

**LAPORAN UJIAN TENGAH SEMESTER
MATA KULIAH BIG DATA
Jobsheet 7 – Spark Docker**



**Dosen Pengampu:
M. Hasyim Ratsanjani, S.Kom., M.Kom.**

Disusun Oleh:

Yuma Rakha Samodra Sikayo

NIM.2241720194

**PROGRAM STUDI D4 TEKNIK INFORMATIKA
JURUSAN TEKNOLOGI INFORMASI
POLITEKNIK NEGERI MALANG
2025**

Langkah Langkah

1. Pull Image Spark Resmi

```
Terminal

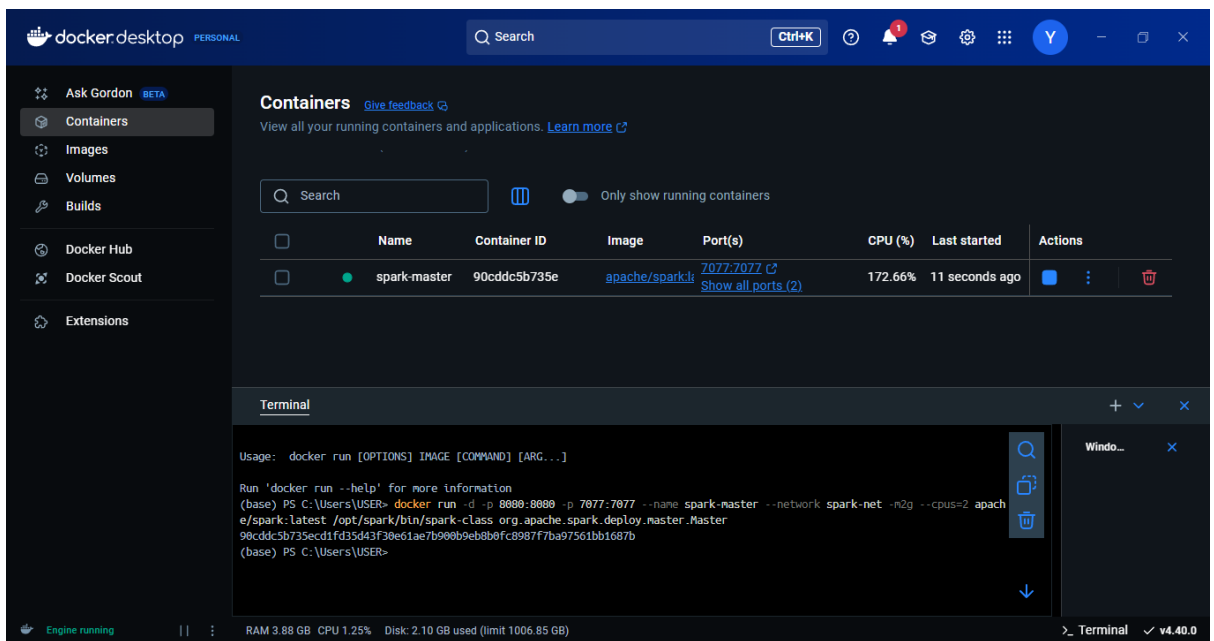
f937e0a2086c: Pull complete
0f3083818c14: Pull complete
b072aa17899d: Pull complete
5762a181dda2: Pull complete
1ba3910f6ba2: Pull complete
4f4fb700ef54: Pull complete
391ef20df327: Pull complete
Digest: sha256:39321d67b23e2e0953f81b60778f74bf40c40a18dfb0e881e6a38593af60afa1
Status: Downloaded newer image for apache/spark:latest
docker.io/apache/spark:latest
(base) PS C:\Users\USER>
```

2. Menjalankan Spark Master

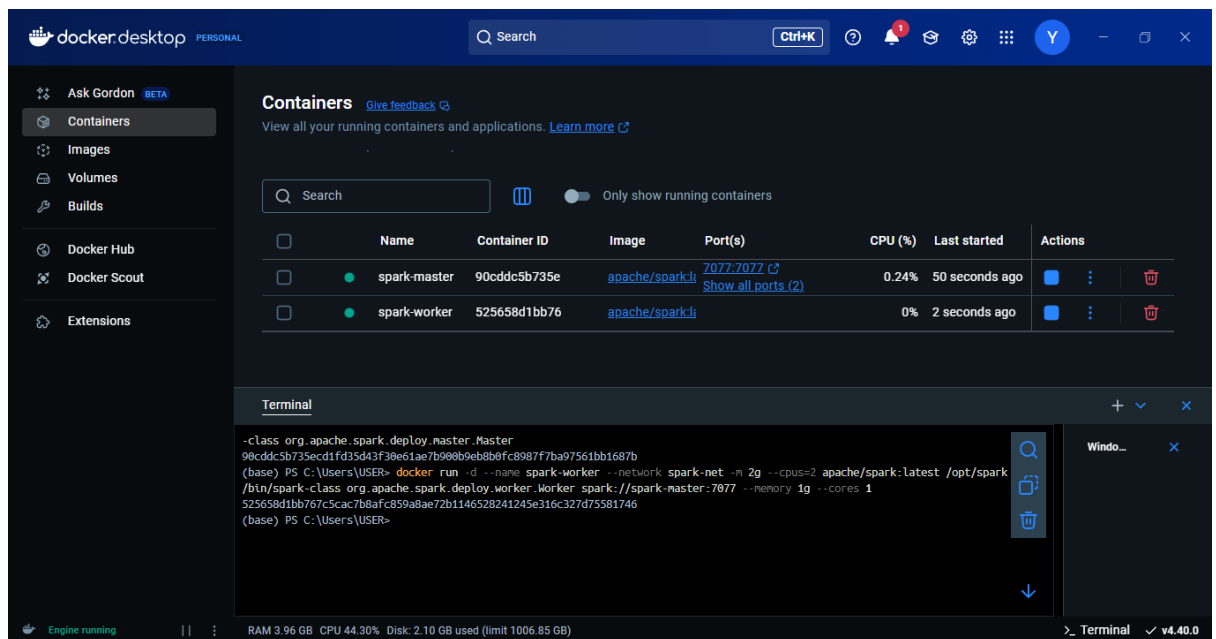
Sebelumnya buat docker network sebagai berikut

```
(base) PS C:\Users\USER> docker network create spark-net
49470ceb7a55fd2299e618f5295aa12ac3b15dc54adb430be494d897dbbb823f
(base) PS C:\Users\USER>
```

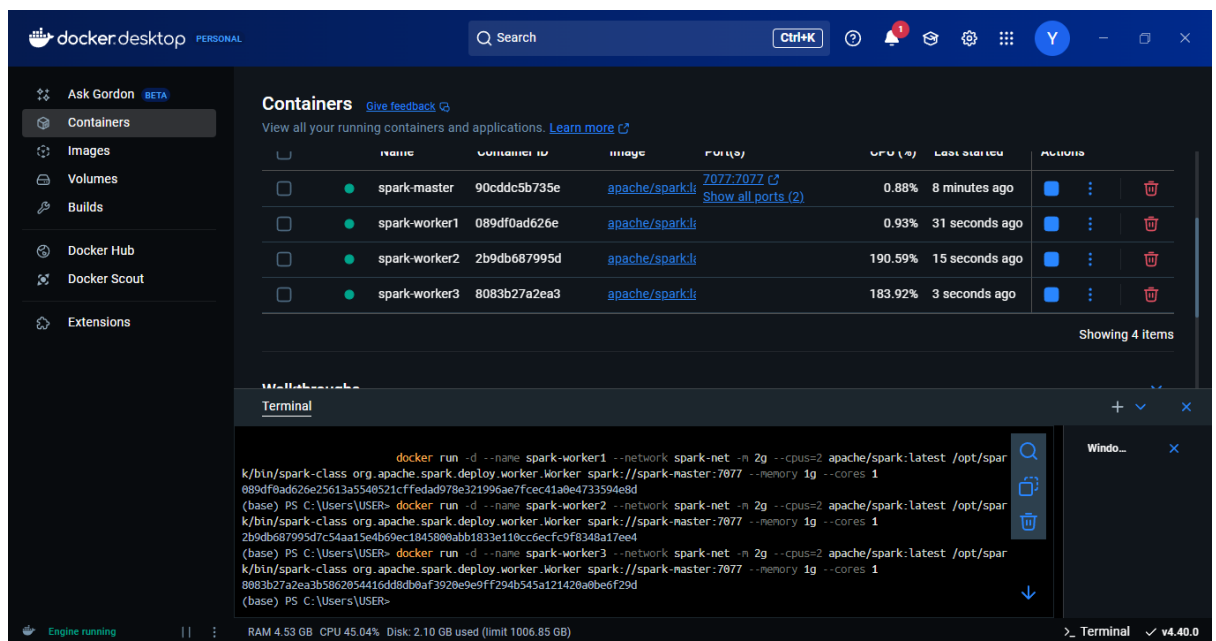
Menjalankan spark-master dalam network tersebut



3. Menjalankan Spark Worker



Membuat 3 worker



4. Mengakses Spark Web UI

The screenshot shows the Spark Master web interface. At the top, it says "Spark Master at spark://172.18.0.2:7077". Below this, there are several status metrics:

- URL:** spark://172.18.0.2:7077
- Alive Workers:** 3
- Cores in use:** 3 Total, 0 Used
- Memory in use:** 3.0 GiB Total, 0.0 B Used
- Resources in use:**
- Applications:** 0 Running, 0 Completed
- Drivers:** 0 Running, 0 Completed
- Status:** ALIVE

Below the metrics, there is a section titled "Workers (4)" which contains a table listing the workers.

Worker Id	Address	State	Cores	Memory	Resources
worker-20250422100150-172.18.0.3-33203	172.18.0.3:33203	DEAD	1 (0 Used)	1024.0 MiB (0.0 B Used)	
worker-20250422100822-172.18.0.3-42831	172.18.0.3:42831	ALIVE	1 (0 Used)	1024.0 MiB (0.0 B Used)	
worker-20250422100839-172.18.0.4-34389	172.18.0.4:34389	ALIVE	1 (0 Used)	1024.0 MiB (0.0 B Used)	
worker-20250422100847-172.18.0.5-45057	172.18.0.5:45057	ALIVE	1 (0 Used)	1024.0 MiB (0.0 B Used)	

Below the workers table, there is a section titled "Running Applications (0)".

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

Below the running applications table, there is a section titled "Completed Applications (0)".

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

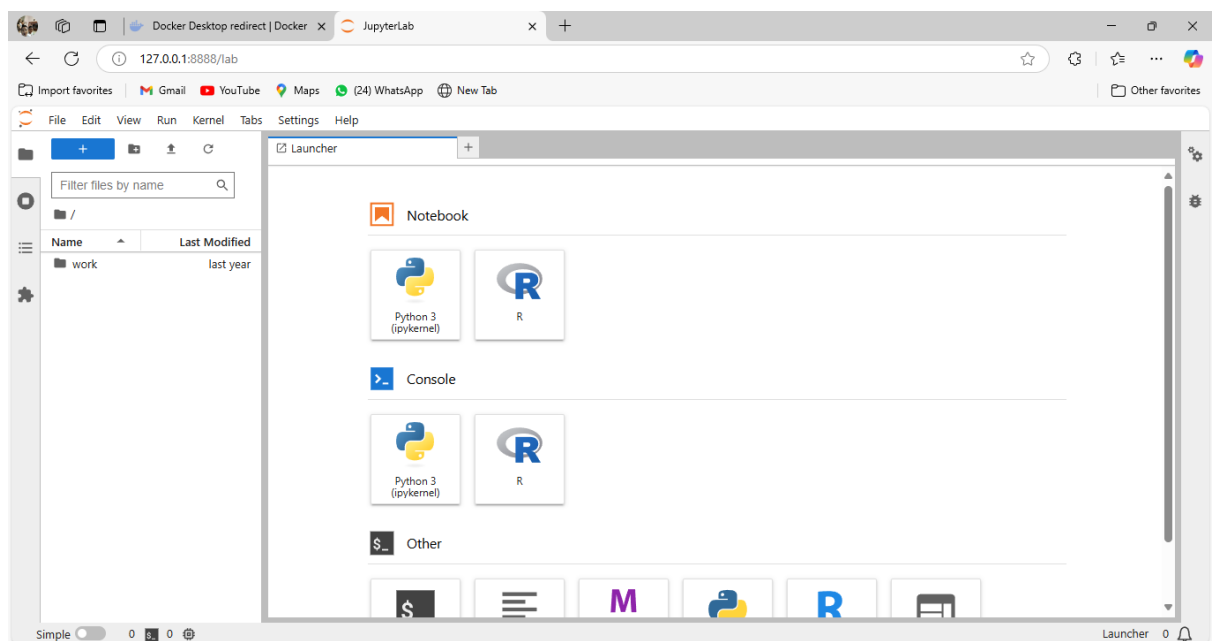
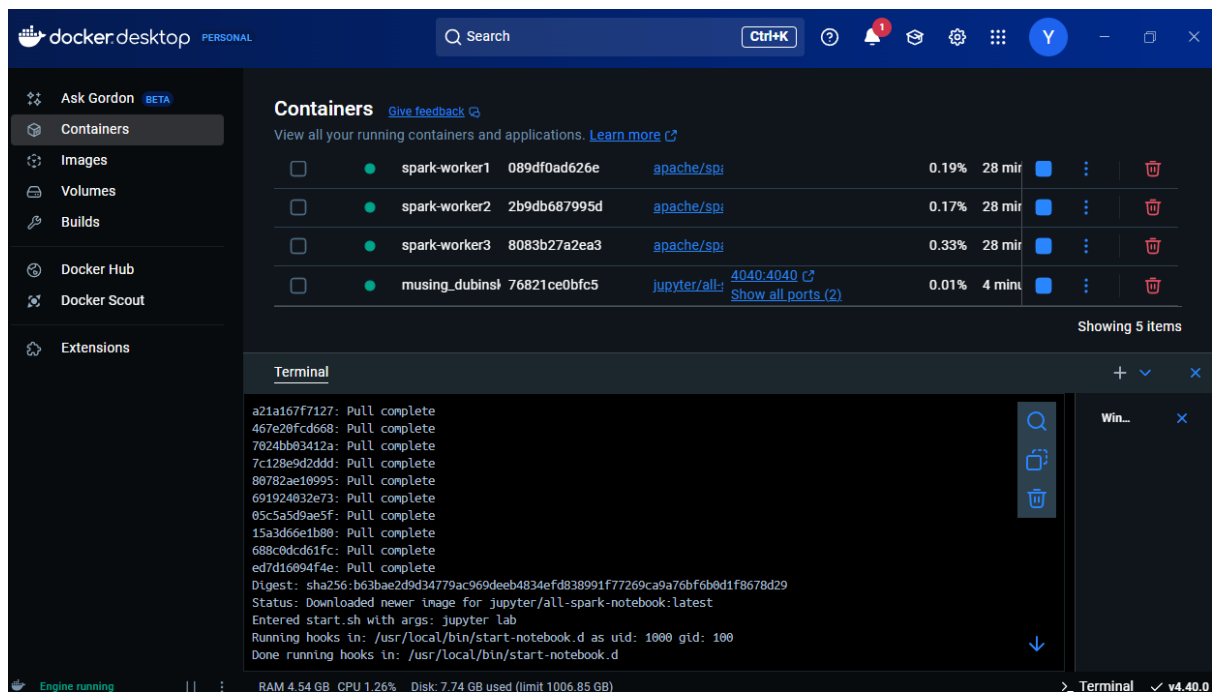
5. Menjalankan Spark Shell

The screenshot shows the Docker Desktop application window. The top bar is dark blue with the Docker logo and 'docker desktop PERSONAL' text. A search bar is on the right. The sidebar on the left contains navigation icons and labels: 'Ask Gordon (BETA)', 'Containers' (selected), 'Images', 'Volumes', 'Builds', 'Docker Hub', 'Docker Scout', and 'Extensions'. The main panel is titled 'Containers' with a subtitle 'View all your running containers and applications. Learn more'. It features a search bar and a toggle switch 'Only show running containers'. Below is a table of containers:

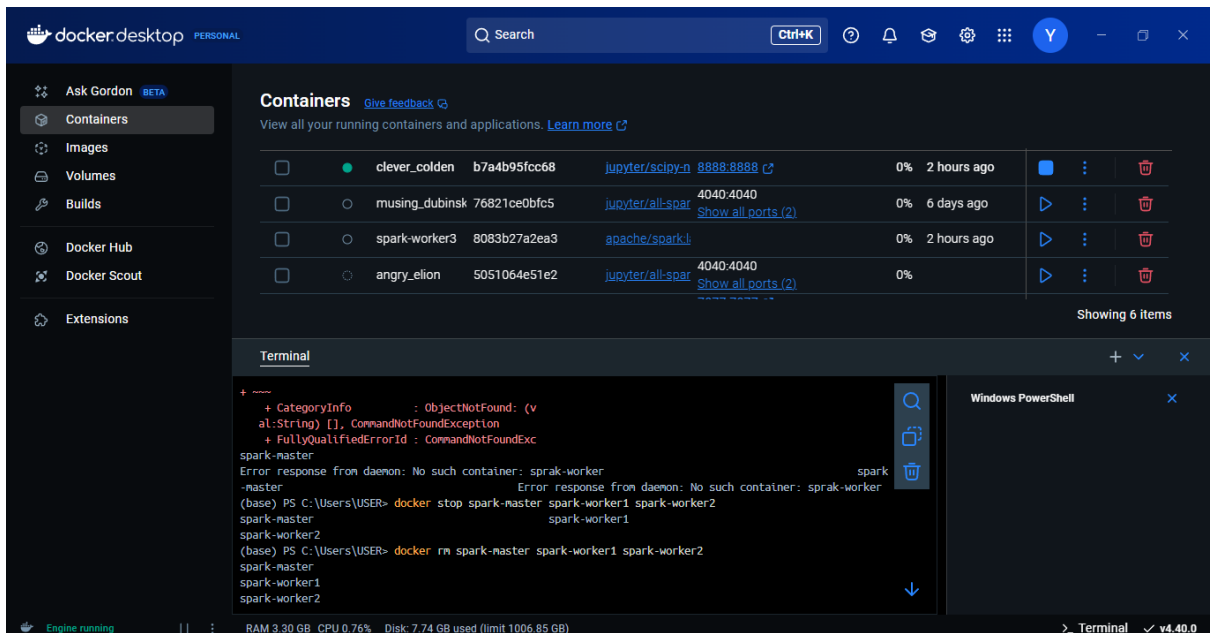
	Name	Container ID	Image	Port(s)	CPU (%)	Last s	Actions
<input type="checkbox"/>	spark-master	90cdc5b735e	apache/spark	7077-7077	0.26%	12 min	[Stop] [Refresh] [Delete]
<input type="checkbox"/>	spark-worker1	089df0ad626e	apache/spark		0.44%	5 min	[Stop] [Refresh] [Delete]
<input type="checkbox"/>	spark-worker2	2b9db687995d	apache/spark		0.51%	5 min	[Stop] [Refresh] [Delete]
<input type="checkbox"/>	spark-worker3	8083b27a2ea3	apache/spark		0.5%	4 min	[Stop] [Refresh] [Delete]

At the bottom, a 'Terminal' tab is active, showing a Scala REPL session with the prompt 'scala>'. The status bar at the very bottom shows system metrics: 'RAM 4.54 GB, CPU 1.52%, Disk: 2.10 GB used (limit 1006.85 GB)' and a 'Terminal' icon.

6. Menggunakan Jupyter Notebook dengan Spark

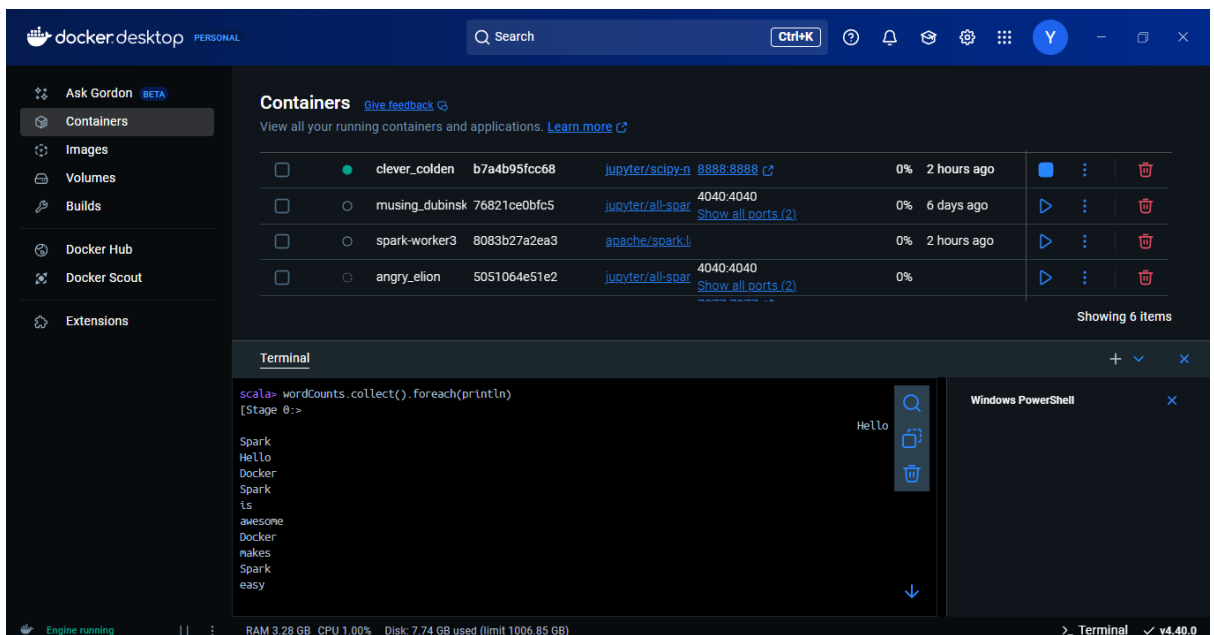


7. Meghentikan container

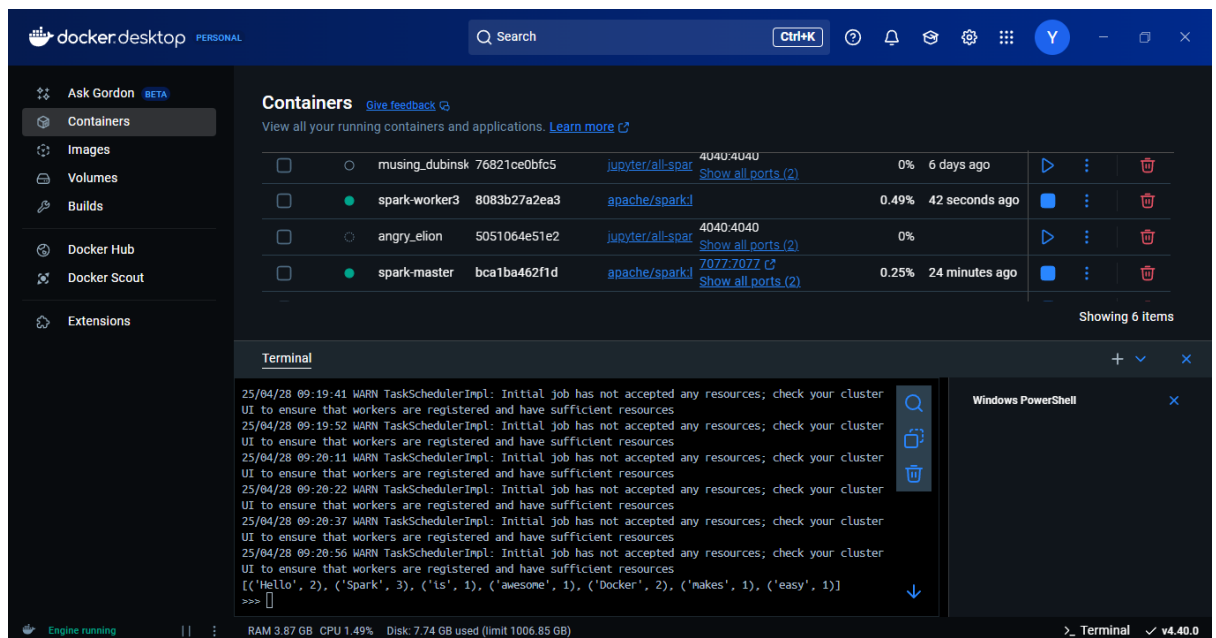
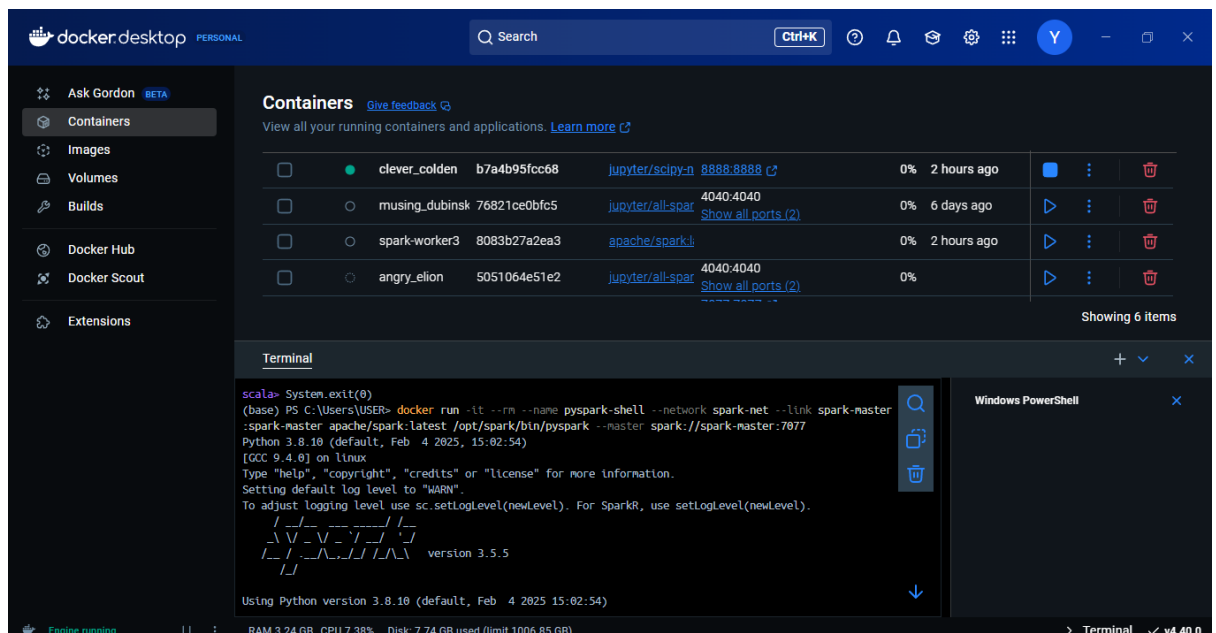


Contoh Program Word Count dengan Spark di Docker

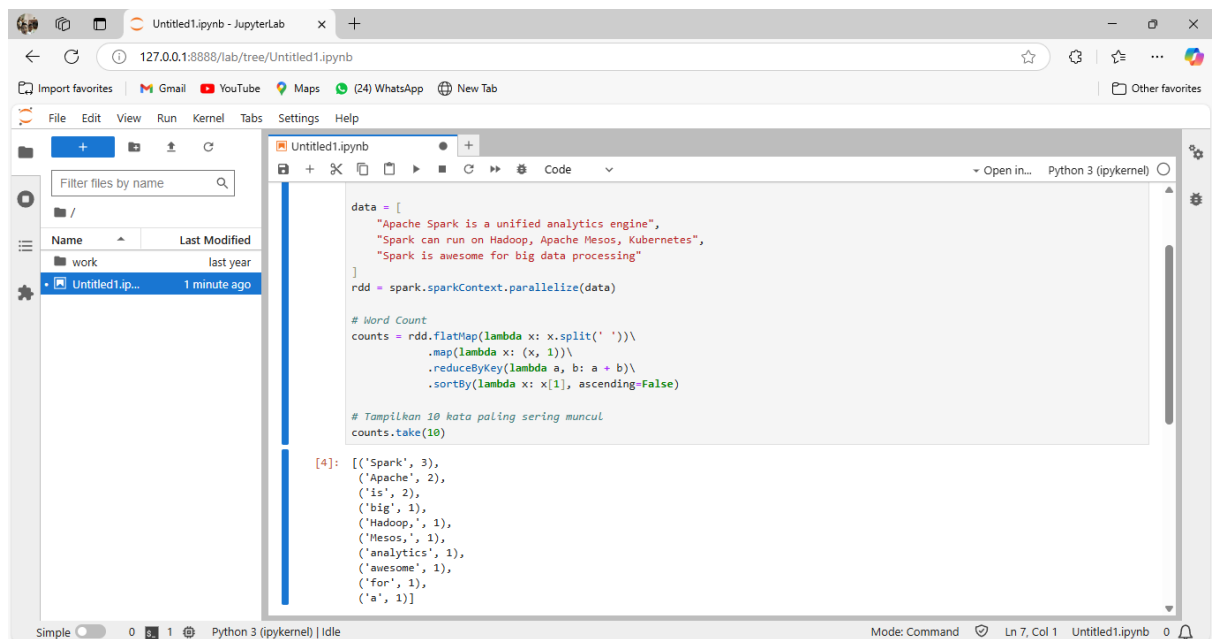
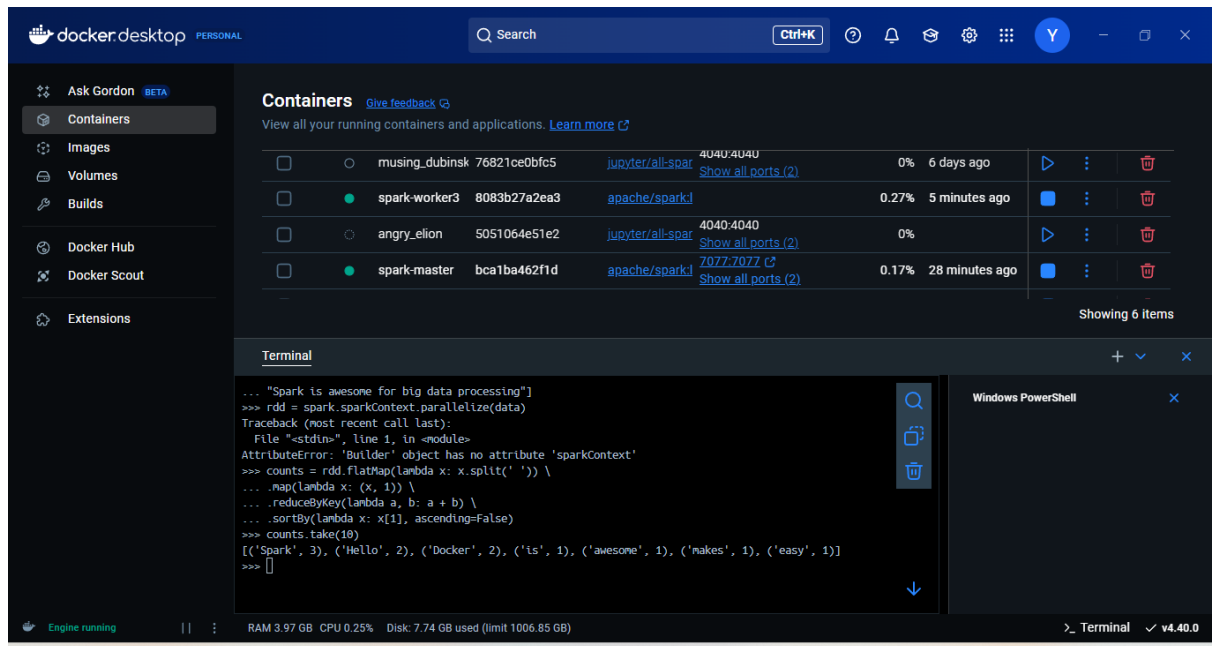
1. Menggunakan PySpark Shell



2. Menggunakan Pyspark(Python)



3. Menggunakan Jupyter Notebook



4. Menjalankan Program sebagai script

docker.desktopPERSONAL

Q Search

Ctrl+K

Ask GordonBETA

Containers

Images

Volumes

Builds

Docker Hub

Docker Scout

Extensions

Containers

Give feedback

View all your running containers and applications. Learn more

	Name	Container ID	Image	Port(s)	CPU (%)	Last started	Actions
<input type="checkbox"/>	spark-jupyter	f37f5800252d	jupyter/all-spar	8888:8888	32.7%	37 minutes ago	
<input type="checkbox"/>	spark-master	35f2e8d1da03	apache/spark-l	7077:7077	0.14%	40 minutes ago	
<input type="checkbox"/>	spark-worker	b415baa21376	apache/spark-l		0%	5 minutes ago	
<input type="checkbox"/>	spark-worker1	148d476292c7	apache/spark-l		0.13%	48 seconds ago	

Showing 7 items

Terminal

25/04/28 12:26:19 INFO DAGScheduler: ResultStage 1 (collect at /app/wordcount.py:16) Finished in 0.634 s

25/04/28 12:26:19 INFO DAGScheduler: Job 0 is finished. Cancelling potential speculative or zombie tasks for this job

25/04/28 12:26:19 INFO TaskSchedulerImpl: Killing all running tasks in stage 1: Stage finished

25/04/28 12:26:19 INFO DAGScheduler: Job 0 finished: collect at /app/wordcount.py:16, took 8.787042 s

Hello: 2

Spark: 2

ts: 1

awesome: 1

Docker: 1

25/04/28 12:26:19 INFO SparkContext: SparkContext is stopping with exitCode 0.

25/04/28 12:26:19 INFO SparkUI: Stopped Spark web UI at http://ae671f55356c:4040

Windo...

Engine running

RAM 3.93 GB CPU 4.23% Disk: 8.67 GB used (limit 1006.85 GB)

Terminal v4.40.0