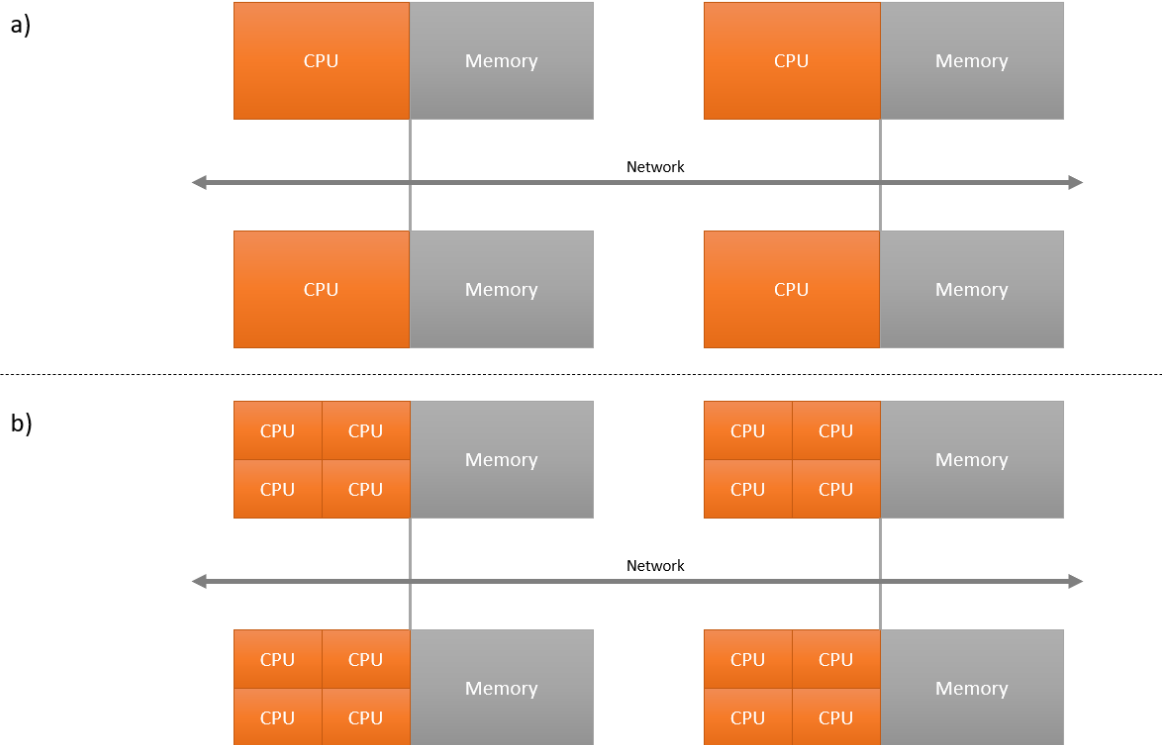


### Big data processing under a parallel programming model on the cloud

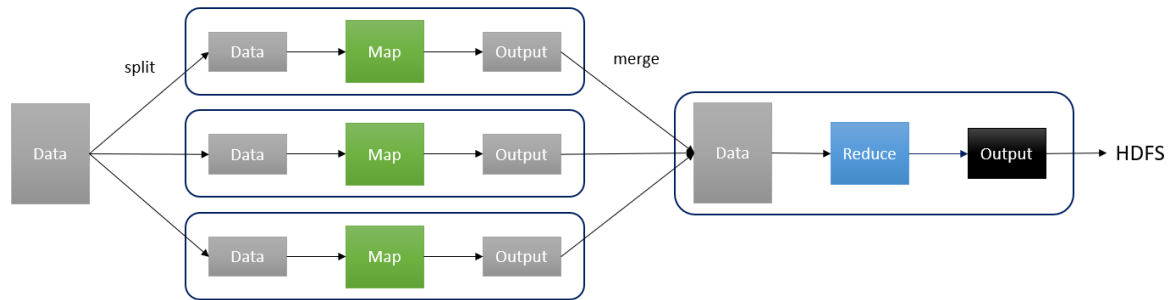
Since the 1980s, one problem that the digital data had is the distribution of information to handle fast and efficient searches. Until then, the data processing was done in a single thread, which limiting the total execution time. Thus, a multi-thread model was proposed to reach this efficient searching process. An example of this is the Message Passing Interface (MPI), this addresses the message-passing parallel programming model: data is moved from the address space of one process to that of another process through cooperative operations on each process.



1. Data parallelism process. a) MPI original architecture.  
b) MPI architecture modified for distributed systems.

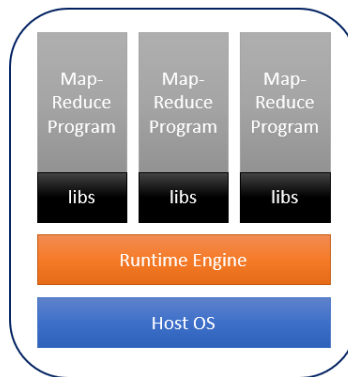
Data parallelism entails partitioning a large data set among multiple processing nodes, with each one operating on an assigned chunk of data, before participating in the process of combining the partial results.

On Big data, a parallelism is done on the Map-Reduce process. The main objective of this process is to let an efficient management of the information. On this, the data is split into different “threads” and executed on parallel to do the Mapping. When the process on each thread is completed, they converge into a single thread to do the Reduce process.



## 2. Map reduce process.

To implement this data parallelism into the cloud, the Containerization process can be used to upload each thread inside a container. These containers will be connected to a runtime engine that will execute the data parallelism.



## 3. Map reduce containerized to the cloud.