**Lab 4**

**Galen J. Yanofsky**

**SDEV 400 7981**

**Dr. James Robertson**

**17 December 2017**

1. **List and describe the five pillars of the AWS well-architected framework. For each pillar, give at least 2 specific examples of AWS services that support the pillar and describe how the services support the pillar.**

The AWS Well-architected framework includes 5 pillars:

1. **Operational Excellence:** This pillar focuses on running and monitoring system of varying size to deliver business value as well as continuous process improvement. Some of the key topics of this pillar include:
   * **Key AWS Services Supporting Operational Excellence**
     + **AWS Cloud Compliance:** enables understanding of the robust controls in place at AWS to maintain security and data protection in the cloud.
     + **AWS Developer Tools:** set of services enabling rapid and safe delivery of software.
2. **Security:** This pillar focuses on protecting information & system. Topics include confidentiality and integrity of data, identifying and managing who can do what with privilege management, protecting systems, and establishing controls to detect security events.
   * **Key AWS Services Supporting Security**
     + **IAM Instance Profiles For EC2 Instances:** Allows user to leverage the Amazon EC2 metadata service and managed, temporary credentials for accessing other AWS APIs.
     + **Amazon Athena:** Can be used to analyze logs, such as CloudTrail logs, to help identify trends and further isolate activity by attribute, such as source IP address or user.
3. **Reliability:** This pillar focuses on the ability to prevent, and quickly recover from failures to meet business and customer demand. Key topics include foundation elements around setup, cross project requirements, recovery planning, and change management.
   * **Key AWS Services Supporting Reliability**
     + **AWS Virtual Private Cloud (VPC):** Which allows allocation of private IP address ranges to either provide non-internet-accessible resources or to extend your data center.
     + **AWS Elastic Cloud Compute (EC2):** Allows implementation of multiple types of services and solutions (Windows & Linux) on top of the VPC.
4. **Performance Efficiency:** This pillar focuses on using IT and computing resources efficiently. Key topics include selecting the right resource types and sizes based on workload requirements, monitoring performance, and making informed decisions to maintain efficiency as business needs evolve.
   * **Key AWS Services Supporting Performance Efficiency**
     + **Amazon EBS:** Provides persistent block storage volumes for use with EC2 Instances.
     + **Amazon Glacier:** Provides an archival platform for long-term data storage and is integrated with Amazon S3 for tiered storage.
5. **Cost Optimization:** Focuses on avoiding un-needed costs. Key topics include understanding and controlling where money is being spent, selecting the most appropriate and right number of resource types, analyzing spending over time, and scaling to meet business needs without overspending.
   * **Key AWS Services Supporting Cost Optimization**
     + **AWS Lambda:** Provides a serverless computing environment that can execute code without a company having to pay for, or manage the underlying infrastructure.
     + **AWS Cost Explorer:** The tool is used to visualize and analyze a user’s costs.
6. **You are about to launch a new Cloud application running multiple EC2 instances. A team member suggests, studying the workload of the application and then determine which EC2 instances to deploy. Should you follow what your team member suggests? Explain your answer.**

Based on the variety EC2 Instance types that are available to the AWS user, it would definitely be beneficial to heed your team member’s advice. There are some important concepts that need to be understood to make the correct decision in choosing the correct instance type. Amazon Machine Image (AMI) is a template that defines one’s operating environment, including the operating system. A single AMI can be used to launch one or thousands of instances.

Instances provide computing power. As mentioned, there are variety of instance types that can be “spun up”. Instances comprise various combinations of CPU, memory, storage, and networking capacity that hives the user the flexibility to choose depending on a specific business requirement. As well, the AMI provides an auto-scaling feature that can either increase or decrease the number of instances based on demand.

The instance types are grouped into different families. They include

**General Purpose (M1, M3):** Provide a balance of CPU, memory, and network resources making them a good choice for many applications.

**Compute-Optimized (C1,CC2):** Geared toward applications that would benefit from high computing power.

**Memory-Optimized (M2,CR1):** Instances that are designed for memory intensive applications. Examples include high performance databases and distributed cache, in-memory analytics, genome assembly, Microsoft Sharepoint, etc.

**Storage-Optimized (HI1, HS1):** Provides user with direct attached storage options for applications with specific disk I/O and storage capacity requirements.

**Micro Instances (T1, T2):** Instances are a very low-cost option providing a small amount of CPU resources. Micro instances are well suited for lower throughput applications like bastion hosts or administrative application for low-traffic websites that require additional compute cycles from time to time.

**GPU Instances (CG1):** Allows users to take advantage of the parallel performance of NVidia Tesla GPUs using the CUDA or OpenCL programming models for GPGPU computing. GPU instances provide high capabilities and support cluster networking.

The availability of multiple instance types, combined with features like EBS-optimization, and cluster networking allow application to be optimized for increased performance, improved application resilience, and lower costs.

1. **We live in an insecure IT era. It is no longer a matter of if a security incident will occur as when the security incident will occur and how best to be prepared to respond to that incident. How do the available AWS services assist in responding to incidents? Be specific by providing the service and how that service may assist.**

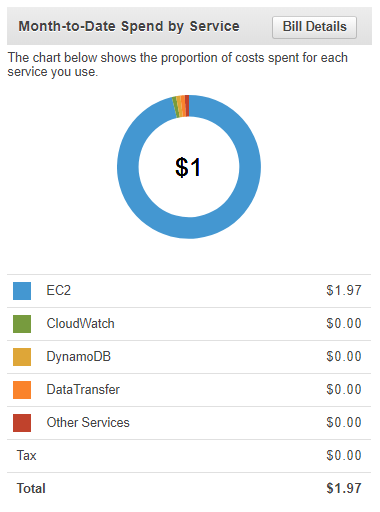
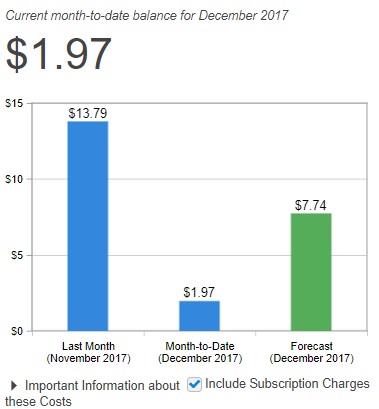
An incident can be classified as an unplanned interruption to an IT service or reduction in the quality of an IT service. For all services, it is important to track, monitor, analyze, and audit events. The following tools can be leveraged in building a cloud-specific incident response plan.

* **AWS CloudTrail:** A service that enables governance, compliance, operational auditing, and risk auditing of your AWS account. With CloudTrail, a user can log, continuously monitor, and retain account activity. CloudTrail provides event history of AWS Account activity by tracking components like AWS Management Console, AWS SDKs, Command Line Tools, etc.
* **Amazon CloudWatch:** A monitoring service for AWS cloud resource and the application you run on AWS. Use Amazon CloudWatch to collect and track metrics, collect and monitor log files, set alarms, and automatically react to changes within AWS resources. CloudWatch allows the AWS user to gain system-wide visibility into resource utilization, application performance, and operational health.
* **AWS Config:** A service that enables the user to assess, audit, and evaluate the configuration of your AWS resources. Config continuously monitors and records AWS resource configurations and allows the user to automate the evaluation of the recorded configurations against desired configurations. AWS Config simplifies compliance auditing, security analysis, change management, operational troubleshooting.
* **AWS Config Rules:** The AWS Config rule system provides AWS users, admins, and developers with the flexibility of launching, configuring, using, and terminate processing different types of storage and networking

All of these services work together to help maintain security, governance, and compliance within an organization.

1. **Show your current AWS cost utilization using the AWS console and clear screen captures. Describe the information in the cost center in AWS and how you can use that information to keep cloud costs under control.**

To show cost utilization, the user has to navigate within the management console to the Billing & Cost Management Dashboard. A screen capture of my current cost utilization is displayed below.



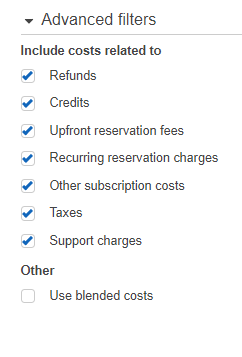
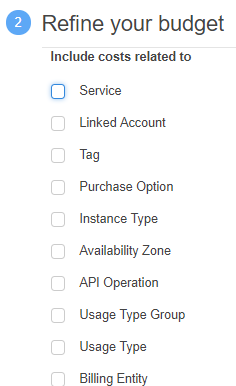
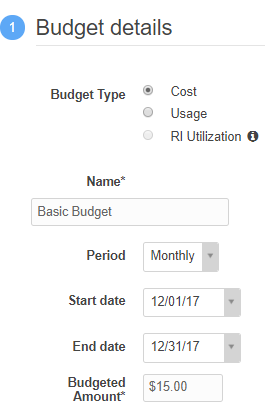
* 1. **What was the most expensive service you used?**

Clearly the most expensive service used was the EC2 Service. For my group’s CMSC495 Project, we decided to host our web application using VPC, RDS, and an EC2 instance. Due to project requirements, our needs fell out of the free-tier and we had to spin up a micro-instance, database, and network connection with a little more computing, storage, and bandwidth respectively.

* 1. **What processes can you use to efficiently monitor your cost? Describe and implement at least one of these processes.**

Cost explorer is a free tool that a user can leverage to view and monitor costs associated with AWS components. A user can view historical data for the last 13 months, forecast future costs, and get instance recommendations to better serve business needs and requirements.

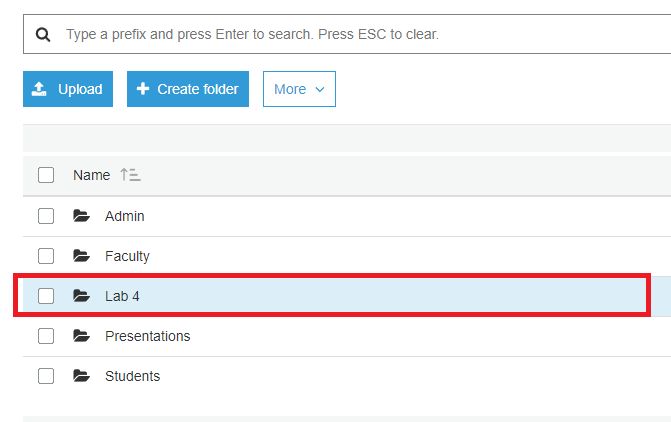
Within Billing & Cost Management Dashboard, the user has the ability to create a budget for services that automatically alerts when costs, usage, and resource utilization breach defined thresholds. In the example below I have created a budget that begins tomorrow and ends December 31st. I have created the budget so that it will alert me I If exceed $15 of cost throughout the specified time period.



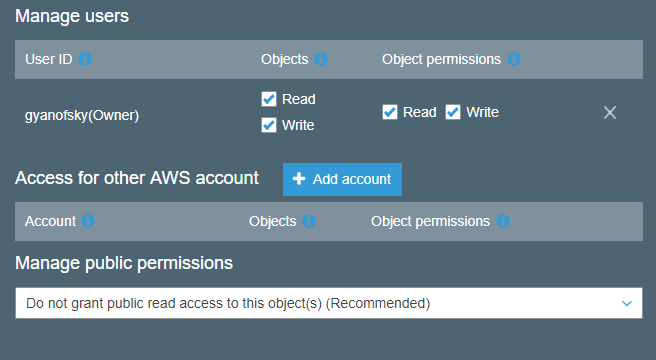
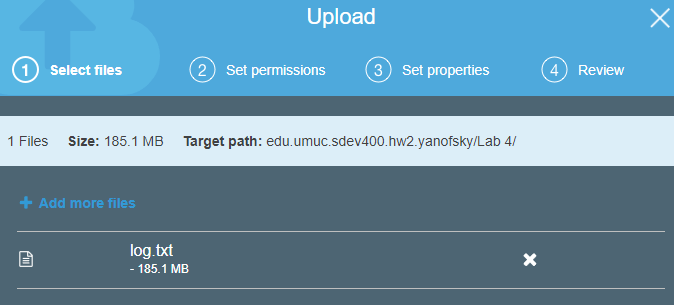
The budget tool even allows for the user to setup Notifications where AWS will notify the user via email if the previously defined threshold has been breached.

1. **Analysis of Apache Log File**

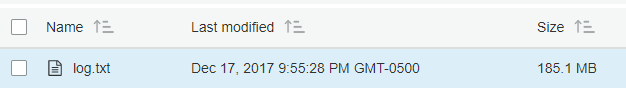
I begin this portion of the lab by creating an S3 container where the log file cane be uploaded. I create and S3 folder within my already created database where I will upload the file. The folder is labeled “Lab 4”.



I then upload the large Apache log file (185.1 MB) to this folder.



I assign the necessary permissions and options for the file. Due to my tremendously throttled internet connection, the file takes about 10 minutes to upload.

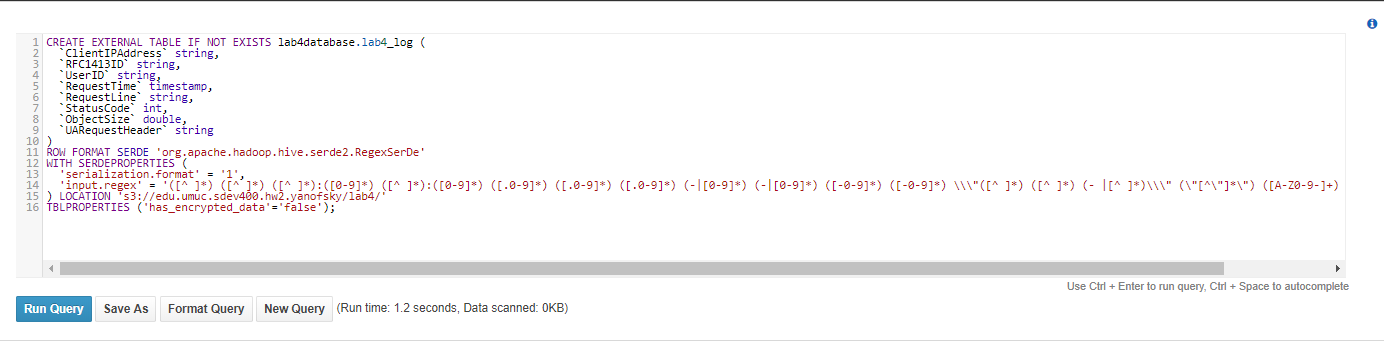


I start the Athena service and create a database. I label the database lab4database. I label the table lab4\_log. I specify the container location “s3://edu.umuc.sdev400.hw2.yanofsky/lab4/”. I then specify the file as an Apache Server Log and assign the default regular expression which will aid in processing the data so that it can be queried. The expression is below.

^(?!#)([^ ]+)\\s+([^ ]+)\\s+([^ ]+)\\s+([^ ]+)\\s+([^ ]+)\\s+([^ ]+)\\s+([^ ]+)\\s+([^ ]+)\\s+([^ ]+)\\s+([^ ]+)\\s+[^\(]+[\(]([^\;]+).\*\%20([^\/]+)[\/](.\*)$

A Perl Compatible Regular Expression (PCRE) is charactering filtering construct that implements pattern matching capabilities using a generalized syntax. The expression above is optimized for data that is stored in Apache Log Files.

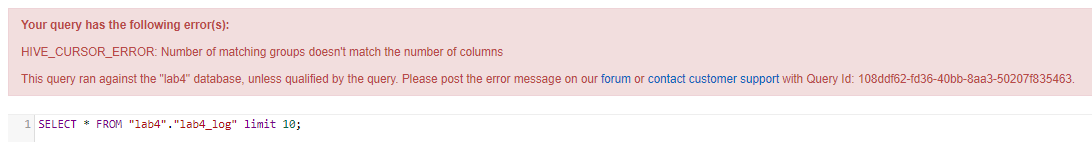
I then specify the table data. Each column has specific name and associated data-type that will optimize querying. The resulting generated table script is below.



Now to analyze the log data that has been organized into a table.

1. **How many unique client IP addresses exist in the log file?**
2. **What are the earliest and latest timestamps in the log file?**
3. **What percentage of records returned a status code of 404 (not found)?**
4. **What were the smallest and largest datafile sizes that returned with the status code of 200?**
5. **Should the supervisor be concerned about a possible botnet attack? If so, explain what pieces of information you used to determine this? Also, if you are concerned about the attack, list the top 5 client IP addresses that attempted a botnet**

Unfortunately, when I try to query my database I receive the following error.



I have tried to recreate the database and table with specified columns but I am unable to return data in table-form via the Athena service. A simple analysis within notepad++ allows me to see that the log file was in fact downloaded correctly as the data exists and should be able to be parsed via the regex.

