Hadoop Installation

- Download Hadoop and jdk tar files in the vm

JAVA INSTALLATION
- tar xzf jdk-8u221-linux-x64.tar.gz
- sudo mkdir /opt/jdk
- sudo cp -rf '/home/cclab/Downloads/jdk1.8.0_221' /opt/jdk
- sudo update-alternatives --install /usr/bin/java java /opt/jdk/jdk1.8.0_221/bin/java 100
- sudo nano /etc/environment
- java -version
- **javac installation** – sudo apt install openjdk-8-jdk

Hadoop Installation
- Extract tar file
- **adduser** - sudo adduser hadoop
- **su** – su - hadoop
- **openssh installation** - sudo apt install openssh-server
  - ssh-keygen -t rsa -P " -f ~/.ssh/id_rsa
  - cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
  - chmod 0600 ~/.ssh/authorized_keys
  - ssh localhost
- **Move Hadoop to Hadoop user** - cp -rf '/home/cclab/Downloads/hadoop' /home/hadoop
- **Edit ~/.bashrc** – nano ~/.bashrc
  export HADOOP_HOME=/home/hadoop/hadoop
  export HADOOP_INSTALL=$HADOOP_HOME
  export HADOOP_MAPRED_HOME=$HADOOP_HOME
  export HADOOP_COMMON_HOME=$HADOOP_HOME
  export HADOOP_HDFS_HOME=$HADOOP_HOME
  export YARN_HOME=$HADOOP_HOME
  export HADOOP_COMMON_LIB_NATIVE_DIR = $HADOOP_HOME/lib/native
  export
PATH=$PATH:$HADOOP_HOME/sbin$HADOOP_HOME/bin
- **Apply changes** – source ~/.bashrc

- Edit $HADOOP_HOME/etc/hadoop/hadoop-env.sh
  - export JAVA_HOME=/opt/jdk/jdk1.8.0_221

Setup Hadoop Configuration Files
- cd $HADOOP_HOME/etc/hadoop

HADOOP PROPERTY

NAME	VALUE
fs.default.name	hdfs://localhost:9000
dfs.replication	1
dfs.name.dir	file:///home/hadoop/hadoopdata/hdfs/namenode
dfs.data.dir	file:///home/hadoop/hadoopdata/hdfs/datanode
mapreduce.framework.name	yarn
yarn.nodemanager.aux-services	mapreduce_shuffle

- Files to be edited
  - 1 – core-site.xml
  - 2,3,4 – hdfs-site.xml
  - 5 – mapred-site.xml
  - 6 – yarn-site.xml

  - **Format Namenode** - hdfs namenode -format
  - **Start Hadoop Cluster**
    - cd $HADOOP_HOME/sbin/
    - ./start-dfs.sh, ./start-yarn.sh
  - In browser,
    - Port 9870 - server
    - Port 8042 – cluster and applications
    - Port 9864 – hadoop node