Table of Contents

1 Task A

1.1 Importing data

'Importing file using "mongoimport -d assignment1 -c RepublicaNepal –type json –file RepublicaNepal.json" command.

Starting mongo using "mongo" command.

Displaying databases using "show dbs" command.

Using database named "assignment1"

```
> show collections
RepublicaNepal
```

Displaying collections using "show collections".

```
> db.RepublicaNepal.count()
510
```

2 Task B

2.1 Analyzing the data

Displaying numbers of documents using "db.commandName.count()" in my collection

```
> db.RepublicaNepal.findOne()
{
        "_id" : ObjectId("6235d2299a8d828903ab06c3"),
        "contributors" : null,
        "truncated" : true,
        "text" : "Republica has found that in Nakkhu of Lalitpur district there is an asphalt plant being run by Shailun
g Group of Co… https://t.co/qqqnhl6LAA",
        "is_quote_status" : false,
        "in_reply_to_status_id" : null,
        "id" : NumberLong("1227244649775013888"),
        "favorite_count" : 1,
        "entities" : {
                "symbols" : [ ],
                "user_mentions" : [ ],
                "hashtags" : [ ],
                "urls" : [
                        {
                                "url" : "https://t.co/qqqnhl6LAA",
                                "indices" : [
                                        117,
                                        140
                                ],
                                "expanded_url" : "https://twitter.com/i/web/status/1227244649775013888",
                                "display_url" : "twitter.com/i/web/status/1…"
                        }
                ]
        },
        "retweeted" : false,
        "coordinates" : null,
        "source" : "<a href=\"https://about.twitter.com/products/tweetdeck\" rel=\"nofollow\">TweetDeck</a>",
        "in_reply_to_screen_name" : null,
        "in_reply_to_user_id" : null,
        "retweet_count" : 0,
        "id_str" : "1227244649775013888",
        "favorited" : false,
        "user" : {
                "follow_request_sent" : false,
                "has_extended_profile" : false,
                "profile_use_background_image" : true,
```

Displaying any one document of the collection using "db.RepublicaNepal.findOne()" command.

```
> db.RepublicaNepal.distinct("lang")
[ "da", "en", "hi", "in", "ne", "nl", "und" ]
> db.RepublicaNepal.distinct("is_quote_status")
[ false, true ]
> db.RepublicaNepal.distinct("text")
[
        "\"CIAA did not file corruption cases against former prime minister Madhav K Nepal and NCP leader Bishnu Poudel,
 thou… https://t.co/SmWC8KsTcA",
        "\"Even if it will be unhappy, we will be able to convince China. For a country nothing is more important than l
ives,… https://t.co/AtjnLVmnpA",
        "\"Government of India remains committed to continue collaborating with Government and people of Nepal for compl
etion… https://t.co/uNXqnOpEmp",
        "\"I have a belief that Nepal and Britain would see a more cordial relation during your tenure.\", Prachanda tol
d the… https://t.co/kHBnPl3csU",
        "\"Suprme Commander of the Council of Ministers is Prime Minister. So, he has the major role on Cabinet decision
. CIA… https://t.co/0P0jnqETZ4",
        "\"The grant assistance will be used to strengthen community resilience to disasters in Madi Municipality, one o
f the… https://t.co/kgzWc0JCu3",
        "\"We request the government to first upgrade it into a regional airport and gradually to international airport,
\" https://t.co/5er5pIj9t3",
        "\"We will use frequency being used by #CDMA for #4G services and eventually terminate CDMA service in the next
fisca… https://t.co/VNoYfnX3CR",
        "#EDITORIAL \nThis is one among the cases of how business people take advantage of their connection with influen
tial… https://t.co/nNEzWcalsN",
        "#EDITORIAL It is unfortunate that the action of the anti-graft body has landed in controversy. \n#LalitaNiwas…
https://t.co/lGMkOuoadz",
        "#Messi proved pivotal in the Catalans levelling right before halftime, delivering a free kick into the path of
Serg… https://t.co/PyIci18iTs",
        "#OPINION \nBy and large, the current Watershed Management Act (1982) is dysfunctional and the authorities worki
ng in… https://t.co/ARlAvbAjSS",
        "#OPINION \nNepal needs to take advantage of emerging favorable investment climate to boost private investment f
rom b… https://t.co/6942XNQzIX",
        "#OPINION \nTrump administration's peace plan for Israel and Palestine embodies none of the ingredients of succe
ssful… https://t.co/eDe8c6UEU3",
        "#OPINION In fact, five of the ten "most polluted countries" in the world with regard to fine particulates (PM2.
5) a… https://t.co/id8fTzrbl2",
```

To show the unique values in one filed using "db.RepublicaNepal.distinct("is_quote_status)" command.

```
Type   It   for   more
> db.RepublicaNepal.find( {lang:"null"} ).pretty();
> db.RepublicaNepal.find( {lang:"en"} ).pretty();
{
        "_id" : ObjectId("6235d2299a8d828903ab06c3"),
        "contributors" : null,
        "truncated" : true,
        "text" : "Republica has found that in Nakkhu of Lalitpur district there is an asphalt plant being run by Shailun
g Group of Co… https://t.co/qqqnhl6LAA",
        "is_quote_status" : false,
        "in_reply_to_status_id" : null,
        "id" : NumberLong("1227244649775013888"),
        "favorite_count" : 1,
        "entities" : {
                "symbols" : [ ],
                "user_mentions" : [ ],
                "hashtags" : [ ],
                "urls" : [
                        {
                                "url" : "https://t.co/qqqnhl6LAA"
```

To show document based on certain criteria using "db.RepublicaNepal.find({lang:"en"}).pretty()" command.

```
> db.RepublicaNepal.find({"user.name":"unes"},{id:1, _id:0}).pretty()
{ "id" : NumberLong("1227151347817967618") }
> db.RepublicaNepal.find({"user.name":"Niraj"},{"lang":"en"}).pretty()
{ "_id" : ObjectId("6235d2299a8d828903ab07d4"), "lang" : "en" }
{ "_id" : ObjectId("6235d2299a8d828903ab0844"), "lang" : "en" }
>
```

to show document based on certain criteria using
"db.RepublicaNepal.find({"user.name":"unes"},{id:1,_id:0}).preety()" command.

```
> db.RepublicaNepal.find({text: { $in: [/^canada/i,/climate/]} },{text: 1})
{ "_id" : ObjectId("6235d2299a8d828903ab06c4"), "text" : "Few developments in Nepal's hydropower sector reflect the impr
oving investment climate in the country.\n#climate… https://t.co/qAy3aoMtKr" }
{ "_id" : ObjectId("6235d2299a8d828903ab06cd"), "text" : "Investment climate: Good or bad? \n'Nepal needs to take advant
age of emerging favorable investment climate to boost… https://t.co/85xGqjKp1G" }
{ "_id" : ObjectId("6235d2299a8d828903ab06dc"), "text" : "Nepal needs to take advantage of emerging favorable investment
 climate to boost private investment from both domest… https://t.co/JussOP0qmY" }
{ "_id" : ObjectId("6235d2299a8d828903ab06ec"), "text" : "#OPINION \nNepal needs to take advantage of emerging favorable
 investment climate to boost private investment from b… https://t.co/6942XNQzIX" }
{ "_id" : ObjectId("6235d2299a8d828903ab076d"), "text" : ""There is a greater need for discussing climate change impacts
 on mountains and the forum will be the best platform… https://t.co/QgsRexQwBF" }
{ "_id" : ObjectId("6235d2299a8d828903ab07a3"), "text" : "DPM Pokharel mentioned that Nepal's mountains, lifestyle and e
cological system had borne the scourge of climate cha… https://t.co/cSKIxjRkll" }
>
```

To search document for certain criteria with expression using " db.RepublicaNepal.find({text: {
$in: [/^canada/i,/climate/]} },{text: 1})" command.

```
> db.RepublicaNepal.find({ text: { $regex: /is.*not/, $options: "si" } }, {_id:0, text:1})
{ "text" : "Furey is not scheduled for Speaker Sapkota.\n#speaker #parliament #canada\nhttps://t.co/yA3kuXGkQz" }
{ "text" : "Bring them back: 'Why is our government not responding to the call of Nepali citizens in China? It should no
t put t… https://t.co/OMcmeJEA01" }
{ "text" : "#Opinion Why is our government not responding to the call of Nepali citizens in China? It should not put the
 lives… https://t.co/stK11jQS7M" }
{ "text" : "Vegetable sellers wait for shoppers to buy radish in Sankhu, Kathmandu. The vendors say their radish was not
 being… https://t.co/nd73iKabCf" }
{ "text" : "Vegetable sellers wait for shoppers to buy radish in Sankhu, Kathmandu. The vendors say their radish was not
 being… https://t.co/dyqDU1cISi" }
{ "text" : "@RepublicaNepal This is nothing more that leftist elitist crap. You can't replace fossil fuel and still expe
ct the… https://t.co/A0wOR7y6a0" }
{ "text" : "CIAA is being criticized for not implicating the then PMs while filing the Lalita Niwas land.\n#corruption…
https://t.co/0cWY9jmCXA" }
{ "text" : "@RepublicaNepal The unpleasant things and struggle is memorable, others are not." }
{ "text" : "CIAA recorded the statements of former prime ministers Madhav Kumar Nepal and Baburam Bhattarai, but it did
not imp… https://t.co/mrED3QgWMD" }
{ "text" : "The season is appropriate for visiting Chitwan but the tourists are not seen significantly in this season.\
.https://t.co/5DyWZ9XGIO" }
{ "text" : ""If the Nepal government is not willing to evacuate us anytime soon, it should stop giving us false hopes a
d mere… https://t.co/smMYhsk0qz" }
>
```

To search document for certain criteria with expression using " db.RepublicaNepal.find({ text: {
$regex: /is.*not/, $options: "si" } }, {_id:0, text:1})" command.

3 Task C

3.1 Command to update data

```
> db.RepublicaNepal.find({_id: ObjectId("6235d2299a8d828903ab06d2")}, {lang:1})
{ "_id" : ObjectId("6235d2299a8d828903ab06d2"), "lang" : "en" }
```

Finding document with object reference value using "db.RepublicaNepal.find({_id: ObjectId("6235d2299a8d828903ab06d2")}, {lang:1})" command.

```
> db.RepublicaNepal.update({_id: ObjectId("6235d2299a8d828903ab06d2")}, {$set: {'lang':'In'}})
WriteResult({ "nMatched" : 1, "nUpserted" : 0, "nModified" : 1 })
```

Updating document value "en" to "In" using "db.RepublicaNepal.update({_id: ObjectId("6235d2299a8d828903ab06d2")}, {$set: {'lang':'In'}})" command.

```
> db.RepublicaNepal.find({_id: ObjectId("6235d2299a8d828903ab06d2")}, {lang:1})
{ "_id" : ObjectId("6235d2299a8d828903ab06d2"), "lang" : "In" }
```

Checking document using "db.RepublicaNepal.find({_id: ObjectId("6235d2299a8d828903ab06d2")}, {lang:1})" command.

```
db.createCollection('newNepal')
"ok" : 1 }
```

Creating new collection named "newNepal".

```
> show collections
RepublicaNepal
kathmandupost
newNepal
```

Displaying new collection named "newNepal".

```
> db.newNepal.insert(
... {
... id:'01',
... text: 'New nepal has updated',
... is_quote_status: 'true',
... user: [
... {
... location:'Kathmandu',
... lang:'np'
... }
... ]
... }
... )
WriteResult({ "nInserted" : 1 })
```

Inserting data in new collection.

```
db.newNepal.find({id:'01'}, { user: {$elemMatch: {lang: "np"}}}).pretty()

        "_id" : ObjectId("62360535ed8b44921221d895"),
        "user" : [
                {
                        "location" : "Kathmandu",
                        "lang" : "np"
                }
        ]
```

Finding new collection's data using query.

```
db.newNepal.find( {}, {"user.location":1, text:1})
"_id" : ObjectId("62360535ed8b44921221d895"), "text" : "New nepal has updated", "user" : [ { "location" : "Kathmandu"
] }
db.newNepal.aggregate([
.. {$project:{"user.location":1, text:1}},
.. {$out:"projnewNepal"}
.. ])
db.projnewNepal.find()pretty()
uncaught exception: SyntaxError: unexpected token: identifier :
@(shell):1:22
db.projnewNepal.find().pretty()

        "_id" : ObjectId("62360535ed8b44921221d895"),
        "text" : "New nepal has updated",
        "user" : [
                {
                        "location" : "Kathmandu"
                }
        ]
```

Reshaping the collection's data.

4 Task D

4.1 Advantage to using this approach for handling Big Data

MongoDB have better indexing and scalability

MongoDB indexing, both primary and secondary, has a lot of advantages. I indexing causes queries to be processed first, which is an important aspect. Shards may be created via indexing. Shards are sub-collections of data that've been grouped using such a shard-key. The shard-key is used to decide where in the eligible shards to search when a query is submitted. If there would be no shards, Big Data might take awhile to analyze because all of the documentation might have to be examined, and the processing may potentially run out before consumers received what they requested.

5 Task E

5.1 Disadvantage to using this approach for handling Big Data

Huge Data is not accurately described

Companies struggle to succeed in their Big Data projects due to a lack of knowledge. People may not understand what data is, how it is stored, processed, and where it comes from. People will not have a clear view of what's going on, even if data specialists do. For example, People who do not recognize the need of data storing, may not be able to maintain a backups of sensitive material. They were unable to correctly save data in databases. As a result, when this critical information is needed, it is difficult to locate.