

## **Report on MapReduce: Simplified Data Processing on Large Clusters**

This paper demonstrates a MapReduce implementation that runs on a large cluster of computers and is extremely scalable. On Google's clusters, MapReduce tasks are used on a daily basis. The author explains a typical MapReduce computing process on thousands of computers in detail. The basic goal of such calculations is to complete the computation of massive input data that is distributed among hundreds or thousands of computers in an acceptable period. The major influences for this work are the Map and Reduce primitives found in Lisp and many other functional languages. They demonstrate the issue with calculations by using a map operation to each logical "Record" in the input to compute the key/value pairs in this article. This work presents a simple and powerful interface for automated parallelization and distribution of large-scale calculations on huge clusters of commodity PCs, resulting in great performance.

By automating the partitioning of the input data into net pf splits, these map invocations are scattered over a large number of computers. These splits are produced in parallel. The master is one of the divided program's copies. The remainder are assigned to be slaves who carry out the master's orders. The memory is buffered with the key/value pairs created in the intermediate. This is then written to the local disk on a regular basis. When the master notices that one of the workers is working less, he calls remote procedure calls to read the buffered data. The reduction function's output is added to an output file. When all of the steps have been performed, the master activates the program.

In a nutshell, Google's map reduction tool is extensively utilized for a variety of reasons. The approach is simple to use even for programmers with little prior expertise, which is one of the key reasons for its widespread adoption. Another important point raised in the study is that the MapReduce program may simply translate any complicated computation into a MapReduce process. The authors have created a MapReduce solution that can handle massive clusters of computers with thousands of machines. Because its approach makes good use of these machine resources, it may be applied to a wide range of bigger computing tasks at Google.