



## Big Data Practice Environment



**BUILDING  
SINGLE NODE  
HADOOP CLUSTER  
ENVIRONMENT SETUP**



# Steps To Build

## Get your Cloud

Sign up for any of the Cloud Services

01

## Understand Compute

2 VCPUs with 8GB of RAM  
22GB of RAM

02

## Get ubuntu on Compute

Setup passwordless login inside Linux

03

04

## Setting up Python

Setting up Pyspark and Jupyterlab and ensure connectivity

05

## Download installers

Hadoop, Hive, Spark3 and Postgres Docker

06

## Initiate the installation

Requires good grasp of Linux commands





# Steps & Actions

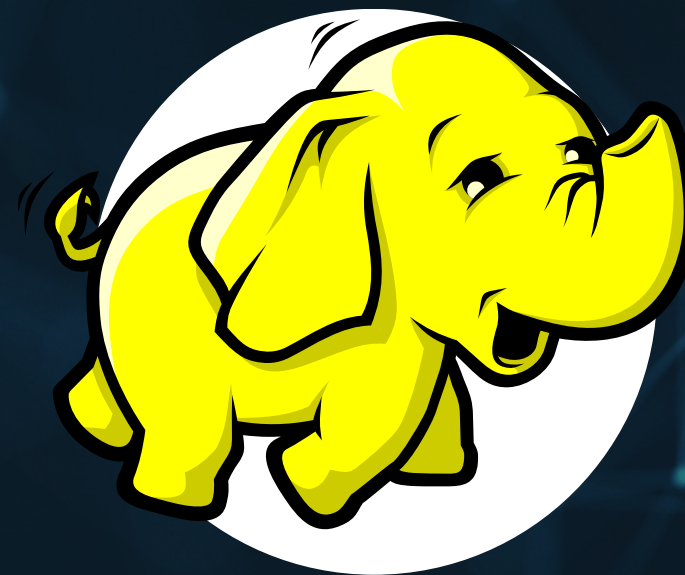
---

- **EC2 instance : T2 Large**
- **SSH connection setup for Ubuntu user**
- **Passwordless login for Ubuntu User**
- **Setting up Python & Docker**
- **Pulling Postgres Docker and configuring**



# The Ecosystem

Its all written by humans for humans...



File System

HADOOP



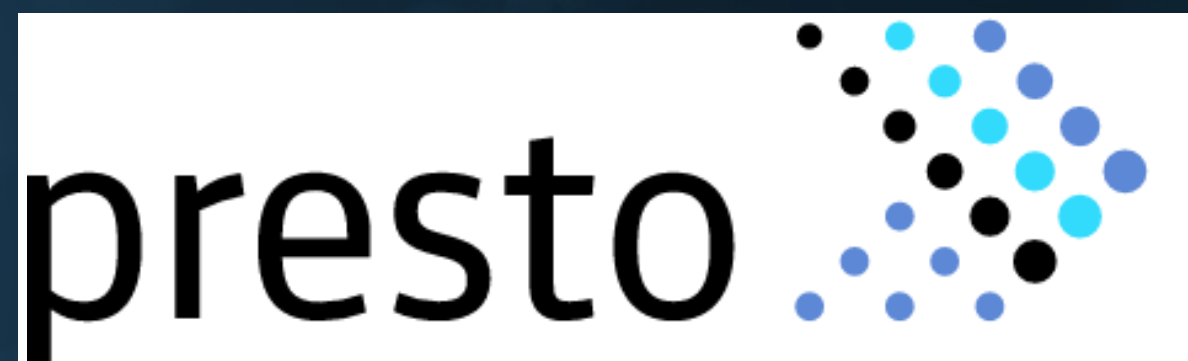
Query Engine

Spark



Programming

Python



# THANKS FOR WATCHING

**PRACTICE**

**PRACTICE**

 **LIKE**

 **SHARE**

 **SUBSCRIBE**

**PRACTICE**

**PRACTICE**