**Installing Docker**

For installing Docker and further using it with our containers for this course, we should have the following requirements.

The minimal requirements for running notebooks on our computer are:

2-core processor

4 GB of RAM

20 GB of free disk space

Installing Docker on Unix

To install Docker on Ubuntu, we should run the following command with sudo privileges:

curl -sSL https://get.docker.com/ | sh

Then optionally we can add our user account to the docker group:

**usermod -aG docker $USER**

This allows us to deploy containers without sudo but we don’t recommend doing this on CentOS for security reasons.

Checking the installation

Now run the test container to check the installation successful:

**docker run hello-world**

We get some Docker deployment logs on the screen and then “Hello from Docker!” message from a successfully deployed container.

Deploying containers

We can find all of them in our Docker Hub repository. Appropriate <container\_name> is specified at the bottom of self-reading to each assignment. E.g. you can see “If you want to deploy the environment on your own machine, please use https://hub.docker.com/r/bigdatateam/all-spark Docker container.”

To run a Juputer notebook on our machine we should do the following steps.

1. Pull the actual version of the container from DockerHub

**docker pull <container\_name>**

2. Deploy the container

**docker run --rm -it -p <your\_port>:8888 <container\_name>**

3. Open Jupyter environment

Now Jupyter can be opened in a browser by typing localhost:<your\_port>. If we were working on Windows, we should open <container\_host>:<your\_port> instead of localhost. We can find <container\_host> in the deployment logs of the container.

So now we have a full environment with installed and adjusted Hadoop services, prepared datasets and demo codes on your own machine.

Working with docker through a terminal

To start the container and work with it via Unix terminal you should do the following steps.

1. To start a Tmux session:

**tmux new -s my\_docker**

2. To run our container in terminal opened:

**docker run --rm -it -p 8888:8888 -p 50070:50070 -p 8088:8088 bigdatateam/hdfs-notebook**

(please, take into account that you can forward some additional ports with Hadoop UIs, not only Jupyter's port).

3. Detach the Tmux session by typing "Ctrl+b" and then "d".

4. Check your container's id: docker ps

We can get our container’s id as output of above command

5. Finally, we can open the terminal within the container by executing:

**docker exec -it <OUR\_CONTAINER\_ID> /bin/bash**

6. Now we've logged in the container as root and we can execute Hadoop shell commands.

root@da1d48ac25fc:~# hdfs dfs -ls /data

Found 1 items

**drwxrwxrwx - jovyan supergroup 0 2017-10-15 16:30 /data/wiki**

root@da1d48ac25fc:~#

**Simultaneously, we can work with Jupyter via browser.**

To stop the docker container we should do the following steps.

1. Attach the Tmux session -- tmux a -t my\_docker

2. To get the name of the Tmux session --- tmux ls

2. Stop the container via Ctrl+C.

3. Exit from the Tmux session.