

Facultade de Informática



UNIVERSIDADE DA CORUÑA

TRABALLO FIN DE GRAO
GRAO EN CIENCIA E ENXEÑARÍA DE DATOS

Perfil completo de artistas musicais: integrando información de Spotify e MusicBrainz

Estudante: Clara Lado Río

Dirección: Fernando Silva Coira

Hilda Romero Velo

A Coruña, xuño de 2024.

À miãa familia, por estar sempre.

Agradecementos

Gracias a todos os amigos e compañeiros que me acompañaron durante estes anos de camiño, sen vós non sería posible. Á miña familia, por darme o apoio necesario día tras día para poder chegar ata aquí. Tamén quero agradecer aos meus dous titores imprescindibles para o desenvolvemento do proxecto, Fernando e Hilda, por facer isto realidade, polo tempo e paciencia adicados. Por último, a Tirso pola confianza depositada.

Resumo

Neste traballo de fin de grao desenvólvese unha ferramenta que leva a cabo a integración de dúas fontes de datos da industria musical: Spotify, ofrece para os artistas a súa discografía completa e unha serie de atributos de valor e MusicBrainz, enciclopedia que complementa a información anterior. A mestura de ambas permite ter un conxunto de datos amplo cunha visión completa do panorama musical. A través de procesos ETL extráese, prepárase e cárgase nun almacén de datos (DW), deseñado en concreto para que se reflicta de maneira eficiente e acorde ca realidade a información correspondente. Para simplificar o proceso de inserción ao usuario implementase unha simple interface. Por último, lévase a cabo unha serie de análises e mostras de información, empregando gráficos, filtros e dashboards interactivos que son de axuda para identificar tendencias ou curiosidades nos datos, permitindo que o coñecemento sexa máis accesible para os usuarios.

Abstract

In this final degree project, a tool that integrates two data sources from the music industry is developed: Spotify which offers the complete discography of artists along with a series of interesting attributes, and MusicBrainz, an encyclopedia that complements the previous information. The combination of both provides a comprehensive dataset with a complete view of the musical landscape. Through ETL (Extract, Transform, Load) processes, data is extracted, prepared, and loaded into a Data Warehouse, specifically designed to efficiently and accurately reflect the corresponding information. To simplify the insertion process for the user, a simple interface is implemented. Finally, a series of analyses and data presentations are done using interactive graphs, filters, and dashboards, in order to help identifying trends or curiosities in the data, making knowledge more accessible to users.

Palabras chave:

- Tableau
- Almacén de Datos
- ETL
- Análise de Datos
- Industria Musical
- Spotify
- MusicBrainz

Keywords:

- Tableau
- Data Warehouse
- ETL
- Data Analysis
- Music Industry
- Spotify
- MusicBrainz

Índice Xeral

1	Introdución	1
1.1	Motivación	1
1.2	Obxectivos	3
1.3	Estrutura da memoria	4
2	Fundamentos tecnolóxicos	5
2.1	Que é un Data Warehouse?	5
2.1.1	Características	5
2.2	Que é un proceso ETL?	6
2.3	Que é unha API?	7
3	Ferramentas e tecnoloxías empregadas	9
3.1	Python	9
3.2	PostgreSQL	10
3.3	Dbeaver	10
3.4	Tableau	11

4	Metodoloxía e planificación	12
4.1	Metodoloxía	12
4.1.1	Metodoloxía Kimball	13
4.2	Planificación do proxecto	14
5	Deseño do Data WareHouse	20
5.1	Definición de requisitos	20
5.2	Deseño do modelo dimensional	22
5.2.1	Escoller os procesos de negocio	22
5.2.2	Determinar o nivel de detalle e granulado	23
5.2.3	Determinar entidades	23
5.3	Deseño do modelo lóxico	25
6	Proceso ETL	31
6.1	Fontes de datos	32
6.2	Extracción	32
6.2.1	APIs e librerías	34
6.3	Transformación	37
6.3.1	Datas	38
6.3.2	Integrantes:	38
6.3.3	Nomes	39
6.4	Carga	39
6.4.1	Táboas	40

6.4.2	Inserción	41
6.4.3	Pasos	43
7	Explotación do Data Warehouse	51
7.1	Grupo	52
7.2	Artista	54
7.3	Especificacións <i>tracks</i> dun álbum	56
7.4	Colaboracións dun artista	57
7.5	Instrumentos	59
8	Conclusións	63
8.1	Traballo futuro	64
A	Material adicional	67
A.1	Adición de intérpretes	67
A.2	Borrador deseño dimensional	68
	Relación de Acrónimos	71
	Glosario	72
	Bibliografía	73

Índice de Figuras

2.1	Fases procesos ETL	8
4.1	Diagrama de Gantt: Estimación inicial	15
4.2	Diagrama de Gantt: Evolución real	18
5.1	Diagrama Entidade Relación (ER)	26
5.2	Diagrama relacional	30
6.1	Extracción e carga da información na base de datos	50
7.1	Fonte de datos de Tableau	52
7.2	Dashboard con información relativa dun grupo e os seus integrantes	54
7.3	<i>Dashboard</i> que engloba información dun artista	58
7.4	<i>Dashboard</i> con conxunto de gráficos informativos dunha track	58
7.5	Dashboard coas colaboracións dun artista	60
7.6	Dashboard coas colaboracións específicas de Bad Bunny con Nicky Jam	60
7.7	<i>Dashboard</i> con conxunto de gráficos informativos dunha track	62

A.1	Interface para a introducción de artistas	69
A.2	Mensaxe tras a correcta introdución	69
A.3	Aproximación inicial do diagrama ER	70

Índice de Táboas

4.1	Planificación e coste do proxecto con retrasos	19
-----	--	----

Introdución

1.1 Motivación

Na actualidade xérase unha grande cantidade de datos diariamente, dos que se pode sacar información de utilidade, tanto con obxectivos económicos como para satisfacer a curiosidade persoal e revelar datos de interese. Esta relevancia é tal que se fixo imprescindible contar con ferramentas e procesos que permitan a súa eficiente xestión. Un exemplo destas son os [Data WareHouse \(DW\)](#), almacéns de datos que permiten realizar a integración e análise de grandes volumes de información obtidos de diferentes fontes. Xunto con eles, cabe destacar a necesidade dos procesos [Extract Transform Load \(ETL\)](#) (extracción, limpeza, transformación e carga), cuxo principal obxectivo é adquirir os datos e preparalos para as necesidades fundamentais co fin do correcto desenvolvemento da tarefa de análise.

A análise de datos tamén xoga un papel crucial na toma de decisións informadas. Permite identificar patróns, tendencias, prever comportamentos futuros e optimizar procesos en múltiples ámbitos. No sector empresarial, por exemplo, as empresas utilizan a análise para entender mellor as necesidades dos seus clientes, ofrecer os produtos e servizos adecuados, mellorar as súas estratexias de mercado mediante [Intelixencia de Negocio \(BI\)](#) ou xestionar recursos de xeito máis eficiente.

A música está máis de moda ca nunca e, a industria musical, ao igual co resto da sociedade, experimentou unha evolución notable ó longo do tempo, a cal se acentuou ca chegada das novas tecnoloxías. A proliferación das plataformas de *streaming* veñen transformando de xeito significativo as maneiras de produci-la, distribuí-la e consumí-la. Plataformas como Spotify [1], Apple Music [2] ou Amazon Music [3], que xeran datos útiles sobre as preferencias e

comportamentos dos usuarios, permiten acceder a unha xigante biblioteca musical de forma instantánea, cando e onde se queira.

Por outra banda, a información dispoñible a través de internet no relativo á música é variada. A través del podemos achegarnos á enciclopedia en liña [MusicBrainz \(MB\)](#) [4], a cal aporta unha serie de coñecemento e datos detallados acerca de todo o panorama musical, dende álbums, gravacións, instrumentos, grupos, integrantes ata produtores, entre outros.

Ambas plataformas, Spotify e MB céntranse nun obxectivo concreto, a primeira na distribución e consumo e a segunda na información. A integración de ambas fontes permite a creación de perfís completos. Este tipo de integración non só enriquece a calidade da información dispoñible, senón que tamén abre a porta a novas oportunidades para o desenvolvemento de aplicacións innovadoras. Por exemplo, unificar datos de consumo de música con información detallada sobre os artistas pode permitir a creación de novas ferramentas de descubrimento musical, aplicacións educativas sobre a historia da música e análises máis profundas das tendencias musicais globais.

En conclusión, ao extraer a información de interese, integrarse e completarse os datos contidos en ambas, pódese chegar a facer unha [Bases de datos](#) completa dos artistas e as súas obras. Mediante a explotación desta, empregando gráficos analíticos é posible obter resultados, revelar detalles e proporcionar perspectivas valiosas, facilitando unha comprensión máis profunda dos datos e das súas implicacións abstraéndoo da complexidade da parte máis analítica. Algunhas das consultas posibles son:

- **Perfís detallados dos artistas:** crear perfís completos que inclúan tanto a súa popularidade, orixe como información detallada sobre as súas obras.
- **Análise das tendencias da música:** descubrir cales son os artistas e xéneros máis populares ou os instrumentos máis comúns.
- **Recoñecemento da produción musical:** identificar as tarefas de produción nas cancións e destacar a todos os traballadores involucrados na súa creación.
- **Evolucións ao longo do tempo:** comparar os valores de determinadas medidas como a acústica das cancións ou reconto de publicacións a través dos anos.

1.2 Obxectivos

Neste traballo recompilarase información concreta que abrangue dende detalles biográficos dos intérpretes ata as súas obras musicais, álbums, colaboracións e participantes nas creacións das cancións. Este enfoque permite obter unha visión global detallada dos artistas solicitados e da súa traxectoria musical, ofrecendo unha ferramenta útil tanto para os profesionais da industria como para o público xeral interesado en coñecer un pouco máis dos seus artistas favoritos. O obxectivo principal é ofrecer unha ferramenta que permita unha visión completa e integrada da información musical, algo esencial neste mundo cada vez máis dixital e interconectado.

Os pasos seguidos ó longo do proxecto para levar a cabo os obxectivos descritos son os seguintes:

- Definición dos requisitos necesarios para realizar a análise, identificación das necesidades, obxectivos específicos e criterios adecuados.
- Deseño e implementación da **BBDD** onde se almacenará a información, cunha estrutura eficiente que permita unha xestión óptima.
- Análise das distintas fontes de datos e selección de atributos de interese, mediante a avaliación de relevancia e calidade dos datos dispoñibles para garantir a precisión e fiabilidade das análises.
- Acceso á información a través da interacción coas **Application Programming Interface (API)** das distintas bases de datos, Spotify e **MB**.
- Deseño e creación dos procesos **ETL** para os distintos conxuntos de datos para cargar a información na **BBDD**, asegurando a integración adecuada dos datos e a súa transformación segundo os requisitos.
- Execucións de análises e creación de cadros de mando para visualizar os resultados obtidos, que permitan chegar a facer interpretacións.
- Observación dos resultados obtidos.

1.3 Estructura da memoria

A memoria está organizada en varios capítulos que detallan os pasos necesarios para a elaboración e entendemento do proxecto, dende a introdución ata as conclusións finais. A división é a seguinte:

- **Capítulo 1 - Introducción:** presenta o traballo a realizar, fundamentándose nas razóns polas que se decidiu levar a cabo e os seus obxectivos. Ofrece unha visión xeral para que o lector saiba que esperar nos seguintes capítulos.
- **Capítulo 2 - Fundamentos tecnolóxicos:** aborda os conceptos teóricos necesarios para asentar unha base de coñecemento e poder comprender o proxecto. Nel defínese que é un Data Warehouse, cales son as súas características principais, o proceso [ETL](#) e as [API](#).
- **Capítulo 3 - Ferramentas e tecnoloxías empregadas ao longo do proxecto:** describe as ferramentas e tecnoloxías utilizadas. Explicase Python, a linguaxe de programación principal empregada para enlazar os procesos, PostgreSQL e Dbeaver para a creación e xestión da base de datos e, por último, Tableau como plataforma utilizada na visualización en análise de datos.
- **Capítulo 4 - Metodoloxía e planificación:** detalla a metodoloxía seguida para o desenvolvemento do proxecto, concretamente unha baseada en Kimball, os pasos levados a cabo e o cronograma da planificación.
- **Capítulo 5 - Deseño do Data Warehouse:** céntrase na definición do Data Warehouse e os requisitos para o seu desenvolvemento.
- **Capítulo 6 - Proceso ETL:** describe o proceso levado a cabo para a obtención e xestión dos datos. Detalla as fontes de datos, o proceso de extracción e as librerías empregadas. A continuación, menciónanse as transformacións aplicadas e, por último, nárrese o procedemento que combina as diferentes etapas.
- **Capítulo 7 - Explotación do Data Warehouse:** explícanse as análises realizadas a partir dos datos almacenados utilizando gráficos e [dashboard](#) interactivos. Nel detállanse cada unha das vistas.
- **Capítulo 8 - Conclusións:** ofrece unha reflexión xeral do proxecto.
- **Apéndices A.1 A.2:** inclúe información da interface e un borrador inicial do deseño.

Fundamentos tecnolóxicos

Neste capítulo dótase dunha base de coñecemento sobre as necesidades dos procesos e sistemas de **BDD** cos que se traballará ao longo do proxecto.

2.1 Que é un Data Warehouse?

Os datos convertéronse nun pilar fundamental para a maioría das empresas actuais, cobrando gran importancia a xestión dos mesmos, así como a calidade, capacidade de obtelos e almacenalos dunha maneira eficiente. Así, o aproveitamento da información contida neles é crucial para tomar decisións clave que permitan optimizar o funcionamento da empresa [5].

Neste contexto, os **DW**, tamén chamados almacéns de datos, son un tipo de sistema para xestionar grandes volumes de datos e dar soporte ás tarefas de intelixencia empresarial, especialmente as analíticas. Trátanse dun entorno de integración de datos que combinan unha serie de tecnoloxías e compoñentes para almacenalos, consúltalos e analízalos, recollendo así a información valiosa oculta neles. Habitualmente, a información que se pretende recompilar atópase dispersa nunha ampla gama de fontes de distintos formatos, a cal debe de ser extraída mediante un elaborado proceso **ETL** [6] [7].

2.1.1 Características

Os almacéns de datos permiten realizar unha análise axeitada sobre os datos, fundaméntanse nas seguintes características [8] [5]:

- Orientados a suxeitos: poden analizarse os datos sobre un tema particular ou área. A información está organizada en torno a suxeitos de interese como poden ser as cancións, os artistas ou as relacións.
- Integrados: os DW xeran uniformidade entre diferentes tipos de datos de fontes heteroxéneas. Unha vez que os datos son cargados no almacén, non se modifican nin se eliminan.
- Administrar grandes cantidades de información: organiza grandes cantidades de información de xeito compacto e eficiente.
- Non volátiles: a información contida é estable e non cambia co paso do tempo.
- Teñen en conta as variacións ao longo do tempo: a analítica dos DW toma en consideración os cambios que se producen co tempo. Están deseñados para almacenar datos históricos e tomar en consideración as variacións, que axudan a entender a evolución e comportamento.

2.2 Que é un proceso ETL?

O ETL (Extract, Transform, Load) é o proceso polo cal se recompilan datos de diferentes fontes, cuxas orixes poden ser moi diversas en tipo, formato, volume e fiabilidade. Este procedemento garante que a información a empregar está limpa, estruturada e preparada para a súa explotación. É necesario procesalos para que sexan informativos e, consecuentemente, de utilidade. Os tres pasos principais son [9] [10]:

- **Extracción:** durante esta etapa identifícanse as diversas fontes e orixes de onde os datos se copian, de modo que se poidan transportar ao almacén de datos de destino. Estes datos poden ser estruturados ou non.
- **Transformación:** é necesario procesar os datos co fin de preparalos e quedarse unicamente cos rexistros de interese no almacén destino. Deben de ser transformados e analizados, é dicir, realizar unha limpeza e corrección para manter a integridade e coherencia no resultado final.
- **Carga:** por último, levarase a cabo a fase de carga, aquela na que se trasladan os datos ao almacén final onde poderán ser explotados. Pode ser en tempo real ou por lotes.

2.3 Que é unha API?

Unha **API** é un conxunto de normas e protocolos que permiten que diferentes programas ou aplicacións se comuniquen entre si. As **API** permiten acceder a funcionalidades e datos doutros servizos de xeito estruturado, facilitando a integración e a extensión das capacidades das aplicacións. Adoitan estruturarse en cliente e servidor [11].

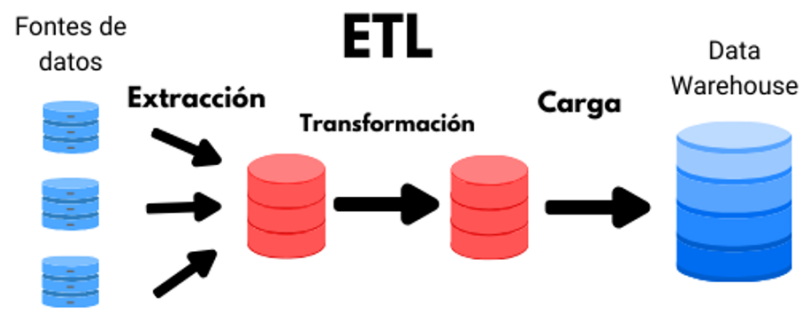


Figura 2.1: Fases procesos ETL

Ferramentas e tecnoloxías empregadas

Neste capítulo dótase dunha definición ás ferramentas e tecnoloxías relacionadas co proxecto de integración e análise.

Para a realización deste traballo precisáronse unha combinación de APIs, linguaxe de programación, ferramentas de xestión de bases de datos, e plataformas de visualización. A continuación, describirase cada unha delas para poder comprender os pasos realizados e o seu papel no proxecto.

3.1 Python

É unha linguaxe de programación con estruturas de datos de alto nivel eficientes e un sinxelo sistema de programación orientado a obxectos. A súa elegante sintaxe e tipado dinámicos, xunto ca súa natureza interpretada convérteno nunha linguaxe ideal para o *scripting* e desenvolvemento rápido de aplicación en diversas áreas, para a maioría de plataformas.

As librarías son coleccións de módulos e funcións desenvolvidas para facilitar a realización de tarefas comúns. Para traballar con APIs, Python ofrece unha ampla variedade de librarías que simplifican a interacción con estas interfaces. Dúas librarías claves neste proxecto son *Spotipy* e *musicbrainzngs* [12] [13].

- **Spotipy**: é unha librería de Python que intermedia de maneira sinxela coa API de Spo-

tify simplificando deste xeito as interaccións. Permite realizar buscas, obter información sobre álbums, artistas e listas de reprodución, así como acceder a datos sobre as preferencias dos usuarios [14].

- **Musicbrainzngs**: librería de Python que permite acceder directamente á **BBDD** de **MB**. Esta librería é unha **API** propia da plataforma. Facilita a obtención de información detallada sobre álbums, artistas, gravacións e outros datos relacionados coa música, evitando a necesidade de ter que cargala no propio equipo [15].
- **Psycopg2**: librería de Python que proporciona unha interface para interactuar cunha **BBDD** PostgreSQL permitindo así unhas rápidas conexións, insercións e consultas ca base de interese.

3.2 PostgreSQL

É un sistema de xestión de bases de datos de código aberto que destaca pola súa capacidade de adaptación e cumprimento co estándar **Structured Query Language (SQL)**. Emprégase para almacenar, organizar e recuperar información de maneira estruturada e segura. Algunha das súas características relevantes son [16]:

- Extensibilidade.
- Soporta grandes cantidades de traballo.
- Transaccións **ACID**: garantindo a fiabilidade dos datos.
- Compatibilidade con linguaxes procesuais, como python.
- Soporte para **JavaScript Object Notation (JSON)**.
- De balde.

3.3 Dbeaver

É unha aplicación software e ferramenta de bases de datos, deseñada para traballar con diversos sistemas de xestión como PostgreSQL, Oracle ou MySQL, entre outros. É de software libre e código aberto. Tamén presenta unha interface intuitiva dende a que permite a visualización dos datos[17].

3.4 Tableau

É unha ferramenta software para BI que permite analizar e explotar os datos de varias fontes [18]. Permite crear mapas, gráficos e análises en visualizacións interactivas en forma de libros ou *dashboards*. Grazas a isto, pódese obter coñecemento de grandes cantidades de información que axuden a un organismo ou particular na toma de decisións ou resolución de problemas. Algunhas das características que a fan tan interesante son as seguintes:

- Sinxela integración.
- Análise en tempo real.
- Conectividade de datos.
- Interface intuitiva.

Metodoloxía e planificación

Neste capítulo explícase a metodoloxía empregada durante a realización do proxecto e a súa planificación, a cal se divide en varias fases que se corresponden coas diferentes tarefas a realizar.

4.1 Metodoloxía

Un proxecto pasa por unha serie de fases e pasos ca intención de chegar a cumprir un determinado obxectivo. Por iso, é un proceso complexo que require unha planificación coitada e efectiva, de xeito que, realizando comprobacións periódicas dos resultados pódase asegurar o seu correcto desenvolvemento. A chave está en seguir unha metodoloxía adecuada para a organización dos pasos a facer, a xestión do tempo, dos recursos e a avaliación da calidade de resultados. Investir tempo e esforzo en desenvolver a metodoloxía adecuada é crucial para garantir o éxito do traballo.

Neste caso, empregouse unha metodoloxía baseada en Kimball, adaptándoa máis ás necesidades específicas do proxecto. Isto débese a que as táboas empregadas non se corresponden a simples dimensións, feitos e joins. Conteñen unha mestura de información e relacións cruzadas que precisan un enfoque máis flexible. As relacións entre as entidades suxiren un tratamento máis orientado a grafos ou redes de información, onde parte da súa complexidade radica en comprender como están conectadas entre si.

4.1.1 Metodoloxía Kimball

Esta caracterízase por basearse nun modelo dimensional, é dicir, que os datos se organizan en torno ás dimensións e os feitos, e por empregar os procesos ETL. Primeiro obteranse os datos das fontes, os cales pasan unha serie de transformacións e limpeza no caso de ser necesario e, por último, cargaranse os datos ao almacén. [19] [20]

Tamén se fundamenta nun desenvolvemento iterativo e incremental. O desenvolvemento do almacén de datos realízase mediante unha serie de etapas ou iteracións sucesivas, o cal permite unha entrega de valor rápida e capacidade de ir mellorando o deseño. Gracias a isto, finalmente devolve unha solución completa acorde coas expectativas e requisitos requiridos.

Adoitan empregarse para o deseño, construción e análise dos **Data Warehouse (DW)**. As fases de traballo que a comprenden son as seguintes:

1. **Planificación do proxecto:** nesta primeira fase defínense claramente os obxectivos do proxecto, alcance, recursos que serán necesarios ademais da organización do cronograma. Son necesarias reunións iniciais para poder marcar as metas e os recursos necesarios entre os involucrados no proxecto.
2. **Definición de requisitos:** na segunda fase recompílase a información das necesidades e intereses dos potenciais clientes ou usuarios finais do servizo ou proxecto. Serán necesarias intervencións, nas cales se poderán reforzar os coñecementos e entender os requisitos funcionais e non funcionais.
3. **Deseño dimensional:** neste paso creárase un deseño dimensional que facilite a análise dos datos e relacione todos os conceptos necesarios. Algúns dos pasos necesarios son:
 - Escoller os procesos de negocio: a directiva decidirá centrarse nunha área ou proceso a modelar. Basearanse nos requisitos definidos na fase anterior.
 - Determinar o nivel de detalle e granularidade: escóllese como de específica debe de ser a **BBDD** e en que temas se quere afondar máis.
 - Escoller as dimensións: identificar as dimensións que proporcionarán o contexto necesario para a análise dos feitos. Basearanse no nivel de detalle e requisitos requiridos.
 - Determinar os feitos medibles: identifica as medidas e feitos que serán almacenados nas táboas. Proporcionan información relevante para as análises.

4. **Deseño do modelo físico:** tras a identificación de táboas de feitos e dimensións, determínase a estrutura que conformará a unión entre elas para que sexa funcional, é dicir, defínirase o modelo dimensional.
5. **Deseño do sistema de ETL:** defínirase nesta fase cales son as fontes de datos de interese e o deseño de fluxos de datos. Deberanse defínir as regras de transformación e limpeza e, polo tanto, cales son valores válidos e de interese. Por último, deberase preparar a carga dos datos no almacén necesario.
6. **Deseño da arquitectura do almacén de datos:** determinar a infraestrutura técnica apropiada, o DW para poder soportar a arquitectura e as táboas requiridas. É importante a selección dun software e hardware axeitados.
7. **Implantación:** constrúese o DW segundo os deseños establecidos no paso anterior. Os procesos ETL estarán operativos e as táboas de feitos e de dimensións iranse enchendo. Nesta fase realízanse moitas probas.
8. **Desenvolvemento de aplicacións front-end:** esta fase céntrase na creación de interfaces de usuario que facilitarán a interacción ca aplicación ou cos datos almacenados no DW.
9. **Explotación dos datos:** empregaranse ferramentas de Intelixencia de Negocio que permitan a realización de informes e *dashboards* para a análise exhaustiva dos datos. Permiten aos usuarios finais acceder ós datos dun xeito comprensible.

4.2 Planificación do proxecto

Para lograr o éxito nun proxecto, é fundamental contar cunha planificación axeitada e organizada en fases ou etapas establecidas que sirvan de guía durante todo o proceso. Isto garante unha execución ordenada e a identificación de problemas de maneira eficiente. Unha boa planificación é a clave para evitar imprevistos, minimizar riscos e asegurar a calidade do proxecto. A causa diso, definiuse un diagrama de Gantt, como se mostra na figura 5.2, onde se atopan as diferentes tarefas do proxecto estimadas en tempo dependendo da súa dificultade e densidade.

O traballo foi dividido en 9 etapas e, a maiores, a redacción da memoria.

En primeiro lugar, leváronse a cabo dúas reunións iniciais entre o alumno e os dous titores para establecer o alcance e requisitos do proxecto. Nelas valoráronse as diferentes fontes de

información e ferramentas a empregar.

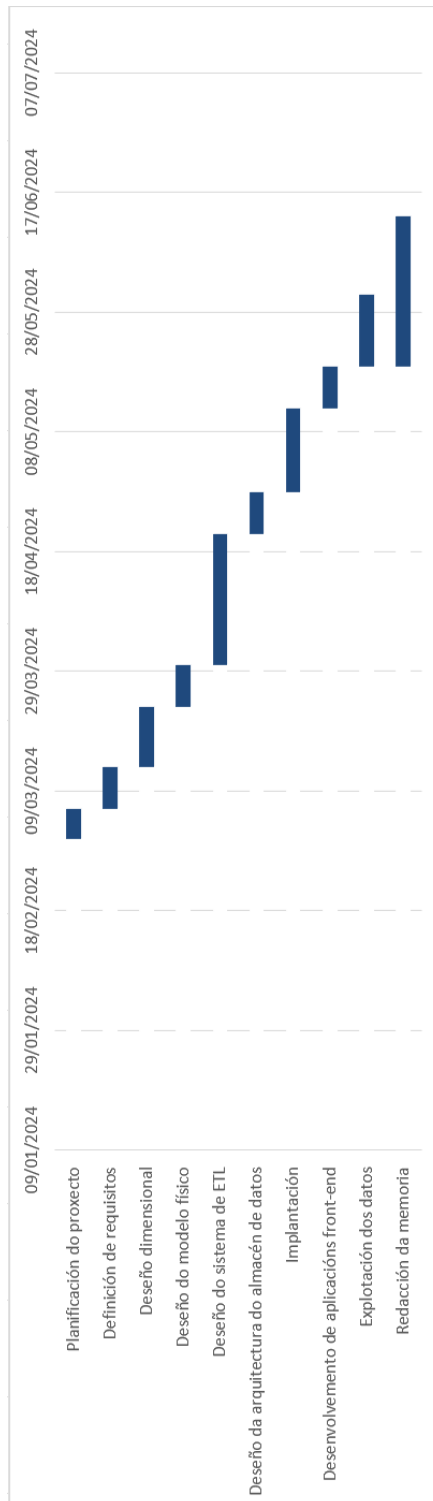


Figura 4.1: Diagrama de Gantt: Estimación inicial

Seguemento

Realizáronse ao longo da elaboración do traballo reunións periódicas cunha frecuencia de promedio de dúas semanas. Nelas valoráronse os avances e resultados obtidos.

- **Planificación do proxecto:** desenvolveuse adecuadamente cumprindo coas expectativas iniciais.
- **Definición de requisitos:** realizouse un estudo exhaustivo das dúas fontes de datos para compender a súa capacidade e alcance.
- **Deseño dimensional:** nesta fase definíronse as entidades, os seus atributos e as relacións entre elas. Debido a que as necesidades deste proxecto non requiren nin táboas de dimensións nin de feitos non se sigue de xeito puro a metodoloxía de Kimball, mantendo os principios básicos pero adaptada. Nela deseñouse un bosquexo inicial, (exemplo no apéndice A.2) que foi evolucionando co paso do tempo, reflexándose en múltiples versións, ata o modelo definitivo mostrado no capítulo 5.2. Plantexáronse cuestións como cales eran os principais fluxos de información que o sistema debe soportar, as colaboracións ou a posibilidade de que un artista poda ser un traballador. As revisións periódicas permitiron identificar posibles melloras e realizar os axustes necesarios.
- **Deseño do modelo físico:** ideouse a estruturación das táboas e as relacións entre elas. Foi revisada e aprobada nas reunións. Non se presentaron complicacións destacables nesta fase.
- **ETL:** esta foi a etapa máis longa e complexa do proxecto. A heteroxeneidade das fontes de información e a dispersión dos valores de interese supuxo un reto. A estrutura de MB é especialmente enrevesada, sendo así a súa extracción e limpeza un dos principais desafíos. Outros obstáculos atopado foron as actualizacións e modificacións producidas na API de Spotify, aparecendo novos erros no código xa funcional. Solucionouse realizando cambios manuais nas funcións desactualizadas da librería Spotipy.
- **Deseño do almacén de datos:** o deseño da infraestrutura técnica e a selección de ferramentas axeitadas realizáronse sen problemas significativos. As revisións periódicas permitiron garantir que a arquitectura proposta cumpría cos requisitos do proxecto.
- **Implantación:** houbo unha serie de problemas durante a carga de datos á base debido ao exceso de peticións diarias a Spotify, penalizando o proceso e non permitindo execucións nas horas seguintes. Estas poden variar, mais o límite é de aproximadamente

180 solicitudes por minuto, para arranxalo estableceuse un tempo de espera tras varias consultas, ralentizando o proceso.

- **Desenvolvemento de aplicacións front-end:** desenvolveuse segundo o previsto. Aumentáronse as horas de traballo dedicadas diariamente ao proxecto para compensar os retrasos das dúas etapas anteriores e cumprir ca planificación.
- **Explotación dos datos:** produciuse un parón de 5 días no proxecto por parte da alumna causado pola entrega de traballos finais e exames.
- **Redacción da memoria:** paso final do proxecto. A documentación foi supervisada durante as reunións periódicas, garantindo que todos os aspectos do proxecto estivesen ben cubertos e que a memoria reflectise axeitadamente o traballo realizado e os resultados obtidos.

A continuación, na figura 4.1 refléxanse as horas totais empregadas na realización do proxecto e as das reunións cos dous supervisores. No tocante aos materiais e recursos empregados, todos foron de balde, entre eles: ordenador persoal e adquisición de Tableau cunha licenza de estudante.

Os principais portes do traballo veñen derivados do tempo dedicado, tanto polo encargado do desenvolvemento do proxecto como polos seus supervisores. O encargado traballou durante un total de 320 horas. Tamén se realizaron 10 reunións, cada unha dunha hora, con dous supervisores, sumando un total de 20 horas de supervisión. Tras o mencionado, estimando un custo de 15€/h para o traballador do proxecto e de 30€/h para os xefes, trataríase dun proxecto cun presuposto de 5.400€.

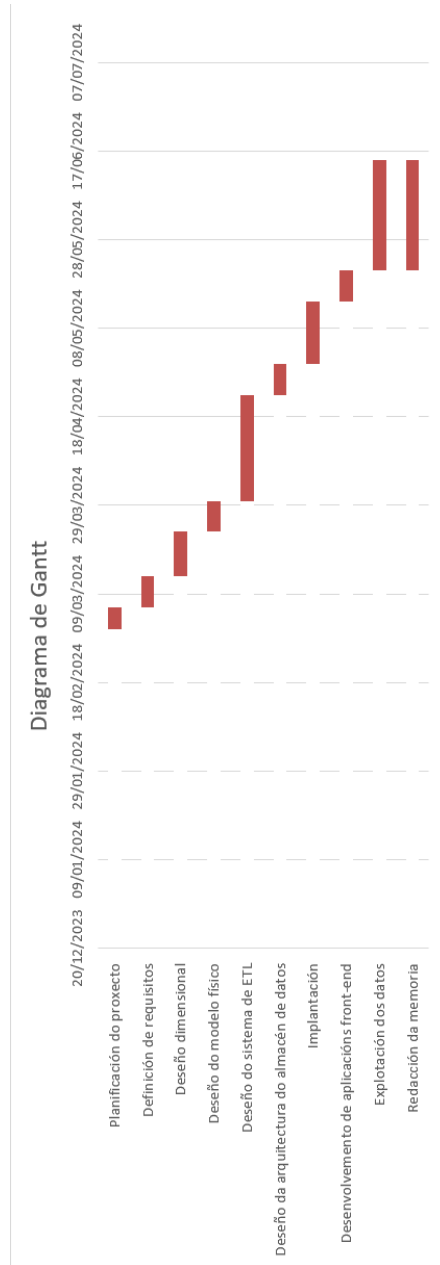


Figura 4.2: Diagrama de Gantt: Evolución real

Tarefa	Data de inicio	Data de finalización	Duración en días	Duración en horas
Planificación do proxecto	01-03-2024	05-03-2024	5	10
Definición de requisitos	06-03-2024	12-03-2024	7	20
Deseño dimensional	13-03-2024	22-03-2024	10	30
Deseño do modelo físico	23-03-2024	29-03-2024	7	20
Deseño do sistema de ETL	30-03-2024	22-04-2024	24	60
Deseño da arquitectura do almacén de datos	23-04-2024	29-04-2024	7	20
Implantación	30-04-2024	13-05-2024	14	40
Desenvolvemento de aplicacións front-end	14-05-2024	20-05-2024	7	20
Explotación dos datos	21-05-2024	14-06-2024	25	30
Redacción da memoria	21-05-2024	14-06-2024	25	50
Total				320
Reunións adicionais			10	20
Custe Total (€)				5.400

Táboa 4.1: Planificación e coste do proxecto con retrasos

Deseño do Data WareHouse

Neste capítulo explórase o deseño do almacén *DW*. Trátase dende a definición dos requisitos, a creación dos modelos dimensional e lóxico ata a implementación das táboas para almacenar e analizar a información detallada da industria musical.

5.1 Definición de requisitos

Como xa se explicou na introdución (Capítulo 1) o obxectivo principal do traballo é recompilar e integrar información detallada da industria musical, os artistas, as súas obras e proporcionar unha visión global e precisa que sexa útil para un público non especializado nas técnicas de análise. Esta combinación non só enriquece a calidade da información dispoñible, senón que tamén abre novas oportunidades para desenvolver ferramentas e aplicacións innovadoras. Algunhas das análises a realizar son as seguintes:

- **Perfil completo dos artistas:** creación dun perfil detallado dun artista no que se poida atopar información acerca da súa biografía, popularidade, orixe, obras, etc. A través del pódese comprender e analizar a súa traxectoria, impacto na industria musical, colaboracións e discografías.
 - **Datos requiridos:**
 - * Datos de obras e popularidade dos artistas en Spotify.
 - * Información biográfica e discográfica de *MB*.
 - * Relacións e colaboracións con outros artistas.

- **Impacto de Producción e Labor Creativa:**
 - **Motivo:** recoñecer o traballo dos produtores, compositores e outros traballadores que participan na creación das cancións, e entender como estas contribucións inflúen no éxito das obras.
 - **Datos requiridos:**
 - * Datos de créditos de produción e composición en [MB](#).
 - * Información sobre participación de artistas colaboradores en Spotify.
 - * Datos sobre os métodos de produción e tecnoloxías utilizadas nas gravacións.
- **Análise das tendencias da música:** observar como os artistas e obras máis vellos foron cambiando co paso do tempo. Descubrir cales son os artistas e xéneros máis populares, así como os instrumentos e as producións máis comúns dependendo da época.
 - **Datos requiridos:**
 - * Datos de obras, categorías, xéneros en Spotify.
 - * Información dos instrumentos en [MB](#).
 - * Datas.
- **Análise da Evolución dos Grupos Musicais:** examinar a evolución dos grupos musicais ao longo do tempo, identificando cambios na súa formación, publicacións e influencia na industria musical. Isto permitirá comprender como os grupos se adaptan ou quedan eclipsados polas carreiras individuais.
 - **Datos requiridos:**
 - * Datos históricos sobre a formación e cambios de membros dos grupos en [MB](#).
 - * Información sobre os álbums e cancións lanzadas por cada grupo en Spotify.
 - * Análise das colaboracións e proxectos con outros artistas e grupos.

Os requisitos funcionais son:

1. Conexión cás [API](#) de Spotify e [MB](#) e extracción: en ambas deberá de realizarse unha autenticación para poder conectarse e acceder aos datos. A continuación, extraeranse os datos.
2. Integración dos datos: combinar a información obtida de ambos e normalizala.
3. Realización dos procesos [ETL](#) requiridos.

4. Almacenar os datos de xeito eficiente nunha **BBDD** coa estrutura adecuada.
5. Funcionalidades de busca e consulta da información contida na base.
6. Visualizacións interactivas mediante gráficos, mapas ou *dashboards* interactivos que permitan visualizacións da popularidade das cancións, álbums, etc.

5.2 Deseño do modelo dimensional

Como xa se explicou con anterioridade, o deseño dimensional, tamén chamado conceptual, serve para facilitar a análise dos datos e relacionar os conceptos nos que se centra o proxecto. Descríbense a continuación cada unha das fases de maneira detallada.

5.2.1 Escoller os procesos de negocio

Os procesos de negocio correspóndense cos principais fluxos de traballo e actividades que o sistema debe soportar. Neste caso, para a **BBDD** musical considéranse os seguintes:

- **Xestión de artistas e de grupos:** rexístranse artistas e grupos na **BBDD** xunto cunha serie de detalles asociados á súa bibliografía e produción. Centrándose nos grupos, tense en conta o período de actividade do mesmo, ademais da duración das relacións cos artistas.
- **Publicación de álbums e cancións:** todas as cancións, tamén denominadas *tracks*, atópanse asociadas a un álbum, aínda que unicamente se encontre conformado por unha. Poderanse engadir novos álbums á **BBDD**, incluíndo detalles como o nome e a imaxe. Da mesma maneira, é posible agregar máis cancións.
- **Xestión dos xéneros musicais:** asociar aos artistas ou grupos aqueles xéneros que mellor os definan. Os xéneros musicais son categorías ou etiquetas que Spotify decidiu crear en función dos criterios que consideren relevantes. Adoitan ser comúns ou relacionados entre, por exemplo, artistas do mesmo grupo ou os seus colaboradores.
- **Xestión de colaboracións:** á hora de publicar un álbum ou unha canción, poden estar asociado a un ou máis grupos ou artistas.

- **Xestión dos traballadores musicais:** habitualmente non aparecen recoñecidos de xeito directo nas cancións, mais as labores de produción e escritura, entre outras, son fundamentais para o éxito da mesma.
- **Análises e métricas:** conxunto de medidas como a duración das cancións, popularidade ou enerxía para poder analízalas e comparalas.
- **Visualización e informes analíticos:** xerar informes detallados e interactivos que acheguen información sobre aspectos determinados da **BBDD**.

Ao longo do tempo poderanse realizar procesos de actualización da información existente como, por exemplo, para os cambios na biografía como a imaxe, a popularidade ou a revisión das formacións e pertenzas ós grupos.

5.2.2 Determinar o nivel de detalle e granulado

O nivel de granulado define a especificade da información a almacenar. Neste caso, o nivel de detalle será bastante específico. Como se explicou no proceso de negocio, almacenarase información concreta dos artistas, álbums, cancións, etc. Por exemplo:

- Para os artistas móstrase información detallada da súa persoa, como o nome propio, imaxe ou popularidade.
- Para os álbums obtense a data de publicación, que tipo son e todas as cancións que a conforman.
- No caso das cancións, tómanse varias características auditivas, os traballadores co papel que desenvolveron e os instrumentos empregados.
- No relativo ás datas úsanse os anos, especificación suficiente para a análise a realizar.

5.2.3 Determinar entidades

Os conceptos distinguibles con existencia propia nun dominio específico denomínanse entidades. No caso do modelo entidade relación (**Entidade Relación (ER)**), as entidades e a maneira na que se relacionan son cruciais porque definen os compoñentes fundamentais da **BBDD**. Permiten realizar un deseño claro e estruturado. Neste caso correspóndense a:

- **Artistas:** todos aqueles cantantes, (*disc-jockey (DJ)*) ou músicos recoñecidos no mundo da música que dispoñen dunha conta de Spotify.
- **Grupos:** agrupacións de varios artistas. Os seus integrantes poden ter carreiras en solitario tamén.
- **Histórico de pertenzas:** relaciona os grupos e os seus integrantes ó longo do tempo.
- **Álbuns:** poden estar conformados por unha ou por un conxunto de cancións. Existen dous tipos: o ‘album’, a opción máis clásica, a cal se refire a unha colección completa de música xeralmente con máis de 7 *tracks*; o ‘single’, que adoita conter unha ou un número moi reducido de cancións (xeralmente 3 como máximo). No caso de Spotify, o *single* tamén serve para designar os ‘extended play’, a medio camiño entre a duración de un *single* e un álbum.
- **Xéneros:** categoría que agrupa as obras musicais que comparten características en común. Poden estar asociados ó estilo musical, como a estrutura da canción, o idioma, o contexto cultural ou a temática e o contido. Na base relaciónanse cos artistas e grupos.
- **Tracks (cancións):** é unha pista de audio individual que forma parte dun álbum ou colección de música. Adoitan ser cancións, pero pódese corresponder tamén a outra forma de gravación de audio.
- **Traballadores:** aquelas persoas participantes no desenvolvemento dun track nalgún dos procesos requiridos. Poden estar involucrados na composición da letra, mestura de sons, produción, tocar instrumentos, ou engadir a súa voz en coros.
- **Instrumentos:** ferramenta creada co propósito de producir sons.

Alguns dos valores analíticos que se poden obter das entidades como álbums, artistas e *tracks* son:

- **Popularidade:** medida baseada na frecuencia do número de escoitas, descargas e gardados.

Para as *tracks* podemos medir os seguintes valores:

- **Valance:** medida da positividade musical transmitida por unha pista.
- **Enerxía:** medida da intensidade e actividade perceptual dunha pista.

- **Acústica:** medida da probabilidade de que unha pista sexa acústica.
- **Instrumentalidade:** medida de probabilidade de que unha pista non conteña letra.
- **Danceability:** medida de como de xeitosa é unha pista de audio para bailar, baseada en varios elementos musicais.
- Cantidade de **colaboracións** entre artistas e *tracks*.
- Número de **roles** e participación de traballadores e instrumentos na produción.

Todo o anteriormente mencionado, na figura 5.1 pódese apreciar no diagrama entidade relación, isto é, como as entidades interactúan entre si para poder reflexar as relacións existentes na realidade. Resumidamente, ponse de manifesto o seguinte:

- Os xéneros caracterizan aos artistas e os grupos, e ambos se poden identificar con varios xéneros.
- Un artista puido pertencer a un grupo en varias ocasións, que son gardadas de maneira independente en función do período temporal no que ocorreu. Os roles axudan a comprender a actividade que desenvolvían no grupo, por exemplo: ‘orixinal’: refírese a un integrante que estivo dende o inicio, é dicir, fundou o grupo; “guitarrista”: se o seu rol era tocar a guitarra, etc. É posible ter varios.
- Os artistas e os grupos poden publicar varios álbums e cancións de xeito individual ou en colaboracións.
- Cada track pode ter traballadores, tanto implicados na produción como tocando instrumentos.
- Os traballadores poden ser artistas.

5.3 Deseño do modelo lóxico

Para o correcto deseño da base de datos débese realizar o deseño do modelo lóxico. Baseándose no conceptual previamente explicado, as táboas e os seus atributos son definidas. A maneira na que se relacionan entre elas definirá a estrutura do DW. Búscase que esta reflexe, da maneira máis acorde á realidade, óptima e coherente as relacións para que o xestor da BDD poida aproveitar os recursos dunha forma eficiente.

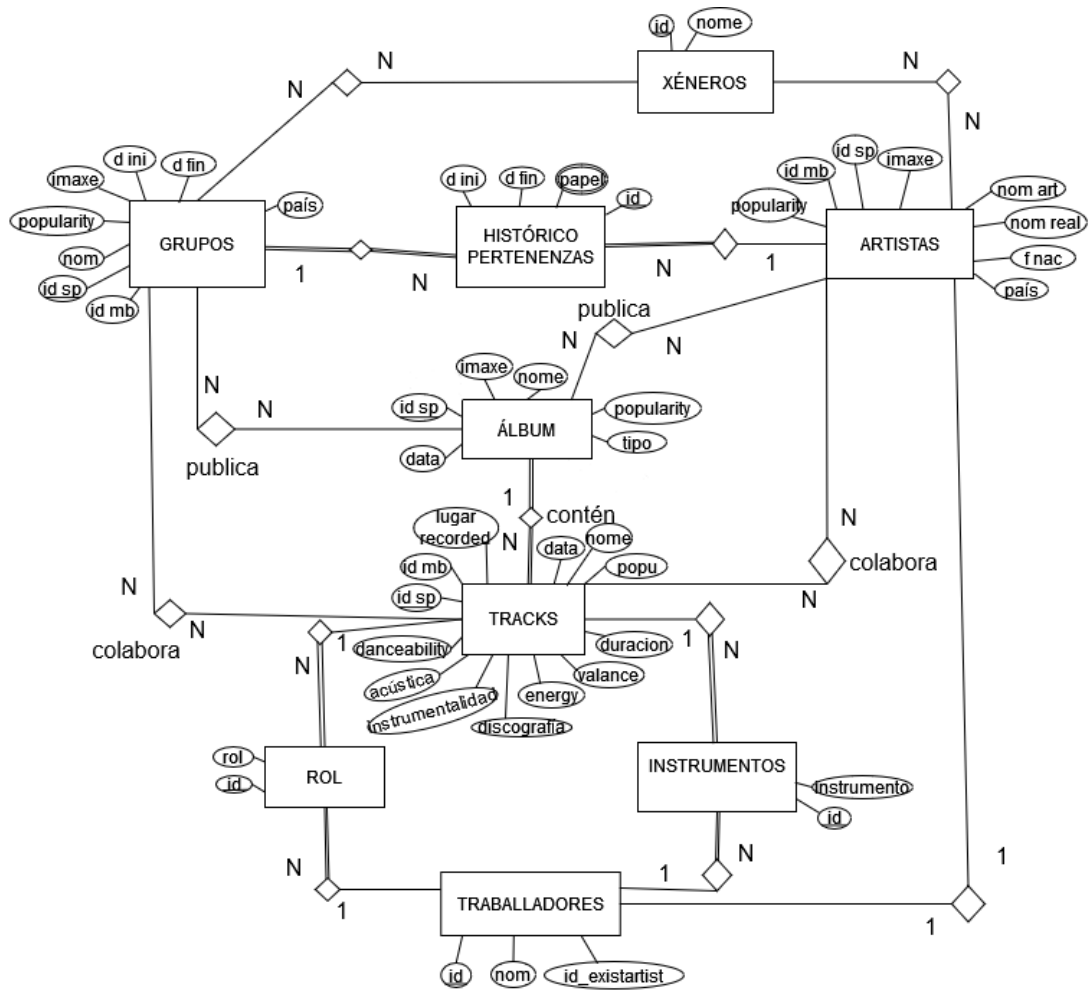


Figura 5.1: Diagrama Entidade Relación (ER)

As táboas son dotadas de atributos, como as chaves foráneas, para poder ser ligadas entre elas. A necesidade de unir varias táboas de dimensións, tales como as colaboracións dos diferentes artistas ou grupos á hora de publicar un álbum ou as colaboracións nas cancións, xera un novo tipo de táboas chamadas táboas ponte.

En consecuencia, quedan da seguinte forma:

- **ARTISTAS:**

- Id_mb: identificador propio da BBDD de MusicBrainz, único para cada persoa.
- Id_spoty: identificador propio da BBDD de Spotify, único para cada artista.
- Nome artístico: pseudónimo empregado polo artista para crear unha identidade propia e máis recoñecible. Pode corresponderse co nome real.
- Nome real: nome real da persoa vinculada á personalidade artística.
- Data de nacemento: ano de nacemento da persoa vinculada ao artista.
- País: país no que naceu ou co que se vincula maioritariamente o artista.
- Imaxe: imaxe de Spotify máis recente do artista.
- Popularidade: medida de recoñecemento dun artista. Explicado no apartado 5.2.3

- **GRUPOS:**

- Id_spoty: identificador propio da BBDD de Spotify único para cada grupo.
- Id_mb: identificador propio da BBDD de MB único para cada grupo.
- Nome da agrupación: nome baixo o cal se denomina ó conxunto de artistas.
- Data de inicio: ano no que o grupo se formou.
- Data do fin: ano no que o grupo decidiu separarse e cesar a súa actividade.
- País: país co que se asocia o grupo, pode ser aquel no que comezou a súa actividade, ou o país ao que pertencen a maioría dos seus integrantes.
- Imaxe: imaxe de Spotify máis recente do grupo.
- Popularidade: medida de recoñecemento dun artista. Explicado no apartado 5.2.3

- **ÁLBUMS:**

- Id_spoty: identificador propio da BBDD de Spotify, único para cada álbum.
- Nome: nome co que se designa o álbum.
- Tipo: pode ser álbum ou *single*, como xa se aclarou previamente.

- Imaxe: imaxe da portada do álbum. Adoita asociarse como imaxe dos *tracks* tamén.
 - Popularidade: medida de recoñecemento dun artista. Explicado no apartado 5.2.3
 - Data de publicación: ano de publicación do álbum.
- **XÉNEROS:**
 - id: identificador único do xénero.
 - nome: nome do xénero.
- **TRACKS:**
 - Id_spotify: identificador propio da BBDD de Spotify, único para cada track.
 - Id_mb: identificador propio da BBDD de MB, único para cada track.
 - Id_album: identificador do álbum ao que pertence o track.
 - Nome: nome do track.
 - Data de publicación: ano de publicación do álbum.
 - Lugar de gravación: lugar onde foi gravado o track.
 - Duracion: duración en milisegundos.
 - Popularidade, Valance, Enerxía: medidas de recoñecemento, positividade e actividade dunha canción. Explicado no apartado 5.2.3
 - Discografía: compañía discográfica baixo a que se produxo e/ou publicou o track.
 - Acústica, Instrumentalidade, Danceability: probabilidades de utilización de instrumentos acústicos, que no teña letra, solo conteña instrumentos e de que sexa bailable. Explicados no apartado 5.2.3
- **TRABALLADOR:**
 - Id_mb: identificador propio da BBDD de MB, único para cada traballador.
 - Nome: nome do traballador.
 - Id_existartist: identificador do artista existente se aplica.
- **ROL:**
 - Id: identificador único do rol.
 - Id_track: identificador do track no que o rol é relevante.
 - Rol: descrición do rol desempeñado.

- **INSTRUMENTAL:**
 - Id: identificador único do instrumento.
 - Id_track: identificador do track no que se usou o instrumento.
 - Id_trab: identificador do traballador que tocou o instrumento.
 - Instrumento: descrición do instrumento, explicado no apartado 5.2.3
- **HISTÓRICO DE PERTENZAS:**
 - Data de inicio: data de inicio da pertenza.
 - Data de fin: data de fin da pertenza.
 - Rol: rol desempeñado durante a pertenza. Poden ser varios.
- **GROUP_XEN** (táboa ponte):
 - Id_grupo: identificador do grupo.
 - Id_xen: identificador do xénero.
- **ART_XEN** (táboa ponte):
 - Id_art: identificador do artista.
 - Id_xen: identificador do xénero.
- **PUBLICA_ARTISTA** (táboa ponte):
 - Id_artista: identificador do artista.
 - Id_album: identificador do álbum.
- **PUBLICA_GRUPO** (táboa ponte):
 - Id_grupo: identificador do grupo.
 - Id_album: identificador do álbum.
- **COLABORA_GRUPO** (táboa ponte):
 - Id_grupo: identificador do grupo.
 - Id_track: identificador do track.
- **COLABORA_ARTISTA** (táboa ponte):
 - Id_artista: identificador do artista.
 - Id_track: identificador do track.

Finalmente xorden 16 táboas, como se pode ver na figura 5.2, das cales 10 correspóndense a entidades e 6 a derivadas táboas ponte.

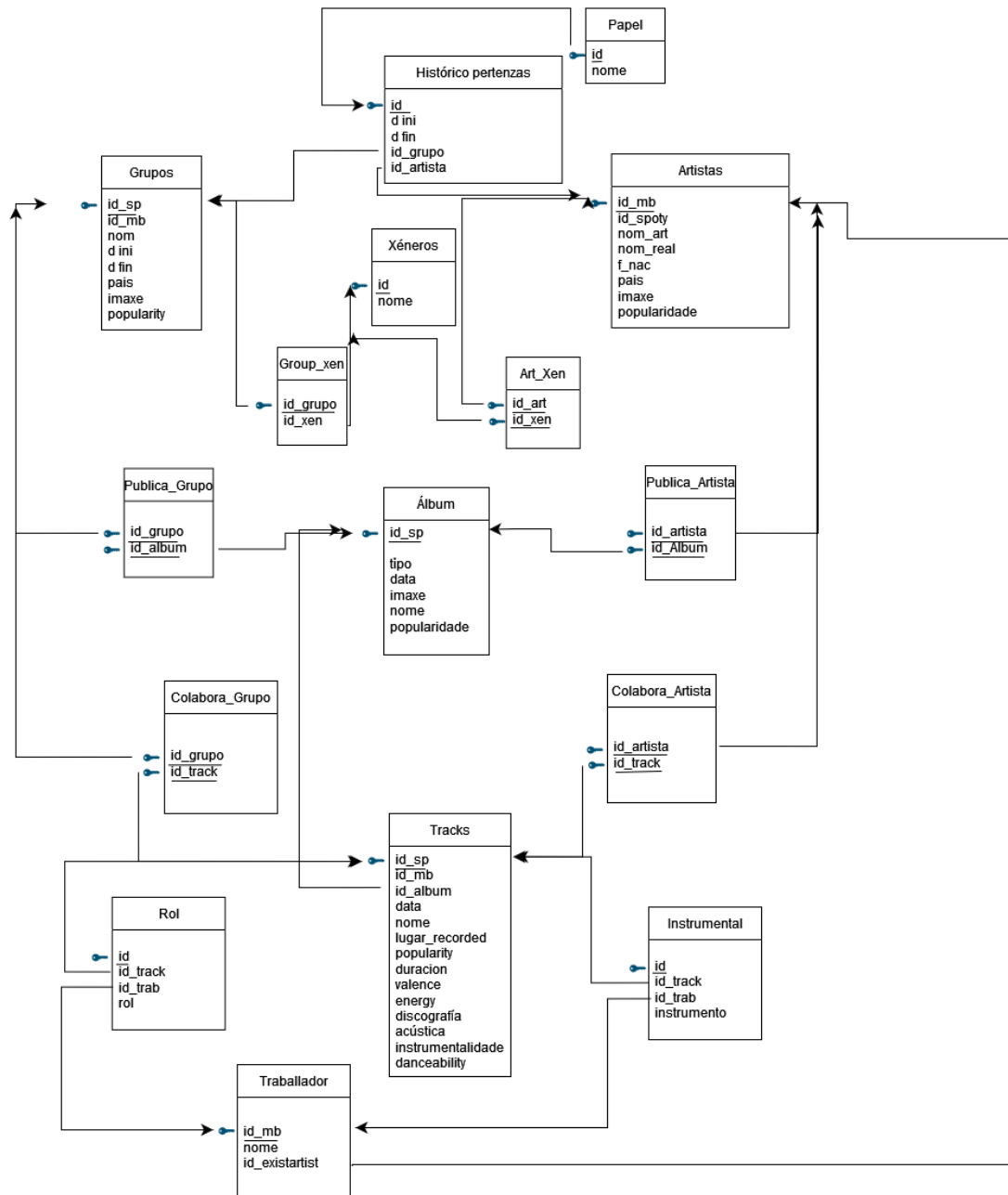


Figura 5.2: Diagrama relacional

Proceso ETL

Neste capítulo descríbese o proceso **ETL**, detallando cada unha das súas fases. Inclúense especificacións das ferramentas e tecnoloxías empregadas, como son as **API**, Python, PostgreSQL e Dbeaver.

O proceso **ETL**, é dicir, a extracción, transformación e limpeza dos datos son claves nun proxecto que requira empregar información de fontes diferentes. É un paso laborioso de realizar debido á complexidade de extraer os datos dos múltiples lugares nos que se encontran espallados; a transformación, para que estean dotados dun sentido global e se correspondan con valores válidos e, por último, os sucesivos procesos de conexións e cargas de datos ca base de datos. Este conxunto asegura que a información a empregar está limpa, estruturada e preparada para a súa explotación. Ter uns datos de calidade é crucial para a análise pola:

- **Precisión:** permiten obter resultados fiables, e consecuentemente, tomar as decisións máis apropiadas co debido coñecemento.
- **Eficiencia:** uns datos ben estruturados facilitan uns procesos máis rápidos permitindo aforrar tempo e recursos.

Python foi esencial nas tres fases: actuou como enlace entre todos os procesos, interactuando coas **API** para a extracción de datos, limpando e organizando os datos recollidos, e conectándose coa base de datos para executar consultas e insercións nas táboas nunha **BBDD** PostgreSQL. O código empregado atópase dividido en 3 ficheiros, un contén as funcións para realizar as conexións, outro xestiona as relacións e filtrados coa **BBDD** de **MB** e un terceiro e principal que chama os anteriores, onde se atopan as xestións da información de Spotify.

6.1 Fontes de datos

Neste caso, as dúas fontes principais son dúas bases de datos: [MB](#) e Spotify. Como se mencionou no capítulo 1, cada unha ofrece servizos diferentes na industria especializados nun sector.

Spotify é unha plataforma de distribución de música online orientada ó consumo, facendo de intermediaria entre artistas e consumidores. Esta asenta unha boa base para relacionar aos artistas e o seu catálogo, mais non afonda nin nos intérpretes nin na realización das cancións. Dela obtéñense a discografía completa, as colaboracións e atributos interesantes, tales como a popularidade tanto para artistas, cancións e álbums ou unha imaxe ilustrativa. Unha das desvantaxes que presenta é que nela non se distingue entre artistas individuais ou grupos e tampouco se indican os membros que o conforman.

Pola outra banda, a [BBDD MB](#) é unha enciclopedia que se centra na recompilación de información específica de todos os campos da industria musical. Xuntando a información de ambas chégase a unha visión completa do panorama musical, a cal permite unha análise máis profunda e exhaustiva das colaboracións, grupos e relacións entre os datos. Un conxunto de datos completo e diverso é esencial para a creación de análises e informes precisos.

Nos seguintes apartados afondarase en cada un dos ciclos da [ETL](#).

6.2 Extracción

Tal e como se detallou na sección 2.2, a extracción comprende o proceso de chegar ás diferentes fontes onde se atopan os datos para collelos e poder comezar a traballar con eles.

Spotify

Spotify é unha plataforma de música en liña, que contén a discografía da ampla maioría dos artistas e grupos do mundo [1]. A web [API](#) de Spotify é unha interface que autoriza ós desenvolvedores a acceder a algúns dos seus datos e funcionalidades [14]. As súas funcionalidades principais son:

- Recuperar información musical: obter detalles dos artistas, *tracks*, álbums...

- Xestionar listas de reprodución: crear, modificar e eliminar listas de reprodución para os usuarios.
- Controlar a reprodución: escolla da canción ou álbum, play, pausa ou salto.
- Acceder a información dos perfís de usuario: poder ver as cancións e artistas máis escoitados, a música gardada e obter recomendacións.

Para o proxecto a máis relevante é a primeira xa que o fin do seu uso é recadar toda a información relevante sobre os artistas ou agrupacións musicais e as súas creacións.

A información e accións estrutúranse en diferentes conxuntos pola súa temática: artistas, álbums, *tracks*, audiolibros, categorías, capítulos, episodios, xéneros, mercados, reprodutor, listas de reprodución, buscador de usuarios e categorías. Este estudo enfocarase en:

- **Álbums:** almacena a información dos álbums e a relación co usuario. Cabe destacar que devolve as *tracks* que o conforman e os seus autores.
- **Artistas:** almacena información dos artistas, podendo obter o seu xénero, imaxes, popularidade, tipo, os seus álbums, as súas cancións máis exitosas e recomendacións doutros artistas semellantes.
- **Tracks:** almacena a información das *tracks* e as relacións co usuario. É moi completo e dispón de información moi específica das características sonoras e análises de audio.
- **Buscador:** mediante a súa interacción ao pasarlle o nome e categoría do que se quere, devolve o identificador de Spotify.

MusicBrainz

[MusicBrainz \(MB\)](#) é unha BBDD aberta de música moi completa que recolle información de todo o relacionado coa industria musical. Nela pódese atopar información da bibliografía completa dos artistas, detalles moi técnicos das gravacións, compilacións de audios, eventos musicais, documentación, lugares, etc. É mantida por unha comunidade de voluntarios e é utilizada como unha fonte de datos fiable para outros servizos [4].

Esta agrupa a información nunha cuantiosa cantidade de táboas que se relacionan entre si de xeito complexo. As principais táboas de interese son as seguintes:

- **Recording:** refírese a unha canción individual. Permite obter os datos da produción, músicos, mixers, instrumentos involucrados e afíns.
- **Work:** contén a información máis técnica relacionada coa composición e escritura.
- **Release:** similar a un álbum, adoita conter a información das compilacións musicais.
- **Area:** refírese aos lugares, onde puideron gravar as cancións, nacer os artistas, etc.
- **Artist:** pode referirse a un artista, conxunto (grupo) ou traballadores da profesión (produtores, enxeñeiros). Nesta pódese obter a información detallada das biografías.

A información recadada de Spotify fará de fío condutor durante toda a extracción, que se irá completando coa información de [MB](#).

6.2.1 APIs e librerías

Como se expuxo no capítulo 3.1, empregaremos a linguaxe python para realizar o proceso de extracción. Para interaccionar dun xeito máis sinxelo coas fontes e, polo tanto, coas [API](#) xa explicadas, empréganse as dúas librerías mencionadas especializadas en cada unha. Isto permite facilitar tanto as solicitudes como os envíos de datos, autenticación, autorización e manexo de erros.

Spotipy

Esta librería permite unha autenticación sinxela á hora do manexo dos tokens. Tamén abstrae a complexidade e facilita unha a configuración e integración [21].

Funcións de Spotipy

- Token de inicio: o access token é unha cadea de caracteres que representa a autorización concedida a unha aplicación para acceder a recursos específicos. É emitido como parte do fluxo de autorización e é necesario para facer as chamadas a [API](#).

```
1 set_auth(token_info[ 'access_token' ])
```

- Búsqueda do artista polo seu nome:

```
1 search(q='artist:' + name_artist, type='artist')
```

- Obter todos os seus álbums (álbums, *singles* e [Extended Plays](#)):

```
1 artist_albums(info_art_mb['id_spoty'], include_groups=('single,
album'))
```

- Información dos álbums:

```
1 sp.album(album_id_sp)
```

- Obter as *tracks* dun álbum:

```
1 album_tracks(album_id)
```

- Información dunha track:

```
1 sp.track(track_id)
```

- Información máis técnica dos sons:

```
1 sp.audio_features(track_id)
```

Musicbrainzngs

Abstrae a complexidade de facer as solicitudes [Hypertext Transfer Protocol \(HTTP\)](#) e devolve os datos en obxectos Python. Conta con métodos predefinidos para tarefas comúns, sen necesidade de construír manualmente as [Uniform Resource Locators \(URL\)](#), simplemente pasando os parámetros de consulta [15].

Funcións empregadas

- Configurar a aplicación:

```
1 musicbrainzngs.set_useragent("Example music app", "0.1",
"http://127.0.0.1:5000")
```

- Buscar un artista ou grupo polo seu nome (co obxectivo de obter o id):

```
1 musicbrainzngs.search_artists(artist_name)
```

- Información completa do artista:

```
1 musicbrainzngs.get_artist_by_id(artist_id)
```

- Busca do id de MB asociado á gravación da pista:

```
1 musicbrainzngs.search_recordings(artist=artista, recording=cancion,
    limit=1, offset=None)
```

- Información asociada á produción, músicos, mixers e instrumentos:

```
1 musicbrainzngs.get_recording_by_id(id_track,
    includes=['label-rels', 'artist-rels', 'place-rels'])
```

- Trata de obter o nome real do lugar a través do seu identificador:

```
1 musicbrainzngs.get_place_by_id(place_id, includes=['area-rels'],
    release_status=[], release_type=[])
```

- Obter información sobre áreas:

```
1 musicbrainzngs.get_area_by_id(area_id, includes=['area-rels'])
```

- Busca do id de MB asociado ao *work*:

```
1 musicbrainzngs.search_works(artist=artista, work=cancion, limit=0)
```

- Información detallada dunha obra:

```
1 musicbrainzngs.get_work_by_id(work_id, includes=['artist-rels'])
```

En xeral, tómanse unha serie de precaucións como as comprobacións de que os comandos devolven a información requirida. Antes de explotar a información compróbase que existen os elementos solicitados e non se atopan baldeiros para evitar erros durante a execución e sucesivas chamadas. Para xestionar a comunicación coas API é preciso renovar os tokens con frecuencia para manter a conexión e evitar o vencemento dos mesmos. Para xestionalo, definiuse unha función para a súa actualización e prever cortes de conexión.

```
1 TOKEN_INFO= "token_info"
2
3 def create_spotify_oauth():
4     return SpotifyOAuth(
5         client_id = "____",
6         client_secret = "____" ,
```

```
7         redirect_uri="http://127.0.0.1:5000/callback",
8         scope='user-library-read user-top-read',# user-library-read"
9     )
10
11 # Función simplificada para obter token
12 def get_token(sp_oauth):
13     token_info = sp_oauth.get_cached_token()
14     if not token_info:
15         print("Token not found.")
16         return None
17     now = int(time.time())
18     is_expired = token_info['expires_at'] - now < 120
19     if is_expired:
20         token_info =
21         sp_oauth.refresh_access_token(token_info['refresh_token'])
22     return token_info
23
24 def ensure_valid_token(sp_oauth, sp):
25     token_info = sp_oauth.get_cached_token()
26     if not token_info or token_info['expires_at'] -
27     int(time.time()) < 120:
28         token_info =
29         sp_oauth.refresh_access_token(token_info['refresh_token'])
30         sp.set_auth(token_info['access_token'])
31     return sp
32
33 musicbrainzngs.set_useragent("Example music app", "0.1",
34                             "http://127.0.0.1:5000")
```

6.3 Transformación

Tal e como se expuxo previamente na sección 2.2, é necesaria a realización do tratamento e limpeza dos datos recadados no paso anterior. Debido ás áxiles operacións de extracción da fonte e carga na **BDD**, cada comprobación e transformación necesaria é aplicada xusto depois da recolección do valor. Tras a obtención de todos os valores requeridos procédese coa inserción. Seguidamente explicaranse cales foron as precaucións e cambios aplicados para a consistencia dos valores.

6.3.1 Datas

De tódalas datas, tanto para as relativas aos nacementos dos artistas, como para os lanzamentos das cancións ou álbums, ou comezo ou fin do período de actividade dun grupo, extráese unicamente o ano. Isto ven derivado pola variedade de información dispoñible. Dende que a sociedade e a industria musical se volveron máis tecnolóxicas e inmediatas, xunto coa aparición das novas plataformas de streaming, os lanzamentos e rupturas están moi regulados con, por exemplo, datas de estrea fixadas con precisión a nivel de día. Aínda así, non todas son tan concretas, especialmente as de aqueles artistas e publicacións máis antigas, tendo un nivel de detalle inferior. Para as análises a realizar, non é necesario ter un día ou mes concreto, o nivel de ano é suficiente. Por iso, a solución atopada foi extraer mediante expresións regulares o ano, para simplificar o almacenamento e análise dos datos.

Afondando nas datas do inicio e fin dos grupos e as relacións cos seus integrantes tratouse o seguinte:

- **Grupo segue en activo:** a ausencia do valor na data de fin, pero coa presenza da de inicio, xeralmente indica que o grupo segue en activo.
- **Grupo rematou a súa actividade** (é dicir, valor válido de data fin): para os seus integrantes realizáronse unha serie de comprobacións e, consecuentemente, aplicáronse as correccións correspondentes:
 - A relación co integrante ten un inicio anterior ou igual ó fin do grupo (é dicir, o inicio da relación do artista como integrante ocorreu antes de que se separara) e non ten data fin ou é máis tardía coa separación do grupo: substitúese o valor da fin da relación pola data do final do grupo.
 - A relación co integrante ten un inicio posterior ó fin do grupo (é dicir, o inicio da relación do artista como integrante ocorreu despois de que se separara): considérase un reencontro puntual e unicamente terá data de inicio.

6.3.2 Integrantes:

Compróbase que, das relacións obtidas cos grupos, solo é de interese obter os integrantes reais. É necesario, polo tanto, realizar un filtrado, de maneira que o tipo de relación sexa ‘member of band’ e se corresponda cunha persoa. Outro tipo de relacións existentes, as cales

non son de interese neste caso son “guest”, “cover” ou “collaborated with”. Por exemplo, no caso da banda “One direction” houbo que filtrar o seguinte caso:

```
1 {
2   'type': 'member of band',
3   'type-id': '5be4c609-9afa-4ea0-910b-12ffb71e3821',
4   'target': '473c8b1c-5865-44c9-9725-56355308dc28',
5   'direction': 'forward',
6   'begin': '2010',
7   'end': '2010',
8   'ended': 'true',
9   'attribute-list': ['original'],
10  'artist': {
11    'id': '473c8b1c-5865-44c9-9725-56355308dc28',
12    'type': 'Group',
13    'name': 'The X Factor Finalists 2010',
14    'sort-name': 'X Factor Finalists 2010, The'
15  },
16  'attributes': [
17    {
18      'type-id': '4fd3b255-a7d7-4424-9a63-40fa543b601c',
19      'attribute': 'original'
20    }
21  ]
22 }
```

6.3.3 Nomes

Tras a busca inicial do artista a partir do nome artístico cedido polo usuario, a denominación oficial recollida na **BBDD** será a de **MB**. O motivo é que simplifica o traballo á hora de obter os nomes dos integrantes e posteriores consultas. Este criterio evita as posibles discrepancias entre as dúas fontes. Para o resto de nomes, tanto dos álbums (álbum/*single*/*EP*), como das cancións (*tracks*) e grupos tómase como referencia o valor de Spotify.

6.4 Carga

A fase de carga é a etapa final do longo proceso **ETL**. Durante esta os datos, debidamente transformados, son inseridos no seu destino final, permitindo así que se poida acceder e analizar a información de maneira eficiente. Débese asegurar que as relacións e a integridade dos

datos se manteñan intactas cando os datos son cargados nas táboas do **DW**. É crucial porque garante que os datos sexan introducidos na **BBDD** de xeito organizado coherente e eficiente.

O primeiro paso no proceso de carga é a conexión ao sistema de xestión de bases de datos, neste caso **PostgreSQL**. Este ofrece a infraestrutura necesaria para crear a **BBDD** que almacenará os datos transformados .

6.4.1 Táboas

A continuación, é necesario definir as táboas mencionadas na sección 5.3 na **BBDD**. Estas táboas son a estrutura da información. Dende a ferramenta software Dbeaver establécese unha conexión coa **BBDD** anteriormente xerada e comenzase a deseñar as agrupacións de datos mediante consultas **SQL**. Para cada unha realízase unha consulta na que se sinalan as claves primarias, os valores únicos e as relacións existentes mediante as claves foráneas que fan referencias, polo que é importante seguir unha orde.

A xerarquía seguida para a creación das táboas foi a seguinte: “artistas”, “grupos”, “histórico_pertenzas”, “papel”, “xéneros”, “group_xen”, “art_xen”, “albums”, “tracks”, “publica_artista”, “publica_grupo”, “colabora_artista”, “colabora_grupo”, “traballadores”, “rol” e “instrumental”.

Móstranse exemplos de dúas consultas para a creación da táboa para a entidade artistas e, tamén, para a que representa a relación existente entre as colaboracións dos grupos e as cancións. Para os artistas a clave primaria tomada como referencia é a de **MB**, para facilitar a súa relación cos traballadores, pero o identificador de Spotify é declarado como único tamén. No outro caso, xérase unha clave primaria composta por dúas claves foráneas que fan referencia as outras dúas táboas.

```
1 CREATE TABLE artistas (  
2 id_mb VARCHAR(100) PRIMARY KEY,  
3 id_spoty VARCHAR(50) UNIQUE,  
4 nom_art VARCHAR(100),  
5 nom_real VARCHAR(100),  
6 f_nac INTEGER,  
7 pais VARCHAR(50),  
8 imagen TEXT,  
9 popularity INTEGER  
10 );
```

```
1 CREATE TABLE colabora_grupo (  
2 id_grupo VARCHAR(50),  
3 id_track VARCHAR(50),
```

```

4 PRIMARY KEY (id_grupo, id_track), -- Clave primaria composta
5 FOREIGN KEY (id_grupo) REFERENCES grupos(id_spoty),
6 FOREIGN KEY (id_track) REFERENCES tracks(id_spoty)
7 );

```

6.4.2 Inserción

Tras isto, comeza a fase de encher a **BBDD**. Realízase a conexión á base de datos con Python empregando a librería ‘psycopg2’, que é un adaptador de PostgreSQL para a linguaxe que permite interactuar e manipular este tipo de **BBDD** de forma eficiente.

Para realizar as insercións débese de establecer unha conexión, executar a consulta, facer un commit, e, ao rematar, pechar a conexión. Foi definida unha función de inserción para cada unha das táboas de xeito independente que, á hora da inserción, comproba que non exista ningunha fila con esa clave primaria mediante o ‘ON CONFLICT’, para evitar repeticións. Isto é de utilidade na maioría de táboas, contando con todas aquelas que a chave primaria é definida antes da incorporación na **BBDD** (ou o propio valor no caso dos xéneros). Sen embargo, para o histórico de pertenzas, rol e instrumental é máis complexo. Nestes casos realízase unha comprobación valor por valor, previa á súa inserción, que determinará se ten sentido ou non añadilo.

```

1 def insert_rol(cursor, rol_data):
2     insert_query = """
3     INSERT INTO rol (id_track, id_trab, rol)
4     VALUES (%s, %s, %s);
5     """
6     if check_exists(cursor, 'rol', rol_data):
7         pass
8     else:
9         cursor.execute(insert_query, (rol_data['id_track'],
10        rol_data['id_trab'], rol_data['rol']))

```

Para as claves foráneas tamén se verifica que exista a clave primaria na táboa a referenciar para prever fallos e así ser incorporadas.

```

1 def insert_group_xen(cursor, group_xen_data):
2     id_xen = get_xenero_id(cursor, group_xen_data['nome'])
3     if id_xen is None:
4         print(f"No se encontró el ID para el género
5         {group_xen_data['nome']}")
6     return

```

```

6
7     insert_query = """
8     INSERT INTO group_xen (id_grupo, id_xen)
9     VALUES (%s, %s)
10    ON CONFLICT (id_grupo, id_xen) DO NOTHING;
11    """
12    cursor.execute(insert_query, (group_xen_data['id_art'], id_xen))

```

No código defínense as seguintes funcións:

- Conexións:
 - **connect_db**: establece a conexión coa **BBDD** postgresSQL e manexa posibles erros.
 - **close_db**: pecha a conexión coa **BBDD** e asegúrase que os recursos foron liberados correctamente.
- Comprobacións:
 - **check_exist**: comproba se unha liña existe na táboa.
 - **get_xenero_id**: obtén o identificador do xénero debido a que é autoxerado pola **BBDD**.
- Insercións:
 - **insert_artist**, **insert_goup**, **insert_album**, **insert_track**, **insert_trab** insiren os datos de artistas, grupos, álbums, *tracks* e traballadores evitando conflitos de chave primaria mediante “on conflict”. **insert_publica_grupo**, **insert_publica_artista**, **insert_colabora_grupo**, **insert_colabora_artista** almacenan as táboas ponte e revisan a clave primaria dobre empregando o mesmo sistema. **insert_rol**, **insert_instrumental** revísano chamando a función de comprobación.
 - **insert_xeneros** inserta os xéneros evitando conflitos polo nome. **group_xen** e **art_xen** insertan as relacións entre artistas e xéneros, para o que se apoian en chamadas a “get_xenero_id”.
 - **insert_pertenencias**,: inserta as pertenzas chamando á función de comprobación e a **insert_papel** que garda o rol desenvolvido no grupo.

6.4.3 Pasos

Debido ás complexas relacións na base de MB [22], a información recadada de Spotify fai de fío condutor durante toda a extracción, que se irá completando cos datos da primeira. A xerarquía seguida durante as solicitudes e, especialmente para as insercións, ten en conta as dependencias existentes entre as táboas para así efectuar un fluxo coherente. Na imaxe 6.1 (páxina 50) pódese apreciar a orde levada a cabo, que foi a seguinte :

1. **Conexión con Spotify e MB:** configuración da autenticación e conexión coas bases de datos.

```
1 sp.set_auth(token_info['access_token'])
2 musicbrainzngs.set_useragent("Example music app", "0.1",
   "http://127.0.0.1:5000")
```

2. **Conexión coa base de datos propia:** chamada á función deseñada.

```
1 cursor, connection = bd.connect_db()

1 def connect_db():
2     try:
3         # Conecta base de datos
4         connection = psycopg2.connect(
5             host="localhost", database="tfg_database",
6             user="postgres",
7             password="postgres", port="5432"
8         )
9         cursor = connection.cursor()
10        return cursor, connection
11
12    except (Exception, psycopg2.Error) as error:
13        print("Error while connecting to PostgreSQL", error)
```

3. **Busca do artista polo seu nome preciso:** a partir do nome que escribe o usuario, búscase na BBDD de Spotify. O criterio empregado é quedarse coa primeira opción. No resultado desta solicitude, a maiores da obtención do id, tamén devolve atributos de interese como a popularidade, unha imaxe e os seus xéneros.

- En Spotify e obtención do seu ID:

```
1 sp.search(q='artist:' + name_artist, type='artist')
```

4. **Busca en MB de artistas e grupos:** igual que para Spotify, emprégase o nome que dá o usuario para buscalo na BBDD e recuperar a primeira resposta devolta. O obxectivo principal é obter o id do intérprete na BBDD e o seu tipo: “group” ou “person”. Despois, dependendo disto, tomarase un camiño ou outro á hora de recompilar a información.

- Buscar un artista ou grupo polo seu nome (co obxectivo de obter o id):

```
1 musicbrainzngs.search_artists(artist_name)
```

- **Persoa:** co id é posible obter o nome artístico, o nome real, o ano de nacemento e o país.

- Tras obter o identificador do artista obtense información máis completa: o nome artístico, o nome real da persoa, o país e a data de nacemento . Especificamos que queremos que inclúa información acerca dos seus nomes (“alias”) no resultado.

```
1 musicbrainzngs.get_artist_by_id(artist_info['artist']['id'],
    includes=["aliases"])
```

- **Grupo:** co id é posible obter o nome, data de inicio, data de fin, o país, a imaxe, e recadar os seus integrantes e, en consecuencia a información de cada membro individual (a mesma que para un artista tanto en MB como en Spotify), ademais do rol que tiveron e o período de tempo que pertenceron ó grupo. Especificase que o resultado debe incluír as relacións cos artistas:

```
1 musicbrainzngs.get_artist_by_id(group_id,
    includes=['artist-rels'])
```

- Prodúcese a integración correspondente. No caso dos grupos é preciso tamén gardar as relacións con cada un dos seus integrantes, e consecuentemente a información de ditos artistas:

```
1 if individual:
2     bd.funcion_insert(cursor, connection, bd.insert_artist,
    info_art_mb)
3 else:
4     bd.funcion_insert(cursor, connection, bd.insert_group,
    info_art_mb)
5
6     for i, integrante in enumerate(historico_pertenencias):
7         historico_pertenencias[i]['id_grupo'] =
    info_art_mb['id_spoty']
8         _, info_art_mb4, _, xeneros_set4, art_xen4, _ =
    completa_artista(sp,
```

```

musicbrainzngs.get_artist_by_id(integrante['id_artista'])['artist'] ['name']
historico_pertenencias[i]['id_artista'])
9
    bd.funcion_insert(cursor, connection, bd.insert_artist,
10 info_art_mb4)
    bd.funcion_insert(cursor, connection,
11 bd.insert_pertenencias, historico_pertenencias[i])
12
13 art_xen += art_xen4
14 xeneros_set.update(xeneros_set4)
15

```

5. **Busca en Spotify de todos os álbums publicados polo artista:** co id de artista de Spotify realízase unha consulta e obtense unha lista con todos os seus álbums. Para cada un obtéñense: o id, o nome, o tipo (*single*/album), imaxe, popularidade, ano de publicación e artistas colaboradores.

- Co identificador do artista obtemos todos os seus álbums (álbums, *singles* e *EP*). Neste paso consíguese o identificador de Spotify, o nome, tipo e imaxe.

```

1 sp.artist_albums(info_art_mb['id_spoty'], include_groups=('single,
album'))

```

- Extracción de máis información dos álbums: a popularidade e a data de lanzamento.

```

1 sp.album(album_id_sp)

```

- Para os colaboradores novamente obterase toda a información de artista tanto en **MB** como en Spotify. Código de mostra:

```

1 for album in items_albums:
2     album_name = album['name']
3
4     album_mb=[]
5     album_mb['id_spoty']=album_id_sp = album['id']
6     album_mb['name'] = album_name
7     album_mb['tipo'] = album['album_group']
8     album_mb['imagen'] =album['images'][2]['url']
9
10    album_especifico = sp.album(album_id_sp)
11    album_mb['popularity'] = album_especifico['popularity']
12    album_mb['f_publi'] =
mb.extract_year(album_especifico['release_date'])
13    bd.funcion_insert(cursor, connection, bd.insert_album, album_mb)

```



```

14     if individual:         bd.funcion_insert(cursor, connection,
bd.insert_publica_artista, {'id_artista':
info_art_mb.get('id_mb'),'id_album':album_id_sp})
15     else:                 bd.funcion_insert(cursor, connection,
bd.insert_publica_grupo, {'id_grupo':
info_art_mb.get('id_spoty'),'id_album':album_id_sp} )
16
17     for artista in album_especifico['artists']:
18         if artista.get('id') != info_art_mb['id_spoty']:
19
20             individual2, info_art_mb2, _, xeneros2, art_xen2,
group_xen2 = completa_artista(sp, artista.get('name'))
21             if individual2 == 'Error':
22                 pass
23             elif individual2 and info_art_mb2:
24
25                 bd.funcion_insert(cursor, connection,
bd.insert_artist, info_art_mb2)
26                 bd.funcion_insert(cursor, connection,
bd.insert_publica_artista, {'id_artista':
info_art_mb2.get('id_mb'),'id_album':album_id_sp})
27                 art_xen += art_xen2
28             elif info_art_mb2:
29
30                 bd.funcion_insert(cursor, connection,
bd.insert_group, info_art_mb2)
31                 bd.funcion_insert(cursor, connection,
bd.insert_publica_grupo, {'id_grupo':
info_art_mb2.get('id_spoty'),'id_album':album_id_sp})
32                 group_xen += group_xen2
33                 xeneros_set.update(xeneros2)

```

6. Busca de todas as cancións para cada álbum mediante a función de Spotify

- Busca das cancións que pertencen ao álbum a través do seu identificador e obtención dos ids de cada unha.

```
1 sp.album_tracks(album_id)
```

Delas, sabendo o seu id, extráense:

- Nome, duración, popularidade e ano de publicación. Para os colaboradores novamente se obterá toda a información de artista tanto en [MB](#) como en Spotify.

```
1 sp.track(track_id)
```

- Información máis técnica dos sons da track: *valence*, *enerxía*, acústica, instrumentalidade, *danceability*.

```
1 sp.audio_features(track_id)
```

7. **En MB como información complementaria para as *tracks***: do mesmo xeito que se fixo para os álbums, empréganse o nome do artista e da canción (devolto no paso anterior en Spotify). Obtéñense o identificador, o lugar de gravación, a discografía e todos os traballadores participantes na realización cos seus roles ou instrumentos:

- Busca do id de MB asociado ao work polo nome da pista e do artista:

```
1 musicbrainzngs.search_works(artist=artista, work=cancion,
  limit=0).get('work-list', [])
```

- Co id obtéñense algúns dos traballadores que participaron na súa elaboración

```
1 musicbrainzngs.get_work_by_id(work_id,
  includes=['artist-rels'])['work']['artist-relation-list']
```

- Busca do id de MB asociado á gravación da pista:

```
1 musicbrainzngs.search_recordings(artist=artista, recording=cancion,
  limit=1, offset=None)['recording-list']
```

- Tras obter o id de MB asociado á gravación da pista, obtense máis información asociada á discografía, produción, músicos, mixers, instrumentos e lugares de gravación:

```
1 musicbrainzngs.get_recording_by_id(id_track,
  includes=['label-rels', 'artist-rels',
  'place-rels'])['recording']
```

- Trata de obter o nome real do lugar a través do seu identificador:

```
1 musicbrainzngs.get_place_by_id(place_id, includes=['area-rels'],
  release_status=[], release_type=[])
```

- Obter información sobre áreas:

```
1 musicbrainzngs.get_area_by_id(area_id, includes=['area-rels'])
```

Exemplo da extracción da información dos traballadores contidos no *work*:

```
1 try:
2     works =musicbrainzngs.search_works(artist = artista, work=
  cancion, limit=0).get('work-list', [])
```

```

3     if works:
4         work_id=works[0].get('id')
5         work_info= musicbrainzngs.get_work_by_id(work_id,
6             includes=['artist-rels'])['work']['artist-relation-list']
7
8         #print(work_info)
9         for traballa in work_info:
10            traballador_id = traballa['artist'].get('id')
11            #print(traballador)
12
13            rol.append({'id_track': None,      'id_trab':
14                traballador_id,      'rol': traballa.get('type') })
15            if not any(trab['id_mb'] == traballador_id for trab in
16                traballador):
17                traballador.append({'id_mb': traballador_id,
18                    'name': traballa['artist'].get('name'), 'id_existartist': None})
19
20 except KeyError:
21     pass

```

8. Insírese a información de cada track e, no caso de existir, os seus correspondentes traballadores, roles e instrumentos. Repítese o mesmo proceso para as colaboracións das cancións que dos álbums. Exemplos:

```

1 if audio_features:
2     info_track_mb['valence'] =audio_features['valence']
3     info_track_mb['energy'] = audio_features['energy']
4     info_track_mb['acustica'] =audio_features['acousticness']
5     info_track_mb['instrumentalidad']
6     =audio_features['instrumentalness']
7     info_track_mb['danceability']=audio_features['danceability']
8     bd.funcion_insert(cursor, connection, bd.insert_track,
9         info_track_mb)
10
11 for t in traballador:
12     bd.funcion_insert(cursor, connection, bd.insert_trab, t)
13
14 for r in rol: bd.funcion_insert(cursor, connection, bd.insert_rol,
15     r)
16
17 for i in instrumental: bd.funcion_insert(cursor, connection,
18     bd.insert_instrumental, i)

```

9. Ao longo de todo o procedemento fôronse agrupando os xéneros e as súas correspondencias cos artistas e grupos para evitar chamadas repetitivas, que agora ao final serán

engadidos segundo corresponda.

```
1 xeneros = [{'nome': xenero} for xenero in xeneros_set]
2     for x in xeneros: bd.funcion_insert(cursor, connection,
3     bd.insert_xeneros, x)
4     for ax in art_xen: bd.funcion_insert(cursor, connection,
5     bd.insert_art_xen, ax)
6     for gx in group_xen:bd.funcion_insert(cursor, connection,
7     bd.insert_group_xen, gx)
```

10. Finalízase co peche da conexión.

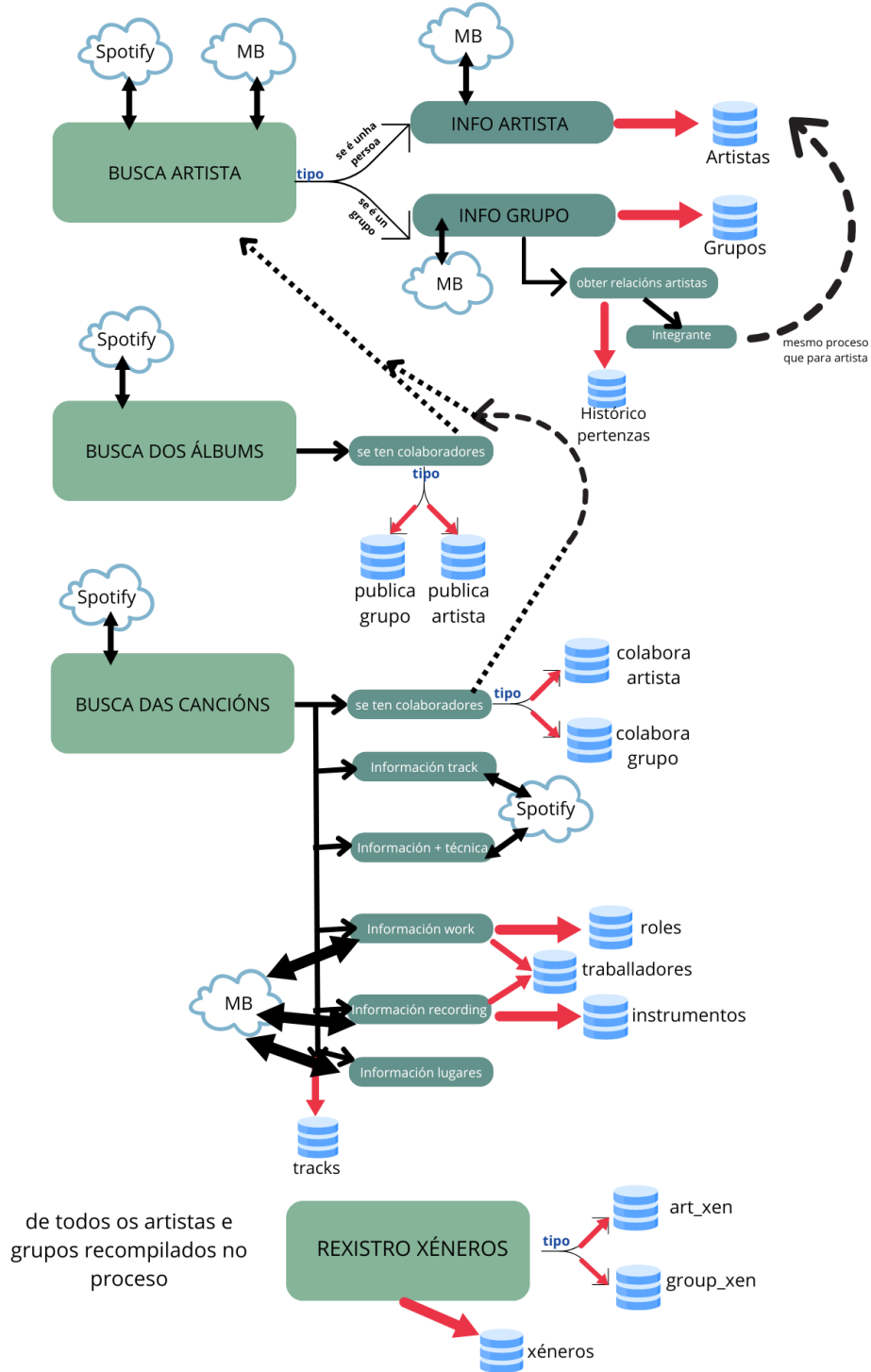


Figura 6.1: Extracción e carga da información na base de datos

Explotación do Data Warehouse

Neste capítulo explórase a explotación do **DW**: os gráficos e análises obtidos a partir dos datos.

Para a realización de análises, 35 artistas foron engadidos directamente á **BBDD**, dos cales 9 correspóndense a unha agrupación. Os criterios empregados polos cales se cargou esta información foron a busca dunha variedade de xéneros, procedencia, época e popularidade. Desta forma, existen intérpretes internacionais como Shakira, coñecida pola súa ampla traxectoria e influencia global; antigos como Luis Miguel, que ten unha carreira consolidada ao longo dos anos e previa ao desenvolvemento das **BBDD**. Tamén hai grupos clásicos como ABBA, cun legado na música indiscutible, mais con pouca granularidade temporal debido á pertenza a épocas pasadas e artistas españois como Quevedo, dun xénero máis actual. Ademais incluíronse grupos máis pequenos como as Tanxugueiras, dos que se ten menos información. En consecuencia, todos os artistas relacionados con eles, a través de pertencenzas ou colaboracións, foron engadidos de xeito indirecto, unicamente coa información biográfica básica.

Cabe destacar que esta é unha ferramenta e que, para o seu uso óptimo, máis información debe de ser engadida, como se explica no capítulo [A.1](#).

Despois de obter e preparar os datos, así como de construír e cargar o almacén, é o momento de levar a cabo a última etapa do proxecto. A explotación do **DW** é a fase na que se levan a cabo os estudos e análises dos datos mediante a realización de gráficos e **dashboards**. A creación de informes detallados axuda a identificar tendencias e patróns nos datos. Tamén permite que o coñecemento sexa máis accesible para os usuarios, agrupando de xeito intuitivo a información máis relevante en determinados temas. Estes gráficos son interactivos, aceptan o uso de filtros e seleccións, facendo máis aberta a participación e permitindo afondar nas

seccións de interese en campos concretos.

A ferramenta software empregada neste paso foi Tableau. A través desta plataforma realizouse a conexión coa **BBDD** PostgreSQL e definíronse varias fontes de datos coas súas relacións pertinentes en función do **dashboard** e contidos a tratar concretos. Tableau non soporta ben as relacións múltiples, polo que, dependendo do caso empregárase unha ou outra, adaptándose as necesidades do análise. Na figura 7.1 pódese ver como neste caso a fonte de datos foi feita arredor das *tracks*, co obxectivo de estudar as colaboracións e traballadores implicados.

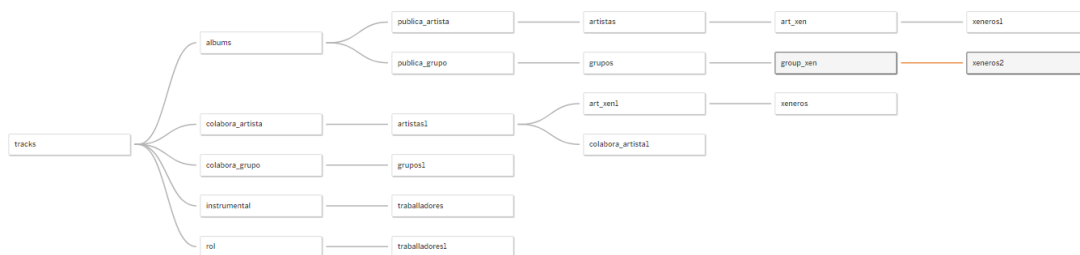


Figura 7.1: Fonte de datos de Tableau

Como xa se expuso con anterioridade, neste traballo recompílese información específica dos artistas, as súas obras, colaboracións e os participantes na produción das cancións. Esta información permite analizar diferentes vistas da industria musical, centrándose en varios dos puntos claves. Entre eles, para os intérpretes ou artistas seleccionados lograr unha perspectiva completa e minuciosa do percorrido musical. Conteñen tanto información máis xeral, como algúns detalles técnicos. Baseándose niso, as análises obtidas foron as seguintes:

7.1 Grupo

Un dos principais intereses do proxecto é poder estudar as agrupacións musicais e os seus integrantes. Os grupos teñen un período de vida limitado; en ocasións, estes disólvense e os artistas deciden emprender carreiras en solitario. Noutros casos, os membros compaxinan estas carreiras durante o tempo de actividade, tendo varias traxectorias dentro da industria.

No *dashboard* pódese explorar tanto o percorrido individual como o grupal. É importante que, para obter toda a información, como o reconto das publicacións solitarias, todos os artistas relacionados estean engadidos na **BBDD**. A continuación, detállanse máis cada unha das gráficas que o compoñen.

Existe un filtro desplegable no panel da dereita para seleccionar ó grupo (só un é posible).

Para a información do grupo escollido:

- Amósase a imaxe que os representa xunto co seu nome e as datas de inicio e fin da súa vida.
- Paralelamente móstrase ao longo do tempo a cantidade de obras publicadas ('albums'), facendo unha diferenza entre álbums (longos) e *singles/EP* (curtos). Non se conta as *tracks* que ten cada álbum e cales pertencen a un *single*. Revela o nivel de actividade dun grupo e cal é o seu produto principal. Algúns artistas publican máis *singles*, que son máis rápidos de facer e consumir. Ao ser unha obra pequena, é máis dixerible e fácil que destaque. Por outra banda, os álbums son pezas máis complexas, que necesitan maior tempo de elaboración e é complicado que todas as súas cancións teñan o mesmo protagonismo.
- Os xéneros de cada un dos integrantes son contabilizados, o cal indica se seguen todos un perfil musical similar ou, por outra banda, é unha mestura chea de diversidade.
- Para cada un dos integrantes indícase o seu nome artístico, papel que desenvolveu no grupo, o tempo durante o que foron membros oficiais, e, a maiores, a súa popularidade de xeito particular. Mediante esta comparación pódese ver se hai algúns membros que triunfan máis que outros segundo os criterios definidos por Spotify, é dicir, a súa popularidade é maior. Os papeis dos artistas son importantes xa que amosan se pertencen ao grupo 'orixinal', se eran cantantes ou se se adicaban a algún instrumento en particular.
- O relato de todas as *tracks* que publicou cada integrante ao longo do tempo. Este gráfico pode revelar dende cando empezaron a traballar de xeito individual, aqueles artistas que máis cancións teñen no mercado e a súa constancia.

É posible interactuar coa gráfica de integrantes, pulsando no artista de interese e, en consecuencia, o relato das publicacións solitarias e os xéneros cinguiranse ao individuo concreto.

Na figura 7.2 podemos ver unha mostra desta análise. Neste caso o grupo escollido é One Direction. Publica tanto *singles* (recopilacións curtas) coma álbums. Pódese entender que, como carta de presentación, comezou con *singles*, desta maneira, ao ser uns artistas xa coñecidos, pódese facer unha avaliación rápida (prodúcense e publícanse en relativamente pouco tempo) da aceptación que pode ter no mercado e cal é o seu público obxectivo. Tivo un inicio forte co que empezou a súa carreira lanzando 4 *singles*. Ao final rematouna co seu derradeiro álbum. Estivo conformado por 5 homes e todos deixaron de formar parte do grupo no mesmo ano. Os seus integrantes, de xeito individual, adícanse a xéneros semellantes, todos

ao 'pop' xenérico e un máis concretamente categorizado como 'uk pop'. O membro menos activo pero máis popular é 'Harry Styles'. En xeral, todos gozan de recoñecemento e seguen desenvolvéndose no mundo da música.

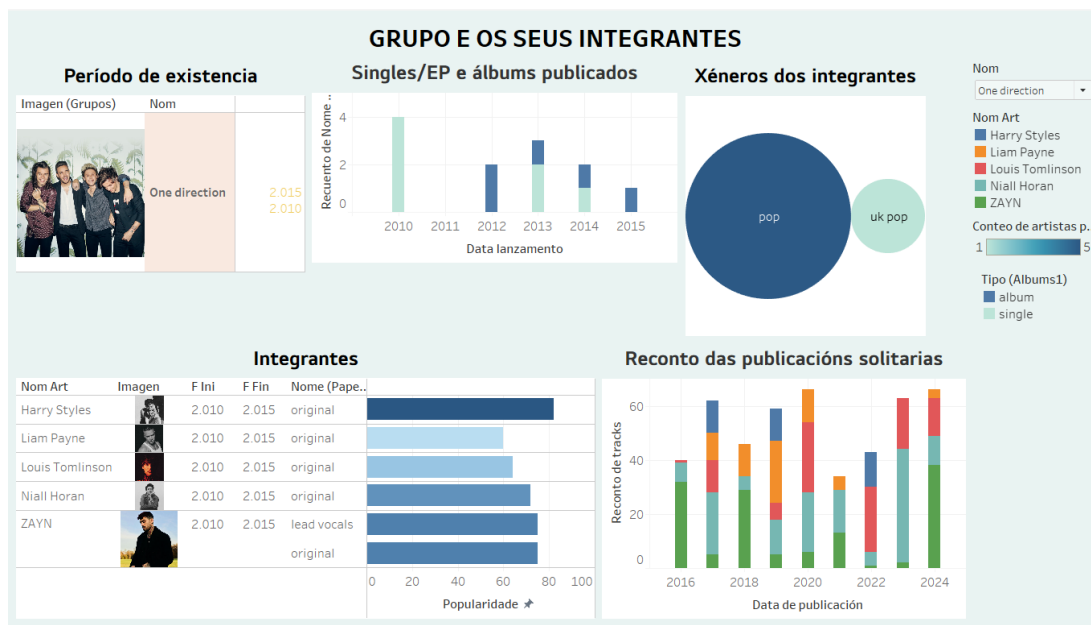


Figura 7.2: Dashboard con información relativa dun grupo e os seus integrantes

7.2 Artista

Outro dos pilares da música e da análise é o tipo de intérpretes máis común, os artistas. Estes están asociados a unha única persoa, teñen a capacidade de expresar a súa visión creativa de maneira máis persoal e íntima. A súa traxectoria pode variar amplamente, dende aqueles que se especializan nun único xénero ou experimentan con múltiples estilos.

A información dispoñible no *dashboard* ofrece un resumen visual do seu perfil ao longo do tempo, con posibilidade de concretar máis en certos aspectos. O artista pode ser seleccionado na esquina superior do panel dereito (únicamente é posible seleccionar un).

- Primeiro, móstrase o nome completo real da persoa vinculada á figura artística, acompañado dunha foto e do seu ano de nacemento.
- A popularidade amósase nun rango do 1 ao 100, segundo Spotify. É importante para facerse unha idea do impacto e recoñecemento do que está dotado.

- Tamén aparecen os xéneros nos que se encadra e o país de procedencia marcado nun mapa de xeito máis ilustrativo.
- Como xa se explicou no apartado anterior é posible publicar dous tipos de álbums (coleccións), os álbums propiamente ditos e os *singles*. Expóñense a porcentaxe de todas as *tracks* que se corresponden ou ben a un álbum ou a un *single*. Nela pódese ver as preferencias do artista, se prefiren o desenvolvemento dun concepto máis complexo ou simple. Pode chegar a ser máis representativa do traballo oculto que simplemente a contabilización de álbums ou *singles* directamente, xa que é un álbum ten moitas máis *tracks*. É unha gráfica interactiva, o feito de clicar nun tipo de álbum afectará aos gráficos seguintes (mostra dos álbums dese tipo, ano de publicación e cancións).
- Lista dos álbums (*singles* e álbums) do artista xunto coa súa popularidade e duración (suma da duración en milisegundos de todas as súas *tracks*). Os álbums tamén son interactivos e, ao escollerse un, marcarase o ano concreto no que foi publicado.
- Reconto das publicacións para cada ano ó longo do tempo, reflectindo a produtividade e constancia. Se se selecciona un álbum concreto indicará o seu ano específico de lanzamento.
- Por último, aparecen todas as cancións publicadas polo artista, ordenadas por popularidade. Como xa se dixo, ás cancións mostradas aféctanlle os filtros anteriores sendo 3 opcións posibles; ver todas, seleccionar *single* ou álbum ver solo as que se corresponden ao tipo concreto ou as que pertencen a un álbum(*single*/álbum) concreto que foi seleccionado polo seu nome.

Concretamente na imaxe 7.3, aparece a información de Beyoncé, de popularidade moi alta, xa que é unha artista internacional e de fala inglesa. É orixinaria de EEUU e os seus xéneros principais son “pop” e “r&b”. Ten máis cancións que pertencen a un álbum que a un *single*, o cal é habitual xa que os álbums conteñen máis. Conta unha ampla carreira musical dende inicios dos 2000 e unha grande cantidade de publicacións as súas costas, tanto cancións como álbums. O seu período de actividade máis forte foi entre o 2005 e 2012, sendo unha das protagonistas da época. Despois baixou o ritmo, máis nos últimos anos esta en aumento. Ten numerosas cancións moi exitosas como “Bodyguard” ou “Texas hold'em”.

7.3 Especificacións *tracks* dun álbum

As *tracks* ou cancións son outro dos elementos fundamentais na análise musical. Cada canción representa unha unidade de creatividade que pode variar enormemente dependendo do estilo, xénero e duración. Non só reflicten o talento dos intérpretes e implicados na súa produción, senón que tamén mostran as tendencias musicais do momento. Poden ser tanto obras individuais ‘*single*’, coleccións pequenas EP ou compilacións máis grandes ‘álbum’. Spotify categoriza *single*/EP ou álbum dentro do concepto “álbum” máis xeral para referirse a estas compilacións.

No panel da dereita seleccionárase o artista e o álbum do que se queren analizar as súas *tracks*. No *dashboard* aparecen os gráficos seguintes:

- Do álbum seleccionado amósase o seu nome, ano de publicación, se se corresponde a un álbum/*single*, o número de cancións que contén e a súa popularidade.
- Debaixo do gráfico atópanse todas as cancións co seu nome, lugar onde foron gravadas e a súa duración en milisegundos. Este panel é interactivo e nel pódese seleccionar unha canción concreta. Se non se indica ningunha a información reflexada nos seguintes gráficos é de carácter xeral para o álbum enteiro.
- 5 características de audio aparecen para as *tracks*, explicadas na sección 5.2.3, nun rango de valores do 0 ao 1.
- Os roles/traballo de produción implicados na elaboración da canción, por número de persoas implicada. É interactivo. Dos roles anteriores indícanse as persoas implicadas.
- Os instrumentos, ao igual cos roles, aparecen tendo en conta por canta xente foron empregados na track. É interactivo. Tamén se mostran os músicos que tocaron os instrumentos.

Na imaxe 7.4 vese un exemplo do *dashboard* mencionado. As *tracks* pertencen ao álbum ‘Motomami’, lanzado no 2022 conformado por 16 cancións. Practicamente todas foron gravadas en Los Ángeles e dúas cabe descatacar que dúas son máis curtas co resto. Isto pódese deber a que son *tracks* máis experimentais, rompendo ca clásica duración de radio máis comercial. O álbum, en xeral, é bastante bailable e conta cunha variedade de instrumentos na súa produción. O tipo de traballadores que máis requiren as súas tracks son os compositores e produtores, algúns deles adicanse a diferentes tarefas. Os instrumentos máis frecuentes no

álbum son a batería e o piano, empregándose 7 e 4 veces. Ao concretar en determinadas cancións apréciase que en todas se emprega algún instrumento, mais nunca máis de 5 diferentes.

7.4 Colaboracións dun artista

As colaboracións entre artistas son un elemento enriquecedor da industria musical. Estas asociacións permiten a mestura de estilos, xéneros e influencias, creando pezas únicas que van máis alá das capacidades individuais dos artistas. Poden ser entre artistas en solitario, grupos, membros dun grupo, etc. As colaboracións tamén poden impulsar a popularidade dos artistas, permitíndolles chegar a novos públicos e mercados.

Analizalas é importante para entender as redes de relacións na música. No *dashboard* móstrase información detallada sobre as colaboracións dun artista. No panel da dereita pódese escoller entre que sexa máis concreta, referida a álbums, *singles* ou ambos:

- Reconto das colaboracións nas súas cancións publicadas ao longo do tempo para un artista. Diferéncianse se pertencen a un álbum ou *single*. É máis habitual que os álbums sexan máis ‘proprios’ dun artista, é dicir, que conteñan máis obras solitarias. Por outro lado, os *singles* son máis frecuentes como colaboracións puntuais entre diferentes intérpretes que deciden crear un elemento común. É posible que no inicio da carreira se leven a cabo máis colaboracións para tratar de ‘despegar’ chegando a un público maior e afín (ao doutros artistas similares). É interactiva, ao seleccionarse, ou ben as colaboracións dun *single* ou dun álbum de determinado ano mostraranse:
 - * Os álbums/*singles* publicados nese ano con colaboracións.
 - * As *tracks* que se correspondan e os colaboradores.
 - * Os países.
- Para cada álbum (*single*/álbum) reconto das colaboracións totais, é dicir, as colaboracións de todas as *tracks* que as conforman. É interactiva, existindo a posibilidade de pulsar nun dos álbums, e filtrar a información correspondente as outras gráficas. Afecta a:
 - O reconto de colaboracións: mostra únicamente o álbum e, polo tanto indica o ano no que se publicou.
 - Os países dos colaborados do álbum.
 - As *tracks* dos álbums que se mostran.

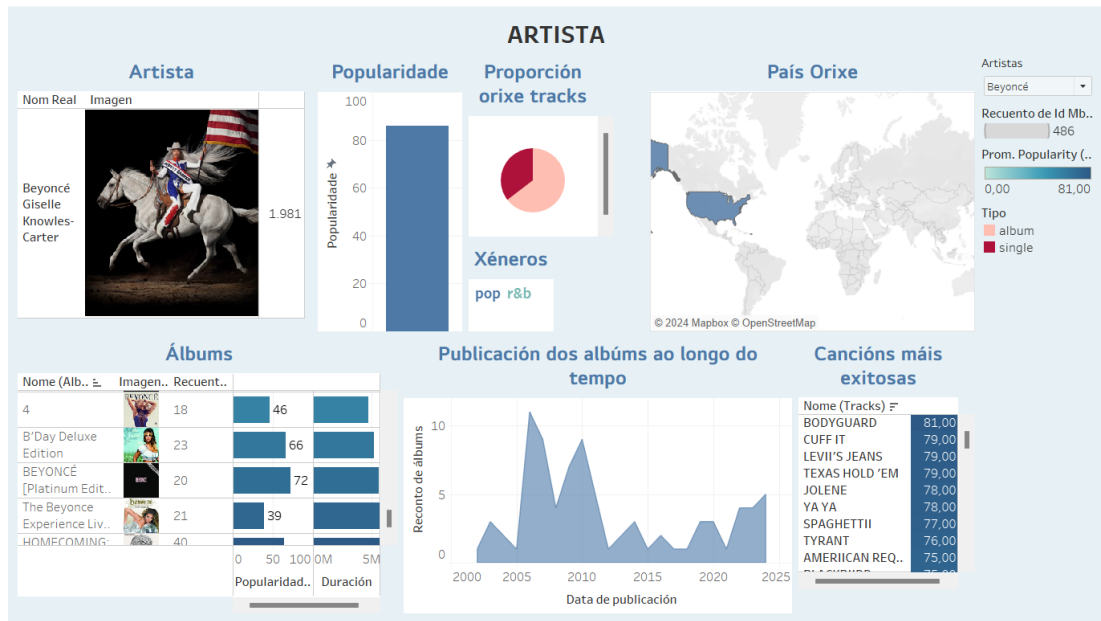


Figura 7.3: Dashboard que engloba información dun artista

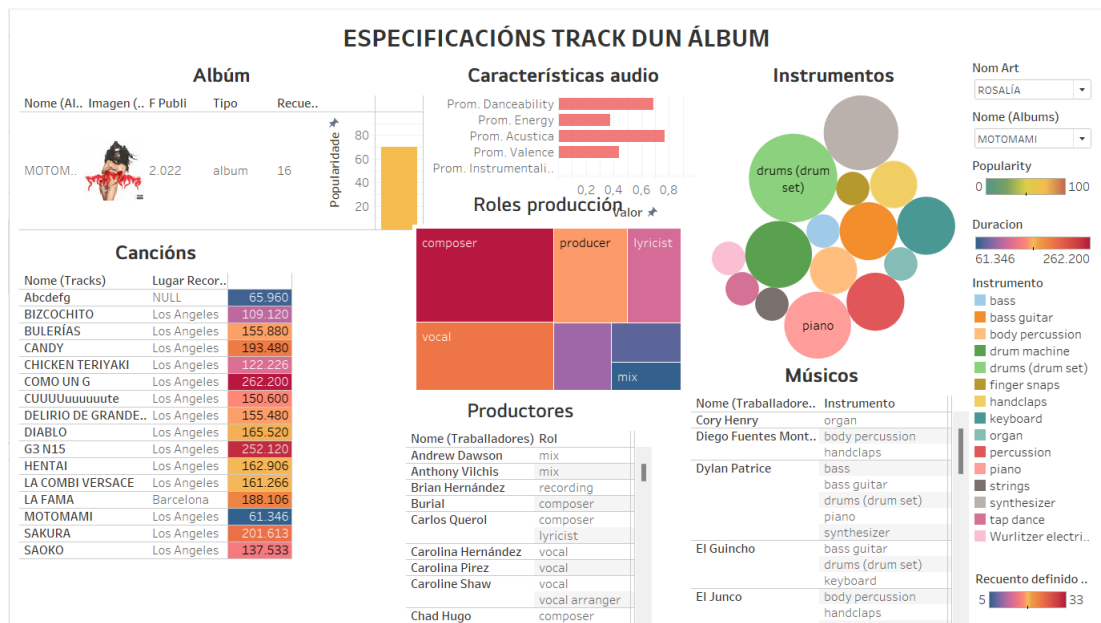


Figura 7.4: Dashboard con conxunto de gráficos informativos dunha track

- Os países dos colaboradores individuais marcados nun mapa. É de utilidade para saber por qué zonas ou continentes se move máis. No caso dos artistas angloparlantes habitúan facer máis colaboracións con EEUU, Canadá, Reino Unido ou Australia, entre outros. Para os latinos, en xéneros como o reggaeton ou máis urbanos é frecuente traballar con artistas de América ou España. É interactiva, marcando un país filtrará e mostrará as colaboracións cos artistas individuais desa procedencia: cantas e en que anos, os álbums e tracks.
- Amósanse as cancións dos álbums e os artistas que colaboraron nelas cunha foto, o seu nome e número de colaboracións nas que participou en total con calquer artista (é importante aclarar que se contan aquelas que existen na **BBDD**).
- Tamén aparecen as cancións con colaboracións, pero neste caso asociadas aos grupos.

Por último, no panel da dereita, tamén se pode pinchar nun artista colaborador (individual) concreto e ver cando colaborou, en que álbums e en que anos.

Para o caso concreto da imaxe 7.5, amósase información das colaboracións do artista Bad Bunny ó longo da súa carreira musical. Comezou no ano 2016 e, dende ese momento, engadiu a moitos colaboradores ás súas cancións. É unha técnica para poder darse a coñecer ao público do outro artista. Tamén, cando unha canción triunfa adoitan ser remix con outras persoas para sacarlle máis proveito. Unha vez xa instaurada a súa carreira pódese ver como o número se reduxo considerablemente. Ao ter máis prestixio pode ter máis criterio de selección de con quen colaborar. Prácticamente todos os álbums conteñen algunha colaboración, a maioría un par, existindo un par de excepcións que chegan ata os 12 intérpretes implicados. A procedencia dos principais artistas colaboradores son EEUU, Cánada, América latina, España e Reino Unido, é dicir, as principais potencias musicais de fala inglesa e hispana. Isto reflexa un impacto forte, chegando a expandirse a outras culturas. Cada canción non adoita ter máis de 3 colaboradores. Na gráfica 7.6 mostrase un exemplo filtrando polo artista colaborador específico Nicky Jam. Neste caso, marca o país do que procede e as 3 cancións que fixeron xuntos ambos, das cales dúas foron publicadas no mesmo ano.

7.5 Instrumentos

Os instrumentos musicais son un elemento esencial na produción musical, proporcionando a base sonora sobre a que se constrúen as cancións. A variedade dos empregados pode



Figura 7.5: Dashboard coas colaboracións dun artista



Figura 7.6: Dashboard coas colaboracións específicas de Bad Bunny con Nicky Jam

influír significativamente no estilo e melodía. A elección de instrumentos e intérpretes pode revelar detalles sobre a versatilidade dos músicos, a innovación e a experimentación da obra.

No *dashboard* pódese acceder a información detallada sobre frecuencia de uso de cada instrumento e como contribúen ao son xeral das súas producións. Esta análise é útil para poder comprender a complexidade e a riqueza das composicións musicais, así como para apreciar a diversidade e creatividade no uso de recursos instrumentais na industria musical.

- Reconto dos instrumentos empregados nas cancións máis comúns na base de datos. Tense en conta, aínda que fose na mesma gravación, o feito de que fose interpretado ou empregado por intérpretes diferentes. É interactiva, podendo seleccionar un instrumento concreto que afectará en consecuencia aos seguintes gráficos.
- Cantos instrumentos se empregan ao longo dos anos e cantas veces, para as cancións das que se dispoñe información na BBDD. No caso de que se seleccione un instrumento no gráfico anterior, aparecerá o reconto dese en concreto.
- Promedio da instrumentalidade e popularidade das cancións que conteñen ditos instrumentos.
- Os músicos, intérpretes ou produtores que traballaron co instrumento ou son nalgunha canción. Aparece o número de instrumentos diferentes que empregaron ao longo da súa carreira e en cantas cancións participaron. Hai que ter en conta que os *remixes* suben os valores das cancións. É interactiva e condiciona a todas as gráficas do *dashboard*.
- Promedio anual para todas as tracks, ou para aquelas que conteñen un instrumento concreto, se é seleccionado, dos seguintes valores: a instrumentalidade (probabilidade de que conteña solo melodía sen voces), acústica (unha puntuación alta indica uso de elementos acústicos e menos electrónica) e enerxía (mestura de intensidade, actividade, volume..).

No *dashboard 7.7*, amósase a información concreta para o instrumento seleccionado, o teclado. Pódese ver como foi adquirindo popularidade co paso dos anos, sendo máis empregado nas cancións. Tamén é de interese analizar os cambios dos valores das cancións co paso do tempo. Pódese apreciar como aquelas publicacións máis vellas eran máis acústicas e, co paso dos anos foise reducindo considerablemente, impulsado pola aparición de novos instrumentos e tendencias. A bailabilidade é menor especialmente na década dos 70. Para as tracks da base de datos a instrumentalidade mantívose máis ou menos constante, é dicir, teñen letra.



Figura 7.7: Dashboard con conxunto de gráficos informativos dunha track

Conclusiones

Neste capítulo sintetízanse os aspectos finais do proxecto como son as funcionalidades máis importantes, a súa utilidade ao mundo real e o grado de cumprimento dos obxectivos. Tamén se tratan as posibles aplicacións e melloras que realizar.

Tal e como se explicou no inicio deste traballo, hoxe en día existe unha inmensa cantidade de datos en todos os sectores, incluído o musical, os cales están dotados de gran relevancia. A pesar disto, para que realmente sexan de utilidade é preciso revelar a información oculta neles coa finalidade de poder extraer coñecemento. Por iso, é necesario deseñar ferramentas que axuden aos usuarios na súa xestión, almacenaxe e interpretación. Neste documento expuxéronse os fundamentos necesarios para comprender os conceptos básicos do mundo das bases de datos, así como as infraestruturas requiridas para a súa creación. Tamén se detallaron os procesos e fases necesarias a seguir para desenvolver unha proposta de valor neste ámbito. Os obxectivos propostos desenvolvéronse con éxito, en concreto a creación de perfís completos, tanto dos artistas, os grupos, produtores e outros tipos de intérpretes. Foi preciso acadar unha serie de submetas durante o proceso:

- A organización e planificación das etapas para levar a cabo o plan definido.
- O deseño e creación dun Data Warehouse: implicou realizar un deseño adaptado ás necesidades da industria musical que represente todas as conexións ocultas entre os seus piares fundamentais, dende as unidades máis básicas como as cancións, a súa produción, interpretación e a maneira de agrupalas.
- Extracción dos datos das súas respectivas fontes mediante a interacción con interfaces externas e a limpeza xunto coas adecuadas comprobacións para a concordancia da

información.

Otra das metas propostas básicas é a obtención de análises de interese para un perfil variado de usuarios. Usáronse ferramentas de BI para o estudo e visualización dos datos, como foi o caso de Tableau, facilitando a creación de paneis interactivos. Neles é posible examinar as seguintes cuestións:

- Evolucións ao longo do tempo: comparar e obter conclusións das métricas e valores obtidos para diferentes elementos ao longo do tempo. Por exemplo, no caso do uso de instrumentos, a acústica das cancións ou número de lanzamentos.
- Exploración das colaboracións realizadas: análise de todas as relacións existentes entre artista a través das súas obras.
- Comparativas: recompilación dunha serie de atributos e posibilidade de realizar comparacións mediante diferentes tipos de gráficos, mapas e cores. Como é para os casos da popularidade de artistas, álbums e cancións, métricas técnicas das *tracks* ou recontos de xéneros e roles.
- Apreciación dos traballos na produción musical: identificación, análise e posta en valor das labores relacionadas coa produción das cancións como os traballadores implicados.

Por último, simplificouse a interacción do usuario coa base de datos para chegar a perfís inexpertos, creando unha interface básica coa única funcionalidade de engadir máis información ao almacén. Isto reduce os erros posibles no procedemento. En conclusión, acadouse a funcionalidade principal deste proxecto: creouse unha ferramenta robusta para a introdución de información relevante de datos musicais e a súa análise.

8.1 Trabajo futuro

Aínda que os obxectivos principais do proxecto foron acadados, é posible realizar unha serie de melloras sobre o traballo. A continuación, describíranse algunhas delas:

- Integrar información doutras bases de datos: podería ser interesante ter máis datos dos artistas e cancións. Isto podería incluír detalles máis específicos sobre os xéneros nos que se encadra cada canción, así como a evolución das súas reproducións ao longo do

tempo. Non só enriquecería o Data Warehouse con máis datos senón que permitiría realizar análises máis profundas e detalladas da evolución musical.

- Análises personalizadas: unha opción interesante sería conectar a conta persoal de Spotify coa base de datos. Deste xeito, en función dos artistas ou cancións máis escoitados poderíanse mostrar una serie de perfís e análises correspondentes. Esta funcionalidade permitiría ofrecer vistas personalizadas aos usuarios en función dos seus datos.
- Proceso de automatización: para manter unha base de datos rica e variada unha opción interesante é automatizar a adición de información ao almacén. Poderían realizarse as tarefas de extracción, limpeza e carga minimizando os erros humanos mediante scripts que exploren as plataformas existentes. Ademais reflectiría as últimas tendencias e cambios na industria.

Apéndices

Material adicional

Neste capítulo explicase información extra, como engadir información de novos artistas á base de datos ou contidos ilustrativos.

A.1 Adición de intérpretes

Ao longo de todo o proxecto mencionáronse os obxectivos principais a levar a cabo, como a creación de perfís completos ca información integrada das dúas fontes co fin de obter unha base de datos completa na que poder realizar análises. Para poder conseguir isto, é importante engadir o coñecemento de cada un dos artistas, a través do proceso ETL.

Deseñouse unha interface para facilitar o traballo de inserción. Esta fai de intermediaria entre o usuario e o arquivo de código principal. Permite a execución do mesmo a través dunha sinxela acción: a escritura do nome do artista nunha caixa e pulsar o botón, como se amosa na imaxe [A.1](#) (páxina 69). Esta interface non só simplifica o proceso de inserción para o usuario, senón que tamén minimiza os erros que poderían ocorrer cunha inserción manual e directa na base de datos. Está deseñada para ser intuitiva e accesible, deste xeito, os usuarios de diferentes niveis técnicos poden participar no proceso de adición de novos artistas.

Unha vez introducido o artista e pulsado o botón, a interface inicia o procedemento de extracción e inserción de datos, interactuando co código principal para recoller toda a información dispoñible sobre o artista nas diferentes fontes. O ficheiro principal, a súa vez, como se explicou na sección [2.2](#), comunicase con outros dous, aquel que se encarga da información de MusicBrainz e o das conexións ca base.

Este proceso inclúe:

- A información do artista ou grupo indicado, no caso dos grupos tamén dos seus respectivos integrantes e relacións que mantiveron.
- Os seus álbums, tracks e colaboracións, tamén o perfil básico dos intérpretes (grupos/artistas) implicados.
- Os roles e instrumentos implicados na produción da canción en relación cos seus traballadores.

Tras a execución, cando remate o longo proceso de extracción e insercións (dependendo do artista pode rondar a media hora) , o usuario recibirá unha confirmación de que os datos foron engadidos correctamente á base de datos, como se pode ver na imaxe [A.2](#) (páxina 69). No caso de que ocorrea algún problema mostrará 'Error'.

A.2 Borrador deseño dimensional

Durante o deseño dimensional definíronse varias aproximacións para tratar de representar o mellor posible os conceptos escollidos e as súas conexións, as cales, pouco a pouco, foron pulidas permitindo así que o deseño evolucionase ata o explicado no apartado 5.2. Nesta fase partiuse do bosquejo inicial, que se pode ver a continuación:

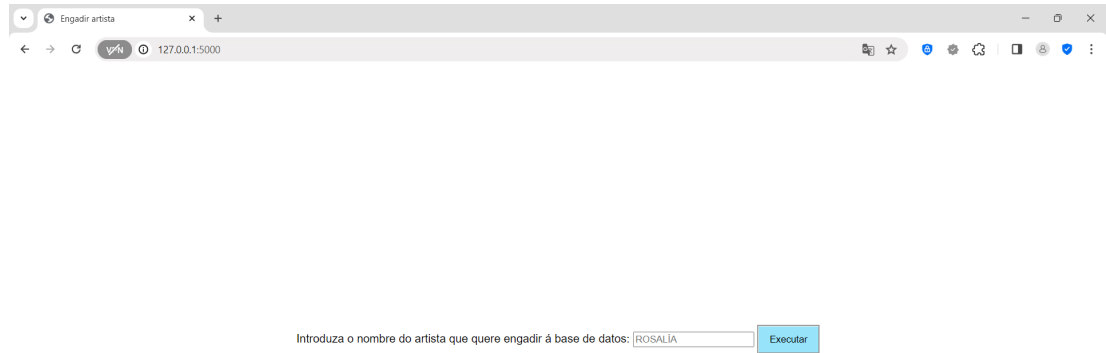
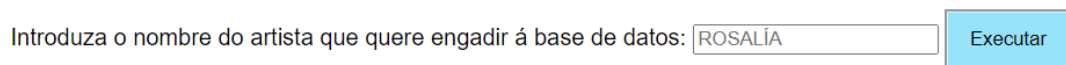


Figura A.1: Interface para a introducción de artistas



Resultado:

Datos completados para ROSALÍA

Figura A.2: Mensaxe tras a correcta introducción

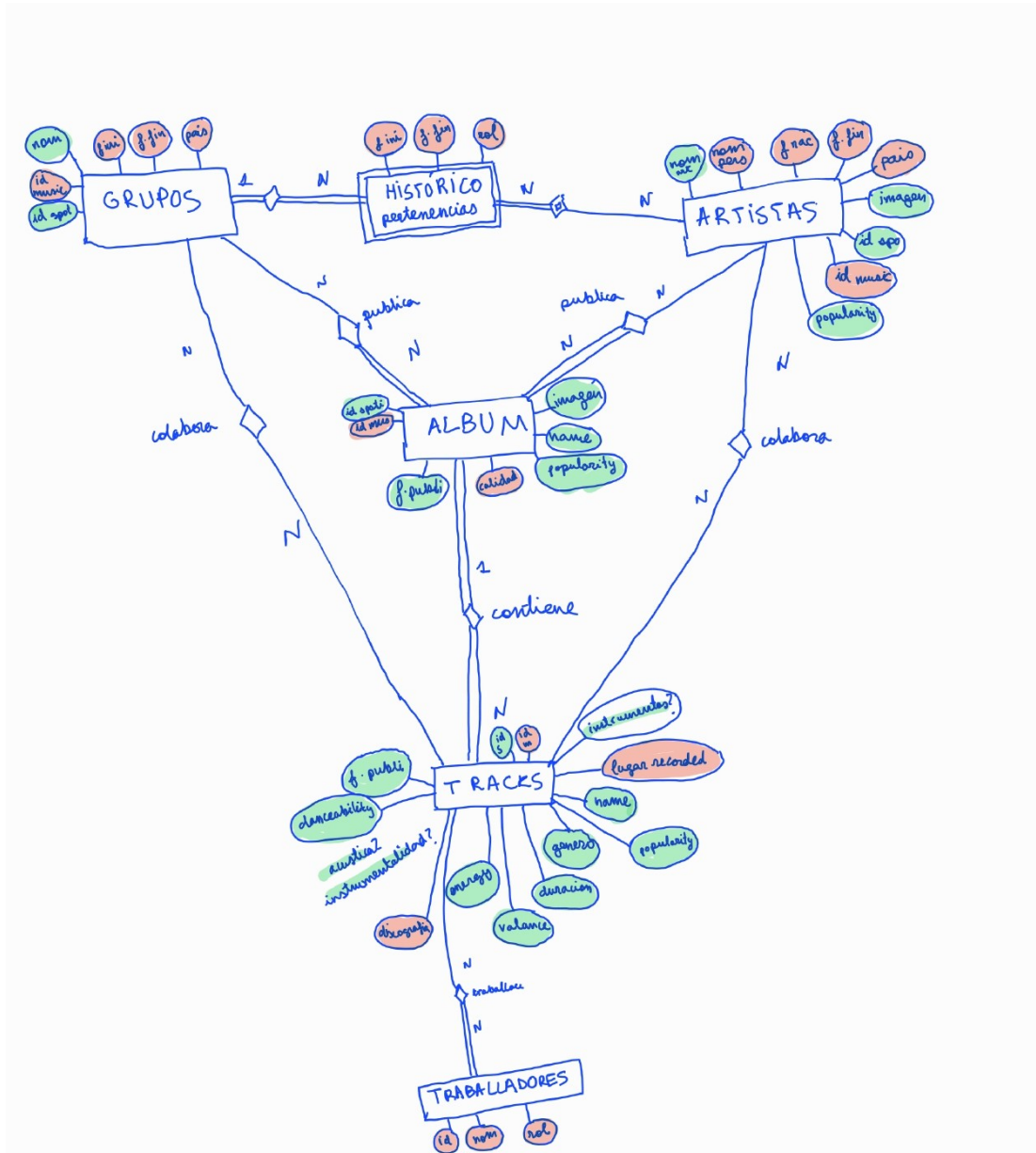


Figura A.3: Aproximación inicial do diagrama ER

Relación de Acrónimos

- ACID** Atomicity, Consistency, Isolation and Durability. 10
- API** Application Programming Interface. 3, 4, 7, 9, 10, 21, 31, 32, 34, 36
- BBDD** Bases de datos. 2, 3, 5, 10, 13, 22, 23, 25, 27, 28, 31–33, 37, 39–44, 51, 52, 59, 61
- BI** Inteligencia de Negocio. 1, 11, 14, 64
- DJ** disc-jockey. 24
- DW** Data WareHouse. 1, 5, 6, 13, 14, 20, 25, 40, 51
- EP** Extended Play. 35, 39, 45, 53, 56
- ER** Entidade Relación. iv, 23, 26
- ETL** Extract Transform Load. iv, 1, 3–6, 8, 13, 14, 21, 31, 32, 39
- HTTP** Hypertext Transfer Protocol. 35
- JSON** JavaScript Object Notation. 10
- MB** MusicBrainz. 2, 3, 10, 16, 20, 21, 27, 28, 31–34, 36, 39, 40, 43–47
- SQL** Structured Query Language. 10, 40
- URL** Uniform Resource Locators. 35

Glosario

dashboard Panel de control. É unha ferramenta visual que proporciona unha vista resumida e gráfica das métrica. [4](#), [11](#), [14](#), [22](#), [51](#), [52](#), [54](#), [56](#), [57](#), [61](#)

Bibliografía

- [1] Spotify Ltd., “Spotify,” accessed: 9 de xuño de 2024. [En liña]. Dispoñible en: <https://open.spotify.com/intl-es>
- [2] Apple Inc., “Apple music,” accessed: 9 de xuño de 2024. [En liña]. Dispoñible en: <https://www.apple.com/es/apple-music/>
- [3] Amazon Inc., “Amazon music,” accessed: 9 de xuño de 2024. [En liña]. Dispoñible en: <https://music.amazon.es/>
- [4] MetaBrainz Foundation, “Musicbrainz,” accessed: 29 de maio de 2024. [En liña]. Dispoñible en: <https://musicbrainz.org/>
- [5] H. S. G. y Prakash C. Rao, *La integración de información para la mejor toma de decisiones*. MExico: A Simon Schuster Company, 1996.
- [6] Oracle Corporation, “¿qué es un data warehouse?” accessed: 30 de maio de 2024. [En liña]. Dispoñible en: <https://www.oracle.com/es/database/what-is-a-data-warehouse/>
- [7] Nunsys, “Data warehouse,” accessed: 30 de maio de 2024. [En liña]. Dispoñible en: <https://www.nunsys.com/data-warehouse/>
- [8] Iván de Souza, “Data warehouse: ¿qué es y cómo funciona?” accessed: 30 de maio de 2024. [En liña]. Dispoñible en: <https://rockcontent.com/es/blog/data-warehouse/#:~:text=El%20Data%20Warehouse%20permite%20el,normativas%2C%20objetivos%20y%20plantear%20riesgos.>
- [9] IBM Corporation, “Etl (extract, transform, and load),” accessed: 31 de maio de 2024. [En liña]. Dispoñible en: <https://www.ibm.com/es-es/topics/etl>
- [10] R. Kimball and J. Caserta, *The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*. Indianapolis: Wiley, 2004.

- [11] Amazon Web Services, “What is an api?” accessed: 22 de maio de 2024. [En liña].
Dispoñible en: <https://aws.amazon.com/es/what-is/api/>
- [12] Aprende Python Hoy, “Librerías en python y su utilidad para programar,” accessed: 27 de maio de 2024. [En liña]. Dispoñible en:
<https://aprendepythonhoy.com/librerias-en-python-y-su-utilidad-para-programar#:~:text=%C2%BFQu%C3%A9%20son%20las%20librer%C3%ADas%20en%20Python%3F%20Las%20librer%C3%ADas,ahorrar%20tiempo%20y%20esfuerzo%20al%20momento%20de%20programar.>
- [13] Python Software Foundation, “Python documentation,” accessed: 27 de maio de 2024.
[En liña]. Dispoñible en: <https://docs.python.org/es/3/tutorial/>
- [14] Spotify Ltd., “Spotify web api documentation,” accessed: 25 de maio de 2024. [En liña].
Dispoñible en: <https://developer.spotify.com/documentation/web-api>
- [15] MusicBrainz NGS Developers, “Python-musicbrainzngs api documentation,” accessed:
25 de maio de 2024. [En liña]. Dispoñible en: https://python-musicbrainzngs.readthedocs.io/en/v0.7.1/api/#musicbrainzngs.musicbrainz.VALID_RELEASE_TYPES
- [16] Platzi, “Qué es postgresql,” accessed: 27 de maio de 2024. [En liña]. Dispoñible en:
<https://platzi.com/blog/que-es-postgresql/>
- [17] Wikipedia, “Dbeaver,” accessed: 27 de maio de 2024. [En liña]. Dispoñible en:
<https://es.wikipedia.org/wiki/DBeaver>
- [18] Tableau, “Tableau,” accessed: 8 de xuño de 2024. [En liña]. Dispoñible en:
<https://www.tableau.com/es-es>
- [19] R. Kimball and M. Ross, *The Data Warehouse Toolkit: The Definitive Guide to dimensional Modeling*. Indianapolis: Wiley, 2013, 3° edition.
- [20] MediaWiki, “Medología de kimball,” accessed: 22 de maio de 2024. [En liña].
Dispoñible en: <https://inteligenciadenegociosval.blogspot.com/2014/01/metodologia-de-kimball.html#:~:text=La%20Metodolog%C3%ADa%20Kimball%2C%20es%20una,de%20decisiones%20en%20la%20entidad>
- [21] Spotipy Developers, “Spotipy documentation,” accessed: 25 de maio de 2024. [En liña].
Dispoñible en: <https://spotipy.readthedocs.io/en/2.24.0/>
- [22] MetaBrainz Foundation, “Diagrama completo tablas e relacións de musicbrainz,”
accessed: 13 de marzo de 2024. [En liña]. Dispoñible en: <https://wiki.musicbrainz.org/images/5/5e/soup.svg>