# Ecole Polytechniuqe Montreal
# LOG8415 : Lab 1
# Selecting VM instances in the Cloud through benchmarking

**Abstract**

In this assignment, first you will learn about a Cloud provider computing service. Then you will perform extensive benchmarking on several virtual machines instance types. After this first step, you will be asked to select appropriate VM instances for the deployment of a cloud application. Finally, you are asked to prepare and submit a report on Moodle. Your report should contain the results of the benchmarking of VMs and a justification for your selection of the VM instances.

Keywords: Benchmarking; Cloud Computing; Amazon EC2; Cloud Providers, Virtual Machines

## 1 Objectives

Amazon Web Services (AWS) is a leading infrastructure as a service Cloud provider. In this lab assignment, you will be comparing the performance of different AWS EC2 instances in terms of CPU, Memory, IO and network performance. You will be reporting your findings by producing a LaTeX article. During the Lab session, some guidelines on how to use the provider will be presented to you. Amazon Web Services [1] is a collection of remote computing services, also called web services, that make up a cloud-computing platform offered by Amazon.com. These services operate from 12 geographical regions across the world for building, deploying and managing applications and services through a global network of hosted data centers. It provides both PaaS and IaaS services and supports many different programming languages, tools and frameworks, including both AWS-specific and third-party software and systems.

During the first laboratory session, the Lab instructor will present an introduction to Cloud Computing and AWS EC2. He will provide necessary guidelines to create instances, how to set them up and use SSH to connect to them. Once you get familiar with the Cloud provider, you should create instances and start benchmarking. It is important to stop EC2 instances when you are not working on them. You will have 200 CAD credit provided by AWS for three assignments of

this course. You should manage your expense carefully.

The overall goals of this lab assignment are:

- Benchmarking various Amazon EC2 Virtual Machines to compare their performances.
- Understand the requirements of a concrete cloud problem in terms of resources usages.
- Report your findings in an article typeset using LaTeX.

## 2 Benchmarking

The performance of applications deployed in cloud environments is influenced by the policy decisions of cloud providers: bandwidth rate-limits, VM packing decisions, or CPU scheduling policies. It is highly recommended to analyze the performances of a cloud computing service before taking a decision to migrate applications. In order to run benchmarks, login to your AWS EC2, and create Ubuntu (latest version) virtual machines according to the figure below. In the following, we provide you with a step by step training on cloud benchmarking:

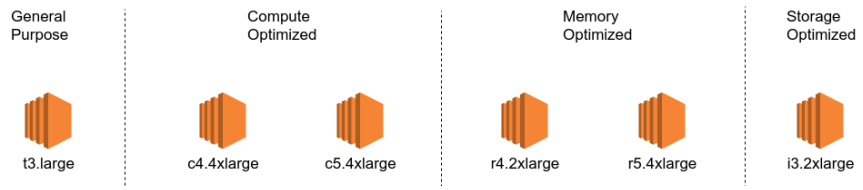


Figure 1: Design Architecture Elements

After you are done practicing launching and stopping virtual machine instances, you will need to perform some benchmarking. Table 1 contains some standard benchmarks that are widely used to evaluate virtual servers using some metrics. You can extend this table by proposing others metrics of comparison and explaining your choices. In addition to running these benchmarks, you should also report the characteristics of your virtual machines for CPU, Memory, Disk and Network speed. Create a table, along with figures in your report and summarize them.

Table 1: Benchmarks

| Performances of | Benchmark |
|---|---|
| IO | dd if=/dev/zero of=sb-io-test bs=1M count=1k conv=fdatasync |
| CPU Memory Disk | Phoronix Test Suite |
| IOPS | Bonnie++ |
| Memory | stress-ng |
| Disk | hdparm |
| Network Throughput | SpeedTest |

Please note that in most cases, benchmarking must be done using regression analysis to find the boundaries of input parameters. In this assignment, Lab instructor will help you define the average input parameters, and there is no need to perform any extra analysis.

— Make sure that you attend the lab session to better understand how to determine the values required in your benchmarking. It will be explained during the lab session only.

### 2.1 Method

We recommend that you put commands in a script for running and gathering the test results for all the benchmarks you select. Once you are satisfied with your script, you have to perform all benchmarking on the different Amazon EC2 instances separately. Instruction on how to choose VMs from the list of services will be given to you during our first lab session. For each instance type, you should execute your selected benchmarks at least 5 times and take **the average**.

Make a plot

### 2.2 Analysis

You should provide a fair comparison between your results of benchmarking for EC2 instances. For example, you can generate a plot showing the results of CPU benchmarking. This plot can provide the differences between the family types.

**Resources**

- Phoronix Test Suite [4]
- Stress-ng on Ubuntu OS; [6]
- Performance Evaluation of Virtualization in Cloud Data Center; [3]
- hdparam on Ubuntu; [2]
- Web Server Performance Leader board [5]

*Above links are just examples and it is not mandatory to follow instructions represented on them. You must perform extensive research on the web in order to understand each and every benchmarking tool.*

## 3 Problem

An online retailer company (Beta X) wants to benefit from data science to obtain insights on how the business changes over time, what are the major questions that can be addressed and consequently how they can improve their business. In order to do so, they would like to consult engineers to build data pipe-lines to extract data and load into business intelligent models. In the rapidly changing environment of Cloud Computing, an enterprise must gain the most out of data science to understand business obstacles and provide data driven information to its

executives. You have been recruited as consultant to assist the company with this difficult task. bellow is the design of the cloud-based application that the company is building.

- 1) Storage: Historical data will be kept in these VMs (storage focused)
- 2) Data extraction and load into a clustered Database (extensive I/O)
- 3) Business intelligent services (heavy computations) for visualization

Please see the application design below ( Figure 2) to understand the overall concepts. You are asked to define which instance is the best with regard to disk speed, storage capacity, IO speed and cost?

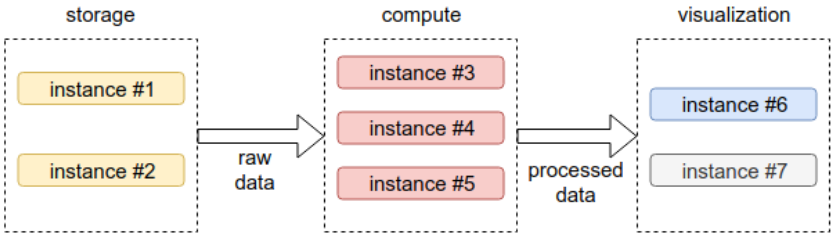


Figure 2: Design Architecture Elements

Please note that your solution:

- Must consider AWS Cloud provider EC2 Ubuntu machines.
- Should be based on your benchmarking results.
- Should contain maximum of three instances per element depicted in design figure.
- Should include virtual machines running Ubuntu::latest version.

# 4 Working in Group

You should work in groups of two or three. It is recommended to use either a Bash script and/or Python to write your benchmarking script. Your final submission guidelines are provided at the end of this document.

# 5 Paper

A submission per group is required for this assignment. To present your findings, you have to write a lab report on your benchmarks using LaTeX template. This lab description document is an example of a LaTeX document.

In your report, answer the following questions:

- 1) What are the CPU, Memory, and Disk characteristics of each of the VM instances? Include a table that summarizes these properties.
- 2) What benchmark applications did you choose and why? Specifically, what do their results tell us about the system?
- 3) What issues if any did you have in performing these benchmarks? How did you overcome these challenges?
- 4) What were the results of your benchmarks? For each benchmark show a graph or plot of the results and describe what happened. Did you encounter any surprising results?
- 5) How does the VM selection fit into the benchmarking results? For each VM instance type in the required architecture (Figure 2), explain in details how did the results help you understand the best choice between the family types of EC2 machines?

Need to answer these questions

One submission per group is required for this assignment. You must submit only one ZIPPED file named **tp1.zip** which includes:

  - **tp1.pdf** Containing the name of the team members and all the material of the report. You can design the format and sections at your convenience. Feel free to use any type of representation such as figures, tables etc.

  - **scripts.sh** Which will include all the necessary commands to run the benchmarks. It should have enough comments and description to understand every single section of it.

Python?

# 6 Evaluation

A single final submission per group for this assignment is due on the date assigned on Moodle web site.

Your assignment will be graded on content as following:

- 3 pts: Description of machines characteristics, benchmarks and metrics;
- 10 pts: Results and analysis of your benchmarks performed in EC2;
- 4 pts: Explain and justify the selection of VM instances for the application;
- 2 pts: Scripts and commands (correctness, structured, ...);
- 1 pts: General presentation and the quality of the report. It is important to respect the format of the submission.

- Please note that all team members must participate in all the step of this assignment. Should you have any questions, please do not hesitate to contact the

Lab instructor: a.abtahizadeh [at] polymtl [dot] ca

*Required skill sets for assignment no.1 : Linux, Bash Scripts, LaTeX*

**Remember to check the forum page on Moodle regularly as any kind of announcement/help/guide will be put in there:**
https://moodle.polymtl.ca/mod/forum/view.php?id=291432

# 7 ACKNOWLEDGMENT

We would like to thank Amazon for supporting us through their AWS in Education plan.

# References

[1] Amazon web services. `http://aws.amazon.com`. Accessed: 2020-01-10.

[2] hdparam on ubuntu. `http://manpages.ubuntu.com/manpages/trusty/man8/hdparm.8.html`. Accessed: 2020-01-10.

[3] Performance evaluation. `http://www.diva-portal.org/smash/get/diva2:563427/FULLTEXT01.pdf`. Accessed: 2020-01-10.

[4] The phoronix test suite. `https://github.com/phoronix-test-suite/phoronix-test-suite/`. Accessed: 2020-01-10.

[5] Server bear. `http://serverbear.com/benchmarks`. Accessed: 2020-01-10.

[6] Stress-ng on ubuntu. `http://kernel.ubuntu.com/~cking/stress-ng/`. Accessed: 2020-01-10.