

# Tier One U.S. Wireless Operator Automates Hadoop Cluster Management and Deployment With StackIQ



## CASE STUDY

### Challenge:

- Deploy a 53-node pilot Hadoop cluster with HP servers to compare server management solutions and test analytics applications
- Deploy a 150-node Hadoop cluster with Cisco UCS servers in a converged infrastructure and then, two months later, scale up with an additional 150 nodes. Add another 145-node Hadoop cluster with HP servers
- Run analytics applications seamlessly across the heterogeneous cluster infrastructure
- Introduce simplified, scalable, repeatable processes for Hadoop deployment, monitoring, and patches based on automation and centralized management

### Solution and Benefits

#### With StackIQ Cluster Manager:

- Deployed and managed Hadoop clusters with heterogeneous hardware without increasing IT headcount, with a design that can be easily scaled and managed
- Fully automated, easy-to-use solution that has become the standard for cluster management within the company
- Centralized control of clusters is now possible to provide customized configurations, fixes, patches, changes, and real-time high resolution device utilization monitoring

### The company discovered the benefits of a centralized, automated, easy-to-use platform to manage its heterogeneous HP and Cisco cluster environment without increasing IT headcount.

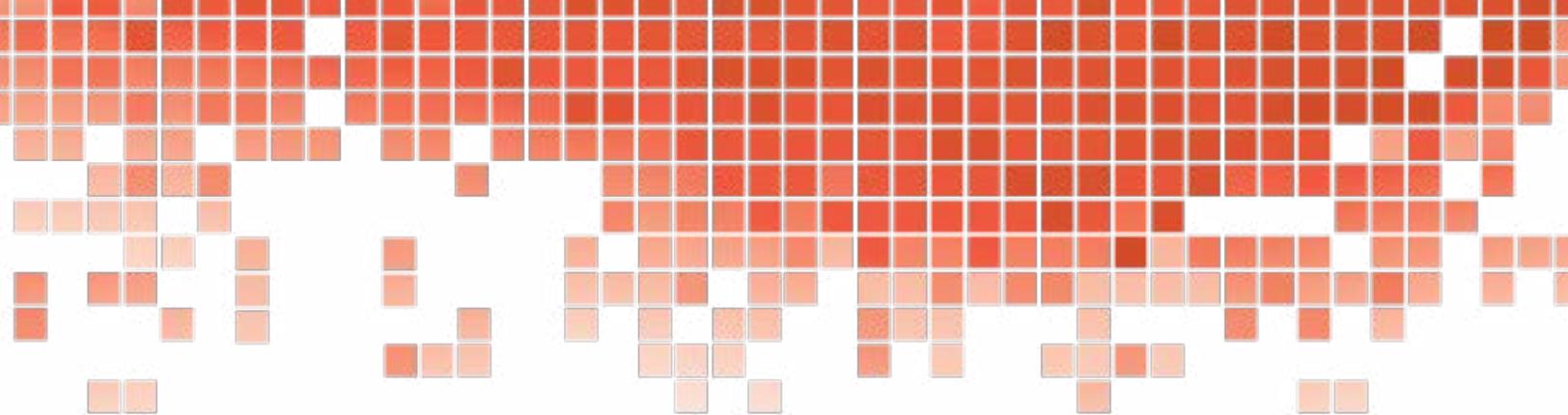
Building and managing individual servers versus clusters? It's about as different a challenge as coaching an individual athlete versus an entire team. IT System administrators at one of the largest wireless operators were tasked with rapidly and efficiently installing and managing their first Hadoop cluster with Cisco Unified Computing System™ (Cisco UCS®) servers. This would be preceded by a proof of concept cluster of 53 HP servers. Later they were going to more than double the number of nodes in the first 150-node production cluster and then add another large cluster with new HP servers. More clusters were to follow. The complexity of the project was new and daunting. It was determined that legacy tools like BladeLogic which require extensive training or professional services to administer, and tools like Puppet Enterprise and Chef, which only address the application configuration layer and require manual scripting for integration, were not acceptable options. Searching for an alternative, IT management reached out to a trusted reseller for advice. They were soon directed to StackIQ, the creators of StackIQ Cluster Manager, a highly scalable, cost-effective, software platform that automates the deployment and management of Big Data, OpenStack and Enterprise Linux environments.

### Challenge: Aggressive Timetable to Set up a Big Data Analytics Environment

Wireless operators are at the epicenter of the evolving realm of Big Data analytics. Traffic on smartphones, tablets, laptops, and other mobile devices is expected to increase 13-fold between 2012 and 2017, according to research by Cisco. That's 11.2 exabytes a month of data traffic by 2017. Wireless operators must process millions of services per second, perform forensics on dropped and poor quality calls, gather and analyze data from transmission towers and from bandwidth and service consumption records, and capture data to enhance sales and marketing.

Management wanted to start reaping the benefits of Big Data as soon as possible so IT was given a very aggressive timeline. A 53-node pilot cluster with HP servers was the first order of business. This would be followed by a 150-node cluster with Cisco UCS servers, networking gear, and storage. In a second phase, 150 more nodes would be added to that cluster. Then another cluster with 145 HP nodes would be built.

"The server administrators looked at the task before them and determined that if they used their existing tools, which rely heavily on manual script writing, the IT headcount would have to double to get the job done on time and to provide the required ongoing cluster management," says Greg Bruno, Ph.D., StackIQ Co-Founder and Vice President of Engineering. "This was their first experience with Hadoop and they didn't have any cluster experience. They knew that clusters are a different beast. Deploying individual servers is a challenge with linear complexity while deploying clusters is a challenge with quadratic complexity. One reason for this



## CASE STUDY

increased complexity is that each cluster node must be aware of all the other cluster nodes in order for all the nodes to solve a common problem together.”

In search of an enterprise tool that could handle cluster setup and management across their heterogeneous server infrastructure for the lifetime of the systems, the wireless operator was referred to StackIQ.

“Many enterprises don’t understand until late in the game that the common data center tools are not sufficient for a constantly changing and scalable cluster environment,” says Dr. Bruno. “Manual scripting and cumbersome processes with those tools introduce points of failure and error. So when we first told them about StackIQ Cluster Manager, IT management at the wireless operator determined that if the product worked as described it would save them a lot of time and money.”

### Solution: StackIQ Cluster Manager

StackIQ Cluster Manager handles the day-to-day operation of the entire software stack for clusters. It provides heterogeneous hardware support, bare metal OS provisioning, programmatic disk and network management, and integration with native Hadoop management APIs. It manages all of the software that sits between bare metal and a cluster application like Hadoop.

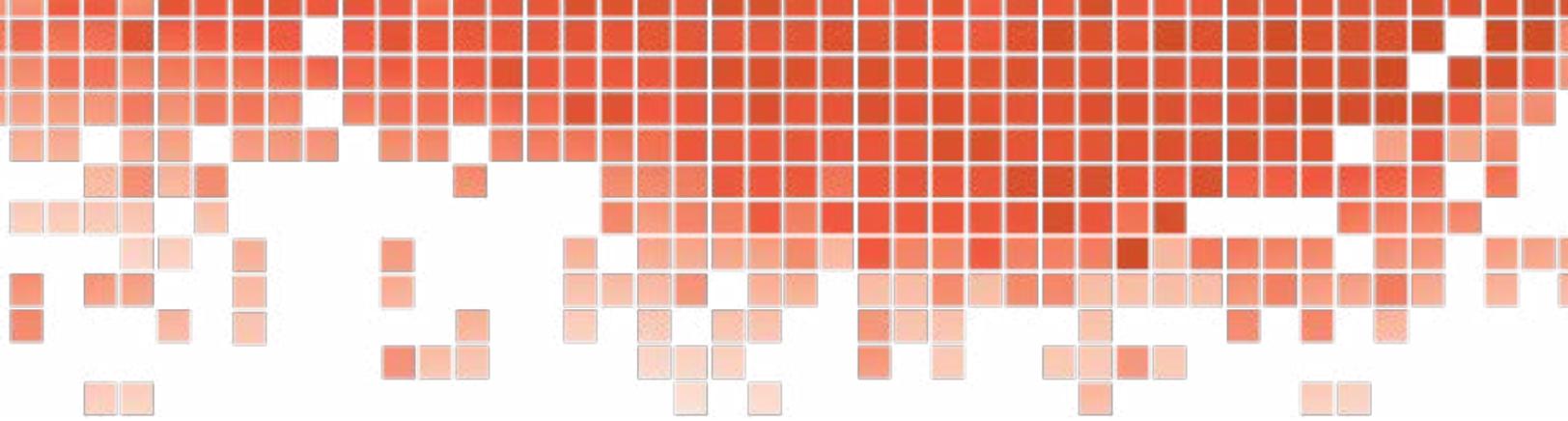
The decision was made to use StackIQ Cluster Manager for the 53-node pilot cluster that would validate the proof of concept for the new analytics systems. HP servers were then delivered to the data center. When StackIQ Cluster Manager was installed its asset inventory capability detected that nodes specified for one purpose had been interposed with nodes designed to handle other functions. StackIQ Cluster Manager helped the company’s IT staff to map out the correct cluster configuration and direct administrators at the data center exactly how to physically swap the machines on the racks into their proper location. Subsequently, because StackIQ Cluster Manager maintains cluster configuration parameters in a dynamic database used for machine configuration, the

preferred server configuration (including disk and controller configuration) could be achieved by simply powering on each server in the cluster. The combination of StackIQ’s asset inventory, automated host configuration, application definition, cluster topology, and optimized RAID controller configuration turned what would have otherwise been an unsuccessful delivery into a solution delivered ahead of the original timeline. Completed successfully, the pilot cluster led to the choice of StackIQ Cluster Manager as the product for managing Big Data infrastructure.

Soon after, the 150-node production cluster for Big Data analytics was physically ready to install (rack, stacked, cabled). A 12-hour window was given to deploy the cluster due to the start of the high-traffic holiday period. No extra network activities were to be introduced during peak production hours, so the allotted window was set for 8:00 PM to 8:00 AM.

That night, at 8:00 PM, StackIQ engineers were on a call with the wireless operator’s IT team, planning for the cluster go live. A StackIQ Cluster Manager server was running but had not been connected to the network, so it couldn’t install or manage the backend nodes in the database. Unbeknownst to the IT team, during the call one of the administrators at the data center turned on the network links to the cluster. So instead of bringing up a single data node, then a gateway node, and then a master node, to slowly and methodically introduce the cluster to the production network and monitor the activity, the first 85 nodes came online within 10 minutes and the remaining 65 were online soon thereafter. The 12-hour window was collapsed to 30 minutes and the infrastructure was perfectly configured.

“StackIQ Cluster Manager had all of the details for the back end nodes within its database and it was ready to go,” says Dr. Bruno. “It knew where all of the machines were, what software stack was being used, and took care of the installation on each node. The following day, the Hortonworks Data Platform was configured, tuned, validated, and handed over to the applications team.”



The wireless operator chose Hortonworks' distribution for Hadoop (Hortonworks Data Platform). This open, scalable, and extensible data platform is based on popular and essential Hadoop projects for storing, processing, and analyzing large volumes of structured and unstructured data. StackIQ Cluster Manager was delivered with its Hortonworks Roll. StackIQ's Rolls are pre-packed software modules that provide an end-to-end software stack while also integrating software components for site-specific requirements. In this case the chosen Roll included Hortonworks Data Platform and automated the configuration of Ambari. StackIQ also supports other major Hadoop distributions, including Cloudera and MapR, and other cluster applications like OpenStack as "out of the box" solutions.

Since the nodes in the wireless operator's cluster were behind a firewall, installing any Hadoop distribution without StackIQ Cluster Manager would have been problematic because to do a standard install, a dedicated connection to the server via the public network is required. Once the network links are established it can take days to install and configure all the software. In contrast, since StackIQ Cluster Manager Rolls are prepackaged with the intelligence to install and configure native Hadoop management tools, installation is simple, fast and reliable.

## Ongoing ROI and Peace of Mind

Since the original Hadoop cluster went live, its nodes were doubled and two more clusters have been added. An additional two more have also been built. Using StackIQ Cluster Manager, the wireless operator's IT department was also able to experiment with Cloudera, another Hadoop distribution.

"The administrators now have the right tools to respond swiftly to requests from the business users," says StackIQ's Dr. Bruno. "Using StackIQ Cluster Manager, they set up the new server racks, add a profile, and turn the new cluster on. It's nearly as easy as flipping a switch."

## About StackIQ

StackIQ is a leading provider of software that automates the deployment and management of Big Infrastructure. Based on open source Rocks cluster-management software, StackIQ simplifies the installation and management of the hardware and software that provides the infrastructure for large-scale environments with tens, hundreds, or thousands of servers supporting Big Data, cloud (including OpenStack), enterprise Linux applications, and HPC.