Q1: What is the difference between EBS and S3?

S3 is a storage repository, and is primarily for archiving and storage to avoid purchasing the hardware to do the safe and secure object storage of any amount of data. S3 will automatically archive data that isn’t accessed very frequently to lower cost.

EBS is meant to be used as a computing hard drive for data that is actively being used by AWS EC2 computing instances. For this reason, it is particularly useful for database-intensive applications that frequently encounter many random reads and writes across the dataset.

Q2: When would you consider a cloud infrastructure for your data science tasks?

If you had a large amount of data that was frequently accessed and you don’t wish to bog down your local hard drive with stored data for applications that might upscale very quickly and/or unpredictably. It is also useful if you want to maximize a customer experience or ensure separation of computing and storage and provide data redundancy, so that no one failure loses data.

Q3: What is the difference between spot instances and reserved instances?

Spot instances are compute capacity purchased from Amazon AWS with a set maximum hourly price when needed for specific applications with no up-front costs or commitments. Spot instances are for specific needs rather than a steady demand, and it is essentially purchasing excess computing capacity on an as available basis for a discounted price. If the market price goes above the set maximum, AWS will shut the instance down until resumed later when the price goes back down.

Reserved instances are ideal for steady, predictable demand. Processing capacity is guaranteed for a specific zone and instance type for a set hourly price in advance, at a discount from the on-demand rate. Ideal for predictable, steady applications that can’t afford to have any gaps in processing capacity.

Q4: List the names of four software packages installed on the UCB AMI

Python

Ruby

Perl

Java