

---

## EAD Q2

---

# PRÁCTICA DATA WAREHOUSE

### Grupo 3

---

Pérez Paredes, Alexandre - alexandre.perez1@udc.es

Armenteros López, Ana - ana.armenteros@udc.es

## Contenido

---

Descripción del proceso de negocio.....	1
Nivel de detalle de la tabla de hechos.....	1
Dimensiones seleccionadas y sus atributos.....	1
Festival.....	1
Artista.....	2
Recinto.....	2
Hora de inicio y Hora de fin.....	3
Fecha.....	3
Clima.....	3
Ejemplos de consultas analíticas.....	3
Diagrama DFM.....	5
Descartes de atributos y métricas.....	6
Dimensiones que cambian lentamente en el tiempo.....	7
Dimensión Recinto.....	7
Estrategia por seguir.....	7
Métricas Aditivas.....	7
Métricas no aditivas.....	7
Diseño lógico.....	8
Consultas analíticas.....	10

## Figuras

---

Figura 1. Diagrama DFM.....	5
Figura 2. Dimensión fecha.....	6
Figura 3 . Modelo lógico.....	9

## Descripción del proceso de negocio.

---

La empresa *Festival Company* se dedica a la organización de **festivales de música**. Para cada festival, se decide la fecha, el lugar y el artista para cada una de las distintas **actuaciones** que se celebrarán en él. Además, se encarga de proporcionar un servicio de barra que se hará cargo de la venta de consumiciones a lo largo de todo el festival.

Se desea contar con la información necesaria para decidir quién va a realizar una actuación, cuándo y dónde se va a celebrar, tratando de conseguir el máximo beneficio posible (vender más entradas y más consumiciones durante las actuaciones).

## Nivel de detalle de la tabla de hechos.

---

La tabla de hechos corresponderá a cada **actuación** que se celebra en un festival, en un día y hora concreta. Cada actuación será realizada por un artista (en solitario o como parte de una banda), en un recinto concreto y tendrá un número determinado de asistentes. Durante una actuación, cada asistente podrá salir del recinto en cualquier momento que desee. Además, se controlarán las consumiciones totales y los ingresos que generen; los posibles altercados que puedan ocurrir; la ocupación del recinto respecto a su capacidad y el clima durante la actuación.

## Dimensiones seleccionadas y sus atributos.

---

A continuación, se enumeran las dimensiones y los atributos pertenecientes a cada una de ellas, junto a una descripción y algún ejemplo como justificación de estas.

### Festival.

Nos permite conocer en qué festival tuvo lugar una actuación en concreto y, por tanto, en qué país tuvo lugar y cuántas personas asistieron en total al festival. Sus atributos son:

- **Nombre.** Nos sirve para identificar fácilmente un festival.
- **Ciudad, país, provincia, región.** Indica en qué espacio geográfico se celebró el festival. Comprobaremos si la asistencia a una actuación de un artista depende de su lugar de origen. (p.ej. puede que para un artista alemán acuda más gente si el festival se celebra en Alemania o en otro país donde sea especialmente conocido).
- **Número de asistentes.** Indica cuántas personas asistieron al festival, independientemente de las actuaciones a las que hayan asistido (si solamente ha asistido a una actuación en todo el festival, se cuenta como asistente). Nos servirá para calcular el % de asistencia para cada actuación. Se tendrán diferentes **rangos** de valores para hacer una aproximación de cuantas personas asistieron en total al festival y serán los siguientes:
  - [0, 5000]
  - (5000, 10000]
  - (10000, 20000]

- (20000, 30000]
- (30000, 40000]
- (40000, 50000]
- (50000, 60000]
- (70000, 80000]
- (80000, 90000]
- (90000, 100000]
- (100000, 250000]
- (250000, 500000]
- (500000, Inf]
- **Año.** Indica el año o edición del festival. Nos sirve para comparar los datos de un mismo festival para distintos años.

### Artista.

Corresponde al artista o banda que va a realizar dicha actuación. Sus atributos son:

- **Nombre.** Corresponde al nombre del artista o de la banda.
- **Ciudad, país, provincia, región y país.** Indica cuál es el espacio geográfico de origen del artista. Lo comparemos respecto al lugar de celebración del festival para comprobar cuánto puede influir en la asistencia.
- **Género musical.** Es un atributo multivaluado, puesto que un artista o banda puede tener varios géneros musicales. Resulta útil para conocer si el número de asistencia, el número de altercados, el ingreso total de consumiciones y el país depende de él. (p.ej. en Finlandia es probable que tengan más éxito los artistas cuyo género sea el metal)

### Recinto.

Corresponde al lugar específico donde se va a realizar una actuación. Sus atributos son:

- **Nombre.** Nos sirve para identificar fácilmente un recinto.
- **Tipo.** Podrá ser un recinto cerrado o al aire libre. Puede tener relación con el tiempo de permanencia de los asistentes, según el clima que haya, etc.
- **Capacidad.** Se refiere al número de asistentes que puede albergar. Se tendrán diferentes **rangos** de valores para almacenar la capacidad de cada recinto, que serán los siguientes:
  - [0, 5000]
  - (5000, 10000]
  - (10000, 20000]
  - (20000, 30000]
  - (30000, 40000]
  - (40000, 50000]
  - (50000, 60000]
  - (70000, 80000]
  - (80000, 90000]
  - (90000, 100000]
  - (100000, Inf]

## Hora de inicio y Hora de fin.

Corresponden a la hora en la que empieza la actuación y a la que termina, respectivamente. Pertenecen a una jerarquía compartida **Hora**. Son útiles como referencia temporal.

## Fecha.

Corresponde al día del año en el que se celebra una actuación, siendo útil como referencia temporal. Sus atributos son:

- **Día, mes y año.** Nos sirve para aplicar las métricas en días concretos.
- **Día de la semana.** La asistencia puede variar según el día de la semana (p.ej. es probable que la asistencia aumente si es fin de semana).
- **Estación del año.** La asistencia puede variar según la estación del año (p.ej. es probable que la asistencia aumente si es verano).
- **Festivo.** Especifica si la fecha correspondía a un festivo o no. La asistencia puede variar según sea festivo o no (p.ej. es probable que la asistencia aumente si es festivo).

## Clima.

- **Temperatura.** Se tendrán diferentes rangos de valores para almacenar la temperatura que hizo durante una actuación. El tiempo de permanencia puede variar según los grados que hubo durante la actuación (p.ej. es probable que la permanencia disminuya en un recinto cerrado si hace mucho calor). Se tendrán los siguientes **rangos**:
  - $(-\text{Inf}, -5^{\circ}]$
  - $(-5^{\circ}, 0^{\circ}]$
  - $(0^{\circ}, 5^{\circ}]$
  - $(5^{\circ}, 10^{\circ}]$
  - $(10^{\circ}, 15^{\circ}]$
  - $(15^{\circ}, 20^{\circ}]$
  - $(20^{\circ}, 25^{\circ}]$
  - $(25^{\circ}, 30^{\circ}]$
  - $(30^{\circ}, \text{Inf}]$
- **Tipo.** Puede ser soleado, nuboso, lluvioso, tormenta, nevoso o ventoso. El tiempo de permanencia puede variar según el tipo de día que hace (p.ej. es probable que la permanencia aumente en un recinto cerrado si está lluvioso).

## Ejemplos de consultas analíticas.

- 1) Queremos realizar una consulta para obtener la media de asistencia:
  - Por artista/grupo, en cada actuación de cada año en cada país.
  - Por artista/grupo en cada país.
  - Por artista/grupo en cada provincia de cada país.
  - Por artista/grupo en cada región de cada país.
  - Por artista/grupo en cada ciudad de cada provincia de cada país.

- Por artista/grupo en cada ciudad de cada región de cada país.
- Por artista/grupo en cada recinto.

De esta forma, la consulta nos permite conocer el “tirón” que tiene cada artista/grupo según el espacio geográfico en el que realicen la actuación.

2) Queremos realizar una consulta para obtener el número de altercados:

- Por artista/grupo.
- Por género musical del artista/grupo.
- Por la temperatura y tipo de clima que hubo durante la actuación.
- Por el tipo de recinto y su capacidad.
- Por el día de la semana y según si es festivo o no.
- Por el porcentaje de ocupación del recinto.

De esta forma, la consulta nos permite conocer si algún artista/grupo “promueve” de cierta manera que ocurran más altercados, así como el género musical, y también si influye la hora, la temperatura y el tipo de recinto y su capacidad (p. ej. una situación “sofocante” podría aumentar el número de altercados). También se tiene en cuenta si el fin de semana o un festivo influye.

3) Queremos realizar una consulta para obtener el número de consumiciones y los ingresos por consumiciones:

- Por género musical del artista/grupo en cada actuación.
- Por la hora de fin de la actuación.
- Según si el día de la actuación es festivo o no.
- Según el día de la semana.
- Por el ciudad, provincia, región y país donde se celebra el festival.
- Por temperatura media y tipo de recinto.

De esta forma, la consulta nos permite conocer si las consumiciones aumentan por un género musical en concreto, por si es festivo o fin de semana, por la hora en la que termina dicha actuación, por el espacio geográfico donde se celebra y por la temperatura media en un recinto determinado (p. ej. si es una situación sofocante puede que aumente el número de consumiciones; si el género musical es electrónica y ocurre a la noche, puede aumentar los ingresos debido a una mayor ingesta de alcohol).

4) Queremos realizar una consulta para obtener la valoración media:

- Por artista/grupo.
- Por tipo de clima y tipo de recinto.
- Por estación del año.
- Por el recinto.
- Por el porcentaje de ocupación del recinto.

De esta forma, podemos conocer qué artistas son mejor valorados y tratar de organizar sus actuaciones de manera que los más valorados actúen cuando el clima, la estación del año y el recinto no sean favorables.

5) Queremos realizar una consulta para obtener la media de minutos de permanencia:

- Por artista/grupo.

- Por género del artista y lugar de celebración.
- Por temperatura media, tipo de clima y tipo de recinto.

De esta forma, podremos conocer qué factores influyen en que un asistente permanezca más en una actuación determinada.

## Diagrama DFM.

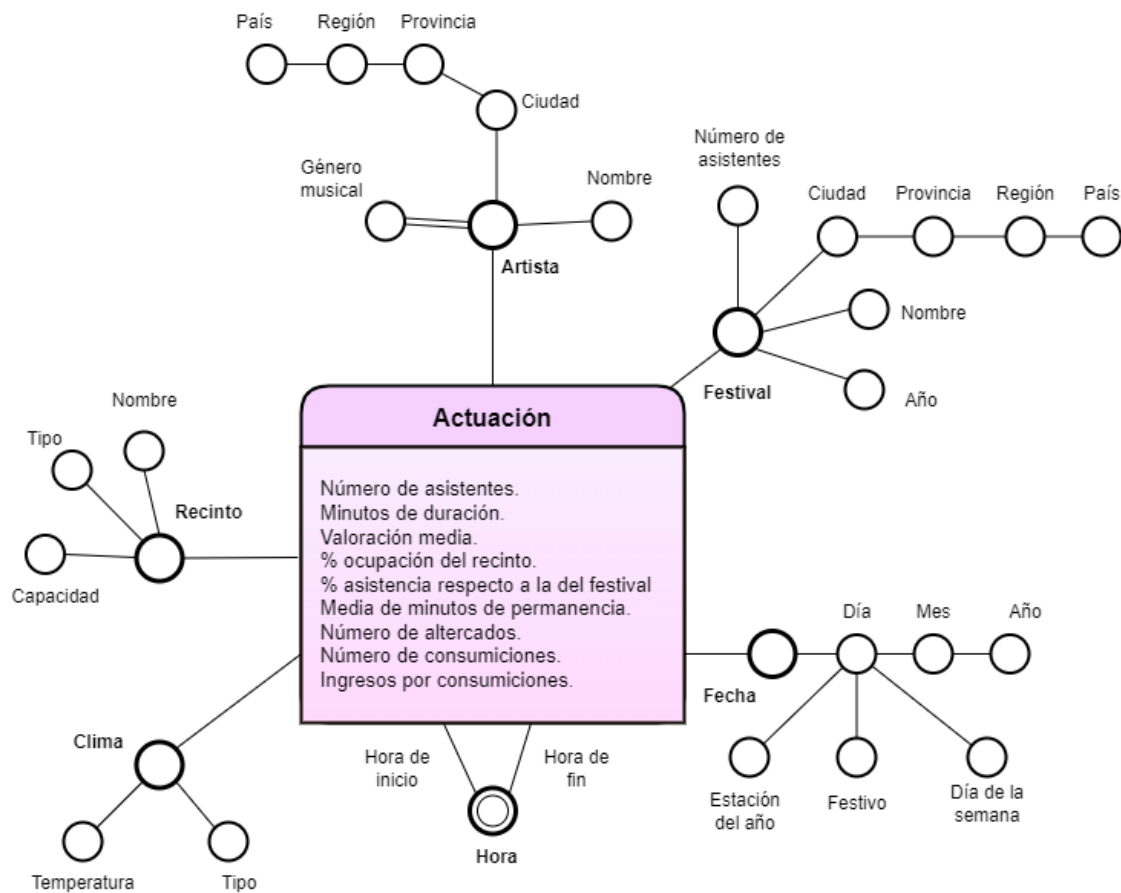


Figura 1. Diagrama DFM.

### Especificación utilizada finalmente en la dimensión fecha:

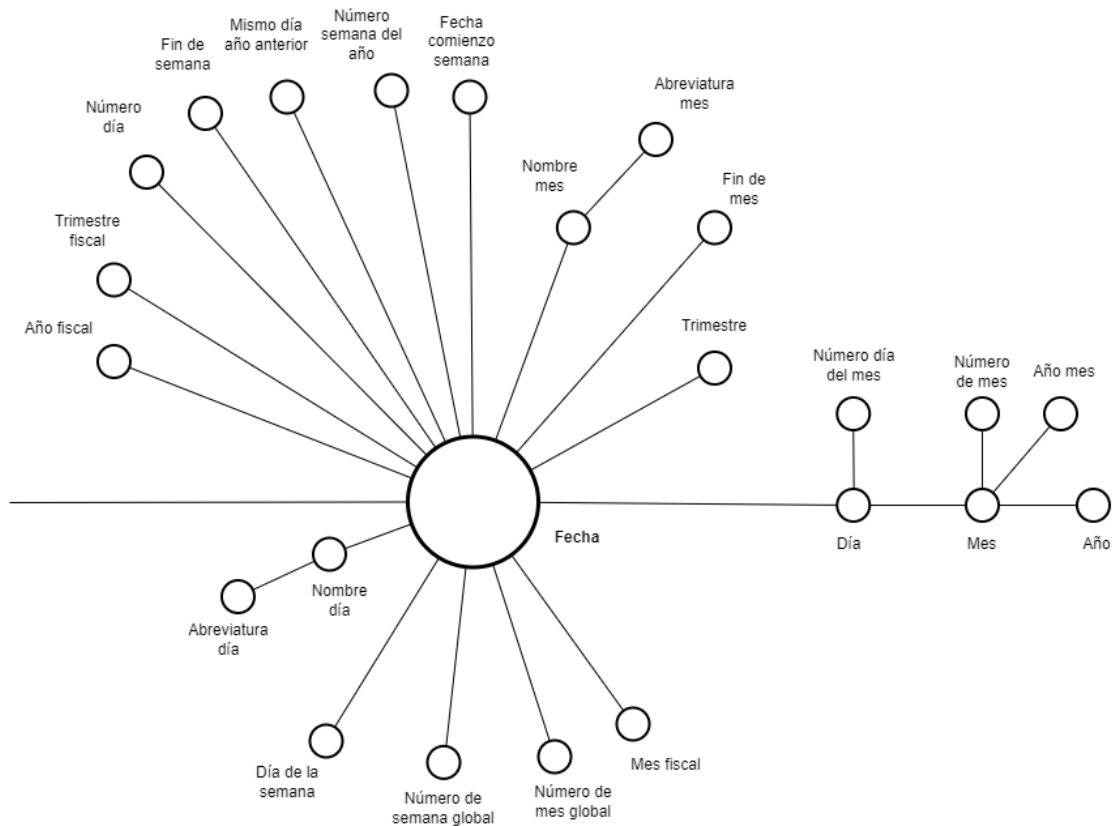


Figura 2. Dimensión fecha.

### Descartes de atributos y métricas.

El número de hombres y de mujeres que han asistido como métrica que, aunque parezca un dato interesante, consideramos que no aporta nada a la toma de decisiones en cuanto a la organización de las actuaciones dentro de un festival.

Fecha de inicio y fecha de fin del festival. Pertenece a un atributo compartido Fecha. Un festival puede durar varios días. No nos aporta más información útil, con conocer la fecha y hora de la actuación es suficiente.

El tipo de abono como atributo de festival (nombre y precio de éste). Especifica los tipos de abonos que se venderán para un festival en concreto. No nos aporta más información útil para este caso.

La media de los minutos en los que los asistentes llegan antes del comienzo de la actuación (adelanto) y la media de los minutos en que los asistentes llegan después del comienzo de la actuación (retraso). No nos aporta más información útil para este caso.



## Dimensiones que cambian lentamente en el tiempo.

La siguiente dimensión puede sufrir de algún cambio en algún momento del tiempo, pero será poco frecuente.

Dimensión Recinto.

La **capacidad** de un recinto podrá cambiar lentamente en el tiempo debido a normas, legislaciones o reformas que impliquen un cambio.

Estrategia por seguir.

La estrategia por seguir frente a algún cambio para las dimensiones anteriores será la **SCD2**, añadiendo una nueva fila con el nuevo valor y dejando sin modificar la fila ya existente, puesto que todos los valores anteriores nos siguen interesando para los análisis.

## Métricas Aditivas.

En esta categoría entrar los casos en los que la suma de dicha métrica corresponda al resultado total de un festival.

- Los **minutos de duración** de dicha actuación. Se considera que no se celebran diferentes actuaciones al mismo tiempo. En caso de que en un futuro se permitiese celebrar más de una actuación (en diferentes recintos) en el mismo momento, la métrica pasaría a ser no aditiva al no corresponderse con la duración real del festival.
- El **número de altercados** registrados.
- El **número de consumiciones** vendidas durante la actuación.
- Los **ingresos obtenidos por la venta de consumiciones** durante la actuación. Consideramos que es interesante medir los ingresos aparte del número de consumiciones; p.ej. en el caso de tener un número aproximado de consumiciones en dos actuaciones, puede existir una gran diferencia en los ingresos si en una actuación se han consumido más bebidas alcohólicas (generalmente más costosas) o no.

## Métricas no aditivas.

Todas las **medias** y **porcentajes** realizados, al no tratarse del resultado de la suma para todas las actuaciones de un festival, pertenecerán a esta categoría.

- El **número de personas que asistieron** a la actuación (asistentes). La suma de los asistentes de cada actuación no corresponde al número total de asistentes de un festival.
- La **media de las valoraciones** obtenidas por los asistentes. Esta valoración consiste en una puntuación, otorgada por un asistente, según le haya gustado más o menos la actuación.
- La **media de los minutos en que los asistentes han permanecido** en la actuación (permanencia).

- El **porcentaje de ocupación del recinto** respecto a su capacidad. Se calcula con el número de asistentes que tuvo la actuación entre la capacidad del recinto.
- El **porcentaje de asistencia respecto a la del festival**. Nos sirve para conocer a qué actuación asistieron más personas respecto a la asistencia total del propio festival. Se calcula con el número de asistentes que tuvo la actuación entre el número de asistentes del festival.

## Diseño lógico.

---

Se ha escogido un **esquema multidimensional en estrella** dado que solo vamos a tener una tabla de hechos y va a estar directamente relacionada con todas las dimensiones mostradas en el DFM, con la particularidad de que tiene una tabla “bridge” para un atributo multivaluado.

Para reflejar que “**genero**” es un atributo multivaluado (puesto que cada artista puede tener más de un género) se ha decidido crear una tabla “genero” conforme a los géneros musicales existentes y una tabla *bridge* llamada “**generoartista**” para poder unir a los artistas con varios de estos géneros. Esta tabla también incluirá un atributo **peso**, que refleja la ponderación del género.

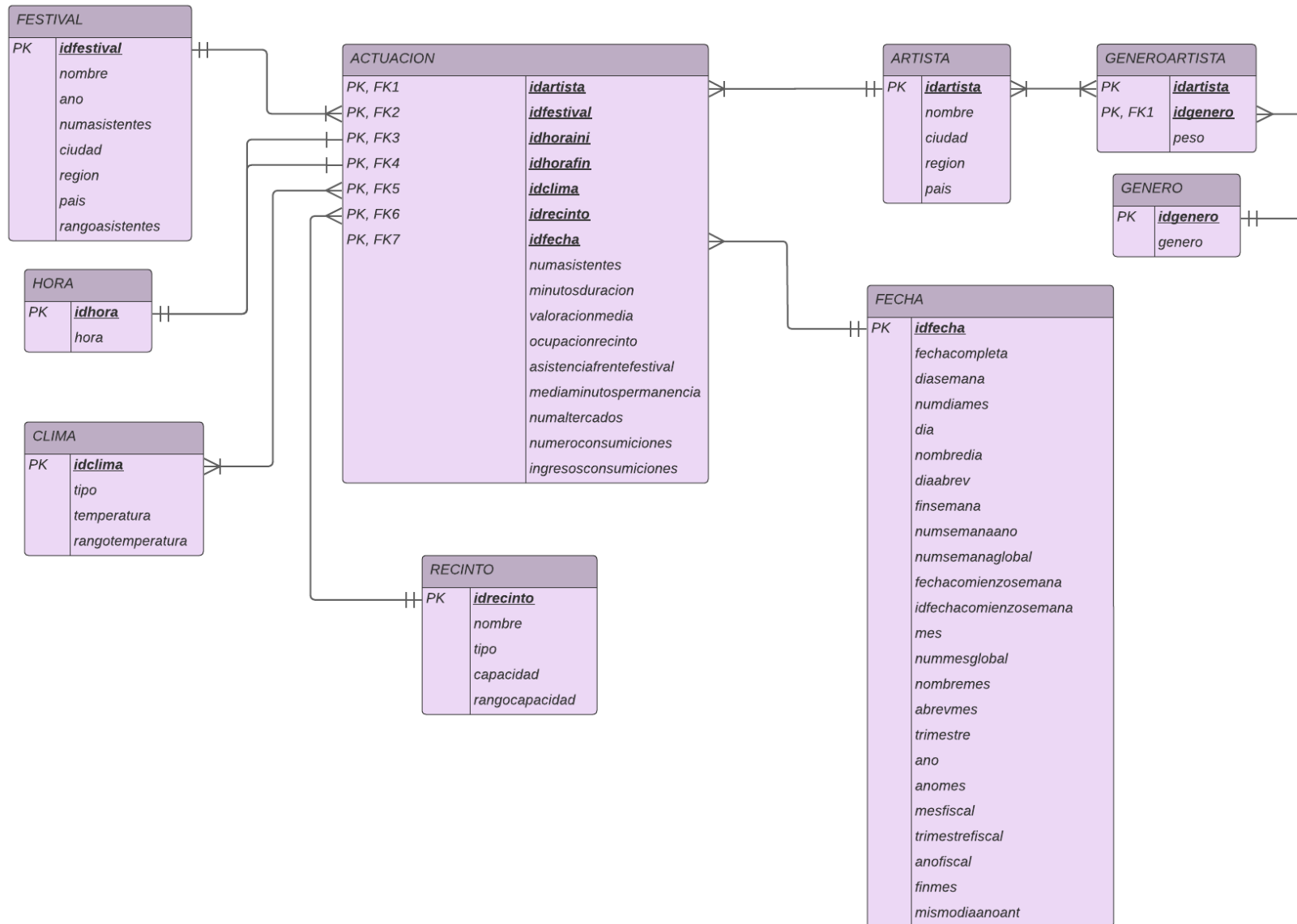
En cuanto a la jerarquía compartida “**hora**” por “**horaini**” y “**horafin**”, hemos creado una tabla a la que harán referencia “idhoraini” e “idhorafin” desde la tabla “actuacion”.

En aquellas dimensiones con atributos que mostraríamos en rangos, se les añade una variable adicional “rango<nombreatributo>”. De esta forma, para las tablas “clima”, “recinto” y “festival” se ha añadido “**rangotemperatura**”, “**rangocapacidad**” y “**rangoasistentes**”.

Para la dimensión “**fecha**” se ha decidido añadir múltiples atributos para poder hacer consultas mucho más precisas en el tiempo que ocurrieron las actuaciones.

Por simplificación, se han quitado de este diseño los atributos “**festivo**” y “**estacion**”, de la dimensión fecha, y el atributo “**provincia**” de las dimensiones “festival” y “artista” para simplificar la inserción de los datos (en este último caso, las provincias irían en el campo región).

Figura 3 . Modelo lóxico.



## Consultas analíticas.

---

La implementación de las consultas está en el fichero *consultas.sql*. A continuación, haremos una breve explicación de cada una:

- **Primera consulta, usando PIVOT.** Para cada artista estadounidense, queremos ver el número de actuaciones que han realizado dependiendo del tipo de recinto (o cerrado o aire libre). De esta forma, podríamos ver si los artistas de este país tienen alguna preferencia para el tipo de recinto para sus actuaciones, pudiendo tenerlo en cuenta para futuros conciertos.
- **Segunda consulta, usando funciones analíticas.** Para cada actuación de cada festival, donde la temperatura fue mayor de 20 grados, queremos que se muestre la fecha de la actuación, el artista que realiza el concierto, el festival al que pertenece la actuación, el número de consumiciones que hubo en la actuación, el número de consumiciones mínimo del mes que se ha realizado dicha actuación y el número de consumiciones mínimo histórico para dicho mes hasta ese momento. Obtenemos así información de gran valor para conocer los meses del año en que los asistentes consumen menos bebidas, planificando para estos menos conciertos.
- **Tercera consulta, usando CUBE.** Para aquellos conciertos con una valoración media mayor que 5, queremos ver el número de asistentes según el artista, el número de asistentes por festival, el número de asistentes por artista y festival y el número total de asistentes. Dicho resultado queremos verlo ordenado por el mayor número de asistentes. Así, obtenemos información sobre el éxito de asistencia que tiene un festival y un artista concreto en un festival pudiendo potenciar estos para llegar a cotas superiores.