Splunk implementation

Our experiences throughout the 3 year journey

About us

- Harvard University – *University Network Services Group*
  - Serving over 2500 faculty and more than 18,000 students

- Jim Donn Management Systems
  - Architect and implement Management solutions
  - Deliver fault notifications
  - Previously with HSBC
  - 13 years in IT from NOC -> Sr. Engineer

- Tim Hartmann Systems Administrator
  - Architect and implement Authentication solutions
  - Troubleshoot various server related issues
  - Previously with another division within the University
  - 11 Years in IT from Help Desk -> Sr. Engineer

Our Interests

- Share our experiences with others
- Collaborating with like minded people
- Discuss strategies to tackle common issues
- Share solutions / code
- Endorse community activity!

Day 0

- Network and Systems team have very similar needs – centralized logging.
- Teams belong to the same department, but historically act independently.
- 2 independent Syslog-NG implementations.

- Jim and Tim break the mold and talk to each other!

Network Management Systems Drivers

- New tools must scale with the rebuild of Enterprise Network Management Systems
- Syslog needs:
  - Syslog aggregation
  - Reliable event forwarding
  - Easy to use web interface
  - Centralized log viewer
  - Correlation and alerting engine*

Systems Team Drivers

- Need to track down and resolve issues faster
- Syslog needs:
  - Centralized logging
  - Web based search viewer
  - Role based access to logs
  - Alerting
  - Reporting
  - Trend Analysis

Evaluation

- Tim leads Splunk evaluation, sets up server
  - Simple installation
- Tim and Jim point Syslog-NG envs at Splunk
- Develop User Roles strategies
  - Net Eng, NOC, Security, and Server teams
- Develop data separation strategies (KISS)
  - Host names
  - Sourcetypes
  - Indexes

Installation stats

- 400 Linux, Solaris, and Windows servers
- 700 Switches and Routers
- 2300 Wireless Access Points
- TACACS+ authentication logs
- VPN access logs
- DNS and DHCP logs
- 50 registered Splunk users, half are regular users

Phase 1 Hardware and Strategies

What it runs on

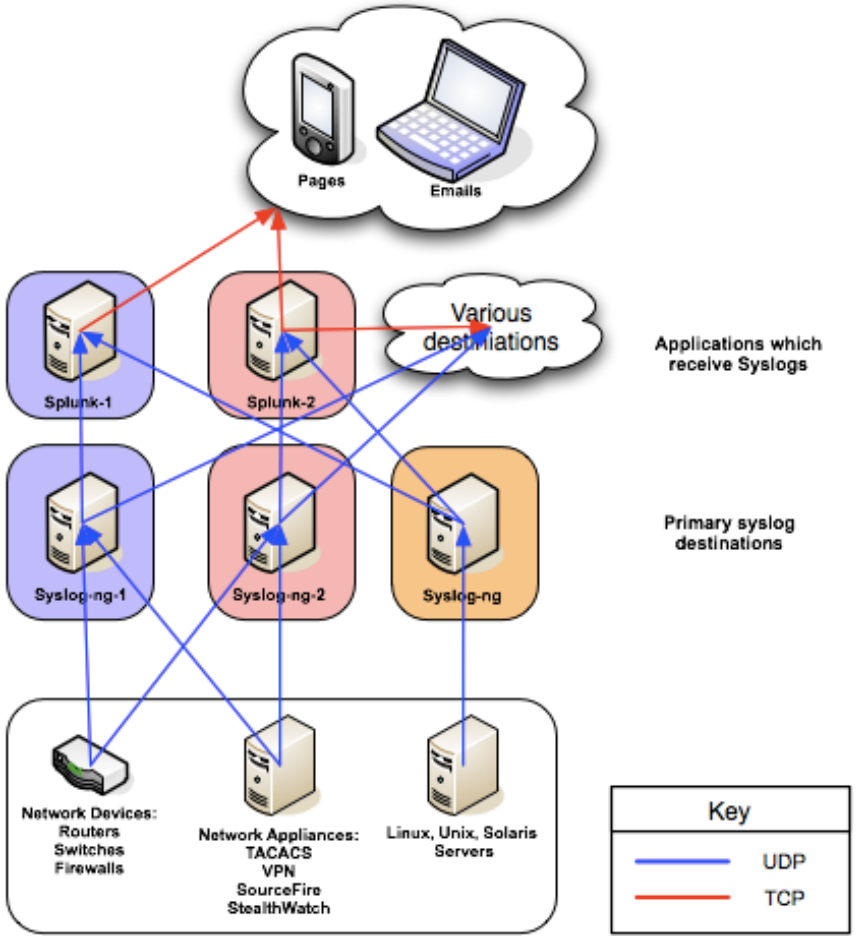
- RHEL 5 – 64 bit
- Commodity HW
- 15k local disk
  - RAID 5 1.6T
- 2 x 4 Core Processors (3.00 GHz)
- 16 GB RAM
- Custom Yum Repo for software Deployment

Strategies

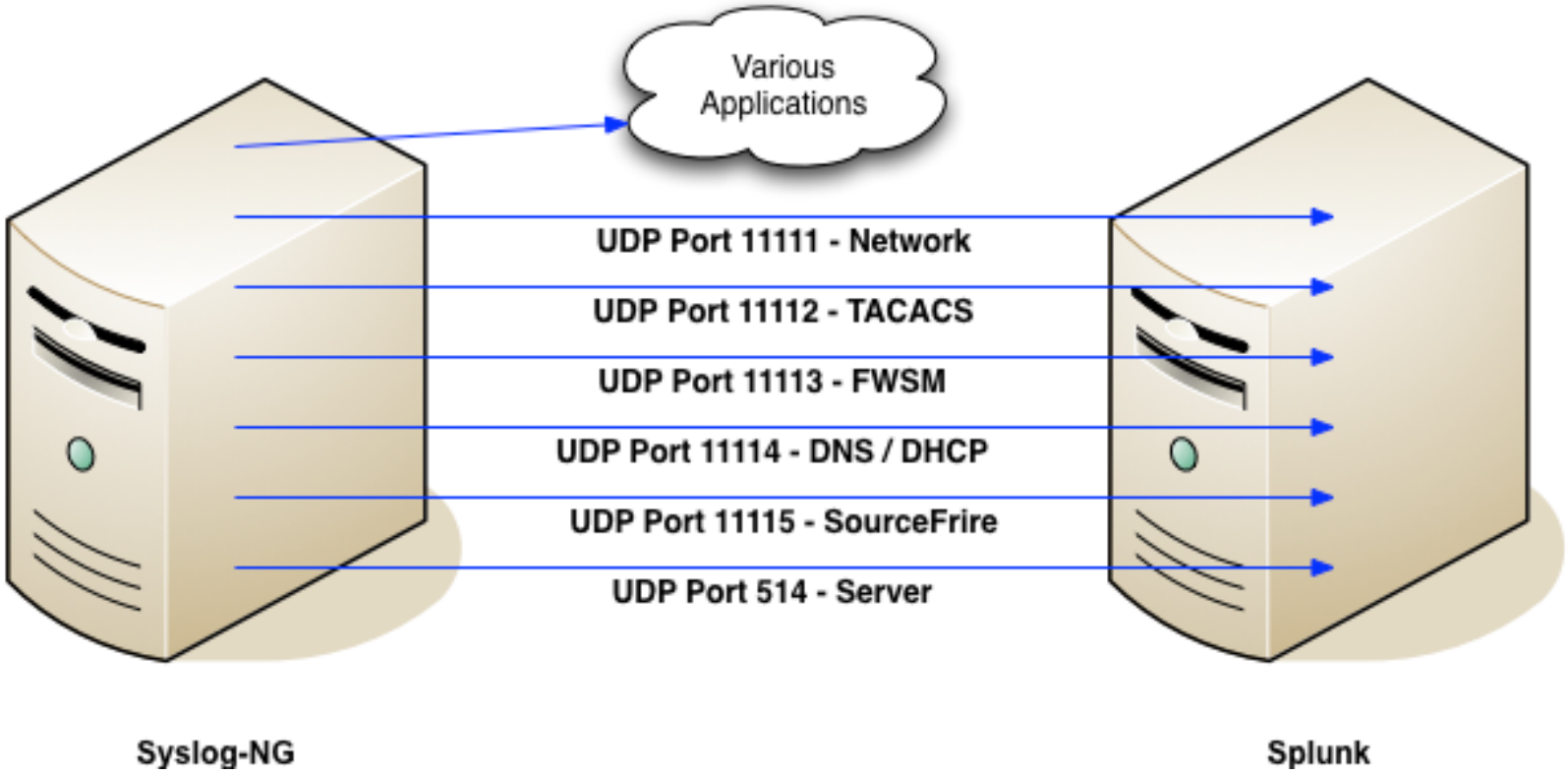
- Two of everything
- Fast disk
- Wherever possible we made our configurations independent of other services (SAN/NAS)
- Simplicity keeps it maintainable

Phase 1 – Basic syslog, “just get it in”

- Very few agents
- All UDP
- Sourcetype based roles
- Dual role servers (search & index)
- Hot / Hot HA architecture
- 1.6 Terabytes of useable disk each
- Splunk v 3.x

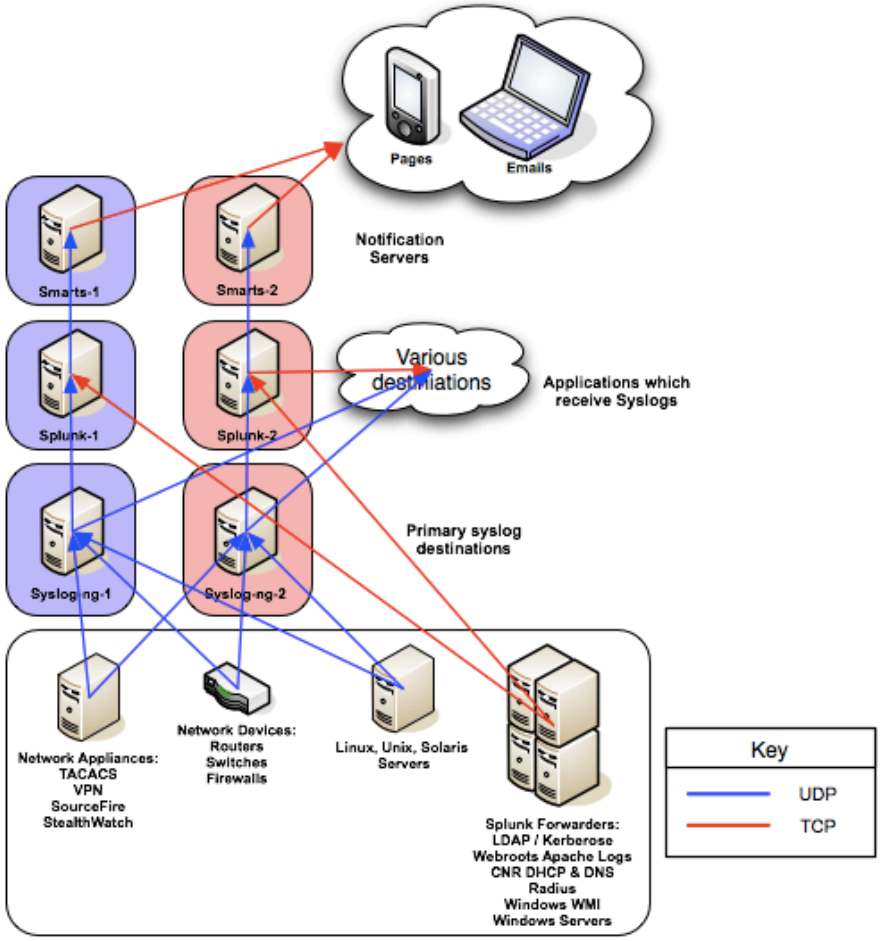


Closer look at Syslog-NG



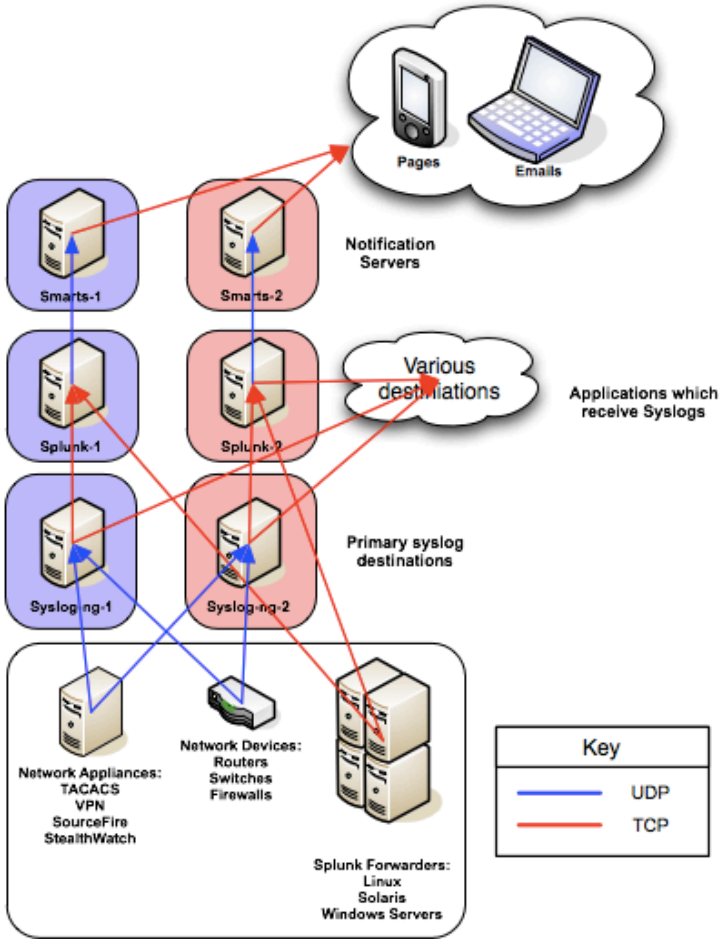
Phase 2 – More logs!

- Merge Syslog-NG servers
- Start to introduce more Splunk agents to grab difficult logs
- Add more departments
- Splunk integrated with event notification path
  - Replaces syslog adapter in EMC Smarts
- Splunk v 3.x

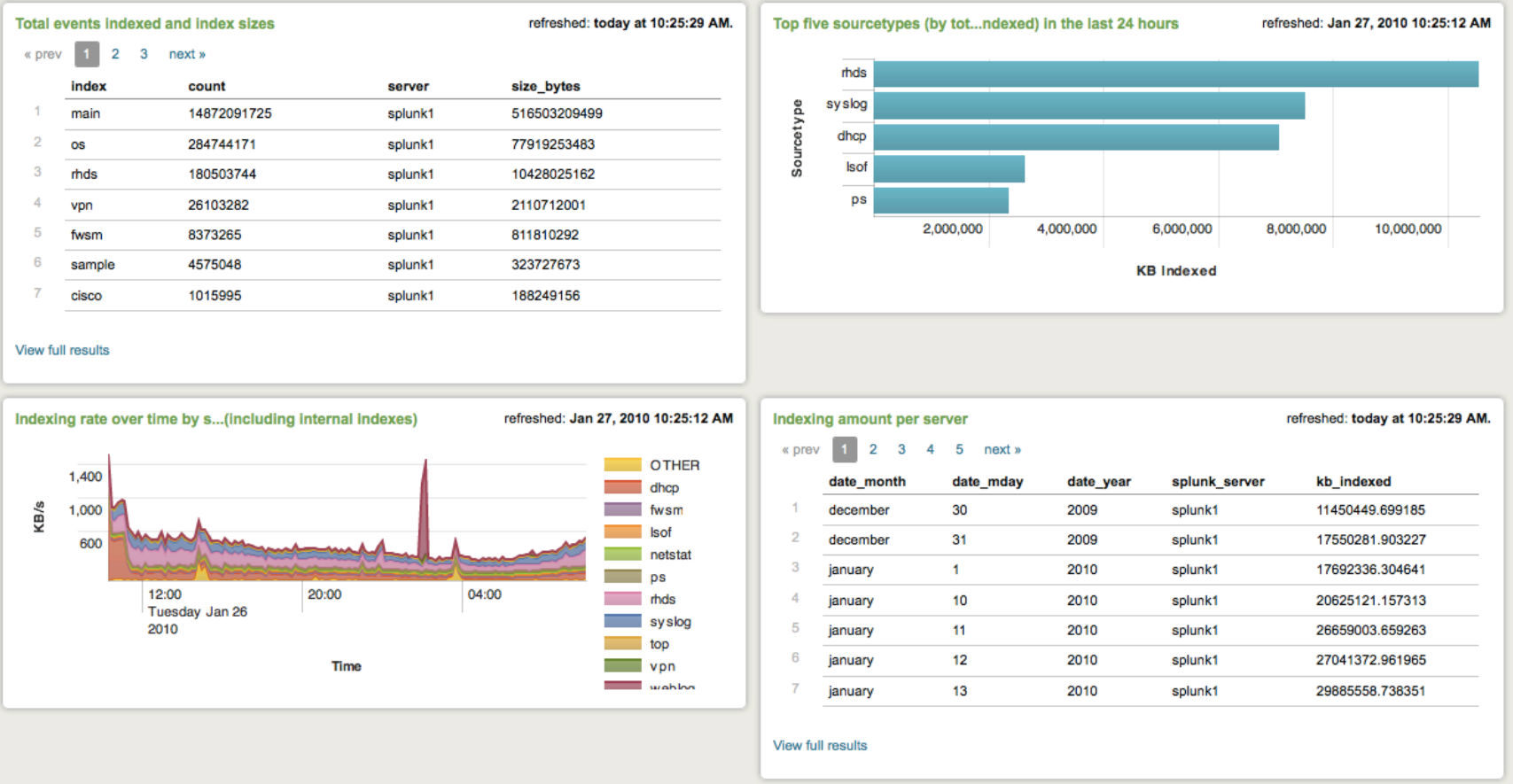


Phase 3 – Agents and Indexes

- More and more Splunk agents
  - Windows servers migrated
- TCP forwarding of syslogs
- Multiple indexes
  - Index based roles
  - Faster searches
- Replace Smarts DB with Splunk
  - Hardware is now available for Splunk expansion
- Splunk begins to fill monitoring gaps, acts as “glue”
- Splunk v 4.x
  - Apps now available
  - Free Unix & Windows Apps
  - First round of developing our own



Snapshot after implementing more indexes



Total events indexed and index sizes — refreshed: today at 10:25:29 AM.

« prev 1 2 3 next »

	index	count	server	size_bytes
1	main	14872091725	splunk1	516503209499
2	os	284744171	splunk1	77919253483
3	rhds	180503744	splunk1	10428025162
4	vpn	26103282	splunk1	2110712001
5	fwsm	8373265	splunk1	811810292
6	sample	4575048	splunk1	323727673
7	cisco	1015995	splunk1	188249156

View full results

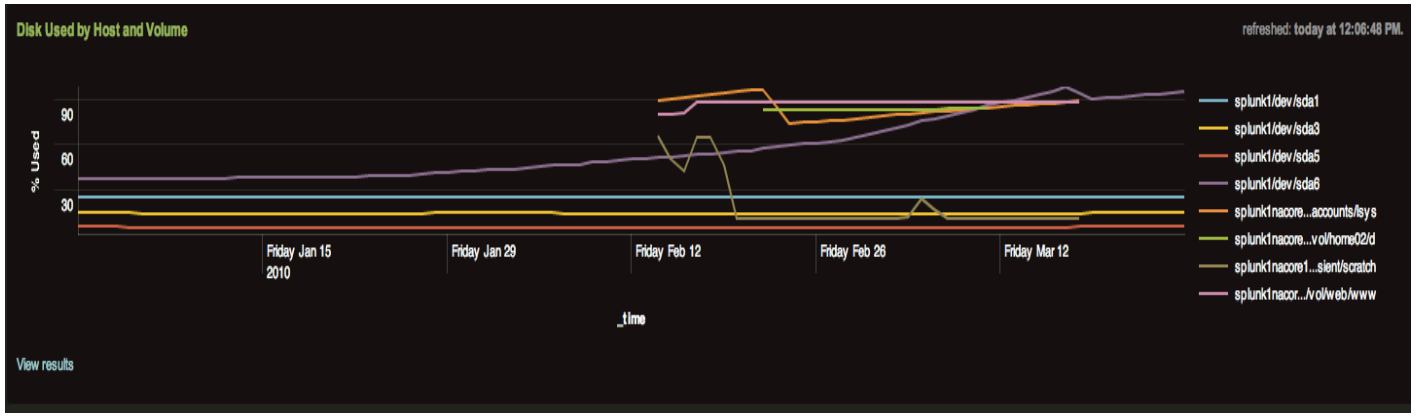
Indexing amount per server — refreshed: today at 10:25:29 AM.

« prev 1 2 3 4 5 next »

	date_month	date_mday	date_year	splunk_server	kb_indexed
1	december	30	2009	splunk1	11450449.699185
2	december	31	2009	splunk1	17550281.903227
3	january	1	2010	splunk1	17692336.304641
4	january	10	2010	splunk1	20625121.157313
5	january	11	2010	splunk1	26659003.659263
6	january	12	2010	splunk1	27041372.961965
7	january	13	2010	splunk1	29885558.738351

View full results

Splunk growth around the same time



- Organic growth with other departments
- Steady growth of indexed data
  - Introduction of new indexes
- Security mandate to have Splunk on all servers

Phase 4 Hardware and Strategies

What new Indexers runs on

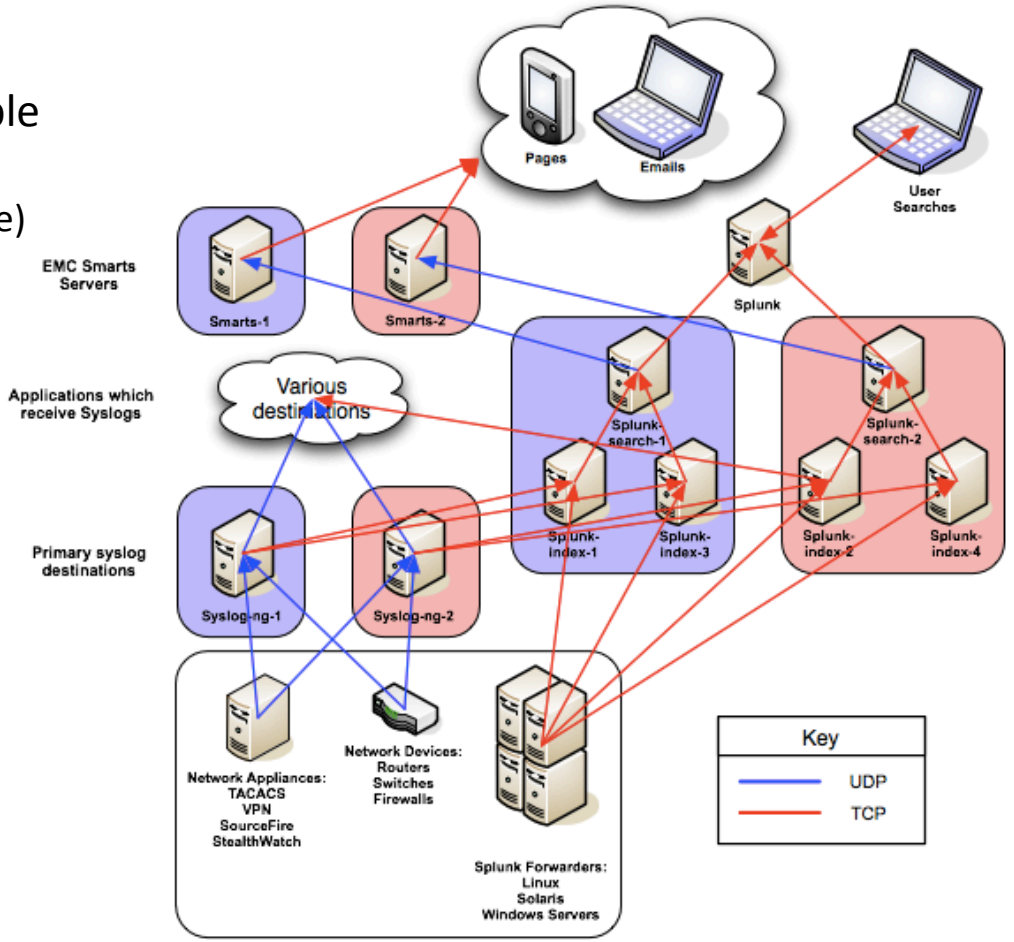
- RHEL 5 – 64 bit
- Commodity HW
- 15k Direct Attached Array
  - RAID 5 1 TB
  - Room for more drives
- 2 x 4 Core Processors (3.00 GHz)
- 12 GB RAM
- Custom Yum Repo for software Deployment

Strategies

- Horizontal expansion
  - Search Heads
- Two of everything
  - Keep the hardware specs close as possible
- Fast disk
  - Use of Linux LVM to grow additional disk
- Wherever possible we made our configurations independent of other services (SAN/NAS)
- Simplicity keeps it maintainable

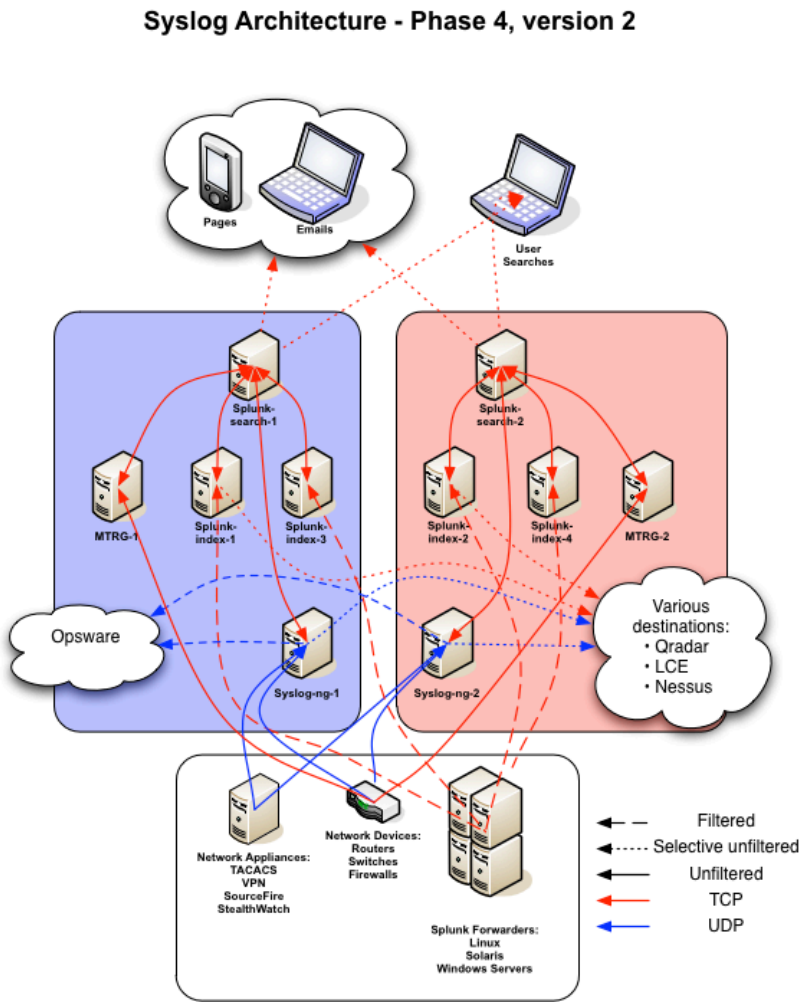
Phase 4 – Apps and Security

- Migrate unified alerting
- Remove UDP everywhere possible
- New Splunk Architecture!
  - Horizontal expansion (map reduce)
  - Search Heads
  - Scheduled search server
  - Automated sync
  - More disk!
  - Load balanced VIP?
- Agents, agents, agents
  - Support for apps
  - Custom inputs
  - Scripted output
- Splunk Agent on Syslog-NG
- Deployment Server



Phase 4, v. 2 - Apps

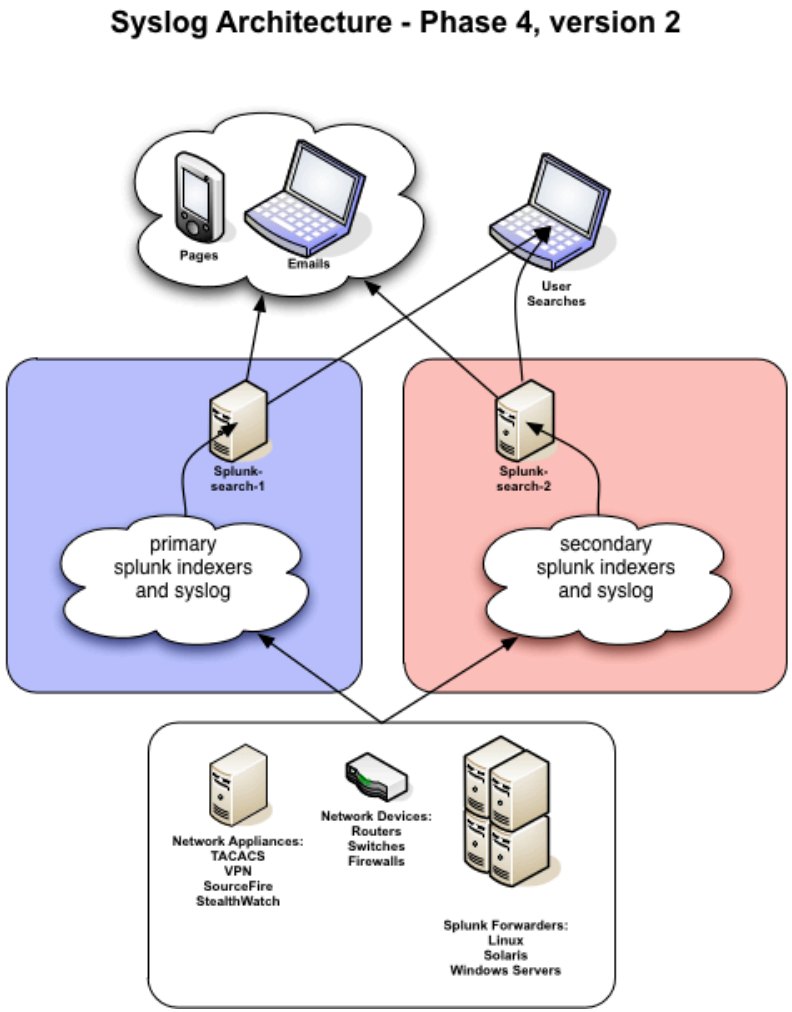
- Same as v. 1 but…
- Collapse Apps into Splunk infrastructure:
  - MRTG?
  - Syslog-NG?
  - Splunk-data-gatherer hybrid?
- Deployment Server:
  - Use Puppet
  - Use SVN



From a users perspective

Search heads have access to all indexers:

Two of everything for automatic redundancy



Home Brewed Splunk Apps / Usage

- Xen server status
- Replace legacy monitoring scripts
- Transaction based alerts for Linux and Windows
- Scripted inputs provide visibility into Network device port status (CLI only data)

Future Apps

- Security App?

- Manager of Managers
  - Add Net-SNMP trap receiver
  - Migrate most MRTG graphs (Non-RRD)
  - Replace Cacti (RRD)
  - Trend all EMC Smarts / snmpoll data

Additional info

Contact info

james_donn@harvard.edu

tim_hartman@harvard.edu

Community

http://answers.splunk.com

https://listserv.uconn.edu/cgi-bin/wa?A0=SPLUNK-L