

Architecting Splunk Enterprise Deployments Lab Exercises

Background

You are a consultant who has been hired to help Buttercup Games implement Splunk. You will be meeting with Pat Livingston, the director of IT Operations at Buttercup Games, who is serving as both the project manager and the sponsor. During each lab exercise, you will be given some Q&A to increase your understanding of the company.

Module 1 Lab Exercise: Getting Started

Description

The goal of this lab exercise is to familiarize you with the Buttercup Games infrastructure and the planning tools available.

Throughout this course, you will be asked to complete small portions of the deployment design. For the final lab exercise, you will submit these exercises to your instructor via email. You can complete the exercises in a single document or use multiple files.

You may complete your exercises in any tool that you wish, including pen and paper. However, you will need to convert your exercises into a form readable by your instructor. Allowable formats:

- PDF
- Rich Text Format (RTF) or Microsoft Word
- Microsoft PowerPoint
- Microsoft Excel
- JPEG (for diagrams)

Steps

Task 1: Download the planning tools.

- Data Source inventory (data_inventory.rtf)
- Index Planning and Sizing (data_sizing.xlsx)
- Splunk Icon Library (SplunkIconLibrary.ppt)
- Use Case checklist (use_case.pdf)

Task 2: Read initial interview questions.

Q: Can you provide some **background** on your company?

A: Buttercup Games is a privately held company headquartered in San Francisco, CA with offices in Boston, MA and London, England. Our product line encompasses a variety of games and accessories. We sell our product through two channels: directly through an online e-commerce website and through a worldwide chain of independently owned and operated “brick and mortar” retail stores.

Q: What are the goals for the deployment?

A: We want to:

- Consolidate syslog and Windows-based logs into a single integrated data store.
- Correlate user activity and access management across all logs.
- Create ad-hoc and routine reports to meet our needs.
- Be able to grow the environment easily.
- Provide a high quality service to our users, with data integrity and minimum downtime.

Q: What are your **motivations for this project**?

A: Here is some history in regards to the project. We instituted our current logging infrastructure to meet the PCI compliance requirements requested by our electronic payment partners. The logging system that was implemented to address those requirements is now at the saturation point. We are running out of resources to effectively store and analyze our logs. This is directly affecting our ability to monitor our operational, security and compliance logs. We have decided that an integrated approach is required to provide a solid foundation for our enterprise operational data.

Q: What is your **current IT environment**?

A: Our core IT systems are:

- Email is Microsoft Exchange using an Internet-based hosting service.
- HR, ERP and CRM systems are all cloud-based.
- The San Francisco, Boston and London offices are part of the enterprise VPN.
- The online store is a combination of Apache web servers and a homegrown e-commerce system, all running on Linux.
- We use Active Directory user authentication.
- File servers (Linux) are used to store log files as well as for file sharing.
- Symantec Endpoint Protection is used on all of our Windows servers.
- Each of our three offices has a badge reader system.

Exhibit 1: IT Environment

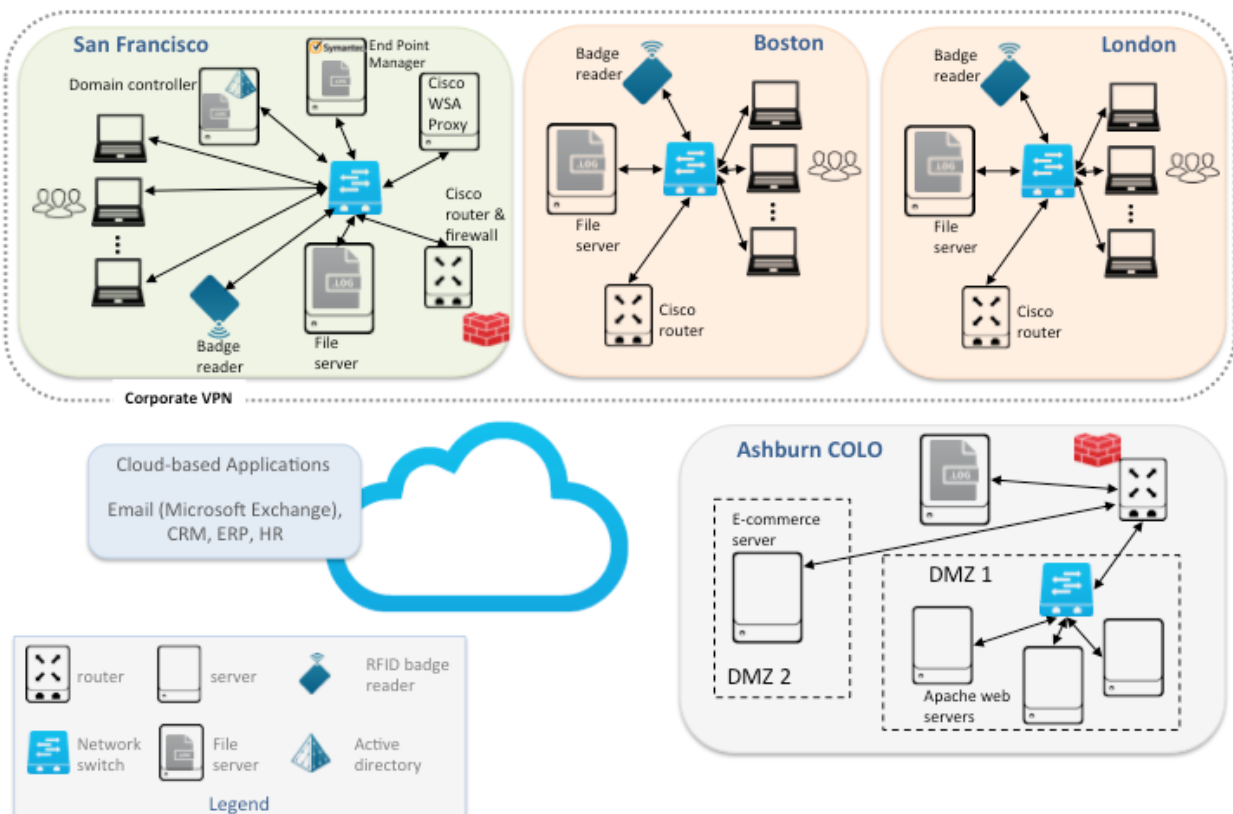
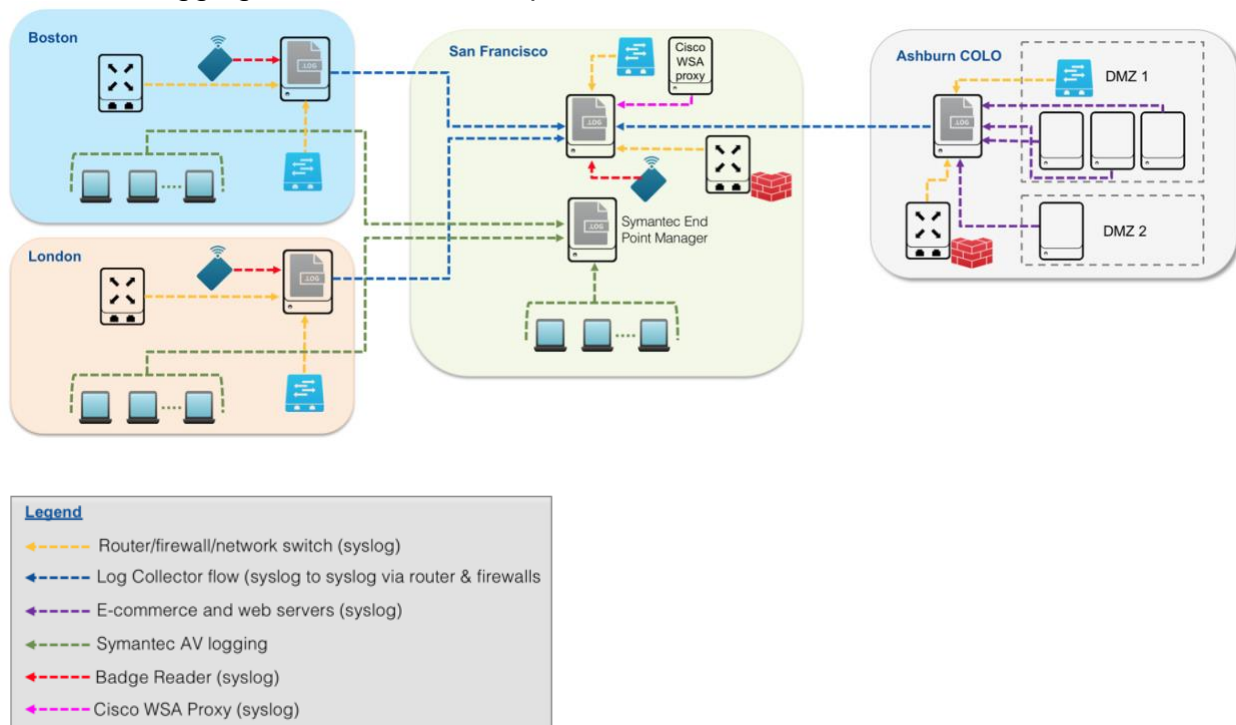


Exhibit 2: Logging Environment – Corporate VPN



Module 2 Lab Exercise: Project Requirements

Description

Using the interview as a guide, complete the Data Source Inventory. To expedite the process, we will use an abbreviated form of the inventory that is already partially completed.

Q: What is your **current logging infrastructure**?

A: The current environment is split between two logging formats: **syslog** and **Windows**. The following systems feed into a Linux-based log server, either directly or through an intermediate log collector using syslog:

- Badge reader log
- Cisco WSA Proxy
- Cisco ASA Firewall
- e-commerce
- Webservers
- Linux OS

The logs generated by the Windows servers and Active Directory are logged as Windows Events. For anti-virus software, we have Symantec Endpoint Protection; the Symantec Endpoint Protection Manager runs on a Windows server.

Q: What logs will be ingested into Splunk?

A: The **data sources** will include the firewall logs from the Cisco ASA devices, Linux logs, webserver logs, Cisco WSA proxy log, Window logs, Symantec logs and e-commerce transaction logs.

Q: What are the **retention periods** for each of those logs you mentioned?

A: The e-commerce transaction logs must be accessible for one year. The WSA proxy logs are kept for only a month. The badge reader logs are available for 6 months. Everything else is kept for 3 months. We would like to keep all the logs much longer, but we don't have a large budget for storage.

Q: What **levels of access** will be required to the data?


A: I will start with the easy one first: the Security group will require access to all logs. The IT Operations group will need access to the Cisco WSA proxy log, Linux logs, Cisco ASA firewall logs, Windows logs, Symantec logs and webserver logs. Our Sales Support group will need access to only the webserver and e-commerce logs.


Steps


Task: Complete the data source inventory.

For each data source, complete the following information. Note that you may not have all the information for all sources at this time. You may make assumptions or leave items blank.


Data source 1	Cisco ASA firewall logs
Average data per day	
Retention	90 days
Visibility	Security, IT Ops
Collection method	UDP:514 maybe? (It is okay to leave this blank if you are unsure)
Location	San Francisco, Ashburn COLO


Data source 2	Badge reader log
Average data per day	
Retention	
Visibility	
Collection method	
Location	San Francisco, Boston, London


Data source 3	Cisco WSA Proxy
Average data per day	
Retention	
Visibility	
Collection method	
Location	San Francisco

Data source 4	Linux Logs
Average data per day	
Retention	
Visibility	
Collection method	
Location	All locations

Data source 5	Windows Logs
Average data per day	
Retention	
Visibility	
Collection method	
Location	San Francisco

Data source 6	Symantec logs
Average data per day	
Retention	
Visibility	
Collection method	All end user systems interact with the Symantec EndPoint (EP) manager
Location	San Francisco

Data source 7	Web logs
Average data per day	
Retention	
Visibility	
Collection method	
Location	Ashburn COLO

Data source 8	e-commerce logs
Average data per day	
Retention	
Visibility	
Collection method	
Location	Ashburn COLO

Module 3 Lab Exercise: Index Design

Description

The goal of this exercise is to identify the indexes needed and to estimate the disk space required for Splunk. Here a list of the daily volumes for each data source, but in real life you might collate this information from the data source inventory.

Q: Do you have an estimate of the **daily amounts for your data sources**?

A: Yes, let me give you a list.


Source	Daily Amount
Cisco ASA firewall logs	25 GB
Badge reader log	15 GB
Cisco WSA Proxy	30 GB
Linux logs	20 GB
Windows logs	35 GB
Symantec logs	5 GB
Web logs	110 GB
e-commerce logs	75 GB

Steps

Task: Estimate the disk space requirements.

Assume the following compression factors for all of the data sources:

- 15% compression for rawdata
 - 35% compression for index files
- Now that you have the daily ingestion values, return to lab 2 and fill in the *Average data per day* fields.
 - Estimate the size of each data source on disk.
(You may choose to use the `data_sizing.xlsx` handout to complete this task.)
 - Identify the indexes that will be needed based on the criteria from this module and the information in the data source inventory. Complete the last column in the table below, choosing the names for the indexes.

Data source	GB per day	Retention (days)	Est. size on disk	Visible to	Assigned to index
Cisco ASA firewall logs	25	90		Security, IT Ops	
Badge reader log	15	180		Security	
Cisco WSA proxy logs	30	30		Security, IT Ops	
Linux logs	20	90		Security, IT Ops	
Windows logs	35	90		Security, IT Ops	
Symantec logs	5	90		Security, IT Ops	
Web logs	110	90		Security, IT Ops, Sales Support	
e-commerce logs	75	365		Security, Sales Support	
Total					

4. How many indexes will be needed? Why did you choose this number?



5. What is your estimate for the Splunk license? Why did you choose this license?



Challenge Task: Identify potential apps.

Examine Splunkbase (<http://splunkbase.com>) for any apps that might be relevant to the customer's environment.

Module 4 Lab Exercise: Resource Planning


Description

Based on the previous exercises, we can now begin to create our Splunk infrastructure. We will continue to revise and enhance the design throughout the remainder of class.

Steps

Task: Size the infrastructure.

How many instances of each type of Splunk server will be needed? Are all of these needed? Can any of these be combined or virtualized? Reference Appendix 1 for additional information.

Splunk Server	Number of servers and why	Recommend Virtualize?	Combine?
Indexer			
Search head			
Deployment server			
License master			
Monitoring Console			





Module 5 Lab Exercise: Clustering Overview








Description

Based on the previous exercise, update your deployment to include indexer clustering for data replication.

Steps

Task: Incorporate indexer clustering in your environment.

1. Assume that the ecommerce data is critical. The customer wants to ensure that the Sales Support team can always access this data. Should any of the indexes be replicated?

2. What should the *minimum* replication factors be if the customer wants to ensure both data integrity and availability during a single indexer failure?

3. What effect will this have on the disk space requirements?

4. What effect would indexer clustering have on the overall infrastructure? How would the number and type of instances change?

5. Update the sizing table to include indexer clustering.

Splunk Server	Number of servers and why	Recommend Virtualize?	Combine?
Indexer			
Search head			
Master Node	 		
Deployer			
Deployment server			
License master			
Monitoring Console			

Module 6 Lab Exercise: Forwarding and Deployment Best Practices

Description

In this exercise, describe the topology of your Splunk deployment, including the forwarders. Also, consider the effects of deploying an app.

Steps

Q: Where are the systems that produce these logs?

A: The webserver and e-commerce transaction logs are on production servers at our data center co-location site in Ashburn, Virginia. We have two firewalls between them and our Headquarters in San Francisco. We have Linux and Windows servers at HQ in San Francisco and Linux-based file servers at our branch offices in London and Boston. The Cisco ASA firewalls and WSA proxy write their logs to a syslog port. The end user systems are Windows machines with Symantec Endpoint Protection; they interact with the Symantec Endpoint Protection Manager (EPM) and the EPM stores the consolidated antivirus logs. There are topology diagrams of both the current IT and logging environments. (See Exhibits 1, 2 and 3)

Q: Will this deployment be completed all at once?

A: No, we will utilize a **two-phase approach**.

The first phase will focus on making the logs available in Splunk as quickly as possible. We will use the existing logging infrastructure as much as possible, particularly the existing Windows logs and Linux-based syslog servers. The Symantec Endpoint Protection Manager is the collection point for anti-virus status and logs across the organization; we want to continue to leverage that function. We also want to get some initial custom reports and dashboards. Finally, we need to import our historical logs into the new Splunk deployment to meet our retention requirements.

During the second phase, we want to discontinue the use of our current syslog infrastructure. We intend to migrate as much as possible into Splunk.

Assumptions:

- All data that needs indexing resides in San Francisco
- The following data WILL NOT be indexed:
 - Employee workstations
 - Cloud applications such as HR, CRM, etc.
- Make reasonable assumptions about sizing the number of search heads (i.e. 10 users)

Steps

Task 1: Create a Phase 1 topology diagram.

Using the tool of your choice, create a topology diagram that shows where the Splunk instances will be installed. Include the appropriate production and logging infrastructure machines in your diagram. (You can omit the PCs and end users from your diagram.)

Include the data collection mechanisms in the deployment, whether you choose to use forwarder(s) or agentless input(s). Your topology should reflect Phase 1 of the deployment.

If you wish, you may create this diagram on paper and submit a JPEG picture of it. If you choose to use a tool, you may submit the topology as a separate file with your lab exercises. Be sure that the output file is in one of the formats identified in Lab Exercise 1. Ask your instructor if you are unsure how to submit your lab exercises.

Task 2: Analyze apps and understand the deployment issues.

1. Choose one of the following apps that might be appropriate for the customer environment.
 - Cisco Networks App
 - Splunk for Symantec End Point Protection

Consult Splunkbase (<http://splunkbase.com>) for more information on the app. You may want to download the app to your local machine (laptop, etc.) and unzip it to examine its contents. Then answer the following questions:

- Does the app have an add-on available?
- Does the app define an index or inputs?
- How will you deploy the app? What configuration files belong on the different instances in your infrastructure?



Task 3: Design a test environment.

2. How many Splunk servers are needed for the test environment if the customer is not interested in performance testing? Please explain your reason for selecting the number of servers for the test environment.
3. Can the test environment be used to test the deployment process?

Module 7 Lab Exercise: Create a Phase 2 Topology Diagram

Description

Use the Monitoring Console to gather information from a test environment before finalizing the production deployment.

Steps

Task: Review the forwarder configuration and create a Phase 2 topology diagram.



1. In the previous section, you created a topology diagram. Edit the forwarder design if needed.
2. Update the topology to reflect Phase 2 of the deployment.
Should all the syslog data collectors be replaced with forwarders?

Module 8 Lab Exercise: Performance Monitoring and Tuning

Description

Use the Monitoring Console to gather information from a test environment before finalizing the production deployment.

Steps



Task 1: Determine the resource usage for the test environment.

1. What is the average maximum load average for the last 7 days?
2. Are there specific days with high disk usage? Can you determine what caused the high disk usage (high number of users, ad-hoc reports, long searches, etc.)?
3. What is the maximum CPU usage deployment-wide?
4. Is there a spike in CPU usage during the week?

Steps

Task 2: Determine the performance of the indexes.

5. Which indexes are getting the most usage?
6. What is the average indexing rate per instance deployment-wide?

Steps

Task 3: Determine Search performance.

7. Are there any searches taking a long time to run?
8. Are there any reports that are taking a long time to run?

Module 9 Discussion - Use Cases

Description

Based on the use cases, how would you answer these questions on how to architect their Splunk deployment.

Questions

Use Case 1: Energy Service Provider

1. What is the minimum number of indexes required?
2. How many indexers do you recommend?
3. Would you have the master node and license master on the same Splunk instance?
4. In the long term would you look to replace the syslog servers and just listen on TCP ports for the data?
5. How do you get data from ISeries into Splunk?

Use Case 2: Financial Service Provider

6. Would you treat this as one Splunk deployment or two?
7. How would you collect data from the IBM Mainframe?
8. The customer decides to use 50 indexers and 5 search heads. Assuming an equal split of users/data at each site, how would you build out the architecture?

Use Case 3: Retail Company

9. How many indexers and search heads do you suggest for this customer?
10. Would you use index or search head clustering for this deployment?
11. Is a one or two stage approach recommended for the customer?
12. The customer is on a virtualization campaign. Which Splunk instances do you recommend for virtualization?

Final Lab Exercise: Submit Your Work

Description

In order to receive credit for this course, you must submit your lab exercises to your instructor. Your instructor will provide an email address for your lab exercises.

1. Make sure that all the documents containing your lab exercises use one of these technologies; otherwise, your instructor may be unable to score your exercises. Allowed formats:
 - PDF
 - Microsoft PowerPoint
 - Microsoft Excel
 - Rich Text Format (RTF) or Microsoft Word
 - JPEG (for diagrams)
2. If you are using multiple files, you may zip them before emailing them.
3. If you are sending pictures of documents, please ensure they are readable. This is especially true of hand-written drawings.
4. If you have any questions about your lab exercise submission, please ask your instructor before the end of class.

All lab exercises must be submitted on the final day of class. The instructor is required to submit Pass/Fail information for each student at the end of the final day of class.

Appendix 1

List of Employees by Department and Splunk Role:

Employee Name	Employee Position	Department (Data Access)	Splunk Deployment Functional Role
Pat Livingston	Director	IT Operations	Project Manager
Gabriel Voronff	Manager	IT Operations	Knowledge Manager, Designated Support Contact
Lien Teng	Team Member	IT Operations	Administrator, Designated Support Contact
Enrique Maxwell	Team Member	IT Operations	Administrator, Developer
Naomi Sharpe	Team Member	IT Operations	Search Expert
Amanda Curry	Manager	Security	Basic User
Alan Dombrowski	Team Member	Security	Developer
Nigella Pearce	Team Member	Security	Search Expert
Louis Sagers	Team Member	Security	Basic User
Cerys Farrell	Manager	Sales Support	Basic User
Placido Toscani	Team Member	Sales Support	Search Expert
Ian King	Team Member	Sales Support	Developer
Yanto Owen	Team Member	Sales Support	Basic User
Finlay Brya	Team Member	Sales Support	Search Expert

Description of Typical Splunk Roles and Skills

Functional Role	Responsibility	Skill Set	Recommended Splunk Training	Splunk Role Capabilities Required
Project Manager	<ul style="list-style-type: none"> Define project goals Set requirements Manage project plan Coordinate cross-department efforts 	<ul style="list-style-type: none"> PM skills Business problem expert Scope focus 	<ul style="list-style-type: none"> What is Splunk? (e-learning) 	<ul style="list-style-type: none"> User role capabilities Adopts one of the other user access roles
Splunk Administrator	<ul style="list-style-type: none"> Deploy Splunk Servers Deploy Forwarders Manage Splunk Manage users Helpdesk for system 	<ul style="list-style-type: none"> *Nix or Windows Sys Admin Good networking background Familiar with infrastructure 	<ul style="list-style-type: none"> All courses required for Administration Certification Administration Certification 	<ul style="list-style-type: none"> Admin role capabilities Can create saved searches and dashboards
Splunk Architect	<ul style="list-style-type: none"> Design Architecture Plan Splunk Deployment Deploy Splunk 	<ul style="list-style-type: none"> *Nix or Windows Sys Admin Good networking background Familiar with infrastructure 	<ul style="list-style-type: none"> All courses required for Architect Certification Architect Certification 	<ul style="list-style-type: none"> Admin role capabilities
Splunk Knowledge Manager	<ul style="list-style-type: none"> Oversight of knowledge object creation & usage Normalization of event data Create data models Create & enforce naming conventions 	<ul style="list-style-type: none"> Working with Splunk Apps Understands event processing Familiar with users and roles 	<ul style="list-style-type: none"> All courses required for Knowledge Manager Certification Knowledge Manager Certification 	<ul style="list-style-type: none"> Power or Admin role capabilities
Splunk Developer	<ul style="list-style-type: none"> Develop Splunk Apps Changes to .conf files Create management dashboards Write hooks to other data sources 	<ul style="list-style-type: none"> Splunk search Splunk UI Splunk SDK Python or other scripting External inputs 	<ul style="list-style-type: none"> Using Searching & Reporting Developing Apps Developing with SDKs 	<ul style="list-style-type: none"> Admin role capabilities Can create saved searches and dashboards Can develop Apps
Splunk Power User	<ul style="list-style-type: none"> Custom searches Forensics Helpdesk for searching 	<ul style="list-style-type: none"> Splunk search Source specific knowledge 	<ul style="list-style-type: none"> Using Searching & Reporting Adv. Searching & Reporting 	<ul style="list-style-type: none"> Can create saved searches and dashboards Power role capabilities
Splunk Basic User	<ul style="list-style-type: none"> Runs saved searches Views predefined dashboards 	<ul style="list-style-type: none"> Basic Splunk User Training Domain/Source knowledge 	<ul style="list-style-type: none"> Using 	<ul style="list-style-type: none"> Only views saved searches and dashboards User role capabilities