Module 8 - Lesson 8: DASK Hands On

Pick at least two sections under USER INTERFACE and SCHEDULING to examine. Please answer the following questions after examining these features:

1. What do you think the most useful DASK feature is?
   Users sometimes just use the internals of DASk for the benefit of the Scheduler where it provides users the ability to create their own custom computations.

   What I believe to be the most useful aspect of DASK is that it just enhances the capabilities of pandas, numpy and scikit learn libraries. These three libraries are used to analyze, wrangle, and visualize data, in addition to providing machine learning coding. DASK enhances the capacity of data being processed through the same code.

   DASK simply needs to be installed through pip install or conda install in order to be available for us. As DASK already has pandas, numpy, and scikit learn “under the hood.” DASK is also typically already installed through Anaconda.

   With Python being the leading language in data science and one of the top languages for many in the tech field, DASK facilitates the processing of big data without having to learn new lingo, language, or process for processing.

2. Why is the advent of DASK so important?
   As stated above, it facilitates the processing of big data, without having to learn a new system. DASK enhances the capacity of well know libraries/packages in Python that many already use for data science and analytics and DASK is just an added package for use, which provide for more effective and efficient processing of data.

   DASK provides parallel computing which can be done through a single computer or connected through the cloud or other cluster. Parallel computing allows for large problems to be divided into small ones that can be solved at the same time. The facilitation of connecting DASK to these different types of data outputs and be able to break it down for processing is extremely beneficial and enables performance at scale for big data.

3. What would you like to learn more about?
   I would like to learn more about the ease in using DASK and how it scales down big data for easier processing and analyzing. What are the optimal uses of DASK and how can it efficiently and effectively facilitate my work when implementing it into my python code processing.