1. **Why would you want to use the Data API?**

A.

What are the benefits of splitting a large dataset into multiple files?

1. During training, how can you tell that your input pipeline is the bottleneck? What can you do to fix it?
2. Can you save any binary data to a TFRecord file, or only serialized protocol buffers?
3. Why would you go through the hassle of converting all your data to the Example protobuf format? Why not use your own protobuf definition?
4. When using TFRecords, when would you want to activate compression? Why not do it systematically?
5. Data can be preprocessed directly when writing the data files, or within the tf.data pipeline, or in preprocessing layers within your model, or using TF Transform. Can you list a few pros and cons of each option?