

Version 4.0.1 Release Notes

What's New in Version 4.0.1

The 4.0.1 release of the MapR Distribution for Apache Hadoop contains the following new features:

MapR-DB

MapR-DB is a high-performance NoSQL database that supports both operational and analytic applications. This database is integrated into the MapR Distribution for Hadoop and used for big data applications. MapR-DB comes with a number of performance, reliability, and availability innovations. (In earlier versions of the MapR Distribution, MapR-DB was named M7.)

The **MapR Community Edition** now includes support for MapR-DB. You can develop applications using HBase APIs and deploy an unlimited number of nodes. MapR-DB applications requiring high availability (HA) features such as mirroring, snapshot, and NFS HA should migrate to **MapR Enterprise Database Edition**.

The **MapR Enterprise Edition** is for Hadoop and HBase applications. Migrating to MapR Enterprise Edition will make MapR-DB tables read-only.

Support for Hadoop 2.4.1

The MapR Distribution is built on the Hadoop 2.4.1 code base, including YARN 2.4.1.

MapR clusters support the YARN framework. In addition to Hadoop YARN functionality, MapR provides these features:

- The MapR Warden service manages node memory resources for the NodeManager, ResourceManager, and HistoryServer services. It also manages YARN container resources based on CPU, memory, and disks available on the node.
- The MapR Control System (MCS) includes views for the ResourceManager and JobHistoryServer user interfaces.
- High availability for the ResourceManager.

Heterogeneous Processing

A node in a MapR cluster can run MapReduce v1 jobs, MapReduce v2 applications, and other applications that run on YARN. Warden distributes CPU, memory, and disk resources between the TaskTracker and NodeManager.

YARN Enhancements

Version 4.0.1 includes the following YARN enhancements.

Label-Based Scheduling

The MapR label-based scheduling feature works with the following Hadoop YARN services: ResourceManager and NodeManager.

Wire-Level Security (WLS)

WLS support is extended to include YARN and MapReduce v1. WLS features work with the ResourceManager and NodeManager YARN services.

Resource Allocation

You can define the number of disks available to process YARN containers.

Central Configuration

The MapR Central Configuration feature can update configuration files for YARN applications: MapReduce v2 and other applications that can run on YARN.

HBase 0.98 Client Support for MapR-DB

MapR-DB now provides client support for Apache HBase Version 0.98.

MapR client support for HBase is at the API level. Additional HBase functionality, such as reverse scans and cell level ACLs, is not supported.

Quick Installer

The MapR quick installer adds support for the following new features:

- Installation of the [Spark](#) and [Hive](#) open source components by using the [configuration file](#).



Scala 2.10.3 or later is a pre-requisite for Spark installation. Verify that Scala is installed on nodes where you plan to install Spark.

- Installation of HiveServer2, the Derby-based Hive Metastore, and the Hive client. Multiple Hive servers are supported, but only one Metastore node can be installed.
- Configuration of the number of disks in a storage pool, known as the *stripe width*. The default stripe width is 3.
- Installation of MapReduce version 1 and MapReduce version 2 on the same node.
- Installation support with local repository: no Internet connectivity required.

MapR Interoperability Matrix

See the [Interoperability Matrix](#) page for detailed information about MapR server, JDK, client, and ecosystem compatibility.

Ecosystem Support

Note that the ecosystem components are hosted in a new ecosystem repository that is specific to Version 4.x: <http://package.mapr.com/releases/ecosystem-4.x>

To see a list of components supported in Version 4.0.1, see [Ecosystem Support Matrix](#).

For the latest ecosystem information, see [Hadoop Component Release Notes](#).

Unavailable in this Release

- Amazon EMR installation
- Rolling upgrades to Version 4.0.1

Change in MapR-FS Memory Allocation

By default, Warden allocates 35 percent of node memory to MapR-FS. However, when you specify the `-noDB` option with the `configure.sh` script, Warden changes the node memory allocation to 20 percent.

Installation with CentOS Version 6.3 and Earlier

MapR installations on Version 6.3 and earlier may fail because of an unresolved dependency on the `redhat-lsb-core` package.

This problem occurs when the CentOS repository points to vault.centos.org, rather than mirror.centos.org. If you encounter this problem, use one of the following workarounds:

- Add this repository: http://mirror.centos.org/centos/6/os/x86_64/
- Manually download and install the RPM:
 1. `wget http://mirror.centos.org/centos/6/os/x86_64/Packages/redhat-lsb-core-4.0-7.el6.centos.x86_64.rpm`
 2. `yum localinstall redhat-lsb-core-4.0-7.el6.centos.x86_64.rpm`

Known Issues

You may encounter the following known issues after upgrading to Version 4.0.1.

14907: When several jobs are submitted and the ResourceManager is using the ZKRMStateStore for failover, the cluster may experience ZooKeeper timeouts and instability. MapR recommends that customers always use the FileSystemRMStateStore to support ResourceManager HA. See [Configuring the ResourceManager State Store](#).

14947: When you configure multiple ResourceManagers in a cluster that runs on a virtual private cloud, `configure.sh` may not set the value of `yarn.resourcemanager.ha.id` correctly. This property is required for ResourceManager high availability. Workaround: Verify that the `yarn-site.xml` on each ResourceManager node contains the following:

- A unique ID (serviceID) in the `yarn.resourcemanager.ha.id` property. Each ResourceManager node should not have a serviceID equal to `rm1`.
- The ResourceManager serviceID for each ResourceManager in the cluster should be listed in the `yarn.resourcemanager.ha.rm-ids` property.

14696/15100: When ResourceManager HA is enabled and a job is submitted with impersonation turned ON by a user without impersonation privileges, the job submission eventually times out instead of returning an appropriate error. This behavior does not affect standard ecosystem services such as HiveServer because they are configured to run as the `mapr` user (with impersonation allowed). However, this problem does affect non-ecosystem applications or services that attempt to submit jobs with impersonation turned ON. MapR recommends that customers add the user in question to the impersonation list so that the job can proceed. Alternatively, wait for the timeout error to be logged (indicating that the job is not allowed on the cluster).

15096: A misleading alarm displays in the MCS when the HistoryServer addresses are not identical in the `mapred-site.xml` and the value set by the `configure.sh -HS` parameter. Workaround: Run `configure.sh` with the `-HS <hostname>` option to define the node that runs the HistoryServer.

15201: The Quick Installer installation logs print "Configuring Hive" and "Configuring Spark" messages even when these components were not configured.

Resolved Issues

The following issues are resolved in Version 4.0.1.

Installation

14477: When `configure.sh` is run with the `-R` option, the Installer no longer runs a disk space check.

MapR Control System (MCS)

8506: Multiple email addresses are now allowed when you configure alerts.

12953: A Root Directory Permissions option now exists when you create a volume, corresponding to the `maprccli -rootdirperms` argument.

14288: The Forget Node option now removes the node from the NFS Nodes view.

14430: The Job Metrics view no longer shows `Running` status for any jobs that have already completed.

Metrics Database

13158: The `hoststats` service was generating core files on several nodes at regular intervals.

14228: The `hoststats` service no longer truncates network interface metrics.

14349: A memory leak in the `hoststats` service caused a gradual increase in memory usage.

14279: A MySQL password containing an ampersand (&) could not be parsed. In Version 4.0.1, such a password is replaced with the `&` string in the `hibernate.cfg.xml` file.

MapR-DB

13166: Compression could not be set for MapR tables from the hbase shell.

14880: Gets against MapR tables sporadically returned incorrect data.

14312: Full-table scans were being used when the scan had a prefix filter. In Version 4.0.1, scans start at an appropriate key.

14558: A client application that was using AsyncHBase APIs to access MapR tables leaked memory.

YARN

13766: For Hive queries, containers were not correctly distributed across nodes in the cluster.

14023: The MapR-FS Scheme method returned an unsupported operation exception.

CLDB

12387, 14396: Null pointer exceptions (NPEs) were fixed in the CLDB.

9275: When a MapR client was behind a NAT router, the RPC layer on the CLDB rejected the client's connection attempts.

Security

14494: An option was added to disable replay detection when applications connect to the cluster using Kerberos.

12938: A new `maprlogin generateticket` command was added to support service accounts.

MapReduce

13265: Using a `hadoop job -kill` command on a streaming MapReduce job did not kill the running task processes for the job.

12722: Error messages are now logged when the `hadoop distcp` command returns an NPE.

File Client

14553: File client error messages now contain the file ID (FID).

File Server

14508: High file server memory alarms occurred on multiple nodes, with MFS memory increasing until a restart was required.

NFS

13444: You can mount a directory and its subdirectories by specifying the top-level directory. The export list does not need to have a separate entry for subdirectories within a path unless you are mounting to multiple nodes.

14447: When virtual IP failover occurs, NFS writes no longer fail with an I/O error.

14448: I/O operations no longer generate an NFS core dump.

14183: A `mkdir` command over NFS failed with a "permission denied" error for Hadoop streaming jobs submitted by users that are only part of one group.

JobTracker/TaskTracker

10927: The JobTracker became unresponsive and marked a large number of TaskTrackers as lost.

14167: TaskTracker memory was not calculated correctly when non-default map and reduce heap sizes were set.

Ecosystem

14583: AsyncHBase for MapR tables ignored the "bufferable" setting for client put requests.

12969: A Pig script with a skewed join failed with an IllegalArgumentException.

Build and Package

14267: The `amazon-s3.jar` was added to the MapR Maven repository for compatibility with the Spring Hadoop framework.

14865: An existing MapR license was disabled when a patch was installed on an EMC build.