



PADDYPOWER.

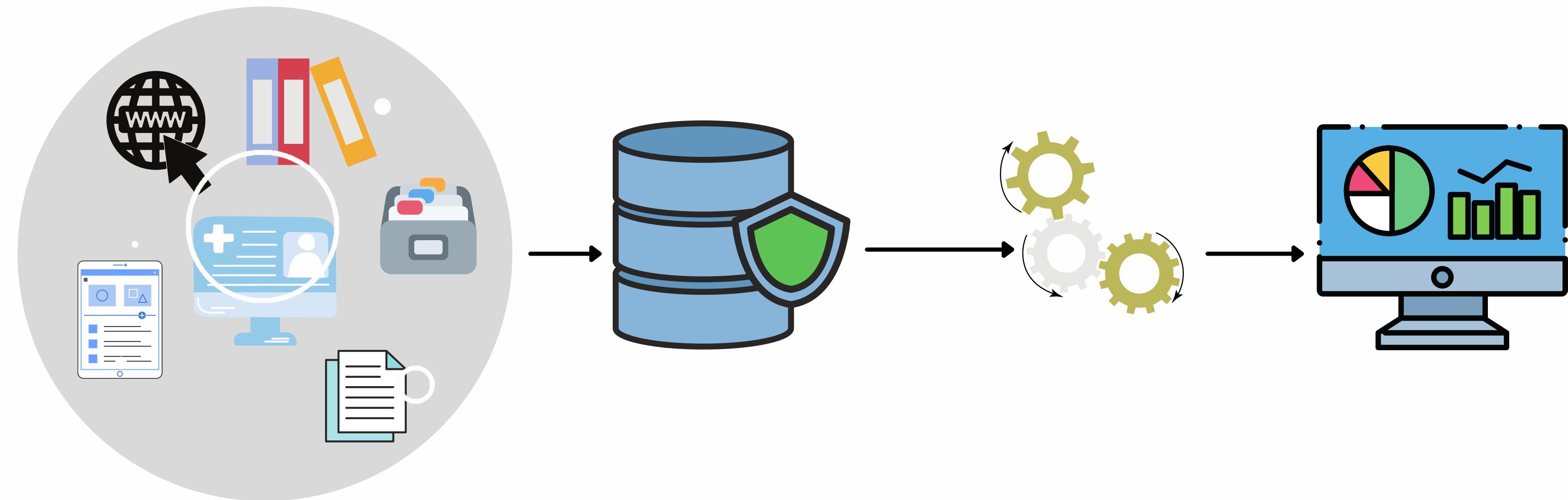
 betfair

# Planificarea

PADDYPOWER.

betfair

Traseul datelor de la sursă la raportare



# Arhitectura bazei de date



- Structura obiectelor și relațiile dintre ele
  - Chei primare / unice
  - Chei străine
- Tipuri de date
  - Indexări
- Triggers

# Arhitectura bazei de date



Account

Date

Song

Streams

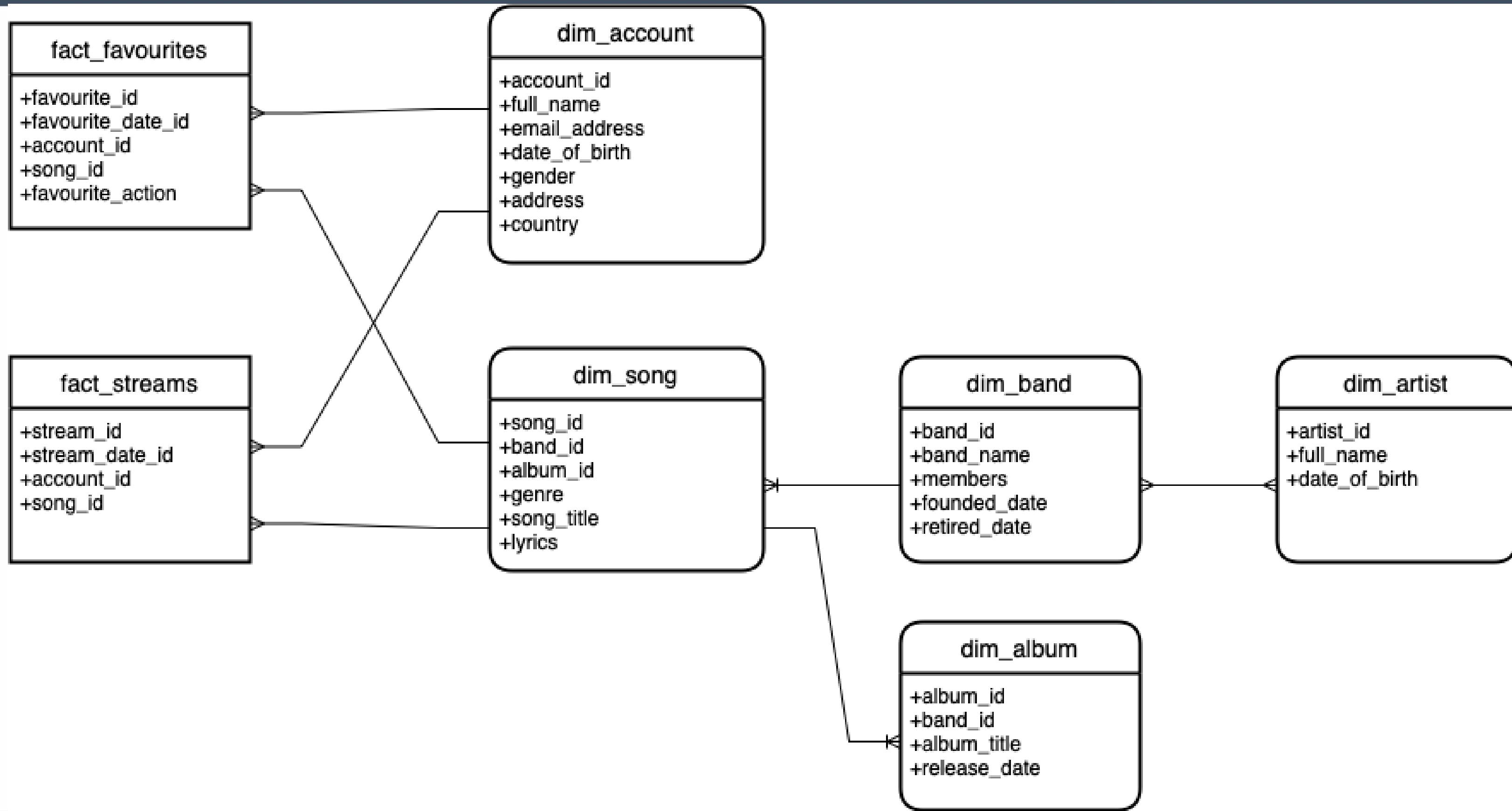
Favourites

Band

Album

Artist

# Arhitectura bazei de date



# Crearea obiectelor in BQ



- Proiect nou în Google Cloud Platform
- Dataset nou în BigQuery
  - Tabele noi în dataset -> pe baza de schemă sau completarea fiecărei coloane
- (Optional) Un bucket nou in Google Storage
- Popularea tabelelor cu date de test
  - Mockaroo -> încărcare fișiere csv în Google Storage
  - Faker -> generare date de test cu ajutorul librăriei Faker

# Crearea obiectelor in BQ



Viewing pinned projects.

music-streaming-332016



Open



Create dataset



## Create dataset

Project ID

music-streaming-332016

CHANGE

Dataset ID \*

star\_schema

Letters, numbers, and underscores allowed

Data location

europe-west6 (Zurich)



### Default table expiration

Enable table expiration

Default maximum table age

Days

### Encryption

Google-managed encryption key

No configuration required

Customer-managed encryption key (CMEK)

Manage via Google Cloud Key Management Service

CREATE DATASET

CANCEL

# Crearea obiectelor in BQ



Viewing pinned projects.

music-streaming-332016



Open

Delete

Create table



Create table



## Source

Create table from  
Empty table

## Destination

Project \* music-streaming-332016 [BROWSE](#)

Dataset ID \* star\_schema

Table name \* dim\_account

Unicode letters, marks, numbers, connectors, dashes or spaces allowed.

Table type Native table

## Schema

Edit as text

Field name	Type	Mode	Description
account_id	INTEGER	REQUIRED	
first_name	STRING	NULLABLE	
last_name	STRING	NULLABLE	

CREATE TABLE

CANCEL

# Generarea datelor de test

PADDYPOWER.

betfair

Optiunea 1: Mockaroo + Încărcare fisier csv in Google Storage

The Mockaroo interface allows users to define the schema for generated test data. The current schema is defined as follows:

Field Name	Type	Options
id	Row Number	blank: 0 % <input type="checkbox"/> $\Sigma$ <input type="checkbox"/>
first_name	First Name	blank: 0 % <input type="checkbox"/> $\Sigma$ <input type="checkbox"/>
last_name	Last Name	blank: 0 % <input type="checkbox"/> $\Sigma$ <input type="checkbox"/>
email	Email Address	blank: 0 % <input type="checkbox"/> $\Sigma$ <input type="checkbox"/>
gender	Gender	blank: 0 % <input type="checkbox"/> $\Sigma$ <input type="checkbox"/>
ip_address	IP Address v4	blank: 0 % <input type="checkbox"/> $\Sigma$ <input type="checkbox"/>

Below the schema, there are settings for the number of rows (1000), file format (CSV), line ending (Unix (LF)), and inclusion of header and BOM. At the bottom, there are buttons for DOWNLOAD DATA, PREVIEW, SAVE THIS SCHEMA, and MORE.

## Create table

### Source

Create table from  
Google Cloud Storage

Select file from GCS bucket \*

music-streaming/dim\_account.csv

BROWSE

File format  
CSV

Source Data Partitioning

### Destination

Project \*  
music-streaming-332016

Dataset ID \*  
star\_schema

Table name \*  
dim\_account

Unicode letters, marks, numbers, connectors, dashes or spaces allowed.

Table type  
Native table

### Schema

Auto detect

Edit as text

# Generarea datelor de test



Optiunea 2: generarea datelor de test cu ajutorul librăriei Faker

În acest caz, trebuie să încarcăm un fișier ajutător în Google Storage care conține librăria și să creăm funcțiile de generare de date.

The screenshot shows the Google Cloud Storage interface. On the left, there's a sidebar with icons for Cloud Storage (selected), Browser (highlighted in blue), Monitoring, and Settings. The main area displays a bucket named "music-streaming". The bucket details show it's located in "europe-west6 (Zurich)", has a "Standard" storage class, "Not public" public access, and no protection. Below the details, there are tabs for "OBJECTS", "CONFIGURATION", "PERMISSIONS", "PROTECTION", and "LII". Under the "OBJECTS" tab, it shows the path "Buckets > music-streaming" and provides options to "UPLOAD FILES", "UPLOAD FOLDER", "CREATE FOLDER", and "MANAGE HOLDS". There are also filters for "Filter by name prefix only" and "Filter objects and folders". A table lists the contents of the bucket, showing one item: "faker.js" (Size: 52 B, Type: application/x-javascript).

Name	Size	Type
faker.js	52 B	application/x-javascript

# Generarea datelor de test

PADDYPOWER.

betfair

Optiunea 2: generarea datelor de test cu ajutorul librăriei Faker



```
1  /** Generator function */
2  CREATE TEMP FUNCTION dummy_account()
3  RETURNS STRUCT<account_id Int,
4          full_name String,
5          email_address String,
6          date_of_birth Timestamp,
7          gender String,
8          address Struct<street_name String,
9                  street_number Int,
10                 city String
11                >,
12                country String
13              >
14 LANGUAGE js
15 AS """
16 var faker = getFaker()
17 var dummy_account = {};
18
19 dummy_account.account_id = faker.random.number();
20 dummy_account.full_name = faker.name.findName();
21 dummy_account.email_address = faker.internet.email();
22 dummy_account.date_of_birth = faker.date.past();
23 let genders = [ 'F' , 'M' ];
24 dummy_account.gender = faker.random.arrayElement(genders);
25 dummy_account.country = faker.address.country();
26
39    /** Insert 1000 dummy records */
40    INSERT INTO
41      `music-streaming-332016.star_schema.dim_account`
42      (SELECT dummy_account().* from UNNEST(GENERATE_ARRAY(1, 1000)));
43
44    select * from `music-streaming-332016.star_schema.dim_account` ;
45
```

Toate resursele sunt disponibile aici:  
[https://github.com/autrefois/bi\\_resources](https://github.com/autrefois/bi_resources)