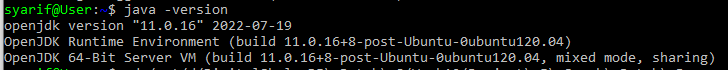
OS yang digunakan Linux Ubuntu 20.04

# Install Java terlebih dahulu

sudo apt update -> update ubuntu terlebih dahulu

sudo apt install default-jdk -> install java di ubuntu

java –version -> cek versi JAVA



# Install spark

wget <https://dlcdn.apache.org/spark/spark-3.3.0/spark-3.3.0-bin-hadoop3.tgz> -> download spark

tar xvf spark-3.3.0-bin-hadoop3.tgz -> unzip/extract spark yang sudah di download sebelumnya

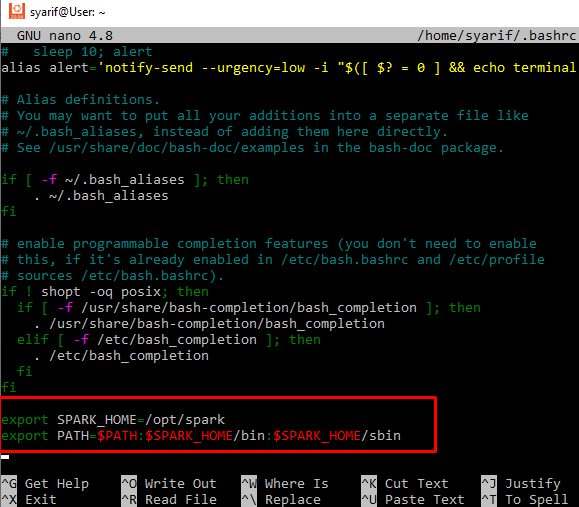
sudo mv spark-3.3.0-bin-hadoop3 /opt/spark -> move/pindahkan folder spark hasil extract ke folder /opt

nano ~/.bashrc -> edit file bashrc dan tambahkan

export SPARK\_HOME=/opt/spark

export PATH=$PATH:$SPARK\_HOME/bin:$SPARK\_HOME/sbin



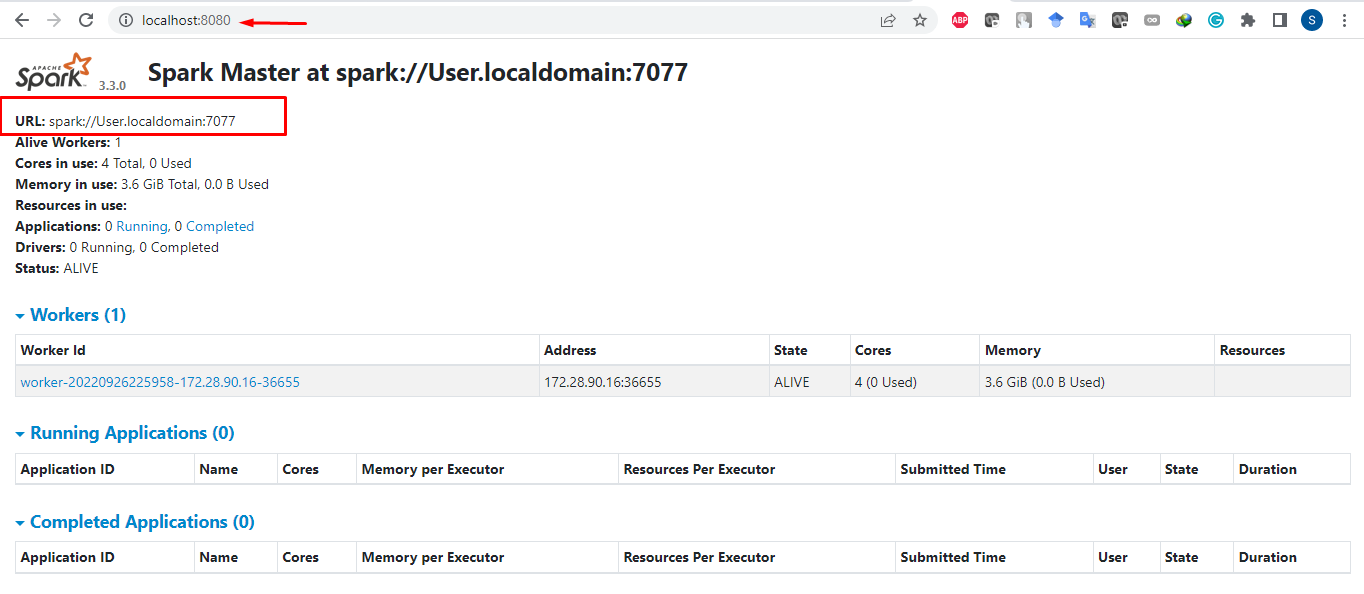


source ~/.bashrc -> mengaktifkan perubahan



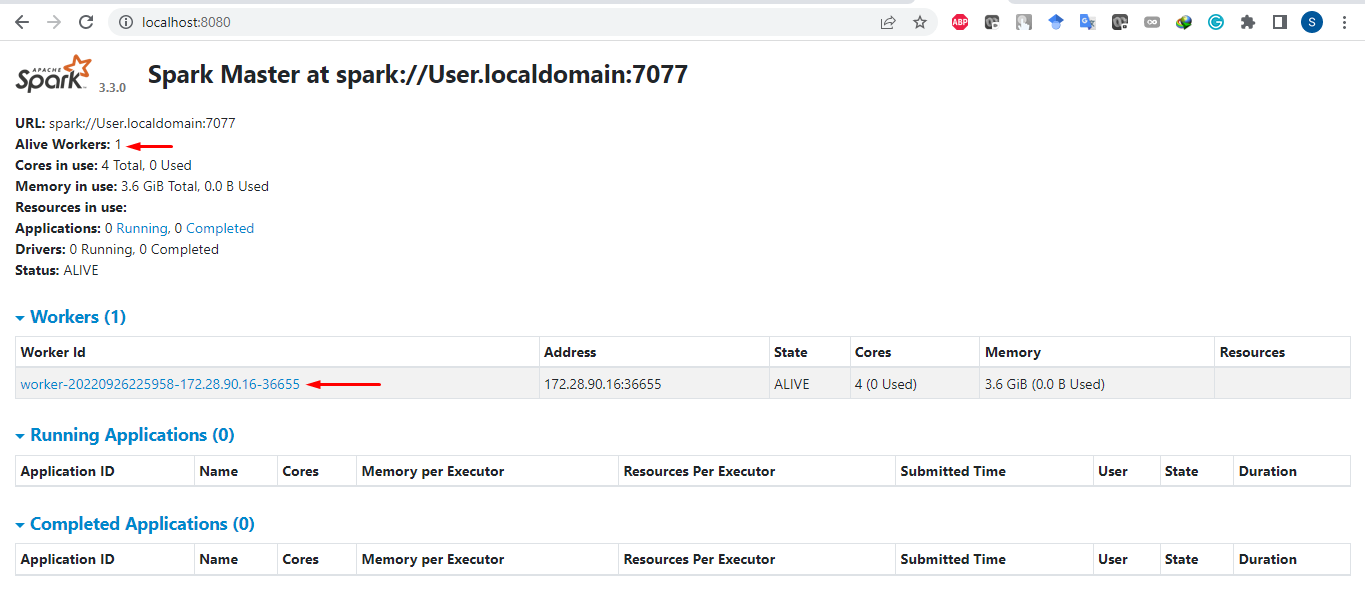
start-master.sh -> Start a standalone master server





start-worker.sh spark://User.localdomain:7077 -> jalankan worker spark





Jalankan Program Pyhon Bacth Processing

Upload ke git

git config --global user.mail [syarifudinzuhri@yahoo.com](mailto:syarifudinzuhri@yahoo.com)

git config --global user.name "syarifudinzuhri"

git init

git add .

git commit -m "Initial Commit"

git remote add origin https://github.com/syarifudinzuhri/de8\_project5\_spark\_batch\_processing.git

git branch -M main

