Questions:

1. What is TensorFlow? Which company is the leading contributor to TensorFlow?
    a. *TensorFlow is a free and open-source software library for machine learning.*
    b. *Google Brain team. Originally for internal use, later released under Apache License 2.0 in 2015*
2. What is TensorRT? How is it different from TensorFlow?
    a. *The core of NVIDIA® TensorRT™ is a C++ library that facilitates high-performance inference on NVIDIA graphics processing units (GPUs). It is designed to work in a complementary fashion with training frameworks such as TensorFlow, Caffe, PyTorch, MXNet, etc. It focuses specifically on running an already-trained network quickly and efficiently on a GPU for the purpose of generating a result (a process that is referred to in various places as scoring, detecting, regression, or inference).*
    b. *TensorRT is a runtime optimization that exists on top of tensor flow models, such that we can achieve higher performance on NVIDIA trained devices.*
3. What is ImageNet? How many images does it contain? How many classes?
    a. *The ImageNet project is a large visual database designed for use in visual object recognition software research.*
    b. *More than 14 million images have been hand-annotated by the project to indicate what objects are pictured and in at least one million of the images, bounding boxes are also provided.*
    c. *ImageNet has 1000 classes*
4. Please research and explain the differences between MobileNet and GoogleNet (Inception) architectures.
    a. *The Inception module computes multiple different transformations over the same input map in parallel, connecting the results into a single output. For each layer, it does a 5x5 convolution, 3x3 convolution, and max pooling, each carries different information, which of course is computationally costly. Therefore the authors of Inception decided to overcome this problem by introducing the dimension reductions.The idea behind MobileNet is to use depth wise separable convolutions to build lighter deep neural networks. In a regular convolutional layer, the convolution kernel or filter is applied to all of the channels of the input image, by doing weighted sum of the input pixels with the filter and then slides to the next input pixels across the images.*

Reference:

https://en.wikipedia.org/wiki/TensorFlow
https://docs.nvidia.com/deeplearning/tensorrt/developer-guide/index.html
https://blog.tensorflow.org/2019/06/high-performance-inference-with-TensorRT.html
https://en.wikipedia.org/wiki/ImageNet
https://towardsdatascience.com/an-intuitive-guide-to-deep-network-architectures-65fdc477db41
https://medium.com/@fransiska26/the-differences-between-inception-resnet-and-mobilenet-e97736a709b0