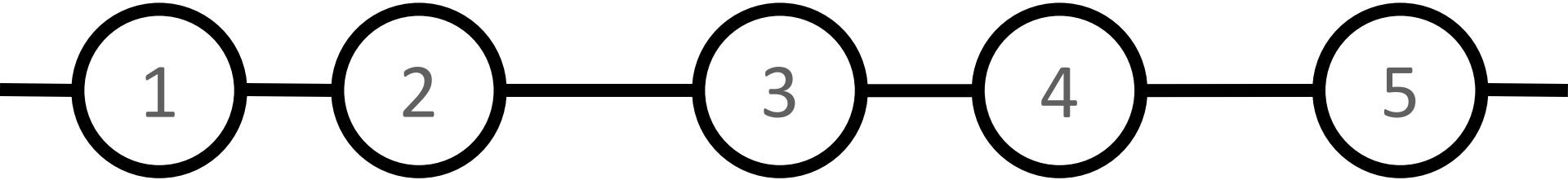
Part 4

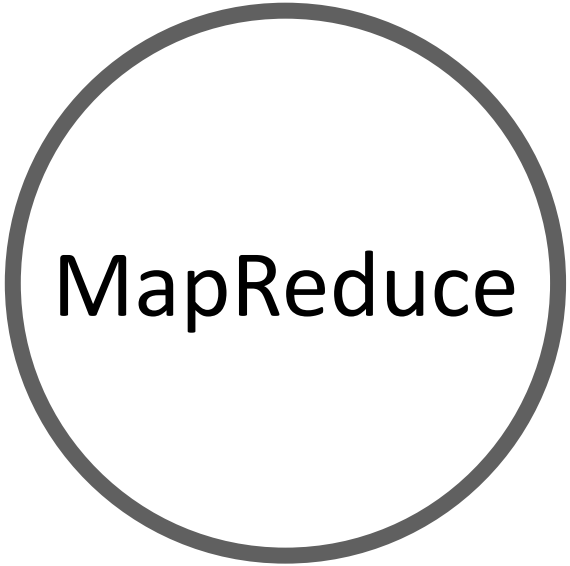
Ahmed Ramadan, ary00@fayoum.edu.eg

AGENDA



1. Map Reduce Overll
2. Combiner
3. With Vs Without Combiner
4. Inverted index Problem
5. End
   Questions

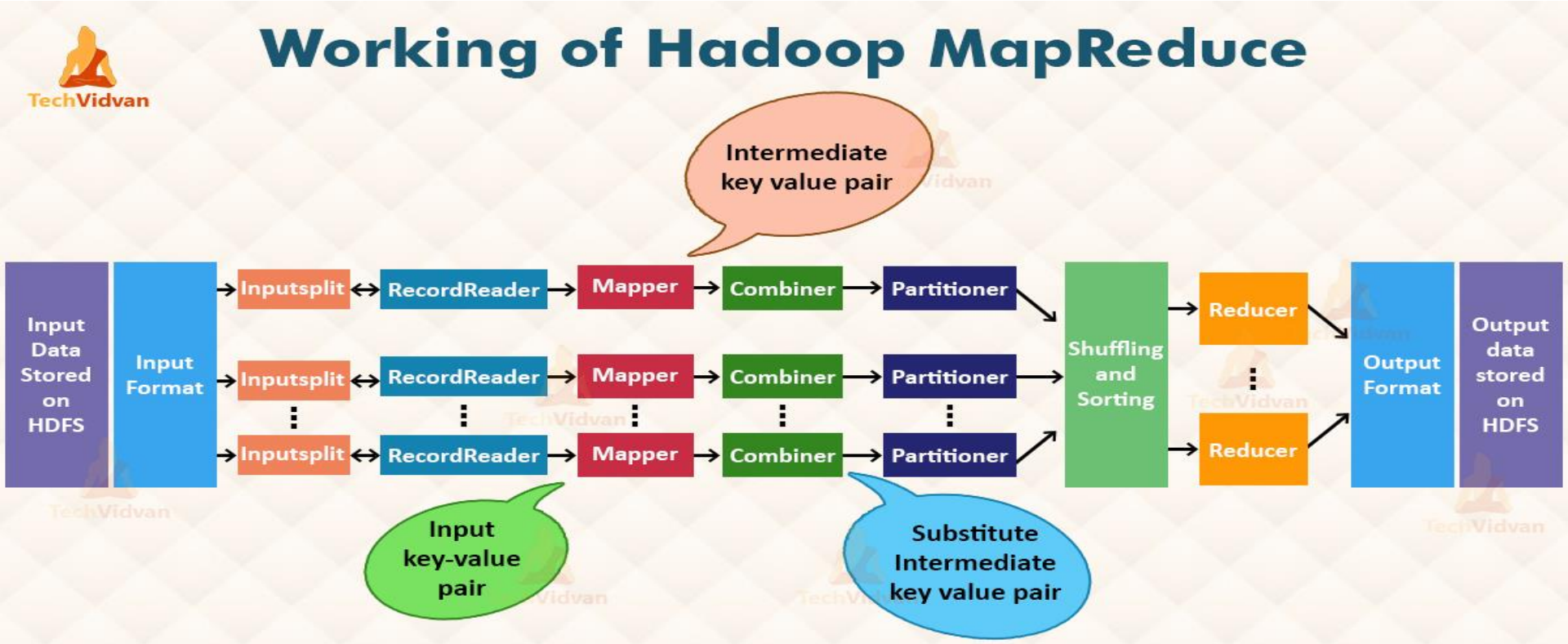
Ahmed Ramadan, ary00@fayoum.edu.eg



Map Reduce Stages

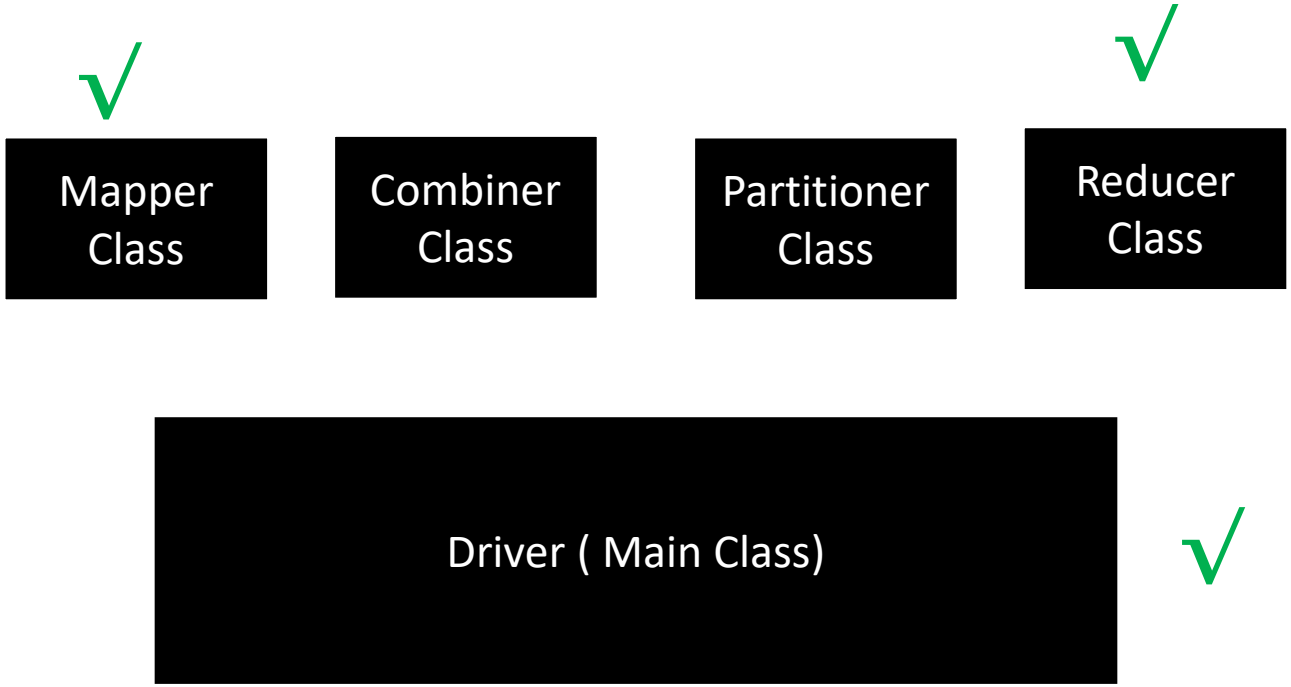
Ahmed Ramadan, ary00@fayoum.edu.eg

Overall stages

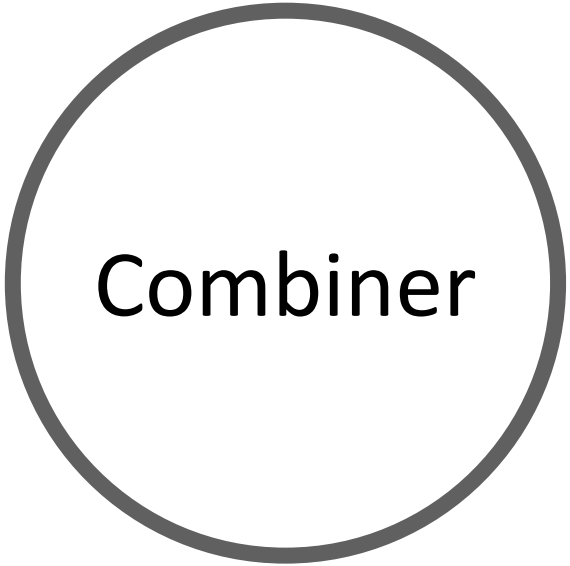


Ahmed Ramadan, ary00@fayoum.edu.eg

Hadoop Client Job



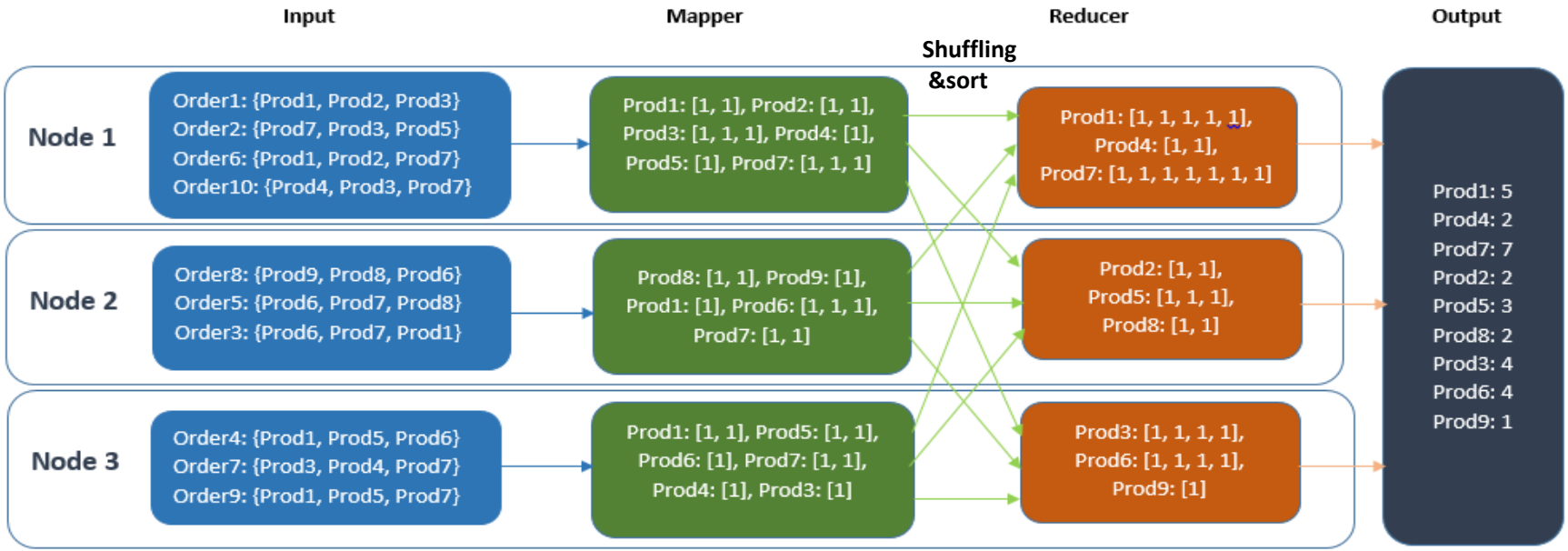
Ahmed Ramadan, ary00@fayoum.edu.eg



Combiner Stage

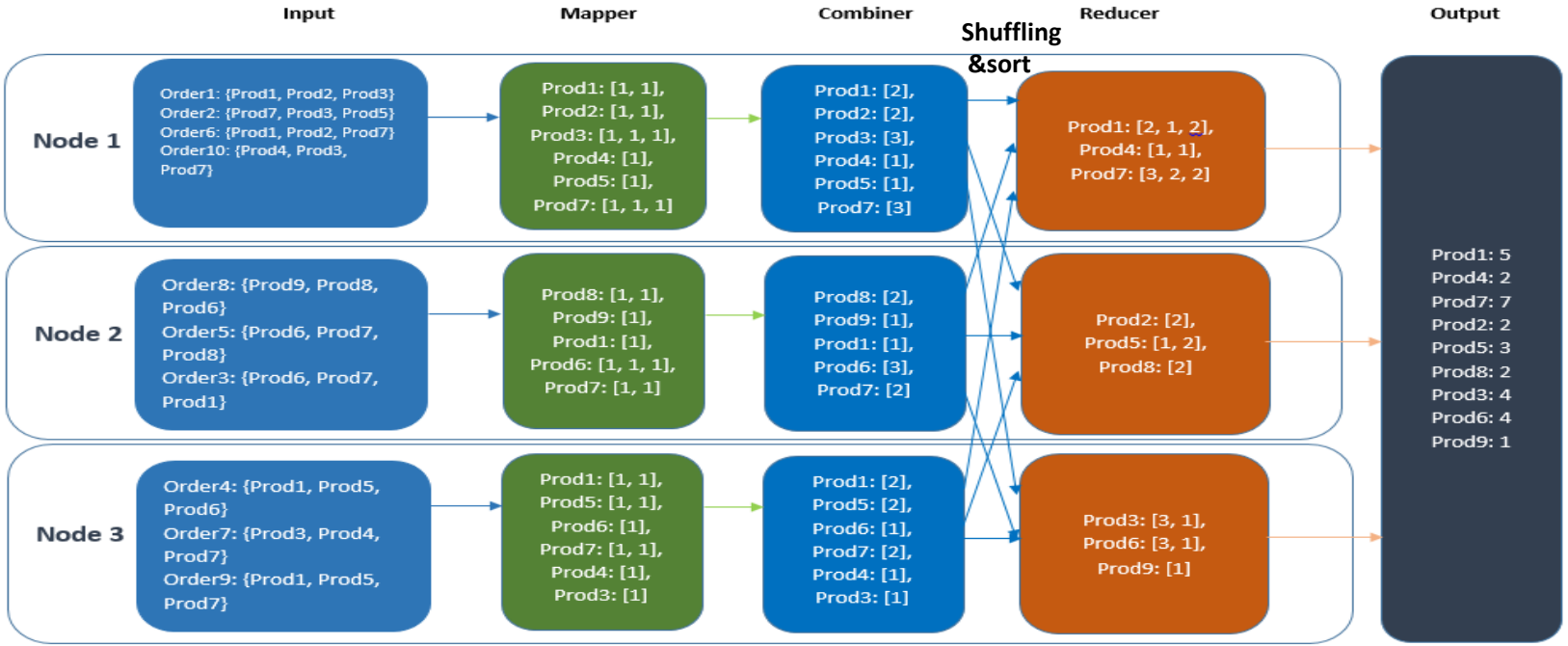
Ahmed Ramadan, ary00@fayoum.edu.eg

MapReduce without Combiner:



Ahmed Ramadan, ary00@fayoum.edu.eg

MapReduce with Combiner:



Ahmed Ramadan, ary00@fayoum.edu.eg

Combiner

- Moving data in cluster is one of the bottleneck in DS.
- We need to find a way to reduce the amount of data movement in cluster.
- Mapper produce a large amount of intermediate data.
- Solution: Using mini reducer (Combiner)

Ahmed Ramadan, ary00@fayoum.edu.eg

Combiner

- The Combiner class inherit the reducer class and override reduce method.
- The combiner run on the **Map-Side**.
- Combiner process on each map output <Key,Value> (Same as the reducer).
- Combiner and Reducer must have the **identical** input and output data types.
- The operation must be **associative and commutative**.

Ahmed Ramadan, ary00@fayoum.edu.eg

Combiner – Reducer

- Allow Combiner:
  - Sum operation.
  - Max operation.
  - Min operation.

- Disallow Combiner:
  - Average operation .
  - Mod operation.

```
job.setMapperClass(TokenizerMapper.class);
job.setCombinerClass(IntSumReducer.class);
job.setReducerClass(IntSumReducer.class);
```


Ahmed Ramadan, ary00@fayoum.edu.eg

Inverted index Problem

https://programmer.group/inverted-index-for-mapreduce-programming-development.html

Ahmed Ramadan, ary00@fayoum.edu.eg

QUESTIONS

Ahmed Ramadan, ary00@fayoum.edu.eg

THANK YOU!

Ahmed Ramadan, ary00@fayoum.edu.eg