# TABLE OF CONTENTS

Introduction …….....1

How to Install …….....2

Images/Screenshots/Installation ……..3

Importance of Hadoop ……4

Stored Data to Data Node ……..5

# INTRODUCTION

Apache Hadoop is a collection of open-source software utilities that facilitates using a network of many computers to solve problems involving massive amounts of data and computation. It provides a software framework for distributed storage and processing of big data using the MapReduce programming model. Hadoop was originally designed for computer clusters built from commodity hardware, which is still the common use. It has since also found use on clusters of higher-end hardware. All the modules in Hadoop are designed with a fundamental assumption that hardware failures are common occurrences and should be automatically handled by the framework.

The core of Apache Hadoop consists of a storage part, known as Hadoop Distributed File System (HDFS), and a processing part which is a MapReduce programming model. Hadoop splits files into large blocks and distributes them across nodes in a cluster. It then transfers packaged code into nodes to process the data in parallel. This approach takes advantage of data locality, where nodes manipulate the data they have access to. This allows the dataset to be processed faster and more efficiently than it would be in a more conventional supercomputer architecture that relies on a parallel file system where computation and data are distributed via high-speed networking.

# HOW TO INSTALL HADOOP

**Step 1:** go to https://hadoop.apache.org/releases.html and click 3.3.0 Binary Download

**Step 2:** Click any of the available mirror links to download Hadoop

**Step 3:** Make sure you have Java Developer Kit (JDK) installed. If it is not installed, go to https://www.oracle.com/java/technologies/javase/javase-jdk8-downloads.html.

**Step 4:** Create a new Environment Variable. You can input any name on Variable Name, input the path of your folder in Variable Value, and copy the file path into System Variable.

**Step 5:** Open CMD and enter the following command: cd Program Files\Java\jdk-13.0.2\bin. Afterwards, type javac.

**Step 6:** After installing JDK, you may proceed to install Hadoop. Simply extract the files outside the archive and transfer them to Local C.

**Step 7:** After extracting the files, find the following files in the etc folder:

-- core-site-- xml mode

-- hadoop-env-- cmd mode

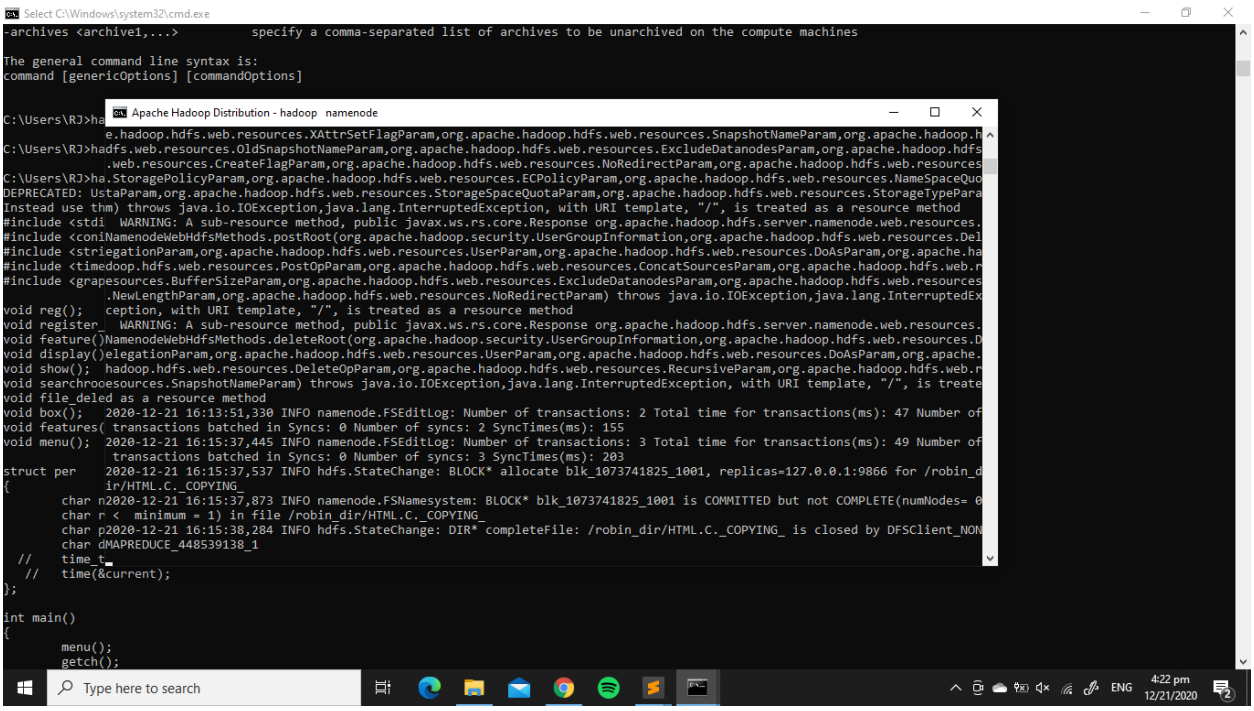
-- mapred-site -- xml mode

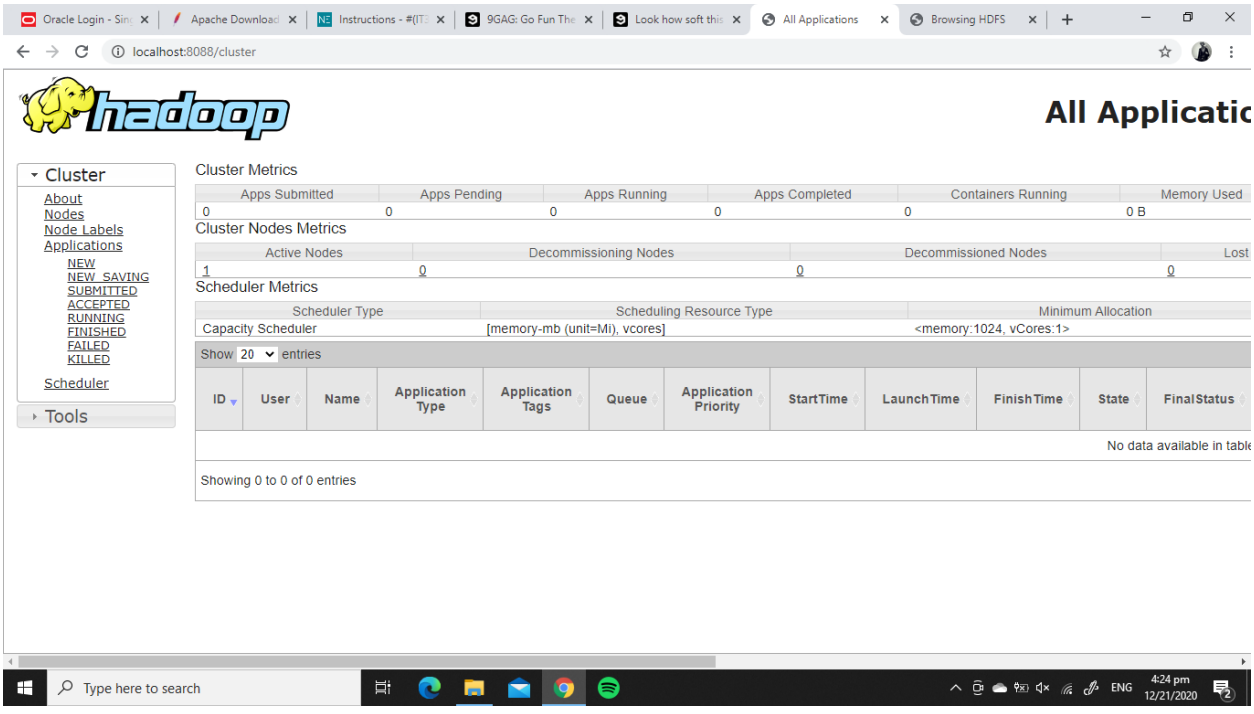
-- yarn site -- xml mode

-- hdfs site - xml mode

**Step 8:** Create a folder named “data” in Hadoop. In data, create two folders, datanode and namenode.

# IMAGES/SCREENSHOTS/INSTALLATION



# IMPORTANCE OF HADOOP

Why is Hadoop important?

- **Ability to store and process huge amounts of any kind of data, quickly.** With data volumes and varieties constantly increasing, especially from social media and the Internet of Things (IoT), that's a key consideration.
- **Computing power.** Hadoop's distributed computing model processes big data fast. The more computing nodes you use, the more processing power you have.
- **Fault tolerance.** Data and application processing are protected against hardware failure. If a node goes down, jobs are automatically redirected to other nodes to make sure the distributed computing does not fail. Multiple copies of all data are stored automatically.
- **Flexibility.** Unlike traditional relational databases, you don’t have to preprocess data before storing it. You can store as much data as you want and decide how to use it later. That includes unstructured data like text, images and videos.
- **Low cost.** The open-source framework is free and uses commodity hardware to store large quantities of data.
- **Scalability.** You can easily grow your system to handle more data simply by adding nodes. Little administration is required.

# STORING DATA IN DATA NODE

**Step 1:** Open CMD

**Step 2:** Create a file directory inside datanode by enter the command:

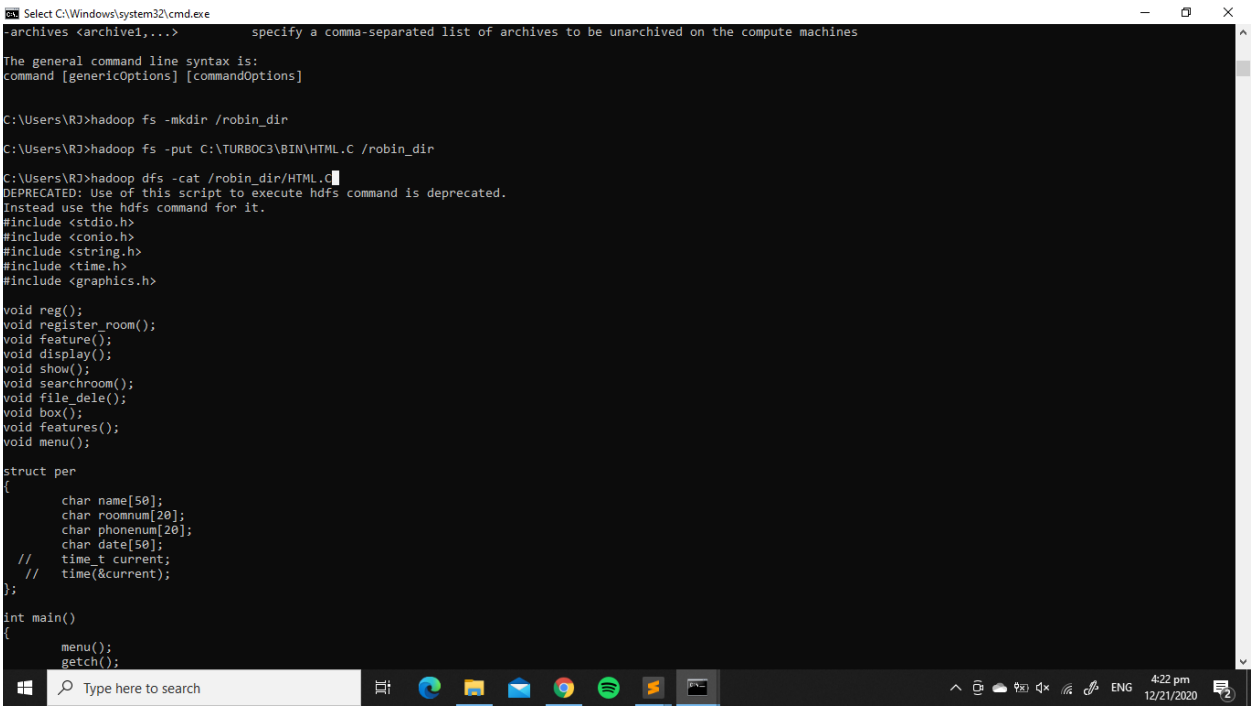
C:\Users\RJ >hadoop fs –mkdir /robin_dir

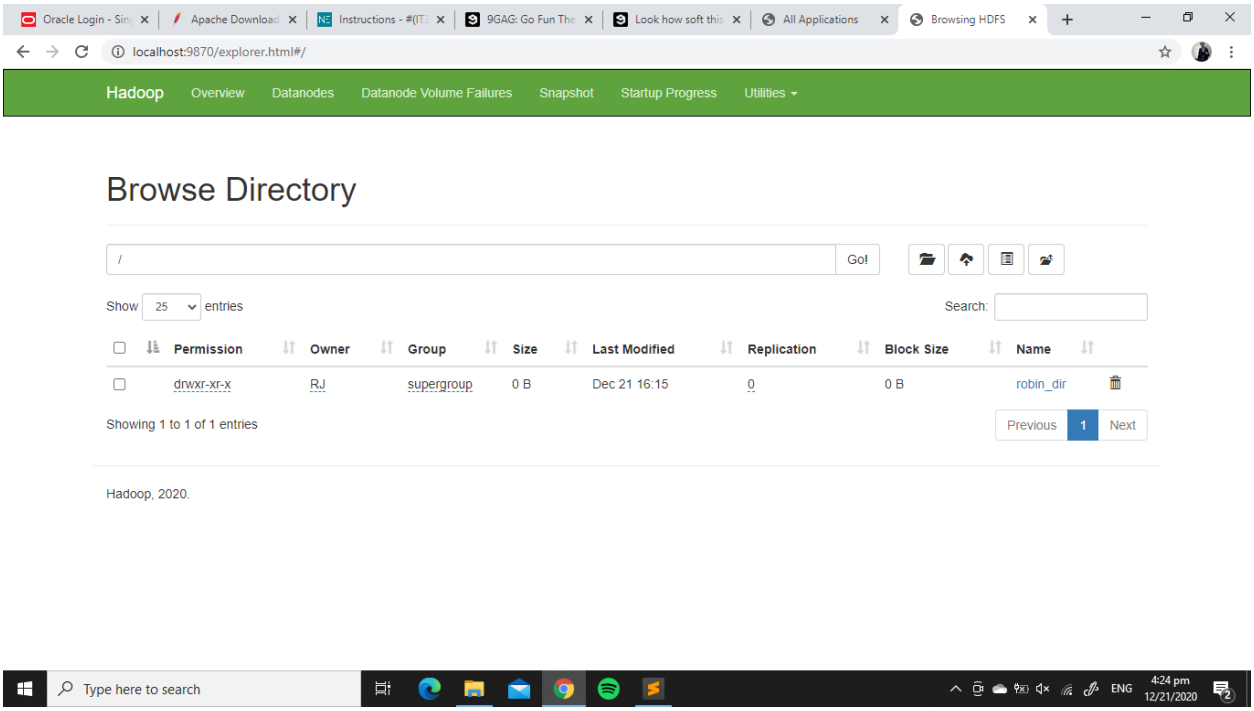
**Step 3:** Insert a file into robin_dir (or whatever you named it) using the command:

C:\Users\RJ>hadoop fs –put [Directory of the file you want to insert] /robin_dir

**Step 4:** After the file has been successfully inserted, you can have it displayed using this command:

C:\Users\RJ>hadoop fs –cat /robin_dir/[Name of the inserted file]




Oracle Login - Sin
Apache Download
Instructions - #(IT
9GAG: Go Fun The
Look how soft this
All Applications
Browsing HDFS
localhost:9870/explorer.html#/
Hadoop
Overview
Datanodes
Datanode Volume Failures
Snapshot
Startup Progress
Utilities
Browse Directory
/
Go!
Show 25 entries
Search:
Permission
Owner
Group
Size
Last Modified
Replication
Block Size
Name
drwxr-xr-x
RJ
supergroup
0 B
Dec 21 16:15
0
0 B
robin_dir
Showing 1 to 1 of 1 entries
Previous
1
Next
Hadoop, 2020.
Type here to search
ENG
4:24 pm
12/21/2020