



Home



My Network



Jobs



Run Apache Spark without Cloudera VM or any public cloud service like Dataproc, EMR or HDInsight

Published on September 9, 2021



Archit Soni

Senior Consultant | Data Engineer | Spark | Scala | Hadoop | Python | Azure Certified

2 articles

[+ Follow](#)

You saw it right, you can practice Spark without having the headache of running an Oracle Virtual Box, neither you have to sign up for yet another free trial on GCP/AWS/Azure to get your hands dirty in Spark.

This can be achieved by just following some boilerplate code in **Google Colab** and within minutes you have your Spark sandbox ready for rocking!



Sumit M



Master Big Data | Data Engineerin...
youtube.com

Google Co
colab.research.goo

Write a message...

Run Google account

```
# import Spark library from apache.org.spark
!wget https://www.apache.org/dyn/closer.lua/spark/spark-3.1.2-bin-hadoop3.2.tgz
# untar the file you just downloaded
!tar -xvf spark-3.1.2-bin-hadoop3.2.tgz
# check if the file is unzipped
!ls
# Google Colab has Java 11 available, test it using below
!ls /usr/lib/jvm
# now we need to setup environment variables for our Spark
import os
os.environ["JAVA_HOME"] = "/usr/lib/jvm/java-11-openjdk-amd64"
os.environ["SPARK_HOME"] = "/content/spark-3.1.2-bin-hadoop3.2"
# install findspark package
!pip install findspark
# time to initiate a Spark Session
import findspark
findspark.init()
from pyspark.sql import SparkSession
spark = SparkSession.builder.master("local[*]").getOrCreate()
# lets see Spark UI, for this we need to do few more steps
# Install ngrok
!wget https://bin.equinox.io/c/4VmDzA7iaHb/ngrok-stable-linux-amd64.zip
!unzip ngrok-stable-linux-amd64.zip
# to get public URL
get_ipython().system_raw('./ngrok http 4050 &')
# wait for 15 seconds for it to host the Spark UI
import time
time.sleep(15)
```

Sumit M

Hope you liked this short but informative article!

Report this

Published by



Archit Soni

Senior Consultant | Data Engineer | Spark | Scala | Hadoop | Python | Azure Certified
Published • 1h

2

articles

[+ Follow](#)

== Get going with an Apache Spark sandbox in less than 5 mins for FREE ==

With the increasing interest in distributed computing and Spark, you have a number of options to get started - you can download a Cloudera VM or you can signup for AWS, Azure or GCP and use their services to spin a cluster and start following the tutorials.

These are great options but more often than not, they are overkill when we are just learning the nuances of technology, I don't care how many executors are working on our Spark job, while we are just dealing with just a few rows to do some transformations and displaying the final result on the console.

Read the article to get the bare-bones to get started in Spark.



Tell me in comments, how many executors will this sandbox have ;)

[#spark](#) [#colab](#) [#big](#)






Sumit M





HomeMy NetworkJobs

Reactions



+5





1 Comment


Most relevant ▾



Master Big Data | Data Engineerin...
youtube.com


Thanks

Write a message...

Send

Add a comment...




**Sumit M** • Following55m ...

Founder & CEO of Trendytech | #1 Big Data Trainer | Ex-Cisco | Ex-VMware | MCA @ NIT Trichy | #SumitTeaches | New Batch Starting on 4th September

very insightful

Like | Reply





Archit Soni

Senior Consultant | Data Engineer | Spark | Scala | Hadoop | Python| Azure Certified

+ Follow

More from Archit Soni

**Apache Spark - Pinpoint Skewness in partitions - Code snippet**
Archit Soni on

Sumit M

in

Home

My Network

Jobs

About

Community Guidelines

Privacy & Terms

Sales Solutions

Safety Center

Accessibility

Careers

Ad Choices

Mobile

LinkedIn Corporation © 2021

Master Big Data | Data Engineerin...

youtube.com

Thanks

Si

Write a message...

Send

Questions?

our Help Center.

Manage your account

privacy

o your Settings.

Select Language

English (English)

Sumit M