

alluxio安装和 结合cdh使用，alluxio和spark整合

访问官网下载地址：<http://www.alluxio.org/download>，选择你想下载的对版本。
如果使用的Hadoop环境是CDH的，那么下载相应的CDH版本就好。这样兼容性更好。

```
wget http://downloads.alluxio.org/downloads/files/
```

第二步：安装好JDK8, 解压Alluxio并配置环境变量。

```
export JAVA_HOME=/opt/core/jdk1.8.0_102
export ALLUXIO_HOME=/opt/core/alluxio
export PATH=.:$JAVA_HOME/bin:$ALLUXIO_HOME/bin:$PATH
```

第三步：配置alluxio配置文件。

- 1) 在目录\$ALLUXIO_HOME/conf的目录下面找到配置文件alluxio-env.sh
编辑以及更改一下配置项：

```
ALLUXIO_RAM_FOLDER=/opt/data/ramdisk
```

解释：内存映射到磁盘下目录（读取文件的时候读的是内存，内存！！不是磁盘。）

```
ALLUXIO_UNDERFS_ADDRESS=hdfs://serviceName/alluxio
```

解释：Alluxio底层存储的文件系统，可以是s3以及其他的文件系统，这里用HDFS。（官方配置的是namenode的地址和服务端口，这里直接用的serviceName, 目的是支持多个Namenode HA的情况）

```
ALLUXIO_WORKER_MEMORY_SIZE=90GB
```

解释：这个是每台机器上面的worker节点所占用的内存大小，根据需求自动调整。

```
ALLUXIO_JAVA_OPTS="-
```

```
Dalluxio.zookeeper.enabled=true -
```

```
Dalluxio.zookeeper.address=zk1:2181, zk2:2181, zk3:2181 -
```

```
Dalluxio.master.journal.folder=hdfs://serviceName/alluxio/journal"
```

解释：启用Alluxio高可用的配置项，启动用户必须有权限在HDFS上面写数据。这个配置项是全局的，也就是在master, worker, client上面都起作用。

```
ALLUXIO_USER_JAVA_OPTS="-Dalluxio.zookeeper.enabled=true -  
Dalluxio.zookeeper.address=zk1:2181, zk2:2181, zk3:2181"
```

解释：配置客户端的配置项，全局已经配置，这里可以省略。

2) 更加高级的配置项可以在\$ALLUXIO_HOME/conf/alluxio-site.properties定制化配置。

如：启用多个存储策略（目前支持mem, SSD, disk），也可以定制存储的算法。

同时也可以配置用户的权限。

其他的配置项目可以参考官方网址：

第四步：把hadoop下面的core-site.xml, hdfs-site.xml拷贝到\$ALLUXIO_HOME/conf/的目录下面。若没有拷贝可能造成解析不了hdfs的serviceName，最后把worker的配置文件配置你想要配置的worker节点。

第五步：启动master, worker节点。

```
alluxio-start.sh master  
alluxio-start.sh worker
```

使用HDFS在本地运行Alluxio

配置完成后，你可以在本地启动Alluxio，观察一切是否正常运行：

1. \$./bin/alluxio format
2. \$./bin/alluxio-start.sh local

该命令应当会启动一个Alluxio master和一个Alluxio worker，可以在浏览器中访问<http://localhost:19999>查看master Web UI。

接着，你可以运行一个简单的示例程序：

```
$ ./bin/alluxio runTests
```

运行成功后，访问HDFS Web UI <http://localhost:50070>，确认其中包含了由Alluxio创建的文件和目录。在该测试中，创建的文件名称应像这

样：/alluxio/data/default_tests_files/BasicFile_STORE_SYNC_PERSIST。

运行以下命令停止Alluxio：

```
$ ./bin/alluxio-stop.sh all
```

测试：

```
sudo -u spark spark-shell --jars /opt/alluxio/client/alluxio-1.7.1-client.jar

hadoop fs -put -f LICENSE hdfs://192.168.33.101:8020/alluxio/LICENSE
val s = sc.textFile("alluxio://localhost:19998/LICENSE")
val double = s.map(line => line + line)
double.saveAsTextFile("alluxio://localhost:19998/LICENSE5")
```

ALLuxio使用常见问题

问题：出现“No FileSystem for scheme: alluxio”这种错误信息是什么原因？

解决办法：当你的应用（例如MapReduce、Spark）尝试以HDFS兼容文件系统接口访问Alluxio，而又无法解析alluxio://模式时会产生该异常。要确保HDFS配置文件core-site.xml（默认在hadoop安装目录，如果为Spark自定义了该文件则在spark/conf/目录下）包含以下配置：

```
<configuration>
  <property>
    <name>fs.alluxio.impl</name>
    <value>alluxio.hadoop.FileSystem</value>
```

```
</property>
</configuration>
```

问题：出现“java.lang.RuntimeException: java.lang.ClassNotFoundException: Class alluxio.hadoop.FileSystem not found”这种错误信息是什么原因？

解决办法：当你的应用（例如MapReduce、Spark）尝试以HDFS兼容文件系统接口访问Alluxio，并且alluxio://模式也已配置正确，但应用的classpath未包含Alluxio客户端jar包时会产生该异常。用户通常需要通过设置环境变量或者属性的方式将Alluxio客户端jar包添加到所有节点上的应用的classpath中，这取决于具体的计算框架。以下是一些示例：

- 对于MapReduce应用，可以将客户端jar包添加到\$HADOOP_CLASSPATH:

```
export HADOOP_CLASSPATH=/opt/alluxio/client/alluxio-1.7.1-
client.jar:${HADOOP_CLASSPATH}
```

- 对于Spark应用，可以将客户端jar包添加到\$SPARK_CLASSPATH:

```
export SPARK_CLASSPATH=/opt/alluxio/client/alluxio-1.7.1-
client.jar:${SPARK_CLASSPATH}
```

除了上述方法，还可以将以下配置添加到spark/conf/spark-defaults.conf中：

```
spark.driver.extraClassPath /opt/alluxio/client/alluxio-1.7.1-client.jar
spark.executor.extraClassPath /opt/alluxio/client/alluxio-1.7.1-client.jar
```

如果已经设置相关的classpath，但是异常仍然存在，用户可以这样检测路径是否有效：

```
ls /opt/alluxio/client/alluxio-1.7.1-client.jar
```

临时方法：

```
sudo -u hdfs spark-shell --jars /opt/alluxio/client/alluxio-1.7.1-
client.jar
```

错误二

```
java.io.IOException: alluxio.exception.AccessControlException:  
Permission denied: user=hdfs, access=-w-, path=/LICENSE2/_temporary/0:  
failed at /, inode owner=root, inode group=root, inode mode=rwxr-xr-x
```

解决方法:

```
vi conf/alluxio-site.properties
```

增加

```
alluxio.security.login.username=hdfs
```

错误三

最新版需要jdk1.8 ，修改spark用jdk1.8



参考官网文档:

<https://www.alluxio.org/docs/master/cn/Getting-Started.html>