EMBnet.journal
Bioinformatics in Action

Mattila et al. (2017) EMBnet.journal 22, e879
http://dx.doi.org/10.14806/ej.22.0.879

Technical Notes

Set up your own bioinformatics server: Chipster in EGI Federated Cloud

Kimmo Mattila[1], Diego Scardaci[2], Marica Antonacci[3], Catalin Condurache[4]

[1]CSC - IT Center For Science, Espoo, Finland; [2]EGI Foundation; [3]INFN, Bari, Italy; [4]Rutherford Appleton Laboratory, Oxfordshire, United Kingdom.

Abstract

Chipster is an easy to use data analysis platform for bioinformatics. It provides an uniform graphical interface for over 360 commonly used bioinformatics tools including several R/Bioconductor-based tools and standalone programs (*e.g.* BWA, TopHat). Chipster is based on a client-server system where the user runs locally a Chipster-client that submits analysis tasks to a Chipster server. Even though Chipster is an open source tool, there is no public Chipster server that would be open for everybody. Due to that, a researcher needs to have an access to some of the existing Chipster servers to be able to use this platform. Alternatively, a researcher can set up your his own Chipster server. In this paper, we describe how a Chipster server can be launched EGI Federated Could environment, that provides resources for all European researchers. With the instructions provided here, any European researcher can launch and manage his own Chipster server, suited for needs of a small research group or a bioinformatics course. The setup described here is based on a collaboration of several European instances. Chipster[1] is developed by CSC – IT Center for Science Ltd. in Finland. European Grid Infrastructure (EGI) has fitted Chipster to cloud environment[2] and provides the cloud computing resources. Finally, Rutherford Appleton Laboratory hosts the CVMFS server[3] that provides the scientific tools and data sets for the Chipster servers running in EGI federated cloud.

1. Preparatory steps

The EGI Federated cloud environment can be used from Linux or Mac OSX machines. In order to launch a Chipster server in EGI Federated Cloud, the machine that is used to manage the Chipster server must have the following tools and files installed: a valid personal X.509 certificate; rOCCI command line client for managing cloud computing environment; `voms-proxy-init` command to create proxy certificates; settings files to connect the VOMS server hosting chipster.csc.fi VO.

In addition, the manager of the Chipster server must join the chipster.csc.fi Virtual Organisation. The manager needs to do these preparatory steps only once. After that, Chispster servers can be managed with the FedCloud_Chipster_manager tool. Note that the end-users who wish just to use the Chipster server running in EGI Federated Cloud, do not need to do any of these preparatory steps.

Programme

EGI Federated Cloud uses X.509 certificates for user authentication. Researchers from the member countries of GÉANT network[4] can use the DigiCert certificate service[5] to obtain a personal grid certificate. Users from other countries should use their local certification authorities. Once you have a grid certificate installed in your browser, you can join the chipster.csc.fi Virtual Organisation(VO) in the VO home page[6].

Article history
Received: 18 January 2017
Published: 25 January 2017

1.2. Installing rOCCI and VOMS client

The management of Federated Cloud resources is done using rOCCI, a ruby based implementation of OCCI standard. The authentication in EGI federated cloud is done using proxy certificates generated with command voms-proxy-init. The instructions to install these tools to your local machine can be found from EGI wiki site[7]. Once you have installed the rOCCI and `voms-proxy-init` commands, you must still define the connection to Chipster.csc.fi VO management server (VOMS). To do this, first create directory "/etc/grid-security/vomsdir/chipster.csc.fi" and go to this directory:

```
mkdir /etc/grid-security/vomsdir/
chipster.csc.fi
cd /etc/grid-security/vomsdir/chipster.
csc.fi
```

Then create a file "voms.fgi.csc.fi.lsc" that contains the following 2 lines:

```
/O=Grid/O=NorduGrid/CN=host/voms.fgi.
csc.fi
/O=Grid/O=NorduGrid/CN=NorduGrid
Certification Authority
```

[1] http://chipster.csc.fi
[2] https://www.egi.eu/federation/egi-federated-cloud
[3] https://www.gridpp.ac.uk/wiki/RAL_Tier1_CVMFS
[4] http://www.geant.org
[5] https://www.digicert.com/sso
[6] https://voms.fgi.csc.fi:8443/voms/chipster.csc.fi
[7] https://wiki.egi.eu/wiki/Fedcloud-tf:CLI_Environment

© 2017 Mattila *et al.*; the authors have retained copyright and granted the Journal right of first publication; the work has been simultaneously released under a Creative Commons Attribution Licence, which allows others to share the work, while acknowledging the original authorship and initial publication in this Journal. The full licence notice is available at http://journal.embnet.org.

If you already have file "/etc/vomses", move the file "/etc/vomses" to "/etc/vomses/old_vomses" (vomses will be a directory now). Create a file "chipster.csc.fi-voms.fgi.csc.fi" in "/etc/vomses" and write inside the following line:

```
"chipster.csc.fi"  "voms.fgi.csc.fi"  "15010"  "/O=Grid/O=NorduGrid/CN=host/voms.fgi.csc.fi"  "chipster.csc.fi"
```

1.3 Obtaining keys and FedCloud_chipster_manager

FedCloud_chipster_manager is a help tool that can be used to manage Chipster instances to EGI Federated Cloud. It can be downloaded from the Chipster git-hub[8]. Some of the FedCloud_chipster_manager operations require that user provides encryption key pair that is used to access the virtual machine. The key pair can be created for example with command:

```
ssh-keygen -t rsa -b 2048 -f FedCloudKey
```

2. Managing Chipster server2.1 Setting up VOMS proxy

Before launching or managing virtual Chipster servers, you have to create a temporary proxy certificate that is used to authenticate to EGI Federated Cloud environment. If you have the voms-proxy-init command installed and a valid X.509 certificate in your ".globus" directory, you can create a temporary proxy certificate with command:

```
voms-proxy-init --voms chipster.csc.fi --rfc --dont_verify_ac
```

The command above asks the password of your certificate and creates a proxy certificate that is valid for 12 hours. Note that voms-proxy-init requires that you are using OpenJDK-based Java environment. Other Java environments cause error messages like: "Credentials couldn't be loaded".

2.2 Launching a Chipster server

Once you have done all the preparations, you can launch a new Chipster Virtual Server with command (assuming you have the FedCloud_chipster_manager tool in your current working directory):

```
./FedCloud_chipster_manager -key keyfile -launch
```

This launching command uses default values, for resources and user accounts linked to the Chipster. Option `-volume_size` can be added to modify the size of the data volume (in gigabytes) that is used to store the data during the computing. The default size of the volume is only 20 GB, which is enough for testing, but for real usage a bigger data volume may be needed. By default, only one Chipster account (user: chipster, password: chipster) is created to a new Chipster server. A list of user accounts for a new Chipster server can be defined with option -users. The argument for this option should be a file containing a list of accounts in format:

```
user_name:password:expiration_date
```

The expiration date is defined with format: yyyy-mm-dd. For example file "accounts.txt" could look like following:

```
trng1:4eoU8hmx:2017-05-15
trng2:4eoU8hmx:2017-05-15
```

Note that these accounts are just Chipster server accounts, not Linux accounts that could be used to open terminal connections to the virtual machine. Launching a Chipster server with these accounts and 100 GB storage size could be done with command:

```
./FedCloud_chipster_manager -launch -key FedCloudKey -volume_size 100 -users accounts.txt
```

The launching process can take tens of minutes. In the end the launching process prints out information about how the server can be accessed. For example:

```
----------------------------------------
You can now connect your virtual machine with command:

ssh -i FedCloudkey ubuntu@90.147.102.41

The Chipster server can be connected with URL: http://90.147.102.41:8081
```

The users can now use the URL to use the Chipster server while the ssh connection is intended for managing the Chipster server. Note that each Chipster server will get a unique IP address. The address is assigned by the cloud environment and it can't be set or modified by the user.

2.3 Other management tasks

In addition to launching Chipster servers, FedCloud_chipster_manager tool can be used to manage an existing server. You can use FeCloud_chipster_manager with option `-list`, to list your virtual Chipster servers running in the EGI Federated Cloud. Option: `-status` makes FedCloud_chipster_manager to look for Chipster VMs launched by the user, and to check the status of the Chipster server running in the VMs found. In this case, you must also use the `-key` option to define the

[8] https://raw.githubusercontent.com/chipster/chipster/master/src/main/admin/fedcloud/FedCloud_chipster_manager

Mattila et al. (2017) EMBnet.journal 22, e879
http://dx.doi.org/10.14806/ej.22.0.879

key file, that was used to launch the server. The password for the key file is asked for each server to be connected. The option `-restart` makes FedCloud_chipster_manager restart the Chipster server running in the given Federated Cloud VM instance. This option can be used for example to fix the server if the Chipster server is using internal IP address instead of public IP address. For example, restarting the Chipster server running in instance `/compute/86b97ed5-e256-4bce-83b5-aa3a41920975` can be done with command:

```
./FedCloud_chipster_manager -key
FedCloudKey -restart ⧵
/compute/86b97ed5-e256-4bce-83b5-
aa3a41920975
```

To completely delete the virtual machine running in EGI Federated Cloud you can use option `-delete`:

```
./FedCloud_chipster_manager -delete
instance-ID
```

For more detailed management, you can open a terminal connection to the virtual machine and apply the instructions in the Chipster technical manual[9].

[9] https://github.com/chipster/chipster/wiki/TechnicalManual

3. Using your Chipster server in EGI Federated Cloud

The Chipster Virtual organisation can provide only limited resources for the Chipster user community. By default, the FedCloud_chipster_manager starts a Chipster server on a virtual machine that has 4 computing cores with a total of 8 gigabytes of memory. This is not much, but it should be enough to serve the needs of a small research group (only a few simultaneous users). If you wish to use a larger virtual machine, please contact the Chipster VO manager. Once launched, the server can be kept up and running as long as the data processing continues. This can be weeks or months, but finally the Chipster server should be shut down by the owner of the server.

If your server has been running longer than 4 months, the VO manager can ask the owner of Chipster server to send a report about the usage of the server. When using the Chipster in EGI Federated Cloud, you should remember that the intermediate data at the servers is not back-upped. If you need to rebuild your Chipster server, the data in the previous version will be lost when the old version is removed. Further, you should remember that current setup for running Chipster in EGI Federated Cloud is still under testing and development. We do not guarantee uninterrupted access to the resources at all times.

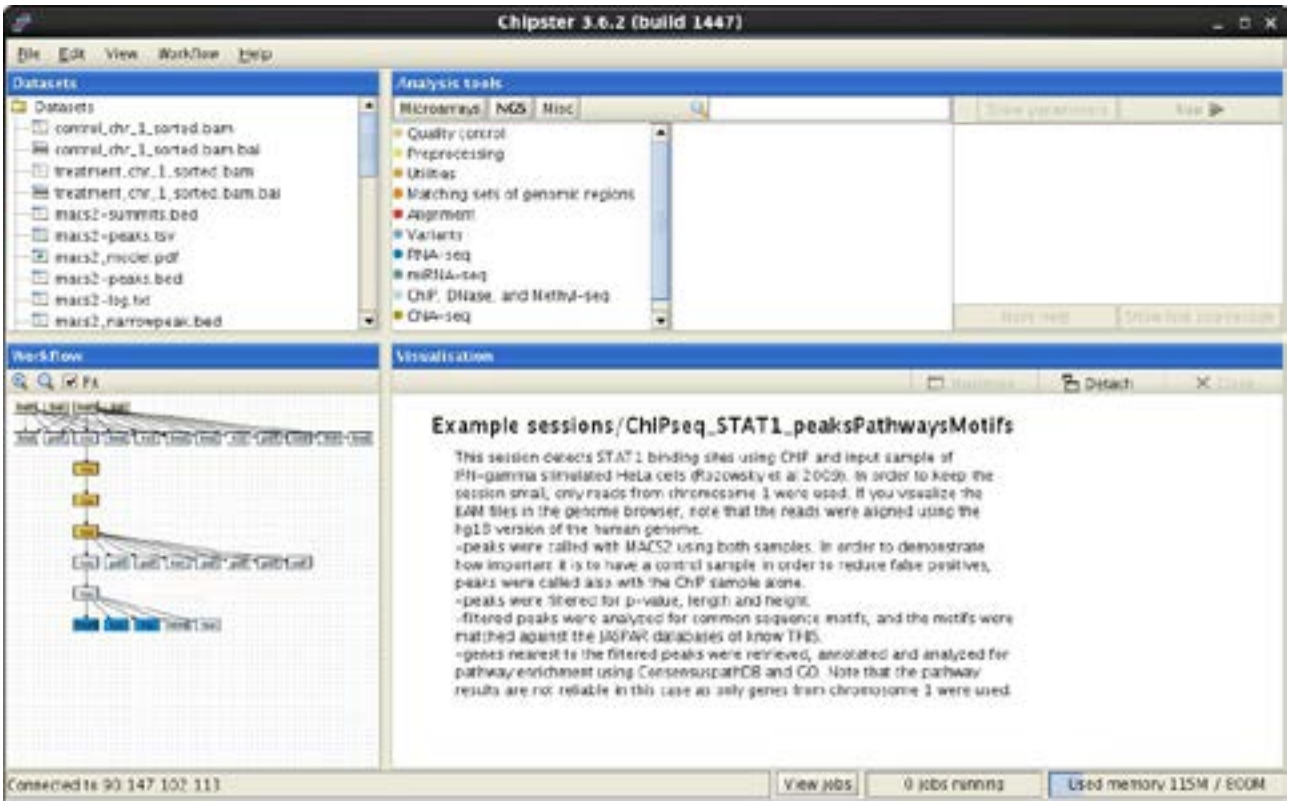

Figure 1. Chipster client started from a Chipster server running in EGI Federated Cloud.