

Big Data

Pull Bitnami

Dosen :

M. Hasyim Ratsanjani, S.Kom., M.Kom.



Dibuat Oleh :

Nurhaliza Anindya Putri





2241720016

Kelas : TI – 3D

POLITEKNIK NEGERI MALANG

2025

## 1. Implementasi Python Spark (di Dockers)

No	Gambar	Deskripsi
1		Pull bitnami/spark image
2		Menjalankan Master
3		Menjalankan Worker
4		Notebook
5	<pre> from pyspark.sql import SparkSession import gc  # Stop session sebelumnya jika ada try:     if 'spark' in locals() and isinstance(spark, SparkSession):         print("Menghentikan SparkSession yang sudah ada...")         spark.stop() except:     pass gc.collect()  # Inisialisasi SparkSession spark = SparkSession.builder \     .appName("Spark Cluster Test") \  .master("spark://spark-master:7077 ") \     .getOrCreate()  print("✅ SparkSession berhasil dibuat") print(f"📌 App Name: {spark.sparkContext.appName}") print(f"📌 Master: {spark.sparkContext.master}") </pre>	<p>Upload data ke notebook</p> <p>Copy ke direktory /opt/spark_data/</p>

	<pre># Load CSV df = spark.read.csv("ecommerce_transactions_1000.csv", header=True, inferSchema=True) df.show(5)  # Hitung jumlah baris print("Jumlah baris dalam DataFrame:") print(df.count())</pre>	
--	--	--