3. Explain why companies like Google and Amazon were among the first to address the Big Data problem. (10 points)

Because they were the first companies to grow into a large online presence, they were the first companies to have a need to address Big Data, and because they continue to grow as a large online presence, they will likely be the first to address Big Data when the definition of Big Data changes.

4. Explain the difference between scaling up and scaling out. (10 points)

Scaling up involves improving the specs of current hardware (replacing storage with larger capacity devices, installing higher performance processors and busses, getting better servers). Scaling out involves adding more hardware (divvying data storage and computations amongst more servers).

7. Explain why veracity, value, and visualization can also be said to apply to relational databases as well as Big Data. (10 points)

Veracity is important no matter how much data you have. If your data isn’t trustworthy, you might as well not even have it.

In relational databases, the value of the data determines whether or not you chose to store that data. Useless data is useless.

While visualization may be more difficult in Big Data, it’s still important in relational databases to better understand how the data is organized and reveal possible design flaws.

17. Explain why graph databases tend to struggle with scaling out. (10 points)

Data isn’t organized or collected into groups around central topics. This means related data is more likely to be physically separated within a scaled-out environment, making aggregation more difficult.

18. Explain what it means for a database to be aggregate aware. (10 points)

The database choses a single, central entity around which it collects data.