

MongoDB Case study

Review: Group 11

```
> db.delayData.find().sort({year:1}).limit(1).pretty();
{
    "_id" : ObjectId("625d88a82c8be4ce18263bf5"),
    "year" : 2003,
    "month" : 6,
    "carrier" : "AA",
    "carrier_name" : "American Airlines Inc.",
    "airport" : "ABQ",
    "airport_name" : "Albuquerque, NM: Albuquerque International Sunport",
    "arr_flights" : 307,
    "arr_del15" : 56,
    "carrier_ct" : 14.68,
    "weather_ct" : 10.79,
    "nas_ct" : 19.09,
    "security_ct" : 1.48,
    "late_aircraft_ct" : 9.96,
    "arr_cancelled" : 1,
    "arr_diverted" : 1,
    "arr_delay" : 2530,
    "carrier_delay" : 510,
    "weather_delay" : 621,
    "nas_delay" : 676,
    "security_delay" : 25,
    "late_aircraft_delay" : 698,
    "" : ""
}
```

FINDING THE MINIMUM
AND MAXIMUM YEAR
RECORDED IN THE
DATABASE

QUERY 1:
`db.delayData.find().sort({year:1}).limit(1).pretty();`

```
> db.delayData.find().sort({year:-1}).limit(1).pretty();
{
    "_id" : ObjectId("625d88cb2c8be4ce1828be69"),
    "year" : 2013,
    "month" : 1,
    "carrier" : "9E",
    "carrier_name" : "Pinnacle Airlines Inc.",
    "airport" : "DFW",
    "airport_name" : "Dallas/Fort Worth, TX: Dallas/Fort Worth International",
    "arr_flights" : 139,
    "arr_del15" : 37,
    "carrier_ct" : 13.04,
    "weather_ct" : 0.8,
    "nas_ct" : 13.92,
    "security_ct" : 0,
    "late_aircraft_ct" : 9.24,
    "arr_cancelled" : 2,
    "arr_diverted" : 5,
    "arr_delay" : 1860,
    "carrier_delay" : 789,
    "weather_delay" : 75,
    "nas_delay" : 451,
    "security_delay" : 0,
    "late_aircraft_delay" : 545,
    "" : ""
}
```

QUERY 2:

```
db.delayData.find().sort({year:-1}).limit(1).pretty();
```

```
> db.delayData.aggregate([{$group : {_id : "$carrier", MinYear : {$min : "$year"}}}]);  
{ "_id" : "VX", "MinYear" : 2012 }  
{ "_id" : "XE", "MinYear" : 2006 }  
{ "_id" : "AQ", "MinYear" : 2006 }  
{ "_id" : "F9", "MinYear" : 2005 }  
{ "_id" : "OH", "MinYear" : 2004 }  
{ "_id" : "HA", "MinYear" : 2003 }  
{ "_id" : "9E", "MinYear" : 2007 }  
{ "_id" : "WN", "MinYear" : 2003 }  
{ "_id" : "US", "MinYear" : 2003 }  
{ "_id" : "AA", "MinYear" : 2003 }  
{ "_id" : "AS", "MinYear" : 2003 }  
{ "_id" : "TZ", "MinYear" : 2003 }  
{ "_id" : "DH", "MinYear" : 2003 }  
{ "_id" : "B6", "MinYear" : 2003 }  
{ "_id" : "UA", "MinYear" : 2003 }  
{ "_id" : "CO", "MinYear" : 2003 }  
{ "_id" : "MQ", "MinYear" : 2003 }  
{ "_id" : "EV", "MinYear" : 2003 }  
{ "_id" : "DL", "MinYear" : 2003 }  
{ "_id" : "NW", "MinYear" : 2003 }  
Type "it" for more
```

FINDING THE YEARS OF SERVICE [AGGREGATE FUNCTIONS]

QUERY 3:

```
db.delayData.aggregate([{$group : {_id : "$carrier", MinYear : {$min : "$year"}}}]);
```

```
> db.delayData.aggregate([{$group : {_id : "$carrier", MaxYear : {$max : "$year"}}}]);  
{ "_id" : "VX", "MaxYear" : 2013 }  
{ "_id" : "XE", "MaxYear" : 2011 }  
{ "_id" : "AQ", "MaxYear" : 2008 }  
{ "_id" : "F9", "MaxYear" : 2013 }  
{ "_id" : "OH", "MaxYear" : 2010 }  
{ "_id" : "HA", "MaxYear" : 2013 }  
{ "_id" : "9E", "MaxYear" : 2013 }  
{ "_id" : "WN", "MaxYear" : 2013 }  
{ "_id" : "US", "MaxYear" : 2013 }  
{ "_id" : "AA", "MaxYear" : 2013 }  
{ "_id" : "AS", "MaxYear" : 2013 }  
{ "_id" : "TZ", "MaxYear" : 2006 }  
{ "_id" : "DH", "MaxYear" : 2005 }  
{ "_id" : "B6", "MaxYear" : 2013 }  
{ "_id" : "UA", "MaxYear" : 2013 }  
{ "_id" : "CO", "MaxYear" : 2011 }  
{ "_id" : "MQ", "MaxYear" : 2013 }  
{ "_id" : "EV", "MaxYear" : 2013 }  
{ "_id" : "DL", "MaxYear" : 2013 }  
{ "_id" : "NW", "MaxYear" : 2009 }  
Type "it" for more
```

QUERY 4:

```
db.delayData.aggregate([{$group : {_id :  
"$carrier", MaxYear : {$max : "$year"}}}]);
```

```
> db.delayData.aggregate([{$group : {_id : "$carrier", TotalDelay: {$sum : "$carrier_ct"}}}]);  
{ "_id" : "VX", "TotalDelay" : 2740.08 }  
{ "_id" : "XE", "TotalDelay" : 102311.54 }  
{ "_id" : "AQ", "TotalDelay" : 3257.98 }  
{ "_id" : "F9", "TotalDelay" : 37074.63 }  
{ "_id" : "OH", "TotalDelay" : 146193.59 }  
{ "_id" : "HA", "TotalDelay" : 27486.12 }  
{ "_id" : "9E", "TotalDelay" : 65592.78 }  
{ "_id" : "WN", "TotalDelay" : 523048.34 }  
{ "_id" : "US", "TotalDelay" : 229268.93 }  
{ "_id" : "AA", "TotalDelay" : 367637.12 }  
{ "_id" : "AS", "TotalDelay" : 88999.15 }  
{ "_id" : "TZ", "TotalDelay" : 6563.67 }  
{ "_id" : "DH", "TotalDelay" : 26079.83 }  
{ "_id" : "B6", "TotalDelay" : 103185.42 }  
{ "_id" : "UA", "TotalDelay" : 243557.23 }  
{ "_id" : "CO", "TotalDelay" : 112686.43 }  
{ "_id" : "MQ", "TotalDelay" : 266857.85 }  
{ "_id" : "EV", "TotalDelay" : 294554.02 }  
{ "_id" : "DL", "TotalDelay" : 312964.66 }  
{ "_id" : "NW", "TotalDelay" : 198141.92 }
```

TO FIND THE TOTAL DELAY
OCCURRED DUE TO
CARRIER LOAD-UNLOAD

Query 5:

```
db.delayData.aggregate([{$group : {_id : "$carrier", TotalDelay: {$sum : "$carrier_ct"}}}]);
```

```
> db.delayDB.insert([{"year" : 2022, "month" : 4, "carrier" : "AAA", "carrier_name" : "Divesh Airlines Inc.", "airport" : "DSM", "airport_name" : "Divesh International", "arr_flights" : 422, "arr_del15" : 15, "carrier_ct" : 3.39, "weather_ct" : 1.57, "nas_ct" : 5.07, "security_ct" : 0, "late_aircraft_ct" : 6.97, "arr_cancelled" : 0, "arr_diverted" : 0, "arr_delay" : 0, "carrier_delay" : 212, "weather_delay" : 50, "nas_delay" : 259, "security_delay" : 0, "late_aircraft_delay" : 345, "" : ""}]);  
BulkWriteResult({  
    "writeErrors" : [ ],  
    "writeConcernErrors" : [ ],  
    "nInserted" : 1,  
    "nUpserted" : 0,  
    "nMatched" : 0,  
    "nModified" : 0,  
    "nRemoved" : 0,  
    "upserted" : [ ]  
})
```

INSERTING NEW RECORD

QUERY 6: db.delayDB.insert([{"year" : 2022, "month" : 4, "carrier" : "AAA", "carrier_name" : "Divesh Airlines Inc.", "airport" : "DSM", "airport_name" : "Divesh International", "arr_flights" : 422, "arr_del15" : 15, "carrier_ct" : 3.39, "weather_ct" : 1.57, "nas_ct" : 5.07, "security_ct" : 0, "late_aircraft_ct" : 6.97, "arr_cancelled" : 0, "arr_diverted" : 0, "arr_delay" : 0, "carrier_delay" : 212, "weather_delay" : 50, "nas_delay" : 259, "security_delay" : 0, "late_aircraft_delay" : 345, "" : ""}]);

```
> var mapCheck1 = function(){ emit( this.carrier_name, this.carrier_delay );}
> var reduceCheck1 = function( key, valArr ) {return Array.sum( valArr );}
> db.delayData.mapReduce( mapCheck1, reduceCheck1,{out:"outputMR"});
{
    "result" : "outputMR",
    "timeMillis" : 2446,
    "counts" : {
        "input" : 173657,
        "emit" : 173657,
        "reduce" : 4272,
        "output" : 25
    },
    "ok" : 1
}
```

MAP-REDUCE QUERY 1

QUERY 7: Defining Mapper and Reducer

- var mapCheck1 = function(){ emit(this.carrier_name, this.carrier_delay);}
- var reduceCheck1 = function(key, valArr) {return Array.sum(valArr);}
- db.delayData.mapReduce(mapCheck1, reduceCheck1,{out:"outputMR"});

Loading Data

```
> db.outputMR.find().sort({year:-1});
[{"_id": "ATA Airlines d/b/a ATA", "value": 405504},
 {"_id": "AirTran Airways Corporation", "value": 4028057},
 {"_id": "Alaska Airlines Inc.", "value": 5062626},
 {"_id": "Aloha Airlines Inc.", "value": 144922},
 {"_id": "America West Airlines Inc.", "value": 1699448},
 {"_id": "American Airlines Inc.", "value": 22870499},
 {"_id": "American Eagle Airlines Inc.", "value": 16095940},
 {"_id": "Atlantic Coast Airlines", "value": 1064001},
 {"_id": "Atlantic Southeast Airlines", "value": 13837566},
 {"_id": "Comair Inc.", "value": 8415020},
 {"_id": "Continental Air Lines Inc.", "value": 6480851},
 {"_id": "Delta Air Lines Inc.", "value": 19040800},
 {"_id": "ExpressJet Airlines Inc.", "value": 14492140},
 {"_id": "Frontier Airlines Inc.", "value": 1721862},
 {"_id": "Hawaiian Airlines Inc.", "value": 1355530},
 {"_id": "Independence Air", "value": 304026},
 {"_id": "JetBlue Airways", "value": 6512832},
 {"_id": "Mesa Airlines Inc.", "value": 8681157},
 {"_id": "Northwest Airlines Inc.", "value": 11745027},
 {"_id": "Pinnacle Airlines Inc.", "value": 4365161}]
Type "it" for more
> it
[{"_id": "SkyWest Airlines Inc.", "value": 17016279},
 {"_id": "Southwest Airlines Co.", "value": 22184643},
 {"_id": "US Airways Inc.", "value": 11378650},
 {"_id": "United Air Lines Inc.", "value": 14375218},
 {"_id": "Virgin America", "value": 154025}]
```

QUERY 8:

Loading the output with sorted order according to the requirement

```
> var mapCheck2 = function(){ emit( this.carrier_name, this.weather_delay );}
> var reduceCheck2 = function( key, valArr ) {return Array.sum( valArr );}
> db.delayData.mapReduce( mapCheck2, reduceCheck2, {out:"NewOutputMR"});
{
    "result" : "NewOutputMR",
    "timeMillis" : 992,
    "counts" : {
        "input" : 173657,
        "emit" : 173657,
        "reduce" : 4272,
        "output" : 25
    },
    "ok" : 1
}
```

MAP-REDUCE QUERY 2

QUERY 9: Defining Mapper and Reducer

- var mapCheck2 = function(){ emit(this.carrier_name, this.weather_delay);}
- var reduceCheck2 = function(key, valArr) {return Array.sum(valArr);}
- db.delayData.mapReduce(mapCheck2, reduceCheck2, {out:"NewOutputMR"});

```
> db.NewOutputMR.find().sort({year:-1})
{ "_id" : "ATA Airlines d/b/a ATA", "value" : 23136 }
{ "_id" : "AirTran Airways Corporation", "value" : 306101 }
{ "_id" : "Alaska Airlines Inc.", "value" : 345748 }
{ "_id" : "Aloha Airlines Inc.", "value" : 2297 }
{ "_id" : "America West Airlines Inc.", "value" : 71713 }
{ "_id" : "American Airlines Inc.", "value" : 5228234 }
{ "_id" : "American Eagle Airlines Inc.", "value" : 4157053 }
{ "_id" : "Atlantic Coast Airlines", "value" : 317277 }
{ "_id" : "Atlantic Southeast Airlines", "value" : 4318712 }
{ "_id" : "Comair Inc.", "value" : 4849900 }
{ "_id" : "Continental Air Lines Inc.", "value" : 1414448 }
{ "_id" : "Delta Air Lines Inc.", "value" : 2215918 }
{ "_id" : "ExpressJet Airlines Inc.", "value" : 2546530 }
{ "_id" : "Frontier Airlines Inc.", "value" : 183900 }
{ "_id" : "Hawaiian Airlines Inc.", "value" : 23615 }
{ "_id" : "Independence Air", "value" : 34186 }
{ "_id" : "JetBlue Airways", "value" : 635836 }
{ "_id" : "Mesa Airlines Inc.", "value" : 952900 }
{ "_id" : "Northwest Airlines Inc.", "value" : 2342139 }
{ "_id" : "Pinnacle Airlines Inc.", "value" : 669742 }
```

Loading Data

QUERY 10:

Loading the output with sorted order according to the requirement.

MongoDB Review

P KOUSHIK [CB.EN.U4CSE19449]

1. Updating the carrier name from the database.

- db.delayData.update({carrier_name : "American Airlines Inc."}, {\$set:{carrier_name : "Russian Airlines Inc."}}, {multi:true})

```
> db.delayData.update({carrier_name : "American Airlines Inc."}, {$set:{carrier_name : "Russian Airlines Inc."}}, {multi:true})
WriteResult({ "nMatched" : 9775, "nUpserted" : 0, "nModified" : 9775 })
> db.delayData.find()
[{"_id": ObjectId("625d7d8a46c1a4eb66438604"), "year": 2003, "month": 6, "carrier": "AA", "carrier_name": "Russian Airlines Inc.", "airport": "ABQ", "airport_name": "Albuquerque, NM: Albuquerque International Sunport", "arr_flights": 307, "arr_del15": 56, "carrier_ct": 14.68, "weather_ct": 10.79, "nas_ct": 19.09, "security_ct": 1.48, "late_aircraft_ct": 9.96, "arr_cancelled": 1, "arr_diverted": 1, "arr_delay": 2530, "carrier_delay": 510, "weather_delay": 621, "nas_delay": 676, "security_delay": 25, "late_aircraft_delay": 698, "": ""}, {"_id": ObjectId("625d7d8a46c1a4eb66438605"), "year": 2003, "month": 6, "carrier": "AA", "carrier_name": "Russian Airlines Inc.", "airport": "ANC", "airpor
t_name": "Anchorage, AK: Ted Stevens Anchorage International", "arr_flights": 90, "arr_del15": 27, "carrier_ct": 7.09, "weather_ct": 2, "nas_ct": 10.75, "security_ct": 0, "late_aircraft_ct": 7.16, "arr_cancelled": 0, "arr_diverted": 0, "arr_delay": 1390, "carrier_delay": 271, "weather_delay": 83, "nas_delay": 581, "security_delay": 0, "late_aircraft_delay": 455, "": ""}, {"_id": ObjectId("625d7d8a46c1a4eb66438606"), "year": 2003, "month": 6, "carrier": "AA", "carrier_name": "Russian Airlines Inc.", "airport": "ATL", "airpo
rt_name": "Atlanta, GA: Hartsfield-Jackson Atlanta International", "arr_flights": 752, "arr_del15": 186, "carrier_ct": 33.99, "weather_ct": 27.82, "nas_ct": 104.76, "security_ct": 1.9, "late_aircraft_ct": 17.53, "arr_cancelled": 5, "arr_diverted": 0, "arr_delay": 8314, "carrier_delay": 1367, "weather_delay": 1722, "nas_delay": 3817, "security_delay": 139, "late_aircraft_delay": 1269, "": ""}, {"_id": ObjectId("625d7d8a46c1a4eb66438607"), "year": 2003, "month": 6, "carrier": "AA", "carrier_name": "Russian Airlines Inc.", "airport": "AUS", "airpo
rt_name": "Austin, TX: Austin - Bergstrom International", "arr_flights": 842, "arr_del15": 174, "carrier_ct": 60.24, "weather_ct": 20.54, "nas_ct": 47.78, "security_ct": 4.69, "late_aircraft_ct": 40.75, "arr_cancelled": 9, "arr_diverted": 1, "arr_delay": 8344, "carrier_delay": 3040, "weather_delay": 1032, "nas_delay": 1835, "security_delay": 115, "late_aircraft_delay": 2322, "": ""}, {"_id": ObjectId("625d7d8a46c1a4eb66438608"), "year": 2003, "month": 6, "carrier": "AA", "carrier_name": "Russian Airlines Inc.", "airport": "BDL", "airpo
rt_name": "Hartford, CT: Bradley International", "arr_flights": 383, "arr_del15": 55, "carrier_ct": 14.9, "weather_ct": 8.91, "nas_ct": 14.57, "security_ct": 0, "late_aircraft_ct": 16.61, "arr_cancelled": 0, "arr_diverted": 0, "arr_delay": 3137, "carrier_delay": 915, "weather_delay": 574, "nas_delay": 555} ]
```

2. Finding the maximum number of flights cancelled for that airport

- db.delayData.aggregate([{\$group : {_id : "\$airport", number : {\$max : "\$arr_cancelled"}}}])

```
type "it" for more
> db.delayData.aggregate([{$group : {_id : "$airport", number : {$max : "$arr_cancelled"}}}])
{ "_id" : "PPG", "number" : 0 }
{ "_id" : "CIU", "number" : 5 }
{ "_id" : "SHD", "number" : 0 }
{ "_id" : "AZA", "number" : 1 }
{ "_id" : "ISN", "number" : 7 }
{ "_id" : "ESC", "number" : 7 }
{ "_id" : "BRD", "number" : 4 }
{ "_id" : "HOB", "number" : 8 }
{ "_id" : "GRI", "number" : 4 }
{ "_id" : "MVY", "number" : 5 }
{ "_id" : "GUM", "number" : 2 }
{ "_id" : "ABR", "number" : 10 }
{ "_id" : "UTM", "number" : 3 }
{ "_id" : "MHK", "number" : 13 }
{ "_id" : "MWH", "number" : 5 }
{ "_id" : "SAF", "number" : 10 }
{ "_id" : "ITH", "number" : 9 }
{ "_id" : "OTH", "number" : 8 }
{ "_id" : "LMT", "number" : 10 }
{ "_id" : "YKM", "number" : 1 }
Type "it" for more
>
```

3. Finding the flights data which ends with a Strings

- db.delayData.find({airport :/D\$/});

```
> db.student.find({airport :/D$/});
> db.delayData.find({airport :/D$/});
{ "_id" : ObjectId("625d7d8a46c1a4eb6643861f"), "year" : 2003, "month" : 6, "carrier" : "AA", "carrier_name" : "Russian Airlines Inc.", "airport" : "IAD", "airpo
rt_name" : "Washington, DC: Washington Dulles International", "arr_flights" : 445, "arr_del15" : 76, "carrier_ct" : 17.58, "weather_ct" : 15.86, "nas_ct" : 32.62
, "security_ct" : 0, "late_aircraft_ct" : 9.94, "arr_cancelled" : 2, "arr_diverted" : 3, "arr_delay" : 3661, "carrier_delay" : 997, "weather_delay" : 759, "nas_d
elay" : 1225, "security_delay" : 0, "late_aircraft_delay" : 680, "" : "" }
{ "_id" : ObjectId("625d7d8a46c1a4eb66438621"), "year" : 2003, "month" : 6, "carrier" : "AA", "carrier_name" : "Russian Airlines Inc.", "airport" : "IND", "airpo
rt_name" : "Indianapolis, IN: Indianapolis International", "arr_flights" : 278, "arr_del15" : 50, "carrier_ct" : 17.92, "weather_ct" : 5.17, "nas_ct" : 14.94, "s
ecurity_ct" : 1, "late_aircraft_ct" : 10.97, "arr_cancelled" : 3, "arr_diverted" : 0, "arr_delay" : 2861, "carrier_delay" : 758, "weather_delay" : 800, "nas_dela
y" : 743, "security_delay" : 46, "late_aircraft_delay" : 514, "" : "" }
{ "_id" : ObjectId("625d7d8a46c1a4eb6643863c"), "year" : 2003, "month" : 6, "carrier" : "AA", "carrier_name" : "Russian Airlines Inc.", "airport" : "ORD", "airpo
rt_name" : "Chicago, IL: Chicago O'Hare International", "arr_flights" : 7628, "arr_del15" : 1309, "carrier_ct" : 225.53, "weather_ct" : 48.83, "nas_ct" : 763.26,
"security_ct" : 9.15, "late_aircraft_ct" : 262.23, "arr_cancelled" : 41, "arr_diverted" : 14, "arr_delay" : 72172, "carrier_delay" : 19608, "weather_delay" : 34
40, "nas_delay" : 34502, "security_delay" : 403, "late_aircraft_delay" : 14219, "" : "" }
{ "_id" : ObjectId("625d7d8a46c1a4eb66438641"), "year" : 2003, "month" : 6, "carrier" : "AA", "carrier_name" : "Russian Airlines Inc.", "airport" : "PVD", "airpo
rt_name" : "Providence, RI: Theodore Francis Green State", "arr_flights" : 118, "arr_del15" : 28, "carrier_ct" : 6.15, "weather_ct" : 0.75, "nas_ct" : 13.51, "se
curity_ct" : 0, "late_aircraft_ct" : 7.59, "arr_cancelled" : 0, "arr_diverted" : 0, "arr_delay" : 1216, "carrier_delay" : 274, "weather_delay" : 64, "nas_delay"
: 424, "security_delay" : 0, "late_aircraft_delay" : 454, "" : "" }
{ "_id" : ObjectId("625d7d8a46c1a4eb6643866c"), "year" : 2003, "month" : 6, "carrier" : "AS", "carrier_name" : "Alaska Airlines Inc.", "airport" : "IAD", "airpor
t_name" : "Washington, DC: Washington Dulles International", "arr_flights" : 60, "arr_del15" : 14, "carrier_ct" : 4.79, "weather_ct" : 0, "nas_ct" : 7.77, "secu
rity_ct" : 0.5, "late_aircraft_ct" : 0.94, "arr_cancelled" : 0, "arr_diverted" : 0, "arr_delay" : 561, "carrier_delay" : 201, "weather_delay" : 0, "nas_delay" : 2
```

4. Skip:

- db.delayData.find().limit(6).skip(1).sort({security_delay: 1});

```
> db.delayData.find().limit(6).skip(1).sort({security_delay: 1});
{ "_id" : ObjectId("625d7d8a46c1a4eb66438608"), "year" : 2003, "month" : 6, "carrier" : "AA", "carrier_name" : "Russian Airlines Inc.", "airport" : "BDL", "airport_name" : "Hartford, CT: Bradley International", "arr_flights" : 383, "arr_del15" : 55, "carrier_ct" : 14.9, "weather_ct" : 8.91, "nas_ct" : 14.57, "security_ct" : 0, "late_aircraft_ct" : 16.61, "arr_cancelled" : 0, "arr_diverted" : 0, "arr_delay" : 3137, "carrier_delay" : 815, "weather_delay" : 574, "nas_delay" : 555, "security_delay" : 0, "late_aircraft_delay" : 1193, "" : "" }
{ "_id" : ObjectId("625d7d8a46c1a4eb6643860c"), "year" : 2003, "month" : 6, "carrier" : "AA", "carrier_name" : "Russian Airlines Inc.", "airport" : "BUR", "airport_name" : "Burbank, CA: Bob Hope", "arr_flights" : 119, "arr_del15" : 27, "carrier_ct" : 7.49, "weather_ct" : 4.65, "nas_ct" : 7.81, "security_ct" : 0, "late_aircraft_ct" : 7.04, "arr_cancelled" : 1, "arr_diverted" : 0, "arr_delay" : 1187, "carrier_delay" : 326, "weather_delay" : 191, "nas_delay" : 285, "security_delay" : 0, "late_aircraft_delay" : 385, "" : "" }
{ "_id" : ObjectId("625d7d8a46c1a4eb6643860d"), "year" : 2003, "month" : 6, "carrier" : "AA", "carrier_name" : "Russian Airlines Inc.", "airport" : "CLE", "airport_name" : "Cleveland, OH: Cleveland-Hopkins International", "arr_flights" : 172, "arr_del15" : 29, "carrier_ct" : 5.51, "weather_ct" : 3.21, "nas_ct" : 11.76, "security_ct" : 0, "late_aircraft_ct" : 8.51, "arr_cancelled" : 0, "arr_diverted" : 0, "arr_delay" : 2037, "carrier_delay" : 598, "weather_delay" : 250, "nas_delay" : 418, "security_delay" : 0, "late_aircraft_delay" : 771, "" : "" }
{ "_id" : ObjectId("625d7d8a46c1a4eb6643860e"), "year" : 2003, "month" : 6, "carrier" : "AA", "carrier_name" : "Russian Airlines Inc.", "airport" : "BWI", "airport_name" : "Baltimore, MD: Baltimore/Washington International Thurgood Marshall", "arr_flights" : 593, "arr_del15" : 101, "carrier_ct" : 17.56, "weather_ct" : 20.49, "nas_ct" : 38.26, "security_ct" : 0, "late_aircraft_ct" : 24.69, "arr_cancelled" : 1, "arr_diverted" : 1, "arr_delay" : 5698, "carrier_delay" : 1058, "weather_delay" : 1332, "nas_delay" : 1708, "security_delay" : 0, "late_aircraft_delay" : 1600, "" : "" }
{ "_id" : ObjectId("625d7d8a46c1a4eb66438610"), "year" : 2003, "month" : 6, "carrier" : "AA", "carrier_name" : "Russian Airlines Inc.", "airport" : "CMH", "airport_name" : "Columbus, OH: Port Columbus International", "arr_flights" : 231, "arr_del15" : 29, "carrier_ct" : 8.6, "weather_ct" : 5.12, "nas_ct" : 12.65, "security_ct" : 0, "late_aircraft_ct" : 2.63, "arr_cancelled" : 1, "arr_diverted" : 1, "arr_delay" : 1314, "carrier_delay" : 292, "weather_delay" : 302, "nas_delay" : 522, "security_delay" : 0, "late_aircraft_delay" : 198, "" : "" }
{ "_id" : ObjectId("625d7d8a46c1a4eb66438613"), "year" : 2003, "month" : 6, "carrier" : "AA", "carrier_name" : "Russian Airlines Inc.", "airport" : "DAY", "airport_name" : "Dayton, OH: James M Cox/Dayton International", "arr_flights" : 59, "arr_del15" : 14, "carrier_ct" : 5.28, "weather_ct" : 2.44, "nas_ct" : 2.15, "security_ct" : 0, "late_aircraft_ct" : 4.13, "arr_cancelled" : 1, "arr_diverted" : 0, "arr_delay" : 805, "carrier_delay" : 326, "weather_delay" : 223, "nas_delay" : 57, "security_delay" : 0, "late_aircraft_delay" : 199, "" : "" }
```

5. To find how many flights are there in carrier by using Count:

- db.delayData.count({carrier:"AA"});

```
> db.delayData.count({carrier:"AA"});  
9775
```

CASESTUDY: AIRPORTS ANALYTICS

Analysis with MongoDB

19CSE357 | Group 11

BIGDATA ANALYTICS

RAVELLA ABHINAV



CB.EN.U4CSE19453

1. Find the airports that which flights count in between 500 and 1000?

```
> db.delayData.findOne({arr_flights : {$lt :1000, $gt : 500}})  
{  
  "_id" : ObjectId("625d6adf78e0b10c9e8d4055"),  
  "year" : 2003,  
  "month" : 6,  
  "carrier" : "AA",  
  "carrier_name" : "American Airlines Inc.",  
  "airport" : "ATL",  
  "airport_name" : "Atlanta, GA: Hartsfield-Jackson Atlanta International",  
  "arr_flights" : 752,  
  "arr_del15" : 186,  
  "carrier_ct" : 33.99,  
  "weather_ct" : 27.82,  
  "nas_ct" : 104.76,  
  "security_ct" : 1.9,  
  "late_aircraft_ct" : 17.53,  
  "arr_cancelled" : 5,  
  "arr_diverted" : 0,  
  "arr_delay" : 8314,  
  "carrier_delay" : 1367,  
  "weather_delay" : 1722,  
  "nas_delay" : 3817,  
  "security_delay" : 139,  
  "late_aircraft_delay" : 1269,
```

Query:

- db.delayData.find({arr_flights : {\$lt :1000, \$gt : 500}})

To find one document:

- db.delayData.findOne({arr_flights : {\$lt :1000, \$gt : 500}})

2. Find the airline name that starts with "Continental".

```
db.delayData.find({carrier_name:/^Continental/}).limit(10).pretty();

  "_id" : ObjectId("625d6adf78e0b10c9e8d40ee"),
  "year" : 2003,
  "month" : 6,
  "carrier" : "CO",
  "carrier_name" : "Continental Air Lines Inc.",
  "airport" : "ABQ",
  "airport_name" : "Albuquerque, NM: Albuquerque International Sunport",
  "arr_flights" : 120,
  "arr_del15" : 22,
  "carrier_ct" : 5.57,
  "weather_ct" : 0.51,
  "nas_ct" : 10.04,
  "security_ct" : 0,
  "late_aircraft_ct" : 5.88,
  "arr_cancelled" : 0,
  "arr_diverted" : 0,
  "arr_delay" : 1277,
  "carrier_delay" : 279,
  "weather_delay" : 29,
  "nas_delay" : 544,
```

Query:

- **db.delayData.find({carrier_name:/^Continental/}).limit(10).pretty();**

3. Find the count of flights that are arriving for airports.

```
> db.delayData.aggregate([{$group : {_id : "$airport", Fights : {$sum : "$arr_flights"}}}])  
{ "_id" : "PPG", "Fights" : 60 }  
{ "_id" : "CIU", "Fights" : 391 }  
{ "_id" : "SHD", "Fights" : 60 }  
{ "_id" : "AZA", "Fights" : 131 }  
{ "_id" : "ISN", "Fights" : 1463 }  
{ "_id" : "ESC", "Fights" : 626 }  
{ "_id" : "BRD", "Fights" : 697 }  
{ "_id" : "HOB", "Fights" : 1257 }  
{ "_id" : "GRI", "Fights" : 1454 }  
{ "_id" : "MVY", "Fights" : 343 }  
{ "_id" : "GUM", "Fights" : 958 }  
{ "_id" : "ABR", "Fights" : 1042 }  
{ "_id" : "UTM", "Fights" : 208 }  
{ "_id" : "MHK", "Fights" : 5581 }  
{ "_id" : "MWH", "Fights" : 675 }  
{ "_id" : "SAF", "Fights" : 5175 }  
{ "_id" : "ITH", "Fights" : 1995 }  
{ "_id" : "OTH", "Fights" : 5561 }  
{ "_id" : "LMT", "Fights" : 5674 }  
{ "_id" : "YKM", "Fights" : 747 }  
Type "it" for more  
> █
```

Query:

- db.delayData.aggregate([{\$group : {_id : "\$airport", Fights : {\$sum : "\$arr_flights"}}}, {"\$sort": {"Fights: 1}}])

4. Find the carrier names that belongs to ANC and IND airports

```
{  
    "_id" : ObjectId("625d6adf78e0b10c9e8d429b"),  
    "year" : 2003,  
    "month" : 6,  
    "carrier" : "HP",  
    "carrier_name" : "America West Airlines Inc.",  
    "airport" : "IND",  
    "airport_name" : "Indianapolis, IN: Indianapolis International",  
    "arr_flights" : 60,  
    "arr_del15" : 9,  
    "carrier_ct" : 3.61,  
    "weather_ct" : 0,  
    "nas_ct" : 3.69,  
    "security_ct" : 0,  
    "late_aircraft_ct" : 1.7,  
    "arr_cancelled" : 0,  
    "arr_diverted" : 0,  
    "arr_delay" : 272,  
    "carrier_delay" : 135,  
    "weather_delay" : 0,  
    "nas_delay" : 72,  
    "security_delay" : 0,  
    "late_aircraft_delay" : 65,  
    "" : ""  
}
```

Query:

- `db.delayData.find({airport: {$in:['ANC','IND']}}).limit(10).pretty();`

5. Find the total security of IND airport with airlines having security delays

```
> db.delayData.aggregate({$match:{airport:"IND"}},{$group:{_id:"$carrier",Total:{$sum:"$security_delay"}}});  
{ "_id" : "XE", "Total" : 942 }  
{ "_id" : "F9", "Total" : 0 }  
{ "_id" : "FL", "Total" : 0 }  
{ "_id" : "OO", "Total" : 610 }  
{ "_id" : "OH", "Total" : 474 }  
{ "_id" : "AA", "Total" : 432 }  
{ "_id" : "MQ", "Total" : 24 }  
{ "_id" : "CO", "Total" : 1358 }  
{ "_id" : "UA", "Total" : 0 }  
{ "_id" : "EV", "Total" : 452 }  
{ "_id" : "DL", "Total" : 80 }  
{ "_id" : "NW", "Total" : 1470 }  
{ "_id" : "HP", "Total" : 320 }  
{ "_id" : "YV", "Total" : 104 }  
{ "_id" : "RU", "Total" : 346 }  
{ "_id" : "DH", "Total" : 140 }  
{ "_id" : "TZ", "Total" : 626 }  
{ "_id" : "9E", "Total" : 206 }  
{ "_id" : "US", "Total" : 612 }  
{ "_id" : "WN", "Total" : 2786 }  
> █
```

Query:

- **db.delayData.aggregate({\$match:{airport:"IND"}},{\$group:{_id:"\$carrier",Total:{\$sum:"\$security_delay"}}});**

6. Find the number of flight diversions that occur in St Louis

- var m1 = function(){
 var rval = { airport: this.airport_name,
 year: this.year };
 if(this.airport=="STL") {
 emit(rval, this.arr_diverted);
 }
};
- var r1 = function(key, valArr) {
 return Array.sum(valArr);
}
- db.delayData.mapReduce(m1, r1, {out:'tmp'});
- db.tmp.find().sort({'_id.year':-1})

MapReduce Output

```
> var m1 = function(){
...   var rval = { airport: this.airport_name,
...             year: this.year };
...   if( this.airport=="STL" ) {
...     emit( rval, this.arr_diverted );
...   }
... };
>
> var r1 = function( key, valArr ) {
...   return Array.sum( valArr );
... }
>
> db.delayData.mapReduce( m1, r1, {out:'tmp'});
{
  "result" : "tmp",
  "timeMillis" : 751,
  "counts" : {
    "input" : 173657,
    "emit" : 1601,
    "reduce" : 135,
    "output" : 11
  },
  "ok" : 1
}
```

```
> db.tmp.find().sort( {'_id.year':-1} )
[{"_id": {"airport": "St. Louis, MO: Lambert-St. Louis International", "year": 2013}, "value": 94},
 {"_id": {"airport": "St. Louis, MO: Lambert-St. Louis International", "year": 2012}, "value": 100},
 {"_id": {"airport": "St. Louis, MO: Lambert-St. Louis International", "year": 2011}, "value": 150},
 {"_id": {"airport": "St. Louis, MO: Lambert-St. Louis International", "year": 2010}, "value": 85},
 {"_id": {"airport": "St. Louis, MO: Lambert-St. Louis International", "year": 2009}, "value": 83},
 {"_id": {"airport": "St. Louis, MO: Lambert-St. Louis International", "year": 2008}, "value": 165},
 {"_id": {"airport": "St. Louis, MO: Lambert-St. Louis International", "year": 2007}, "value": 144},
 {"_id": {"airport": "St. Louis, MO: Lambert-St. Louis International", "year": 2006}, "value": 119},
 {"_id": {"airport": "St. Louis, MO: Lambert-St. Louis International", "year": 2005}, "value": 116},
 {"_id": {"airport": "St. Louis, MO: Lambert-St. Louis International", "year": 2004}, "value": 135},
 {"_id": {"airport": "St. Louis, MO: Lambert-St. Louis International", "year": 2003}, "value": 109}]
```

MongoDB

CB.EN.U4CSE19459

SHANTHAN REDDY

Dataset

Label/field name	Meaning
year	The year the month occurred
month	The month for which data was collected
carrier	The airline code
carrier_name	The airline name
airport	The airport code
airport_name	The airport name
arr_flights	# of flights that arrived at the airport
arr_del15	# of flights that arrived >= 15 minutes late
carrier_ct	# of flights delayed due to the carrier
weather_ct	# of flights delayed due to weather
nas_ct	# of flights delayed due to national air system
security_ct	# of flights delayed due to security
late_aircraft_ct	# flights delayed because a previous flight using the same aircraft was late
arr_cancelled	# of canceled arrivals
arr_diverted	# of scheduled arrivals that were diverted
arr_delay	Sum of the delay minutes
carrier_delay	Total minutes of delays due to carriers
weather_delay	Total minutes of delays due to weather
nas_delay	Total minutes of delays due to natl. air service
security_delay	Total minutes of delays due to security
late_aircraft_delay	Similar to late_aircraft_ct. The total minutes of delay due to a previous flight using the same aircraft arriving late.

+

```
"carrier" : "AA",
"carrier_name" : "American Airlines Inc.",
"airport" : "ABQ",
"airport_name" : "Albuquerque, NM: Albuquerque International Sunport",
"arr_flights" : 307,
"arr_del15" : 56,
"carrier_ct" : 14.68,
"weather_ct" : 10.79,
"nas_ct" : 19.09,
"security_ct" : 1.48,
"late_aircraft_ct" : 9.96,
"arr_cancelled" : 1,
"arr_diverted" : 1,
"arr_delay" : 2530,
"carrier_delay" : 510,
"weather_delay" : 621,
"nas_delay" : 676,
"security_delay" : 25,
"late_aircraft_delay" : 698,
"" : ""

"_id" : ObjectId("625d7ad10c34204c60928a68"),
"year" : 2003,
"month" : 6,
"carrier" : "AA",
"carrier_name" : "American Airlines Inc.",
"airport" : "ANC",
"airport_name" : "Anchorage, AK: Ted Stevens Anchorage International",
"arr_flights" : 90,
"arr_del15" : 27,
"carrier_ct" : 7.09,
"weather_ct" : 2,
"nas_ct" : 10.75,
"security_ct" : 0,
"late_aircraft_ct" : 7.16,
"arr_cancelled" : 0,
"arr_diverted" : 0,
"arr_delay" : 1390,
"carrier_delay" : 271,
"weather_delay" : 83,
"nas_delay" : 581,
"security_delay" : 0,
"late_aircraft_delay" : 455,
"" : ""
```

QUERY 1:

```
db.delayData.find({carrier_name:"American Airlines Inc."}).pretty();
```

QUERY 2:

```
var m1 = function(){  
...   emit( this.airport_name, this.arr_diverted );  
... };  
  
var r1 = function( key, valArr ) {  
...   return Array.sum( valArr );  
... }  
  
db.delayData.mapReduce( m1, r1,  
{out:'tmp'});
```

```
"late_aircraft_delay" : 698,  
"" : ""  
}  
{  
  "_id" : ObjectId("625d7ad10c34204c60928a68"),  
  "year" : 2003,  
  "month" : 6,  
  "carrier" : "AA",  
  "carrier_name" : "American Airlines Inc.",  
  "airport" : "ANC",  
  "airport_name" : "Anchorage, AK: Ted Stevens Anchorage International",  
  "arr_flights" : 90,  
  "arr_del15" : 27,  
  "carrier_ct" : 7.09,  
  "weather_ct" : 2,  
  "nas_ct" : 10.75,  
  "security_ct" : 0,  
  "late_aircraft_ct" : 7.16,  
  "arr_cancelled" : 0,  
  "arr_diverted" : 0,  
  "arr_delay" : 1390,  
  "carrier_delay" : 271,  
  "weather_delay" : 83,  
  "nas_delay" : 581,  
  "security_delay" : 0,  
  "late_aircraft_delay" : 455,  
  "" : ""  
}  
Type "it" for more  
> var m1 = function(){  
...   ...   emit( this.airport_name, this.arr_diverted );  
... };  
> var r1 = function( key, valArr ) {  
...   ...   return Array.sum( valArr );  
... };  
> db.delayData.mapReduce( m1, r1, {out:'tmp'});  
{  
  "result" : "tmp",  
  "timeMillis" : 900,  
  "counts" : {  
    "input" : 173657,  
    "emit" : 173657,  
    "reduce" : 16838,  
    "output" : 363  
  },  
  "ok" : 1  
}> |
```

```
> db.tmp.find().sort( {value:1} )
{ "_id" : "Clarksburg/Fairmont, WV: North Central West Virginia", "value" : 0 }
{ "_id" : "Columbus, MS: Columbus AFB", "value" : 0 }
{ "_id" : "Dickinson, ND: Dickinson - Theodore Roosevelt Regional", "value" : 0 }
{ "_id" : "Greenville, MS: Mid Delta Regional", "value" : 0 }
{ "_id" : "Guam, TT: Guam International", "value" : 0 }
{ "_id" : "Gustavus, AK: Gustavus Airport", "value" : 0 }
{ "_id" : "Iron Mountain/Kingsfd, MI: Ford", "value" : 0 }
{ "_id" : "Kansas City, MO: Charles B. Wheeler Downtown", "value" : 0 }
{ "_id" : "Moses Lake, WA: Grant County International", "value" : 0 }
{ "_id" : "Pago Pago, TT: Pago Pago International", "value" : 0 }
{ "_id" : "Palmdale, CA: Palmdale USAF Plant 42", "value" : 0 }
{ "_id" : "Phoenix, AZ: Phoenix - Mesa Gateway", "value" : 0 }
{ "_id" : "Saipan, TT: Francisco C. Ada Saipan International", "value" : 0 }
{ "_id" : "Salem, OR: McNary Field", "value" : 0 }
{ "_id" : "Staunton, VA: Shenandoah Valley Regional", "value" : 0 }
{ "_id" : "Visalia, CA: Visalia Municipal", "value" : 0 }
{ "_id" : "Yakima, WA: Yakima Air Terminal/McAllister Field", "value" : 0 }
{ "_id" : "Del Rio, TX: Del Rio International", "value" : 1 }
{ "_id" : "Houston, TX: Ellington", "value" : 1 }
```

QUERY 3:

db.tmp.find().sort({value:1})

QUERY 4:

```
db.delayData.find({arr_cancelled:{$lte:5}}).pretty();
```

```
"carrier" : "AA",
"carrier_name" : "American Airlines Inc.",
"airport" : "ELP",
"airport_name" : "El Paso, TX: El Paso International",
"arr_flights" : 269,
"arr_del15" : 78,
"carrier_ct" : 27.29,
"weather_ct" : 9.68,
"nas_ct" : 27.28,
"security_ct" : 1.38,
"late_aircraft_ct" : 12.38,
"arr_cancelled" : 2,
"arr_diverted" : 0,
"arr_delay" : 3589,
"carrier_delay" : 1113,
"weather_delay" : 564,
"nas_delay" : 982,
"security_delay" : 22,
"late_aircraft_delay" : 908,
"" : ""

{
  "_id" : ObjectId("625d7ad10c34204c60928a6c"),
  "year" : 2003,
  "month" : 6,
  "carrier" : "AA",
  "carrier_name" : "American Airlines Inc.",
  "airport" : "FAT",
  "airport_name" : "Fresno, CA: Fresno Yosemite International",
  "arr_flights" : 77,
  "arr_del15" : 21,
  "carrier_ct" : 4.09,
  "weather_ct" : 3.53,
  "nas_ct" : 7.84,
  "security_ct" : 0,
  "late_aircraft_ct" : 5.54,
  "arr_cancelled" : 0,
  "arr_diverted" : 0,
  "arr_delay" : 906,
  "carrier_delay" : 337,
  "weather_delay" : 99,
  "nas_delay" : 216,
  "security_delay" : 0,
  "late_aircraft_delay" : 254,
  "" : ""
}
```

```
"month" : 2,
"carrier" : "TZ",
"carrier_name" : "ATA Airlines d/b/a ATA",
"airport" : "TPA",
"airport_name" : "Tampa, FL: Tampa International",
"arr_flights" : 0,
"arr_del15" : 0,
"carrier_ct" : 0,
"weather_ct" : 0,
"nas_ct" : 0,
"security_ct" : 0,
"late_aircraft_ct" : 0,
"arr_cancelled" : 0,
"arr_diverted" : 0,
"arr_delay" : 0,
"carrier_delay" : 0,
"weather_delay" : 0,
"nas_delay" : 0,
"security_delay" : 0,
"late_aircraft_delay" : 0,
"" : ""

{
  "_id" : ObjectId("625d7ad10c34204c6092b910"),
  "year" : 2004,
  "month" : 3,
  "carrier" : "EV",
  "carrier_name" : "Atlantic Southeast Airlines",
  "airport" : "BNA",
  "airport_name" : "Nashville, TN: Nashville International",
  "arr_flights" : 0,
  "arr_del15" : 0,
  "carrier_ct" : 0,
  "weather_ct" : 0,
  "nas_ct" : 0,
  "security_ct" : 0,
  "late_aircraft_ct" : 0,
  "arr_cancelled" : 0,
  "arr_diverted" : 0,
  "arr_delay" : 0,
  "carrier_delay" : 0,
  "weather_delay" : 0,
  "nas_delay" : 0,
  "security_delay" : 0,
  "late_aircraft_delay" : 0,
}
```

QUERY 5:

```
db.delayData.find().sort({arr_flights:1}).pretty().limit(10);
```

MongoDB Review

CASESTUDY

19CSE357
BIGDATA ANALYTICS

CSE - E

DIVESH KOSURI

12-04-2022

CB.EN.U4CSE19422

• • •

FINDING THE MINIMUM AND MAXIMUM YEAR RECORDED IN THE DATABASE

QUERY 1: db.delayData.find().sort({year:1}).limit(1).pretty();

```
> db.delayData.find().sort({year:1}).limit(1).pretty();
{
    "_id" : ObjectId("625d88a82c8be4ce18263bf5"),
    "year" : 2003,
    "month" : 6,
    "carrier" : "AA",
    "carrier_name" : "American Airlines Inc.",
    "airport" : "ABQ",
    "airport_name" : "Albuquerque, NM: Albuquerque International Sunport",
    "arr_flights" : 307,
    "arr_del15" : 56,
    "carrier_ct" : 14.68,
    "weather_ct" : 10.79,
    "nas_ct" : 19.09,
    "security_ct" : 1.48,
    "late_aircraft_ct" : 9.96,
    "arr_cancelled" : 1,
    "arr_diverted" : 1,
    "arr_delay" : 2530,
    "carrier_delay" : 510,
    "weather_delay" : 621,
    "nas_delay" : 676,
    "security_delay" : 25,
    "late_aircraft_delay" : 698,
    "" : ""
}
```

QUERY 2: db.delayData.find().sort({year:-1}).limit(1).pretty();

```
> db.delayData.find().sort({year:-1}).limit(1).pretty();
{
    "_id" : ObjectId("625d88cb2c8be4ce1828be69"),
    "year" : 2013,
    "month" : 1,
    "carrier" : "9E",
    "carrier_name" : "Pinnacle Airlines Inc.",
    "airport" : "DFW",
    "airport_name" : "Dallas/Fort Worth, TX: Dallas/Fort Worth International",
    "arr_flights" : 139,
    "arr_del15" : 37,
    "carrier_ct" : 13.04,
    "weather_ct" : 0.8,
    "nas_ct" : 13.92,
    "security_ct" : 0,
    "late_aircraft_ct" : 9.24,
    "arr_cancelled" : 2,
    "arr_diverted" : 5,
    "arr_delay" : 1860,
    "carrier_delay" : 789,
    "weather_delay" : 75,
    "nas_delay" : 451,
    "security_delay" : 0,
    "late_aircraft_delay" : 545,
    "" : ""
}
```

FINDING THE YEARS OF SERVICE [AGGREGATE FUNTIONS]

QUERY 3: db.delayData.aggregate([{\$group : {_id : "\$carrier", MinYear : {\$min : "\$year"}}}]);

```
> db.delayData.aggregate([{$group : {_id : "$carrier", MinYear : {$min : "$year"}}}]);
{ "_id" : "VX", "MinYear" : 2012 }
{ "_id" : "XE", "MinYear" : 2006 }
{ "_id" : "AQ", "MinYear" : 2006 }
{ "_id" : "F9", "MinYear" : 2005 }
{ "_id" : "OH", "MinYear" : 2004 }
{ "_id" : "HA", "MinYear" : 2003 }
{ "_id" : "9E", "MinYear" : 2007 }
{ "_id" : "WN", "MinYear" : 2003 }
{ "_id" : "US", "MinYear" : 2003 }
{ "_id" : "AA", "MinYear" : 2003 }
{ "_id" : "AS", "MinYear" : 2003 }
{ "_id" : "TZ", "MinYear" : 2003 }
{ "_id" : "DH", "MinYear" : 2003 }
{ "_id" : "B6", "MinYear" : 2003 }
{ "_id" : "UA", "MinYear" : 2003 }
{ "_id" : "CO", "MinYear" : 2003 }
{ "_id" : "MQ", "MinYear" : 2003 }
{ "_id" : "EV", "MinYear" : 2003 }
{ "_id" : "DL", "MinYear" : 2003 }
{ "_id" : "NW", "MinYear" : 2003 }
Type "it" for more
```

QUERY 4: db.delayData.aggregate([{\$group : {_id : "\$carrier", MaxYear : {\$max : "\$year"}}}]);

```
> db.delayData.aggregate([{$group : {_id : "$carrier", MaxYear : {$max : "$year"}}}]);
{ "_id" : "VX", "MaxYear" : 2013 }
{ "_id" : "XE", "MaxYear" : 2011 }
{ "_id" : "AQ", "MaxYear" : 2008 }
{ "_id" : "F9", "MaxYear" : 2013 }
{ "_id" : "OH", "MaxYear" : 2010 }
{ "_id" : "HA", "MaxYear" : 2013 }
{ "_id" : "9E", "MaxYear" : 2013 }
{ "_id" : "WN", "MaxYear" : 2013 }
{ "_id" : "US", "MaxYear" : 2013 }
{ "_id" : "AA", "MaxYear" : 2013 }
{ "_id" : "AS", "MaxYear" : 2013 }
{ "_id" : "TZ", "MaxYear" : 2006 }
{ "_id" : "DH", "MaxYear" : 2005 }
{ "_id" : "B6", "MaxYear" : 2013 }
{ "_id" : "UA", "MaxYear" : 2013 }
{ "_id" : "CO", "MaxYear" : 2011 }
{ "_id" : "MQ", "MaxYear" : 2013 }
{ "_id" : "EV", "MaxYear" : 2013 }
{ "_id" : "DL", "MaxYear" : 2013 }
{ "_id" : "NW", "MaxYear" : 2009 }
Type "it" for more
```

• • •

TO FIND THE TOTAL DELAY OCCURRED DUE TO CARRIER LOAD-UNLOAD

Query 5: db.delayData.aggregate([{\$group : {_id : "\$carrier", TotalDelay: {\$sum : "\$carrier_ct"}}}]);

```
> db.delayData.aggregate([{$group : {_id : "$carrier", TotalDelay: {$sum : "$carrier_ct"}}}]);
{ "_id" : "VX", "TotalDelay" : 2740.08 }
{ "_id" : "XE", "TotalDelay" : 102311.54 }
{ "_id" : "AQ", "TotalDelay" : 3257.98 }
{ "_id" : "F9", "TotalDelay" : 37074.63 }
{ "_id" : "OH", "TotalDelay" : 146193.59 }
{ "_id" : "HA", "TotalDelay" : 27486.12 }
{ "_id" : "9E", "TotalDelay" : 65592.78 }
{ "_id" : "WN", "TotalDelay" : 523048.34 }
{ "_id" : "US", "TotalDelay" : 229268.93 }
{ "_id" : "AA", "TotalDelay" : 367637.12 }
{ "_id" : "AS", "TotalDelay" : 88999.15 }
{ "_id" : "TZ", "TotalDelay" : 6563.67 }
{ "_id" : "DH", "TotalDelay" : 26079.83 }
{ "_id" : "B6", "TotalDelay" : 103185.42 }
{ "_id" : "UA", "TotalDelay" : 243557.23 }
{ "_id" : "CO", "TotalDelay" : 112686.43 }
{ "_id" : "MQ", "TotalDelay" : 266857.85 }
{ "_id" : "EV", "TotalDelay" : 294554.02 }
{ "_id" : "DL", "TotalDelay" : 312964.66 }
{ "_id" : "NW", "TotalDelay" : 198141.92 }
```

INSERTING NEW RECORD

QUERY 6: db.delayDB.insert([{"year" : 2022, "month" : 4, "carrier" : "AAA", "carrier_name" : "Divesh Airlines Inc.", "airport" : "DSM", "airport_name" : "Divesh International", "arr_flights" : 422, "arr_del15" : 15, "carrier_ct" : 3.39, "weather_ct" : 1.57, "nas_ct" : 5.07, "security_ct" : 0, "late_aircraft_ct" : 6.97, "arr_cancelled" : 0, "arr_diverted" : 0, "arr_delay" : 0, "carrier_delay" : 212, "weather_delay" : 50, "nas_delay" : 259, "security_delay" : 0, "late_aircraft_delay" : 345, "" : ""}]);

```
> db.delayDB.insert([{"year" : 2022, "month" : 4, "carrier" : "AAA", "carrier_name" : "Divesh Airlines Inc.", "airport" : "DSM", "airport_name" : "Divesh International", "arr_flights" : 422, "arr_del15" : 15, "carrier_ct" : 3.39, "weather_ct" : 1.57, "nas_ct" : 5.07, "security_ct" : 0, "late_aircraft_ct" : 6.97, "arr_cancelled" : 0, "arr_diverted" : 0, "arr_delay" : 0, "carrier_delay" : 212, "weather_delay" : 50, "nas_delay" : 259, "security_delay" : 0, "late_aircraft_delay" : 345, "" : ""}]);
BulkWriteResult({
  "writeErrors" : [ ],
  "writeConcernErrors" : [ ],
  "nInserted" : 1,
  "nUpserted" : 0,
  "nMatched" : 0,
  "nModified" : 0,
  "nRemoved" : 0,
  "upserted" : [ ]
})
```

MAP-REDUCE QUERY 1

QUERY 7: Defining Mapper and Reducer

- var mapCheck1 = function(){ emit(this.carrier_name, this.carrier_delay);}
- var reduceCheck1 = function(key, valArr) {return Array.sum(valArr);}
- db.delayData.mapReduce(mapCheck1, reduceCheck1,{out:"outputMR"});

```
> var mapCheck1 = function(){ emit( this.carrier_name, this.carrier_delay );}
> var reduceCheck1 = function( key, valArr ) {return Array.sum( valArr );}
> db.delayData.mapReduce( mapCheck1, reduceCheck1,{out:"outputMR"});
{
  "result" : "outputMR",
  "timeMillis" : 2446,
  "counts" : {
    "input" : 173657,
    "emit" : 173657,
    "reduce" : 4272,
    "output" : 25
  },
  "ok" : 1
}
```

QUERY 8: Loading the output with sorted order according to the requirement:

```
> db.outputMR.find().sort({year:-1});
{ "_id" : "ATA Airlines d/b/a ATA", "value" : 405504 }
{ "_id" : "AirTran Airways Corporation", "value" : 4028057 }
{ "_id" : "Alaska Airlines Inc.", "value" : 5062626 }
{ "_id" : "Aloha Airlines Inc.", "value" : 144922 }
{ "_id" : "America West Airlines Inc.", "value" : 1699448 }
{ "_id" : "American Airlines Inc.", "value" : 22870499 }
{ "_id" : "American Eagle Airlines Inc.", "value" : 16095940 }
{ "_id" : "Atlantic Coast Airlines", "value" : 1064001 }
{ "_id" : "Atlantic Southeast Airlines", "value" : 13837566 }
{ "_id" : "Comair Inc.", "value" : 8415020 }
{ "_id" : "Continental Air Lines Inc.", "value" : 6480851 }
{ "_id" : "Delta Air Lines Inc.", "value" : 19040800 }
{ "_id" : "ExpressJet Airlines Inc.", "value" : 14492140 }
{ "_id" : "Frontier Airlines Inc.", "value" : 1721862 }
{ "_id" : "Hawaiian Airlines Inc.", "value" : 1355530 }
{ "_id" : "Independence Air", "value" : 304026 }
{ "_id" : "JetBlue Airways", "value" : 6512832 }
{ "_id" : "Mesa Airlines Inc.", "value" : 8681157 }
{ "_id" : "Northwest Airlines Inc.", "value" : 11745027 }
{ "_id" : "Pinnacle Airlines Inc.", "value" : 4365161 }
Type "it" for more
> it
{ "_id" : "SkyWest Airlines Inc.", "value" : 17016279 }
{ "_id" : "Southwest Airlines Co.", "value" : 22184643 }
{ "_id" : "US Airways Inc.", "value" : 11378650 }
{ "_id" : "United Air Lines Inc.", "value" : 14375218 }
{ "_id" : "Virgin America", "value" : 154025 }
```

• • •

MAP-REDUCE QUERY 2

QUERY 9: Defining Mapper and Reducer

```
var mapCheck2 = function(){ emit( this.carrier_name, this.weather_delay );}
var reduceCheck2 = function( key, valArr ) {return Array.sum( valArr );}
db.delayData.mapReduce( mapCheck2, reduceCheck2, {out:"NewOutputMR"});
```

```
> var mapCheck2 = function(){ emit( this.carrier_name, this.weather_delay );}
> var reduceCheck2 = function( key, valArr ) {return Array.sum( valArr );}
> db.delayData.mapReduce( mapCheck2, reduceCheck2, {out:"NewOutputMR"} );
{
  "result" : "NewOutputMR",
  "timeMillis" : 992,
  "counts" : {
    "input" : 173657,
    "emit" : 173657,
    "reduce" : 4272,
    "output" : 25
  },
  "ok" : 1
}
```

QUERY 10: Loading the output with sorted order according to the requirement:

```
> db.NewOutputMR.find().sort({year:-1})
{ "_id" : "ATA Airlines d/b/a ATA", "value" : 23136 }
{ "_id" : "AirTran Airways Corporation", "value" : 306101 }
{ "_id" : "Alaska Airlines Inc.", "value" : 345748 }
{ "_id" : "Aloha Airlines Inc.", "value" : 2297 }
{ "_id" : "America West Airlines Inc.", "value" : 71713 }
{ "_id" : "American Airlines Inc.", "value" : 5228234 }
{ "_id" : "American Eagle Airlines Inc.", "value" : 4157053 }
{ "_id" : "Atlantic Coast Airlines", "value" : 317277 }
{ "_id" : "Atlantic Southeast Airlines", "value" : 4318712 }
{ "_id" : "Comair Inc.", "value" : 4849900 }
{ "_id" : "Continental Air Lines Inc.", "value" : 1414448 }
{ "_id" : "Delta Air Lines Inc.", "value" : 2215918 }
{ "_id" : "ExpressJet Airlines Inc.", "value" : 2546530 }
{ "_id" : "Frontier Airlines Inc.", "value" : 183900 }
{ "_id" : "Hawaiian Airlines Inc.", "value" : 23615 }
{ "_id" : "Independence Air", "value" : 34186 }
{ "_id" : "JetBlue Airways", "value" : 635836 }
{ "_id" : "Mesa Airlines Inc.", "value" : 952900 }
{ "_id" : "Northwest Airlines Inc.", "value" : 2342139 }
{ "_id" : "Pinnacle Airlines Inc.", "value" : 669742 }
```

BIG DATA MONGODB:

CB.EN.U4CSE19449

P. Sai Koushik

1. Update

```
db.delayData.update({carrier_name : "American Airlines Inc."},{$set:{carrier_name : "Russian Airlines Inc."}}, {multi:true})
```

```
> db.delayData.update({carrier_name : "American Airlines Inc."},{$set:{carrier_name : "Russian Airlines Inc."}}, {multi:true})
WriteResult({ "nMatched" : 9775, "nUpserted" : 0, "nModified" : 9775 })
> db.delayData.find()
[{"_id" : ObjectId("625d7d8a46c1a4eb66438604"), "year" : 2003, "month" : 6, "carrier" : "AA", "carrier_name" : "Russian Airlines Inc.", "airport" : "ABQ", "airport_name" : "Albuquerque, NM: Albuquerque International Sunport", "arr_flights" : 307, "arr_del15" : 56, "carrier_ct" : 14.68, "weather_ct" : 10.79, "nas_ct" : 19.09, "security_ct" : 1.48, "late_aircraft_ct" : 9.96, "arr_cancelled" : 1, "arr_diverted" : 1, "arr_delay" : 2530, "carrier_delay" : 510, "weather_delay" : 621, "nas_delay" : 676, "security_delay" : 25, "late_aircraft_delay" : 698, ":" : "" }, {"_id" : ObjectId("625d7d8a46c1a4eb66438605"), "year" : 2003, "month" : 6, "carrier" : "AA", "carrier_name" : "Russian Airlines Inc.", "airport" : "ANC", "airpo rt_name" : "Anchorage, AK: Ted Stevens Anchorage International", "arr_flights" : 98, "arr_del15" : 27, "carrier_ct" : 2, "nas_ct" : 16.75, "security_ct" : 0, "late_aircraft_ct" : 7.16, "arr_cancelled" : 0, "arr_diverted" : 0, "arr_delay" : 1390, "carrier_delay" : 271, "weather_delay" : 83, "nas_delay" : 581, "security_delay" : 0, "late_aircraft_delay" : 455, ":" : "" }, {"_id" : ObjectId("625d7d8a46c1a4eb66438606"), "year" : 2003, "month" : 6, "carrier" : "AA", "carrier_name" : "Russian Airlines Inc.", "airport" : "ATL", "airpo rt_name" : "Atlanta, GA: Hartsfield-Jackson Atlanta International", "arr_flights" : 752, "arr_del15" : 186, "carrier_ct" : 33.99, "weather_ct" : 27.82, "nas_ct" : 104.76, "security_ct" : 1.9, "late_aircraft_ct" : 17.53, "arr_cancelled" : 5, "arr_diverted" : 0, "arr_delay" : 8314, "carrier_delay" : 1367, "weather_delay" : 1722, "nas_delay" : 3817, "security_delay" : 139, "late_aircraft_delay" : 1269, ":" : "" }, {"_id" : ObjectId("625d7d8a46c1a4eb66438607"), "year" : 2003, "month" : 6, "carrier" : "AA", "carrier_name" : "Russian Airlines Inc.", "airport" : "AUS", "airpo rt_name" : "Austin, TX: Austin - Bergstrom International", "arr_flights" : 842, "arr_del15" : 174, "carrier_ct" : 60.24, "weather_ct" : 28.54, "nas_ct" : 47.78, "security_ct" : 4.69, "late_aircraft_ct" : 40.75, "arr_cancelled" : 9, "arr_diverted" : 1, "arr_delay" : 8344, "carrier_delay" : 3040, "weather_delay" : 1032, "nas_delay" : 1835, "security_delay" : 115, "late_aircraft_delay" : 2322, ":" : "" }, {"_id" : ObjectId("625d7d8a46c1a4eb66438608"), "year" : 2003, "month" : 6, "carrier" : "AA", "carrier_name" : "Russian Airlines Inc.", "airport" : "BDL", "airpo rt_name" : "Hartford, CT: Bradley International", "arr_flights" : 383, "arr_del15" : 55, "carrier_ct" : 14.9, "weather_ct" : 8.91, "nas_ct" : 14.57, "security_ct" : 0, "late_aircraft_ct" : 16.61, "arr_cancelled" : 0, "arr_diverted" : 0, "arr_delay" : 3137, "carrier_delay" : 515, "weather_delay" : 574, "nas_delay" : 655 }
```

2. Count:

```
db.delayData.count({carrier:"AA"});
```

```
> db.delayData.count({carrier:"AA"});
9775
```

3. Aggregate(Max):

```
db.delayData.aggregate([{$group : {_id : "$airport", number : {$max : "$arr_cancelled"}}}])
```

```
> db.delayData.aggregate([{$group : {_id : "$airport", number : {$max : "$arr_cancelled"}}}])
[{"_id": "PPG", "number": 0},
 {"_id": "CIU", "number": 5},
 {"_id": "SHD", "number": 0},
 {"_id": "AZA", "number": 1},
 {"_id": "ISN", "number": 7},
 {"_id": "ESC", "number": 7},
 {"_id": "BRD", "number": 4},
 {"_id": "HOB", "number": 8},
 {"_id": "GRI", "number": 4},
 {"_id": "MVY", "number": 5},
 {"_id": "GUM", "number": 2},
 {"_id": "ABR", "number": 10},
 {"_id": "UTM", "number": 3},
 {"_id": "MHK", "number": 13},
 {"_id": "MWH", "number": 5},
 {"_id": "SAF", "number": 10},
 {"_id": "ITH", "number": 9},
 {"_id": "OTH", "number": 8},
 {"_id": "LMT", "number": 10},
 {"_id": "YKM", "number": 1}]
Type "it" for more
> ■
```

4. Strings

```
db.delayData.find({airport :/D$/});
```

```
> db.student.find({airport :/D$/});
> db.delayData.find({airport :/D$/});
[{"_id": ObjectId("625d7d8a46cia4eb6643861f"), "year": 2003, "month": 6, "carrier": "AA", "carrier_name": "Russian Airlines Inc.", "airport": "IAD", "airpot_name": "Washington, DC: Washington Dulles International", "arr_flights": 445, "arr_del15": 76, "carrier_ct": 17.58, "weather_ct": 15.86, "nas_ct": 32.62, "security_ct": 6, "late_aircraft_ct": 9.94, "arr_cancelled": 2, "arr_diverted": 3, "arr_delay": 3661, "carrier_delay": 997, "weather_delay": 759, "nas_delay": 1225, "security_delay": 0, "late_aircraft_delay": 680, ":" : "" },
 {"_id": ObjectId("625d7d8a46cia4eb66438621"), "year": 2003, "month": 6, "carrier": "AA", "carrier_name": "Russian Airlines Inc.", "airport": "IND", "airpot_name": "Indianapolis, IN: Indianapolis International", "arr_flights": 278, "arr_del15": 58, "carrier_ct": 17.92, "weather_ct": 5.17, "nas_ct": 14.94, "security_ct": 1, "late_aircraft_ct": 10.97, "arr_cancelled": 3, "arr_diverted": 0, "arr_delay": 2861, "carrier_delay": 758, "weather_delay": 800, "nas_delay": 743, "security_delay": 46, "late_aircraft_delay": 514, ":" : "" },
 {"_id": ObjectId("625d7d8a46cia4eb6643863c"), "year": 2003, "month": 6, "carrier": "AA", "carrier_name": "Russian Airlines Inc.", "airport": "ORD", "airpot_name": "Chicago, IL: Chicago O'Hare International", "arr_flights": 7628, "arr_del15": 1389, "carrier_ct": 225.53, "weather_ct": 48.83, "nas_ct": 763.26, "security_ct": 9.15, "late_aircraft_ct": 262.23, "arr_cancelled": 41, "arr_diverted": 14, "arr_delay": 72172, "carrier_delay": 19608, "weather_delay": 3440, "nas_delay": 34502, "security_delay": 403, "late_aircraft_delay": 14219, ":" : "" },
 {"_id": ObjectId("625d7d8a46cia4eb66438641"), "year": 2003, "month": 6, "carrier": "AA", "carrier_name": "Russian Airlines Inc.", "airport": "PVD", "airpot_name": "Providence, RI: Theodore Francis Green State", "arr_flights": 118, "arr_del15": 28, "carrier_ct": 6.15, "weather_ct": 0.75, "nas_ct": 13.51, "security_ct": 0, "late_aircraft_ct": 7.59, "arr_cancelled": 0, "arr_diverted": 0, "arr_delay": 1216, "carrier_delay": 274, "weather_delay": 64, "nas_delay": 424, "security_delay": 0, "late_aircraft_delay": 454, ":" : "" },
 {"_id": ObjectId("625d7d8a46cia4eb6643866c"), "year": 2003, "month": 6, "carrier": "AS", "carrier_name": "Alaska Airlines Inc.", "airport": "IAD", "airpot_name": "Washington, DC: Washington Dulles International", "arr_flights": 66, "arr_del15": 14, "carrier_ct": 4.79, "weather_ct": 0, "nas_ct": 7.77, "security_ct": 0, "late_aircraft_ct": 0, "arr_cancelled": 0, "arr_diverted": 0, "arr_delay": 561, "carrier_delay": 784, "weather_delay": 0, "nas_delay": 0}]
```

5. Skip:

```
db.delayData.find().limit(6).skip(1).sort({security_delay: 1});
```

```
> db.delayData.find().limit(6).skip(1).sort({security_delay: 1});
{
  "_id": ObjectId("625d7d8a46cia4eb66438608"),
  "year": 2003, "month": 6, "carrier": "AA", "carrier_name": "Russian Airlines Inc.", "airport": "BDL", "airpo
rt_name": "Hartford, CT Bradley International", "arr_flights": 383, "arr_del15": 55, "carrier_ct": 14.9, "weather_ct": 8.91, "nas_ct": 14.57, "security_ct
": 0, "late_aircraft_ct": 16.61, "arr_cancelled": 0, "arr_diverted": 0, "arr_delay": 3137, "carrier_delay": 815, "weather_delay": 574, "nas_delay": 555,
  "security_delay": 0, "late_aircraft_delay": 1193, "": "" }
{
  "_id": ObjectId("625d7d8a46cia4eb6643860c"),
  "year": 2003, "month": 6, "carrier": "AA", "carrier_name": "Russian Airlines Inc.", "airport": "BUR", "airpo
rt_name": "Burbank, CA: Bob Hope", "arr_flights": 119, "arr_del15": 27, "carrier_ct": 7.49, "weather_ct": 4.65, "nas_ct": 7.81, "security_ct": 0, "late_a
irrcraft_ct": 7.04, "arr_cancelled": 1, "arr_diverted": 0, "arr_delay": 1187, "carrier_delay": 326, "weather_delay": 191, "nas_delay": 285, "security_delay"
: 0, "late_aircraft_delay": 385, "": "" }
{
  "_id": ObjectId("625d7d8a46cia4eb6643860d"),
  "year": 2003, "month": 6, "carrier": "AA", "carrier_name": "Russian Airlines Inc.", "airport": "CLE", "airpo
rt_name": "Cleveland, OH: Cleveland-Hopkins International", "arr_flights": 172, "arr_del15": 29, "carrier_ct": 5.51, "weather_ct": 3.21, "nas_ct": 11.76,
  "security_ct": 0, "late_aircraft_ct": 8.51, "arr_cancelled": 0, "arr_diverted": 0, "arr_delay": 2037, "carrier_delay": 598, "weather_delay": 250, "nas_de
lay": 418, "security_delay": 0, "late_aircraft_delay": 771, "": "" }
{
  "_id": ObjectId("625d7d8a46cia4eb6643860e"),
  "year": 2003, "month": 6, "carrier": "AA", "carrier_name": "Russian Airlines Inc.", "airport": "BWI", "airpo
rt_name": "Baltimore, MD: Baltimore/Washington International Thurgood Marshall", "arr_flights": 593, "arr_del15": 101, "carrier_ct": 17.56, "weather_ct": 20
.49, "nas_ct": 38.26, "security_ct": 0, "late_aircraft_ct": 24.69, "arr_cancelled": 1, "arr_diverted": 1, "arr_delay": 5698, "carrier_delay": 1058, "weat
her_delay": 1332, "nas_delay": 1708, "security_delay": 0, "late_aircraft_delay": 1600, "": "" }
{
  "_id": ObjectId("625d7d8a46cia4eb66438610"),
  "year": 2003, "month": 6, "carrier": "AA", "carrier_name": "Russian Airlines Inc.", "airport": "CMH", "airpo
rt_name": "Columbus, OH: Port Columbus International", "arr_flights": 231, "arr_del15": 29, "carrier_ct": 8.6, "weather_ct": 5.12, "nas_ct": 12.65,
  "security_ct": 0, "late_aircraft_ct": 2.63, "arr_cancelled": 1, "arr_diverted": 1, "arr_delay": 1314, "carrier_delay": 292, "weather_delay": 302, "nas_delay": 22
, "security_delay": 0, "late_aircraft_delay": 198, "": "" }
{
  "_id": ObjectId("625d7d8a46cia4eb66438613"),
  "year": 2003, "month": 6, "carrier": "AA", "carrier_name": "Russian Airlines Inc.", "airport": "DAY", "airpo
rt_name": "Dayton, OH: James M Cox/Dayton International", "arr_flights": 59, "arr_del15": 14, "carrier_ct": 5.28, "weather_ct": 2.44, "nas_ct": 2.15, "secu
rity_ct": 0, "late_aircraft_ct": 4.13, "arr_cancelled": 1, "arr_diverted": 0, "arr_delay": 805, "carrier_delay": 326, "weather_delay": 223, "nas_delay": 57,
  "security_delay": 0, "late_aircraft_delay": 199, "": "" }
```

MongoDB:

Queries:

Dataset loading

```
hduser@abhinav:~$ cd MongoDB/
hduser@abhinav:~/MongoDB$ ls
airlineDelay.csv
hduser@abhinav:~/MongoDB$ mongoimport -d userDB -c delayData --type csv --file airlineDelay.csv --headerline
2022-04-18T19:12:55.282+0530      connected to: localhost
2022-04-18T19:12:58.272+0530      [#####.....] userDB.delayData      13.8MB/29.4MB (46.9%)
2022-04-18T19:13:01.272+0530      [#####.....] userDB.delayData      25.6MB/29.4MB (87.0%)
2022-04-18T19:13:02.237+0530      [#####.....] userDB.delayData      29.4MB/29.4MB (100.0%)
2022-04-18T19:13:02.237+0530      imported 173657 documents
hduser@abhinav:~/MongoDB$ 
• > db.delayData.count()
• 173657
```

1. Find the airports that which flights count in between 500 and 1000

- db.delayData.find({arr_flights : {\$lt :1000, \$gt : 500}})
- To find one document:

```
db.delayData.findOne({arr_flights : {$lt :1000, $gt : 500}})
```

```
> db.delayData.findOne({arr_flights : {$lt :1000, $gt : 500}})
{
    "_id" : ObjectId("625d6adf78e0b10c9e8d4055"),
    "year" : 2003,
    "month" : 6,
    "carrier" : "AA",
    "carrier_name" : "American Airlines Inc.",
    "airport" : "ATL",
    "airport_name" : "Atlanta, GA: Hartsfield-Jackson Atlanta International",
    "arr_flights" : 752,
    "arr_del15" : 186,
    "carrier_ct" : 33.99,
    "weather_ct" : 27.82,
    "nas_ct" : 104.76,
    "security_ct" : 1.9,
    "late_aircraft_ct" : 17.53,
    "arr_cancelled" : 5,
    "arr_diverted" : 0,
    "arr_delay" : 8314,
    "carrier_delay" : 1367,
    "weather_delay" : 1722,
    "nas_delay" : 3817,
    "security_delay" : 139,
    "late_aircraft_delay" : 1269,
    "" : ""
}
```

2. Find the airline name that starts with “Continental”.

- db.delayData.find({carrier_name:/^Continental/}).limit(10).pretty();

```

db.delayData.find({carrier_name:/^Continental/}).limit(10).pretty();
-- 
{
  "_id" : ObjectId("625d6adf78e0b10c9e8d40ee"),
  "year" : 2003,
  "month" : 6,
  "carrier" : "CO",
  "carrier_name" : "Continental Air Lines Inc.",
  "airport" : "ABQ",
  "airport_name" : "Albuquerque, NM: Albuquerque International Sunport",
  "arr_flights" : 120,
  "arr_del15" : 22,
  "carrier_ct" : 5.57,
  "weather_ct" : 0.51,
  "nas_ct" : 10.04,
  "security_ct" : 0,
  "late_aircraft_ct" : 5.88,
  "arr_cancelled" : 0,
  "arr_diverted" : 0,
  "arr_delay" : 1277,
  "carrier_delay" : 279,
  "weather_delay" : 29,
  "nas_delay" : 544,
  "security_delay" : 0,
  "late_aircraft_delay" : 425,
  "" : ""
}
•
> db.delayData.count({carrier_name:/^Continental/})
6917
•

```

3. Find the count of flights that are arriving for airports

- db.delayData.aggregate([{\$group : {_id : "\$airport", Fights : {\$sum : "\$arr_flights"}}, {"\$sort": {Fights: 1}}]])

```

> db.delayData.aggregate([{$group : {_id : "$airport", Fights : {$sum : "$arr_flights"}}}])
{
  "_id" : "PPG", "Fights" : 60 },
{
  "_id" : "CIU", "Fights" : 391 },
{
  "_id" : "SHD", "Fights" : 60 },
{
  "_id" : "AZA", "Fights" : 131 },
{
  "_id" : "ISM", "Fights" : 1463 },
{
  "_id" : "ESC", "Fights" : 626 },
{
  "_id" : "BRD", "Fights" : 697 },
{
  "_id" : "HOB", "Fights" : 1257 },
{
  "_id" : "GRI", "Fights" : 1454 },
{
  "_id" : "MVY", "Fights" : 343 },
{
  "_id" : "GUM", "Fights" : 958 },
{
  "_id" : "ABR", "Fights" : 1042 },
{
  "_id" : "UTM", "Fights" : 208 },
{
  "_id" : "MHK", "Fights" : 5581 },
{
  "_id" : "MNH", "Fights" : 675 },
{
  "_id" : "SAF", "Fights" : 5175 },
{
  "_id" : "ITH", "Fights" : 1995 },
{
  "_id" : "OTH", "Fights" : 5561 },
{
  "_id" : "LMT", "Fights" : 5674 },
{
  "_id" : "YKM", "Fights" : 747 }
Type "it" for more
>

```

4. Find the carrier names that belongs to ANC and IND airports

- db.delayData.find({airport:{\$in:['ANC','IND']}},).limit(10).pretty();

Output:

```

    {
        "_id" : ObjectId("625d6adf78e0b10c9e8d429b"),
        "year" : 2003,
        "month" : 6,
        "carrier" : "HP",
        "carrier_name" : "America West Airlines Inc.",
        "airport" : "IND",
        "airport_name" : "Indianapolis, IN: Indianapolis International",
        "arr_flights" : 60,
        "arr_del15" : 9,
        "carrier_ct" : 3.61,
        "weather_ct" : 0,
        "nas_ct" : 3.69,
        "security_ct" : 0,
        "late_aircraft_ct" : 1.7,
        "arr_cancelled" : 0,
        "arr_diverted" : 0,
        "arr_delay" : 272,
        "carrier_delay" : 135,
        "weather_delay" : 0,
        "nas_delay" : 72,
        "security_delay" : 0,
        "late_aircraft_delay" : 65,
        "" : ""
    }

```

-

5. Find the total security delay of IND airport with airlines having security delays

- db.delayData.aggregate({\$match:{airport:"IND"}},{\$group:{_id:"\$carrier",Total:{\$sum:"\$security_delay"}}});

```

> db.delayData.aggregate({$match:{airport:"IND"}},{$group:{_id:"$carrier",Total:{$sum:"$security
delay"}}});
{ "_id" : "XE", "Total" : 942 }
{ "_id" : "F9", "Total" : 0 }
{ "_id" : "FL", "Total" : 0 }
{ "_id" : "OO", "Total" : 610 }
{ "_id" : "OH", "Total" : 474 }
{ "_id" : "AA", "Total" : 432 }
{ "_id" : "MQ", "Total" : 24 }
{ "_id" : "CO", "Total" : 1358 }
{ "_id" : "UA", "Total" : 0 }
{ "_id" : "EV", "Total" : 452 }
{ "_id" : "DL", "Total" : 80 }
{ "_id" : "NW", "Total" : 1470 }
{ "_id" : "HP", "Total" : 320 }
{ "_id" : "YV", "Total" : 104 }
{ "_id" : "RU", "Total" : 346 }
{ "_id" : "DH", "Total" : 140 }
{ "_id" : "TZ", "Total" : 626 }
{ "_id" : "9E", "Total" : 206 }
{ "_id" : "US", "Total" : 612 }
{ "_id" : "WN", "Total" : 2786 }
> █

```

6. Find the number of flight diversions that occur in St Louis

Query:

- var m1 = function(){
 var rval = { airport: this.airport_name,
 year: this.year };
 if(this.airport=="STL") {
 emit(rval, this.arr_diverted);
 }
 };
- var r1 = function(key, valArr) {
 return Array.sum(valArr);
 }

- db.delayData.mapReduce(m1, r1, {out:'tmp'});
- db.tmp.find().sort({'_id.year':-1})

Result:

```
> var m1 = function(){
...   var rval = { airport: this.airport_name,
...             year: this.year };
...   if( this.airport=="STL" ) {
...     emit( rval, this.arr_diverted );
...   }
... };
>
> var r1 = function( key, valArr ) {
...   return Array.sum( valArr );
... }
>
> db.delayData.mapReduce( m1, r1, {out:'tmp'});
{
  "result" : "tmp",
  "timeMillis" : 751,
  "counts" : {
    "input" : 173657,
    "emit" : 1601,
    "reduce" : 135,
    "output" : 11
  },
  "ok" : 1
}
```

```
> db.tmp.find().sort( {'_id.year':-1} )
[{"_id": {"airport": "St. Louis, MO: Lambert-St. Louis International", "year": 2013}, "value": 94},
 {"_id": {"airport": "St. Louis, MO: Lambert-St. Louis International", "year": 2012}, "value": 100},
 {"_id": {"airport": "St. Louis, MO: Lambert-St. Louis International", "year": 2011}, "value": 150},
 {"_id": {"airport": "St. Louis, MO: Lambert-St. Louis International", "year": 2010}, "value": 85},
 {"_id": {"airport": "St. Louis, MO: Lambert-St. Louis International", "year": 2009}, "value": 83},
 {"_id": {"airport": "St. Louis, MO: Lambert-St. Louis International", "year": 2008}, "value": 165},
 {"_id": {"airport": "St. Louis, MO: Lambert-St. Louis International", "year": 2007}, "value": 144},
 {"_id": {"airport": "St. Louis, MO: Lambert-St. Louis International", "year": 2006}, "value": 119},
 {"_id": {"airport": "St. Louis, MO: Lambert-St. Louis International", "year": 2005}, "value": 116},
 {"_id": {"airport": "St. Louis, MO: Lambert-St. Louis International", "year": 2004}, "value": 135},
 {"_id": {"airport": "St. Louis, MO: Lambert-St. Louis International", "year": 2003}, "value": 109}]
```

7.

BIG DATA MONGODB:

CB.EN.U4CSE19459

SHANTHAN REDDY

DATABASE:

Label/field name	Meaning
year	The year the month occurred
month	The month for which data was collected
carrier	The airline code
carrier_name	The airline name
airport	The airport code
airport_name	The airport name
arr_flights	# of flights that arrived at the airport
arr_del15	# of flights that arrived >= 15 minutes late
carrier_ct	# of flights delayed due to the carrier
weather_ct	# of flights delayed due to weather
nas_ct	# of flights delayed due to national air system
security_ct	# of flights delayed due to security
late_aircraft_ct	# flights delayed because a previous flight using the same aircraft was late
arr_cancelled	# of canceled arrivals
arr_diverted	# of scheduled arrivals that were diverted
arr_delay	Sum of the delay minutes
carrier_delay	Total minutes of delays due to carriers
weather_delay	Total minutes of delays due to weather
nas_delay	Total minutes of delays due to natl. air service
security_delay	Total minutes of delays due to security
late_aircraft_delay	Similar to late_aircraft_ct. The total minutes of delay due to a previous flight using the same aircraft arriving late.

```
hduser@shanthan-VirtualBox:~/Downloads$ mongo
MongoDB shell version v3.6.8
connecting to: mongodb://127.0.0.1:27017
Implicit session: session { "id" : UUID("e14d94d4-f692-4071-9291-e81f377fde04") }
MongoDB server version: 3.6.8
Server has startup warnings:
2022-04-18T17:23:56.669+0530 I STORAGE  [initandlisten]
2022-04-18T17:23:56.669+0530 I STORAGE  [initandlisten] ** WARNING: Using the XFS filesystem is strongly recommended with the WiredTiger storage engine
2022-04-18T17:23:56.669+0530 I STORAGE  [initandlisten] **             See http://dochub.mongodb.org/core/prodnotes-filesystem
2022-04-18T17:24:06.331+0530 I CONTROL  [initandlisten]
2022-04-18T17:24:06.331+0530 I CONTROL  [initandlisten] ** WARNING: Access control is not enabled for the database.
2022-04-18T17:24:06.332+0530 I CONTROL  [initandlisten] **             Read and write access to data and configuration is unrestricted.
2022-04-18T17:24:06.332+0530 I CONTROL  [initandlisten]
> show dbs;
admin  0.000GB
config  0.000GB
local   0.000GB
userDB  0.026GB
> use userDB
switched to db userDB
> show collections
delayData
tmp
>
```

QUERY 1:

```
db.delayData.find({carrier_name:"American Airlines Inc."}).pretty();
```

```

"carrier" : "AA",
"carrier_name" : "American Airlines Inc.",
"airport" : "ABQ",
"airport_name" : "Albuquerque, NM: Albuquerque International Sunport",
"arr_flights" : 307,
"arr_del15" : 56,
"carrier_ct" : 14.68,
"weather_ct" : 10.79,
"nas_ct": 19.09,
"security_ct" : 1.48,
"late_aircraft_ct" : 9.96,
"arr_cancelled" : 1,
"arr_diverted" : 1,
"arr_delay" : 2530,
"carrier_delay" : 510,
"weather_delay" : 621,
"nas_delay" : 676,
"security_delay" : 25,
"late_aircraft_delay" : 698,
"" : ""

"_id" : ObjectId("625d7ad10c34204c60928a68"),
"year" : 2003,
"month" : 6,
"carrier" : "AA",
"carrier_name" : "American Airlines Inc.",
"airport" : "ANC",
"airport_name" : "Anchorage, AK: Ted Stevens Anchorage International",
"arr_flights" : 90,
"arr_del15" : 27,
"carrier_ct" : 7.09,
"weather_ct" : 2,
"nas_ct": 10.75,
"security_ct" : 0,
"late_aircraft_ct" : 7.16,
"arr_cancelled" : 0,
"arr_diverted" : 0,
"arr_delay" : 1390,
"carrier_delay" : 271,
"weather_delay" : 83,
"nas_delay" : 581,
"security_delay" : 0,
"late_aircraft_delay" : 455,
"" : ""

```

QUERY 2:

```

#MapReduce

var m1 = function(){
... emit( this.airport_name, this.arr_diverted );
... };

var r1 = function( key, valArr ) {
... return Array.sum( valArr );
... }

db.delayData.mapReduce( m1, r1, {out:'tmp'});

```

```

        "late_aircraft_delay" : 698,
        "" : ""
    }
{
    "_id" : ObjectId("625d7ad10c34204c60928a68"),
    "year" : 2003,
    "month" : 6,
    "carrier" : "AA",
    "carrier_name" : "American Airlines Inc.",
    "airport" : "ANC",
    "airport_name" : "Anchorage, AK: Ted Stevens Anchorage International",
    "arr_flights" : 90,
    "arr_del15" : 27,
    "carrier_ct" : 7.09,
    "weather_ct" : 2,
    "nas_ct" : 10.75,
    "security_ct" : 0,
    "late_aircraft_ct" : 7.16,
    "arr_cancelled" : 0,
    "arr_diverted" : 0,
    "arr_delay" : 1390,
    "carrier_delay" : 271,
    "weather_delay" : 83,
    "nas_delay" : 581,
    "security_delay" : 0,
    "late_aircraft_delay" : 455,
    "" : ""
}
Type "it" for more
> var m1 = function(){
...   ... emit( this.airport_name, this.arr_diverted );
...   ... };
> var r1 = function( key, valArr ) {
...   ... return Array.sum( valArr );
...   ... }
> db.delayData.mapReduce( m1, r1, {out:'tmp'});
{
    "result" : "tmp",
    "timeMillis" : 900,
    "counts" : {
        "input" : 173657,
        "emit" : 173657,
        "reduce" : 16838,
        "output" : 363
    },
    "ok" : 1
}
C> ■

```

QUERY 3:

```
db.tmp.find().sort( {value:1} )
```

```

> db.tmp.find().sort( {value:1} )
{ "_id" : "Clarksburg/Fairmont, WV: North Central West Virginia", "value" : 0 }
{ "_id" : "Columbus, MS: Columbus AFB", "value" : 0 }
{ "_id" : "Dickinson, ND: Dickinson - Theodore Roosevelt Regional", "value" : 0 }
{ "_id" : "Greenville, MS: Mid Delta Regional", "value" : 0 }
{ "_id" : "Guam, TT: Guam International", "value" : 0 }
{ "_id" : "Gustavus, AK: Gustavus Airport", "value" : 0 }
{ "_id" : "Iron Mountain/Kingsfd, MI: Ford", "value" : 0 }
{ "_id" : "Kansas City, MO: Charles B. Wheeler Downtown", "value" : 0 }
{ "_id" : "Moses Lake, WA: Grant County International", "value" : 0 }
{ "_id" : "Pago Pago, TT: Pago Pago International", "value" : 0 }
{ "_id" : "Palmdale, CA: Palmdale USAF Plant 42", "value" : 0 }
{ "_id" : "Phoenix, AZ: Phoenix - Mesa Gateway", "value" : 0 }
{ "_id" : "Saipan, TT: Francisco C. Ada Saipan International", "value" : 0 }
{ "_id" : "Salem, OR: McNary Field", "value" : 0 }
{ "_id" : "Staunton, VA: Shenandoah Valley Regional", "value" : 0 }
{ "_id" : "Visalia, CA: Visalia Municipal", "value" : 0 }
{ "_id" : "Yakima, WA: Yakima Air Terminal/McAllister Field", "value" : 0 }
{ "_id" : "Del Rio, TX: Del Rio International", "value" : 1 }
{ "_id" : "Houston, TX: Ellington", "value" : 1 }

```

QUERY 4:

```
db.delayData.find({arr_cancelled:{$lte:5}}).pretty();
```

```
"carrier" : "AA",
"carrier_name" : "American Airlines Inc.",
"airport" : "ELP",
"airport_name" : "El Paso, TX: El Paso International",
"arr_flights" : 269,
"arr_del15" : 78,
"carrier_ct" : 27.29,
"weather_ct" : 9.68,
"nas_ct" : 27.28,
"security_ct" : 1.38,
"late_aircraft_ct" : 12.38,
"arr_cancelled" : 2,
"arr_diverted" : 0,
"arr_delay" : 3589,
"carrier_delay" : 1113,
"weather_delay" : 564,
"nas_delay" : 982,
"security_delay" : 22,
"late_aircraft_delay" : 908,
"" : ""

"_id" : ObjectId("625d7ad10c34204c60928a6c"),
"year" : 2003,
"month" : 6,
"carrier" : "AA",
"carrier_name" : "American Airlines Inc.",
"airport" : "FAT",
"airport_name" : "Fresno, CA: Fresno Yosemite International",
"arr_flights" : 77,
"arr_del15" : 21,
"carrier_ct" : 4.09,
"weather_ct" : 3.53,
"nas_ct" : 7.84,
"security_ct" : 0,
"late_aircraft_ct" : 5.54,
"arr_cancelled" : 0,
"arr_diverted" : 0,
"arr_delay" : 906,
"carrier_delay" : 337,
"weather_delay" : 99,
"nas_delay" : 216,
"security_delay" : 0,
"late_aircraft_delay" : 254,
"" : ""
```

QUERY 5:

```
db.delayData.find().sort({arr_flights:1}).pretty().limit(10);
```

```
"month" : 2,
"carrier" : "TZ",
"carrier_name" : "ATA Airlines d/b/a ATA",
"airport" : "TPA",
"airport_name" : "Tampa, FL: Tampa International",
"arr_flights" : 0,
"arr_del15" : 0,
"carrier_ct" : 0,
"weather_ct" : 0,
"nas_ct" : 0,
"security_ct" : 0,
"late_aircraft_ct" : 0,
"arr_cancelled" : 0,
"arr_diverted" : 0,
"arr_delay" : 0,
"carrier_delay" : 0,
"weather_delay" : 0,
"nas_delay" : 0,
"security_delay" : 0,
"late_aircraft_delay" : 0,
"" : ""

"_id" : ObjectId("625d7ad10c34204c6092b910"),
"year" : 2004,
"month" : 3,
"carrier" : "EV",
"carrier_name" : "Atlantic Southeast Airlines",
"airport" : "BNA",
"airport_name" : "Nashville, TN: Nashville International",
"arr_flights" : 0,
"arr_del15" : 0,
"carrier_ct" : 0,
"weather_ct" : 0,
"nas_ct" : 0,
"security_ct" : 0,
"late_aircraft_ct" : 0,
"arr_cancelled" : 0,
"arr_diverted" : 0,
"arr_delay" : 0,
"carrier_delay" : 0,
"weather_delay" : 0,
"nas_delay" : 0,
"security_delay" : 0,
"late_aircraft_delay" : 0,
```