

实验环境搭建——Storm

软件准备：Storm 1.2.2 安装包

环境：Python 2.7, Hadoop 2.8.5, ZooKeeper 3.4.12

1、Python 安装

首先确定系统中已安装 Python 2.6 以上，通过下述命令来验证版本。

```
$ python --version
```

若 python 未安装，可以使用命令一键安装。

```
$ sudo apt install python
```

2、Hadoop + Zookeeper 安装

详见 Hadoop 完全分布

3、安装 Storm

Storm 的安装方式与 Hadoop, ZooKeeper 等类似，直接解压 Storm 安装包

```
$ tar zxvf <storm 安装包>
```

然后把解压的整个目录移到指定位置。

4、配置 Storm

Storm 默认配置就是单机版配置，所以单机版 Storm 部署，无需修改任何配置。

分布式配置方式需要先配置 ZooKeeper，ZooKeeper 的配置方式见 Hadoop 完全分布。配置完 ZooKeeper 并保证能够正常启动。

Storm 需要配置两个文件 storm-env.sh 和 storm.yaml

storm-env.sh 的配置如下，设置 JAVA_HOME 的绝对路径。

```
# The java implementation to use.
export JAVA_HOME=/usr/local/jdk1.8.0_181
```

storm.yaml 的配置内容如下：

```
##### These MUST be filled in for a storm configuration
storm.zookeeper.servers:
  - "hadoop1"
  - "hadoop2"
  - "hadoop3"

nimbus.seeds: ["hadoop1", "hadoop2"]

storm.local.dir: "/home/hadoop/bigdata/storm-1.2.2/local"
storm.log.dir: "/home/hadoop/bigdata/storm-1.2.2/logs"

worker.childopts: "-Xmx1024m"

supervisor.slots.ports:
  - 6700
  - 6701
  - 6702
  - 6703
```

- i. storm.zookeeper.servers 表示配置 Zookeeper 集群地址。如果 zookeeper 集群中使用的不是默认端口，则还需要配置 storm.zookeeper.port
- ii. nimbus.seeds 配置了主控节点，可以有多个
- iii. Nimbus 和 Supervisor 守护进程需要在本地存储少量状态，storm.local.dir 定义了目录位置
- iv. workder.childopts 指定 Storm 启动 worker 进程时 JVM 的最大内存
- v. supervisor.slots.ports 用来设置该节点运行 worker 的个数，和分别的端口，每个 worker 使用一个端口接收

```
## Locations of the drpc servers
drpc.servers:
  - "hadoop1"
#   - "server2"
```

- vi. 设置 drpc server 为 hadoop1 节点

部署完成以后，在 storm 安装目录下新建 local 和 logs 目录。

复制配置好的 storm 到其他两台主机

```
$ scp -r storm-1.2.2/ hadoop2:/home/hadoop/bigdata/
```

```
$ scp -r storm-1.2.2/ hadoop3:/home/hadoop/bigdata/
```

5、测试

Storm 集群中包含了两类节点，主控节点（Master Node）就是我们配置的（hadoop1 和 hadoop2）和工作节点（Work Node）（3 台主机），在启动 Storm 时首先启动主控节点，其次启动工作节点。

启动主控节点服务（hadoop1，hadoop2）

```
$ bin/storm nimbus >> logs/nimbus.out 2>&1 &
```

启动主控节点 UI （hadoop1，hadoop2）

```
$ bin/storm ui >> logs/ui.out 2>&1 &
```

启动工作节点（hadoop1，hadoop2，hadoop3）

```
$ bin/storm supervisor >> logs/supervisor.out 2>&1 &
```

启动 logviewer（hadoop1，hadoop2，hadoop3）

```
$ bin/storm logviewer >> logs/logviewer.out 2>&1 &
```

验证是否启动完成，使用 jps 命令查看，访问 <http://hadoop1:8080>

启动成功后，输入网址可以看到 storm 的启动信息。

The screenshot shows the Storm UI web interface in a browser. The address bar shows 'hadoop1:8080/index.html'. The page title is 'Storm UI'. There are two main sections: 'Cluster Summary' and 'Nimbus Summary'.

Cluster Summary

Version	Supervisors	Used slots	Free slots	Total slots	Executors	Tasks
1.2.2	3	0	12	12	0	0

Nimbus Summary

Host	Port	Status	Version	UpTime
hadoop1	6627	Leader	1.2.2	42m 46s
hadoop2	6627	Not a Leader	1.2.2	42m 5s

Showing 1 to 2 of 2 entries

Topology Summary

Name	Owner	Status	Uptime	Num workers	Num executors	Num tasks	Replication count	Assigned Mem (MB)	Scheduler Info
wordcount	hadoop	ACTIVE	7s	0	0	0	2	0	

Showing 1 to 1 of 1 entries

运行 wordcount 的例子，进入 examples/storm-starter 目录下，打包程序为 jar 包

```
$ mvn package
```

```
hadoop@hadoop1:~/bigdata/storm-1.2.2$ bin/storm.jar examples/storm-starter/storm-starter-1.2.2.jar org.apache.storm.starter.WordCountTopology wordcount
```

可以在网页上查看运行的日志信息。