BDT, Praktikumsbericht 2

Gruppe 21: Maximilian Neudert, Kai Pehns

Aufgabe 1

Postgres

```
insert into public.user values (generate_series(1,1000000));

\copy public.movie FROM '/pgpool/movielens/adjusted/1m/movies.dat' with (format
csv, delimiter ';');

\copy public.genre FROM '/pgpool/movielens/adjusted/1m/genres.dat' with (format
csv, delimiter ';');

\copy public.rating FROM '/pgpool/movielens/adjusted/1m/ratings.dat' with (format
csv, delimiter ';');
```

MongoDB

```
mongoimport \
-u prak21 -p prak21 \
--db prak21 \
--collection movies \
--file /mnt/datasets/Movielens/JSON/movies_1m.json
```

Couchbase

```
/opt/couchbase/bin/cbdocloader \
-u prak21 -p prak21 \
-b prak21 \
-n 127.0.0.1:8091 -v -m 100 \
/mnt/datasets/Movielens/couchbase/movies_1m.zip
```

Aufgabe 2

Ergebnis ist ein Auszug aus dem Output, wahlweise der `count`, wenn man die query um `count` erweitert.

Postgres

1. Ergebnis: `3883 rows affected.`

```
select * from movies
left join(
  select mid, string_agg(genre, ',') as "genres"
  from genres group by mid
) genres
using (mid)
left join(
  select mid, array_agg((uid, ratings.rating)) as "ratings"
  from ratings group by mid
) ratings
using (mid);
```

2. Ergebnis: `1000209`

```
select count(*) from ratings;
```

3. Ergebnis: `"American Beauty (1999)"`

```
select titleyear
from movies
join (
  select mid
  from ratings
  group by mid
  order by count(rating) desc limit 1
) q
using (mid);
```

4. Ergebnis: `"430 rows affected"`

```
select titleyear, rating_avg
from movies
join (
  select mid, round(avg(rating), 2) as rating_avg
  from ratings
  group by mid
) subq
using (mid)
where rating_avg >= 4
order by rating_avg desc;
```

5. Ergebnis: `"Seven Samurai (The Magnificent Seven) (Shichinin no samurai) (1954)"`

```
select titleyear
from movies
join (
  select mid, round(avg(rating), 2) as rating_avg
  from ratings
  group by mid
  having count(rating) >= 100
) subq
using (mid)
where rating_avg >= 4
order by rating_avg desc
limit 1;
```

6. Ergebnis: `"503 rows affected"`

```
select titleyear
from movies
join (
  select mid, genre
  from genres
  where genre like 'Action'
) subq
using (mid)
```

7. Ergebnis: `"Meet the Parents (2000)"`

```
select titleyear
from movies
join (
  select *
  from ratings
  where uid = 10
) subq
using (mid)
```

8. Ergebnis: `4169`

```
select uid
from ratings
group by uid
order by count(rating) desc
limit 1;
```

MongoDB

1. Ergebnis: `3883`

```
db.movies.find();
```

2. Ergebnis: `{ "_id" : 1, "total" : 1000209 }`

```
db.movies.aggregate([
  {
    $group: {
      _id: 1,
      total: {
        $sum: {
          $size: "$ratings"
        }
      }
    }
  }
]);
```

3. Ergebnis: `{ "_id" : 2858, "title" : "American Beauty (1999)", "ratingAmount" : 3428 }`

```
db.movies.aggregate([
  {
    $project: {
      _id: "$_id",
      title: "$title",
      ratingAmount: {
        $size: "$ratings"
      }
    }
  },
  {
    $sort: {
      ratingAmount: -1
    }
  },
  {
    $limit: 1
  }
]);
```

4. Ergebnis: `{ "count" : 430 }`

```
db.movies.aggregate([
  {
    $project: {
      _id: "$_id",
      title: "$title",
      averageRating: {
        $avg: "$ratings.rating"
      }
    }
  },
  {
    $match: {
      averageRating: {
        $gte: 4
      }
    }
  }
]);
```

5. Ergebnis: { "_id" : 2019, "title" : "Seven Samurai (The Magnificent Seven) (Shichinin no samurai) (1954)", "averageRating" : 4.560509554140127, "ratingAmount" : 628 }

```
db.movies.aggregate([
  {
    $project: {
      _id: true,
      title: true,
      averageRating: {
        $avg: "$ratings.rating"
      },
      ratingAmount: {
        $size: "$ratings"
      }
    }
  },
  {
    $match: {
      ratingAmount: {
        $gte: 100
      }
    }
  },
  {
    $sort: {
      averageRating: -1
    }
  },
  {
    $limit: 1
  }
]);
```

6. Ergebnis: { "count" : 503 }

```
db.movies.aggregate([
  {
    $match: {
      genres: "Action"
    }
  }
]);
```

7. Ergebnis: `{ "count" : 401 }`

```
db.movies.aggregate([
  {
    $match: {
      "ratings.userId": 10
    }
  }
]);
```

8. Ergebnis: `{ "_id" : 4169, "count" : 2314 }`

```
db.movies.aggregate([
  {
    $unwind: "$ratings"
  },
  {
    $group: {
      _id: "$ratings.userId",
      count: {
        $sum: 1
      }
    }
  },
  {
    $sort: {
      count: -1
    }
  },
  {
    $limit: 1
  }
]);
```

Couchbase

1. Ergebnis: `count: 3883`

```
select * from prak21;
```

2. Ergebnis: `countRatings: 1000209`

```
SELECT SUM(ARRAY_LENGTH(ratings)) as countRatings FROM prak21;
```

3. Ergebnis: `"American Beauty (1999)"`

```
select raw title from prak21 order by array_length(ratings) desc limit 1;
```

4. Ergebnis: `"rating_avg": 4.073059360730594, "title": "Best in Show (2000)"`

```
select title, rating_avg
from prak21
let rating_avg = (select raw avg(ratings.rating)
  from prak21.ratings as ratings)[0]
where rating_avg >= 4;
```

5. Ergebnis: `"rating_avg": 4.560509554140127, "title": "Seven Samurai (The Magnificent Seven) (Shichinin no samurai) (1954)"`

```
select title, rating_avg
from prak21
let rating_avg = array_avg(
  array tags.rating for tags in ratings end)
where array_length(
  array tags.rating for tags in ratings end) >= 100
order by rating_avg desc
limit 1;
```

6. Ergebnis: `"Get Carter (2000)"`

```
select raw title
from prak21
where "Action" in genres;
```

7. Ergebnis: `"Meet the Parents (2000)"`

```
select raw title
from prak21
where 10 in array tags.userId
for tags in ratings end;
```

8. Ergebnis: `"$1": 2314, "userId": 4169"`

```
select rating.userId, count(rating.userId)
from prak21 as d
unnest d.ratings as rating
group by rating.userId
order by count(rating.userId) desc
limit 1;
```