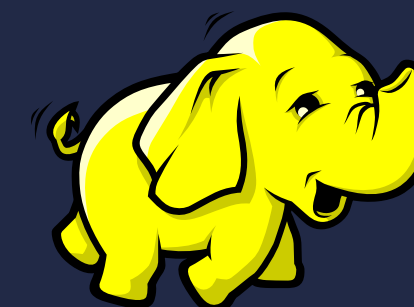


Find Next Hit: Your Playlist from million songs.

Team 1 - Bunyod Suvonov, Meng Zi, Pan Haoyun, Wang Zhiyuan



JOINT INSTITUTE
交大密西根学院

Reference

ECE4720J Team01 Su23. Win Or Lose?

Music speaks louder than words!

These two QR codes are *Soul Deep* and its recommended song *True Lovers*. Try these two country music! Really similar, right?

Soul Deep - The Box Tops



用手机扫一扫,了解更多信息

True Lovers - Ruby Hunter



用手机扫一扫,了解更多信息

MapReduce v.s. PySpark

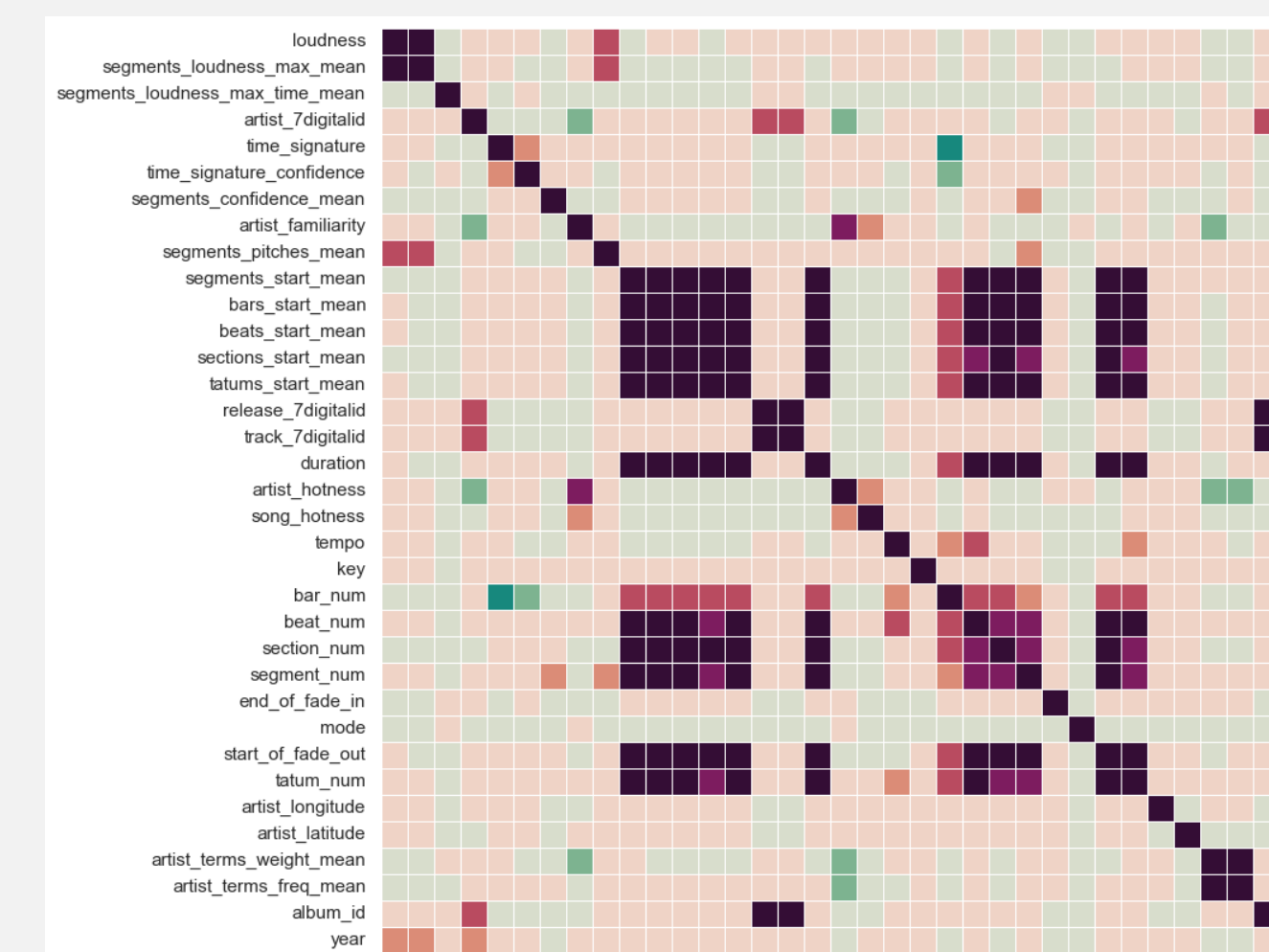
PySpark performs more than **twice** as fast as MapReduce!

Performance summary:

- **MapReduce**: 77.905s
- **Pyspark**: 28.160s

Million in One

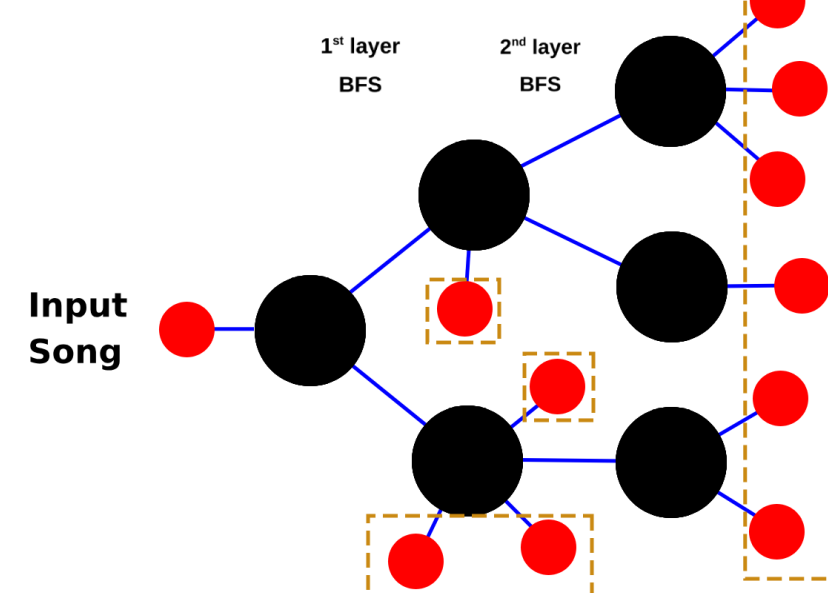
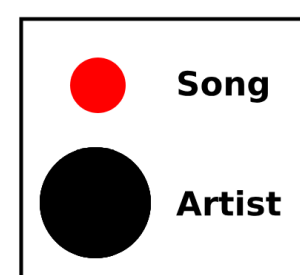
We decoded one million of *.h5* files for your best experience, and we use *.avro* to store them for high performance. At the very beginning, we can use **pandas** library to read data and here is some result from EDA:



So, we found those variables with correlationship to year:

- loudness
- segments_loudness_max_mean
- artist_7digitalid
- album_id
- release_7digitalid
- track_7digitalid
- time_signature_confidence

Big Data Recommendation



Or an alternative ML strategy!

2 cluster algorithm:

- **K-means**
- **DBSCAN**

For algorithm demonstration, check this:
<https://www.naftaliharris.com/blog/>

Distance based Recommendation using Parallel BFS

- Process every BFS level in **Parallel**
- Suggest the hottest song discovered

Result:

| title | album_name | Cos_correlation_factor |
|----------------|-----------------|------------------------|
| b'Soul Deep' | b'Dimensions' | 1.000000 |
| b'True Lovers' | b'Feeling Good' | 0.999996 |

Nostalgia! Year Prediction of a Song

Setup: **PCA**, a dimension reduction technique; **Linear Regression**

On **Spark**: in-memory processing, able to handle the large dataset efficiently (4.49s)

MSE: 97.35 High quality! Can predict the year of a song within 10 years.

Interesting finding: As the years increase, people tend to give songs shorter names.

Trivia you might want to know...

Our platform maintain an SQL query service on **Drill**, so if you are curious with certain tag, artist or year, you can smoothly find and filter them!

Did you know...

- ... that the age of the songs covered from 13 years to 102 years?
- ... that Jingle Bell Rock is the hottest song that is the shortest and shows highest energy with lowest tempo?
- ... that First Time In A Long Time: The Reprise Recordings is the album with most tracks (85 tracks)?
- ... that the band *Mystic Revelation of Rastafari* has recorded *Grounation* with the highest duration?
- ... that In Sydney With Bob Rogers is the hottest song from the Beatles?