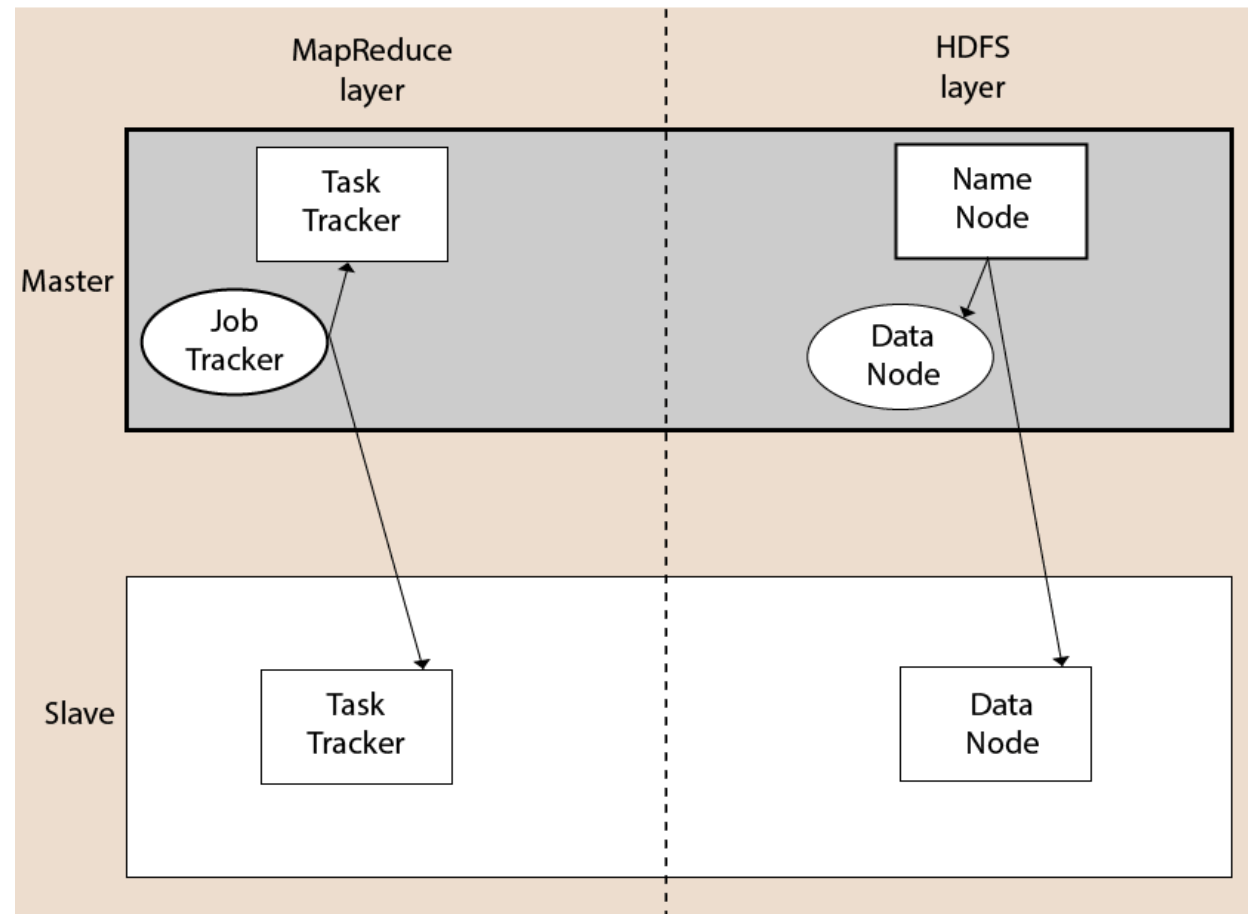


BML 621 – Cloud Computing for Big Data

Hadoop Framework

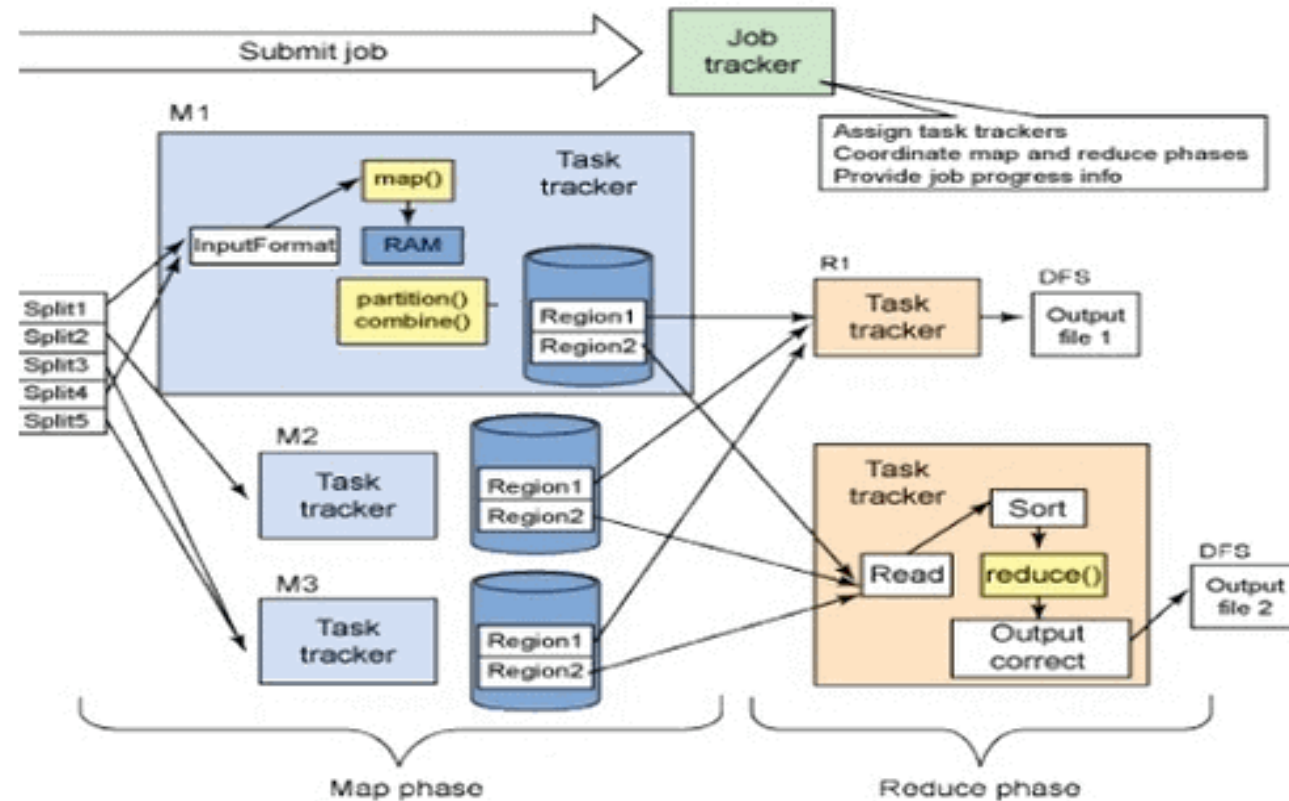


Recap



Recap

Hadoop Architecture



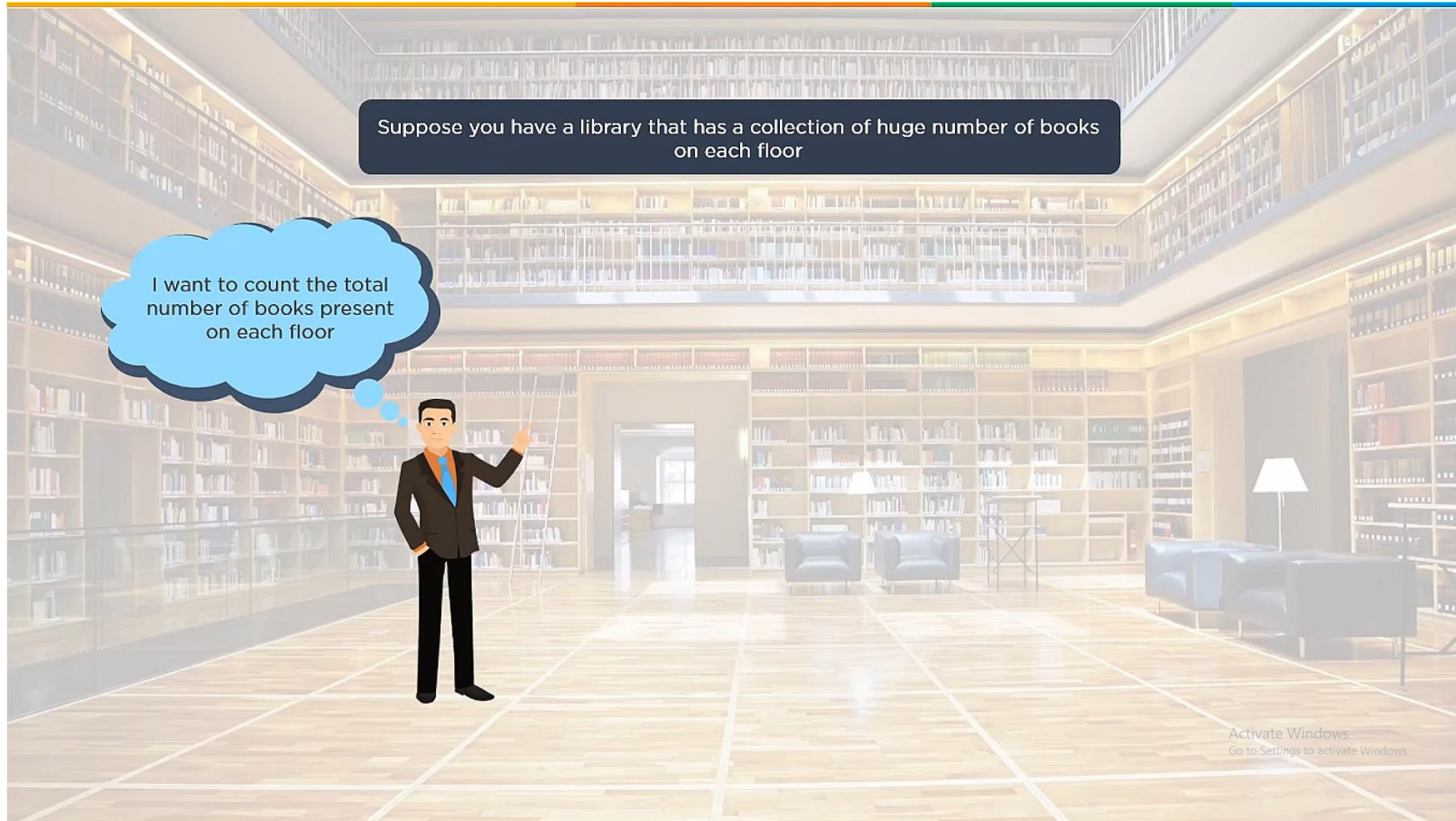
Hadoop

1. HDFS

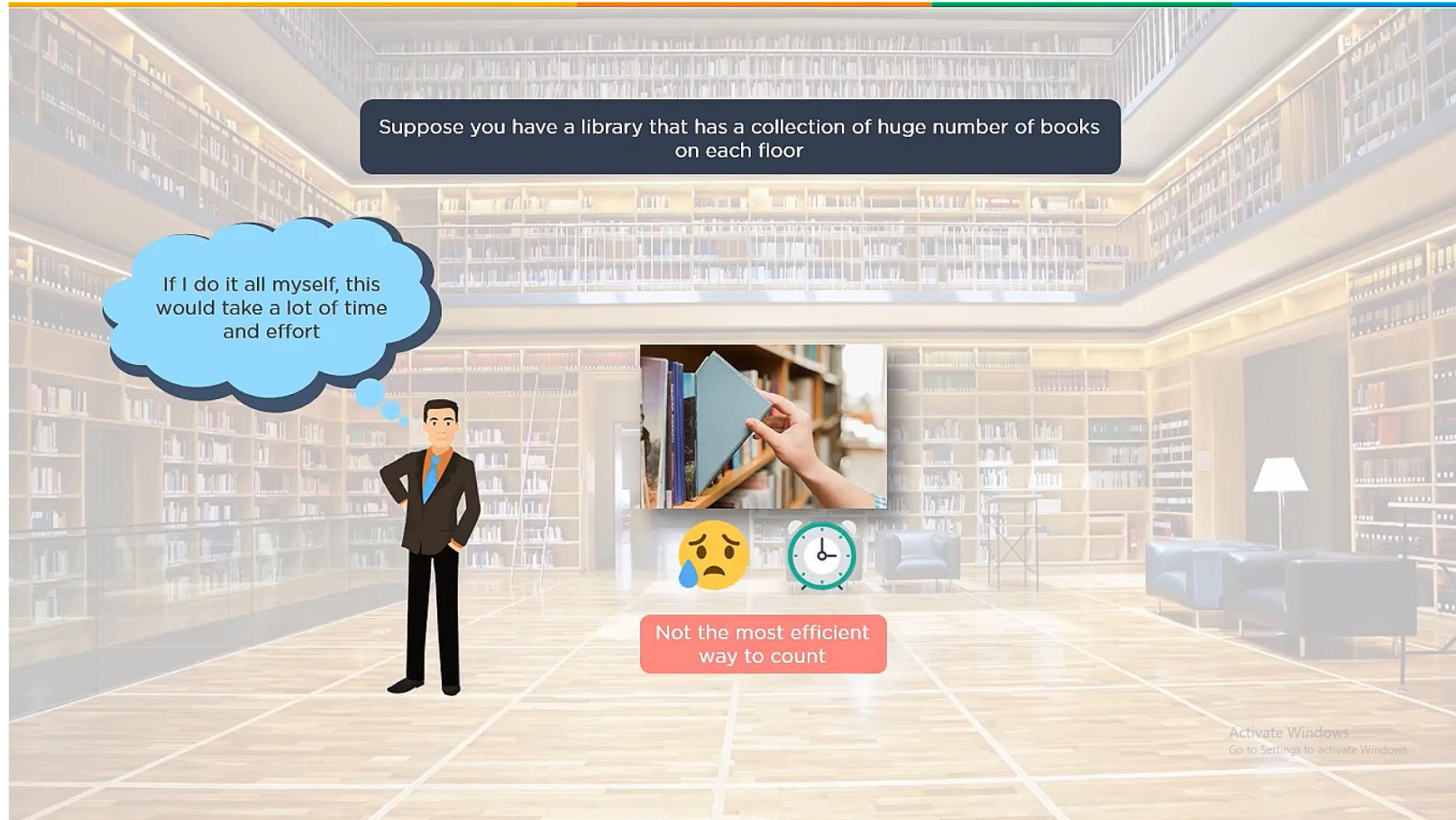
[HDFS architecture, Blocks, Replication, Rack awareness, Hadoop Architecture]

2. MR ?

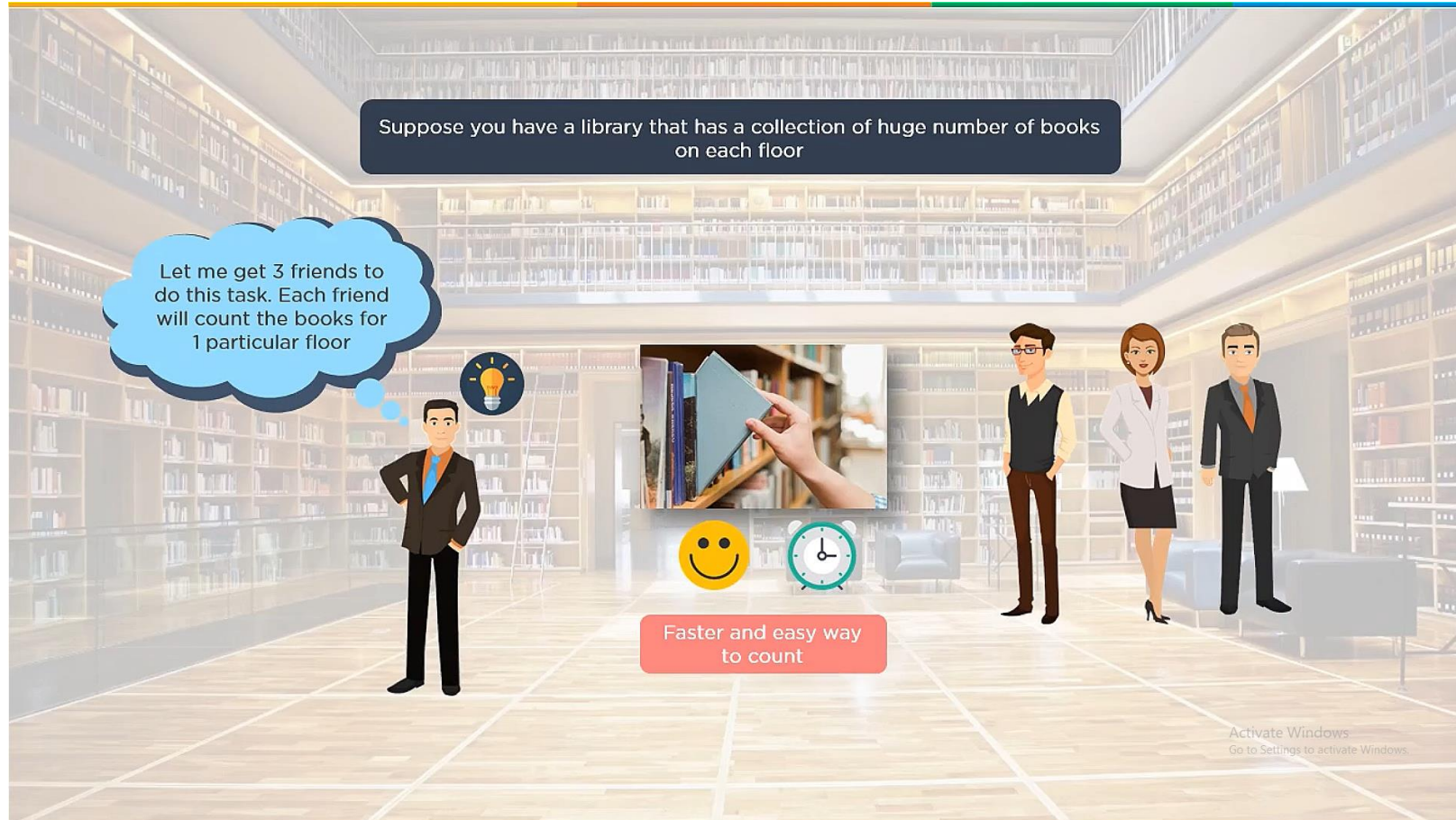
Apache MapReduce



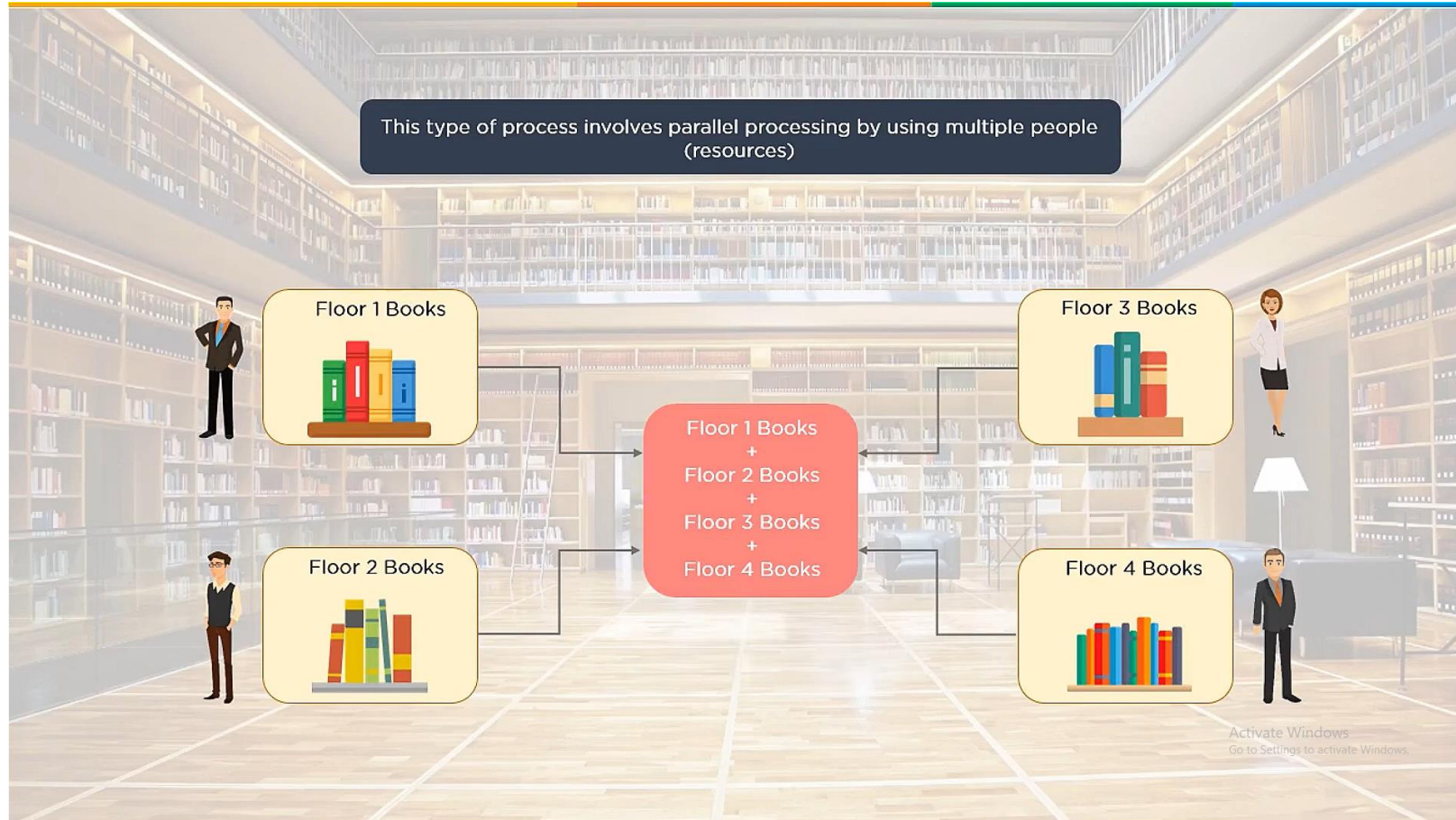
Scenario ... Contd.



Scenario ... Contd.



Scenario ... Contd.



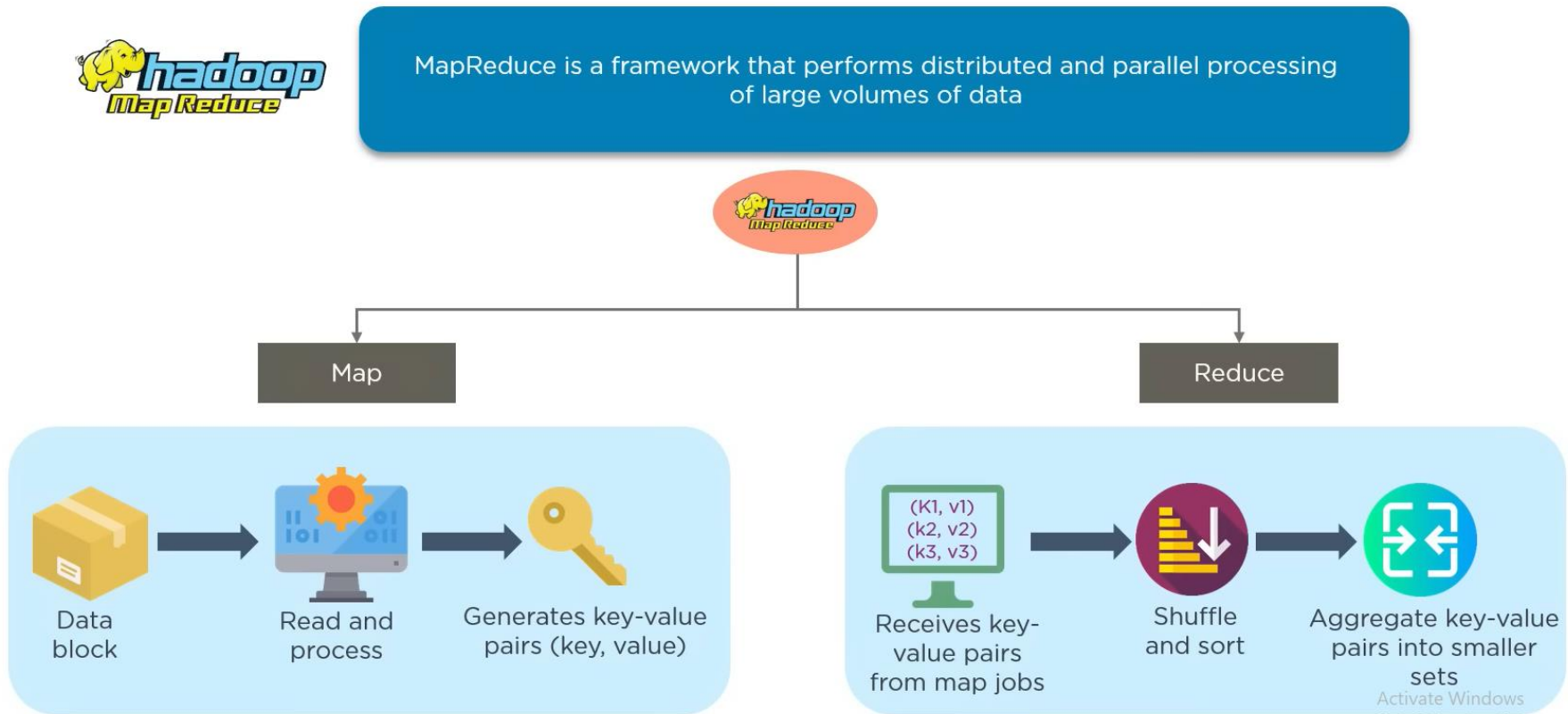
Scenario ... Contd.



Scenario ... Contd.



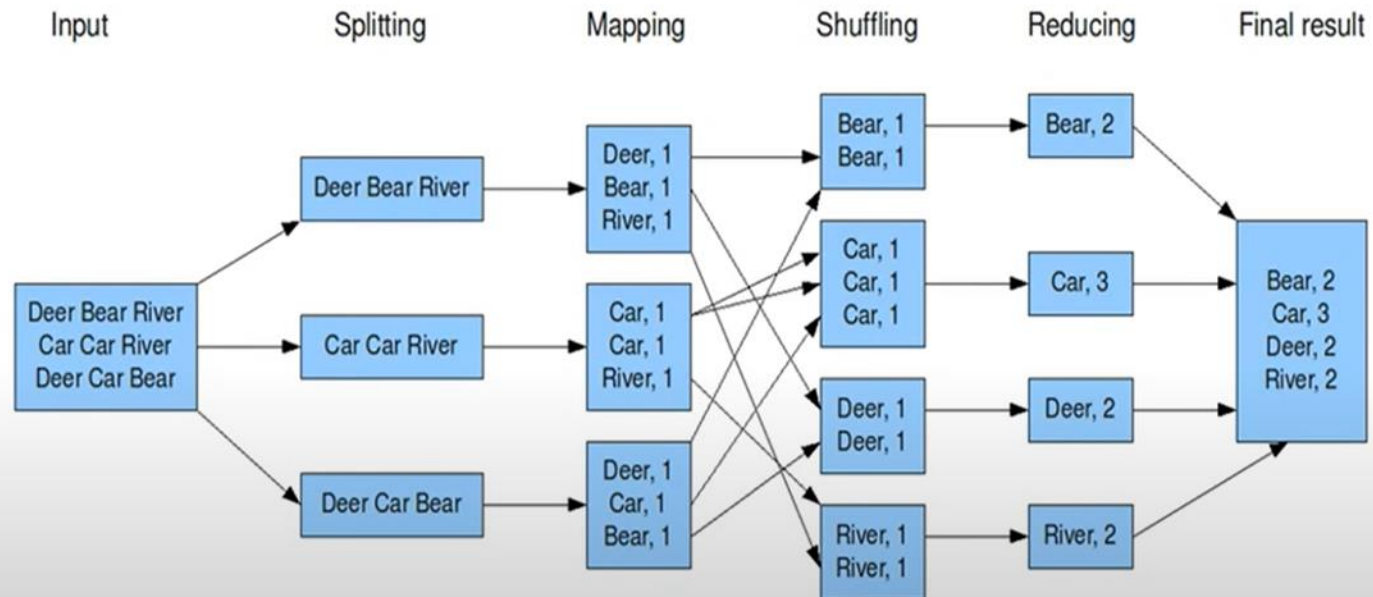
Hadoop - MapReduce



MapReduce - Example

Word Count Example

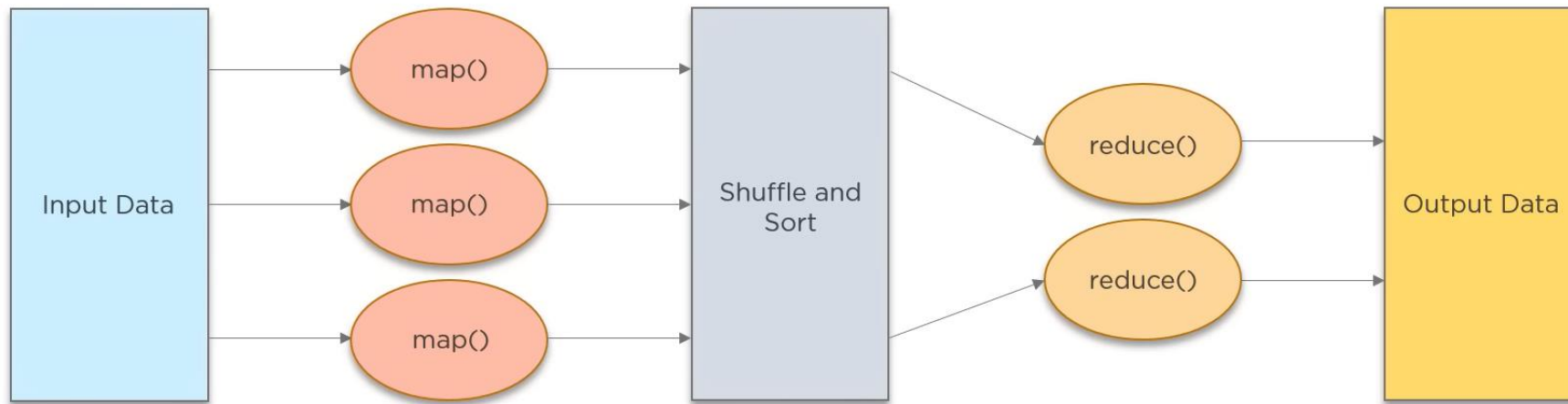
The overall MapReduce word count process



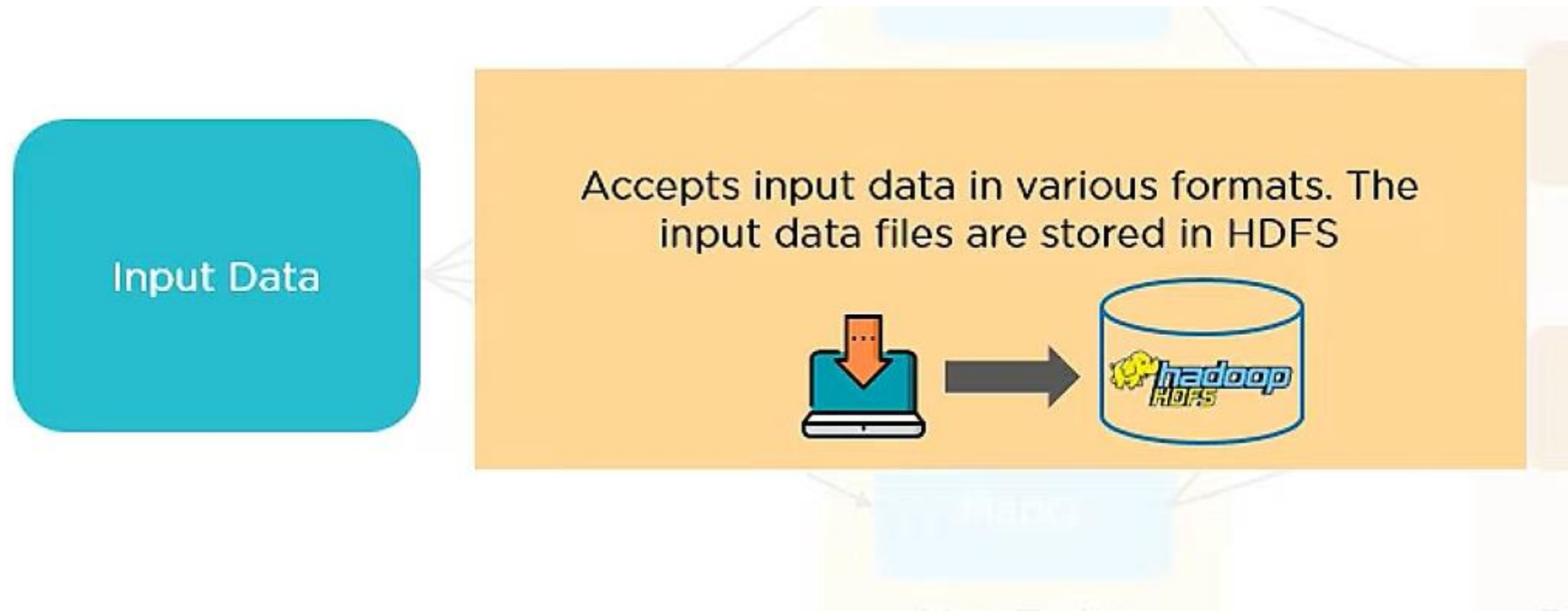
Hadoop - MapReduce



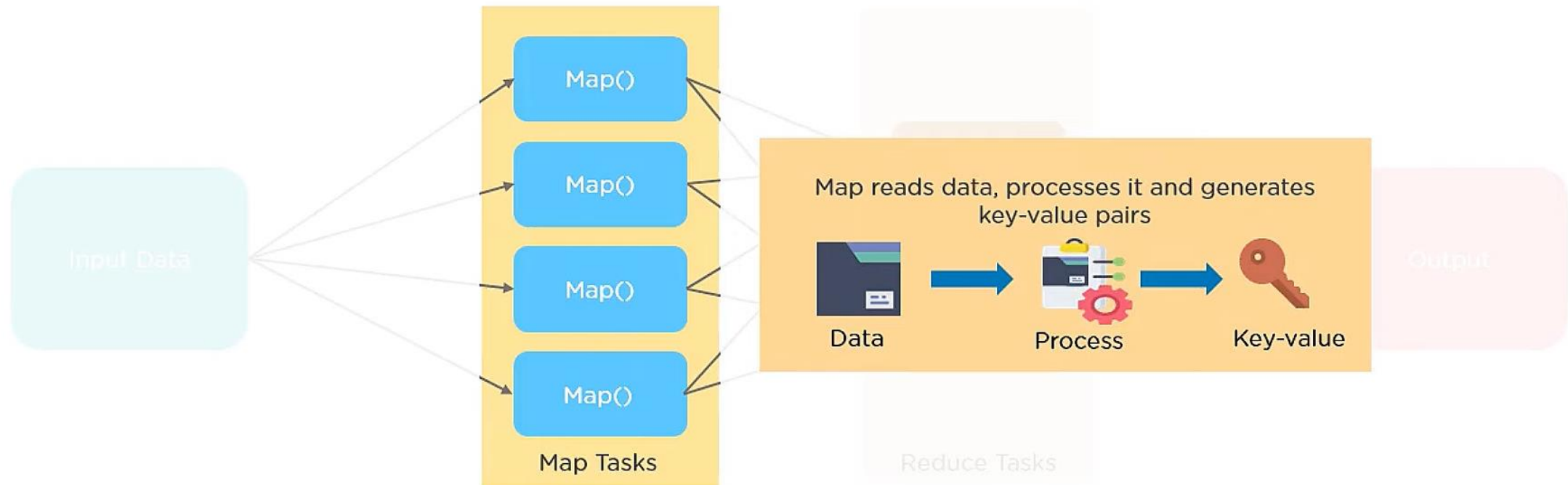
MapReduce is a framework that performs distributed and parallel processing of large volumes of data



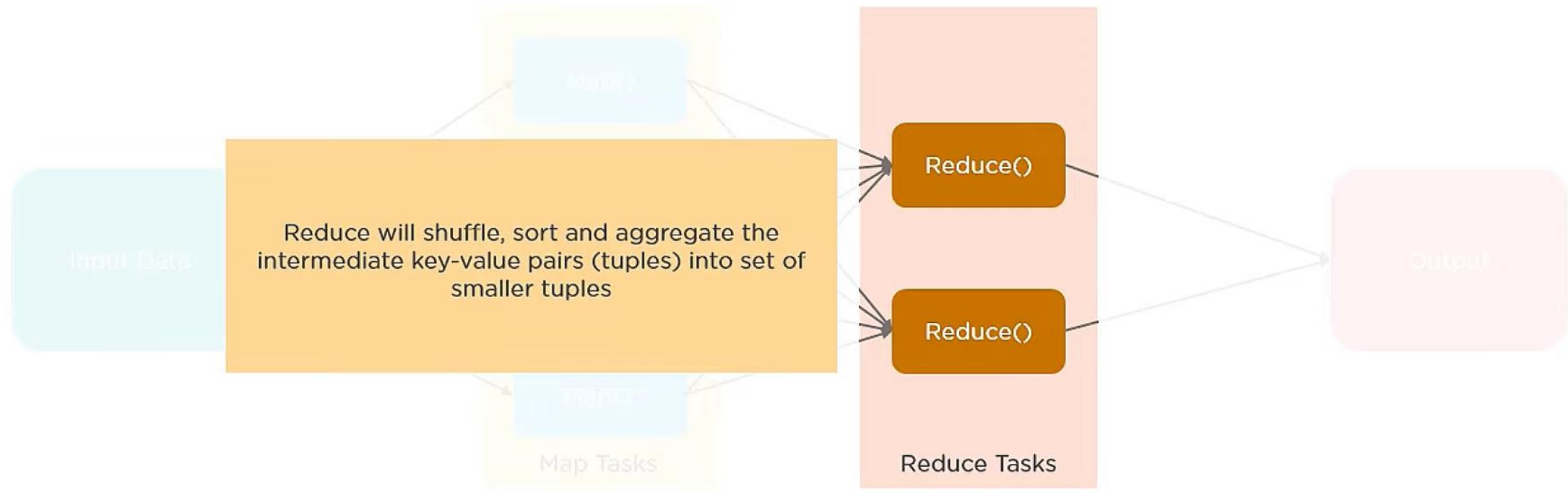
MapReduce – Input data



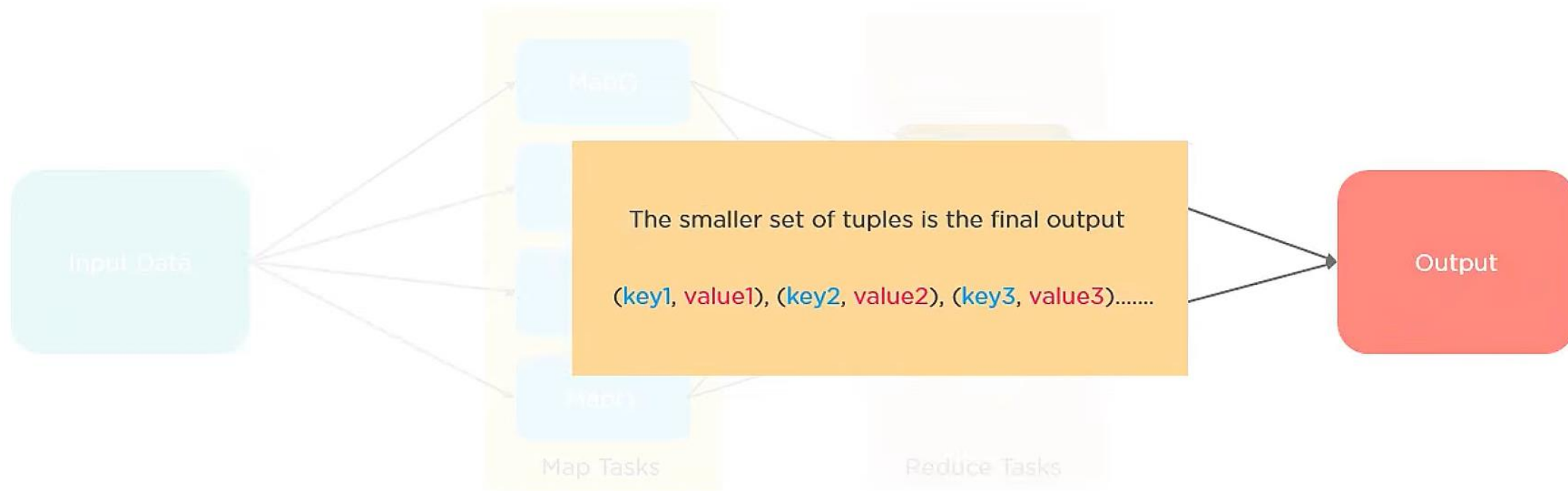
MapReduce – Mapper



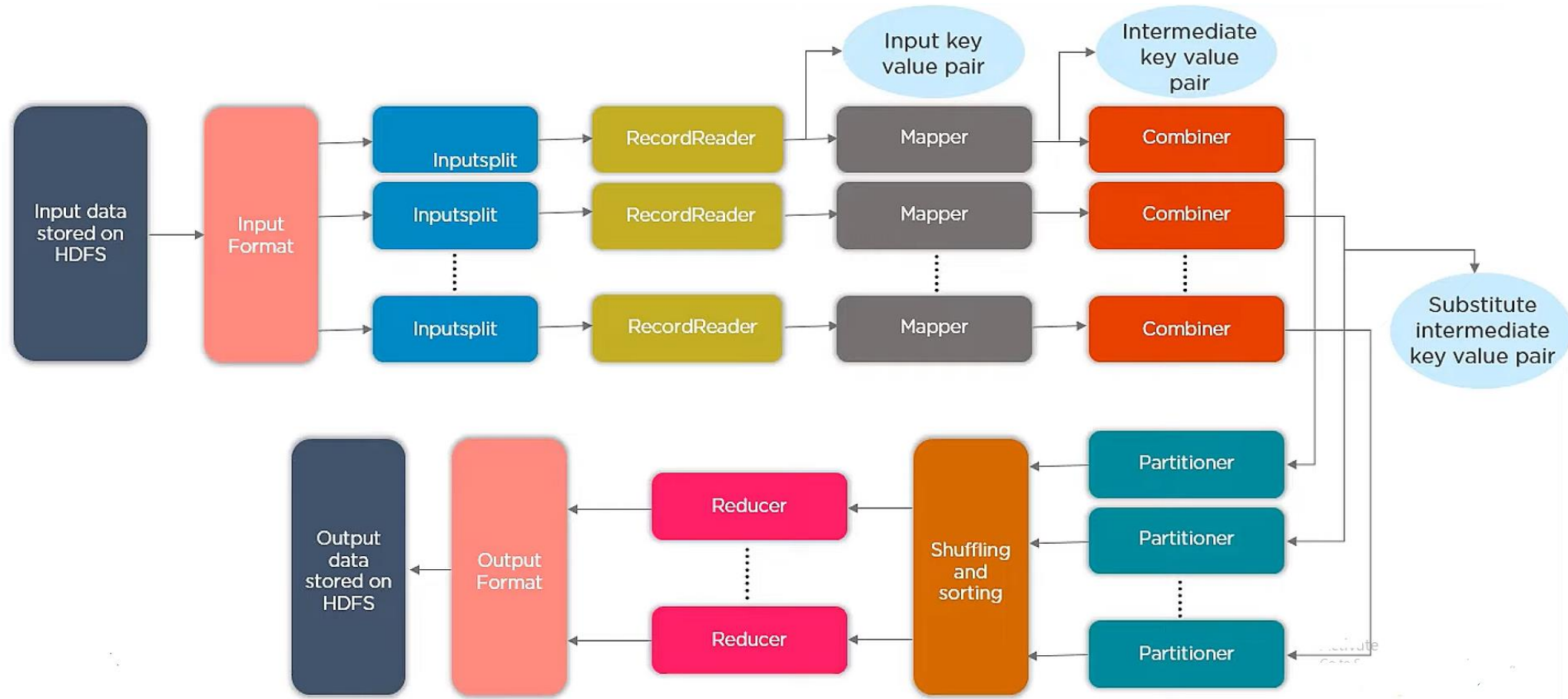
MapReduce – Reducer



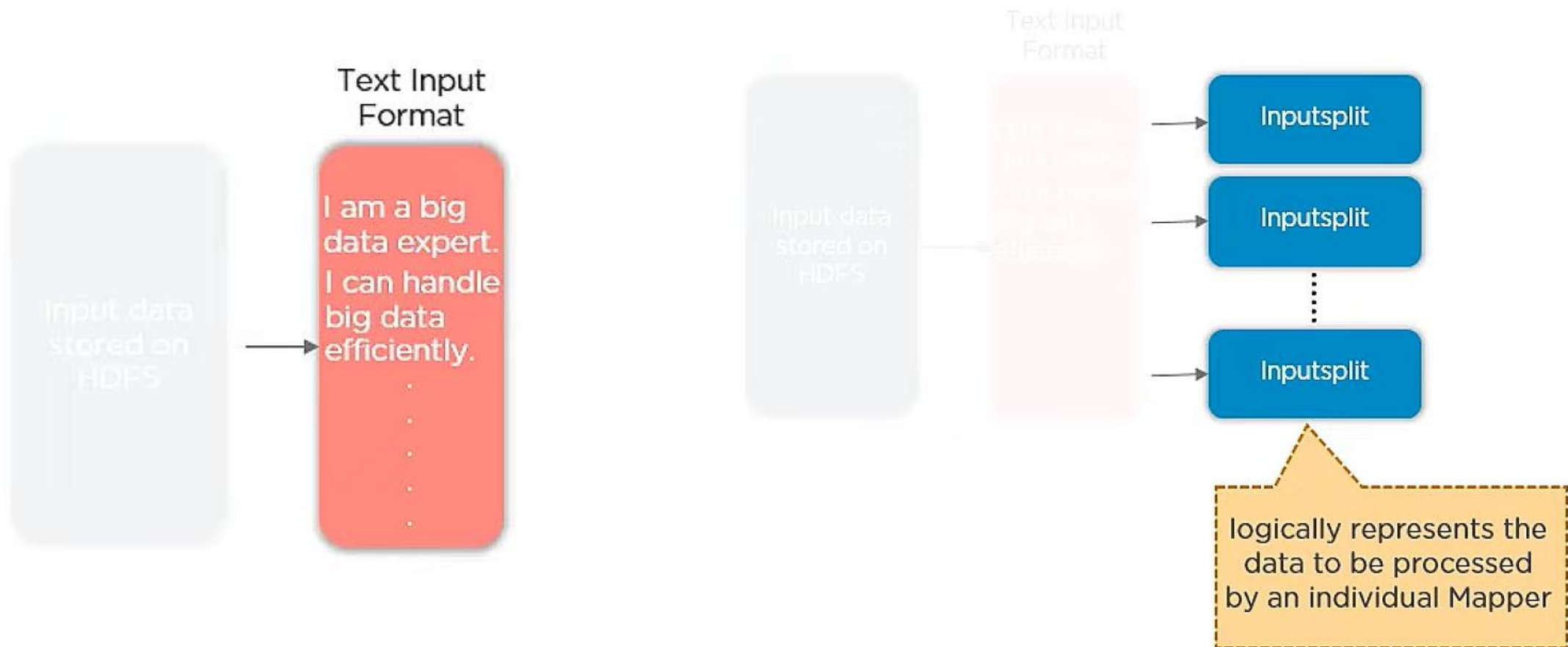
MapReduce – Result (Output)



MapReduce – Workflow – Overall



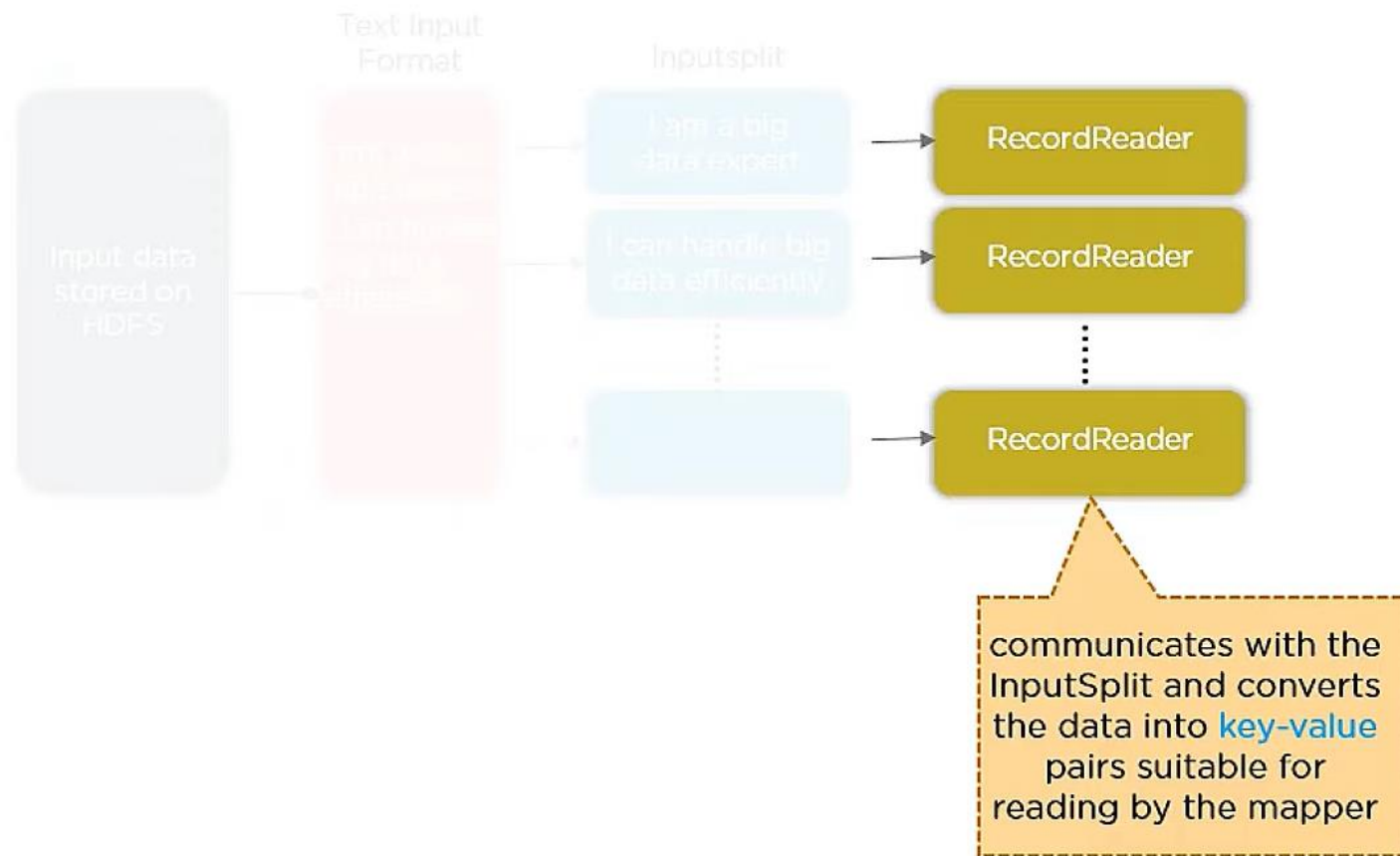
MapReduce – Workflow – Input data



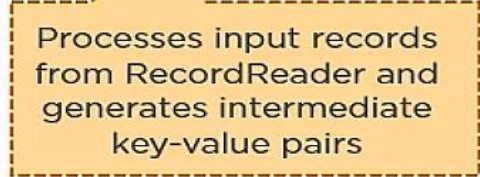
MapReduce – Workflow - Input data



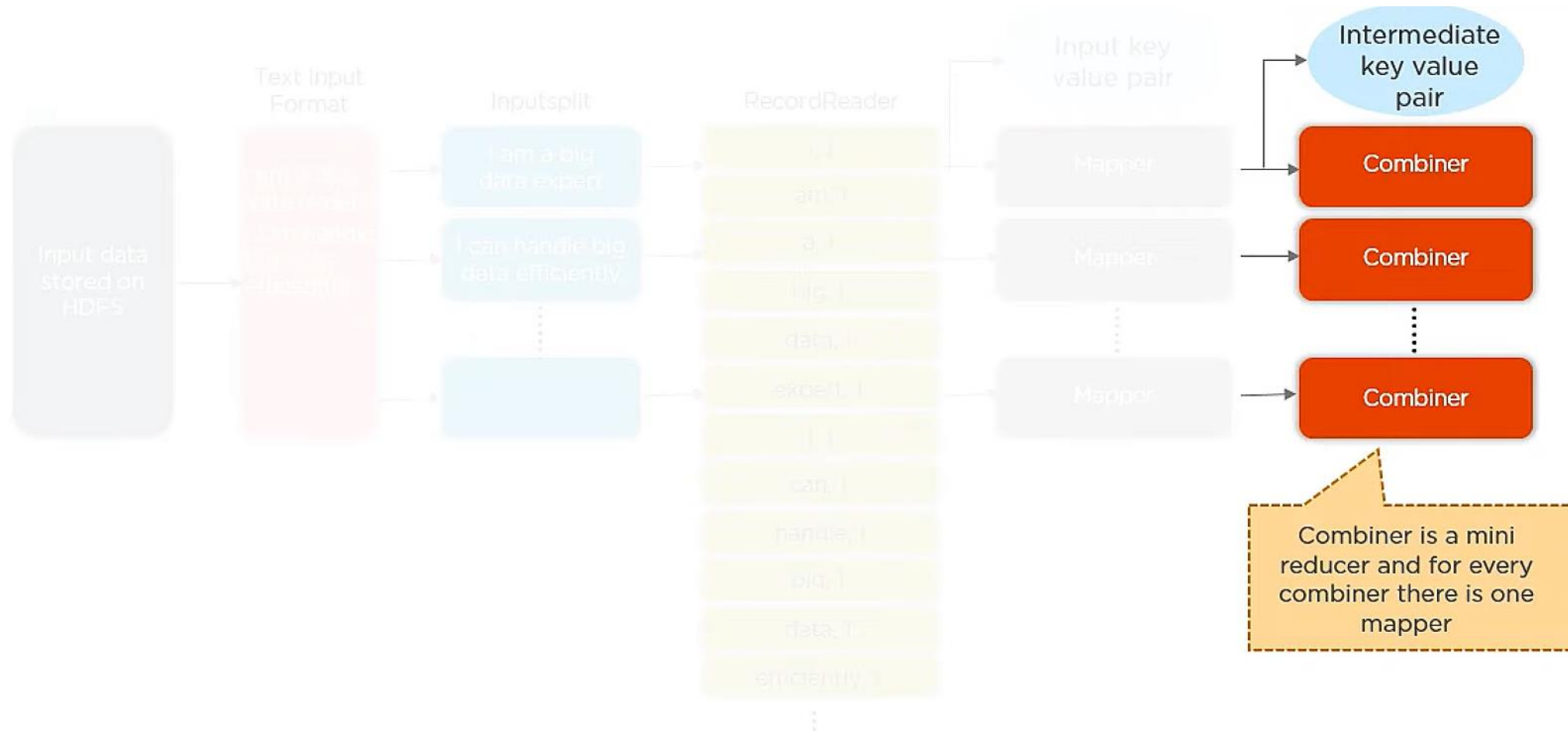
MapReduce – Workflow – Record Reader



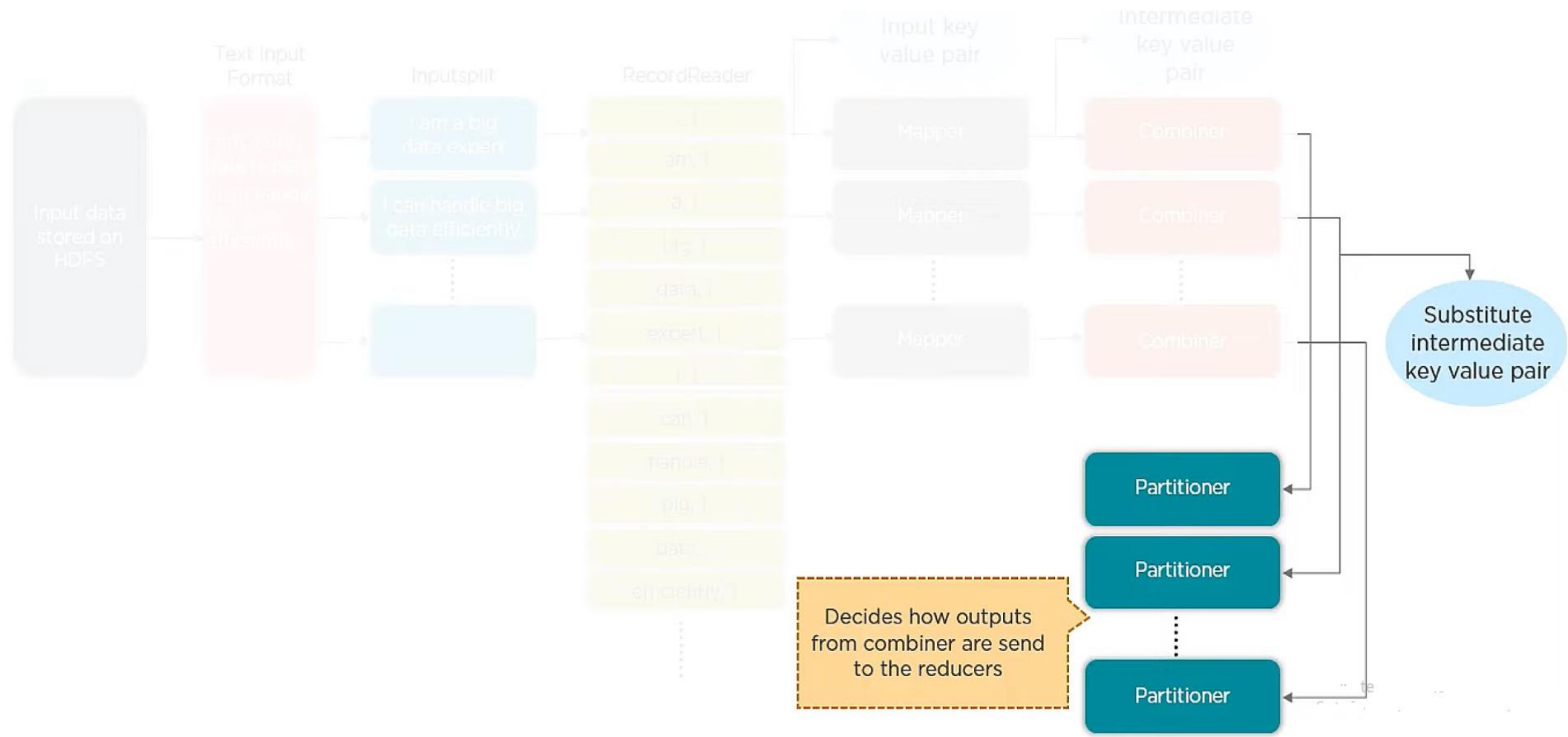
MapReduce – Workflow – Record Reader



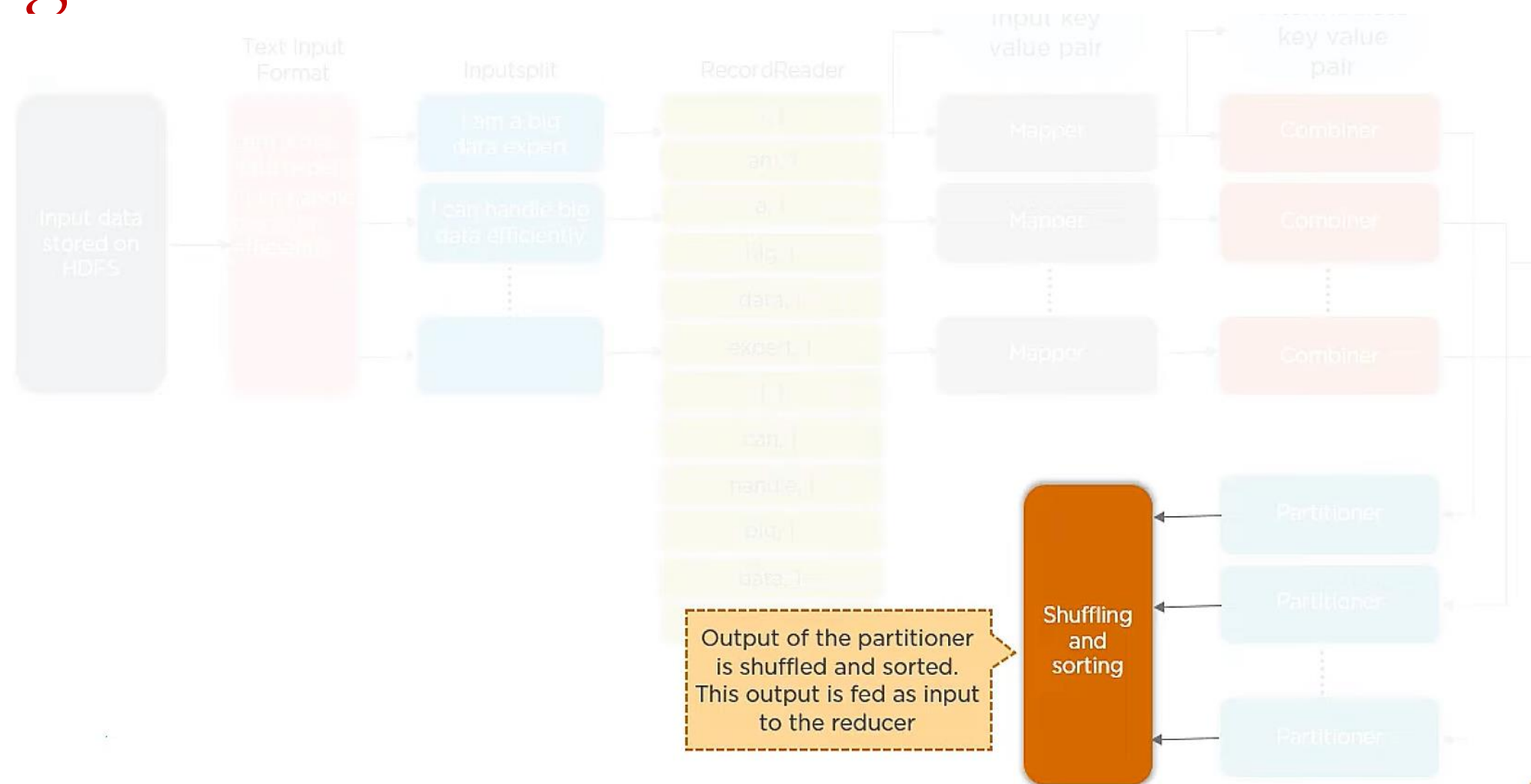
MapReduce – Workflow – Combiner



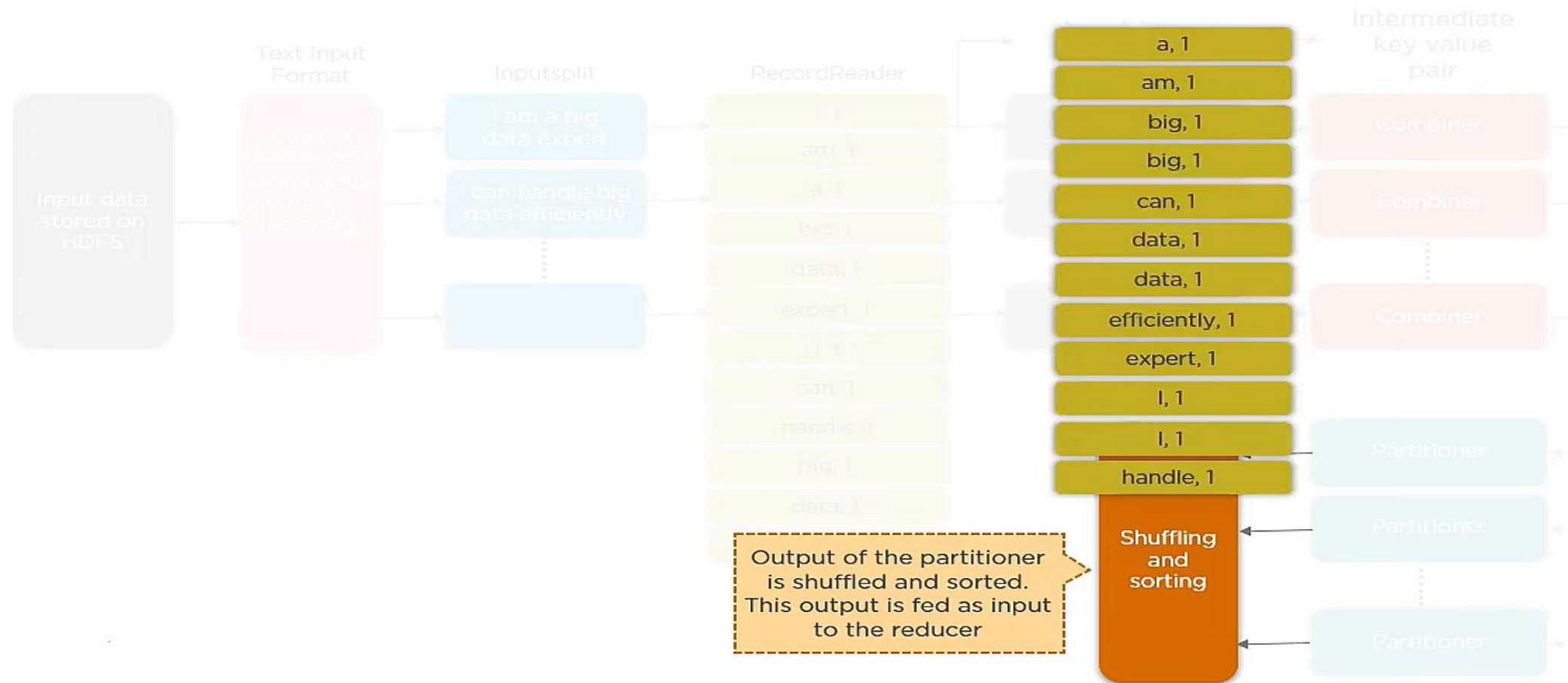
MapReduce – Workflow – Partitioner



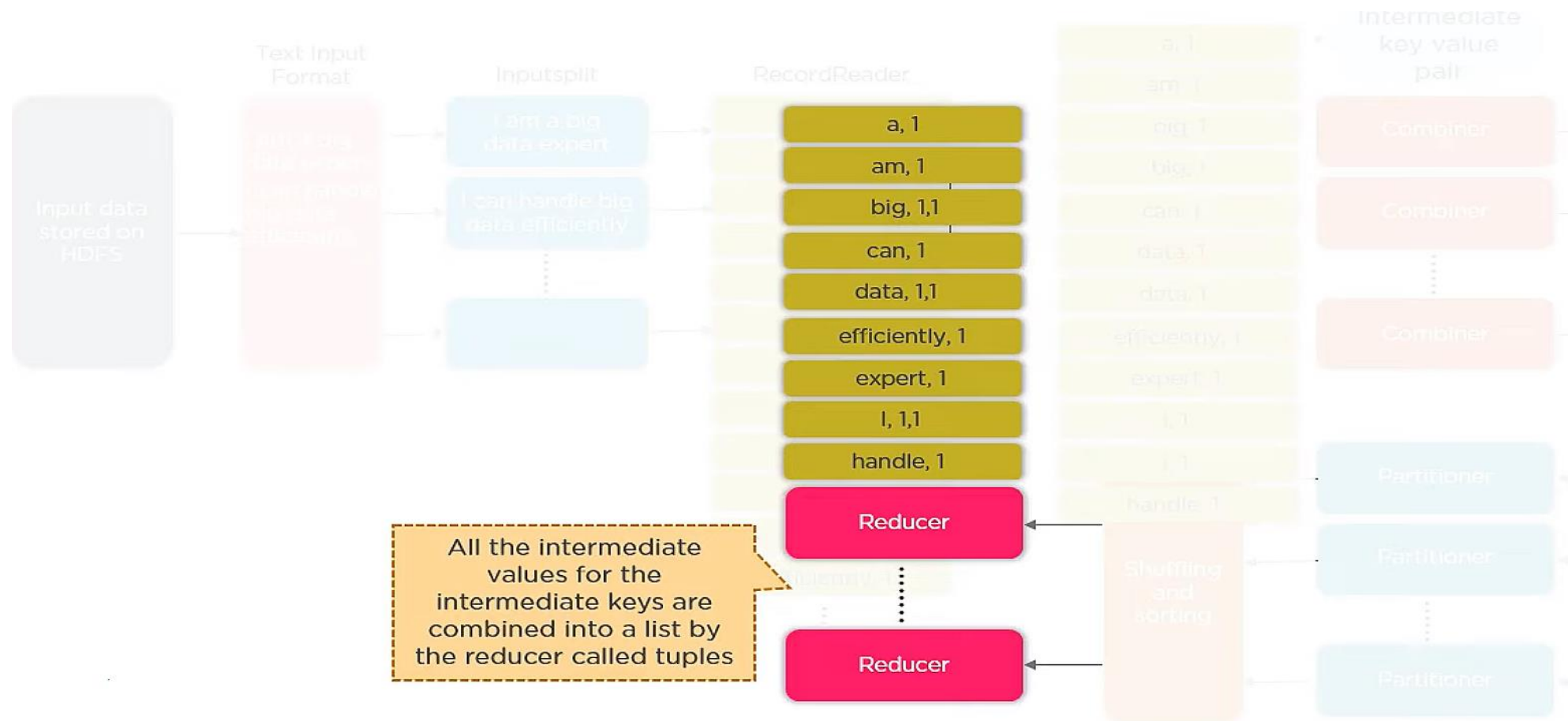
MapReduce – Workflow – Shuffling & Sorting



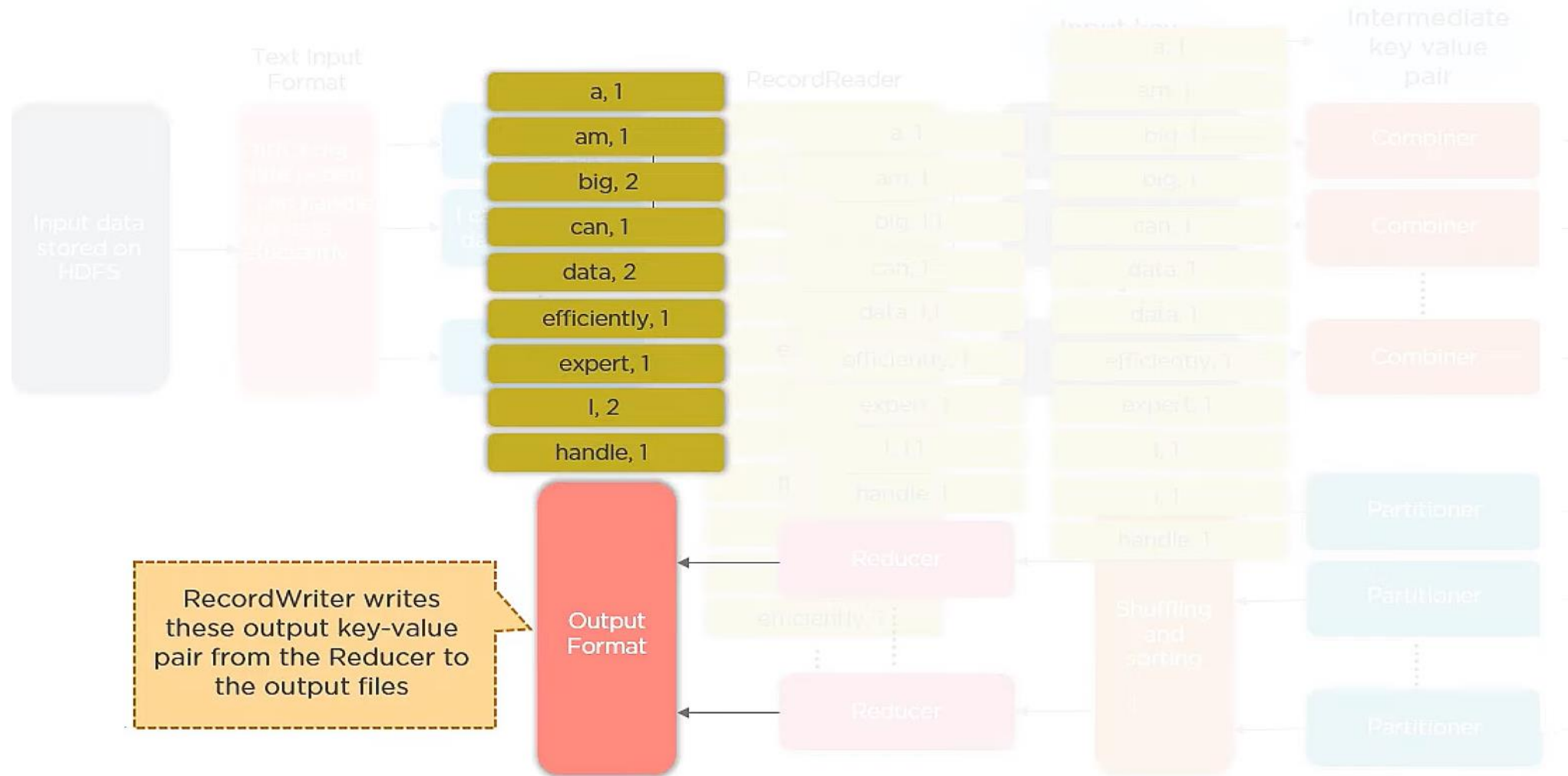
MapReduce – Workflow – Shuffling & Sorting



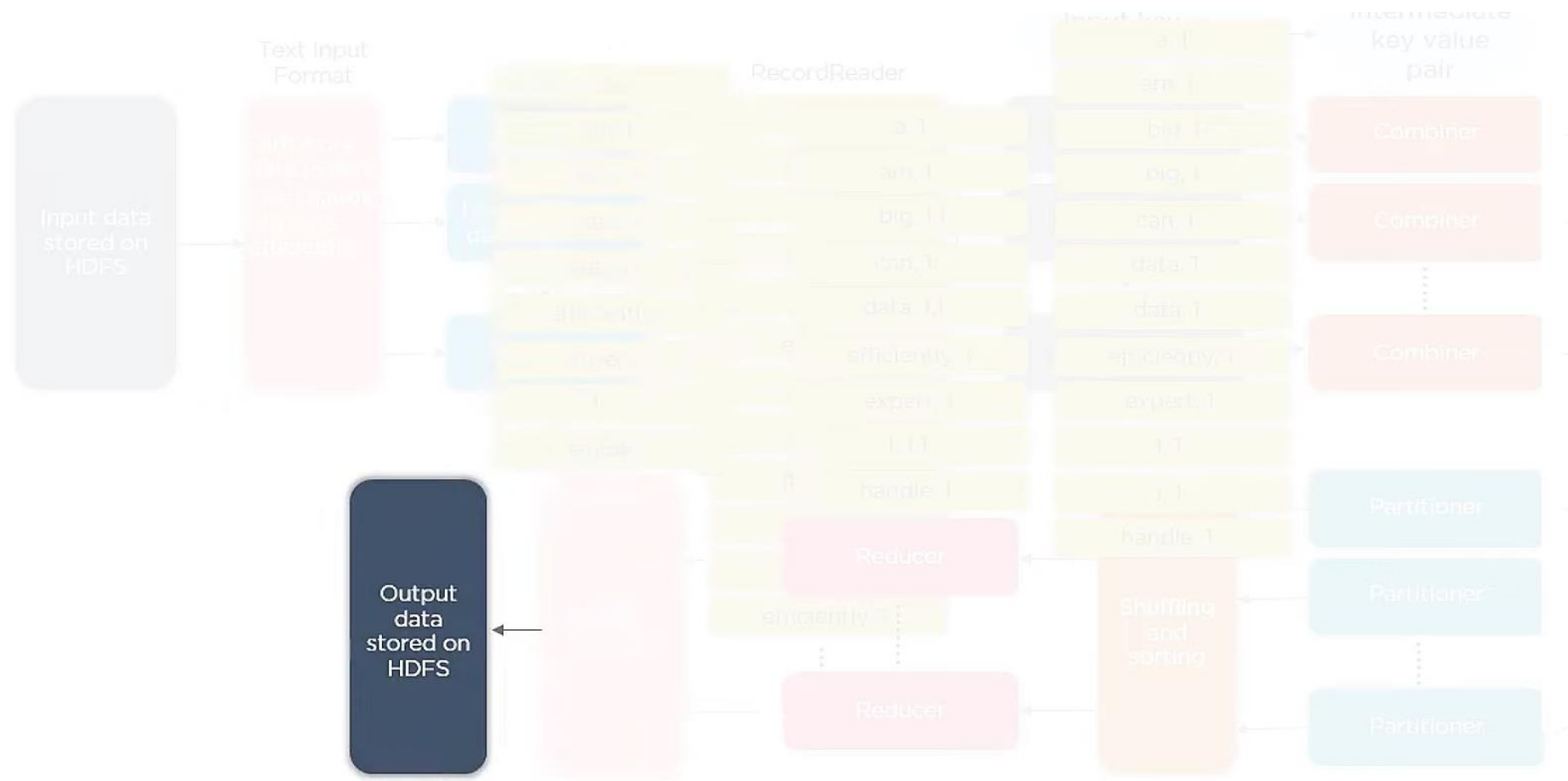
MapReduce – Workflow – Reducer



MapReduce – Workflow – Reducer

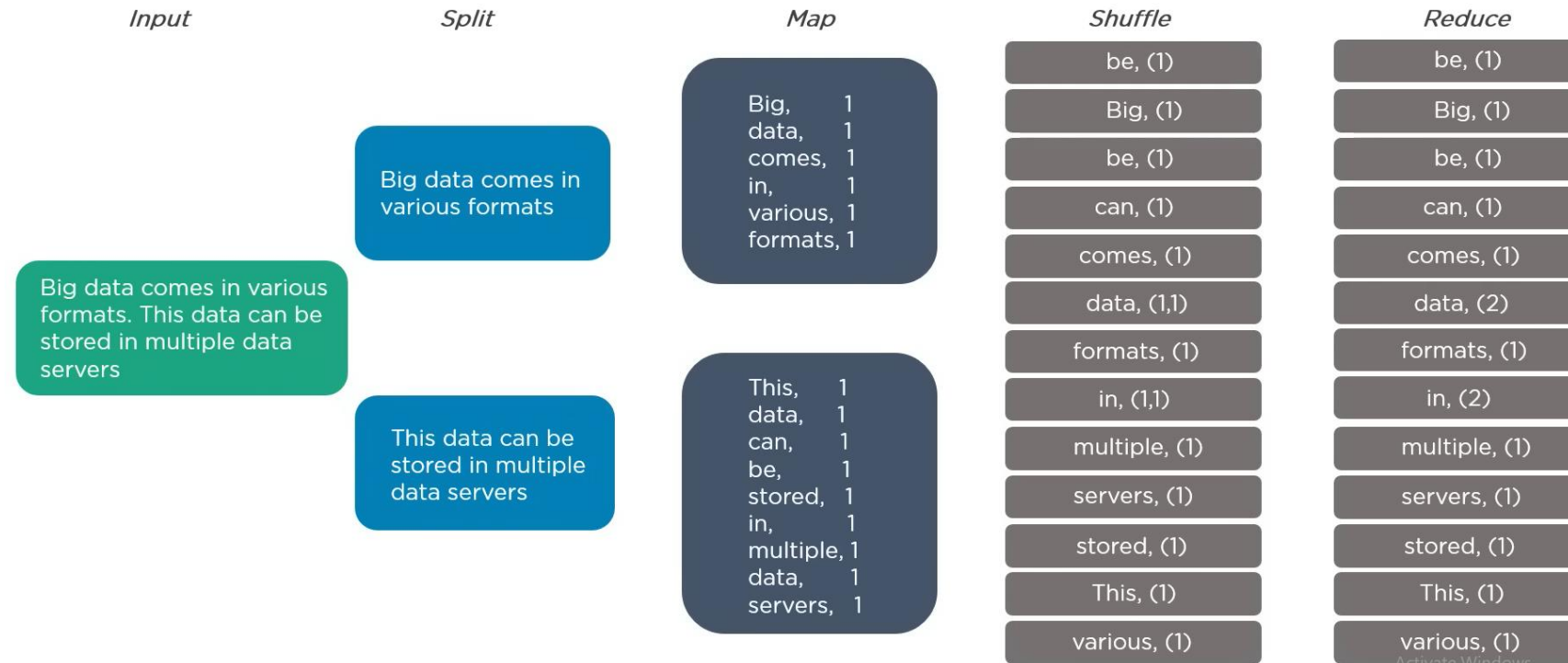


MapReduce – Workflow – Output

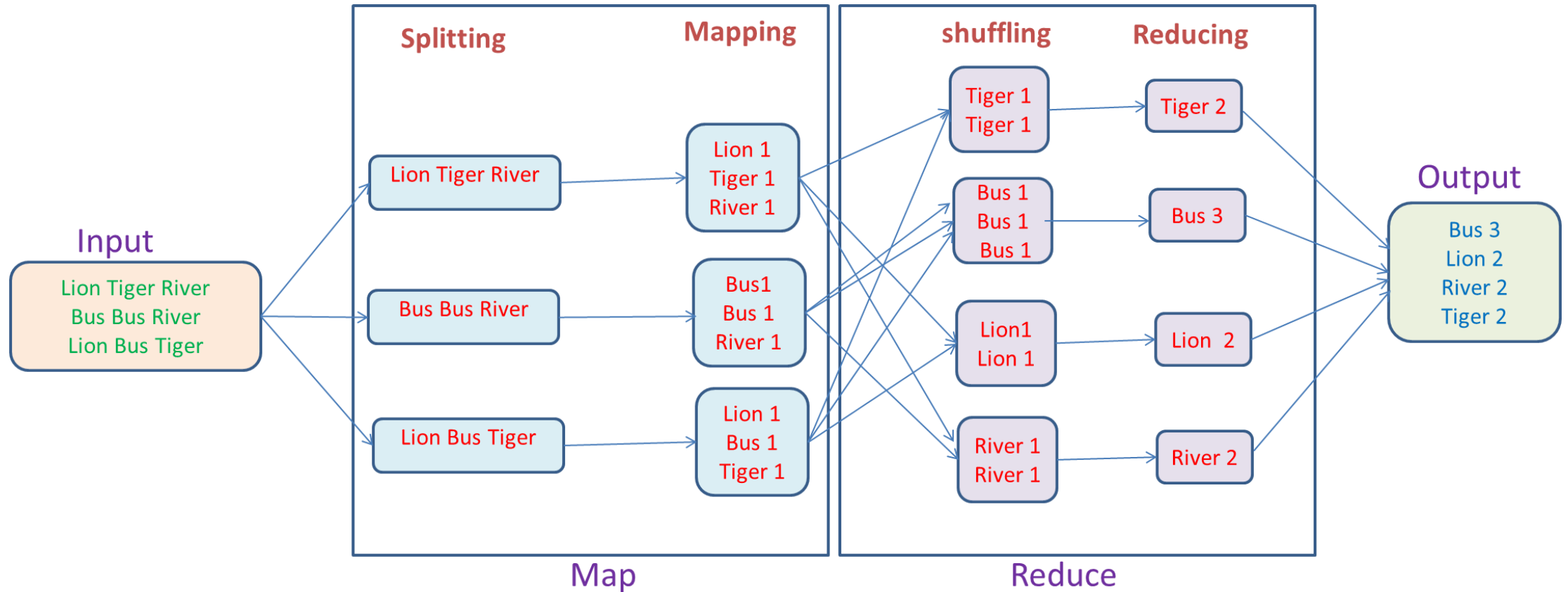


1-Mapper-Reducer

MapReduce Example

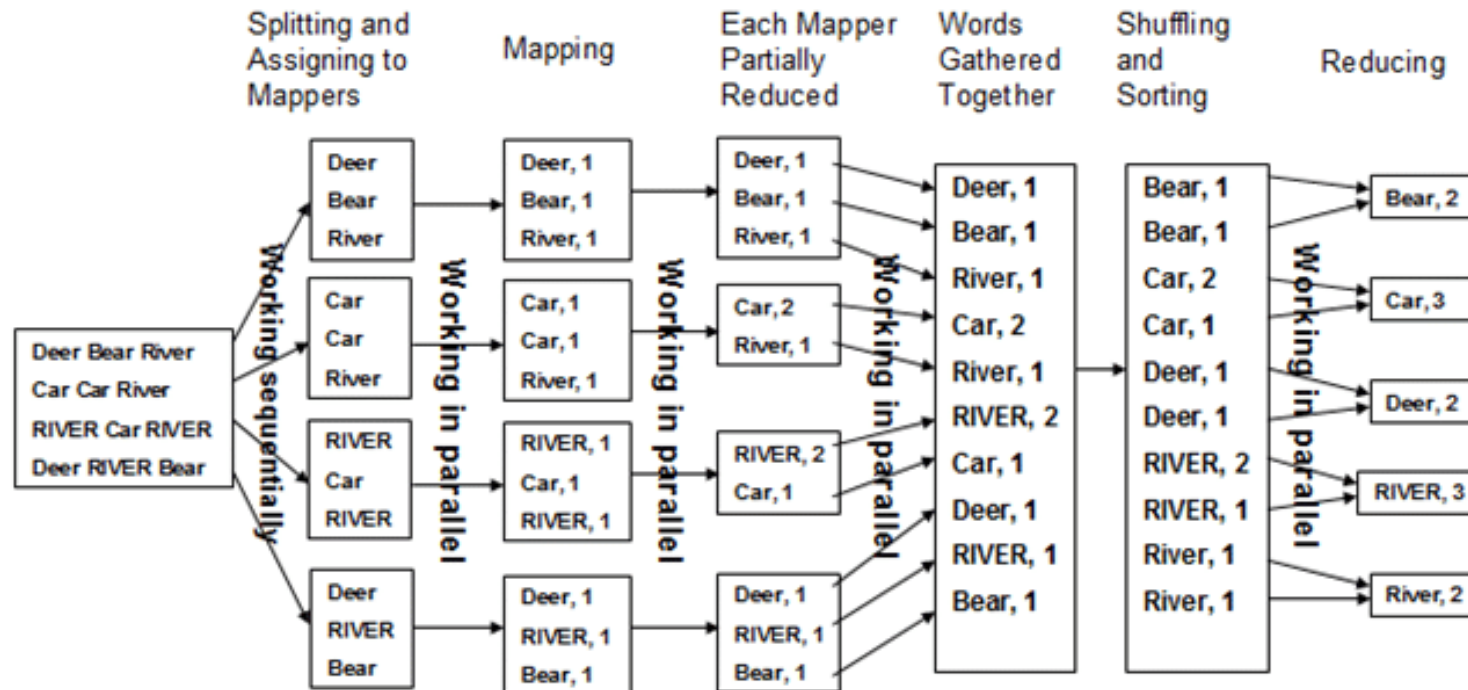


1-Mapper-Reducer



2-Mapper-Combiner-Reducer

Word Count with four Mappers/Splits



Thank You