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	PS C:\Users\USER> docker pull bitnami/spark:3.5.0 3.5.0: Pulling from bitnami/spark 4c098073a180: Pull complete Digest: Aba\$5:c68f458ea2e32c2128cd16906e50070dbd7182f5a6b26c014196825cb0f4ece9 Status: Downloaded newer inage for bitnami/spark:3.5.0 docker.to/bitnami/spark:3.5.0 PS C:\Users\USER>[]	Pull bitnami/spark image
2	oS C. Ulbers IUSED - <mark>Oxder: ran -d</mark>	Menjalankan Master
3	OS CUIDAND/BIDDS declar sin — anno spork-senter)minoris spork-sent - \$9900_MRSTBU_UBL-spork-//no ork-senter 7027 -> *-forest data /-more flower datamore 20more 20mo	Menjalankan Worker
4	FS C-Ubernitotio decker on .t. a place data .a devo data .a service spack and .a cor root .a SHAM_MSIGN-spack.//pack and ter: PVVV = CMMS_SMOS_MSS -/-look data /petrops_data spacer_tala_pack analogod data. The root consoler for deame. For service services post Top 6.8 -6.8490 = 127.6 6.1-6. Listen top 0.6.6 -6.8490 = 127.6 6	Notebook
5	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}")	Upload data ke notebook Copy ke direktory /opt/spark_data/

Load CSV
df =
spark.read.csv("ecommerce_transac
tions_1000.csv", header=True,
inferSchema=True)
df.show(5)

Hitung jumlah baris
print("Jumlah baris dalam
DataFrame:")
print(df.count())