



Partner Brief

## Hadoop's Most Powerful Real-Time Stream Analytics Platform



Many organizations need to process Big Data in real-time – for real-time analytics, and to be able to respond to events instantaneously. With Hadoop becoming the de-facto solution for Big Data,

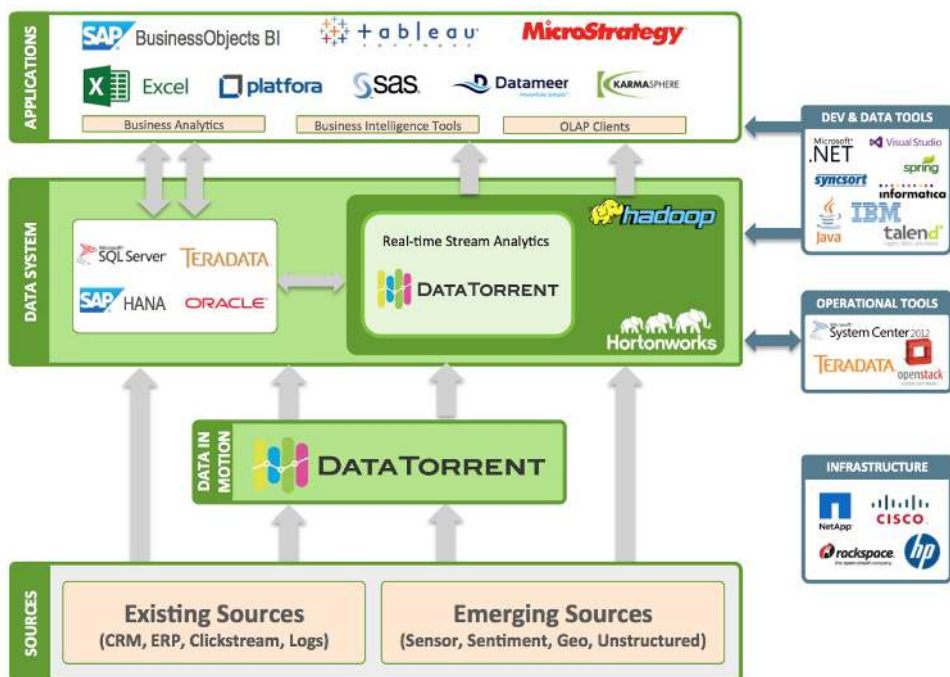
organizations want to leverage their Hadoop investment for real-time computations as well.

As a native Hadoop solution, DataTorrent supports today's most demanding, mission-critical, big-data applications. It enables organizations to develop applications that ingest massive amounts of data from various real time sources, and perform highly scalable, complex computations in real-time.

Processing **billions of events per second** with sub-second latency, and with features such as linear scalability and built-in fault tolerance, DataTorrent's unparalleled performance and enterprise-grade capabilities make it the most powerful platform for real-time analytics on a massive scale.

### The Modern Data Architecture

This diagram shows how Hadoop and DataTorrent fit with existing applications, databases and new and emerging types of data.



### Features & Benefits

- Leverage Hadoop for real-time
- Plug-in any data source
- Support any business logic
- Easily integrate with your existing data flow
- Scale linearly – processing billions of events per second
- Sub-second latency
- Built-in high-availability and state management
- All runtime operations automatically handled by the platform
- Sustain any future changes without having to change your stack or code
- Reduce cost of ownership – both of development efforts and operations.

## DataTorrent Simplifies the Development and Runtime of Big Data Real-time Applications:

- **Linear Scalability:** DataTorrent automatically scales to accommodate any data size and processing need. Linear scalability with sub-second latency - even while processing billions of events per second.
- **High Performance:** DataTorrent support massive throughput- with all computations done in-memory, with sub second latency. Per container, DataTorrent allows for massive ingestion and computation, resulting in better utilization of your infrastructure.
- **High Availability:** Applications self-heal with no data loss, no state loss, and no human intervention. Highly efficient and distributed automatic state snapshot enables check-pointing with minimal impact on latency.
- **Easy Data Integration:** Easily integrate any data flow with pre-configured input/output adapters to message buses and databases. Automatically integrate your real-time applications with your technology stack.
- **Develop Faster:** Shorten time to market with DataTorrent's extensive **open-source library** of pre-configured Operators, UI widgets, and application templates released under project **Malhar**. A rich set of tools, interactive dashboards, and webservices are provided for monitoring and debugging.

### DataTorrent - a Certified Technology Partner

DataTorrent is a Certified Technology Partner on Hortonworks Data Platform (HDP2). The Hortonworks Certified Technology Program reviews and certifies technologies for architectural best practices, validated against a comprehensive suite of integration test cases, benchmarked for scale under varied workloads and comprehensively documented.



### For additional questions, contact:

DataTorrent, Inc.  
+1 (949) 302-8814  
[info@datatorrent.com](mailto:info@datatorrent.com)  
[Datatorrent.com](http://Datatorrent.com)

## Industry Successes

A DataTorrent customer monitors the state and health of thousands of devices distributed worldwide to detect anomalies and predict possible degradations of services. The machine and sensor data was being collected at a rate of 10 million events per second. **DataTorrent was able to linearly scale to over a billion events per second, real-time.**

*"We are pleased to be joining forces Hortonworks, offering our robust real-time stream analytics platform for Big Data to Hortonworks customers," said Phu Hoang, CEO and Co-founder of DataTorrent. "Now, Hadoop customers can leverage their existing infrastructure for real-time computations on a massive scale – supporting linear scalability of 100s of millions of events per second, built-in fault tolerance and ease of development."*

### About Hortonworks

Hortonworks develops, distributes and supports the only 100-percent open source distribution of Apache Hadoop explicitly architected, built and tested for enterprise-grade deployments.



3460 West Bayshore Road  
Palo Alto, CA 94303 USA

US: 1.855.846.7866  
International: 1.408.916.4121  
[www.hortonworks.com](http://www.hortonworks.com)