

به نام خدا

تکلیف پنجم درس رایانش ابری - Hadoop and MapReduce

محمد جلالی و علیرضا صالحی

9725113

(الف)

فایل کد پیوست شده است.

(ب)

با توجه به خروجی لاگ:

#### File System Counters

FILE: Number of bytes read=510388922  
FILE: Number of bytes written=639254698  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=127986788  
HDFS: Number of bytes written=391865  
HDFS: Number of read operations=15  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=4  
HDFS: Number of bytes read erasure-coded=0

#### Map-Reduce Framework

Map input records=8616  
Map output records=10600367  
Map output bytes=106394862  
Map output materialized bytes=127595602  
Input split bytes=108  
Combine input records=0  
Combine output records=0  
Reduce input groups=32437  
Reduce shuffle bytes=127595602  
Reduce input records=10600367  
Reduce output records=32437  
Spilled Records=31801101  
**Shuffled Maps =1**  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=47

Total committed heap usage (bytes)=1393557504

و اینکه به طور دیفالت ما یک Reducer داریم پس Suffled Maps تعداد Map هایمان را نشان میدهد که برابر ۱ است.

(ج)

با توجه به خروجی لاگ:

#### Map-Reduce Framework

Map input records=86160  
Map output records=106003670  
Map output bytes=1063948620  
Map output materialized bytes=1275956020  
Input split bytes=1080  
Combine input records=0  
Combine output records=0  
Reduce input groups=32437  
Reduce shuffle bytes=1275956020  
Reduce input records=106003670  
Reduce output records=32437  
Spilled Records=318011010  
**Shuffled Maps =10**  
Failed Shuffles=0  
Merged Map outputs=10  
GC time elapsed (ms)=1307  
Total committed heap usage (bytes)=12327059456

و پیشفرض داشتن یک Reducer پس یعنی ما ۱۰ تا Map داشتیم که هر کدام برای یک فایل بودند.

(د) طبق لاگ و توضیحات قبلی ۱ Reducer و ۵ تا مپر داریم.

Map input records=86160  
Map output records=106003670  
Map output bytes=1063948620  
Map output materialized bytes=1275955990  
Input split bytes=575  
Combine input records=0  
Combine output records=0  
Reduce input groups=32437  
Reduce shuffle bytes=1275955990  
Reduce input records=106003670  
Reduce output records=32437  
Spilled Records=318011010  
**Shuffled Maps =5**  
Failed Shuffles=0  
Merged Map outputs=5  
GC time elapsed (ms)=621  
Total committed heap usage (bytes)=6138888192

ه) برای فایل ۳۰۰ مگابایتی داریم:

```
Map input records=43080
Map output records=53001835
Map output bytes=531974310
Map output materialized bytes=637977998
Input split bytes=336
Combine input records=0
Combine output records=0
Reduce input groups=32437
Reduce shuffle bytes=637977998
Reduce input records=53001835
Reduce output records=32437
Spilled Records=159005505
Shuffled Maps =3
Failed Shuffles=0
Merged Map outputs=3
GC time elapsed (ms)=363
Total committed heap usage (bytes)=3359113216
```

پس تعداد reducer ها برابر ۱ و mapper ها برابر ۳ است.

و) تعداد ۲ را برای reducer ها ست کردیم و همانطور که میبینیم Shuffled maps برابر ۶ شده یعنی  $6/2$  که برابر ۳ است میسر داشتیم. تغییری در تعداد میسر ها رخ نداده.

```
Map input records=43080
Map output records=53001835
Map output bytes=531974310
Map output materialized bytes=637978016
Input split bytes=336
Combine input records=0
Combine output records=0
Reduce input groups=32437
Reduce shuffle bytes=637978016
Reduce input records=53001835
Reduce output records=32437
Spilled Records=159005505
Shuffled Maps =6
Failed Shuffles=0
Merged Map outputs=6
GC time elapsed (ms)=356
```

ز) طبق جستجو ها و نتایج مشاهده شده تعداد مپ ها بر اساس سایز hdfs بلاک ها هستند که با تغییر آن ها میتوانیم این تعداد را تغییر دهیم همینطور به تعداد فایل هم بستگی داریم طبق آزمایشاتمان که تعداد فایل ها کوچک ولی ده تا بود و تعداد مپ ها ده تا شد. نکته ی دیگر این بود که اگر تعداد reducer ها را زیاد کنیم تغییری ایجاد نشد.