Splunk for Akamai Cloud Monitor

Pierre Pellissier

Leela Kesireddy

Performance Management

PayPal, Inc.

splunk>

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.

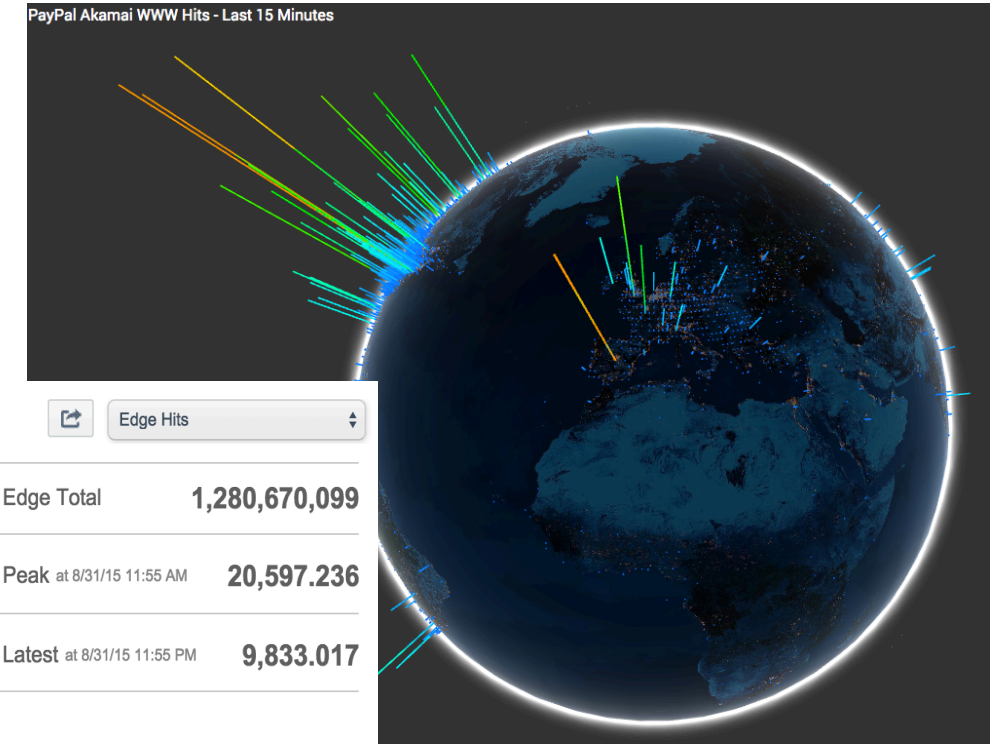
In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not, be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.

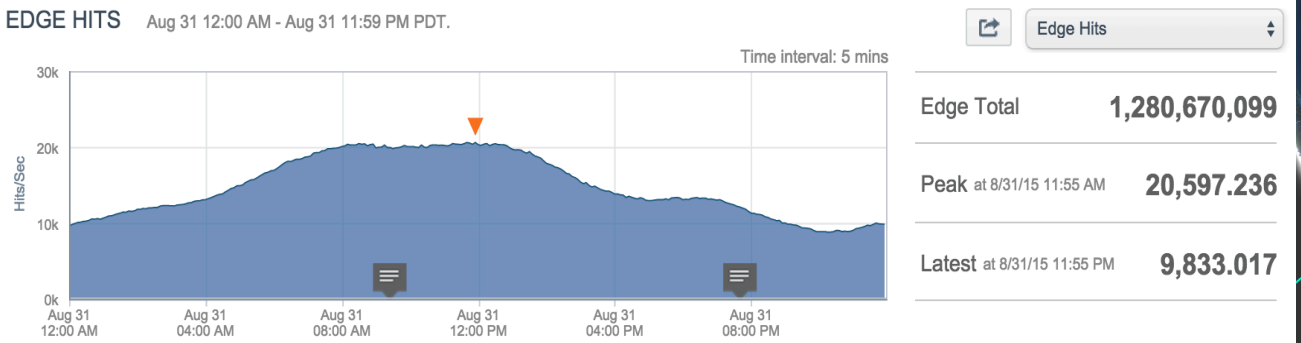
Agenda

- Why we need Log Analysis
- What we had was too slow
- Real Time logging Option
- Getting to Real Time
- Benefits & Challenges
- Configuration Details
- Akamai Splunk Application
- Final Analysis and Q & A

Why we need log analysis

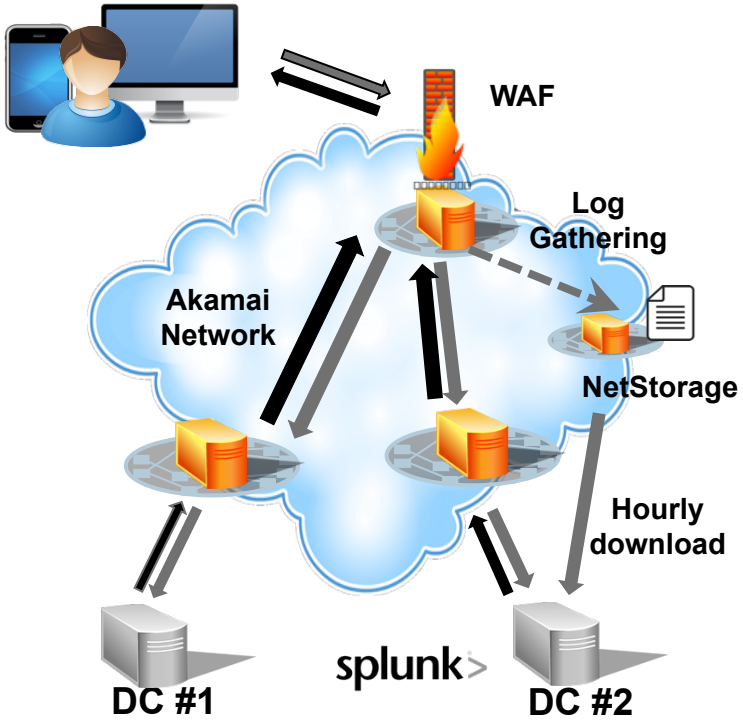
- Customer interactions on Akamai exceed 1 BN/day
- It is business critical that we :
  - Know the state of the site
  - Track the impact of changes
  - Analyze the customer experience





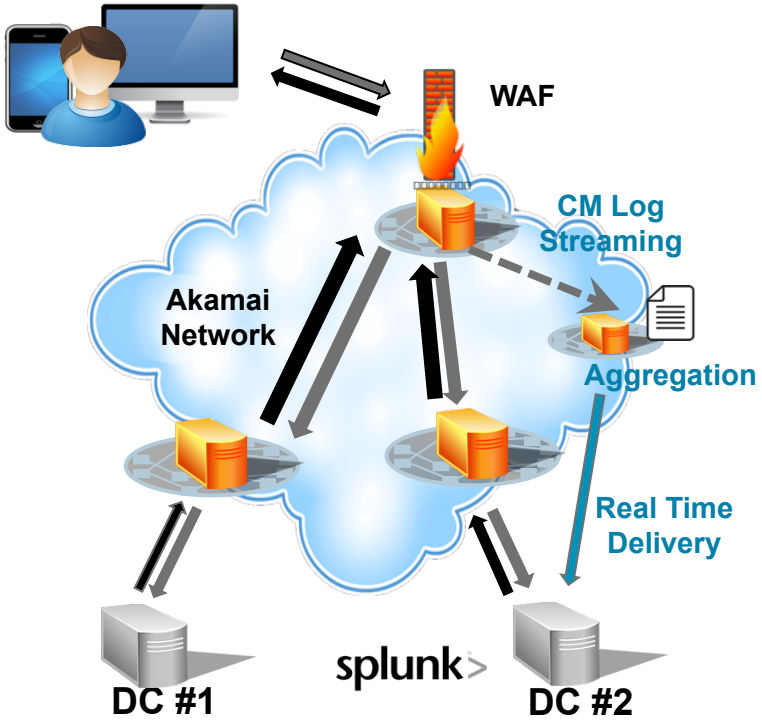
What we had was too slow

- Akamai for content delivery, web acceleration and security
- Akamai log collection takes hours to get logs into NetStorage.
- Downloading logs hourly and indexing them in Splunk makes them available for analysis.
- But this is not a real time solution.



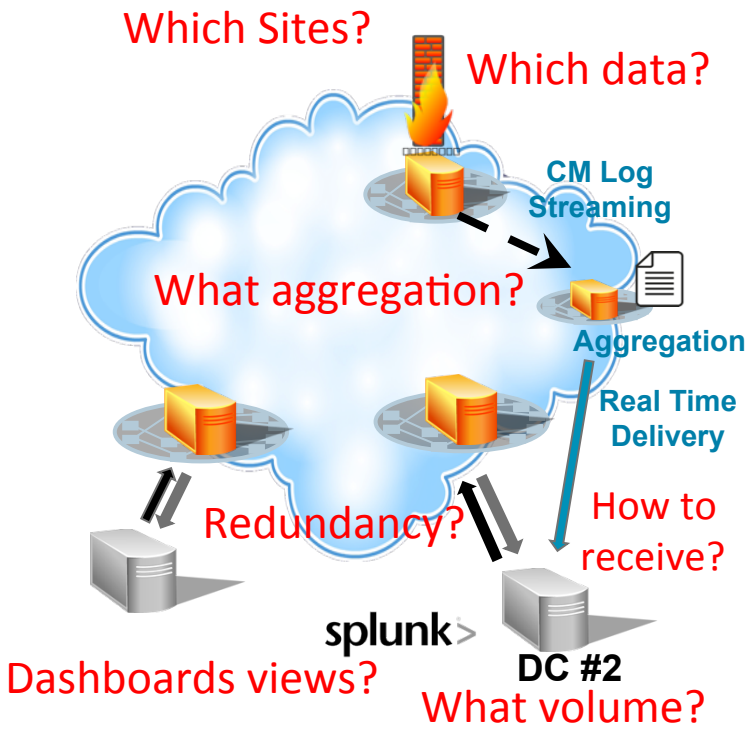
Real time logging option

- Akamai Cloud Monitor delivers logs in < 60 sec
- Faster delivery enables real time operational monitoring and easier analytics by internal users.
- Lots of configuration decisions.
- The results are worth the effort.
- Ultimately easier than expected.



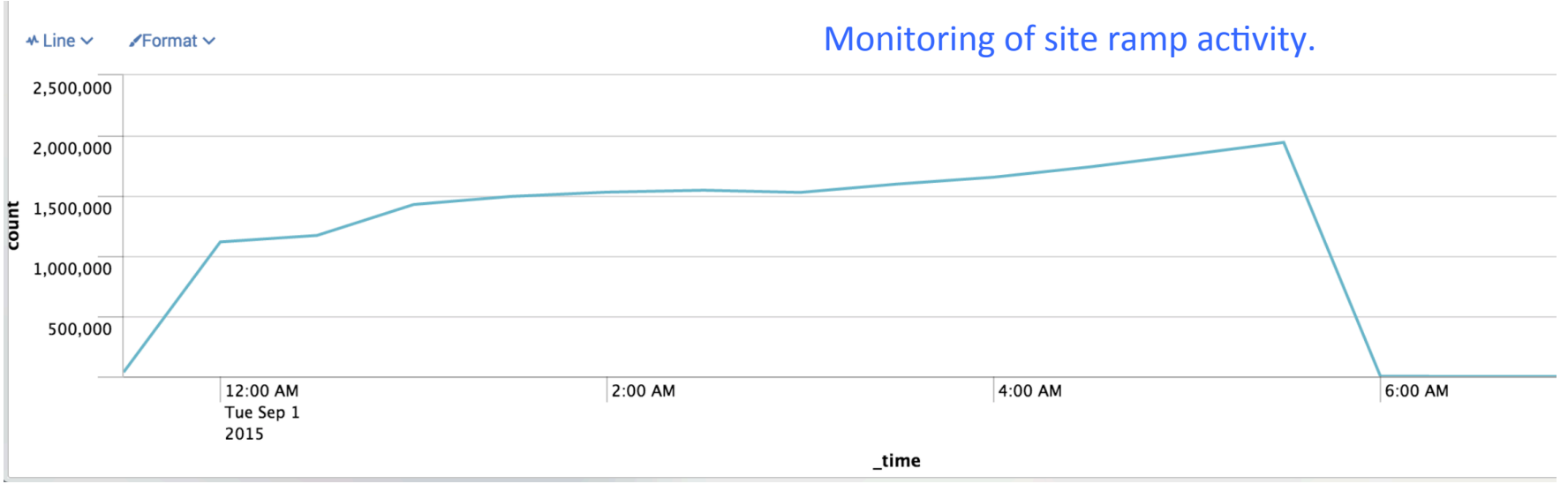
Getting to real time

- Operational goals required:
  - Real time visibility into site
  - Improved usability of logs
- Lots of design decisions needed to be made
  - How will we receive the logs?
  - How do we configure CM on Akamai?
  - How will this change the indexed volume?
  - How do we ensure we get the logs?
  - How will the logs be used?
- The initial benefits were achieved quickly
  - Multiple “bonus” benefits were also realized
  - Some challenges needed to be dealt with



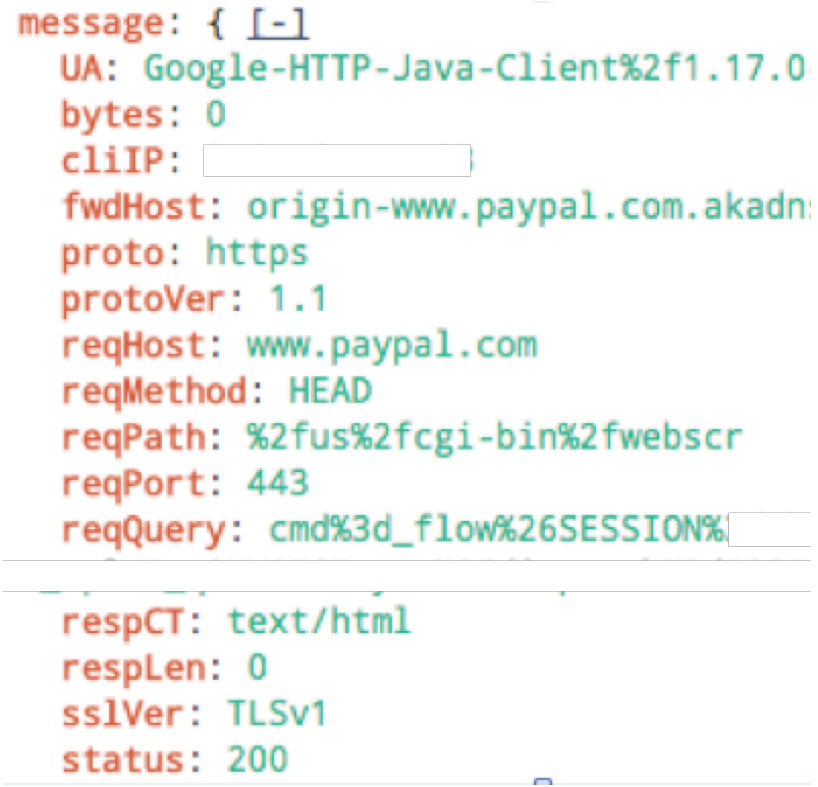
Initial benefits – site control

- Receiving the logs in <60 sec. makes them usable for real time analysis
  - The NOC can immediately see the effects of site changes
  - The CDN team can see the effects of Akamai configuration changes in real time



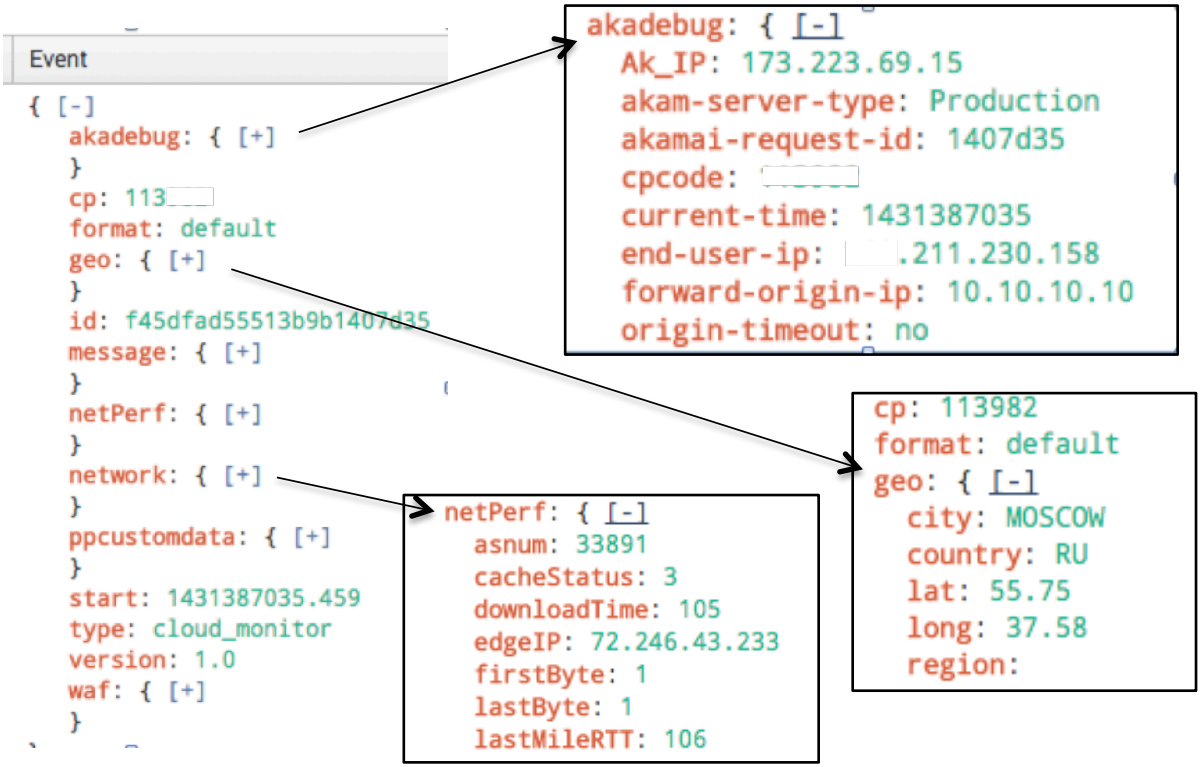
Initial benefits – log usability

- NetStorage log parsing needed improving
- Akamai CM logs use JSON formatting
- Splunk TA for CM logs put them into CIM format for ES consumption
- Internal customers can work with the JSON formatted logs more intuitively



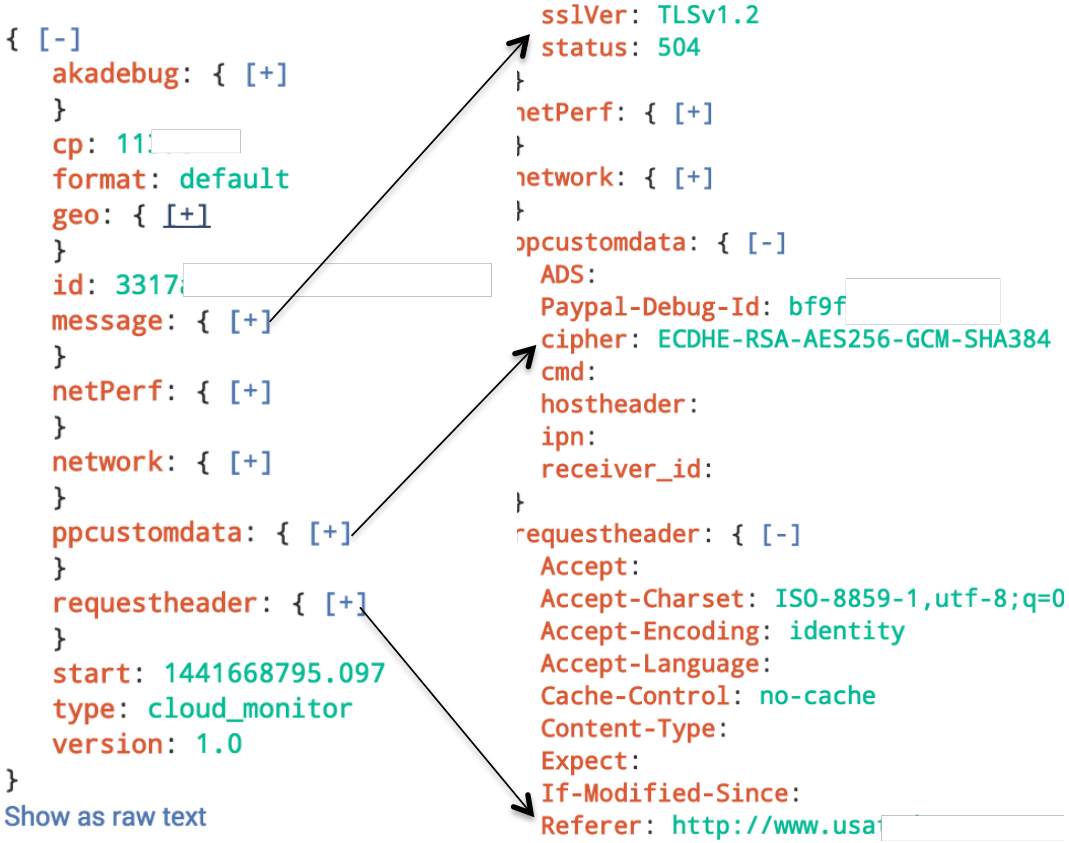
Additional benefits – optional CM data

- Optional data fields enable rich analytics for:
  - Operational Teams
  - Performance
  - Security
  - Fraud
- Adding more fields results in higher indexed volume



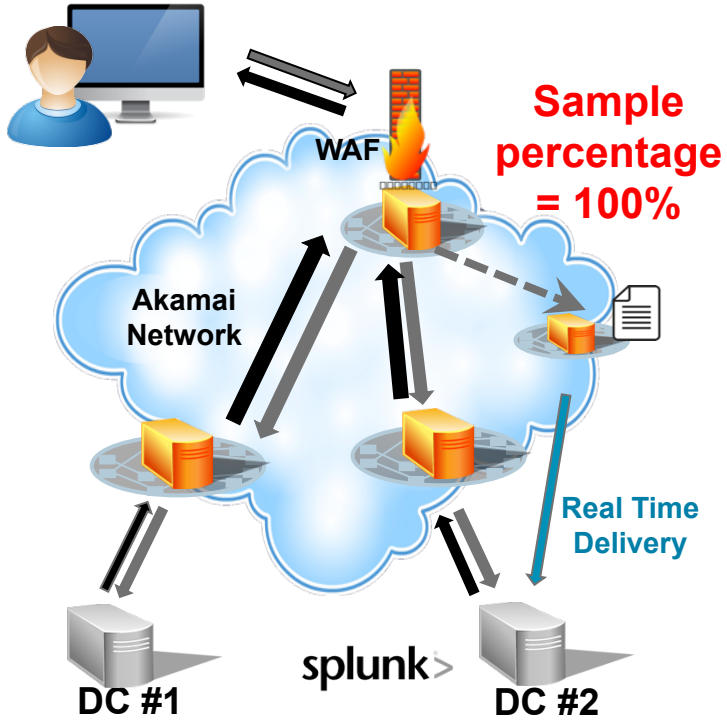
Additional benefits – custom fields

- Data from the connection, HTTP header or payload can be inserted into CM
- Enables powerful insight for multiple teams:
  - Crypto use analysis
  - Customer Experience
  - Referer Analytics
  - Troubleshooting



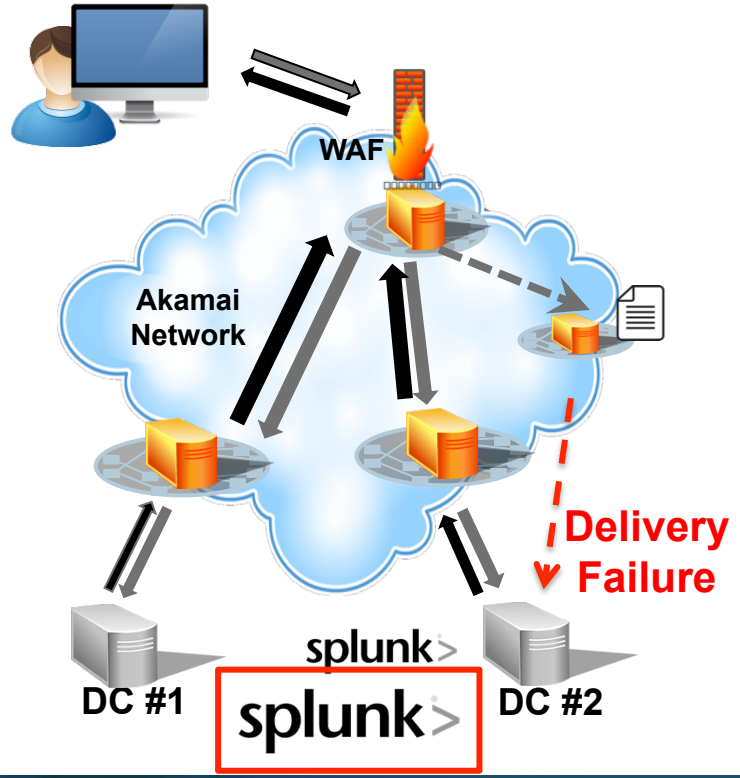
Additional benefits – more complete data

- For real time analytics we chose to receive 100% of the CM logs
  - Some use cases may only need 10%
- NetStorage log collection does not capture 100% of the logs
- The CM log count is ~5% greater than the NetStorage log count
  - This increases our confidence that we are seeing the entire picture



Challenges

- Greater log volume with CM
  - ~50% more data indexed
  - Partly due to JSON formatting
  - Mostly due to additional CM data
- Initial configuration of log receiver did not provide redundancy
  - CM Logs were lost when it was unavailable
- Log forwarding logic took several iterations to get it right
  - A POC got us close and then we adjusted after we were in production



Configuration Details

splunk>

CM & splunk checklist

Initial configurations:

- ❑ Add Cloud Monitor to the Akamai contract
- ❑ Set up receiver VIP, DNS, SSL Cert and redundant servers
- ❑ Choose the property and CM data sets to log
- ❑ Build the initial CM and Splunk configurations

Trial Run:

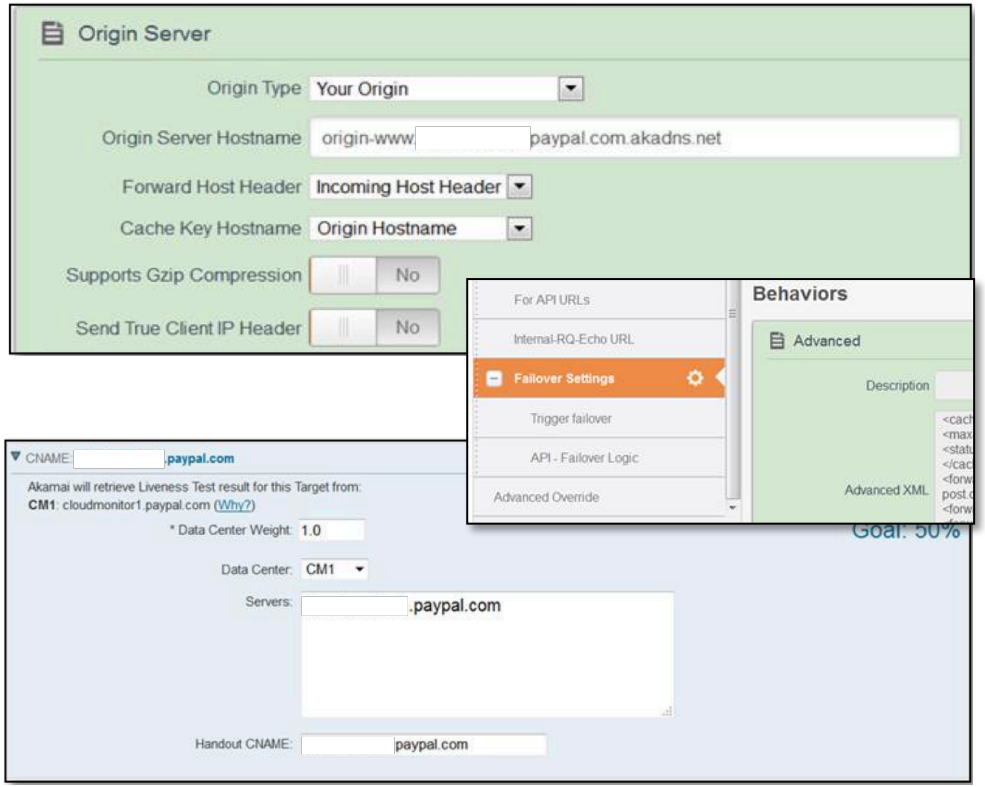
- ❑ Configure for a short duration of logging at 10% to verify receiver is working
- ❑ Verify the CM data is being mapped to Splunk CIM
- ❑ Verify that the receiver and Splunk can support the expected load
- ❑ Verify the data is what you need and expect

Plan on doing several rounds of configurations, testing and tuning

- ❑ It is easy to iterate and change the configurations as needed

CM data delivery

- Select Data Sets to Include
- Configure Delivery Endpoint
- Configure the Aggregation options:
  - Time: 60 seconds max
  - Line Count: Max 3000 records
  - Message Size: Max 900 KB of data
- Configure Distribution & Failover Options:
  - Primary receiver gets 100%
  - Secondary receiver gets 100% if Primary receiver is unavailable
  - NetStorage gets 100% if Primary & Secondary are unavailable
    - With scheduled FTP download hourly

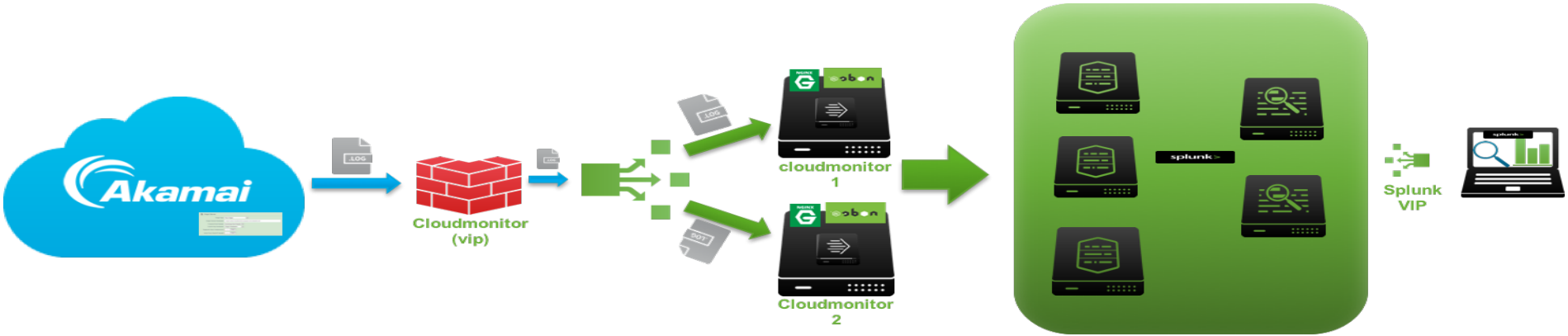


CM configurations

- Default data sets:
  - CP, Format, Message, Type, Version
- Optional data sets:
  - Akadebug, network, netPerf, Geo, WAF, PPCustomData
    - Can not opt out of data elements within an optional data set
    - Results in duplicates of some data or unwanted data
- Custom Data field:
  - Selected HTTP Header data is included
  - Other header info is excluded- like large cookies

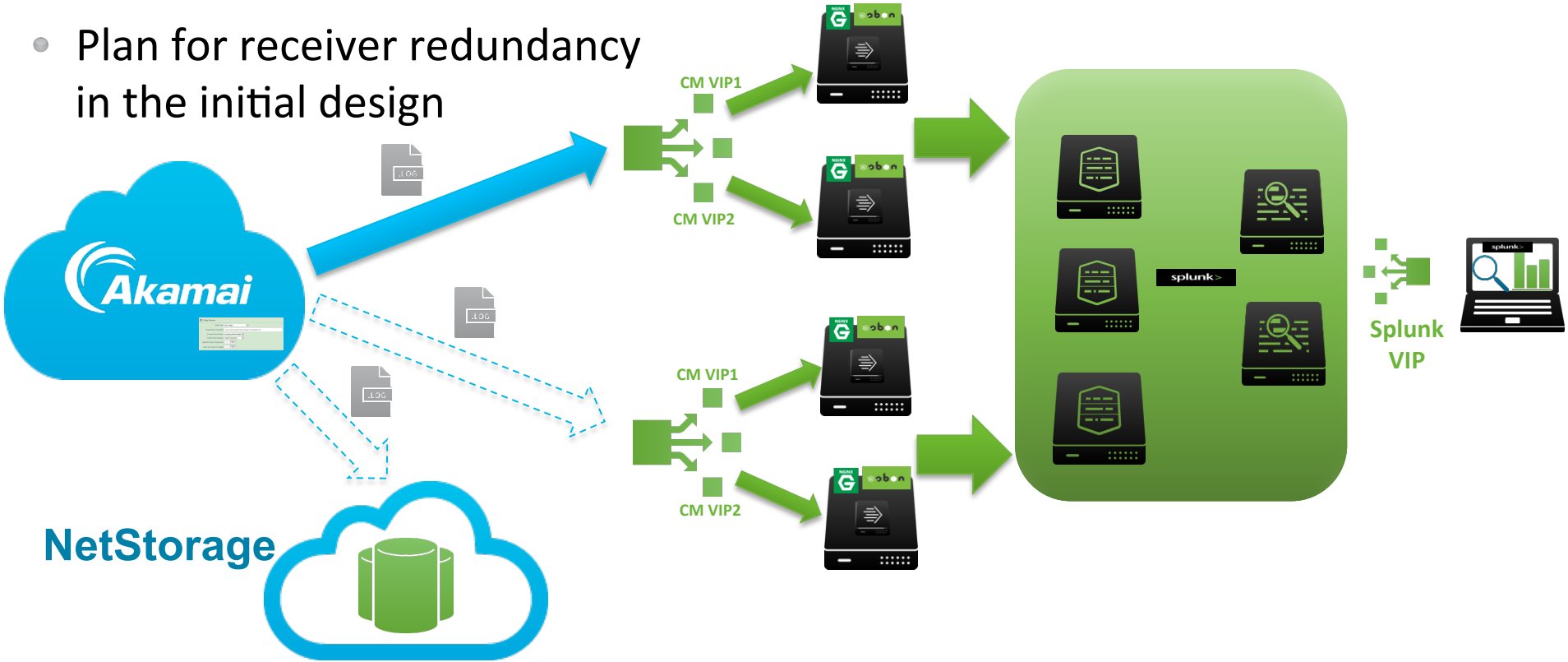
CM receiver

- Receiver build
  - Akamai Posts CM data to receiver VIP
  - SSL cert required for secure connection by Akamai
  - Multiple Linux servers in pool running nginx & node.js
  - Writes logs to a local data file
  - Splunk Universal Forwarder monitors logs and forwards them to Indexers



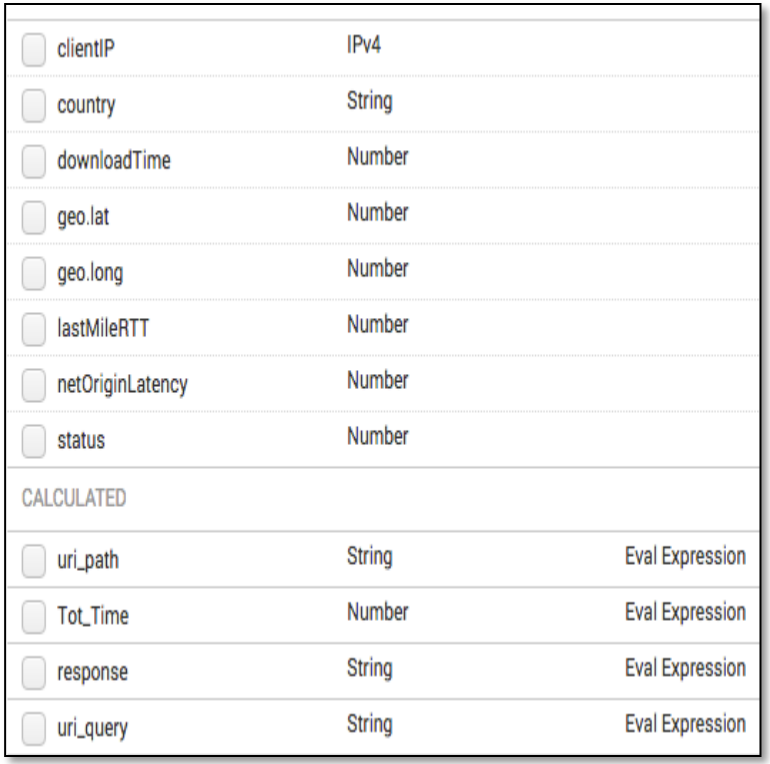
CM receiver (future state)

- Plan for receiver redundancy in the initial design



Splunk configuration

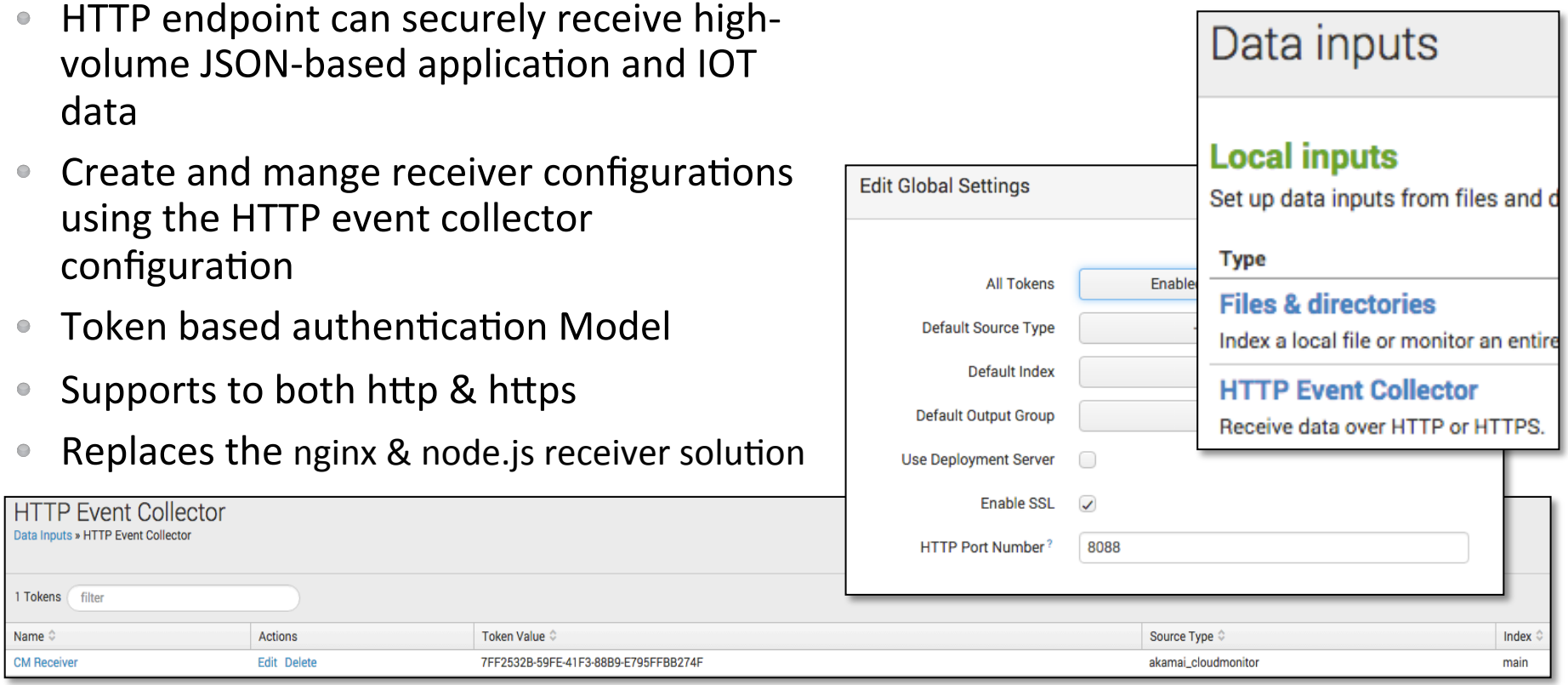
- Index configured with 100 day retention
- Built Add-on's with Field renames & Evaluations
- Normalized data to be CIM Complaint
- LDAP group configured for accessing CM data
- Apps set up for internal teams to build and share searches
- Common Akamai CM Dashboard for all teams
  - Performance & Availability Dashboard for each property
  - Starting point for internal teams to view data but with click through access to granular data and search bar



clientIP	IPv4	
country	String	
downloadTime	Number	
geo.lat	Number	
geo.long	Number	
lastMileRTT	Number	
netOriginLatency	Number	
status	Number	
CALCULATED		
uri_path	String	Eval Expression
Tot_Time	Number	Eval Expression
response	String	Eval Expression
uri_query	String	Eval Expression

New- Splunk 6.3 HTTP event collector

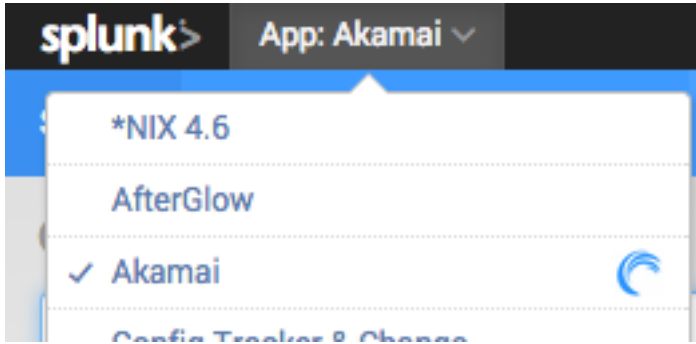
- HTTP endpoint can securely receive high-volume JSON-based application and IOT data
- Create and mange receiver configurations using the HTTP event collector configuration
- Token based authentication Model
- Supports to both http & https
- Replaces the nginx & node.js receiver solution



Akamai Application

splunk>

Akamai App



- Monitor the Real Time Traffic
- Insights by Property
- Monitor Health & Performance of Origin
- Monitor Heath & Performance of Akamai Edge
- Observe User's Pain Points
- Inspect App / URL level Issues & Performance
- Investigate Issues

Hurdles

- Performance Issue
  - Searches Were Slow
  - Dashboards taking longer to load
  - Greater than 25K events/sec
- Summary Index
  - Loses the Rich information in data
- Report Acceleration
  - Acceleration is suspended as the summary reaches 10% of its total size
  - No Control on Summary Schedule

Akamai Application

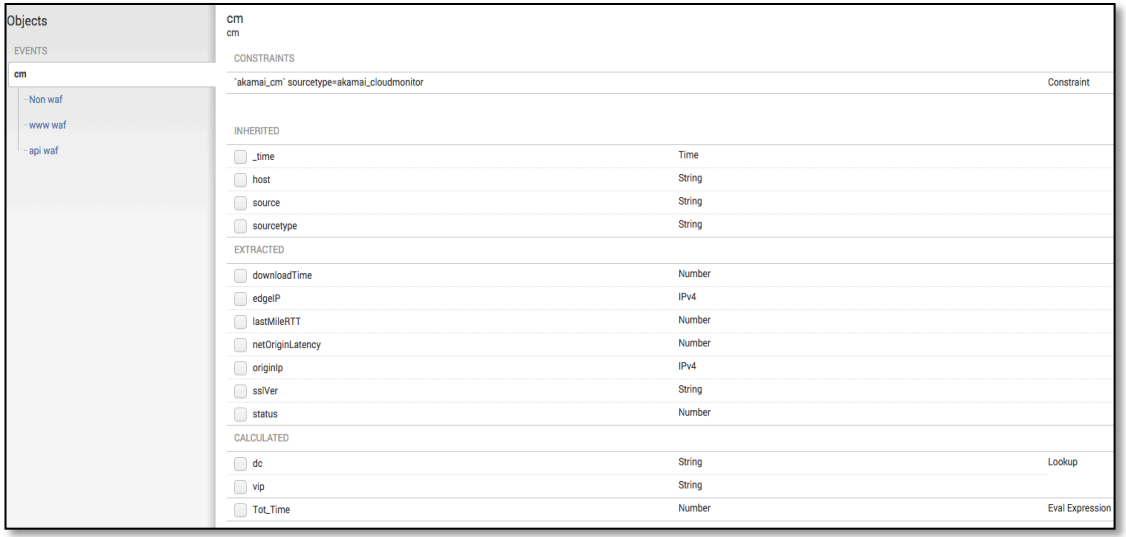
Data Models & Acceleration

Came to Rescue

Data Models

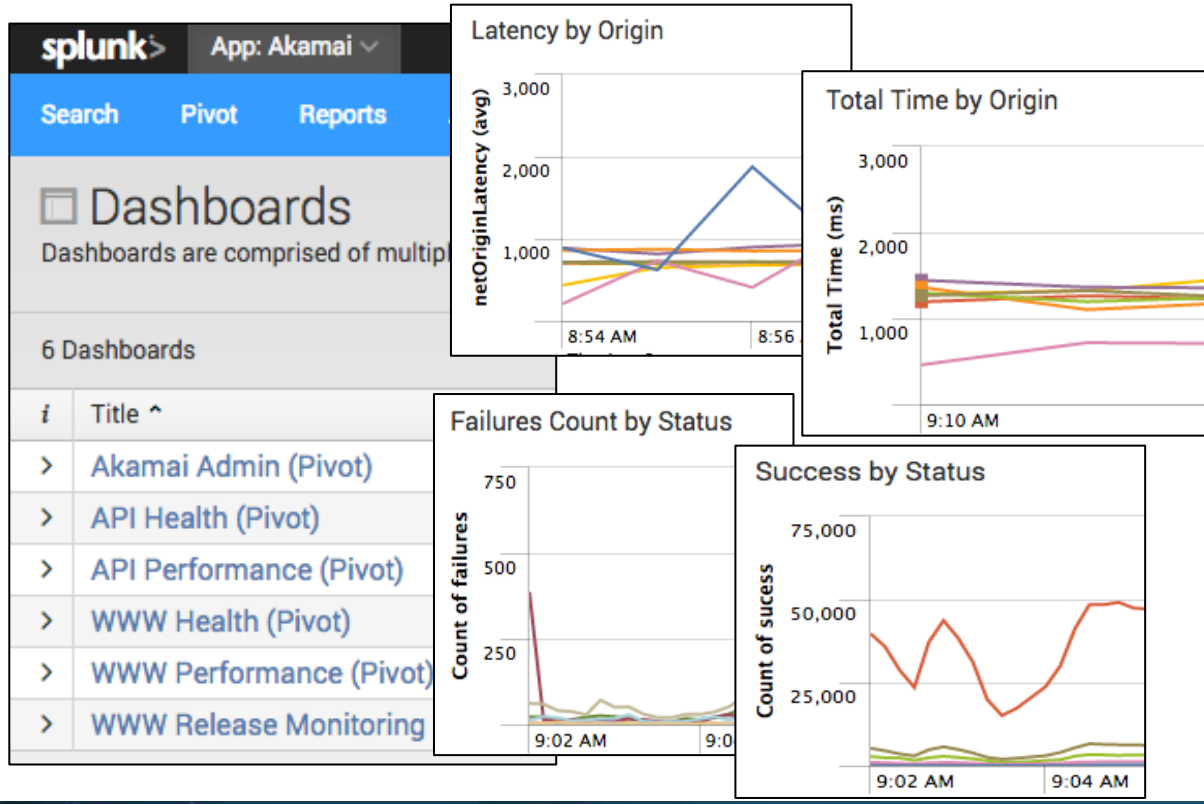
Data Model's facilitated to format, evaluate fields and work with necessary event data for each data set and also accelerate the model

- Data models are made up of object types like lookups, transactions, fields extractions & Calculated fields.
- Data models encodes only necessary knowledge objects.
- Data Model for each Property or group of Properties.
- Objects by Success or Failure (HTTP Status)
- Add additional fields using regular expressions, eval expressions and lookups.



Dashboards

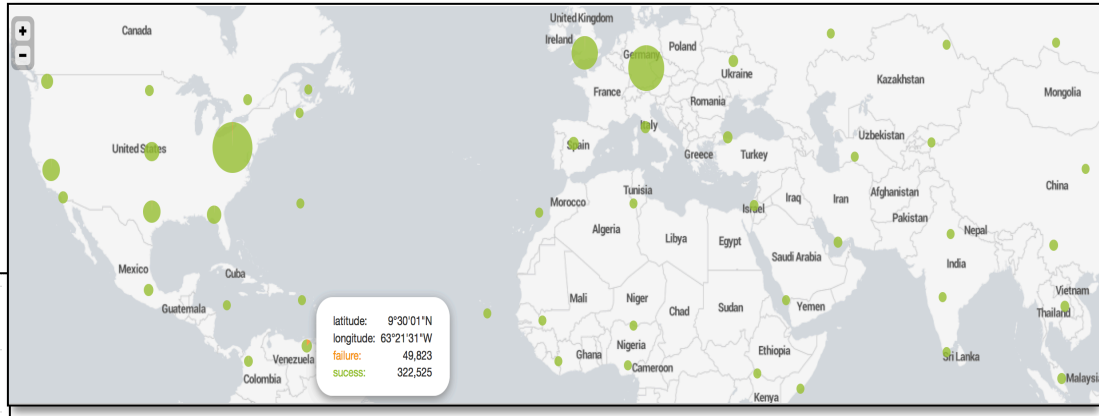
- Real Time Availability & Performance
- Origin Insights
- Edge Overview
- Security
- Release Monitoring

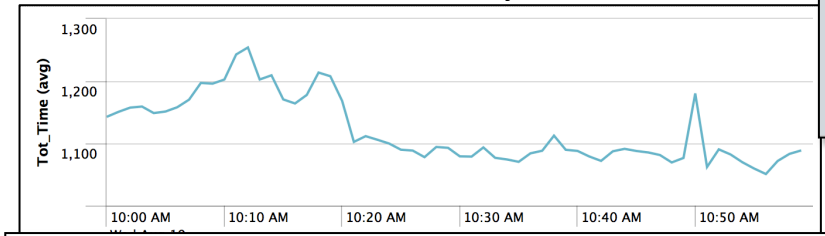


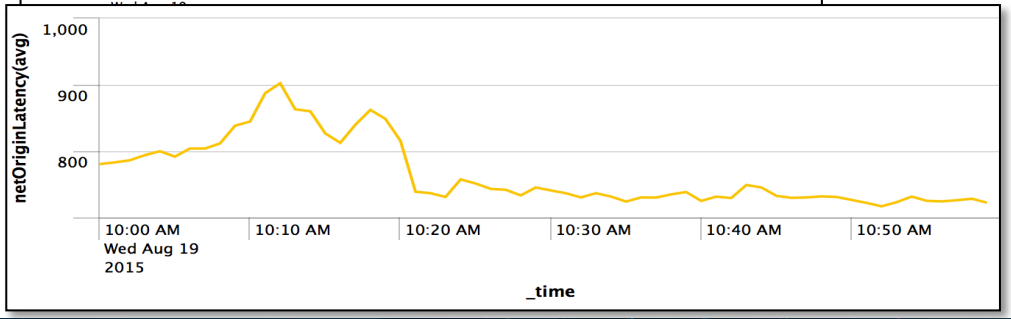
Availability & Performance

Availability

- Volume of Hits.
- Success & Failure Volume by HTTP Status.
- Success & Failure by GEO







Performance

- Track total time for transactions.
- Origin Latency
- Last Mile Round Trip Time

Origin & Edge

Origin

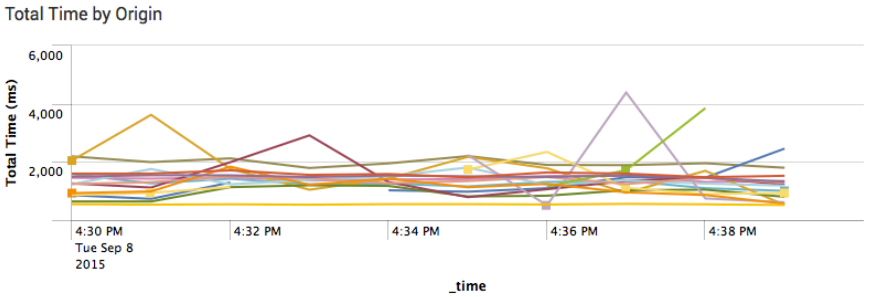
- Monitor traffic distribution.
- Error Monitoring.
- Monitor Latency.

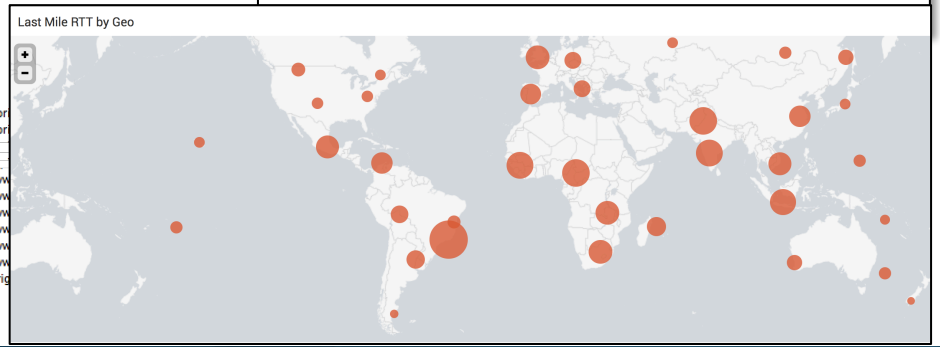
Edge

- Edge Errors
- Edge Performance stats
- Last Mile RTT

HTTP Errors by Edge Server — 17m ago

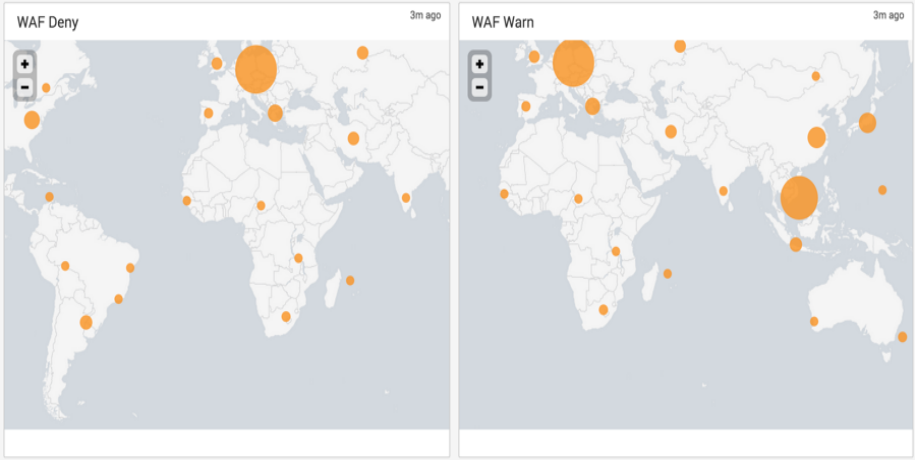
edgeIP	Count of Non waf
187.59.4.147	3389
23.74.2.7	2917
23.74.2.12	2625
202.7.177.76	1424
23.74.2.15	663
125.252.224.31	414
23.74.2.29	342
23.74.2.22	333
72.246.43.217	314

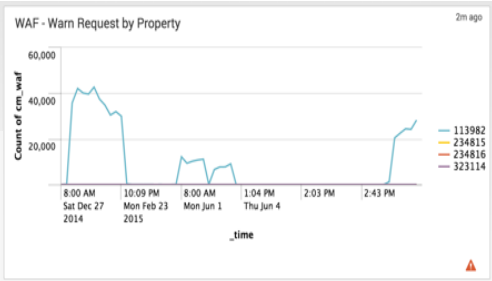


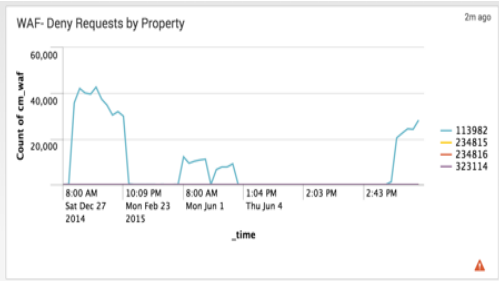


Security

- Monitor the WAF denies and warnings.
- Monitor Top Deny Rules.
- Report on the WAF warning triggered.
- Top deny & warning URL's.
- Deny & warn tracking by Geo
- Top Denied Client IP's







Top Warnings Rules — 2m ago

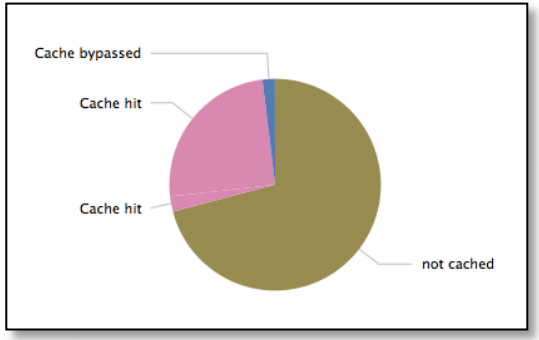
	warnRule	warn_count
1	950109	2742
2	981173	2742
3	INVALID-HTTP-ANOMALY	2742
4	615379	1961
5	950005	1039
6	950010	973
7	981173	973
8	950109	285
9	981173	285
10	981318	285

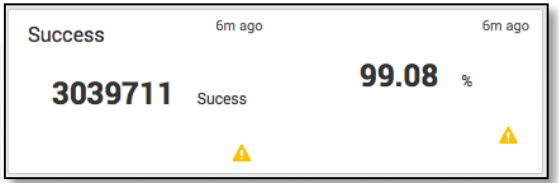
Top Deny Rules — 2m ago

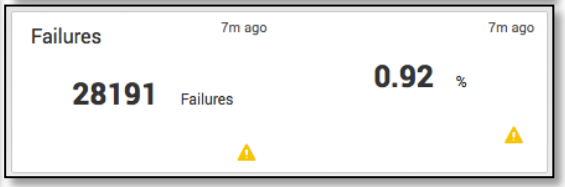
	denyRule	deny_count
1	618746	319033
2	618899	40653
3	IPBLOCK	14461
4	IPBLOCK-SUMMARY8-6581716-3	12488
5	IPBLOCK-BURST4-6342054-3	9226
6	IPBLOCK-BURST4-6581716-3	8761
7	IPBLOCK-SUMMARY8-6581709-3	7285
8	IPBLOCK-BURST4-6498226-3	5167
9	IPBLOCK-BURST4-6342042-3	4060
10	IPBLOCK-SUMMARY8-6498232-3	3677

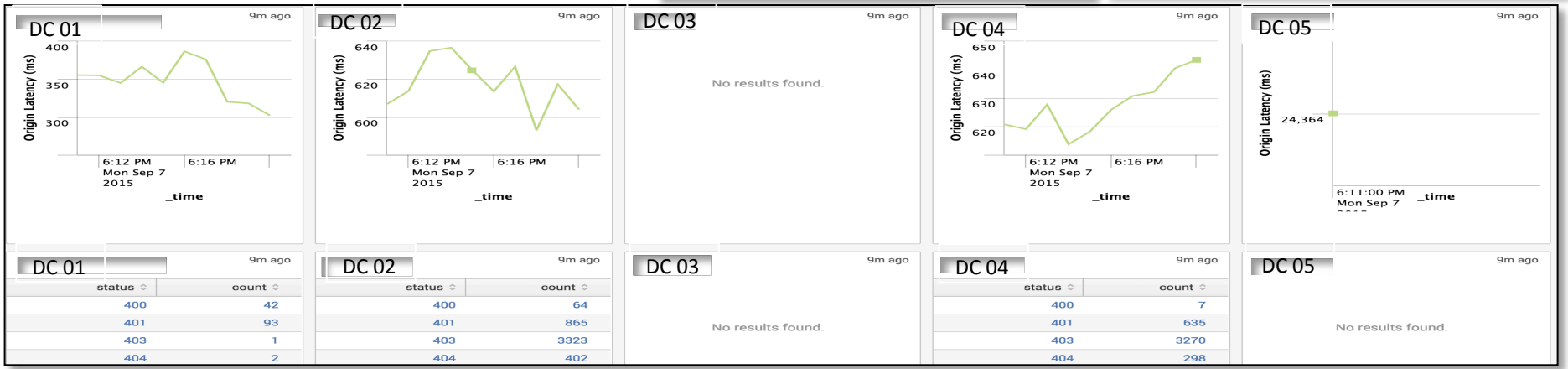
Release Monitoring

- Monitor Traffic by Origin / DC
- Monitor Issue's & Performance by Property
- Performance by Origin
- Last Mile Time by Edge
- Performance by GEO









Final Analysis and Q & A

splunk>

Final analysis

- The initial business needs for Real Time logging have been met:
  - Real time monitoring enables us to know the current state of the site
  - Dashboards allow us to track the effect of site changes
  - JSON data formatting makes it easier to do analytics on the customer experience
- The additional benefits are extremely valuable:
  - Custom fields to help troubleshoot site and code issues
  - Header information providing customer experience analytics data
  - Rich data set about customer crypto to help with SHA256 migration
  - Performance data on a regional and network level
- Continuing to find new ways to leverage the CM data in Splunk
- Q & A

THANK YOU

splunk>