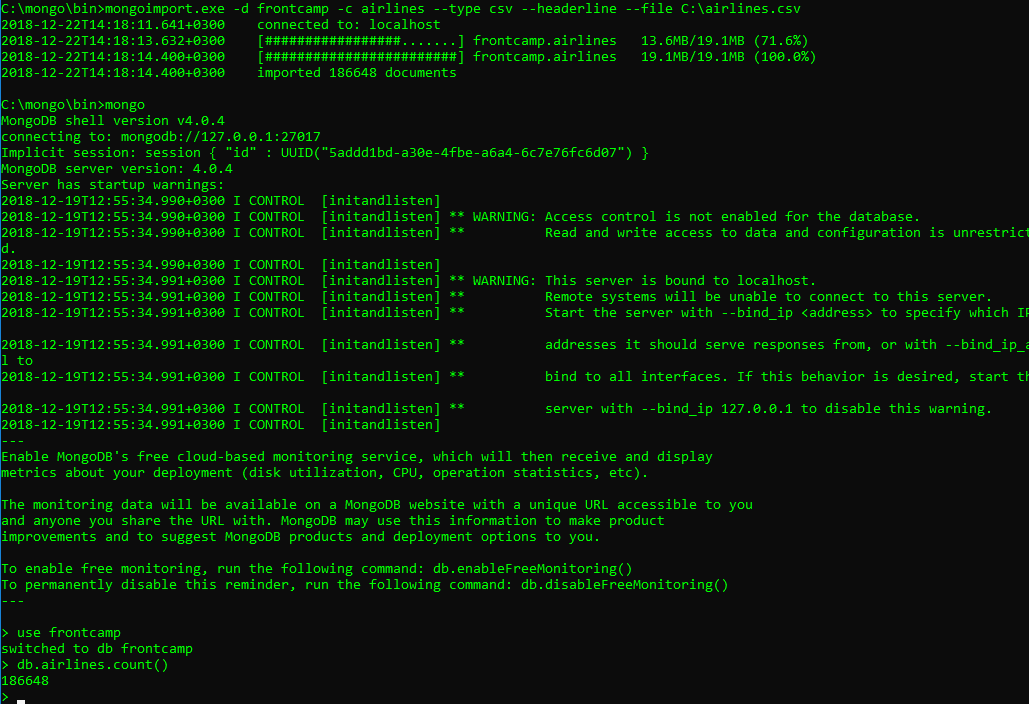
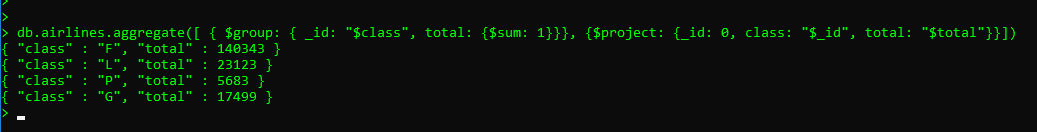
**MongoDB part2.**

Start MongoDB and import airlines.csv



1. How many records does each airline class have? Use $project to show result as { class: "Z", total: 999 } ?

**  
Query:**

>db.airlines.aggregate([

{ $group: { \_id: "$class", total: {$sum: 1}}},

{$project: {\_id: 0, class: "$\_id", total: "$total"}}

])

**Result:**

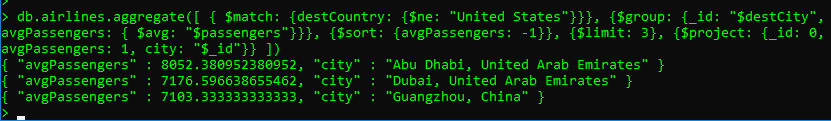
{ "class" : "F", "total" : 140343 }

{ "class" : "L", "total" : 23123 }

{ "class" : "P", "total" : 5683 }

{ "class" : "G", "total" : 17499 }

2. What are the top 3 **destination cities outside** of the United States (destCountry field, not included) with the **highest average** passengers count? Show result as { "avgPassengers" : 2312.380, "city" : "Minsk, Belarus" } ?



**Query:**

> db.airlines.aggregate([

{$match: {destCountry: {$ne: "United States"}}},

{$group: {\_id: "$destCity", avgPassengers: { $avg: "$passengers"}}},

{$sort: {avgPassengers: -1}}, {$limit: 3},

{$project: {\_id: 0, avgPassengers: 1, city: "$\_id"}}

])

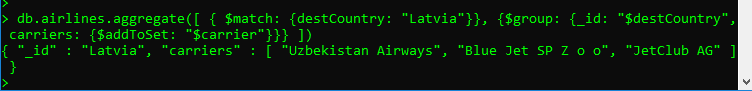
**Result:**

{ "avgPassengers" : 8052.380952380952, "city" : "Abu Dhabi, United Arab Emirates" }

{ "avgPassengers" : 7176.596638655462, "city" : "Dubai, United Arab Emirates" }

{ "avgPassengers" : 7103.333333333333, "city" : "Guangzhou, China" }

3. Which carriers provide flights to Latvia (destCountry)? Show result as one document { "\_id" : "Latvia", "carriers" : [ "carrier1", " carrier2", …] }



**Query:**

> db.airlines.aggregate([

{ $match: {destCountry: "Latvia"}},

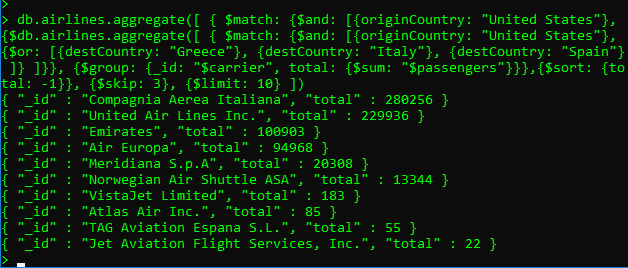
{$group: {\_id: "$destCountry", carriers: {$addToSet: "$carrier"}}}

])

**Result:**

{ "\_id" : "Latvia", "carriers" : [ "Uzbekistan Airways", "Blue Jet SP Z o o", "JetClub AG" ] }

4. What are the carriers which flue the most number of passengers from the United State to either Greece, Italy or Spain? Find top 10 carriers, but provide the last 7 carriers (do not include the first 3). Show result as { "\_id" : "<carrier>", "total" : 999}



**Query:**

>db.airlines.aggregate([

{$match: {$and: [

{originCountry: "United States"},{$or: [

{destCountry: "Greece"},

{destCountry: "Italy"},

{destCountry: "Spain"}

]}

]}},

{$group: {\_id: "$carrier", total: {$sum: "$passengers"}}},

{$sort: {total: -1}},

{$skip: 3},

{$limit: 10},

])

**Result:**

{ "\_id" : "Compagnia Aerea Italiana", "total" : 280256 }

{ "\_id" : "United Air Lines Inc.", "total" : 229936 }

{ "\_id" : "Emirates", "total" : 100903 }

{ "\_id" : "Air Europa", "total" : 94968 }

{ "\_id" : "Meridiana S.p.A", "total" : 20308 }

{ "\_id" : "Norwegian Air Shuttle ASA", "total" : 13344 }

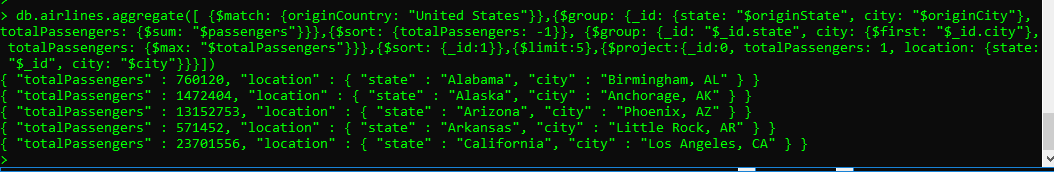
{ "\_id" : "VistaJet Limited", "total" : 183 }

{ "\_id" : "Atlas Air Inc.", "total" : 85 }

{ "\_id" : "TAG Aviation Espana S.L.", "total" : 55 }

{ "\_id" : "Jet Aviation Flight Services, Inc.", "total" : 22 }

5. Find the city (originCity) with the highest sum of passengers for each state (originState) of the United States (originCountry). Provide the city for the first 5 states ordered by state alphabetically (you should see the city for Alaska, Arizona and etc). Show result as { "totalPassengers" : 999, "location" : { "state" : "abc", "city" : "xyz" } }



**Query:**  
> db.airlines.aggregate([

{$match: {originCountry: "United States"}},

{$group: {\_id: {state: "$originState", city: "$originCity"}, totalPassengers: {$sum: "$passengers"}}},

{$sort: {totalPassengers: -1}},

{$group: {\_id: "$\_id.state", city: {$first: "$\_id.city"}, totalPassengers: {$max: "$totalPassengers"}}},

{$sort: {\_id:1}},

{$limit:5},

{$project:{\_id:0, totalPassengers: 1, location: {state: "$\_id", city: "$city"}}}

])

**Response:**

{ "totalPassengers" : 760120, "location" : { "state" : "Alabama", "city" : "Birmingham, AL" } }

{ "totalPassengers" : 1472404, "location" : { "state" : "Alaska", "city" : "Anchorage, AK" } }

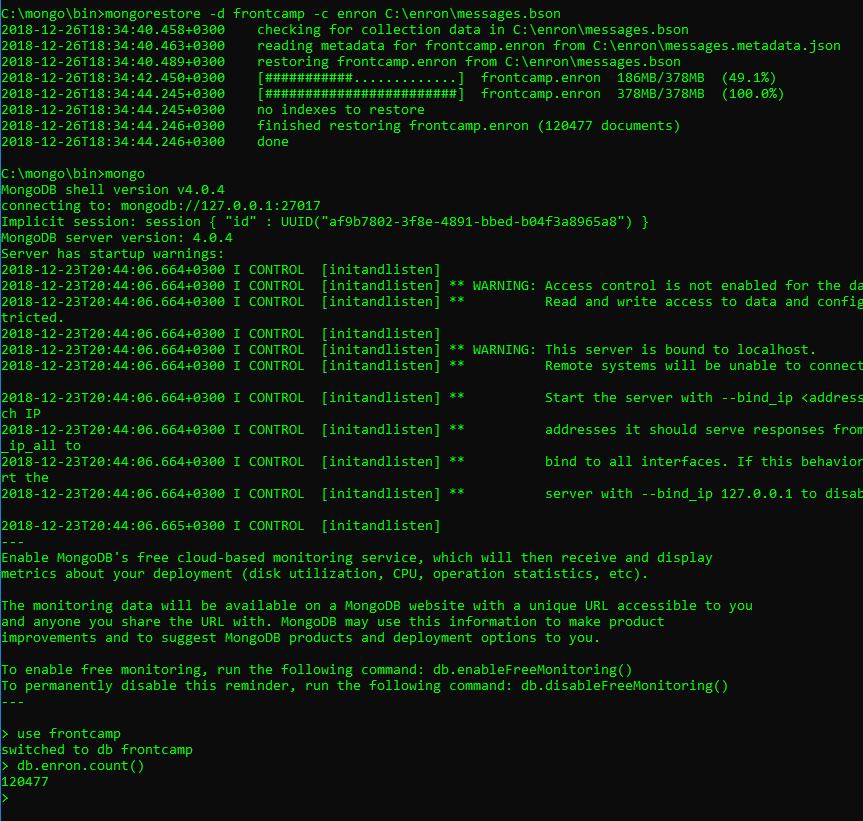
{ "totalPassengers" : 13152753, "location" : { "state" : "Arizona", "city" : "Phoenix, AZ" } }

{ "totalPassengers" : 571452, "location" : { "state" : "Arkansas", "city" : "Little Rock, AR" } }

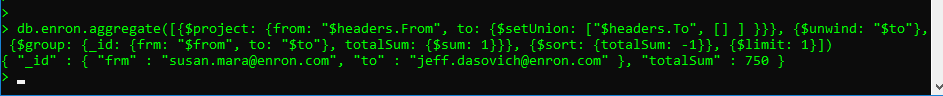
{ "totalPassengers" : 23701556, "location" : { "state" : "California", "city" : "Los Angeles, CA" } }

6. Enron task

Start **MongoDB** and restore **enron.bson**



Which pair of people have the greatest number of messages in the dataset?



**Query:**  
> db.enron.aggregate([

{$project: {from: "$headers.From", to: {$setUnion: ["$headers.To", [] ] }}},

{$unwind: "$to"},

{$group: {\_id: {from: "$from", to: "$to"}, totalSum: {$sum: 1}}},

{$sort: {totalSum: -1}},

{$limit: 1}

])

**Response:**

{ "\_id" : { "from" : "susan.mara@enron.com", "to" : "jeff.dasovich@enron.com" }, "totalSum" : 750 }