

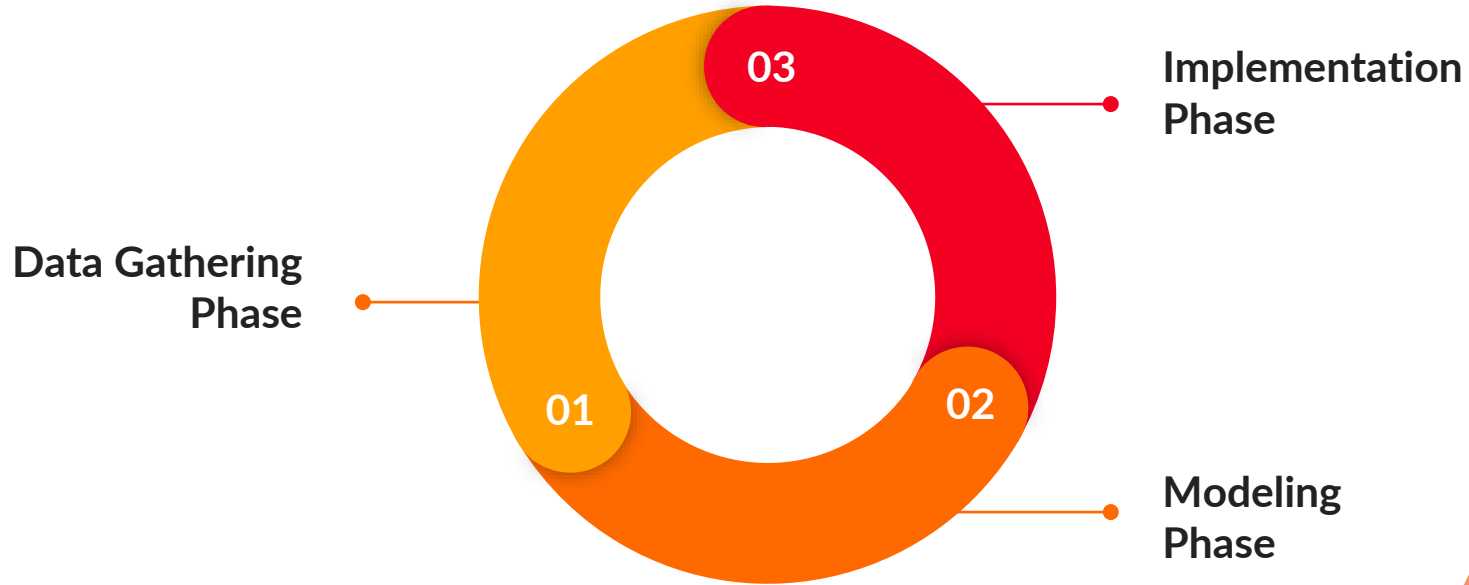
Information Integration

Project Presentation

Valentina Sisti

1952657

Main Phases



Phase 1.

Data Gathering



Domain of Interest

🎵 Music & Spotify

- Songs
- Artists
- Albums



- Genres
- Lyrics
- Charts

Data Collection Phase



8+M. Spotify Tracks
Dataset



tracks.csv



artists.csv



albums.csv



audio_features.csv



r_albums_artists.csv



r_albums_tracks.csv



r_artist_genre.csv



r_tracks_artist.csv

Data Collection Phase



billboard-hot-100-tracks.csv

Billboard Hot 100 -
Full History Dataset



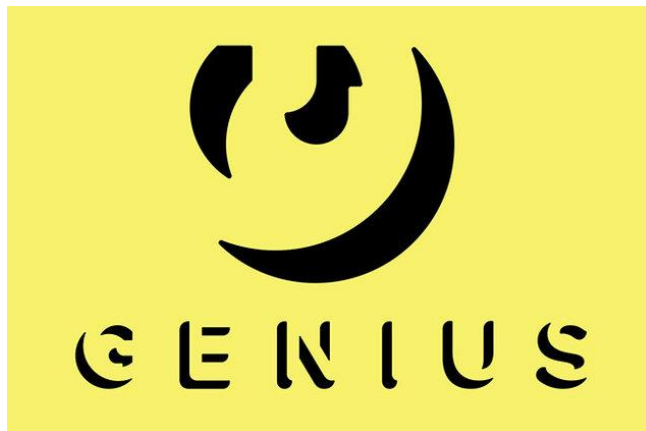
Data Collection Phase



5 Millions Genius
Lyrics Dataset



genius_lyrics.csv



Data Collection Phase



influence_data.csv

The Influence of
Music, Spotify and
AllMusic Data



Data Collection Phase



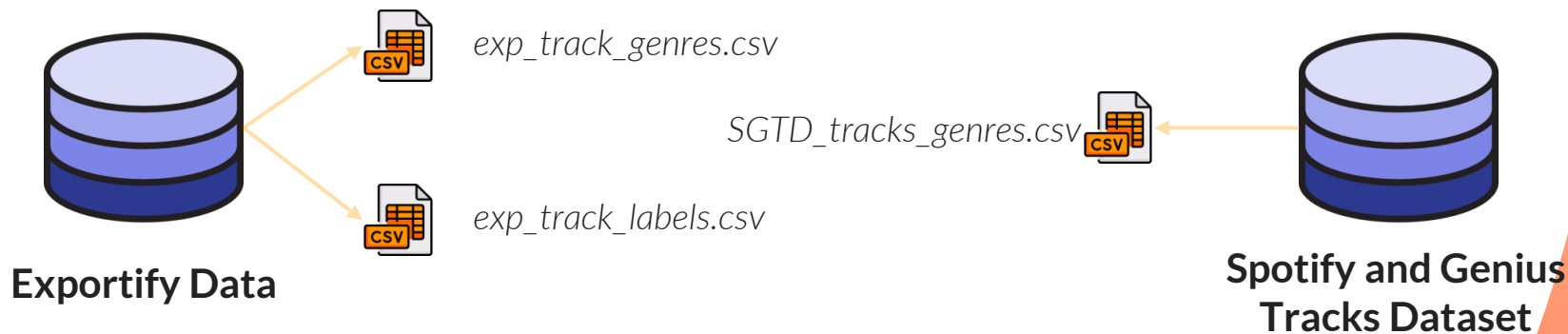
record_labels.csv

Record Labels - All
Universal Music
Group Artists



UNIVERSAL MUSIC GROUP

Data Collection Phase





7 Datasets



15 CSV Files

Tasks definition

- Find out which are the most popular **artists**, **songs** or **albums**
- Find out songs of which **genre** are most likely to reach the Billboard Hot 100, based on the history
- Find out if there are some **labels** that have very popular artists and in turn very popular songs
- Find out which is the **influence** that very important artists have had on other artists

Problems of the Data

- **Conceptual heterogeneity**
 - Concept of «id» may refer to Spotify ID, Billboard ID or custom ID internally defined
- ***track_id*** and ***audio_feature_id*** are the same in *tracks.csv*
- Songs' ***duration*** stored in milliseconds
- Albums' ***release_date*** stored in epoch format

Problems of the Data (2)

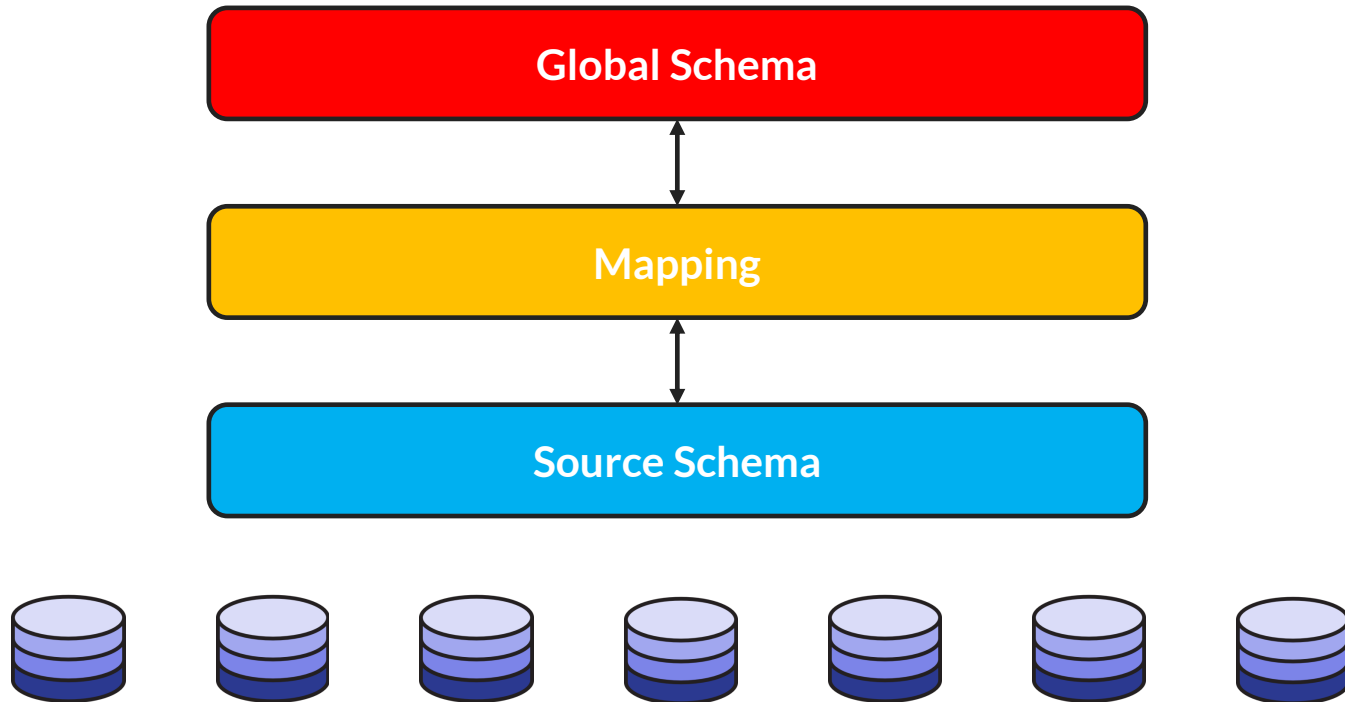
- Some data sources contain a **huge volume** of data
- A great part of it refers to quite **unknown artists** and songs
- In order to gain interesting insights, we want to restrict our discussion to just **most relevant** artists and songs

Phase 2.

Modeling

The background features a series of overlapping, wavy shapes in shades of orange and red, creating a dynamic, abstract design that flows from the bottom left towards the top right.

Information Integration System



Source Schema

Artists8M(name, id, popularity, followers)

Songs8M(id, disc_number, duration, explicit, audio_feature_id, name, preview_url, track_number, popularity, is_playable)

Albums8M(id, name, album_group, album_type, release_date, popularity)

TracksArtists8M(track_id, artist_id)

AlbumsTracks8M(album_id, track_id)

AlbumsArtists8M(album_id, artist_id)

ArtistsGenres8M(artist_id, genre)

Source Schema (2)

GeniusLyrics(title, tag, artist, year, views, features, lyrics, id)

BillboardFullHistory(billboard_id, rank, date, artist, title, last_week, weeks_on_chart, peak_rank)

Influence(influencer_name, influencer_main_genre, influencer_active_start, follower_name, follower_main_genre, follower_active_start)

AudioFeatures8M(id, acousticness, danceability, duration, energy, instrumentalness, key, liveness, loudness, mode, speechiness, tempo, time_signature)

Source Schema (3)

SGTDTracksGenres(*artist_popularity*, *followers*, *genre*, *id*, *name*,
track_id, *track_name_prev*, *type*)

ExportifyTracksGenres(*track_URI*, *track_name*, *artist_id*, *artist_name*, *album_URI*,
album_name, *track_id*, *genre*)

ExportifyTracksLabels(*track_URI*, *track_name*, *artist_id*, *artist_name*, *track_id*, *genre*)

RecordLabels(*id*, *group*, *label*, *artist_name*)

Global Schema

*Song*_{/7}(*id*, *name*, *popularity*, *track_number*, *duration*, *explicit*, *preview_url*)

*Artist*_{/4}(*id*, *name*, *popularity*, *followers*)

*Album*_{/6}(*id*, *name*, *album_type*, *popularity*, *release_date*)

*BillboardHot100Song*_{/5}(*song_id*, *rank*, *date*, *weeks_on_chart*, *peak_rank*)

*AudioFeature*_{/12}(*song_id*, *acousticness*, *danceability*, *energy*,
instrumentalness, *key*, *liveness*, *loudness*, *mode*, *speechiness*,
tempo, *time_signature*)

Global Schema (2)

*isSongOf*_{/2}(*song_id*, *artist_id*)

Song with «*song_id*» is performed by Artist with «*artist_id*»

*isSongInAlbum*_{/2}(*song_id*, *album_id*)

Song with «*song_id*» is in Album with «*album_id*»

*isAlbumOf*_{/2}(*album_id*, *artist_id*)

Album with «*album_id*» is of Artist with «*artist_id*»

Global Schema (3)

*Genre*_{/1}(*genre_name*)

*doesGenre*_{/2}(*artist_id*, *genre_name*)

Artist with «*artist_id*» does the music Genre «*genre_name*»

*hasInfluenced*_{/2}(*artist_id1*, *artist_id2*)

Artist with «*artist_id1*» has influenced Artist with «*artist_id2*»

*hasLyrics*_{/2}(*song_id*, *lyrics*)

Song with «*song_id*» has Lyrics «*lyrics*»

Global Schema (4)

*Label*_{/1}(*label_name*)

*LabelGroup*_{/1}(*group_name*)

*hasLabel*_{/2}(*artist_id*, *label*)

Artist «*artist_id*» has published a Song under the Label «*label*»

*isLabelInGroup*_{/2}(*label*, *group*)

Label «*label*» belongs to the Label Group «*group*»

Conjunctive GAV Mapping

Artist:

$\forall id, n, p, f. (\text{Artist8M}(p, f, id, n) \wedge (p \geq 50)) \rightarrow \text{Artist}(id, n, p, f)$

Song:

$\forall id, n, d, e, pu, p, tn. (\exists dn, afi, ap, af, aid, an. \text{Song8M}(\mathbf{id}, dn, d, e, afi, n, pu, p, tn) \wedge \text{TrackArtist8M}(\mathbf{id}, \mathbf{aid}) \wedge \text{Artist8M}(ap, af, aid, an) \wedge (ap \geq 50)) \rightarrow \text{Song}(id, n, p, tn, d, e, pu)$

Album:

$\forall id, n, at, p, rd. (\exists ap, af, aid, an. \text{Album8M}(\mathbf{id}, n, ag, at, rd, p) \wedge \text{AlbumArtist8M}(\mathbf{id}, \mathbf{aid}) \wedge \text{Artist8M}(ap, af, aid, an) \wedge (ap \geq 50)) \rightarrow \text{Album}(id, n, at, p, rd)$

Conjunctive GAV Mapping (2)

isSongOf:

$$\forall sid, aid. (\exists p, f, n. \text{TrackArtist8M}(sid, \text{aid}) \wedge \text{Artist8M}(p, f, \text{aid}, n) \wedge (p \geq 50)) \rightarrow \text{isSongOf}(sid, aid)$$

isSongInAlbum:

$$\forall sid, albid. (\exists p, f, aid, n. \text{AlbumTrack8M}(\text{sid}, albid) \wedge \text{TrackArtist8M}(\text{sid}, \text{aid}) \wedge \text{Artist8M}(p, f, \text{aid}, n) \wedge (p \geq 50)) \rightarrow \text{isSongInAlbum}(sid, albid)$$

isAlbumOf:

$$\forall albid, aid. (\exists p, f, n. \text{AlbumArtist8M}(albid, \text{aid}) \wedge \text{Artist8M}(p, f, \text{aid}, n) \wedge (p \geq 50)) \rightarrow \text{isAlbumOf}(albid, aid)$$

Conjunctive GAV Mapping (3)

AudioFeature:

$\forall sid, a, d, e, i, k, li, lo, m, s, t, ts. (\exists ap, af, aid, an, du.$
 $AudioFeature8M(sid, a, d, du, e, i, k, li, lo, m, s, t, ts) \wedge TrackArtist8M(sid, aid) \wedge$
 $Artist8M(ap, af, aid, an) \wedge (ap \geq 50)) \rightarrow AudioFeature(sid, a, d, e, i, k, li, lo, m, s, t, ts)$

Genre:

$\forall genre. (\exists tu, tn, aid, an, alu, aln, tnb.$
 $ExportifyTracksGenres(tu, tn, aid, an, alu, aln, tnb, genre) \rightarrow Genre(genre)$

$\forall genre. (\exists p, f, id, n, tid, tnp, t. SGTDTTracksGenres(p, f, genre, id, n, tid, tnp, t)) \rightarrow Genre(genre)$

$\forall genre. (\exists artist_id. ArtistGenre8M(artist_id, genre)) \rightarrow Genre(genre)$

Conjunctive GAV Mapping (4)

doesGenre:

$\forall \text{artist_id, genre. } (\exists \text{ap, af, artist_name, tu, tn, alu, aln, tnb.}$
 $\text{ExportifyTracksGenres}(\text{tu, tn, artist_id, artist_name, alu, aln, tnb, genre}) \wedge$
 $\text{Artist8M}(\text{ap, af, artist_id, artist_name}) \wedge (\text{ap} \geq 50)) \rightarrow \text{doesGenre}(\text{artist_id, genre})$

$\forall \text{artist_id, genre. } (\exists \text{p, f, id, ap, af, track_id, dn, d, e, afi, n, pu, tp, tn.}$
 $\text{SGTDTracksGenres}(\text{p, f, genre, id, n, track_id, tnp, t}) \wedge \text{Song8M}(\text{track_id, dn, d, e, afi, n, pu, tp, tn}) \wedge$
 $\text{TrackArtist8M}(\text{sid, aid}) \wedge \text{Artist8M}(\text{ap, af, artist_id, artist_name}) \wedge (\text{ap} \geq 50)) \rightarrow \text{doesGenre}(\text{artist_id, genre})$

$\forall \text{artist_id, genre. } (\exists \text{ap, af, artist_name. } \text{ArtistGenre8M}(\text{artist_id, genre}) \wedge$
 $\text{Artist8M}(\text{ap, af, artist_id, artist_name}) \wedge (\text{ap} \geq 50)) \rightarrow \text{doesGenre}(\text{artist_id, genre})$

Conjunctive GAV Mapping (5)

hasLyrics:

$\forall sid, lyrics. (\exists ap, af, aid, an, title, t, y, v, ft, i, dn, d, e, afi, pu, sp, tn.$
 $Lyrics5M(title, t, an, y, v, ft, lyrics, i) \wedge Song8M(sid, dn, d, e, af, title, pu, sp, tn) \wedge$
 $TrackArtist8M(sid, aid) \wedge Artist8M(ap, af, aid, an) \wedge (ap \geq 50)) \rightarrow hasLyrics(sid, lyrics)$

BillboardHot100Song:

$\forall sid, r, d, woc, pr. (\exists bid, a_name, title, lw, ap, af, aid, dn, du, e, afi, pu, sp, tn.$
 $BillboardFullHistory(bid, r, d, a_name, title, lw, woc, pr) \wedge Song8M(sid, dn, du, e, afi, title, pu, sp, tn) \wedge$
 $TrackArtist8M(sid, aid) \wedge Artist8M(ap, af, aid, an) \wedge (p \geq 50))$
 $\rightarrow BillboardHot100Song(sid, r, d, woc, pr)$

Conjunctive GAV Mapping (6)

hasInfluenced:

$\forall aid1, aid2. (\exists infl_name, img, ias, foll_name, fmg, fas, ap1, ap2, af1, af2.$
 $Influence(infl_name, img, ias, foll_name, fmg, fas) \wedge Artist8M(ap1, af1, aid1, infl_name) \wedge (ap1 \geq 50) \wedge$
 $Artist8M(ap2, af2, aid2, foll_name) \wedge (ap2 \geq 50)) \rightarrow hasInfluenced(aid1, aid2)$

Label:

$\forall label. (\exists tu, tn, aid, an, tnb. ExportifyTracksLabels(tu, tn, aid, an, tnb, label)) \rightarrow Label(label)$

$\forall label. (\exists id, group, artist_name. RecordLabels(id, group, label, artist_name)) \rightarrow Label(label)$

Group:

$\forall group. (\exists id, label, artistname. RecordLabels(id, group, label, artist_name)) \rightarrow Group(group)$

Conjunctive GAV Mapping (7)

hasLabel:

$\forall \text{label}, \text{artist_id}. (\exists \text{ap}, \text{af}, \text{aid}, \text{an}, \text{tu}, \text{tn}, \text{tnb}.$
 $\text{ExportifyTracksLabels}(\text{tu}, \text{tn}, \text{artist_id}, \text{an}, \text{tnb}, \text{label}) \wedge$
 $\text{Artist8M}(\text{ap}, \text{af}, \text{artist_id}, \text{an}) \wedge (\text{ap} \geq 50)) \rightarrow \text{hasLabel}(\text{artist_id}, \text{label})$

$\forall \text{label}, \text{artist_id}. (\exists \text{ap}, \text{af}, \text{aid}, \text{artist_name}, \text{id}, \text{group}.$
 $\text{RecordLabels}(\text{id}, \text{group}, \text{label}, \text{artist_name}) \wedge \text{Artist8M}(\text{ap}, \text{af}, \text{artist_id}, \text{artist_name}) \wedge$
 $(\text{ap} \geq 50)) \rightarrow \text{hasLabel}(\text{artist_id}, \text{label})$

isLabelInGroup:

$\forall \text{label}, \text{group}. (\exists \text{ap}, \text{af}, \text{aid}, \text{artist_name}, \text{id}, \text{group}.$
 $\text{RecordLabels}(\text{id}, \text{group}, \text{label}, \text{artist_name}) \wedge \text{Artist8M}(\text{ap}, \text{af}, \text{aid}, \text{artist_name}) \wedge$
 $(\text{ap} \geq 50)) \rightarrow \text{isLabelInGroup}(\text{label}, \text{group})$

The Query Answering Problem

- It can be proved that, given the Retrieved Global Database, in the case of Conjunctive GAV Mapping, we can compute *certain answers* for CQ and UCQ.
- Now, in order to compute and obtain the RGDB, we can exploit the Materialization-based approach in the GAV context, simply by materializing the global schema views.
- In order to do this in practice, we can exploit a **data integration tool** and this leads us to the final phase of the project: the Implementation Phase.

Phase 3.

Implementation

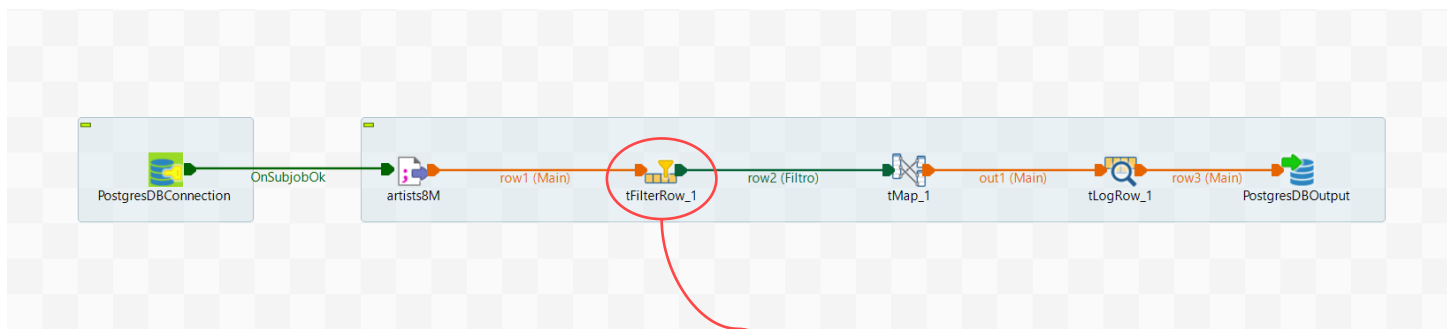
The background features a series of overlapping, wavy shapes in shades of orange and red, creating a dynamic, flowing effect that transitions from the bottom left towards the top right.

Data Integration Tool: **Talend**

- Talend is the *leader tool* in cloud and big data integration
- Fully *open source*
- Embeds *existing* Java code libraries and allows to create *custom* components
- Allows to *easily* integrate data from different sources thanks to a GUI



Artists



tFilterRow_1

Schema: Integrato Edit schema Sync columns

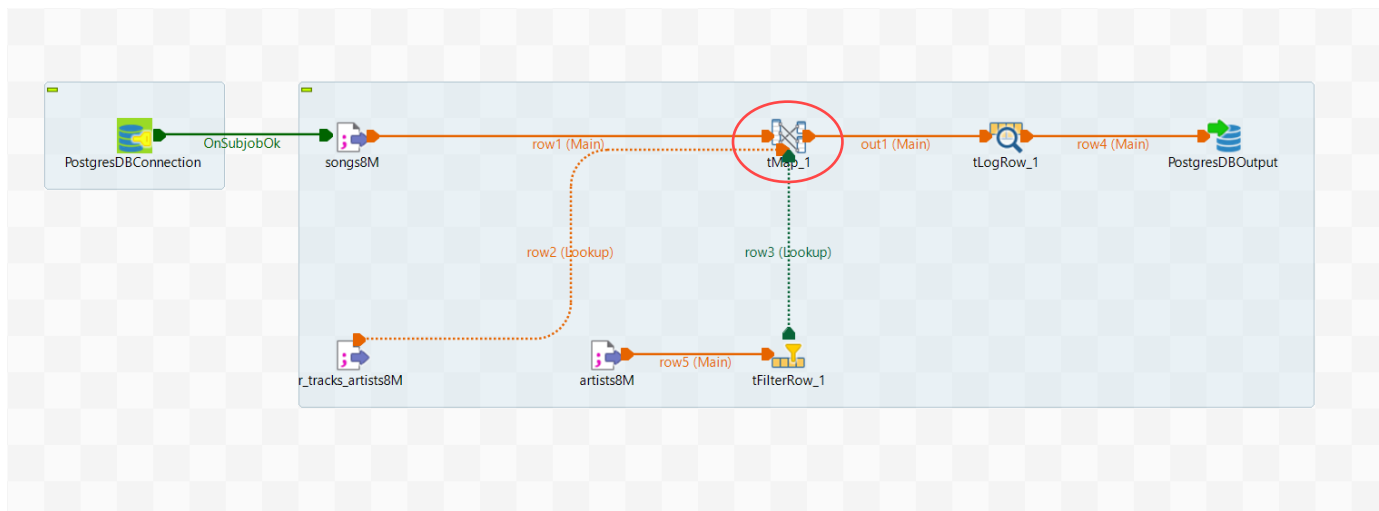
Operazione logica usata per combinare condizioni: e

Condizioni

Colonna input	Funzione	Operatore	Valore
popularity	Vuoto	>=	50

+ × ↑ ↓ [] []

Songs



Songs - tMap

```
public class MillisecondsConversion {  
  
    public static String getDateFromMillis(long millis) throws ParseException{  
  
        String timeInHHMMSS = "";  
  
        if(millis != 0) {  
            Duration duration = Duration.ofMillis(millis);  
            long HH = duration.toHours();  
            long MM = duration.toMinutesPart();  
            long SS = duration.toSecondsPart();  
            timeInHHMMSS = String.format("%02d:%02d:%02d", HH, MM, SS);  
        }  
        return timeInHHMMSS;  
    }  
}
```

row1

Column
id
disc_number
duration
explicit
audio_feature_id
name
preview_url
track_number
popularity
is_playable

row2

Property	Value
Lookup Model	Carica una volta
Match Model	Corrispondenza unica
Join Model	Inner Join
Store temp data	false

Expr. key

Column
track_id
artist_id

row3

Property	Value
Lookup Model	Carica una volta
Match Model	Corrispondenza unica
Join Model	Inner Join
Store temp data	false

Expr. key

Column
name
id
popularity
followers

Var

Espressione	Tipo	Variable
-------------	------	----------

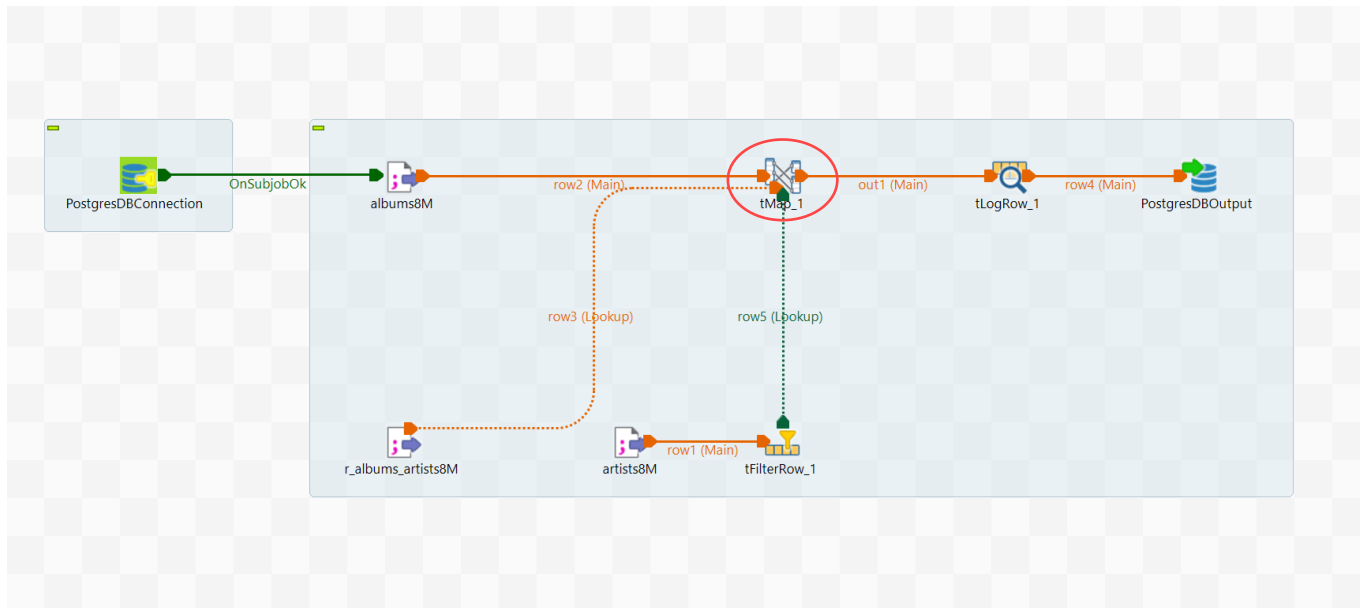
out1

Espressione
row1.id
row1.name
row1.popularity
row1.track_number
<u>SqlDate.parseDate("HH:mm:ss", routines.MillisecondsConversion.getDateFromMillis(row1.duration))</u>
row1.explicit
row1.preview_url

Column

Column
id
name
popularity
track_number
duration
explicit
preview_url

Albums



Albums - tMap

```
public class EpochToDate {  
    public static Date getDateFromEpoch(Long timeAsLong, String javaDatePattern) throws ParseException{  
        String returnVal = "";  
        Date returnDateVal = null;  
  
        if(timeAsLong!=null){  
            SimpleDateFormat sdf = new SimpleDateFormat(javaDatePattern);  
            returnVal = sdf.format(new Date(timeAsLong));  
            returnDateVal = sdf.parse(returnVal);  
        }  
        return returnDateVal;  
    }  
}
```

The screenshot displays a data integration tool interface with three main panels. The left panel shows a data flow diagram with three rows (row2, row3, row5) and a central 'Var' panel. The right panel shows the configuration for the 'out1' output, which is a table with columns: id, name, album_type, popularity, and release_date. The expression for the 'release_date' column is set to 'routines.EpochToDate.getDateFromEpoch(row2.release_date, "dd/MM/yyyy")', which is circled in red. A red arrow points from this expression to the 'EpochToDate' class code shown above.

row2

Column
id
name
album_group
album_type
release_date
popularity

row3

Property	Value
Lookup Model	Carica una volta
Match Model	Corrispondenza unica
Join Model	Inner Join
Store temp data	false
Expi. key	
row2.id	Column album_id artist_id

row5

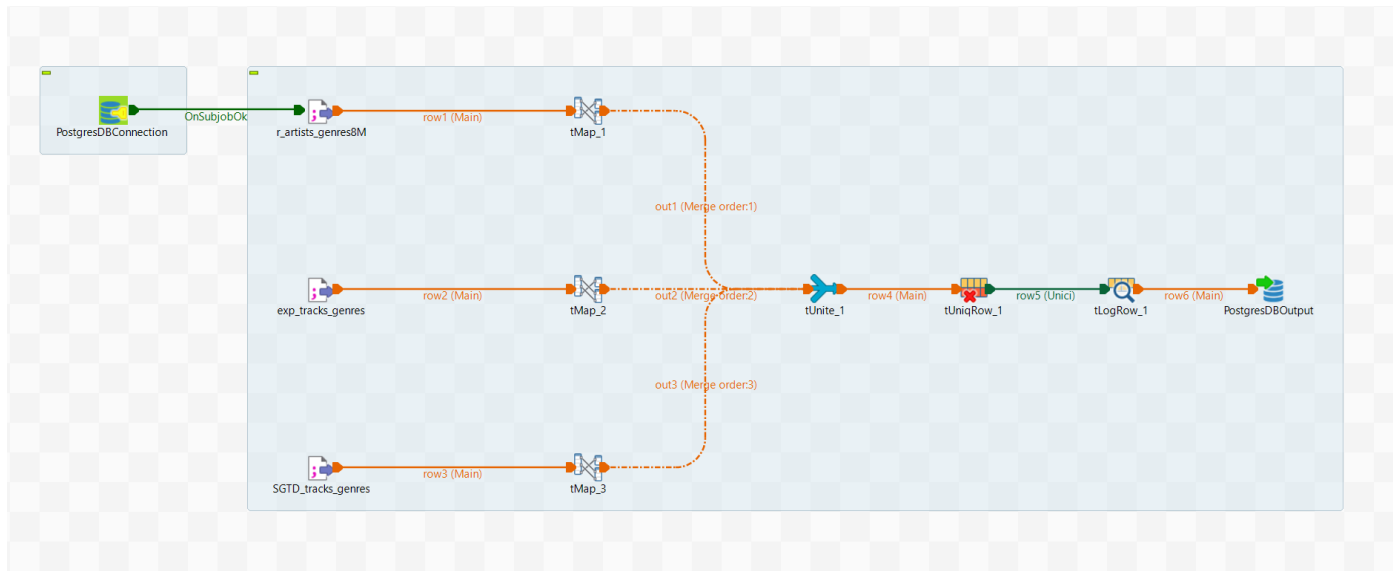
Property	Value
Lookup Model	Carica una volta
Match Model	Corrispondenza unica
Join Model	Inner Join
Store temp data	false
Expi. key	
row3.artist_id	Column name id popularity followers

Var

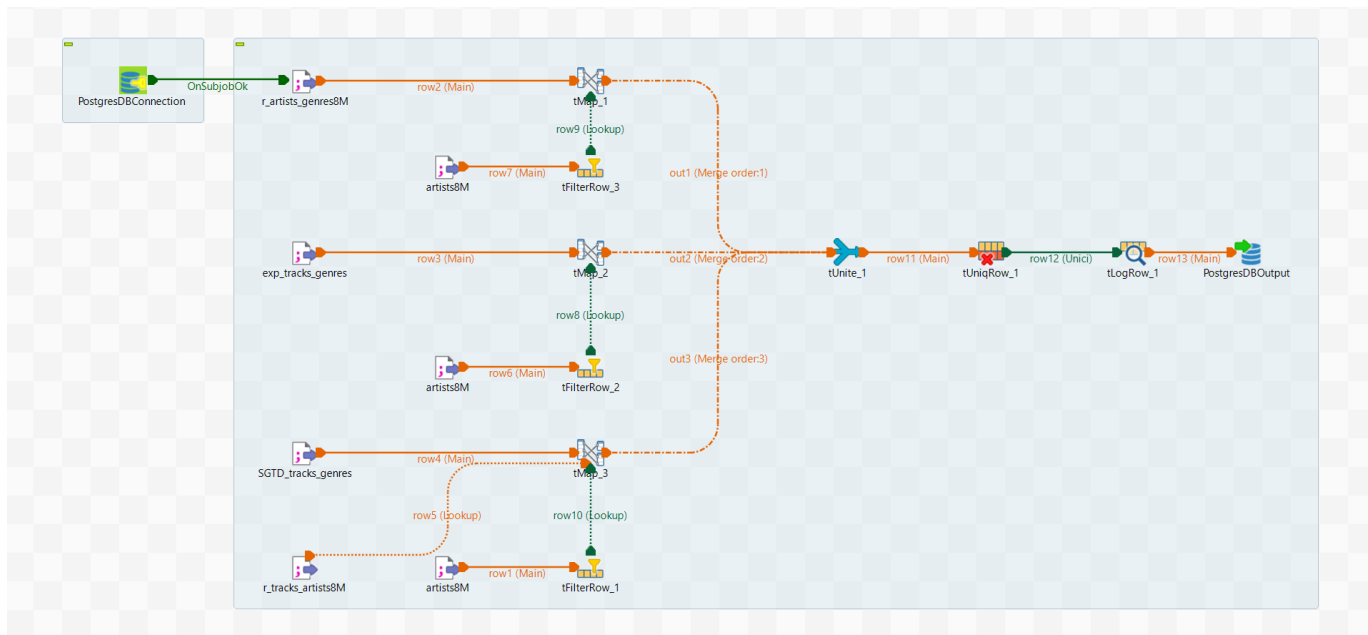
out1

Espressione	Column
row2.id	id
row2.name	name
row2.album_type	album_type
row2.popularity	popularity
routines.EpochToDate.getDateFromEpoch(row2.release_date, "dd/MM/yyyy")	release_date

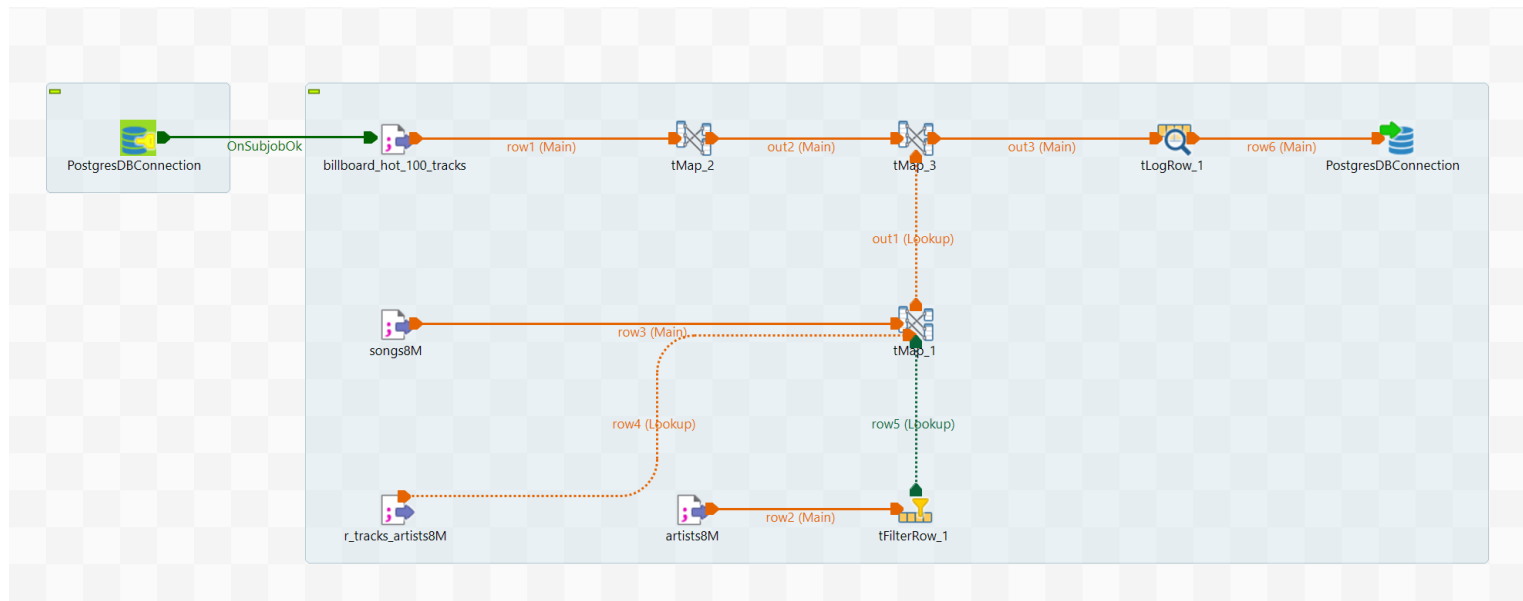
Genres



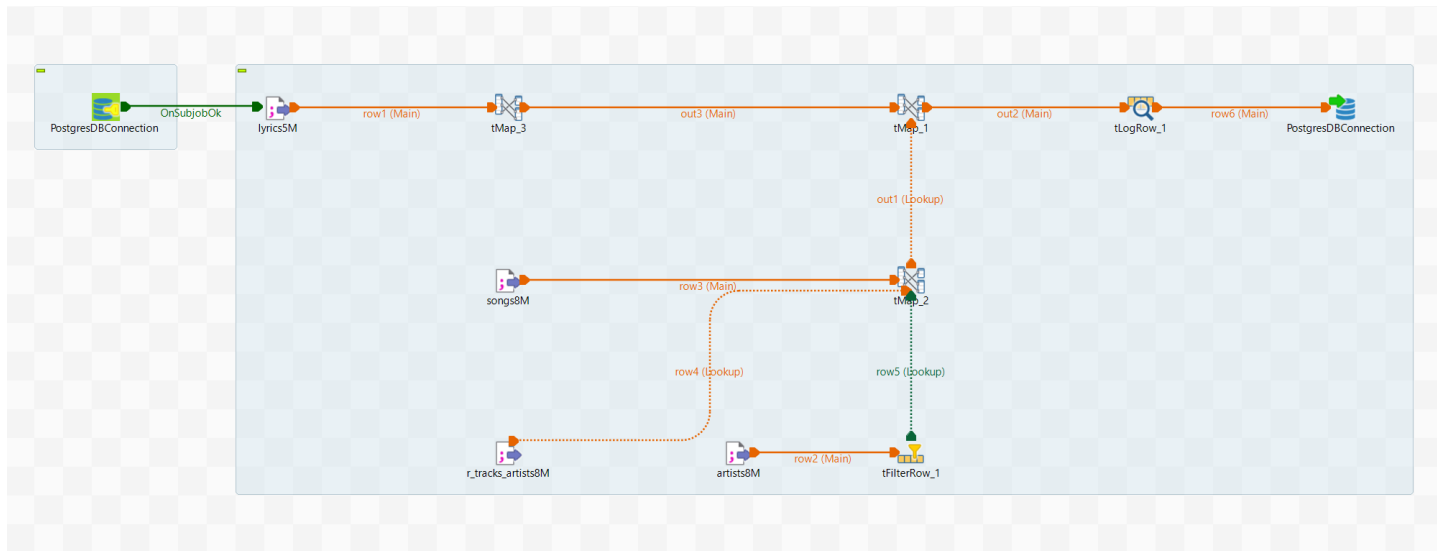
doesGenre



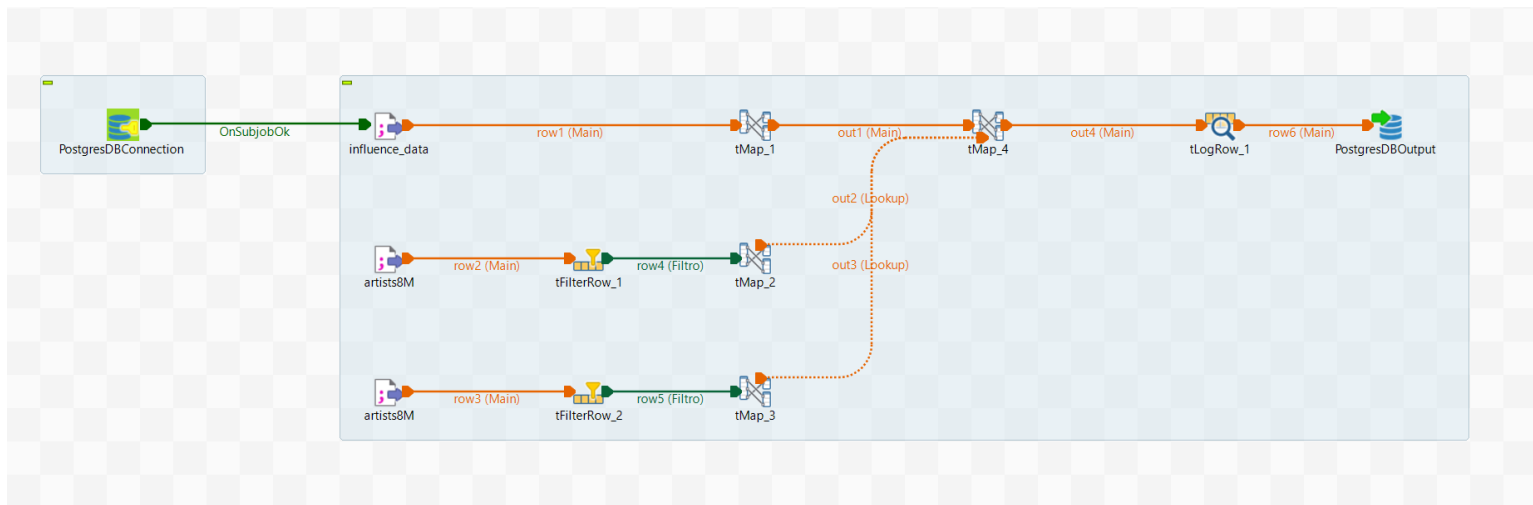
BillboardHot100Song



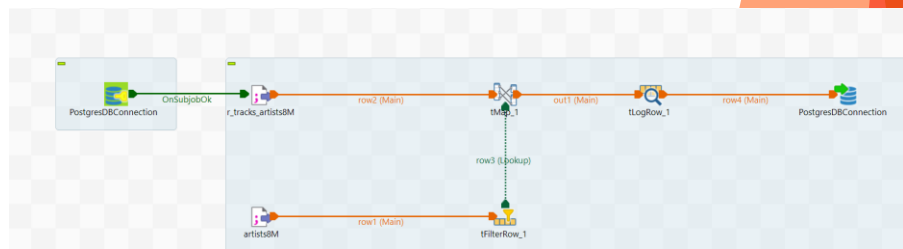
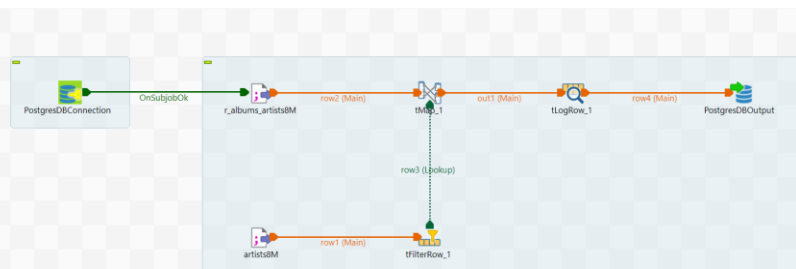
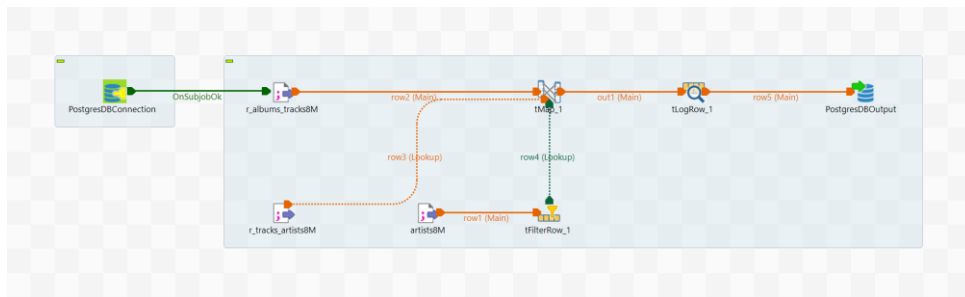
hasLyrics



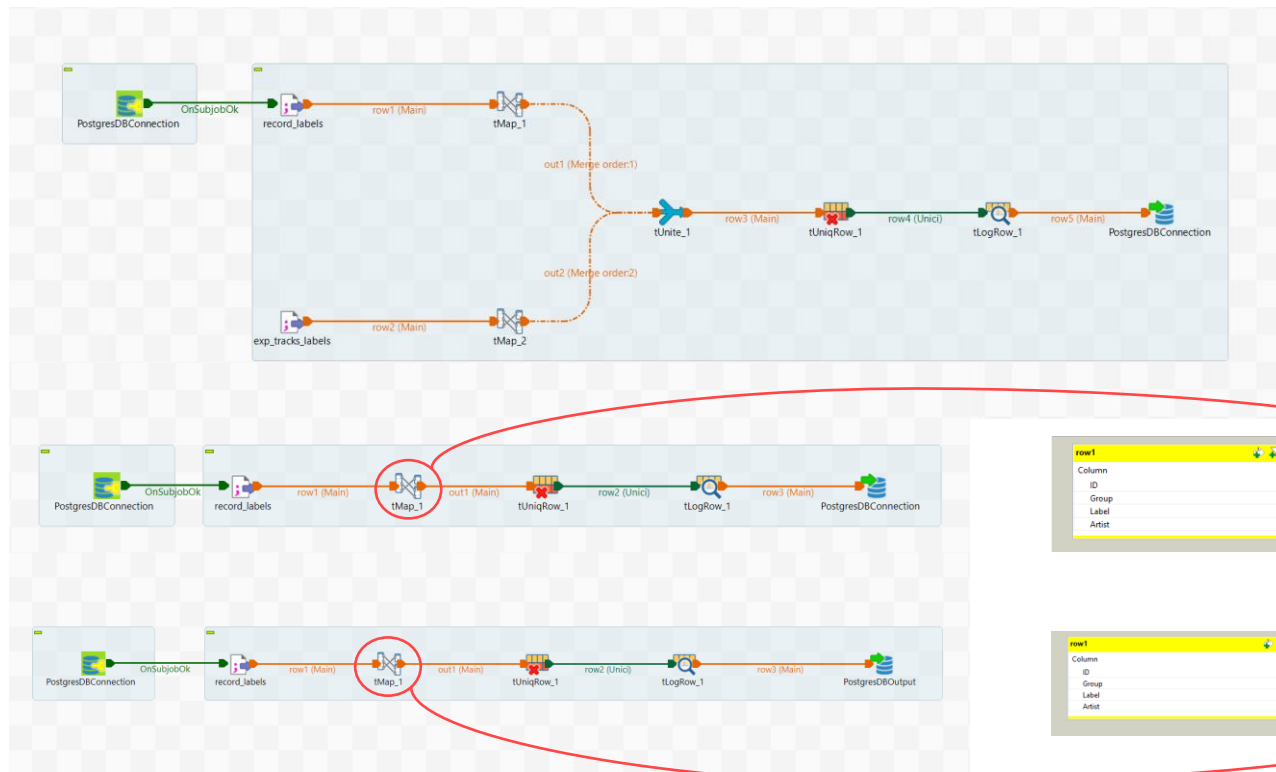
hasInfluenced



isSongOf – isAlbumOf – isSongInAlbum



Label - Group - isLabelInGroup



row1	Var	out1
Column		Espressione
ID		row1.Group
Group		Column
Label		group_name
Artist		

row1	Var	out1
Column		Espressione
ID		row1.Label
Group		label_name
Label		group_name
Artist		

Query 1 - FOL

Find Albums of Artists with more than 500k followers that contain Songs that are present in the Billboard Hot 100 Chart at least 10 years after their release

$$q = \{(\mathbf{album_name}, \mathbf{artist_name}, \mathbf{song_name}, \mathbf{billboard_date}, \mathbf{release_date}, \mathbf{rank}, \mathbf{peak_rank}) \mid$$
$$\begin{aligned} &\exists al_id, al_type, al_pop, ar_id, ar_pop, foll, s_id, s_pop, t_num, dur, exp, prev_url, woc . \\ &Album(al_id, \mathbf{album_name}, 'album', al_pop, \mathbf{release_date}) \quad \wedge \\ &AlbumOf(al_id, ar_id) \quad \wedge \\ &Artist(ar_id, \mathbf{artist_name}, ar_pop, foll) \quad \wedge \\ &SongInAlbum(s_id, al_id) \quad \wedge \\ &Song(s_id, \mathbf{song_name}, s_pop, t_num, dur, exp, prev_url) \quad \wedge \\ &BillboardSong(s_id, \mathbf{rank}, \mathbf{billboard_date}, woc, \mathbf{peak_rank}) \quad \wedge \\ &(foll > 500000) \quad \wedge \quad ((billboard_date / release_date) \geq 10)\} \end{aligned}$$

Query 1 - SQL

Find Albums of Artists with more than 500k followers that contain Songs that are present in the Billboard Hot 100 Chart at least 10 years after their release

```
SELECT DISTINCT al.name as album_name, ar.name as artist_name,
                s.name as song_name, al.release_date,
                bhs.date as billboard_date, bhs.peak_rank, bhs.rank
FROM albums as al, artists as ar, is_album_of as iao,
     is_song_in_album as isia,
     songs as s, billboard_hot100_songs as bhs
WHERE al.id = iao.album_id AND
      ar.id = iao.artist_id AND
      al.id = isia.album_id AND
      s.id = isia.track_id AND
      s.id = bhs.track_id AND
      (DATE_PART('year', bhs.date) -
       DATE_PART('year', al.release_date)) ≥ 10 AND
      ar.followers > 500000 AND
      al.album_type = 'album'
ORDER BY al.release_date DESC
```

	album_name character varying (500)	artist_name character varying (500)	song_name character varying (1000)	release_date date	billboard_date date	peak_rank integer	rank integer
1	Cole World: The Sideline Story	J. Cole	Interlude	2011-09-27	2021-05-22	8	8
2	Cole World: The Sideline Story	J. Cole	Interlude	2011-09-27	2021-05-29	8	17
3	Cole World: The Sideline Story	J. Cole	Interlude	2011-09-27	2021-06-05	8	71
4	Merry Christmas II You	Mariah Carey	All I Want For Christmas Is You	2010-01-01	2020-01-04	1	1
5	Merry Christmas II You	Mariah Carey	All I Want For Christmas Is You	2010-01-01	2020-11-28	1	29
6	Merry Christmas II You	Mariah Carey	All I Want For Christmas Is You	2010-01-01	2020-12-05	1	14
7	Merry Christmas II You	Mariah Carey	All I Want For Christmas Is You	2010-01-01	2020-12-12	1	2
8	Merry Christmas II You	Mariah Carey	All I Want For Christmas Is You	2010-01-01	2020-12-19	1	1
9	Merry Christmas II You	Mariah Carey	All I Want For Christmas Is You	2010-01-01	2020-12-26	1	2
10	Merry Christmas II You	Mariah Carey	All I Want For Christmas Is You	2010-01-01	2021-01-02	1	1
11	Merry Christmas II You	Mariah Carey	All I Want For Christmas Is You	2010-01-01	2021-01-09	1	9
12	Merry Christmas II You	Mariah Carey	Oh Santa!	2010-01-01	2020-12-19	76	76
13	Relapse: Refill	Eminem	Forever	2009-05-15	2020-03-14	64	64
14	The Definition of X: Pick Of The Litter	DMX	Ruff Ryders' Anthem	2007-06-12	2021-04-24	16	16
15	The Definition of X: Pick Of The Litter	DMX	X Gon' Give It To Ya	2007-06-12	2021-04-24	46	46
16	Johnny B. Goode/His Complete '50s Chess Recordings	Chuck Berry	Run Rudolph Run	2007-01-01	2019-01-05	45	45
17	Johnny B. Goode/His Complete '50s Chess Recordings	Chuck Berry	Run Rudolph Run	2007-01-01	2020-01-04	36	36
18	Johnny B. Goode/His Complete '50s Chess Recordings	Chuck Berry	Run Rudolph Run	2007-01-01	2020-12-19	36	40
19	Johnny B. Goode/His Complete '50s Chess Recordings	Chuck Berry	Run Rudolph Run	2007-01-01	2020-12-26	29	29
20	Johnny B. Goode/His Complete '50s Chess Recordings	Chuck Berry	Run Rudolph Run	2007-01-01	2021-01-02	10	10
21	Johnny B. Goode/His Complete '50s Chess Recordings	Chuck Berry	Run Rudolph Run	2007-01-01	2021-01-09	10	29
22	Songs for Christmas	Nat King Cole	Deck the Halls	2004-01-01	2020-12-26	47	47
23	Songs for Christmas	Nat King Cole	Deck the Halls	2004-01-01	2021-01-02	43	43
24	Meteora	Linkin Park	Numb	2003-03-24	2017-08-12	11	34
25	Meteora (Bonus Edition)	Linkin Park	Numb	2003-03-24	2017-08-12	11	34
26	A Day Without Rain	Enya	Only Time	2000-11-11	2013-12-07	10	43
27	Hybrid Theory	Linkin Park	In the End	2000-10-24	2017-08-12	2	37
28	Hybrid Theory (Bonus Edition)	Linkin Park	In the End	2000-10-24	2017-08-12	2	37
29	It's Dark And Hell Is Hot	DMX	Ruff Ryders' Anthem	1998-05-01	2021-04-24	16	16
30	The Deram Anthology 1966 - 1968	David Bowie	Space Oddity	1997-06-09	2016-01-30	15	42
31	HiStory - PAST, PRESENT AND FUTURE - BOOK I	Michael Jackson	Billie Jean	1995-06-16	2014-06-07	1	14
32	HiStory - PAST, PRESENT AND FUTURE - BOOK I	Michael Jackson	Thriller	1995-06-16	2013-11-16	4	42
33	HiStory - PAST, PRESENT AND FUTURE - BOOK I	Michael Jackson	Thriller	1995-06-16	2014-11-15	4	35
34	HiStory - PAST, PRESENT AND FUTURE - BOOK I	Michael Jackson	Thriller	1995-06-16	2015-11-21	4	45
35	HiStory - PAST, PRESENT AND FUTURE - BOOK I	Michael Jackson	Thriller	1995-06-16	2018-11-10	4	31
36	HiStory - PAST, PRESENT AND FUTURE - BOOK I	Michael Jackson	Thriller	1995-06-16	2019-11-09	4	44
37	HiStory - PAST, PRESENT AND FUTURE - BOOK I	Michael Jackson	Thriller	1995-06-16	2020-11-14	4	48
38	Merry Christmas	Mariah Carey	All I Want for Christmas Is You	1994-11-01	2012-12-22	29	29
39	Merry Christmas	Mariah Carey	All I Want for Christmas Is You	1994-11-01	2012-12-29	25	25

Query 2 - FOL

Find acousticness, danceability, energy and loudness of Songs of Artists that do the 'rock' Genre and that reached the #1 of the Billboard Hot 100 Chart

$$q = \{(song_name, artist_name, acousticness, danceability, energy, loudness, billboard_date) \mid$$
$$\begin{aligned} &\exists ar_id, ar_pop, foll, genre, s_id, s_pop, t_num, dur, exp, prev_url, woc, peak_rank, i, k, l, m, s, t, ts. \\ &Artist(ar_id, artist_name, ar_pop, foll) \quad \wedge \\ &doesGenre(ar_id, 'rock') \quad \wedge \\ &isSongOf(s_id, ar_id) \quad \wedge \\ &Song(s_id, song_name, s_pop, t_num, dur, exp, prev_url) \quad \wedge \\ &BillboardSong(s_id, 1, billboard_date, woc, peak_rank) \quad \wedge \\ &AudioFeature(s_id, acousticness, danceability, energy, i, k, l, loudness, m, s, t, ts)\} \end{aligned}$$

Query 2 - SQL

Find acousticness, danceability, energy and loudness of Songs of Artists that do the 'rock' Genre and that reached the #1 of the Billboard Hot 100 Chart

```
SELECT DISTINCT s.name as song_name, ar.name as artist_name,
                af.acousticness, af.danceability, af.energy,
                af.loudness, bhs.rank, bhs.date as week_of_date
FROM artists as ar, songs as s, is_song_of as iso,
     billboard_hot100_songs as bhs, does_genre as dg,
     audio_features as af
WHERE ar.id = iso.artist_id AND
      s.id = iso.track_id AND
      s.id = bhs.track_id AND
      dg.artist_id = ar.id AND
      af.track_id = s.id AND
      dg.genre = 'rock' AND
      bhs.rank = 1
ORDER BY week_of_date DESC
```

	song_name character varying (1000)	artist_name character varying (500)	acousticness real	danceability real	energy real	loudness double precision	rank integer	date date
1	Havana	Young Thug	0.186	0.768	0.517	-4.32299995422363	1	2018-01-27
2	With Arms Wide Open	Creed	0.0037	0.414	0.533	-8.44499969482422	1	2000-11-11
3	With Arms Wide Open	Creed	0.00425	0.41	0.539	-8.41199970245361	1	2000-11-11
4	Smooth	Santana	0.189	0.521	0.815	-15.0869998931885	1	2000-01-08
5	Smooth	Santana	0.189	0.521	0.815	-15.0869998931885	1	2000-01-01
6	Smooth	Santana	0.189	0.521	0.815	-15.0869998931885	1	1999-12-25
7	Smooth	Santana	0.189	0.521	0.815	-15.0869998931885	1	1999-12-18
8	Smooth	Santana	0.189	0.521	0.815	-15.0869998931885	1	1999-12-11
9	Smooth	Santana	0.189	0.521	0.815	-15.0869998931885	1	1999-12-04
10	Smooth	Santana	0.189	0.521	0.815	-15.0869998931885	1	1999-11-27
11	Smooth	Santana	0.189	0.521	0.815	-15.0869998931885	1	1999-11-20
12	Smooth	Santana	0.189	0.521	0.815	-15.0869998931885	1	1999-11-13
13	Smooth	Santana	0.189	0.521	0.815	-15.0869998931885	1	1999-11-06
14	Smooth	Santana	0.189	0.521	0.815	-15.0869998931885	1	1999-10-30
15	Smooth	Santana	0.189	0.521	0.815	-15.0869998931885	1	1999-10-23
16	Have You Ever Really Loved A Woman?	Bryan Adams	0.535	0.566	0.477	-10.1829996109009	1	1995-07-01
17	Have You Ever Really Loved A Woman?	Bryan Adams	0.536	0.57	0.478	-10.1850004196167	1	1995-07-01
18	Have You Ever Really Loved A Woman?	Bryan Adams	0.55	0.563	0.545	-7.76200008392334	1	1995-07-01
19	Have You Ever Really Loved A Woman?	Bryan Adams	0.535	0.566	0.477	-10.1829996109009	1	1995-06-24
20	Have You Ever Really Loved A Woman?	Bryan Adams	0.536	0.57	0.478	-10.1850004196167	1	1995-06-24
21	Have You Ever Really Loved A Woman?	Bryan Adams	0.55	0.563	0.545	-7.76200008392334	1	1995-06-24
22	Have You Ever Really Loved A Woman?	Bryan Adams	0.535	0.566	0.477	-10.1829996109009	1	1995-06-17
23	Have You Ever Really Loved A Woman?	Bryan Adams	0.536	0.57	0.478	-10.1850004196167	1	1995-06-17
24	Have You Ever Really Loved A Woman?	Bryan Adams	0.55	0.563	0.545	-7.76200008392334	1	1995-06-17
25	Have You Ever Really Loved A Woman?	Bryan Adams	0.535	0.566	0.477	-10.1829996109009	1	1995-06-10
26	Have You Ever Really Loved A Woman?	Bryan Adams	0.536	0.57	0.478	-10.1850004196167	1	1995-06-10
27	Have You Ever Really Loved A Woman?	Bryan Adams	0.55	0.563	0.545	-7.76200008392334	1	1995-06-10
28	Have You Ever Really Loved A Woman?	Bryan Adams	0.535	0.566	0.477	-10.1829996109009	1	1995-06-03

Query 3 - FOL

Top 2 songs on Billboard in the same week of two different artists that belong to two different labels but of the same Group of labels.

$$q = \{(\mathbf{an1}, \mathbf{sn1}, \mathbf{r1}, \mathbf{an2}, \mathbf{sn2}, \mathbf{r2}, \mathbf{group}, \mathbf{week_of_date}) \mid$$
$$\exists sid1, woc1, pr1, sid2, woc2, pr2, arid1, arid2, sp1, tn1, d1, e1, pu1, sp2, tn2, d2, e2, pu2, ap1, f1, ap2, f2, label1, label2 .$$
$$BillboardSong(s_id1, \mathbf{r1}, \mathbf{week_of_date}, woc1, pr1) \wedge BillboardHot100Song(s_id2, \mathbf{r2}, \mathbf{week_of_date}, woc2, pr2)$$
$$isSongOf(s_id1, ar_id1) \wedge isSongOf(s_id2, ar_id2) \wedge$$
$$Song(s_id1, \mathbf{sn1}, sp1, tn1, d1, e1, pu1) \wedge Song(s_id2, \mathbf{sn2}, sp2, tn2, d2, e2, pu2) \wedge$$
$$Artist(ar_id1, \mathbf{an1}, ap1, f1) \wedge Artist(ar_id2, \mathbf{an2}, ap2, f2)$$
$$hasLabel(a_id1, label1) \wedge hasLabel(a_id1, label2) \wedge$$
$$isLabelInGroup(label1, \mathbf{group}) \wedge isLabelInGroup(label2, \mathbf{group})$$
$$(r1 = 1) \wedge (r2 = 2) \wedge \neg(ar_id1 = ar_id2) \wedge \neg(label1 = label2)\}$$

Query 3 - SQL

Top 2 songs on Billboard in the same week of two different artists that belong to two different labels but of the same Group of labels.

```
SELECT DISTINCT ar1.name as artist1_name, s1.name as song_name, bhs1.rank as s1_rank, ar2.name as artist2_name,
                s2.name as song2_name, bhs2.rank as s2_rank2, ilig1.group_name, bhs1.date as week_of_date
FROM billboard_hot100_songs as bhs1, billboard_hot100_songs as bhs2, is_song_of as iso1, is_song_of as iso2,
     songs as s1, songs as s2, artists as ar1, artists as ar2, has_label as hl1, has_label as hl2,
     is_label_in_group as ilig1, is_label_in_group as ilig2
WHERE bhs1.rank = 1 AND
      bhs2.rank = 2 AND
      bhs1.track_id = s1.id AND
      bhs2.track_id = s2.id AND
      iso1.track_id = s1.id AND
      iso2.track_id = s2.id AND
      iso1.artist_id = ar1.id AND
      iso2.artist_id = ar2.id AND
      hl1.artist_id = iso1.artist_id AND
      hl2.artist_id = iso1.artist_id AND
      ilig1.label_name = hl1.label_name AND
      ilig2.label_name = hl2.label_name AND
      ilig1.group_name = ilig2.group_name AND
      ar1.id ≠ ar2.id AND
      hl1.label_name ≠ hl2.label_name AND
      bhs1.date = bhs2.date
ORDER BY week_of_date DESC
```

	artist1_name character varying (500)	song_name character varying (1000)	s1_rank integer	artist2_name character varying (500)	song2_name character varying (1000)	s2_rank2 integer	group_name character varying (50)	date
1	Eminem	Lose Yourself	1	Kelly Rowland	Dilemma	2	Interscope Geffen A&M	2002-11-09
2	Eminem	Lose Yourself	1	Nelly	Dilemma	2	Interscope Geffen A&M	2002-11-09
3	50 Cent	In Da Club	1	Jennifer Lopez	All I Have	2	Interscope Geffen A&M	2003-03-08
4	50 Cent	In Da Club	1	LL Cool J	All I Have	2	Interscope Geffen A&M	2003-03-08
5	50 Cent	In Da Club	1	Jennifer Lopez	All I Have	2	Interscope Geffen A&M	2003-03-15
6	50 Cent	In Da Club	1	LL Cool J	All I Have	2	Interscope Geffen A&M	2003-03-15
7	50 Cent	In Da Club	1	Jennifer Lopez	All I Have	2	Interscope Geffen A&M	2003-03-22
8	50 Cent	In Da Club	1	LL Cool J	All I Have	2	Interscope Geffen A&M	2003-03-22
9	50 Cent	In Da Club	1	R. Kelly	Ignition	2	Interscope Geffen A&M	2003-03-29
10	50 Cent	In Da Club	1	R. Kelly	Ignition	2	Interscope Geffen A&M	2003-04-05
11	50 Cent	In Da Club	1	R. Kelly	Ignition	2	Interscope Geffen A&M	2003-04-12
12	50 Cent	In Da Club	1	R. Kelly	Ignition	2	Interscope Geffen A&M	2003-04-19
13	50 Cent	In Da Club	1	R. Kelly	Ignition	2	Interscope Geffen A&M	2003-04-26
14	50 Cent	In Da Club	1	Sean Paul	Get Busy	2	Interscope Geffen A&M	2003-05-03
15	Sean Paul	Get Busy	1	50 Cent	In Da Club	2	Island Records	2003-05-10
16	Sean Paul	Get Busy	1	50 Cent	21 Questions	2	Island Records	2003-05-17
17	Sean Paul	Get Busy	1	Nate Dogg	21 Questions	2	Island Records	2003-05-17
18	Sean Paul	Get Busy	1	50 Cent	21 Questions	2	Island Records	2003-05-24
19	Sean Paul	Get Busy	1	Nate Dogg	21 Questions	2	Island Records	2003-05-24
20	50 Cent	21 Questions	1	Sean Paul	Get Busy	2	Interscope Geffen A&M	2003-05-31
21	50 Cent	21 Questions	1	Sean Paul	Get Busy	2	Interscope Geffen A&M	2003-06-07
22	50 Cent	21 Questions	1	Sean Paul	Get Busy	2	Interscope Geffen A&M	2003-06-14
23	50 Cent	21 Questions	1	Sean Paul	Get Busy	2	Interscope Geffen A&M	2003-06-21

Query 4 - FOL

Find which albums contain a song that is a featuring between «Machine Gun Kelly» and another artist who belongs to his same label

$$\begin{aligned} q = \{ & (\mathbf{aname2}, \mathbf{sname}, \mathbf{albname}, \mathbf{aname3}, \mathbf{releasedate}) \mid \\ & \exists aid1, p, f, sid1, sid2, aid2, p2, f2, label1, label2, sp, tn, d, e, pu, albid, at, abp, aid3, p3, f3. \\ & Artist(aid1, 'Machine Gun Kelly', p, f) \wedge isSongOf(sid1, aid1) \wedge isSongOf(sid2, aid2) \wedge \\ & (sid1 = sid2) \wedge \neg(aid1 = aid2) \wedge Artist(aid2, \mathbf{aname2}, p2, f2) \wedge \\ & hasLabel(aid1, label1) \wedge hasLabel(aid2, label2) \wedge (label1 = label2) \wedge \\ & Song(sid1, \mathbf{sname}, sp, tn, d, e, pu) \wedge isSongInAlbum(sid1, albid) \wedge \\ & Album(albid, \mathbf{albname}, at, abp, \mathbf{releasedate}) \wedge \\ & isAlbumOf(albid, aid3) \wedge Artist(aid3, \mathbf{aname3}, p3, f3) \} \end{aligned}$$

Query 4 - SQL

Find which albums contain a song that is a featuring between «Machine Gun Kelly» and another artist who belongs to his same label

```
SELECT DISTINCT ar2.name as artist_name, s.name as song_name, al.name as album_name,
                ar3.name as album_of, al.release_date
FROM is_song_of AS iso1, is_song_of AS iso2, artists AS ar1, artists AS ar2, has_label AS hl1
    has_label AS hl2, songs as s, albums as al, is_song_in_album as isia, is_album_of as iao, artists as ar3
WHERE ar1.name = 'Machine Gun Kelly' AND
    ar1.id = iso1.artist_id AND
    iso1.artist_id ≠ iso2.artist_id AND
    iso1.track_id = iso2.track_id AND
    iso1.track_id = s.id AND
    iso1.track_id = isia.track_id AND
    isia.album_id = al.id AND
    isia.album_id = iao.album_id AND
    iao.artist_id = ar3.id AND
    ar2.id = iso2.artist_id AND
    hl1.artist_id = iso1.artist_id AND
    hl2.artist_id = iso2.artist_id AND
    hl1.label_name = hl2.label_name
```

	artist_name character varying (500)	song_name character varying (1000)	album_name character varying (500)	album_of character varying (500)	release_date
1	blackbear	e.z. (feat. Machine Gun Kelly)	cybersex	blackbear	2017-11-27
2	blackbear	End Of The Road	Lace Up (Deluxe)	Machine Gun Kelly	2012-01-01
3	blackbear	my ex's best friend (with blackbear)	Tickets To My Downfall	Machine Gun Kelly	2020-09-25
4	blackbear	my ex's best friend (with blackbear)	Tickets To My Downfall (SOLD OUT Deluxe)	Machine Gun Kelly	2020-09-29
5	Papa Roach	Sunrise Trailer Park (feat. MGK)	Crooked Teeth (Deluxe Version)	Papa Roach	2017-05-19
6	X Ambassadors	Home (with Machine Gun Kelly, X Ambassadors & Bebe Rexha)	Home (with Machine Gun Kelly, X Ambassadors & Bebe Rexha)	Bebe Rexha	2017-11-18
7	X Ambassadors	Home (with Machine Gun Kelly, X Ambassadors & Bebe Rexha)	Home (with Machine Gun Kelly, X Ambassadors & Bebe Rexha)	Machine Gun Kelly	2017-11-18
8	X Ambassadors	Home (with Machine Gun Kelly, X Ambassadors & Bebe Rexha)	Home (with Machine Gun Kelly, X Ambassadors & Bebe Rexha)	X Ambassadors	2017-11-18
9	YUNGBLUD	acting like that (feat. Machine Gun Kelly)	acting like that (feat. Machine Gun Kelly)	Machine Gun Kelly	2020-12-02
10	YUNGBLUD	acting like that (feat. Machine Gun Kelly)	acting like that (feat. Machine Gun Kelly)	YUNGBLUD	2020-12-02
11	YUNGBLUD	acting like that (feat. Machine Gun Kelly)	weird!	YUNGBLUD	2020-12-04
12	YUNGBLUD	body bag (feat. YUNGBLUD & Bert McCracken of The Used)	Tickets To My Downfall (SOLD OUT Deluxe)	Machine Gun Kelly	2020-09-29
13	YUNGBLUD	I Think I'm OKAY (with YUNGBLUD & Travis Barker)	Hotel Diablo	Machine Gun Kelly	2019-07-05

Query 5 - FOL

Artists that have been influenced by «Nirvana» and that have done a featuring between them that reached the Billboard Hot 100

$q = \{(song_name, foll1_name, foll2_name, peak_rank) \mid$
 $\exists infl_aid, ip, if, foll1_aid, foll2_aid, sid1, sid2, sp, tn, d, e, pu, fp1, fp2, ff1, ff2, r, wod, woc.$
 $Artist(infl_aid, 'Nirvana', ip, if) \wedge hasInfluenced(infl_aid, foll1_aid) \wedge$
 $hasInfluenced(infl_aid, foll2_aid) \wedge \neg(foll1_aid = foll2_aid) \wedge$
 $isSongOf(sid1, foll1_aid) \wedge isSongOf(sid2, foll2_aid) \wedge (sid1 = sid2) \wedge$
 $Song(sid1, song_name, sp, tn, d, e, pu) \wedge$
 $Artist(foll1_aid, foll1_name, fp1, ff1) \wedge$
 $Artist(foll2_aid, foll2_name, fp2, ff2) \wedge$
 $BillboardHot100Song(sid1, r, wod, woc, peak_rank)\}$

Query 5 - SQL

Artists that have been influenced by «Nirvana» and that have done a featuring between them that reached the Billboard Hot 100

```
SELECT DISTINCT s.name as song_name, ar1.name as ar1_name, ar2.name as ar2_name, bhs.peak_rank
FROM has_influenced as hi1, has_influenced as hi2, is_song_of as iso1, is_song_of as iso2, artists as ar1,
     artists as ar2, songs as s, artists as influencer_artist, billboard_hot100_songs as bhs
WHERE influencer_artist.name = 'Nirvana' AND
      hi1.influencer_id = influencer_artist.id AND
      hi2.influencer_id = influencer_artist.id AND
      hi1.follower_id ≠ hi2.follower_id AND
      iso1.artist_id = hi1.follower_id AND
      iso2.artist_id = hi2.follower_id AND
      iso1.track_id = iso2.track_id AND
      iso1.track_id = s.id AND
      hi1.follower_id = ar1.id AND
      hi2.follower_id = ar2.id AND
      iso1.track_id = bhs.track_id
```

	song_name character varying (1000) 🔒	ar1_name character varying (500) 🔒	ar2_name character varying (500) 🔒	peak_rank integer 🔒
1	Under Pressure	My Chemical Romance	The Used	41
2	Under Pressure	My Chemical Romance	The Used	54
3	Under Pressure	The Used	My Chemical Romance	41
4	Under Pressure	The Used	My Chemical Romance	54

Query 6 - FOL

Lyrics of songs with popularity ≥ 80 , of artists with popularity ≥ 80 , that do either «*alternative rock*» or «*alternative metal*» or «*punk*» music genre

$$q = \{(aname, apopularity, sname, spopularity, lyrics) \mid$$
$$\begin{aligned} &(\exists aid, af, sid, tn, d, e, pu. \\ &Artist(aid, aname, apopularity, af) \wedge (apopularity \geq 80) \wedge isSongOf(sid, aid) \wedge \\ &Song(sid, sname, spopularity, tn, d, e, pu) \wedge (spopularity \geq 80) \wedge \\ &doesGenre(aid, "alternative rock") \wedge hasLyrics(sid, lyrics)) \\ &\vee \\ &(\exists aid, af, sid, tn, d, e, pu. \\ &Artist(aid, aname, apopularity, af) \wedge (apopularity \geq 80) \wedge isSongOf(sid, aid) \wedge \\ &Song(sid, sname, spopularity, tn, d, e, pu) \wedge (spopularity \geq 80) \wedge \\ &doesGenre(aid, "alternative metal") \wedge hasLyrics(sid, lyrics)) \\ &\vee \\ &(\exists aid, af, sid, tn, d, e, pu. \\ &Artist(aid, aname, apopularity, af) \wedge (apopularity \geq 80) \wedge isSongOf(sid, aid) \wedge \\ &Song(sid, sname, spopularity, tn, d, e, pu) \wedge (spopularity \geq 80) \wedge \\ &doesGenre(aid, "punk") \wedge hasLyrics(sid, lyrics))\} \end{aligned}$$


```

(SELECT DISTINCT ar.name as artist_name, ar.popularity as artist_popularity,
                 s.name as song_name, s.popularity as song_popularity, hl.lyrics
FROM songs as s, artists as ar, is_song_of as iso, has_lyrics as hl, does_genre as dg
WHERE ar.popularity ≥ 80 AND
      iso.artist_id = ar.id AND
      iso.track_id = s.id AND
      s.popularity ≥ 80 AND
      iso.artist_id = dg.artist_id AND
      dg.genre = 'alternative rock' AND
      hl.track_id = s.id
)
UNION
(SELECT DISTINCT ar.name as artist_name, ar.popularity as artist_popularity,
                 s.name as song_name, s.popularity as song_popularity, hl.lyrics
FROM songs as s, artists as ar, is_song_of as iso, has_lyrics as hl, does_genre as dg
WHERE ar.popularity ≥ 80 AND
      iso.artist_id = ar.id AND
      iso.track_id = s.id AND
      s.popularity ≥ 80 AND
      iso.artist_id = dg.artist_id AND
      dg.genre = 'alternative metal' AND
      hl.track_id = s.id
)
UNION
(SELECT DISTINCT ar.name as artist_name, ar.popularity as artist_popularity,
                 s.name as song_name, s.popularity as song_popularity, hl.lyrics
FROM songs as s, artists as ar, is_song_of as iso, has_lyrics as hl, does_genre as dg
WHERE ar.popularity ≥ 80 AND
      iso.artist_id = ar.id AND
      iso.track_id = s.id AND
      s.popularity ≥ 80 AND
      iso.artist_id = dg.artist_id AND
      dg.genre = 'punk' AND
      hl.track_id = s.id
) ORDER BY artist_name

```

Query 6 - SQL

Lyrics of songs with popularity ≥ 80 , of artists with popularity ≥ 80 , that do either «*alternative rock*» or «*alternative metal*» or «*punk*» music genres

	artist_name character varying (500)	artist_popularity integer	song_name character varying (1000)	song_popularity integer	lyrics character varying (10000000)
1	blink-182	81	All The Small Things	82	[Verse 1: Tom DeLonge] All the small things True care, truth brings I'll
2	Foo Fighters	81	Everlong	80	[Verse 1] Hello, I've waited here for you Everlong Tonight, I throw myse
3	Green Day	83	American Idiot	81	[Verse 1] Don't wanna be an American idiot Don't want a nation under
4	Green Day	83	Basket Case	81	[Verse 1] Do you have the time To listen to me whine About nothing an
5	Linkin Park	87	Numb	83	[Verse 1: Chester Bennington] I'm tired of being what you want me to b
6	Linkin Park	87	In the End	85	[Verse 1: Mike Shinoda & Chester Bennington] (It starts with one) One
7	Nirvana	84	Smells Like Teen Spirit	83	[Verse 1] Load up on guns, bring your friends It's fun to lose and to pre
8	Radiohead	81	Creep	83	[Verse 1] When you were here before Couldn't look you in the eye You
9	Red Hot Chili Peppers	86	Otherside	80	[Chorus] How long, how long will I slide? Well, separate my side I don't
10	Red Hot Chili Peppers	86	Under the Bridge	82	[Instrumental Intro] [Verse 1] Sometimes I feel like I don't have a partn
11	Red Hot Chili Peppers	86	Californication	83	[Verse 1] Psychic spies from China try to steal your mind's elation And
12	The Killers	81	Mr. Brightside	83	[Verse] Coming out of my cage and I've been doing just fine Gotta, got
13	The Strokes	80	The Adults Are Talking	80	[Verse 1] They've been sayin' you're sophisticated They're complainin'

Thanks!