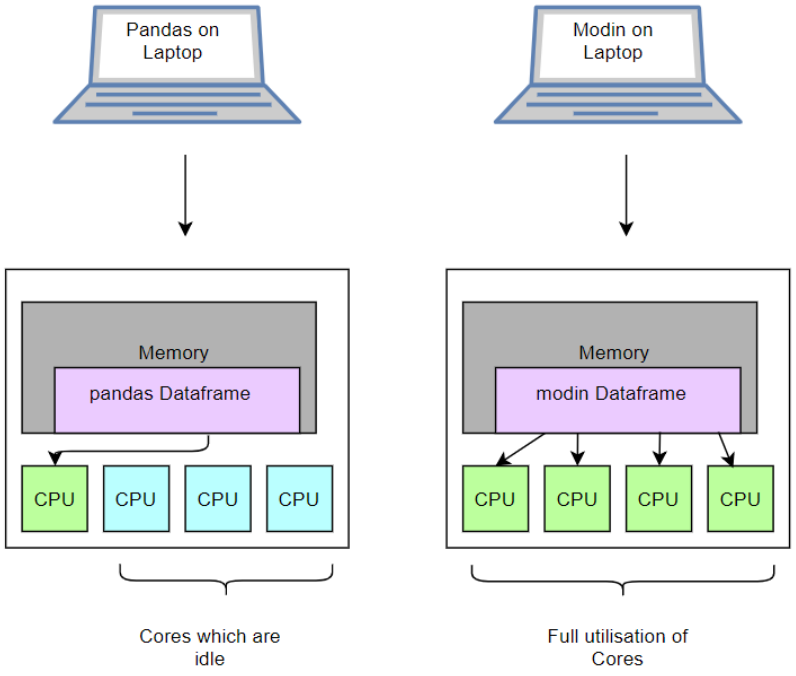
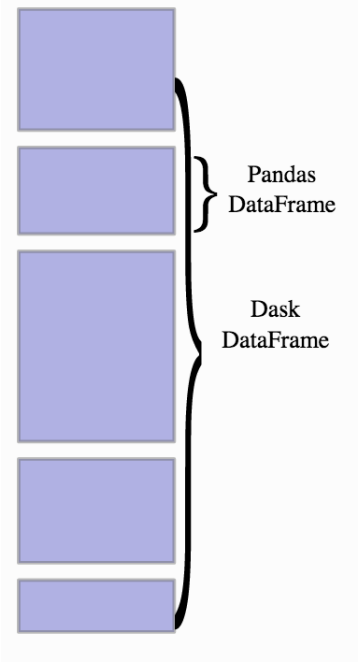
Report

1. **Pandas:** Its development started in 2008 and by the end of 2009 it was declared as an open source and is maintained by a community and regularly updated.
   Pandas is used for data analysis and is a strong library and is highly optimized as it's back-end code is purely written in C and python. Pandas uses only a single core CPU. This library is good in reading small datasets like few 100 mb's but as the size goes big it's hard to process with a single core CPU and processing time increases drastically. Even your PC/laptop has multiple cores pandas simple can't use it.
2. **Modin:** Is a python library that is used to deal with large datasets using parallelization. Its syntax is similar to that of pandas and is better than pandas in processing large number of datasets. Modin uses multiple core CPU's for its processing which makes it more efficient than pandas.



3. **Dask:** A flexible library used in parallel computing in python. It's useful in reading large datasets. The main advantage dask has over pandas is that it can also use all of the CPU cores.
   Dask introduces 3 parallel collections that can store data even bigger than RAM, namely Dataframes, Bags and arrays. Each of these types can use data partitioned between RAM and hard disk. Dask data frame is made up of smaller split up Pandas data frame and therefore it allows a subset of pandas query matrix.



4. **Datatable:** Its mimic library of data.table in R and is widely used to read and process large data frames. The main advantage of datatable over all libraries' is that it can use all the CPU and GPU cores. Many people are not friendly with the syntax of datatable so they can read the dataset with datatable then convert it to pandas for further processing, but for overall efficiency its recommended to work in datatable only.