

Hướng dẫn cài đặt và sử dụng Sahara

[1. Cài đặt sahara](#)

[1.1. Cài đặt](#)

[1.2. Cấu hình](#)

[2. Cài đặt UI Sahara](#)

[3. Hướng dẫn sử dụng](#)

[4. Một số lỗi hay gặp phải](#)

1. Cài đặt sahara

1.1. Cài đặt

Tham khảo từ trang chủ http://docs.openstack.org/developer/sahara/userdoc/installation_guide.html

```
sudo apt-get install python-setuptools python-virtualenv python-dev
```

```
virtualenv sahara-venv
```

```
sahara-venv/bin/pip install sahara
```

```
# Hoặc có thể sử dụng
```

```
# sahara-venv/bin/pip install 'http://tarballs.openstack.org/sahara/sahara-master.tar.gz'
```

```
# Bạn có thể vào http://tarballs.openstack.org/sahara/ chọn phiên bản khác
```

```
mkdir sahara-venv/etc
```

```
# Có 2 file config cơ bản sahara.conf.sample-basic và sahara.conf.sample. Khuyến nghị nên dùng  
# file sample-basic.
```

```
cp sahara-venv/share/sahara/sahara.conf.sample-basic sahara-venv/etc/sahara.conf
```

```
# Cấu hình (xem mục 1.2)
```

```
# Nếu bạn sử dụng mysql, sửa file my.sql:
```

```
...  
[mysqld]  
...  
max_allowed_packet = 256M
```

```
# Tạo database schema
```

```
sahara-venv/bin/sahara-db-manage --config-file sahara-venv/etc/sahara.conf upgrade head
```

```
# Khởi động Sahara:
```

```
sahara-venv/bin/sahara-all --config-file sahara-venv/etc/sahara.conf
```

1.2. Cấu hình

Dưới đây là file cấu hình mà chúng tôi đã thiết lập. Bạn có thể thay đổi, tham khảo tại:

http://docs.openstack.org/developer/sahara/userdoc/configuration_guide.html

```

nano sahara.conf
[DEFAULT]
os_auth_host=127.0.0.1
os_auth_port=35357
os_admin_username=admin
os_admin_password=$pass_admin
os_admin_tenant_name=admin

use_floating_ips=true
use_neutron=true
use_namespaces=false

log_dir=/var/log/sahara
log_file=sahara.log

[database]
connection=sqlite:///tmp/sahara.db
# Nếu bạn muốn sử dụng MySQL, thì thay thế bằng dòng dưới đây. Nhớ là phải tạo
# database sahara và gán quyền truy cập cho nó trong MySQL nhé.
# connection=mysql://sahara:pass_saharadb@127.0.0.1/sahara

[keystone_authtoken]
auth_uri=http://127.0.0.1:5000/v2.0/
identity_uri=http://127.0.0.1:35357/

```

Có thể thiết lập thêm các thông số verbose và debug.

2. Cài đặt UI Sahara

Tham khảo từ trang chủ http://docs.openstack.org/developer/sahara/horizon/installation_guide.html

```
sudo pip install sahara-dashboard
```

Nếu bạn muốn cài bản development của Sahara UI hãy chạy:

```
# sudo pip install http://tarballs.openstack.org/sahara-dashboard/sahara-dashboard-master.tar.gz
```

Thêm 'sahara' vào phần HORIZON_CONFIG của file settings.py

```
nano /usr/share/openstack-dashboard/openstack_dashboard/settings.py
```

```
HORIZON_CONFIG = {
```

```
'dashboards': ('nova', 'syspanel', 'settings', ..., 'sahara'),
```

```
INSTALLED_APPS = (  
    'saharadashboard',  
    ....
```

Sửa file /usr/share/openstack-dashboard/openstack_dashboard/local/local_settings.py

Thêm vào file 3 dòng sau:

```
SAHARA_USE_NEUTRON = True  
AUTO_ASSIGNMENT_ENABLED = False  
SAHARA_URL = 'http://localhost:8386/v1.1'
```

Bạn cũng có thể bỏ dòng SAHARA_URL = 'http://localhost:8386/v1.1'. Thay vào đó chạy:

`keystone service-create --name sahara --type data_processing --description "Sahara Data Processing"`

```
keystone endpoint-create --service sahara --region RegionOne \  
--publicurl "http://10.0.0.2:8386/v1.1/%(tenant_id)s" \  
--adminurl "http://10.0.0.2:8386/v1.1/%(tenant_id)s" \  
--internalurl "http://10.0.0.2:8386/v1.1/%(tenant_id)s"
```

Khởi động lại apache

`sudo service apache2 restart`

Bạn vào lại giao diện Openstack, sẽ thấy có thêm 1 mục Sahara như sau:

The screenshot shows the OpenStack Sahara web interface. The top navigation bar includes the OpenStack logo, a user dropdown menu (admin), and a 'Sign Out' button. The left sidebar contains a menu with 'Project', 'Admin', and 'Sahara' sections. Under 'Sahara', there are links for 'Clusters', 'Cluster Templates', 'Node Group Templates', 'Job Executions', 'Jobs', 'Job Binaries', 'Data Sources', 'Image Registry', and 'Plugins'. The main content area is titled 'Overview' and 'Usage Summary'. It prompts the user to 'Select a period of time to query its usage:' with input fields for 'From: 2014-07-01' and 'To: 2014-07-12', and a 'Submit' button. Below this, it displays usage statistics: 'Active Instances: 3', 'Active RAM: 3GB', 'This Period's VCPU-Hours: 18.56', and 'This Period's GB-Hours: 185.61'. A 'Download CSV Summary' button is also present. The 'Usage' section features a table with the following data:

| Project Name | VCPUs | Disk | RAM | VCPU Hours | Disk GB Hours |
|--------------|-------|------|-----|------------|---------------|
| admin | 3 | 30 | 3GB | 18.56 | 185.61 |

Below the table, it indicates 'Displaying 1 item'.

3. Hướng dẫn sử dụng

Ở đây chúng tôi sẽ hướng dẫn bạn tạo một cụm cluster hadoop thông qua giao diện UI Sahara.

Bước 1: Tải về file image và up lên glance. Ở đây chúng tôi chọn phiên bản trên ubuntu:

`wget http://sahara-files.mirantis.com/sahara-icehouse-vanilla-1.2.1-ubuntu-13.10.qcow2`

`glance image-create --name=sahara-icehouse-vanilla-1.2.1-ubuntu-13.10 --disk-format=qcow2 --container-format=bare < ./sahara-icehouse-vanilla-1.2.1-ubuntu-13.10.qcow2`

Bước 2: Đăng ký một image

Chọn Image Registry

Image: *

Select Image

User Name: *

ubuntu

Description:

vanilla 1.2.1

Register tags required for the Plugin with specified Hadoop Version

| Plugin | Version |
|---------|---------|
| vanilla | 1.2.1 |

Add all

Add custom tag

Cancel

Done

Image Registry tool:

Image Registry is used to provide additional information about images for Sahara

Specified username will be used by Sahara to apply configs and manage processes on instances.

Tags are used for filtering images suitable for each plugin and each hadoop version. To add required tags, select a plugin with Hadoop version and click "Add all" button.

You may also add any custom tag.

Unnecessary tags may be removed by clicking a cross near tag's name.

- Chọn sahara-icehouse-vanilla-1.2.1-ubuntu-13.10.qcow2
- Đặt tên image: Đặt là "ubuntu"
- Phần plugin chọn vanilla và version 1.2.1
- Cộng 3 tag vanilla, 1.2.1 và ubuntu
- Done

Bước 3: Tạo các Node Group Templates

Ở đây ta sẽ tạo 2 templates: 1 master và 1 worker

Configure Node Group Template *

Template Name: *

Description:

OpenStack Flavor: *

m1.tiny

Storage location: *

Ephemeral Drive

Floating IP pool:

ext_net

Processes: *

☐ namenode
☐ datanode
☐ secondarynamenode
☐ oozie
☐ tasktracker
☐ jobtracker
☐ hiveserver

This Node Group Template will be created for:

Plugin: vanilla

Hadoop version: 1.2.1

The Node Group Template object should specify processes that will be launched on each instance. Also an OpenStack flavor is required to boot VMs.

Sahara provides different storage location options. You may choose Ephemeral Drive or a Cinder Volume to be attached to instances.

When processes are selected, you may set **node** scoped Hadoop configurations on corresponding tabs.

Cancel

Create

- Đặt tên cho template
- Chọn flavor cho template
- Phần Storage location bạn có thể chọn Ephemeral Drive
- Phần Floating IP pool, bạn có thể chọn external network
- Với Processes bạn chọn:
Worker: tasktracker, datanode
Master: jobtracker, namenode

Bước 4: Tạo 1 cluster template

Đặt tên cho cluster template

Create Cluster Template

Details *

Node Groups

General Parameters

HDFS Parameters

MapReduce Parameters

Template Name: *

Description:

Use anti-affinity groups for :

☐ namenode
☐ datanode
☐ secondarynamenode
☐ oozie
☐ tasktracker
☐ jobtracker
☐ hiveserver

This Cluster Template will be created for:

Plugin: vanilla
Hadoop version: 1.2.1

The Cluster Template object should specify Node Group Templates that will be used to build a Hadoop Cluster. You can add Node Groups using Node Group Templates on a "Node Groups" tab.

You may set **cluster** scoped Hadoop configurations on corresponding tabs.

The Cluster Template object may specify a list of processes in anti-affinity group. That means these processes may not be launched more than once on a single host.

Cancel

Create

Tiếp theo, khai báo số node groups. Bạn có thể điều chỉnh số lượng các node bằng cách ấn + hoặc -

Create Cluster Template

Details *

Node Groups

General Parameters

HDFS Parameters

MapReduce Parameters

test-worker-templ

+

| Group Name | Template | Count | |
|-------------------|-------------------|-------|---|
| test-master-templ | test-master-templ | 1 | <div>-</div> <div>+</div> <div>Remove</div> |
| test-worker-templ | test-worker-templ | 2 | <div>-</div> <div>+</div> <div>Remove</div> |

Cancel

Create

Bước 5: Tạo cluster

Launch Cluster

Cluster Name: *

Description:

Cluster Template: *

demo-cluster-template

Base Image: *

sahara-icehouse-vanilla-1.2.1-ubuntu-13.10

Keypair:

mykey

+

Neutron Management Network: *

int_net

This Cluster will be started with:

Plugin: vanilla

Hadoop version: 1.2.1

Cluster can be launched using existing Cluster Templates.

The Cluster object should specify OpenStack Image to boot instances for Hadoop Cluster.

User has to choose a keypair to have access to clusters instances.

Cancel

Create

Ở đây, bạn khai báo:

- Tên cluster muốn đặt
- Chọn cluster template đã tạo ở bước 4
- Keypair
- Neutron Management Network

Khi bạn click Create, vào phần Instances bạn sẽ thấy các máy sẽ được tạo ra:

Instances

Instances

Filter

Filter

+ Launch Instance

Soft Reboot Instances

Terminate Instances

| <input type="checkbox"/> | Instance Name | Image Name | IP Address | Size | Key Pair | Status | Availability Zone | Task | Power State | Uptime | Actions |
|--------------------------|-----------------------------------|--|-------------------------------|---|----------|--------|-------------------|------|-------------|-------------------|--|
| <input type="checkbox"/> | demo-cluster-test-worker-tmpl-002 | sahara-icehouse-vanilla-1.2.1-ubuntu-13.10 | 10.20.10.13 192.168.50.173 | m1.hadoop 1GB RAM 1 VCPU 10.0GB Disk | mykey | Active | nova | None | Running | 1 day, 2 hours | <div>Create Snapshot</div> <div>More ▾</div> |
| <input type="checkbox"/> | demo-cluster-test-worker-tmpl-001 | sahara-icehouse-vanilla-1.2.1-ubuntu-13.10 | 10.20.10.12 192.168.50.175 | m1.hadoop 1GB RAM 1 VCPU 10.0GB Disk | mykey | Active | nova | None | Running | 1 day, 2 hours | <div>Create Snapshot</div> <div>More ▾</div> |
| <input type="checkbox"/> | demo-cluster-test-master-tmpl-001 | sahara-icehouse-vanilla-1.2.1-ubuntu-13.10 | 10.20.10.11 192.168.50.174 | m1.hadoop 1GB RAM 1 VCPU 10.0GB Disk | mykey | Active | nova | None | Running | 1 day, 2 hours | <div>Create Snapshot</div> <div>More ▾</div> |

Displaying 3 items

Bây giờ bạn có thể ssh vào và trải nghiệm:
`ssh -i your_keypair.pem ubuntu@ip_master_hadoop`

4. Một số lỗi hay gặp phải

4.1. Command 'gcc' failed with exit status 1

Khi cài đặt sahara: [sahara-venv/bin/pip install sahara](#), nếu gặp lỗi:

```
ERROR: /bin/sh: 1: xslt-config: not found
```

```
...
```

```
In file included from src/lxml/lxml.etree.c:346:0:
```

```
/root/sahara-venv/build/lxml/src/lxml/includes/etree_defs.h:9:31: fatal error: libxml/xmlversion.h:
```

```
No such file or directory
```

```
...compilation terminated.
```

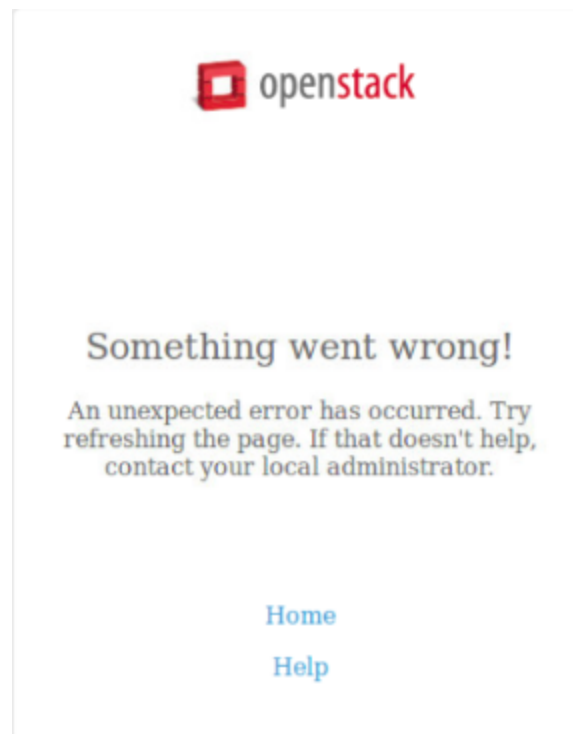
```
error: command 'gcc' failed with exit status 1
```

Cách khắc phục: Cài thêm các gói libxml2-dev và libxslt-dev:

[sudo apt-get install libxml2-dev libxslt-dev](#)

4.2. Something went wrong

Nếu bạn nhận được thông báo lỗi:



Hãy kiểm tra xem bạn đã khởi động Sahara hay chưa, nếu chưa hãy khởi động nó:

[sahara-venv/bin/sahara-api --config-file sahara-venv/etc/sahara.conf](#)

4.3. Lỗi DB khi reboot lại server

Vì một lý do nào đó server cài Openstack của bạn reboot lại. Khi bạn start sahara-api lên hiện thông báo lỗi:

```
ERROR sahara.openstack.common.periodic_task [-] Error during
SaharaPeriodicTasks.update_job_statuses: (OperationalError) no such table: job_executions
u'SELECT job_executions.created_at AS job_executions_created_at,
job_executions.updated_at AS job_executions_updated_at, job_executions.id AS
job_executions_id, job_executions.tenant_id AS job_executions_tenant_id,
job_executions.job_id AS job_executions_job_id, job_executions.input_id AS
job_executions_input_id, job_executions.output_id AS job_executions_output_id,
job_executions.start_time AS job_executions_start_time, job_executions.end_time AS
job_executions_end_time, job_executions.cluster_id AS job_executions_cluster_id,
job_executions.info AS job_executions_info, job_executions.progress AS
job_executions_progress, job_executions.oozie_job_id AS job_executions_oozie_job_id,
job_executions.return_code AS job_executions_return_code, job_executions.job_configs AS
job_executions_job_configs, job_executions.extra AS job_executions_extra \nFROM
job_executions \nWHERE job_executions.end_time IS NULL' ()
```

Hãy kiểm tra lại file sahara.conf, nếu bạn đang đặt là: connection=sqlite:///tmp/sahara.db

Nguyên nhân: Do file db sahara bạn đặt trong thư mục /tmp/ nên khi reboot lại server, file này bị clean dẫn đến lỗi trên.

Cách khắc phục: Hãy thay bằng một đường dẫn khác.

4.3. Hadoop không chạy khi reboot lại cluster

Khi bạn khởi động lại cluster (<=> khởi động lại VMs), sẽ gặp vấn đề đó là tất cả các processes của hadoop không "tự khởi động".

Nguyên nhân: Do sahara không add các processes hadoop vào startup script.

Cách khắc phục: Trên thực tế không có một tập tin để khôi phục lại tất cả các processes. Chúng tôi khuyến cáo bạn nên khởi động lại cluster (VMs hadoop). Quá trình phục hồi phụ thuộc vào phiên bản hadoop bạn sử dụng.

Đối với bản vanilla 1.2.1, bạn chạy:

`sudo su -c "/usr/sbin/hadoop-daemon.sh start X" hadoop` trong đó X là processes hadoop.