Big Data Practice Environment

CONFIGURING SINGLE NODE HADOOP CLUSTER

INSTALLATION SETUP

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

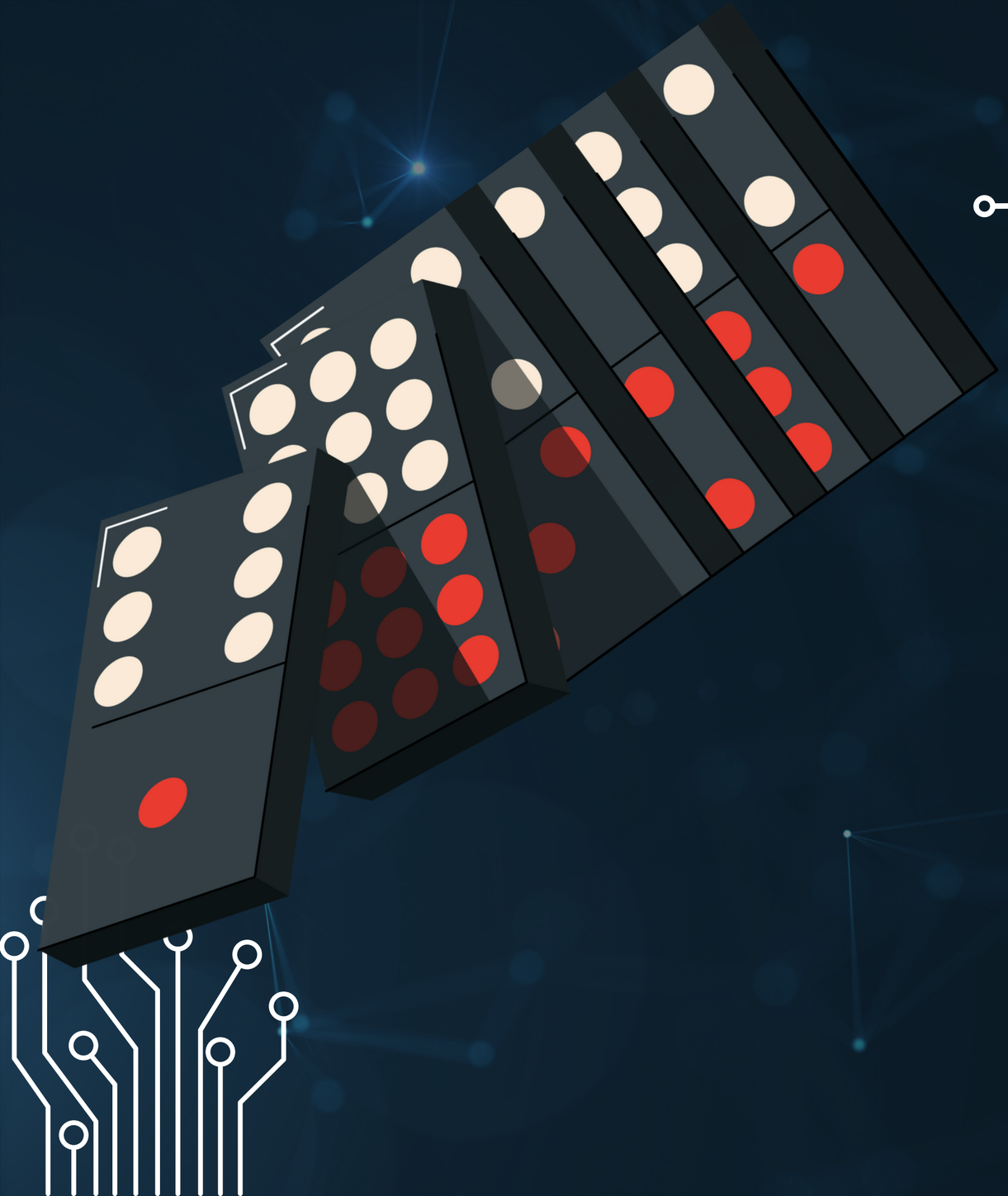
05 Download installers

Hadoop, Hive, Spark3 and Postgres Docker

06 Initiate the installation

Requires good grasp of Linux commands

HTTPS://GITHUB.COM/INSIGHTBUILDER

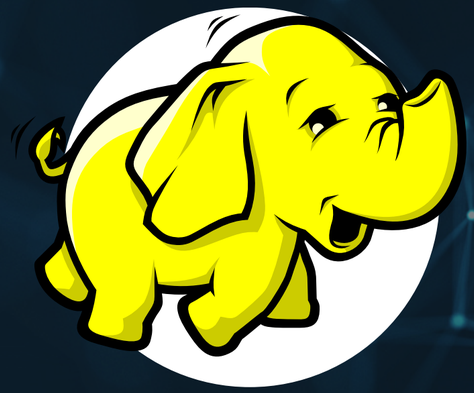


Steps & Actions

- Download Hadoop, Hive and Spark executables
- Start HDFS installation and Format
- Complete Hive installation
- Complete Spark Installation with Pyspark3
- Configure JupyterLab

The Ecosystem

Its all written by humans for humans...



File System

HADOOP

Query Engine

Spark

Programming

Python

HTTPS://GITHUB.COM/INSIGHTBUILDER

THANKS FOR WATCHING

PRACTICE

PRACTICE

LIKE

SHARE

SUBSCRIBE

PRACTICE

PRACTICE

HTTPS://GITHUB.COM/INSIGHTBUILDER