

The Essential Guide to Data

An update to *The Essential Guide to Machine Data*, exploring how to bring data to every question, decision and action

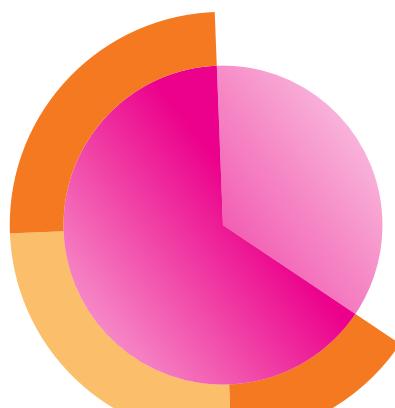


011010

Time-Series Data. Streaming Data. Dark Data.

It's no secret that data remains underused and undervalued in most organizations all over the world. Despite the constant talk of data-driven decisions, organizations of all sizes are still missing the mark on how to effectively capture and use the troves of data being generated every day, whether it comes from users, outside industry resources, or their own networked devices. In fact, most business and IT decision makers estimate that **55% of their data is dark data**, information you don't know you have, or can't fully tap.

This is a big missed opportunity. Important insights across IT, security and your organization lie hidden in this data. Data holds the definitive record of all activity and behavior of your customers and users, transactions, applications, servers, networks, mobile devices and more. Critical information on everything from configurations, APIs, message queues, diagnostic outputs, sensor data of industrial systems and more is all there — you just have to tap into it the right way.



With the right approach, data makes it simple to:

- Make better informed decisions about every part of your business.
- Run your operations more efficiently.
- Optimize user and customer experiences.
- Detect the fingerprints of fraud — or prevent it altogether.
- Uncover potential disasters before they happen.
- Find hidden trends that help your company leapfrog the competition.
- Make everyone who uses it look like a hero.
- ... and so much more.

The challenge with leveraging the vast quantity of data that most companies collect is that it comes in a dizzying range of formats that traditional data monitoring and analysis tools aren't designed to handle. Many tools can't keep up with the varying data structures, sources or time scales. And it goes well beyond just machine data as well. But the upside to tapping into your data is tremendous, and this is where Splunk comes in.

With Splunk, you can bring data to every question, decision and action in your organization to create meaningful outcomes. Unlike any other platform, Splunk is truly able to take any data from any source and drive real action to benefit the business — from IT infrastructure and security monitoring to DevOps and application performance monitoring and management.

Turn Data Into Doing in Practice

Use data to:



The organizations that get the most value out of their data are those able to take disparate data types, enrich them and extract answers. But not knowing what data to ingest can stop businesses before they start.

Familiarizing yourself with general use cases in security, IT operations, business analytics, DevOps, the Internet of Things (IoT) and more — including the data types and sources involved — can get you on track right away.

Here's an example:

1. A customer's order didn't go through
2. The customer called support to resolve the issue
3. After too much time on hold, the customer gave up and tweeted a complaint about the company

What Does Machine Data Look Like?

Sources	
Order Processing	ORDER, 05-21T14:04:12.484,10098213, 569281734,67,17,10,12,43CD1A7B8322,SA-2100 MAY 21 14:04:12.996 wl-01.acme.com Order 569281734 failed for customer 10098213. Exception follows: weblogic.jdbc.extensions.ConnectionDeadSQLException: weblogic.common.resourcepool.ResourceDeadException: Could not create pool connection. The DBMS driver exception was: [BEA]Oracle JDBC Driver Error establishing socket to host and port: ACMEDB-01:1521, Reason: Connection refused
Middleware Error	05/21 16:33:11.238 [CONNEVENT] Ext 1207130 (0192033): Event 20111, CTI Num:ServID:Type 0:19:9, App 0, ANI T7998#1, DNIS 5555685981, SerID 40489a07-7f6e-4251-801a-13ae51a6d092 05/21 16:33:11.242 [SCREENPOPEVENT] SerID 40489a07-7f6e-4251-801a-13ae51a6d092 CUSTID 10098213 05/21 16:37:49.732 [DISCEVENT] SerID 40489a07-7f6e-4251-801a-13ae51a6d092
Care IVR	(actor:displayName: "Go team!",followersCount:1366,friendsCount:789,link: http://dallascowboys.com/location [displayName:Dallas, TX] objectType:place"), objectType:person,preferredUsername:'BoysF@n80',statusesCount:6072),body: "Can't buy this device from @ACME. Site doesn't work! Called, gave up on waiting for them to answer! RT if you hate @ACME!",objectType:activity",postedTime:'05-21T16:39:40.647-0600')
Twitter	

Figure 1: Data can come from any number of sources, and at first glance, can look like random text.

Machine Data Contains Critical Insights

Sources	Customer ID	Order ID	Product ID
Order Processing	10098213	569281734	67,17,10,12,43CD1A7B8322,SA-2100
Middleware Error	MAY 21 14:04:12.996 wl-01.acme.com Order 569281734 failed for customer 10098213. Exception follows: weblogic.jdbc.extensions.ConnectionDeadSQLException: weblogic.common.resourcepool.ResourceDeadException: Could not create pool connection. The DBMS driver exception was: [BEA]Oracle JDBC Driver Error establishing socket to host and port: ACMEDB-01:1521, Reason: Connection refused	Order ID	adSQLException
Care IVR	05/21 16:33:11.238 [CONNEVENT] Ext 1207130 (0192033): Event 20111, CTI Num:ServID:Type 0:19:9, App 0, ANI T7998#1, DNIS 5555685981, SerID 40489a07-7f6e-4251-801a-13ae51a6d092 05/21 16:33:11.242 [SCREENPOPEVENT] SerID 40489a07-7f6e-4251-801a-13ae51a6d092 CUSTID 10098213	Customer ID	Time waiting on hold
Twitter	(actor:displayName: "Go team!",followersCount:1366,friendsCount:789,link: http://dallascowboys.com/location [displayName:Dallas, TX] objectType:place"), objectType:person,preferredUsername:'BoysF@n80',statusesCount:6072),body: "Can't buy this device from @ACME. Site doesn't work! Called, gave up on waiting for them to answer! RT if you hate @ACME!",objectType:activity",postedTime:'05-21T16:39:40.647-0600')	Customer's Tweet!	Customer's Twitter ID

Figure 2: The value of data is hidden in this seemingly random text.

The Essential Guide to Data

Machine Data Contains Critical Insights

Sources

- Order Processing
- Middleware Error
- Care IVR
- Twitter

Customer ID	Order ID	Product ID
ORDER_05-21T14:04:12.484 10098213 56926 734 67.17.10.12.43CD1A7B8322 SA-2100		
MAY 21 14:04:12.996 wl-01.acme.com Order 569281734 failed to customer 10098213 Exception follows: weblogic.jdbc.extensions.OrderIDSQLException weblogic.common.resourcepool.ResourceDeadException: Could not create pool connection. The DBMS driver exception was: [BEA]Oracle JDBC Driver Error establishing socket to host and port: ACMEDB-01:1521. Reason: Connection refused	Order ID	Customer ID
05/21 16:33:11.238 [CONNEVENT] Ext 1207130 (015-233) Event 20111, CTI Num:ServID:Type 0:199, App 0, ANI T7998#1, DNIS 5555685981, ServID 40489a07-7f6c-4251-801a-13ae51a6d092 CUSTID 10098213 Customer ID 05/21 16:37:49.732 [DISCEVENT] ServID 40489a07-7f6c-4251-801a-13ae51a6d092		
{actor: {displayName: "Go team!!", followersCount: 1366, friendsCount: 789, link: "http://dallascowboys.com/location", Customer's Twitter ID}, objectType: "place"}, objectType: "person", preferredUsername: "BoysFC@n80", statusesCount: 6072}, body: "Can't buy this device from @ACME. Site doesn't work. Called, gave up on waiting for them to answer! RT if you hate @ACME!", objectType: "activity", postedTime: "05-21T16:39:40.647-0600"	Customer's Twitter ID	Customer's Tweet
Company's Twitter ID		

Figure 3: By correlating different types of data together, you can start to gain real insight into what's going on in your infrastructure, see security threats or even use the insights to drive better business decisions.

By taking all the data involved in the process — i.e., pulling information from order processing, middleware, interactive voice response systems and Twitter — an organization can get a full view of the customer experience problem.

In this guide, we provide a high-level overview of the most common types of data in organizations of all sizes. While needs may differ from business to business based on vendors, products and infrastructure or mission, we've outlined the details that are key as you look for value in your data.

Most of the data sources and types in this guide can support multiple use cases — a major driver of value for data — so we've also coded the sections with icons and colors for easy reading. Enjoy reaping the benefits of Splunk whatever your need.



Security and Compliance



IT Ops, App Delivery and DevOps



Internet of Things



Business Analytics

Table of Contents

User Data.....	6
Virtual Private Networks (VPN).....	6
Authentication Data.....	7
Application Data.....	8
Antivirus	8
Application Performance Management (APM) Tool Data.....	9
Automation, Configuration, Deployment Tools (Platforms)	9
Binary Repositories.....	10
Build Systems (Platforms).....	10
Code Management.....	11
Container Logs and Metrics.....	11
Container Orchestration Metrics	12
CRM, ERP and Other Business Applications	13
Custom Application and Debug Logs	14
Distributed Tracing Tools.....	15
Mail Server	16
Test Coverage Tools.....	16
Serverless Monitoring.....	17
Vulnerability Scanning.....	18
Middleware Data	19
Application Server.....	19
Middleware.....	20
Mobile Device Data.....	20
Web Server	21
Network Data	22
Deep Packet Inspection Data.....	22
DHCP	23
DNS.....	23
Endpoint	24
Firewall.....	24
FTP.....	25
Intrusion Detection/Prevention	25
Load Balancer	26
Network Access Control (NAC).....	26
Network Protocols.....	27
Network Routers.....	28
Network Switches.....	28

Proxies.....	29
VoIP	29
SNMP	30
Operating System Data.....	31
System Logs.....	31
System Performance.....	32
Virtual Infrastructure Data	33
AWS Services.....	33
Google Cloud Platform (GCP).....	34
Microsoft Azure.....	34
Pivotal Cloud Foundry (PCF).....	35
VMware Server Logs, Configuration Data and Performance Metrics.....	36
Physical Infrastructure Data	37
Backup.....	37
Environmental Sensors.....	38
Industrial Control Systems (ICS).....	38
Mainframe	39
Medical Devices	39
Metric Line Protocols.....	40
Patch Logs.....	41
Physical Card Readers.....	41
Point-of-Sale Systems (POS).....	42
RFID/NFC/BLE.....	43
Sensor Data.....	44
Server Logs.....	45
Smart Meters.....	45
Storage.....	46
Telephony	46
Transportation.....	47
Wearables.....	47
Additional Data Sources.....	48
Database.....	48
Business Service Transaction and Business Service Performance Data.....	49
Human Resources	50
Social Media Feeds.....	50
Third-Party Lists	51





User Data

Virtual Private Networks (VPN)

Use Cases: Security and Compliance

Examples: Citrix NetScaler Nitro, Citrix NetScaler IPFIX, Cisco

Virtual private networks (VPNs) are a way of building a secure extension of a private network over an insecure, public one. VPNs can be established either between networks, routing all traffic between two sites, or between a client device and a network. Network-to-network VPNs typically are created using strong credentials such as certificates on each end of the connection. Client-to-network VPNs rely on user authentication, which can be as simple as a username and password. VPNs use network tunneling protocols such as IPSec, OpenVPN plus SSL or L2TP with cryptographically strong algorithms to scramble information in transit and ensure end-to-end data integrity.

Use Cases

Security and Compliance: VPN logs help analyze users coming onto the network. This information can be used in a number of ways, including situational awareness, monitoring foreign IP subnets, and compliance monitoring of browsers and applications of connected hosts. VPN data can also help identify:

- Activities from different locations, such as changes in location within a given amount of time.
- Access from risky countries or locations.
- User sessions at odd times, such as late evenings or weekends.
- User land speed violations.
- Abnormal frequency of sessions based on each user profile.

Authentication Data

Use Cases: Security and Compliance, IT Operations, Application Delivery

Examples and Data Sources: Active Directory, LDAP, Identity Management, Single-Sign On

Authentication data provides insight into users and identity activity. Common authentication data sources include:

- **Active Directory:** A distributed directory in which organizations define user and group identities, security policies and content controls.
- **LDAP:** An open standard defined by the Internet Engineering Task Force (IETF) and is typically used to provide user authentication (name and password). It has a flexible directory structure that can be used for a variety of information such as full name, phone numbers, email and physical addresses, organizational units, workgroup and manager.
- **Identity Management:** Identity management is the method of linking the users of digital resources — whether people, IoT devices, systems or applications — to a verifiable online ID.
- **Single Sign-On (SSO):** A process of using federated identity management to provide verifiable, attestable identities from a single source to multiple systems. SSO significantly increases security by tying user credentials to a single source, allowing changes to user rights and account status to be made once, and reflected in every application or service to which the user has access. SSO is particularly important for users with elevated security rights such as system or network administrators that have access to a large number of systems.

Use Cases

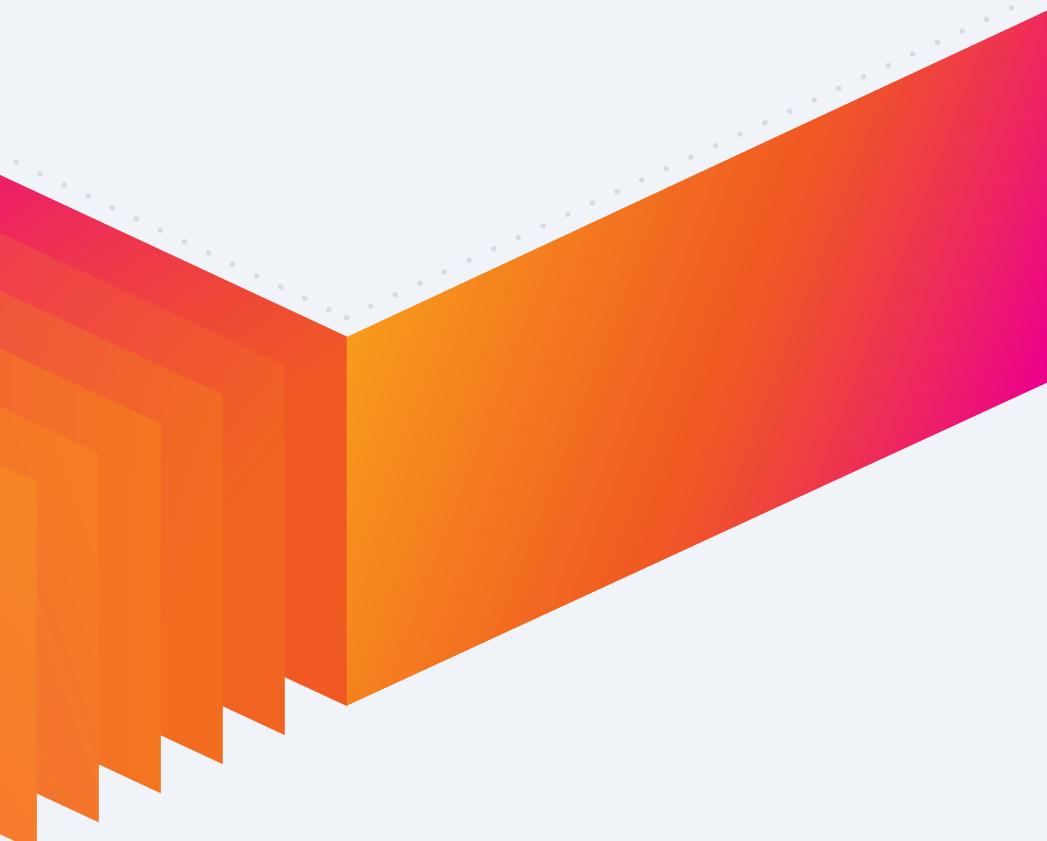
Security and Compliance: For security, authentication data provides a wealth of information about user activity, such as multiple login failures or successes to multiple hosts in a given time window, activities from different locations within a given amount of time, and brute force activities. Specifically:

- Active Directory domain controller logs contain information regarding user accounts, such as privileged account activity, as well as the details on remote access, new account creation and expired account activity.
- LDAP logs include a record of who, when and where users log in to a system and how information is accessed.
- Identity Management data shows access rights by user, group and job title (e.g., CEO, supervisor or regular user). This data can be used to identify access anomalies that could be potential threats — for example, the CEO accessing a low-level networking device or a network admin accessing the CEO's account.

IT Ops and Application Delivery: Authentication data supports IT operations teams as they troubleshoot issues related to authentication. For example, application support can be tied to logins, enabling IT operations to see whether users are struggling to log in to applications. For IT operations teams that support Active Directory, logs can be used to troubleshoot and understand the health of Active Directory.



Application Data



Antivirus

Use Cases: Security and Compliance

Examples: Kaspersky, McAfee, Norton Security, F-Secure, Avira, Panda, Trend Micro

The weakest link in corporate security are individuals, and antivirus is one way to protect them from performing inadvertently harmful actions. Whether it's clicking on an untrustworthy web link, downloading malicious software or opening a booby-trapped document (often one sent to them by an unsuspecting colleague), antivirus can often prevent, mitigate or reverse the damage.

So-called advanced persistent threats (APTs) often enter through a single compromised machine attached to a trusted network. While not perfect, antivirus software can recognize and thwart common attack methods before they can spread.

Use Cases

Security and Compliance: Antivirus logs support the analysis of malware and vulnerabilities of hosts, laptops and servers; and can be used to monitor for suspicious file paths. It can help identify:

- Newly detected binaries, file hash, files in the filesystem and registries.
- When binaries, hash, or registries match threat intelligence.
- Unpatched operating systems.
- Known malware signatures.

Application Performance Management (APM) Tool Data

Use Cases: Security and Compliance, IT Operations, Application Delivery

Examples: Dynatrace, New Relic, AppDynamics, MMSoft Pulseway, LogicMonitor, Stackify, Idera, Ipswitch

Application performance management software provides end-to-end measurement of complex, multitier applications to provide performance metrics from an end user's perspective. APM logs also provide event traces and diagnostic data that can assist developers in identifying performance bottlenecks or error conditions. The data from APM software provides both a baseline of typical application performance and record of anomalous behavior or performance degradation. Carefully monitoring APM logs can provide an early warning to application problems and allow IT and developers to remediate issues before users experience significant degradation or disruption. APM logs also are required to perform post-hoc forensic analysis of complex application problems that may involve subtle interactions between multiple machines, network devices or both.

Use Cases

Security and Compliance: Security teams can use APM logs to perform post-hoc forensic analysis of incidents that span multiple systems and exploit vulnerabilities. The data can be used to correlate security indications between the system and application activities. It also helps to identify SQL/API calls/CMD made in relation to suspicious activity, or abnormal amounts of sessions or CPU load in relation to security activity.

IT Ops and Application Delivery: By providing end-to-end measurement of complex, multi tier applications, APM logs can show infrastructure problems and bottlenecks that aren't visible when looking at each system individually, such as slow DNS resolution causing a complex web app to bog down as it tries to access content and modules on many different systems.

Automation, Configuration, Deployment Tools (Platforms)

Use Case: Application Delivery and DevOps

Examples: Puppet Enterprise, Ansible Tower, Chef, SaltStack, Rundeck machine data ingested through APIs, webhooks or run logs

Automated configuration and deployment tools, also known as infrastructure as code, allow IT and DevOps practitioners to practice continuous application delivery in the cloud or on premises. When infrastructure is treated as code, it's easy to share, collaborate, manage version control, perform peer unit testing, automate deployments, check the status of deployment and more.

Tools like Rundeck are platforms that take automation frameworks like Salt Stack and enable teams to automate states or playbooks to make sure the code is released and reported back to a central reporting tool.

Use Cases

Application Delivery and DevOps: Automation and configuration machine data monitoring helps application delivery teams deliver applications faster without sacrificing stability or security.

Binary Repositories

Use Case: Application Delivery and DevOps

Examples: Data from Nexus, Artifactory, delivered through APIs, webhooks; Yum, Pacman and Aptly data delivered through logs

A binary repository is a tool for downloading and storing binary files used and created in software development. It's used to store software binary packages, artifacts and their corresponding metadata. They're different from source code repositories, as binary repositories do not store source files. Searching through these repositories is possible by analyzing associated metadata.

Use Cases

Application Delivery and DevOps: Analyzing binary repository data helps application delivery teams and release managers to ensure that the final deployment of code to production is successful.

Build Systems (Platforms)

Use Case: Application Delivery and DevOps

Examples: Jenkins, Bamboo, TravisCI, TeamCity machine data ingested through APIs, logs, webhooks

Build platforms, like Jenkins and Bamboo, enable a continuous integration practice that allows application delivery teams — including developers, DevOps practitioners, QA and release engineering — to build artifacts, trigger new builds and environments, automate tests and more.

Use Cases

Application Delivery and DevOps: Build systems monitoring helps release managers, test and QA teams understand the health of their build environment, the status of tests, get insights into stack traces and build queues. This visibility helps remediate build or test bottlenecks and increase the application delivery velocity and quality.

Code Management

Use Case: Application Delivery

Examples: Github, GitLab

For all but the most trivial implementations, application source code consists of dozens if not hundreds of interrelated files. The complexity and volatility of code — particularly when using agile development methodologies and changes are made daily — makes keeping track of it virtually impossible without a structured, automated source code management and revision control system.

Originally built as client-server applications where developers checked in code to a central repository, today's systems (such as Git) are often distributed, with each developer working from a local copy of the full repository and changes synchronized across all subscribers to a particular project. Code management systems provide revision control (the ability to back out changes to an earlier version), software build automation, configuration status records and reporting, and the ability to branch or fork all or part of a source-code tree into a separate subproject with its own versioning.

Use Cases

Application Delivery: The version records of code management can help IT operations teams identify application changes that are causing system problems, such as excessive resource consumption or interference with other applications.

Container Logs and Metrics

Use Case: Application Delivery and DevOps

Examples: Docker

Container logs are an efficient way to acquire logs generated by applications running inside a container. By utilizing logging drivers, output that is usually logged is redirected to another target. Since logging drivers start and stop when containers start and stop, this is the most effective way of capturing machine data, given the often limited lifespan of a container.

Container metrics contain details related to CPU, memory, I/O and network metrics generated by a container. By capturing this data, you have the opportunity to spot specific containers that appear to consume more resources than others — enabling faster, more precise troubleshooting.

Use Cases

Application Delivery and DevOps: Acquiring container log files gives developers and operations teams insight on errors, issues and availability of applications running inside containers. Logs and metrics at the container level also call attention to containers whose performance is outside of expected parameters. As a result, admins can “kill” or “stop” a container instance, and “run” a new container in its place.

Container Orchestration Metrics

Use Case: Application Delivery and DevOps

Examples: Kubernetes, Amazon ECS2, Azure Container Services, Docker Swarm, Google Container Engine

Container orchestration tools provide an enterprise-level framework for automating container deployments and integrating and managing containerized applications at scale. Container orchestration tools like Kubernetes are important for ensuring the speed, availability, scaling and networking of containerized environments. Like container metrics, it's important to collect container orchestration metrics at high-resolutions due to their self-healing, ephemeral nature.

The most popular container orchestration platform is Kubernetes. Kubernetes metrics contain details related to the inventory, health and performance of container resources (cluster maps, node state, pod status, container status, namespace status, workload deployments details, etc.) along with aggregated system metrics (CPU, disk, memory, network) across nodes. By visualizing and correlating this data, you have the opportunity to keep track of infrastructure inventory, capacity, and cost and investigate underlying issues across your Kubernetes environment leading to failures — expediting troubleshooting.

Use Cases

Application Delivery and DevOps: Acquiring Kubernetes metrics gives developers and operations teams insights across all layers of their Kubernetes environment and the underlying infrastructure. This broad view helps operators monitor and manage the health of containerized environments, oversee services migrating to Kubernetes, and quickly diagnose any issues with the infrastructure, the orchestration platform itself, or the container.

For example, operators can look into an under-performing pod then to the metrics for the workload running in that pod and view its neighbors allowing for more context than just container level metrics and logs. Since particular problems in container environments can often be hard to find, this context is critical for teams to correlate patterns — reducing mean time to clue and expediting root cause analysis. This is particularly helpful during troubleshooting when DevOps teams need to quickly pinpoint which service is causing a sudden spike in latency or error rate and why. This comprehensive view also assists with resource optimization and capacity planning.

CRM, ERP and Other Business Applications

Use Cases: Security and Compliance, Application Delivery, Business Analytics

Examples: SAP, SFDC, SugarCRM, Oracle, Microsoft Dynamics

Business Applications can create a wealth of data as part of normal operations. Two examples are CRM and ERP applications:

Customer relationship management (CRM) systems have become an essential part of every organization, providing a central database of all customer contact information, communications and transaction details. CRM systems have evolved from simple contact management systems to platforms for customer support and engagement by providing personalized sales and support information. The same customer support data repository can be used to develop customized marketing messages and sales promotions. CRM systems are also useful for application support and enhancement by recording details about customer problems with a particular system or application along with their eventual solution — details that can inform future application or service updates.

Enterprise resource planning (ERP) applications are a critical back-office IT service that provides systematic, automated collection and analysis of a variety of product, supply chain and logistics data. ERP is used in product planning, tracking purchases of components and supplies, inventory management, monitoring and regulating manufacturing processes, managing logistics, warehouse inventory and shipping, and to monitor and measure the effectiveness of sales and marketing campaigns.

ERP software also integrates with CRM, HR, finance/accounting/payroll and asset management systems, with bidirectional data flows that provide consistent information across back-end digital business processes. ERP systems are typically built on a relational database management system with a variety of modules and customizations for specific functions such as supplier relationship management or supply chain management. Due to their complexity, ERP systems often are installed and managed by product specialists.

Use Cases

Security and Compliance: CRM records can help security teams unravel incidents that involve multiple customers and problem episodes over a long time span. They can also provide evidence of a breach, should records be modified outside normal business processes. In addition, the data can be used to audit access records of customer or internal user information.

Application Delivery: CRM, ERP, and other business applications are often mission-critical systems that facilitate a variety of front and back office processes. The performance of these applications can impact internal operations. Business application logs can be used to determine the health of those operations.

Business Analytics: CRM, ERP, and other business applications facilitate a variety of front and back office processes that span other systems as well. As part of an end-to-end view of those complex business processes, business application data can help provide insights into the health of business operations.

Custom Application and Debug Logs

Use Cases: Security and Compliance, IT Operations, Application Delivery

Examples: Custom applications

Best practices for application developers require the inclusion of debugging code in applications that can be enabled to provide minute details of application state, variables and error conditions or exceptions. Debug output is typically logged for later analysis that can expose the cause of application crashes, memory leaks, performance degradation and security holes. Furthermore, since the events causing a security or performance problem may be spaced over time, logs — along with the problem software — can help correlate and trace temporally separated errors to show how they contribute to a larger problem.

Application debug logs provide a record of program behavior that is necessary to identify and fix software defects, security vulnerabilities or performance bottlenecks. While test logs record the output results of application usage, debug logs provide information about an application's internal state, including the contents of variables, memory buffers and registers; a detailed record of API calls; and even a step-by-step trace through a particular module or subroutine. Due to the performance overhead and amount of data produced, debug logs typically are enabled only when a problem can't be identified via test or event logs.

Use Cases

Security and Compliance: Security breaches are often the result of improper handling of unexpected inputs, such as buffer overflow exploits or data injection used in cross-site scripting attacks. This type of low-level vulnerability is almost impossible to detect without logging the internal state of various application variables and buffers.

Similar to APM logs, custom application and debug logs can be used to correlate security indications between the system and application activities. It also helps to identify SQL/API calls/CMD made in relation to suspicious activity, or abnormal amounts of sessions or CPU load in relation to security activity.

IT Ops and Application Delivery: Debug output can expose application behavior that causes inefficient use of system resources or application failures that can be addressed by developers and operations teams. Debug output is useful for unraveling the internal state of an application that exhibits performance problems or has been shown to have security vulnerabilities, and the data can be helpful in identifying root cause.

Distributed Tracing Tools

Use Case: IT Operations, Application Delivery and DevOps

Examples: SignalFx, OpenTelemetry, Zipkin, Jaeger, fluentd

Distributed tracing is a method used to monitor how requests flow through your microservices applications by mapping transaction paths and duration as they propagate across services through trace and span data.

Popular open source distributed tracing instrumentation tools like OpenTelemetry record and publish operation data useful for finding sources of latency and errors within a distributed system — illuminating the relationship between user-visible behavior and the complex mechanics of the microservices underneath. APM software tools metricize information collected through these instrumentation tools to provide actionable insights on performance problems drilling down into specific service-level details.

Traces contain a lot of information about the method, operation, or block of code that it captures like the operation name, the start time of the operation, how long the operation took to execute, the logical name of the service on which the operation took place, the IP address of the service instance on which the operation took place, and trace context propagation. These are often represented as RED (Rate, Errors, Duration) metrics for monitoring purposes.

Distributed tracing along with APM Tools provide a context rich, complete view of service transactions that exist in complex distributed systems so IT and developers can understand user-visible latency, SLAs, and perform root-cause analysis with preserved traces that serve as anomaly benchmarks.

Use Cases

IT Operations, Application Delivery and DevOps: By providing end-to-end measurement of complex, multi-tier applications, tracing data can show microservices problems and bottlenecks that aren't visible when looking at each application individually, especially through service mapping, such as slow DNS resolution causing a complex web app to bog down as it tries to access content and modules on many different systems.

Distributed tracing allows DevOps teams to see all traces and spans for an API call and fix underperforming APIs. This helps teams improve system performance in real-time, before downstream effects impact customers. APM tools can expose which transaction spans deviated from the norm while showing correlation to code and infrastructure for deeper root cause analysis and troubleshooting. Since teams can visualize tracing data in real-time, this information improves time to market by making it easy to immediately see how updates and rollouts to services impact applications.

Mail Server

Use Cases: Security and Compliance, IT Operations

Examples: Exchange, Office 365

Email remains the primary form of formal communication in most organizations. As such, mail server databases and logs are some of the most important business records. Due to their size and tendency to grow without bounds, email data management typically requires both data retention and archival policies so that only important records are held and inactive data is moved to low-cost storage.

Use Cases

Security and Compliance: Mail server data can help identify malicious attachments, malicious domain links and redirects, emails from known malicious domains, and emails from unknown domains. It can also be used to identify emails with abnormal or excessive message sizes, and abnormal email activities times.

IT Ops: Email messages and activity logs can be required to maintain compliance with an organization's information security, retention and regulatory compliance processes. Mail server transaction and error logs also are essential debugging tools for IT problem resolution and also may be used for usage-based billing.

Test Coverage Tools

Use Case: Application Delivery and DevOps

Examples: Static Analysis and Unit Testing logs (SonarQube, Tox, PyTest, RubyGem MiniTest, Bacon, Go Testing), build server logs and performance metrics

Typical test coverage includes functional, statement, branch and conditional coverage. The idea is to match what percentage of code can be exercised by a test suite of one or more coverage criteria. Coverage tests are usually defined by rule or requirements. In addition to coverage testing, software delivery teams can utilize machine data to understand the line count, code density and technical debt.

Use Cases

Application Delivery and DevOps: Test coverage data monitoring helps release managers, application owners and others understand:

- How much technical debt and issues are they resolving?
- How ready is their next release?
- From unit testing — how many tests were performed per hour and what tests are being run?

If test coverage data is combined with build data, release managers can start monitoring build and release performance and start understanding the release quality. They can understand the trends in error percentage and make decisions on if the build is ready for production. Understanding code quality can also help support teams get prepared for any additional volume of calls or any particular issues that may arise.



Serverless Monitoring

Use Case: Application Delivery and DevOps

Examples: AWS Lambda, Google Cloud Functions (GCF), Azure Functions, OpenShift Serverless

Event-driven, serverless computing platforms also known as functions-as-a-service (FaaS) allow IT and DevOps practitioners to practice continuous application delivery without the need to perform administrative tasks required to provision and manage infrastructure. With FaaS, developers write single-purpose functions that are triggered and scaled on demand by events emitted from services so teams can focus on writing and delivering business critical applications. It makes it easy to automate processes, control costs, autoscale services and APIs, and promote collaboration across teams writing specialized applications in different languages. However, the “statelessness” and ephemerality of functions make monitoring their performance almost impossible without real-time, contextual solutions.

Use Cases

Application Delivery and DevOps: Serverless monitoring helps DevOps teams, application owners and others understand:

- Availability of applications running on serverless with point in time information about current state of functions like average latency and total number of function cold starts.
- Usage on concurrency for availability and cost planning. Teams can increase the amount of concurrency during times of high demand and lower it, or completely turn it off, when demand decreases in real-time.
- Errors with visibility and insights into failed invocations so developers can remediate issues before users are impacted.
- Compute duration — time from when your function code starts executing as the result of an invocation to when it stops executing for deeper understanding into costs.
- How functions are supporting business and customer experience including user requests, checkout abandonment, revenue per location, etc.
- Trends and breakdowns of functions by account, region, etc. for deeper root cause analysis.

Data from functions can also be monitored via distributed tracing for granular visibility into the performance of serverless applications along with end-to-end transaction views into invocations of multiple functions and all services.



Vulnerability Scanning

Use Case: Security and Compliance

Examples: ncircle IP360, Nessus

An effective way to find security holes is to examine infrastructure from the attacker's point of view. Vulnerability scans probe an organization's network for known software defects that provide entry points for external agents. These scans yield data about open ports and IP addresses that can be used by malicious agents to gain entry to a particular system or entire network.

Systems often keep network services running by default, even when they aren't required for a particular server. These running, unmonitored services are a common means of external attack, as they may not be patched with the latest OS security updates. Broadscale vulnerability scans can reveal security holes that could be leveraged to access an entire enterprise network.

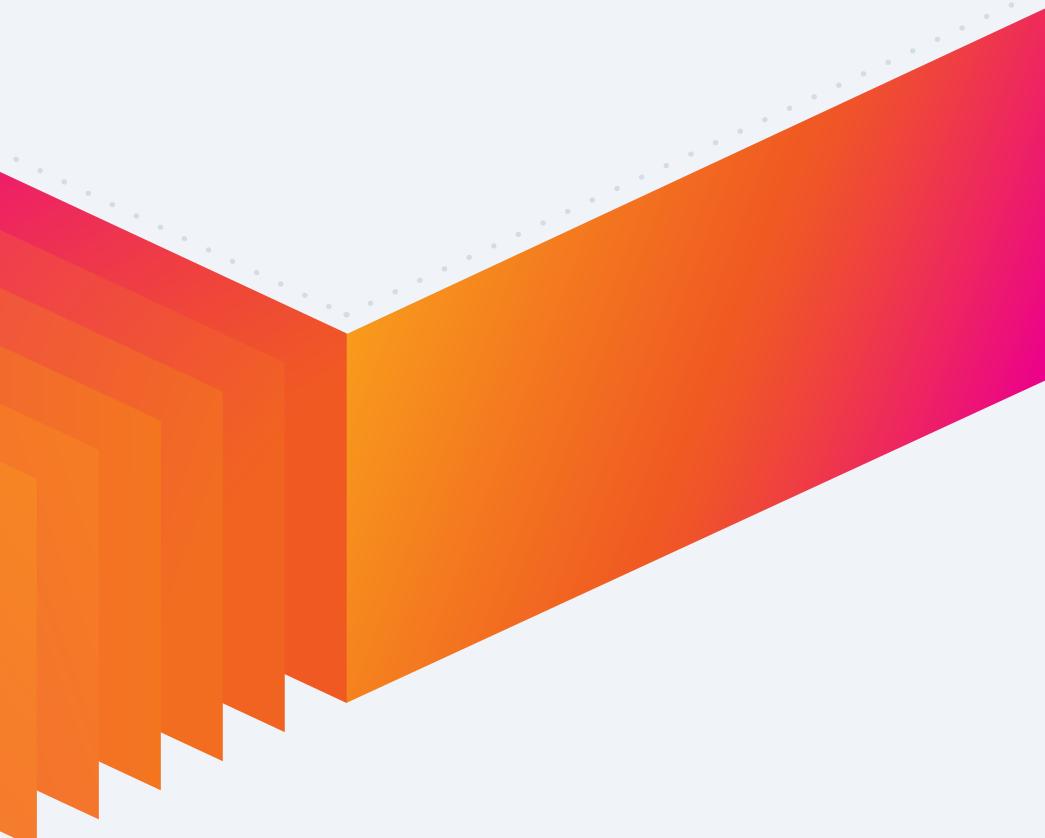
Use Cases

Security and Compliance: Vulnerability scans yield data about open ports and IP addresses that can be used by malicious agents to gain entry to a particular system or entire network. The data can be used to identify:

- System misconfiguration causing security vulnerability.
- Outdated patches.
- Unnecessary network service ports.
- Misconfigured filesystems, users or applications.
- Changes in system configuration.
- Changes in various user, app or filesystem permissions.



Middleware Data



Application Server

Use Cases: Security and Compliance, IT Operations, Application Delivery

Examples: log4j, log4php

Whether building a multi-tier web application or using a traditional client-server design, application servers run the backend software that handles user requests. Today, these are typically deployed as virtual machines on a multi-tenant hypervisor.

Use Cases

Security and Compliance: Security breaches are often the result of improper handling of unexpected inputs, such as buffer overflow exploits or data injection used in cross-site scripting attacks. This type of low-level vulnerability is almost impossible to detect without logging the internal state of various application variables and buffers. Since the events causing a security or performance problem may be spaced over time, logs, along with the problem software, can help correlate and trace temporally separated errors to show how they contribute to a larger problem. Anomalies in the logs can indicate potential failures or compromised attempts. The data can also help:

- Monitor user or customer transactions.
- Identify abnormal volume/amount/session of transactions.
- Identify unknown user interaction with third accounts, users or both.
- Sequence the exact transaction patterns matching fraudulent profiles.

IT Ops and Application Delivery: The value of application server logs depends on what they collect; however, these may include customer information useful in troubleshooting or application state transitions similar to, but less verbose than debug output that can provide clues to application crashes, memory leaks and performance problems.





Middleware

Use Cases: Security and Compliance, IT Operations, Application Delivery

Examples: Tibco, Software AG, Apache Active MQ, Kafka, AMQP, MQTT

Middleware describes a software layer of the prototypical three-tier enterprise application that typically implements data transformations, analysis and business logic. Middleware accesses databases for persistent storage and relies on web apps for the user interface. Middleware is often developed on the J2EE platform.

Use Cases

Security and Compliance: Since middleware generally accesses network services and sensitive databases, security teams can use log data to vet application integrity, identify suspicious behavior and specific vulnerabilities. It can also be used for user and customer transaction monitoring and to identify abnormal transactions, unknown user interaction with third party accounts, and the sequence of exact transaction patterns that match known fraudulent profiles.

IT Ops and Application Delivery: Middleware data can help operations teams diagnose problems with three-tier applications that involve the interaction between web, middleware and database servers.



Mobile Device Data

Use Cases: Security and Compliance, IT Operations, Application Delivery

Given the array of always-active sensors on mobile devices, they are veritable gushers of data that can include:

- Physical parameters such as location, network MAC ID, device GUID, device type and OS version.
- Network settings such as address, AP or cell-base station location, link performance.
- Application-specific telemetry such as time in app, features used and internal state and debug parameters similar to those provided by conventional application servers.

Use Cases

Security and Compliance: Security teams can expand the threat landscape by monitoring mobile device data for abnormal activity in regards to authentication, location and application usage.

IT Ops and Application Delivery: Since mobile apps invariably connect to one or more backend services, data from the client's point of view can provide insight into the app's condition and state when investigating issues such as crashes, performance degradation or security leaks. Mobile data shows the sequence of events and the application conditions leading up to and during a problem. If the source of the problem is the mobile application itself, getting insight on mobile application data can help developers deliver a better performing mobile app.

Web Server

Use Cases: Security and Compliance, IT Operations, Application Delivery

Examples: Java J2EE, Apache, Application Usage Logs, IIS logs, nginx

Web servers are the backend application behind every website that delivers all content seen by browser clients. Web servers access static HTML pages and run application scripts in a variety of languages that generate dynamic content and call other applications such as middleware.

Web servers can vary widely, and can include:

- **Java – J2EE:** Java is [the most popular programming language](#) due to its versatility, relative ease of use and rich ecosystem of developer tools. Via the J2EE platform, which includes APIs, protocols, SDKs and object modules, Java is widely used for enterprise apps including web applets, middle-tier business logic and graphic front ends. Java is also used for native Android mobile apps.
- **Apache:** Apache is one of the oldest and most-used web servers on the internet, powering millions of enterprise, government and public sites. Apache keeps detailed records of every transaction: every time a browser requests a web page, Apache log details include items such as the time, remote IP address, browser type and page requested. Apache also logs various error conditions such as a request for a missing file, attempts to access a file without appropriate permissions or problems with an Apache plug-in module. Apache logs are critical in debugging both web application and server problems, but are also used to generate traffic statistics, track user behavior and flag security attacks such as attempted unauthorized entry or DDoS.
- **Application Usage Logs:** Like Apache web logs, collecting application usage can provide valuable information to multiple stakeholders including developers, IT, sales and marketing. Depending on how granular the measurement, usage tracking can assist developers in identifying application features that are most

and seldom used, those that users have trouble with and areas for future enhancement. For customer-facing applications, usage logs provide sales and marketing teams insight into the effectiveness of online and app-based sales channels and promotions, data about sell-through and transaction abandonment, and information for potential cross-sales promotions.

Use Cases

Security and Compliance: Web logs record error conditions such as a request to access a file without appropriate permissions and also track user activity that can flag security attacks such as attempted unauthorized entry or DDoS. It can also help to identify SQL injections and support correlating fraudulent transactions.

- Since Java apps frequently access network services and sensitive databases, security teams can use log data to vet the integrity of J2EE apps, identify suspicious application behavior and application vulnerabilities.
- Apache web logs can alert to security attacks such as attempted unauthorized entry, XSS, buffer overflows or DDoS.
- Like web logs, generic application usage logs can alert security teams to unauthorized access such as someone consuming more resources than normal, or using applications at odd hours.

IT Ops and Application Delivery: Web logs are critical in debugging both web application and server problems, but also are used to generate traffic statistics that are useful in capacity planning. Web server data can provide varying information for IT operations teams:

- J2EE data can help operations teams diagnose problems with three-tier applications that involve the interaction between web, middleware and database servers.
- In aggregate, Apache web logs can show activity of a web service. Drilling into details can reveal infrastructure bottlenecks and indicate downstream issues.
- Application usage logs can help IT operations teams with infrastructure capacity planning, optimization, load balancing and usage-based billing by providing detailed records of resource consumption.



Network Data

Deep Packet Inspection Data

Use Cases: Security and Compliance, IT Operations

Examples: Stream, PCAP, bro

Deep Packet Inspection Data (DPI) is a fundamental technique used by firewalls to inspect headers and the payload of network packets before passing them down the network subject to security rules. DPI provides information about the source and destination of the packet, the protocol, other IP and TCP/UDP header information and the actual data.

Use Cases

Security and Compliance: Packet Capture logs (PCAP) see everything traversing a network and are required to identify security attacks and incidents such as advanced persistent threats, data exfiltration, DDoS and malware. DPI also can be used to filter content subject to an organization's terms of service. PCAP data can also be used to provide and identify:

- DNS session analysis for malicious domain communications from each endpoint.
- Abnormal amounts of traffic or sessions.
- Abnormal amounts of domain and host communications.
- Known malicious traffic from a host.
- Expired SSL certification analysis.
- Abnormal host communications (internal and external).

IT Ops: Data on the network wire is authoritative and difficult to spoof (although encryption, steganography and advanced deception techniques can evade DPI). For example, DPI provides raw information of everything transmitted over a network, including things that aren't necessarily part of or difficult to extract from a log, such as database query results.

DHCP

Use Cases: Security and Compliance, IT Operations

Examples: DHCP Insight, Linux DHCP

DHCP is the network protocol most client devices use to associate themselves with an IP network. Implemented via a DHCP server, which could be standalone or embedded in a router or other network appliance, DHCP provides network clients with critical network parameters including IP address, subnet mask, network gateway, DNS servers, WINS or other name servers, time servers (NTP), a host and domain name and the address of other optional network services.

Use Cases

Security and Compliance: DHCP logs show exactly which systems are connecting to a network, their IP and MAC addresses, when they connect and for how long. This information is useful in establishing the state of a network when a security incident occurs and tracing an attacker's address back to a time of access and type of device by looking at the MAC ID and vendor identification string. The data can also be used to support user network access verification.

IT Ops: DHCP logs can be used when troubleshooting a client device that is having network problems, since it provides a definitive record of the device's primary IP parameters. The data may show that the DHCP server itself is at fault; for example, by not properly vending addresses, renewing IP leases or giving the same address to two separate devices.

DNS

Use Cases: Security and Compliance, IT Operations

Examples: BIND, PowerDNS, Unbound, DnsMasq, Erl-DNS

The domain name system (DNS) is the internet's phone book, providing a mapping between system or network resource names and IP addresses. DNS has a hierarchical name space that typically includes three levels: a top-level domain (TLD) such as .com, .edu or .gov; a second-level domain such as "google" or "whitehouse;" and a system level such as "www" or "mail." DNS nameservers operate in this hierarchy either by acting as authoritative sources for particular domains, such as a company or government agency, or by acting as caching servers that store DNS query results for subsequent lookup by users in a specific location or organization; for example, a broadband provider caching addresses for its customers.

Use Cases

Security and Compliance: Security teams can use DNS logs to investigate client address requests such as correlating lookups with other activity, whether requests are made for inappropriate or otherwise suspicious sites and relative popularity of individual sites or domains. Since DNS servers are a frequent target of DDoS attacks, logs can reveal an unusually high number of requests from external sources. Likewise, since compromised DNS servers themselves are often used to initiate DDoS attacks against other sites, DNS logs can reveal whether an organization's servers have been compromised. DNS data can also provide detection of unknown domains, malicious domains and temporary domains.

IT Ops: DNS server logs provide operations teams with a record of traffic, the type of queries, how many are locally resolved either from an authoritative server or out of cache, and a picture of overall system health.

Endpoint

Use Case: Security and Compliance

Examples: McAfee ePO, Symantec SEP

Endpoint security is used to protect corporate networks from inadvertent attacks by compromised devices using untrusted remote networks such as hotspots. By installing clients on laptops or other wireless and mobile devices, endpoint security software can monitor activity and provide security teams with warnings of devices attempting to spread malware or pose other threats.

In this context, endpoint refers to the security client software or agent installed on a client device that logs security-related activity from the client OS, login, logout, shutdown events and various applications such as the browser (Explorer, Edge), mail client (Outlook) and Office applications. Endpoints also log their configuration and various security parameters (certificates, local anti-malware signatures, etc.), all of which is useful in post-hoc forensic security incident analysis.

Use Cases

Security and Compliance: Endpoint data can be used for a variety of security uses, including identifying newly detected binaries, file hash, files in the filesystem, and registries. It can also help with identifying binary and hash registries that match threat intelligence, as well as unpatched operating systems and binaries, and to detect known malware.

Firewall

Use Cases: Security and Compliance, IT Operations

Examples: Palo Alto, Cisco, Check Point

Firewalls demarcate zones of different security policies. By controlling the flow of network traffic, firewalls act as gatekeepers collecting valuable data that might not be captured in other locations due to the firewall's unique position as the gatekeeper to network traffic. Firewalls also execute security policy and thus may break applications using unusual or unauthorized network protocols.

Use Cases

Security and Compliance: Firewall logs provide a detailed record of traffic between network segments, including source and destination IP addresses, ports and protocols, all of which are critical when investigating security incidents. The data may also reveal gaps in security policy that can be closed with tighter construction of firewall rules. Firewall data can help identify and detect:

- Lateral movement
- Command and Control traffic
- DDoS traffic
- Malicious domain traffic
- Unknown domain traffic
- Unknown locations traffic

IT Ops: When network applications are having communication problems, network security policies may be the culprit. Firewall data can provide visibility into which traffic is blocked and which traffic has passed through — helping identify if you have an app or network issue.



FTP

Use Cases: Security and Compliance, IT Operations

Examples: OSSEC, Getwatchlist, UTBox, Security Onion, iSeries - AS400, Traffic Ray

FTP is one of the oldest and most rudimentary network protocols for copying data from one system to another. Before websites and HTTP, FTP was the best way to move large files across the internet. FTP is still used in organizations that need reliable, deterministic internet file transfer.

Use Cases

Security and Compliance: Analyzing FTP servers can help security teams identify when compromised credentials are used, when abnormal traffic is coming from different locations or at odd times, and when sensitive files and documents are being accessed.

IT Ops: FTP traffic logs record the key elements of a file transmission, including source (client) name and address and remote user name if the destination is password-protected. This and other data are crucial when troubleshooting FTP problems, regardless of the application.

Intrusion Detection/ Prevention

Use Case: Security and Compliance

Examples: Tipping Point, Juniper IDP, Netscreen Firewall, Juniper NSM IDP, Juniper NSM, Snort, McAfee IDS

IDS and IPS are complementary, parallel security systems that supplement firewalls — IDS by exposing successful network and server attacks that penetrate a firewall, and IPS by providing more advanced defenses against sophisticated attacks. IDS is typically placed at the network edge, just inside a perimeter firewall, although some organizations also put a system outside the firewall to provide greater intelligence about all attacks. Likewise, IPS is typically placed at the network perimeter, although it also may be used in layers at other points inside the network or on individual servers. IPS usually works by dropping packets, resetting network connections and blacklisting specific IP addresses or ranges.

Use Cases

Security and Compliance: IDS logs provide security teams detailed records of attacks including the type, source, destination and port(s) used that provide an overall attack signature. Special signatures may trigger alarms or other mitigating actions. IPS provide the same set of attack signature data, but also may include a threat analysis of bad network packets and detection of lateral movement. This data can also detect command and control traffic, DDoS traffic, and malicious or unknown domain traffic.



Load Balancer

Use Cases: IT Operations

Examples: Local Traffic Manager, Cisco Load Balancer, Citrix, Kemp Technologies, Radware AppDirector OnDemand

Load balancers allocate external network traffic bound for a particular server or application across multiple redundant instances. There are two categories of load balancer: local, in which all resources in a load-balanced pool are on the same subnet; and global or distributed, where the resource pool is spread across multiple sites. Load balancers use several user-selectable algorithms to allocate traffic including:

- Round robin (systems get an equal number of connections allocated sequentially).
- Weighted round robin (where the load is assigned according to the percentage weight assigned each system in a pool).
- Least connections (where new connections go to the system with the fewest number of existing clients).
- Weighted least connections (where the connection handling capacity of each system is taken into account when determining the least busy system for new connections).
- Random (connections are randomly assigned to each member of a pool).

Use Cases

IT Ops: Load balancer logs provide operations teams with a record of overall traffic to systems or particular applications and provide indicators of each system's traffic-handling capacity and health, along with the status and health of the load balancer itself.

Network Access Control (NAC)

Use Case: Security and Compliance

Examples: Aruba ClearPass, Cisco ACS

Network access or admission control is a form of client/endpoint security that uses a locally installed software agent to pre-authorize connections to a protected network. NAC screens client devices for contamination by known malware and adherence to security policies such as running an approved OS with the most recent patches. Clients failing NAC screens are rerouted to an isolated quarantine network until any detected problems are corrected.

Use Cases

Security and Compliance: NAC software collects data about the connecting clients such as an inventory of installed client software, compliance with security policies, OS and application patch versions, accessibility by remote access clients and user access to protected networks. NAC logs provide security teams with a detailed profile of a client's state and activity. It can provide details into unauthorized device connections and be used to correlate users/IP to a physical network location.



Network Protocols

Use Cases: Security and Compliance, IT Operations

Examples: HTTP, Cisco NetFlow, Ntop, Flow-tools, FlowScan, EHNT, BPFT

Network protocols describe the structure of data that flows through networks. In most cases, network ports are assigned to specific protocols for both security and performance reasons. Some protocols operate at a lower level of the computing stack and are used to direct packet routing, such as TCP, UDP or IP. Other protocols, such as HTTP, HTTPS and TNS describe how packets are structured for applications — such as web services, databases and a wide range of client-based applications. By capturing, decrypting and analyzing network protocol data, you can better understand the kinds of applications, their usage, performance and even payload (content of the data) of applications. Since this data can be gathered directly from a network tap, or with specialized software, it provides a perspective on applications and how they interoperate that may not be otherwise available.

Use Cases

Security and Compliance: Network protocols are an important source for identifying advanced persistent threats, analyzing traffic flows for unusual activity and identifying potential data exfiltration. Aggregating and analyzing flow records also can show anomalous traffic patterns and flow destinations that are indicative of a breach, such as an APT phoning home to a command and control server for instructions, additional malware code, or copying large amounts of data to an attacker's system. The data can also be used to detect traffic related to DDoS, malicious domains, and unknown domains or locations.

IT Ops: Network protocol traffic analysis can help determine the network's role in overall availability and performance of critical services. Application traffic can be monitored for usage, performance, availability and can provide visibility into specific user data. For applications that cannot be instrumented on the servers, network traffic may be the only way to acquire performance data.

Network Routers

Use Cases: Security and Compliance, IT Operations

Examples: Routers from Cisco, Juniper, Linksys, Arista, Extreme Networks, Avaya

If switches are network intersections, then routers are the signal lights and traffic cops — the devices responsible for ensuring that traffic goes to the right network segment. Unlike switches that operate at Layer 2, routers work at Layer 3, directing traffic based on TCP/IP address and protocol (port number). Routers are responsible for particular Layer 3 address spaces and manage traffic using information in routing tables and configured policies. Routers exchange information and update their forwarding tables using dynamic routing protocols.

Use Cases

Security and Compliance: Routers collect the same sort of traffic logs and statistics as switches; thus, their data is equally valuable to security teams as a source for flagging advanced persistent threats, analyzing traffic flows for unusual activity and identifying potential data exfiltration. As a wire-level data source, router statistics are almost impossible to spoof and thus a critical source of security data. Router data can also be used to detect configuration changes, and error or failure alerts correlating with security indicators.

IT Ops: Network engineers use router logs and statistics to monitor traffic flow and ensure that traffic is being correctly forwarded between network segments. Data from routing protocol updates can show whether your routers are appropriately exchanging route tables with other locations, that external traffic can reach you, and that internal traffic is correctly forwarded to external routers.

Network Switches

Use Cases: Security and Compliance, IT Operations

Examples: Ethernet Switch, Virtual Switches

Switches are network intersections, places where packets move from one network segment to another. In their purest form, switches work within a particular IP subnet and can't route Layer 3 packets to another network. Modern data center designs typically use a two-tier switch hierarchy: top-of-rack (ToR) switches connecting servers and storage arrays at the edge and aggregation or spine switches connecting to the network core. Although ethernet switches are far more widespread, some organizations also use fiber channels or infiniband for storage area networks or HPC interconnects, each of which has its own type of switch.

Use Cases

Security and Compliance: Switch data, often captured as NetFlow records, is a critical data source for flagging advanced persistent threats, analyzing traffic flows for unusual activity and identifying potential data exfiltration. As a wire-level data source, switch statistics are almost impossible to spoof and thus a crucial source of security data. This data can also be used to correlate users or IP addresses to a physical network location.

IT Ops: Operations teams use switch logs to see the state of traffic flow, such as source and destination, class of service and causes of congestion. Logs also can show traffic statistics in the aggregate, by port and by client, and whether particular ports are congested, failing or down.

Proxies

Use Cases: Security and Compliance, IT Operations

Examples: Blue Coat, Fortinet, Juniper IDP, Netscreen Firewall, Palo Alto Networks, Palo Alto Networks config, Palo Alto Networks system, Palo Alto Networks threat, Palo Alto Networks traffic, nginx

Network proxies are used in several ways in IT infrastructure: as web application accelerators and intelligent traffic direction, application-level firewalls, and content filters. By acting as a transparent ‘bump-in-the-wire’ intermediary, proxies see the entire Layer 7 network protocol stack, which allows them to implement application-specific traffic management and security policies.

Use Cases

Security and Compliance: Security teams are interested in proxies as application-layer firewalls. Here, proxy records can identify details about specific content traversing network control points including file names, types, source and destination, and metadata about the requesting client such as OS signature, application and username/ID (depending on the proxy implementation). The data can also be used to help detect command and control traffic, malicious domain traffic and unknown domain traffic.

Web proxies and some next generation firewalls may act in a transparent or explicit mode communicating with HTTP(S) servers on behalf of a client. Using a number of related technologies, the request and response can be inspected and permitted, or blocked, based on user role, site or resource category or attack indicator. Data logged in the events can potentially be used in detective correlation.

IT Ops: Operations teams often use proxies embedded in an application delivery controller (ADC), a more advanced, Layer 7-aware version of a load balancer. In this context, proxy logs can provide information about incoming requests and traffic distribution among available resources.

VoIP

Use Cases: Security and Compliance, IT Operations

Examples: Asterisk CDR, Asterisk event, Asterisk messages

Voice over IP protocol refers to several methods for transmitting real-time audio and video information over an IP-based data network. Unlike traditional phone systems using dedicated, point-to-point circuits, VoIP applications use packet-based networks to carry real-time audio streams that are interspersed with other ethernet data traffic. Since TCP packets may be delivered out of order due to data loss and retransmission, VoIP includes features to buffer and reassemble a stream. Similarly, VoIP packets are usually tagged with quality of service (QoS) headers to prioritize their delivery through the network.

Use Cases

Security and Compliance: VoIP deployments may expose organizations to potential security threats, and analyzing VoIP logs can help identify and prevent these exploits.

IT Ops: VoIP logs provide troubleshooting and usage data similar to that of other network applications. Details include source, destination, time and duration of calls, call quality metrics (e.g., packet loss, latency, audio fidelity/bit rate) and any error conditions. Integrating VoIP source/destination records with an employee database such as AD or LDAP and a DHCP database allows linking call records to actual people and IP addresses to physical locations; information that can assist in troubleshooting and billing.

SNMP

Use Cases: Security and Compliance, IT Operations

Examples: LogicMonitor, ManageEngine, Spiceworks, Ruckus Idera, Ipswitch

The simple network management protocol (SNMP) is one of the oldest, most flexible and broadly adopted IP protocols used for managing or monitoring networking devices, servers and virtual appliances. This includes network devices such as routers and switches, as well as non-networking equipment such as server hardware or disk arrays.

SNMP supports two different methods of obtaining data.

- **SNMP Traps** are essentially alerts, set to send an alert on a state change, critical threshold, hardware failure, and more. Traps are initiated by the SNMP device, and the trap is sent to an SNMP collector.
- **SNMP Polling** is an interactive query/response approach. Unlike traps, polling is initiated by the SNMP collector in the form of a request for certain, or all, SNMP data available on the SNMP device.

Although many now provide vendor-specific APIs for remote management and data collection, SNMP is still valuable in troubleshooting due to its ubiquity (nearly every device supports it) and inherently centralized design (a single instance of SNMP management software can collect data from every device on an internal network, even across route domains).

Use Cases

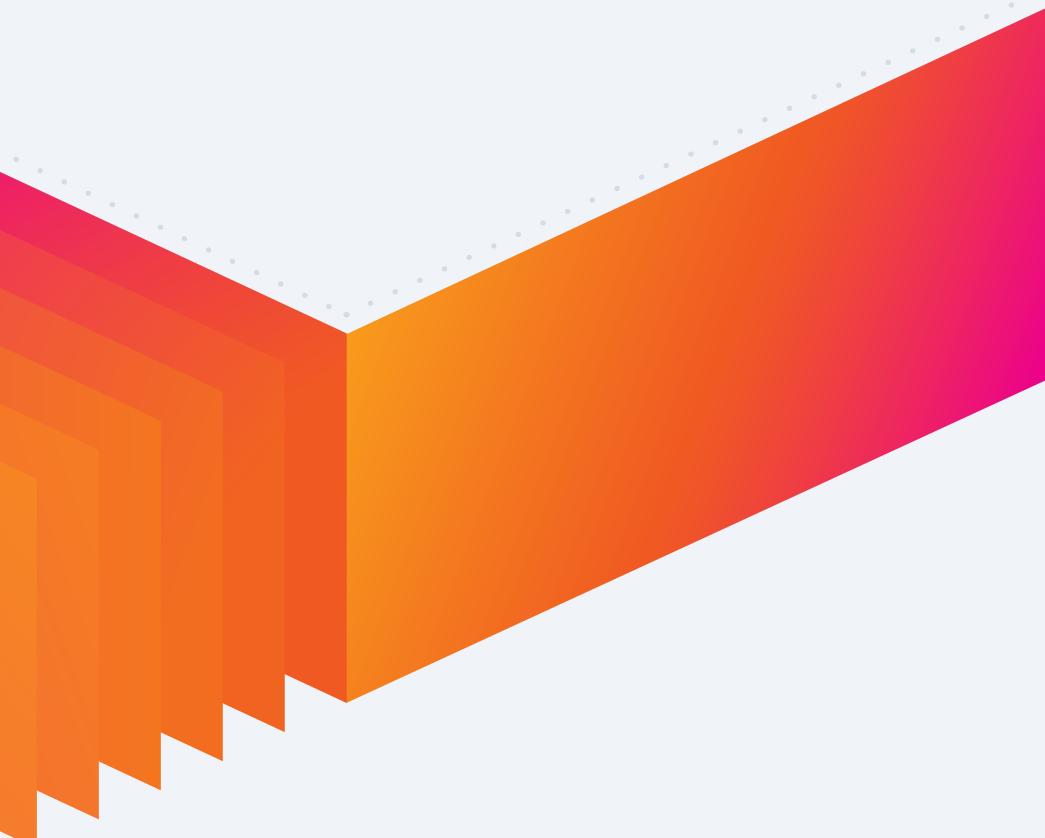
Security and Compliance: SNMP traps and alerts from network devices can help security teams identify abnormal activity over the network. SNMP Polling helps a security analyst to see the data transmission rates for a network-connected device that is suspected of malicious activity.

The data can also help identify abnormal amounts of traffic to a certain site or domain, an abnormal amount of specific SNMP traps from a certain host, and an abnormal number of unique SNMP traps from hosts compared to normal profiles.

IT Ops: SNMP data can provide current information about performance, configuration and current state. This allows the monitoring of the “normal” state of the environment, which is vital when using a service-level approach to monitoring the health of any environment. This could include current speed of all of the ports on a switch, the number of bytes sent (per port or in aggregate) through a router, the CPU temperature of a server, and any other information made available by the vendor per the SNMP MIBs for that device.

Many environments rely on SNMP traps for alerting when a critical state is reached (e.g., CPU temperature is critical) or when a failure occurs (e.g., RAID disk failure). SNMP traps are not only sent by devices to monitoring systems, in some environments SNMP traps are the de-facto method for multiple monitoring and alerting systems to aggregate errors to a single console.

Operating System Data



System Logs

Use Cases: Security and Compliance, IT Operations, Application Delivery

Examples: Unix, Windows, Mac OS, Linux

Every OS records details of its operating conditions and errors, and these time-stamped logs are the fundamental and authoritative source of system telemetry. Depending on the OS, there may be separate logs for different classes of events, such as routine informational updates, system errors, boot loader records, login attempts and debug output. Error logs often aggregate records from multiple subsystems and OS services or daemons, and, thus, are a definitive source of troubleshooting information.

Use Cases

Security and Compliance: System logs include a variety of security information such as attempted logins, file access and system firewall activity. These entries can alert security teams to network attacks, a security breach or compromised software. They also are an invaluable source of information in forensic analysis of a security incident. For example, the data can be used to identify changes in system configurations and commands executed by users or privileged users.

IT Ops and Application Delivery: System logs often are the first place operations teams turn when troubleshooting system problems, whether with the OS, hardware or various I/O interfaces. Since a particular problem often manifests itself with errors in multiple subsystems, correlating log entries is one of the best ways of identifying the root cause of a subtle system failure.

System Performance

Use Cases: Security and Compliance, IT Operations, Application Delivery

Examples: PERFMON, Windows Events Logs, sar, vmstat, iostat, statsd

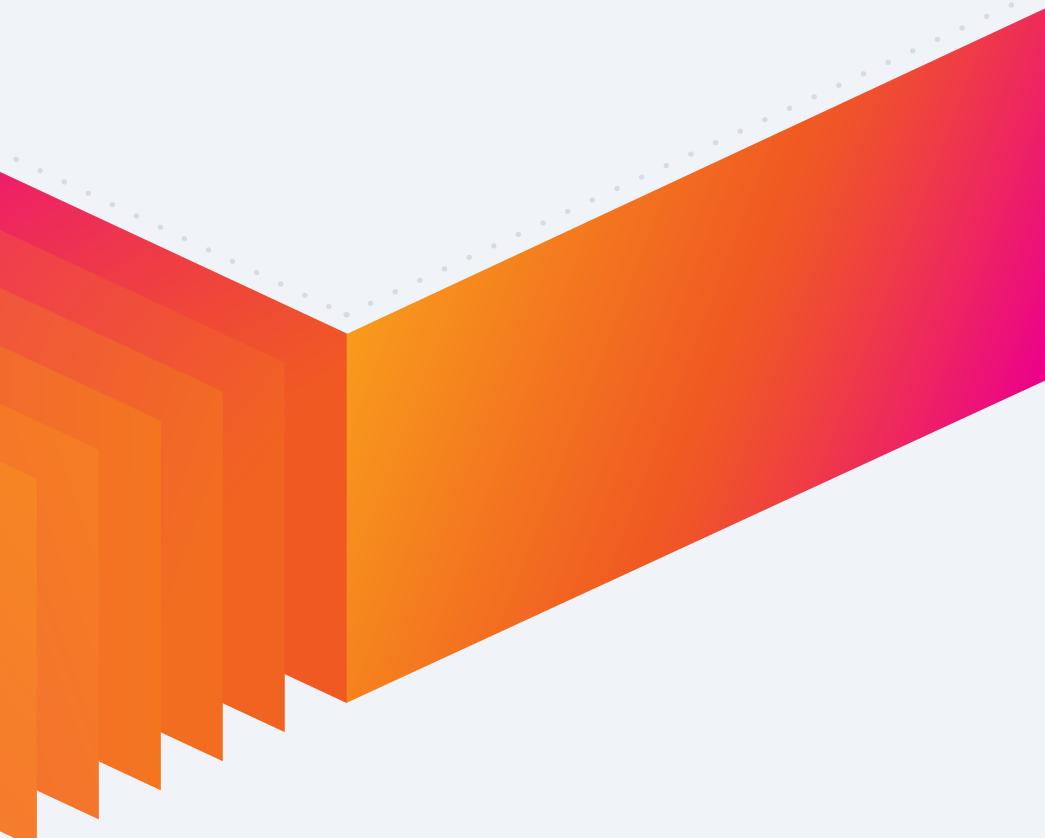
Measures of system activity such as CPU load, memory and disk usage, and I/O traffic are the IT equivalent of EKGs to a doctor: the vital signs that show system health. Recording these measures provides a record of system activity over time that shows normal, baseline levels and unusual events. By registering myriad system parameters, performance logs also can highlight mismatches between system capacity and application requirements, such as a database using all available system memory and frequently swapping to disk.

Use Cases

Security and Compliance: While primarily used for keeping infrastructure up and running, monitoring system performance can also be used to uncover potential security incidents by detecting abnormal activity in performance. One example is abnormal system resource usage in correlation with a security indication.

IT Ops and Application Delivery: Performance logs provide a real-time indication of system health by showing resource usage that, when compared with historical norms, flags performance problems. When measurements deviate from standard or typical parameters, it's a warning for IT admins to do further investigation.

Virtual Infrastructure Data



AWS Services

Use Cases: Security and Compliance, IT Operations

Examples: CloudTrail, CloudWatch, Config, S3

AWS is the largest and most widely used public cloud infrastructure, providing on-demand compute, storage, database, big data and application services with consumption-based pricing. AWS can be used to replace traditional enterprise virtual server infrastructure in which software runs on individual virtual machines (VM) or to host cloud-native applications built from a collection of AWS services. AWS includes a host of service management, automation, security, network and monitoring services used to deploy, scale, decommission, audit and administer one's AWS environment, subscriptions and hosted applications.

Use Cases

Security and Compliance: Security data from AWS services includes login and logout events and attempts, API calls and logs from network and web application firewalls.

IT Ops: AWS services provide similar types of system and service data as traditional IT infrastructure, much of which is consolidated by the CloudWatch service. These include service monitoring, alarms and dashboards for metrics, logs and events generated by other AWS resources and applications. Typical events and measures include when instances are instantiated and decommissioned, CPU usage, network traffic and storage consumption.



Google Cloud Platform (GCP)

Use Cases: Security and Compliance, IT Operations

Examples: Stackdriver

GCP is a popular and widely used public cloud infrastructure, providing on-demand compute, storage, database, big data and application services with consumption-based pricing. GCP can be used to replace traditional enterprise virtual server infrastructure in which software runs on individual VMs, or to host cloud-native applications built from a collection of GCP services. GCP includes a host of service management, automation, security, network and monitoring services used to deploy, scale, decommission, audit and administer one's GCP environment, subscriptions and hosted applications.

Use Cases

Security and Compliance: Security data from GCP services includes login and logout events and attempts, API calls and logs from network and web application firewalls.

IT Ops: GCP services provide similar types of system and service data as traditional IT infrastructure, much of which is consolidated by Stackdriver. These include service monitoring, alarms and dashboards for metrics, logs and events generated by other GCP resources and applications. Typical events and measures include when instances are instantiated and decommissioned, CPU usage, network traffic and storage consumption.

Microsoft Azure

Use Cases: Security and Compliance, IT Operations

Examples: WADLogs, WADEventLogs, WADPerformanceCounter, WADDiagnosticInfrastructure

Azure is a popular and widely used public cloud infrastructure, providing on-demand compute, storage, database, big data and application services with consumption-based pricing. Azure can be used to replace traditional enterprise virtual server infrastructure in which software runs on individual VMs, or to host cloud-native applications built from a collection of Azure services. Azure includes a host of service management, automation, security, network and monitoring services used to deploy, scale, decommission, audit and administer one's Azure environment, subscriptions and hosted applications.

Use Cases

Security and Compliance: Security teams can use Azure service logs to audit and attest to compliance with established policies. Log data also is invaluable for incident forensic analysis, such as identifying unauthorized access attempts from access logs, tracking resources and configuration change events and identifying vulnerabilities in hosts or firewalls.

IT Ops: Azure services provide detailed metrics and logs for monitoring one's infrastructure across the entire technology stack, VMs, containers, storage and application services. The data is useful in maintaining application delivery quality and service levels, measuring user behavior, resource utilization and for capacity planning and cost management.



Pivotal Cloud Foundry (PCF)

Use Cases: IT Operations and DevOps

Examples: Loggregator, PCF Healthwatch

Pivotal Cloud Foundry is a platform-as-a-service (PaaS) built on top of Cloud Foundry, an open source cloud computing platform that allows developers to easily deploy, operate and scale cloud-native applications. Enterprises can manage the entire application lifecycle, from packaging to deployment to execution, as Cloud Foundry supports many cloud frameworks and application languages. With PCF, the installation and administration of cloud-native applications is simplified with capabilities around infrastructure management and provisioning, OS patching, container orchestration, security and more.

Use Cases

IT Ops and DevOps: Operations teams can use PCF metrics, much of which is consolidated via the Loggregator Firehose to gain insights into deployment health, capacity needs and application health before end users are impacted by degraded performance. Since PCF allows DevOps to run their applications on any cloud rapidly and to scale on demand, PCF data is critical for teams to get the end-to-end visibility into the entire lifecycle and visibility between each individual component. When it comes to operating PCF deployments at scale, understanding performance relies on dependencies among the various layers within the app, container and larger architecture.



VMware Server Logs, Configuration Data and Performance Metrics

Use Cases: Security and Compliance, IT Operations

Examples: vCenter, ESXi

VMware vSphere ESXi is the most commonly used enterprise server virtualization platform. The VMware management platform, whether one of the vSphere products or standalone hypervisor, produce a variety of data and fall into four main categories:

- **vCenter Logs:** vCenter is the “control center” of a vSphere environment. The vCenter logs show information including: who is logging in to make changes, which individuals made changes and authentication failures.
- **ESXi Logs:** Every vSphere environment includes one or more ESXi hypervisors; these are the systems that host the virtual machines. ESXi logs contain information that is useful when troubleshooting hardware and configuration issues.
- **Inventory Information:** the vCenter environment tracks configuration about a number of configuration items including: hypervisors, virtual machines, datastores, clusters and more. This includes the configuration of each item, and how a given item relates to any other. This information is not represented in the log files from either the vCenter or ESXi servers. This information can be viewed using the vSphere client or by using vSphere APIs to pull this information. In both cases this information is pulled from the vCenter servers.

- **Performance Information:** for each configuration item, the vCenter server tracks a number of performance metrics about that item. Datastore latency, virtual or physical CPU utilization, and over 100 other metrics fall into this category. As with the inventory information, this information is not present in the log files and must be viewed through the vSphere client or polled through the vSphere API.

Use Cases

Security and Compliance: The uncoupled nature of virtual resources and underlying physical hardware can cause complex challenges during incident investigations, capacity analyses, change tracking and security reporting. One common security use case for VMware data comes from the vCenter logs, which audit the activity of individuals using the vSphere interface to re-assign user permissions within the VMware environment.

IT Ops: Operations teams can use VMware data to measure the health of the overall hypervisor environment and underlying guest operating systems. Admins can use this data for capacity planning, and for troubleshooting of ongoing performance issues, such as datastore latency issues.

This data also records hardware resource usage that can be used to optimize VM deployments across a server pool to maximize resource consumption without having workloads overwhelm any given server.



Physical Infrastructure Data



Backup

Use Case: IT Operations

Despite the use of data replication to mirror systems, databases and file stores, data backup remains an essential IT function by providing for long-term, archival storage of valuable information, much of which has legal and regulatory requirements regarding its preservation. Backups also can be used to store multiple versions of system images and data, allowing organizations to reverse changes, accidental deletions or corrupted data quickly, restoring the system or database to a known good state. Backup software can use different types of storage media depending on the likelihood of needing the data: external disks or virtual tape libraries for active data and tape, optical disks or a cloud service for long-term storage.

Use Cases

IT Ops: Backup systems routinely log activity and system conditions, recording information such as job history, error conditions, backup target and a detailed manifest of copied files or volumes. This data allows operations teams to monitor the health of backup systems, software and jobs; triggers alerts in the case of errors; and assists in debugging backup failures. It also allows teams to locate where specific data may be stored, when a recovery is required.



Environmental Sensors

Use Cases: Internet of Things, Business Analytics

Examples: Bosch Sensortec, Mouser Electronics, Raritan, Schneider Electric, TSI, Vaisala

Environmental sensors provide data on barometric air pressure, humidity, ambient air temperature and air quality. They are applied in everything from combating pollution and detecting gasses to keeping data centers from overheating.

Use Cases

Internet of Things: Environmental sensors are a class of smart meters that have been optimized to monitor the environment. In some instances, such as a data center, the information provided by these sensors is used to automatically alter temperature setting and heat flow.

Business Analytics: Environmental sensor data collect can be used in retail applications capable of answering predictive questions, such as “what impact inclement weather might have on foot traffic in a mall?”

Industrial Control Systems (ICS)

Use Cases: Security and Compliance, Internet of Things, Business Analytics

Examples: ABB, Emerson Electric, GE, Hitachi, Honeywell, Rockwell Automation, Siemens, Toshiba

Within the context of a manufacturing environment, industrial control systems make use of programmable logic controllers to both acquire data and execute supervisory functions. Much of the process automation employed in a manufacturing facility is enabled by the industrial control systems.

Use Cases

Security and Compliance: Industrial control systems play a critical role in delivering services to industry and municipalities across the world. These systems live on top of traditional IT infrastructure and — while typically separate from enterprise IT — digital transformation is driving organizations to provide connectivity to these systems, increasing exposure to attacks. These systems tend to be unmanned from a security perspective. Regardless of how ICS might get attacked or infected, data from ICS devices can provide visibility and can be used to analyze and identify malicious activity and potential threats. This visibility enables companies to measure impact and risk, and associate them with business processes.

Internet of Things: Machine data from ICS can be used to gain real-time visibility into the uptime and availability of critical assets. This enables companies to detect an issue, perform root cause analysis and take preventive action to prevent certain events from happening in the future. Companies are also leveraging machine data from ICS systems to secure these mission-critical assets.

Business Analytics: Organizations can apply machine learning algorithms against the machine data created by industrial control systems to increase productivity, uptime and availability. ICS data can also drive visibility into complex manufacturing processes, helping identify bottlenecks and remove inefficiencies.



Mainframe

Use Cases:

IT Operations

Mainframes are the original business computer: large, centralized systems housing multiple processors, system memory (RAM) and I/O controllers. Despite their 60-year legacy, mainframes still are widely used for mission-critical applications, particularly transaction processing. Although they usually run a proprietary OS, mainframes also can be virtualized to run Unix and Linux or, with add-on processor cards, Windows Server. Mainframes are valued for their bulletproof reliability and security, using highly redundant hardware and resilient, stringently tested software. As such, they appeal to organizations wanting to consolidate workloads onto a small number of systems and that need the added reliability and versatility.

Use Cases

IT Ops: Like other servers, mainframes measure and log numerous system parameters that show their current status, configuration and overall health. Since most mainframe subsystems are redundant, system logs also show non-disruptive hardware failures or anomalous behavior that is predictive of an impending failure. Due to their use for critical applications, mainframes often record application performance data such as memory usage, I/O and transaction throughput, processor utilization and network activity.

Medical Devices

Use Cases:

Internet of Things, Business Analytics

Examples: Abbott Laboratories, Apple, Baxter, Boston Scientific, GE, Siemens, St. Jude Medical

Everything from intensive care units to wearable devices generates multiple types of machine data. In fact, just about every aspect of patient care inside and out of a hospital setting can be instrumented. While the primary goal is to save lives, a crucial secondary goal is to reduce healthcare costs by reducing both the number of potential visits to a hospital as well as the length of stay.

Use Cases

Internet of Things: Most devices inside a hospital are connected to local monitoring applications. But it's possible to monitor patient care remotely using sensors that communicate with either a wearable device or some other system for monitoring patients in their homes.

Business Analytics: Machine data also makes it simpler for medical professionals to analyze both patient and anonymous data across a broader range of geographically distributed regions—for example, to see how certain diseases are affecting a group of people more than another.



Metric Line Protocols

Use Cases: IT Operations, Application Delivery, Internet of Things

Examples: collectd, statsd

Metrics are measurements generated by a process running on a system that provide a regular data point around a given metric, such as CPU utilization. Metrics data sources generate measurements on regular intervals and generally consist of:

- Timestamp
- Metric Name
- Measurement (a data point)
- Dimensions (that often describe the host, kind of instance, or other attributes that you might want to filter or sort metrics on)

Metrics are typically generated by a daemon (or process) that runs on a server (OS), container, application. Each data measurement is delivered by a network protocol, such as UDP or HTTP, to a server that indexes and analyzes that information.

Metrics are particularly useful for monitoring. For example, a heart monitor that regularly checks a patient's pulse, metrics provide insight into trends or problems that affect the performance and availability of infrastructure and application. However, a heart monitor won't tell you why a patient has a sudden issue with their heart rate - you need other means to quickly identify the cause and stabilize the patient. It's the same with machine data. When combined with other data sources, usually logs, you gain insight into both what's going on, and why it's happening.

Examples of Metric Line Protocols

collectd: Collectd is a protocol that involves an agent running on a server that is configured to measure specific attributes and transmit that information to a defined destination. Collectd is an extensible measurement engine, so you can collect a wide range of data. Currently, collectd is most often used for core infrastructure monitoring insights, such as getting insight on the workload, memory usage, I/O, and storage of servers and other infrastructure components. Collectd is part of the open source community, and you can learn much more about collectd by visiting <http://collectd.org>.

statsd: is a network daemon that runs on node.js. It has gained popularity with windows administrators, application performance experts and others. Statsd provides some capabilities that allow for metrics to be delivered in batch, and while it uses the less dependable UDP network method, many administrators like how easy it is to deploy. Much like collectd, statsd is focused on collecting metrics, mostly involving the usage and performance of applications and application components, and sending them via the network to a tool that can collect and analyze that information.

Use Cases

IT Ops and Application Delivery: Metrics Line Protocols provides usage, performance and availability data across operating systems, storage devices, applications and other components of IT infrastructure. Metrics are particularly useful for the monitoring portion of IT Operations and Application Delivery, where trends can help identify where there is a problem. Once trends and thresholds illustrate performance issues, other data sources are often correlated to determine the root cause of the problem.

Internet of Things: As devices become more intelligent, more metrics based telemetry will be on board. Metrics line protocols represent an efficient way for these devices to report their status and performance.

Patch Logs

Use Cases:

Keeping operating systems and applications updated with the latest bug fixes and security patches is an essential task that can prevent unplanned downtime, random application crashes and security breaches. Although commercial apps and operating systems often have embedded patching software, some organizations use independent patch management software to consolidate patch management and ensure the consistent application of patches across their software fleet and to build patch jobs for custom, internal applications.

Patch management software keeps a patch inventory using a database of available updates and can match these against an organization's installed software. Other features include patch scheduling, post-install testing and validation and documentation of required system configurations and patching procedures.

Use Cases

Security and Compliance: Security teams can use patch logs to monitor system updates and determine which assets could be at risk, due to failed or out-of-date patches.

IT Ops: Operations teams use patch logs to verify the timely and correct application of scheduled patches, identify unpatched systems and applications, and alert to errors in the patching process. Correlating errors to patch logs can indicate when an error is due to a patch.

Physical Card Readers

Use Case:

Most organizations use automated systems to secure physical access to facilities. Historically, these have been simple magnetic strips affixed to employee badges; however, locations with stringent security requirements may use some form of biometric reader or digital key. Regardless of the technology, the systems compare an individual's identity with a database and activate doors when the user is authorized to enter a particular location. As digital systems, badge readers record information such as user ID, date and time of entry and perhaps a photo for each access attempt.

Use Cases

Security and Compliance: For IT security teams, the data from card readers provide the same sort of access information for physical locations as a network firewall log. The data can be used to detect attempted breaches and be correlated to system and network logs to identify potential insider threats and provide overall situational awareness. It can also be used to detect access at unusual times and locations or for unusual durations.



Point-of-Sale Systems (POS)

Use Cases: Security and Compliance, Internet of Things, Business Analytics

Examples: IBM, LightSpeed, NCR, Revel Systems, Square, Toshiba, Vend

Point-of-sale systems are most often associated with transactions generated at a retail outlet. However, thanks to the rise of mobile POS solutions, many of these systems are starting to be deployed in temporary locations, such as a community fair or a high school event.

The typical POS system incorporates a cash register based on a PC or embedded system, monitor, receipt printer, display, barcode scanner, and debit/credit card reader. Machine data generated by POS systems provides organizations with real-time insight into everything from what's sold, to the amount of cash being generated per transaction, to what payment methods are being used.

Use Cases

Security and Compliance: POS systems are typically used for financial transactions and are often targeted since they contain account, payment and financial information. Because the POS transaction information is highly sought after for its value to attackers, and the POS can be used as an entry point to the network, it's critical to protect these systems. Furthermore, POS systems are usually unmanned, run an underlying operating system, and versioning/monitoring typically fall outside of IT's purview — adding additional complexity to their security. Visibility and analysis of POS systems and data can provide insights that are critical to protecting financial information, detecting fraud and securing vulnerabilities.

Internet of Things: Historically, POS systems were either not connected or managed on a dedicated private network. Thanks to the rise of the IoT, these systems are being connected directly to cloud platforms that make remotely administering these devices from a central location much simpler. There's no longer a need to dispatch IT personnel to manually update each system. This is critical because a POS failure can result in longer lines that inconvenience customers and potentially lead to lost revenue. A negative customer experience can easily translate to customers opting to shop somewhere else in a retail industry that is intensely competitive.

Business Analytics: POS systems contain information about what's sold, how it's paid for, as well as the pace at which it's being sold. Organizations can use this data to monitor revenue in real time, which can feed into how to better market 1:1 against customers, track product placement and sales in a store, or detect potentially fraudulent transactions in real time. This type of real time Big Data analysis can have a profound impact on customers cross- and up-sell opportunities. POS data also delivers visibility into customer experience such as which coupons are most popular or the combinations of products that are selling together. When enriched with geolocation data, it can also drive valuable insights into location-based analytics.



RFID/NFC/BLE

Use Cases: Internet of Things, Business Analytics

Examples: Alien Technology, BluVision, CheckPoint Systems, Gimbal, MonsoonRF, Radius Networks, STMicroelectronics, TAGSYS RFID, ThingMagic

The two primary wireless methods organizations use today to keep track of objects and interact with customers in retail stores involve two distinct types of wireless communications technologies. The better known is radio-frequency optimization (RFID), which involves the use of tags capable of storing information such as product information or what goods might be loaded in a shipping container.

At the same time, organizations are adopting Bluetooth Low Energy (BLE) wireless connectivity solutions that can broadcast signals to other devices. BLE is used most widely in beacons that are employed, for example, to inform shoppers of new sales in retail stores on their smartphones or update fans on events that might be occurring during a sporting event.

Use Cases

Internet of Things: RFID is arguably one of the first instances of an IoT application. Deployed in place of traditional barcode readers, RFID tags are used in everything from shipping to keeping track of farm animals. IoT deployments make it possible to capture RFID data in a way that makes it simpler to track events involving anything that has an attached RFID tag. Data insights from RFID can help improve overall supply chain, order processing and inventory management.

BLE, meanwhile, is used to engage customers more directly as they move about a specific location, which in turn creates data that can be used to optimize the customer experience.

Business Analytics: Whether it's inventory tracked using RFID tags or customers and employees moving around specific locations, new classes of analytics applications are using the data generated by these devices to serve up actionable business insights in near real time. Retailers can leverage this data for several use cases, such as making sure that inventory is located as close as possible to the locations where customers are most likely to want to purchase.



Sensor Data

Use Cases: Security and Compliance, IT Operations, Internet of Things, Business Analytics

Examples: Binary and numeric values including switch state, temperature, pressure, frequency, flow, from MQTT, AMQP and CoAP brokers, HTTP event collector

Industrial equipment, sensors and other devices often have embedded processors and networking that allows them to record and transmit a vast array of information about operating conditions. Regardless of device, their data provides unprecedented detail about performance parameters and anomalies that can indicate larger problems — for example, a device ready to fail or issues with another system. Aggregating and correlating data from multiple devices and subsystems provides a complete picture of equipment, system, factory or building performance.

Use Cases

Security and Compliance: Sensor data can help protect mission-critical assets and industrial systems against cybersecurity threats by providing visibility into system performance or set points that could put machines or people at risk. Data can also be used to satisfy compliance reporting requirements.

IT Ops: Some of the most important parameters for operations teams to monitor are environmental conditions such as temperature, humidity, airflow and voltage regulation in a data center. Similar readings are available from individual servers and network equipment that, when correlated, can highlight problems in the facility or equipment ready to fail.

Additional Use Cases

Preventative Maintenance and Asset Lifecycle Management:

Sensor data can provide insights into asset deployment, utilization and resource consumption. Operational data can also be used to proactively approach long-term asset management, maintenance and performance.

Monitoring and Diagnostics: Monitoring sensors can help ensure that equipment in the field operates as intended, for example, monitoring and tracking unplanned device or system downtime. The data can also be used to understand the cause of failure on a device to improve efficiency and availability, and to identify outliers and issues in device production or deployment.





Server Logs

Use Cases: Security and Compliance, IT Operations, Application Delivery

Server operating systems routinely record a variety of operational, security, error and debugging data such as system libraries loaded during boot, application processes open, network connections, file systems mounted and system memory usage. The level of detail is configurable by the system administrator; however, there are sufficient options to provide a complete picture of system activity throughout its lifetime. Depending on the subsystem, server logs are useful to system, network, storage and security teams.

Use Cases

Security and Compliance: Server logs include data from security subsystems such as the local firewall, login attempts and file access errors that security teams can use to identify breach attempts, track successful system penetrations and plug vulnerabilities. Monitoring server logs such as file access, authentication and application usage can help secure infrastructure components.

IT Ops and Application Delivery: Server logs provide a detailed record of overall system health, and forensic information about the exact time of errors and anomalous conditions that are invaluable in finding the root cause of system problems.

Smart Meters

Use Cases: Internet of Things, Business Analytics

Examples: ABB, GE, Google, eMeter, IBM, Itron, Schneider Electric, Siemens

Smart meters record consumption of energy, usage of water, or usage of natural gas so that the information can be continually processed and shared. Typically, smart meters allow for bi-directional communication in real time in a way that allows a gauge of some type to be adjusted.

Use Cases

Internet of Things: Smart meters are deployed across critical systems at large utilities companies, for example, power, gas and water utilities. These systems are the lifeblood of infrastructure and failure can lead to catastrophic outcomes. Real time monitoring of smart meters can help organizations better analyze failures remotely, by way of detecting remotely line down failures. Equally important is securing the devices from tampering that could lead to malicious attacks and breaches.

Energy companies and water utilities make extensive use of smart sensors to track everything from oil reserves to the quality of the water supply.

Business Analytics: A wide variety of industries are applying analytics to the data being collected by smart meters to optimize service. For example, an oil or gas company no longer needs to physically send a worker to a location to read a meter. The provider already knows how much fuel has been consumed and how much remains.

Smart meters in the future will be used in everything from modern traffic control systems to defense systems designed to protect critical infrastructure. Aggregating data from these smart meters can give utilities critical insights into the demand. Heavily regulated utilities are required to meet established SLA's during demand response events, and machine data from smart meters can drive visibility into how they are responding.



Storage

Use Case: IT Operations

Examples: EMC, Netapp, IBM, Amazon EBS

Data center storage is provisioned in two general ways: built into servers and shared using various network storage protocols, or via a dedicated storage array that consolidates capacity for use by multiple applications that access it using either a dedicated storage area network (SAN) or ethernet LAN file-sharing protocol. The activity of internal, server-based storage is typically recorded in system logs, however storage arrays have internal controllers/storage processors that run a storage-optimized OS and log a plethora of operating, error and usage data. Since many organizations have several such arrays, the logs often are consolidated by a storage management system that can report on the aggregate activity and capacity.

Use Cases

IT Ops: Shared storage logs record overall system health (both hardware and software), error conditions (such as a failed controller, network interface or disks) and usage (both capacity used per volume and file or volume accesses). Collectively, the information can alert operations teams to problems, the need for more capacity and performance bottlenecks.

Telephony

Use Cases: IT Operations

Examples: Cisco Unified Communications Manager, Shoretel, Twilio

Real-time business communications are no longer limited to voice calls provided by plain old telephone service (POTS); instead, voice, video, text messaging and web conferences are IP applications delivered over existing enterprise networks. Unlike traditional client-server or web applications, telephony and other communications applications have strict requirements on network quality of service, latency and packet loss, making service quality and reliability much more sensitive to network conditions and server responsiveness. Traditional POTS has conditioned people to expect immediate dial tone when picking up the phone and be intolerant of noise, echo or other problems that can plague IP telephony; as such, the systems and supporting infrastructure require careful monitoring and management to assure quality and reliability.

Use Cases

IT Ops: Like VoIP, telephony logs provide an overview of system health along with troubleshooting and usage data similar to that of other network applications. Details include source, destination, time and duration of voice/video calls, web conferences and text messages, call-quality metrics (e.g., packet loss, latency, audio fidelity/bit rate), error conditions and user attendance at web conferences. By integrating telephony records of source/destination address with an employee database such as AD or LDAP and a DHCP database, organizations can link call records to actual user IDs and IP addresses to physical locations; information that can assist in troubleshooting and billing. Logs also can reveal any network segments experiencing congestion or other performance problems that may indicate equipment problems or the need for an upgrade.



Transportation

Use Cases: Internet of Things, Business Analytics

Examples: Boeing, BMW, Ford, GE, General Motors, Daimler-Benz, John Deere, Volkswagen

Vehicles of all sizes and types generate massive amounts of machine data every day that can be used to gain real-time visibility into the health and performance of an asset, and to drive predictive maintenance applications. Armed with that data, an airplane or automobile manufacturer can follow a maintenance regime that is data driven rather than driven “by the book.”

That information can then be used to improve availability and reliability, and extend the life cycle of a vehicle that has not been extensively used or, conversely, replace components that have seen extensive wear and tear sooner.

Use Cases

Internet of Things: Vehicle manufacturers are attaching sensors to every mechanical and electronic component they use. This allows companies to gain a unified view of assets to quickly identify and diagnose operational issues, and to monitor, track and avoid unplanned asset downtime. This helps to ensure that equipment is operating as intended. They can also detect anomalies and deviations from normal behavior to take corrective action — improving uptime, asset reliability and longevity.

Business Analytics: With access to machine data, vehicle manufacturers are applying analytics in ways that fundamentally changes their business models. Instead of selling a vehicle, manufacturers increasingly prefer to lease vehicles based on actual usage. The longer that vehicle can be used between repairs, the more profitable that leasing service becomes. The key to providing this type of service economically is advanced analytics, which are applied to all the aggregate data that's collected.

Wearables

Use Cases: Internet of Things, Business Analytics

Examples: ARM, Intel, Lenovo, Microsoft, Samsung

From smartwatches that double as fitness aids to medical devices that enable physicians to remotely monitor vital statistics, wearable devices have proven they are here to stay. Wearable devices are one of the most recognizable parts of the Internet of Things.

Use Cases

Internet of Things: Beyond merely syncing with smartphones, the latest generation of smartwatches is taking advantage of geo-positioning systems and application programming interfaces to give device owners an optimal application experience that includes both their location and often time of day.

Going forward, there soon will be whole new classes of wearable devices taking advantage of everything from virtual reality applications delivered via a headset to sensors embedded in the latest fashion.

Business Analytics: As more people become comfortable with sharing data via wearable devices, many are experiencing the power of analytics firsthand. Developers of applications optimized for wearables are making recommendations concerning everything from how to improve life expectancy to where to find a meal. Analytics from wearables can help improve user experience and drive product innovation. For example, product managers can understand how consumers are interacting with devices to build better features.





Additional Data Sources

Database

Use Cases: Security and Compliance, IT Operations, Application Delivery

Examples: MySQL, Postgres, Other Relational Databases

Databases are the fundamental elements of information collection, storage and analysis of digital information. Databases are categorized as either relational, in which data is organized in spreadsheet-like tables of columns and rows, or NoSQL (non-relational), where information is organized purely by columns (column store) as key-value pairs, by unstructured documents or interconnected graphs linking related data elements.

Use Cases

Security and Compliance: Database logs provide security teams information about the accounts or systems accessing tables or other database elements. Correlating database access and transaction logs with identity management system records can flag unauthorized access or access attempts to databases. Database logs can also expose security holes such as open ports or dormant, unused admin accounts, and help identify abnormal queries or users, and abnormal database/table access.

IT Ops and Application Delivery: Database logs can be aggregated and analyzed to show the overall performance of a particular database system, and also provide visibility into database issues. Metrics useful to IT operations teams include queries per second and query response time, both measured against a baseline standard made from historical data.

Business Service Transaction and Business Service Performance Data

Use Case: Security and Compliance, IT Operations, Application Delivery

Examples: Payments Status, Batch Upload Status, Customer Order Status, Requests per Customer, SLA Tracking, Business KPI Tracking, SLO and SLI insights

Transaction records provide an auditable trail of activity for every part of every business process. Whether for financial transactions such as payments and orders, or tasks such as customer support and service calls, business process logs are required to verify activity in case of disputes, to certify compliance with regulations and terms of service, and to provide detailed evidence of business transactions. A technique called business process mining uses sophisticated software to analyze logs and identify process, control, data, organizational and social structures. These might include mapping the flow of patients through a hospital or customer problems through a support organization to optimize process flow, measure performance and identify outlier incidents for further investigation. Tracking real-time business performance metrics like response times or even shoes sold per minute during a new product launch, allows teams to provide a consistent customer experience and benchmark for future releases.

Use Cases

Security and Compliance: Hackers are good at covering their tracks by altering common log files, but business process logs that track activity across multiple systems used in a particular process can highlight anomalies that may indicate security issues.

IT Ops and Application Delivery: IT can use process logs to identify flaws in their support or admin processes, or problems that have fallen through gaps in existing process flows. IT can use business performance metrics for understanding system baselines and comparing performance to SLIs and SLOs to ensure customer requirements are met. These baselines are great for capacity planning and enabling faster release cycles thanks to increased predictability. Monitoring custom business metrics can also provide real-time insights, especially during high-volume service spikes, so teams can mitigate customer experience issues before the business is affected.



Human Resources

Use Cases: Security and Compliance

Examples: BambooHR, Fairsail HRMS, Namely, Zenefits

Human Resources records include information relating to the entire employee life cycle. HR records provide the definitive source of employee information for identity management systems and enterprise directories, making them an important source for authentication and authorization data. Although HR data traditionally has been textual, it increasingly includes images and biometric information such as an employee's portrait, fingerprints and iris scans.

Use Cases

Security and Compliance: HR records can show if someone no longer employed still has active accounts, and can also provide evidence of disciplinary action that might be useful in security investigations.

Social Media Feeds

Use Cases: IT Operations

Social networks are some of the most heavily trafficked sites on the internet. By allowing users to communicate and share information among friends and colleagues, social media has become an important outlet for news, entertainment, photo-sharing and real-time reaction to public events. As such, social media feeds are an increasingly effective advertising medium and source of customer contact, feedback and support.

Use Cases

IT Ops: Due to their interactivity, convenience and ubiquity, social media feeds provide organizations with an unfiltered and instantaneous view of customer opinion. By analyzing feeds from the most popular sites, organizations can quickly identify potential problems with a product or service, mishandled customer support incidents or other sources of customer dissatisfaction about an organization's products or online presence. Proactively addressing these online complaints allows the organization to turn unhappy and potentially lost customers into delighted and loyal ones.





Third-Party Lists

Use Case: Security and Compliance

Examples: Threat Lists, OS Blacklist, IP Blacklists, Vulnerability Lists, Google Analytics

One of the methods that IT security vendors use to detect and flag security problems is one or more databases of known threats and vulnerabilities. These include malware code signatures, OS and application patch versions, the source IP address of previous attacks and spam and reputation databases using real-time aggregation of malware, spam and compromised websites collected from millions of users. Third-party lists provide an early-warning system for new methods or sources of attack.

Use Cases

Security and Compliance: By aggregating data from users around the world, third-party security lists provide security teams with real-time information about nascent threats and vulnerabilities that allow updating security policies, firewall rules and vulnerable software before an attack. Lists also are used to identify known sources of spam, both commercial and malware-infested, to improve the effectiveness of filters on internal email systems.

About Splunk.

Splunk Inc. turns data into doing. Splunk technology is designed to investigate, monitor, analyze and act on data at any scale. Join millions of passionate users by trying Splunk for free.

Free Trial

splunk>

Splunk, Splunk> and Turn Data Into Doing are trademarks and registered trademarks of Splunk Inc. in the United States and other countries. All other brand names, product names or trademarks belong to their respective owners. © 2022 Splunk Inc. All rights reserved.

22-13476-Splunk-Essential-Guide-to-Data-EB-111

