Glassbeam integrates with Apache Spark for in-memory machine data processing

Analyst: Jason Stamper

25 Nov, 2014

Glassbeam says that with the latest version of its SCALAR data processing engine, it is prepared for the IoT, which will require real-time data analytics able to handle up to billions of sensor readings. It already had a fast analytics platform, but integration with Apache Spark enables real-time analytics, as well as predictive analytics and machine learning.

The 451 Take

We're seeing growing demand for real-time analytics as organizations seek to deliver richer insights to decision-makers and their partners and customers more quickly. While the IoT may still be in its infancy, that too will require rapid analytics and machine-learning capabilities. Since it is already tracking 1.2 billion sensor readings per day, Glassbeam has some expertise in this field, and we see its integration with the Apache Spark data processing engine as another step in the right direction.

Context

Machine-data-analytics firm Glassbeam was founded in 2009 by Puneet Pandit. Pandit was founder and CEO of Orchesys, where he incubated Glassbeam. Prior to Orchesys, Puneet was senior director at Network Appliance, leading the database and business applications solutions group. Before NetApp, Puneet worked at Ernst & Young strategic advisory services and Tata Unisys as a management consultant. All told, he has more than 20 years of global IT experience.

Today, the company has about 40 staff and has raised $8.1m from angel investors in two rounds. The latest was a $2m round in September. Kumar Malavalli, cofounder, chairman and key investor, assumed the chief strategy officer role effective Q3 2014; Sashi Reddi, founder and CEO of Softlab (sold to CSC in 2012) was nominated to the board of directors around the same time.

Technology

The company started with Glassbeam Analytics for standard and custom analytics on machine-generated data, and then introduced Glassbeam Explorer for search and exploratory analysis in 2013. It later added a back-end platform to its portfolio, dubbed SCALAR, which is designed to bring performance and scalability to the process of analyzing data from the IoT and is delivered via the cloud in a SaaS model. The company recently announced a new version of the SCALAR platform, which it says is tightly integrated with Apache Spark, an open source data processing engine that runs on Hadoop, Mesos, either stand-alone or in the cloud. It is said to be able to run programs 100 times faster than Hadoop MapReduce in memory, or 10 times faster on disk.

Glassbeam has already been helping product manufacturers and other companies monitor and analyze their machine data – it says it tracks 1.2 billion sensor readings per day for its customers. With the data stored in the cloud, Glassbeam needed to be mindful of security, and it has been, with granular security controls based on role and no root logins.

In addition to integrating with Apache Spark to deliver order-of-magnitude speed increases, the firm has added new machine-learning and predictive-analytics capabilities to its core platform. For example, such capabilities will allow product manufacturers to be more proactive in preventing part failures, or to predict which machines or parts are susceptible to higher failure rates in the future. As a result, manufacturers will be able to prescribe solutions to problems before they happen in the field, the firm says. Apache Spark also gives Glassbeam MLlib library integration with its machine-learning algorithms that can perform predictive analytics on large sets of machine data in the cloud. Furthermore, implementing Apache Spark SQL directly on Cassandra data – which SCALAR was already based on – will allow real-time analytics on data as it is streaming in and getting parsed and transformed through the Glassbeam platform.

To begin using the platform, Glassbeam develops the structured product labeling (SPL) code for a particular device and data format to ingest the data and do data parsing and transformation. Today, this is done by Glassbeam staff with an in-house tool called Glassbeam Studio. The company has plans to expose the tool to customers and partners in early 2015 so they can write their own SPL

and get up and running faster. Glassbeam also has a Rules and Alerts engine in place with a front-end application that lets users define the rules. The plan is to expand that functionality to allow end users to define more complex rules on incoming streaming data to make it more real-time, thus taking action on both data in motion and data at rest.

Glassbeam supports the ingestion of multiple data types and formats, including text, XML, JSON and .CSV. The REST API allows access to the data via third-party tools for further visualization, such as Tableau. The application comes out of the box with a number of applications: a log vault to store all log data; Explorer for full text and faceted search across logs; Workbench to play around with parsed structured data; and Rules and Alerts to define rules on machine data. Today, the SaaS application runs on Amazon Web Services, but it has also been hosted by Glassbeam partners Dimension Data and Cisco. The company says it may offer it on IBM SoftLayer, too, in order to give customers more choice.

The company has also announced the formation of a machine learning 'center of excellence' with Tiger Analytics, a 'big data' and data science services provider. Although still a work in progress, the idea is to build a community of people interested in machine data analysis that can share their experience and ideas. We understand Glassbeam is also in talks with research institutions about joining the 'center.'

Customers

Glassbeam counts IBM, HDS, Aruba Networks, Dimension Data and Meru Networks among its customers. Violin Memory, a provider of all-flash storage arrays, recently selected the Glassbeam platform to deliver key performance and health-check analytics for its upcoming client portal. Violin Memory's storage products generate complex machine data as call-home, and it uses Glassbeam to analyze this and provide proactive and predictive information to its customers.

Competition

Due to its focus on machine data analytics Glassbeam will be compared with Splunk, but it says it can ingest more data types than Splunk – in addition to syslogs, it says it can ingest multi-structured logs, configuration data, statistical data and static data. Others in the log management space that it will likely be compared with include Sumo Logic, Loggly, Logentries, X15 Software and TIBCO (through its acquisition of LogLogic).

Other players still are focusing on machine data around the IoT, focusing slightly less on log data and more on data created by sensors or data streaming in from smart devices. In this camp, we'd

put the likes of ParStream, IBM InfoSphere Streams, DataTorrent, SQLstream and Software AG Apama. Potential customers might also be considering AWS's Kinesis stream-processing service, which is likely to become an increasingly popular choice.

SWOT Analysis

Strengths

Glassbeam has the ability to ingest multiple data types and formats and perform relatively fast analytics in the cloud. It comes with plenty of applications out of the box and has granular security controls to keep data safe.

Weaknesses

As a five-year-old company, it's perhaps understandable that it doesn't have a huge number of reference customers, but we'd like to see a few more added to that list in coming quarters. Today, Glassbeam helps with the initial setup, but users will get more control as the firm releases more tools.

Opportunities

With interest in the IoT growing, proactive and predictive analytics are going to become increasingly more valuable. Glassbeam was fast before, but hitching its cart to Apache Spark should help it to offer real-time analytics.

Threats

Because it runs on its own cloud and integrates with other AWS offerings such as the Redshift warehouse, S3 storage and Amazon Elastic MapReduce, Kinesis is a potential threat, as are the rival log management and streaming companies (of which there are many).