

# Lab 1

Course: Big Data and Machine Learning

Instructors:

Salman Toor, [Salman.Toor@it.uu.se](mailto:Salman.Toor@it.uu.se)

## Introduction:

The aim of this Assignment is to give you hands-on experience with the computing infrastructure used in this course, to give you some experience with working in the Linux command prompt on the Virtual Machines, and to introduce the Hadoop framework. In the lab we will use SNIC Science Cloud (SSC). SSC is a national resource (a community cloud) that provides Infrastructure-as-a-Service (IaaS). It is built using OpenStack cloud software and Ceph storage and offers the following services:

1. Instance management
2. Storage management
3. Identity management
4. Image management
5. Network management

## Important links:

1. Information page: <https://cloud.snic.se/>
2. User guide: <https://docs.openstack.org/horizon/latest/user/>
3. Dashboard: <https://cloud.snic.se/>

Your goal in this assignment is to, using the user-guide, complete the tasks below. Each task has two sections, "Hands-On" and Questions. We expect brief answers to the questions, and they should be uploaded in writing as a properly formatted pdf-document.

Good Luck!

## Task-1 (Instance management)

### Hands-on:

1. Start one instance of flavors *small*.
  2. Assign floating IPs to each of the instances.
  3. Access them both via VNC and SSH client.
  4. Open port *5000* on one of the instance (by adding a new security group).
  5. Create a snapshot of the instance.
- 

Note: For Windows using here are some details how to access Linux instance:

<https://support.rackspace.com/how-to/logging-in-with-an-ssh-private-key-on-windows/>

---

### Questions:

1. What is the difference between the local network (SNIC 2019/10-9 Internal IPv4 Network) and external network (*public*)?
  2. Can a VM access the Internet without assigning a floating IP? Explain the role of the floating IP.
  3. What is the difference between an Image, Instance and a Snapshot?
- 

## Task-2 (Storage management)

### Hands-on:

1. Create a volume of size 1GB.
2. Attach your newly created volume to your instance.
3. Access the volume and copy files to the attached volume.
4. Extend the size of the volume created in step 1.

### Questions:

1. What is the technology used to provide volumes in OpenStack? Is it RAID or LVM?
  2. What is LVM? Explain the advantage(s) of using LVM?
  3. Can one volume be attached to multiple instances or vice versa?
-

*Clean up: Terminate all your running instances, delete the volume, and delete the snapshot.*

---

## Task-3 (Object Storage)

### Hands-on:

1. Find out the public url of the Container "Lab1".
2. Download the object NIST.pdf from the Horizon dashboard.
3. Try to create a container named "testcontainer", did it work? If not, why?

### Questions:

1. Explain the difference between a folder on your UNIX filesystem and a pseudo-folder inside a container?
  2. The corresponding system in Amazon Web Services is called "S3". Is there a principal difference between an "S3 bucket" and a container in OpenStack's object store?
  3. What is the name of the OpenStack service providing Object storage?
- 

## Task-4 (Cowsay-as-a-Service)(Optional)

### Hands-on:

In this task you will deploy a simple service to the benefit to the world.

Access your VM and start by installing the program "cowsay" (use "apt-get"). Then proceed to use "git" to clone the repository

```
https://www.github.com/TDB-UU/csaas
```

cd into the folder csaas/cowsay and then execute

```
$ screen python app.py
```

If you get any messages about missing packages, just go ahead and install them using "pip" (a Python package management system)

Test that things are working by executing (from your client)

```
$ curl -i http://<your\_public\_ip>:5000/cowsay/api/v1.0/saysomething
```

## Questions:

1. Examine the code in app.py. What Python framework is used to provide the (extremely simplistic) RESTful service?
2. What problem does “screen” solve?
3. Is SSC a Public, Community, Private or Hybrid cloud, and why?