



UNIVERSIDAD NACIONAL DE LA MATANZA

INTELIGENCIA DE NEGOCIOS

Resolución TP

Caso SoundWave

Docentes: ING. LORENA R. MATTEO/LIC. MARCELO N. SAIED

CASO PLATAFORMA STREAMING SOUNDWAVE



PROCESO DE DISEÑO DIMENSIONAL



CASO PLATAFORMA STREAMING SOUNDWAVE

PROCESO DE DISEÑO DIMENSIONAL



Pasos

- 1 Elegir el **proceso de negocios** a modelar (ventas diarias, manejo de stock, etc.).
- 2 Elegir las **dimensiones** que van a intervenir (tener en claro los **atributos**, **jerarquías** y elementos de atributos).
- 3 Elegir la **granularidad** (nivel de detalle) del proceso de negocios con que se van a guardar los datos en el DW.
- 4 Elegir los **hechos** y **medidas** que se van a utilizar en la tabla de hechos (básicas y calculadas).

CASO PLATAFORMA STREAMING SOUNDWAVE

PROCESO DE DISEÑO DIMENSIONAL



Ejercicio Caso Práctico - Modelado Dimensional y Físico Plataforma de Streaming SoundWave

La plataforma de streaming de música SoundWave ha experimentado un crecimiento exponencial desde su lanzamiento hace tres años. Con sede en Buenos Aires, **ofrece a sus usuarios acceso a más de 50 millones de canciones de diversos géneros musicales bajo diferentes modelos de suscripción:** Free (con publicidad), Basic (sin publicidad), Premium (sin publicidad y con descargas ilimitadas), y Family (hasta 6 usuarios). La plataforma está disponible en Argentina, Chile, Uruguay, Paraguay, Brasil, Colombia, México y España.

En SoundWave, los usuarios pueden crear playlists personalizadas, seguir a artistas, compartir música con amigos, y descubrir nuevas canciones a través del algoritmo de recomendación. Los artistas y sellos discográficos reciben regalías basadas en la cantidad de reproducciones que obtienen sus canciones.

El sistema de operaciones actual contiene información sobre cada reproducción de **canción**, con los siguientes datos: **ID de reproducción, timestamp de inicio, duración de la escucha, ID de usuario, ID de canción, dispositivo utilizado (móvil Android, móvil iOS, web desktop, smart TV, smartwatch), si la reproducción fue completa o parcial, si se realizó online o como archivo descargado, y si formaba parte de una playlist.**

También cuenta con información sobre las canciones: **ID de canción, título, artista principal, artistas colaboradores (si los hay), álbum, género principal, subgénero, duración en segundos, fecha de lanzamiento, sello discográfico, y país de origen del artista principal.**

Sobre los **usuarios**, el sistema almacena: **ID de usuario, nombre, edad, género, país, ciudad, tipo de suscripción, fecha de registro en la plataforma, y dispositivos registrados.**

Mariana Suárez, recién nombrada como Directora de Análisis de Datos, identificó que a pesar de contar con abundante información, la empresa tiene dificultades para obtener insights accionables que mejoren la experiencia de usuario y las decisiones de negocio. Actualmente, el equipo de análisis depende del departamento de TI para ejecutar consultas SQL complejas, lo que genera demoras y frustraciones.

CASO PLATAFORMA STREAMING SOUNDWAVE

PROCESO DE DISEÑO DIMENSIONAL



Tras varias reuniones donde Mariana presentó las limitaciones del enfoque actual y las posibilidades que ofrecería un sistema de BI, Santiago finalmente accedió a realizar una prueba. Mencionó varios análisis que le gustaría poder realizar fácilmente: patrones de escucha por **hora del día**, **días de la semana** y **temporadas**; tendencias de **géneros musicales** por **región**; comportamiento de usuarios según **tipo de suscripción**; **artistas** emergentes con crecimiento acelerado; **canciones** con mayor **retención de escucha**; impacto de las **playlists** en el descubrimiento de nuevos **artistas**; y proyección de **ingresos por tipo de suscripción**.

Santiago hizo énfasis en poder analizar tendencias por periodos de **tiempo** específicos (**años, trimestres, meses, semanas, días**), por **ubicación geográfica** a diferentes niveles (**país, región/provincia, ciudad**), por características del **usuario** (**tipo de suscripción, rango etario, género**), por características de la **música** (**género, subgénero, artista, sello discográfico**) y por **dispositivo de acceso**.

Mariana contactó a la consultora especializada en BI donde usted trabaja para llevar adelante este proyecto. Como consultor principal, usted debe diseñar e implementar la solución que transformará la manera en que SoundWave utiliza sus datos para tomar decisiones estratégicas.

Retencion de escucha = Porcentaje de Reproduccion
Temporadas = Agrupacion de Semanas, meses , Estaciones

CASO PLATAFORMA STREAMING SOUNDWAVE



ANALISIS DESEABLES

Analisis deseables

- Patrones de escucha por hora del día, días de la semana y temporadas;
- Tendencias de géneros musicales por región
- Comportamiento de usuarios según tipo de suscripción;
- Artistas emergentes con crecimiento acelerado
- Canciones con mayor retención de escucha
- Impacto de las playlists en el descubrimiento de nuevos artistas
- Patrones de escucha por tipo de suscripción

- Por Periodos de tiempo específicos (años, temporadas ,trimestres, meses, semanas, días)
- Por ubicación geográfica a diferentes niveles (país, región/provincia, ciudad),
- Por características del usuario (tipo de suscripción, rango etario, género),
- Por características de la música (género, subgénero, artista, sello discográfico)
- Por dispositivo de acceso
- Por Pais del Artista
- Por tipo de suscripción

CASO PLATAFORMA STREAMING SOUNDWAVE

PROCESO DE DISEÑO DIMENSIONAL



A partir del enunciado anterior, se pide:

- a) ¿Cuál es el objetivo de negocios a cumplir?
- b) Liste las posibles dimensiones.
- c) Liste las medidas o hechos, indicando cuáles corresponden a medidas básicas y calculadas. ¿En qué proceso de negocio englobaría los hechos detectados?
- d) Construya el modelo dimensional conceptual indicando la granularidad del modelo.
- e) Identifique atributos para cada dimensión con sus jerarquías.
- f) Construya el modelo dimensional lógico.
- g) Arme el modelo físico del Data Warehouse en Esquema Estrella (Star)
- h) Arme el modelo físico del Data Warehouse en Esquema Copo de Nieve (Snowflake)

CASO PLATAFORMA STREAMING SOUNDWAVE



A) ¿CUÁL ES EL OBJETIVO DE NEGOCIOS A CUMPLIR?

El objetivo de negocios principal es transformar la toma de decisiones en SoundWave de un enfoque basado en reportes estáticos y dependientes de TI a un modelo ágil, autónomo y basado en datos en tiempo real, permitiendo a los diferentes departamentos obtener insights accionables para:

Mejorar la experiencia del usuario: Entender patrones de escucha, preferencias musicales, impacto de las recomendaciones y playlists para personalizar la oferta y aumentar la retención.

Optimizar las decisiones de negocio: Identificar tendencias de mercado, evaluar el rendimiento de artistas y sellos, optimizar estrategias de marketing y contenido, y proyectar ingresos.

Democratizar el acceso a la información: Empoderar a los diferentes departamentos para realizar sus propios análisis sin depender del equipo de TI, agilizando la identificación de oportunidades y la resolución de problemas.

En esencia, el objetivo es convertir los datos brutos de SoundWave en inteligencia de negocio para impulsar el crecimiento y la competitividad de la plataforma.

CASO PLATAFORMA STREAMING SOUNDWAVE



B) LISTE LAS POSIBLES DIMENSIONES.

- **Tiempo:** Para analizar tendencias a lo largo del tiempo.
- **Usuario:** Para segmentar y analizar el comportamiento de los usuarios.
- **Canción:** Para analizar las características de la música (genero, Tipo de reproduccion, Estado) reproducida.
- **Artista:** Para analizar el rendimiento de los artistas.
- **Sello Discográfico:** Para analizar el rendimiento de los sellos.
- **Ubicación:** Para analizar el comportamiento por región geográfica.
- **Dispositivo:** Para analizar el uso de la plataforma según el dispositivo.
- **Suscripción:** Para analizar el comportamiento según el tipo de suscripción.
- **Playlist:** Para analizar si la reproducción provino de una playlist.
- **Estado de Reproduccion:** Para analizar si el esatado de la reproduccion es Completa / parcial
- **Tipó de Reproduccion:** Para analizar si el Tipo de la reproduccion es Online / Downloaded

CASO PLATAFORMA STREAMING SOUNDWAVE



C) LISTE LAS MEDIDAS O HECHOS, INDICANDO CUÁLES CORRESPONDEN A MEDIDAS BÁSICAS Y CALCULADAS. ¿EN QUÉ PROCESO DE NEGOCIO ENGLOBARÍA LOS HECHOS DETECTADOS?

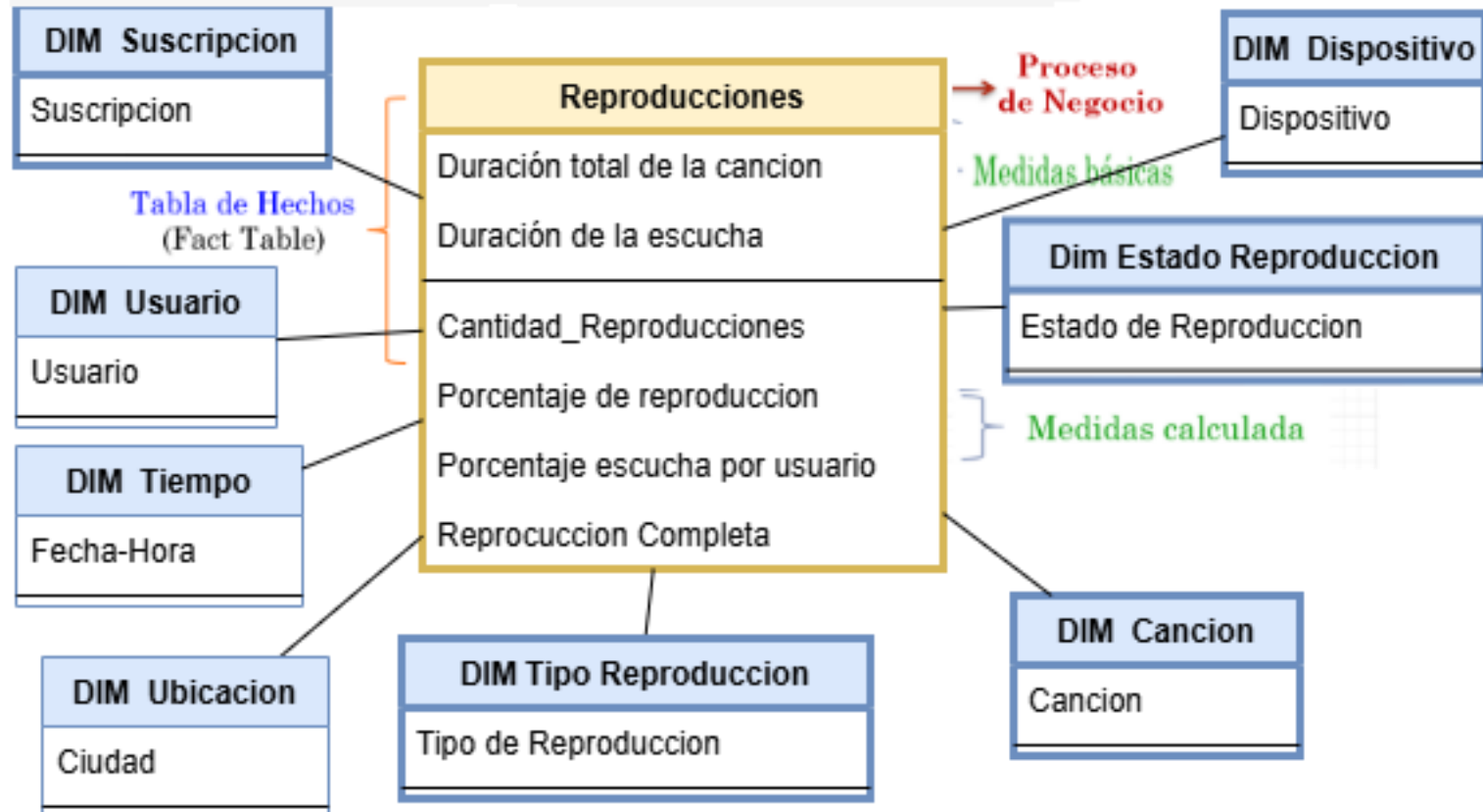
| Medida | Descripción | Tipo | Fórmula | Forma de consolidación |
|-------------------------------------|--|----------|--|------------------------|
| Duración Total de la Canción | Tiempo Total de una canción | B | | SUM |
| Duración total de escucha | Tiempo total de escucha en segundos | B | Suma de duraciones | SUM |
| Reproducciones completas | Número de reproducciones que se completaron | C | Contar reproducciones completas | SUM |
| Porcentaje de reproducción | Porcentaje de la canción que fue escuchada | C | $(\text{Duración de escucha} / \text{Duración total de la canción}) * 100$ | AVG |
| Promedio escucha por usuario | Tiempo promedio de escucha por usuario | C | Suma de duraciones / Cantidad de usuarios | AVG |
| Cantidad de reproducciones | Número total de veces que se reprodujo una canción | C | Contar | SUM |

CASO PLATAFORMA STREAMING SOUNDWAVE



D) CONSTRUYA EL MODELO DIMENSIONAL CONCEPTUAL INDICANDO LA GRANULARIDAD DEL MODELO.

STREAMING SOUNDWAVE (Conceptual)



Granularidad de la tabla de Hechos:

**Usuario x Ciudad x Suscripción x Dispositivo x Canción x Playlist x
Estado_Reproducción x Tipo Reproducción x Fecha-hora**

CASO PLATAFORMA STREAMING SOUNDWAVE



D)CONSTRUYA EL MODELO DIMENSIONAL CONCEPTUAL INDICANDO LA GRANULARIDAD DEL MODELO.

Granularidad de la tabla de Hechos:

**Usuario x Ciudad x Suscripcion x Dispositivo x Cancion x Playlist
x Estado_Reproduccion X Tipo_Reproduccion X Fecha-hora**

Granularidad del Modelo:

La granularidad de la tabla de hechos de "Reproducciones" será a nivel de cada evento individual de reproducción de una canción por un usuario en un momento específico. Esto significa que cada fila en la tabla de hechos representará una única reproducción. Esta granularidad permitirá realizar análisis detallados a nivel de reproducción y luego agregarlos a niveles superiores utilizando las dimensiones.

CASO PLATAFORMA STREAMING SOUNDWAVE



E) IDENTIFIQUE ATRIBUTOS PARA CADA DIMENSIÓN CON SUS JERARQUÍAS.

Dimensión Usuario:

Atributos:

- Rango etario
- (ej: 13-17,18-24,25-34,...)
- Género
- País
- Region/Provincia
- Ciudad
- Tipo de suscripción

Jerarquías:

- Rango etario -> Usuario
- Genero -> Usuario

Dimensión Estado

Reproduccion:

Atributos:

- Estado de Reproduccion (Completa/Parcial)

Jerarquías:

- Estado de Reproduccion

Dimensión Tiempo:

Atributos:

- Año
- Temporada (Verano, Otoño, Invierno, etc..)
- Trimestre
- Mes
- Semana
- Día semana
- Fecha – Hora

Jerarquías:

- Año-> Trimestre -> Mes -> Fecha-Hora
- Año-> Temporada ->Semana-> Día samana-> Fecha-Hora

Dimensión Dispositivo:

Atributos:

- Tipo de dispositivo

Jerarquías:

- Tipo de dispositivo

Dimensión Tipo Reproduccion:

Atributos:

- Tipo de Reproduccion (Online/Descarga)

Jerarquías:

- Tipo de Reproduccion

CASO PLATAFORMA STREAMING SOUNDWAVE



E) IDENTIFIQUE ATRIBUTOS PARA CADA DIMENSIÓN CON SUS JERARQUÍAS.

Dimensión Canción:

Atributos:

- Título de la canción
- Nombre del artista
- Pais Artista
- Género Principal
- Subgénero
- Sello Discografico

Jerarquías:

- Género Principal -> Subgénero -> Canción
- Sello Discográfico -> Artista -> Canción

Dimensión Ubicación:

Atributos:

- País
- Región/Provincia
- Ciudad

Jerarquías:

- País -> Región/Provincia -> Ciudad

Dimensión Suscripción:

Atributos:

- Tipo de suscripción
(Free/Premium/Family)

Jerarquías:

- Tipo de suscripción

Dimensión Playlist:

Atributos:

- Playlist

Jerarquías:

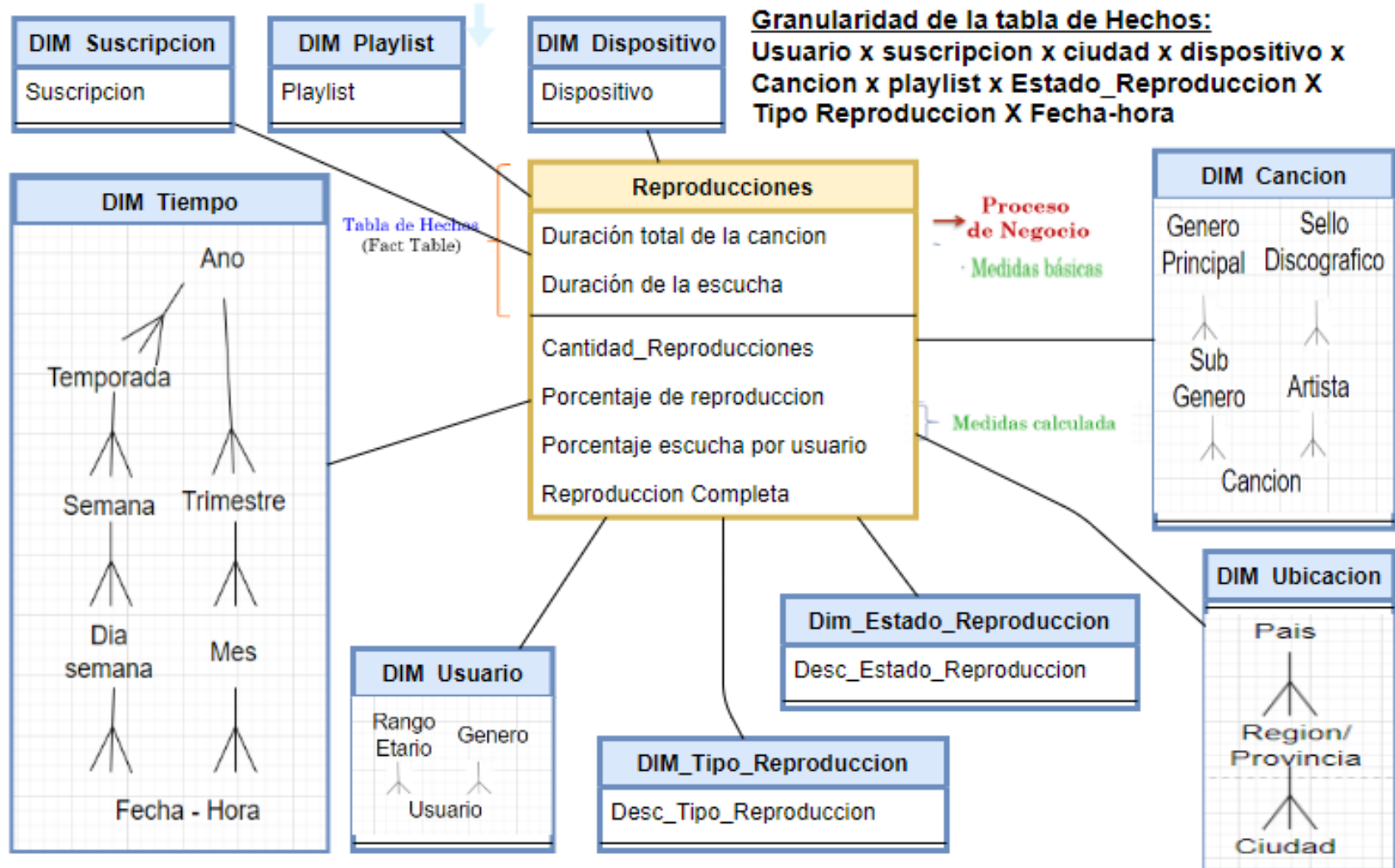
- Playlist

CASO PLATAFORMA STREAMING SOUNDWAVE



F) CONSTRUYA EL MODELO DIMENSIONAL LÓGICO.

STREAMING SOUNDWAVE (Logico)



CASO PLATAFORMA STREAMING SOUNDWAVE



G) ARME EL MODELO FÍSICO DEL DATA WAREHOUSE EN ESQUEMA ESTRELLA (STAR)

STREAMING SOUNDWAVE (Físico - Estrella)

Granularidad de la tabla de Hechos:

**Usuario x suscripcion x ciudad x dispositivo x
Cancion x playlist x
Estado_Reproduccion X Tipo Reproduccion X
Artista x Fecha-hora**

Alias:

Ciudad Artista = Ciudad Usuario

| LK_Cancion | |
|------------|-----------------------|
| PK/SK | ID_Cancion |
| | Titulo_cancion |
| | ID_Artista |
| | Nombre_Artista |
| FK | ID_Pais_Artista |
| FK | Pais_Artista |
| | ID_Subgenero |
| | Subgenero |
| | ID_Genero_principal |
| | Genero_principal |
| | ID_Sello_Discografico |
| | Sello_discografico |
| | Duracion |

| BT_Reproducciones | |
|-------------------|------------------------|
| PK/SK | ID_Reproduccion |
| FK | ID_Fecha |
| FK | ID_Usuario |
| FK | ID_Cancion |
| FK | ID_Dispositivo |
| FK | ID_Playlist |
| FK | ID_Suscripcion |
| FK | ID_Estado_Reproduccion |
| FK | ID_Tipo_Reproduccion |
| | Duracion_Escucha |
| | Diracion_Cancion |

| LK_Estado_Reproduccion | |
|------------------------|--------------------------|
| PK/SK | ID_Estado_Reproduccion |
| | Desc_Estado_Reproduccion |

| LK_Tipo_Reproduccion | |
|----------------------|------------------------|
| PK/SK | ID_Tipo_Reproduccion |
| | Desc_Tipo_Reproduccion |

| LK_Usuario | |
|------------|----------------|
| PK/SK | ID_Usuario |
| | Usuario_Nombre |
| | Edad |
| | Rango_Etario |
| FK | ID_Ciudad |
| | ID_Genero |
| | Genero |

| LK_Tiempo | |
|-----------|---------------|
| PK/SK | Fecha_ID |
| | Fecha-Hora |
| | ID_Mes |
| | Mes |
| | ID_Dia_Semana |
| | Dia_Semana |
| | ID_Semana |
| | Semana |
| | ID_Temporada |
| | Temporada |
| | ID_Trimestre |
| | Trimestre |
| | ID_Ano |
| | Año |

| LK_Ubicacion | |
|--------------|---------------------|
| PK | ID_Ciudad |
| | Ciudad |
| | ID_Region_provincia |
| | Region_Provincia |
| | ID_Pais |
| | Pais |

| LK_Playlist | |
|-------------|-----------------|
| PK/SK | ID_DPlaylist |
| | Playlist_Nombre |

| LK_Dispositivo | |
|----------------|--------------------|
| PK/SK | ID_Dispositivo |
| | Dispositivo_Nombre |

| LK_Suscripcion | |
|----------------|------------------|
| PK/SK | ID_Suscripcion |
| | tipo_suscripcion |

CASO PLATAFORMA STREAMING SOUNDWAVE



G) ARME EL MODELO FÍSICO DEL DATA WAREHOUSE EN ESQUEMA ESTRELLA (STAR)

Descripción del Modelo Físico en Esquema Copo de Nieve:

- **Tabla de Hechos (fact_reproducciones):** Permanece como la tabla central, conteniendo las medidas y las claves foráneas a las dimensiones de nivel superior.
- **Dimensiones Normalizadas:** Las dimensiones que en el esquema estrella eran tablas únicas y potencialmente grandes (como dim_usuario, dim_cancion, dim_sello_discografico, dim_ubicacion) ahora se han normalizado en múltiples tablas relacionadas.
- **dim_usuario y dim_ubicacion:** La información de ubicación del usuario (país, región, ciudad) se ha movido a una sub-dimensión dim_ubicacion, que a su vez se relaciona con tablas dim_pais, dim_region y dim_ciudad. Esto reduce la redundancia si muchos usuarios comparten la misma ubicación.
- **dim_artista y dim_ubicacion:** Similar a los usuarios, la información de ubicación del artista se normaliza a través de dim_ubicacion, dim_pais, dim_region y dim_ciudad.

CASO PLATAFORMA STREAMING SOUNDWAVE



G) ARME EL MODELO FÍSICO DEL DATA WAREHOUSE EN ESQUEMA ESTRELLA (STAR)

Descripción del Modelo Físico en Esquema Estrella:

- **Tabla de Hechos (fact_reproducciones):** Esta es la tabla central del esquema estrella. Contiene las medidas o hechos (como duracion_escucha) y las claves foráneas que referencian las claves primarias de cada tabla de dimensión. La clave primaria de esta tabla suele ser una clave artificial o una combinación de claves foráneas (aunque aquí usamos un id_reproduccion explícito).

Puntos Clave del Diseño Físico:

- **Claves Primarias (PK):** Cada tabla de dimensión tiene una clave primaria única para identificar cada registro. La tabla de hechos también tiene una clave primaria.
- **Claves Foráneas (FK):** La tabla de hechos contiene claves foráneas que enlazan cada registro de hecho con los registros correspondientes en las tablas de dimensiones. Esto asegura la integridad referencial y permite realizar joins para el análisis.
- **Tipos de Datos:** Se han utilizado tipos de datos comunes en bases de datos relacionales para Data Warehousing (INT, VARCHAR, BOOLEAN, TIMESTAMP, DATE, SMALLINT, TINYINT). La elección final del tipo de dato puede depender del sistema de base de datos específico.
- **Desnormalización:** Las tablas de dimensiones están típicamente desnormalizadas para optimizar las consultas de análisis. Esto significa que pueden contener atributos que podrían estar en tablas separadas en un modelo transaccional (como el nombre del artista dentro de la tabla dim_cancion).
- **Tabla de Ubicación:** Se ha creado una tabla dim_ubicacion separada y se ha referenciado desde dim_usuario, dim_artista y dim_sello_discografico. Esto permite una gestión más centralizada de la información geográfica y facilita el análisis por ubicación.
- **Tabla de Playlist:** Se ha creado una dimensión simple para indicar si la reproducción provino de una playlist. Este modelo físico en esquema estrella simplifica la estructura del Data Warehouse y optimiza las consultas de análisis, ya que las joins se realizan principalmente entre la tabla de hechos central y las tablas de dimensiones directamente conectadas. Esto facilita la comprensión del modelo y mejora el rendimiento de las consultas de Business Intelligence.

CASO PLATAFORMA STREAMING SOUNDWAVE



G) ARME EL MODELO FÍSICO DEL DATA WAREHOUSE EN ESQUEMA COPO DE NIEVE (SNOWFLAKE)

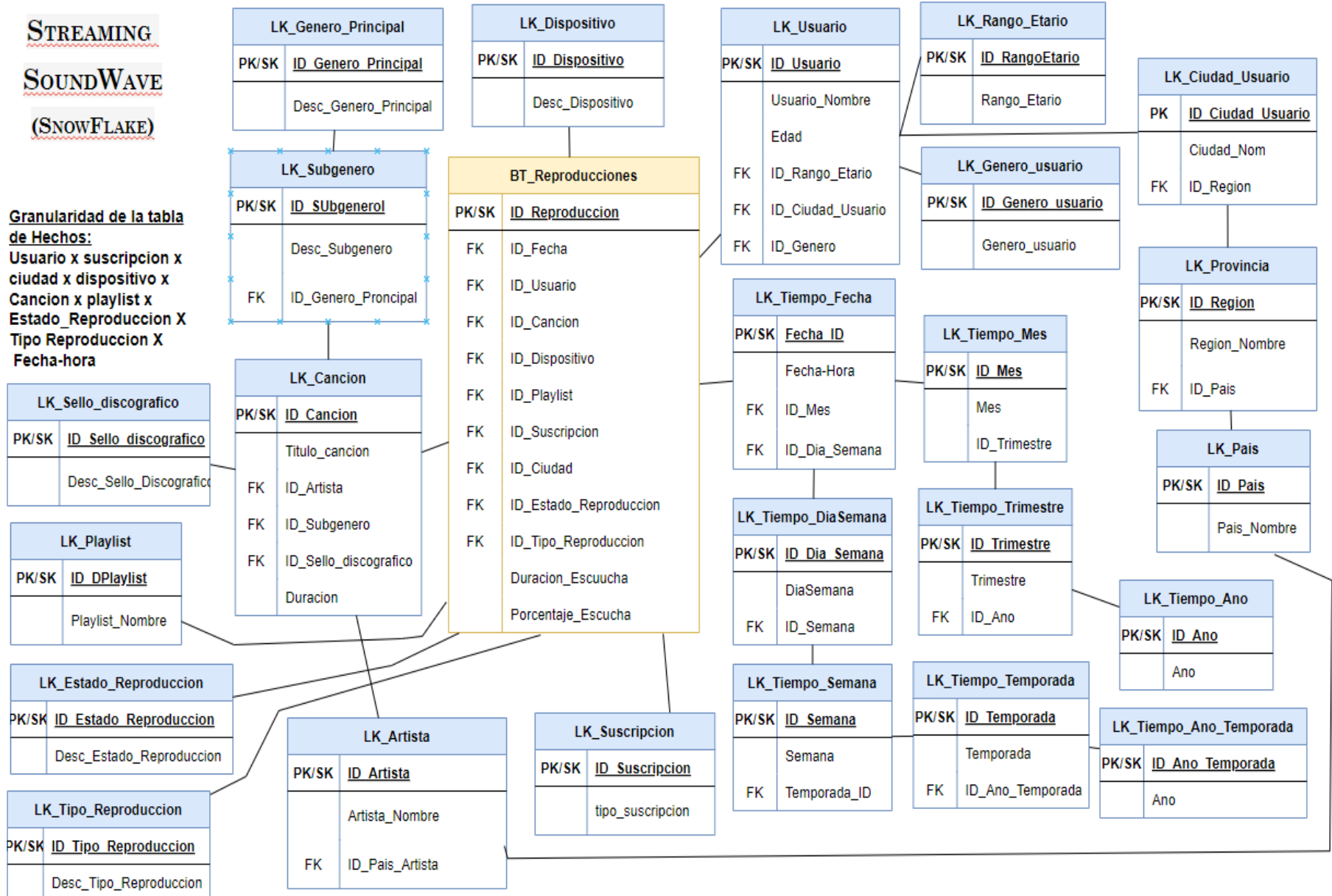
STREAMING

SOUNDWAVE

(SNOWFLAKE)

Granularidad de la tabla de Hechos:

Usuario x suscripcion x ciudad x dispositivo x Cancion x playlist x Estado_Reproduccion X Tipo Reproduccion X Fecha-hora



CASO PLATAFORMA STREAMING SOUNDWAVE



G) ARME EL MODELO FÍSICO DEL DATA WAREHOUSE EN ESQUEMA COPO DE NIEVE (SNOWFLAKE)

Ventajas del Esquema Copo de Nieve:

- **Reducción de la redundancia de datos:** Al normalizar las dimensiones, se evita la repetición de atributos, lo que puede ahorrar espacio de almacenamiento.
- **Mejora de la integridad de los datos:** Los cambios en los atributos de las dimensiones solo necesitan realizarse en una única tabla.
- **Mayor flexibilidad para cambios en las dimensiones:** Modificar la estructura de una dimensión normalizada puede ser más sencillo sin afectar directamente a la tabla de hechos.

Desventajas del Esquema Copo de Nieve:

- **Consultas más complejas:** Para obtener la información completa de una dimensión, a menudo se requieren más joins entre las tablas normalizadas, lo que puede aumentar la complejidad de las consultas y potencialmente afectar el rendimiento (aunque las bases de datos modernas suelen optimizar esto).
- **Mayor complejidad del modelo:** El esquema con más tablas puede ser más difícil de entender y mantener inicialmente.

CASO PLATAFORMA STREAMING SOUNDWAVE

G) ARME EL MODELO FÍSICO DEL DATA WAREHOUSE EN ESQUEMA COPO DE NIEVE (SNOWFLAKE)



Descripción del Modelo Físico en Esquema Copo de Nieve:

- **Tabla de Hechos (fact_reproducciones):** Permanece como la tabla central, conteniendo las medidas y las claves foráneas a las dimensiones de nivel superior.
- **Dimensiones Normalizadas:** Las dimensiones que en el esquema estrella eran tablas únicas y potencialmente grandes (como dim_usuario, dim_cancion, dim_artista, dim_sello_discografico, dim_ubicacion) ahora se han normalizado en múltiples tablas relacionadas.
- **dim_usuario y dim_ubicacion:** La información de ubicación del usuario (país, región, ciudad) se ha movido a una sub-dimensión dim_ubicacion, que a su vez se relaciona con tablas dim_pais, dim_region y dim_ciudad. Esto reduce la redundancia si muchos usuarios comparten la misma ubicación.
- **dim_artista y dim_ubicacion:** Similar a los usuarios, la información de ubicación del artista se normaliza a través de dim_ubicacion, dim_pais, dim_region y dim_ciudad.
- **dim_sello_discografico y dim_ubicacion:** La ubicación del sello discográfico también se normaliza de la misma manera.

· **Otras Dimensiones:** Dimensiones como dim_tiempo, dim_suscripcion, dim_dispositivo y dim_playlist pueden permanecer más desnormalizadas si la redundancia en sus atributos no es significativa o si la complejidad adicional de la normalización no se justifica por los beneficios.

Ventajas del Esquema Copo de Nieve:

- **Reducción de la redundancia de datos:** Al normalizar las dimensiones, se evita la repetición de atributos, lo que puede ahorrar espacio de almacenamiento.
- **Mejora de la integridad de los datos:** Los cambios en los atributos de las dimensiones solo necesitan realizarse en una única tabla.
- **Mayor flexibilidad para cambios en las dimensiones:** Modificar la estructura de una dimensión normalizada puede ser más sencillo sin afectar directamente a la tabla de hechos.

Desventajas del Esquema Copo de Nieve:

- **Consultas más complejas:** Para obtener la información completa de una dimensión, a menudo se requieren más joins entre las tablas normalizadas, lo que puede aumentar la complejidad de las consultas y potencialmente afectar el rendimiento (aunque las bases de datos modernas suelen optimizar esto).
- **Mayor complejidad del modelo:** El esquema con más tablas puede ser más difícil de entender y mantener inicialmente.

