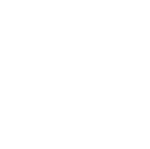
**B.M.S. COLLEGE OF ENGINEERING BENGALURU**

Autonomous Institute, Affiliated to VTU



Lab Record

**Big Data Analytics**

*Submitted in partial fulfillment for the 6th Semester Laboratory*

Bachelor of Technology

in

Computer Science and Engineering

*Submitted by:*

**Ankit Kesar**

1BM18CS150

Department of Computer Science and Engineering

B.M.S. College of Engineering

Bull Temple Road, Basavanagudi, Bangalore 560 019

Mar-June 2021

**B.M.S. College of EngineerinG**

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

****

***CERTIFICATE***

This is to certify that the Big Data Analytics (20CS6PEBDA) laboratory has been carried out by Ankit Kesar (1BM18CS150) during the 6th Semester Mar-June-2021.

Signature of the Faculty Incharge:

Bhoomika :

Department of Computer Science and Engineering

B.M.S. College of Engineering, Bangalore

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| Sr. No | Programs | Page No. |
| 1. | Cassandra  (Program 1 – Program 2) | 4 - 7 |
| 2. | MongoDB  (Program 3) | 7 - 10 |
| 3. | Hadoop  (Program 4 – Program 5) | 11 – 15 |
| 4. | Map Reduce  (Program 6 – Program 8) | 16 – 20 |
| 5. | Spark  (Program 9 – Program 10) | 21 – 26 |

**Program 1.** Perform the following DB operations using Cassandra.  
  
1. Create a keyspace by name Employee  
2. Create a column family by name Employee-Info with attributes  
Emp\_Id Primary Key, Emp\_Name, Designation, Date\_of\_Joining, Salary,  
Dept\_Name  
3. Insert the values into the table in batch  
4. Update Employee name and Department of Emp-Id 121  
5. Sort the details of Employee records based on salary  
6. Alter the schema of the table Employee\_Info to add a column Projects which stores a set of Projects done by the corresponding Employee.  
7. Update the altered table to add project names.  
8 Create a TTL of 15 seconds to display the values of Employees.

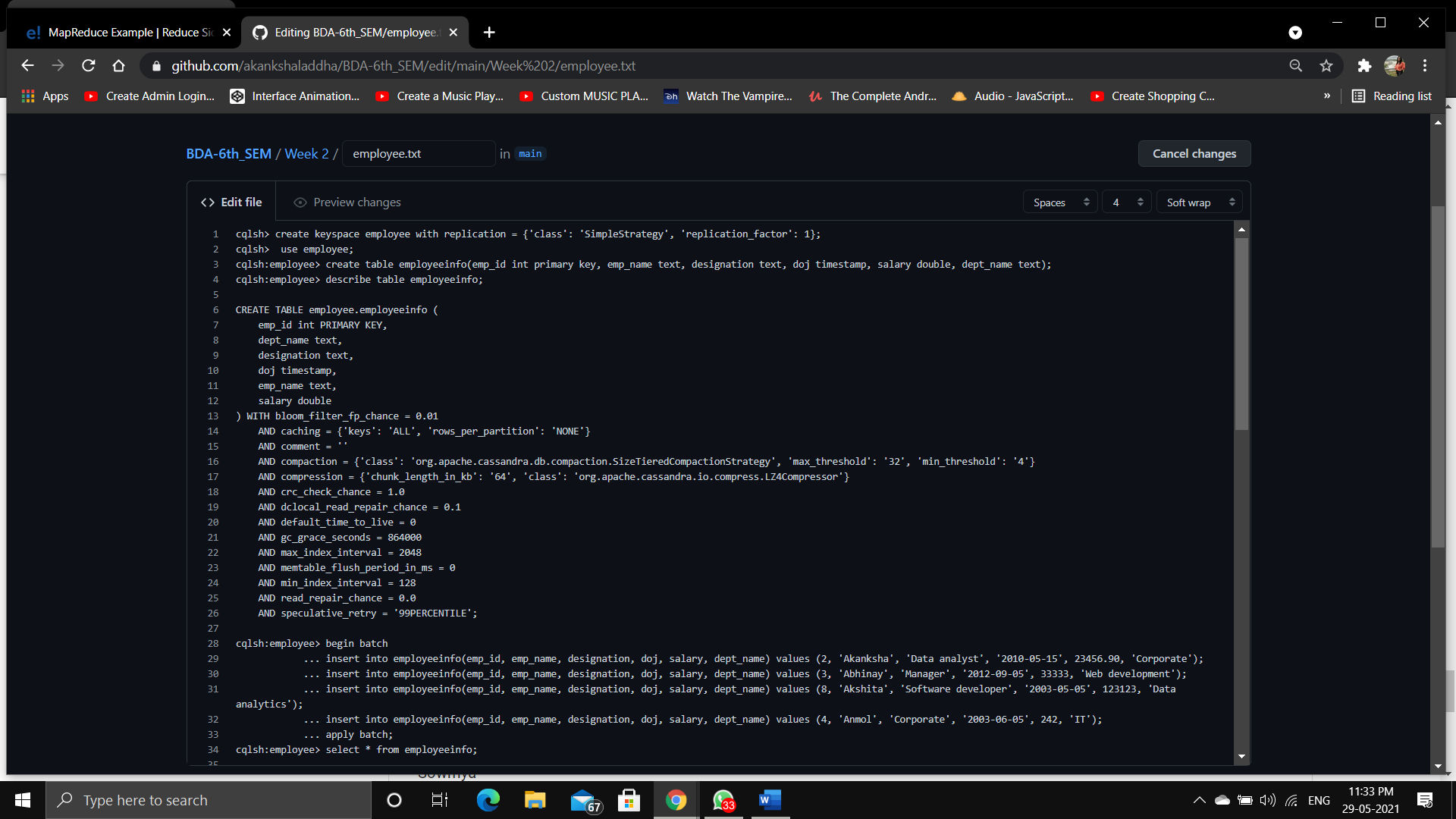


Fig 1.1



Fig 1.2

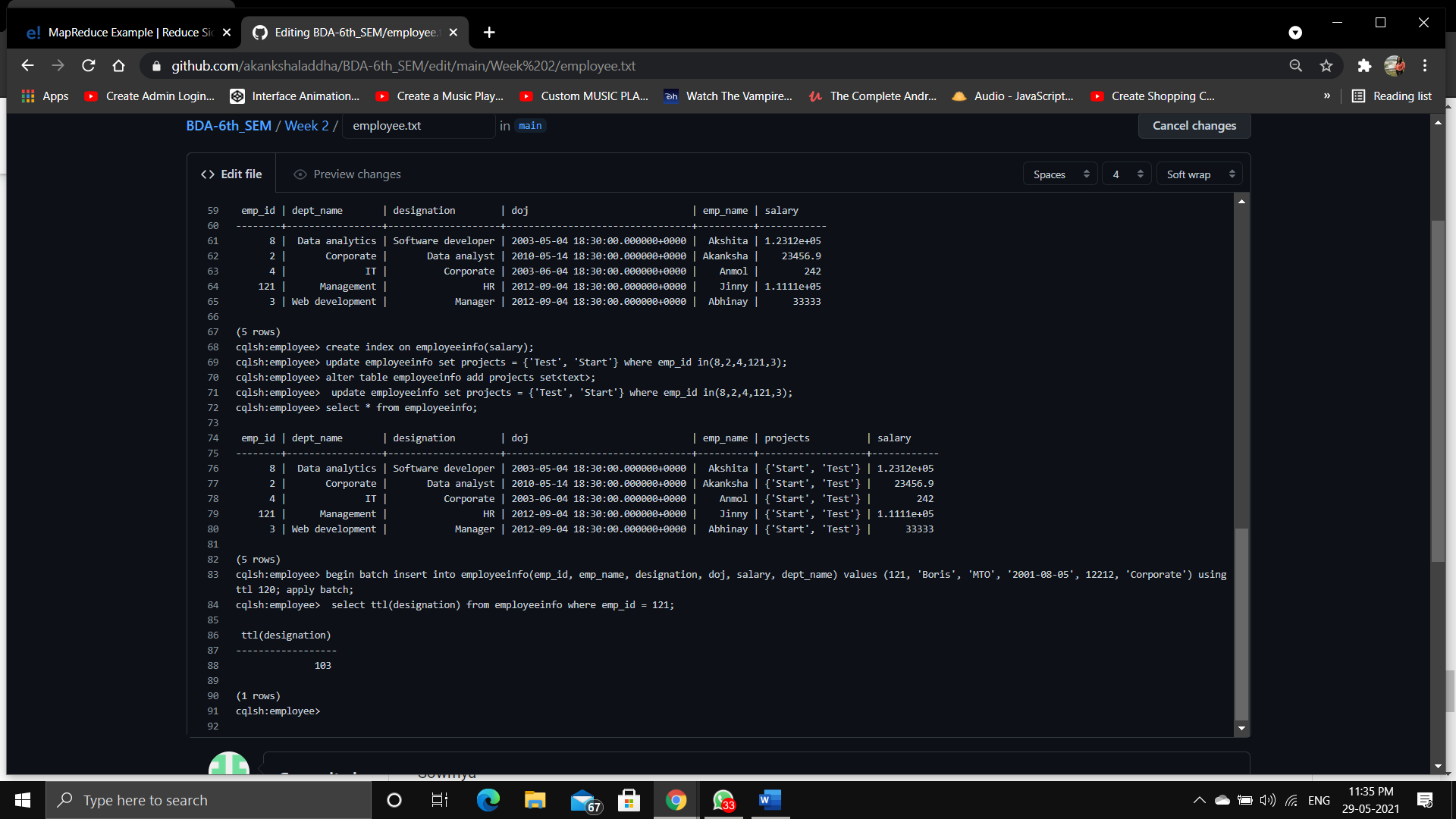


Fig 1.3

**Program 2.**  Perform the following DB operations using Cassandra.

1 Create a keyspace by name Library  
2. Create a column family by name Library-Info with attributes  
Stud\_Id Primary Key,  
Counter\_value of type Counter,  
Stud\_Name, Book-Name, Book-Id, Date\_of\_issue  
3. Insert the values into the table in batch  
4. Display the details of the table created and increase the value of the counter  
5. Write a query to show that a student with id 112 has taken a book “BDA” 2 times.  
6. Export the created column to a csv file  
7. Import a given csv dataset from local file system into Cassandra column family

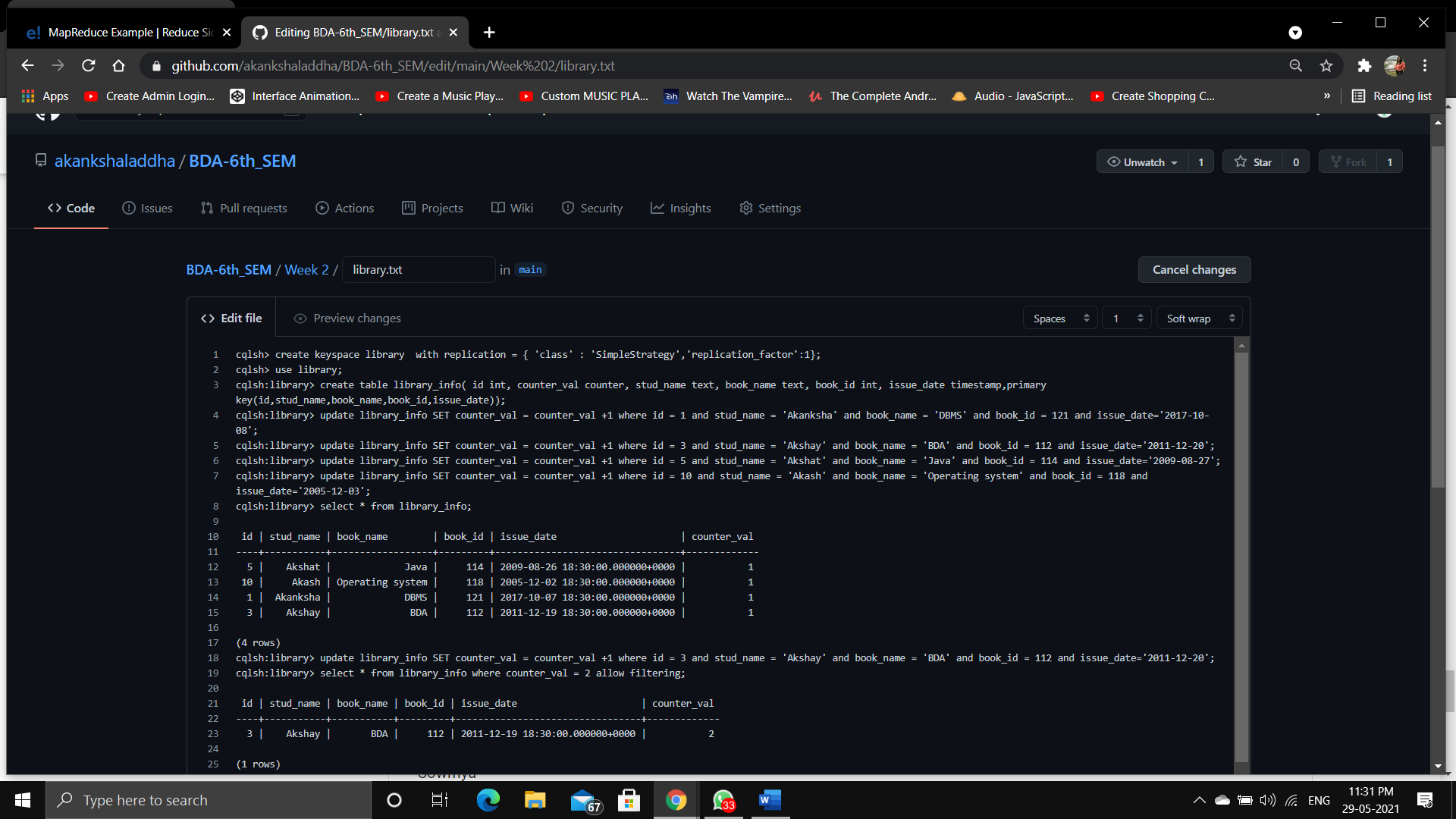


Fig 2.1

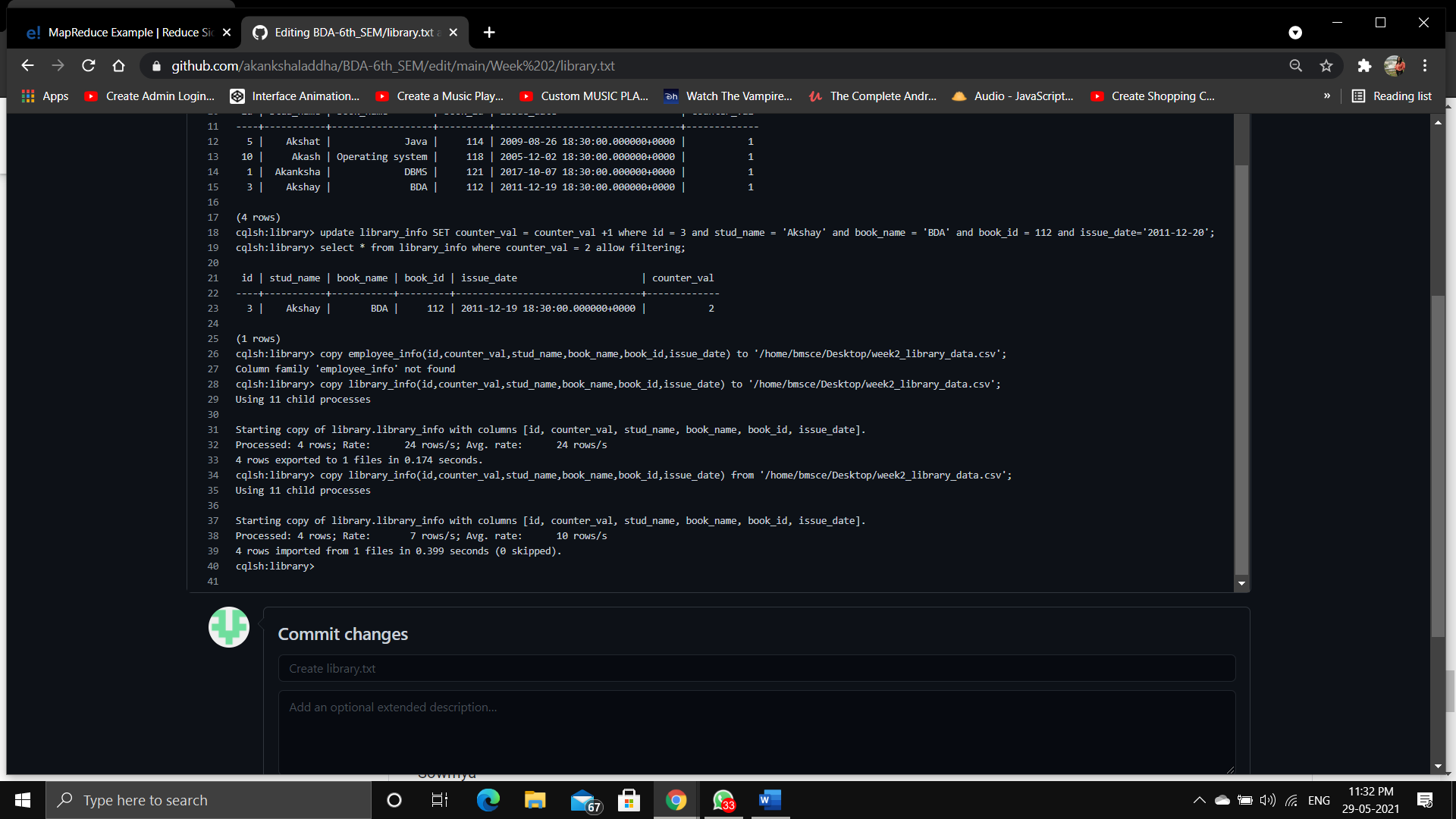


Fig 2.2

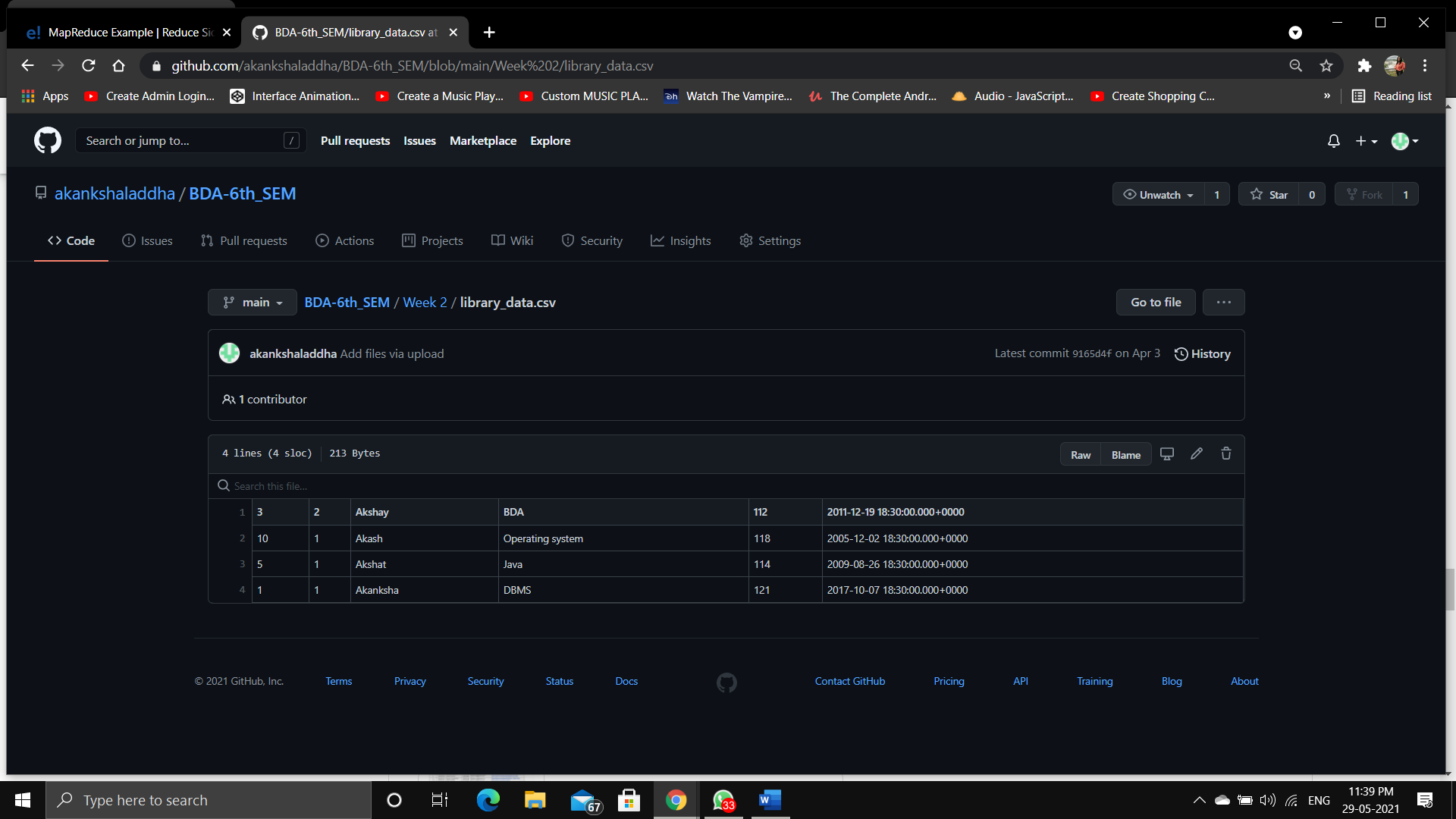


Fig 2.3

**Program 3**. Perform the following DB operations using MongoDB.1. Create a database “Student” with the following attributes Rollno, Age, ContactNo, Email-Id.2. Insert appropriate values3. Write query to update Email-Id of a student with rollno 10.4. Replace the student name from “ABC” to “FEM” of rollno 11.5. Export the created table into local file system6. Drop the table7. Import a given csv dataset from local file system into mongodb collection.

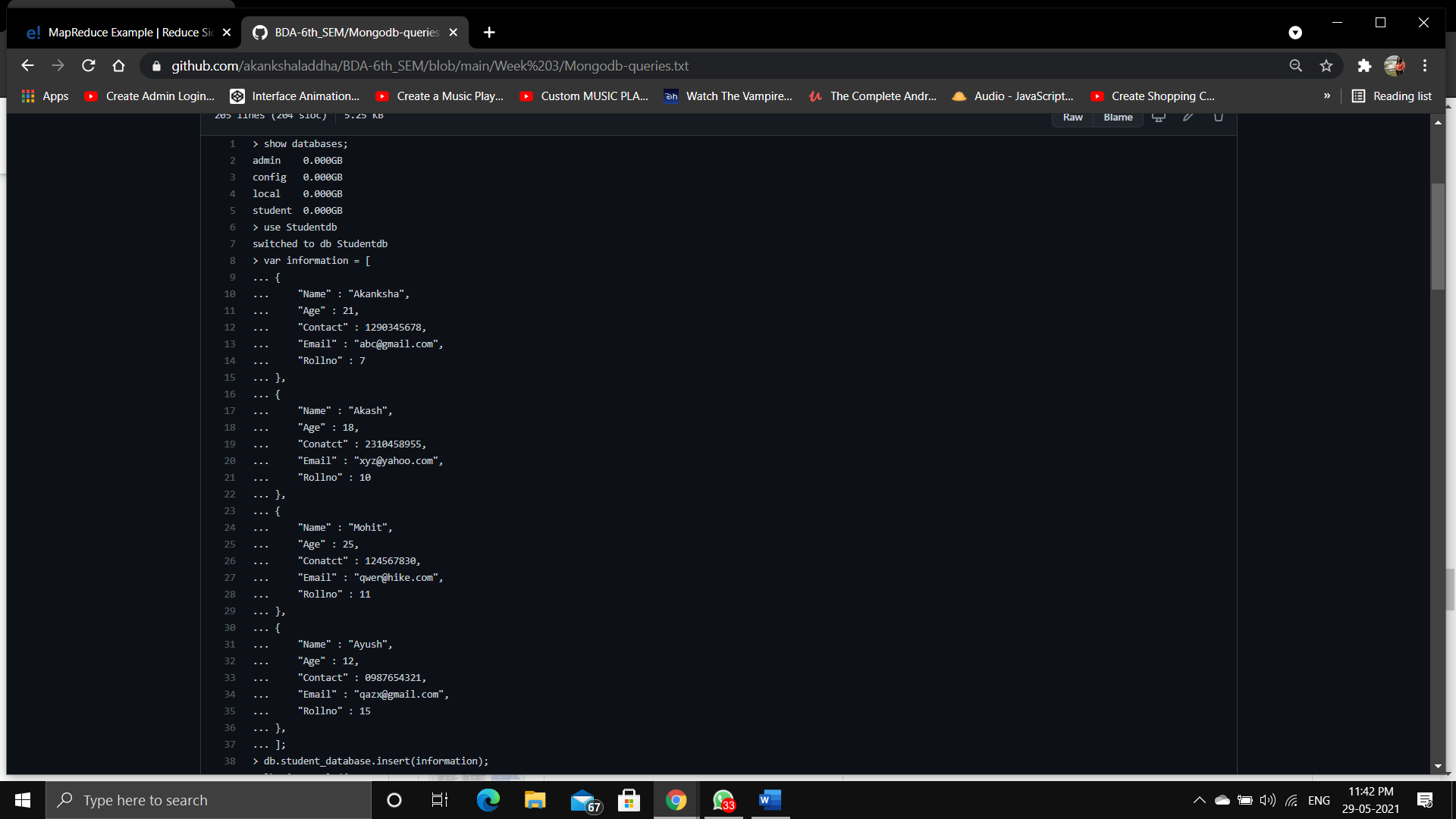


Fig 3.1

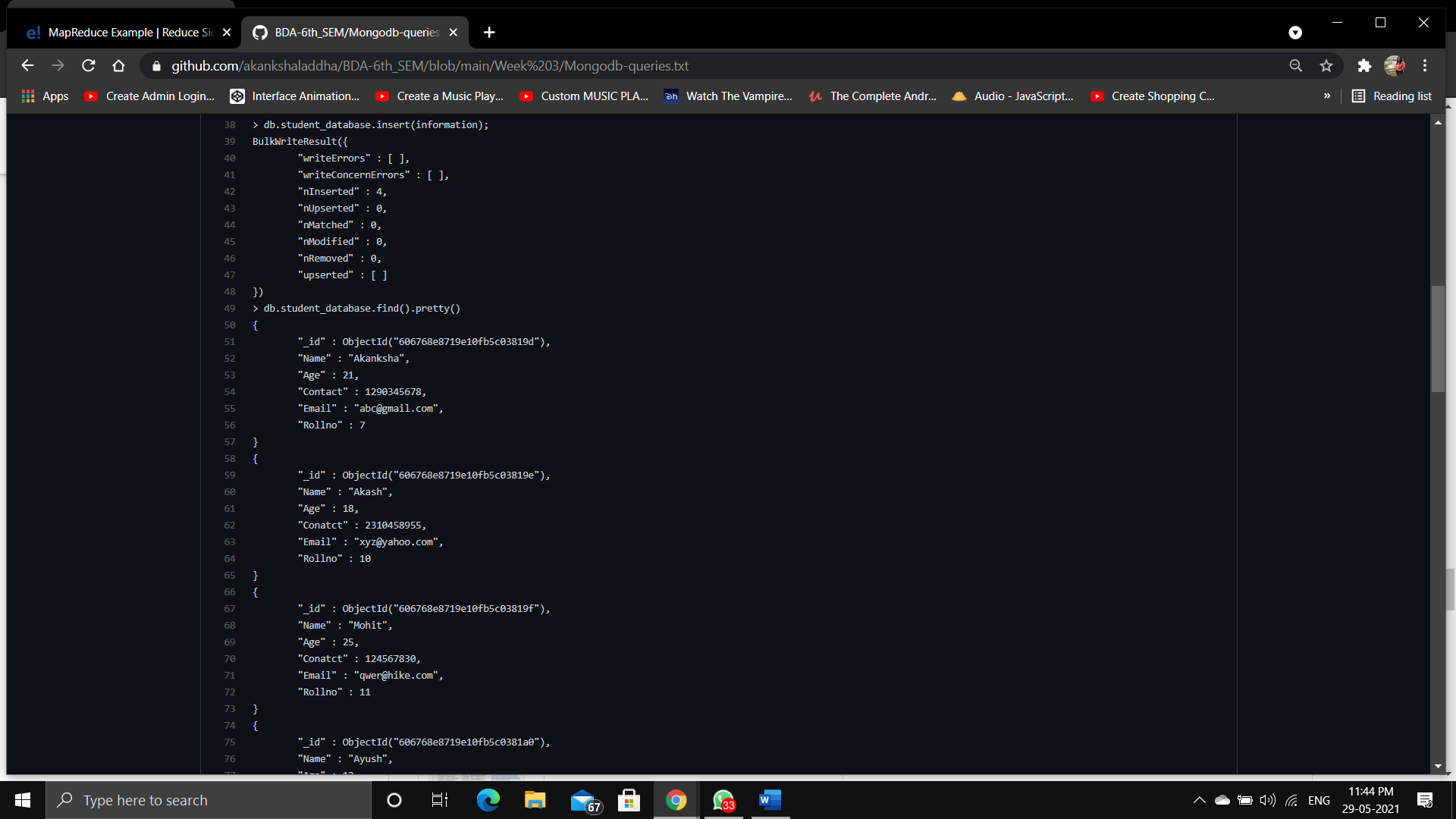


Fig 3.2

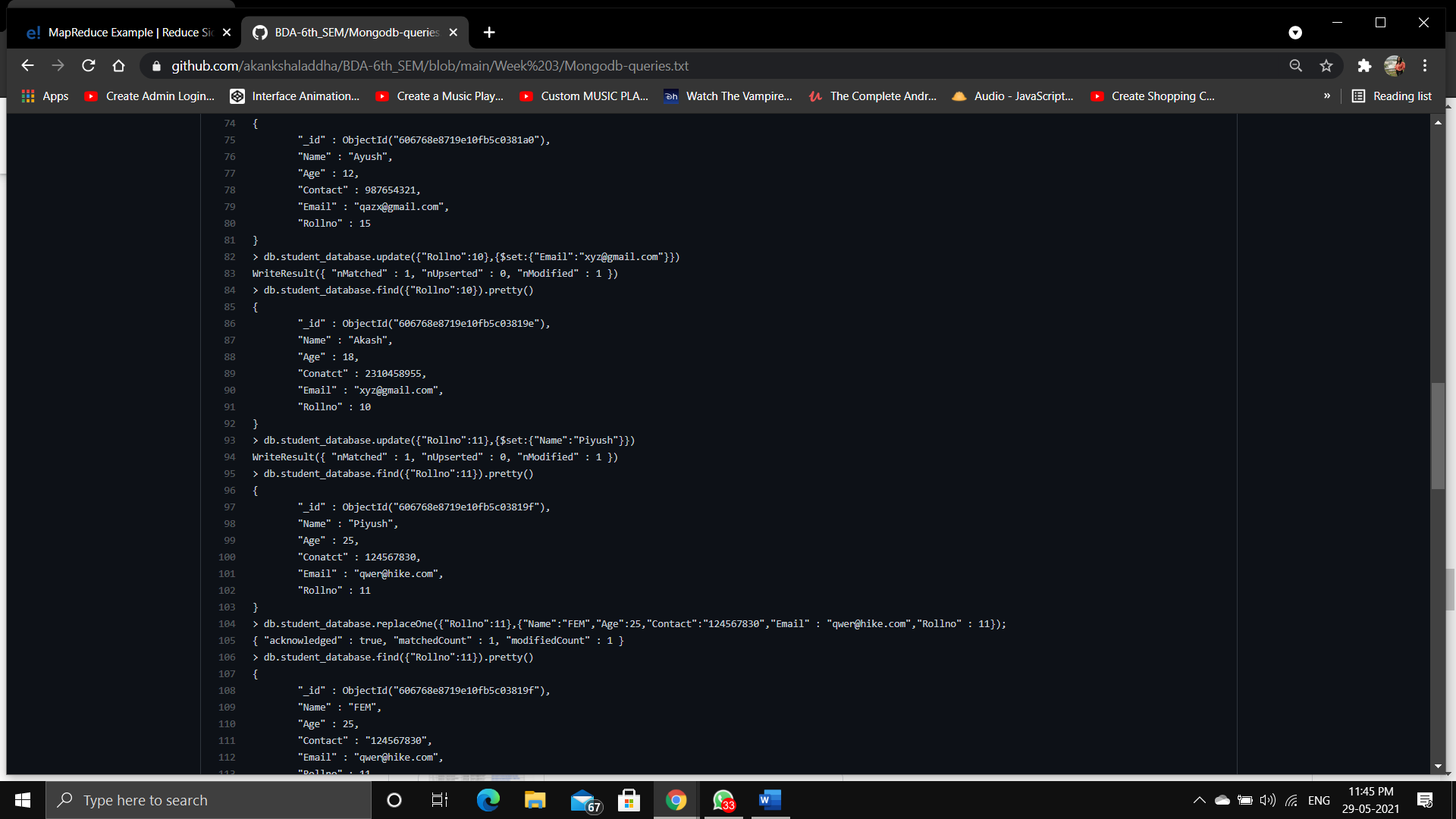


Fig 3.3

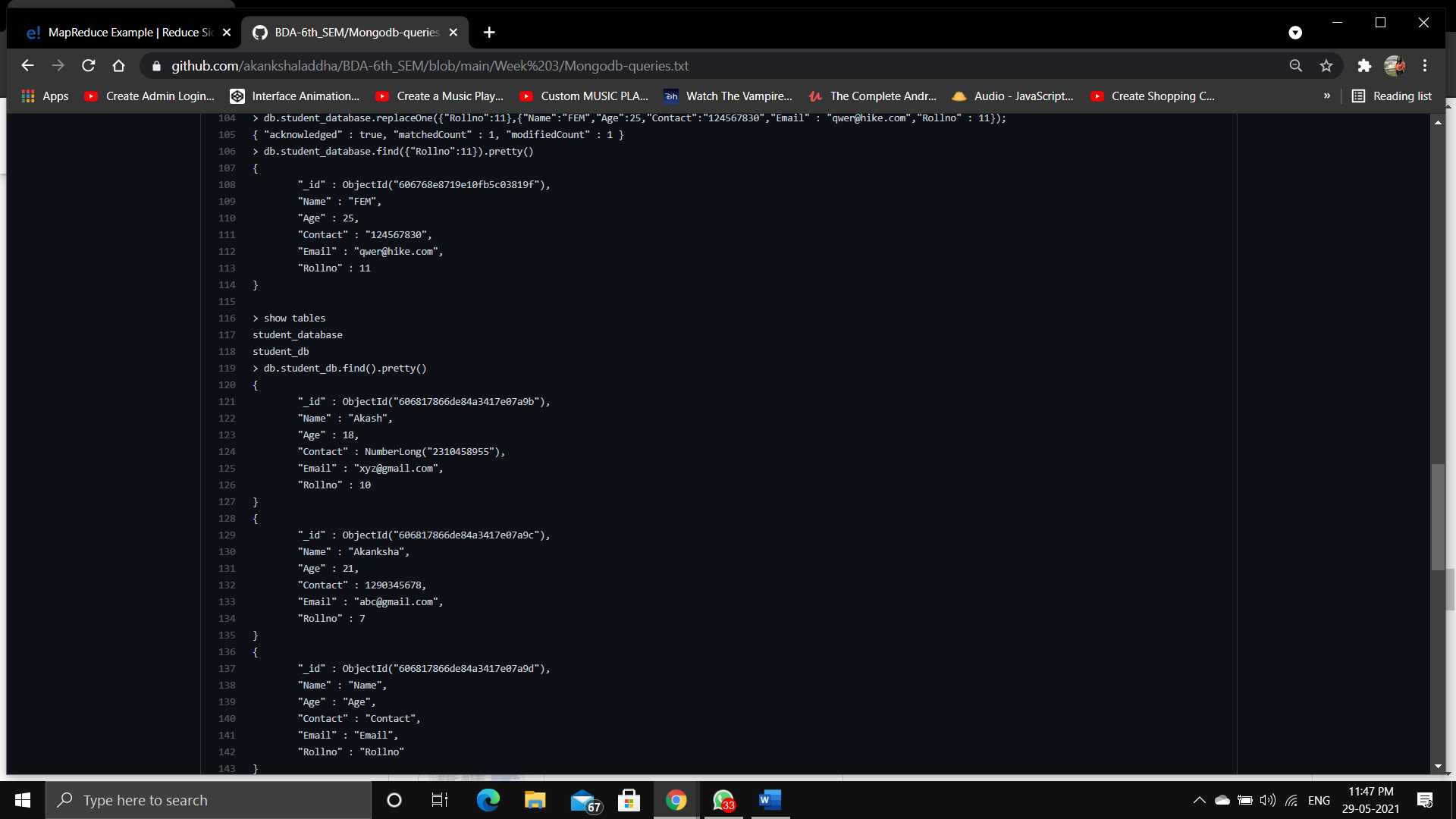


Fig 3.4

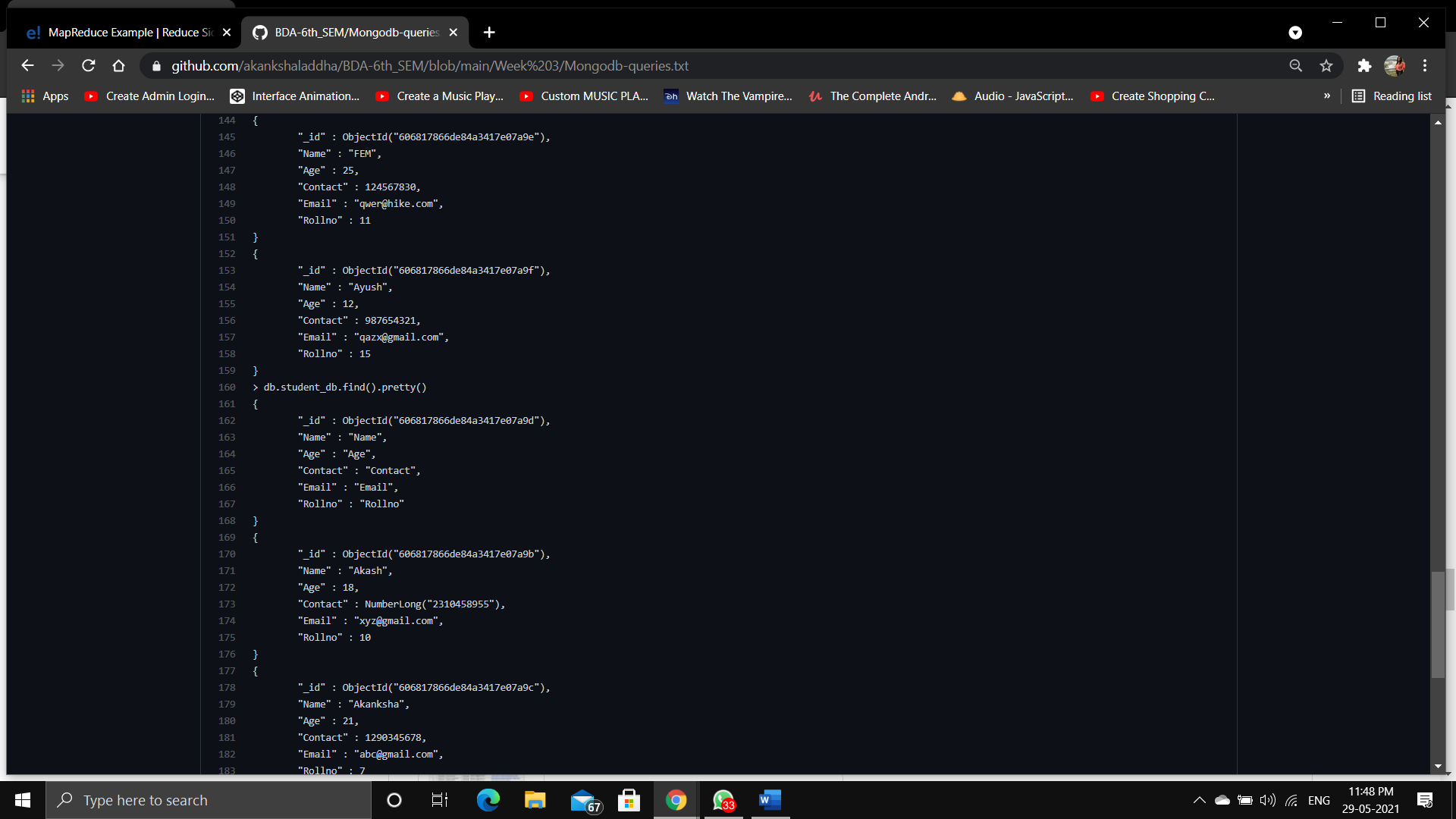


Fig 3.5

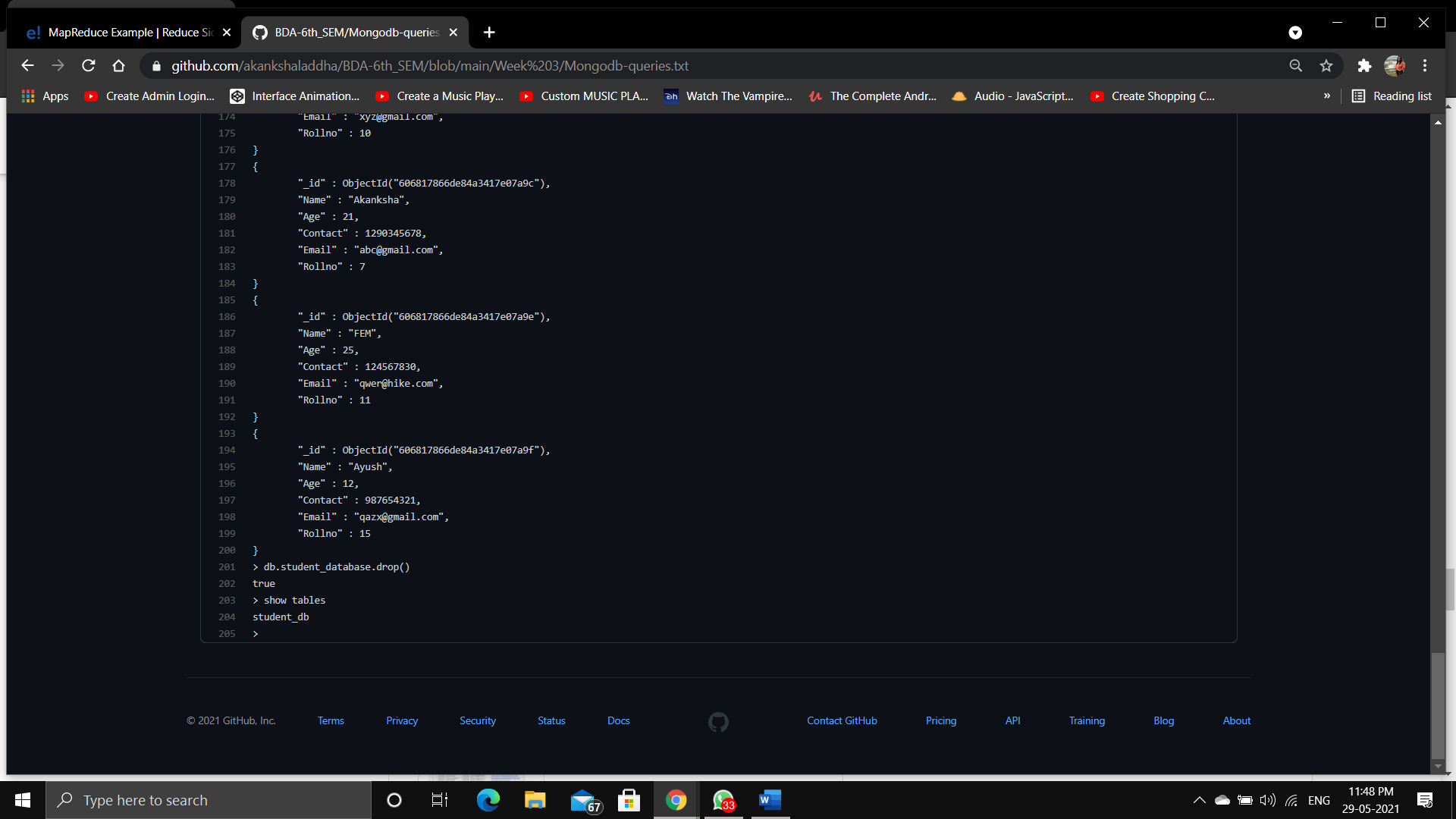


Fig 3.6

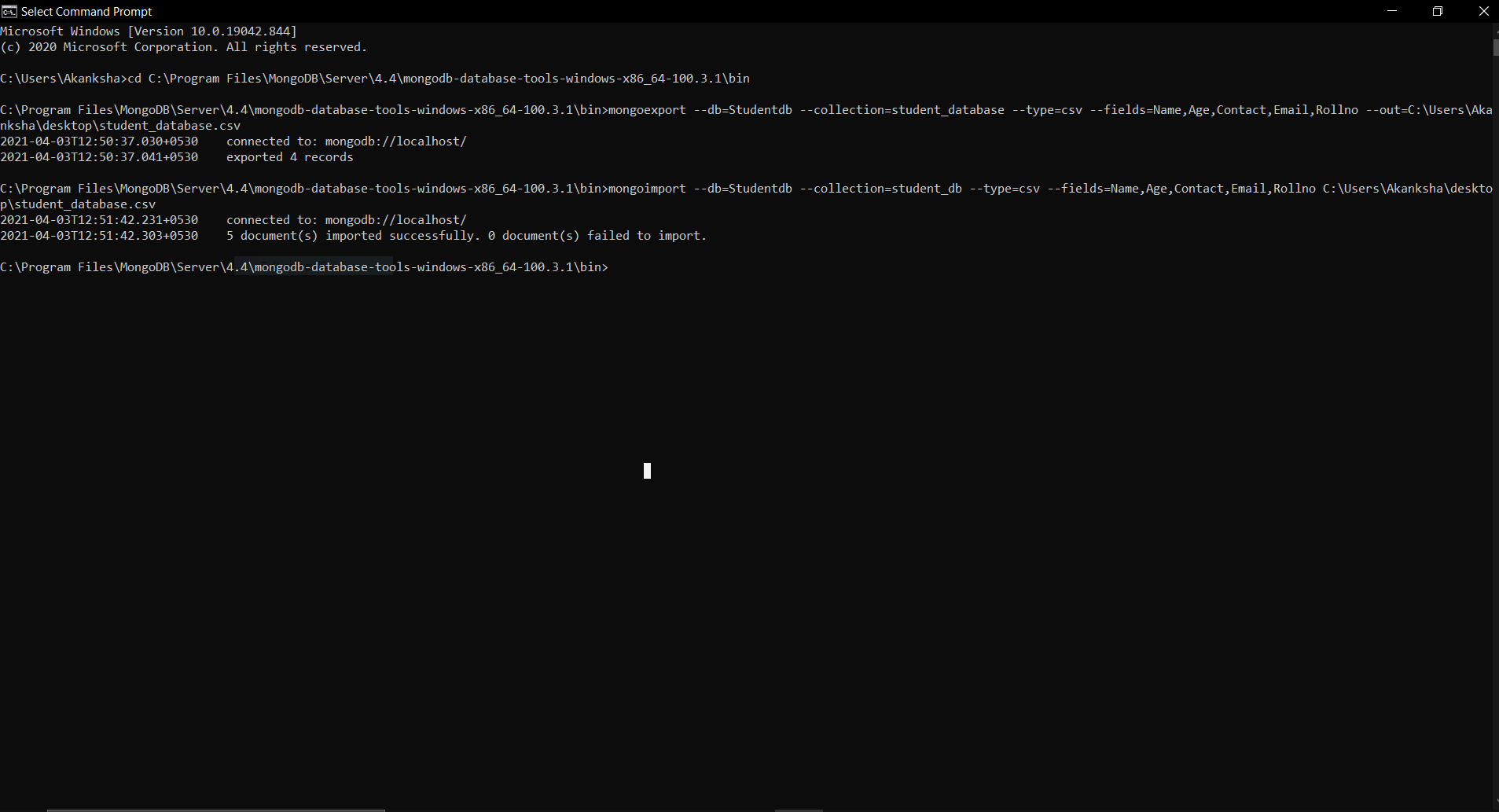


Fig 3.7

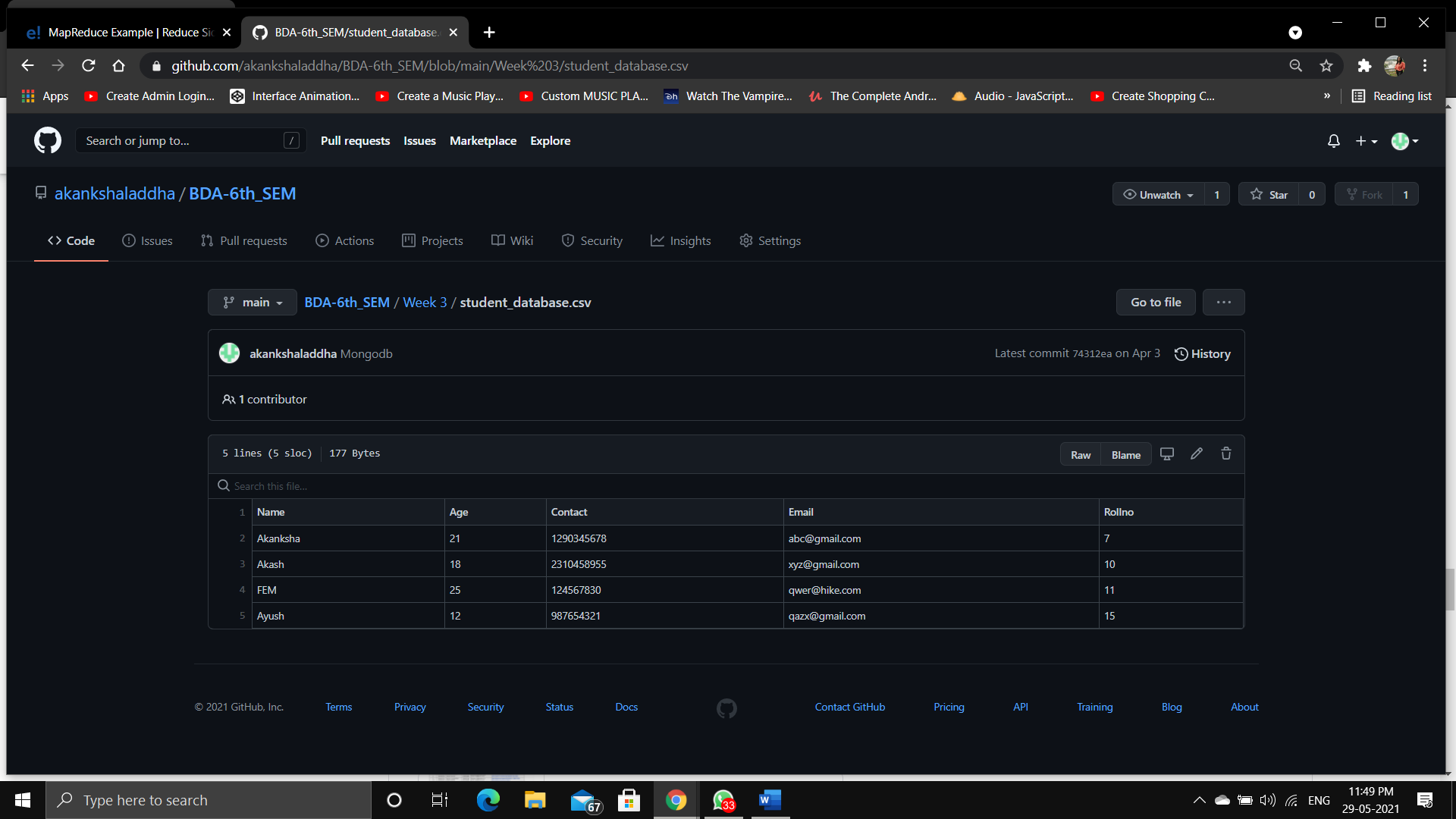


Fig 3.8

**Program 4.** Screenshot of Hadoop installed

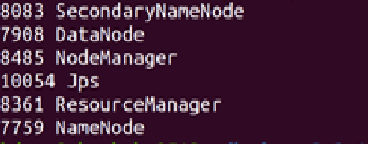
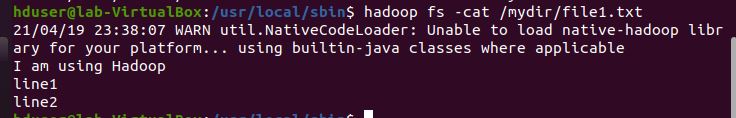
****

Fig 4.1

**Program 5.** Execution of HDFS Commands for interaction with Hadoop Environment. (Minimum 10 commands to be executed).



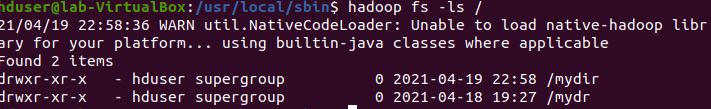
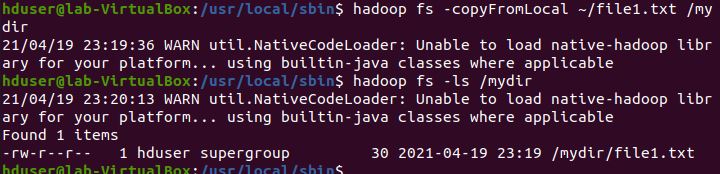


Fig 5.1



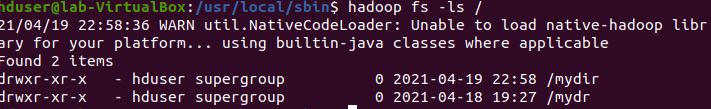
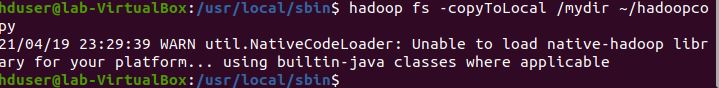


Fig 5.2



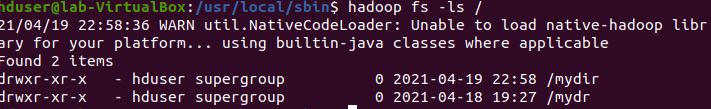


Fig 5.3

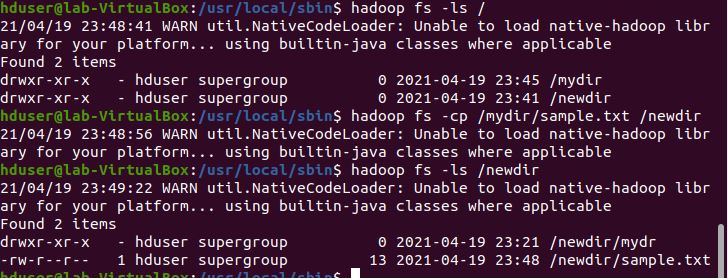


Fig 5.4



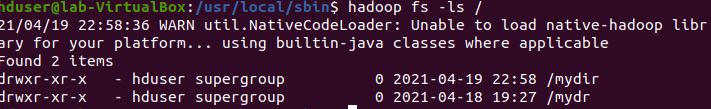


Fig 5.5

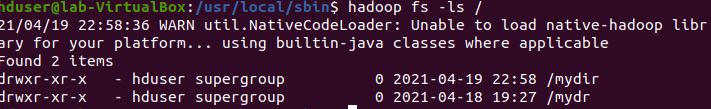


Fig 5.6

**Program 6.** From the following link extract the weather data

<https://github.com/tomwhite/hadoop-book/tree/master/input/ncdc/all> .Create a Map Reduce program to

i) find average temperature for each year from NCDC data set.

ii) find the mean max temperature for every month.

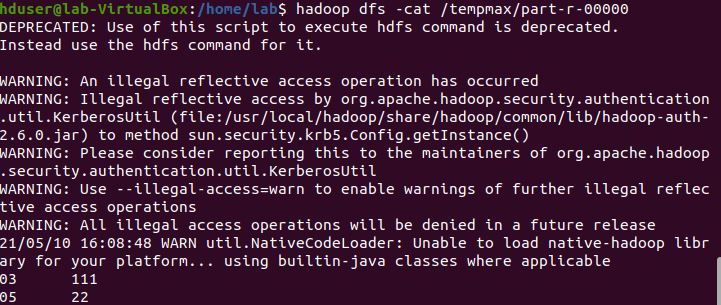


Fig 6.1.1

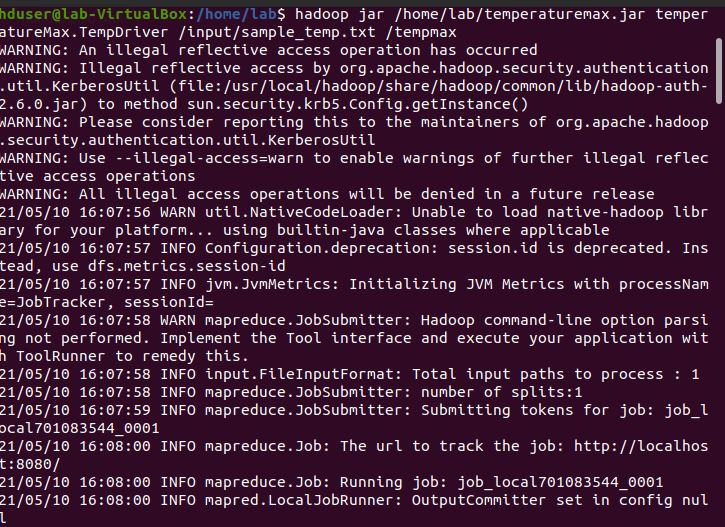


Fig 6.1.2

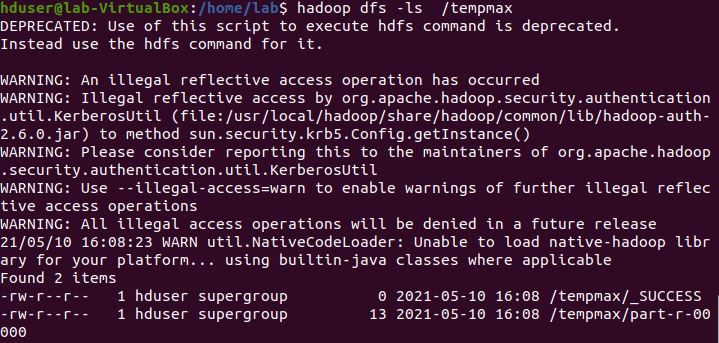


Fig 6.2

**Program 7.** For a given Text file, create a Map Reduce program to sort the

content in an alphabetic order listing only top 10 maximum occurrences of words.

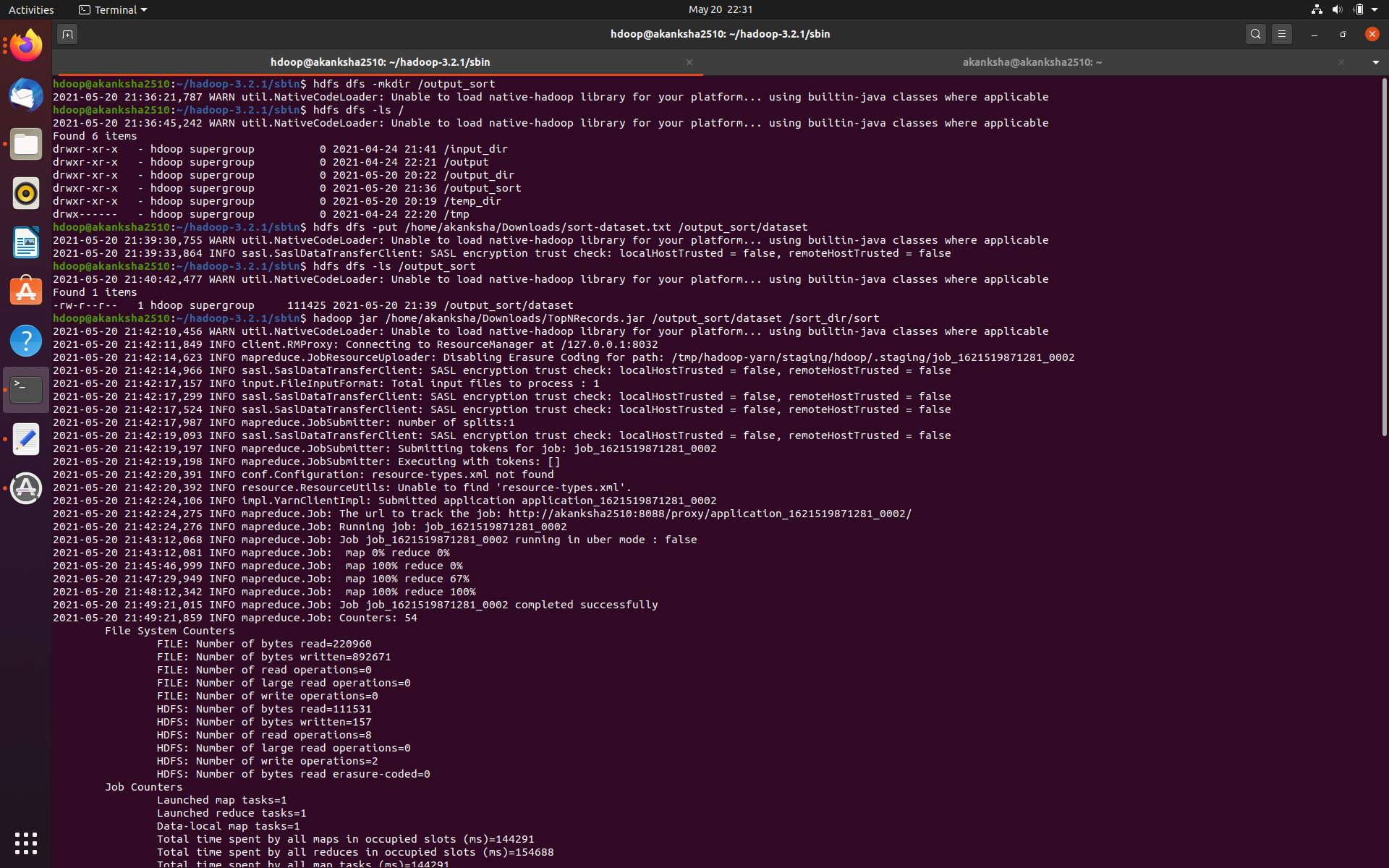


Fig 7.1



Fig 7.2

**Program 8.** Create a Map Reduce program to combine information from the users file along with Information from the posts file by using the concept of join and display user\_id, Reputation and Score**.**

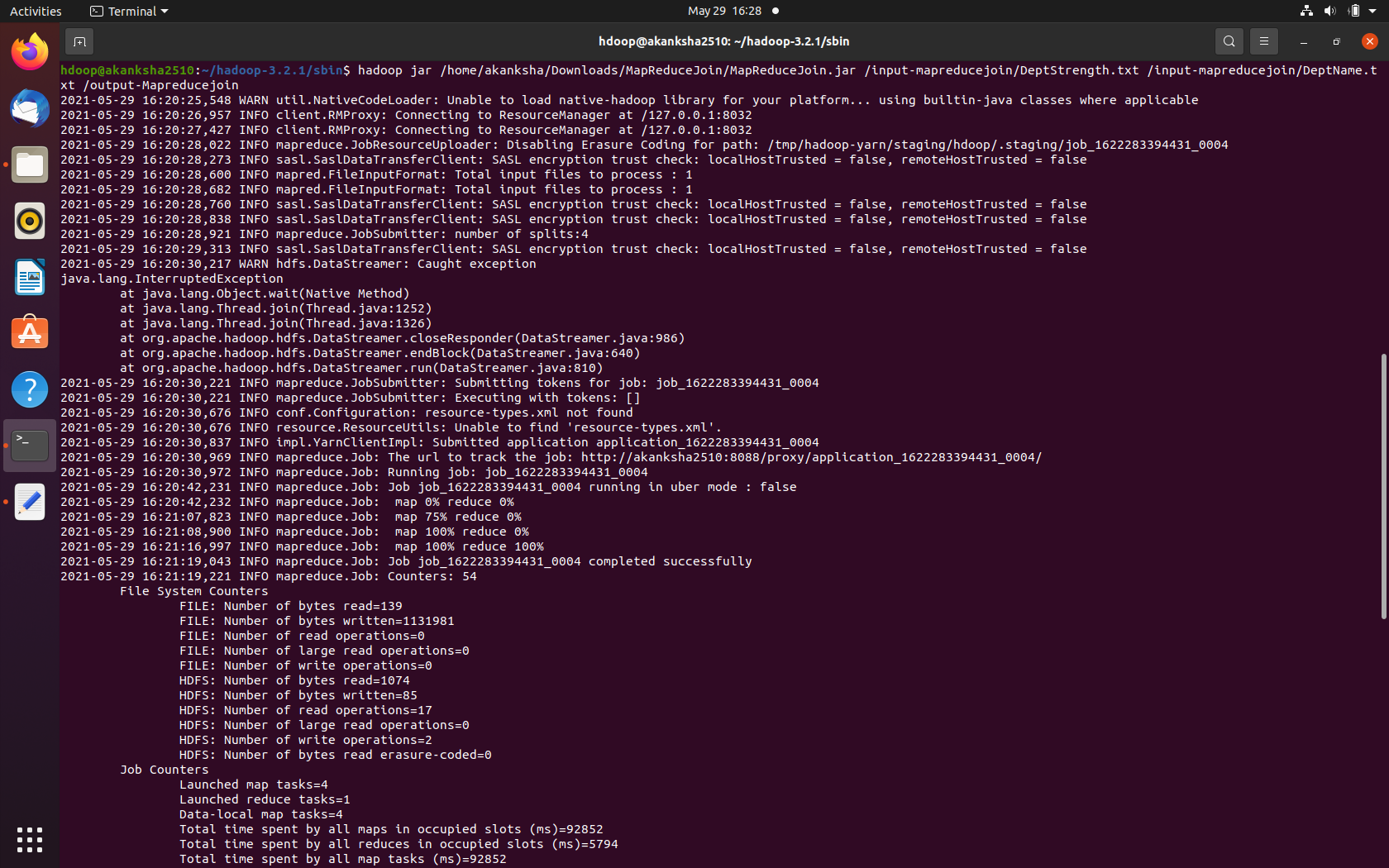


Fig 8.1

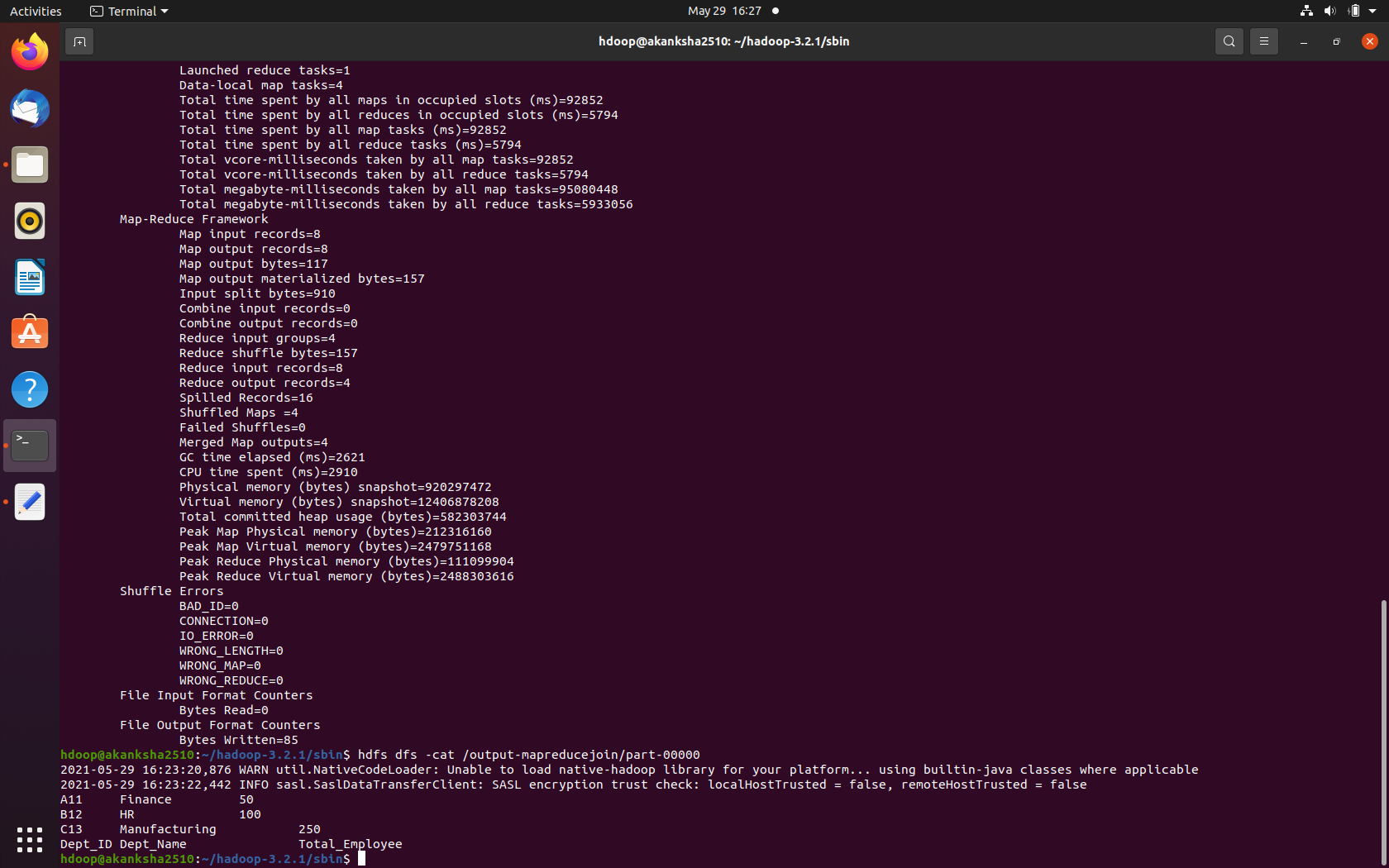


Fig 8.2

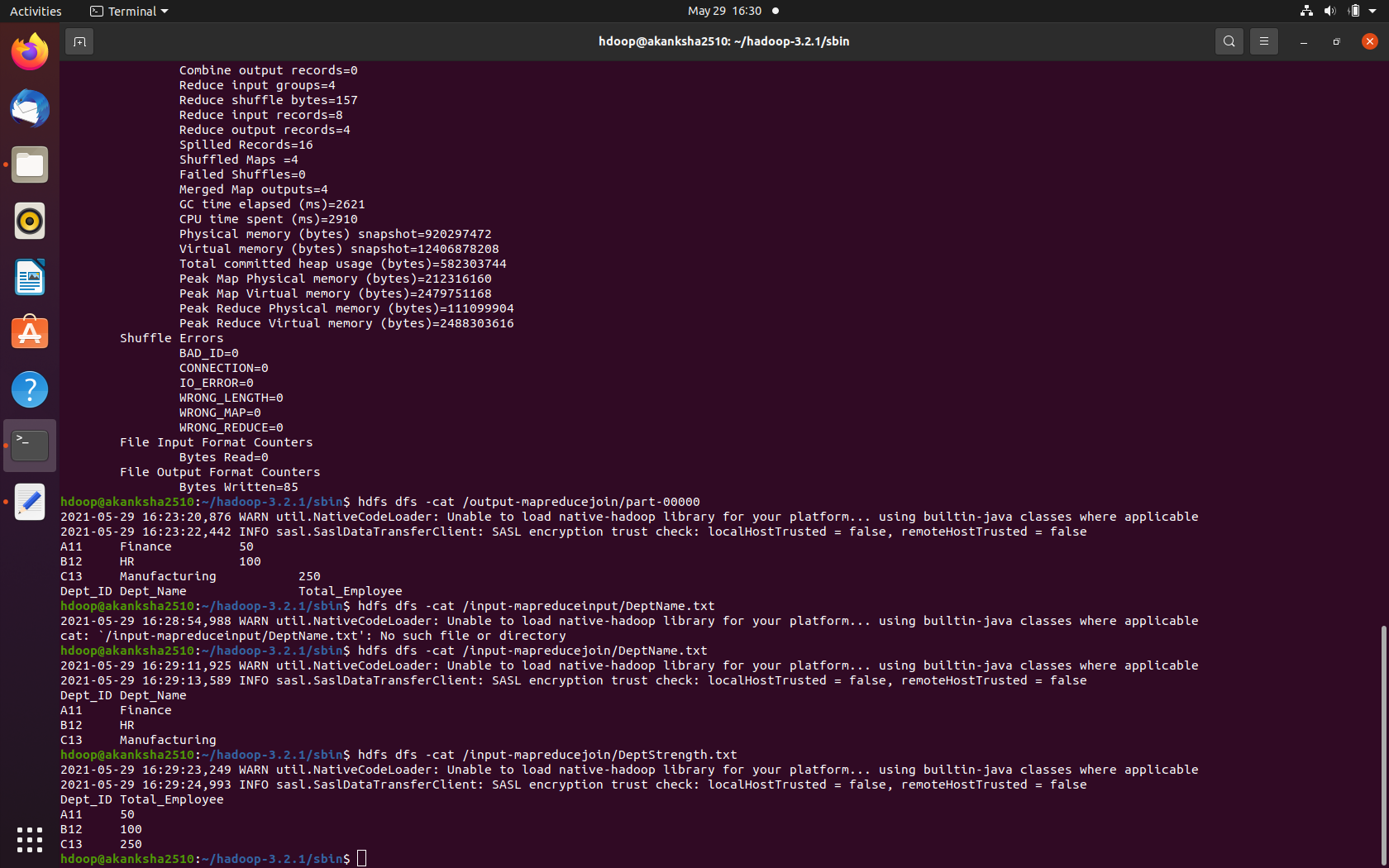


Fig 8.3

**Program 9.**  Scala programs and Screenshot of Spark Installed

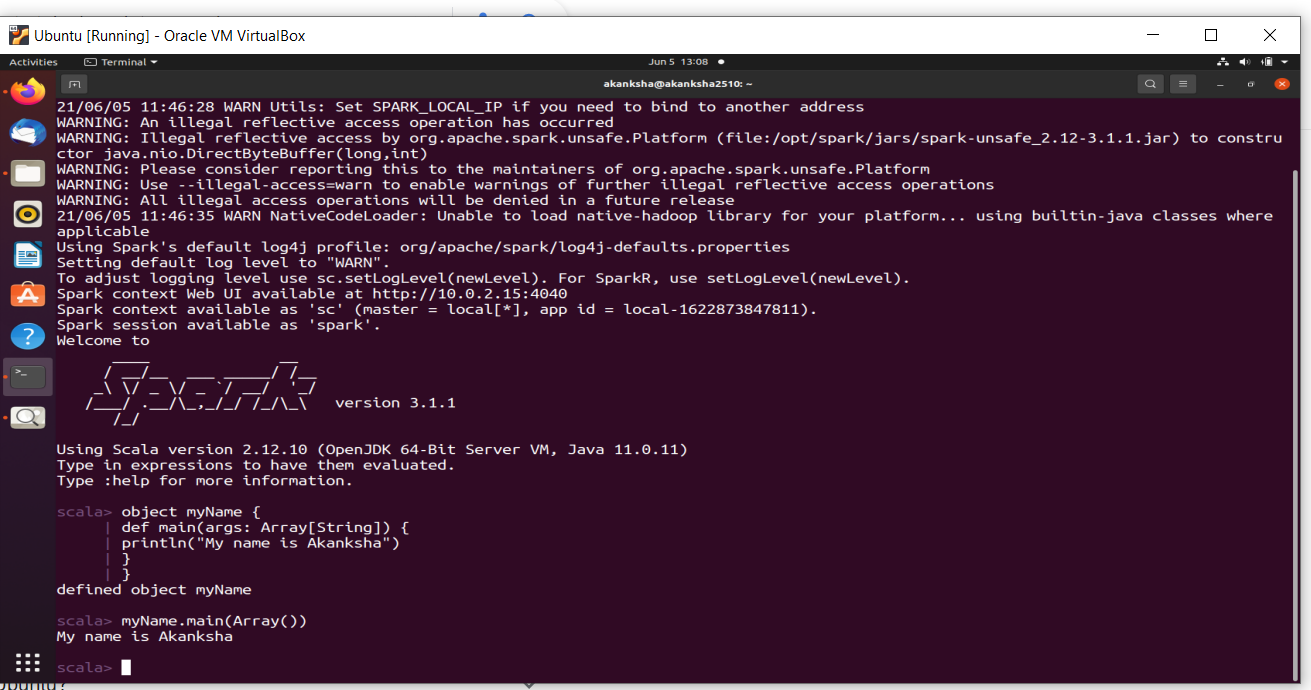


Fig 9.1

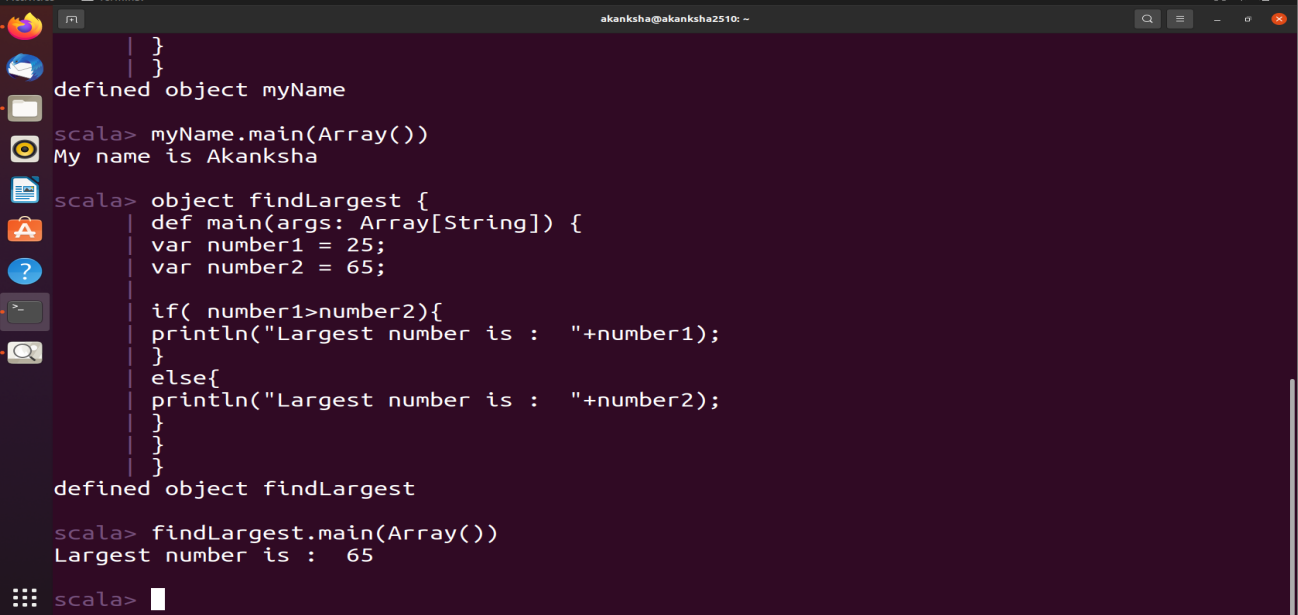


Fig 9.2



Fig 9.3

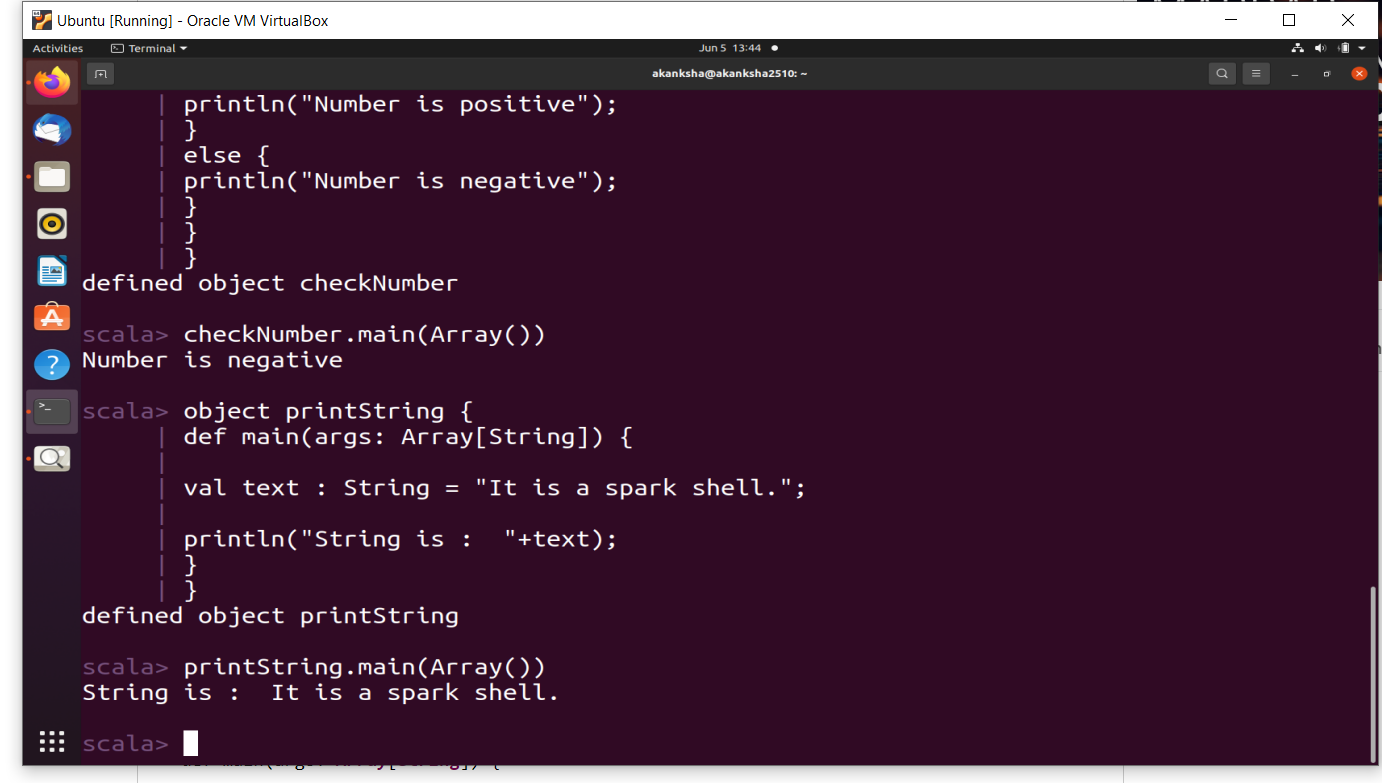


Fig 9.4

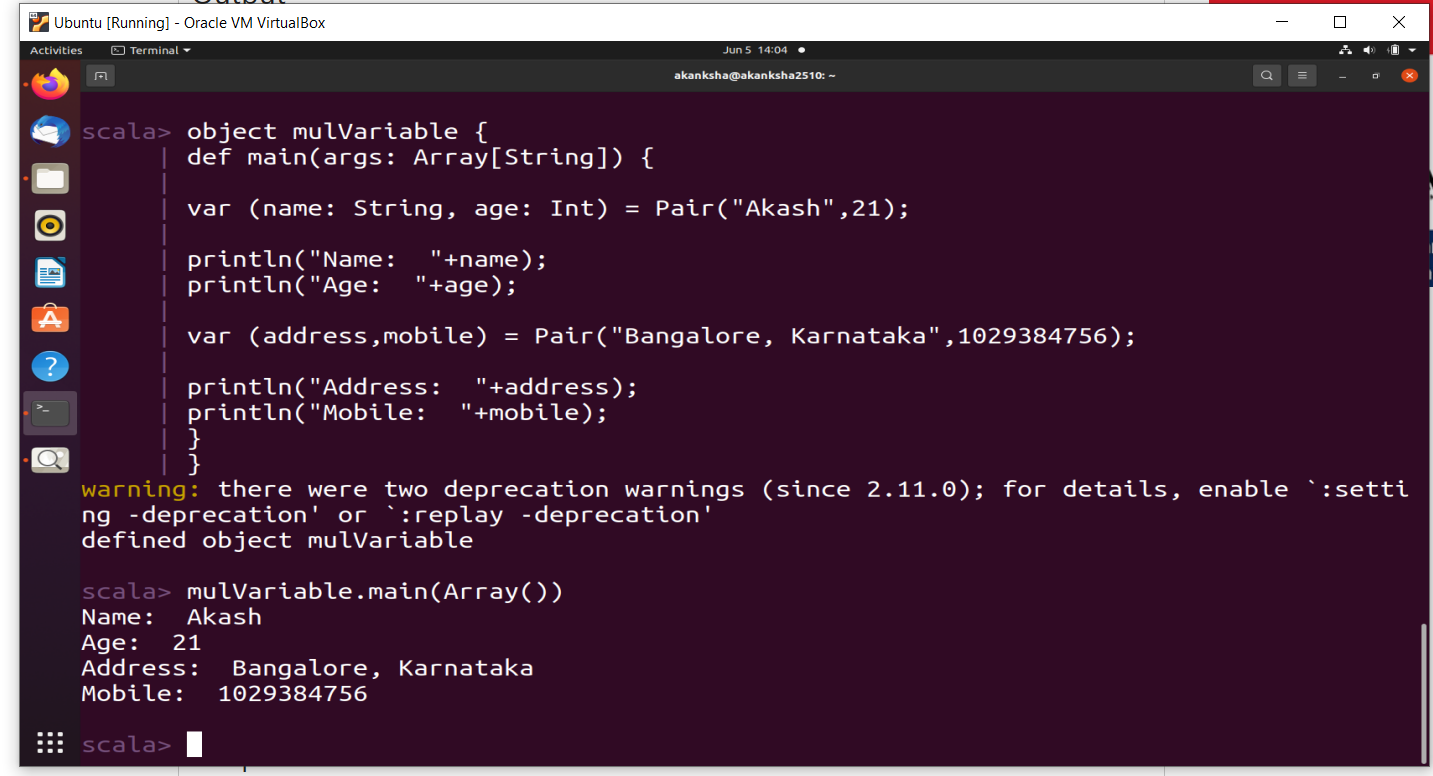


Fig 9.5

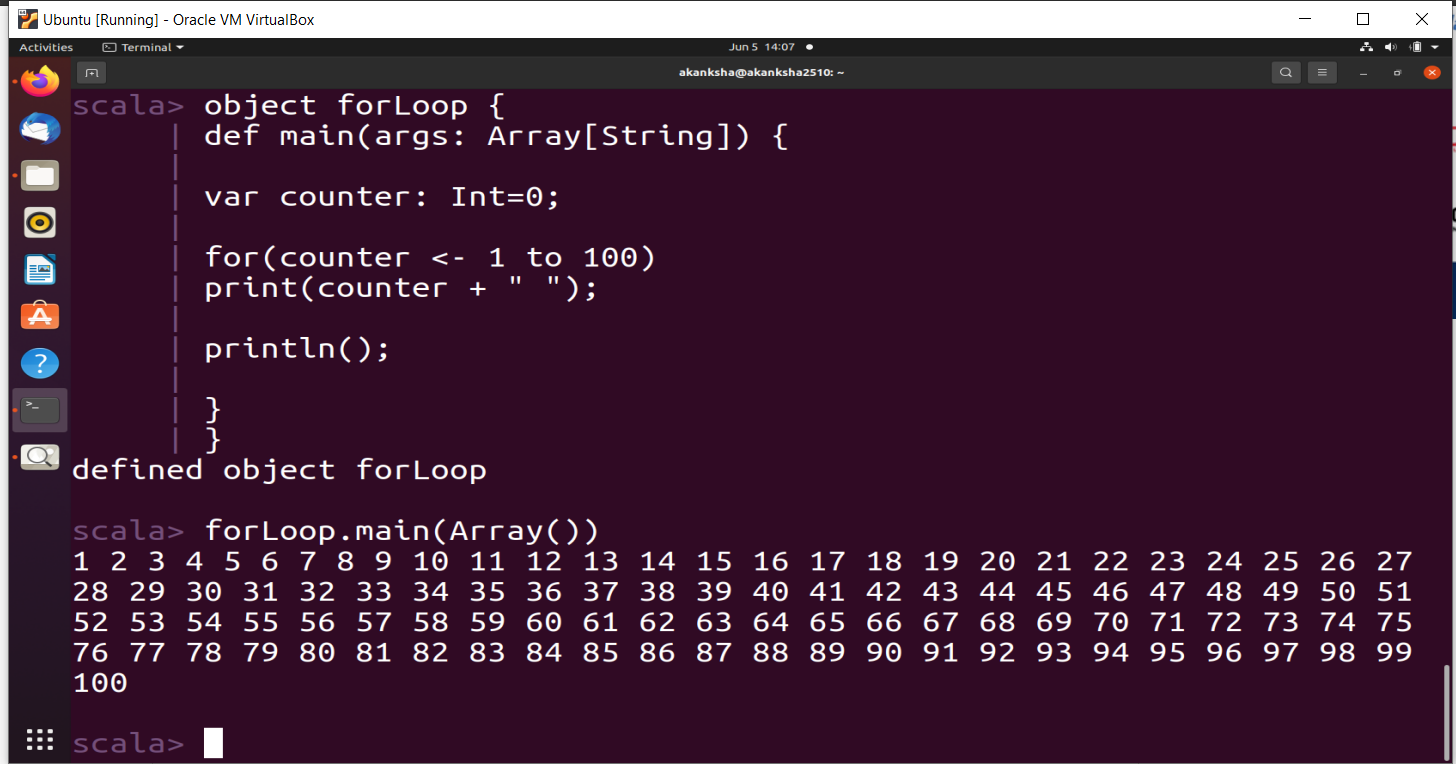


Fig 9.6

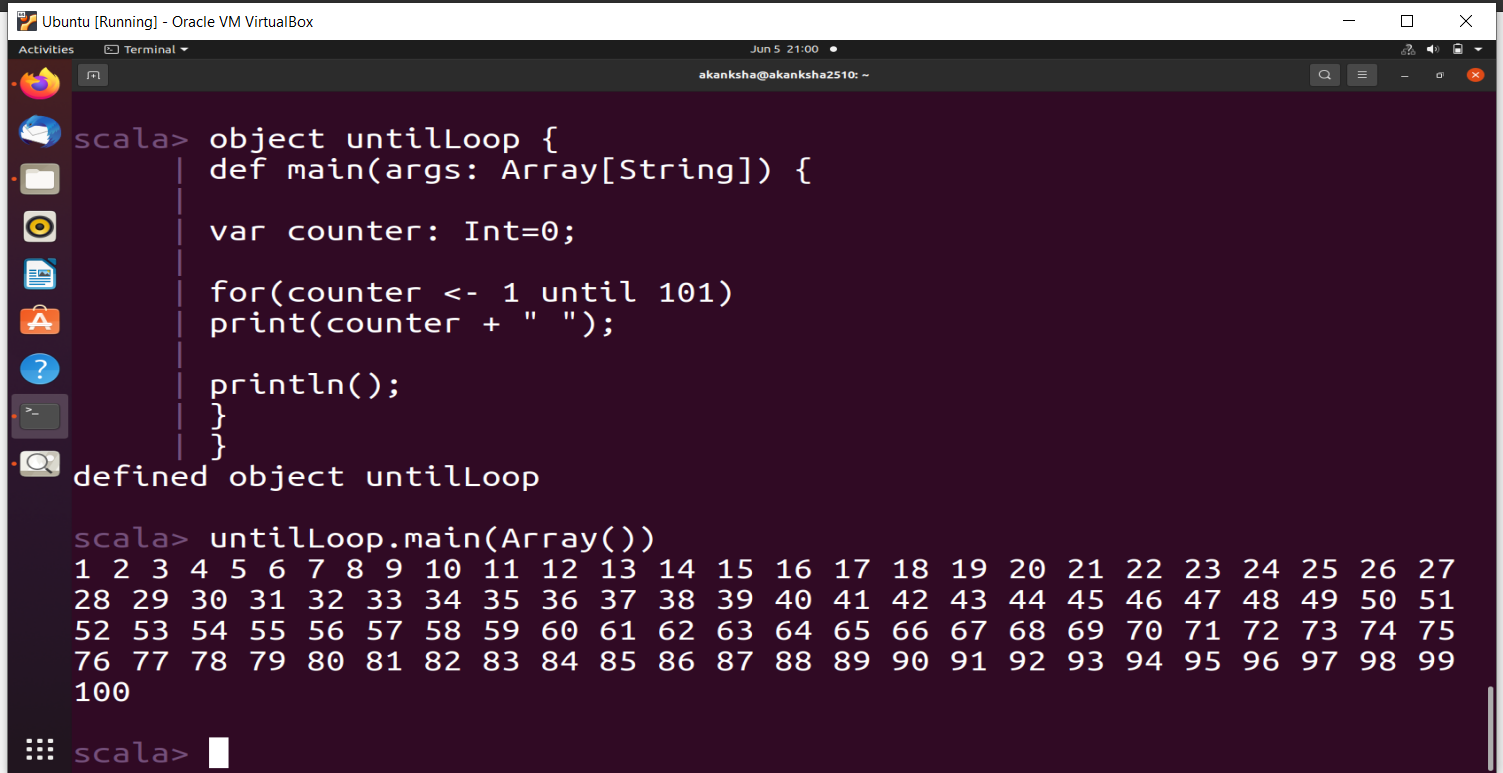


Fig 9.7

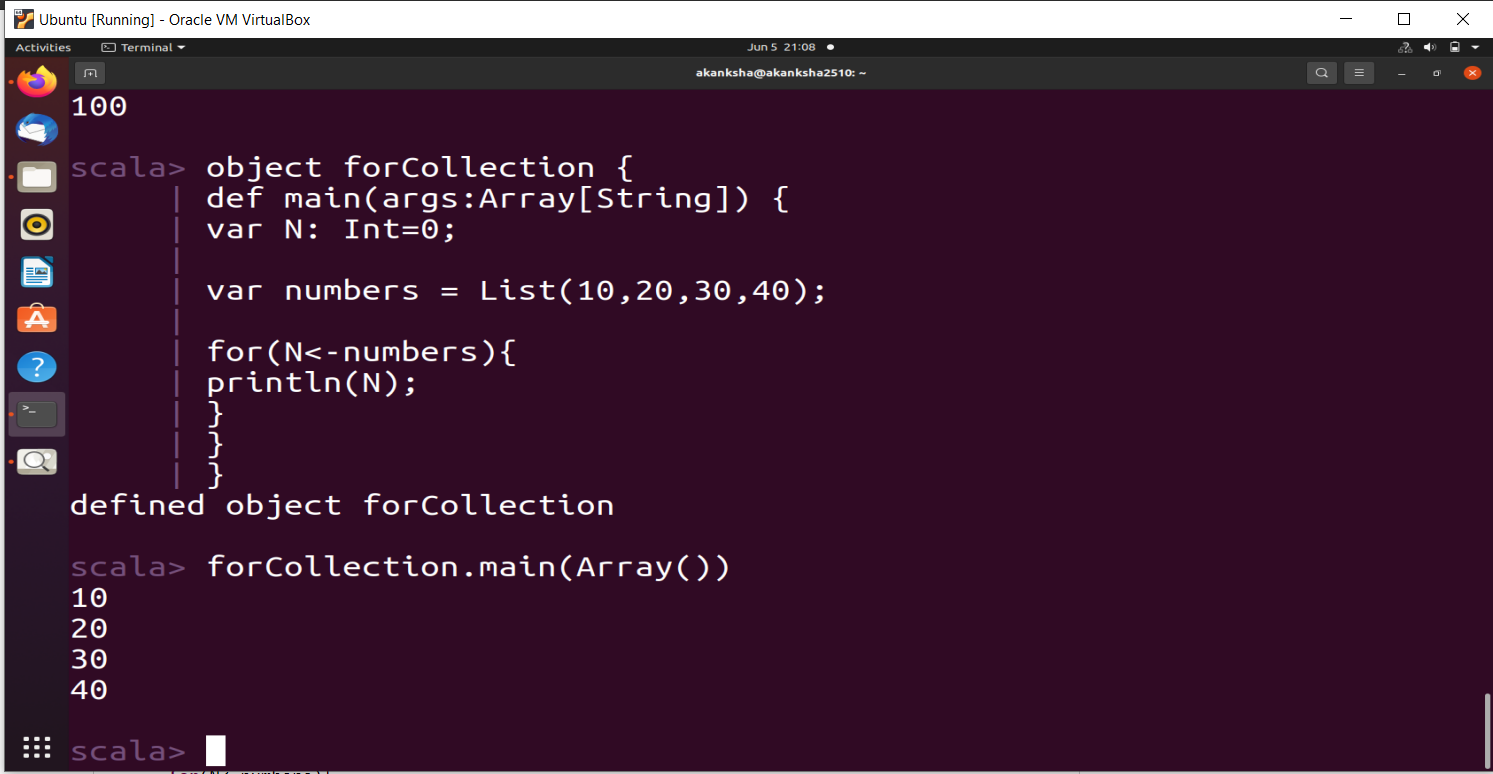


Fig 9.8

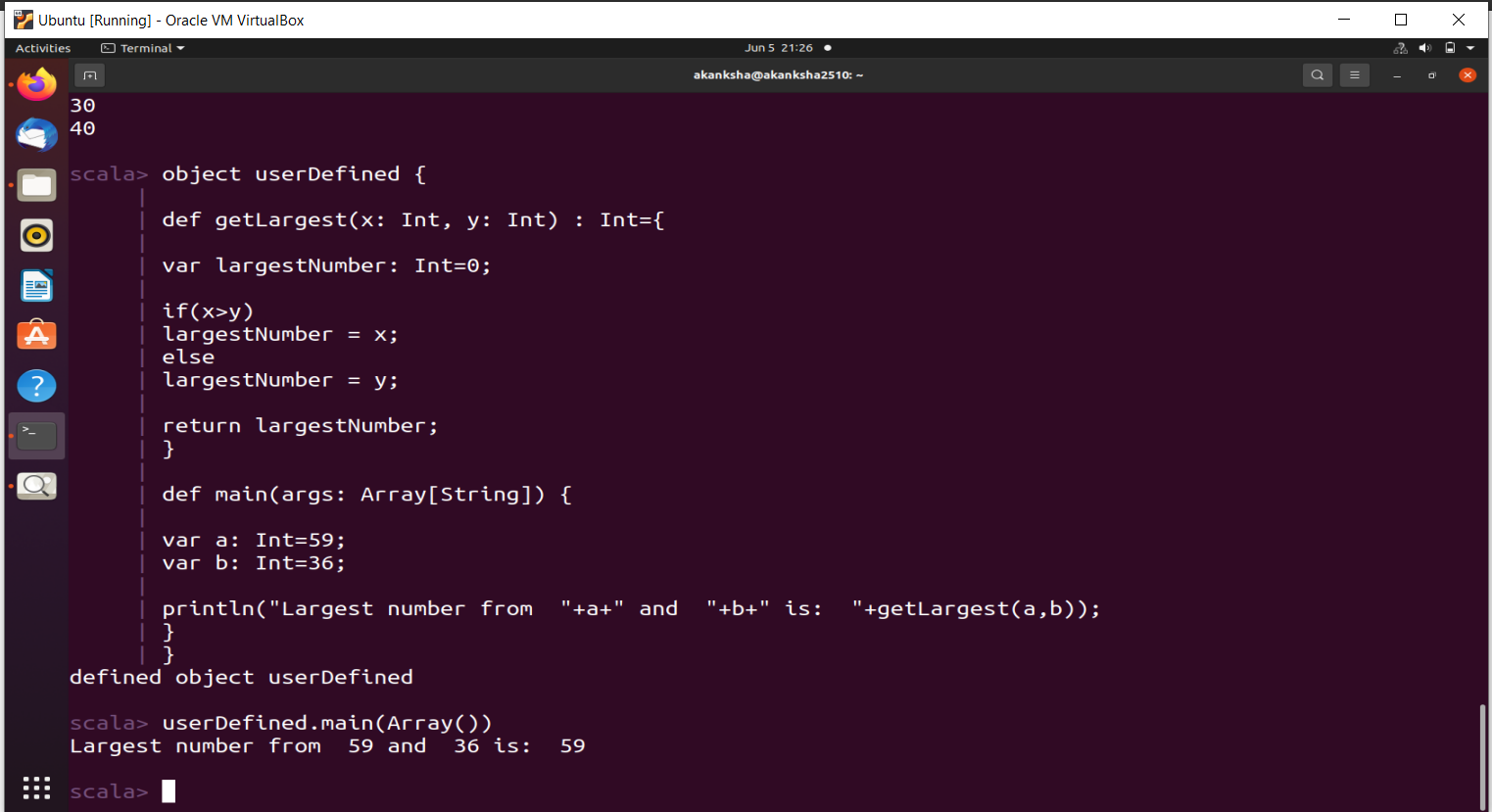


Fig 9.9

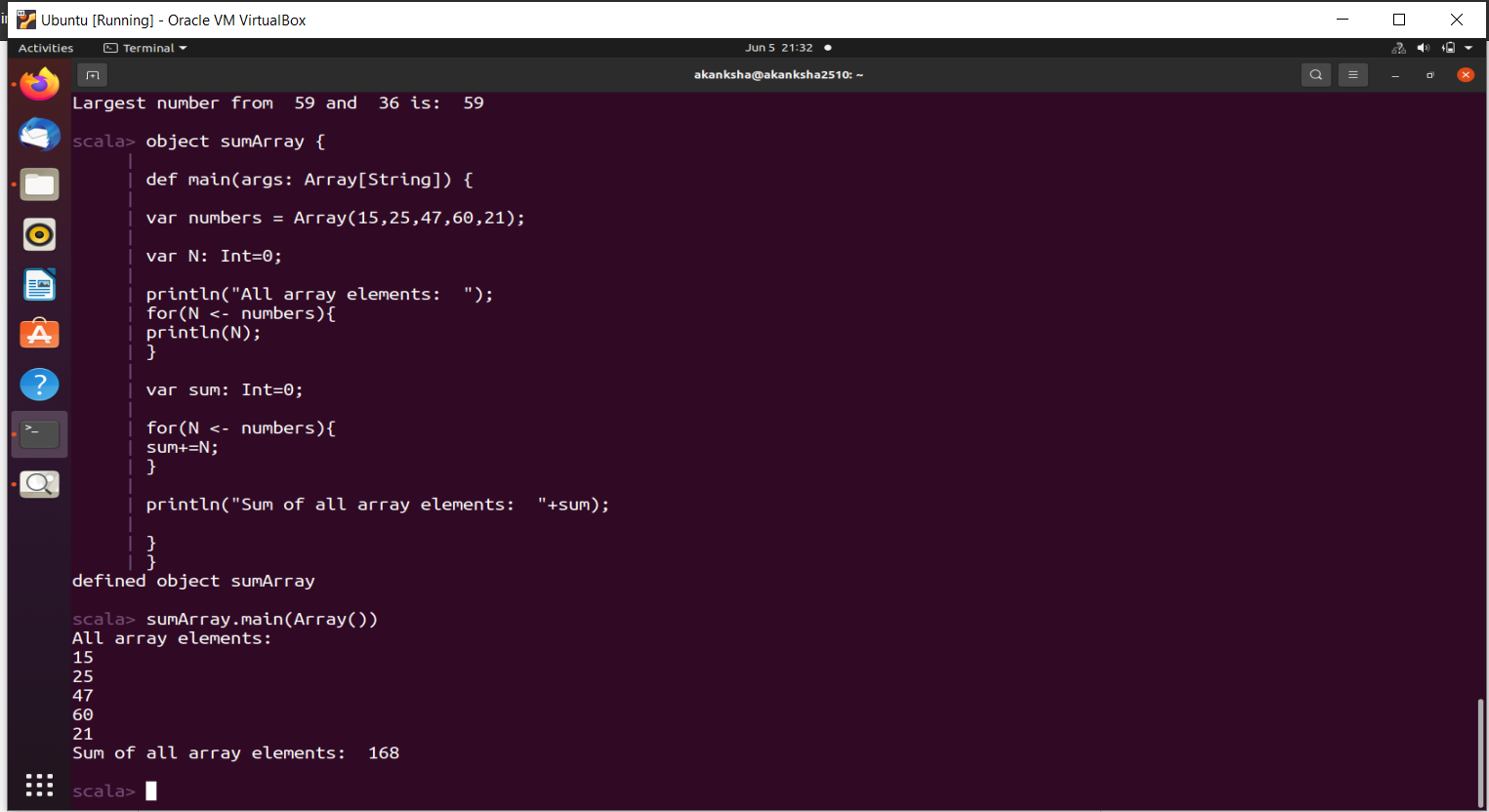


Fig 9.10

**Program 10.** Using RDD and Flat Map count how many times each word appears in a file and write out a list of words whose count is strictly greater than 4 using Spark.

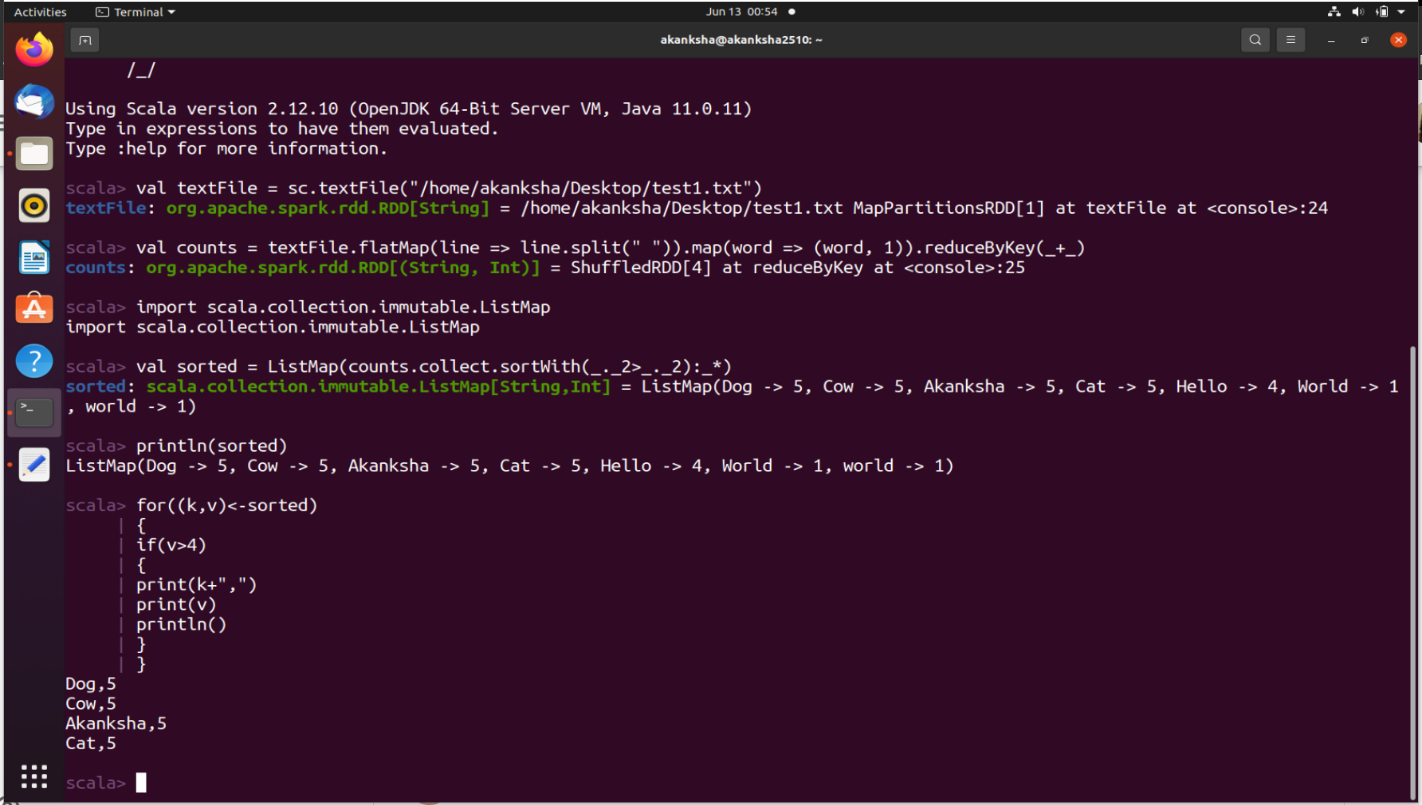


Fig 10.1