Learning plan1_gakyeong_bae

Learning plan1

Detail

- Study 1 hour on M, W, F

week 4- 5 : Introduction to Docker

Goal :

Understand Docker fundamentals and how to containerize applications.

Topic :

What is Docker & Why Use It?
Installing Docker (Windows/Linux/Mac)
Key Concepts: Containers, Images, Volumes, Networks Running Containers (docker run, docker ps, docker stop) Docker CLI Commands Writing a Dockerfile
Building & Running Custom Images
Managing Multiple Containers

Resource :

- main resource: https://www.youtube.com/watch?v=i7ABlHngi1Q
- Additional resource:
- https://docker-curriculum.com/
- https://www.linkedin.com/learning/azure-spark-databricks-essential-training/optimize-data-pipelines?u=2153100 (Databricks)
- https://docs.databricks.com/en/index.html (Databricks)

week 6-7 : Introduction to pySpark

Goal :

Learn to run PySpark inside Docker for distributed data processing.

Topic :

Running PySpark in a Docker Container
Using Docker Compose for PySpark + Jupyter
Mounting Local Files to a PySpark Container Networking Between Containers

Resource :

- main resource: https://learning.oreilly.com/course/introduction-to-pyspark/9781771375535/
- Additional resource:
- W3schools pyspark tutorial : https://www.w3schools.com/python/pyspark_intro.php
- pyspark-examples: https://github.com/spark-examples/pyspark-examples