

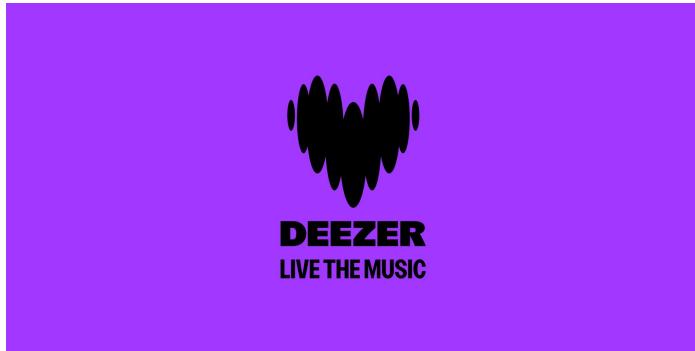
The Alliance of Algorithms and Cloud Native Technologies in the Service of the Music Experience at Deezer

CNCF Lorient - December 2024

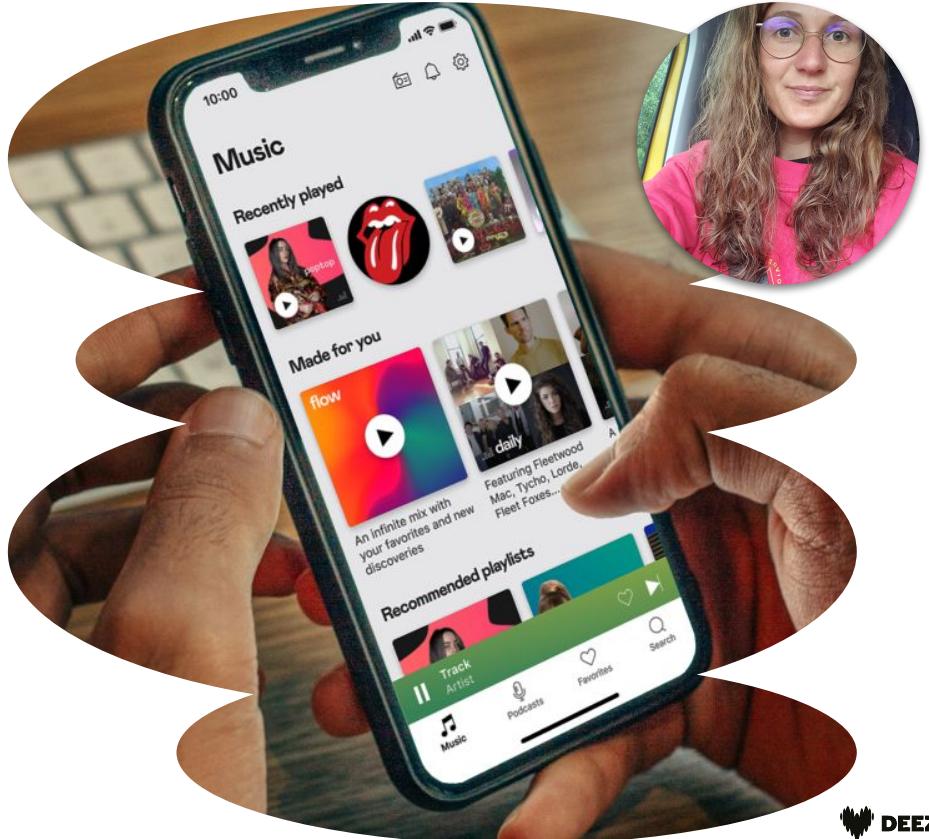
Léa Briand
Data Scientist / ML Engineer



Léa Briand - Data Scientist / ML Engineer @Deezer



- Working at Deezer for ~7 years
- Remotely from Lorient for ~4 years



AGENDA

1

Context

2

Collaborative Filtering for recommendation

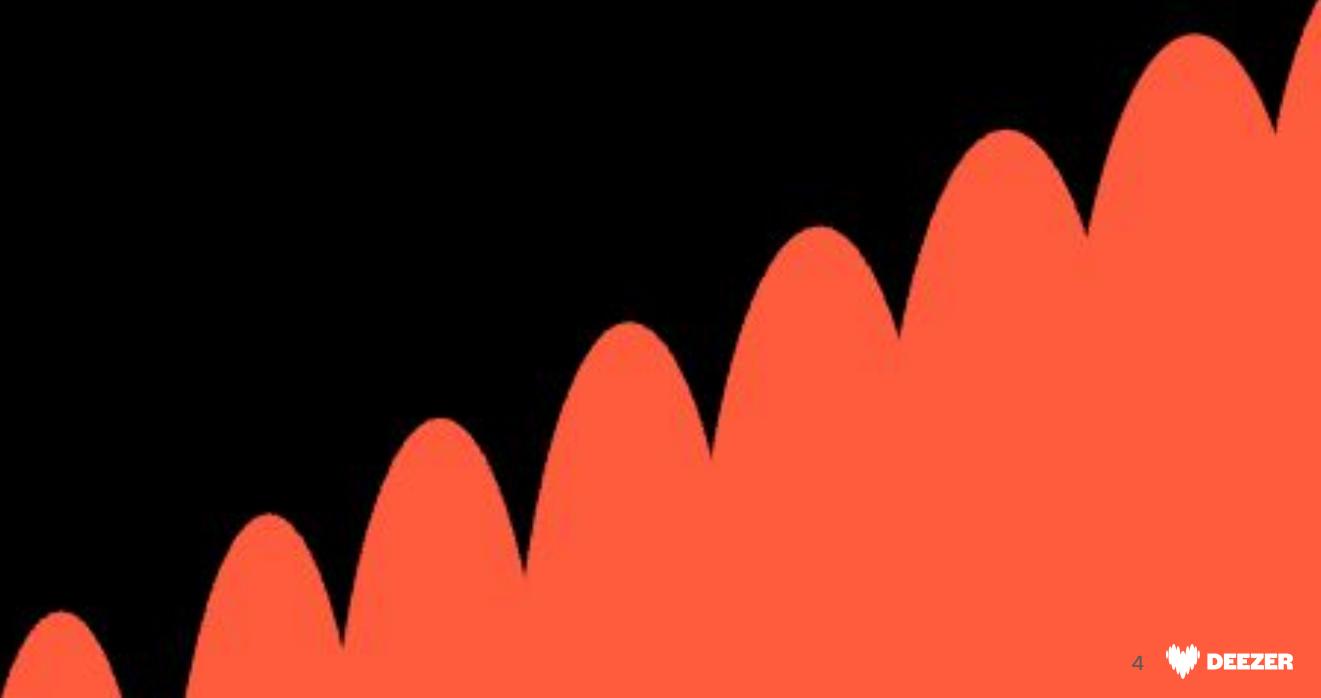
3

Cloud Native technologies & on-premise infrastructures for scaling

4

Concrete examples

1. Context



Deezer in Numbers



10M

Global monthly active users



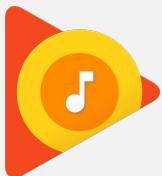
180+

Countries with **Deezer** available



600

Employees worldwide



Millions of

Tracks available on **Deezer**

Digital streaming platforms has liberated the access to the content

The collage includes:

- A Netflix interface showing the top 10 TV series in France.
- A Deezer logo.
- A list of US comedies.
- A YouTube interface showing a search for "TOUR DU STUDIO".

Millions of Tracks available on **Deezer**

Toute la musique dont tu as toujours rêvé

Profite de plus de 120 millions de morceaux, découvre les nouvelles sorties musicales, les tendances albums et des playlists créées par nos experts.

Playlists displayed:

- Décence Jul
- Chambre 140 (Part.1) PLK
- APRES MINUIT Kofs
- Pourvu qu'il pleuve Shay
- Du peu que j'ai eu, Du mieux que j'ai pu Lesram

But it has also created a new **problem :**

the paradox of choice

**Listening to music can be discouraging
when there is too much choice**

**The more options you have, the more you
take time to “respond”.**

Where to start ?



If you know what you want to listen ...

The Search



Q |Artists, tracks, podcasts...

Search history

yard act

La Femme – La Femme ressort

Clear

•

•

Your library



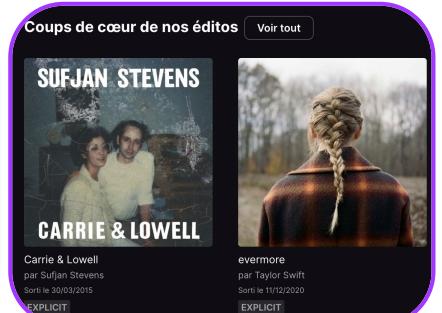
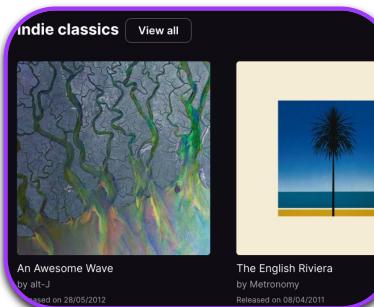
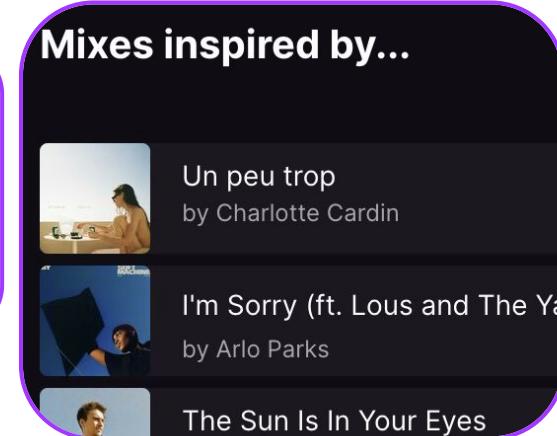
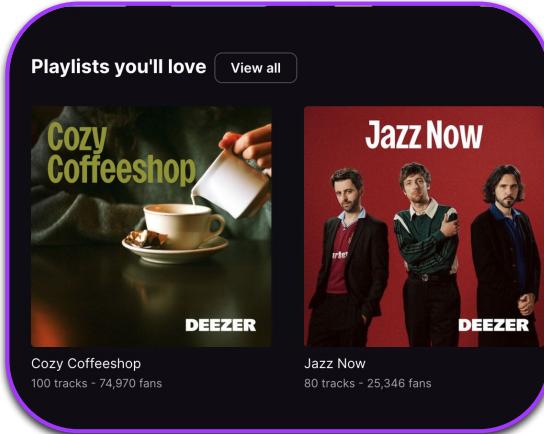
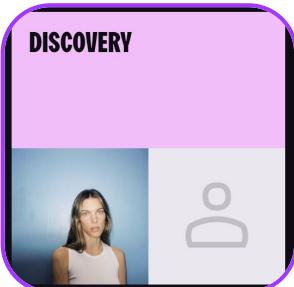
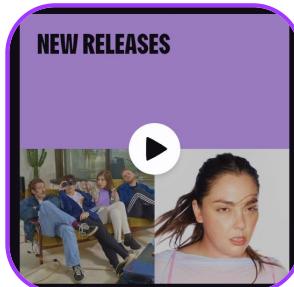
Favoris

75 followers | 65 following

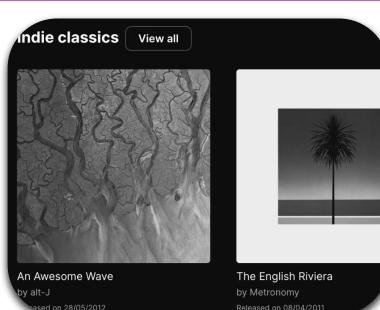
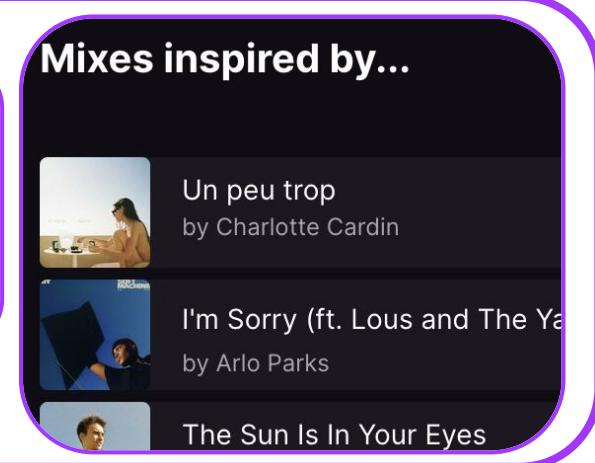
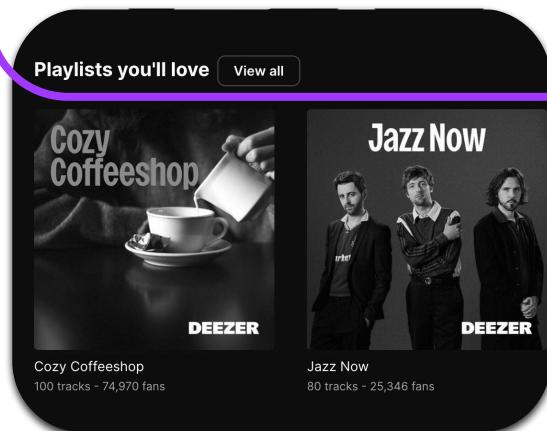
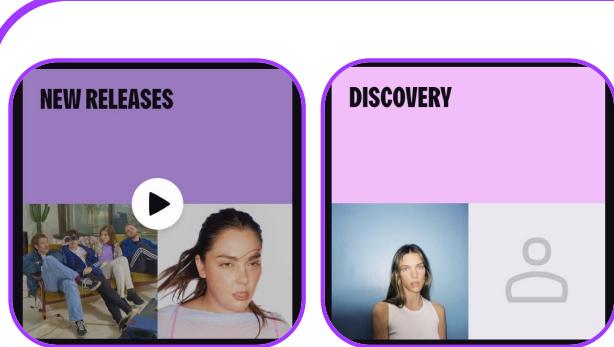
Ma musique en aléatoire

Vue d'ensemble Coups de cœur Playlists Albums Artistes

... But if you don't really know : algorithmic & editorial content



Focus on algorithmic content



Our mission at Recommendation team

Enhance user experience by delivering personalized recommendations that inspire music discovery and deepen listener engagement through algorithmic content

Pure algorithmic content



Mixes inspired by...

- Un peu trop by Charlotte Cardin
- I'm Sorry (ft. Lous and The Y... by Arlo Parks
- The Sun Is In Your Eyes

DAILY



NEW RELEASES



"Algotorial"

Playlists you'll love [View all](#)

Cozy Coffeeshop 100 tracks - 74,970

Jazz Now

Indie classics [View all](#)

An Awesome Wave by alt-J Released on 28/05/2012

The English Riviera by Metronomy Released on 08/04/2011

Order all these contents on the homepage

play how you feel

Flow

New releases

Workout

Party

Chill

Sed

Love

Focus

Discover

DAILY

Featuring Belles, Art Brut, SOLES, MN

Featuring Royal Republic, Marmozets, Black Honey, Damon Albarn

Featuring Arctic Monkeys, MAC MILLER, The Phoenix, Coldplay, System of a Down

Mixes inspired by...

Heartbreak by alt-J

Darkwave by The Cranberries

Love Bi Fidels by Arctic Monkeys

My Little Pony by Arctic Monkeys

Shallow by Ed Sheeran

I Will Wait For You by Arctic Monkeys

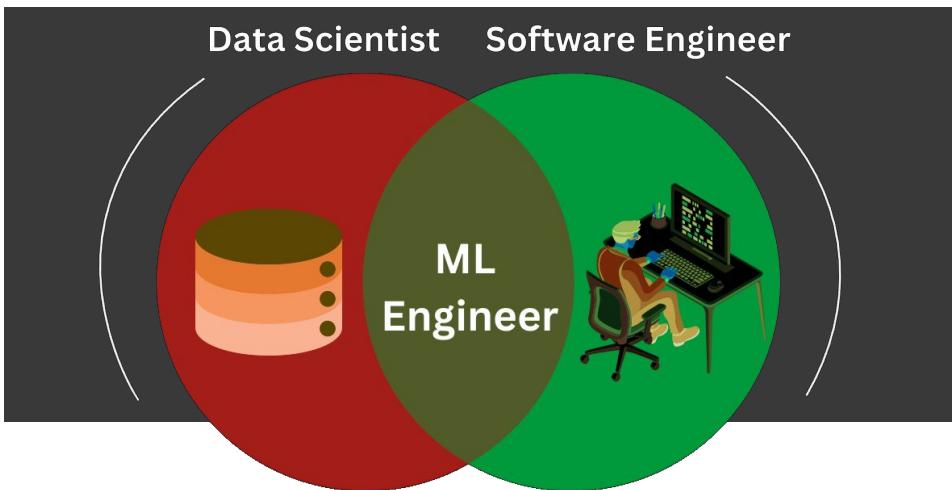
grasp by Arctic Monkeys

The Cranberries' cult albums [View all](#)

Our mission at Recommendation team

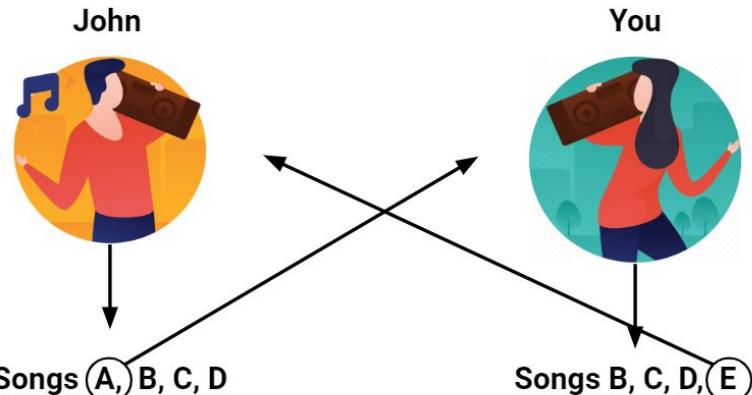
What is a Data Scientist / Machine Learning Engineer ?

- Bridges algorithms and practical software
- Designs and implements scalable ML systems and algorithm for production
- Combines software engineering practices with data science



Personalized recommendations at scale :

Collaborative Filtering for recommendation



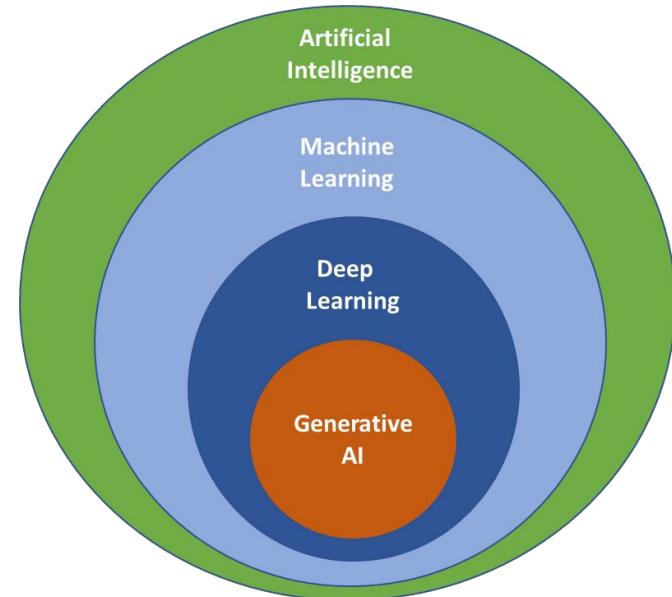
Cloud Native technologies & on-premise infrastructures for scaling



2. Collaborative Filtering for recommendation

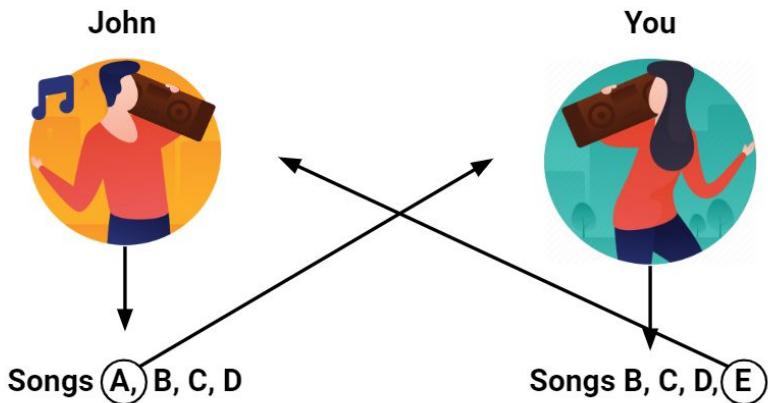
What is Artificial Intelligence (AI) and what is the link with Machine Learning (ML) ?

- **Artificial Intelligence (AI):** technology that simulates human behavior, allowing machines to learn and execute tasks without explicit programming
- **Machine learning (ML):** A branch of AI where systems learn autonomously, using data and algorithms to make predictions.
- **Deep learning:** A specialized area of ML that employs multilayered neural networks to process large datasets and perform complex predictive tasks.
- **Generative AI:** A form of deep learning that generates new content, such as text, code, or images, based on input descriptions.



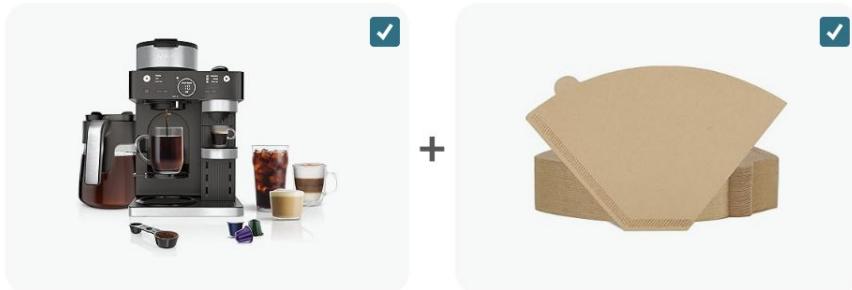
[Source](#)

Collaborative Filtering (CF) for recommendation



Bundle picks from related brands

Based on brands customers bought together



Sponsored ⓘ

This item: Ninja Espresso & Coffee Maker & Barista System, 3 Espresso Brew Styles, Single-...

4.3 ★★★★★ 1,872

-24% \$189⁹⁵ (\$15.83/Count)

List: \$249.99

160 Count #4 Cone Coffee Filters
8-12 Cup Disposable Coffee Filters 4 Cone Paper Unbleache...

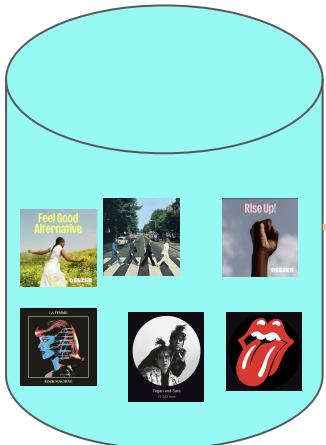
4.7 ★★★★★ 452

\$11⁹⁹ (\$0.07/Count)

✓prime

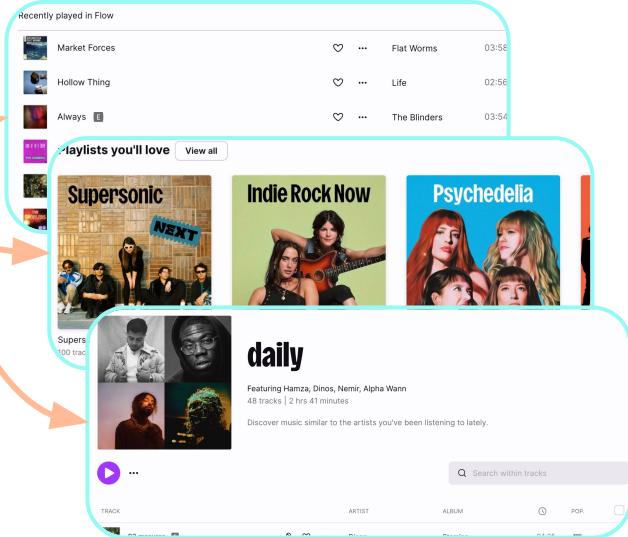
The recommendation engine filters out the catalog ...

Catalog
Millions of tracks,
artists, albums,
playlists...

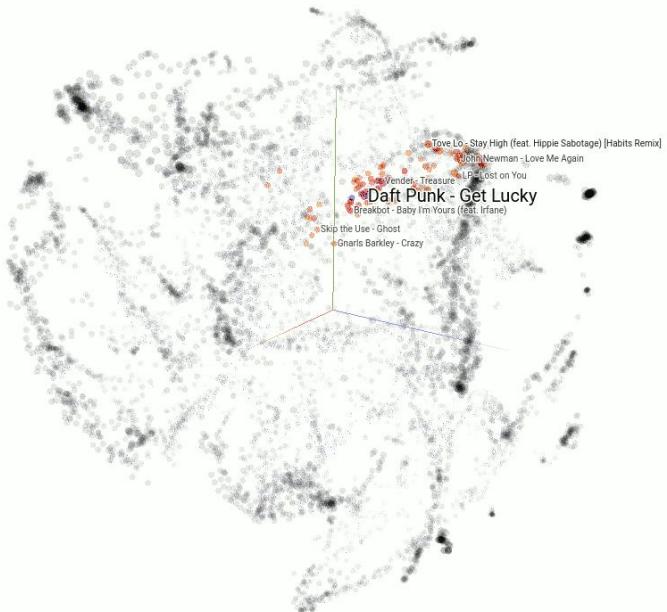


This is the recommendation engine

Recommendations
Tailored list of items



Inside the engine: embeddings representation for recommendation



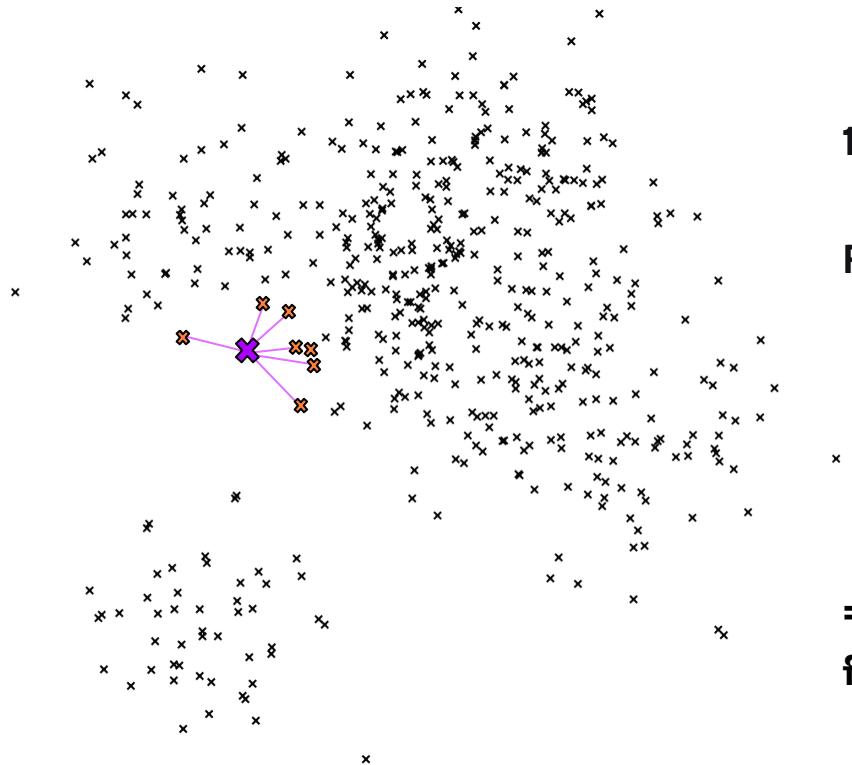
Dataset of user-made playlists
(~5 millions playlists)

Build a track co-occurrence matrix
(~2 millions tracks)

Factorize it using Singular Value Decomposition (SVD)

1 track = 1 embedding vector

Inside the engine: embeddings representation for recommendation



1 track = 1 embedding vector

From track embeddings, we compute:

- Artist, Album, Playlist embeddings;
- User embeddings.

Proximity \Leftrightarrow Similarity

=> We can generate personalized lists from any point in the space

Technical challenge

Exact Neighbours search would require 10M users \times 2M songs
= **20 trillion comparisons.**

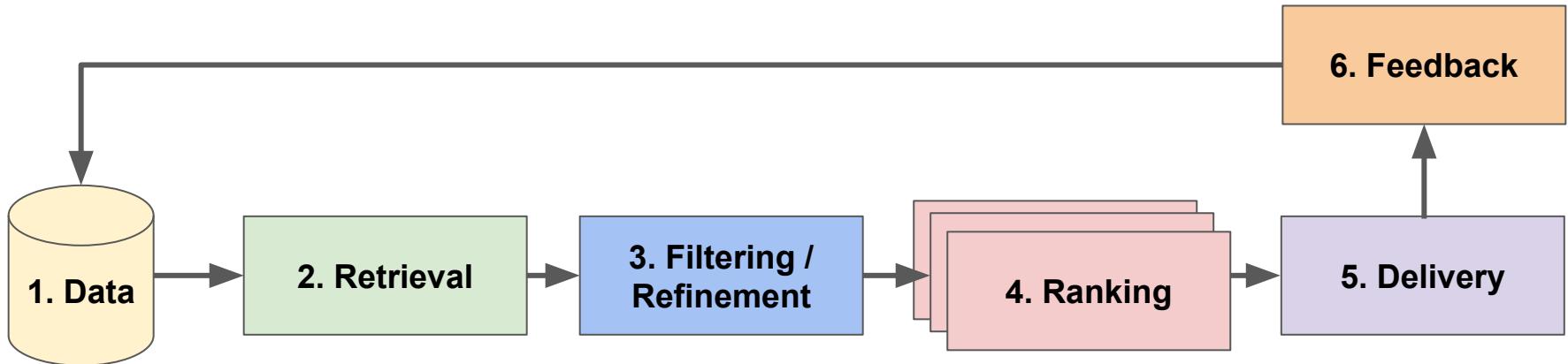
Computationally infeasible for real-time recommendations.

Approximate nearest neighbour
(ANN) search for recommendation :
generating personalized list of items.

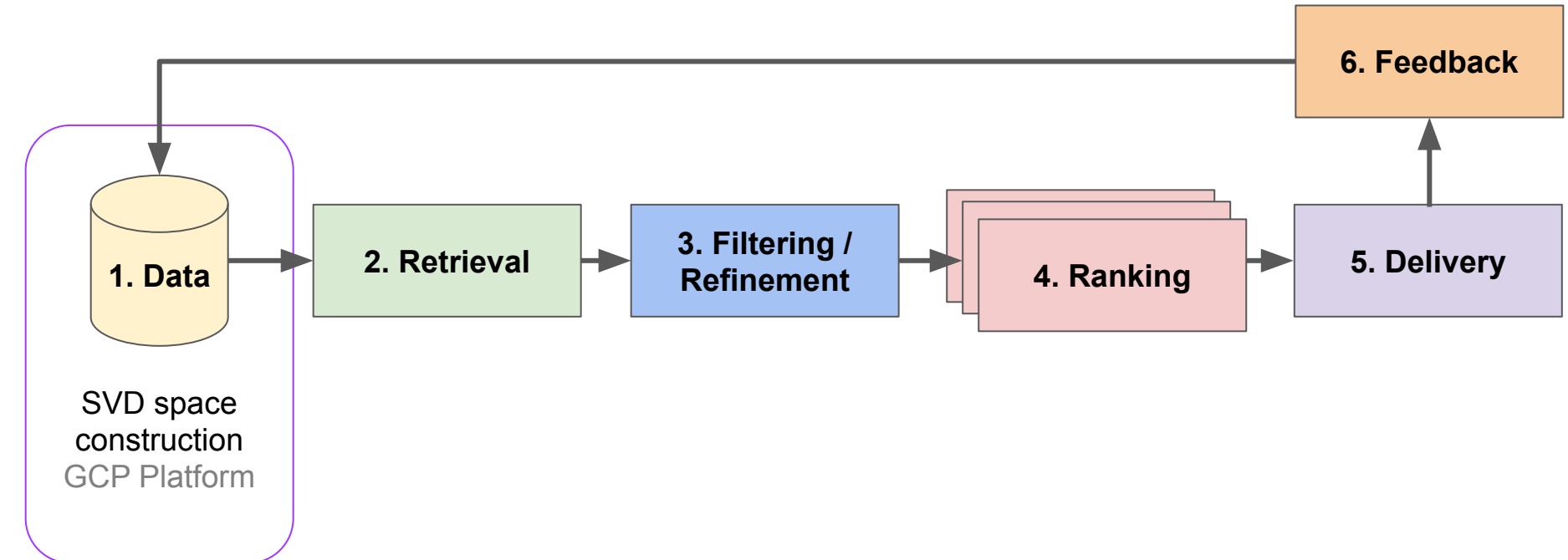
Using **Faiss** library in our **recovoisine** microservice

3. Cloud Native technologies & on-premise infrastructures for scaling

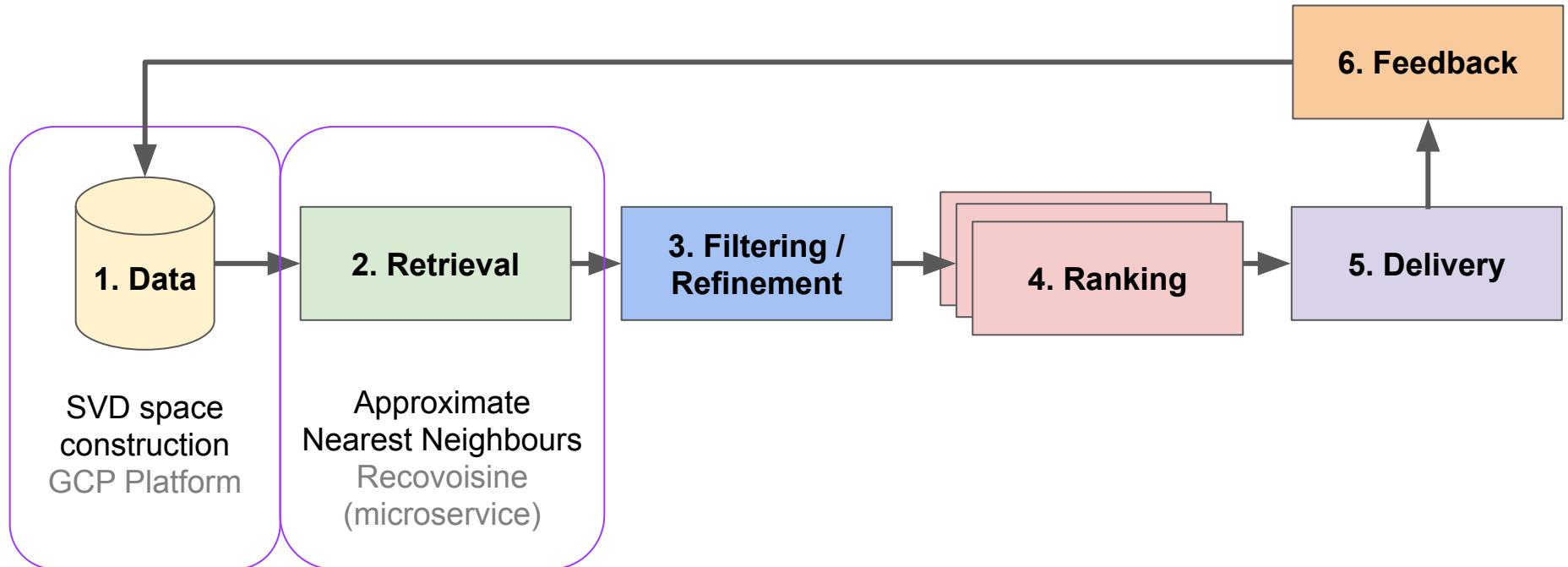
How we deliver the algorithm outputs to millions of users



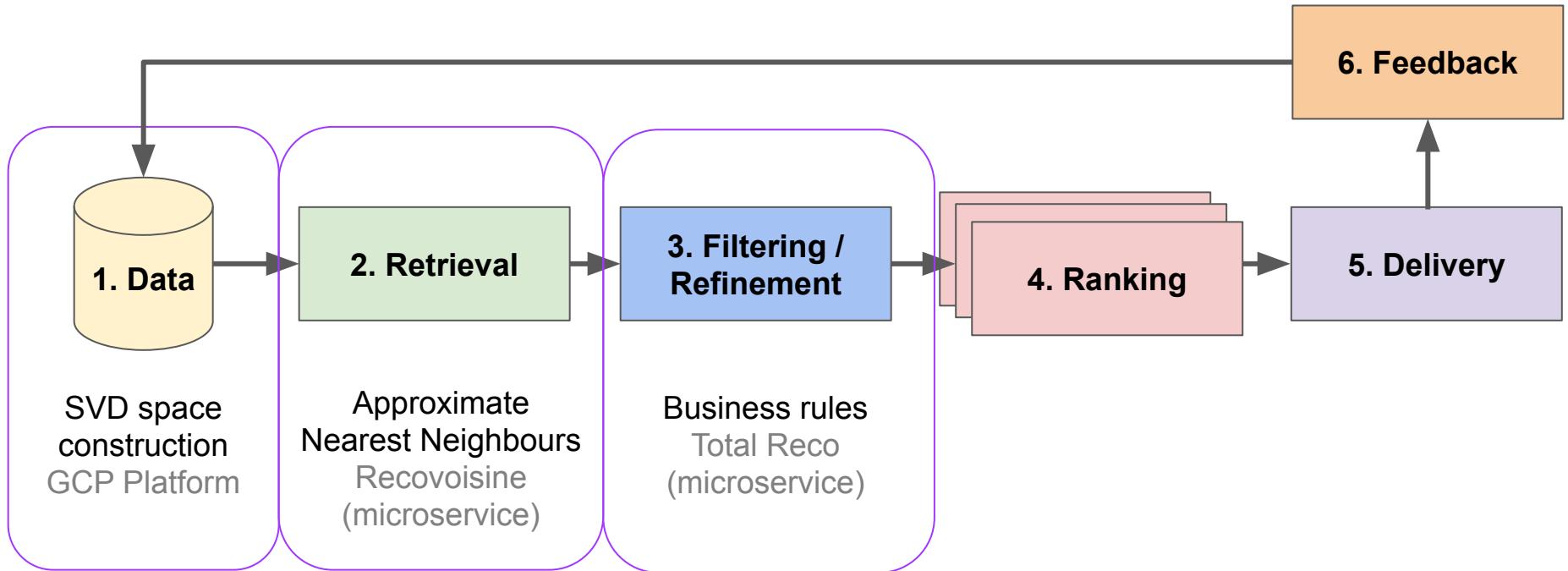
How we deliver the algorithm outputs to millions of users



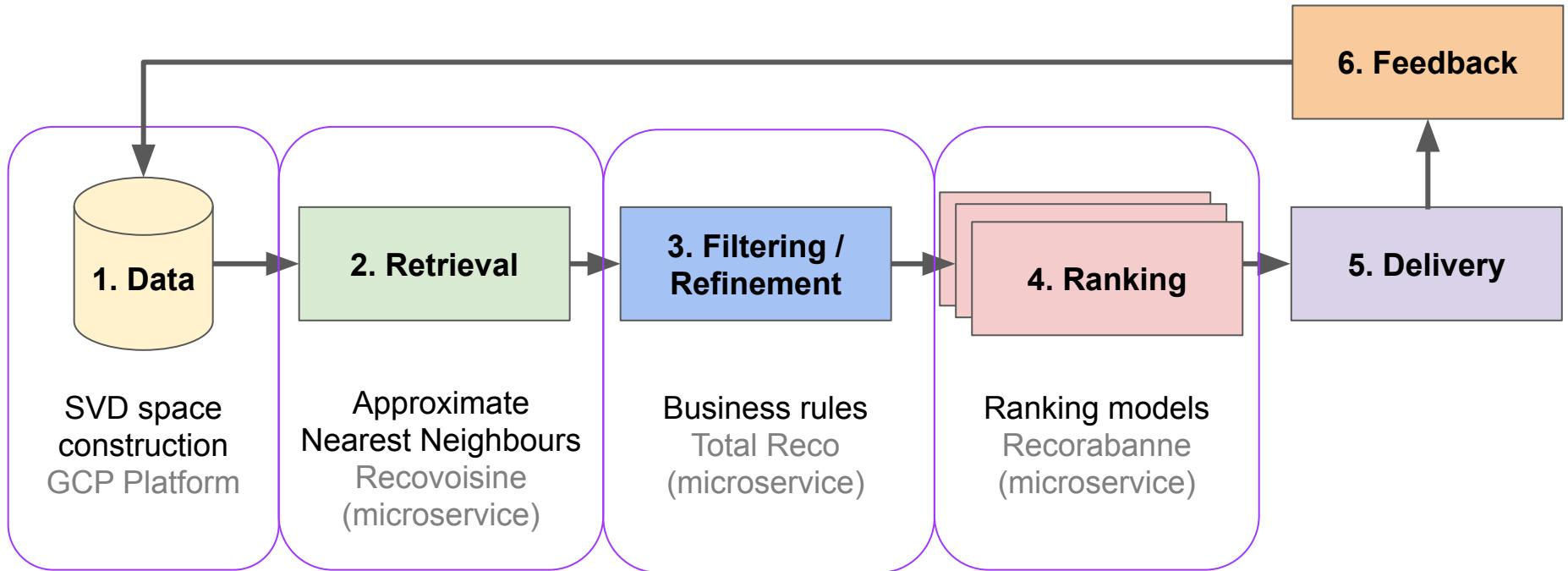
How we deliver the algorithm outputs to millions of users



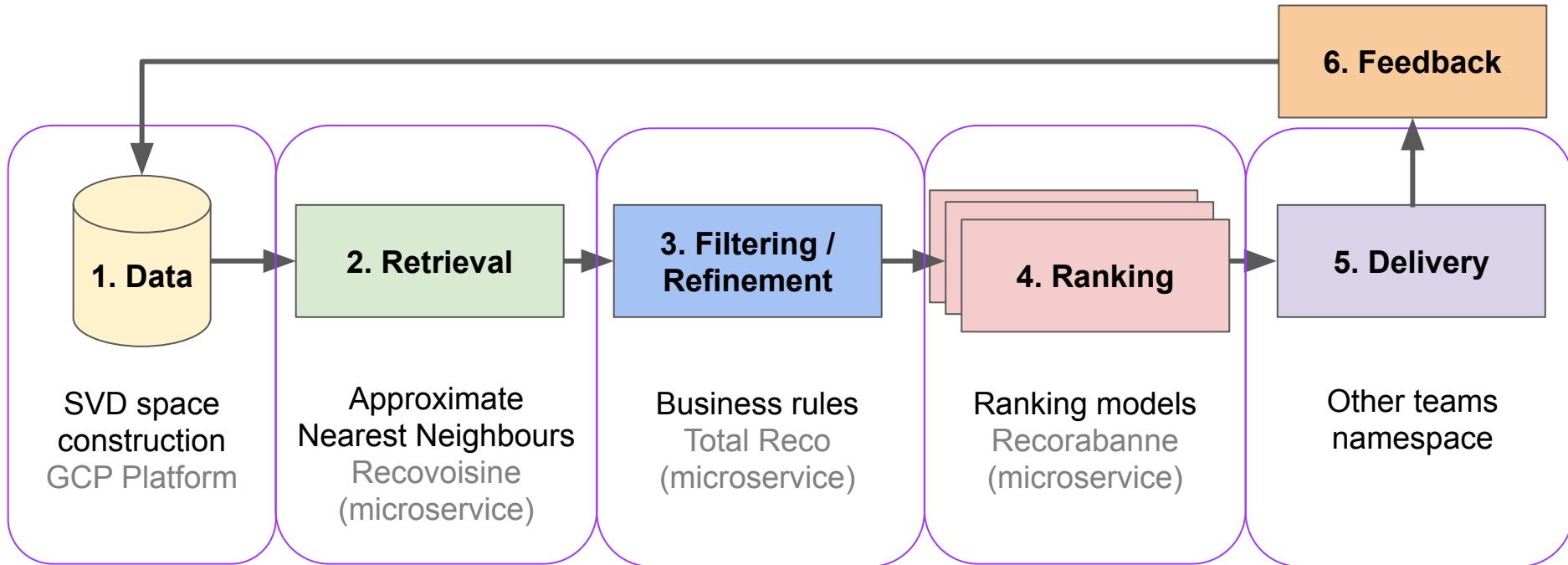
How we deliver the algorithm outputs to millions of users



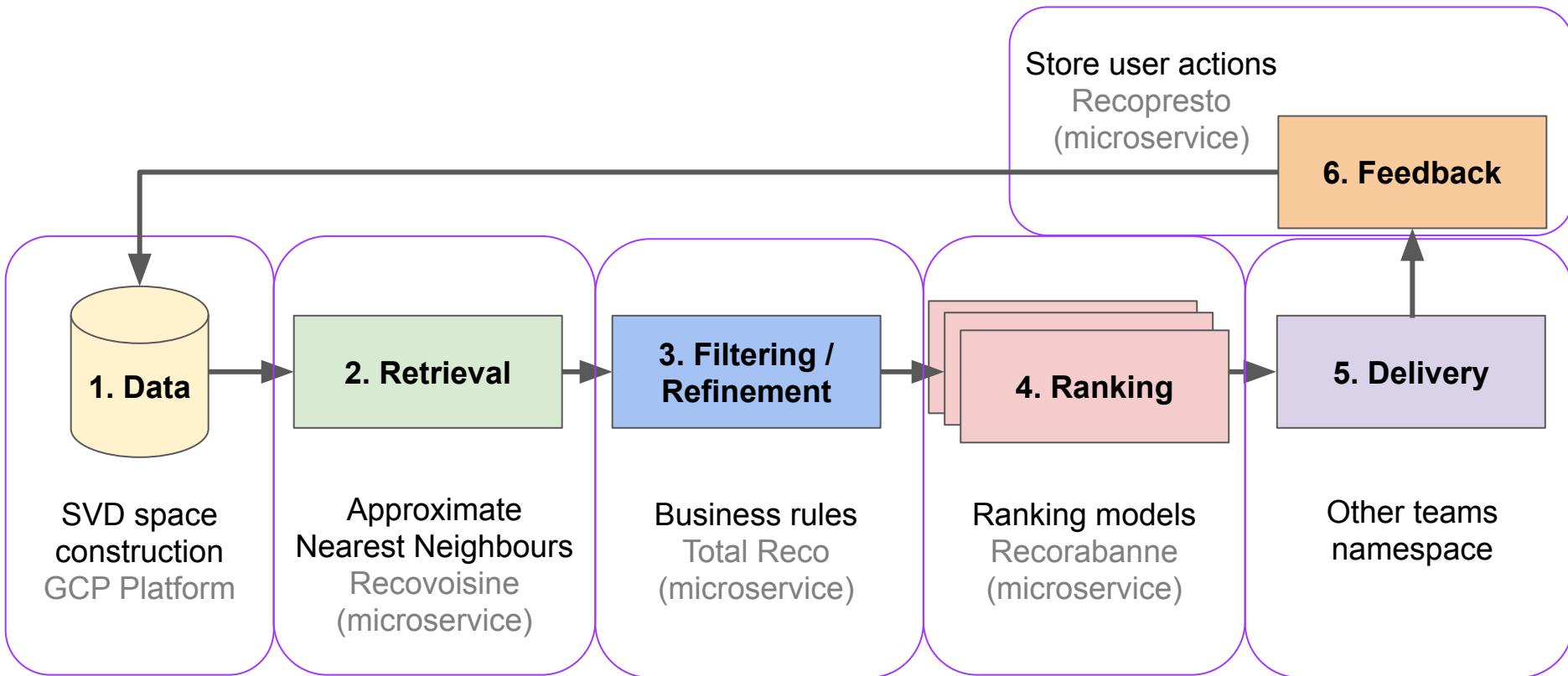
How we deliver the algorithm outputs to millions of users



How we deliver the algorithm outputs to millions of users

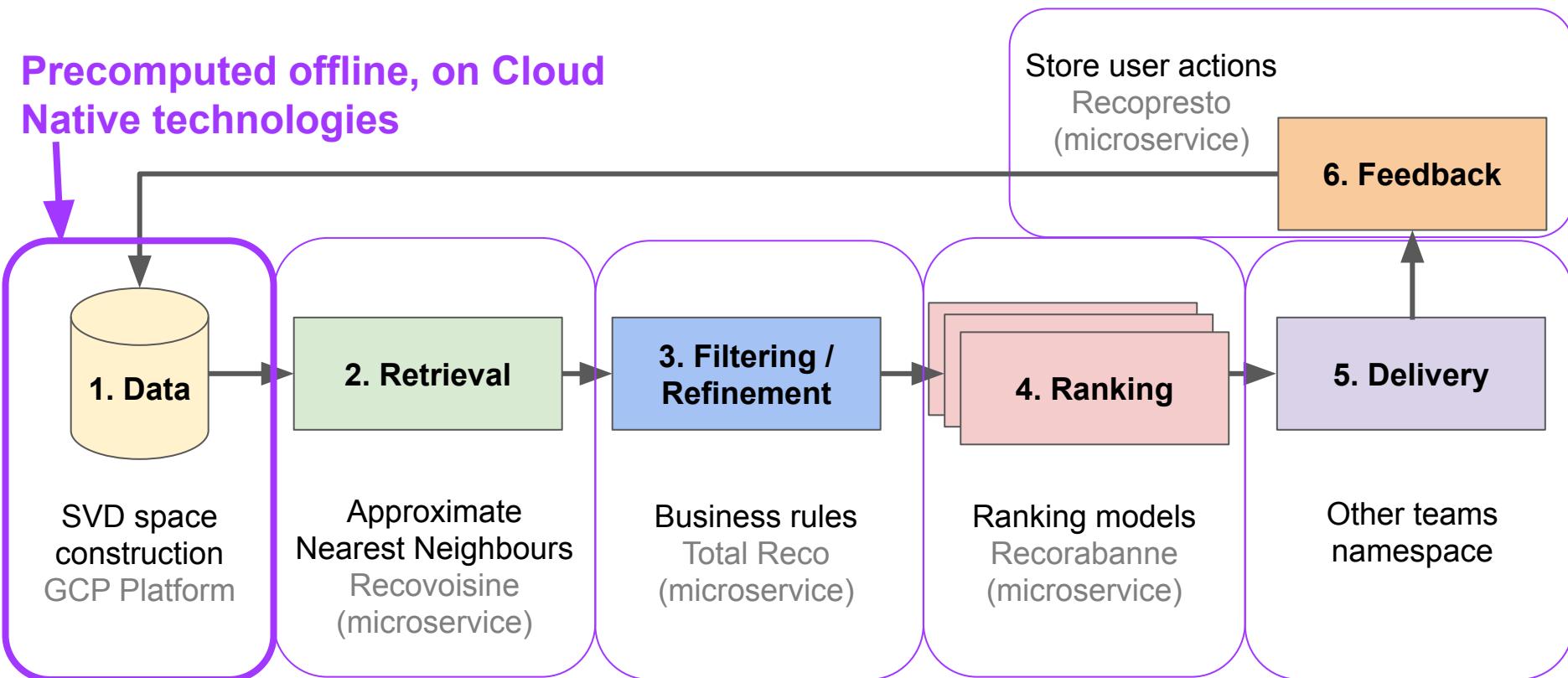


How we deliver the algorithm outputs to millions of users

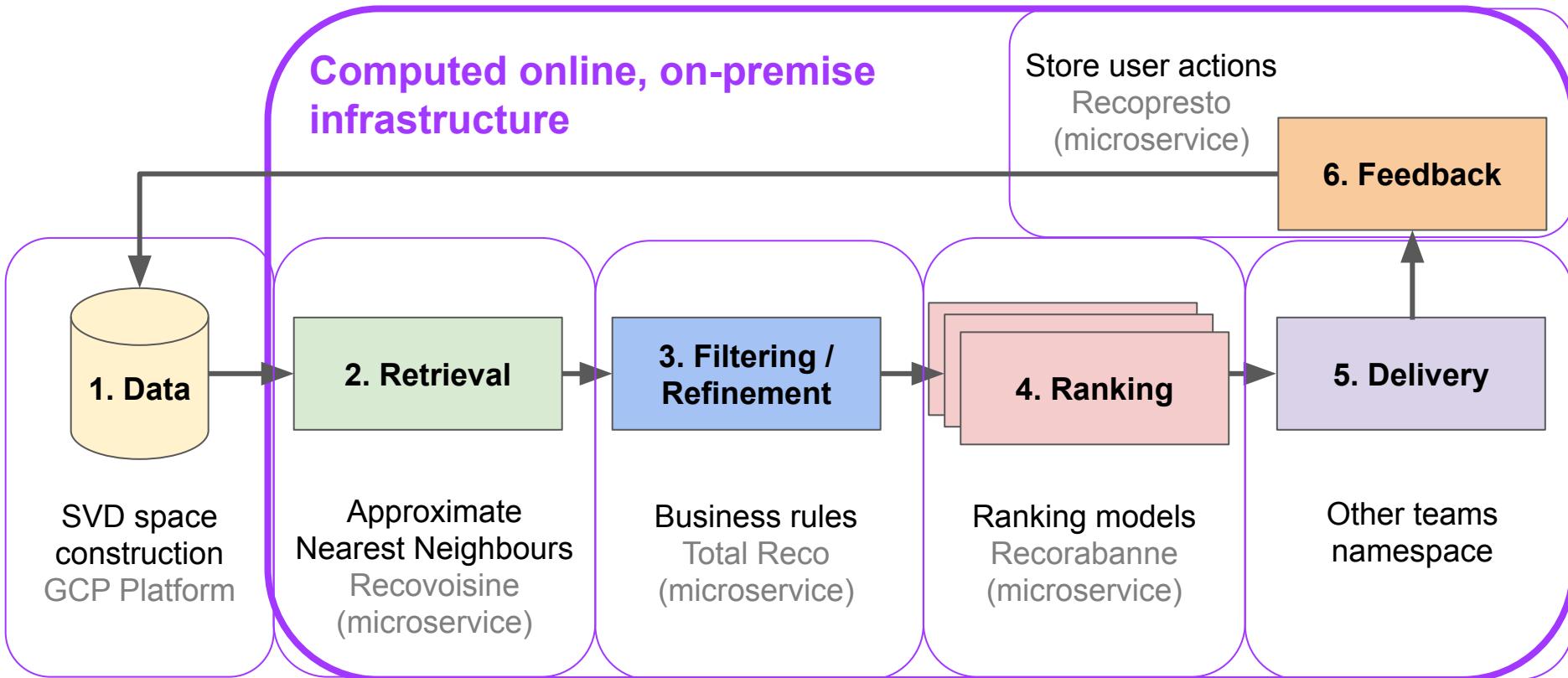


How we deliver the algorithm outputs to millions of users

Precomputed offline, on Cloud
Native technologies



How we deliver the algorithm outputs to millions of users



Used almost everywhere @Deezer

Pure algorithmic content



Mixes inspired by...

- Un peu trop by Charlotte Cardin
- I'm Sorry (ft. Lous and The Yards) by Arlo Parks
- The Sun Is In Your Eyes

DAILY



NEW RELEASES



"Algotorial"

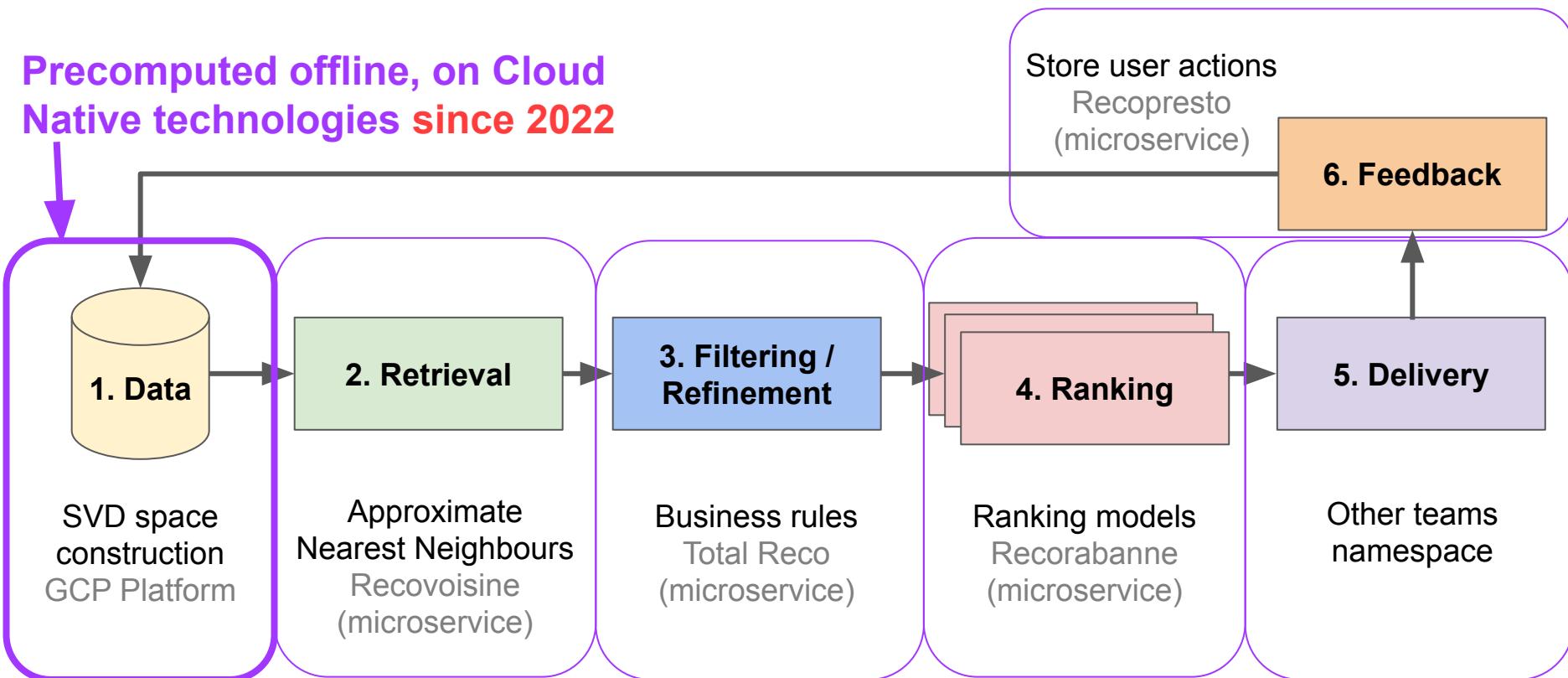
The first screen shows a "Playlists you'll love" section with "Cozy Coffeeshop" and "Jazz Now".
The second screen shows "indie classics" with "An Awesome Wave" by alt-J and "The English Riviera" by Metronomy.
The third screen shows a "Discover" section with various tracks and artists.

Order all these contents on the homepage

The homepage features a "play how you feel" section with categories like Flow, New releases, Workout, Party, CHILL, Sed, Love, and Focus. Below this are sections for "Discover" (Daily mixes), "Mixes inspired by...", and "The Cranberries' cult albums".

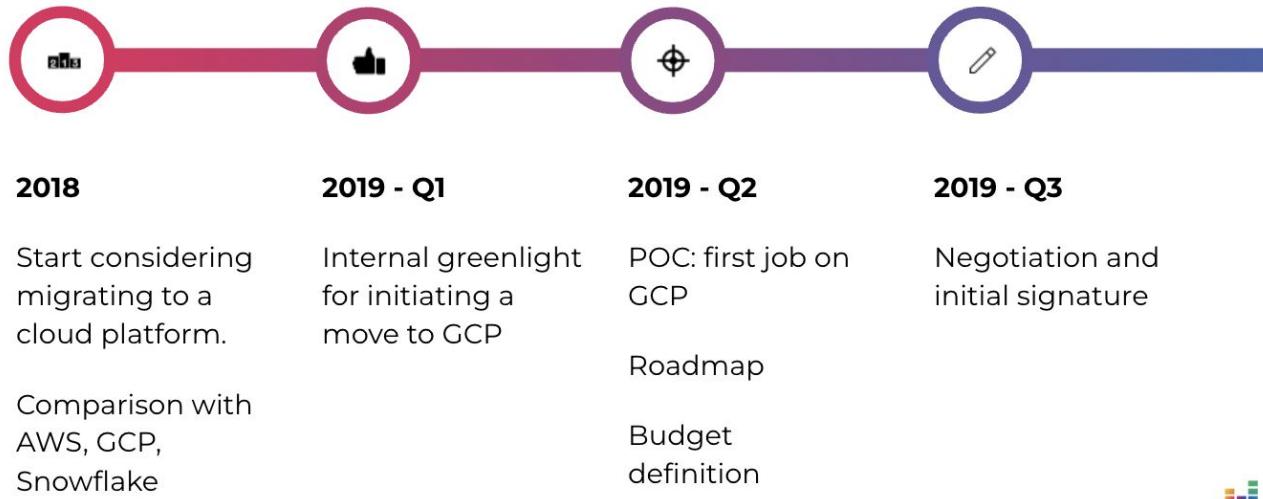
How we deliver the algorithm outputs to millions of users

Precomputed offline, on Cloud
Native technologies since 2022



4 years of GCP migration :

Migration Timeline 1/2



4 years of GCP migration :

Migration Timeline 2/2



2020

First jobs in production on GCP

Data exported in GCS and BigQuery

2021

3 years contract signed with Google

Teams started migrating to GCP

May 2022

End of the migration.

Cleaning to be performed, machines to be decommissioned



Key changes:



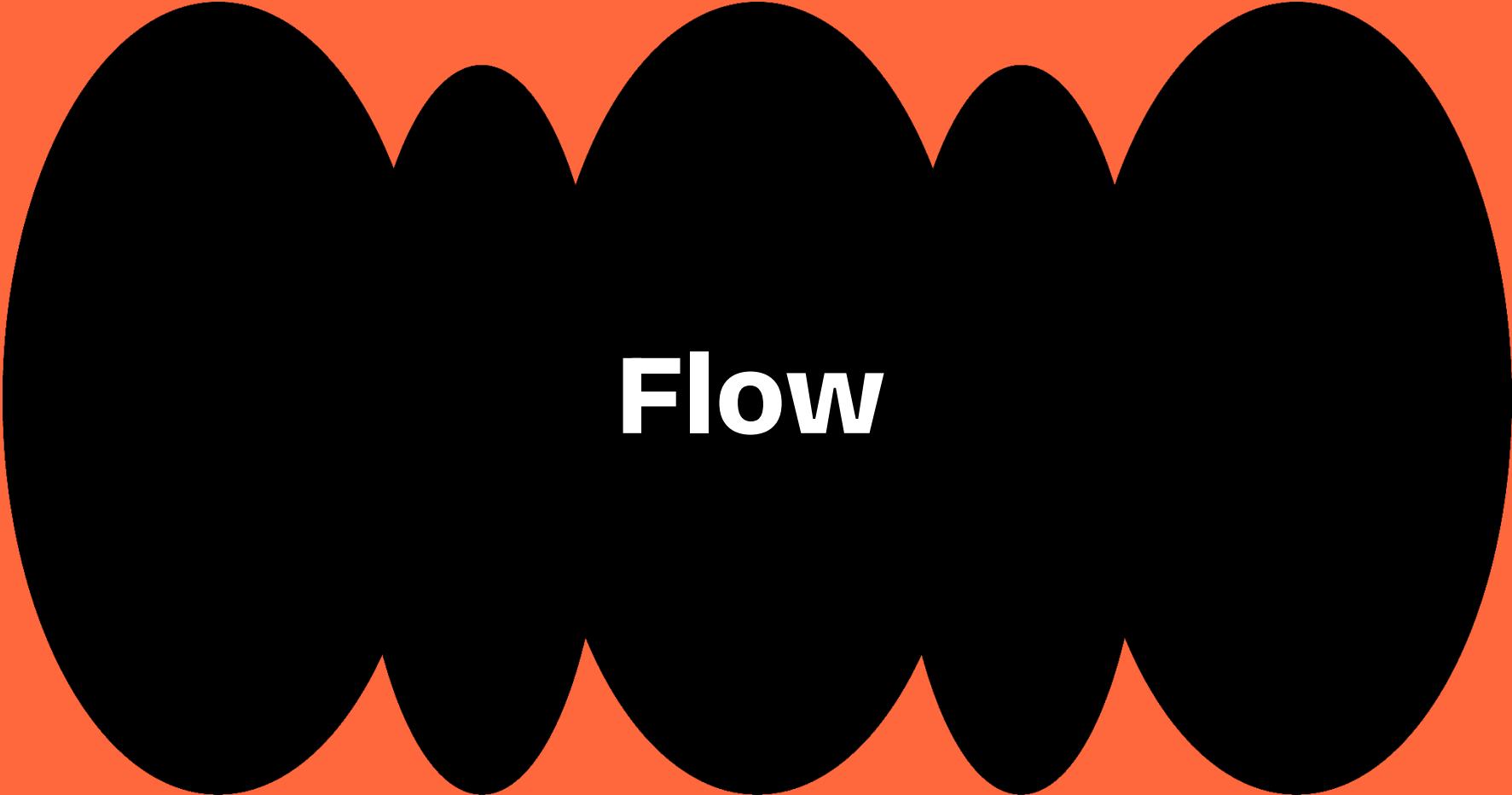
- ✗ Scalable: 3 months to add a node
- ✗ Easy maintenance
- ✗ Cost control: expensive workforce
- ✗ Flexible: 1 cluster = 1 environnement, hard to upgrade
- ✗ Fast analytics - Hive



- ✓ Scalable: **managed** data platform that can scale on demande, in minutes
- ✓ Easy maintenance: none on our part
- ✓ Cost control: 1 price = 1 level of service
- ✓ Flexible: configurable clusters on demand
- ✓ Fast analytics - BigQuery
- ✓ Lots of tools and services available out of the box



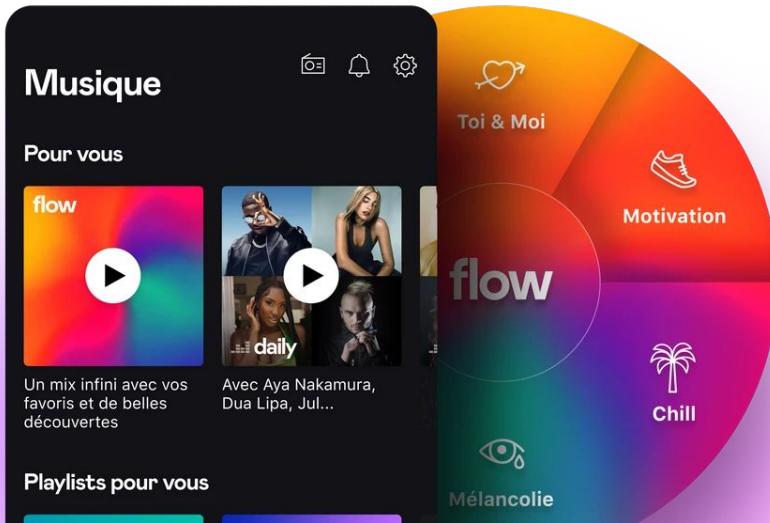
4. Concrete examples



Flow

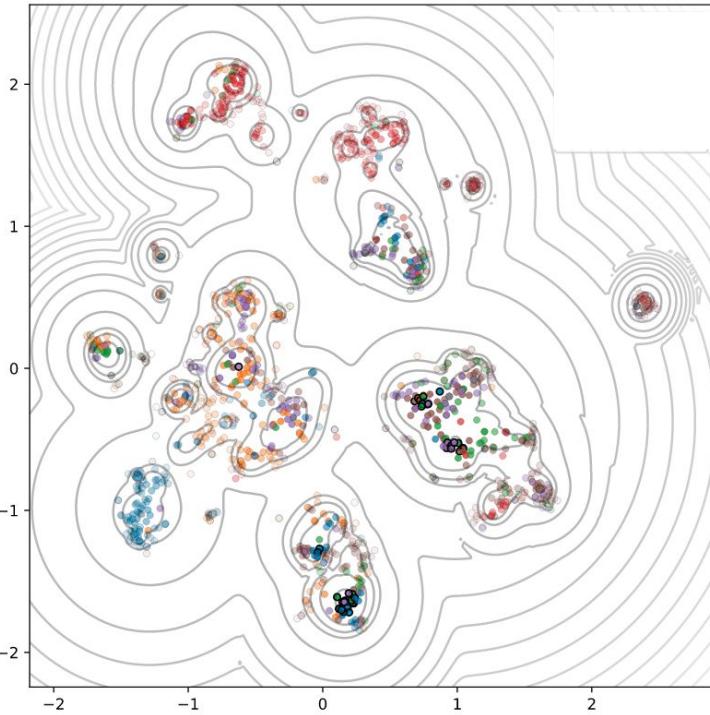
What is the Flow ?

*An Infinite Mix with **favorites** and **discoveries***



- Flagship reco feature of Deezer
- Introduced in 2014
- Millions of monthly active users
- Your personal and unique soundtrack
- Mixing familiar songs, discoveries, and forgotten gems

Flow: an exploration of the map centered on the user profile





Flow Mood

Give the user some control

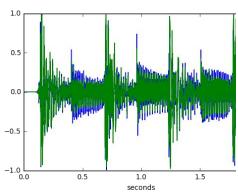
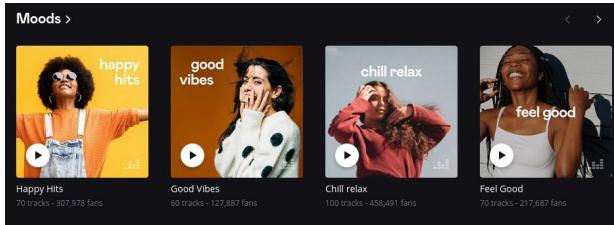
It's impossible to read people's mind.

Giving some **control** allow deeper **discoveries**

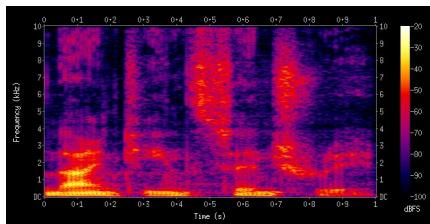


Detecting some moods & genres from audio

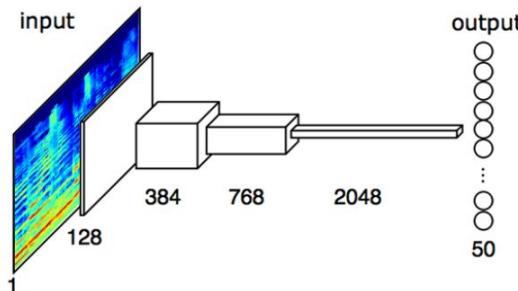
Annotated data



raw audio



spectrogram

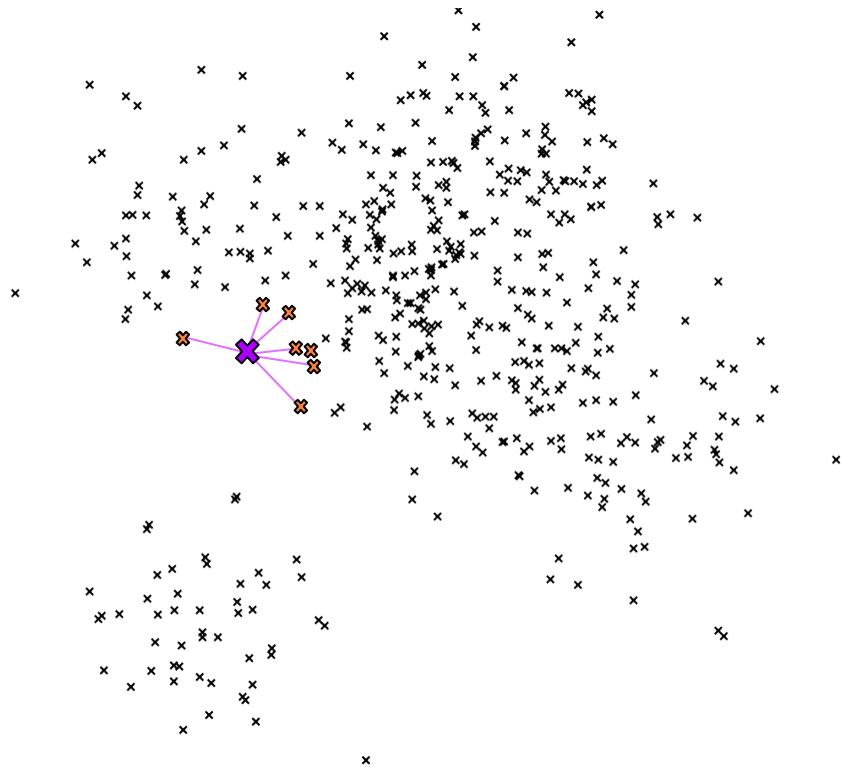


deep learning model to predict moods probabilities



New Releases

Coldstart challenge

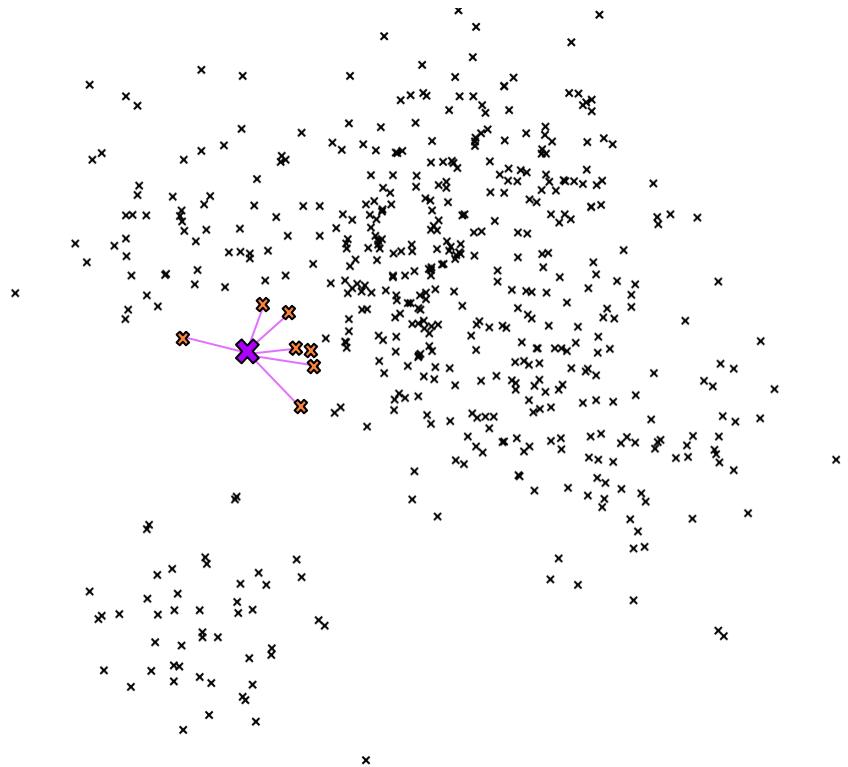


1 track = 1 **embedding** vector

From track embeddings, we compute:

- **Artist, Album, Playlist embeddings;**
- **User embeddings.**

Coldstart challenge



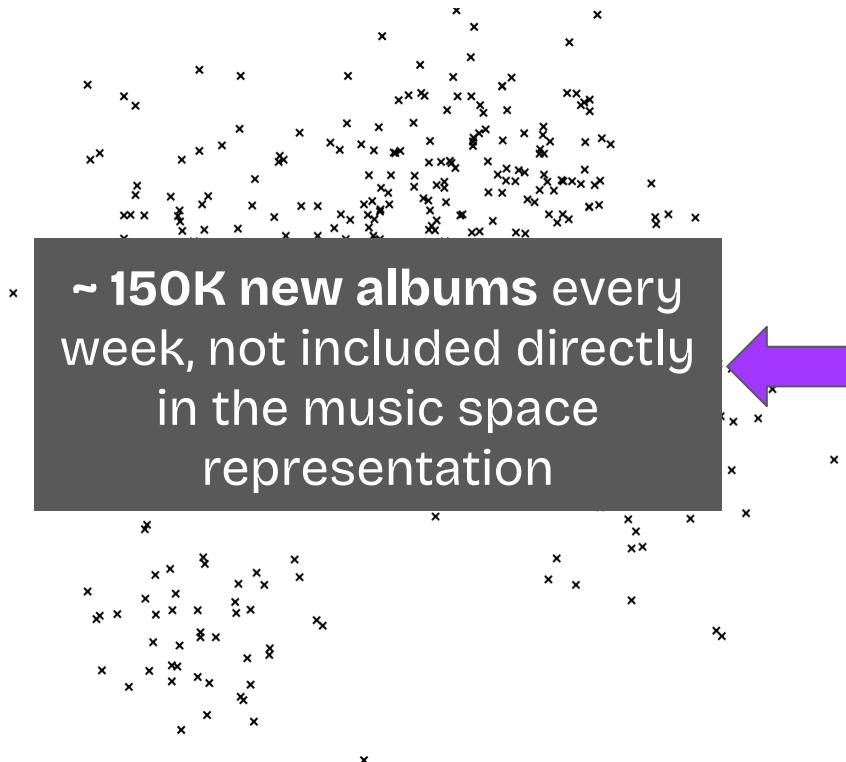
1 track = 1 embedding vector

From track embeddings, we compute:

- Artist, Album, Playlist embeddings;
- User embeddings.

For music items with
enough usage data at the
time of the space
computation

Coldstart challenge



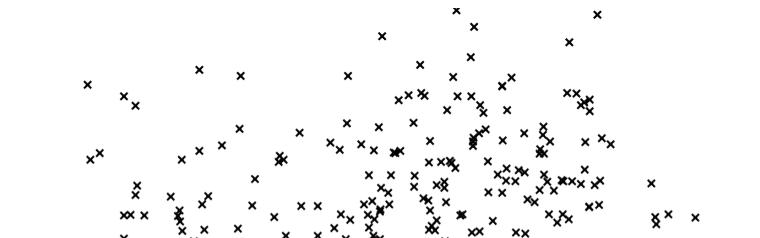
1 track = 1 embedding vector

From track embeddings, we compute:

- Artist, Album, Playlist embeddings;
- User embeddings.

For music items with enough usage data at the time of the space computation

Coldstart challenge



~ 150K new albums every week, not included directly in the music space representation



1 track = 1 embedding vector

From track embeddings, we compute:

- Artist, Album, Playlist embeddings;
- User embeddings.

For music items with enough usage data at the time of the space computation

Coldstart challenge:

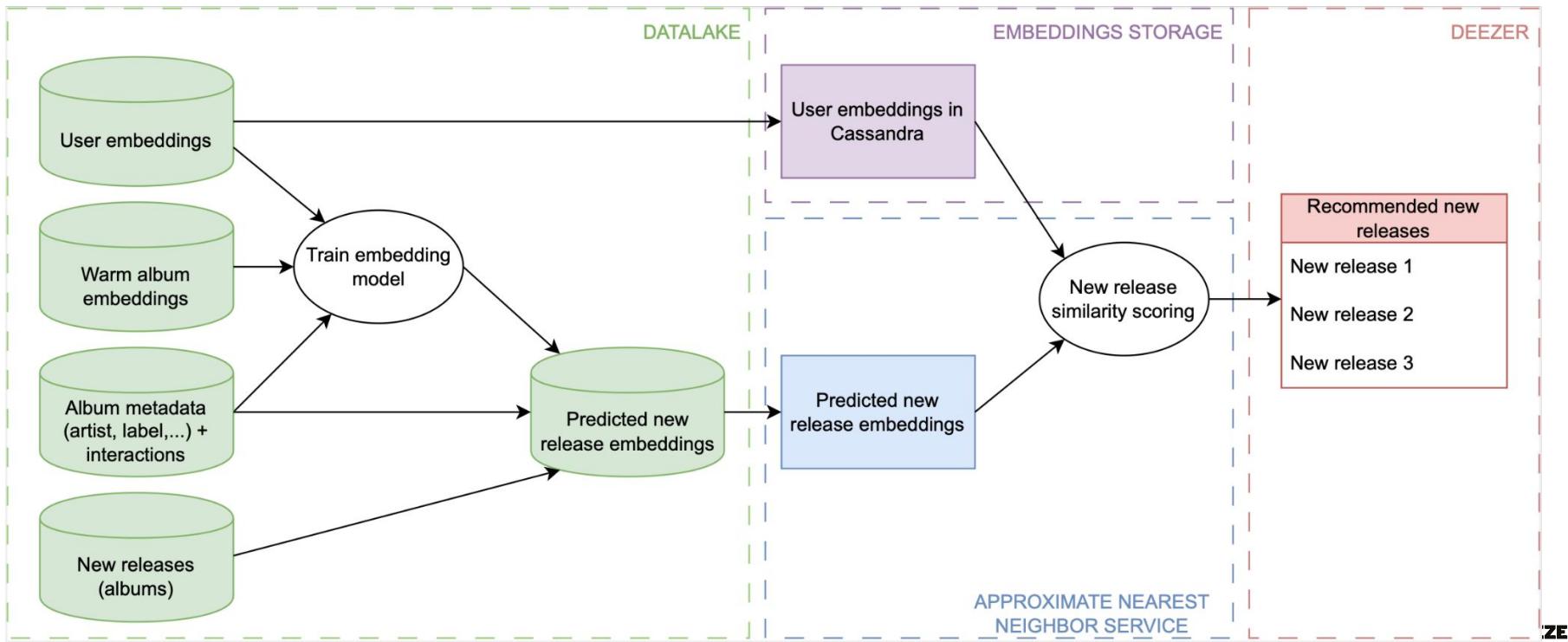
How to project new releases in this space, despite having few/no usage data?

Let's Get It Started: Fostering the Discoverability of New Releases on Deezer

Léa Briand^{1,*}, Théo Bontempelli¹, Walid Bendada^{1,2},
Mathieu Morlon¹, François Rigaud¹, Benjamin Chapus¹,
Thomas Bouabça¹, and Guillaume Salha-Galvan¹

[Read the paper to know more about it !](#)

[ECIR 2024 Industry paper \(Glasgow, UK\)](#)
[RECSYS 2024 MuRS workshop \(Bari, Italy\)](#)



Providing fresh reco @deezer

Dedicated feature for new releases : New Releases For You carousel on the dynamic Homepage

New releases for you > < >



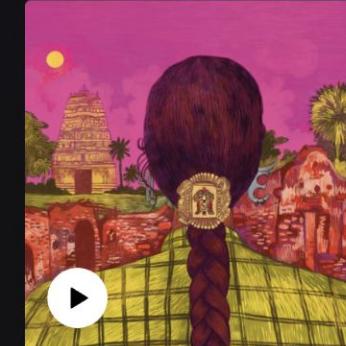
Suburban House
by Holly Macve
Released on 20/10/2023



Dancer
by IDLES
Released on 18/10/2023



I Want to Die
by Marc Rebillet
Released on 20/10/2023

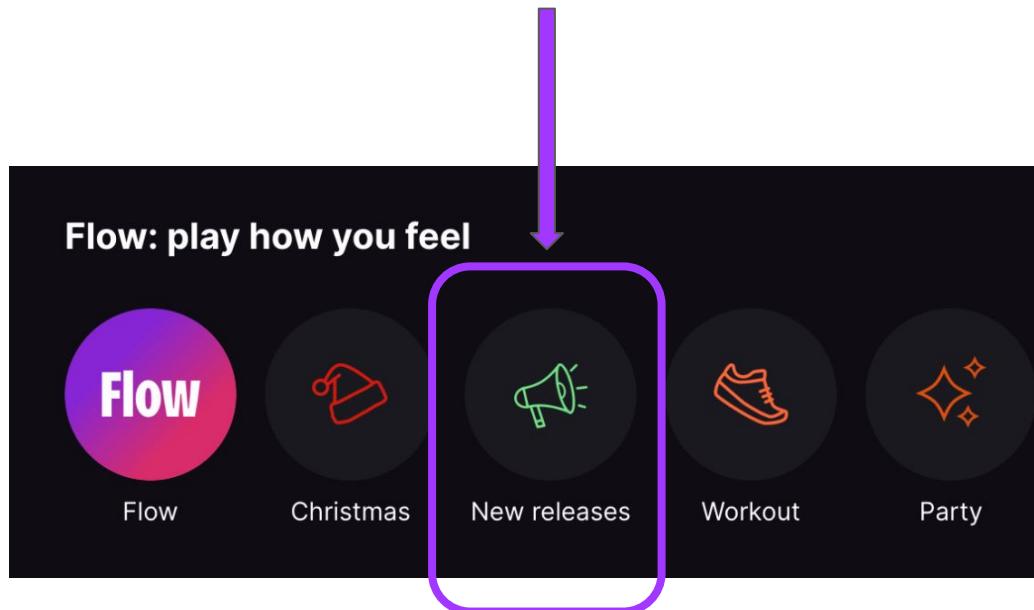


Santhosam
by Priya Ragu
Released on 20/10/2023

Providing fresh reco @deezer

New feature : infinite playlist for new releases tailored for you !

Flow New Releases



Other challenges

In the future @Deezer



Continue to improve user modeling

- Better take into account real time actions
- Repetition affinity
- ...

In the future @Deezer



Continue to improve user modeling

- Better take into account real time actions
- Repetition affinity
- ...



Enrich music items representation

- Metadata (biography, reviews, tags...)
- Lyrics
- ...

In the future @Deezer



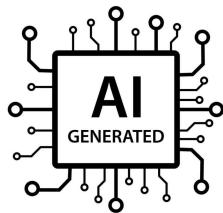
Continue to improve user modeling

- Better take into account real time actions
- Repetition affinity
- ...



Enrich music items representation

- Metadata (biography, reviews, tags...)
- Lyrics
- ...



Leveraging generative AI (LLM)

- For cold start or "long tail" items
- New features
- ...

Stay tuned !

Créer ma playlist



Playlist

Add your favorite songs



AI Playlist **BETA**

Generate a playlist from an idea

Generative AI playlist tool
BETA test

Purple Club

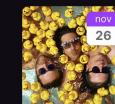
En tant qu'abonné·e, gagne tes places pour des concerts, des évènements exclusifs et bien plus.

Concours du moment



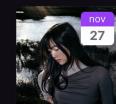
Concerts

Around you



Naïve New Beaters

26 novembre • L'Olympia



Sarah Kinsley, Linn Koch-Emmery

27 novembre • La Cigale



Shaka Ponk

27 novembre • AccorHotels Arena

Gagne tes places pour le concert de THEODORT

Concert • 06/03/25 • Paris

Purple Door Club
Live experiences

Q & A

Thank you

