

Large Financial Services Company Solves Cluster Management Challenges

Case Study

Challenge

- Deploy new **363**-node Hadoop cluster as part of a multi-cluster analytics system totaling **500+** nodes
- Introduce simplified, scalable, repeatable processes for Hadoop deployment, monitoring, and fixes based on automation and centralized management

Solution and Benefits

With StackIQ Enterprise:

- All **363** nodes in the new cluster were online in two days
- Extensions to the product to incorporate the company's specific configurations were completed in one week
- Simplified, automated, consistent, and dependable management of the entire cluster, not just the Hadoop layer, was introduced to the organization
- Tight integration with native Hadoop management tools (MapR Control System in this case)
- Other benefits include a lower total cost of ownership, faster deployment, and reduced downtime due to fewer configuration errors

After working for four months to stabilize a new **363**-node Hadoop cluster, a large financial services company turned to StackIQ. Days later, the cluster was in production. Today's financial systems must be agile and scalable to support rapid change and technological innovation. Nowhere is this more evident than in analytics systems for the financial industry, where Big Data infrastructure provides valuable information and insights that can be monetized to great advantage. At a large financial services company, Big Data analytics is a top priority. When a team from IT needed to add a large Hadoop cluster to three others in production, a new approach had to be found. Fast. Enter StackIQ Cluster Manager, a highly scalable, cost-effective, automated, software suite that optimizes the management of cluster infrastructure large and small. The product brings software-defined automation to the installation, configuration and management of Linux clusters, whether private clouds, Hadoop, NoSQL, or other enterprise Linux applications.

Challenge: When the Old Ways Don't Scale, a New Hadoop Implementation and Administration Solution Must Be Found

Looking for a high-performance, state-of-the-art platform for their large, new server cluster, the financial services company chose the Cisco Unified Computing System™ (Cisco UCS®). This converged infrastructure, which integrates network, compute, and virtualization resources into one cohesive system, greatly simplifies setup and just-in-time resource provisioning. Cisco UCS integrates a low-latency unified network fabric with enterprise-class **x86**-architecture servers. Converged resources are provisioned and managed in a unified management domain.

The **363** Cisco UCS server nodes were to be added to the mostly homogenous, legacy server environment of the

CASE STUDY

other clusters. Additionally, the new cluster used the MapR distribution for Hadoop, a high availability solution with no single points of failure across the entire stack. A project requirement was that any chosen cluster management tools must integrate with, and not replace, native Hadoop management tools from MapR, such as the MapR Control System.

Work began in early 2013, with the goal of putting the new cluster into production mid-year. However, it became clear that using a combination of Red Hat Satellite Server along with home-grown tools to handle provisioning, systems management, network management, and disk management was not getting the job done. The team had been working 14-hour days and had yet to deliver the system. Using commercial software and writing scripts had worked before, when the clusters totaled about 100 nodes.

Now it was clear that the approach, leaning heavily on time-consuming, manual coding, did not scale. As configurations changed in the future, it would be an ongoing challenge to implement those changes.

Solution: Stacki Enterprise

Stacki Enterprise manages the day-to-day operation of the entire software stack running on the clusters. It provides heterogeneous hardware support, bare metal OS provisioning, programmatic disk and network management, and an interactive Hadoop management system. In other words, Stacki Enterprise manages all of the software that sits between bare metal and a cluster application like Hadoop. Cluster configuration parameters are kept in a dynamic database, which is used for machine configuration, software deployment using a unique BitTorrent-like Avalanche peer-to-peer installer, management, and monitoring. Stacki Enterprise is based on StackIQ's open source Linux cluster provisioning and management solution, Rocks, which was originally developed by researchers at the San Diego Supercomputer Center at the University of California, San Diego. Stacki Enterprise was used with MapR, which enables simplified integration of Hadoop with data architectures. Components include Hadoop software such as MapR-FS, MapReduce, Pig,

Hive, HBase, and Zookeeper.

Armed with these highly automated tools for Hadoop deployment and management, the financial services company saw their new analytics cluster go online in two days. In the next week, a consulting team from StackIQ further helped to encapsulate the company's environment, writing extensions to incorporate customer-specific configurations which were coded into Stacki Enterprise. Within two weeks, the new cluster was in production.

Ongoing Benefits

Seeing the result of the new approach to cluster configuration and management, the financial services company is now deploying Stacki Enterprise on all of their other clusters. The company has fully embraced the fully automated solution from StackIQ, with its easy-to-use graphical interface and powerful verb-based command-line, providing a centralized tool to provide customized configurations, fixes, patches, and changes.

From now on, the company can look forward to increased reliability and greatly reduced time to production, utilizing StackIQ's script-free configuration and deployment features for the complete lifecycle management of clusters, including hardware, operating system, networks, and disks. The fast peer-to-peer Avalanche installer quickly builds clusters from bare metal. Operation through a single pane of glass simplifies ongoing visibility into and support of the cluster environment at the financial services company.

With Stacki Enterprise, many other software components are available for different types of workloads (e.g., additional Hadoop distributions, private cloud APIs, and general purpose Linux applications), if needed. The team at the financial services company has already added a Solr appliance, an open source enterprise search feature, to the StackIQ framework. The product's deployment and management engines were designed for scale, so expanding a cluster or creating new ones will be easy and fast, with no scripts to edit or configuration guessing.

The close collaboration between engineers at StackIQ,

CASE STUDY

Cisco and MapR added credibility to the infrastructure and solutions from the beginning. The ultimate success of the project has resonated among other organizations grappling with their own Big Data infrastructure challenges.

About StackIQ

StackIQ helps customers build, run, and manage large distributed systems and private cloud infrastructure with a complete automation platform. To date, the company has helped over **150** organizations automate over **1 Million** Linux servers thereby removing over **560** years worth of manual installation and configuration tasks. Learn more about us at StackIQ.com



420 Stevens Avenue, Suite 100, Solana Beach, CA 92075
Phone: 858.380.2020 | info@stackiq.com | StackIQ.com
