

@SID - 14TH JUNE 2022

Learning add-on technologies

ESTABLISHED TECHNOLOGIES USED BY THE MODERN
DATA SCIENTIST

Agenda for discussion

1. New channel planning.
 2. Git and GitHub and their uses in real world.
 3. Use of vm and steps to create, its importance.
 4. Choose the right os for vm and learning spree for linux.
 5. Container applications and its benefit of use.
-

Git and GitHub remote hosting repository

- Github is a remote repository for storing and sharing of code applications and coding journey.
 - This gives a good impression and a great way to monitor the progress and shows the importance to prospective employer.
 - Every developer must have an account or 2 and manage the code files.
 - Learning the process of creating and managing repositories
-

Linux, vms and the use with preferred os for data scientist.

- Linux is the os used by the entire production systems and data centres around the world. Its a great os, every developer or coder and or tech person must know.
 - Pms are a great way to manage the various development machines for different projects and environments. It maintains individuality.
 - Choosing the beginner friendly os is a great way to learn faster and use of cheatsheet is friendly all it needs is the use.
-

Container orchestration and deployment

- Containers are a great way to develop and deploy the applications anywhere and anytime.
 - They have significant benefits in the development to deployment environment.
 - Its a unique way to build uniformity.
 - Docker and kubernetes are a great way to start with as these are the most powerful and widely used in the industry.
-

Big data tools and technologies

- Big data refers to bigger projects and heavy application development.
 - Some of the technology involves - apache kafka, apache spark, apache hadoop, Cassandra,
 - Mongo db, sql db,
 - Tools like weka, auto ml using weka, power BI, open source technologies.
 - Knime analytics, and other tools.
-