Topic: the role "environment" plays in a data scientist's workflow

Establishing python environment can be a very critical part of workflow for data scientists. As the author of first article point out that there are a couple of situations that we need to build different python environments. For example, when certain applications, which used to be working perfectly, stop running right now, or when we need to run our applications on other team members' computer, or when we are delivering applications to our clients. There are two most popular python package managers: PIP and Conda. To install Conda, there are total 3 installers: Anaconda, Miniconda, and Anaconda Enterprise platform. Each one of these has its pros and cons. In addition, when managing our python working environment, we have to take into consideration of three other dimensions, such as operating system, bit-counts, and python version.

There is another tool called Docker, which is very useful to launch multiple data science environment. The technology that Docker exploits is called containerization. Containerization can allow applications to run in an isolated environment independent of the operating system (like a virtual machine). Moreover, Docker allows data scientists to download and run images "which contain a preconfigured lightweight OS, set of libraries, and application specific packages" according to second article. Some advantages of using Docker includes the ability to get stated quickly, consistent across platform, and ability to checkpoint and restore.