

JOBSHEET 6

Deep Dive into Apache Spark

MATA KULIAH

BIG DATA



KELAS SIB 3E

Oleh:

Agta Fadjrin Aminullah 2241760072

PROGRAM STUDI D4 SISTEM INFORMASI BISNIS

POLITEKNIK NEGERI MALANG

2025

Tugas Praktikum

Mengkonfigurasi environment variables .bashrc

```
# this, if it's already enabled in /etc/bash.bashrc and /etc/profile
# sources /etc/bash.bashrc).
if ! shopt -oq posix; then
  if [ -f /usr/share/bash-completion/bash_completion ]; then
    . /usr/share/bash-completion/bash_completion
  elif [ -f /etc/bash_completion ]; then
    . /etc/bash_completion
  fi
fi

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=192.168.109.158

hadoopuser@hadoop-datanode1:~$ _
```

Penjelasan: Baris pertama digunakan untuk menentukan lokasi folder instalasi Spark di sistem, “PATH=” agar bisa jalankan perintah spark-submit, spark-shell, “LD_LIBRARY_PATH=” agar spark bisa akses library native milik hadoop, “HADOOP_CONF_DIR=...” agar Spark tahu konfigurasi Hadoop berada di mana, “SPARK_MASTER_HOST=” IP dari master.

Konfigurasi spark-env.sh

```
# You might get better performance to enable these options if using native BLAS (see SPARK-21005).
# - 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=$(readlink -f /usr/bin/java | sed "s:bin/java::")
export SPARK_MASTER_HOST=192.168.109.158
export HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoop
export SPARK_WORKER_CORES=2
export SPARK_WORKER_MEMORY=4g
export SPARK_DRIVER_MEMORY=2g
export SPARK_EXECUTOR_MEMORY=2g

hadoopuser@hadoop-datanode1:~$
```

Penjelasan: “JAVA_HOME=” otomatis mendeteksi lokasi instalasi Java. “SPARK_MASTER_HOST=” Alamat node utama Spark, “HADOOP_CONF_DIR=” Lokasi konfigurasi Hadoop, “SPARK_WORKER_CORES=2” Setiap worker pakai 2 CPU core, “SPARK_WORKER_MEMORY=4g” Setiap worker punya 4 GB Ram, “SPARK_DRIVER_MEMORY=2g” Memori untuk Spark driver, “SPARK_EXECUTOR_MEMORY=2g” Memori tiap executor yang memproses data.

Menjalankan start-worker.sh spark://192.168.109.158:7077 pada hadoop dan menjalankan <http://192.168.109.158:8080> pada browser untuk cek worker node.

← → ↺

Not Secure 192.168.109.158:8080

80%

🔒 ⬇️ 👤 📄 ☰

Kotak Masuk (2.928) - ...

ChatGPT

Perplexity

SmallTalk

LeetCode

E-Learning

IELTS Online Practice T...

IELTS Listening Practic...

MySkill | Platform Pen...

IBM Data Analyst Prof...

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

Penjelasan: Dapat terlihat pada hasil di atas jika sudah saling terhubung antara master ataupun worker node di alamat IP Master 192.168.109.158.