Hadoop is a framework that allows us to store and process large data sets in parallel and distributed fashion. HDFS is the storage layer of Hadoop Ecosystem, while MapReduce is the processing layer of the ecosystem. All the data in Hadoop is stored in Hadoop Distributed File System.

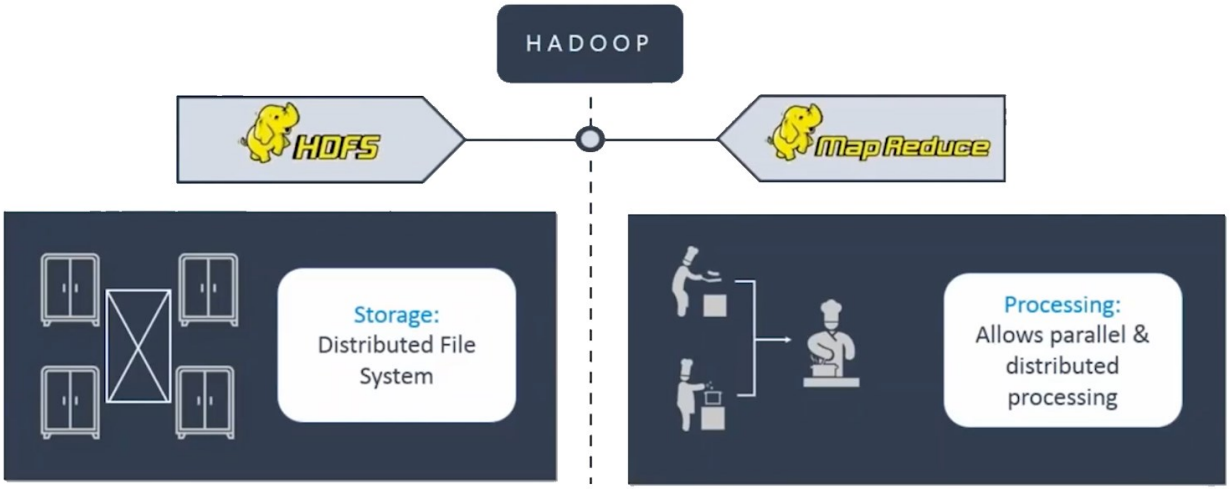

Figure 1: Hadoop, HDFS and MapReduce

For example, Let us suppose their there is multiple orders in a restaurant but there is only one chef and one food shelf. To fulfil all the orders very quickly, it requires multiple food shelf and multiple chefs otherwise it will be very time consuming for one chef. Distributing multiple food shelf for the chefs refers to HDFS and Cooking all the foods at the same time by multiple chefs refers to MapReduce. HDFS means distributed file storage systems and MapReduce means parallel and distributed processing system.

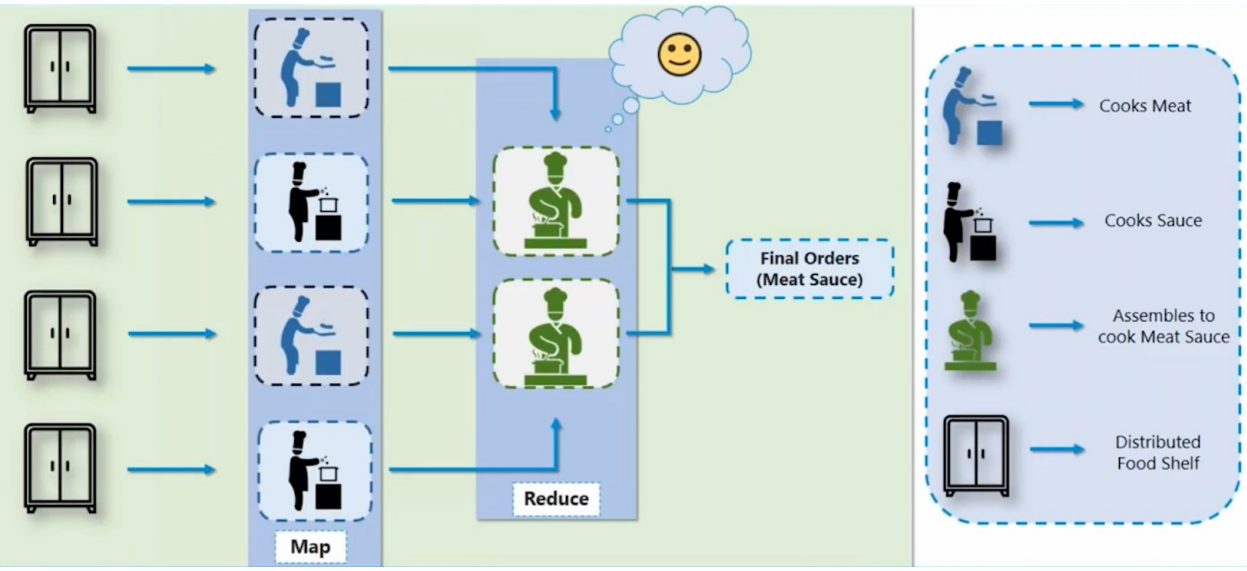

Figure 2: Distributed and Parallel Approach for Food Orders