549 Final Project

In this project, we will do several things. First, we will write a paper review on multiple MapReduce papers. Second, we will utilize Google Cloud to implement Map Reduce paper as well as run several Map Reduce examples on it. Third, we will utilize our own machine that implements a simple MapReduce framework

Part One: Paper Review

Due to the physics limitation, the computation power of a single machine is limited. Therefore, instead of focusing on increasing the computing power of a single machine, people start to research how to coordinate multiple machines together to finish the same job. MapReduce, developed by Google, is one of the most popular parallel and distributed computing models. It addresses the challenge of efficiently processing vast amounts of data by breaking down complex tasks into smaller, more manageable subtasks and distributing them across a cluster of interconnected computers.

1. MapReduce Structure

A diagram of a user program

Description automatically generated

MapReduce works on key/value pairs.d

Reference: