## Task 1

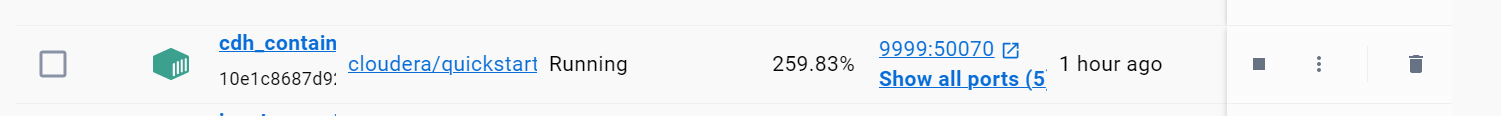
**Download and install Cloudera Quickstart using Docker Desktop (see the manual)**

I’ve already have docker desktop in WSL 2 based engine. When I tried to run cloudera/quickstart with command docker run --name=cdh\_container --hostname=quickstart.cloudera --privileged=true -t -i -v **C:\hadoop**:/src -p 8888:8888 -p 7180:7180 -p 80:80 -p 9999:50070 -p 8088:8088 cloudera/quickstart /usr/bin/docker-quickstart

container it is always finish with code 137. So, I found a solution to turn off wsl2 based engine.

A screenshot of a computer screen

Description automatically generated

After solving this issue, I got running container Download and install 

After starting Cloudera manager by running sudo /home/cloudera/cloudera-manager –express I got an issue

A screenshot of a computer error message

Description automatically generated

To solve it I need to restart ntpd

service ntpd restart

And it is solve the issue A screenshot of a computer

Description automatically generated

After container restart I need to restart cloudera service and agent using these commands

service cloudera-scm-server restart  
service cloudera-scm-agent restart

## Task 2

* **Download the data set :** [**https://www.kaggle.com/c/expedia-hotel-recommendations/data**](https://www.kaggle.com/c/expedia-hotel-recommendations/data)
* **Unzip these files**
* **Copy the files from host to the Cloudera Container using cmd commands:**

**Run: docker ps to get the running Container ID**

**Run: docker cp** **<host\_file\_directory> <Container\_ID>:<container\_file\_directory>**

[**Docker Commands Tutorial | Top 15 Docker Commands | Edureka**](https://www.edureka.co/blog/docker-commands/)

* **Put the files in HDFS system through cloudera HDFS/HUE WEB UI / using bash commands (hdfs dfs). Target directory: user/hive/warehouse. Please create separate folder for each file with appropriate name**

[**Hadoop Commands | Learn Top 23 Useful Hadoop Commands (educba.com)**](https://www.educba.com/hadoop-commands/)

* **Make screenshots of your Hue file browser.**

I’ve downloaded and unzip files in C:\Users\Anastasiya\_Viktarovi\Desktop\BigData\expedia-hotel-recommendations folder.

Copy the files from host to the Cloudera Container

To get id of container: docker ps

To copy files to the Cloudera Container: docker cp C:\Users\Anastasiya\_Viktarovi\Desktop\BigData\expedia-hotel-recommendations 73a1cbfe7dfd:expedia-hotel-recommendations

Check files A screenshot of a computer

Description automatically generated

* Put the files in HDFS system through cloudera HDFS/HUE WEB UI / using bash commands (hdfs dfs). Target directory: user/hive/warehouse. Please create separate folder for each file with appropriate name

Create folders by running this commands:

hdfs dfs -mkdir -p /user/hive/warehouse/destinations

hdfs dfs -mkdir -p /user/hive/warehouse/sample\_submission

hdfs dfs -mkdir -p /user/hive/warehouse/test

hdfs dfs -mkdir -p /user/hive/warehouse/train

copy files:

hdfs dfs -put /expedia-hotel-recommendations/destinations.csv /user/hive/warehouse/destinations/

hdfs dfs -put /expedia-hotel-recommendations/sample\_submission.csv /user/hive/warehouse/sample\_submission/

hdfs dfs -put /expedia-hotel-recommendations/test.csv /user/hive/warehouse/test/

hdfs dfs -put /expedia-hotel-recommendations/train.csv /user/hive/warehouse/train/

* Make screenshots of your Hue file browser.

A screenshot of a login

Description automatically generated

A screenshot of a web page

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a web page

Description automatically generated

A screenshot of a login page

Description automatically generated

## Task 3

**Using Python convert data from task2 into Parquet and save as a copy (For example for Java check** [**https://hadoop.apache.org/docs/r2.7.5/api/org/apache/hadoop/fs/FileSystem.html**](https://hadoop.apache.org/docs/r2.7.5/api/org/apache/hadoop/fs/FileSystem.html) **).**

**Copy scripts to the archive. Make screenshot of parquet contents.**

**While uploading large files directly in HDFS via standard tools you'll get some problems due to file size. The workaround is to use batch loading (requires additional scripts in Python/Java/Scala)**

[**https://towardsdatascience.com/why-and-how-to-use-pandas-with-large-data-9594dda2ea4c**](https://towardsdatascience.com/why-and-how-to-use-pandas-with-large-data-9594dda2ea4c)

## **EXPECTED OUTPUTS:**

**· ZIP-ed folder with your screenshots**

## Task 4

**After tasks completion, make a screenshot of YARN applications and charts as presented on the webpage at the link below:**

[**https://docs.cloudera.com/documentation/enterprise/latest/topics/cloudera\_manager.html**](https://docs.cloudera.com/documentation/enterprise/latest/topics/cloudera_manager.html)

A screenshot of a computer

Description automatically generated