Lecture 2 - Overview of HPC and the Cloud

DSE 512

Drew Schmidt
2022-01-27

From Last Time

- Assignment 1 due tonight

HPC

HPC

- High Performance Computing
- Usually used interchangeably with "supercomputing"
- Summit, ISAAC, ...
- The Cloud?
- Your desktop???
- Your phone?????

High Performance?

Supercomputing

Characteristics

- Big
- Parallel file system
- Multiple nodes
- High speed interconnect
- Batch programming

Non-characteristics

- Heterogeneous/homogeneous
- Big FLOPS

Summit

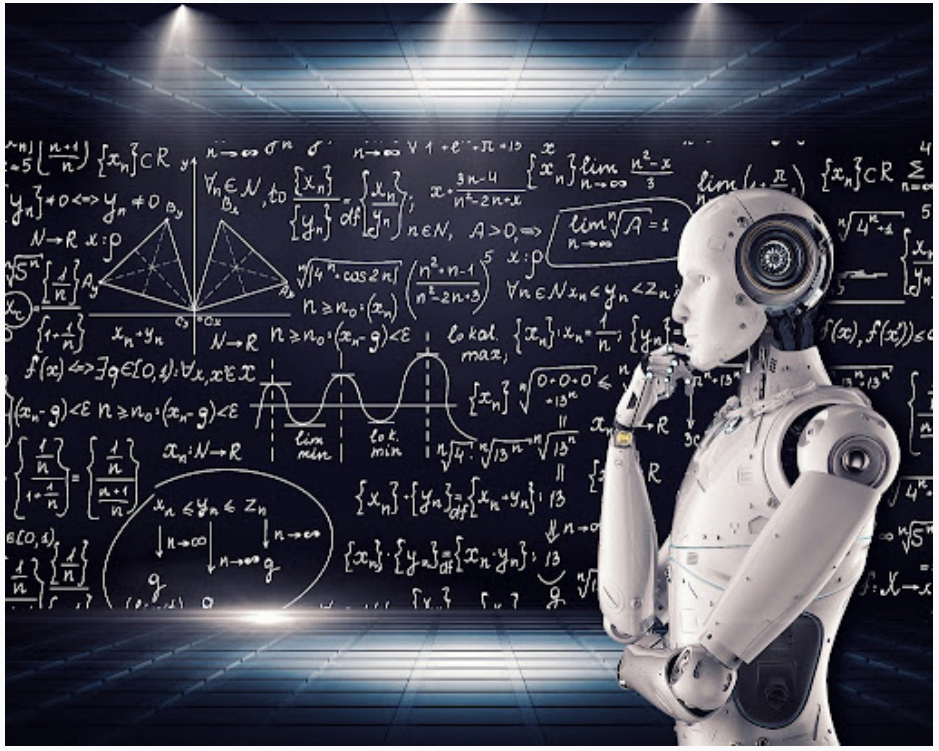
- 4608 Nodes
  - 2x22-core IBM P9 CPUs
  - 6 NVIDIA V100 GPUs
- 250 PB storage
- Mellanox interconnect
  - 200 Gb/s InfiniBand
  - Non-blocking fat-tree
- 13 MW of power

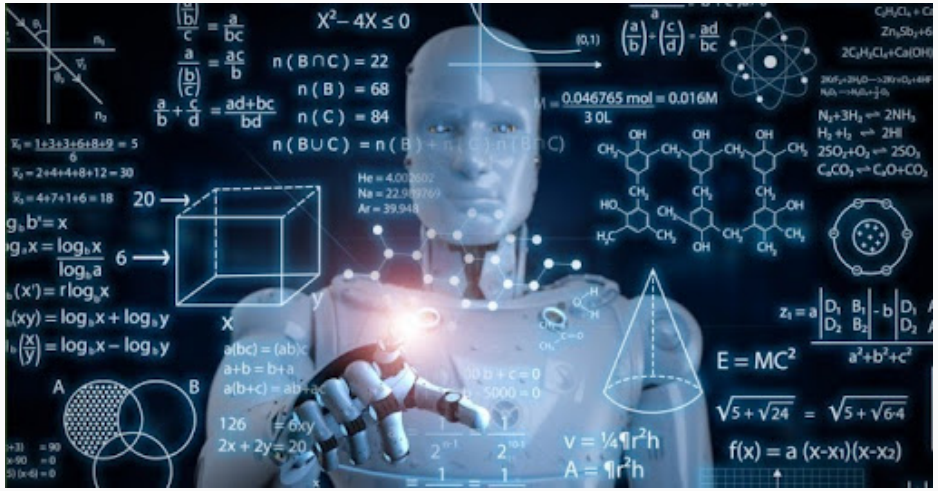
Interactive vs Batch

```
$ python
>>> 1+1

2

>>> exit()

$
```

```
$ bsub myjob.bs

<time passes>

$ cat myjob.out

2

$
```

"AI"





Some Common Terms

- **gemm** - matrix-matrix multiply
- **BLAS** - Basic Linear Algebra Subprograms; matrix library
- **FLOPS** - Floating Point Operations Per Second (adds and multiplies)
- **LINPACK** - important benchmark
- **TOP500** - list of computers ranked by LINPACK benchmark

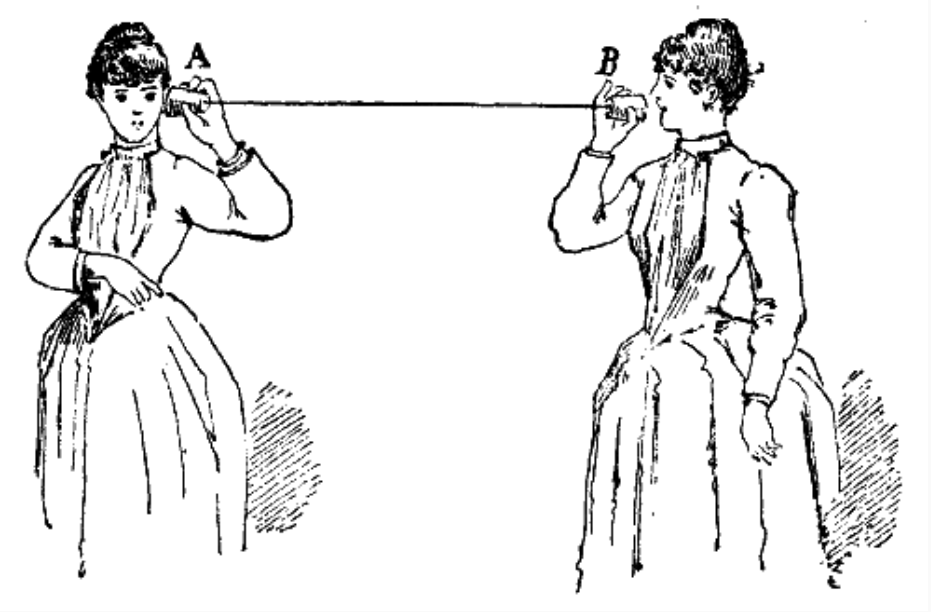
Some Notable Benchmarks

- **TOP500** - solve $Ax = b$
- **Green500** - LINPACK, sorted by energy efficiency ("FLOPS per watt")
- **Graph500** - Graph benchmark; BFS and SSSP
- **HPCG** - High Performance Conjugate Gradient; SpMV, dot product, ...

Year	Supercomputer	Peak speed (Rmax)	Location
2018	IBM Summit	122.3 PFLOPS	Oak Ridge, U.S.
2016	Sunway TaihuLight	93.01 PFLOPS	Wuxi, China
2013	NUDT Tianhe-2	33.86 PFLOPS	Guangzhou, China
2012	Cray Titan	17.59 PFLOPS	Oak Ridge, U.S.
2012	IBM Sequoia	17.17 PFLOPS	Livermore, U.S.
2011	Fujitsu K computer	10.51 PFLOPS	Kobe, Japan
2010	Tianhe-IA	2.566 PFLOPS	Tianjin, China
2009	Cray Jaguar	1.759 PFLOPS	Oak Ridge, U.S.
2008	IBM Roadrunner	1.026 PFLOPS	Los Alamos, U.S.
		1.105 PFLOPS	

MPI

- Message Passing Interface
- Distributed programming standard
- Implementations
  - OpenMPI
  - MPICH
  - MPT
  - Spectrum
- NCCL
- HPC/Data convergence



Using Video Game Hardware to Multiply Matrices

- AKA GPGPU
- Not just for video games and mining bitcoin anymore!
- Major players
  - NVIDIA
  - AMD
  - Intel…?!?!
- Pros:
  - Fast
  - When you give up, you can mine bitcoin Cons:
  - Hard to program
  - Expensive

DOE Jargon

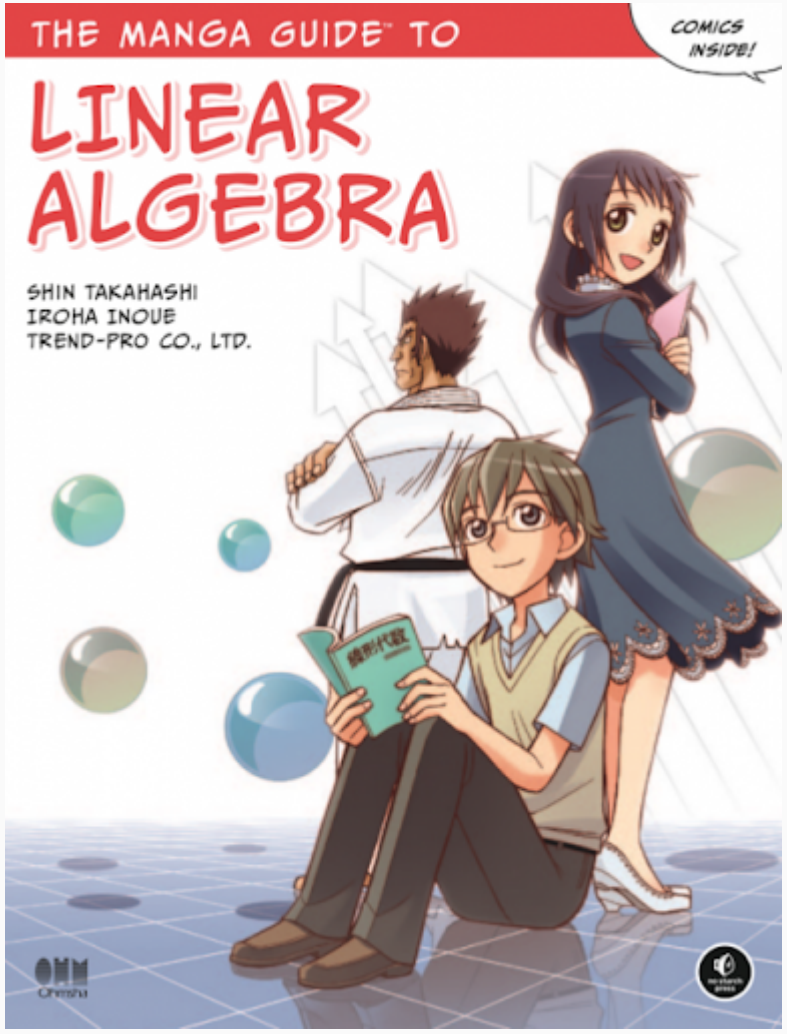
- leadership - really big jobs
- capability - big jobs
- capacity - many jobs
- Allocations
  - INCITE - leadership
  - ALCC - things DOE likes
  - DD - small jobs the center likes

Some Resources

- https://www.hpcwire.com/
- https://insidehpc.com/
- https://twitter.com/HPC_Guru

Linear Algebra

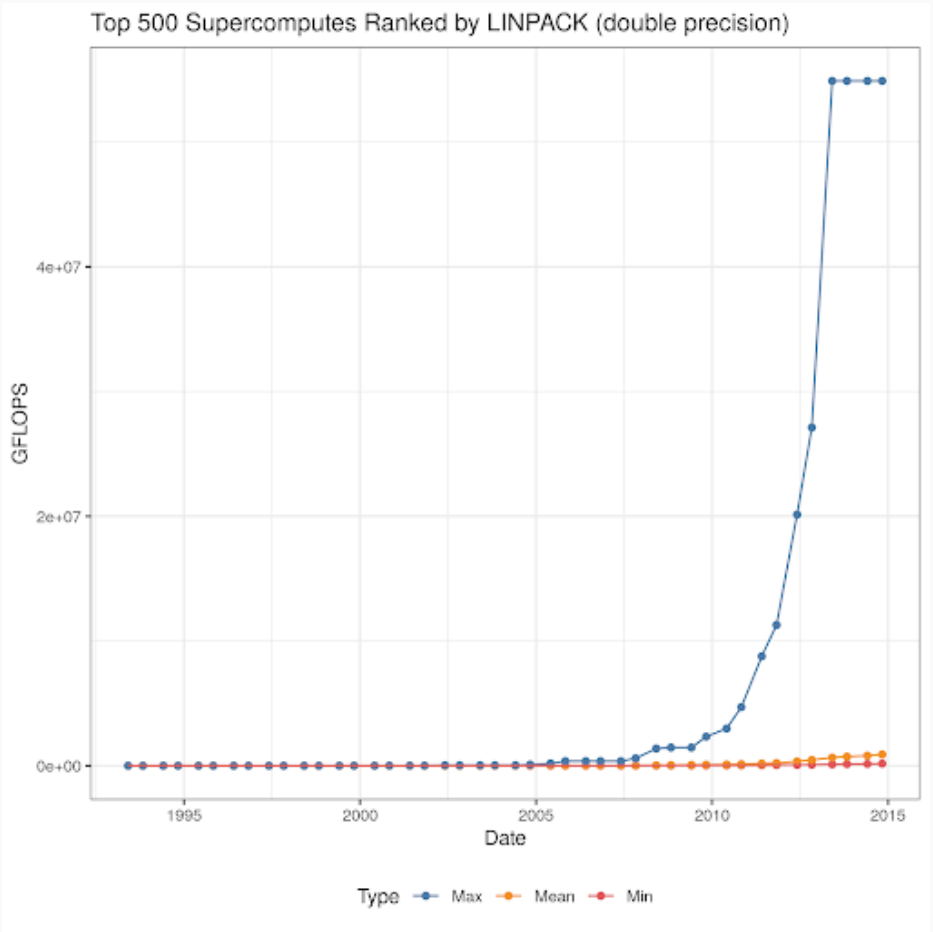
- LA dominates scientific and data computing
- Some uses in data:
  - PCA - SVD
  - Linear Models - QR
  - Covariance/correlation - gemm/syrk
  - Inverse - Cholesky, LU
- 1970's: LINPACK (not that one)
- 1980's: BLAS, LAPACK
- 1990's: ScaLAPACK
- 2000's: PLASMA, MAGMA
- 2010's: ~~DPLASMA~~ SLATE

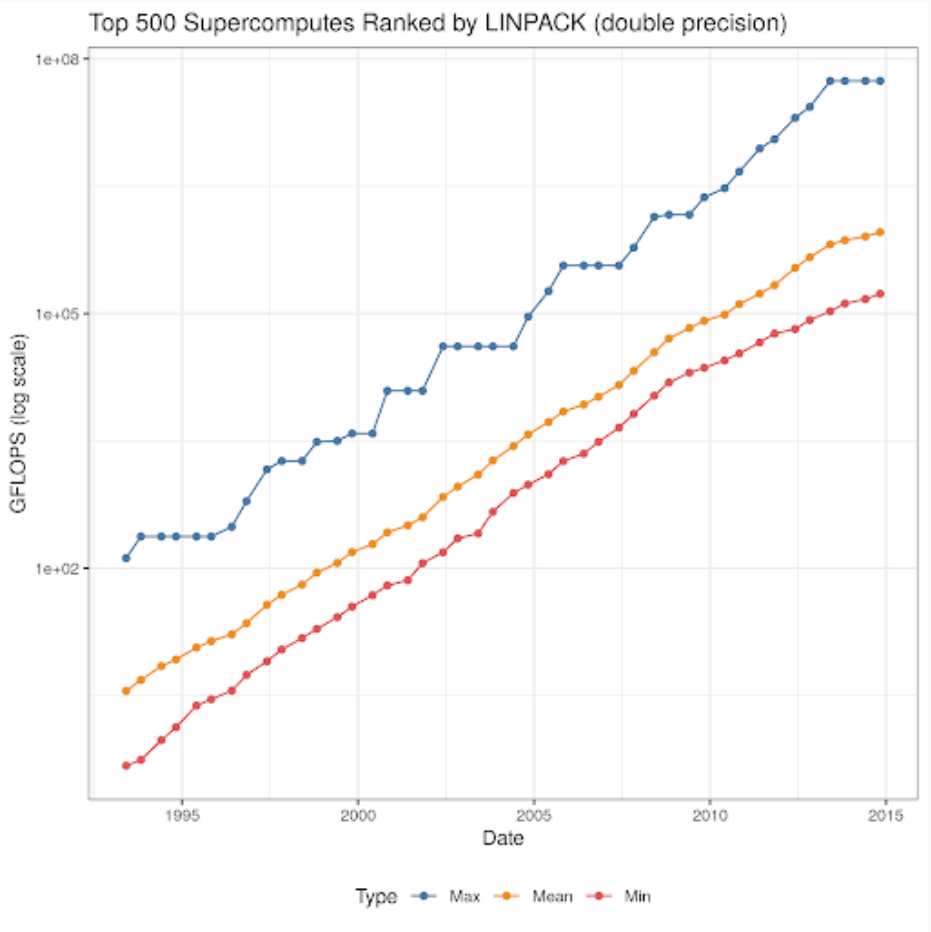


The LINPACK Benchmark

- Solve the system $Ax = b$
  - A- $n \times n$ matrix (you choose n)
  - Double precision
  - Must use LU with partial pivoting
    - $A = LU$
    - $b = Ax = LUx$
- $\frac{2}{3}n^3 + 2n^2$ operations
- Solution must satisfy some accuracy conditions.

Top 500 Rankings





LINPACK on my Desktop

CPU

- double
  - Best n=39000 t=200.030
  - Theoretical Peak 217 GFLOPS
  - LINPACK 197.715 GFLOPS
- float
  - Best n=45000 t=156.066
  - Theoretical Peak 434 GFLOPS
  - LINPACK 389.285 GFLOPS

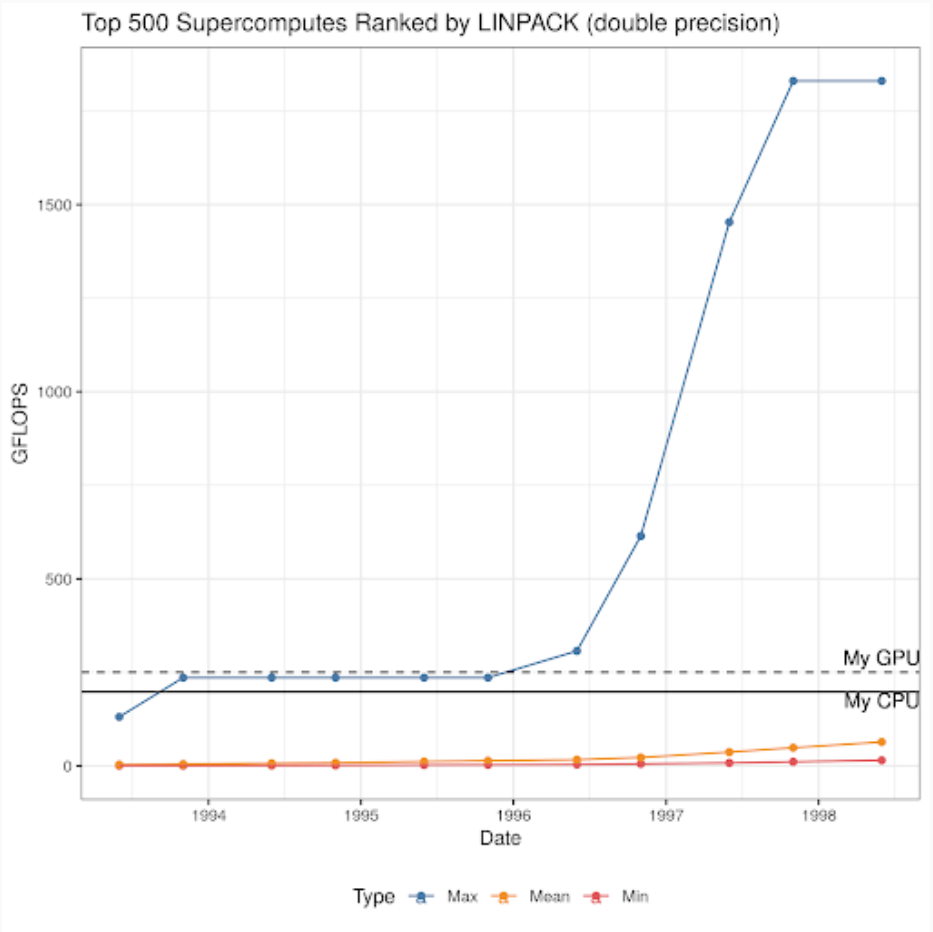
GPU

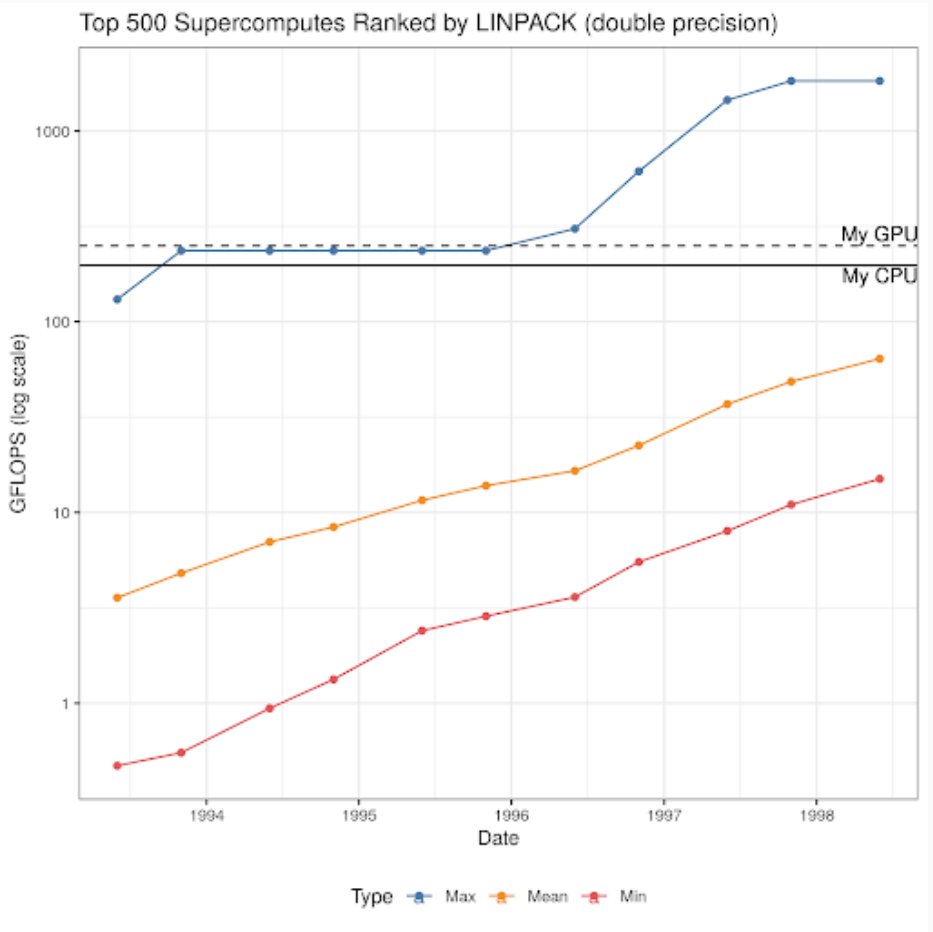
- double n=28000 t=58.444 GFLOPS=250.433
- float n=36000 t=4126.614 GFLOPS=4126.614

LINPACK on my Desktop

How Do We Rank?

How Do We Rank?





The Cloud

History

- 1960's: Mainframe time sharing
- 1970-80's: Dumb terminals
- 1990's: Virtualization invented
- 2006: AWS launched
- 2015: Docker created

We will discuss virtualization and Docker in depth in a later lesson.

So What Even Is The Cloud?

Someone else's computer?

Something else?

ZDNet Menu US

Stop saying the cloud is just someone else's computer - because it's not

Cloud and 'as a service' models have changed the way computing works. Get over it.

Written by **Mary Branscombe**,
Freelance blogger

Posted in **500 words into the future** on July 12, 2016 | Topic: **Cloud**

SERVICES, SERVICES, SERVICES!

Compute
Amazon EC2
Amazon Elastic Container Service
Amazon Elastic Container Service for Kubernetes
Amazon Elastic Container Registry
Amazon Lightsail
AWS Batch
AWS Elastic Beanstalk
AWS Fargate
AWS Lambda
AWS Serverless Application Repository
Auto Scaling
Elastic Load Balancing
VMware Cloud on AWS

Storage
Amazon Simple Storage Service (S3)
Amazon Elastic Block Storage (EBS)
Amazon Elastic File System (EFS)
Amazon Glacier
AWS Storage Gateway
AWS Snowball
AWS Snowball Edge
AWS Snowmobile

Database
Amazon Aurora
Amazon RDS
Amazon DynamoDB

Networking & Content Delivery
Amazon VPC
Amazon CloudFront
Amazon Route 53
Amazon API Gateway
AWS Direct Connect
Elastic Load Balancing

Developer Tools
AWS CodeStar
AWS CodeCommit
AWS CodeBuild
AWS CodeDeploy
AWS CodePipeline
AWS Cloud9
AWS X-Ray
AWS Tools & SDKs

Management Tools
Amazon CloudWatch
AWS CloudFormation
AWS CloudTrail
AWS Config
AWS OpsWorks
AWS Service Catalog
AWS Systems Manager
AWS Trusted Advisor
AWS Personal Health Dashboard
AWS Command Line Interface
AWS Management Console

Machine Learning
Amazon SageMaker
Amazon Comprehend
Amazon Lex
Amazon Polly
Amazon Rekognition
Amazon Machine Learning
Amazon Translate
Amazon Transcribe
AWS DeepLens
AWS Deep Learning AMIs
Apache MXNet on AWS
TensorFlow on AWS

Analytics
Amazon Athena
Amazon EMR
Amazon CloudSearch
Amazon Elasticsearch Service
Amazon Kinesis
Amazon Redshift
Amazon QuickSight
AWS Data Pipeline
AWS Glue

Security, Identity & Compliance
AWS Identity and Access Management (IAM)
Amazon Cloud Directory
Amazon Cognito

AR & VR
Amazon Sumerian

Application Integration
Amazon MQ
Amazon Simple Queue Service (SQS)
Amazon Simple Notification Service (SNS)
AWS AppSync
AWS Step Functions

Customer Engagement
Amazon Connect
Amazon Pinpoint
Amazon Simple Email Service (SES)

Business Productivity
Alexa for Business
Amazon Chime
Amazon WorkDocs
Amazon WorkMail

Desktop & App Streaming
Amazon WorkSpaces
Amazon AppStream 2.0

Internet of Things
AWS IoT Core
Amazon FreeRTOS
AWS Greengrass
AWS IoT 1-Click

SERVICES, SERVICES, SERVICES!

- Think of a sequence of letters and numbers
- It's probably an Amazon product
- (Ungraded) Homework:
  - Find a random AWS service
  - Try to understand what it does
  - (Bonus) No crying

In This Class

We will restrict attention to EC2 ("someone else's computer")

HPC vs The Cloud

How Are They Similar?

- Somebody else's computer
- Lots of jargon
- Entire career paths
- Can be hard to use
- Tech stacks more similar than most realize

How Are They Different?

- Privileges (user vs root)
- Salaries, job growth/potential, etc
- Academia vs Industry
- "Free" (taxes) vs you pay
- Some non-intersecting tech stacks
  - cloud: HDFS, databases, web, ...
  - HPC: HDF5, binary files, Fortran, ...

Other Compute Models

- Your computer
- "The office quasi-cluster"
- Edge computing

Next Time

- Computing on remote systems
  - ~30 minutes lecture
  - live tutorial component on ISAAC and/or AWS
- Assignment 1 due tonight
- No new assignments

Questions?