BEStack 01A Installing External on Linux
Tod Casasent
2018-03-03-1100

1 Introduction

These instructions are aimed at people familiar with Linux administration and GitHub, Docker and Docker-Compose, with root access to their Linux environment. Installation is for external users.

2 Target Operating System and Installation

These instructions were tested on Debian 9.1 installed on Oracle VirtualBox. These instructions with appropriate modifications should work as a basis for installing MBatch on other distributions.

Perform the installs in the order given in this document. Some steps are dependent on previous steps.

3 Operating System Prerequisites

Docker and Docker-Compose are required installs.

Docker versions used in development and testing were:

Version: 17.12.0-ce or 17.06.0-ce

API version: 1.35 or 1.30

Go version: go1.9.2 or go1.8.3

Docker-Compose versions used in development and testing were:

docker-compose version 1.16.1 or 1.14.0

docker-py version: 2.5.1 or 2.3.0

CPython version: 2.7.13

4 Clone the GitHub Repository

Clone the GitHub repository with a shallow clone, since you will not be checking anything back in. This clone call grabs the newest version from master.

```
git clone --depth 1 https://github.com/MD-Anderson-Bioinformatics/BatchEffectsInterfaceStack.git
```

5 Setup Local Information

5.1 Docker-Compose Paths and Files

Since Docker Containers do not save internal modifications between runs, the Batch Effects Interface Stack expects external locations for three directories. At this time, for these instructions, these paths need to be on a Linux system.

The properties files for BEI are stored by default in an external directory /data/PROPS.

The output (job information) files are stored by default in an external directory /data/OUTPUT.

The website information (used to display results to users) are stored by default in an external directory /data/batcheffects.

All three directory paths can be changed by editing setupBEI_linux.bash and putting the new paths into the DEST_PROPS, OUTPUT_DIR, and WEBSITE_DIR variables at the top of the file.

```
# directory for properties files
DEST_PROPS=/data/PROPS

# directory for output from MBatch
OUTPUT_DIR=/data/OUTPUT

# directory for links to output to be displayed by Batch Effects Website
WEBSITE_DIR=/data/batcheffects
```

These paths will be used in the docker-compose.yml file.

5.2 Entries for gdc.properties

Since machines have different amounts of memory, we provide a configurable memory setting for downloading GDC data. This settings is also in the setupBEI_linux.bash file. The MEMSIZE variable should be 64GB to download any currently supported/known data set. (Data sets may increase in size over time.)

Recommended MEMSIZE settings are from 10G to 20G or 64G.

```
# memory size to use for running GDC downloads
# 10-20G works for smaller data sets. Larger datasets may use up to 64G
MEMSIZE=20G
```

The MEMSIZE entry is automatically updated in the gdc.properties when the setup script is run.

5.3 Manual Entries for bei.properties

Similary, the bei.properties file can be manually configured if needed. The provided file uses the settings for Development Values.

Key	Description	Development Value	Production Value
maxDscThreads	The number of threads to be used calculating DSC values. This requires memory and CPU time.	5	10
maxBoxplotGeneCount	The maximum number of genes to use in the boxplot. We strongly recommend a maximum of 5000 for most genomic data sets.	2500	5000
minBoxplotGeneCount	The minimum number of genes to use in the boxplot. We strongly recommend using the maximum value, unless you are simply testing the system.	500	1000
maxDscGeneCount	The number of genes used for calculating the DSC. We strongly recommend using at least a maximum of 10,000 for production systems.	5000	10000
maxDscPermutations	The number of permutations used for calculating the DSC p-value. We strongly recommend using 2000.	2000	2000
filterMaxAllowed	The maximum number of cells (rows times columns) in a matrix used for processing Batch Effects. This is useful for systems limited by memory.	4000000	8000000
filterMinAllowed	The minimum number of cells (rows times columns) in a matrix used for processing Batch Effects. This is useful for systems limited by memory. NOTE: A value of 0 means "use all data".	1000000	0

6 Run Setup Script

Run setupBEI_linux.bash to populate the local directories and setup the docker-compose.yml file. Please note, the local directories are set to 777 (chmod -R 777) because the processes in the docker container run using UID 2222. Since the processes create, edit, and delete files and directories, they need full access to the external directories.

```
chmod u+x setupBEI_linux.bash
./setupBEI_linux.bash
```

There may be some warnings about chmoding the /data directory, depending on your configuration.

7 Starting and Stopping the Docker Compose Stack

In the directory with the docker-compose.yml file, the containers (stack) are started using:

```
docker-compose -p EXT -f docker-compose.yml up -d
```

The EXT may be varied if needed on your system to ensure unique ids for the stack.

The Docker Compose Stack can be stopped using:

```
docker-compose -p EXT -f docker-compose.yml down
```

8 Debugging and File Maintenance

To connect to the different containers as root, using the following:

```
docker exec -t -i -u 0 bei_cont_outside /bin/bash
docker exec -t -i -u 0 gdc_cont_outside /bin/bash
docker exec -t -i -u 0 bei_cont_outside /bin/bash
```

To connect as the default user (2222 docker_tcga), use:

```
docker exec -t -i bei_cont_outside /bin/bash
docker exec -t -i gdc_cont_outside /bin/bash
docker exec -t -i bei_cont_outside /bin/bash
```

The various tomcat logs are in: /opt/tomcat/logs/. A future version will place these logs outside the container.

Another useful command is to use the container root to set access to directories and files created by the container to 777.

```
cd /BEI
chmod -R 777 *
```

9 Testing the Docker Compose Stack and URL

The URL for your install will be something like:

http://your-server.your-company.com:9999/BatchEffectsInterface/

To test the Docker Compose Stack, see the BEStack 02A Using Batch Effects Interface Assessments document and follow the Downloading GDC Data, Configuring Assessments, and running the job sections.