Cray Today...

Cray Inc.
- Nasdaq: CRAY
- 950+ employees across 30 countries
- Headquartered in Seattle, WA

Serving
- Fortune 1000 Companies
- Governments
- Research Consortia
- Companies such as yours

Business
- High Performance & Cluster Computing
- Storage and Data Management
- Big Data Analytics Solutions – YarcData Company
The Cray Mission

Building tools to solve the world's most challenging problems
Modeling The World
Cray Supercomputers solving “grand challenges” in science, engineering and analytics

Data Models
Integration of datasets and math models for search, analysis, predictive modeling and knowledge discovery

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

Data-Intensive Processing
High throughput event processing & data capture from sensors, data feeds and instruments
Integrated Computational Environments that enable shortened product development cycles, improve product quality and lower product costs.
Fusion of Big Data and Big Compute is characterized by highly complex approaches that are multi-disciplinary, multi-dimensional, multi-scale and multi-physics.

**Discover**
- Graph and Search Appliances
- Examples:
  - Precision Medicine
  - New Drug Repurposing
  - Human Brain Project
  - Cybersecurity

**Capture & Manage**
- Storage, Data Migration, and Archiving Solutions
- Examples:
  - Square Kilometer Array
  - Digital Oil Fields

**Model**
- Simulation and Analysis
- Examples:
  - Real-Time Weather
  - Earth System Informatics
  - Drug Discovery
  - High Fidelity Simulation
  - Next Generation Sequencing
Cray HPC Solutions For Manufacturing

- **Cray’s performance optimized environments for virtual prototyping and advanced analytics enable:**
  - Greater simulation fidelity, combined with greater system throughput, increases the value of simulation in the product design process
  - Shortened product development cycles, improved product quality and lower product costs
  - Companies to make smarter decisions at the earliest stages of design process and improve overall competitiveness

*Image Courtesy of Altair Engineering, Inc.
CFD by AcuSolve*
CAE trends driving HPC requirements

- “Extreme Fidelity”
  - Enhanced physics and larger models, e.g., 1 Billion cell models
  - Large models scale better across compute cores
- Design optimization methods
  - Many simulations required to explore the design space
  - Multiple runs can require 100x compute power
- Robust design
  - Looking for the “best solution”

“Future performance depends on highly scalable parallel software”

* ref: ANSYS CFD presentation
## Industrial Innovation Every Day

<table>
<thead>
<tr>
<th>Aircraft design</th>
<th>Consumer products</th>
<th>Engine cycle-to-cycle variation</th>
<th>Jet engine efficiency</th>
<th>Li-ion batteries</th>
<th>Long-haul truck fuel efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulating takeoff and landing scenarios improved a critical code for estimating characteristics of commercial aircraft, including lift, drag, and controllability</td>
<td>Leadership computing and molecular dynamics software advanced understanding of chemical processes that can limit product shelf life</td>
<td>Emerging model of engine cyclic variation will apply thousands of processors to a challenging problem</td>
<td>Accurate predictions of atomization of liquid fuel by aerodynamic forces enhance combustion stability, improve efficiency, and reduce emissions</td>
<td>New classes of solid inorganic Li-ion electrolytes could deliver high ionic and low electronic conductivity and good electrochemical stability</td>
<td>Simulations reduced by 50% the time to develop a unique system of add-on parts that increases fuel efficiency by 7–12%</td>
</tr>
</tbody>
</table>
The “Big Data” Landscape

Data Warehouses + Extensions
(Oracle, Teradata, Greenplum, DB2)

NoSQL Databases
(MongoDB, CouchBase, DynamoDB, AsterData)

Big Data Solutions

Hadoop / MapReduce
(Cloudera, HortonWorks, MapR, Intel)

Graph Analytics
( Neo4j, AllegroGraph, Objectivity, Virtuoso)

These solutions can compete, but also can be very complementary as each has strengths & weaknesses
The Baby Elephant in the Room

Current Perception of Hadoop
- Synonymous with Big Data and openness
- Excellent for the three “V’s” (Volume, Velocity, & Variety)
- Capable of huge scale with ad-hoc infrastructure

Current Reality of Hadoop
- Many experimenting (64% of F1000, in 2012)
- Much expertise in Warehousing – little beyond that
- Bottlenecks Data Scientist, so performance not yet an issue

Current Trajectory of Hadoop
- Industry Momentum—Vendors, analysts, organizations, etc.
- More Users – Beyond Data scientists, Business & Ops users
- More Complexity – Near real-time, complex algorithms, etc.

Hadoop widely perceived as high potential, not yet high value, but that’s about to change…
Untapped Hadoop Potential: Most Organizations just scratching the Surface

In 2013 – Most organizations, using Hadoop, are approximately here.

High Value Hadoop - Ideally, most organizations aspire to be here, but...

In 2013 – Most organizations, using Hadoop, are approximately here.
## Realizing Hadoop Potential: Increasing Value adds Complexity

<table>
<thead>
<tr>
<th>Type or Users</th>
<th>Store</th>
<th>Report</th>
<th>Analyze</th>
<th>Monitor</th>
<th>Predict</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type or Users</strong></td>
<td>Data Scientists</td>
<td>Data Scientists</td>
<td>Data Scientists</td>
<td>Data Scientists</td>
<td>Data Scientists</td>
</tr>
<tr>
<td><strong># of Users</strong></td>
<td>Few</td>
<td>Few</td>
<td>Medium</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td><strong>Algorithms</strong></td>
<td>Few Crude</td>
<td>Few Basic</td>
<td>Complex</td>
<td>Many Complex</td>
<td>Many Advanced</td>
</tr>
<tr>
<td><strong>Latency</strong></td>
<td>Infrequent Batch</td>
<td>Frequent Batch</td>
<td>Frequent Batch</td>
<td>Frequent Batch</td>
<td>Frequent Batch</td>
</tr>
<tr>
<td><strong>Data Types</strong></td>
<td>Unstructured</td>
<td>Unstructured</td>
<td>Unstructured</td>
<td>Unstructured</td>
<td>Unstructured</td>
</tr>
<tr>
<td><strong>Volume</strong></td>
<td>Medium-High</td>
<td>Medium-High</td>
<td>High Value</td>
<td>High Value</td>
<td>High Value</td>
</tr>
<tr>
<td><strong>Value</strong></td>
<td>Low</td>
<td>Medium-High Value</td>
<td>High Value</td>
<td>High Value</td>
<td>High Value</td>
</tr>
</tbody>
</table>
Cray Cluster Supercomputers for Hadoop: Purpose-Built, Turnkey, Hadoop Solutions

Best Hadoop Distribution
- Security – Comprehensive, and fast, encryption
- Performance – Faster Hive, Cache acceleration, etc.
- Management – Intel Manager for Hadoop Software

Performance of a Cray
- Proven HPC – Cray HPC technology and expertise
- Vast Scale – Grow to meet any mission requirements
- Holistic Design – Balanced: Compute, networking, & Storage

Turnkey Solution
- Reliable – Rapid ROI… runs as-advertised
- Support – One throat to choke, for the whole stack
- Maintenance – Update & evolve, without concerns

High Value Hadoop
- Performance – Power to accommodate current & future goals
- Reliability – Will meet any challenge, without surprises
- Maintenance – Easy to maintain & accommodate change
Current Big Data approaches are based on SEARCH...

**Search:**
Needle-in-a-Haystack Paradigm

- When we know what we’re looking for
- We can break up the problem

**Current Big Data Approaches:**
Partition and Scale out
**YarcData focuses on DISCOVERY — not search**

Needle-in-a-Needlestack

When we don’t know what we’re looking for…

We can’t break up the problem

Current Big Data Approaches do NOT enable Discovery

“...when the purpose of the system is discovery of relationships, not extracting information from already known interrelations, achieving satisfactory performance is difficult.”

“When the relationships among data are mysterious, and the nature of the inquiries unknown, no meaningful scheme for partitioning the data is possible.”

(Source: Gartner Report, YarcData’s Urika™ Shows Big Data Is More Than Hadoop and Data Warehouses, September 2012)
"We are also impressed with... its product strategy, which combines the advantages of a pre-integrated hardware appliance with the flexibility of a subscription model."
Scalable Storage & Data Management Solutions

- **Cray Cluster Connect**
  - Complete storage and data management solutions for cluster environments

- **Cray Data Management Platform**
  - Lustre File Systems by Cray
  - Cray Development and Login
  - Cray Management Services
  - Cray Migration and Archive

- **Cray Storage Systems**
  - Cray Sonexion™
  - Lustre Files Systems by Cray

- **Cray Monitoring and Management Tools**
building computational tools that help change the world!