Applied Convergence of Supercomuting and Analytics

Rob Vesse and Vincent Pel

rvesse@cray.com
vpel@cray.com

About Cray

Seymour Cray founded Cray Research in 1972

- Cray Inc. formed in April 2000

Cray Inc.

- Headquartered in Seattle, WA
- NASDAQ: CRAY
- Over 1,300 employees across 30 countries

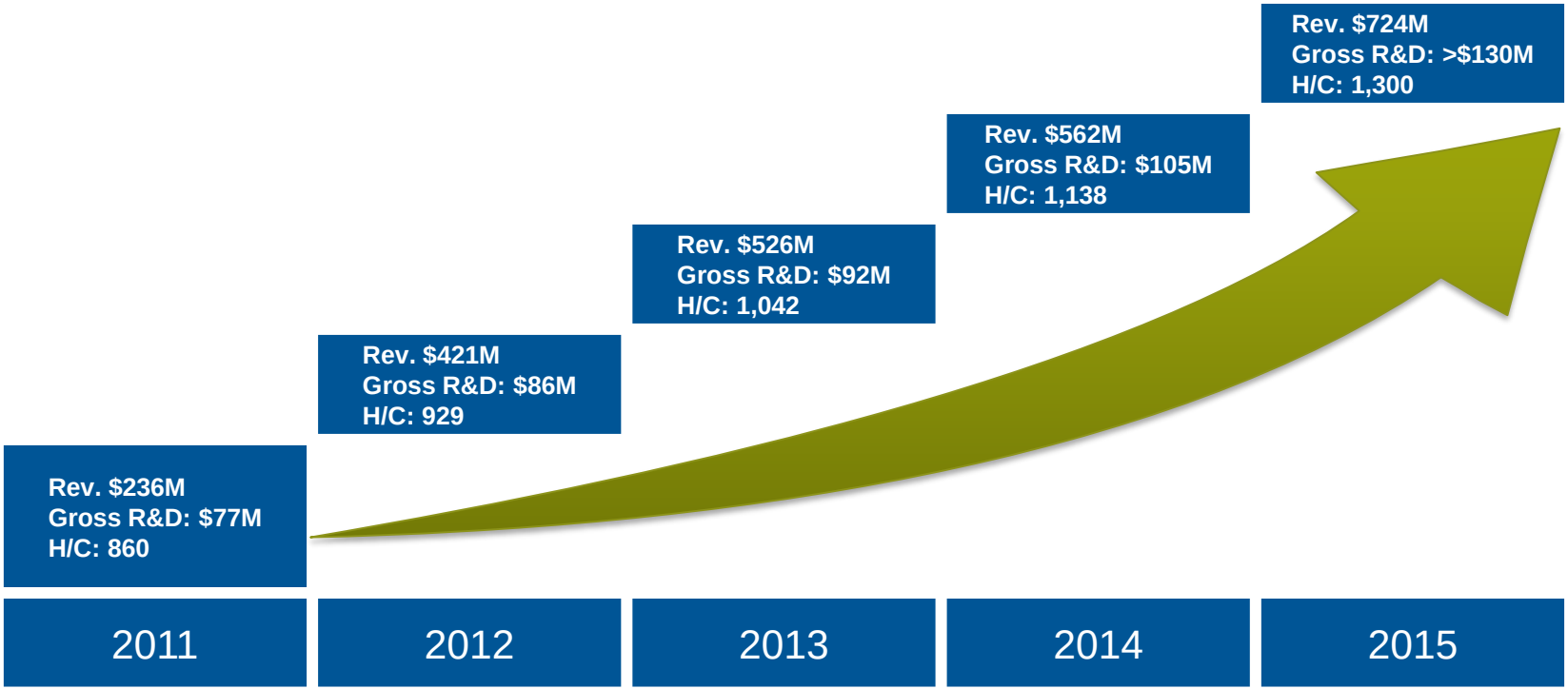
Three Focus Areas

- Computation
- Storage
- Analytics

Seven Major Development Sites:

- Austin, TX
- Chippewa Falls, WI
- Pleasanton, CA
- Bristol, UK
- San Jose, CA
- Seattle, WA
- St. Paul, MN

Continuing Financial Strength



Supercomputing Leadership

Top 500 Supercomputers in the World
June 2016

TOP500 SUPERCOMPUTER SITES

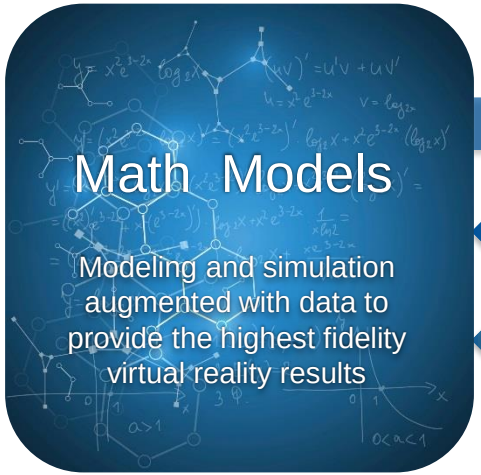
	Top 10	Top 50	Top 100	Top 500
Cray Systems	5	18	30	60
Vendor Rank	#1	#1	#1	#3

Cray Customers by Verticals

Cray's Vision:

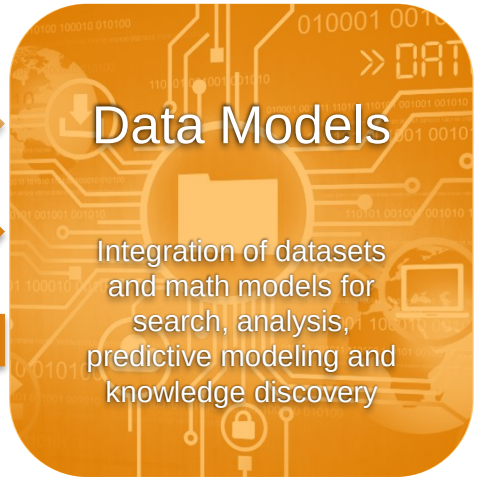
The Fusion of Supercomputing and Big & Fast Data

Modeling The World



Data-Intensive Processing

Modeling and simulation augmented with data to provide the highest fidelity virtual reality results

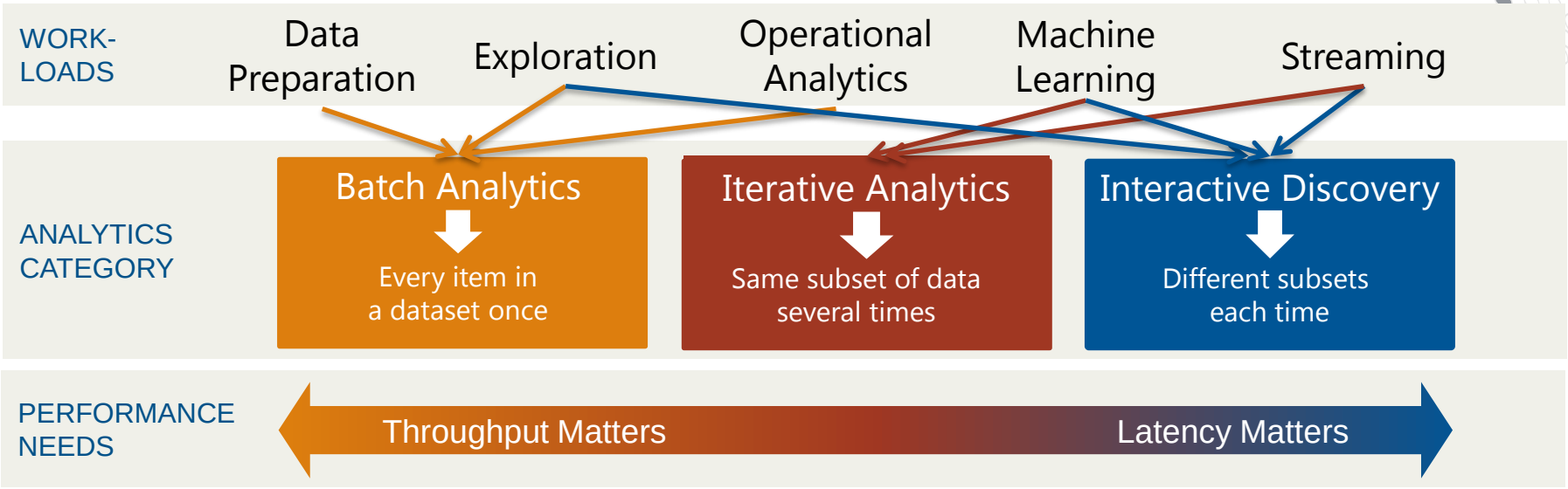


Compute | **Store** | **Analyze**


Applied Convergence of
Supercomputing and Analytics

Spectrum of Analytics Workloads

Need the Agility to Efficiently Run Mixed Workloads



Traditional Big Data Solutions have:

- Standalone frameworks
- Silo'd data and functions
- Costly data integration
- Poor performance at scale
- Single use

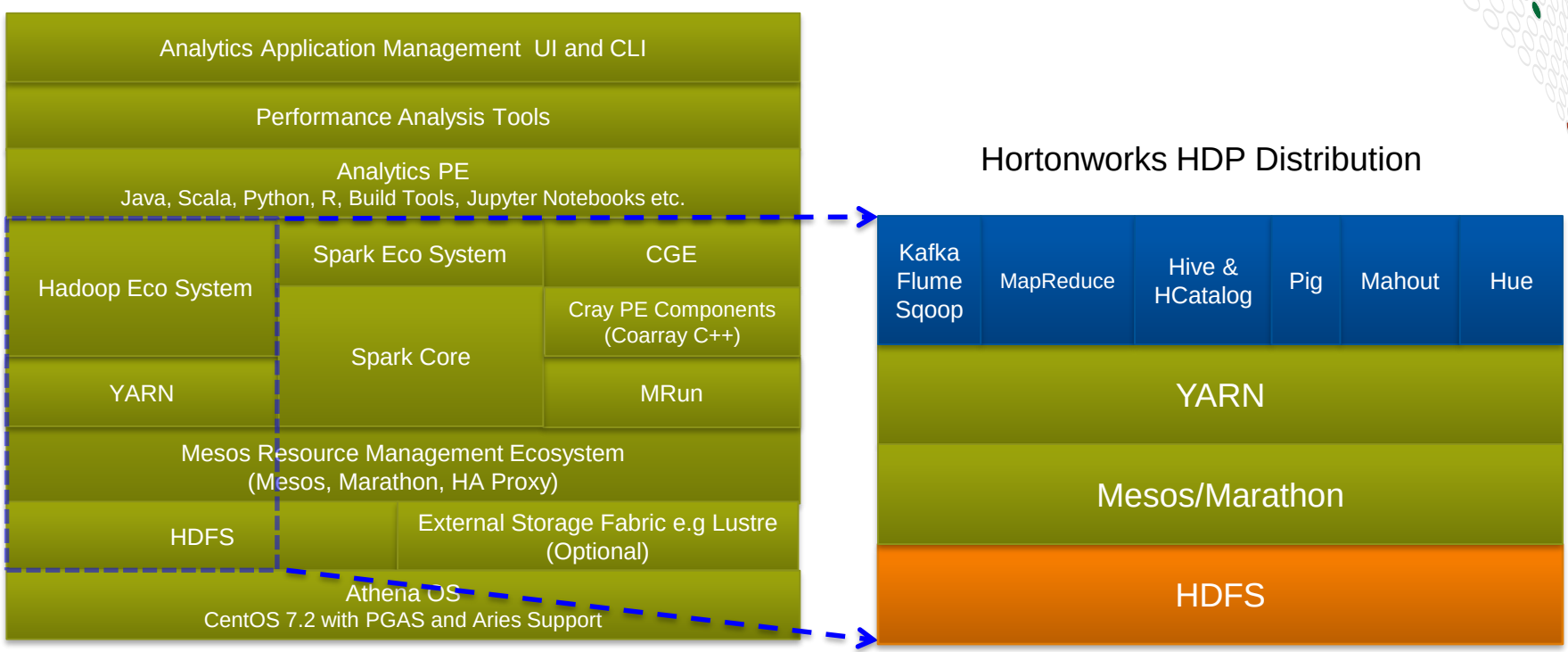
Why a single platform?

- Increasingly real-world workloads span the spectrum of workload styles
- Our vision is for big data and HPC workloads to co-exist on the same system
- At scale a lot of big data problems look very similar to HPC problems
  - Similar hardware and software requirements
  - Similar scaling challenges
- Applying HPC technologies let us accelerate and expand the capabilities of big data technologies

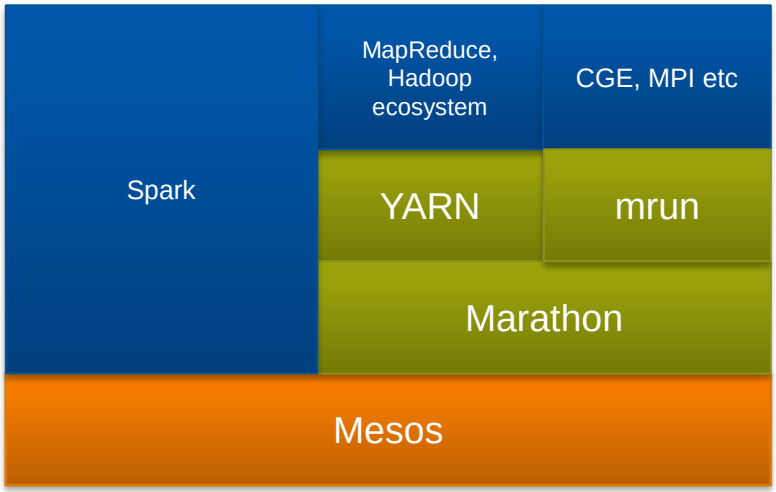
Urika-GX Hardware

- 42U Rack
- Up to 48 total nodes
  - Up to 44 Compute nodes
  - 2 Login nodes
  - 2 IO nodes
- Up to 1584 processor cores
- Up to 22TB RAM
- Up to 352TB Node Local Storage
  - Evenly split between spinning and flash disks
- Support for external storage via dedicated Infiniband or Ethernet links
- Cray Aries™ Interconnect
  - Support RDMA for HPC applications
  - IP over Aries to accelerate normal network traffic
- Gigabit Ethernet operational network plus dual 10 Gigabit Ethernet links to external network

Software Overview



Cluster Resource Management

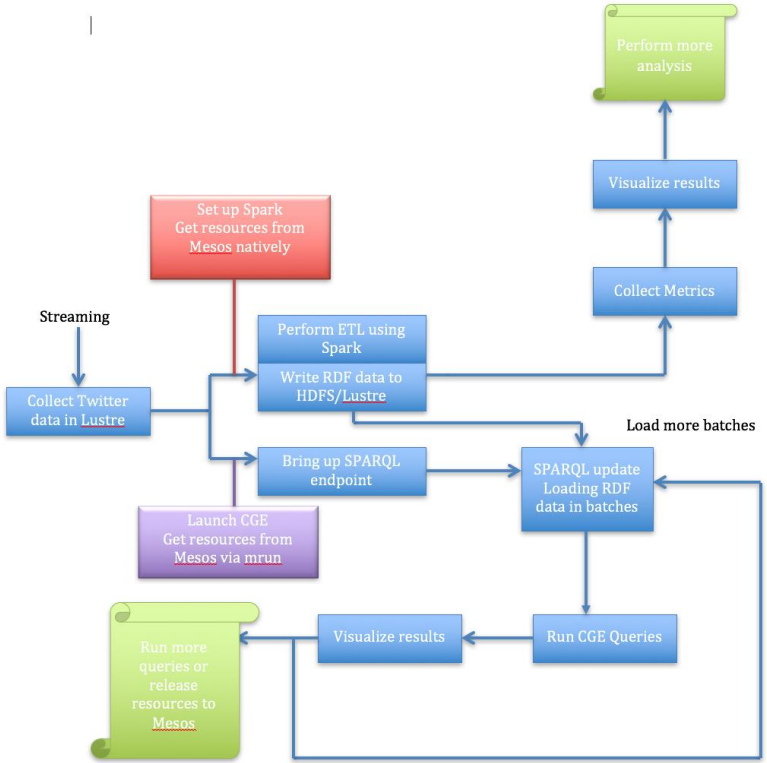


- Analytics Frameworks require a variety of resource managers
  - No one size fits all
  - Hadoop Ecosystem built around YARN
  - Spark can work with Mesos directly
  - CGE and HPC applications need Aries support
- Marathon is used for long running Mesos jobs
  - YARN sub-cluster dynamically created
  - mrun creates HPC sub-cluster

Use Case – Social Media Sentiment Analysis

- Analyse social media sentiment and determine the key influencers for targeted marketing
  - Customer use case from retail/entertainment sector
- Input data
  - Twitter firehose
- Spark Streaming for ETL
  - Filter for desired search terms, hash tags and corporate account mentions
  - Transform Data format for further analysis
- CGE for Graph Analytics
  - Use graph algorithms to extract communities and their key influencers

Use Case – Social Media Sentiment Analysis



Use Case - Cybersecurity

- Large enterprises face increasing risks of cyber attack with a potential to cause large harm to the business
- Vast amounts of data are gathered, challenge is meaningfully analysing it to identify threats and attacks ASAP
  - Machine Logs
  - Firewall Logs
  - Netflow
  - Intrusion Detection Systems (IDS)
  - Black/White lists
- Different kinds of data need different kinds of analytics
- But need to combine data to spot increasingly sophisticated threats

Questions?

rvesse@cray.com
vpel@cray.com

Come talk to us at Cray Booth #26