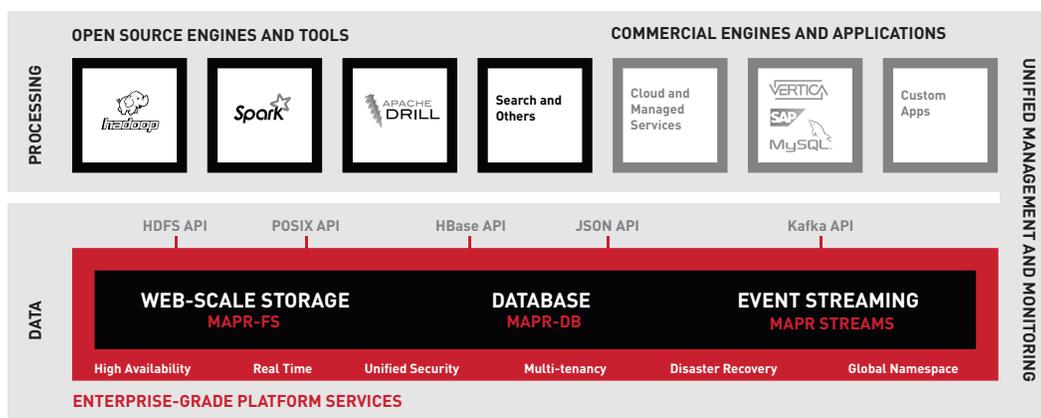## MAPR CONVERGES

- Distributed File System (MapR-FS)
- NoSQL (MapR-DB)
- Event Streams (MapR Streams)
- Apache Hadoop/Apache Spark
- ANSI SQL (Apache Drill)
- External compute engines requiring persistent storage

MAPR CONVERGED DATA PLATFORM

A platform engineered for next-generation applications

The MapR Converged Data Platform solves the crisis of complexity that results from continually deploying workload-specific data silos. Within a single platform on a single codebase, it converges the key technologies that make up a modern data architecture, including a distributed file system, a multi-model NoSQL database, a publish/subscribe event streaming engine, ANSI SQL, and a broad set of open source data management and analytics technologies.

The MapR Converged Data Platform delivers speed, scale, and reliability, driving both operational and analytical workloads in a single platform. It is architected with many performance optimizations to get the most out of your hardware. It efficiently scales horizontally ("scale out") on commodity hardware to cost-effectively expand or contract your computing power as your load changes, even to exabyte levels. It provides mission-critical high availability (HA), disaster recovery (DR), and data recovery features to maximize uptime and reduce risk of data loss.

The MapR Converged Data Platform

- Performance-optimized architecture for faster data processing and analytics
- Configurable memory usage for in-memory speeds
- Scalability to exabytes of data, trillions of files/tables/rows/documents/events
- Mission-critical business continuity features (high availability, disaster recovery, data recovery)
- Snapshots for exact point-in-time views of the cluster, without separate data copies
- MapR volumes for big data governance around security, retention, and quotas
- Direct Access NFS™ for easy data ingestion and interoperability with third-party systems
- Kerberos/LDAP integration, plus easy-to-configure username/password authentication
- Access Control Expressions (ACE) for fine-grained authorization
- Comprehensive auditing for data accesses and administrative actions, stored as JSON
- MapR Control System (MCS) for easy cluster-wide and cross-cluster administration
- MapR Installer to deploy and modify clusters quickly
- MapR Monitoring with a customizable and extensible architecture, based on popular open source tools
- Multi-master, real-time MapR-DB table replication
- Row/document-level ACID transactions in MapR-DB
- Reliable, geo-replication of event data streams

THE MAPR CONVERGED DATA PLATFORM INCLUDES THE FOLLOWING COMPONENTS

Web-Scale Storage. MapR-FS is a distributed POSIX file system with full read-write semantics, which can scale to exabytes of data and trillions of files in a single cluster.

NoSQL Database. MapR-DB is a multi-model NoSQL database that natively supports JSON document and wide column data models with high performance, consistent low latency, strong consistency, multi-master replication, granular security, and completely automatic self-tuning.

Event Streaming. MapR Streams is a publish-subscribe, event stream transport engine for reliably delivering ordered messages at high volumes and velocities.

Apache™ Hadoop®. MapR provides open source ecosystem projects to handle a variety of big data management tasks. Projects include Apache Storm™, Apache Pig, Apache Hive™, Apache Mahout™, YARN, Apache Sqoop™, Apache Flume™, and more.

Apache Spark™. MapR provides the full stack of the popular Spark tool set for fast, in-memory processing of big data.

ANSI SQL. Apache Drill™ is a SQL query engine that provides low-latency results, using familiar business intelligence (BI) tools. It also queries "schemaless" data such as JSON to enable self-service data exploration and analytics.

Third-party compute engines and custom apps. Due to interoperability features built into the system, a wide variety of third-party compute engines can run on MapR to take advantage of its speed, scale, and reliability at the data storage level.

ANALYTICAL APPLICATIONS

Business Insight

NEXT-GEN APPLICATIONS

OPERATIONAL APPLICATIONS

Business Performance

In a traditional environment, you have a separation of analytical applications from operational applications. The analytical applications dealt with historical data, and operational applications dealt with current data. Since the workload requirements between these two classes of applications differed greatly, you needed to have distinct data silos run by distinct technologies.

With MapR, you can run your analytical workloads with your operational workloads in the same cluster. This helps you avoid the complexity and risk for error with distinct security models, distinct administration frameworks, ongoing data movement, and resource allocation.

RECORD-SETTING SPEED

MapR officially set the MinuteSort record by sorting 1.5 TB of data in under a minute on Google Compute Engine. A MapR customer has since exceeded that record by sorting 1.65 TB, with one-seventh the number of servers of the highest non-MapR record.

EXTREME SCALE

MapR clusters scale linearly and incrementally, and can handle:

- Thousands of nodes and exabytes of data in a single cluster
- Trillions of files and MapR-DB tables
- Billions of streams and topics in MapR Streams
- Trillions of rows/documents per MapR-DB table
- Trillions of data events in MapR Streams

MISSION-CRITICAL RELIABILITY

High availability (HA). MapR eliminates single points of failure, avoiding data and job loss, even upon multiple node failures in the cluster. No special configuration is required, as all HA features are built in.

Disaster recovery (DR). MapR natively includes incremental, bandwidth-aware, cross-data center mirroring and multi-master table replication to enable low recovery point objectives (RPO) and low recovery time objectives (RTO) for your disaster recovery strategy.

Point-in-time recovery. Consistent snapshots create exact point-in-time views of the data without making separate distinct copies, to recover from accidental deletions, overwrites, or corruption.

INTEGRATED SECURITY

Regulatory standards. MapR complies with regulatory standards including PCI, HIPAA, NIST 800-53, GDPR, FIPS 140-2, FISMA, FedRAMP, and ISO 27001.

Access controls. Access Control Expressions (ACEs) control permissions at various levels, including file, directory, volume, column, document, and element by user, group, and/or role.

Kerberos and LDAP integration. MapR can authenticate users with Kerberos and/or LDAP.

Native authentication. MapR also offers a standards-based authentication system as a simpler alternative to Kerberos that leverages Linux Pluggable Authentication Modules (PAM) to provide the widest registry support.

Comprehensive auditing. MapR auditing logs help to analyze user behavior as well as meet regulatory compliance requirements. MapR uses the JSON format to log accesses at various levels, including the column and sub-document levels. MapR also audits at the administrative, authentication, and file levels.

Performant wire-level encryption. MapR encrypts data sent between nodes and applications to ensure data privacy, using Intel AES-NI capabilities where available.

MANAGEMENT AND MONITORING

MapR Control System. MapR includes a browser-based interface to manage, administer, expand, and monitor your cluster. You can immediately view the status of your cluster via heatmaps, and drill into specific issues to investigate any problems. Alarms proactively notify you if potential problems arise.

Unified administration. The MapR Control System (MCS) handles cluster administration across the components in the system. A command line interface (CLI) and REST API are also available.

MapR Monitoring. MapR leverages popular open source tools for a customizable and extensible monitoring framework to summarize cluster metrics and logs, critical for cluster management and expansion planning.

Automatic optimizations. MapR-DB handles region splits (i.e., sharding) automatically and eliminates compaction (defragmentation) delays. MapR-DB is self-optimizing and does not require application-level database administration code.

MULTI-TENANCY

MapR supports advanced multi-tenancy to let distinct user groups, data sets, and applications coexist in the same cluster. Tenants may be completely isolated from each other, or may allow some levels of information sharing for certain segments of their data where desired.

Volumes. MapR supports the logical grouping of files and directories on which policies (permissions, replication factors, quotas, etc.) can be set.

Security. MapR authentication and authorization controls provide another level of user and data isolation. The Whole-Volume ACE feature is an extra safeguard to guarantee that any given data set cannot inadvertently be made accessible to an unauthorized user.

INTEROPERABILITY

Open source ecosystem tool support. MapR supports all Hadoop APIs and Hadoop data processing tools to access Hadoop data. You can easily move data from MapR to other distributions, and vice versa.

Standards-based file access. MapR provides POSIX Network File System (NFS) capabilities. MapR Direct Access NFS™ provides standard file system access (via a single NFS mount point), to copy data into and out of MapR easily at high rates, or to access data using common command line tools and desktop applications. The optional add-on MapR POSIX Client provides authenticated NFS access from remote nodes, with compression and parallelization for faster throughput.

Industry standards. MapR fully supports additional industry-standard APIs, including ODBC/JDBC, LDAP, Kerberos, HBase, JSON, HDFS, NFS, and more.

Third-party tool ecosystem. The entire ecosystem of third-party tools (BI, ETL, etc.) built for use on Hadoop work on MapR. Examples of certified tools are available at the **MapR App Gallery** mapr.com/appgallery.

Visit **mapr.com/get-started-with-mapr** to get started with the MapR Converged Data Platform. Your options include downloading the software (the free MapR Converged Community Edition is available for internal testing use), downloading the MapR Sandbox virtual machine, or running MapR in the cloud.

For more information visit mapr.com