Dockerizing Splunk at Scale 2:

The Container Strikes Back

October 2018 | Version 1.0

There will be technical details

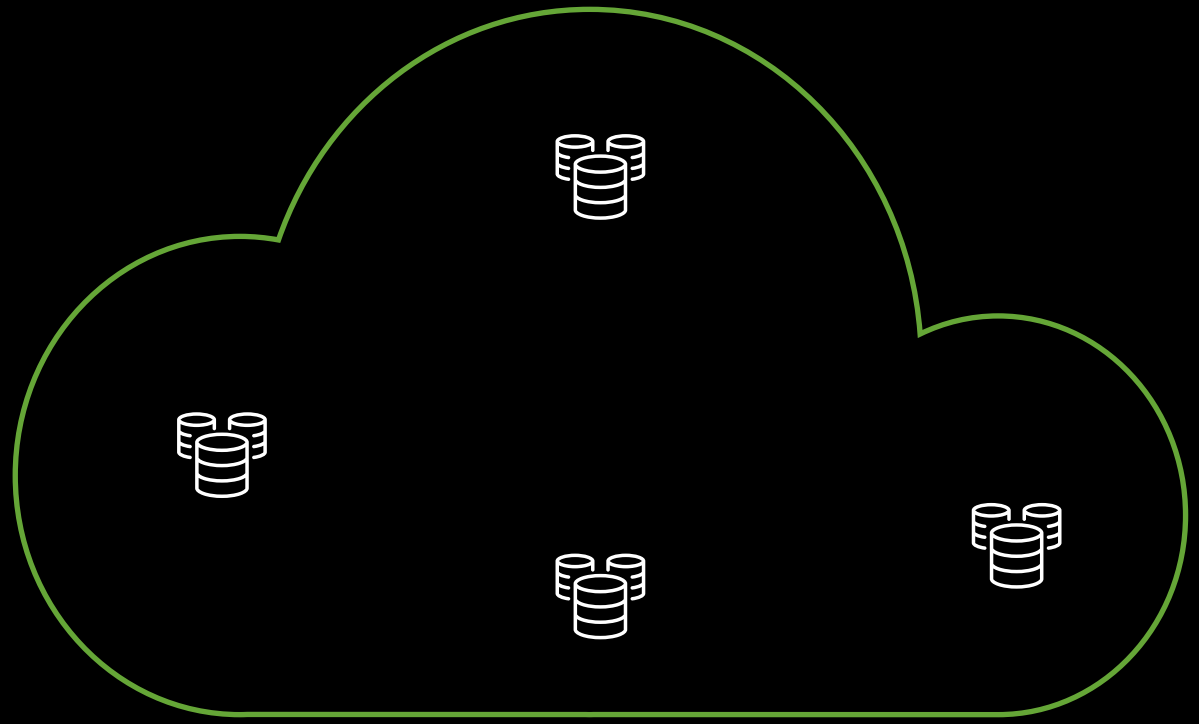
Forward-Looking Statements

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.

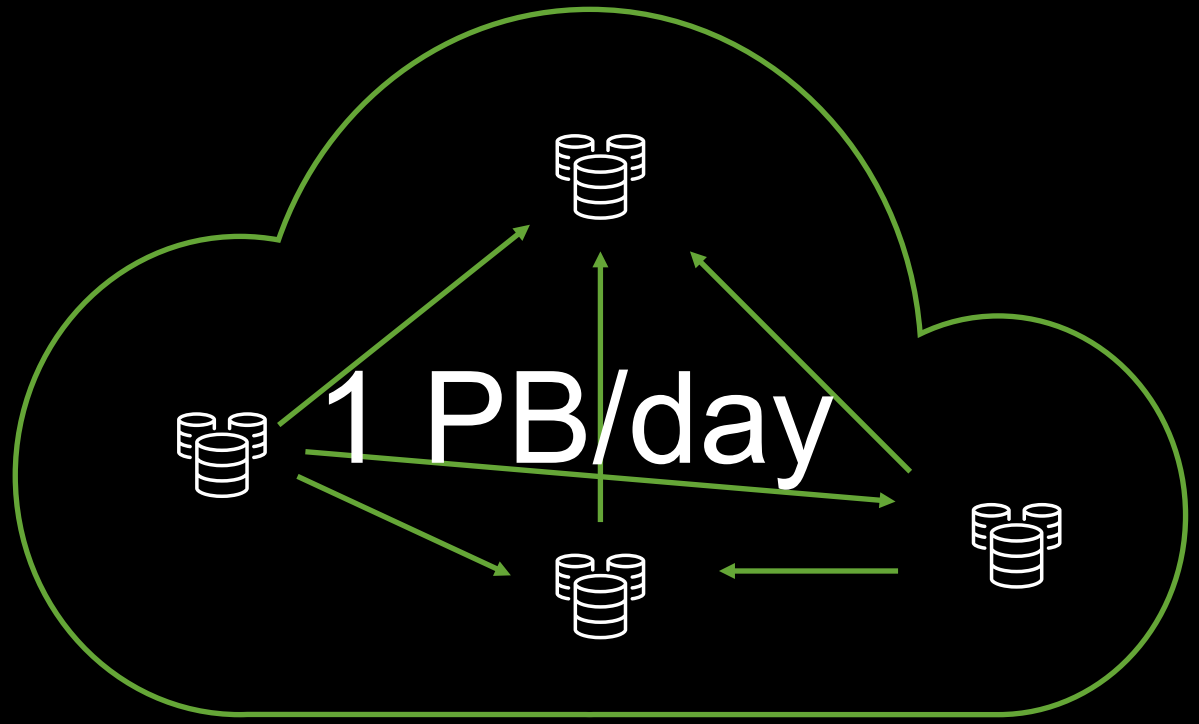
splunk> .conf18

What does Splunk at scale look like?



What does Splunk at scale look like?

- How about over 1 PB/day of ingestion via event generation?



What does Splunk at scale look like?

1500 Servers

What does Splunk at scale look like?

- 20 Week Testing Cycles

What is Splunk at Scale?

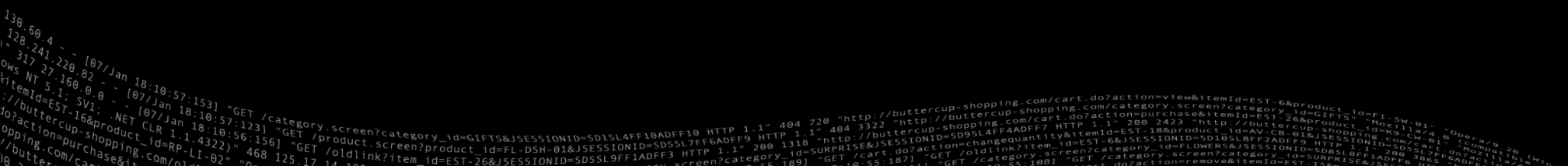
Guess what… it grew

- 1 PB / day of ingestion / event generation
- 1500 combined EC2 / On prem servers
- 20 Week Testing Cycles for performance
- 24 HR tests for QA / Development
- 1 MILLION+ containers / services created and destroyed per 24hrs

How did we do it?

So let's begin at the beginning

- About Us!
- Our Challenges
- Part 2: The Container Strikes Back
- Splunk and Containers from the inside
- Splunk and Containers from the outside
- Open sourcing is easy
- Question and Answer



About Us!

I Like to Pretend

(that I'm good at disc golf….)

- Developer at Splunk for 7 years
  - Splunk Apps -
  - ITSI (Up until 1.0)
  - VMWare (2.0-4.0)
  - NetApp (Guidance only)
  - ES (only on 2.1)
- Infrastructure
  - ORCA
  - Eventgen

A Little about Brent

- Working in technology since 2001
- Splunker since 2013
  - ITSI
  - Splunk for VMWare
  - Splunk for NetApp
  - ORCA (current project)
- Enjoys rock climbing.
- Star Wars fan

Our Challenges

Splunk is still hard!

But really, how scary can it be?

How scary can it be?

- Rapidly going from a single standalone instance to a cluster is easy

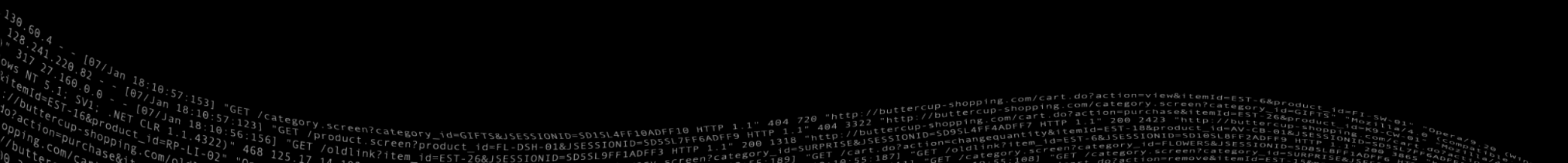
splunk> .conf18

How scary can it be?

- ~~Rapidly going from a single standalone instance to a cluster is easy~~

How scary can it be?

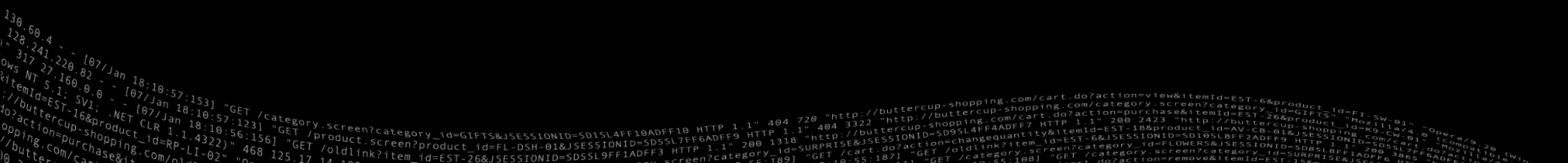
- ~~Rapidly going from a single standalone instance to a cluster is easy~~
- Writing apps for all the different Splunk architectures is straightforward



splunk> .conf18

How scary can it be?

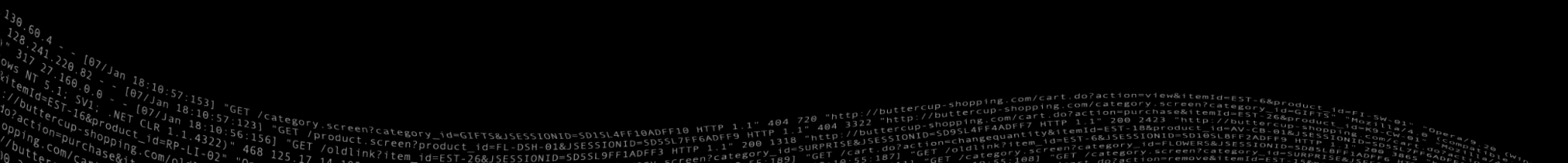
- ~~Rapidly going from a single standalone instance to a cluster is easy~~
- ~~Writing apps for all the different Splunk architectures is straightforward~~



splunk> .conf18

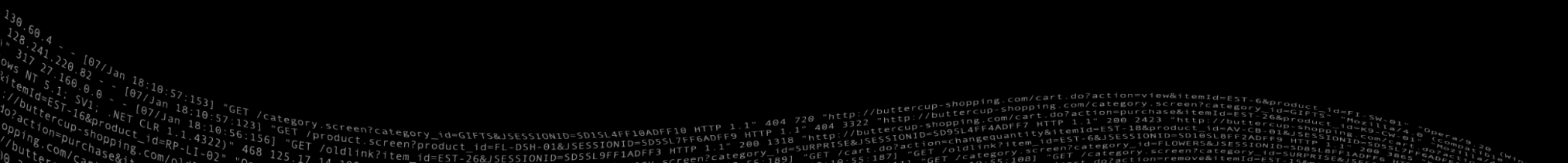
How scary can it be?

- ~~Rapidly going from a single standalone instance to a cluster is easy~~
- ~~Writing apps for all the different Splunk architectures is straightforward~~
- Splunk is easy to learn and has straightforward, easy to understand configurations



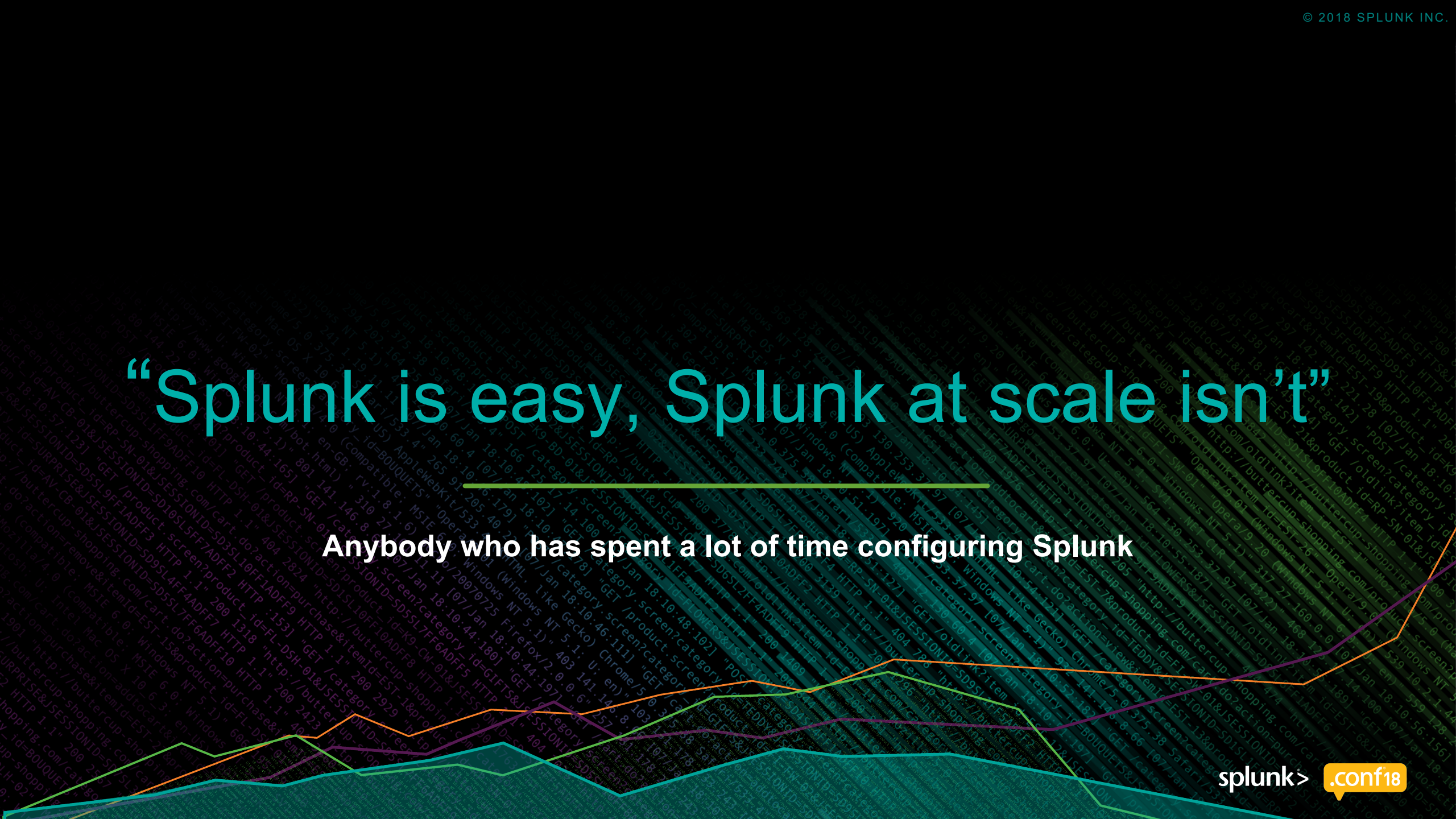
How scary can it be?

- ~~Rapidly going from a single standalone instance to a cluster is easy~~
- ~~Writing apps for all the different Splunk architectures is straightforward~~
- ~~Splunk is easy to learn and has straightforward, easy to understand configurations~~



“Splunk is easy, Splunk at scale isn’t”

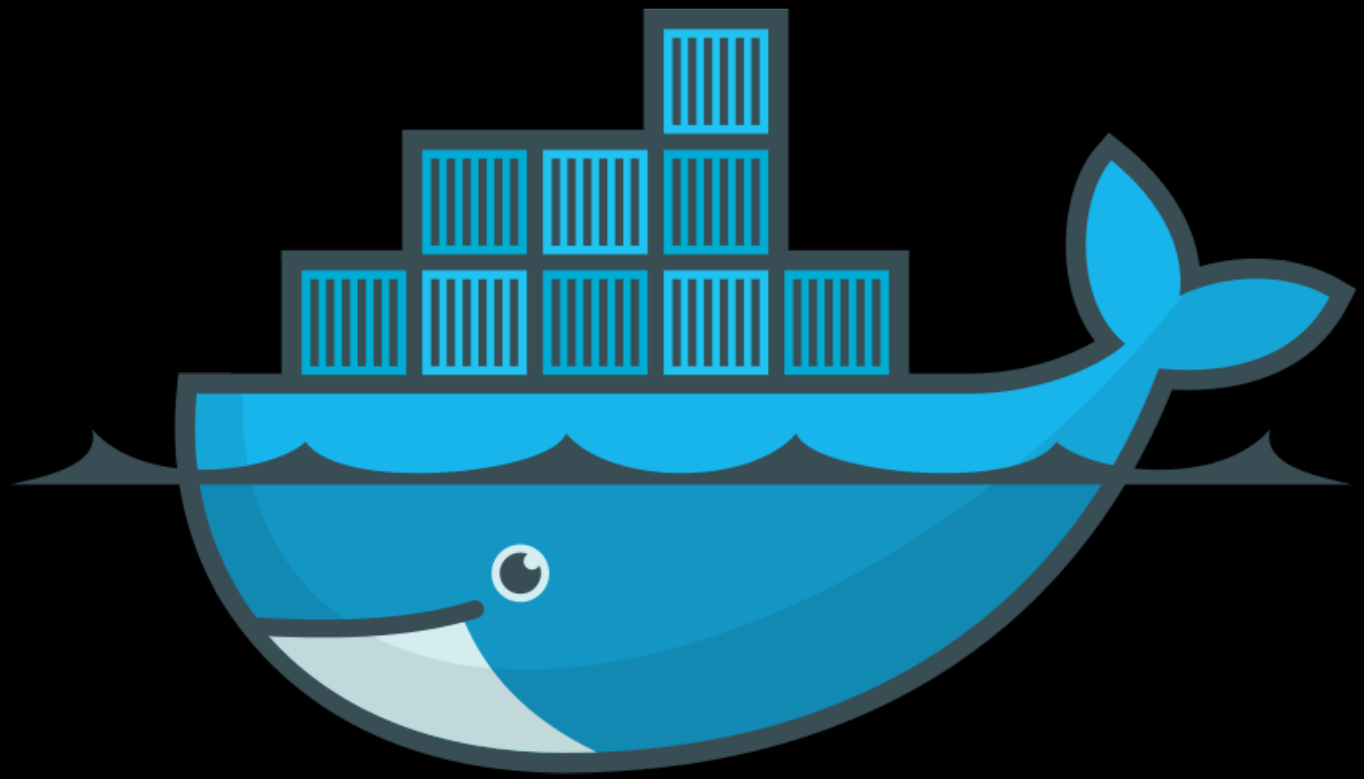
Anybody who has spent a lot of time configuring Splunk



Splunk and Containers

Major Platforms

Contain yourselves



Major Platforms

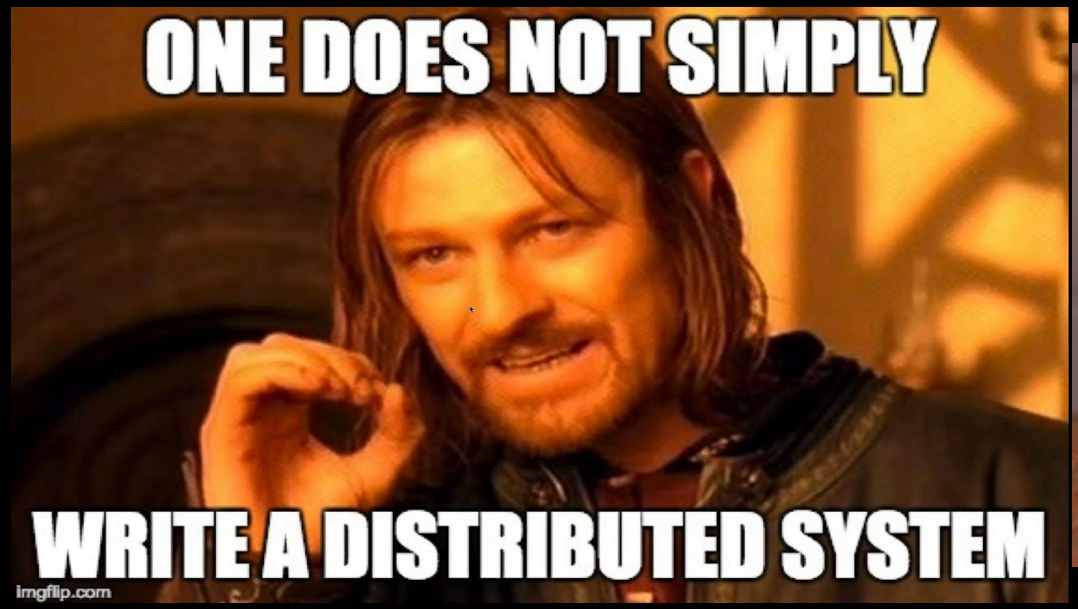
It's a bit of the wild west out there

Docker != Docker Swarm != Docker UCP != Kubernetes

People use containers differently

Major Platforms

We want to run splunk like…..



Local laptops

- Docker container on a local laptop
  - Test Splunk as a standalone environment
  - Play around with the product and the configuration
  - Usually promote the configurations to bare-metal or orchestrator installs

People use containers differently!

- Docker Swarm
  - Installing Splunk in a smaller clustered environment
  - Want support for distributed (high availability) Splunk

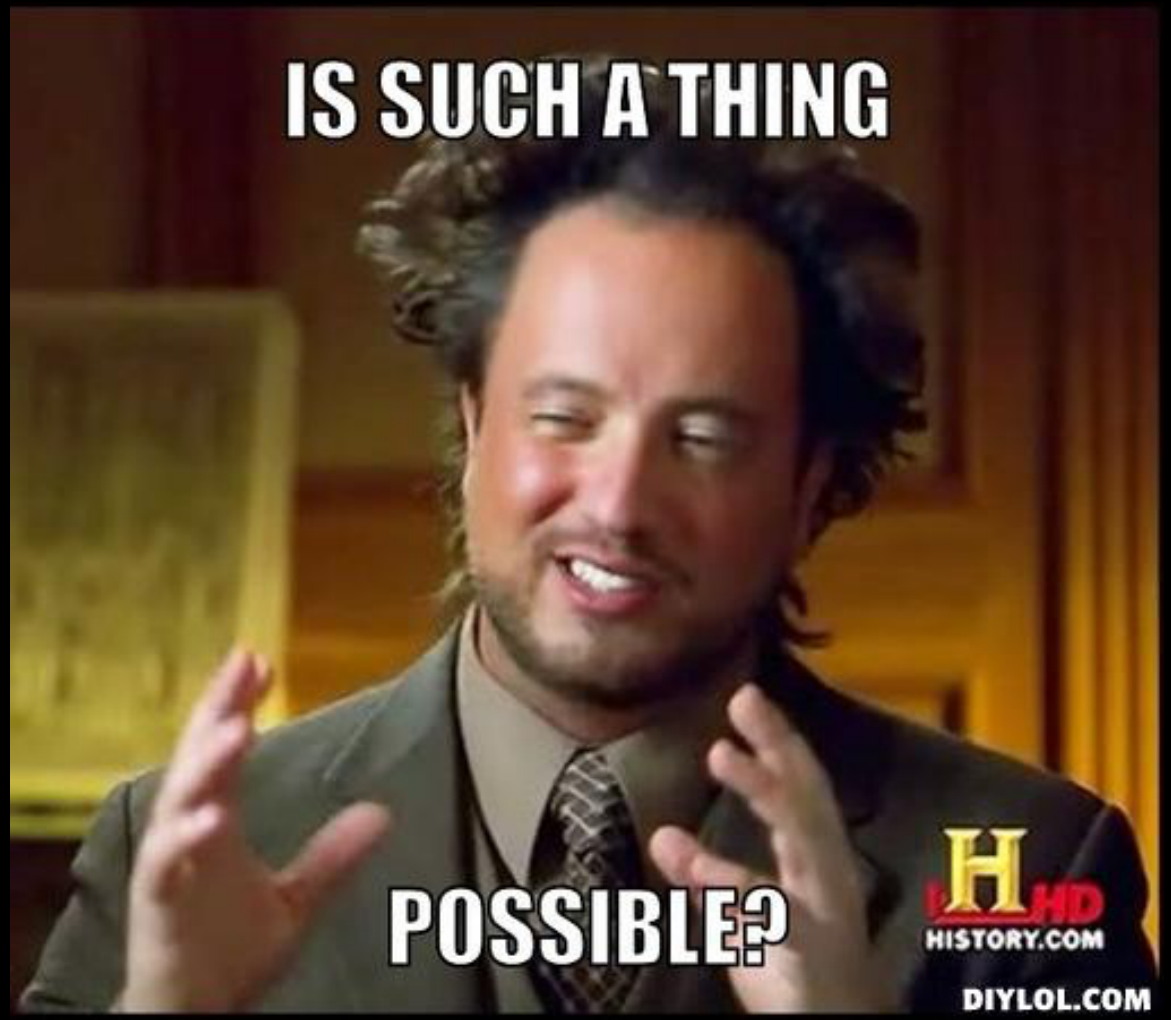
People use containers differently!

- Kubernetes / Docker UCP
  - Installing Splunk in a larger clustered environment
  - Want support for distributed (high availability) Splunk
  - Want the complex functionality of dynamically scaling Splunk
  - Usually complicated and unpredictable/unknown data volumes

People use containers differently!

- Just a few requirements:
  - Make Splunk easy to install on 1 environment, and have that same container be supported in bigger environments
  - Allow splunk to be deployed in any of the complex configurations based on the users need
  - Allow splunk to change configurations as I grow
  - Allow all the upgrade paths of docker containers to function
  - Persist my data for recovery / use on new containers
  - Get the same supported package that Splunk support blesses today, inside of the container
  - Do all of this, without me having to have in-depth knowledge of the product
  - One more thing, make it work on Kubernetes….

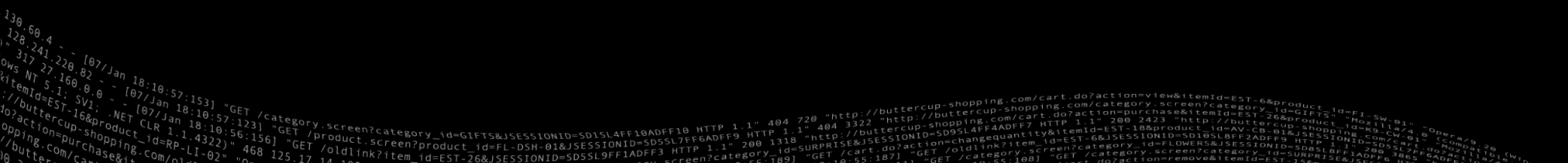
Major Platforms



So this is where we tell you about ORCA?

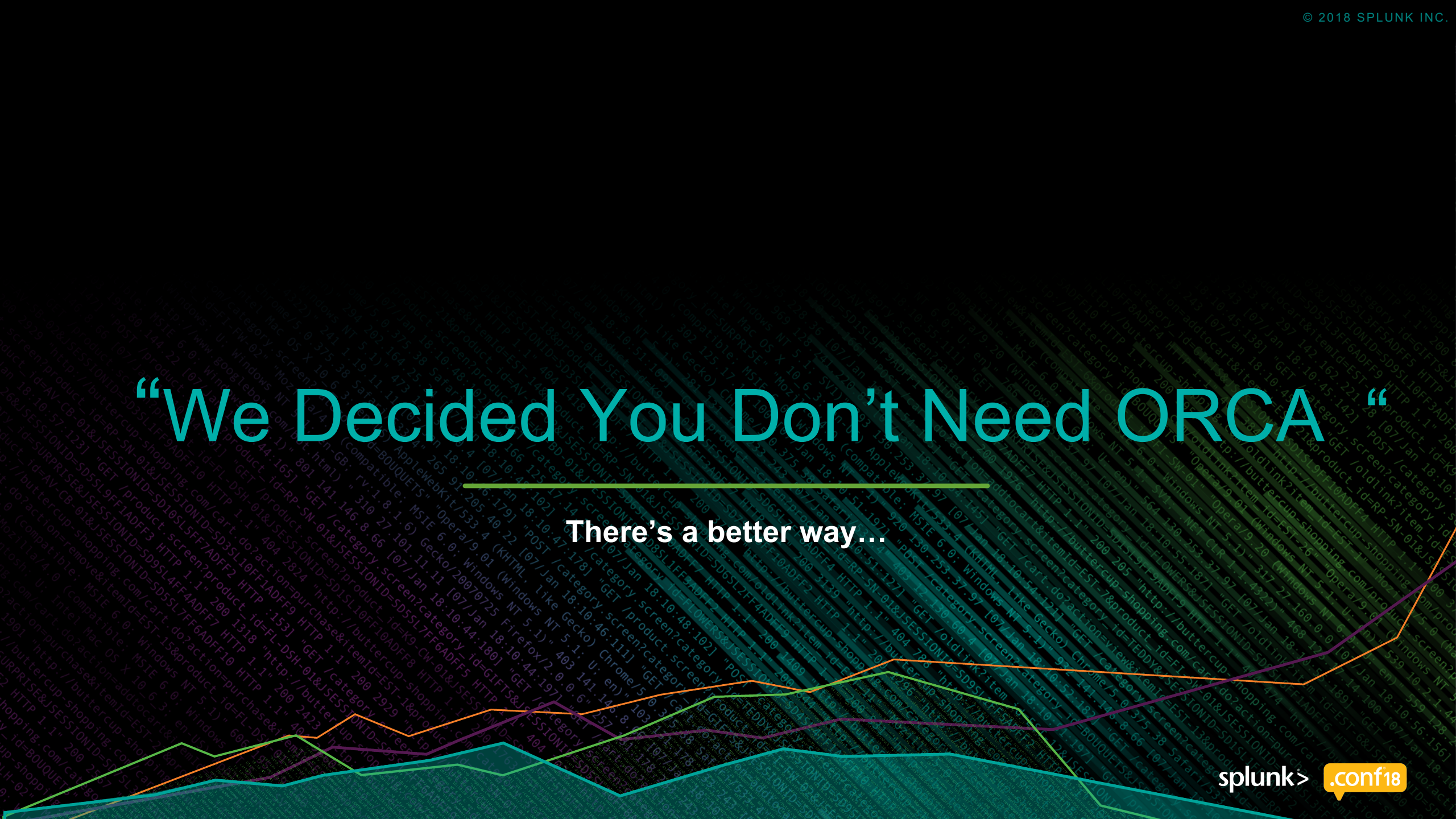
- Last year we told you about our internal tool for configuring splunk and deploying to ec2, openstack, docker ucp, etc, and demo'd some cool stuff… And you all wanted it.
- We also had mentioned that it was going open source, and would be available hopefully soon. You know… a year ago could be soon, right?

- WHATS UP WITH THAT?



“We Decided You Don’t Need ORCA “

There’s a better way…



splunk> .conf18

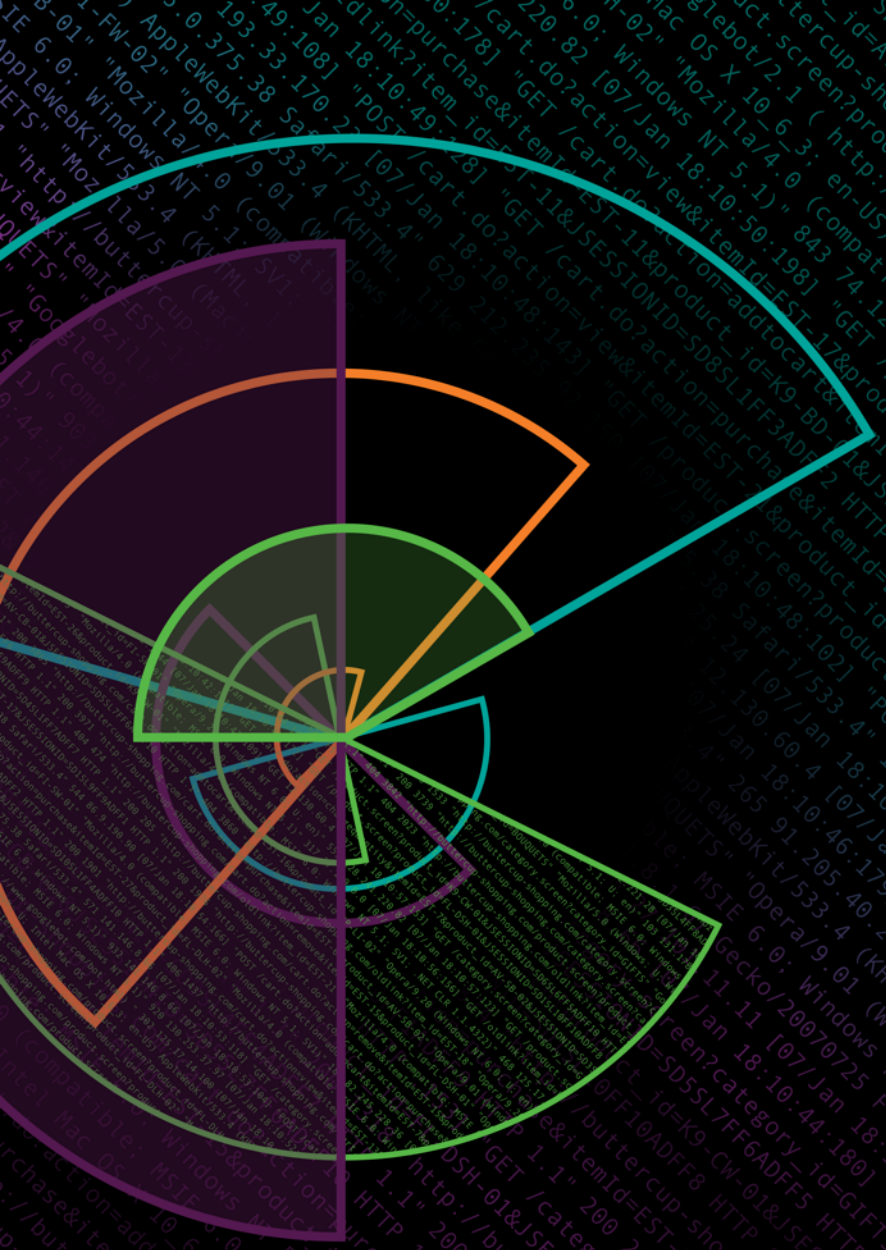
Part 2: The Container Strikes Back

The New Shiny

splunk> .conf18

Docker Container 1.0

- The new docker container allows you to download and deploy on any major container software that supports docker images.
- The simplest configuration can be for standalone deployment, while the most complex supports search head clustering with index clustering.
- Service objects can now upgrade containers with newer versions, while still maintaining custom mount points.
- You can expand your deployments.
- You can mount your volumes to almost any directory.
- You can bring up cluster with docker-compose (or virtually any other orchestrator's deployment tool)
- We tried to maintain backwards support for the unsupported docker images.
  - So hopefully your investment to the old open source container will still work
- Yes, it runs on Kubernetes.



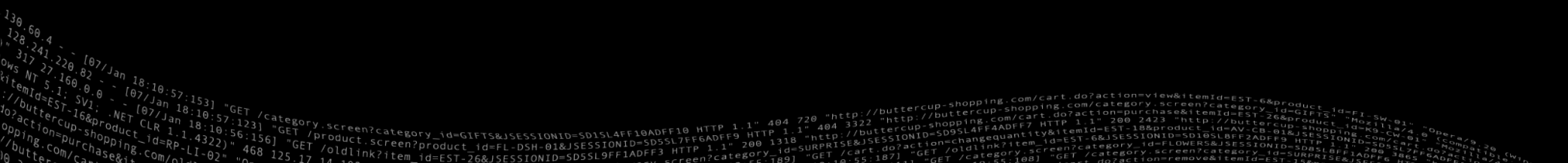
Demo

Installation and Configuration

How do I use the darn thing?

Requirements

- Linux Kernel 4.9 or higher
  - Centos 7.5
  - RHEL 7.5
  - Fedora 25
- Docker Engine 17.06.02 or higher
- UCP Enterprise Engine 2.2.11 or higher
- x86-x64 Chipsets



Starting the instance from the command line

- Command line
  docker run -d -p 8000:8000 -e "SPLUNK_START_ARGS=--accept-license" -e "SPLUNK_PASSWORD=Chang3d!" splunk/splunk:latest
- You'll see the container start
- Splunk Enterprise will not be available until the startup script finishes executing
- Use docker logs to follow the progress of the container

Configuring a Standalone Instance

- Docker Compose Format
- Use environment variables to pass in arguments
  - SPLUNK_START_ARGS - command line arguments
- Mount a defaults directory
  - Contains container specific information
  - Use the defaults mount to pass in sensitive information
    - Default password
    - HEC Tokens

```
version: "3.6"

networks:
  splunknet:
    driver: bridge
    attachable: true

services:
  so1:
    image: splunk/splunk:latest
    hostname: so1
    container_name: so1
    environment:
      - SPLUNK_START_ARGS=--accept-
license
    ports:
      - 8000
    volumes:
      - ./defaults:/tmp/defaults
```

splunk> .conf18

Configuring a Cluster

- General Environment Variables
  - SPLUNK_START_ARGS
  - SPLUNK_ROLE - search head, indexer, cluster master.
  - SPLUNK_LICENSE_URI - URI containing the Splunk Enterprise License
- Role specific environment variables
  - SPLUNK_INDEXER_URL
  - SPLUNK_SEARCH_HEAD_URL
  - SPLUNK_DEPLOYER_URL
  - SPLUNK_SEARCH_HEAD_CAPTAIN_URL
  - SPLUNK_HEAVY_FORWARDER_URL
- Many more roles in the documentation
- All roles and their values are documented.

Configuring Search Heads and Indexers

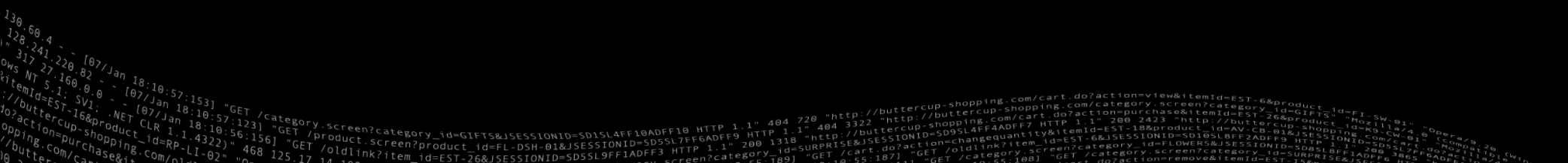
- Search Head Docker Compose Configuration
- Note the additional Environment Variables.

```
version: "3.6"

networks:
  splunknet:
    driver: bridge
    attachable: true

services:
  sh1:
    networks:
      splunknet:
        aliases:
          - sh1
    image: splunk/splunk:latest
    command: start
    hostname: sh1
    container_name: sh1
    environment:
      - SPLUNK_START_ARGS=--accept-license
      - SPLUNK_INDEXER_URL=idx1,idx2
      - SPLUNK_SEARCH_HEAD_URL=sh1,sh2
      - SPLUNK_ROLE=splunk_search_head
      - SPLUNK_LICENSE_URI=http://splunk-license-uri/splunk-license
    ports:
      - 8000
    volumes:
      - ./defaults:/tmp/defaults
```

Configuring Search Heads and Indexers

- Search Head Docker Compose Configuration
- Note the additional Environment Variables

```
sh2:
  networks:
    splunknet:
      aliases:
        - sh2
  image: splunk/splunk:latest
  command: start
  hostname: sh2
  container_name: sh2
  environment:
    - SPLUNK_START_ARGS=--accept-license
    - SPLUNK_INDEXER_URL=idx1,idx2
    - SPLUNK_SEARCH_HEAD_URL=sh1,sh2
    - SPLUNK_ROLE=splunk_search_head
    - SPLUNK_LICENSE_URI=http://splunk-license-uri/splunk-license
  ports:
    - 8000
  volumes:
    - ./defaults:/tmp/defaults
```



splunk> .conf18

Configuring Search Heads and Indexers

- Indexer Docker Compose Configuration
- Note the additional Environment Variables

```
idx1:
  networks:
    splunknet:
      aliases:
        - idx1
  image: splunk/splunk:latest
  command: start
  hostname: idx1
  container_name: idx1
  environment:
    - SPLUNK_START_ARGS=--accept-license
    - SPLUNK_INDEXER_URL=idx1,idx2
    - SPLUNK_SEARCH_HEAD_URL=sh1,sh2
    - SPLUNK_ROLE=splunk_indexer
    - SPLUNK_LICENSE_URI=http://splunk-license-uri/splunk-license
  ports:
    - 8000
  volumes:
    - ./defaults:/tmp/defaults
```

```
idx2:
    networks:
      splunknet:
        aliases:
          - idx2
    image: splunk/splunk:latest
    command: start
    hostname: idx2
    container_name: idx2
    environment:
      - SPLUNK_START_ARGS=--accept-license
      - SPLUNK_INDEXER_URL=idx1,idx2
      - SPLUNK_SEARCH_HEAD_URL=sh1,sh2
      - SPLUNK_ROLE=splunk_indexer
      - SPLUNK_LICENSE_URI=http://splunk-license-uri/splunk-license
    ports:
      - 8000
    volumes:
      - ./defaults:/tmp/defaults
```

Cluster Configurations

- Deployer and Cluster Master Configurations

```
dep1:
  networks:
    splunknet:
      aliases:
        - dep1
  image: splunk/splunk:latest
  command: start
  hostname: dep1
  container_name: dep1
  environment:
    - SPLUNK_START_ARGS=--accept-license
    - SPLUNK_INDEXER_URL=idx1,idx2,idx3
    - SPLUNK_SEARCH_HEAD_URL=sh2,sh3
    - SPLUNK_SEARCH_HEAD_CAPTAIN_URL=sh1
    - SPLUNK_CLUSTER_MASTER_URL=cm1
    - SPLUNK_ROLE=splunk_deployer
    - SPLUNK_DEPLOYER_URL=dep1
    - SPLUNK_LICENSE_URI=/tmp/defaults/splunk-license
  ports:
    - 8000
  volumes:
    - ./defaults:/tmp/defaults
```

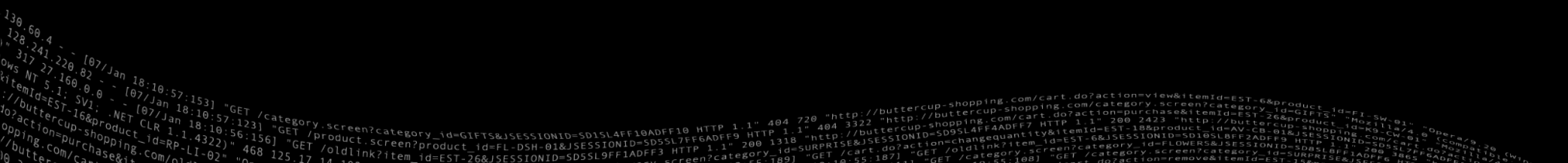
```
cm1:
  networks:
    splunknet:
      aliases:
        - cm1
  image: splunk/splunk:latest
  command: start
  hostname: cm1
  container_name: cm1
  environment:
    - SPLUNK_START_ARGS=--accept-license
    - SPLUNK_INDEXER_URL=idx1,idx2,idx3
    - SPLUNK_SEARCH_HEAD_URL=sh2,sh3
    - SPLUNK_SEARCH_HEAD_CAPTAIN_URL=sh1
    - SPLUNK_CLUSTER_MASTER_URL=cm1
    - SPLUNK_ROLE=splunk_cluster_master
    - SPLUNK_DEPLOYER_URL=dep1
    - SPLUNK_LICENSE_URI=/tmp/defaults/splunk-license
  ports:
    - 8000
  volumes:
    - ./defaults:/tmp/defaults
```

Inside of the Container

The gory details

Dockerfile

- The whirlwind tour of good to know pieces of our Dockerfile
  - Please refer to the open source files
- Based on debian:stretch-slim
- We install splunk and do some basic static configurations
- Includes a container health check
  - Pings the endpoint at 8000



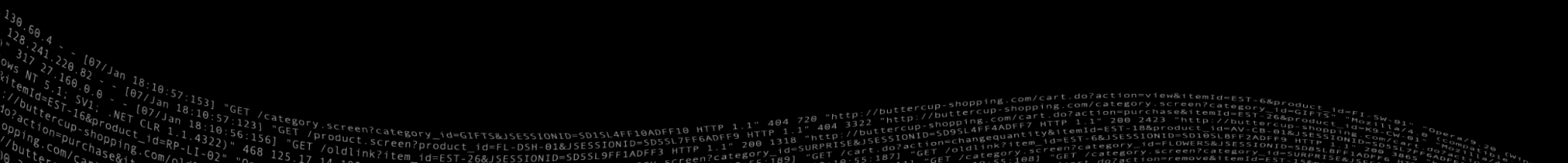
Dockerfile

- The file contents may change between when the slides were created and the date of the presentation

```
FROM base-debian-9:latest

ARG SPLUNK_FILENAME
ARG SPLUNK_BUILD_URL
ARG SPLUNK_DEFAULTS_URL

ENV SPLUNK_HOME=/opt/splunk \
    SPLUNK_GROUP=splunk \
    SPLUNK_USER=splunk \
    SPLUNK_ROLE=splunk_standalone \
    SPLUNK_FILENAME=${SPLUNK_FILENAME} \
    SPLUNK_DEFAULTS_URL=${SPLUNK_DEFAULTS_URL}

# Setup users and download Splunk
RUN groupadd -r ${SPLUNK_GROUP} \
    && useradd -r -m -g ${SPLUNK_GROUP} ${SPLUNK_USER} \
    && usermod -aG sudo ${SPLUNK_USER} \
    && sed -i -e 's/%sudo\s\+ALL=(ALL\(:ALL\)\?)\s\+ALL/%sudo ALL=NOPASSWD:ALL/g' /etc/sudoers \
    && echo "Downloading Splunk and validating the checksum at: ${SPLUNK_BUILD_URL}" \
    && wget -qO /tmp/${SPLUNK_FILENAME} ${SPLUNK_BUILD_URL} \
    && wget -qO /tmp/${SPLUNK_FILENAME}.sha512 ${SPLUNK_BUILD_URL}.sha512 \
    && (cd /tmp && sha512sum -c ${SPLUNK_FILENAME}.sha512) \
    && mv /tmp/${SPLUNK_FILENAME} /tmp/splunk.tgz \
    && rm -rf /tmp/${SPLUNK_FILENAME}.sha512

USER ${SPLUNK_USER}
COPY splunk-ansible /opt/ansible
COPY [ "splunk/debian-9/entrypoint.sh", "splunk/debian-9/checkstate.sh", "/sbin/" ]

EXPOSE 4001 8000 8065 8088 8089 8191 9887 9997
VOLUME [ "/opt/splunk/etc", "/opt/splunk/var" ]

HEALTHCHECK --interval=30s --timeout=30s --start-period=3m --retries=5 CMD /sbin/checkstate.sh || exit 1

ENTRYPOINT [ "/sbin/entrypoint.sh" ]
CMD [ "start-service" ]
```

Docker Container 1.0

- How do we configure the container on the inside?
  - Ansible
    - Ansible is configuration management software
  - We've been using it in house for 2 years
  - All plays are run inside of the containers themselves
  - More resilient than BASH scripts
  - Easier to understand than BASH scripts
  - More flexible than BASH scripts
  - Please do not ask us to "write the configuration in BASH"



What happens under the hood?

- The container starts
- An entrypoint script is executed, that starts splunk and then runs ansible
- We expect a default.yml file to contain the basic configuration
- Included in the official documentation are examples and templates

What happens under the hood?

- The contents of default.yml

```
---
retry_num: 100
splunk:
    opt: /opt
    home: /opt/splunk
    user: splunk
    group: splunk
    exec: /opt/splunk/bin/splunk
    pid: /opt/splunk/var/run/splunk/splunkd.pid
    password: "{{ splunk_password }}"
    svc_port: 8089
    s2s_port: 9997
    http_port: 8000
    hec_port: 8088
    hec_disabled: 0
    hec_enableSSL: 1
    hec_token: 00000000-0000-0000-0000-000000000000
    app_paths:
        default: /opt/splunk/etc/apps
        shc: /opt/splunk/etc/shcluster/apps
        idxc: /opt/splunk/etc/master-apps
        httpinput: /opt/splunk/etc/apps/splunk_httpinput

# Search Head Clustering
  shc:
      enable: false
      label: shc_label
      secret: SoS3cr3t!
      label: pickle_rick
      replication_factor: 3
      replication_port: 4001

  # Indexer Clustering
  idxc:
      label: idxc_label
      secret: SoS3cr3t!
      search_factor: 2
      label: evil_morty
      replication_factor: 3
      replication_port: 9887
```

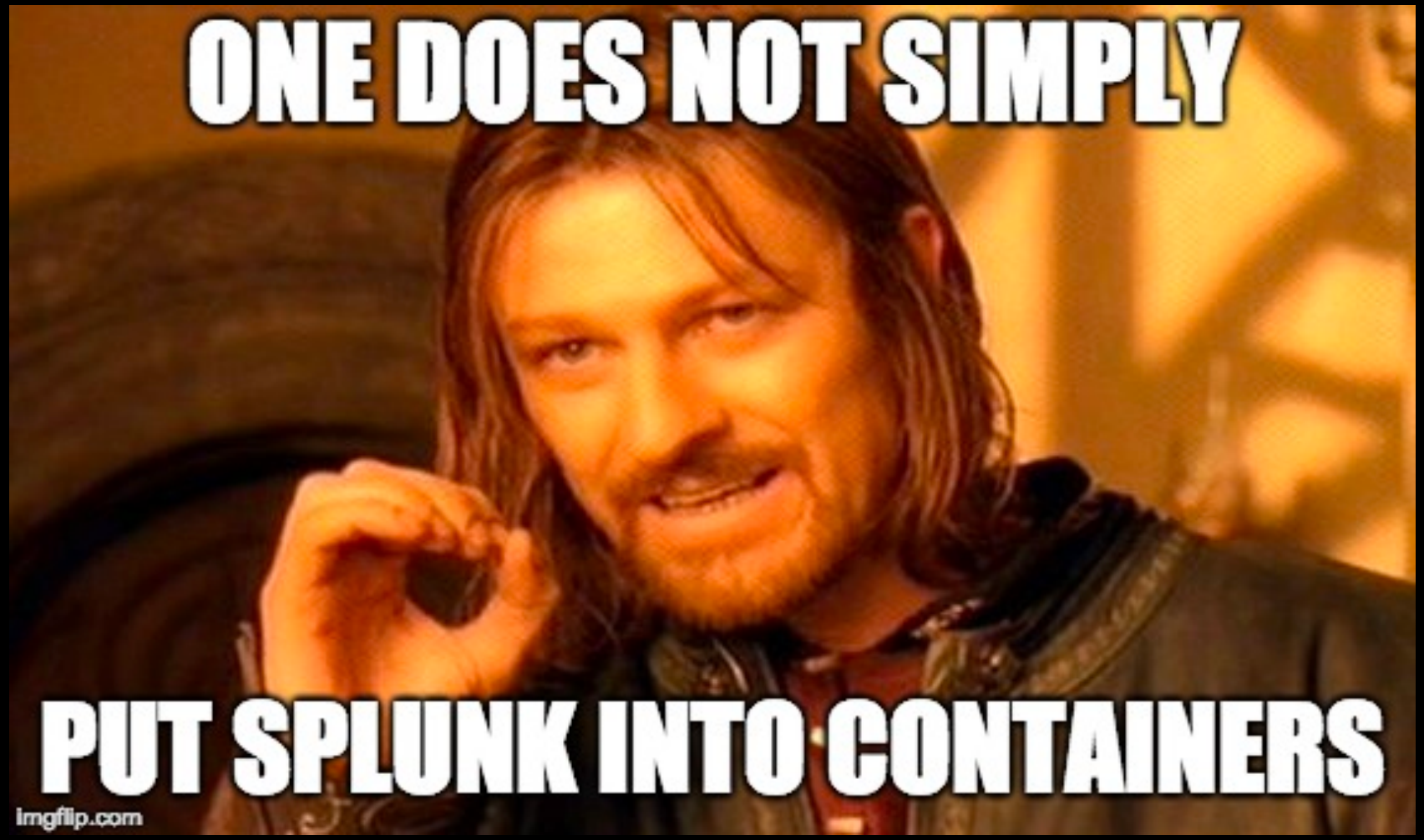
Ansible Technical Details

- Another 30 minutes of very technical details.

Data Persistence

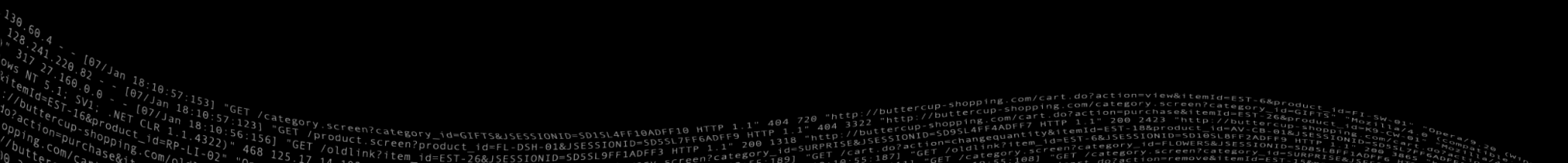
Did you want to keep that?

Data Storage



Data Storage

- Splunk is a stateful monolith
- Right now data storage is tightly coupled to the configuration
- Containers favor stateless micro-services
- We’re still evaluating the best way to persist data
- The data volumes are specifically coupled to the containers role
  - Ever had an index just disappear?
  - Ever had your search artifacts just disappear?
- Specific issues may arise only within a container



Data Storage

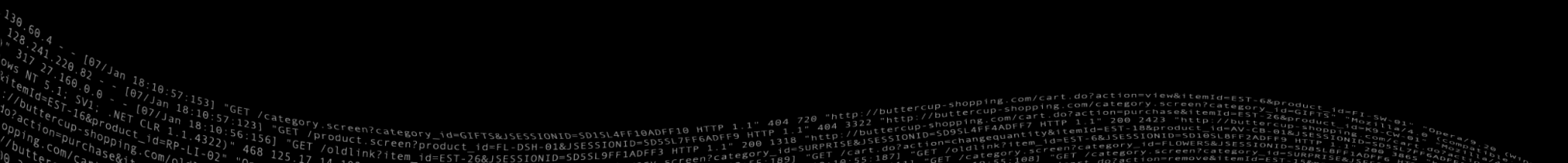
- Mount your data persistence volume to /opt/splunk/var
  - Indexes
  - Search Artifacts
  - Logs
- Docker and UCP
  - Use node affinities to restrict containers to certain nodes
  - Search heads tend to be memory bound
  - Indexers and search peers tend to be IO bound
- Look into how your orchestration layer handles volume mounts

Troubleshooting

We can’t all be perfect

Common Issues

- Getting your docker environment right is the first step
- Are all of your ports mapped and exposed properly?
  - If you're interacting with more external Splunk Enterprise elements, you will need more ports open
- Is your network configured properly?
  - Are all hosts reachable from the other hosts?
- Are all of the containers actually running?
  - OOM killers can terminate containers
  - Make sure that your hosts have adequate resources for your workloads
  - Set memory limits based on your workloads



Simple Debugging Commands

- All of our ansible plays send their information to stdout/stderr on the container
  - Use docker logs
- You should see ansible output like this

```
PLAY [localhost] ***************************************************************

TASK [Gathering Facts] *********************************************************
Wednesday 29 August 2018  09:27:06 +0000 (0:00:00.070)       0:00:00.070 ******
ok: [localhost]

TASK [include_role : splunk_upgrade] *******************************************
Wednesday 29 August 2018  09:27:08 +0000 (0:00:02.430)       0:00:02.501 ******

TASK [include_role : {{ splunk_role }}] ****************************************
Wednesday 29 August 2018  09:27:09 +0000 (0:00:00.137)       0:00:02.638 ******

TASK [splunk_common : Install Splunk] ******************************************
Wednesday 29 August 2018  09:27:09 +0000 (0:00:00.378)       0:00:03.016 ******
changed: [localhost]
```

Typical Output

- If all goes well…

```
PLAY RECAP *********************************************************************
localhost                  : ok=12   changed=6    unreachable=0    failed=0
`

Wednesday 29 August 2018  09:31:56 +0000 (0:00:01.435)       0:04:49.684 ******
===============================================================================
splunk_common : Start Splunk ------------------------------------------ 105.37s
splunk_common : Download Splunk license ------------------------------- 83.32s
splunk_common : Install Splunk ---------------------------------------- 80.31s
splunk_common : Apply Splunk license ---------------------------------- 6.31s
splunk_common : Enable the Splunk-to-Splunk port ---------------------- 6.26s
Gathering Facts ------------------------------------------------------- 2.43s
splunk_cluster_master : Set indexer discovery ------------------------- 1.44s
splunk_common : include_tasks ----------------------------------------- 1.42s
splunk_common : Generate user-seed.conf ------------------------------- 0.69s
splunk_common : Set license location ---------------------------------- 0.59s
include_role : {{ splunk_role }} -------------------------------------- 0.37s
splunk_common : include_tasks ----------------------------------------- 0.22s
splunk_cluster_master : Get indexer count ----------------------------- 0.18s
splunk_cluster_master : Get default replication factor ---------------- 0.18s
splunk_common : Set as license slave ---------------------------------- 0.17s
include_role : splunk_upgrade ----------------------------------------- 0.14s
splunk_common : Install Splunk (Windows) ------------------------------ 0.12s
splunk_cluster_master : Lower indexer search/replication factor --------- 0.10s
Stopping splunkd...
Shutting down. Please wait, as this may take a few minutes.
..
Stopping splunk helpers...
```

splunk> .conf18

Typical Output

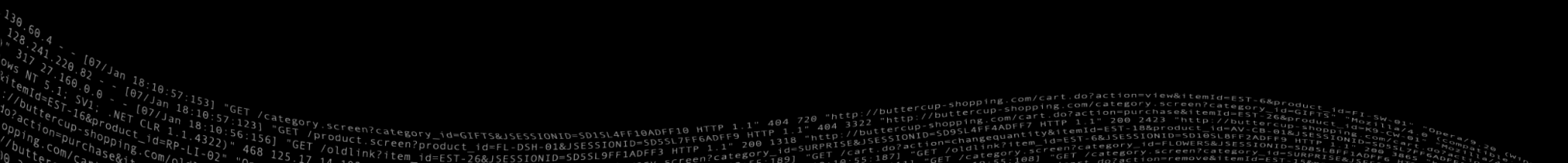
- Pay attention to failures

```
PLAY RECAP *********************************************************************
localhost                  : ok=12   changed=6    unreachable=0    failed=1
`
Wednesday 29 August 2018  09:31:56 +0000 (0:00:01.435)       0:04:49.684 ******
===============================================================================
splunk_common : Start Splunk ------------------------------------------ 105.37s
splunk_common : Download Splunk license ------------------------------- 83.32s
splunk_common : Install Splunk ---------------------------------------- 80.31s
splunk_common : Apply Splunk license ---------------------------------- 6.31s
splunk_common : Enable the Splunk-to-Splunk port ---------------------- 6.26s
Gathering Facts ------------------------------------------------------- 2.43s
splunk_cluster_master : Set indexer discovery ------------------------- 1.44s
splunk_common : include_tasks ----------------------------------------- 1.42s
splunk_common : Generate user-seed.conf ------------------------------- 0.69s
splunk_common : Set license location ---------------------------------- 0.59s
include_role : {{ splunk_role }} -------------------------------------- 0.37s
splunk_common : include_tasks ----------------------------------------- 0.22s
splunk_cluster_master : Get indexer count ----------------------------- 0.18s
splunk_cluster_master : Get default replication factor ---------------- 0.18s
splunk_common : Set as license slave ---------------------------------- 0.17s
include_role : splunk_upgrade ----------------------------------------- 0.14s
splunk_common : Install Splunk (Windows) ------------------------------ 0.12s
splunk_cluster_master : Lower indexer search/replication factor --------- 0.10s
Stopping splunkd...
Shutting down. Please wait, as this may take a few minutes.
..
Stopping splunk helpers...
```

splunk> .conf18

Shelling into the container

- Command line
  docker run -d -p 8000:8000 -e "SPLUNK_START_ARGS=--accept-license" -e "SPLUNK_PASSWORD=Chang3d!" --entrypoint=/bin/bash splunk/splunk:latest
- Our default entrypoint is /sbin/entrypoint.sh start



splunk> .conf18

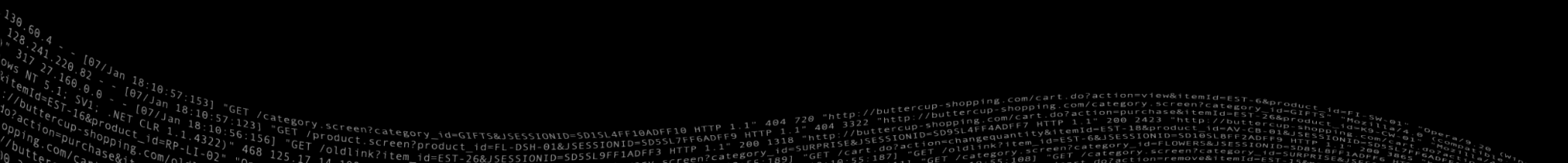
Q&A

Brian Bingham | Principal Engineer
Brent Boe | Principal Engineer

splunk> .conf18

Please leave

- It was great having you here, but our time is over



Thank You

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