Cray and Cloudera Simplify Big Data Analytics

The Cray® Urika-XA™ platform certified with Cloudera delivers high performing analytics capability in a single pre-integrated, open environment

Big data today holds immense value. But the complexity of standing up a Hadoop cluster at scale poses a major obstacle in getting to analytic insight. Many organizations do not maximize their value from big data because too much time and effort is spent on solving Hadoop infrastructure challenges.

The Cray Urika-XA extreme analytics platform simplifies high performance analytics. Pre-integrated with Cloudera, the Urika-XA system combines the benefits of a turnkey analytics appliance with a flexible, open platform that users can modify with their choice of complementary tools. Optimized for compute-heavy, memory-centric analytics, the Urika-XA platform incorporates memory-storage hierarchies and fast interconnects for excellent performance at scale on both current and emerging analytics applications. This single platform design reduces the analytics footprint and total cost of ownership — and delivers valuable business insights now and into the future.

Cray Urika-XA Platform Overview

Turnkey Platform

Choosing an analytics infrastructure typically comes down to a choice between a rigid appliance or the investment of time and money required to build a custom solution. With the Urika-XA platform, customers get a robust, turnkey platform that eliminates both lock-in and the complexity of integrating individual components — and delivers immediate advantages such as rapid time to value and increased productivity. Stand up a production Hadoop cluster quickly and manage it without an army of experts. Cray Urika-XA provides:

- Integrated hardware and software platform
- Performance-tuned Cloudera software stack
- Single point of support from Cray
- Unified management pane for hardware and software
- Datacenter standards compliant

Reduced Datacenter Footprint

Scale-out Hadoop clusters have the potential to consume precious datacenter space. Building multiple clusters to accommodate a variety of analytic workloads is another contributor to cluster sprawl. The high-density Urika-XA analytics system is a single, multi-use platform designed to consolidate analytics workloads, supporting a variety of processing requirements and reducing the footprint and TCO of the analytics environment:

- 48-node Hadoop cluster configured in a rack
- Replace multiple standard racks with one Urika-XA rack
- Multi-purpose platform for high-throughput and low-latency workloads
- Includes both Hadoop (HDFS) and POSIX-compliant storage
Benefits of the Cray® Urika-XA™ Extreme Analytics Platform

Rapid Time to Value
- Turnkey, integrated hardware and software platform with pre-configured Cloudera stack

Ease of Management
- Unified system management pane and single point-of-support from Cray

Smaller Datacenter Footprint
- High compute and memory density allows the replacement of multiple standard racks

Future-Proof Architecture
- Designed for emerging latency-sensitive, in-memory workloads

Benefits of Cloudera
- Powerful – Store, process, and analyze all your data to drive competitive advantage
- Efficient – Hadoop unifies compute and data to improve operational efficiency
- Open – 100% open source: CDH is the world’s most popular open source distribution powered by Apache Hadoop
- Simple – Easy to deploy and operate with centralized administration
- Compatible – Leverage your existing investments for rapid adoption and lower TCO
- Economical – Rethink the economics of data management with an open source platform on industry standard hardware - up to 90% more cost effective than traditional solutions
- Enterprise Ready – Equipped with critical capabilities to support mission-critical operations

Future-Proof Architecture
Analytics on big data is trending towards lower-latency processing options such as interactive SQL, stream processing, and in-memory computing. Traditional big data architecture was originally set up for batch processing and cannot support these emerging analytics workloads. Designed for next-generation analytics, the Urika-XA platform brings the versatility required by the ever-changing big data landscape:
- Architected for emerging latency-sensitive, in-memory workloads with large memory size and unique combination of InfiniBand performance networking and high-speed SSD storage
- Open platform allows users to install analytics tools of choice

<table>
<thead>
<tr>
<th>Component</th>
<th>Specifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compute nodes</td>
<td>48 compute nodes</td>
</tr>
<tr>
<td></td>
<td>• Intel® Xeon® processors</td>
</tr>
<tr>
<td></td>
<td>• 1,536 cores, 6TB RAM per rack</td>
</tr>
<tr>
<td></td>
<td>• 800GB PCIe SSD per node</td>
</tr>
<tr>
<td>Interconnect</td>
<td>FDR InfiniBand</td>
</tr>
<tr>
<td>Storage</td>
<td>206TB total storage per rack</td>
</tr>
<tr>
<td></td>
<td>• 120TB Cray Sonexion® 900 POSIX storage</td>
</tr>
<tr>
<td></td>
<td>• 38TB SSD</td>
</tr>
<tr>
<td></td>
<td>• 48TB HDD</td>
</tr>
<tr>
<td>Software</td>
<td>• Cloudera Enterprise</td>
</tr>
<tr>
<td></td>
<td>• Apache Spark</td>
</tr>
<tr>
<td></td>
<td>• Urika-XA Management System</td>
</tr>
</tbody>
</table>

About Cloudera
Cloudera is revolutionizing enterprise data management by offering the first unified Platform for big data, an enterprise data hub built on Apache Hadoop. Cloudera offers enterprises one place to store, access, process, secure, and analyze all their data, empowering them to extend the value of existing investments while enabling fundamental new ways to derive value from their data. Cloudera’s open source big data platform is the most widely adopted in the world, and Cloudera is the most prolific contributor to the open source Hadoop ecosystem. As the leading educator of Hadoop professionals, Cloudera has trained over 40,000 individuals worldwide. Over 1,600 partners and a seasoned professional services team help deliver greater time to value. Finally, only Cloudera provides proactive and predictive support to run an enterprise data hub with confidence. Leading organizations in every industry plus top public sector organizations globally run Cloudera in production. www.cloudera.com

About Cray
Supercomputing leader Cray provides high-performance data analytics and discovery platforms that help organizations derive value from big data. Cray’s analytics products accelerate time to insight by unifying hardware, software and management into turnkey, converged platforms. Organizations enjoy superior analytic productivity through Cray’s implementation of innovative memory and storage strategies and use of performance networking in its systems. Founded in 1972, Cray has focused exclusively on supercomputing and analytics technologies for over 40 years. www.cray.com