University of Tours

B.D.M.A

Big Data Management Analytics

Report Phase 1 Decisional Project

Author:
Akaichi, Ines
Cissé, Ismaila
de Saint Ceran, Louis
F. Nascimento Filho, Jessé

September 22, 2018

Contents

1 Introduction

Not long ago, there was a strong belief that the internet was killing the music industry. For years now, the industry has failed to keep up with the rapid pace of technological advancement and barely had any understanding of their audience, who was buying their CDs, or cassettes.[9]

With streaming services taking over, these companies found ways to track the listening habits of users and have access to detailed information such as when, how, where, and who is listening to what.

The aim of the industry, now, is to use these customer behavior insights together with knowledge of the music itself, which is made possible only with Big Data. The raw music that is produced is essentially like unstructured data. In the digital era, this raw music can be easily digitized and analyzed.

As part of our master degree 's program , we choose to work on the design and implementation of a business intelligence system[10] supporting the analysis of songs. Our objective is to analyze the current music streaming industry and more precisely the songs and the artists' popularity. We decided to bring the subject closer to us by focusing on French artists.

2 Management Methodology

Our team is a small multicultural and autonomous group, for this reason we found a common equation to process with ours tasks project life cycle. The result from this equation become our management methodology that is based in a lean solid and agile solution called Scrum. For this purpose we decided to follow the best practices of Scrum combined with a Kanban approach.

2.1 Scrum

The Scrum arise as a process framework to manage complex projects since the early 1990s. The essence of Scrum is to provide effective iterations and incremental knowledge transfer to the success of a project [11].

Our Project will be composed of 4 sprints during this semester:

- Sprint 0 — Study : The list of requirements and project planning;

- Sprint 1 — Model : Conception & modeling of data warehouse;
- Sprint 2 — ETL : Data Extraction, Transformation and Loading;
- Sprint 3 — DEMO : B.I system demonstration of an initial version;
- Sprint 4 — DEFENSE : Oral project presentation;

2.1.1 Schedule

[1]

2.1.2 Sprint 0

The name SPRINT 0 has been learned to describe the preparation phase which precedes the launching of the project. The term SPRINT 0 is being simpler to use than the preparation or inception phase, it is increasingly used in SCRUM projects. Sprint 0 does not diminish the flexibility of our project. On the contrary, it will allow us to anticipate certain actions and have an overview that will facilitate the management of changes that will emerge at the following sprints. [8]

In this Sprint we will be able to :

1. Share a clear vision of the project;
2. Identify users need;
3. Identify the preliminary workload resulting of the users need;
4. Prepare the project management plan;

2.1.3 Sprint 1

the preliminary specification of the workload in sprint 0 will help us in this sprint in modeling our data warehouse,thus the formalization of the entire workload. In addition, in this phase it is essential to maintain an active technological watch to choose our Essential BI tools used in next sprints.

2.1.4 Sprint 2

in this sprint we will be able to define our data warehouse 's architecture, assess the data quality and implement the designed ETL system .

2.1.5 Sprint 3

in this sprint we will be able to visualize our data using the BI restitution tools .

2.1.6 Sprint 4

in this final sprint , We will be able to prepare an oral presentation where we summarize all the steps that we have gone through when developing our project 's data warehouse and present our work to our professors .

2.2 Kanban

In addition of Scrum methodology we choose to use Kanban approach, that means "visual card" in Japanese, to help us simplify the sprints workload. We going to make a visual work-flow using Trello for create and manage all cards with micro tasks, it will result in each sprints deliveries milestones.

2.2.1 Trello

Trello[7] is a project management software that utilizes the concept of boards to represent projects and within boards, cards to represent tasks. Trello supports Team Collaboration enabling members to discuss a project in real-time. It keeps everybody informed through task assignments, activity log, and e-mail notifications.[3]

3 Preliminary Workload

3.1 Data Sources

Open Data sources:
MusicBrainz Database And Schema[5]
Lastfm API [4]

Limited Data Sources:
echo Nest API [2]
Spotify API [6]

3.2 User Needs

- General statistics about songs, disaggregating by genre.
- General statistics about artists, disaggregating by year.
- Number of songs by year, label and location.
- Number of songs or artists that achieved a certain rating.
- Performance of artists (average rating of the songs where they participated).
- The must popularity type of song in France for doing physical activity.(statistics about fitness songs of playlist)
- General statistics about tempo
- What is the best revenue by genre? General statistics about market
- What is the unless popular songs by region of France?
- General statistics (global rating) about songs produced in each region of France.
- What is the impact of cover art on success of album?
- Artist that most engaged of the public long of the last years.
- General statistics by genre
- General statistics
- What makes a tube
- Statistics about top performer

References

[1] Coolors pattern. https://coolors.co/1a535c-20a39e-f7fff7-ff6b6b-ffe66d. accessed: 21.09.2018.

[2] echo nest. http://the.echonest.com/. accessed: 21.09.2018.

[3] Kanban. https://www.tutorialspoint.com/kanban/kanban_tutorial.pdf. accessed: 21.09.2018.

[4] Lastfm api. https://labrosa.ee.columbia.edu/millionsong/lastfm#api. accessed: 21.09.2018.

[5] Musicbrainz database and schema. https://musicbrainz.org/doc/MusicBrainz_Database/Schema. accessed: 21.09.2018.

[6] Spotify api. https://developer.spotify.com/documentation/web-api/. accessed: 21.09.2018.

[7] Trello. https://trello.com/. accessed: 21.09.2018.

[8] Christian DESTREMAU. Méthode scrum partie 2 : définition de la méthode. https://www.supinfo.com/articles/single/6054-methode-scrum-partie-2-definition-methode. accessed: 21.09.2018.

[9] Avantika Monnappa. Predicting the next big hit - big data and the music industry. https://www.simplilearn.com/big-data-science-in-music-industry-article. accessed: 21.09.2018.

[10] L.T. Moss and S. Atre. *Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications*. Addison-Wesley Information Technology Series. Pearson Education, 2003.

[11] Ken Schwaber and Jeff Sutherland. The scrum guide™ : The definitive guide to scrum: The rules of the game. https://www.scrumguides.org/docs/scrumguide/v2017/2017-Scrum-Guide-US.pdf. accessed: 21.09.2018.