Assignment 3 - Part A (Utilizing AIR for Data Processing)

Name: Shivankar Pilligundla
Roll No: IMT2020016

Similar workflow as YSB is used to do the regex pattern match processing. Initially event generator generates a random 50 char pattern. Then its passed to an event filter within the same rank. This filters out the patterns based on regex and forwards the events to partial aggregator of same rank. Here window id's are calculated based on event time and count with in the rank is computed. Now these partial aggregators share their events with full aggregators of every possible rank(i.e all dataflows) and they unitedly perform final aggregation. Then these are forwarded to a collector at rank 0 that aggregates all the results and computes the avg latencies.

Results:

For the throughput of 100 the following latencies have been reported:

```
~/AIR/Release master !10 ?7 > mpirun -np 4 ./AIR YSB 100

*******************AIR (c) 2020 Uni.lu*******************

AIR INSTANCE AT RANK 2/4 | TP: 100 | MSG/SEC/RANK: 2 | AGGR_WINDOW: 10000ms
AIR INSTANCE AT RANK 3/4 | TP: 100 | MSG/SEC/RANK: 2 | AGGR_WINDOW: 10000ms
AIR INSTANCE AT RANK 4/4 | TP: 100 | MSG/SEC/RANK: 2 | AGGR_WINDOW: 10000ms
AIR INSTANCE AT RANK 1/4 | TP: 100 | MSG/SEC/RANK: 2 | AGGR_WINDOW: 10000ms
 WID: 0 count: 144      latency: 1001 SIZE 0 CAP 2400

  #1 COUNT: 144 AVG_LATENCY: 1001     N=1

 WID: 1 count: 384      latency: 1 SIZE 0 CAP 2400

  #2 COUNT: 384 AVG_LATENCY: 501      N=1

 WID: 2 count: 240      latency: 1 SIZE 0 CAP 2400

  #3 COUNT: 240 AVG_LATENCY: 334      N=1

 WID: 3 count: 384      latency: 1001 SIZE 0 CAP 2400

  #4 COUNT: 384 AVG_LATENCY: 501      N=1

 WID: 4 count: 288      latency: 1001 SIZE 0 CAP 2400

  #5 COUNT: 288 AVG_LATENCY: 601      N=1

 WID: 5 count: 96       latency: 1 SIZE 0 CAP 2400

  #6 COUNT: 96  AVG_LATENCY: 501      N=1

 WID: 6 count: 288      latency: 2001 SIZE 0 CAP 2400

  #7 COUNT: 288 AVG_LATENCY: 715      N=1

 WID: 7 count: 48       latency: 6001 SIZE 0 CAP 2400

  #8 COUNT: 48  AVG_LATENCY: 1376     N=1

 WID: 8 count: 336      latency: 2001 SIZE 0 CAP 2400

  #9 COUNT: 336 AVG_LATENCY: 1445     N=1

 WID: 9 count: 240      latency: 1001 SIZE 0 CAP 2400

  #10 COUNT: 240        AVG_LATENCY: 1401       N=1

 WID: 10       count: 240      latency: 1 SIZE 0 CAP 2400

  #11 COUNT: 240        AVG_LATENCY: 1273       N=1

 WID: 11       count: 336      latency: 1001 SIZE 0 CAP 2400

  #12 COUNT: 336        AVG_LATENCY: 1251       N=1

 WID: 12       count: 384      latency: 1001 SIZE 0 CAP 2400

  #13 COUNT: 384        AVG_LATENCY: 1231       N=1

 WID: 13       count: 288      latency: 1 SIZE 0 CAP 2400

  #14 COUNT: 288        AVG_LATENCY: 1143       N=1

 WID: 14       count: 288      latency: 1 SIZE 0 CAP 2400

  #15 COUNT: 288        AVG_LATENCY: 1067       N=1

 WID: 15       count: 480      latency: 1 SIZE 0 CAP 2400

  #16 COUNT: 480        AVG_LATENCY: 1001       N=1
```

When the throughput is increased to 10000:

```
~/AIR/Release master !10 ?7 > mpirun -np 4 ./AIR YSB 10000

*******************AIR (c) 2020 Uni.lu*******************

AIR INSTANCE AT RANK 1/4 | TP: 10000 | MSG/SEC/RANK: 2 | AGGR_WINDOW: 10000ms
AIR INSTANCE AT RANK 2/4 | TP: 10000 | MSG/SEC/RANK: 2 | AGGR_WINDOW: 10000ms
AIR INSTANCE AT RANK 3/4 | TP: 10000 | MSG/SEC/RANK: 2 | AGGR_WINDOW: 10000ms
AIR INSTANCE AT RANK 4/4 | TP: 10000 | MSG/SEC/RANK: 2 | AGGR_WINDOW: 10000ms
 WID: 0 count: 35000    latency: 4 SIZE 0 CAP 2400

  #1 COUNT: 35000        AVG_LATENCY: 4  N=1

 WID: 1 count: 40000    latency: 2002 SIZE 0 CAP 2400

  #2 COUNT: 40000        AVG_LATENCY: 1003       N=1

 WID: 2 count: 40000    latency: 4 SIZE 0 CAP 2400

  #3 COUNT: 40000        AVG_LATENCY: 670        N=1

 WID: 3 count: 45000    latency: 1002 SIZE 0 CAP 2400

  #4 COUNT: 45000        AVG_LATENCY: 753        N=1

 WID: 4 count: 35000    latency: 2002 SIZE 0 CAP 2400

  #5 COUNT: 35000        AVG_LATENCY: 1002       N=1

 WID: 5 count: 40000    latency: 4 SIZE 0 CAP 2400

  #6 COUNT: 40000        AVG_LATENCY: 836        N=1

 WID: 6 count: 30000    latency: 4 SIZE 0 CAP 2400

  #7 COUNT: 30000        AVG_LATENCY: 717        N=1

 WID: 7 count: 40000    latency: 1002 SIZE 0 CAP 2400

  #8 COUNT: 40000        AVG_LATENCY: 753        N=1

 WID: 8 count: 45000    latency: 4 SIZE 0 CAP 2400

  #9 COUNT: 45000        AVG_LATENCY: 669        N=1

 WID: 9 count: 30000    latency: 2002 SIZE 0 CAP 2400

  #10 COUNT: 30000       AVG_LATENCY: 803        N=1

 WID: 10        count: 35000    latency: 1002 SIZE 0 CAP 2400

  #11 COUNT: 35000       AVG_LATENCY: 821        N=1

 WID: 11        count: 35000    latency: 4 SIZE 0 CAP 2400

  #12 COUNT: 35000       AVG_LATENCY: 753        N=1

 WID: 12        count: 30000    latency: 4 SIZE 0 CAP 2400

  #13 COUNT: 30000       AVG_LATENCY: 695        N=1

 WID: 13        count: 25000    latency: 4 SIZE 0 CAP 2400

  #14 COUNT: 25000       AVG_LATENCY: 646        N=1

 WID: 14        count: 25000    latency: 4 SIZE 0 CAP 2400

  #15 COUNT: 25000       AVG_LATENCY: 603        N=1

 WID: 15        count: 30000    latency: 4 SIZE 0 CAP 2400

  #16 COUNT: 30000       AVG_LATENCY: 565        N=1
```

Throughput(50k) - this was the max:

```
~/AIR/Release master !10 ?7 > mpirun -np 4 ./AIR YSB 50000

*******************AIR (c) 2020 Uni.lu*******************

AIR INSTANCE AT RANK 4/4 | TP: 50000 | MSG/SEC/RANK: 2 | AGGR_WINDOW: 10000ms
AIR INSTANCE AT RANK 2/4 | TP: 50000 | MSG/SEC/RANK: 2 | AGGR_WINDOW: 10000ms
AIR INSTANCE AT RANK 1/4 | TP: 50000 | MSG/SEC/RANK: 2 | AGGR_WINDOW: 10000ms
AIR INSTANCE AT RANK 3/4 | TP: 50000 | MSG/SEC/RANK: 2 | AGGR_WINDOW: 10000ms
 WID: 0 count: 225000   latency: 32 SIZE 0 CAP 2400

  #1 COUNT: 225000       AVG_LATENCY: 32 N=1

 WID: 1 count: 75000    latency: 15 SIZE 0 CAP 2400

  #2 COUNT: 75000        AVG_LATENCY: 23 N=1

 WID: 2 count: 175000   latency: 1006 SIZE 0 CAP 2400

  #3 COUNT: 175000       AVG_LATENCY: 351        N=1

 WID: 3 count: 200000   latency: 16 SIZE 0 CAP 2400

  #4 COUNT: 200000       AVG_LATENCY: 267        N=1

 WID: 4 count: 175000   latency: 16 SIZE 0 CAP 2400

  #5 COUNT: 175000       AVG_LATENCY: 217        N=1

 WID: 5 count: 125000   latency: 3007 SIZE 0 CAP 2400

  #6 COUNT: 125000       AVG_LATENCY: 682        N=1

 WID: 6 count: 125000   latency: 1007 SIZE 0 CAP 2400

  #7 COUNT: 125000       AVG_LATENCY: 728        N=1

 WID: 7 count: 175000   latency: 25 SIZE 0 CAP 2400

  #8 COUNT: 175000       AVG_LATENCY: 640        N=1

 WID: 8 count: 125000   latency: 1007 SIZE 0 CAP 2400

  #9 COUNT: 125000       AVG_LATENCY: 681        N=1

 WID: 9 count: 200000   latency: 16 SIZE 0 CAP 2400

  #10 COUNT: 200000      AVG_LATENCY: 614        N=1

 WID: 10        count: 150000   latency: 2011 SIZE 0 CAP 2400

  #11 COUNT: 150000      AVG_LATENCY: 741        N=1

 WID: 11        count: 100000   latency: 1007 SIZE 0 CAP 2400

  #12 COUNT: 100000      AVG_LATENCY: 763        N=1

 WID: 12        count: 150000   latency: 1007 SIZE 0 CAP 2400

  #13 COUNT: 150000      AVG_LATENCY: 782        N=1

 WID: 13        count: 100000   latency: 16 SIZE 0 CAP 2400

  #14 COUNT: 100000      AVG_LATENCY: 727        N=1

 WID: 14        count: 125000   latency: 16 SIZE 0 CAP 2400

  #15 COUNT: 125000      AVG_LATENCY: 680        N=1

 WID: 15        count: 200000   latency: 27 SIZE 0 CAP 2400

  #16 COUNT: 200000      AVG_LATENCY: 639        N=1
```

Experimenting with different dataflows. The max throughput supported by a single dataflow was 10k. So I've decided to benchmark different no of dataflows using 10k as throughput:

1 dataflow:

```
[~/AIR/Release master !10 ?7 > mpirun -np 1 ./AIR YSB 10000

******************AIR (c) 2020 Uni.lu******************

AIR INSTANCE AT RANK 1/1 | TP: 10000 | MSG/SEC/RANK: 2 | AGGR_WINDOW: 10000ms
 WID: 0 count: 25000    latency: 12 SIZE 0 CAP 2400

 #1 COUNT: 25000        AVG_LATENCY: 12 N=1

 WID: 1 count: 10000    latency: 8005 SIZE 0 CAP 2400

 #2 COUNT: 10000        AVG_LATENCY: 4008       N=1

 WID: 2 count: 35000    latency: 11 SIZE 0 CAP 2400

 #3 COUNT: 35000        AVG_LATENCY: 2676       N=1

 WID: 3 count: 25000    latency: 19 SIZE 0 CAP 2400

 #4 COUNT: 25000        AVG_LATENCY: 2011       N=1

 WID: 4 count: 50000    latency: 11 SIZE 0 CAP 2400

 #5 COUNT: 50000        AVG_LATENCY: 1611       N=1

 WID: 5 count: 35000    latency: 12 SIZE 0 CAP 2400

 #6 COUNT: 35000        AVG_LATENCY: 1345       N=1

 WID: 6 count: 30000    latency: 12 SIZE 0 CAP 2400

 #7 COUNT: 30000        AVG_LATENCY: 1154       N=1

 WID: 7 count: 50000    latency: 12 SIZE 0 CAP 2400

 #8 COUNT: 50000        AVG_LATENCY: 1011       N=1

 WID: 8 count: 25000    latency: 12 SIZE 0 CAP 2400

 #9 COUNT: 25000        AVG_LATENCY: 900        N=1

 WID: 9 count: 35000    latency: 1005 SIZE 0 CAP 2400

 #10 COUNT: 35000       AVG_LATENCY: 911        N=1
```

5 dataflows:

```
[~/AIR/Release master !10 ?7 > mpirun -np 5 ./AIR YSB 10000

******************AIR (c) 2020 Uni.lu******************

AIR INSTANCE AT RANK 1/5 | TP: 10000 | MSG/SEC/RANK: 2 | AGGR_WINDOW: 10000ms
AIR INSTANCE AT RANK 2/5 | TP: 10000 | MSG/SEC/RANK: 2 | AGGR_WINDOW: 10000ms
AIR INSTANCE AT RANK 3/5 | TP: 10000 | MSG/SEC/RANK: 2 | AGGR_WINDOW: 10000ms
AIR INSTANCE AT RANK 4/5 | TP: 10000 | MSG/SEC/RANK: 2 | AGGR_WINDOW: 10000ms
AIR INSTANCE AT RANK 5/5 | TP: 10000 | MSG/SEC/RANK: 2 | AGGR_WINDOW: 10000ms
 WID: 0 count: 25000    latency: 5 SIZE 0 CAP 2400

  #1 COUNT: 25000       AVG_LATENCY: 5   N=1

 WID: 1 count: 30000    latency: 3003 SIZE 0 CAP 2400

  #2 COUNT: 30000       AVG_LATENCY: 1504       N=1

 WID: 2 count: 30000    latency: 3003 SIZE 0 CAP 2400

  #3 COUNT: 30000       AVG_LATENCY: 2003       N=1

 WID: 3 count: 15000    latency: 4003 SIZE 0 CAP 2400

  #4 COUNT: 15000       AVG_LATENCY: 2503       N=1

 WID: 4 count: 15000    latency: 4003 SIZE 0 CAP 2400

  #5 COUNT: 15000       AVG_LATENCY: 2803       N=1

 WID: 5 count: 25000    latency: 3003 SIZE 0 CAP 2400

  #6 COUNT: 25000       AVG_LATENCY: 2836       N=1

 WID: 6 count: 55000    latency: 5 SIZE 0 CAP 2400

  #7 COUNT: 55000       AVG_LATENCY: 2432       N=1

 WID: 7 count: 25000    latency: 2003 SIZE 0 CAP 2400

  #8 COUNT: 25000       AVG_LATENCY: 2378       N=1

 WID: 8 count: 30000    latency: 5 SIZE 0 CAP 2400

  #9 COUNT: 30000       AVG_LATENCY: 2114       N=1

 WID: 9 count: 20000    latency: 3003 SIZE 0 CAP 2400

  #10 COUNT: 20000      AVG_LATENCY: 2203       N=1
```

10 dataflows:

```
[~/AIR/Release master !10 ?7 > mpirun —np 10 ./AIR YSB 10000

******************AIR (c) 2020 Uni.lu******************

AIR INSTANCE AT RANK 8/10  | TP: 10000 | MSG/SEC/RANK: 2 | AGGR_WINDOW: 10000ms
AIR INSTANCE AT RANK 9/10  | TP: 10000 | MSG/SEC/RANK: 2 | AGGR_WINDOW: 10000ms
AIR INSTANCE AT RANK 1/10  | TP: 10000 | MSG/SEC/RANK: 2 | AGGR_WINDOW: 10000ms
AIR INSTANCE AT RANK 2/10  | TP: 10000 | MSG/SEC/RANK: 2 | AGGR_WINDOW: 10000ms
AIR INSTANCE AT RANK 3/10  | TP: 10000 | MSG/SEC/RANK: 2 | AGGR_WINDOW: 10000ms
AIR INSTANCE AT RANK 4/10  | TP: 10000 | MSG/SEC/RANK: 2 | AGGR_WINDOW: 10000ms
AIR INSTANCE AT RANK 5/10  | TP: 10000 | MSG/SEC/RANK: 2 | AGGR_WINDOW: 10000ms
AIR INSTANCE AT RANK 6/10  | TP: 10000 | MSG/SEC/RANK: 2 | AGGR_WINDOW: 10000ms
AIR INSTANCE AT RANK 7/10  | TP: 10000 | MSG/SEC/RANK: 2 | AGGR_WINDOW: 10000ms
AIR INSTANCE AT RANK 10/10 | TP: 10000 | MSG/SEC/RANK: 2 | AGGR_WINDOW: 10000ms
 WID: 0 count: 20000    latency: 1014 SIZE 0 CAP 2400

  #1 COUNT: 20000        AVG_LATENCY: 1014       N=1

 WID: 1 count: 25000    latency: 2032 SIZE 0 CAP 2400

  #2 COUNT: 25000        AVG_LATENCY: 1523       N=1

 WID: 2 count: 30000    latency: 9 SIZE 0 CAP 2400

  #3 COUNT: 30000        AVG_LATENCY: 1018       N=1

 WID: 3 count: 30000    latency: 37 SIZE 0 CAP 2400

  #4 COUNT: 30000        AVG_LATENCY: 773        N=1

 WID: 4 count: 25000    latency: 12 SIZE 0 CAP 2400

  #5 COUNT: 25000        AVG_LATENCY: 620        N=1

 WID: 5 count: 30000    latency: 8 SIZE 0 CAP 2400

  #6 COUNT: 30000        AVG_LATENCY: 518        N=1

 WID: 6 count: 40000    latency: 6 SIZE 0 CAP 2400

  #7 COUNT: 40000        AVG_LATENCY: 445        N=1

 WID: 7 count: 35000    latency: 8 SIZE 0 CAP 2400

  #8 COUNT: 35000        AVG_LATENCY: 390        N=1

 WID: 8 count: 40000    latency: 1066 SIZE 0 CAP 2400

  #9 COUNT: 40000        AVG_LATENCY: 465        N=1

 WID: 9 count: 35000    latency: 8 SIZE 0 CAP 2400

  #10 COUNT: 35000       AVG_LATENCY: 420        N=1

 WID: 10       count: 25000    latency: 1014 SIZE 0 CAP 2400
```