RED HAT SUMMIT

BOSTON, MA
JUNE 23-26, 2015

SPLUNK ON RHGS FOR THE WIN

David Yaffe
Technical Analyst at SaskTel
Jacob Shucart
Sr. Cloud Storage Solution Architect at Red Hat

Thursday, June 25th, 2015

redhat.

WHO ARE WE AND WHO DO WE REPRESENT?

Who are we and what do we do?

David Yaffe - SaskTel

- Technical Analyst
- Splunk Expert
- Storage Expert
- ITF Tae Kwon Do Black Belt

Jacob Shucart – Red Hat

- Storage Architect
- Linux Expert
- Owns a Leather Black Belt

About SaskTel

Saskatchewan Telecommunications Holding Corporation (SaskTel) is the leading, full-service communications provider in Saskatchewan. Based in Regina, SaskTel and its wholly owned subsidiaries have a workforce of ove 4,000 full-time equivalent employees.

As a full-service provider, SaskTel offers a wide range of communications products and services, including competitive voice, data, Internet, entertainment, security monitoring, messaging, cellular, wireless data, and directory services.

In addition, SaskTel International offers software solutions and project consulting in countries around the world. Generating $1.2 billion in annual revenue, SaskTel has over 1.44 million customer connections, including 608,000 wireless accesses, 492,000 wireline network accesses, 250,000 Internet and data accesses, and 100,000 maxTV™ subscribers.

SO, WHAT ARE PEOPLE DOING FOR DATA ANALYTICS?

What solutions exist in the market?

For analytics

- The ELK Stack
- Palantir
- Sumo Logic for cloud
- Potentially Solr
- Custom scripts and point solutions

For storage

- EMC, NetApp, Hitachi, others
- Local disk

How do you choose?

For analytics

- What are you analyzing?
- Who needs it?
- Why do you need it?

For storage

- Quantity of sources
- Lifespan of the data
- Performance requirements

SPOILER ALERT – SASKTEL SOLVED THIS PROBLEM

Our Original Use Case

Background

- SaskTel launched 3G
- Many customers subscribed to unlimited plans
- Many heavy users also frequently roamed
- SaskTel had roaming agreements with partners

Use case

- Meet roaming agreement commitments through accurate network analysis

How did SaskTel choose Splunk?

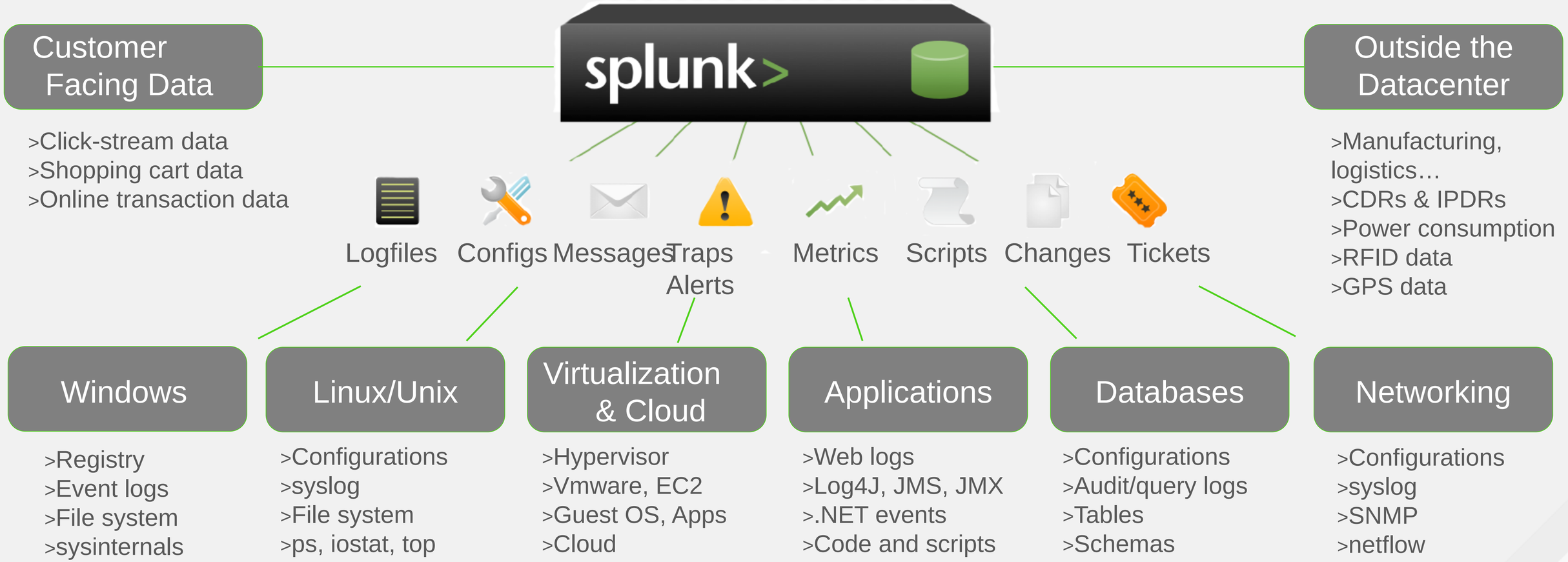
Wireless Network Support group needed a solution

- Performance analysis of CDMA and HSPA networks
- Analysis of AAA server and wireless network device logs
- Initially 100GB/day

Splunk was chosen as it was the best fit

- Existing tools could not handle load
- Wireless Support liked Splunk
- Wireless Support wanted someone else to administer it

What is Splunk? Analytics for Machine Data



It’s a big sandbox, so everybody gets to join in!

It started with Wireless Support, but…

- IS had syslog needs
- Corporate Audits and Security had infrastructure needs
- Network Support had device needs

As a result of the initial adoption, other groups started to join in

- IS adopted Splunk
- Followed by Mobility and Security
- And other groups as well

What was life like before The Solution?

In the beginning, there were silos

- Silos for jobs and for data
- Silos create access and performance issues

Many groups needed the ability to do log and data analysis

- Enterprise IT and Security
- Corporate Audits and Network Support

Most organizations outgrow traditional tools quickly.

- Standard Linux tools don’t scale
- Scripts are difficult to maintain

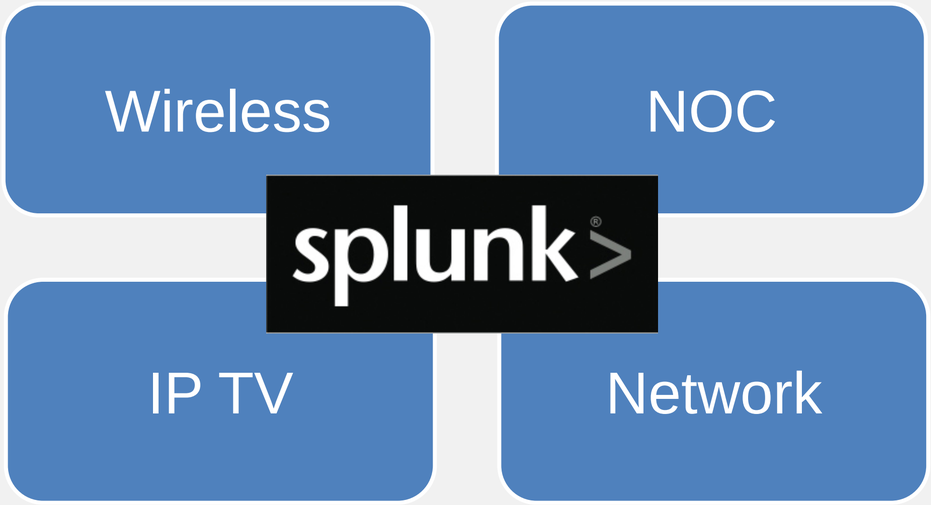
Early Adopters

Wireless Support

IPTV Head End for syslog

Network Support for core routing.

Network Operations Center for switch logs



The Second Wave of Adopters

As solution proved itself out, use proliferated

- Internal Network Support
- *nix Systems Administration
- Server Operations Center
- Corporate Security

And the more groups that used it, the more storage was needed

And, what about storage?

It started out small

- Local storage
- Then EMC, but that started becoming expensive

The search for inexpensive and scalable storage began

- OrangeFS, PvFS, GlusterFS, and others.
- Good on paper, but not well-supported

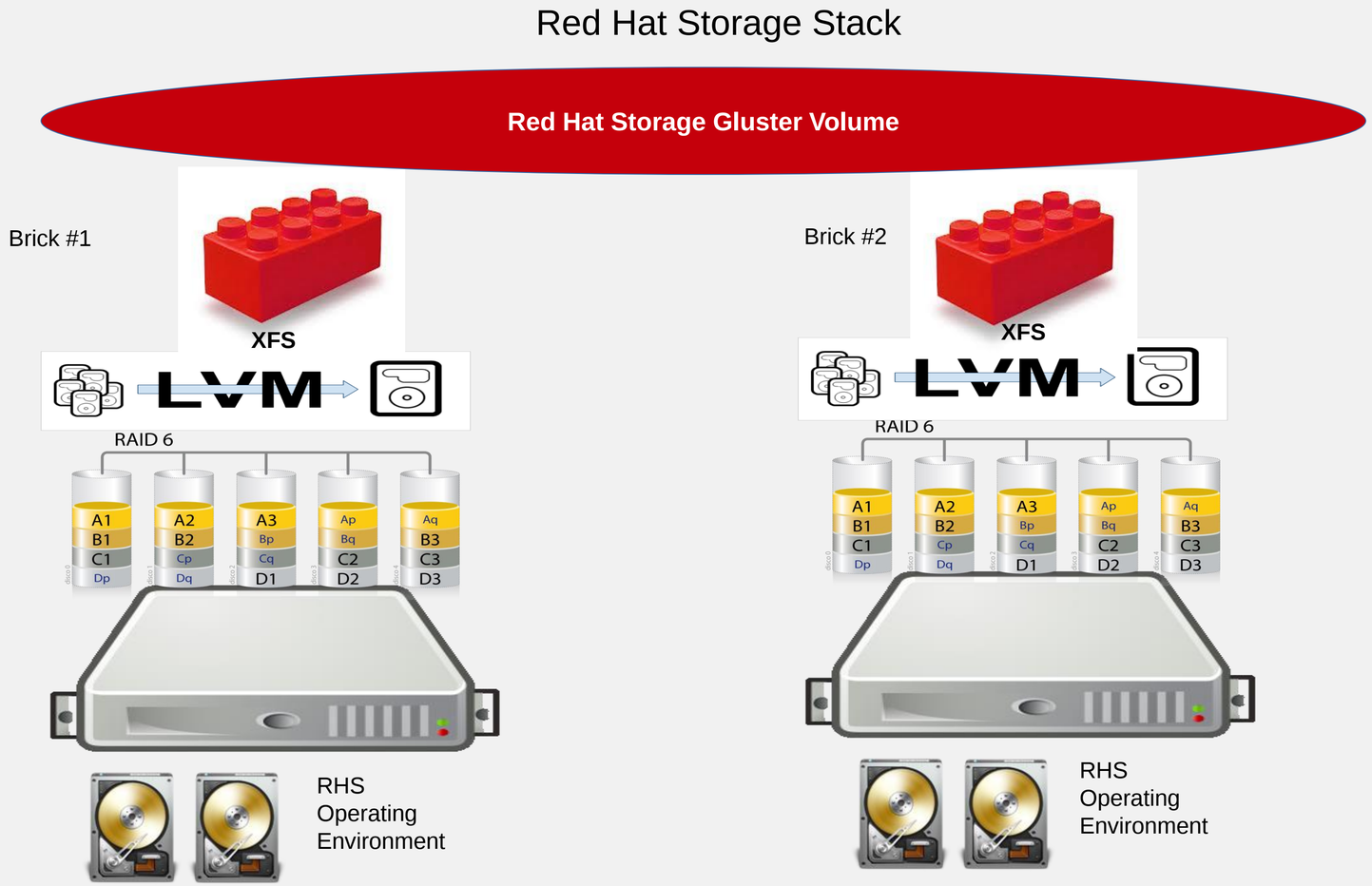
Then Red Hat announced the Gluster acquisition in 2011

What is Red Hat Gluster Storage?

Connects DAS from group of x86 servers
into a single pool of storage.

- Single, shared namespace with petabyte scalability
- File, object, and virtual block storage interfaces.
- Data protection and HA at disk, server, and site levels
- Seamlessly extensible and self-healing

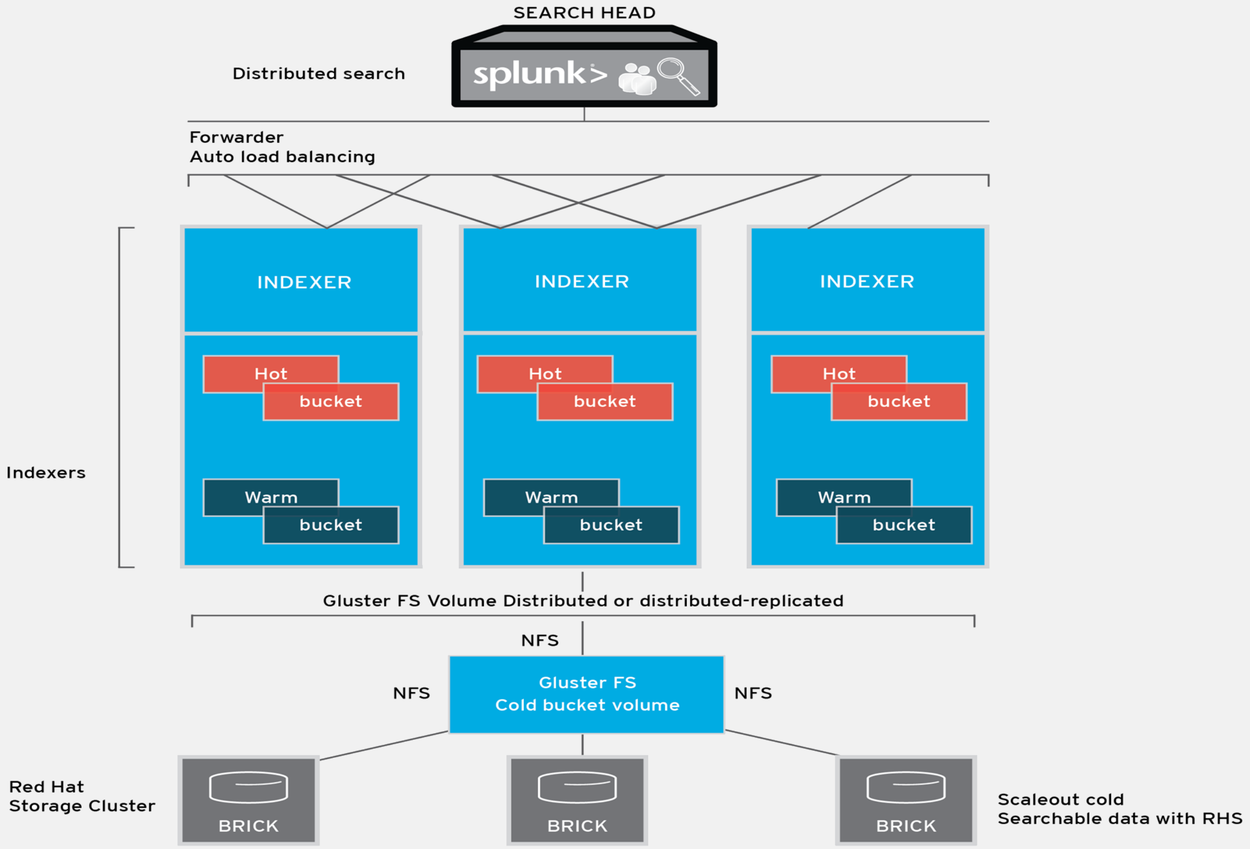
RGHS Architecture



- Software-defined storage
- Use with x86 servers
- Self-healing and managing
- Expand as needed

Less expensive to procure, deploy, maintain and support!!!

What does the combination of the two look like?



- Support for clustered and non-clustered configurations of Splunk
- Highly available storage back end
- Fast performance with data always on-line
- Grow storage as cold storage needs increase

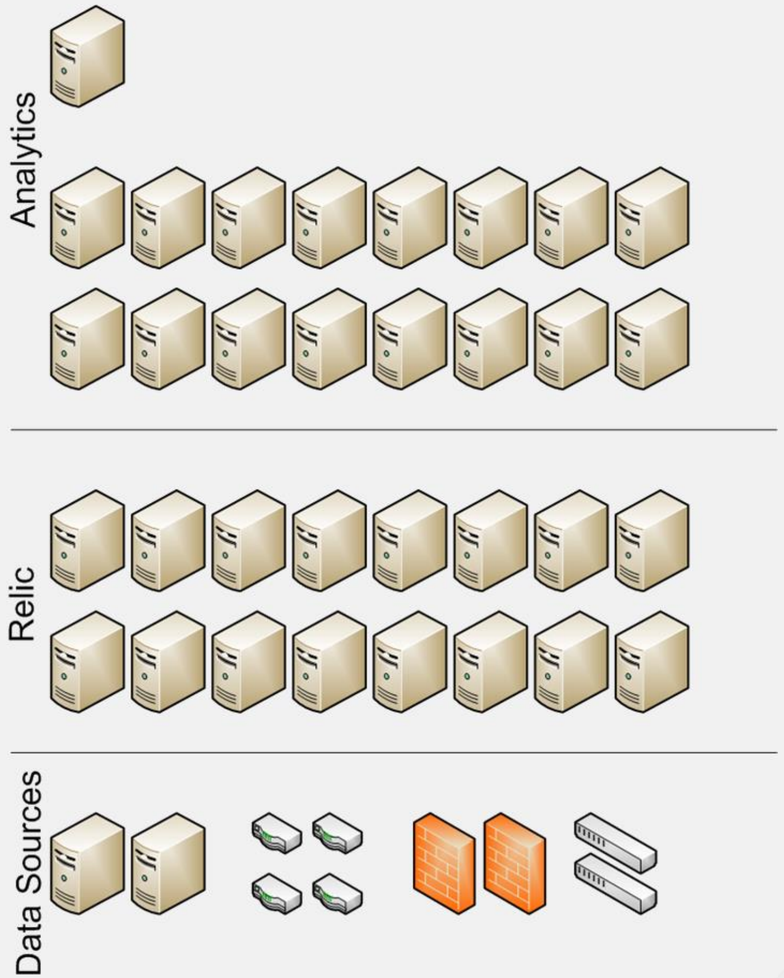
GOSH THAT SOUNDS GREAT, BUT HOW DID YOU DO IT?

SaskTel Analytics Environment

3 Tier Approach
- Data sources
- Relic
- Splunk

Splunk cold goes back to Relic

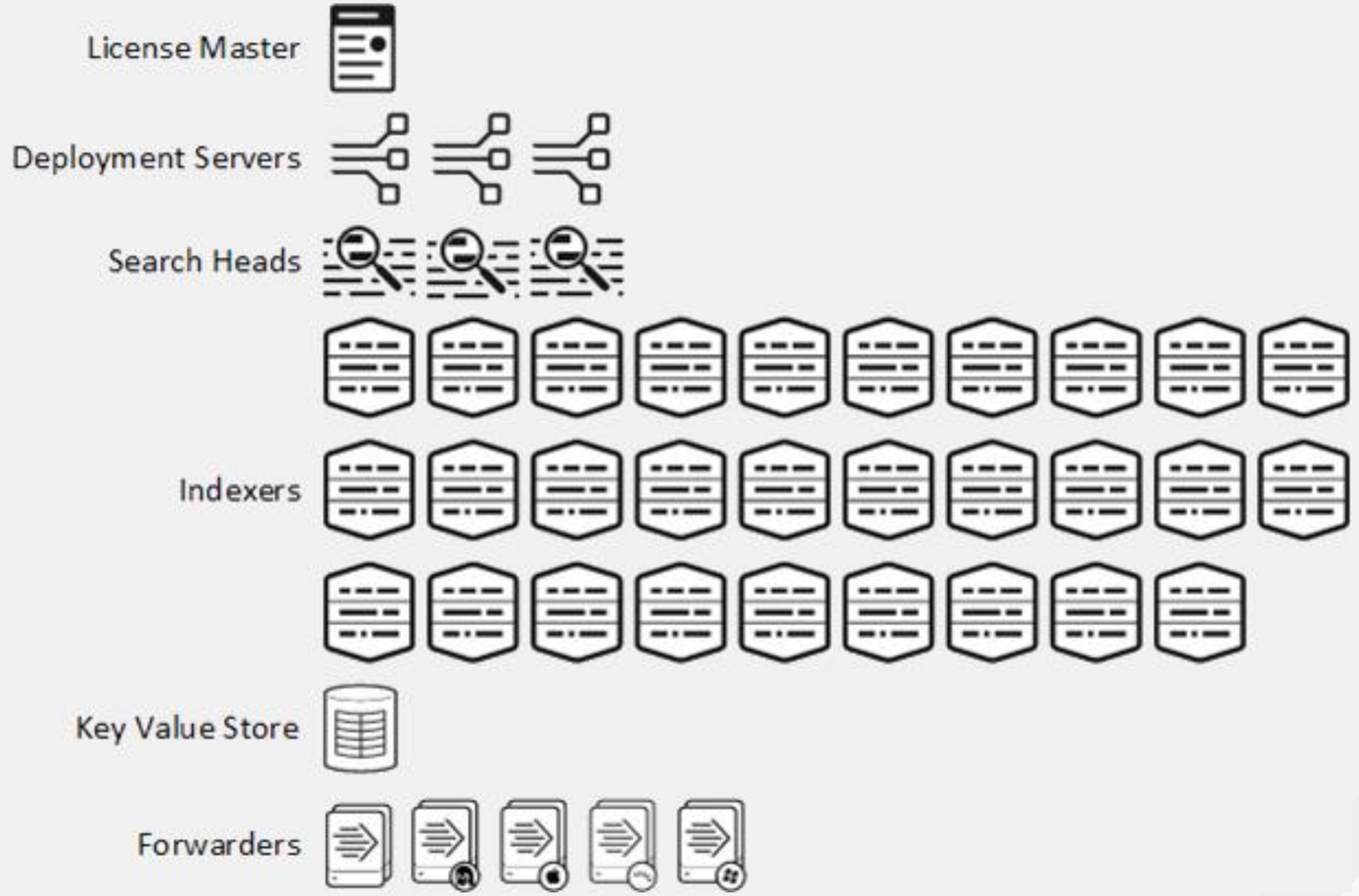
Relic is secure and provides dat access
- Red Hat Gluster Storage is a key component of Relic



SaskTel Splunk Deployment Overview

Splunk and Relic have grown

1 – License Master
3 – Deployment Server
3 – Search Heads (Clustered)
29 – Splunk Indexers
Lots – Splunk Forwarders
4 – Red Hat Gluster storage servers
2 – Web Servers



The ~~Magic~~ *Glue* that Binds Everything Together

There is no magic involved

- Rsyslog for most logs
- Splunk forwarders where needed
- Rsync(+ssh) and standard tools in some cases
- SSL where possible

Be methodical

- Define standards and stick with them!
- Integrity check everything

A Strong Foundation is Important

Understand the requirements BEFORE YOU IMPLEMENT

- All decisions should support the end goal

Identify your data sources

- What data exists, and who owns it?
- What governance and compliance rules exist?

Assess your consumers

- Who needs it?
- What needs integration?
- What else might need it?

Tips and Tricks

Plan for overcoming the gaps in your final solution

- Out of the box compatible?
- Can it handle multiple data sources or be extended?

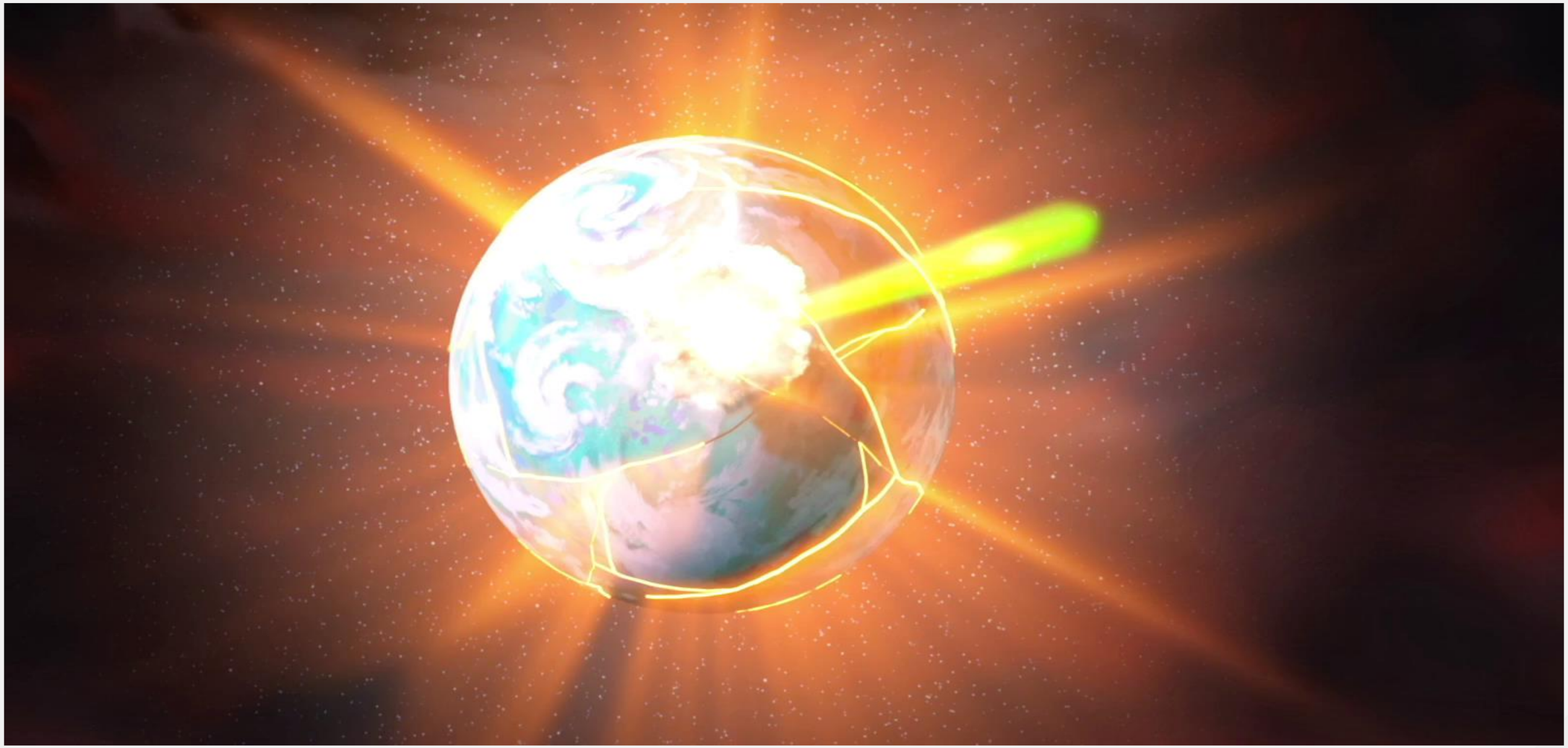
Centralize your data as a gold master

- Push, pull, poll, centralize, and prevent sprawl
- Protect and control access to data

Create and enforce standards

- Standard naming conventions for all devices in environment
- Central control of users and security AND…

Whatever you do – Don’t name your syslog server Alderaan – It will end poorly!



IN SUMMARY

Red Hat Gluster storage is a powerful tool to answer your storage requirements

- Distributed, replicated storage on commodity hardware
- Scales easily – Scale up or Scale out
- Geo-replication & self heal capabilities
- Large open source community behind the commercial product.
- To get started with Gluster visit http://www.redhat.com/storage

Splunk is a very powerful analysis platform

- Splunk Community area on the splunk.com web site is your friend
- Extensive Documentation (text and tutorial videos) for the application
- Answers website is like Stack Exchange but for Splunk
- Splunk apps allow you to easily extend Splunk to your specific use case.
- .conf is Splunk's version of Summit
- To get started with Splunk visit http://www.splunk.com/en_us/download.html

Questions and Answers?

Contact Information:

David Yaffe
david.yaffe@sasktel.com

Jacob Shucart
jshucart@redhat.com

LEARN. NETWORK.
EXPERIENCE OPEN SOURCE.