Title of Paper: MapReduce: Simplified Data Processing on Large Clusters

Authors: Jeffrey Dean and Sanjay Ghemawat

The paper introduces MapReduce, a programming model that addresses the needs of big-data processing which masks the intricate details of parallel and distributed computing. Simple abstraction are provided for the programmer as users are only required to implement a map operation to compute a set of intermediate key/value pairs from input, and another reduce operation to combine the values that belongs to the same key. The programming system will then handle parallelization, fault-tolerance, data distribution and load balancing through an architecture of master and work nodes. MapReduce has been successfully used at Google for multiple purposes, and is easily expressible for a large variety of problems