

**LAPORAN  
BIG DATA  
JOBSHEET 6**



Disusun Oleh :

**Zanuar Aldi Syahputra  
2241760062  
SIB-3E**

**JURUSAN TEKNOLOGI INFORMASI  
PROGRAM STUDI D-IV SISTEM INFORMASI BISNIS  
POLITEKNIK NEGERI MALANG  
2025**

## Konfigurasi *~/.bashrc*

```
GNU nano 7.2 /home/hadoopuser/./bashrc
#alias dir='dir --color=auto'
#alias vdir='vdir --color=auto'

alias grep='grep --color=auto'
alias fgrep='fgrep --color=auto'
alias egrep='egrep --color=auto'
fi

# colored GCC warnings and errors
#export GCC_COLORS='error=01:31:warning=01:35:note=01:36:caret=01:32:locus=01:quote=01'

# some more ls aliases
alias ll='ls -alF'
alias la='ls -A'
alias l='ls -CF'

# Add an "alert" alias for long running commands.  Use like so:
# sleep 10; alert
alias alert='notify-send --urgency=low -i "${?} = 0" && echo terminal || echo error)' "${history|tail -n1|sed -e '\s/\s*[0-9]\s+\s*//;s/[:&]\s*alert$//'\s

# Alias definitions.
# You may want to put all your additions into a separate file like
# ~/.bash_aliases, instead of adding them here directly.
# See /usr/share/doc/bash-doc/examples in the bash-doc package.

if [ -f ~/.bash_aliases ]; then
    . ~/.bash_aliases
fi

# enable programmable completion features (you don't need to enable
# 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
-
[?] Help [?] Write Out [?] Where Is [?] Cut [?] Execute [?] Location [?] Undo [?] Set Mark [?] To Bracket [?] Previous
[?] Exit [?] Read File [?] Replace [?] Paste [?] Justify [?] Go To Line [?] Redo [?] Copy [?] Where Was [?] Next
```

## Konfigurasi *spark-env.sh*

```
GNU nano 7.2 /opt/spark/conf/spark-env.sh
# - YARN_CONF_DIR, to point Spark towards YARN configuration files when you use YARN

# Options for the daemons used in the standalone deploy mode
# - SPARK_MASTER_HOST, to bind the master to a different IP address or hostname
# - SPARK_MASTER_PORT / SPARK_MASTER_WEBUI_PORT, to use non-default ports for the master
# - SPARK_MASTER_OPTS, to set config properties only for the master (e.g. "-Dx=y")
# - SPARK_WORKER_CORES, to set the number of cores to use on this machine
# - SPARK_WORKER_MEMORY, to set how much total memory workers have to give executors (e.g. 1000m, 2g)
# - SPARK_WORKER_PORT / SPARK_WORKER_WEBUI_PORT, to use non-default ports for the worker
# - SPARK_WORKER_DIR, to set the working directory of worker processes
# - SPARK_WORKER_OPTS, to set config properties only for the worker (e.g. "-Dx=y")
# - SPARK_DAEMON_MEMORY, to allocate to the master, worker and history server themselves (default: 1g).
# - SPARK_HISTORY_OPTS, to set config properties only for the history server (e.g. "-Dx=y")
# - SPARK_SHUFFLE_OPTS, to set config properties only for the external shuffle service (e.g. "-Dx=y")
# - SPARK_DAEMON_JAVA_OPTS, to set config properties for all daemons (e.g. "-Dx=y")
# - SPARK_DAEMON_CLASSPATH, to set the classpath for all daemons
# - SPARK_PUBLIC_DNS, to set the public dns name of the master or workers

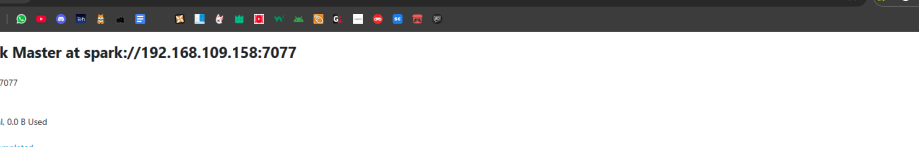
# Options for launcher
# - SPARK_LAUNCHER_OPTS, to set config properties and Java options for the launcher (e.g. "-Dx=y")

# 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=$(readlink -f /usr/bin/java | sed "s:bin/java::")
export SPARK_MASTER_HOST=192.168.109.158
export HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoop
export SPARK_WORKER_CORES=2
export SPARK_WORKER_MEMORY=4g
export SPARK_DRIVER_MEMORY=2g
export SPARK_EXECUTOR_MEMORY=2g
-
[?] Help [?] Write Out [?] Where Is [?] Cut [?] Execute [?] Location [?] Undo [?] Set Mark [?] To Bracket [?] Previous
[?] Exit [?] Read File [?] Replace [?] Paste [?] Justify [?] Go To Line [?] Redo [?] Copy [?] Where Was [?] Next
```

## Menjalankan Spark di browser



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-20250424081105-0001	Spark shell	6	2.0 GiB		2025/04/24 08:11:06	hadoopuser	FINISHED	59 s
app-20250424081109-0000	Spark shell	6	2.0 GiB		2025/04/24 08:10:09	hadoopuser	FINISHED	38 s

## Hasil uji coba Spark

```
hadoopuser@hadoop-namenode:~$ spark-shell --master spark://192.168.109.158:7077
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
25/04/24 08:11:06 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Spark context Web UI available at http://hadoop-namenode:4040
Spark context available as 'sc' (master = spark://192.168.109.158:7077, app id = app-20250424081106-0001).
Spark session available as 'spark'.
Welcome to

  ____      _
 / ___|    / \
| |  | |  / _ \
| |  | | / ___ \
| |  | |/_/   \_\
| |  | |
| |  | |
|_|  |_|

 version 3.5.5

Using Scala version 2.12.18 (OpenJDK 64-Bit Server VM, Java 1.8.0_442)
Type in expressions to have them evaluated.
Type :help for more information.

scala> _
```