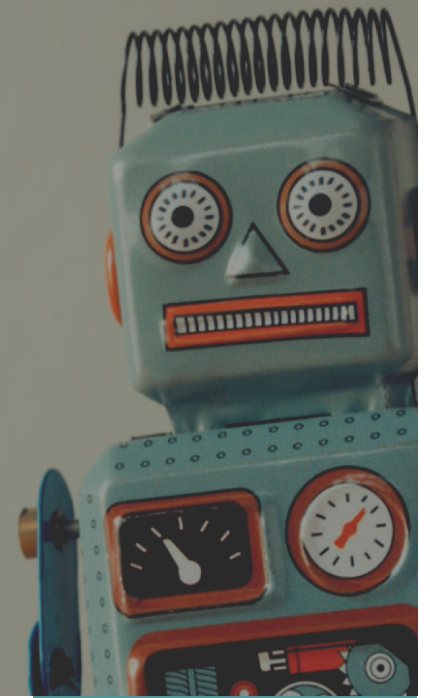# *INTERVIEW ASSIGNMENT*

Cloud Business Office, VMware

ROOT CAUSE ANALYSIS FOR AWS EVENTS

#BIG DATA
#DATA SCIENCE
#PUBLIC CLOUD
#PROGRAMMING

Introduction:

Large enterprises have hundreds of public cloud accounts and thousands of users with millions of dollars in costs. Company X is aiming to transform traditional public cloud monitoring by adding a layer of intelligence that filters the noise. Public cloud usage and costs monitoring create millions of data points every day. Company X is building a homegrown tool that aims to understand what data points are relevant/actionable and creates an alert. Given the complexity of the public cloud, it is not always easy to understand the reason for a cost spike. CloudBot will help break down underlying data to understand the reason for the sudden spike to help the user understand the next course of action. There is hardly a day where usage of the previous looks like today. This means that there will be plenty of variances that may not necessarily be important for the user to know about.

CONT.

Where do I start?

As a starting point, you can think of events that occur within an account (lineitem_accountid), per AWS service (product_productname). Use lineitem_unblendedcost and lineitem_usageamount columns as a numeric measure for variance in each row/datapoint.

Tip: You can use perspective_servicename - which is a business grouping within VMware to see if accounting under a certain type of perspective behaves similarly. Further perspective_COGS is an indication of an account being in production or under development. They both will behave quite differently.

What questions should I answer?

Your assignment is to think through this problem, review the sample data set, and document your thought process for a solution.

Think along the lines of:

1. How would I define an event for a particular account/AWS service combination? Is it a percentage change from yesterday to today? Can I learn from history what that percentage should qualify as an event?

    1. How do I communicate the root cause of an event in a simple manner? What algorithms would help me identify the root cause? Can It be multiple things or one single thing that is the root cause for an event?

    2. Am I identifying too many events in a day? Can all of them be useful to the user? If not, how do I ensure I provide only useful event alerts going forward? (Tip: feedback loop)

What is expected of me during the interview?

You are not expected to have a fully thought-out solution during your interview. We are looking to understand your thought process. Please feel free to ask clarifying questions, it helps us understand what you're thinking. The sample data set is only to help you understand the kind of data that is available to solve the problem. You are not expected to run any models on it. However, If you would like to document your thoughts about the problem as a document, on a whiteboard or presentation, that would be fine (not expected but welcome).

Most importantly, it is the interviewer's job is to set you up for success and try to bring out the best in you during your interview. Relax and bring a clear mind, the rest will follow.