

# Birla Institute of Technology & Science, Pilani

## Work Integrated Learning Programmes

### Division First Semester 2022-2023

#### Assignment I

Course No.: DSECL ZG522

Course Title: Big Data Systems

---

#### Group 93

Sr.No.	Name	BITS ID	Contribution (%)
1	Sherwin Philip	2021FC04106	100
2	Avinash Kumar	2021FC04746	100
3	Sidharth vij	2021fc04507	100
4	Saurabh Arunrao Dhande	2021fc04700	100

Sr.No.	Software Used	Version	Comment
1	MongoDBCompass	MongoDB 3.6	MongoDB Compass is a powerful GUI for querying, aggregating, and analyzing your MongoDB data in a visual environment
2	Studio 3T	Studio 3T 2022.10.0	Studio 3T is made for growing professional teams, offering a variety of ways to view and interrogate data collections, including sophisticated aggregations, native Mongo JSON extensions, traditional SQL queries, and a drag and drop query builder.

1]The number of new cases, new deaths and new recovered

Answer - 228693, 5415, 174623

```
db = db.getSiblingDB("BITSMONGODB");
```

```
db.getCollection("Country").aggregate(
```

```
[
```

```

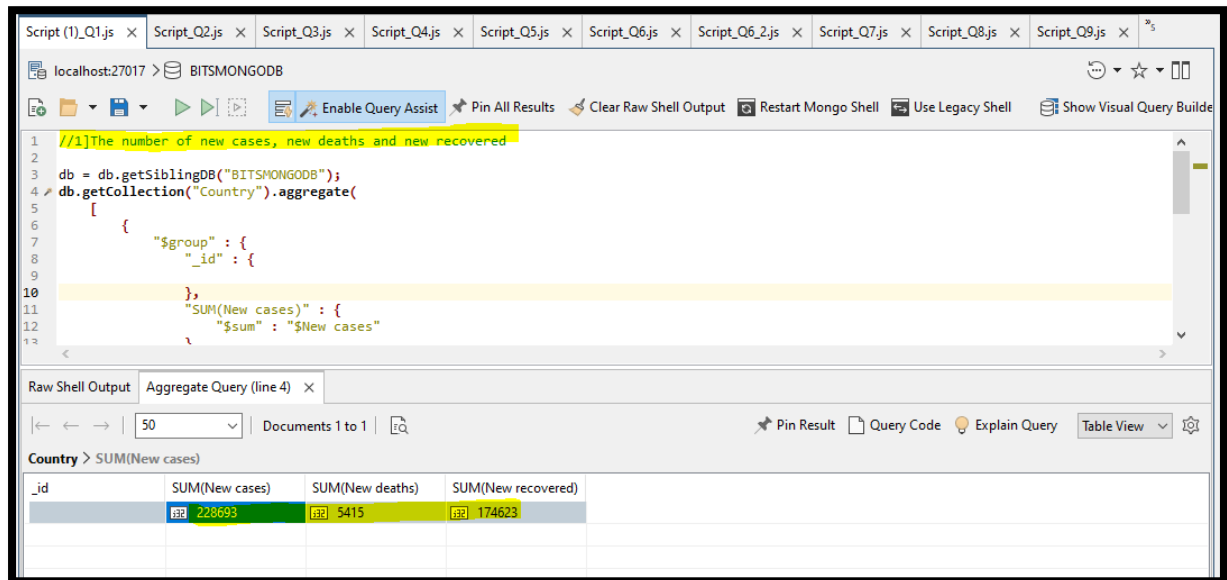
{
  "$group" : {
    "_id" : {

    },
    "SUM(New cases)" : {
      "$sum" : "$New cases"
    },
    "SUM(New deaths)" : {
      "$sum" : "$New deaths"
    },
    "SUM(New recovered)" : {
      "$sum" : "$New recovered"
    }
  }
},
{
  "$project" : {
    "SUM(New cases)" : "$SUM(New cases)",
    "SUM(New deaths)" : "$SUM(New deaths)",
    "SUM(New recovered)" : "$SUM(New recovered)",
    "_id" : NumberInt(0)
  }
}
],
{
  "allowDiskUse" : true
}

```

```
}  
);
```

### Screenshot:



2]The number of death cases in each country of continent Asia and also the corresponding WHO regions

### Answer:

```
db = db.getSiblingDB("BITSMONGODB");
```

```
db.getCollection("World").find(  
  
  {  
  
    "Continent" : "Asia"  
  
  },  
  
  {  
  
    "Country/Region" : "$Country/Region",  
  
    "WHO Region" : "$WHO Region",  
  
    "TotalDeaths" : "$TotalDeaths",  
  
    "_id" : NumberInt(0)
```

```
}  
);
```

## Screenhot:

The screenshot shows the MongoDB Shell interface with a query and its results. The query is as follows:

```
//2]The number of death cases in each country of continent Asia and also the corresponding WHO regions  
db = db.getSiblingDB("BITSMONGODB");  
db.getCollection("World").find(  
  {  
    "Continent" : "Asia"  
  },  
  {  
    "Country/Region" : "$Country/Region",  
    "WHO Region" : "$WHO Region",  
    "TotalDeaths" : "$TotalDeaths",  
    "_id" : NumberInt(0)  
  })
```

The results are displayed in a table with the following columns: Country/Region, WHO Region, and TotalDeaths. The results are as follows:

Country/Region	WHO Region	TotalDeaths
India	South-EastAsia	41638
Iran	EasternMediterranean	17976
Saudi Arabia	EasternMediterranean	3055
Pakistan	EasternMediterranean	6035
Bangladesh	South-EastAsia	3306
Turkey	Europe	5798
Iran	EasternMediterranean	5161

## 3]The number of deaths that occurred on 12-02-2020

### Answer - 1118

```
db = db.getSiblingDB("BITSMONGODB");
```

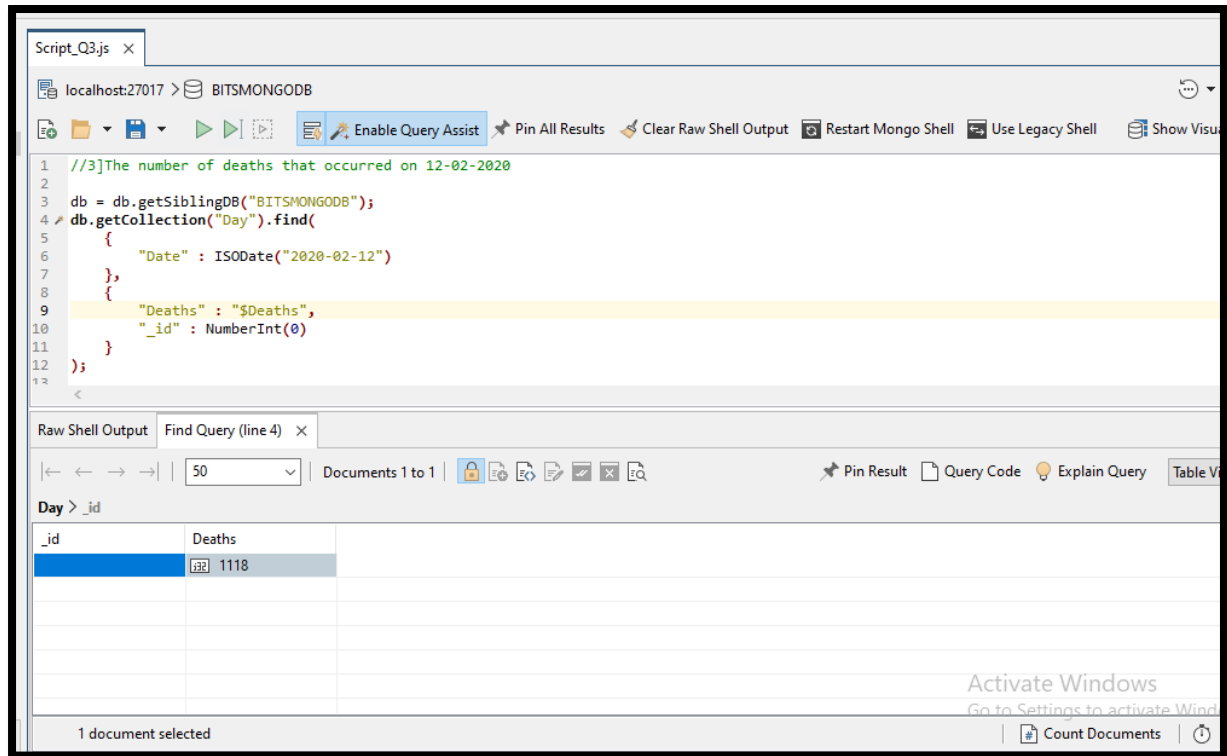
```
db.getCollection("Day").find(  
  {  
    "Date" : ISODate("2020-02-12")  
  },  
  {  
    "Deaths" : "$Deaths",  
    "_id" : NumberInt(0)  
  })
```

```

    "_id" : NumberInt(0)
  }
);

```

**Screenshot:**



**4]The number of active new cases (new cases-(new death+new recovered)) in a reverse sorted order based on the country name**

**Answer:**

```

db = db.getSiblingDB("BITSMONGODB");
db.getCollection("Country").aggregate(
[
  {
    "$group" : {
      "_id" : {

```

```

    "Country/Region" : "$Country/Region"
  },
  "SUM(New deaths)" : {
    "$sum" : "$New deaths"
  },
  "SUM(New recovered)" : {
    "$sum" : "$New recovered"
  },
  "SUM(New cases)" : {
    "$sum" : "$New cases"
  }
}
},
{
  "$project" : {
    "Country/Region" : "$_id.Country/Region",
    "SUM(New deaths)" : "$SUM(New deaths)",
    "SUM(New recovered)" : "$SUM(New recovered)",
    "SUM(New cases)" : "$SUM(New cases)",
    'ActiveNewCases': { '$subtract': [ '$SUM(New cases)', { '$add': [ '$SUM(New
deaths)', '$SUM(New recovered)' ] } ] } },
    "_id" : NumberInt(0)
  }
},

{ "$sort":
{

```

```

        "Country/Region" : NumberInt(-1)
    }
}

],

{
    "allowDiskUse" : true
}

);

```

### Explanation:

The number of active new cases (new cases-(new death+new recovered)) in a reverse sorted order based on the country name

Observation - The number of active new cases are showing up as negative for few countries. This is because the sum of "new death" and "new recovered" is more than "new cases".

### Screenshot:

The screenshot shows a MongoDB Shell window with the following aggregate query:

```

//4]The number of active new cases (new cases-(new death+new recovered)) in a reverse sorted order based on the country name
db = db.getSiblingDB("BITSMONGODB");
db.getCollection("Country").aggregate(
{
  "$group" : {
    "_id" : {
      "Country/Region" : "$Country/Region"
    },
    "SUM(New deaths)" : {
      "$sum" : "$New deaths"
    }
  }
}
)

```

The results are displayed in a table view with the following columns: Country/Region, SUM(New deaths), SUM(New recovered), SUM(New cases), and ActiveNewCases. The data is sorted in descending order by Country/Region.

Country/Region	SUM(New deaths)	SUM(New recovered)	SUM(New cases)	ActiveNewCases
Zimbabwe	2	24	192	166
Zambia	1	465	71	-395
Yemen	4	36	10	-30
Western Sahara	0	0	0	0
West Bank and...	2	0	152	150
Vietnam	0	0	11	11
Venezuela	4	212	525	309

1 document selected

---

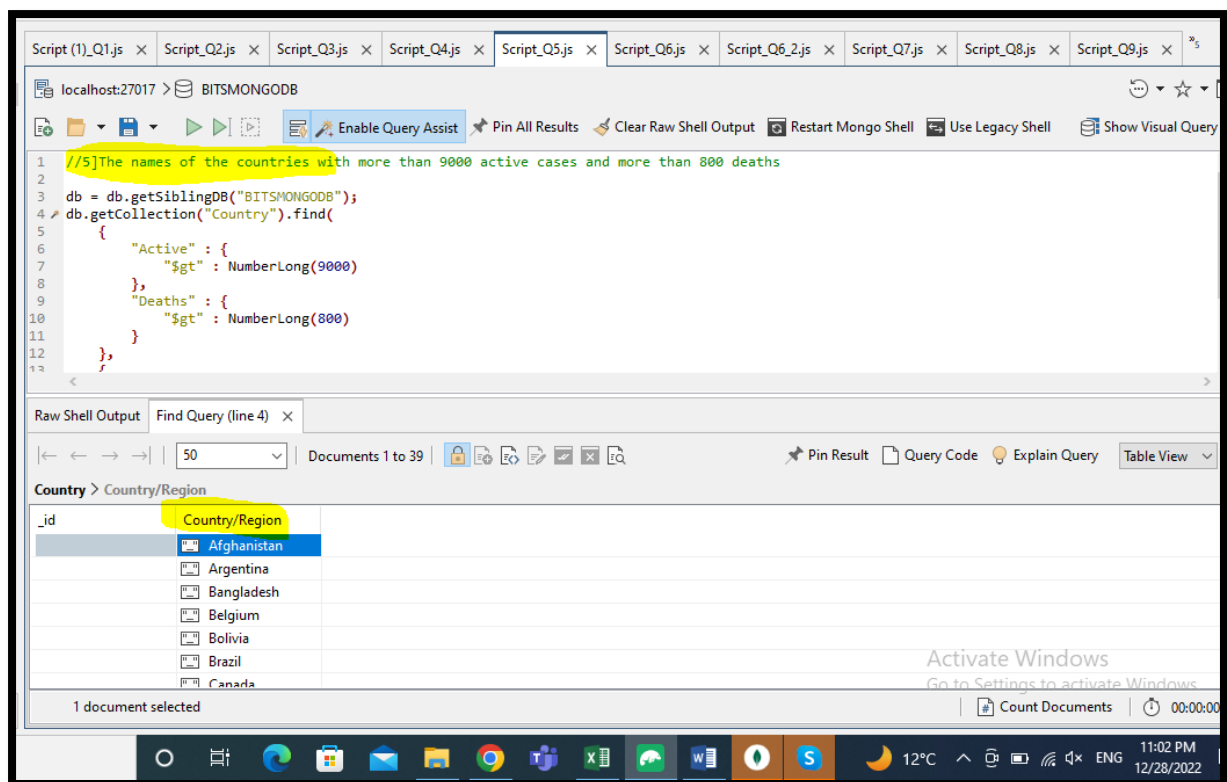
**5]The names of the countries with more than 9000 active cases and more than 800 deaths**

**Answer:**

```
db = db.getSiblingDB("BITSMONGODB");
db.getCollection("Country").find(
  {
    "Active" : {
      "$gt" : NumberLong(9000)
    },
    "Deaths" : {
      "$gt" : NumberLong(800)
    }
  },
  {
    "Country/Region" : "$Country/Region",
    "_id" : NumberInt(0)
  }
);
```

**Screenshot :**





**6] The country with the highest number of active cases and also with second highest death rate**

**Answer:** - US, UNITED KINGDOM

## HIGHEST ACTIVE

```
db = db.getSiblingDB("BITSMONGODB");
```

```
db.getCollection("Country").find({},
```

```
{
```

```
  "Country/Region" : "$Country/Region",
```

```
  "_id" : NumberInt(0)
```

```
}
```

```
).sort(
```

```
{
```

```
        "Active" : NumberInt(-1)
    }
).limit(1);
```

## **SECOND HIGHEST DEATH RATE**

```
db = db.getSiblingDB("BITSMONGODB");
db.getCollection("Country").find({
    {
        "Country/Region" : "$Country/Region",
        "_id" : NumberInt(0)
    }
}).sort(
    {
        "Deaths / 100 Cases" : NumberInt(-1)
    }
).skip(1).limit(1);
```

**Screenshot:**

## **HIGHEST ACTIVE**

Script (1)\_Q1.js x Script\_Q2.js x Script\_Q3.js x Script\_Q4.js x Script\_Q5.js x Script\_Q6.js x Script\_Q6\_2.js x Script\_Q7.js x Script\_Q8.js x Script\_Q9.js x

localhost:27017 > BITSMONGODB

Enable Query Assist Pin All Results Clear Raw Shell Output Restart Mongo Shell Use Legacy Shell Show Visual Query

```
1 //6] The country with the highest number of active cases and also with second highest death rate
2
3 db = db.getSiblingDB("BITSMONGODB");
4 db.getCollection("Country").find({},
5   {
6     "Country/Region" : "$Country/Region",
7     "_id" : NumberInt(0)
8   }
9 ).sort(
10  {
11    "Active" : NumberInt(-1)
12  }
13 ).limit(1);
```

Raw Shell Output Find Query (line 4) x

50 Documents 1 to 1 Pin Result Query Code Explain Query Table View

Country > Country/Region

_id	Country/Region
US	US

1 document selected

Count Documents 00:00

11:06 PM 12/28/2022

## SECOND HIGHEST DEATH RATE

Script (1)\_Q1.js x Script\_Q2.js x Script\_Q3.js x Script\_Q4.js x Script\_Q5.js x Script\_Q6.js x Script\_Q6\_2.js x Script\_Q7.js x Script\_Q8.js x Script\_Q9.js x

localhost:27017 > BITSMONGODB

Enable Query Assist Pin All Results Clear Raw Shell Output Restart Mongo Shell Use Legacy Shell Show Visual Query

```
1 //6] The country with the highest number of active cases and also with second highest death rate
2
3 db = db.getSiblingDB("BITSMONGODB");
4 db.getCollection("Country").find({},
5   {
6     "Country/Region" : "$Country/Region",
7     "_id" : NumberInt(0)
8   }
9 ).sort(
10  {
11    "Deaths / 100 Cases" : NumberInt(-1)
12  }
13 ).limit(1);
```

Raw Shell Output Find Query (line 4) x

50 Documents 1 to 1 Pin Result Query Code Explain Query Table View

Country > Country/Region

_id	Country/Region
United Kingdom	United Kingdom

1 document selected

Count Documents 00:00

11:07 PM 12/28/2022

**7] The total number of deaths all around the world**

**Answer - 654036**

```
db = db.getSiblingDB("BITSMONGODB");
```

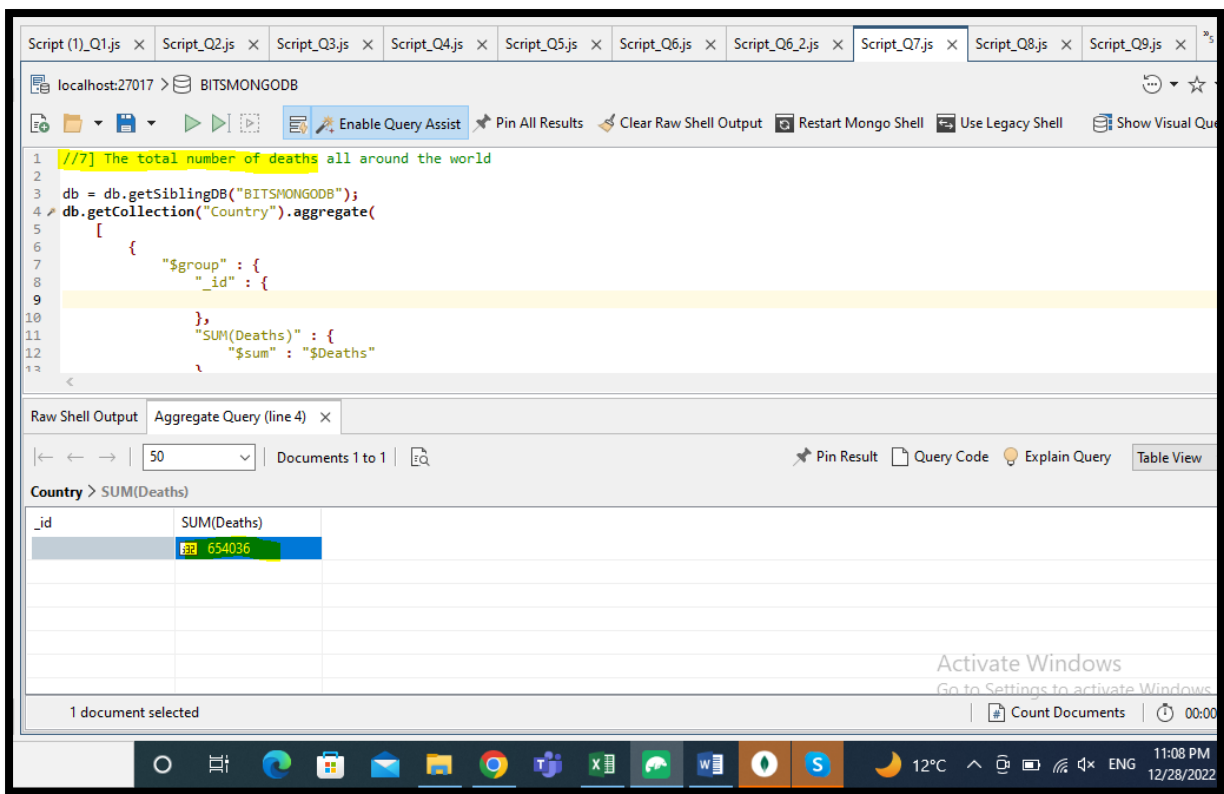
```
db.getCollection("Country").aggregate(
```

```
[
  {
    "$group" : {
      "_id" : {

      },
      "SUM(Deaths)" : {
        "$sum" : "$Deaths"
      }
    }
  },
  {
    "$project" : {
      "SUM(Deaths)" : "$SUM(Deaths)",
      "_id" : NumberInt(0)
    }
  }
],
{
  "allowDiskUse" : true
}
```

);

## Screenshot:



## 8] The number of death cases and active cases between 28-01-2020 and 21-02-2020

Answer- 2168, 73279

```
db = db.getSiblingDB("BITSMONGODB");
```

```
db.getCollection("Day").aggregate([
```

```
[
  {
    "$match": {
      "Date": {$gte:ISODate("2020-01-28"),$lte:ISODate("2020-02-21")}
    }
  },
])
```

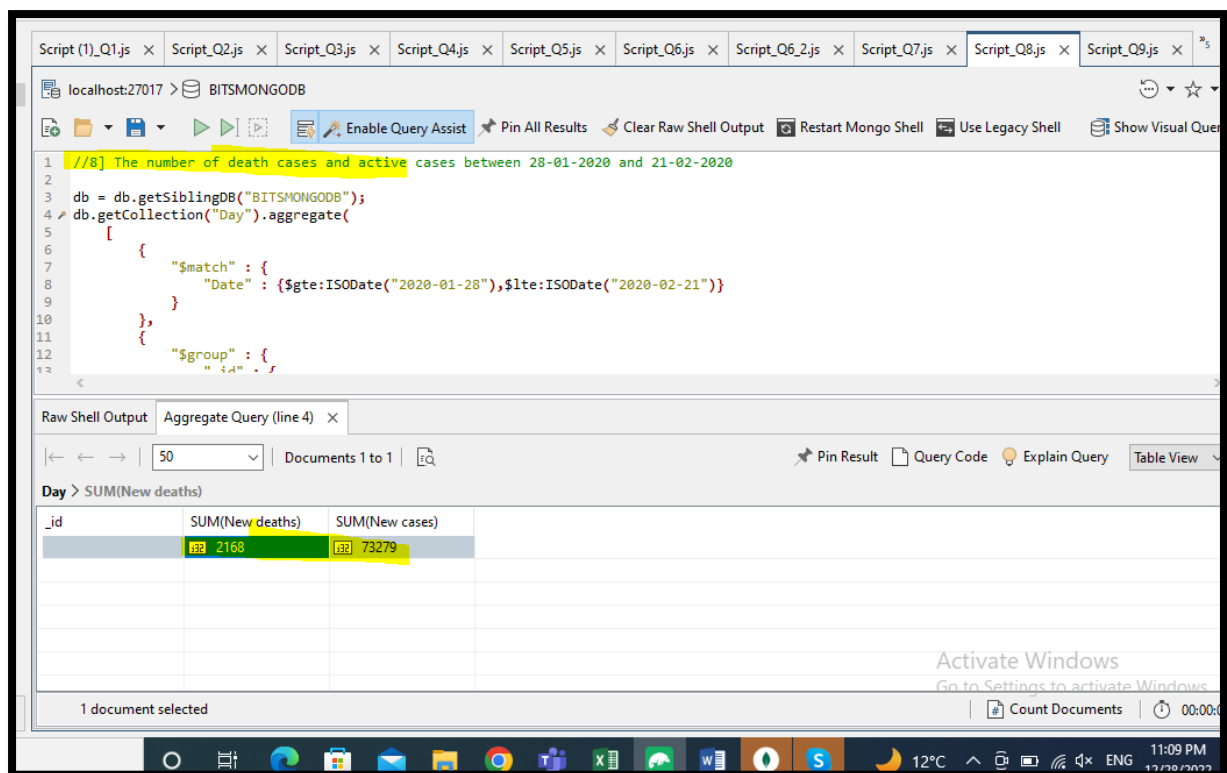
```

{
  "$group" : {
    "_id" : {

    },
    "SUM(New deaths)" : {
      "$sum" : "$New deaths"
    },
    "SUM(New cases)" : {
      "$sum" : "$New cases"
    }
  }
},
{
  "$project" : {
    "SUM(New deaths)" : "$SUM(New deaths)",
    "SUM(New cases)" : "$SUM(New cases)",
    "_id" : NumberInt(0)
  }
}
],
{
  "allowDiskUse" : true
}
);

```

**Screenshot:**



## 9] The latitude and longitude of countries ending with “ia” and the number of countries

**Answer:**

```
db = db.getSiblingDB("BITSMONGODB");
```

```
db.getCollection("Covid19").find(
```

```
{
```

```
  "Country/Region" : /^.*ia$/i
```

```
},
```

```
{
```

```
  "Country/Region" : "$Country/Region",
```

```
  "Lat" : "$Lat",
```

```
  "Long" : "$Long",
```

```
  "_id" : NumberInt(0)
```

```
}
```

);

---

Answer part 2 - 38

```
db = db.getSiblingDB("BITSMONGODB");
db.getCollection("Covid19").aggregate(
[
  {
    "$match" : {
      "Country/Region" : /^.*ia$/i
    }
  },
  {
    "$group" : {
      "_id" : {

      },
      "COUNT_DISTINCT(Country/Region)" : {
        "$addToSet" : "$Country/Region"
      }
    }
  },
  {
    "$project" : {
      "COUNT(Country/Region)" : { $size: "$COUNT_DISTINCT(Country/Region)" },
      "_id" : NumberInt(0)
    }
  }
]
```



```
        }  
    }  
],  
{  
    "allowDiskUse" : true  
}  
);
```

**Explanation:**

Please note that "country" is not the lowest level of hierarchy in this data. We have "province/state" which is one level lower in hierarchy compared to "country". Hence, you will see there are multiple records for countries such as "Australia" that has many "Lat" and "Long" values. This is because "Australia" has many data "province/state" each with its own "Lat" and "Long" values.

Another point to note here is that when we count such number of "countries" ending with "ia", we should do a distinct count of such countries and not take in to account "province/state" column.

**Screenshot:****Part 1:**

Script (1)\_Q1.js x Script\_Q2.js x Script\_Q3.js x Script\_Q4.js x Script\_Q5.js x Script\_Q6.js x Script\_Q6\_2.js x Script\_Q7.js x Script\_Q8.js x Script\_Q9.js x

localhost:27017 > BITSMONGODB

Enable Query Assist Pin All Results Clear Raw Shell Output Restart Mongo Shell Use Legacy Shell Show Visual Query

```

1 //9] The latitude and longitude of countries ending with "ia" and the number of countries
2
3 db = db.getSiblingDB("BITSMONGODB");
4 db.getCollection("Covid19").find(
5   {
6     "Country/Region" : /.*ia$/i
7   },
8   {
9     "Country/Region" : "$Country/Region",
10    "Lat" : "$Lat",
11    "Long" : "$Long",
12    "_id" : NumberInt(0)
13  }

```

Raw Shell Output Find Query (line 4) x

Documents 1 to 50

Covid19 > Country/Region

_id	Country/Region	Lat	Long
	Albania	41.1533	20.1683
	Algeria	28.0339	1.6596
	Armenia	40.0691	45.0382
	Australia	-35.4735	149.0124
	Australia	-33.8688	151.2093
	Australia	-12.4634	130.8456
	Australia	-27.4608	153.0251

1 document selected

Count Documents 00:00

Activate Windows  
Go to Settings to activate Windows.

12°C 11:13 PM 12/28/2022

## Part 2:

Script\_Q2.js x Script\_Q3.js x Script\_Q4.js x Script\_Q5.js x Script\_Q6.js x Script\_Q6\_2.js x Script\_Q7.js x Script\_Q8.js x Script\_Q9.js x Script\_Q9\_2.js x

localhost:27017 > BITSMONGODB

Enable Query Assist Pin All Results Clear Raw Shell Output Restart Mongo Shell Use Legacy Shell Show Visual Query

```

1 //9] The latitude and longitude of countries ending with "ia" and the number of countries
2
3 db = db.getSiblingDB("BITSMONGODB");
4 db.getCollection("Covid19").aggregate(
5   [
6     {
7       "$match" : {
8         "Country/Region" : /.*ia$/i
9       }
10    },
11    {
12      "$group" : {
13        "_id" : "COUNT(Country/Region)"
14      }
15    }
16  ]

```

Raw Shell Output Aggregate Query (line 4) x

Documents 1 to 1

Covid19 > COUNT(Country/Region)

_id	COUNT(Country/Region)
	38

1 document selected

Count Documents 00:00

Activate Windows  
Go to Settings to activate Windows.

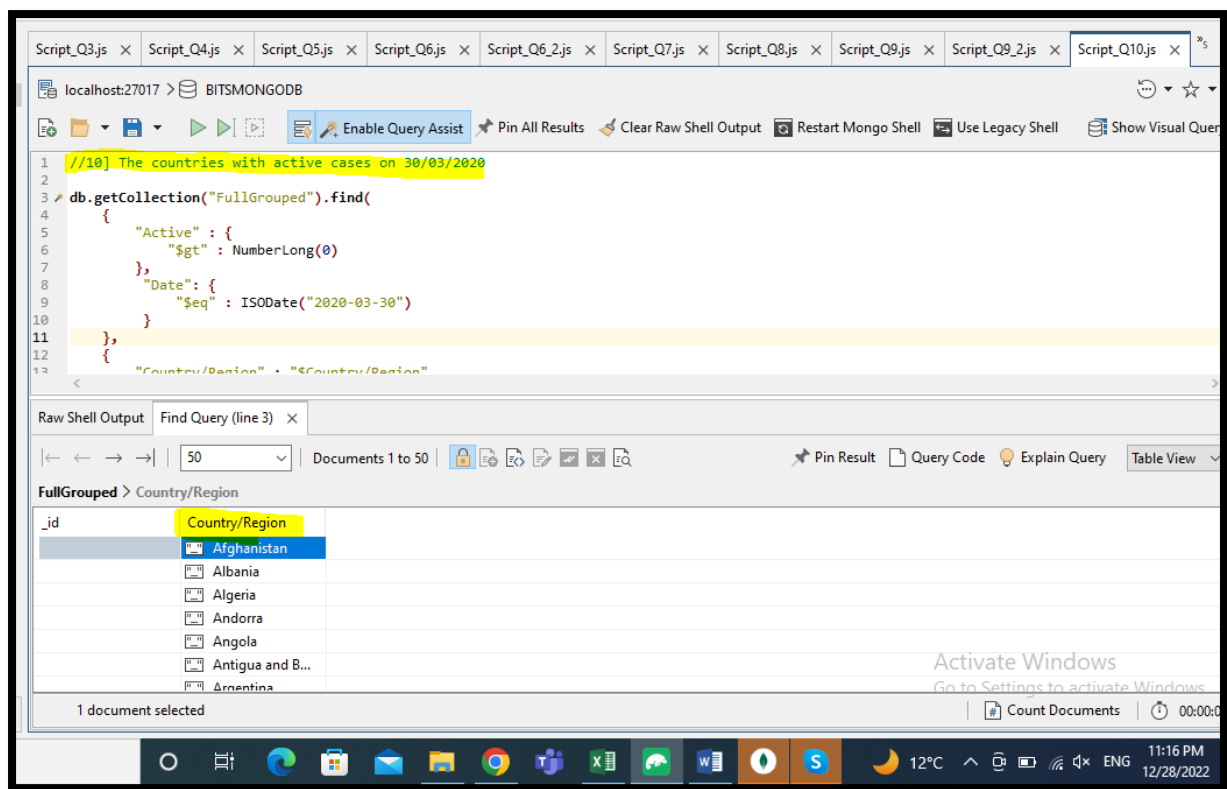
12°C 11:14 PM 12/28/2022

### 10] The countries with active cases on 30/03/2020

**Answer:**

```
db.getCollection("FullGrouped").find(
{
  "Active" : {
    "$gt" : NumberLong(0)
  },
  "Date": {
    "$eq" : ISODate("2020-03-30")
  }
},
{
  "Country/Region" : "$Country/Region",
  "_id" : NumberInt(0)
}
);
```

**Screenshot:**



**11] The latitude and longitude of those countries which are having active cases greater than 100**

**Answer:**

```
db = db.getSiblingDB("BITSMONGODB");
```

```
db.getCollection("Covid19").find(
```

```
{
```

```
  "Active" : {
```

```
    "$gt" : NumberLong(100)
```

```
  }
```

```
},
```

```
{
```

```
  "Country/Region" : "$Country/Region",
```

```
  "Province/State" : "$Province/State",
```

```

    "Lat" : "$Lat",

    "Long" : "$Long",

    "_id" : NumberInt(0)

}

);

```

### Screenshot:

The screenshot shows a MongoDB Shell interface with the following query and results:

```

1 //11] The latitude and longitude of those countries which are having active cases greater than 100
2
3 db = db.getSiblingDB("BITSMONGODB");
4 db.getCollection("Covid19").find(
5   {
6     "Active" : {
7       "$gt" : NumberLong(100)
8     }
9   },
10  {
11    "Country/Region" : "$Country/Region",
12    "Province/State" : "$Province/State",
13    "Lat" : "$Lat",
14    "Long" : "$Long"
15  }
16 )

```

The results are displayed in a table view with the following columns: Country/Region, Province/State, Lat, and Long. The table shows 7 documents, all from China, with various provinces and states listed. The first document is highlighted in blue.

Country/Region	Province/State	Lat	Long
China	Hubei	30.9756	112.2707
China	Hubei	30.9756	112.2707
China	Hubei	30.9756	112.2707
China	Hubei	30.9756	112.2707
China	Guangdong	23.3417	113.4244
China	Hubei	30.9756	112.2707
China	Zhejiang	29.1832	120.0934

1 document selected

12]The countries and respective dates in which maximum increase of active cases occurred.

Answer:

```

db = db.getSiblingDB("BITSMONGODB");

db.getCollection("FullGrouped").aggregate(

[

```

```

{
  $setWindowFields: {
    partitionBy: "$Country/Region",
    sortBy: { Date: 1 },
    output: {
      lag: {
        $shift: {
          output: "$Active",
          by: -1,
          default: NumberInt(0)
        }
      },
    }
  }
}

,

{
  "$project" : {
    "Country/Region" : "$Country/Region",
    "Date" : "$Date",
    'ActiveNewCases': { '$subtract': ['$Active', '$lag']},
    "_id" : NumberInt(0)
  }
},

```

```
{
$group: {
  "_id": "$Country/Region",
  "max_active_new_cases": {
    $max: "$ActiveNewCases"
  },
  "records": {
    $push: "$$ROOT"
  }
}
},
{
  "$project": {
    items: {
      "$filter": {
        "input": "$records",
        "as": "records",
        "cond": {
          $eq: [
            "$$records.ActiveNewCases",
            "$max_active_new_cases"
          ]
        }
      }
    }
  }
}
```

```

    }
  },
  {
    $unwind: "$items"
  },
  {
    $replaceWith: "$items"
  }
],
{
  "allowDiskUse" : true
}
);

```

### **Explanation:**

This is a complex question. To solve this question, we need to find the increase of cases on a each date with respect to the previous date. Once we have this increase, we should find the max value of this increase for each country and show the result.

Step 1 - first we do a "lag" on "active" cases.

Step 2 - subtract "active" cases and cases on previous date - "lag" value.

Step 3 - group the data based country and find the max value of the subtracted value we recieved in step 2

### **Screenshot:**



Script\_Q5.js x Script\_Q6.js x Script\_Q6\_2.js x Script\_Q7.js x Script\_Q8.js x Script\_Q9.js x Script\_Q9\_2.js x Script\_Q10.js x Script\_Q11.js x Script\_Q12.js x

localhost:27017 > BITSMONGODB

Enable Query Assist Pin All Results Clear Raw Shell Output Restart Mongo Shell Use Legacy Shell Show Visual Query

```
1 //12]The countries and respective dates in which maximum increase of active cases occurred.
2
3 db = db.getSiblingDB("BITSMONGODB");
4 db.getCollection("FullGrouped").aggregate(
5   [
6
7     {
8       $setWindowFields: {
9         partitionBy: "$Country/Region",
10        sortBy: { Date: 1 },
11        output: {
12          lag: {
13            $shift: 1
```

Raw Shell Output Aggregate Query (line 4) x

← → | 50 | Documents 1 to 50 | 🔍 Pin Result Query Code Explain Query Table View

FullGrouped > Country/Region

_id	Country/Region	Date	ActiveNewCases
	Somalia	2020-05-12T00:...	76
	Yemen	2020-06-15T00:...	46
	Namibia	2020-07-18T00:...	125
	Gambia	2020-07-25T00:...	61
	Malaysia	2020-06-04T00:...	249
	Peru	2020-07-27T00:...	8484
	Philippines	2020-07-08T00:...	2279

1 document selected

Count Documents 00:00:0

Activate Windows  
Go to Settings to activate Windows.

12°C 11:20 PM 12/28/2022