18-847F: Special Topics in Computer Systems

Foundations of Cloud and Machine Learning Infrastructure

Lecture 1: Introduction and Logistics

Foundations of Cloud and Machine Learning Infrastructure

Graduate Seminar Class

(Almost) no lectures

Reading research papers

Student presentations

Class Discussions

Final Research Project (No Exams!)

Learning Objectives

o Know the state-of-the-art frameworks in cloud and machine learning and their theoretical foundations

o Read and provide constructive criticism of research papers

o Present to an audience, and answer their questions

o Do creative, collaborate research

Why study Cloud and ML infrastructure?

What are the largest words after 'Big Data'?

Big Data Gold Rush

Who got rich in the
California gold rush?

Big Data Gold Rush
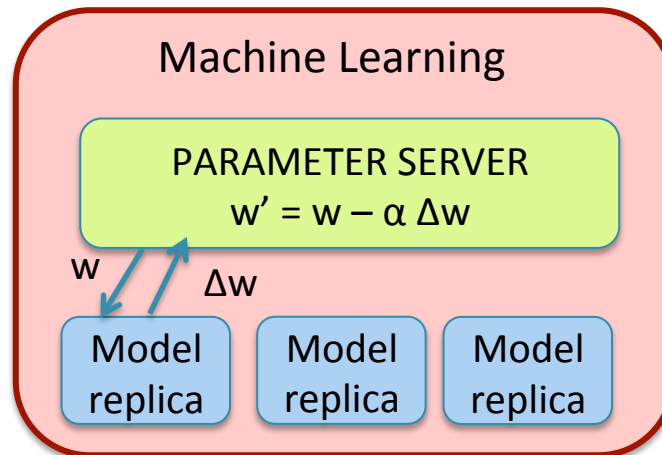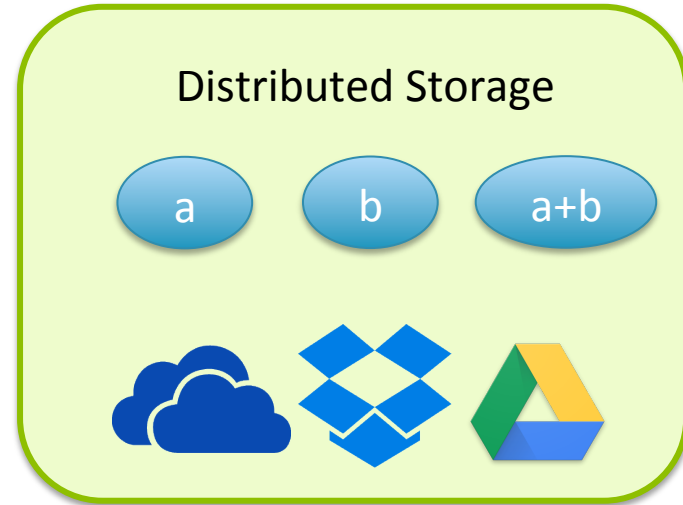

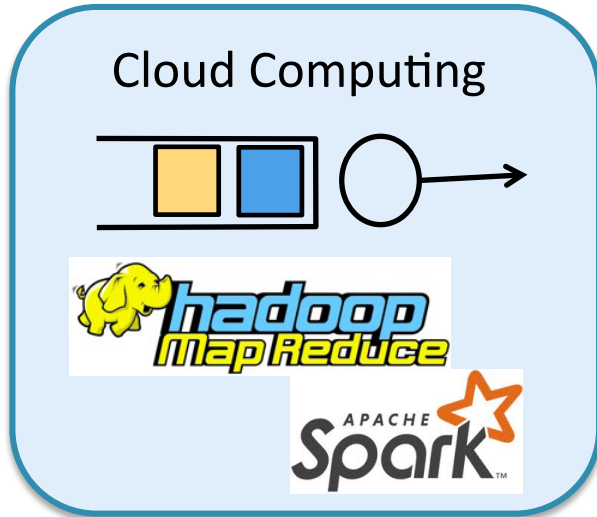Who got rich in the California gold rush?

In the Big Data rush, it's the infrastructure companies

Topics Covered

Cloud Computing

Distributed Storage

a   b   a+b

Machine Learning

PARAMETER SERVER
$w' = w - \alpha \Delta w$

w Δw

Model replica   Model replica   Model replica

Topics Covered

Cloud Computing
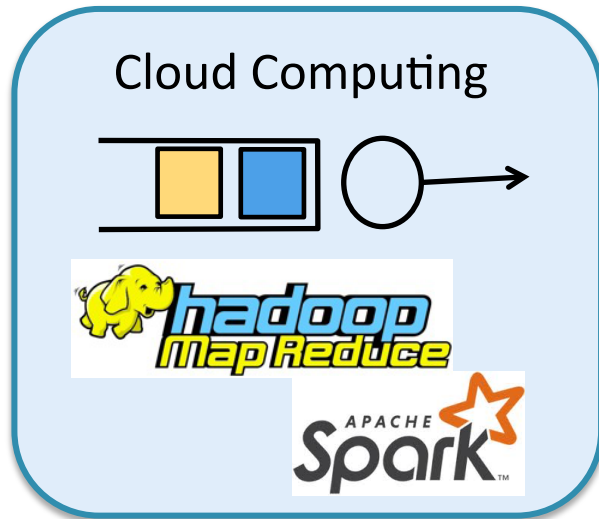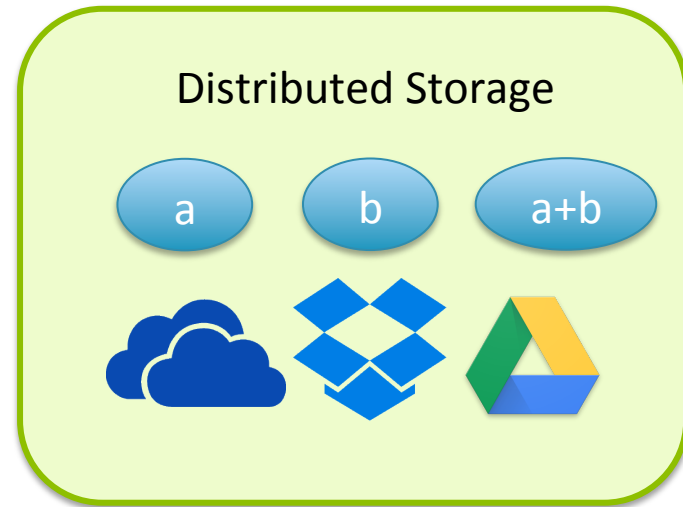
o   Scheduling in Parallel Computing

o   MapReduce, Spark

o   Straggler Replication

o   Task Replication in Queueing Systems

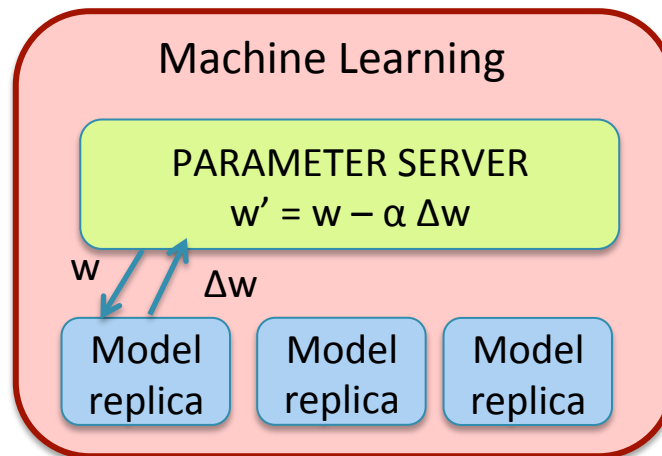o   Coded MapReduce

Topics Covered

o   Coding for locality/repair

o   Systems implementation of codes

o   Reducing latency in content

download

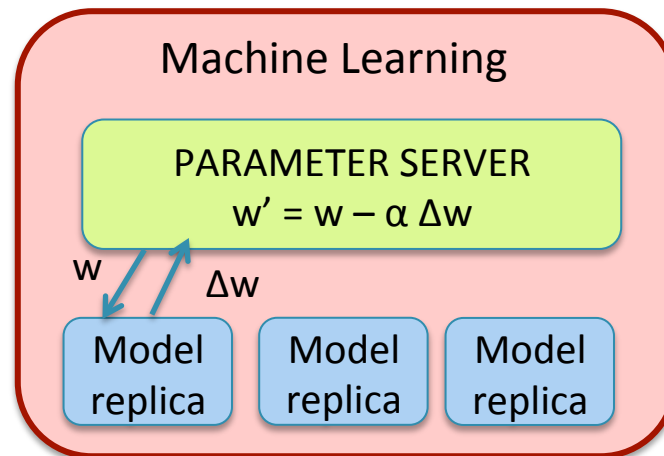

Distributed Storage

a   b   a+b

Topics Covered

o SGD and Support Vector Machines

o Backpropagation, LeNet, AlexNet, GoogleNet

o Distributed Gradient Descent

o Hyper-parameter tuning

o GANs, Deep reinforcement learning

Machine Learning

PARAMETER SERVER
$w' = w - \alpha \, \Delta w$

w Δw

Model replica   Model replica   Model replica

Topics Covered

Cloud Computing

Distributed Storage

a   b   a+b

Machine Learning

PARAMETER SERVER
$$w' = w - \alpha \, \Delta w$$

w Δw

Model replica   Model replica   Model replica
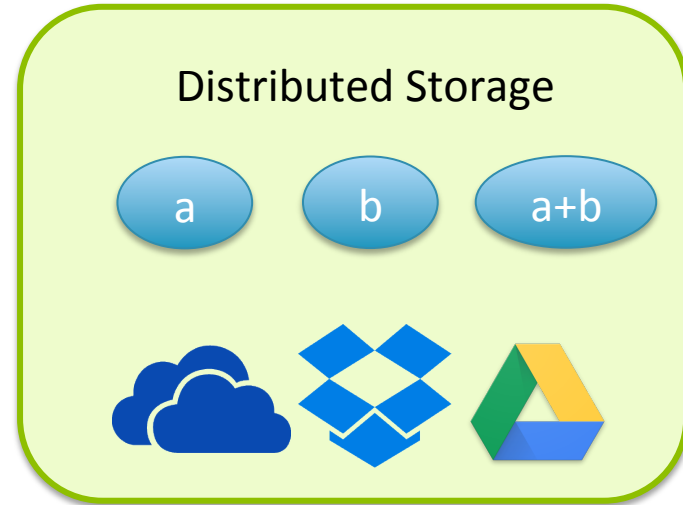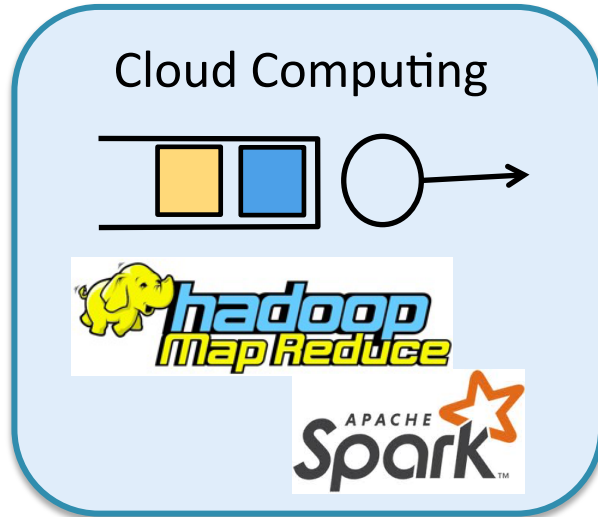
Instructor: Gauri Joshi

B.Tech+M.Tech
2005-2010

SM + PhD
2010-2016

Research Staff Member
2016-2017

Assistant Professor
Fall 2017 -

Internships

Have worked in all these areas

Cloud Computing

Distributed Storage

a   b   a+b

Machine Learning

PARAMETER SERVER
$w' = w - \alpha\, \Delta w$

w   Δw

Model replica   Model replica   Model replica
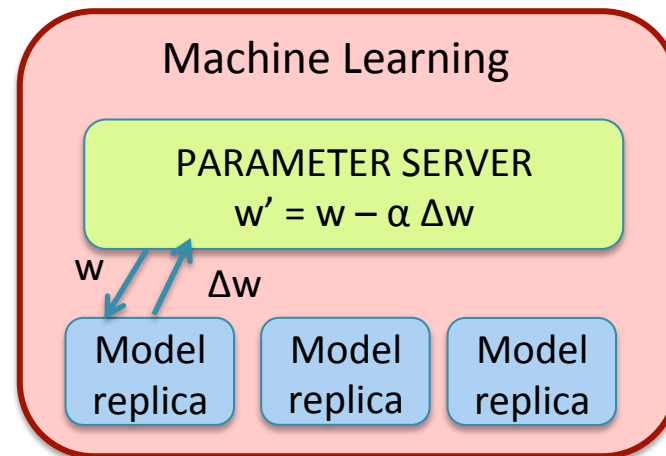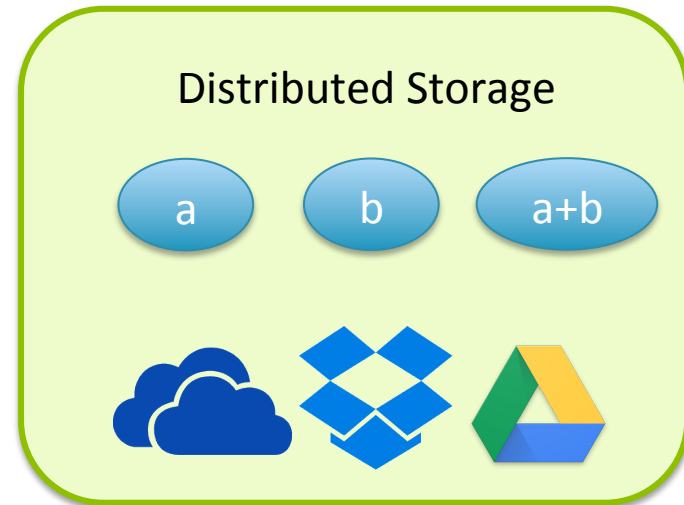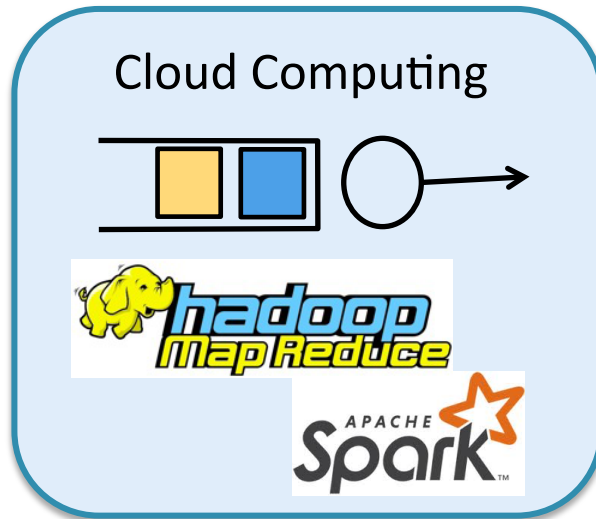
Student Introductions

o Name?

o Department?

o Masters or PhD?

o Previous related classes (if any)?

o What you are looking to learn from this class?

Class Hours and Website(s)

o When: Mon, Wed 4:30-6:00 pm

o Where: Scaife Hall 222

o Class Website (Readings, Schedule): https://www.andrew.cmu.edu/user/gaurij/18-847F-Fall-2017.html

o Canvas Site (Readings, Assignments, Projects): https://canvas.cmu.edu/

Reading Material

Papers will be posted on the class website or on Canvas

o   Book chapters

o   Survey papers

o   Theory papers (Scheduling, Queuing, Coding, Optimization)

o   Systems papers (Cloud, Machine Learning)

Additional reference books listed in the syllabus

Instructor and Office Hours

Prof. Gauri Joshi, ECE Dept

Email: gaurij [AT] andrew.cmu.edu

Office Location: CIC 4105

Office Hours: Wed 2:00-3:00pm or by appointment

Graduate Seminar Class

(Almost) no lectures

Reading research papers

Student presentations

Class Discussions

Final Research Project

Lectures

o   Next week: Deeper Overview of course topics

o   3-4 Guest lectures during the semester by authors of papers relevant to this class

Graduate Seminar Class

(Almost) no lectures

Reading research papers

Student presentations

Class Discussions

Final Research Project

Homeworks (45%)

o Class divided into two groups; read the paper assigned to your group

o Submit paper review (due 9:00 am before class)
   o Two reviews per week (advise: finish them together!)

o Discussion with your group is okay, but write reviews in your own words. List collaborators in the homework submission

Paper Review Format

o Brief Summary of the paper

    o Reflects your understanding of the paper

    o No judgments (positive or negative) here

o High-level technical/writing comments

    o Significance & correctness of results (Don't be mean!)

    o Paper organization

o Low-level technical/writing comments

    o Smaller clarifications, corrections, typos

o Discussion Questions for Class (at least 2)

    o Confusions about the paper, open research directions

Tentative Grading Rubric (Total: 10 pts)

- o Clarity, Organization (1 pt)

- o Understanding of the paper (5 pts)

- o High-level comments (2 pts)

- o Low-level comments (2 pt)

Graduate Seminar Class

(Almost) no lectures

Reading research papers

Student presentations

Class Discussions

Final Research Project

Class Presentations (20%)

o One student from each group signs up for presentation, at least 1 week in advance

o Each student will present ~2 times in the semester

o 20 min presentation, followed by 25 min discussion

- o Motivation and Related work

- o Summary of main results

- o Your views on the paper

Tentative Grading Rubric (Total: 10 pts)

o Motivation (2 pts)

o Clarity (2 pts)

o Correctness (2 pts)

o Engaging the audience (2 pts)

o Extra research, going beyond the paper (2 pts)

Graduate Seminar Class

(Almost) no lectures

Reading research papers

Student presentations

Class Discussions

Final Research Project

Tentative Grading Rubric (Total: 5 pts)

o Discussion questions (2 pt)

o Attendance and attention (1 pt)

o Speaking up in class (1 pt)

o Insightful Questions/Comments (1 pt)

Graduate Seminar Class

(Almost) no lectures

Reading research papers

Student presentations

Class Discussions

Final Research Project

Research Project (25%)

o Groups of 2 or 3

o Original research on a topic of your choice
- o Topics aligned with your research allowed and encouraged
- o If you can't think of topics, come talk to me!

o Possible Project Types:
- o New theoretical analysis
- o Implementation using one of the frameworks discussed
- o In-depth literature survey of a particular topic

Timeline

o   1-page proposal due Sept 27

o   Publishable quality report (max 5 pg) in ACM format
   o   Initial draft due: Nov 22
   o   Final report due: Dec 8

o   Peer-review 2 other reports

o   Last 2 weeks of class: Presentations (20 min each)

Tentative Grading Rubric (Total: 20 pts)

- o Originality (1 pts)

- o Review of Related Work (1 pts)

- o Writing and Organization (2 pts)

- o Technical Results (5 pts)

- o Peer-Review of Other reports (1 pts)

- o Final presentation (10 pts)

In Summary..

o   Paper Reading

o   Submitting Reviews

o   Class Presentations (~2 in the semester)

o   Final Project

Might seem like a lot of work but..

o   You will get fast and efficient at reading papers

o   The project will be a fun, collaborative exercise

o   No exams!

TO DO

o Form groups for class readings

o Sign-up for presentation

o Form groups for class projects

o Start reading the papers

o Start thinking about projects

Next Class

History and overview of cloud and
machine learning infrastructure

Will give you additional time for the
first paper reviews and presentations