Nothing Splunked, Nothing Gained

Profiles of Splunk® Customer Success

splunk>

Published by Splunk Inc., 250 Brannan St., San Francisco, CA 94107

Editor/Analyst: Splunk Inc.
Copyeditor: Splunk Inc.
Production Editor: Splunk Inc.
Cover: Splunk Inc.
Graphics: Splunk Inc.

Second Edition: March 2015

Disclaimer

This book is intended as a text and reference book for reading purposes only. The actual use of Splunk's software products must be in accordance with their corresponding software license agreements and not with anything written in this book. The documentation provided for Splunk's software products, and not this book, is the definitive source for information on how to use these products. Although great care has been taken to ensure the accuracy and timeliness of the information in this book, Splunk does not give any warranty or guarantee of the accuracy or timeliness of the information and Splunk does not assume any liability in connection with any use or result from the use of the information in this book. The reader should check at docs.splunk.com for definitive descriptions of Splunk's features and functionality

A Note About Customer Names and Examples

The stories in this book accurately convey how actual customers achieved compelling value using Splunk software. In order to tell the stories at a granular level of detail, we do not name the customer or organization. Furthermore, in some cases, when a customer is so unique or so dominant in its industry that it would be easy to guess its name, we have altered certain identifying information in our descriptions, such as geographic location, or the precise business involved. Rarely, we have merged the stories of similar customers or use cases in order to further conceal a customer's identity.

In certain stories we have recreated dashboards and correlations based on customer interviews or customer presentations. When possible, we have updated the examples to reflect the most recent version of Splunk software. Specifics such as value information, (e.g., MTTR improvements and similar statistics), the approximate size of the company and the details of the Splunk Enterprise deployment are all factually accurate.

Table of Contents

Introduction

My favorite part of the job is meeting with our customers all over the world. They consistently tell me that stories of Splunk customers' success often inspire their own organizations to gain insight and business value from machine-generated data. Customers and partners also tell me that they want more in-depth, technical information on how the Splunk platform is deployed in real-world environments to solve problems, improve operations and provide ROI. They seek valuable details on common use cases, including typical data sources and exactly how Splunk software is being used.

I'm pleased to announce that these requests were an inspiration for our new E-book ***Nothing Splunked, Nothing Gained: Profiles of Splunk Customer Success***. Written for both business and technology leaders, this is a collection of in-depth customer use cases that describes the business problems, technical solutions and the return on investment customers have achieved with Splunk software.

Each chapter of the book highlights the value that a customer has gained from collecting, analyzing and visualizing machine data generated by its IT systems and technology infrastructures. Each chapter introduces a different business challenge and shows how leveraging machine data and Splunk software in new and interesting ways can drive powerful business and operational outcomes.

I hope this book provides greater awareness and understanding of what the Splunk platform can do for you. Although each chapter profiles a use case specific to a particular customer and its industry, Splunk's underlying principles and approaches can apply to organizations in any industry. We'll continue to add chapters to share the paths customers take as they implement and benefit from Splunk software.

Best regards,

Godfrey R. Sullivan
Chairman & CEO
Splunk Inc.

CHAPTER 1

Understanding the Digital Audience

How Splunk Software is Used to Find the Needle and See the Whole Haystack

USE CASES

Marketing Analytics & Reporting
Improving Operational Efficiencies

Executive Summary

In the early days of Splunk® Enterprise, its major appeal was delivering centralized searches for log and event data. The initial benefits were powerful because manual processes and tedious procedures were eliminated. With access to such data from a single location, users were empowered to find the proverbial "needle in the haystack." Virtually all of Splunk's early customers were network, security or application developers desperate for a tool that enabled "Google for their logs" to find error messages, failed logins or the thrown exception.

As Splunk software evolved beyond a search engine for machine data into a platform for operational intelligence, it has begun to attract the attention of people interested in looking at the "whole haystack." Data scientists, business analysts, digital marketers and others who analyze large amounts of data have begun to incorporate Splunk software into their organizations' data management architecture.

One Splunk customer—a business analyst for a national media company—discovered the solution when searching for a way to report on the company's "whole haystack" of digital audio and video distribution. Once Splunk software was in the company's environment, the analyst found that it enabled her to address many critical challenges faced by other analysts:

- **Track usage of new platforms.** Ten years ago, a digital presence simply meant a website; now it means a website, a mobile site, mobile applications, partnerships with other sites and a presence on social media platforms. Traditional tools, which are designed on legacy technology frameworks to track and analyze each of these different platforms, haven't kept pace. Because Splunk software can flexibly collect and index any type of machine data, it is perfectly suited to track new platforms as soon as they come online.
- **Present the big picture.** Traditional tools are designed with databases that require data normalization to conform to a certain schema. This means that they can report on some types of audience activity (such as page views to a website), but not easily on others (such as tweets). Splunk software imposes no schema and doesn't rely on a database, so it can index, correlate and report on any type of user activity.
- **Answer unexpected questions.** Traditional solutions offer a collection of pre-defined or out-of-the-box reports, possibly including a limited search and correlation capability. For many users, a select few of these out-of-the-box reports are useful, but many are irrelevant. The few out-of-the-box reports that are useful address very specific questions (e.g., "How did that story do?"). Unfortunately, they are limited in supporting iterative analytics or answering subsequent questions (e.g., "Did it fare differently on different platforms in different cities?"). Thanks to its flexibility, Splunk software can easily adapt and apply correlation and statistical analysis to the data, enabling analysts to answer unforeseen questions.

Many Lenses, Little Insight

As a business analyst, the customer was exposed to solutions from many vendors. There was no shortage of business intelligence platforms, dashboards and tools designed to meet one or two of the company's needs. The customer regularly got emails and phone calls from firms selling products for tracking the company's main site, its mobile applications, its overall mobile site traffic, brand sentiment as expressed in social media sites, and the audio and video it distributed on its site. These vendors offered historical analysis, real-time dashboards and everything in between.

Enter Splunk

The company had invested in a traditional web analytics platform that relied on a JavaScript-based, client-side tracking

Business Benefits at a Glance

Splunk Value	How Value Is Measured	Business Impact
Greater and more timely visibility into all operational data	• Better business intelligence • More informed decisions • Increased customer engagement	Grew its audience by 17% within a single year
Aggregate metrics from various sources with a single platform	Reduce the use of multiple solutions to track digital audio and video traffic	Saving approximately $100,00 each year in log tracking service costs
Deeper insight into system performance	Optimize API performance	• Sped API performance by 50% • Reduced infrastructure upgrades • Enhanced user experience

methodology. This was sufficient for tracking traffic to the website, where the company could include the requisite JavaScript into its web pages. However, a critical component of the company's digital strategy included distributing audio and video through channels it couldn't "tag" such as iTunes.

To report on that kind of traffic, the analyst needed insight into the access logs where each download request was recorded. After much research, the analyst concluded that Splunk Enterprise was the best option to parse and analyze that data quickly.

However, once Splunk software was acquired to solve the problem of reporting on audio and video traffic, the company found that the solution helped to solve other issues by empowering its stakeholders with the data they need to make critical decisions.

A single analytics tool couldn't paint the whole picture. The company had a website, three different mobile applications, a mobile site, and audio and video content distributed through third-party apps including iTunes. It used three different types of tracking tools to record and report on traffic to each of these various channels.

Thus, when executives asked for a dashboard that provided a holistic view, analysts would have to extract reports from each tool, normalize the data in Excel and then craft a dashboard that summarized the data into high-level numbers that stakeholders could understand at a glance.

This process was tedious, time-consuming and prone to errors. As such, reports couldn't be created more frequently than once a month. In the news business, content strategy discussions are not monthly—they happen at least once a day. Because the handcrafted reports lacked up-to-the-moment insights, they ultimately served as historical records.

The tracking tools couldn't keep pace with technology. Virtually all digital intelligence tools rely on client-side tracking to gather user information. This approach allows the tool to record a user's every point and click—very useful for user interface designers.

However, the predominant requirement for web traffic analysis is to understand how a user engages with the content. For this type of analysis, server-side tracking is just as useful as client-side tracking, if not more so.

Client-side tracking has some disadvantages: it can only track traffic to platforms the customer can "tag," which means that third-party apps, social media sites, and direct audio and video downloads are excluded. Also, for those platforms that can be tagged, it's a time-consuming process to code everything properly and start recording the right data in the most useful way. If a platform is launched—a new app for an Android tablet, for example—and the tagging wasn't done properly, the information is lost.

Company stakeholders couldn't ask complex questions. One of the selling points frequently heard from vendors was the large number of pre-designed reports their tool would provide. Indeed, most analytics tools came with an excess of 100 reports. Unfortunately, most of these pre-designed reports weren't useful to the product team.

For example, they could generate a report that showed traffic broken down by operating system (Macintosh, iOS, Windows XP, Android, etc.), but at one point the product team needed to break iOS traffic down by version—something traditional tools just couldn't do.

In other words, traditional tools could give stakeholders enough basic information to approximate the size of their audience and how they engage with the content, but these systems were limited in their ability to help the team to thoroughly understand the complexities of the audience, which impeded innovation.

The Splunk Solution: Answering Key Questions

Once the Splunk platform was indexing data, the customer was able to generate the basic reports to meet stakeholders' needs. All stakeholders—from the content producers to the media sales department to the executive suite—wanted answers to several very key questions:

"How did my programs do last week?" Before having direct access to the raw logs, answering even this basic question was challenging. Since programs were distributed on a variety of platforms—including websites accessed on desktops, mobile sites and mobile apps—stakeholders usually had to piece together reports from different tracking tools by cutting and pasting numbers into Excel. This meant that reports were laborious, delayed and not generated very frequently.

Once they had designed some basic dashboards in Splunk Enterprise, stakeholders were able to automate the delivery of weekly reports for anyone who requested one. For example, reports on audio and video downloads were based on standard access logs (see Figure 1).

They associated each program ID with the program name (in this case "Products") using a lookup table (see Figure 2), so that the final report would be easily readable by anyone.

First, the query filtered out events that didn't represent completed downloads:

```
> sourcetype="access_combined"
status<300
```

Then, they added a timechart command to group downloads into increments of one day, broken down by program name:

```
> sourcetype="access_combined"
status<300 | timechart span=1day count
by ProgramName
```

Audio File Download Event

```
69.133.25.145 - - [04/Oct/2012:07:49:14
-0500] "GET /pum/pum092112pod.mp3 HTTP/1.1"
206 68056 "-" "AppleCoreMedia/1.0.0.10A403
(iPhone; U; CPU OS 6_0 like Mac OS X; en_us)"
```

Figure 1. A standard access log highlighting the status field and program ID.

Lookup Tables

Lookup tables allow the user to expand on raw data by associating a value within the event with other information. To accomplish this, the user creates a comma-separated file and delineates the column headers on the first line of the file. Each subsequent line's first value matches the value in the raw data.

ProgramID,ProgramName,ProgramCategory
pum,Products,Long Form
des,Design,Short Form
hum,Humor,Short Form

Once a lookup table is configured through Splunk's web interface, anytime the raw value is found in the data, the other values on the corresponding line of the lookup table are added as if they, too, were part of that event.

Figure 2. Lookup table to associate program ID with program name.

They used the visualization tool in Splunk Enterprise to create a stacked column graph that displayed both overall traffic and each program's relative contribution to the total (see Figure 3).

What programs are trending on what platforms? All media organizations were investing heavily in multi-platform distribution, but few were able to make truly informed decisions about how each platform could best be used.

The stakeholders for the national media company could only suspect that certain types of content were more popular on certain platforms—long-form music programs fared better on the iPad, for example, while short news segments were more popular on desktop browsers. They were basing their decisions mostly on gut instinct—until they had a platform for machine data.

To track which audio files were accessed on which platforms, they used the section of each event that detailed the user agent, in this case, "CompanyApp/2.3.3 (iPhone; U; CPU OS 6_0 like Mac OS X; en_us)" (see Figure 4).

However, because the user agent is defined by the software running on the user's device, it tends to be very detailed and

Audio File Download Event

```
69.133.25.145 - - [04/Oct/2012:07:49:14
-0500] "GET /pum/pum092112pod.mp3 HTTP/1.1"
206 68056 "-" "CompanyApp/2.3.3 (iPhone; U;
CPU OS 6_0 like Mac OS X; en_us)"
```

Figure 4. A standard access log highlighting the user agent.

variable. In this sample line, for example, the user agent includes the software title and version, the device, operating system, and language. This represented a challenge for the customer, because a simple report showing events by user agent would have generated thousands of results.

To solve this problem, stakeholders created event types (see Figure 5) to group similar user agents into broader categories. For example, to learn when a user downloaded a file from the company's mobile application using an iPhone, they created an event type called "Mobile App—iPhone" that filtered for any event where the user agent contained the text "CompanyApp" and "iPhone." The end result was a tidy, readable chart that enabled stakeholders to understand how users were engaging on different device/application combinations (see Figure 6).

What percentage of users have upgraded? Although stakeholders found it useful to group user agent data into larger categories for general reports, there were times when being able to filter events to answer specific strategic questions was also very useful.

For example, at one point the company's mobile product team was considering adding a new feature to its iPad application. This feature would only work for users who had upgraded their iOS operating system to the latest version, and adding the feature entailed a significant investment of development time. Would that investment be worth it? Could most of their users actually take advantage of the new feature?

The product team wondered if there was any way it could estimate what percentage of users had upgraded. Fortunately, although that type of analysis required filtering events by an obscure substring within the user agent field (see Figure 7), Splunk's approach to indexing the entire raw event and extracting

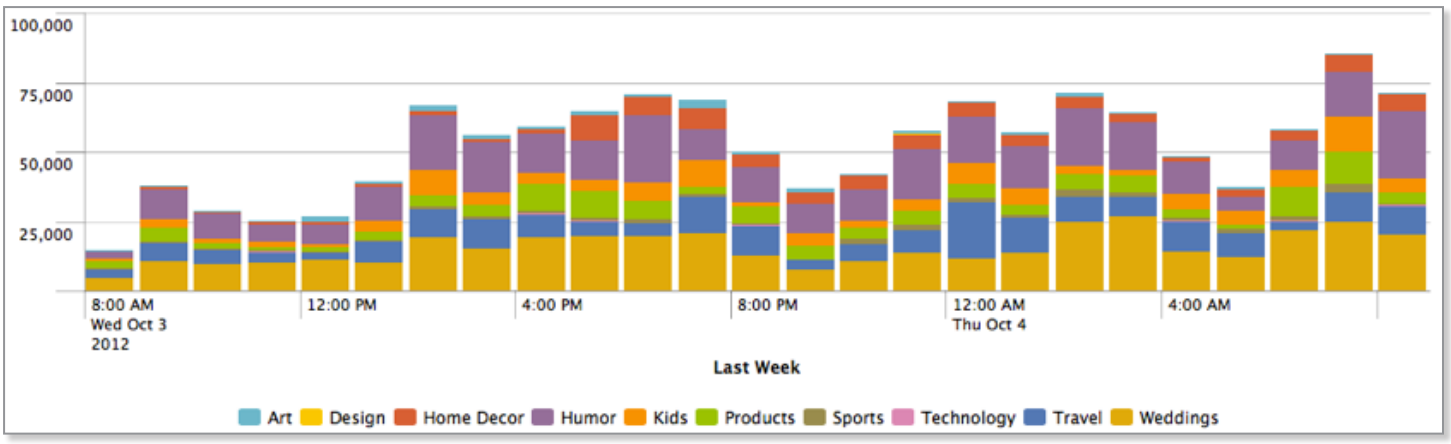

Figure 3. Visualization in Splunk showing total traffic and each program's contribution to the overall traffic.

Event Types

Event types allow the user to group individual events into categories using standard Splunk search language. For example, the user could create an event type called "Bad Request" and define the search string as:

status>=300 OR error* OR fail*

Then, a search for

eventtype="Bad Request"

would match any event that fit the described search string. Even more useful is events can be grouped by event types. Thus,

* | stats count by eventtype

would group all events by their associated event types. For example, if a user wanted to understand what percent of his events were good requests and what percent were bad, he would create an event type describing "Good Request" and one describing "Bad Request" and then apply the "top" function, as in:

* | top eventtype

The result would look like:

	eventtype	count	percent
1	Good Request	37141	99.328733
2	Bad Request	251	0.671267

Figure 5. Event types group similar users agents into broader categories.

subsets of the data into fields only at search time enabled the customer to perform the analysis.

The customer used the **rex** command in the Splunk software to extract the portion of the user agent field relevant to the analysis:

```
* | rex field=useragent "OS
(?<version>\d)" | stats count by
version
```

This query generated results that indicated that more than 70% of users had, indeed, upgraded and could enjoy the new feature. By being able to perform such a specific analysis quickly, the customer enabled stakeholders to make tactical and strategic decisions informed by timely operational data.

How many people have seen my story on Twitter? Social media platforms are powerful promotional tools for news organizations and have become a major driver of traffic. While the customer had a wide variety of social media tracking tools to choose from, most offered metrics that weren't valuable—in particular, they focused on mysteriously calculated sentiment scores, while failing to give access to more basic indicators.

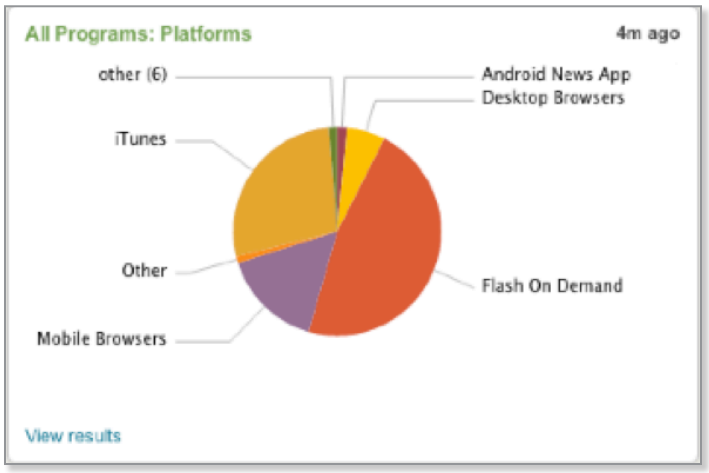

Figure 6. Chart that shows how users engage on different platforms.

Using a scripted input (see Figure 8), the business analyst indexed tweets containing hashtags and keywords related to the company. When tweets included a link to a particular story on the company's site, they were able to extract the URL to create a report that listed the day's most tweeted stories. They were even able to calculate a reach for each story by adding the number of followers of each tweeter at the time they tweeted the story link.

Finding the Needle While Seeing the Whole Haystack

As with so many customers, once this news organization's business analyst brought in Splunk software, its usage spread throughout the enterprise. While the production team was using Splunk Enterprise to better understand the digital audience, application development teams were using the software to diagnose errors in their custom-built content management system and still others were using Splunk software to measure usage of the company's publicly available API.

The results were forthcoming and substantial. First, the company grew its overall audience. By using Splunk software to better understand how audiences engage in different platforms with different content, it was able to optimize the overall experience. In the year after implementing Splunk Enterprise, its digital listening grew 17%.

Second, by using Splunk Enterprise to track digital audio and video traffic, it saved approximately $100,000 per year in log tracking service costs.

Audio File Download Event

```
69.133.25.145 - - [04/Oct/2012:07:49:14
-0500] "GET /pum/pum092112pod.mp3 HTTP/1.1"
206 68056 "-" "CompanyApp/2.3.3 (iPhone; U;
CPU OS 6_0 like Mac OS X; en_us
```

Figure 7. A standard access log highlighting a specific substring.

Moreover, the company used Splunk software to accelerate API performance by 50%, reducing the need for infrastructure upgrades, improving its audience's overall digital experience and increasing user satisfaction.

In other words, rather than simply searching for a needle in the haystack, this company uses Splunk Enterprise to see the whole haystack and, ultimately, to improve its operations and efficiencies.

With Splunk, They Captured the Big Picture

In this use case, we explored how Splunk software can enable operational intelligence for a classic business analytics challenge: "How do I extract value from the information across a variety of disparate sources and third-party applications?" This use case demonstrated:

- **Elimination of data silos.** Because Splunk software indexes all kinds of data, the customer's reliance on multiple reports and formats from varied systems was eliminated and the customer was able provide real-time information to stakeholders.
- **Correlations drive analytics.** Because Splunk Enterprise correlates different types of data, including structured and unstructured data, the customer was able to perform critical analytics and gain business intelligence such as determining whether users had upgraded their devices.
- **Flexible analytics powered by a read-time schema.** Because Splunk software collects data in full fidelity without any filtering, the customer doesn't lose any potential value by making its data fit in a schema. This means the customer can engage in on-the-fly analysis to determine the success of its programming.
- **Value generation across multiple use cases.** In addition to gaining insights into its digital audience, the company accelerated API performance by 50%, curtailing the need for infrastructure upgrades and providing a better user experience.

Scripted Inputs

Splunk software can index data one of three ways: by monitoring a file or folder, by monitoring the output of a port, or by indexing the output of a script. When indexing the output of a script, Splunk can be configured to trigger the script at a frequency determined by the user.

Among other things, the scripted input is useful for indexing data "scraped" from a website or an API, such as the Twitter API.

Figure 8. Scripted inputs allow users to index valuable data.

CHAPTER 2

Building a Better Business Process

How Splunk Software is Used to Provide Real-time Visibility Into Sales and Marketing Data

USE CASES

Marketing Analytics & Reporting
Improving Operational Efficiencies

Executive Summary

What would an executive team say if you told them you could provide real-time visibility into sales activity while reducing infrastructure downtime? While most executives have access to an abundance of sales reports, transaction data, web traffic statistics and marketing projections, those reams of data rarely come together to provide a real-time pulse of the business.

Executives at one Splunk customer, a national wireless carrier, found that using Splunk® Enterprise to create real-time dashboards provides a never-before-seen world of customer data for its IT, sales, marketing and executive teams. This clarity provided unique benefits that traditional reporting tools couldn't match, including an overview dashboard displaying transactions broken out by zip code, by rate plan type or provider. Subsequent screens displayed activations over time, average revenue per activation, and sales of devices and rate plans. The marketing team was even able to monitor which artists and songs are trending on the company's music download service.

These real-time dashboards addressed needs that traditional business intelligence systems are not designed to fulfill:

- **Delivered a real-time view of sales activity.** Once the customer had real-time visibility into subscriber behavior, they were able to get a vivid view of sales activity that they couldn't get from historical reports. Are we growing the way we'd like in the Southwest? Has that promotion for the new handset been effective? What device types are driving the most music downloads?
- **Reduce downtime to improve the bottom line.** Conversely, real-time visibility provided a newfound agility in correcting issues as they arose. In one case, the company was able to identify a dangerous revenue hole—a technical glitch that enabled some customers to use their phones without paying for service—before the issue had a material impact on the bottom line.
- **Simplify manual business processes.** By displaying every transaction on a single screen in real time, the sales and marketing teams no longer had to manually review reports and projections for information and were empowered to literally look at the big picture.

Finding Business Insight in a Complex Architecture

As a US wireless provider with $2B in annual revenue and over 5M subscribers, this customer had a number of sales and marketing reporting systems. Every component of the organization's billing, point of sale and customer management solutions generated data that staff could collate into periodic reports.

As anyone who has bought a mobile phone can attest, the purchase and activation process is not simple. The user's credit must be checked, the device has to be provisioned and the financial transaction must be completed. Existing users may want to add or remove features, disconnect other phones or change rate plans. In the case of this wireless provider, the process was complicated further by the desire to supply a variety of rate plans and network providers for subscribers.

Needless to say, the resulting transaction architecture was extremely complex, comprising hundreds of APIs executing dozens of business processes between both proprietary and third-party billing systems.

The customer had invested in a middleware system that coordinated and collated these transactions from various point of sale and billing systems and passed the information to its internal reporting systems as well as its customer-facing website. This middleware acted as the hub for the entire enterprise and thus was ripe for analysis.

Traditional BI and the Constraints of Non-traditional Use Cases

Business intelligence systems serve a traditional reporting function in this organization. Batch-oriented reports and dashboards delivered on a daily, weekly, monthly or quarterly schedule provide the organization with snapshot views across a variety of areas such as sales, marketing and merchandising. These systems, however, imposed serious limitations.

Business Benefits at a Glance

Splunk Value	How Value Is Measured	Business Impact
Provide a precise, customized view of operations	Time to respond to outages before cascading throughout the system	• Reduced downtime by 15% • Estimated annual savings of $1.2M
Gain real-time visibility into marketing initiatives	Efficiency and agility of the marketing team to respond to user feedback	New ability to escalate and de-escalate promotions based on immediate feedback
Understand customer usage and satisfaction	Improve customer satisfaction to reduce churn	Lowered the high cost of losing and/or regaining customers

Reliance on batch reporting hinders agility. Reporting in batch mode impacts the pace and thus the quality of decision making.

Relying on data that's a few hours old can create doubt. Has that special promotion kicked in since these numbers were pulled? Do people in that region of the country shop less on Fridays than those in other regions? In many cases, circumstances can change by the time a report has been produced.

Database schemas are designed to support only specific upfront queries. When an executive does find a glimmer of insight in a report and asks a follow-up question, analysts face major challenges in quickly providing the answer. In this scenario, the data is either too difficult to correlate or simply not in the system.

Traditional business intelligence tools rely on databases, and databases rely on tables with strictly defined rows and columns. Within this structure, relationships in the data that might not seem important at the time of schema definition can often be critical to understanding a specific business opportunity. For example, understanding the relationship between a transaction type and the location of a store can significantly help executives understand the business impact. All of those relationships must be considered in advance when using systems that require upfront definition via a schema.

In addition to these limitations, the customer's combination of off-the-shelf and homegrown systems used to cobble together its sales and marketing picture was very costly to maintain. Keeping the homegrown systems functioning properly, for example, demanded staff to manage databases and modify schemas.

Ultimately, the company's business intelligence systems failed to provide real-time insights into its customers and operations and were expensive to deploy and operate.

The Splunk Solution: Real-time Monitoring and Analytics

When the director of application operations learned of Splunk Enterprise and its ability to index and analyze large volumes of unstructured data, he used the solution to tap into the middleware's JMS messaging system—the clearinghouse for every transaction. In his words, "It was just 20 minutes of development work" to write a rule that simply copied those messages into a file.

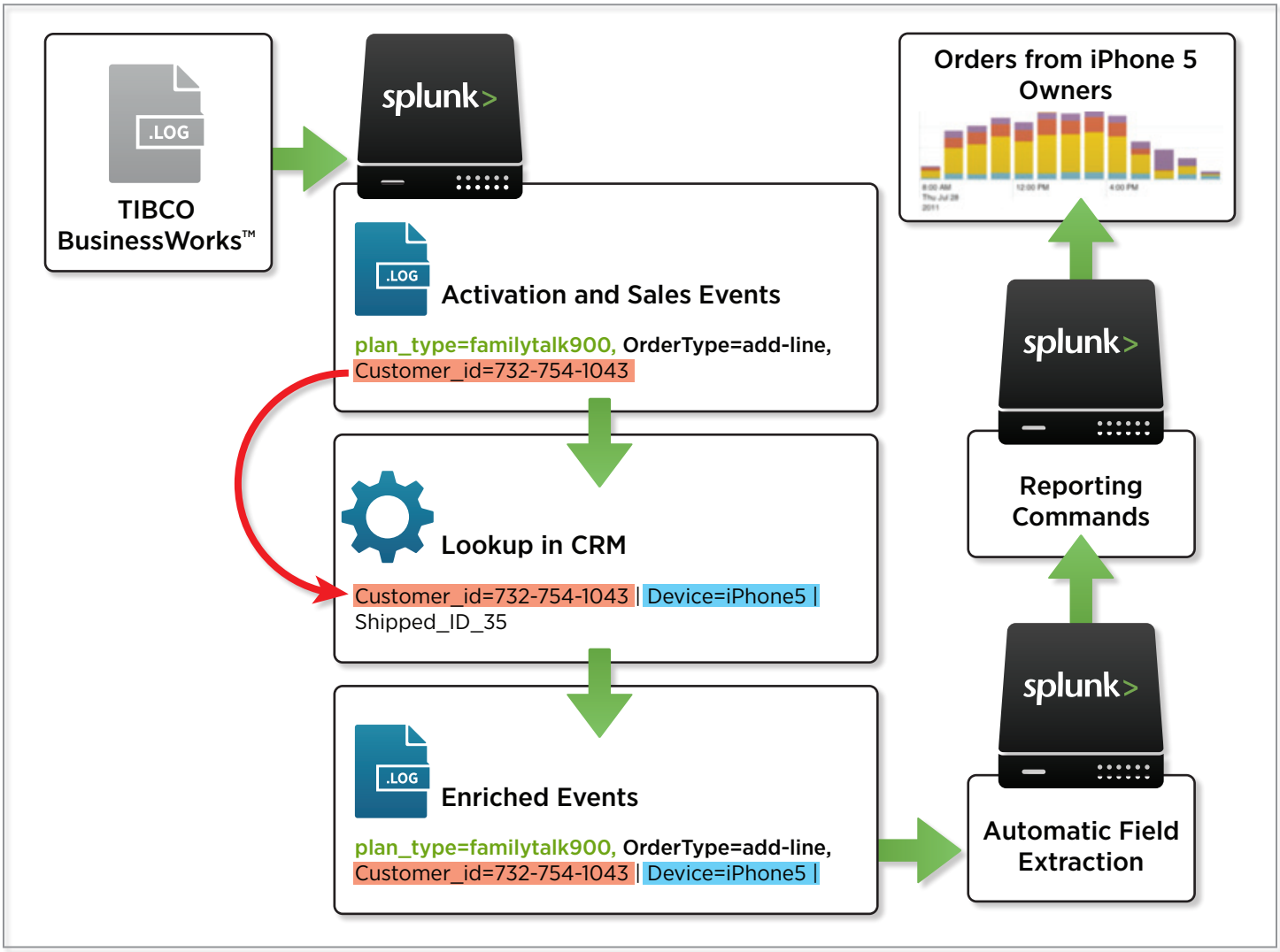

Figure 1. Transaction data's journey from middleware to Splunk Enterprise.

Step #1—Gathering and correlating the data. Splunk Enterprise continually monitored this file and created an event for each transaction as it happened. As shown in Figure 1, these events included details like the customer ID, the date and time of the transaction, and the transaction type. The company further enriched the data by connecting Splunk software to its custom-built CRM system, enabling them to append every event with information on each customer's city, history and device type.

Step #2—Use real-time search to create insight. Once these detailed transactions were available in Splunk Enterprise, the company was able to provide dashboards for the operations, sales, marketing and customer care divisions. Each set of dashboards drew from the same data, but provided each group with an individual lens of relevant metrics.

The company's executives immediately recognized the value of real-time searches in filling a critical gap in the sales and marketing teams' ability to monitor incoming transactions. By utilizing a few simple searches, they created a dashboard that enabled them to view activations by rate plan and average revenue per phone (see Figure 2).

The dashboard was modified to allow the end user to filter the data by network provider and plan type, and to choose the timeframe for the report—any range from a rolling 30-second window to "all time." Thus, users could use the same dashboard to continually monitor sales in real time and generate a weekly or monthly report for executives who prefer high-level views.

The marketing team now has the agility enabled by real-time business intelligence. It can analyze marketing initiatives as they occur, escalating or de-escalating promotions, for example, based on customer usage. Thanks to greater insights, the company enhanced the satisfaction of its customers and reduced churn and the costs of losing and regaining them.

Google Maps integration. While it would have been simple enough for the customer to create a static report listing sales by city and state, this would have lacked the clarity and impact of a dashboard that actually mapped transactions to a location when and where they occurred.

Splunk software's flexible visualization framework allows users to bring in external elements and integrate them with other data collected. By far, the most popular use of this feature is incorporating data with geographic details into a Google Map (see Figure 3).

In this case, the map was fed by a simple search for new activations. The subscriber's zip code had been appended by the CRM at the time the event was indexed, and the latitude and longitude of the zip code were appended with the help of Splunk Enterprise's lookup table feature. By default, the dashboard shows the entire United States, but the user can zoom into a particular region for more granular views (see Figure 4).

Soft ROI: Visibility Into Every Facet of a Modern Enterprise

In many cases, gaps in business insight can be attributed to a technical limitation or a decision made without the right support. Before this customer deployed Splunk software, it was unable to view key performance indicators like revenue and activation count in real time. Its systems lacked the flexibility and functionality to collate and analyze data with anything like real-time agility. Because no tool or platform had been able to deliver that kind of visibility before, the sales and marketing teams were unaware of how useful such optics would be for them and the rest of the organization.

Moreover, this real-time visibility extends to the company's IT

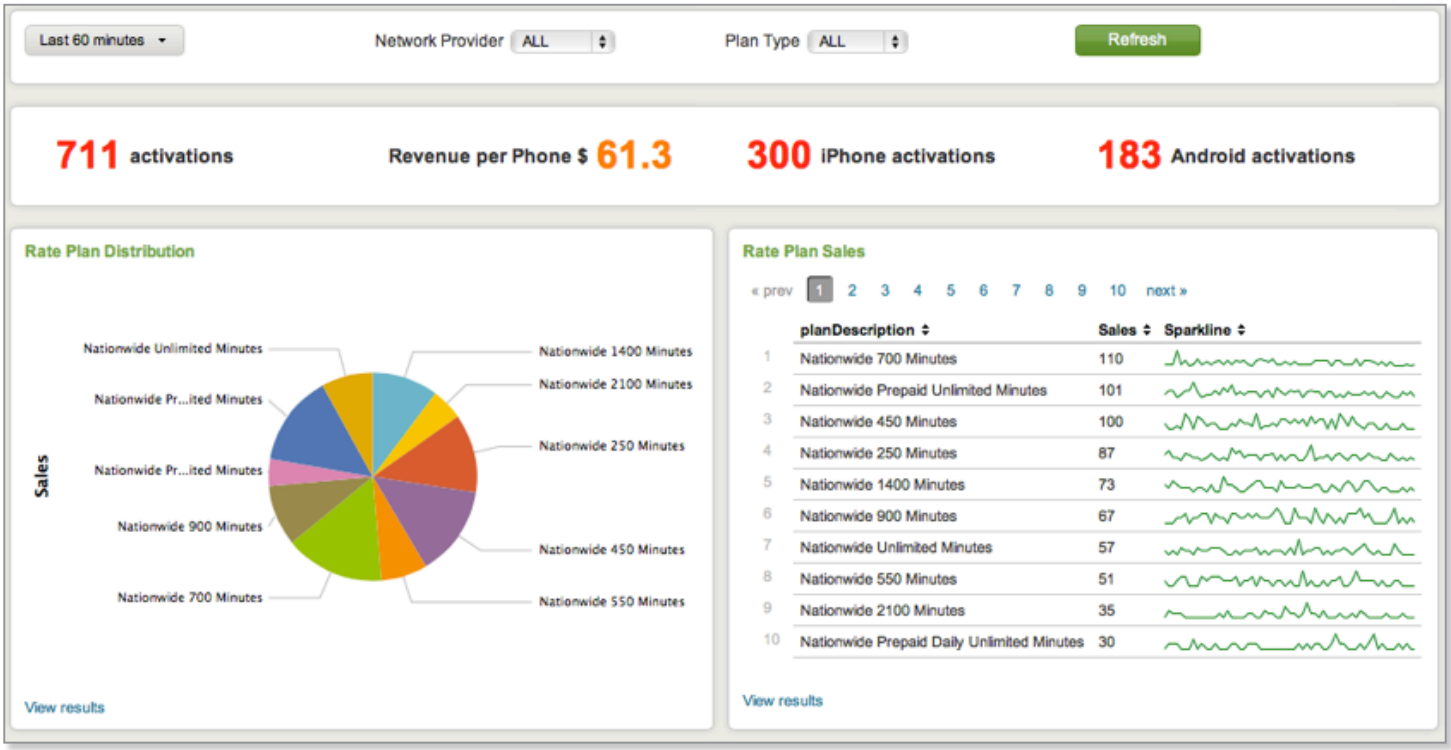

Figure 2. Sales by rate plan for the last hour.

Figure 3. Splunk software and Google Maps integration.

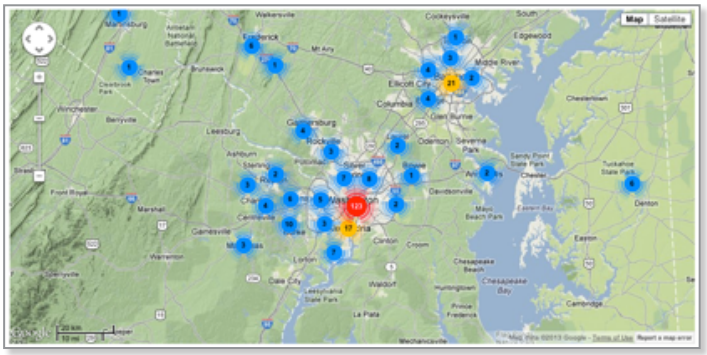

Figure 4. Zoomed region of Google Maps.

infrastructure. By tracking the performance of its systems, staff can address potential outages before they impact operations. As a result, the company reduced downtime by 15%, saving an estimated $1.2M each year.

The effectiveness of real-time search and analysis isn't limited to telecommunications companies. As the customer said in an interview, "[telecommunication companies] have every facet of a modern enterprise. They have a supply chain, they have inventory, they have retail and they have a huge technology infrastructure to run the service they're actually selling."

That's another way of saying that the value of combining Splunk Enterprise's real-time search with enriched mapping and database integration can have a major impact on nearly any type of company.

Using Splunk Software, They Got the Vision They Needed

In this use case, we explored how Splunk software enables operational intelligence for a classic business analytics problem: how to provide the best possible data to executives so they can make informed business decisions? This use case demonstrated:

- **Elimination of data silos.** Because Splunk software indexes all kinds of data, the customer's reliance on multiple reports and formats from varied systems was eliminated and the customer was able to provide a single source of truth to executives.
- **Correlations drive analytics.** Because Splunk software correlates different types of data, the customer could link data from its CRM system to provide information on a granular basis, including visual views on a Google Map.
- **Flexible analytics powered by a read-time schema.** Because Splunk software keeps the data unstructured, the customer doesn't have to break any connections in the data to make it fit in a schema. The customer no longer has to spend significant time or resources on answering follow-up questions for executives.

CHAPTER 3

Determining the Highest Value Leads

How Splunk Software Helps to Effectively Convert Sales Leads

USE CASES

Marketing Analytics & Reporting
Bolstering Business Efficiencies

Executive Summary

When a customer walks into a retail store, there is an immediate opportunity for salespeople to offer service and help drive sales. For an online store, hundreds, if not thousands, of potential customers might come in each hour. Imagine if an online retailer could automatically determine who is interested and who is "just browsing," and dispatch salespeople only to the serious buyers?

One Splunk customer—an online software as a service (SaaS) company—provides cloud-based development, test and training environments for technology customers. Prospective customers sign up to try out the company's cloud software during a free trial period. The sales team then follows up with these leads and attempts to convert them into paying customers. Sales, however, lacked any insight into which were the most promising leads or the most effective marketing campaigns. The company also lacked the optics to prevent some users from misusing the free trial period, which fraudulently consumed its IT resources.

By deploying Splunk® Enterprise, the SaaS provider gained the visibility and metrics to deploy a scoring system to identify and prioritize potential customers, as well as to evaluate the effectiveness of its marketing measures and reduce abuse of its free trials. The company addressed these core challenges:

- **Qualify sales leads from a large base of trial users.** The quality of leads coming from the free trials was poor, leading to inefficient targeting by sales. The sales team was unable to distinguish high-potential leads from unlikely customers, resulting in low sales conversion rates. The company now uses intelligence gleaned from Splunk Enterprise to score and prioritize the most likely customers.
- **Measure the effectiveness of marketing initiatives.** The marketing team was unable to track the results of campaigns or promotions. Using Splunk software, the team saw immediate benefits by identifying the most effective initiatives.
- **Reduce fraud.** Although trial environments are limited to a few testers for a short time, some customers were defrauding the system by conducting lengthy classes for large numbers of attendees, which drove up IT infrastructure expenditures. Splunk software allowed the company to track fraud in real time, reducing inappropriate usage and its costs.

Business Problem: Finding the High Value Customer

This SaaS company provides cloud-based software that is used to conduct virtual training. For example, a typical customer holds 300 virtual training classes for 1,500 students in one year. Learners log into virtual classrooms to receive lectures from live instructors and complete online labs and quizzes. By using this cloud service, customers derive enormous benefit by avoiding the time and expense of managing their own training infrastructure.

Before using Splunk software, the company knew that 10,000 prospective customers were signing up for free trials each month. Typically, each salesperson received 400 or more leads per day. These leads were tracked in a CRM system (see Figure 1), and the company's salespeople followed up on as many prospects as they could via mass emails or cold calls. However, salespeople complained about the quality of the leads. A trial user that never logged into a virtual classroom was represented in the CRM the same way as a trial user that logged into multiple virtual classrooms. Without any information on user engagement, sales had to give all prospects the same time and attention. Worse, many of the leads wasted salespeople's time because the sign-up information provided was spurious.

Business Benefits at a Glance

Splunk Value	How Value Is Measured	Business Impact
Identify high-value sales leads by scoring trial users	• Ability to prioritize leads • Conversion rates improve	• Sales targets the top 20% leads • Improved conversion rate from 2% to 25%
Evaluate the effectiveness of marketing initiatives	• Ability to identify top leads and referral generating campaigns • Better resource allocations	• Marketing targets the sites and campaigns that best generate paying customers
Reduced fraud and abuse	• Detect free trial abusers • Reduced IT costs for fraudulent use	• Alerts when users violate the free trial agreement • Conversion of 50% of fraudulent users to paying customers • Reduced compute costs

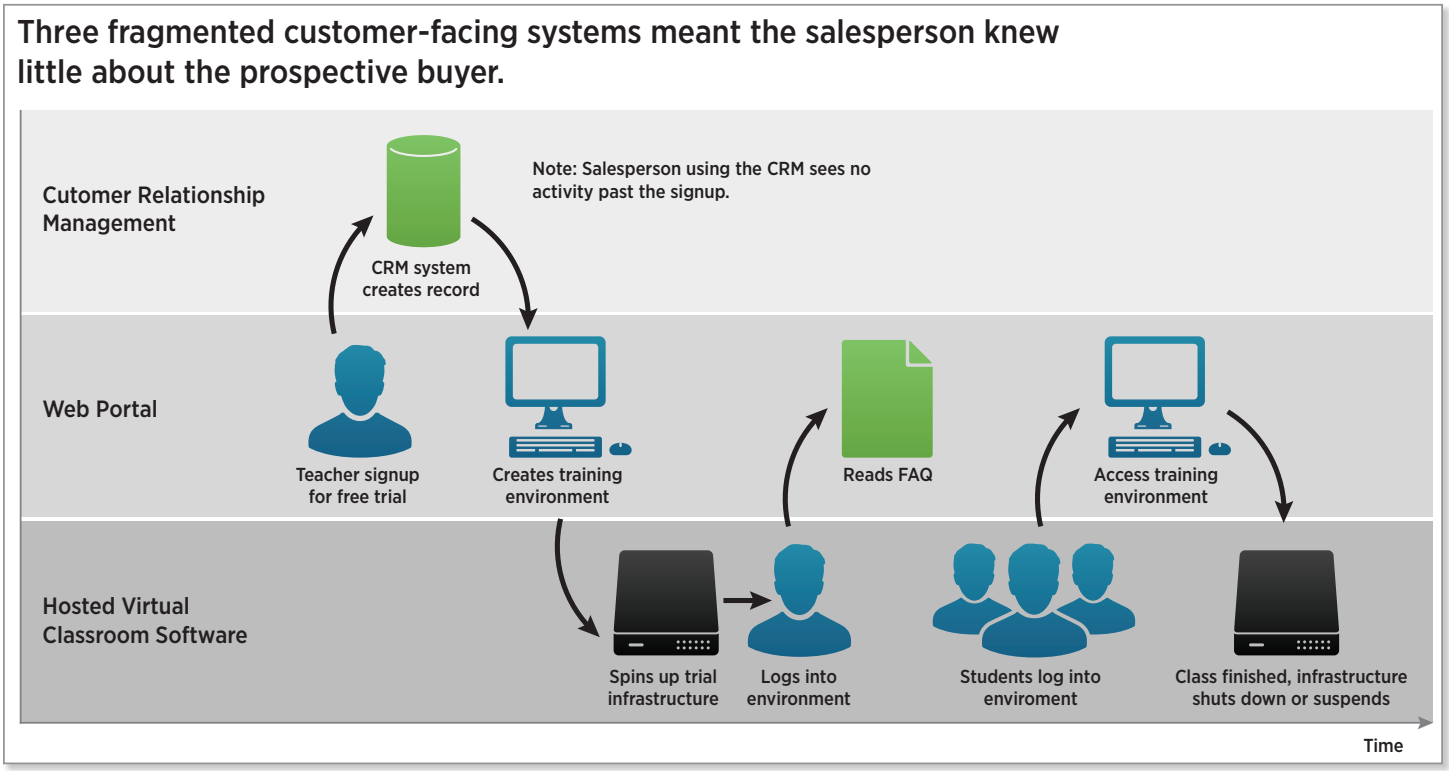

Figure 1. The typical process flow of a trial user.

Due to resource capacity, 10–25% of the leads were going untouched. As much as half of the contacted prospects were unreachable or uninterested. The conversion rate from a trial user to a paying customer was very low, typically around 2%. To increase sales efficiency, the salespeople needed higher quality customer information.

Meanwhile, the overhead cost of the hardware infrastructure was increasing in a way that did not correlate with sales and revenue. Management feared that free trials were unsustainable as a lead-generation engine because of the growing cost of hosting instances. To counteract this, they added an auto-suspend to environments that were unused for two hours, but suspected some users of scripting around this.

Management came to suspect that fraud and abuse of the system contributed to the disconnect between free trial users and subsequent sales. A number of users were gaming the free trial. In fact, about a dozen paying customers admitted to first using the free trial in a full production capacity.

Enter Splunk

Qualifying the most valuable sales leads

The company was already using various marketing automation products to track and value its leads from marketing campaigns across multiple online and in-person channels. However, it was hard for these tools to capture behavior and data at a granular level from online trials. These resources also required changes to IT systems and data output formats. Lastly, they were expensive. Unsure how valuable the leads from free trials truly were, the company was unwilling to invest further in expensive, dedicated marketing tools.

The company had already deployed Splunk Enterprise to help monitor its IT operations. It uses the solution to capture, index and integrate log event data from its infrastructure's systems and applications and then to graphically display all kinds of performance metrics in dashboards. Management realized that with some simple but creative extension, Splunk Enterprise could help solve the challenge of finding the most valuable sales leads.

To improve insights into trial users, the company needed more information on how they interacted with their web portal. As the first touchpoint between a potential customer and the company, the web portal allows a visitor to register for a free trial, process payments, read FAQs and access other support content. Trial customers can then launch and access virtual trial classroom environments via the web portal. All of these activities leave a trail of log event (machine) data (see Figure 2).

```
FirstName=John
FirstSessionLength=0
HasCampaigns=False
HasOpenCancelRequest=False
IsBlocked=False
IsCollaborator=False
IsEmailVerified=False
IsEntApp=False
IsPhoneNumberVerified=False
IsPro=True
IsRoot=False
IsTrusted=False
JobTitle=None
LastLoginTime="2013-01-24 00:24:22.637000
LastName=Doe
LoginCount=1
LoginCount30Days=1
MainUsage=Windows
MarketingId=None
Name="John Doe"
NextBillingDate=None
NextPaymentDate=None
NumEnvsAsCollaborator=0
```

Figure 2. Sample raw log data from web activity.

Raw data on trial customer activity levels is converted into a table

	Actor	AppUser	TemplateType	FirstSessionLength	TotalSessionsLength	ActionsInFirstSession	AllActions	NumSessions
1	07c31a0516@gmail.com	Lost	Sharepoint	11	164	Create Login	Create Login Logout Resume View	8
2	12345@abcde.com	Lost	Sharepoint	25	70	Create Login View	Create Login Resume View	6
3	17abhish@gmail.com	Lost	Sharepoint	31	1224	Login View	Create Login Logout Resume	14

Sample considerations contributing to the MQL score:

Is the area code of a trial user's phone number from a blacklisted country?

summary index - MQLPoints from the country phone number

Search *

```
| inputlookup users | search ProLevel="Trial" | fields User ProLevel PhoneNumber | phone2country PhoneNumber | fields
PhoneNumber_country User | eval
MQLPoints=case(PhoneNumber_country="India",-10,PhoneNumber_country="Brazil",-10,PhoneNumber_country="Iran",-20,PhoneNumber_
country="Egypt",-20,PhoneNumber_country="Turkey",-10,1=1,0) | where MQLPoints!=0 | table User MQLPoints
```

Is the user's virtual classroom environment receiving a lot of bad Internet traffic?

summary index - MQLPoints from networking

Search *

```
| set union [search index="ranking-summary" source="block_udp" | stats max(Rank) as Rank by EnvOwner Category]
 [ search index="ranking-summary" source="block_drop" | stats max(Rank) as Rank by EnvOwner Category] | stats sum(Rank) as
Rank by EnvOwner | rename EnvOwner as User | eval MQLPoints=-2*Rank
```

How did the user get to the website?

summary index - MQLPoints from referring sites

Search *

```
|inputlookup marketing_data |  eval landingPageRefferer=LandingPageRefferer | `normalize_marketing_data` | lookup referring-
sites landingPageRefferer as landingPageRefferer | where not isnull(MQLPoints) |  table User MQLPoints
```

Does the email domain match the company name? Have they logged in at least twice? What type of virtual classroom template did they spin up or log into? Do they use a webmail domain, e.g., Gmail?

summary index - MQLPoints from miscellaneous parameters

Search *

```
| inputlookup users
| eval
TemplatePoints=case(TemplateType="Sharepoint",10,TemplateType="Desktop",-2,TemplateType="Server",2,like(TemplateType,"%Linux%"),-4,1=
1,0)
| eval MainUsagePoints=case(MainUsage="Sharepoint",5,MainUsage="Windows",-2,MainUsage="Linux",-5,1=1,0)
| eval TrustedPoints=if(IsTrusted="True",15,0)
| eval LoginsPoints=if(LoginCount>2,4,0)
| rex field=User "@(?<Domain>.*)" | eval Domain=lower(Domain) | lookup webmail Domain as Domain | fillnull UseWebMail
| eval WebMailPoints=if(UseWebMail=0,2,0)
| eval MQLPoints=TemplatePoints+MainUsagePoints+TrustedPoints+LoginsPoints+WebMailPoints
| table User MQLPoints TemplatePoints MainUsagePoints TrustedPoints LoginsPoints WebMailPoints
```

Figure 3. Calculating the MQL with Splunk software.

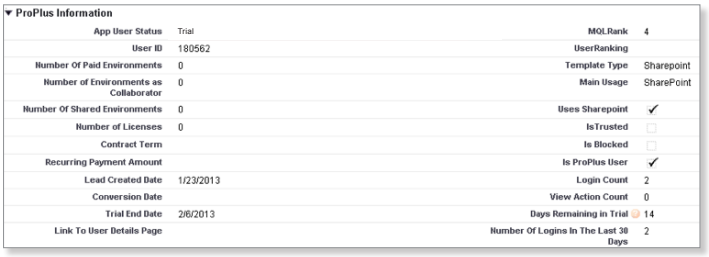

Figure 4. Profile and MQL Score for a trial user.

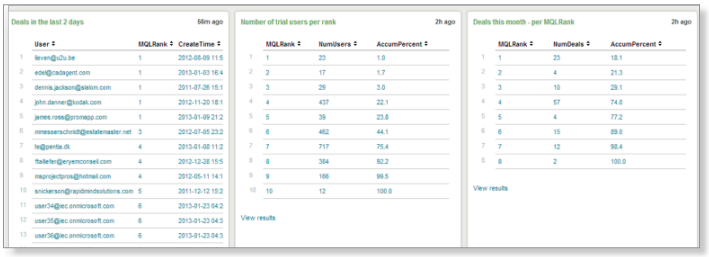

Figure 5. Prioritized customer leads, according to MQL.

After converting the raw data into easily parseable tables using Splunk Enterprise, the customer then uses the software's analytics capabilities to calculate an MQL (Marketing Quality Lead) score for each trial user (see Figure 3). The score considers such factors as the validity of the user's phone number and the amount of time spent on the trial website. An MQL score of 1 suggests a user with a high likelihood of purchase, while an MQL of 7 means the user is unlikely to become a paying customer. Figure 4 shows a sample profile and MQL score developed.

The company now applies this scoring methodology to all of its trial customers, using Splunk software to quickly and easily build dynamic tables and charts. The result is a collection of dashboards and leads lists for its salespeople. The ranked list of trial customers by MQL is uploaded to the CRM system, enabling the sales team to prioritize high-potential leads. Figure 5 shows a prioritized sales leads list, while Figure 6 is an executive dashboard reflecting new trial customers and how they distribute among the MQL scoring range of 1–7.

Using Splunk software, this SaaS company now makes sense of its large trial user-base and identifies the most valuable users—those who are likely to become paying customers. The company has concluded that a trial user with MQL of 1 has a 50% chance of becoming a paying user. By focusing on trial users with MQL scores equaling 1 or 2, the sales team dramatically improved its overall leads-to-sales conversion rate from 2% to 25%. The lower scoring leads—those with MQL from 3 through 7—are contacted via automated campaigns, saving valuable time and effort.

Gauging the effectiveness of marketing campaigns

To further improve its efficiencies, the SaaS company uses Splunk software to determine the effectiveness of its online marketing campaigns, which comprise search engine advertisements, direct outreach to current and potential customers, and referrals through partner websites. By analyzing application and web data with Splunk software, the company now identifies the most successful marketing campaigns by searching for such items as referring codes and how many referred customers become paying customers. Figure 7 shows the marketing campaign conversion dashboard within the company's Splunk Enterprise deployment.

Detecting fraudulent use of free trials

The company identified another issue that could be rectified with better intelligence. Prior to the deployment of Splunk software, fraud contributed to a dramatic increase in hardware infrastructure costs. Trial users perpetrated fraud in several ways:

- Keeping the virtual class open beyond the allowed trial duration
- Inviting more than the maximum number of allowed trial guests to a class by sharing permalinks
- Opening multiple virtual classrooms simultaneously

Many of these practices were hard to detect in real time, and all of them consumed costly resources by imposing a disproportionate load onto the company's IT infrastructure.

Again, Splunk software provided a solution. Splunk Enterprise monitors various aspects of the user trial system through web and application logs. When it detects unusual patterns, either known or unknown, it sends an alert to sales. These alerts have actionable information on the users who are logged in and the classes they are conducting or attending. Because the most effective fraud prevention and sales conversion happens if the salesperson calls while the violation is in progress, this real-time ability greatly reduces fraud while increasing sales. In fact, some 50% of fraudulent users are being converted to paying customers. Moreover, without inappropriate usage, the company's compute resources are now dedicated to serving customers and potential customers. Figure 8 shows the abuse detection dashboard.

Insight

This SaaS company delved into its machine data with Splunk Enterprise to empower its salespeople with insights into prospective buyers, allowing sales to be more proactive and efficient. By using the Splunk platform to address three main challenges—qualifying leads, assessing marketing campaigns, and identifying and reducing fraud—the company improves operational efficiencies and resource allocations. It leverages the data its infrastructure generates to obtain actionable intelligence with which to effectively manage its business processes without the expense of numerous diagnostic tools or outside consultants.

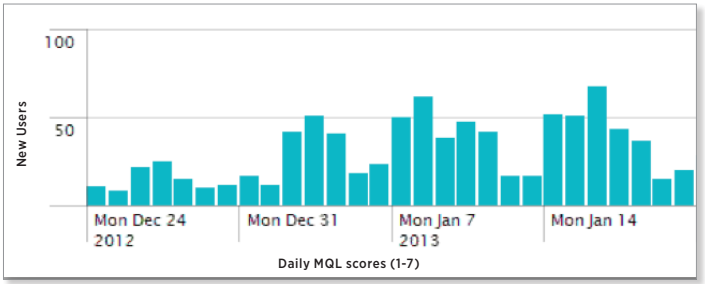

Figure 6. Distribution of MQL scores of new trial users across four Mondays.

With Splunk, They Took the Lead

In this use case, we explored how Splunk software can provide a solution for a classic sales and operational challenge: "How do I qualify my leads into high-value paying customers?" This use case demonstrates:

- **Elimination of data silos.** Because Splunk software indexes all kinds of data, the customer ended its reliance on multiple reports and formats from varied systems and has gained operational insight for sales and marketing.
- **Lead generation analytics.** By having Splunk Enterprise correlate both structured and unstructured data, the customer can link data from its CRM system to calculate an MQL score for each trial user.
- **Flexible analytics powered by a read-time schema.** Because Splunk software collects data in full fidelity without filtering, the customer avoids losing any potential value in the data just to make the information fit in a schema. This means the customer can engage in on-the-fly analysis to determine the success of its marketing campaigns.
- **Value generation.** In addition to gaining insights into prospects to prioritize the most promising, the customer reduces costly fraud and converts many of these users into paying customers.

The bottom line is that without undue cost or effort, the company gains optics into its core revenue-generating processes to bolster its sales, competitiveness and success.

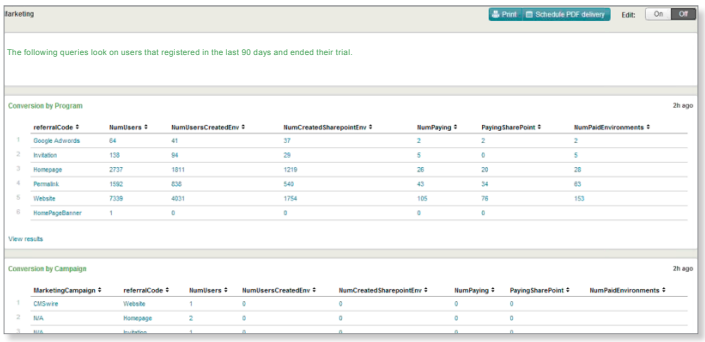

Figure 7. Leads conversion by marketing campaigns.

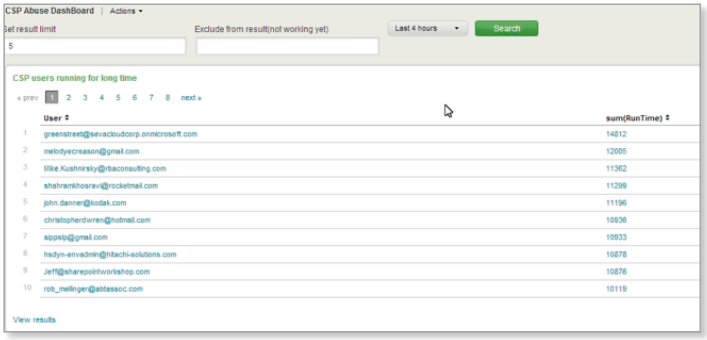

Figure 8. Catching trial users who abuse the system.

CHAPTER 4

Automating Healthcare Claim Processing

How Splunk Software Helps to Manage and Control Both Processes and Costs

USE CASES

Troubleshooting Services Delivery
Streamlining Internal Processes

Executive Summary

As the pressure on healthcare practitioners and insurers to streamline their operations rises, many are strategically deploying Electronic Data Interchange (EDI) claim submission solutions to automate and accelerate claims processing. With an EDI solution, healthcare insurance companies spend less time figuring out what is covered and what is not, providers get paid faster and the cost of claims processing decreases for insurers.

But what if the automated claims process is not up to speed? What if it is failing?

One major US healthcare insurer confronted this exact problem. The company's EDI solution had a very high error rate, forcing the insurer to spend more time and money each quarter to manually re-process claims. IT staff also suffered from limited visibility into the insurer's infrastructure, causing the enterprise to spend even more time and money to troubleshoot the system's performance problems. This scenario stood in stark contrast to the cost-saving efficiencies promised by EDI automation.

To rectify this situation, the insurer turned to Splunk® Enterprise. Using the solution, the insurer gained an end-to-end view of its entire EDI claims processing chain, enabling its IT staff to quickly pinpoint and remediate system errors and bottlenecks. The company is gaining the visibility to identify improperly submitted claims by partners and other providers, reducing error rates further. As an added benefit, Splunk software is streamlining system reporting, web analytics and regulatory compliance, presenting an opportunity to retire costly third-party reporting tools.

With the Splunk platform capturing, indexing and displaying data on its claims processing infrastructure, the insurer will finally realize the cost-efficiencies and elevated service levels promised by EDI.

- **Reduced submission error rates.** Automated claims processing routinely failed to meet the insurer's "first pass" quality goals, requiring expensive and laborious re-processing. Using Splunk Enterprise, the company captured and indexed the logs for all systems involved in claims processing. This allowed IT staff to correlate data across various applications with requisite information, such as sessionID and userID, providing granular optics into the infrastructure. The IT team is now able to rapidly identify and correct the source of errors and reduce the rate of improperly processed claims, saving the insurer substantial costs.
- **Realized cost savings and efficiencies.** The insurer previously relied on a hodgepodge of costly reporting tools that offered limited insight into its EDI environment. Performance and processing issues persisted and the company's IT teams spent 17-plus hours per day troubleshooting and producing mandatory reports. With the introduction of the Splunk platform, the teams created

Business Benefits at a Glance

Splunk Value	How Value Is Measured	Business Impact
Insight into issues in the automated claim review process	• Number of claims correctly processed per day/month/year • Greatly reduced costs for manual re-processing	• 29% potential reduction in failed claims • Expected to eliminate 200K hours of manual review, the equivalent of over 100 full-time employees • Estimated $14 million of savings in a single year
Easy correlation and real-time monitoring of events across all layers and components of the claim processing chain	• Reduced hours spent identifying and addressing performance issues • Adherence to uptime and performance service levels	• Over $650,000 per year expected savings through improved troubleshooting • "First pass" performance objectives will be quickly met
Rapid and automated production of regulatory compliance verification reports	• Reduction in hours spent manually generating reports • Improved accuracy of reports • Decreased spending on unsatisfactory analytics tools	• Eliminating 17-plus hours per day generating reports, leading to more than $300,000 in savings • Reports are timely and more precise • Projected savings of thousands of dollars per year in licensing fees for analytics tools

dashboards to continuously monitor the operations of all systems. For the first time, they can measure performance levels, expedite claims processing and reduce manual re-processing costs.

- **Verify compliance reporting.** With sparse views into its application and network components, the insurer found that the creation of regulatory compliance reports was incredibly time consuming. Capitalizing on Splunk software, the IT team is building dashboards and alerts to enable real-time compliance verification and reporting across disparate enterprise platforms and applications. They can even automatically generate and deliver reports as PDFs.

The Mystery of Failed Claims Processing

Like many healthcare companies, this large insurance company invested substantial time and money to automate claim processing in order to make its operations faster and more efficient. It built a complex chain of components capable of evaluating claims, processing their data and accurately making accept/decline decisions. The system accepts submissions from healthcare providers via the web and other sources (fax, email), and sends them through the various claims review and approval applications based on their content. It then returns the approval/denial of coverage information back to the providers or aggregators as quickly as possible.

If something is wrong with the submission, the initial processing of the claim fails, which requires an analyst to review the claim and determine whether it should be accepted or rejected. Unfortunately, due to the intricacies of the system's many interlocking components (see Figure 1), the rate of electronically submitted claims that failed on the first pass was 14 percent or almost 1.5 million failures a year. Each failed claim demanded a half hour to manually re-process, which totaled over 730,000 hours of manual labor each year spent on investigation and resolution—and that was just for the initial claims submission acceptance.

In a complex environment, IT/Operations staff and system analysts can spend as much time trying to determine the cause of performance and quality problems as it took to deploy the systems. Figure 2 illustrates the complexity of the insurer's monitoring architecture. Its application integration, web platform, WebSphere MQ/Message Broker and EDI processing support personnel each had areas of responsibility, but with no end-to-end visibility into the claims processing workflow and a cumbersome reporting infrastructure, they lacked the ability to assess the root cause of the performance issues. Troubleshooting problems took too long, resulting in failures to meet targeted services levels. As a result, resources continued to be squandered on manually assessing failed claims.

Evaluating the performance of the business-critical TriZetto Facets application, for example, was impossible with the company's existing tools. Because the application traverses many layers of infrastructure and the network, support staff were unable to determine the cause of claims processing problems. Without this knowledge, senior management could not understand key issues and, therefore, could not improve overall claims throughput.

Moreover, existing reporting systems were unwieldy and fragile. Analysts spent far too much time gathering and manipulating data by hand. Reports took too long to generate and were of limited value because the tools were unable to fully correlate customer experience and other metrics. The bottom line was the company's business model was compromised. The situation looked bleak.

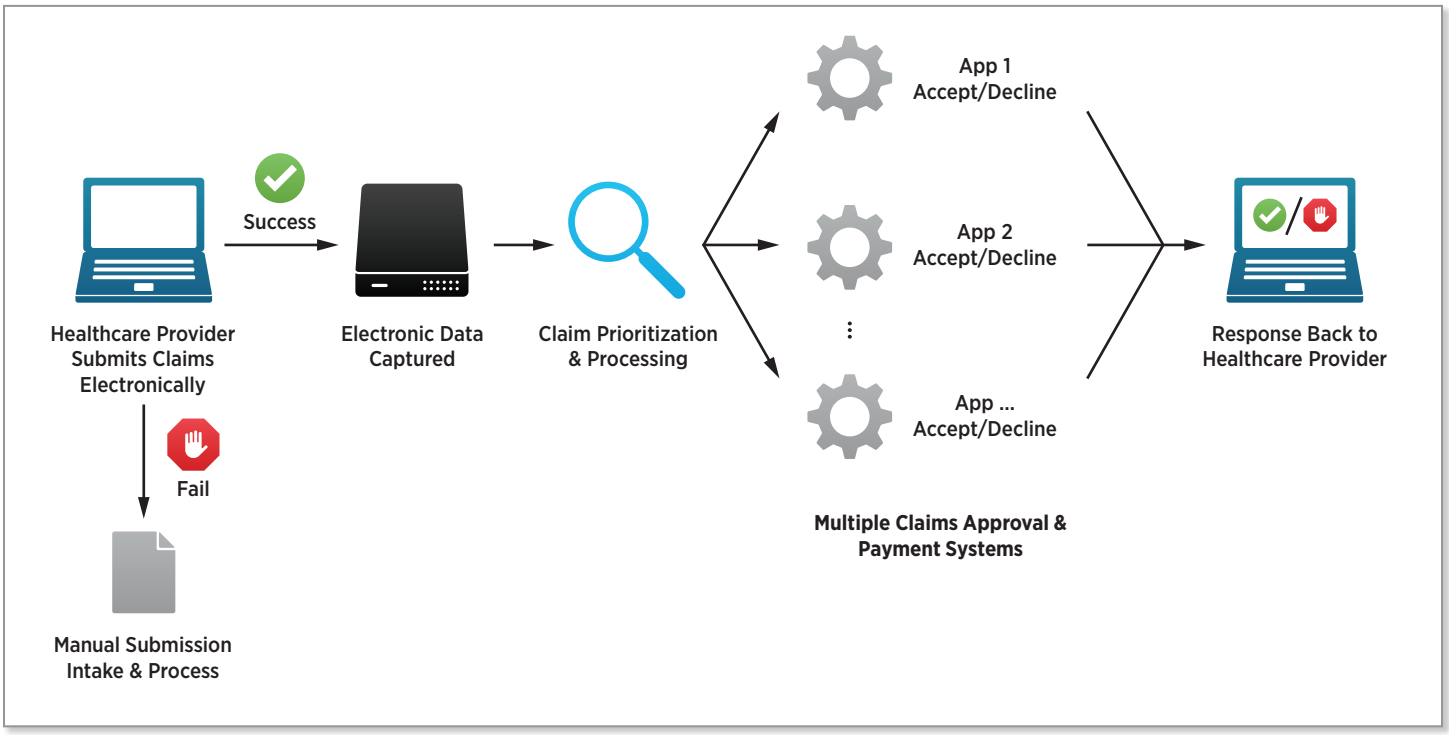

Figure 1. The flow of claims processing, including manual re-processing of failed submissions.

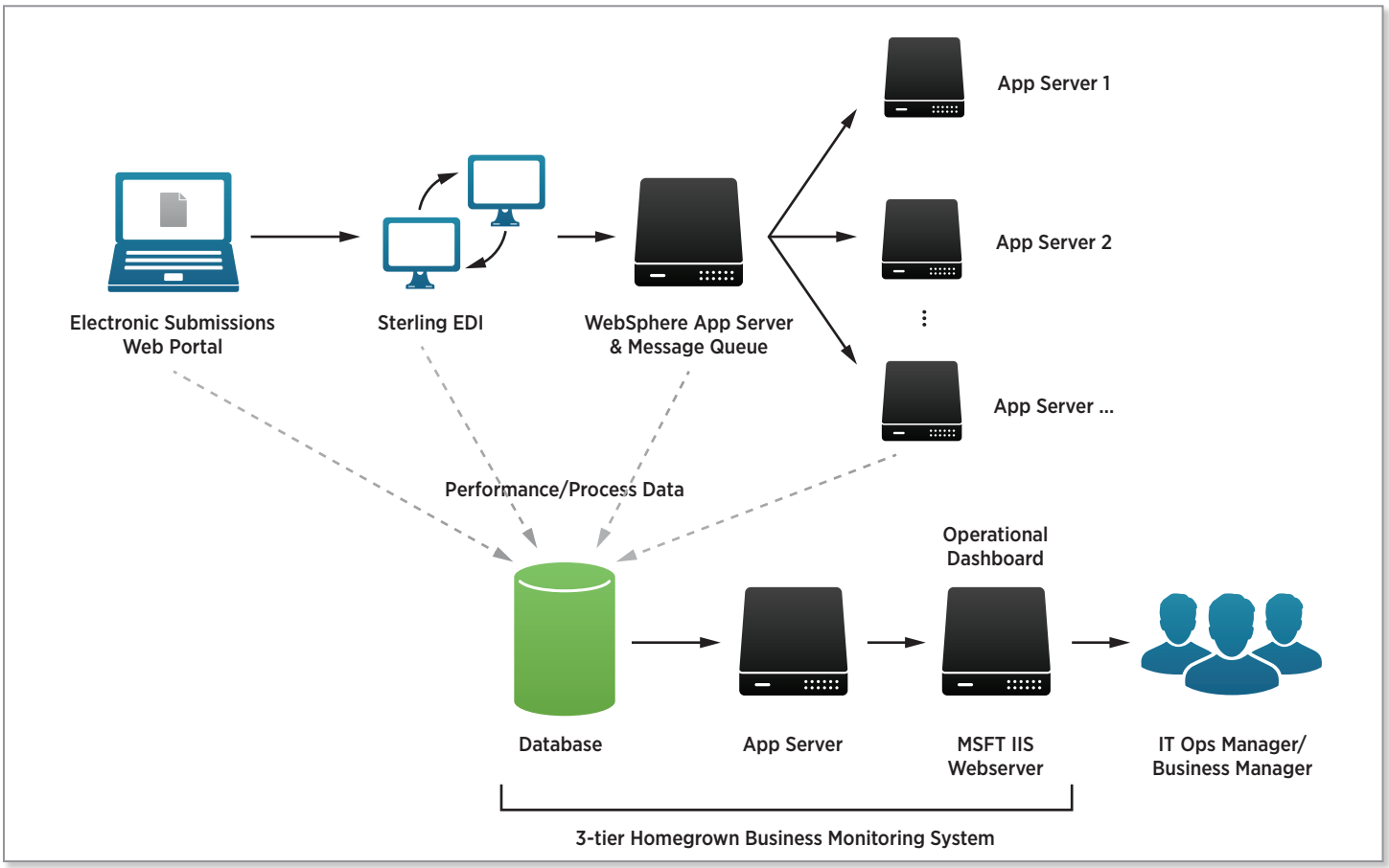

Figure 2. The complex monitoring and reporting system.

Enter Splunk

The Splunk team approached the insurer to demonstrate the ability of Splunk Enterprise to capture, index, collate, and visualize logs and data generated by the many components of the EDI claims processing chain. The engineers set up the Splunk software in a test environment and within just one week, simply used WebEx meetings to show each developer and IT staff member how to query the solution to obtain precisely the data they needed. The IT staff was particularly impressed by Splunk's ability to access WebSphere MQ data and how the Splunk DB Connect application riding on top of the platform can retrieve and integrate structured data from databases.

In the end, the insurer's own IT team, rather than Splunk representatives, presented the Splunk solution to upper management.

The insurer's staff funneled the following data sources into Splunk Enterprise and correlated the data across applications by using fields such as sessionID and userID:

- WebSphere Application Server
- MS Internet Information Server (IIS)
- WebSphere MQ administrative logs
- WebSphere Queue manager analytics
- Sterling File Gateway system logs
- Sybase Facets performance analytics.

Once data collection was set up and the Splunk platform was extracting meaningful data from the EDI systems, the company's various IT teams created a number of dashboards so they and business users could graphically view, in real time, the number and kinds of claims going through the system.

Viewing Every Link in the Chain From Start to Finish

The Splunk Enterprise implementation immediately afforded the insurer visibility into its malfunctioning claims processing, as well as comprehensive views into its entire infrastructure, allowing IT teams to identify the sources of errors and quickly resolve them. Figure 3, for example, shows the traffic of EDI claims into the Sterling File Gateway broken down by provider/aggregator, as well as failures to route claims correctly within the WebSphere application servers by source/provider.

This and other insights allow the company to identify the primary sources of misfiled EDI claims, which in turn enables staff to rapidly address their root causes. Optics into the failures by source also permit the insurer to target partners and other providers with high error rates for additional training on properly submitting claims.

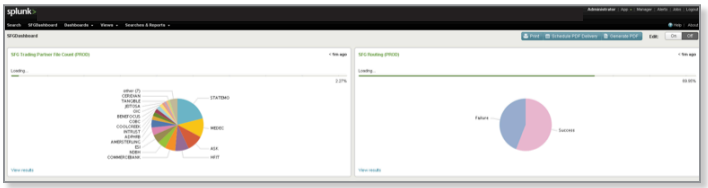

Figure 3. EDI claims and the failure rate of routed claims within WebSphere app servers.

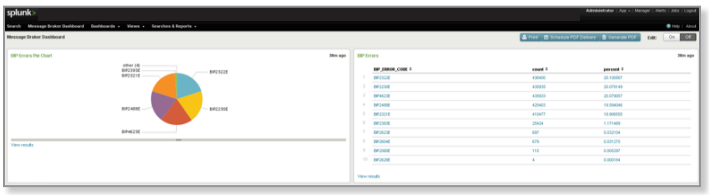

Figure 4. The Message Broker dashboard showing the WebSphere MQ performance by queue and broker.

In another example, staff used two types of logs for the WebSphere MQ monitoring and troubleshooting dashboards:

- Administrative logs: number of elements in queue, depth of queue, queue performance
- Manager logs: connections into the queue, clients connecting to this queue server, errors in connections

The Message Broker dashboard (see Figure 4) shows WebSphere MQ performance by queue and broker, with the WebSphere Message Broker error messages, known as BIP errors, representing the brokers' aggregate errors. BIP errors mean that other apps in the workflow cannot get their events into the queues, impacting performance of the overall system. IT investigators and application developers now can drill down into a given failure to quickly identify its cause.

Figure 5. The performance characteristics of each queue.

The MQ dashboard (see Figure 5) shows the performance characteristics of the queues. Each line is the number of events in each stage of the processing process.

The web analytics dashboard (see Figure 6) shows all employer, broker and provider claims submission activity at a glance, as well as information about how they access the site. This provides awareness of resource usage by submitter and enables staff to prioritize which platforms need improved user interfaces to simplify deployment of their portal and minimize errors.

The WebSphere dashboard (see Figure 7) covers network connectivity and performance. The first panel shows the connection establishment times for providers connecting into the EDI processing queue. The second displays connectivity between all the various Java apps that make up the EDI processing queue. When a queue begins to back up, alerts are issued and the appropriate application support team is notified to address the issue.

Ensuring Investments in Automation Pay Off

Investing in EDI automation is a wise long-term investment, but it is not enough to build it and hope "they will come." These processes are complex and require a high level of visibility to troubleshoot and manage their components. As this healthcare insurance company uses Splunk software, the visibility delivered by the platform exposes opportunities that will save millions of

Figure 6. The panels of the web analytics dashboard.

Figure 7. The panels of the WebSphere dashboard.

dollars a year in personnel-hours spent trying to identify and resolve the root causes of processing errors and freeing those analysts, IT investigators and application developers to develop better functionality for the company's customers.

The insurer expects to reduce failed claims by 29 percent and to meet its "first pass" performance objectives. As a result, the company is on its way to saving over 210,000 hours of manual labor and $14 million in cost annually. Moreover, its newfound ability to quickly identify root causes of errors is projected to save $650,000 per year in troubleshooting costs.

In addition to the greatly improved submission success rates, the Splunk solution also simplified the company's internal performance and web analytics reporting as well as external compliance reporting, allowing it to begin phasing out third-party reporting tools. This will save thousands of dollars each year in licensing fees.

With Splunk, They Got Up To Speed

This use case illustrates how Splunk software can improve operational efficiencies, particularly in distributed transactional environments that deliver mission-critical business processes. Although the customer had a sophisticated, highly complex infrastructure of interlocking applications and systems, Splunk Enterprise provided the optics to manage and control both processes and costs.

- **Elimination of data silos.** By indexing a wide variety of data, both structured and unstructured, the Splunk platform provides the customer's many IT teams with holistic views across the diverse components and application stacks comprising the environment.
- **Correlations drive analytics.** Splunk software correlates different types of data to enable the insurer to link error messages in one system's logs to useful evidence in other data to resolve issues.
- **Insight into complex business processes.** Splunk software's enterprise-wide visibility offers in-depth reporting and effective compliance verification to meet governance and regulatory demands.
- **Operational intelligence.** Splunk Enterprise delivers insights not only into the performance of business processes, but also into the health and availability of the underlying IT infrastructure supporting them.
- **Knowledge and control.** This use case illustrates how Splunk software can offer knowledge into business operations. With this knowledge comes the control to achieve objectives while ensuring efficiencies and service levels.

CHAPTER 5

Finding Order(s) in the Chaos

How Splunk Software is Used to Troubleshoot Transactions

USE CASES

Troubleshooting Services Delivery
Streamlining Internal Processes

Executive Summary

From the simple storefronts and shopping carts of the 1990s, e-commerce has grown into complicated supply chains and partner networks. The capturing, processing and fulfilling of orders has also grown, presenting a unique set of challenges. Losing orders within a complex network of internal and/or partner systems is one of the biggest nightmares for online retailers. Orders that fall between the cracks not only directly impact the revenue stream, they also damage customer satisfaction, the company's reputation and long-term relationships with consumers.

With a complex stack of dozens of different software components handling up to four million online orders per hour, one Splunk customer, a large online retailer, found that Splunk® Enterprise provided the visibility into its environment that it required to catch issues with orders before they could affect customers. Founded in the 1980s, this rapidly expanding retailer has $25B in annual revenue and more than 90,000 employees in over 25 countries.

Upon deploying the Splunk platform, the customer could visualize its online processes in dashboards for end-to-end visibility into order transactions across the enterprise. Specifically, the retailer defines transactions in Splunk software to monitor customer orders as they travel through the IT data integration backbone, a complex environment which consists of multiple, disparate applications. The retailer also gained visibility into its systems to better understand usage and trends and to enable its support and operation teams to detect problems before they impact customers. As a result, the retailer met its business-critical needs:

- **Seeing beyond the complexity of the software stack.** The retailer's multiple, custom-built internal- and external-facing ERP applications do not use a single unique identifier for each transaction to correlate events from one system to another. Consequently, debugging issues that impeded transactions was extremely challenging. The support team learned of problems only when customers called to report unfulfilled orders. Splunk Enterprise now provides comprehensive optics across all systems to track transactions and identify issues.
- **Simplify the troubleshooting process.** Once an issue was reported, multiple members of the support and operations staff were frequently needed to identify and resolve the problem. This involved tracking, tracing and correlation of data manually through millions of lines of inconsistently formatted log entries, and trying to link unique identifiers from each individual system and silo. Tickets were passed from group to group while customers awaited resolution. Now that Splunk software coheres log events to display both holistic and granular views of the transaction chain, support staff can address issues rapidly and cost-effectively.
- **Improve the customer experience.** Company executives had been concerned about customer satisfaction issues. Support staff was always in triage mode, fighting fires and never getting ahead of the problem. Now with greater insight into its transaction processing, the retailer has improved the efficiencies of its online business to better meet customer expectations.

The Nightmare of Online Retailing—Orders Lost in the System

To support a business that has grown over the years to millions of transactions per hour, this online retailer built a multi-step, loosely decoupled ordering process and a corresponding IT architecture to underpin it (see Figures 1 & 2).

Business Benefits at a Glance

Splunk Value	How Value Is Measured	Business Impact
End-to-end visibility on customer orders	• Decrease in lost orders • Faster resolution when customers inquire about orders	• $900,000 annually in recovered revenue by preventing lost orders • Higher customer satisfaction and loyalty
Integrated, yet granular views across siloed systems	• Reduced Mean Time to Repair (MTTR) • Reduced SLA violations for order processing and fulfillment	• More efficient order fulfillment • Issue analysis and resolution accelerated by 25% • Reduction of lost revenue
High business value, low IT pain	• No IT systems require re-architecting	• Rapid ROI via operational efficiencies, recovered revenue and a renewed focus on high value projects

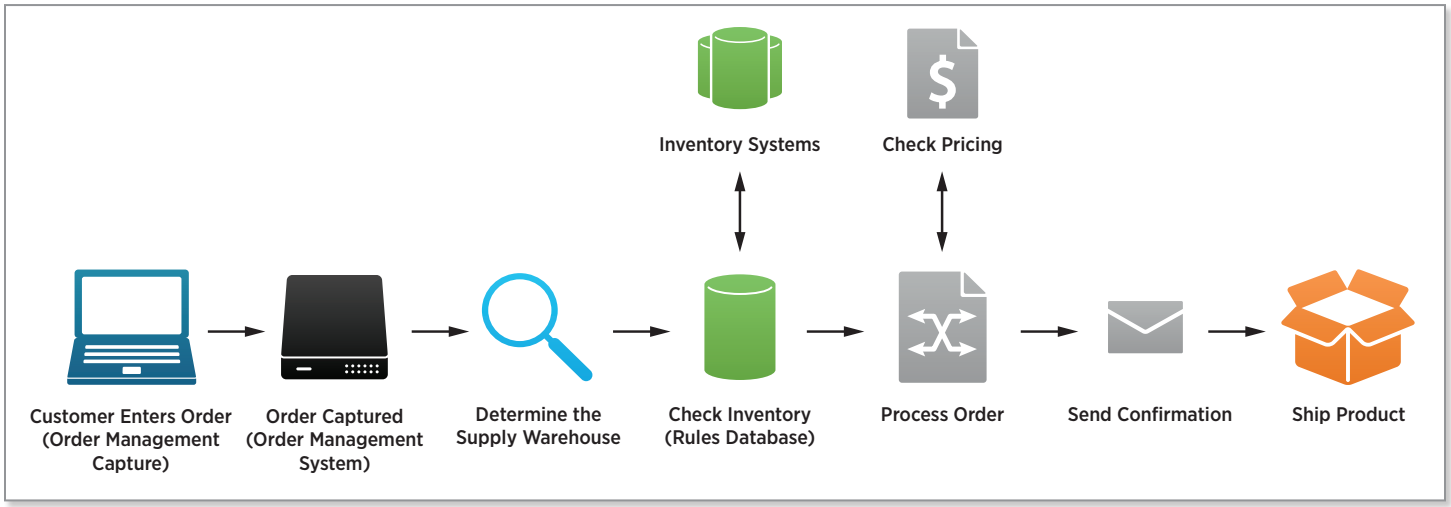

Figure 1. Multi-step business process and systems.

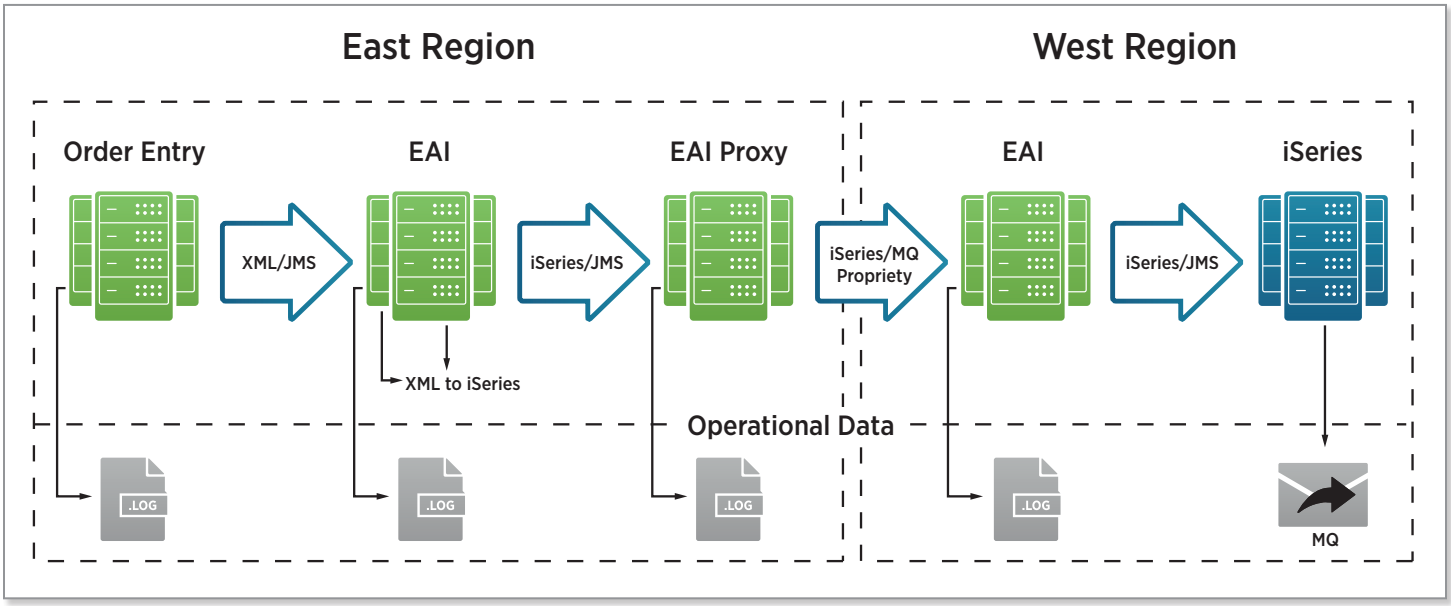

Figure 2. The high-level IT architecture supporting the ordering process.

Whenever orders went missing, troubleshooting issues through such a complex stack was a laborious task that required extensive IT resources over long periods of time. Since the IT systems involved—SOA, EAI-connected and platform-based services—were loosely integrated and the underlying data generated were siloed, tracing a particular customer's transaction end-to-end was nearly impossible. Multiple applications with common fields but different identifiers made correlating data extremely difficult, which complicated matters further. As a result, issue tracking and resolution often took weeks or longer, which undermined productivity and customer satisfaction.

Enter Splunk

One reason the retailer sought out Splunk Enterprise as a solution was its desire to minimally impact its loosely coupled IT architecture and its complex, siloed data systems, which had taken years to build. Starting over to create a more tightly integrated system that could easily provide an end-to-end view of business transactions simply was not feasible. Instead, Splunk Enterprise allowed the retailer to leave the existing architecture unchanged while attacking the problem "bottom up" by aggregating, indexing and analyzing machine data across the entire environment (see Figure 3).

For a view of what this data can look like, Figure 4 is a screenshot from the retailer's key process orchestration system, the webMethods Service Bus. The top and bottom left are sample webMethods data and fields. The bottom right shows how the Splunk REGEX command pulls out special fields such as **ServiceName** and **OrderNum**, which are particularly useful for tracking orders.

Another key source of data is Tuxedo, a high-performance transaction processing system the retailer uses to commit order data into its ERP order systems. In Figure 5, log messages from

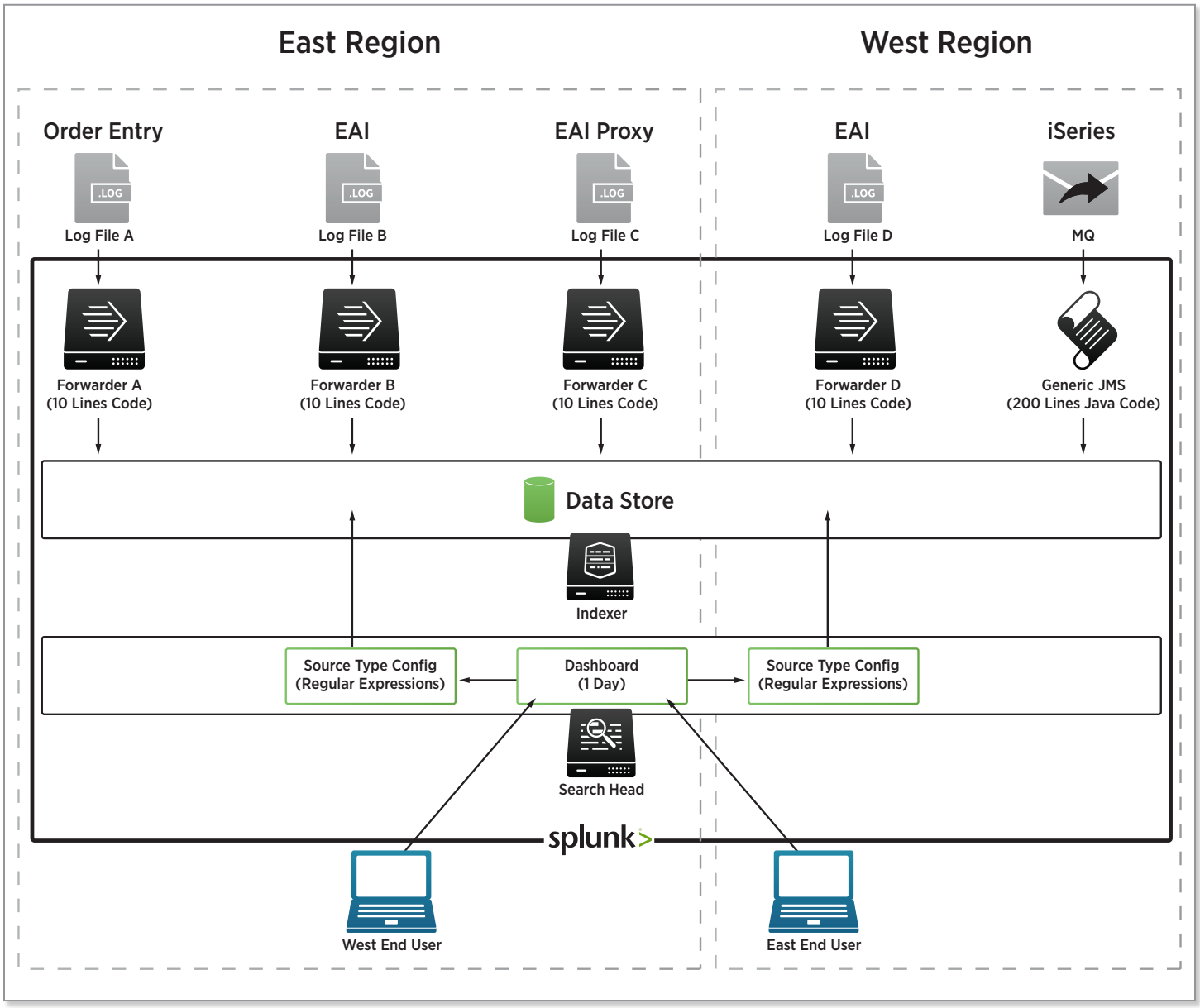

Figure 3. Collecting log data from all systems into Splunk Enterprise.

Tuxedo are in the left column, while the right column shows how the REGEX command extracts important data fields such as the application transaction number (**appl_tran_no**) and transaction error description (**xtrn_error_desc**).

After pulling relevant data from such disparate sources as webMethods and Tuxedo logs, the retailer connects the information using the Splunk Search Processing Language (SPL™), specifically the transaction command. This step groups events from webMethods and Tuxedo systems together and builds an end-to-end trace for purchase orders.

```
> (sourcetype=webMethods_log OR
sourcetype=tux_log) AND action=purchase
| eval order_id = coalesce(orderNum,
appl_tran_no) | transaction order_id
```

This particular set of search commands looks through extracted webMethods and Tuxedo logs and informs the system that **orderNum** from webMethods and **appl_tran_no** from Tuxedo are the same data fields that uniquely identify each order. Splunk Enterprise then coalesces these two fields into a single new field **order_id** and, lastly, uses **order_id** to group and trace events in webMethods and Tuxedo.

The Splunk Solution: Complete Order Visualization

Using Splunk Enterprise to extract and group meaningful data from disparate systems, the retailer then created a number of dashboards that enable its IT and business users to visualize orders as they travel across the production environment.

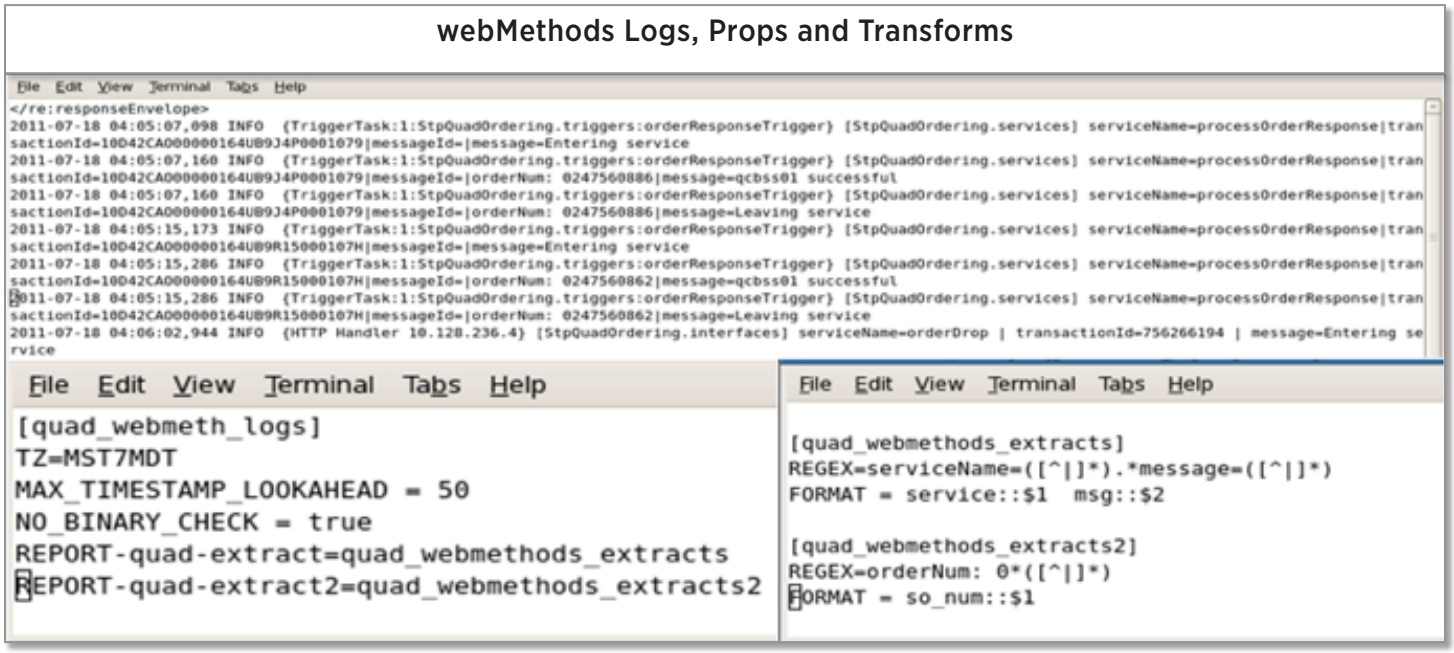

Figure 4. Pulling useful data from webMethods logs.

A dashboard was created for IT operations to monitor transactions going through the retailer's integration backbone system (see Figure 6).

This dashboard shows a pie chart of the count of services that have sent any events in six single value metrics, breaking down the number of transactions versus failures versus SLA violations, as well as a table of these results. The retailer's IT operations team can then drill down into any particular service name in the chart for granular details. From the dashboard above, users can track a specific transaction via a unique transaction ID across all the disparate systems.

Figure 7 represents the transaction path dashboard, which allows a user to view details about a particular transaction. A successful transaction log shows the hops it went through, as well as the raw events. If there was a failure, the dashboard also provides a link to the integration page. This was a huge win for the IT developers because it enabled them to trace any problem to the error origin and even the error cause.

The retailer also uses Splunk Enterprise to aggregate transactional data for real-time optics into the volume of orders going through the system. Figure 8, for example, is an executive dashboard featuring a stack chart broken out by orders over a 30-day period by geography and business unit.

In addition, Splunk Enterprise has helped IT monitor the overall trends and health of the entire stack. The trends dashboard (see Figure 9) shows comparisons of services between days of the week, for entire weeks, and quarterly, as well as a stack chart of all the services during the previous month. When executives request this type of information, it can be made available instantly, enabling them to holistically view service levels and anticipate future needs.

Orders Found, Time and Revenue Regained

As retailers grow, they inevitably reach a stage where there are multiple in-house and partner systems that integrate to fulfill customer orders. One by-product of this growth and these multiple systems is order loss due to complex data integration. Order loss can significantly impact a retailer in terms of reduced revenue and customer satisfaction as well as expensive technical troubleshooting involving numerous resources and man-hours.

Rather than go through the costly, lengthy and unpredictable process of tightening systems integration or data warehousing, this particular retailer deployed Splunk Enterprise to trace orders throughout their life cycle. As with most Splunk customers, the retailer saw immediate value from Splunk software. By using the solution to track customer orders, the retailer has prevented up to 100 lost orders per week, with a corresponding $900,000 in annual revenue recovery. In addition, the retailer has enhanced

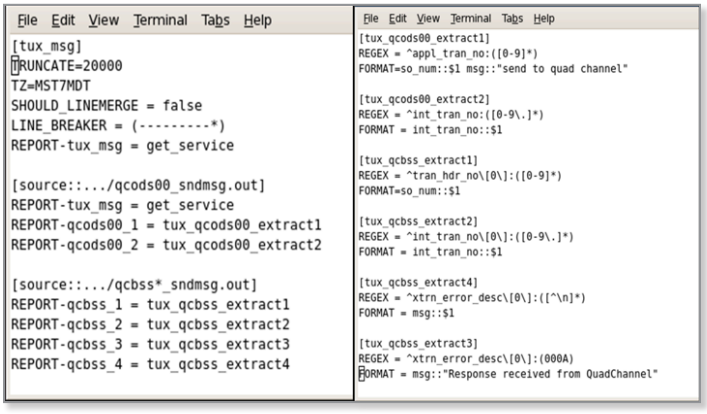

Figure 5. Pulling useful data from Tuxedo logs.

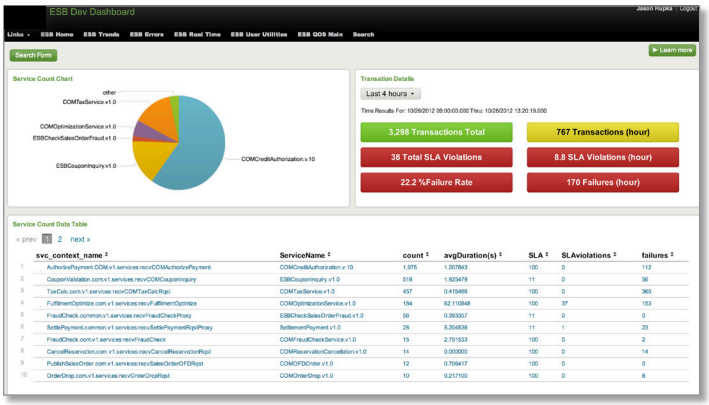

Figure 6. Integration dashboard monitors transactions and SLAs.

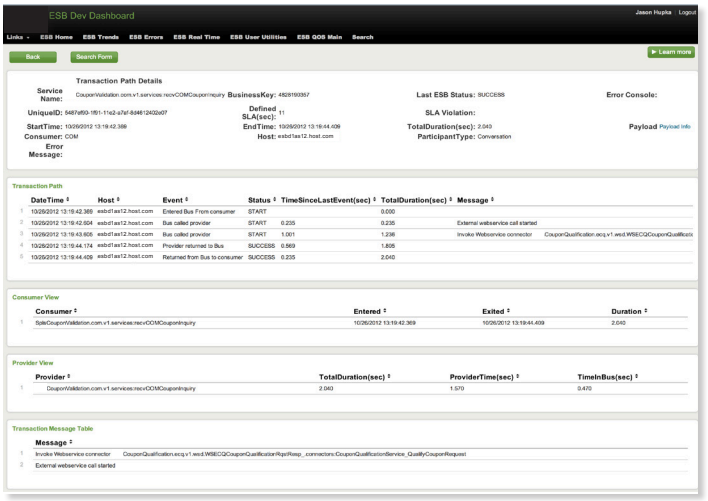

Figure 7. Detailed order transaction tracing.

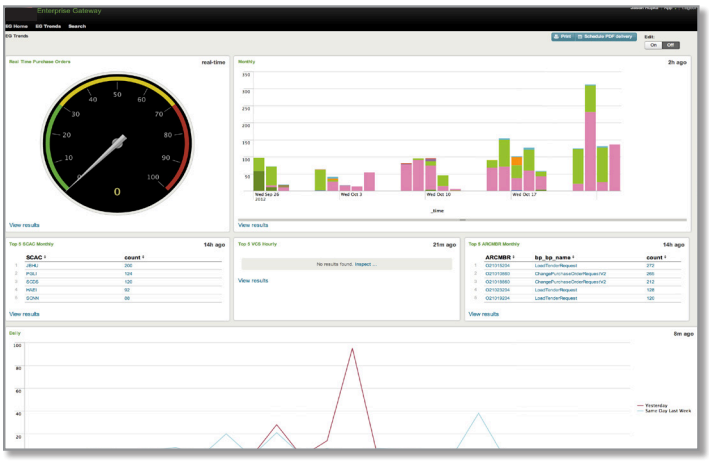

Figure 8. Real-time order dashboard.

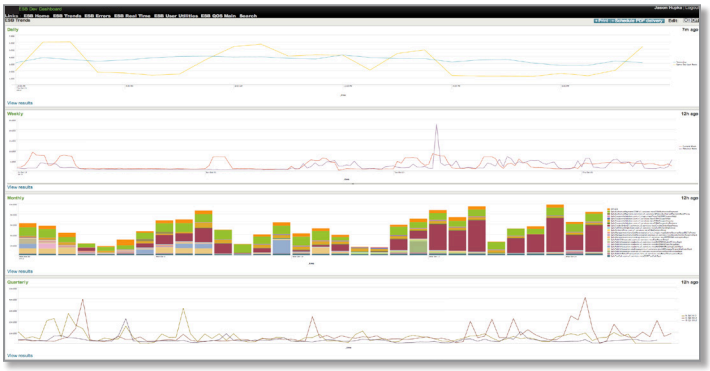

Figure 9. Transactions trends.

operational efficiency, saving numerous man-hours for the IT support and development operations teams involved in troubleshooting. Since deploying Splunk software, these teams have found and addressed issues up to 25% faster, freeing up time for higher value projects.

Splunk software not only allows this retailer to extract critical business insights out of its application and systems data, the solution also enables the customer to improve overall operational efficiency and increase customer satisfaction.

With Splunk, They Found Order(s) in Chaos

In this use case, we explored how Splunk software can enable operational intelligence for a classic transaction/workflow challenge. Despite a daunting stack of dozens of different software components handling four million orders per hour, an online retailer has the optics to figure out when something has gone wrong and how to fix it. This use case demonstrated:

- **Elimination of data silos.** Because Splunk software indexes all kinds of machine data, the customer's reams of siloed data no longer hinder its ability to quickly find lost orders and rectify underlying issues.
- **Correlations drive analytics.** Because Splunk software correlates different types of data, the customer links error messages in one system's logs to useful evidence in other logs to resolve issues.
- **Flexible analytics powered by a read-time schema.** Because Splunk software keeps the data unstructured, the customer views the data through the lens that matters at the time it queries the data, not the lens that mattered to the system at the time it generated the event.
- **Significantly reduced manual intervention and labor cost.** Because Splunk Enterprise provides real-time visibility of end-to-end data flow through the customer's business process, the retailer tracks orders across disparate systems and reduces MTTR.

CHAPTER 6

Detecting Insider Threats

How Splunk Software is Used to Safeguard Financial Data

USE CASES

Fortifying Internal Security
Streamlining Internal Processes

Executive Summary

Financial institutions increasingly rely on outside contractors to program and maintain applications, manage projects, and perform analyses and assessments. Unfortunately, whether intentionally or not, these contract employees can pose a serious insider threat. How can organizations effectively monitor contract employees to detect potential security breaches before they can impact the confidentiality of customers' data?

The security teams at a major North American bank struggled with just this issue. Application developers approaching the end of their contracts were data flight risks and threatened the bank's intellectual property. Monitoring the activities of all these contractors using commercial tools, however, was far too costly and resource intensive. The bank also had no way to correlate data from its various monitoring tools to assign risk levels to contractors and it lacked an automated methodology to respond quickly to risky behavior.

To resolve these challenges, the bank turned to Splunk® Enterprise. The software rapidly integrated data from the bank's various monitoring systems to provide graphical, holistic views of its threat assessment environment. With this insight, the financial institution assigned risk levels to contract employees based on their roles at the bank and the expiration of their contracts. When a risk level is exceeded, the Splunk platform issues alerts and enables the security team to take timely and appropriate actions.

With the Splunk software capturing and displaying all relevant security data, the bank cost-effectively monitors its contract employees and safeguards its assets from both intentional and inadvertent wrongdoing.

- **Integrate disparate data sources into holistic views.** The bank's monitoring systems provided data about risky sites and malware and tracked which websites contractors visited, but these solutions were compartmentalized. Piecing together a threat matrix was far too manual and time consuming. The Splunk platform collects, indexes, and visualizes these unstructured data streams, offering managers coherent views of such questionable behaviors as improper site visits or downloads.
- **Identify high-risk behavior.** Managers were unable to correlate multiple data sources that contained indicators of high-risk behaviors with employee role data to identify which contract workers posed legitimate security risks. Deploying more granular monitoring tools was prohibitively expensive. Splunk software correlates threat data with contractors' role data to identify workers more likely to pose hazards to the bank.
- **Automated targeted responses.** The bank's costly monitoring tools were unable to automatically target contract employees who exceed predefined risk ratings and pose a security threat. The Splunk platform coheres all threat and employee data, so when workers exceed risk ratings, alerts notify managers of potential breaches and restrictive actions are triggered to defuse the threat. The bank preserves its security cost-effectively and neutralizes potential threats presented by employees identified as "high-risk."

Data Flight: Much More Than a Financial Concern

Banks run on money and software. They require hundreds, sometimes thousands, of contract application developers, project managers, systems analysts and other technical staff to deliver functionality to customers across all banking services. In addition, banks must adhere to stringent security mandates to safeguard highly confidential financial data. Yet, in spite of considerable

Business Benefits at a Glance

Splunk Value	How Value Is Measured	Business Impact
Correlate multiple data sources to identify high-risk behaviors by employees and contractors	• Elimination of data silos • Integrated, holistic views of data from multiple tools	• Critical bank source code and data are protected • Visibility into potentially damaging employee behavior
Trigger alerts on individuals with high-risk behaviors	• Flexibly align high-risk activities with security policies	• Proactive enforcement of security policies
Automatically deploy more focused monitoring or loss prevention measures	• Rapid response to identified risks	• Intellectual assets and financial data safeguarded from temporary employees • Reduced cost of monitoring

security measures, the threat of data flight is significant and difficult to to proactively identify. As contractors approach the end of their terms, they typically attempt three things:

1. Securing their next contract
2. Obtaining a copy of the source code that they've written for their portfolios and subsequent jobs
3. Getting enough sample data to exercise that code

Unfortunately, these activities can result in significant financial and legal impacts for any employer, but particularly for financial institutions. While employees' motives for copying source code, process information or data might not be ill-intentioned, their actions can place the enterprise in peril and be in violation of multiple federal laws.

Moreover, temporary employees often spend company time searching for their next contracts, at the expense of their current work and productivity. Additionally, contractors who download code can inadvertently introduce malware to the corporate network. While the great majority of contract employees are trustworthy, it only takes one to cause significant llegal, financial and publicized problems.

This major bank confronted these challenges as contract employees' termination dates neared. Although lost productivity from contractors seeking new assignments was problematic, the bank's chief concern was data exfiltration. Many contractors had access to application source code that revealed how the bank's underwriting and loan applications systems worked. Such knowledge, if leaked to competitors or identity thieves, could cause severe financial losses and damage the bank's reputation.

The financial institution maintained a robust security posture, but it was vulnerable to these internal threats. It required a way to identify potentially risky employee behavior and respond automatically with appropriate measures, such as deploying more granular monitoring tools.

How to Extract Value From Available Data?

The bank considered available solutions, including a leading security information and event management (SIEM) solution that proved unable to scale effectively to the institution's needs. However, the information needed to identify potentially risky behavior was already available in the bank's existing machine-generated data. The data simply needed to be integrated for a complete picture. There were, however, substantial challenges:

- **The data needed to identify high-risk behaviors was spread across multiple systems.** Microsoft® Active Directory provided detailed information about each employee's business roles and contract terms. BlueCoat listed potentially risky websites, recorded when and which contract employee visited them, and even identified specific job listings accessed by the contractor. FireEye provided threat intelligence about malware and questionable sites. While these disparate systems continually issued logs that tracked internal activities, unfortunately there was no system in place to cohere the data streams to draw timely conclusions.
- **Once risky behavior was identified, deploying more invasive tools to monitor or restrict the target employees was time-consuming and complex.** The bank was unable to automate this process with existing tools, so even if triggering behaviors were noticed, the bank took days to deploy the appropriate monitoring.
- **The bank could not efficiently and selectively monitor based on risk.** Deploying tools like keystroke monitoring to watch employees at a more granular level would have been extremely costly. The bank would need to monitor every contract employee constantly to determine whether any engaged in questionable behavior, which would incur substantial licensing costs. Moreover, keystroke monitoring would consume network resources like CPU cycles, slowing down computing performance across the enterprise.

```
2013-08-30 11:20:18 172.16.200.9 SG-SSL-Proxy-Service shaun 172.16.210.74 www.dice.com
74.115.248.15

443 http://www.dice.com/job/results?caller=basic&q=big+data&x=all&p= "Job_Search/Careers" 200
text/html "Mozilla/5.0 (Macintosh; Intel Mac OS X 10.7; rv:6.0) Gecko/20100101 Firefox/6.0" www.
dice.com 986 494 TUNNELED OBSERVED ICAP_NOT_SCANNED -

2013-08-30 11:20:18 172.16.200.9 SG-SSL-Proxy-Service shaun 172.16.210.74 www.dice.com
74.115.248.15

443 http://www.dice.com/job/result/10451979/268944?src=19&q=big%20data "Job_Search/Careers" 200
text/html "Mozilla/5.0 (Macintosh; Intel Mac OS X 10.7; rv:6.0) Gecko/20100101 Firefox/6.0" www.
dice.com 986 494 TUNNELED OBSERVED ICAP_NOT_SCANNED -

2013-09-02 15:10:45 172.16.200.9 SG-SSL-Proxy-Service greg 172.16.211.32 jobsearch.monster.com
208.71.195.72 443 http://jobsearch.monster.com/search/?q=big-data "Job_Search/Careers" 200 text/
html Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.8.1.14)" - 961 1516 TUNNELED OBSERVED
ICAP_NOT_SCANNED -

2013-09-02 15:10:45 172.16.200.9 SG-SSL-Proxy-Service greg 172.16.211.32 jobsearch.monster.com
208.71.195.72 443 http://jobview.monster.com/Big-Data-Hadoop-Engineer-Job-San-Jose-CA-124644251.
aspx "Job_Search/Careers" 200 text/xml%20charset=utf-8 "Mozilla/5.0 (Windows; U; Windows NT 5.1;
en-US; rv:1.8.1.14)" - 961 1516 TUNNELED OBSERVED ICAP_NOT_SCANNED -
```

Figure 1. Data reveals a contractor searching for jobs on external sites.

Enter Splunk

Looking for a solution, the bank downloaded a copy of Splunk Enterprise and discovered that the software can capture and index unstructured, machine-generated data. Splunk Enterprise could monitor and analyze logs from disparate application servers at a scale that the bank required and their current SIEM could not deliver. The Splunk Search Processing Language (SPL) enabled the bank's security team to quickly query the data and graphically display the findings in dashboards in real time.

Since its deployment, Splunk Enterprise has allowed the bank to harness its existing data to proactively identify employees who warrant additional monitoring. The software correlates the data about user behaviors coming from BlueCoat with the FireEye information about potentially risky sites, and then ties it all together with employee information from Active Directory. This allows the bank to define a risk rating and identify employees who surpass it. The risk rating takes into account employees' roles at the company and contract end dates from their Active Directory profiles. For example, employees working on multiple code projects (such as project managers) have access to more data than those working on single projects. Splunk software also can identity high risk behaviors such as data downloads, which might contain malware, as well as external website visits after working hours.

Using the correlated data, Splunk Enterprise triggers an alert whenever an employee exceeds a predefined risk rating. As part of the alert firing, Splunk Enterprise is configured to take action automatically based on the data. For example, depending on the severity of the alert, Splunk Enterprise can do one or more of the following: turn on keylogging, block or track USB storage devices, and/or send a warning email to the contractor's manager. Other actions include setting a flag in the user's profile that blocks access to such online storage services as Dropbox, Yahoo! email, Amazon S3 or iCloud.

```
Index="bluecoat" cs_host="*.dice.com"OR
cs_host="jobsearch.monster.com" cs_
username=* | rex "(FREE_TEXT|q)=(?"<"Search
String">"[^\w\+(\)\%\'\.]+)" | stats count by
SearchString, cs_host, cs_username | rename
cs_host AS "Web Site" cs_username AS "User
Name" | sort _time
```

Figure 2. A Splunk search to help identify high-risk contractors.

To Catch a Potential Data Thief

How does Splunk software perform the analysis? Data indexed from the Bluecoat Proxy can be searched (and alerted on) for suspicious activity. Figure 1 shows a contractor visiting sites such as dice.com and monster.com to search for "Big Data Hadoop Engineer jobs" during working hours.

To identify contractors who may warrant more monitoring, the bank's security team crafted Splunk searches like the one in Figure 2.

This search examines BlueCoat data to identify contractors accessing job search sites over a 24-hour period and displays the search terms used broken down by contractor name and site, as shown in the bottom panel of Figure 3. Splunk's **rex** search command automatically extracts the search term (in this case, the type of jobs the person is searching). Figure 3's top panel displays when the searches were made over this 24-hour period.

Results from this search are then enhanced with FireEye data and used to drive the automated deployment of additional monitoring to contractors with "risk rating" values above the defined level. In addition to this example, the bank has over 200 other types of data collected from 11 other applications by over 3,000 Splunk forwarders linked to application servers and web devices. This information feeds close to 200 security dashboards and reports, which are images of static dashboards sent by email or PDF.

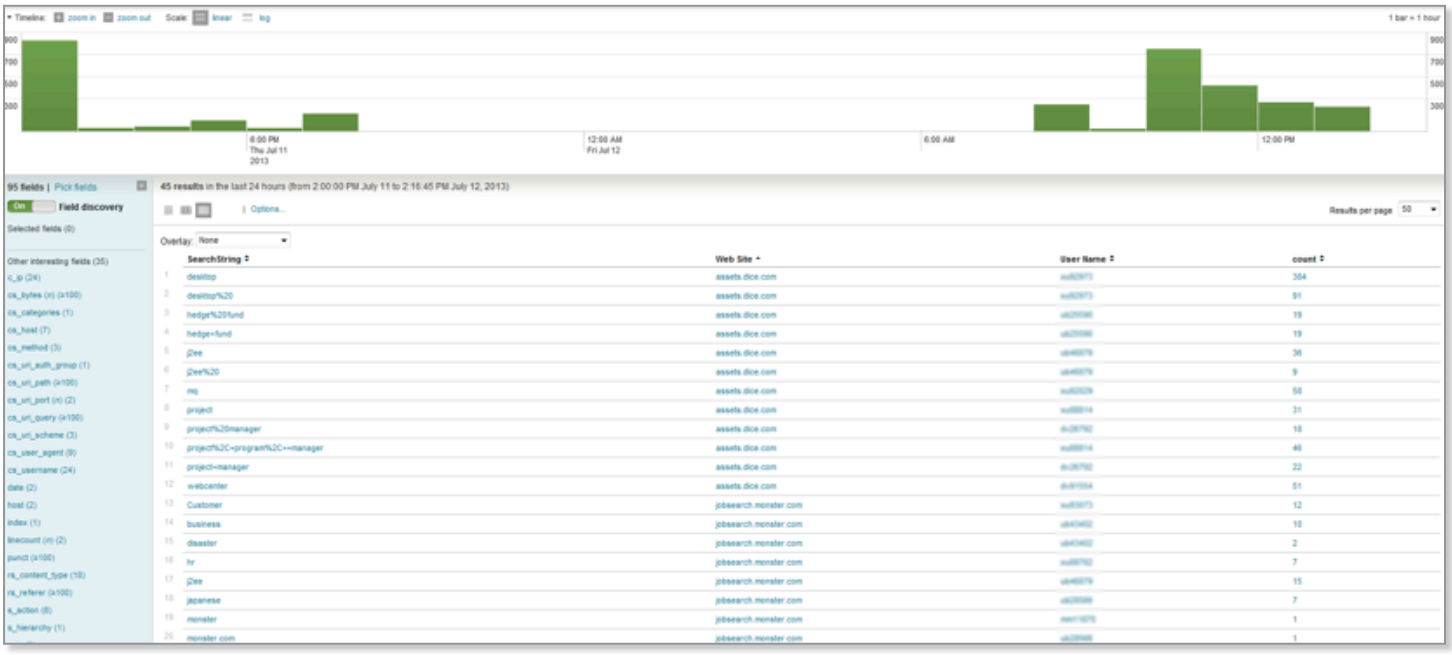

Figure 3. Contractors accessing job search sites over 24 hours (bottom) and the times when the searches were conducted (top).

Clamping Down on Internal Threats

While companies typically invest heavily in technologies that protect them from external security threats, internal breaches are often the harder challenge. This problem is exacerbated when numerous temporary contractors are working on key projects and have access to sensitive systems and data. Individual device-level monitoring can be expensive, foster an atmosphere of distrust, and may be ineffective because they capture only a single facet of risky behavior, such as keystrokes. Companies need to leverage data sources from all IT systems and applications. Correlating across these data sources can provide insights to risky behavior that prefaces security breaches.

Splunk software analyzes and correlates data and helps customers understand the behavior of internal users. Specifically, the Splunk solution allowed this particular bank to engage in proactive enforcement of its security policies to protect its intellectual assets and potentially avoid multi-million dollar fines.

With Splunk, They Are Safe and Secure

Enterprise data can reveal much to a business. This use case demonstrates how machine-generated data provides the knowledge to ensure the internal security of a financial institution. What is essential, however, is that data produced by many disparate systems is captured, integrated, and displayed. This use case demonstrated:

- **Elimination of data silos.** Because Splunk software indexes all kinds of data, the customer's siloed data no longer hinders its ability to obtain coherent views of potential security problems.
- **Correlations can bolster security.** Because Splunk software correlates different types of data, the customer links messages in the logs of multiple systems into useful evidence for potential issues.
- **Flexible analytics powered by a read-time schema.** Because Splunk Enterprise collects data in full fidelity without any filtering, the customer does not lose any potential value by making its data fit in a schema. The customer can engage in near real-time analysis of questionable employee behavior.
- **Significantly reduced manual intervention and labor cost.** Because the Splunk platform provides real-time visibility of online actions and can trigger alerts when thresholds are exceeded, appropriate responses are automated, providing very rapid measures to curb unwanted actions.

CHAPTER 7

Solving the Disappearing Test Problem

How Splunk Software Is Used to Track and Score Millions of Exams

USE CASES

Troubleshooting Services Delivery
Improving Operational Efficiencies

Executive Summary

What do you do when 100 "widgets" go into a machine and only 99 come out? And what if those widgets are standardized tests, the "machine" is a $7B exam processing company and a multitude of government and commercial customers—not to mention anxious test-takers—are waiting eagerly on the other end, expecting results?

Executives at one Splunk customer struggled with this exact situation and were considering a variety of costly solutions to address an even costlier problem: missing exams meant broken contracts, lost business opportunities and a tarnished company reputation.

After reviewing the available options, the executives determined that Splunk® Enterprise had the three key components necessary to diagnose and resolve their issue. Splunk software now enables them to leverage their machine data to:

- **Gain visibility across data silos.** They knew the data necessary to find the "lost" exams was somewhere in the enterprise, but they had no way to search across the various data silos efficiently enough to be useful. Splunk software let users search across data silos with one query.
- **Sift data in the ways that mattered.** Log data was naturally organized by system, but technicians needed to filter the data by exam. Because Splunk Enterprise doesn't force data into pre-determined structures, they were able to filter data from a variety of systems to pinpoint the activity surrounding a specific exam.
- **Automate and simplify manual time-consuming processes.** The visibility enabled by Splunk software allowed the IT staff to access machine data from across the infrastructure and all of its systems. Staff no longer had to write custom scripts, manually search through the machine data to locate a bottleneck or determine a root cause.
- **Clarity into compliance and customer service.** Additionally, this customer found Splunk software to be useful in other ways. Because they could track human graders' activities by time, they identified employees who were overworking and violating compliance standards. They also used Splunk Enterprise to alert their technicians to problems before customers were affected and the support staff was overwhelmed.

The Disappearing Test Problem

Turning a standardized test into a carefully graded and fully validated document is a significant technical challenge. Over several years, this customer had custom-built a system consisting of five Java applications running on JBoss middleware and an Oracle database. This system performed over 25 separate functions, including matching exams with other registration information, comparing exams to answer keys and even processing results from human-graded portions of exams.

At one time, the system could process as many as five million exams over a single weekend, but unrelated organizational and technical demands led the customer to switch from a centralized, mainframe-based system to a distributed architecture.

While this change solved old problems, it created a new problem of exams going missing, forcing technicians to reduce batch sizes from the original five million exams per weekend to as few as 10,000. Even with these smaller batches, the customer found that for every 10,000 exams that went into the system, 9,997 would come out fully processed. The others would get lost in the system, initiating a time-consuming manual investigation to determine the root cause and find the exams.

The Business Process: Exam In, Grade Out

The customer's complex, distributed grading system consisting of the 25+ step exam evaluation process can be grouped into five basic categories (see Figure 1):

1. **Import exam.** Exams in both paper and digital media are imported into the evaluation system.

Business Benefits at a Glance

Splunk Value	How Value Is Measured	Business Impact
Visibility to the entire test-scoring workflow and its systems	Time to locate missing test in complex system	Reduced from weeks to minutes
Reduce manual diagnosing and searching for lost tests	Number of worker days and corresponding costs saved	Eliminated the need for a 12-person team for an estimated savings of $360K annually
Labor law compliance	Number of out-of-compliance graders	Reduced from 30–40 daily-out-of-compliance graders to near zero

2. **Match exam to registered student.** Exams are matched to examinee registration data and validated to ensure the right examinee took the right exam.
3. **Machine score the exam.** An answer key is loaded and compared to the answers provided on the test, and then a score is generated.
4. **Human score the exam.** For exams requiring more than machine evaluation, the system queues and presents the test for a grader and processes the results.
5. **Aggregate and publish the scores.** Once an exam is fully reviewed, scores are aggregated and published on a web-based portal for customers and examinees to review.

A small percentage of the exams that passed through this multistep grading process had issues. Customer test-takers would check the online portal for their test scores only to discover the exam results were "still processing" days after they should have been available.

This led many customers to call the company's help line. To resolve these issues, a dozen-person, cross-functional team often worked weeks to manually dig through complex, multi-tiered systems. This team represented datacenter infrastructure, hosting, database administration, application development, the software architecture board, customer-facing groups and at least one business executive. This ongoing manual effort cost the company hundreds of thousands of dollars.

The Investigation: So Much Data, So Little Time

Technicians troubleshooting the missing exams had no alternative but to manually sift through the various system events. This was problematic for a variety of reasons:

The data was not centralized. They had volumes of data from each of their five Java applications, their Oracle database and their JBoss middleware system. They also had logs tracking each exam as it went through the system and logs tracking the overall enterprise service bus. The logs were of different formats and were located in different files on different servers. When exams went missing, technicians were forced to comb through each set of logs manually, a process that was tedious at best and futile at worst.

Error messages did not help solve the problem. In some cases, technicians would find helpful error messages. For example, if a database error occurred, they might find a message like "record locked for update." That helped them to pinpoint the culprit—the database—but they could not easily connect the error message to a specific exam. They were still preventing from finding the lost exam and resolving the customer's issue.

They could not correlate the data that would be most useful. Logs were naturally organized by application or system, but the technicians were tracking down lost exams, not lost systems. Their system logs included test IDs, but isolating those test IDs and viewing the progress of exams across all systems was essential to their process and yet impossible with their existing tools.

The Splunk Solution: Operational Timesheets

Using Splunk Enterprise, the customer's team focused on the three critical pieces of information necessary to find a lost test:

- The test ID
- The processes the test had completed
- The time when the test completed each process.

All of these data elements were present in the logs, but the technicians could not easily gather and correlate them. Splunk software enabled them to capture and visualize a test's journey through the system (see Figure 2).

Step #1—Gathering and correlating the data. As indicated in Figure 2, each of the five Java applications wrote the test ID and its own workflow ID to an "authoritative transaction log" when it began work on a given test. For example, when the first application (known as "A1") began work on test "jdoe1," it wrote the following line to a log:

```
2011-10-02 09:01:17.12 Start attempt for
import WFID=A1 TestID=jdoe1
```

Once those logs were indexed into the Splunk platform, the technicians had enough information to track a test all the way through the system. To view all the data related to test ID "jdoe1," for example, the Splunk query is simply:

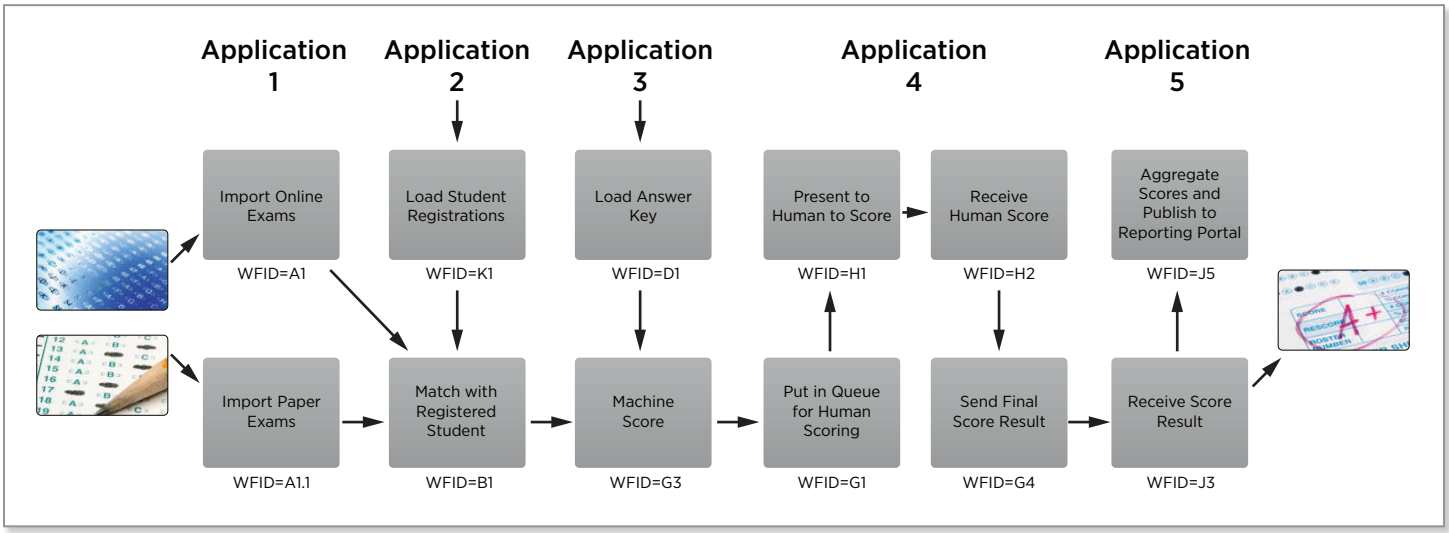

Figure 1. A snapshot of the scoring process.

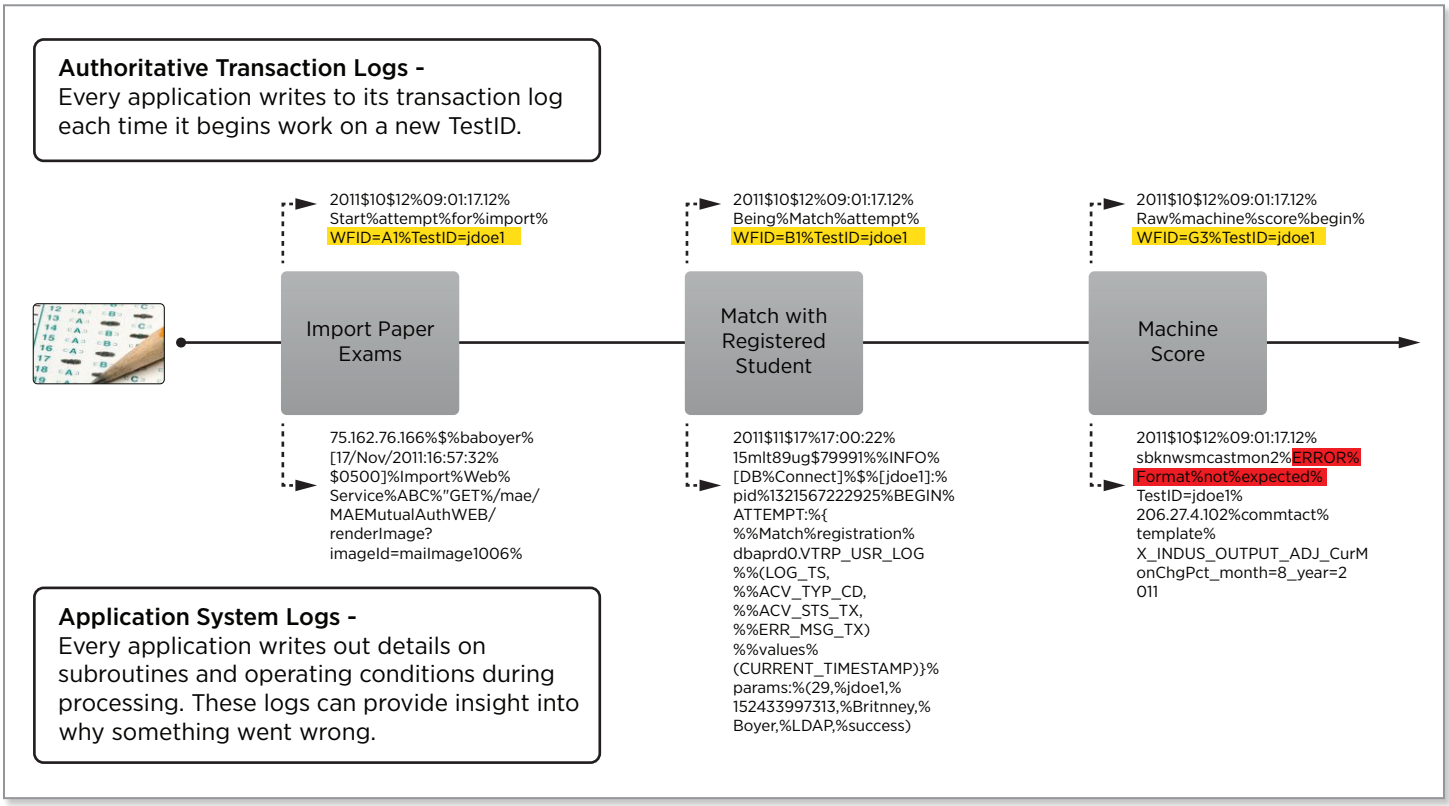

Figure 2. Gathering and correlating the data.

```
> TestID=jdoe1
```

If they wanted to see how far test ID "jdoe1" had progressed through the system, they could use the statistical analysis command "last," as in:

```
> TestID=jdoe1 | stats last(WFID)
```

and Splunk would display the most recent workflow ID (WFID) reached by test "jdoe1."

If they wanted to identify those tests that hadn't completed the entire cycle (as evidenced by reaching workflow ID "J5"), the command would be:

```
> * | stats last(WFID) as LastWorkFlow by
TestID | search LastWorkFlow!=J5
```

Finally, if they wanted to investigate an issue, they could select one of the test IDs from the above search and do another simple search for all data associated with that test ID. The search would reveal all data related to that test and would identify the precise time when that test got lost in the system. After isolating the search to that second or minute, the technicians could search across all their data for issues reported during that narrow timeframe, reducing the technicians' analysis time from hours and even days to seconds.

Once the process was automated, the customer no longer needed the standing cross-functional team to search for lost tests, reducing resolution time from weeks to minutes and saving an estimated $360,000 per year.

Step #2—Using scheduled searches and lookup tables to expedite resolution. Once technicians discovered that even the most basic of Splunk searches could dramatically improve their ability to locate lost exams and identify issues, they began to discover Splunk software's other benefits.

Scheduled Search. Scheduled Search is a Splunk feature that allows the user to create and save a query and then ask Splunk software to trigger that query on a scheduled basis. In this case, technicians created a query to isolate and correlate the three key pieces of data (test ID, workflow ID, and timestamp) and then ran this query against their data early each morning, when other users were unlikely to be making demands on the system.

Lookup Tables. Lookup tables are flat, comma-separated files that enable the user to provide extra information about a piece of indexed data. For example, the raw indexed data might only contain "employee=Sondra." The user could instruct Splunk to link that raw indexed data to a lookup table with email address and other information. If the final lookup table looked like this:

```
employee,email,role,location

Sondra,srussell@splunk.com,Sales
Engineer,Washington D.C.
```

the following search would find the raw data containing "employee=Sondra"

```
> location="Washington D.C."
```

Lookup tables were used in an unusual way to efficiently track the state of any given exam at any given time (see Figure 3).

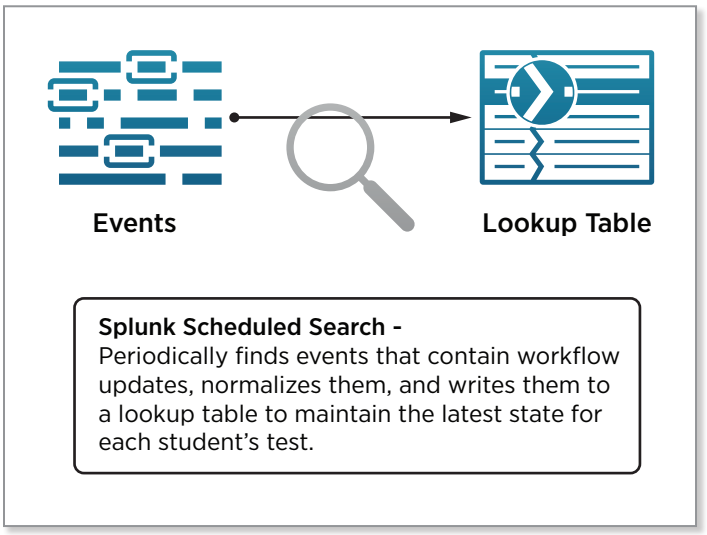

Figure 3. Using scheduled searches.

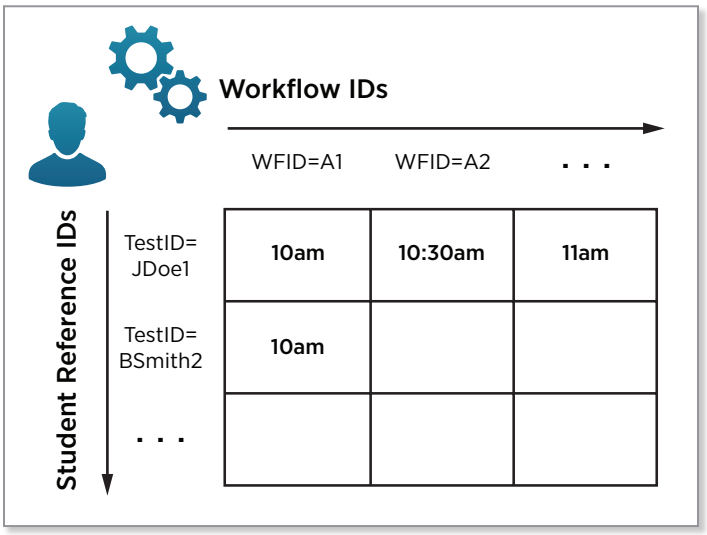

Figure 4. Workflow ID lookup table.

Using the following scheduled search (query),

```
> * | table WFID over TestID | outputcsv
```

the customer created and continually updated the lookup table (see Figure 4).

The final lookup table read like a timesheet for each exam and gave the customer a powerful, easily searched snapshot of the entire system status that it could link with raw data when needed. This combination of high-level system state summary and low-level system data allowed technicians to both quickly resolve tactical, one-off issues and answer strategic, organizational-level questions, including:

- Where is each exam in the system at this moment?
- What, if any, exams are "stuck" in the system?
- Which workflow stages take longer to process than expected?
- How long does the average exam take to fully process?

Continually analyzing the data from this perspective enabled the customer to move from a reactive organization—where users were forced to reduce batch sizes and devote extra hours to researching technical issues—to a proactive organization with the information at hand to improve service and accurately forecast capacity.

Soft ROI: HR Compliance and Proactive Customer Service

As with many Splunk customers, this company soon found other uses for its data, including reducing the demand for technical support and identifying hourly workers who are overworking.

Reducing costly escalations and improving the customer experience. The company provides exam takers with access to a web portal to check the status of their exams. Before its Splunk deployment, the company had difficulty identifying which exams were lost and therefore mostly relied on test-takers to call technical support and self-identify as having a problem.

Once Splunk Enterprise was in place, the customer used the alert feature to trigger a Python script. The alert identifies tests that needed to be resubmitted and the Python script sends an email to the examinee requesting the resubmission. This process yields the triple benefits of speeding up the resubmission process, reducing the inbound call volume to technical support and improving the overall customer experience.

Preventing HR labor violations. The customer employs hourly workers to hand-grade exams that require a high level of evaluation. Since these graders typically work from home, the customer's human resources department found it difficult to ensure that employees were complying with standard regulations, including the requirement that workers take periodic breaks and that they not work between 1 a.m. and 6 a.m.

By querying the data, the operations team was easily able to identify employees who weren't in compliance. The team reported a dramatic improvement—its investigation identified and counseled over 30 out-of-compliance employees and today it reports only rare instances of non-compliance.

Using Splunk Software, They Made the Grade

In this use case, we explored how Splunk software can enable operational intelligence for a classic workflow problem: how do you find stray objects in a system with a single expected path to successful completion? This use case demonstrated:

- **Elimination of data silos.** Because Splunk software indexes all kinds of data, the customer's siloed data no longer hinders its staff from quickly identifying lost exams and rectifying the issues.
- **Correlations drive analytics.** Because Splunk software correlates different types of data, the customer could link error messages in one system's logs to useful evidence in other logs and finally resolve issues.
- **Flexible analytics powered by a read-time schema.** Because Splunk Enterprise keeps the data unstructured, the customer could view the data through the lens that

mattered at the time it queried the data, not the lens that mattered to the system at the time it generated the log.

- **Significantly reduce manual intervention and labor costs.** Because the Splunk platform provides real-time visibility of end-to-end data flow through the customer's business process, when an error or bottleneck occurs, staff can diagnose and resolve the issue immediately, often in an automated way.

CHAPTER 8

No More IT War Rooms

How Splunk Software Keeps Critical Customer-Facing Systems Healthy And Profitable

USE CASES

Troubleshooting Services Delivery
Streamlining Internal Processes

Executive Summary

One Splunk customer, a national home improvement retail chain, prides itself on its range of products and outstanding customer service. The retail chain built customer-facing systems both online and in its stores across the U.S. in order to provide customers with the best purchasing experience possible. These systems were essential for generating sales, but were intricate webs of applications, databases and servers spread across several datacenters and thousands of stores. A small failure or slowdown of any component could cripple the entire stack. As a result, the retail chain frequently suffered brownouts—if not outright outages—of critical systems, resulting in lost sales, disgruntled customers and frustrated employees.

When a vital application stack like a customer ordering system performed sluggishly or crashed, it had to be fixed and returned to service as quickly as possible. Because the components in each stack were owned and managed by different teams, obtaining holistic views of any system was very difficult. The retail chain's IT staff was challenged with the laborious, complex task of determining whether an issue was the root cause of a problem or a symptom of a failure elsewhere in the stack.

By deploying Splunk Enterprise, the retail chain gained comprehensive visibility into its IT infrastructure. The Splunk platform aggregates logs and other machine-generated data to provide end-to-end performance and health metrics of all components. Internal users query the data and display the results in customized dashboards, enabling them to visualize the inner workings of key systems. The company now has the insight to rapidly redress brownouts and outages, as well as to proactively fix potential issues before they impact sales and customers. Some of the benefits since deploying Splunk software include:

- **Holistic views of siloed infrastructure.** Due to the breadth and complexity of its IT infrastructure, the retailer assigned teams responsibility for specific components in its systems. This let staff specialize in the technologies, but it divided stacks into silos. When a system crashed or underperformed, all the teams had to meet in a "war room" to laboriously share logs and data to identify the root cause of the problem. The Splunk platform collects and indexes logs from all systems and provides views of application stacks that are both end-to-end and granular.
- **Avoiding costly issues.** Without panoramic optics, determining which component was at fault when a critical system failed was time consuming. The process could take hours and the retailer lost sales with every minute of downtime. Moreover, this lack of optics hindered the company's ability to identify and remediate potential issues before they became expensive problems. With insights offered by Splunk Enterprise as well as alerts when predefined thresholds are exceeded, the retailer's IT teams can react far more quickly to a faulty application or component and correct shortcomings before they can escalate into crippling problems.
- **Obtaining clarity amidst the complexity.** The retailer's CIO and IT executives had no way of knowing the general health and performance of the company's IT infrastructure. They now use Splunk dashboards to monitor the performance of all systems in near real time and alert on latency and availability issues.
- **Improving the customer experience.** At any time in any store, a customer-facing application could fail, undermining revenues and customer loyalty. The retailer now uses Splunk software to identify problem apps and proactively prevent future outages. By ensuring availability and avoiding latency, the company is able to consistently deliver a superior customer experience.

Business Benefits at a Glance

Splunk Value	How Value Is Measured	Business Impact
Proactive monitoring and alerting of the health of key sales applications	• Fewer application outages	• Reduced the number of high-severity application outages by over 40%
Discover and resolve root causes	• Mean Time To Resolution (MTTR) when applications are down	• MTTR is reduced from hours to minutes
Find faulty apps to prevent future problems	•Improved quality of in-house, custom-built applications	• CIO-level reports showing error rates for home-grown applications • Error rates for some apps reduced from 80–90% to less than 10%

Fighting a Losing Battle

Imagine a hypothetical situation. A customer at a major home improvement retail store has spent an hour selecting custom flooring and carpeting and calculating the necessary square footage. When she takes her notes to the sales representative, however, he tells her that the order system terminal is down. In fact, all terminals in that store and 200 other stores in the region are down. Without the system, he cannot place the customer's order nor schedule necessary installation services. He asks the customer to return the following day, hoping he does not lose a sale to a competitor.

Meanwhile, the store manager is on the phone with a corporate IT war room, along with fellow managers from across the region. They are painfully aware that custom orders are their most profitable product segment, so daily and weekly sales projections are in jeopardy due to this outage.

Nearly the entire IT operations department ends up mobilized across the company's war rooms and the CIO is pulled into the fray. Teams onsite and across the country work on multiple systems to diagnose the problem, spending multiple man-hours on the issue. Every minute that the system is down affects the bottom line, as well as the overall customer experience. The cost is tremendous—time, money and the company's reputation are at stake.

Discovering Root Cause

The outage situation described above was, unfortunately, far from hypothetical at this particular national home improvement retail chain. The company's customer ordering system is comprised of more than 20 subsystems—applications, databases and servers. Some are deeply integrated; some are connected together "on the glass" at store kiosks. Some systems are custom-built by the company's application development team; others are an integration of business data and information from partners. IT operations analysts need to pull log data from all of these systems to piece together the full picture and identify the problem.

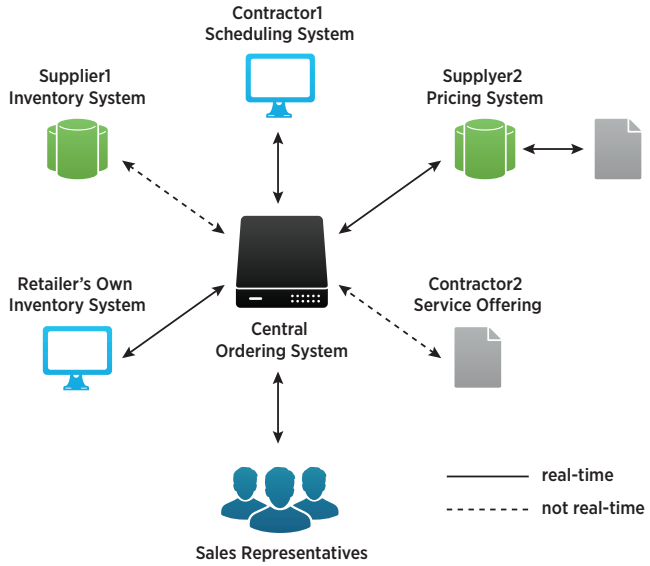

Figure 1. Central ordering system connected to internal and partner systems.

The home improvement company initially evaluated the Splunk platform to obtain a single pane-of-glass solution for logging and application troubleshooting. It sought to reduce application brownouts, which were common, as well as outages, or blackouts. The applications at risk were both customer-facing systems and those used for internal business processes. When issues arose, customers were unable to place orders or productivity was impaired. To meet customer expectations and minimize business impact, administrators needed to decrease the number of incidences and lessen their severity or longevity.

To do so, administrators required centralized visibility into system logs to troubleshoot the behavior of devices and systems in complex application stacks (see Figure 1). However, prior to deploying Splunk software, the organization lacked standardized logging capabilities and its log data was siloed by application and owner. Without holistic

tomcat_access	netscaler	idm_usage	com_sterling_ ScheduleAgent	memcached
syslog.linux	Perfmon:SQLServerPRDI29	microstrategy_report	com_sterling_ ItemAssortmentServer	splunk_util_netstat
WMI:SMS_PackageStatus	com_sterling_ ProcessOrderUpdatesServer	com_sterling_ POReprocessAgent	com_sterling_ ProcessReceiptServer	appl_log.mainframe
mobility	javabatch	eai_score_log	com_sterling_ InventorySyncServer	tomcat_sso
iis.Emp_Portal	store.broadband	istore_kiosk	urlcheck	cpuinfo
bluelog	access_common	com_sterling_ ItemDownloadServer	apache_server_status	WinEventLog:System
netcool_event	com_sterling_jboss	bpm_access	assortplan.usage	WinEventLog:Application
macaddr.store	bizlink_log4j	Cctv_log	abinitio	esx.iscsi

Figure 2. Select data source types.

views of system performance across stacks, many administrators—each responsible for particular applications—needed to hastily confer in the datacenters or virtual conferences to share log data and figure out the root causes of problems.

Additionally, aggregating logs in the company's two datacenters from hundreds of stores and multiple applications congested the company's network. This further slowed down troubleshooting.

Enter Splunk

The company implemented Splunk Enterprise in its two datacenters and across its application portfolio. Staff are now able to search for root cause immediately because the software continuously captures and indexes all relevant log information. The retailer's IT teams now have all the data in hand, avoiding congested links and slashing mean time to resolution from hours to minutes.

The size of the Splunk deployment is impressive. The volume of data the company collects and indexes has grown from 350GB every day to 500GB. As more and more systems send data to be monitored, the volume of data will increase to approximately 750GB daily. The data represents over 200 types of sources (see Figure 2) and more than 600 applications. Logs are collected from the employee portal, the customer order management system, network devices, store support systems and the e-commerce site. Data sources range from Tomcat web servers and a Sterling Commerce engine to Microsoft SQL Server database and closed circuit TV logs.

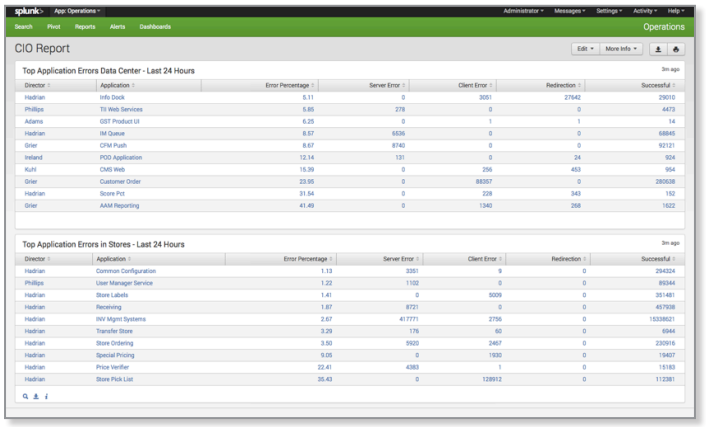

Figure 3. The CIO's dashboard displays the most error-prone applications.

To send logs from endpoints to Splunk Enterprise, over 25,000 small applications called forwarders are deployed on various systems and applications, and more than 120,000 searches are conducted on new and historical data every day. Eventually, as many as 200,000 forwarders will be installed when all store systems and devices are equipped with them.

Building Dashboards Across the Enterprise

The Splunk platform displays performance metrics in highly customizable dashboards; once the deployment was mature, demand for Splunk dashboards grew across the organization. More than 200 IT professionals routinely search for valuable data and insights, and executive dashboards keep IT and business leaders informed of the overall health of enterprise systems. The CIO's dashboard shows the status of key business and IT systems across the organization, and highlights the most error-prone and vulnerable applications (see Figure 3).

IT operations dashboards monitor over 600 applications in real time and batch mode across 2,000 stores. They track usage metrics and identify unusual patterns that may cause outages (see Figure 4). Hourly and daily operational reports and notifications are delivered and integrated with alert systems and IT ticket systems.

Dashboards are typically six to nine panel varieties with combinations of single value buttons; count charts, trend charts, time charts and tables. Executives find business trend comparisons over time to be quite useful, such as this week versus last week or this month versus the same month last year.

A Splunk IT operations team of 2.5 full-time employees builds the dashboards. They take requests from different groups, from the CIO on down, and develop dashboards for lines of businesses, end users, applications and systems owners.

Since putting Splunk Enterprise into production, the team has delivered approximately five dashboards per week. This rapid deployment resulted in an accumulation of approximately 5,000 searches per hour to keep the dashboards continually updated and for alerts when anomalous behavior is detected.

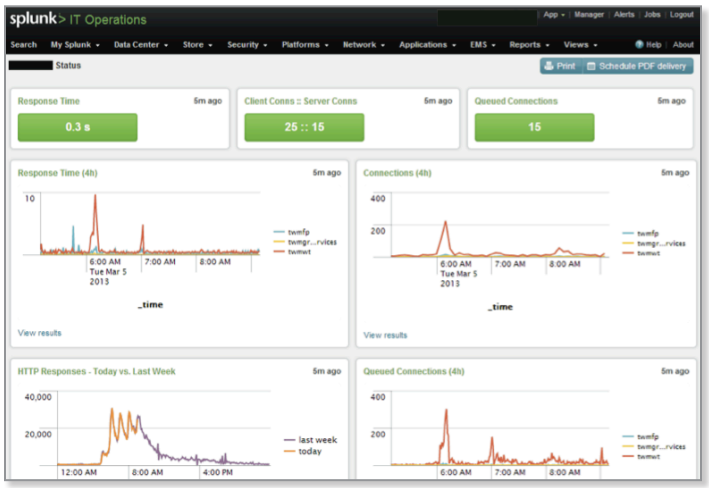

Figure 4. IT operational dashboard shows key performance metrics for an enterprise application.

Security is Intelligence

The Splunk solution benefits other operational domains at this national home improvement retailer. The security team uses the Splunk App for Enterprise Security, which runs on top of Splunk Enterprise. The application aggregates data from the company's Windows infrastructure, such as from Active Directory and domain controllers, as well as from antivirus and intrusion detection systems, to provide the monitoring, alerting and analytics needed to detect and address all manner of threats. The security team also feeds data collected by Splunk software into HP ArcSight, a security information and event management (SIEM) system.

Finding the Problem Application

In addition to using Splunk software to recover from system outages, the company is proactively using the solution to prevent downtime. In the past, IT operations lacked visibility into application quality. Some poorly architected and built apps can consume disproportionate amount of CPU, memory and network resources. The symptoms of problem applications manifest not only in their error messages, but also in the slow performance of nearby applications. For example, a Java Virtual Machine (JVM) where the ToolRental application is running has been restarted 61 times, impacting 32 stores. It is likely that with such frequency, there are unusual technical issues with that application. Customers in these stores would not have a smooth and efficient experience when renting home improvement tools, possibly causing them to rent elsewhere.

Via application and systems logs, the retailer's IT teams are using Splunk Enterprise to diagnose the behavior of applications. The results have been illuminating. Many custom-build applications are Java applications that access back-end databases. In the case of one problem application, 98% of its database calls were incomplete or failing, and it was making thousands and thousands of calls per minute to the back-end database. This slowed down the application itself, as well as other applications accessing the same database, resulting in poor user experience. Developers did not discover the problem in routine application testing and only saw it on the Splunk report and dashboard. With CIO-level visibility to such error reports, applications such as this were quickly recoded and most error-prone applications now have an error rate under 10%

"Splunk Craft" at Work—Deploying Splunk

How does the retailer deploy the Splunk platform? The architecture team utilized a fault tolerant storage array at each datacenter to run the Splunk solution, similar to other enterprise-scale software (see Figure 5). In the event of an outage at one datacenter, failover procedures will redirect traffic to the other.

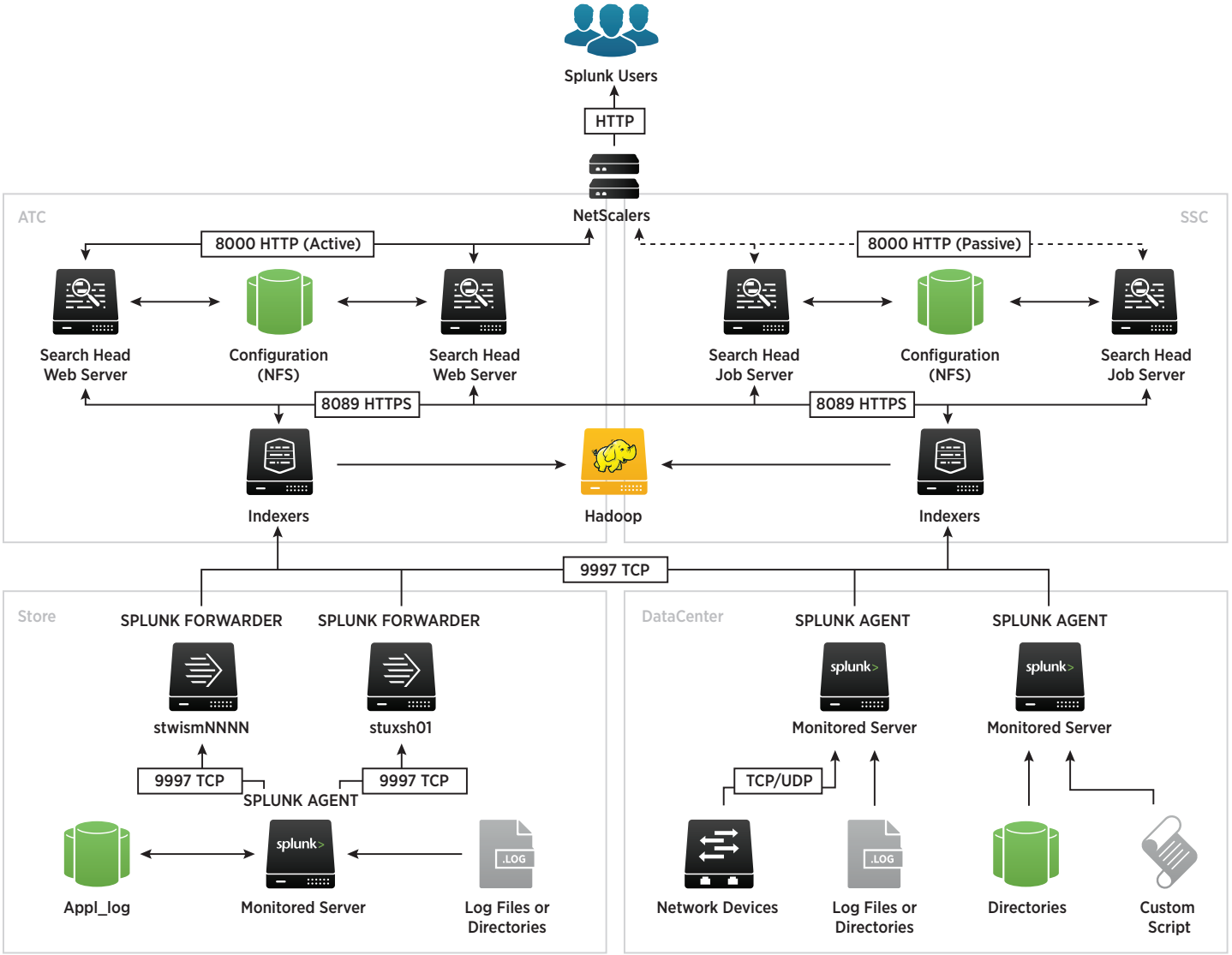

Figure 5. Deployment architecture.

All data is collected company-wide and indexed. Administrators keep the data needed for operational visibility in Splunk for 90 days, and store historical data within Hadoop for data mining, long-term trending and market analyses. Three search heads serve three core business domains—the IT operations team, the security team and the team responsible for e-commerce processes.

In addition to ad hoc queries when a system slows down or fails, search use cases include:

- Application response times
- Application requests per second
- Network bandwidth used between stores and datacenters
- Numbers and lists of users connected to wireless access point at stores
- Check the status of PC image replication at stores
- Get Citrix NetScaler statistics for detailed, aggregated reporting in Splunk Enterprise
- Get Tivoli Workload Scheduler logs to monitor nightly batch jobs run at stores.

Splunk deployment servers are used to manage configuration files on the Splunk instances. The primary deployment servers are located in the datacenters. The secondary deployment servers are deployed at each store location and service the forwarders in the store. A caching server between the primary and secondary deployment servers aids scalability.

The company uses a cookie-cutter approach for the IT systems in each of the 2,000 stores. Everything is identical. So a Splunk configuration change for forwarders will have to go to 2,000 stores. With the caching server in the middle, the primary deployment server can deliver the updated configuration package once, the first time, through the caching server and the caching server delivers the copy the remaining 1,999 times. Updates, as a result, are automated, saving substantial time and labor.

With Splunk, They Won the War Without the War Room

This use case examined how a national chain uses the Splunk platform to preserve the customer experience by ensuring customer-facing systems function effectively. It exemplifies how retail stores can no longer depend only on friendly, hard-working customer service reps to make sales. Today, informative and reliable point-of-sales and custom-order systems are critical to retail success. While millions of IT dollars have been spent on hardware, software, network and experts to harden these systems, the reality is that they are complicated and often siloed. When they fail, the downtime can result in the loss of thousands of sales and millions of dollars in revenue, not to mention the impact to the company's reputation. By extracting intelligence from machine-generated data, the Splunk solution can help resolve the issue faster, help prevent downtime and provide panoramic view of all systems involved. This use case demonstrated:

- **Elimination of data silos.** Because Splunk software indexes all kinds of data, the retailer's siloed data no longer impedes its ability to holistically view its systems to rapidly diagnose root causes of slowdowns and outages.
- **Correlations drive analytics.** Because Splunk software correlates different types of data, the company gains unprecedented visibility into its complex application stacks that ranges from panoramic, end-to-end views to granular optics into specific components.
- **Consistent customer experience.** The company not only has the insights needed to minimize the frequency and duration of application failures, it can proactively identify potentially troublesome components and take corrective measures, thus preventing future disruptions to customer service and revenue generation.
- **A vigilant eye on security.** The company uses the Splunk platform not only for operations and business processes, but also for detailed views into its security posture. By indexing and displaying data from systems involved with internal and perimeter security, administrators are aware of all defenses and are alerted to any threats.

CHAPTER 9

Solving the Compliance Challenge

How Splunk Software Is Used To Meet Audit Requirements And Prevent Insider Fraud At An International Bank

USE CASES

Fortifying Internal Security
Streamlining Regulatory Compliance

Executive Summary

You might expect that someone embezzling money, engaging in illegal stock trades or misdirecting funds would want to take time off to enjoy the bounty. Over the past decade, however, regulators discovered that several high profile, multimillion-dollar financial crimes had a common element; perpetrators of these crimes never took time off. They appeared to be model employees. They came in early and left late. They rarely took the vacation time that they had accrued. It turns out that in cases of financial fraud, perpetrators usually have to stay at work in order to deflect inquiries, hide the evidence and prevent others from noticing questionable transactions.

Because of this recent discovery, new compliance mandates have been enacted in most major economic markets around the globe, requiring that certain employees and contractors take at least two contiguous weeks off each year. The two weeks can begin and start at any time during the year, according to employee preference. During this leave, employees are not permitted to use a mobile device, laptop or tablet to log in to their work accounts in any way. While these employees are on vacation, auditors and management examine their electronic books and other work products in an attempt to discover any unusual activity.

In the U.S., vacation monitoring is a highly recommended internal control by the FDIC, the SEC and the FINRA. In Europe, vacation monitoring is mandatory; the European Banking Authority (EBA) maintains these guidelines. Because it is an unbroken block of time, this compliance regulation is known in various jurisdictions as "Mandatory Block Leave" or "Block Leave Monitoring." This regulation is so effective for averting fraud that national regulatory agencies have fined financial institutions millions of dollars in penalties for lax implementation of these internal controls. For example, in Hong Kong, one bank was fined $6 million by the Securities & Futures Commission for failing to properly implement Block Leave Monitoring that could have averted significant fraud.

When a European Banking Authority auditor contacted one Splunk customer to ensure that it was compliant to this new mandated control, the customer was deeply concerned. Although the customer is an international bank with billions in assets, a significant portion of its workforce did not use a consistent system to record time off. Vacation time was tracked by administrative staff, sometimes with pen and paper. With a global footprint, there was no standard method used between divisions or locations. Without enough time to change how vacations were tracked, how could the compliance team know which two-week period to analyze for "no activity"? And, how—and where—do you look for "nothing"? Employees might log on to a dozen systems a day, each system capturing the login credentials in a different way. It would be tedious and time consuming to look in the logs for every possible system for every relevant user and for every potential two-week timespan. Like proving a negative, this seemed like an impossible task.

The Information Technology team at the bank had recently replaced an existing SIEM with Splunk Enterprise. They had invested in training for key staff, and set up a Splunk Center of Excellence to share Splunk expertise and best practices. They continually discovered additional, valuable use cases for Splunk Enterprise that went far beyond the capabilities of their previous SIEM solution. Because of this expertise, the IT team quickly realized that Splunk software could be used to resolve this

Business Benefits at a Glance

Splunk Value	How Value Is Measured	Business Impact
State-of-the-art user interface and toolset	Speed to create meaningful reports for auditors; easy transfer of existing solutions to new use case	• $300K/year saved by avoiding the need to hire new personnel to meet compliance needs • The bank avoided millions of dollars in regulatory penalties
Holistic view of security; ability to ingest multiple types of data	Ability to see across multiple systems to detect and alert on unusual activity, revealing activity that had been masked by too much data "noise"	$1.5M/year staff costs redirected positively toward solutions rather than sleuthing
Ability to create alerts on high-risk behaviors	• Comprehensive visibility into fraudulent activity from its source to its intended target • Discovering and mitigating fraud before it is detected and reported to the public by regulatory agencies	• Ongoing loan fraud was detected before fraudulent loans were funded, preventing millions of dollars in lost funds • The bank avoided significant and costly damage to its reputation, which would have resulted in lost customers and decreased market share

compliance crisis. The team used the capabilities of Splunk software, adding key data to data that had already been ingested, as well as correlating relevant fields, in order to see activity across different systems by user. Block Leave Monitoring dashboards were created to reveal which employees had taken their two weeks off and when. The customer was able to quickly develop detailed, understandable, up-to-date reports for the auditor that demonstrated this compliance and prevented penalties. But Splunk software also exposed user activity beyond the compliance requirements.

Typical for the industry, the block leave search revealed that there were several rogue traders and rogue loan officers who were skipping their vacation in order to cover questionable trades or loan activity. But applying Splunk software also revealed this: some employees never seemed to log off, even those who could prove that they had been on vacation. Further investigation, made possible because of Splunk capabilities, revealed that employees had been using their vacationing colleagues' privileged IDs to misdirect funds, sometimes for years. Thanks to using Splunk software to look at both real-time and historical data, the customer was able to stop the fraud quickly, find out who was committing this fraud and learn the loopholes that they were exploiting to do so. Meeting the compliance needs and finding fraud prevented future legal issues, substantial penalties, large future losses and potential damage to the customer's reputation. And it didn't take a team of programmers to do it.

With Splunk software, the customer was able to:

- **Consolidate information from dozens of disparate systems to prove compliance.** Splunk software successfully found relationships across data from multiple systems, producing reports that either validated compliance or alerted staff that further investigation was warranted. Easy-to-build, easy-to-use dashboards presented a picture that would have been impossible to see any other way.
- **Detect and stop fraudulent activity.** By using Splunk software to compare machine locations of the valid employee with other login activity, compliance staff was able to hone in on the exact physical location of the fraud perpetrators.
- **Measure and mitigate fraudulent activity.** Once staff knew which internal devices and desktops to monitor, they used Splunk software to track exactly what the perpetrator was doing with the stolen credentials. They could use this knowledge to discover patterns, they could immediately shut down access and they would now have an understanding of what needed to be done to both stop ongoing damage and prevent future fraud.

Validating Vacations: Looking for Nothing

A government auditor demanded reports proving that one Splunk customer, an international financial institution, was compliant with Block Leave Monitoring. The Director of Compliance for the bank knew that he had no way to produce the reports because:

- For a significant portion of employees, administrative staff tracked vacation time manually, sometimes with pen and paper. Even for those using HRIS software to track time off, there was no standard method used between divisions or locations.
- Multiple systems had multiple logins; it was difficult to match employee name to the appropriate login.
- Several roles and employee levels were included in the compliance regulations, which meant that in the normal course of their work, employees used multiple systems for various functions, making anomalous behavior difficult to spot.

There are hundreds of sources of machine data at this global financial institution, with a handful available to each employee. Without the ability to change how leave was tracked historically, and no time to implement a new system, the compliance team was challenged to prove that they could analyze and report on "no activity" across multiple financial systems. Proving a negative is a difficult task. The Director of Compliance of the bank was desperate. Meeting the auditors' requests seemed impossible. He calculated that to validate Mandatory Block Leave for thousands of employees with the tools he had on hand, it would take new head count—half a dozen new employees—and thousands of tedious hours to organize and implement.

Enter Splunk

The bank's Director of Information Systems had recently used Splunk Enterprise to replace a Security Information and Event Management (SIEM) solution. The SIEM had not been able to handle the increasing volumes of data produced by the bank, but Splunk software had more than met this challenge, and the Information Technology group realized it could do far more. Specifically, it could be the solution to the Director of Compliance's crisis. Because the financial institution already had expertise with Splunk software, and had already ingested most of the machine data it needed into Splunk, the team was able to quickly determine how to create a Block Monitoring dashboard from a set of custom searches.

The first challenge was to correlate every possible way that a user might log in so that they could validate periods of "no activity" in all of these disparate systems for the specified employee. Since the field for capturing user identity was different in each system (for example, "username" in one, "userid" in another) the ability of Splunk software to correlate field names allowed the team to aggregate multiple login IDs into a single entity, revealing how any single entity was logging into and accessing any bank system. IT staff correlated user names from dozens of machine data sources, adding lookups to relevant content data sources for context:

- Email addresses from all mail and mobile systems
- Active Directory (to find those employees with the roles subject to the Block Leave requirement)
- Login names from all relevant financial systems
- Logins to all mobile and desktop systems
- ID badge swipe records.

To test the validity of the new dashboards, IT staff entered in vacation time for an employee that they knew was legitimately out during a specific date range. The results can be seen in the dashboard shown in Figure 1.

Next, they decided to find employees that did not have a break in activity. The customer used a search such as the one shown

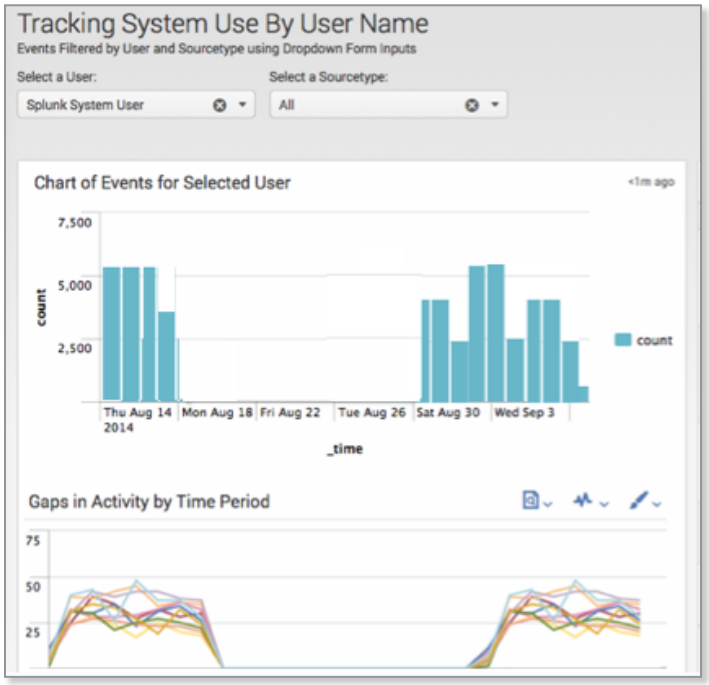

Figure 1. Validating an employee's Block Leave—"No Activity" (Each line represents a different system).

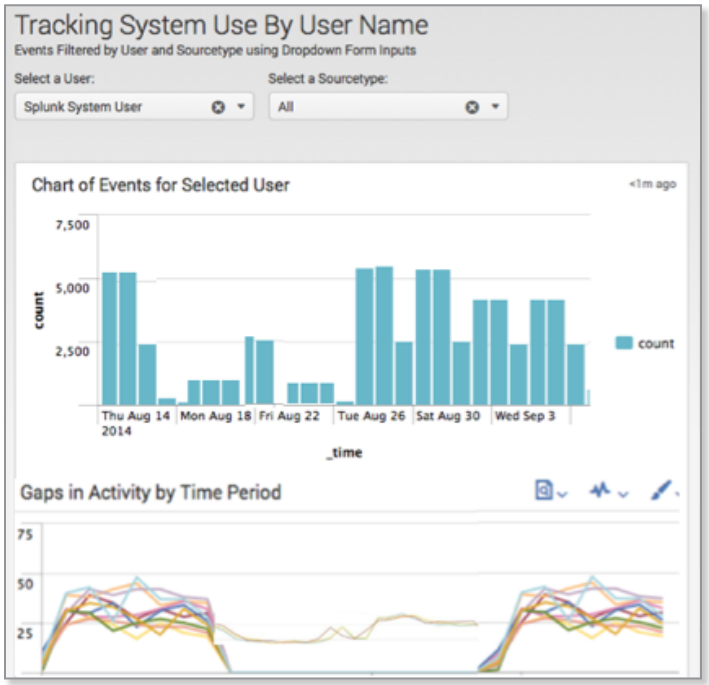

Figure 2. The Tracking dashboard for an employee who seemed to never fully disengage from the bank's systems.

below to obtain a list of users who never had more than a 14-day break between logins. In the example below, "userid" had previously been defined to the login field for each system. The result of this search would be a listing of users whose "userid" had not been idle for more than two contiguous weeks.

```
Sourcetype=logins_to_systems
| streamstats global=f window=2 current=t
    earliest(_time) as previous_login_time
    latest(_time) as current_login_time
  by userid
| eval time_between_logins=current_login_
time - previous_login_time
| stats max(time_between_logins) as longest_
break by userid
| where longest_break < (14*24*60*60)
```

The result for an example specific user is displayed in the Tracking System dashboard shown in Figure 2. Further examination of this "never vacationing" employee dashboard reveals some interesting anomalies.

As you can see in Figure 3, activity sharply dropped off for this employee during a two-week period, but it didn't stop entirely. There is a noticeable drop off over several weeks, except for access at a "normal rate" on some systems. While the original goal had been to find the few employees who had simply not taken the required leave, it was expected that this would mean that they were working continuously. However, this result indicates that employees were taking time off but continuing to access specific systems.

Armed with these results, Internal Auditing contacted the employees who had apparently never fully logged off, which did uncover a few cases of people who simply couldn't disengage even on vacation, whether for innocent or suspicious reasons. But seeing the results on the Splunk dashboard made it clear: there were imposters fraudulently using the IDs of traders and loan officers—year round! Because of Splunk software, the bank was able to:

- Meet compliance requirements by clearly validating and reporting which employees had not logged in during the required leave.
- Detect internal fraud and violators of the Mandatory Block Leave mandate by identifying those employees who were in fact logging in during their required leave and uncovering activity that these employees were trying to hide.
- Detect an unexpected type of ongoing fraud by discovering that imposters were fraudulently using the IDs of the vacationing employees—continuously, not only when the employees were on vacation.
- Clearly see which systems the imposters were targeting, providing clues to the motive behind the spoofing.

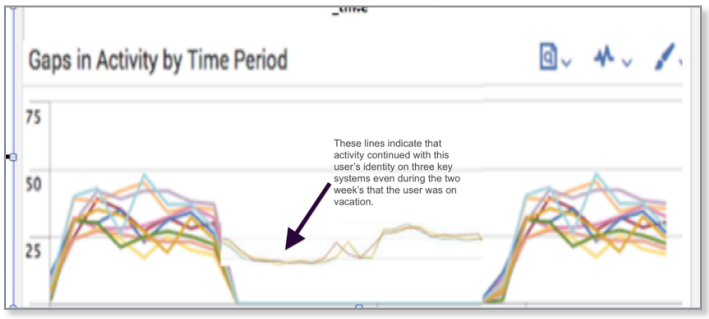

Figure 3. Magnified section from Figure 2. The arrow points to continuing activity despite the legitimate ID owner's reported vacation.

ID'ing the ID ring: When Crime Doesn't Play

The unusual intermittent access revealed that imposters were using the credentials of vacationing employees. The fraudulent activity was as clear as a Splunk dashboard and no longer buried in a fog of endless unstructured data. Until Splunk software revealed the fraud, it was masked by the legitimate activities of the rightful owners of the ID credentials.

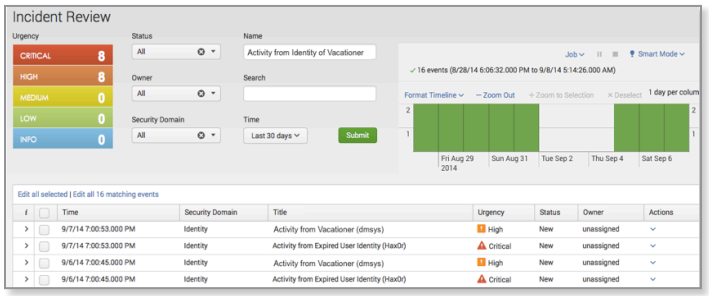

Figure 4. Incident Review from the Splunk App for Enterprise Security, modified to include Vacationing ID with other Identity Tracking.

Figure 4 displays an incident review dashboard that the bank configured to include suspicious use of vacationing employees' IDs with other suspicious identity events. Clicking on an event displayed the dashboard shown in Figure 5. When the IT team found a desktop or mobile device that had fraudulently used the ID of a vacationer (seen in the list in Figure 4), they used Splunk's drilldown capabilities to investigate further, by clicking on the line with the suspicious activity. This reveals event details, also shown in Figure 5.

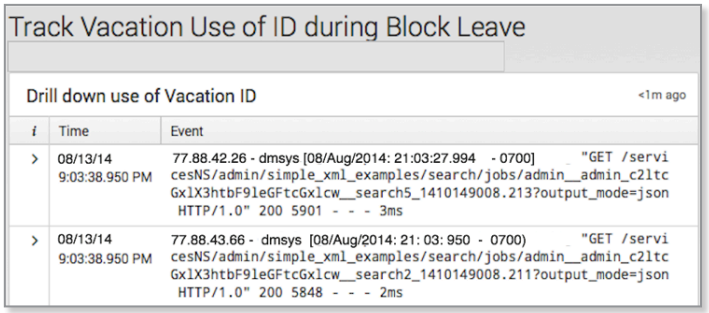

Figure 5. Drill down to discover the origin of the fraudulent access.

For an earlier security use case at the bank, Splunk software had been configured to ingest machine data related to the bank's physical assets, such as mobile devices, laptops, desktop PCs, etc. Correlations were created that tied each device's unique "MAC Address" and static IP address to valid users. The IT team leveraged this earlier work to pinpoint the identity of those fraudulently logging with vacationers' IDs. This took the investigators right to the desktop of the ID spoofer. Now the team could investigate further, discover which systems were being accessed with the ill-gotten IDs and determine the ultimate motive.

With the help of Splunk software, the team discovered:

- Which systems were fraudulently accessed
- How long this fraud had been going on
- Who was the likely perpetrator
- What systems they were trying to access

Instead of slogging through log files to validate compliance to Mandatory Block Leave, which would have taken many months and taxed its resources, the bank's IT staff was able to creatively use Splunk software to find the answers they needed, in weeks. And because of Splunk software, they had visibility into additional suspicious activity that had previously been masked by a deluge of data.

Discovering the Motive: Unmasking the Loan Arrangers

To prevent an employee who was reviewing a credit application from being more generous than rational, the bank had a complex multi-step approval process requiring multiple electronic sign-offs for mortgages or credit cards, unusual requests and other consumer loans or investments. In a simplified example, employees at Level One would do an initial screening, then electronically pass the file up to Level Two (a supervisor) for a second approval. The Level Two approval process was more robust and designed to filter out risky loans. But Level Two also had more authority and could be used to grant exceptions to loan applications for legitimate reason or with irregularities.

In the course of complying with Mandatory Block Leave, the customer discovered that some users never logged off, even during vacation leave. They soon discovered that the vacationing user's ID was being fraudulently used. Further sleuthing with Splunk software revealed why: several Level One approvers had somehow learned their Level Two approvers' login credentials. The Level One approvers were using the stolen Level Two login to sign off on financial transactions that their supervisor never saw.

This type of fraud had the following impact:

- Granting loans without the proper procedures caused instability and exposed the bank to severe penalties from governments in every jurisdiction.
- Risk-mitigating steps that were normally taken to ensure that certain loans were a good bet for the bank were being totally ignored.
- Loans were granted to non-existent/fraudulent accounts, making them impossible to collect.
- Credit limits were increased on existing accounts that were not qualified to get the increased credit.

The fraudsters had been using the stolen IDs to sign off on transactions for years. The fraud ranged from serious multi-person rings processing hundreds of thousands of dollars worth of loans, to occasional perpetrators, such as the employee who ensured she would be popular at the next family reunion by upping the credit limit for relatives who did not have the required income. Or the employee who reveled in the recognition he received because he successfully processed more loans than his coworkers who followed the appropriate procedures. All of the fraud, of any size, put the bank's reputation and assets at risk. Finding it quickly, before external auditors or the media discovered it, was paramount. Based on compliance alone, bank management stated that Splunk software paid for itself almost as soon as it was launched. But because Splunk software also revealed significant fraud, Splunk software positively impacted the bottom line.

Using the Splunk App for Enterprise Security

To meet the goals of the audit, the Information Technology team was able to accomplish everything it needed by using the dashboards and searches they created with Splunk Enterprise. However, when the bank subsequently added the Splunk App for Enterprise Security (Splunk App for ES), the team realized that this app could be used to place identity and asset management—including Block Leave Monitoring and other compliance requirements—into a larger security context. (The screenshots shown in this story show modifications to the Splunk App for ES.) For example, as shown in Figure 6, the Identity Notables dashboard from Splunk App for ES was designed to alert whenever an ID of a terminated employee logs in or when a non-authorized user attempts to access privileged systems.

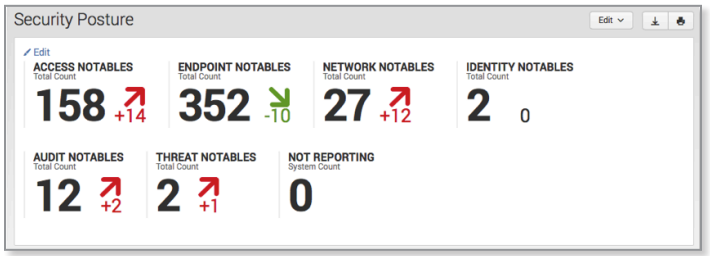

Figure 6. Vacation Identity Tracking was added to Identity Notables, and now it could be seen as part of the entire security picture.

To build on these capabilities, and to take full advantage of the Asset and Identity Center features provided by the app, the bank tailored identity dashboards and searches to include "Vacationing Employee Identity" activity tracking for relevant employees. At first, the bank's IT team added a control to make it easy to manually enter vacation time as reported by staff; later, the team devised a way to collect and feed the vacation dates in via a spreadsheet lookup table they had built. Finally, the bank required vacations to be entered into the Payroll/HRIS systems so that vacation dates of relevant employees could be automatically monitored for ID activity when the employee was on mandatory leave.

With Splunk, the Bank's Investment Paid Dividends

In this use case, we explored how Splunk software provided audit-quality visibility into bank operations, enabled easier reporting and provided unprecedented insight into day-to-day operations of key employees, causing the bank to discover fraud that had been concealed by a mask of disconnected data.

This use case demonstrated:

- **Making sense of disparate data.** Because Splunk software indexed data from multiple login IDs across multiple systems, the bank saved months of time that would have been required to manually comb through data, and was able provide real-time information to regulators, providing effective compliance verification to meet internal and external governance demands.
- **Solving big problems with easy-to-build, easy-to-use dashboards.** Vacation monitoring, which first seemed nearly impossible, became easy with targeted dashboards that the team created in Splunk Enterprise. When the customer later implemented the Splunk App for Enterprise Security, the IT team incorporated vacation monitoring into the app's identity monitoring dashboards, creating a holistic view of identity management tailored for the bank's compliance needs.
- **Knowledge and control.** This use case illustrates how Splunk software can use drilldowns to get to the source of a problem, then use these insights to create relevant alerts and stop issues before they happen.
- **Value generation across multiple use cases.** In addition to creating quick compliance reports, the customer discovered serious ongoing fraud. Prior to this, the company had more than recouped its investment in Splunk by gradually phasing out SIEMs and replacing them with Splunk software.

By building on its IT team's Splunk platform expertise and maximizing data that had already been ingested, this customer realizes escalating value from its Splunk implementation. It's like buying solar for your house and then realizing you have enough capacity to plug in a car. Once you've paid for the investment, the rest is profit. The more use cases you give Splunk software, the more it gives you back.

One Splunk. Many Uses.

While the business problems discussed in this case were specific to this customer and its industry, and the solution made creative use of Splunk software's features to solve these particular problems, the underlying theories apply to many business use cases.

With the right data, Splunk software can quickly find why website visitors are encountering difficulties, historically or as they happen. Once staff is alerted to these difficulties, Splunk dashboards lead to their source, exposing issues and leading to solutions. Finding and resolving these issues leads to increased customer retention and increased profits. While the possibilities are endless, the process is simple.

Next Steps

To learn more about Splunk customer success, customer snapshots, ROI stories, customer profiles and more, please visit: http://www.splunk.com/view/customer-case-studies/SP-CAAABB2

Splunk software is also available as a free download. Download Splunk Enterprise and get started today: http://www.splunk.com/download

If you would like to speak to a salesperson, please use our online form: http://www.splunk.com/index.php/ask_expert/2468/3117

CHAPTER 10

Business Insights On-the-Fly

How Splunk Software Helped a Newly Merged Airline Take Off

USE CASES

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do

Executive Summary

Imagine that your company's revenue depends on a complex, ever-changing network of variable prices, discounts, fees, regulatory requirements, taxes and international politics. There is a finite amount of product that you can sell, and the price of this product changes daily due to competitive pressure. Imagine forces beyond your control affecting the demand and supply of your product. A volcano, a tornado, a hurricane, a snowstorm—literally anywhere on the planet—and everything can change at once. You have entered the world of the airline industry.

One Splunk customer, a major international airline, had recently merged with another, forming a new airline with over $5 billion in annual revenue. The newly formed airline did not have visibility into which of its flights were earning money and which were losing money. It had routes around the globe, with no insight into the best path to profits. Worse yet, customer satisfaction was rapidly plunging. The airline's website, one of the most heavily trafficked in the world, was losing customers—pages took a long time to load and bookings would disappear right before a potential passenger could confirm a purchase. As a result, would-be passengers were taking flight—as in, fleeing the site before booking—and the airline's analytics team could not determine why. Revenue was decreasing. Market share was falling. The number of calls to customer support—and the time it took staff to resolve problems—was steadily increasing. All too often, the support team could not mitigate a problem in time to save the booking. Seats were empty that could have been filled, unrecoverable profit flying away forever. The airline needed real-time, actionable visibility across multiple systems in order to intercept problems before customers disembarked from the website, as well as insight into customer behavior to maximize the value of each passenger trip.

Once the airline deployed Splunk software, it gained the visibility into patterns of activity across airline-specific logs and other homegrown machine data needed to diagnose and repair the issues that were leading potential passengers to leave the site. Because Splunk software can follow transactions across virtually disconnected systems, it is now possible for the airline to answer questions that impact customer satisfaction such as "Why did this potential passenger drop off while a flight was in the shopping cart, ready to be purchased?" and questions that can lead to higher revenue such as "Which flights are making the most profit, and how can we target potential offers to passengers most likely to choose them?" As a result, the airline received the operational visibility it needed to resolve its website issues. Moreover, it developed the tools to determine the most profitable flight paths, and also to shape customer behavior toward the most profitable choices. Splunk software enabled insights to materialize from data that had been impossible to understand, insights so critical to the airline that its initial investment was recouped in the first four months. The return on investment reached 300 percent by the end of the first year.

Splunk software helped the customer gain:

- **Digital customer experience insight.** Because Splunk dashboards provide visibility to see website performance and troubleshoot issues in real time, the airline analytics team continually improves the site. Improved uptime and response time result in increased user engagement. The airline has a pilot's view of its website infrastructure, improving the monitoring of applications and critical processes, and determining the fastest flight path to profits.

Business Benefits at a Glance

Splunk Value	How Value Is Measured	Business Impact
Quick insight into website issues that halt sales	Improving look-to-book ratio for website	60 percent improvement in converting site visitors to ticket purchasers, increasing revenue $1.3M/year
Detailed user and usage analysis of customer sessions	• Customer retention and adoption rate • More effective website design	Improved site experience and availability; improved customer retention and adoption rates, increasing revenue $2.9M/year
Rapid incident investigation and root-cause determination	Customer support call escalations	Reduced escalations by 80 percent, saving $800K/year
Sophisticated real-time correlation of clickstream and other data	• IT costs for problem resolutions • Customer support calls and research time for disputed and fraudulent transactions	• Problem resolution costs were reduced by 50 percent, saving more than $1.2M/year • Eliminated 50 percent of inbound phone calls regarding disputed or fraudulent transactions

- **Accelerated troubleshooting.** Splunk dashboards cut the cost of avoiding and resolving problems in half, improving customer satisfaction and retention, and leading to repeat sales. The airline's analytics team now receives real-time alerts for problems so that they can intercept them before the client ever sees the issues. In a business where customers can change providers with the click of a mouse, real-time corrections prevent $millions in lost fares.
- **Business analytics across the board.** Splunk software helped the airlines on both sides of the merger get insights into critical aspects of the newly joined entity. Splunk dashboards reveal which flights make the most profit, which promotions work and strategies that can increase revenue.

Unneccessary Losses

In most industries, if you miss an opportunity to sell your product, you can recoup some of those losses. Perhaps you can sell them at a discount or you can give them away in a promotion. But there is no way to recoup the value of an empty airline seat on a flight that has already flown. And, a customer lost because of one frustrating web experience could become a competitor's customer forever.

The newly merged airline had many customer issues related to the merger: escalating support phone calls, dropping revenue, and rampant negative comments in social media. But the airline had no way to diagnose what was causing all of these problems.

The airline's two major challenges revolved around selling every seat. First, the airline didn't want to lose a sale that was sitting in a shopping cart because a technical or user experience issue caused the purchaser to give up. And secondly, the airline wanted to send passengers to their destination along the path that was the most profitable. But the airline had two problems:

- **There were multiple front-ends, silos and systems behind the ordering process.** The airline had multiple websites that were customized to language preference. There were dozens of systems recording web clicks and determining user path, working with airline-specific data—too many places to look to find out why customers were dropping off the website.
- **Lack of insight into flight profitability.** When a passenger enters "origin" and "destination" during a booking process, the most profitable approach for the airline is to offer fares that combine the best revenue for the airline (good for profit margins) with the best value for the passenger (good for passenger retention and adoption). But fuel prices, local tariffs, competitor prices and allowable fees shift daily in this industry. The airline had no real-time visibility into which flights were most profitable at a given point in time.

It was clear that visitors were being successfully attracted to the website, but leaving without purchasing flights. The airline suspected that speed or stalled processes might be the culprits, but the only way to understand exactly what was happening was to analyze log files from many different systems, including Sabre, an airline industry system. The first goal was to identify which process or step consumed most of the time of the stalled transactions. But to do this, the airline would have to see patterns from all of the different logs. This was not only a technical issue, it was also an organizational issue. Systems that interacted on the website had totally unrelated purposes and totally unrelated log file structures—and they were managed by different organizations within the airline. If one team needed the logs of a system managed by another, it could take days to get access to the appropriate logs. Even knowing how to frame the question was difficult—how could you be sure to get the logs you needed? And diagnosing anything in real time—in time to prevent unsold passenger seats? Impossible.

Enter Splunk

When the team discovered that Splunk software could ingest log file data from different systems, they realized that the platform was the answer to their two biggest challenges—optimizing the website to improve customer retention and enhancing per-trip revenue. They immediately signed up to explore a trial version and realized that it was the answer to every merger's dilemma—how to make sense of disparate systems with intersecting processes but disconnected log files. After seeing what Splunk software could do, the airline decided to move forward. Suddenly, due to Splunk's schema-on-the-fly, patterns were visible that had been invisible before. No one had to wait days to see what was happening in another part of a process.

Stopping the Flight of Website Visitors

Splunk software ingests log files from Sabre, log4j, Apache, and other homegrown application and business systems, and correlates similar fields between different systems. With Splunk Enterprise, the airline's analytics team ties together all of the pieces of a transaction in a way that makes them visible. Now they can measure total end-to-end transaction times as well as times for a given segment.

Figures 1 and 2 show snippets of log files. The Sabre log file (Figure 1) records one part of a customer session, while a homegrown log file records another part. Without Splunk software, making a connection between the two files would be time consuming and labor-intensive. Making a connection between thousands of such files at once, in real time, would be impossible.

The customer can ingest dozens of logs like those shown in Figures 1 and 2 into Splunk Enterprise to be indexed. By correlating related fields so that all relevant information about each customer session is available, the airline could see the entire picture of what happened to customers from the moment they first clicked on the site through the time they purchased a flight—or left to visit a competitor. Splunk dashboards also revealed big-picture information, cutting through the details to reveal which issues were causing the most customer frustrations.

For example, like most global companies, the airline personalizes web experiences based on language preference or point of origin. Until Splunk software arrived at the airline, this factor was just another variable that the airline suspected impacted customer experience, but could not know for sure. For example, certain marketing promotions seemed to work wonderfully on one site, but not at all on another, even though other research had revealed sufficient demand. After deploying Splunk software, the airline found that the

```
user=
apid=
reqid=hg03b20193336798&8741.667008
sid=xy2464613679856&1_J17IUKPUAR
Executing Query: INSERT INTO booking_engine_var.sabre_vcr5_info
(id_cashier, holding_client, pnr)

level=WARN
dom=.SplunkAirlineCustomer.com
log_order=159
script_name=cashier_notification.cgi
url=/cgi-bin/cashier/notification.cgi
user=
apid=
reqid=hg03b20197136798&8741.667008
sid=xy2464613679856&1_J17IUKPUAR
Ticketer - getVcr5List - Missing holdingClient for the cashier [2&87469231]
```

Figure 1. Example Sabre log displays an error that prevents a successful purchase.

```
1.55 0.00 (1333745259 &93359): $VAR1 = {
          'GAConfigAndVar' => {
                                'Config' => {
                                              'hide_component' => 1,
                                              'step' => 'step2',
                                              'application' => 'buy',
                                              'debug' => 1,
                                              'class' => 'Purchase',
                                              'mx_content_cap' => 'Rd_$1'

'Description' => 'Google_Analytics::Collector->should_hide_component',
                                              'Var' => {}
                                            }
        };
1.55 0.00 (1353745859 &94989): clicktale_start component
1.55 0.00 (1353745859 &95202): clicktale_start component draw_to_buffer
1.56 0.00 (1353745859 &95664): Application::Session - Session Manager user
starts
1.56 0.00 (1353744859 &97818): Application::Session - Loading Parameters...
```

Figure 2. Example custom log reveals the time it takes to complete a process.

high unsold seat rate could not be fixed by tweaking the marketing campaign; it found out that there was a remarkably different purchasing experience for customers around the globe. Figure 3 illustrates that customers in the process of purchasing a ticket spent many minutes more on the transaction depending on the website interface they accessed.

The color of the bars indicates trips with the same origin-destination (OD). The dashboard illustrates that passengers who purchased a ticket using the German-based website spent six minutes on the purchase, while passengers using the Spanish-based interface to the website spent 17 minutes for the same purchase. This helped the business understand a key reason why certain websites were selling more tickets than others, something that had seemed inexplicably unrelated to marketing campaigns. The airline put resources toward fixing the underperforming sites immediately. Customer support calls began to drop quickly and the customer experience between globally customized websites improved.

It is helpful to know where you have problems in your customers' experience, but what if you need more information about why the problems are happening in order to fix them? Figure 4, an "Unsold Passenger Seats" dashboard, breaks down the reasons for delays and unsold seats in even more detail. Now the airline can understand the scope and causes of the problem of unsold seats, organized by OD. This dashboard, designed for the operations group, reveals reasons that seats were left unsold in a way that makes it possible to correct the problems systematically and strategically. Here are some of the issues that Splunk dashboards revealed were causing customers to leave the site:

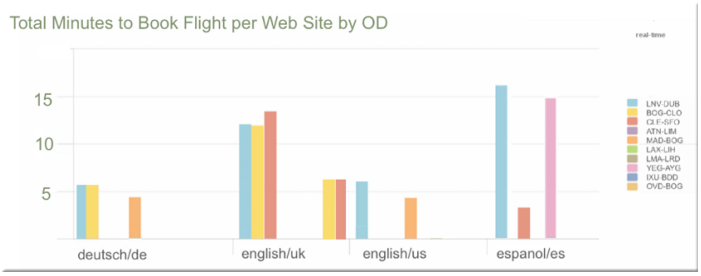

Figure 3. Time differences for similar trips based on geographic front-end.

- Flight unavailable after clicking through all the choices
- Dates selected early on in the process were in fact not valid
- Website dropping customers mid-order
- Frequent stalls and hangs.

Splunk dashboards revealed the impact, in unsold seats, of website errors and issues. Once the team could see where the issues were, they were able to triage on the websites with the most expensive problems, turning around the losses quickly in time to sell the empty seats. Also because they had real-time insight to problems, they could alert the business to tie a particularly empty route to a web-based promotion after fixing issues that had stalled purchases of that flight. This way, if a flight was undersold due to a web issue, the business learned about it in time to quickly devise a web-based promotion to sell the seats that would have otherwise flown empty.

Learning the exact nature and location of the website issues allowed the teams to drill down through Splunk dashboards and events to discover root causes. Within the first three months of deploying Splunk Enterprise, dropped sessions had drastically decreased, and customer satisfaction and retention had already improved. By the end of the airline's first year using Splunk software, calls to support and expensive escalations had dropped off by 80 percent. Airline business analysts attributed over $3M in additional revenue to Splunk software, just in the first year of use.

Figure 4. A comprehensive look at causes and impact of unsold passenger seats.

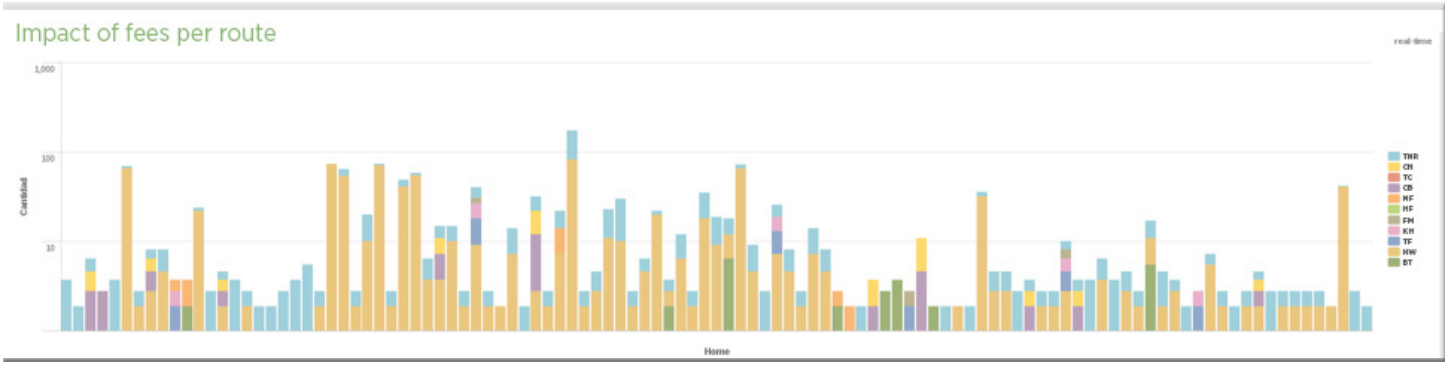

Figure 5. Splunk dashboard illustrating impact of fees and taxes on revenue per route.

Solving the customer retention problem was the first challenge that the customer addressed with Splunk software. Now that the airline had made a significant difference to this challenge, it wanted to address the other challenge: discovering the most profitable routes to promote.

Calculating Profit Margins Was Taxing

For international flights and even some domestic ones, each time an airliner lands or takes off, it is subject to taxes and fees. Some of the fees are charged directly to the passenger based on class of service, but others are assessed to the airline directly. Some fees are based on the airspace through which a flight passes; other fees are assessed based on airports where the plane lands and the time it stays at the airport. Rerouting connecting flights through alternative airports can make the difference between profit and loss.

The airline realized that it could not figure out the actual profit or cost from flights because the information was not easy to see across a widely dispersed and disparate infrastructure. The relevant information changed frequently and came from different parts of the organization. The airline's executive staff didn't know which flights to promote and which flights to consider rerouting or canceling. It was becoming impossible to figure out how profits worked out per flight segment, per consumer, per flight.

Once key IT staff was freed from endlessly fighting support issues, they had the time to look at the problem of taxes and regulations. To find out more about how taxes and regulations impacted flights, web analytics staff asked IT staff to ingest data into Splunk Enterprise that would give them specific details about the fees that might apply to travel for the airline's numerous routes. In partnership, the two teams designed meaningful dashboards that explained where revenue was coming in at every level of drilldown that the business needed. Figure 5 shows one Splunk dashboard that the airline created to show the effects of taxes based on route traveled.

Business staff and business executives can now find out the value of each trip, or see the big picture about the true costs of channeling traffic through a specific hub. This has helped them refine pricing and promotions, so that the flights that appear first to a consumer searching for fares on the website are the ones that are the best value for both the airline and the consumer. The airline can be more competitive and build loyalty by offering better pricing to its customers.

Based on the most profitable route, the business now designs promotions to persuade customers to select routes that pay off for everyone—in real time, on the fly! It is often a win-win, as customers may have no preference about the location of a stopover, but causing a fuller flight to go to the stopover with fewer fees can increase profits for the airline or make the price more competitive for the passenger. The airline now has the insight to set different fares, manage promotions that maximize profits and increase revenues to the airline. This also gives business executives at the highest level the opportunity to add critical new variables to big picture decisions. They can use insights from machine data to optimize flight plans, determine new markets to enter, new alliances to forge, and which offerings to drop, refine or promote. This visibility has replaced uncertainty with a clear competive advantage.

With Splunk, the Acquisition Took Off

In this use case, we explored how Splunk software brought clarity to the disparate systems that resulted when two major airlines merged. When executives first proposed the airline merger, they saw great promise because it vastly, and literally, expanded the reach of both airlines. The merger held the promise of simplifying complex trips, for both domestic and international travelers. But at first, before Splunk software entered the picture, the merger only led to plunging customer retention, rising support costs and lack of visibility into the true cost of passenger trips. When Splunk software was deployed to address these issues, the synergy that the executives had originally envisioned began to be a reality.

The airline realized that the value it obtained from Splunk only increased the more data was ingested—the amount of ingested data doubled after the first year. Payback for the Splunk investment was calculated at three months. Return on investment has already reached $5 million in the first year of deployment, and is projected to be over $15 million by the third year.

This use case demonstrated:

- **Elimination of data silos**. Because Splunk software could index so many types of data without the need to alter it, there was no need to phase out systems just to get clarity. This made merging the information technology departments much easier.
- **Correlations drive analytics.** Because Splunk Enterprise correlates so many types of machine data, the customer could follow transactions across different systems to get a true understanding of the customer experience.

- **Flexibile analytics powered by a read-time schema.** When the website logs were designed, they were focused on capturing the information it took to obtain a ticket and close a sale. But when the analysts looked at the data through Splunk dashboards, they could search it in novel ways that revealed insights.
- **Value generation across multiple use cases.** In addition to significantly improving MTTR and customer website experience, Splunk Enterprise allowed the airline to determine the best route to increased profits.

One Splunk. Many Uses.

While the business problems discussed in this case were specific to this customer and its industry, and the solution made creative use of Splunk software's features to solve these particular problems, the underlying theories apply to many business use cases.

With the right data, Splunk software can quickly find why website visitors are encountering difficulties, historically or as they happen. Once staff is alerted to these difficulties, Splunk dashboards lead to their source, exposing issues and leading to solutions. Finding and resolving these issues leads to increased customer retention and increased profits. While the possibilities are endless, the process is simple.

Next Steps

To learn more about Splunk customer success, customer snapshots, ROI stories, customer profiles and more, please visit: http://www.splunk.com/view/customer-case-studies/SP-CAAABB2

Splunk software is also available as a free download. Download Splunk Enterprise and get started today: http://www.splunk.com/download

If you would like to speak to a salesperson, please use our online form: http://www.splunk.com/index.php/ask_expert/2468/3117

Find Out More

To learn more about Splunk customer success, customer snapshots, ROI stories, customer profiles and more, please visit: http://www.splunk.com/view/customer-case-studies/SP-CAAABB2

Splunk software is also available as a free download. Download Splunk Enterprise and get started today: http://www.splunk.com/download

If you would like to speak to a salesperson, please use our online form: http://www.splunk.com/index.php/ask_expert/2468/3117

Index

250 Brannan St., San Francisco, CA 94107 info@splunk.com | sales@splunk.com 866-438-7758 | 415-848-8400 splunkbase.splunk.com

splunk> listen to your data™

www.splunk.com

Item # CS-X-Splunk-ProfileBook-118 HY