

Ubuntu安装Flink-1.4.0

1. 下载

Flink可以运行在Linux, Mac OS X和Windows上。这边文章是在ubuntu上运行Flink, JDK的版本是1.8。Windows 用户, 请查看 Flink在Windows上的安装指南。

下载页面: <http://flink.apache.org/downloads.html>

这里flink的版本一定要和hadoop的版本匹配, 所以选择hadoop2.8的下载地址如下:

下载地址: http://www.apache.org/dyn/closer.lua/flink/flink-1.4.0/flink-1.4.0-bin-hadoop28-scala_2.11.tgz

2. 解压、安装以及配置环境变量

解压安装包, 进入conf配置文件目录下, 主要配置文件为flink-conf.yaml和slaves,配置flink-conf.yaml解析如下:

(1) 基本配置

jobmanager.rpc.address: localhost1 --jobManager 的IP地址

jobmanager.rpc.port: 6123 --jobManager 的端口, 默认为6123

jobmanager.heap.mb --jobManager 的JVM heap大小

taskmanager.heap.mb --taskManager的jvm heap大小设置

taskmanager.numberOfTaskSlots --taskManager中taskSlots个数, 最好设置成work节点的CPU个数相等

parallelism.default --并行计算数

fs.default-scheme --文件系统来源

fs.hdfs.hadoopconf --hdfs置文件路径

jobmanager.web.port -- jobmanager的页面监控端口

(2) 内存管理配置

flink默认上分配taskmanager.heap.mb配置值得70%留它管理, 内存的管理让flink批量处理效果很高; 并且flink不会出现OutMemoryException的问题, 因为flink知道预留多少内存来执行程序; 如果flink运行的程序所需要的内存超过了它所管理的内存, Flink就可以利用磁盘; 总而言之, flink的内存管理提高了鲁棒性和系统的速度; 下面就介绍管理内存的配置文件:

taskmanager.memory.fraction --管理内存的百分比, 默认0.7

taskmanager.memory.size --taskManager 具体管理内存的大小; 此配置重写

taskmanager.memory.fraction的配置

taskmanager.memory.segment-size --内存管理器所使用的内存缓冲区的大小和网络堆栈字节

taskmanager.memory.preallocate --taskmanager是否启动时管理所有的内存

(3) slaves 中配置节点机器的ip或主机名

(4) 环境变量配置

[plain]

```
01. ## Setting Flink
02. export FLINK_HOME=/usr/local/software/flink-1.4.0
03. export PATH=${FLINK_HOME}/bin:$PATH
```

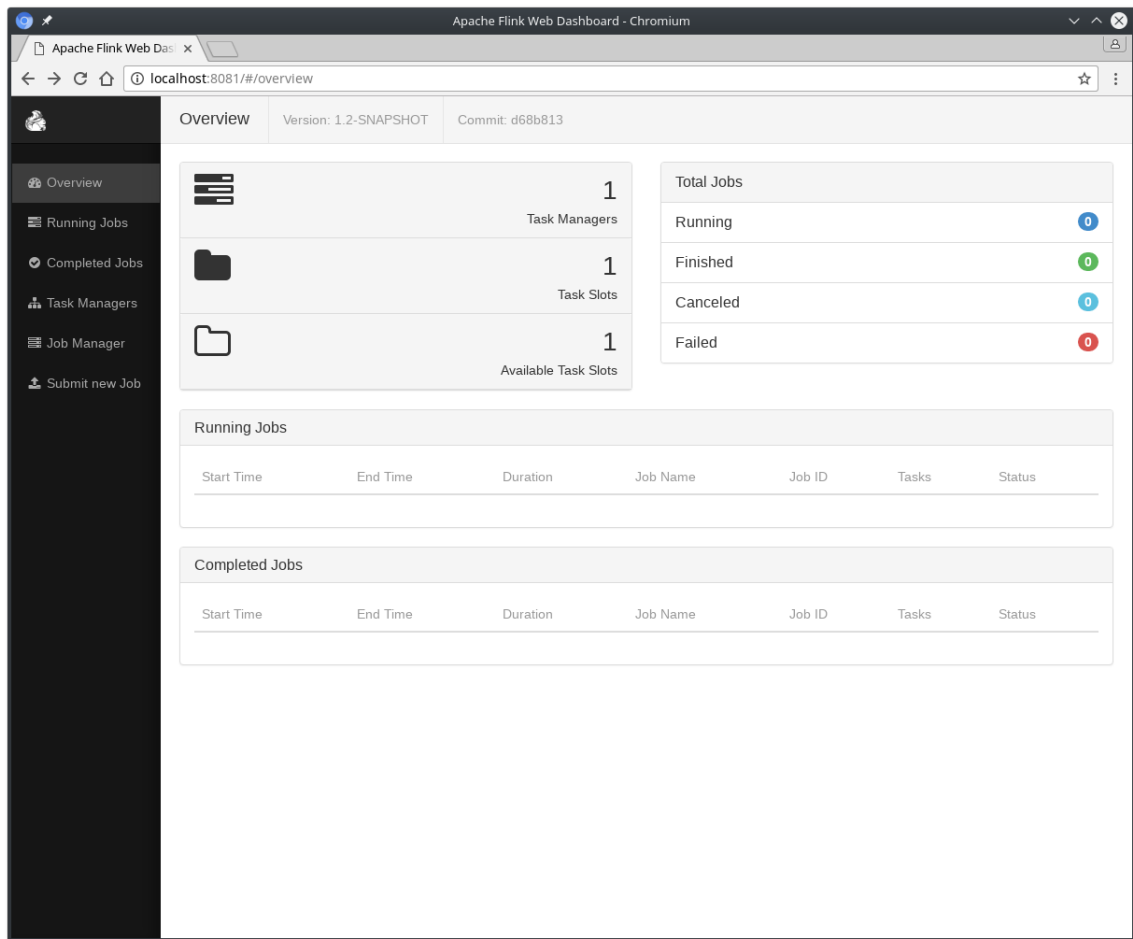
3. 启动

使用如下命令启动flink

```
[sql]
```

```
01. hadoop@s01:~/opt/flink-1.4.0$ ./bin/start-local.sh
02. Starting jobmanager daemon on host yoona.
```

通过访问 <http://localhost:8081> 检查JobManager网页,确保所有组件都启动并已运行。网页会显示一个有效的TaskManager实例。



你也可以通过检查日志目录里的日志文件来验证系统是否已经运行。

4. 运行Example

(1)首先,我们可以通过netcat命令来启动本地服务:

```
[plain]
```

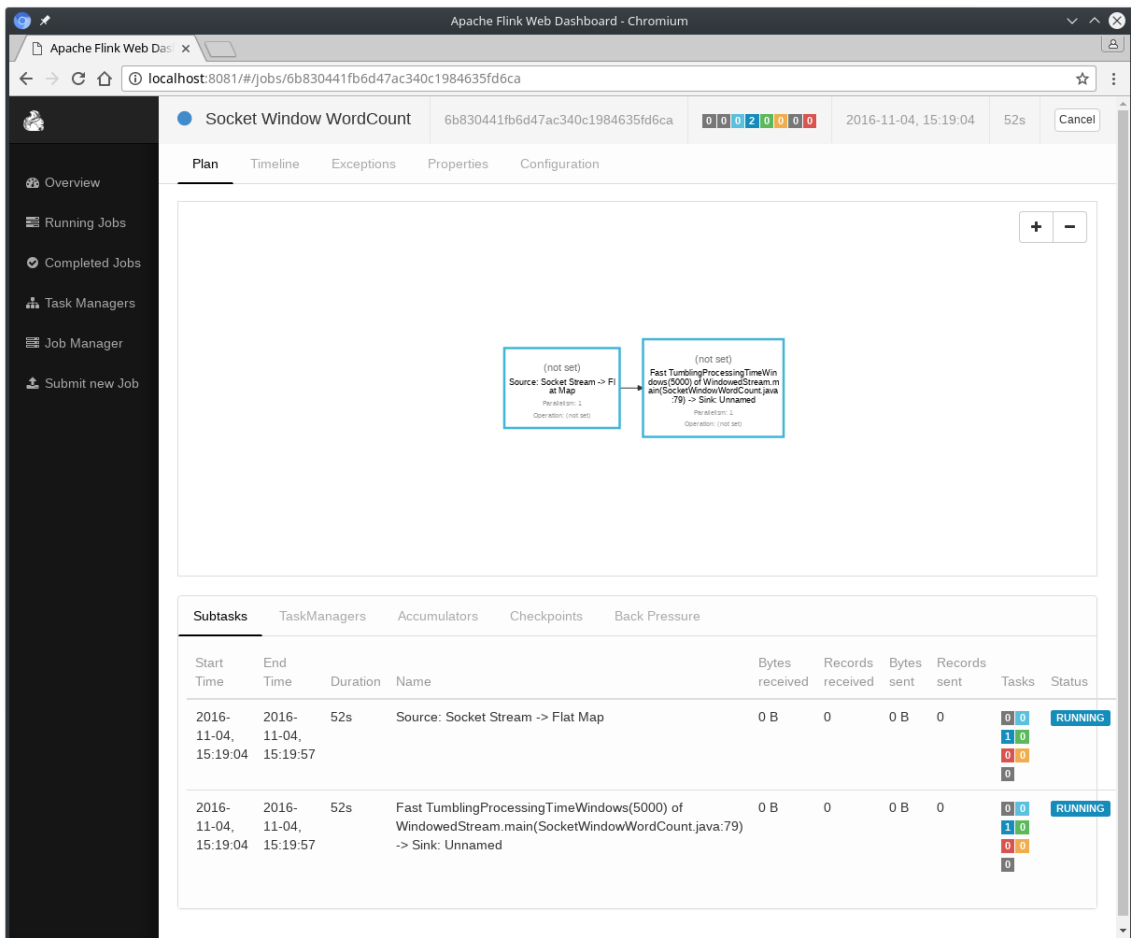
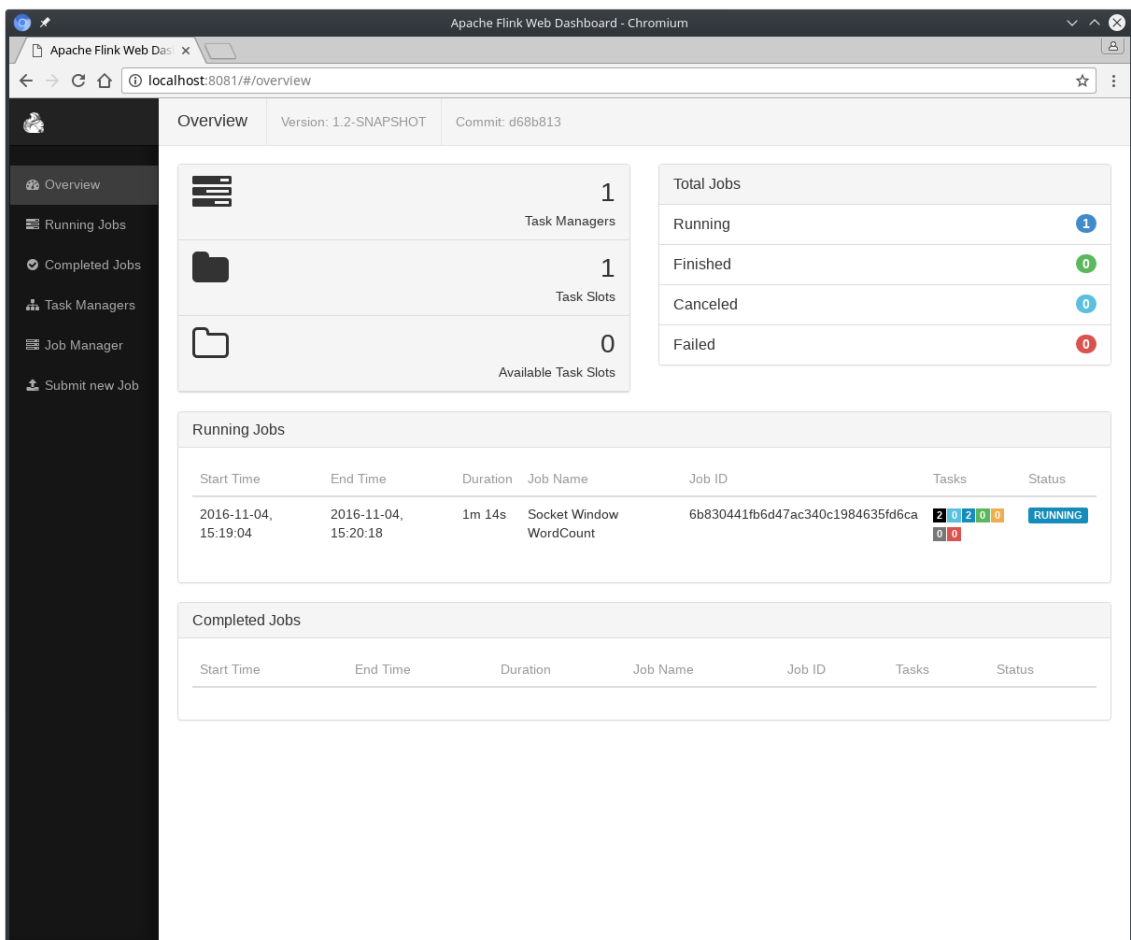
```
01. nc -l 9000
```

(2)提交Flink程序, 这里运行的是flink自带的example:

```
[plain]
```

```
01. ./bin/flink run examples/streaming/SocketWindowWordCount.jar --port 9000
```

(3)应用程序连接socket并等待输入, 你可以通过web界面来验证任务期望的运行结果:



单词的数量在5秒的时间窗口中进行累加（使用处理时间和tumbling窗口），并打印在stdout。监控JobManager的输出文件，并在nc写一些文本(回车一行就发送一行输入给Flink)：

```
hadoop@s01:/usr/local/software/flink-1.4.0$ sudo nc -l 9000
[sudo] password for hadoop:
aaa
bbb
java
the sdd
123
aaa
c++
python
scala
scala
```

.out文件将在每个时间窗口截止之际打印每个单词的个数，打开文件

`${FLINK_HOME}/log/flink-hadoop-taskmanager-0-s01.out`

```
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/software/flink-1.4.0/lib/slf4j-log4j12-1.7.7.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/software/hadoop-2.8.1/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
aaa : 1
123 : 1
sdd : 1
the : 1
java : 1
bbb : 1
aaa : 1
python : 1
c++ : 1
scala : 2
```

5.停止flink

使用以下命令来停止flink:

[plain]

```
01. ./bin/stop-local.sh
```