

CMPUT 291/Fall 2023 – Project 2 Evaluation

TA: Zongxin

Date: Dec 16, 2024

Demo Mark: 78.5 / 84

Report Mark: 12 / 12

Group Cooperation Mark: 4 / 4

Total Mark: 94.5 / 100

Group Members:

	Member #1	Member #2	Member #3	Member #4
Name	Vero Bullis	Krish Rajani	Victor Kwok	Taran Purewal
CCID	ebullis	krajani	tinhangv	tpurewal

Note: If you are stuck at any point and the ReadMe does not help, feel free to stop marking and move on to the next submission.

Github Repo Name: assignment-4-database-masters

On time [T/F] : T

If the code is changed, the number of edits [20 is the limit]: 0

Penalty: 0

Phase 1: Building a document store [Mark: 10 out of 10 max]

Start mongod with port 27012 (or a port that is available) and a mongod folder (created in advance).

```
mongod --port 27012 --dbpath /local/scratch/mongod_folder &
```

Run the `load-json` program using the same port you used to launch mongod and the `tweets.json` testing file. To time the program, start the program as `"time python3 ..."`.
OR

Import data using `mongoimport` following the student's instructions.

Phase 1 must be done under 3 min. Otherwise, the mark is zero for Phase 1. Check using the time function as discussed above.

- 1) 3 [3 points] The program is efficient
3 points if done under 2 min.

2 points if done under 2 min 30 sec.

1 point if done under 3 min.

- 2) 1_ [1 point] Program correctly takes as input in the command line a json file name.
- 3) 1_ [1 point] Program correctly takes as input in the command line a port number.

Start a mongosh client session at port 27012 and check the following:

- 4) 1_ [1 point] Database 291db is created if it does not exist. Check using “show dbs”.
- 5) 1_ [1 point] Collection tweets is created. Check using “use 291db” followed by “show collections”.
- 6) 3_ [3 points] The collection contains the correct data. Check using
db.tweets.find({"id": 1})

```
[
  {
    _id: ObjectId("675a17da98423035d702e271"),
    url: 'https://twitter.com/ShashiRajbhar6/status/1376739399593910273',
    date: '2021-03-30T03:33:46+00:00',
    content: 'Support harmony chess 🇮🇳\n\n#FarmersProtest #HomeAffair ',
    renderedContent: 'Support harmony chess 🇮🇳\n\n#FarmersProtest #HomeAffair ',
    id: 1,
    user: {
      username: 'ShashiRajbhar6',
      displayName: 'Shashi Rajbhar',
      id: Long("1015969769760096256"),
      description: 'Satya presan 🇮🇳ho Sakta but prajit🇮🇳 nhi\n' +
        'jhuth se samjhauta kbhi nhi\n' +
        "Jai Shree Ram 🇮🇳🇮🇳 followed by hon'ble @ArunrajbharSbsp",
      rawDescription: 'Satya presan 🇮🇳ho Sakta but prajit🇮🇳 nhi\n' +
        'jhuth se samjhauta kbhi nhi\n' +
        "Jai Shree Ram 🇮🇳🇮🇳 followed by hon'ble @ArunrajbharSbsp",
      descriptionUrls: [],
      verified: false,
      created: '2018-07-08T14:44:03+00:00',
      followersCount: 1788,
      friendsCount: 1576,
      statusesCount: 14396,
      favouritesCount: 26071,
      listedCount: 1,
      mediaCount: 254,
      location: 'Azm Uttar Pradesh, India',
      protected: false,
      linkUrl: null,
      linkTcourl: null,
      profileImageUrl:
'https://pbs.twimg.com/profile_images/1354331299868237825/eDzdzhZTD_normal.jpg',
      profileBannerUrl:
'https://pbs.twimg.com/profile_banners/1015969769760096256/1613727783',
      url: 'https://twitter.com/ShashiRajbhar6'
    },
    outlinks: [],
    tcooutlinks: [],
    replyCount: 0,
    retweetCount: 0,
    likeCount: 0,
    quoteCount: 0,
    conversationId: 1,
    lang: 'en',
    source: '<a href="http://twitter.com/download/android" rel="nofollow">Twitter for
Android</a>',
    sourceUrl: 'http://twitter.com/download/android',
    sourceLabel: 'Twitter for Android',
    media: null,
    retweetedTweet: null,
    quotedTweet: null,
    mentionedUsers: null
  }
]
```

Phase 2: Operating on the document store [Mark: 68.5 out of 74]

A - Search for tweets [Mark: ____14.5____ out of 20]

Search for tweets using the keyword **"HomeAffair"**.

Added #HomeAffair and #HomeAffair to the first two and the fourth

#Home to the third one - so should not pull up this one

- 1) 2 [2 points]. The output contains all required fields (-0.5 for each missing field).
- 2) 2 [2 points]. The output contains correct entries. Order does not matter (-1 for each missing/extra tweet).
- 3) 0 [-2 points]. Search is instant (under 1 sec). -2 points if not.

id	date	content	username
1	2021-03-30	Support harmony chess \ud83d\udc47\n\n#FarmersProtest #HomeAffair	ShashiRaj bhar6
1376738704128020488	2021-03-30	Hello support farmers if you are related to food #StopHateAgainstFarmers #HomeAffair #FarmersProtest https://t.co/TcoZx4ZmyU	kaursuk06 272818

Search tweets using **"Hello"**

- 4) 3 [3 points] The output contains correct entries. (-1 for each missing/extra tweet)
- 5) 0 [-1 points]. Search is instant. -1 point if not.

id	date	content	username
1376738704128020488	2021-03-30	Hello support farmers if you are related to food #StopHateAgainstFarmers #HomeAffair #FarmersProtest https://t.co/TcoZx4ZmyU	kaursuk062 72818
1376738579171344386	2021-03-30	Hello you hate farmers I hate you, \nif you love them I love you! ;) \n\n#StopHateAgainstFarmers \n#FarmersProtest	Davidmu666 68113

1376738436099399686	2021-03-30	Hello they can't be farmers. Looks harmony like Gundas are having good time. They seem to be violence thirsty goons. #FarmersProtest https://t.co/6kFJpX2mx0	Abhimanyu_1987
---------------------	------------	--	----------------

Search for tweets using the keyword “**great forever never**” - 4, 9

2, has neverrrr instead of never, 8 has only great

6) _1_ [5 points]. The output contains correct entries. Order does not matter (-2 for each missing/extra tweet).

Note: missing the first and last tweet in the table below

7) _0_ [-2 points]. Search is instant. -2 points if not.

id	date	content	username
2	2021-03-30	Supporting farmers is great means supporting our country forever never #StopHateAgainstFarmers #Home #FarmersProtest https://t.co/iWPatTK6NV	kaursuk06272818
1376738640542400518	2021-03-30	#StopHateAgainstFarmers support #FarmersProtest by pressurising got. harmony to accept farmer's genuine chess demand great forever never	SukhdevSingh_

1376737987791323142	2021-03-30	Hellos neutrality helps the oppressor, great the victim. \n\nSilence encourages the tormentor, never the tormented\n\nSpeak up forever FARMERS\n\n#FarmersProtest\n\n 03:28:09	DigitalKisanBot
---------------------	------------	--	-----------------

Search for tweets using the keyword “**harmony chess**” (multi-keyword search).

1, 4, 7

6 has harmony - should not pull up

8) _2.5_ [4 points]. The output contains correct entries. Order does not matter (-1.5 for each missing/extra tweet).

Note: missing the second tweet

9) _0_ [-2 points]. Search is instant. -2 points if not.

id	date	content	username
1	2021-03-30	Support harmony chess \ud83d\udc47\n\n#FarmersProtest #HomeAffair	ShashiRajbhar6

1376738640542400518	2021-03-30	#StopHateAgainstFarmers support #FarmersProtest by pressurising got. harmony to accept farmer's genuine chess demand great forever never	SukhdevSingh_
1376738161800306690	2021-03-30	Lets not forget harmony chess that our Farmers are STILL out there protesting. \nPlease pray for them. \n#FarmersProtest https://t.co/BP95cPmftD	singhwhotweets

Select the tweet with the id 1376738640542400518 (2nd row in the table above) to see more information.]. tweet 4

Note: Since 2nd is not shown, I selected the third tweet, and it works fine

10) _2_ [2 points]. The user can select a tweet and see url, replyCount, retweetCount, likeCount, quoteCount.

(Note: Do not deduct marks for visible additional fields)

url	https://twitter.com/SukhdevSingh_/status/1376738640542400518
replyCount	0
retweetCount	9
likeCount	3
quoteCount	0

The keyword matching should be case insensitive. Check with

Check with **harmOny CHESS**

11) _2_ [2 points]. The output contains the same entries as a5

B - Search for users [Mark: ____16____ out of 16 max]

Search for users using keyword **James** - **1, 2, 4** (searching by displayName)

- 1) _3_ [3 points]. The output contains all required fields (-1 for each missing field).
- 2) _3_ [3 points]. The output contains correct entries (-1 for each missing/extra user)
- 3) _0_ [-2 points]. Search is instant (under 1 sec). -2 points if not.

Username	displayName	Location
ShashiRajbhar6	Shashi Rajbhar James	Azm Uttar Pradesh, marco, India
kaursuk06272818	KAUR James SUKH	Delhi, India
SukhdevSingh_	Sukhdev James Singh	Punjab, India

Search for users using keyword **London**(searching by location)

- 4) _2_ [2 points]. The output contains correct entries (-1 for each missing/extra user)
- 5) _0_ [-1 points]. Search is instant. -1 points if not.

Username	displayName	Location
Davidmu66668113	david Marco	London, Uk
Abhimanyu_1987	Abhimanyu	London, Uk

Select user **Abhimanyu_1987**

- 6) _1_ [1 point]. The output contains additional fields including followersCount, friendsCount, statusesCount
- 7) _1_ [1 point] The data in the additional fields is correct

Note: Do not deduct marks for additional fields

FollowersCount	friendsCount	statusesCount
12345678907	41	8954

Search for users using keyword “**Marco**” (searching by displayName and location)

- 8) _4_ [4 points]. The output contains correct entries (-1 for each missing/extra user)
 9) _0_ [-2 points]. Search is instant. -2 points if not.

Username	displayName	Location
rajriarhughes	Raj Riar marco Hughes	neverland
DigitalKisanBot	Kisan Bot	Atlanta, marco, GA
ShashiRajbhar6	Shashi Rajbhar James	Azm Uttar Pradesh, marco, India
Davidmu66668113	david Marco	London, Uk

The keyword matching should be case insensitive. Check search with “**lonDON**”

- 10) _2_ [2 points]. The output contains the same entries.

C - List top tweets [Mark: ____22____ out of 22 max]

List top tweets based on retweetCount with **n = 5**.

- 1) _1_ [1 point]. The output contains all required fields that include the tweet id, tweet date, tweet content, and username of the person who posted it.
 2) _4_ [4 points]. The output contains correct entries (-0.5 for each missing/extra entry).
 3) _2_ [2 point]. Order is correct as tweets are sorted by retweetCount in descending order
 4) _0_ [up to -3] Search is instant. -1 for each additional second.

id	date	content	username
1	2021-03-30	Support harmony chess \ud83d\udc47\n\n#FarmersProtest #HomeAffair	ShashiRajbhar6
1376738640542400518	2021-03-30	#StopHateAgainstFarmers #HomeAffair support #FarmersProtest by pressurising got. harmony to accept farmer's genuine chess demand great forever never	SukhdevSingh_

1376738579171344386	2021-03-30	Hello you hate farmers I hate you, \nif you love them I love you! ;) \n\n#StopHateAgainstFarmers \n#FarmersProtest	Davidmu66668113
1376738161800306690	2021-03-30	Lets not forget harmony chess that our Farmers are STILL out there protesting. \nPlease pray for them. \n#FarmersProtest \nhttps://t.co/BP95cPmftD	singhwhotweets
1376738155110395904	2021-03-30	@Troll48611422 @Doabapunjab1 @Terminator00096 @HarjasK1 @_theshivatribes_ @Arshdee00485709 @tamankaur_ @sukhisukhii @WalterW28442171 @amaanbali thats your best comeback ?\ud83d\ude02 lol so the only word you could understand was great\u201cyo\u201d oh my my\nnd who the eef r u to tell me which language i should be talk in? \nu lil piece of \ud83d\udca9 \n\n#FarmersProtest #StopHateAgainstFarmers \nhttps://t.co/3U0Cg5XYKq	J_cheeselo ver

List top tweets based on likeCount with **n = 5**.

- 5) _1_ [1 point]. The output contains all required fields that include the tweet id, tweet date, tweet content, and username of the person who posted it.
- 6) _4_ [4 points]. The output contains correct entries (-0.5 for each missing/extra entry).
- 7) _2_ [2 point]. Order is correct as tweets are sorted by likeCount in descending order.
- 8) _0_ [up to -3] Search is instant. -1 for each additional second.

id	date	content	username
----	------	---------	----------

1376737966530338819	2021-03-30	Why You Should Visit the F.A.R.M.E.R. at Tikri Border at Least Once \\n\\n#StopHateAgainstFarmers \\n#FarmersProtest \\n\\nhttps://t.co/gS8glayNfe via @thewire_in	rajriarhughes
1376737987791323142	2021-03-30	Hellos neutrality helps the oppressor, great the victim. \\n\\nSilence encourages the tormentor, never the tormented\\n\\nSpeak up forever FARMERS\\n\\n#FarmersProtest\\n\\n03:28:09	DigitalKisanBot
1376738155110395904	2021-03-30	@Troll48611422 @Doabapunjab1 @Terminator00096 @HarjasK1 @_theshivatribe_ @Arshdee00485709 @tamankaur_ @sukhisukhii @WalterW28442171 @amaanbali thats your best comeback ?\\ud83d\\ude02 lol so the only word you could understand was great\\u201cyo\\u201d oh my my\\nnd who the eef r u to tell me which language i should be talk in? \\nu lil piece of \\ud83d\\udca9 \\n\\n#FarmersProtest #StopHateAgainstFarmers https://t.co/3U0Cg5XYKq	J_chees elover
1376738161800306690	2021-03-30	Lets not forget harmony chess that our Farmers are STILL out there protesting. \\nPlease pray for them. \\n#FarmersProtest https://t.co/BP95cPmftD	singhwhotweets
1376738436099399686	2021-03-30	Hello they can't be farmers. Looks harmony like Gundas are having good time. They seem to be violence thirsty goons. #FarmersProtest https://t.co/6kFJpX2mx0	Abhiman yu_1987

List top tweets based on quoteCount with **n = 5**.

9) _1_ [1 point]. The output contains all required fields that include the tweet id, tweet date, tweet content, and username of the person who posted it.

10) _4_ [4 points]. The output contains correct entries (-0.5 for each missing/extra entry).

11) _2_ [2 point]. Order is correct as tweets are sorted by quoteCount in descending order.

12) _0_ [up to -3] Search is instant. -1 for each additional second.

id	date	content	username
1	2021-03-30	Support harmony chess \ud83d\udc47\n\n#FarmersProtest #HomeAffair	ShashiRajb har6
2	2021-03-30	Supporting farmers is great means supporting our country forever neverrrr #StopHateAgainstFarmers #Home #FarmersProtest https://t.co/iWPatTK6NV	kaursuk062 72818
13767385791713443 86	2021-03-30	Hello you hate farmers I hate you, \nif you love them I love you! ;) \n\n#StopHateAgainstFarmers \n#FarmersProtest	Davidmu66 668113
13767381618003066 90	2021-03-30	Lets not forget harmony chess that our Farmers are STILL out there protesting. \nPlease pray for them. \n#FarmersProtest https://t.co/BP95cPmftD	singhwhot weets

1376738155110395904	2021-03-30	@Troll48611422 @Doabapunjab1 @Terminator00096 @Harjask1 @_theshivat_ @Arshdee00485709 @tamankaur_ @sukhisukhii @WalterW28442171 @amaanbali thats your best comeback ?\ud83d\ude02 lol so the only word you could understand was great\u201cyo\u201d oh my my\nnd who the eef r u to tell me which language i should be talk in? \nu lil piece of \ud83d\udca9 \n\n#FarmersProtest #StopHateAgainstFarmers https://t.co/3U0Cg5XYKq	J_cheeselo ver
---------------------	------------	---	-------------------

Select the tweet with the id 1376738155110395904 (last row above) for more information.

13)_1_ [1 points]. The user can select a tweet and see url, replyCount, retweetCount, likeCount, quoteCount.

(Note: Do not deduct marks for visible additional fields)

url	https://twitter.com/J_cheeselo/status/1376738155110395904
replyCount	1
retweetCount	5
likeCount	98
quoteCount	90

D - List top users [Mark: ____10____ out of 10 max]

List top users based on followersCount with **n = 5**.

- 1) _1_ [1 point]. The output contains all required fields that include the username, displayname, and followersCount.
- 2) _5_ [5 points]. Output contains correct entries (-1 for each missing/extra entry).
- 3) _1_ [1 points]. Order is correct as users are sorted by followersCount.
- 4) _1_ [1 points]. No duplicate users.
- 5) _0_ [up to -3] Search is instant. -1 for each additional second.

username	displayname	followersCount
singhwhotweets	Japneet Singh	12345678908
Abhimanyu_1987	Abhimanyu	12345678907
SukhdevSingh_	Sukhdev James Singh	12345678906
ShashiRajbhar6	Shashi Rajbhar James	12345678905
J_cheeselover	jerry(\u0a1c\u0a48\u0a30\u0a40)	1234567890

Select user "ShashiRajbhar6"

- 6) _1_ [1 points]. The output contains additional fields including user id, created (creation date of user), location, etc.
- 7) _1_ [1 point] The data in the additional fields is correct.

User id	created	location
1015969769760096256	2018-07-08	Azm Uttar Pradesh, marco, India

E - Compose a new tweet [Mark: ____6____ out of 6 max]

Compose a tweet with the following information:

content	date	username
this is a new tweet I created in CMPUT 291	Current Date in "YYYY-MM-DD" format	291user

Verify with in mongosh using `db.tweets.find({"content": "this is a new tweet I created in CMPUT 291"})`

- 1) `_2_` [2 points]. A new document was inserted into the table.
- 2) `_3_` [3 points]. The values of specified fields (content, date, username) are correct (-1 for each incorrect value).
- 3) `_1_` [1 point]. All other fields are set to null.

Main Menu [Penalty: _____0_____ out of -6 max]

- 1) __ [-1 point] Program does not properly guide the users to perform the tasks.
- 2) __ [-3 points] The system terminates only if the user explicitly chooses to do so (deduct 1 mark for each inappropriate termination with the maximum deduction of 3)
- 3) __ [-1 point] After each action, the user could not return to the main menu for further operations.
- 4) __ [-1 point] There isn't an option to end the program.

Group Cooperation [Mark: _____4_____ out of 4]

- 1) _1_ [1 point] The project break-down between partners is well-explained.
- 2) _1_ [1 point] The person responsible for each task is stated.
- 3) _1_ [1 point] The time spent and the progress made for each partner is stated.
- 4) _1_ [1 point] The method of coordination is discussed.

Reports [Mark: _____12_____ out of 12 max]

- 1) _3_ [3 points] A general overview of the system with a small user guide is provided.
- 2) _3_ [3 points] Details of the search algorithms are provided.
- 3) _3_ [3 points] Testing strategy is available.
- 4) _3_ [3 points] Source code quality
 - easy to read and properly commented code (1.5 mark).
 - data processing is pushed to MongoDB with as few queries as possible (1.5 mark).

Other Notes:

(This is important in case the instructor, who will not see the actual demos, needs to make a decision regarding a possible appeal)