

Instituto Tecnológico de Costa Rica
Centro Académico de Alajuela
Bases de Datos II

Grupo 8:

Marco Rodríguez Vargas – Carné: 2022149445

Kevin Carranza Jiménez – Carné: 2015100260

Jorge Esteban Benavides Castro – Carné: 2022230697

Profesor:

Alberto Shum Chan

Fecha de entrega:

8/9/2023 a las 16:59 pm

II Semestre, 2023

Tema: Inteligencia de negocios, diseño

Entrega en el TecDigital:

Deben presentar un archivo .PDF

Forma de trabajo en grupos de 3 personas.

Objetivos:

- Identificar dimensiones y medidas a partir de un caso de negocio
- Aplicar principios básicos de modelado dimensional
- Diseñar de forma general un proceso de ETL (extracción, transformación y carga).

Ejercicio práctico descripción

Utilizando la información sobre las fuentes de datos y el contexto de negocios que se presenta a continuación, debe especificar un modelo dimensional indicando: dimensiones identificadas y posibles atributos, medidas y granularidad.

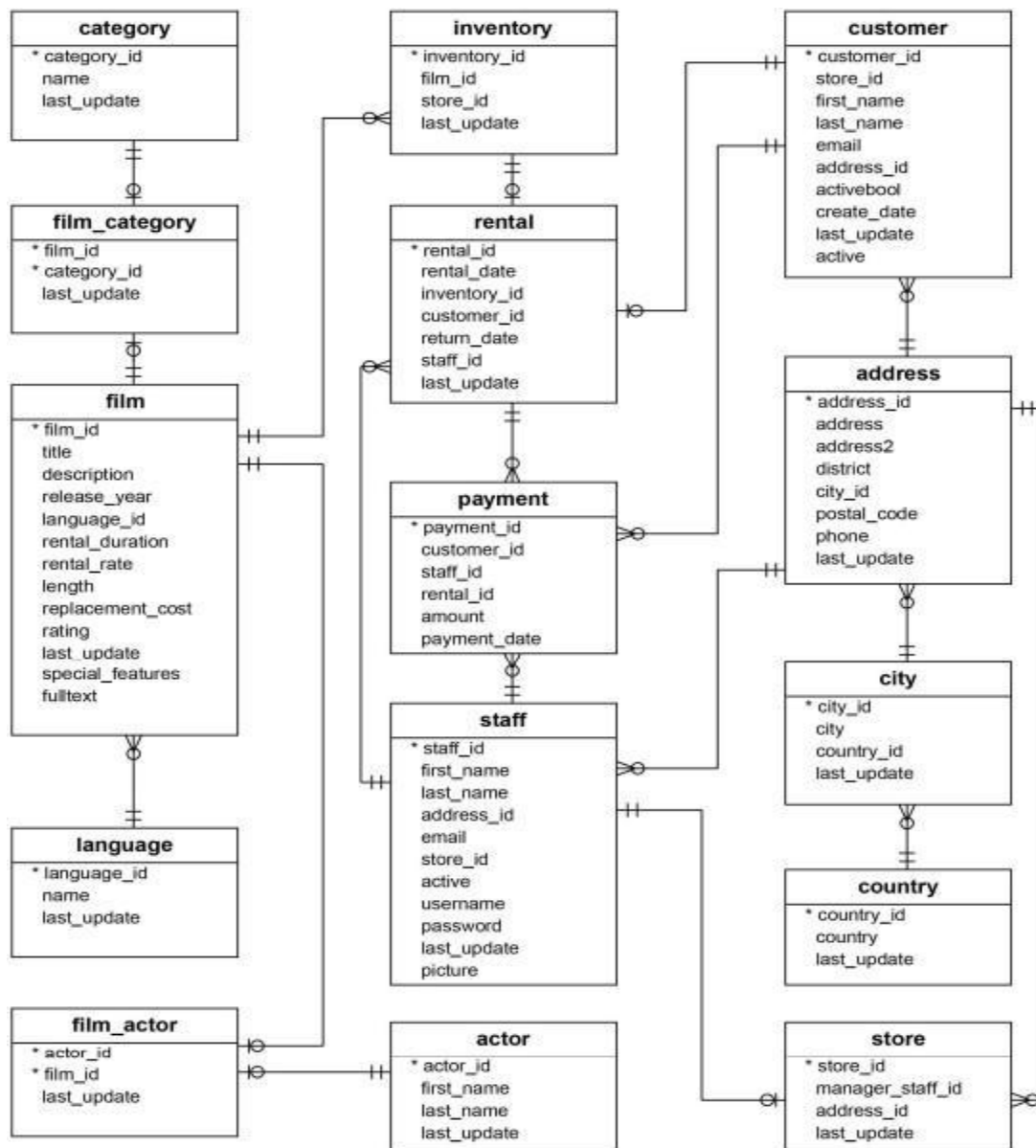
Se debe identificar también las fuentes de datos (orígenes) desde dónde se tomará la información para poblar el modelo dimensional (ie. tablas de dimensión y hechos).

Fuentes de datos disponibles: Para el ejercicio se utilizará la base de datos para alquiler de películas descrita en la página PostgreSQL Sample Database (<https://www.postgresqltutorial.com/postgresql-sample-database/>). Dicha base de datos consiste de 15 tablas para el control de préstamos de películas:

actor	almacena datos sobre actores y actrices
film	almacena datos de interés sobre diferentes películas
film_actor	almacena la asociación entre películas y actores
category	codifica los diferentes tipos de género de películas
film_category	almacena la asociación entre películas y categorías
store	contiene la diferente información sobre las tiendas
inventory	almacena los datos del inventario físico de discos con películas
rental	datos sobre los diferentes alquileres que se realizan. El alquiler de una película se realiza a un cliente en una fecha particular. El campo return_date es solo informativo, no debe ser tomado en cuenta para el cálculo del monto a pagar por el cliente.
payment	almacena los pagos de los diferentes clientes cuando retornan las películas alquiladas. El campo payment_date permite calcular el monto a cobrar por el alquiler de una película (ie. Sacando la diferencia entre la fecha de

	alquiler y la fecha de retorno real de la película). Vamos a asumir que no hay multas por entrega tardía de las películas.
staff	almacena los datos de los empleados
customer	almacena los datos de los clientes
address	almacena los datos de las direcciones de los clientes, empleados o tiendas
city	codifica el nombre de las ciudades
country	codifica el nombre de los países

Modelo de la base de datos



Con el fin de analizar la información acumulada de los alquileres se debe implementar un datawarehouse que permita responder las siguientes consultas:

La empresa espera que el sistema le permita hacer un rastreo del número de películas alquiladas y el total cobrado. Para calcular el total cobrado, la empresa utiliza el monto fijo de **500 colones por día de préstamo** de una película.

Algunas de las medidas de interés para el negocio son:

- ❑ El número de películas y el total cobrado por mes por cliente (consultan los datos por nombre del cliente).
- ❑ El número de películas y el total cobrado por mes por película (los informes deben incluir, en algunos casos, el idioma, el nombre de la película y los actores)
- ❑ El número de películas y el total cobrado por día por sucursal
- ❑ Comparativos de los totales cobrados anualmente.

Debe realizar lo siguiente utilizando la metodología presentada en el capítulo del libro estudiado [1]:

1. Identificar y documentar las dimensiones y sus respectivos atributos. (10 puntos)
2. Identificar las medidas por analizar y documentar su forma de cálculo (medidas a incorporar en la tabla de hechos) (10 puntos).
3. Identificar la granularidad del diseño dimensional (2 puntos)
4. Diseñar un esquema estrella que incluya las tablas identificadas en los puntos 1 y 2 (10 puntos)
5. Para cada atributo identificar la fuente de datos (tabla y atributo origen) (5 puntos)

Referencias

[1] Reeves, L (2009). A Manager Guide's to Data Warehousing (Capítulo 7). El capítulo está disponible en el TecDigital o el libro se puede bajar de:

<https://www.pdfdrive.com/a-managers-guide-to-data-warehousing-e23005825.html>

1. Identificar y documentar las dimensiones y sus respectivos atributos. (10 puntos)

Dimension Cliente

Attribute name	Definition	Sample Values
Customer identifier	Unique numeric identifier for a specific client	2, 5
Customer first name	Represents the first name of a customer.	Esteban, Bruce
Customer last name	Represents the last name of a customer	Vargas, Cruz

Dimensión Película

Attribute name	Definition	Sample Values
Film identifier	Unique numeric identifier for a specific movie	3, 8
Film title	Name of a particular movie	Matrix, Tiburón
Film language	Language in which the movie was originally filmed	Spanish, English
Movie Actor	An actor or actress that conformed part of the movie cast.	Tom Holland, Robert Downey Jr.

Dimensión Sucursal

Attribute name	Definition	Sample Values
Store identifier	Unique numeric identifier for a specific store	6, 9
Address identifier	Unique numeric identifier for a specific store	7, 14

Dimensión Fecha

Attribute name	Definition	Sample Values
Rental date	The date the movie is rented by a client	26/08/2022, 02/02/2022
Return date	The date the movie is returned to the store	30/08/2022, 07/02/2022
Payment date	Date of the payment of a rental	30/08/2022, 07/02/2022
Calendar Day	A calendar day used for reporting and tracking purposes	26/08/2022, 02/02/2022
Calendar Month	The calendar month used for reporting purposes	08/2022, 02/2022
Calendar Year	The calendar year used for reporting and comparison purposes	2021, 2022

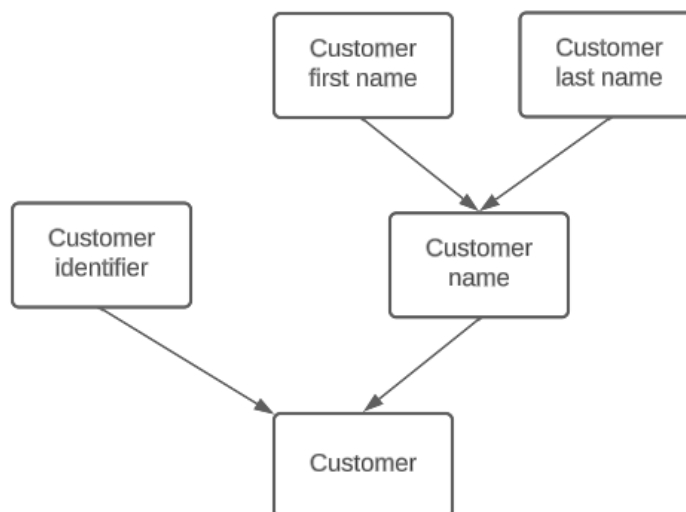
2. Identificar las medidas por analizar y documentar su forma de cálculo (medidas a incorporar en la tabla de hechos) (10 puntos).

Tabla de hechos - Alquileres

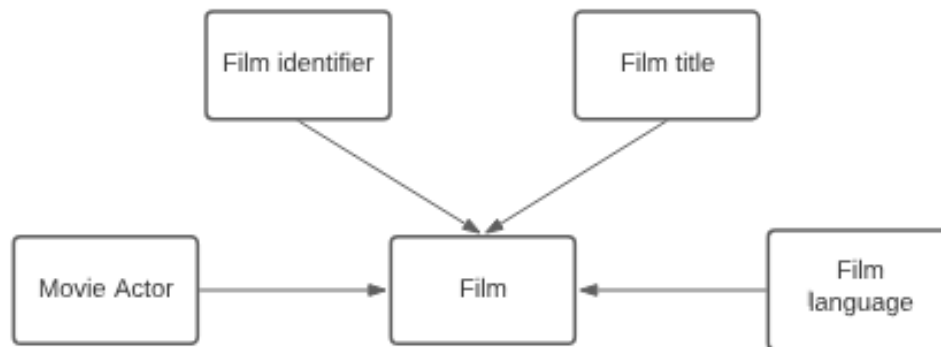
Fact name	Fact definition	Aggregation rule
Rental charge	Amount that the client has to pay for the rental of a film (500 colones per day)	Sum
Number of films per month	Number of films rented by a client, store or film per month	Count
Total charged per month	Total charged of a rental of a client, store or film per month	Sum
Number of films per day	Number of films rented by a client, store or film per day	Count
Total charged per day	Total charged of a rental of a client, store or film per day	Sum

3. Identificar la granularidad del diseño dimensional (2 puntos)

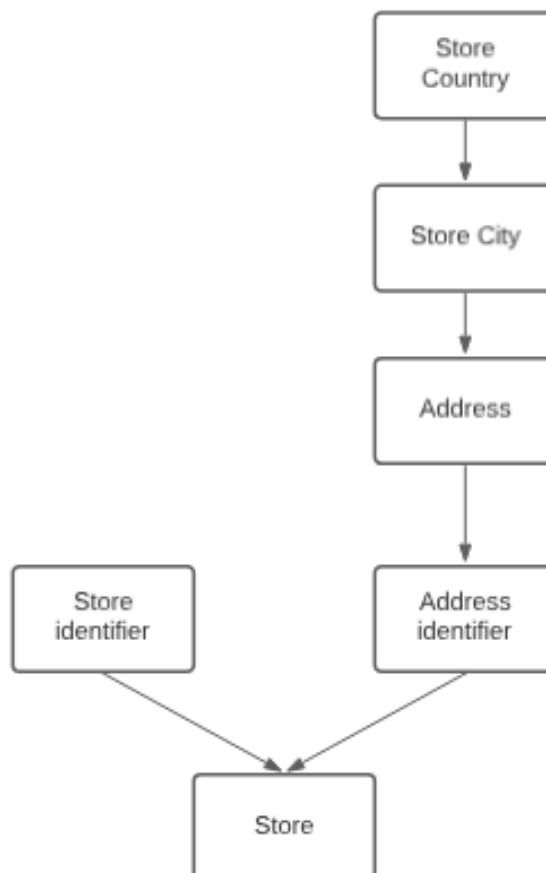
Granularidad de la dimensión cliente



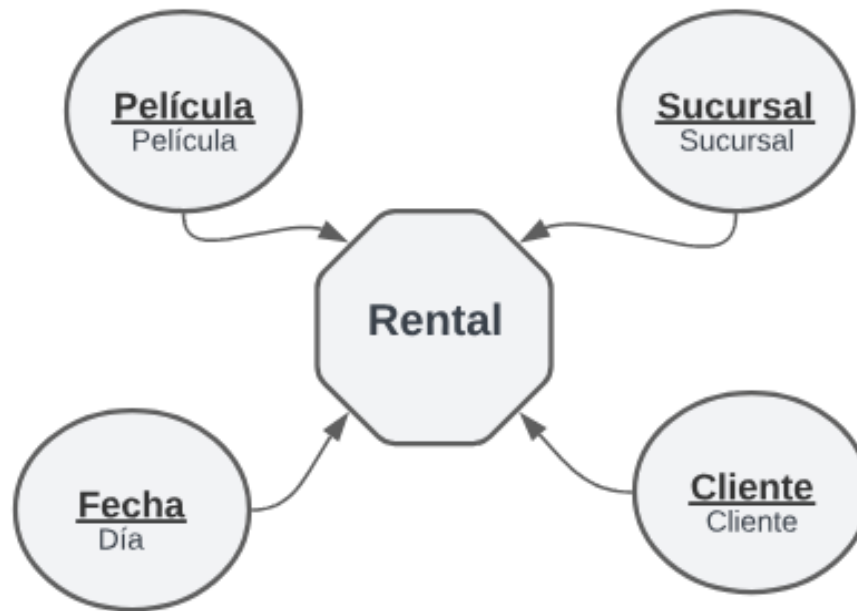
Granularidad de la dimensión película



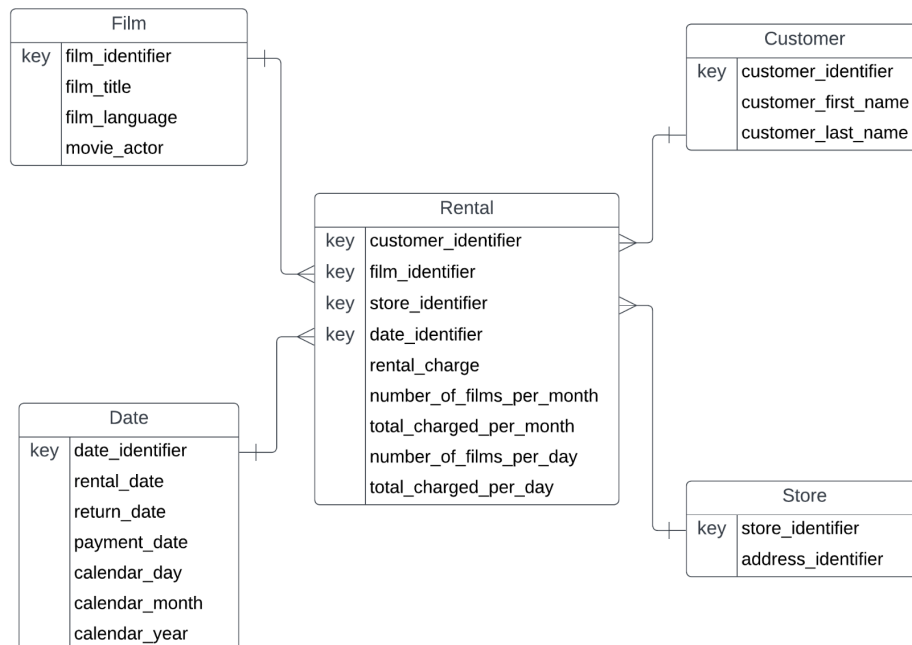
Granularidad de la dimensión sucursal



4. Diseñar un esquema estrella que incluya las tablas identificadas en los puntos 1 y 2 (10 puntos)



Esquema Estrella



5. Para cada atributo identificar la fuente de datos (tabla y atributo origen) (5 puntos)

Tabla de la dimension Cliente

Attribute name	Table	Origin of the attribute
customer_identifier	customer	customer_id
customer_first_name	customer	first_name
customer_last_name	customer	last_name

Tabla de la dimensión Película

Attribute name	Table	Origin of the attribute
film_identifier	film	film_id
film_title	film	title
film_language	language	name
movie_actor	film_actor	actor_id

Dimensión Sucursal

Attribute name	Table	Origin of the attribute
store_identifier	store	store_id
address_identifier	address	address

Dimensión Fecha

Attribute name	Table	Origin of the attribute
date_id	payment	payment_id
rental_date	rental	rental_date
return_date	rental	return_date
payment_date	payment	payment_date
calendar_day	rental	rental_date
calendar_month	rental	rental_date
calendar_year	rental	rental_date