# **Information System Management Report** Léandre BROSSIER, Mathys DECKER, Benoît HUA, Kylie WU - ESILV A4 CDOF3

Find the \DATABASE\WMLMS\_Step\_3\_Database\_Setup.sql in the zip file to get all our sql queries displayed in this document!

# Chapter 1: Preparing the relational schema

#### 1. Data search

https://www.kaggle.com/datasets/yamaerenay/spotify-dataset-19212020-600k-tracks, with "artists.csv" and "tracks.csv" (more than 500 000 lines in each one). We chose to split "tracks.csv" into "tracks.csv"

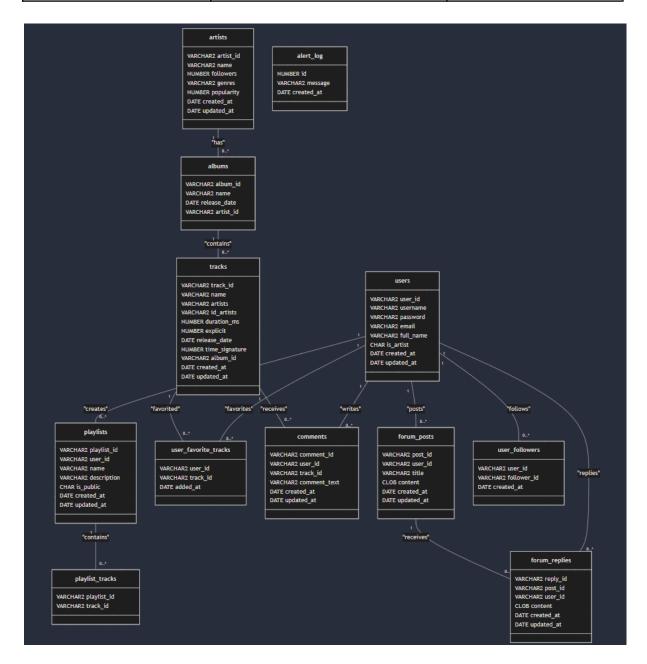
The dataset used for this project is the Spotify Dataset from Kaggle:

and "tracks\_audio\_features.csv" thanks to a python script (available in the Github repository), with the columns ['id', 'name', 'artists', 'id\_artists', 'duration\_ms', 'explicit', 'release\_date', 'time\_signature'] for tracks.csv and ['id', 'danceability', 'energy', 'key', 'loudness', 'mode', 'speechiness', 'acousticness', 'instrumentalness', 'liveness', 'valence', 'tempo'] for "tracks\_audio\_features.csv". We also created and added additional tables: albums, playlists, users, user\_favorite\_tracks (and some more throughout the project, see the board below).

# 2. Design the relational schema

| Table name            | Primary key  | Foreign keys      |  |
|-----------------------|--|-------------------|--|
| artists               | artist_id VARCHAR2(36)   | -                 |  |
| albums                | album_id VARCHAR2(36)  | artist_id         |  |
| tracks                | track_id VARCHAR2(36)  | album_id          |  |
| tracks_audio_features | track_id VARCHAR2(36)  | track_id          |  |
| users                 | user_id VARCHAR2(36)   | -                 |  |
| playlists             | playlist_id VARCHAR2(36)                                       | user_id           |  |
| user_favorite_tracks  | user_id VARCHAR2(36)<br>track_id VARCHAR2(36)                  | user_id, track_id |  |
| alert_log             | id NUMBER  | -                 |  |
| playlist_tracks       | playlist_tracks playlist_id VARCHAR2(36) track_id VARCHAR2(36) |                   |  |
| comments              | comment_id VARCHAR2(36)  | user_id, track_id |  |

| forum_posts    | post_id VARCHAR2(36)                             | user_id              |
|----------------|--|----------------------|
| forum_replies  | reply_id VARCHAR2(36)                            | post_id, user_id     |
| user_followers | user_id VARCHAR2(36)<br>follower_id VARCHAR2(36) | user_id, follower_id |
| logs           | log_id VARCHAR2(36)                              | user_id              |



#### Relationships:

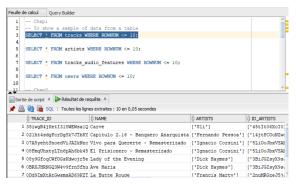
- The artists table is independent and serves as a reference for albums
- Each album is associated with one artist through the artist\_id foreign key
- Tracks are linked to albums via the album\_id foreign key
- Tracks\_audio\_features contains additional information for each track, linked by track\_id

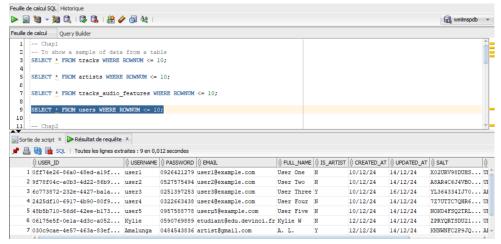
- Users table stores user information independently
- Playlists are created by users, hence the user\_id foreign key
- User favorite tracks is a junction table linking users to their favorite tracks
- Playlist tracks links playlists to tracks, forming a many-to-many relationship
- Comments are associated with both users and tracks
- Forum posts are created by users
- Forum\_replies are linked to forum\_posts and created by users
- User followers establishes a self-referential relationship within the users table
- Logs track user activities and are linked to the users table

#### Integrity constraints include:

- Primary keys are unique and not null
- Username and email in the users table are unique
- Generate uuid function for unique identifiers (in the IDs)
- Date fields have default values of SYSDATE or SYSTIMESTAMP
- Boolean fields use CHAR(1) with 'Y' or 'N' values
- ON DELETE CASCADE is used for some foreign keys to maintain data consistency
- Some tables use composite primary keys (i.e. user\_followers, playlist\_tracks)
- Text fields like comment\_text and event\_description have specified maximum lengths
- CLOB data type is used for longer text content in forum posts and replies
- Populating the tables (lines 170-325 in the WMLMS\_tables\_creation.sql)

We imported the .csv data into our database according to the defined schema, and populated the tables needed.





### Chapter 2: Security and user management

### 1. Creating users

Application user (app): manages application operations, application user with general privileges

Administrator (admin): administrator with full access, for database administration tasks

Regular user (regular user): standard user with limited read-only access to specific views

## 2. Assigning privileges

Privileges are granted specifically to each type of user:

#### For the app user:

- Has CONNECT, RESOURCE and ALL privileges
- Can select from app view tracks, app view playlists, and app view user playlists

#### For the admin user:

- Has CONNECT and DBA privileges
- Has UNLIMITED TABLESPACE privilege
- Full access to all tables (SELECT, INSERT, UPDATE, DELETE)
- Can SELECT from all admin views (admin\_view\_users, admin\_view\_artists, admin\_view\_tracks) and app views (app\_view\_tracks, app\_view\_playlists, app\_view\_user\_playlists)

### For the regular\_user:

- Has CONNECT privilege
- Read-only (SELECT) access to app\_view\_tracks, app\_view\_playlists, and app\_view\_user\_playlists

## 3. Securing the system

- a) There is a generate\_uuid function for unique identifiers, and a trigger-based UUID assignment, which ensures UUID generation when inserting new users.
- b) The use of views (app\_view\_tracks, app\_view\_playlists...) allows limiting access to sensitive data and precisely controlling which information is accessible to each type of user.
- c) Passwords are hashed and salted before storage:
  - We modified the users table to store hashed passwords (ALTER TABLE users MODIFY password VARCHAR2(256))
  - A salt field was added to enhance hashing security (ALTER TABLE users ADD (salt VARCHAR2(64));)
  - custom hash function (custom hash) to generate a unique hash for each password
  - Function to hash passwords with salt (hash\_password), combining the password and salt for increased security
  - The insertion trigger for users (trg\_users\_before\_insert) was modified to automatically generate a salt, hash the password, and store both in the users table
  - A similar trigger (trg\_users\_before\_update) was created to handle password updates, ensuring that passwords are always properly hashed and salted when changed

### Chapter 3: Queries and optimization

1. Designing SQL queries (lines 436 to 499 in WMLMS\_tables\_creation.sql)

JOIN between tables to show the tracks names, artists, albums and playlists for the tracks that are more than 5 minutes long.

```
EXPLAIN PLAN FOR
SELECT t.name AS track_name, a.name AS artist_name, al.name AS album_name,
p.name AS playlist_name
FROM tracks t
JOIN artists a ON INSTR(t.id_artists, a.artist_id) > 0
LEFT JOIN albums al ON t.album_id = al.album_id
LEFT JOIN user_favorite_tracks uft ON t.track_id = uft.track_id
LEFT JOIN users u ON uft.user_id = u.user_id
LEFT JOIN playlists p ON u.user_id = p.user_id
WHERE t.duration_ms > 300000;
```

```
SELECT * FROM TABLE(DBMS_XPLAN.DISPLAY); -- 0,313s
```

Analytical function AVG() to compute the average energy of the tracks per year or release

```
EXPLAIN PLAN FOR
SELECT EXTRACT(YEAR FROM t.release_date) AS release_year, AVG(taf.energy) AS
avg_energy
FROM tracks t
JOIN tracks_audio_features taf ON t.track_id = taf.track_id
GROUP BY EXTRACT(YEAR FROM t.release_date)
ORDER BY release_year DESC;
```

```
SELECT * FROM TABLE(DBMS_XPLAN.DISPLAY); -- 0,101s
```

Analytical function COUNT() to count the number of tracks per artist and how only those who have more than 5 tracks

```
SELECT a.name AS artist_name, COUNT(t.track_id) AS track_count
FROM artists a
JOIN tracks t ON INSTR(t.id_artists, a.artist_id) > 0
GROUP BY a.name
HAVING COUNT(t.track_id) > 5
ORDER BY track_count DESC;
```

Analytical function SUM() to have the total number of followers for each genre of artist

```
SELECT a.genres, SUM(a.followers) AS total_followers
FROM artists a
GROUP BY a.genres
```

```
ORDER BY total_followers DESC;
```

Nested query to retrieve artists whose popularity is above the average of all artists

```
SELECT name, popularity FROM artists
WHERE popularity > (SELECT AVG(popularity) FROM artists);
```

Subquery to identify popular artists compared to the average number of followers in the database

```
SELECT name, followers FROM artists
WHERE followers > (SELECT AVG(followers) FROM artists);
```

Window function to rank artists by number of followers

```
SELECT name, followers, RANK() OVER (ORDER BY followers DESC) AS follower_rank FROM artists;
```

Window function to rank artists with more than 1 million followers

```
SELECT name, followers, RANK() OVER (ORDER BY followers DESC) AS follower_rank FROM artists
WHERE followers > 1000000;
```

Window function to rank artists based on the total number of tracks they have, and display the number of tracks for each artist

```
SELECT a.name AS artist_name, COUNT(t.track_id) AS track_count, RANK() OVER (ORDER BY COUNT(t.track_id) DESC) AS track_rank
FROM artists a
JOIN tracks t ON INSTR(t.id_artists, a.artist_id) > 0
GROUP BY a.name;
```

## **2. Query optimization** (lines 502 to 540)

With the creation of indexes on frequently used columns in joins and WHERE conditions, and materialized views to speed up searches, then rewriting certain queries for greater efficiency:

```
CREATE INDEX idx_user_favorite_tracks_user_id ON user_favorite_tracks(user_id);
CREATE INDEX idx_tracks_album_id ON tracks(album_id);
CREATE INDEX idx_tracks_id_artists ON tracks(id_artists);
CREATE INDEX idx_playlists_user_id ON playlists(user_id);
```

Rewriting of some queries to enhance their performances

Show the tracks names, artists, albums and related playlists for the tracks that are more than 5 minutes long

```
EXPLAIN PLAN FOR
SELECT
    t.track_id,
    t.name AS track_name,
    a.name AS artist_name,
    al.name AS album_name,
    u.username AS user_name,
    p.name AS playlist_name
FROM tracks t
JOIN artists a ON INSTR(t.id_artists, a.artist_id) > 0
LEFT JOIN albums al ON t.album_id = al.album_id
LEFT JOIN user_favorite_tracks uft ON t.track_id = uft.track_id
LEFT JOIN users u ON uft.user_id = u.user_id
LEFT JOIN playlists p ON u.user_id = p.user_id
WHERE t.duration_ms > 300000;
```

```
SELECT * FROM TABLE(DBMS_XPLAN.DISPLAY); -- 0,034s
```

#### Materialized view:

### Get the average energy per year:

```
CREATE MATERIALIZED VIEW mv_avg_energy_per_year AS

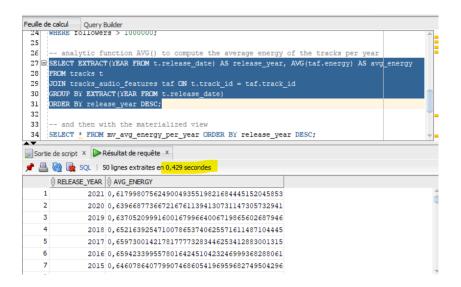
SELECT EXTRACT(YEAR FROM t.release_date) AS release_year, AVG(taf.energy) AS

avg_energy

FROM tracks t

JOIN tracks_audio_features taf ON t.track_id = taf.track_id

GROUP BY EXTRACT(YEAR FROM t.release_date);
```



Use the materialized view mv\_avg\_energy\_per\_year in a query for instance :

```
SELECT * FROM mv_avg_energy_per_year ORDER BY release_year DESC; -- 0,009s
```

# **Chapter 4: Automation of the Information System**

### 1. Triggers

## a) Triggers for UUID generation:

```
CREATE OR REPLACE TRIGGER trg_artists_id
BEFORE
INSERT ON artists
FOR EACH ROW
BEGIN
    If :NEW.artist_id IS NULL THEN
        :NEW.artist_id := generate_uuid();
    END IF;
END;
//
```

```
CREATE OR REPLACE TRIGGER trg_albums_id
BEFORE INSERT ON albums
FOR EACH ROW
BEGIN
    If :NEW.album_id IS NULL THEN
        :NEW.album_id := generate_uuid();
    END IF;
END;
/
```

We also have similar scripts for trg\_tracks\_id, trg\_users\_id, trg\_playlists\_id, trg\_comments\_id, trg\_forum\_posts\_id, trg\_forum\_replies\_id or trg\_logs\_id.

b) Triggers for table updates (lines 547 to 578):

```
CREATE OR REPLACE TRIGGER trg_update_artists_updated_at
BEFORE UPDATE ON artists
FOR EACH ROW
BEGIN
   :NEW.updated_at := SYSDATE;
END;
/
```

```
CREATE OR REPLACE TRIGGER trg_update_tracks_updated_at
BEFORE UPDATE ON tracks
FOR EACH ROW
BEGIN
   :NEW.updated_at := SYSDATE;
END;
/
```

```
CREATE OR REPLACE TRIGGER trg_update_users_updated_at
BEFORE UPDATE ON users
FOR EACH ROW
BEGIN
   :NEW.updated_at := SYSDATE;
END;
/
```

```
CREATE OR REPLACE TRIGGER trg_update_playlists_updated_at
BEFORE UPDATE ON playlists
FOR EACH ROW
BEGIN
   :NEW.updated_at := SYSDATE;
END;
/
```

To add an user in the artists table when their artist status is updated :

```
CREATE OR REPLACE TRIGGER trg_insert_artist

AFTER UPDATE OF is_artist ON users

FOR EACH ROW

BEGIN

IF :NEW.is_artist = 'Y' AND :OLD.is_artist = 'N' THEN

INSERT INTO artists (artist_id, name, followers, genres, popularity, created_at, updated_at)

VALUES (generate_uuid(), :NEW.full_name, 0, NULL, 0, SYSDATE, SYSDATE);

END IF;

END;
/
```

c) Triggers for passwords hashing:

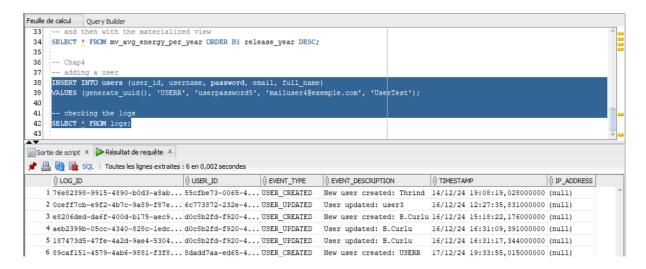
```
CREATE OR REPLACE TRIGGER trg_users_before_insert
BEFORE INSERT ON users
FOR EACH ROW
BEGIN
    IF :NEW.user_id IS NULL THEN
        :NEW.user_id := generate_uuid();
    END IF;
    :NEW.salt := DBMS_RANDOM.STRING('X', 64);
    :NEW.password := hash_password(:NEW.password, :NEW.salt);
END;
//
```

```
CREATE OR REPLACE TRIGGER trg_users_after_insert
AFTER INSERT ON users
FOR EACH ROW
BEGIN
```

```
add_log(:NEW.user_id, 'USER_CREATED', 'New user created: ' || :NEW.username,
NULL);
END;
/
```

Some additional details are available in the sql source file.

For instance, if we add a new user (or update), and then check the logs, we get the table updated:



## 2. Stored functions and procedures

Function to compute the total length of a playlist :

```
CREATE OR REPLACE FUNCTION get_playlist_duration(p_playlist_id VARCHAR2)
RETURN NUMBER
IS
    v_total_duration NUMBER;
BEGIN
    SELECT SUM(t.duration_ms)
    INTO v_total_duration
    FROM tracks t
    JOIN playlist_tracks pt ON t.track_id = pt.track_id
    WHERE pt.playlist_id = p_playlist_id;

RETURN NVL(v_total_duration, 0); -- returns 0 if no track is found
END;
//
```

Function to generate a monthly report on the number of added users :

```
CREATE OR REPLACE FUNCTION generate_monthly_report RETURN VARCHAR2 IS v_report VARCHAR2(4000);
BEGIN
```

```
SELECT "Nombre d utilisateurs ajoutés ce mois-ci : " || COUNT(*)
INTO v_report
FROM users
WHERE EXTRACT(MONTH FROM created_at) = EXTRACT(MONTH FROM SYSDATE)
AND EXTRACT(YEAR FROM created_at) = EXTRACT(YEAR FROM SYSDATE);

RETURN v_report;
END;
//
```

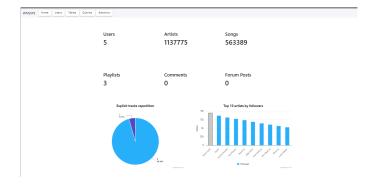
To see and check the results, we could use : SELECT generate\_monthly\_report() FROM DUAL;

Automatic execution of tasks (PL/SQL procedure):

This job will automatically execute the generate\_monthly\_report procedure at the beginning of each month, allowing for regular, automated reporting.

### 3. Advanced automation (bonus)

We have a dedicated page on our admin panel to display KPI and charts and gather important information automatically, it automatically refreshes and updates the displayed values by executing queries automatically. On this page, the administrator will have indicators on the status of its database like the number of registered users, the number of registered artists, the numbers of active forums, comments and songs. This statistics panel (in the administration tool) also provide two charts to display some characteristics of the dataset:



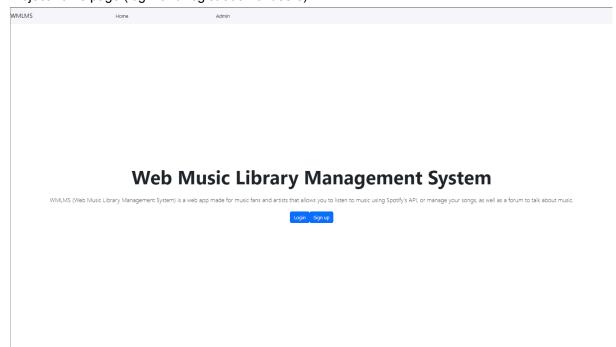
# Chapter 5 (bonus): Graphical interface

# 1. Creating a graphical interface

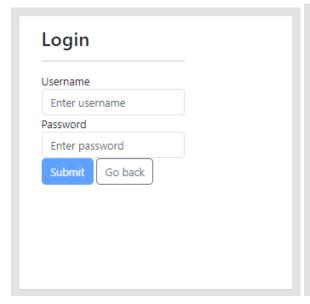
You will have access to the presentation video in the Presentation folder of the project.

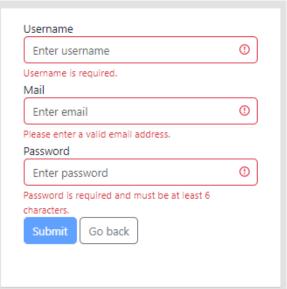
Our graphical interface is built around administration tasks on the dataset such as table visualisation, user details modification, queries executions and database statistics visualisation. Here is some of the main interfaces you will encounter on this web-app ( further user functions are not included in the project since it is a database administration project ):

## Project home page (login and registration of users):

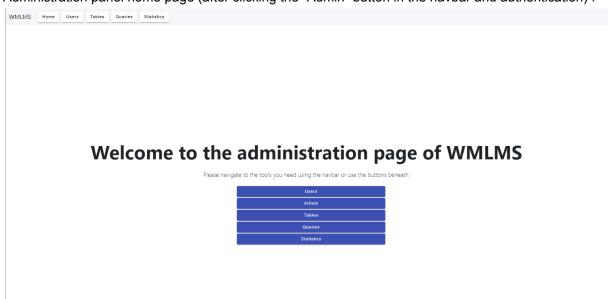


# User login and registration:

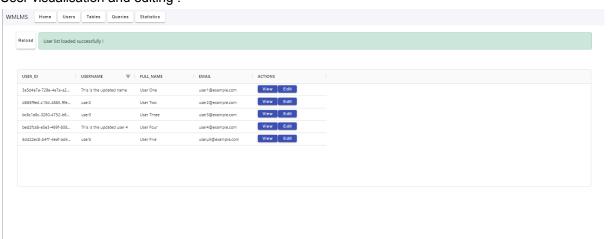




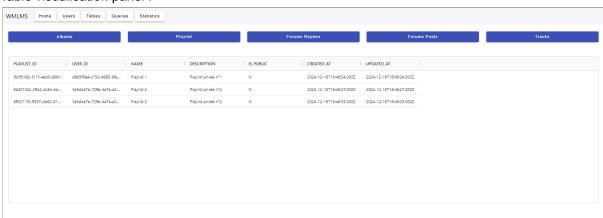
Administration panel home page (after clicking the "Admin" button in the navbar and authentication):



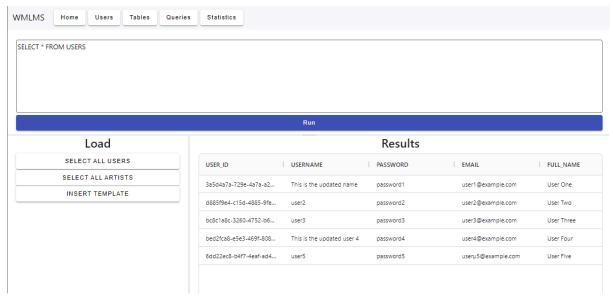
# User visualisation and editing:



#### Table visualisation panel:



# Queries panel:



# Statistics panel:

