



The United States of Digital Search

Kristerpher Henderson
Principal Engineer | Comcast



COMCAST
splunk> .conf19

Prologue

Who is this guy?

.conf19

splunk>

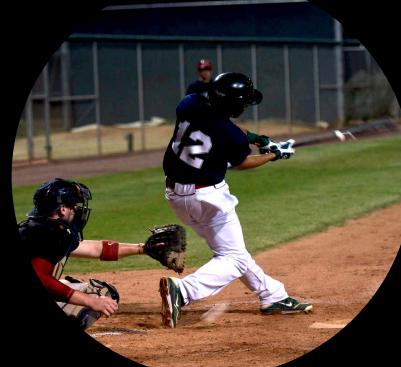


Kristerpher

Technology | Arts | Baseball | Photography



Technologist



Semi-Pro
Baseball
Player



Silent Actor in
Opera



Open Source
and Code for
Philly Activist



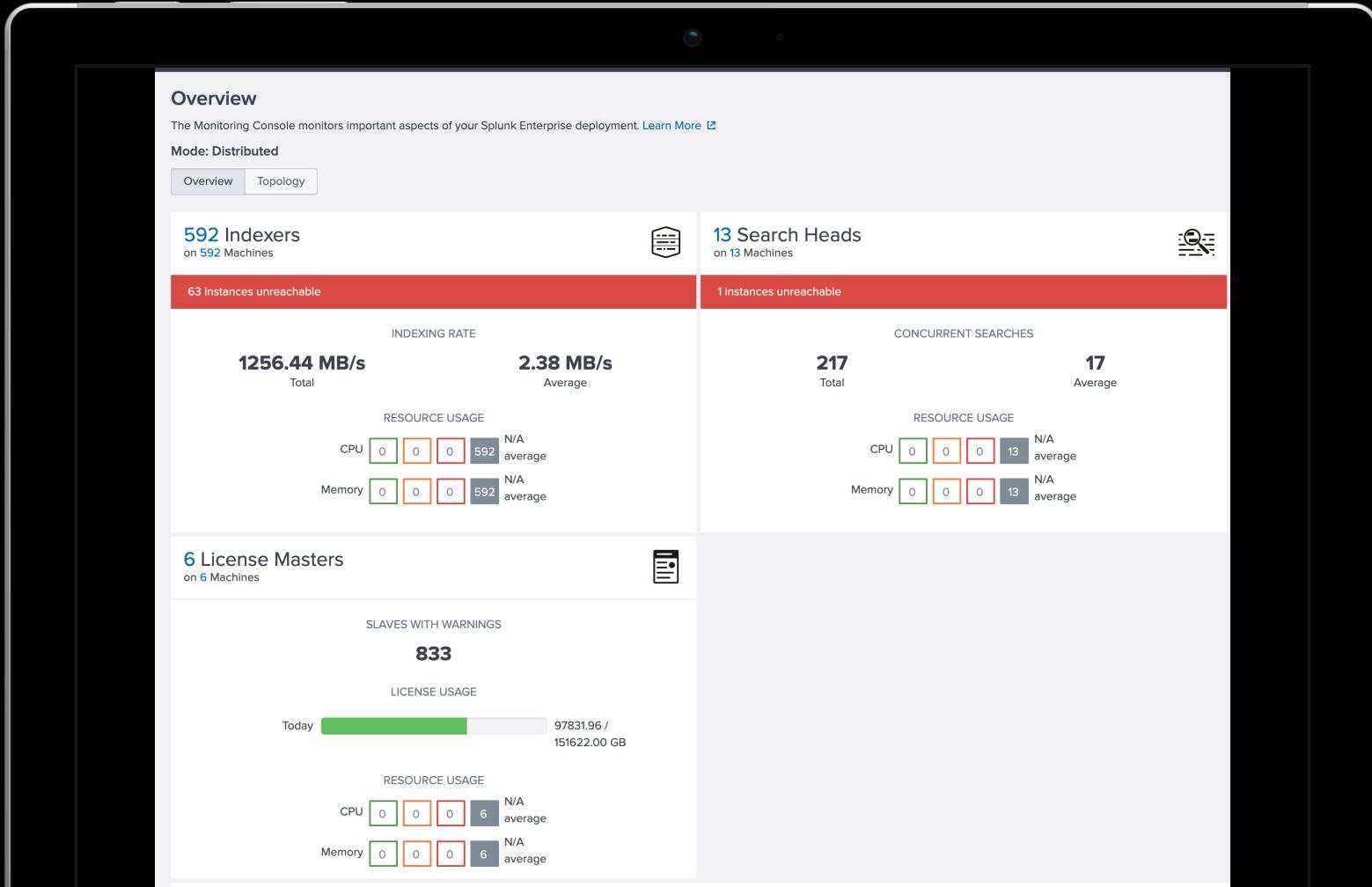
You're
friendly
neighborhood
splunker



splunk> .conf19

Splunk at Comcast

All the data...



- ▶ Over 750 indexers
- ▶ 12 Node SHC across multiple AWS Regions
- ▶ Few indexes but lots of sourcetypes

PROLOGUE

What's in there?

Technology Product and Experience

The ecosystem uses voice remote info to trigger box-to-device application through a central API endpoint. The Xfinity TV app and its kid app are developed using the same interface with the help of a central API endpoint. The Xfinity TV app can interact with other products in the ecosystem.

- ▶ 20+ Million Subscribers STB
 - ▶ XRE + RDK
 - ▶ Guide, Voice and Home
 - ▶ Support Services



COMCAST

splunk> .conf19

Our biggest failures have lead to our greatest successes:

How we learn and scale

splunk>enterprise Apps ▾

Administrator 1002 Messages ▾ Settings ▾ Activity ▾ Help ▾ Find ▾

Monitoring Console

Overview Health Check Instances Indexing ▾ Search ▾ Resource Usage ▾ Forwarders ▾ Settings ▾ Run a Search

Indexing Performance: Deployment

Group: All Indexers Hide Filters

Select views: All Snapshot Historical

Snapshots

Overview of Indexing Performance

541	1,239,929 KB/s	2,292 KB/s	2,077 KB/s
Indexers	Total Indexing Rate	Average Indexing Rate	Median Indexing Rate

Indexing Performance by Instance

541 instances

Instance	Pipeline Set Count	Indexing Rate (KB/s)	Status	Parsing Queue Fill Ratio (%)	Aggregation Queue Fill Ratio (%)	Typing Queue Fill Ratio (%)	Indexing Queue Fill Ratio (%)
spl-ch-c0052-g.ch.tvx.comcast.com	1	17422	normal	99.94	64.17	99.94	58.94
spl-ch-6026-g.ch.tvx.comcast.com	2	15382	normal	pset0: 0.00 pset1: 0.00	pset0: 0.00 pset1: 0.00	pset0: 0.00 pset1: 0.00	pset0: 0.00 pset1: 0.04
spl-as-1084-g.as.tvx.comcast.com	1	11585	normal	0.00	0.00	0.00	0.00
spl-as-1089-g.as.tvx.comcast.com	1	9629	normal	0.01	0.00	0.00	0.67
spl-as-1086-g.as.tvx.comcast.com	1	8879	normal	7.30	99.97	98.75	99.97
ctv-splix-ntlk-a-01.railroadave.il.chicago.comcast.net	2	7749	normal	pset0: 0.04 pset1: 0.00	pset0: 0.09 pset1: 0.04	pset0: 0.00 pset1: 0.10	pset0: 9.47 pset1: 0.00
ctv-dnvr-splix-02.highfield.co.denver.comcast.net	2	7692	normal	pset0: 0.00 pset1: 0.00	pset0: 0.00 pset1: 0.00	pset0: 0.29 pset1: 0.00	pset0: 0.00 pset1: 0.00
spl-as-1087-g.as.tvx.comcast.com	1	7486	normal	0.00	0.00	0.00	0.18
spl-ch-6025-g.ch.tvx.comcast.com	1	7382	normal	0.00	0.00	0.00	0.00
spl-as-1012-g.as.tvx.comcast.com	1	7056	normal	0.00	0.00	0.00	0.00

Click instance name for more details.

Indexing rate measured over 30 seconds every 30 seconds (available with Splunk Enterprise 6.2 or later indexers).

< prev 1 2 3 4 5 6 7 8 9 10 next >



COMCAST
splunk> .conf19

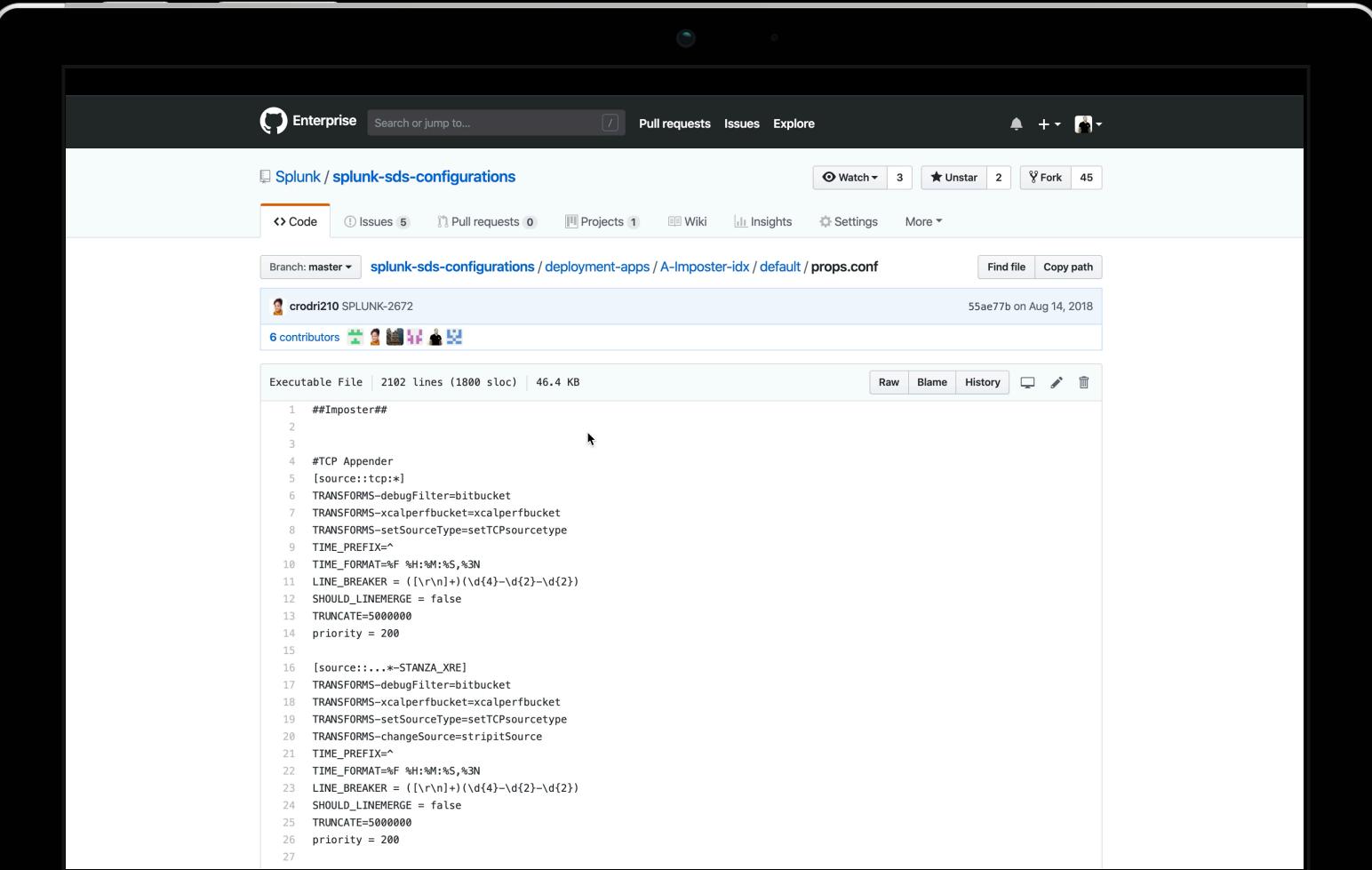
How we learn and scale

- ▶ Ensure hardware is spec'd for ingest, not search.
 - ▶ Buy for growth, not requirements.
 - ▶ Automation...
 - ▶ ...Automation
 - ▶ AUTOMATION
 - ▶ Locality matters
 - ▶ Do not support the forwarders...
 - ▶ ...support the process to enable the forwarders.

How we learn and scale

- ▶ Welcome new data...
- ▶ ...but how do we handle all of it.
- ▶ IMPOSTER

Imposter



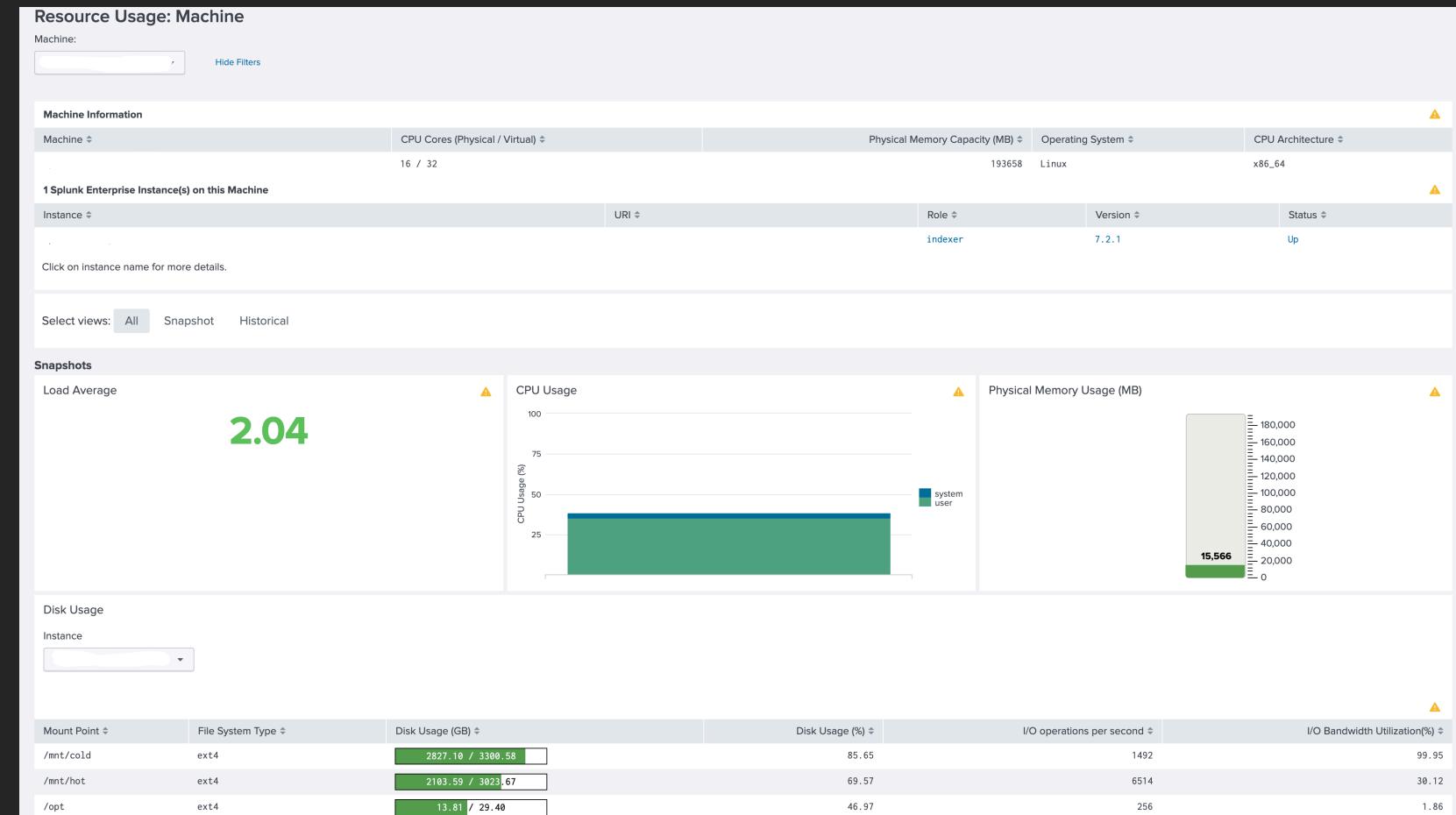
The screenshot shows a GitHub repository page for 'Splunk / splunk-sds-configurations'. The repository has 3 stars, 2 forks, and 45 issues. The 'Code' tab is selected, showing the 'props.conf' file. The file content is as follows:

```
1  ##Imposter##
2
3
4  #TCP Appender
5  [source::tcp:*]
6  TRANSFORMS-debugFilter=bitbucket
7  TRANSFORMS-xcalperfbucket=xcalperfbucket
8  TRANSFORMS-setSourceType=setTCPsourcetype
9  TIME_PREFIX=
10 TIME_FORMAT=%F %H:%M:%S,%3N
11 LINE_BREAKER = ([\r\n]+)(\d{4}-\d{2}-\d{2})
12 SHOULD_LINEMERGE = false
13 TRUNCATE=5000000
14 priority = 200
15
16 [source::...*-STANZA_XRE]
17 TRANSFORMS-debugFilter=bitbucket
18 TRANSFORMS-xcalperfbucket=xcalperfbucket
19 TRANSFORMS-setSourceType=setTCPsourcetype
20 TRANSFORMS-changeSource=stripItSource
21 TIME_PREFIX=
22 TIME_FORMAT=%F %H:%M:%S,%3N
23 LINE_BREAKER = ([\r\n]+)(\d{4}-\d{2}-\d{2})
24 SHOULD_LINEMERGE = false
25 TRUNCATE=5000000
26 priority = 200
27
```

- ▶ An Imposter App is deployed to all Indexers.
- ▶ It allows for us to match specific data formats without needing to restart the indexers.

What's under the hood?

Check the headlight
fluid



COMCAST
splunk> .conf19

How fast really is fast...

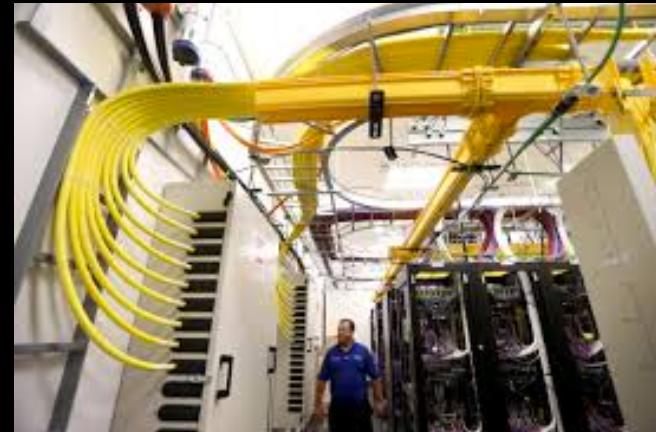
- ▶ Direct Attached Storage
- ▶ IOPS over 3000 on multiple channels
- ▶ Spec'd for 250GB of ingest per day.
- ▶ Total install may approach over 1.5GB/sec at any given point...
 - ▶ ~140 TB (with a T)

But where is your data?

All over the place...



National Data Centers



Regional Data Centers



AWS... kind of...

Types? You got them?

Yes.

- ▶ Unstructured
- ▶ Structured
- ▶ Lambda/Kinesis
- ▶ The entire TV Guide...



COMCAST
splunk> .conf19

Other cool things?

Of course.

- ▶ Real-time monitoring
- ▶ Customer Experience Dashboard
- ▶ Ingest Automation (more on that later...)
- ▶ End-to-End view of customer's experience on the X1 Platform

THE FOLLOWING SCREENS ARE INTERNAL
EXPERIMENTAION THAT ARE NOT RELATED TO COMCAST
BUISNESS, BUT SOLELY REPRESENTED FOR
DEMONSTRATION ON WHAT SPLUNK CAN DO WITH DATA.

xHome Customer Account Lookup

Edit More Info ↴ ↵

Account Number

during Dec 2016

Customer Experience

Over the last 30 days

-3

Number of Calls	IVR Call Summary	Number of Truck Rolls								
4	IVR Call Summary <table border="1"> <thead> <tr> <th>Issue Type</th> <th>Issue Name</th> </tr> </thead> <tbody> <tr> <td>Appointment</td> <td></td> </tr> <tr> <td>Payments</td> <td></td> </tr> <tr> <td>Technical</td> <td></td> </tr> </tbody> </table>	Issue Type	Issue Name	Appointment		Payments		Technical		0
Issue Type	Issue Name									
Appointment										
Payments										
Technical										

Customer's Error for xHome

N/A

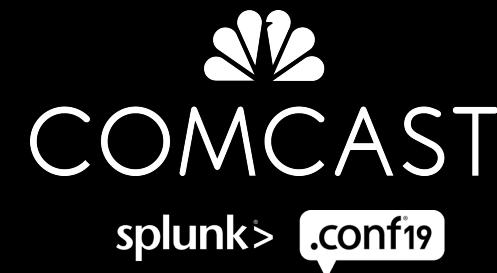
<1 ago

Scoring:
Did they call (-3)
Did they have an error (-1)
Did we visit them (-6)



COMCAST

splunk> .conf19



ACT I

Indexer and SHC Cluster Architecture

Distributed Search on a National Scale

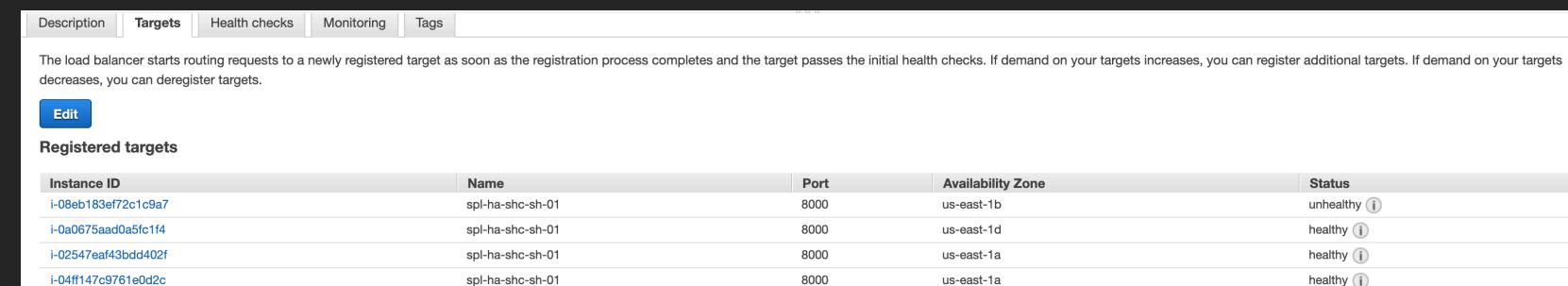
When automation doesn't handle a server's problem we are alerted in many different ways.

- ▶ 12 Node Search Head Cluster
- ▶ Distributed Search
- ▶ Splunk Deployment Server (Cluster)
- ▶ Monitoring via OP5/Zabbix



Multi-Region Search Head Cluster

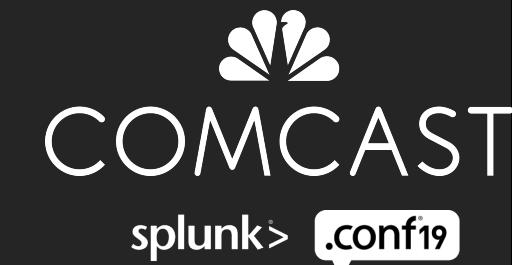
Spec'd to survive a complete data center outage.



The load balancer starts routing requests to a newly registered target as soon as the registration process completes and the target passes the initial health checks. If demand on your targets increases, you can register additional targets. If demand on your targets decreases, you can deregister targets.

Registered targets

Instance ID	Name	Port	Availability Zone	Status
i-08eb183ef72c1c9a7	spl-ha-shc-sh-01	8000	us-east-1b	unhealthy ⓘ
i-0a0675aad0a5fc1f4	spl-ha-shc-sh-01	8000	us-east-1d	healthy ⓘ
i-02547eaf43bdd402f	spl-ha-shc-sh-01	8000	us-east-1a	healthy ⓘ
i-04ff147c9761e0d2c	spl-ha-shc-sh-01	8000	us-east-1a	healthy ⓘ



3 AWS Regions

US-EAST-1 (4)

US-EAST-2 (Waiting on DX)

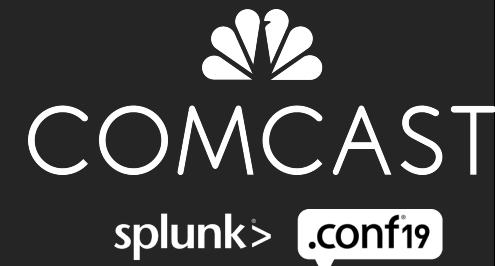
US-WEST-2 (8)



The load balancer starts routing requests to a newly registered target as soon as the registration process completes and the target passes the initial health checks. If demand on your targets increases, you can register additional targets. If demand on your targets decreases, you can deregister targets.

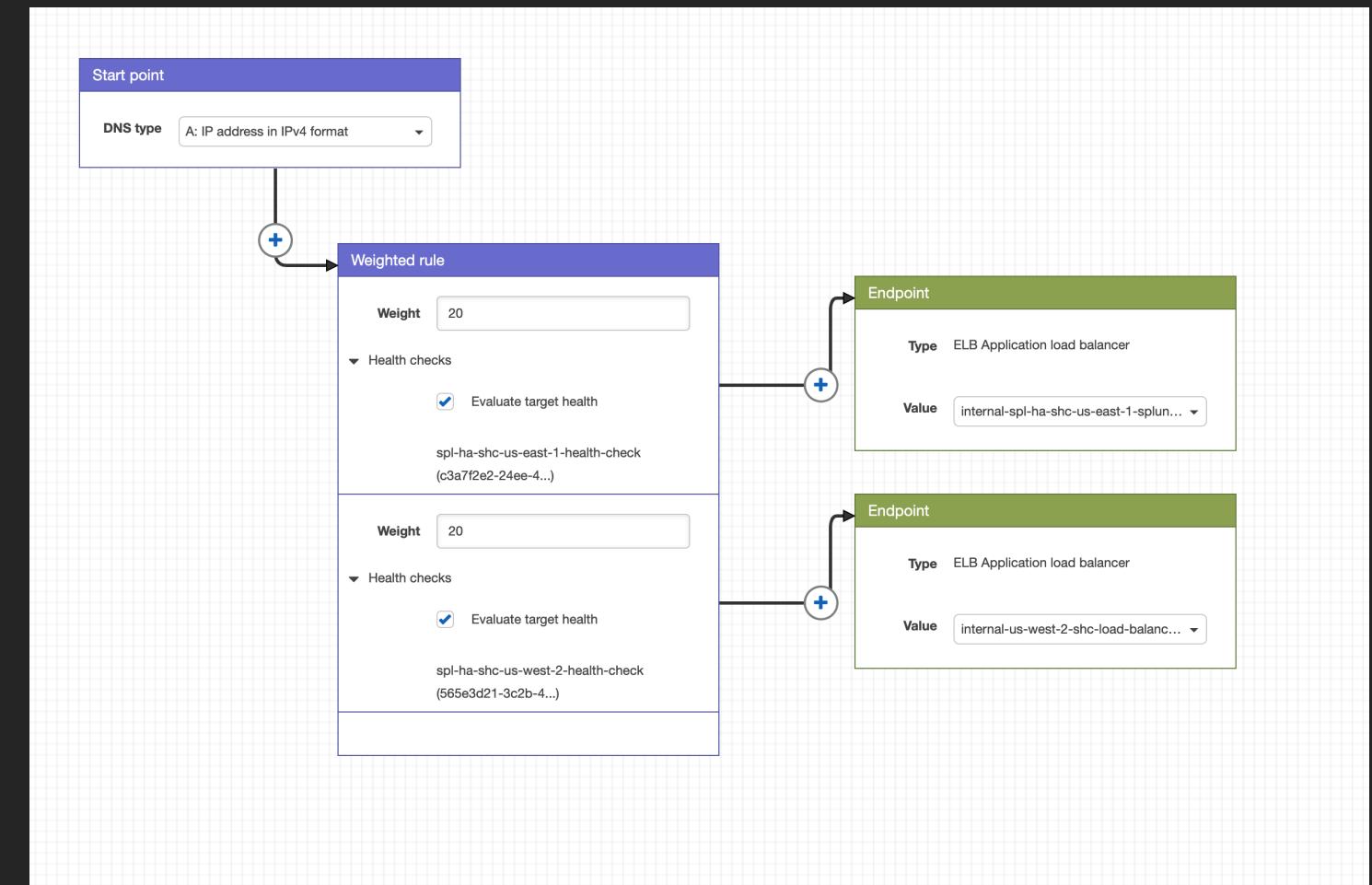
Registered targets

Instance ID	Name	Port	Availability Zone	Status
i-08eb183ef72c1c9a7	spl-ha-shc-sh-01	8000	us-east-1b	unhealthy ⓘ
i-0a0675aad0a5fc1f4	spl-ha-shc-sh-01	8000	us-east-1d	healthy ⓘ
i-02547eaf43bdd402f	spl-ha-shc-sh-01	8000	us-east-1a	healthy ⓘ
i-04ff147c9761e0d2c	spl-ha-shc-sh-01	8000	us-east-1a	healthy ⓘ

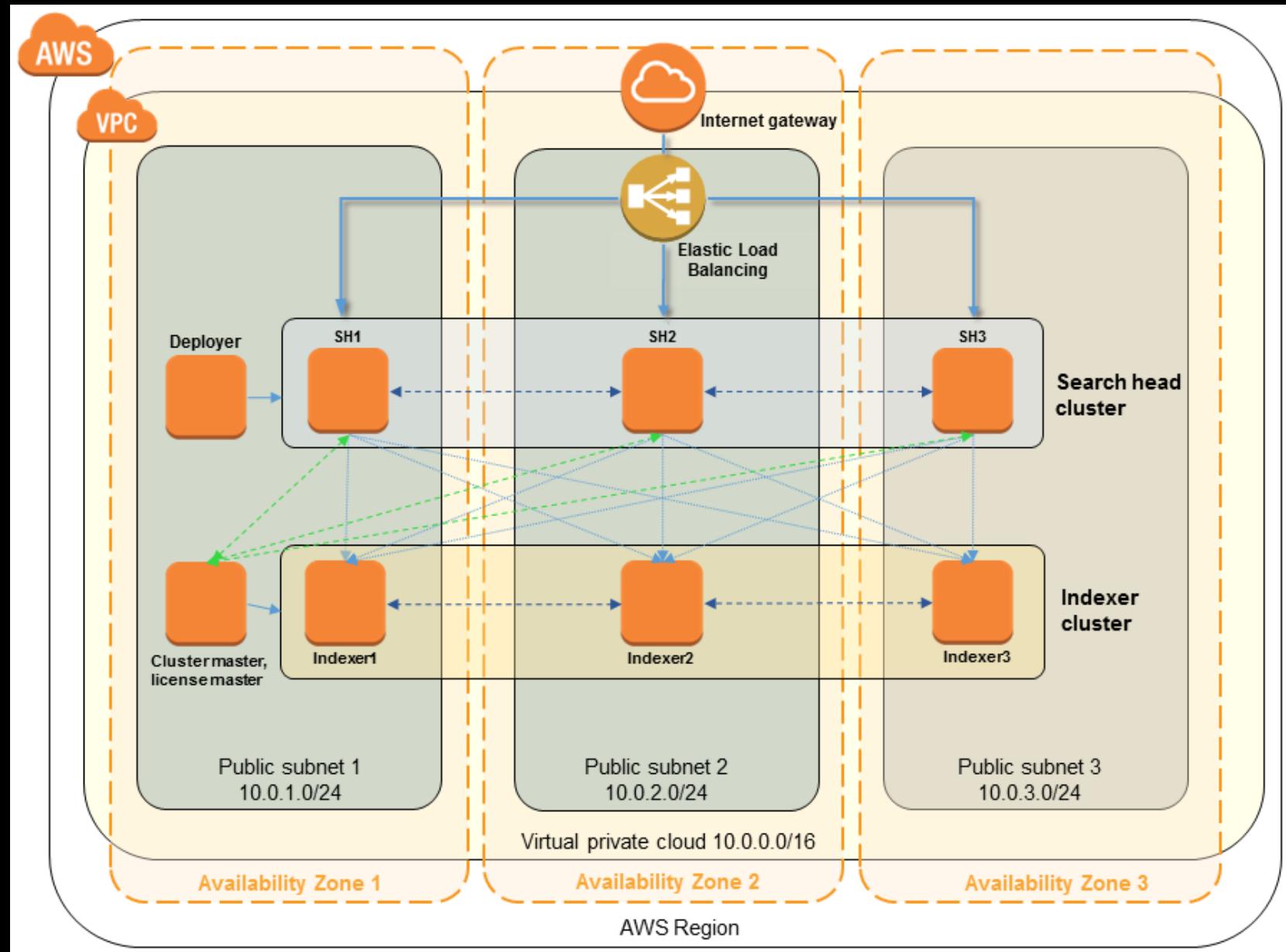


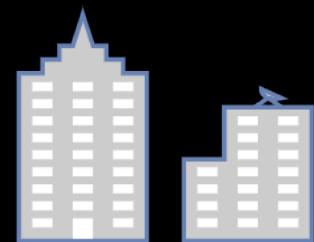
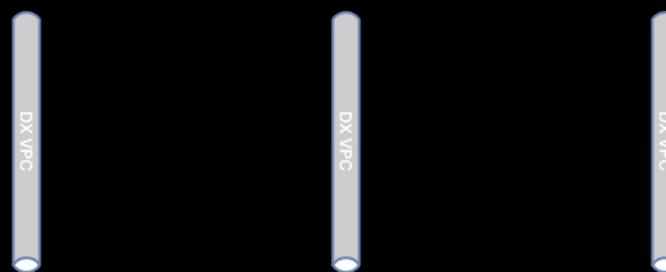
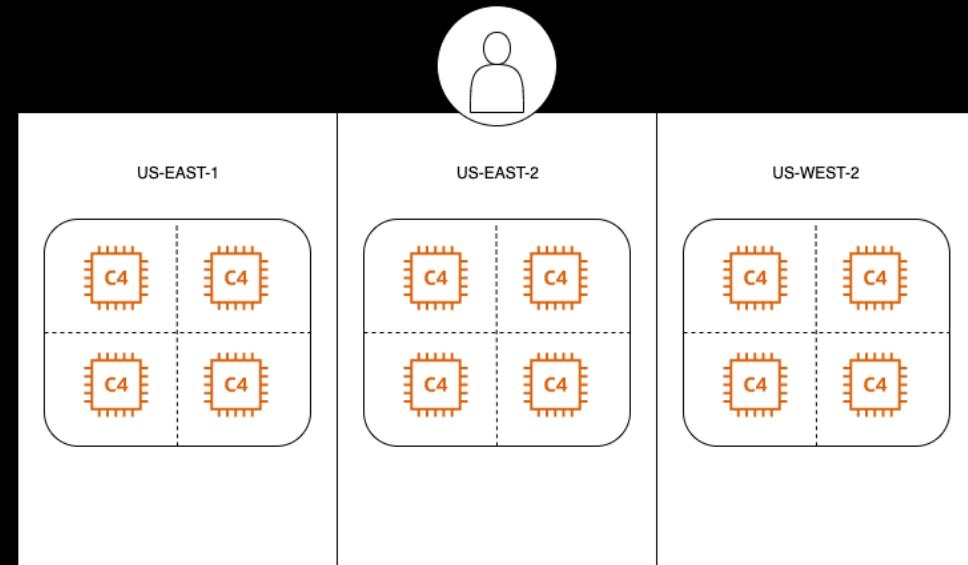
Routing Policy

Zone Records are controlled via AWS



COMCAST
splunk> .conf19





SHC Connectivity Workflow

- ▶ Users use one URL for UI, one for API
- ▶ AWS Routes them based on the geographic location.
- ▶ Private DX VPC will allow connectivity between the indexers and the search heads.
- ▶ Data that needs to use cross region connectivity, will use the Comcast Backbone to travel (~78ms Latency)

DOES IT WORK?

Sure, you just need to fine tune it...



COMCAST

splunk> .conf19

Timeouts and Connectivity

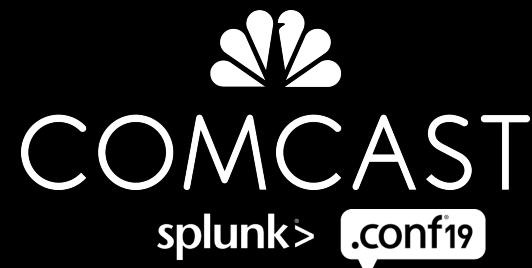
The Presentation Layer and The Splunk Layer

- ▶ The presentation layer has shorter timeouts.

Timeouts and Connectivity

The Presentation Layer and The Splunk Layer

- ▶ Shorter timeouts means, less abuse from cURL



Timeouts and Connectivity

The Presentation Layer and The Splunk Layer

- ▶ Splunk needs longer timeouts

Timeouts and Connectivity

The Presentation Layer and The Splunk Layer

- ▶ Snapshots and Bundles...

The Problems

The Quick...

- ▶ Long running searches keep connections open too long.
- ▶ Tune them to be shorter, but use search policy to help.

The Problems

...and the dirty...

```
[splunker@deployer default]# cat distsearch.conf
```

```
[replicationSettings]
```

```
replicationThreads = auto
```

```
connectionTimeout = 120
```

```
sendRcvTimeout = 120
```

```
concerningReplicatedFileSize = 90
```

```
[distributedSearch]
```

```
#peerResolutionThreads = 32 - no longer supported
```

```
#trySSLFirst = false - no longer supported
```

```
statusTimeout = 120
```

```
authTokenConnectionTimeout = 120
```

```
authTokenReceiveTimeout = 120
```

```
authTokenSendTimeout = 120
```

```
bestEffortSearch = true
```

```
connectionTimeout = 120
```

```
sendTimeout = 120
```

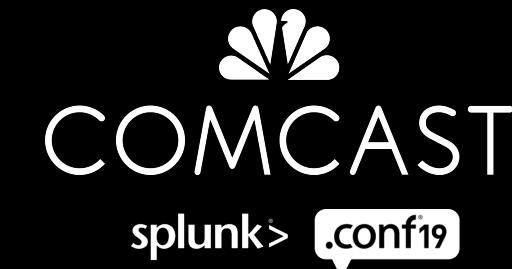
```
[replicationBlacklist]
```

```
userApps = users/*/*/lookups/*.csv
```

Lessons... Learned?

- ▶ Sometimes networking doesn't notify you of outages.
- ▶ Users are going to beat you to problems, let it be.
- ▶ Don't be a hero, be part of the community and focus on the hotspots.

INTERMISSION



ACT II

Automation

I'm a REALLY Lazy Engineer

With over 3000+ unique users, we need to be sure that Splunk has their data. Looking for more ways to self-heal is part of the mission. It's a shared resource.

- ▶ Let them eat cake: Self-Service Ingestion
- ▶ Bots!
- ▶ RBAC and Power Users
- ▶ Community First Goals



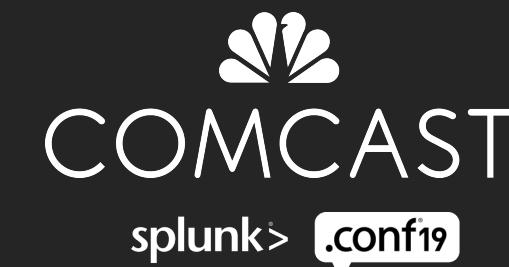
COMCAST

splunk> .conf19

Splunk Deployment Server end-to- end Configuration Management

git push and PR

The screenshot shows a Splunk Deployment Server build dashboard for build #496. The build was successful, completed on 21 Oct 2019 at 7:50:15 PM. The summary indicates a duration of 5 minutes and no labels. The revision is 406940e... and it was successful since build #493. The build included stages for Checkout Configs and btool, Check Config, Verify Deployment Servers, and SDS Configs Check against btool. The included in deployment project section shows Splunk Deployment Server Service > Release: deploy-SPL-SDS-496-deployment-446. The environment table shows Production and Production (Reload) both with SUCCESS status. The code commits table lists three commits from authors Sai Kalvala, catherine_rodriguez@cable.comcast.com, and hdave699. The Jira issues table shows one issue, SPLUNK-4338, with the message "Could not obtain issue details from Jira".



SDS Deployment

Fully Automatic Baby!

Job details Docker Tasks Requirements Artifacts Miscellaneous

Tasks

A task is a piece of work that is being executed as part of the build. The execution of a script, a shell command, an Ant Task or a Maven goal are only few examples of Tasks. Learn more about tasks.

You can use runtime, plan and global variables to parameterize your tasks.

Type	Description	Actions
Script	GHE Pending Status	
Source Code Checkout	Checkout Default Repository	
Command	make staging directory	
Command	copy splunk configs to tmp	
Command	create symbolic link for deployment-apps	
Command	create symbolic link for system	
Script	btool check source files and update GHE	
Command	btool list source files in /tmp/splunk/staging	
Command	copy conf to splunk directory	
Command	copy deployment-apps to splunk directory	
Command	btool check source files on splunk path	
Final tasks	Are always executed even if a previous task fails	
Command	clean up /tmp/splunk/	

Add task

No task selected
Select a task from the list on the left to configure it.

1 agent has the capabilities to run this job

Build dashboard / splunk
Splunk Deployment Server master Branch status

Used for updating and deploying configurations to the deploy server.

Plan summary Branches Recent failures History Tests Issues Deployments

Related deployments

Deployment project	Environment	Release	Release branch	Build result used	Result	Completed	Actions
Splunk Deployment Server Service	Production	deploy-SPL-SDS-496-deployment-446	master	② #496	Logs	21 Oct 2019 08:27 PM	↻ ↻
	Production (Reload)	deploy-SPL-SDS-496-deployment-446	master	② #496	Logs	21 Oct 2019 08:27 PM	↻ ↻

SDS Deployment

Fully Automatic Baby!

Job details Docker Tasks Requirements Artifacts Miscellaneous

Tasks

A task is a piece of work that is being executed as part of the build. The execution of a script, a shell command, an Ant Task or a Maven goal are only few examples of Tasks. Learn more about tasks.

You can use runtime, plan and global variables to parameterize your tasks.

Type	Description	Actions
Script	GHE Pending Status	
Source Code Checkout	Checkout Default Repository	
Command	make staging directory	
Command	copy splunk configs to tmp	
Command	create symbolic link for deployment-apps	
Command	create symbolic link for system	
Script	btool check source files and update GHE	
Command	btool list source files in /tmp/splunk/staging	
Command	copy conf to splunk directory	
Command	copy deployment-apps to splunk directory	
Command	btool check source files on splunk path	
Final tasks	Are always executed even if a previous task fails	
Command	clean up /tmp/splunk/	

Add task

No task selected
Select a task from the list on the left to configure it.

1 agent has the capabilities to run this job

Build dashboard / splunk
Splunk Deployment Server master Branch status

Used for updating and deploying configurations to the deploy server.

Plan summary Branches Recent failures History Tests Issues Deployments

Related deployments

Deployment project	Environment	Release	Release branch	Build result used	Result	Completed	Actions
Splunk Deployment Server Service	Production	deploy-SPL-SDS-496-deployment-446	master	② #496	Logs	21 Oct 2019 08:27 PM	↻ ↻
	Production (Reload)	deploy-SPL-SDS-496-deployment-446	master	② #496	Logs	21 Oct 2019 08:27 PM	↻ ↻

Can you approve my PR?

Sure, it's green for all checks!

Week 43 - Monday #846

Merged [skalva293 merged 2 commits into master from release](#) 2 days ago

Conversation 0 Commits 2 Checks 0 Files changed 1 +10 -0

crodr210 commented 2 days ago Member

No description provided.

hdave699 and others added some commits 2 days ago

SPLUNK-4338:Added new sourcetype for Recording Validations service f8f1792

Merge pull request #845 from hdave699/SPLUNK-4338 ... 94c29be

crodr210 requested review from khende810, skalva293 and tyenda357 2 days ago

skalva293 approved these changes 2 days ago View changes

skalva293 merged commit 406940e into master 2 days ago Hide details Revert

1 check passed

Splunk Configuration Check Build Passed! Details

Write Preview

Leave a comment

Attach files by dragging & dropping, selecting them, or pasting from the clipboard.

Styling with Markdown is supported Comment

ProTip! Add comments to specific lines under Files changed.

Reviewers

- skalva293 ✓
- khende810 ⚡
- tyenda357 ⚡

Assignees

No one—assign yourself

Labels

None yet

Projects

None yet

Milestone

No milestone

Notifications

Unsubscribe

You're receiving notifications because your review was requested.

3 participants

Lock conversation

You can opt in to our Beta for issue deletion. Try it.

Making the Community Responsible

You do it, you own it.

- ▶ Users can ingest on their own time frame.
 - ▶ Errors can be easily tracked.
- ▶ 12 Node Deployment Server Cluster can handle the nodes.



COMCAST
splunk> .conf19

Bots

Slack lets you interact...

11:40  Need help in creating a user access request in splunk Jira. It gives me an error. Pl advise how to get access to splunk.ccp.cable.comcast.com

11:40 **splunker** APP Do you need help with access to Splunk?

11:40  Yes.



11:40  **splunker** APP Yea, sure.

You will need to submit a request to offshore security via [JIRA] (<https://www.teamccp.com/jira/secure/CreateIssue!default.jspa>), and select the XOPS Project then Access Request.

11:41  Thanks will try.



COMCAST
splunk> .conf19

Bots

Slack lets you interact...

14:19 SDS Bamboo APP

splunk > Splunk Deployment Server > release, #118 passed. Changes by Vaneet Singla and Atlassian Bamboo <bamboo@example.com> Responsible Users: Vaneet Singla, Atlassian Bamboo

14:52 Kristerpher Henderson 🌱 @vaneet_singla Sorry was in dispose. Checking it now :)

14:55 github APP [splunk-sds-configurations:master] 7 new commits by Vaneet Singla and 2 others:

- 77b23d0 SPLUNK-744 - Vaneet Singla
- 6bdf5ca [bamboo] Automated branch merge (from develop:77b23d011f50828eee16362448ad72af351ee55) - Atlassian Bamboo
- aae345b SPLUNK-748 & SPLUNK-749 - Vaneet Singla
- f6c729d [bamboo] Automated branch merge (from develop:aae345b8f0b920b81a46218d673eed685f117bd3) - Atlassian Bamboo
- 8128ff0 SPLUNK-748 & SPLUNK-749 - Vaneet Singla [Show more...](#)

14:55 splunker APP

SPLUNK-744
Onecloud cluster logs adding into Splunk

Status	Assignee
IN PROGRESS	Singla, Vaneet
Reporter	Watchers
Chandra Pisati	Nazareth, Binoy, Chandra Pisati, Sasmita, Dania, Singla, Vaneet

[Watch](#) [Assign to me](#) [Dev Ready](#) [In Progress](#) [Rank Top](#)

SPLUNK-748
Need Splunk configured to ingest files for the xh_storm cluster. storm_nimbus

Status	Assignee
Needs more information	Singla, Vaneet
Reporter	Watchers
Bender, Kyle	Cramasta, Joe, Bender, Kyle, Singla, Vaneet

[Watch](#) [Assign to me](#) [Dev Ready](#) [In Progress](#) [Rank Top](#)

SPLUNK-749
Need Splunk configured to ingest files for the xh_storm cluster. storm_supervisor

Status	Assignee
Needs more information	Singla, Vaneet
Reporter	Watchers
Bender, Kyle	Cramasta, Joe, Bender, Kyle

[Watch](#) [Assign to me](#) [Dev Ready](#) [In Progress](#) [Rank Top](#)

14:55 Kristerpher Henderson 🌱 ^ so boss...

Security and RBAC

Program Groups get indexes.

- ▶ Each role in Splunk has an AD Group
- ▶ There are even groups for function, like Login
- ▶ Apps have roles too, and continue the hierarchy

Nesting Groups

Streamlines onboarding.

- ▶ Login
- ▶ -> User is part of Login
- ▶ -> Admin is part of User

Community First

Things we do to help.

- ▶ Publish Root Cause Analyses quickly.
- ▶ Help users “graduate” into better searches.
 - ▶ Guard Rails
 - ▶ Office Hours

Fin!

Thank You!

