

自动化参数调优模型操作手册

安装部署流程

1. 下载安装JDK 1.8，安装配置说明[here](#)
2. 安装配置Hadoop2.7，设置一台主机为master节点，其余为worker节点，安装配置流程见[here](#)
3. 在.bash_profile中对JDK和Hadoop的环境变量进行设置

```
#java environment
export JAVA_HOME=/home/gurong1228/jdk1.8.0_221
export PATH=$PATH:$JAVA_HOME/bin:$JAVA_HOME/jre/bin
export CLASSPATH=.:$JAVA_HOME/lib:$JAVA_HOME/jre/lib:$CLASSPATH

#hadoop environment
export HADOOP_HOME=/home/gurong1228/hadoop-2.7.7
export PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin
```

4. 下载安装Alluxio 2.1，设置与Hadoop相同的master节点，其余为worker节点，安装配置流程见[here](#)
5. 在.bash_profile中对Alluxio的环境变量进行配置

```
#alluxio environment
export ALLUXIO_HOME=/home/gurong1228/alluxio-2.1.0
export PATH=$PATH:$JAVA_HOME/bin:$ALLUXIO_HOME/bin
```

6. 将Hadoop底层文件系统HDFS作为Alluxio的底层存储，具体配置见[here](#)

运行流程

1. 在正式运行模型之前，需要保证上述软件JDK、Hadoop、Alluxio等成功安装，并且Alluxio的底层存储为HDFS。
2. 找到脚本文件Job.sh。其中Job.sh为主文件，它会调用ruleModel.sh文件；ruleModel.sh执行参数调优功能，它会调用资源采样脚本getSample.sh和资源监测脚本getResource.sh。
3. 修改自定义配置，视实际情况修改。其中，需要注意取消作业重复执行带来的副作用（如：File exists）。

```
sourceFileName="resource.txt"
# 脚本文件中均有该项配置，指定了脚本间信息交流的方式
jobcommand="alluxio fs copyFromLocal /home/gurong1228/bigfile/bigfile_10GB /"
# 指定了需要执行的作业命令，在Job.sh和ruleModel.sh中进行配置
properties_name='/home/gurong1228/alluxio-2.1.0/conf/alluxio-site.properties'
# 在ruleModel.sh中进行配置，指定了Alluxio配置文件的位置
netPort='em1';netSpeed=125;
# 在ruleModel.sh中进行配置，分别指定了监测的网络端口和网络带宽的值
runningTime=90;diet=0.05;
# 在ruleModel.sh中进行配置，分别指定了模型为每个参数分配的作业运行时间和资源使用的提升比例

#此外，可能存在某项资源利用率低但提升较困难的情况，因此，本模型可以手动设置针对某项资源的调参，设置
相应的CPU_chosen、Memory_chosen、Network_chosen为true即可。
```

4. 根据实际情况完成上述操作后，即可运行脚本Job.sh。下图表明了未优化参数配置下平均的资源使用情况。

```
[gurong1228@slave205 resourceInfo]$ ./Job.sh
/bigfile_10GB has been removed
nohup: redirecting stderr to stdout
1.217 7.300 115.975
nohup: redirecting stderr to stdout
26.883 7.800 113.78
nohup: redirecting stderr to stdout
15.533 7.700 113.517
nohup: redirecting stderr to stdout
12.683 7.700 114.807
nohup: redirecting stderr to stdout
11.183 7.700 116.037
nohup: redirecting stderr to stdout
9.783 7.700 114.805
nohup: redirecting stderr to stdout
9.100 7.700 114.272
nohup: redirecting stderr to stdout
8.617 7.700 115.499
nohup: redirecting stderr to stdout
8.617 7.800 116.367
nohup: redirecting stderr to stdout
8.267 7.800 115.337
Average rate: 11.19 7.69 115.04
under default parameter : 11.19 7.69 115.04
judge finished
```

5. 在脚本中设置针对CPU的优化，从下图中可以看出，已经开始针对候选值进行尝试并对每个采样点输出当前资源的使用情况。如果资源利用率并未超出一定的比例（ruleModel.sh中设置的diet）则不会停止搜索。

```

optimize from CPU
testing alluxio.user.network.netty.worker.threads with value 0
alluxio-site.properties
alluxio-site.properties
/bigfile* does not exist.
nohup: ignoring input and redirecting stderr to stdout
1.217 7.300 113.272
nohup: ignoring input and redirecting stderr to stdout
26.383 7.800 113.936
nohup: ignoring input and redirecting stderr to stdout
16.717 7.800 114.204
nohup: ignoring input and redirecting stderr to stdout
13.350 7.800 116.51
nohup: ignoring input and redirecting stderr to stdout
11.417 7.800 116.543
nohup: ignoring input and redirecting stderr to stdout
10.267 7.800 114.725
nohup: ignoring input and redirecting stderr to stdout
9.483 7.800 114.805
nohup: ignoring input and redirecting stderr to stdout
8.883 7.800 114.314
nohup: ignoring input and redirecting stderr to stdout
8.867 7.900 114.96
nohup: ignoring input and redirecting stderr to stdout
8.467 7.900 116.532
Average rate: 11.51 7.77 114.98
bigfile erased

```

6. 由下图可以看出，当资源使用率超过一定限度后，本模型会停止搜索，并将当前的参数配置作为推荐值，并重新执行作业。

```

testing alluxio.user.network.netty.worker.threads with value 1
alluxio-site.properties
alluxio-site.properties
/bigfile_10GB has been removed
nohup: ignoring input and redirecting stderr to stdout
1.217 7.300 113.637
nohup: ignoring input and redirecting stderr to stdout
27.550 7.800 114.493
nohup: ignoring input and redirecting stderr to stdout
17.617 7.800 114.607
nohup: ignoring input and redirecting stderr to stdout
13.800 7.800 114.426
nohup: ignoring input and redirecting stderr to stdout
11.750 7.800 116.668
nohup: ignoring input and redirecting stderr to stdout
10.533 7.800 114.67
nohup: ignoring input and redirecting stderr to stdout
9.633 7.900 115.039
nohup: ignoring input and redirecting stderr to stdout
8.950 7.900 115.26
nohup: ignoring input and redirecting stderr to stdout
8.917 7.900 114.649
nohup: ignoring input and redirecting stderr to stdout
8.483 7.900 116.484
Average rate: 11.85 7.79 114.99
bigfile erased
Successfully change the alluxio.user.network.netty.worker.threads's value to 1
restart the job after the parameter optimization
/bigfile_10GB has been removed

```

7. 在重新执行作业的同时，也会再次对作业进行采样分析，以验证参数的有效性。

```

Copied file:///home/gurong1228/bigfile/bigfile_10GB to /
the average resource occupied partly are (some samples from the beginning):
CPU 11.83 Memory 7.85 Network 115.03

```

