Lesson 3
MapReduce Programming Tasks Execution

Figure 4.2 MapReduce Programming Model

"Big Data Analytics ", Ch.04 L03: MapReduce, Hive and Pig
Raj Kamal, and Preeti Saxena © McGraw-Hill Education (India)
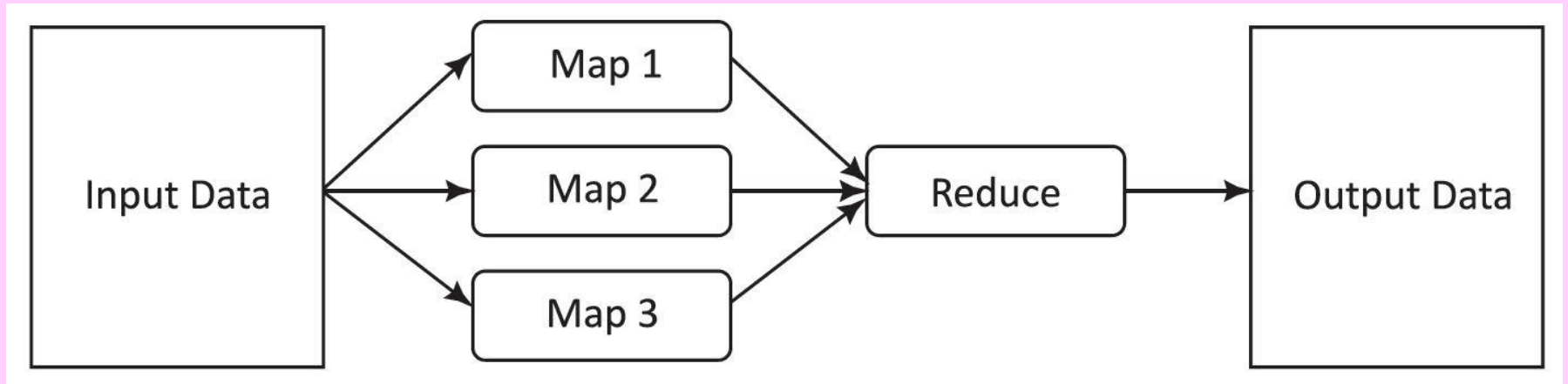
MapReduce Programming Model

Refer Figure 4.2

- Parallel processing of large data sets, based on a YARN-based system

- Each job consisting of number of Map and Reduce tasks, running in parallel

- Java Programming Framework

MapReduce Programming Model

Refer Figure 4.2 Data Flow

- Input Data flow to Map Task

- MapTask maps, shuffles, sorts and other operations

- Sends output to Reduce Task

"Big Data Analytics ", Ch.04 L03: MapReduce, Hive and Pig
Raj Kamal, and Preeti Saxena © McGraw-Hill Education (India)

MapReduce Programming Model

Refer Figure 4.2 Data Flow

- Reduce task produces results using combiner, aggregation, count or other reduce functions

- Output of Reducer Tasks flows to ─ Application Support APIs and Application

Map task

- Implements a map (), which runs user application codes for each key-value pair (**k1**, **v1**).

- Key **k1**─ a set of keys.

- **k1** maps to a group of data values **v1**. (Section 3.3.1)

- Values **v1**─ a large string which is read from the input file(s)

"Big Data Analytics ", Ch.04 L03: MapReduce, Hive and Pig
Raj Kamal, and Preeti Saxena © McGraw-Hill Education (India)

Figure 4.3 MapReduce process on client submitting a job

"Big Data Analytics ", Ch.04 L03: MapReduce, Hive and Pig
Raj Kamal, and Preeti Saxena © McGraw-Hill Education (India)
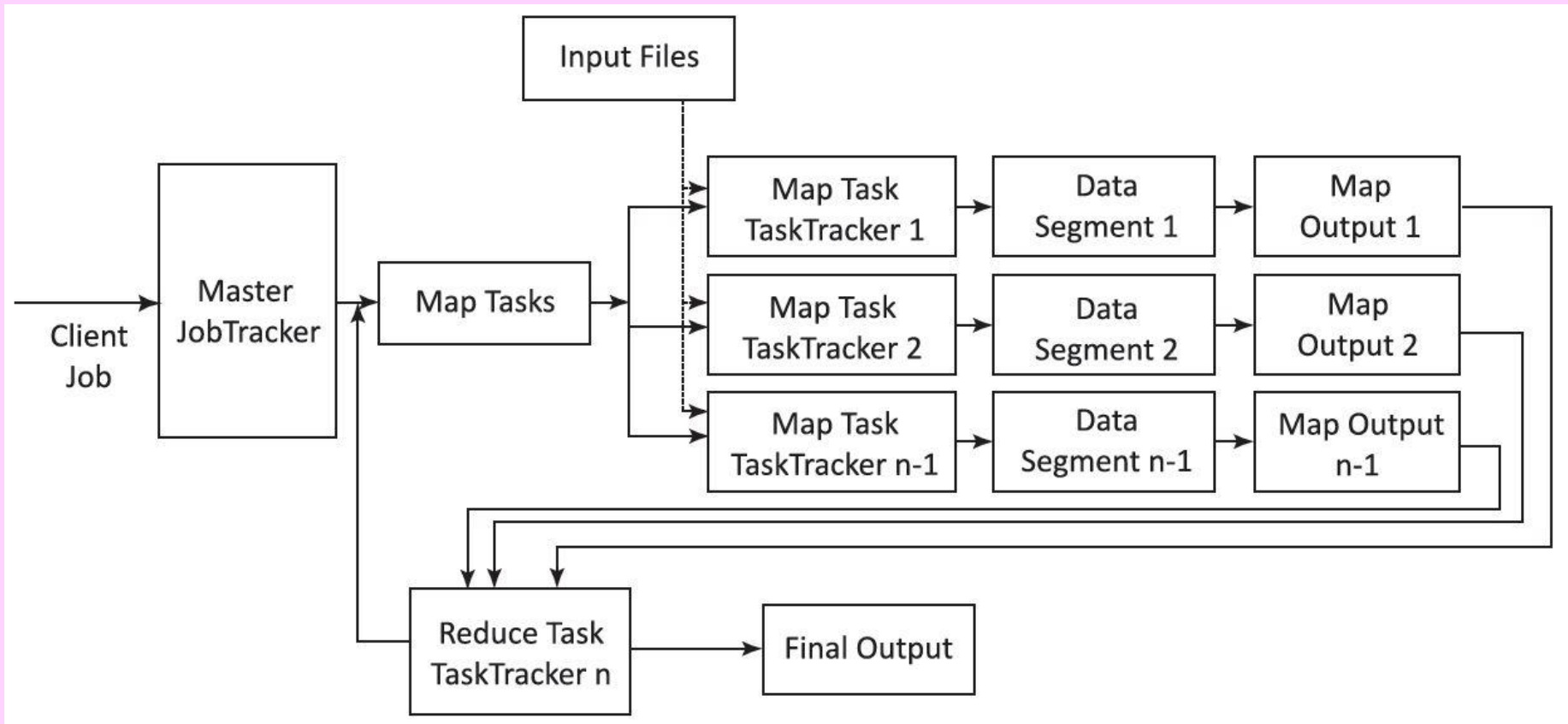
Figure 4.4 Logical view of functioning of map()

map(Key1, value1)   ⟶   List (Key2, value2)

Input in the form of key-value pair.

Zero or intermediate output, key-value pairs produced.
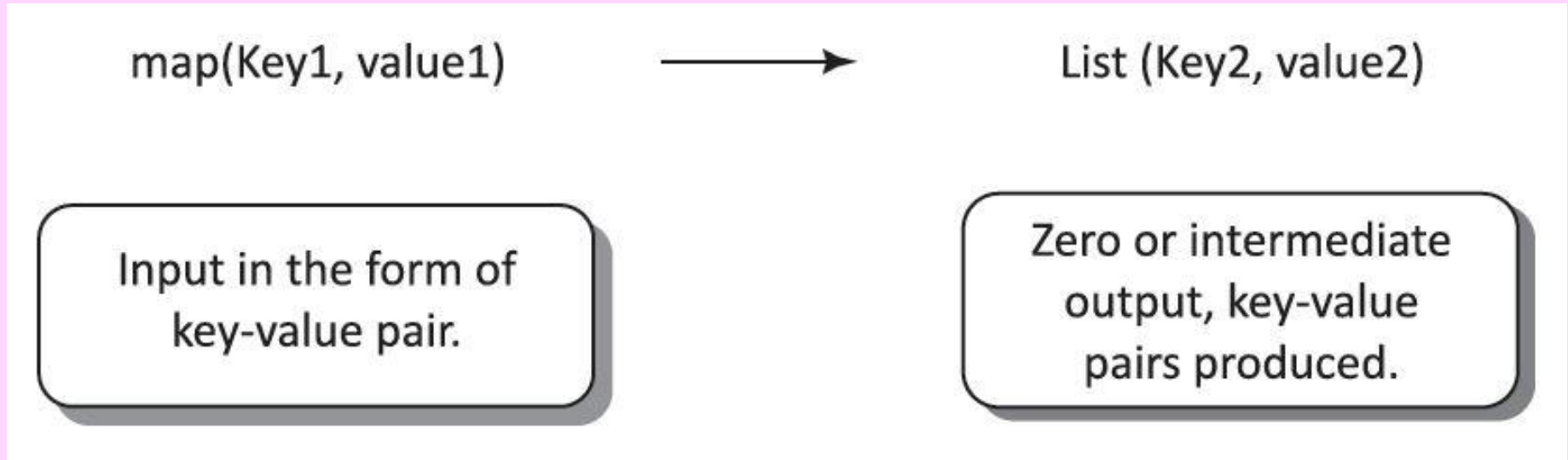
Reduce task

- Takes the output **v2** from the map

- *Combiner* combines the input those data pieces into a smaller set of data using reducer functions

- Reducer always execute after the Mapper

Example 1

- Number of Car models (Jaguar Land Rover, Zest, …) sold by number of showrooms of the company (ID1, ID2, ..) on different dates in a year (Date1, Date2)

"Big Data Analytics ", Ch.04 L03: MapReduce, Hive and Pig
Raj Kamal, and Preeti Saxena © McGraw-Hill Education (India)

Key-Value-Pairs

- Consider key-value pairs which are input from client APIs

- (showRoomID1, Date1, JaguarSales); (showRoomID2, Date1, ZestSales); (showRoomID2, Date2, JaguarSales); (showRoomID1, Date2, JaguarSales); (showRoomID1, Date1, ZestSales); .

  …..

Figure 4.5 Key-value pairing in MapReduce
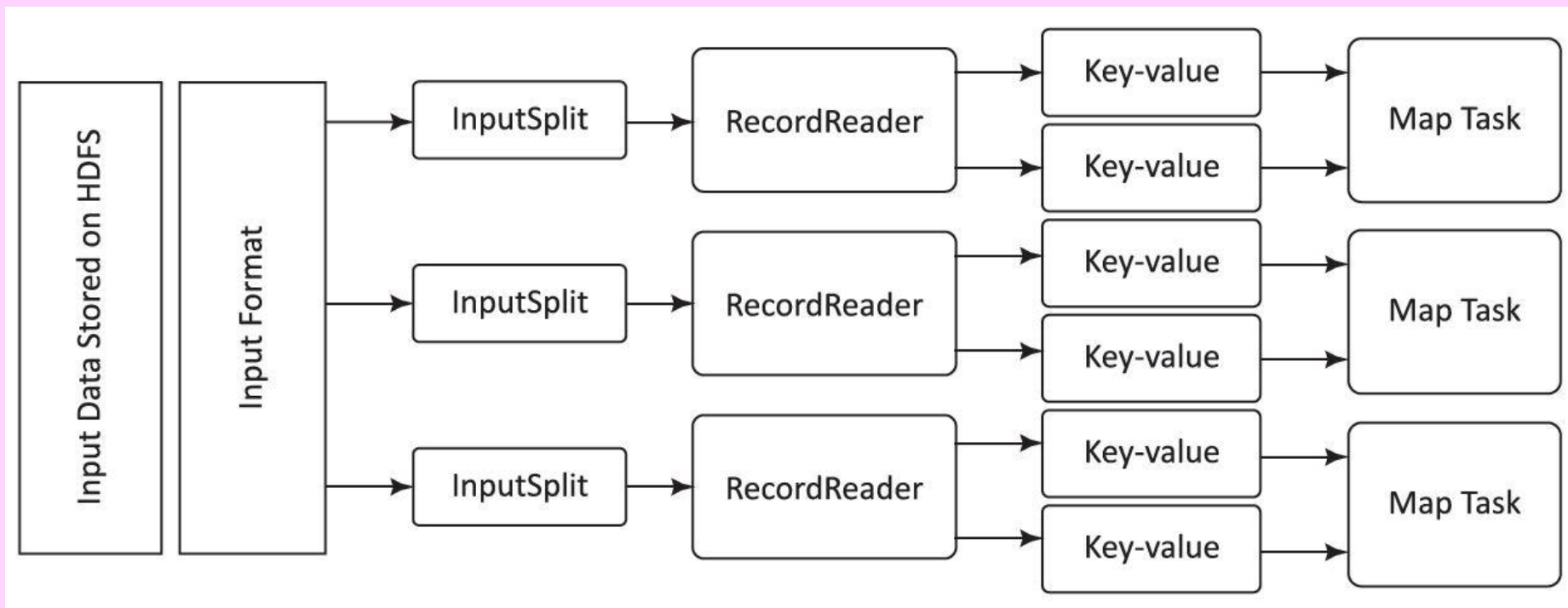
Summary

We learnt :

- Application Client API submits a Job
- Job consists of map and reduce tasks
- Execution in parallel of the map ()

"Big Data Analytics ", Ch.04 L03: MapReduce, Hive and Pig
Raj Kamal, and Preeti Saxena © McGraw-Hill Education (India)

Summary

We learnt :

- Reduce () executes after map () using combiners and reducer functions

- Response (output data) flows to Client API

"Big Data Analytics ", Ch.04 L03: MapReduce, Hive and Pig
Raj Kamal, and Preeti Saxena © McGraw-Hill Education (India)

End of Lesson 3 on MapReduce Programming Tasks Execution

"Big Data Analytics ", Ch.04 L03: MapReduce, Hive and Pig
Raj Kamal, and Preeti Saxena © McGraw-Hill Education (India)