Cross Device Ad Targeting at Scale

jerry ye

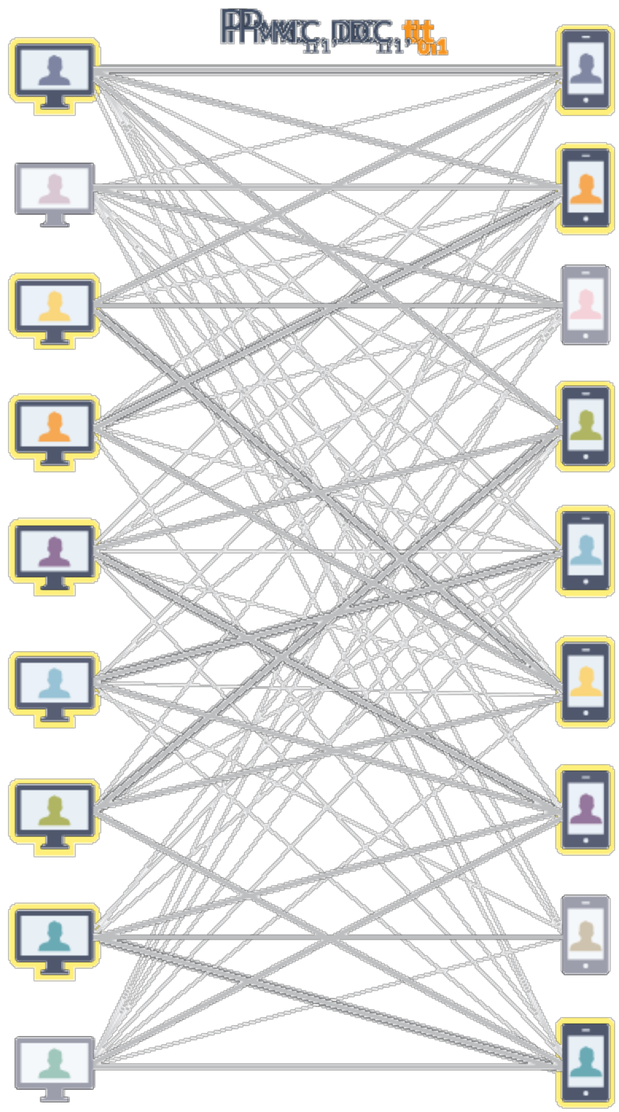
drawbridge

Agenda

- Cross Device Ad Targeting
  - Overview
  - Infrastructure
- Ad CTR, CVR, Revenue prediction
- Data vs. Algorithm
- Machine learning on Hadoop
  - MapReduce
  - MPI
- Case Study: GBDT
- Large scale machine learning at Drawbridge
- Vowpal Wabbit
- Results

Sequoia Capital, Kleiner Perkins, and Northgate Capital Funded
$20.5MM raised
Demand Side Partner for Cross Device Ad Targeting
Over 475MM devices probabilistically matched
Non Personally Identifiable Information
Reach users on any device

Matching Devices

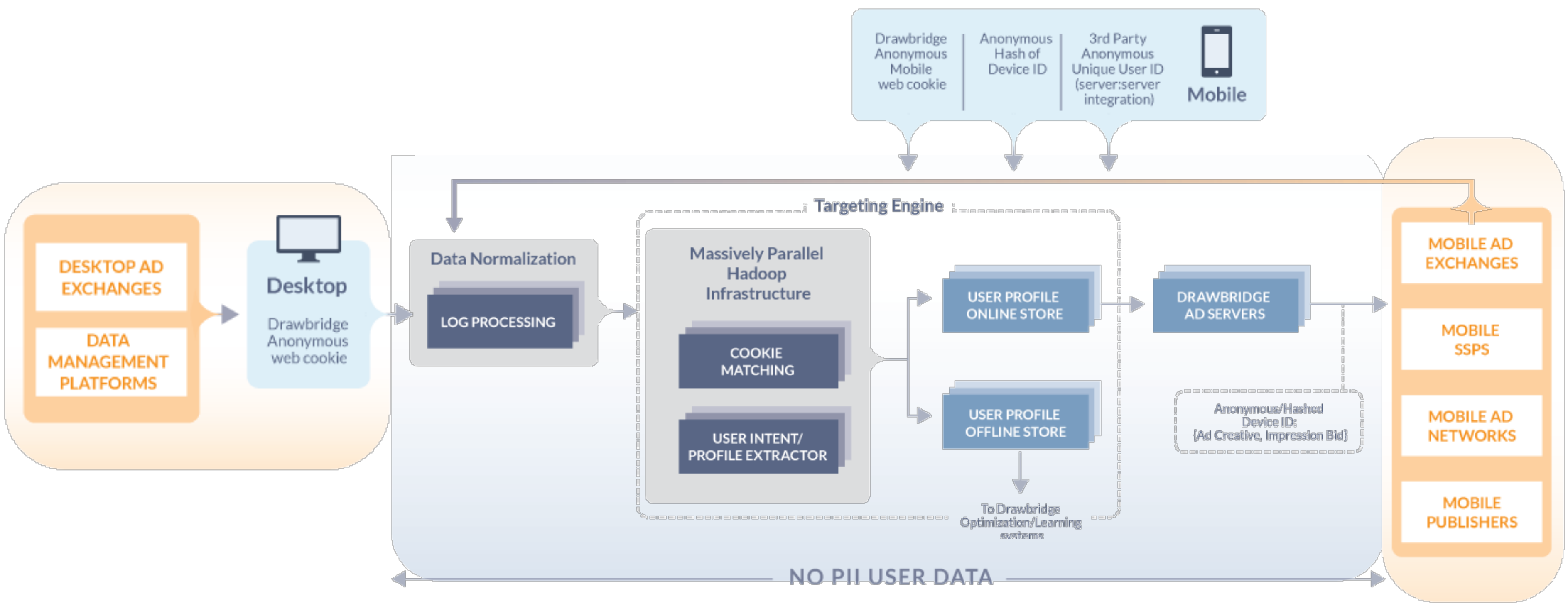


User Profile

After bridging devices, generate user profile:

- Non-personally identifiable information
- Demographic
- Location
- Interests
- Retargeting
- Bridged devices

Drawbridge Infrastructure



Supply and Demand

Drawbridge is a Demand Side Partner

Supply comes from RTB exchanges as well as direct partnerships

Ability to target desktop segments on mobile and vice versa

Given a request, within milliseconds:

- select campaign
- price
- place bid

Optimize for

- Client ROI
- Revenue/Profit

Prediction

In the case of revenue, given a request, predict:

Cost Per Click:

Predicted Revenue = pctr*cpc

Cost Per Acquisition:

Predicted Revenue = pctr*pcvr*cpa

Given revenue, pick bidding strategy

Data

Months of historic ad performance data

- Click/Conversion

Per impression data

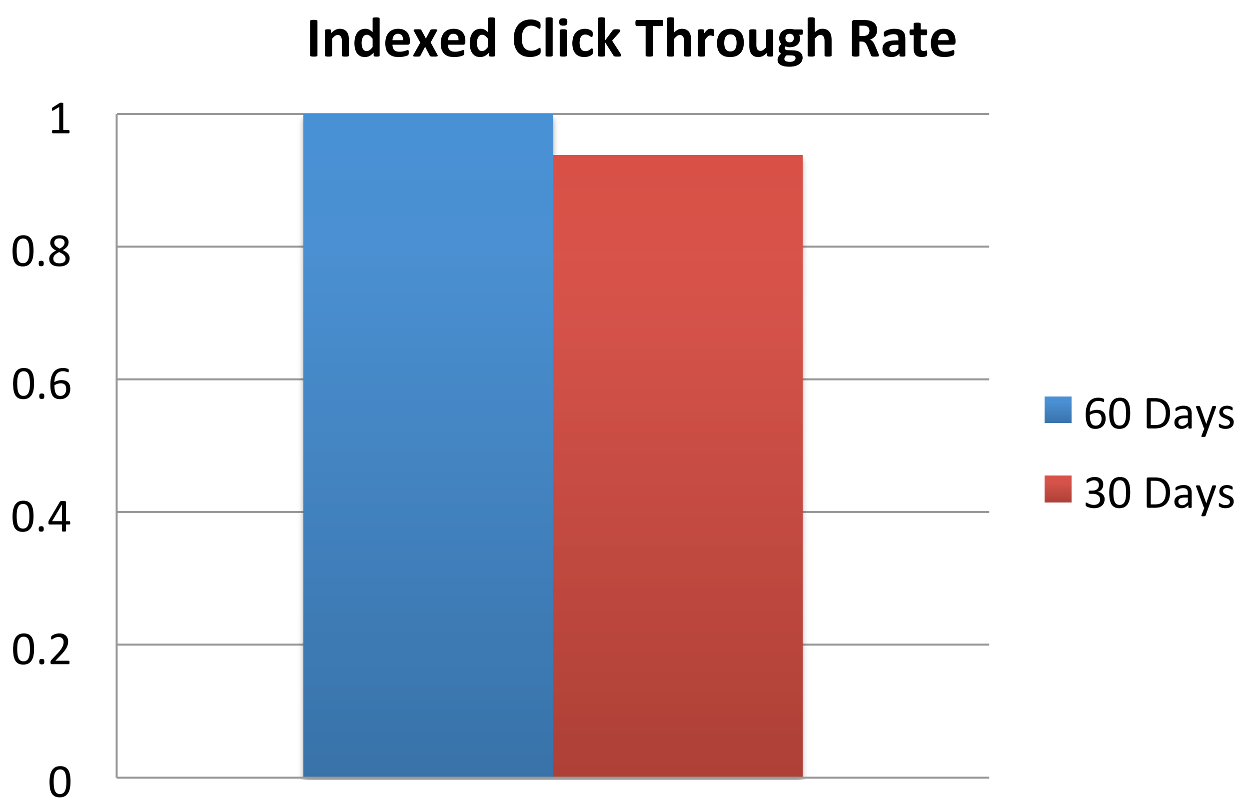
- Temporal
- Contextual
- Anonymized User Features

Training Data

- 1.5 TB
- 1MM dimensions
- 2B samples
- Unsampled, trained on everything

Advantage of More Data

drawbridge



6.6% Increase in CTR from doubling data

Data versus Algorithm

More data?

Better algorithm?

Why not both?

Difficult to scale ML

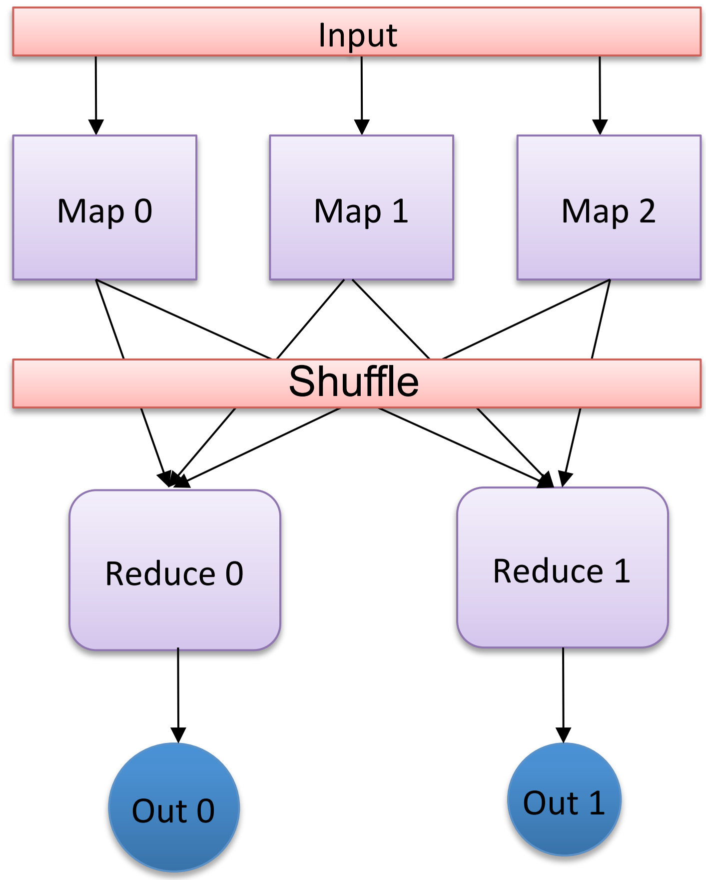
Most use R, Matlab, Weka

Hadoop?

MapReduce

drawbridge

- Hadoop is an open source implementation of MapReduce
- MapReduce designed for data intensive applications
- Relatively simple to use
- Wide range of applications
- Streaming or MapReduce
- No means of communication
- Great for data prep



Scaling Up Machine Learning

Most ML is iterative (gradient descent)

Optimizes for a set of parameters over the same data

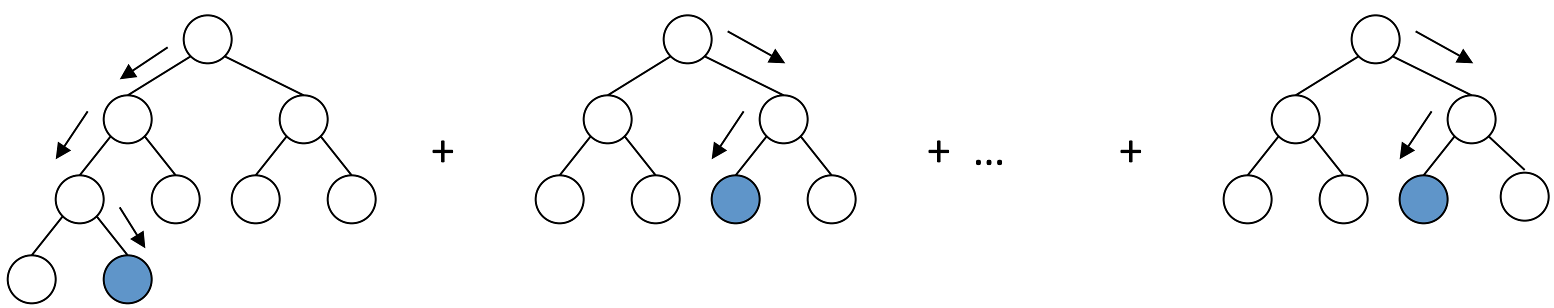
Updated parameters are usually aggregated

Naïve ML implementations in MapReduce do not work

Case Study: Gradient Boosted Decision Trees at Yahoo!

Gradient Boosted Decision Trees

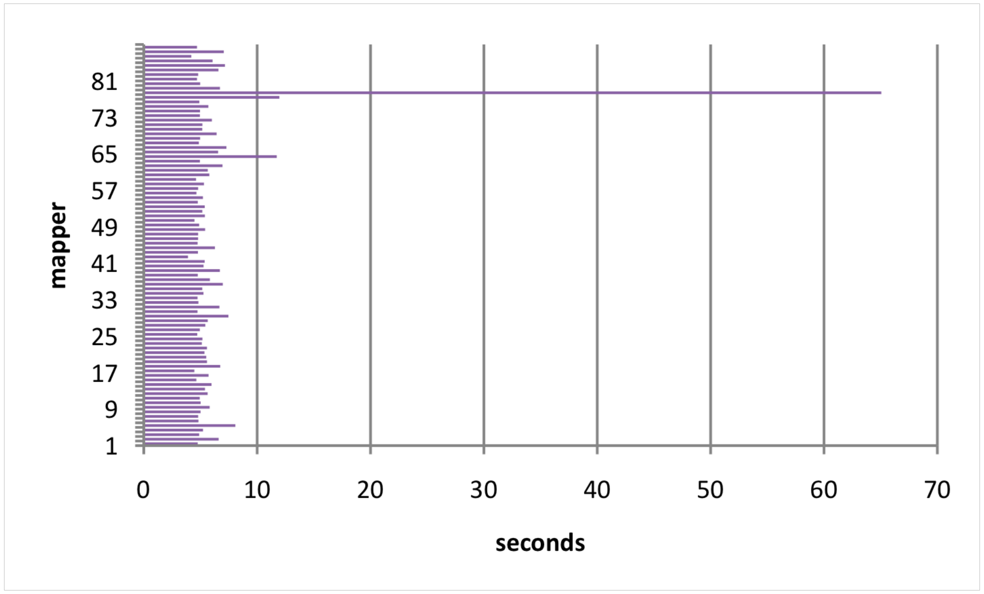
- Gradient Boosted Decision Trees was introduced by Jerome Friedman in 1999
- An additive regression model over an ensemble of trees, fitted to current residuals, gradients of the loss function, in a forward step-wise manner
- Favors many shallow trees (e.g., 6 nodes, 2000 trees)
- Numerous applications within Yahoo!
- Blender in Bellkor's winning Netflix solution



Motivation

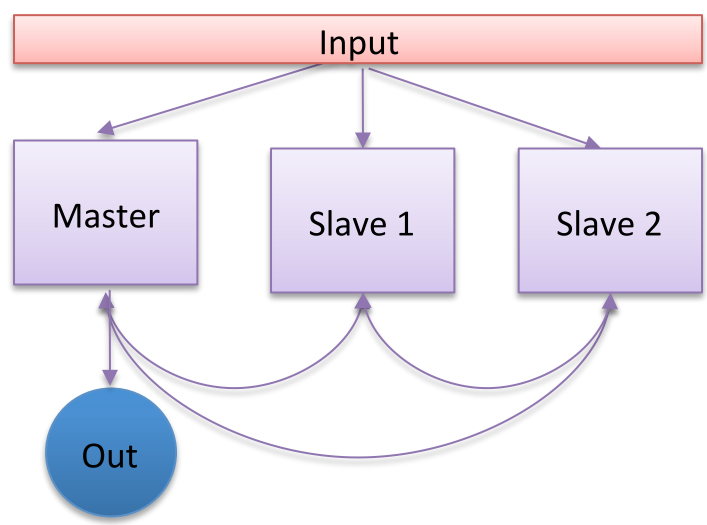
drawbridge

- High memory tasks
- HDFS lag, experience with GBDT
- Certain tasks would get performance bump if written with MPI
- Yahoo! has invested heavily in Hadoop
- Multiple research and production clusters with thousands of machines each
- Expensive to deploy MPI specific clusters
- Utilize existing Hadoop clusters for MPI jobs



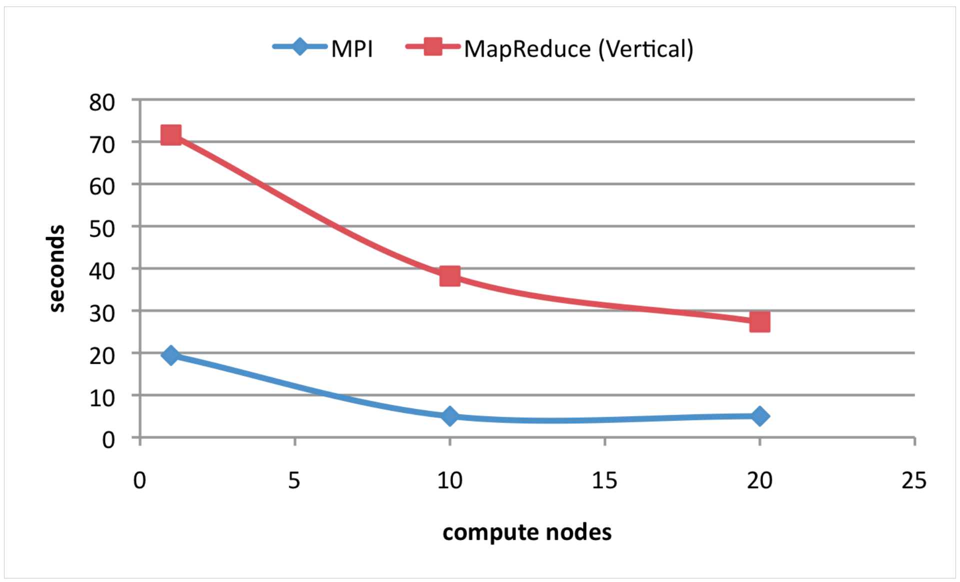
Message Passing Interface

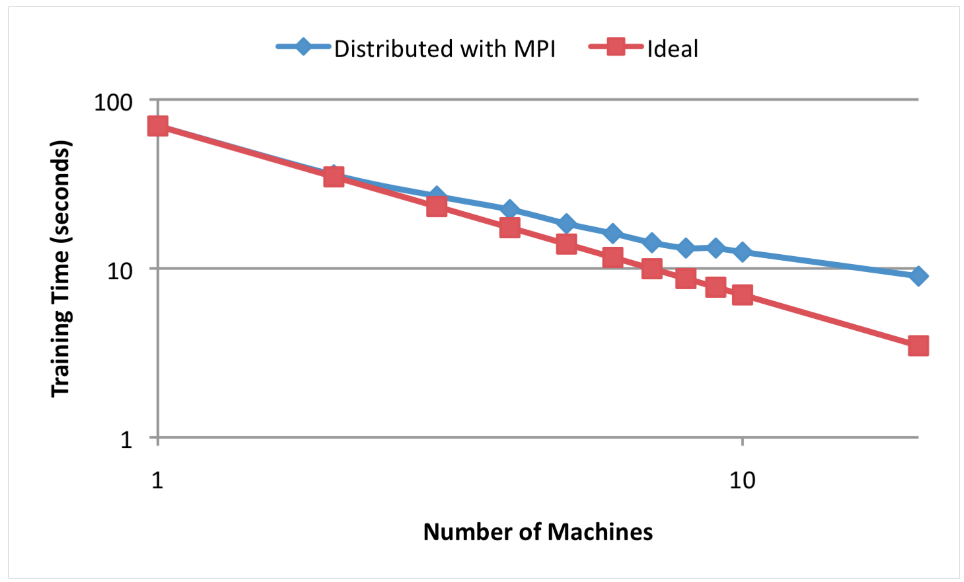
- Message Passing Interface (MPI) allows many computers to communicate with each other.
- Dominant model in high performance computing
- Scalable, portable
- Distributed shared memory for high RAM jobs
- OpenMPI is an open source implementation of MPI
- Low level and can be complicated to use
- Modified OpenMPI to run on Hadoop
- Fault tolerance



Speedup

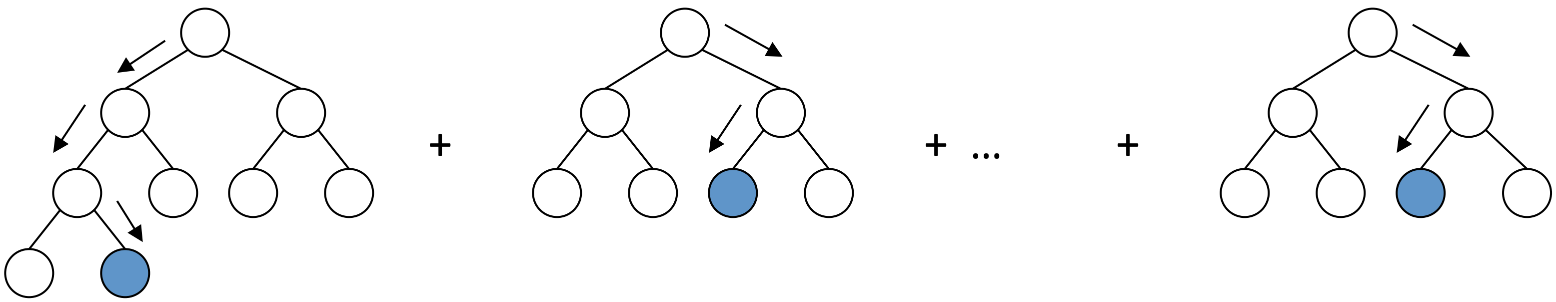
- MPI implementation of GBDT magnitudes faster than horizontal MapReduce implementation
- MPI implementation 6X faster than vertical MapReduce implementation
- Communications cost dramatically reduced
- MPI close to ideal initially, but communications not free
- MapReduce vs. MPI implementation on left
- Scalability of MPI implementation on right





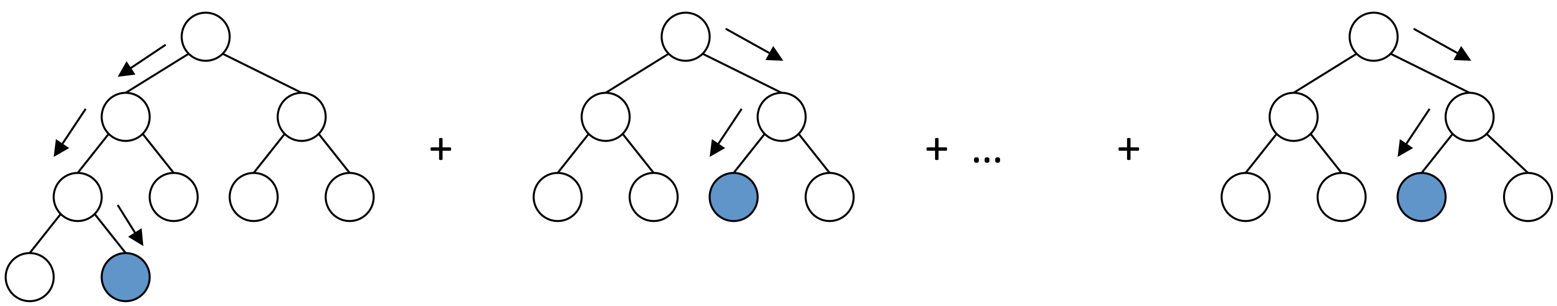
GBDT Improvement

MapReduce



Horizontal: 211 minutes x 2500 trees = 366 days x 100 machines
Vertical: 28 seconds x 2500 trees = 19.4 hours x 20 machines

MPI



5 seconds x 2500 trees = 3.4 hours x 10 machines

1800% less node hours!

Machine Learning @ Drawbridge

- Pig, Hive, Hbase
- Clean data, abuse detection
- Missing values
- Normalization
- Binning
- Formatting
- Vowpal Wabbit

Vowpal Wabbit

- John Langford at Microsoft Research (previously Yahoo! Labs)
- http://hunch.net/~vw
- State of the art in scalable, fast, efficient machine learning. Magnitudes faster and more scalable than any MapReduce based learner
- VW has been shown to reliably scale to 1000+ machines.
- AllReduce implementation similar to MPI's
- Runs on Hadoop
- @Drawbridge, we predict ctr, cvr, win rate

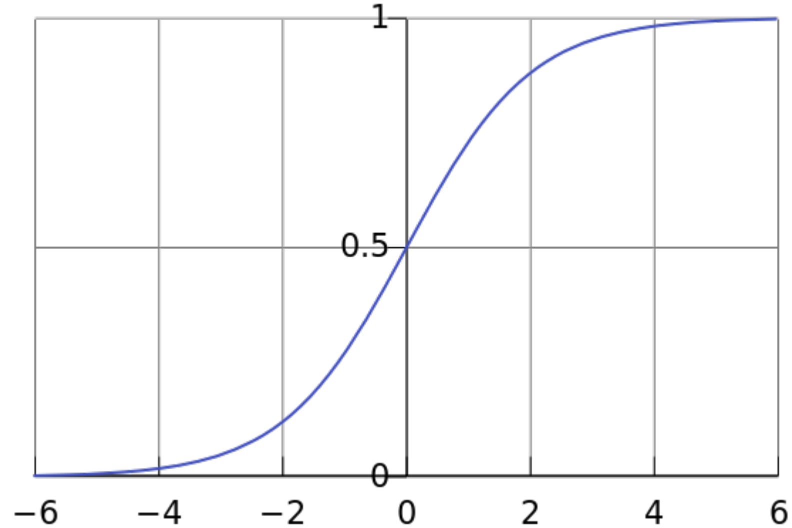
Logistic Regression

drawbridge

- Predict probability of click/conversion [0,1]
- Solve for θ

$$f(x) = \frac{1}{1+e^{-z}}$$

$$z = \beta + \sum_{i=0}^{n} \theta_i x_i$$



What is Theta?

drawbridge

- Θ are weights for our features
- How important is
  - User frequency cap?
  - Time of day?
  - Device?
  - Etc

$$z(x) = \beta + \theta_0(\text{user frequency count}) + \theta_1(Sunday) + \theta_2(\text{iPhone})$$

A/B Testing

- Logistic Regression vs. Method of Moments
- Bucket tests on percentage of traffic
- Several models released since July of last year
- 10-30% improvement in CTR, CVR each time
- Lower cost, higher revenue

Conclusions

drawbridge

- Cross Device Ad Targeting
- Ad CTR, CVR, Revenue prediction
- Data AND algorithm
- Machine learning on Hadoop
  - MapReduce
  - MPI
  - GBDT much faster in MPI
- Large scale machine learning at Drawbridge
- Vowpal Wabbit