# HADOOP WORD COUNT

C:\hadoop_new\share\hadoop\mapreduce>hadoop jar
hadoop-mapreduce-examples-3.1.0.jar
An example program must be given as the first argument.
Valid program names are:
  aggregatewordcount: An Aggregate based map/reduce program that counts the words in the input files.
  aggregatewordhist: An Aggregate based map/reduce program that computes the histogram of the words in the input files.
  bbp: A map/reduce program that uses Bailey-Borwein-Plouffe to compute exact digits of Pi.
  dbcount: An example job that count the pageview counts from a database.
  distbbp: A map/reduce program that uses a BBP-type formula to compute exact bits of Pi.
  grep: A map/reduce program that counts the matches of a regex in the input.
  join: A job that effects a join over sorted, equally partitioned datasets
  multifilewc: A job that counts words from several files.
  pentomino: A map/reduce tile laying program to find solutions to pentomino problems.
  pi: A map/reduce program that estimates Pi using a quasi-Monte Carlo method.
  randomtextwriter: A map/reduce program that writes 10GB of random textual data per node.
  randomwriter: A map/reduce program that writes 10GB of random data per node.
  secondarysort: An example defining a secondary sort to the reduce.
  sort: A map/reduce program that sorts the data written by the random writer.
  sudoku: A sudoku solver.
  teragen: Generate data for the terasort
  terasort: Run the terasort
  teravalidate: Checking results of terasort
  wordcount: A map/reduce program that counts the words in the input files.
  wordmean: A map/reduce program that counts the average length of the words in the input files.
  wordmedian: A map/reduce program that counts the median length of the words in the input files.

wordstandarddeviation: A map/reduce program that counts the standard deviation of the length of the words in the input files.

C:\hadoop_new\share\hadoop\mapreduce>hdfs dfs -mkdir /input

C:\hadoop_new\share\hadoop\mapreduce>hdfs dfs -put E:\Desktop\input.txt \input

C:\hadoop_new\share\hadoop\mapreduce>hdfs dfs -ls \input
Found 1 items
-rw-r--r--   1 Admin supergroup       45 2021-04-26 14:46 /input/input.txt

C:\hadoop_new\share\hadoop\mapreduce>hdfs dfs -cat \input\input.txt
deer bear river
car car river
deer car bear

C:\hadoop_new\share\hadoop\mapreduce>hadoop jar hadoop-mapreduce-examples-3.1.0.jar wordcount \input \firstExampleOutput
2021-04-26 16:04:57,079 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2021-04-26 16:04:59,541 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/Admin/.staging/job_1619433198405_0001
2021-04-26 16:05:02,640 INFO input.FileInputFormat: Total input files to process : 1
2021-04-26 16:05:04,138 INFO mapreduce.JobSubmitter: number of splits:1
2021-04-26 16:05:04,450 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
2021-04-26 16:05:05,111 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1619433198405_0001
2021-04-26 16:05:05,114 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-04-26 16:05:05,609 INFO conf.Configuration: resource-types.xml not found
2021-04-26 16:05:05,614 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.

2021-04-26 16:05:06,486 INFO impl.YarnClientImpl: Submitted application application_1619433198405_0001
2021-04-26 16:05:06,666 INFO mapreduce.Job: The url to track the job: http://DESKTOP-CK14PFH:8088/proxy/application_1619433198405_0001/
2021-04-26 16:05:06,668 INFO mapreduce.Job: Running job: job_1619433198405_0001
2021-04-26 16:05:39,704 INFO mapreduce.Job: Job job_1619433198405_0001 running in uber mode : false
2021-04-26 16:05:39,806 INFO mapreduce.Job:  map 0% reduce 0%
2021-04-26 16:06:04,188 INFO mapreduce.Job:  map 100% reduce 0%
2021-04-26 16:06:41,797 INFO mapreduce.Job:  map 100% reduce 100%
2021-04-26 16:06:46,907 INFO mapreduce.Job: Job job_1619433198405_0001 completed successfully
2021-04-26 16:06:47,311 INFO mapreduce.Job: Counters: 53
        File System Counters
                FILE: Number of bytes read=50
                FILE: Number of bytes written=427745
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=147
                HDFS: Number of bytes written=28
                HDFS: Number of read operations=8
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Launched reduce tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=19887
                Total time spent by all reduces in occupied slots (ms)=34624
                Total time spent by all map tasks (ms)=19887
                Total time spent by all reduce tasks (ms)=34624
                Total vcore-milliseconds taken by all map tasks=19887
                Total vcore-milliseconds taken by all reduce tasks=34624
                Total megabyte-milliseconds taken by all map tasks=20364288
                Total megabyte-milliseconds taken by all reduce tasks=35454976

Map-Reduce Framework
     Map input records=3
     Map output records=9
     Map output bytes=80
     Map output materialized bytes=50
     Input split bytes=102
     Combine input records=9
     Combine output records=4
     Reduce input groups=4
     Reduce shuffle bytes=50
     Reduce input records=4
     Reduce output records=4
     Spilled Records=8
     Shuffled Maps =1
     Failed Shuffles=0
     Merged Map outputs=1
     GC time elapsed (ms)=192
     CPU time spent (ms)=3855
     Physical memory (bytes) snapshot=454619136
     Virtual memory (bytes) snapshot=598560768
     Total committed heap usage (bytes)=245891072
     Peak Map Physical memory (bytes)=269107200
     Peak Map Virtual memory (bytes)=346247168
     Peak Reduce Physical memory (bytes)=185511936
     Peak Reduce Virtual memory (bytes)=252313600
Shuffle Errors
     BAD_ID=0
     CONNECTION=0
     IO_ERROR=0
     WRONG_LENGTH=0
     WRONG_MAP=0
     WRONG_REDUCE=0
File Input Format Counters
     Bytes Read=45
File Output Format Counters
     Bytes Written=28

```
C:\hadoop_new\share\hadoop\mapreduce>hdfs dfs -ls \input
Found 1 items
-rw-r--r--   1 Admin supergroup        45 2021-04-26 14:46 /input/input.txt

C:\hadoop_new\share\hadoop\mapreduce>hdfs dfs -ls \firstExampleOutput
Found 2 items
-rw-r--r--   1 Admin supergroup         0 2021-04-26 16:06
/firstExampleOutput/_SUCCESS
-rw-r--r--   1 Admin supergroup        28 2021-04-26 16:06
/firstExampleOutput/part-r-00000

C:\hadoop_new\share\hadoop\mapreduce>hdfs dfs -cat
\firstExampleOutput\part-r-00000
bear   2
car    3
deer   2
river  2
```