Sergiu Buhatel

BlazingSQL on NVIDIA DGX A100

Table of Contents

[1. Introduction 1](#_Toc192239332)

[1 Installation 1](#_Toc192239333)

[2 Run 2](#_Toc192239334)

[2.1 Launch 2](#_Toc192239335)

[2.2 Stop 2](#_Toc192239336)

[2.3 Copy files 3](#_Toc192239337)

[2.4 Copy files 3](#_Toc192239338)

[2.5 GPUs usage 4](#_Toc192239339)

[3 References 4](#_Toc192239340)

# Introduction

BlazingSQL is a SQL database that leverages the usage of GPUs instead of the CPUs.

This document describes the minimal steps to install and run BlazingSQL on NVIDIA DGX A100 machine.

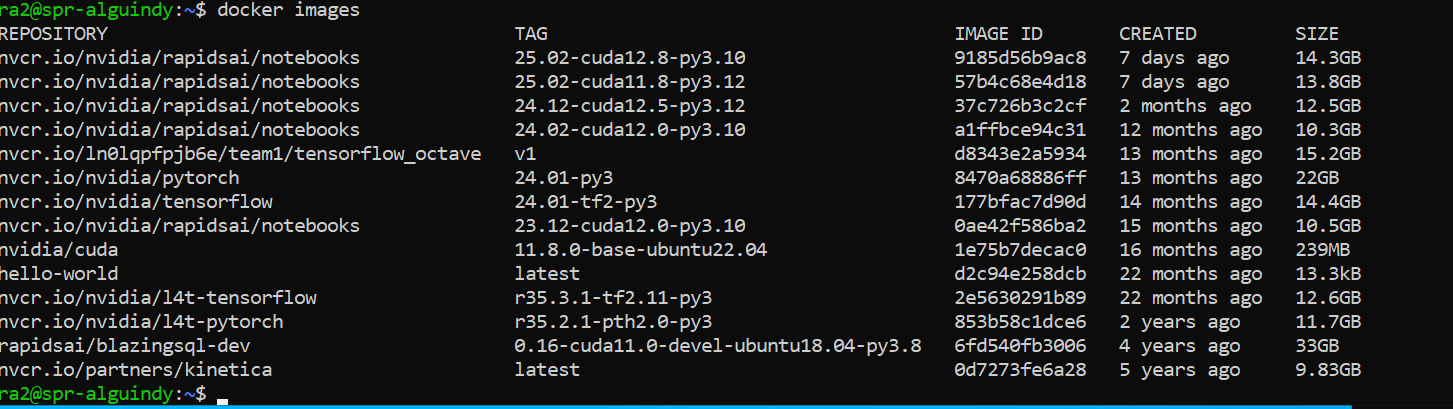
# Installation

Pull the image to the machine

docker pull rapidsai/blazingsql-dev:0.16-cuda11.0-devel-ubuntu18.04-py3.8

Check if the image was already pulled

docker images



# Run

## Launch

docker run --gpus all -v /home/ra2:/mnt/data -it --rm rapidsai/blazingsql-dev:0.16-cuda11.0-devel-ubuntu18.04-py3.8 bash

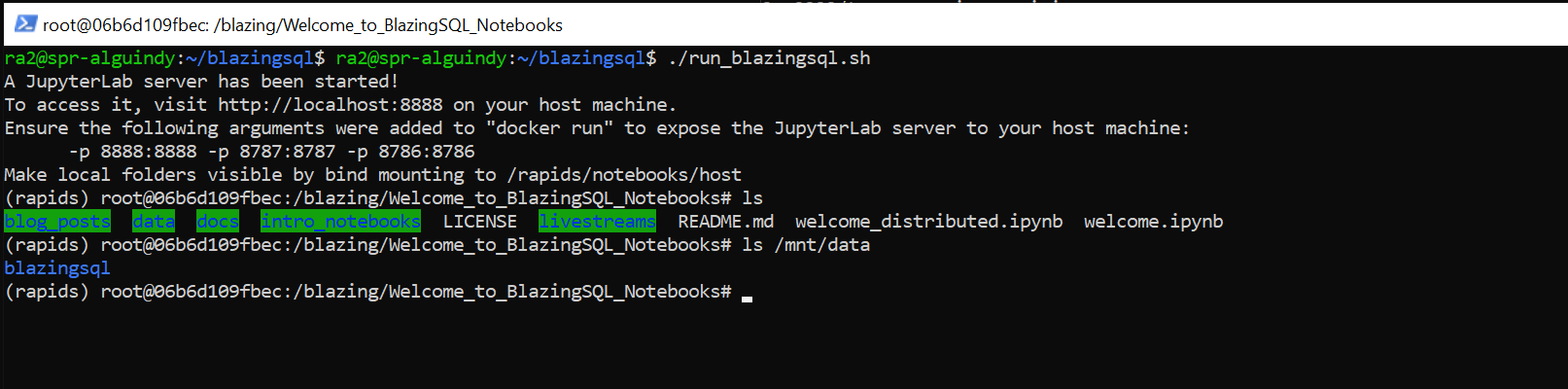
/home/ra2 is a directory on the machine, and /mnt/data is a mount inside the running container.

The above command line can be store in a .sh file, for easy usage. For example, is stored inside /home/ra2/blazingsql/run\_blazingsql.sh

Can be launched

cd /home/ra2/blazingsql

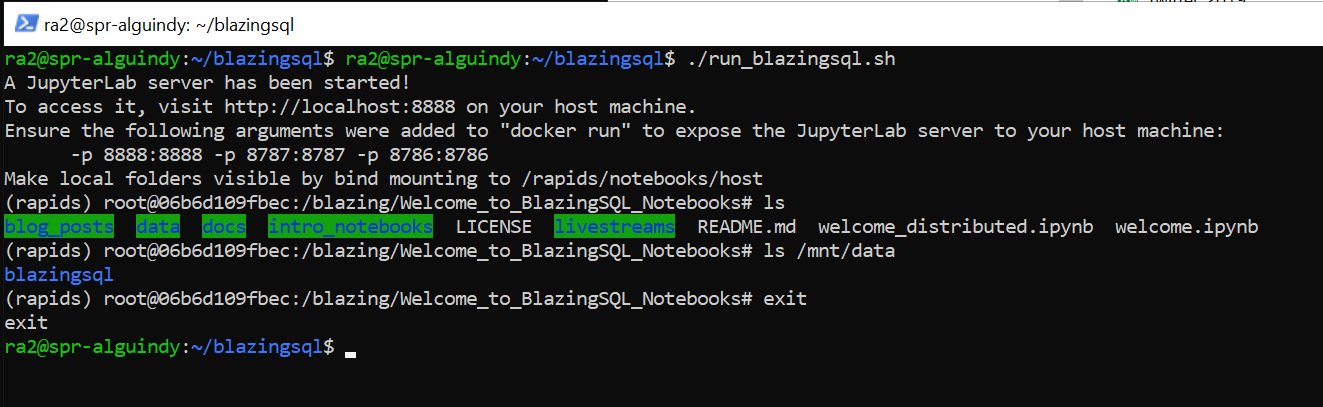
./run\_blazingsql.sh



As we can see, /mnt/data inside the running container, is mirroring /home/ra2 on the machine.

## Stop

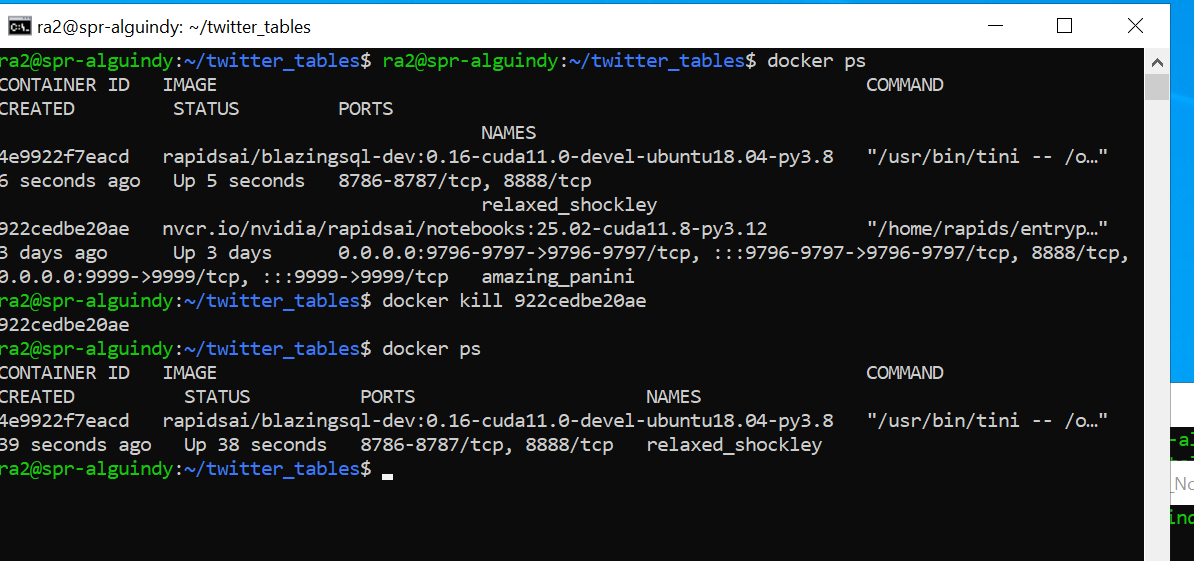
To exit the running container, type exit.



Another way is to find the running container in a different command shell and stop it explicitly.

docker ps

docker kill <container id>

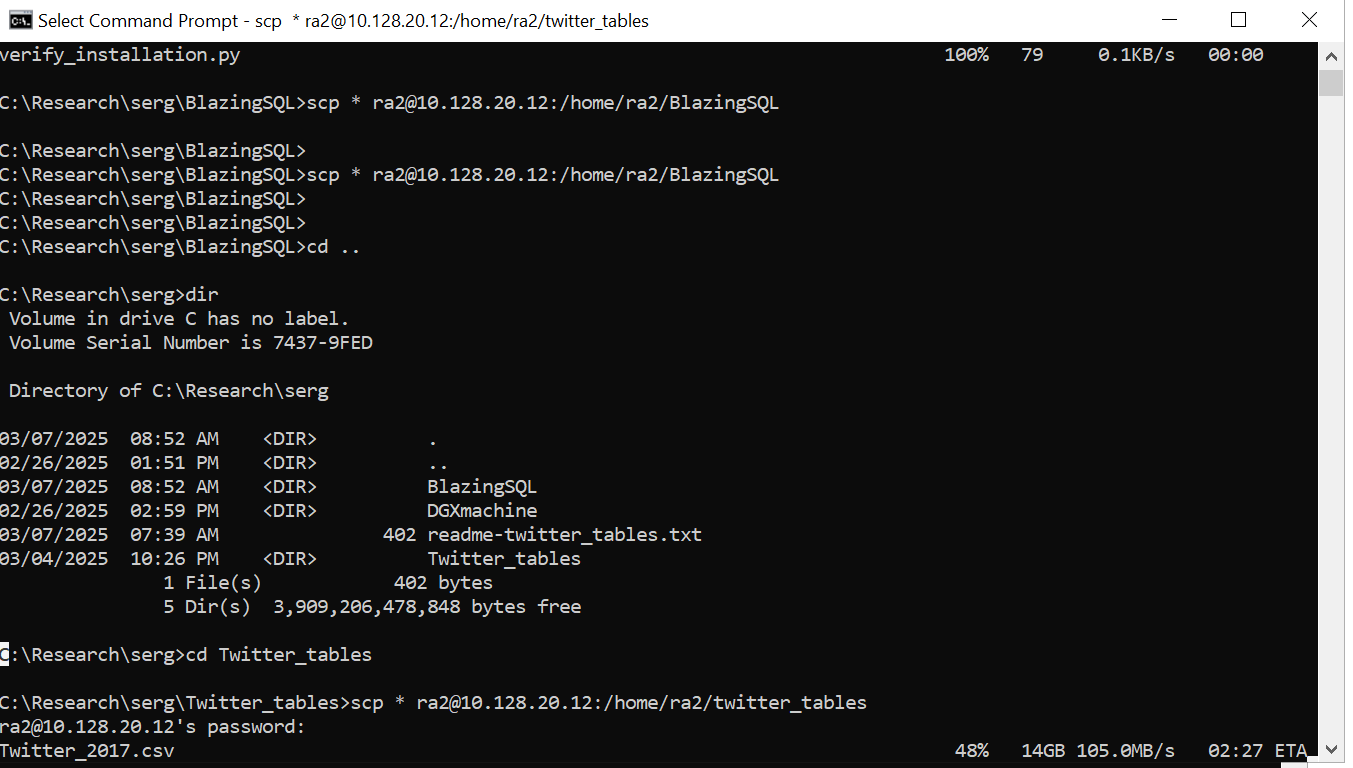


## Copy files

To copy files from Windows machine to Nvidia DGX machine, open a command prompt on windows and change directory to the directory containing the files you want to copy.

cd C:\Research\serg\Twitter\_tables

scp \* ra2@10.128.20.12:/home/ra2/twitter\_tables



## GPUs usage

To continuously check GPUs usage, open a shell command and type:

watch -n 1 nvidia-smi

Press Ctrl/C to exit.

## Import data

from blazingsql import BlazingContext

import cudf # RAPIDS cuDF for GPU-accelerated dataframes

bc = BlazingContext()

bc.create\_table('Twitter\_2022\_table', '/mnt/data/twitter\_tables/Twitter\_2022.csv')

result = bc.sql("SELECT \* FROM Twitter\_2022\_table LIMIT 5")

print(result)

# References

<https://hub.docker.com/r/rapidsai/blazingsql-dev>

<https://towardsdatascience.com/intro-to-blazingsql-in-5-minutes-172162d3a95/>

<https://developer.nvidia.com/blog/beginners-guide-to-querying-data-using-sql-on-gpus-in-python/>