LAPORAN
TUGAS PRAKTIKUM BIG DATA
(Apache Spark)

Disusun oleh:

Fitria Nur Sholikah 2241760004

PROGRAM STUDI D-IV SISTEM INFORMASI BISNIS

JURUSAN TEKNOLOGI INFORMASI POLITEKNIK

NEGERI MALANG

2025

Tugas Praktikum

1.Instalasi Apache Spark

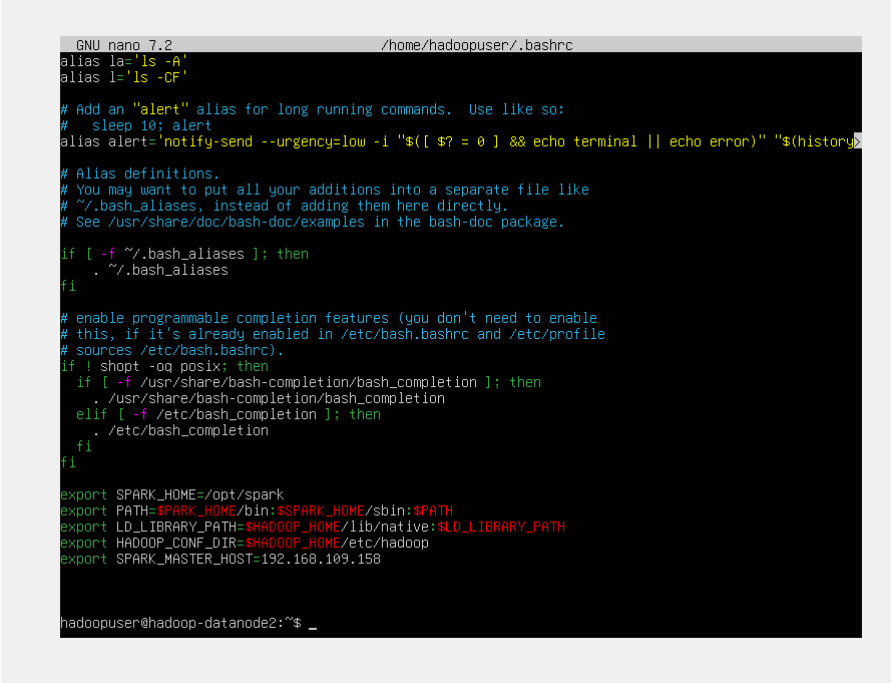
- Silakan gunakan Cluster Hadoop dari hasil kuis sebelumnya di VBox kelompok Anda.
- Lakukan instalasi Apache Spark.
- Unduh versi terbaru Spark dari situs resmi atau gunakan *wget* dari dalam namenode vbox Anda:
- *wget https://downloads.apache.org/spark/spark-3.5.5/spark-3.5.5-bin hadoop3.tgz*

1. Instalasi Apache Spark (2)

- Ekstrak dan pindahkan direktori tar -xvzf spark-3.4.1-bin-hadoop3.tgz sudo mv spark-3.4.1-bin-hadoop3 /opt/spark

2. Konfigurasi Apache Spark

- Konfigurasi environment variables. Edit .bashrc atau .profile :
  nano ~/.bashrc
- Tambahkan baris berikut:
  export SPARK_HOME=/opt/spark export
  PATH=$SPARK_HOME/bin:$SPARK_HOME/sbin:$PATH export
  LD_LIBRARY_PATH=$HADOOP_HOME/lib/native:$LD_LIBRARY_PATH export
  HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoop export
  SPARK_MASTER_HOST=<IP_MASTER_NODE>
  Kemudian jalankan:
  source ~/.bashrc
  Hasil:

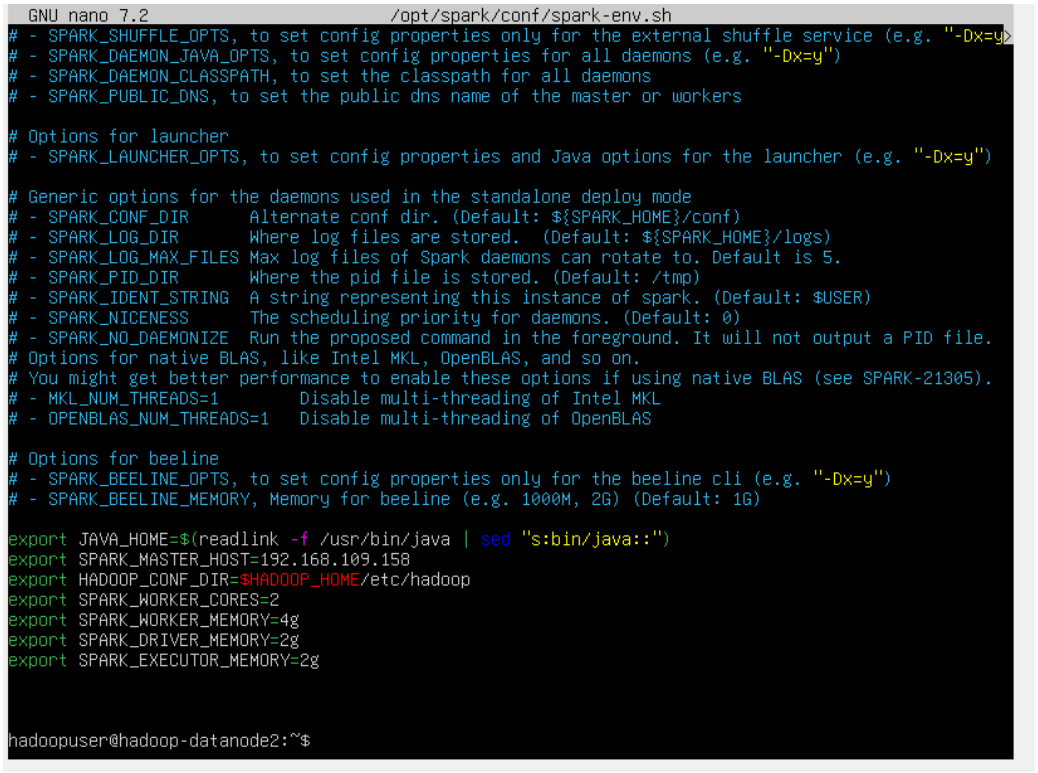


2. Konfigurasi Apache Spark (2)
    - Konfigurasi spark-env.sh
    - Salin templat dan edit:

    cp /opt/spark/conf/spark-env.sh.template /opt/spark/conf/spark-env.sh
    nano /opt/spark/conf/spark-env.sh
    - Tambahkan:

    export JAVA_HOME=$(readlink -f /usr/bin/java | sed "s:bin/java::") export SPARK_MASTER_HOST=<IP_MASTER_NODE> export HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoop export SPARK_WORKER_CORES=2 export SPARK_WORKER_MEMORY=4g export SPARK_DRIVER_MEMORY=2g export SPARK_EXECUTOR_MEMORY=2g

    Hasil:



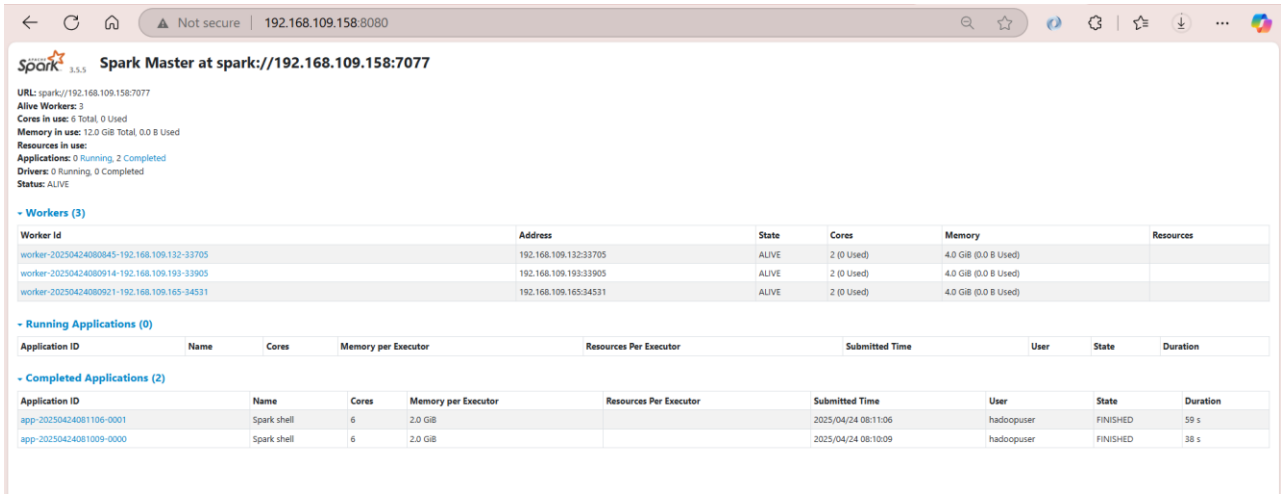
3. Menjalankan Apache Spark di Cluster Hadoop
    - Jalankan Spark Master di namenode (Master Node), jalankan: start-master.sh
    - Buka di browser: http://<IP_MASTER>:8080
    - Di setiap Worker Node (data node-pastikan spark sudah setup), jalankan:

    start-worker.sh spark://<IP_MASTER>:7077
4. Uji Apache Spark
    - Cek apakah Spark bekerja dengan baik:

    spark-shell --master spark://<IP_MASTER>:7077
    - Atau jalankan contoh aplikasi:

    /opt/spark/bin/run-example SparkPi 10

    Hasil:


Not secure | 192.168.109.158:8080
Spark 3.5.5
Spark Master at spark://192.168.109.158:7077
URL: spark://192.168.109.158:7077
Alive Workers: 3
Cores in use: 6 Total, 0 Used
Memory in use: 12.0 GiB Total, 0.0 B Used
Resources in use:
Applications: 0 Running, 2 Completed
Drivers: 0 Running, 0 Completed
Status: ALIVE
Workers (3)
Worker Id | Address | State | Cores | Memory | Resources
worker-20250424080845-192.168.109.132-33705 | 192.168.109.132:33705 | ALIVE | 2 (0 Used) | 4.0 GiB (0.0 B Used)
worker-20250424080914-192.168.109.193-33905 | 192.168.109.193:33905 | ALIVE | 2 (0 Used) | 4.0 GiB (0.0 B Used)
worker-20250424080921-192.168.109.165-34531 | 192.168.109.165:34531 | ALIVE | 2 (0 Used) | 4.0 GiB (0.0 B Used)
Running Applications (0)
Application ID | Name | Cores | Memory per Executor | Resources Per Executor | Submitted Time | User | State | Duration
Completed Applications (2)
Application ID | Name | Cores | Memory per Executor | Resources Per Executor | Submitted Time | User | State | Duration
app-20250424081106-0001 | Spark shell | 6 | 2.0 GiB | | 2025/04/24 08:11:06 | hadoopuser | FINISHED | 59 s
app-20250424081009-0000 | Spark shell | 6 | 2.0 GiB | | 2025/04/24 08:10:09 | hadoopuser | FINISHED | 38 s