Using Splunk to Manage AWS

Gaining Transparency into Cloud Computing

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.

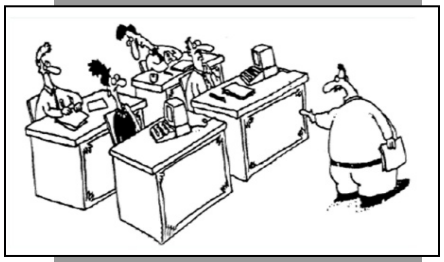
Who We Are

- FINRA—the Financial Industry Regulatory Authority—is an independent, non-governmental regulator for all securities firms doing business with the public in the United States
- FINRA protects investors by regulating brokers and brokerage firms and by monitoring trading on U.S. stock markets
- FINRA monitor over 6 billion shares traded on the stock market each day
- FINRA handles more 'Big Data' on a daily basis than the Library of Congress or Visa®—to build a holistic picture of the trading market
- FINRA – Deter, Detect, Discipline

Investor Protection

Historical View

- Cyclic Processes
  - POC – Budget Approval – SDLC - Maintenance
- Defined Roles
  - Coders Code
  - Managers Manage
  - Administrators Administer
- Agile Development/Cloud Computing
  - Developers Make These Decisions:
    - Security
    - Financial
    - Architecture
  - And It's All Point and Click
- Hacking Redefined Security
  - Defensive Coding
  - Baked In, Not Painted On



"You guys start coding, I'll go find out what the users want."

Same Challenges/Different Environment

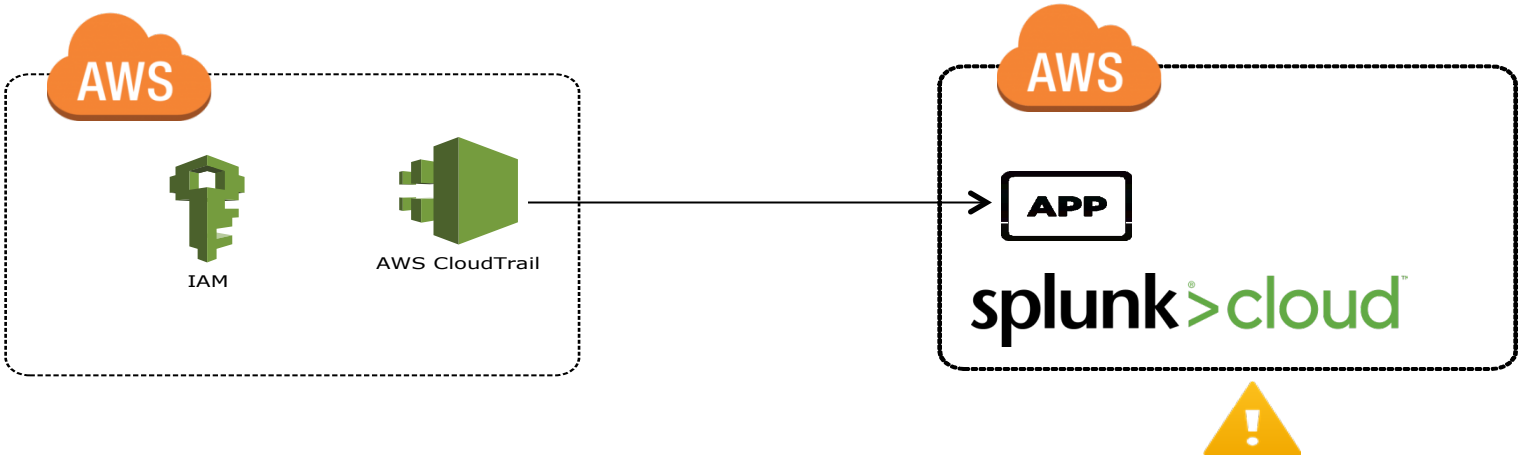
- Security
  - Engaged All Necessary AWS Security Features
  - Are we Firewalled Correctly
- Compliance
  - Followed All Published Standards
- Networking
  - Placed Servers on the Correct Network
- Finance
  - Stayed within Budget
- Capacity Planning
  - Used Resources Optimally
- But, Now in a Decentralized Model
  - It's déjà vu all over again…*Yogi Berra*

"With great power comes great responsibility."

AWS Security – Identity Access Management

- AWS Shared Responsibility Model
  - AWS – “Security of the Cloud”
  - YOU – “Security in the Cloud”
- Identity Access Management (IAM)
  - Critical security-related “service”
- Best Practices
  - Lock away your AWS account (root) access keys
    - Console Access
  - Rotate Credentials Regularly
    - Time vs Event Driven
  - Remove Unnecessary Credentials
    - Unused
  - Grant Least Privilege
  - Keep a History of

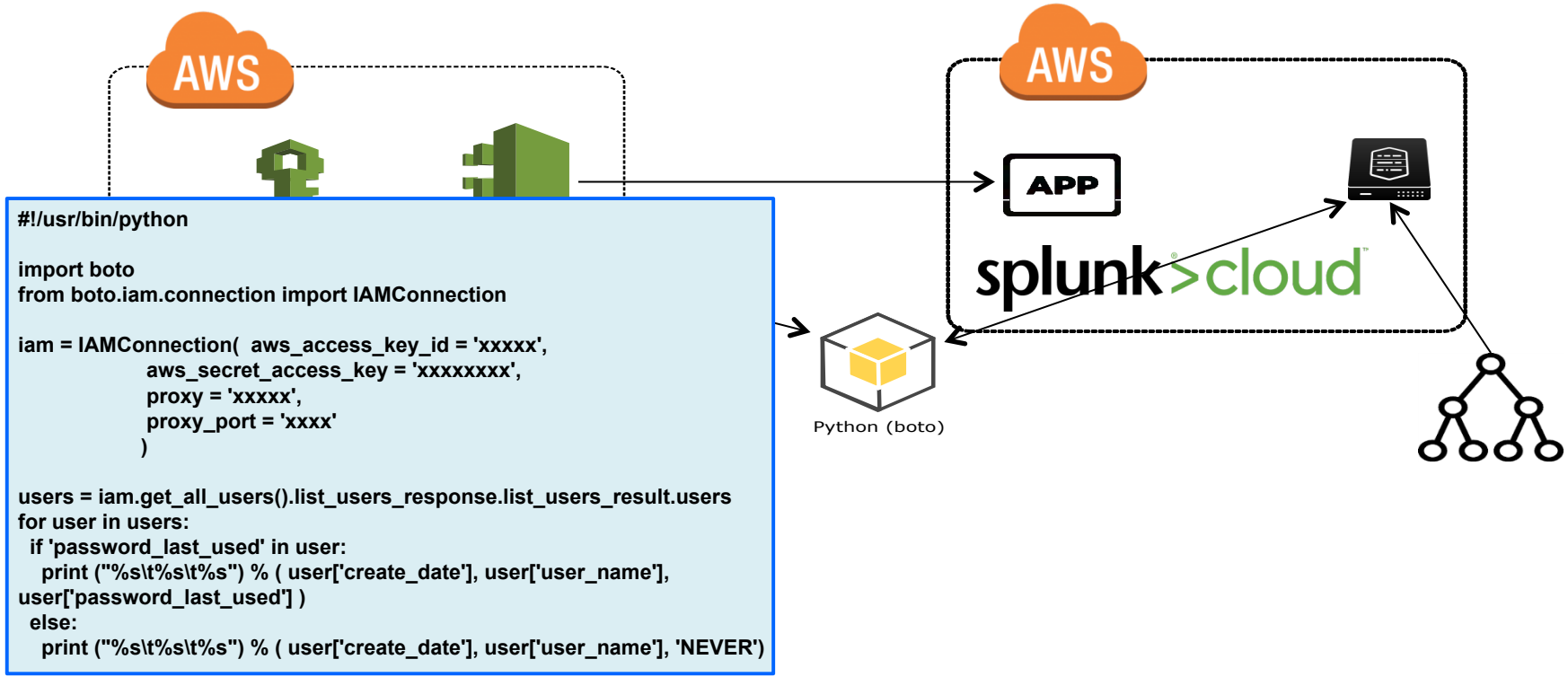
Lock Away Your AWS Account (Root) Access Keys



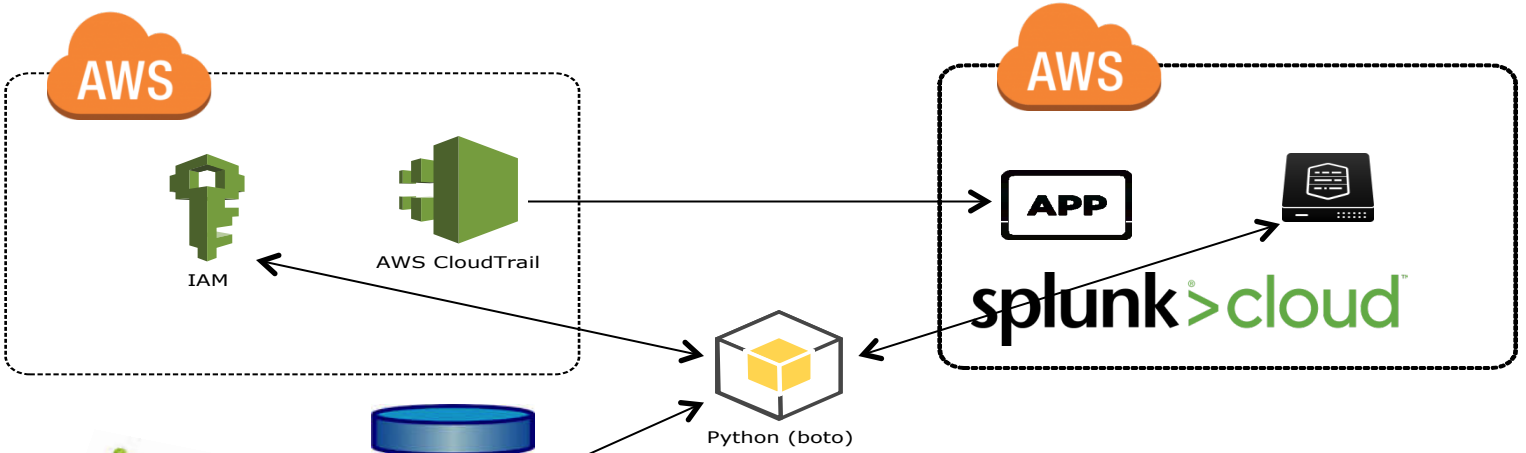

```
index=aws-cloudtrail sourcetype=aws-cloudtrail "requestParameters.userName"=root eventName=CreateAccessKey
```

```
index=aws-cloudtrail "userIdentity.arn"="*:root" eventName=Console* |
table _time eventName responseElements.ConsoleLogin awsRegion sourceIPAddress eventSource userIdentity.arn userIdentity.type
userIdentity.accountId userAgent
```

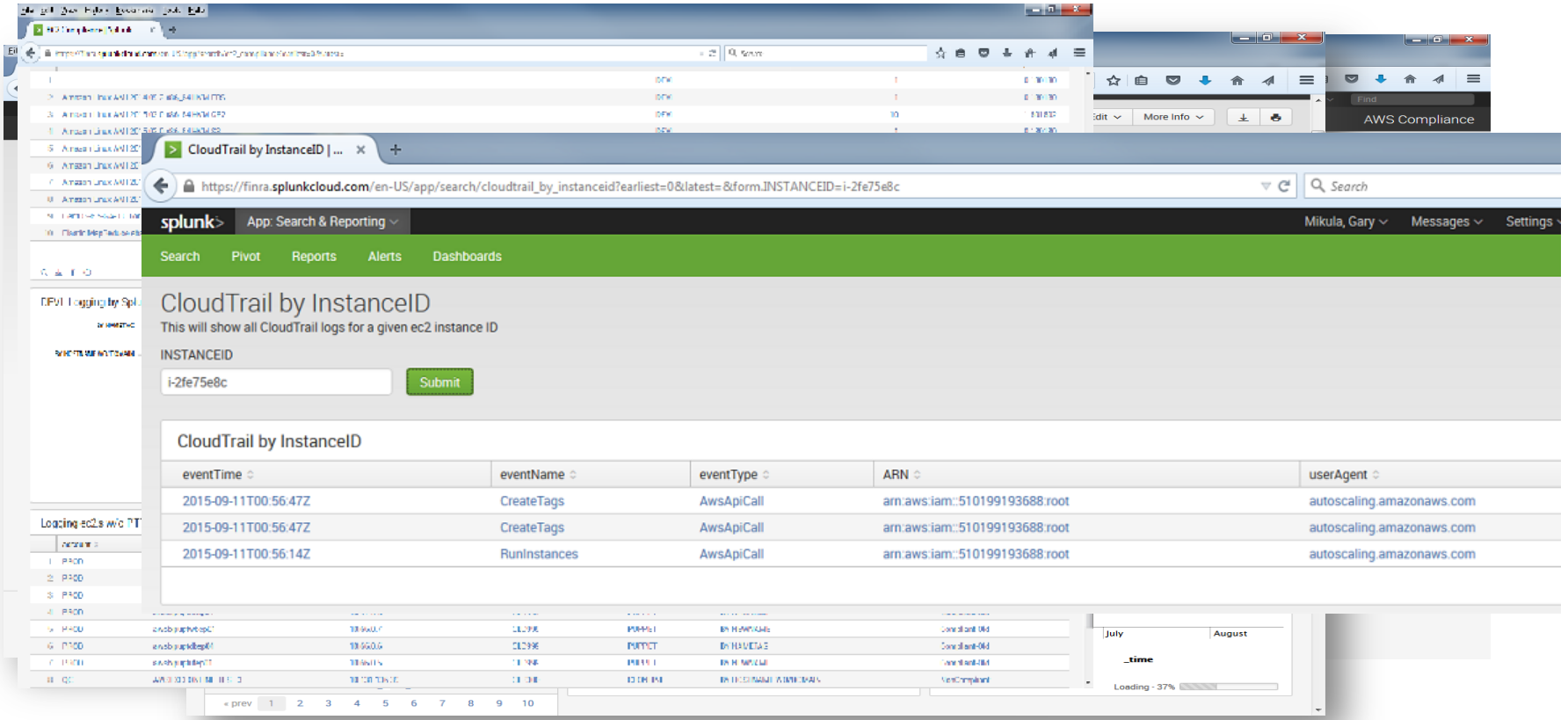
Credentials (Rotation/Unnecessary)



Grant Least Privilege



ROLE ABC
autoscaling:DeletePolicy:Deny
autoscaling:DeleteScheduledAction:Deny
autoscaling:DeleteTags:Deny
autoscaling:DescribeAdjustmentTypes:Allow
autoscaling:DescribeAutoScalingGroups:DontCare
autoscaling:DescribeAutoScalingInstances:Allow


CloudTrail by InstanceID | ...
https://finra.splunkcloud.com/en-US/app/search/cloudtrail_by_instanceid?earliest=0&latest=&form.INSTANCEID=i-2fe75e8c
Search
splunk>
App: Search & Reporting
Mikula, Gary
Messages
Settings
Search
Pivot
Reports
Alerts
Dashboards
CloudTrail by InstanceID
This will show all CloudTrail logs for a given ec2 instance ID
INSTANCEID
i-2fe75e8c
Submit
CloudTrail by InstanceID
eventTime
eventName
eventType
ARN
userAgent
2015-09-11T00:56:47Z
CreateTags
AwsApiCall
arn:aws:iam::510199193688:root
autoscaling.amazonaws.com
2015-09-11T00:56:47Z
CreateTags
AwsApiCall
arn:aws:iam::510199193688:root
autoscaling.amazonaws.com
2015-09-11T00:56:14Z
RunInstances
AwsApiCall
arn:aws:iam::510199193688:root
autoscaling.amazonaws.com
AWS Compliance
More Info
July
August
_time
Loading - 37%
« prev
1
2
3
4
5
6
7
8
9
10

Beyond IAM Best Practices

- Tagging Compliance
- Security Group
- Logging
- S3/EBS Encryption
- S3 Bucket Policies
- AMI White/Blacklisting
- EC2 Role
- Naming Conventions
- New AWS Features will Supplant
  - VPC Flows
  - Config

Project Cost Management in AWS

Harnessing the Power of Splunk

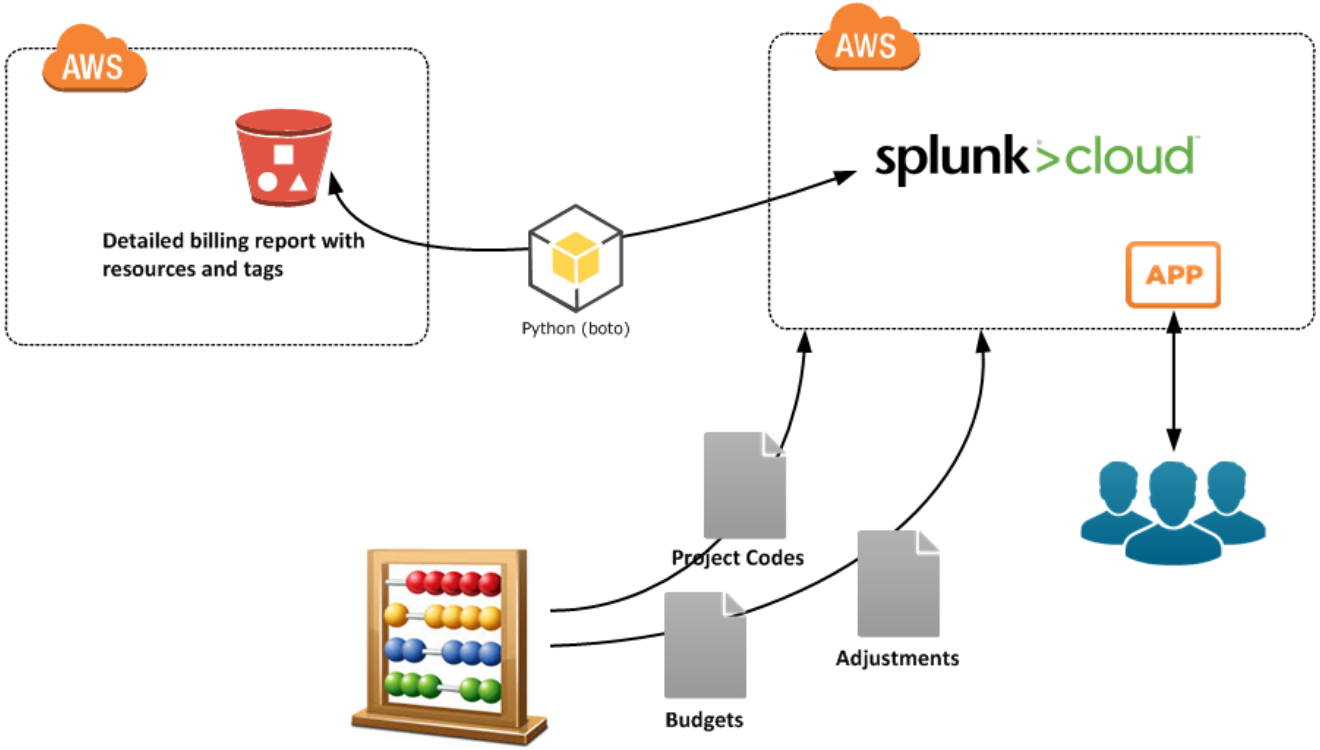
Where We Were

- Traditional Financial Review Cycles Too Long
  - Quarterly Reviews
- AWS Detailed Billing Reports Are Daunting
  - Over 10 Million Line Items
- Project Managers Need Focus
  - Am I Below My Budget?
  - Where Are My Costs Going?
  - Who's Spending Them?
- Manual Compilation of Reports
  - Integrate FINRA Data

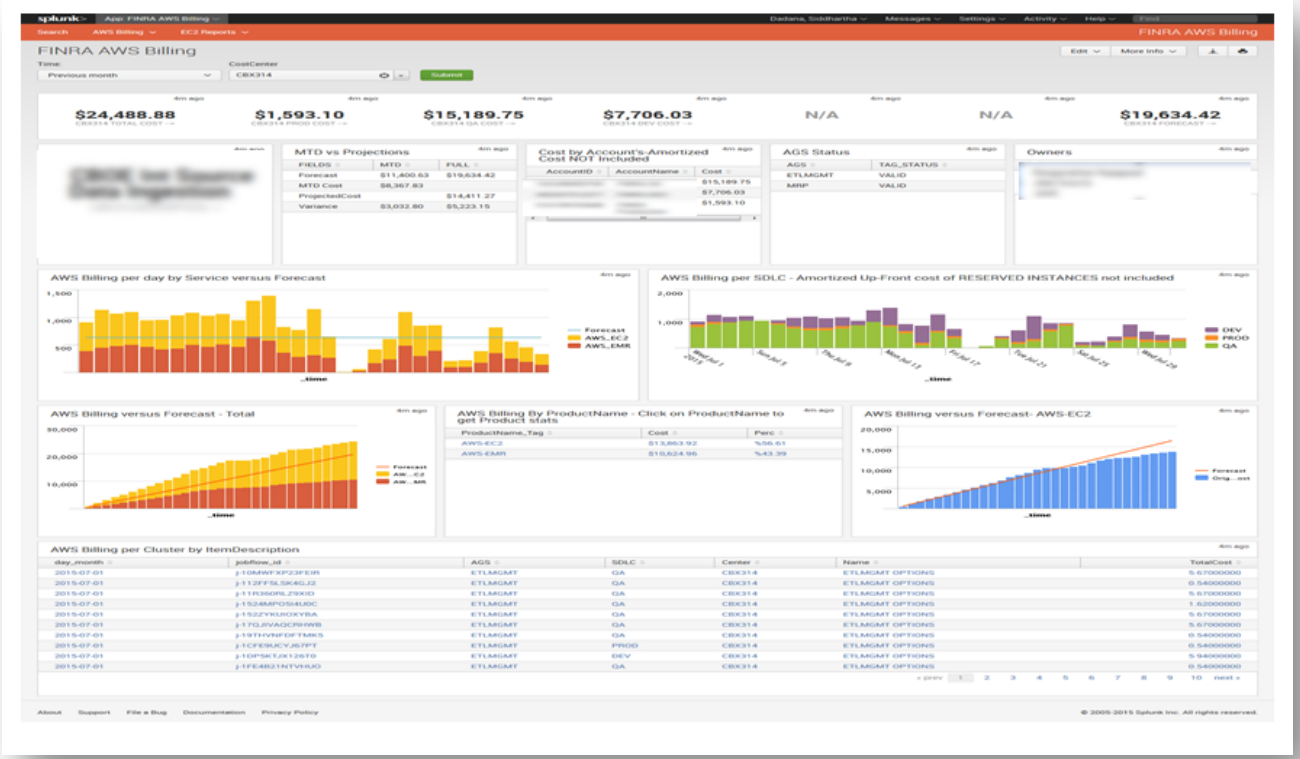
Approach Chosen

- Use Splunk as Process/Delivery System
  - Ability to Collect/Analyze/Visualize
- Collect AWS Billing Data in Splunk
  - Billing Data from S3 bucket (Daily Load)
  - Detailed Line Items w/Resources & Tags
- Data Enrichment
  - Project Code Lookups
  - Forecast Projections
  - Billing Adjustments
- Build Interfaces
  - FINRA AWS Billing

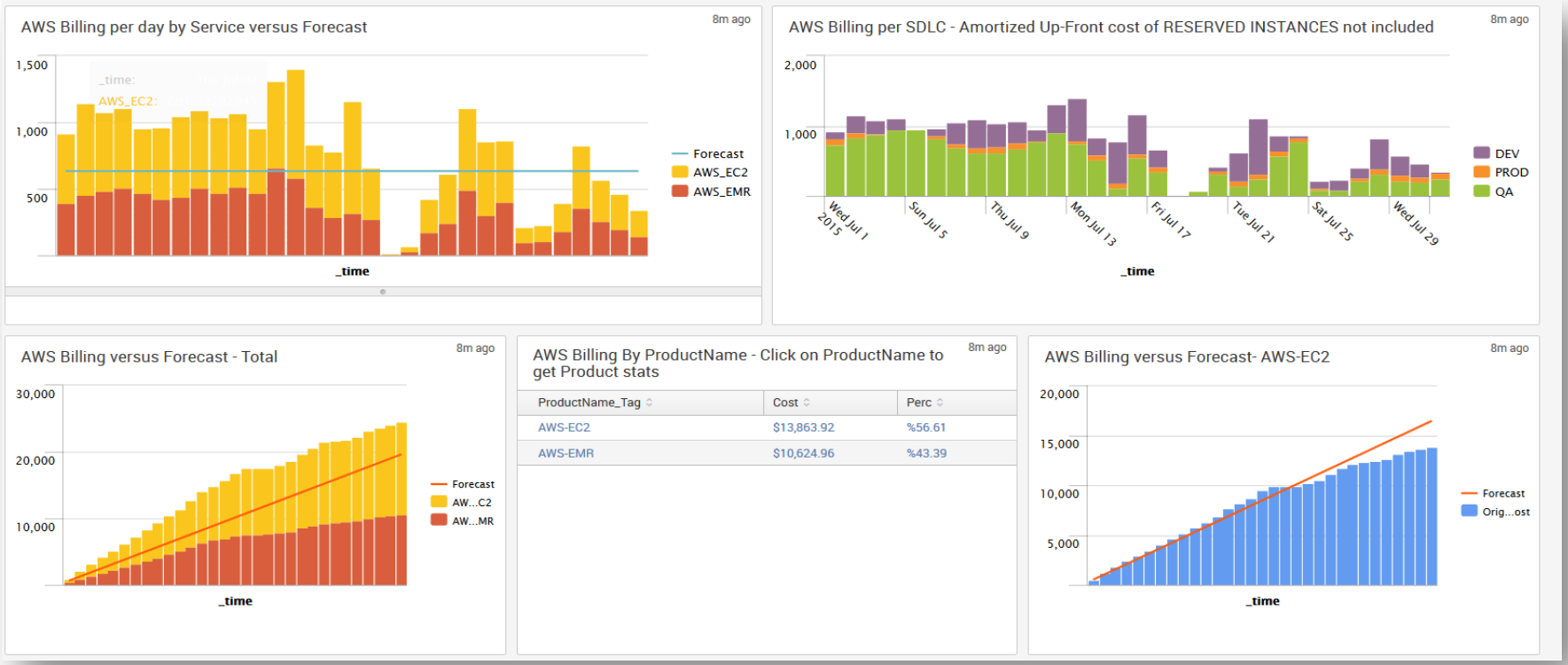
How We Did It


AWS
Detailed billing report with resources and tags
Python (boto)
AWS
splunk>cloud
APP
Project Codes
Budgets
Adjustments

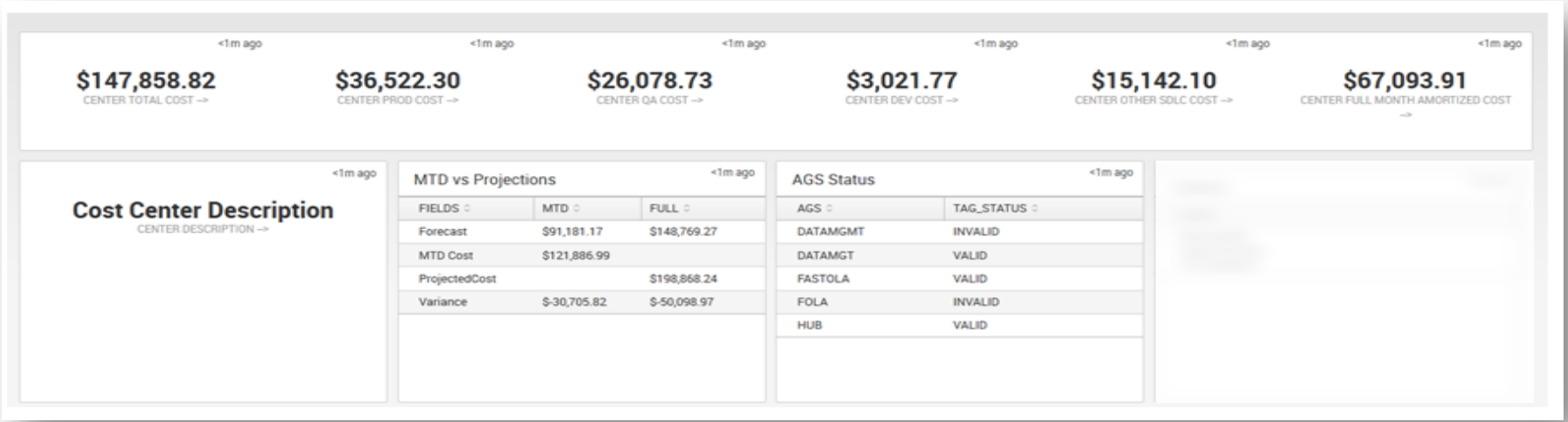
FINRA AWS Billing App



AWS Billing App



AWS Billing App



Impact – Reduced Costs

- Focus on Low Hanging Fruit
  - Shutting Down Services over Weekends/Evenings
  - Storage Sun Setting/Dormant EC2
  - Identify AWS Services with Highest Spending
  - Projects Over Budget
- Results
  - 13.5% Reduction in Billing Line Items in 1 Month
- Better Forecast Projections
  - Feedback and Control

Futures

- Users Want Even Shorter Cycles
- Back Tagging
- ‘Free Rider’ Services
- Drill Down Analytics

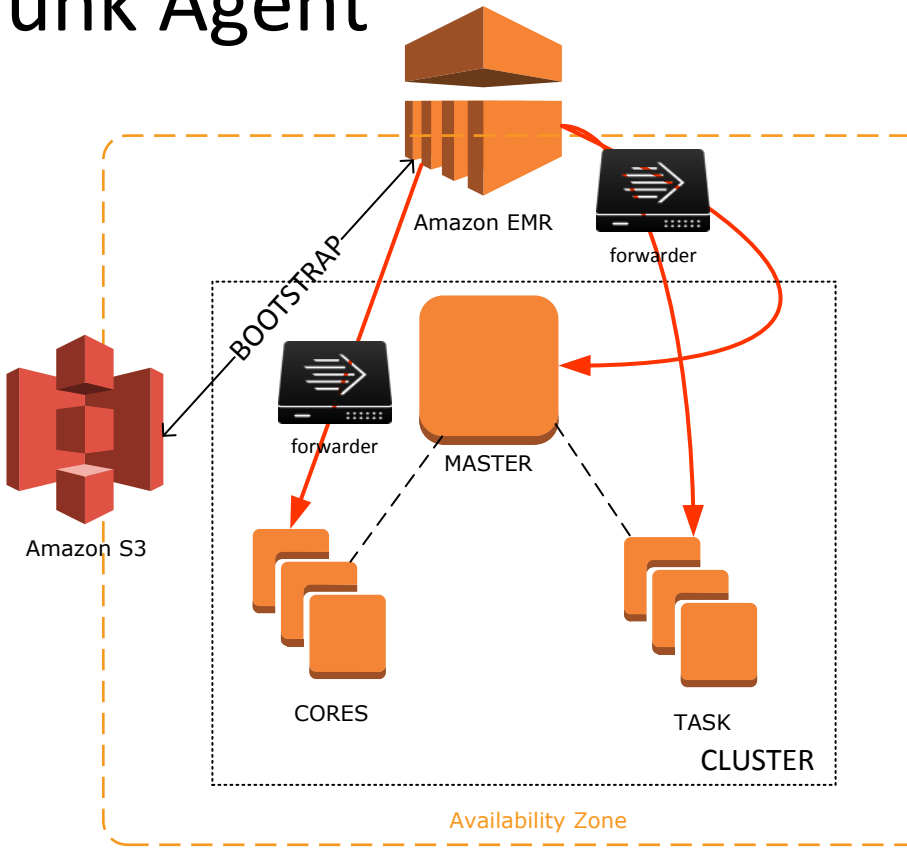
Splunking in AWS EMR
Gaining Transparency Into PaaS

What Stood In Our Way

- PaaS and IaaS are not Equal
- Instance Fingerprinting
  - Identify Nodes
  - Instance Role
  - User Tags
- Data Retention
- Collection Delay
- ➢ Bootstrap Splunk Agent

Bootstrap Splunk Agent

- Store bsx in S3
  - Splunk rpm
  - Deploymentclient.conf
  - Discovery Scripts*
- Execute Bootstrap
- Master Installation
- Core & Task Installation
- AWS was Extremely Supportive of this Method



Instance Fingerprinting

1. Determine the Identity of the Node:
   - EC2 nodes have “metadata” service @ http://169.254.169.254
     - INSTANCE_ID=`curl http://169.254.169.254/1.0/meta-data/instance-id/`
     - echo $INSTANCE_ID i-8f0d4c75
     - IP=`curl -s http://169.254.169.254/2014-11-05/meta-data/local-ipv4/`
2. Grab the User Defined Tags
   - AGS_TAG=`/opt/aws/apitools/ec2/bin/ec2-describe-tags|grep $INSTANCE_ID|grep AGS|cut -f5`
   - SDLC=`/opt/aws/apitools/ec2/bin/ec2-describe-tags|grep $INSTANCE_ID|grep SDLC|cut -f5|awk '{print substr($0,1,1)}'`

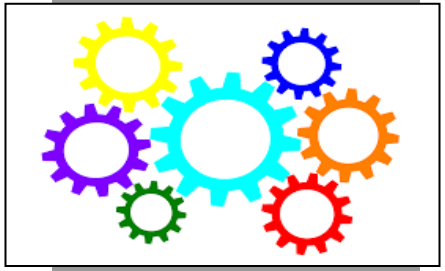
Instance Fingerprinting

3. Get the jobFlowID (EMR Cluster ID) for the node:
   - JOBFLOW_ID=`awk -F"[,:]" '{for(i=1;i<=NF;i++){if($i~/jobFlowId\042/){print $(i+1)} } }' /mnt/var/lib/info/job-flow.json|tr -d [\"\ ]`
4. Determine Role Node in the EMR cluster:
   - INSTANCE_ROLE=`cat /mnt/var/lib/info/job-flow.json | ruby -e "require 'rubygems'; require 'json'; require 'pp'; igs=JSON [STDIN.read]['instanceGroups']; igid=ENV['INSTANCEGROUP_ID'];ig = igs.find {|i| i['instanceGroupId'] == igid};puts ig['instanceRole']"|awk '{print toupper($0)}'`
5. Update Splunk Config Files
   - Deploymentclient.conf
   - Inputs.conf

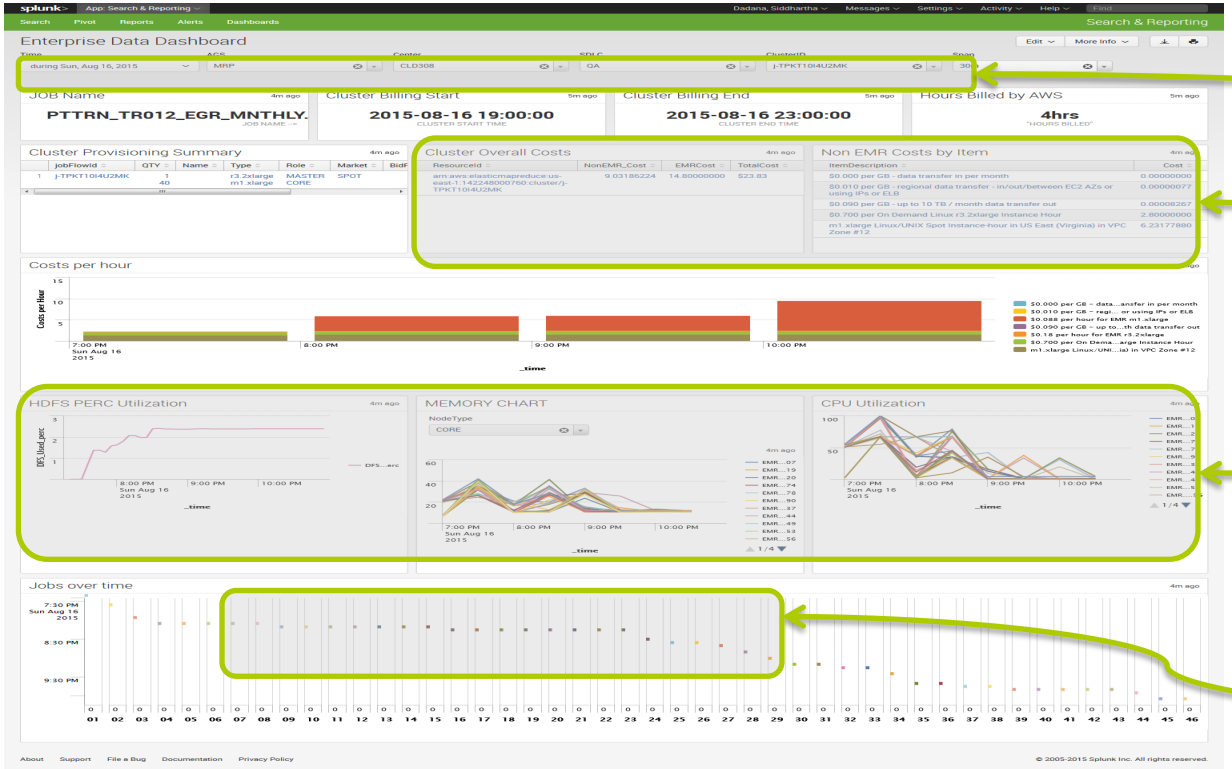
Inputs.conf Example

inputs.conf

```
[default]
host = EMRAPP1-MASTERED-j-YQQU43YEHG4X:1.1.1.1
_meta=jobflow-id::"j-YQQU43YEHG4X" instancegroup-
id::"ig-3JBIDJGYVSD4N" instance-id::"i-22f7f3f0"
instance-role::"MASTER" ags_hostname::"EMRAPP1-
MASTERED-j-YQQU43YEHG4X:1.1.1.1"  ags::“ABCD"
cost_center::"CLDABC" creator::“APP1_Team"
name::"AWSLXAPP1-CLED01-YR" owner::“USERIDABC"
purpose::“APP1_QUERY_CL" sdlc::"DEV"
```



EMR Cluster Analyzer –> Summary Dashboard



SELECT THE CLUSTER

AWS Billing

Splunk for *NIX

Jobs running

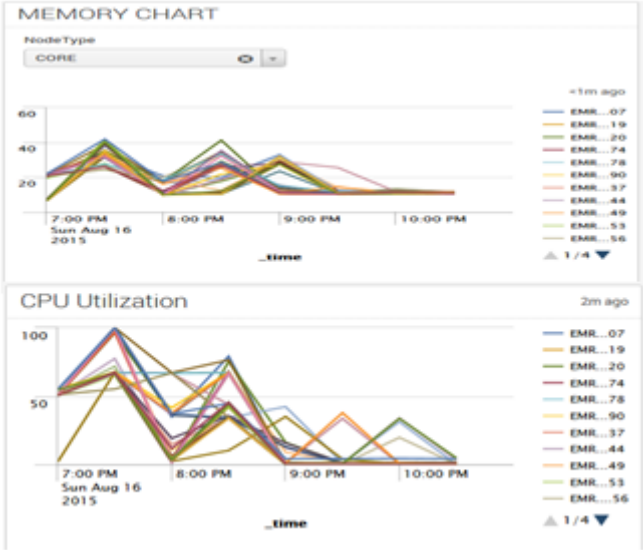
Potential EC2 Provisioning Cost Savings

- m1.xlarge
  - 4 vCPUs with 8 ECUs
  - 15 GB memory
  - 4 x 420 GB disk
  - $0.35 per hour

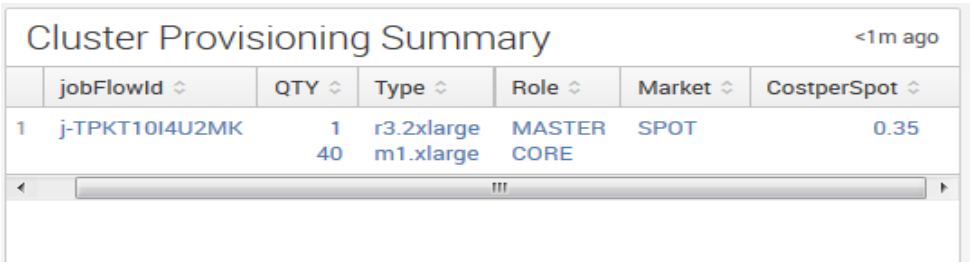
- c3.xlarge
  - 4 vCPUs with 14 ECUs
  - 7.5 GB memory
  - 2 x 40 GB disk (SSD)
  - $0.27 per hour

Potential 23% savings in EMR costs

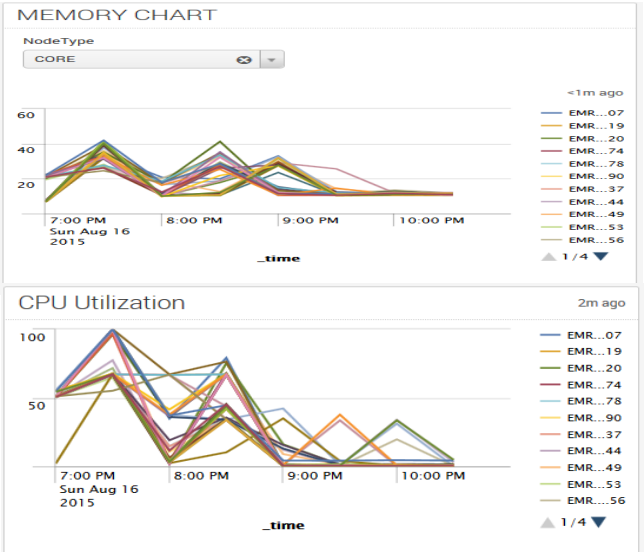
With additional ECUs and SSD disks, the c3.xlarge may be more performant than the m1.xlarge instances at a better price point
RESIZING***



Resizing Analysis



Summary provides
the information of number of Nodes



1. Less than 80% utilization overall

2. After 70mins utilization at less than 50% (cpu)

Conclusion:
The cluster be resized after the two hours

Resizing Analysis – Scenario 1

All instances running for full duration of the job

1st Hour	2nd Hour	3rd Hour	4th Hour
40	40	40	40

Cost Analysis:

Cost = (Price per instance) * (No of Instances) * (No of hours)

Price = 0.35 * 40 * 4 = $56

Resizing Analysis – Scenario 2

Resizing after the 1st hour and 2nd hour

1st Hour	2nd Hour	3rd Hour	4th Hour
40	35	20	20

Cost Analysis:

Cost = (Price per instance) * (No of Instances) * (No of hours)

Price = (0.35*40*1) + (0.35*35*1)+(0.35*20*2) = 14 + 12.25 + 14 = $40.25

Savings Compared to Scenario 1 = 28.2%

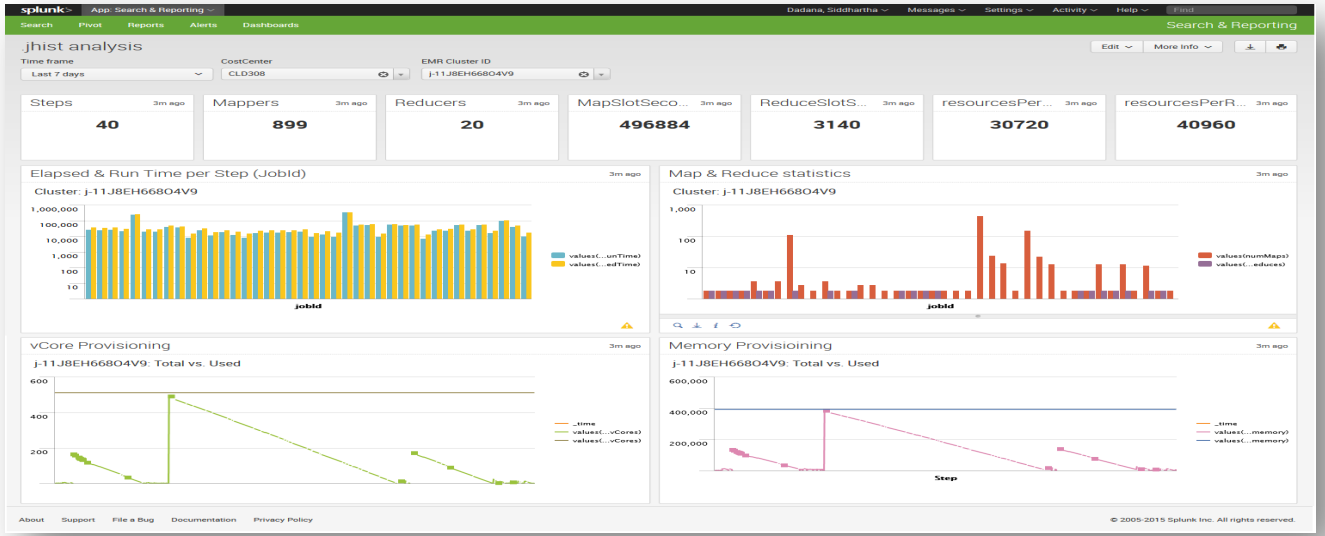
Combined EC2 Provisioning & Resizing Analysis

- Combined Savings
  - Original Cost = $56
  - Price After Combined Analysis = $31.05
  - Job Savings = $24.95 = 55.44%
- Job Runs 5x/day ($24.95 * 5 = $124.75)
- Every Business day/week ($124.75 * 5 = $623.75)
- Every Week of the Year ($623.75 * 52 = $32,435)
- And….We Haven't Affected Performance
  - Just More Efficient Provisioning

Job History

Cluster Summary

Per Cluster



.jhist file Analysis:

- Cluster level statistics
- Run Time, Map/Reduce Stats per Steps

Futures/Other Uses

- Grade Clusters
  - Identify under utilized clusters for faster resizing
- ITSI Integration
  - KPI's based auto analysis on Cloud
- Additional Input Variables
  - Size of Data Sets
  - Number of Runs
- Metrics Correlation
  - Analyze Steps
  - Jobs Impact on System

THANK YOU

splunk>