The United States of Digital Search

Kristerpher Henderson
Principal Engineer | Comcast

COMCAST

splunk> .conf19

COMCAST

Prologue

Who is this guy?

.conf19
splunk>

Kristerpher

Technology | Arts | Baseball | Photography

COMCAST

NOAA

pjm

Technologist

Semi-Pro Baseball Player

Silent Actor in Opera

CODE FOR PHILLY

Open Source and Code for Philly Activist

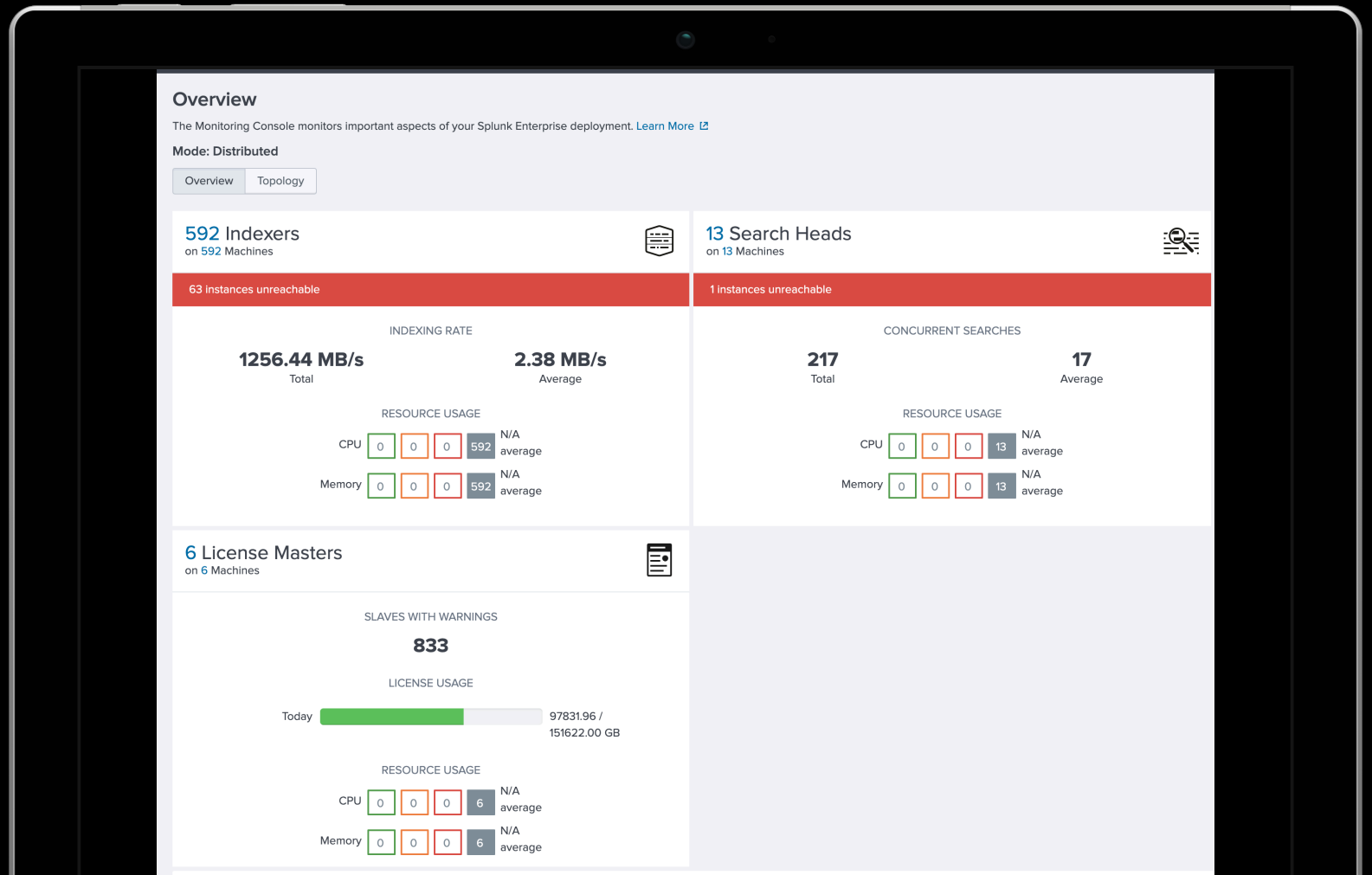
You're friendly neighborhood splunker

COMCAST

splunk> .conf19

Splunk at Comcast

All the data…

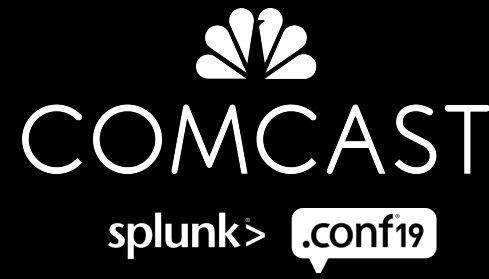


- Over 750 indexers
- 12 Node SHC across multiple AWS Regions
- Few indexes but lots of sourcetypes

COMCAST

splunk> .conf19

PROLOGUE



What's in there?

Technology Product and Experience

[illegible]
products respond to customer actions.

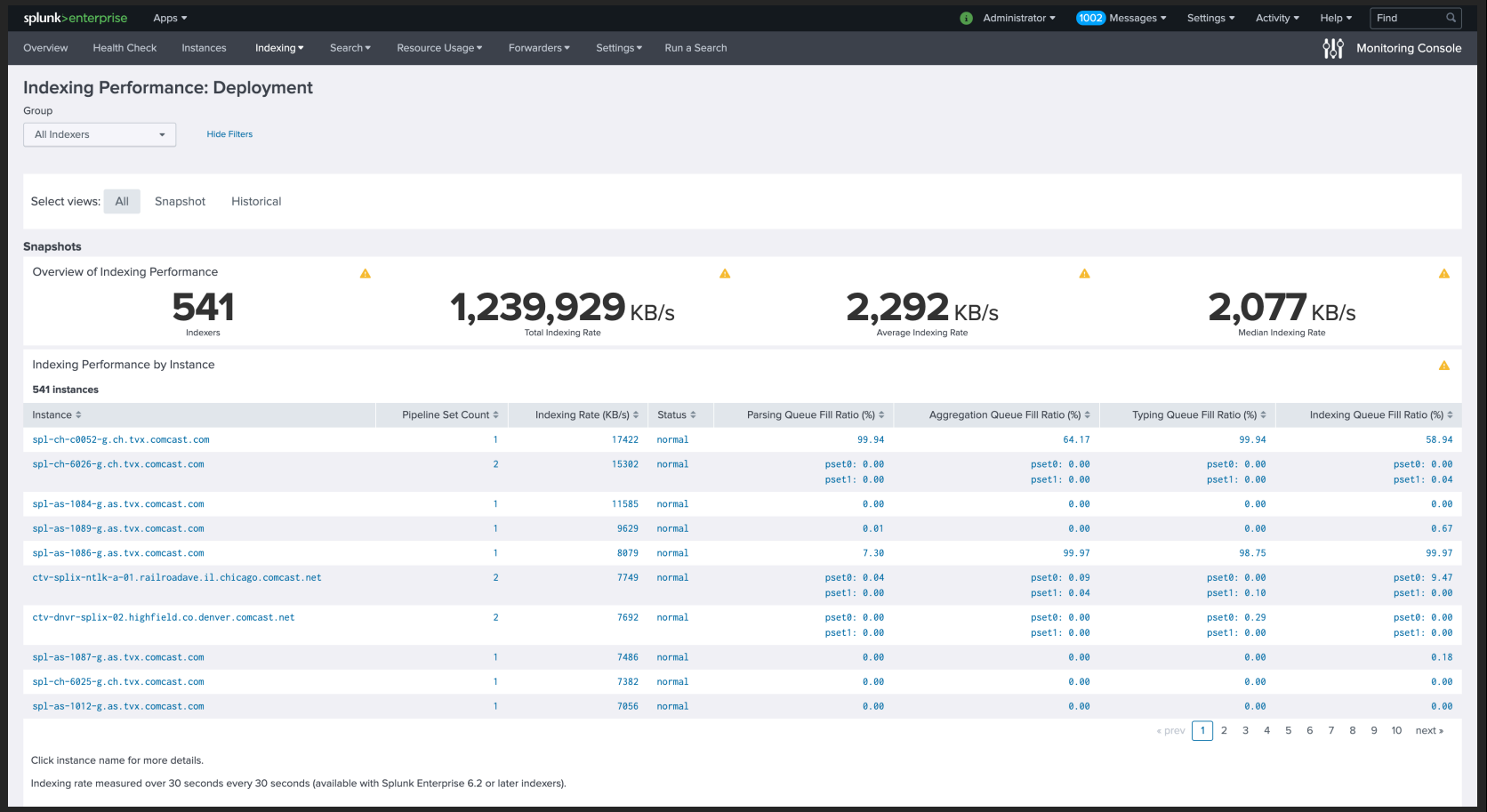
- 20+ Million Subscribers STB
- XRE + RDK
- Guide, Voice and Home
- Support Services

COMCAST

splunk> .conf19

Our biggest failures have lead to our greatest successes:

How we learn and scale



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.

COMCAST

splunk> .conf19

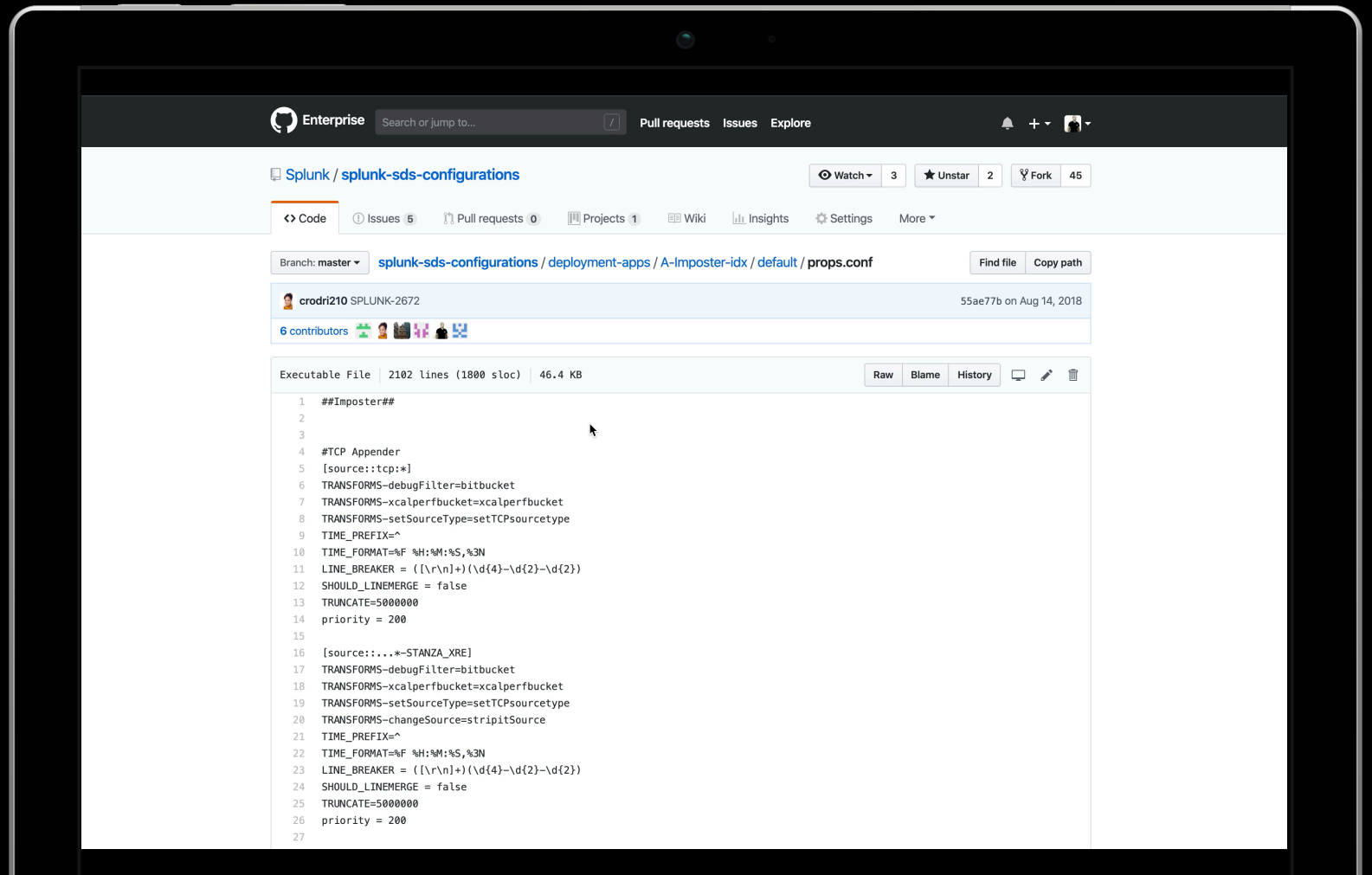
How we learn and scale

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

COMCAST

splunk> .conf19

Imposter



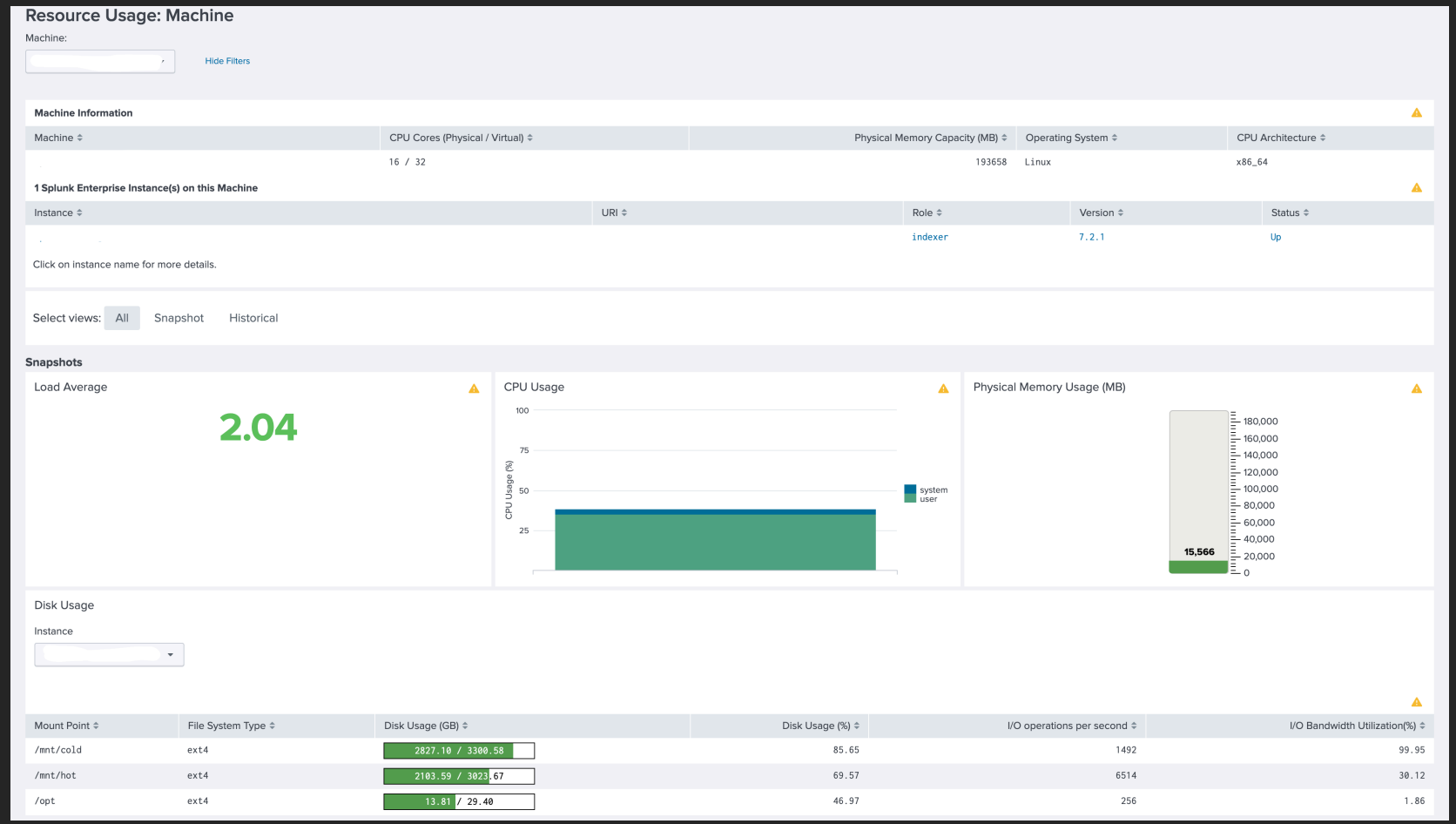
- An Imposter App is deployed to all Indexers.
- It allows for us to match specific data formats without needing to restart the indexers.

COMCAST

splunk> .conf19

What's under the hood?

Check the headlight fluid



COMCAST

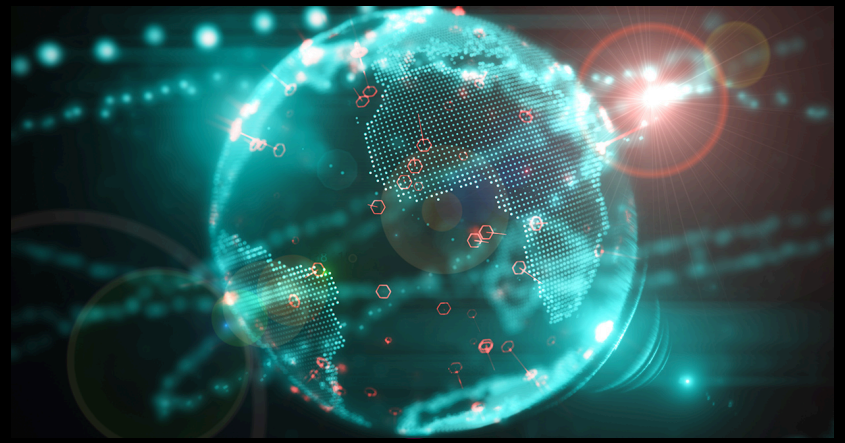
splunk> .conf19

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…

COMCAST

splunk> .conf19

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

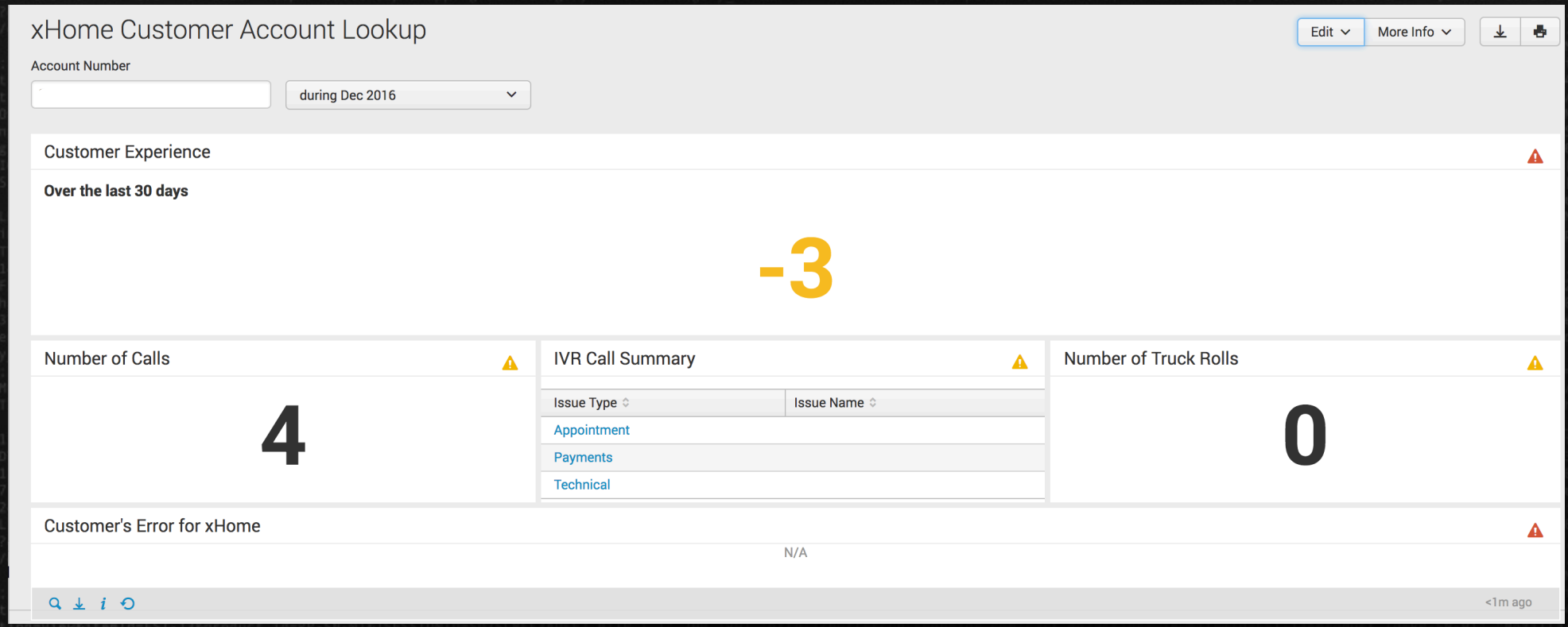
COMCAST

splunk> .conf19

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.

COMCAST

splunk> .conf19



Scoring:

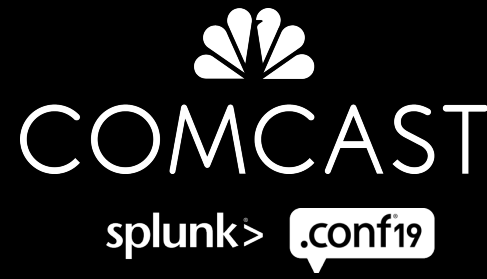
Did they call (-3)

Did they have an error (-1)

Did we visit them (-6)

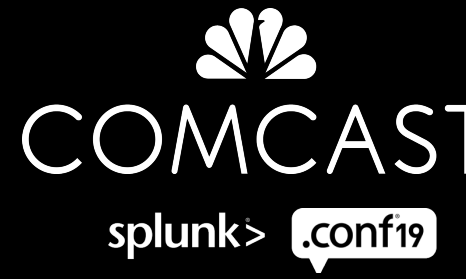
COMCAST

splunk> .conf19



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.

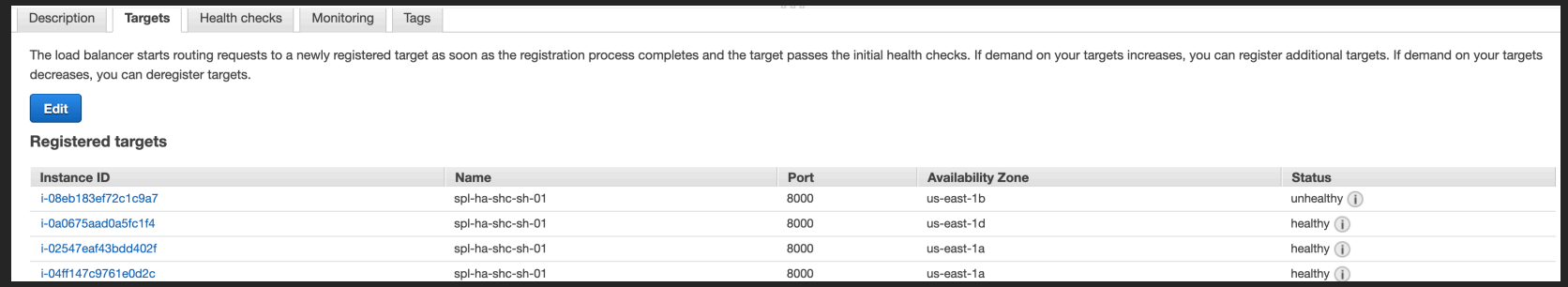
- 12 Node Search Head Cluster
- Distributed Search
- Splunk Deployment Server (Cluster)
- Monitoring via OP5/Zabbix

COMCAST

splunk> .conf19

Multi-Region Search Head Cluster

Spec'd to survive a complete data center outage.



Description | **Targets** | Health checks | Monitoring | Tags

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.

Edit

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

COMCAST

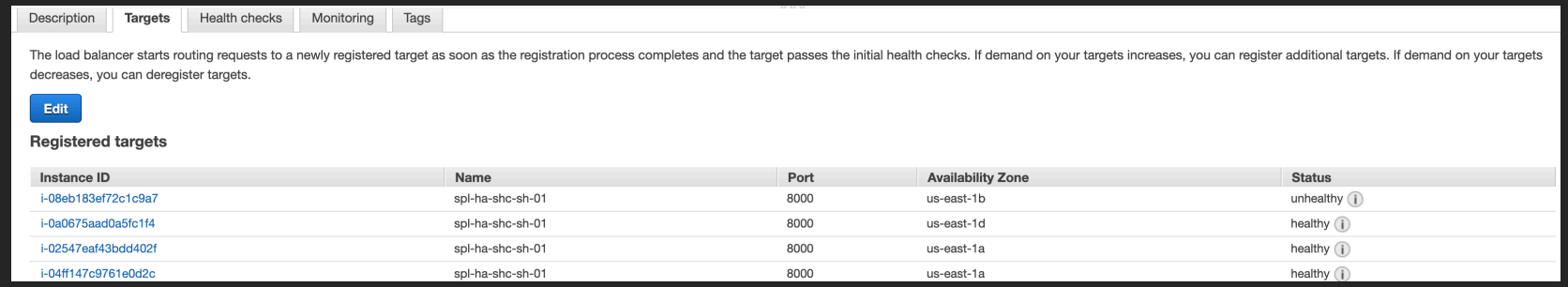
splunk> .conf19

3 AWS Regions

US-EAST-1 (4)

US-EAST-2 (Waiting on DX)

US-WEST-2 (8)



Description | **Targets** | Health checks | Monitoring | Tags

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.

Edit

Registered targets

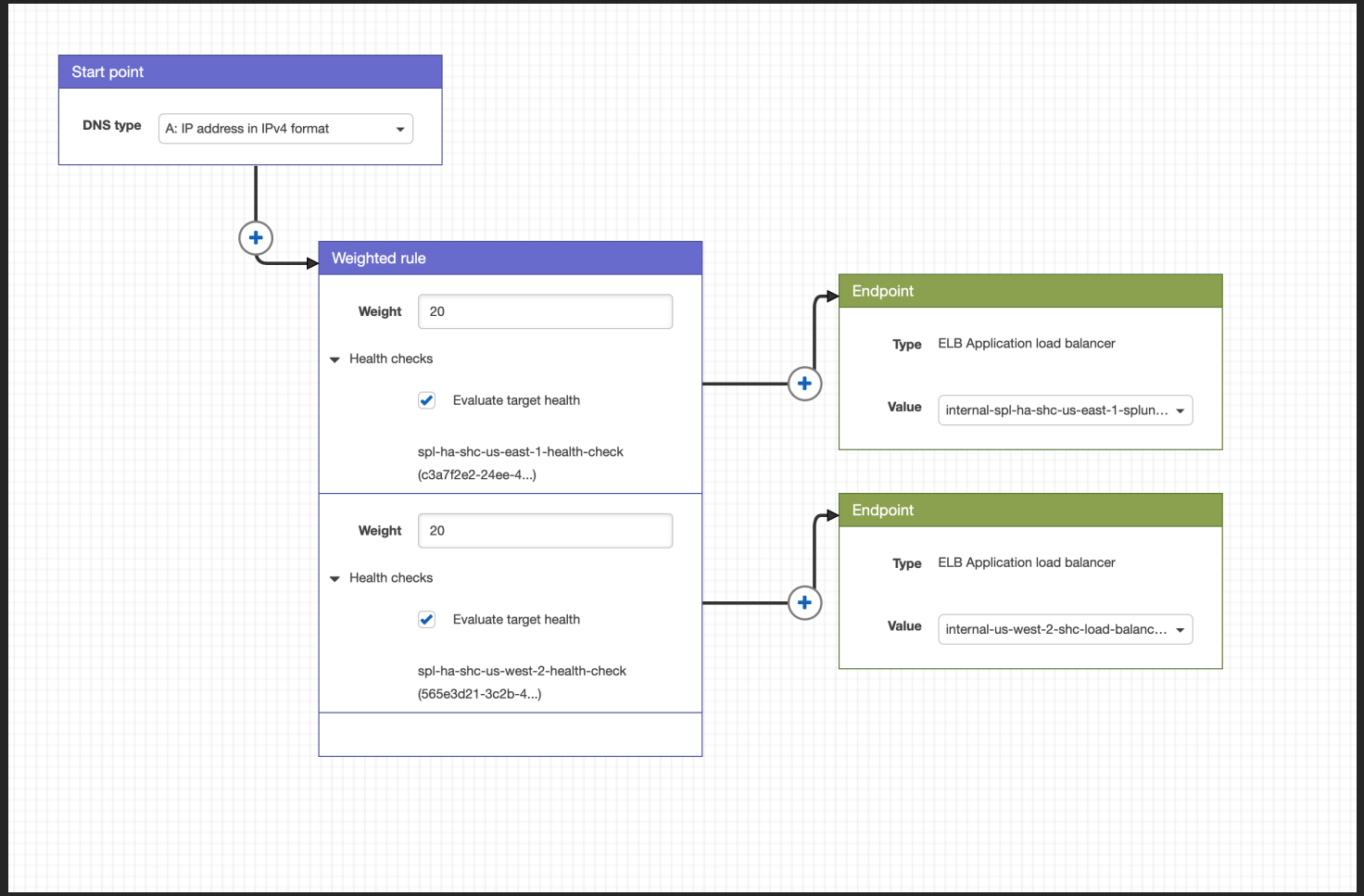
Instance ID	Name	Port	Availability Zone	Status
i-08eb183ef72c1c9a7	spl-ha-shc-sh-01	8000	us-east-1b	unhealthy
i-0a0675aad0a5fc1f4	spl-ha-shc-sh-01	8000	us-east-1d	healthy
i-02547eaf43bdd402f	spl-ha-shc-sh-01	8000	us-east-1a	healthy
i-04ff147c9761e0d2c	spl-ha-shc-sh-01	8000	us-east-1a	healthy

COMCAST

splunk> .conf19

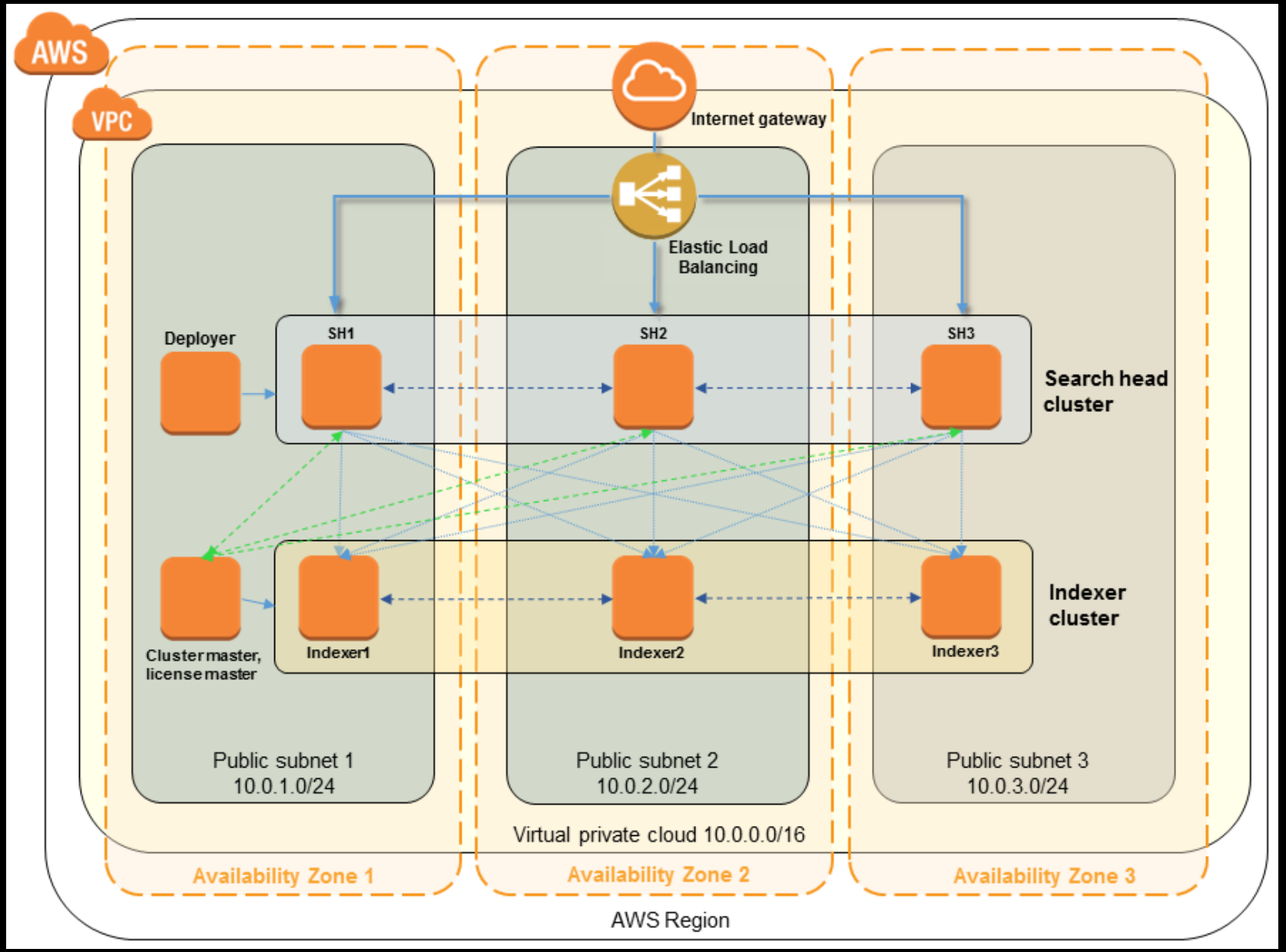
Routing Policy

Zone Records are controlled via AWS



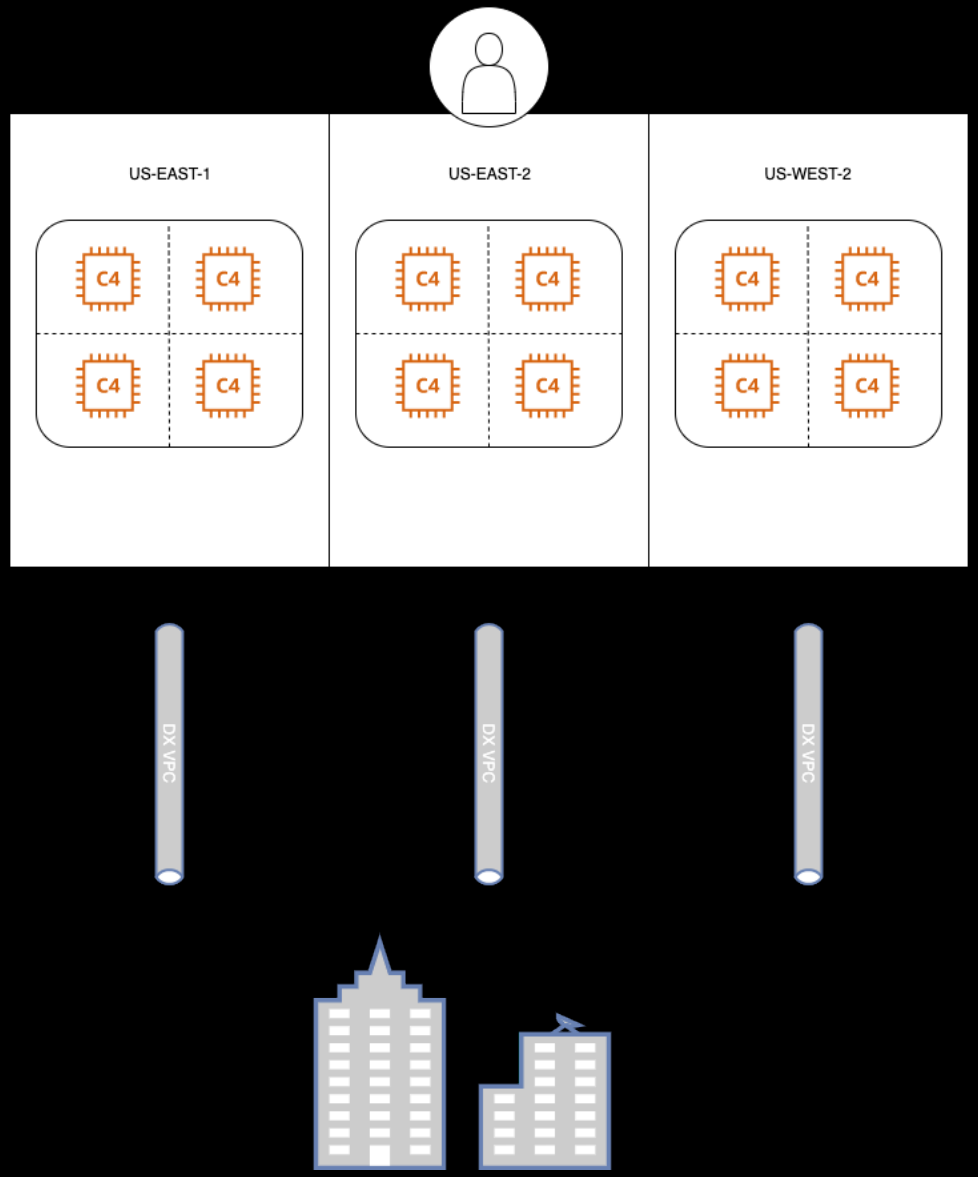
COMCAST

splunk> .conf19



COMCAST

splunk> .conf19

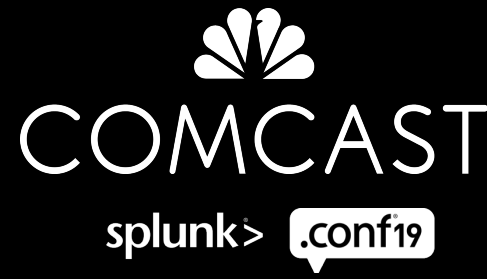


COMCAST

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.

COMCAST

splunk> .conf19

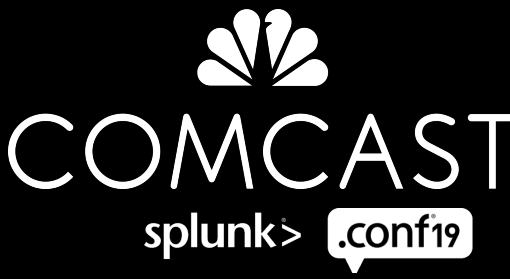
The Problems

…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
```

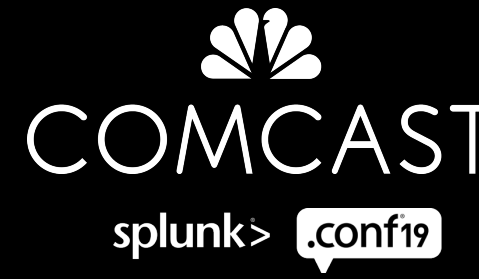


Lessons…

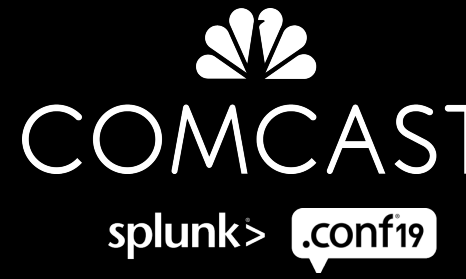
Learned?

- 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.

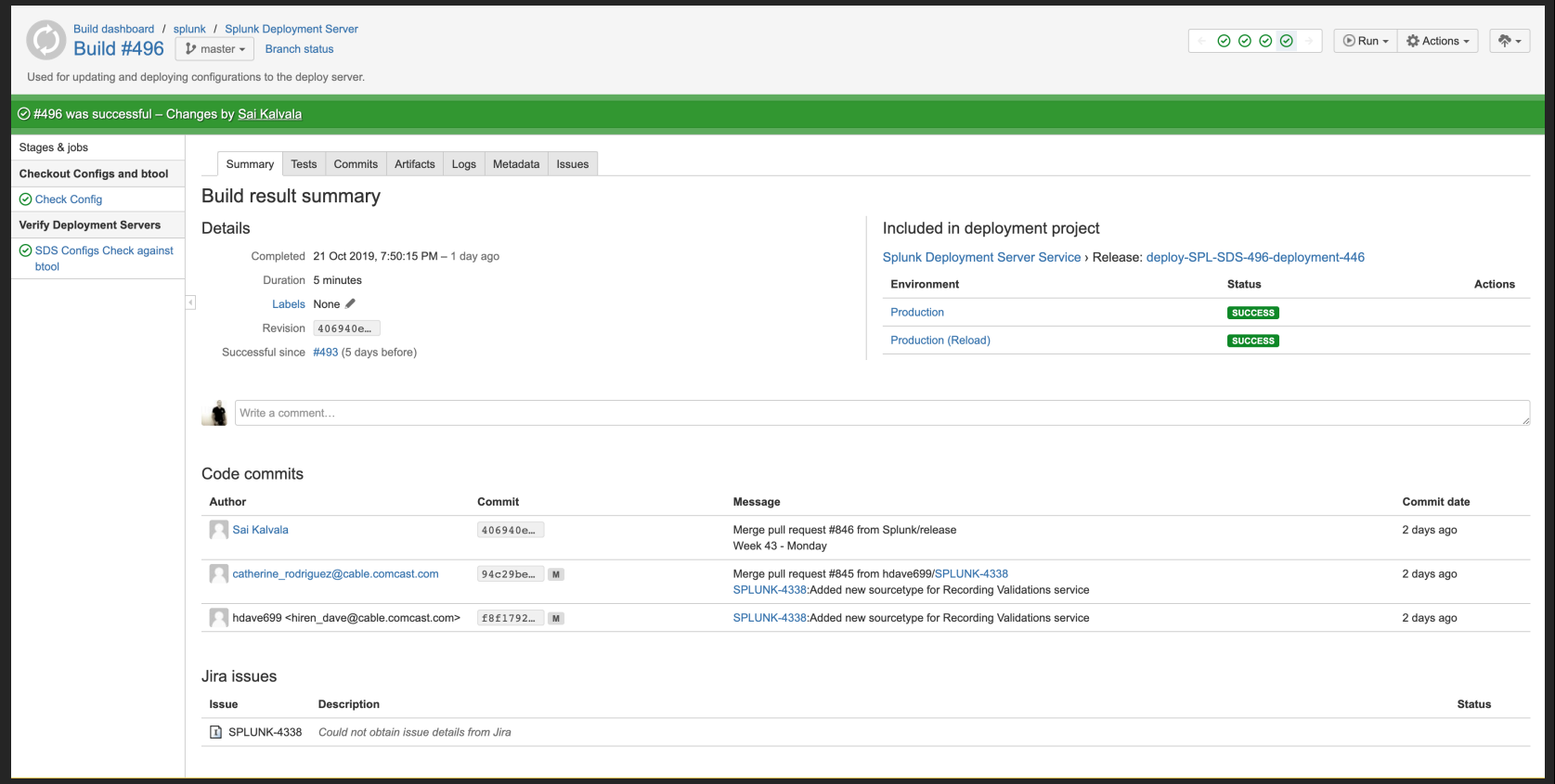
- Let them eat cake: Self-Service Ingestion
- Bots!
- RBAC and Power Users
- Community First Goals

COMCAST

splunk> .conf19

Splunk Deployment Server end-to-end Configuration Management

git push and PR

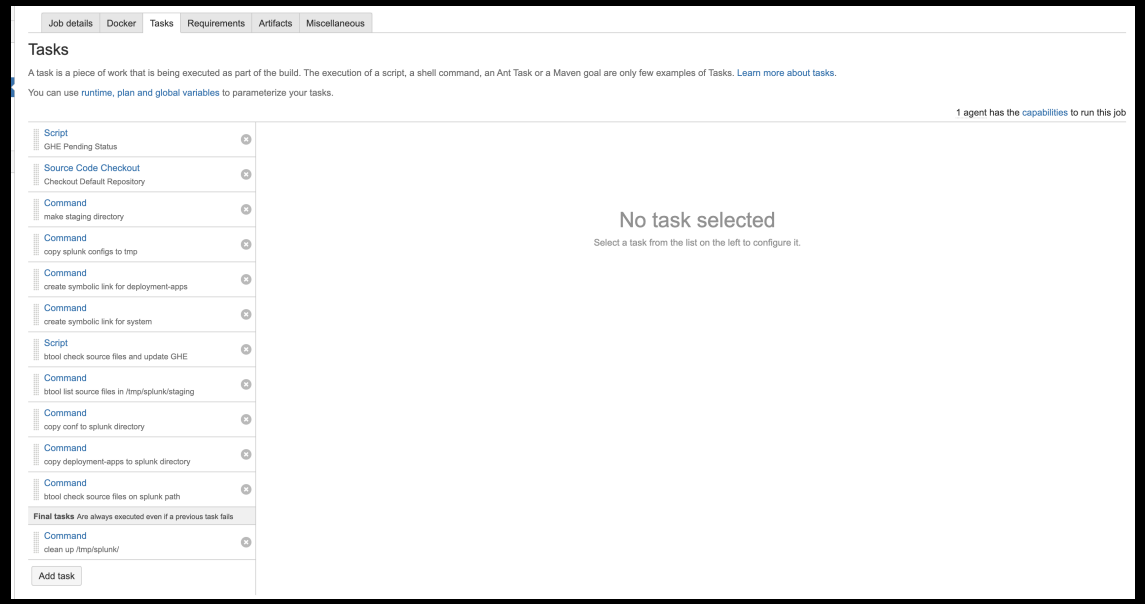


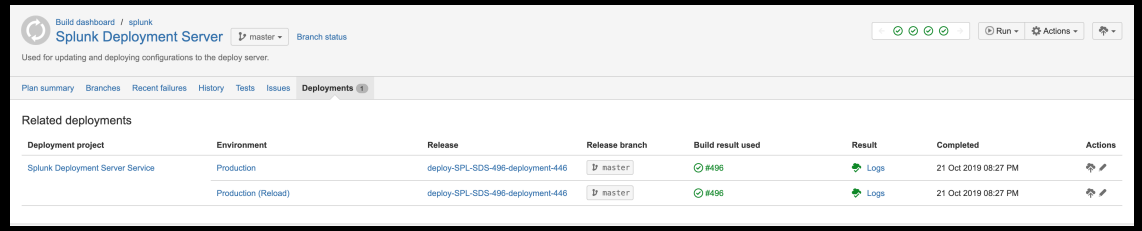
COMCAST

splunk> .conf19

SDS Deployment

Fully Automatic Baby!



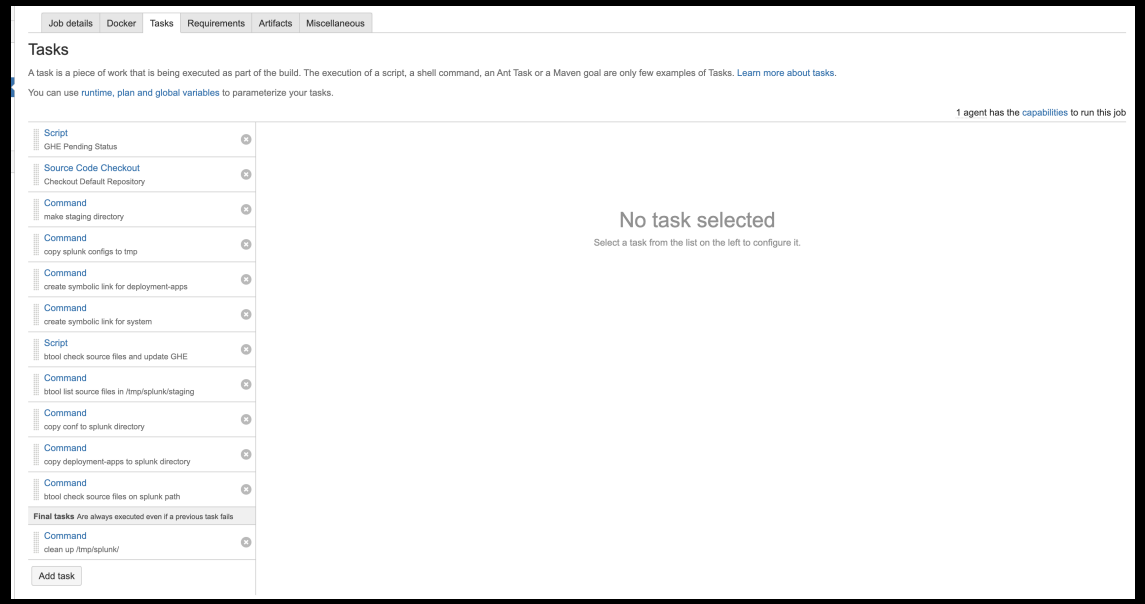


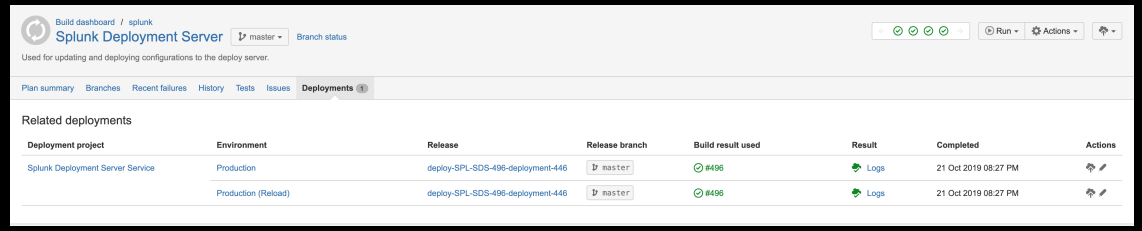
COMCAST

splunk> .conf19

SDS Deployment

Fully Automatic Baby!



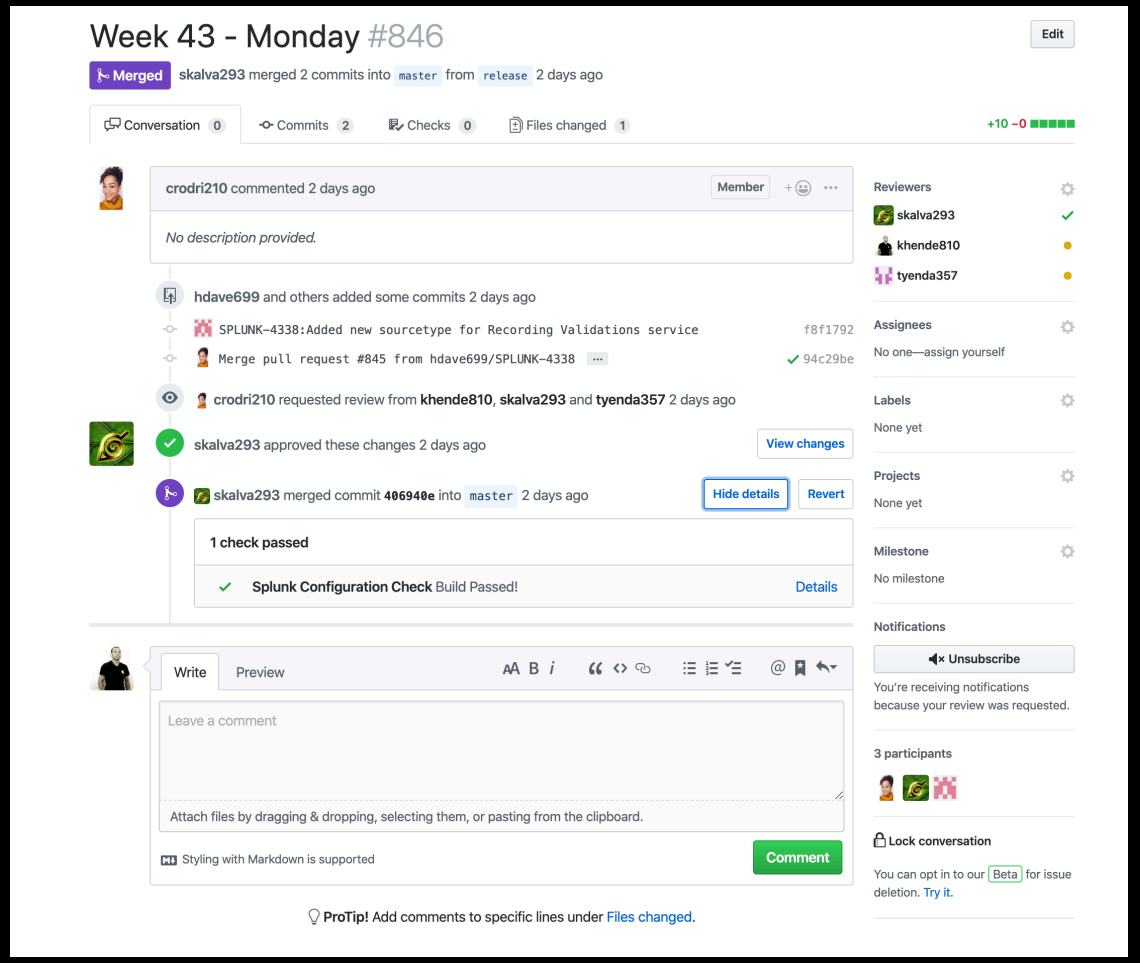


COMCAST

splunk> .conf19

Can you approve my PR?

Sure, it's green for all checks!



COMCAST

splunk> .conf19

Making the Community Responsible

You do it, you own it.

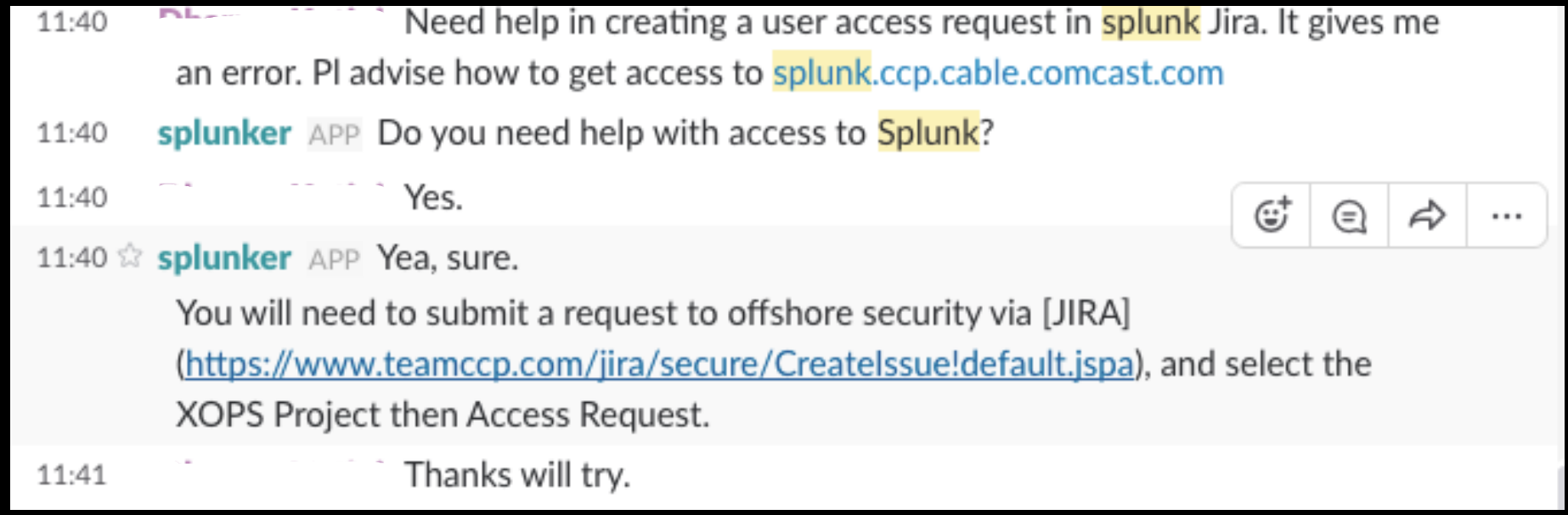
- Users can ingest on their own time frame.
  - Errors can be easily tracked.
- 12 Node Deployment Server Cluster can handle the nodes.

COMCAST

splunk> .conf19

Bots

Slack lets you interact…



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:77b23d011f50828eeea16362448ad72af351ee55) - 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** IN PROGRESS — **Assignee** Singla, Vaneet
> **Reporter** Chandra Pisati — **Watchers** Nazareth, Binoy, Chandra Pisati, Sasmita, Daniar, 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** Needs more information — **Assignee** Singla, Vaneet
> **Reporter** Bender, Kyle — **Watchers** 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** Needs more information — **Assignee** Singla, Vaneet
> **Reporter** Bender, Kyle — **Watchers** Cramasta, Joe, Bender, Kyle
> Watch | Assign to me | Dev Ready | In Progress | Rank Top

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

COMCAST

splunk> .conf19

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

COMCAST

splunk> .conf19

Fin!

Thank You!

splunk> .conf19