In [18]:

```
import sys
import os
import pandas as pd
from pyspark.context import SparkContext
from pyspark.sql.session import SparkSession
```

Installing Apache Spark

In [3]:

```
!apt-get install openjdk-8-jdk-headless -qq > /dev/null
!wget -q https://archive.apache.org/dist/spark/spark-3.0.0/spark-3.0.0-bin-hadoop3.2.tgz
!tar xf spark-3.0.0-bin-hadoop3.2.tgz
```

In [15]:

```
os.environ["JAVA_HOME"] = "/usr/lib/jvm/java-8-openjdk-amd64"
os.environ["SPARK_HOME"] = "/content/spark-3.0.0-bin-hadoop3.2"
```

In [16]:

```
!pip install findspark
```

```
Requirement already satisfied: findspark in /usr/local/lib/python3.7/dist-packages (1.4.2
)
```

In [17]:

```
import findspark
findspark.init("/content/spark-3.0.0-bin-hadoop3.2")
```

Importing CSV file in pandas

In [7]:

```
data = pd.read_csv('mobile.csv')
print(data)
```

```
        id  battery_power  blue  ...  three_g  touch_screen  wifi
0        1           1043     1  ...        0             1     0
1        2            841     1  ...        1             0     0
2        3           1807     1  ...        0             1     1
3        4           1546     0  ...        1             1     0
4        5           1434     0  ...        1             0     1
..     ...            ...   ...  ...      ...           ...   ...
995    996           1700     1  ...        1             1     0
996    997            609     0  ...        0             1     1
997    998           1185     0  ...        1             0     0
998    999           1533     1  ...        0             1     0
999   1000           1270     1  ...        1             0     1

[1000 rows x 21 columns]
```

Dualsim data in the format of list

In [8]:

```
data = list(data['dual_sim'])
print(data)
```

```
[1, 1, 0, 1, 0, 1, 0, 1, 1, 0, 0, 0, 1, 1, 0, 0, 1, 0, 1, 0, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1,
1, 1, 1, 1, 0, 0, 1, 1, 0, 0, 0, 1, 1, 1, 0, 0, 1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0,
1, 1, 0, 1, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1,
1, 1, 1, 0, 1, 0, 1, 1, 0, 0, 1, 0, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 1, 0,
```

```
0, 1, 1, 0, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 0, 0, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 0, 0,
0, 1, 1, 1, 0, 0, 1, 1, 0, 1, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0,
0, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 0, 1, 0, 0, 1,
0, 0, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 0, 0, 1, 0, 0, 1, 0, 0, 1, 1, 1, 1, 0, 0, 1, 0, 0,
0, 1, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 1, 0, 0, 0, 1, 0, 1, 1, 0, 0,
1, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 1, 1,
1, 0, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 1, 0, 1, 1, 1, 1,
1, 1, 0, 1, 1, 0, 0, 0, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 1,
0, 0, 1, 1, 1, 0, 0, 0, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 0, 0, 0, 1,
0, 1, 1, 0, 0, 1, 1, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0,
0, 0, 0, 1, 1, 0, 1, 1, 0, 0, 1, 1, 1, 0, 0, 0, 1, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0,
1, 0, 1, 0, 1, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 1, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1,
0, 1, 0, 1, 1, 1, 0, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0, 0, 0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1,
1, 1, 1, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 1, 1, 0, 1, 0, 1, 0, 0, 1, 1, 0, 0, 0, 1, 0, 1,
1, 1, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 0,
0, 1, 1, 1, 0, 1, 0, 0, 1, 0, 1, 0, 1, 1, 0, 0, 1, 0, 1, 1, 1, 1, 0, 0, 1, 1, 0, 0, 0, 0,
0, 1, 0, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 1, 1, 0,
1, 0, 1, 0, 0, 0, 1, 1, 1, 0, 0, 1, 0, 1, 1, 1, 1, 0, 1, 0, 0, 1, 1, 1, 1, 1, 0, 1, 0, 1,
1, 0, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 0, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 0, 1,
1, 1, 1, 0, 1, 1, 1, 1, 0, 0, 0, 1, 1, 0, 0, 1, 1, 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 1, 0, 1,
0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 0, 0,
0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0,
0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 1, 0, 1, 0, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 1,
0, 0, 1, 1, 1, 1, 0, 0, 0, 1, 0, 1, 1, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 0, 0, 1, 0,
1, 0, 1, 0, 0, 1, 0, 1, 1, 0, 0, 1, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 0, 1, 0, 0, 1, 1, 0, 1,
0, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 0, 0, 0, 0,
0, 0, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 1,
1, 1, 0, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 0, 0, 1, 0, 0, 1, 0, 1,
0, 0, 1, 1, 1, 0, 1, 0, 1, 0]
```

By using Spark

In [19]:

```
spark = SparkSession(SparkContext.getOrCreate())

mobileConfig = spark.read.csv(
                  "mobile.csv",
                  header = "true"
                )
mobileConfig.show()
```

```
+---+-------------+----+-----------+--------+---+------+----------+-----+---------+----
---+---+---------+--------+----+----+----+---------+-------+------------+----+
| id|battery_power|blue|clock_speed|dual_sim| fc|four_g|int_memory|m_dep|mobile_wt|n_core
s| pc|px_height|px_width| ram|sc_h|sc_w|talk_time|three_g|touch_screen|wifi|
+---+-------------+----+-----------+--------+---+------+----------+-----+---------+----
---+---+---------+--------+----+----+----+---------+-------+------------+----+
|  1|         1043|   1|        1.8|       1| 14|     0|         5|  0.1|      193|
3| 16|      226|    1412|3476|  12|   7|         2|      0|           1|   0|
|  2|          841|   1|        0.5|       1|  4|     1|        61|  0.8|      191|
5| 12|      746|     857|3895|   6|   0|         7|      1|           0|   0|
|  3|         1807|   1|        2.8|       0|  1|     0|        27|  0.9|      186|
3|  4|     1270|    1366|2396|  17|  10|        10|      0|           1|   1|
|  4|         1546|   0|        0.5|       1| 18|     1|        25|  0.5|       96|
8| 20|      295|    1752|3893|  10|   0|         7|      1|           1|   0|
|  5|         1434|   0|        1.4|       0| 11|     1|        49|  0.5|      108|
6| 18|      749|     810|1773|  15|   8|         7|      1|           0|   1|
|  6|         1464|   1|        2.9|       1|  5|     1|        50|  0.8|      198|
8|  9|      569|     939|3506|  10|   7|         3|      1|           1|   1|
|  7|         1718|   0|        2.4|       0|  1|     0|        47|    1|      156|
2|  3|     1283|    1374|3873|  14|   2|        10|      0|           0|   0|
|  8|          833|   0|        2.4|       1|  0|     0|        62|  0.8|      111|
1|  2|     1312|    1880|1495|   7|   2|        18|      0|           1|   1|
|  9|         1111|   1|        2.9|       1|  9|     1|        25|  0.6|      101|
5| 19|      556|     876|3485|  11|   9|        10|      1|           1|   0|
| 10|         1520|   0|        0.5|       0|  1|     0|        25|  0.5|      171|
3| 20|       52|    1009| 651|   6|   0|         5|      1|           0|   1|
| 11|         1500|   0|        2.2|       0|  2|     0|        55|  0.6|       80|
7|  6|      503|    1336|3866|  13|   7|        20|      0|           1|   0|
| 12|         1343|   0|        2.9|       0|  2|     1|        34|  0.8|      171|
```

```
3|  6|      235|    1671|3911|  15|   8|        8|      1|           1|   1|
| 13|          900|   1|        1.4|       1|  0|     0|        30|    1|       87|
2|  3|      829|    1893| 439|   6|   2|       20|      1|           0|   0|
| 14|         1190|   1|        2.2|       1|  5|     0|        19|  0.9|      158|
5| 15|      227|    1856| 992|  13|   0|       16|      1|           1|   0|
| 15|          630|   0|        1.8|       0|  8|     1|        51|  0.9|      193|
8|  9|     1315|    1323|2751|  17|   6|        3|      1|           1|   0|
| 16|         1846|   1|          1|       0|  5|     1|        53|  0.7|      106|
8|  7|      185|    1832| 563|   9|   5|       10|      1|           0|   1|
| 17|         1985|   0|        0.5|       1| 14|     1|        26|    1|      163|
2| 17|      613|    1511|2083|  13|   3|       14|      1|           1|   0|
| 18|         1042|   0|        2.9|       0|  5|     1|        48|  0.2|      186|
4| 15|      335|     532|2187|   9|   2|        5|      1|           0|   0|
| 19|         1231|   1|        1.7|       1|  2|     1|        37|  0.2|      194|
2|  3|       82|    1771|3902|  19|  12|       15|      1|           0|   1|
| 20|         1488|   0|        2.6|       0|  9|     0|        37|  0.7|      189|
4| 20|       47|     559|2524|   5|   0|        6|      0|           0|   0|
+---+-------------+----+----------+--------+---+------+----------+-----+---------+----
---+---+---------+--------+----+----+----+---------+-------+------------+----+
only showing top 20 rows
```

Reading the dual_sim column in the pyspark

In [20]:

```
sc = SparkContext.getOrCreate()
spark = SparkSession(sc)

mobileConfig = spark.read.csv(
                  "mobile.csv",
                  header = "true"
                )
mobileConfig.select("dual_sim").show()
```

```
+--------+
|dual_sim|
+--------+
|       1|
|       1|
|       0|
|       1|
|       0|
|       1|
|       0|
|       1|
|       1|
|       0|
|       0|
|       0|
|       1|
|       1|
|       0|
|       0|
|       1|
|       0|
|       1|
|       0|
+--------+
only showing top 20 rows
```

Dataframe to list convertion

In [21]:

```
dual_sim = mobileConfig.select("dual_sim").rdd.flatMap(lambda x: x).collect()
print(dual_sim)
```

```
['1', '1', '0', '1', '0', '1', '0', '1', '1', '0', '0', '0', '1', '1', '0', '0', '1', '0'
```

', '1', '0', '0', '1', '1', '1', '1', '1', '1', '0', '1', '1', '1', '1', '1', '1', '0', '0
', '1', '1', '0', '0', '0', '1', '1', '1', '0', '0', '1', '1', '1', '0', '1', '0', '1', '
0', '1', '1', '0', '1', '1', '0', '1', '1', '0', '1', '1', '0', '0', '1', '0', '0', '1',
'0', '0', '0', '1', '1', '1', '1', '1', '1', '1', '0', '0', '1', '1', '0', '1', '0', '1',
'1', '1', '1', '1', '0', '1', '0', '1', '1', '0', '0', '1', '0', '1', '1', '1', '0', '1',
'0', '1', '1', '1', '1', '1', '1', '1', '0', '0', '0', '1', '0', '0', '1', '1', '0', '0',
'0', '1', '1', '0', '0', '1', '1', '0', '0', '1', '0', '0', '0', '1', '1', '1', '1', '0',
'1', '1', '1', '1', '0', '0', '0', '0', '1', '1', '1', '0', '0', '1', '1', '0', '1', '1',
'0', '0', '1', '0', '1', '0', '0', '0', '0', '0', '0', '0', '1', '1', '0', '1', '0', '0',
'0', '0', '0', '0', '0', '0', '1', '1', '1', '1', '0', '0', '0', '0', '1', '0', '0', '1',
'0', '1', '1', '1', '0', '1', '1', '1', '0', '1', '0', '0', '1', '0', '0', '0', '1', '1',
'0', '1', '1', '0', '1', '0', '1', '1', '0', '0', '1', '0', '0', '1', '0', '0', '1', '1',
'1', '1', '0', '0', '1', '0', '0', '0', '1', '0', '0', '1', '0', '0', '1', '0', '1', '0',
'0', '0', '1', '0', '1', '1', '0', '1', '1', '1', '0', '0', '0', '1', '0', '1', '1', '0',
'0', '1', '0', '1', '0', '0', '0', '1', '0', '1', '0', '1', '0', '1', '1', '1', '0', '1',
'1', '0', '0', '0', '1', '0', '0', '1', '0', '1', '0', '1', '1', '1', '0', '1', '0', '0',
'1', '1', '1', '1', '1', '1', '1', '0', '1', '1', '1', '1', '0', '1', '1', '0', '0', '1',
'1', '1', '0', '1', '1', '1', '1', '1', '1', '0', '1', '1', '0', '0', '0', '1', '1', '1',
'0', '1', '1', '1', '0', '1', '1', '1', '0', '0', '0', '1', '0', '0', '1', '0', '0', '1',
'1', '0', '0', '1', '1', '1', '0', '0', '0', '0', '1', '0', '1', '1', '1', '1', '1', '1',
'1', '1', '1', '1', '1', '0', '1', '1', '0', '0', '0', '0', '1', '0', '1', '1', '0', '0',
'1', '1', '0', '0', '0', '0', '1', '1', '1', '0', '0', '0', '0', '1', '1', '0', '0', '1',
'0', '0', '1', '1', '0', '0', '0', '0', '0', '0', '1', '1', '0', '1', '1', '0', '0', '1',
'1', '1', '0', '0', '0', '1', '1', '0', '1', '1', '0', '0', '0', '0', '0', '0', '0', '0',
'0', '1', '0', '1', '0', '1', '1', '0', '0', '1', '1', '0', '0', '0', '0', '0', '0', '0',
'1', '1', '0', '1', '0', '1', '1', '0', '0', '0', '0', '0', '0', '0', '0', '0', '0', '0',
'0', '0', '0', '0', '0', '1', '0', '1', '1', '1', '0', '0', '1', '0', '1', '0', '0', '0',
'1', '0', '0', '0', '0', '0', '1', '0', '1', '0', '1', '1', '1', '0', '0', '1', '0', '1',
'0', '1', '0', '1', '1', '0', '0', '0', '1', '1', '1', '0', '1', '1', '0', '1', '1', '0',
'1', '1', '1', '1', '0', '0', '0', '0', '1', '1', '1', '0', '0', '0', '0', '1', '1', '0',
'1', '0', '1', '0', '0', '1', '1', '0', '0', '0', '1', '0', '1', '1', '1', '1', '0', '1',
'0', '1', '0', '0', '0', '1', '0', '0', '0', '1', '1', '1', '1', '0', '1', '1', '1', '0',
'1', '1', '1', '1', '0', '1', '0', '0', '1', '1', '1', '0', '1', '0', '0', '1', '0', '1',
'0', '1', '1', '0', '0', '1', '0', '1', '1', '1', '1', '0', '0', '1', '1', '0', '0', '0',
'0', '0', '1', '0', '1', '1', '1', '1', '1', '0', '0', '0', '0', '0', '0', '1', '0', '0',
'0', '0', '1', '1', '1', '0', '0', '1', '1', '0', '1', '1', '0', '1', '0', '1', '0', '0',
'0', '1', '1', '1', '0', '0', '1', '0', '1', '1', '1', '1', '0', '1', '0', '0', '1', '1',
'1', '1', '1', '0', '1', '0', '1', '1', '0', '1', '0', '0', '1', '1', '0', '1', '0', '1',
'1', '0', '0', '1', '1', '0', '1', '1', '0', '1', '0', '1', '1', '1', '0', '1', '1', '0',
'1', '1', '1', '1', '0', '1', '1', '1', '1', '0', '0', '0', '1', '1', '0', '0', '1', '1',
'0', '1', '1', '0', '0', '0', '0', '1', '1', '0', '1', '0', '1', '0', '0', '1', '0', '1',
'0', '0', '0', '0', '0', '0', '0', '0', '0', '1', '0', '1', '1', '1', '1', '0', '1', '1',
'0', '1', '0', '1', '1', '0', '0', '0', '1', '1', '1', '1', '1', '0', '1', '1', '1', '0',
'0', '1', '1', '1', '1', '0', '1', '1', '0', '0', '1', '1', '0', '1', '0', '0', '0', '0',
'0', '0', '0', '0', '0', '1', '0', '0', '0', '1', '1', '0', '1', '0', '1', '0', '1', '1',
'0', '0', '1', '1', '1', '1', '1', '1', '0', '0', '0', '0', '1', '0', '0', '1', '1', '1',
'1', '0', '0', '0', '1', '0', '1', '1', '1', '1', '0', '1', '1', '0', '1', '0', '1', '1',
'1', '1', '1', '0', '0', '1', '0', '1', '0', '1', '0', '0', '1', '0', '1', '1', '0', '0',
'1', '1', '1', '0', '1', '0', '1', '1', '1', '0', '1', '0', '1', '0', '0', '1', '1', '0',
'1', '0', '0', '0', '1', '1', '0', '0', '1', '1', '0', '0', '1', '1', '1', '0', '1', '0',
'1', '1', '1', '0', '1', '1', '0', '1', '1', '0', '0', '0', '0', '0', '0', '1', '0', '1',
'0', '1', '0', '1', '1', '1', '0', '1', '1', '0', '0', '1', '1', '0', '0', '0', '0', '0',
'1', '1', '1', '0', '0', '0', '1', '1', '1', '0', '1', '1', '1', '0', '1', '1', '0', '0',
'1', '1', '0', '1', '0', '1', '1', '1', '1', '1', '1', '0', '0', '1', '0', '0', '1', '0',
'1', '0', '0', '1', '1', '1', '0', '1', '0', '1', '1', '0', '1', '0', '1', '1', '1', '0']

Using map and reduceByKey to implement the Mapreduce method

In [22]:

```
frequent_words = (mobileConfig.select("dual_sim").rdd.flatMap(lambda x: x).map(lambda w:
(w, 1)).reduceByKey(lambda x,y: x+y))
print(frequent_words.collect())
```

```
[('1', 517), ('0', 483)]
```

Observation :
Dualsim Phone - 517
Singlesim Phone - 483