CS 122A: Introduction to Data Management – Spring 2019

Homework 8: NoSQL (100 points)

Due Date: **Thursday, June 6 (5:00 PM)**

Last Name: ZHOU **First Name: YIHUA** **Student ID: 83186710**

1. [10 pts]

In 1NF: Observer, Observations, Users, Interest. These datasets only include atomic attributes

Not in 1NF: Thought, Event. These dataset contain non-atomic attributes.

2. [10pts]

Comparing both Schema and NoSchema, NoSchema has fewer attribute than Schema in datatype. There is no difference in the way both version showing the data. Able to query other attributes that is not defined in the NoSchema file just like it defined.

3. [10 pts]

***Users**:*
We can include both PHLogger and Instructor entity in the User entity, we have utype to distinguish the user that belongs to. Therefore, we can use utype to query User separately depends on type we want to have. We include the nested/repeating memberOf attribute for each User in one attribute rather than separate them.

***Events**:*
We include the nested/repeating relatedObservations attribute for each Event in one attribute rather than separate them.

4. [12 pts] [Result size: 1]

Sample output: { "id": 97, "name": "Gwenn Dooley", "email": "gwenn.dooley@uci.com", "rate": 21, "kind": "watch", "startTimestamp": "2019-04-15T13:00:00.000Z" }

[8 pts] Query (Please also submit this query in Question_4.txt within your zip file to EEE):

```
Select  s.rate, u.id, u.name, u.email, r. kind, s.startTimestamp from Users as u, Observations as s, Observers as r where u.id = 97 and  u.id = r.ownedBy and r.id = s.observerId ORDER BY s.startTimestamp DESC LIMIT 10;
```

[4 pts] Result:

Output

Results:

```
{ "id": 97, "email": "gwenn.dooley@uci.com", "kind": "glucose mo
nitor", "startTimestamp": "2019-03-21T16:00:00.000Z", "name": "G
wenn Dooley", "rate": 62 }
```

```
Duration of all jobs: 0.472 sec
```

Success: Query Complete

5. [14 pts] [Result size: 3]

Sample output: { "id": 2, "utype": "PHLogger", "contact": [ ], "memberOf": ["alcoholism 1", "exercise 2", "asthma 1"], "address": { "state": "NC", "city": "Murazikfort", "street": "Timmy Club", "zipcode": "61183" }, "passwd": "aa8ed233b0aa9b7492bd58910c5f1392", "email": "bethany.macgyver@uci.com", "name": "Bethany MacGyver" }...

[7 pts] Query (Please also submit this query in Question_5.txt within your zip file to EEE):

```
SELECT VALUE U
FROM Users U
ORDER BY Len(u.memberOf) DESC
LIMIT 3;
```

[4 pts] Result:

Output

Results:


```
{ "id": 6, "passwd": "40df8316008b824b73d5f2441ada
8c50", "email": "genevieve.stamm@uci.com", "utype"
: "PHLogger", "contact": [ "100-081-7299", "413-35
6-5368" ], "memberOf": [ "HIV 0", "depression 1",
"exercise 0" ], "address": { "state": "NM", "city"
: "East Ardis", "street": "Roberts Hollow", "zipco
de": "79831-8928" }, "name": "Genevieve Stamm" }
{ "id": 8, "passwd": "801b6bd9016febe4203c187aa981
f16d", "email": "homer.kunze@uci.com", "utype": "P
HLogger", "contact": [ "095-367-1678" ], "memberOf
": [ "HIV 0", "alcoholism 0" ], "address": { "stat
e": "MT", "city": "Bertieton", "street": "Angelena
 Trail", "zipcode": "63662-6198" }, "name": "Homer
 Kunze" }
{ "id": 9, "passwd": "9394443e715dbf35bd0327edbe5a
d9ca", "email": "jenifer.schmitt@uci.com", "utype"
: "PHLogger", "contact": [ "093-011-7538" ], "memb
erOf": [ "alcoholism 2", "diabetes 0", "alcoholism
 0" ], "address": { "state": "ME", "city": "Edwinp
ort", "street": "Stacie River", "zipcode": "09867-
6558" }, "name": "Jenifer Schmitt" }
```


```
Duration of all jobs: 0.373 sec
```

Success: Query Complete

[3 pts] Now try the same query on the **NoSchema** version of the PHLogger dataverse (**'NoSchema'**).

a) Did it work?

   YES

b) Were the results different?

   NO

c) What does this tell you?

   Different way to find the same result, Can query non-defined attribute like pre-defined ones

6. [14 pts] [Result size: 4]
Sample output: { "id": 9 }...

[10 pts] Query (Please also submit this query in Question_6.txt within your zip file to EEE):

```
USE HW8;
WITH H AS (SELECT  ov.ownedBy AS id
FROM Observations AS o, Observers AS ov
WHERE ov.id = o. observerId
AND o.kind = "heart_rate"
GROUP BY ov.ownedBy
HAVING AVG(o.rate) > 85),
D AS (SELECT u.id AS id
FROM Users AS u
WHERE SOME e IN u.memberOf SATISFIES e IN (SELECT VALUE i.name FROM Interests AS i WHERE i.topic
= "diabetes" ))
SELECT did.id
FROM H AS hid, D AS did
WHERE hid.id = did.id;
```

[4 pts] Result:

Results:

```
{ "id": 9 }
{ "id": 21 }
{ "id": 25 }
{ "id": 32 }
```

```
Duration of all jobs: 0.076 sec
```

Success: Query Complete

7. [18 pts] [Result size: 1]

Sample output: { "thought_texts": ["Today I went to an AA meeting."], "blood_pressure": [{ "diastolic": 93, "systolic": 128 }, { "diastolic": 103, "systolic": 171 }], "heart_rates": [70, 94], "id": 11, "utype": "PHLogger", "contact": [ ], "memberOf": ["HIV 2"], "address": { "state": "SC", "city": "New Dovie", "street": "Lind Roads", "zipcode": "36502" }, "passwd": "b643b952877ba9468e0a40734626f8e7", "email": "jewell.greenfelder@uci.com", "name": "Jewell Greenfelder" }

[10 pts] Query (Please also submit this query in Question_7.txt within your zip file to EEE):

```
USE HW8;
SELECT Users.*, (SELECT VALUE t.text FROM Thoughts AS Th WHERE Th.phlid = 85) AS thought_texts,(SELECT o.diastolic, o.systolic FROM Observations AS o, Observers AS ov WHERE ov. ownedBy = 85 AND ov.id = o. observerId AND o.kind = "blood_pressure") AS blood_pressure, (SELECT VALUE o.rate FROM Observations AS o, Observers AS ov WHERE ov. ownedBy = 85 AND ov.id = o. observerId AND o.kind = "heart_rate") AS heart_rates FROM Users WHERE Users.id = 85;
```

[4 pts] Result:

Results:


```
- {
  - thought_texts: [
      null,
      null,
      null,
      null
    ],
  - blood_pressure: [
    + { 2 fields },
    + { 2 fields },
    + { 2 fields }
    ],
  - heart_rates: [
      69,
      96,
      86
    ],
    id: 85,
    passwd: "8fd4f4773894d0316332121d0fa60c8c",
    email: "ammie.gleason@uci.com",
    utype: "PHLogger",
  - contact: [
      "640-577-0013"
    ],
  - memberOf: [
      "alcoholism 3"
    ],
  - address: {
      state: "CT",
      city: "East Elishamouth",
      street: "Elbert Fields",
      zipcode: "98199"
    },
    name: "Ammie Gleason"
}
```

[4 pts] As NoSQL provides nested attributes, we could have pushed all the thoughts, observations into the Users dataset. Briefly explain what could be the reason(s) that we didn't do that.

Because change on one table usually won't affect other table, but putting everything into one table then table become larger when searching for something will be a big complexity, also when try to change data, it likely make mistake to other data.

8. [12 pts] Write a query to print the **user id** and **the number of thoughts** of users who have posted at least 9 thoughts. [Result size: 6]
Sample output: { "id": 1, "cnt": 1 }...

[8 pts] Query (Please also submit this query in Question_8.txt within your zip file to EEE):

```
USE HW8;
SELECT t.phlid AS i, COUNT(*) AS c
FROM Thoughts AS t, Users AS u
Where  t.phlid = u.id
GROUP BY t.phlid
HAVING COUNT(*) >= 9;
```

[4 pts] Result:

Results:


```
- {
    i: 37,
    c: 9
  }
- {
    i: 49,
    c: 9
  }
- {
    i: 79,
    c: 9
  }
- {
    i: 92,
    c: 9
  }
- {
    i: 31,
    c: 9
  }
- {
    i: 78,
    c: 9
  }
```


```
Duration of all jobs: 0.029 sec
```

Success: Query Complete