



CLOUD
PHYSICS

Collective Intelligence for Data Center Operations Management

Xiaojun (XJ) Liu, Chief Scientist

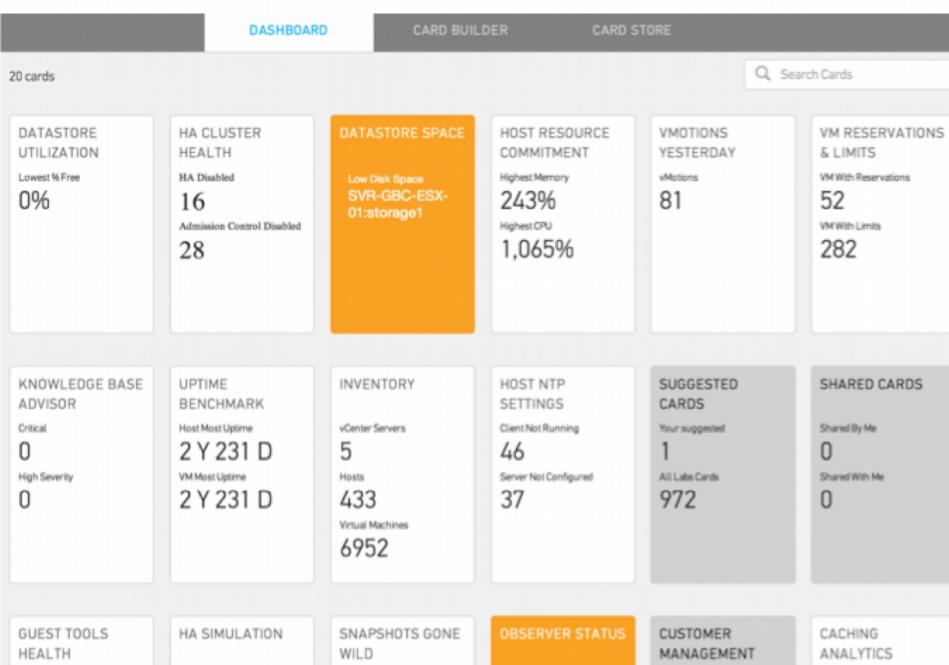
Spark Summit 2013

Outline

- Operations Management SaaS
- Experiences with Spark
- Ongoing Work

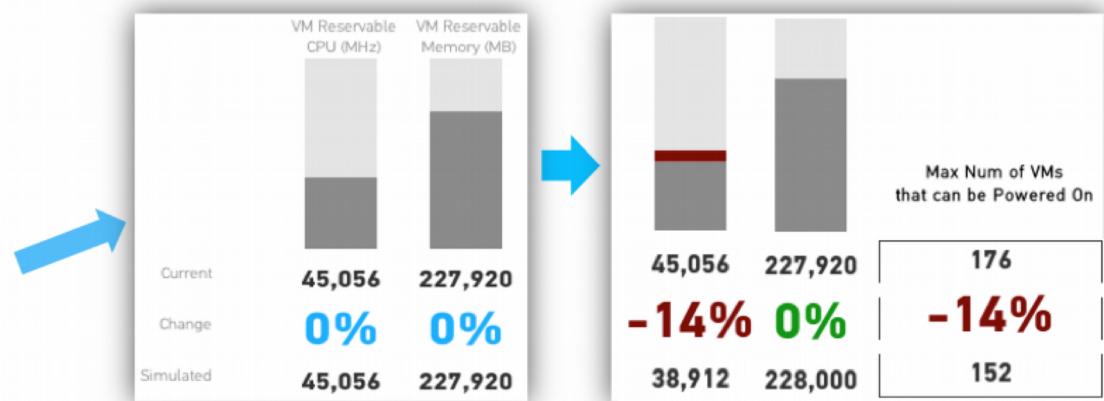
Operations Management SaaS

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

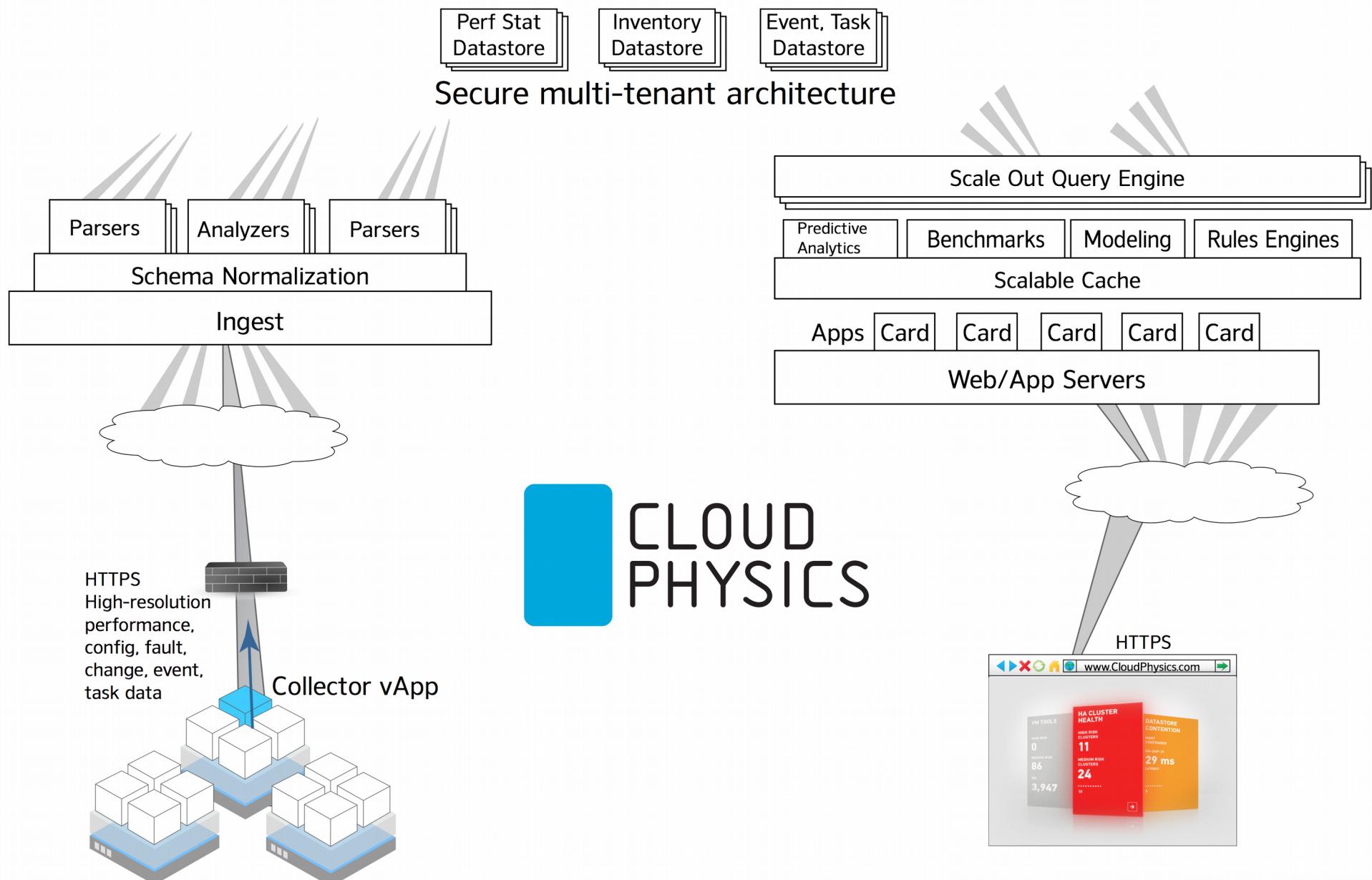


The dashboard displays the following cards:

- DATASTORE UTILIZATION**: Lowest % Free 0%
- HA CLUSTER HEALTH**: HA Disabled 16, Admission Control Disabled 28
- DATASTORE SPACE**: Low Disk Space SVR-GBC-ESX-01:storage1
- HOST RESOURCE COMMITMENT**: Highest Memory 243%, Highest CPU 1,065%
- VMOTIONS YESTERDAY**: VMotions 81
- VM RESERVATIONS & LIMITS**: VM With Reservations 52, VM With Limits 282
- KNOWLEDGE BASE ADVISOR**: Critical 0, High Severity 0
- UPTIME BENCHMARK**: Host Most Uptime 2 Y 231 D, VM Most Uptime 2 Y 231 D
- INVENTORY**: vCenter Servers 5, Hosts 433, Virtual Machines 6952
- HOST NTP SETTINGS**: Client Not Running 46, Server Not Configured 37
- SUGGESTED CARDS**: Your suggested 1, All Labs Cards 972
- SHARED CARDS**: Shared By Me 0, Shared With Me 0
- GUEST TOOLS HEALTH**
- HA SIMULATION**
- SNAPSHOTS GONE WILD**
- OBSERVER STATUS**
- CUSTOMER MANAGEMENT**
- CACHING ANALYTICS**



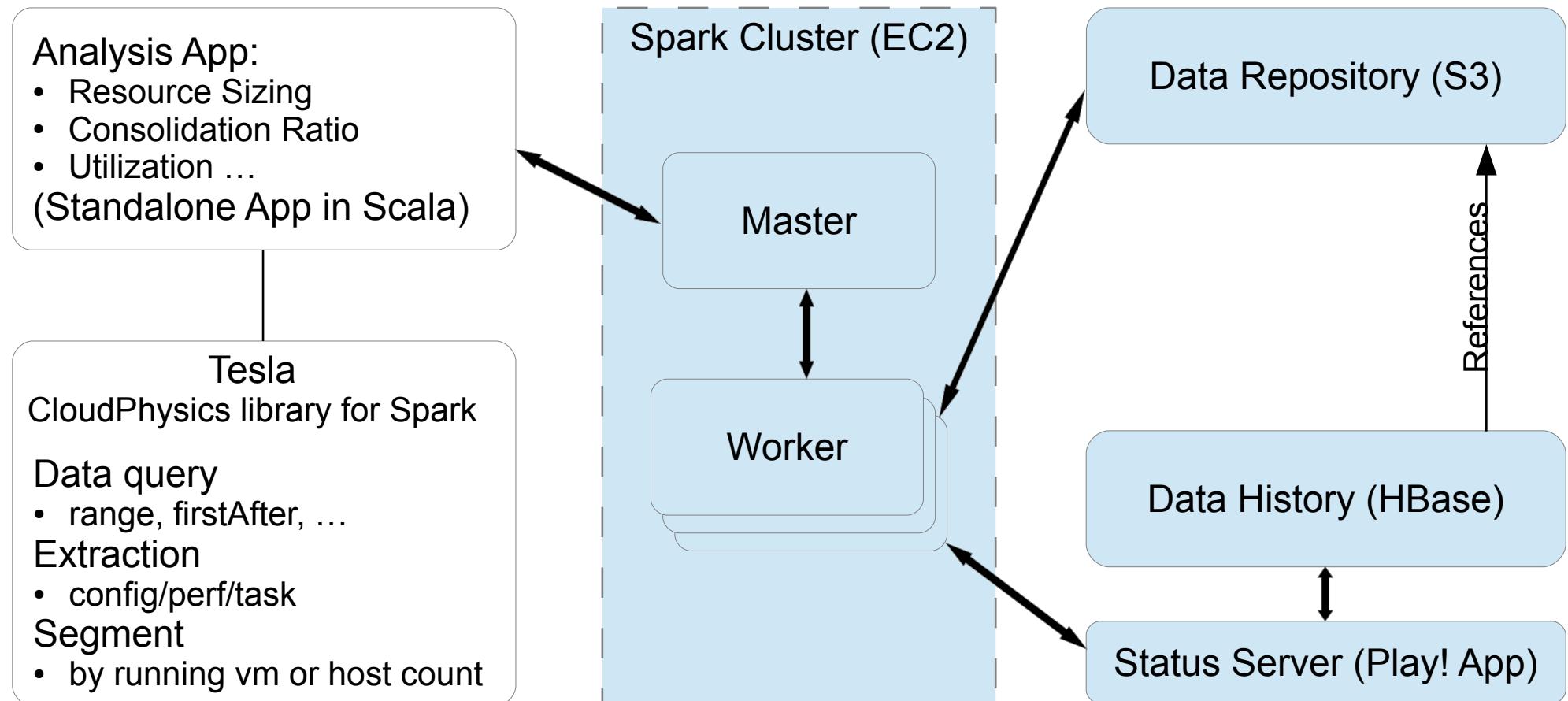
Data Pipeline



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

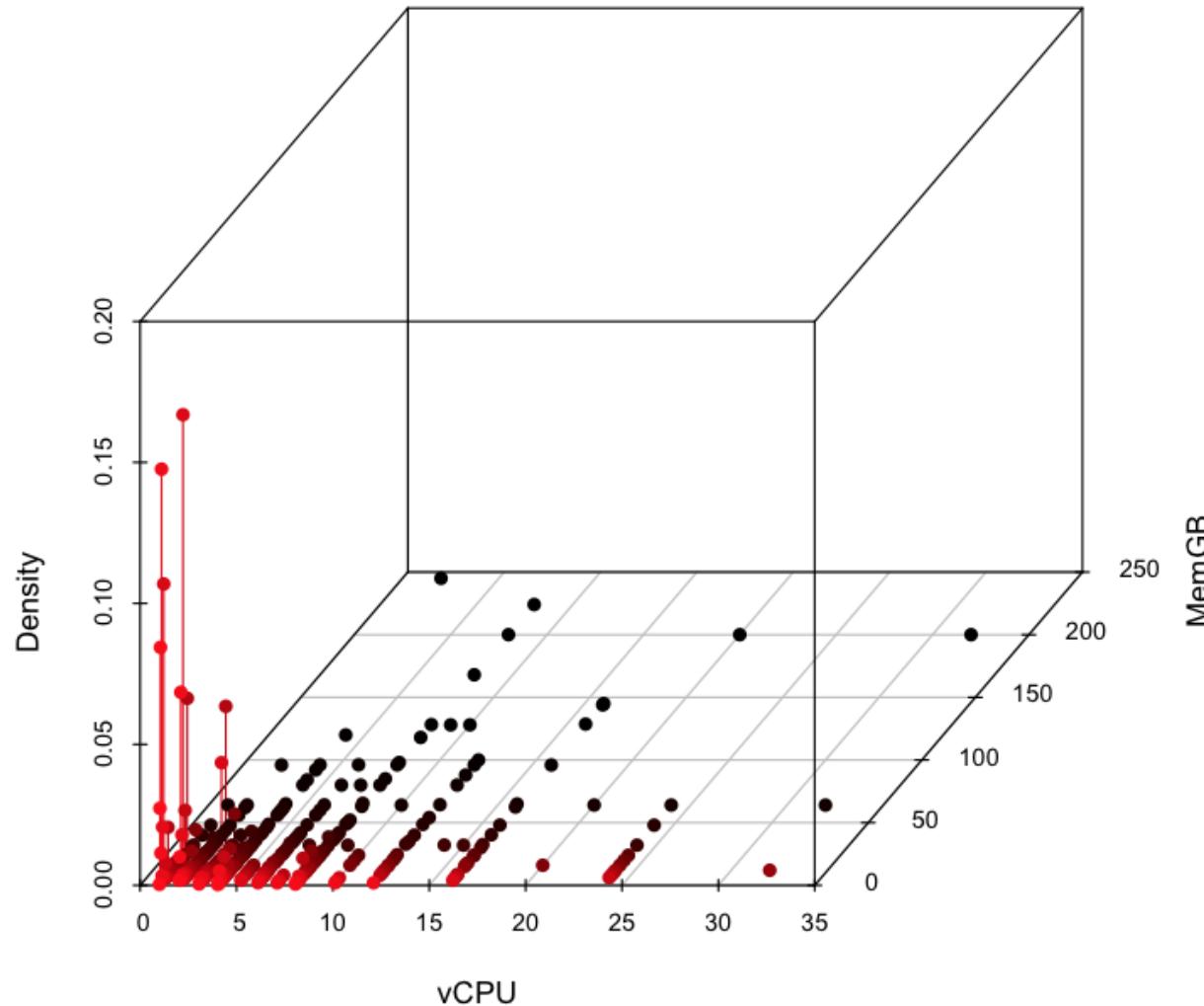
Cross-User Analysis Setup



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