Big Data Analytics – CS7070
Programming Project #2
Submitted By: Ankit Pandey
M13435273

1. **Program for TinyDataSet**

   Source Code:

```
%spark.pyspark

# read input text file to RDD
rdd = sc.textFile("/tmp/data/tinyDataset.txt")

#Define the List with the list constructor
list_rdd=list()

#Store the rdd in List with the help of collect
list_rdd=rdd.collect()

#Iterating through the Loop to display the graph taken as Input
print("TinyDataSet Graph: (List of edges) as Input:")
for x in range(len(list_rdd)):
   print(list_rdd[x])

#Splitting the RDD at spaces
rdd2=rdd.map(lambda x: x.split())

#FlatMap will flatten multiple list into single list and storing the verices of the graph in two-way form assuming it to be a undirected graph
rdd3=rdd2.flatMap(lambda y:[[y[0],y[1]],[y[1],y[0]]])

# reduceByKey Merges the values for each key. It will perform the merging locally on each mapper before sending results to a reducer, similarly to a “combiner” in MapReduce.
rdd4=rdd3.reduceByKey(lambda k,v:k+"-"+v)
list_rdd4=list()
list_rdd4=rdd4.collect()
tup1=()
print("Output- For each node  list of all the nodes to which the key node is connected.")
for l in range(len(list_rdd4)):
   tup1=list_rdd4[l]
   print("Node:"+tup1[0]+"\t "+"\tList of Nodes it is connected to: {"+tup1[1]+"}")
```

Tiny Dataset Graph: (List of edges) as Input:

7 10
7 8
7 4
8 9
8 5
9 5
9 10
10 6
4 5
5 6
4 6
1 4
1 3
2 3
2 6
3 4
3 6

Output- For each node list of all the nodes to which the key node is connected.

Node:10 List of Nodes it is connected to: {7-9-6}
Node:8 List of Nodes it is connected to: {7-9-5}
Node:4 List of Nodes it is connected to: {7-5-6-1-3}
Node:9 List of Nodes it is connected to: {8-5-10}
Node:1 List of Nodes it is connected to: {4-3}
Node:7 List of Nodes it is connected to: {10-8-4}
Node:5 List of Nodes it is connected to: {8-9-4-6}
Node:6 List of Nodes it is connected to: {10-5-4-2-3}
Node:3 List of Nodes it is connected to: {1-2-4-6}
Node:2 List of Nodes it is connected to: {3-6}

2. **Program for Small Data Set**

Source Code:

```
%spark.pyspark

# read input text file to RDD
rdd = sc.textFile("/tmp/data/SmallDataSet.txt")

#Define the List with the list constructor
list_rdd=list()

#Store the rdd in List with the help of collect
list_rdd=rdd.collect()

#Iterating through the Loop to display the graph taken as Input
print("SmallDataSet Graph: (List of edges) as Input:")
for x in range(len(list_rdd)):
   print(list_rdd[x])

#Splitting the RDD at spaces
rdd2=rdd.map(lambda x: x.split())

#FlatMap will flatten multiple list into single list and storing the verices of the graph in two-way form assuming it to be a undirected graph
rdd3=rdd2.flatMap(lambda y:[[y[0],y[1]],[y[1],y[0]]])

# reduceByKey Merges the values for each key. It will perform the merging locally on each mapper before sending results to a reducer, similarly to a “combiner” in MapReduce.
rdd4=rdd3.reduceByKey(lambda k,v:k+"-"+v)
list_rdd4=list()
list_rdd4=rdd4.collect()
tup1=()
print("Output- For each node  list of all the nodes to which the key node is connected.")
for l in range(len(list_rdd4)):
   tup1=list_rdd4[l]
   print("Node:"+tup1[0]+"\t "+"\tList of Nodes it is connected to: {"+tup1[1]+"}")
```

SmallDataSet Graph: (List of edges) as Input:

1 2
2 3
3 4
4 5
1 3
3 5
1 10
1 6
2 6
2 7
2 11
3 7
3 12
3 8
4 8
5 4
4 9
5 9
6 7
8 7
8 9
9 14
9 13
4 13
8 13
8 12
7 12
7 11
11 6
20 19
5 33
14 34
6 10
13 14
13 12
12 11
11 10
28 29
30 28
31 28
32 28
32 31

30 31
29 30
29 31
32 30
24 28
29 24
20 29
25 29
30 25
21 30
26 30
26 31
27 31
32 27
32 23
23 27
27 22
22 31
22 26
26 27
26 21
21 25
25 20
20 24
24 19
19 28
10 19
19 39
10 15
28 40
15 19
40 39
11 15
16 15
16 11
16 12
16 20
16 21
21 17
17 22
17 16
17 12
17 13
18 13

14 18
14 23
18 23
18 22
23 22
21 22
20 21
33 34
23 35
34 35
32 36
35 36
1 37
37 38
10 38
38 39

Output- For each node list of all the nodes to which the key node is connected.

Node:1 List of Nodes it is connected to: {2-3-10-6-37}
Node:4 List of Nodes it is connected to: {3-5-8-5-9-13}
Node:10 List of Nodes it is connected to: {1-6-11-19-15-38}
Node:12 List of Nodes it is connected to: {3-8-7-13-11-16-17}
Node:8 List of Nodes it is connected to: {3-4-7-9-13-12}
Node:9 List of Nodes it is connected to: {4-5-8-14-13}
Node:14 List of Nodes it is connected to: {9-34-13-18-23}
Node:20 List of Nodes it is connected to: {19-29-25-24-16-21}
Node:19 List of Nodes it is connected to: {20-24-28-10-39-15}
Node:33 List of Nodes it is connected to: {5-34}
Node:34 List of Nodes it is connected to: {14-33-35}
Node:29 List of Nodes it is connected to: {28-30-31-24-20-25}
Node:24 List of Nodes it is connected to: {28-29-20-19}
Node:21 List of Nodes it is connected to: {30-26-25-16-17-22-20}
Node:26 List of Nodes it is connected to: {30-31-22-27-21}
Node:22 List of Nodes it is connected to: {27-31-26-17-18-23-21}
Node:40 List of Nodes it is connected to: {28-39}
Node:16 List of Nodes it is connected to: {15-11-12-20-21-17}
Node:17 List of Nodes it is connected to: {21-22-16-12-13}
Node:32 List of Nodes it is connected to: {27-23-36-28-31-30}
Node:27 List of Nodes it is connected to: {32-23-22-26-31}
Node:23 List of Nodes it is connected to: {32-27-14-18-22-35}
Node:31 List of Nodes it is connected to: {22-28-32-30-29-26-27}
Node:25 List of Nodes it is connected to: {21-20-29-30}
Node:28 List of Nodes it is connected to: {19-40-29-30-31-32-24}

Node:39 List of Nodes it is connected to: {19-40-38}
Node:15 List of Nodes it is connected to: {10-19-11-16}
Node:11 List of Nodes it is connected to: {15-16-2-7-6-12-10}
Node:13 List of Nodes it is connected to: {17-18-9-4-8-14-12}
Node:18 List of Nodes it is connected to: {13-14-23-22}
Node:35 List of Nodes it is connected to: {23-34-36}
Node:36 List of Nodes it is connected to: {32-35}
Node:37 List of Nodes it is connected to: {1-38}
Node:38 List of Nodes it is connected to: {37-10-39}
Node:2 List of Nodes it is connected to: {1-3-6-7-11}
Node:3 List of Nodes it is connected to: {2-4-1-5-7-12-8}
Node:5 List of Nodes it is connected to: {4-3-4-9-33}
Node:6 List of Nodes it is connected to: {1-2-7-11-10}
Node:7 List of Nodes it is connected to: {2-3-6-8-12-11}
Node:30 List of Nodes it is connected to: {28-31-29-32-25-21-26}