

Instituto Politécnico Nacional
Escuela Superior de Cómputo

Trabajo Terminal No. 2019-B052

Herramienta para transformar un modelo
entidad-relación a un modelo no relacional

Presentan:

Aparicio Quiroz Omar
Martínez Acosta Eduardo

Directoras:

M. en C. Ocotitla Rojas Nancy
D. en C. Chavarria Baez Lorena

Índice general

1. Introducción	7
1.1. Descripción del problema	8
1.2. Propuesta de solución	8
1.3. Justificación	9
1.4. Objetivo	10
1.5. Alcance y limitaciones	10
2. Estado del Arte	12
2.1. Herramientas similares	12
2.1.1. Kashliev Data Modeler	12
2.1.2. NoSQL Workbench for Amazon DynamoDB	13
2.1.3. HacKolade	13
2.1.4. Mortadelo	14
2.1.5. NoSE: Schema Design for NoSQL Applications	21
2.1.6. Conclusiones	21
3. Marco Teórico	24
3.1. Modelos de datos para bases de datos	24
3.1.1. Modelo entidad-relación	24
3.1.1.1. Entidades	25
3.1.1.2. Atributos	26
3.1.1.3. Relaciones	27
3.1.1.4. Resumen de la notación para los diagramas ER	29
3.1.2. Modelo relacional	30
3.1.2.1. Relaciones	31
3.1.2.2. Claves	32
3.1.2.3. Restricciones de integridad	32

3.1.2.4.	Propiedades de las relaciones	33
3.1.2.5.	ACID	33
3.1.2.6.	Structured Query Language	34
3.1.3.	Modelos NoSQL	36
3.1.3.1.	Teorema CAP	36
3.1.3.2.	Clave-Valor	38
3.1.3.3.	Orientado a documentos	39
3.1.3.4.	Orientado a columnas	40
3.1.3.5.	Orientado a grafos	41
3.2.	Tecnologías a usar	42
3.2.1.	Hypertext Transfer Protocol	42
3.2.2.	HTML 5	44
3.2.3.	Style Sheet Language: CSS vs SASS	45
3.2.4.	Web Frameworks	47
3.2.5.	JavaScript Web Frameworks: Vue/Nuxt vs React vs Angular	49
3.2.6.	CSS Frameworks: Vuetify vs Bootstrap	51
3.2.7.	MongoDB vs Apache CouchDB	52
3.2.8.	Bibliotecas JavaScript para diagramado: GoJS vs Fabric.js vs D3.js	54
3.2.9.	Python 3	55
3.2.10.	Django vs Flask	57
3.3.	Conclusiones	59
4.	Análisis y Diseño del Sistema	60
4.1.	Metodología	60
4.1.1.	Scrum	61
4.1.1.1.	Historias de usuario	62
4.1.1.2.	Lista de producto (<i>product backlog</i>)	68
4.2.	Factibilidad técnica	71
4.3.	Factibilidad económica	71
4.3.1.	COCOMO II	72
4.3.1.1.	El modelo de composición de aplicación COCOMO II	72
4.4.	Factibilidad operativa	75
5.	Algoritmos	77

5.1. Validaciones entidad-relación	77
5.2. Modelo entidad-relación a relacional	77
5.3. Obtención de esquema SQL desde modelo relacional	80
5.4. Modelo entidad-relación a Generic Data Metamodel	80
5.5. Generic Data Metamodel a modelo lógico NoSQL	80
5.6. Modelo lógico NoSQL a modelo físico en MongoDB	80
A. Apéndice	81
A.1. Unified Modeling Language	81
A.2. El modelo entidad-relación extendido	84
A.3. Álgebra relacional	86

Índice de figuras

2.1. Mortadelo	14
2.2. Mortadelo	15
2.3. Notación textual del GDM	16
2.4. Modelo lógico orientado a documentos de Mortadelo	17
2.5. Access Tree - Modelo lógico orientado a documentos	18
2.6. Access query Q4	18
3.1. Tabla en el modelo Relacional	31
3.2. <i>Bucket</i> en el almacenamiento clave-valor	38
3.3. Documento en el almacenamiento de documentos	39
3.4. Familia de columnas	41
3.5. modelo conceptual orientado a grafos	41
A.1. Asociación	82
A.2. Agregación	83
A.3. Asociación rombo sin rellenar y composición rombo negro	83

Índice de tablas

3.1. Notación para los diagramas ER	29
3.2. Comandos de SQL.	34
4.1. Lista de producto	71

Resumen

Capítulo 1

Introducción

En los últimos años, los sistemas *Not only SQL* o NoSQL han surgido como alternativa a los sistemas de bases de datos relacionales y se enfocan, principalmente, en buscar resolver problemáticas inherentes al usarlas en algunos escenarios.

Los sistemas NoSQL admiten diferentes modelos de datos, como los de clave-valor, documentos, columnas y grafos; con ello se enfocan en guardar datos de una manera apropiada de acuerdo a los requisitos actuales para la gestión de datos en la web o en la nube, enfatizando la escalabilidad, la tolerancia a fallos y la disponibilidad a costa de la consistencia.

De acuerdo a Google Trends[1], en 2009 aumentó el interés en los sistemas NoSQL y desde entonces, también empezaron a resaltar algunas problemáticas propias de estos sistemas, porque si bien resuelven algunos dilemas de las bases de datos relacionales, por enfocarse en la ya mencionada escalabilidad, la tolerancia a fallos o la flexibilidad para realizar cambios en el esquema de la base de datos, la heterogeneidad de los sistemas NoSQL ha llevado a una amplia diversificación de las interfaces de almacenamiento de datos, provocando la pérdida de un paradigma de modelado común o de un lenguaje de consultas estándar como SQL.

Respecto al modelado de datos, en el diseño de las bases de datos tradicionales, el modelo entidad-relación[2] es el modelo conceptual más usado y tiene procedimientos bien establecidos como resultado de décadas de investigación; sin embargo, para los sistemas NoSQL los enfoques tradicionales de diseño de bases de datos no proporcionan un soporte adecuado para satisfacer el modelado de sus diferentes modelos de datos y para abordar esta problemática se han creado varias metodologías de diseño para sistemas NoSQL en los últimos años, porque los diseñadores de bases de datos deben tener en cuenta no solo qué datos se almacenarán en la base de datos, sino también cómo se accederán a ellos para modelar estos sistemas [3]-[5].

Para tener en cuenta cómo se accederán a los datos se debe conocer cómo se realizarán las consultas de los mismos y de acuerdo al trabajo de Mosquera[6], que es una investigación de 1376 *papers* sobre el modelado de sistemas NoSQL, se muestra que de las metodologías propuestas la mayoría usa el lenguaje de modelado UML (*Unified Modeling Language*), mientras que algunos proponen su propio modelo conceptual y en cada propuesta presentan una manera de representar las consultas de los datos; en resumen, en la literatura sobre el tema solo existen cinco herramientas propuestas para el modelado conceptual y, en general, no hay una tendencia en el

modelado de datos NoSQL.

Lo que resta del capítulo está organizado de la siguiente manera: primero se muestra la problemática a resolver, después la propuesta de solución, la justificación, el objetivo del proyecto y por último se menciona el alcance con las limitaciones del mismo.

1.1. Descripción del problema

Como se ha visto en la introducción, son pocas las herramientas propuestas en la literatura para el modelado de sistemas NoSQL en sus tres niveles de abstracción (nivel conceptual, lógico y físico).

Solo en el modelado conceptual, el modelo entidad-relación puede considerarse como una tendencia, sin embargo, el modelo entidad-relación por sí mismo no es suficiente para representar cómo se consultarán los datos ni tampoco con qué frecuencia se accederán a ellos; por eso, para modelar sistemas NoSQL es necesario conocer enfoques de desarrollo como el *query-driven design*, el *domain-driven design*, el *data-driven design* o el *workload-driven design*.

Ahora bien, respecto a los modelos de datos NoSQL, no hay un modelo estándar ni tampoco existe un acuerdo sobre la mejor definición de reglas de transformación entre modelos en los tres niveles de abstracción. Además, en la literatura sobre el tema se pueden encontrar unos 36 estudios que proponen diferentes enfoques para las transformaciones.

Asimismo, por los pocos años que han pasado desde que el interés sobre el tema aumentó y que son usados estos sistemas de bases de datos, aún no se ven reflejados en los programas de estudio o solo se da una introducción del tema a los estudiantes de licenciatura en ingeniería en sistemas o carreras afines.

En consecuencia, para el estudiante es un problema tener que enfrentarse a diseñar un sistema NoSQL, porque no se le enseña ninguna metodología de desarrollo para sistemas NoSQL, tampoco conoce los distintos modelos de datos NoSQL que hay, mucho menos sus ventajas o desventajas para poder elegir el modelo más apropiado para su aplicación y, como resultado, no pueden visualizar el diseño de una base de datos no relacional a partir de los conocimientos adquiridos en la asignatura de Bases de Datos.

Además, la escasez de herramientas CASE para el diagramado de esquemas no relacionales o herramientas que apoyen la migración de un modelo entidad-relación o relacional a uno no relacional aumenta la complejidad para que el estudiante haga uso de una base de datos no relacional.

1.2. Propuesta de solución

Desarrollar una aplicación web que permita:

- 1.º: la edición de un diagrama entidad-relación o diagrama ER.

-
- 2.º: realizar la validación del diagrama ER.
 - 3.º: transformar el diagrama ER a un diagrama del modelo relacional o diagrama Relacional.
 - 4.º: obtener el esquema de la base de datos en sentencias SQL del diagrama Relacional.
 - 5.º: obtener el esquema conceptual de un modelo de datos NoSQL orientado a documentos
 - 6.º: obtener el esquema de datos NoSQL en un NoSQL RDBMS (*NoSQL Database Management System*).

Respecto al primer punto, la edición del diagrama ER se implementará con alguna biblioteca para diagramado, en donde al usuario se le permitirá arrastar los elementos del diagrama ER desde una “paleta”(como una entidad, un atributo, etc.) a la zona para diagramar, conocida también como “canvas”.

Para el segundo punto, la validación se realizará con eventos de escucha en el “canvas” o zona de diagramado y será de acuerdo a reglas de validación que se expondrán más adelante.

Para el tercer punto, la transformación del diagrama ER al diagrama Relacional se realizará con alguno de los algoritmos bien conocidos en la literatura del tema y así como en el segundo punto, también se expondrá más adelante el algoritmo a usar.

Asimismo, para el cuarto punto se generarán las sentencias SQL desde el diagrama Relacional.

Para el quinto punto, dado que hay varias propuestas en la literatura sobre modelos conceptuales NoSQL, se optará por usar la propuesta de Alfonso de la Vega[7], ya que el modelo que propone, conocido como Generic Data Metamodel, es un modelo conceptual que sirve para describir una base de datos orientada a documentos, a columnas, de clave-valor y orientada a grafos. Además, es una notación que está probada con varios casos de estudio.

Finalmente, para el sexto punto se generará el esquema de una base de datos NoSQL desde el Generic Data Metamodel del punto anterior.

1.3. Justificación

Las pocas herramientas para modelado conceptual NoSQL no ofrecen un modelado desde un modelo conceptual relacional, porque se enfocan en un modelo conceptual específico para cada tipo sistemas NoSQL (clave-valor, orientado a documentos, columnas o grafos). Tampoco ofrecen validaciones para sus diagramas ni dan la posibilidad de obtener el esquema SQL, ya que no están enfocados en sistemas relacionales.

Asimismo, por los pocos años que han pasado desde que el interés y el uso de los sistemas NoSQL aumentó, estos sistemas todavía no están reflejados en los programas de estudio.

En consecuencia, para abordar esta problemática en la que el estudiante pueda diseñar un sistema NoSQL sin haber aprendido ninguna metodología de desarrollo para dichos sistemas y sin poder visualizar en primera instancia el diseño de una base de datos no relacional a partir de lo que ve en la asignatura de Bases de Datos, se propone una herramienta CASE que pueda apoyar al estudiante a diseñar una base de datos NoSQL a partir de los conocimientos adquiridos del estudiante.

Se usará como modelo conceptual el modelo entidad-relación por ser un modelo bien conocido para el estudiante de licenciatura en sistemas, que está probado en el área de las bases de datos y es también el más usado en las investigaciones sobre el modelado conceptual de sistemas NoSQL.

Respecto al modelo físico que obtenga la herramienta a desarrollar, se probará en el esquema orientado a documentos en una base de datos NoSQL orientada a documentos.

1.4. Objetivo

Desarrollar una aplicación web que permita la edición de un diagrama de bases de datos bajo el modelo entidad-relación y realice la validación del mismo.

El diagrama podrá ser transformado al modelo relacional con la posibilidad de obtener el esquema de la base de datos en sentencias SQL, o bien, obtendrá el esquema de base de datos en un modelo de datos no relacional y tendrá la posibilidad de implementar el modelo no relacional en un NoSQL RDBMS (*NoSQL Database Management System*).

1.5. Alcance y limitaciones

1. Se usará la notación oficial en los modelos de datos conceptuales entidad-relación y relacional, es decir, la notación de Chen y la notación de Codd, respectivamente.
2. Para la generación de *scripts* de bases de datos relacionales se usará el lenguaje de consultas SQL.
3. Asimismo, los *scripts* generados se probarán con MySQL, ya que es el RDBMS (*Database Management System*) que se usa en la asignatura de Base de Datos y es con el que están más familiarizados los alumnos.
4. Para la transformación del modelo de datos conceptual del modelo ER al modelo relacional no se realizará ninguna normalización, porque se necesita un diagrama no normalizado para la transformación al modelo conceptual de datos NoSQL.

-
5. Los *scripts* NoSQL generados por la aplicación solo serán probados y ejecutados en un único NoSQL RDBMS.
 6. Respecto al registro de los usuarios, se podrá acceder a la aplicación con un correo válido o uno que no lo sea, teniendo en cuenta de que si no crea su cuenta con un correo válido, no le llegará notificación de correo que le indica que se ha registrado exitosamente junto con su contraseña.
 7. El diagramador ER solo podrá editar elementos del modelo entidad-relación, no se podrá diagramar elementos del modelo entidad-relación extendido ni de ningún otro modelo.
 8. No se podrá editar el diagrama generado para el modelo relacional.
 9. Tampoco se podrá editar el diagrama conceptual del modelo NoSQL orientado a documentos.
 10. El modelo elegido para desarrollar la propuesta de modelo conceptual será independiente del modelo de datos NoSQL o enfocado en el modelo NoSQL orientado a documentos.
 11. Asimismo, aunque la propuesta de modelo conceptual sea independiente del modelo de datos NoSQL, la generación de *scripts* será únicamente en el modelo de datos orientado a documentos.
 12. El grado de partición en el modelo entidad-relación será máximo de dos entidades.

Capítulo 2

Estado del Arte

Para la elaboración de la presente propuesta de Trabajo Terminal, ha sido necesaria una investigación de herramientas similares en el ámbito académico y empresarial.

Lo que resta de este capítulo está organizado de la siguiente manera: se presenta cada oferta con una breve descripción y se finaliza con una tabla comparativa.

2.1. Herramientas similares

2.1.1. Kashliev Data Modeler

De acuerdo al sitio web de KDM[8], Kashliev Data Modeler es una herramienta de modelado desarrollada a partir del 2015 por Andrii Kashliev, profesor asistente del Departamento de informática de la Universidad del Oeste Michigan, que automatiza el diseño de esquemas para Apache Cassandra, una base de datos NoSQL orientada a columnas.

Por medio de una aplicación web, KDM ayuda al usuario en el modelado de datos, comenzando con un modelo conceptual de datos (de una notación familiar al modelo entidad-relación) y generando un modelo físico de datos o *script* CQL (Cassandra Query Language).

KDM automatiza:

1. El mapeo conceptual a lógico
2. El mapeo lógico a físico
3. La generación de *script* CQL

El modelo de datos usado en KDM es el propuesto por Chebotko[4] y por medio de un algoritmo, KDM automatiza el proceso de pasar del modelo de datos a un modelo de datos físicos.

Además, KDM cuenta con una versión de prueba de tiempo indefinido con características limitadas que permite la generación de un modelo lógico y guardar los proyectos de KDM.

2.1.2. NoSQL Workbench for Amazon DynamoDB

De acuerdo al sitio web de Amazon[9], NoSQL Workbench for Amazon DynamoDB es una herramienta desarrollada por Amazon que proporciona funciones de modelado de datos, visualización de datos y desarrollo de consultas para ayudar a diseñar, crear, consultar y administrar bases de datos NoSQL orientadas a columnas.

NoSQL Workbench for Amazon DynamoDB usa el modelo conceptual propuesto por Chebotko[4].

También se pueden crear nuevos modelos de datos o diseñar modelos basados en modelos de datos existentes que satisfagan los requerimientos de acceso a datos de su aplicación, así como importar y exportar el modelo de datos diseñado al final del proceso.

El visualizador de modelo de datos proporciona un lienzo donde puede asignar consultas y visualizar las facetas (parte de la base de datos) de la aplicación sin tener que escribir código.

Cada faceta corresponde a un patrón de acceso diferente en DynamoDB; puede agregar datos manualmente a su modelo de datos o importar datos desde MySQL.

Cuenta con un generador de operaciones para ver, explorar y consultar conjuntos de datos.

Por último, admite la proyección, la declaración de expresiones condicionales y permite generar código de muestra en varios idiomas.

2.1.3. HacKolade

De acuerdo al sitio web de Hackolade[10], es un *software* con capacidad de representar objetos JSON anidados; la aplicación combina la representación gráfica de colecciones (término usado en las bases de datos orientados a documentos) y vistas en un diagrama.

La aplicación está basada en la desnormalización, el polimorfismo y matrices anidadas JSON; la representación gráfica de la definición del esquema JSON de cada colección está en una vista de árbol jerárquica.

Hackolade genera dinámicamente *scripts* MongoDB a medida que construye un modelo de datos, deriva modelos de datos por medio de ingeniería inversa de instancias MongoDB, dando facilidad para obtener descripciones, propiedades y restricciones.

Lamentablemente, en el sitio web de HacKolade no hay información sobre qué modelo conceptual está basado su aplicación, así que se asume que usan una propuesta propia para el modelo conceptual del que hacen uso.

Es una herramienta de modelado de datos para MongoDB, Neo4j, Cassandra,

Couchbase, Cosmos DB, DynamoDB, Elasticsearch, HBase, Hive, Google BigQuery, Firebase/Firestore, MarkLogic, entre otros.

2.1.4. Mortadelo

De acuerdo con de la Vega[7], Mortadelo está basado en el *model driven*; es decir, que para diseñar bases de datos NoSQL necesita de modelos conceptuales bien definidos.

Como menciona este autor, un modelo conceptual que especifique solo qué entidades conforman el sistema no es suficiente, porque para los sistemas NoSQL, es clave conocer cómo estas entidades serán consultadas y actualizadas en el día a día de la implementación de la base de datos.

Asimismo, en trabajos como los de Chebotko[4], de Lima[11] o Mior[12] los modelos conceptuales tradicionales como el modelo entidad-relación o el modelo relacional son complementados con propuestas para los patrones de acceso a los datos.

Lo anterior implica usar dos modelos diferentes pero interrelacionados que añada complejidad al modelado de la base datos; por esta razón de la Vega propone el Generic Data Metamodel (GDM), que es un modelo conceptual NoSQL donde la estructura (entidades, atributos, relaciones entre entidades) y patrones de acceso (cómo se consultarán los datos) están integradas en un mismo modelo conceptual.

Cabe destacar que el GDM es un modelo conceptual independiente del paradigma NoSQL; es decir, puede representar una base de datos NoSQL de clave-valor, orientado a documentos, orientado a columnas u orientado a grafos.

Tal como está en la figura 2.1, Mortadelo empieza con un GDM, realiza una transformación “modelo a modelo”(M2M) para generar el modelo lógico NoSQL y realiza una transformación “modelo a texto”(M2T) para generar el modelo físico correspondiente.

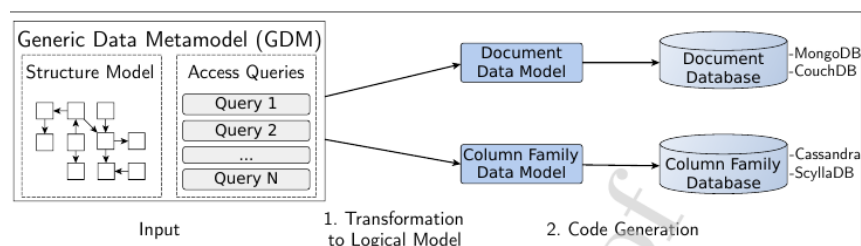


Figura 2.1: Mortadelo

Mortadelo: modelo conceptual (Generic Data Metamodel)

Como se puede ver en la figura 2.2, el modelo conceptual que propone de la Vega contiene en un mismo modelo con notación UML los elementos de la estructura (*structure model elements*) y el cómo se realizarán las consultas (*access queries elements*).

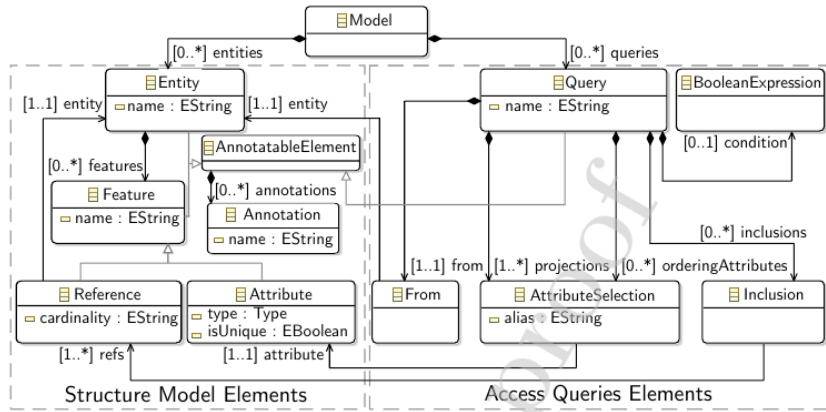


Figura 2.2: Mortadelo

A continuación se da una explicación de cada elemento de la figura 2.2 donde la clase *Model* es para indicar que un modelo GDM puede tener n entidades y n consultas donde $n = 0, 1, \dots, n$:

Structure model elements

- Clase *Entity*: puede tener *features* y solo puede ser referenciada directamente solo desde las clases *From* y *Reference*.
- Clase *Feature*: es una clase abstracta, puede ser una *reference* o un *attribute*; asimismo, puede heredar de la clase *AnnotatableElement* para que la instancia de la clase pueda ser comentada con indicadores de texto que proveen sobre información extra para generar el modelo lógico.
- Clase *Reference*: empieza con la palabra clave “ref”, un nombre de tipo de entidad, una cardinalidad y un nombre de referencia. Por ejemplo: “ref Category[*] categories” define una referencia llamada *categories*, del tipo de entidad *Category* con una cardinalidad de cero a varios.
- Clase *Attribute*: contiene un tipo y un identificador de unicidad.
- Clase *AnnotatableElement*: clase abstracta para permitir que una clase contenga anotaciones.
- Clase *Annotation*: es un indicador de texto que proveen sobre información extra para generar el modelo lógico.

Access queries elements

- Clase *Query*: puede tener solo un elemento de la clase *From*, n elementos de la clase *AttributeSelection*, n elementos de la clase *Inclusion* y tener o no un único elemento de la clase *BooleanExpression*.
- Clase *From*: es la que asocia una clase *Query* con la clase *Entity*; es la clase que permite referenciar un tipo de entidad.

- Clase *AttributeSelection*: es con la que se acceden a los atributos de la *Entity* referenciada por la clase *From* o la clase *Inclusion*.
- Clase *BooleanExpression*: se puede expresar una expresión booleana para declarar alguna restricción.
- Clase *Inclusion*: se puede acceder en una query a los atributos de otras tipos de entidad.

```
// Entities
entity Product {
  id productId
  text name
  text description
  number price
  ref Category[*]
    categories
  ref Provider[1]
    provider
}

entity Category {
  id categoryId
  text name
  text description
}

entity Provider {
  id providerId
  text name
}

// Queries
query Q1_productsById:
  select prod.productId,
    prod.name, prod.price,
    prod.description
  from Product as prod
  where prod.productId = "?"

query Q2_productsAndCategoryByName:
  select prod.name, prod.price,
    prod.description,
    cat.name
  from Product as prod
  including prod.categories as cat
  where prod.name = "?"

query Q3_productsByCategory:
  select prod.name, prod.price,
    prod.description,
    cat.name
  from Product as prod
  including prod.categories as cat
  where cat.name = "?"
  order by prod.price
```

Figura 2.3: Notación textual del GDM

En la figura 2.3 se puede ver una instancia del modelo conceptual GDM en su notación textual en la que se puede notar que, por ejemplo, en la tercera consulta se accede a los elementos de la entidad *Category* a través del elemento ref de la entidad *Product*.

Mortadelo: modelo lógico orientado a documentos

Las bases de datos orientadas a documentos tienen como objetivo almacenar jerarquías de objetos que son probables que se consultarán juntas; estas jerarquías de objetos se conocen como documentos y se agrupan en colecciones; asimismo, una colección almacena documentos de la misma entidad y por lo general un documento está compuesto de pares clave-valor y puede contener otros documentos incrustados.

En la figura 2.4 se muestra el modelo lógico orientado a documentos de la herramienta Mortadelo y a continuación se explica cada clase que conforma el modelo.

- Clase *DocumentDataModel*: está compuesta de n *Collections*.
- Clase *Collection*: tiene un nombre que la identifica y en una colección se almacenan documentos que, en general, comparten la misma estructura.

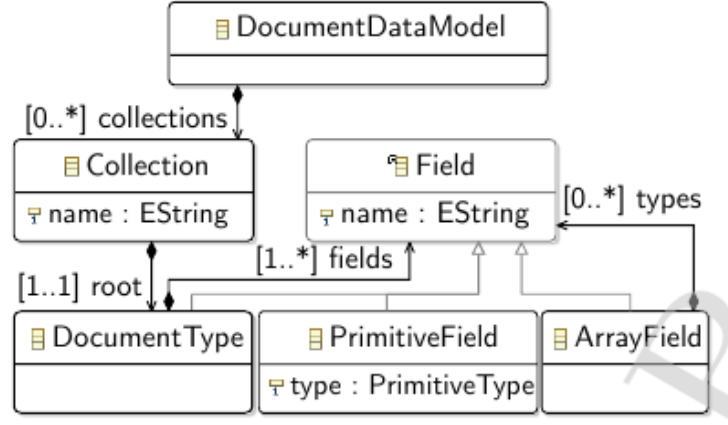


Figura 2.4: Modelo lógico orientado a documentos de Mortadelo

- Clase *DocumentType*: define la estructura de los documentos a guardar con un conjunto de instancias de la clase *Field*.
- Clase *Field*: puede contener instancias de las clases *PrimitiveField* (atributos simples), *ArrayField* (más instancias de la clase *Field*) o *DocumentType* (otras estructuras de documentos), permitiendo guardar elementos anidados.
- Clase *PrimitiveField*: contiene un tipo primitivo de dato
- Clase *ArrayField*: permite anidar más instancias de la clase *Field*

Transformación M2M Un *DocumentType* es la estructura principal que se guardará en la colección.

Sea Q el conjunto de todas las queries q_n en una instancia GDM:

$$Q = \{q_1, q_2, \dots, q_n\} \text{ donde } n = 1, 2, \dots$$

Sea E_f el conjunto de entidades formado al consultar el único elemento from de cada q_n ($q_{n.from}$ de Q):

$$E_f = \{e_{f1}, e_{f2}, \dots, e_{fn}\} \text{ donde } n = 1, 2, \dots$$

Sea C el conjunto de colecciones c_n donde cada c_n corresponde a un e_{fn} de E_f (es decir que por cada e_{fn} se crea un c_n en la que cada c_n contiene un único elemento *DocumentType*, llamado “documento raíz”):

$$C = \{c_1, c_2, \dots, c_n\} \text{ donde } n = 1, 2, \dots$$

Finalmente, por cada c_n se generan los contenidos de cada documento raíz:

Como cada colección c_n es creada a partir de una única entidad e_{fn} , se necesita buscar del conjunto Q todas las q_n que contengan esa e_{fn} como entidad de búsqueda principal (es decir, que e_{fn} esté en el elemento *from* de la q_n).

Dicho de otra manera, sea el conjunto Q_E el subconjunto propio de Q en el que cada q_{en} contiene en su elemento *from* la e_{fn} .

$$Q_E = \{q_{e1}, q_{e2}, \dots, q_{en}\} \text{ donde } Q_E \subseteq Q | q_{en}.from == e_{fn}$$

Por cada q_{en} se consulta todos sus elementos *including*, a cada elemento *including* le llamamos r_n para formar el conjunto R de referencias.

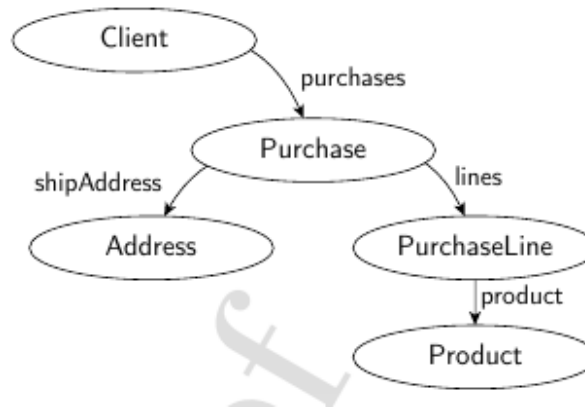


Figura 2.5: Access Tree - Modelo lógico orientado a documentos

```

query Q4_clientPurchasesNearChristmas:
select client.clientId, client.name,
       client.nationality,
       purchases.purchaseId, purchases.year,
       purchases.month, purchases.day,
       lines.quantity, lines.unitPrice,
       product.name, address.postalCode
from Client as client
including client.purchases as purchases,
         client.purchases.lines as lines,
         client.purchases.lines.product as product,
         client.purchases.shipAddress as address
where client.clientId = "71" and purchases.year = "72"
and purchases.month >= 10
order by purchases.month, product.price
  
```

Figura 2.6: Access query Q4

Creamos un árbol de acceso con el conjunto R y con ese árbol de acceso generamos los documentos de cada documento raíz de cada c_n que tiene una única e_{fn} .

La figura 2.5 es un ejemplo del árbol de acceso de la consulta q_4 de la figura 2.6.

Añadimos todos los atributos simples de la e_{fn} .

Cada r_n lo añadimos al documento raíz de acuerdo a 1) la cardinalidad de la referencia y 2) la entidad objetivo de la referencia

Si la entidad destino está dentro del árbol de consultas implica que datos de esa entidad serán requeridos en la consulta, por lo que el nuevo *field* deberá ser incluido como un subdocumento o una colección de documentos, dependiendo de su cardinalidad.

En contraparte, si la entidad destino no está en el árbol de consultas, se podría quitar esa referencia, pero se agregará para mejorar la escalabilidad de la base de datos para futuras consultas. Esta referencia tendría el valor del identificador de la entidad referenciada cuando la cardinalidad es 1. Si la cardinalidad es n, será un arreglo de identificadores.

Cuando una referencia es transformada a un subdocumento se vuelve a generar su árbol de acceso en una llamada recursiva para generar sus contenidos.

Por último, el autor menciona dos optimizaciones si se quiere reducir el nivel de denormalización, las que se pueden consultar en su investigación[7].

En resumen, el algoritmo quedaría de la forma:

Algorithm 1: Transformación del modelo conceptual GDM al modelo lógico orientado a documentos

Input : una instancia del modelo GDM, *gdm*
Output: un modelo lógico orientado a documentos, *ddm*

```

1 mainEntities  $\leftarrow$  gdm.queries.collect((q)|q.from);
2 foreach me  $\in$  mainEntities do
3   | collection  $\leftarrow$  newCollection();
4   | collection.name  $\leftarrow$  me.name;
5   | accessTree  $\leftarrow$  allQueryPaths(me, gdm.queries);
6   | collection  $\leftarrow$  populateDocumentType(collection.root, accessTree);
7   | ddm.collections.add(collection);
8 end
```

Donde la función *populateDocumentType()* es otro algoritmo de la forma:

Para la comprobación de su validez, se ha implementado una herramienta que admite la generación de almacenes de datos de columnas con transformaciones concretas para Cassandra, así mismo ofrece soporte preliminar de los basados en documentos.

Algorithm 2: Generar el contenido de un *DocumentType* dado un árbol de acceso

Input : Un “document type”, *dt*
Output: un nodo del arbol de acceso

```
1 nodeAttributes  $\leftarrow$  node.entity.features.select(f|f.isTypeOf(Attribute))
2 nodeReferences  $\leftarrow$ 
   node.entity.features.select(f|f.isTypeOf(Reference))
3 foreach attr  $\in$  nodeAttributes do
4   | pf  $\leftarrow$  newPrimitiveField()
5   | pf.name  $\leftarrow$  attr.name
6   | pf.type  $\leftarrow$  attr.type
7   | dt.fields.add(pf)
8 end
9 foreach ref  $\in$  nodeReferences do
10  | targetNode  $\leftarrow$  node.arcs.find(a|a.name = ref.name).target
11  | if exists(targetNode) then
12  |   | baseType  $\leftarrow$  newDocumentType()
13  |   | populateDocumentType(baseType, targetNode)
14  | else
15  |   | baseType  $\leftarrow$  newPrimitiveField()
16  |   | baseType.type  $\leftarrow$  findIdType(ref.entity)
17  | end
18  | baseType.name  $\leftarrow$  ref.name
19  | if ref.cardinality == 1 then
20  |   | dt.field.add(baseType)
21  | else
22  |   | arrayField  $\leftarrow$  newArrayField()
23  |   | arrayField.type  $\leftarrow$  baseType
24  |   | dt.fields.add(arrayField)
25  | end
26 end
```

2.1.5. NoSE: Schema Design for NoSQL Applications

De acuerdo con Michael J. Mior[5], NoSE usa un modelo conceptual de los datos y una descripción de cómo se accederán a ellos, que es conocido como *workload* y es usado para generar un modelo físico de una base de datos NoSQL orientado a columnas.

Para el esquema generado, NoSE debe tener un modelo conceptual que describa la información que se almacenará, por ello NoSE espera este modelo conceptual en forma de un grafo de entidad.

Los grafos de entidad son un tipo restringido del modelo de entidad-relación. Cada cuadro representa un tipo de entidad, cada borde es una relación entre entidades y la cardinalidad asociada de la relación (uno a muchos, uno a uno o muchos a muchos). Las entidades tienen atributos, uno o más de los cuales sirven como clave.

Respecto al *workload*, se describe como un conjunto de consultas parametrizadas y declaraciones de “actualización”. Cada consulta y actualización está asociada con un peso que indica su frecuencia relativa en la carga de trabajo.

2.1.6. Conclusiones

Herramienta	Creador	Objetivo	Licencia	Sistema operativo	Publicación	Modelo			Modelado		
						Conceptual	Lógico	Físico	Transformación entre niveles de abstracción	Lenguaje	Metodología
KDM	Andrii Kashliev.	Comercial	Software privativo	Web based GUI	2018	E-R	Columnas	Cassandra DB	Reglas de mapeo. Flujo de trabajo de la aplicación. Patrones de mapeo reglas de mapeo proceso de normalización y desnormalización	Diagramas query-driven a nivel logico	
Hackolade	IntegrIT SA / NV,	Comercial	Software privativo	Windows, Mac, Linux	2016	E-R	Multi-paradigma	Multi-plataforma		No definido	model-driven

Continuación de Tabla 3.2

Continuación de tabla	Amazon	En desarrollo	Software privativo	Windows, Mac	2019	E-R	llave-valor	Dynamo-DB	Reglas de mapeo	No definido	query-driven
NoSQL Workbench for Amazon DynamoDB	Alfonso de la Vega, Diego García-Saiz, Carlos Blanco, Marta Zorrilla, and Pablo Sánchez	Investigación	implementación de software bajo licencia libre	-	2019	GDM	Columnas, documento	Cassandra, MongoDB	Reglas de mapeo	UML	model-driven
NoSE (Schema Design for NoSQL Applications)	Michael J. Mior, Kenneth Salem, Ashraf Aboulnaga and Rui Liu	Investigación	implementación de software bajo licencia libre	-	2017	E-R	Columnas	Cassandra	Carga de trabajo. Limitación de espacio	N/A	query-driven
Fin de Tabla											

Capítulo 3

Marco Teórico

El capítulo está organizado de la siguiente manera: primero se muestra cada tema que el lector debe conocer para entender el desarrollo de la propuesta de solución, después se mostrará cada tecnología a usar y finalmente se tendrá un apartado de conclusión en que estará el porqué se ha elegido cada tecnología.

3.1. Modelos de datos para bases de datos

De acuerdo a la bibliografía de Elmasri [13], un modelo de datos es una colección de conceptos que describen una estructura de una base de datos.

Los modelos de datos de alto nivel o conceptuales ofrecen conceptos visuales simples que representan un modelo de datos, mientras que los modelos de datos de bajo nivel o físicos ofrecen conceptos que describen los detalles de cómo se implementa el almacenamiento de los datos en el sistema de la base de datos.

Los modelos de datos conceptuales utilizan conceptos como entidades, atributos y relaciones o, en el caso de ser bases de datos no relacionales, no hay un estándar definido para modelar conceptualmente este tipo de bases de datos.

Una entidad representa un objeto o concepto del mundo real, un atributo representa alguna propiedad que describe a una entidad y una relación es una asociación entre entidades.

A continuación se presentan varios de los modelos de datos existentes.

3.1.1. Modelo entidad-relación

De acuerdo a Elmasri [13], el modelo entidad-relación, que fue creado y formalizado por Peter Chen en 1976[14], se utiliza con frecuencia para el diseño conceptual de las aplicaciones de base de datos.

En esta sección se describen los conceptos básicos y las restricciones del modelo ER.

3.1.1.1. Entidades

El objeto básico representado por el modelo ER es una entidad, que es una cosa del mundo real con una existencia independiente.

Una entidad puede ser un objeto con una existencia física (por ejemplo, una persona en particular, un coche, una casa o un empleado) o puede ser un objeto con una existencia conceptual (por ejemplo, una empresa, un trabajo o un curso universitario).

Tipo de entidad

Un tipo de entidad define una colección (o conjunto) de entidades que tienen los mismos atributos.

La colección de todas las entidades de un tipo de entidad en particular de la base de datos se denomina conjunto de entidades; se usa el mismo nombre del tipo de entidad para hacer referencia al conjunto de entidades.

Tipos de entidades débiles

Los tipos de entidad que no tienen atributos clave propios se denominan tipos de entidad débiles. En contraposición, los tipos de entidad regulares que tienen un atributo clave se denominan tipos de entidad fuertes.

Las entidades que pertenecen a un tipo de entidad débil, se identifican como relacionadas con entidades específicas de otro tipo de entidad (llamada entidad identificado o propietario) en combinación con uno de sus valores de atributo.

Este tipo de relación que relaciona un tipo de entidad débil con su propietario lo podemos llamar relación identificativa del tipo de entidad débil.

Un tipo de entidad débil siempre tiene una restricción de participación total (dependencia de existencia) respecto a su relación identificativa, porque una entidad débil no puede identificarse sin una entidad propietaria. No obstante, no toda dependencia de existencia produce un tipo de entidad débil.

Un tipo de entidad débil normalmente tiene una clave parcial, que es el conjunto de atributos que pueden identificar sin lugar a dudas las entidades débiles que están relacionadas con la misma entidad propietaria.

En los diagramas ER, tanto el tipo de la entidad débil como la relación identificativa, se distinguen rodeando sus cuadros y rombos mediante unas líneas dobles.

El atributo de clave parcial aparece subrayado con una línea discontinua o punteada. Los tipos de entidades débiles se puede representar a veces como atributos complejos (compuestos, multivalor).

En general, se puede definir cualquier cantidad de niveles de tipos de entidad débil; un tipo de entidad propietaria puede ser ella misma un tipo de entidad débil.

Además, un tipo de entidad débil puede tener más de un tipo de entidad identificativa y un tipo de relación identificativa de grado superior a dos.

3.1.1.2. Atributos

Cada entidad tiene atributos, que son propiedades particulares que la describen y en el modelo ER hay varios tipos: simple frente a compuesto, monovalor frente a multivalor, almacenado frente a derivado y nulo.

Atributos compuestos frente a atributos simples Los atributos compuestos se pueden dividir en subpartes más pequeñas que representan atributos más básicos con significados independientes.

Los atributos que no son divisibles se denominan atributos simples o atómicos, mientras que los atributos compuestos pueden formar una jerarquía.

El valor de un atributo compuesto es la concatenación de los valores de sus atributos simples.

Atributos monovalor y multivalor La mayoría de los atributos tienen un solo valor para una entidad en particular; dichos atributos reciben el nombre de monovalor o de un solo valor.

En algunos casos, un atributo puede tener un conjunto de valores para la misma entidad y se denominan multivalor.

Un atributo multivalor puede tener límites superior e inferior para restringir el número de valores permitidos para cada entidad individual.

Atributos almacenados y derivados El atributo derivado puede calcularse u obtenerse a partir de otro atributo, que se denomina almacenado.

Atributos complejos Los atributos complejos son los atributos compuestos y multivalor que se anidan arbitrariamente.

Podemos representar el anidamiento arbitrario agrupando componentes de un atributo compuesto entre paréntesis () separando los componentes con comas, y mostrando los atributos multivalor entre llaves {}.

Por ejemplo, si una persona puede tener más de una residencia y cada residencia puede tener una sola dirección y varios teléfonos, el atributo TlfDir de una persona se puede especificar Como

```
{TlfDir({Tlf(CodÁrea,NumTlf)},
Dir(DirCalle(Número,Calle,NumApto),
Ciudad,Provincia,CP))}
```

Los atributos Tlf y Dir son compuestos.

Atributos clave de un tipo de entidad Una restricción importante de las entidades de un tipo de entidad es la clave o restricción de unicidad de los atributos.

Un tipo de entidad normalmente tiene un atributo cuyos valores son distintos para cada entidad del conjunto de entidades.

Los valores de un atributo en que se pueden utilizar para identificar cada entidad inequívocamente se denomina atributo clave.

En la notación diagramática ER cada atributo clave tiene su nombre subrayado dentro del óvalo y algunos tipos de entidad tienen más de un atributo clave.

Un tipo de entidad que carece de clave se le denomina tipo de entidad débil (que se explicará más adelante).

Conjuntos de valores (dominios) de atributos Cada atributo simple de un tipo de entidad está asociado con un conjunto de valor (o dominio de valores) que especifica el conjunto de los valores que se pueden asignar a ese atributo por cada entidad individual.

Los conjuntos de valores no se muestran en los diagramas ER; normalmente se especifican mediante los tipos de datos básicos disponibles en la mayoría de los lenguajes de programación, como entero, cadena, booleano, flotante, tipo enumerado, subrango, etcétera.

También se emplean otros tipos de datos adicionales para representar la fecha, la hora y otros.

Atributos de los tipos de relación Los tipos de relación también pueden tener atributos; los atributos de los tipos de relación 1:1 o 1:N se pueden trasladar a uno de los tipos de entidad participantes.

En el caso de un tipo de relación 1:N, un atributo de relación solo se puede migrar al tipo de entidad que se encuentra en el lado N de la relación.

Para los tipos de relación M:N, algunos atributos pueden determinarse mediante la combinación de entidades participantes en una instancia de relación, no mediante una sola relación. Dichos atributos deben especificarse como atributos de relación.

3.1.1.3. Relaciones

Un tipo de relación R entre n tipos de entidades E_1, E_2, \dots, E_n define un conjunto de asociaciones (o un conjunto de relaciones) entre las entidades de esos tipos de entidades.

Como en el caso de los tipos de entidades y los conjuntos de entidades, normalmente se hace referencia a un tipo de relación y su correspondiente conjunto de relaciones con el mismo nombre R .

Matemáticamente, el conjunto de relaciones R es un conjunto de instancias de relación r_i , donde cada r_i asocia n entidades individuales (e_1, e_2, \dots, e_n) y cada entidad e_j de r_i es un miembro del tipo de entidad $E_j, 1 \leq j \leq n$. Por tanto, un tipo de relación es una relación matemática en E_1, E_2, \dots, E_n .

De forma alternativa, se puede definir como un subconjunto del producto cartesiano $E_1 \times E_2 \times \dots \times E_n$. Se dice que cada uno de los tipos de entidad E_1, E_2, \dots, E_n participa en el tipo de relación R ; de forma parecida, cada una de las entidades individuales e_1, e_2, \dots, e_n se dice que participa en la instancia de relación $r_i = (e_1, e_2, \dots, e_n)$.

En los diagramas ER, los tipos de relaciones se muestran mediante rombos, conectados a su vez mediante líneas a los rectángulos que representan los tipos de entidad participantes y el nombre de la relación se muestra dentro del rombo.

Grado de relación, nombres de rol y relaciones recursivas

Grado de un tipo de relación El grado de un tipo de relación es el número de tipos de entidades participantes. Un tipo de relación de grado dos se denomina binario y uno de grado tres ternario. Las relaciones pueden ser generalmente de cualquier grado, pero las más comunes son las relaciones binarias.

Nombres de rol y relaciones recursivas Cada tipo de entidad que participa en un tipo de relación juega un papel o rol particular en la relación.

El nombre de rol hace referencia al papel que una entidad participante del tipo de entidad juega en cada instancia de relación y ayuda a explicar el significado de la relación.

Los nombres de rol no son técnicamente necesarios en los tipos de relación donde todos los tipos de entidad participantes son distintos, puesto que cada nombre de tipo de entidad participante se puede utilizar como participación.

Cuando un tipo de entidad se relaciona consigo misma, se tiene una relación recursiva y es necesario indicar los roles que juegan los miembros en la relación.

Restricciones en los tipos de relaciones Los tipos de relaciones normalmente tienen ciertas restricciones que limitan las posibles combinaciones entre las entidades que pueden participar en el conjunto de relaciones correspondiente.

Estas restricciones están determinadas por la situación del minimundo representado por las relaciones.

Podemos distinguir dos tipos principales de restricciones de relación: razón de cardinalidad y participación.

Razones de cardinalidad para las relaciones binarias La razón de cardinalidad de una relación binaria especifica el número máximo de instancias de relación en las que una entidad puede participar.

Las posibles razones de cardinalidad para los tipos de relación binaria son 1:1, 1:N, N:1 y M:N.

1. uno a uno: una relación R de X a Y es uno a uno si cada entidad en X se asocia con cuando mucho una entidad en Y e, inversamente, cada entidad en Y se asocia con cuando mucho una entidad en X.
2. uno a muchos: una relación R de X a Y es uno a muchos si cada entidad en X se puede asociar con muchas entidades en Y, pero cada entidad en Y se asocia con cuando mucho una entidad en X.

3. muchos a uno: una relación R de X a Y es muchos a uno si cada entidad en X se asocia con cuando mucho una entidad en Y , pero cada entidad en Y se puede asociar con muchas entidades en X .
4. muchos a muchos: una relación R de X a Y es muchos a muchos si cada entidad en X se puede asociar con muchas entidades en Y y cada entidad en Y se puede asociar con muchas entidades en X .

Restricciones de participación y dependencias de existencia La restricción de participación especifica si la existencia de una entidad depende de si está relacionada con otra entidad a través de un tipo de relación.

Esta restricción especifica el número mínimo de instancias de relación en las que puede participar cada entidad y en ocasiones recibe el nombre de restricción de cardinalidad mínima.

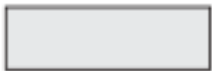




Hay dos tipos de restricciones de participación, total y parcial; la participación total también se conoce como dependencia de existencia.








1. Participación total: si todo miembro de un conjunto de entidades debe participar en una relación, es una participación total del conjunto de entidades en la relación. Esto se denota al dibujar una línea doble desde el rectángulo de entidades hasta el rombo de relación.
2. Participación parcial: una línea sencilla indica que algunos miembros del conjunto de entidades no deben participar en la relación.

Nos referiremos a la razón de cardinalidad y a las restricciones de participación, en conjunto, como restricciones estructurales de un tipo de relación.

3.1.1.4. Resumen de la notación para los diagramas ER

Tabla 3.1: Notación para los diagramas ER

Notación para los diagramas ER	
Símbolo	Significado
	Entidad
	Entidad débil
	Relación
	Relación de identificación
	Atributo

Continuación de Tabla 3.2	
Continuación la notación para los diagramas ER	
	Atributo clave
	Atributo multivalor
	Atributo compuesto
	Atributo derivado
	Participación total
	Razón de cardinalidad
	Restricción estructural
Fin de Tabla	

3.1.2. Modelo relacional

De acuerdo a la bibliografía de Elmasri[13], el modelo relacional introducido por Ted Codd en 1970[2] utiliza el concepto de una relación matemática como bloque de construcción básico y tiene su base teórica en la teoría de conjuntos y la lógica del predicado.

La lógica de predicado, utilizada ampliamente en matemáticas, proporciona un marco en el que una afirmación (declaración de hecho) puede verificarse como verdadera o falsa.

La teoría de conjuntos es una ciencia matemática que trata con conjuntos o grupos de cosas y se utiliza como base para la manipulación de datos en el modelo relacional.

El modelo relacional representa la base de datos como una colección de relaciones. Cuando una relación está pensada como una tabla de valores, cada fila representa una colección de valores relacionados.

Asimismo, cada fila de la tabla representa un hecho que, por lo general, corresponde con una relación o entidad real. El nombre de la tabla y de las columnas se utiliza para ayudar a interpretar el significado de cada uno de los valores de las filas.

En terminología formal, una fila recibe el nombre de tupla, una cabecera de columna es un atributo y el nombre de la tabla una relación. El tipo de dato que describe los valores que pueden aparecer en cada columna está representado por un dominio de posibles valores.

Basado en estos conceptos, el modelo relacional tiene tres componentes bien definidos:

1. Una estructura de datos lógica representada por relaciones.

2. Un conjunto de reglas de integridad para garantizar que los datos sean consistentes.
3. Un conjunto de operaciones que define cómo se manipulan los datos.

3.1.2.1. Relaciones

Estructuras de datos relacionales Su usan tablas con relación entre ellas.

Tablas En este modelo, las tablas se usan para contener información acerca de los objetos a representar en la base de datos. Al usar los términos del modelo Entidad-Relación, los conjuntos de entidades y de relaciones se muestran usando tablas.

Una relación o esquema de relación se representa como una tabla bidimensional en la que las filas de la tabla corresponden a registros individuales y las columnas corresponden a atributos.

Formalmente, un esquema de relación R , denotado por $R(A_1, A_2, \dots, A_n)$ está constituido por un nombre de relación R y una lista de atributos A_1, A_2, \dots, A_n .

Cada atributo A_i es el nombre de un papel jugado por algún dominio D en el esquema de relación R . Se dice que D es el dominio de A_i y se especifica como $dom(A_i)$.

Un esquema de relación se utiliza para describir una relación; se dice que R es el nombre de la misma. El grado de una relación es el número de atributos n de la misma.

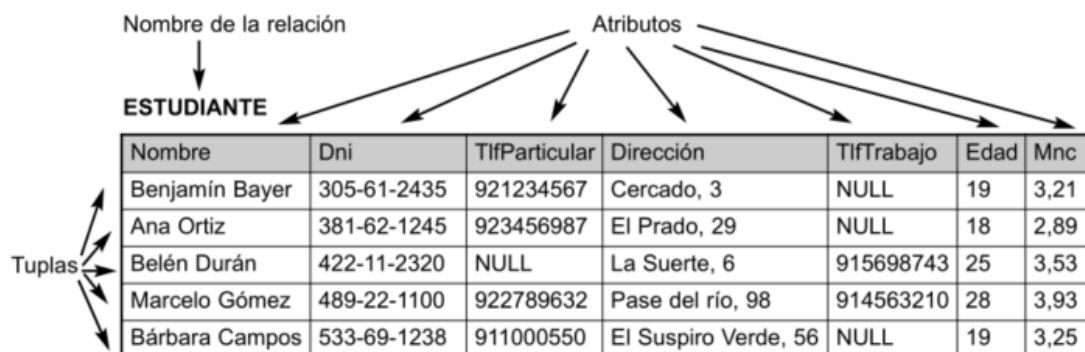


Figura 3.1: Tabla en el modelo Relacional

Cada fila de la tabla corresponde a un registro individual o instancia de entidad. En el modelo relacional cada fila se llama tupla y la tabla que representa una relación tiene las siguientes características:

- Cada celda de la tabla contiene solo un valor.
- Cada columna tiene un nombre distinto, que es el nombre del atributo que representa.
- Todos los valores en una columna provienen del mismo dominio, pues todos son valores del atributo correspondiente.

-
- Cada tupla o fila es distinta; no hay tuplas duplicadas.
 - El orden de las tuplas o filas es irrelevante.

Relaciones y tablas de bases de datos Una relación (o estado de relación) r del esquema $R(A_1, A_2, \dots, A_n)$, también especificado como $r(R)$, es un conjunto de n -tuplas $r = t_1, t_2, \dots, t_m$.

Cada tupla t es una lista ordenada de n valores $t = \langle v_1, v_2, \dots, v_n \rangle$, donde v_i , $1 \leq i \leq n$, es un elemento de $dom(A_i)$ o un valor especial NULL.

El i -ésimo valor de la tupla t , que se corresponde con el atributo A_i , se referencia como $t[A_i]$ o $t[i]$ si utilizamos una notación posicional.

3.1.2.2. Claves

En el modelo relacional, las claves son importantes porque aseguran que cada fila en una tabla sea unívocamente identificable.

También son usadas para establecer relaciones entre tablas y asegurar la integridad de los datos.

Una clave es un atributo o grupo de atributos que pueden identificar los valores de otros atributos.

Clave compuesta Una clave compuesta es una clave que se compone de más de un atributo. Un atributo que forma parte de una clave se denomina atributo clave.

Superclave Un atributo o atributos que identifican de manera única cualquier fila de una tabla

3.1.2.3. Restricciones de integridad

Integridad de dominio La integridad de dominio es la validez de las restricciones que debe cumplir una determinada columna de la tabla.

Integridad de entidad Todas las claves principales son únicas, y ninguna clave primaria puede ser nula.

Integridad referencial Una clave externa puede ser nula siempre que no sea parte de la clave principal de su tabla, o puede tener el valor que coincida con el valor de la clave primaria en una tabla con la que está relacionada (cada valor de clave externa no nula debe hacer referencia a un valor de clave primaria existente).

3.1.2.4. Propiedades de las relaciones

Grado El número de columnas en una tabla se llama grado de la relación. Una relación con una sola columna es de grado uno y se llama relación unaria. Una relación con dos columnas se llama binaria, una con tres columnas se llama ternaria y, después de ella, por lo general se usa el término n-aria. El grado de una relación es parte de la intensidad de la relación y nunca cambia.

Cardinalidad La cardinalidad de una relación es el número de entidades a las que otra entidad puede mapear bajo dicha relación.

3.1.2.5. ACID

El modelo relacional en las transacciones cumple con las propiedades de ACID, que es el acrónimo de *Atomicity* (atomicidad), *Consistency* (consistencia), *Isolation* (aislamiento) y *Durability* (durabilidad).

Atomicidad Requiere que se completen todas las operaciones (solicitudes SQL) de una transacción; si no, la transacción se cancela.

Si una transacción T_1 tiene cuatro solicitudes SQL, las cuatro solicitudes deben completarse con éxito; de lo contrario, se anula toda la transacción.

En otras palabras, una transacción se trata como una unidad de trabajo única, indivisible y lógica.

Consistencia Indica la permanencia del estado consistente de la base de datos. Una transacción lleva una base de datos de un estado consistente a otro.

Cuando se completa una transacción, la base de datos debe estar en un estado coherente. Si alguna de las partes de la transacción viola una restricción de integridad, se anula la transacción completa.

Aislamiento Significa que los datos utilizados durante la ejecución de una transacción no pueden ser utilizados por una segunda transacción hasta que se complete la primera.

En otras palabras, si la transacción T_1 se está ejecutando y está utilizando el elemento de datos X , ninguna otra transacción puede acceder a ese elemento de datos ($T_2...T_n$) hasta que finalice T_1 .

Esta propiedad es particularmente útil en entornos de bases de datos multiusuario porque varios usuarios pueden acceder y actualizar la base de datos al mismo tiempo.

Durabilidad Garantiza que una vez que se realizan y confirman los cambios en la transacción, no se pueden deshacer ni perder, incluso en el caso de una falla del sistema.

3.1.2.6. Structured Query Language

Structured Query Language o SQL está basado en el álgebra relacional, en el cálculo relacional y es un lenguaje de manipulación de datos, un lenguaje de definición de datos, un lenguaje de control de transacciones y un lenguaje de control de datos.

Lenguaje de manipulación de datos (DML) Un *Data Manipulation Language* o DML incluye comandos para insertar, actualizar, eliminar y recuperar datos dentro de las tablas de la base de datos.

Lenguaje de definición de datos (DDL) Un *Data Definition Language* o DDL incluye comandos para crear objetos de base de datos como tablas, índices y vistas, así como comandos para definir accesos a objetos de la base de datos.

Lenguaje de control de transacciones (TCL) Los comandos de un *Transaction Control Language* o TCL se ejecutan dentro del contexto de una transacción, que es una unidad lógica de trabajo compuesta por una o más instrucciones SQL.

SQL proporciona comandos para controlar el procesamiento de estas transacciones atómicas.

Lenguaje de control de datos (DCL) Los comandos de un *Data Control Language* o DCL se utilizan para controlar el acceso a los objetos de datos, como otorgar a un usuario permiso para ver solo una tabla y otorgar a otro usuario permiso para cambiar los datos de la misma tabla.

Tabla 3.2: Comandos de SQL.

Comandos de manipulación de datos	
Comando	Descripción
SELECT	Selecciona atributos de filas en una o más tablas o vistas
FROM	Especifica las tablas de las que se deben recuperar los datos
WHERE	Restringe la selección de filas en función de una expresión condicional
GROUP BY	Agrupar las filas seleccionadas en función de uno o más atributos
HAVING	Restringe la selección de filas agrupadas en función de una condición
ORDER BY	Ordena las filas seleccionadas en función de uno o más atributos
INSERT	Inserta filas en una tabla
UPDATE	Modifica los valores de un atributo en una o más filas de la tabla
DELETE	Elimina una o más filas de una tabla

Continuación de Tabla 3.2	
Continuación de comandos SQL	
Operadores de comparación =, <, >, ≤, ≥, <>, !=	Usados en expresiones condicionales
Operadores lógicos AND, OR, NOT	Usados en expresiones condicionales
Operadores especiales BETWEEN	Comprueba si un valor de atributo está dentro de un rango
IN	Comprueba si un valor de atributo coincide con algún valor dentro de una lista de valores
LIKE	Comprueba si un valor de atributo coincide con un patrón de cadena dado
IS NULL	Comprueba si un valor de atributo es nulo
EXIST	Comprueba si una subconsulta devuelve alguna fila
DISTINCT	Limita que los valores sean únicos
Comandos de definición de datos CREATE SCHEMA	Crea el esquema de la base de datos
CREATE TABLE	Crea una nueva tabla en el esquema de la base de datos del usuario
NOT NULL	Asegura que una columna no tendrá valores nulos
UNIQUE	Asegura que una columna no tendrá valores duplicados
PRIMARY KEY	Define una clave primaria para una tabla
FOREIGN KEY	Define una clave externa para una tabla
DEFAULT	Define un valor predeterminado para una columna (cuando no se proporciona ningún valor)
CHECK	Valida datos en un atributo
CREATE INDEX	Crea un índice para una tabla
CREATE VIEW	Crea un subconjunto dinámico de filas y columnas a partir de una o más tablas
ALTER TABLE	Modifica la definición de una tabla (agrega, modifica o elimina atributos o restricciones)
CREATE TABLE AS	Crea una nueva tabla basada en una consulta en el esquema de la base de datos del usuario
DROP TABLE	Elimina permanentemente una tabla (y sus datos)

Continuación de Tabla 3.2	
Continuación de comandos SQL	
DROP INDEX	Elimina permanentemente un índice
DROP VIEW	Elimina permanentemente una vista
Lenguaje de control de transacciones	
COMMIT	Guarda de manera permanente los cambios en los datos
ROLLBACK	Restaura los datos a sus valores anteriores
Lenguaje de control de datos	
GRANT	Le da un usuario permiso de hacer una acción de sistema o de acceder a datos de un objeto
REVOKE	Le quita el privilegio a un usuario de hacer algunas operaciones
Fin de Tabla	

3.1.3. Modelos NoSQL

De acuerdo a la bibliografía de Catherine [15], el término NoSQL significa *not only SQL* y se usa para agrupar sistemas de bases de datos diferentes a los relacionales.

Por los nuevos requerimientos en la época actual como disponibilidad total, tolerancia a fallos, almacenamiento de penta bytes de información distribuida en miles de servidores, la necesidad de nodos con escalabilidad horizontal, entre otros, surge la necesidad de sistemas de bases de datos no relacionales.

Estos tipos de sistemas no requieren esquemas fijos, son fáciles y rápidos en la instalación, usan lenguajes no declarativos, ofrecen alto rendimiento y disponibilidad, evitan operaciones de juntas, soportan paralelismo y escalan principalmente en forma horizontal soportando estructuras distribuidas que no necesariamente cumplen las propiedades ACID[15], sino que se enfocan en el modelo de consistencia de datos BASE.

3.1.3.1. Teorema CAP

En el Simposio de Principios de Computación Distribuida (PODC, en inglés) en el año 2000[16], el Dr. Eric Brewer declaró en su presentación que “en cualquier sistema de datos altamente distribuido hay tres propiedades comúnmente deseables: *Consistency* (consistencia), *Availability* (disponibilidad) y *Partition tolerance* (tolerancia al particionado). Sin embargo, es imposible que un sistema proporcione las tres propiedades al mismo tiempo”.

El acrónimo CAP representa las tres propiedades deseables:

Consistencia En una base de datos distribuida, la consistencia tiene el papel más importante. Todos los nodos deben ver los mismos datos al mismo tiempo, lo que significa que las réplicas deben actualizarse inmediatamente. Sin embargo, esto implica lidiar con la latencia y los atrasos de la red.

No hay que confundir la consistencia en la gestión de transacciones con la consistencia del teorema CAP. La consistencia de la gestión de transacciones se refiere al resultado cuando la ejecución de una transacción en una base de datos cumple con todas las restricciones de integridad.

La consistencia en CAP se basa en la suposición de que todas las transacciones tienen lugar al mismo tiempo en todos los nodos, como si se estuvieran ejecutando en una base de datos de un solo nodo (todos los nodos ven los mismos datos al mismo tiempo).

Disponibilidad En términos simples, el sistema siempre cumple una solicitud. Ninguna solicitud recibida se pierde y este es un requisito fundamental para todas las organizaciones centradas en la web.

Tolerancia al particionado El sistema continúa funcionando incluso en caso de falla de un nodo. Esto es equivalente a la transparencia de fallas en bases de datos distribuidas. El sistema fallará solo si fallan todos los nodos.

Aunque el teorema CAP se centra en sistemas basados en la web altamente distribuidos, sus implicaciones están muy extendidas para todos los sistemas distribuidos, incluidas las bases de datos.

En los sistemas de bases de datos, las propiedades ACID aseguran que todas las transacciones exitosas den como resultado un estado de base de datos consistente, uno en el que todas las operaciones de datos siempre devuelven los mismos resultados.

Para bases de datos distribuidas centralizadas y pequeñas, la latencia no es un problema, pero para una base de datos altamente distribuida el garantizar transacciones ACID sin pagar un alto precio en latencia de red o en conflictos de datos.

La relación entre consistencia y disponibilidad ha generado un nuevo tipo de sistemas de datos distribuidos, diferente al ACID, denominados BASE, *Basically Available* (básicamente disponibles), *Soft state* (estado suave), *Eventually consistent* (eventualmente consistente).

BASE BASE se refiere a un modelo de consistencia de datos en el que los cambios de datos no son inmediatos, sino que se propagan lentamente a través del sistema hasta que todas las réplicas sean consistentes.

En la práctica, la aparición de bases de datos distribuidas NoSQL proporciona un espectro de consistencia que va desde lo altamente consistente (ACID) hasta lo eventualmente consistente (BASE).

3.1.3.2. Clave-Valor

De acuerdo a la bibliografía de Coronel[17] Una base de datos de clave-valor, o almacén de clave-valor, es un paradigma de almacenamiento de datos diseñado para almacenar, recuperar y administrar arreglos asociativos.

Comúnmente se usa un diccionario o tabla hash que contienen una colección de registros anidados, secuenciada de bits, que se almacenan y se recuperan utilizando una clave que identifica de manera única el registro y se utiliza para encontrar rápidamente los datos dentro de la base de datos.

No obstante, es responsabilidad de las aplicaciones que hagan uso de este tipo de base de datos interpretar el significado de los datos. No hay claves foráneas y las relaciones no pueden rastrearse entre las claves. Esta simplificación en el RDBMS de clave-valor permite que sean rápidas y escalables.

Los pares de clave-valor generalmente se organizan en *buckets*. Todas las claves dentro un *bucket* deben ser únicas, pero pueden repetirse en otros *buckets* y todas las operaciones, incluidas las consultas, se basan en el *bucket* + la clave.

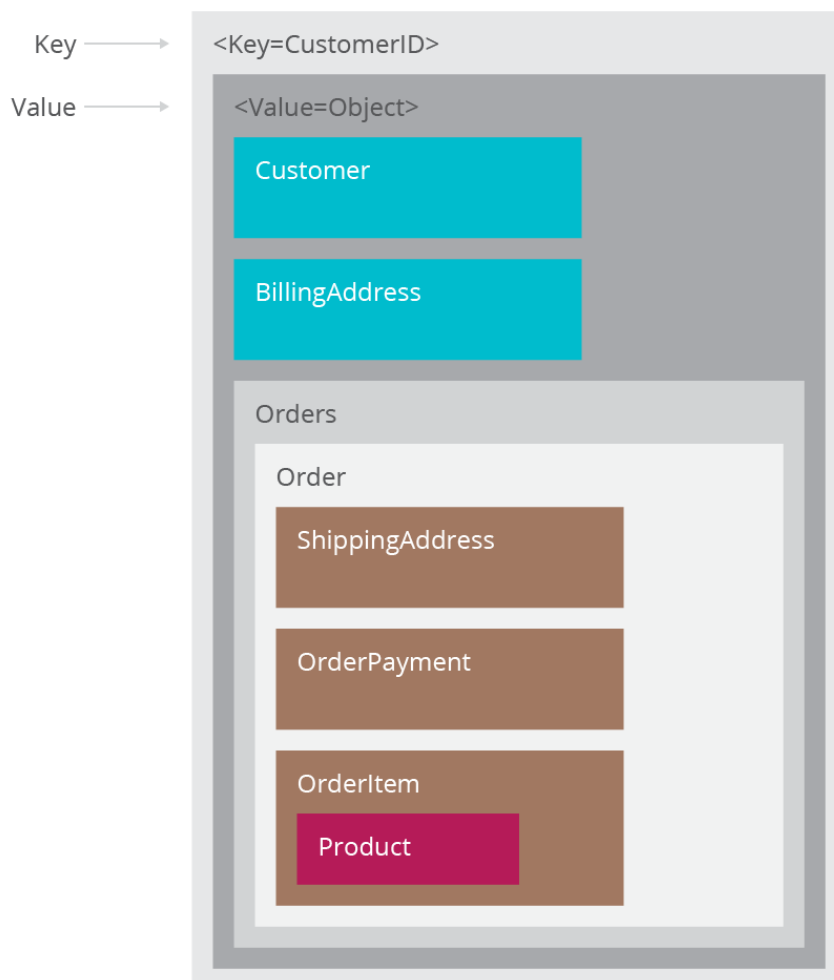


Figura 3.2: *Bucket* en el almacenamiento clave-valor

Respecto a las operaciones de este tipo de bases de datos, se usan las operaciones de *get*, *store* y *delete*. La operación *get* o *fetch* es usada para obtener el valor de un

par. El operador de *store* almacena datos en una clave. Si la combinación de *bucket* + clave no existe, se añade como un nuevo par de clave-valor. En cambio, en caso de existir la combinación de *bucket* + clave, el valor es reemplazado por el nuevo. El operador de *delete* es para eliminar un par de clave-valor.

Algunas de las bases de datos de clave-valor populares son Riak, Redis, Memcached DB, Berkeley DB, HamsterDB, Amazon DynamoDB y Project Voldemort (una implementación de código abierto de Amazon DynamoDB)[18].

3.1.3.3. Orientado a documentos

Una base de datos de documentos es una base de datos NoSQL que almacena datos en documentos etiquetados en pares clave-valor.

A diferencia de una base de datos clave-valor donde el componente de valor puede contener cualquier tipo de datos, una base de datos de documentos siempre almacena un documento en el componente de valor. El documento puede estar en cualquier formato codificado, como XML, JSON o BSON.

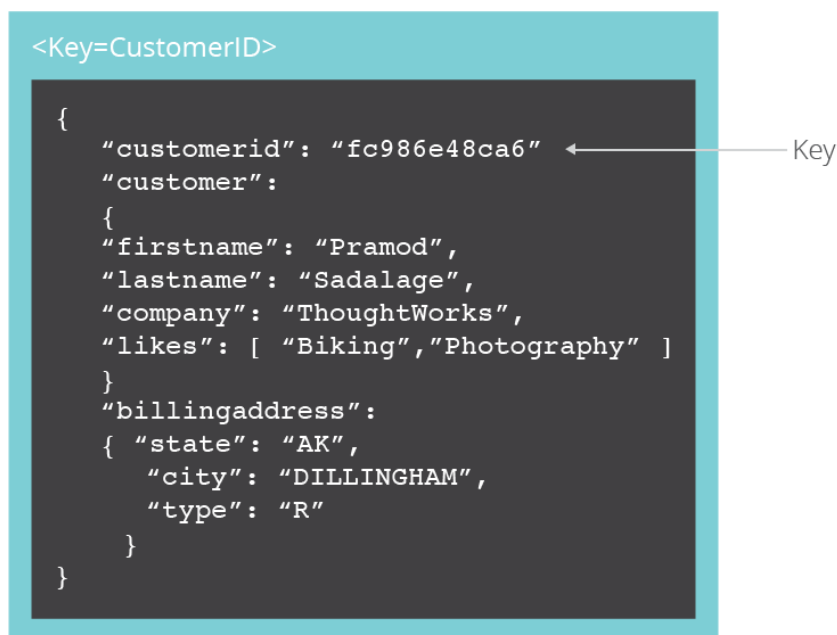


Figura 3.3: Documento en el almacenamiento de documentos

Otra diferencia importante es que, si bien las bases de datos clave-valor no intentan comprender el contenido del componente de valor, las bases de datos de documentos sí lo hacen. Las etiquetas representan porciones de un documento.

Por ejemplo, un documento puede tener etiquetas para identificar qué texto en el documento representa el título, el autor y el cuerpo del documento.

Dentro del cuerpo del documento, puede haber etiquetas adicionales para indicar capítulos y secciones. A pesar del uso de etiquetas en los documentos, las bases de datos de documentos se consideran sin esquema, es decir, no imponen una estructura predefinida en los datos almacenados.

Para una base de datos de documentos, no tener esquemas significa que aunque

todos los documentos tienen etiquetas, no todos los documentos deben tener las mismas etiquetas, por lo que cada documento puede tener su propia estructura.

Las etiquetas en una base de datos de documentos son extremadamente importantes porque son la base de la mayoría de las capacidades adicionales que tienen las bases de datos de documentos sobre las bases de datos clave-valor.

Las etiquetas dentro del documento son accesibles para el SGBD, lo que hace posible consultas complejas. Al igual que las bases de datos clave-valor agrupan pares clave-valor en grupos lógicos llamados *buckets*, las bases de datos de documentos agrupan documentos en grupos lógicos llamados colecciones.

Si bien se puede recuperar un documento especificando la colección y la clave, también es posible realizar consultas en función del contenido de las etiquetas.

Las bases de datos de documentos tienden a funcionar bajo el supuesto de que un documento es independiente, que no está en diferentes tablas como en una base de datos relacional.

Una base de datos de documentos asume que todos los datos relacionados de una orden estén en un solo documento.

Por lo tanto, cada orden en una colección contendría datos sobre el cliente, el pedido en sí y los productos comprados en esa orden.

Las bases de datos de documentos no almacenan relaciones como se hace en el modelo relacional y generalmente no tienen soporte para operaciones como la unión.

3.1.3.4. Orientado a columnas

Este modelo de base de datos se originó con el BigTable de Google. Otras bases de datos orientadas a columnas incluyen HBase, Hypertable y Cassandra.

Cassandra comenzó como un proyecto en Facebook, pero Facebook lo lanzó a la comunidad de código abierto, que ha seguido desarrollando a Cassandra en una de las bases de datos orientadas a columnas más populares.

Afortunadamente, las bases de datos de la familia de columnas son conceptualmente simples y conceptualmente lo suficientemente cercanas al modelo relacional para que su comprensión del modelo relacional pueda ayudarlo a comprender el modelo familiar de columnas.

El término base de datos orientada a columnas puede referirse a dos conjuntos diferentes de tecnologías que a menudo se confunden entre sí.

En cierto sentido, el término de base de datos orientada a columnas puede usarse para tecnologías de bases de datos relacionales tradicionales que usan almacenamiento enfocado en columnas en lugar de almacenamiento en filas.

Por otro lado, en un sistema de base de datos NoSQL describe un tipo de base de datos que organiza datos en pares nombre-valor donde el nombre actúa también como la clave.

Cada par de clave-valor representa una columna y siempre contiene una fecha que sirve para resolver conflictos de escritura, datos expirados entre otras cosas.

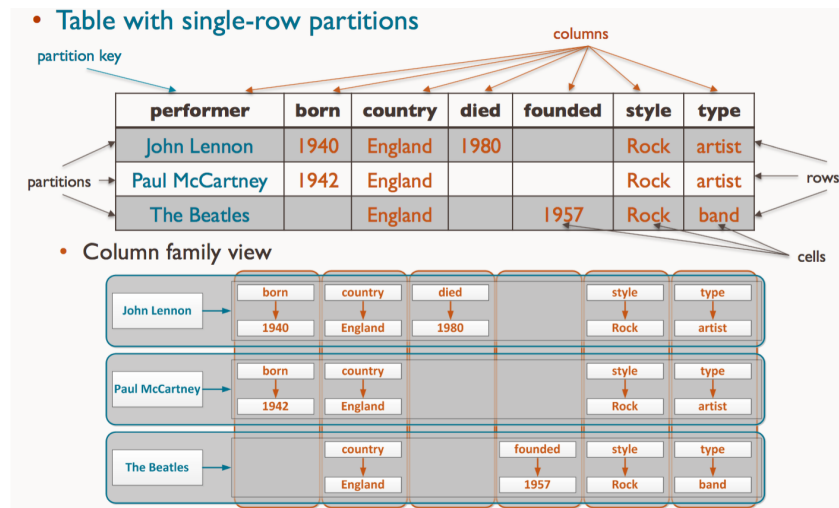


Figura 3.4: Familia de columnas

Al ser una bases de datos NoSQL, no requiere que los datos se ajusten a estructuras predefinidas ni admite SQL para consultas.

3.1.3.5. Orientado a grafos

Una base de datos de grafos es una base de datos NoSQL basada en la teoría de grafos para almacenar datos sobre entornos ricos en relaciones.

La teoría de grafos es un campo matemático y de ciencias de la computación que modela relaciones, o aristas, entre objetos llamados nodos.

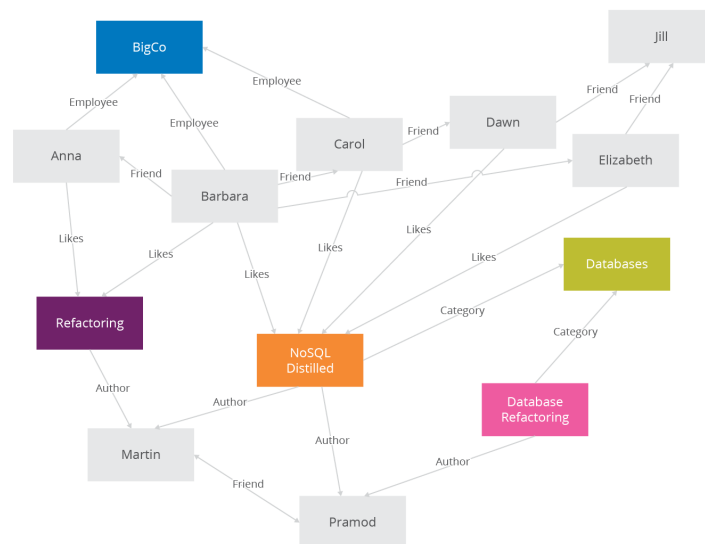


Figura 3.5: modelo conceptual orientado a grafos

Modelar y almacenar datos sobre relaciones es el enfoque de las bases de datos de grafos y la teoría en la que se basa es un campo de estudio bien establecido que se remonta a cientos de años.

Como resultado, la creación de un modelo de base de datos basado en la teoría

de grafos proporciona de inmediato una fuente rica de algoritmos y aplicaciones que han ayudado a las bases de datos de grafos a ganar terreno.

Los componentes principales de las bases de datos de grafos son nodos, aristas y propiedades, como se muestra en la figura 3.5. Un nodo corresponde a la idea de una instancia de entidad relacional.

El nodo es una instancia específica de algo sobre lo que queremos mantener datos. Cada nodo (círculo) en la figura 3.5 representa un solo agente.

Las propiedades son como atributos; son los datos que necesitamos almacenar sobre el nodo. Todos los nodos de agente pueden tener propiedades como nombre y apellido, pero no todos los nodos deben tener las mismas propiedades.

Un borde es una relación entre nodos. Los bordes (mostrados como flechas en la figura 3.5) pueden estar en una dirección, o pueden ser bidireccionales.

Para las consultas, la terminología correcta sería atravesar el grafo. Las bases de datos de grafos los recorridos se enfocan en las relaciones entre nodos, como la ruta más corta y el grado de conexión.

La base de datos de grafos comparte algunas características con otras bases de datos NoSQL en que las bases de datos de grafos no obligan a los datos a ajustarse a estructuras predefinidas, no admiten SQL y están optimizados para proporcionar velocidad de procesamiento, al menos para datos intensivos en relaciones.

Sin embargo, otras características clave no se aplican a las bases de datos de grafos. Las bases de datos de grafos no se escalan muy bien a los clústeres debido a las diferencias en los datos agregados.

3.2. Tecnologías a usar

Para seleccionar las tecnologías a usar para el desarrollo de la aplicación web, se ha optado por hacer un estudio y comparación de tecnologías similares.

Lo que resta de esta sección está organizado de la siguiente manera: primero se muestra cada tecnología usar. En caso de que sea la única opción se describirá qué es la herramienta y en caso de que haya varias opciones, se explicará cada opción y se tendrá un apartado al final de cada comparación sobre la herramienta que se ha elegido.

Finalmente, en la subsección de conclusiones se pondrá un resumen de cada tecnología escogida y el porqué.

3.2.1. Hypertext Transfer Protocol

De acuerdo a la W3C[19], Hypertext Transfer Protocol, HTTP, o protocolo de transferencia de hipertexto, es el protocolo de comunicación que permite las transferencias de información en la World Wide Web.

HTTP fue desarrollado por el World Wide Web Consortium y la Internet Engineering Task Force, colaboración que culminó en 1999 con la publicación de una

serie de RFC, siendo el más importante de ellos el RFC 2616 que especifica la versión 1.1. HTTP define la sintaxis y la semántica que utilizan los elementos de software de la arquitectura web (clientes, servidores, proxies) para comunicarse.

HTTP es un protocolo sin estado, es decir, no guarda ninguna información sobre conexiones anteriores. El desarrollo de aplicaciones web necesita frecuentemente mantener estado. Para esto se usan las cookies, que es información que un servidor puede almacenar en el sistema cliente.

Es un protocolo orientado a transacciones y sigue el esquema petición-respuesta entre un cliente y un servidor. Al cliente se le suele llamar “agente de usuario.” *user agent*, que realiza una petición enviando un mensaje con cierto formato al servidor. Al servidor se le suele llamar servidor web y envía un mensaje de respuesta.

Los mensajes HTTP son en texto plano lo que lo hace más legible y fácil de depurar. Esto tiene el inconveniente de hacer los mensajes más largos.

Los mensajes tienen la siguiente estructura:

1. Línea inicial (termina con retorno de carro y un salto de línea).
 - a) Para las peticiones: la acción requerida por el servidor (método de petición) seguido de la URL del recurso y la versión HTTP que soporta el cliente.
 - b) Para respuestas: La versión del HTTP usado seguido del código de respuesta (que indica que ha pasado con la petición seguido de la URL del recurso) y de la frase asociada a dicho retorno.
2. Las cabeceras del mensaje que terminan con una línea en blanco. Son metadatos. Estas cabeceras le dan gran flexibilidad al protocolo.
3. Cuerpo del mensaje: es opcional. Su presencia depende de la línea anterior del mensaje y del tipo de recurso al que hace referencia la URL. Típicamente tiene los datos que se intercambian cliente y servidor. Por ejemplo para una petición podría contener ciertos datos que se quieren enviar al servidor para que los procese. Para una respuesta podría incluir los datos que el cliente ha solicitado.

HTTP define una serie predefinida de métodos de petición que pueden utilizarse. El protocolo tiene flexibilidad para ir añadiendo nuevos métodos y para así añadir nuevas funcionalidades. El número de métodos de petición se ha ido aumentando según se avanzaba en las versiones.

Cada método indica la acción que desea que se efectúe sobre el recurso identificado. Lo que este recurso representa depende de la aplicación del servidor. Por ejemplo, el recurso puede corresponderse con un archivo que reside en el servidor.

1. GET: el método GET solicita una representación del recurso especificado. Las solicitudes que usan GET solo deben recuperar datos y no deben tener ningún otro efecto.

-
2. HEAD: pide una respuesta idéntica a la que correspondería a una petición GET, pero en la respuesta no se devuelve el cuerpo. Esto es útil para poder recuperar los metadatos de los encabezados de respuesta, sin tener que transportar todo el contenido.
 3. POST: envía los datos para que sean procesados por el recurso identificado. Los datos se incluirán en el cuerpo de la petición. Esto puede resultar en la creación de un nuevo recurso o de las actualizaciones de los recursos existentes o ambas cosas.
 4. PUT: sube, carga o realiza un upload de un recurso especificado (archivo o fichero) y es un camino más eficiente ya que POST utiliza un mensaje multiparte y el mensaje es decodificado por el servidor. En contraste, el método PUT permite escribir un archivo en una conexión socket establecida con el servidor. La desventaja del método PUT es que los servidores de alojamiento compartido no lo tienen habilitado.

3.2.2. HTML 5

De acuerdo a la documentación de Mozilla[20], HyperText Markup Language, abreviado HTML, o lenguaje de marcado de hipertextos, es la pieza más básica para la construcción de la web y se usa para definir el sentido y estructura del contenido en una página web.

Es un estándar a cargo del World Wide Web Consortium (W3C) o Consorcio WWW, organización dedicada a la estandarización de casi todas las tecnologías ligadas a la web.

HTML hace uso de enlaces que conectan las páginas web entre sí, ya sea dentro de un mismo sitio web o entre diferentes sitios web.

Un elemento HTML se separa de otro texto en un documento por medio de “etiquetas”, las cuales consisten en elementos rodeados por “<,>”.

Ejemplos de estas etiquetas son <head>, <title>, <body>, <header>, <p>, , , , y otros más.

HTML 5 (HyperText Markup Language, versión 5) es la quinta revisión importante del lenguaje básico de la World Wide Web, HTML.

HTML5 establece elementos y atributos que reflejan el uso típico de los sitios web modernos. Algunos de ellos son técnicamente similares a las etiquetas <div> y , pero tienen un significado semántico, como por ejemplo <nav> (bloque de navegación del sitio web) y <footer>.

Características:

1. Incorpora etiquetas: *canvas* 2D y 3D, audio, vídeo con codecs para mostrar los contenidos multimedia. Actualmente hay una lucha entre imponer codecs libres (WebM + VP8) o privados (H.264/MPEG-4 AVC).
2. Etiquetas para manejar grandes conjuntos de datos: Datagrid, Details, Menu

y Command. Permiten generar tablas dinámicas que pueden filtrar, ordenar y ocultar contenido en cliente.

3. Mejoras en los formularios: Nuevos tipos de datos (*email*, *number*, *url*, *datetime*) y facilidades para validar el contenido sin JavaScript.
4. Visores: MathML (fórmulas matemáticas) y SVG (gráficos vectoriales); en general se deja abierto a poder interpretar otros lenguajes XML.
5. Drag & Drop: nueva funcionalidad para arrastrar objetos como imágenes.

Respecto a la compatibilidad con los navegadores, la mayoría de elementos de HTML5 son compatibles con Firefox 19, Chrome 25, Safari 6 y Opera 12 en adelante.

3.2.3. Style Sheet Language: CSS vs SASS

De acuerdo con Wikipedia[21], un *style sheet language* es un lenguaje que representa los estilos o elementos visuales de documentos estructurados.

CSS

De acuerdo a la documentación de la W3C[22], Cascading Style Sheets, CSS, o hojas de estilo en cascada, es un lenguaje de diseño gráfico para definir y crear la presentación de un documento estructurado escrito en un lenguaje de marcado.

CSS está diseñado principalmente para marcar la separación del contenido del documento y la presentación del mismo, con características como las capas o layouts, los colores y las fuentes.

La separación entre el contenido del documento y la presentación busca mejorar la accesibilidad, proveer más flexibilidad y control, permitir que varios documentos HTML compartan un mismo estilo usando una sola hoja de estilos separada en un archivo .css y reducir la complejidad y la repetición de código en la estructura del documento.

La especificación CSS describe un esquema prioritario para determinar qué reglas de estilo se aplican si más de una regla coincide para un elemento en particular. Estas reglas son aplicadas con un sistema llamado de cascada, de modo que las prioridades son calculadas y asignadas a las reglas, así que los resultados son predecibles.

La especificación CSS es mantenida por el World Wide Web Consortium (W3C). El MIME type text/css está registrado para su uso por CSS descrito en el RFC 23185. El W3C proporciona una herramienta de validación de CSS gratuita para los documentos CSS.

CSS se ha creado en varios niveles y perfiles. Cada nivel de CSS se construye sobre el anterior, generalmente añadiendo funciones al nivel previo.

Los perfiles son, generalmente, parte de uno o varios niveles de CSS definidos para un dispositivo o interfaz particular. Actualmente, pueden usarse perfiles para dispositivos móviles, impresoras o televisiones.

La última versión del estándar, CSS3.1, está dividida en varios documentos separados, llamados “módulos”.

Cada módulo añade nuevas funcionalidades a las definidas en CSS2, de manera que se preservan las anteriores para mantener la compatibilidad.

Los trabajos en el CSS3.1 comenzaron a la vez que se publicó la recomendación oficial de CSS2 y los primeros borradores de CSS3.1 fueron liberados en junio de 1999.

Debido a la modularización del CSS3.1, diferentes módulos pueden encontrarse en diferentes estados de su desarrollo, de forma que hay alrededor de cincuenta módulos publicados, tres de ellos se convirtieron en recomendaciones oficiales de la W3C en 2011: “Selectores”, “Espacios de nombres”, y “Color”.

Respecto al soporte de los navegadores web, cada navegador web usa un motor de renderizado para renderizar páginas web, y el soporte de CSS no es exactamente igual en ninguno de los motores de renderizado. Ya que los navegadores no aplican el CSS correctamente, muchas técnicas de programación han sido desarrolladas para ser aplicadas por un navegador específico (comúnmente conocida esta práctica como *CSS hacks* o *CSS filters*).

Además, CSS3 definen muchas queries, entre las cuales se provee la directiva @supports que permite a los desarrolladores especificar navegadores con soporte para alguna función en específico directamente en el CSS3.1.

Sass

De acuerdo a la documentación de Sass[\[23\]](#), Sass es un lenguaje de preprocesado que genera hojas de estilo en cascada (CSS) y consta de dos sintaxis.

Características

Variables Las variables comienzan con un signo de dólar y la asignación de valor se realiza con dos puntos.

SassScript permite 4 tipos de datos:

1. Números (incluyendo las unidades).
2. Strings (con comillas o sin ellas).
3. Colores (código, o nombre).
4. Booleanos.

Las variables pueden ser resultados o argumentos de varias funciones disponibles. Durante el proceso de traducción, los valores de las variables son insertados en el documento CSS de salida.

Anidamiento CSS soporta anidamiento lógico, pero los bloques de código no son anidados. Sass permite que el código anidado sea insertado dentro de cualquier otro bloque.³

Mixins Como CSS no soporta *mixins*, cualquier código duplicado debe ser repetido en cada lugar. Un *mixin* en Sass es una sección de código que contiene código Sass.

Cada vez que se llama un *mixin* en el proceso de conversión el contenido del mismo es insertado en el lugar de la llamada. Los *mixin* permiten una solución limpia a las repeticiones de código, así como una forma fácil de alterar el mismo.

Elección

Tomando en cuenta la experiencia del equipo con CSS, además de que el proyecto no se enfocará en hacer muchas hojas de estilo para cada componente, página o vista de la aplicación web, se usará CSS en lugar de Sass.

3.2.4. Web Frameworks

De acuerdo con stack overflow^[24], los web *frameworks* más populares están escritos en JavaScript/TypeScript, por ello se realizó una investigación del lenguaje más apto para el proyecto.

JavaScript

De acuerdo con la documentación de Mozilla^[25], JavaScript es una marca registrada con licencia de Sun Microsystems (ahora Oracle) que se usa para describir la implementación del lenguaje de programación JavaScript.

Debido a problemas de registro de marcas en la Asociación Europea de Fabricantes de Computadoras, la versión estandarizada del lenguaje tiene el nombre de ECMAScript, sin embargo, en la práctica se conoce como lenguaje JavaScript.

Abreviado como JS, es un lenguaje ligero e interpretado, orientado a objetos, basado en prototipos, imperativo, débilmente tipado y dinámico; es usado en node.js, Apache CouchDB y Adobe Acrobat.

EL núcleo del lenguaje JavaScript está estandarizado por el Comité ECMA TC39 como un lenguaje llamado ECMAScript. La última versión de la especificación es ECMAScript 6.0.

El estándar ECMAScript define

1. Sintaxis: reglas de análisis, palabras clave, flujos de control, inicialización literal de objetos.
2. Mecanismos de control de errores: *throw*, *try/catch*, habilidad para crear tipos de Errores definidos por el usuario.

-
3. Tipos: *boolean*, *number*, *string*, *function*, *object*.
 4. Objetos globales: en un navegador, los objetos globales son los objetos de la ventana, pero ECMAScript solo define API no específicas para navegadores, como *parseInt*, *parseFloat*, *decodeURI*, *encodeURI*.
 5. Mecanismo de herencia basada en prototipos.
 6. Objetos y funciones incorporadas.
 7. Modo estricto.

La sintaxis básica es similar a Java y C++ con la intención de reducir el número de nuevos conceptos necesarios para aprender el lenguaje. Las construcciones del lenguaje, tales como sentencias *if*, bucles *for* y *while*, bloques *switch* y *try catch* funcionan de la misma manera que en estos lenguajes (o casi).

JavaScript puede funcionar como lenguaje procedimental y como lenguaje orientado a objetos. Los objetos se crean programáticamente añadiendo métodos y propiedades a lo que de otra forma serían objetos vacíos en tiempo de ejecución, en contraposición a las definiciones sintácticas de clases comunes en los lenguajes compilados como C++ y Java. Una vez se ha construido un objeto, puede usarse como modelo (o prototipo) para crear objetos similares.

Las capacidades dinámicas de JavaScript incluyen construcción de objetos en tiempo de ejecución, listas variables de parámetros, variables que pueden contener funciones, creación de scripts dinámicos (mediante *eval*), introspección de objetos (mediante *for ... in*), y recuperación de código fuente (los programas de JavaScript pueden decompilar el cuerpo de funciones a su código fuente original).

Desde el 2012, todos los navegadores modernos soportan completamente ECMAScript 5.1. Los navegadores más antiguos soportan por lo menos ECMAScript 3.

El 17 de Julio de 2015, ECMA International publicó la sexta versión de ECMAScript, la cual es oficialmente llamada ECMAScript 2015 y fue inicialmente nombrada como ECMAScript 6 o ES6. Desde entonces, los estándares ECMAScript están en ciclos de lanzamiento anuales.

TypeScript

De acuerdo con TypeScript Publishing[26], TypeScript es por definición es JavaScript para el desarrollo de aplicaciones, siendo también un superconjunto del mismo.

TypeScript es un lenguaje compilado orientado a objetos. Fue diseñado por Anders Hejlsberg (diseñador de C#) en Microsoft. TypeScript es tanto un lenguaje como un conjunto de herramientas. TypeScript es un superconjunto de JavaScript que genera código JavaScript.

Características

- Compilación: cuenta con un transpilador para la verificación de errores si hay errores de compilación, cosa que no es posible con JavaScript.
- Típo estático fuerte: provee un sistema opcional de típo estático y de inferencia de tipos a través del TypeScript Language Service, lo que permite inferir el tipo de una variable declara sin tipo en función de su valor.
- Definiciones de tipo: permite la extensión del lenguaje con bibliotecas externas JavaScript.
- Programación orientada a objetos: admite conceptos como clases, interfaces, herencia, etc.

Elección

De acuerdo con el sitio [stack overflow\[27\]](#), JavaScript es el lenguaje más popular de 2019 y aunque TypeScript es de los lenguajes que tienen un mayor nivel de aceptación, se usará JavaScript no solo por ser el lenguaje más popular y, en consecuencia, con más compatibilidad y material de ayuda, sino también porque el equipo está acostumbrado a este lenguaje y tiene experiencia con web *frameworks* escritos en JavaScript.

3.2.5. JavaScript Web Frameworks: Vue/Nuxt vs React vs Angular

De acuerdo con Wikipedia[\[28\]](#), un web *framework* es un conjunto de *software* que permite el desarrollo de una aplicación web y en el lenguaje JavaScript hay varias opciones, incluidas las más populares Vue, React y Angular.

React

De acuerdo con Wikipedia[\[29\]](#), React es una biblioteca Javascript de código abierto diseñada para crear interfaces de usuario con el objetivo de facilitar el desarrollo de *single page applications*.

Características

1. Virtual DOM: React usa un virtual DOM propio en lugar del navegador.
2. Props: son definidos como los atributos de configuración para dicho componente.
3. Estado de cada componente: lleva un registro de las propiedades y atributos del componente.
4. Ciclos de vida: son la serie de estados por los cuales pasan los componentes *statefull* a lo largo de su existencia.

Angular

De acuerdo con Wikipedia[30], Angular es un web *framework* desarrollado en TypeScript de código abierto mantenido por Google que se utiliza para crear y mantener *single page applications*.

1. Generación de código
2. Componentes
3. Ciclos de vida de componentes

Vue/Nuxt

De acuerdo con la documentación de Vue.js[31], Vue es un *framework* progresivo para desarrollar interfaces de usuario. A diferencia de otros *frameworks* monolíticos, Vue está diseñado desde cero para ser utilizado incrementalmente.

La librería central está enfocada solo en la capa de visualización y es fácil de utilizar e integrar con otras librerías o proyectos existentes. Por otro lado, Vue también es perfectamente capaz de impulsar sofisticadas *single-page applications* cuando se utiliza en combinación con librerías de apoyo.

Comparación con React React y Vue comparten muchas similitudes; ambos utilizan un DOM virtual, proporcionan componentes de vista reactivos y componentes, mantienen el enfoque en la librería central, con temas como el enrutamiento y la gestión global del estado manejadas por librerías asociadas.

Tanto React como Vue ofrecen un rendimiento comparable en los casos de uso más comunes, con Vue normalmente un poco por delante debido a su implementación más ligera del DOM virtual.

En Vue, las dependencias de un componente se rastrean automáticamente durante su renderizado, por lo que el sistema sabe con precisión qué componentes deben volver a renderizarse cuando cambia el estado. Se puede considerar que cada componente tiene un *shouldComponentUpdate* automáticamente implementado.

En general, esto elimina la necesidad de toda una clase de optimizaciones de rendimiento a los desarrolladores y les permite centrarse más en la construcción de la aplicación en sí misma a medida que va escalando.

Comparación con Angular En términos de rendimiento, ambos *frameworks* son excepcionalmente rápidos y no hay suficientes datos de casos de uso en el mundo real para hacer un veredicto.

Vue es mucho menos intrusivo en las decisiones del desarrollador que Angular, ofreciendo soporte oficial para una variedad de sistemas de desarrollo, sin restricciones sobre cómo estructurar su aplicación. Muchos desarrolladores disfrutaban de esta libertad, mientras que algunos prefieren tener solo una forma correcta de desarrollar cualquier aplicación.

Para empezar con Vue, todo lo que se necesita es familiarizarse con HTML y ES5 JavaScript; con estas habilidades básicas, se puede empezar a desarrollar aplicaciones no triviales.

La curva de aprendizaje de Angular es mucho más pronunciada. La superficie de la API del framework es enorme y como usuario necesitará familiarizarse con muchos más conceptos antes de ser productivo. La complejidad de Angular se debe en gran medida a su objetivo de diseño de apuntar solo a aplicaciones grandes y complejas - pero eso hace que el *framework* sea mucho más difícil de entender para los desarrolladores con menos experiencia.

Nuxt.js

De acuerdo a la documentación de Nuxt[32], Nuxt.js es un *framework* progresivo basado en Vue.js para crear aplicaciones web. Se basa en Vue.js y herramientas de desarrollo como webpack, Babel y PostCSS. El objetivo de Nuxt es hacer que el desarrollo web en Vue.js sea eficaz.

Características

1. Manejo de archivos Vue (*.vue)
2. División automática de código
3. Representación del lado del servidor
4. Potente sistema de enrutamiento con datos asincrónicos
5. Servicio de archivos estáticos
6. Soporte sintaxis ES2015+ (Javascript ES6)
7. Gestión del elemento <head> (<title>, <meta>, etc.)
8. Preprocesador: Sass, Less, Stylus, etc.

Elección

Se usará Vue/Nuxt por ser el *framework* con el que el equipo está más acostumbrado, además de ser el más flexible de las opciones expuestas.

3.2.6. CSS Frameworks: Vuetify vs Bootstrap

De acuerdo con Wikipedia[33], un CSS *framework* es una biblioteca de estilos genéricos que puede ser usada para implementar diseños web y aportan una serie de utilidades que pueden ser aprovechadas frecuentemente en los distintos diseños web.

Vuetify

De acuerdo a la documentación de Google[34], Material Design es un lenguaje visual que sintetiza los principios clásicos del buen diseño respecto a las ideas de Google y en estos principios está basado Vuetify.

El objetivo de Material Design es crear un lenguaje visual que sintetice los principios clásicos del buen diseño, unificar el desarrollo de un único sistema subyacente que unifique la experiencia del usuario en plataformas y dispositivos, así como personalizar el lenguaje visual de Material Design.

Asimismo, de acuerdo con la documentación de Vuetify[35], este CSS *framework* está integrado para ser usado en los componentes de Vue/Nuxt, desde botones, barras de navegación, *layouts* y demás.

Bootstrap

De acuerdo a la documentación de Bootstrap[36], es un CSS *framework* orientado al diseño responsivo de una aplicación web.

Tiene *templates* para botones, barras de navegación, estilos de tipografía entre otros. Es de fácil integración con React, AngularJS o Vue y tiene una comunidad extensa por los años y popularidad que tiene.

Elección

Se ha elegido usar en una primera instancia Vuetify porque es un CSS *framework* que está integrado en las tecnologías asociadas de Vue, como Vue Router, Vue Meta. Asimismo, sus componentes son simples de entender y de implementar.

3.2.7. MongoDB vs Apache CouchDB

MongoDB

De acuerdo a la documentación de Wikipedia[37], MongoDB (del inglés humongous, “enorme”) es un sistema de base de datos NoSQL, orientado a documentos y de código abierto.

En lugar de guardar los datos en tablas, tal y como se hace en las bases de datos relacionales, MongoDB guarda estructuras de datos BSON (una especificación similar a JSON) con un esquema dinámico, haciendo que la integración de los datos en ciertas aplicaciones sea más fácil y rápida.

Características

1. Consultas ad hoc: MongoDB soporta la búsqueda por campos, consultas de rangos y expresiones regulares.

-
2. Indexación: cualquier campo en un documento de MongoDB puede ser indexado, al igual que es posible hacer índices secundarios.
 3. Replicación: MongoDB soporta el tipo de replicación primario-secundario.
 4. Balanceo de carga; MongoDB puede escalar de forma horizontal usando el concepto de *sharding*.
 5. Almacenamiento de archivos: MongoDB puede ser utilizado como un sistema de archivos, aprovechando la capacidad de MongoDB para el balanceo de carga y la replicación de datos en múltiples servidores.
 6. Agregación: MongoDB proporciona un framework de agregación que permite realizar operaciones similares al GROUP BY de SQL.

Apache CouchDB

De acuerdo con Wikipedia[38], se trata de una base de datos NoSQL que emplea JSON para almacenar los datos, JavaScript como lenguaje de consulta por medio de MapReduce y HTTP como API.

Características

1. Almacenamiento de documentos: CouchDB almacena los datos como “documentos”, esto es, uno o más pares campo/valor expresados en JSON.
2. Semántica ACID: CouchDB provee una semántica de atomicidad, consistencia, aislamiento y durabilidad.
3. Vistas e índices Map/Reduce: los datos almacenados se estructuran por medio de vistas.
4. Arquitectura distribuida con replicación: CouchDB se diseñó con teniendo en mente la replicación bidireccional (o sincronización) y la operación *offline*.
5. Interfaz REST: todos los ítems tienen una URI única que queda expuesta vía HTTP. REST usa los métodos HTTP POST, GET, PUT y DELETE para las cuatro operaciones básicas CRUD (*Create, Read, Update, Delete*).

Elección

De acuerdo con Mosquera[6], MongoDB es la base de datos NoSQL orientada a documentos más popular y usada en los *papers* de investigación. Por ello se usará MongoDB como base de datos.

3.2.8. Bibliotecas JavaScript para diagramado: GoJS vs Fabric.js vs D3.js

GoJS

De acuerdo a la documentación de GoJS[39], GoJS es una biblioteca de JavaScript y TypeScript para crear diagramas y gráficos interactivos.

GoJS le permite crear todo tipo de diagramas y gráficos para sus usuarios, desde simples diagramas de flujo y organigramas hasta diagramas industriales altamente específicos, diagramas SCADA y BPMN, diagramas médicos como genogramas y diagramas de modelos de brotes, y más.

GoJS facilita la construcción de diagramas JavaScript de nodos complejos, enlaces y grupos con plantillas y diseños personalizables.

GoJS ofrece muchas funciones avanzadas para la interactividad del usuario, como arrastrar y soltar, copiar y pegar, edición de texto en el lugar, información sobre herramientas, menús contextuales, diseños automáticos, plantillas, enlace de datos y modelos, gestión de estado y deshacer transaccional, paletas, descripciones generales, controladores de eventos, comandos, herramientas extensibles para operaciones personalizadas y animaciones personalizables.

GoJS se implementa en TypeScript y puede usarse como una biblioteca de JavaScript o incorporarse a su proyecto desde fuentes de TypeScript.

GoJS normalmente se ejecuta completamente en el navegador, renderizando a un elemento HTML Canvas o SVG sin ningún requisito del lado del servidor.

Fabric.js

De acuerdo con la documentación de Fabric[40], Fabric.js es una biblioteca de JavaScript que proporciona un modelo para trabajar sobre un canvas HTML5 para poder agregar objetos como rectas, circunferencias, rectángulos, etc.

Características

1. Drag & Drop integrado en cada objeto de Fabric.js
2. Permite la especialización de clases para crear objetos personalizados.

D3.js

De acuerdo con Wikipedia[41], D3.js es una biblioteca de JavaScript para producir a partir de datos infogramas dinámicos e interactivos en navegadores web.

Características

1. Selecciones: se pueden seleccionar elementos del documento HTML y asignarle propiedades.

-
2. Transiciones: permiten interpolar en el tiempo valores de atributos, lo que produce cambios visuales en los infogramas.
 3. Asociación de datos: se asocia a cada elemento un objeto SVG con propiedades (forma, colores, valores) y comportamientos (transiciones, eventos).

Elección

Se llevó a la práctica en el prototipo funcional las tres opciones antes expuestas junto con algunas otras y se decidió usar GoJS para el proyecto por ser la biblioteca JavaScript más completa para diagramado y que genere los diagramas con un JSON simple para parsear esos datos y realizar la conversión.

3.2.9. Python 3

De acuerdo con Mark Lutz[42], Python es un lenguaje de programación interpretado, interactivo y orientado a objetos. Incorpora módulos, excepciones, tipeo dinámico, tipos de datos dinámicos y clases.

Tiene sintaxis clara, interfaces para muchas llamadas de sistema y bibliotecas, así como para varios sistemas de ventanas. Además, es extensible en C o C++.

También se puede usar como un lenguaje de extensión para aplicaciones que necesitan una interfaz programable y es portátil: se ejecuta en muchas variantes de Unix, en Mac y en Windows 2000 y versiones posteriores.

La biblioteca estándar del lenguaje, cubre áreas como el procesamiento de cadenas (expresiones regulares, Unicode, cálculo de diferencias entre archivos), protocolos de Internet (HTTP, FTP, SMTP, XML-RPC, POP, IMAP, programación CGI), ingeniería de software (pruebas unitarias, registro, creación de perfiles, análisis del código Python) e interfaces del sistema operativo (llamadas al sistema, sistemas de archivos, sockets TCP/IP).

Fortalezas

Es orientado a objetos y funcional Python es un lenguaje orientado a objetos; su modelo de clase admite nociones avanzadas como el polimorfismo, la sobrecarga del operador y la herencia múltiple; sin embargo, en el contexto de la simple sintaxis y escritura de Python, la programación orientada a objetos es fácil de aplicar.

Además de servir para la estructuración y reutilización de código, la naturaleza orientada a objetos de Python lo hace ideal como herramienta de secuencias de comandos para otros lenguajes de sistemas orientados a objetos. Por ejemplo, con el código apropiado, los programas Python pueden especializar clases implementadas en C++, Java y C#.

No obstante, la programación orientada a objetos es una opción en Python. Al igual que C++, Python admite modos de programación tanto procedimentales como orientados a objetos. Las herramientas orientadas a objetos se pueden aplicar siempre que las restricciones lo permitan.

Además de sus paradigmas originales de procedimientos (basados en declaraciones) y orientados a objetos (basados en clases), Python en los últimos años ha adquirido soporte incorporado para la programación funcional, un conjunto que incluye generadores, comprensiones, cerraduras, mapas, decoradores, funciones anónimas y lambdas.

Es extensible Su conjunto de herramientas lo ubica entre los lenguajes de *scripting* tradicionales como Tcl, Scheme y Perl; y los lenguajes de desarrollo de sistemas como C, C++ y Java.

Python proporciona toda la simplicidad y facilidad de uso de un lenguaje de programación, junto con herramientas de ingeniería de software más avanzadas que normalmente se encuentran en lenguajes compilados.

A diferencia de algunos lenguajes de secuencias de comandos, esta combinación hace que Python sea útil para proyectos de desarrollo a gran escala. Algunas de las herramientas de Python son:

Escritura dinámica Python realiza un seguimiento de los tipos de objetos que utiliza su programa cuando se ejecuta; eso no requiere declaraciones complicadas de tipo y tamaño en su código. De hecho, no existe una declaración de tipo o variable en Python.

Debido a que el código Python no restringe los tipos de datos, también se aplica automáticamente a toda una gama de objetos.

Gestión automática de la memoria Python asigna automáticamente objetos y los reclama el recolector de basura cuando ya no se usan y la mayoría puede crecer y reducirse según la demanda. Es decir, Python realiza un seguimiento de los detalles de la memoria de bajo nivel.

Tipos de objetos incorporados Python proporciona estructuras de datos de uso común como listas, diccionarios y cadenas como partes intrínsecas del lenguaje. Son flexibles y fáciles de usar. Por ejemplo, los objetos integrados pueden crecer y reducirse según demanda, pueden anidarse arbitrariamente para representar información compleja, y más.

Herramientas incorporadas Para procesar todos esos tipos de objetos, Python viene con operadores potentes y estándar, que incluyen concatenación (unir colecciones), segmentar (extraer secciones), ordenar, mapear y más.

Utilidades de biblioteca Para tareas más específicas, Python también viene con una gran colección de herramientas de biblioteca precodificadas que admiten todo, desde la coincidencia de expresiones regulares hasta la creación de redes. Una vez que aprende el lenguaje en sí, las herramientas de la biblioteca de Python son donde ocurre gran parte de la acción a nivel de aplicación.

Utilidades de terceros Debido a que Python es de código abierto, los desarrolladores pueden contribuir con herramientas precodificadas que admitan tareas que aún no son herramientas estándar; en la Web, encontrará soporte gratuito para COM, imágenes, programación numérica, XML y acceso a bases de datos.

Elección

Se ha elegido Python para desarrollar los algoritmos del proyecto dado que es multiplataforma y es de fácil integración con el *framework* web elegido para el back end.

3.2.10. Django vs Flask

Para elegir el lenguaje a usar para el back end se realizó un estudio de lenguajes apropiados para usar con un JavaScript web *framework*.

Flask

Flask es un micro *web framework* escrito en Python. Se clasifica como micro porque no requiere herramientas o bibliotecas particulares.

No tiene capa de abstracción de base de datos, validación de formularios ni ningún otro componente donde las bibliotecas de terceros preexistentes brinden funciones comunes. Sin embargo, Flask admite extensiones que pueden agregar características de la aplicación como si se implementaran en el propio Flask.

Existen extensiones para mapeadores relacionales de objetos, validación de formularios, manejo de carga, varias tecnologías de autenticación abiertas y varias herramientas relacionadas con el marco común. Las extensiones se actualizan con mucha más frecuencia que el programa central Flask.

Ventajas y desventajas de Flask Está basado en la especificación WSGI de Werkzeug y el motor de templates Jinja2; además, tiene una licencia BSD.

Entre las ventajas y desventajas, destacamos:

Ventajas

1. Es un framework que se destaca en instalar extensiones o complementos de acuerdo al tipo de proyecto que se va a desarrollar, es decir, es perfecto para el prototipado rápido de proyectos.
2. Incluye un servidor web, así podemos evitamos instalar uno como Apache o Nginx. Además, nos ofrece soporte para pruebas unitarias y para Cookies de seguridad (sesiones del lado del cliente), apoyándose en el motor de plantillas Jinja2.

-
3. Su velocidad es mejor a comparación de Django. Generalmente el desempeño que tiene Flask es superior debido a su diseño minimalista que tiene en su estructura.
 4. Flask permite combinarse con herramientas para potenciar su funcionamiento, por ejemplo: Jinja2, SQLAlchemy, Mako y Peewee entre otras.

Desventajas

1. Su sistema de autenticación de usuarios es muy básico, a comparación del potente sistema de autenticación que utiliza Django, este puede crear un sistema de Login API sencillo para aplicaciones más pequeñas.
2. Su representación de Plugins no es tan extensa como la tiene Django.
3. Es complicado en las pruebas unitarias o migraciones.
4. El ORM (Mapeo objeto relacional) para conectar con las bases de datos, SQLAlchemy es externo.

Django

De acuerdo con Wikipedia[43], Django es un framework de desarrollo web de código abierto, escrito en Python, que respeta el patrón de diseño conocido como MVC (Modelo–Vista–Controlador).

Características

1. Aplicaciones “enchufables” que pueden instalarse en cualquier página gestionada con Django.
2. Una API de base de datos robusta.
3. Un sistema incorporado de “vistas genéricas” que ahorra tener que escribir la lógica de ciertas tareas comunes.
4. Un sistema extensible de plantillas basado en etiquetas, con herencia de plantillas.
5. Un despachador de URL basado en expresiones regulares.
6. Un sistema “middleware” para desarrollar características adicionales.
7. Documentación incorporada accesible a través de la aplicación administrativa.

Elección

Se ha elegido Flask por ser un *framework* web conocido por el equipo y que ha dado resultados.

3.3. Conclusiones

1. Tomando en cuenta la experiencia del equipo con CSS, además de que el proyecto no se enfocará en hacer muchas hojas de estilo para cada componente, página o vista de la aplicación web, se usará CSS en lugar de Sass.
2. De acuerdo con el sitio stack overflow[27], JavaScript es el lenguaje más popular de 2019 y aunque TypeScript es de los lenguajes que tienen un mayor nivel de aceptación, se usará JavaScript no solo por ser el lenguaje más popular y, en consecuencia, con más compatibilidad y material de ayuda, sino también porque el equipo está acostumbrado a este lenguaje y tiene experiencia con web *frameworks* escritos en JavaScript.
3. Se usará Vue/Nuxt por ser el *framework* con el que el equipo está más acostumbrado, además de ser el más flexible de las opciones expuestas.
4. Se ha elegido usar en una primera instancia Vuetify porque es un CSS *framework* que está integrado en las tecnologías asociadas de Vue, como Vue Router, Vue Meta. Asimismo, sus componentes son simples de entender y de implementar.
5. De acuerdo con Mosquera[6], MongoDB es la base de datos NoSQL orientada a documentos más popular y usada en los *papers* de investigación. Por ello se usará MongoDB como base de datos.
6. Se llevó a la práctica en el prototipo funcional las tres opciones antes expuestas junto con algunas otras y se decidió usar GoJS para el proyecto por ser la biblioteca JavaScript más completa para diagramado y que genere los diagramas con un JSON simple para parsear esos datos y realizar la conversión.
7. Se ha elegido Python para desarrollar los algoritmos del proyecto dado que es multiplataforma y es de fácil integración con el *framework* web elegido para el back end.
8. Se ha elegido Flask por ser un *framework* web conocido por el equipo y que ha dado resultados.

Capítulo 4

Análisis y Diseño del Sistema

Aquí falta poner: El capítulo está organizado de la siguiente manera blablabla y tiene tal, tal y tal.

4.1. Metodología

De acuerdo a la Universidad Católica los Ángeles[44], en el campo del desarrollo de *software* hay dos grupos de metodologías, las tradicionales y las ágiles.

Las tradicionales suelen ser rígidas en cuanto a la documentación, exigiendo que sea exhaustiva y centrándose en cumplir con un plan de trabajo establecido en la etapa inicial del proyecto, mientras que las ágiles permiten realizar cambios en los requerimientos conforme a los avances del mismo.

Dado que cualquier cambio en el proceso de una metodología tradicional genera la necesidad de una reconstrucción del plan de trabajo y requiere invertir tiempo de desarrollo, surgieron las llamadas metodologías ágiles que permiten realizar cambios en los requerimientos conforme a los avances en el proyecto, tomando en cuenta la habilidades del equipo de desarrollo y una relación con el cliente; esta forma de trabajo permite mostrar avances funcionales en el producto en un periodo de tiempo corto para realizar una evaluación y, en caso de ser requerido, se sugieran cambios.

Se han propuesto muchos modelos ágiles de proceso y están en uso en toda la industria. Entre ellos se encuentran los siguientes:

- Desarrollo adaptativo de software (DAS).
- Scrum.
- Método de desarrollo de sistemas dinámicos (MDSD).
- Cristal.
- Desarrollo impulsado por las características (DIC).
- Desarrollo esbelto de software (DES).
- Modelado ágil (MA).

-
- Proceso unificado ágil (PUA).

4.1.1. Scrum

De acuerdo con *The Scrum Guide*[45], Scrum es un marco de trabajo para la entrega de productos incrementales y de máximo valor productivo.

Un artefacto es un elemento que garantiza la transparencia, es el registro de la información fundamental del proceso Scrum y a continuación se describen los 4 artefactos principales de Scrum:

- Lista de producto (*product backlog*): es el listado de todas las tareas que necesita el proyecto para alcanzar su realización. Al iniciar el desarrollo del proyecto esta lista no se encuentra completa y conforme avanzan los *sprints* se le añaden tareas para solventar las necesidades que van surgiendo gracias a la retroalimentación del cliente.
- Lista de pendientes del *sprint* (*sprint backlog*): es la lista de tareas seleccionadas del product backlog que se planifica realizar durante el periodo del *sprint* y se definen a los responsables de cada tarea.
- *Sprint*: es el corazón de Scrum; es de un periodo de tiempo determinado de un mes o menos, donde el equipo completa conjuntos de tareas incluidas en el *backlog* para crear un incremento del producto utilizable.
- Incremento: es la suma de todos los elementos de la lista de producto completados durante un *sprint* unido con los incrementos de los *sprints* anteriores. Al finalizar el *sprint* el nuevo incremento debe estar en condiciones de ser utilizado.

La metodología nos permite definir un periodo de hasta un mes para cada *sprint* y se ha optado por un periodo de 30 días, contemplándose un total de ocho sprints, donde al término de cada uno se tendrá un avance del sistema.

Asimismo, Scrum tiene cuatro eventos formales contenidos dentro del *sprint* para la inspección y adaptación del producto que se describen a continuación:

- Planificación del *sprint* (*sprint planning*): es una reunión con todo el equipo Scrum con una duración máxima de 8 horas para el *sprint* de un mes en la que el Scrum master es el encargado de que los asistentes entiendan el propósito de dicha reunión.
- Scrum diario (*daily scrum*): es una reunión de máximo 15 minutos en la cual el equipo de desarrollo expone sus actividades y planifica las tareas de las próximas 24 horas.
- Revisión del *sprint* (*sprint review*): al finalizar cada *sprint* se lleva a cabo una reunión para la revisión del incremento del producto y en caso de ser necesario realizar ajustes a la lista de producto.

-
- Retrospectiva del *sprint* (*sprint retrospective*): es el momento para el equipo Scrum de pensar en mejoras para el próximo *sprint* que contribuyan al proyecto.

El equipo Scrum consiste en los siguientes roles:

- El dueño de producto (*product owner*): es la persona responsable de maximizar el valor del producto y el trabajo del equipo de desarrollo; es la única responsable de gestionar la lista de producto y cualquier cambio a esa lista debe ser revisada y aprobada por él.
- El equipo de desarrollo (*development team*): es el conjunto de profesionales que realizan el trabajo para la entrega de un incremento en producto en cada *sprint*; son un grupo autoorganizado y multifuncional donde cada miembro del equipo tiene habilidades especializadas pero que la responsabilidad de las tareas completadas, incrementos del producto o retrasos recaen en el equipo como un todo.
- El Scrum master: es la persona responsable de asegurar que Scrum es entendido y adoptado por todos los involucrados en el proyecto, asegurándose de ayudar a las personas externas al equipo Scrum a entender qué interacciones con el equipo pueden ser útiles y cuáles no.

4.1.1.1. Historias de usuario

De acuerdo con Scrum México[46], las historias de usuario son la técnica por la que el usuario especifica de manera general los requerimientos que el sistema debe cumplir.

Normalmente estas redacciones se llevan a cabo en tarjetas de papel donde se describen brevemente las funciones que el producto final debe poseer sean o no requisitos funcionales.

El tratamiento de las historias de usuario es flexible y dinámico, cada una de ellas es lo suficientemente detallada y delimitada para que el equipo de desarrollo pueda implementarla durante la duración del *sprint*.

Es habitual que se siga una plantilla para estas tarjetas como la siguiente:

- Como <Usuario>
- Quiero <algún objetivo>
- Para <motivo>

Una de sus grandes ventajas dado el caso de que un usuario no sea lo suficientemente detallista con la historia, esta se puede partir en historias más pequeñas antes de que el equipo empiece a trabajar en ella.

Un ejemplo de historia de usuario para el desarrollo sería:

-
- Como usuario
 - Quiero poder ingresar al sistema con mi correo y contraseña
 - Para tener acceso a sus funciones

Otra forma de darle detalle a las historias de usuario es mediante el añadido de un criterio de aceptación. Un criterio de aceptación es una prueba que será cierta cuando el equipo de desarrollo complete la descripción de la tarjeta.

A continuación se listan las principales historias de usuario que se consideraron para el desarrollo de la propuesta de solución; tenga en cuenta que algunas de ellas tienen criterios de aceptación pero en otras no se consideró necesarias porque son explícitamente claras.

-
- No. 1
 - Como usuario
 - Quiero poder ingresar al sistema con mi correo y contraseña
 - Para tener acceso a sus funciones
 - Criterios de aceptación
 - El usuario recibirá un correo electrónico de confirmación de su alta en el sistema con el correo y contraseña que ingresó para tenerlos de respaldo.

-
- No. 2
 - Como usuario
 - Quiero poder recuperar mi contraseña en caso de olvidarla
 - Para no perder el trabajo realizado en el sistema
 - Criterios de aceptación
 - El usuario podrá ingresar una nueva contraseña, siempre y cuando recuerde el correo electrónico con el que se dio de alta en el sistema.
 - Al ingresar una nueva contraseña, recibirá un correo de confirmación del cambio de contraseña y sus datos permanecerán intactos.

-
- No. 3
 - Como usuario del sistema quiero darme de alta con una contraseña facil de recordar
-

-
- Pero que esté segura en la base de datos,
 - Para no tener comprometidos los diagramas que yo pueda generar en el sistema.
 - Criterios de aceptación
 - Asegurarse que usuario ingrese una contraseña de al menos 8 caracteres.
 - Se le solicitará al usuario que ingrese 2 veces la misma contraseña para asegurarse que le es fácil recordarla y que efectivamente es la misma.
 - Antes de guardar la contraseña, esta deberá pasar por un método que la haga ilegible para el usuario (algún algoritmo de digestión o cifrado).
-

- No. 4
 - Como usuario quiero poder crear un diagrama ER arrastrando y soltando elementos de una “paleta”
 - Para hacerlo de manera mas fácil e intuitiva.
 - Criterios de aceptación
 - El usuario podrá empezar un nuevo diagrama al seleccionar la opción de “diagramador ER”.
 - Tendrá a su disposición una paleta con los elementos permitidos en un diagrama ER estándar.
 - Podrá arrastrar y soltar los elementos de la paleta a un área delimitada para empezar con el diseño de su diagrama.
-

- No. 5
 - Como usuario quiero poder guardar mi último trabajo realizado en el diagramador ER.
 - Para poder consultarlo en otro momento.
 - Criterios de aceptación
 - Dispondrá de un botón para poder guardar en la base de datos el diagrama que esté creando/editando.
 - Antes de almacenar el diagrama en el *canvas* o zona de diagramado, se le mostrará un mensaje de confirmación para guardar su diagrama actual.
-

- No. 6
-

-
- Como usuario me gustaría poder ver mi último trabajo que realicé
 - Cuando seleccione la opción “Entidad-Relación”
 - Para poder modificar el diseño.
-

- No. 7
 - Como usuario quiero tener la opción de cargar un diagrama a partir de un archivo
 - Para hacer modificación de dicho diagrama y guardarlo de ser necesario.
 - Criterios de aceptación
 - El usuario tendrá un botón “cargar.”^{en} el menú del diagramador ER para poder importar un archivo con extensión .json
 - Al importar el archivo, este pasará por un proceso de validación para asegurarse que es un archivo JSON válido.
 - Durante el proceso de validación se verificará que el contenido del archivo es un diagrama compatible con la estructura de los generados por el diagramador ER.
 - Al contener una información compatible, se mostrará en la zona de diagramado el contenido del archivo.
-

- No. 8
 - Como usuario quiero poder descargar el diagrama que esté visible en la página web
 - Para poder distribuirlo como yo desee.
 - Criterios de aceptación
 - El usuario dispondrá de un botón “Descargar.”^{en} el diagramador ER para obtener un archivo con el contenido del diagrama visible en la zona de diagramado.
 - El archivo generado será de extensión .json con la información necesaria para que el diagramador pueda cargarlo en otro momento.
-

- No. 9
 - Como usuario quiero tener una forma de validar mi diagrama ER
-

-
- Por que es importante saber si el diagrama que estoy creando es un diagrama válido del modelo ER.
 - Criterios de aceptación
 - El usuario tendrá disponible un botón que al darle click iniciará un proceso de validación del diagrama actual de la zona de diagramado.
 - Al término del proceso de validación se le mostrará un mensaje al usuario indicando si el diagrama cumple o no las reglas del modelo ER.
-

- No. 10
 - Como usuario en caso de tener un diagrama ER válido
 - Me gustaría poder tranformar el diagrama ER en su versión del modelo relacional.
 - Para poder ver el equivalente del diagrama ER en el modelo Relacional.
 - Criterios de aceptación
 - El usuario dispondrá de la opción de tranformar al modelo relacional solamente despues de haber validado que el diagrama ER cumple con las reglas.
 - Posterior a la validación se le mostrará al ususario en mensaje de confirmación y un botón para disparar el proceso de transformación a su equivalente relacional.
 - Al terminar el proceso de transformación equivalente se le redirijira al menú “Relacional”donde podrá visualizar el equivalente al modelo relacional.
-

- No. 11
 - Como usuario después de observar el diagrama relacional
 - Quiero poder obtener las sentencias SQL equivalentes
 - Para poder crear el esquema de base de datos relacional en un SBGB
 - Criterios de aceptación
 - Las sentencias SQL solo podrán ser descargadas en el menú “Relacional.”^a un archivo con extension .sql dando click a un botón con la leyenda “Descargar SQL”.
 - Solo podrán ser obtenidas las sentencias SQL de un diagrama ER creado y/o validado por el sistema.
-

-
- No. 12
 - Como usuario quiero poder transformar mi diagrama ER en su equivalente modelo concepcual NoSQL
 - Para poder observar el cambio entre modelos.
 - Criterios de aceptación
 - El usuario dispondrá de la opción de tranformar al modelo NoSQL solamente después de haber validado que el diagrama ER cumple con las reglas.
 - Posterior a la validación se le mostrará al usuario un mensaje de confirmación y un botón para disparar el proceso de transformación al modelo conceptual NoSQL.
 - Una vez validado el diagrama ER e iniciado el proceso para la transformación al modelo conceptual NoSQL se le indicará al usuario que el proceso puede ser tardado.
 - Al término del proceso de transformación se le redirigirá al menú “No Relacional” donde observará el modelo conceptual NoSQL equivalente a su diagrama ER.
-

- No. 13
 - Como usuario quiero poder obtener el *script* desde el modelo conceptual NoSQL
 - Para poder generar la base de datos en un gestor de base de datos NoSQL orientado a documentos
 - Criterios de aceptación
 - El usuario dispondrá de la opción de obtener el *script* solamente después de haber validado que el diagrama ER cumple con las reglas.
 - Posterior a la validación se le mostrará al usuario un mensaje de confirmación y un botón para disparar el proceso de generación de *scripts* para el gestor de base de datos orientado a documentos.
 - Una vez empezado el proceso de generación de *scripts* se le indicará al usuario que el proceso puede ser tardado.
 - Al término del proceso de transformación se le redirigirá al menú “No Relacional” donde observará los *scripts* NoSQL.
-

- No. 14
-

-
- Como usuario me gustaría tener un reporte técnico y
 - Quiero que la redacción sea legible y referenciada
 - Para compartirlo en el futuro con equipos de desarrollo y ver la posibilidad de agregar nuevas funciones al sistema.
-

Teniendo en cuenta que se está trabajando con una metodología ágil, estas historias de usuario pueden aumentar o en su defecto dividirse en historias más pequeñas dependiendo de los criterios del equipo de desarrollo durante el proceso de la implementación de cada historia.

4.1.1.2. Lista de producto (*product backlog*)

De acuerdo a Trigas Gallego[47], la lista de producto es una lista ordenada de todo lo que podría ser necesario en el producto y es la fuente de requisitos para cualquier cambio a realizarse en el producto, enumera todas las características, funcionalidades, requisitos, mejoras y correcciones que constituyen cambios a realizarse sobre el producto para entregas futuras.

Se consideran la siguiente lista como la lista de producto con las tareas necesarias para cumplir con todas las historias de usuario mencionadas en la sección anterior.

Considerando que estas pueden cambiar conforme avancen los *sprints* y añadirse nuevas tareas.

No. Historia de Usuario	requerimiento/Tarea	Responsable
14	Investigación Bases de datos relacionales.	Eduardo/Omar
14	Redacción y selección de las tecnologías a utilizar para el desarrollo de la plataforma.	Eduardo
14	Investigación Bases de datos relacionales.	Eduardo/Omar
14	Redacción bases de datos relacionales en el documento técnico.	Eduardo
14	Investigación Bases de datos no relacionales.	Eduardo/Omar
14	Redacción bases de datos no relacionales en el documento técnico.	Eduardo
14	Investigación y selección del modelo de base de datos no relacional a utilizar junto a las tecnologías a utilizar.	Eduardo/Omar
14	Análisis y diseño de la arquitectura web.	Eduardo/Omar
1	Desarrollo de la estructura básica del <i>backend</i> .	Omar

Continuación de tabla de lista de producto

1	Desarrollo de la estructura básica del <i>frontend</i> .	Eduardo
1	Agregar servicio <i>backend</i> para registrar un usuario.	Omar
1	Agregar formulario para captura de datos de registro de un usuario en el <i>frontend</i> .	Eduardo
2	Agregar servicio <i>backend</i> para recuperar contraseña del usuario.	Omar
2	Agregar servicio <i>backend</i> para envío de email al usuario registrado y de recuperación de contraseña.	Omar
2	Agregar vista con formulario para recuperación de contraseña del usuario en el <i>frontend</i> .	Eduardo
2	Integración de los servicios de registro y recuperación de contraseña en el <i>frontend</i> .	Eduardo
3	Agregar servicio <i>backend</i> para hacer ilegible la contraseña del usuario en la base de datos.	Omar
4	Planteamiento de escenarios de los esquemas entidad-relación.	Eduardo
4	Agregar a la interfaz gráfica de la aplicación web el menú “Entidad-Relación”.	Eduardo
4	Agregar iconos de los elementos básicos de un diagrama ER en el diagramador.	Eduardo
5	Agregar servicio <i>backend</i> para guardar un diagrama ER en formato JSON en la base de datos e integrarlo al <i>frontend</i> .	Omar
5	Agregar servicio <i>backend</i> para recuperar el diagrama guardado del usuario de la base de datos y regresarlo en formato JSON.	Omar
6	Recuperar el último diagrama del usuario del <i>backend</i> y mostrarlo en el <i>frontend</i> .	Eduardo
6	Manejar el estado de la interfaz web para no perder el diagrama ER que está editando el usuario.	Eduardo
6	Definición de las reglas del modelo entidad-relación.	Eduardo
4	Implementar la edición de diagramas ER en el <i>frontend</i> .	Eduardo
7	Habilitar la carga de un archivo en la aplicación web.	Omar
7	Agregar la función para validar el contenido del archivo JSON y pintarlo en la zona de diagramado.	Eduardo
8	Agregar la descarga del diagrama visible en la zona de diagramado a un archivo JSON.	Eduardo
9	Agregar botón de validar al <i>frontend</i> y mostrar el <i>loader</i> mientras se procesa el diagrama ER.	Eduardo/Omar
9	Agregar servicio <i>backend</i> para la validación del diagrama entidad-relación.	Eduardo/Omar

Continuación de tabla de lista de producto

9	implementación de algoritmo para validación del diagrama ER en el <i>backend</i> .	Eduardo/Omar
9	Pruebas de captura de distintos diagramas entidad-relación.	Eduardo/Omar
9	Pruebas para validar el algoritmo de validación.	Eduardo/Omar
10	Agregar servicio al <i>backend</i> para transformación del esquema entidad-relación al modelo relacional.	Omar
10	Implementación del algoritmo de transformación ER -> relacional	Eduardo/Omar
10	Agregar menú relacional a la interfaz gráfica.	Eduardo/Omar
10	Prueba de transformación de distintos diagramas E-R al modelo relacional.	Eduardo/Omar
10	Visualización de la transformación del modelo ER al modelo Relacional.	Eduardo
14	Revisión de la redacción del reporte técnico para presentación de TT1	Eduardo
11	Agregar servicio <i>backend</i> para la descarga del archivo .sql con las sentencias equivalentes.	Eduardo/Omar
11	Pruebas de coherencia de las sentencias equivalentes de distintos en el SGBD.	Eduardo/Omar
12	Definición de las reglas de transformación al modelo NoSQL.	Eduardo/Omar
12	Pruebas de distintos escenarios del modelo relacional al modelo NoSQL.	Eduardo/Omar
12	Agregar servicio al <i>backend</i> para transformación del esquema relacional al modelo NoSQL.	Omar
12	Agregar servicio al <i>backend</i> para guardar el modelo NoSQL en la base de datos.	Omar
12	Agregar menú no relacional a la interfaz gráfica.	Omar
12	Implementación del algoritmo de transformación de modelo relacional al modelo conceptual NoSQL.	Eduardo
12	Comprobación de la coherencia de la transformación entre modelos relacional a no relacional.	Eduardo/Omar
13	Agregar servicio <i>backend</i> para transformación del modelo ER al modelo no relacional.	Eduardo/Omar
13	Ajustar la interfaz del menú ER para mostrar mensaje transformación la modelo NoSQL.	Omar
13	Manejar el estado del diagrama ER y redireccionar al menú no relacional al terminar la transformación.	Eduardo
13	Pruebas de caso de estudio para verificar la correcta transformación y coherencia en los datos.	Eduardo/Omar
14	Revisión de la redacción del reporte técnico para presentación de TT2	Eduardo

Tabla 4.1: Lista de producto

4.2. Factibilidad técnica

Para la factibilidad técnica se realiza una evaluación de las herramientas de hardware y software que el equipo de trabajo tiene disponible.

Recursos técnicos de software

Equipo	Tipo	Sistema operativo	Procesador	RAM
Toshiba Satellite Radius P55W-B	Laptop	Windows 10	Intel Core i7-10U	8GB
Dell Laptop Latitude 3400	Laptop	Windows 10	Intel Core I5 8265	8GB
HP Pavilion15"	Laptop	Windows 10	Core i5-9300H	8GB
DeskJet Ink Advantage 3775	Impresora Multifuncional	-	-	-

Recursos técnicos de hardware

Nombre	Tipo
Windows 10	Sistema operativo
Ubuntu	Sistema operativo
MongoDB Atlas	Base de datos en la nube
Visual Studio Code	IDE
Flask	Framework
Vue.js	Framework
Nuxt	Framework

Con todo esto, se puede concluir que se cuenta con todo el software y hardware necesario para la realización del proyecto.

4.3. Factibilidad económica

Estimación

Para la mayoría de los proyectos, el mayor costo es el primer rubro. Debe estimarse el esfuerzo total (en meses-hombre) que es probable se requiera para completar el trabajo de un proyecto. Desde luego, se cuenta con datos limitados para realizar tal valoración, de manera que habrá que hacer la mejor evaluación posible y a continuación agregar contingencia significativa (tiempo y esfuerzo adicionales) en caso de que la estimación inicial sea optimista.

4.3.1. COCOMO II

El modelo COCOMO (por sus acrónimo en inglés COnstructive COnst MOdel) original se convirtió en uno de los modelos de estimación de costo más ampliamente utilizados y estudiados en la industria. Evolucionó hacia un modelo de estimación más exhaustivo, llamado COCOMO II. Como su predecesor, COCOMO II en realidad es una jerarquía de modelos de estimación que aborda las áreas siguientes: [Roger S. Pressman, Ph.D. **Ingeniería del software UN ENFOQUE PRÁCTICO** 7 edición, ed. McGRAW-HILL]

1. **Modelo de composición de aplicación.** Se usa durante las primeras etapas de la ingeniería de software, cuando son primordiales la elaboración de prototipos de las interfaces de usuario, la consideración de la interacción del software y el sistema, la valoración del rendimiento y la evaluación de la madurez de la tecnología.
2. **Modelo de etapa temprana de diseño.** Se usa una vez estabilizados los requisitos y establecida la arquitectura básica del software.
3. **Modelo de etapa postarquitectónica.** Se usa durante la construcción del software.

Como todos los modelos de estimación para software, los modelos COCOMO II requieren información sobre dimensionamiento. Como parte de la jerarquía del modelo, están disponibles tres diferentes opciones de dimensionamiento: puntos objeto, puntos de función y líneas de código fuente.

4.3.1.1. El modelo de composición de aplicación COCOMO II

El esfuerzo necesario para concretar un proyecto de desarrollo de software, cualquiera sea el modelo empleado, se expresa en meses/persona (PM) y representa los meses de trabajo de una persona fulltime, requeridos para desarrollar el proyecto. Para derivar una estimación del esfuerzo debemos utilizar la siguiente relación.

$$Esfuerzo_{estimado} = \frac{NOP}{PROD}$$

donde:

- Esfuerzo estimado (PM) es la estimación del esfuerzo en meses-hombre.
- NOP es el número de puntos de aplicación en el sistema entregado.
- PROD es la productividad del punto de aplicación

Para determinar NOP se necesita utilizar el punto de objeto el cual una medida de software indirecta que se calcula usando conteos del número de:

1. pantallas (en la interfaz de usuario)

2. reportes
3. componentes que probablemente se requieran para construir la aplicación.

Cada instancia de objeto (por ejemplo, una pantalla o reporte) se clasifica en uno de tres niveles de complejidad (simple, medio o difícil), usando criterios sugeridos por Boehm [Boehm, B., "Anchoring the Software Process", IEEE Software, vol. 13, núm. 4, julio 1996, pp. 73-82]

Tipo de objeto	Cantidad	Peso de complejidad			Subtotal
		Simple	Media	Difícil	
Pantalla	4	1	2	<u>3</u>	12
Reporte	9	10	<u>5</u>	8	45
Componente	10			10	100
Total OP					157

Determinado el conteo de puntos de objeto se multiplican el número original de instancias de objeto por el factor de ponderación y se suman para obtener un conteo total de puntos de objeto.

$$NOP = (Puntos_de_objeto) \left(\frac{100 - \%reutilizacion}{100} \right)$$

Dentro de este punto, no se encuentra código el cual reutilizar, por lo que:

$$NOP = (Puntos_de_objeto)$$

$$NOP = (157)$$

Para la tasa de productividad se utiliza la figura [referenciar tabla], para diferentes niveles de experiencia del desarrollador y de madurez del entorno de desarrollo.

Experiencia del desarrollador	Muy baja	Baja	Nominal	Alta	Muy alta
Madurez/ capacidad del entorno	Muy baja	Baja	Nominal	Alta	Muy alta
PROD	4	7	10	25	50

Una vez determinados los nuevos puntos de objeto y la tasa de productividad, se puede calcular una estimación del esfuerzo del proyecto

$$Esfuerzo_{estimado} = \frac{NOP}{PROD}$$

$$Esfuerzo_{estimado} = \frac{157}{7}$$

$$Esfuerzo_{estimado} = 22.42$$

$$Esfuerzo_{estimado} = 23 \text{ Meses} - \text{hombre}$$

Una vez obtenidos estos valores, se pueden sumar los costos derivados del desarrollo de software

Factibilidad económica

Dentro de la factibilidad económica se encuentran los costos necesarios que llevará consigo el desarrollo de la aplicación.

Costo de hardware

Costo	Equipo
\$10,000.00	Toshiba Satellite Radius P55W-B, Windows 10, Intel Core i7-10U, 8GB RAM
\$13,999.00	Dell Laptop Latitude 3400, Ubuntu, Intel Core I5 8265, 8GB RAM.
\$16,499.00	HP Pavilion15”, Windows 10, Core i5-9300H, 8GB RAM.
\$1,599.00	DeskJet Ink Advantage 3775

Nombre	Costo
Windows 10	\$3,599.00
Ubuntu	Gratuito
MongoDB Atlas	
Visual Studio Code	Gratuito
Flask	Software libre
Vue.js	Software libre
Nuxt	Software libre

Es necesario detallar que el costo de la herramienta no será determinado puramente con los costos de software y hardware descritos dentro de esta sección, ya que son requeridos valores tales como el tiempo de desarrollo y el esfuerzo requeridos que serán especificados gracias al Modelo Constructivo de Costos (COCOMO II)

Gastos de operación

Gasto de operación	Gasto mensual (MXN)	Gasto total (MXN)
Luz	200	
Agua	200	
Internet	499	
Plan de datos	299	
Papelería	150	
Recursos humanos		

4.4. Factibilidad operativa

Factibilidad operativa

La herramienta va dirigida hacia cualquier estudiante que se encuentre trabajando con modelos de bases de datos desde un enfoque conceptual como lo requieren las primeras etapas del aprendizaje sobre el tema, buscando principalmente lograr un impacto en los estudiantes que comienzan a tener una aproximación a los modelos no relacionales de bases de datos.

Se cuenta con una interfaz intuitiva para que el estudiante pueda visualizar la información de una manera ya conocida, como lo son los diagramas E-R y las opciones que ésta les brinde de manera comprensible, no requiriendo de un usuario o personal especializado para su utilización. El usuario contará con registro ligado a su cuenta de correo electrónico que le permitirá acceder a la herramienta de forma única brindándole un almacenamiento de los proyectos que con esta puedan haberse realizado.

Dentro de la herramienta se encuentran tres funciones principales, en un desarrollo primario se cuenta con un diagramador entidad-relación el cual permite una visualización abstracta de la construcción de una base de datos que apoya en su correcta descripción gracias a mensajes sobre la conexión de elementos por medio de la validación del mismo.

El segundo bloque podrá recibir el diagrama entidad-relación anteriormente tratado, o bien, trabajar sobre uno cuyo desarrollo y carga se haya realizado con antelación. Posteriormente el diagrama podrá ser transformado al modelo relacional con la posibilidad de obtener el esquema de la base de datos en sentencias SQL. Finalmente si es requerido, se obtendrá el esquema de base de datos en un modelo de datos no relacional y tendrá la posibilidad de implementar el modelo no relacional a un SGBDNR (Sistema Gestor de Base de Datos No Relacional).

Basándonos en las necesidades y tecnologías mencionadas a utilizar para el desarrollo de la herramienta, dentro del personal se deberá contar con un perfil como se describe a continuación:

Analista de sistemas:

Se entiende por analista a la persona que trabaja con los requerimientos, en el diseño global y en el diseño detallado. Los principales atributos que deberían considerarse en un analista son:

- Liderazgo
- Dedicación
- Habilidad para el diseño,
- Análisis de requerimientos
- Correcta comunicación
- Cooperación entre sus pares.

El salario promedio para un puesto de Analista de sistema en México es de \$11,708 al mes. Las estimaciones de salarios se basan en 564 salarios que empleados y usuarios que trabajan de Analista de sistema enviaron a Indeed de forma anónima, y en los salarios que recopilamos de los anuncios de empleo que se publicaron en Indeed en los últimos 36 meses. [<https://www.indeed.com.mx/salaries/analista-de-sistema-Salaries>]

Desarrollador

Las tendencias actuales siguen enfatizando la importancia de la capacidad de los analistas. Sin embargo, debido a que la productividad se ve afectada notablemente por la habilidad del programador en el uso de las herramientas actuales, existe una tendencia a darle mayor importancia a la capacidad del programador.

- **Front end.**

El desarrollador front-end es un especialista encargado de diseñar la interfaz de usuario de los sitios web, es decir, es el encargado de traducir las definiciones de diseño y estilo visual realizadas en etapas previas a código semántico. Diseña la estructura, la tipografía, la colorimetría, imágenes, banners, etc.

El salario promedio de Front End en México es de \$27.500-Mes Basado en 777 salarios [<https://neuvo.com.mx/salario/?job=front+end>]

- **Back end.**

El desarrollador back-end es el encargado de implementar la interacción entre el usuario y el sitio web. Se trata del responsable de la programación de un sitio web y todos sus componentes, coordinando páginas, formularios, funcionalidades, bases de datos y servidores web, evitando problemas en las capas más profundas del proyecto.

El salario promedio de Back End en México es de \$29.000-Mes Basado en 429 salarios [<https://neuvo.com.mx/salario/?job=back+end>]

Capítulo 5

Algoritmos

5.1. Validaciones entidad-relación

5.2. Modelo entidad-relación a relacional

Elmasri[13] propone un algoritmo en siete pasos para convertir las estructuras de un modelo ER básico con tipos de entidades fuertes y débiles, relaciones binarias con distintas restricciones estructurales, relaciones n-ary y atributos (simples, compuestos y multivalor) en relaciones.

Paso 1: Mapeado de los tipos de entidad regulares

Por cada entidad (fuerte) regular E del esquema ER cree una relación R que incluya todos los atributos simples de E .

Incluya únicamente los atributos simples que conforman un atributo compuesto; seleccione uno de los atributos clave de E como clave principal para R .

Si la clave elegida de E es compuesta, el conjunto de los atributos simples que la forman constituirán la clave principal de R .

Si durante el diseño conceptual se identificaron varias claves para E , la información que describe los atributos que forman cada clave adicional conserva su orden para especificar las claves (únicas) secundarias de la relación R .

El conocimiento sobre las claves también es necesario para la indexación y otros tipos de análisis.

Paso 2: Mapeado de los tipos de entidad débiles

Por cada tipo de entidad débil W del esquema ER con el tipo de entidad propietario E , cree una relación R e incluya todos los atributos simples (o componentes simples de los atributos compuestos) de W como atributos de R .

Además, incluya como atributos de la *foreign key* de R , los atributos de la o las

relaciones que correspondan al o los tipos de entidad propietarios; esto se encarga de identificar el tipo de relación de W .

La clave principal de R es la combinación de las claves principales del o de los propietarios y la clave parcial del tipo de entidad débil W si la hubiera.

Si hay un tipo de entidad débil E_2 cuyo propietario también es un tipo de entidad débil E_1 , E_1 debe asignarse antes que E_2 para determinar primero su clave principal.

Paso 3: Mapeado de los tipos de relación 1:1 binaria

Por cada tipo de relación 1:1 binaria R del esquema ER, identifique las relaciones S y T que corresponden a los tipos de entidad que participan en R . Hay tres metodologías posibles: (1) la metodología de la *foreign key*, (2) la metodología de la relación mezclada y (3) la metodología de referencia cruzada o relación de relación. La primera metodología es la más útil y la que debe seguirse salvo que se den ciertas condiciones especiales, como las que explicamos a continuación:

- 1 **Metodología de la *foreign key*:** seleccione una de las relaciones (por ejemplo, S) e incluya como *foreign key* en S la clave principal de T . Lo mejor es elegir un tipo de entidad con participación total en R en el papel de S . Incluya todos los tributos simples (o los componentes simples de los atributos compuestos) del tipo de relación 1:1 R como atributos de S .
- 2 **Metodología de la relación mezclada:** una asignación alternativa de un tipo de relación 1:1 es posible al mezclar los dos tipos de entidad y la relación en una sola relación. Esto puede ser apropiado cuando las dos participaciones son totales.
- 3 **Metodología de referencia cruzada o relación de relación:** consiste en configurar una tercera relación R con el propósito de crear una referencia cruzada de las claves principales de las relaciones S y T que representan los tipos de entidad. Como veremos, esta metodología es necesaria para las relaciones $M : N$ binarias. La relación R se denomina relación de relación (y, en algunas ocasiones, tabla de búsqueda), porque cada tupla de R representa una instancia de relación que relaciona una tupla de S con otra de T .

Paso 4: Mapeado de tipos de relaciones 1:N binarias

Por cada relación 1 : N binaria regular R , identifique la relación S que representa el tipo de entidad participante en el lado N del tipo de relación.

Incluya como *foreign key* en S la clave principal de la relación T que representa el otro tipo de entidad participante en R ; hacemos esto porque cada instancia de entidad en el lado N está relacionada, a lo sumo, con una instancia de entidad del lado 1 del tipo de relación.

Incluya cualesquiera atributos simples (o componentes simples de los atributos compuestos) del tipo de relación 1 : N como atributos de S .

De nuevo, una metodología alternativa es la opción de relación de relación (referencia cruzada), como en el caso de las relaciones 1:1 binarias.

Creamos una relación R separada cuyos atributos son las claves de S y T , y cuya clave principal es la misma que la clave de S . Esta opción puede utilizarse si pocas tuplas de S participan en la relación para evitar excesivos valores NULL en la *foreign key*.

Paso 5: Mapeado de tipos de relaciones M:N binarias

Por cada tipo de relación $M : N$ binaria R cree una nueva relación S para representar a R .

Incluya como atributos de la *foreign key* en S las claves principales de las relaciones que representan los tipos de entidad participantes; su combinación formará la clave principal de S .

Incluya también cualesquiera atributos simples del tipo de relación $M : N$ (o los componentes simples de los atributos compuestos) como atributos de S .

No podemos representar un tipo de relación $M : N$ con un atributo de *foreign key* en una de las relaciones participantes (como hicimos para los tipos de relación $1 : 1$ o $1 : N$) debido a la razón de cardinalidad $M : N$; debemos crear una relación de relación S separada.

Siempre podemos asignar las relaciones $1 : 1$ o $1 : N$ de un modo similar a las relaciones $M : N$ utilizando la metodología de la referencia cruzada (relación de relación), como explicamos anteriormente.

Esta alternativa es particularmente útil cuando existen pocas instancias de relación, a fin de evitar valores NULL en las *foreign keys*.

En este caso, la clave principal de la relación de relación sólo será una de las *foreign keys* que hace referencia a las relaciones de entidad participantes.

Para una relación $1 : N$, la clave principal de la relación de relación será la *foreign key* que hace referencia a la relación de la entidad en el lado N .

En una relación $1 : 1$, cada *foreign key* se puede utilizar como la clave principal de la relación de relación, siempre y cuando no haya entradas NULL en la relación.

Paso 6: Mapeado de atributos multivalor

Por cada atributo multivalor A , cree una nueva relación R .

Esta relación incluirá un atributo correspondiente a A , más el atributo clave principal K (como *foreign key* en R) de la relación que representa el tipo de entidad o tipo de relación que tiene A como un atributo.

La clave principal de R es la combinación de A y K . Si el atributo multivalor es compuesto, incluimos sus componentes simples.

Paso 7: Mapeado de los tipos de relación n-ary

Por cada tipo de relación n-ary R , donde $n > 2$, cree una nueva relación S para representar R .

Incluya como atributos de la *foreign key* en S las claves principales de las relaciones que representan los tipos de entidad participantes.

Incluya también cualesquiera atributos simples del tipo de relación n-ary (o los componentes simples de los atributos compuestos) como atributos de S .

Normalmente, la clave principal de S es una combinación de todas las *foreign keys* que hacen referencia a las relaciones que representan los tipos de entidad participantes.

No obstante, si las restricciones de cardinalidad en cualquiera de los tipos de entidad E que participan en R es 1, entonces la clave principal de S no debe incluir el atributo de la *foreign key* que hace referencia a la relación E' correspondiente a E .

5.3. Obtención de esquema SQL desde modelo relacional

5.4. Modelo entidad-relación a Generic Data Metamodel

5.5. Generic Data Metamodel a modelo lógico NoSQL

5.6. Modelo lógico NoSQL a modelo físico en MongoDB

Apéndice A

Apéndice

A.1. Unified Modeling Language

De acuerdo con Wikipedia[48], un diagrama de clases en Lenguaje Unificado de Modelado (UML) es un tipo de diagrama de estructura estática que describe la estructura de un sistema mostrando las clases del sistema, sus atributos, operaciones (o métodos), y las relaciones entre los objetos.

Miembros

UML proporciona mecanismos para representar los miembros de la clase, como atributos y métodos, así como información adicional sobre ellos.

Visibilidad

Para especificar la visibilidad de un miembro de la clase (es decir, cualquier atributo o método), se coloca uno de los siguientes signos delante de ese miembro:

1. + Público
2. - Privado
3. # Protegido
4. / Derivado (se puede combinar con otro)
5. Paquete

Ámbitos

UML especifica dos tipos de ámbitos para los miembros: instancias y clasificadores y estos últimos se representan con nombres subrayados.

-
1. Los miembros clasificadores se denotan comúnmente como “estáticos” en muchos lenguajes de programación. Su ámbito es la propia clase.
 - a) Los valores de los atributos son los mismos en todas las instancias.
 - b) La invocación de métodos no afecta al estado de las instancias.
 2. Los miembros instancias tienen como ámbito una instancia específica.
 - a) Los valores de los atributos pueden variar entre instancias.
 - b) La invocación de métodos puede afectar al estado de las instancias (es decir, cambiar el valor de sus atributos).

Para indicar que un miembro posee un ámbito de clasificador, hay que subrayar su nombre. De lo contrario, se asume por defecto que tendrá ámbito de instancia.

Relaciones

Una relación es un término general que abarca los tipos específicos de conexiones lógicas que se pueden encontrar en los diagramas de clases y objetos. UML presenta las siguientes relaciones:

Enlace

Un enlace es la relación más básica entre objetos.

Asociación

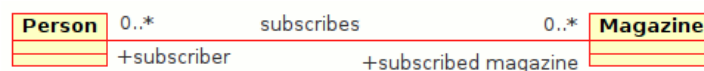


Figura A.1: Asociación

Una asociación representa a una familia de enlaces. Una asociación binaria (entre dos clases) normalmente se representa con una línea continua. Una misma asociación puede relacionar cualquier número de clases. Una asociación que relacione tres clases se llama asociación ternaria.

A una asociación se le puede asignar un nombre, y en sus extremos se puede hacer indicaciones, como el rol que desempeña la asociación, los nombres de las clases relacionadas, su multiplicidad, su visibilidad, y otras propiedades.

Hay cuatro tipos diferentes de asociación: bidireccional, unidireccional, agregación (en la que se incluye la composición) y reflexiva. Las asociaciones unidireccional y bidireccional son las más comunes.

Por ejemplo, una clase vuelo se asocia con una clase avión de forma bidireccional. La asociación representa la relación estática que comparten los objetos de ambas clases.

Agregación

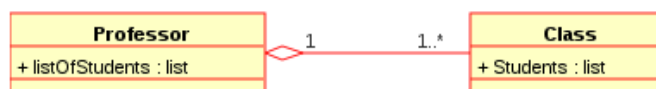


Figura A.2: Agregación

La agregación o agrupación es una variante de la relación de asociación “tiene un”: la agregación es más específica que la asociación. Se trata de una asociación que representa una relación de tipo parte-todo o parte-de.

Como se puede ver en la imagen del ejemplo (en inglés), un Profesor ‘tiene una’ clase a la que enseña.

Al ser un tipo de asociación, una agregación puede tener un nombre y las mismas indicaciones en los extremos de la línea. Sin embargo, una agregación no puede incluir más de dos clases; debe ser una asociación binaria.

Una agregación se puede dar cuando una clase es una colección o un contenedor de otras clases, pero a su vez, el tiempo de vida de las clases contenidas no tienen una dependencia fuerte del tiempo de vida de la clase contenedora (de el todo). Es decir, el contenido de la clase contenedora no se destruye automáticamente cuando desaparece dicha clase.

En UML, se representa gráficamente con un rombo hueco junto a la clase contenedora con una línea que lo conecta a la clase contenida. Todo este conjunto es, semánticamente, un objeto extendido que es tratado como una única unidad en muchas operaciones, aunque físicamente está hecho de varios objetos más pequeños.

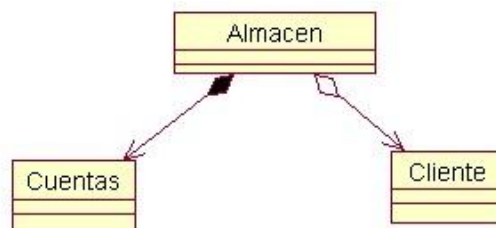


Figura A.3: Asociación rombo sin rellenar y composición rombo negro

Composición

La representación en UML de una relación de composición es mostrada con una figura de diamante relleno del lado de la clase contenedora, es decir al final de la línea que conecta la clase contenido con la clase contenedor.

Diferencias entre Composición y Agregación

Relación de Composición

-
1. Cuando intentamos representar un todo y sus partes. Ejemplo, un motor es una parte de un coche.
 2. Cuando se elimina el contenedor, el contenido también es eliminado. Ejemplo, si eliminamos una universidad eliminamos igualmente sus departamentos.

Relación de Agrupación

1. Cuando representamos las relaciones en un software o base de datos. Ejemplo, el modelo de motor MTR01 es parte del coche MC01. Como tal, el motor MTR01 puede hacer parte de cualquier otro modelo de coche, es decir si eliminamos el coche MC01 no es necesario eliminar el motor pues podemos usarlo en otro modelo.
2. Cuando el contenedor es eliminado, el contenido usualmente no es destruido. Ejemplo, un profesor tiene estudiantes, cuando el profesor muere los estudiantes no mueren con él o ella.

A.2. El modelo entidad-relación extendido

De acuerdo a la bibliografía de Catherine[\[49\]](#), el modelo entidad-relación extendido (EE-R) extiende el modelo ER para permitir la inclusión de varios tipos de abstracción, y para expresar restricciones más claramente. A los diagramas ER estándar se agregan símbolos adicionales para crear diagramas EE-R que expresen estos conceptos.

Especialización

Con frecuencia, un conjunto de entidades contiene uno o más subconjuntos que tienen atributos especiales o que participan en relaciones que otros miembros del mismo conjunto de entidades no tiene.

El método de identificar subconjuntos de conjuntos de entidades existentes, llamado especialización, corresponde a la noción de herencia de subclase y clase en el diseño orientado a objetos, donde se representa mediante jerarquías de clase.

El círculo que conecta a la superclase con las subclases se llama círculo de especialización. Cada subclase se conecta al círculo mediante una línea que tiene un símbolo de herencia, un símbolo de subconjunto o copa, con el lado abierto de frente a la superclase. Las subclases heredan los atributos de la superclase y opcionalmente pueden tener atributos locales distintos.

Dado que cada miembro de una subclase es miembro de la superclase, al círculo de especialización a veces se le conoce como relación isa.

En ocasiones una entidad tiene solo un subconjunto con propiedades o relaciones especiales de las que quiere tener información. Solo contiene una subclase para una especialización. En este caso, en el diagrama EE-R se omite el círculo y simplemente se muestra la subclase conectada mediante una línea de subconjunto a la superclase.

Las subclases también pueden participar en relaciones locales que no se apliquen a la superclase o a otras subclases en la misma jerarquía.

Generalización

Además de la especialización, también se pueden crear jerarquías de clase al reconocer que dos o más clases tienen propiedades comunes e identificar una superclase común para ellas, un proceso llamado generalización. Estos dos procesos son inversos uno de otro, pero ambos resultan en el mismo tipo de diagrama jerárquico.

Restricciones

Las subclases pueden ser traslapantes (*overlapping*), lo que significa que la misma instancia de entidad puede pertenecer a más de una de las subclases, o desarticuladas (*disjoint*), lo que significa que no tienen miembros en común. A esto se le refiere como restricción de desarticulación (*disjointness*) y se expresa al colocar una letra adecuada, *d* u *o*, en el círculo de especialización. Una *d* indica subclases de desarticulación y una *o* indica subclases de traslapamiento.

Una especialización también tiene una restricción de completud (*completeness*), que muestra si todo miembro del conjunto de entidades debe participar en ella.

Si todo miembro de la superclase debe pertenecer a alguna subclase, se tiene una especialización total. Si a algunos miembros de la superclase no se les puede permitir pertenecer a alguna subclase, la especialización es parcial.

En algunas jerarquías de especialización es posible identificar la subclase a la que pertenece una entidad al examinar una condición o predicado específico para cada subclase, es decir, es una especialización definida por predicado, pues la membresía a la subclase está determinada por un predicado.

Algunas especializaciones definidas por predicado usan el valor del mismo atributo en el predicado definitorio para todas las subclases. Estas se llaman especializaciones definidas por atributo.

Las especializaciones que no están definidas por predicado se dice que son definidas por el usuario, pues el usuario es el responsable de colocar la instancia de entidad en la subclase correcta.

Jerarquías múltiples y herencia

Cuando el mismo conjunto de entidades puede ser una subclase de dos o más superclases, se dice que tal clase es una subclase compartida y tiene herencia múltiple de sus superclases.

Unión

Mientras que una subclase compartida representa un miembro de todas sus superclases y hereda atributos de todas ellas, una subclase se puede relacionar con

la de una colección, llamada unión o categoría de superclases, en vez de pertenecer a todas ellas. En este caso, una instancia de la subclase hereda solo los atributos de una de las superclases, dependiendo de a cuál miembro de la unión pertenece.

Las categorías pueden ser parciales o totales, dependiendo de si cada miembro de los conjuntos que constituyen la unión participan en ella.

A.3. Álgebra relacional

De acuerdo a Elmasri[13], el álgebra relacional es muy importante por varias razones. La primera, porque proporciona un fundamento formal para las operaciones del modelo relacional. La segunda razón, y quizá la más importante, es que se utiliza como base para la implementación y optimización de consultas en los RDBMS. Tercera, porque algunos de sus conceptos se han incorporado al lenguaje estándar de consultas SQL para los RDBMS.

Operaciones relacionales unarias

Selección

SELECCIÓN se emplea para seleccionar un subconjunto de las tuplas de una relación que satisfacen una condición de selección. Se puede considerar esta operación como un filtro que mantiene sólo las tuplas que satisfacen una determinada condición.

SELECCIÓN puede visualizarse también como una partición horizontal de la relación en dos conjuntos de tuplas: las que satisfacen la condición son seleccionadas y las que no, descartadas.

En general, SELECCIÓN está designada como:

$$\sigma \langle \text{condición de selección} \rangle (R)$$

donde el símbolo σ (sigma) se utiliza para especificar el operador de SELECCIÓN, mientras que la condición de selección es una expresión lógica (o booleana) especificada sobre los atributos de la relación R .

Observe que R es, generalmente, una expresión de álgebra relacional cuyo resultado es una relación: la más sencilla de estas expresiones es sólo el nombre de una relación de base de datos. El resultado de SELECCIÓN tiene los mismos atributos que R .

SELECCIÓN es unaria, es decir, se aplica a una sola relación. Además, esta operación se aplica a cada tupla individualmente; por consiguiente, las condiciones de selección no pueden implicar a más de una tupla.

El grado de la relación resultante de una operación SELECCIÓN (su número de atributos) es el mismo que el de R . El número de tuplas en la relación resultante es siempre menor que o igual que el número de tuplas en R .

Proyección

PROYECCIÓN selecciona ciertas columnas de la tabla y descarta otras. Si sólo estamos interesados en algunos atributos de una relación, usamos la operación PROYECCIÓN para planear la relación sólo sobre esos atributos.

Por consiguiente, el resultado de esta operación puede visualizarse como una partición vertical de la relación en otras dos: una contiene las columnas (atributos) necesarias y otra las descartadas.

La forma general de la operación PROYECCIÓN es:

$$\Pi \langle \text{lista de atributos} \rangle (R)$$

donde Π (π) es el símbolo usado para representar la operación PROYECCIÓN, mientras que $\langle \text{lista de atributos} \rangle$ contiene la lista de campos de la relación R que queremos.

De nuevo, observe que R es, en general, una expresión de álgebra relacional cuyo resultado es una relación, cuyo caso más simple es obtener sólo el nombre de una relación de base de datos.

El resultado de la operación PROYECCIÓN sólo tiene los atributos especificados en $\langle \text{lista de atributos} \rangle$ en el mismo orden a como aparecen en la lista. Por tanto, su grado es igual al número de atributos contenidos en $\langle \text{lista de atributos} \rangle$.

Si la lista de atributos sólo incluye atributos no clave de R , es posible que se dupliquen tuplas. La operación PROYECCIÓN elimina cualquier tupla duplicada, por lo que el resultado de la misma es un conjunto de tuplas y, por consiguiente, una relación válida. Esto se conoce como eliminación de duplicados.

Renombrar

Podemos definir una operación RENOMBRAR como un operador unario. Una operación RENOMBRAR aplicada a una relación R de grado n aparece denotada de cualquiera de estas tres formas

$$\rho_S(B_1, B_2, \dots, B_n) \text{ o } \rho_S(R) \text{ o } \rho_{(B_1, B_2, \dots, B_n)}(R)$$

donde el símbolo ρ (rho) se utiliza para especificar el operador RENOMBRAR, S es el nombre de la nueva relación y B_1, B_2, \dots, B_n son los de los nuevos atributos.

La primera expresión renombra tanto la relación como sus atributos, la segunda solo lo hace con la relación y la tercera solo con los atributos. Si los atributos de R son (A_1, A_2, \dots, A_n) por este orden, entonces cada A_i es renombrado como B_i .

Operaciones de álgebra relacional de la teoría de conjuntos

Para combinar los elementos de dos conjuntos se utilizan varias operaciones de la teoría de conjuntos, como la UNIÓN, la INTERSECCIÓN y la DIFERENCIA DE CONJUNTOS (llamada también a veces MENOS, o MINUS). Todas ellas son operaciones binarias, es decir, se aplican a dos conjuntos de tuplas.

Cuando se refieren a las bases de datos relacionales, las relaciones sobre las que se aplican estas tres operaciones deben tener el mismo tipo de tuplas; esta condición recibe el nombre de compatibilidad de unión.

Dos relaciones $R(A_1, A_2, \dots, A_n)$ y $S(B_1, B_2, \dots, B_n)$ se dice que son de unión compatible si tienen el mismo grado n y si $\text{dom}(A_i) = \text{dom}(B_i)$ para $1 \leq i \leq n$.

Esto significa que ambas relaciones tienen el mismo número de atributos y que cada par correspondiente cuenta con el mismo dominio.

Podemos definir las tres operaciones UNIÓN, INTERSECCIÓN y DIFERENCIA DE CONJUNTO en dos relaciones de unión compatible R y S del siguiente modo:

- UNIÓN El resultado de esta operación, especificada como $R \cup S$, es una relación que incluye todas las tuplas que están tanto en R como en S o en ambas, R y S . Las tuplas duplicadas se eliminan.
- INTERSECCIÓN El resultado de esta operación, especificada como $R \cap S$, es una relación que incluye todas las tuplas que están en R y en S .
- DIFERENCIA DE CONJUNTO (o MENOS). El resultado de esta operación, especificada como $R - S$, es una relación que incluye todas las tuplas que están en R pero no en S .

Producto cartesiano

Se trata también de una operación de conjuntos binarios, aunque no es necesario que las relaciones en las que se aplica sean una unión compatible. En su forma binaria produce un nuevo elemento combinando cada miembro (tupla) de una relación (conjunto) con los de la otra.

En general, el resultado de $R(A_1, A_2, \dots, A_n) \times S(B_1, B_2, \dots, B_m)$ es una relación Q con un grado de $n + m$ atributos $Q(A_1, A_2, \dots, A_n, B_1, B_2, \dots, B_m)$, en este orden.

La relación resultante Q tiene una tupla por cada combinación de éstas (una para R y otra para S). Por tanto, si R tiene n_R tuplas (indicado como $|R| = n_R$), y S cuenta con n_S tuplas, $R \times S$ tendrá $n_R * n_S$ tuplas.

La operación PRODUCTO CARTESIANO n -ario es una extensión del concepto indicado más arriba que produce nuevas tuplas concatenando todas las posibles combinaciones de tuplas desde n relaciones subyacentes.

Es útil cuando va seguida por una selección que correlacione los valores de los atributos procedentes de las relaciones componentes.

Operaciones relacionales binarias

Concatenación

CONCATENACIÓN, especificada mediante \bowtie , se emplea para combinar tuplas relacionadas de dos relaciones en una sola. Esta operación es muy importante para

cualquier base de datos relacional que cuente con más de una relación, ya que nos permite procesar relaciones entre relaciones.

La forma general de una CONCATENACIÓN en dos relaciones $R(A_1, A_2, \dots, A_n)$ y $S(B_1, B_2, \dots, B_m)$ es:

$$R \bowtie \langle \text{condición de conexión} \rangle S$$

El resultado de la CONCATENACIÓN es una relación Q de $n + m$ atributos $Q(A_1, A_2, \dots, A_n, B_1, B_2, \dots, B_m)$ por este orden; Q tiene una tupla por cada combinación de éstas (una para R y otra para S) siempre que dicha combinación satisfaga la condición de conexión.

Ésta es la principal diferencia existente entre el PRODUCTO CARTESIANO y la CONCATENACIÓN. En la CONCATENACIÓN sólo aparecen en el resultado las combinaciones de tuplas que satisfacen la condición de conexión, mientras que en el PRODUCTO CARTESIANO se incluyen todas las combinaciones de tuplas.

La condición de conexión está especificada sobre los atributos de las dos relaciones R y S y es evaluada para cada combinación de tuplas, incluyéndose en la relación Q resultante en forma de una única tupla combinada sólo aquéllas cuya condición de conexión se evalúe como VERDADERO.

División

La DIVISIÓN, especificada mediante \div , es útil para cierto tipo de consultas que a veces se realizan en aplicaciones de bases de datos.

Para que una tupla t aparezca en el resultado T de la DIVISIÓN, los valores de aquélla deben aparecer en R en combinación con cada tupla en S .

Observe que en la formulación de la operación DIVISIÓN, las tuplas de la relación denominador restringen la relación numerador seleccionando aquellas tuplas del resultado que sean iguales a todos los valores presentes en el denominador.

La operación DIVISIÓN está definida por conveniencia para gestionar las consultas que implican una cuantificación universal.

Bibliografía

- [1] Google, *Google Trends - NoSQL*, 2020. dirección: <https://trends.google.es/trends/explore?date=all&q=NoSQL>.
- [2] E. F. Codd, «A relational model of data for large shared data banks», *Communications of the ACM*, vol. 13, DOI: [10.1145/362384.362685](https://doi.org/10.1145/362384.362685).
- [3] C. Li, «Transforming relational database into HBase: A case study», en *2010 IEEE International Conference on Software Engineering and Service Sciences*, jul. de 2010, págs. 683-687. DOI: [10.1109/ICSESS.2010.5552465](https://doi.org/10.1109/ICSESS.2010.5552465).
- [4] A. Chebotko, A. Kashlev y S. Lu, «A Big Data Modeling Methodology for Apache Cassandra», en *2015 IEEE International Congress on Big Data*, jun. de 2015, págs. 238-245. DOI: [10.1109/BigDataCongress.2015.41](https://doi.org/10.1109/BigDataCongress.2015.41).
- [5] M. J. Mior, K. Salem, A. Aboulnaga y R. Liu, «NoSE: Schema Design for NoSQL Applications», *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, n.º 10, págs. 2275-2289, oct. de 2017, ISSN: 2326-3865. DOI: [10.1109/TKDE.2017.2722412](https://doi.org/10.1109/TKDE.2017.2722412).
- [6] D. Martinez-Mosquera, R. Navarrete y S. Lujan-Mora, «Modeling and Management Big Data in Databases—A Systematic Literature Review», *Sustainability*, 2020. DOI: <https://doi.org/10.3390/su12020634>.
- [7] A. de la Vega, D. García-Saiz, C. Blanco, M. Zorrilla y P. Sánchez, «Mortadelo: Automatic generation of NoSQL stores from platform-independent data models», *Future Generation Computer Systems*, vol. 105, págs. 455-474, 2020, Publisher: Elsevier.
- [8] datafluent, *Kashliev Data Modeler*, 2020. dirección: <https://www.datafluent.org/>.
- [9] Amazon, *NoSQL Workbench for Amazon DynamoDB*, 2020. dirección: <https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/workbench.html>.
- [10] Hackolade, *Hackolade*, 2020. dirección: <https://hackolade.com/>.
- [11] C. de Lima y R. dos Santos Mello, «A workload-driven logical design approach for NoSQL document databases», en *Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services*, 2015, págs. 1-10.
- [12] M. J. Mior, K. Salem, A. Aboulnaga y R. Liu, «NoSE: Schema design for NoSQL applications», *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, n.º 10, págs. 2275-2289, 2017, Publisher: IEEE.

-
- [13] Ramez Elmasri, *Fundamentos de SIstemas de Bases de Datos*. Pearson, ISBN: 84-7829-085-0.
- [14] P. Chen, *The entity-relationship model: Toward a unified View of data*. dirección: <https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.523.6679>.
- [15] Cristina Marta Bender, Claudia Deco, Juan Sebastián González Sanabria, María Hallo y Julio César Ponce Gallegos, *Tópicos Avanzados de Bases de Datos*, 1.^a ed. Iniciativa Latinoamericana de Libros de Texto Abiertos.
- [16] E. Brewer, «Towards robust distributed systems», ene. de 2000. DOI: [10.1145/343477.343502](https://doi.org/10.1145/343477.343502).
- [17] C. Coronel y S. Morris, *Database Systems: Design, Implementation, and Management*, 13.^a ed. Cengage, ISBN: 978-1-337-62790-0.
- [18] P. J. Sadalage y M. Fowler, *Nosql Distilled: A Brief Guide to the Emerging World of Polyglot Persistence*, 1.^a ed. Pearson Education, ISBN: 978-0-321-82662-6.
- [19] *HTTP - Hypertext Transfer Protocol*. dirección: <https://www.w3.org/Protocols/>.
- [20] *HTML*. dirección: <https://developer.mozilla.org/es/docs/Web/HTML>.
- [21] Wikipedia contributors, *Style sheet language* — *Wikipedia, The Free Encyclopedia*. 2020. dirección: https://en.wikipedia.org/w/index.php?title=Style_sheet_language&oldid=937209434.
- [22] *What is CSS?* Dirección: <https://www.w3.org/standards/webdesign/htmlcss#whatcss>.
- [23] *Documentation Sass*. dirección: <https://sass-lang.com/documentation>.
- [24] *Most Loved, Dreaded, and Wanted Web Frameworks* - *stack overflow*. dirección: <https://insights.stackoverflow.com/survey/2019#technology--most-loved-dreaded-and-wanted-web-frameworks>.
- [25] *JavaScript*. dirección: <https://developer.mozilla.org/es/docs/Web/JavaScript>.
- [26] T. Publishing, *TypeScript Programming Language*, 1.^a ed. Independently published, nov. de 2019, ISBN: 1-70883-980-1.
- [27] *stack overflow - Developer Survey Results 2019 - Most Popular Technologies*. dirección: <https://insights.stackoverflow.com/survey/2019#most-popular-technologies>.
- [28] Wikipedia contributors, *Web framework* — *Wikipedia, The Free Encyclopedia*. 2020. dirección: https://en.wikipedia.org/w/index.php?title=Web_framework&oldid=956641850.
- [29] Wikipedia, *React* — *Wikipedia, La enciclopedia libre*. 2020. dirección: <https://es.wikipedia.org/w/index.php?title=React&oldid=125579300>.
- [30] —, *Angular (framework)* — *Wikipedia, La enciclopedia libre*. 2020. dirección: [https://es.wikipedia.org/w/index.php?title=Angular_\(framework\)&oldid=122940933](https://es.wikipedia.org/w/index.php?title=Angular_(framework)&oldid=122940933).
-

-
- [31] ¿Qué es Vue.js? Dirección: <https://es.vuejs.org/v2/guide/>.
- [32] What is NuxtJS? Dirección: <https://nuxtjs.org/guide>.
- [33] Wikipedia contributors, *CSS framework* — *Wikipedia, The Free Encyclopedia*. 2020. dirección: https://en.wikipedia.org/w/index.php?title=CSS_framework&oldid=952876853.
- [34] *Introduction Material Design*. dirección: <https://material.io/design/introduction#goals>.
- [35] *Vuetify Component API Overview*. dirección: <https://vuetifyjs.com/en/components/api-explorer/>.
- [36] *Documentation Bootstrap*. dirección: <https://getbootstrap.com/docs/4.5/getting-started/introduction/>.
- [37] Wikipedia, *MongoDB* — *Wikipedia, La enciclopedia libre*. 2020. dirección: <https://es.wikipedia.org/w/index.php?title=MongoDB&oldid=125700084>.
- [38] —, *CouchDB* — *Wikipedia, La enciclopedia libre*. 2020. dirección: <https://es.wikipedia.org/w/index.php?title=CouchDB&oldid=125555652>.
- [39] *GoJs*. dirección: <https://gojs.net/latest/index.html>.
- [40] *Fabric Documentation*, 2020. dirección: <http://fabricjs.com/fabric-intro-part-1>.
- [41] Wikipedia contributors, *D3.js* — *Wikipedia, The Free Encyclopedia*. 2020. dirección: <https://en.wikipedia.org/w/index.php?title=D3.js&oldid=956132961>.
- [42] M. Lutz, *Learning Python*, 5.^a ed. O'Really, jun. de 2013, ISBN: 978-1-4493-5573-9.
- [43] Wikipedia, *Django (framework)* — *Wikipedia, La enciclopedia libre*. 2020. dirección: [https://es.wikipedia.org/w/index.php?title=Django_\(framework\)&oldid=125055379](https://es.wikipedia.org/w/index.php?title=Django_(framework)&oldid=125055379).
- [44] U. C. los Ángeles, *Metodología de desarrollo de software*, 2020. dirección: <https://www.uladech.edu.pe/images/stories/universidad/documentos/2018/metodologia-desarrollo-software-v001.pdf>.
- [45] T. S. Guide, *The Definitive Guide to Scrum: The Rules of the Game*, 2020. dirección: <https://www.scrumguides.org/docs/scrumguide/v2017/2017-Scrum-Guide-US.pdf#zoom=100>.
- [46] S. México, *SCRUM México - Escribiendo Historias de Usuario*, 2020. dirección: <https://www.scrum.mx/informate/historias-de-usuario>.
- [47] M. T. Gallego, *Metodología SCRUM*, 2020. dirección: <http://openaccess.uoc.edu/webapps/o2/bitstream/10609/17885/1/mtrigasTFC0612memoria.pdf>.
- [48] Wikipedia, *Diagrama de clases* — *Wikipedia, La enciclopedia libre*. 2020. dirección: https://es.wikipedia.org/w/index.php?title=Diagrama_de_clases&oldid=126211624.
- [49] Catherine M. Ricardo, *Bases de datos*. McGraw-Hill, ISBN: 978-970-10-7275-2.
-