Report Of Asg-2

1.

Bu çalışmada pairs ve strips mapreduce job ları için oluşturulan frekanslara PMI fonksiyonu uygulanmıştır, ben bu uyguların herbirini ard arda iki job olarak gerçekleştirdim;

Pairs pmi değeri için iki farklı eclipse dosyası oluşuturup (entegre çalışacak şekilde) ilk java kodu ile pairs sayısı ve pairler hesaplanıp, pairsden aldığı çıktı ile pmi hesaplayan ikinci bir java kodu oluşturdum.

Stripes pmi değeri için iki farklı eclipse dosyası oluşuturup (entegre çalışacak şekilde) ilk java kodu ile stripes sayısı ve stripeslar hesaplanıp, stripesdan aldığı çıktı ile pmi hesaplayan ikinci bir java kodu oluşturdum.

Yapılan mapreduce joblarının ayrıntıları aşağıdadır;

İlk olarak pairs işlemi için veriler;

Map-Reduce Framework

Map input records=124788

Map output records=6538544

Map output bytes=79914301

Map output materialized bytes=92991395

Input split bytes=119

Combine input records=0

Combine output records=0

Reduce input groups=1016446

Reduce shuffle bytes=92991395

Reduce input records=6538544

Reduce output records=34052

Spilled Records=19615632

Shuffled Maps =1

Failed Shuffles=0

Merged Map outputs=1

GC time elapsed (ms)=68

Total committed heap usage (bytes)=681050112

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=5589917

File Output Format Counters

Bytes Written=845797

Pairs işlemi için yapılan pmi verileri;

Map-Reduce Framework

Map input records=34052

Map output records=68104

Map output bytes=1349311

Map output materialized bytes=1485525

Input split bytes=142

Combine input records=0

Combine output records=0

Reduce input groups=36421

Reduce shuffle bytes=1485525

Reduce input records=68104

Reduce output records=34052

Spilled Records=136208

Shuffled Maps =1

Failed Shuffles=0

Merged Map outputs=1

GC time elapsed (ms)=9

Total committed heap usage (bytes)=574095360

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=845797

File Output Format Counters

Bytes Written=1704431

Stripes işlemi için veriler;

Map-Reduce Framework

Map input records=124788

Map output records=1006624

Map output bytes=47676213

Map output materialized bytes=49689467

Input split bytes=119

Combine input records=0

Combine output records=0

Reduce input groups=29543

Reduce shuffle bytes=49689467

Reduce input records=1006624

Reduce output records=29542

Spilled Records=2013248

Shuffled Maps =1

Failed Shuffles=0

Merged Map outputs=1

GC time elapsed (ms)=59

Total committed heap usage (bytes)=672137216

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=5589917

File Output Format Counters

Bytes Written=856981

Stripes pmi işlemi için veriler;

Map-Reduce Framework

Map input records=29542

Map output records=68104

Map output bytes=1333543

Map output materialized bytes=1469757

Input split bytes=137

Combine input records=0

Combine output records=0

Reduce input groups=36421

Reduce shuffle bytes=1469757

Reduce input records=68104

Reduce output records=34052

Spilled Records=136208

Shuffled Maps =1

Failed Shuffles=0

Merged Map outputs=1

GC time elapsed (ms)=11

Total committed heap usage (bytes)=576716800

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=856981

File Output Format Counters

Bytes Written=999501

2. Pairs ve Stripes için işlem süreleri;

Pairs job 1 için; 20.735 seconds

Pairs job 2 için; 2.692 second

Stripes job 1 için; 9.034 seconds

Stripes job 2 için; 2.678 seconds

3. Ard arda iki job kullandığım ve kullandığım algoritma yapısından dolayı combiner kullanmadım.

4. 68109 pmi değeri elde edilmiştir.

5. En yüksek elde edilen pmi değeri;

hangs the -0.001177273685707848 kelimeleri içindir.