.initConfig(LocalLogger.java:72)
        at com.supermap.geotools.LocalLogger.<clinit>(LocalLogger.java:14)
        at com.supermap.geotools.Convertor.<clinit>(Convertor.java:53)
        at com.supermap.bsp.rdd.BufferFeatureRDD.compute(BufferFeatureRDD.scala:27)
        at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:306)
        at org.apache.spark.rdd.RDD.iterator(RDD.scala:270)
        at org.apache.spark.scheduler.ResultTask.runTask(ResultTask.scala:66)
        at org.apache.spark.scheduler.Task.run(Task.scala:89)
        at org.apache.spark.executor.Executor$TaskRunner.run(Executor.scala:227)
        at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:114
5)
        at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:61
5)
        at java.lang.Thread.run(Thread.java:745)
```

答：使用 iObjects Java 中 libmawt.so 替换 jdk1.7.0_80/jre/lib/amd64/headless 中的 libmawt.so