



(<http://hpcwire.com/>)

(<http://www.enterprisetech.com>)



(<http://www.hpcwire.jp/>)

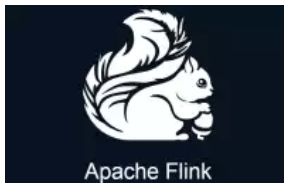
Top News from  
Leading Solution  
Providers

(<http://www.datanami.com/>)

January 12, 2015

## Apache Flink Takes Its Own Route to Distributed Data Processing

Alex Woodie



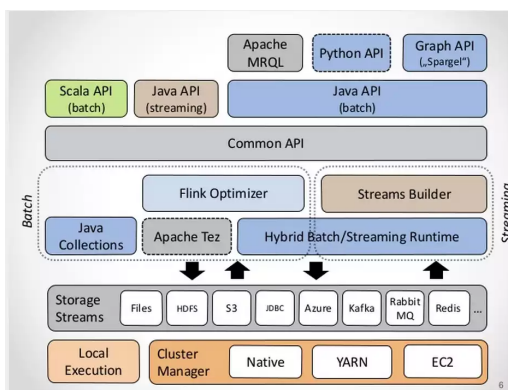
Apache Flink, a distributed in-memory data processing framework project born out of Germany, this week graduated the Apache Incubator stage and became a Top-Level Project at the open source software foundation, paving the way for an acceleration of Flink development and deployment.

Originating from the Stratosphere research project at the Technical University of Berlin in

2009, the Flink framework is designed to run batch and streaming analytic workloads in a parallel manner. In a Hadoop cluster, Flink sits atop YARN and provides yet another alternative to MapReduce as the processing engine. But Flink doesn't have to run atop Hadoop, and in fact can run on laptops, in the cloud, or even as a containerized Web app.

The creators of Apache Flink (<http://flink.apache.org/>) concentrated their efforts on building a scalable and performant framework. To that end, they developed unique type serialization, memory management, and query optimizers in Flink. These innovations continue to be the calling card for Flink and the source of the performance advantage over MapReduce-powered Hadoop touted by its founders.

According to Flink's backers, the open source framework is particularly good at iterative processing, which is something that MapReduce has always struggled with. During tests, Flink ran the PageRank algorithm on a graph of ([http://2s7qir373w3x22if92z99mgm5w.wpengine.netdna-cdn.com/wp-content/uploads/2015/01/apache-flink\\_3.png](http://2s7qir373w3x22if92z99mgm5w.wpengine.netdna-cdn.com/wp-content/uploads/2015/01/apache-flink_3.png)) Twitter data nearly three times faster than MapReduce, with twice the number of iterations.



Currently, Flink offers programming APIs in Java and Scala, as well as Spargel, a specialized API that implements a Pregel programming model. The Flink community is currently working on a Python API, as well as an API for streaming data.

Most of the creators behind Flink are at data Artisans (<http://data-artisans.com/>), a Berlin-based company founded to further the development of the software and to provide technical support for companies adopting the software. Kostas Tzoumas, co-founder and CEO of data Artisans, says becoming a Top-Level Project in the Apache Software Foundation is a great milestone for Flink, and reflects how quickly the community is growing.

"The community is currently working on some exciting new features that make Flink even more powerful and accessible to a wider audience, and several companies around the world are including Flink in their data infrastructure," Tzoumas says.

This Just In

Most Read

May 15, 2015

PCTEL Selects Cloudera as its Platform Provider (<http://www.datanami.com/this-just-in/pctel-selects-cloudera-as-its-platform-provider/>)

May 14, 2015

Accenture Launches Advanced Analytics

## Sponsored Whitepapers

**SpotlightON: Accelerating Time to Value of Hadoop Data**  
(<http://www.datanami.com/whitepaper/spotlighton-accelerating-time-to-value-of-hadoop-data/>)

**SpotlightON: OLAP on Hadoop**  
(<http://www.datanami.com/whitepaper/spotlighton-olap-on-hadoop/>)

View the Whitepaper Library  
(<http://www.datanami.com/whitepaper/>)

## Job Bank

View the Job Bank  
(<http://www.datanami.com/job-bank/>)

## Contributors



Alex Woodie  
Editor in Chief



George Leopold  
Contributing Editor



Thomas Ayres  
Contributing Editor

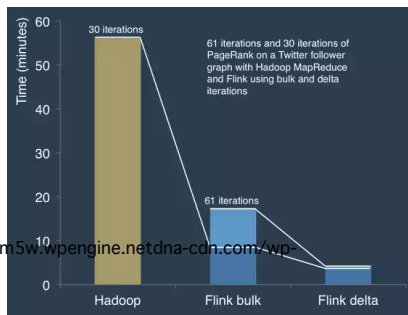


Steve Conway  
IDC



Tiffany Trader  
Contributing Editor

## Featured Events



(http://2s7gjr373w3x22jf92z99mgn5w.wpengine.netdna-cdn.com/wp-

content/uploads/2015/01/apache-flink\_2.png)

There don't appear to be many wide-scale Flink deployments at the moment, which is understandable considering how young Flink is. But there are several prominent companies are experimenting with Flink, including the European travel giant Amadeus, streaming music site Spotify, and ResearchGate, a gateway for open scientific research.

"At Amadeus, we continually seek for better improvement in our analytic platform and our experiments with Apache Flink for analytics on our travel data show a lot of potential in the system for our production use," says Denis Arnaud, head of data science development Amadeus Travel Intelligence.

"I have been experimenting with Flink, and we are very excited to hear that Flink is becoming a top-level Apache project," said Anders Arpteg, Analytics Machine Learning Manager at Spotify.

ResearchGate uses Apache Flink in its production data infrastructure, according to Ijad Madisch, co-founder and CEO of ResearchGate. "We are happy all around and excited that Flink provides us with the opportunity for even better developer productivity and testability, especially for complex data flows," Madisch says. "It's with good reason that Flink is now a top-level Apache project."

#### Related Items:

Hadoop Alternative Is Faster and Lighter, Proponents Say  
(http://www.datanami.com/2014/09/29/hadoop-alternative-faster-lighter-proponents-say/)

Hadoop Alternative Hydra Re-Spawns as Open Source  
(http://www.datanami.com/2014/03/12/hadoop\_alternative\_hydra\_re-spawns\_as\_open\_source/)

Hadoop Alternative Lands on Amazon's Cloud  
(http://www.datanami.com/2011/12/02/hadoop\_alternative\_lands\_on\_amazon\_s\_cloud/)

#### Share this:

ⓧ (http://www.datanami.com/2015/01/12/apache-flink-takes-route-distributed-data-processing/?share=twitter)

f (http://www.datanami.com/2015/01/12/apache-flink-takes-route-distributed-data-processing/?share=facebook)

g+ (http://www.datanami.com/2015/01/12/apache-flink-takes-route-distributed-data-processing/?share=google-plus-1)

in (http://www.datanami.com/2015/01/12/apache-flink-takes-route-distributed-data-processing/?share=linkedin)

♥ (http://www.datanami.com/2015/01/12/apache-flink-takes-route-distributed-data-processing/?share=pocket)

📺 (http://www.datanami.com/2015/01/12/apache-flink-takes-route-distributed-data-processing/?share=reddit)

p (http://www.datanami.com/2015/01/12/apache-flink-takes-route-distributed-data-processing/?share=pinterest)

t (http://www.datanami.com/2015/01/12/apache-flink-takes-route-distributed-data-processing/?share=tumblr)

📌 (http://www.datanami.com/2015/01/12/apache-flink-takes-route-distributed-data-processing/?share=stumbleupon)

✉ (http://www.datanami.com/2015/01/12/apache-flink-takes-route-distributed-data-processing/?share=email)

Tags: Apache Flink (http://www.datanami.com/tag/apache-flink/), Hadoop (http://www.datanami.com/tag/hadoop/), mapreduce (http://www.datanami.com/tag/mapreduce/)

Only registered users may comment. Register using the form below.

Check off newsletters you would like to receive \*

- ☐ HPCwire
- ☐ EnterpriseTech
- ☒ Datanami

**Provider Forum on Financing & Investing in Data Center & Cloud Services Infrastructure**  
(http://www.datanami.com/event/financing-investing-in-data-centers-forum/)

**Big Data Expo New York**  
(http://www.datanami.com/event/data-expo-new-york/)

**IMN's DATA CENTERS FORUM (EAST)**  
MAY 28-29, 2015 | NYC  
May 28 - May 29  
New York NY United States

**BIG DATA & ANALYTICS FOR PHARMA Summit**  
(http://www.datanami.com/event/big-data-analytics-for-pharma-summit-2/)

**Philadelphia 2015**  
June 10 @ 8:00 am - June 11 @ 5:00 pm  
Philadelphia PA United States

**CIO Energy Summit**  
(http://www.datanami.com/event/energy-summit/)

August 27 - August 28  
Houston TX United States

View More...  
(http://www.datanami.com/events/)

#### Sponsored Multimedia

(http://www.datanami.com/2015/01/12/how-to-realize-analytics-value-faster/)

**How to Realize Analytics Value Faster**  
(http://www.datanami.com/2015/01/12/how-to-realize-analytics-value-faster/)

**Webcast Sponsored By: Hortonworks and TERADATA**  
(http://www.datanami.com/2015/01/12/top-5-ways-reduce-complexity-hadoop-data-warehouse/)

**Webcast: Top 5 Ways to Reduce Complexity Between Hadoop and the Data Warehouse**  
(http://www.datanami.com/2015/01/12/top-5-ways-reduce-complexity-hadoop-data-warehouse/)

- ☐ Technology Conferences & Events
- ☐ Advanced Computing Job Bank
- ☐ Technology Product Showcase

**Email \***

**Name \***

First

Last