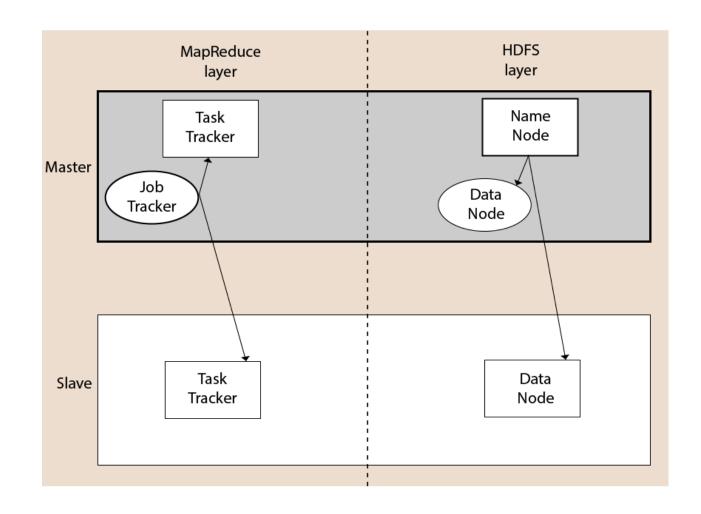


BML 621 – Cloud Computing for Big Data Hadoop Framework



Recap

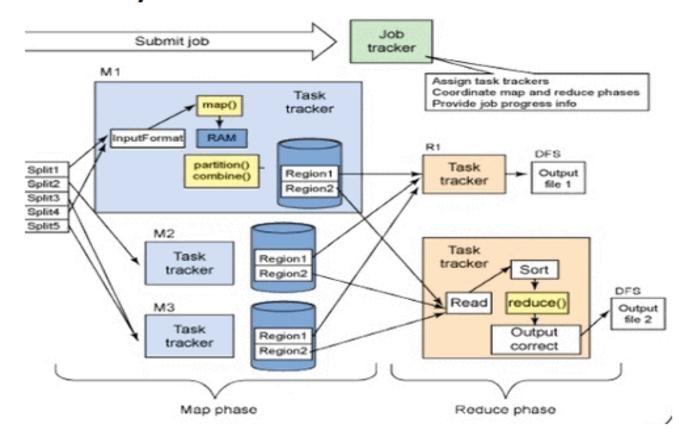




Recap



Hadoop Architeture



Hadoop



Hadoop

1. HDFS

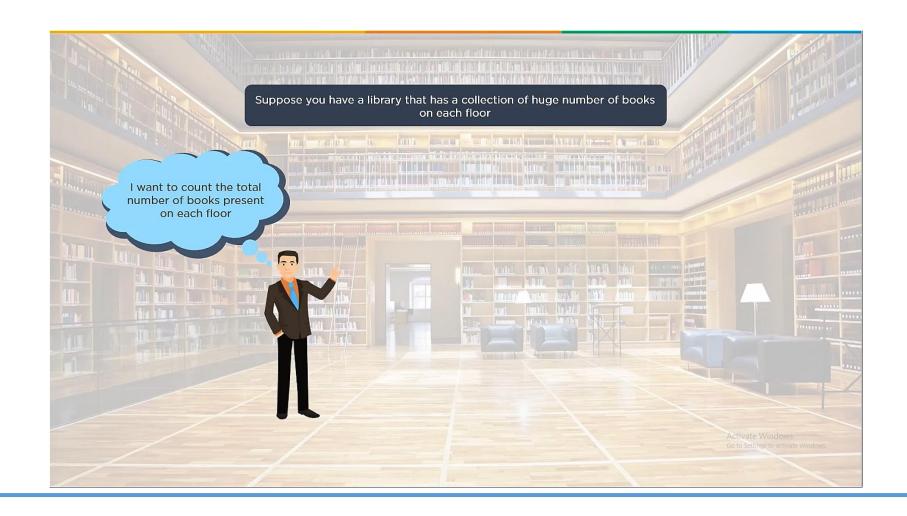
[HDFS architecture, Blocks, Replication, Rack awareness, Hadoop Architecture]

2. MR?

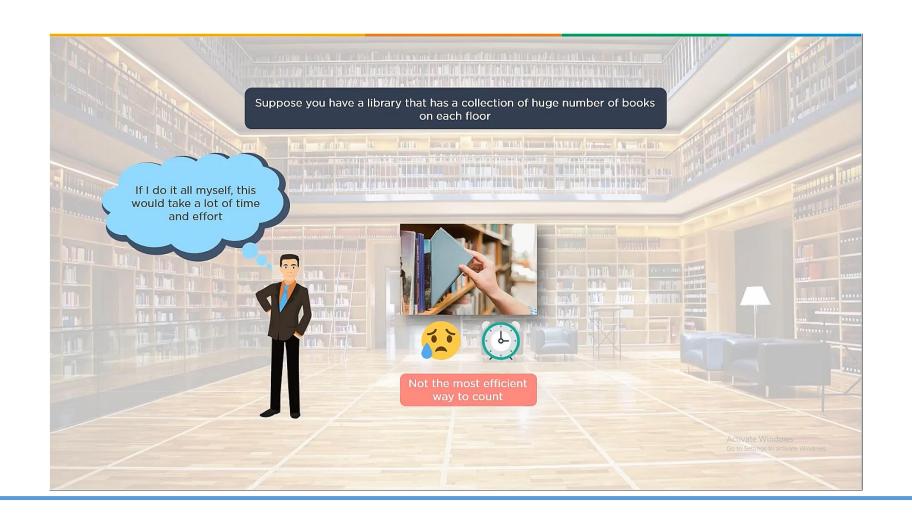
Apache MapReduce





















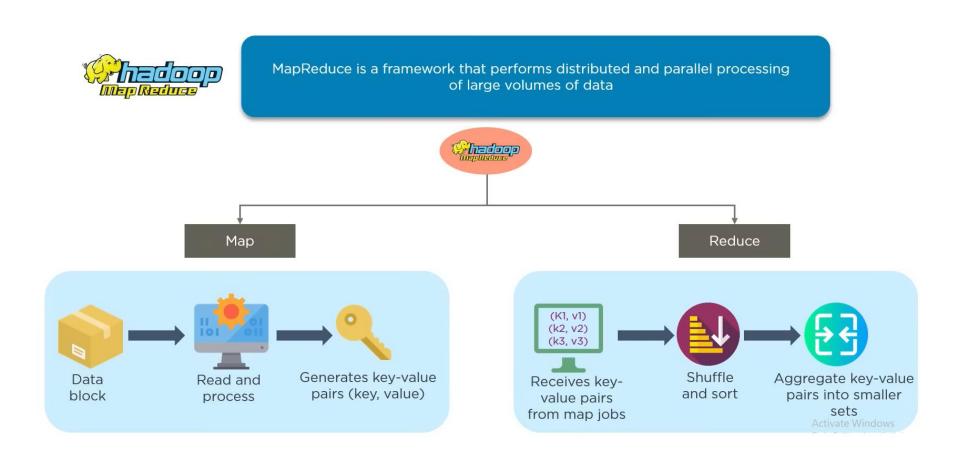






Hadoop - MapReduce

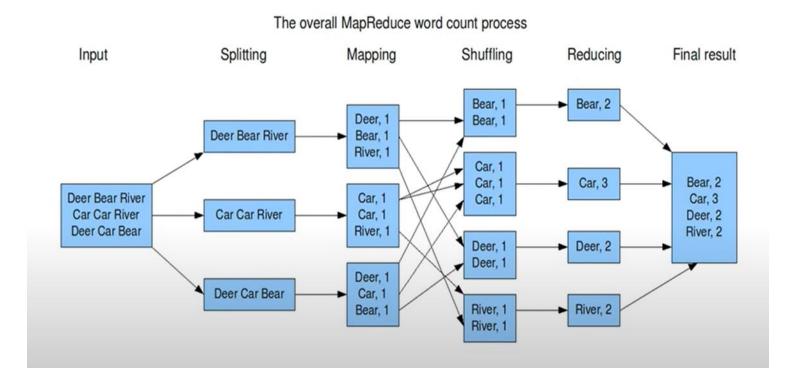




MapReduce - Example



Word Count Example

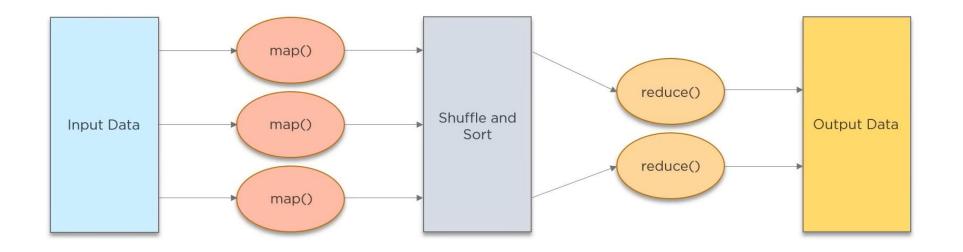


Hadoop - MapReduce





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

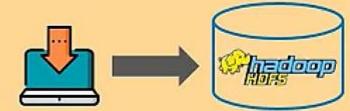


MapReduce – Input data



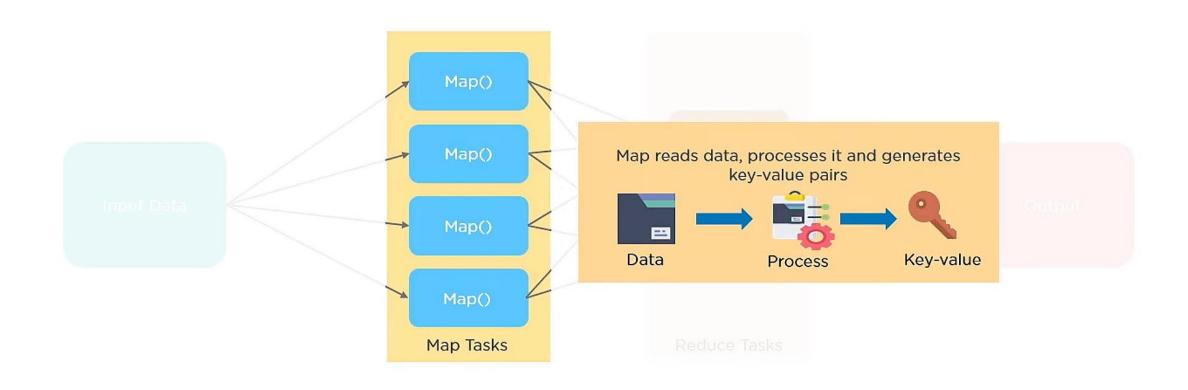
Input Data

Accepts input data in various formats. The input data files are stored in HDFS



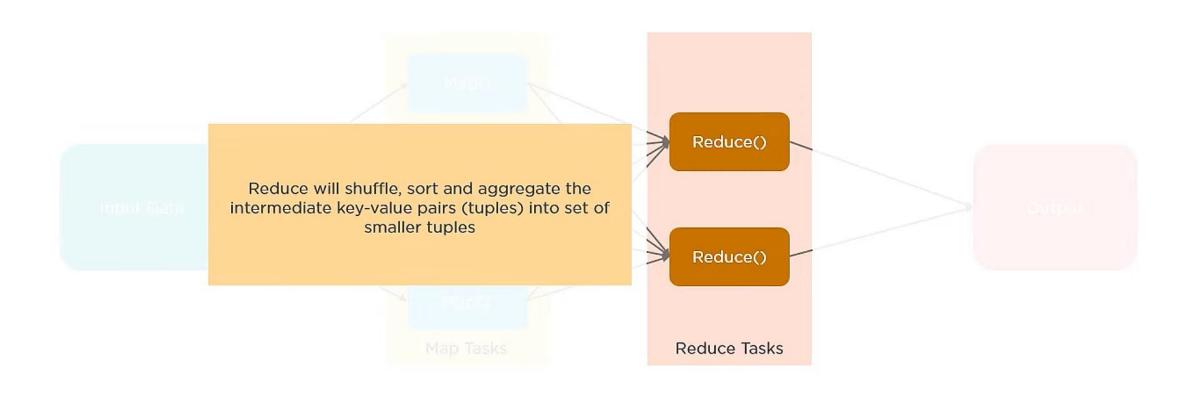
MapReduce – Mapper





MapReduce – Reducer





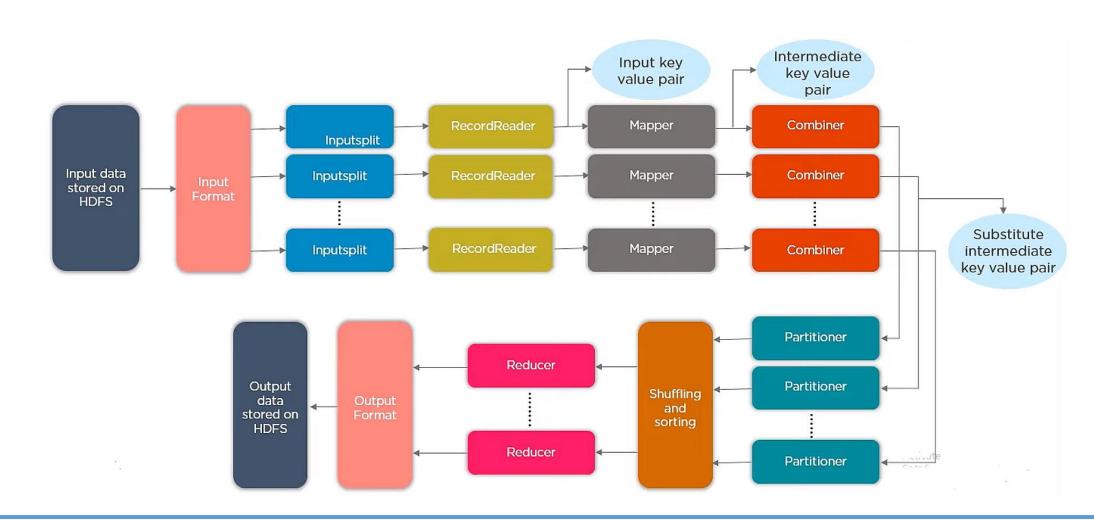
MapReduce – Result (Output)





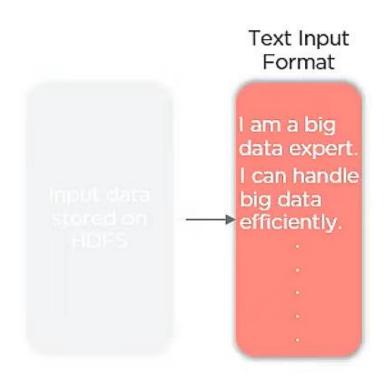
MapReduce – Workflow – Overall

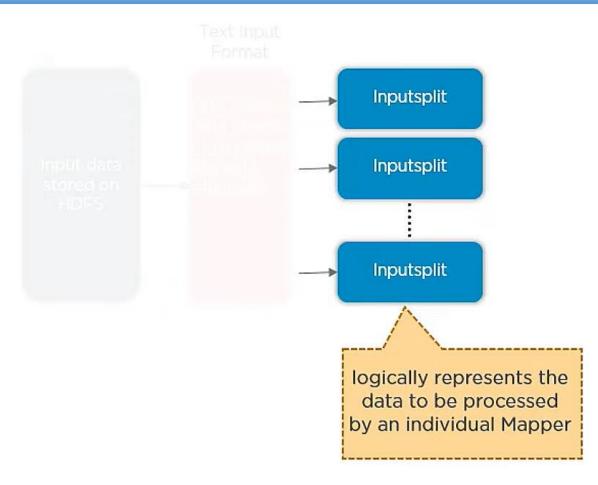




MapReduce – Workflow – Input data





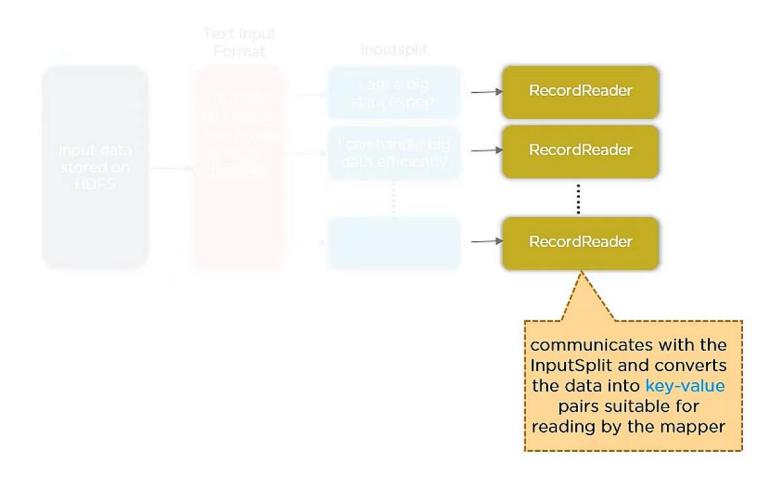


MapReduce – Workflow - Input data





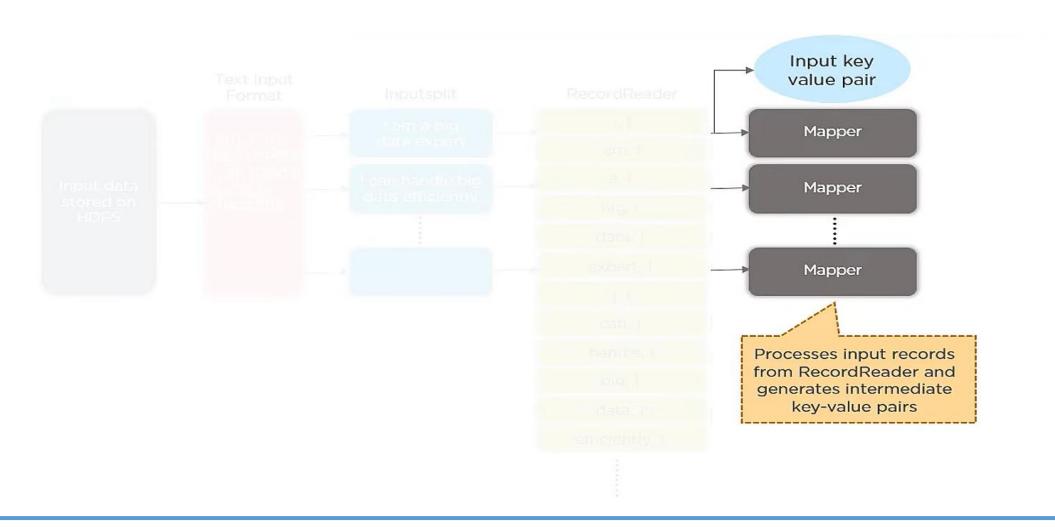
MapReduce – Workflow – Record Reader Information Technology Kottavam



MapReduce – Workflow – Record Reader Information Technology Kottayam

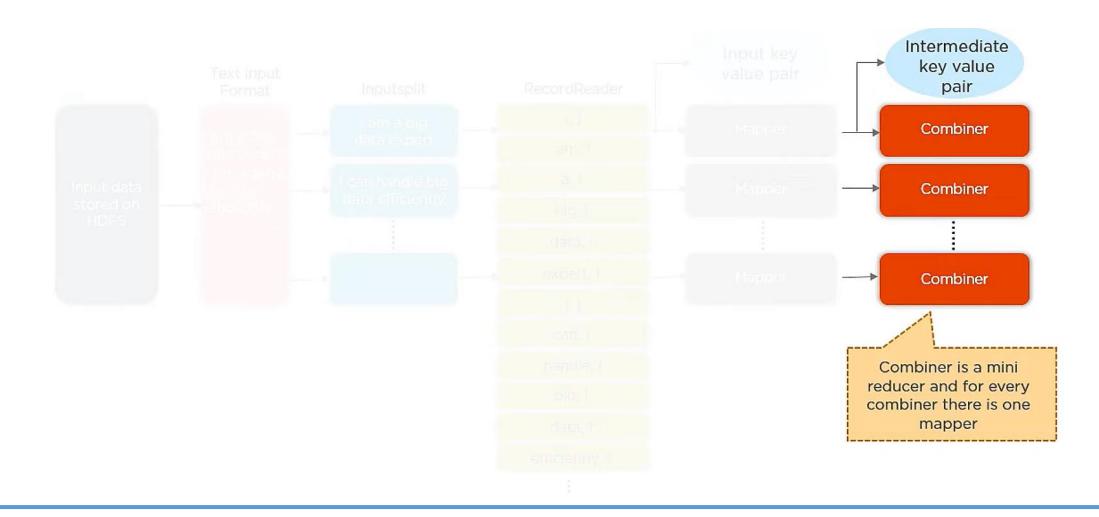
MapReduce – Workflow – Mapper





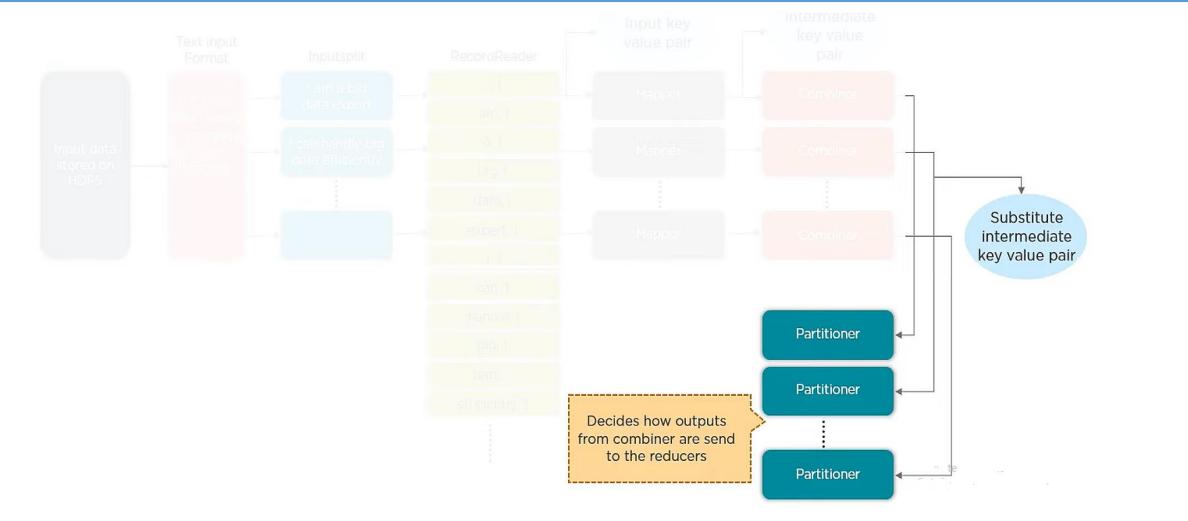
MapReduce – Workflow – Combiner





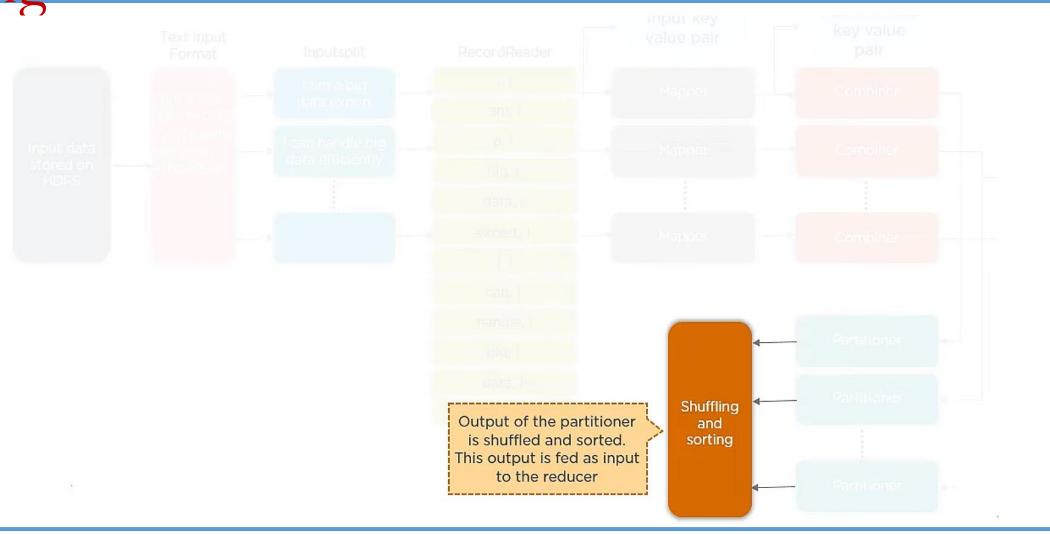
MapReduce – Workflow – Partitioner





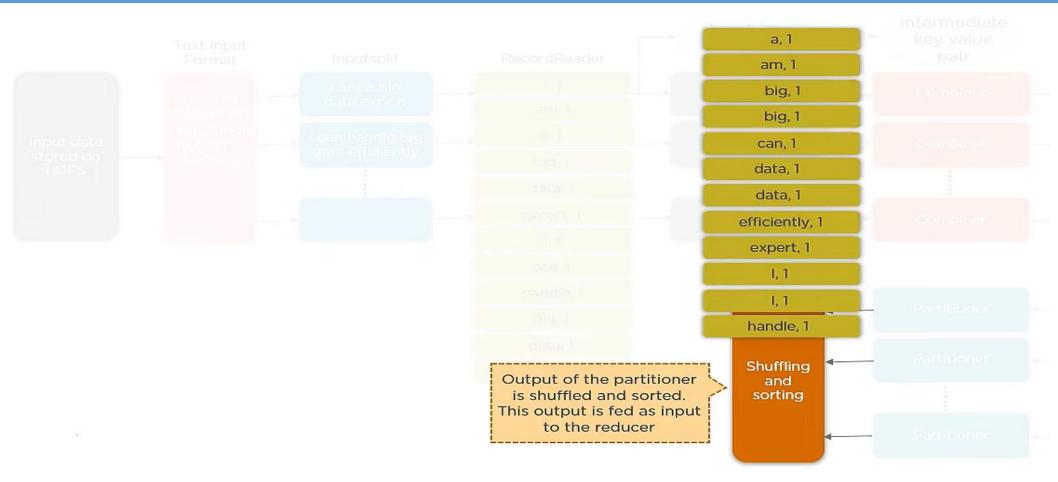
MapReduce – Workflow – Shuffling & Sorting





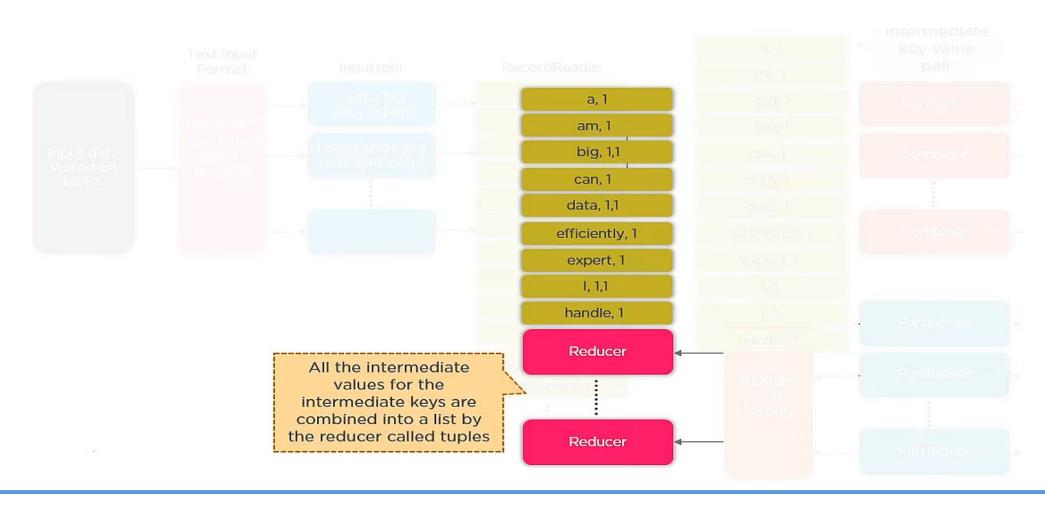
MapReduce – Workflow – Shuffling & Sorting





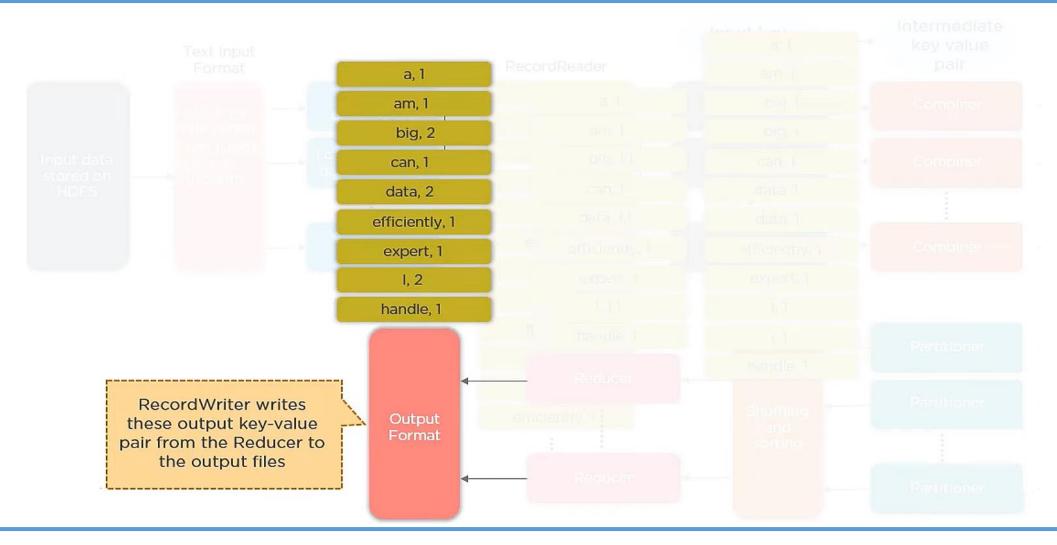
MapReduce – Workflow – Reducer





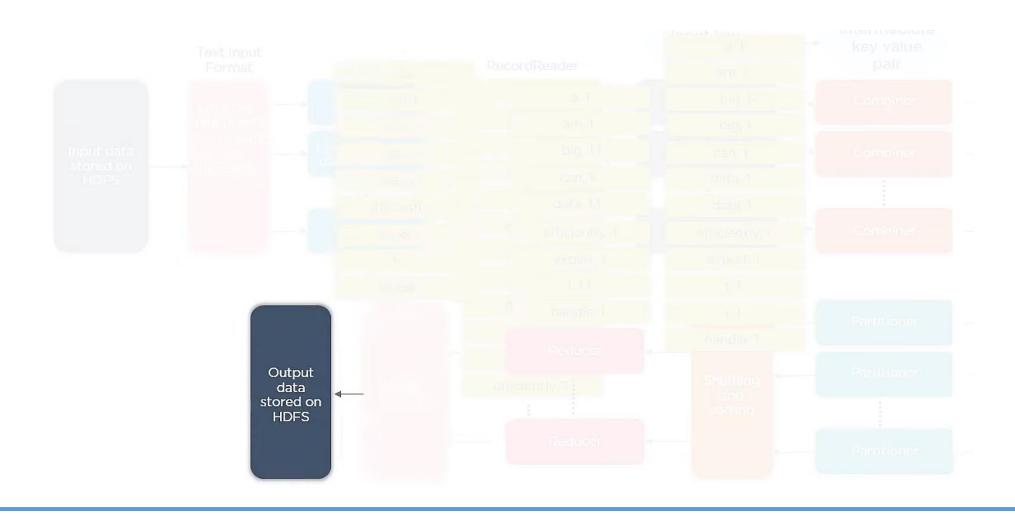
MapReduce – Workflow – Reducer





MapReduce – Workflow – Output





1-Mapper-Reducer

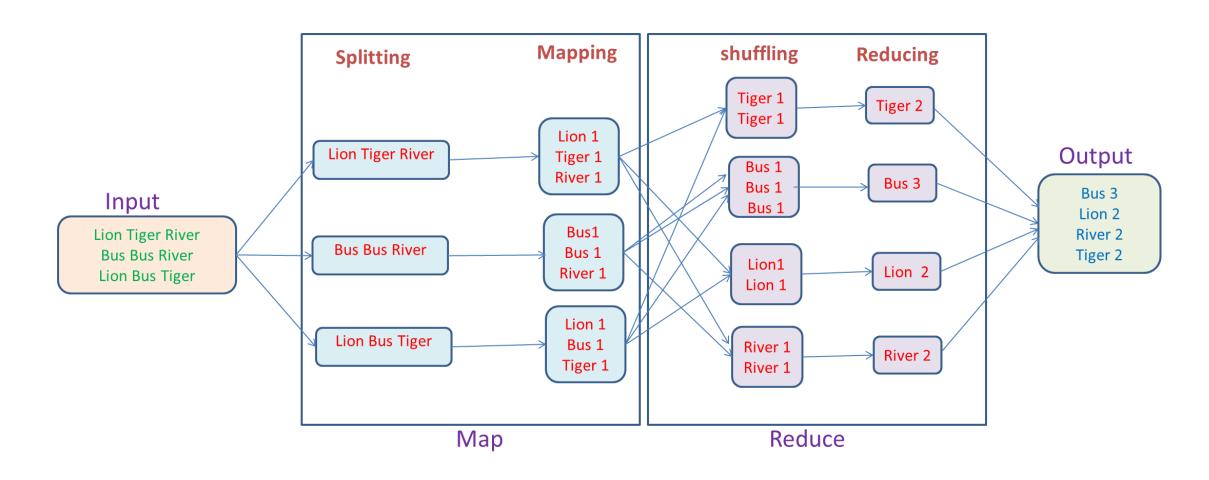


MapReduce Example

Input Split Мар Shuffle Reduce be, (1) be, (1) Big, Big, (1) Big, (1) data, be, (1) be, (1) comes, 1 Big data comes in various formats can, (1) can, (1) various, 1 formats, 1 comes, (1) comes, (1) Big data comes in various data, (1,1) data, (2) formats. This data can be stored in multiple data formats, (1) formats, (1) servers This, in, (1,1) in, (2) data, This data can be can, multiple, (1) multiple, (1) stored in multiple be, data servers servers, (1) stored, 1 servers, (1) stored, (1) stored, (1) multiple, 1 data, This, (1) This, (1) servers, 1 various, (1) various, (1)

1-Mapper-Reducer

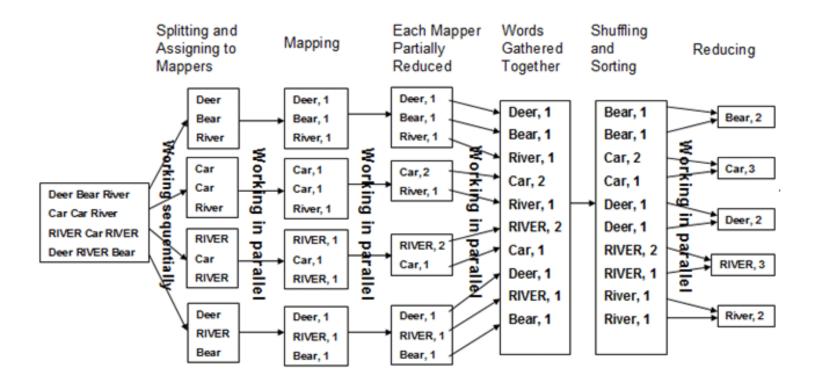




2-Mapper-Combiner-Reducer



Word Count with four Mappers/Splits





Thank You