Go Big or Go Home

Sean Delaney – Specialist SE
Mustafa Ahamed – Director, Product Management

splunk>

Agenda

- 3 Tier Approach
- Design the Forwarding Tier
- Design the Indexing Tier
- Design the Search Tier
- Best Practices in Scaling

3 Tier Approach

splunk>

Pop Quiz

Designing a scalable environment is:

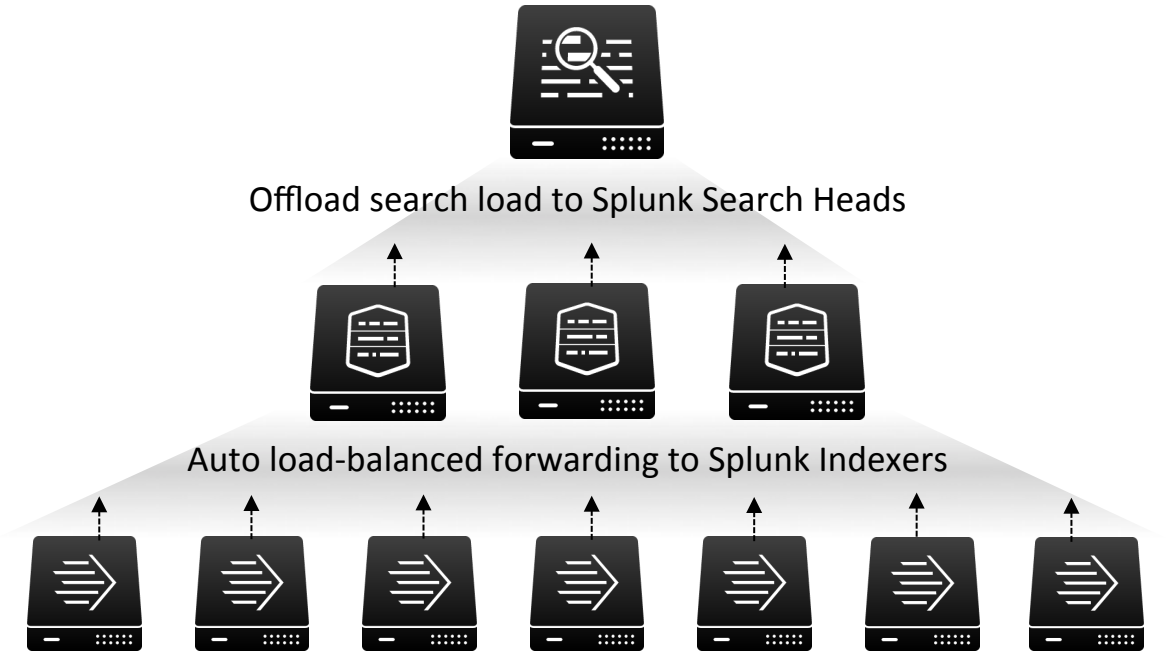
- ❑ Easy (If planned well in advance)
- ❑ Hard (With no planning)

Sizing Considerations

- Vital Info
  - Amount of incoming data
  - Amount of indexed (stored) data
  - Number of concurrent users
  - Number of saved searches
  - Types of searches
  - Specific Splunk apps
- http://docs.splunk.com/Documentation/Splunk/latest/Installation/Performancechecklist

Splunk Enterprise 6.3

Enterprise-class Scale, Resilience and Interoperability



Send data from thousands of servers using any combination of Splunk forwarders

Forwarding Architecture

splunk>

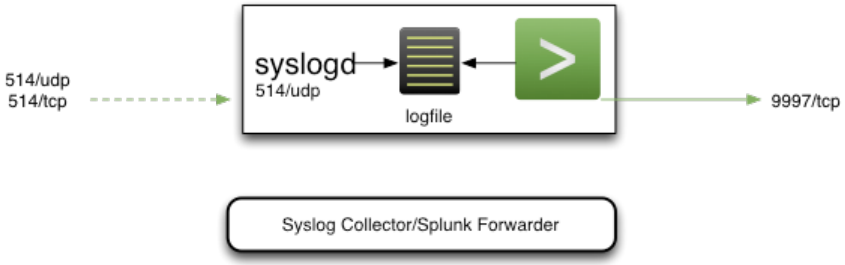
Forwarding Tier

Design Factors

- Syslog Collectors (HA)
- DBConnect Inputs
  - McAfee EPO data
- TA Inputs
  - CheckPoint
- Assorted Inputs
  - Microsoft AD logs
  - MicroSoft Exchange Server
  - Microsoft Sharepoint logs
  - Log4j, Linux, IIS

Syslog Collectors

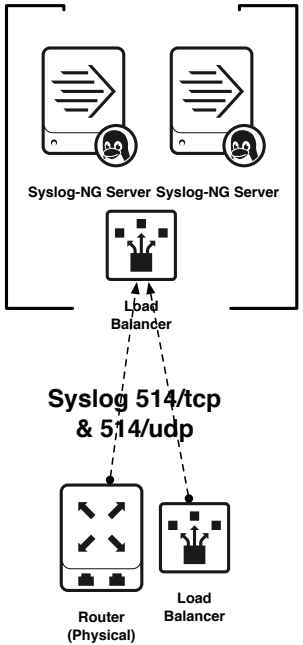
- Best practice to use dedicated syslog servers
- Syslog-NG/rSyslog recommended
- Syslog can write events to dedicated log files allowing for easy sourcetype classification on inputs



Syslog Collectors

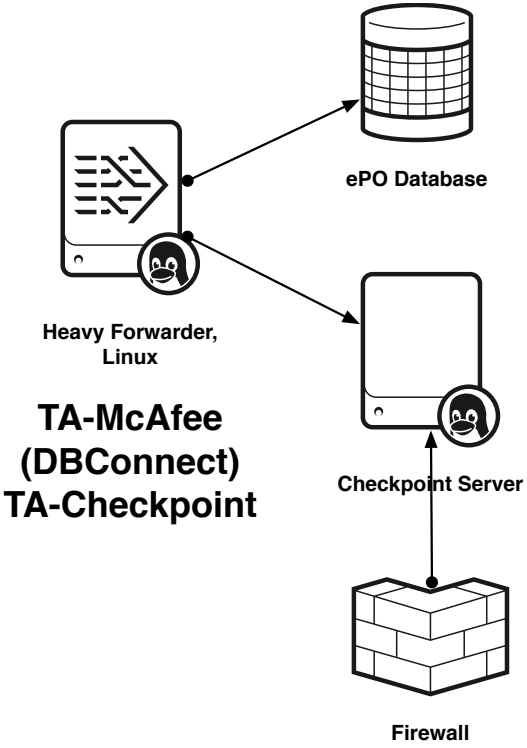
High Availability

- Using a Load Balancer/VIP with Linux Heartbeat to provide failover for the syslog listener
- Syslog-NG PE Client-side failover



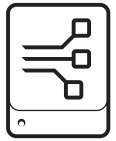
Forwarder for TA's

- TA-McAfee requires DBConnect to pull endpoint events
- TA-Checkpoint uses the LEA Client to retrieve Firewall log events
- Not a HA design, but could be hosted on a VM to standby or failover



Deployment Server

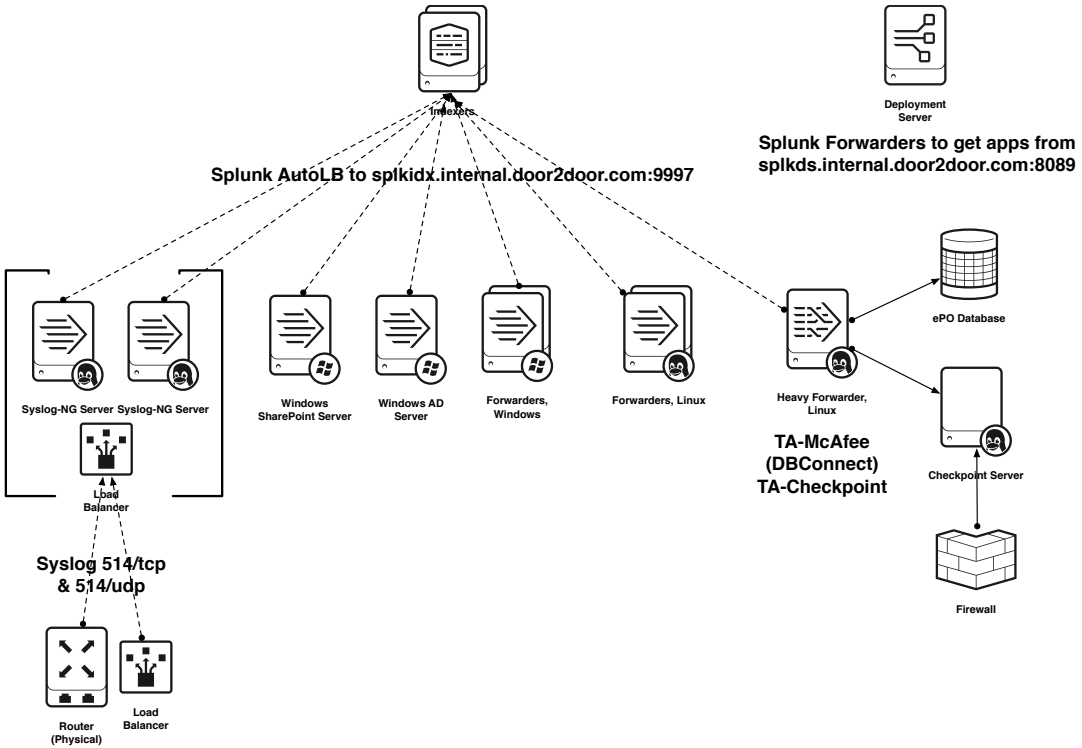
- Deployment Server to manage Linux and Windows forwarders
- Not a HA design, but could be hosted on a VM to standby or failover



Deployment Server

Splunk Forwarders to get apps from splkds.internal.door2door.com:8089

Forwarding Tier



Forwarding Tier Design Best Practices

- Use a Syslog Server for Syslog data
- Be careful with Intermediate forwarders
  - They can introduce bottlenecks
  - Reduce the distribution of events across Indexers
- AutoLB will spread over all available indexers, but don't assume evenly!
  - Enable forceTimebasedAutoLB
- May need to increase UF thruput setting for high velocity sources
  - [thruput]
  - maxKBps

Indexing Architecture

splunk>

Indexing Tier

Design Factors

- 1 Tb/day (1000Gb/day) peak ingest
- High Availability – Indexer Replication (RF=3/SF=2)
- 10% Disk Space Contingency
- 90 days minimum data retention
- Cluster Sizing Calculator
- http://splunk-sizing.appspot.com

Storage Calculations

- RAID Configuration
  - Amount of raw disk
  - Fault tolerance
  - Available IOPS
- Filesystem Overhead
  - inodes consume space
- Wiggle room
  - Additional replicated buckets when a node fails
  - Unbalanced replicated buckets
- Splunk internal logs, Summary Indexes, Report Acceleration, Accelerated Data Models

Indexer IOPS

- 800+

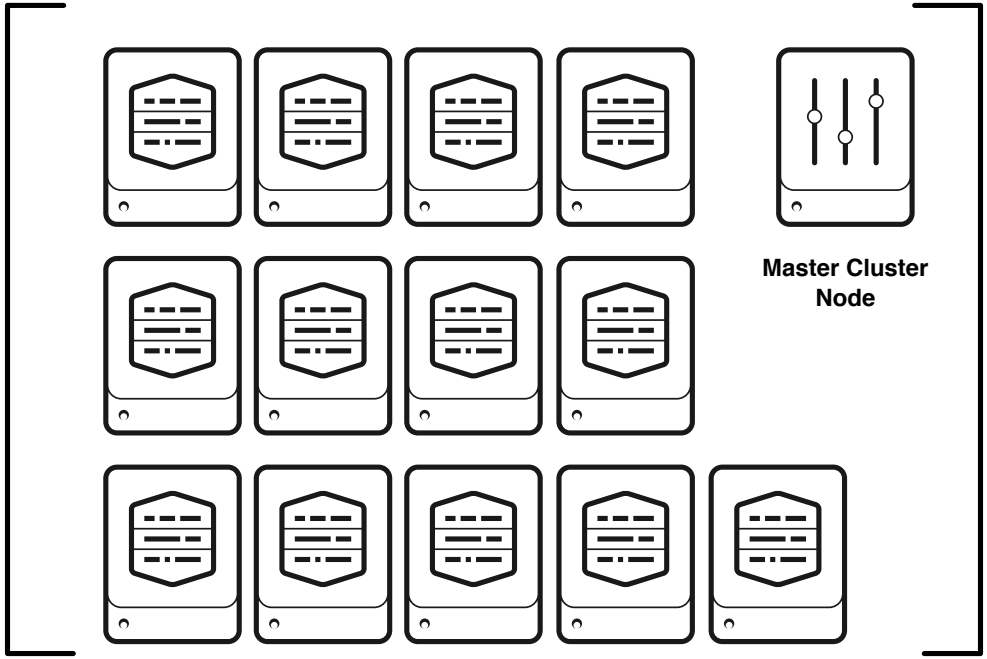
Storage Types

- Local vs Direct Attached vs SAN vs NAS
- SSD/Flash vs Spinning Disk
  - SSDs offer much higher IOPS with no latency
  - Significant performance increases with Sparse Searches

Cluster Master Server

- Indexer Apps are deployed via CM
- Not a HA design, but could be hosted on a VM to standby or failover

Indexing Tier



Indexing Tier Design Best Practices

- Depending on Searchload 150 – 300 Gb max/idx/day***
- Max # of Indexes (indices) when clustering is enabled

How Clustering Affects Sizing

- Increased storage:
  - 15% of raw usage for every replica copy
  - 35% MORE to make that searchable
- Increased processing
  - Incoming data to indexer is streamed to indexing peers to satisfy required number of copies
- More hosts
  - Need “replication factor” + 2 (search head, cluster master)

Benefits of Clustering

- Data redundancy
- Data availability
- Indexer resiliency
- Simpler management of indexers
- Simpler setup of distributed search
- Multi-site clustering allows site-specific search to reduce WAN traffic

Search Architecture

splunk>

Search Tier

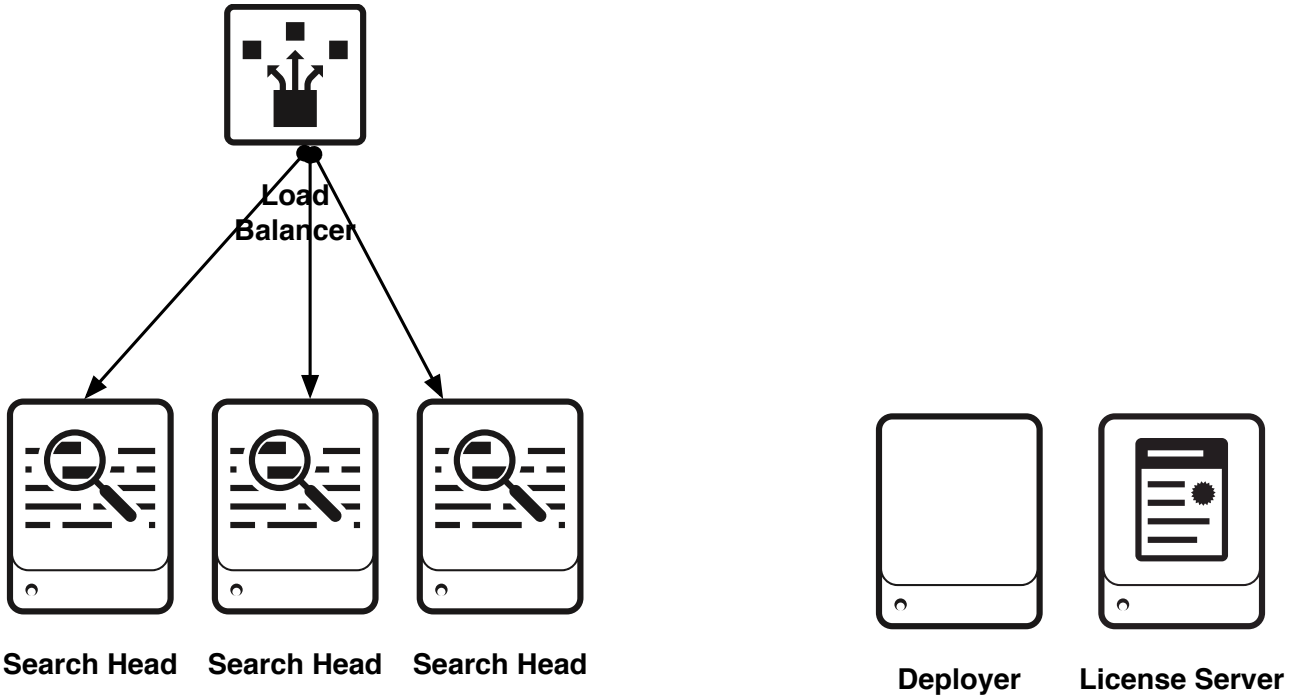
Design Factors

- High Availability
- Search Head Clustering
- # users
- # concurrent searches
- Forward all data to indexers

SHC & Deployer

- Search Head Cluster Apps need to be installed by the Deployer
- A minimum of 3 Search Heads are required for a SHC
- No exchange, no dbx with SHC

Search Tier


Load Balancer
Search Head
Search Head
Search Head
Deployer
License Server

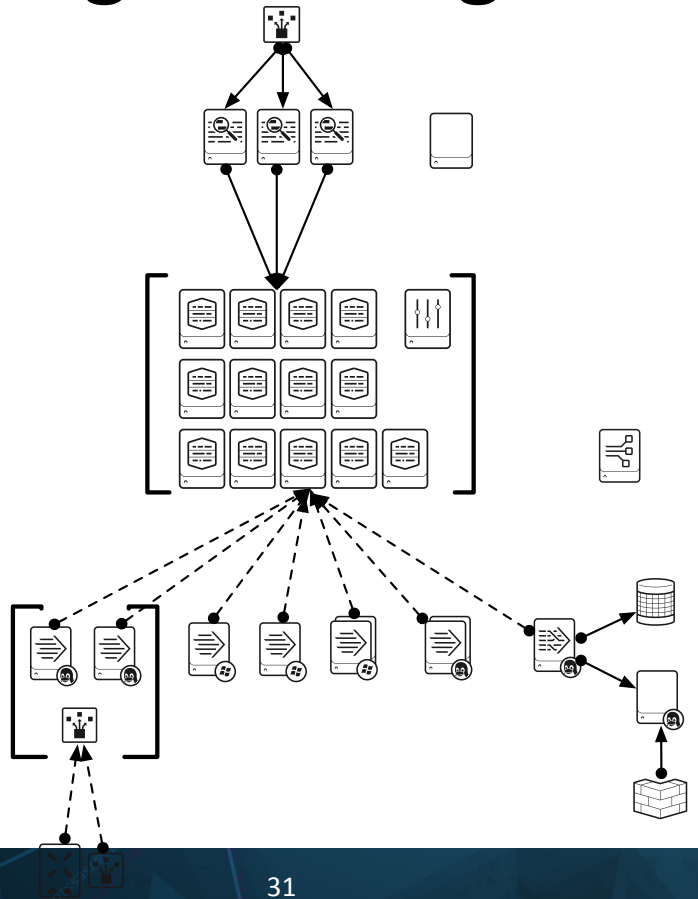
Search Tier Design Best Practices

- ES will still require a separate Search Head or dedicated SHC
- Use LDAP/AD/SSO for user Authentication
- Load Balancer configured for sticky sessions

Final Design

splunk>

Putting It All Together



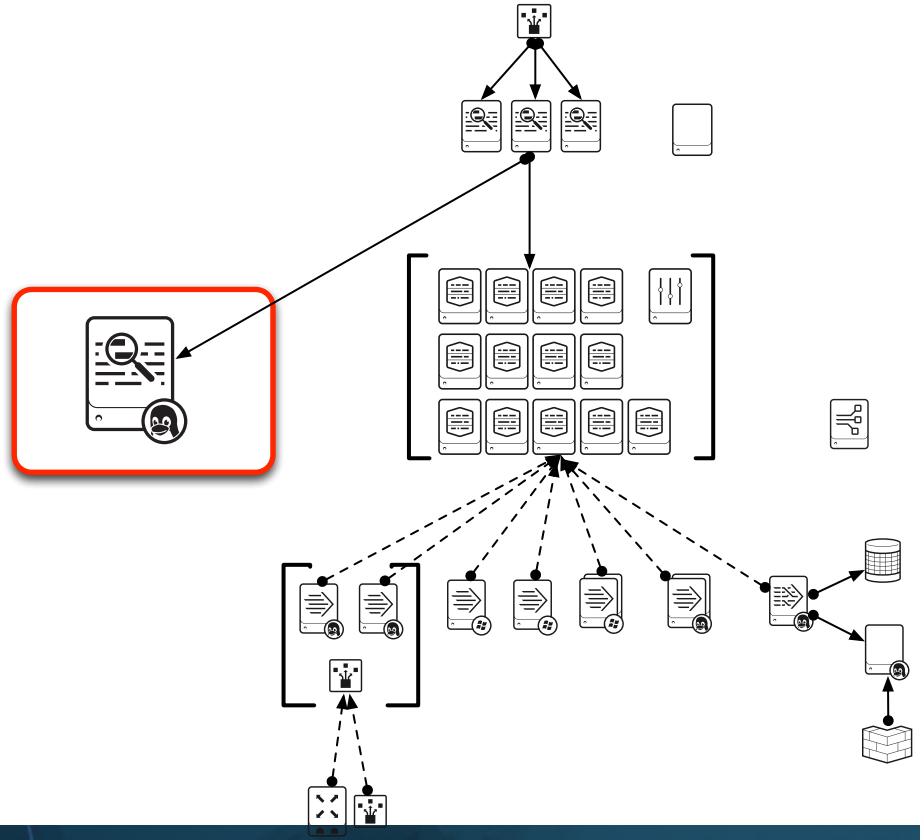
.conf2015

Migration

splunk>

Hybrid Approach

- Add the existing Splunk instance as a search peer until the data retention period has expired
- Disable scheduled searches on the old instance
- Migrate any Summary Index data to new Indexers



Review

splunk>

Top 5 Things To Consider

- Indexer Storage requirements – Size and IOPS
- Minimum buy-in for a SHC is 3
- Use VMs for CM/LS/DS/Deployer if possible
- Consider a dedicated SH for a Distributed Management Console
- When in doubt – add another Indexer

Required Reading

- Distributed Deployment Manual
  - http://docs.splunk.com/Documentation/Splunk/latest/Deploy/Distributedoverview
- Highlights
  - Reference hardware specs
  - How searches affect performance
    - Dense / Rare / Sparse
  - App considerations
  - Summary table

Best Practices

splunk>