Assignment 3

Bella Fu jf70

Task 1

Java code included in the file "task1-java-code"

Result

2013-01-08	6086401.280000469
2013-01-11	7309754.250000235
2013-01-19	6861240.609999935
2013-01-22	6859188.610000252
2013-01-01	6155808.070000131
2013-01-09	6340631.130000349
2013-01-12	6794477.000000252
2013-01-23	7141795.610000398
2013-02-02	7480307.670000353
2013-02-13	7541222.270000333
2013-02-24	6734030.250000294
2013-03-03	7164542.680000728
2013-03-14	8332600.810000411
2013-03-25	6242181.430000134
2013-04-04	7892240.750000184
2013-04-15	6829723.230000164
2013-04-26	8162598.720000312
2013-05-05	7308989.850000491
2013-05-16	8141080.420000477
2013-05-27	5203819.5800001975
2013-05-30	7773610.690000756
2013-06-06	7990652.720000534
2013-06-17	6601498.390000233
2013-06-20	7727357.320000615
2013-06-28	7823081.670000429
2013-07-07	5353565.250000039
2013-07-10	7053278.450000221
2013-07-18	7727945.430000308
2013-07-21	6250172.660000221
2013-07-29	6252667.090000473
2013-08-08	6512086.030000453
2013-08-11	3069384.85000003

2013-08-19	6164400.170000227
2013-08-22	7184582.000000339
2013-09-01	5619155.650000318
2013-09-09	6881140.260000429
2013-09-12	7916396.880000437
2013-09-23	6616708.470000272
2013-10-07	6508816.200000132
2013-10-10	7928990.740000219
2013-10-18	8151395.510000445
2013-10-21	6678712.330000285
2013-10-29	6912935.060000494
2013-11-08	8284354.440000461
2013-11-11	6455445.770000158
2013-11-19	7072677.4000001345
2013-11-22	8034678.870000388
2013-12-01	6253203.70000069
2013-12-09	6924651.610000448
2013-12-12	8256267.260000355
2013-12-23	5657237.3999999575
2013-01-02	5597586.19000029
2013-01-13	6337345.520000491
2013-01-24	7394285.4800002165
2013-01-03	6063858.900000115
2013-01-14	6247518.030000225
2013-01-25	7591697.2200000845
2013-02-04	6465668.770000438
2013-02-15	7872468.330000689
2013-02-26	6990771.510000217
2013-03-05	7033434.130000625
2013-03-16	8052750.150000352
2013-03-27	7014765.390000323
2013-03-30	6864070.360000182
2013-04-06	7761401.130000321
2013-04-17	7426764.830000225
2013-04-20	7889826.780000055
2013-04-28	7073965.570000082
2013-05-07	7272448.350000596
2013-05-10	8200131.330000756
2013-05-18	7942858.820000603
2013-05-21	7657783.0200002855
2013-05-29	7179023.220000537
2013-06-08	6933761.410000358
2013-06-11	7115482.590000324
2013-06-19	7476068.830000555

2013-06-22	6790382.830000521
2013-07-01	6251501.600000258
2013-07-09	6757271.850000111
2013-07-12	7513843.17000032
2013-07-23	6953065.7400000375
2013-08-02	3530800.630000363
2013-08-13	6662944.510000315
2013-08-24	6513604.34000038
2013-09-03	6491695.110000047
2013-09-14	7305007.4700001385
2013-09-25	7240695.170000383
2013-10-01	6836625.290000091
2013-10-09	7426570.660000361
2013-10-12	7753415.9200002635
2013-10-23	7525775.590000441
2013-11-02	7668629.24000078
2013-11-13	7411807.630000057
2013-11-24	7030430.580000366
2013-12-03	7134861.170000482
2013-12-14	7281088.110000375
2013-12-25	3465875.3099999367
2013-01-04	6436856.210000185
2013-01-15	6904804.110000155
2013-01-26	7380616.140000331
2013-02-05	6728588.430000651
2013-02-16	7286060.730000584
2013-02-27	7278291.400000172
2013-03-06	7299922.8700008495
2013-03-17	7213525.390000344
2013-03-20	7273280.520000249
2013-03-28	7653224.730000114
2013-03-31	6430424.840000082
2013-04-07	7094588.600000548
2013-04-10	7324328.76000038
2013-04-18	7921211.370000094
2013-04-21	7110201.060000192
2013-04-29	6769091.820000659
2013-05-08	7715748.710000588
2013-05-11	7465048.090000532
2013-05-19	7409152.300000518
2013-05-22	7754988.380000457
2013-06-01	7481803.2700008955
2013-06-09	6888553.510000441
2013-06-12	7494382.580000324

2013-06-23	6570007.290000441
2013-07-02	6362945.3600005
2013-07-13	6587855.010000025
2013-07-24	7237750.5700006075
2013-08-03	3182263.760000206
2013-08-14	7552538.050000365
2013-08-25	6193454.910000368
2013-09-04	6761444.710000122
2013-09-15	7059224.600000193
2013-09-26	7627537.7900004275
2013-10-02	7193015.640000156
2013-10-13	7190288.140000477
2013-10-24	7993033.550000352
2013-11-03	7192686.050000614
2013-11-14	7789295.340000308
2013-11-25	6486676.800000244
2013-12-04	7452656.220000424
2013-12-15	6771079.610000316
2013-12-26	4601206.359999932
2013-01-05	6284187.200000157
2013-01-16	7099953.430000305
2013-01-27	6691499.990000044
2013-01-30	6819942.91000031
2013-02-06	7124564.7400005525
2013-02-17	6845261.53000049
2013-02-20	7137929.2200006675
2013-02-28	7666926.94000045
2013-03-07	7653469.5600003535
2013-03-10	6897237.450000272
2013-03-18	6668808.510000594
2013-03-21	7368950.290000269
2013-03-29	7433572.3300002795
2013-04-08	6456388.9500002265
2013-04-11	7942809.0600001365
2013-04-19	8108821.010000268
2013-04-22	6715859.190000024
2013-05-01	7515615.840000708
2013-05-09	8170779.440000853
2013-05-12	6934527.770000359
2013-05-23	7831608.79000004
2013-06-02	6883789.75000066
2013-06-13	7822149.280000436
2013-06-24	6540394.860000241
2013-07-03	6418768.160000504

2013-07-14	6487142.680000114
2013-07-25	7461431.370000593
2013-08-04	3073608.61000036
2013-08-15	7309690.670000275
2013-08-26	5948970.699999955
2013-09-05	6849652.470000072
2013-09-16	6851916.240000133
2013-09-27	7851086.62000021
2013-09-30	6582569.770000837
2013-10-03	7691451.380000134
2013-10-14	6096785.550000316
2013-10-25	8123852.110000397
2013-11-04	6920247.070000162
2013-11-15	7970699.000000397
2013-11-26	6882872.590000152
2013-12-05	7814472.9600006575
2013-12-16	7031713.510000603
2013-12-27	5608426.559999982
2013-12-30	5956747.680000558
2013-01-06	5713649.169999975
2013-01-17	7369913.010000426
2013-01-20	6308348.880000001
2013-01-28	6318050.780000441
2013-01-31	7611473.570000514
2013-02-07	7810428.980000311
2013-02-10	6491805.029999984
2013-02-18	5719261.550000275
2013-02-21	7594376.810000539
2013-03-08	7736197.170000756
2013-03-11	6603637.270000187
2013-03-19	6856158.650000527
2013-03-22	8026893.840000632
2013-04-01	6263647.910000198
2013-04-09	6927704.38000021
2013-04-12	8256000.260000265
2013-04-23	7268278.210000054
2013-05-02	7915287.5200008415
2013-05-13	6921766.400000614
2013-05-24	7725439.820000494
2013-06-03	6793589.040000501
2013-06-14	7748373.750000326
2013-06-25	7055273.1600002255
2013-07-04	4575369.319999892
2013-07-15	6608855.280000186

2013-07-26	7313389.780000516
2013-08-05	6657059.8900002055
2013-08-16	7137331.870000365
2013-08-27	6234402.300000014
2013-08-30	7002455.110000105
2013-09-06	7543070.940000139
2013-09-17	7184226.6800001515
2013-09-20	8145315.140000101
2013-09-28	7703734.320000281
2013-10-04	7939072.64000022
2013-10-15	6582799.590000392
2013-10-26	7677650.65000019
2013-11-05	6896607.9800002165
2013-11-16	7721349.440000324
2013-11-27	6483025.480000399
2013-11-30	6057614.490000115
2013-12-06	8341385.070000729
2013-12-17	7176713.170000495
2013-12-20	8183023.320000826
2013-12-28	5782093.000000052
2013-12-31	6272547.980000513
2013-01-07	5709195.480000225
2013-01-10	6923892.1800003275
2013-01-18	7753118.5399999395
2013-01-21	5409171.510000268
2013-01-29	6661626.40000036
2013-02-08	5642667.740000279
2013-02-11	6600228.500000704
2013-02-19	6530409.830000613
2013-02-22	7967313.820000698
2013-03-01	8139945.370000561
2013-03-09	7843434.580000635
2013-03-12	7151880.2700003255
2013-03-23	7645868.220000511
2013-04-02	7035845.220000216
2013-04-13	7625592.250000173
2013-04-24	7381586.470000276
2013-05-03	8211510.980000812
2013-05-14	7265246.020000626
2013-05-25	6327689.1600002125
2013-06-04	7237697.450000622
2013-06-15	7076438.010000271
2013-06-26	7454240.260000361
2013-07-05	4803163.460000097

```
2013-07-16    6914721.620000269
2013-07-27    6562832.940000306
2013-07-30    6500185.240000375
2013-08-06    6576576.120000195
2013-08-17    6524048.620000127
2013-08-20    6468051.680000117
2013-08-28    6666909.240000105
2013-08-31    6239629.399999892
2013-09-07    7410219.290000074
2013-09-10    7333647.010000246
2013-09-18    7477339.480000224
2013-09-21    7786891.859999999
2013-09-29    7122464.470000353
2013-10-05    7557050.500000589
2013-10-16    7309766.840000507
2013-10-27    7194001.180000275
2013-10-30    7278987.040000562
2013-11-06    7394702.590000286
2013-11-17    7033678.510000097
2013-11-20    7469779.110000168
2013-11-28    4748377.520000076
2013-12-07    7961737.7300004605
2013-12-10    7101578.31000045
2013-12-18    7730028.710000701
2013-12-21    7176444.250000439
2013-12-29    5918205.399999988
```

Hadoop

```
[hadoop@ip-172-31-15-253 ~]$ hadoop jar task1.jar -r 8
s3://chrisjermainebucket/comp330_A3/ out
22/10/24 17:28:25 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-15-
253.us-east-2.compute.internal/172.31.15.253:8032
22/10/24 17:28:25 INFO client.AHSProxy: Connecting to Application History server at ip-172-31-
15-253.us-east-2.compute.internal/172.31.15.253:10200
22/10/24 17:28:25 WARN mapreduce.JobResourceUploader: Hadoop command-line option
parsing not performed. Implement the Tool interface and execute your application with
ToolRunner to remedy this.
22/10/24 17:28:27 INFO input.FileInputFormat: Total input files to process : 12
22/10/24 17:28:27 INFO lzo.GPLNativeCodeLoader: Loaded native gpl library
22/10/24 17:28:27 INFO lzo.LzoCodec: Successfully loaded & initialized native-lzo library
[hadoop-lzo rev 049362b7cf53ff5f739d6b1532457f2c6cd495e8]
```

22/10/24 17:28:27 INFO mapreduce.JobSubmitter: number of splits:298

22/10/24 17:28:27 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1666631367881_0004

22/10/24 17:28:27 INFO conf.Configuration: resource-types.xml not found

22/10/24 17:28:27 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.

22/10/24 17:28:27 INFO resource.ResourceUtils: Adding resource type - name = memory-mb, units = Mi, type = COUNTABLE

22/10/24 17:28:27 INFO resource.ResourceUtils: Adding resource type - name = vcores, units = , type = COUNTABLE

22/10/24 17:28:27 INFO impl.YarnClientImpl: Submitted application application_1666631367881_0004

22/10/24 17:28:27 INFO mapreduce.Job: The url to track the job: http://ip-172-31-15-253.us-east-2.compute.internal:20888/proxy/application_1666631367881_0004/

22/10/24 17:28:27 INFO mapreduce.Job: Running job: job_1666631367881_0004

22/10/24 17:28:33 INFO mapreduce.Job: Job job_1666631367881_0004 running in uber mode : false

22/10/24 17:28:33 INFO mapreduce.Job: map 0% reduce 0%

22/10/24 17:28:45 INFO mapreduce.Job: map 1% reduce 0%

22/10/24 17:28:53 INFO mapreduce.Job: map 2% reduce 0%

22/10/24 17:28:55 INFO mapreduce.Job: map 3% reduce 0%

22/10/24 17:29:10 INFO mapreduce.Job: map 4% reduce 0%

22/10/24 17:29:13 INFO mapreduce.Job: map 5% reduce 0%

22/10/24 17:29:25 INFO mapreduce.Job: map 6% reduce 0%

22/10/24 17:29:30 INFO mapreduce.Job: map 7% reduce 0%

22/10/24 17:29:36 INFO mapreduce.Job: map 8% reduce 0%

22/10/24 17:29:47 INFO mapreduce.Job: map 9% reduce 0%

22/10/24 17:29:51 INFO mapreduce.Job: map 10% reduce 0%

22/10/24 17:29:57 INFO mapreduce.Job: map 11% reduce 0%

22/10/24 17:30:09 INFO mapreduce.Job: map 12% reduce 0%

22/10/24 17:30:10 INFO mapreduce.Job: map 13% reduce 0%

22/10/24 17:30:27 INFO mapreduce.Job: map 14% reduce 0%

22/10/24 17:30:30 INFO mapreduce.Job: map 15% reduce 0%

22/10/24 17:30:40 INFO mapreduce.Job: map 16% reduce 0%

22/10/24 17:30:47 INFO mapreduce.Job: map 17% reduce 0%

22/10/24 17:30:50 INFO mapreduce.Job: map 18% reduce 0%

22/10/24 17:31:03 INFO mapreduce.Job: map 19% reduce 0%

22/10/24 17:31:07 INFO mapreduce.Job: map 20% reduce 0%

22/10/24 17:31:13 INFO mapreduce.Job: map 21% reduce 0%

22/10/24 17:31:23 INFO mapreduce.Job: map 22% reduce 0%

22/10/24 17:31:27 INFO mapreduce.Job: map 23% reduce 0%

22/10/24 17:31:40 INFO mapreduce.Job: map 24% reduce 0%

22/10/24 17:31:44 INFO mapreduce.Job: map 25% reduce 0%

22/10/24 17:31:46 INFO mapreduce.Job: map 26% reduce 0%

22/10/24 17:32:01 INFO mapreduce.Job: map 27% reduce 0%

```
22/10/24 17:32:04 INFO mapreduce.Job:  map 28% reduce 0%
22/10/24 17:32:12 INFO mapreduce.Job:  map 29% reduce 0%
22/10/24 17:32:20 INFO mapreduce.Job:  map 30% reduce 0%
22/10/24 17:32:24 INFO mapreduce.Job:  map 31% reduce 0%
22/10/24 17:32:37 INFO mapreduce.Job:  map 32% reduce 0%
22/10/24 17:32:41 INFO mapreduce.Job:  map 33% reduce 0%
22/10/24 17:32:48 INFO mapreduce.Job:  map 34% reduce 0%
22/10/24 17:32:57 INFO mapreduce.Job:  map 35% reduce 0%
22/10/24 17:33:02 INFO mapreduce.Job:  map 36% reduce 0%
22/10/24 17:33:13 INFO mapreduce.Job:  map 37% reduce 2%
22/10/24 17:33:20 INFO mapreduce.Job:  map 38% reduce 2%
22/10/24 17:33:30 INFO mapreduce.Job:  map 39% reduce 2%
22/10/24 17:33:35 INFO mapreduce.Job:  map 40% reduce 2%
22/10/24 17:33:44 INFO mapreduce.Job:  map 41% reduce 2%
22/10/24 17:33:52 INFO mapreduce.Job:  map 42% reduce 2%
22/10/24 17:34:03 INFO mapreduce.Job:  map 43% reduce 2%
22/10/24 17:34:06 INFO mapreduce.Job:  map 44% reduce 2%
22/10/24 17:34:15 INFO mapreduce.Job:  map 45% reduce 2%
22/10/24 17:34:24 INFO mapreduce.Job:  map 46% reduce 2%
22/10/24 17:34:35 INFO mapreduce.Job:  map 47% reduce 2%
22/10/24 17:34:37 INFO mapreduce.Job:  map 48% reduce 2%
22/10/24 17:34:46 INFO mapreduce.Job:  map 49% reduce 2%
22/10/24 17:34:54 INFO mapreduce.Job:  map 50% reduce 2%
22/10/24 17:35:04 INFO mapreduce.Job:  map 51% reduce 2%
22/10/24 17:35:09 INFO mapreduce.Job:  map 52% reduce 2%
22/10/24 17:35:16 INFO mapreduce.Job:  map 53% reduce 2%
22/10/24 17:35:25 INFO mapreduce.Job:  map 54% reduce 2%
22/10/24 17:35:35 INFO mapreduce.Job:  map 55% reduce 2%
22/10/24 17:35:41 INFO mapreduce.Job:  map 56% reduce 2%
22/10/24 17:35:47 INFO mapreduce.Job:  map 57% reduce 2%
22/10/24 17:35:59 INFO mapreduce.Job:  map 58% reduce 2%
22/10/24 17:36:07 INFO mapreduce.Job:  map 59% reduce 2%
22/10/24 17:36:13 INFO mapreduce.Job:  map 60% reduce 2%
22/10/24 17:36:20 INFO mapreduce.Job:  map 60% reduce 3%
22/10/24 17:36:21 INFO mapreduce.Job:  map 61% reduce 3%
22/10/24 17:36:31 INFO mapreduce.Job:  map 62% reduce 3%
22/10/24 17:36:36 INFO mapreduce.Job:  map 63% reduce 3%
22/10/24 17:36:45 INFO mapreduce.Job:  map 64% reduce 3%
22/10/24 17:36:52 INFO mapreduce.Job:  map 65% reduce 3%
22/10/24 17:37:03 INFO mapreduce.Job:  map 66% reduce 3%
22/10/24 17:37:07 INFO mapreduce.Job:  map 67% reduce 3%
22/10/24 17:37:15 INFO mapreduce.Job:  map 68% reduce 3%
22/10/24 17:37:24 INFO mapreduce.Job:  map 69% reduce 3%
22/10/24 17:37:34 INFO mapreduce.Job:  map 70% reduce 3%
```

22/10/24 17:37:38 INFO mapreduce.Job: map 71% reduce 3%
22/10/24 17:37:47 INFO mapreduce.Job: map 72% reduce 3%
22/10/24 17:37:56 INFO mapreduce.Job: map 73% reduce 3%
22/10/24 17:38:05 INFO mapreduce.Job: map 74% reduce 3%
22/10/24 17:38:10 INFO mapreduce.Job: map 75% reduce 3%
22/10/24 17:38:16 INFO mapreduce.Job: map 76% reduce 3%
22/10/24 17:38:26 INFO mapreduce.Job: map 77% reduce 3%
22/10/24 17:38:31 INFO mapreduce.Job: map 78% reduce 3%
22/10/24 17:38:39 INFO mapreduce.Job: map 79% reduce 3%
22/10/24 17:38:47 INFO mapreduce.Job: map 80% reduce 3%
22/10/24 17:38:55 INFO mapreduce.Job: map 81% reduce 3%
22/10/24 17:39:02 INFO mapreduce.Job: map 82% reduce 3%
22/10/24 17:39:09 INFO mapreduce.Job: map 83% reduce 3%
22/10/24 17:39:19 INFO mapreduce.Job: map 84% reduce 3%
22/10/24 17:39:26 INFO mapreduce.Job: map 85% reduce 4%
22/10/24 17:39:33 INFO mapreduce.Job: map 86% reduce 4%
22/10/24 17:39:41 INFO mapreduce.Job: map 87% reduce 4%
22/10/24 17:39:49 INFO mapreduce.Job: map 88% reduce 4%
22/10/24 17:39:57 INFO mapreduce.Job: map 89% reduce 4%
22/10/24 17:40:03 INFO mapreduce.Job: map 90% reduce 4%
22/10/24 17:40:12 INFO mapreduce.Job: map 91% reduce 4%
22/10/24 17:40:21 INFO mapreduce.Job: map 92% reduce 4%
22/10/24 17:40:29 INFO mapreduce.Job: map 93% reduce 4%
22/10/24 17:40:35 INFO mapreduce.Job: map 94% reduce 4%
22/10/24 17:40:44 INFO mapreduce.Job: map 95% reduce 4%
22/10/24 17:40:52 INFO mapreduce.Job: map 96% reduce 4%
22/10/24 17:40:59 INFO mapreduce.Job: map 97% reduce 4%
22/10/24 17:41:05 INFO mapreduce.Job: map 98% reduce 4%
22/10/24 17:41:13 INFO mapreduce.Job: map 99% reduce 4%
22/10/24 17:41:19 INFO mapreduce.Job: map 100% reduce 4%
22/10/24 17:41:22 INFO mapreduce.Job: map 100% reduce 13%
22/10/24 17:41:23 INFO mapreduce.Job: map 100% reduce 25%
22/10/24 17:41:25 INFO mapreduce.Job: map 100% reduce 38%
22/10/24 17:41:26 INFO mapreduce.Job: map 100% reduce 50%
22/10/24 17:41:27 INFO mapreduce.Job: map 100% reduce 63%
22/10/24 17:41:28 INFO mapreduce.Job: map 100% reduce 75%
22/10/24 17:41:30 INFO mapreduce.Job: map 100% reduce 100%
22/10/24 17:41:31 INFO mapreduce.Job: Job job_1666631367881_0004 completed successfully
22/10/24 17:41:32 INFO mapreduce.Job: Counters: 55
        File System Counters
            FILE: Number of bytes read=48973
            FILE: Number of bytes written=67968879
            FILE: Number of read operations=0
            FILE: Number of large read operations=0

FILE: Number of write operations=0
    HDFS: Number of bytes read=34643
    HDFS: Number of bytes written=10590
    HDFS: Number of read operations=620
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=16
    S3: Number of bytes read=19750006390
    S3: Number of bytes written=0
    S3: Number of read operations=0
    S3: Number of large read operations=0
    S3: Number of write operations=0
Job Counters
    Killed map tasks=1
    Launched map tasks=298
    Launched reduce tasks=8
    Data-local map tasks=298
    Total time spent by all maps in occupied slots (ms)=316582080
    Total time spent by all reduces in occupied slots (ms)=100695552
    Total time spent by all map tasks (ms)=3297730
    Total time spent by all reduce tasks (ms)=524456
    Total vcore-milliseconds taken by all map tasks=3297730
    Total vcore-milliseconds taken by all reduce tasks=524456
    Total megabyte-milliseconds taken by all map tasks=10130626560
    Total megabyte-milliseconds taken by all reduce tasks=3222257664
Map-Reduce Framework
    Map input records=173179771
    Map output records=173179759
    Map output bytes=3290415421
    Map output materialized bytes=113629
    Input split bytes=34643
    Combine input records=173179759
    Combine output records=4509
    Reduce input groups=365
    Reduce shuffle bytes=113629
    Reduce input records=4509
    Reduce output records=365
    Spilled Records=9018
    Shuffled Maps =2384
    Failed Shuffles=0
    Merged Map outputs=2384
    GC time elapsed (ms)=143327
    CPU time spent (ms)=3249630
    Physical memory (bytes) snapshot=299237203968
    Virtual memory (bytes) snapshot=1450167406592

```
        Total committed heap usage (bytes)=289102888960
    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=19750006390
    File Output Format Counters
        Bytes Written=10590
[hadoop@ip-172-31-15-253 ~]$ hadoop fs -ls
Found 1 items
drwxr-xr-x   - hadoop hdfsadmingroup        0 2022-10-24 17:41 out
[hadoop@ip-172-31-15-253 ~]$ hadoop fs -ls out
Found 9 items
-rw-r--r--   1 hadoop hdfsadmingroup        0 2022-10-24 17:41 out/_SUCCESS
-rw-r--r--   1 hadoop hdfsadmingroup     1422 2022-10-24 17:41 out/part-r-00000
-rw-r--r--   1 hadoop hdfsadmingroup     1364 2022-10-24 17:41 out/part-r-00001
-rw-r--r--   1 hadoop hdfsadmingroup     1299 2022-10-24 17:41 out/part-r-00002
-rw-r--r--   1 hadoop hdfsadmingroup     1250 2022-10-24 17:41 out/part-r-00003
-rw-r--r--   1 hadoop hdfsadmingroup     1219 2022-10-24 17:41 out/part-r-00004
-rw-r--r--   1 hadoop hdfsadmingroup     1276 2022-10-24 17:41 out/part-r-00005
-rw-r--r--   1 hadoop hdfsadmingroup     1336 2022-10-24 17:41 out/part-r-00006
-rw-r--r--   1 hadoop hdfsadmingroup     1424 2022-10-24 17:41 out/part-r-00007
```

Task 2

Java code included in the file "task2-java-code"

Result

```
1EDF99EE9DAC182027330EF48828B54A   242656.10000000003
D85749E8852FCC66A990E40605607B2F   246370.08999999997
E4F99C9ABE9861F18BCD38BC63D007A9   563445.9400000003
CFCD208495D565EF66E7DFF9F98764DA   615174.6499999993
```

664927CDE376A32789BA48BF55DFB7E3   728594.3299999998

Hadoop

22/10/25 01:07:36 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-40-62.us-east-2.compute.internal/172.31.40.62:8032
22/10/25 01:07:36 INFO client.AHSProxy: Connecting to Application History server at ip-172-31-40-62.us-east-2.compute.internal/172.31.40.62:10200
22/10/25 01:07:36 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
22/10/25 01:07:38 INFO input.FileInputFormat: Total input files to process : 12
22/10/25 01:07:38 INFO lzo.GPLNativeCodeLoader: Loaded native gpl library
22/10/25 01:07:38 INFO lzo.LzoCodec: Successfully loaded & initialized native-lzo library [hadoop-lzo rev 049362b7cf53ff5f739d6b1532457f2c6cd495e8]
22/10/25 01:07:38 INFO mapreduce.JobSubmitter: number of splits:298
22/10/25 01:07:38 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1666646922232_0020
22/10/25 01:07:38 INFO conf.Configuration: resource-types.xml not found
22/10/25 01:07:38 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
22/10/25 01:07:38 INFO resource.ResourceUtils: Adding resource type - name = memory-mb, units = Mi, type = COUNTABLE
22/10/25 01:07:38 INFO resource.ResourceUtils: Adding resource type - name = vcores, units = , type = COUNTABLE
22/10/25 01:07:38 INFO impl.YarnClientImpl: Submitted application application_1666646922232_0020
22/10/25 01:07:38 INFO mapreduce.Job: The url to track the job: http://ip-172-31-40-62.us-east-2.compute.internal:20888/proxy/application_1666646922232_0020/
22/10/25 01:07:38 INFO mapreduce.Job: Running job: job_1666646922232_0020
22/10/25 01:07:43 INFO mapreduce.Job: Job job_1666646922232_0020 running in uber mode : false
22/10/25 01:07:43 INFO mapreduce.Job: map 0% reduce 0%
22/10/25 01:07:56 INFO mapreduce.Job: map 1% reduce 0%
22/10/25 01:08:05 INFO mapreduce.Job: map 2% reduce 0%
22/10/25 01:08:08 INFO mapreduce.Job: map 3% reduce 0%
22/10/25 01:08:26 INFO mapreduce.Job: map 4% reduce 0%
22/10/25 01:08:27 INFO mapreduce.Job: map 5% reduce 0%
22/10/25 01:08:41 INFO mapreduce.Job: map 6% reduce 0%
22/10/25 01:08:46 INFO mapreduce.Job: map 7% reduce 0%
22/10/25 01:08:52 INFO mapreduce.Job: map 8% reduce 0%
22/10/25 01:09:04 INFO mapreduce.Job: map 9% reduce 0%
22/10/25 01:09:07 INFO mapreduce.Job: map 10% reduce 0%
22/10/25 01:09:16 INFO mapreduce.Job: map 11% reduce 0%

22/10/25 01:09:27 INFO mapreduce.Job: map 13% reduce 0%
22/10/25 01:09:47 INFO mapreduce.Job: map 14% reduce 0%
22/10/25 01:09:48 INFO mapreduce.Job: map 15% reduce 0%
22/10/25 01:09:59 INFO mapreduce.Job: map 16% reduce 0%
22/10/25 01:10:06 INFO mapreduce.Job: map 17% reduce 0%
22/10/25 01:10:11 INFO mapreduce.Job: map 18% reduce 0%
22/10/25 01:10:21 INFO mapreduce.Job: map 19% reduce 0%
22/10/25 01:10:28 INFO mapreduce.Job: map 20% reduce 0%
22/10/25 01:10:32 INFO mapreduce.Job: map 21% reduce 0%
22/10/25 01:10:46 INFO mapreduce.Job: map 22% reduce 0%
22/10/25 01:10:47 INFO mapreduce.Job: map 23% reduce 0%
22/10/25 01:11:05 INFO mapreduce.Job: map 24% reduce 0%
22/10/25 01:11:07 INFO mapreduce.Job: map 26% reduce 0%
22/10/25 01:11:24 INFO mapreduce.Job: map 27% reduce 0%
22/10/25 01:11:26 INFO mapreduce.Job: map 28% reduce 0%
22/10/25 01:11:40 INFO mapreduce.Job: map 29% reduce 0%
22/10/25 01:11:45 INFO mapreduce.Job: map 30% reduce 0%
22/10/25 01:11:51 INFO mapreduce.Job: map 31% reduce 0%
22/10/25 01:12:03 INFO mapreduce.Job: map 32% reduce 0%
22/10/25 01:12:04 INFO mapreduce.Job: map 33% reduce 0%
22/10/25 01:12:14 INFO mapreduce.Job: map 34% reduce 0%
22/10/25 01:12:23 INFO mapreduce.Job: map 35% reduce 0%
22/10/25 01:12:24 INFO mapreduce.Job: map 36% reduce 0%
22/10/25 01:12:36 INFO mapreduce.Job: map 37% reduce 0%
22/10/25 01:12:40 INFO mapreduce.Job: map 37% reduce 2%
22/10/25 01:12:46 INFO mapreduce.Job: map 38% reduce 2%
22/10/25 01:12:54 INFO mapreduce.Job: map 39% reduce 2%
22/10/25 01:12:58 INFO mapreduce.Job: map 40% reduce 2%
22/10/25 01:13:09 INFO mapreduce.Job: map 41% reduce 2%
22/10/25 01:13:18 INFO mapreduce.Job: map 42% reduce 2%
22/10/25 01:13:27 INFO mapreduce.Job: map 43% reduce 2%
22/10/25 01:13:33 INFO mapreduce.Job: map 44% reduce 2%
22/10/25 01:13:42 INFO mapreduce.Job: map 45% reduce 2%
22/10/25 01:13:49 INFO mapreduce.Job: map 46% reduce 2%
22/10/25 01:14:00 INFO mapreduce.Job: map 47% reduce 2%
22/10/25 01:14:06 INFO mapreduce.Job: map 48% reduce 2%
22/10/25 01:14:13 INFO mapreduce.Job: map 49% reduce 2%
22/10/25 01:14:22 INFO mapreduce.Job: map 50% reduce 2%
22/10/25 01:14:31 INFO mapreduce.Job: map 51% reduce 2%
22/10/25 01:14:38 INFO mapreduce.Job: map 52% reduce 2%
22/10/25 01:14:46 INFO mapreduce.Job: map 53% reduce 2%
22/10/25 01:14:54 INFO mapreduce.Job: map 54% reduce 2%
22/10/25 01:15:02 INFO mapreduce.Job: map 55% reduce 2%
22/10/25 01:15:10 INFO mapreduce.Job: map 56% reduce 2%

```
22/10/25 01:15:19 INFO mapreduce.Job:  map 57% reduce 2%
22/10/25 01:15:27 INFO mapreduce.Job:  map 58% reduce 2%
22/10/25 01:15:33 INFO mapreduce.Job:  map 59% reduce 2%
22/10/25 01:15:41 INFO mapreduce.Job:  map 60% reduce 2%
22/10/25 01:15:51 INFO mapreduce.Job:  map 61% reduce 3%
22/10/25 01:15:58 INFO mapreduce.Job:  map 62% reduce 3%
22/10/25 01:16:05 INFO mapreduce.Job:  map 63% reduce 3%
22/10/25 01:16:14 INFO mapreduce.Job:  map 64% reduce 3%
22/10/25 01:16:23 INFO mapreduce.Job:  map 65% reduce 3%
22/10/25 01:16:32 INFO mapreduce.Job:  map 66% reduce 3%
22/10/25 01:16:37 INFO mapreduce.Job:  map 67% reduce 3%
22/10/25 01:16:47 INFO mapreduce.Job:  map 68% reduce 3%
22/10/25 01:16:55 INFO mapreduce.Job:  map 69% reduce 3%
22/10/25 01:17:03 INFO mapreduce.Job:  map 70% reduce 3%
22/10/25 01:17:10 INFO mapreduce.Job:  map 71% reduce 3%
22/10/25 01:17:19 INFO mapreduce.Job:  map 72% reduce 3%
22/10/25 01:17:29 INFO mapreduce.Job:  map 73% reduce 3%
22/10/25 01:17:33 INFO mapreduce.Job:  map 74% reduce 3%
22/10/25 01:17:43 INFO mapreduce.Job:  map 75% reduce 3%
22/10/25 01:17:49 INFO mapreduce.Job:  map 76% reduce 3%
22/10/25 01:17:56 INFO mapreduce.Job:  map 77% reduce 3%
22/10/25 01:18:05 INFO mapreduce.Job:  map 78% reduce 3%
22/10/25 01:18:14 INFO mapreduce.Job:  map 79% reduce 3%
22/10/25 01:18:20 INFO mapreduce.Job:  map 80% reduce 3%
22/10/25 01:18:29 INFO mapreduce.Job:  map 81% reduce 3%
22/10/25 01:18:38 INFO mapreduce.Job:  map 82% reduce 3%
22/10/25 01:18:48 INFO mapreduce.Job:  map 83% reduce 3%
22/10/25 01:18:52 INFO mapreduce.Job:  map 84% reduce 3%
22/10/25 01:19:02 INFO mapreduce.Job:  map 85% reduce 3%
22/10/25 01:19:05 INFO mapreduce.Job:  map 85% reduce 4%
22/10/25 01:19:10 INFO mapreduce.Job:  map 86% reduce 4%
22/10/25 01:19:20 INFO mapreduce.Job:  map 87% reduce 4%
22/10/25 01:19:27 INFO mapreduce.Job:  map 88% reduce 4%
22/10/25 01:19:34 INFO mapreduce.Job:  map 89% reduce 4%
22/10/25 01:19:44 INFO mapreduce.Job:  map 90% reduce 4%
22/10/25 01:19:55 INFO mapreduce.Job:  map 91% reduce 4%
22/10/25 01:20:02 INFO mapreduce.Job:  map 92% reduce 4%
22/10/25 01:20:09 INFO mapreduce.Job:  map 93% reduce 4%
22/10/25 01:20:17 INFO mapreduce.Job:  map 94% reduce 4%
22/10/25 01:20:27 INFO mapreduce.Job:  map 95% reduce 4%
22/10/25 01:20:35 INFO mapreduce.Job:  map 96% reduce 4%
22/10/25 01:20:41 INFO mapreduce.Job:  map 97% reduce 4%
22/10/25 01:20:51 INFO mapreduce.Job:  map 98% reduce 4%
22/10/25 01:20:58 INFO mapreduce.Job:  map 99% reduce 4%
```

22/10/25 01:21:04 INFO mapreduce.Job: map 100% reduce 4%
22/10/25 01:21:12 INFO mapreduce.Job: map 100% reduce 8%
22/10/25 01:21:18 INFO mapreduce.Job: map 100% reduce 13%
22/10/25 01:21:19 INFO mapreduce.Job: map 100% reduce 24%
22/10/25 01:21:20 INFO mapreduce.Job: map 100% reduce 36%
22/10/25 01:21:21 INFO mapreduce.Job: map 100% reduce 38%
22/10/25 01:21:34 INFO mapreduce.Job: map 100% reduce 48%
22/10/25 01:21:36 INFO mapreduce.Job: map 100% reduce 60%
22/10/25 01:21:37 INFO mapreduce.Job: map 100% reduce 73%
22/10/25 01:21:39 INFO mapreduce.Job: map 100% reduce 75%
22/10/25 01:21:51 INFO mapreduce.Job: map 100% reduce 86%
22/10/25 01:21:52 INFO mapreduce.Job: map 100% reduce 97%
22/10/25 01:21:53 INFO mapreduce.Job: map 100% reduce 98%
22/10/25 01:21:55 INFO mapreduce.Job: map 100% reduce 100%
22/10/25 01:21:55 INFO mapreduce.Job: Job job_1666646922232_0020 completed successfully
22/10/25 01:21:55 INFO mapreduce.Job: Counters: 56
        File System Counters
                FILE: Number of bytes read=626459834
                FILE: Number of bytes written=1779299523
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=34643
                HDFS: Number of bytes written=2148498
                HDFS: Number of read operations=620
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=16
                S3: Number of bytes read=19749820717
                S3: Number of bytes written=0
                S3: Number of read operations=0
                S3: Number of large read operations=0
                S3: Number of write operations=0
        Job Counters
                Killed map tasks=1
                Killed reduce tasks=1
                Launched map tasks=298
                Launched reduce tasks=9
                Data-local map tasks=298
                Total time spent by all maps in occupied slots (ms)=331396992
                Total time spent by all reduces in occupied slots (ms)=126283200
                Total time spent by all map tasks (ms)=3452052
                Total time spent by all reduce tasks (ms)=657725
                Total vcore-milliseconds taken by all map tasks=3452052
                Total vcore-milliseconds taken by all reduce tasks=657725

Total megabyte-milliseconds taken by all map tasks=10604703744
Total megabyte-milliseconds taken by all reduce tasks=4041062400
Map-Reduce Framework
Map input records=173179771
Map output records=173179759
Map output bytes=7100370119
Map output materialized bytes=1085096722
Input split bytes=34643
Combine input records=0
Combine output records=0
Reduce input groups=43191
Reduce shuffle bytes=1085096722
Reduce input records=173179759
Reduce output records=43191
Spilled Records=346359518
Shuffled Maps =2384
Failed Shuffles=0
Merged Map outputs=2384
GC time elapsed (ms)=143329
CPU time spent (ms)=3530250
Physical memory (bytes) snapshot=299890036736
Virtual memory (bytes) snapshot=1449987616768
Total committed heap usage (bytes)=287309824000
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=19749820717
File Output Format Counters
Bytes Written=2148498
22/10/25 01:21:55 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-40-62.us-east-2.compute.internal/172.31.40.62:8032
22/10/25 01:21:55 INFO client.AHSProxy: Connecting to Application History server at ip-172-31-40-62.us-east-2.compute.internal/172.31.40.62:10200
22/10/25 01:21:55 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
22/10/25 01:21:55 INFO input.FileInputFormat: Total input files to process : 8
22/10/25 01:21:55 INFO mapreduce.JobSubmitter: number of splits:8

22/10/25 01:21:55 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1666646922232_0021
22/10/25 01:21:55 INFO impl.YarnClientImpl: Submitted application application_1666646922232_0021
22/10/25 01:21:55 INFO mapreduce.Job: The url to track the job: http://ip-172-31-40-62.us-east-2.compute.internal:20888/proxy/application_1666646922232_0021/
22/10/25 01:21:55 INFO mapreduce.Job: Running job: job_1666646922232_0021
22/10/25 01:22:00 INFO mapreduce.Job: Job job_1666646922232_0021 running in uber mode : false
22/10/25 01:22:00 INFO mapreduce.Job: map 0% reduce 0%
22/10/25 01:22:07 INFO mapreduce.Job: map 25% reduce 0%
22/10/25 01:22:10 INFO mapreduce.Job: map 75% reduce 0%
22/10/25 01:22:11 INFO mapreduce.Job: map 100% reduce 0%
22/10/25 01:22:14 INFO mapreduce.Job: map 100% reduce 13%
22/10/25 01:22:15 INFO mapreduce.Job: map 100% reduce 38%
22/10/25 01:22:17 INFO mapreduce.Job: map 100% reduce 50%
22/10/25 01:22:18 INFO mapreduce.Job: map 100% reduce 63%
22/10/25 01:22:19 INFO mapreduce.Job: map 100% reduce 75%
22/10/25 01:22:21 INFO mapreduce.Job: map 100% reduce 88%
22/10/25 01:22:22 INFO mapreduce.Job: map 100% reduce 100%
22/10/25 01:22:22 INFO mapreduce.Job: Job job_1666646922232_0021 completed successfully
22/10/25 01:22:22 INFO mapreduce.Job: Counters: 51
        File System Counters
            FILE: Number of bytes read=2126
            FILE: Number of bytes written=3552298
            FILE: Number of read operations=0
            FILE: Number of large read operations=0
            FILE: Number of write operations=0
            HDFS: Number of bytes read=2149730
            HDFS: Number of bytes written=255
            HDFS: Number of read operations=48
            HDFS: Number of large read operations=0
            HDFS: Number of write operations=16
        Job Counters
            Killed map tasks=1
            Launched map tasks=8
            Launched reduce tasks=8
            Data-local map tasks=7
            Rack-local map tasks=1
            Total time spent by all maps in occupied slots (ms)=4515264
            Total time spent by all reduces in occupied slots (ms)=4357056
            Total time spent by all map tasks (ms)=47034
            Total time spent by all reduce tasks (ms)=22693
            Total vcore-milliseconds taken by all map tasks=47034

Total vcore-milliseconds taken by all reduce tasks=22693
Total megabyte-milliseconds taken by all map tasks=144488448
Total megabyte-milliseconds taken by all reduce tasks=139425792
Map-Reduce Framework
Map input records=43191
Map output records=40
Map output bytes=2220
Map output materialized bytes=3163
Input split bytes=1232
Combine input records=0
Combine output records=0
Reduce input groups=1
Reduce shuffle bytes=3163
Reduce input records=40
Reduce output records=5
Spilled Records=80
Shuffled Maps =64
Failed Shuffles=0
Merged Map outputs=64
GC time elapsed (ms)=2133
CPU time spent (ms)=10280
Physical memory (bytes) snapshot=6513569792
Virtual memory (bytes) snapshot=95766466560
Total committed heap usage (bytes)=6352797696
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=2148498
File Output Format Counters
Bytes Written=255