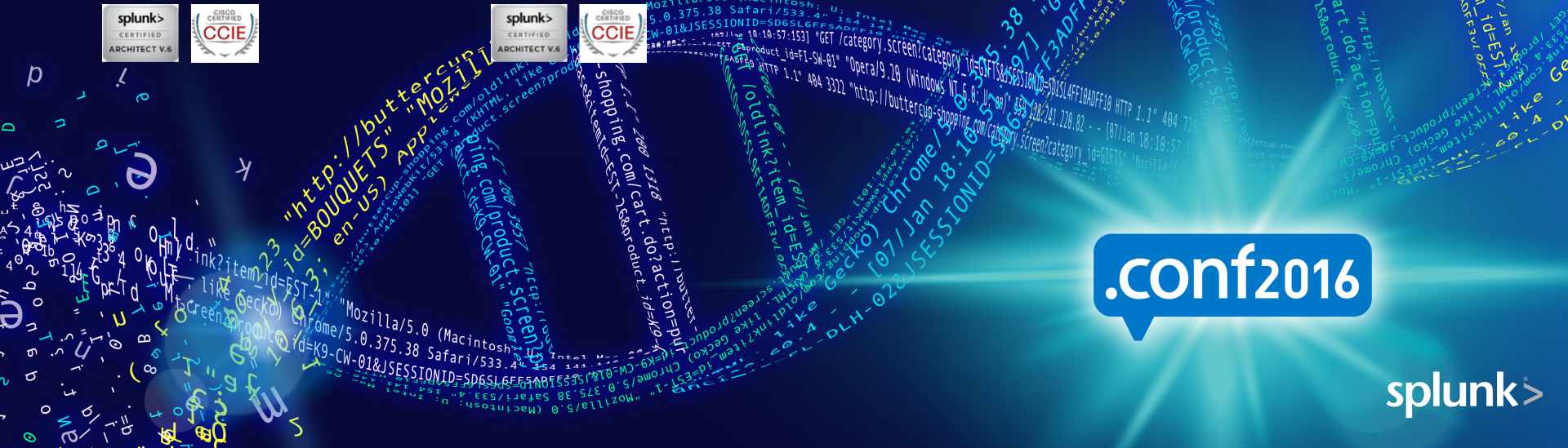
Infrastructure Analytics: Driving Outcomes through Practical Uses and Applied Data Science at Cisco

Matt Birkner
Distinguished Engineer, Cisco

Ian Hasund
Chief Architect, Cisco

Robert Novak
Big Data Partner CSE, Cisco



Disclaimer

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 the 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.

Agenda

- Introductions
- Our View Of Infrastructure Analytics
- Data Sources We Collect & Typical Problems We Care About
- Splunk Apps We Have Built
- Examples Of Use Cases
- Demo & Summary
- Cisco and Splunk Partnership

Applied Data Science At Cisco

2016: Inflection Point In The Network

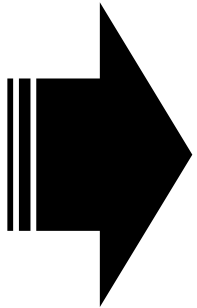
- Analytics Platforms, Capacity Planning and Traffic modeling have matured
- Telemetry is key focus area, streaming data, data at rest, structured, unstructured
- SDN Controllers have become network application development platforms
- Resource elasticity and virtualization enabled by NfV
- Interoperable network programming and collection standards arriving
- Predictive Analytics, Data Correlation are happening and accelerating – opening the door to Self Learning Networks (SLN)
- We can drive outcomes and network intelligence with all of the collected network data
- We use Splunk to help our Engineers solve customer issues

5

Our Customer Business Challenges Drove Us...

Growth, Scale, Customer Experience--Reduce MTTR, Increase Performance

Business Needs
Cost Reduction
Business Continuity
Service Assurance

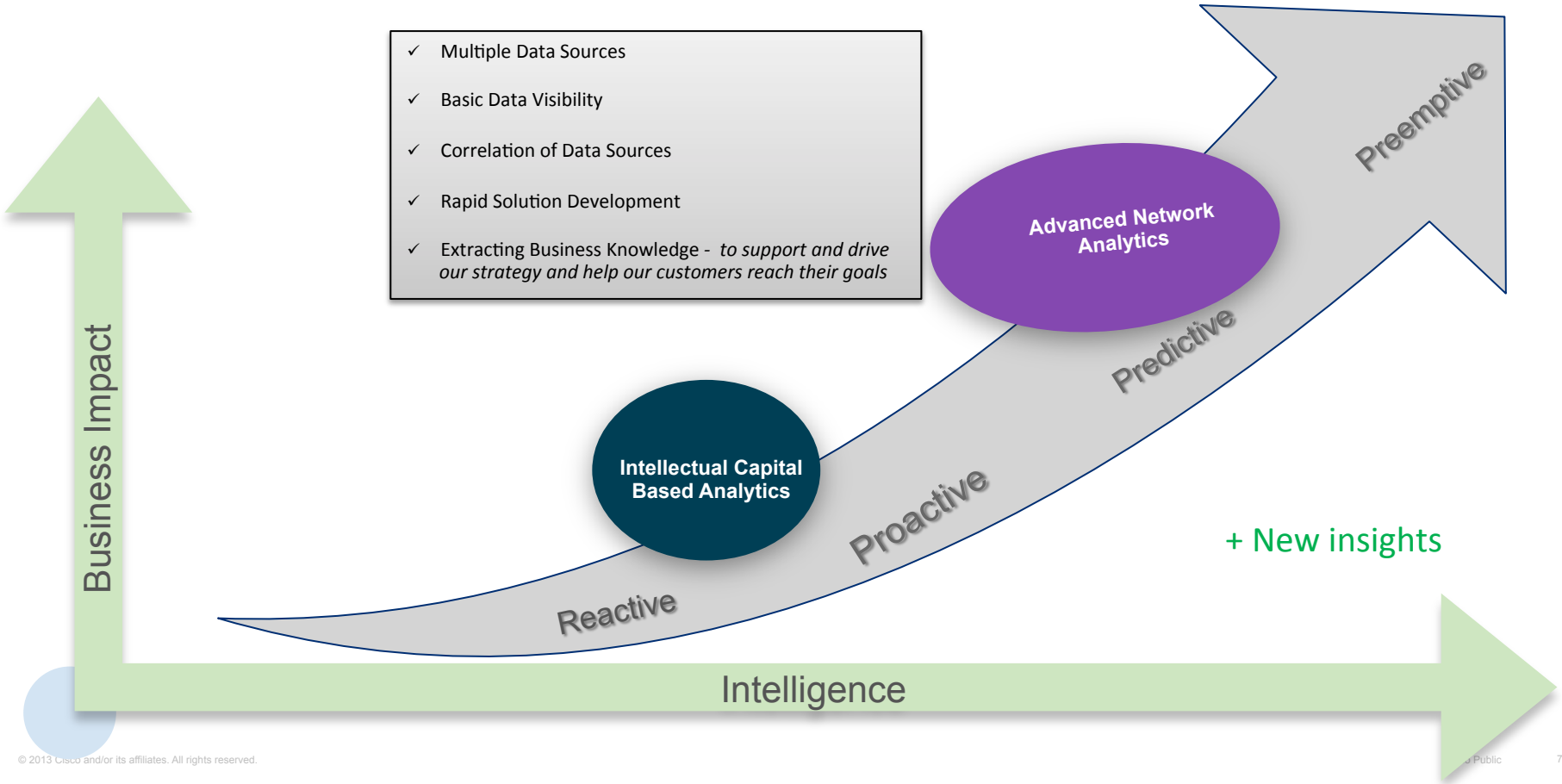


Analytics Capability
Performance & Capacity Analytics
Topology Analytics
Fault Analytics
Device Level Analytics

Exec: “How do I measure how well my business is running?....”

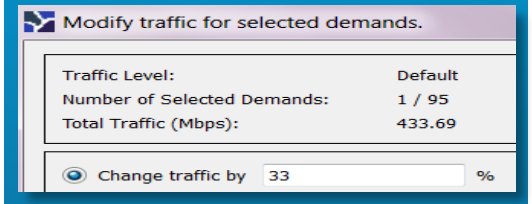
CCIE + Splunk Architect:
”Translate to Network KPI!..”

Cisco's Infrastructure Analytics Journey: Reactive → Proactive → Predictive → Preemptive (RP3)



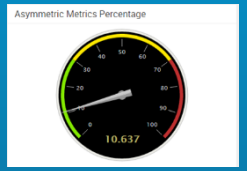
Multiple Teams Demand Infrastructure Analytics

Planning



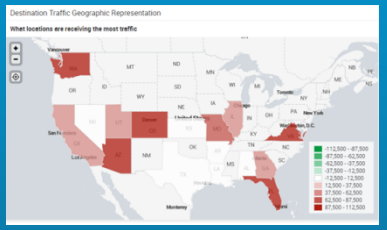
- Growth Forecasts
- Upgrade Analysis
- New Service Impact
- SLA planning
- Areas of Over-Capacity
- Areas of Under-Capacity
- Best Place to add a new Link
- Multi chassis vs. Single Chassis

Engineering/ Architecture



- Balancing Traffic
- Optimal Topology Design
- RSVP, QoS, Multicast Design
- The need for MPLS TE or not?
- Poorly defined Metrics
- Design Validation
- Asymmetric Metrics & Loads
- Inadequate QoS & Fragmented QoS

Operations



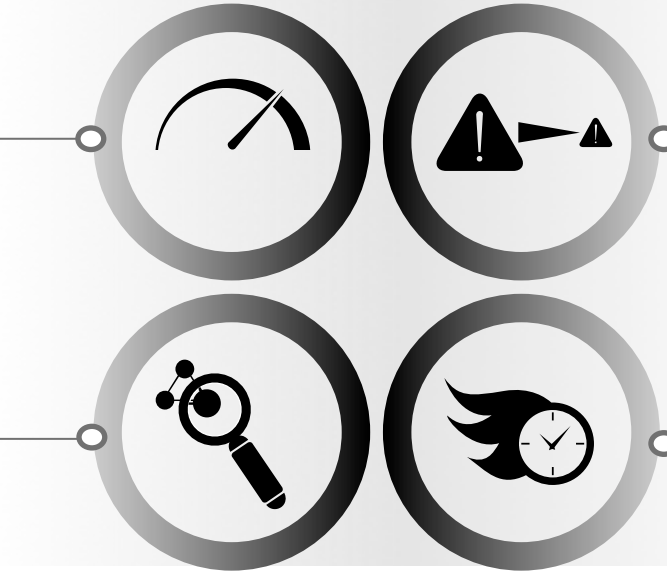
- Network Health and Traffic Trends
- Maintenance Planning
- Troubleshooting
- Congestion Mitigation
- Failure Analysis (RCA)
- Fault Analytics

Cisco Internal Infrastructure Apps We Have Built

Trending – Anomaly Detection - Recommendations

KPI Portal

Global
Universe
Benchmarking

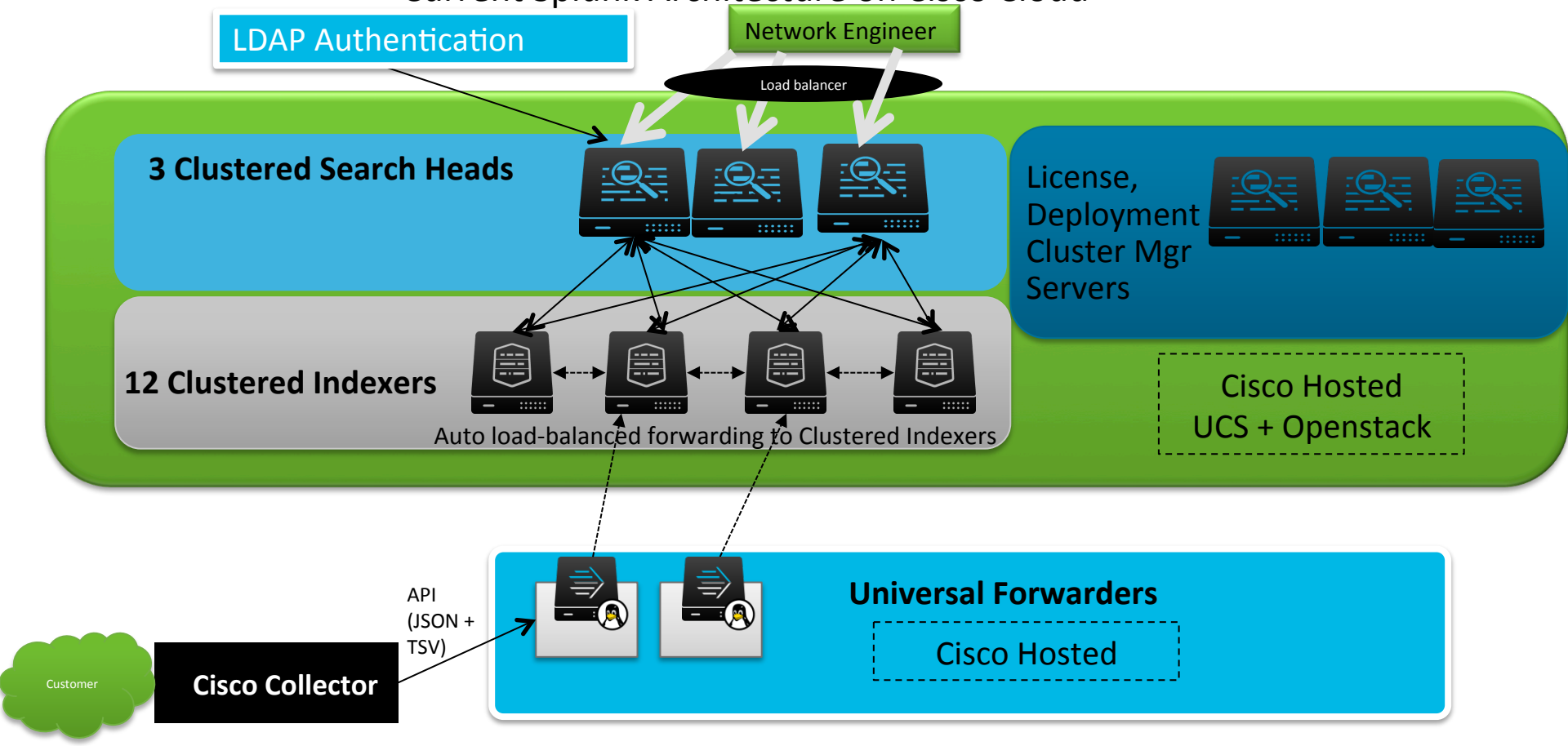


Syslog

Root Cause
Analysis

Good Morning Screen

Current Splunk Architecture on Cisco Cloud



Which Raw Data Do We Start With ?

```
[root@cariden-live flow]# more *
::::::::::::::
flow_matrix_file-latest
::::::::::::::
ETYPE,SRC_AS,DST_AS,AS_PATH,PEER_DST_AS,PEER_SRC_IP,PEER_DST_IP,IN_IFACE,OUT_IFACE,TOS,SAMPLING_RATE,PACKETS,FLOWS,BYTES
800,4444,2906,2906,2906,11.31.128.77,11.31.128.218,37,45,0,1,705346,0,32445916
800,2014,2906,2906,2906,11.31.128.77,11.31.128.218,35,45,0,1,4210857,0,193699422
800,2906,2014,2014,2014,11.31.128.218,11.31.128.77,30,103,0,1,146643,0,6745578
800,2906,4444,4444,4444,11.31.128.218,11.31.128.77,30,103,0,1,149161,0,6861406
800,2906,2014,,0,11.31.128.218,11.31.128.77,30,103,0,1,37783,0,1738018
800,2906,4444,,0,11.31.128.218,11.31.128.77,30,103,0,1,38174,0,1756004
```

Example Netflow Data

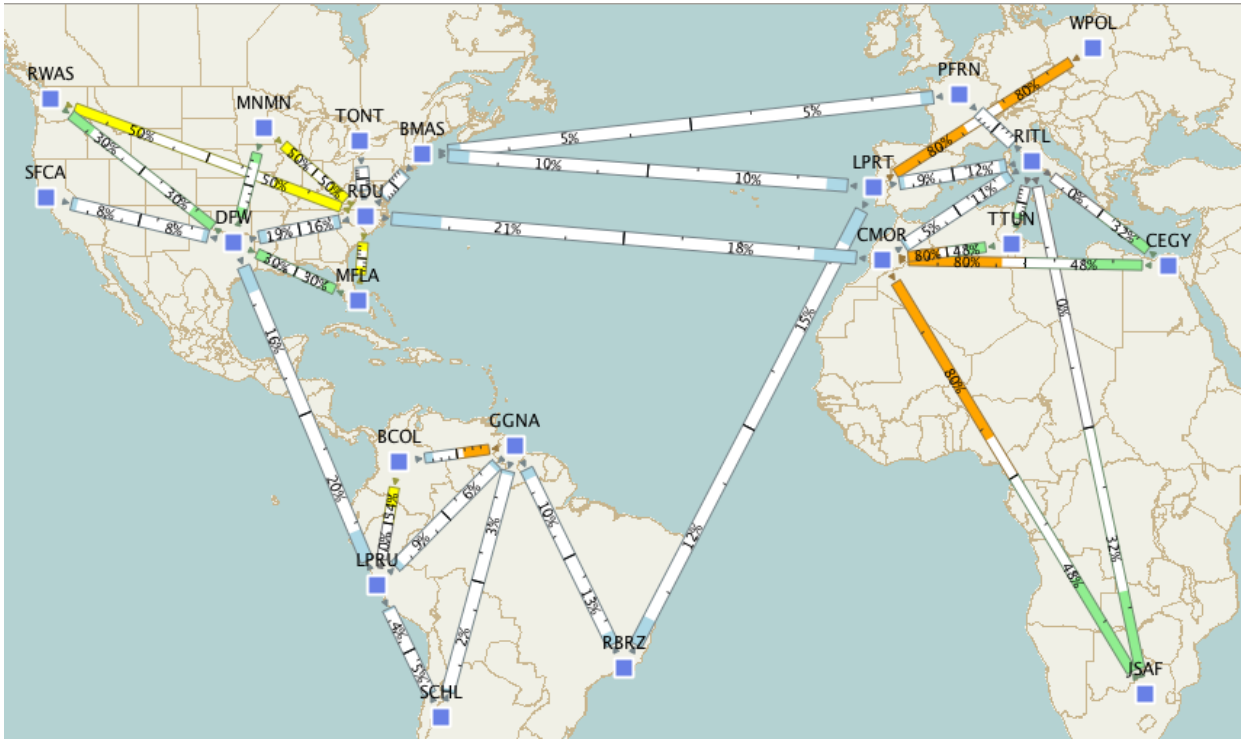
```
65000 AFI: all VRF: global
Apr 18 13:22:46 11.31.128.207 368: Apr 18 17:22:45.825: %BGP-5-ADJCHANGE: neighbor 11.31.5.21 Down User reset
Apr 18 13:22:46 11.31.128.207 369: Apr 18 17:22:45.825: %BGP_SESSION-5-ADJCHANGE: neighbor 11.31.5.21 IPv4 Unicast topology base removed from sessi
User reset
Apr 18 13:22:46 11.31.128.207 370: Apr 18 17:22:45.826: %BGP-5-ADJCHANGE: neighbor 11.31.128.45 Down User reset
Apr 18 13:22:46 11.31.128.207 371: Apr 18 17:22:45.827: %BGP_SESSION-5-ADJCHANGE: neighbor 11.31.128.45 VPNv4 Unicast topology base removed from se
n  User reset
Apr 18 13:22:46 11.31.128.207 372: Apr 18 17:22:45.827: %BGP-5-ADJCHANGE: neighbor 11.31.128.46 Down User reset
Apr 18 13:22:46 11.31.128.207 373: Apr 18 17:22:45.827: %BGP_SESSION-5-ADJCHANGE: neighbor 11.31.128.46 VPNv4 Unicast topology base removed from se
```

Example Syslog Data

```
<NetIntInterASFlows>
SourceAS  FromNeighborAS  IngressNode IngressInterface    EgressNode  EgressInterface ToNeighborAS  ASPath  DestinationAS  IPProtocolVersion  Traffic
2906      2906    asr9k-b TenGigE0/1/0/1  FF-B    TenGigE0/3/1/0  2014    2014    2014        101.82
2906      2906    asr9k-b TenGigE0/1/0/1  FF-B    GigabitEthernet0/3/0/0  4444    4444    4444        102.884
2014      2014    FF-B    TenGigE0/3/1/0  asr9k-b TenGigE0/1/0/1  2906    2906    2906        2282.413
4444      4444    FF-B    GigabitEthernet0/3/0/0  asr9k-b TenGigE0/1/0/1  2906    2906    2906        380.403
```

Aggregated Netflow Data

Network Topology Of ACME (Example)



Description:

- Global Provider of Voice and Data Solutions
- Points of Presence in Americas, Europe and Africa.
- Mixture of 10GE and 1GE interfaces
- Traffic Matrix for Voice and Data

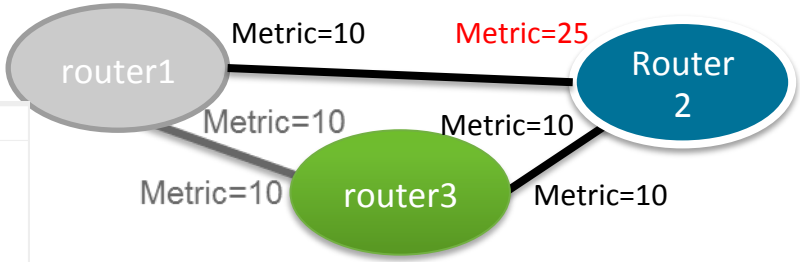
Goal is to Identify:

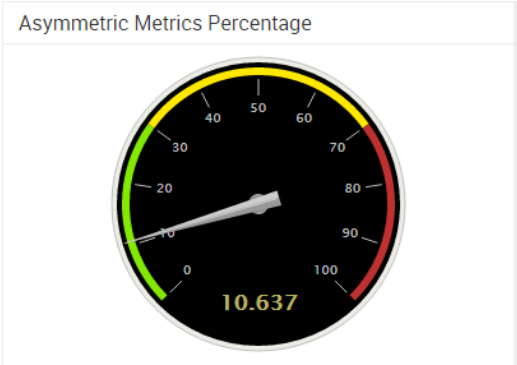
- Single Points of Failure
- Capacity Hot Spots
- QoS Analysis for VOICE Class
- Identify Suboptimal and High Latency Routing
- Load Balancing Optimization
- Topology Potential Improvements including circuit Upgrade Recommendations

Analytics Objective

Proactive: Topology Analytics

Identifying Outlier Metrics





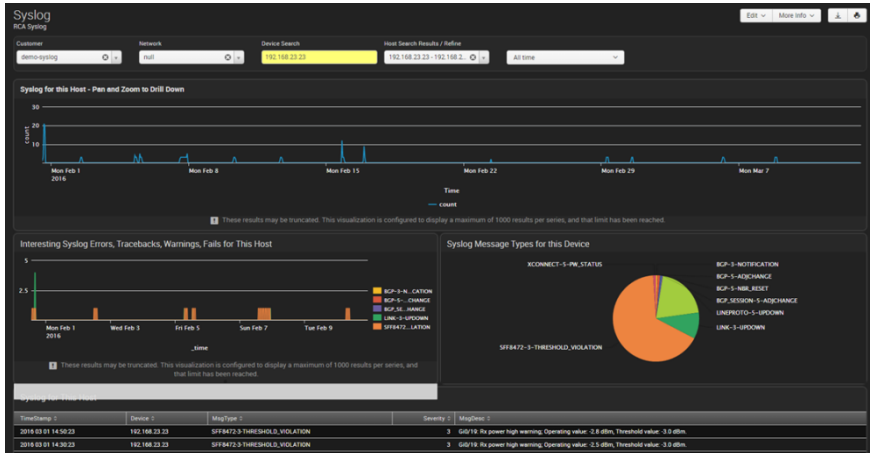
Asymmetry may lead to Potential Incorrect Routing

And can help reduce OPEX

Reactive: Root Cause Analytics

Stakeholders: Cisco and Customer Team

- Situation
  - Something “bad” happened to a device, but why …? What were the leading indicators prior to the failure? Typically Cisco has to call customer for more data
- Solution
  - Cisco uses data driven approach to identify Syslog messages, field notices, PSIRT, and Best practice alerts prior to the event using our RCA app.
- Outcome
  - Suggestions of potential root causes around and leading up to the time of failure MTTR. Saves time when creating a RFO and getting customer network restored

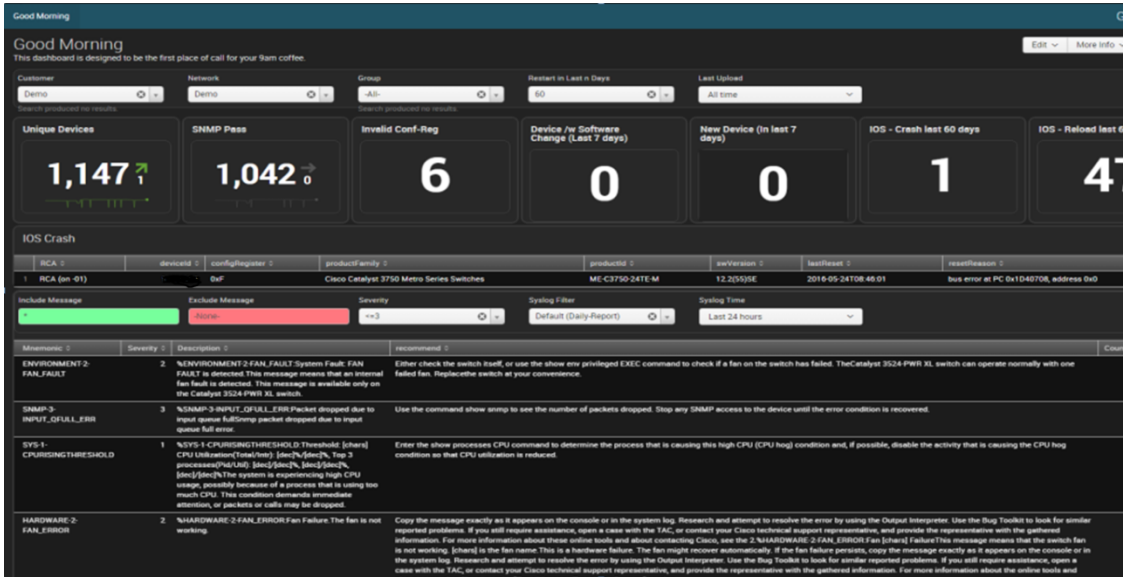


High priority Syslog messages, HW/SW version changes, last reset, PSIRT, Field Notices, EOX, Best Practices, Vulnerabilities in the timeframe desired

Proactive Engineer : What Changed?

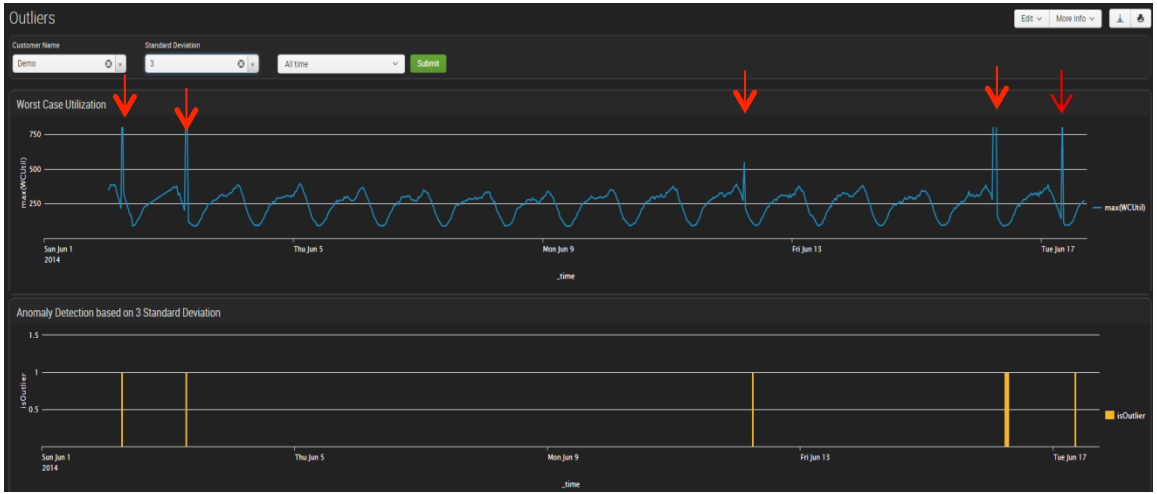
Stakeholder: Cisco and Customer Team

- Situation
  - What changed since yesterday?
  - What should I notify my customer about?
- Solution
  - AS team created “good morning” and “KPI” app to show data trends in great detail around faults to show KPIs that need attention right away with recommended course of action
- Outcome
  - Better NCE visibility to data in more meaningful way.
  - By changing config registers, standardizing code versions, implementing a best practices we prevented an outage



Proactive: Anomaly Detection

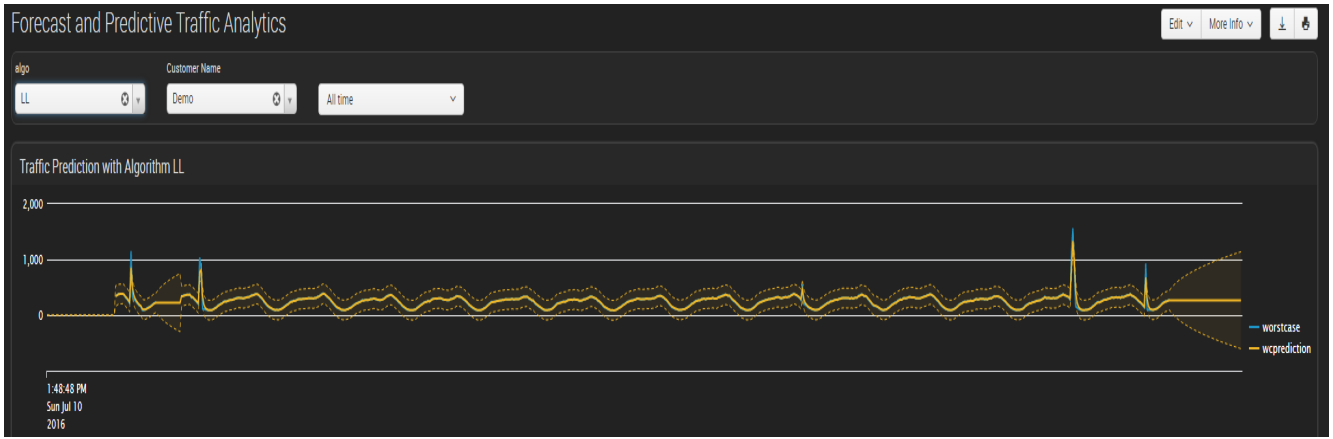
- A deviation in expected behavior (time series sensitive)
- Worst case utilization shown as a “Spike” in normal pattern
- Alerts can be generated for each anomaly detected
- Standard deviations can be tweaked
- Could also be:
  - Utilization spike
  - Number of Syslogs
  - Number of Reboots
  - QoS Policy or Config Change



Detect Numeric Outliers and find values that differ significantly from previous values.

Predictive: Link Capacity Forecasting

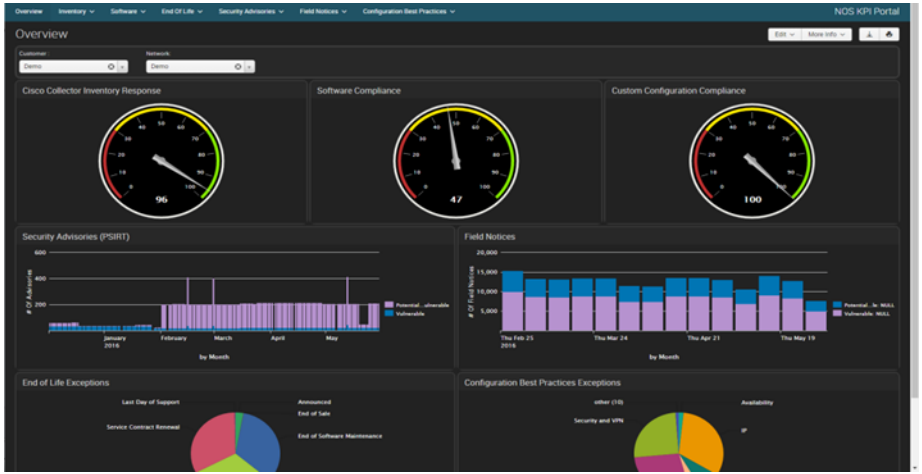
- Forecasting: estimation of future values based on past and seasonal patterns
- Great for Traffic Prediction for capacity planning
- Important for anomaly detection



Forecast Time Series:
Forecast future values given past values of a metric (numeric time series).

Proactive: Hardware, Software, Compliance

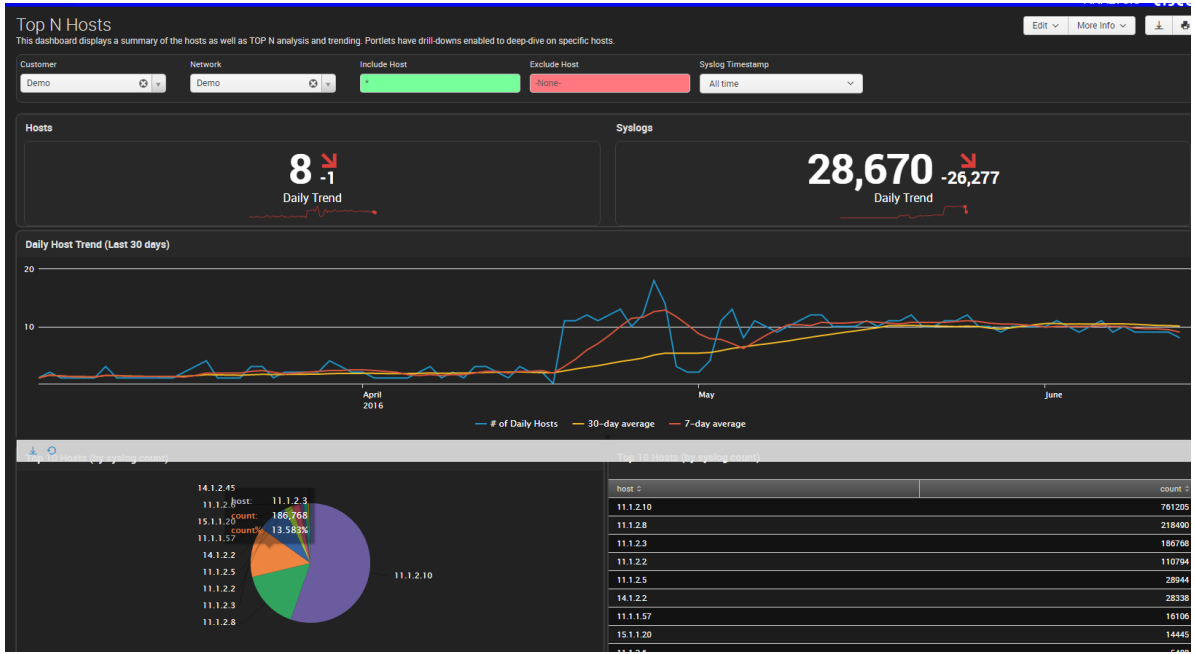
- Situation
  - What Hardware / Software compliance risks should I notify my customer about?
- Solution
  - AS team created “KPI” portal app to show data trends in great detail around compliance against Field Notices, Configuration Best Practices, PSIRT to show KPIs that need attention right away with recommended course of action
- Outcome
  - Better visibility to data in more meaningful way
  - Key Indication of overall health of a customer network



Network Dashboard

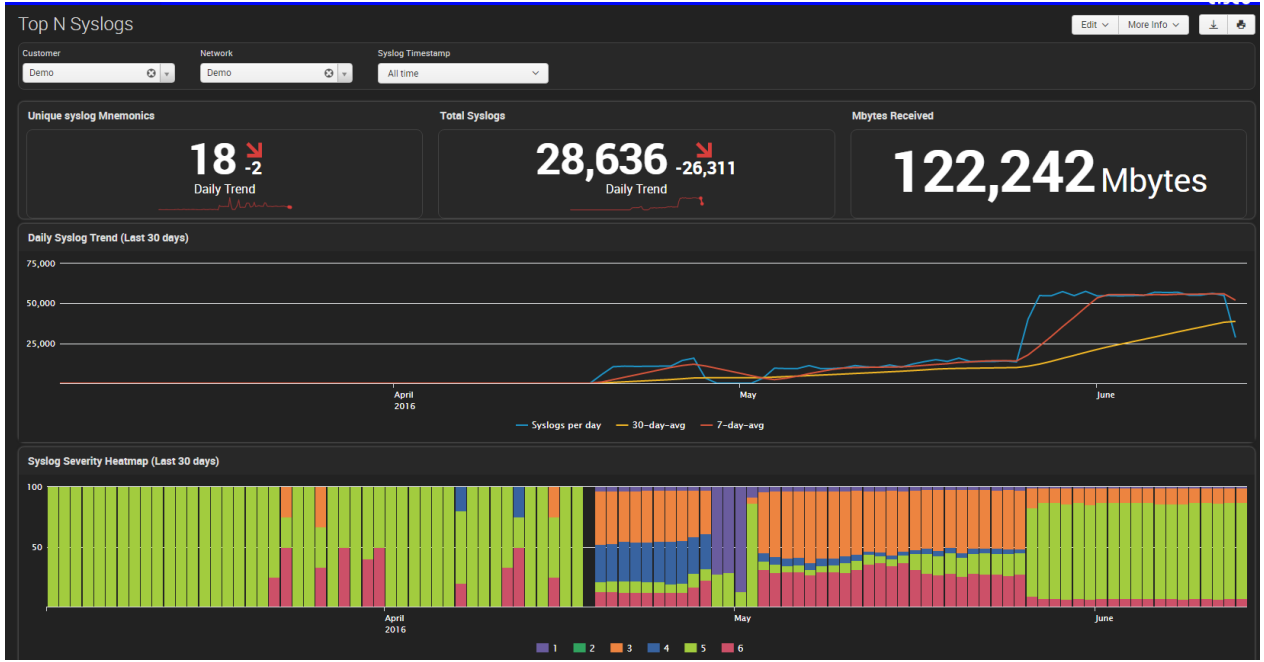
Understanding “Steady State”, Creating a Baseline

- Quick View of the last 180 days
- Determine which hosts are missing
- Identify Outliers leveraging Host/Syslog Trend
- Identify Network Issues through Syslog Severity Heat map
- Alerting for Anomalies
- Severity Heat map



Syslog Severity Distribution

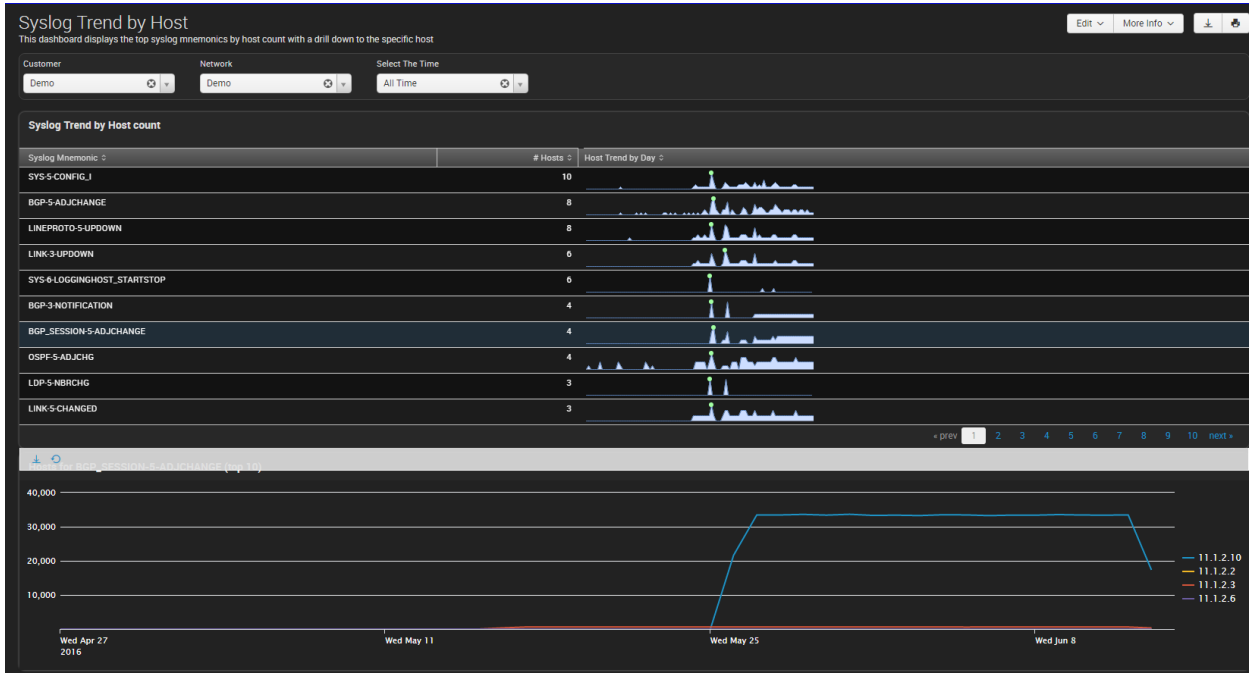
- Top Syslogs by # of Hosts
- Drill-down to get host heat map
- Critical for Baseline creation and proactive actions



Syslog Trend By Host

Where to focus “Level 3 Support”

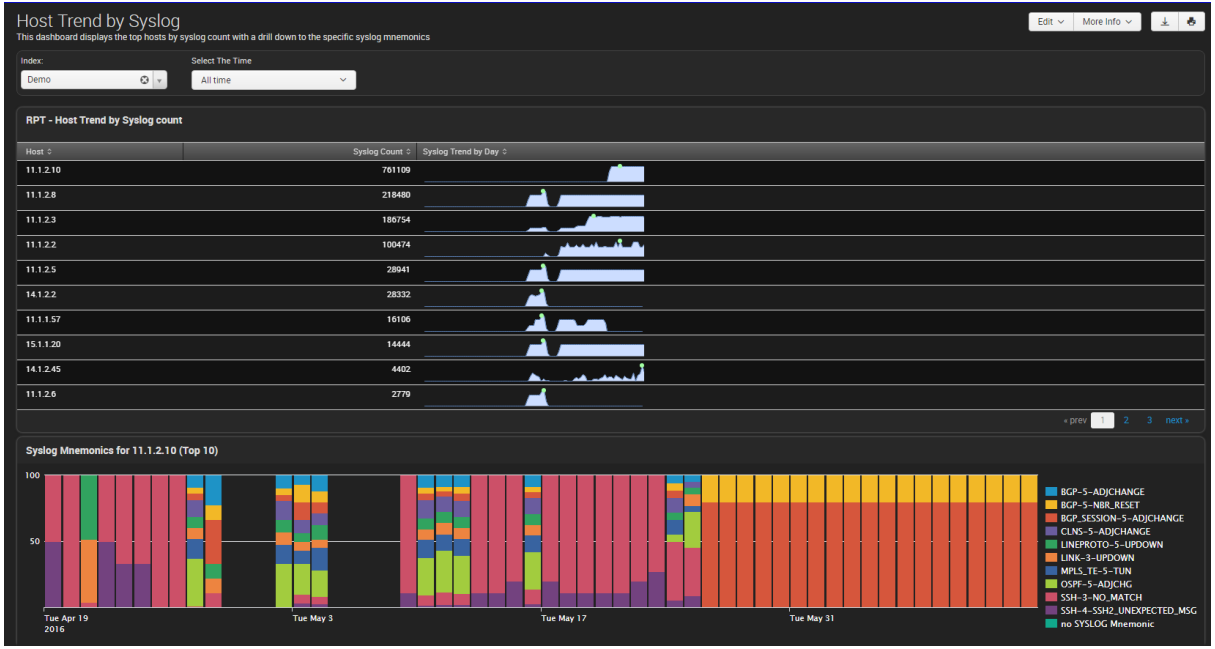
- Identify which Syslogs and Hosts have largest potential problems
- Top Syslogs by # of Hosts
- Drill-down to get host heat map
- Sparkline to show quick glance trend
- Possible to correlate to capacity Issues



Rare Syslog Trend By Host

Finding “The Needle in The Haystack”

- Identify Rare Syslogs and Hosts based on host impact
- Top Syslogs by # of Hosts
- Drill-down to get host heatmap



Looking At Cisco And Splunk Better Together

What is Cisco's Unified Computing System (UCS)?

Unified Management: UCS Manager uses policy-based configuration to ensure consistent deployments

Unified Fabric: Integrated 10/40 Gigabit Ethernet and Storage Networking (FCoE/ iSCSI)

Service Profiles: Maintain consistency across batches of servers and multiple applications. Deploy and expand in record time.

Performance: Built with 10GbE and 40GbE at the core, repeatable configurations and performance, and over 100 benchmark records

Why Splunk On Cisco UCS?

Time to Deployment: Spin up a mutually validated, pre-tested environment in hours rather than days or weeks

Total Cost of Ownership: Integrated networking and management reduce customer cost and effort to migrate, deploy, and expand

Time to Grow: Expand servers and network capacity quickly and consistently

Why Does Hardware Still Matter?

- Cisco customer big data pools tend to grow 2-3x/year
- Cisco customer IT staff doesn't grow as fast
- The Cisco Unified Computing System (UCS) provides scalable, repeatable, predictable, and manageable deployments across dozens to thousands of servers for any application deployment
- Pallet to production in hours, not days or weeks
- Deep engineering integration between Cisco and Splunk with tested and proven configurations

More on this later…

Why Cisco UCS For Any Hardware Deployment? Ten Second Edition

- Scalability
- Manageability
- Performance

(longer version at rsts11.com)

Why Cisco UCS For Any Hardware Deployment?

- Single point of management and access control for thousands of servers
- Centralized host/network/storage/lights-out firmware management built in
- High performance networking at the core
- Flexible network configuration with vNICs for security and scalability
- Open XML API for automation and third party integration
- Fully functional remote console (including virtual media) *at no extra cost*

Why Do These Matter For Big Data Specifically?

- Single point of management and access control for thousands of servers
- Centralized host/network/storage/lights-out firmware management built in
- High performance networking at the core
- Flexible network configuration with vNICs for security and scalability
- Open XML API for automation and third party integration
- Fully functional remote console (including virtual media) *at no extra cost*

- **Big Data environments grow faster than most platforms. Ever added 100 Oracle servers?**
- **Big Data environments tend to grow 2x-3x (or more) within two years. IT staff do not.**
- **More data moving around means heavier pressure on the network to perform**
- **New software models may require different networking and storage**
- **Larger companies have existing management infrastructures to work with**
- **Some vendors nickel and dime for management features and licenses.**

Cisco UCS + Splunk = Better Together

Seamless Scalability Facilitates Rapid Growth

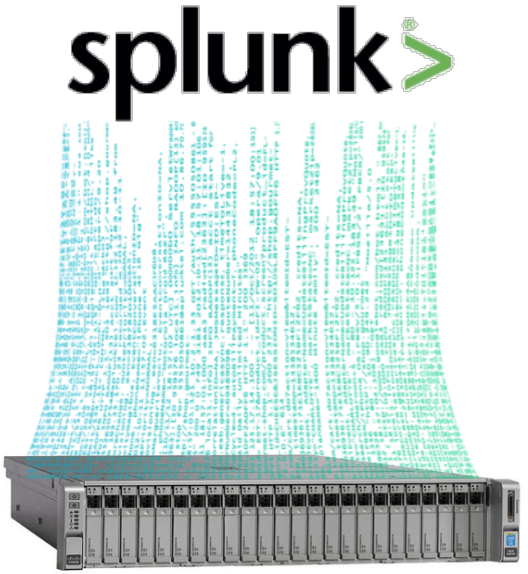
- Scale Splunk from a single server to distributed/clustered deployment
- Grow your clusters efficiently and consistently
- Runs on the same UCS C-Series servers as other big data platforms

Split Second Response Times

- Exceptional performance for "needle-in-a-haystack" searches
- **Consistent performance** as simultaneous users increase

Simplified **Repeatable Deployments**

- Four pre-tested UCS Reference Architectures
- Retention (Capacity) or Performance optimization
- NEW! Cisco Validated Design (CVD) with HA and Archiving



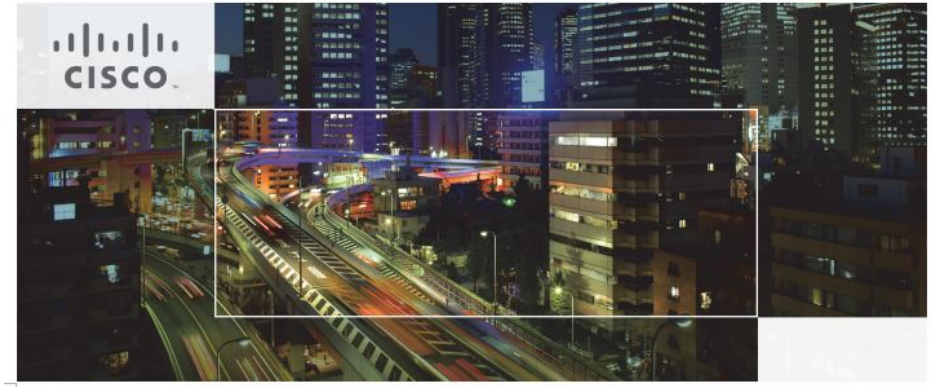
Cisco UCS Reference Architectures

	Single Server	Clustered Deployment
Retention optimized	250 GB indexed per day 4 months retention	Up to 4TB indexed per day 1 year Retention
Performance optimized	250 GB indexed per day 1 month retention	Up to 4TB indexed per day 3 months Retention

Reference Architecture Solution Brief [cisco.com]

Cisco Validated Design (CVD) For Splunk

- Developed by Cisco & Splunk engineers in mid-2015 (Update coming soon!)
- 250+ page guide to design and deployment, pallet to production
- Based on UCS C-Series (C220, C240, C3160) servers and Splunk Enterprise software
- Includes high availability & data archiving
- Download for free at cisco.com/go/bigdata_design



THANK YOU

splunk>