Nama : Syava Aprilia Puspitasari

NIM : 2241760129 / 18

# Interaksi dengan Spark di Lingkungan Windows Menggunakan Docker

Dalam praktikum ini kita akan menjalankan Apache Spark di Windows menggunakan Docker dan mencoba membuat job sederhana dengan berbagai macam alternatif cara.

Prasyarat

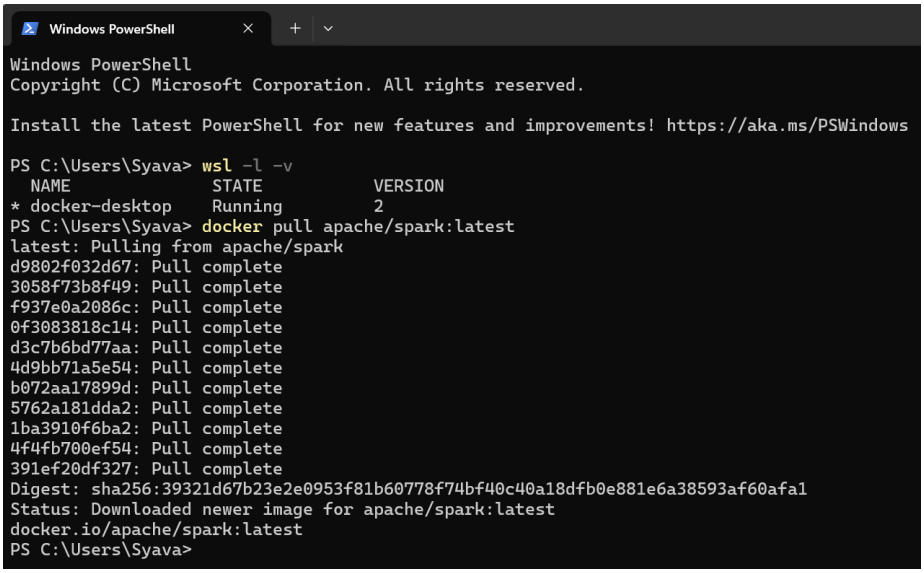
1. Windows 10/11 (64-bit) dengan versi Pro, Enterprise, atau Education

2. Docker Desktop untuk Windows diinstal dan berjalan

3. WSL 2 (Windows Subsystem for Linux versi 2) diaktifkan

Langkah-langkah

1. Pull Image Spark Resmi



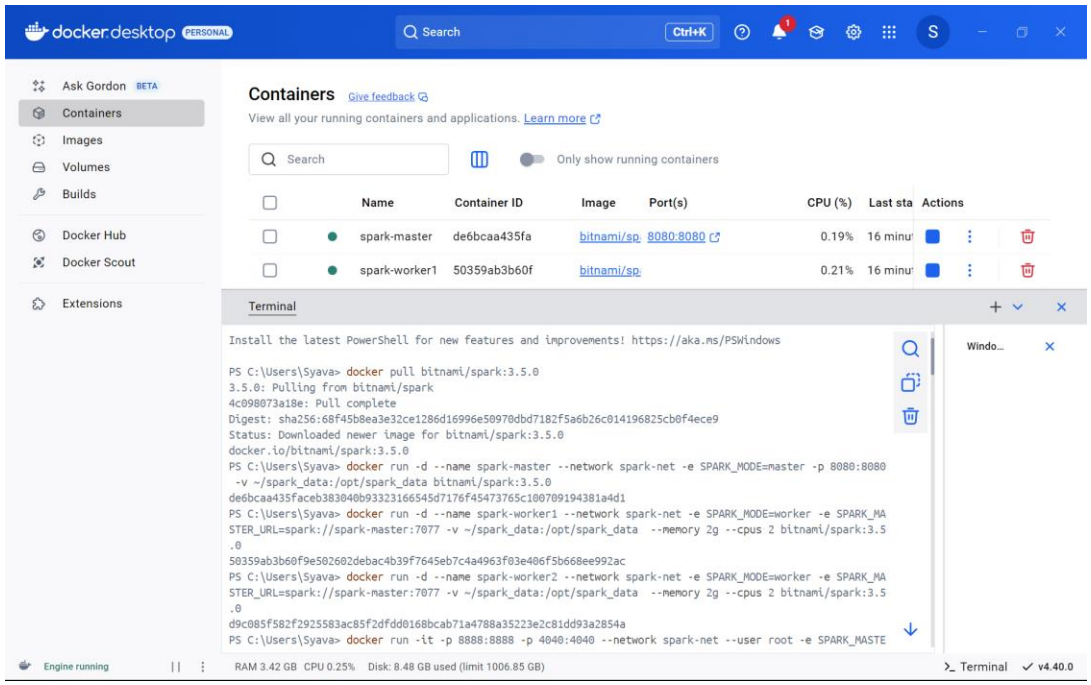

## Menjalankan Master

docker run -d --name spark-master --network spark-net -e SPARK_MODE=master -p 8080:8080 -v ~/spark_data:/opt/spark_data bitnami/spark:3.5.0

## Menjalankan Worker

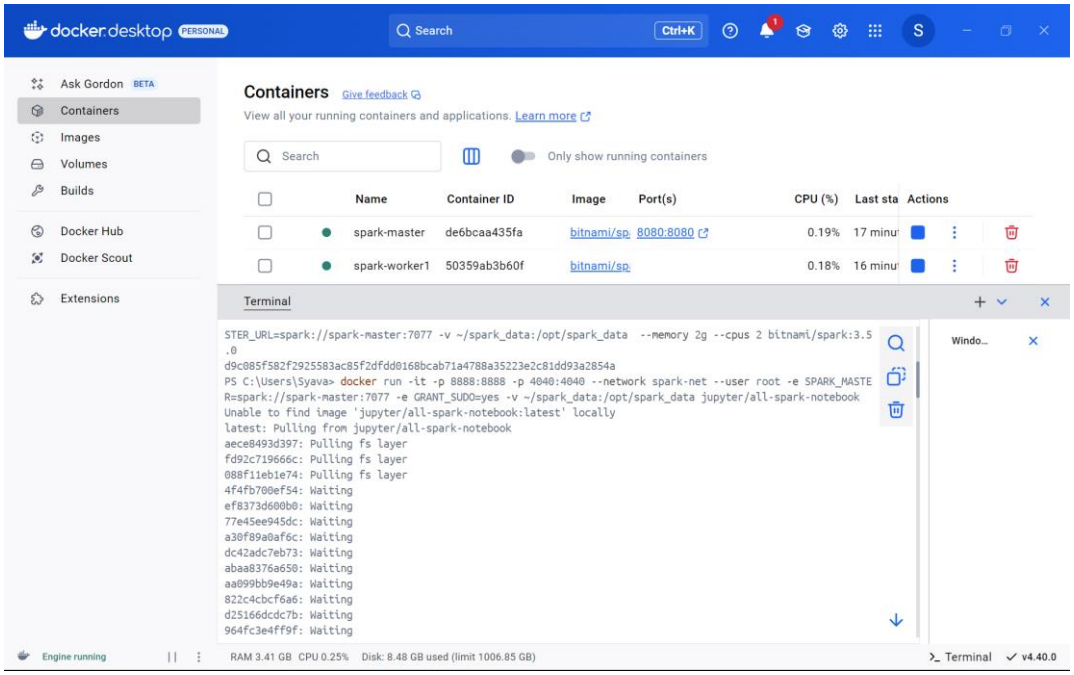
docker run -d --name spark-worker1 --network spark-net -e SPARK_MODE=worker -e SPARK_MASTER_URL=spark://spark-master:7077 -v ~/spark_data:/opt/spark_data --memory 2g --cpus 2 bitnami/spark:3.5.0

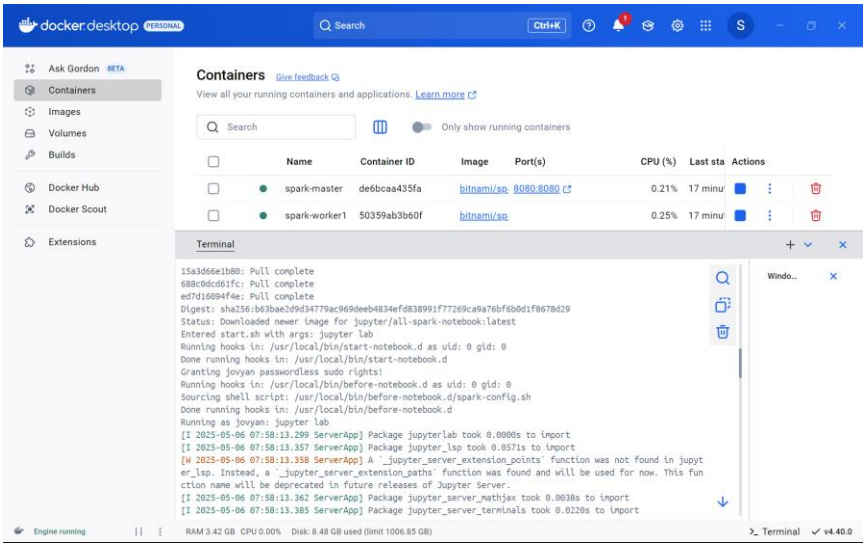
docker run -d --name spark-worker2 --network spark-net -e SPARK_MODE=worker -e SPARK_MASTER_URL=spark://spark-master:7077 -v ~/spark_data:/opt/spark_data  --memory 2g --cpus 2 bitnami/spark:3.5.0

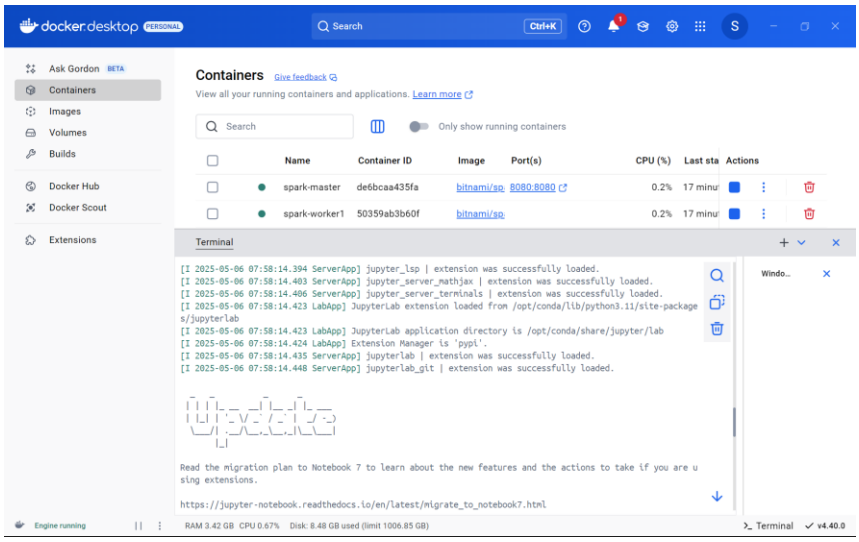


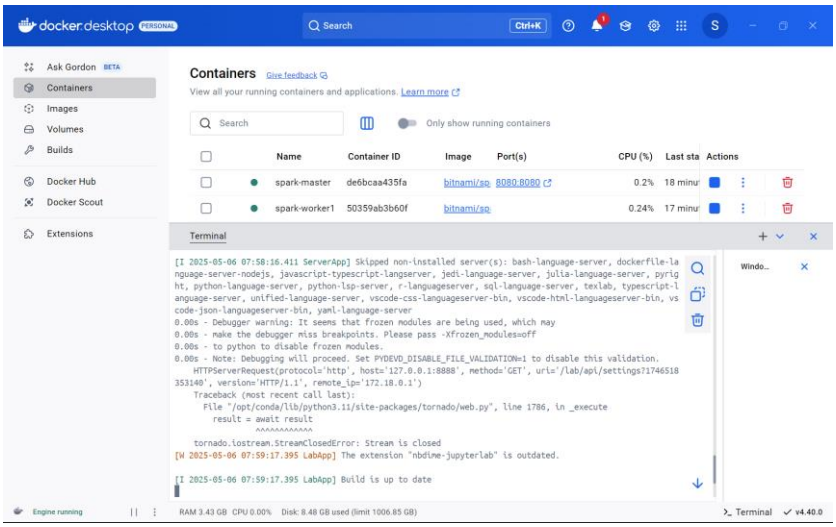

Menjalankan Spark Shell lalu Menggunakan Jupyter Notebook dengan Spark

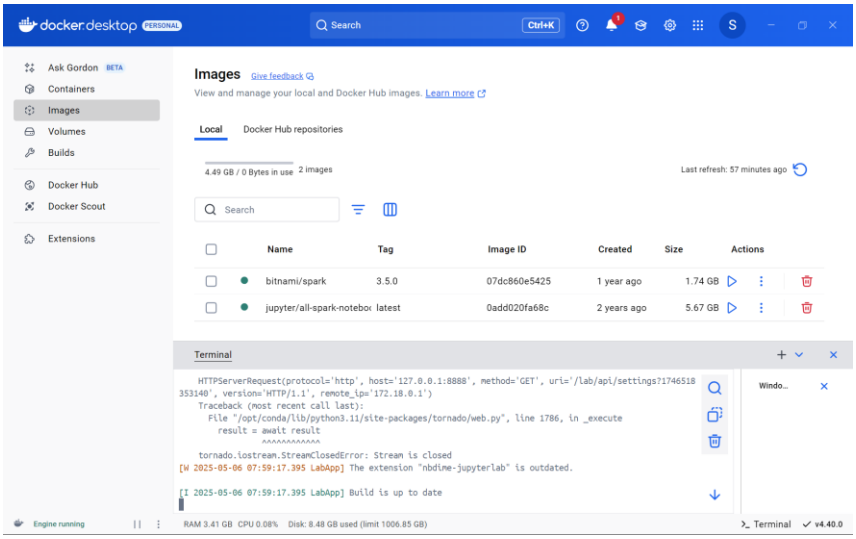
docker run -it -p 8888:8888 -p 4040:4040 --network spark-net --user root -e SPARK_MASTER=spark://spark-master:7077 -e GRANT_SUDO=yes -v ~/spark_data:/opt/spark_data jupyter/all-spark-notebook

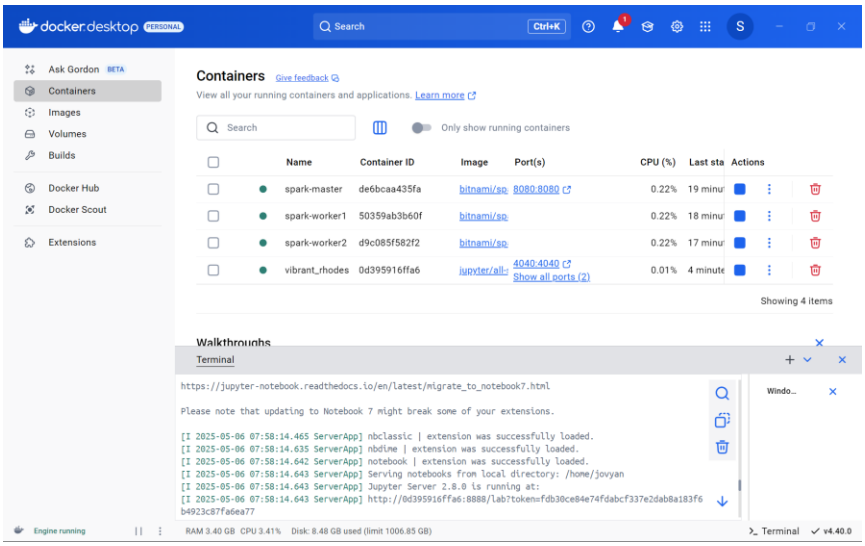

docker desktop PERSONAL
Search Ctrl+K
Ask Gordon BETA
Containers
Images
Volumes
Builds
Docker Hub
Docker Scout
Extensions
Containers Give feedback
View all your running containers and applications. Learn more
Search
Only show running containers
Name Container ID Image Port(s) CPU (%) Last sta Actions
spark-master de6bcaa435fa bitnami/sp 8080:8080 0.19% 17 minu
spark-worker1 50359ab3b60f bitnami/sp 0.18% 16 minu
Terminal
STER_URL=spark://spark-master:7077 -v ~/spark_data:/opt/spark_data --memory 2g --cpus 2 bitnami/spark:3.5
.0
d9c085f582f2925583ac85f2dfdd0168bcab71a4788a35223e2c81dd93a2854a
PS C:\Users\Syava> docker run -it -p 8888:8888 -p 4040:4040 --network spark-net --user root -e SPARK_MASTE
R=spark://spark-master:7077 -e GRANT_SUDO=yes -v ~/spark_data:/opt/spark_data jupyter/all-spark-notebook
Unable to find image 'jupyter/all-spark-notebook:latest' locally
latest: Pulling from jupyter/all-spark-notebook
aece8493d397: Pulling fs layer
fd92c719666c: Pulling fs layer
088f11eb1e74: Pulling fs layer
4f4fb700ef54: Waiting
ef8373d600b0: Waiting
77e45ee945dc: Waiting
a30f89a0af6c: Waiting
dc42adc7eb73: Waiting
abaa8376a650: Waiting
aa099bb9e49a: Waiting
822c4cbcf6a6: Waiting
d25166dcdc7b: Waiting
964fc3e4ff9f: Waiting
Windo...
Engine running
RAM 3.41 GB CPU 0.25% Disk: 8.48 GB used (limit 1006.85 GB)
Terminal v4.40.0


docker desktop PERSONAL
Search Ctrl+K
Ask Gordon BETA
Containers
Images
Volumes
Builds
Docker Hub
Docker Scout
Extensions
Containers Give feedback
View all your running containers and applications. Learn more
Search
Only show running containers
Name Container ID Image Port(s) CPU (%) Last sta Actions
spark-master de6bcaa435fa bitnami/sp 8080:8080 0.21% 17 minu
spark-worker1 50359ab3b60f bitnami/sp 0.25% 17 minu
Terminal
15a3d66e1b80: Pull complete
688c0dcd61fc: Pull complete
ed7d16094f4e: Pull complete
Digest: sha256:b63bae2d9d34779ac969deeb4834efd838991f77269ca9a76bf6b0d1f8678d29
Status: Downloaded newer image for jupyter/all-spark-notebook:latest
Entered start.sh with args: jupyter lab
Running hooks in: /usr/local/bin/start-notebook.d as uid: 0 gid: 0
Done running hooks in: /usr/local/bin/start-notebook.d
Granting jovyan passwordless sudo rights!
Running hooks in: /usr/local/bin/before-notebook.d as uid: 0 gid: 0
Sourcing shell script: /usr/local/bin/before-notebook.d/spark-config.sh
Done running hooks in: /usr/local/bin/before-notebook.d
Running as jovyan: jupyter lab
[I 2025-05-06 07:58:13.299 ServerApp] Package jupyterlab took 0.0000s to import
[I 2025-05-06 07:58:13.357 ServerApp] Package jupyter_lsp took 0.0571s to import
[W 2025-05-06 07:58:13.358 ServerApp] A `_jupyter_server_extension_points` function was not found in jupyt
er_lsp. Instead, a `_jupyter_server_extension_paths` function was found and will be used for now. This fun
ction name will be deprecated in future releases of Jupyter Server.
[I 2025-05-06 07:58:13.362 ServerApp] Package jupyter_server_mathjax took 0.0038s to import
[I 2025-05-06 07:58:13.385 ServerApp] Package jupyter_server_terminals took 0.0220s to import
Windo...
Engine running
RAM 3.42 GB CPU 0.00% Disk: 8.48 GB used (limit 1006.85 GB)
Terminal v4.40.0

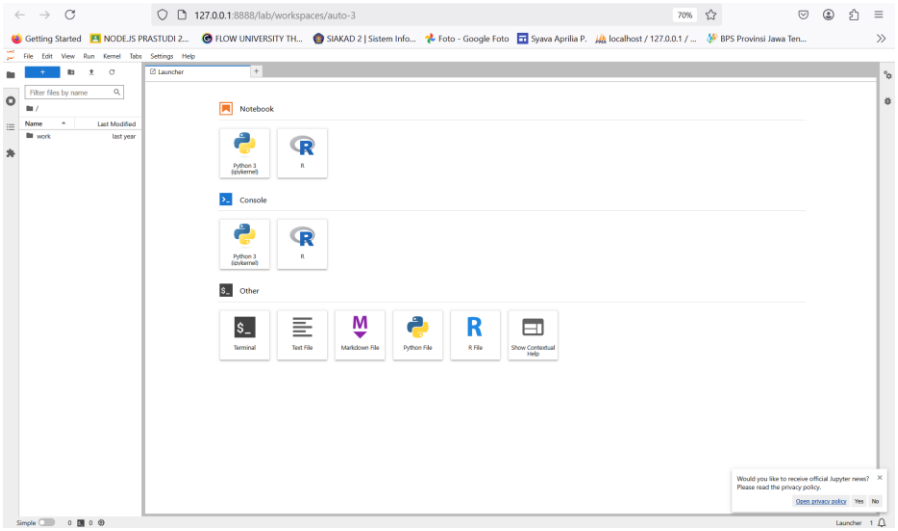

docker desktop PERSONAL
Search
Ctrl+K
Ask Gordon BETA
Containers
Images
Volumes
Builds
Docker Hub
Docker Scout
Extensions
Containers
Give feedback
View all your running containers and applications. Learn more
Search
Only show running containers
Name
Container ID
Image
Port(s)
CPU (%)
Last sta
Actions
spark-master
de6bcaa435fa
bitnami/sp
8080:8080
0.2%
17 minu
spark-worker1
50359ab3b60f
bitnami/sp
0.2%
17 minu
Terminal
[I 2025-05-06 07:58:14.394 ServerApp] jupyter_lsp | extension was successfully loaded.
[I 2025-05-06 07:58:14.403 ServerApp] jupyter_server_mathjax | extension was successfully loaded.
[I 2025-05-06 07:58:14.406 ServerApp] jupyter_server_terminals | extension was successfully loaded.
[I 2025-05-06 07:58:14.423 LabApp] JupyterLab extension loaded from /opt/conda/lib/python3.11/site-package
s/jupyterlab
[I 2025-05-06 07:58:14.423 LabApp] JupyterLab application directory is /opt/conda/share/jupyter/lab
[I 2025-05-06 07:58:14.424 LabApp] Extension Manager is 'pypi'.
[I 2025-05-06 07:58:14.435 ServerApp] jupyterlab | extension was successfully loaded.
[I 2025-05-06 07:58:14.448 ServerApp] jupyterlab_git | extension was successfully loaded.
Read the migration plan to Notebook 7 to learn about the new features and the actions to take if you are u
sing extensions.
https://jupyter-notebook.readthedocs.io/en/latest/migrate_to_notebook7.html
Windo...
Engine running
RAM 3.42 GB CPU 0.67% Disk: 8.48 GB used (limit 1006.85 GB)
Terminal
v4.40.0


docker desktop PERSONAL
Search
Ctrl+K
Ask Gordon BETA
Containers
Images
Volumes
Builds
Docker Hub
Docker Scout
Extensions
Containers
Give feedback
View all your running containers and applications. Learn more
Search
Only show running containers
Name
Container ID
Image
Port(s)
CPU (%)
Last sta
Actions
spark-master
de6bcaa435fa
bitnami/sp
8080:8080
0.2%
18 minu
spark-worker1
50359ab3b60f
bitnami/sp
0.24%
17 minu
Terminal
[I 2025-05-06 07:58:16.411 ServerApp] Skipped non-installed server(s): bash-language-server, dockerfile-la
nguage-server-nodejs, javascript-typescript-langserver, jedi-language-server, julia-language-server, pyrig
ht, python-language-server, python-lsp-server, r-languageserver, sql-language-server, texlab, typescript-l
anguage-server, unified-language-server, vscode-css-languageserver-bin, vscode-html-languageserver-bin, vs
code-json-languageserver-bin, yaml-language-server
0.00s - Debugger warning: It seems that frozen modules are being used, which may
0.00s - make the debugger miss breakpoints. Please pass -Xfrozen_modules=off
0.00s - to python to disable frozen modules.
0.00s - Note: Debugging will proceed. Set PYDEVD_DISABLE_FILE_VALIDATION=1 to disable this validation.
HTTPServerRequest(protocol='http', host='127.0.0.1:8888', method='GET', uri='/lab/api/settings?1746518
353140', version='HTTP/1.1', remote_ip='172.18.0.1')
Traceback (most recent call last):
File "/opt/conda/lib/python3.11/site-packages/tornado/web.py", line 1786, in _execute
result = await result
^^^^^^^^^^^
tornado.iostream.StreamClosedError: Stream is closed
[W 2025-05-06 07:59:17.395 LabApp] The extension "nbdime-jupyterlab" is outdated.
[I 2025-05-06 07:59:17.395 LabApp] Build is up to date
Windo...
Engine running
RAM 3.43 GB CPU 0.00% Disk: 8.48 GB used (limit 1006.85 GB)
Terminal
v4.40.0


docker desktop PERSONAL
Search
Ctrl+K
Ask Gordon BETA
Containers
Images
Volumes
Builds
Docker Hub
Docker Scout
Extensions
Images
Give feedback
View and manage your local and Docker Hub images. Learn more
Local
Docker Hub repositories
4.49 GB / 0 Bytes in use
2 images
Last refresh: 57 minutes ago
Search
Name
Tag
Image ID
Created
Size
Actions
bitnami/spark
3.5.0
07dc860e5425
1 year ago
1.74 GB
jupyter/all-spark-notebo
latest
0add020fa68c
2 years ago
5.67 GB
Terminal
HTTPServerRequest(protocol='http', host='127.0.0.1:8888', method='GET', uri='/lab/api/settings?1746518
353140', version='HTTP/1.1', remote_ip='172.18.0.1')
Traceback (most recent call last):
File "/opt/conda/lib/python3.11/site-packages/tornado/web.py", line 1786, in _execute
result = await result
^^^^^^^^^^^
tornado.iostream.StreamClosedError: Stream is closed
[W 2025-05-06 07:59:17.395 LabApp] The extension "nbdime-jupyterlab" is outdated.
[I 2025-05-06 07:59:17.395 LabApp] Build is up to date
Windo...
Engine running
RAM 3.41 GB CPU 0.08% Disk: 8.48 GB used (limit 1006.85 GB)
Terminal
v4.40.0




Setelah itu, akses Jupyter Notebook di: http://localhost:8888



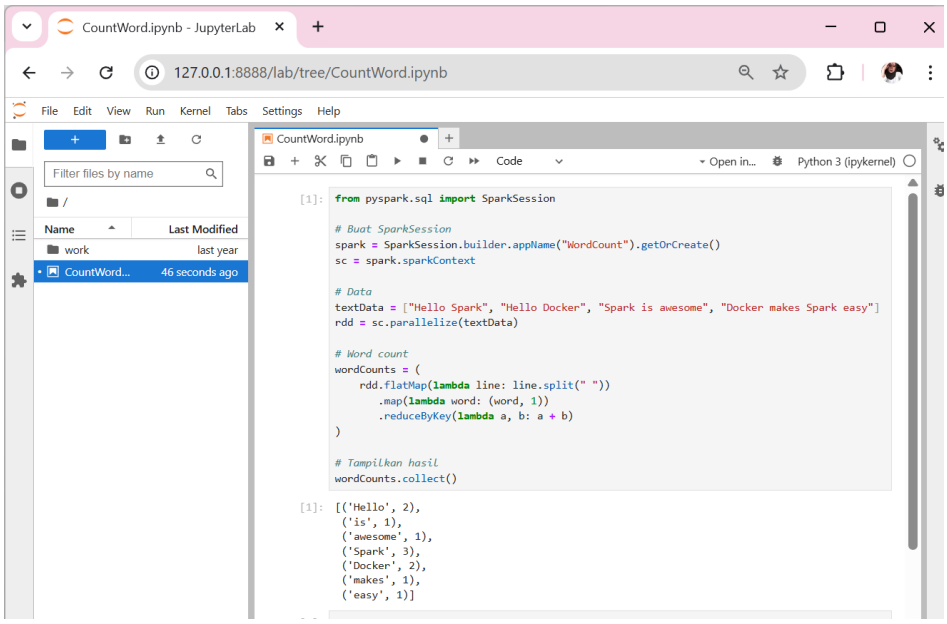

**Contoh Program Word Count dengan Spark di Docker**

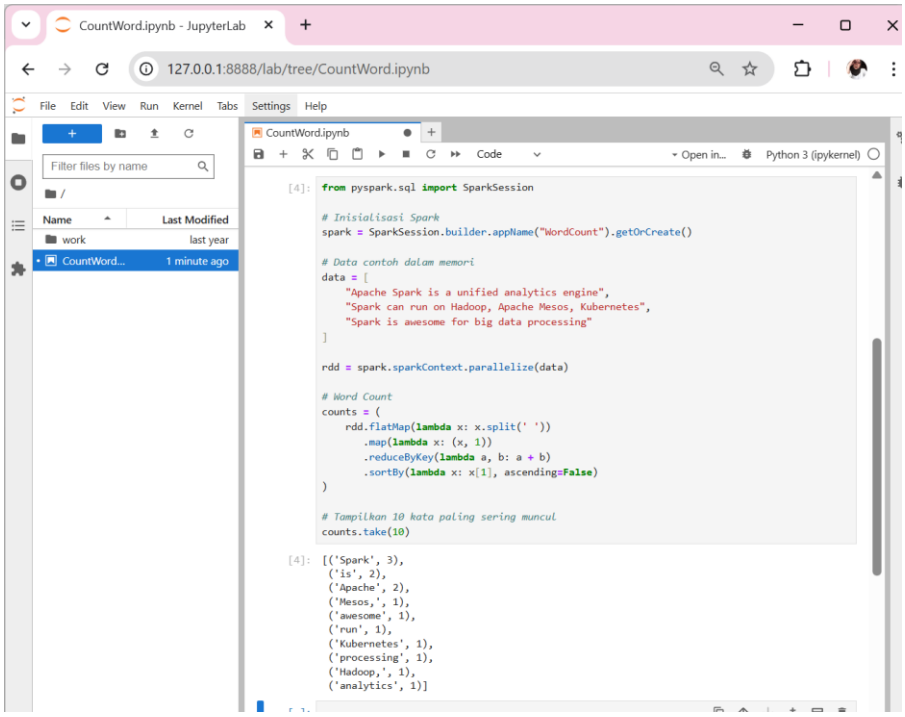
Berikut adalah contoh program Word Count (menghitung kemunculan kata) menggunakan Apache Spark

yang bisa dijalankan di lingkungan Docker:

**Cara 3: Menggunakan Jupyter Notebook**

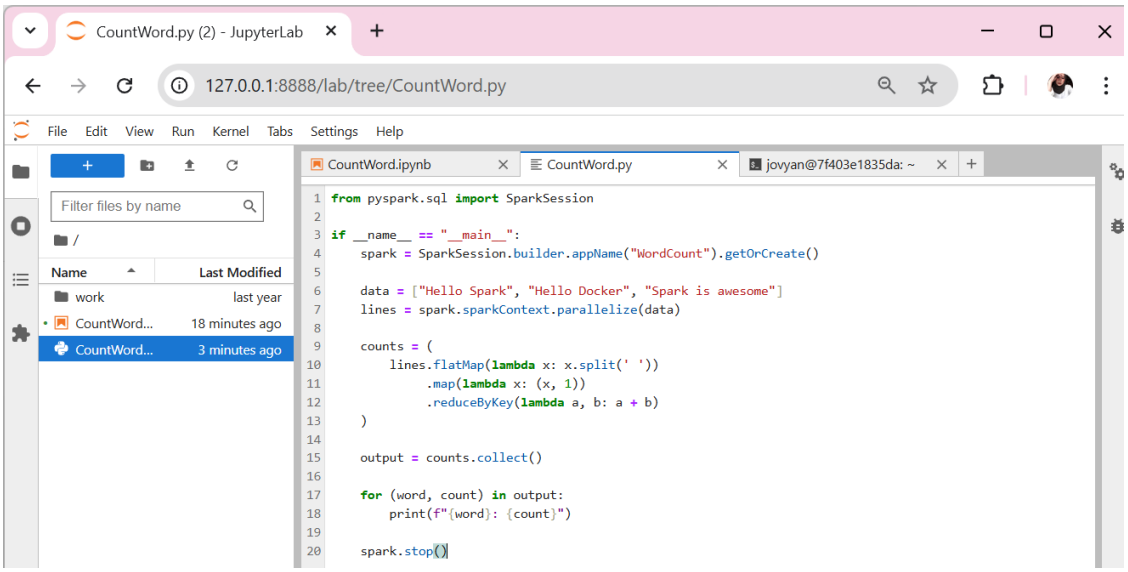
Jika Anda menggunakan Jupyter Notebook (seperti di container jupyter/all-spark-notebook):




Menjalankan Program sebagai Script

1. Buat file wordcount.py dengan isi berikut:




2. Jalankan jika sudah memastikan skrip di atas benar:
   - Simpan file CountWord.py
   - Jalankan:

     spark-submit CountWord.py