The Hitchhiker's Guide to Splunk Validated Architectures

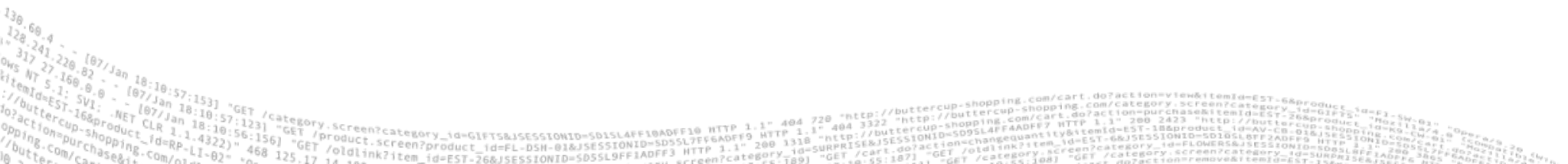
Stefan Sievert, Principal Architect
Eric Six, Staff Architect

October 2018 I Version 1

Forward-Looking Statements

During the course of this presentation, we may make forward-looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC.

The forward-looking statements made in this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward-looking statements we may make. In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.



Who are these guys?!

STEFAN SIEVERT

Principal Architect (Seattle)

ERIC SIX

Staff Architect (Tokyo)

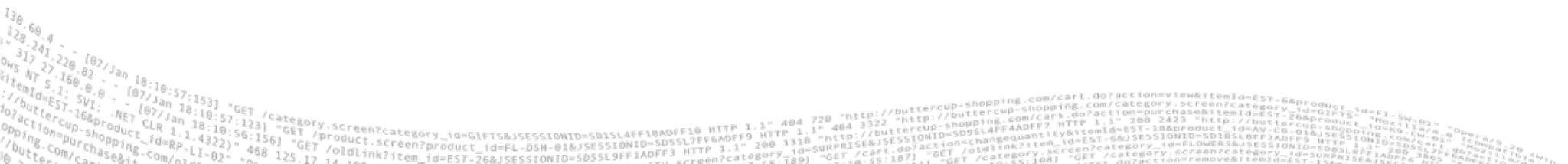
This solution ain't gonna architect itself...

Agenda

Am I in the right session?

Agenda

- What are SVAs and why do they exist
- Where you can find the content
- Scope and Structure of SVAs
- Example Scenarios
- Future Plans for SVAs
- Q&A



Splunk Validated Architectures

A refresher on “what” and “why”

splunk> .conf18

“Provide guidance on selecting proven reference architectures for stable, efficient and repeatable Splunk deployments.”

Topologies & their characteristics, and deployment best practices

splunk> .conf18

Primary Goals

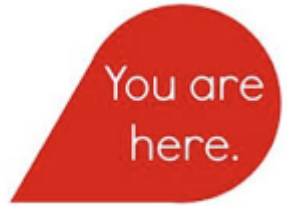
Why we care

1. Put your Splunk deployment on a solid architectural foundation
2. Meet your requirements exactly at minimal TCO
3. Focus resources on value realization, instead of fixing Splunk

Phased Development Approach

SVAs will evolve over time

- Phase 1: Focused on Indexing and Search Tiers



- Phase 2: Added Data Collection Tier components
- Phase 3: Making SVAs interactive
- Phase 4: Integrate with sizing calculators

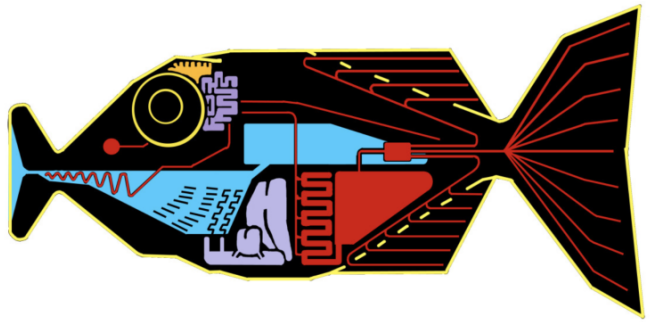
Problem & Solution

And who can benefit

Why SVAs?

The problem that needs solving

- Splunk deployments need a proven architectural foundation
  - Supportable
  - Scalable
  - Meeting Requirements
  - Best TCO



- “Issues” impede adoption and value realization
- “Issues” costing resources to troubleshoot and fix

SVAs to the rescue!

How we address the problem

- SVAs…
  - …provide step-by-step guides and best practices, complementing our product documentation
  - …empower you to architect, deploy and operate according to proven best practices
  - …result in more stable, efficient and repeatable deployments…
  - ...that meet your requirements
  - …will ensure your success with optimal TCO
  - …will allow you to shift your resources from “analyze & fix” to “adopt & delight”

Who should care?

The target audience for SVAs

- Everyone… wins!
  - Customer Staff
  - Professional Services Staff
  - Splunk Partners
  - Splunk Staff

SVA Content Location

When the search engine fails…

splunk> .conf18

Didn’t bring your towel??

- “Splunk Validated Architectures white paper site:splunk.com” OR
- https://www.splunk.com/pdfs/white-papers/splunk-validated-architectures.pdf

Splunk Validated Architectures Proven reference architectures for ...
https://www.splunk.com/pdfs/white-papers/splunk-validated-architectures.pdf
No information is available for this page.
Learn why

Splunk | References, Whitepapers, Solution Guides, Data Sheets, Fact ...
https://www.splunk.com/en_us/resources.html ▾
Analyst Reports, E-Books, **White Papers** and More. Explore .... **Splunk Validated Architectures** ...
The Accenture Cyber Defense Platform - **Architectural** Overview.

Scope & Structure

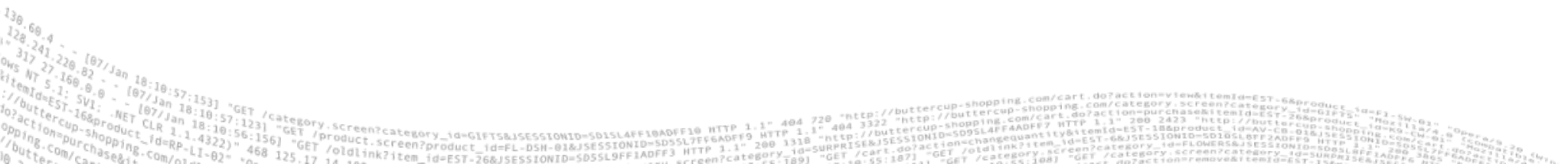
What SVAs cover and how to read them

SVA Scope

A quick tour

- SVAs DO…
  - ...address topology selection for indexing and search tiers
  - ...address data collection mechanism selection and architecture
  - ...provide best practices and design principles for implementation

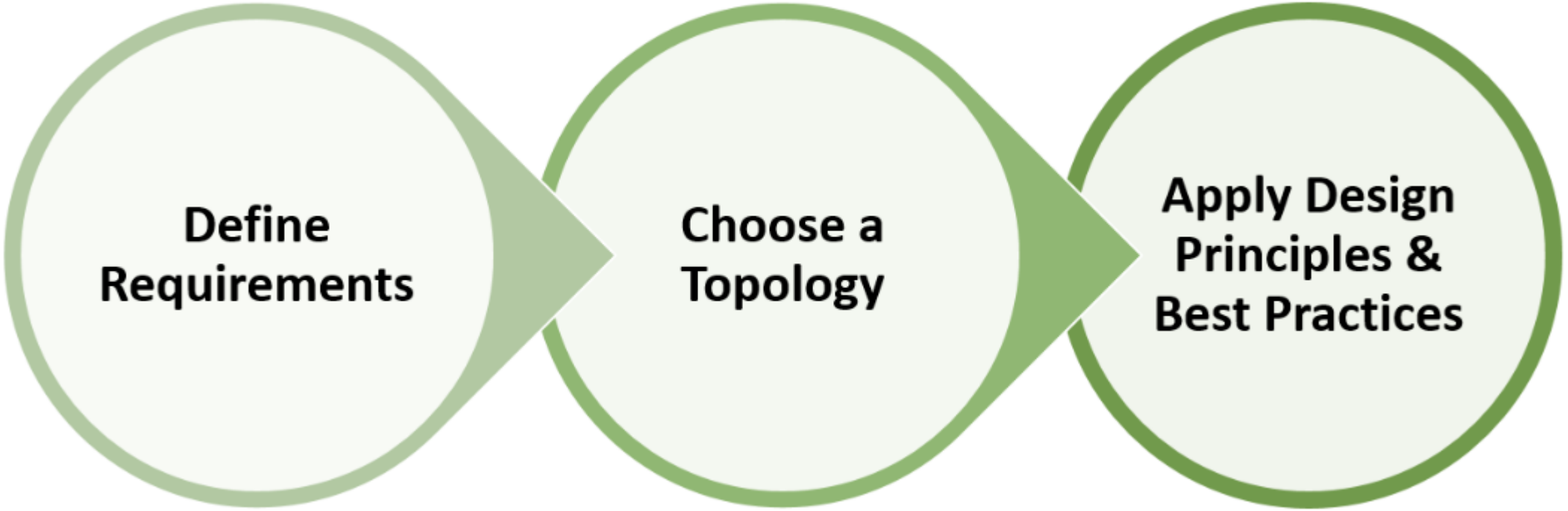
- SVAs DO NOT...
  - ...contain sizing guidelines
  - ...address detailed deployment choices (OS, Cloud, etc)
  - ...cover every possible customer scenario



SVA Structure

A quick tour

- The SVA document will guide you through a 3-step process:

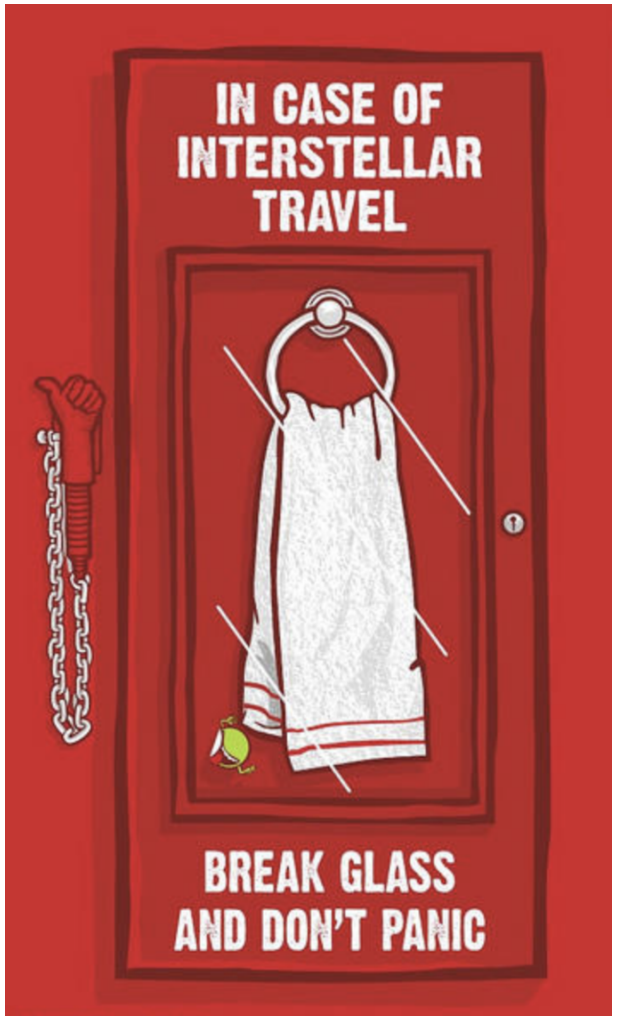


Example Scenarios

Walking through real-world examples

Use Case 1: New Deployment

- Brand New! Deployment..
  - 100gb/day
  - Network Data Sources
  - No HA / DR for anything



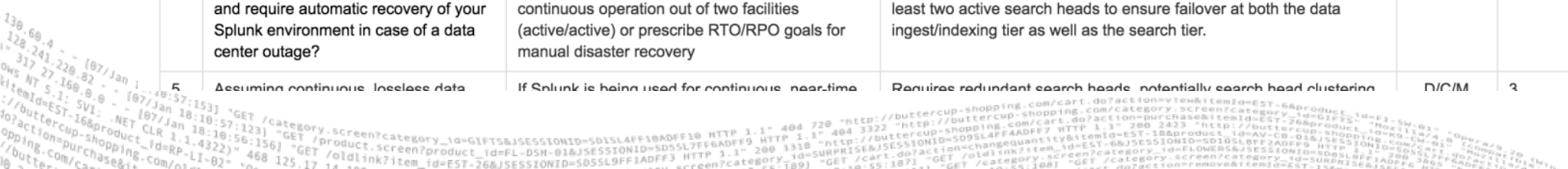
Use Case 1: Topology Selection

https://www.splunk.com/pdfs/white-papers/splunk-validated-architectures.pdf

Questionnaire 1: Defining Your Requirements for Index and Search Tiers

♦ *See the key above for an explanation of topology category codes. If you answer "yes" to multiple questions, use the topology category code for the highest numbered question.*

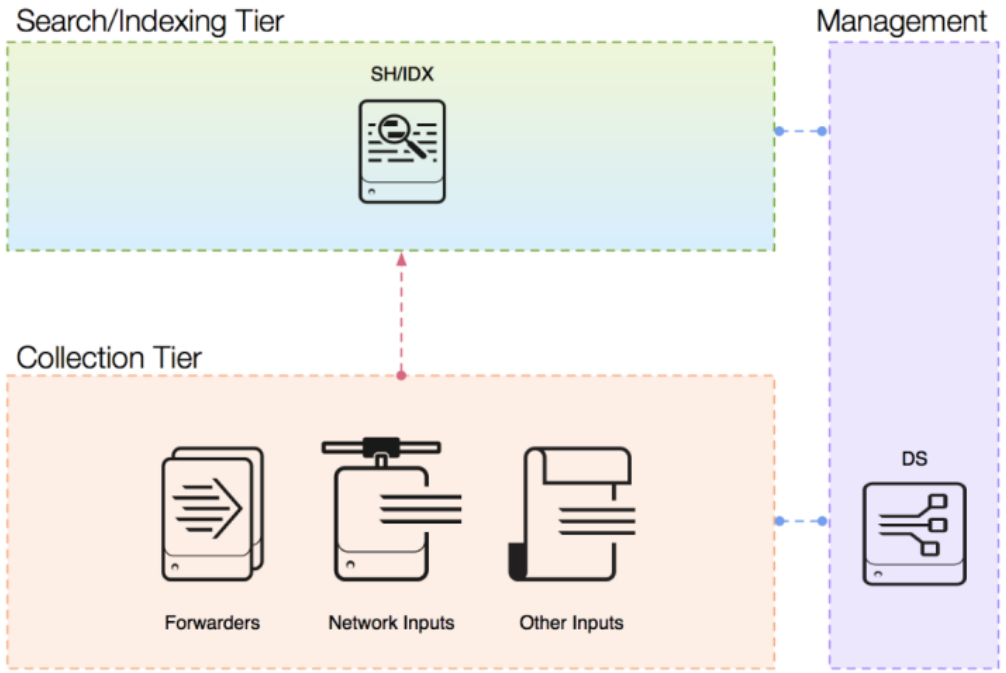
#	Question	Considerations	Impact on Topology	Indexer Tier Topology Category ♦	Search Tier Topology Category ♦
1	Is your expected daily data ingest less than ~300GB/day?	Consider short-term growth in the daily ingest (~6-12 month)	Candidate for a single server deployment, depending on answers to availability-related questions	S	1
2	Do you require high availability for data collection/indexing?	If you are not planning on using Splunk for monitoring use cases that require continuous data ingest, a temporary interruption of the inbound data flow may be acceptable; assuming no log data is lost.	Requires distributed deployment to support continuous ingest	D	1
3	Assuming an available Search Head to run a search: Does your data need to be completely searchable at all times, i.e. you cannot afford any impact to search result completeness?	If your use case is calculating performance metrics and general usage monitoring using aggregate functions, for example, a single indexer outage may not materially affect the calculation of statistics over a large number of events. if your use case is security auditing and threat detection, blind spots in search results are very likely undesirable	Requires clustered indexers with a replication factor of at least two (2). Note: While a replication factor of 2 provides minimal protection against a single indexer node failure, the recommended (and default) replication factor is 3.	C	1
4	Do you operate multiple data centers and require automatic recovery of your Splunk environment in case of a data center outage?	Disaster recovery requirements may dictate continuous operation out of two facilities (active/active) or prescribe RTO/RPO goals for manual disaster recovery	Continuous operation will require multi-site indexer clustering and at least two active search heads to ensure failover at both the data ingest/indexing tier as well as the search tier.	M	2
5	Assuming continuous, lossless data	If Splunk is being used for continuous, near-time	Requires redundant search heads, potentially search head clustering	D/C/M	3



splunk> .conf18

Use Case 1: Topology

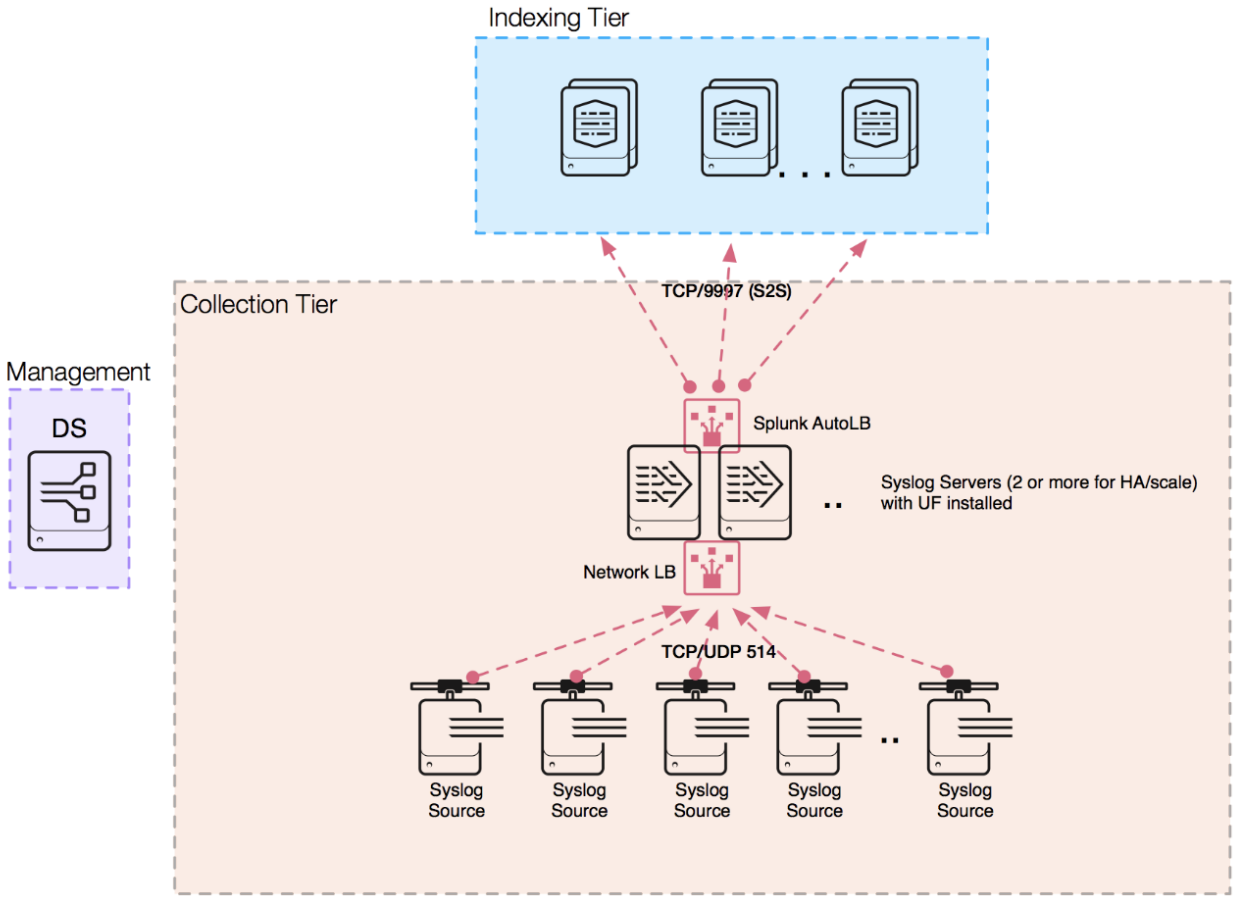
Single Server Deployment (S1)



Description of the Single Server Deployment (S1)	Limitations
This deployment topology provides you with a very cost-effective solution if your environment meets all of the following criteria: a) you do not have any requirements to provide high-availability or automatic disaster recovery for your Splunk deployment, b) your daily data ingest is under 300GB/day, and c) you have a small number of users with non-critical search use cases. This topology is typically used for smaller, non business-critical use-cases (often departmental in nature). Appropriate use cases include data onboarding test environments, small DevOps use cases, application test and integration environments, and similar scenarios. The primary benefits of this topology include easy manageability, good search performance for smaller data volumes, and a fixed TCO.	• No High Availability for Search/Indexing • Scalability limited by hardware capacity (straightforward migration path to a distributed deployment)

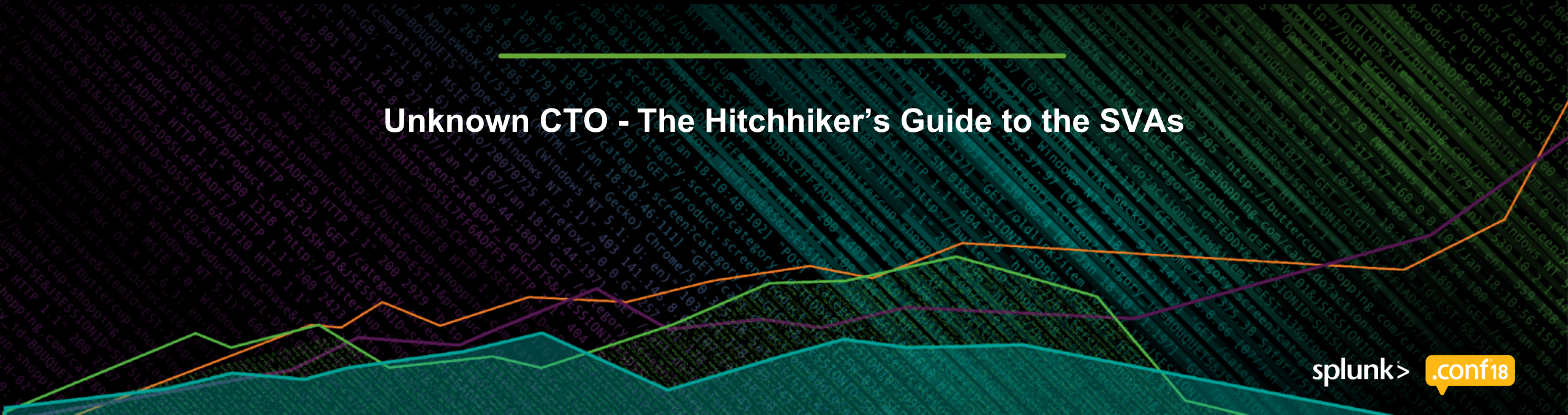
Use Case 1: Data Collection

- Identify required collection components via the second questionnaire
  - for this use case, let's assume you need Syslog data collection only



“ One of the things I always found hardest to understand about Splunk Architects was their habit of continually stating and repeating the very very obvious, as in ‘This Gin is amazing’, or ‘Adding more Indexers will fix that problem’.”

Unknown CTO - The Hitchhiker’s Guide to the SVAs

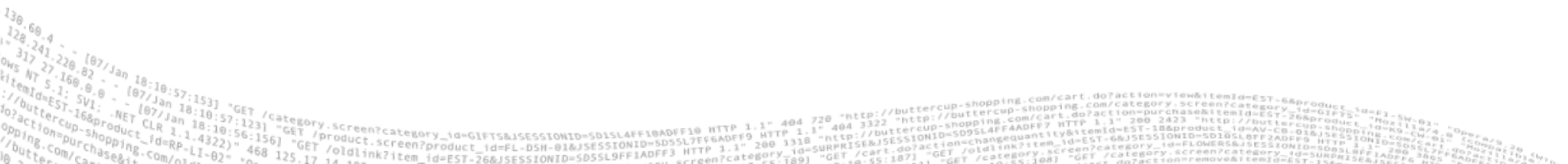


Use Case 2: New “Large Deployment”

- Brand New, “Large” Deployment
  - 5tb/day, HA + DR requirements, Multi-DataCenter deployment
  - 5000+ UFs / Endpoints

Use Case 2: Topology Selection

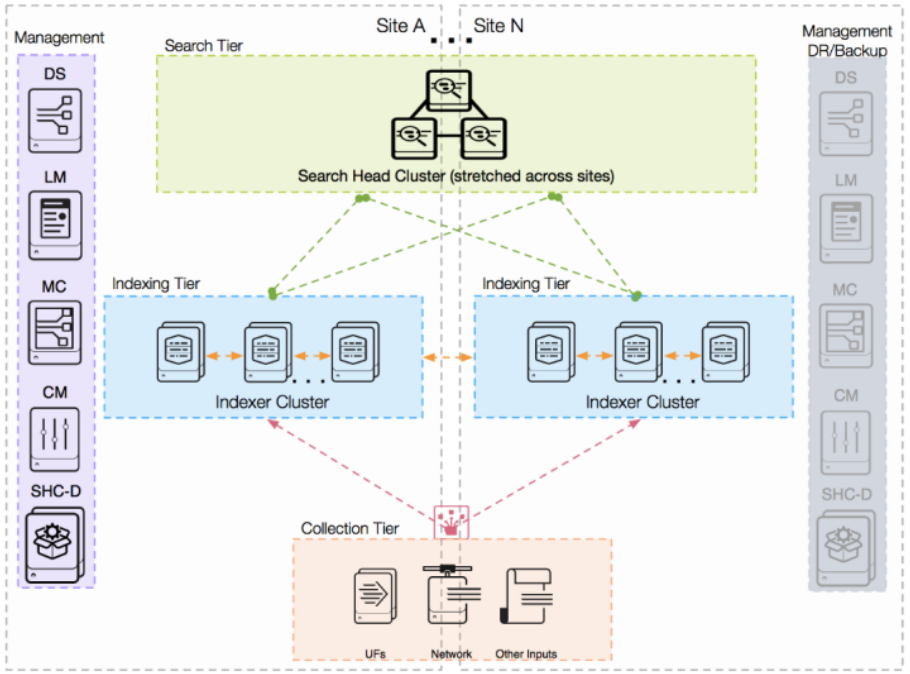
#	Question	Considerations	Impact on Topology	Indexer Tier Topology Category ♦	Search Tier Topology Category ♦
	and require automatic recovery of your Splunk environment in case of a data center outage?	continuous operation out of two facilities (active/active) or prescribe RTO/RPO goals for manual disaster recovery	least two active search heads to ensure failover at both the data ingest/indexing tier as well as the search tier.		
5	Assuming continuous, lossless data ingest, do you require HA for the user-facing search tier?	If Splunk is being used for continuous, near-time monitoring, interruptions in the search tier are likely not tolerable. This may or may not be true for other use cases.	Requires redundant search heads, potentially search head clustering	D/C/M	3
6	Do you need to support a large number of concurrent users and/or a significant scheduled search workload?	Requirements for more than ~50 concurrent users/searches typically require horizontal scaling of the search tier	May require a topology that uses a search head cluster in the search tier	D/C/M	3
7	In a multi-data center environment, do you require user artifacts (search results, searches, dashboards and other knowledge objects) to be synchronized between sites?	This will decide whether users will have a current and consistent experience in case of a site outage.	Requires a "stretched" search head cluster across sites with appropriate configuration	M	4



splunk> .conf18

Use Case 2: Topology

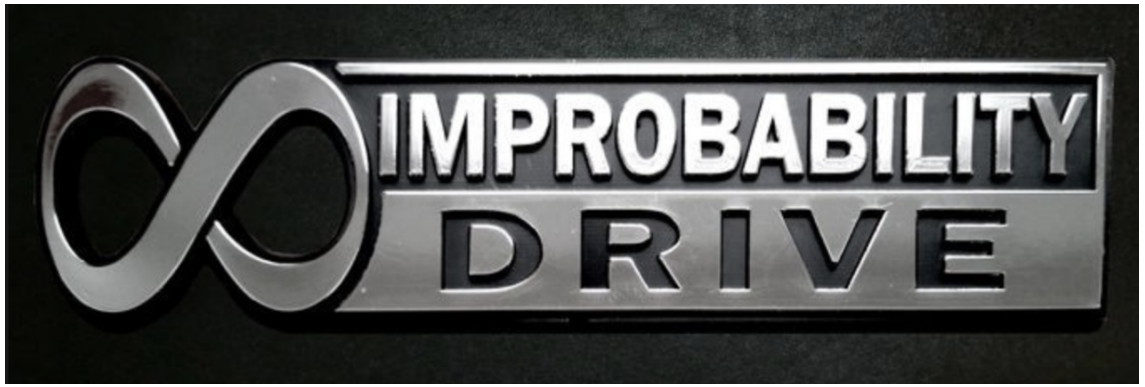
Distributed Clustered Deployment + SHC - Multi-Site (M4 / M14)



Description of Distributed Clustered Deployment + SHC - Multi-Site (M4 / M14)	Limitations
This is the most complex validated architecture, designed for deployments that have strict requirements around high-availability and disaster recovery. We strongly recommend involving Splunk Professional Services for proper deployment. When properly deployed, this topology provides continuous operation of your Splunk infrastructure for data collection, indexing, and search. This topology involves implementation of a "stretched" search head cluster that spans one or more sites. This provides optimal failover for users in case of a search node or data center failure. Search artifacts and other runtime knowledge objects are replicated in the SHC. Careful configuration is required to ensure that replication will happen across sites, as the SHC itself is not site-aware (i.e. artifact replication is non-deterministic). Site-affinity can be configured to ensure the WAN link between sites is utilized only in cases when a search cannot be satisfied locally. **Note:** If your category code is M14 (i.e. you intend to deploy the Splunk App for Enterprise Security), a single dedicated search head cluster (also stretched across sites) is required to deploy the app (this is not pictured in the topology diagram). A network load-balancer is required in front of the SHC members to ensure proper load balancing of users across the cluster.	• Network latency across sites must be within documented limits • Failover of the SHC may require manual steps if only a minority of cluster members survive

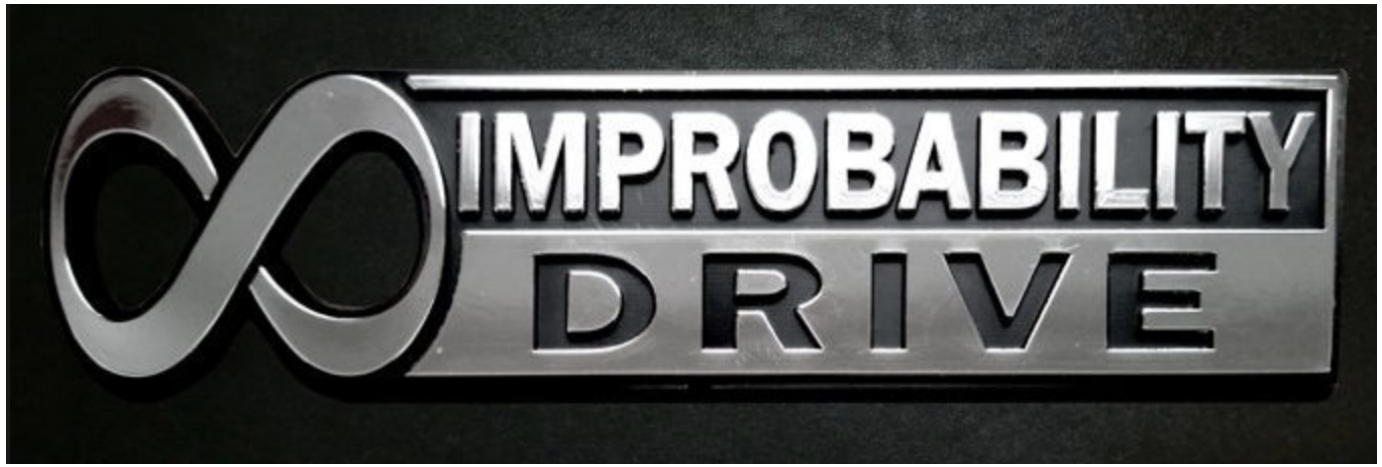
Use Case 2: Data Collection

- Large Deployments like this typically require multiple different mechanisms to acquire and collect log data
- Identify those mechanisms using the second questionnaire
- Pay attention to the critical considerations, as they will affect performance
- Consider management approach for data collection tier configuration



Use Case 3: Inherited “Buckets O’ Fun”

- Existing Deployment
  - Maybe-- 2tb/day
  - 3000+ UFs / Endpoints
  - Syslog Aggregates
  - DBX, AWS, OPSec LEA



The Future

What's next for SVAs?

What's in the | for SVAs

- Turning SVAs into an interactive tool
  - Go to a URL
  - Answer questions
  - Get your deployment diagram!
- Sizing Calculator Integration
  - Integration with Core and Premium Sizing Calculators
  - Generating a detailed deployment diagram you can use to work off of

Q&A

Got questions?

splunk>

DON'T PANIC
AND CARRY A TOWEL

USES FOR YOUR TOWEL:

1 - Wave as a distress signal
2 - Soak it and use as a weapon
3 - Cover your face against fumes
4 - Hide from bugblatters
5 - Use as a sail
6 - Small blanket
7 - Dry off

Thank You

Don't forget to rate this session in the .conf18 mobile app