ALLUXIO

# Alluxio (formerly Tachyon)

## Open Source Memory Speed Virtual Distributed Storage

Haoyuan Li

CEO, Alluxio, Inc.

# About Alluxio

- Team
  - Alluxio Creators and Top Developers/Committers (all top 8 committers).



Carnegie Mellon

Google

Palantir

UBER

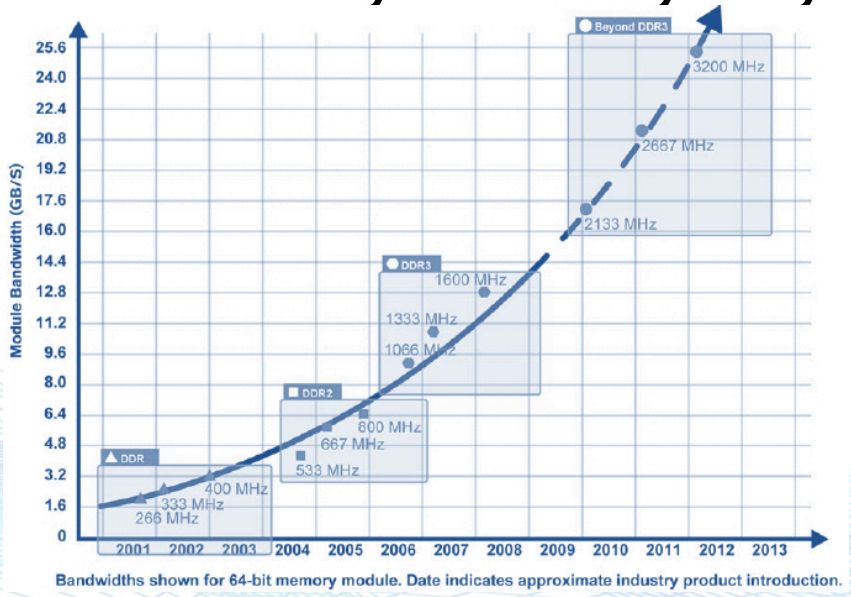
YAHOO!

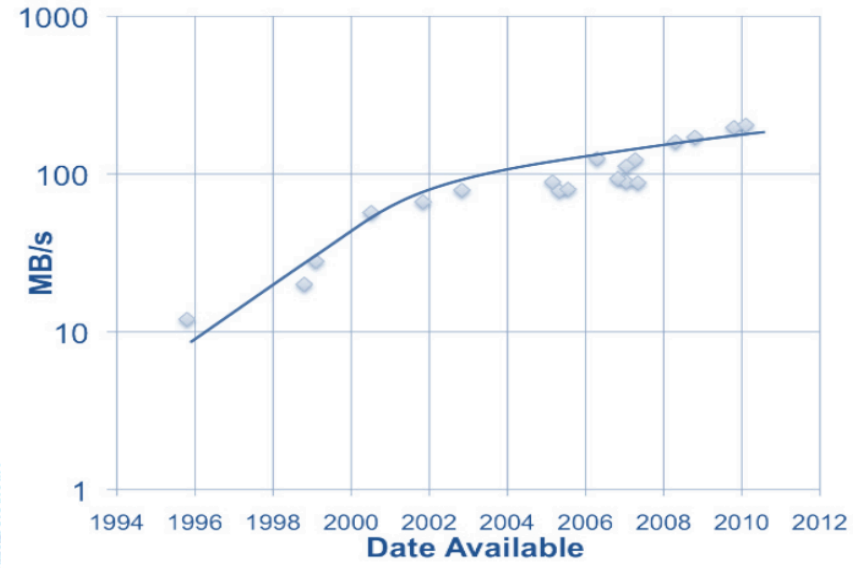
- Investors

ANDREESSEN HOROWITZ

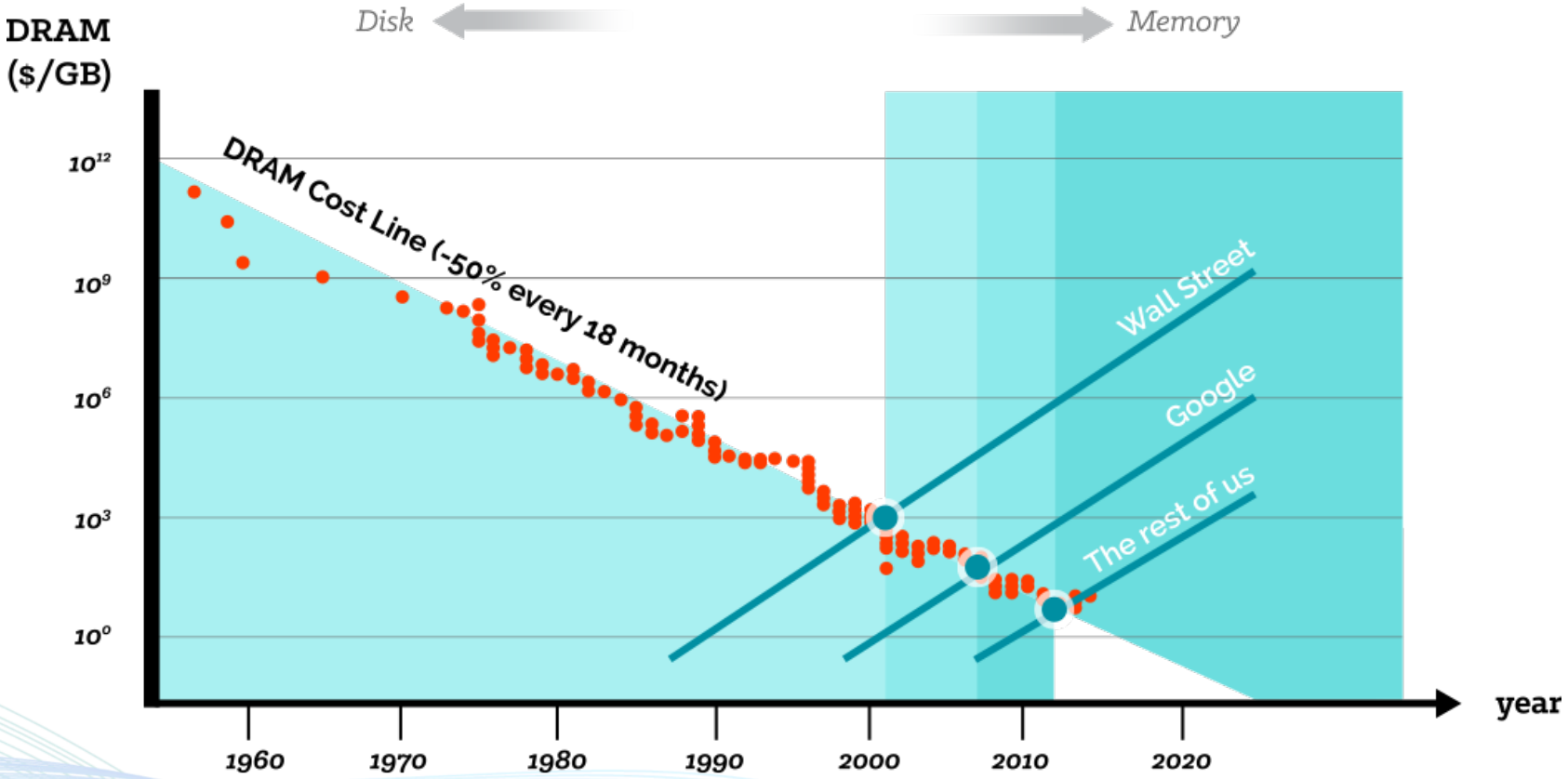
# Performance Trend: Memory is Fast

- RAM throughput increasing exponentially
- Disk throughput increasing slowly
- Memory-locality key to interactive response



# Price Trend: Memory is Cheaper




Source: jcmit.com

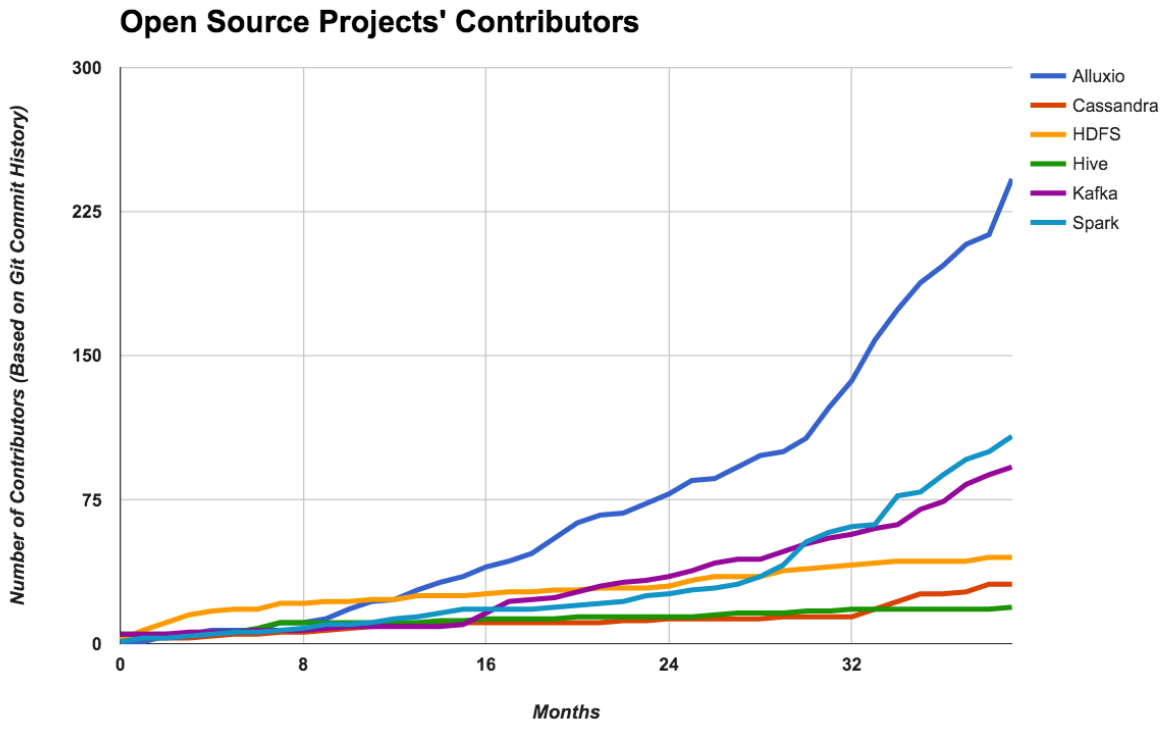
# The Big Data Ecosystem Today

# What is Alluxio?

- Alluxio: Memory Speed Virtual Distributed Storage
- Enables Virtualized Data Across Multiple Types of Storage

# Open Source Alluxio System

- The fastest growing open source project in big data
- Over 250 contributors from over 100 organizations



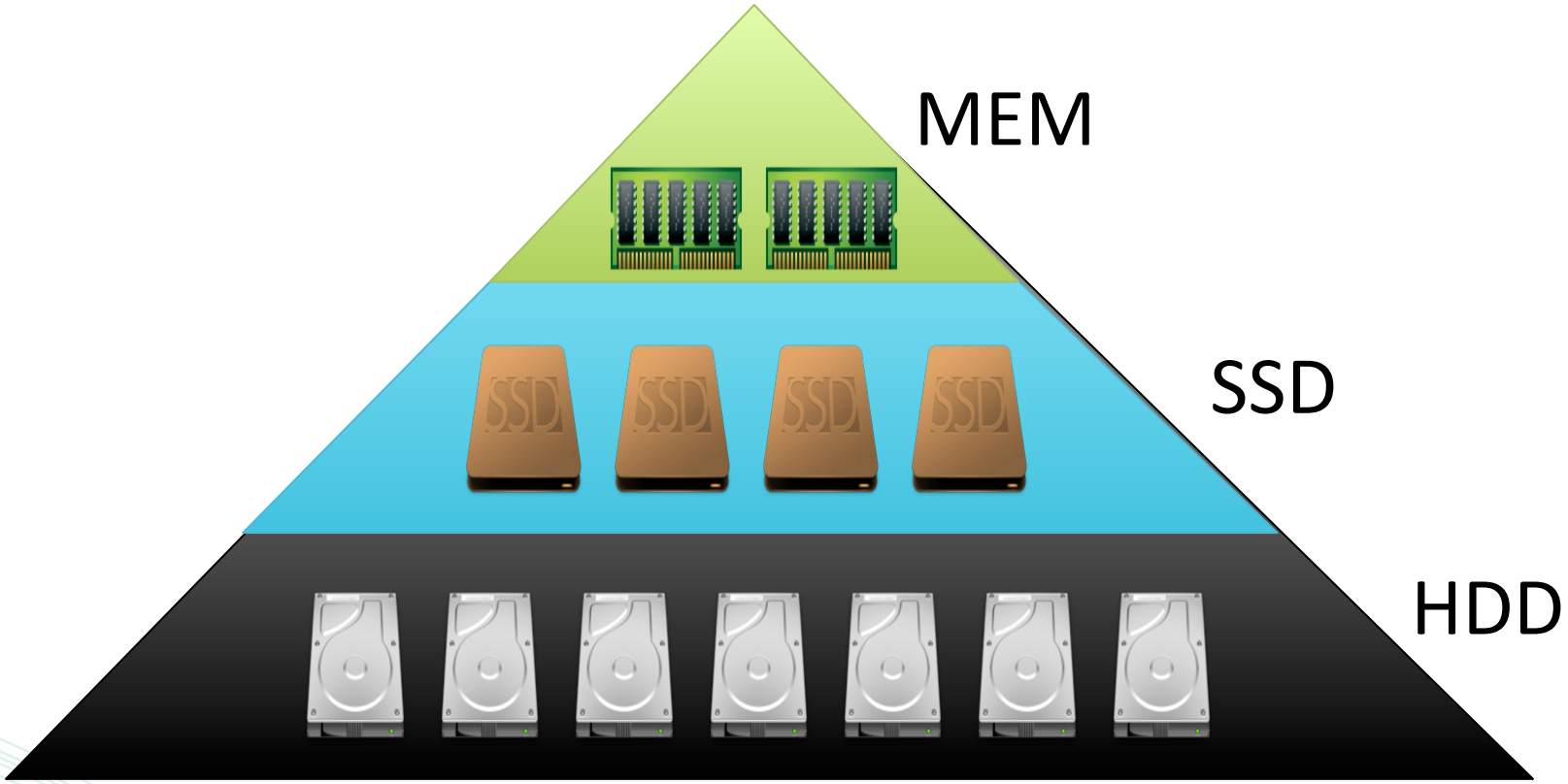
# Alluxio Benefits

- Flexibility
  - Enable new workloads across any storage systems
  - Unified Name Space enable application to access data in any storage system
- Agility
  - Work with the framework of your choice
  - Work with the storage of your choice
- Performance
  - High performance data access
- Cost
  - Grow Storage and Compute independently
- Any application accesses any data from any storage at memory speed.

# New Features

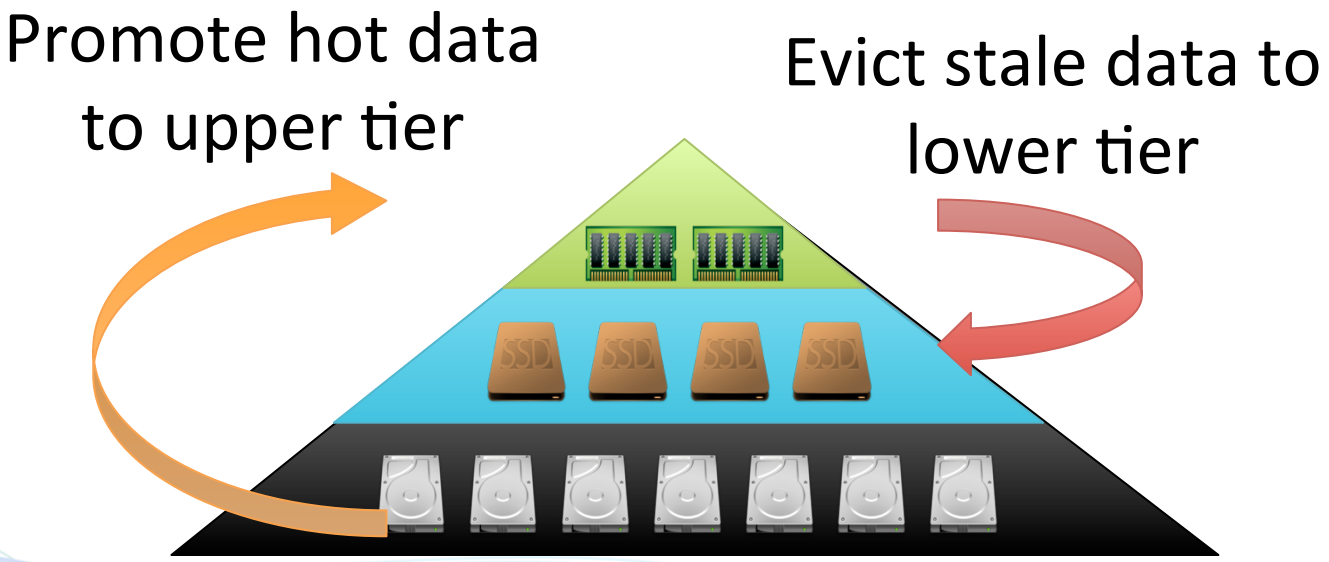
- Tiered Storage
- Transparent Naming
- Unified Namespace
- Native Amazon S3, Google Cloud Storage, Open Stack Swift, Alibaba OSS integrations
- Fuse Connector, K/V Interface
- One Command Cluster Deployment
- Metrics Reporting

# The Storage Tier Hierarchy



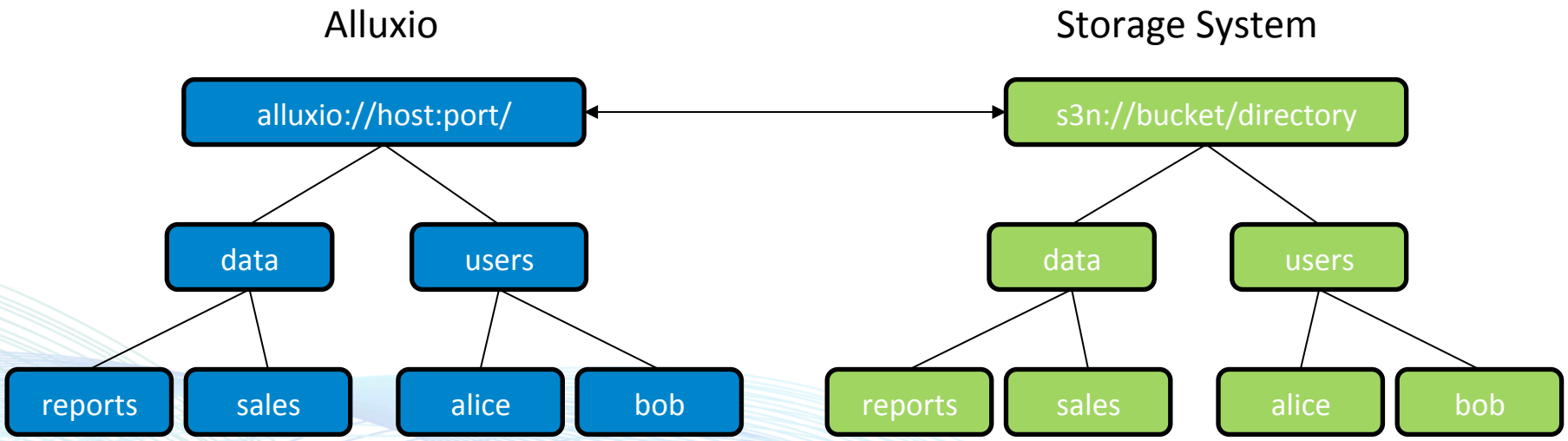
# Automatic Data Migration

- Data can be evicted to lower layers if it is “cooling down”
- Data can be promoted to upper layers if it is “warming up”



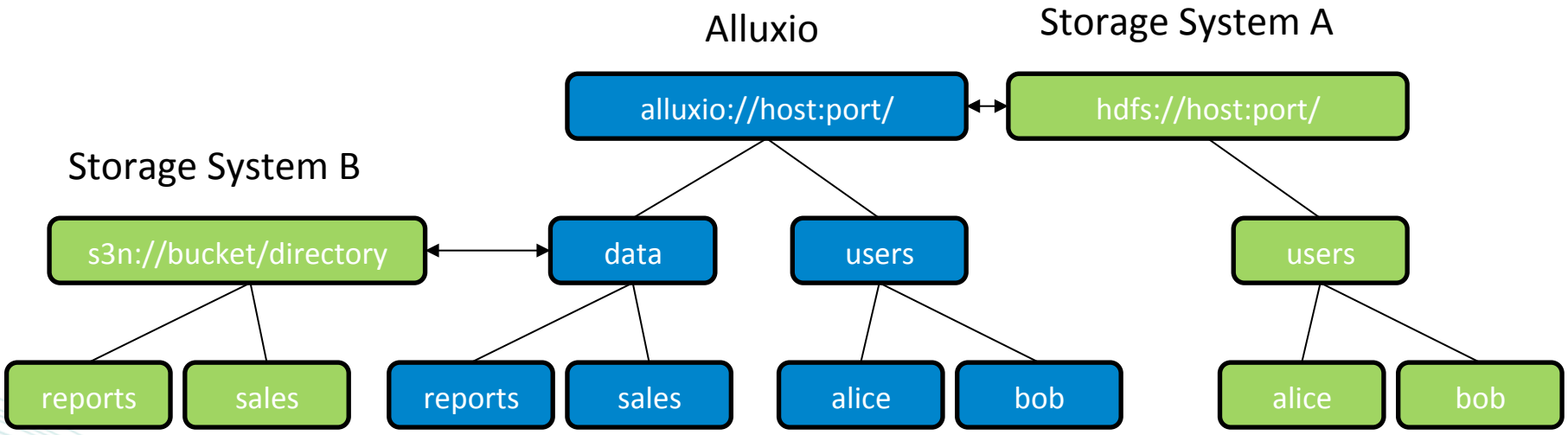
# Transparent Naming

- Applications can transparently and efficiently interact with remote storage through Alluxio.
- Applications do not need to use different APIs for interacting with different storage systems.



# Unified Namespace

- Applications can read and write different storage systems
- Decouples data location from application



ALLUXIO

# Use Cases

- Accelerate access to remote storage
- Share data across jobs at memory speed
- Transparently manage data across different storage systems

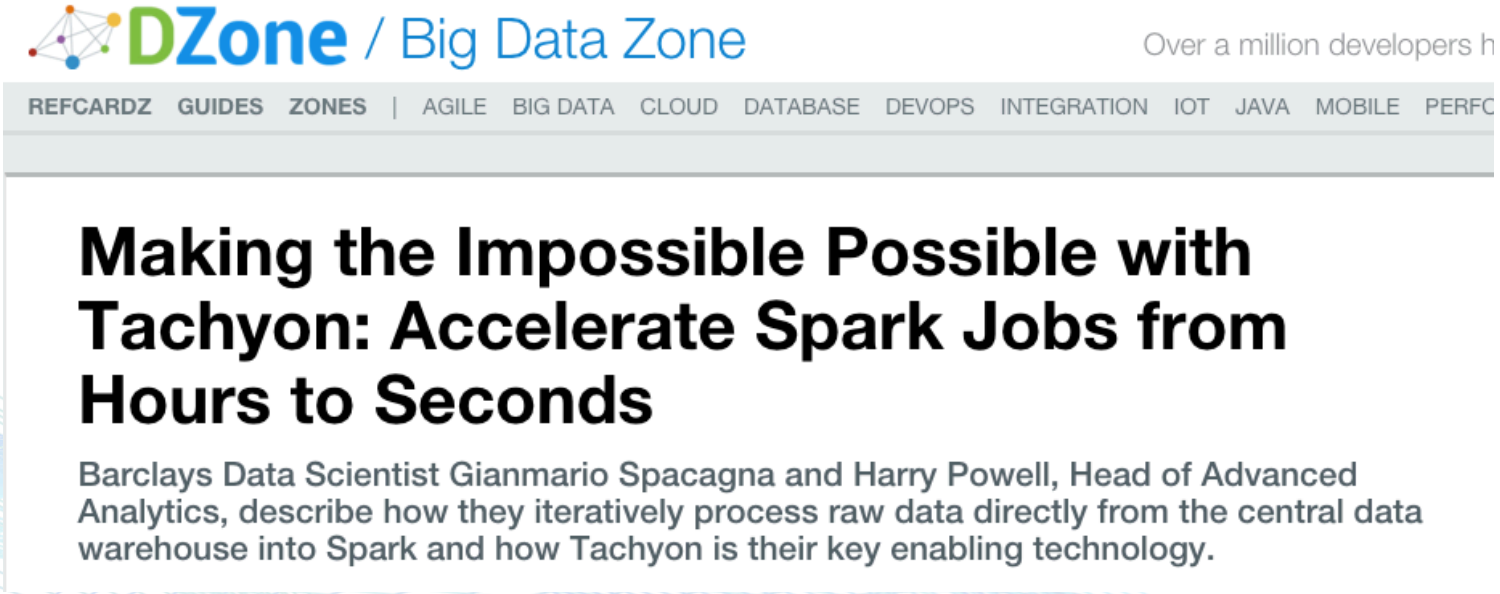



- Framework: Spark
- Under Storage: Baidu's File System
- Storage Media: MEM + HDD
- 200+ nodes deployment

Baidu Queries Data 30 Times Faster with Alluxio

BARCLAYS ALLUXIO

- Framework: Spark
- Storage Media: MEM
- Improvement from Hours to Seconds



去哪儿? Qunar.Com 聪明你的旅行

ALLUXIO

ALLUXIO

- Framework: Spark Streaming & Hive
- Under Storage: HDFS & Ceph
- Storage Media: MEM + HDD
- 200 nodes deployment
- Alluxio enables previously impossible jobs to finish
- 300x Performance Improvement

ALLUXIO

# Contacts

- Alluxio Project: www.alluxio.org
- Alluxio Inc: www.alluxio.com
- Development: www.github.com/Alluxio/alluxio
- Meet Friends: www.meetup.com/Alluxio
- Contact: info@alluxio.com ; haoyuan@alluxio.com

ALLUXIO

# Thank You