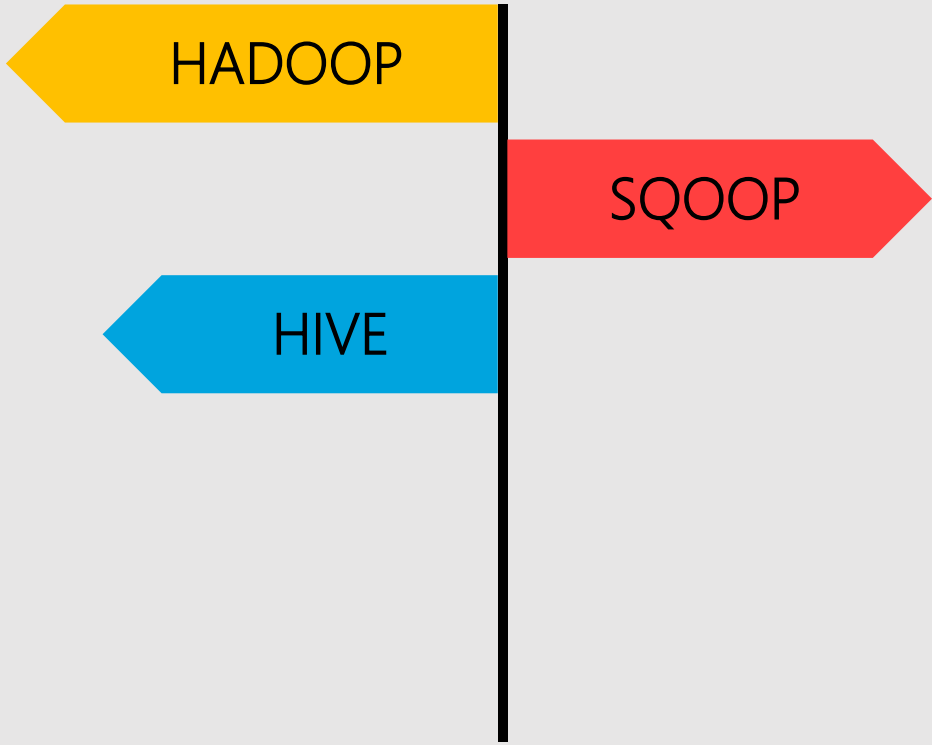
BIG DATA Analytics Using Apache Hadoop

Agenda

- Introducion to Hadoop & Configuration
- Assign 1. MapReduce

- Introduction to Sqoop & Configuration
- Assign 2. Sqoop

- Introduction to Hive & Configuration
- Assign 3. Hive



Apache Hadoop

- What is Hadoop ?

- Formally speaking, Hadoop is an open source framework for writing and running distributed applications that process large amounts of data

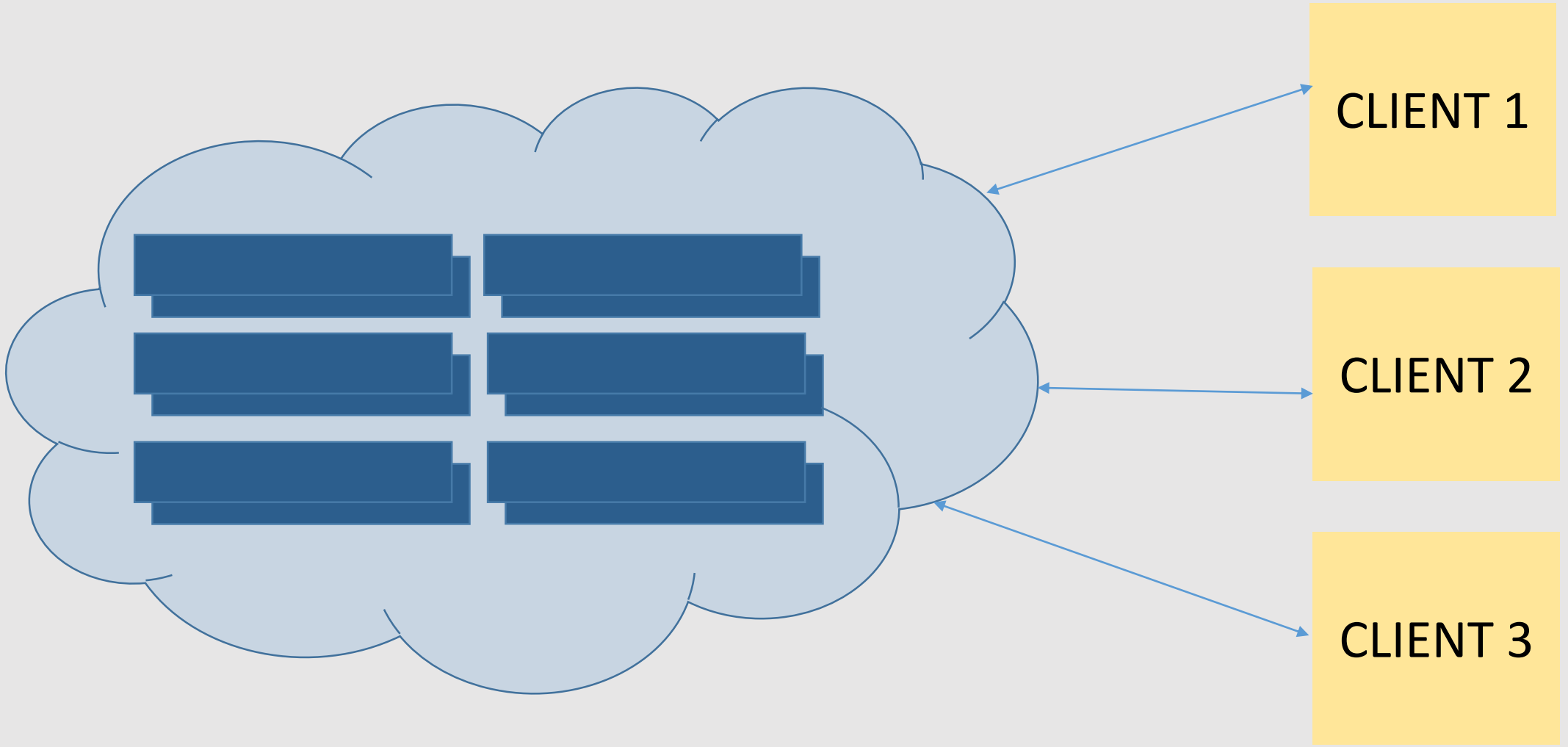
- Offline Line Batch processing.

Hadoop vs RDBMS

- Scale-Out instead of Scale-Up
- Key-Value pairs instead of Tables
- MapReduce Jobs instead of SQL

- Offline batch Processing instead of Online Transactions

Hadoop Cluster

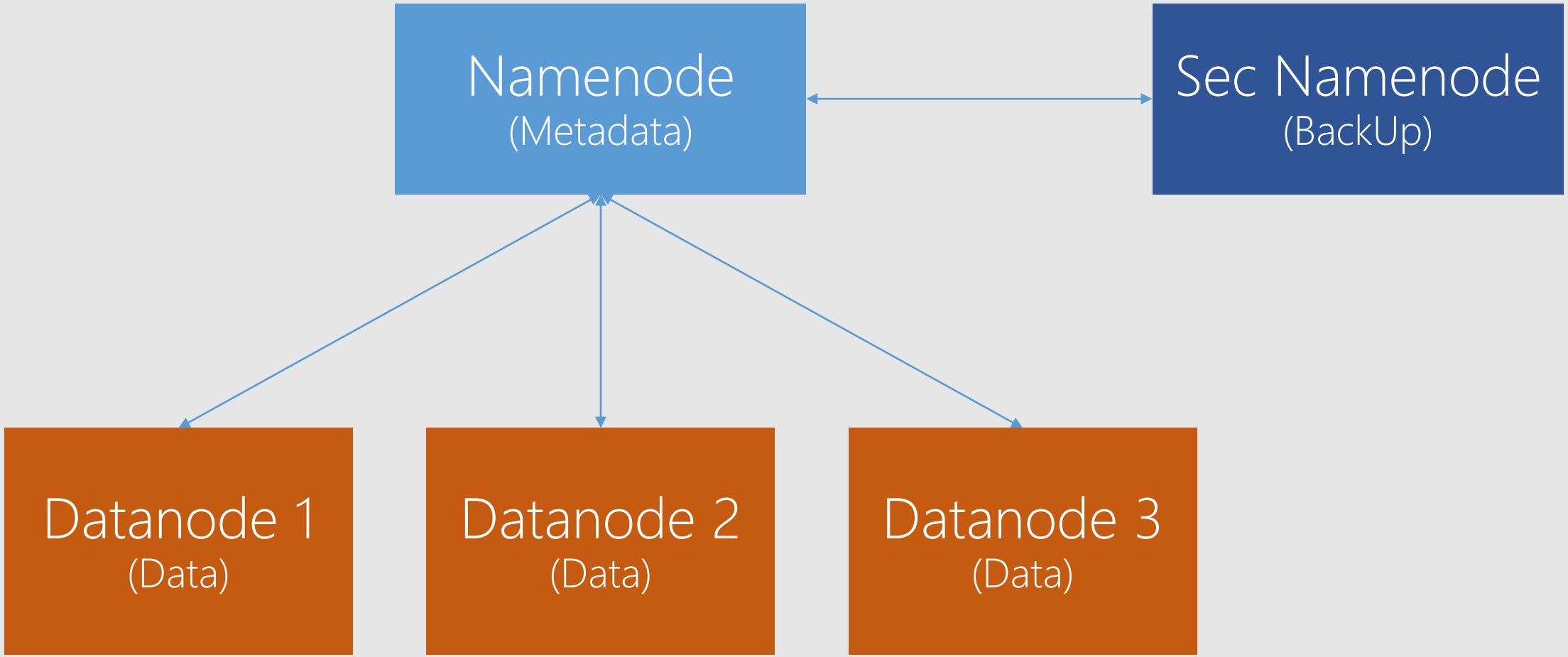


Components Of Hadoop (Daemons)

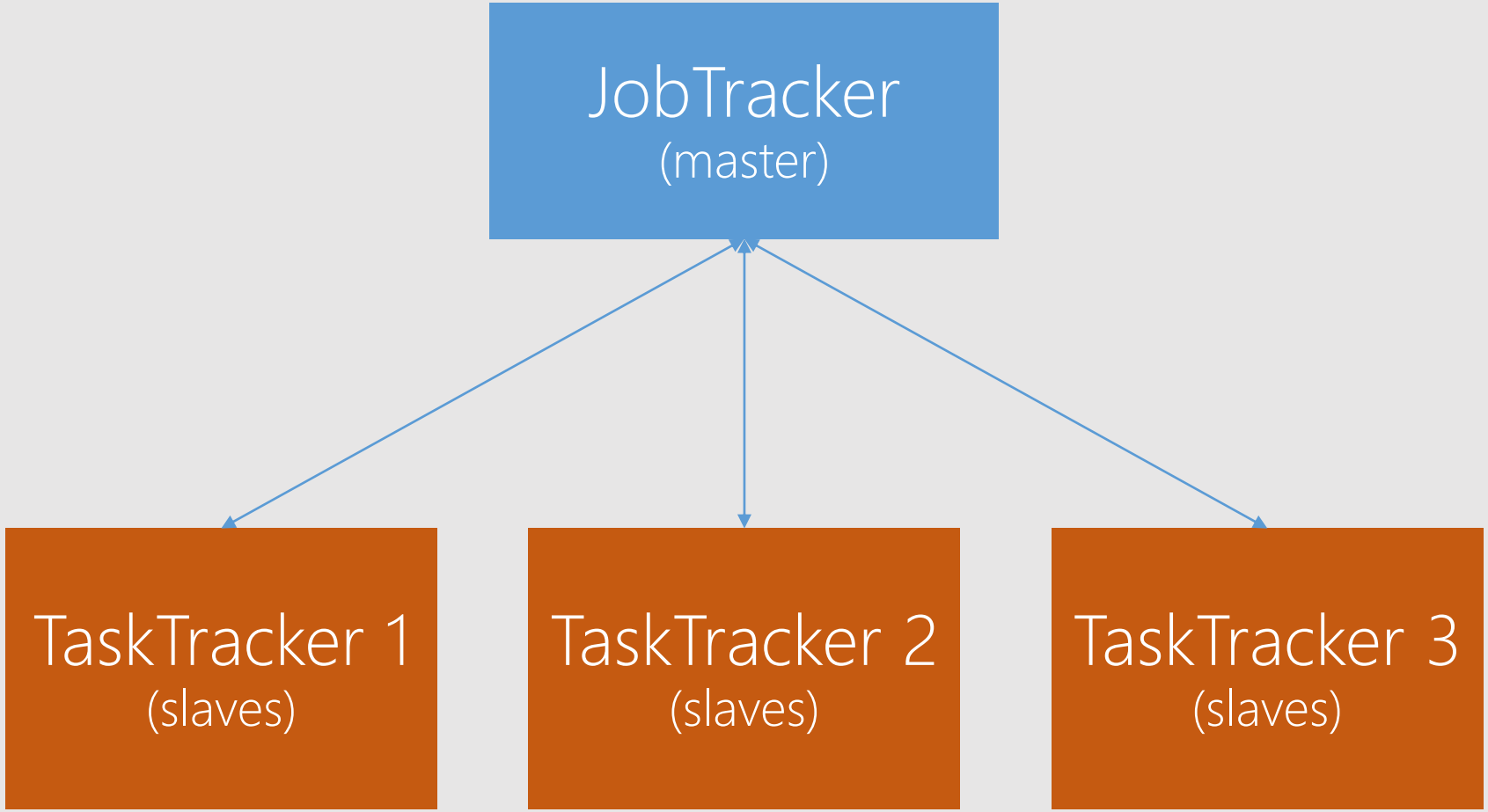
- Namenode
- Secondary namenode
- Datanode

- JobTracker / ResourceManager
- TaskTracker / NodeManager

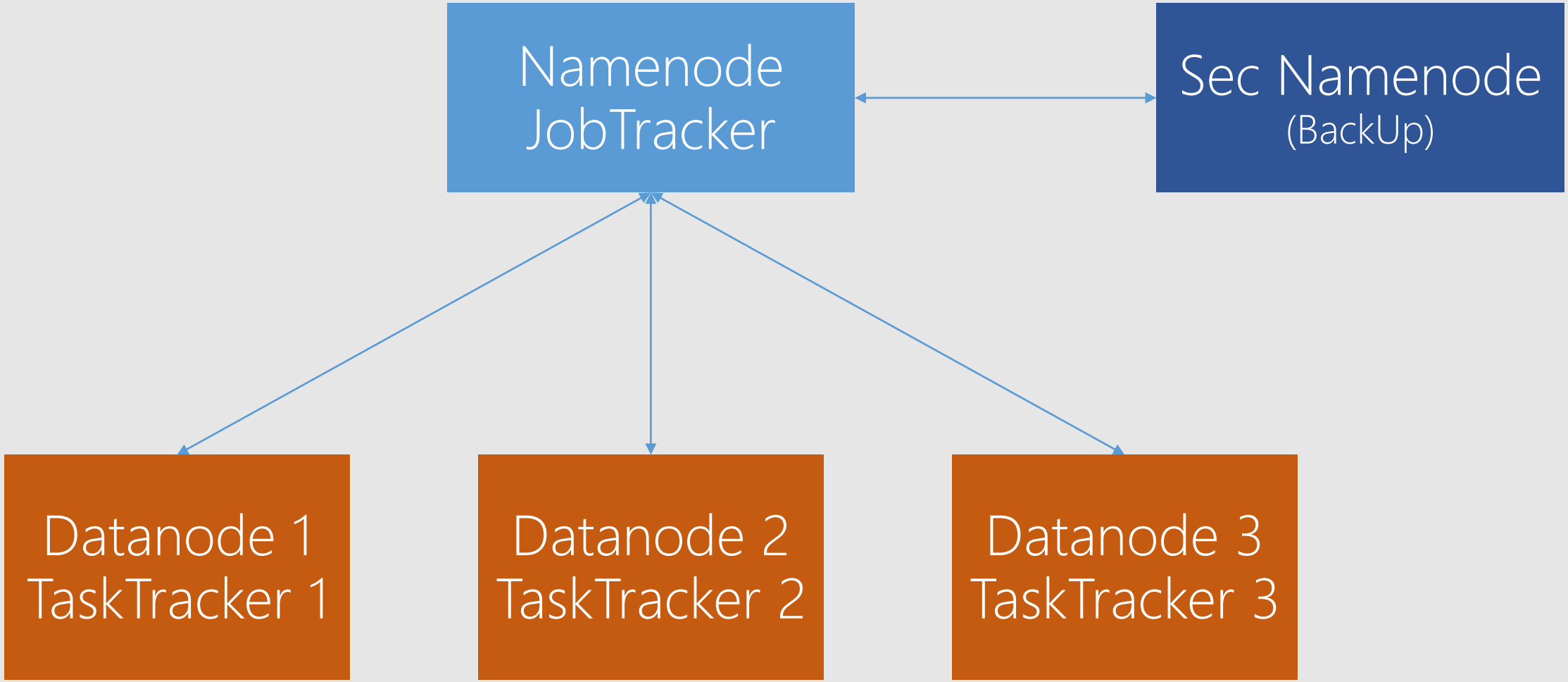
HDFS Daemons



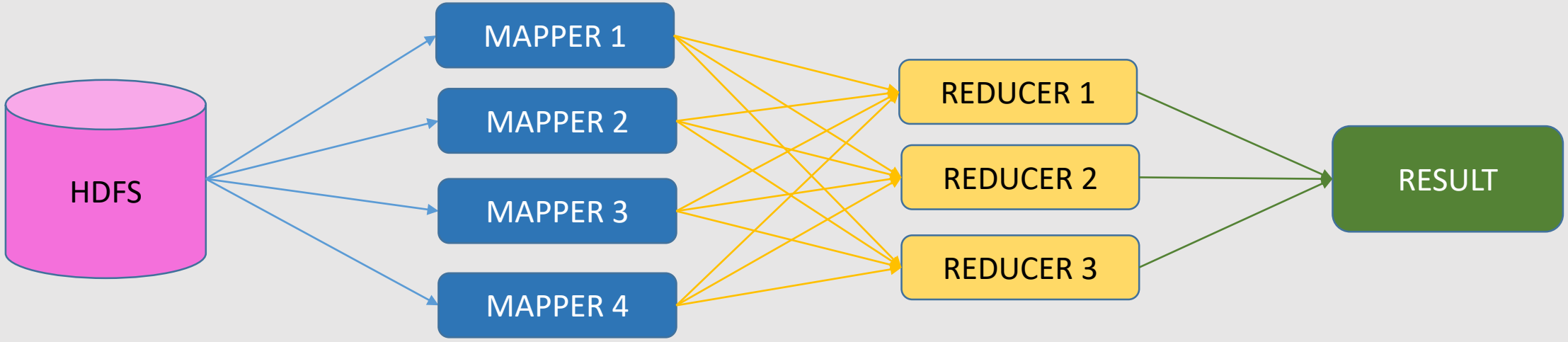
MapReduce Daemons



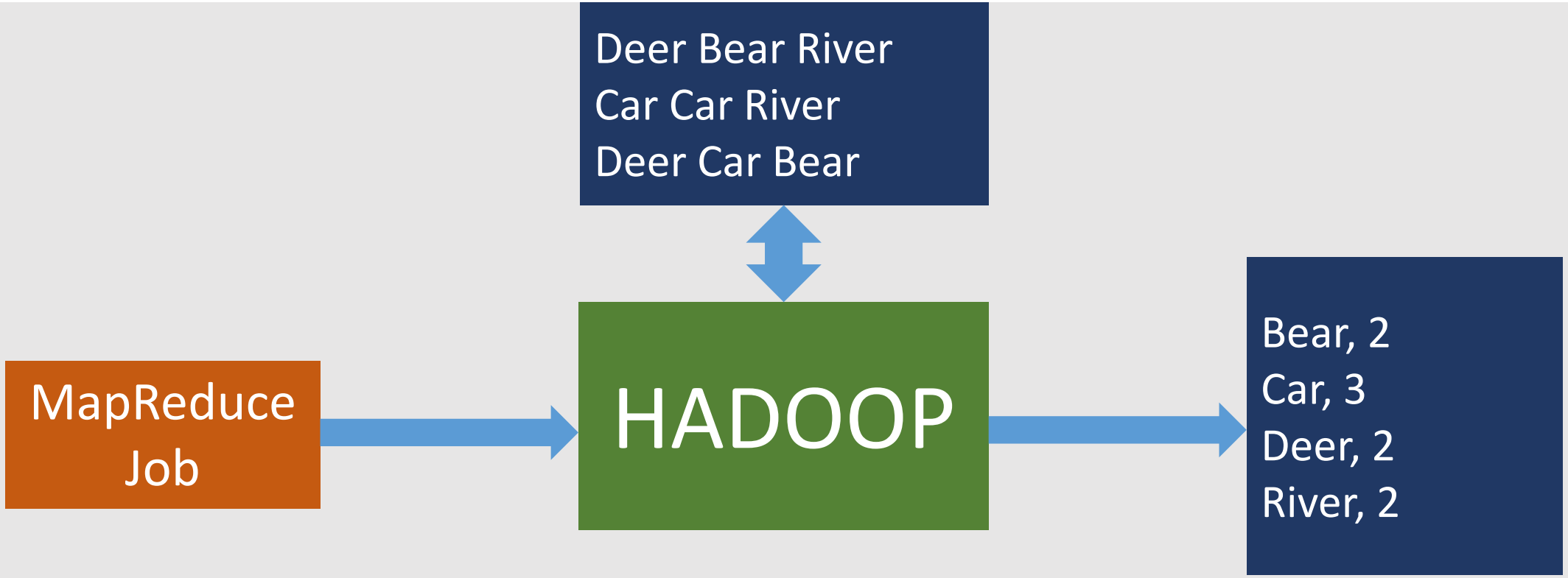
Real Cluster Scenario

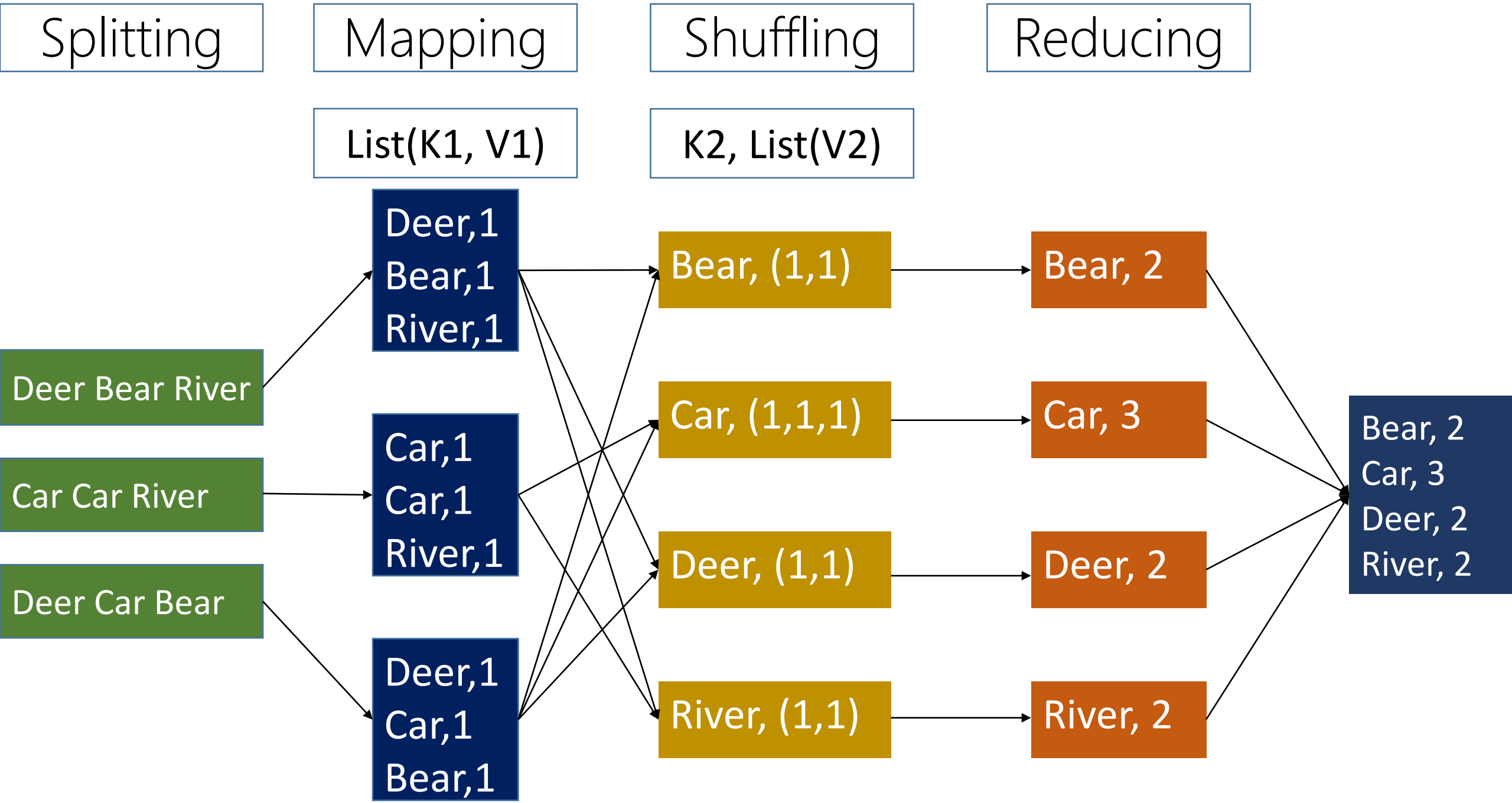


Hadoop In Action (MapReduce)



Executing MapReduce Job




Splitting
Mapping
Shuffling
Reducing
List(K1, V1)
K2, List(V2)
Deer Bear River
Car Car River
Deer Car Bear
Deer,1
Bear,1
River,1
Car,1
Car,1
River,1
Deer,1
Car,1
Bear,1
Bear, (1,1)
Car, (1,1,1)
Deer, (1,1)
River, (1,1)
Bear, 2
Car, 3
Deer, 2
River, 2
Bear, 2
Car, 3
Deer, 2
River, 2

Mapper for word count program

```
public static void map(key, value, outputCollector)
{
    String line = value.toString();
    StringTokenizer st = new StringTokenizer(line);
    while(st.hasNext()){
        outputCollector.collect(st.nextToken(), 1);
    }
}
```

Reducer for word count program

```
public static void reduce(key, values, outColl)
{
    int sum=0;
    while(values.hasNext()){
        sum += values.next();
    }
    outColl.collect(key, sum);
}
```

Apache Sqoop

unlocking Hadoop for RDBMS

Apache Hive

Data warehouse & Query Language for hadoop

References

- ‘Hadoop In Action’ , by Chuck Lam, dreamtech press

- Fundamentals of Hadoop MapReduce | Understanding Hadoop MapReduce | Hadoop Tutorial @

  https://youtu.be/-G8N_sjJZ4g?list=PL9ooVrP1hQOFrYxqxb0NJCdCABPZNo0pD

Thank You !