Rushi Samirbhai Patel rs397441@dal.ca
B00886157

Assignment 4

Problem-2

Task 1 : Data Processing using Spark

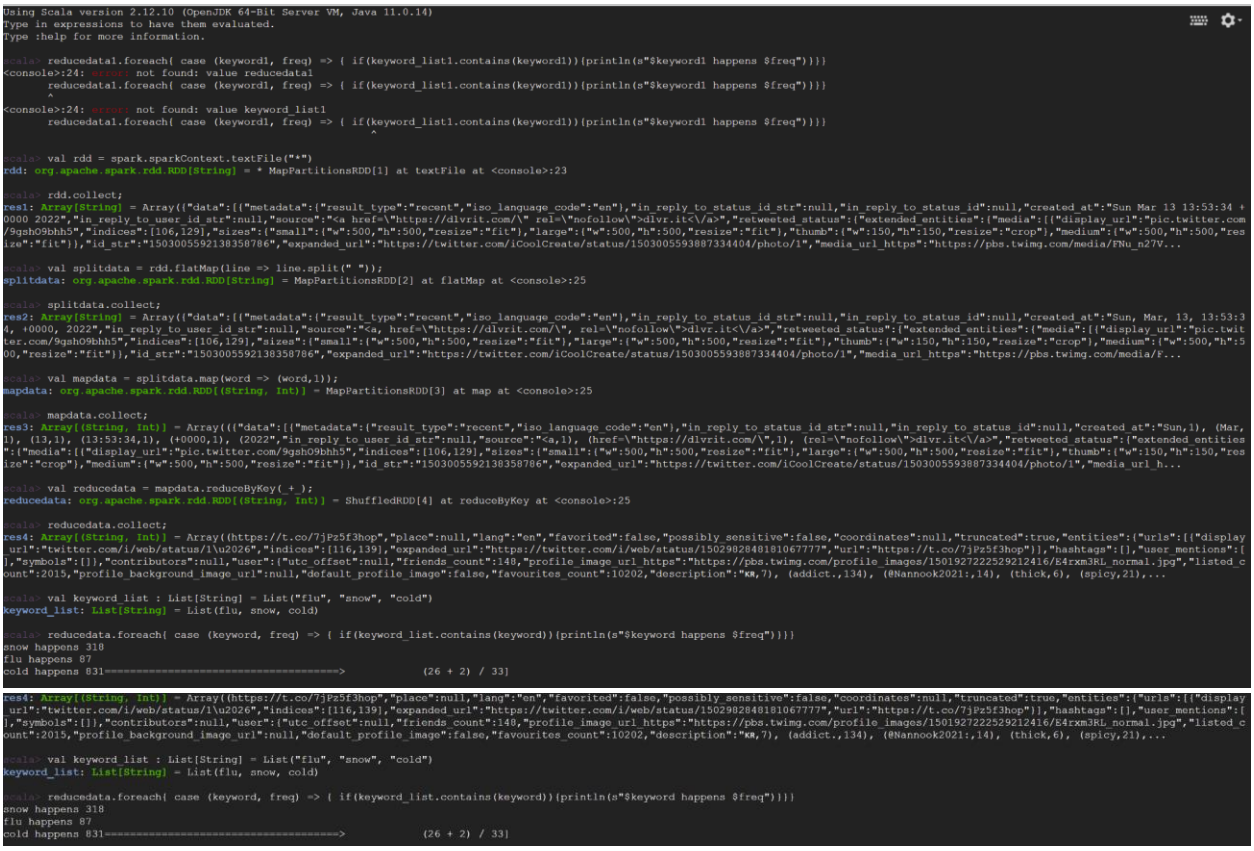
Mapreduce Algorithm:

Step 1: Connect to the instance created on GCP during task 1.
Step 2: Push all the text files to the gitlab repository and clone this repository in the GCP instance to retrieve all the text files containing the Twitter data.
Step 3: open spark shell from this directory and create RDD containing all the files with the command : val rdd = spark.sparkContext.textFile(“*”).
Step 4: Split the data by space to fetch all the words.
val splitdata = data.flatMap(line => line.split(" "));
Step 5: Perform the map operation. val mapdata = splitdata.map(word => (word,1));
Step 6: Perform the reduce operation. val reducedata = mapdata.reduceByKey(_+_);
Step 7: Create a list of words that we need the word count for : val keyword_list : List[String] = List("flu", “snow”, “cold”).
Step 8: Fetch the word count by iterating in a for each loop and matching the data with the list of words created in Step 7.

Word count:
Snow: 318
Flu: 87
Cold: 831

Rushi Samirbhai Patel rs397441@dal.ca
B00886157

Screenshot of the word count:



Task 2 : Data Visualization using Graph Database

Cypher queries:

//Creating Flu nodes

CREATE (n:FluNode{name:'flu',types:'Influenza A,Influenza B,Influenza C'})
CREATE (n:flu{name:'Influenza A', fluProperty:'antigenic',occurenceInTweets:'13', types:'hemagglutinin,neuraminidase', symptoms: 'cough,runny nose,fever,fatigue'})
CREATE (n:flu{name:'Influenza B', fluProperty:'antigenic',occurenceInTweets:'25', types:'types: Victoria,Yamagata', symptoms:'fever,chills,sore throat,cough'})

Rushi Samirbhai Patel rs397441@dal.ca
B00886157

CREATE (n:flu{name:'Influenza C', fluProperty:'genetic and antigenic',occurenceInTweets:'29', types:'hemagglutinin,neuraminidase', symptoms:'dry cough, muscle pain,achiness'})

//Creating Snow nodes

CREATE (n:SnowNode{name:'snow',types:'Blizzard,Snow storm,Snow Burst'})
CREATE (n:snow{typeOfSnowFall:'Blizzard', name:'Snowflakes', location:'Russia and central and northeastern Asia, northern Europe, Canada, the northern United States, and Antarctica', inches:'36'})
CREATE (n:snow{typeOfSnowFall:'Snow storm', name:'Polycrystals', location:'Northeastern United States, Eastern United States, Tibet, Eastern Canada', inches:'39'})
CREATE (n:snow{typeOfSnowFall:'Snow Burst', name:'Graupel', location:'United States, Europe and Canada', inches:'31'})

//Creating Cold nodes

CREATE (n:ColdNode{name:'cold',types:'Freezing Weather,Diseases'})
CREATE (n:cold{type:'freezing weather',name:'Freezing Weather',location:'Eastern Antarctic Plateau,Russia Greenland,Coastal regions',windSpeed:'>3mph',season:'winter'})
CREATE (n:cold{type:'diseases',name:'Pneumonia',symptoms:'Chest pain,Cough,Fever,Shortness of breath',types:'Bacterial,Fungal,Viral'})
CREATE (n:cold{type:'diseases',name:'Sinusitis',symptoms:'nasal inflamation,swelling,pain,blocked nose',types:'Accute,Subaccute,Chronic'})

//Generating relationships with cold nodes

MATCH (a:ColdNode), (b:cold) WHERE a.name="cold" AND b.name="Freezing Weather" CREATE (b)-[r:CONTEXT_OF]->(a)
MATCH (a:ColdNode), (b:cold) WHERE a.name="cold" AND b.name="Pneumonia" CREATE (b)-[r:TYPE_OF]->(a)
MATCH (a:ColdNode), (b:cold) WHERE a.name="cold" AND b.name="Sinusitis" CREATE (b)-[r:TYPE_OF]->(a)

//Generating relationships with snow nodes

MATCH (a:SnowNode), (b:snow) WHERE a.name="snow" AND b.typeOfSnowFall="Blizzard" CREATE (b)-[r:TYPE_OF]->(a)
MATCH (a:SnowNode), (b:snow) WHERE a.name="snow" AND b.typeOfSnowFall="Snow storm" CREATE (b)-[r:TYPE_OF]->(a)

Rushi Samirbhai Patel rs397441@dal.ca
B00886157

MATCH (a:SnowNode), (b:snow) WHERE a.name="snow" AND b.typeOfSnowFall="Snow Burst" CREATE (b)-[r:TYPE_OF]->(a)

//Generating relationships with flu nodes

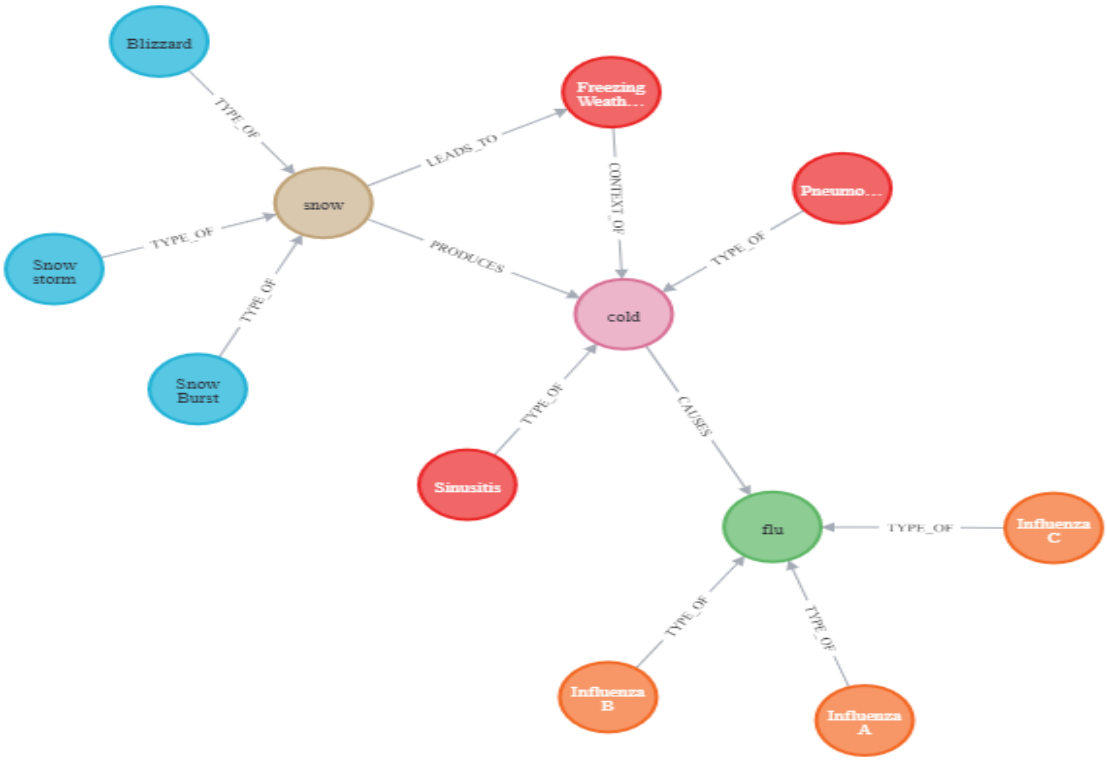
MATCH (a:FluNode), (b:flu) WHERE a.name="flu" AND b.name="Influenza A" CREATE (b)-[r:TYPE_OF]->(a)
MATCH (a:FluNode), (b:flu) WHERE a.name="flu" AND b.name="Influenza B" CREATE (b)-[r:TYPE_OF]->(a)
MATCH (a:FluNode), (b:flu) WHERE a.name="flu" AND b.name="Influenza C" CREATE (b)-[r:TYPE_OF]->(a)

//Generating overall relationships

MATCH (a:SnowNode), (b:ColdNode) WHERE a.name="snow" AND b.name="cold" CREATE (a)-[r:PRODUCES]->(b)
MATCH (a:FluNode), (b:ColdNode) WHERE a.name="flu" AND b.name="cold" CREATE (b)-[r:CAUSES]->(a)

MATCH (a:SnowNode), (b:cold) WHERE a.name="snow" AND b.name="Freezing Weather" CREATE (a)-[r:LEADS_TO]->(b)

Graph:



Rushi Samirbhai Patel rs397441@dal.ca
B00886157

References:

[1] "Apache Spark Word Count Example - Javatpoint", *www.javatpoint.com*, 2022. [Online]. Available: https://www.javatpoint.com/apache-spark-word-count-example. [Accessed: 13- Mar- 2022].