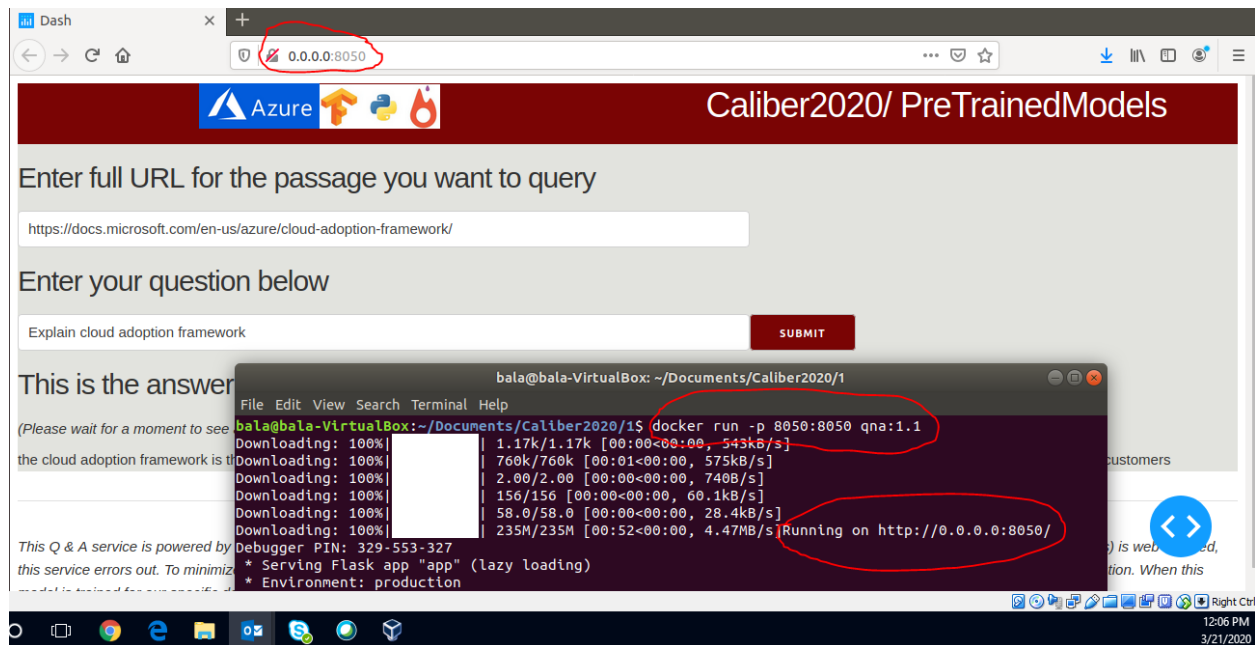


Screen shot of webapp running in a Docker container:



Typical business context:

Pointed questions about a System Integrator's (SI) capabilities/ products need to be answered during deal pursuit. For example, which RDBMS is supported by one's healthcare insurance Core Administration platform. To answer such queries, SI have a dedicated a "Pursuit Advisory" team. But, even for this team, it takes significant effort to find the right answer.

Wider challenge:

Such reading comprehension challenges exist across the world and is the very subject of the SQuAD challenge by Stanford University. According to the [SQuAD website](#),

What is SQuAD?

Stanford Question Answering Dataset (SQuAD) is a reading comprehension dataset, consisting of questions posed by crowdworkers on a set of Wikipedia articles, where the answer to every question is a segment of text, or span, from the corresponding reading passage, or the question might be unanswerable.

SQuAD2.0 combines the 100,000 questions in SQuAD1.1 with over 50,000 unanswerable questions written adversarially by crowdworkers to look similar to answerable ones. To do well on SQuAD2.0, systems must not only answer questions when possible, but also determine when no answer is supported by the paragraph and abstain from answering.

Solution approach:

The use of transfer learning with pre-trained models for vision and natural language processing is a well-established approach that drastically reduces training time and the training dataset needed. Google

Research's [BERT](#) (Bidirectional Encoder Representations from Transformers) was a breakthrough model that is described as,

BERT is designed to pretrain deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context in all layers. As a result, the pre-trained BERT model can be finetuned with just one additional output layer to create state-of-the-art models for a wide range of tasks, such as question answering and language inference, without substantial task specific architecture modifications.

The leaderboard for the SQuAD challenge shows [ALBERT](#) is currently (March 2020) the dominant solution architecture. ALBERT is an improvement on BERT aimed at reducing the number of parameters and hence the training time needed for BERT.

An ALBERT model, finetuned on the SQuAD2.0 dataset described above, has been open sourced by Hugging Face [here](#).

It is clearly understood that we have to finetune an ALBERT pretrained model on the specific Q&A dataset we wish to query.

