Making our MapReduce more scalable

Our implementation of MapReduce simplifies the original design because the workers only communicate with the controller, they do not communicate directly with each other. In a real MapReduce implementation, the output from the map phase is sent directly from the mappers to the reducers.

This requires the mappers to know which reducer should handle a given key. After performing the mapping computation, the mapper sends each (key, intermediate value) pair to the reducer responsible for the key. It then notifies the controller that it has completed.

Meanwhile, the reducers are collecting the results associated with each key that they are responsible for. When the controller learns that the map phase is entirely complete, it signals the reducers to begin the reduce phase on the intermediate values they have collected. They then perform the reduce computation and send the results back to the controller.

Although this is not an app, you may implement a more sophisticated MapReduce implementation in lieu of your elective app.