











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

Published on September 9, 2021



Archit Soni

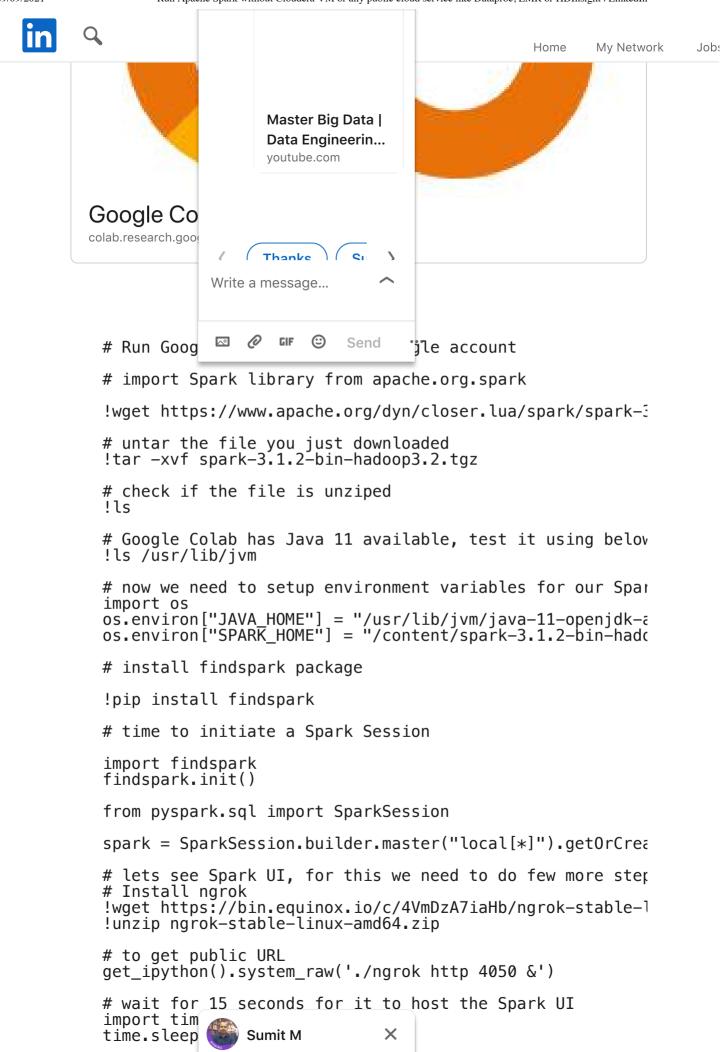
Senior Consultant | Data Engineer | Spark | Scala | Hadoop | Python| Azure

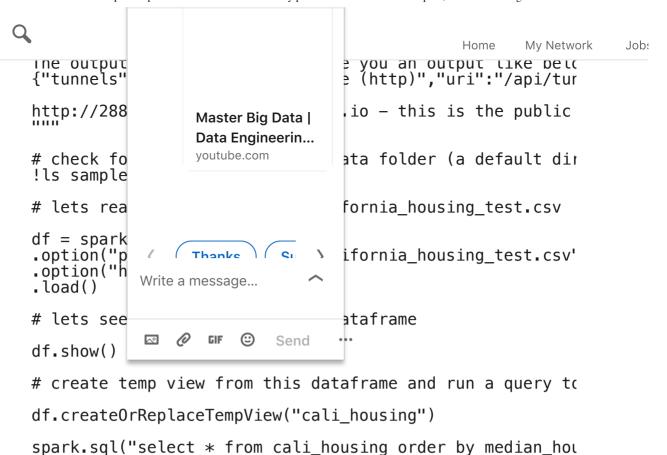
2 articles



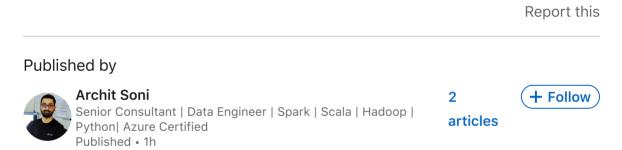
You saw it right, you can practice Spark without having the headache of running an Oracle Virtual Box, neither you have to sign up for yet another free trial on GCP/AWS/Azure to get your hands dirty in Spark.

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





Hope you liked this short but informative article!



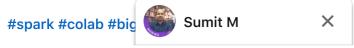
== Get going with an Apache Spark sandbox in less than 5 mins for FREE ==

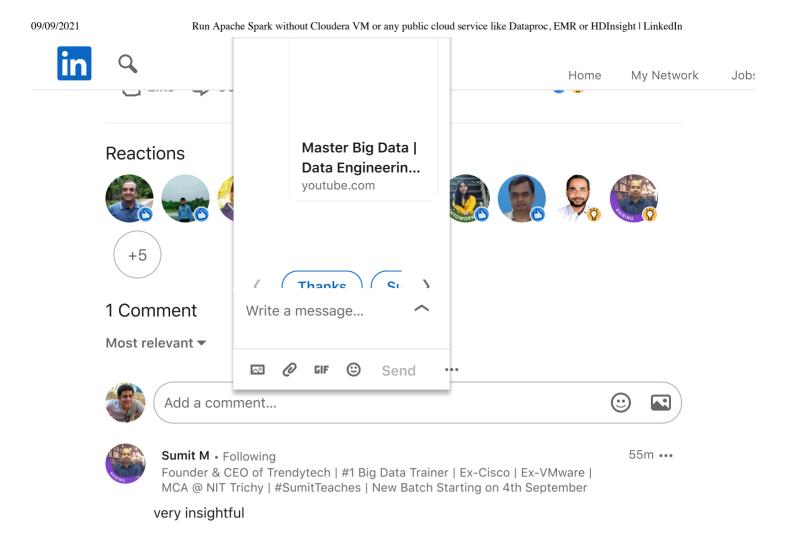
With the increasing interest in distributed computing and Spark, you have a number of options to get started - you can download a Cloudera VM or you can signup for AWS, Azure or GCP and use their services to spin a cluster and start following the tutorials.

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 ;)





Archit Soni

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



More from Archit Soni

Like

Reply

