

Arquitectura de Datos

PO2.1: DISEÑO EN MONGODB

Curso 2025/2026



Grupo de prácticas: 10

Nombre	NIA	Correo Electrónico	Grupo
Javier Rosales Lozano	100495802	100495802@alumnos.uc3m.es	81
Alonso Rios Guerra	100495821	100495821@alumnos.uc3m.es	81
Nahuel Sebastián Vargas	100495715	100495715@alumnos.uc3m.es	81
Alejandro Rodríguez García	100495745	100495745@alumnos.uc3m.es	81

Fecha de entrega: 13/10/2025

1. Introducción.

La siguiente memoria documenta el proceso de desarrollo de la segunda práctica correspondiente a la asignatura Arquitectura de Datos del curso 2025/26. La práctica se basa en un contexto de diseño de un servicio de suscripción de televisión bajo demanda (películas y series).

El propósito de la empresa **StreamIt S.A.** es explotar información de la aplicación que ofrece para obtener valor de datos y responder preguntas clave del negocio, como conocer los hábitos de consumo del cliente o detectar patrones culturales diferenciados de cada país por residencia, entre otros.

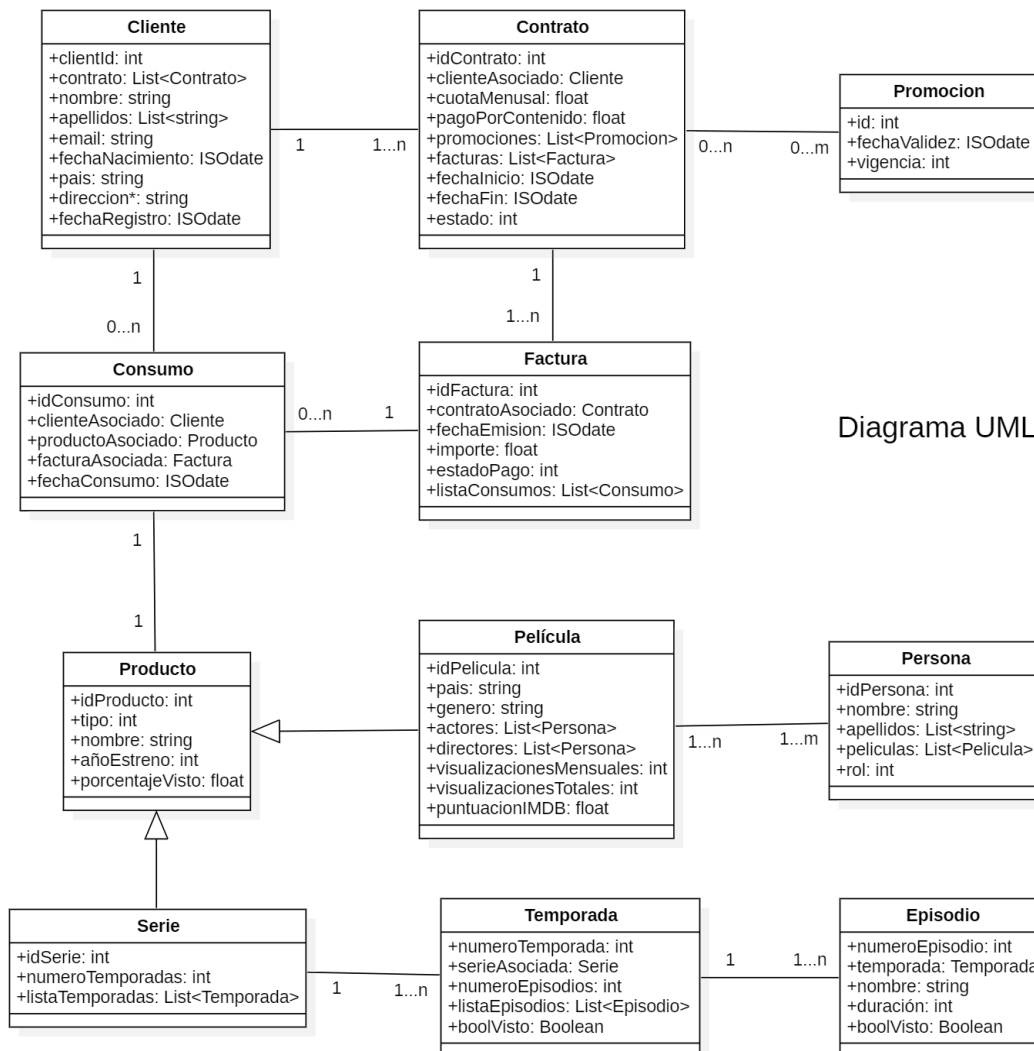
Los **casos de uso** que se tratan en la práctica son los siguientes:

1. **CU1: Historial de consumos de un cliente:** Un responsable de atención al cliente necesita consultar el historial completo de consumos de un cliente concreto (películas y series), con detalle de fechas, títulos y porcentaje visto. También se requiere poder acceder a los datos de contrato vigentes en el momento de cada consumo.
2. **CU2: Fidelización en series:** El área de marketing quiere identificar a los llamados "Serial Lovers", que son los clientes que han completado al menos una temporada completa de una serie. Para ello, es necesario poder contar los episodios vistos y compararlos con el número total de episodios de la temporada.
3. **CU3: Análisis por país de residencia:** El área de investigación sociocultural desea elaborar un informe de apellidos más comunes por país. Por tanto, se requiere mantener de forma consistente el país y los datos de cliente asociados, así como normalizar el tratamiento de apellidos (que pueden venir como string o array).
4. **CU4: Ranking de actores y películas:** El área de contenidos quiere obtener información sobre: o las parejas de actores más vistas en películas o el conjunto de "clásicos modernos", definido como películas del primer decil de antigüedad (el 10% de las más antiguas) que al mismo tiempo están en el top 5% de visualizaciones mensuales. Esto implica almacenar metadatos de películas y relacionarlos con los visionados de clientes.
5. **CU5: Resumen mensual de facturación:** El departamento financiero necesita un resumen mensual de facturación y consumos, por contrato y por cliente, de manera que puedan analizar la evolución de ingresos y correlacionarla con el uso de contenidos.

El ejercicio consiste en diseñar un **modelo de datos** en MongoDB que permita explotar eficazmente los casos de uso especificados en el enunciado para distintos departamentos, estructurando el modelo conceptual y lógico de clases, y justificando las estrategias de agregados y reglas de consistencia en la base de datos no estructurada.

2. Modelo UML

El siguiente diagrama es el resultado de analizar el contexto y los casos de uso según el diseño de clases y asociaciones:



2.1. Documentación de supuestos

- Cada **cliente** mantiene uno o varios contratos activos, y cada **contrato** está asociado a **un producto** con condiciones específicas (cuota mensual, pago por contenido y promociones activas). Cada **contrato** pertenece a un **único cliente**, pero **un cliente** puede tener **varios contratos** a lo largo del tiempo. **Un contrato** puede generar **varias facturas** mensuales, y cada factura está asociada a un único periodo de tiempo. Las facturas agrupan los consumos realizados por el cliente durante el mes correspondiente, tanto de películas como de series. Los documentos de factura y consumo son inmutables una vez generados, salvo corrección por duplicados. En caso de facturas duplicadas, se conserva la versión más reciente según la fecha de emisión. Las promociones pueden aplicarse a varios contratos, y un contrato puede beneficiarse de varias promociones durante su vigencia, la cual se controla mediante el atributo fechaValidez.

- **Película y Serie** son subclases de Producto, mediante una relación de herencia (generalización) en el modelo UML. Los consumos de películas se registran a nivel de producto (una película completa), mientras que los consumos se registran a nivel de episodio, aunque la relación en el modelo se mantenga con la clase Producto. Cada episodio pertenece a una única temporada, y cada temporada pertenece a una única serie.
- La **entidad Persona** se utiliza tanto para actores como para directores, diferenciados mediante el atributo rol, el cual puede tomar los valores actor, director o ambos, según la participación en una película. Una película puede estar asociada a varias personas, y una persona puede intervenir en múltiples películas.
- Cada cliente tiene un único país de residencia, almacenado en minúsculas y sin tildes. Los apellidos de los clientes se almacenan como un array de strings, normalizados y con capitalización correcta.
- Todas las **fechas** se representan en formato ISODate; los campos faltantes se almacenan con el valor NULL, no se eliminan; los nombres de campos siguen la notación camelCase; y los géneros, países e idiomas se guardan en minúsculas y sin tildes.

2.2. Entidades y atributos

Las clases y relaciones del diagrama UML permiten cubrir los principales casos de uso definidos para el servicio:

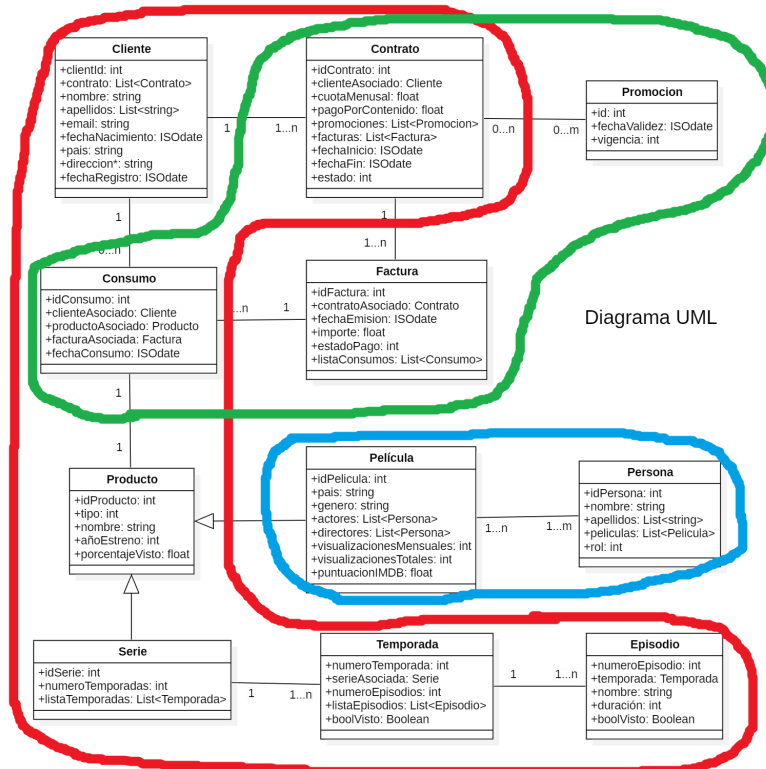
- **Cliente:** representa al usuario del servicio, almacenando datos personales, país, fecha de registro y todos sus contratos activos y pasados. Permite consultar el historial de consumos y facturación asociados a cada cliente.
- **Contrato:** modela la suscripción entre cliente y plataforma, registrando posibles promociones y todas las facturas generadas durante su vigencia.
- **Promoción:** define promociones aplicables a los contratos.
- **Factura:** registra los importes mensuales y el estado de pago, relacionándose con el contrato correspondiente y almacenando todos los consumos realizados en el periodo. Cubre el análisis de ingresos y uso mensual.
- **Consumo:** cada vez que un cliente visualiza un contenido, se representa como un consumo; se incluyen referencias al cliente, producto consumido y dónde se factura el consumo. Permite un seguimiento detallado por fecha y porcentaje visto.
- **Producto:** es una clase general que engloba cualquier contenido audiovisual diferenciando el tipo. Sirve como punto de unión para rankings y porcentaje visto.
- **Película:** especifica los metadatos de cada título cinematográfico, como año, país, género, reparto de personas (actores y directores), etc. Es esencial para los análisis de popularidad mensual y relaciones entre actores.
- **Persona:** unifica actores y directores, relacionando cada persona con todas las películas en las que participa y su rol.
- **Serie, Temporada y Episodio:** reflejan la jerarquía de los contenidos seriados, permitiendo verificar cuántos episodios y temporadas ha visto un cliente.

2.3. Formato de atributos

Los atributos de cada clase se han diseñado para soportar consultas y análisis requeridos. Lo **más destacable** de la arquitectura propuesta puede considerarse como el uso de enteros para clasificar estados de los atributos (por ejemplo, los roles de las personas: 0 “actor”, 1 “director”, 2 “ambos”) y el uso de arrays de instancias para las consultas.

3. Diseño de agregados (mongoDB)

A continuación, se despliega la justificación del diseño de agregados:



Agregado DatosCliente (Rojo) - Casos de uso CU1, CU2 y CU3	
Entidades	Cliente, Contrato, Consumo, Producto, Serie, Temporada, Episodio
Raíz	Cliente
Perímetro	Centraliza toda la información relevante de un cliente, incluyendo sus contratos activos e históricos, facturas, consumos audiovisuales, y la estructura jerárquica de serie, temporada y episodio. Optimiza consultas sobre historial de consumo, fidelización en series y análisis cultural por país y apellidos.
Entidades Embebidas	Contratos con Facturas y Consumos embebidos; estructura jerárquica de Serie, Temporada, Episodio embebida en cada Consumo para facilitar análisis y consultas rápidas.
Entidades Referenciadas	Producto, cuando se requiere información extendida; Serie, si se necesita información adicional no embebida.
Redundancias	Se duplican datos clave de producto (nombre, tipo, porcentaje visto) y episodios en los consumos para acelerar la consulta y evitar operaciones costosas de join.

Reglas de consistencia	Fechas en ISODate, apellidos normalizados como array, país en minúsculas y sin tildes, campos faltantes como NULL, nombres de campo en camelCase. Duplicados: en facturas repetidas, se conserva la versión más reciente según fechaEmision, integridad entre consumos y contrato vigente.
-------------------------------	--

Agregado EstadísticasReparto (Azul) - Caso de uso CU4

Entidades	Película, Persona
Raíz	Película
Perímetro	Gestiona los metadatos completos de cada película, incluyendo reparto, visualizaciones, puntuaciones y relaciones para rankings y análisis de popularidad. Optimiza consultas sobre rankings de actores, películas y "clásicos modernos".
Entidades Embebidas	Lista de personas en películas para facilitar consultas de reparto y rankings.
Entidades Referenciadas	Persona, para información extendida o biográfica.
Redundancias	Se almacenan visualizaciones mensuales/totales y nombres de personas en películas para acelerar rankings y evitar joins costosos.
Reglas de consistencia	Rol de persona como entero (actor, director, ambos), puntuaciones y visualizaciones validadas periódicamente, fechas en ISODate, país/género en minúsculas y sin tildes, campos faltantes como NULL.

Agregado Facturación (Verde) - Caso de uso CU5

Entidades	Factura, Contrato, Consumo, Promoción
Raíz	Factura
Perímetro	Abarca el resumen financiero mensual, asociando cada factura a su contrato, los consumos facturados y las promociones aplicadas. Optimiza el análisis de ingresos y correlación con el uso de contenidos.
Entidades Embebidas	Consumos facturados (con detalles clave) y Promociones aplicadas en la Factura para facilitar análisis financiero.
Entidades Referenciadas	Contrato, para información extendida; Promoción, si se prefiere evitar duplicidad; Producto, si se requiere información adicional de los consumos.
Redundancias	Importe total y estado de pago almacenados en factura; duplicación de datos clave de consumos para evitar cálculos repetidos.
Reglas de consistencia	Fechas en formato ISODate, facturas inmutables salvo corrección por duplicados, sincronización entre consumos y contratos, control de promociones vigentes por fecha, nombres en camelCase, campos faltantes como NULL, país/género en minúsculas y sin tildes.

4. Declaración de uso de IA

Para el desarrollo de esta práctica, se ha partido de la ayuda de IA generativa como soporte con el fin de resolver algunas dudas sobre conceptos teóricos y prácticos. También nos ha sido de gran ayuda para obtener feedback sobre nuestro diseño, lo que nos ha permitido mejorar iterativamente nuestro trabajo.

En general, el uso de la IA utilizada (Perplexity Pro) ha aportado feedback y ayuda en la toma de decisiones para algunos de los prompts relacionados con la revisión de cardinalidades y asociaciones del diagrama de clases, o la explicación de cómo realizar un modelado de agregados según éste.