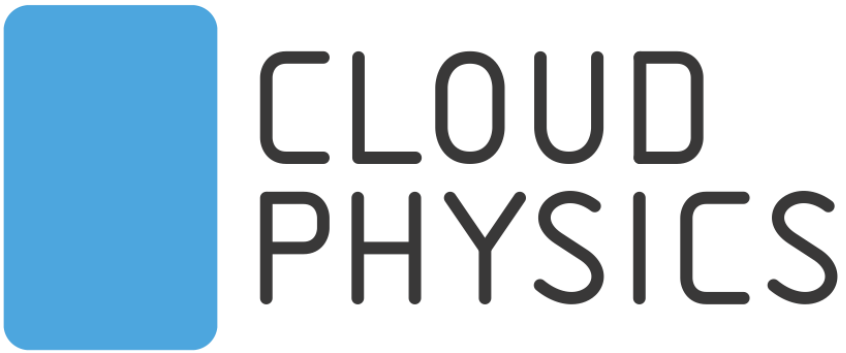


Collective Intelligence for Data Center Operations Management

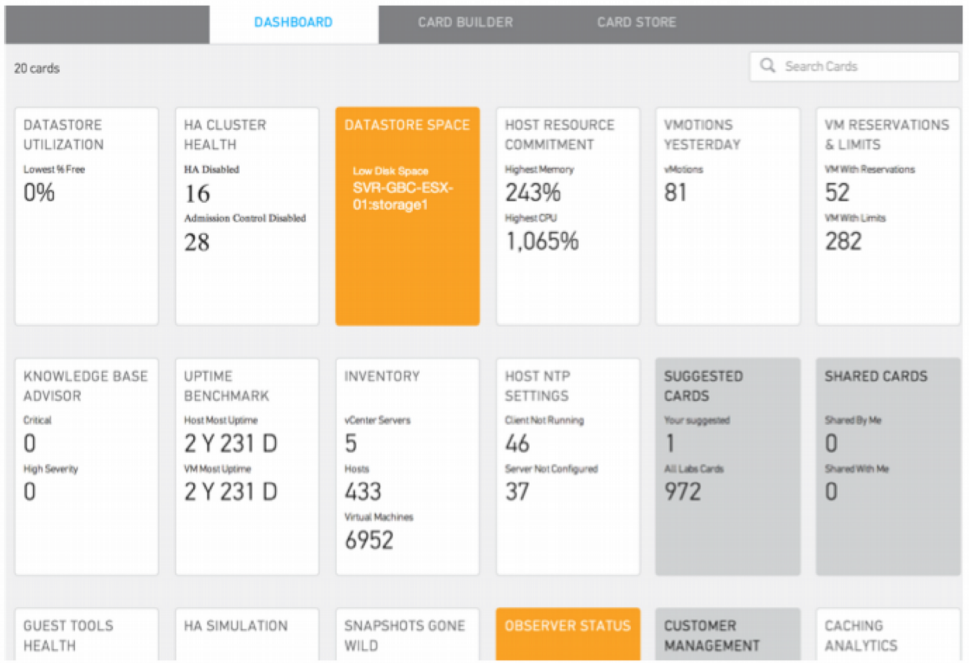
Xiaojun (XJ) Liu, Chief Scientist

Spark Summit 2013

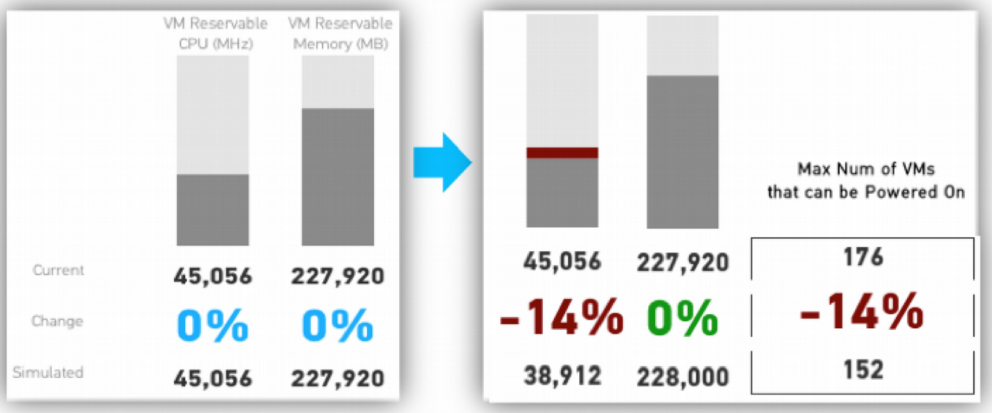
Outline

- Operations Management SaaS
- Experiences with Spark
- Ongoing Work

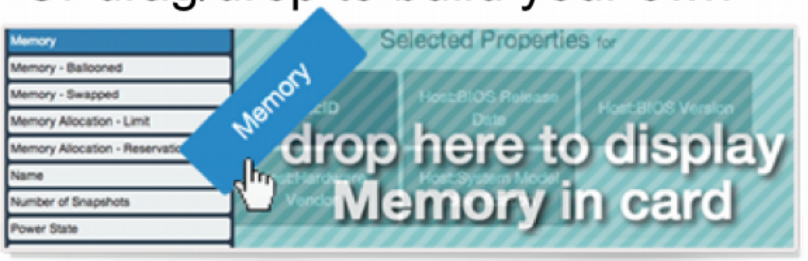
Operations Management SaaS



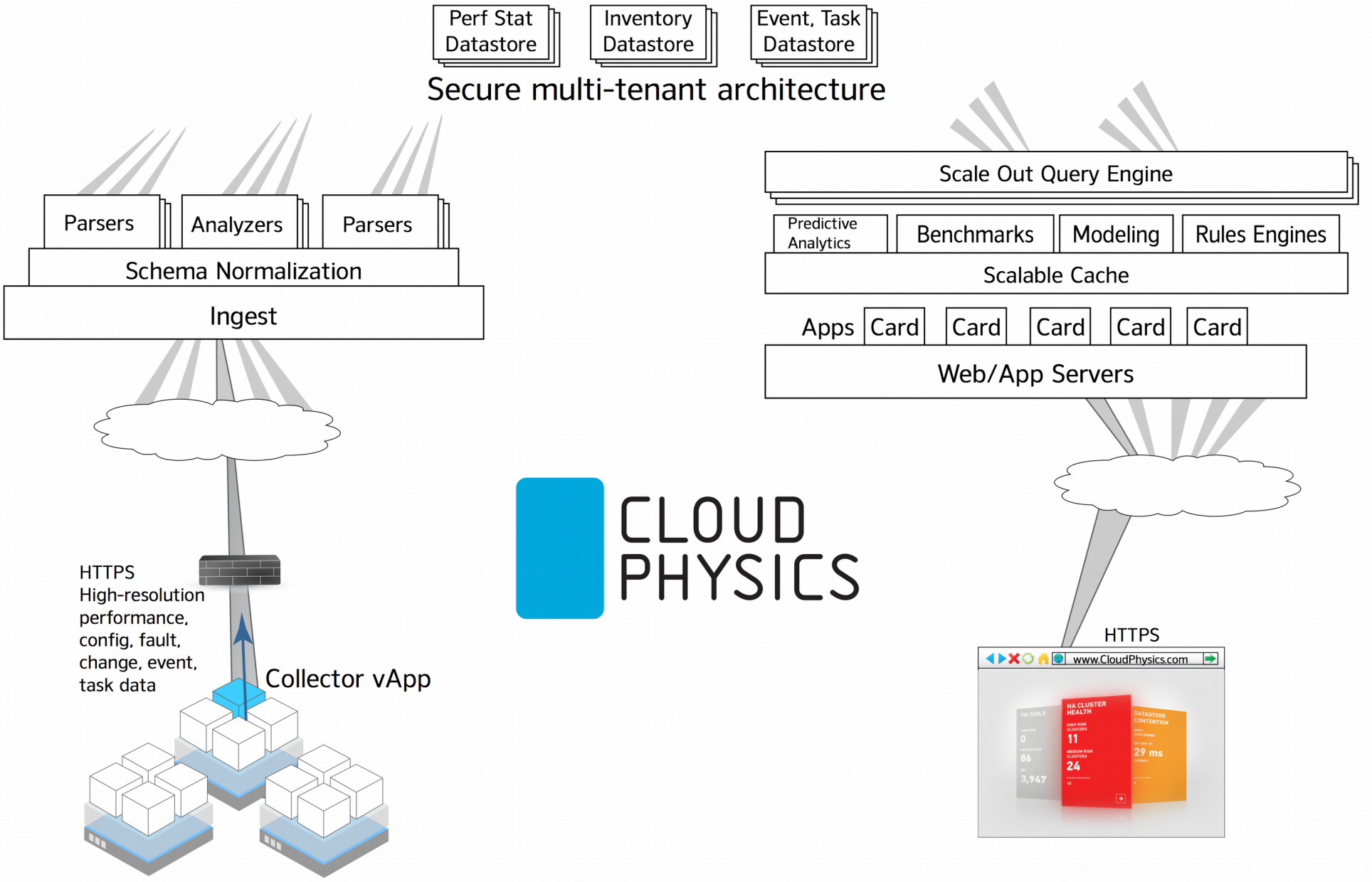
Use an existing card (e.g. HA Sim)



Or drag/drop to build your own



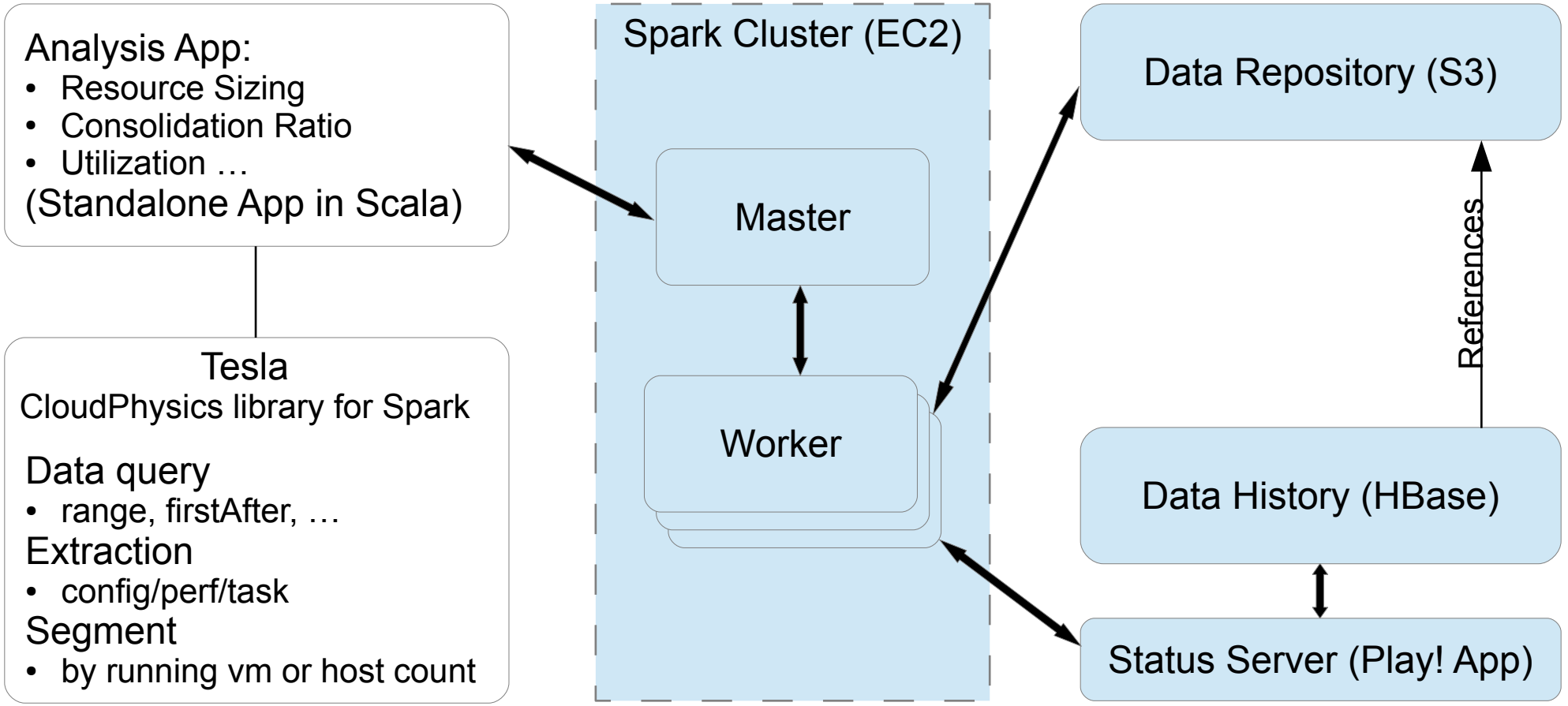
Data Pipeline



The Data We Collect

- Configuration, performance, tasks and more
- From virtual machines, servers, networks and storage
- 100 billion+ metric samples per day
- On average 1.3 million properties per datacenter

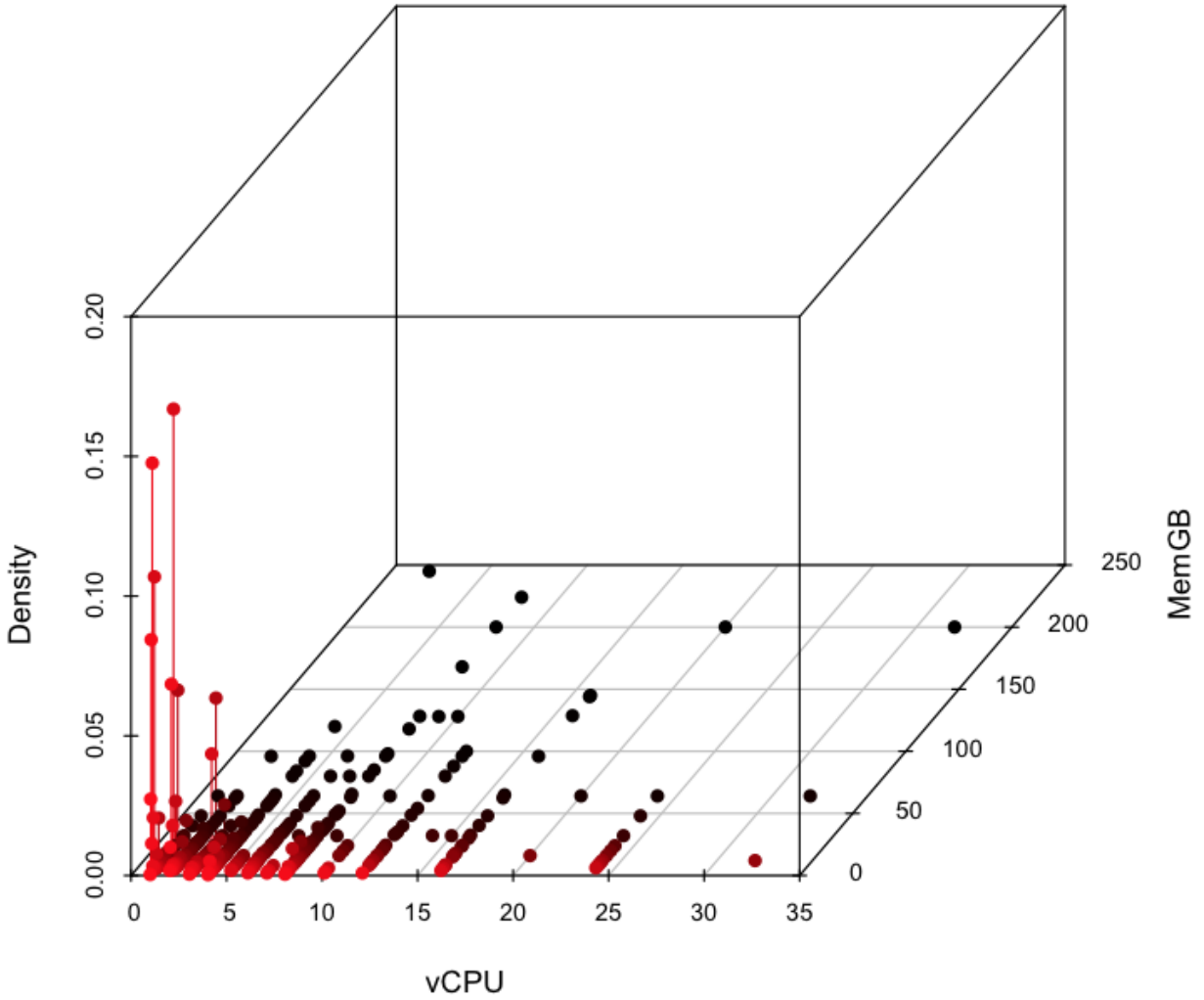
Cross-User Analysis Setup



Analyses We Have Done

- VM configured resources and utilization
- Server CPU and memory utilization distribution
- Storage array and interconnect adoption
- Virtualization product feature adoption

VM vCPU and Memory Size Distribution



Experiences with Spark on EC2

+ Quick getting up to speed
+ Great EC2 support
+ Tolerating variations in task execution time
- Variations in EC2 instance performance

Next Steps

- Create RESTful API for frequently used analyses to make update and consumption easier
- Utilize Shark and MLbase for cross-user data analysis
- Utilize Spark Streaming for near-realtime performance data analysis

Thank You!

www.cloudphysics.com

info@cloudphysics.com

@cloudphysics

Xiaojun (XJ) Liu

xj@cloudphysics.com

