Why you need

IMPROVED OPERATIONAL INTELLIGENCE

for

BIG DATA

splunk>

Industry Perspective

EXECUTIVE SUMMARY

Government agencies are creating more data than ever before, yet they often fail to capitalize on all of the information they're collecting. A recent Forrester study found that organizations are only analyzing 12 percent of their data, leaving the rest to sit idle, never to be capitalized for better decision-making.

Today, a big data solution offers hope for governments to unlock potential from all the data they collect, store and manage. Through better data management, governments can re-imagine their business processes. At a time when public sector resources and budgets are shrinking and citizens are demanding improved services, big data promises much needed relief for government agencies.

Now is the time for agencies to capitalize on their data. For years, government has been developing data lakes, or locations where they aggregate and store massive amounts of data they collect, all resting on the promise that someday the information will be valuable.

That day is today. The technology now exists to quickly process and analyze data. Organizations can finally derive value from their data lakes as well as from their machine data, or data that is created without the intervention of humans, from transactions, APIs, call centers, sensors, and more. Machine data is growing at an exponential rate, and requires organizations to take a new approach to manage all the information and extract value.

In this industry perspective, we will provide an overview of how big data is reshaping government, the challenges agencies face in extracting value from this information, and ways to work around common big data roadblocks.

We will also highlight how Splunk, an operational intelligence company for government, is helping government organizations leverage their data in new ways. Specifically, this report will:

- Provide an overview of the current landscape for big data
- Explain how Splunk can help agencies unlock new value from their information
- Describe what machine data is and the role it plays for big data programs
- Highlight an interview with William Von Alt, Sales Engineering Manager at Splunk
- Show how big data can mitigate the risks of insider threats

Public sector managers now realize that their data has high potential to transform their business, and they are ready to turn their data into insights.

"There is a lot of value that could be extracted from this data, and agencies are trying to figure out where the value is and how to gain knowledge from the information," said William Von Alt, Sales Engineering Manager at Splunk.

For government to reach the goals of their complex missions, they must learn to extract knowledge from their data to find new trends, patterns and identify business value from information. But in order to get there, agencies must first understand what big data is, and how to overcome some common challenges.

Understanding the Big Data Landscape

Many organizations still struggle to define big data, especially in the context of what a big data solution can do to transform their agency. To make the task of characterizing data even more challenging, every agency, city and municipality defines big data differently. Due to various missions and organizational needs, how the Department of Defense (DoD) defines big data is different than Department of Commerce, and how the City of Hartford is going to describe big data is different than how Baltimore County will.

But there are some commonalities across governments, like the four Vs of big data:

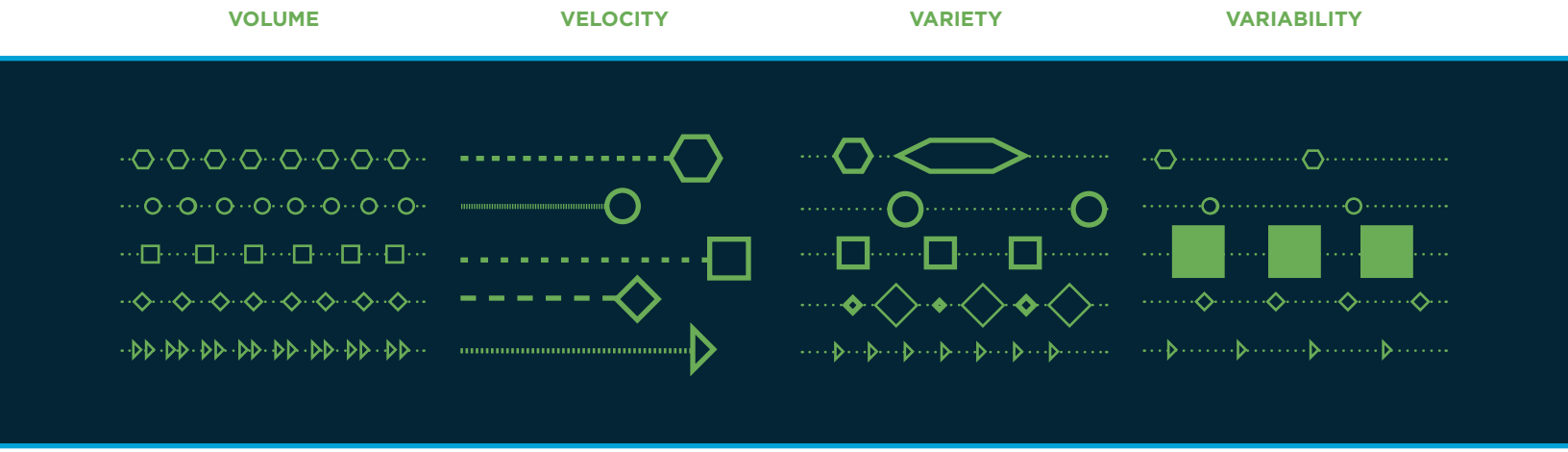
VOLUME	The quantity of data that your agency collects.
VELOCITY	The speed at which data is created.
VARIETY	The various data types that your agency has access to.
VARIABILITY	The authoritative nature of government data.

One of the most essential things to remember about big data is that the definition will continue to evolve.

AS AGENCIES CONTINUE TO COLLECT MORE DATA, WHAT IS CONSIDERED BIG TODAY MIGHT BE SMALL TOMORROW. SO AGENCIES MUST NOT FOCUS SOLELY ON THE SIZE OF THEIR DATA, BUT RATHER, WHAT THEY CAN DO WITH IT.

The promise of big data lies in its ability to unlock new insights, trends and patterns from your information to drive improved business and mission outcomes. In order to do that, agencies must overcome one of the most common barriers of big data: selecting the right problem to solve.

"Sometimes governments don't know where to begin, and how to use the data they have at their disposal," said Von Alt. "Governments are trying to figure out how to integrate their many data sources and query that data to extract value. This is so they can know what business or mission insights they should be getting out of their data."



As governments continue to deploy big data, and look for new ways to leverage information, Von Alt believes that industry needs to create not just the technical components for the public sector, but also work diligently to educate customers on the power of information and lead them down a path towards success.

"What Splunk aims to do is to make the challenges of big data more manageable, helping lead the customers down the path of here's what you need to do and here's how we can help you accomplish that technically," said Von Alt.

Another important element for success in big data is that you must be able to collect and manage all data, regardless of format. "You don't want to have a business question or a security question in your organization and find out after the fact that you can't answer it because you have discarded the data point that contained the answer," said Von Alt.

"Often you don't know in advance what questions you will have to ask of your data, so it is impossible to determine in advance what data is relevant to you, or which will be important later on," he added.

SPLUNK SOFTWARE OPERATES AS AN OPERATIONAL INTELLIGENCE PLATFORM, AND WITH IT, ORGANIZATIONS CAN BEGIN TO ASK QUESTIONS ABOUT THEIR DATA THEY PREVIOUSLY COULD NOT.

However, if data is removed, deleted or not consolidated, valuable insights might be missed.

"We help our customers determine which data is very low return on investment, but prior to that, you shouldn't take that data off the table. You just don't know what might be valuable, and so trying to retain as much data as possible is key," said Von Alt.

What is Machine Data?

Understanding machine data is a critical component to developing your big data strategy. Machine data is growing faster than any kind of data, and must be harnessed to maximize value from the information your agency creates. Machine data is created from the various ways you engage with citizens and key stakeholders. So any transactions that are made — application data, call center records or sensor data from systems — are all commonly referred to as machine data.

Machine data often arrives in ways that traditional government infrastructure were not designed to handle. So for government agencies to leverage the information, they need a new way to collect, store and manage machine data. That's where Splunk comes into play to help you understand all your data and drive improved processes.

Splunk allows you to collect all your machine data, from wherever it is created. You can also query, monitor and conduct analysis on your information in real time. This will allow you to gain new operational intelligence into your business. **Splunk provides a chart that highlights some of the different kinds of machine data.**

Every environment has its own unique footprint of machine data. Here are a few examples:

DATA TYPE	WHERE TO FIND IT	WHAT IT CAN TELL YOU
Application Logs	Local log files, log4j, log4net, Weblogic, WebSphere, JBoss, .NET, PHP	User activity, fraud detection, application performance
Business Process Logs	Business process management logs	Customer activity across channels, purchases, account changes, trouble reports
Call Detail Records	Call detail records (CDRs), charging data records, event data records logged by telecoms and network switches	Billing, revenue assurance, customer assurance, partner settlements, marketing intelligence

Examples cont'd.

DATA TYPE	WHERE TO FIND IT	WHAT IT CAN TELL YOU
Clickstream Data	Web server, routers, proxy servers, ad servers	Usability analysis, digital marketing and general research
Configuration Files	System configuration files	How an infrastructure has been set up, debugging failures, backdoor attacks, time bombs
Database Audit Logs	Database log files, audit tables	How database data was modified over time and who made the changes
Filesystem Audit Logs	Sensitive data stored in shared filesystems	Monitoring and auditing read access to sensitive data
Management and Logging APIs	Checkpoint firewalls log via the OPSEC Log Export API (OPSEC LEA) and other vendor specific APIs from VMware and Citrix	Management data and log events
Message Queues	JMS, RabbitMQ, and AquaLogic	Debug problems in complex applications and as the backbone of logging architectures for applications
Operating System Metrics, Status and Diagnostic Commands	CPU and memory utilization and status information using command-line utilities like ps and iostat on Unix and Linux and performance monitor on Windows	Troubleshooting, analyzing trends to discover latent issues and investigating security incidents
Packet/Flow Data	tcpdump and tcpflow, which generate pcap or flow data and other useful packet-level and session-level information	Performance degradation, timeouts, bottlenecks or suspicious activity that indicates that the network may be compromised or the object of a remote attack
SCADA Data	Supervisory Control and Data Acquisition (SCADA)	Identify trends, patterns, anomalies in the SCADA infrastructure and used to drive customer value
Sensor Data	Sensor devices generating data based on monitoring environmental conditions, such as temperature, sound, pressure, power, water levels	Water level monitoring, machine health monitoring and smart home monitoring
Syslog	Syslogs from your routers, switches and network devices	Troubleshooting, analysis, security auditing
Web Access Logs	Web access logs report every request processed by a web server	Web analytics reports for marketing
Web Proxy Logs	Web proxies log every web request made by users through the proxy	Monitor and investigate terms of service and the data leakage incidents
Windows Events	Windows application, security and system event logs	Detect problems with business critical applications, security information and usage patterns.
Wire Data	DNS lookups and records, protocol level information including headers, content and flow records	Proactively monitor the performance and availability of applications, end-user experiences, incident investigations, networks, threat detection, monitoring and compliance

Mitigating Insider Threats With Big Data

INSIDER THREATS PRESENT A UNIQUE CHALLENGE TO GOVERNMENT AGENCIES, IN PART BECAUSE WHAT MAY BE NORMAL ACTIVITY FOR ONE EMPLOYEE COULD BE ABNORMAL FOR ANOTHER.

In large organizations, this problem only grows in scope, making it very difficult to prevent and stop a potential threat. But what is common across all activities is that when an insider threat strikes, there was some kind of anomalous activity, where someone was accessing information or conducting activities that are outside the scope of their work.

To help prevent insider threats, Splunk software leverages data from many sources — anything from server logs, files, workstation data, network data, and the entire universe of machine data. Splunk can then use that information to create a baseline of what looks normal within an organization. They can assess each user and every IP address to predict and identify events that are outliers.

By using this technology, agencies can instantly know what activities are abnormal based on historical behavior. The behavior may not be specifically illegal or bad, but certainly something that is different outside normal business operations for an individual. With Splunk technology, this process requires no coding, and no establishing of pre-determined conditions to look for and report on. The software can provide a unique view across the data and enable the orgnization to spot abnormalities in behavior, providing managers new insights and tools they previously did not have. Once an abnormal activity takes place, the Splunk system automatically alerts administrators. This can keep agencies' confidential data more secure, and mitigates the impact of insider threats with early detection.

"You can only address insider threats through big data, because otherwise you would have a traditional database management system that stores very specific data, and then determine the data needed, identify what is abnormal activity, and have specific rules or signatures or program language to look for certain events. With Splunk you avoid that process, because you look across all sources. Agencies are also looking across historical trends and norms in real time, so you will know right away that something doesn't look quite right."

How Splunk Can Help

"Our mission for years has been to break down data silos and aggregate information together," said Von Alt.

Splunk software enables agencies to identify real-time insights from their machine data, and helps them to unlock the value of big data. Splunk also provides:

- Collection and indexing of machine data from any source
- Analysis of real-time and historical data with search language processing
- Real-time analytics and monitoring that send alerts when abnormalities occur
- Dashboards and custom views for users
- Role-based security and access control
- A platform to build big data applications

Many government agencies have data stored in a number of different places, and are beginning to understand that they must avoid creating data silos. That's why many have looked to Hadoop, an open source platform designed to store the massive volumes of structured and unstructured data. In many cases, Hadoop has become synonymous with big data. However, while Hadoop is a solid tool for low-cost data storage, it struggles as an analytics tool.

In order to provide the necessary analytics on data in Hadoop, Splunk created Hunk, an integrated analytics platform. Hunk allows agencies to visualize data in Hadoop and NoSQL data stores. "Splunk has simplified leveraging Hadoop for our customers by developing the Hunk product offering," said Von Alt. "Hunk is essentially Splunk analytics on top of Hadoop, and it offers the ability to answer queries and extract Operational Intelligence regardless of where data is stored, whether that be in a Splunk index or in your Hadoop file system."

Moving forward, organizations can anticipate even more data being created, and more value to be found by creating a strong data driven strategy. Von Alt believes that the industry is moving towards five key areas in terms of big data.

SECURITY USE CASES There will be more uses of big data to fight waste, fraud and abuse, along with the use of public safety data to combat crimes.

LEVERAGING HADOOP Hadoop will continue to grow in applications and uses, and when used in partnership with Splunk, organizations can gain even more insight into their data.

INTERNET OF THINGS (IOT) The current trends on IoT show the convergence of mobile, social and sensor data, all creating more machine data that an agency can conduct analysis on.

"As far as Splunk is concerned, data is data," said Von Alt. "We don't necessarily care if it is coming from an end-user workstation, a GPS-enabled device on a soldier in theatre, or a back-office management system. Splunk can ingest it and help you leverage it."

MOBILE DEVICE Mobile devices will continue to flood the market and allow organizations to create new services and produce more data.

DEVELOPMENT AND OPERATIONS Organizations will use data to create custom applications and create dashboards and tools to monitor key metrics and indicators at their agency.

About Splunk

Splunk Inc. provides the leading platform for Operational Intelligence. Splunk® software searches, monitors, analyzes and visualizes machine-generated big data from websites, applications, servers, networks, sensors and mobile devices. Organizations use Splunk software to deepen business and customer understanding, mitigate cybersecurity risk, improve service performance and reduce costs.

www.splunk.com

splunk>

About GovLoop

GovLoop's mission is to "connect government to improve government." We aim to inspire public-sector professionals by serving as the knowledge network for government. GovLoop connects more than 150,000 members, fostering cross-government collaboration, solving common problems and advancing government careers. GovLoop is headquartered in Washington, D.C., with a team of dedicated professionals who share a commitment to connect and improve government.

For more information about this report, please reach out to Catherine Andrews, GovLoop Director of Content, at **Catherine@govloop.com**.

1101 15th St NW, Suite 900
Washington, DC 20005

Phone: (202) 407-7421 | Fax: (202) 407-7501
www.govloop.com
Twitter: **@GovLoop**

govloop

1101 15th St NW, Suite 900
Washington, DC 20005

Phone: (202) 407-7421 | Fax: (202) 407-7501

www.govloop.com
Twitter: @GovLoop