



Ciências  
ULisboa

# DATA MODELLING AND QUERYING

## ADVANCED DATABASES

**GROUP 1:** DANIELA VIEIRA, JOÃO RAIMUNDO, JOÃO RATO, MARIA VIEIRA

**PROFESSOR:** CÁTIA PESQUITA

18 October 2021



# INTRODUCTION



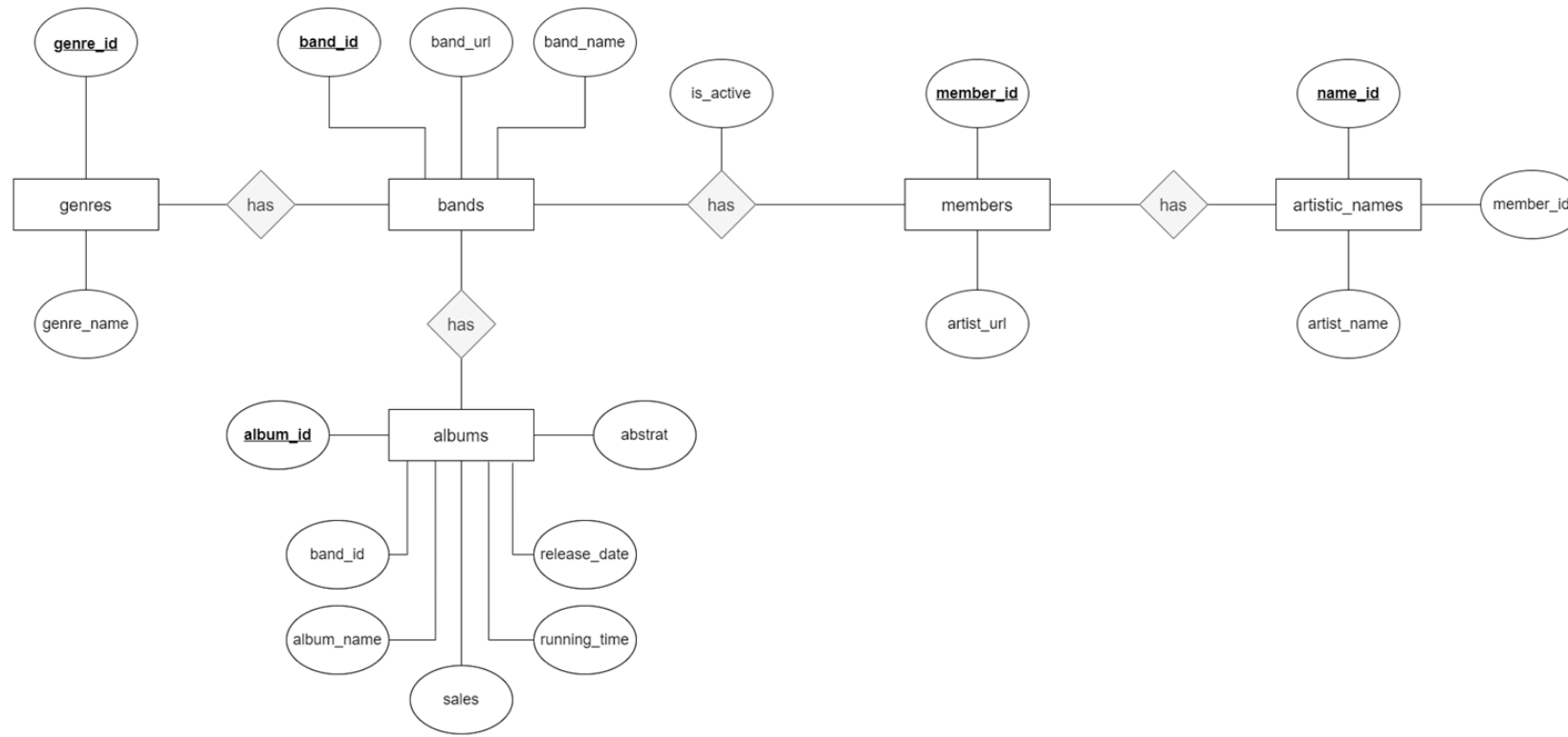
# IMPLEMENTATION



# RELATIONAL DATABASE

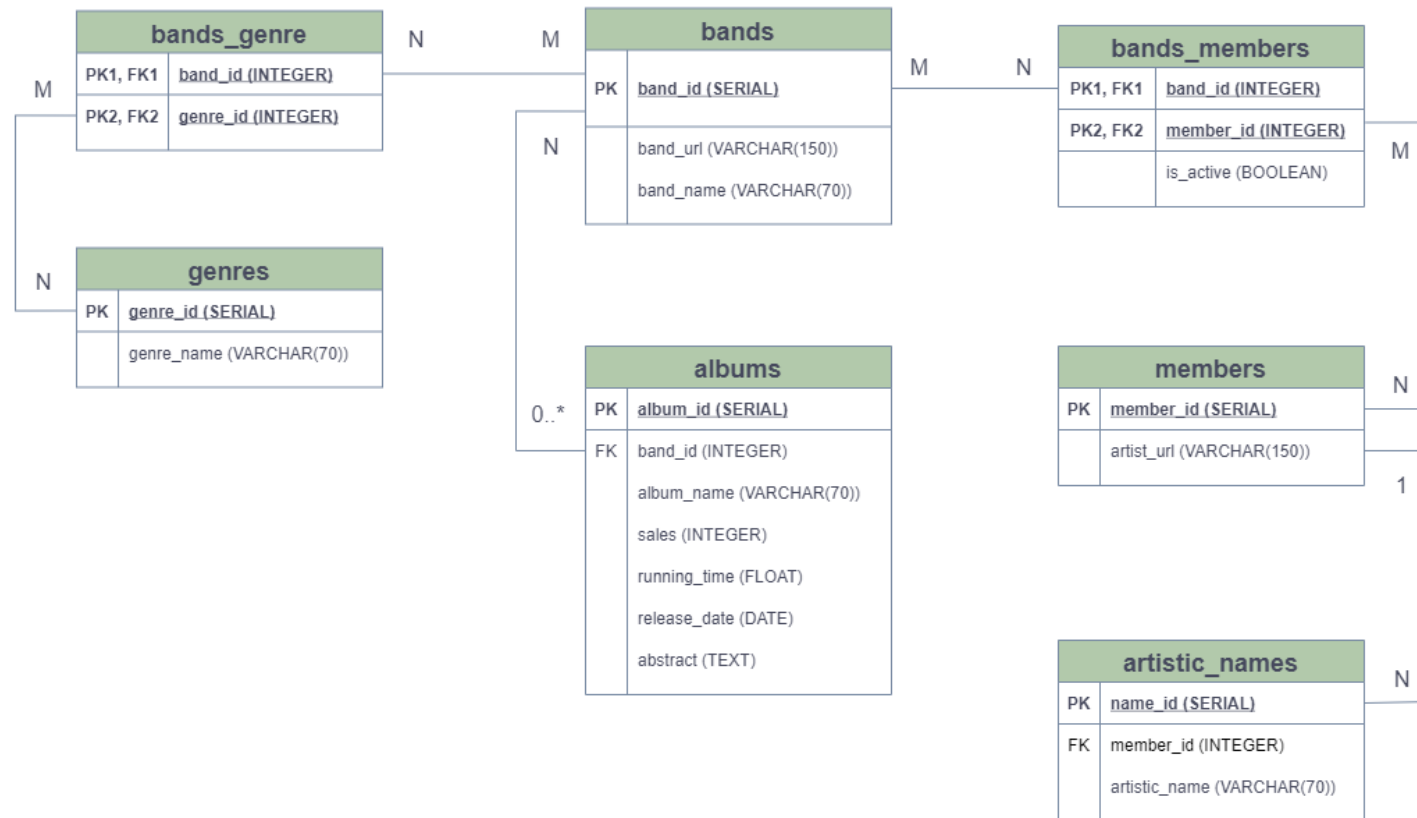
# RELATIONAL DATABASE

- Entity-Relationship Model



# RELATIONAL DATABASE

## ○ Relation Model





# DATA PROCESSING FOR RELATIONAL DATABASE

- Python for data cleanup;
- All tables have serial keys as primary keys;
- Merged data;
- Format changes -> YMD;
- Avoiding duplicate information;
- Abstract issues were removed;

# CREATE SQL TABLES

```
CREATE TABLE bands (  
    band_id SERIAL,  
    band_url VARCHAR(150) NOT NULL,  
    band_name VARCHAR(70) NOT NULL,  
    PRIMARY KEY(band_id),  
    UNIQUE(band_url)  
);  
  
CREATE TABLE albums (  
    album_id SERIAL,  
    band_id INTEGER,  
    album_name VARCHAR NOT NULL,  
    sales INTEGER,  
    running_time FLOAT(8) NOT NULL,  
    release_date DATE NOT NULL,  
    abstract TEXT,  
    PRIMARY KEY(album_id),  
    FOREIGN KEY(band_id) REFERENCES bands  
);  
  
CREATE TABLE members (  
    member_id SERIAL,  
    artist_url VARCHAR NOT NULL,  
    PRIMARY KEY(member_id),  
    UNIQUE (artist_url)  
);
```

```
CREATE TABLE bands_members (  
    band_id INTEGER,  
    member_id INTEGER,  
    is_active BOOLEAN NOT NULL,  
    PRIMARY KEY(band_id,member_id,is_active),  
    FOREIGN KEY(band_id) REFERENCES bands,  
    FOREIGN KEY(member_id) REFERENCES members  
);  
  
CREATE TABLE bands_members (  
    band_id INTEGER,  
    member_id INTEGER,  
    is_active BOOLEAN NOT NULL,  
    PRIMARY KEY(band_id,member_id,is_active),  
    FOREIGN KEY(band_id) REFERENCES bands,  
    FOREIGN KEY(member_id) REFERENCES members  
);  
  
CREATE TABLE artistic_names (  
    name_id SERIAL,  
    member_id INTEGER,  
    artistic_name VARCHAR(70) NOT NULL,  
    PRIMARY KEY(name_id),  
    FOREIGN KEY(member_id) REFERENCES members  
);
```

```
CREATE TABLE genres (  
    genre_id SERIAL,  
    genre_name VARCHAR(70) NOT NULL,  
    PRIMARY KEY(genre_id)  
);  
  
CREATE TABLE bands_genre (  
    band_id INTEGER,  
    genre_id INTEGER,  
    PRIMARY KEY(band_id,genre_id),  
    FOREIGN KEY (band_id) REFERENCES bands,  
    FOREIGN KEY (genre_id) REFERENCES genres  
);
```



# POPULATING THE DATABASE

```
-- Populate bands table
COPY "bands" FROM '~/bands_data/bands.csv' DELIMITER E'\t' CSV HEADER;

-- Populate genres table
COPY "genres" FROM '~/bands_data/genres.csv' DELIMITER E'\t' CSV HEADER;

-- Populate bands_genre table
COPY "bands_genre" FROM '~/bands_data/bands_genre.csv' DELIMITER E'\t' CSV HEADER;

-- Populate bands_genre table
COPY "bands_genre" FROM '~/bands_data/bands_genre.csv' DELIMITER E'\t' CSV HEADER;

-- Populate albums table
COPY "albums" FROM '~/bands_data/albums.csv' DELIMITER E'\t' CSV HEADER;

-- Populate members table
COPY "members" FROM '~/bands_data/members.csv' DELIMITER E'\t' CSV HEADER;

-- Populate bands_members table
COPY "bands_members" FROM '~/bands_data/bands_members.csv' DELIMITER E'\t' CSV HEADER;

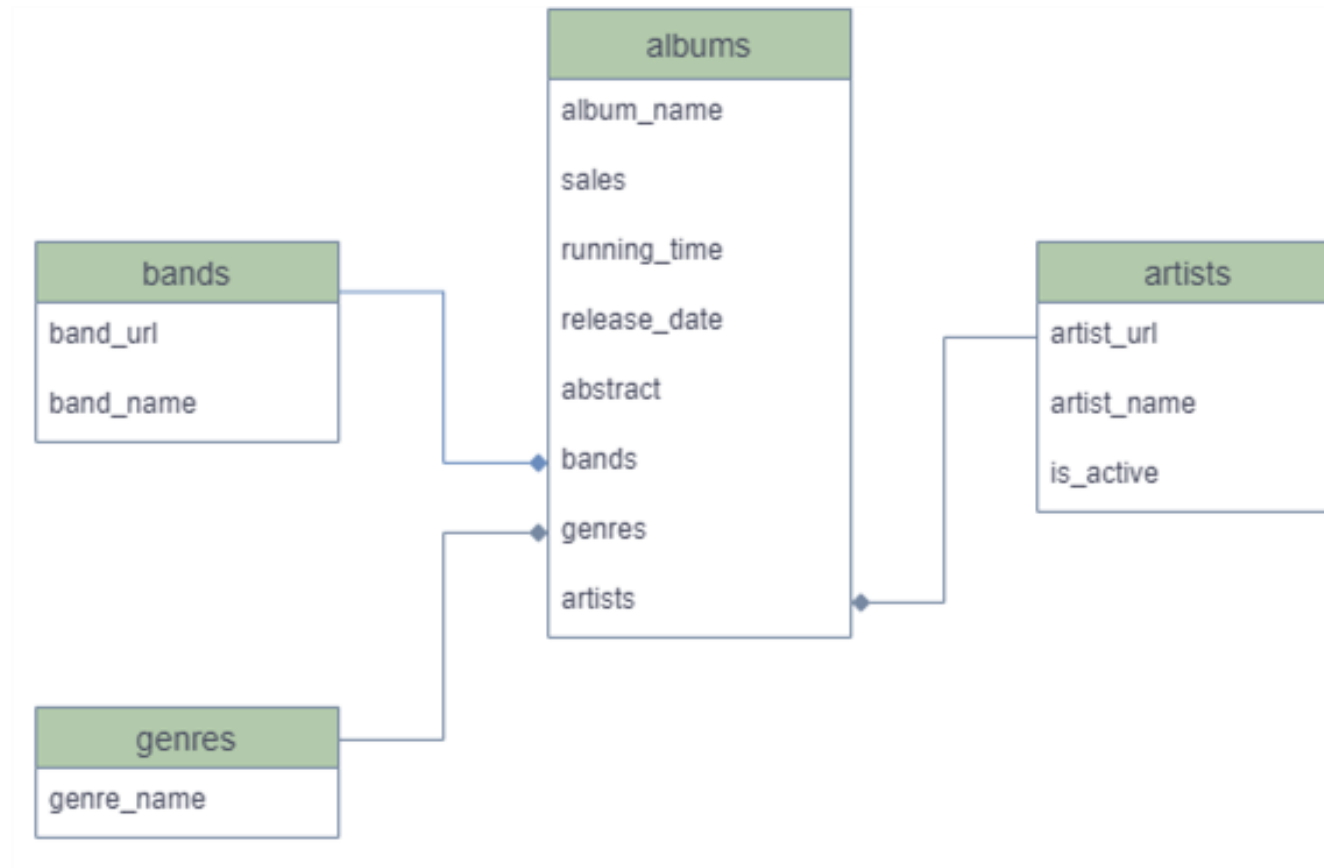
-- Populate artistic_names table
COPY "artistic_names" FROM '~/bands_data/artistic_names.csv' DELIMITER E'\t' CSV HEADER;
```



# NoSQL DATABASE

# NoSQL DATABASE

- Aggregation schema



# CREATING ARRAYS FOR BANDS, GENRES AND ARTISTS

```
# get bands
def getBand(url):
    # add bands
    name = None
    with open("mongo_csv/bands_mongoDB.txt","r",encoding="utf-8") as file:
        for line in file:
            line = line.split("\t")
            band_url = line[0]
            band_name = line[1].strip('\n')
            if url == band_url:
                name = band_name

    # verify if the band associated with the band_url exist for the album's band,
    # otherwise the key for the bands list/array in the albums dictionary will not
    # be created, and the band list/array will not be appended to the albums dictionary.

    return name if name else None
```

# CREATING ALBUMS COLLECTION

```
# _____ Albums Dictionary _____ //
```

```
# import albums data
albums_albums = open("mongo_csv/albums_mongoDB.txt", "r", encoding="utf-8")

# open a list that will contain one dictionary for each band
albums = []

# open a set of unique band URLs
album_set = set()
|
# add albums. Create a dictionary for the albums collection
for line in albums_albums:
    line = line.split("\t")

    # transform realease date str into datetime
    year = int(line[4].split("-")[0])
    month = int(line[4].split("-")[1])
    day = int(line[4].split("-")[2])
    date_str = line[4]
    date = datetime.strptime(date_str, '%Y-%m-%d')

    band_url = line[0]
    album_name = line[1].strip('\n')
    sales = int(line[2])
    running_time = float(line[3])
    release_date = date
    abstract = line[5].strip('\n').strip('\n\n')

    # add new band URL to albums set and create a dictionary entry
    dict_album = {
        "band_url": band_url,
        "album_name": album_name,
        "sales": sales,
        "running_time": running_time,
        "release_date": release_date,
        "abstract": abstract,
    }

    # in case of the inexistence of the bands, genres and artists list/array/dictionary associated with the band_url
    # for an album, the list/array/dictionary will not be appended to the album dictionary, aka document,
    # and the a key will not be created.

    # get bands
    bands = getBand(band_url)
    if bands:
        dict_album["bands"] = bands

    # get genre
    genres = getGenre(band_url)
    if genres:
        dict_album["genres"] = genres

    # get artists
    artists = getArtists(band_url)
    if artists:
        dict_album["artists"] = artists

    # add an album dictionary to albums list, for each band
    albums.append(dict_album)
```

# POPULATING NoSQL DATABASE

```
# connect to Mongo Atlas DB (replace <password>)
ca = certifi.where()
cluster = MongoClient("mongodb+srv://bandsDB:<password>@bandscluster.1eit9.mongodb.net/myFirstDatabase?retryWrites=true&w=majority",tlsCAFile=ca)

db = cluster["bandsDB"]
collection = db["albums"]

#_____Populate MongoDB with albums dictionaries_____//

for album in albums:
    collection.insert_one(album)|
```




# QUERIES

## QUERY 1

Update to 1980 the release date of all albums for the genre Math Rock, which were released in the 90's with an abstract over 200 characters, and that had the most sales.

## QUERY 2

Update to 0 the sales from the album with the most sales in the first decade of the year 2000, and which the running time is longer than 45 minutes.





# SQL QUERIES AND RESULTS





# QUERY 1

## Result before Update:

band_id	release_date	sales	genres	abstract
6171	1994-10-24	9779	Math rock	At Action Park is the first full-length record by Shellac, released in 1994. The title is unrelated to the infamous New Jersey theme park, Action Park, which closed in 1996 due to numerous fatalities. The drummer, Todd Trainer, came up with the title of the fictional park because it sounded cool.

## Update:

```
UPDATE albums
SET release_date = '1980/01/01'
WHERE albums.release_date = (SELECT albums.release_date
FROM (((albums
INNER JOIN bands ON albums.band_id = bands.band_id)
INNER JOIN bands_genre ON bands.band_id = bands_genre.band_id)
INNER JOIN genres ON bands_genre.genre_id = genres.genre_id)
WHERE genres.genre_name = 'Math rock'
AND albums.release_date >= '1990/01/01'
AND albums.release_date <= '1999/12/31'
AND LENGTH(albums.abstract) > 200
ORDER BY albums.sales
DESC
LIMIT 1
);
```

## Result after Update:

band_id	release_date	sales	genre_name	abstract
6171	1980-01-01	9779	Math rock	At Action Park is the first full-length record by Shellac, released in 1994. The title is unrelated to the infamous New Jersey theme park, Action Park, which closed in 1996 due to numerous fatalities. The drummer, Todd Trainer, came up with the title of the fictional park because it sounded cool.

# QUERY 2

Result before Update:

band_id	sales	release_date	running_time
464	12410000	2000-11-13	76.933334

Update:

```
UPDATE albums
SET sales = 0
WHERE albums.sales = (SELECT A.sales
                      FROM albums AS A
                      WHERE A.running_time > '45'
                      AND A.release_date >= '2000/01/01'
                      AND A.release_date <= '2010/12/31'
                      ORDER BY A.sales
                      DESC
                      LIMIT 1
);
```

Result after Update:

band_id	sales	release_date	running_time
1664	0	2002-06-25	46.1




# NOSQL QUERIES AND RESULTS

# QUERY 1

Select before update: `db.albums.find({`  
    `$and :`  
    `[`  
        `{genres: "Math rock"},`  
        `{release_date: {$gte: ISODate("1990-01-01T00:00:00.000+00:00")}},`  
        `{release_date: {$lte: ISODate("1999-12-31T00:00:00.000+00:00")}},`  
        `{abstract: /^[\\s\\S]{200,}$/ }`  
    `]`  
    `}).sort({sales: -1}).limit(1)`

Result:

```
{ _id: ObjectId("6168bd6e416ed4db14e46417"),
  band_url: 'http://dbpedia.org/resource/Shellac_(band)',
  album_name: 'At Action Park',
  sales: 9779,
  running_time: 37.05,
  release_date: 1994-10-24T00:00:00.000Z,
  abstract: 'At Action Park is the first full-length record by Shellac, released in 1994.
            The title is unrelated to the infamous New Jersey theme park, Action Park,
            which closed in 1996 due to numerous fatalities. The drummer, Todd Trainer,
            came up with the title of the fictional park because it sounded cool.',
  bands: 'Shellac (band)',
  genres: [ 'Noise rock', 'Math rock', 'Post-hardcore' ],
  artists:
    [ { artist_url: 'http://dbpedia.org/resource/Todd_Trainer',
        artistic_name: [ 'Todd Trainer' ],
        is_active: 'True' },
      { artist_url: 'http://dbpedia.org/resource/Bob_Weston',
        artistic_name: [ 'Bob Weston' ],
        is_active: 'True' },
      { artist_url: 'http://dbpedia.org/resource/Steve_Alбини',
        artistic_name: [ 'Steve Alбини' ],
        is_active: 'True' } ] }
```



# QUERY 1


Update:

```
> db.albums.updateOne({_id: ObjectId("6168bd6e416ed4db14e46417")},
                        {$set: {release_date: ISODate("1980-01-01T00:00:00.000+00:00")}}
                        )
```

```
< { acknowledged: true,
    insertedId: null,
    matchedCount: 1,
    modifiedCount: 1,
    upsertedCount: 0 }
```

```
db.albums.find({_id: ObjectId("6168bd6e416ed4db14e46417")})
```

```
{ _id: ObjectId("6168bd6e416ed4db14e46417"),
  band_url: 'http://dbpedia.org/resource/Shellac_(band)',
  album_name: 'At Action Park',
  sales: 9779,
  running_time: 37.05,
  release_date: 1980-01-01T00:00:00.000Z,
  abstract: 'At Action Park is the first full-length record by Shellac,
            released in 1994. The title is unrelated to the infamous
            New Jersey theme park, Action Park, which closed in 1996 due
            to numerous fatalities. The drummer, Todd Trainer, came up with
            the title of the fictional park because it sounded cool.',
  bands: 'Shellac (band)',
  genres: [ 'Noise rock', 'Math rock', 'Post-hardcore' ],
  artists:
    [ { artist_url: 'http://dbpedia.org/resource/Todd_Trainer',
        artistic_name: [ 'Todd Trainer' ],
        is_active: 'True' },
      { artist_url: 'http://dbpedia.org/resource/Bob_Weston',
        artistic_name: [ 'Bob Weston' ],
        is_active: 'True' },
      { artist_url: 'http://dbpedia.org/resource/Steve_Alбини',
        artistic_name: [ 'Steve Alбини' ],
        is_active: 'True' } ] }
```



# QUERY 2

Select before update:

```
db.albums.find({
  $and :
  [
    {running_time: {$gte: 45}},
    {release_date: {$gte: ISODate("2000-01-01T00:00:00.000+00:00")}},
    {release_date: {$lte: ISODate("2010-12-31T00:00:00.000+00:00")}}
  ]
}).sort({sales: -1}).limit(1)
```

Result:



```
{ _id: ObjectId("6168b8f6416ed4db14e40d38"),
  band_url: 'http://dbpedia.org/resource/The_Beatles',
  album_name: '1 (Beatles album)',
  sales: 12410000,
  running_time: 79.13333333,
  release_date: 2000-11-13T00:00:00.000Z,
  abstract: '1 is a compilation album by English rock band the Beatles, originally
    released on 13 November 2000. The album features virtually every number-one
    single the band achieved in the United Kingdom and United States from 1962 to 1970.
    Issued on the 30th anniversary of the band\'s break-up, it was their first compilation
    available on only one compact disc. 1 was a commercial success and topped the charts worldwide.
    It has sold over 31 million copies. In addition, 1 is the fourth best-selling album in the US
    since Nielsen SoundScan began tracking US album sales in January 1991 and the best-selling album
    of the decade (2000 to 2009) in the US, as well as the best-selling album of the decade worldwide.
    1 was remastered and reissued in September 2011, and was remixed and reissued again in several
    different deluxe editions in November 2015, the most comprehensive of which is a three-disc set
    entitled 1+, which includes video discs of the band\'s music videos. As of June 2015, 1 is the
    sixth best-selling album of the 21st century in the UK, having sold over 3.1 million copies.',
  bands: 'The Beatles',
  genres: [ 'Pop music', 'Rock music' ],
  artists:
  [ { artist_url: 'http://dbpedia.org/resource/George_Harrison',
      artistic_name: [ 'George Harrison' ],
      is_active: 'False' },
    { artist_url: 'http://dbpedia.org/resource/John_Lennon',
      artistic_name: [ 'John Lennon' ],
      is_active: 'False' },
    { artist_url: 'http://dbpedia.org/resource/Ringo_Starr',
      artistic_name: [ 'Ringo Starr' ],
      is_active: 'False' },
    { artist_url: 'http://dbpedia.org/resource/Paul_McCartney',
      artistic_name: [ 'Paul McCartney' ],
      is_active: 'False' },
    { artist_url: 'http://dbpedia.org/resource/Paul_McCartney',
      artistic_name: [ 'Sir Paul McCartney' ],
      is_active: 'False' } ] }
```

# QUERY 2

## Update:

```
> db.albums.updateOne({_id: ObjectId("6168b8f6416ed4db14e40d38")},
                      {$set: {sales: 0 }}
                      )
```

```
< { acknowledged: true,
    insertedId: null,
    matchedCount: 1,
    modifiedCount: 1,
    upsertedCount: 0 }
```



```
db.albums.find({_id: ObjectId("6168b8f6416ed4db14e40d38")})
```

```
{ _id: ObjectId("6168b8f6416ed4db14e40d38"),
  band_url: 'http://dbpedia.org/resource/The_Beatles',
  album_name: '1 (Beatles album)',
  sales: 0,
  running_time: 79.13333333,
  release_date: 2000-11-13T00:00:00.000Z,
  abstract: '1 is a compilation album by English rock band the Beatles,
            originally released on 13 November 2000. The album features
            virtually every number-one single the band achieved in the
            United Kingdom and United States from 1962 to 1970.
            Issued on the 30th anniversary of the band\'s break-up, it was
            their first compilation available on only one compact disc.
            1 was a commercial success and topped the charts worldwide.
            It has sold over 31 million copies. In addition, 1 is the fourth
            best-selling album in the US since Nielsen SoundScan began tracking
            US album sales in January 1991 and the best-selling album of the
            decade (2000 to 2009) in the US, as well as the best-selling album
            of the decade worldwide. 1 was remastered and reissued in September 2011,
            and was remixed and reissued again in several different deluxe editions
            in November 2015, the most comprehensive of which is a three-disc set
            entitled 1+, which includes video discs of the band\'s music videos.
            As of June 2015, 1 is the sixth best-selling album of the 21st century
            in the UK, having sold over 3.1 million copies.',
  bands: 'The Beatles',
  genres: [ 'Pop music', 'Rock music' ],
  artists:
    [ { artist_url: 'http://dbpedia.org/resource/George_Harrison',
        artistic_name: [ 'George Harrison' ],
        is_active: 'False' },
      { artist_url: 'http://dbpedia.org/resource/John_Lennon',
        artistic_name: [ 'John Lennon' ],
        is_active: 'False' },
      { artist_url: 'http://dbpedia.org/resource/Ringo_Starr',
        artistic_name: [ 'Ringo Starr' ],
        is_active: 'False' },
      { artist_url: 'http://dbpedia.org/resource/Paul_McCartney',
        artistic_name: [ 'Paul McCartney' ],
        is_active: 'False' },
      { artist_url: 'http://dbpedia.org/resource/Paul_McCartney',
        artistic_name: [ 'Sir Paul McCartney' ],
        is_active: 'False' } ] }
```



# CONCLUSION