



# QUANTIFYING THE VALUE OF MAPR

## TABLE OF CONTENTS

- Introduction
- Ease of Development
- Ease of Administration
- High Availability
- Data Protection
- Performance Value

## INTRODUCTION

MapR makes development, administration and end-user file access and insight much simpler and faster. Developed specifically for high availability and data protection, MapR provides assurance with 100% uptime for your business analytics process, recovery from user and application errors and strong protection against lost data. MapR makes life easier for developers, administrators, and the analytics team allowing them all to spend their time productively. Once MapR's strong foundation is in place, businesses can start realizing significant value from their data.

## EASE OF DEVELOPMENT

MapR breaks the batch paradigm with its unique Storage Services layer that includes Direct Access NFS. With other Hadoop distributions, you have to write log files to an intermediate location, load the data to Hadoop with a batch program, then close the file and perform batch analysis. The underlying storage for other distributions is "append only" and analytics can be performed only after the file is closed. With MapR, the cluster can appear as storage and mounted through NFS. Log files can be streamed directly to the cluster and analyzed at any point because MapR storage services are completely random read/write. There is no need to perform a close. This is significant because it makes analysis easier for end users and allows applications to function seamlessly.

The ease at which you can develop and get data in and out is as simple as dragging and dropping files with full random read/write support using mounted Direct Access NFS™. Unlike other Hadoop distributions where the user must copy data out of the cluster to use the standard tools, MapR dramatically reduces costs by performing these functions on the same disk and further reduces network overhead and disk footprint with automatic client-side compression.

## EASE OF DEVELOPMENT, CONTINUED

Data is logged with random/read write directly to the disk and can be manipulated using standard Linux commands and tools from a file browser. MapR is built with applications in mind and the real time nature of the batch-less data means that you will have access to a wide variety of applications. Furthermore, MapR has 100% integrated API compatibility with Hadoop, MapReduce, HDFS and HBase, eliminating cross-distribution bottlenecks. All of this means that you can now use your standard and familiar tools, commands and interfaces with MapR's automated, non-batch, streaming compression on an intuitive interface.

VALUE: MapR saves you time and improves system productivity, giving you the ability to leverage existing tools, commands, applications and procedures without having to re-write them or extract data. These features plus automatic compression of your data translate to significant time and savings.

## EASE OF ADMINISTRATION

With MapR, installation and configuration of your system is simple. Unlike aggravating manually intensive legacy systems, MapR has been engineered for the lights out, automated datacenter. For you this means no more building workarounds and manual processes in an attempt to keep your Hadoop cluster up and running -- it just works.

MapR has built in all the functions that you need and wish you had before. For the administrator, MapR provides alerts, alarms and insights through MapR's Heatmap™ that provides a clear view of cluster health and performance. MapR Volumes simplify data security, retention, placement and quota management and the MapR Control System provides visual insight into node health, service status, resource utilization organized by cluster topology (datacenters and racks).

The momentum of big data requires tools that go beyond just visibility and MapR has the self-healing features you need into the system so that your system can automatically recover from errors. To ensure data integrity, MapR has built in end-to-end check summing.

Designed with scalability in mind, MapR intuitively manages large clusters with thousands of nodes and the MapR Heatmap shows the health of the entire cluster at a glance. To make your life easier, MapR eliminates the complexity of managing Hadoop - complete distribution includes HBase, Pig, Hive, Mahout, HBase, Cascading, Zookeeper, and more. The administrator has comprehensive tools for setting quotas for users and then track capacity usage — across one node or thousands.

VALUE: No more wasted time building manual workarounds to unsuccessfully try to keep your Hadoop cluster up, everything you need and more is built into MapR. Alerts and an intuitive graphical interface give you cluster health at a glance. Self-healing features help your system automatically recover from all software and hardware errors with full data protection. All of this is built into a package that scales with your needs to thousands of nodes.

## HIGH AVAILABILITY

MapR has been designed from the start for high availability and is the only Hadoop distribution with advanced HA features, which function at different layers, to protect your SLAs. For instance, MapR's "No NameNode" architecture provides automated stateful failover capability that protects against data loss or downtime even in the face of multiple disk or node failures without any manual intervention.

MapR also provides automated, stateful failover for all software and hardware errors. Automated re-replication of data means that your system will work through any errors without issues. MapR guarantees no more lost jobs with JobTracker HA and improves recovery time without any administrative intervention or action. MapR's rolling upgrades guarantee high availability during routine hardware and software patches as Big Data advances and demands new functionality.

**VALUE:** MapR's fully reliable system is built on a High Availability architecture, which means that you no longer have to worry about SLAs. With "no NameNode", there is no data loss or downtime even with multiple disk or node failures.

## DATA PROTECTION

MapR is built to give full data protection with Mirroring and Snapshots – features designed to efficiently maintain data integrity and business continuity across clusters and sites. MapR helps you set policies to automatically snapshot and mirror your data within your cluster and between clusters and sites eliminating the need for manual backup. This is significant because the replication that other Hadoop distributions use does not protect against user or application errors that are replicated across a cluster but with MapR you are fully protected. MapR makes future proofing a reality with data protection built in. Furthermore, you will experience zero performance lost on writing to original during snapshot, a petabyte snapshot can be performed in only seconds, giving you full range of data availability and analytic capabilities.

Successful big data use also requires tools that intelligently spot and fix errors as they occur, and MapR has the monitoring and automated self-healing features you need so that your system can automatically recover from all hardware and software errors.

**VALUE:** With MapR your data is fully protected. MapR's efficient design means that snapshots are easy to perform with no performance penalty and with no additional storage. While other Hadoop distributions lack any protection against user or application errors, MapR expands the viable use cases that are supported including those that are mission critical to your business. MapR is poised to play an integral role in your data recovery strategy.

## PERFORMANCE VALUE

MapR has built the fastest Hadoop distribution in the world so that you can interpret your data quicker and make better informed decisions earlier. MapR's patented engineering increases the speed and scale of Hadoop so you'll have confidence that your system will be able to keep up with the momentum of big data. To eliminate file contention, the main scaling issue to Hadoop, MapR's patent pending Lockless Storage Services provides a distributed read/write layer that eliminates the HDFS file limitations, Java issues and the overhead and restrictions associated with a single name node. MapR's underlying storage services architecture reduces network overhead with automatic and transparent compression and enables highly optimized shuffle operations for further performance advantages. Because of these technical advancements, MapR excels at random I/O throughput with faster performance than any other Hadoop distribution. While other distributions can scale only to a maximum of 100-200 million files, MapR keeps going to supports over a trillion files at a speed over 15,000 file creates a second. No other distribution comes close.

VALUE: MapR is the fastest and most scalable Hadoop distribution in the world. With screaming fast speeds and impressive disk-efficient architecture, you can have confidence that MapR will excel on even your most intensive analytics tasks. MapR's automatic file compression reduces storage overhead and helps you get the most out of your hardware. As the big data explosion continues, you can be assured that MapR's distribution will keep your business analytics ahead of the pack.

VALUE: Greater hardware utilization provides many benefits, the most apparent being a strong ROI that encompasses hard dollar CAPEX and OPEX savings. This benefit is not at a sacrifice of performance, uptime or ease of use but in addition to them. MapR fully utilizes your network and leverages commodity and low cost hardware to its full potential, delivering outstanding value.