

Deploying Hadoop for Strategic Gain with SGI and Cloudera

Modern enterprises must increasingly capture, store, and analyze truly massive amounts of unstructured data in order to expose key insights that help them compete and innovate. The use of Apache Hadoop in particular has grown dramatically, and with it, the need for scalable, cost-effective, and flexible systems. Together with Cloudera and select analytics partners, SGI delivers highly effective systems for Hadoop giving organizations the confidence to deploy Big Data solutions that meet their specific needs.



Harnessing the Massive Growth of Unstructured Data

Once the province of on-line firms, harnessing data for strategic analytics is increasingly essential to diverse organizations and missions:

- Manufacturers need to optimize process yields based on sensor data and improve products by analyzing customer telematics.
- The intelligence community needs to coordinate diverse intelligence sources and trend data to predict events.
- In the life sciences, sequence data from bioinformatics applications needs to be mapped to reference genomes.
- Financial services firms need to conduct market analysis as well as fraud identification, risk analysis, and processing for compliance purposes.
- Retailers need to perform customer buying analysis and inventory optimization.

To address the needs of these and myriad other disciplines, SGI and Cloudera have partnered to deliver solutions that help take the time and risk out of deploying Apache Hadoop. SGI has extensive experience deploying large data management systems, and fully understands the end-to-end data management lifecycle. Certified by Cloudera, SGI Hadoop solutions are based on SGI Rackable servers powered by the Intel® Xeon® processor E5 family, and include:

- A complete and flexible Hadoop ecosystem based on Cloudera's Distribution including Apache Hadoop (CDH) or Cloudera Enterprise

- A choice of flexible deployment options that let organizations get Apache Hadoop up and running quickly
- A choice of commercial and community Linux distributions and data analytics solutions through industry-leading partners

SGI Big Data Leadership

SGI has been one of the leading suppliers of Apache Hadoop servers since the technology was first introduced – including both large commercial enterprise deployments and optimized clusters in key government agencies. SGI has also been a leader in complementary high speed, high-capacity data ingest and large-scale in-memory analytics using SGI UV and previous-generation technologies for over 15 years. Perhaps most importantly, SGI is ideally positioned to deliver clusters that fit an organization's unique requirements, and has built and deployed Apache Hadoop installations as large as 40,000 nodes and individual clusters as large as 4,000 nodes.

- **Delivering unprecedented speed and scale.** With a strong background in high-performance and technical computing, SGI pushes the envelope in terms of both performance and scalability, as evidenced by continuous completion of world-record benchmarks. SGI provides a choice of high-performance Ethernet or InfiniBand backbone networking, and superior cluster manageability at scale using SGI Management Center.

- **End-to-end solutions for data analytics.** With a broad range of systems and storage solutions, SGI provides systems for end-to-end data analytics. SGI can provide complementary solutions for essential functions such as data ingest and archive, and systems with both horizontal (scale-out) and vertical (scale up) scalability.
- **The world's only power-optimized Hadoop cluster.** With growing power costs, data centers increasingly require solutions that optimize performance at the lowest possible power consumption. SGI provides the world's only power-optimized Hadoop cluster, offering fine-grained, dynamic, and policy-driven power management that avoids the hassle and risk of powering down individual servers.
- **Plug In and Go!** SGI understands that organizations need solutions that are ready to deploy and that come with enterprise-class 7x24 services. SGI provides true factory-integrated Hadoop solutions that include servers, storage, networking, and installed Cloudera software, delivered ready to run.

A Flexible and Complete Hadoop Solution

SGI systems for Apache Hadoop combine enterprise ready distributions from Cloudera with flexible deployment options and an ecosystem of data analytics applications.

CDH from Cloudera – The Enterprise-Ready Hadoop Stack and an Expanding SGI Analytics Ecosystem

Individual organizations have their own unique analytics objectives that must be served by their Apache Hadoop infrastructure. By partnering with Cloudera, SGI provides the premier Apache Hadoop®-based distribution in both commercial and non-commercial environments.

- CDH is Cloudera's market-leading, 100% open source Hadoop platform that's built for the enterprise. CDH delivers the core elements of Hadoop – scalable storage and distributed computing – as well as a number of other enterprise capabilities such as security, high availability, and integration with a broad range of hardware and software solutions. CDH is the most widely deployed distribution of Hadoop and is currently run at scale in production environments across a variety of different industries and use cases.
- Cloudera Enterprise is a subscription offering that includes Cloudera Manager, combined with technical support. Cloudera Enterprise enables data-driven enterprises to deploy & manage CDH cost-effectively in production with repeatable success.

- Complementing Cloudera's offerings, SGI has expanding partnerships with industry-leading analytics, visualization, and tool vendors. Full feature trial applications from Datameer, Kitenga, Quantum4D, and Pentaho are all pre-installed on SGI Hadoop solutions, and can be easily extended for production use.

Multiple Deployment Options

In addition to flexible software choices, SGI lets organizations choose deployment options that best fit their needs, (Table 1).

Table 1. SGI provides flexible Hadoop cluster options

SGI Hadoop Cluster Option	Capacity	Number of Servers
Half Rack Starter Kit	• 160 TB	• 20
Full Rack Starter Kit	• 320 TB	• 40
Multi-rack Reference Implementation	• Petabytes	• Any
Design to Order	• Petabytes	• Any

- **SGI Hadoop Starter Kits.** Offering the fastest time to production, SGI Hadoop Starter Kits are pre-configured, factory integrated, and certified by Cloudera. With both half-rack and full-rack options based on SGI Rackable servers using the highest-performance Intel® Xeon® processor E5 family, these kits are purpose-designed to run Apache Hadoop. Multi-rack systems using SGI's reference implementation are also available.
- **Design to Order Hadoop Clusters.** For optimal multi-rack deployments, SGI Design to Order (DTO) clusters are custom designed to meet exact customer needs. Custom considerations can include performance, density, power, cooling, networking, and cost requirements.

Conclusion

Together with Cloudera and other key analytics partners, SGI delivers Hadoop leadership at speed and scale. This integrated approach gives organizations the confidence to rapidly deploy Apache Hadoop while also drastically reducing both cost and risk. A choice of deployment options helps ensure that systems are designed and deployed to meet the needs of each individual organization.

Ultimately, SGI delivers Apache Hadoop as a part of a complete end-to-end data management solution that can encompass data ingest, data consolidation, data analytics and visualization, data archiving and storage. To get started right away, contact SGI to configure an easy-to-deploy SGI Hadoop Starter kit, or speak with an SGI representative about specifying a Design to Order Hadoop cluster.

URL: <http://www.sgi.com/products/hadoop/>

Global Sales and Support: sgi.com/global

©2012 Silicon Graphics International Corp. All rights reserved. SGI and the SGI logo are registered trademarks or trademarks of Silicon Graphics International Corp. or its subsidiaries in the United States and/or other countries. All other trademarks are property of their respective holders. 13072012 4373

cloudera

