Towards a Unified Architecture for in-RDBMS Analytics

Shivaram Venkataraman

CS 294: Big Data Systems Research

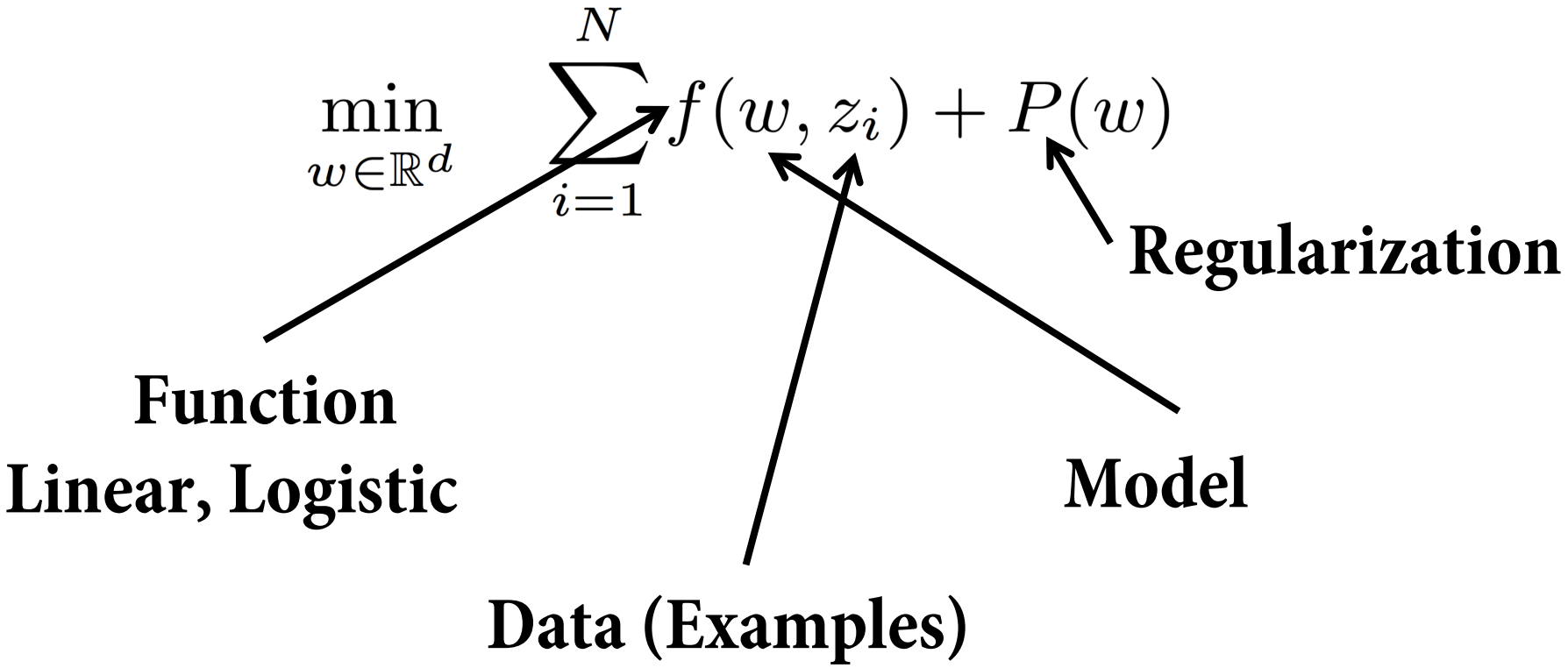
Motivation

Advanced Analytics

Classification

Recommendation

Convex Optimization



Convex Optimization

$$\min_{w \in \mathbb{R}^d} \quad \sum_{i=1}^{N} f(w, z_i) + P(w)$$

What is convex ?
Linear Regression, Linear SVM
Kernel SVMs, Logistic Regression,

What is not convex ?
Graph mining, Deep Learning

Gradient Descent

$$w^{(k+1)} = w^{(k)} - \alpha_k \nabla f(w^{(k)})$$

Initialize w

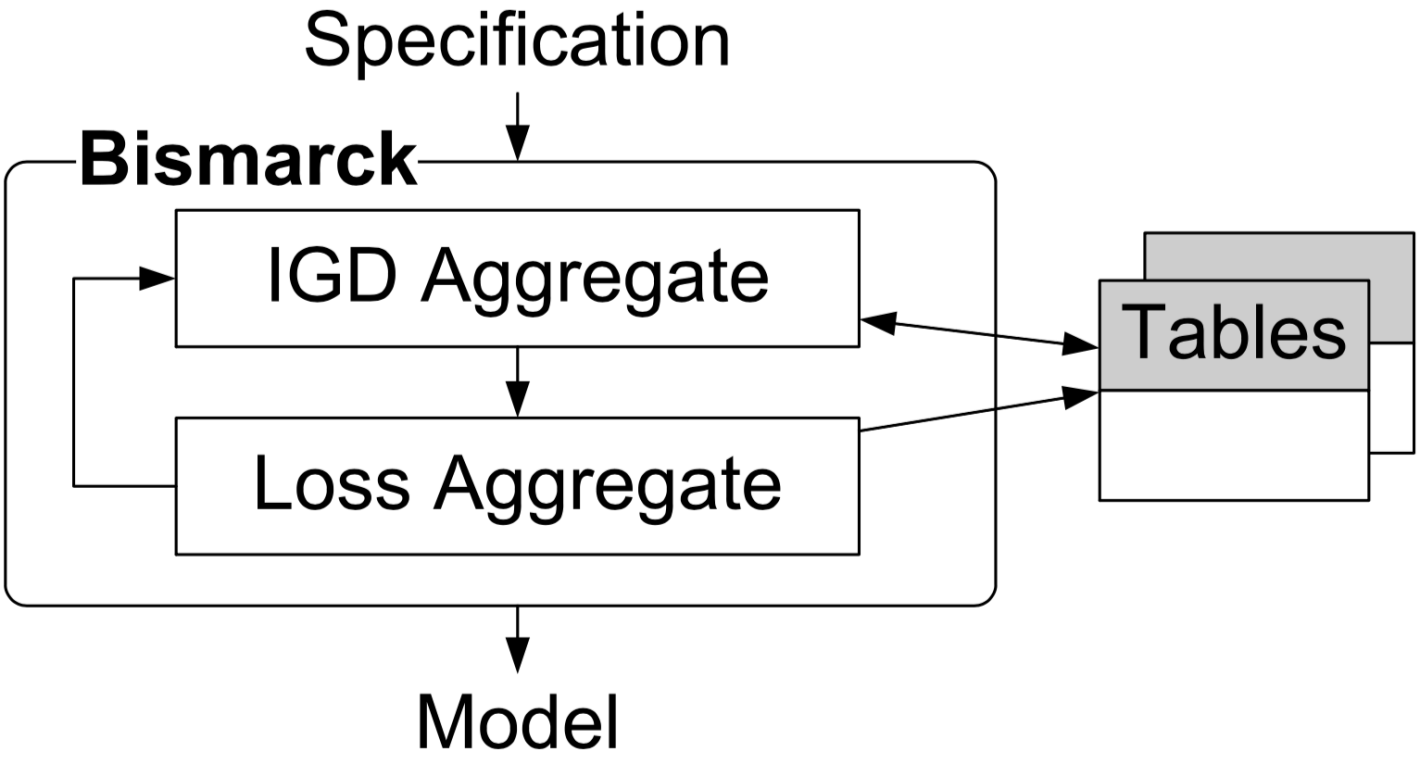
For many iterations:

Compute Gradient

Update model

End

Bismarck Architecture



Spark MLlib

Gradient

```
(data, weights) -> gradient
```

Updater

```
(weights,grad)->newWeights
```

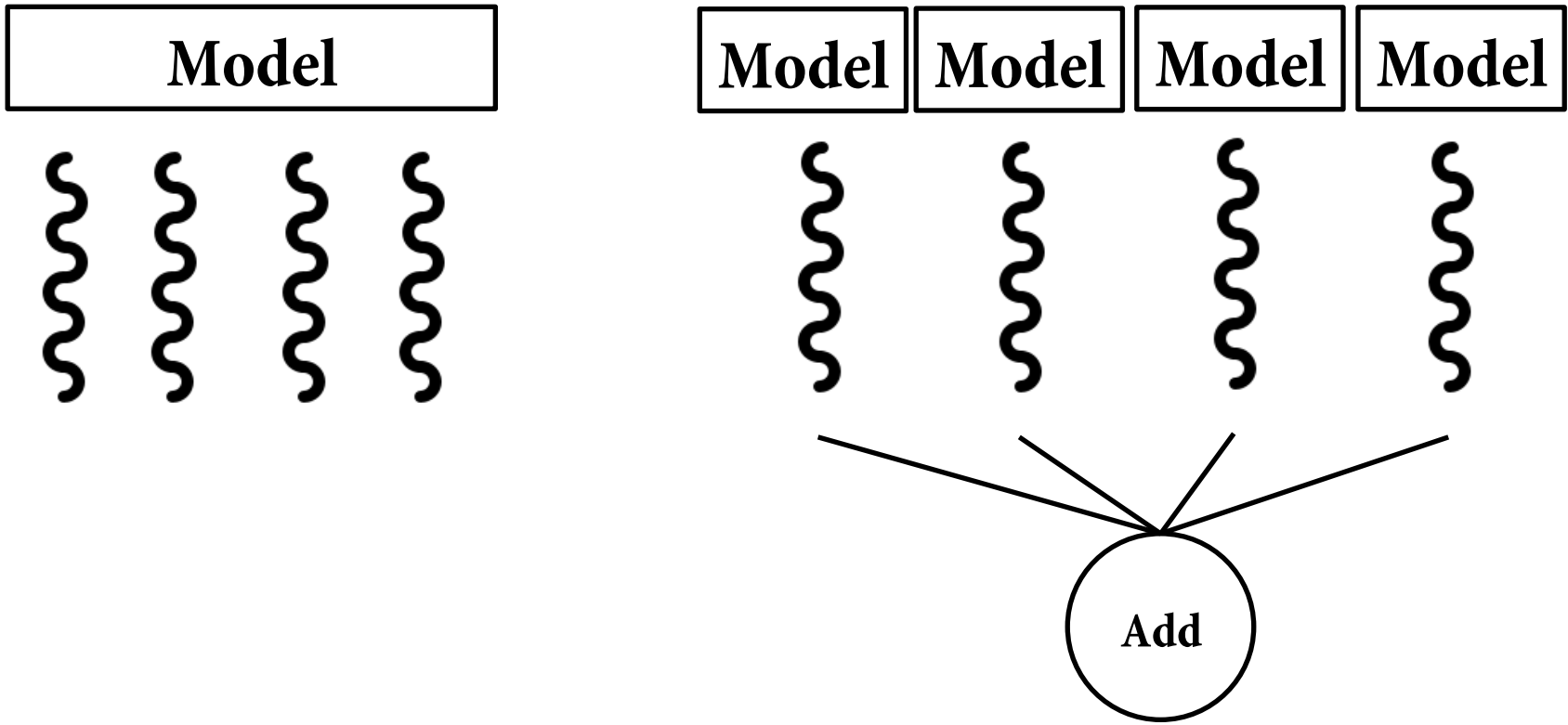
Gradient Descent

```
for (i <- 1 to ITERS) {
  val gradSum = data.sample(fraction).map { x =>
   gradient.compute(x, weights)
  }.reduce(_ + _)

  weights = updater.compute(weights, gradSum)
}
```

Discussion

(1) Shared Memory vs. Shared Nothing



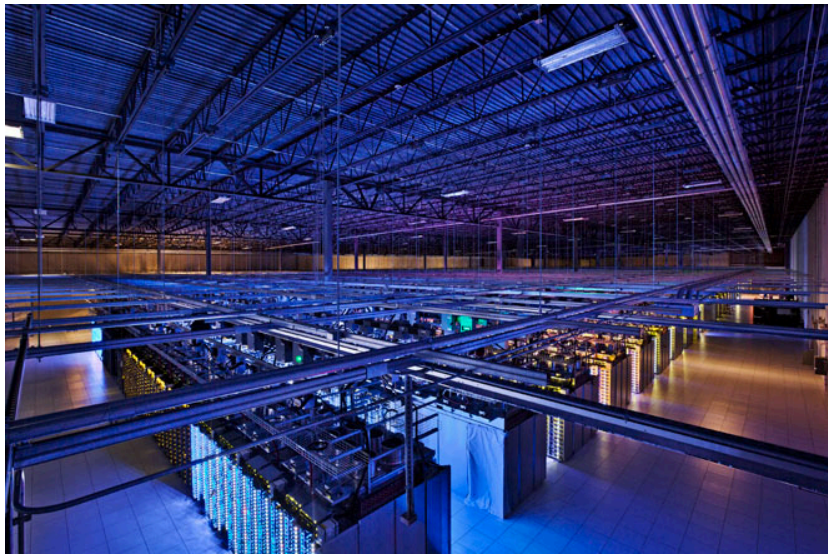
Bismarck: Shared memory beats Shared Nothing

(1) Shared Memory vs. Shared Nothing

Better Algorithms	Better Systems
AdaGrad	Parameter Server
Dual Averaging	AllReduce (VW)
…	…

Fundamental trade-off between delay & convergence

(2) Single disk vs. Cluster



Compute Bandwidth	4-8 GB/s ? (PCIe)	60 GB/s * N
Update Latency (1MB)	30 microseconds	800 microseconds * N (10 Gbps)

Colin Scott http://www.eecs.berkeley.edu/~rcs/research/interactive_latency.html

(2) Single disk vs. Cluster

Example: 1M data points, 128k features, 1 class

Data size: 1TB

Cluster: 8 machines ~ 128GB / machine

Single disk: 8 SSDs

	Single Disk	Cluster
1 Epoch Gradient	128s	2.5s
Updates	30 microseconds	6.4 milliseconds

(2) Single disk vs. Cluster

More gradient computations

vs.

Frequent Updates

Depends on

- **Cost of gradient computation**
- **Model Size**
- **Data characteristics (condition number)**

(3) IGM vs. L-BFGS vs. SDCA etc.

One architecture to rule them all ?

Momentum, L-BFGS – Need for model history

Block Coordinate – Residuals (data size)

Dual Coordinate Ascent – Store dual coefficients

…

(4) Efficient Sampling

Single node

- **Shuffle once (Bismarck)**
- **Shuffle every epoch**

Cluster

- **Bernoulli Sample per-batch**
  - **Need to look at entire data**
- **Shuffle + Sample blocks (KMN)**