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.

Splunk, Splunk>, Listen to Your Data, The Engine for Machine Data, Splunk Cloud, Splunk Light and SPL are trademarks and registered trademarks of Splunk Inc. in the United States and other countries. All other brand names, product names, or trademarks belong to their respective owners. © 2017 Splunk Inc. All rights reserved.
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 charges (29M), tweets (0.5B), and 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

What’s in it for YOU?
One Simple Query

Function="Splunk-SendMessage" | transaction RequestId startswith=START endswith=STOP keepEvicted=1
Find The Long Running Process
Payback for the Hard Work
Other Useful Commands
Information at Your Fingertips

- `Function="Splunk-SendMessage" | transaction RequestId startswith=START endswith=STOP keepevicted=1 | search closed_txn=0`
  - All Lambda runs that haven’t completed…...gracefully

- `Function="Splunk-SendMessage" | transaction RequestId startswith=START endswith=STOP | mvexpand Severity | Severity <= 3`
  - All Lambda runs that produced ERROR/CRITICAL/ALERT/EMERGENCY messages

- `Function="Splunk-SendMessage" | transaction RequestId startswith=START endswith=STOP keepevicted=1 | search RequestId > 1`
  - All Lambda runs that had automatic restarts

- `Function="Splunk-SendMessage" | 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

<table>
<thead>
<tr>
<th>MB</th>
<th>COST</th>
</tr>
</thead>
<tbody>
<tr>
<td>128</td>
<td>0.000000208</td>
</tr>
<tr>
<td>192</td>
<td>0.000000313</td>
</tr>
<tr>
<td>256</td>
<td>0.000000417</td>
</tr>
<tr>
<td>320</td>
<td>0.000000521</td>
</tr>
<tr>
<td>384</td>
<td>0.000000625</td>
</tr>
<tr>
<td>448</td>
<td>0.000000729</td>
</tr>
<tr>
<td>512</td>
<td>0.000000834</td>
</tr>
<tr>
<td>576</td>
<td>0.000000938</td>
</tr>
<tr>
<td>640</td>
<td>0.000001042</td>
</tr>
<tr>
<td>704</td>
<td>0.000001146</td>
</tr>
<tr>
<td>768</td>
<td>0.000001250</td>
</tr>
<tr>
<td>832</td>
<td>0.000001354</td>
</tr>
<tr>
<td>896</td>
<td>0.000001459</td>
</tr>
<tr>
<td>960</td>
<td>0.000001563</td>
</tr>
<tr>
<td>1024</td>
<td>0.000001667</td>
</tr>
<tr>
<td>1088</td>
<td>0.000001771</td>
</tr>
<tr>
<td>1152</td>
<td>0.000001875</td>
</tr>
<tr>
<td>1216</td>
<td>0.000001980</td>
</tr>
<tr>
<td>1280</td>
<td>0.000002084</td>
</tr>
<tr>
<td>1344</td>
<td>0.000002188</td>
</tr>
<tr>
<td>1408</td>
<td>0.000002292</td>
</tr>
<tr>
<td>1472</td>
<td>0.000002396</td>
</tr>
<tr>
<td>1536</td>
<td>0.000002501</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>MB</th>
<th>NormalizedTime</th>
<th>COST</th>
<th>UnitPrice</th>
</tr>
</thead>
<tbody>
<tr>
<td>320</td>
<td>203.35</td>
<td>0.000000521</td>
<td>0.000106</td>
</tr>
<tr>
<td>384</td>
<td>159.30</td>
<td>0.000000625</td>
<td>0.000096</td>
</tr>
<tr>
<td>448</td>
<td>156.82</td>
<td>0.000000729</td>
<td>0.000114</td>
</tr>
</tbody>
</table>

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 HAPPENS
- EVENT INTO CLOUD TRAIL BUCKET
- POLLER PROCESSES EVENT OBJECT
- DATA INGESTED INTO SPLUNK
- CONTROLLABLE DELAY
- POLLING CODE STARTS
- AWS DELAY
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
S3 bucket
Lambda function
HEC
Indexers

AWS SNS
SNS topic
Lambda function
HEC
Indexers
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?

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

\[
\frac{415,000 \text{ objects}}{\text{Month}} \times \frac{2300 \text{ ms}}{\text{Lambda Run}} \times \frac{\$0.000000417}{100 \text{ ms}} = \$3.98/\text{Month}
\]

\$1 runs 104,264 \(\lambda\) functions
## Other Search Methods

A brief look at Other Collection & Search Methods

<table>
<thead>
<tr>
<th>AWS Athena</th>
<th>AWS CloudSearch</th>
<th>Manual</th>
</tr>
</thead>
<tbody>
<tr>
<td>▶️ Un/Semi/Structured Data</td>
<td>▶️ Structured Data</td>
<td>▶️ Download Files</td>
</tr>
<tr>
<td>▶️ S3 Objects as Data Feed</td>
<td>▶️ Manual/Scripted Upload</td>
<td>▶️ Unzip and Analyze</td>
</tr>
<tr>
<td>▶️ Database Tables</td>
<td>▶️ JSON/XML</td>
<td>▶️ Difficult</td>
</tr>
<tr>
<td>▶️ Limited Data Formats</td>
<td>▶️ Enriching Data</td>
<td>▶️ Not Cost Effective</td>
</tr>
<tr>
<td>▶️ Enrichment of Data</td>
<td>▶️ Pay Hourly per Instance</td>
<td></td>
</tr>
<tr>
<td>▶️ Reporting &amp; Alerting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>▶️ Pay per Search</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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?
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!!

Waivers

Compliance Results

splunk>cloud

DBX

Waivers
<table>
<thead>
<tr>
<th>i</th>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7/18/17</td>
<td>2:05:33 PM</td>
</tr>
</tbody>
</table>

- **AllocatedStorage**: 30
- **AutoMinorVersionUpgrade**: True
- **AvailabilityZone**: us-east-1c
- **BackupRetentionPeriod**: 14
- **CACertificateIdentifier**: rd5-ca-2015
- **CopyTagsToSnapshot**: false
- **DBInstanceArn**: arn:aws:rds:us-east-1:150199192688:db:wiki
- **DBInstanceClass**: db.m4.xlarge
- **DBInstanceIdentifier**: wiki
- **DBInstanceStatus**: available
- **DBName**: wiki
- **DBParameterGroups**: [ ]
- **DBSecurityGroups**: [ ]
- **DBSubnetGroup**: [ ]
- **DBInstancePort**: 0
- **DBInstanceStorage**: 1024
- **DBParameterGroup**: RD5_1024
- **DBTagOptions**: [ ]
- **DBSubnetGroup**: [ ]
- **DBInstanceDescription**: ""
Security

LicenseModel: postgresql-license
MasterUsername: xlddeployuser
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: Marcelo Carbo
  Purpose: DEPLOY3
  SDLC: QA
  aws:cloudformation:logical-id: RDS
  aws:cloudformation:stack-name: XLDEPLOY-RDS-DEPLOY3
VpcSecurityGroups: [ [-] ]
  [-]
    Status: active
    VpcSecurityGroupId: sg-f87cb782

Show as raw text
DBName = xlddeploy
DBName = xlddeploy
Engine = postgres
Engine = postgres
ResourceId = xlddeploy3
ResourceId = xlddeploy3
StorageEncrypted = true
StorageEncrypted = true
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
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
Cost Analysis

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

\[
\left( \frac{4\text{M}}{\text{Config Records}} \right) \times \left( \frac{192\text{MB}}{100\text{ms}} \right) \times \left( \frac{\$0.000000313}{\text{Lambda}} \right) = \$1.2
\]
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
Thank You

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