After doing some researches, the topic that caught my attention is "Efficient development of high performance data analytics in Python", whose authors are Javier Álvarez Cid-Fuentes, Pol Álvarez, Ramon Amela, Kuninori Ishii, Rafael K. Morizawa, Rosa M. Badia. The purpose of the research is to introduce a framework for python called PyCOMPSs. As I am also studying python, one of the most popular programming languages, I think it would be beneficial to explore different frameworks of this language. According to the authors, high performance computing (HPC) programming models lack productivity and unable to process a large amount of data. Thus, big data processing tools are not completely compatible with current HPC environments. As a result, the authors recommended PyCOMPSs as the better solution to big data processing in HPC. By developing two machine learning algorithms, Cascade SVM and K-means, with PyCOMPSs the authors evaluated the efficiency of PyCOMPSs based on these algorithms when working on an HPC cluster with 1536 cores and 320 million input vectors. The results showed that PyCOMPSs not only achieves similar performance and scalability to Message Passing Interface (MPI) in HPC infrastructure but also provides a better productive interface that helps promote the development of data analytics algorithms. Therefore, PyCOMPSs can be used in any research areas like genomics, high energy physics and astronomy when large-scale data processing is undoubtedly important.