|  |  |
| --- | --- |
| **IMG_256** | **D:\Big Data Analytics - IIT\Semester 1\logo.pnglogo** |

**INFORMATICS INSTITUTE OF TECHNOLOGY**

**In Collaboration With**

**ROBERT GORDON UNIVERSITY ABERDEEN**

Big Data Programming Report

MSc. in Big Data Analytics

2019/2020

By

Dulan Manujaya

|  |  |
| --- | --- |
| IIT No : **2019528** | RGU Id : **1913768** |

CMM705

Big Data Programming

Contents

[Section 1 4](#_Toc29762625)

[Section 2.1 5](#_Toc29762626)

[2.1.1 5](#_Toc29762627)

[6](#_Toc29762628)

[6](#_Toc29762629)

[2.1.2 7](#_Toc29762630)

[7](#_Toc29762631)

[8](#_Toc29762632)

[Section 2.2 10](#_Toc29762633)

[2.2.1 10](#_Toc29762634)

[2.2.2 11](#_Toc29762635)

[11](#_Toc29762636)

[2.2.3 12](#_Toc29762637)

[12](#_Toc29762638)

[Section 2.3 13](#_Toc29762639)

[2.3.1 13](#_Toc29762640)

[2.3.2 14](#_Toc29762641)

[2.3.3 15](#_Toc29762642)

[Section 3 16](#_Toc29762643)

[16](#_Toc29762644)

[Section 4 18](#_Toc29762645)

[4.1 & 4.2 18](#_Toc29762646)

[18](#_Toc29762647)

[4.3 19](#_Toc29762648)

[4.4 20](#_Toc29762649)

[4.5 21](#_Toc29762650)

[4.6 22](#_Toc29762651)

[4.7 23](#_Toc29762652)

[Appendix 24](#_Toc29762653)

[Source Code Repository 24](#_Toc29762654)

# Section 1

A screenshot of a cell phone

Description automatically generated

Data Source: - Airbnb Website

Data Collection Layer: - ETL

Storage Layer: - HDFS, Cloud (S3), NoSQL

Processing Layer: - Batch Processing

Querying Layer: - Hive, Pig, SparkSQL, MapReduce

Analytics Layer: - Machine Learning, Predictive Modelling, Statistical Analytics

Visualization Layer: - JavaScript Webpage

# Section 2.1

## 2.1.1



# 

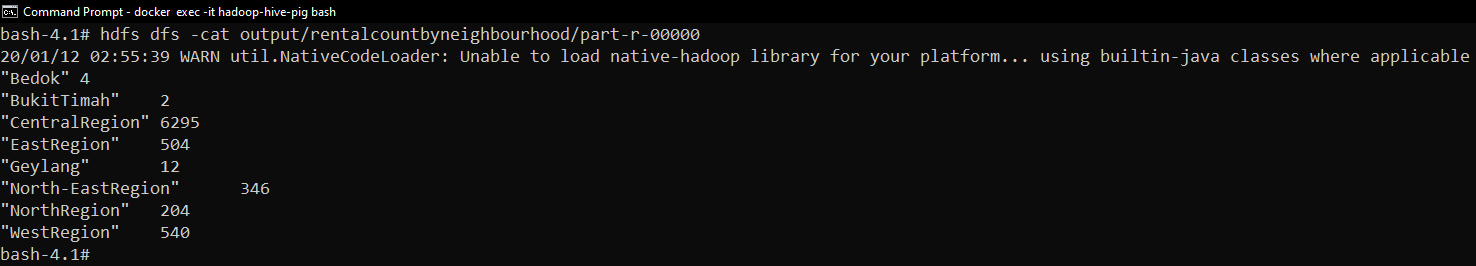
# 

## 2.1.2

## 

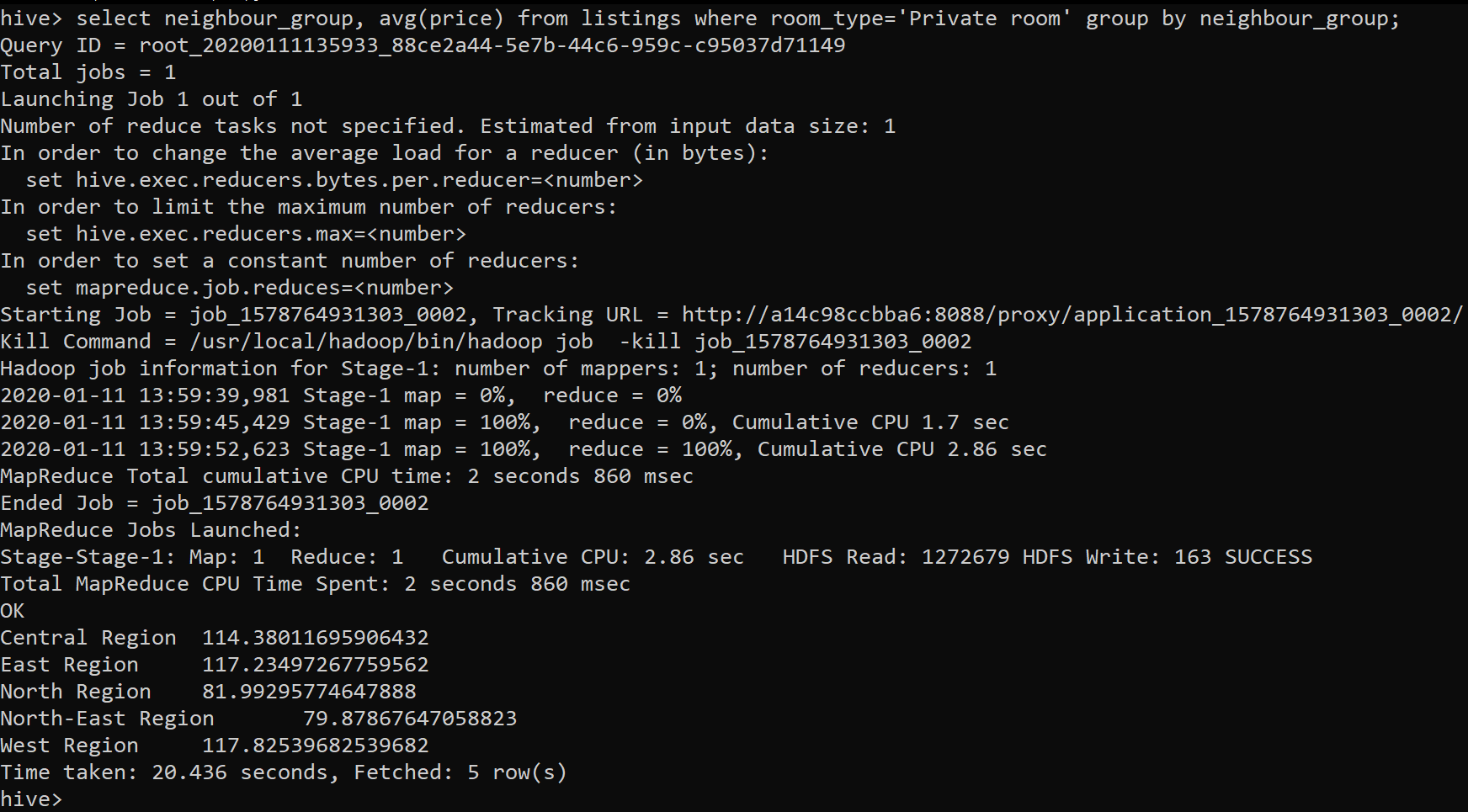
# 





# Section 2.2

## 2.2.1



## 2.2.2

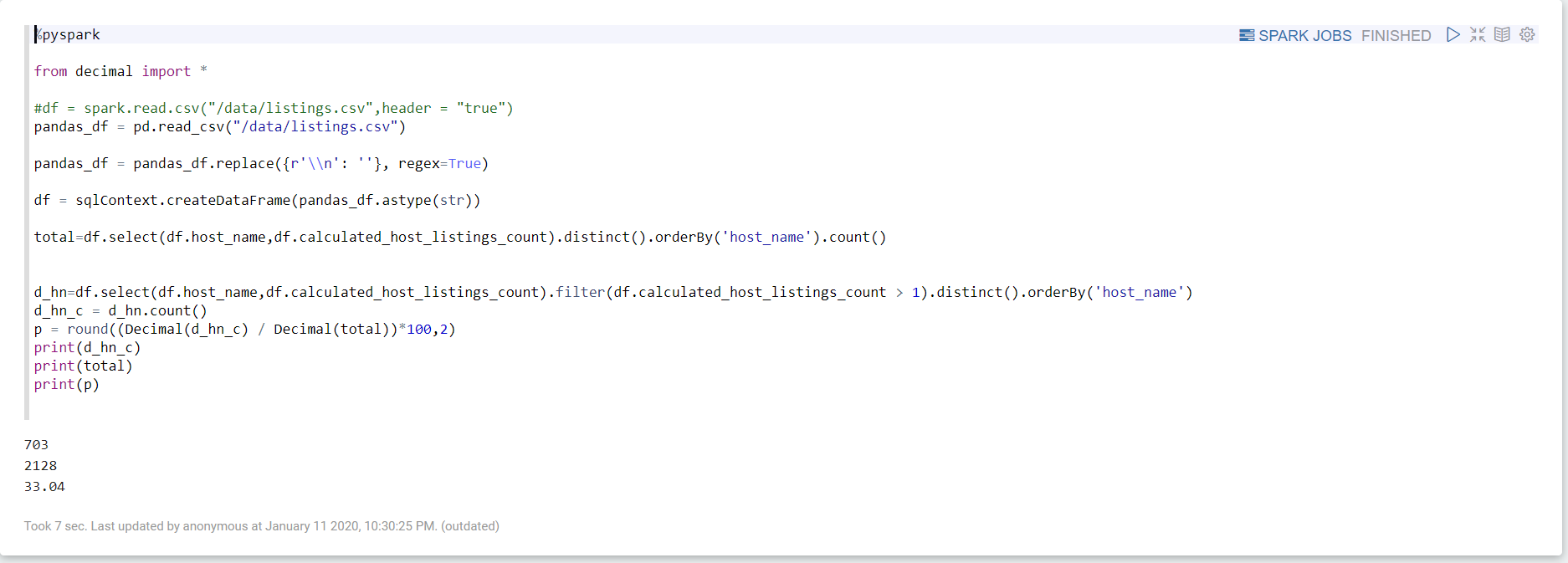
## 

## 2.2.3

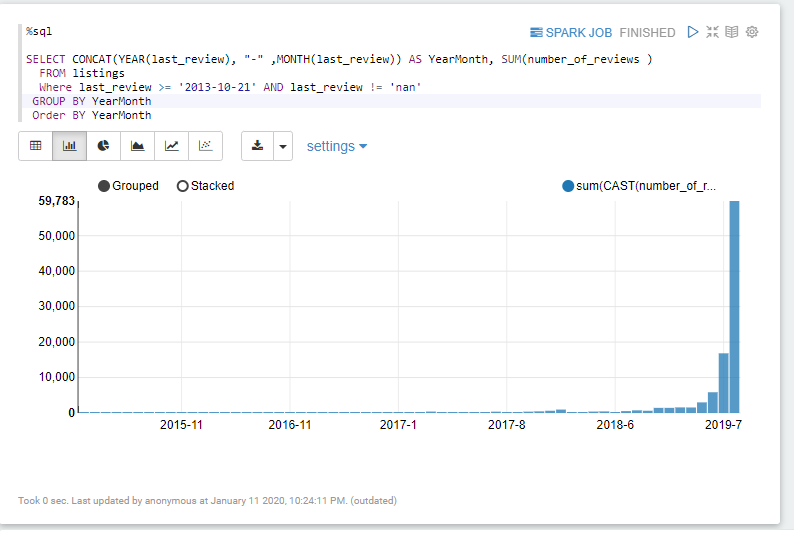
## 

# Section 2.3

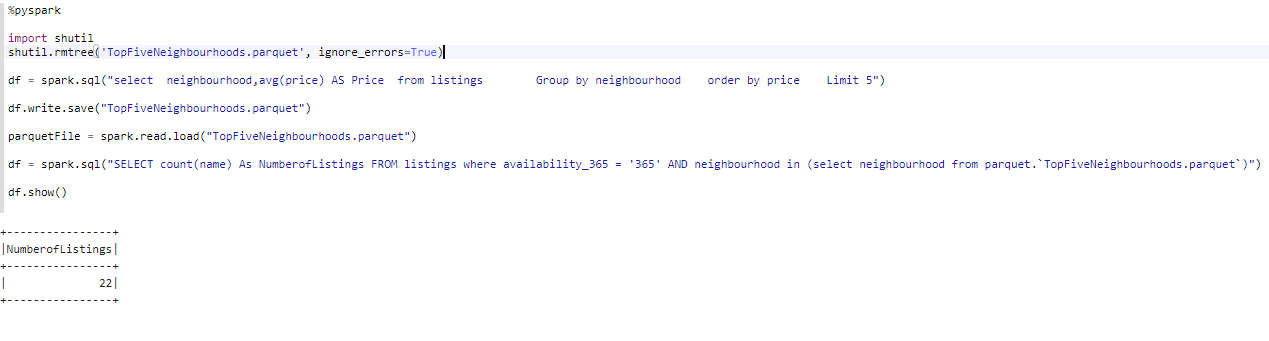
## 2.3.1



## 2.3.2



## 2.3.3



# Section 3

# 



# Section 4

# 4.1 & 4.2

# A screenshot of a cell phone Description automatically generated

## 4.3

A screenshot of a cell phone

Description automatically generated

## 4.4

A screenshot of a cell phone

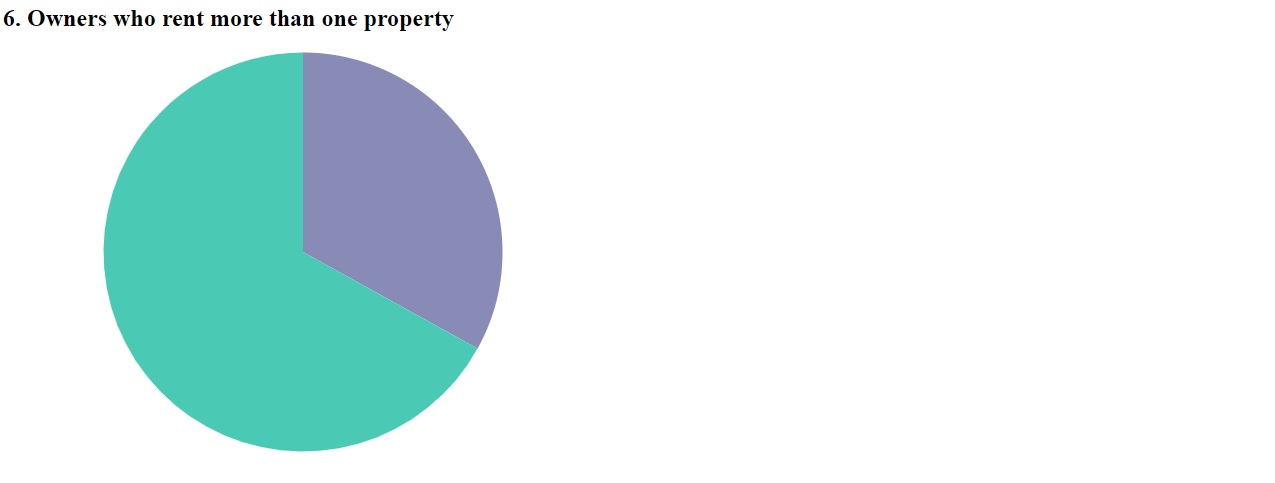
Description automatically generated

## 4.5

A screenshot of a social media post

Description automatically generated

## 4.6



## 4.7

A screenshot of a social media post

Description automatically generated

# Appendix

## Source Code Repository

<https://github.com/gotukola/hive-spark-mapreduce-javascript>