1. **What is Big Data? Give a brief definition.**

Big Data refers to a collection of data that is either too large, or too complex to be stored, processed, and analyzed in a traditional way.

Oracle.com refers to it as:

> "...the incredible amount of structured and unstructured information that humans and machines generate—petabytes every day, according to PwC. It's the social posts we mine for customer sentiment, sensor data showing the status of machinery, financial transactions that move money at hyperspeed. It's also too massive, too diverse, and comes at us way too fast for old-school data processing tools and practices to stand a chance."

2. **What are the traditional 3 Vs of big data? Briefly define each.**

Volume - The quantity of data being stored. As the volume grows systems need to increase their storage as well, leading to them having to choose between scaling up, or scaling out.

Velocity - The rate at which new data is entering the system, as well as the rate it must be processed at. With so much raw data the amount entering a system and the processing that needs to be done on said data grows astronomically.

Variety - The different formats and structures in which the data may be captured. The data could be well structured from forms, unstructured , or semi-structured.

3. **Explain why companies like google and Amazon were among the first to address the big data problem.**

Companies like Google or amazon were among the first to address big data for a multitude of reasons. For one they were one of the first largest companies who ran into the issues of needing to handle big data. As an example for google they wanted to address the issue of indexing the internet to help with their search engine and results. Another reason is they were one of the first companies who had the means to collect and process this data, with the growing amount of people using these products every day, thay had a large amount of data being entered into their systems. And finally there is money. As two of the big tech giants they have the money to use for these projects, where smaller companies would not be able to fund these projects at the same scale.

4. **Explain the difference between scaling up and scaling out.**

Scaling up refers to adding more compute resources to your system. This could be better CPU, more RAM, or more disk capacity.

Scaling out refers to addressing the same limitations but instead of adding more compute power to the system, incorporates more nodes into the system to split the work loads up into more manageable jobs.

5. **What is Stream processing and why is it sometimes necessary?**

Stream processing is the act of collecting, analyzing, and processing data in real time as it is produced. Stream processing allows companies to gain a wide range of data on customer and business activity as it is happening, handling the large amounts of volume from multiple sources at various speeds

6. **How is Stream processing different from feedback loop processing?**

Stream processing is designed and focused on continuous and immediate handling/ processing of data. The difference here is that feedback loop processing behaves in a more iterative way, creating feedback, responding to it, then putting it to use and continuing in a cycle until the objective is achieved.