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

Mengkonfigurasikan environment variables .bashrc

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

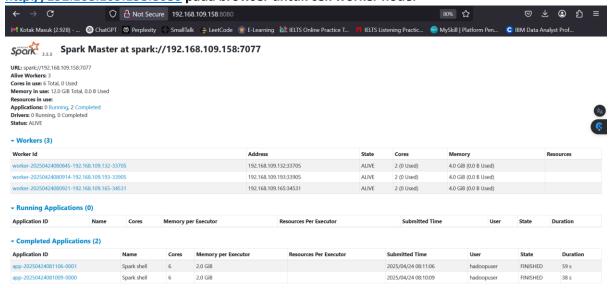
```
# -MKL_NUM_THREADS=1 Disable multi-threading of Intel MKL
# - OPENBLAS_NUM_THREADS=1 Disable multi-threading of OpenBLAS

# 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_MORKER_CORES=2
export SPARK_MORKER_MEMORY=4g
export SPARK_DRIVER_MEMORY=2g
export SPARK_EXECUTOR_MEMORY=2g
hadoopuser@hadoop-datanode1:~$
```

Penjelasan: "JAVA_HOME=" otomatis mendteksi 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.



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