

# On the Radar: DataTorrent, aiming to package streaming applications

---

Publication Date: 24 Aug 2017 | Product code: IT0014-003325

Tony Baer

---



## Summary

### Catalyst

The explosion of mobile and Internet of Things (IoT) data and, more importantly, tangible use cases have combined to put stream processing back on the front burner. Not surprisingly, a wealth of new technologies and solutions, both proprietary and open source, has emerged alongside longstanding stream processing tools to create an increasingly crowded market landscape. DataTorrent, whose founders created the real-time streaming technology that is now the Apache Apex open source project, is seeking to stand out from the crowd with a more use case/application-centric approach.

### Key messages

- DataTorrent developed and delivers commercial support for the Apache Apex streaming engine.
- The company's biggest challenge is differentiating itself in a crowded landscape where several rivals have enjoyed earlier starts.
- DataTorrent is differentiating itself by taking more of an application- and business use case-driven approach that differs from the technology focus of most of its rivals.

### Ovum view

In a crowded landscape, DataTorrent is distinguishing itself with a more business-focused approach centering on use cases demanding real-time analytics. While the underlying Apex streaming engine provides competitive and in some cases superior performance, throughput, and reliability compared to other streaming engines, the company has made the shrewd choice that streaming analytics is not purely a technology sale.

## Recommendations for enterprises

### Why put DataTorrent on your radar?

Consider DataTorrent if your analytics use cases require real-time performance with rollback and recovery capabilities, and especially if your use cases revolve around areas such as fraud detection and prevention, adtech, healthcare, telecom, IoT, and other areas where the company has cultivated expertise from early engagements.

## Highlights

### Background

The forerunner of streaming analytics – event processing – appeared prior to the emergence of open source, smart devices, inexpensive bandwidth, and powerful commodity infrastructure. That has broadened addressable markets as streaming solutions become accessible and affordable to organizations outside the original bastion of financial services.

DataTorrent was founded in 2012 by several ex-Yahoo employees who sought to develop a system that would complement big data platforms such as Hadoop; where the streaming technology would analyze data in motion, the underlying Hadoop platform would handle more complex analytics (such as building machine learning and/or predictive models) from data at rest. Although the technology was developed as a proprietary offering, it was open sourced and became a top-level Apache project in 2015. Apex's best-known reference is its use as the stream processing engine behind the GE Predix IoT cloud-based analytics platform.

From a technology perspective, Apex was designed as a streaming engine (the ability to handle a single event at a time), which makes it similar to other open source engines like Flink, Storm, or Heron (but not Spark Streaming, which was developed for microbatch). Although it operates through time-based windows, it can process events before the time interval has expired. Like Flink, it also supports stateful processing, meaning streams can be rolled back to specific events, and exactly-once sequential processing (the most stringent form of reliability for event processing that ensures no duplicate or out-of-order events are processed). DataTorrent claims that Apex delivers 10–100x lower latency than the microbatch-based Spark Streaming engine. Like Kafka Streams and Spark Streaming, Apex can autoscale, meaning that processing can more readily ramp up and down (which can be highly useful if paired with elastic computing in the cloud). But the biggest distinction for Apex is that unlike other open source streaming engines, it was designed (not adapted) to run under YARN, Hadoop's resource manager. While other streaming engines can work with YARN, Apex's YARN-friendly design makes it well suited for operating alongside Hadoop, where more complex, offline analytics for building predictive models or scoring models can be run.

An accompanying open source project, Malhar, provides "operators" or functions for Apex that are equivalent to the libraries that Spark also provides for Spark Streaming. That means that streaming and other operations supported by Malhar run on the same code base with minimal loss of performance, compared to other streaming engines that rely on third-party connectors. That also provides benefits when taking advantage of features such as Apex's underlying fault tolerance – there is no impedance mismatch. In early 2017, DataTorrent introduced a library of architectural templates for data ingestion, preparation, and data lake synchronization.

## Current position

As noted above, the streaming landscape has become very crowded; Ovum counts roughly 20 open source and proprietary tools, engines, and frameworks vying for attention. And so competing on technology alone will not distinguish Apex, which lacks the community size of the Spark computing project. That, along with competition from the major cloud platform providers (each of which offers its own streaming and data flow services), is DataTorrent's greatest challenge.

DataTorrent has raised a relatively modest \$23m in three funding rounds, with the last having come back in 2015. By comparison, Databricks, whose founders created Spark and which currently offers a Spark cloud service, has raised roughly 10x that amount. With a recent change of management, DataTorrent is forging a different path, and it is now starting to focus on business use cases rather than raw technology. It has created an AppFactory marketplace of DataTorrent and partner-developed content that starts with the typical mix of reference architectures, code stubs, and architectural templates; what will make the AppFactory unique, however, will be a growing library of streaming *applications* that is atypical of streaming engine providers. DataTorrent has just released its first such streaming application for real-time detection and prevention of omnichannel fraud; on the horizon, it

will expand this to online account protection to ward off third parties for hijacking ownership of accounts through false identities. As DataTorrent's installed base also counts adtech, IoT, healthcare, utilities, retail, telecom, and connected transportation providers, we expect that there will be additions to the DataTorrent AppFactory in those areas down the road.

## Data sheet

### Key facts

<b>Product name</b>	DataTorrent RTS platform	<b>Product classification</b>	Real-time streaming analytics
<b>Version number</b>	3.9	<b>Release date</b>	August 2017
<b>Industries covered</b>	Financial services, manufacturing, digital service providers, retail, adtech, and others	<b>Geographies covered</b>	Americas, EMEA
<b>Relevant company sizes</b>	Midsize to large	<b>Licensing options</b>	Subscription
<b>URL</b>	www.datatorrent.com	<b>Routes to market</b>	Direct
<b>Company headquarters</b>	San Jose, California, US	<b>Number of employees</b>	60

Source: Ovum

## Appendix

### On the Radar

On the Radar is a series of research notes about vendors bringing innovative ideas, products, or business models to their markets. Although On the Radar vendors may not be ready for prime time, they bear watching for their potential impact on markets and could be suitable for certain enterprise and public sector IT organizations.

### Further reading

*Fast Data 2015–16: The Rebirth of Streaming Analytics*, IT0014-003064 (October 2015)

"Fast data analytics requires even faster governance," IT0014-003158 (October 2016)

*Apache Kafka: Enterprise Messaging in a Scale-Out World*, IT0014-003276 (June 2017)

### Author

Tony Baer, Principal Analyst, Information Management

[tony.baer@ovum.com](mailto:tony.baer@ovum.com)

## Ovum Consulting

We hope that this analysis will help you make informed and imaginative business decisions. If you have further requirements, Ovum's consulting team may be able to help you. For more information about Ovum's consulting capabilities, please contact us directly at [consulting@ovum.com](mailto:consulting@ovum.com).

## Copyright notice and disclaimer

The contents of this product are protected by international copyright laws, database rights and other intellectual property rights. The owner of these rights is Informa Telecoms and Media Limited, our affiliates or other third party licensors. All product and company names and logos contained within or appearing on this product are the trademarks, service marks or trading names of their respective owners, including Informa Telecoms and Media Limited. This product may not be copied, reproduced, distributed or transmitted in any form or by any means without the prior permission of Informa Telecoms and Media Limited.

Whilst reasonable efforts have been made to ensure that the information and content of this product was correct as at the date of first publication, neither Informa Telecoms and Media Limited nor any person engaged or employed by Informa Telecoms and Media Limited accepts any liability for any errors, omissions or other inaccuracies. Readers should independently verify any facts and figures as no liability can be accepted in this regard – readers assume full responsibility and risk accordingly for their use of such information and content.

Any views and/or opinions expressed in this product by individual authors or contributors are their personal views and/or opinions and do not necessarily reflect the views and/or opinions of Informa Telecoms and Media Limited.

## **CONTACT US**

[www.ovum.com](http://www.ovum.com)

[analystsupport@ovum.com](mailto:analystsupport@ovum.com)

## **INTERNATIONAL OFFICES**

Beijing

Dubai

Hong Kong

Hyderabad

Johannesburg

London

Melbourne

New York

San Francisco

Sao Paulo

Tokyo

