A Manual on Installation of Tensorflow Serving in Ubuntu Docker

Prarthana Bataju

2018.02.09

1.Install Docker CE

Install using the repository

Before you install Docker CE for the first time on a new host machine, you need to set up the Docker repository. Afterward, you can install and update Docker from the repository.

SET UP THE REPOSITORY AND INSTALL

1. Update the `apt` package index:

```
$ sudo apt-get update
```

2. Install packages to allow `apt` to use a repository over HTTPS:

```
$ sudo apt-get install
    apt-transport-https ¥
    ca-certificates ¥
    curl ¥
    software-properties-common
```

3. Add Docker's official GPG key:

```
$ curl -fsSL https://download.docker.com/linux/ubuntu/gpg
| sudo apt-key add -
```

Verify that you now have the key with the fingerprint `9DC8 5822 9FC7 DD38 854A E2D8 8D81 803C 0EBF CD88`, by searching for the last 8 characters of the fingerprint.

```
$ sudo apt-key fingerprint 0EBFCD88
```

4. Use the following command to set up the **stable** repository.

```
$ sudo add-apt-repository ¥
   "deb [arch=amd64]
https://download.docker.com/linux/ubuntu ¥
   $(lsb_release -cs) ¥
   stable"
```

5. Update the apt package index.

```
$ sudo apt-get update
```

6. Install the latest version of Docker CE, or go to the next step to install a specific version. Any existing installation of Docker is replaced.

```
$ sudo apt-get install docker-ce
```

7. Add docker to the admin group(not compulsary)
    - Add docker group if it does not already exist
    - Add the connected user $USER to the docker group

```
$ sudo groupadd docker
$ sudo usermod -aG docker $USER
```

2. Create Docker image and run container for TensorFlow Serving

GET THE TENSORFLOW SERVING

```
$ cd ~
$ git clone --recurse-submodules
https://github.com/tensorflow/serving.git
```

BUILD DOCKER IMAGE AND RUN CONTAINER

The configuration of the Docker image is defined via a Dockerfile. TensorFlow Serving provides two of them—one for CPU build and one for GPU build. You can find both under *serving/tensorflow_serving/tools/docker*:

- *Dockerfile.devel* for CPU build
- *Dockerfile.devel-gpu* for GPU build

MODIFICATION OF GPU DOCKERFILE

I could not run the GPU Dockerfile out-of-the-box and modified it slightly. I commented out the execution of Bazel build, so the last lines of *Dockerfile.devel-gpu* look like that:

```
# Build TensorFlow Serving and Install it in
/usr/local/bin
WORKDIR /serving
#RUN bazel build -c opt – config=cuda ¥
# --crosstool_top=@local_config_cuda//crosstool:toolchain
¥
# tensorflow_serving/model_servers:tensorflow_model_server
&& ¥
# cp bazel-
bin/tensorflow_serving/model_servers/tensorflow_model_serv
er /usr/local/bin/ && ¥
# bazel clean --expunge
CMD [“/bin/bash”]
```

Hint: since I used GPU Dockerfile, I refer to it in the next steps. With CPU Dockerfile you would work in a same manner.

Let’s create our Docker image:

```
$ cd serving
$ docker build --pull -t $USER/tensorflow-serving-devel-
gpu -f tensorflow_serving/tools/docker/Dockerfile.devel-
gpu .
```

It takes 10–20 minutes to download dependencies and build the image.

Now we can run Docker container:

```
$ docker run --name=tf_container_gpu -it $USER/tensorflow-
serving-devel-gpu
```

If everything works as expected, congratulations! You are now in the Shell of TensorFlow Serving Docker container.

Hint: when you leave container Shell with *exit* command, it shuts down. To start it again execute:

```
$ docker start -i tf_container_gpu
```

Do not remove the container! Otherwise all changes you made will gone.

In Docker container all necessary components such as Python, Bazel, etc. are already installed. That is a power of containerization—we define once, what we need and then we get it every time automatically.

3.Build TensorFlow Serving in Docker container and deploy the model

In created Docker container we will build and start Tensorflow Serving.

GET THE SOURCES (ONLY FOR CPU DOCKERFILE!)

If you use Dockerfile for TensorFlow CPU build, then you need to clone TensorFlow Serving. From the container Shell get the sources (same operation as we done on our PC):

```
$ git clone --recurse-submodules
https://github.com/tensorflow/serving.git
```

You do not need it if you use GPU Dockerfile as I did. The cloning of a repository is already done during the Docker image build.

BUILD TENSORFLOW SERVING

The next step—build all we need:

```
$ cd /serving/tensorflow
$ ./configure
```

Here we configure our build. You will be asked a couple of question, just accept all defaults. Start the build process…

```
$ cd ..
$ bazel build -c opt tensorflow_serving/...
```

… and wait 30–40 minutes. When the process completes, you should be able to execute

```
$ bazel-bin/tensorflow_serving/model_servers/tensorflow_model_server
```

and see a usage information.

DEPLOY THE MODEL

Next, you copy the exported model (see Part 1) into TensorFlow container. From Shell on you **PC**:

```
$ cd <path to project>
$ docker cp ./<model_directory> <containerId>:/serving
```

You should have the exported model folder in the **container**. From its Shell:

```
$ cd /serving
$ ls ./<model_directory>/1
```

You should get *variables* folder and *saved_model.pb* file.

START SERVING

Hurrah! We are ready to host our model by TensorFlow :-)

```
$ bazel-
bin/tensorflow_serving/model_servers/tensorflow_model_serv
er --port=9000 --model_name=inception --
model_base_path=/serving/<model_directory> &>
<log_file_name> &
```

If you check the log…

```
$ cat gan_log
```

… You should see something like this in the last string:I

```
“tensorflow_serving/model_servers/main.cc:298] Running
ModelServer at 0.0.0.0:9000”
```

CHECK GRPC NETWORK

```
$ sudo docker network inspect bridge | grep IPv4Address
```

CALL PREDICTION FROM CLIENT

```
$ cd <path to your project>
$ python client.py --server= 12.17.0.2:9000 --image=<path
to your image>
```

REFERENCES

1. https://towardsdatascience.com/how-to-deploy-machine-learning-models-with-tensorflow-part-2-containerize-it-db0ad7ca35a7
2. https://github.com/tensorflow/tensorflow/pull/15855
3. https://github.com/tensorflow/serving/issues/449