Session 14: Scala Basics 1 – Assignment 1

1. Start the hadoop in the VM and check all the daemons are running using the command JPS:
   $start-all.sh
   $jps

```
File  Edit  View  Search  Terminal  Help
[acadgild@localhost ~]$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
18/10/10 15:57:56 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Starting namenodes on [localhost]
localhost: starting namenode, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadgild-namenode-localhost.l
ocaldomain.out
localhost: starting datanode, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadgild-datanode-localhost.l
ocaldomain.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadgild-secondaryname
node-localhost.localdomain.out
18/10/10 15:58:19 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
starting yarn daemons
starting resourcemanager, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/yarn-acadgild-resourcemanager-localhost.
localdomain.out
localhost: starting nodemanager, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/yarn-acadgild-nodemanager-localho
st.localdomain.out
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ jps
11440 DataNode
11313 NameNode
12563 Jps
11892 NodeManager
11637 SecondaryNameNode
11787 ResourceManager
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

2. Now start the spark terminal as follows:
   $spark-shell

```
[acadgild@localhost ~]$ spark-shell
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
18/10/10 16:11:34 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
18/10/10 16:11:35 WARN util.Utils: Your hostname, localhost.localdomain resolves to a loopback address: 127.0.0.1; using 192.
168.0.13 instead (on interface eth16)
18/10/10 16:11:35 WARN util.Utils: Set SPARK_LOCAL_IP if you need to bind to another address
Spark context Web UI available at http://192.168.0.13:4040
Spark context available as 'sc' (master = local[*], app id = local-1539168098297).
Spark session available as 'spark'.
Welcome to
      ____              __
     / __/__  ___ _____/ /__
    _\ \/ _ \/ _ `/ __/  '_/
   /___/ .__/\_,_/_/ /_/\_\   version 2.2.1
      /_/

Using Scala version 2.11.8 (Java HotSpot(TM) 64-Bit Server VM, Java 1.8.0_151)
Type in expressions to have them evaluated.
Type :help for more information.

scala>
```

Task 1:
Given a list of strings – ***List[String](“alpha”, “gamma”, “omega”, “zeta”, “beta”)***

1. **Find count of all strings with length 4.**

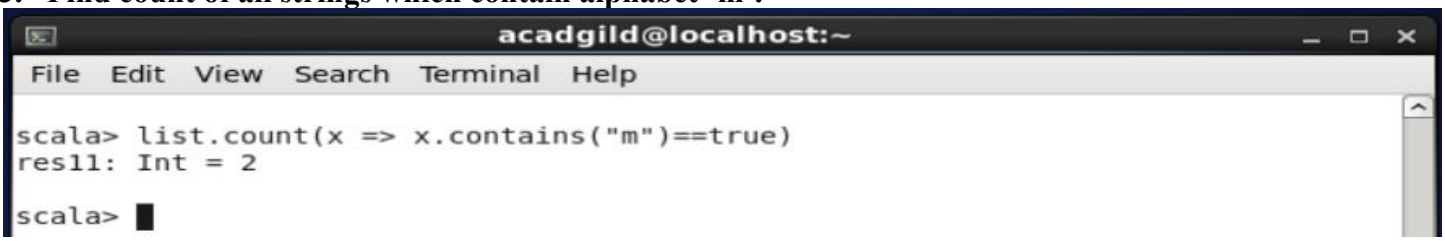
```
acadgild@localhost:~
File  Edit  View  Search  Terminal  Help
scala> val list = List[String] ("alpha", "gamma", "omega", "zeta", "beta")
list: List[String] = List(alpha, gamma, omega, zeta, beta)

scala> list.count(x => x.length == 4)
res0: Int = 2

scala>
```

We can observe that there are two strings with the length of 4.

2. **Convert the list of string to a list of integers, where each string is mapped to its corresponding length.**

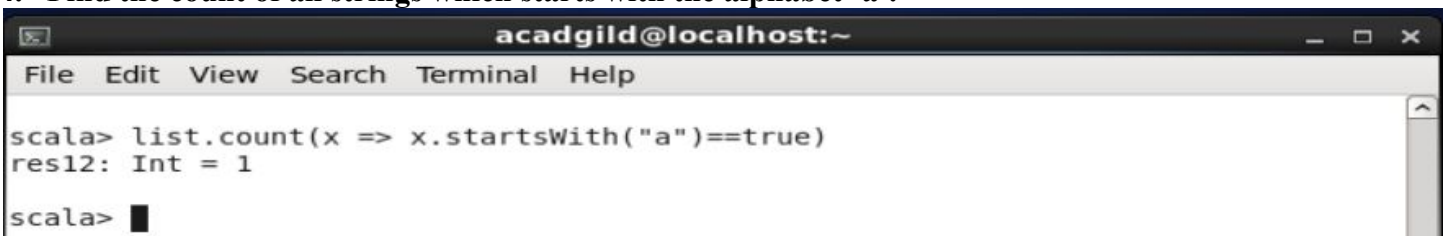
We can observe that the length of the each string that is present in the list of strings.

3. **Find count of all strings which contain alphabet 'm'.**



We can observe that there are two strings which contains the letter 'm'.

4. **Find the count of all strings which starts with the alphabet 'a'.**

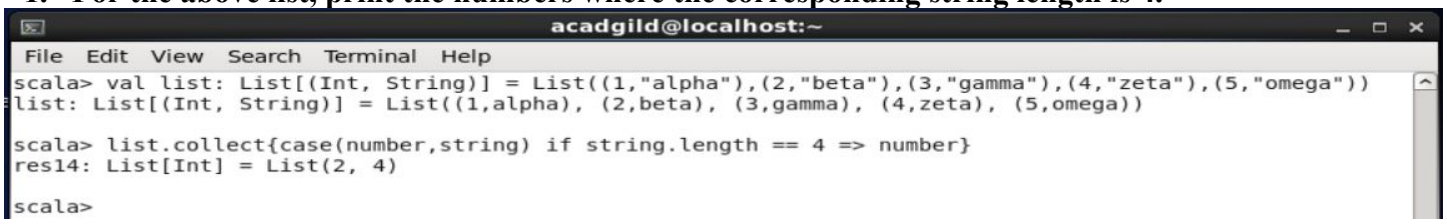


We can observe that there are only one string which starts with the letter 'a'.

Task 2

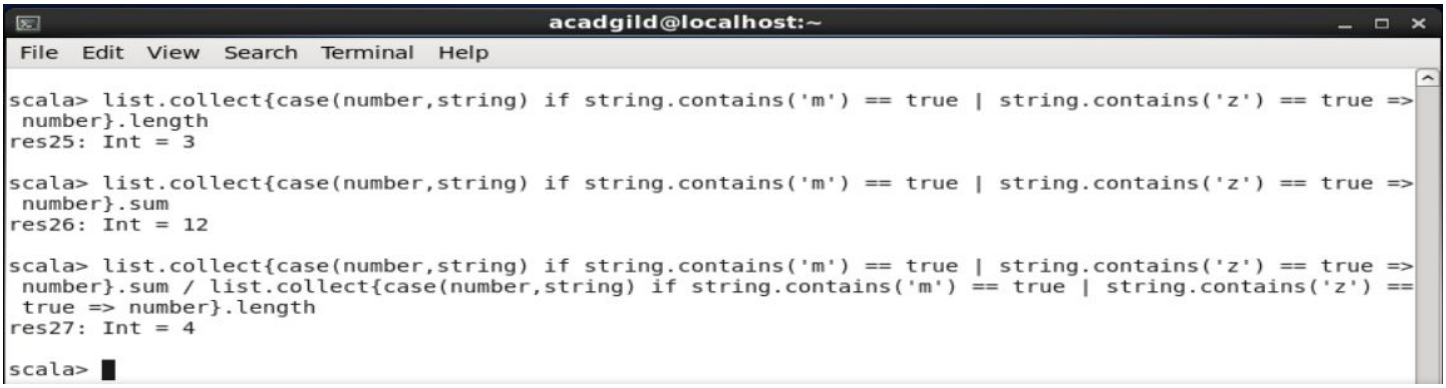
Create a list of tuples, where the 1st element of the tuple is an int and the second element is a string. Example: ((1, 'alpha'), (2, 'beta'), (3, 'gamma'), (4, 'zeta'), (5, 'omega'))

1. **For the above list, print the numbers where the corresponding string length is 4.**



With this we can observe that there are two strings with tuple numbers 2 and 4 having the length of 4.

2. **Find the average of all numbers, where the corresponding string contains alphabet 'm' or alphabet 'z'.**



First count the number of strings with letters ‘m’ or ‘z’. In the first line of output, we will get 3 elements, i.e., gamma, zeta and omega.

In these second output we are printing the sum of those key values, corresponding to the characters ‘m’ or ‘z’ present in the strings. So, we got sum as 12.

Now in the third step we will combine the above statements to get average of the key values. i.e, **sum / length.**

We will get the average value of 4.