**Motivation – Memory is King**
- RAM throughput increasing exponentially.
- Disk throughput increasing slowly.

Memory-locality key to interactive response time

**Problem Solved?**
- Data sharing is the bottleneck in analytics pipeline, due to slow I/O to storage systems.
- Cache lose when process crashes
- In-memory data duplication and Java Garbage Collection

**Solution Overview**
- Feature 1: memory-centric storage architecture
- Feature 2: push lineage down to storage layer

**Stack**

<table>
<thead>
<tr>
<th>Spark</th>
<th>MapReduce</th>
<th>H2O</th>
<th>Shark</th>
<th>GraphX</th>
<th>Impala</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Tachyon**

<table>
<thead>
<tr>
<th>HDFS</th>
<th>S3</th>
<th>GlusterFS</th>
<th>OrangeFS</th>
<th>NFS</th>
<th>Ceph</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Memory-centric Storage Architecture**

**Lineage**

**Question 1: How long to get missing data back?**
Lineage enables **Asynchronous Checkpointing.**

**Edge Algorithm**
- Checkpoint leaves
- Checkpoint hot files
- Bounded recovery cost

**Question 2: How to allocate recomputation Resource?**
- Priority Based Scheduler
- Fair Sharing Based Scheduler

**Evaluation**

Comparison with in Memory HDFS

**Workflow Improvement**

**Recomputation Resource Consumption**

**Open Source Status**
- Apache 2.0 License, Version 0.5.0 (July 2014).
- More than 50 contributors from over 15 companies.
- Deployed at tens of companies.
- Spark and MapReduce can use it without changing any code:
  - Tachyon is in Fedora 20 (Thanks to Redhat)
  - Commercially supported by Atigeo, run in dozens of customers clusters.
  - Tachyon is the default off-heap storage solution for Apache Spark.
- Supports Apache HDFS, GlusterFS, Amazon S3, Ceph, OrangeFS, NFS

**Summary**
- Tachyon is a MEMORY-CENTRIC distributed storage system with LINEAGE as a first class citizen.
- Website: [http://tachyon-project.org/](http://tachyon-project.org/)
- Github: [https://github.com/amplab/tachyon](https://github.com/amplab/tachyon)