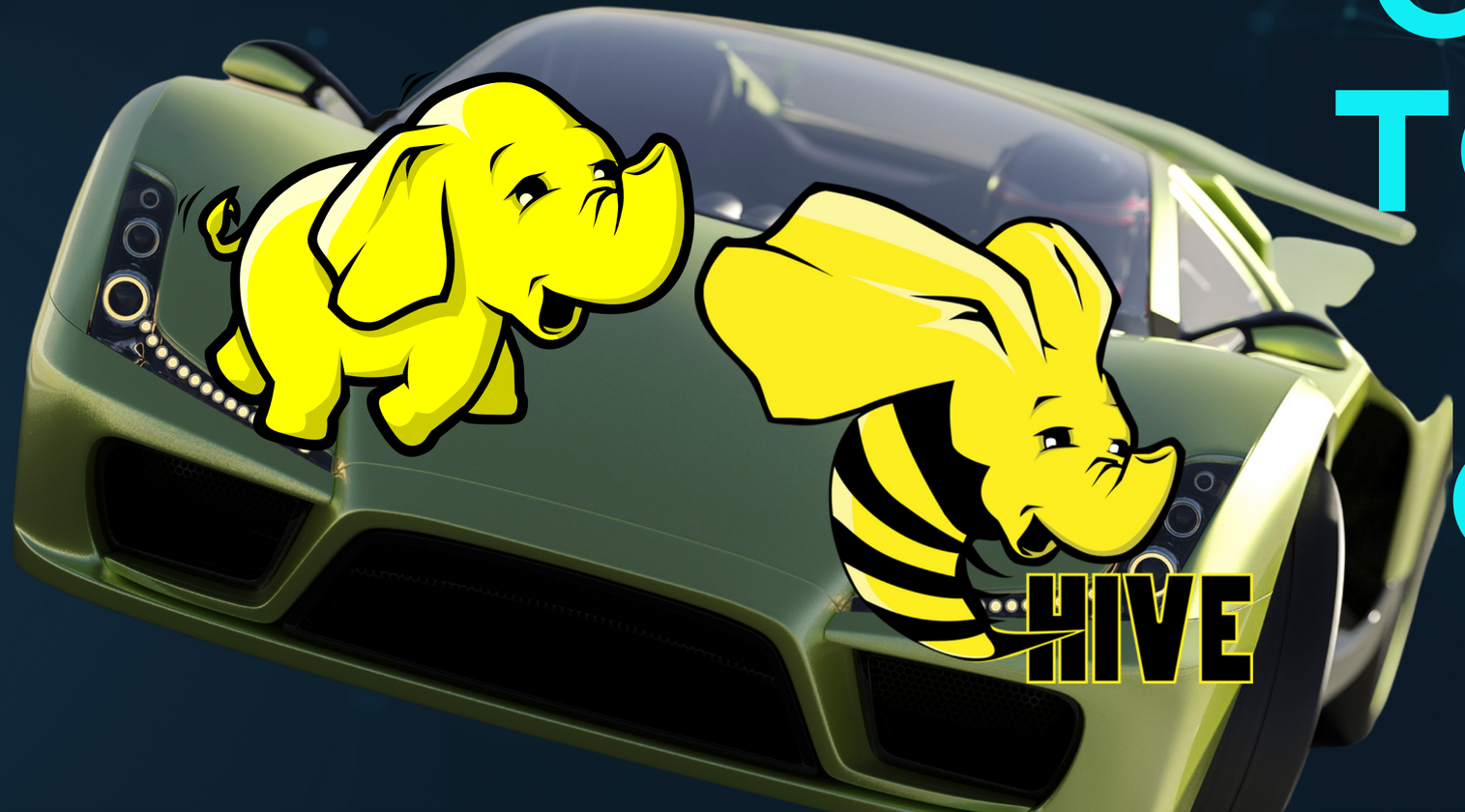
EXPLORING THE POWER OF SINGLE NODE CLUSTER



10 HADOOP COMMAND TO MASTER YOUR CLUSTER

LETS GET HOLD OF DATASET

PIP -Q INSTALL KAGGLE

GET YOUR KAGGLE KEY FROM YOUR KAGGLE.COM PROFILE

CHMOD 600 ~/.KAGGLE/KAGGLE.JSON

KAGGLE DATASETS DOWNLOAD -D INSERT_DATASET_SUFFIX_ -P LOCATION_WHERE_TO_DOWNLOAD

VISIT THE KAGGLE DATASET PAGE & COPY API COMMAND

KAGGLE DATASETS DOWNLOAD -D THEDEVASTATOR/AIRBNB-PRICES-IN-EUROPEAN-CITIES

TYPICALLY THE CLUSTER CONTAIN 3 TYPES OF NODES.

GATEWAY NODES OR CLIENT NODES OR EDGE NODES

MASTER NODES (CONTROLLER)

WORKER NODES

HADOOP COMMANDS

HDFS IS THE MAIN COMMAND TO MANAGE ALL THE COMPONENTS OF HDFS.
DFS IS THE SUB COMMAND TO MANAGE FILES IN HDFS
SUDO -U HDFS HDFS DFS ==> MAKES THE USER AS HDFS

LS = FILE SYSTEM COMMAND

MKDIR /DIR = MAKE DIRECTORY
CHOWN -R GROUP:USER /DIR = CHANGE OWNERSHIP
LS /DIR |GREP WORD = GLOBAL SEARCH

COPYFROMLOCAL OR HDFS DFS -PUT=:TO COPY FILES OR DIRECTORIES FROM LOCAL FILESYSTEM INTO HDFS

COPYTOLOCAL OR HDFS DFS -GET = TO COPY FILES OR DIRECTORIES FROM HDFS TO LOCAL FILESYSTEM.

RM -R -SKIPTRASH /DIR :DELETE DIRECTORY RECURSIVELY

TAIL : CAN BE USED TO PREVIEW LAST 1 KB OF THE FILE
CAT : CAN BE USED TO PRINT THE WHOLE CONTENTS OF THE FILE ON THE SCREEN

HELP STAT

HDFS FSCK DIR/FILE -FILES -BLOCKS -LOCATIONS (FS CONSISTENCY CHECK)

DF :TO GET THE CURRENT CAPACITY AND USAGE OF HDFS.
DU :TO GET THE SIZE OCCUPIED BY A FILE OR FOLDER.

CHMOD : CHANGE THE PERMISSIONS OF FILES.

OWNER : RWX, WHILE MEMBERS OF GROUP & OTHERS HAVE R-X.

RWX STANDS : READ, WRITE AND EXECUTE

CHMOD -R U+W DIR : OWNER GETS READ ACCESS

CHMOD -R 755 DIR : OWNER FULL ACCESS , GROUP & OTHER RW

WE CAN SET THESE PERMISSIONS FOR OWNER, GROUP AND OTHER.

BINARY VALUE	PERMISSIONS MODE	DECIMAL VALUE
000	---	0
001	--X	1
010	-W-	2
011	-WX	3
100	R--	4
101	R-X	5
110	RW-	6
111	RWX	7

DDFS.BLOCKSIZE=64M -DDFS.REPLICATION=3 -PUT /LOCAL-DIR /HDFS-DIR

THANKS FOR WATCHING

PRACTICE

PRACTICE

LIKE

SHARE

SUBSCRIBE

PRACTICE

PRACTICE

HTTPS://GITHUB.COM/INSIGHTBUILDER