

Analiza radova na putu u SAD (2016 - 2021)

SISTEMI BAZA PODATAKA

IN45/2019 OGNJEN POZNANOVIĆ

IN35/2019 LAZAR FRATRIĆ

Predlozi agregacija

Ognjen

1. Pronaći 10 država sa najdužim prosečnim vremenom trajanja radova.
2. Naći 10 država sa najvećim temperaturama za svaku godinu.
3. Naći za svaku državu po godinama koliko je radova obavljeno na levoj i desnoj strani.
4. Naći prosečnu temperaturu, vlažnost i pritisak za sve vremenske zone po godinama.
5. Naći koliko je svaka država obnovila milja puteva za svaku kategoriju radova kada je vreme izvođenja manje od godinu dana.

Predlozi agregacija

Lazar

6. Prikazati broj radova i prosečan broj milja po radu za svaki grad u radijus od 500km od NYC, sortirani u opadajućem redosledu.
7. U kojoj saveznoj državi, se najviše izvodilo radova po kategorijama, po godinama?
8. Koje je prosečno vreme trajanja radova u blizini pešačkog prelaza, za svaki grad?
9. Pronaći šifru radova koji su odradili najviše milja po gradovima.
10. Koja je ukupna dužina radova na autoputu za grad?

Inicijalna logička šema

```
{
  "_id" : ObjectId("647f4779a3cd254107b612eb"),
  "ID" : "C-2079942",
  "Severity" : NumberInt(4),
  "Start_Time" : ISODate("2020-11-09T05:46:55.000+0000"),
  "End_Time" : ISODate("2020-11-16T07:59:00.000+0000"),
  "Start_Lat" : 40.740911,
  "Start_Lng" : -96.644236,
  "End_Lat" : "40.74262",
  "End_Lng" : "-96.644244",
  "Distance(mi)" : 0.118081396671,
  "Description" : "Closed road from Pine Lake Rd to Madalyn Rd due to roadwork.",
  "Number" : null,
  "Street" : "S 56th St",
  "Side" : "R",
  "City" : "Lincoln",
  "County" : "Lancaster",
  "State" : "NE",
  "Zipcode" : "68516",
  "Country" : "US",
  "Timezone" : "US/Central",
  "Airport_Code" : "KLNK",
  "Weather_Timestamp" : ISODate("2020-11-09T09:54:00.000+0000"),
  "Temperature(F)" : "66.0",
  "Wind_Chill(F)" : "66.0",
  "Humidity(%)" : "78.0",
  "Pressure(in)" : "28.5",
  "Visibility(mi)" : "10.0",
  "Wind_Direction" : "S",
  "Wind_Speed(mph)" : "23.0",
  "Precipitation(in)" : "0.02",
  "Weather_Condition" : "Cloudy / Windy",
  "Amenity" : false,
  "Bump" : false,
  "Crossing" : true,
  "Give_Way" : false,
  "Junction" : false,
  "No_Exit" : false,
  "Railway" : false,
  "Roundabout" : false,
  "Station" : false,
  "Stop" : false,
  "Traffic_Calming" : false,
  "Traffic_Signal" : true,
  "Turning_Loop" : false,
  "Sunrise_Sunset" : "Night",
  "Civil_Twilight" : "Night",
  "Nautical_Twilight" : "Night",
  "Astronomical_Twilight" : "Day",
  "Start_day" : 9.0,
  "Start_month" : 11.0,
  "Start_year" : 2020.0,
  "End_day" : 16.0,
  "End_month" : 11.0,
  "End_year" : 2020.0
}
```

Izmene nad bazom

- Iz baze su izbačeni svi dokumenti koji su imali null vrednost u nekom od polja
- Polja Start_time i End_time su razbijeni na polja Start_year, Start_month, Start_day i End_year, End_month, End_day

Agregacije



Primer agregacije

Naći 10 država sa najdužim prosečnim vremenom trajanja radova - vreme izvršavanja 55s

```
db.radovi.aggregate([
  {
    $project: {
      State: 1,
      trajanje: {
        $subtract: [
          { $add: [{ $multiply: ["$End_year", 360] }, { $multiply: ["$End_month", 30] }, "$End_day" ] },
          { $add: [{ $multiply: ["$Start_year", 360] }, { $multiply: ["$Start_month", 30] }, "$Start_day" ] }
        ]
      }
    }
  },
  {
    $group: {
      _id: "$State",
      prosečno_trajanje: { $avg: "$trajanje" }
    }
  },
  {
    $sort: {
      prosečno_trajanje: -1
    }
  },
  {
    $limit: 10
  },
  {
    $project: {
      _id: 0,
      drzava: "$_id",
      prosečno_trajanje: { $round: ["$prosečno_trajanje", 2] }
    }
  }
])
```

Primer agregacije

Naći 10 država sa sa najvećim temperaturama za svaku godinu - vreme izvršavanja 9s

```
db.radovi.aggregate([
  {
    $group: {
      _id: {
        State: "$State",
        Start_year: "$Start_year"
      },
      maxTemperatura: { $max: "$Temperature(F)" }
    }
  },
  {
    $group: {
      _id: "$_id.State",
      maxTemperature: {
        $push: {
          Start_year: "$_id.Start_year",
          maxTemperatura: "$maxTemperatura"
        }
      }
    }
  }
])
```

```
{
  $project: {
    _id: 0,
    State: "$_id",
    maxTemperature: 1
  },
  {
    $sort: {
      "maxTemperature.maxTemperatura": -1
    }
  },
  {
    $limit: 10
  }
])
```


Primer agregacije

Naći za svaku državu po godinama koliko je radova obavljeno na levoj i desnoj strani.

```
db.radovi.aggregate([
  {
    $group: {
      _id: {
        State: "$State",
        Start_year: "$Start_year",
        strana: {
          $cond: {
            if: { $eq: ["$Side", "L"] },
            then: "Leva",
            else: "Desna"
          }
        }
      },
      brojRadova: { $sum: 1 }
    }
  },
])
```

```
{
  $group: {
    _id: {
      State: "$_id.State",
      Start_year: "$_id.Start_year"
    },
    radovi: {
      $push: {
        strana: "$_id.strana",
        brojRadova: "$brojRadova"
      }
    },
    ukupanBrojRadova: { $sum: "$brojRadova" }
  },
  {
    $sort: {
      "_id.State": 1
    }
  }
})
```

Primer agregacije

Naći prosečnu temperaturu, vlažnost i pritisak za sve vremenske zone po godinama

```
db.radovi.aggregate([
  {
    $group: {
      _id: {
        Timezone: "$Timezone",
        Start_year: "$Start_year"
      },
      avgTemperatura: { $avg: "$Temperature(F)" },
      avgHumidity: { $avg: "$Humidity(%)" },
      avgPressure: { $avg: "$Pressure(in)" }
    }
  },
  {
    $group: {
      _id: "$_id.Timezone",
      vrmenski_uslovi: {
        $push: {
          Start_year: "$_id.Start_year",
          avgTemperatura: "$avgTemperatura",
          avgHumidity: "$avgHumidity",
          avgPressure: "$avgPressure",
        }
      }
    }
  }
])
```

```
{
  $project: {
    _id: 0,
    Timezone: "$_id",
    vrmenski_uslovi: 1
  },
  {
    $unwind: "$vrmenski_uslovi"
  },
  {
    $sort: {
      "vrmenski_uslovi.avgTemperatura": -1
    }
  }
})
```

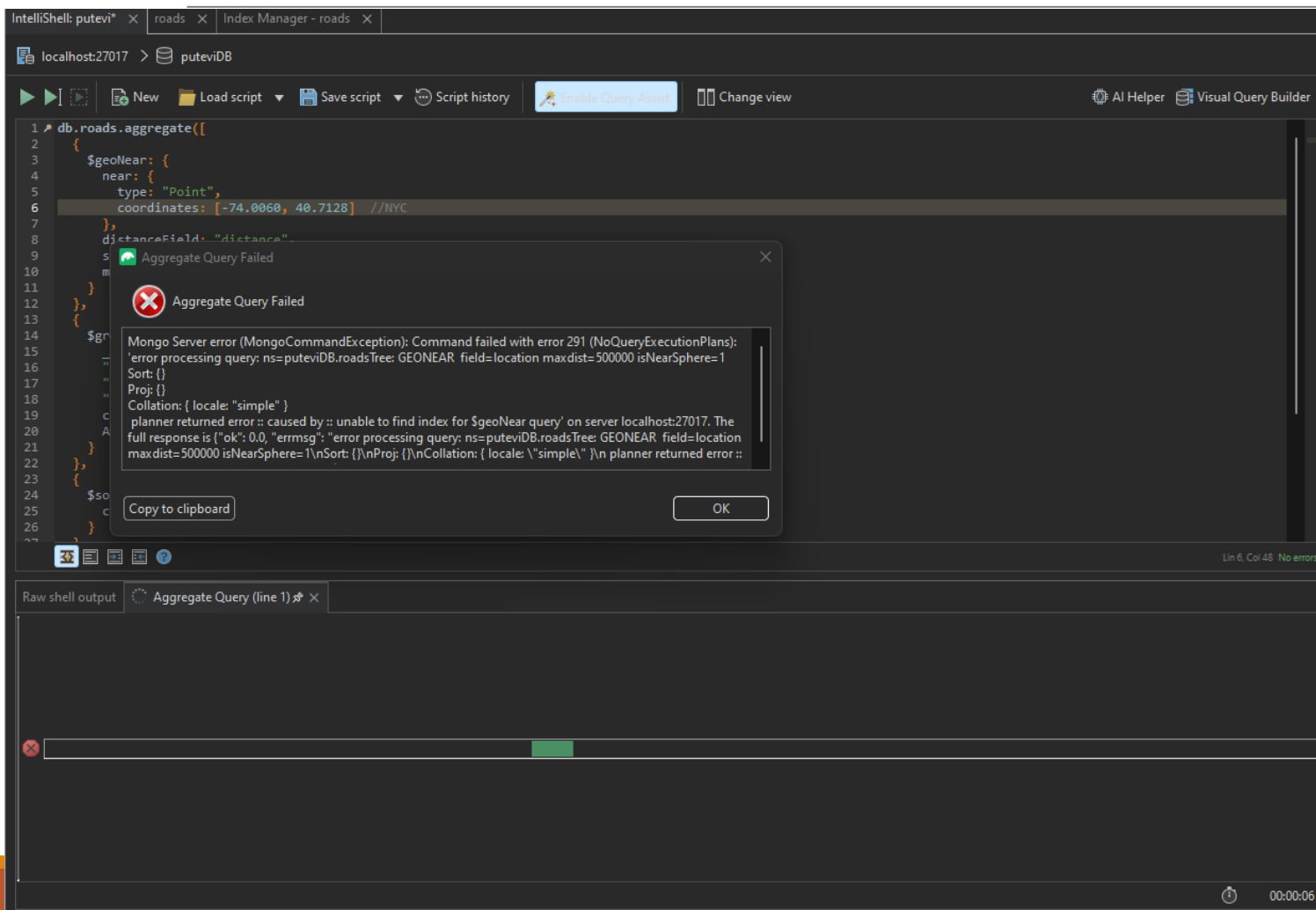
Primer agregacije

Naći koliko je svaka država obnovila milja puteva za svaku kategoriju radova kada je vreme izvođenja manje od godinu dana - vreme izvršavanja 52s

```
db.radovi.aggregate([
  {
    $project: {
      State: 1,
      Severity: 1,
      "Distance(mi)": 1,
      trajanje: {
        $subtract: [
          { $add: [{ $multiply: ["$End_year", 360] }, { $multiply: ["$End_month", 30] }, "$End_day" ] },
          { $add: [{ $multiply: ["$Start_year", 360] }, { $multiply: ["$Start_month", 30] }, "$Start_day" ] }
        ]
      }
    }
  },
  {
    $match: {
      trajanje: { $lt: 366 }
    }
  }
])
```

```
{
  $group: {
    _id: {
      State: "$State",
      Severity: "$Severity"
    },
    uradjeno_milja: { $sum: "$Distance(mi)" },
  },
  {
    $group: {
      _id: "$_id.State",
      radovi: {
        $push: {
          Severity: "$_id.Severity",
          uradjeno_milja: "$uradjeno_milja"
        }
      }
    }
  },
  {
    $project: {
      _id: 0,
      State: "$_id",
      radovi: 1
    }
  }
])
```

Primer agregacije



The screenshot shows the MongoDB IntelliShell interface. The top bar indicates the connection to 'localhost:27017' and the database 'puteviDB'. The main editor displays an aggregate query for the 'roads' collection, using the '\$geoNear' operator to find roads near NYC coordinates. An 'Aggregate Query Failed' dialog box is open, displaying a detailed error message from the MongoDB server. The error message states: 'Mongo Server error (MongoCommandException): Command failed with error 291 (NoQueryExecutionPlans): 'error processing query: ns=puteviDB.roadsTree: GEONEAR field=location maxdist=500000 isNearSphere=1 Sort: {} Proj: {} Collation: { locale: "simple" } planner returned error :: caused by :: unable to find index for \$geoNear query' on server localhost:27017. The full response is {"ok": 0.0, "errmsg": "error processing query: ns=puteviDB.roadsTree: GEONEAR field=location maxdist=500000 isNearSphere=1\nSort: {}\nProj: {}\nCollation: { locale: \"simple\" }\nplanner returned error ::'. Below the error message, there are 'Copy to clipboard' and 'OK' buttons. The bottom of the interface shows a 'Raw shell output' panel, which is currently empty, and a status bar at the bottom right indicating '00:00:06'.

```
1 db.roads.aggregate([
2   {
3     $geoNear: {
4       near: {
5         type: "Point",
6         coordinates: [-74.0060, 40.7128] //NYC
7       },
8       distanceField: "distance",
9       s
10      m
11    }
12  },
13  {
14    $gr
15  },
16  {
17    "
18    Proj: {}
19    Collation: { locale: "simple" }
20  },
21  {
22    $so
23  },
24  {
25    c
26  }
27 ])
```

Aggregate Query Failed

Mongo Server error (MongoCommandException): Command failed with error 291 (NoQueryExecutionPlans): 'error processing query: ns=puteviDB.roadsTree: GEONEAR field=location maxdist=500000 isNearSphere=1 Sort: {} Proj: {} Collation: { locale: "simple" } planner returned error :: caused by :: unable to find index for \$geoNear query' on server localhost:27017. The full response is {"ok": 0.0, "errmsg": "error processing query: ns=puteviDB.roadsTree: GEONEAR field=location maxdist=500000 isNearSphere=1\nSort: {}\nProj: {}\nCollation: { locale: \"simple\" }\nplanner returned error ::'.

Copy to clipboard OK

Raw shell output Aggregate Query (line 1) ✖

00:00:06

6. Prikazati broj radova i prosečan broj milja po radu za svaki grad u radijus od 500km od NYC, sortirani u opadajućem redosledu.

Vreme izvršavanja: Upit se ne može izvršiti bez dodavanja indeksa!

Primer agregacije

The screenshot shows the IntelliJShell IDE interface. The top panel displays a MongoDB aggregation query. The bottom panel shows the results of the query in a table view.

```
9 { $add: [ { $multiply: [ $start_year, 360 ] }, { $multiply: [ $start_month, 30 ] }, { $start_day } ] }
10 }
11 }
12 }
13 { $group: {
14   "_id": { "year": "$start_year", "cat": "$severity" },
15   "year": { "$first": "$start_year" },
16   "cat": { "$first": "$severity" },
17   "state": { "$first": "$state" },
18   "avg_work": { $avg: "$trajanje" },
19   "max_work": { $max: "$trajanje" }
20 }
21 }
22 { $sort: {
23   "_id": 1
24 }
25 }
26 { $project: {
27   "_id": 0,
28   "year": 1,
29   "cat": 1,
30   "state": 1,
31   "avg_work": { $round: [ "$avg_work", 1 ] },
32   "max_work": 1
33 }
34 }
```

Raw shell output: Aggregate Query (line 1) ✖

50 Documents 1 to 19

roads > year

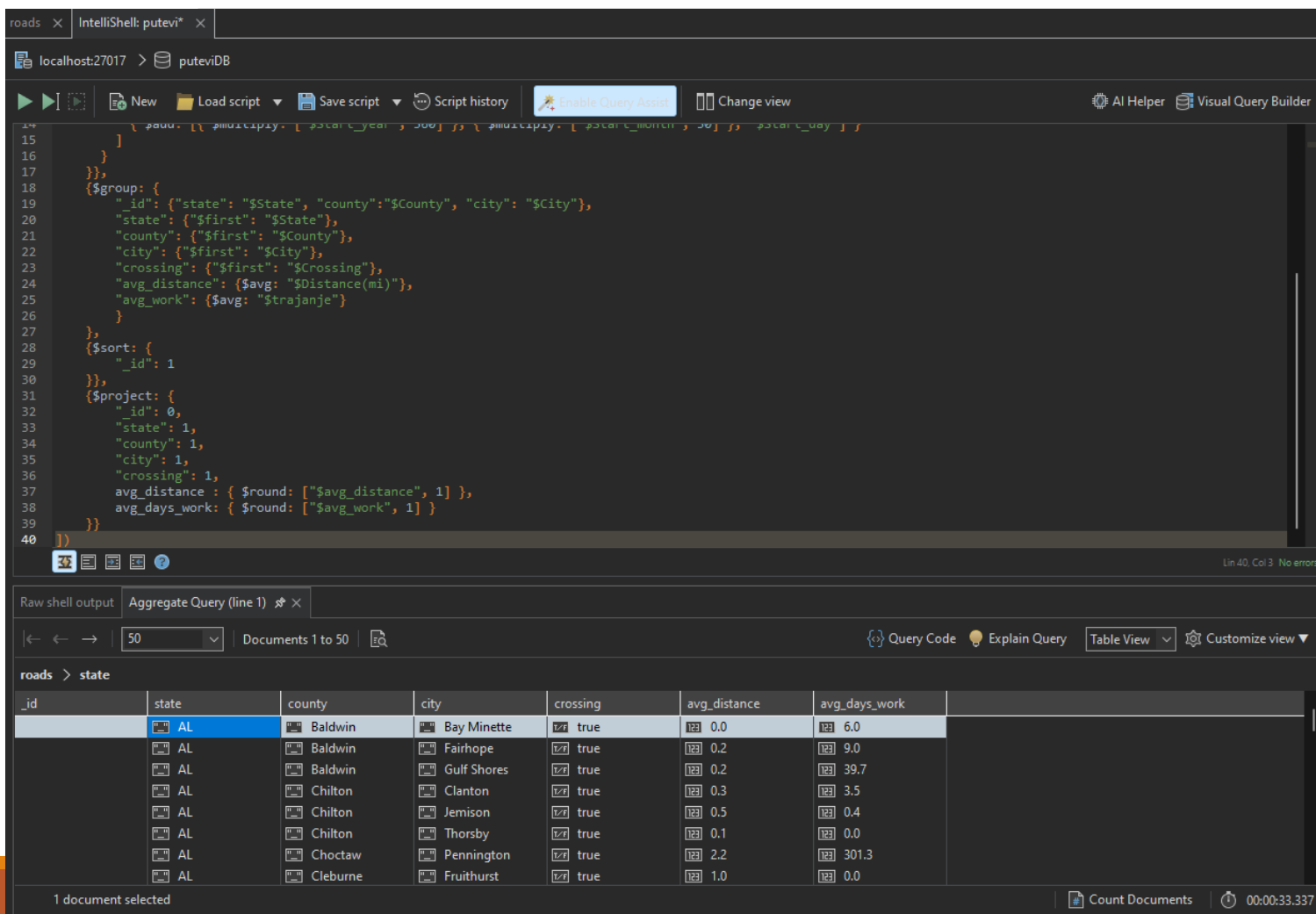
_id	year	cat	state	max_work	avg_work
123	2016	2	TN	1980.0	286.2
123	2016	3	CA	536.0	57.4
123	2016	4	OH	1865.0	74.6
123	2017	2	PA	1758.0	395.9
123	2017	3	IL	236.0	53.2
123	2017	4	WA	1615.0	301.5
123	2018	2	OR	1411.0	132.6
123	2018	3	TX	879.0	28.9

1 document selected | Count Documents | 00:01:30.112

7. U kojoj saveznoj državi, se najviše izvodilo radova po kategorijama, po godinama?

Vreme izvršavanja: 90s

Primer agregacije



The screenshot shows the IntelliJShell IDE interface. The top panel displays a MongoDB aggregation query. The bottom panel shows the results of the query in a table view.

```
15 }
16 }
17 }
18 {$group: {
19   "_id": {"state": "$State", "county": "$County", "city": "$City"},
20   "state": {"$first": "$State"},
21   "county": {"$first": "$County"},
22   "city": {"$first": "$City"},
23   "crossing": {"$first": "$Crossing"},
24   "avg_distance": {"$avg": "$Distance(mi)"},
25   "avg_work": {"$avg": "$Trajanje"}
26 }},
27 {$sort: {
28   "_id": 1
29 }},
30 }},
31 {$project: {
32   "_id": 0,
33   "state": 1,
34   "county": 1,
35   "city": 1,
36   "crossing": 1,
37   "avg_distance": { $round: [ "$avg_distance", 1 ] },
38   "avg_days_work": { $round: [ "$avg_work", 1 ] }
39 }},
40 }
```

Raw shell output: Aggregate Query (line 1) ✖

50 Documents 1 to 50

Query Code Explain Query Table View Customize view

_id	state	county	city	crossing	avg_distance	avg_days_work
AL	AL	Baldwin	Bay Minette	true	0.0	6.0
AL	AL	Baldwin	Fairhope	true	0.2	9.0
AL	AL	Baldwin	Gulf Shores	true	0.2	39.7
AL	AL	Chilton	Clanton	true	0.3	3.5
AL	AL	Chilton	Jemison	true	0.5	0.4
AL	AL	Chilton	Thorsby	true	0.1	0.0
AL	AL	Choctaw	Pennington	true	2.2	301.3
AL	AL	Cleburne	Fruithurst	true	1.0	0.0

1 document selected

Count Documents 00:00:33.337

8. Koje je prosečno vreme trajanja radova u blizini pešačkog prelaza, za svaki grad, po okrugu, po saveznoj državi?

Vreme izvršavanja: 33s

Primer agregacije

The screenshot shows the IntelliJShell IDE interface. The top panel displays a MongoDB-style aggregate query. The query uses a \$group stage to aggregate by state, county, and city, and a \$sort stage to sort by _id. The bottom panel shows the results of the query in a table view, with columns for _id, state, county, city, ID, max_miles, and days_worked. The first row is highlighted, showing results for state AL, county Autauga, and city Deatsville.

```
12 { $add: [ { $multiply: [ { $state_year: 2000 }, { $multiply: [ { $state_month: 20 }, { $state_day: 1 } ] } ] } ] }
13 }
14 }
15 }
16 { $group: {
17   "_id": { "state": "$State", "county": "$County", "city": "$City" },
18   "state": { "$first": "$State" },
19   "county": { "$first": "$County" },
20   "city": { "$first": "$City" },
21   "ID": { "$first": "$ID" },
22   "max_miles": { $max: "$Distance(mi)" },
23   "days_worked": { "$first": "$trajanje" }
24 }
25 },
26 { $sort: {
27   "_id": 1
28 } },
29 { $project: {
30   "_id": 0,
31   "state": 1,
32   "county": 1,
33   "city": 1,
34   "ID": 1,
35   "max_miles": 1,
36   "days_worked": 1
37 } }
38 }
```

roads > state

_id	state	county	city	ID	max_miles	days_worked
	AL	Autauga	Deatsville	C-17097	3.48199151351	229.0
	AL	Autauga	Marbury	C-36918	11.7727568823	1.0
	AL	Autauga	Prattville	C-236467	11.7727568823	0.0
	AL	Baldwin	Bay Minette	C-366365	5.12036415838...	0.0
	AL	Baldwin	Daphne	C-35471	3.94077170053...	0.0
	AL	Baldwin	Fairhope	C-54327	3.82005434082	0.0
	AL	Baldwin	Foley	C-245979	0.250121021251	167.0
	AL	Baldwin	Gulf Shores	C-131567	0.725136957517	104.0

1 document selected

9. Pronaci šifru radova na kojima se odradilo najviše milja u tom gradu, po okrugu, po saveznoj državi.

Vreme izvršavanja: 46s

Primer agregacije

The screenshot shows the IntelliJShell IDE interface. The top panel displays a MongoDB aggregation query in the script editor. The query is as follows:

```
41 avg_wind_speed(mph): { $avg: "$avg_wind_speed(mph)" },
42 "avg_precipitation(in)": { $avg: "$Precipitation(in)" }
43 },
44 },
45 },
46 { $sort: {
47   "_id": 1
48 } },
49 },
50 { $project: {
51   "_id": 0,
52   "state": 1,
53   "county": 1,
54   "city": 1,
55   "description": 1,
56   "avg_miles": { $round: ["$avg_miles", 1] },
57   "avg_days_worked": { $round: ["$avg_days_worked", 1] },
58   "avg_temp(F)": { $round: ["$avg_temp(F)", 1] },
59   "avg_wind_chill(F)": { $round: ["$avg_wind_chill(F)", 1] },
60   "avg_temp_difference(F)": { $round: ["$avg_temp_difference(F)", 1] },
61   "avg_humidity(%)": { $round: ["$avg_humidity(%)", 1] },
62   "avg_pressure(in)": { $round: ["$avg_pressure(in)", 1] },
63   "avg_visibility(mi)": { $round: ["$avg_visibility(mi)", 1] },
64   "avg_wind_speed(mph)": { $round: ["$avg_wind_speed(mph)", 1] },
65   "avg_precipitation(in)": { $round: ["$avg_precipitation(in)", 1] },
66 } },
67 }
```

The bottom panel shows the results of the query in a table view. The table has 8 columns: `_id`, `state`, `county`, `city`, `description`, `avg_miles`, `avg_days_worked`, and `avg_temp(F)`. The first row is selected, showing data for Alabama (AL), Baldwin County, and Fairhope.

_id	state	county	city	description	avg_miles	avg_days_worked	avg_temp(F)
123	AL	Baldwin	Fairhope	Closed road from N Section St to N Summit St due to roadwork.	0.2	9.0	55.0
123	AL	Baldwin	Orange Beach	Closed road in Perdido Beach Blvd due to roadwork.	0.1	7.3	77.7
123	AL	Colbert	Leighton	Closed road from Turkey Farm Rd to Colburn Mountain Rd due to roadwork.	1.6	0.0	33.0
123	AL	Colbert	Tuscumbia	Closed road from Spring Valley Rd to Tri-Cities due to roadwork.	0.6	4.5	39.5
123	AL	Elmore	Wetumpka	Closed road from Love Ln to US Highway 231 due to roadwork.	0.1	0.0	76.0
123	AL	Etowah	Boaz	Closed road from Crickett Ln to Joe Whitt Dr due to roadwork.	0.2	379.0	52.0
123	AL	Houston	Dothan	Closed road from Ross Clark Cir to W Main St due to roadwork.	0.1	52.4	77.5

The status bar at the bottom indicates "1 document selected" and "Count Documents".

10. Koja je ukupna dužina radova na autoputu za svaki grad?

Vreme izvršavanja: 10s

Logička šema
prilagođena
agregacija



Logička šema prilagođena agregacija

Početna kolekcija je podeljena na dve nove “roadsPrimary” i “roadsSecondary”

```
{
  "_id" : ObjectId("6486b94fd75c133bf6ffe678"),
  "Severity" : NumberInt(4),
  "Start_Time" : ISODate("2019-04-05T16:00:00.000+0000"),
  "End_Time" : ISODate("2020-09-29T11:53:57.000+0000"),
  "Start_Lat" : 32.838359999999994,
  "Start_Lng" : -93.152378,
  "End_Lat" : 32.85074,
  "End_Lng" : -93.164388,
  "Distance(mi)" : 1.10349651559,
  "Description" : "Construction on LA-534 WB near EDMONDS LOOP Road closed. Take alternate route.",
  "Number" : NumberInt(4200),
  "Street" : "Highway 534",
  "Side" : "R",
  "City" : "Haynesville",
  "County" : "Claiborne",
  "State" : "LA",
  "Zipcode" : "71038-7130",
  "Country" : "US",
  "Timezone" : "US/Central",
  "Airport_Code" : "KMNE",
  "Weather_Timestamp" : ISODate("2019-04-05T15:55:00.000+0000"),
  "Temperature(F)" : NumberInt(75),
  "Wind_Chill(F)" : NumberInt(75),
  "Humidity(%)" : NumberInt(58),
  "Pressure(in)" : NumberInt(29),
  "Visibility(mi)" : NumberInt(10),
  "Wind_Direction" : "S",
  "Wind_Speed(mph)" : NumberInt(3),
  "Precipitation(in)" : NumberInt(0),
  "Weather_Condition" : "Fair",
  "Start_day" : NumberInt(5),
  "Start_month" : NumberInt(4),
  "Start_year" : NumberInt(2019),
  "End_day" : NumberInt(29),
  "End_month" : NumberInt(9),
  "End_year" : NumberInt(2020),
  "Days_worked" : NumberInt(534),
  "ID" : "C-1"
}
```

```
{
  "_id" : ObjectId("6486d3f1caef744b7e429b36"),
  "Amenity" : false,
  "Bump" : false,
  "Crossing" : false,
  "Give_Way" : false,
  "Junction" : false,
  "No_Exit" : false,
  "Railway" : false,
  "Roundabout" : false,
  "Station" : false,
  "Stop" : false,
  "Traffic_Calming" : false,
  "Traffic_Signal" : false,
  "Turning_Loop" : false,
  "Sunrise_Sunset" : "Day",
  "Civil_Twilight" : "Day",
  "Nautical_Twilight" : "Day",
  "Astronomical_Twilight" : "Day",
  "ID" : "C-1"
}
```

Logička šema prilagođena agregacija

- Dodato polje Days_worked
- Dodato polje location
- Dodat Geospatial indeks
- Dodat indeks nad poljem Days_worked i Temperature(F)

Logička šema prilagođena agregacija

Dodato polje Days_worked i location

- Polje Days_worked je dodato na osnovu polja Start_year, Start_month, Start_day i End_year, End_month, End_day
- Polje location:

```
1 db.roadsPrimary.aggregate([
2   {
3     $addFields: {
4       location: {
5         type: "Point",
6         coordinates: [
7           "$Start_Lng",
8           "$Start_Lat"
9         ]
10      }
11    },
12  ],
13  {
14    $out: "roadsPrimary"
15  }
16 ])
17
```

Logička šema prilagođena agregacija

Dodat Geospatial indeks, indeks nad "Days_worked", "State" i "Temperature(F)"

```
1 db.roadsPrimary.createIndex({ location: "2dsphere" })
```

```
2
```

```
3
```

```
db.roadsPrimary.createIndex({ Days_worked: -1 })
```

```
1 db.roadsPrimary.createIndex({ State: -1 })
```

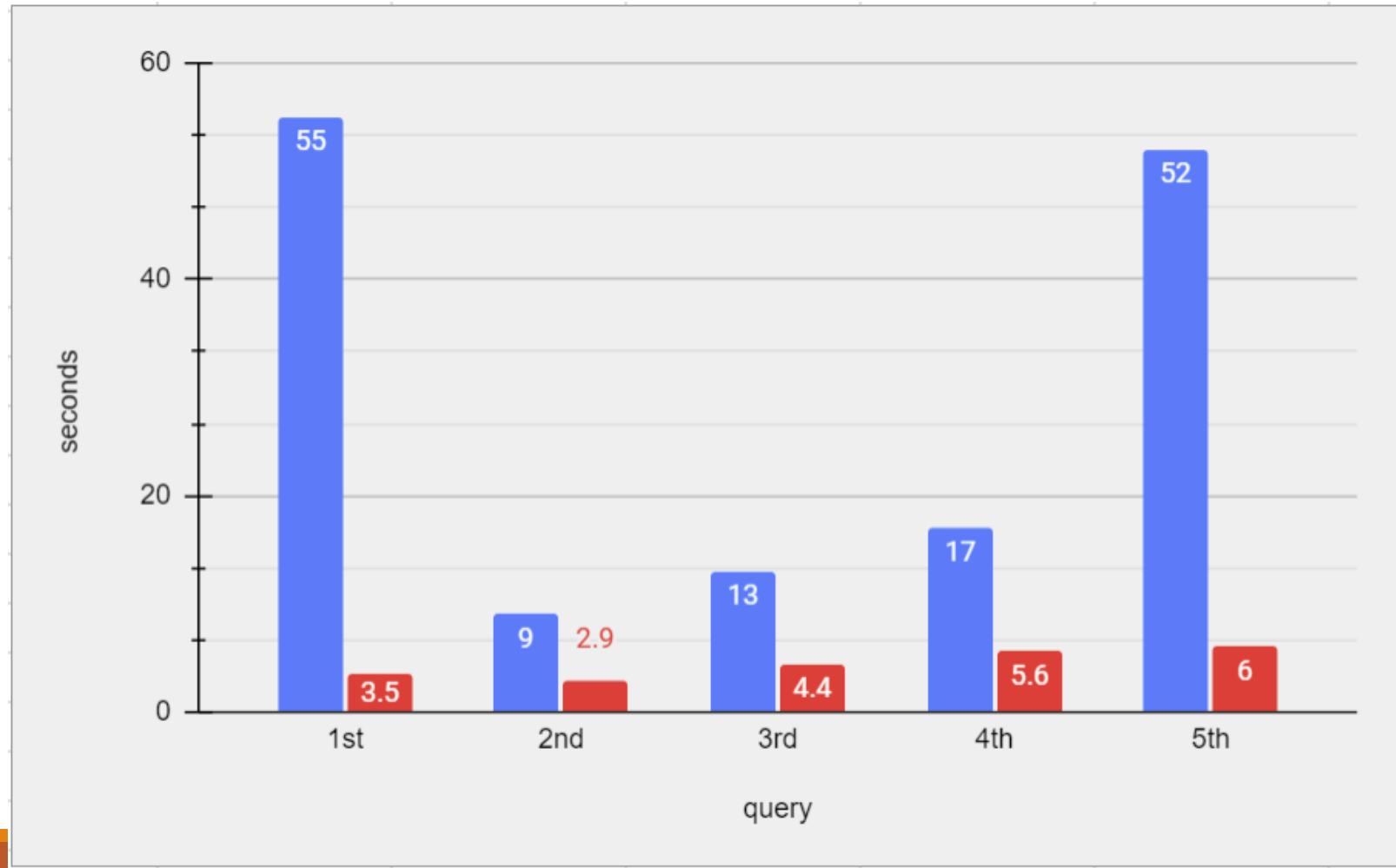
```
2
```

```
db.roadsPrimary.createIndex({ Temperature(F) : -1 })
```

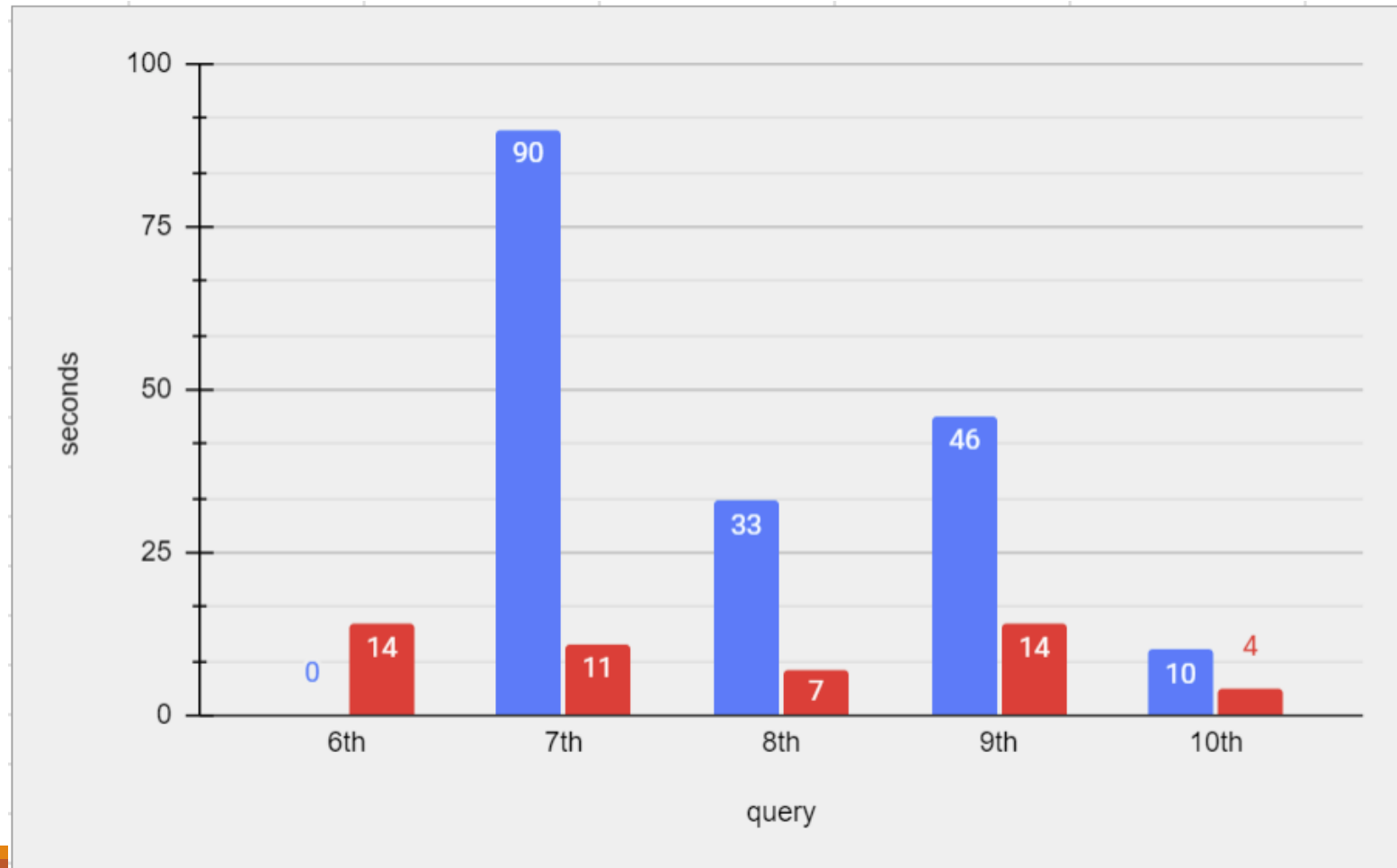
Poređenje performansi



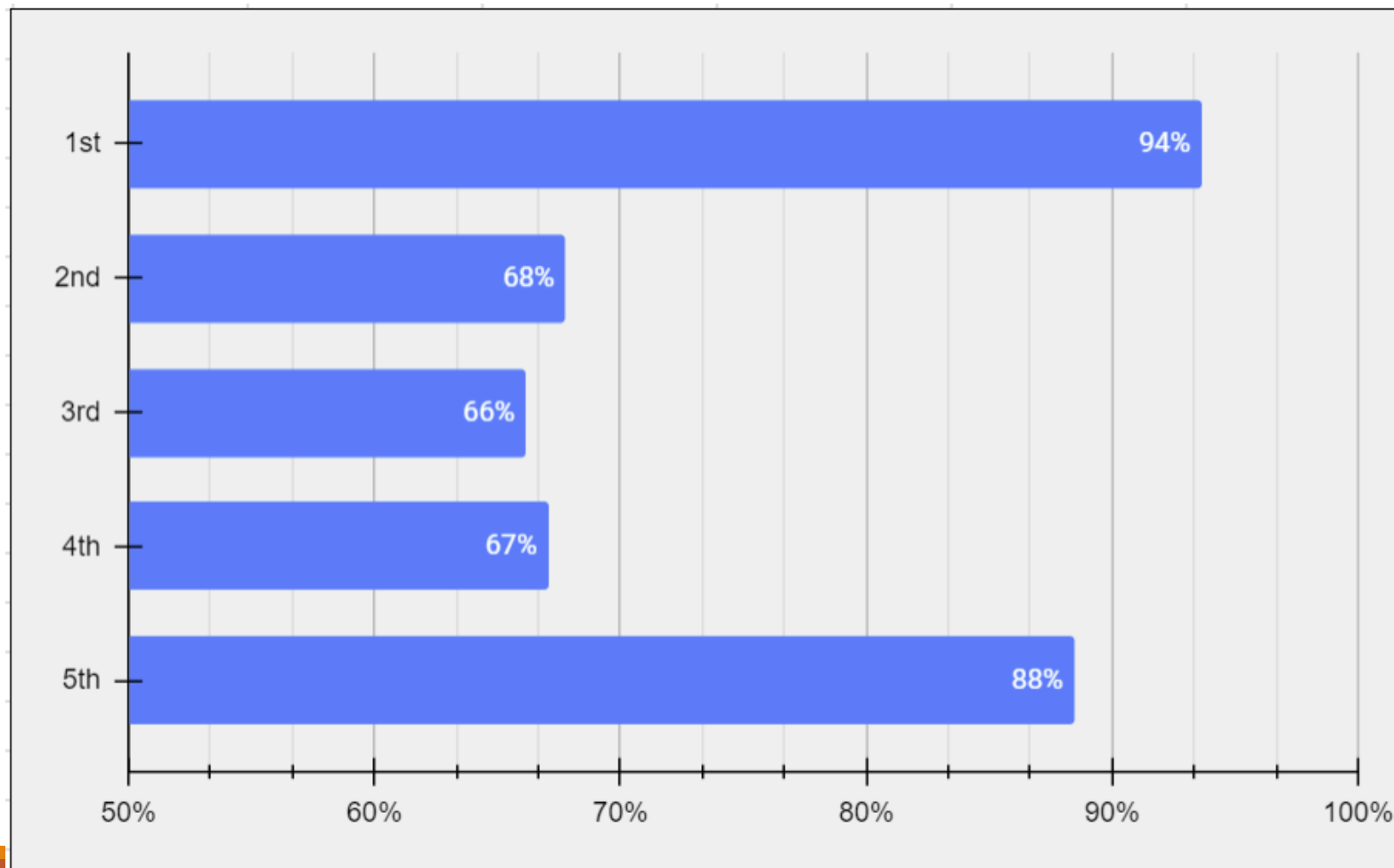
Poređenje performansi – u sekundama



Poređenje performansi – u sekundama



Poređenje performansi – u procentima



Poređenje performansi – u procentima

