**Install Operating System: (Level 0)**

1. Download DAT Linux from: <https://datlinux.com/download/>

([*datlinux-1.0.2-lxqt-desktop-amd64.iso*](https://sourceforge.net/projects/dat-linux/files/datlinux-1.0.2-lxqt-desktop-amd64.iso/download) as of 2023-12-17)

### Use balenaEtcher from <https://etcher.balena.io/> to make a installation USB

### Boot PC from USB (Lenovo Press F1) and perform the installation (Create super user u1/u1)

### Install Open SSH:

### <https://linuxize.com/post/how-to-enable-ssh-on-ubuntu-20-04/>

### Install Remote Desktop:

### <https://phoenixnap.com/kb/ubuntu-remote-desktop-from-windows>

**Install and configure Hadoop: (Level 1)**

(Refer to: <https://www.geeksforgeeks.org/how-to-install-hadoop-in-linux/?ref=gcse>)

1. Run setup\_Hadoop\_1.sh (Use **u1** as password for hadoopusr)
2. Copy .bashrc, hadoop-env.sh and all .xml files
3. Run setup\_Hadoop\_2.sh
4. Run setup\_Hadoop\_3.sh
5. Check if all run well such as Name node, 2th Name node and data node.

Note:

Choose openjdk-8-jdk as default JAVA

**Install and configure Hive/MySql: (Level 2-3)**

(Refer to: <https://www.geeksforgeeks.org/apache-hive-installation-and-configuring-mysql-metastore-for-hive/?ref=header_search> )

1. Run setup\_hive\_1.sh (Use root as password for root mysql)
2. Copy hive-env.sh and hive-sote.xml to conf
3. Run Run setup\_hive\_2.sh
4. Run Python3 and following from command line:

* from pyhive import hive
* conn = hive.connect(host='localhost',port=10000, username='hadoopusr',database='default')

**Install Jupyter Hub: (Level 0)**

(Refer to: https://jupyterhub.readthedocs.io/en/1.2.1/installation-guide-hard.html )

1. Run setup\_jupyter.sh
2. Run install\_conda.sh
3. 3. Go to localhost/jupyter
4. Install Development Tools

* sudo apt update
* sudo apt install build-essential

1. Use Hive\_Conn.ipynb to test connection

**Install Airflow: (Level 1)**

(Refer to: <https://medium.com/international-school-of-ai-data-science/setting-up-apache-airflow-in-ubuntu-324cfcee1427> )

1. Run setup\_airflow\_1.sh
2. New Terminal Run setup\_airflow\_2.sh
3. Go to localhost/8080

**GitHub Integration: (Level 2)**

(Refer to: <https://medium.com/international-school-of-ai-data-science/setting-up-apache-airflow-in-ubuntu-324cfcee1427> )

1. Run setup\_airflow\_1.sh
2. New Terminal Run setup\_airflow\_2.sh
3. Go to localhost/8080

Spark

NAS (Share Files)

CDP

JDBC/ODBC Driver