

# The Splunk Guide to Operational Intelligence

Use Splunk and Your Machine Data to Deliver New Levels of Visibility and Insight for IT and the Business

## What is Splunk® Enterprise™?

Splunk is the engine for machine data. It collects, indexes and harnesses the machine data generated by your IT systems and infrastructure—physical, virtual and in the cloud. Use Splunk and your machine data to deliver new levels of visibility and intelligence for IT and the business.

### The Machine Data Opportunity

All your IT applications, systems and infrastructure generate data every millisecond of every day. This **machine data** contains a definitive record of user transactions, customer behavior, machine behavior, security threats, fraudulent activity and more. It's also dynamic, unstructured and non-standard and makes up the majority of the data in your organization.

Machine data is an incredibly valuable resource, but organizations rarely get the value they need from it. Existing data analysis, management and monitoring solutions are simply not engineered for this type of data.

Take information management. Data warehouses and relational database management systems are based on rigid schemas and designed for structured, consistent data. They provide historical analysis but not real-time visibility. Enterprise Search is designed for human-generated data, such as documents and Web pages. This data is very different than machine data. Machine data has an order of magnitude greater volume and diversity than traditional, structured data.

IT management tools and security information and event management on the other hand are siloed and designed for one level of the organization. They provide a narrow view of the underlying data and are hard-wired for specific data types and sources. Or they monitor across systems, with serious gaps in the data they collect. They also don't provide historical context.

The fact is finding a better way to sift, distill and understand the vast amounts of machine data can transform how IT organizations manage, secure and audit IT. It can also provide valuable insights for the business on trends and behaviors of their customers and services.

### The Splunk Approach

Splunk is the engine for machine data. It was developed to solve the whole machine data challenge and collects, indexes and

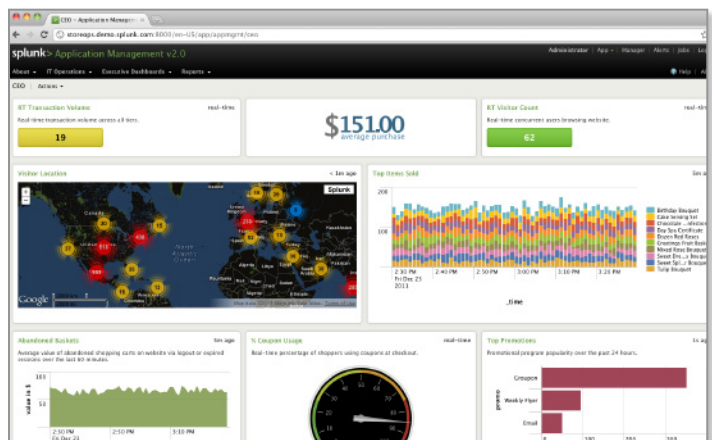
harnesses your unstructured, time-series machine data. Splunk can read data from just about any source imaginable, such as network traffic, Web servers, custom applications, application servers, hypervisors, GPS systems, stock market feeds, social media and preexisting structured databases. It delivers a real-time understanding of what's happening and deep analysis of what's happened across your IT systems and infrastructure. It turns your machine data into the insights you need to make informed decisions.

Splunking machine data has many uses for IT and the business:

- **Application Management:** troubleshoot across application environments; monitor for performance degradation
- **Security and Compliance:** provide rapid incident response, correlation and in-depth monitoring across data sources
- **Infrastructure and Operations Management:** proactively monitor to ensure uptime; rapidly pinpoint and resolve problems
- **Web and Business Analytics:** gain visibility and intelligence on customers, services and transactions; identify trends and patterns in real time

Finding and fixing problems, following the trail of an attacker, reporting for compliance and analyzing customer behavior requires a complete view.

Troubleshooting problems often means correlating Web server logs, SOA messages, database transactions, virtual performance and configuration changes.



Build custom dashboards in just a few clicks for IT and business users.

Investigating security incidents demands both the analysis of events from server logs, firewalls and IDS scans, in addition to application events, configurations and scripts to understand what's happened.

Meeting compliance requires systematic reviews and long-term data retention from across the infrastructure, placing more barriers to accessing this data for day-to-day operational needs.

When the business seeks better intelligence, this may require real-time correlation and analysis of transactions and events from many IT sources, potentially combined with business data.

Splunk arms network engineers, system administrators, security and compliance analysts, developers, support/service desk staff and business users alike with new levels of visibility all from one solution. We call this delivering Operational Intelligence.

## How is Splunk Different?

Splunk is different from previous approaches to managing, auditing, securing and gathering intelligence from IT systems. Here's how.

**Immediate results without the risk.** Splunk is enterprise software made easy. Users can download Splunk for free, install it in minutes, feed it any machine data, and immediately get productive. No more armies of consultants, or a DBA to make it work. The proof is immediate. Most users download and install Splunk while they're under fire. A serious service problem or security incident can now be investigated in a few minutes, versus the hours or days it used to take.

**Based on high-performance indexing and search technology.** Every day millions of people search and navigate billions of Web pages served by computers all over the world. Search is flexible, intuitive and delivers immediate results. Splunk has at its core powerful indexing and search technology, bringing a whole new meaning to speed and responsiveness. With it you can search billions of events in seconds and start seeing results immediately.

**Designed for time-series, unstructured data.** Machine data is unstructured data. It's time-series based, dynamic and non-standard. It captures all machine-to-machine and human-to-machine interactions, generated at volumes that far surpass structured enterprise data. It's also growing at an exponential rate. Splunk uses no predefined schema, so can read data in any format and from just about any source imaginable.

**Indexes data from any source.** System management, SIEM, CEP/ECA and log management products require weeks or months developing or configuring custom connectors for each data source. Splunk directly collects data from tens of thousands of sources, bringing it securely to a central location in real time. In situations where the data you need isn't available over the network, you can deploy Splunk forwarders. Forwarders are lightweight and provide secure, distributed, real-time universal data collection. Monitor local application log files, capture the output of status commands on a schedule, grab performance metrics from virtual or non-virtual sources or watch the file system for configuration, permissions and attribute changes.

**Enables analysis of real-time and historical data.** Traditional IT systems force a decision between real-time monitoring or historical analysis. With Splunk you can search and analyze real-time streaming and historical data from one solution. This means you can identify and respond to patterns of behavior or activity of interest before it's too late.

**Software that users want to use.** It used to make sense to manage your IT infrastructure in silos. But with today's distributed, scaled out computing and the proliferation of complex Web-based applications and virtualization, this just doesn't work anymore. Splunk breaks down the IT silos. Search, report, monitor and analyze all your data from every application, server and device—all from one place in real time. Easily integrate with existing enterprise management, security and compliance tools. Finding and fixing problems, following the trail of an attacker, tracing transactions and gaining new insights from your operational data is suddenly a whole lot faster and easier.

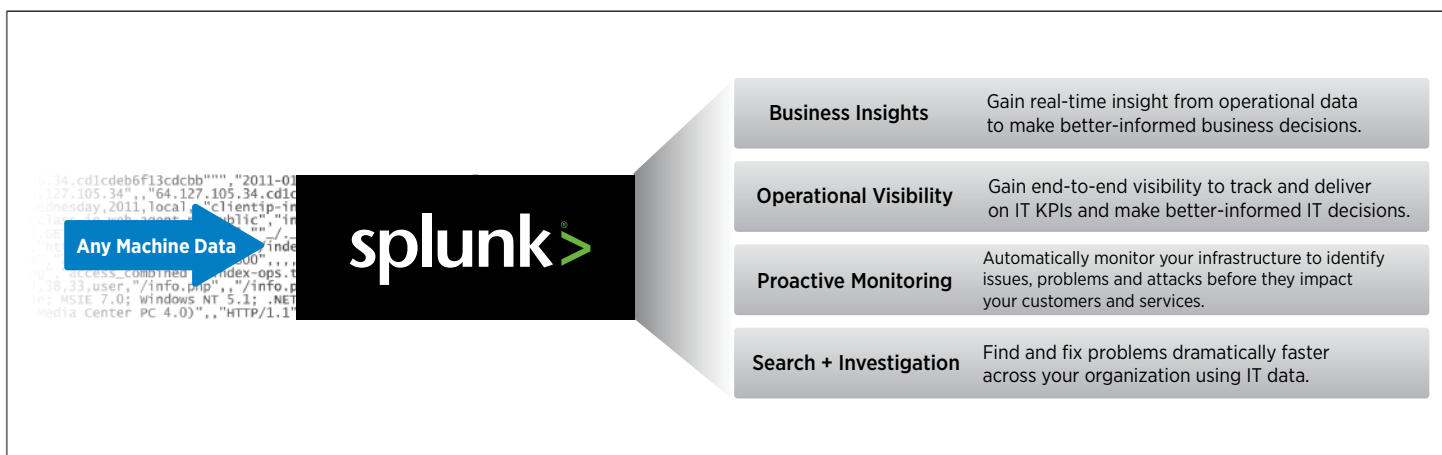
**Create custom dashboards and views.** Splunk helps make sense of huge volumes of machine data to satisfy the needs of different users and groups in the organization. Quickly create custom dashboards that integrate multiple charts and views of your real-time data and view them from your desktop or mobile device. Personalize dashboards for different users in your organization—managers, business analysts, security analysts, auditors, developers and sysadmins. Users can edit dashboards using a simple drag-and-drop interface. Integrated charting controls mean they can change chart types on-the-fly.

**Do more with Splunk Apps.** Create apps on Splunk that deliver a targeted user experience for specific use cases and technologies. You can share and reuse apps within your organization and the rest of the Splunk community. There are a growing number of apps available on our community site ([www.splunkbase.com](http://www.splunkbase.com)), built by our community, partners and Splunk. Apps for Enterprise Security, or for compliance; apps for different platforms, such as Windows, Linux and Unix; and apps for different technologies, such as networking, virtualization and more.

**Keeps up with change.** The only constant in today's complex, virtualized and hybrid IT environments is change. What we think we know is often wrong. Traditional IT management and security approaches assume users know all the possible failures and risks up front and that data formats won't change. This just isn't the case anymore. In fact, most IT organizations spend more time customizing and maintaining their tools than they do using them.

Splunk doesn't rely on brittle schemas that limit flexibility and break when data formats change. Splunk indexes all the data you point it at in real time, all the time. Any interpretation of the data you need, such as extracting a field, tagging a subset of hosts, can be easily done on-the-fly—as you search.

**Scales from the laptop to the datacenter.** During a time when every expense is scrutinized, organizations don't always have the resources they need. That's why Splunk is priced and built to fit all environments. It can be downloaded and run on a laptop in under 5 minutes and the same software can be scaled out across the largest global infrastructures, indexing tens of terabytes of data a day.



Splunk delivers Operational Intelligence.

## Delivering the Key Capabilities for Operational Intelligence

- **Universally collect and index** machine data, from virtually any source
- Enable freeform **search and incident investigations** from one place
- Automatically **discover knowledge** from the data and let users add their own
- **Monitor** your data and provide **real-time alerts** when specific conditions arise
- Provide powerful **reporting and analysis**
- Provide the ability to **create custom dashboards and views** for different roles
- **Scale** efficiently using commodity hardware
- Provide granular **role-based security and access controls**
- Support **multi-tenancy** and be flexibly deployed

## Universal Indexing

Individual components in your infrastructure generate hundreds of events per second. A datacenter can log many terabytes of data per day. You'll probably begin wondering how you're going to access all this data in all the different formats and locations. Splunk offers a variety of flexible input methods and doesn't need special connectors for specific data formats. So you can immediately index logs, clickstream data, configurations, traps and alerts, messages, scripts, performance data and statistics from your applications, servers and network devices—physical, virtual and in the cloud.

**Flexible data input.** Splunk collects and indexes data from just about any source imaginable, such as network traffic, web servers, custom applications, application servers, hypervisors, GPS systems, stock market feeds, social media and preexisting structured databases. No matter how you get the data, or what format it's in, it's indexed the same way—without any specific parsers or connectors to write or maintain.

**Forwards data from remote systems.** Splunk forwarders can be deployed in situations where the data you need isn't available

over the network or visible to the server where Splunk is installed. Splunk forwarders deliver secure, distributed, real-time universal data collection for tens of thousands of sources. They can monitor local application log files, capture the output of status commands on a schedule, grab performance metrics from virtual or non-virtual sources or watch the file system for configuration, permissions and attribute changes. They are lightweight, can be deployed quickly and at no additional cost.

**Real-time indexing.** IT people depend on up-to-date information for troubleshooting, security incident investigations, compliance reporting and other valuable tasks. Splunk continually indexes machine data in real time—your logs, configuration data, change events, the output of diagnostic commands, data from APIs and message queues, even logs from your custom applications.

**Captures everything.** Splunk stores both the raw data and the rich index in an efficient, compressed, filesystem-based datastore, with optional data signing and auditing to prove data integrity.

**No rigid schemas.** Splunk has no predefined schema. Solutions that rely on brittle schemas have limited flexibility and break when data formats change. Any interpretation you need to do on the data, such as extracting a common field, or tagging a subset of hosts—is done at search time.

**Automates chronology.** All this streaming data means extracting, and normalizing timestamps is very important. Splunk automatically determines the time of any event—even with the most atypical or non-traditional formats. Data missing timestamps can be handled by inferring timestamps based on context.

## Search and Investigation

Splunk lets users search and navigate their data from one place.

**Search and investigate anything.** Freeform search supports intuitive Boolean, nested, quoted string and wildcard searches familiar to anyone comfortable on the Web. This allows users to quickly iterate and refine their searches without knowing anything about specific data formats.

**Real-time search.** Searching real-time streaming data and indexed historical data from the same interface is best-in-class. With Splunk you can analyze behavior and activity in real time and see the historical context.

**Time search.** Given the large volume and repetitive nature of machine data, users often start by narrowing their search to a specific time range. With the focus on when events happen, Splunk lets users combine time and term searches. This ability to search across every tier of your infrastructure for errors and configuration changes in the seconds before a system failure occurs, is incredibly fast and powerful.

**Interactive results.** Compared to command line scripts and tools, an interactive interface dramatically improves the user's experience and the speed with which tasks can be accomplished. Zoom in and out on a timeline of results to quickly reveal trends, spikes and anomalies. Click to drill down into your results and eliminate noise to get to the needle in the haystack. Whether you're troubleshooting a customer problem or investigating a security alert, you'll get to the answer in seconds or minutes rather than hours or days.

**Transaction search.** Sending an email, placing an order on a Website or connecting a VOIP call will create a number of events across different IT components. Often you'll want to search for these collections of events that are all part of the same transaction. For example, find all the sendmail events with the same user-ID, between a login and a logout, that occur within 10 minutes.

Splunk lets you correlate events by finding common characteristics and then saving that search as a transaction so you can find the same type of transactions again for different search parameters.

## Add Knowledge

Splunk automatically discovers knowledge from your data and lets users add their own, unlocking your data's full potential. Knowledge about events, fields, transactions, patterns and statistics can be added to your data. You can identify, name and tag this data as well. Go from finding all events with a particular username, to instantly getting statistics on specific user activities. You can also correlate and name transactions that span multiple data sources. Splunk marries the flexibility of freeform search with the power of working with your data, in a way you've never experienced before.

**Map knowledge at search time.** Splunk avoids the problems caused by traditional approaches, by mapping knowledge to data at search time, rather than attempting to normalize the data into a brittle database schema up front. And there's no more need for the complex management of custom parsers and connectors. Easily enrich your machine data with information from external asset management databases, configuration management systems and user directories. Now you have a flexible way to manage your data, so as it changes, you don't have to.

**Work smarter.** Splunk lets every user add their own knowledge as they go. As you're saving searches, identifying different types of fields, events and transactions you make the system smarter for everyone else. And that knowledge doesn't walk out the door when someone leaves.

## Monitor and Alert

Rather than use search to simply react to ad hoc incidents or problems, you want to be proactive. Splunk provides flexible alerting capabilities that improve your monitoring coverage. And because it works across your entire IT infrastructure, it's the most flexible monitoring solution in your arsenal.

**Turn searches into real-time alerts.** Searches can be saved and scheduled for continual monitoring and can trigger alerts via email or RSS. You can even kick off a script to take remedial actions, send an SNMP trap to your system management console or generate a service desk ticket. Scheduling alerts is a great way to complete the investigation of a problem or security incident by proactively looking for similar occurrences in the future.

**Correlate complex events.** Splunk lets you correlate complex events from multiple data sources across your IT infrastructure so you can monitor more meaningful events. For example, you can track a series of related events as a single transaction to measure duration or status.

**Monitor for specific conditions.** Alerts can be based on a variety of threshold and trend-based conditions and to any level of granularity. The search language goes beyond simple Boolean searches into fielded searches, statistical searches and sub-searches, you can correlate on anything you want and alert on complex patterns such as abandoned shopping carts, brute force attacks and fraud scenarios.

## Report and Analyze

If you've ever wanted to generate a report on-the-fly from hard-to-understand machine data, you'll love Splunk. Create powerful, information-rich reports to do analysis, without an advanced knowledge of search commands. You can schedule delivery of any report via PDF and share it with management, business users or other IT stakeholders.

**Report on search results.** Easily build advanced graphs, charts and sparklines from search results and visualize important trends, see highs and lows, summarize top values and report on the most and least frequent types of conditions. The simplicity of analyzing massive amounts of data will amaze you (and your boss). For example, a report can show the total bytes sent by IP address from firewall activity events; a table showing bytes per protocol per IP address; or a chart illustrating firewall traffic by hour for a specific employee's laptop. Virtually any field can be used as reporting criteria. And remember, because fields are identified as your search you can specify new fields without re-indexing your data.

**Analyze correlated events.** Splunk supports five types of correlation. Time-based correlations, to identify relationships based on time, proximity or distance. Transaction-based correlations to trace transactions that span multiple silos, systems and data sources so you can report on and analyze important activities, such as duration to complete a new service transaction, or determine whether a complex transaction actually completed or not. Sub-searches, taking the results of one search and using them in another. Lookups, correlating with external data sources outside of Splunk. Joins, to support SQL-like inner and outer joins.

**Plays well with others.** Now your entire organization can leverage the value of machine data. Reports can be saved and shared with management or other colleagues in secure, read-only formats, such as PDF and even integrated into dashboards.

## Custom Dashboards and Views

Make more sense of the huge volumes of data at your disposal. Splunk lets you create custom dashboards and views for different types of users, technical and non-technical. Integrate reports, search results and even data from external applications. Edit dashboards using a simple drag-and-drop interface; integrated charting controls mean you can change chart types on-the-fly. Doing this all through the Splunk UI means that you can empower business users to do the same.

**Live dashboards.** Dashboards integrate multiple charts, views and reports of live and historical data to satisfy the needs of different users. Best-in-class engines offer the ability to personalize dashboards for management, business or security analysts, auditors, developers and sysadmins.

**Mashups with other apps.** Splunk provides the ability to create mashups with other web-based apps, such as Tivoli, SAP, security consoles and more to provide a seamless view across silos.

**Dashboards anytime, anywhere.** Charts and timelines in Splunk don't use Flash, which means dashboards can be viewed and edited on the go on mobile devices, or in browsers that do not have Flash installed.

## Create and Download Splunk Apps

Now you're indexing and making use of all your machine data, you can make use of apps that let you do even more.

**Innovate on your own.** Splunk makes it easy to create apps that deliver a targeted user experience for different roles and use cases. The Splunk App framework provides the ability to develop and package Apps through a single user interface. Deliver a user experience tailored to a specific use case or augment existing vendor technologies.

**Share and download apps.** You can share and reuse apps within your organization and the rest of the Splunk community. There are a growing number of apps available on our community site [www.splunkbase.com](http://www.splunkbase.com), built by our community, partners and Splunk. You can find apps that help visualize data geographically, or that support specific use cases, such as enterprise security or PCI compliance. There are also apps for different operating systems and third-party technologies, such as Windows, Linux, Blue Coat, Cisco, WebSphere and F5 Networks.

**Easy management.** Once installed you apply role-based access controls and deploy apps with a tailored user experience across the organization, extending the value of your data to different users.

## Massive Scalability

With Splunk you can scale your installation from a single commodity Windows, Linux or Unix server, to the largest most complex multi-geography, multi-datacenter infrastructures indexing tens of terabytes of data per day. The Splunk architecture is based on the MapReduce framework and scales linearly across commodity servers to unlimited data volumes. You'll find a wide range of options to access data, store it, search it and route it to other systems.

**Easy installation.** A self-contained software package with no dependencies on third-party programs makes Splunk easy-to-install and get running. It works on all major operating systems and hardware platforms. And because Splunk is software, it can operate across physical or virtual infrastructures rather than requiring dedicated hardware, power and rack space.

**Analyzes big data.** Your datacenter generates more machine data than you probably ever imagined. A single production server can generate hundreds of megabytes of data a day. Firewalls and Web servers can each generate many times that amount. In fact, machine data is one of the fastest, most complex segments of big data.

This volume of data is also subject to retention requirements ranging from a few days for incident response, to months and years for compliance.

Based on the MapReduce framework, Splunk scales linearly across commodity hardware. When considering performance and comparing approaches to collect, index and harness your machine data here are some things to look for and consider:

**Indexing throughput.** Events-per-second (EPS) is a common throughput measurement, but consider that event sizes can vary from a few hundred bytes to a megabyte or more. EPS ratings are usually calculated at whatever size is optimal for one specific vendor's appliance or solution. Look for vendors that index every byte in your data, without the need for custom parsers or connectors. If the vendor is unable or unwilling to quote you EPS figures based on this criteria, move on and find someone who will.

**Search speed.** Searches of any type should return results in seconds, not minutes or hours. Based on a distributed computing framework, Splunk automatically converts searches into a parallel program providing the ability to quickly retrieve and analyze massive data sets. A single commodity server will support searching of billions of events in seconds.

**Storage efficiency.** Measured as a percentage of the original data stream size, storage efficiency determines the amount of storage capacity you'll need to retain your data and the associated indexes. A good solution will require 25% to 50% of the original data size to retain your data and a useful set of indexes. Beware of solutions that claim 10% or less of original data size. That indicates just the storage of compressed data and no indexing.

**Archiving.** Eventually you may decide to tier the storage of your data. Tiered storage can provide lower cost and better redundancy. Archiving data based on disk utilization or age will come in handy for building a multi-tiered data store. Make sure your solution lets you set up an archiving policy based on datastore size or age and restore your archives on demand.

**Linear scalability.** You can scale Splunk horizontally and vertically by simply adding more computing power. You can run a distributed configuration on different physical servers, a combination of virtual and non-virtual servers, or on a large multi-core, multi-processor machine. Splunk lets you balance workloads by configuring multiple indexers and search engines across your configuration.

**Distributed search.** Often it won't be feasible to physically centralize all your data in one place. You will likely need to search across multiple installations and data stores in different technology or geographic silos.

**Data routing and cloning.** With all the data streams to manage, you'll want the ability to route data based on characteristics and content. This will be important to scale and secure your Splunk installation. And as you come to depend on Splunk as a mission critical part of your IT infrastructure you'll probably want to clone important data to multiple servers for high availability.

**Integration.** If you're like most IT shops, you've made significant investments in other management tools, monitoring tools and analysis tools. Wouldn't it be great if you could integrate Splunk with all of them? Imagine launching in-context searches from your network management console, sending Splunk alerts to your system management console, or automating trouble ticket creation when unusual activity occurs. Splunk provides multiple integration points and a robust, documented API.

## Security

You'll need to keep your machine data secure. Especially as you realize what a valuable information asset you have. Splunk provides secure data handling, access controls, audit-ability, assurance of data integrity and integration with enterprise single sign-on solutions.

**Secure data access and transport.** Machine data can be sensitive. Splunk supports advanced anonymization to mask confidential data from results. Private consumer or corporate information also requires secure access, transport and storage. You should evaluate potential solutions for encrypted access to data streams using something like TCP/SSL. Make sure user access is secured using something like HTTPS or SSH for command-line access.

**Granular access control.** Of course you also need the ability to control the actions users can take and what data, tools and dashboards they can access. You don't necessarily want to allow the application development team access to your IDS scans, alerts and firewall logs. Splunk is a flexible, role-based system that lets you build your own roles to map to your organization's policies for different classes of users.

In some environments, like multi-tenant services, you may need to physically control access to data. The ability to route select data to distinct Splunk installations will let you physically separate data in different data stores. You'll also want to integrate with LDAP and Active Directory and map groups to different roles

**Single sign-on.** If you're using access controls internally and have organizational access control policies, you'll want to make sure you can integrate your Splunk solution with your

authentication system whether it's LDAP, Active Directory, e-Directory or another authentication system.

**Audit capability.** Once you have your access controls set-up, you need to monitor who's doing what. Splunk logs administrative and user activities so you can audit who's accessing what data and when.

**Data integrity.** You'll also need to ensure the integrity of your data. How do you know the search results or report you're viewing is based on data that hasn't been tampered with? With Splunk, individual events can be signed and streams of events block signed. Splunk also provides message integrity measures that prove nobody has inserted or deleted events from the original stream.

**Hardened deployment.** Keeping an audit trail and signing events is worthless if the server running Splunk can be compromised. Be sure your vendor provides hardening guidelines.

## ROI and Splunk

Splunk customers typically achieve an ROI measured in weeks or months, sometimes even before being deployed into production. Splunk users can troubleshoot application problems and investigate security incidents in minutes instead of hours or days, dramatically improve service levels, reduce outages and deliver compliance reporting at a lower cost. This visibility, typically unavailable prior to Splunk, delivers organizations a fast ROI, new productivity and powerful insights. Here are a few examples:

- A leading provider of healthcare management solutions avoided a \$100K SLA penalty—found during the Splunk evaluation phase. This same customer achieved an annual ROI of over \$700,000.
- One of the world's largest business publishers replaced their old server monitoring software with Splunk and other open source software. This eliminated maintenance fees and reduced operations costs by \$1.6 million/year.
- A major communications manufacturer avoided a \$1.5M software license upgrade for their existing SIEM, reassigned 5 full-time analysts to other duties (\$600,000/year) and now monitors new data sources to identify previously unknown attacks.
- The world's largest B2B poker provider, hosting 25 of the industry's top brands and up to 45,000 concurrent players at peak hours, reduced downtime by 30% and quantified an annual savings of \$1.9 MM (16x ROI in the 1st year).
- One of the world's largest online travel sites demonstrated an annual ROI over \$14 million. This ROI was a combination of tools consolidation, retired licenses, outage avoidance and troubleshooting efficiencies gained using Splunk.

### Free Download

Download Splunk for free. You'll get a Splunk Enterprise license for 60 days and you can index up to 500 megabytes of data per day. After 60 days, or anytime before then, you can convert to a perpetual Free license or purchase an Enterprise license by contacting [sales@splunk.com](mailto:sales@splunk.com).

## Seeking a best-in-class solution for managing your machine data? Here's what to look for:

<b>1</b>	<b>Index Live Data</b>
<b>a</b>	Indexes any machine data generated by applications, servers or network devices including logs, clickstream data, configurations, messages, traps and alerts, metrics and performance data without custom parsers or connectors for specific formats (includes virtual and non-virtual environments).
<b>b</b>	Flexible real-time and on-demand access to data from files, network ports and databases and custom APIs and interfaces.
	Listens to TCP and UDP network ports to receive syslog, syslog-ng and other network inputs.
	Consumes archive files.
	Captures new events in live log files in real time.
	Monitors files for changes.
	Queries database tables via DBI.
	Monitors Windows events remotely via WMI.
	Natively accesses the Windows event API.
	Monitors the Windows registry for changes.
	Connects to OPSEC LEA and other key security event protocols.
	Subscribes to message queues such as JMS.
	Captures the output of Unix/Linux system status commands like ps, top and vmstat.
	Remotely copies files via scp, rsync, ftp and sftp.
	Extensible via scripted inputs to capture the output of new status commands, connect to new event APIs and subscribe to different kinds of message queues.
<b>c</b>	Universally indexes data in virtually any format without custom parsers or connectors for specific data formats.
	Identifies events in single line, multi-line and complex XML structures.
	Recognizes and normalizes timestamps. Handles bad or missing timestamps through contextual inference.
	Captures and indexes the structure of each event.
	Tracks and indexes the host and source of each event.
	Classifies source formats dynamically.
<b>d</b>	Densely indexes every term in the original data.
<b>e</b>	Retains original, unaltered machine data.
<b>f</b>	Builds an unstructured index on disk without schema.
<b>g</b>	Supports forwarding and receiving of data from remote hosts for load balancing, failover and distributed deployments.

2	Search and Investigate
a	Search events across components in multiple formats at once.
b	Search live and historical data from the same interface and automatically backfill historical data for real-time windowed searches.
c	Fast results from searches on terms instead of queries optimized for specific fields/columns in a persistent schema.
d	Free form ad-hoc search on any term in the original events with support for Booleans, nesting, quoted strings and wildcards.
e	Precise searches using fields identified within the data at search time. Supports multiple schema views into the same data without redundant storage or re-indexing.
f	Type-ahead suggestions to make it easy to discover what to search.
g	Navigate to related events and refine searches by clicking on fields or terms within the search results.
h	Search by time across multiple data formats.
i	Visualize trends and navigate results using interactive time-based charts, histograms, sparklines and summaries.
j	Search for transactions across different data sources and components.
k	Persist searches as event and transaction types and search, filter and summarize by event and transaction type.
l	Discover fields, event types and transactions interactively at search time.
m	Save searches in reports, dashboards or views to simplify routine search scenarios.
n	Browser based, interactive AJAX user interface. No plug-ins required.
o	Optional scriptable CLI interface for both real-time and historical search.

3	Add Knowledge
a	Enable the system and the user to automatically add semantic meaning to machine data.
b	Automatically discovers knowledge from the machine data, such as timestamp, name/value pairs, headers, etc.
c	Let users add additional knowledge about the events, fields, transactions, and patterns in their machine data.
d	Assign tags to field values to help search groups of events with related field values more efficiently.
e	Identify and classify transactions by correlating events across multiple data sources.
f	Save searches that return interesting results by either saving the search string (to run the search later) or the search results (to review the results later).
g	Share and promote saved searches, saved reports and event types with other authorized users.

4	Monitor and Alert
a	Run time-based search on a schedule and set alerting conditions based on thresholds and deltas in the number and distribution of results.
b	Trigger alerts via email, RSS, SNMP or scripts.
c	Take automated corrective or follow-on actions via scripted alerts.
d	Embed sophisticated correlation rules in alerts via sub-searches.

5	Report and Analyze
a	Build summary reports based on the results of any search interactively by clicking on available fields and statistics.
b	Create reports using fields and schemas identified at search time. Supports multiple schema views into the same data without redundant storage or re-indexing.
c	Supports sophisticated statistical and summary analysis by pipelining advanced search commands together in a single search.
d	View report results in a tabular form.
e	View report results as interactive line, bar, pie, scatterplot and heat map charts.
f	Pivot or drill down into any field or term.
g	Schedule searches or report for automated delivery via email or RSS.
h	Cache the results of scheduled reports for re-use.
i	Create real-time reports based on live streaming data sources.
j	Schedule the delivery reports via PDF.

6	Create Custom Dashboards and Views
a	Create and edit dashboards that combine searches, reports, charts and tables using a visual dashboard editor.
b	Build sophisticated dashboards with entirely custom user interfaces and rich visualizations, including mashups with other applications and data from external sources.
c	Provide pre-packaged dashboards depicting key information and user activity—such as admin activity, search activity, index activity and inputs activity.
d	Provide summary indexing to efficiently report on the very large volumes of data, e.g., long-term trends.
e	Expand or restrict the role-based read and write permissions for a dashboard.
f	Create composite dashboards based on live and historical data sources. Deploy dashboards to devices and web browsers that do not support Flash.
g	Schedule the delivery of any dashboard via PDF.

7	Build and Deploy Apps
a	Provide the ability to build and deploy apps on top of the machine data engine for specific use cases.
b	Package custom dashboards and configurations ranging from scripts, knowledge objects and back-end settings as apps.
c	Easily browse and dynamically switch between apps running on the machine data engine instance by using an app launcher interface. Instantly see all installed apps on the machine data engine instance that the user has permissions to see.
d	Provide a powerful framework to support the creation of robust apps at all levels.
e	Expand or restrict the role-based read and write permissions to the app.

8	Integration
a	Provide APIs to enable the quick integration with other applications, IT management tools and systems.
b	Minimum interface requirements should include, command-line Interface, DBI, data routing, documented SDKs, REST, scripted alerts, scripted inputs.

9	Scale and Deploy
a	The machine data engine server is a self-contained software package with no dependencies on third-party programs. IT runs in virtualized server and storage environments.
b	Native packages (rpm, deb, pkg, dmg, msi, etc.) and archive format distributions (.tgz., .zip, .tar.Z) are available for most widely-deployed operating systems including Linux, Windows, Solaris, HP-UX, AIX, Free BSD and Mac OSX.
c	Servers work together support both centralized and decentralized models for machine data management across the organization.
d	Provides real-time centralization of machine data from production servers with reliable data transport over TCP.
e	Distributed architecture to support highly available configurations with integrated failover and load balancing.
f	Policy-based data routing among servers and to third-party systems.
g	Linear scaling to terabytes per day via distributed search and data balancing based on the MapReduce technique.
h	Single view across silos via distributed search.
i	Maintains a complete, signed audit trail of administrative actions and search history.
j	Monitors its own configurations for unauthorized change.
k	Centralized, policy-based configuration management across servers in a distributed deployment.

<b>l</b>	REST API enables quick integration with other IT management tools and systems.
<b>m</b>	Tunable indexing levels can be set for different sources or events.
<b>n</b>	Extremely fast search speed, delivers results in seconds across billions of events.
<b>o</b>	Highly efficient compressed storage - 12-48% of the original data size typical for syslog depending on indexing level.
<b>p</b>	Datastore uses local or network storage and is compatible with incremental file system back-up utilities.
<b>q</b>	Index is segregated by time to support extended retention times without impact to search performance.
<b>r</b>	Configurable archiving and data retirement policy by age or size.
<b>s</b>	Archive and restore compressed or fully indexed data on demand. Facilitates maintaining oldest data using lower cost nearline storage for extended retention times.
<b>t</b>	Integrated use of MapReduce to enable scaling of real-time and historic search functions across commodity hardware.

<b>10</b>	<b>Secure</b>
<b>a</b>	Flexible roles for controlled user and API access. Supports granular data access and capabilities by role. Enables restricted access to specific data sources, data types, time periods, specific views, reports or dashboards.
<b>b</b>	Authentication and authorization integration with Active Directory, eDirectory and other LDAP-compliant implementations.
<b>c</b>	Integration to enterprise single sign-on solutions enabling pass-through authentication of third party credentials.
<b>d</b>	Real-time remote indexing of data to minimize the opportunity for alteration of audit trails on compromised hosts.
<b>e</b>	Secure data stream access and distributed functionality via SSL/TCP. Secure user access via HTTPS.
<b>f</b>	Block-signs events to demonstrate data integrity.
<b>g</b>	Maintains a complete, signed audit trail of administrative actions and search history.
<b>h</b>	Monitors its own configurations for unauthorized change.