

Carrier delay processing

Using MapReduce

Iteration 01

```
hadoop@ip-172-31-67-69:~  
hive> SELECT Year, avg((CarrierDelay /ArrDelay)*100) from delay_flights GROUP BY Year;  
Query ID = hadoop_20240303032049_044f1e3c-8b3d-4560-bfb5-3d1dabf661e4  
Total jobs = 1  
Launching Job 1 out of 1  
Tez session was closed. Reopening...  
Session re-established.  
Session re-established.  
Status: Running (Executing on YARN cluster with App id application_1709434981938_0002)
```

	VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	1	1	0	0	0	0	0
Reducer 2	container	SUCCEEDED	2	2	0	0	0	0	0

VERTICES: 02/02 [=====>>] 100% ELAPSED TIME: 9.59 s

```
OK  
2004 43.64459443230066  
2009 28.33058554239575  
NULL NULL  
2003 24.557549755575373  
2005 28.01977637202288  
2006 30.453296261292596  
2007 19.850007017971283  
2008 28.88346981456985  
2010 21.89310246015957  
Time taken: 14.932 seconds, Fetched: 9 row(s)
```

Iteration 02

```
hive> SELECT Year, avg((CarrierDelay /ArrDelay)*100) from delay_flights GROUP BY Year;  
Query ID = hadoop_20240303032140_ff84ea81-0f59-4524-8bbf-f35804f72ffe  
Total jobs = 1  
Launching Job 1 out of 1  
Status: Running (Executing on YARN cluster with App id application_1709434981938_0002)
```

	VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	1	1	0	0	0	0	0
Reducer 2	container	SUCCEEDED	2	2	0	0	0	0	0

VERTICES: 02/02 [=====>>] 100% ELAPSED TIME: 6.77 s

```
OK  
2004 43.64459443230066  
2009 28.33058554239575  
NULL NULL  
2003 24.557549755575373  
2005 28.01977637202288  
2006 30.453296261292596  
2007 19.850007017971283  
2008 28.88346981456985  
2010 21.89310246015957  
Time taken: 7.103 seconds, Fetched: 9 row(s)
```

Iteration 03

```
hive> SELECT Year, avg((CarrierDelay /ArrDelay)*100) from delay_flights GROUP BY Year;
Query ID = hadoop_20240303032153_82066c8b-ale7-433e-864d-a95ca7alcc42
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1709434981938_0002)
```

	VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	1	1	0	0	0	0	0
Reducer 2	container	SUCCEEDED	2	2	0	0	0	0	0

VERTICES: 02/02 [=====>>] 100% ELAPSED TIME: 6.84 s

```
OK
2004 43.64459443230066
2009 28.33058554239575
NULL NULL
2003 24.557549755575373
2005 28.01977637202288
2006 30.453296261292596
2007 19.850007017971283
2008 28.88346981456985
2010 21.89310246015957
Time taken: 7.094 seconds, Fetched: 9 row(s)
```

Iteration 04

```
hive> SELECT Year, avg((CarrierDelay /ArrDelay)*100) from delay_flights GROUP BY Year;
Query ID = hadoop_20240303032207_24818404-9fd4-48a4-a9c1-4c3c26dcceaf
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1709434981938_0002)
```

	VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	1	1	0	0	0	0	0
Reducer 2	container	SUCCEEDED	2	2	0	0	0	0	0

VERTICES: 02/02 [=====>>] 100% ELAPSED TIME: 6.58 s

```
OK
2004 43.64459443230066
2009 28.33058554239575
NULL NULL
2003 24.557549755575373
2005 28.01977637202288
2006 30.453296261292596
2007 19.850007017971283
2008 28.88346981456985
2010 21.89310246015957
Time taken: 6.787 seconds, Fetched: 9 row(s)
```

Iteration 05

```
hive> SELECT Year, avg((CarrierDelay /ArrDelay)*100) from delay_flights GROUP BY Year;
Query ID = hadoop_20240303032217_c9b24952-1390-4db6-b761-fb097db024cf
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1709434981938_0002)
```

```
-----
      VERTICES      MODE      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container  SUCCEEDED    1         1         0         0         0         0
Reducer 2 ..... container  SUCCEEDED    2         2         0         0         0         0
-----
VERTICES: 02/02  [=====>>] 100%  ELAPSED TIME: 1.01 s
-----
OK
2004      43.64459443230066
2009      28.33058554239575
NULL      NULL
2003      24.557549755575373
2005      28.01977637202288
2006      30.453296261292596
2007      19.850007017971283
2008      28.88346981456985
2010      21.89310246015957
Time taken: 1.243 seconds, Fetched: 9 row(s)
```

Carrier Delay processing with Spark-SQL

Iteration 01, 02, 03

```
spark-sql (default)> SELECT Year, avg((NASDelay /ArrDelay)*100) from delay_flights GROUP BY Year;
NULL      NULL
2003      29.686276314267346
2004      18.24570061769958
2005      16.638688053731286
2006      18.1193123299377
2007      30.625925917941924
2008      30.165525625941324
2009      37.63093330628511
2010      33.87351363404216
Time taken: 0.317 seconds, Fetched 9 row(s)
spark-sql (default)> SELECT Year, avg((CarrierDelay /ArrDelay)*100) from delay_flights GROUP BY Year;
NULL      NULL
2003      24.557549755575373
2004      43.64459443230066
2005      28.01977637202288
2006      30.453296261292596
2007      19.850007017971283
2008      28.88346981456985
2009      28.33058554239575
2010      21.893102460159568
Time taken: 0.374 seconds, Fetched 9 row(s)
spark-sql (default)> SELECT Year, avg((CarrierDelay /ArrDelay)*100) from delay_flights GROUP BY Year;
NULL      NULL
2004      43.64459443230066
2005      28.01977637202288
2006      30.453296261292596
2007      19.850007017971283
2008      28.88346981456985
2009      28.33058554239575
2010      21.893102460159568
2003      24.557549755575373
Time taken: 0.395 seconds, Fetched 9 row(s)
```

Iteration 04 and 05

```
spark-sql (default)> SELECT Year, avg((CarrierDelay /ArrDelay)*100) from delay_flights GROUP BY Year;
NULL      NULL
2004      43.64459443230066
2005      28.01977637202288
2006      30.453296261292596
2007      19.850007017971283
2008      28.88346981456985
2009      28.33058554239575
2010      21.893102460159568
2003      24.557549755575373
Time taken: 0.311 seconds, Fetched 9 row(s)
spark-sql (default)> SELECT Year, avg((CarrierDelay /ArrDelay)*100) from delay_flights GROUP BY Year;
NULL      NULL
2003      24.557549755575373
2004      43.64459443230066
2005      28.01977637202288
2006      30.453296261292596
2007      19.850007017971283
2008      28.88346981456985
2009      28.33058554239575
2010      21.893102460159568
Time taken: 0.302 seconds, Fetched 9 row(s)
spark-sql (default)> SELECT Year, avg((CarrierDelay /ArrDelay)*100) from delay_flights GROUP BY Year;
NULL      NULL
2003      24.557549755575373
2004      43.64459443230066
2005      28.01977637202288
2006      30.453296261292596
2007      19.850007017971283
2008      28.88346981456985
2009      28.33058554239575
2010      21.893102460159568
Time taken: 0.308 seconds, Fetched 9 row(s)
```