**Project 1: Interview Questions**

***Domain: Logging and Monitoring***

***Question 2: Challenges of Collecting Large Amounts of Log Data***

***What are the challenges of collecting huge amounts of log data? How do security analysts deal with them?***

1. Restate the Problem
   1. How do security analysts deal with collecting huge amounts of data.
2. Provide a Concrete Example Scenario

* In Project 1, when did you deal with log data?
  + I dealt with log data on day 3 of the project when we were running certain tests on our VMs. First, I tried to login incorrectly from my JumpBox to my VMs without using a container. This caused a lot of log entries for denied connection. Next, I caused my CPU stress using the stress command. This generated logs and metrics for CPU usage. Finally, I ran a high amount of web requests to my servers. I did this to make sure that Kibana was correctly monitoring and logging this data.
* What kind(s) of data did you investigate?
  + I investigated logs and metrics.

* How much data were you dealing with?
  + In comparison to what data analysts are dealing with it was nothing. However, I was looking over the data from 2 virtual machines.
* What were you looking for?
  + Specifically, CPU usage, memory usage, and streamed live logs looking for failed login attempts.

1. Explain the Solution Requirements

* What information did you need to find what you were looking for?
  + I was looking to see if I had put my VM under stress. I checked for this by looking at the metrics for CPU usage.
  + I was looking to see if I could find failed login attempts for a failed SSH connection.
* What does an analyst need to analyze large amounts of log data to find this information?
  + Analysts need to analyze large amounts of data so that they can know what is normal and what is abnormal. Every company has a different normal and abnormal.
* In Project 1, what tools did you use to analyze log data?
  + I used Kibana as a tool to analyze log and metric data.

1. Explain the Solution Details

* How did you use these tools to find the log data? E.g., which charts, graphs, etc. were useful for parsing the logs?

1. Identify Advantages and Disadvantages of the Solution

* What kinds of data did you not inspect during Project 1?
  + I did not analyze the key performance indicators. This helps with determining how well the machine is performing.
* Would having access to this additional data have changed your process or conclusions? If so, how?
  + This may have helped to confirm that my machine was connected, but it would not have changed my process or conclusions.