

Slovenská technická univerzita v Bratislave
Fakulta informatiky a informačných technológií

Zadanie 3
Databázové systémy
Adam Jurčišin

Dátum : 24. 3. 2024

Obsah

1.	ÚVOD.....	2
2.	1. ÚLOHA	2
3.	2. ÚLOHA.....	5
4.	3. ÚLOHA.....	7
5.	4. ÚLOHA.....	8
	ZÁVER	9

1. Úvod

V tomto zadání máme za úlohu vytvoriť 4 úlohy, ktoré vykonávajú SQL dopyty na PostgreSQL databázu. Pre každú úlohu, ktorú vytvoríme ukážeme implementovaný endpoint, SQL dopyt, príklad volania endpointu a výsledok z databázy.

Zadanie budeme vytvárať v programovacom jazyku Python a SQL.

2. 1. Úloha

V prvej úlohe sme implementovali takýto endpoint:

```
@router.get("/v3/users/{user_id}/badge_history")
async def posts(user_id: int):
    connection = psycopg2.connect(
        dbname=settings.DATABASE_NAME,
        user=settings.DATABASE_USER,
        password=settings.DATABASE_PASSWORD,
        host=settings.DATABASE_HOST,
        port=settings.DATABASE_PORT
    )
```

Obrázok 1

Tento SQL dopyt kombinuje informácie o odznakoch a príspevkoch pre daného používateľa. Výsledok zahŕňa identifikátor, názov, typ (odznak alebo príspevok), dátum vytvorenia a pozíciu riadku. Používa sa funkcia ROW_NUMBER() na priradenie pozície riadku podľa typu a dátumu vytvorenia. Dotaz obsahuje vnorené poddotazy na získanie dát o odznakoch a príspevkoch a využíva funkcie LEAD() a LAG() na identifikáciu nasledujúceho a predchádzajúceho typu záznamu. Výsledné záznamy sú filtrované na základe kombinácie typov (príspevok a odznak) a usporiadajú sa podľa pozície riadku a typu zostupne.

```
cursor.execute("""
SELECT id, title, type, date AS created_at, ROW_NUMBER() OVER (PARTITION BY type ORDER BY date) AS position
FROM (
    SELECT id, title, type, date,
           LEAD(type) OVER (ORDER BY date) AS next_type,
           LAG(type) OVER (ORDER BY date) AS prev_type
    FROM (
        SELECT b.id, b.name AS title, 'badge' AS type, to_timestamp(to_char(b.date AT TIME ZONE 'UTC', 'YYY
        FROM badges AS b
        WHERE b.userid = %s
        UNION
        SELECT p.id, p.title AS title, 'post' AS type, to_timestamp(to_char(p.creationdate AT TIME ZONE 'UT
        FROM posts AS p
        WHERE p.owneruserid = %s
        ORDER BY date, id
    ) AS subquery
) AS subquery2
WHERE type = 'post' AND next_type = 'badge' OR type = 'badge' AND prev_type = 'post'
ORDER BY position, type DESC
""", (user_id, user_id))
```

Obrázok 2

Príklad volania endpointu:

GET /v3/users/120/badge_history

Obrázok 3

Výsledok z databázy:

```
{
  "items": [
    {
      "id": 7744,
      "title": "How do I make Firefox remember its window size?",
      "type": "post",
      "created_at": "2009-07-18T03:33:08.597000+02:00",
      "position": 1
    },
    {
      "id": 5453,
      "title": "Student",
      "type": "badge",
      "created_at": "2009-07-18T03:47:30.730000+02:00",
      "position": 1
    },
    {
      "id": 8957,
      "title": null,
      "type": "post",
      "created_at": "2009-07-20T02:27:58.430000+02:00",
      "position": 2
    },
    {
      "id": 6095,
      "title": "Teacher",
      "type": "badge",
      "created_at": "2009-07-20T02:32:30.713000+02:00",
      "position": 2
    },
    {
      "id": 14860,
      "title": "How to remove iso 9660 from USB?",
      "type": "post",
      "created_at": "2009-07-29T03:52:34.903000+02:00",
      "position": 3
    },
    {
      "id": 9326,
      "title": "Commentator",
      "type": "badge",
      "created_at": "2009-07-29T04:32:31.143000+02:00",
      "position": 3
    }
  ]
}
```

Obrázok 4

3. 2. Úloha

V druhej úlohe sme implementovali takýto endpoint:

```
@router.get("/v3/tags/{tag}/comments")
async def get_tag_comments(tag: str, count: int = Query(None)):
    connection = psycopg2.connect(
        dbname=settings.DATABASE_NAME,
        user=settings.DATABASE_USER,
        password=settings.DATABASE_PASSWORD,
        host=settings.DATABASE_HOST,
        port=settings.DATABASE_PORT
    )
```

Obrázok 5

Tento SQL dopyt analyzuje komentáre v príspevkoch označených určitým tagom a s počtom komentárov vyšším ako určená hranica. Výsledok obsahuje identifikátor príspevku, názov príspevku, meno používateľa, text komentáru, čas vytvorenia príspevku a čas vytvorenia komentáru. Využíva funkciu LAG() na vypočítanie časového rozdielu medzi aktuálnym a predchádzajúcim komentárom v rámci toho istého príspevku. Pre každý príspevok je vypočítaný priemerný časový rozdiel medzi komentármi pomocou funkcie AVG() s oknom, ktoré sa pohybuje od začiatku až po aktuálny riadok. Celkový dotaz sa skladá z niekoľkých vnorených poddotazov, ktoré zabezpečujú získanie potrebných dát a spracovanie výsledkov. Nakoniec je výsledok zoskupený do jedného súboru údajov pre ďalšie spracovanie alebo zobrazenie.

```
SELECT post_id, title, displayname, text, post_created_at, created_at, diff,
       AVG(CASE WHEN diff IS NOT NULL THEN diff END) OVER (PARTITION BY post_id ORDER BY created_at ROWS BETWEEN
FROM(
  SELECT post_id, title, displayname, text, post_created_at, created_at,
         CASE
           WHEN diff IS NULL THEN (created_at - post_created_at)
           ELSE diff
         END AS diff
  FROM(
    SELECT post_id, title, displayname, text, post_created_at, created_at,
           (created_at - LAG(created_at) OVER (PARTITION BY post_id ORDER BY created_at)) AS diff
    FROM (
      SELECT DISTINCT p.id AS post_id, p.title, u.displayname, c.text,
                     to_timestamp(to_char(p.creationdate AT TIME ZONE 'UTC', 'YYYY-MM-DD"T"HH24:MI:SS.US"Z"'), 'YY
                     to_timestamp(to_char(c.creationdate AT TIME ZONE 'UTC', 'YYYY-MM-DD"T"HH24:MI:SS.US"Z"'), 'YY
      FROM comments AS c
      JOIN users AS u ON c.userid = u.id
      JOIN posts AS p ON c.postid = p.id
      JOIN post_tags AS pt ON p.id = pt.post_id
      JOIN tags AS t ON pt.tag_id = t.id
      WHERE t.tagname = %s AND p.commentcount > %s
      ORDER BY created_at ASC
    ) AS subquery
  ) AS subquery2
) AS subquery3
(tag, count))
```

Obrázok 6

Príklad volania endpointu:

```
GET /v3/tags/networking/comments?count=40
```

Obrázok 7

Výsledok z databázy:

```
{
  "items": [
    {
      "post_id": 1034137,
      "title": "Did I just get hacked?",
      "displayname": "Jonno",
      "text": "Yeah that doesn't look too good. I'm not an expert in Linux by any means, but somethings definitely tried to execute on there. I'm not quite sure how though as it looks like it attempted to log in as root and failed. Are there any other logs in your auth.log? Any other means of remote admin? I've seen Mac's with VNC server enabled get hacked before via that, although this looks like an SSH attempt. Looks like the IPs it was downloading from are hosted in China somewhere.",
      "post_created_at": "2016-02-01T10:21:48.690000+01:00",
      "created_at": "2016-02-01T10:25:02.610000+01:00",
      "diff": 193.92,
      "avg": 193.92
    },
    {
      "post_id": 1034137,
      "title": "Did I just get hacked?",
      "displayname": "David Schwartz",
      "text": "The attack actually came from China.",
      "post_created_at": "2016-02-01T10:21:48.690000+01:00",
      "created_at": "2016-02-01T10:30:45.310000+01:00",
      "diff": 342.7,
      "avg": 268.31
    },
    {
      "post_id": 1034137,
      "title": "Did I just get hacked?",
      "displayname": "vaid",
      "text": "Yes but what is a Microsoft owned IP doing trying to breach a device across the internet?",
      "post_created_at": "2016-02-01T10:21:48.690000+01:00",
      "created_at": "2016-02-01T10:37:58.037000+01:00",
      "diff": 432.727,
      "avg": 323.115667
    },
    {
      "post_id": 1034137,
      "title": "Did I just get hacked?",
      "displayname": "Journeyman Geek",
      "text": "You got brute forced. This is why one does not leave a ssh server on the internet, even if you have a password. Anything short of key based auth is not secure enough these days.",
      "post_created_at": "2016-02-01T10:21:48.690000+01:00",
      "created_at": "2016-02-01T10:41:17.843000+01:00",
      "diff": 199.806,
      "avg": 292.28825
    }
  ],
}
```

Obrázok 8

4. 3. Úloha

Tretia úloha má aplikovaný nasledujúci endpoint:

```
@router.get("/v3/tags/{tagname}/comments/{position}")
async def get_tag_comments_by_position(tagname: str, position: int, limit: int = Query(None)):

    connection = psycopg2.connect(
        dbname=settings.DATABASE_NAME,
        user=settings.DATABASE_USER,
        password=settings.DATABASE_PASSWORD,
        host=settings.DATABASE_HOST,
        port=settings.DATABASE_PORT
    )
```

Obrázok 9

Tento SQL dopyt vyberá identifikátor komentára, meno používateľa, telo komentára, text komentára, skóre a pozíciu. Výber sa uskutočňuje z poddotazu, ktorý vytvára rady čísel pre každý komentár zoradený podľa dátumu vytvorenia vzostupne. Vnútorňý poddotaz zabezpečuje získanie jedinečných dát o komentároch, vrátane dátumu vytvorenia príspevku, počtu komentárov, identifikátora komentára, mena používateľa, tela komentára, textu komentára a skóre. Výsledok sa filtruje pomocou row_number na výber každej k-tej pozície komentára, kde k je hodnota parametru position. Limituje sa počet vrátených záznamov na hodnotu definovanú parametrom limit.

```
SELECT id, displayname, body, text, score, %s AS position
FROM (
    SELECT commentcount, id, displayname, body, text, score, ROW_NUMBER() OVER(ORDER BY creationdate ASC) AS
    FROM (
        SELECT DISTINCT p.creationdate, p.commentcount, c.id, u.displayname, p.body, c.text, c.score
        FROM comments AS c
        JOIN users AS u ON c.userid = u.id
        JOIN posts AS p ON c.postid = p.id
        JOIN post_tags AS pt ON p.id = pt.post_id
        JOIN tags AS t ON pt.tag_id = t.id
        WHERE t.tagname = %s AND p.commentcount >= %s
    ) AS distinct_comments
) AS subquery
WHERE row_num %% %s = 0
LIMIT %s
```

Obrázok 10

Príklad volania endpointu:

```
GET /v3/tags/linux/comments/2?limit=1
```

Obrázok 11

Výsledok z databázy:

```
{
  "items": [
    {
      "id": 745427,
      "displayname": "Oliver Salzburg",
      "body": "\u003Cp\u003EI am running Kubuntu Hardy Heron, with a dual monitor setup, and have VirtualBox on it running Windows XP in seamless mode.\u003C/p\u003E\n\n\u003Cp\u003EMy problem is, I can't get VirtualBox to extend to the second monitor. \nHow can this be achieved?\u003C/p\u003E\n",
      "text": "http://ubuntuforums.org/showthread.php?t=433359",
      "score": 0,
      "position": 2
    }
  ]
}
```

Obrázok 12

5. 4. Úloha

Štvrtá úloha má aplikovaný nasledujúci endpoint:

```
@router.get("/v3/posts/{postid}")
async def get_post(postid: int, limit: int = Query(None)):

    connection = psycpg2.connect(
        dbname=settings.DATABASE_NAME,
        user=settings.DATABASE_USER,
        password=settings.DATABASE_PASSWORD,
        host=settings.DATABASE_HOST,
        port=settings.DATABASE_PORT
    )
```

Obrázok 13

Tento SQL dopyt vyberá meno používateľa, telo príspevku a dátum vytvorenia príspevku v čase UTC pre príspevkov s daným identifikátorom alebo s rodičovským identifikátorom. Používa sa funkcia AT TIME ZONE 'UTC' na konverziu dátumu vytvorenia príspevku na časovú zónu UTC. Výsledky sú zoradené vzostupne podľa dátumu vytvorenia príspevku a limituje sa počet vrátených záznamov na hodnotu definovanú parametrom limit.

```
cursor.execute("""
    SELECT u.displayname, p.body, p.creationdate AT TIME ZONE 'UTC' AS created_at
    FROM posts AS p
    JOIN users AS u ON u.id = p.owneruserid
    WHERE p.id = %s OR p.parentid = %s
    ORDER BY p.creationdate ASC
    LIMIT %s
""", (postid, postid, limit))
```

Obrázok 14

Príklad volania endpointu:

GET /v3/posts/2154?limit=2

Obrázok 15

Výsledok z databázy:

```
{
  "items": [
    {
      "displayName": "Eugene M",
      "body": "\u003Cp\u003E\u003ESo, I'm a technology guy and sometimes I have to troubleshoot a home network, including my own. I make sure the wires are in securely and that the lights suggest there's an actual internet connection. Usually after that point I just reset the router( and possibly the cable modem) and that fixes things most of the time.\u003C/p\u003E\u003E\n\n\u003Cp\u003E\u003EThe problem is I'd like to know what sort of issue I could possibly be fixing by resetting the router.\u003C/p\u003E\u003E\n\n\u003Cp\u003E\u003EEDIT: Just to clarify, I was speaking more about reset as in turning the router off and on. Still, any information about a hard reset(paperclip in the hole) is useful. So the more accurate term would probably be restarting \u003C/p\u003E\u003E\n\n\u003Cp\u003E\u003EAlso, personally I usually have to deal with D-Link or Linksys home routers. I generally only bother messing around with stuff if I can't make a connection to the internet at all.\u003C/p\u003E\u003E\n",
      "created_at": "2009-07-15T12:51:57.340000"
    },
    {
      "displayName": "Ólafur Waage",
      "body": "\u003Cp\u003E\u003EEvery router has it's original firmware stored somewhere on it.\u003C/p\u003E\u003E\n\n\u003Cp\u003E\u003EWhen you reset the router you overwrite the current firmware and config with the original one. What usually is fixing the problem is that the config is overwritten with the original one. But in some cases you have an updated router that isn't working for some reason.\u003C/p\u003E\u003E\n",
      "created_at": "2009-07-15T12:54:48.507000"
    }
  ]
}
```

Obrázok 16

Záver

Implementovali sme úspešne úlohy 1-4, ktoré vykonávajú SQL dopyty ktoré sú im zadané.