



ARTIST INVESTMENT RESEARCH WITH SPOTIFY

RESEARCH QUESTIONS

How to find a growing artist to invest on?

Where are the most interesting markets to promote the artist?



WHAT IS THE PROCESS LIKE?

SECOND TASK

Data storage & distribution

VISUALIZATIONS

Interactive tool & data visualization

FIRST TASK

Scraper & data acquisition

ANALYTICS

Data cleaning &
search for insights

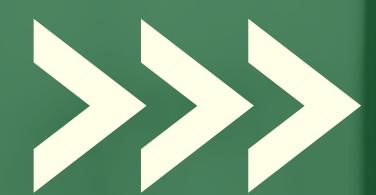
CONCLUSIONS

Commenting the achieved
results

FIRST TASK - VARIETY



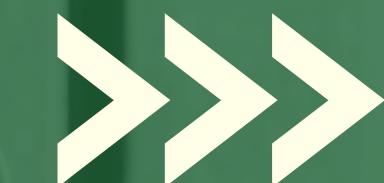
SCRAPER



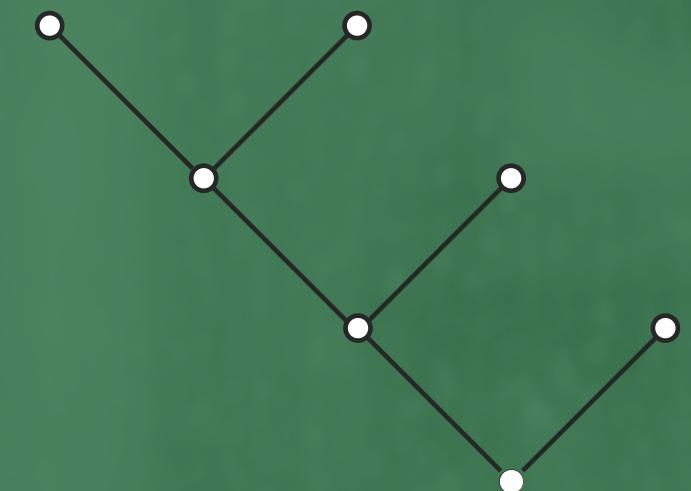
Spotify

Wikipedia

Genius



Composite
Integration



SECOND TASK - VOLUME

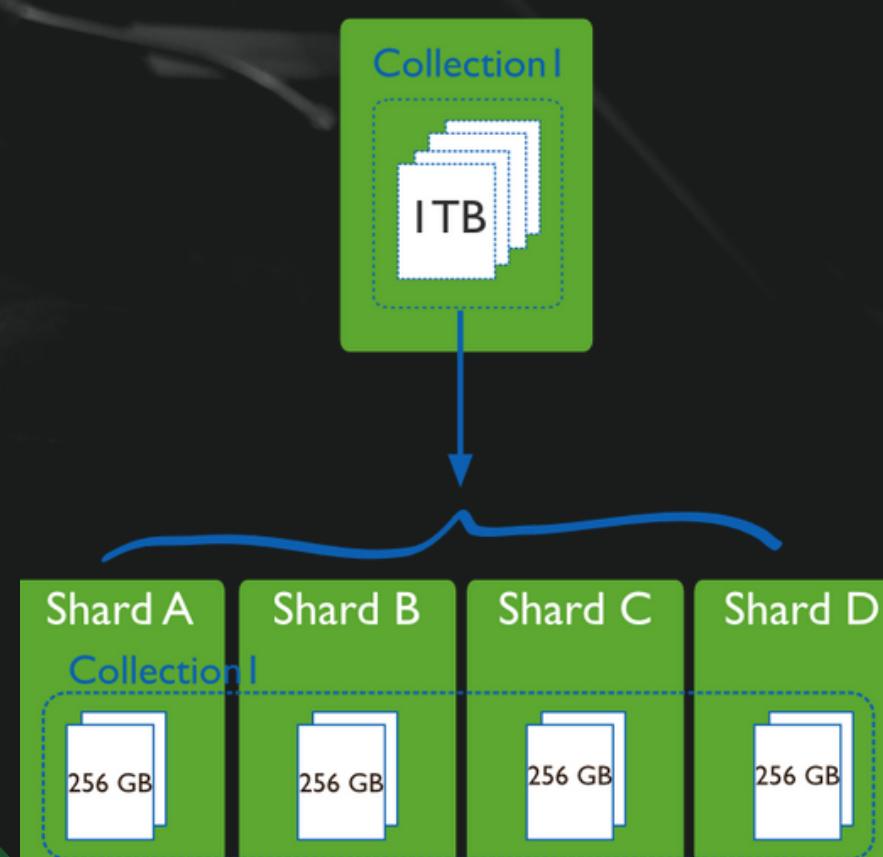
Storage -
MongoDB

Frankfurt Data
Center

Sharding -
Horizontal Scaling



MongoDB



MATCHING PROBLEMS

Names didn't match at first between Spotify and Wikipedia.

Solution = Edit distance

ANAGRAPHIC INFORMATION

Mainly the age field required correction for past artists and groups.

MUSIC GENRE

The granularity in this field required a lot of cleaning.

ABNORMAL LYRICS

The presence of podcasts required the lyrics to be cleaned.

LOTS OF NULL

All the cleaning couldn't solve the high percentage of null in some fields.

ANALYTICS - DATA CLEANING

ANALYTICS - INSIGHTS

GENDER INEQUALITY

Women generate far less streams than men.

GRANULARITY OF THE GENRES

Genres granularity changes a lot, for example pop has a low granularity while rock is diversified in many subgenres.

CHANGE IN MOST PRODUCED GENRE

Most of the pop music was produced in the last decade, infact previous to that was rock the dominating genre.

SIMILAR CULTURES LISTENS TO SIMILAR MUSIC

R&B is listened mainly in America, Northern Europe, and Australia.

Without music, life would be a mistake.

FRIEDRICH NIETZSCHE

Visualizations plus Conclusions can be found on Tableau at this link:

<https://public.tableau.com/profile/davide.b#/>