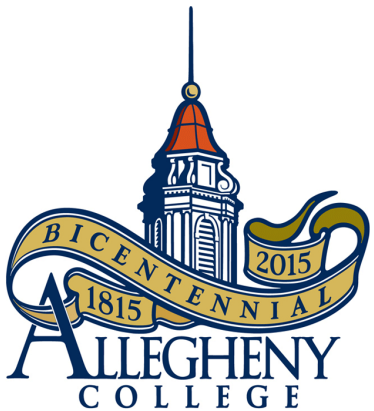
CS402 - Cloud Computing

An Introduction

Aravind Mohan

Allegheny College

August 31, 2020



Meeting Time

- Lecture Session:
  - Monday and Wednesday
    10:20 AM - 11:10 AM, Alden 109 (Online)
    rotation basis
- Practical Session:
  - Friday 10:20 am - 11:10 am, (Online)
- Lab Session:
  - Monday 3:00 PM - 4:50 PM, (Online)

Please read the **Syllabus** for more details on the rotation schedule.

Professor's Office Hours

- Monday, Wednesday, and Friday:
  11:15 am - 12:15 pm
- Tuesday and Thursday:
  10:00 am - 11:30 am

Office hours will be virtual this semster.

To schedule an office hours time slot, please visit my website [teaching page] and click on the **Schedule Meeting** link located on the top right-hand corner to schedule 15 mins slots.
Let us connect and learn from each other...

Website Details

- Professor's Website:

  `https://www.cs.allegheny.edu/sites/amohan/`

- Course Website:

  `https://www.cs.allegheny.edu/sites/amohan/course.php?cid=MTU=`

Textbooks

- There is not a single required textbook for this class.
- Reading assignments will be provided as required.

Administrative Stuff!

- No Lab this week.

  First lab next week on Wednesday, 9th Sep 2020.

Administrative Stuff!

- Practical Assignments
- Laboratory Assignments
- Skill Tests **(3)**
- Course Project
- Class Participation

Please read the **Syllabus** to get an overview of the course.

Administrative Stuff!

Practical Assignments	**15%**
Laboratory Assignments	**35%**
Skill Tests	**15%**
Course Project	**25%**
Class Participation	**10%**

Gradebook will be shared through Canvas. More details in **Syllabus**.

Tips for Success

- Attentively listen to classes and try to participate in all class discussions.
- Bring a notebook with you and start making detailed notes during every class period.
- Clarify with the Professor, if a lesson is confusing.
- Complete all the reading assignments thoroughly.
- Participate in all the in-class activities.

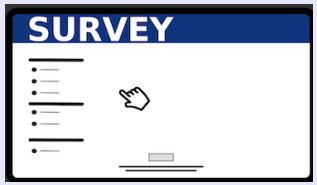
Be ready to **think**, **process**, and **implement** scalable compute and data inten

Interaction between us ...

- Any question is a valid question. No question is good or bad. So, questions are always welcome.
- Interaction is the best way to get rid of long lectures. So, let us try to interact more so that communication is a two-way stream and the class is not boring.

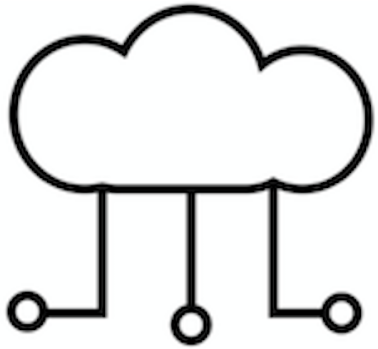
Ask your questions in Google Meet chat window.

Things To Do



Read **Syllabus** before next class!

360° view on computing



So, far we might have done computing on our local machines. But,

- Can we compute (at scale/distributedly) on your machine?
- Can we compute on a different machine?
- Can we compute on a bunch of machines?
- Can we do all these at the same time?

We will find it out in this course!

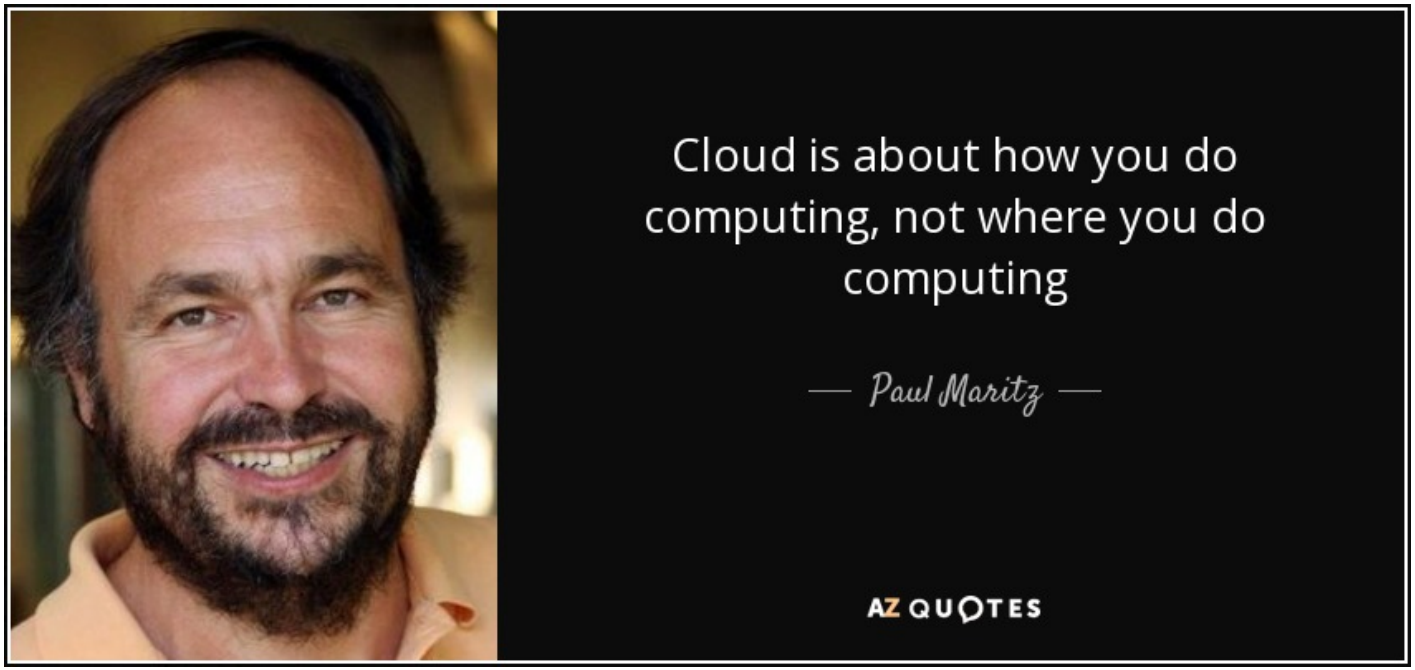
360° view on datasets

So, far we might have used some small scale datasets while doing our computing. But,

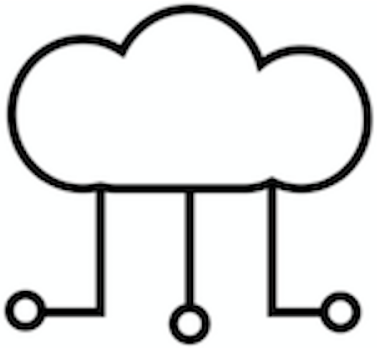
- Can we collect, streamline, and share large scale datasets from websites, such as blogs, online forums, etc ... ?
- Can we collect, streamline, and share highly scalable datasets from social media sites such as twitter, spotify, etc ...?

We will find it out in this course!

In a nutshell



This is so true from an end user's perspective.



Learning Objectives

The four major learning objectives in this course are summarized below:

- **Lo_1** - What are the fundamental principles behind cloud-based distributed systems?
- **Lo_2** - How to collect and share web 2.0 data?
- **Lo_3** - What is the scheduling problem in Cloud Computing? and how do we solve it?
- **Lo_4** - How to set up and integrate open source Cloud platforms?

Large Scale Computation + Basic Networking + Datasets + Cloud

Learning Goals

By the end of this course, you'll in general

- Master a variety of distributed computation techniques, dataset development strategies, and integrating cloud services.
- Be well equipped to learn advanced techniques in the field of Cloud Computing, Big Data, Data Science, and so on ...
- Be prepared to take on bigger challenges on your senior thesis and at work after graduation.

What do we do in Labs?

- Combination of individual and team-based labs.
- Solve compute intensive problems in a scalable (distributed) manner.
- Develop charts to conduct experimental study.
- Implement one or more strategies to develop datasets.
- Integrate one or more networked services in the Cloud.

We will mainly program in Java in this course!

A Practical Example

Compute: Find the total number of primes between 0 and a given number α

- Let us suppose that, we are also given a number β, which represents the total number of workers.
- To scale the computation, we are required to **divide** the work equally between each worker, and **solve** the problem in parallel and fast.
- Assume α and β is unknown, what is the range for each worker? **(formalize the sequence)**
- Look at the (Prime.java) code file in the class repo.

A Practical Example

Example 1:

$$\alpha = 100 \ \& \ \beta = 5$$

(0 - 20)	(20 - 40)	(40 - 60)	(60 - 80)	(80 - 100)
w_1	w_2	w_3	w_4	w_5

Example 2:

$$\alpha = 240 \ \& \ \beta = 4$$

(0 - 60)	(60 - 120)	(120 - 180)	(180 - 240)
w_1	w_2	w_3	w_4

Getting to know each other!

Brainstorm with your peers in Slack & come up with ideas to solve this problem?

- Add a reflection markdown file to the repository.
- Commit and Push changes.

Next Class:

Threads:

- How to scale computation within your machine?
- How to execute computation in parallel within your machine?

Questions?

Please ask if there are any Questions!