

UniversidadeVigo

ESCOLA SUPERIOR DE ENXEÑARÍA INFORMÁTICA

Código Grupo: 2.2

Título: ESCUDERIA

Alumna/o: Álvaro Nóvoa Fernández

Alumna/o: Miguel Fernández Taboada

Alumna/o: Rubén Gómez Martínez

ÍNDICE

1. Descripción del proceso	2
2. Selección de la granularidad	6
3. Identificación de las dimensiones	7
4. Selección de las medidas asociadas al hecho	8
5. Almacenamiento de valores precalculados en la tabla de hechos	9
6. Terminación de las tablas de dimensión	10
7. Selección de la duración de la base de datos	14
8. Control de las dimensiones lentamente cambiantes	14
9. Modos de consulta	15

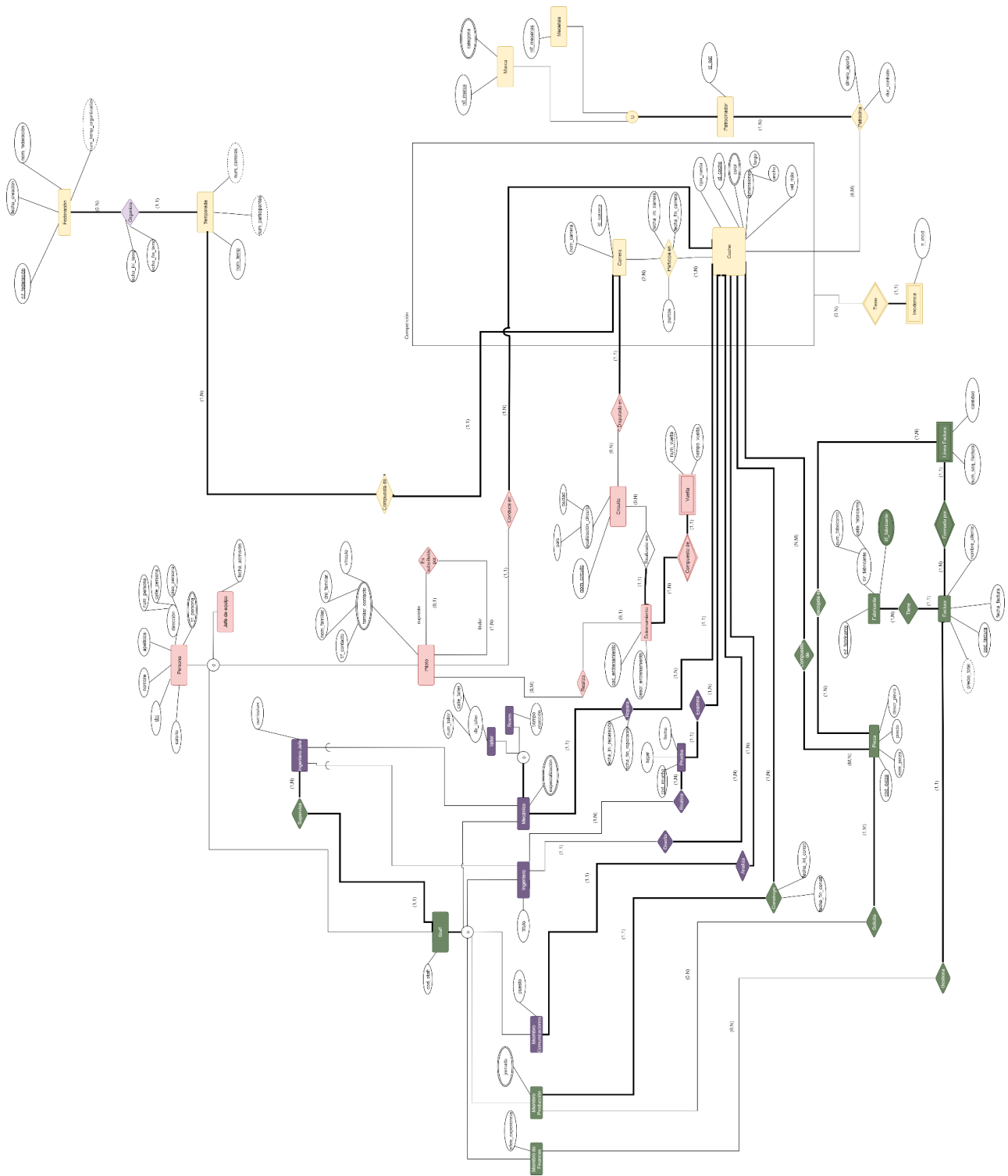
1. Descripción del proceso

1.1 Descripción textual de la organización

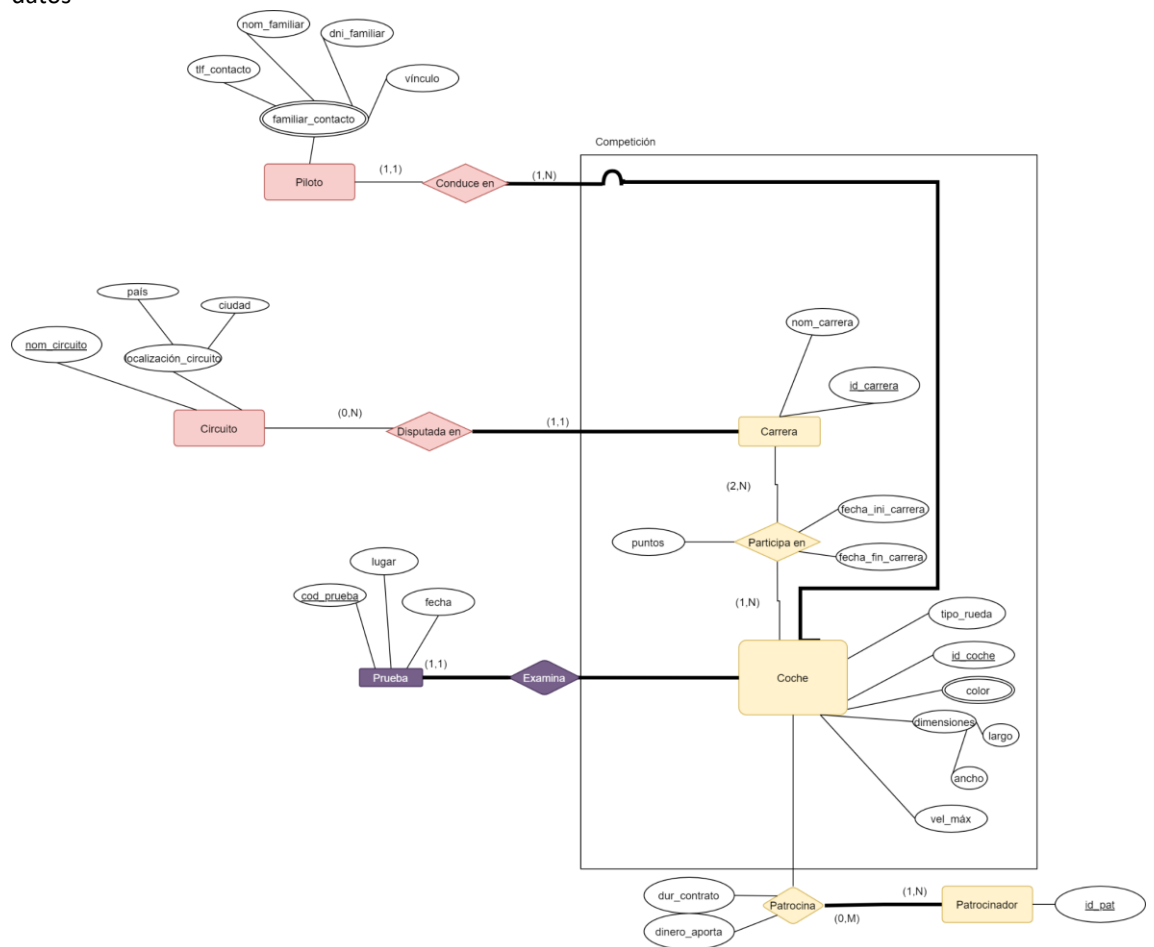
Una escudería de Fórmula 1 es una organización altamente especializada y compleja que se dedica a competir en el campeonato mundial de Fórmula 1. Consta de **pilotos** con sus características físicas y con sus respectivos **coches**, los cuales serán puestos a **prueba** en diversos **circuitos** en diferentes condiciones meteorológicas.

1.2 Modelo E/R original correspondiente al sistema OLTP preexistente

MODELADO MULTIDIMENSIONAL



1.3 Modelo E/R con las entidades que almacenan información de interés para el mercado de datos



1.3.1 Explicar la necesidad de incorporación de nuevos atributos a los preexistentes en la organización, en su caso.

Para un mayor análisis y una mejor obtención de resultados a posteriori hemos añadido atributos:

- **Circuito**

- Num_Curvas: atributo discretizado para saber si el circuito tiene pocas, algunas o muchas curvas.

- **Piloto**

- Peso: atributo para ver cómo influye en el rendimiento y resultado de las carreras

- anhos_experiencia: atributo para ver cómo influye en el rendimiento y resultado de las carreras

- **Coche**

- Potencia: atributo para ver cómo influye en el rendimiento y resultado de las carreras

1.4 Descripción de las fuentes externas

FUENTE	Fichero JSON creado con Mockaroo	
Descripción	Organización meteorológica de todas las ciudades del mundo	
ATRIBUTO	TIPO	DESCRIPCIÓN
Precipitación	decimal	Probabilidad en porcentaje de precipitación que habrá durante la carrera
Temperatura	decimal	Temperatura estimada en la ciudad de la carrera en una fecha
Viento	decimal	Velocidad a la que sopla el viento en KM/h durante la carrera

FUENTE	Fichero CSV creado con Mockaroo	
Descripción	Página con información de los diferentes circuitos de la competición	
ATRIBUTO	TIPO	DESCRIPCIÓN
Longitud	decimal	Distancia en KM de las pistas
Desgaste Frenos	char	Información de cuantos se desgastan los frenos en el circuito (BAJO, MEDIO o ALTO)
id_circuito	char	Id del circuito

1.5 Descripción textual de la actividad a modelar

La actividad a modelar es la eficiencia, eficacia y rendimiento de cada piloto y su respectivo coche en los circuitos de cada temporada. Observando cómo influyen tanto factores internos (pruebas a los coches, experiencia del piloto ...) como factores externos (número de patrocinadores y el dinero que aportan, clima ...).

2. Selección de la granularidad

2.1 Descripción del gránulo de la actividad a modelar

- Se desea almacenar **cuánto dinero** ha aportado **cada patrocinador** y durante **cuánto tiempo**.
- Se desea almacenar el **número de curvas** de **cada circuito** y sobre **qué tiempo** se ha corrido.
- Se desea conocer tanto las **dimensiones** como la **potencia** y **velocidad máxima** del coche que ha corrido en relación a su **porcentaje de pruebas exitosas**.
- Se desea almacenar el **día** de la **semana** del **año** que corre el coche.
- Se desea almacenar el **peso** y la **altura** del piloto que corre en relación a su **experiencia** y **sexo** o **edad**.

2.2 Razonar por qué no se ha optado por una granularidad mayor o menor

Hemos decidido por un nivel de granularidad bastante definido ya que múltiples pequeños factores pueden llegar a afectar en gran medida a deportes de alta velocidad como son las carreras. Estamos seguros de que cualquiera de estos datos influenciará enormemente en los resultados obtenidos.

3. Identificación de las dimensiones

3.1 Descripción textual de las dimensiones (a nivel general, sin detalle de los atributos)

DIM_PILOTO

Responde a la pregunta “Quién corre con el coche”, se refiere específicamente a los pilotos, con características concretas como la edad, los años de experiencia o si es un piloto activo.

DIM_TIEMPO

Responde a la pregunta “Cuándo corre un coche”, nos permite hacer búsquedas según diversas medidas de tiempo tales como carreras en un año, mes o día concreto.

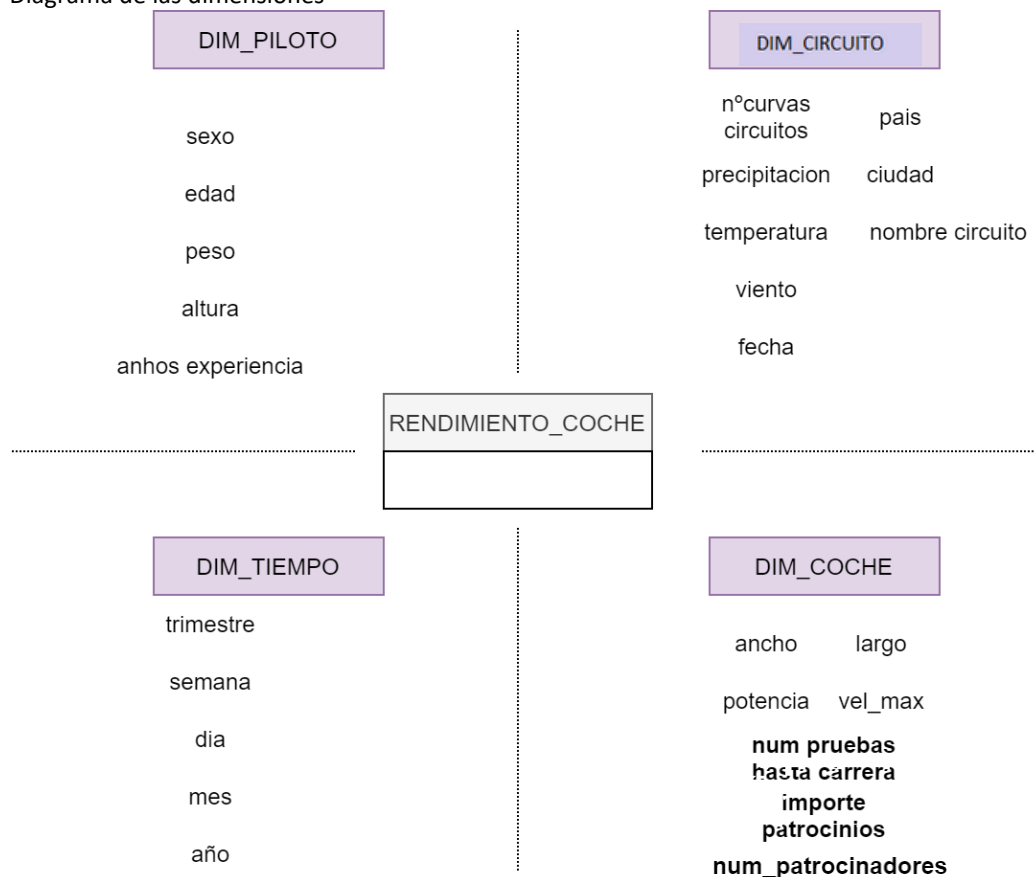
DIM_COCHE

Responde a la pregunta “Que coche corre en una carrera” y se centra en las características físicas del coche tales como la potencia o sus dimensiones.

DIM_CIRCUITO

Responde a la pregunta “Dónde es la carrera” y aúna características del circuito como la meteorología, nº curvas o el país.

3.2 Diagrama de las dimensiones



4. Selección de las medidas asociadas al hecho

4.1 Descripción textual de las medidas y de dónde se obtienen (a qué atributo(s) se corresponde en la BD OLTP).

MODELADO MULTIDIMENSIONAL

MEDIDA	TIPO	DESCRIPCIÓN	TABLA/FUENTE ORIGIN	ATRIBUTO/CA MPO ORIGIN
dinero_patrocinios	decimal	Cantidad total de dinero invertido en el coche para una carrera por todos los patrocinadores	PATROCINA	dinero_aporta
%pruebas	decimal	Porcentaje de pruebas acabadas por el coche antes de una carrera	PRUEBA	id_prueba
mejores vueltas	decimal	Número de veces en el que el coche ha hecho la mejor vuelta de una carrera	PARTICIPA_EN	mejor_vuelta

4.2 Diagrama con la incorporación de las medidas

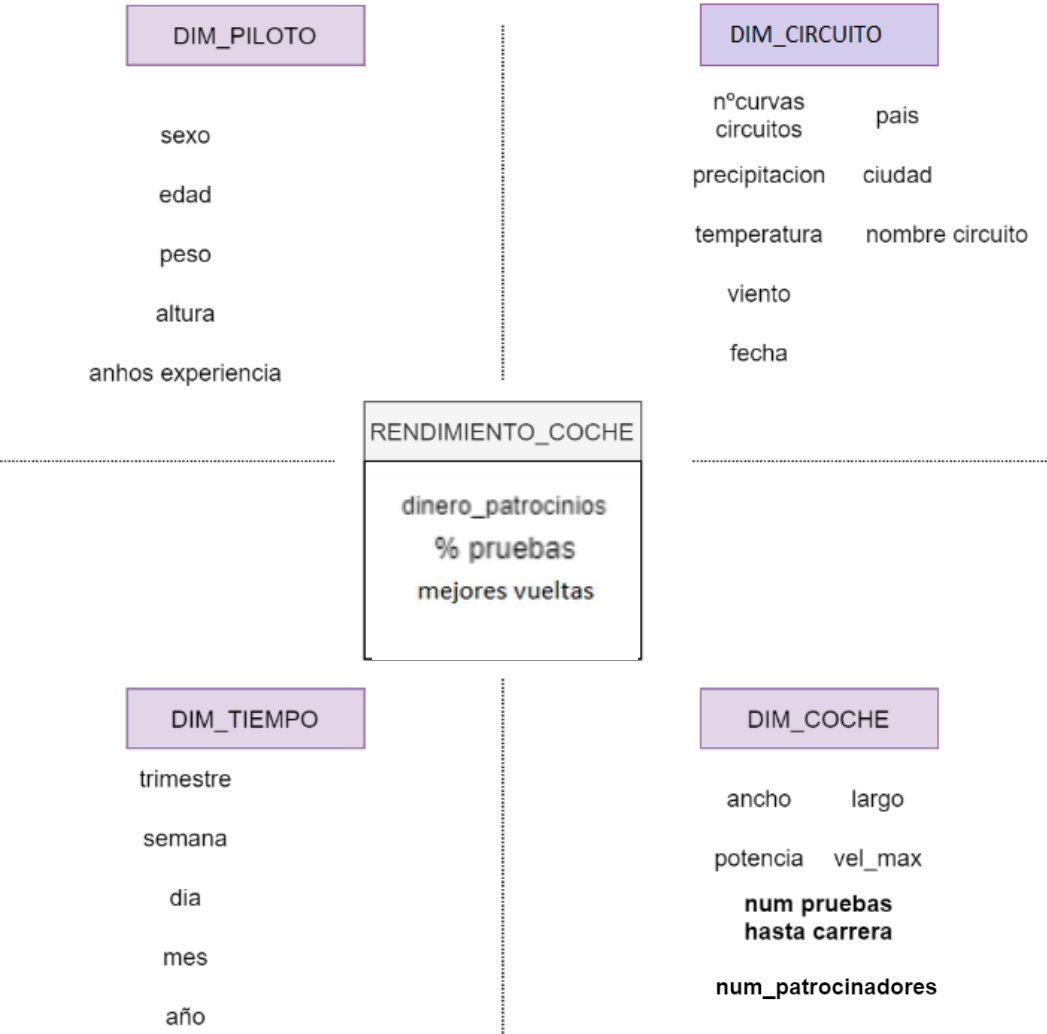


5. Almacenamiento de valores precalculados en la tabla de hechos

5.1 Definición de valores precalculados y descripción acerca de cómo se obtienen

- % de pruebas exitosas: relación entre el número de pruebas exitosas y el número de pruebas totales de un coche en específico para una carrera.
- dinero_patrocinios: suma de todo el dinero aportado por todos los patrocinadores para el coche antes de una carrera
- mejores_vueltas: cuenta de todas las carreras dónde el coche ha obtenido el mejor tiempo

5.2 Diagrama con incorporación de los valores precalculados



6. Terminación de las tablas de dimensión

6.1 Descripción detallada de las dimensiones. Para cada uno de sus atributos indicar:

6.1.1 Cómo se obtiene (atributo(s) específico(s) en OLTP, campo calculado, fuente externa X, etc.)

DIMENSIÓN	DIM_PILOTO			
Descripción de la dimensión	Se refiere específicamente a los pilotos, con características concretas como la edad, los años de experiencia o si es un piloto activo.			
ATRIBUTO	TIPO	DESCRIPCIÓN	TABLA/FUENTE ORIGIN	ATRIBUTO/CAMPO ORIGIN
dni_piloto	Varchar	Clave OLTP	PILOTO	id_piloto
qc_id	smallint	Clave subrogada		
sexo	char	Sexo del piloto	PILOTO	sexo
edad	number	Edad del piloto	PILOTO	edad
peso	number	Peso del piloto	PILOTO	peso
altura	number	Altura del piloto	PILOTO	altura
anho_experiencia	number	Años conducción piloto	PILOTO	experiencia
valido_desde	datetime	Fecha inicio lentamente cambiante		
valido_hasta	datetime	Fecha fin lentamente cambiante		
num_version	smallint	Version para dimensión lentamente cambiante		

DIMENSIÓN	DIM_TIEMPO			
Descripción de la dimensión	Permite hacer búsquedas según diversas medidas de tiempo tales como carreras en un año, mes o día concreto.			
ATRIBUTO	TIPO	DESCRIPCIÓN	TABLA/FUENTE ORIGIN	ATRIBUTO/CAMPO ORIGIN
cc_id	smallint	Clave subrogada		
anho	date	Año carrera		
trimestre	date	Trimestre carrera		
mes	date	Mes carrera		
semana	date	Semana carrera		
día	date	Día carrera		

DIMENSIÓN	DIM_CIRCUITO			
Descripción de la dimensión	Aúna características del circuito como la meteorología, nº curvas de un circuito o el país de un circuito.			
ATRIBUTO	TIPO	DESCRIPCIÓN	TABLA/FUENTE ORIGIN	ATRIBUTO/CAMPO ORIGIN
nom_circuito	smallint	Clave OLTP	CIRCUITO	nom_circuito
dc_id	smallint	Clave subrogada		
n_curvas	varchar	Atributo discretizado de curvas en 3 valores:	CIRCUITO	n_curvas

MODELADO MULTIDIMENSIONAL

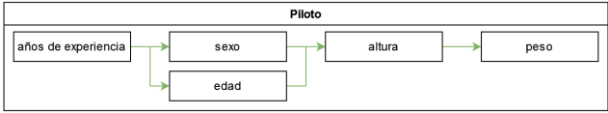
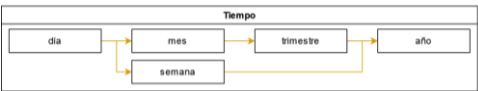
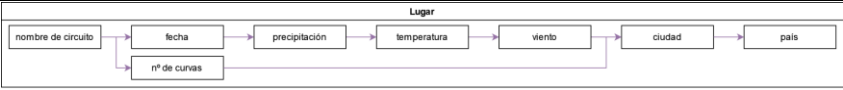
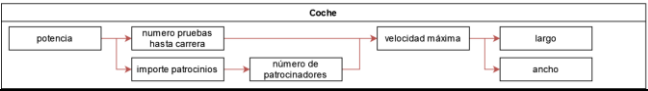
		POCAS, ALGUNAS o MUCHAS		
precipitacion	decimal	Probabilidad en porcentaje de precipitación que habrá durante la carrera	CIRCUITO	meteorología(fuente externa)
temperatura	decimal	Temperatura estimada en la ciudad de la carrera en una fecha	CIRCUITO	meteorología(fuente externa)
viento	decimal	Velocidad a la que sopla el viento en KM/h durante la carrera	CIRCUITO	meteorología(fuente externa)
pais	varchar	País del circuito	CIRCUITO	pais
ciudad	varchar	Ciudad del circuito	CIRCUITO	ciudad
desgaste_frenos	varchar	Cantidad desgaste que da el circuito al coche: BAJO, MEDIO, ALTO		fuentes externa CSV
longitud	number	Longitud km circuito		fuentes externa CSV

DIMENSIÓN	DIM_COCHE			
Descripción de la dimensión	Se centra en las características físicas del coche tales como la potencia o sus dimensiones.			
ATRIBUTO	TIPO	DESCRIPCIÓN	TABLA/FUENTE ORIGIN	ATRIBUTO/CAMPO ORIGIN
id_coche	smallint	Clave OLTP	COCHE	id_coche
qcoche_id	smallint	Clave subrogada		
ancho	decimal	Ancho en centímetros en un coche	COCHE	ancho
largo	decimal	Largo en centímetros de un coche	COCHE	largo
potencia	decimal	Potencia en CV del coche	COCHE	potencia_motor
vel_max	decimal	Velocidad del coche	COCHE	vel_max
n_pruebas_exitosas	decimal	Número de pruebas exitosas que realiza un coche(es calculado de todas las pruebas realizadas)	COCHE	CALCULADO POR NÚMERO DE PRUEBAS QUE HACE UN COCHE
importe_patrocinios	decimal	Presupuesto invertido en el coche	PATROCINADOR	Suma del presupuesto de todos los patrocinadores en un coche
num_patrocinadores	decimal	Número de patrocinadores de un coche	PATROCINADOR	CALCULADO POR EL NÚMERO DE

MODELADO MULTIDIMENSIONAL

				PATROCINADORES DE UN COCHE
--	--	--	--	-------------------------------

6.1.2 Establecimiento de la jerarquía entre los atributos de cada dimensión (p.ej: día, mes, año)

DIMENSIÓN	DIM_PILOTO
Jerarquía	
DIMENSIÓN	DIM_TIEMPO
Jerarquía	
DIMENSIÓN	DIM_CIRCUITO
Jerarquía	
DIMENSIÓN	DIM_COCHE
Jerarquía	

6.1.3 Conversión de datos: numerización/etiquetado, discretización (hacer al menos 2)

DIMENSIÓN	DIM_CIRCUITO	
ATRIBUTO	TIPO DE CONVERSIÓN	DESCRIPCIÓN
n_curvas	De dígito a POCAS, ALGUNAS o MUCHAS	Se clasifica el número de curvas de un circuito en 3 tipos: POCAS(0 a 2), ALGUNAS(3,6) o MUCHAS(más de 7).
dur_contrato	De dígito a CORTO o LARGO	Se clasifica la duración de un contrato en 2 tipos: CORTO(0 a 1 año), LARGO(más de 2 años).

6.1.4 Cómo se debe tratar en caso de datos erróneos

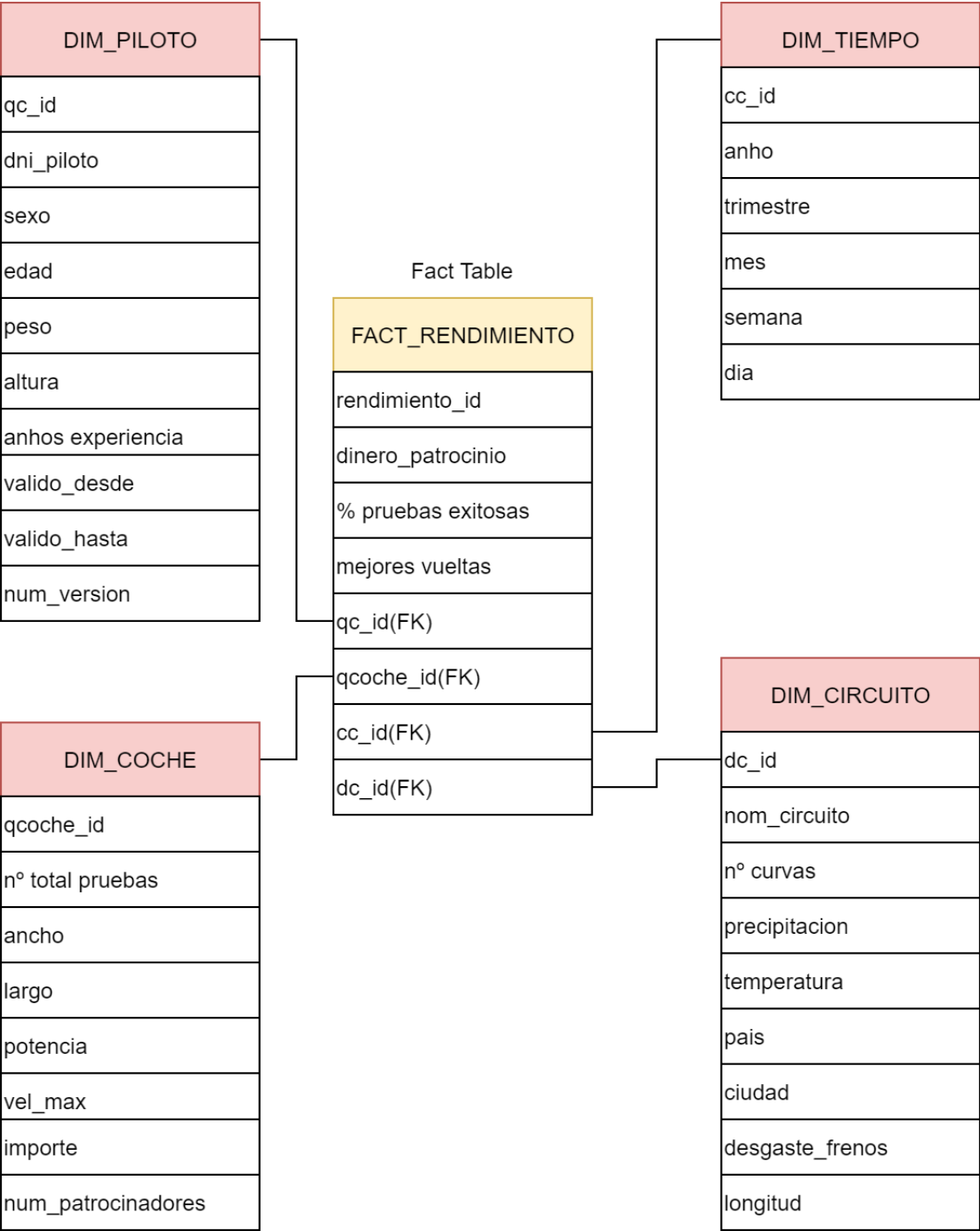
DIMENSIÓN	DÓNDE CORRE
ATRIBUTO	DESCRIPCIÓN DEL TIPO DE ERROR Y CÓMO RESOLVERLO
precipitacion	Si está mal consideramos NADA *
temperatura	En caso de faltar considerar NADA*
viento	En caso de faltar considerar NADA*

6.1.5 Cómo se debe tratar en caso de datos faltantes

DIMENSIÓN	DÓNDE CORRE
ATRIBUTO	DESCRIPCIÓN DEL TIPO DE PROBLEMA Y CÓMO RESOLVERLO
precipitacion	En caso de faltar considerar NADA*
temperatura	En caso de faltar considerar NADA*
viento	En caso de faltar considerar NADA*
nº curvas	En caso de faltar considerar POCAS

* Lo acabaremos eliminando al ir realizando las transformaciones

6.2 Diagrama final de diseño de la BD Datawarehouse versión ROLAP (incorporando claves)



7. Selección de la duración de la base de datos

7.1 Indicación de la duración elegida y justificación

Para evitar problemas de dimensiones lentamente cambiantes y para no dificultar la obtención e interpretación de los datos, decidimos marcar la duración de la base de datos un periodo de tres años. Tiempo suficiente para poder hacer comparaciones útiles con situaciones actuales de la organización y ver su evolución.

8. Control de las dimensiones lentamente cambiantes

8.1 Determinación de las dimensiones lentamente cambiantes y cuáles son los atributos afectados.

DIM_PILOTO: ANHOS_EXPERIENCIA

8.2 Para cada uno de los atributos cambiantes indicar cómo será tratado (Tipo 1, 2 o 3). *Es necesario incluir al menos un atributo cambiante Tipo 2.*

anhos_experiencia- TIPO 2

9. Modos de consulta

9.1 Descripción textual de las consultas OLAP

MEDIDAS: <i>"Mostrar la evolución de..."</i>	
- <i>vel_max alcanzada</i>	- <i>nº puntos</i>
- <i>% pruebas exitosas</i>	- <i>mejores vueltas</i>

QUIÉN CORRE	QUÉ COCHE	CUÁNDO CORRE	DÓNDE CORRE
Frase descriptiva	Frase descriptiva	Frase descriptiva	Frase descriptiva
edad > 18	potencia > 300 cv	en el último trimestre del año	lloviendo
años experiencia >3	nº pruebas > 10	en el año XXXX	en determinada ciudad
peso < 80	vel_max < 400	los viernes	con cierto número de curvas