

JOBSHEET 6

APACHE SPARK

MATA KULIAH BIG DATA



Disusun Oleh:

Angelia Anna P.R	2241760016
Dimas Rosyidin	2241760054
Muhammad Khasbul Hadi Fauzan	2241760039
Muhammad Ivan Yoda Bellamy	2241760094
Rega Dane Wijayanta	2241760113

PROGRAM STUDI D4 SISTEM INFORMASI BISNIS

JURUSAN TEKNOLOGI INFORMASI

POLITEKNIK NEGERI MALANG

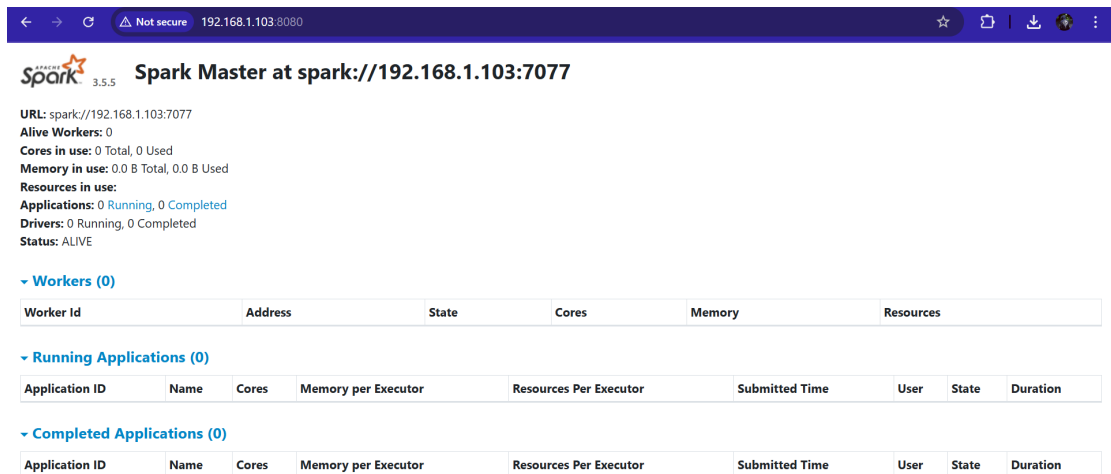
2025

```
GNU nano 7.2 /root/spark/conf/spark-env.sh
# - SPARK_LAUNCHER_OPTS, to set config properties and Java options for the launcher (e.g. "-Dx=y")

# Generic options for the daemons used in the standalone deploy mode
# - SPARK_CONF_DIR      Alternate conf dir. (Default: ${SPARK_HOME}/conf)
# - SPARK_LOG_DIR       Where log files are stored. (Default: ${SPARK_HOME}/logs)
# - SPARK_LOG_MAX_FILES  Max log files of Spark daemons can rotate to. Default is 5.
# - SPARK_PID_DIR       Where the pid file is stored. (Default: /tmp)
# - SPARK_IDENT_STRING   A string representing this instance of spark. (Default: $USER)
# - SPARK_NICENESS       The scheduling priority for daemons. (Default: 0)
# - SPARK_NO_DAEMONIZE   Run the proposed command in the foreground. It will not output a PID file.
# Options for native BLAS, like Intel MKL, OpenBLAS, and so on.
# You might get better performance to enable these options if using native BLAS (see SPARK-21305).
# - MKL_NUM_THREADS=1    Disable multi-threading of Intel MKL
# - OPENBLAS_NUM_THREADS=1  Disable multi-threading of OpenBLAS

# Options for beeline
# - SPARK_BEELINE_OPTS, to set config properties only for the beeline cli (e.g. "-Dx=y")
# - SPARK_BEELINE_MEMORY, Memory for beeline (e.g. 1000M, 2G) (Default: 1G)
export JAVA_HOME=$(readlink -f /usr/bin/java | sed "s:bin/java::")
export SPARK_MASTER_HOST=192.168.1.103
export HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoop
export SPARK_WORKER_CORES=2
export SPARK_WORKER_MEMORY=4g
export SPARK_DRIVER_MEMORY=2g
export SPARK_EXECUTOR_MEMORY=2g
```

3. Menjalankan Apache Spark di Cluster Hadoop



Spark Master at spark://192.168.1.103:7077

URL: spark://192.168.1.103:7077
Alive Workers: 0
Cores in use: 0 Total, 0 Used
Memory in use: 0.0 B Total, 0.0 B Used
Resources in use:
Applications: 0 Running, 0 Completed
Drivers: 0 Running, 0 Completed
Status: ALIVE

▼ Workers (0)

Worker Id	Address	State	Cores	Memory	Resources
-----------	---------	-------	-------	--------	-----------

▼ Running Applications (0)

Application ID	Name	Cores	Memory per Executor	Resources Per Executor	Submitted Time	User	State	Duration
----------------	------	-------	---------------------	------------------------	----------------	------	-------	----------

▼ Completed Applications (0)

Application ID	Name	Cores	Memory per Executor	Resources Per Executor	Submitted Time	User	State	Duration
----------------	------	-------	---------------------	------------------------	----------------	------	-------	----------

belum ada worker dikarenakan kelompok kami belum sempat berkumpul di jaringan yang sama.

4. Uji Apache Spark

```
root@hadoop-namenode: /home/hadoopuser/matrix-mult-scala# ~/spark/bin/run-example SparkPi 10
25/04/25 14:28:35 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
25/04/25 14:28:35 INFO SparkContext: Running Spark version 3.5.5
25/04/25 14:28:35 INFO SparkContext: OS info Linux, 6.8.0-53-generic, amd64
25/04/25 14:28:35 INFO SparkContext: Java version 1.8.0_442
25/04/25 14:28:35 INFO ResourceUtils: =====
25/04/25 14:28:35 INFO ResourceUtils: No custom resources configured for spark.driver.
25/04/25 14:28:35 INFO ResourceUtils: =====
25/04/25 14:28:35 INFO SparkContext: Submitted application: Spark Pi
25/04/25 14:28:35 INFO ResourceProfile: Default ResourceProfile created, executor resources: Map(cores -> name: cores, a
mount: 1, script: , vendor: , memory -> name: memory, amount: 1024, script: , vendor: , offHeap -> name: offHeap, amount
: 0, script: , vendor: ), task resources: Map(cpus -> name: cpus, amount: 1.0)
```

Pengujian menggunakan SparkPi 10 berjalan dengan baik.