Kieren Singh Gill
ksg7699

MapReduce Paper Summary

MapReduce was created by Jeffrey Dean and Sanjay Ghemawat, and they decided to release this paper to share their findings and to make it public. MapReduce is a model that is used for the processing and generating of large datasets. It can be broken down into multiple steps - firstly, the map function creates key/value pairs. This generates another set of intermediate key/value pairs, and the reduce function will merge all intermediate values that are associated with the same intermediate key. The beauty of MapReduce lies in the fact that it is automatically parallelized (which is when many calculations are carried out simultaneously) and it is executed by many commodity machines (cheap or already existing hardware). Hence, it is highly scalable. Furthermore, the system partitions the input data, handles machine failures, and manages inter-machine communication, meaning that programmers with no experience can easily use the system.

The Friendship That Made Google Huge

This article tells the story of Jeff and Sanjay, two of Google's founding engineers that have driven the company to the heights of success. From the very beginning, back to when they solved the flaw in Google's hardware by breaking down code into binary, Jeff and Sanjay stand behind pair programming - they work and code alongside each other to complement one another. In 2003, they came up with MapReduce, which simplifies the task of dividing data, diistributing data, assigning work, and account for hardware failures. With MapReduce, a coder tells the machine how to map,and the coder writes instructions to reduce the machines results. The machine handles the rest. Later on, MapReduce was recreated as Hadoop, and was quickly adopted by large giants like LinkedIn and Netflix. Google also came up with Tensorflow, the AI equivalent of MapReduce. Today, Jeff and Sanjay don't work directly alongside each other, but their work still complements each other while impacting the lives of many others.