

MapR Streams adds high performance global streaming capabilities to the MapR Converged Data Platform.

MapR Streams Innovations

- **Built-in to Hadoop.** Native support for streaming data in the MapR Converged Data Platform
- **Global.** Supports reliable, globally-distributed streaming applications
- **Scalable.** Linearly scale to millions of producers and consumers, and billions of events per second with commodity hardware
- **Flexible.** Direct data access to real-time, batch, and interactive consumers with no data movement
- **Reliable.** Guaranteed message delivery
- **Secure.** Tenant-owned streams have authentication, authorization, and encryption as well as unified policy with all other data services
- **High Throughput.** Millions of messages per second, per node

MapR Streams is part of the MapR Converged Community Edition—for free, unlimited production use without some of the enterprise-grade features above. Please see the *MapR Converged Community Edition* data sheet for more information.

Publish-Subscribe Paradigm: Built for Scale

MapR Streams lets you create breakthrough applications that leverage streaming data for real-time processing with enterprise-grade security and reliability at a global scale. It connects data producers and consumers worldwide in real time, with unlimited scale. MapR Streams scales to billions of events per second, millions of topics, and millions of producer and consumer applications. Geographically dispersed MapR clusters can be joined into a global fabric, passing event messages between producer and consumer applications in any topology, including one-to-one and many-to-many.

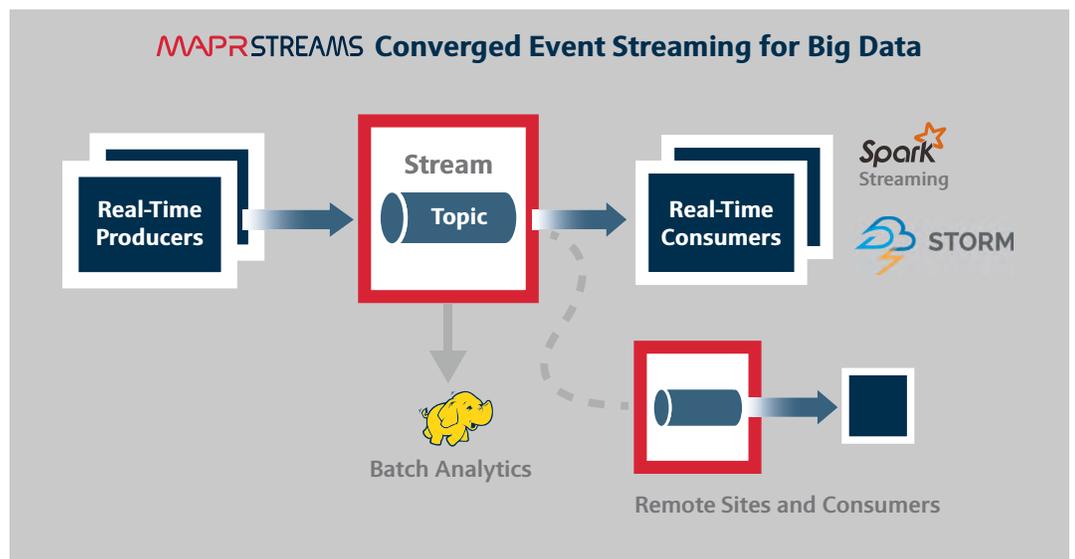
Publish-subscribe is a messaging paradigm where the data producers (referred to as publishers) do not directly send the data to data consumers. Instead, they publish the data to a system that manages “topics.” The data consumers (referred to as subscribers) subscribe to relevant topics to retrieve the data. This model allows publishers and subscribers to publish and subscribe without knowledge of each other, at different rates. This paradigm allows adding/removing data producers and consumers without making any application level changes, allowing flexibility and scale.

MapR Streams: Global Event Streaming System

Many big data sources are continuous flows of data in real time—sensor data, log files, transaction data, etc. Enterprises are struggling to deal with high volume, high velocity data using existing bulk data-oriented tools, which make the data difficult to move and delays time-to-insight.

MapR Streams is the event-oriented service in the MapR Platform and enables events to be ingested, moved, and processed as they happen. Combined with the rest of the MapR Platform, MapR Streams allows organizations to create a centralized, secure, and multi-tenant data architecture, unifying files, database tables, and message topics.

This centralized architecture provides real-time access to streaming data for batch or interactive processing on a global scale with enterprise features including secure access-control, encryption, cross datacenter replication, multi-tenancy, and utility-grade uptime.



MapR Streams is globally scalable, integrated into Hadoop, reliable, and secure.

MapR Streams leverages the technological innovations in the MapR Converged Data Platform to provide the performance, scalability, and strong data consistency to meet your mission-critical requirements.

“Our system analyzes over 65 billion new events a day, and MapR Streams is built to ingest and process these events in real time, opening the doors to a new level of product offerings for our customers.”

Michael Brown
Chief Technology Officer
comScore

Converged Data Platform for Streaming

Single cluster. The MapR Converged Data Platform supports storage and processing of files, database tables, and event streams.

Event persistence. Once events are produced, MapR Streams persists them indefinitely and provides direct data access to batch and interactive frameworks, eliminating data movement.

Reliable. The proven high availability and disaster recovery capabilities of the MapR Platform are inherited by MapR Streams. To ensure the highest levels of uptime, intra-cluster replication guarantees that there are no single points-of-failure and safeguards against multiple failures in a cluster. Events are reliably delivered from producers to consumers with no data loss.

Unlimited scale. Capacity and performance scale linearly as servers are added within a cluster, with each server handling more than 1 million messages per second.

Continuous Data

Publish-subscribe. MapR Streams enables producers and consumers to exchange events in real time via the Apache Kafka 0.9 API.

Out-of-box integration. The system is ready to work with popular stream processing frameworks like Apache Spark Streaming, Apache Storm, Apache Flink, and Apache Apex.

Global

Global stream replication. The MapR Platform provides reliable stream replication between an arbitrary topology, supporting thousands of clusters across the globe. Topologies of connected clusters include one-to-one, one-to-many, many-to-one, many-to-many, star, ring, and mesh. Topology loops are automatically handled to avoid data duplication.

Global metadata replication. Stream metadata is replicated alongside data, allowing producers and consumers to failover between sites for high availability. Data is spread across geographically-distributed locations via cross-cluster replication to ensure business continuity should an entire site-wide disaster occur.

Integrated Security

Access controls. Access Control Expressions (ACEs) control read, write, and administer permissions at the stream level.

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

Native authentication. The MapR Platform 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.

Wire-level encryption. MapR Streams encrypts all data to and from producers and consumers, as well as data replicated between globally-distributed clusters.

MapR provides the industry's only converged data platform that integrates the power of Hadoop and Spark with global event streaming, real-time database capabilities, and enterprise storage, enabling customers to harness the enormous power of their data. Organizations with the most demanding production needs, including sub-second response for fraud prevention, secure and highly available data-driven insights for better healthcare, petabyte analysis for threat detection, and integrated operational and analytic processing for improved customer experiences, run on MapR. A majority of customers achieves payback in fewer than 12 months and realizes greater than 5X ROI. MapR ensures customer success through world-class professional services and with free on-demand training that 45,000 developers, data analysts and administrators have used to close the big data skills gap. Amazon, Cisco, Google, HPE, Microsoft, SAP, and Teradata are part of the worldwide MapR partner ecosystem. Investors include Google Capital, Lightspeed Venture Partners, Mayfield Fund, NEA, Qualcomm Ventures and Redpoint Ventures.

