Anecdotally: Startups with $10^9$ events/day.
Old techniques don’t work

Computers aren’t keeping up

Linear work is unavoidable but even linear time often inadequate.
Core Issues

Efficiency: More efficient algorithms matter.
Core Issues

Efficiency: More efficient algorithms matter.

Parallelism: More data than any machine can handle.

Distribution: Data can’t move.
Core Issues

Efficiency: More efficient algorithms matter.

Parallelism: More data than any machine can handle.

Distribution: Data can’t move.

In the beginning, anyone could create an email server (and I did).
Core Issues

Efficiency: More efficient algorithms matter.

Parallelism: More data than any machine can handle.

Distribution: Data can’t move.

In the beginning, anyone could create an email server (and I did).

Spam became a business model. Individual email servers are typically overwhelmed. Large spam filters at centralized email providers—

How do we **Efficiently** learn to classify Spam in a **Parallel Distributed** environment?
3 Years ago: Samy Bengio, Corinna Cortes, Dennis DeCoste, Francois Fleuret, Ramesh Natarajan, Edwin Pednault, Dan Pelleg, Elad Yom-Tov **Efficient Machine Learning - Overcoming Computational Bottlenecks in Machine Learning**
3 Years ago: Samy Bengio, Corinna Cortes, Dennis DeCoste, Francois Fleuret, Ramesh Natarajan, Edwin Pednault, Dan Pelleg, Elad Yom-Tov *Efficient Machine Learning - Overcoming Computational Bottlenecks in Machine Learning*

2 Years ago: Robert Thibadeau, Dan Hammerstrom, David Touretzky, Tom Mitchell *Parallel Implementations of Learning Algorithms: What have you done for me lately?*
The Wider Context

3 Years ago: Samy Bengio, Corinna Cortes, Dennis DeCoste, Francois Fleuret, Ramesh Natarajan, Edwin Pednault, Dan Pelleg, Elad Yom-Tov Efficient Machine Learning - Overcoming Computational Bottlenecks in Machine Learning

2 Years ago: Robert Thibadeau, Dan Hammerstrom, David Touretzky, Tom Mitchell Parallel Implementations of Learning Algorithms: What have you done for me lately?

1 Year ago: Alex Gray, Alexander Smola, Arthur Gretton, Joseph Gonzalez, Carlos Guestrin Large-Scale Machine Learning: Parallelism and Massive Datasets
The Wider Context

3 Years ago: Samy Bengio, Corinna Cortes, Dennis DeCoste, Francois Fleuret, Ramesh Natarajan, Edwin Pednault, Dan Pelleg, Elad Yom-Tov *Efficient Machine Learning - Overcoming Computational Bottlenecks in Machine Learning*

2 Years ago: Robert Thibadeau, Dan Hammerstrom, David Touretzky, Tom Mitchell *Parallel Implementations of Learning Algorithms: What have you done for me lately?*

1 Year ago: Alex Gray, Alexander Smola, Arthur Gretton, Joseph Gonzalez, Carlos Guestrin *Large-Scale Machine Learning: Parallelism and Massive Datasets*

Us.
The Wider Context

3 Years ago: Samy Bengio, Corinna Cortes, Dennis DeCoste, Francois Fleuret, Ramesh Natarajan, Edwin Pednault, Dan Pelleg, Elad Yom-Tov *Efficient Machine Learning - Overcoming Computational Bottlenecks in Machine Learning*

2 Years ago: Robert Thibadeau, Dan Hammerstrom, David Touretzky, Tom Mitchell *Parallel Implementations of Learning Algorithms: What have you done for me lately?*

1 Year ago: Alex Gray, Alexander Smola, Arthur Gretton, Joseph Gonzalez, Carlos Guestrin *Large-Scale Machine Learning: Parallelism and Massive Datasets*

Us.

Next Year: Large Scale Learning Book with 20+ chapters edited by Ron Bekkerman, Misha Bilenko, & me.
What’s in the book?

Parallel Unsupervised Learning Methods

1. Information-Theoretic Co-Clustering with MPI

2. Spectral Clustering MapReduced

3. K-Means with GPU

4. Latent Dirichlet Analysis with MPI

It’s very hard to compare different results.
... But let's try

![Graph showing Speed per method for various methods and datasets.](image-url)
Ginormous caveat: Prediction performance varies wildly depending on the problem–method pair.
Ground Rules

Ginormous caveat: Prediction performance varies wildly depending on the problem–method pair.

The standard: Input complexity/time.
Ginormous caveat: Prediction performance varies wildly depending on the problem–method pair.

The standard: Input complexity/time.

⇒ No credit for creating complexity then reducing it. (Ouch!)
Ground Rules

Ginormous caveat: Prediction performance varies wildly depending on the problem–method pair.

The standard: Input complexity/time.

⇒ No credit for creating complexity then reducing it. (Ouch!)

Most interesting results reported. Some cases require creative best-effort summary.
Supervised Testing (but not training)

Features/s

Speech-WFST

Inference

MPI-40

Conv. NN

Boosted DT

Conv NN

Caltech

GPU

NIST

FPGA

Images

Asic(sim)

Caltech

Speed per method

parallel

single
Many others defy summarization. Highlights:

1. Feature selection & frequent item systems.
2. Chapters on new parallel computing frameworks of plausible interest to ML people.
The Morning

7:00  Poster setup
7:30  Langford—Intro
8:00  Tsitsiklis—Averaging algorithms and distributed optimization
9:00  Coffee & Posters
9:20  Xiao—Optimal Distributed Online Prediction Using Mini-Batches
9:45  Petrov—MapReduce/Bigtable for Distributed Optimization
10:10 Minitalks
10:30 Posters & Break
The Afternoon

2:00 **Unofficial** Vowpal Wabbit Tutorial
3:30 Guestrin—Machine Learning in the Cloud with GraphLab
4:30 Singh—Distributed MAP Inference for Undirected Graphical Models
4:55 Posters & Break
5:15 Ye—Gradient Boosted Decision Trees on Hadoop
5:40 More Minitalks
6:00 Summary/Panel/Discussant
6:30 Posters & talking
Have Fun!