



Usage instructions:

Launch the product via 1-click. **Please wait until** the instance passes all status checks and is running. You can connect using your Amazon private key and **'ubuntu'** login via SSH client.

- To update software, use: **sudo apt update** and **sudo apt upgrade**

1. To connect to Scala, run the following command:

scala

2. Next run a test. Once inside the **scala>(prompt)** use:

```
println("Testing Scala")
```

You should get the following output:

Testing Scala

3. Press **CTRL+D** to exit from the Scala shell.

To verify Spark Installation, at the **ubuntu prompt** use the following command:

spark-shell

You should see the following output:

```
ubuntu@ip-10-0-0-21:~$ spark-shell
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogL
23/02/03 18:38:57 WARN NativeCodeLoader: Unable to load native-hadoop library
Spark context Web UI available at http://ip-10-0-0-21.ec2.internal:4040
Spark context available as 'sc' (master = local[*], app id = local-167544953)
Spark session available as 'spark'.
Welcome to

      /\_/\_/_/_/_/_/_\_/_/_/_/_\_/_/_/_/_\
     /  V  \_/  _/_/_/_/_/_/_/_/_/_/_/_/_/_\ version 3.3.1
    /_____\//__/_/_/_/_/_/_/_/_/_/_/_/_/_/_\

Using Scala version 2.12.15 (OpenJDK 64-Bit Server VM, Java 11.0.17)
Type in expressions to have them evaluated.
Type :help for more information.
```

Starting Spark

- To start Apache Spark run:

start-master.sh

- To start the Apache Spark Worker service, use the following command:

start-worker.sh spark://your-public-instance-ip:7077

For ex: **start-worker.sh spark://18.206.146.251:7077**

Open your web browser and access the Apache Spark master node using your instance Public IPv4 address.

For ex: **http://323.325.6:8080**

 **Spark Master at spark://ip-10-0-0-54.ec2.internal:7077**

URL: spark://ip-10-0-0-54.ec2.internal:7077
Alive Workers: 1
Cores in use: 2 Total, 0 Used
Memory in use: 2.8 GiB Total, 0.0 B Used
Resources in use:
Applications: 0 Running, 0 Completed
Drivers: 0 Running, 0 Completed
Status: ALIVE

▼ Workers (1)

Worker Id	Address	State	Cores	Memory
worker-20230203200002-10.0.0.54-32929	10.0.0.54:32929	ALIVE	2 (0 Used)	2.8 GiB (0.0 B Used)

If you want to stop the Apache Spark, run the following command:

stop-master.sh

sudo systemctl status spark-master

sudo systemctl status spark-worker

AWS Data

- Data Encryption Configuration: This solution does not encrypt data within the running instance.
- User Credentials are stored: /root/.ssh/authorized_keys & /home/ubuntu/.ssh/authorized_keys
- Monitor the health:
 - Navigate to your Amazon EC2 console and verify that you're in the correct region.
 - Choose Instance and select your launched instance.
 - Select the server to display your metadata page and choose the Status checks tab at the bottom of the page to review if your status checks passed or failed.

Extra Information: (Optional)

Allocate Elastic IP

To ensure that your instance **keeps its IP during restarts** that might happen, configure an Elastic IP. From the EC2 console:

1. Select ELASTIC IPs.
2. Click on the ALLOCATE ELASTIC IP ADDRESS.
3. Select the default (Amazon pool of IPv4 addresses) and click on ALLOCATE.
4. From the ACTIONS pull down, select ASSOCIATE ELASTIC IP ADDRESS.
5. In the box that comes up, note down the Elastic IP Address, which will be needed when you configure your DNS.
6. In the search box under INSTANCE, click and find your INSTANCE ID and then click ASSOCIATE.
7. Your instance now has an elastic IP associated with it.
8. For additional help: <https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/elastic-ip-addresses-eip.html>