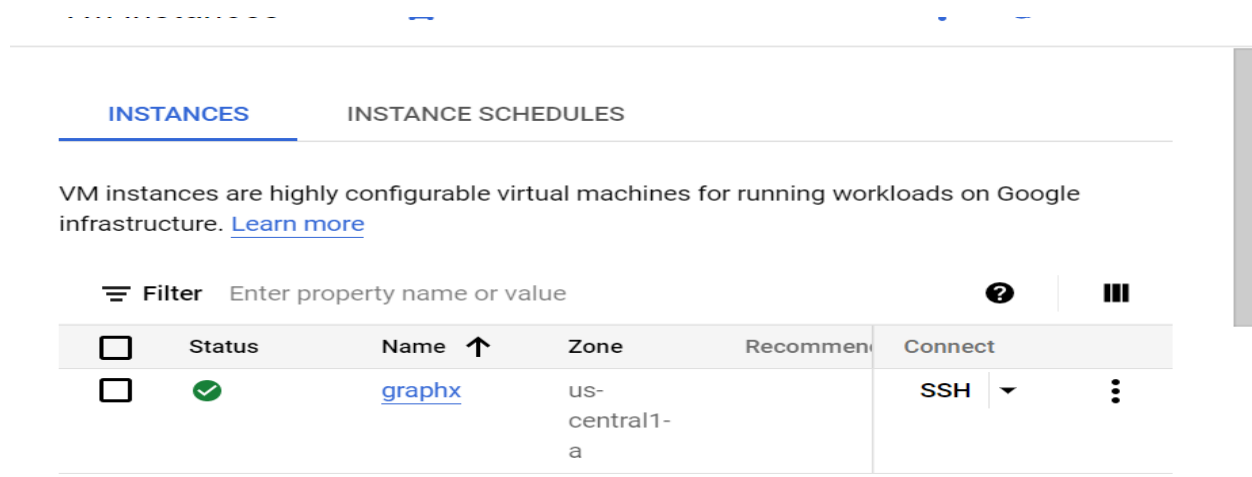


## Create the Virtual Machine on GCP and SSH into it.



## Install Java:

- ★ `sudo apt-get update`
- ★ `sudo apt-get install openjdk-11-jdk`

```
dpandey@graphx:~$ sudo apt install default-jre
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following packages were automatically installed and are no longer required:
  libatasmart4 libblockdev-fs2 libblockdev-loop2 libblockdev-part-err2
  libblockdev-part2 libblockdev-swap2 libblockdev-utils2 libblockdev2
  libmbim-glib4 libmbim-proxy libmm-glib0 libnuma1 libparted-fs-resize0
  libqmi-glib5 libqmi-proxy libudisks2-0 libxmlb2 usb-modeswitch
  usb-modeswitch-data
Use 'sudo apt autoremove' to remove them.
The following additional packages will be installed:
  at-spi2-core ca-certificates-java default-jre-headless fontconfig-config
  fonts-dejavu-core fonts-dejavu-extra java-common libatk-bridge2.0-0
  libatk-wrapper-java libatk-wrapper-java-jni libatk1.0-0 libatk1.0-data
  libatspi2.0-0 libavahi-client3 libavahi-common-data libavahi-common3 libcups2
  libdrm-amdgpu1 libdrm-intel1 libdrm-nouveau2 libdrm-radeon1 libfontconfig1
  libfontenc1 libgif7 libgl1 libgl1-mesa-dri libglapi-mesa libglvnd0
  libglx-mesa0 libglx0 libgraphite2-3 libharfbuzz0b libice6 libjpeg-turbo8
  libjpeg8 liblcms2-2 libllvm12 libpciaccess0 libpcsclite1 libsensors-config
  libpng5 libpng6 libusb-1.0-0 libusb-1.0-dev libusb-1.0-0-dev libusb-1.0-0-libusb2
  libusb-1.0-0-libusb2-dev libusb-1.0-0-libusb2-dev libusb-1.0-0-libusb2-dev
```

## Install Spark and Unpack it:

```
$ wget https://dlcdn.apache.org/spark/spark-3.3.1/spark-3.3.1-bin-hadoop3.tgz
```

```
$ tar -xvf spark-3.3.1-bin-hadoop3.tgz
```

```
$ mv spark-3.3.1-bin-hadoop3 spark
```

```
dpandey@graphx:~$ wget https://dlcdn.apache.org/spark/spark-3.3.1/spark-3.3.1-bin-hadoop3.tgz
--2022-12-18 17:19:36-- https://dlcdn.apache.org/spark/spark-3.3.1/spark-3.3.1-bin-hadoop3.tgz
Resolving dlcdn.apache.org (dlcdn.apache.org)... 151.101.2.132, 2a04:4e42::644
Connecting to dlcdn.apache.org (dlcdn.apache.org)|151.101.2.132|:443... connected
.
HTTP request sent, awaiting response... 200 OK
Length: 299350810 (285M) [application/x-gzip]
Saving to: 'spark-3.3.1-bin-hadoop3.tgz'

spark-3.3.1-bin-hadoop3.tgz 100%[=====>] 28
5.48M 237MB/s in 1.2s

2022-12-18 17:19:37 (237 MB/s) - 'spark-3.3.1-bin-hadoop3.tgz' saved [299350810/299350810]

dpandey@graphx:~$ tar -xvf spark-3.3.1-bin-hadoop3.tgz
spark-3.3.1-bin-hadoop3/
spark-3.3.1-bin-hadoop3/LICENSE
spark-3.3.1-bin-hadoop3/NOTICE
spark-3.3.1-bin-hadoop3/R/
spark-3.3.1-bin-hadoop3/R/lib/
```

Create graphX directory and then prepare person.csv and relationship.csv file into graphX directory

```
$ mkdir graphX
```

```
dpandey@graphx:~$ cat person.csv
id,Name,Age
1,Andrew,45
2,Sierra,43
3,Bob,12
4,Emily,10
5,William,35
6,Rachel,32
dpandey@graphx:~$ cat relationship.csv
src,dst,relation
1,2,Husband
1,3,Father
1,4,Father
1,5,Friend
1,6,Friend
2,1,Wife
2,3,Mother
2,4,Mother
2,6,Friend
3,1,Son
3,2,Son
4,1,Daughter
4,2,Daughter
5,1,Friend
6,1,Friend
6,2,Friend
```

### **Install numpy Library:**

```
$ sudo apt install python3-pip
```

```
$ pip3 install numpy
```

### **Prepare for the code:**

★ **graphX.py**

**Download Link:**

<https://github.com/divyapandey03/Cloud-Computing/blob/main/Apache%20Spark%20%2B%20GraphFramse%20%2B%20GraphX/graphX.py>

### **Download GraphFrames jar file:**

[https://repos.spark-packages.org/graphframes/graphframes/0.8.2-spark3.1-s\\_2.12/graphframes-0.8.2-spark3.1-s\\_2.12.jar](https://repos.spark-packages.org/graphframes/graphframes/0.8.2-spark3.1-s_2.12/graphframes-0.8.2-spark3.1-s_2.12.jar)

### **Submit Pyspark Job:**

```
$ spark-submit --packages graphframes:graphframes:0.8.2-spark3.1-s_2.12 pyspark_graphX.py
```

## Output:

## GraphFrame:

```
22/12/18 21:26:30 INFO CodeGenerator: Code generated in 34.000571 ms
```

```
+---+-----+---+
| id|   Name|Age|
+---+-----+---+
|  1| Andrew| 45|
|  2| Sierra| 43|
|  3|   Bob| 12|
|  4|  Emily| 10|
|  5|William| 35|
|  6| Rachel| 32|
+---+-----+---+
```

```
0 s
```

```
+---+---+-----+
|src|dst|relation|
+---+---+-----+
|  1|  2| Husband|
|  1|  3|  Father|
|  1|  4|  Father|
|  1|  5|  Friend|
|  1|  6|  Friend|
|  2|  1|   Wife|
|  2|  3| Mother|
|  2|  4| Mother|
|  2|  6|  Friend|
|  3|  1|   Son|
|  3|  2|   Son|
|  4|  1| Daughter|
|  4|  2| Daughter|
|  5|  1|  Friend|
|  6|  1|  Friend|
|  6|  2|  Friend|
+---+---+-----+
```

## Triangle Count:

```
4 s
22/12/18 21:26:34 INFO CodeGenerator: Code generated in 42.742765 ms
+---+-----+
| id|degree|
+---+-----+
1 1|    10|
+---+-----+
```

```
22/12/18 21:26:34 INFO FileSourceStrategy: Pushed Filters: IsNotNull(dst).
```

```
22/12/18 21:26:34 INFO DAGScheduler: Job 9 finished: showString a
```

```
5 s
+---+-----+
| id|inDegree|
+---+-----+
1 1|        5|
+---+-----+
```

```
22/12/18 21:26:35 INFO DAGScheduler: Job 11 finished: showString a
24 s
```

```
+---+-----+
| id|outDegree|
+---+-----+
1 1|        5|
+---+-----+
```

```
22/12/18 21:26:37 INFO FileSourceStrategy: Pushed Filters:
```

```
22/12/18 21:26:40 INFO CodeGenerator: Code
+-----+---+-----+---+
|count| id|   Name|Age|
+-----+---+-----+---+
|    3|  1| Andrew| 45|
|    1|  6| Rachel| 32|
|    1|  3|   Bob| 12|
|    0|  5|William| 35|
|    1|  4|  Emily| 10|
1 3|  2| Sierra| 43|
+-----+---+-----+---+
```

```

22/12/18 21:26:44 INFO CodeGenerator: Code generated in 13.037785 ms
+-----+-----+-----+-----+-----+
|count| id|  Name|Age|max_count|
+-----+-----+-----+-----+
|    3|  1|Andrew| 45|        3|
|    3|  2|Sierra| 43|        3|
+-----+-----+-----+-----+

22/12/18 21:26:44 INFO FileSourceStrategy: Pushed Filters:

```

## PageRank:

```

22/12/18 21:26:50 INFO CodeGenerator: Code generated in 21.243891 ms
+-----+-----+-----+-----+
| id|  Name|Age|          pagerank|
+-----+-----+-----+-----+
|  1| Andrew| 45| 1.787923121897472|
|  2| Sierra| 43| 1.406016795082752|
|  6| Rachel| 32| 0.7723665979473922|
|  4|  Emily| 10| 0.7723665979473922|
|  3|   Bob| 12| 0.7723665979473922|
|  5|William| 35| 0.4889602891776001|
+-----+-----+-----+-----+

22/12/18 21:26:51 INFO CodeGenerator: Code generated in 12.548207 ms
+-----+-----+-----+-----+
|src|dst|relation|weight|
+-----+-----+-----+-----+
|  5|  1| Friend| 1.0|
|  3|  1|   Son| 0.5|
|  4|  1|Daughter| 0.5|
|  4|  2|Daughter| 0.5|
|  6|  1| Friend| 0.5|
|  3|  2|   Son| 0.5|
|  6|  2| Friend| 0.5|
|  2|  3| Mother| 0.25|
|  2|  4| Mother| 0.25|
|  2|  1|  Wife| 0.25|
|  2|  6| Friend| 0.25|
|  1|  2| Husband| 0.2|
|  1|  6| Friend| 0.2|
|  1|  3| Father| 0.2|
|  1|  4| Father| 0.2|
|  1|  5| Friend| 0.2|

```

## BFS:

```
89 s
22/12/18 21:26:54 INFO CodeGenerator: Code generated in 17.623646 ms
+-----+-----+-----+-----+-----+
|      from|      e0|      v1|      e1|      to|
+-----+-----+-----+-----+
|{3, Bob, 12}|{3, 1, Son}|{1, Andrew, 45}|{1, 5, Friend}|{5, William, 35}|
+-----+-----+-----+-----+

t-cs570.internal:35549 in memory (size: 7.6 KiB, free: 434.1 MiB)
+-----+-----+-----+-----+-----+
|      from|      e0|      v1|      e1|      to|
+-----+-----+-----+-----+
|{4, Emily, 10}|{4, 1, Daughter}|{1, Andrew, 45}|{1, 6, Friend}|{6, Rachel, 32}|
|{4, Emily, 10}|{4, 2, Daughter}|{2, Sierra, 43}|{2, 6, Friend}|{6, Rachel, 32}|
+-----+-----+-----+-----+

22/12/18 21:26:57 INFO DAGScheduler: Job 62 finished: showString at NativeMethodAccessor1
19 s
+-----+-----+-----+-----+-----+
|      from|      e0|      v1|      e1|      to|
+-----+-----+-----+-----+
|{4, Emily, 10}|{4, 1, Daughter}|{1, Andrew, 45}|{1, 6, Friend}|{6, Rachel, 32}|
| {3, Bob, 12}| {3, 1, Son}|{1, Andrew, 45}|{1, 6, Friend}|{6, Rachel, 32}|
|{4, Emily, 10}|{4, 2, Daughter}|{2, Sierra, 43}|{2, 6, Friend}|{6, Rachel, 32}|
| {3, Bob, 12}| {3, 2, Son}|{2, Sierra, 43}|{2, 6, Friend}|{6, Rachel, 32}|
+-----+-----+-----+-----+
[1]
```