1. Clone repository <https://github.com/jupyter/docker-stacks>

A computer screen with white text

Description automatically generated

1. In docker file add

# Copy jdbc

COPY "mssql-jdbc-12.6.1.jre11.jar" "/home/jovyan/work/mssql-jdbc-12.6.1.jre11.jar"

1. mssql-jdbc-12.6.1.jre11.jar file should be in the same place as dockerfile
2. in gitbash run comand

podman build --rm --force-rm -t my-pyspark-notebook ./for\_git/docker-stacks/images/pyspark-notebook --build-arg spark\_version=3.5.1 --build-arg spark\_checksum=bfe45406c67cc4ae00411ad18cc438f51e7d4b6f14eb61e7bf6b5450897c2e8d3ab020152657c0239f253735c263512ffabf538ac5b9fffa38b8295736a9c387

1. Then run

podman run -p 8888:8888 -p 4040:4040 --user root -e JUPYTER\_ENABLE\_LAB=yes --name pyspark jupyter/pyspark-notebook

A screenshot of a computer screen

Description automatically generated

podman build --rm --force-rm -t my-pyspark-notebook ./for\_git/docker-stacks/images/pyspark-notebook --build-arg spark\_version=3.1.2 --build-arg spark\_checksum=2385CB772F21B014CE2ABD6B8F5E815721580D6E8BC42A26D70BBCDDA8D303D886A6F12B36D40F6971B5547B70FAE62B5A96146F0421CB93D4E51491308EF5D5

podman build --rm --force-rm -t my-pyspark-notebook ./for\_git/docker-stacks/images/pyspark-notebook --build-arg spark\_version=3.2.0 --build-arg spark\_checksum= ea93abaf85d6f15b2630e21f3d84f2765feddd5ab0900d92808d531246df7ed576037bbee8e02c209cdcd698ca32c22d959d926e62c91e52ce474e29bac337d8

podman build --rm --force-rm -t my-pyspark-notebook ./for\_git/docker-stacks/images/pyspark-notebook --build-arg spark\_version=3.0.0 --build-arg spark\_checksum=328D2331516FB8597F7E83A16EE378F6AD541993BC9C89B6B667E91202AAE878379B1BB3E76C85D49A126B2894FFF5AA1ABEAE5F9D3D130452E09769C226B1FB

checksum in this file : https://archive.apache.org/dist/spark/spark-3.1.2/spark-3.1.2-bin-hadoop3.2.tgz.sha512

podman run -p 8888:8888 -p 4040:4040 --user root -e JUPYTER\_ENABLE\_LAB=yes --name pyspark jupyter/pyspark-notebook:spark-3.2.0

podman run -p 8888:8888 -p 4040:4040 --user root -e JUPYTER\_ENABLE\_LAB=yes --name pyspark my-pyspark-notebook

podman run -it --rm -p 8888:8888 jupyter/pyspark-notebook:spark-3.1.2

http://127.0.0.1:8888/lab?token=69c0a35afc2771fa1bd83a3fe409bd5e02681a6a4b462c5a

A screen shot of a computer

Description automatically generated

1. Open notebook

A screenshot of a computer

Description automatically generated

1. Or also from cdm

The next step find Id of container:  
> podman ps

A screenshot of a computer

Description automatically generated

1. Upload Deequ\_puSpark\_skeleton and run command from Deequ\_pySpark\_skeleton\_Vial.ipynbp pip install pydeequ

Pip install --upgrade python-dateutil

A screenshot of a computer program

Description automatically generated

install pydeequ (in git bush)

podman exec -it eb28ed05474b bash -c 'pip install pydeequ'

1. After instalating restart kernel

podman cp mssql-jdbc-12.6.1.jre11.jar eb28ed05474b:/home/jovyan/work

podman machine ssh sh -c 'cat > //home/jovyan/work/ mssql-jdbc-12.6.1.jre11.jar < mssql-jdbc-12.6.1.jre11.jar