Flink 1.13.2 集群安装部署的 3 种方式(建议收藏)

本文作者：在 IT 中穿梭旅行

本文档来自公众号：3 分钟秒懂大数据

大家好，我是土哥。

今天，有位 Flink 初学者问我有没有 Flink 的安装教程，看到这后，土哥二话不说直接安排上。

以下教程全部使用 Flink1.13.2 版本，在普通用户下面部署：

1、Standalone 部署

版本要求：

版本	节点	部署方式
flink-1.13.2-bin-scala_2.11.tgz	192.168.244.129	standalone

1.1 将软件安装包放入集群中

```
[liyaozhou@hlinkui1 lyz]$ ll
总用量 306572
drwxrwxr-x.  3 liyaozhou liyaozhou          84 6月   1 16:29 data
-rw-r--r--.  1 liyaozhou liyaozhou 313922934 9月  17 11:09 flink-1.13.2-bin-scala_2.11.tgz
drwxr-xr-x. 10 liyaozhou liyaozhou         161 9月  16 19:10 hadoop-2.6.4
drwxrwxr-x.  9 liyaozhou liyaozhou         183 9月  16 22:18 hbase-1.2.12
drwxrwxr-x.  4 liyaozhou liyaozhou          37 9月  16 19:13 hdpdata
drwxrwxr-x.  2 liyaozhou liyaozhou          99 5月  17 15:15 jar
drwxr-xr-x.  8 liyaozhou liyaozhou         117 9月   1 16:19 kafka_2.12-2.2.1
drwxrwxr-x.  2 liyaozhou liyaozhou          52 8月   5 16:36 model
drwxr-xr-x. 15 liyaozhou liyaozhou        4096 9月  16 20:36 zookeeper-3.4.14
[liyaozhou@hlinkui1 lyz]$
```

1.2、软件包解压
tar -zxvf flink-1.13.2-bin-scala_2.11.tgz

```
drwxr-xr-x. 15 liyaozhou liyaozhou        4096 9月  16 20:36 zookeeper-3.4.1
[liyaozhou@hlinkui1 lyz]$ tar -zxvf flink-1.13.2-bin-scala_2.11.tgz
flink-1.13.2/
flink-1.13.2/LICENSE
flink-1.13.2/bin/
flink-1.13.2/licenses/
flink-1.13.2/plugins/
flink-1.13.2/NOTICE
flink-1.13.2/examples/
```

1.3、 配置系统环境变量
1、进入目录下
cd flink-1.13.2/

2、查看完整classpsth，然后复制
pwd

 # 3、编辑系统变量
sudo vim /etc/profile

#4、配置变量环境
export FLINK_HOME=/home/liyaozhou/lyz/flink-1.13.2
export PATH=$PATH:$FLINK_HOME/bin

#5 刷新系统变量环境
source /etc/profile

#6 查看是否配置成功
$FLINK_HOME

```
[liyaozhou@hlinkui1 lyz]$ clear
[liyaozhou@hlinkui1 lyz]$ cd flink-1.13.2/        1、进入目录下
[liyaozhou@hlinkui1 flink-1.13.2]$ pwd
/home/liyaozhou/lyz/flink-1.13.2        2、复制
[liyaozhou@hlinkui1 flink-1.13.2]$ sudo vim /etc/profile
[sudo] liyaozhou 的密码：        3、编辑系统变量
```

```
export FLINK_HOME=/home/liyaozhou/lyz/flink-1.13.2
export PATH=$PATH:$FLINK_HOME/bin
```

```
[liyaozhou@hlinkui1 flink-1.13.2]$ source /etc/profile
[liyaozhou@hlinkui1 flink-1.13.2]$ $FLINK_HOME
-bash: /home/liyaozhou/lyz/flink-1.13.2: 是一个目录
[liyaozhou@hlinkui1 flink-1.13.2]$
```

1.4、配置 Flink conf 文件

进入到 flink-1.13.2/conf 目录下

1.4.1 配置 flink-conf.yaml

#1. 配置 jobmanager rpc 地址
jobmanager.rpc.address: 192.168.244.129

#2. 修改 taskmanager 内存大小，可改可不改
taskmanager.memory.process.size: 2048m

#3. 修改一个 taskmanager 中对于的 taskslot 个数，可改可不改
taskmanager.numberOfTaskSlots: 4

#修改并行度，可改可不改
parallelism.default: 4

```
jobmanager.rpc.address: 192.168.244.129

# The RPC port where the JobManager is reachable.

jobmanager.rpc.port: 6123

# The total process memory size for the JobManager.
# Note this accounts for all memory usage within the JobMana

jobmanager.memory.process.size: 1600m

# The total process memory size for the TaskManager.
# Note this accounts for all memory usage within the TaskMana

taskmanager.memory.process.size: 2048m

# To exclude JVM metaspace and overhead, please, use total F
# It is not recommended to set both 'taskmanager.memory.proce
# taskmanager.memory.flink.size: 1280m
# The number of task slots that each TaskManager offers. Each

taskmanager.numberOfTaskSlots: 4

# The parallelism used for programs that did not specify and
parallelism.default: 4
```

1.4.2 配置 master

#修改主节点 ip 地址
192.168.244.129:8081

1.4.3 配置 work

#修改从节点 ip，因为是 standalone，所有主从一样
192.168.244.129

1.4.4 配置 zoo

新建 snapshot 存放的目录，在 flink-1.13.2 目录下建
```
mkdir tmp
cd tmp
mkdir zookeeper
```

#修改 conf 下 zoo.cfg 配置
```
vim zoo.cfg
```

#snapshot 存放的目录
```
dataDir=/home/liyaozhou/lyz/flink-1.13.2/tmp/zookeeper
```

#配置 zookeeper 地址
```
server.1=192.168.244.129:2888:3888
```

```
# The directory where the snapshot is stored.
dataDir=/home/liyaozhou/lyz/flink-1.13.2/tmp/zookeeper

# The port at which the clients will connect
clientPort=2181

# ZooKeeper quorum peers
server.1=192.168.244.129:2888:3888
# server.2=host:peer-port:leader-port
```
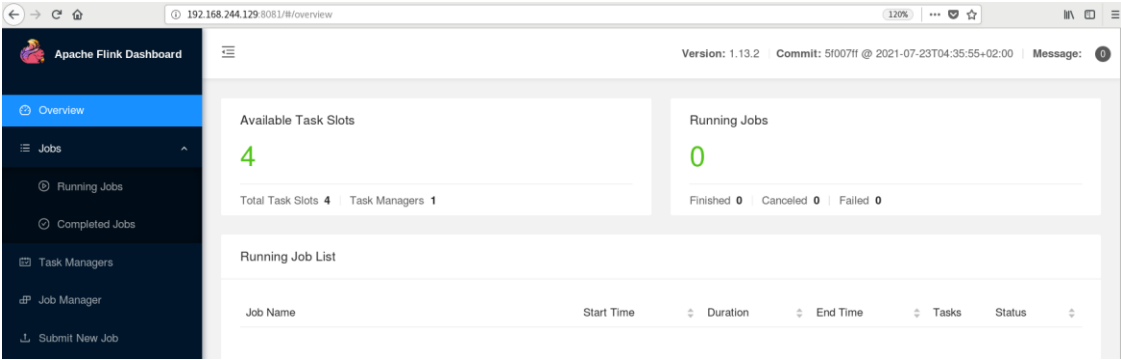
1.5、启动 Flink 集群

进入 flink-1.13.2/bin 目录下

```
./start-cluster.sh
```

2、Standalone-HA 集群部署

集群部署两节点

版本	主节点	从节点	部署方式
flink-1.13.2-bin-scala_2.11.tgz	192.168.244.129	192.168.244.130	standalone-HA
hadoop 2.6.4	192.168.244.129	192.168.244.130	Distributed
zookeeper3.4.14	192.168.244.129	192.168.244.130	Distributed

前提是 zookeeper 和 hadoop 集群全部配置好

2.1、 将软件安装包放入集群中

```
[liyaozhou@hlinkui1 lyz]$ ll
总用量 306572
drwxrwxr-x.  3 liyaozhou liyaozhou        84 6月   1 16:29 data
-rw-r--r--.  1 liyaozhou liyaozhou 313922934 9月  17 11:09 flink-1.13.2-bin-scala_2.11.tgz
drwxr-xr-x. 10 liyaozhou liyaozhou       161 9月  16 19:10 hadoop-2.6.4
drwxrwxr-x.  9 liyaozhou liyaozhou       183 9月  16 22:18 hbase-1.2.12
drwxrwxr-x.  4 liyaozhou liyaozhou        37 9月  16 19:13 hdpdata
drwxrwxr-x.  2 liyaozhou liyaozhou        99 5月  17 15:15 jar
drwxr-xr-x.  8 liyaozhou liyaozhou       117 9月   1 16:19 kafka_2.12-2.2.1
drwxrwxr-x.  2 liyaozhou liyaozhou        52 8月   5 16:36 model
drwxr-xr-x. 15 liyaozhou liyaozhou      4096 9月  16 20:36 zookeeper-3.4.14
[liyaozhou@hlinkui1 lyz]$
```

2.2、软件包解压
tar -zxvf flink-1.13.2-bin-scala_2.11.tgz

```
drwxr-xr-x. 15 liyaozhou liyaozhou      4096 9月   16 20:36 zookeeper-3.4.1
[liyaozhou@hlinkui1 lyz]$ tar -zxvf flink-1.13.2-bin-scala_2.11.tgz
flink-1.13.2/
flink-1.13.2/LICENSE
flink-1.13.2/bin/
flink-1.13.2/licenses/
flink-1.13.2/plugins/
flink-1.13.2/NOTICE
flink-1.13.2/examples/
```

2.3、 配置系统环境变量
1、进入目录下
cd flink-1.13.2/

2、查看完整 classpsth，然后复制
pwd

# 3、编辑系统变量
sudo vim /etc/profile

#4、配置变量环境
export FLINK_HOME=/home/liyaozhou/lyz/flink-1.13.2
export PATH=$PATH:$FLINK_HOME/bin

#5、添加 hadoop_conf classpath
export HADOOP_CONF_DIR=/home/liyaozhou/lyz/hadoop-2.6.4/etc/hadoop

#6 刷新系统变量环境
source /etc/profile

#7 查看是否配置成功
$FLINK_HOME

```
[liyaozhou@hlinkui1 lyz]$ clear
[liyaozhou@hlinkui1 lyz]$ cd flink-1.13.2/          1、进入目录下
[liyaozhou@hlinkui1 flink-1.13.2]$ pwd
/home/liyaozhou/lyz/flink-1.13.2                    2、复制
[liyaozhou@hlinkui1 flink-1.13.2]$ sudo vim  /etc/profile
[sudo] liyaozhou 的密码: █                          3、编辑系统变量
```

```
export FLINK_HOME=/home/liyaozhou/lyz/flink-1.13.2█
export PATH=$PATH:$FLINK_HOME/bin
```

```
[liyaozhou@hlinkui1 flink-1.13.2]$ source /etc/profile
[liyaozhou@hlinkui1 flink-1.13.2]$ $FLINK_HOME
-bash: /home/liyaozhou/lyz/flink-1.13.2: 是一个目录
[liyaozhou@hlinkui1 flink-1.13.2]$ █
```

2.4、配置 Flink conf 文件

进入到 flink-1.13.2/conf 目录下

2.4.1 配置 flink-conf.yaml

```
#1．配置 jobmanager rpc 地址
jobmanager.rpc.address: 192.168.244.129

#2．修改 taskmanager 内存大小，可改可不改
taskmanager.memory.process.size: 2048m

#3．修改一个 taskmanager 中对于的 taskslot 个数，可改可不改
taskmanager.numberOfTaskSlots: 4

#4．修改并行度，可改可不改
parallelism.default: 4

#5．配置状态后端存储方式
state.backend:filesystem

#6．配置启用检查点，可以将快照保存到 HDFS
state.backend.fs.checkpointdir: hdfs://192.168.244.129:9000/flink-check
```

points
#7. 配置保存点，可以将快照保存到 HDFS
state.savepoints.dir: hdfs:*//192.168.244.129:9000/flink-savepoints*

#8. 使用 zookeeper 搭建高可用
high-availability: zookeeper

#9. 配置 ZK 集群地址
high-availability.zookeeper.quorum: 192.168.**244**.129:2181

#10. 存储 JobManager 的元数据到 HDFS
high-availability.storageDir: hdfs:*//192.168.244.129:9000/flink/ha/*

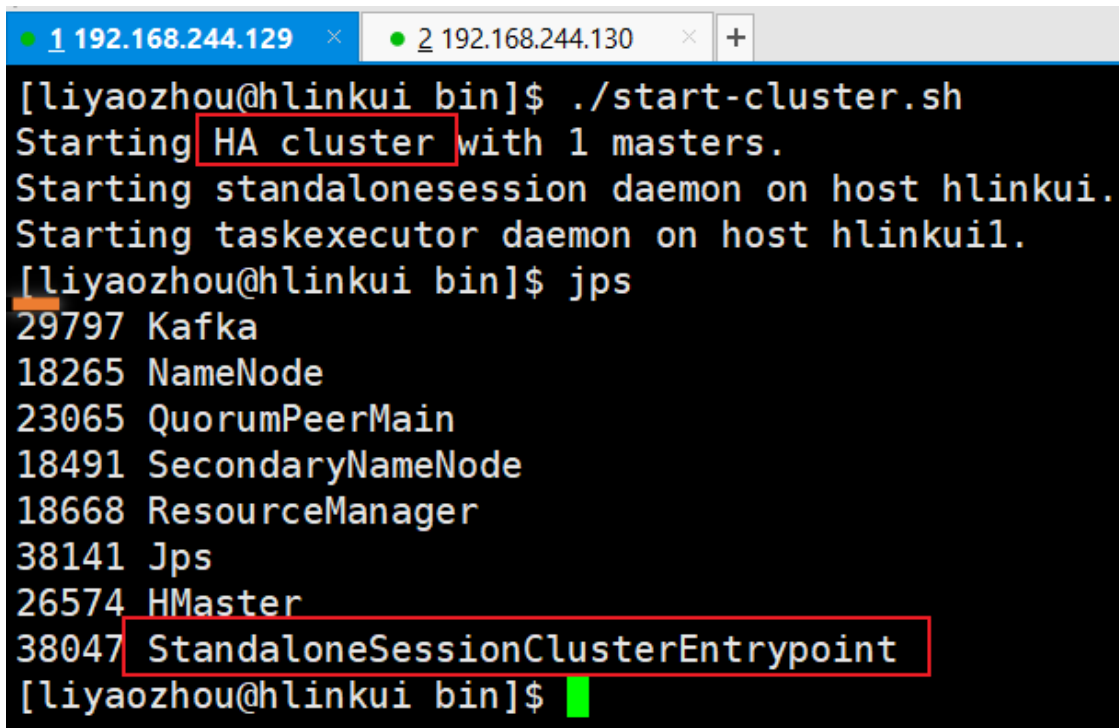**#11.** 配置 zookeeper client 默认是 open，如果 zookeeper security 启用了更改成 creator
high-availability.zookeeper.client.acl: open

```
jobmanager.rpc.address: 192.168.244.129

# The RPC port where the JobManager is reachable.

jobmanager.rpc.port: 6123

# The total process memory size for the JobManager.
# Note this accounts for all memory usage within the JobMana

jobmanager.memory.process.size: 1600m

# The total process memory size for the TaskManager.
# Note this accounts for all memory usage within the TaskMana

taskmanager.memory.process.size: 2048m

# To exclude JVM metaspace and overhead, please, use total Fl
# It is not recommended to set both 'taskmanager.memory.proce
# taskmanager.memory.flink.size: 1280m
# The number of task slots that each TaskManager offers. Eac

taskmanager.numberOfTaskSlots: 4

# The parallelism used for programs that did not specify and
parallelism.default: 4
```
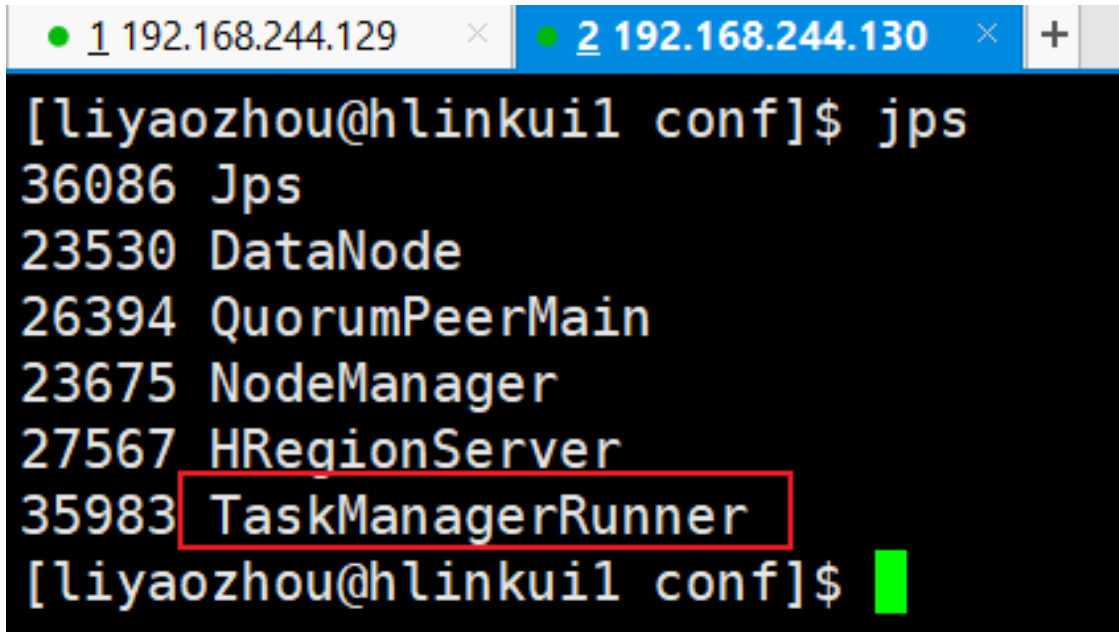
```
state.backend: filesystem
state.backend.fs.checkpointdir: hdfs://192.168.244.129:9000/flink-checkpoints
state.savepoints.dir: hdfs://192.168.244.129:9000/flink-savepoints
high-availability: zookeeper
high-availability.zookeeper.quorum: 192.168.244.129:2181
high-availability.storageDir: hdfs://192.168.244.129:9000/flink/ha/
high-availability.zookeeper.client.acl: open
```

2.4.2 配置 master

#修改主节点 ip 地址
192.168.**244**.129:8081

2.4.3 配置 work

#修改从节点 ip，因为是 standalone-ha，改另一个节点
192.168.**244**.130

2.4.4 配置 zoo

新建 snapshot 存放的目录，在 flink-1.13.2 目录下建
mkdir tmp
cd tmp
mkdir zookeeper

#修改 conf 下 zoo.cfg 配置
vim zoo.cfg

#snapshot 存放的目录
dataDir=/home/liyaozhou/lyz/flink-1.13.**2**/tmp/zookeeper

#配置 zookeeper 地址
server.1=192.168.**244**.129:2888:3888

```
# The directory where the snapshot is stored.
dataDir=/home/liyaozhou/lyz/flink-1.13.2/tmp/zookeeper

# The port at which the clients will connect
clientPort=2181

# ZooKeeper quorum peers
server.1=192.168.244.129:2888:3888
# server.2=host:peer-port:leader-port
```

2.5、下载 hadoop 依赖包

下载地址：https://flink.apache.org/downloads.html#additional-components

将包复制到 flink-1.13.2/lib 目录下

```
[liyaozhou@hlinkui lib]$ ll
总用量 233148
-rw-r--r--. 1 liyaozhou liyaozhou      92314 9月  17 11:53 flink-csv-1.13.2.jar
-rw-r--r--. 1 liyaozhou liyaozhou 115016309 9月  17 11:53 flink-dist_2.11-1.13.2.jar
-rw-r--r--. 1 liyaozhou liyaozhou     148126 9月  17 11:53 flink-json-1.13.2.jar
-rw-r--r--. 1 liyaozhou liyaozhou  36309656 9月  17 13:56 flink-shaded-hadoop-2-uber-2.6.5-10.0.jar
-rw-r--r--. 1 liyaozhou liyaozhou   7709740 9月  17 11:53 flink-shaded-zookeeper-3.4.14.jar
-rw-r--r--. 1 liyaozhou liyaozhou  36420572 9月  17 11:53 flink-table_2.11-1.13.2.jar
-rw-r--r--. 1 liyaozhou liyaozhou  40981118 9月  17 11:53 flink-table-blink_2.11-1.13.2.jar
-rw-r--r--. 1 liyaozhou liyaozhou     67114 9月  17 11:53 log4j-1.2-api-2.12.1.jar
-rw-r--r--. 1 liyaozhou liyaozhou    276771 9月  17 11:53 log4j-api-2.12.1.jar
-rw-r--r--. 1 liyaozhou liyaozhou   1674433 9月  17 11:53 log4j-core-2.12.1.jar
-rw-r--r--. 1 liyaozhou liyaozhou     23518 9月  17 11:53 log4j-slf4j-impl-2.12.1.jar
[liyaozhou@hlinkui lib]$
```

2.6、 文件传输
将主节点 flink 包复制到从节点
scp -r flink-1.13.2 192.168.244.130:/home/liyaozhou/lyz/

```
[liyaozhou@hlinkui flink-1.13.2]$ cd ..
[liyaozhou@hlinkui lyz]$ scp -r flink-1.13.2 192.168.244.130:/home/liyaozhou/lyz/
LICENSE
bash-java-utils.jar
mesos-appmaster-job.sh
stop-zookeeper-quorum.sh
mesos-appmaster.sh
historyserver.sh
flink-daemon.sh
flink
```

修改从节点 flink-conf.yaml rpc 的 ip 地址

```
jobmanager.rpc.address: 192.168.244.130

# The RPC port where the JobManager is read

jobmanager.rpc.port: 6123

# The total process memory size for the Job
```

2.7、启动 Flink 集群

进入 flink-1.13.2/bin 目录下

./start-cluster.sh

可以在登录界面看到，TaskManager 的地址为 192.168.244.130
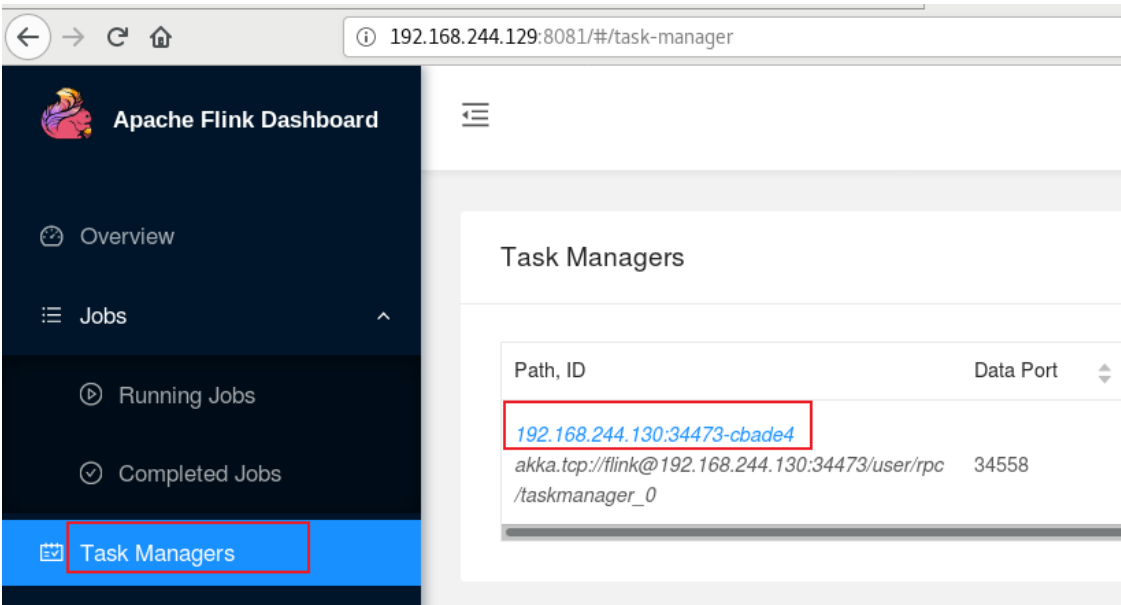
3、Flink On Yarn 集群部署

集群部署两节点

版本	主节点	从节点	部署方式
flink-1.13.2-bin-scala_2.11.tgz	192.168.244.129	192.168.244.130	yarn
hadoop 2.6.4	192.168.244.129	192.168.244.130	Distributed
zookeeper3.4.14	192.168.244.129	192.168.244.130	Distributed

前提是 zookeeper 和 hadoop 集群全部配置好

3.1 修改 Hadoop 集群的 yarn-site.xml 文件

YARN 模式下的 HA 需要注意一点，官方给出建议，必须要增加以下两项配置：YARN 配置，修改 yarn-site.xml

```
<!-- master（JobManager）失败重启的最大尝试次数-->
<property>
  <name>yarn.resourcemanager.am.max-attempts</name>
  <value>4</value>
  <description>
    The maximum number of application master execution attempts.
  </description>
</property>

<!-- 关闭yarn 内存检查 -->
<!-- 是否启动一个线程检查每个任务正使用的虚拟内存量，如果任务超出分配值，则直接将其杀掉，默认为 true -->
```

<!-- 因为对于 flink 使用 yarn 模式下，很容易内存超标，这个时候 yarn 会自动杀掉 job，因此需要关掉-->

```xml
<property>
    <name>yarn.nodemanager.pmem-check-enabled</name>
    <value>false</value>
</property>

<property>
    <name>yarn.nodemanager.vmem-check-enabled</name>
    <value>false</value>
</property>
```

3.2 修改 flink conf 配置

在 flink-conf.yaml 中添加如下两项：

```
#用户提交作业失败时，重新执行次数
yarn.application-attempts: 4

#设置 Task 在所有节点平均分配
cluster.evenly-spread-out-slots: true
```

```
state.backend: filesystem
state.backend.fs.checkpointdir: hdfs://192.168.244.129:9000/flink-checkpoints
state.savepoints.dir: hdfs://192.168.244.129:9000/flink-savepoints
high-availability: zookeeper
high-availability.zookeeper.quorum: 192.168.244.129:2181
high-availability.storageDir: hdfs://192.168.244.129:9000/flink/ha/
high-availability.zookeeper.client.acl: open


#用户提交作业失败时，重新执行次数
yarn.application-attempts: 4

#设置Task在所有节点平均分配
cluster.evenly-spread-out-slots: true
```

3.3 启动测试（Session 模式）

3.3.1 启动 Flink 会话（在 192.168.244.129 上测试）

```
# 主节点中执行
bin/yarn-session.sh -d -jm 1024 -tm 1024 -s 8

# -tm 表示每个 TaskManager 的内存大小
# -s 表示每个 TaskManager 的 slots 数量
# -d 表示以后台程序方式运行
```

```
[liyaozhou@hlinkui flink-1.13.2]$ clear
[liyaozhou@hlinkui flink-1.13.2]$ bin/yarn-session.sh -d -jm 1024 -tm 1024 -s 1
2021-09-17 15:22:36,064 INFO  org.apache.flink.configuration.GlobalConfiguration
68.244.129
2021-09-17 15:22:36,067 INFO  org.apache.flink.configuration.GlobalConfiguration
2021-09-17 15:22:36,067 INFO  org.apache.flink.configuration.GlobalConfiguration
e, 1600m
2021-09-17 15:22:36,067 INFO  org.apache.flink.configuration.GlobalConfiguration
```

3.3.2 登录 yarn 集群页面查看

登录网址：192.168.244.129:8088/cluster

3.3.3 在 yarn 上提交任务 通过 session 模式

注意：此时提交的任务都通过该会话（Session）执行，不会再申请 yarn 资源

（1）创建一个 wordcount.txt 文本，随便早一些数据，然后放到 flink-1.13.2 下面，然后将该文件传到 hdfs 中

```
hadoop  fs  -copyFromLocal  wordcount.txt  /
```

（2）提交任务

192.168.244.129 中执行即可
bin/flink run examples/batch/WordCount.jar --input hdfs://192.168.244.1
29:9000/wordcount.txt

3.3.3 查看 Hadoop 的 ApplicationManager 的 WEB-UI 页面

3.3.4 关闭 Session 模式

yarn application -kill application_1631862788541_0001

3.4 启动测试（Per-job 模式）

3.4.1 直接提交 Job

```
# -m jobmanager 的地址
# -yjm 1024 指定 jobmanager 的内存信息
# -ytm 1024 指定 taskmanager 的内存信息
bin/flink run \
-t yarn-per-job -yjm 1024 -ytm 1024  \
--detached  examples/batch/WordCount.jar  \
--input hdfs://192.168.244.129:9000/wordcount.txt
```

以上就是 Flink 集群安装的讲解内容！觉得好的，点赞，在看，分享三连击，谢谢！！！ 最近整理了一份计算机类的书籍，包含 python、java、大数据、人工智能、算法等，种类特别齐全。

获取方式：关注公众号：**3 分钟秒懂大数据**，回复：**福利**，就可以获得这份超级大礼！

微信搜一搜

3分钟秒懂大数据

扫码加入Flink流计算群
群若过期，加博主微信，拉你进群

3分钟秒懂大数据

Flink技术交流群

扫一扫上面的二维码图案，加我微信

该二维码7天内(9月29日前)有效，重新进入将更新