

MAPR – THE INDUSTRY’S NEXT GENERATION DISTRIBUTION FOR APACHE HADOOP

EASY, DEPENDABLE, FAST, AND FULLY SUPPORTED FOR MISSION CRITICAL USE



DATASHEET

**MAPR IS A COMPLETE DISTRIBUTION FOR APACHE HADOOP** that includes HBase, Pig, Hive, Mahout, Cascading, Sqoop, Flume and more. MapR’s distribution is 100% API compatible with Hadoop (MapReduce, HDFS and HBase). The M5 Edition includes advanced high availability and data protection features such as JobTracker HA, Distributed NameNode HA, Snapshots and Mirroring. The M5 edition, available on a subscription basis, is fully supported for the most demanding enterprise mission critical deployments.

## MAPR’S BREAKTHROUGH INNOVATIONS

- Architecture designed specifically for high availability
- Built-in user controlled data protection and point-in-time recovery
- MapR mirroring to synchronize data across clusters
- Realtime data flows with Direct Access NFS™
- Patent pending Lockless Storage Services™ for faster job execution and greater scalability
- Instant insight from MapR Heatmap™
- MapR Volumes for easier security, placement, retention, and quota management
- Distributed NameNode HA™ for unlimited files support

## FEATURE SPOTLIGHT

**HIGH AVAILABILITY.** With JobTracker HA and Distributed NameNode HA™, MapR’s Lockless Storage Services™ architecture was designed with no single points of failure. Distributed high-availability is also built-in throughout the entire system. The M5 Edition’s JobTracker HA prevents lost jobs and painful restarts. Our Distributed NameNode HA™ eliminates one of the most frustrating failure modes of all other Hadoop distributions and provides both scale and performance advantages.

**MIRRORING FOR BUSINESS CONTINUITY.** The M5 Edition makes disaster recovery easy and built-in. Going far beyond replication, MapR’s mirroring means you can set policies around your Recovery Time objectives (RTO) and mirror your data automatically within your cluster, between two clusters or between two data centers.

**PROTECT YOUR DATA AND SET POINT-IN-TIME RECOVERY WITH SNAPSHOTS.** Replication cannot protect data from application errors or deletion. Serious applications require the ability to recover to a specific point in time. MapR’s Snapshots require very low overhead and can be done on a frequent interval. Recovery is as easy as copying a directory or files from the snapshot directory to the current directory. You can also drag and drop from any file browser supporting NFS.

**SHARE THE CLUSTER SAFELY WITH MULTIPLE JOBS.** We protect the core system by isolating it from user jobs so runaway jobs can’t bring down your entire cluster. Safely sharing the cluster also means higher utilization. Setting quotas for users and then tracking capacity usage is made easy whether you are managing one node or thousands. Automated alerts provide a built-in assistant for you to rely on as you manage the Hadoop environment.

MAPR’S COMPLETE DISTRIBUTION FOR APACHE HADOOP					
MapR Heatmap™	LDAP, NIS Integration	Quotas, Alerts, Alarms	CLI, REST, API		
Hive	Pig	Oozie	Sqoop	HBase	Whirr
Mahout	Cascading	Nagios Integration	Ganglia Integration	Flume	Zookeeper
EASY	DEPENDABLE	FAST			
Direct Deposit NFS™ Realtime Dataflows	Distributed NameNode HA™ JobTracker HA Direct Access NFS™	MapR’s High Performance MapReduce Direct Shuffle™			
MapR Volumes	Mirroring and Snapshots	Data Placement Control Local Mirroring			
MapR’s Lockless Storage Services™					

*MapR is the industry’s fastest, most dependable, easiest to use and 100% API compatible distribution for Apache Hadoop*



## MAPR MAKES HADOOP EASY

- Direct Access NFS™ mounts the cluster as a volume
- Easy graphical provisioning and planning tools
- Instrumented for alerts and alarms
- MapR Heatmap™ provides clear cluster visibility
- User & group usage tracking and quotas
- Complete distribution includes HBase, Pig, Hive, Mahout, Cascading, Sqoop, Flume, and more
- 100% API compatibility with Hadoop, HDFS and HBase
- Integrated, hardened, tested, easy to install



## MAPR MAKES HADOOP DEPENDABLE

- Advanced JobTracker HA prevents lost jobs
- Unique Distributed NameNode™ HA for unlimited scale
- Multi-node HA Direct Access NFS™
- Support for unlimited number of files
- Mirroring for disaster recovery and multi-cloud synchronization
- Snapshots for point-in-time recovery



## MAPR MAKES HADOOP JOBS RUN FASTER

- Faster execution of MapReduce and HBase jobs with half the hardware
- Increased throughput in MB/sec
- Direct block device I/O
- Optimized format on disk
- Optimized shuffle
- Automatic and transparent client-side compression reduces network overhead and footprint
- Lockless architecture designed for large clusters



### TRY IT & LET US KNOW WHAT YOU THINK

Watch demos and videos and download the software at [www.mapr.com](http://www.mapr.com).

#### Join the MapR Forum.

We would love to have you join the MapR community forum and participate. This will soon be the definitive forum for all issues having to do with getting MapReduce to work when your job depends on it.

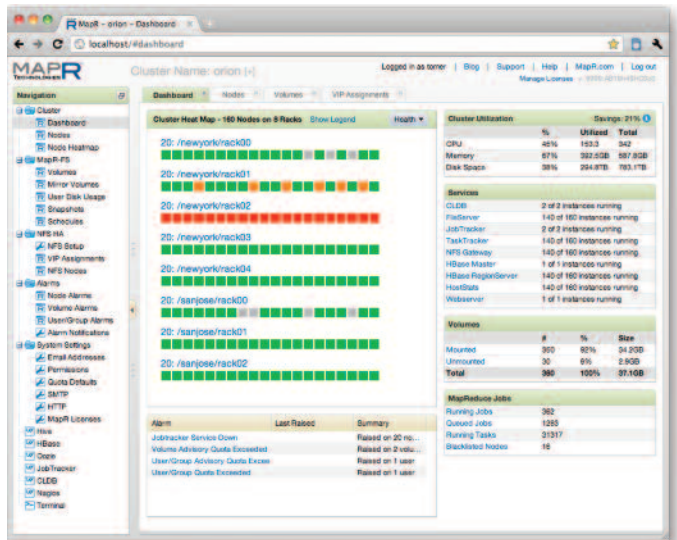


## PROVISION JOBS, USERS AND GROUPS AND PERFORM CAPACITY PLANNING.

MapR is the first distribution to implement Volumes coupled with the ability to provision around users and groups. Volumes allow you to frame your application's business requirements in Hadoop. Elements such as application Recovery Time Objectives, Recovery Point Objectives, required CPU and data resources, expected capacity growth and security requirements can all be defined at the volume level.



**REVOLUTIONARY DIRECT ACCESS NFS™ FOR REALTIME DATAFLOWS.** MapR enables you to simply mount the cluster as an NFS volume giving Hadoop a full read/write storage system to support multiple readers and writers as well as full random read and write. Our patent pending NFS access allows applications to load data directly into the cluster and gain realtime access to the results.



## INSTANT INTUITIVE INSIGHT FROM MAPR HEATMAP™.

MapR provides insight, complete access and power tools to help you understand and operate your cluster with ease. Intuitive views into the most critical issues are automatically provided along with tools to easily provision a single node or thousands. Our graphical and programmatic interfaces are designed to scale to the largest clusters, and of course, we provide CLI and REST access as well.



## MAPR M5 EDITION SUPPORT

- 24x7 phone and email support
- On-demand patches as needed
- Online incident submission and response



MapR Technologies is the creator of the industry's fastest, most dependable and easiest to use distribution for Apache Hadoop. MapR Technologies is dedicated to advancing the Hadoop platform and ecosystem to enable more businesses to harness the power of big data analytics for competitive advantage. For more information, please visit [www.mapr.com](http://www.mapr.com).