Strata+ Hadoop WORLD

PRESENTED BY

O'REILLY cloudera

ADATAO

DATA INTELLIGENCE FOR ALL

Distributed Deep Learning on Spark & Tachyon

Christopher Nguyen, PhD

Vu Pham

Michael Bui, PhD

@adataoinc

@pentagoniac

strataconf.com

#StrataHadoop

The Journey

1. What We Do At Adatao
2. Challenges We Ran Into
3. How We Addressed Them
4. Lessons to Share

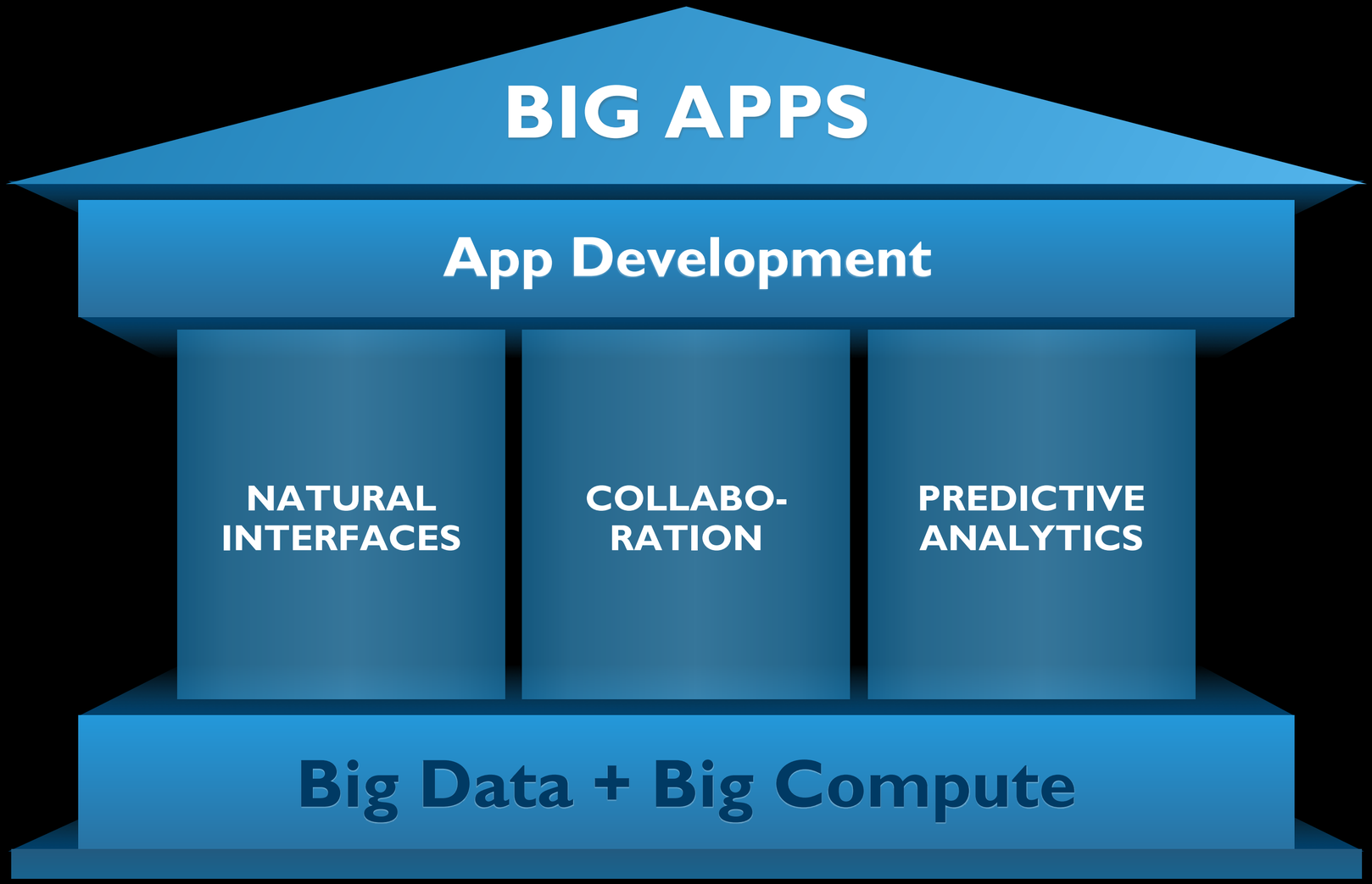
Along the Way, You'll Hear

1. How some interesting things came about
2. Where some interesting things are going
3. How some good engineering/architectural decisions are made

Acknowledgements/Discussions with

- Nam Ma, Adatao
- Haoyuan Li, TachyonNexus
- Shaoshan Liu, Baidu
- Reza Zadeh, Stanford/Databricks

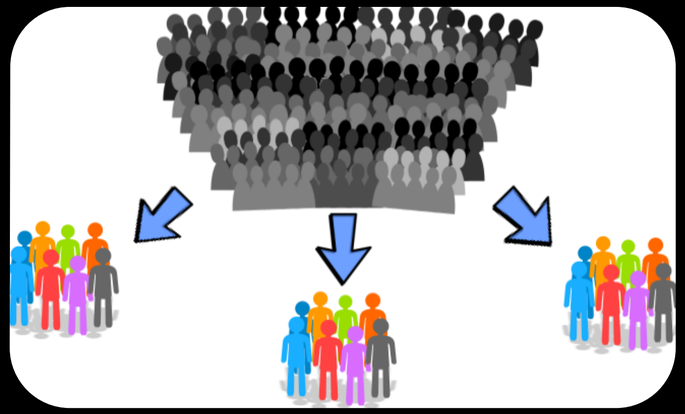
Adatao 3 Pillars



@adataoinc ADATAO DATA INTELLIGENCE FOR ALL @pentagoniac

Deep Learning Use Case

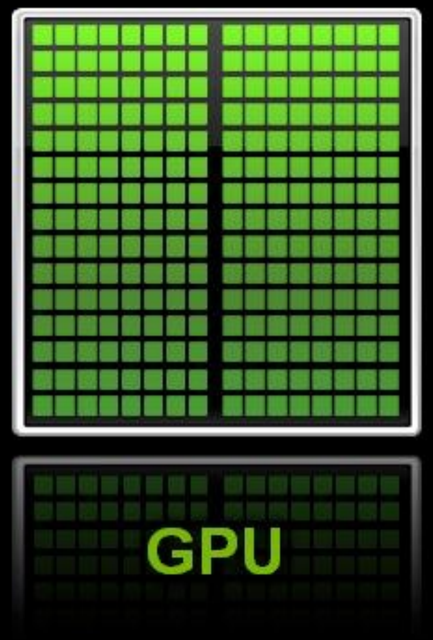
IoT



Customer Segmentation

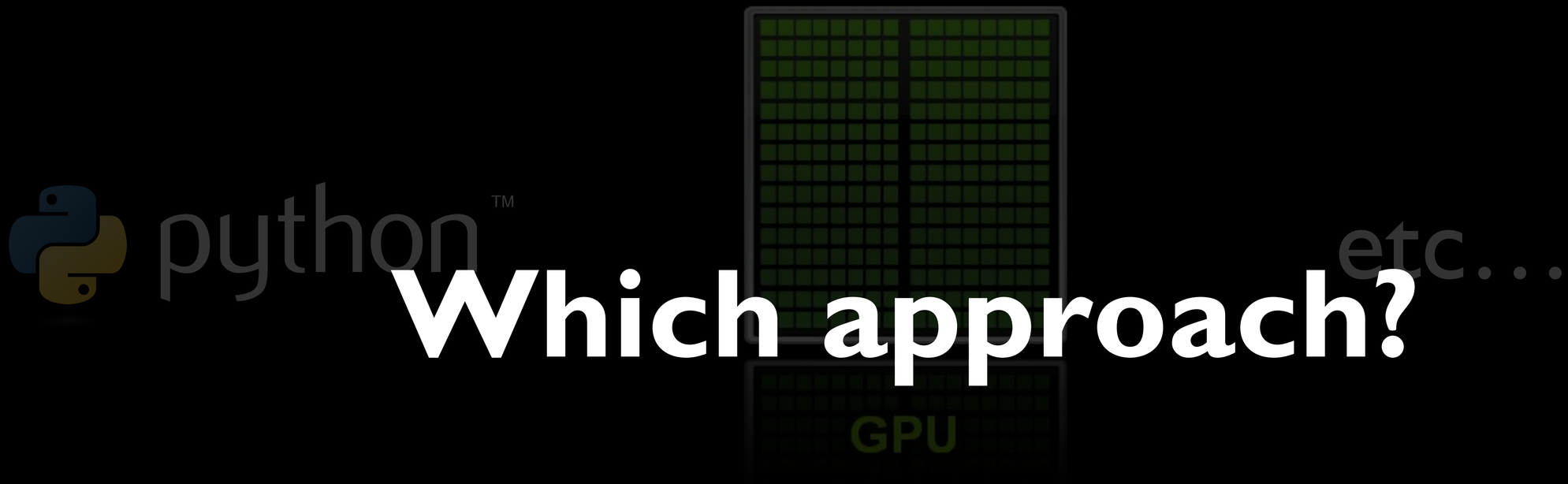
Fraud Detection

Challenge 1: Deep Learning Platform Options



etc…

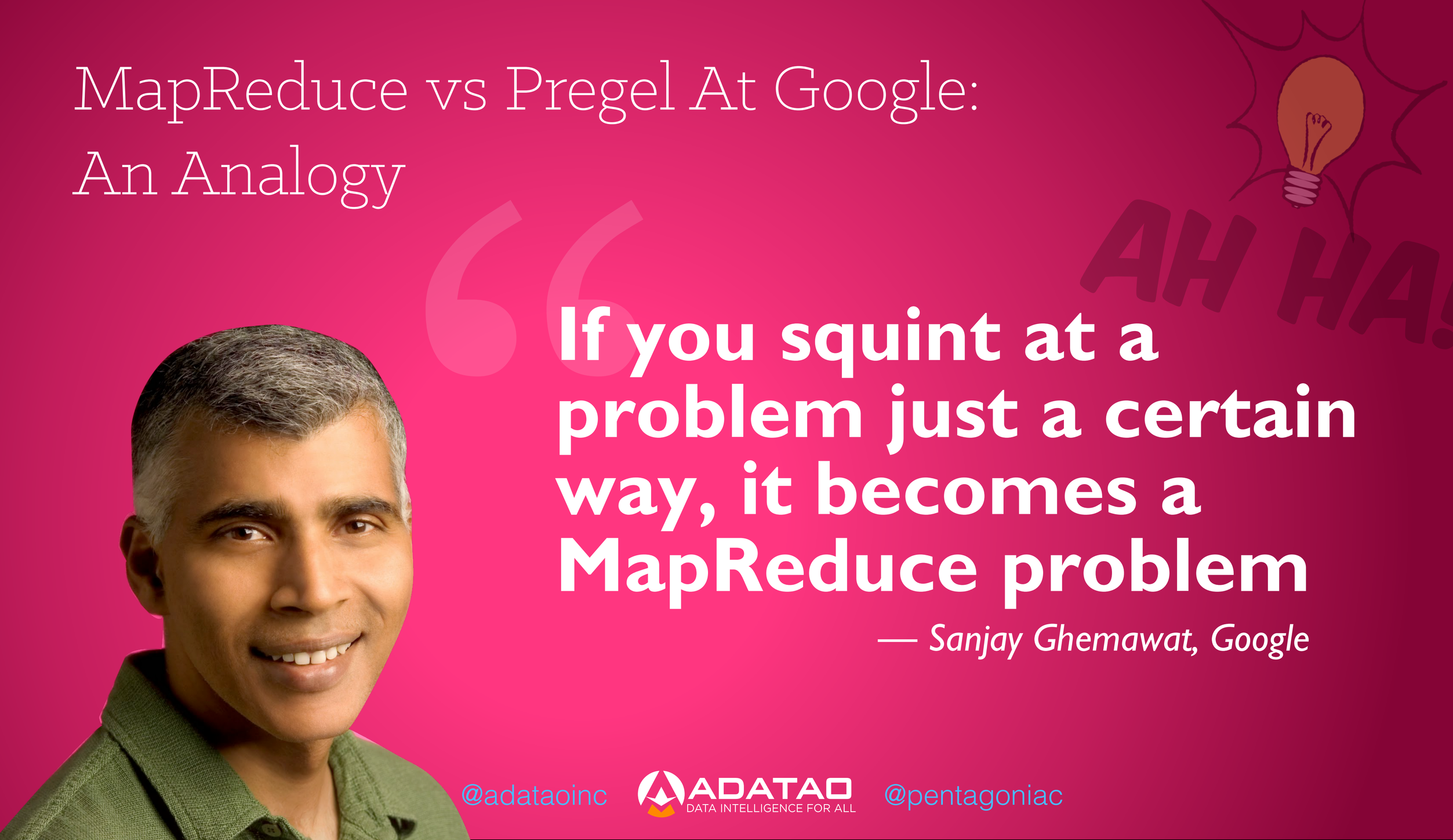
Challenge 1: Deep Learning Platform Options



Which approach?

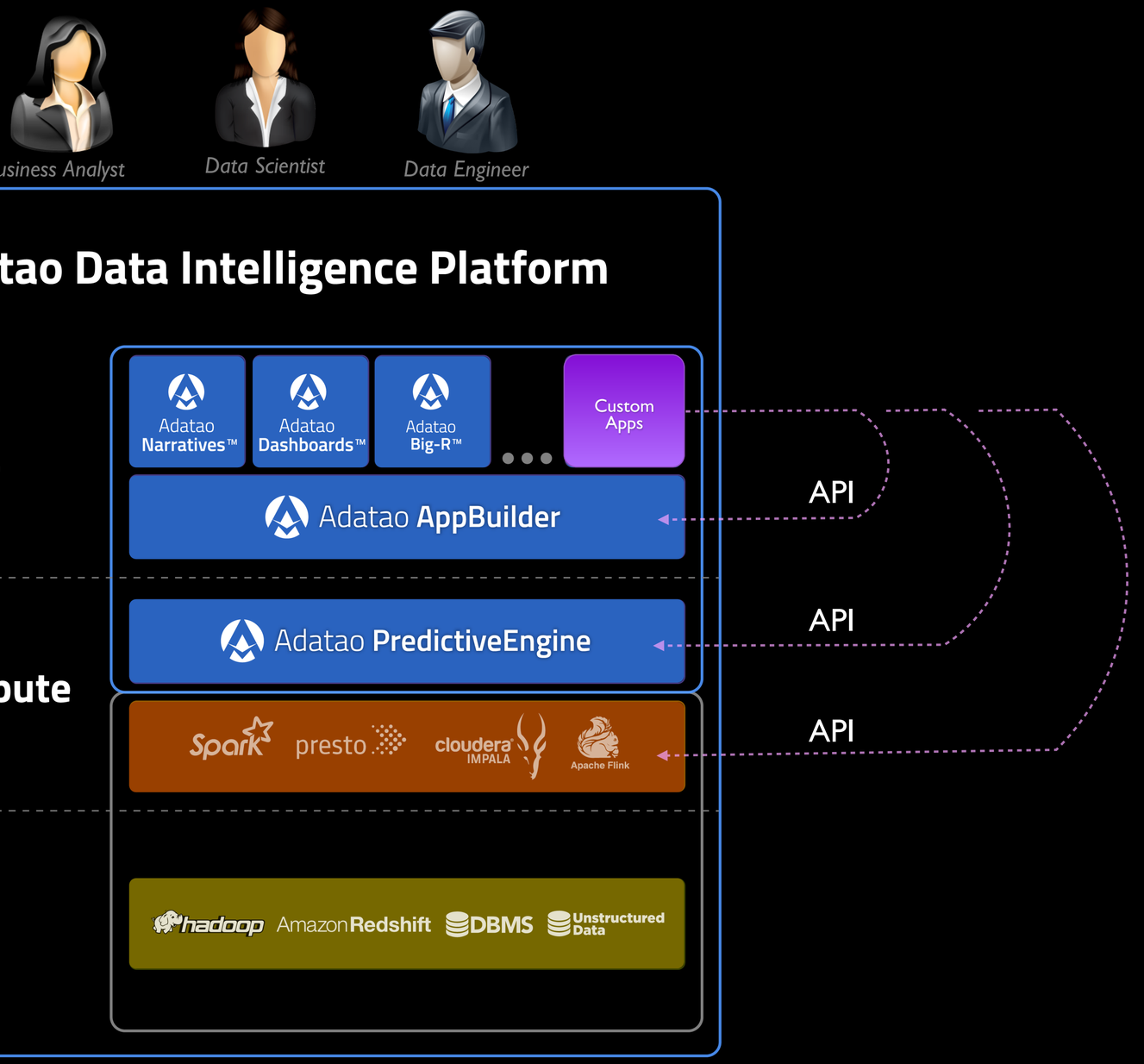
It Depends!

MapReduce vs Pregel At Google: An Analogy



If you squint at a problem just a certain way, it becomes a MapReduce problem

— Sanjay Ghemawat, Google



@adataoinc ADATAO DATA INTELLIGENCE FOR ALL @pentagoniac

Moral: *"And No Religion, Too."*

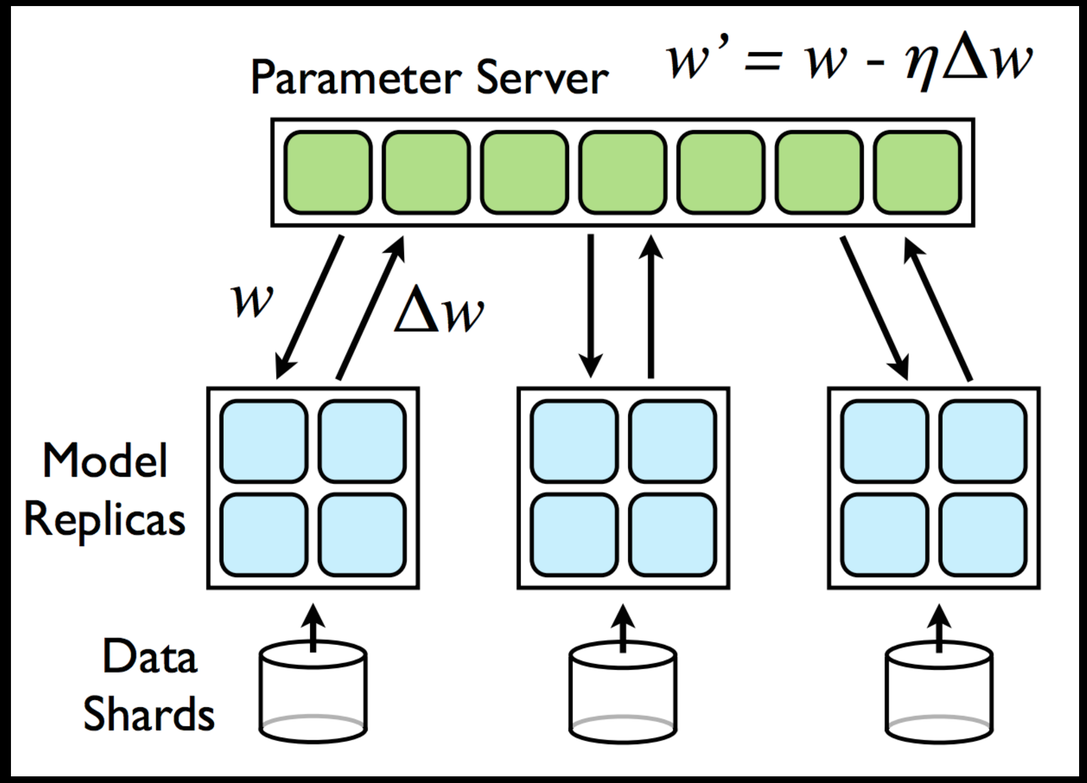
1. Architectural choices are locally optimal.

2. What's best for someone else isn't necessarily best for you. And vice versa.

Challenge 2: Who-Does-What Architecture

DistBelief

1. Compute Gradients



2. Update Params (Descent)

Large Scale Distributed Deep Networks, Jeff Dean et al, NIPS 2012

7 Dwarfs of Parallel Computing

—Phillip Colella, LLBL

The View From Berkeley—Dave Patterson, UC Berkeley

@adataoinc ADATAO DATA INTELLIGENCE FOR ALL @pentagoniac

Unleashing the Potential of Tachyon

today

Memory-Based Filesystem
Datacenter-Scale Distributed Filesystem

tomorrow

Filesystem-Backed Shared Memory
Datacenter-Scale Distributed Memory

Spark & Tachyon Architectural Options

Spark & Tachyon Architectural Options

Spark-Only

Model as Broadcast Variable

Tachyon-Storage

Model Stored as Tachyon File

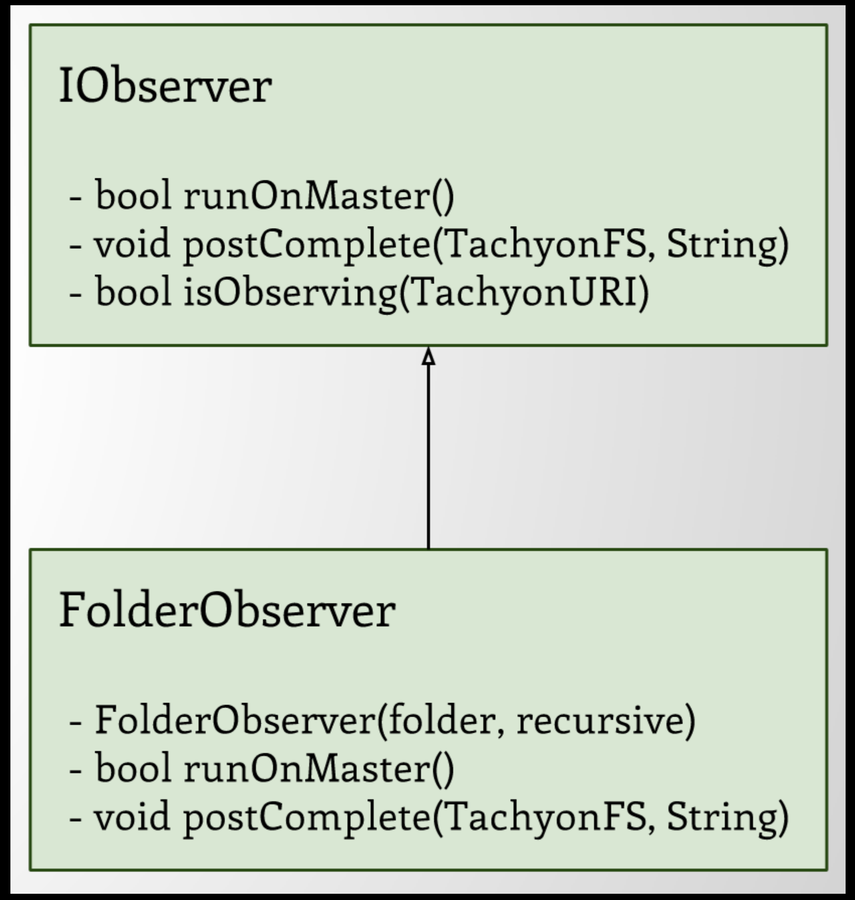
Param-Server

Model Hosted in HTTP Server

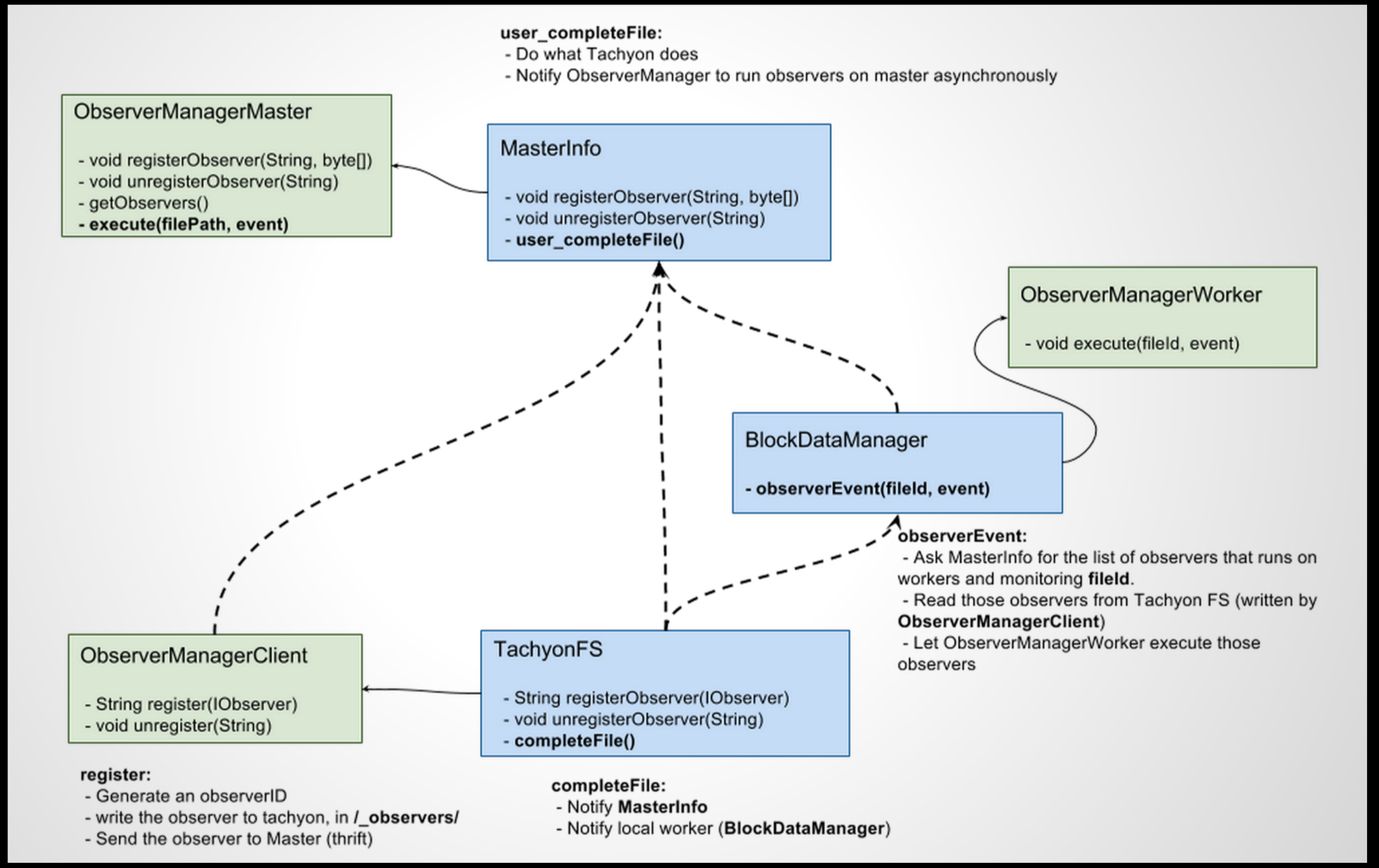
Tachyon-CoProc

Model Stored *and* Updated by Tachyon

Tachyon CoProcessor Concept

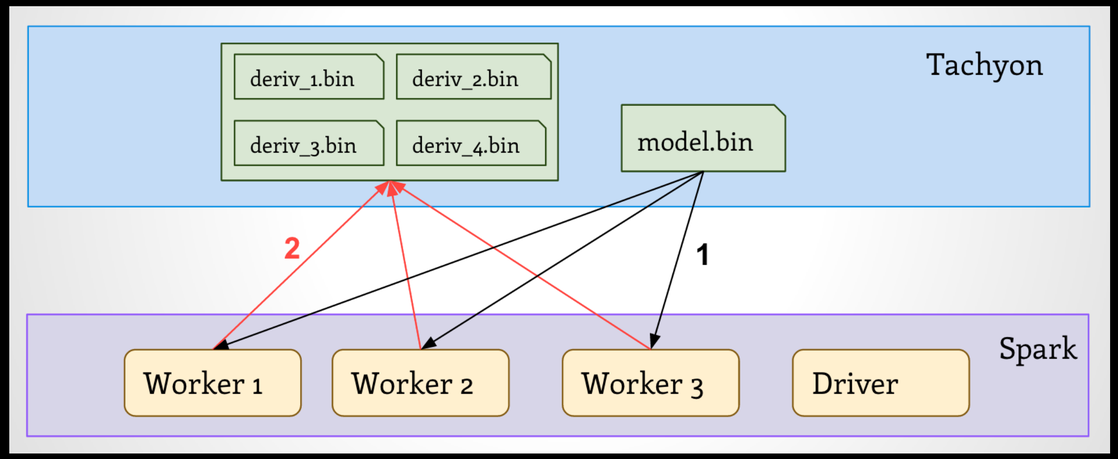


Tachyon CoProcessor

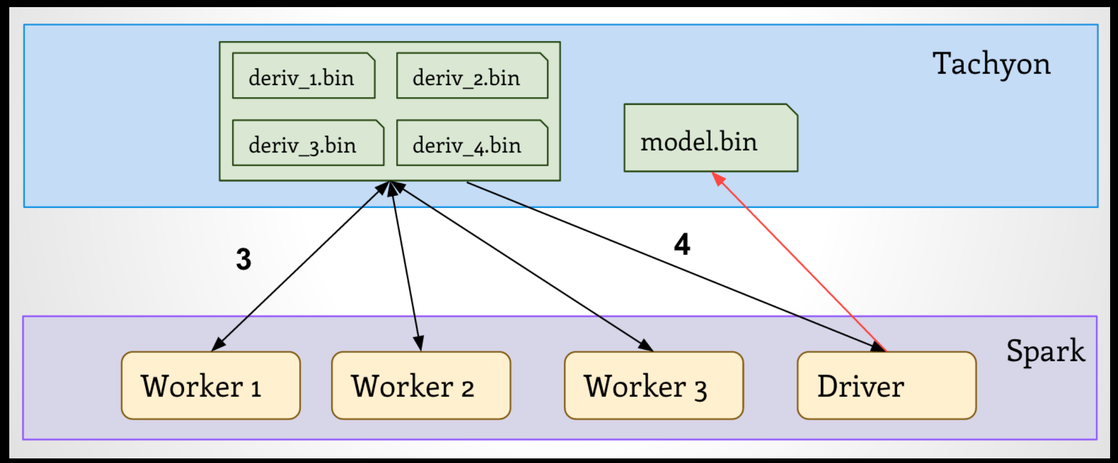


@adataoinc ADATAO DATA INTELLIGENCE FOR ALL @pentagoniac

Tachyon-Storage In Detail



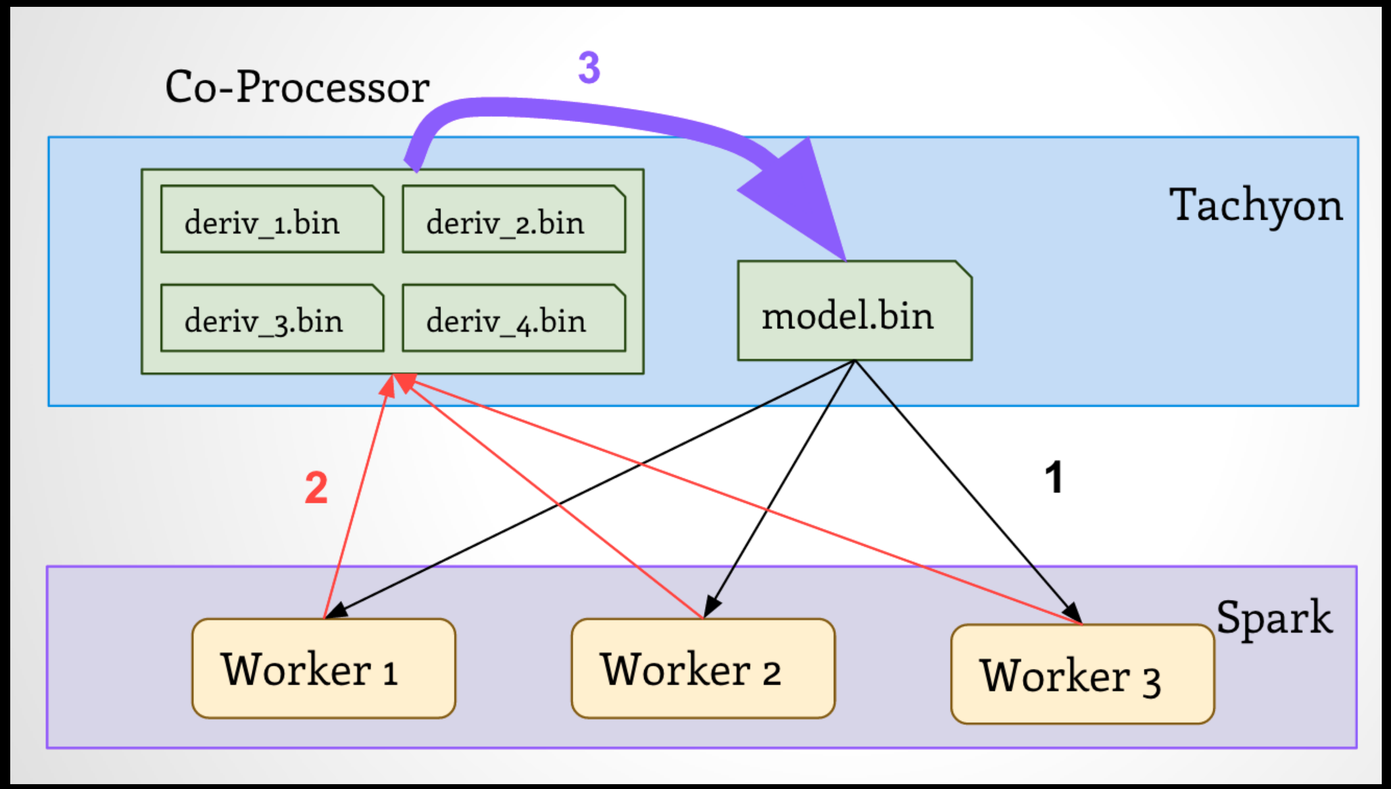
A. Compute Gradients



B. Update Params (Descent)

Tachyon-CoProcessors In Detail

B. Update Params (Descent)



A. Compute Gradients

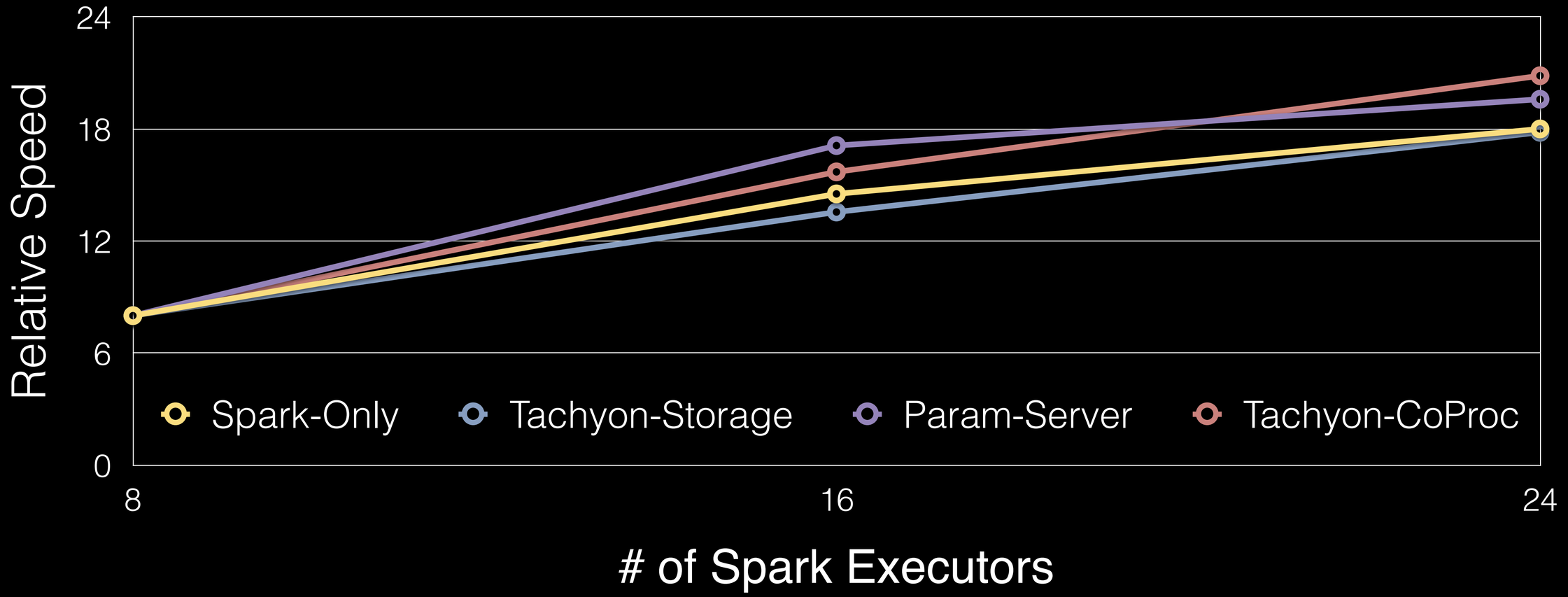
Tachyon-CoProcessors

- Spark workers do Gradients
  - Handle data-parallel partitions
  - Only compute Gradients; freed up quickly
  - New workers can continue gradient compute where previous workers left off (mini-batch behavior)
- Use Tachyon for Descent
  - Model Host
  - Parameter Server

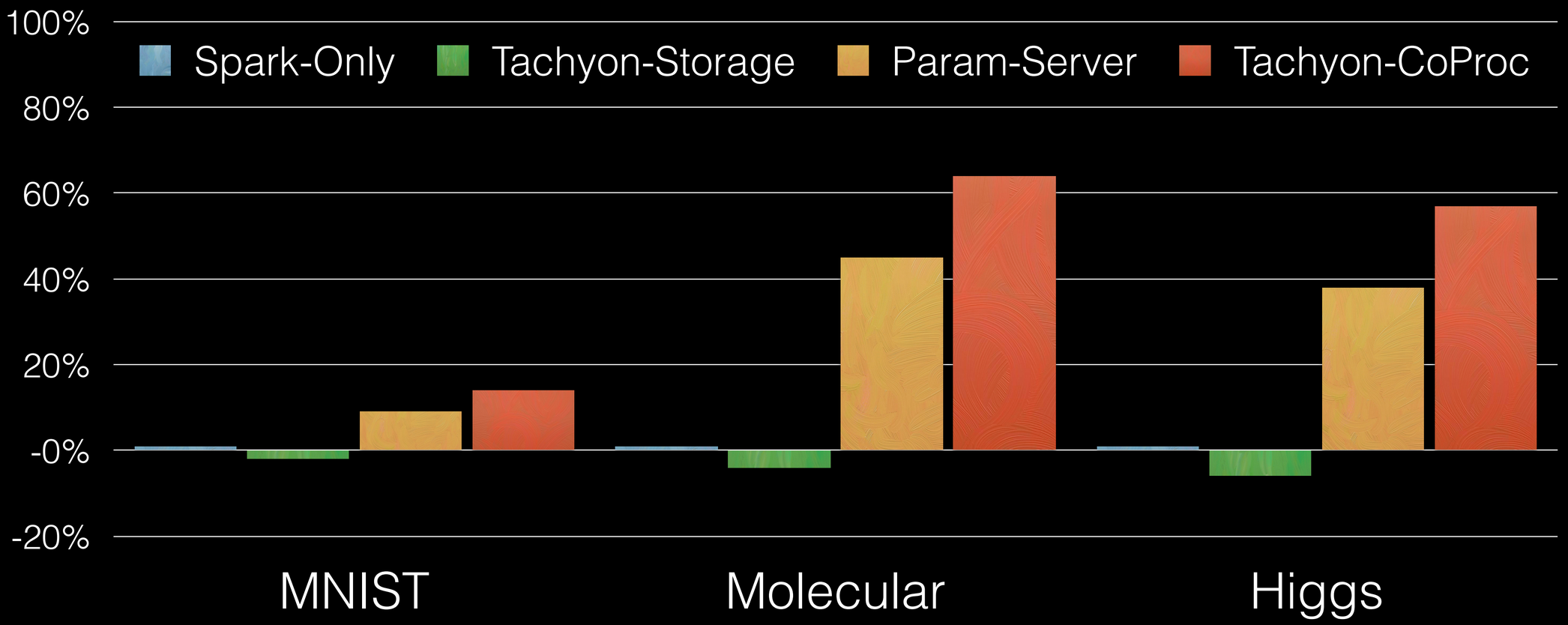
Result

Data Set	Size	Model Size
MNIST	50K x 784	1.8M
Higgs	10M x 28	8.4M
Molecular	150K x 2871	14.3M

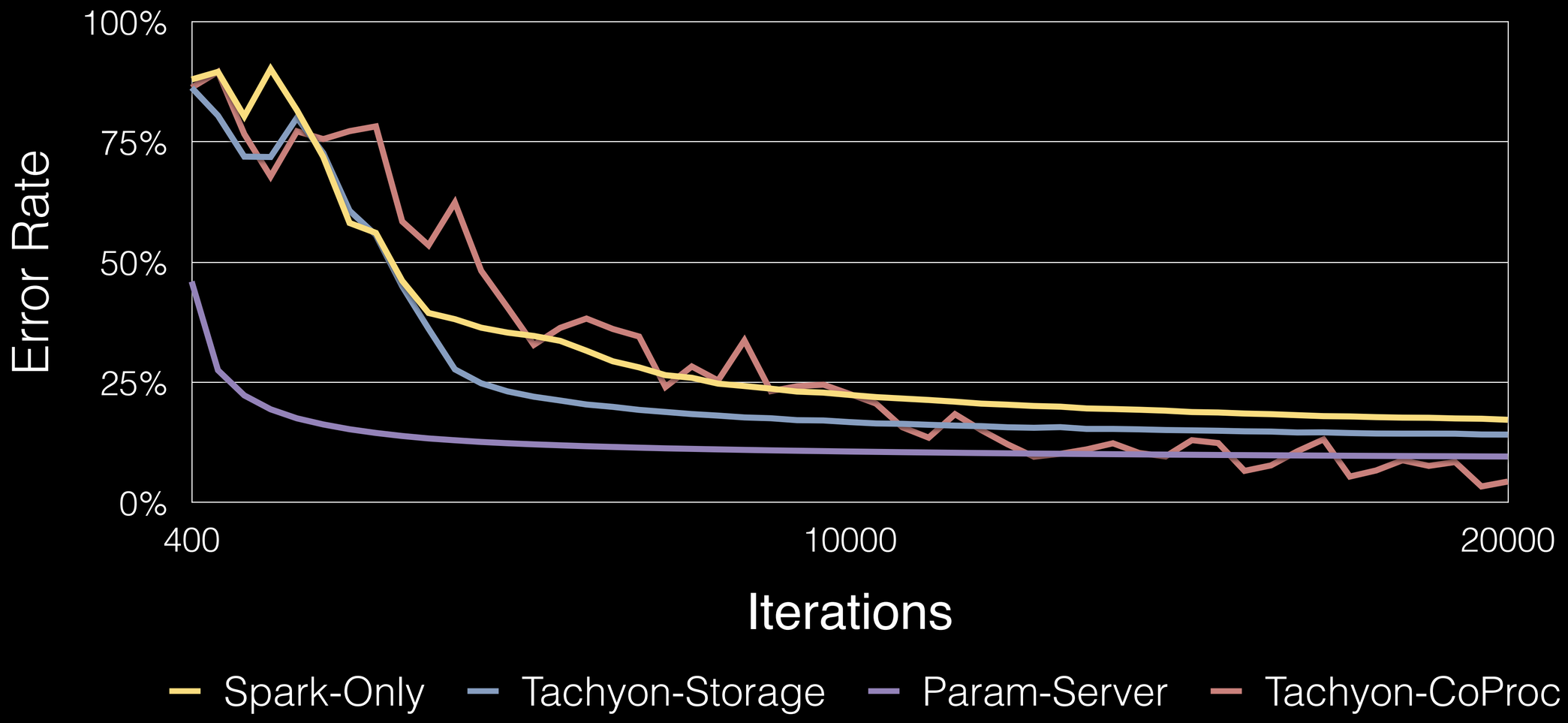
Constant-Load Scaling



Training-Time Speed-Up



Training Convergence



Lessons Learned: Tachyon CoProcessors

- Spark (Gradient) and Tachyon (Descent) can be scaled independently
- The combination gives natural mini-batch behavior
- Up to 60% speed gain, scales almost linearly, and converges faster.

Lessons Learned: Data Partitioning

- Tunable Number of Data Partitions
  - Big partitions: slow convergence, shorter time per epoch
  - Small partitions: faster convergence, longer time per epoch (for network communication)

Lessons Learned: Memory Tuning

- Typically each machine needs:
  - (model_size + batch_size * unit_count) * 3 * 4 * 1.5 * executors
- batch_size matters
- If low RAM capacity, reduce the number of executors

Lessons Learned: GPU vs CPU

- GPU is 10x faster on local, 2-4x faster on Spark
- GPU memory is limited. AWS commonly 4-6 GB of memory
- Better to have multiple GPUs per worker
- On JVM with multi-process accesses, GPUs might fail randomly

Summary

Summary

- Tachyon is much more than memory-based filesystem
  - Tachyon can become filesystem-backed shared-memory
- Combination of Spark & Tachyon CoProcessing yields superior Deep Learning performance in multiple dimensions
- Adatao is open-sourcing both:
  - Tachyon CoProcessor design & code
  - Spark & Tachyon-CoProcessor Deep Learning implementation

Appendices

Design Choices

- Combine the results of Spark workers:
  - Parameter averaging
  - Gradient averaging ✓
  - Best model