Logo

Description automatically generated

**ALY 6110: Data Management and Big Data**

**Cloudera Distribution Hadoop (CDH)**

**CRN: 70633**

**NUID: 002938209**

**Email: jain.abhin@northeastern.edu**

**Submitted by: Abhinav Jain**

**Date: 10/06/2022**

**Submitted to: Prof. Valeriy Shevchenko**

**Abstract**

In this report cloudera distribution Hadoop performed which we help in handling the big data. The most important part we are covering with implementation of cloudera to complete, test and deploy the Apache Hadoop. Here, we have the platform which is 100% free to implement the solution while using Hadoop which offer interactive research, batch processing, and execute SQL queries. Performing the implementation of the cloudera quick start virtual machine which will help in installing Cloudera handling the Big Data.

This report will illustrate about the installation of the VMware to perform the queries which will provide the solution from the big data. There is a step-by-step screenshot through which it will be easier for you to run the queries to implementation for the big data through various platforms like Hadoop, spoop, spark, hive and many more. These platforms will provide the better understanding about the data to deal with such query which are difficult to get the solutions in higher time.

**Task: 1** In this task after clicking the link login the account which will help in entering the VMware azure labs.



Task2 : In this task we open the Cloudera virtual machine by which we can enter in to azure VMware server which has limited time given to us that is 15 hours. By clicking desktop icon we can open the VMware server.

A screenshot of a computer

Description automatically generated

Task 3: In this task enter the credientials

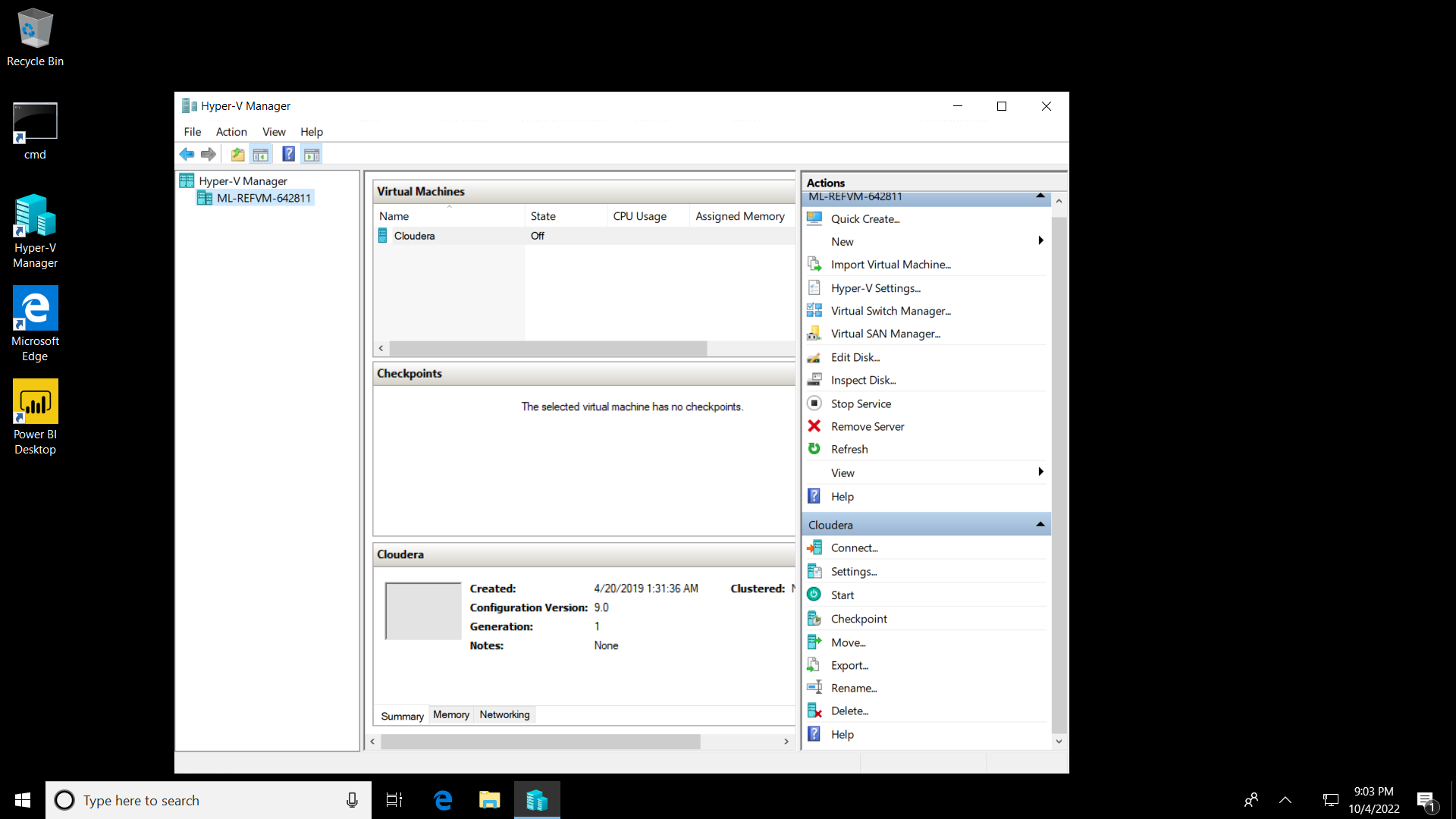
Graphical user interface, application

Description automatically generated

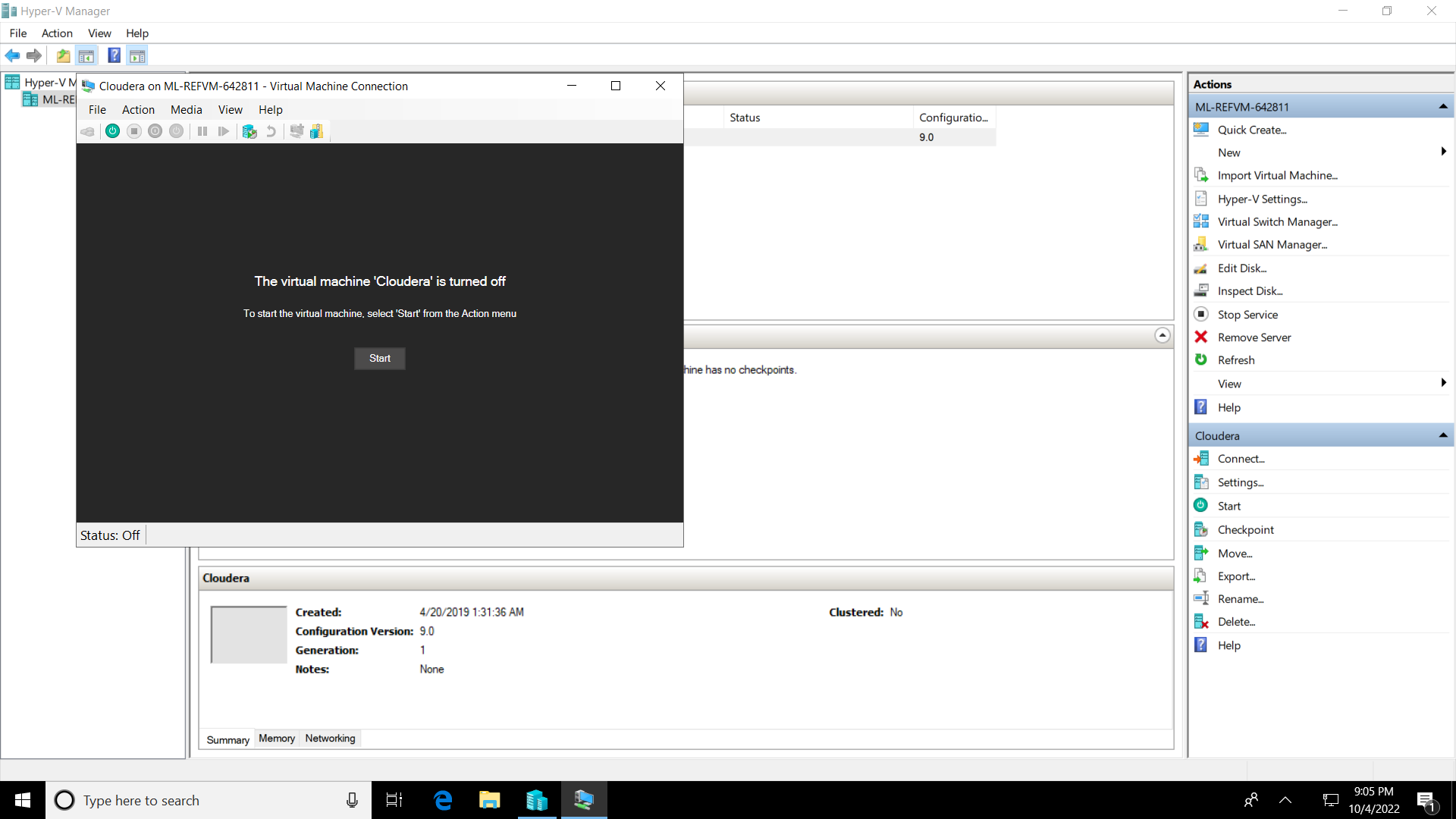
Task 4: In this screenshot you can see don’t forgot to stop your server when it is in unused stage.

Graphical user interface, application

Description automatically generated

Task5: In this task after opening the VMware server find the link HyperV Manager to enter in cloudera where we can see the Start button to enter into it.

Task 6: In this task the virtual machine Cloudera will up in the popup box where we need to start the cloudera big data platform.



Task7: In this task entering the cloudera manager credentials in Moxila firefox

Graphical user interface, application

Description automatically generated

Task 8: In this task we see the welcome to the cloudera to start with the big data handling plateform.

Graphical user interface, text, email

Description automatically generatedGraphical user interface, text, application, email

Description automatically generated

Task 9: In this task we can see the Cloudera quickstart (CDH 5.13.0) version will appear to start with many platforms where we can run queries to handle big data.

Graphical user interface, application

Description automatically generated

Task 10: In this task we activated the service monitor to start the database. It shows when the signal turn green which indicates the start signal. If its not start the we need to restart the particular db.

Graphical user interface, application

Description automatically generated



Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Part 3 :

A screenshot of a social media post

Description automatically generated

A screenshot of a computer

Description automatically generated

Below is the code by which we can find the product revenue.

A screenshot of a social media post

Description automatically generated

Here, we can see the Nike Men’s 5.0 Running Shoe shows 31567642.860603 revenue

Table

Description automatically generated

Conclusion:

While performing the task it was time taking and difficult to access the server to get the results. Big data make the system slow and results taking time to understand the tutorial. There was a problem of maximizing the screen that we need to fix it again and again. Writing query was interesting part of the assignment rest learning about how to install managing the resources was positively executed.

Graphical user interface, text

Description automatically generated

Text

Description automatically generated