

Orchestrator™ High Availability

True High Availability for Hadoop Cluster Environments

Solution Highlights

- Stateful and stateless failover mechanism that handles all the critical services running on the cluster
- No data loss, regardless of which service fails
- Handles all fail-over aspects
 - IP fail-over
 - Synchronizes data with newly elected backup server
 - Instantiates service/backup service automatically
- Failover mechanism is a 1-to-n failover for each key service
- Fail-over mechanism is generic and can be applied to any Unix process (e.g. MongoDB)
- Failover mechanism utilizes Zookeeper and Distributed Replicated Block Device (DRBD)

Hadoop is becoming the go-to technology for storing and processing large amounts of data using arrays of cost-effective servers. Hadoop cluster environments are complex, and require multiple services to productively function. With this complexity comes a challenge, for each service represents a single-point-of-failure. This can be a potential liability for organizations dependent upon their Hadoop deployment for critical applications in financial services, healthcare, and manufacturing, where down-time and data loss are not an option.

Challenge

When it comes to supporting a high-availability environment, open source Hadoop lacks the sophistication of traditional relational database systems, and only provides the option of running two redundant NameNodes in the same cluster in an active/passive configuration with a hot standby. This leaves other services, and as a result the entire cluster, susceptible to failure and subsequent loss of data. For organizations deploying Hadoop as a key element of an analytics solution, failures like this result in increased job run-times, project delays, and lost productivity within the business unit.

The Solution: Orchestrator High Availability and Automated Fail-over for ALL Critical Hadoop Services

Orchestrator delivers enterprise class high-availability across all Hadoop services, eliminating single points of failure that exist in open source Hadoop, and delivering the robust security and compliance capabilities that enterprises expect and need.

Zettaset Orchestrator is not a Hadoop distribution, but operates as a fully-compatible management layer that augments open source distributions including Cloudera CDH and Hortonworks HDP. Orchestrator's distribution-agnostic software simplifies Hadoop deployment, and helps business units get to the analytics phase faster and achieve greater value out of their Big Data projects.

Figure 1: All critical Hadoop services, including Kerberos, can be easily configured for high-availability with Orchestrator's easy-to-use graphical interface.

Critical Hadoop Services

Critical Hadoop services include more than just the NameNode. Failure of any of these services can result in cluster down-time, loss of data, and diminished productivity.

- NameNode (file system metadata)
- SecondaryNameNode (snapshot of the file system metadata)
- JobTracker (MapReduce, Hive, Pig job submissions)
- Oozie (scheduling Hadoop workflows)
- Kerberos (creating, renewing authentication tickets)
- Local LDAP Server (LDAP server provided by Orchestrator)
- Hive metadata DB (read / write Hive metadata)
- Web UI (Orchestrator web interface)

Service Managers

- Hadoop NameNode
- Hadoop Secondary NameNode
- Hadoop JobTracker
- Kerberos 5
- Embedded LDAP
- Hive Metastore

Zettaset Orchestrator Provides True High Availability for Open Source Hadoop Environments.

Here are some examples of how Orchestrator’s software automation boosts high availability for Cloudera CDH and Hortonworks HDP users, while simplifying overall deployment and day-to-day administration of the Hadoop cluster environment.

CLUSTER SERVICE	APACHE HADOOP, CDH, AND HDP WITH ZETTASET ORCHESTRATOR™	CLUDERA MANAGER 4.4	HORTONWORKS HDP 1.3
NameNode, with High Availability	<ul style="list-style-type: none"> • YES – Automatically installed during install process. • Eliminates Single-Point-of-Failure. • Configuration is automated. 	<ul style="list-style-type: none"> • Yes, but requires manual configuration – HA not automatically installed with distribution. Need to enable HA and configure after distro installation is complete. 	<ul style="list-style-type: none"> • NO HA – Failed NameNode must be manually isolated. • Cluster unavailable when NN service is down. • Semi-automatic install requires complex manual configuration.
Job Tracker, with High Availability	<ul style="list-style-type: none"> • YES – Automatically installed during install process. • Eliminates Single-Point-of-Failure. • Configuration is automated. 	<ul style="list-style-type: none"> • Yes, but requires manual configuration – HA not automatically installed with distribution. Need to enable HA and configure after distro installation is complete. 	<ul style="list-style-type: none"> • NO HA – JobTracker is single-point-of-failure. • Significant time to recover from failure. • Requires complex manual configuration.
Kerberos Security, with High Availability	<ul style="list-style-type: none"> • YES – Automatically installed during install process. • Eliminates Single-Point-of-Failure. • Configuration is automated. 	<ul style="list-style-type: none"> • NO – Not available by default. Can only be installed after the cluster is operational, leaving data exposed. Requires manual configuration. 	<ul style="list-style-type: none"> • NO HA – Disabled by default. • Installed after cluster operational, leaving data exposed. • Requires complex manual configuration to deploy.
All Other Services Running in the Cluster	<ul style="list-style-type: none"> • YES – Automatically installed during install process. • Eliminates Single-Point-of-Failure. • Configuration is automated. 	<ul style="list-style-type: none"> • NO – High availability available ONLY for NameNode and JobTracker when they are enabled after the installation. Other services in the cluster remain single points of failure. 	<ul style="list-style-type: none"> • NO HA • Oozie is optional. • Requires complex manual configuration to deploy.

About Zettaset

Zettaset, the leader in secure Big Data management, automates, accelerates, and simplifies Hadoop cluster deployment for the enterprise. Zettaset Orchestrator™ is the only Big Data management solution designed to address enterprise requirements for security, high availability, manageability and scalability in a distributed computing environment.



1945 Landings Drive, Mountain View, CA 94043 // USA: +1.650.314.7920 // Fax: +1.650.314.7950
 sales@zettaset.com // www.zettaset.com