At its recent .conf2016 user conference in Orlando, Splunk sprinkled machine-learning capabilities across much of its product portfolio. It rolled out machine learning in two ways: as an integrated core capability for some Splunk applications, and as a Machine Learning Toolkit app for the core Splunk Enterprise and Splunk Cloud platforms for more expert users.

The 415 Take

Splunk supports prepackaged content and visualizations for a wide variety of use cases, including IT operations, security and business analytics. This is making Splunk-based analytics available to an increasing variety of IT and business users. Now, with the launch of a broad array of machine learning, Splunk provides a comprehensive answer to one of the biggest challenges facing modern organizations: how to harness diverse and increasingly profuse amounts of data to gain valuable business insights. Meanwhile, the company reduced users’ reliance on the Search Processing Language (SPL) to write queries in Splunk thanks to a new GUI, and quietly harmonized searches across Splunk and Hadoop with a Hadoop Data Roll product, which means the former Hadoop-based Hunk analytics is no longer a separate product in its own right. Splunk needs to be mindful,
however, that its GUI enhancements don't scare its loyal following of SPL 'ninjas' – to some extent, it can't afford to negate the need for SPL altogether because to do so would be to disenfranchise the very people that have been so influential in its success.

Context

Founded in 2003 in San Francisco, Splunk went public on Nasdaq in 2012, and in doing so was one of the first companies to be able to capitalize on the hype around big data, and to a lesser extent the Internet of Things. That's because the vendor has an analytics platform that is geared toward the analysis of machine data – the data from logs, systems in datacenters, connected devices and the growing number of device sensors. The logic in doing this is that there is a great deal of value hidden in that huge pool of data, whether to improve the efficiency of the IT operation itself, better understand customer behavior, or even detect fraud and other security- and compliance-related events.

The company has gradually built out its applications on top of the core engine so that it now has specific offerings in three main areas: security, IT operations management and then broader business analytics. It has more than 100 applications all told, and says it has 12,000+ customers and more than 2,700 staff. For its fiscal second quarter ended in July, total revenue increased 43% year-on-year to $212.8m, and it had a GAAP operating loss of $83.6m, with non-GAAP operating income of $8.2m. Operating cash flow was $18.3m. CEO Doug Merritt said the company added 500 new customers in the quarter. Merritt took over the CEO role from former CEO Godfrey Sullivan – who had previously run Hyperion – in November 2015. Merritt had formerly been SVP of field operations at Splunk.

Technology

At its .conf2016 user conference in Orlando, Splunk announced a number of enhancements to its applications and core engine, both in the cloud and on-premises, that center on machine learning. These included machine-learning updates to the core Splunk Enterprise and Splunk Cloud products, as well as
Splunk brings machine learning into the machine-data analytics fold

Splunk IT Service Intelligence (ITSI), Splunk Enterprise Security (ES) and Splunk User Behavior Analytics (UBA) applications. This report focuses on the updates to the core Splunk Enterprise and Splunk Cloud platforms; enhancements to ITSI, ES and UBA will be the focus of separate reports.

Splunk got its hands on some of its machine-learning capabilities through the acquisitions of Metafor Software in IT service management (June 2015) and Caspida in behavioral analytics and threat management (July 2015). By taking some of the work they had done and adding its own skills, it has been able to build machine learning into the core of its Splunk Enterprise and Splunk Cloud products, too.

By way of background, Splunk Enterprise and Splunk Cloud essentially ingest log data, machine and other types of data rapidly into the Splunk platform, where it is indexed, ready for users to perform searches on it and discover new patterns, trends or anomalies. Data is automatically time-stamped, but search results are displayed based on how a query is written, and the Splunk indexers' inherent 'understanding' of the nature of the data itself. Users interact with Splunk by writing search queries in a language called Search Processing Language (SPL), which is unique to Splunk. Its syntax is based on the Unix pipeline and SQL, and it is said to be 'SQL like.'

With the recent general availability of Splunk Cloud and Splunk Enterprise 6.5, the company has added machine-learning-related capabilities into SPL so that users can write queries that also invoke machine-learning algorithms to run across their data. You'll notice that Splunk Enterprise and Splunk Cloud releases are currently in lock step – the company says functionality is almost identical. There are one or two features of Splunk Enterprise that wouldn't make sense to Splunk Cloud users, such as cluster balancing, which is handled automatically in the cloud anyway. It's worth mentioning as an aside that Splunk Cloud only runs in the AWS cloud – AWS is a very close Splunk partner, and Splunk says it has no plans to offer Splunk Cloud on any other clouds, such as Microsoft Azure, Google Cloud Platform or IBM SoftLayer.

With Splunk's machine-learning updates to ITSI, ES and UBA, much of the underlying complexity of the machine learning is hidden from users, and instead, machine-learning suggestions are made to users so they can choose from a number of built-in machine-learning algorithms. These algorithms are unique to
Splunk and not open source.

In the case of the Machine Learning Toolkit for Splunk Enterprise and Splunk Cloud, the company has taken a different approach. Aimed more at expert users who have a far wider range of desired analytic outcomes, the Toolkit offers a choice of 27 machine-learning algorithms that users must understand and choose for themselves. The company says a data scientist isn't necessary, but someone with at least some understanding of statistics is probably going to be needed – the company talks about this role as a 'citizen data scientist' – something we explored in detail (/report-short?entityId=89277) earlier this year.

Splunk assembled these algorithms from a number of open source Python machine-learning libraries. Those libraries offer about 300 algorithms altogether, and Splunk will continue to add support for more of those. Whether or not Splunk supports them in Enterprise or Cloud, they remain open source. Splunk has grouped them in a new library of its own called the Python Data Science Library.

Other machine-learning algorithm libraries, such as Apache Spark MLlib and algorithms associated with the open source statistical language R, are not supported. Splunk told us it is theoretically possible to move data from Splunk to Hadoop, and once there run Spark/MLlib algorithms over it before returning the results to Splunk, but it doesn't know of any customers actually doing this. However, it hasn't ruled out the possibility of adding built-in support for algorithms other than Python to Enterprise and Cloud at some point.

Aside from the addition of the Machine Learning Toolkit to version 6.5 of Enterprise and Cloud, there were a couple of other notable developments. The company introduced the ability for users to interact with Splunk without necessarily being able to write queries in SPL. It does this not by offering support for a different query language (SPL is said to be SQL-like anyway), but instead by offering more in the way of drop-down menus, checkboxes and other graphical elements that can be used to start building queries without SPL. It describes it as an intuitive interface with table data views designed for both specialist and occasional users (for specialists, it could be used as a shortcut, or as a way of checking their SPL code).
The company notes that in many organizations, a lack of SPL skills is holding back the adoption of Splunk. Already, it's typical for a small team of SPL 'ninjas' to use their SPL skills to build dashboards and reports that are digested by a far wider group of business users – whether for IT operations management, security or broader data analytics. With the improvement to its GUI-based capabilities, some of those users can become less reliant on SPL ninjas when they need new reports or changes to their existing dashboards. Second, the company announced it would help users reduce their on-premises TCO through tighter integration with Hadoop. The way it describes this is that organizations can now roll historical data to Hadoop and use hybrid search to analyze all of their data in Splunk. It calls this Hadoop Data Roll.

Since companies license Splunk based on the ingestion of data into the Splunk platform, there is no additional charge if data is then moved to Hadoop for further analysis. Previously, if companies wanted to run Splunk queries on data in Hadoop, they had to license another application, Hunk. Effectively, that application is now free for Splunk Enterprise users who are moving data from Splunk to Hadoop – in fact, as we recently noted (/report-short?entityId=89122), Hunk is no longer even a separate product in its own right.

If the data starts in Hadoop and companies want to analyze it using Splunk, then they need to move it to the Splunk platform for indexing first – and because it's charged at the point of ingestion, there is a cost to do that. In practice, Splunk said all of its Hunk users were already Splunk Enterprise customers, so what it's really done has harmonized Splunk and the former Hunk into one search technology that can be run across data in Splunk or Hadoop, if customers use the Hadoop Data Roll to get the data in Hadoop first.

Customers

The company says all of the users of its ITSI, ES and UBA applications that upgrade to the latest versions will automatically get machine learning built in right out of the box. As discussed, there's more complexity involved in adopting machine-learning algorithms with the Splunk Enterprise and Splunk Cloud products, but the company says there are a handful of enterprises doing that already, including companies in telecommunications, fast food and web search/aggregation.
Competition

Splunk has different competitors depending on which use case companies are trying to tackle: IT operations management, security management/threat detection or broader business analytics. In log management and machine-data analytics, where the technology is often used for IT operations management, competitors include Sumo Logic, Rocana, Loggly, Logentries, LogRhythm, Elastic, Lucidworks, Logzilla, Graylog and X15 Software. Logtrust has just relocated its headquarters to the US from Spain in order to take the fight to Splunk and others.

We think Splunk has probably stolen a march on the competition when it comes to the built-in machine-learning capabilities of 6.5. Competition here will come though from Elastic, which on September 15 announced it had acquired Prelert, a provider of behavioral analytics technology with machine-learning capabilities. In the IT operations management space, we think Datadog’s February pickup of Mortar Data, which tied analytics to monitoring, has some overlap, too. AppDynamics and Moogsoft are also duking it out in this space.

Rivals for Splunk’s Enterprise Security products include RSA Security Analytics from EMC, HPE ArcSight, IBM QRadar, Tripwire and Tier-3/Huntsman Security. In the open source space, some companies might try to piece together their own open source stack using Elasticsearch, Logstash (for log ingest and processing) and Kibana (for visualization). Trying to do so is unlikely to be as inexpensive as it first looks on paper, though, and such a stack would lack Splunk’s machine learning without additional complex integrations into something such as Apache Spark with its MLlib algorithm libraries. In fact, such an open source stack would lack a vast swathe of features that are already built into Splunk, let alone Splunk’s 100-plus applications.

SWOT Analysis

<table>
<thead>
<tr>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
</table>

Although they were already comprehensive log/machine-data-analytics platforms, the addition of machine-learning capabilities to both Splunk Enterprise and Cloud make them even more powerful. Enhancements to the GUI should start to make it possible for less technical users to interact with Splunk.

Today, the only machine-learning algorithms that are supported come via various Python libraries, yet some companies prefer to use Spark with its MLlib or the R language's algorithms.

**Opportunities**

Splunk has only just started down the machine-learning road, but as it becomes a more popular discipline in all sorts of areas from security to customer behavior analytics, it will have an even greater addressable market.

**Threats**

Splunk needs to ensure that its GUI enhancements don’t scare away its loyal following of SPL 'ninjas’ – to some extent, it can’t afford to ever negate the need for SPL altogether because to do so would be to disenfranchise the very people that have been so influential in its success.

---

**Jason Stamper (/analyst-team/analyst/Jason+Stamper)**

Analyst, Data Management and Analytics

---

**M&A ACTIVITY BY SECTOR**

Information management / Data management / Event processing (7)

https://makb.the451group.com/results?basic_selected_sectors=781

**COMPANY MENTIONS (PRIMARY)**

Splunk (/search?company=Splunk)

**COMPANY MENTIONS (OTHER)**

AppDynamics, ArcSight, Amazon Web Services, Caspida, Datadog, Discover Financial Services, Elastic, EMC, Google, Graylog, Hewlett Packard Enterprise, Huntsman Security, Hyperion Solutions, IBM, Logentries, Loggly, LogRhythm, Logstash, Logtrust,
M&A ACTIVITY BY ACQUIER

Amazon Web Services Inc. [aka AWS] [Amazon.com Inc.] (4)
(https://makb.the451group.com/results?basic_acquirers=Amazon+Web Services Inc. [aka AWS] [Amazon.com Inc.])

AppDynamics, Inc. (2)
(https://makb.the451group.com/results?basic_acquirers=AppDynamics,+Inc.)

ArcSight, Inc. (1)
(https://makb.the451group.com/results?basic_acquirers=ArcSight,+Inc.)

EMC Corporation (77)
(https://makb.the451group.com/results?basic_acquirers=EMC+Corporation)

Google Inc. (208)
(https://makb.the451group.com/results?basic_acquirers=Google+Inc.)

Hyperion Solutions Corporation (7)
(https://makb.the451group.com/results?basic_acquirers=Hyperion+Solutions Corporation)

IBM Corporation (166)
(https://makb.the451group.com/results?basic_acquirers=IBM+Corporation)

Microsoft Corporation (157)
(https://makb.the451group.com/results?basic_acquirers=Microsoft+Corporation)

Moog Inc. (9)
(https://makb.the451group.com/results?basic_acquirers=Moog+Inc.)

NASDAQ Inc. [NASDAQ: NDAQ] (16)
(https://makb.the451group.com/results?basic_acquirers=NASDAQ+Inc. [NASDAQ: NDAQ])

RSA Security, Inc. (8)
(https://makb.the451group.com/results?basic_acquirers=RSA+Security, Inc.)

SoftLayer Technologies, Inc. (1)
(https://makb.the451group.com/results?basic_acquirers=SoftLayer+Technologies, Inc.)


CHANNELS

Data Platforms & Analytics (/dashboard?view=channel&channel=6)

SECTORS

All / Information management / Data management / Event processing (/search?sector=781)
<table>
<thead>
<tr>
<th>Tripwire Inc. (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(<a href="https://makb.the451group.com/results?basic_acquirers=Tripwire+Inc">https://makb.the451group.com/results?basic_acquirers=Tripwire+Inc</a>.)</td>
</tr>
</tbody>
</table>

Figures shown indicate number of transactions