

# Cisco UCS Common Platform Architecture with StackIQ: Automated Big Data Infrastructure Solution for the Enterprise

Solution Brief

Together, StackIQ Cluster Manager and Cisco Unified Computing System™ (Cisco UCS®) deliver a automated big data infrastructure solution for the enterprise that automates server and cluster configuration and deployment after the solution has been set up.

### Highlights

#### Reduced Time to Production

- Cisco Unified Computing System™ (Cisco UCS®) makes it easy to build the hardware infrastructure you need for big data, and StackIQ Cluster Manager lets you choose the operating system, big data software, and other software you need and set critical parameters to build your Hadoop cluster automatically using the parallel StackIQ Avalanche installer.

#### Ease of Operation

- Cisco UCS and StackIQ Cluster Manager communicate with each other through the Cisco UCS open XML API and provide all the management tools you need to maintain the cluster, keeping it healthy and operating efficiently.

#### Highly Scalable Platform

- Cisco UCS provides a highly scalable unified computing, networking, and integrated management platform optimized for big data clusters, and StackIQ Cluster Manager provisions and configures the desired big data cluster software on the Cisco UCS platform automatically.

#### Extensibility

- The modular architecture of StackIQ Cluster Manager enables you to customize your modular Cisco UCS cluster with software to meet your particular needs. For example, switching between Hadoop distributions is simple using StackIQ.

#### Enterprise-Class Support

- Reference configurations from Cisco offer confidence and help accelerate implementation of successful deployments with worldwide enterprise-class support from Cisco and StackIQ.

Cisco UCS and StackIQ Cluster Manager work better together to create manageable clustered infrastructure in the data center. The combination of StackIQ's cluster management software and Cisco UCS provides a powerful big data infrastructure solution for enterprise data center environments. Each has its own strengths, and when combined they provide exceptional capability for enterprise-class clustered infrastructure. Cisco UCS is an excellent hardware infrastructure solution, providing flexible network computing and storage. StackIQ's software suite provides a unique, complete cluster management solution that supplements Cisco UCS Manager. StackIQ Cluster Manager integrates transparently with Cisco UCS to install, configure, and deploy all the software in the Cisco UCS cluster.

The Cisco UCS open XML API provides automated, real-time discovery of system hardware configuration information to StackIQ Cluster Manager. That information is used to provision the cluster and automate ongoing hardware change management. The Cisco and StackIQ solution delivers transparent management capability for bare-metal machines, the network, the operating system, and big data software. It addresses the challenge of big data deployments in the enterprise by making clusters reliable and repeatable.

## Enterprise Use Cases

The most common application of clustered infrastructure in the enterprise data center is big data. Today these clusters typically run one of the popular Apache Hadoop distributions or a NoSQL distribution. The StackIQ solution for Cisco UCS supports multiple Hadoop and NoSQL distributions from industry leaders, so customers can choose the big data software that best meets their needs.

Using Hadoop, organizations can move large volumes of complex and relational data into a single repository in which raw data is always available. With its low-cost commodity servers and storage repositories, Hadoop enables this data to be affordably stored and retrieved for a wide variety of analytic applications that can help organizations increase revenue by extracting value such

as strategic insights, solutions to challenges, and ideas for new products and services. By dividing big data into multiple parts, Hadoop allows the simultaneous processing and analysis of each part on servers throughout the cluster, greatly increasing the efficiency of queries and reducing response time.

The use cases for Hadoop are many and varied, including public health, stock and commodities trading, sales and marketing, product development, and scientific research. For the business enterprise, Hadoop use cases include:

- **Data processing:** Hadoop allows IT departments to extract, transform, and load (ETL) data from source systems and to transfer data stored in Hadoop to and from a database management system for the performance of advanced analytics. Hadoop is also used for the batch processing of large quantities of unstructured and semistructured data.
- **Network management:** Hadoop can be used to capture, analyze, and display data collected from servers, storage devices, and other IT hardware to allow administrators to monitor network activity and diagnose bottlenecks and other issues.
- **Retail fraud:** By monitoring, modeling, and analyzing high volumes of data from transactions and extracting features and patterns, retailers can help prevent credit card account fraud.
- **Recommendation engine:** Web companies can use Hadoop to match and recommend users to one another or to products and services based on analysis of user profiles and behavioral data.
- **Opinion mining:** Used in conjunction with Hadoop, advanced text analytics tools analyze the unstructured text of social media and social networking posts, including Tweets and Facebook posts, to determine user sentiment related to particular companies, brands, or products; the focus of this analysis can range from the macro level to the individual user.
- **Financial risk modeling:** Financial firms, banks, and other companies use Hadoop and data warehouses for the analysis of large volumes of transactional data to determine risk and exposure of financial assets, prepare for potential “what-if” scenarios based on simulated market behavior, and rate potential clients for risk.
- **Marketing campaign analysis:** Marketing departments across industries have long used technology to monitor and determine the effectiveness of marketing campaigns. Big data allows marketing teams to incorporate higher volumes of increasingly detailed data, such as click-stream data and call-detail records, to increase the accuracy of analysis.
- **Customer influencer analysis:** Social networking data can be mined to determine which customers have the most influence over others within social networks, to help enterprises determine which customers are most important and influential.
- **Customer experience analysis:** Hadoop can be used to integrate data from previously isolated customer interaction channels (for example, online chats, blogs, and call centers) to gain a complete view of the customer experience. This view enables enterprises to understand the impact that one customer interaction channel has on another so that enterprises can optimize the entire customer lifecycle experience.
- **Research and development:** Enterprises such as pharmaceutical manufacturers use Hadoop to sort through enormous volumes of text-based research and other historical data to assist in the development of new products.
- **Multiple-use clusters:** Enterprise data centers need to maintain agility within the big data infrastructure to meet rapidly changing requirements from the businesses they support. By implementing a dynamic, flexible cluster infrastructure, organizations can accommodate separate instances of Hadoop, NoSQL databases, and other evolving big data applications simultaneously.

## StackIQ Cluster Manager and Cisco UCS: Excellent Big Data Cluster Solution

Organizations of all types are deploying Hadoop and NoSQL solutions to gain a competitive advantage. There is now a range of hardware and software solutions to choose from, and they are being deployed in data centers everywhere. Today's enterprise data center operations require a complete solution to operate effectively. StackIQ combined with Cisco UCS manager provides an integrated solution that automates the deployment and management of Hadoop and NoSQL clusters - from bare metal all the way to a working system.

### One Customer Experience

This use case shows how one of our customers experienced the Cisco UCS and StackIQ solution. This major financial services company asked for a trial of StackIQ Cluster Manager for its big data proof-of-concept project using Cisco UCS. Our team installed the cluster manager and used it to install and configure all the cluster nodes in about 20 minutes. The customer seemed surprised, and when we asked why, the customer said, "How did you do that? We have been struggling to configure one of those machines for over two weeks now and we couldn't get it to install. We've been struggling with the configuration of the LSI controller." Because StackIQ Cluster Manager uses an install-from-scratch model, it had no problem with the errant controller. It did its job and created another satisfied Cisco UCS and StackIQ solution customer.

## Bare-Metal Provisioning

Some solutions assume that you are starting with a cluster that has already been provisioned with an operating system, and that each server was properly configured to work on the network. StackIQ takes a different approach.

It assumes that there is nothing on the servers after Cisco UCS setup is complete. The StackIQ provisioning tool automatically polls Cisco UCS to synchronize its host database by using the Cisco UCS XML open API, and it installs all the software and configures all the services. Starting with bare metal (empty servers), StackIQ

Cluster Manager installs the operating system, libraries, and applications software such as Hadoop. It also configures the network, firewall, disks, and application services, such as MapReduce and Hadoop Distributed File System (HDFS). After this process is complete, each server has the correct software installed on it and is configured with the services it needs to perform its role in the cluster. Table 1 summarizes the process.

**Table 1:** Three-Step Provisioning Process

<b>Step 1</b>	Use Cisco UCS Manager to configure the hardware.
<b>Step 2</b>	Install the StackIQ Cluster Manager node.
<b>Step 3</b>	Power on back-end nodes and let the cluster manager install the software automatically using the information that StackIQ obtained from the Cisco UCS open XML API.

Apart from the need to select the options to install and enter cluster-specific information in the cluster manager, the process is automated, freeing the administrator to perform other tasks.

StackIQ brings enterprise-class management to Hadoop and other big data applications. It was designed from the foundation to deploy and manage large-scale cluster infrastructure. It combines StackIQ's industry-leading cluster management solution with Hadoop management software, providing everything you need to install, configure, deploy, and manage your cluster from bare metal. StackIQ Cluster Manager makes it easier than ever to build a robust, production-class, big data cluster that can reside in any enterprise data center.

## Maintaining the Cluster

After a cluster becomes operational, it will undergo changes. No matter how well planned its deployment was, it will need to be expanded, changes will need to be made, and components will fail. StackIQ Cluster Manager handles all these tasks while maintaining a consistent setup across the cluster. When the cluster is expanded, StackIQ discovers the new nodes and installs them. To make changes, the administrator adds packages to the distribution and installs the target nodes. The cluster handles failures by detecting when replacement hardware is available and automatically setting it up with the desired configuration.

Cisco UCS and StackIQ work together to keep the cluster healthy. Cisco UCS tells StackIQ Cluster Manager when new servers are available, and StackIQ Cluster Manager adds them to its database automatically.

This approach helps ensure that a consistent, reliable description of the cluster is available at all times. After the server is racked and cabled, StackIQ Cluster Manager detects it and installs the correct software automatically on first boot. The result is an automated data center that is easier and cheaper to maintain.

Here is an example of system discovery from the command line:

```
# rocks list ucs host
HOST APPLIANCE STATUS MAC
compute-1-1: compute online 00:25:B5:00:00:5F
compute-1-2: compute online 00:25:B5:00:00:8F
compute-1-3: compute online 00:25:B5:00:00:9F
compute-1-4: compute online 00:25:B5:00:00:6F
compute-1-5: compute online 00:25:B5:00:00:7F
compute-1-6: compute online 00:25:B5:00:00:4F
```

The appliance type is automatically determined by making an API call to Cisco UCS Manager. The MAC address is the hardware address of the network interface card (NIC) that StackIQ will use as the installation and management network. This is all the information required for StackIQ to begin a bare-metal installation.

## Benefits

- **Reduced time to production:** Choose the software you want, set critical parameters, and then sit back and let the parallel StackIQ Avalanche installer build your Hadoop cluster right from bare metal.
- **Ease of operation:** StackIQ provides all the tools you need to keep the cluster healthy and operating efficiently. StackIQ Cluster Manager handles your complete cluster environment from a single pane, providing users with more uptime, efficiency, and performance.
- **Extensibility:** Modular architecture lets you customize your cluster to meet your particular needs. The platform allows the management of any application in your big data environment through the Open Source Rocks framework. A wide variety of software components are readily available, or you can build your own.
- **Reduced time to scale:** When you want to scale out your cluster, StackIQ Cluster Manager makes it easy. Because the deployment and management engines are designed for scalability, expanding your cluster or creating new clusters at other locations is straightforward. You have no scripts to edit and no configuration guesswork.
- **Choice:** Choose your favorite distribution from Hortonworks, MapR, Cloudera, or Apache Hadoop, and more.

## Cisco UCS with StackIQ for Big Data

The Cisco UCS solution for StackIQ is based on the Cisco® Common Platform Architecture (CPA) for Big Data. Cisco CPA is a highly scalable architecture designed to meet a variety of scale-out application demands with transparent data and management integration capabilities built using the following components:

- Cisco UCS 6200 Series Fabric Interconnects provide high-bandwidth, low-latency connectivity for servers, with integrated, unified management provided for all connected devices by Cisco UCS Manager. Deployed in redundant pairs, Cisco fabric interconnects offer the full active-active redundancy, performance, and exceptional scalability needed to support the large number of nodes that are typical in clusters serving big data applications. Cisco UCS Manager enables rapid and consistent server configuration using service profiles, automating ongoing system maintenance activities such as firmware updates across the entire cluster as a single operation. Cisco UCS Manager also offers advanced monitoring with options to raise alarms and send notifications about the health of the entire cluster.
- Cisco Nexus 2000 Series Fabric Extenders extend the network into each rack, acting as remote line cards for fabric interconnects and providing highly scalable and extremely cost-effective connectivity for a large number of nodes.
- Cisco UCS C240 M3 Rack Servers are designed for a wide range of computing, I/O, and storage-capacity demands in a compact two-rack-unit (2RU) design. Cisco UCS C240 M3 servers are powered by dual Intel® Xeon® processor E5-2600 v2 series CPUs and support up to 768 GB of main memory (128 or 256 GB is typical for big data applications). These servers support a range of disk drive options as well as Cisco UCS virtual interface cards (VICs) optimized for high-bandwidth and low-latency cluster connectivity, with support for up to 256 virtual devices.

StackIQ Cluster Manager software runs on a separate management node, or it can share hardware with one of the cluster's data nodes. It serves as the administrator's interface to the cluster for monitoring and management tasks.

## Reference Architecture

Available reference architecture blueprints offer both high-performance and high-capacity options, which you can select according to the specific computing and storage requirements of your organization. StackIQ Cluster Manager is the same, regardless of which Cisco UCS option you select.

- **Performance and Capacity Balanced Option:** The Performance and Capacity Balanced option offers a balance of computing power and I/O bandwidth optimized to achieve an excellent price-to-performance ratio. Equipped for performance, Cisco UCS C240 M3 Rack Servers are powered by two Intel Xeon processor E5-2660 CPUs (16 cores), with 256 GB of memory and twenty-four 1-terabyte (TB) Small Form-Factor (SFF) disk drives.
- **High-Capacity Option:** The high-capacity option is optimized for low cost per terabyte (TB) and is built using Cisco UCS C240 M3 Rack Servers powered by two Intel Xeon processor E5-2640 CPUs (12 cores), with 128 GB of memory and twelve 4-TB Large Form-Factor (LFF) disk drives.



## Architectural Scalability

The single-rack configuration provides two fully redundant Cisco UCS 6248UP 48-Port Fabric Interconnects (to connect up to five racks) or two Cisco UCS 6296UP 96-Port Fabric Interconnects (to connect up to 10 racks and 160 servers), along with two Cisco Nexus® 2232PP 10GE Fabric Extenders and 16 Cisco UCS C240 M3 Rack Servers (either high-performance or high-capacity CPU configurations). Multirack configurations include two Cisco Nexus 2232PP fabric extenders and 16 Cisco UCS C240 M3 servers for every additional rack.

Table 2 summarizes the configurations.

**Table 2:** Configuration Options

Part Number	UCS-SL-CPA2-C	UCS-SL-CPA2-PC
<b>Computing and Storage</b>	16 Cisco UCS C240 M3 Rack Servers, each with: <ul style="list-style-type: none"> <li>• 2 Intel Xeon processors E5-2640 v2 at 2.5 GHz</li> <li>• 128 GB of memory</li> <li>• Cisco UCS VIC 1225</li> <li>• 12 LFF 4-TB 7200-rpm 3.5-inch SAS HDDs</li> <li>• LSI MegaRAID 9271-CV 8i card</li> </ul>	16 Cisco UCS C240 M3 Rack Servers, each with: <ul style="list-style-type: none"> <li>• 2 Intel Xeon processors E5-2660 v2 at 2.9 GHz</li> <li>• 256 GB of memory</li> <li>• Cisco UCS VIC 1225</li> <li>• 24 SFF 1-TB 7200-rpm SFF SATA HDDs</li> <li>• LSI MegaRAID card</li> </ul>
<b>Network</b>	10-Gbps unified fabric supported by: <ul style="list-style-type: none"> <li>• 2 Cisco UCS 6296UP 96-Port Fabric Interconnects</li> <li>• 2 Cisco Nexus 2232PP 10GE Fabric Extenders</li> </ul>	
Part Number	SIQENTDAT-12X5-01-0001	
<b>Software</b>	<ul style="list-style-type: none"> <li>• StackIQ Cluster Manager (supports popular Hadoop and NoSQL distributions, including Cloudera, MapR, and Hortonworks)</li> <li>• StackIQ support</li> </ul>	

## Conclusion

Big data infrastructure is taking its place in the data center, and its use is growing. Choosing a solid foundation on which to build your big data solutions is critical. Using the right tools from the beginning can help ensure success.

StackIQ provides proven technology for building and maintaining healthy cluster infrastructure. It makes big data implementation easy, dependable, and fast for enterprise-ready deployments. StackIQ's engineers have been building cluster management software for more than a decade. The combination of StackIQ Enterprise Data and Cisco UCS creates a consistently dependable deployment and management model that can be implemented rapidly and customized for either high performance or high capacity using Cisco Unified Fabric and powerful and efficient Cisco UCS rack servers. Whether you are deploying a large data center or buying single racks through the Cisco SmartPlay program, the Cisco UCS and StackIQ solution can be sized to meet any big data challenge.



## For More Information

- For more information about the Cisco SmartPlay program, please visit <http://www.cisco.com/go/smartplay>.
- For more information about Cisco UCS big data solutions, please visit <http://www.cisco.com/go/bigdata>.
- For more information about the Cisco CPA for Big Data, please visit <http://blogs.cisco.com/datacenter/cpa/>.
- For more information about StackIQ Cluster Manager, please visit <http://www.stackiq.com/products/>.

© 2014 Cisco and/or its affiliates. All rights reserved. Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. To view a list of Cisco trademarks, go to this URL: [www.cisco.com/go/trademarks](http://www.cisco.com/go/trademarks). Third-party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1110R)

StackIQ and the StackIQ Logo are trademarks of StackIQ Inc. in the United States and/or other countries.

C07-727919-01 09/14