Apache Spark standalone cluster on Windows

Amar Gajbhiye Jun 20, 2019 · 3 min read ★

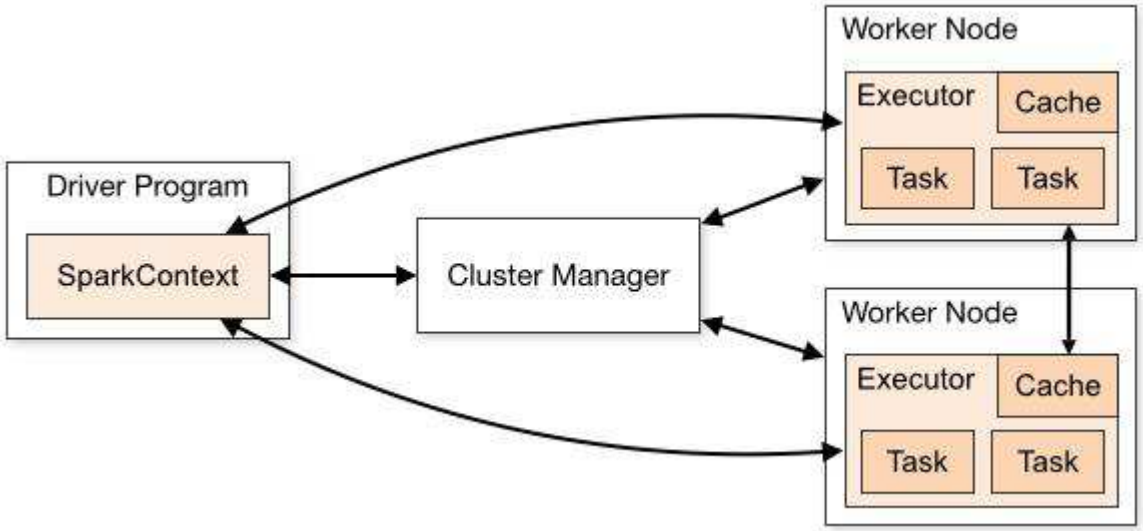
> *Apache Spark is a distributed computing framework which has built-in support for batch and stream processing of big data, most of that processing happens in-memory which gives a better performance. It has built-in modules for SQL, machine learning, graph processing, etc.*

There are two different modes in which Apache Spark can be deployed, **Local** and **Cluster** mode.

Local mode is mainly for testing purposes. In this mode, all the main components are created inside a single process. In cluster mode, the application runs as the sets of processes managed by the driver (SparkContext). The following are the main components of cluster mode.

1. Master
2. Worker
3. Resource Manager

You can visit this link for more details about cluster mode.



Spark cluster overview

Currently, Apache Spark supports **Standalone, Apache Mesos**, **YARN,** and **Kubernetes** as resource managers. Standalone is a spark's resource manager which is easy to set up which can be used to get things started fast.

There are many articles and enough information about how to start a standalone cluster on Linux environment. But, there is not much information about starting a standalone cluster on Windows.

In this article, we will see, how to start Apache Spark using a standalone cluster on the Windows platform.

Few key things before we start with the setup:

1. Avoid having spaces in the installation folder of Hadoop or Spark.
2. Always start Command Prompt with Administrator rights i.e with Run As Administrator option

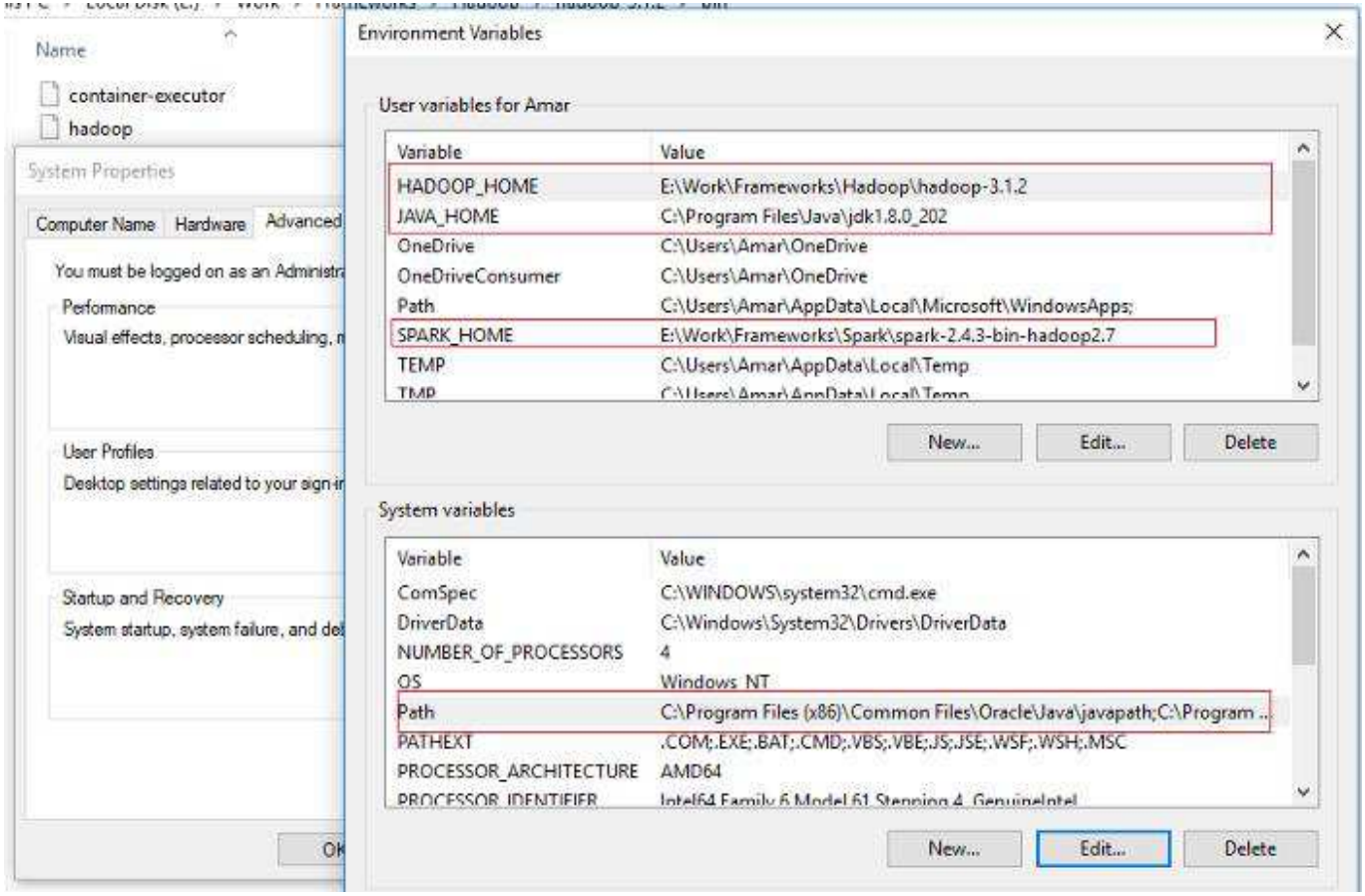
Pre-requisites

1. Download **JDK** and add JAVA_HOME = <path_to_jdk_> as an environment variable.

2. Download **Spark** and add SPARK_HOME=<path_to_spark>. If you choose to download spark pre-built with particular version of hadoop, no need to download it explicitly in step 3.

3. Download **Hadoop** and add HADOOP_HOME=<path_to_hadoop> and add **%HADOOP_HOME%\bin to PATH** variable.

4. Download winutils.exe (for the same Hadoop version as above) and place it under **%HADOOP_HOME%\bin**.

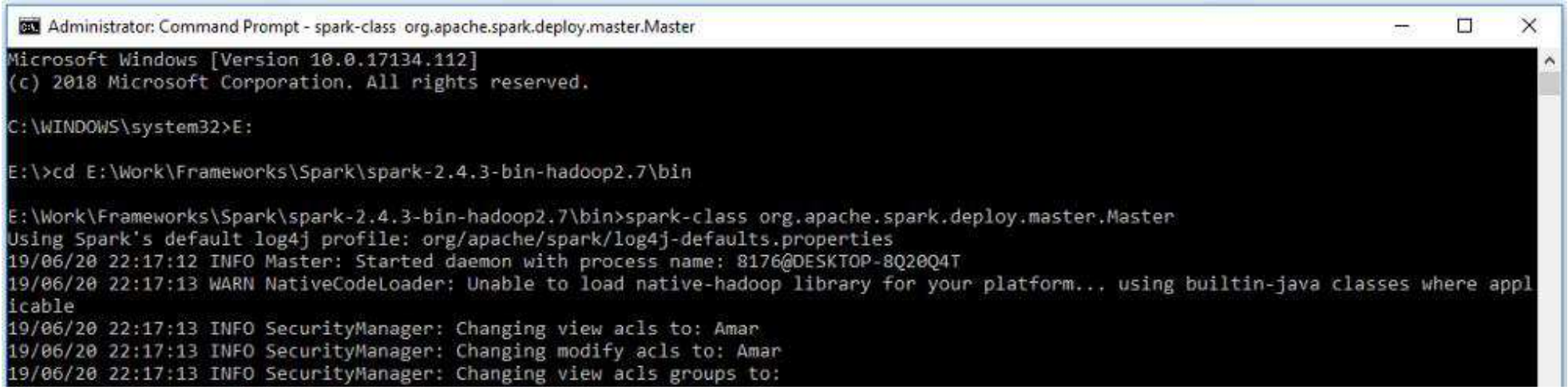


Set up Master Node

Go to spark installation folder, open Command Prompt as administrator and run the following command to start master node.

The host flag (`--host`) is optional. It is useful to specify an address specific to a network interface when multiple network interfaces are present on a machine.

```
bin\spark-class org.apache.spark.deploy.master.Master --host
<IP_Addr>
```



```
19/06/20 22:17:13 INFO SecurityManager: Changing modify acls groups to:
19/06/20 22:17:13 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users  with view permissions: Set(Am
ar); groups with view permissions: Set(); users  with modify permissions: Set(Amar); groups with modify permissions: Set()
19/06/20 22:17:14 INFO Utils: Successfully started service 'sparkMaster' on port 7077.
19/06/20 22:17:15 INFO Master: Starting Spark master at spark://192.168.0.104:7077
19/06/20 22:17:15 INFO Master: Running Spark version 2.4.3
19/06/20 22:17:15 INFO Utils: Successfully started service 'MasterUI' on port 8080.
19/06/20 22:17:15 INFO MasterWebUI: Bound MasterWebUI to 0.0.0.0, and started at http://DESKTOP-8Q20Q4T:8080
19/06/20 22:17:15 INFO Master: I have been elected leader! New state: ALIVE
```

Set up Worker Node

Follow the above steps and run the following command to start a worker node

```
bin\spark-class org.apache.spark.deploy.worker.Worker
spark://<master_ip>:<port> --host <IP_ADDR>
```

```
Administrator: Command Prompt - spark-class org.apache.spark.deploy.worker.Worker spark://192.168.0.104:7077
Microsoft Windows [Version 10.0.17134.112]
(c) 2018 Microsoft Corporation. All rights reserved.

C:\WINDOWS\system32>E:

E:\>cd C:\Windows\System32\cmd.exe
The directory name is invalid.

E:\>cd E:\Work\Frameworks\Spark\spark-2.4.3-bin-hadoop2.7\bin

E:\Work\Frameworks\Spark\spark-2.4.3-bin-hadoop2.7\bin>spark-class org.apache.spark.deploy.worker.Worker spark://192.168.0.104:7077
```

```
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
19/06/20 22:22:50 INFO Worker: Started daemon with process name: 4116@DESKTOP-8Q20Q4T
19/06/20 22:22:50 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
19/06/20 22:22:51 INFO SecurityManager: Changing view acls to: Amar
19/06/20 22:22:51 INFO SecurityManager: Changing modify acls to: Amar
19/06/20 22:22:51 INFO SecurityManager: Changing view acls groups to:
19/06/20 22:22:51 INFO SecurityManager: Changing modify acls groups to:
19/06/20 22:22:51 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users  with view permissions: Set(Amar); groups with vi
ew permissions: Set(); users  with modify permissions: Set(Amar); groups with modify permissions: Set()
19/06/20 22:22:52 INFO Utils: Successfully started service 'sparkWorker' on port 57896.
19/06/20 22:22:52 INFO Worker: Starting Spark worker 192.168.0.104:57896 with 4 cores, 6.7 GB RAM
19/06/20 22:22:52 INFO Worker: Running Spark version 2.4.3
19/06/20 22:22:52 INFO Worker: Spark home: E:\Work\Frameworks\Spark\spark-2.4.3-bin-hadoop2.7
19/06/20 22:22:52 INFO Utils: Successfully started service 'WorkerUI' on port 8081.
19/06/20 22:22:52 INFO WorkerWebUI: Bound WorkerWebUI to 0.0.0.0, and started at http://DESKTOP-8Q20Q4T:8081
19/06/20 22:22:52 INFO Worker: Connecting to master 192.168.0.104:7077...
19/06/20 22:22:52 INFO TransportClientFactory: Successfully created connection to /192.168.0.104:7077 after 47 ms (0 ms spent in bootstraps)
19/06/20 22:22:52 INFO Worker: Successfully registered with master spark://192.168.0.104:7077
```

Your standalone cluster is up with the master and one worker node. And now you can access it from your program using master as `spark://<master_ip>:<port>`.

These two instances can run on the same or different machines.

Spark UI

You can access Spark UI by using the following URL

```
http://<MASTER_IP>:8080
```

Spark UI

If you like this article, check out similar articles here https://www.bugdbug.com

Feel free to share your thoughts, comments.

If you find this article helpful, share it with a friend!