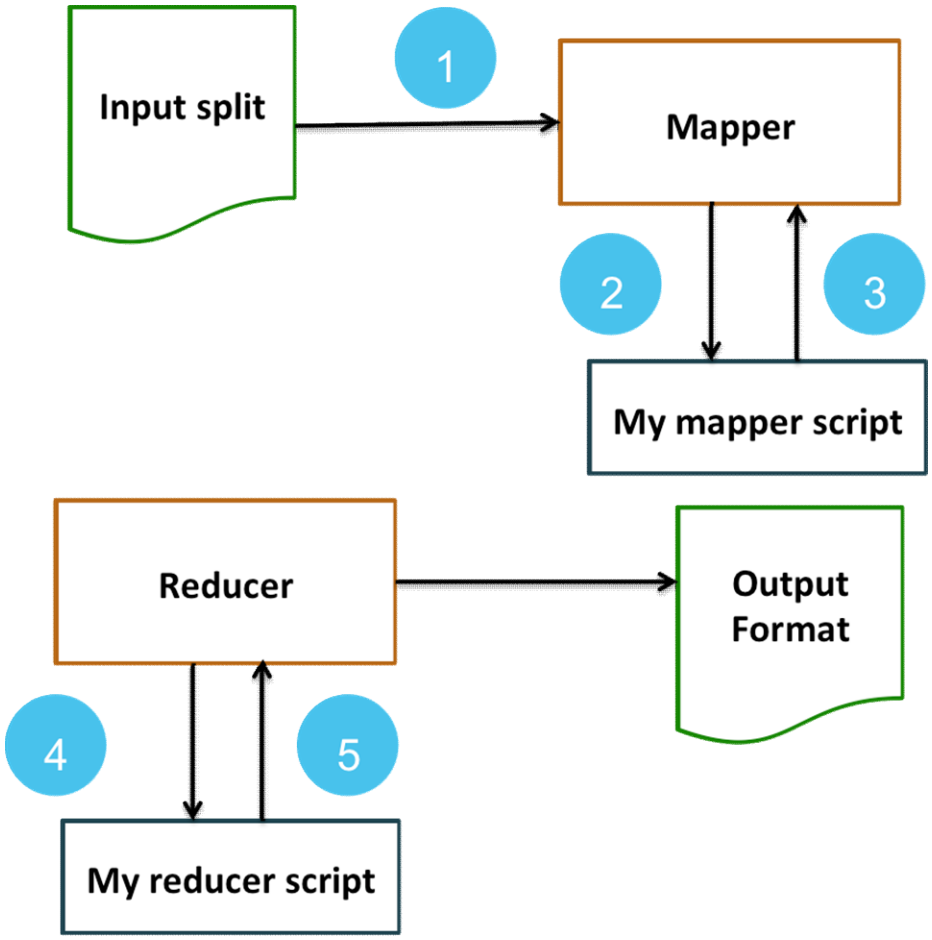

1. The InputFormat generates <k1,v1> pairs
2. Converts the <k1,v1> pairs to lines and sends them to the stdin of the process
3. The stdout of the process is converted into <k2,v2> pairs
4. Converts <k2,(v2,v2,...)> pairs to lines and send them to the stdin of the process
5. The stdout of the process is converted into <k3, v3> pairs

Running a Hadoop Streaming Job

A Streaming job is a MapReduce job defined in the hadoop-streaming.jar file:

```
hadoop  jar $HADOOP_HOME/lib/hadoop-streaming.jar
    -input input_directories
    -output output_directories
    -mapper mapper_script
    -reducer reducer_script
```