splunk> .conf2017

Integrating Splunk And AWS Lambda

Big Results @ Fast-Food Prices

Gary Mikula | Senior Director, Cyber & Information Security

Siddhartha Dadana | Lead Security Engineer

Kuljeet Singh | Lead Security Engineer

September 26, 2017 | Washington, DC

Forward-Looking Statements

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 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.

Who We Are And How We Got Here

FINRA's Roadmap with Splunk and AWS

We Are FINRA

Financial Industry Regulatory Authority

- An independent, non-governmental regulator for all securities firms doing business with the public in the United States
- FINRA protects investors by regulating brokers (641,000) and brokerage firms (3,900) and by monitoring trading on U.S. stock markets
- FINRA monitor over 6 billion shares traded on the stock market each day which translates up to 75 billion transactions analyzed per day
- That more than 20 TIMES the number of VISA charges (29M), 🐦 tweets (0.5B), and f likes and updates (2.7B) per day……combined
- FINRA handles more ‘Big Data’ on a daily basis than the size of the Library of Congress — to build a holistic picture of the trading market

Journey To AWS

Technology meets Necessity

- On-Premise Data Warehouse Solutions
  - Serviceable but Not Scalable
- Intense Proof of Concept (2014)
  - Moved 90% of our Data Volumes & Core Market Surveillance Applications
- Announced Plans to go All In (2015)
- Four Pillars
  - Self Sufficiency
  - Public over Private/Community (Moore's Law)
  - Open Source First
  - No Lift and Shift (DevOps Automation and Security Protection)

Journey With Splunk

Making the Most of the Investment

- Traditional SIEM Vendor Announced Tech-Refresh (2012)
- One of the First Large SplunkCloud Customer (2013)
- 60% Data Intake Increase
- Over 25% of Technology Visits Splunk Every Week
- Mission Critical Tier 2 Application
  - Operations/Security/Development
- Socialization is the Key to ROI
  - Bimonthly Brown Bags (10% of Technology Attends)
  - Find Stewards and Help Them to Grow
  - Democratize The Asset – Become a Data Driven Organization

Security Engineering

Cloud Equals Impact

- ~20 Member Staff of Skilled Engineers with diverse experience
- Build, implement, and maintain controls and analytics to identify, manage, and mitigate threats, risks, and vulnerabilities
- Some Key Responsibilities:
  - Security Compliance
  - Identity and Access Management
    - Administrative Access
  - Security Information and Event Management
  - Insider Risk Technical Controls
- How/Why We Use AWS Lambda with Splunk to Meet These Challenges

Splunk & AWS Lambda

A Developer’s View

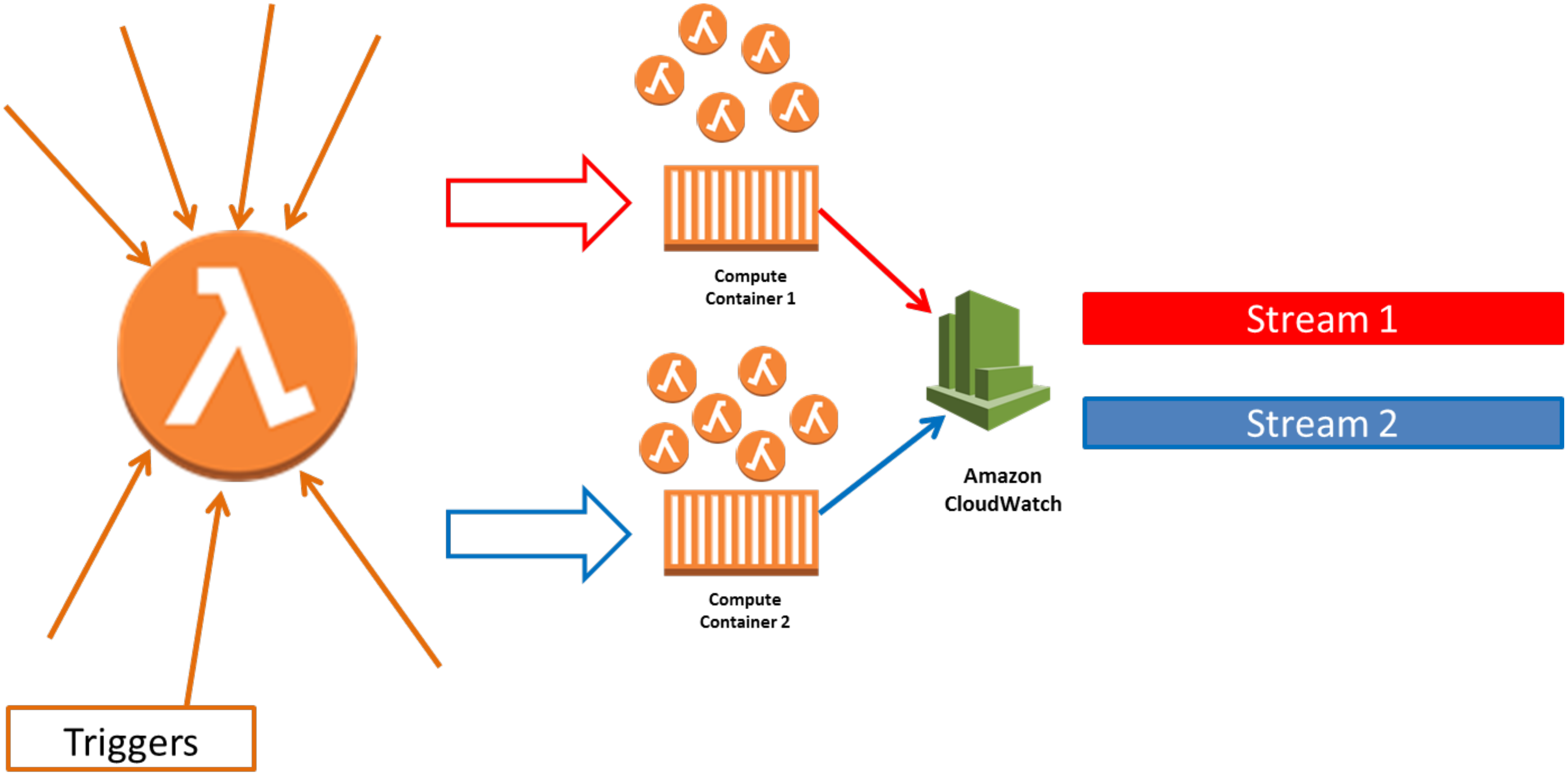
Why “Server-less” Computing

Why FaaS is So Attractive

- Run Your Code on Someone Else’s Computer
- No Infrastructure Worries
  - No Administrators….That You Can See
  - No Patching
  - No Disaster Recovery
- Pay Only for What You Use
  - My Job Only Needs to Run When X Happens
    - What If X Happens, Once a Day/Week/Month, But You Don’t Know When
  - What If When Y Happens, 10,000 X’s Happen?
- Comparatively, AWS Lambda is Quite Affordable

AWS Lambda Native Logging

Where the Fun Begins



Find The Long Running Process

A Test of Perseverance

Three Infrastructure Elements

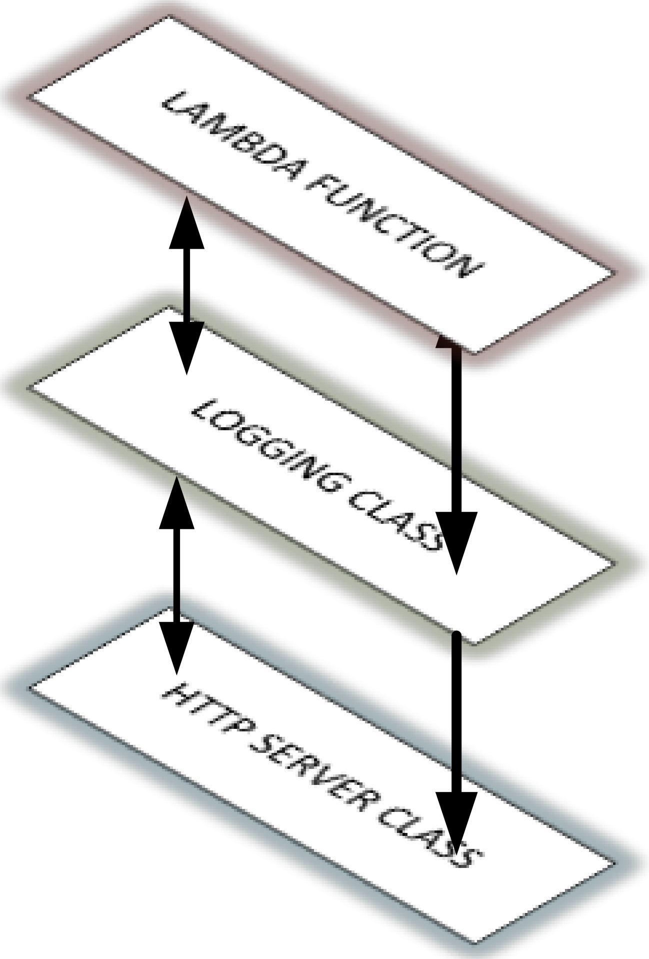
To Empowering Your Development

- Bullet-Proof, Metric-Based, Auto-Scalable, Splunk HTTP Event Collection Service
- Logging Standards
- Enterprise Class Design

Creating An Enterprise Class

Anything Worth Doing…..

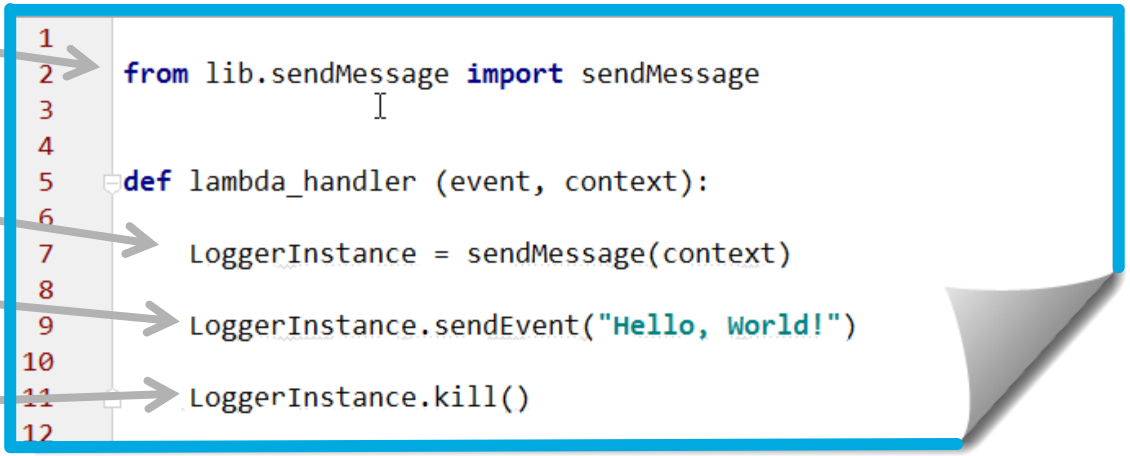
- Lambda Function
  - Zips the Classes into Deployment Package
  - Invokes the Logging Class
- Logging Class
  - Enforces Your Logging Standards
  - Enforces Splunk Keys:Index/Host/Source/Sourcetype
  - Handles HTTP Error Processing
- HTTP Server Class
  - Encapsulates Details of Splunk HEC Interaction
  - Responsible for Reliable Delivery of Log Messages



Making It EASY For Your Developers

Key to Acceptance

- Import the ‘LOGGING’ Class
- Instantiate the Class
- Send an Event
  - Default Severity
- Destroy the Class
  - VERY Important in AWS Lambda



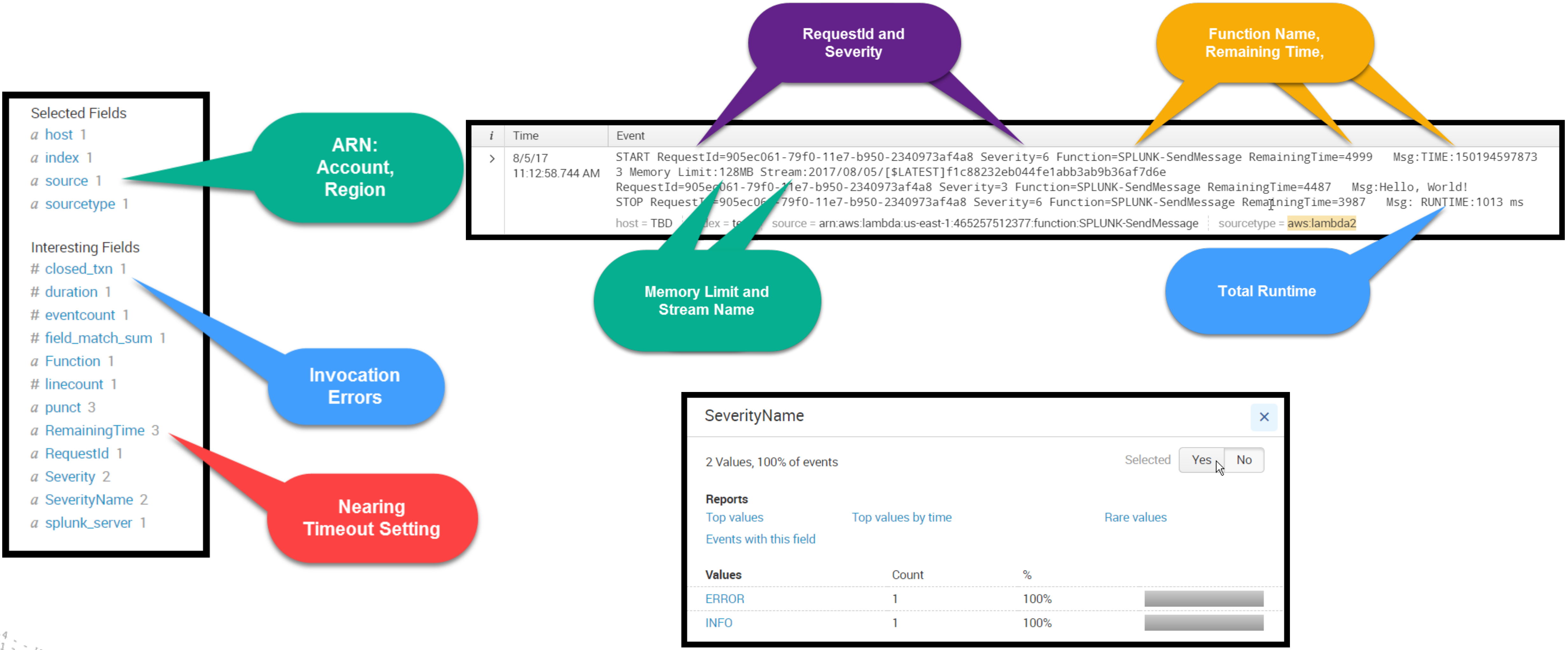
```

from lib.sendMessage import sendMessage


def lambda_handler (event, context):

    LoggerInstance = sendMessage(context)

    LoggerInstance.sendEvent("Hello, World!")

    LoggerInstance.kill()

```

- What’s in it for YOU?

splunk> .conf2017

One Simple Query

Function="Splunk-SendMesage"| transaction RequestId startswith=START endswith=STOP keepevicted=1



Selected Fields
- a host 1
- a index 1
- a source 1
- a sourcetype 1

Interesting Fields
- # closed_txn 1
- # duration 1
- # eventcount 1
- # field_match_sum 1
- a Function 1
- # linecount 1
- a punct 3
- a RemainingTime 3
- a RequestId 1
- a Severity 2
- a SeverityName 2
- a splunk_server 1

i	Time	Event

>	8/5/17
	11:12:58.744 AM
		START RequestId=905ec061-79f0-11e7-b950-2340973af4a8 Severity=6 Function=SPLUNK-SendMessage RemainingTime=4999 Msg:TIME:1501945978733 Memory Limit:128MB Stream:2017/08/05/[$LATEST]f1c88232eb044fe1abb3ab9b36af7d6e
		RequestId=905ec061-79f0-11e7-b950-2340973af4a8 Severity=3 Function=SPLUNK-SendMessage RemainingTime=4487 Msg:Hello, World!
		STOP RequestId=905ec061-79f0-11e7-b950-2340973af4a8 Severity=6 Function=SPLUNK-SendMessage RemainingTime=3987 Msg: RUNTIME:1013 ms
		host = TBD | index = te | source = arn:aws:lambda:us-east-1:465257512377:function:SPLUNK-SendMessage | sourcetype = aws:lambda2

SeverityName ✕

2 Values, 100% of events — Selected: Yes No

Reports

Top values — Top values by time — Rare values

Events with this field

Values	Count	%
ERROR	1	100%
INFO	1	100%

ARN: Account, Region

Invocation Errors

Nearing Timeout Setting

RequestId and Severity

Function Name, Remaining Time,

Memory Limit and Stream Name

Total Runtime

splunk> .conf2017

Find The Long Running Process

Payback for the Hard Work

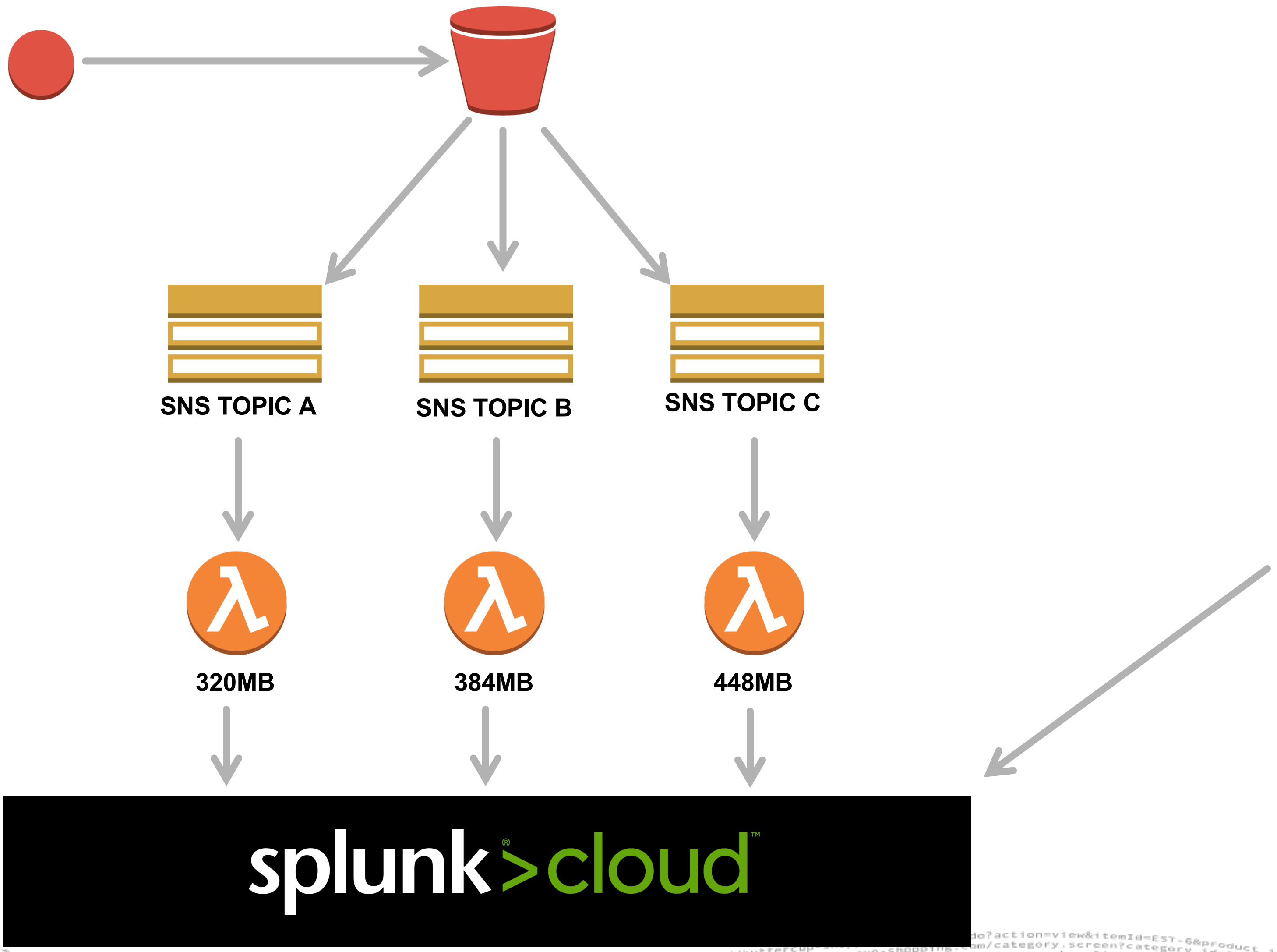
Other Useful Commands

Information at Your Fingertips

- *Function=“Splunk-SendMesage"| transaction RequestId startswith=START endswith=STOP keepevicted=1 | search closed_txn=0*
  - All Lambda runs that haven’t completed……gracefully
- *Function=“Splunk-SendMesage"| transaction RequestId startswith=START endswith=STOP | mvexpand Severity | Severity <= 3*
  - All Lambda runs that produced ERROR/CRITICAL/ALERT/EMERGENCY messages
- *Function=“Splunk-SendMesage"| transaction RequestId startswith=START endswith=STOP keepevicted=1 | search RequestId > 1*
  - All Lambda runs that had automatic restarts
- *Function=“Splunk-SendMesage"| transaction RequestId startswith=START endswith=STOP | stats count by stream*
  - Number of Lambda runs inside of each container
- *Function=“Splunk-SendMessage” RequestId=xx-xx-xx | reverse | delta Time_ms AS DeltaTime*
  - *Show each log line in Chronological Order listing the time each previous step ran*

Blueprint For Optimizing Costs

Facts Beat Guessing… Every Time

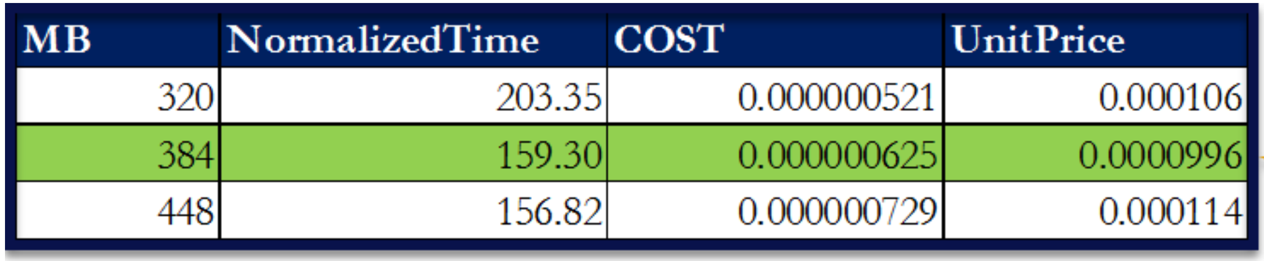


MB	COST
128	0.000000208
192	0.000000313
256	0.000000417
320	0.000000521
384	0.000000625
448	0.000000729
512	0.000000834
576	0.000000938
640	0.000001042
704	0.000001146
768	0.00000125
832	0.000001354
896	0.000001459
960	0.000001563
1024	0.000001667
1088	0.000001771
1152	0.000001875
1216	0.00000198
1280	0.000002084
1344	0.000002188
1408	0.000002292
1472	0.000002396
1536	0.000002501

Analytic Efficiency Equal Cost Savings

May I Have the Envelope Please

- Function=“Splunk-SendMesage"|
  - Transaction RequestId startswith=START endswith=STOP |
  - Rename MemoryLimit AS MB |
  - Stats avg(RunTime) AS NormalizedTime by MB |
  - Lookup LambdaPricing.csv MemoryLimit |
  - Eval UnitPrice=NormalizedTime*COST



MB	NormalizedTime	COST	UnitPrice
320	203.35	0.000000521	0.000106
384	159.30	0.000000625	0.0000996
448	156.82	0.000000729	0.000114

6.42% Cost Savings

Splunk & AWS Lambda

A Security Perspective

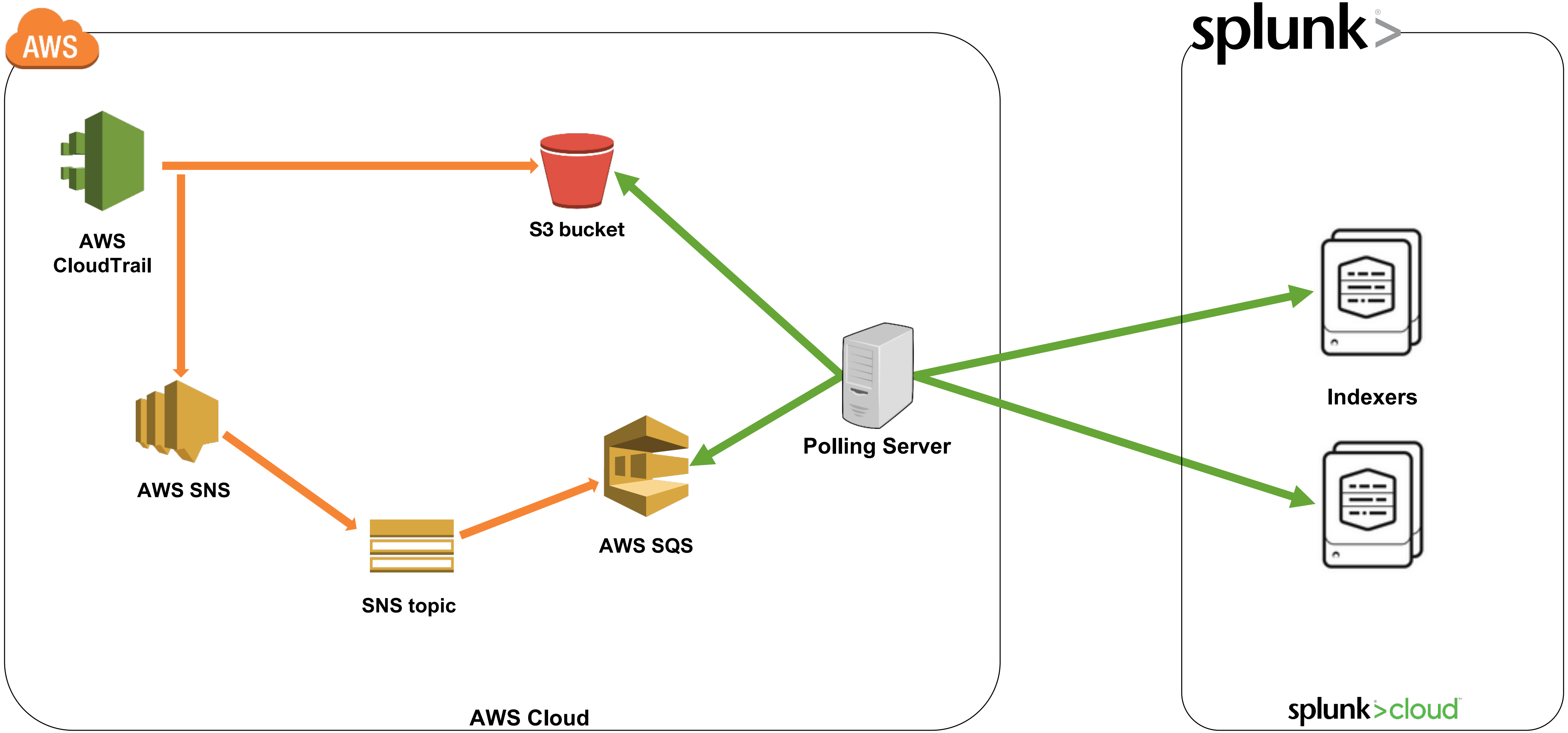
AWS Cloudtrail

What it is & Why you need it?

- Records Every Object Level API Call for your Account
- Is a Regional Service
  - Must be Configured for Each Account/Region pair
- Writes Log Files into an S3 Bucket
- Is required to
  - Perform Security Analysis
  - Detect User Behavior
  - Detect Data Exfiltration on S3 Objects
  - Troubleshoot Operational Issues & Track Resource Changes
  - Alert and Report

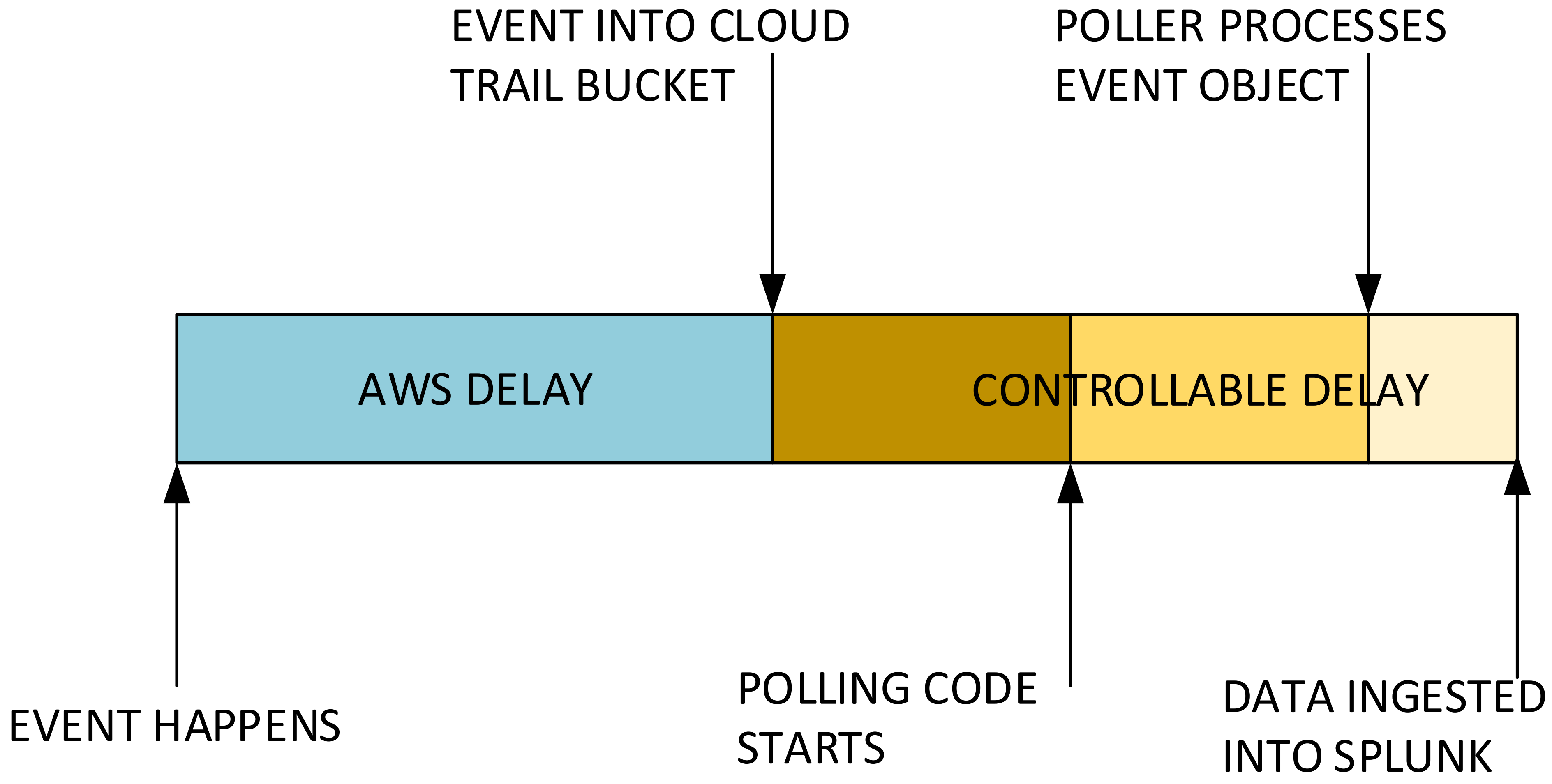
AWS Cloudtrail

Typical Log Collection



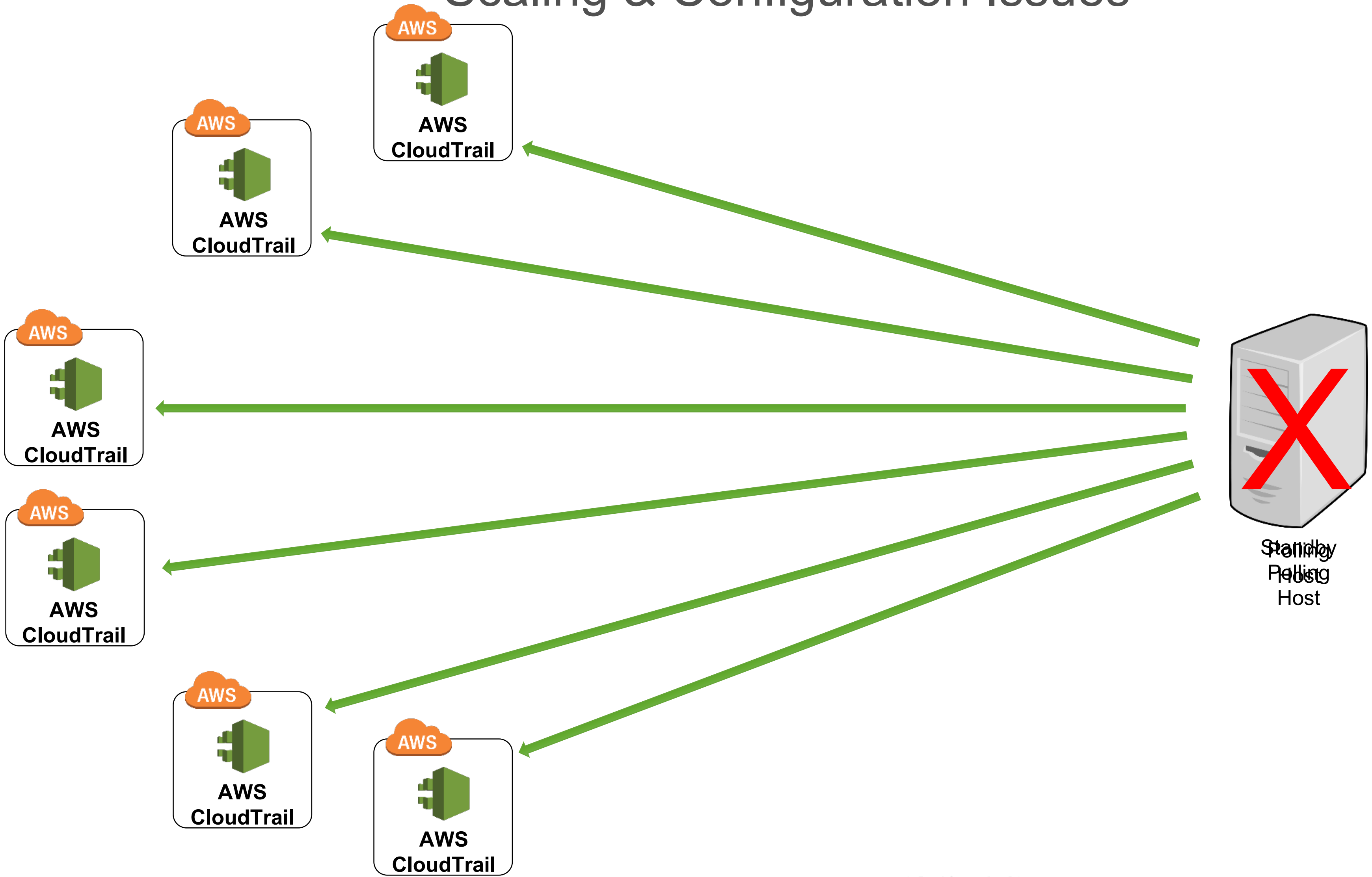
Problems with Cloudtrail Collection

Delay Issues


EVENT INTO CLOUD TRAIL BUCKET
POLLER PROCESSES EVENT OBJECT
AWS DELAY
CONTROLLABLE DELAY
EVENT HAPPENS
POLLING CODE STARTS
DATA INGESTED INTO SPLUNK

Problems With Cloudtrail Collection

Scaling & Configuration Issues



Concurrency Issues

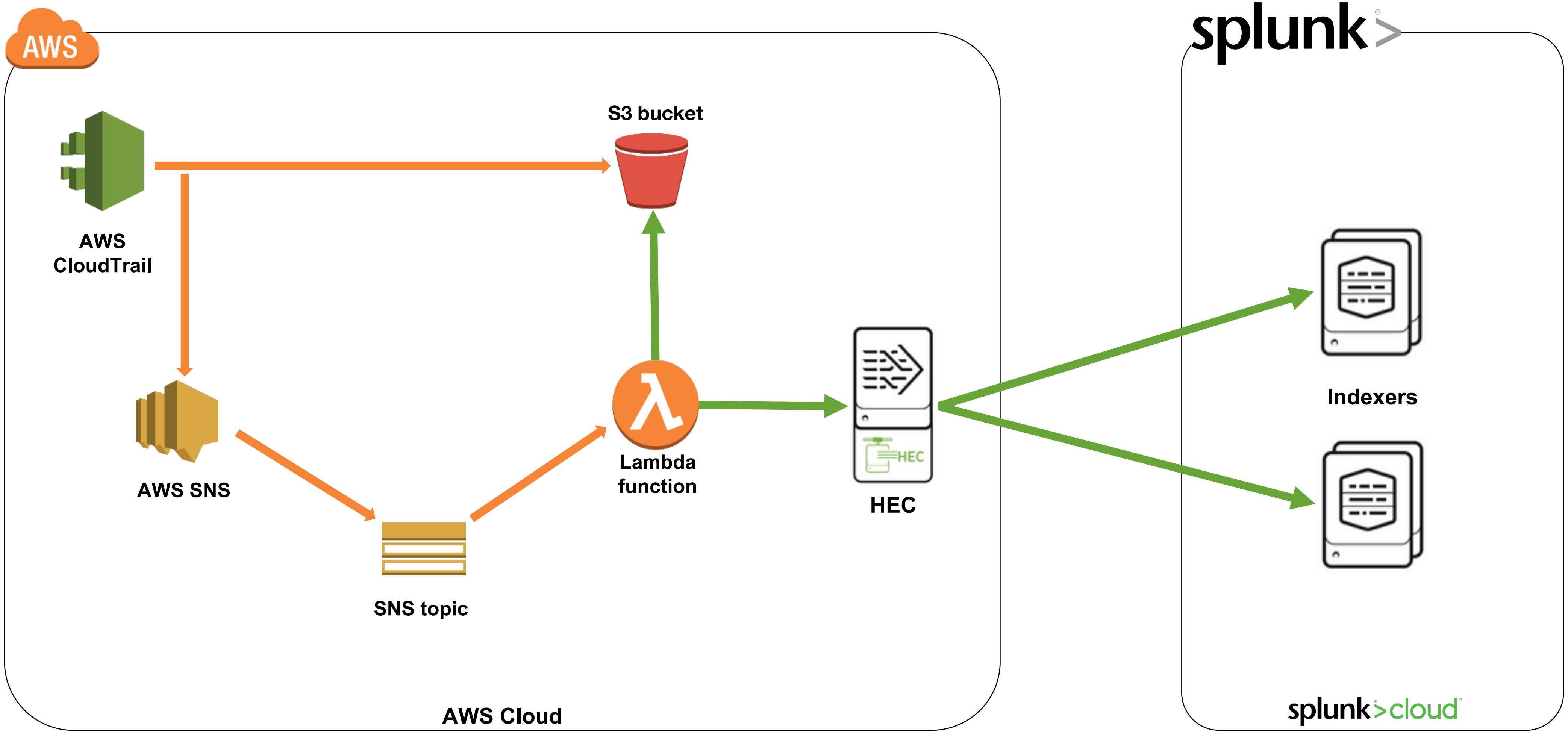
Can’t Lock SQS Messages

- Manually Distribute Load Across Multiple Polling Servers
  - Configuration Maintenance
  - Doesn’t Ensure Load Distribution
  - Manual DR Processes
  - Lots of Idle Time
- Buy a Bigger Polling Server
  - Large Enough to Handle Peak Load, Whenever That May Be
  - Manual DR Process

- With Polling, Each Collector Has to Know What the Other Collector Is Doing

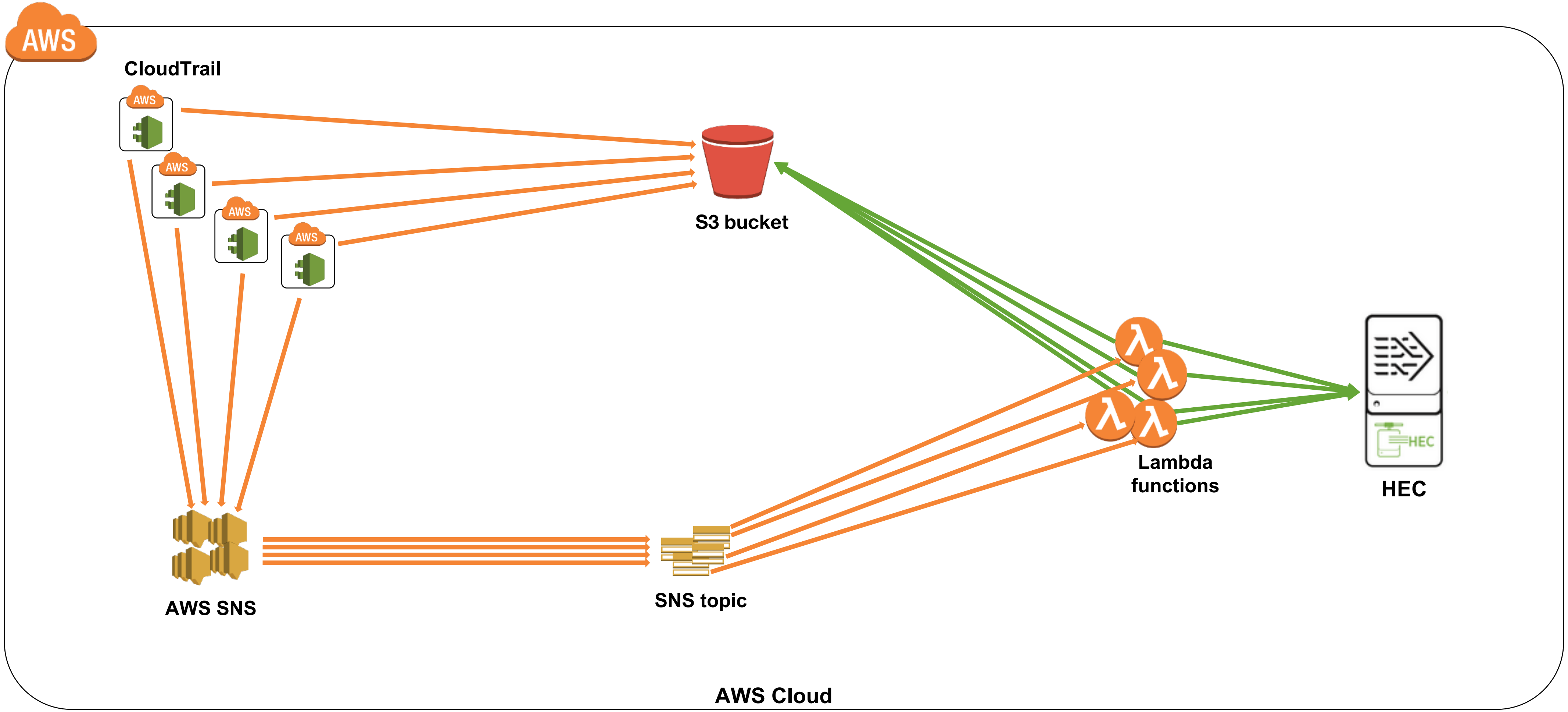
AWS Cloudtrail

How do We Solve Delay?



AWS Cloudtrail

How do We Solve Scaling & Concurrency Issues?



AWS Cloudtrail

Were We Successful?

- On Average, we get the Events to Splunk in 2 seconds
- Zero Server Maintenance
- Zero Polling Server Configuration Maintenance
- NO Manual Fail-Over
  - If we lost all 4 US-EAST-1 regions, make like Horace Greeley
- NO Keys to Maintain
- It scales, 1 Object = 1 Lambda Invocation
  - No Concurrency Issues
- Splunk AWS App still Works!!!

AWS Cloudtrail

Are We Cost Effective?

$$\frac{\$3.99}{\text{1 Big Mac}} \times \frac{5.75\%}{\text{D.C. Sales Tax}} = \$4.22$$

$$\frac{\text{415,000 objects}}{\text{Month}} \times \frac{\text{2300 ms}}{\text{Lambda Run}} \times \frac{\$\ 0.000000417}{\text{100 ms}}$$

$$= \$3.98/\text{Month}$$

$1 runs 104,264 λ functions

splunk> .conf2017

Other Search Methods

A brief look at Other Collection & Search Methods

AWS Athena

- Un/Semi/Structured Data
- S3 Objects as Data Feed
- Database Tables
- Limited Data Formats
- Enrichment of Data
- Reporting & Alerting
- Pay per Search

AWS CloudSearch

- Structured Data
- Manual/Scripted Upload
- JSON/XML
- Enriching Data
- Pay Hourly per Instance

Manual

- Download Files
- Unzip and Analyze
- Difficult
- Not Cost Effective

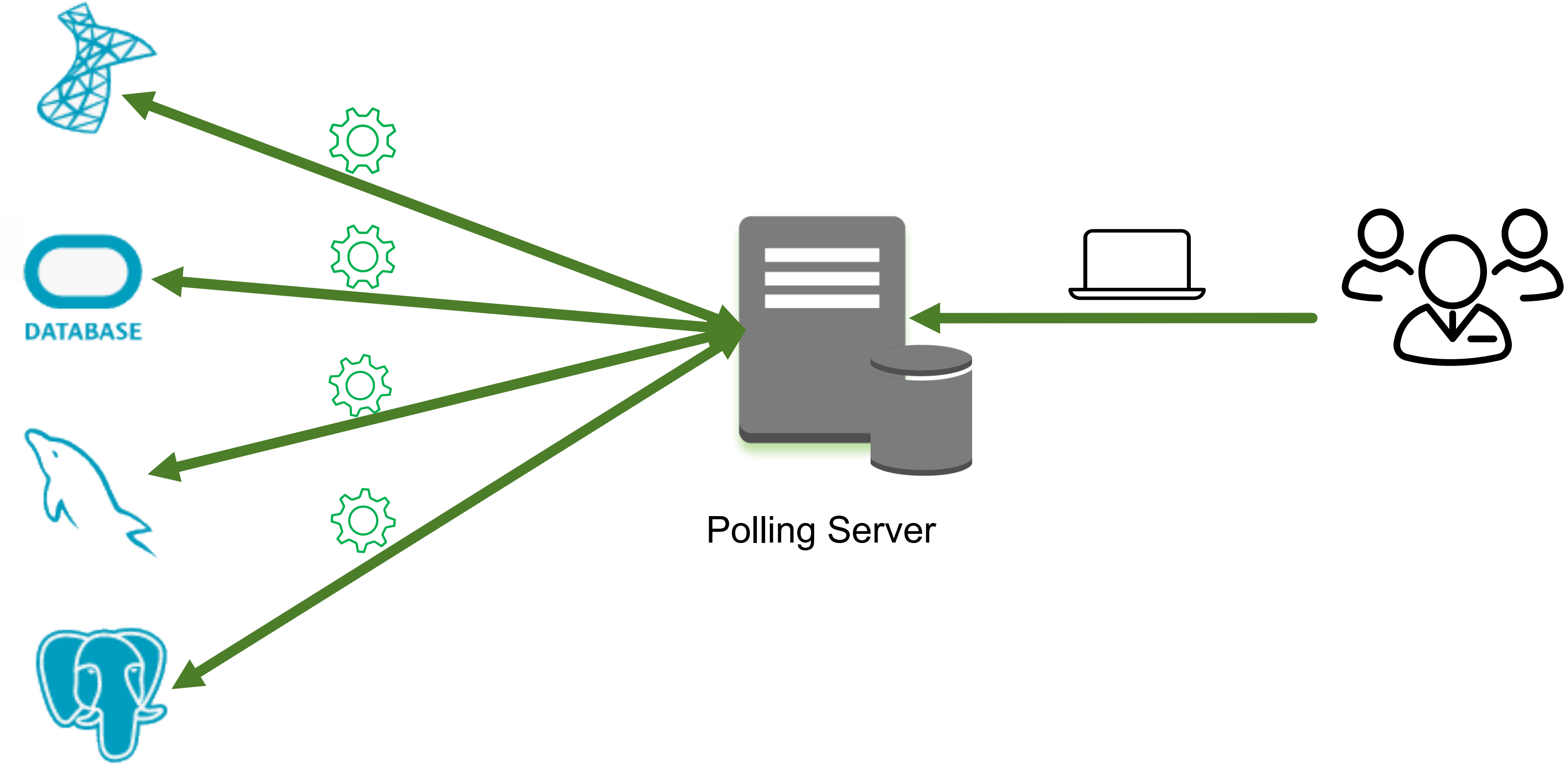
Solving DevOps & Compliance Issues

A DevOps View

What New Hurdles Does Cloud Bring?

- Rapid Deployments
  - From Days To Mins
- Systems Are Transient
  - Monthly Compliance is Woefully Outdated
  - Some Stacks Have Been Re-Built
  - Vendors Have Been Slow To Transition Their Products
- Security Has To Adopt DevOps Automation
  - Security Teams Are Not Traditionally Coders
- DevOps Has To Include Security ‘IN’ the Build
  - Traditionally Added-On
- And This Is Where Splunk & Lambda Come In

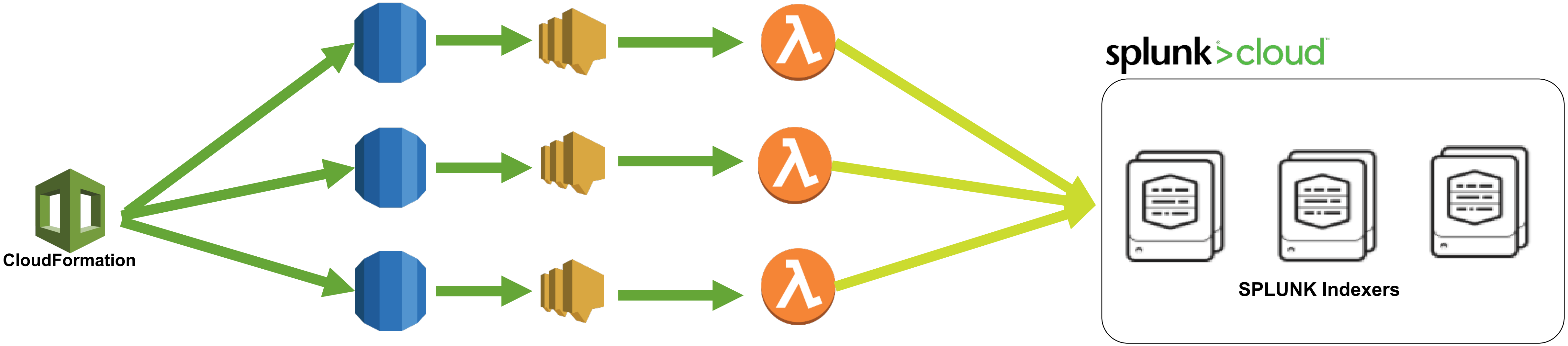
Compliance - Traditional Method



Issues With Traditional Method In Cloud

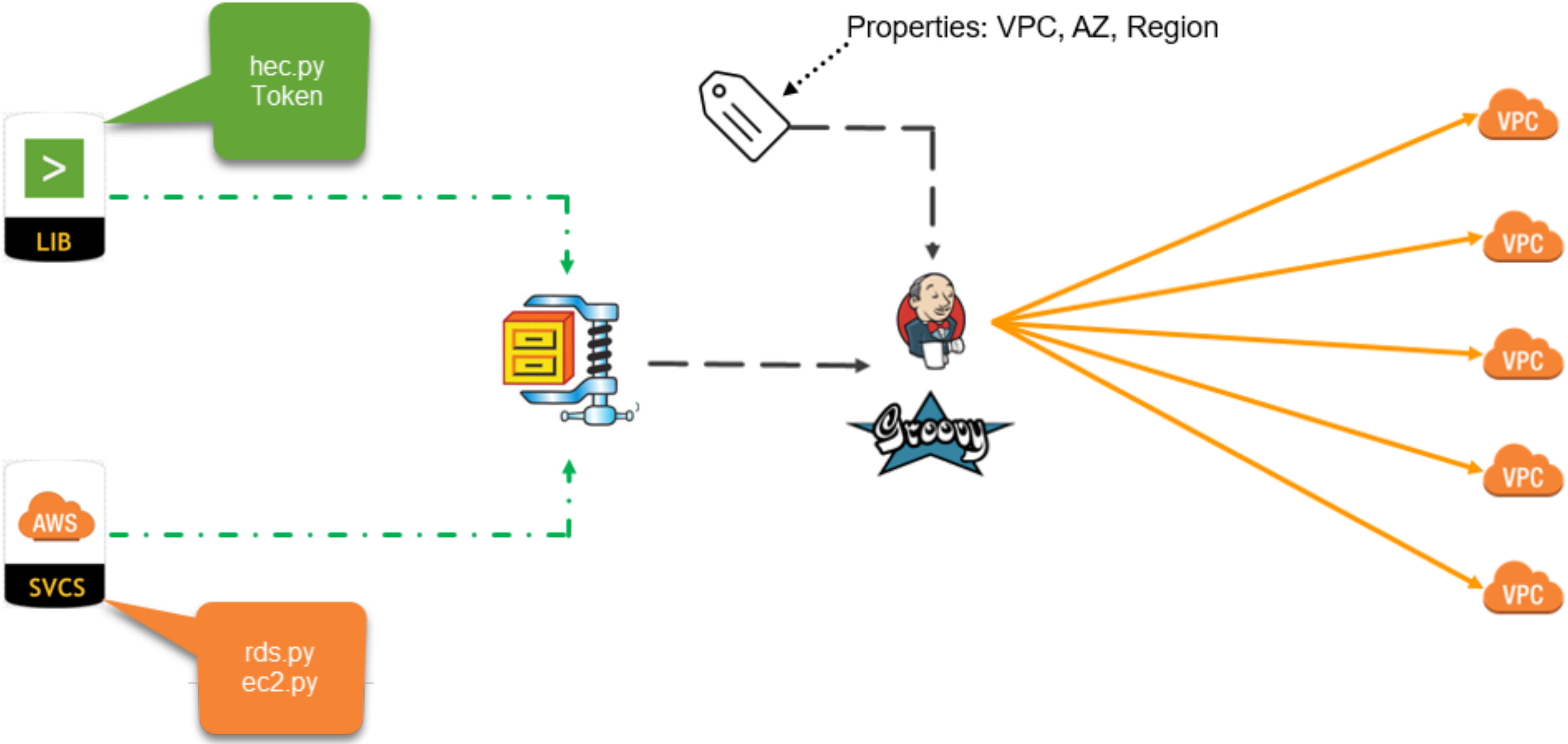
- Collection Scalability
  - Buy A Bigger Polling Device
  - What If 5K Systems Start? 50K? 500K?
- Configuration Scalability
  - Need to Manually Provision Each New VPC
- How Often to Poll
  - Delays in Collection
  - Transient Systems are Missed
- Delay In Collecting Data
  - How Often to Poll
- Relies On Access Keys

Compliance – Using Lambda



- No Access Keys To Manage
- Event Triggered On Every Change
- Near RealTime Data
- No Scaling Issues
- No Provisioning Of Servers
- Any Number Of Accounts, Just A Code Drop Away

How Do We Get There?



hec.py
Token

LIB

AWS

SVCS

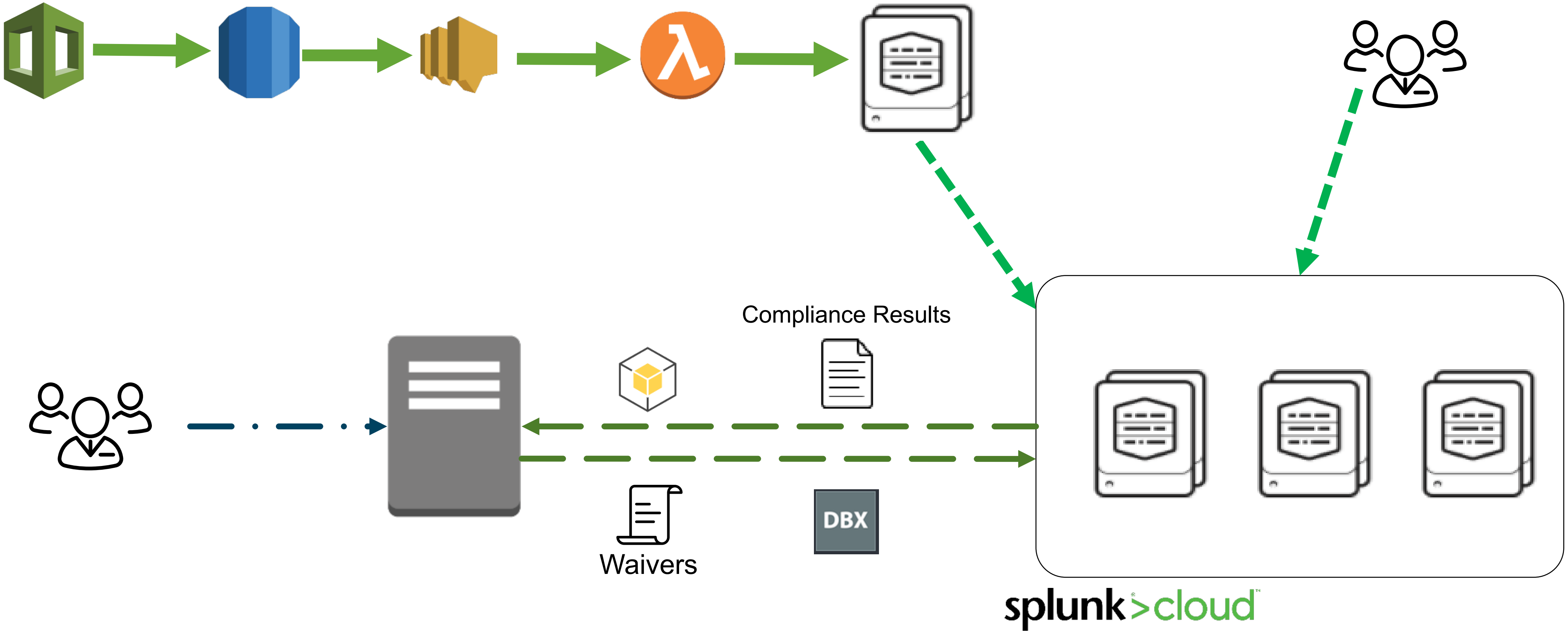
rds.py
ec2.py

Properties: VPC, AZ, Region

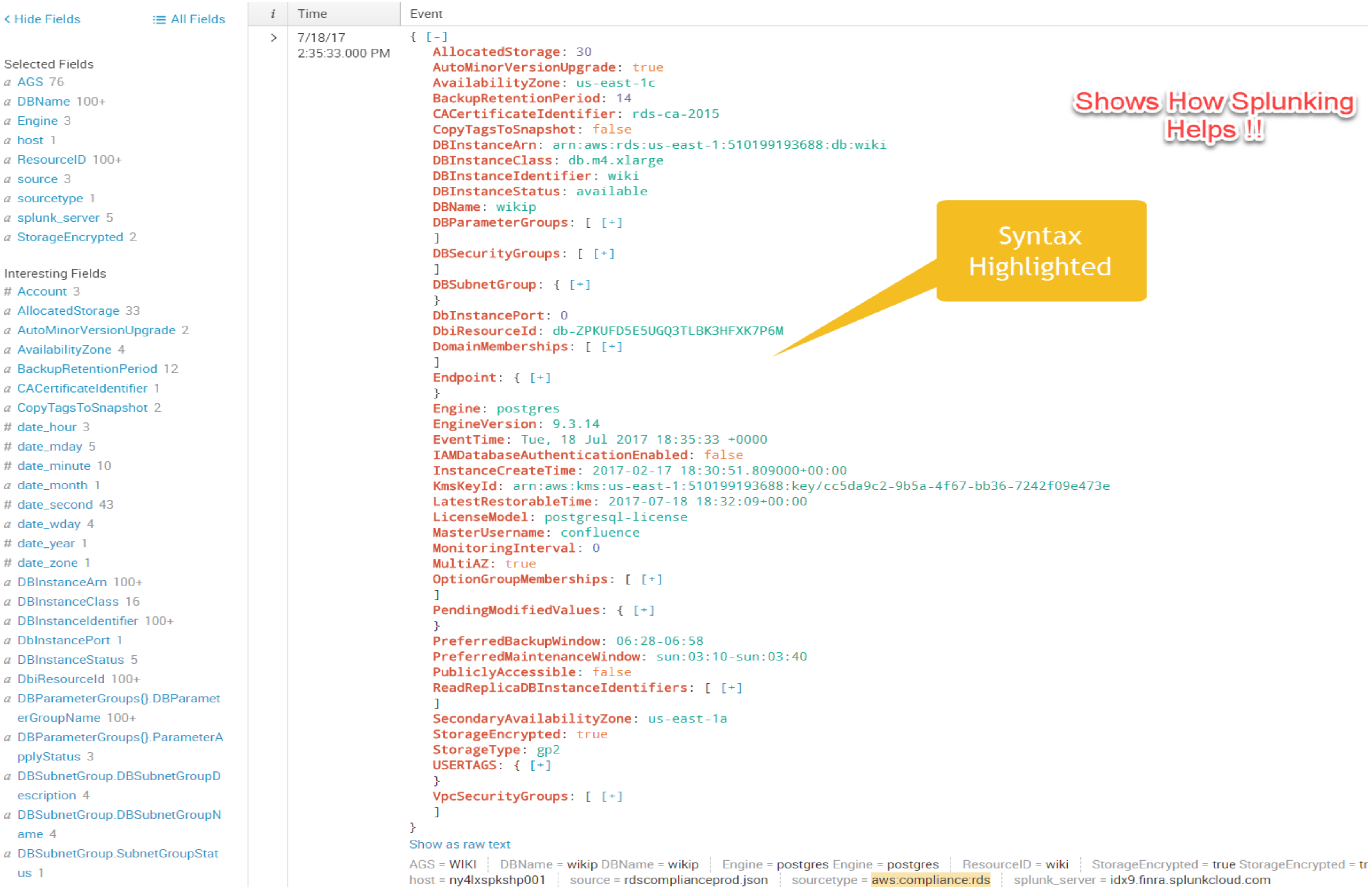
What Additional Data Do We Need?

- Published Standards for AWS Services
  - Define Clear & Specific Checks
  - Include DevOps Early In The Process
  - Try To Cover Major Services First
- Waiver Program
  - Robust & Flexible Waiver Management
    - Reusable Schema Across Services
  - Clear Understanding For APP Teams
    - Waiver Filing & Approval Process
- Integration With UI Platform - SPLUNK
  - One Screen For All Data
  - Enhances User Experience & Enables Faster Adoption
    - Goal Is To Provide Greater Visibility For App Teams

How It Works!!



Magic Of Splunk!



Selected Fields
- *a* AGS 76
- *a* DBName 100+
- *a* Engine 3
- *a* host 1
- *a* ResourceID 100+
- *a* source 3
- *a* sourcetype 1
- *a* splunk_server 5
- *a* StorageEncrypted 2

Interesting Fields
- # Account 3
- *a* AllocatedStorage 33
- *a* AutoMinorVersionUpgrade 2
- *a* AvailabilityZone 4
- *a* BackupRetentionPeriod 12
- *a* CACertificateIdentifier 1
- *a* CopyTagsToSnapshot 2
- # date_hour 3
- # date_mday 5
- # date_minute 10
- *a* date_month 1
- # date_second 43
- *a* date_wday 4
- # date_year 1
- # date_zone 1
- *a* DBInstanceArn 100+
- *a* DBInstanceClass 16
- *a* DBInstanceIdentifier 100+
- *a* DbInstancePort 1
- *a* DBInstanceStatus 5
- *a* DbiResourceId 100+
- *a* DBParameterGroups{}.DBParameterGroupName 100+
- *a* DBParameterGroups{}.ParameterApplyStatus 3
- *a* DBSubnetGroup.DBSubnetGroupDescription 4
- *a* DBSubnetGroup.DBSubnetGroupName 4
- *a* DBSubnetGroup.SubnetGroupStatus 1

> 7/18/17
2:35:33.000 PM

```
{ [-]
   AllocatedStorage: 30
   AutoMinorVersionUpgrade: true
   AvailabilityZone: us-east-1c
   BackupRetentionPeriod: 14
   CACertificateIdentifier: rds-ca-2015
   CopyTagsToSnapshot: false
   DBInstanceArn: arn:aws:rds:us-east-1:510199193688:db:wiki
   DBInstanceClass: db.m4.xlarge
   DBInstanceIdentifier: wiki
   DBInstanceStatus: available
   DBName: wikip
   DBParameterGroups: [ [+]
   ]
   DBSecurityGroups: [ [+]
   ]
   DBSubnetGroup: { [+]
   }
   DbInstancePort: 0
   DbiResourceId: db-ZPKUFD5E5UGQ3TLBK3HFXK7P6M
   DomainMemberships: [ [+]
   ]
   Endpoint: { [+]
   }
   Engine: postgres
   EngineVersion: 9.3.14
   EventTime: Tue, 18 Jul 2017 18:35:33 +0000
   IAMDatabaseAuthenticationEnabled: false
   InstanceCreateTime: 2017-02-17 18:30:51.809000+00:00
   KmsKeyId: arn:aws:kms:us-east-1:510199193688:key/cc5da9c2-9b5a-4f67-bb36-7242f09e473e
   LatestRestorableTime: 2017-07-18 18:32:09+00:00
   LicenseModel: postgresql-license
   MasterUsername: confluence
   MonitoringInterval: 0
   MultiAZ: true
   OptionGroupMemberships: [ [+]
   ]
   PendingModifiedValues: { [+]
   }
   PreferredBackupWindow: 06:28-06:58
   PreferredMaintenanceWindow: sun:03:10-sun:03:40
   PubliclyAccessible: false
   ReadReplicaDBInstanceIdentifiers: [ [+]
   ]
   SecondaryAvailabilityZone: us-east-1a
   StorageEncrypted: true
   StorageType: gp2
   USERTAGS: { [+]
   }
   VpcSecurityGroups: [ [+]
   ]
}
```

Show as raw text

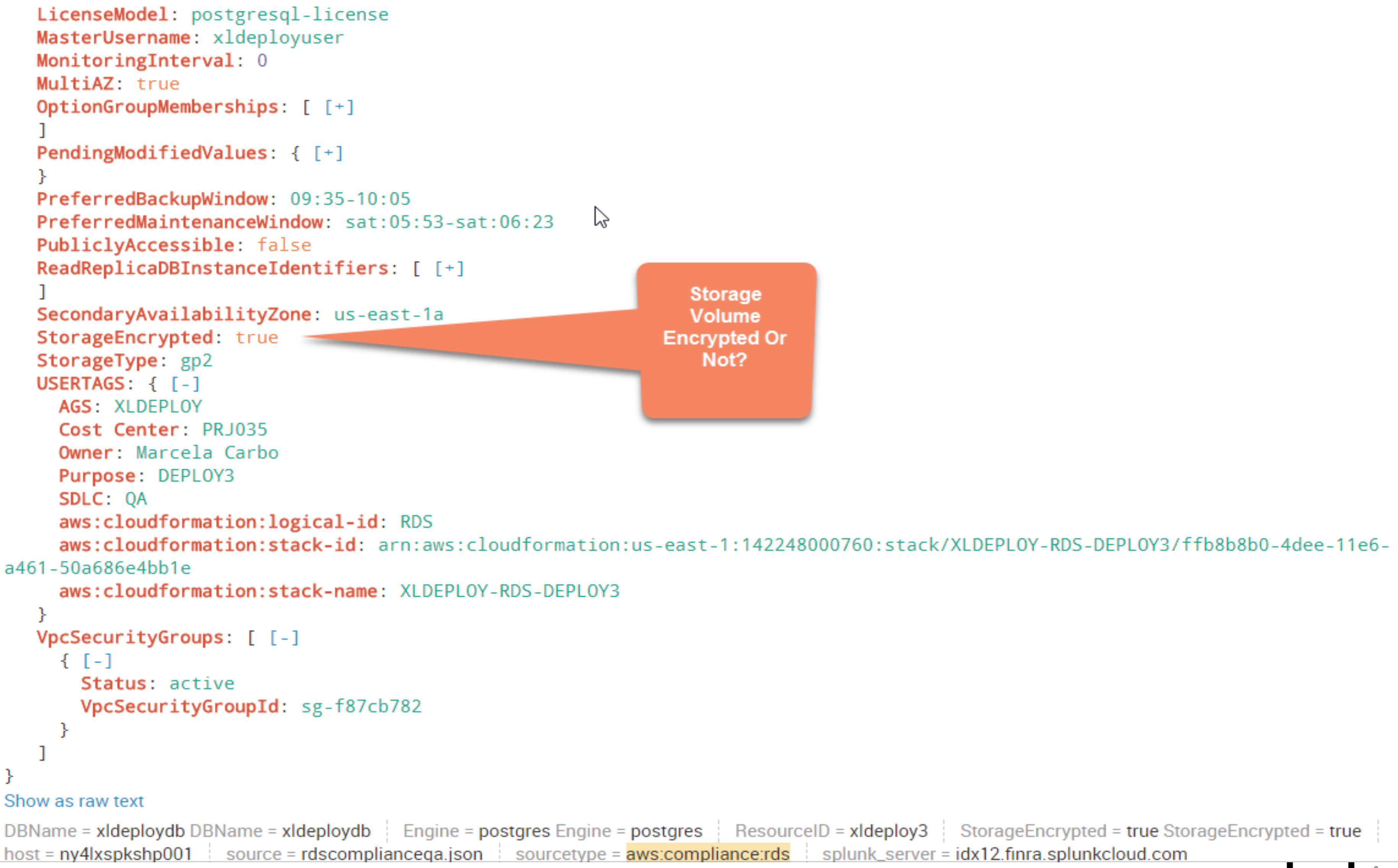
AGS = WIKI | DBName = wikip DBName = wikip | Engine = postgres Engine = postgres | ResourceID = wiki | StorageEncrypted = true StorageEncrypted = tr
host = ny4lxspkshp001 | source = rdscomplianceprod.json | sourcetype = aws:compliance:rds | splunk_server = idx9.finra.splunkcloud.com

Extracted All Fields Automatically

Syntax Highlighted

Shows How Splunking Helps !!

Security



```
   LicenseModel: postgresql-license
   MasterUsername: xldeployuser
   MonitoringInterval: 0
   MultiAZ: true
   OptionGroupMemberships: [ [+]
   ]
   PendingModifiedValues: { [+]
   }
   PreferredBackupWindow: 09:35-10:05
   PreferredMaintenanceWindow: sat:05:53-sat:06:23
   PubliclyAccessible: false
   ReadReplicaDBInstanceIdentifiers: [ [+]
   ]
   SecondaryAvailabilityZone: us-east-1a
   StorageEncrypted: true
   StorageType: gp2
   USERTAGS: { [-]
     AGS: XLDEPLOY
     Cost Center: PRJ035
     Owner: Marcela Carbo
     Purpose: DEPLOY3
     SDLC: QA
     aws:cloudformation:logical-id: RDS
     aws:cloudformation:stack-id: arn:aws:cloudformation:us-east-1:142248000760:stack/XLDEPLOY-RDS-DEPLOY3/ffb8b8b0-4dee-11e6-
a461-50a686e4bb1e
     aws:cloudformation:stack-name: XLDEPLOY-RDS-DEPLOY3
   }
   VpcSecurityGroups: [ [-]
     { [-]
       Status: active
       VpcSecurityGroupId: sg-f87cb782
     }
   ]
}
```

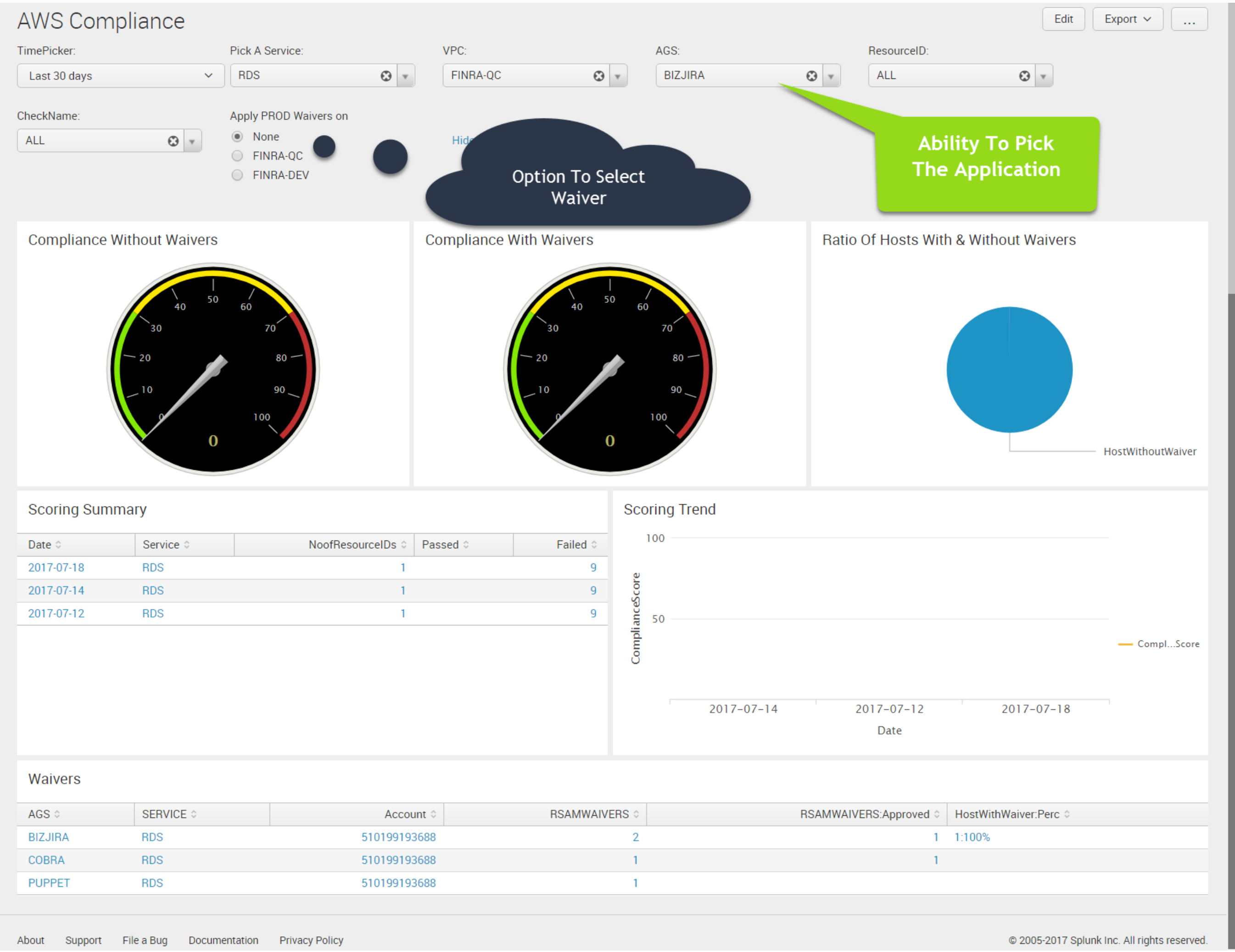
Show as raw text

DBName = xldeploydb DBName = xldeploydb | Engine = postgres Engine = postgres | ResourceID = xldeploy3 | StorageEncrypted = true StorageEncrypted = true
host = ny4lxspkshp001 | source = rdscomplianceqa.json | sourcetype = aws:compliance:rds | splunk_server = idx12.finra.splunkcloud.com

Storage Volume Encrypted Or Not?

Dashboard

- ▶ Calculate Compliance Score For Each Application
- ▶ Build A Simple Dashboard For Users
- ▶ See Near Real-Time Scoring After Deployment
- ▶ Apply Prod Waivers On Test Stacks To Know Their Standing After Production Deployment



Option To Select Waiver

Ability To Pick The Application

AWS Compliance

TimePicker: Last 30 days

Pick A Service: RDS

VPC: FINRA-QC

AGS: BIZJIRA

ResourceID: ALL

CheckName: ALL

Apply PROD Waivers on

● None
○ FINRA-QC
○ FINRA-DEV

Scoring Summary

Date	Service	NoofResourceIDs	Passed	Failed
2017-07-18	RDS	1		9
2017-07-14	RDS	1		9
2017-07-12	RDS	1		9

Waivers

AGS	SERVICE	Account	RSAMWAIVERS	RSAMWAIVERS:Approved	HostWithWaiver:Perc
BIZJIRA	RDS	510199193688	2	1	1:100%
COBRA	RDS	510199193688	1	1	
PUPPET	RDS	510199193688	1		

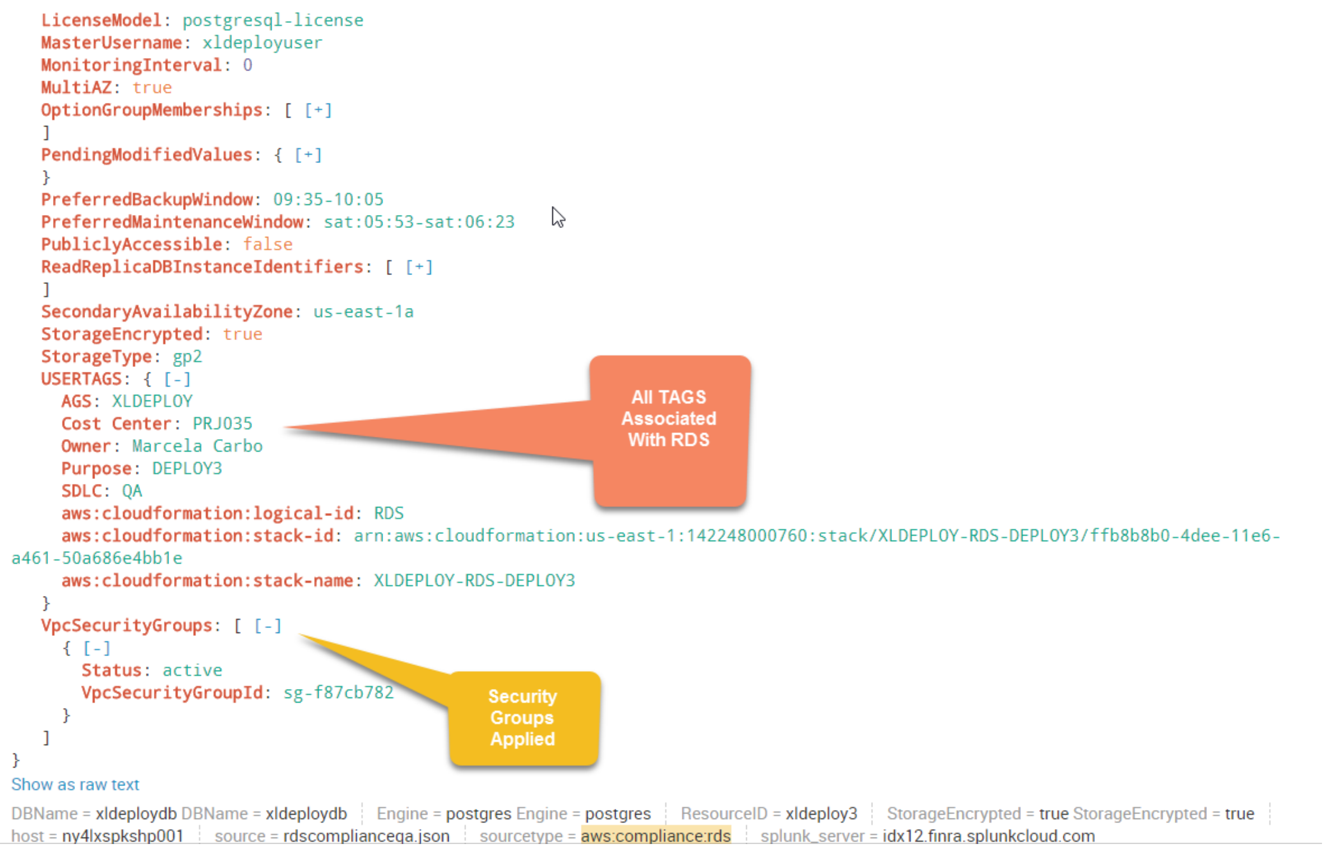
About Support File a Bug Documentation Privacy Policy

© 2005-2017 Splunk Inc. All rights reserved.

DevOps View

Same Data, Different Use Case

- Collect Only Once/Change
- Automate DevOps Checks for Resource Creation
  - Enforce TAGGING
- Provide Metrics to App Teams
  - No of Instances, Usage
- Check for Config Changes
  - Security Group Changes



```
  LicenseModel: postgresql-license
  MasterUsername: xldeployuser
  MonitoringInterval: 0
  MultiAZ: true
  OptionGroupMemberships: [ [+]
  ]
  PendingModifiedValues: { [+]
  }
  PreferredBackupWindow: 09:35-10:05
  PreferredMaintenanceWindow: sat:05:53-sat:06:23
  PubliclyAccessible: false
  ReadReplicaDBInstanceIdentifiers: [ [+]
  ]
  SecondaryAvailabilityZone: us-east-1a
  StorageEncrypted: true
  StorageType: gp2
  USERTAGS: { [-]
    AGS: XLDEPLOY
    Cost Center: PRJ035
    Owner: Marcela Carbo
    Purpose: DEPLOY3
    SDLC: QA
    aws:cloudformation:logical-id: RDS
    aws:cloudformation:stack-id: arn:aws:cloudformation:us-east-1:142248000760:stack/XLDEPLOY-RDS-DEPLOY3/ffb8b8b0-4dee-11e6-
a461-50a686e4bb1e
    aws:cloudformation:stack-name: XLDEPLOY-RDS-DEPLOY3
  }
  VpcSecurityGroups: [ [-]
    { [-]
      Status: active
      VpcSecurityGroupId: sg-f87cb782
    }
  ]
}
```

Show as raw text

DBName = xldeploydb DBName = xldeploydb | Engine = postgres Engine = postgres | ResourceID = xldeploy3 | StorageEncrypted = true StorageEncrypted = true
host = ny4lxspkshp001 | source = rdscomplianceqa.json | sourcetype = aws:compliance:rds | splunk_server = idx12.finra.splunkcloud.com

All TAGS Associated With RDS

Security Groups Applied

Cost Analysis

$$\left(\frac{\text{4M}}{\text{Config Records}}\right) * \left(\frac{\$0.003}{\text{Record}}\right) = \$12{,}000$$

$$\left(\frac{\text{4M}}{\text{Config Records}}\right)\left(\frac{\text{192MB}}{\text{100ms}}\right) * \left(\frac{\$0.000000313}{\text{Lambda}}\right) = \$1.2$$

Wrapping Things Up

Splunk and AWS Lambda

Better Together

3 Key Takeaways

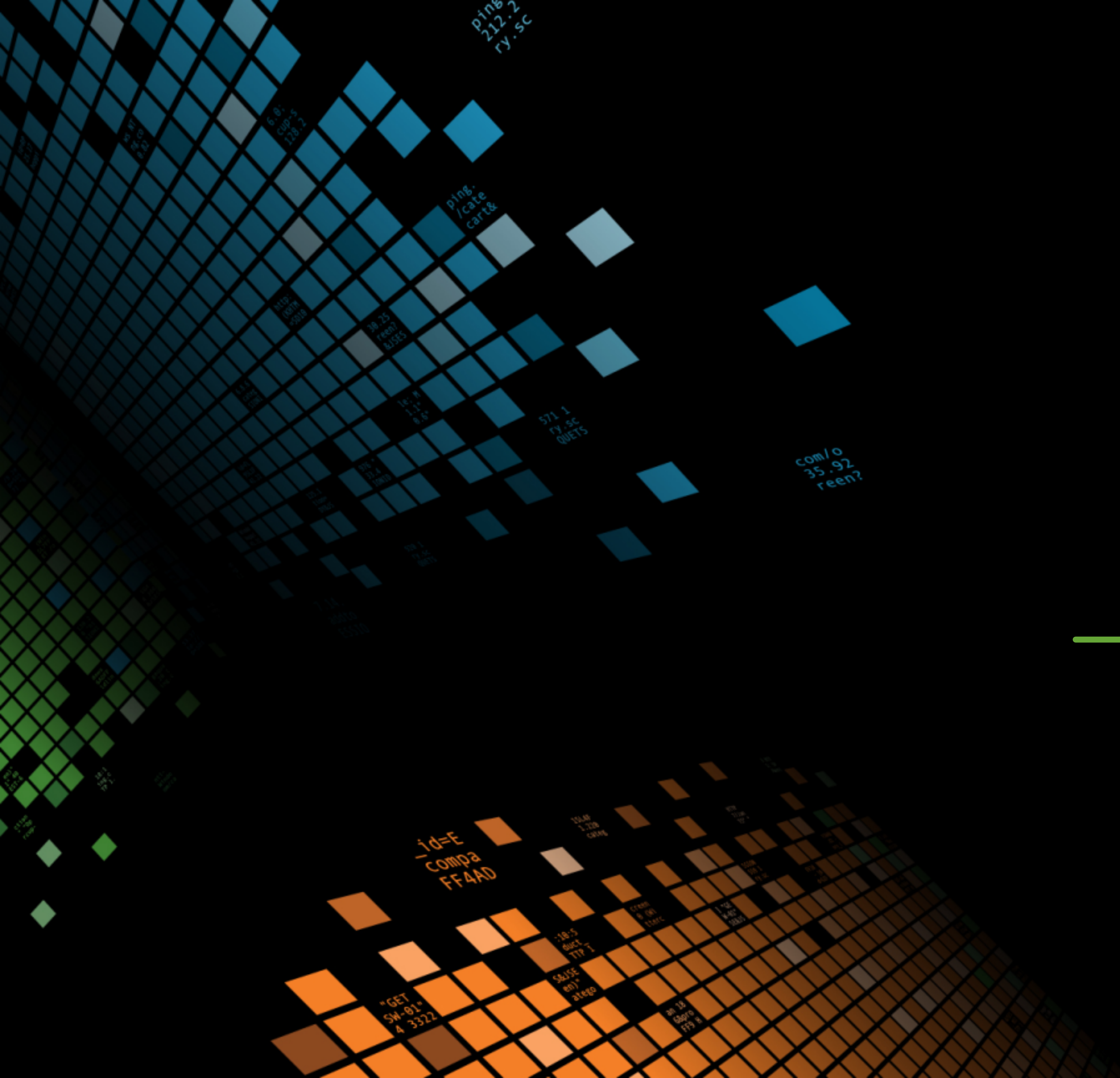
There's Always Three

- **Function As A Service** (FaaS) is Growing in Use Because it is **Affordable** and **Maintenance Free**
- **Integrating with Splunk is Easy** and an **Enterprise Approach** will Enable **Economies of Scale**
- **FaaS Leveraging the Power of Splunk** Leads to Improved **Effectiveness at a Lower Cost** in Many Key Functional Areas: Development, Security, DevOps

What If The Splunk Community?

Had a Forum for Collaboration of Splunk/AWS Lambda Integration

- We Wouldn’t Re-Invent (apologies…)
- We Could just Customize Properties Files
- We Could Deploy Using Our Existing Tools
- Functions Would Deliver AWS Content to Splunk Apps
- We Could Work Together to Build Better Classes
- Work Together to Prioritize HEC Enhancements
- Manual Configuration Would be Replaced by Button Pushes



Questions

Thank You

Don't forget to rate this session in the .conf2017 mobile app

splunk> .conf2017