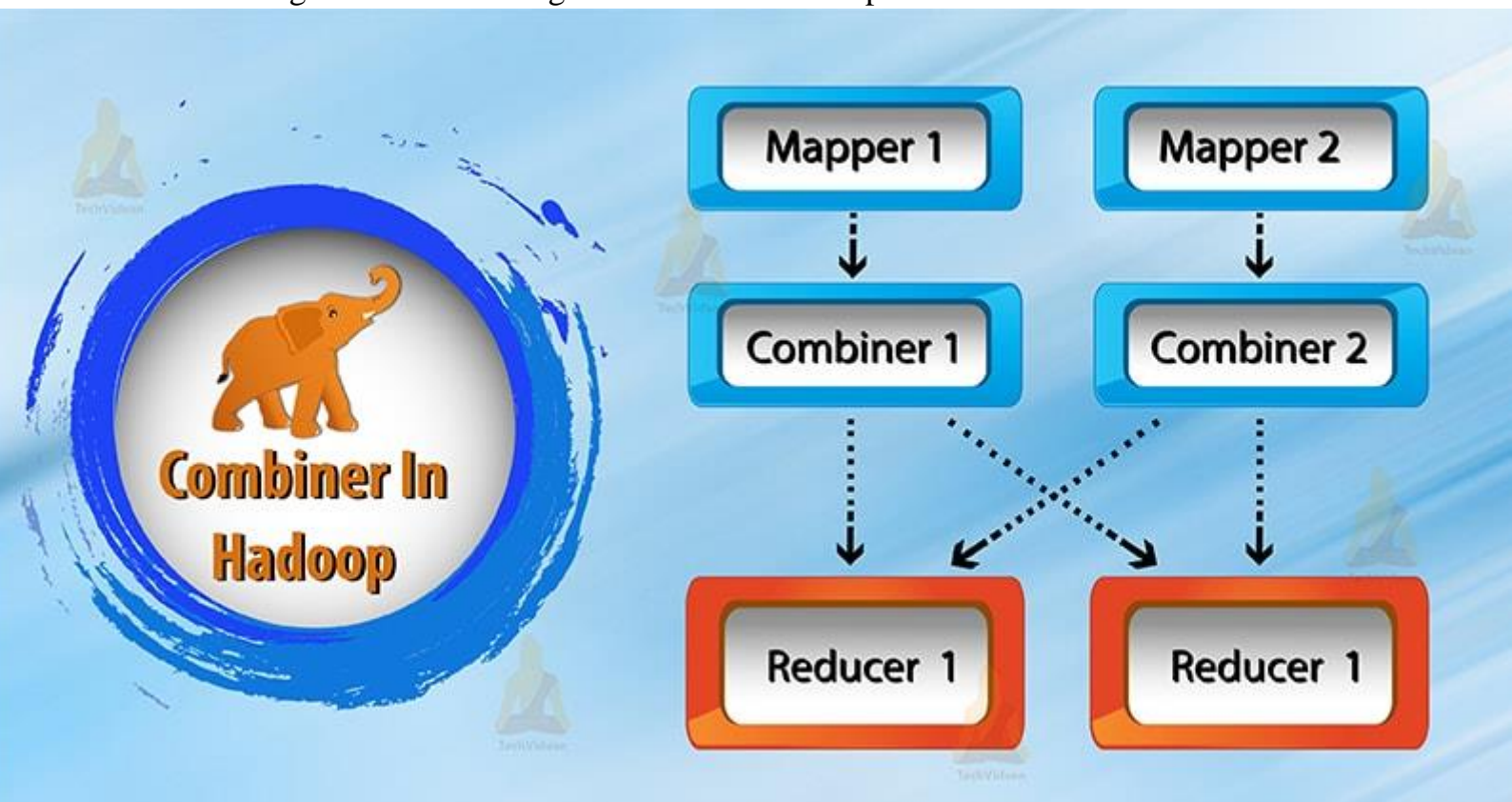


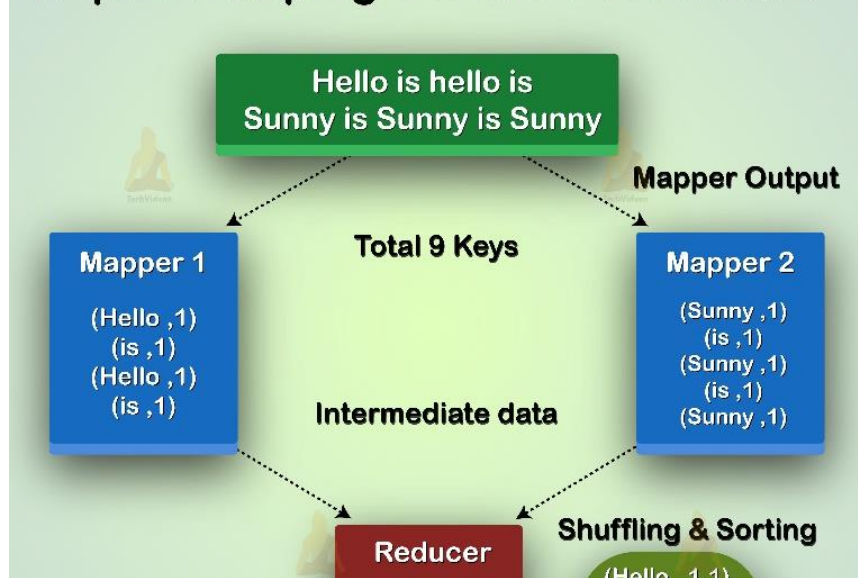
# Hadoop Combiner:

## 1. Objective

In this **Hadoop tutorial**, we will provide you a detailed description of Hadoop Combiner. First of all, we will see what is MapReduce Combiner, what is the key role of Combiner in MapReduce. Then we will discuss the example of MapReduce program with and without combiner in Hadoop. At last, we will also see some advantages and disadvantages of Combiner in MapReduce.



### MapReduce program without Combiner



As we see in above diagram no combiner is there. Input is split into two mappers. The framework generates 9 keys from the mappers.

So, now we have (9 key/value) intermediate data. Further mapper sends this **key-value** directly to the reducer. While sending data to the reducer, it consumes some network bandwidth. It takes more time to transfer data to reducer if the size of data is big.

Now from the above diagram, if we use a combiner in between mapper and reducer. Then combiner will shuffle 9 key/value before sending it to the reducer. And then generates 4 key/value pair as an output.

Now, Reducer needs to process only 4 key/value pair data which are generated from 2 combiners. Therefore reducer gets executed only 4 times to produce the final output. Thus, this increases the overall performance.

## 4. Advantages of Combiner in MapReduce

Let's now discuss the benefits of Hadoop Combiner in MapReduce.

- Use of combiner reduces the time taken for data transfer between mapper and reducer.
- Combiner improves the overall performance of the reducer.
- It decreases the amount of data that reducer has to process.

## 5. Disadvantages of Combiner in MapReduce

There are also some disadvantages of Hadoop Combiner. Let's now discuss the same.

- In the local filesystem, when Hadoop stores the key-value pairs and run the combiner later this will cause expensive disk IO.

- MapReduce jobs can't depend on the combiner execution as there is no guarantee in its execution.

## 6. Conclusion

Hence, Hadoop Combiner plays a key role in reducing network congestion. It improves the overall performance of the reducer by summarizing the output of Mapper. I Hope now you have a clear understanding of Hadoop Combiner. If still you have any query, so, please let us know be leaving a comment in a section below.