

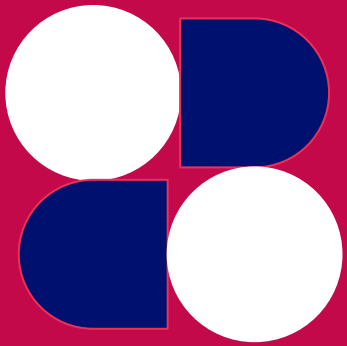


Programa PUCP DATA DRIVEN

Formación tecnológica

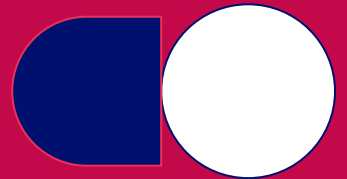
Dirección Académica de Planeamiento y Evaluación
Dirección de Gestión del Talento Humano





SESIÓN 8:

Transformaciones PowerBI





Al finalizar la sesión, lograrás:

- Identificar las diferentes formas de usar SQL en PowerBI
- Usar las consultas SQL manualmente en PowerBI
- Reconocer las diferencias entre los modelos dimensionales
- Implementar un modelo estrella en PowerBI



Contenido de la sesión:

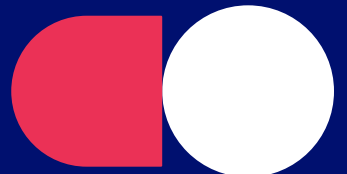


1 Usar SQL en PowerBI

2 Modelo Dimensional



1. Usar SQL en PowerBI





USAR SQL EN POWERBI



Opciones

- Fácil
- No tan fácil
- A pedal (usar un intermediario)





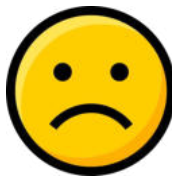
USAR SQL EN POWERBI



Opciones

- **Fácil:**
 - Utilizar una herramienta ETL externa (Como SQL Server) antes de cargar los datos a Power BI

**NO LOS TENEMOS, NO PODEMOS
USAR ESTA OPCIÓN**





Opciones

- **No tan fácil:**
 - Usar Vistas o Tablas alternas dentro de la base de datos para almacenar la información ya filtrada y procesada con SQL, para después “jalarla” directamente a PowerBI Desktop y usar la Transformación de Datos.

- **Definición:**

- Una **VIEW** proporciona un mecanismo para ocultar ciertos datos de la visualización de ciertos usuarios.
- Cualquier relación que no sea del modelo relacional pero que se haga visible para un usuario como una "relación virtual" se denomina **VIEW** (vista)





USAR SQL EN POWERBI



Opciones

- **No tan fácil:**
 - Usar Vistas o Tablas alternas dentro de la base de datos para almacenar la información ya filtrada y procesada con SQL, para después “jalarla” directamente a PowerBI Desktop y usar la Transformación de Datos.

Power BI Service está diseñado para consultar datos y no permite cambiar el esquema (DDL) ni manipular los datos (DML) directamente dentro de un Data Mart. Sería necesario descargar los datos a un motor como PostgreSQL y, desde allí, crear las vistas. NO PODEMOS USAR ESTA OPCIÓN





Opciones

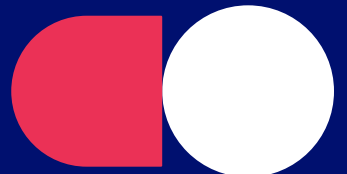
- **A pedal:**
 - Generar las consultas con la información necesaria y descargar las tablas resultado a Excel, para después cargarlas a PowerBI Desktop y crear el modelo dimensional a transformar.

ESTO SI LO PODEMOS HACER





2. Modelo Dimensional



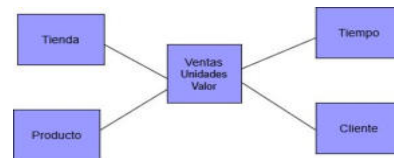


MODELO DIMENSIONAL



Principios

- **Hecho:** Son medidas de negocio
Es lo que se analiza
Normalmente solo contiene medidas numéricas de tipo estadística + relaciones con las dimensiones
Gran cantidad de filas
Ejemplo: Cantidad estudiantes inscritos por clase
- **Dimensión:** Representan los objetos de negocio
Se pueden obtener desde los procesos de negocio
Contiene los criterios de análisis
Jerarquías
Sólo se conecta con otras dimensiones a través de los hechos
Ejemplos: Cuándo, Docentes, Estudiantes, Cursos





MODELO DIMENSIONAL



Las Dimensiones, contienen por lo general, una llave simple y un conjunto de atributos que la describen. Estos son la fuente de gran parte de todas las necesidades empresariales que deben cubrirse, además, sirven de restricciones en la mayoría de las consultas.

Las tablas de hechos, representan los procesos que ocurren en la organización. Cada medida o hecho, se corresponde con una intersección de valores de las dimensiones.



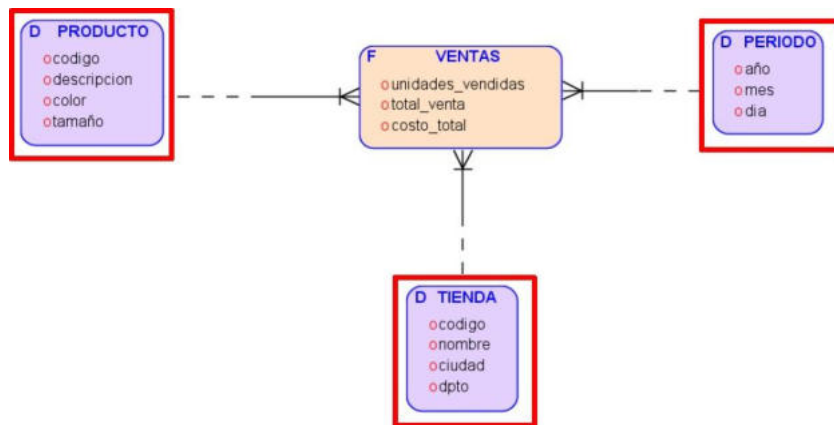


MODELO DIMENSIONAL



Dimensiones

- Están conformadas por atributos dimensionales
- Contienen descripciones textuales de los temas de negocio.
- Las dimensiones son independientes entre sí, no están relacionadas de manera jerárquica entre ellas.



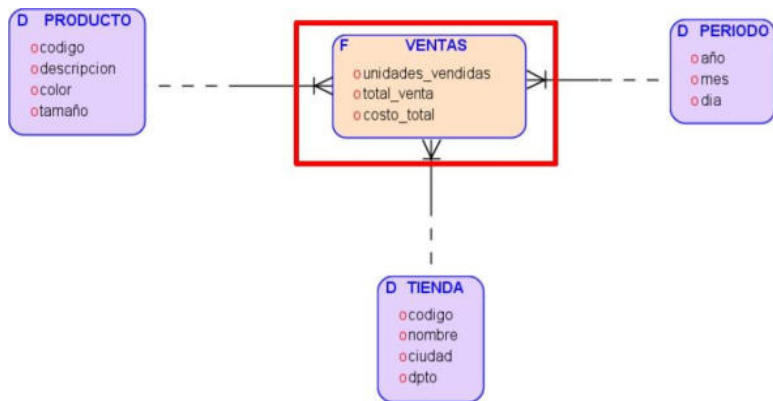


MODELO DIMENSIONAL



Hechos

- Son medidas de negocio
- Por lo general son variables numéricas
- Las medidas describen propiedades numéricas de un hecho o evento



prod	tienda	periodo	unidades - ventas	total_venta	costo_total
TV	Bogotá	01/07/2014	10	20000000	8000000
TV	Medellín	02/07/2014	20	45000000	5000000
DVD	Cali	01/08/2014	15	5000000	3000000
DVD	Bogotá	01/09/2014	30	1000000	2000000
...	...	...	...	...	...

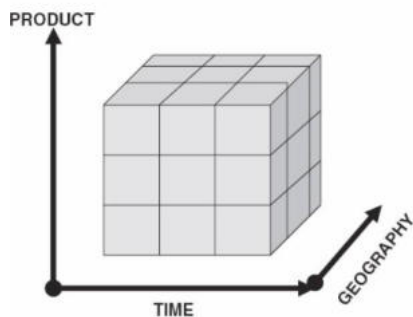


MODELO DIMENSIONAL

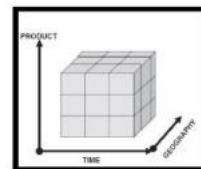


Cubos

- Al generar la descripción explícita del hecho con las traducciones de sus llaves foráneas que lo conforman, se crean los famosos **cubos**



producto	tiempo	geografia	unidades vendidas
TV	martes, 01 de julio de 2014	Chicago	10
TV	domingo, 01 de junio de 2014	Boston	20
DVD	viernes, 01 de agosto de 2014	Chicago	15
DVD	lunes, 01 de septiembre de 2014	Boston	30



- Si tenemos 100 productos, 1825 días y 1000 tiendas; el número máximo de hechos sería igual a: $100 \times 1825 \times 1000 = 182.500.000$



MODELO DIMENSIONAL



Dimensión TIEMPO/CALENDARIO

- Esta es una de las dimensiones que deberían garantizarse en un modelado dimensional.
 - Esta dimensión se constituye en una serie de tiempos.
 - Soporta análisis de tendencias y en general de análisis históricos.
 - Hay que generar sus filas
 - Poner atención a la granularidad
-
- Si el EDW o DM están bien diseñados, esta es la dimensión que permite viajar por el tiempo en los datos

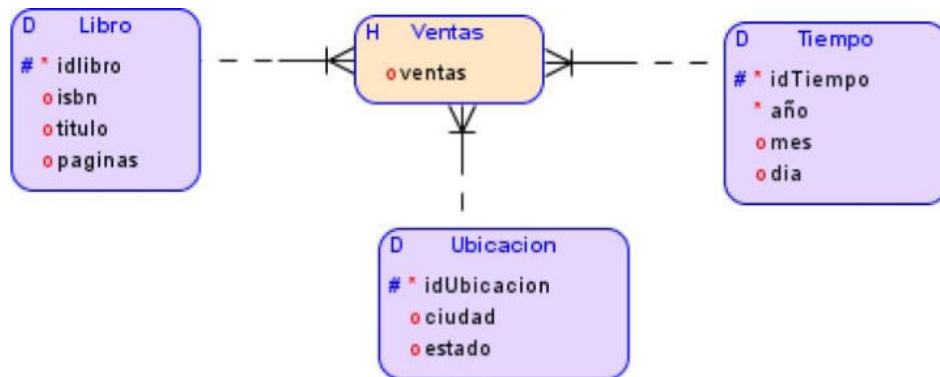
dim time	
time_key	bigint
date	nvarchar(12)
sql_date	datetime
day	tinyint
day_of_the_week	tinyint
day_name	nvarchar(15)
week_number	tinyint
month	tinyint
month_name	nvarchar(9)
short_month_name	nchar(3)
year	smallint
week_day	tinyint



MODELO DIMENSIONAL



Ejemplo



Libro			
IdLibro	ISBN	Titulo	Paginas
1	12548756	Cien años	200
2	257862344	primavera	100

Tiempo				
IdTiempo	Año	Mes	Día	H
1	2014	1	1	
2	2014	1	2	
3	2014	2	1	

Ventas				
IdLibro	IdTiempo	IdUbicacion	Ventas	H
1	3	2	200	
1	3	3	200	
1	4	2	200	
1	4	3	200	
1	5	2	200	
1	5	3	200	
2	3	2	200	
2	3	3	200	
2	4	2	200	
2	4	3	200	
2	5	2	200	
2	5	3	200	

Ubicacion		
IdUbicacion	Ciudad	Estado
2	bogota	cundinamarca
3	medellin	antioquia



MODELO DIMENSIONAL



Jerarquía en las Dimensiones

- Permiten hacer análisis con diferentes niveles de detalle

Year	Quarter	Month	Sales Order Header Count	Freight
Calendar 2006	Quarter 4, 2006	7	282,7964	
	Total	7	282,7964	
Calendar 2007	Quarter 1, 2007	1091	193457,757999999	
	Quarter 2, 2007	1260	243196,1653	
	Quarter 3, 2007	4152	412220,256599997	
	Quarter 4, 2007	5940	379805,159599997	
	Total	12443	1228679,339500002	
Calendar 2008	Quarter 1, 2008	6087	320610,600399998	
	Quarter 2, 2008	6888	406555,602699997	
	Quarter 3, 2008	976	1271,0419	
	Total	13951	728437,245000001	
Grand Total		26401	1957399,380900009	



MODELO DIMENSIONAL



Diseño Físico

- **Esquema en Estrella**
 - Dimensiones completamente desnormalizadas
- **Esquema en Copo de Nieve**
 - Dimensiones normalizadas

Las dos metodologías son igual de poderosas y en costo no hay un impacto significativo. Depende del negocio

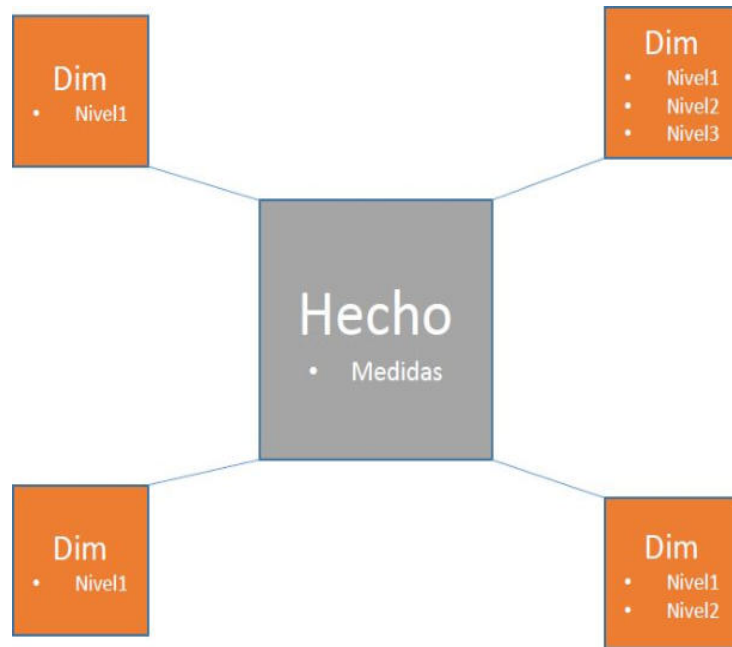


MODELO DIMENSIONAL



Diseño Físico

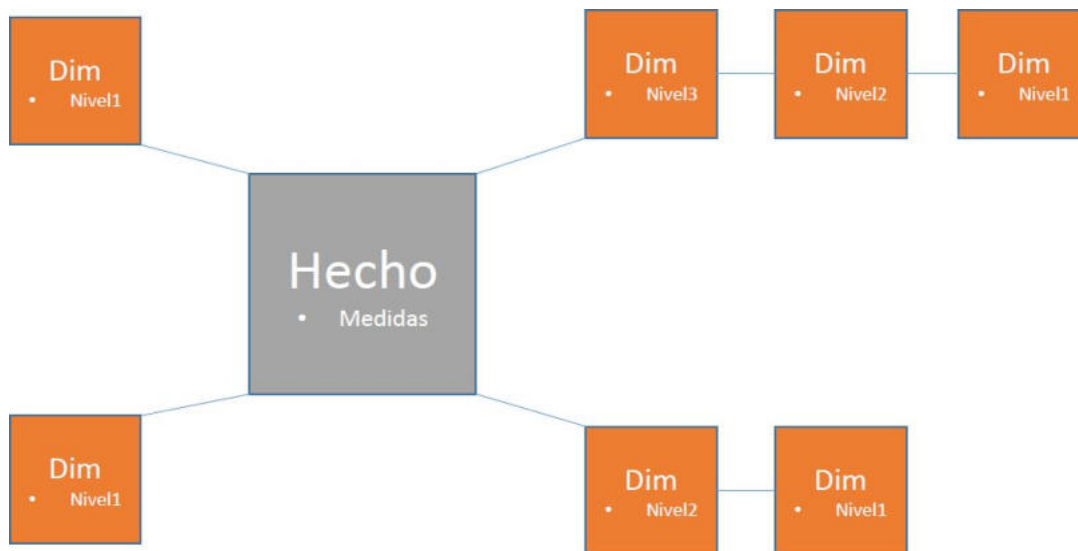
- Esquema en Estrella





Diseño Físico

- Esquema en Copo de Nieve





Ejemplo Estrella - Dimensión de Tiempo desnormalizada

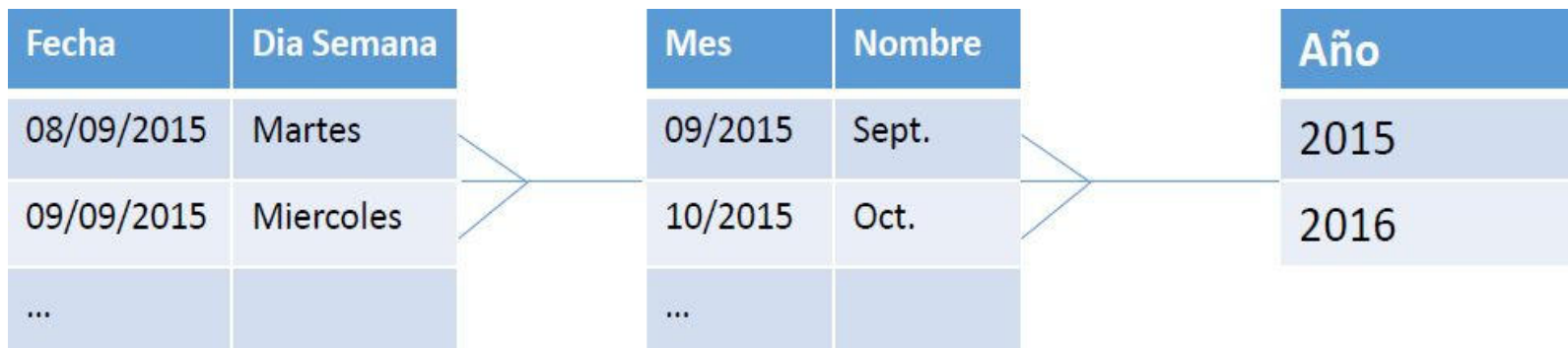
Fecha ID	Fecha	Dia Semana	Número Mes	Mes	Año
12654	09/09/15	Martes	9	Sept.	2015
12655	10/09/15	Miércoles	9	Sept.	2015
...					
12684	06/10/15	Martes	10	Oct	2015



MODELO DIMENSIONAL



Ejemplo Copo de Nieve - dimensión de tiempo





MODELO DIMENSIONAL



Esquema Estrella

- **Dimensiones**
 - Ejemplo de Jerarquía
 - Oficina, Ciudad, Departamento, Región
 - Todos los niveles en una única tabla.
 - Alta redundancia
 - Muy pocos Joins para las consultas



Copo de Nieve

- **Dimensiones**
 - Se mantienen normalizadas las dimensiones
 - Cada nivel de la jerarquía está en una tabla diferente
 - La tabla de hechos se relaciona con el nivel más detallado de la jerarquía
 - Más fácil de actualizar
 - Menos redundancia
 - Más Joins para una consulta agregada

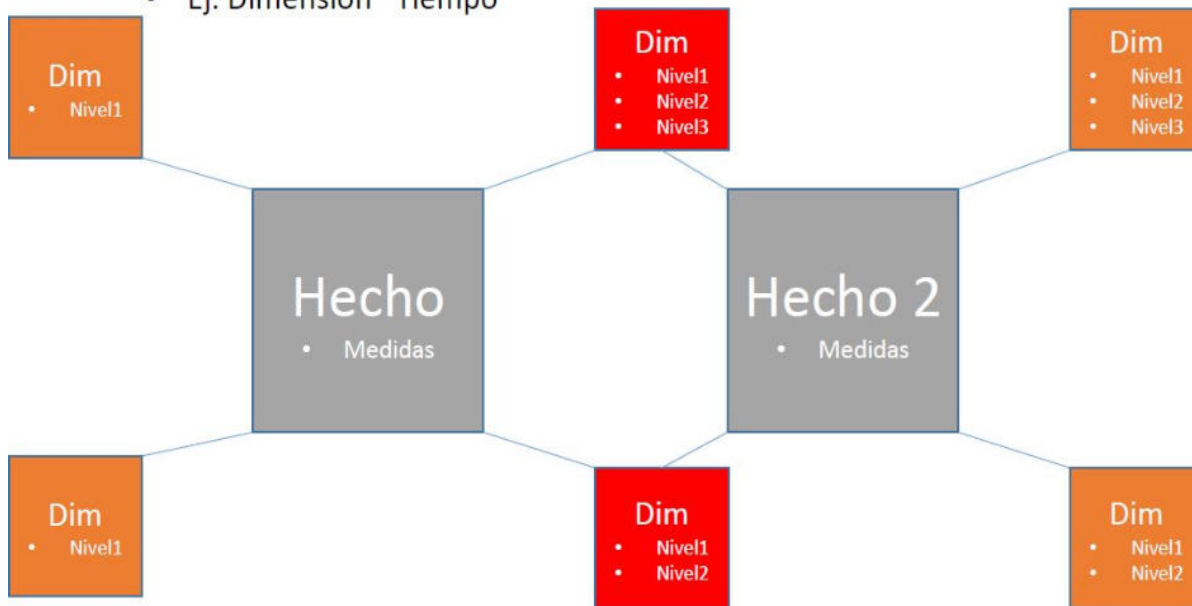


MODELO DIMENSIONAL



Dimensiones Conformadas

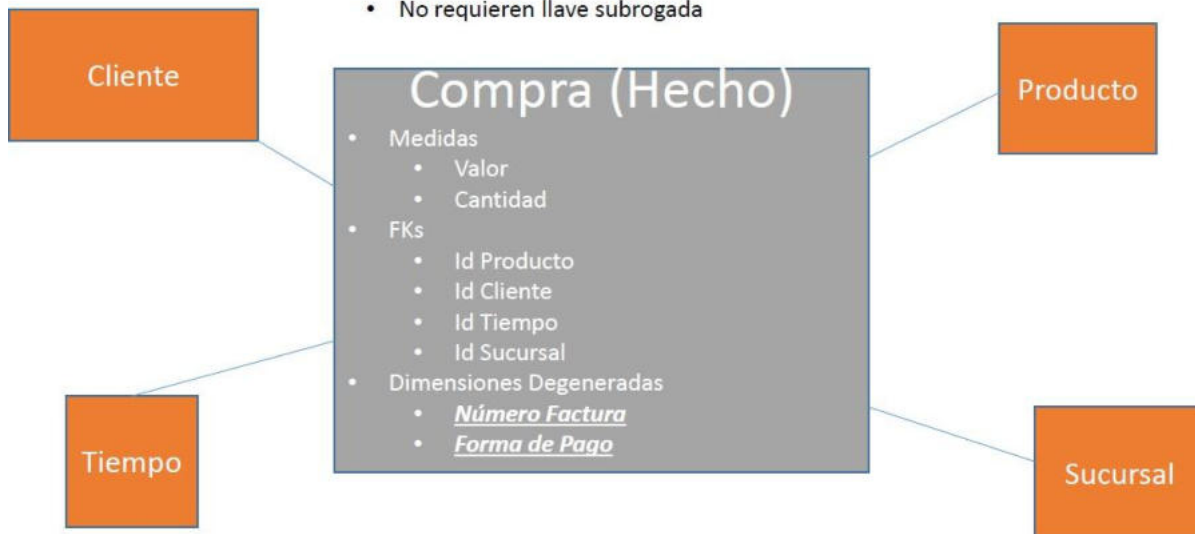
- Dimensiones relacionadas con varios Hechos
 - Ej. Dimensión "Tiempo"





Dimensiones Degeneradas

- Se incluyen en la tabla de hechos
- Muy pocos atributos (Id de negocio, Etiqueta)
- No cambian constantemente
- No requieren llave subrogada





MODELO DIMENSIONAL



OJO:

PowerBI no es una herramienta de analítica y **agradece** el uso de modelos dimensionales o cubos a la hora de generar los reportes





Ejercicio: Vamos a generar un cubo

Pregunta Problema: ¿Cuántos cursos y cuántas horas de clase imparte cada docente, y cómo se distribuyen estos cursos y horarios a lo largo de los ciclos? Extraiga toda la info de docentes, cursos y horarios.

Pistas:

- Necesitan la info del docente
- Necesitan la info de los cursos
- Necesitan la info de los horarios
- Crear una agregación y agrupar



estudiantes		
♂ ♦ #	id_estudiante	int8
♦	codigo_estudiante	text
◇	id_especialidad	int8
♦	apellidos_estudiante	text
♦	nombres_estudiante	text
◇	correo_estudiante	text
◇	dni	text
◇	sexo	text

especialidades		
♂ ♦ #	id_especialidad	int8
♦	nombre_especialidad	text
♦	codigo_especialidad	text
◇	rama	text
◇	seccion	text

inscripciones		
♂ ♦ #	id_inscripcion	int8
◇	id_estudiante	int8
◇	id_horario	int8

cursos		
♂ ♦ #	id_curso	int8
♦	nombre_curso	text
◇	creditaje	int4
◇	ciclo	int4

docentes		
♂ ♦ #	id_docente	int8
♦	codigo_docente	text
♦	nombres_docente	text
♦	apellidos_docente	text
◇	dedicacion_docente	text
◇	categoria_docente	text
◇	departamento_docente	text
◇	correo_docente	text
◇	dni	text
◇	sexo	text

horarios		
♂ ♦ #	id_horario	int8
◇	id_curso	int8
◇	id_docente	int8
◇	horario	text
◇	clave_curso	text
◇	tipo_horario	text
◇	dia	text
◇	hora_inicio	time
◇	hora_fin	time



Gracias

