OWN YOUR BIG DATA CLUSTER IN CLOUD WHY & HOW

What is Big Data

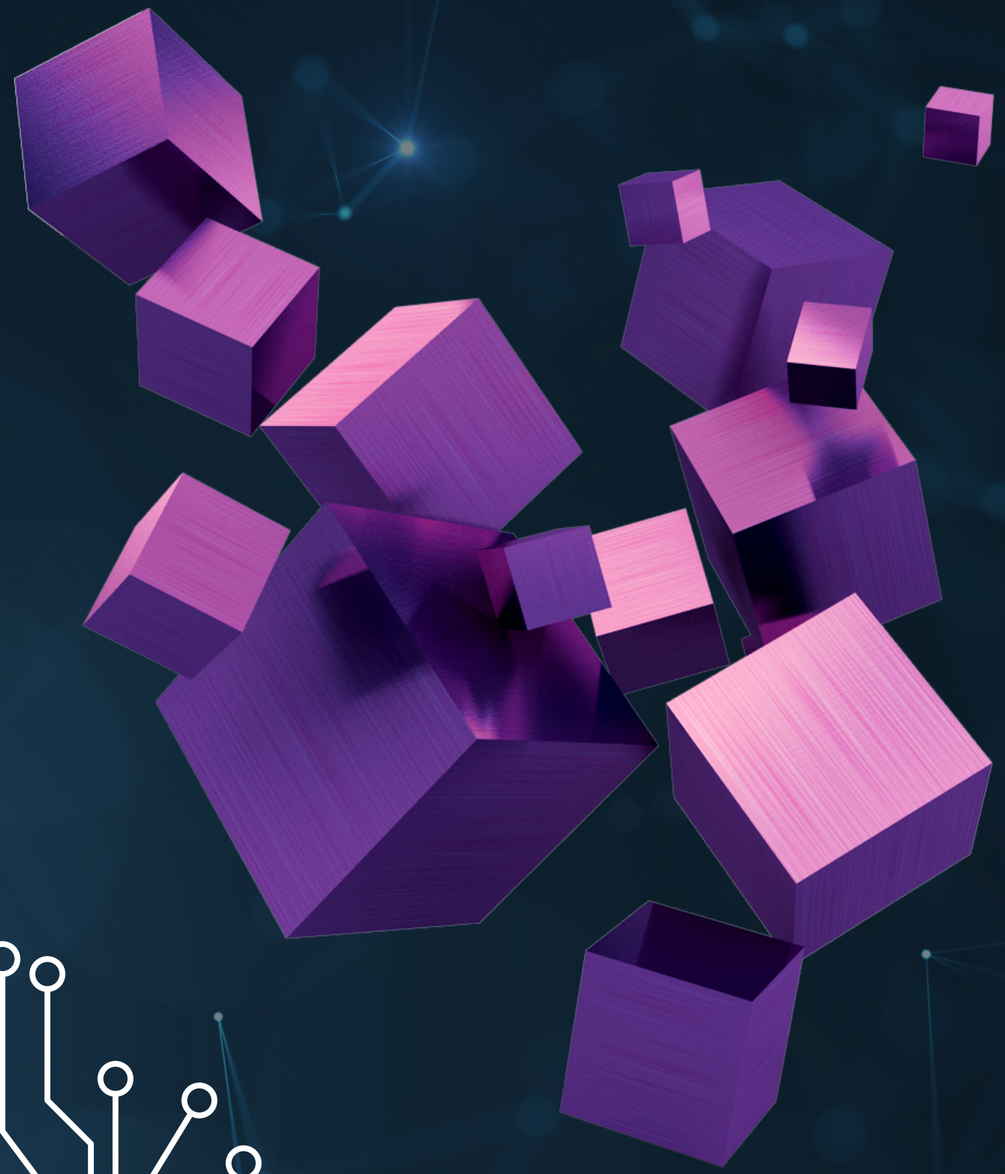
Single Server

small data

A Lay man Definition

A file or Data that cannot be contained inside a RAM of one system

File that is bigger than the RAM will crash the program, when reading the data.



Why BYOB ???

- Clarity on the Ecosystem
- Skillset to build Local Environment
- How each application connect + talk
- Faster Recovery and Troubleshooting
- Can touch and move the raw data
- See through the Technology Abstraction

HTTPS://GITHUB.COM/INSIGHTBUILDER

Steps To Build

01 Get your Cloud

Sign up for any of the Cloud Services

02 Understand Compute

2 VCPUs with 8GB of RAM
22GB of RAM

03 Get ubuntu on Compute

Setup passwordless login inside Linux

04 Setting up Python

Setting up Pyspark and Jupyterlab and ensure connectivity

05 Download installers

Hadoop, Hive, Spark3 and Postgres Docker

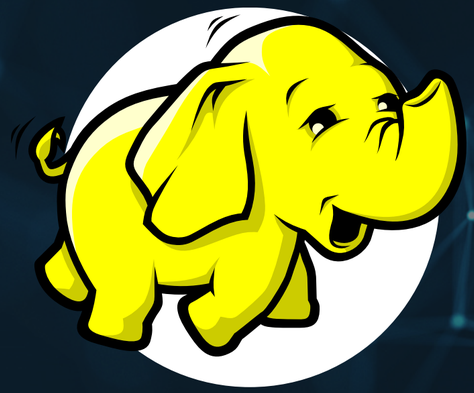
06 Initiate the installation

Requires good grasp of Linux commands

HTTPS://GITHUB.COM/INSIGHTBUILDER

The Ecosystem

Its all written by humans for humans...



File System

HADOOP

Query Engine

Spark

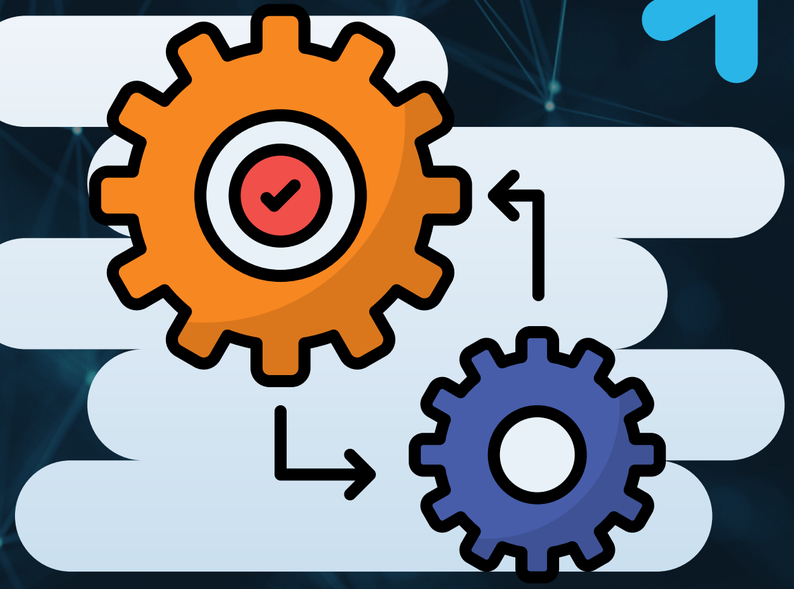
Programming

Python

HTTPS://GITHUB.COM/INSIGHTBUILDER

Multi Purpose + Multi Project

Different Warehouses : Same Architecture



ETL + LEARNING

INTEGRATION

https://github.com/insightbuilder

THANKS FOR WATCHING

PRACTICE

PRACTICE

LIKE

SHARE

SUBSCRIBE

PRACTICE

PRACTICE

HTTPS://GITHUB.COM/INSIGHTBUILDER