



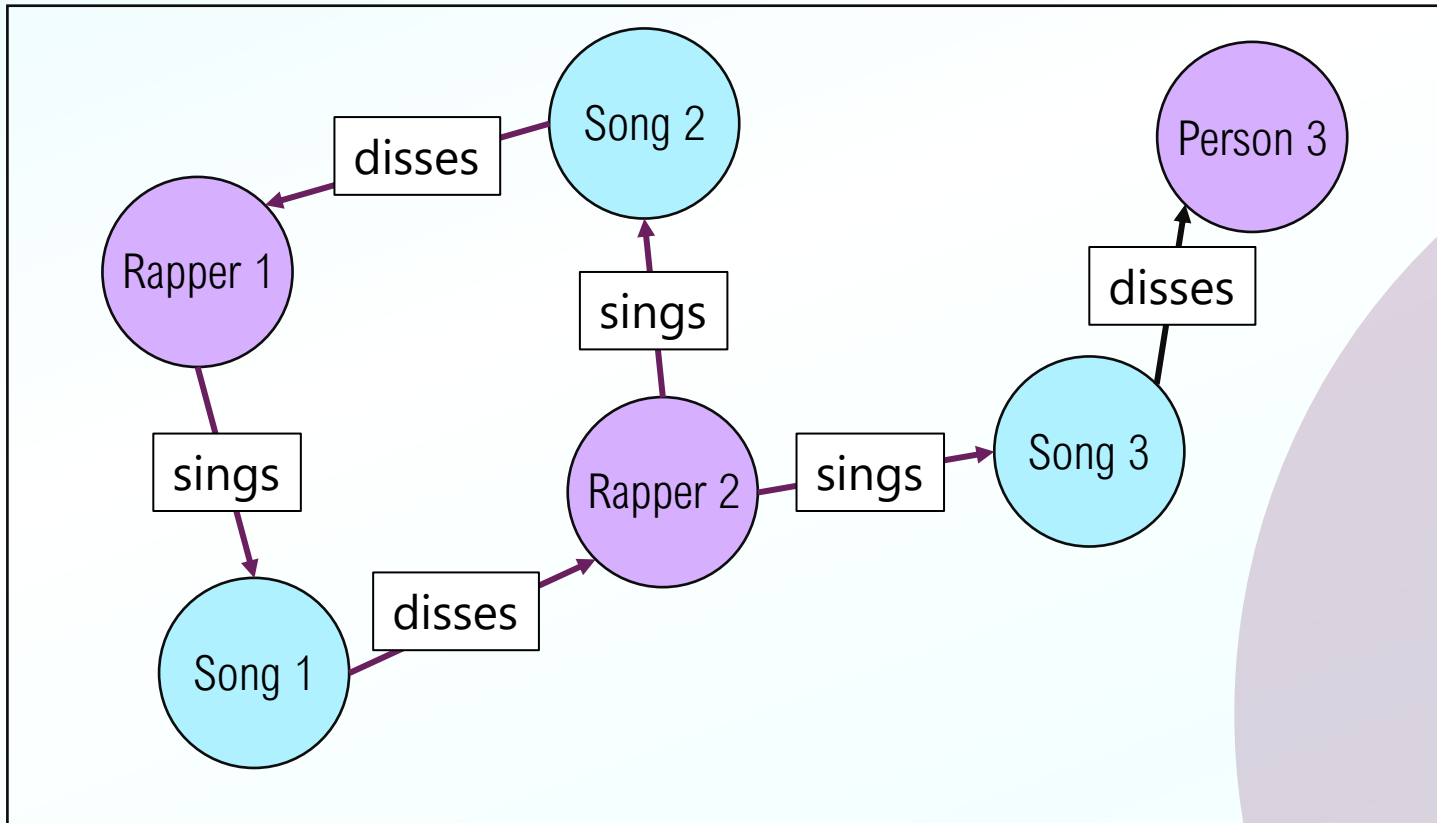
Diss tracks connections

A data pipeline to visualize
connections between diss tracks
and their protagonists

INSA Lyon 5IF
OT 7 : Data Engineering

By Meryem El Halfa and
Emilien Marchet

This is what we want



Questions we want to answer

- Who are those that get dissed the most ?
- Who are the rappers that diss the most ?



This is what we have



1. A Wikipedia table of diss tracks

2. Access to DBPedia

3. Access to Wikidata

4. Potentially other lists from the web



Main page
Contents
Current events
Random article
About Wikipedia
Contact us
Donate

Contribute
Help
Learn to edit
Community portal
Recent changes
Upload file

Tools
What links here
Related changes
Special pages
Permanent link
Page information
Cite this page
Wikidata item

Print/export
Download as PDF
Printable version

Languages
O'zbekcha/Ўзбекча
Edit links

Article Talk

List of diss tracks

From Wikipedia, the free encyclopedia



This article **needs additional citations for verification**. Please help improve this article by adding citations to reliable sources. Unsourced material may be removed and the article may be merged with an existing article. Find sources: "List of diss tracks" – news · newspapers · books · scholar · JSTOR (March 2021) (Learn how and when to remove this template message)

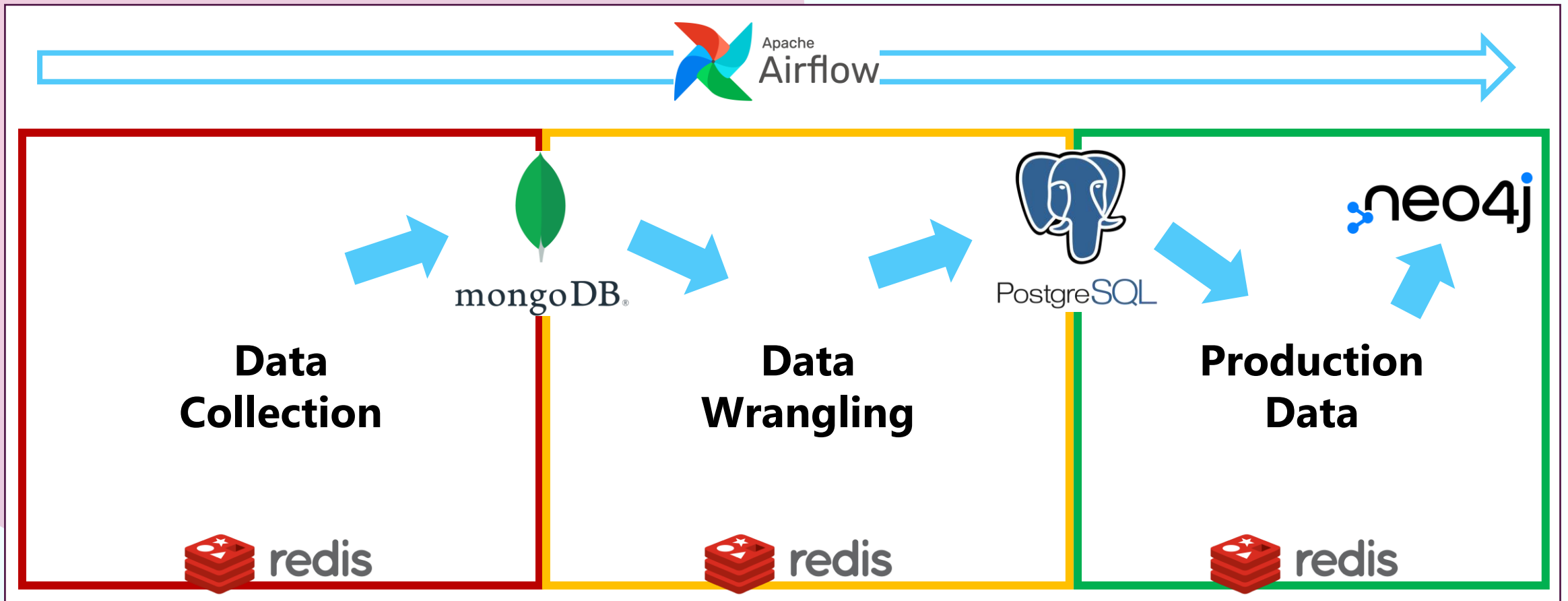
The following is a list of **diss tracks**, songs the primary purpose of which is to verbally attack someone else, usually another artist.

Traditional recordings

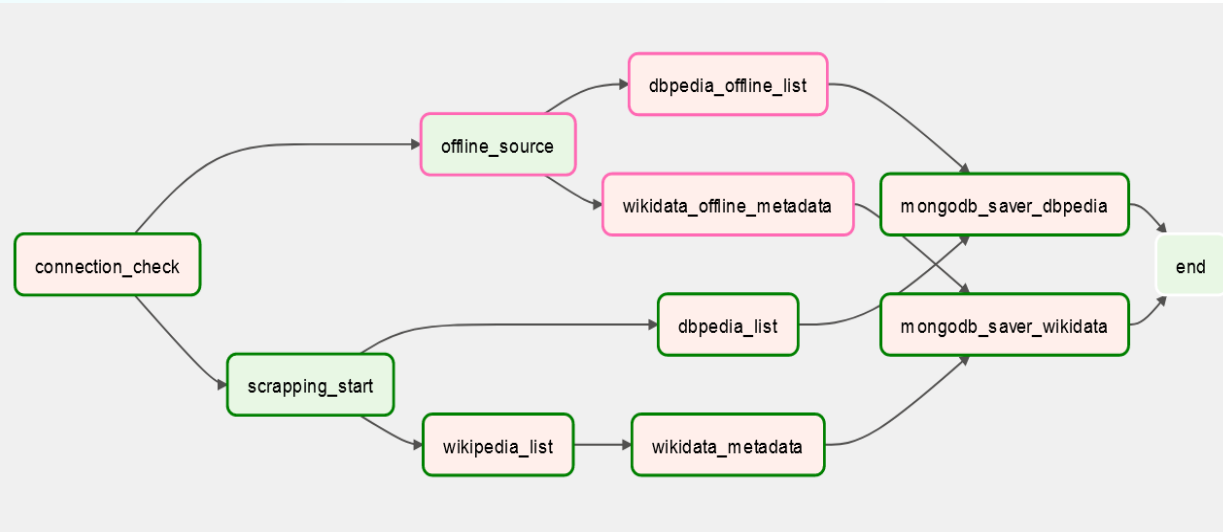
This is a *dynamic list* and may never be able to satisfy particular standards for completeness. You can help by adding missing items with reliable sources.

Date Released	Song Title	Artist(s)	Target(s)	Response to (if applicable)
1754	"Yankee Doodle"	Dr. Richard Schuckberg	George Washington and the colonial "Yankee" troops	American Revolution
1840	"Tippecanoe and Tyler Too"	Whig Party	Martin Van Buren	1840 United States presidential election
1863	"Union Dixie"	Daniel Decatur Emmett	Confederate States of America	American Civil War
1911	"The Preacher and the Slave"	Joe Hill	Salvation Army	In the Sweet By-and-By and Salvation Army recruitment efforts
1938	"La Gota Fría"	Emiliano Zuleta	Lorenzo Morales	
1939	Hitler Has Only Got One Ball	British Army	Adolf Hitler, Hermann Göring, Heinrich Himmler, Joseph Goebbels	
1941	"When That Man Is Dead And Gone"	Al Bowlly	Adolf Hitler, Hermann Göring	The Blitz
1941	"You're a Sap, Mr. Jap"	The Murphy Sisters	Hirohito	The attack on Pearl Harbor
1942	"Njet Molotoff"	Composed by Matti Jurva and was written by Tatu Pekkarinen	Soviet minister for war Vyacheslav Molotov	The Soviet invasion of Finland 1939

Now how do we build our pipeline?



Data collection



What links here
Related changes
Special pages
Permanent link
Page information
Cite this page
Wikidata item

Information [\[edit \]](#)


Ice Cube recorded this song in response to the comments N.W.A made towards him in their album mocking the phrase "hundred miles and running" on "Jackin' for Beats" and ending "I Gotta Say V leads to Ice Cube hanging up on the caller.

The first minute of the song is a reference to N.W.A's "Message to B.A.", in which they call Ice Cube their manager, Jerry Heller.

What we do to populate our Ingestion Database:

- Web scrapping using BeautifulSoup
- Finding Wikidata URI on Wikipedia pages
- Querying Wikidata using SPARQL
- Querying DBPedia using SPARQL
- Add results to MongoDB collections

Data collection



WIKIDATA

Item

Discussion

No Vaseline (Q7045153)

song by Ice Cube

In more languages

Configure

Language	Label	Description	Also known as
English	No Vaseline	song by Ice Cube	
French	No Vaseline	No description defined	
Spanish	No Vaseline	No description defined	
German	No Vaseline	Lied von Ice Cube	

All entered languages

Statements

instance of

musical work/composition

0 references

+ add reference

+ add value

part of

Death Certificate


1 reference

+ add value

form of creative work

song

+ add value



DBpedia

Browse using

Formats

Faceted Browser

About: Diss tracks

An Entity of Type: [Concept](#), from Named Graph: <http://dbpedia.org>, within Data Space: [dbpedia.org](#)

Property	Value
dbpedia:wikiPageID	<ul style="list-style-type: none">47237907 (xsd:integer)
dbpedia:wikiPageRevisionID	<ul style="list-style-type: none">973203124 (xsd:integer)
dbpedia:wikiPageUsesTemplate	<ul style="list-style-type: none">dbt:CatAutoTOC
rdf:type	<ul style="list-style-type: none">skos:Concept
rdfs:label	<ul style="list-style-type: none">Diss tracks (en)
skos:broader	<ul style="list-style-type: none">dbc:Hip_hop_feudsdbc:Song_formsdbc:Songs_about_revengedbc:Works_about_feudsdbc:20th-century_music_genresdbc:21st-century_music_genresdbc:Answer_songs
skos:prefLabel	<ul style="list-style-type: none">Diss tracks (en)
prov:wasDerivedFrom	<ul style="list-style-type: none">wikipedia-en:Category:Diss_tracks?oldid=973203124&ns=14
is dbpedia:wikiPageWikiLink of	<ul style="list-style-type: none">dbc:Always_into_Somethin'dbc:Death_on_Two_Legs_(Dedicated_to_...)dbc:Wah-Wah_(song)dbc:Waitin_for_Youdbc:Wankstadbc:Wanna_Know_(Meek_Mill_song)dbc:War_Paindbc:Diss_(music)dbc:Sexy_Sadiedbc:Shut_Up_(Stormzy_song)

Data collection

data.wikidata_disstracks

```
origin: "traditionally recorded"  
> wikidata_metadata: Object
```

```
_id: ObjectId('637dc495e19bc10f47d8abba')  
Date Released: "1977"  
Song Title: ""Pigs (Three Different Ones)""  
Artist(s): "Pink Floyd"  
Target(s): "Margaret Thatcher and Mary Whitehouse"  
Response to(if applicable): ""  
Notes: "The song's writer, Roger Waters has stated that the second and third v..."  
Ref(s): "[17]"  
Wikipedia endpoint: "/wiki/Pigs_(Three_Different_Ones)"  
Wikidata target id: "Q7416"  
Wikidata song id: "Q636247"  
origin: "traditionally recorded"  
> wikidata_metadata: Object
```

```
_id: ObjectId('637dc495e19bc10f47d8abbb')  
Date Released: ""Idiot Box""  
Song Title: "the Damned"  
Artist(s): "Television and Tom Verlaine"  
Target(s): "Television refusing to allow the Damned to play a show with the"  
Response to(if applicable): ""  
Notes: "[18]"
```

data.dbpedia_disstracks

```
_id: ObjectId('637dc2b4456fa7c3cedee02c')  
> diss: Object  
v name: Object  
  type: "literal"  
  xml:lang: "en"  
  value: "Give It to Me"  
> genre: Object  
> recorded: Object  
> released: Object  
> recordLabel: Object
```

```
_id: ObjectId('637dc2b4456fa7c3cedee02f')  
> diss: Object  
v name: Object  
  type: "literal"  
  xml:lang: "en"  
  value: "Piggy Bank"  
> genre: Object  
> recorded: Object  
> released: Object  
> recordLabel: Object
```



mongoDB®

Data wrangling

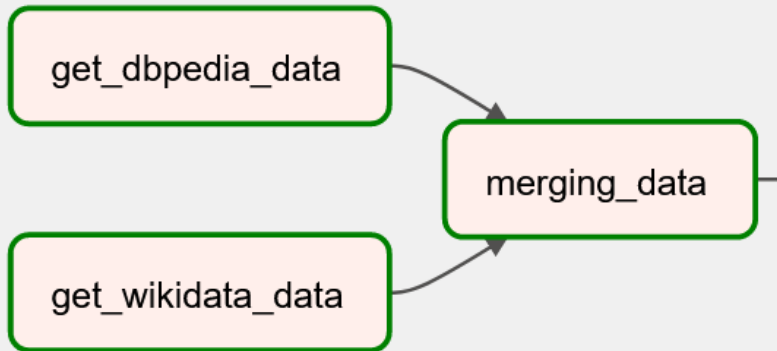
What we do in this dag: (in a nutshell)

- Merging our sources
- Removing unimportant data
- Cleaning some columns
- Building additional standardized columns
- Making new Wikidata requests
- Saving to Postgres tables



Data wrangling

Data merging



What we do in this step:

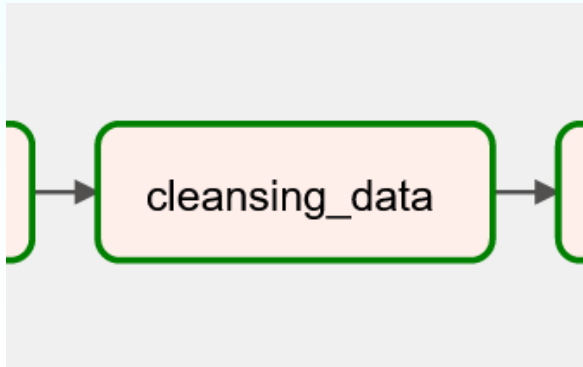
Merging our sources
(Wikipedia/Wikidata and DBPedia)
using a common identifier

Here our identifier is basically the track
title

```
merged_df = pd.merge(wikidata_df,  
dbpedia_df, how="outer", on=["Song  
Title"])
```

Data wrangling

Data cleansing



What we do in this step:

Dropping unimportant columns like recordLabel, url, Wikipedia references...

Building custom columns: Target type and Song Artist

Combining data from our different sources: Artist and Released Date

Data wrangling

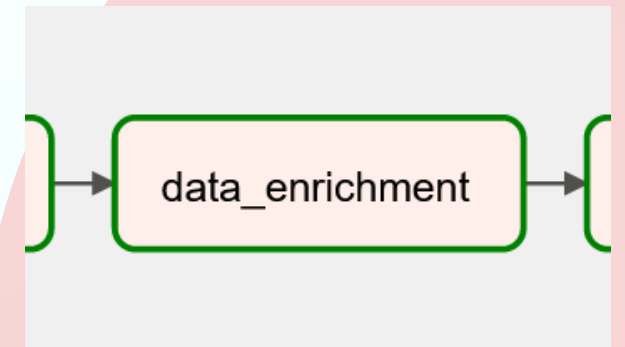
What we do in this step:

Querying additional target information from Wikidata

Querying different information based on target type (person or group)

Storing information on persons and groups in two different DataFrames

Data enrichment



Data wrangling

Example of `postgres.songs`

index	Song Title	Target(s)	Response to(if applicable)	
50	L.A., L.A.	Tha Dogg Pound, Bone Thugs-N-Harmony	"New York, New York"	Q10
115	Chipmunk Reply	Chipmunk, Skepta, Jammer, Lil Wayne	"Pepper Riddim" by Chipmunk	Q10
172	YouTube Stars Diss Track	The general press		Q11
68	Letter To The President	Bill Clinton		Q11
22	Roxanne's Revenge	U.T.F.O.		Q11
146	Intocable (Tiradera pa' Cosculluela)	Cosculluela		Q11
35	Fuck Compton	DJ Quik, Michel'le, N.W.A, Compton's Most Wanted	East Coast–West Coast hip hop rivalry	Q11
42	Def Wish II	DJ Quik	"Tha Last Word" By DJ Quik	Q11
1	Tippecanoe and Tyler Too	Martin Van Buren	1840 United States presidential election	Q11
78	Cry Me a River	Britney Spears		Q11
10	Happy Family	The Beatles	Their break up	Q12
13	How Do You Sleep?	former Beatles bandmate Paul McCartney	Personal slights Lennon felt McCartney had	Q12
24	ohn Fogert	Fogerty's long-standing financial problems	The song was altered and re-titled "John Fogerty"	Q13
8	Njet Molotoff	Soviet minister for war Vyacheslav Molotov	The Soviet invasion of Finland 1939	Q13
144	Not Alike	Machine Gun Kelly	Comments Kelly made about Eminem	Q13
147	Killshot	Machine Gun Kelly, P. Diddy	Machine Gun Kelly's "Rap Devil"	Q13
66	The Agony of Laffitte" / "Laffitte"	Ron Laffitte and Sylvia Rhone		Q13
96	he Gam	"No More Games", "G-Unot Killa"		Q13
31	Miracle Man	Televangelist Jimmy Swaggart	Swaggart's various controversies	Q13

song	
id	SERIAL
title	CHARACTER VARYING
release_date	TIMESTAMP(6) WITHOUT TIME ZONE
record_date	TIMESTAMP(6) WITHOUT TIME ZONE
genre	CHARACTER VARYING
wikidata_id	CHARACTER VARYING
artist_id	INTEGER
target_id	INTEGER

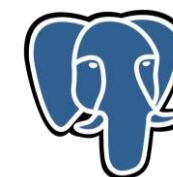
saving_to_postgres

human	
id	INTEGER
occupation	CHARACTER VARYING
first_name	CHARACTER VARYING
last_name	CHARACTER VARYING
birth_place	CHARACTER VARYING

entity	
id	SERIAL
name	CHARACTER VARYING
type	CHARACTER VARYING(50)
wikidata_id	CHARACTER VARYING

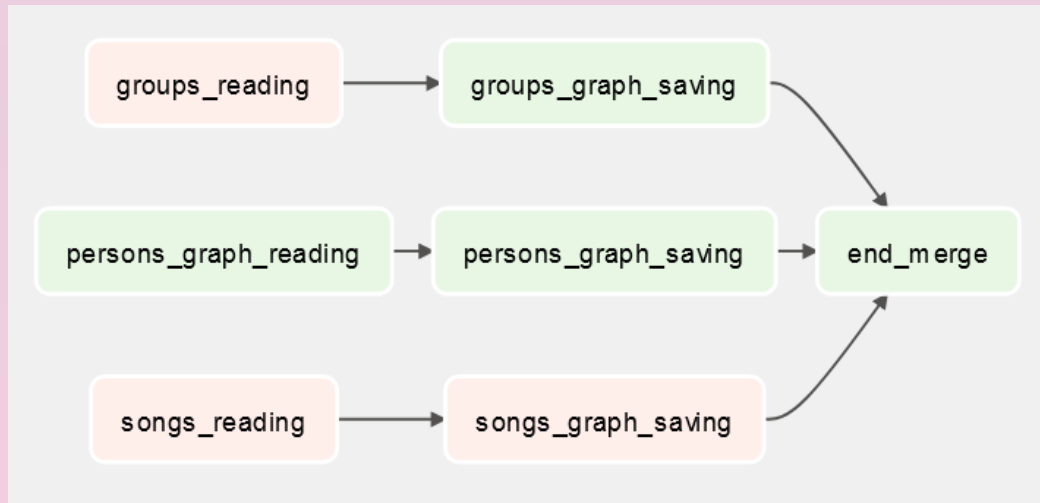
group	
id	INTEGER
country	CHARACTER VARYING
nb_nominations	INTEGER
inception	TIMESTAMP(6) WITHOUT TIME ZONE

other	
id	INTEGER



PostgreSQL

Production data



What we do in this dag:

Build a Neo4j graph based on the data we have

Create a node for each person, group and song

Create the links between persons/groups and songs

➡ Thanks to Neo4j we then have a graph representation of our data, that we can query to answer our initial questions



Thank you

Do you have questions ?



Meryem and Emilien