Assignements 10 + 11
Bloom Filters
MongoDB

Bloom Filter

1. Bloom filters may return a false positive.
   What is a false positive and how can a false positive happen?

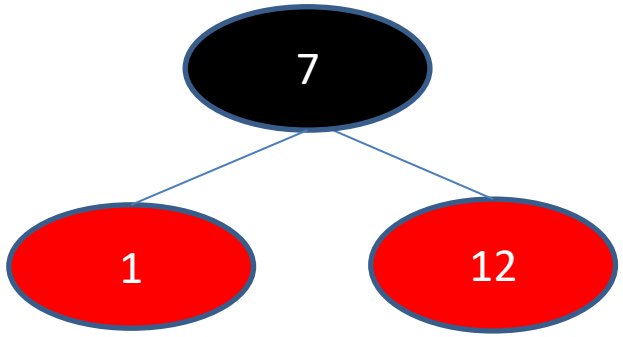
2. Bloom filters cannot be used for range queries. Explain why.

eva.knirsch@kiu.edu.ge

LSM Storage

1: "Mariam"
12: "Otto"
7: "Rudi"

h1(x) = 1x mod 20
h2(x) = 3x mod 20
h3(x) = 7x mod 20



B1 = 0101 1001 0100 1000 1000

- In class, we built the memtable with keys 1,12 and 7.
- We also built the bloom filter B of the 3 keys.
- Memtable is flushed to disk (segment 1), then memtable in memory is discarded.
- Bloom filter is kept in memory.

Task1: Build another memtable and bloomfilter for the following requests:

- store(key: 18, value: "Ana")
- store(key: 8, value: "Guga")
- delete(key: 12)

eva.knirsch@kiu.edu.ge

LSM Storage

Once our memtable reaches its size (in our case: 3 keys) it is flushed to disk to segment 2.

Task 2: Write down the content of segment 1 and segment 2 (keys and values) as stored on disk.

eva.knirsch@kiu.edu.ge

LSM Storage

Task 3: Process the following read requests:

1. get(key: 3)
2. get(key: 12)
3. get(key: 14)

Attention: Neglect the step with the sparse index. A sparse index does not make sense with 3 keys. ☺

eva.knirsch@kiu.edu.ge

LSM Storage

Task 4: Merge segments 1 and 2. Assume that the merged segment can hold more than 3 keys. Assume that segment 1 and 2 are the only segments. Build the new bloom filter.

eva.knirsch@kiu.edu.ge

Postgres Bloom Filter

Task: Get the Postgres Bloom Filter to work and evaluate it.

Activate the Postgres extension and verify:

```
Create extension bloom;
SELECT * FROM pg_extension WHERE extname = 'bloom';
```


eva.knirsch@kiu.edu.ge

Postgres Bloom Filter

Create a test table wanted:

```
CREATE TABLE IF NOT EXISTS public.wanted
(
    id serial,
    f_name text COLLATE pg_catalog."default",
    l_name text COLLATE pg_catalog."default",
    city text COLLATE pg_catalog."default",
    birthday text COLLATE pg_catalog."default",
    passport_number text COLLATE pg_catalog."default",
    issuing_country text COLLATE pg_catalog."default",
    CONSTRAINT wanted_pkey PRIMARY KEY (id)
)

TABLESPACE pg_default;

ALTER TABLE IF EXISTS public.wanted
    OWNER to postgres;
```

Apart from the PK, all columns are of type 'text'. Explain why.

eva.knirsch@kiu.edu.ge

Postgres Bloom Filter

Populate table wanted with 100000 to 200000 rows. Have the content somehow generated, for example:

```
INSERT INTO wanted(f_name, l_name, city, birthday, passport_number, issuing_country)
SELECT
   'First' || trunc(random()*1000)::int,
   'Last' || trunc(random()*1000)::int,
   'City' || trunc(random()*500)::int,
   to_char(date '1950-01-01' + trunc(random()*25000)::int, 'YYYY-MM-DD'),
   'P' || lpad(trunc(random()*10000000)::int::text, 8, '0'),
   CASE WHEN random() < 0.5 THEN 'US' ELSE 'DE' END
FROM generate_series(1, 200000);
```

Run Analyze to update statistics

```
ANALYZE wanted;
```

Run a query: Query for any passport_number that is not in the table. Explain the query execution.

eva.knirsch@kiu.edu.ge

Postgres Bloom Filter

1. Create a bloom filter on 6 columns. Use default parameter settings.

2. Verify that the bloom filter was created.

   ```
   SELECT indexname, indexdef FROM pg_indexes WHERE tablename = 'wanted';
   ```

3. Explain:
   1. how many bit arrays will be created and filled?
   2. length of the bitarrays?
   3. Which "elements" will be hashed into one bitarray?
   4. How many hash functions are used?

4. Run the query on passport_number again and explain its execution.

5. Create a btree index on passport_number.

6. Run the query on passport_number again and explain its execution.

eva.knirsch@kiu.edu.ge

Postgres Bloom Filter

Evaluate the Postgres bloom filter:

1. What main disadvantages do you see?

2. What could be use cases for the Postgres bloom filter?

eva.knirsch@kiu.edu.ge

MongoDB

Basic Queries

Get some practice:

1. How many male teachers are in teacher collection?
2. Return documents of teachers that live in postalcode 4600. Do not show object_id.
3. Return all documents of male teachers in postalcode 4600.
4. Return the documents of teachers Krawinkel and Kaiser (or condition on any other two teachers)
5. Teacher Krawinkel teaches EN, DE and FR. Add the subjects to the document. (updateOne())
6. Teacher Alt teaches MA, PH and CH. Add the subjects to the document.

knirsch@htw-berlin.de

Setting Validation Rules

Set the following validation rules for teacher collection:

1. required fiedls: 't_name', 't_mail', 't_gender'.
2. birthday: needs to be a date field
3. gender needs to be an enum field that allows **only one** of the enum values (one gender!)

4. education needs to be an enum field with values "High School", "Bachelor" and "Master". Multiple values are allowed.

5. subjects needs to be an enum field with a couple of subjects as values. Multiple values are allowed.

knirsch@htw-berlin.de

Insert a Document into a Collection, Verify Validation

Insert the following document into the teacher collection. Use a command that inserts exactly one document.:

t_name: “Smith",
t_mail: “jenny.smith@super.xx",
t_subjects: ["DE", "EN", "FR"],
t_education: ["Bachelor“, "Master"],
t_dob: "1985-03-20",
t_remark: "I have 10 years of experience with private teaching. Students trust me and I will make students more successful. I charge very moderate prices.“,
t_gender: “f”

Insert should fail! Why? Correct the error.

knirsch@htw-berlin.de

Insert a Document into a Collection, Verify Validation

Insert the following document into the teacher collection. Use a command that inserts exactly one document.:

t_name: “Young",
t_mail: “jenny.Young@super.xx",
t_subjects: ["DE", "EN", "FR"], // one of the subjects should NOT be defined.
t_education: ["Bachelor“, "Master"],
t_dob: "1985-03-20",
t_remark: "I have 10 years of experience with private teaching. Students trust me and I will make students more successful. I charge very moderate prices.“,
t_gender: “f”

Insert should fail because one of the array elements not defined. Correct and insert.

knirsch@htw-berlin.de

Document Update: Add / Update an Array

1. The teacher documents imported do not have a subjects' array. Take one of the imported teacher documents, update the document and add a subjects' field.

2. Add another subject to the subjects array. Use function $addToSet and run the same update command twice.

3. Add yet another subject to the subjects array. Use function $push and run the same update command twice.

4. Remove one of the duplicate values out of the array.

Insert Embedded Documents

1. Choose one teacher and add an array of embedded documents holding the students that the teacher teaches:

   Rose Singer, 2014-03-28, 2021-01-15
   Donald Black, 2011-05-05, 2020-01-15
   Donald Blue, 2017-05-05, 2020-01-15

2. Run a query that searches the collection teacher for the student Donald.

knirsch@htw-berlin.de

Create a reference between documents

- Import the student collection
- Create a reference from a student in the student table to two teachers.
- Query the reference to find the teachers of the student in the teacher table.

knirsch@htw-berlin.de