



Ingeniería en Sistemas Computacionales

Bases de Datos No Relacionales

Definición de proyecto

Adriana Díaz Aguayo

Marco Ricardo Cordero Hernández

Carlos Eduardo Rodríguez Castro

Jal., 17 de marzo de 2023

Integrantes del equipo

- Adriana Díaz Aguayo - IS727550
- Marco Ricardo Cordero Hernández - IS727272
- Carlos Eduardo Rodríguez Castro - IS727366

Nombre del proyecto

Proyecto Integrador Bases de Datos No Relacionales || Travel Predictor

Antecedentes

Con el propósito de demostrar los aprendizajes adquiridos a lo largo de la materia actual, se propone un proyecto integrador que tiene como objetivo abordar una situación real: análisis de datos para la toma de decisiones. Y no, no se está hablando de aquel otro curso.

Como bien se ha de saber, los datos por si solos no significan nada más que un conjunto de caracteres apelmazados en un solo lugar, y no es sino hasta que se les da una correcta interpretación cuando se transforman en información que puede llegar a ser vital para una empresa en términos de estudio de mercado, investigaciones de campo, análisis de segmentos de audiencia, etc., esto con el fin de impulsar el negocio o crear nuevas oportunidades en él.

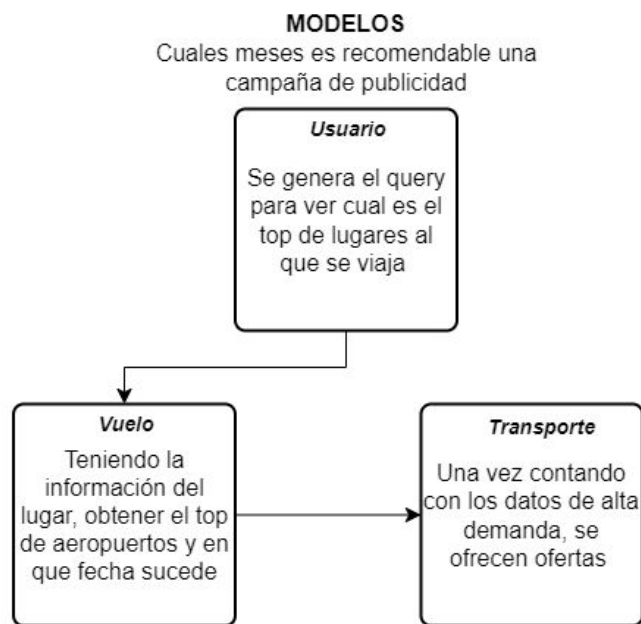
Para el futuro desarrollo, se propone la siguiente colección de datos relacionados al transporte comercial aéreo, y más específicamente, a una instancia de viaje particular:

- Aerolínea (Airline)
- Origen (From)
- Destino (To)
- Día (Day)
- Mes (Month)
- Año (Year)
- Edad (Age)
- Género (Gender) [Hombre, Mujer, Sin especificar, Confidencial]
- Motivo (Reason) [Vacaciones/Placer, Negocios/Trabajo, Regreso a casa]
- Alojamiento (Stay) [Hotel, Estadía en casa foránea de corto plazo, Casa, Amigo/Familia]
- Transporte (Transit) [Taxi de aeropuerto, Renta de automóvil. Servicio de transporte, Transporte público, Aventón, Transporte propio]
- Conexión (Connection) [¿El vuelo es escala/conexión?]
- Espera (Wait) [Si es conexión ¿Cuánto tiempo hay de por medio entre vuelos?]

Modelos propuestos

En base a los campos revisados en la sección anterior, se proponen una serie de modelos para abordar distintas áreas de oportunidad para los grupos aeroportuarios, que pueden ser soportadas por el tipo de datos con los que se cuenta:

- Modelo que sugiere cuáles son los meses en los que es recomendable introducir campañas de publicidad para empresas de renta de carros para cada aeropuerto.



- Modelo que recomienda en qué aeropuertos es recomendable abrir servicios de alimentos/bebidas.

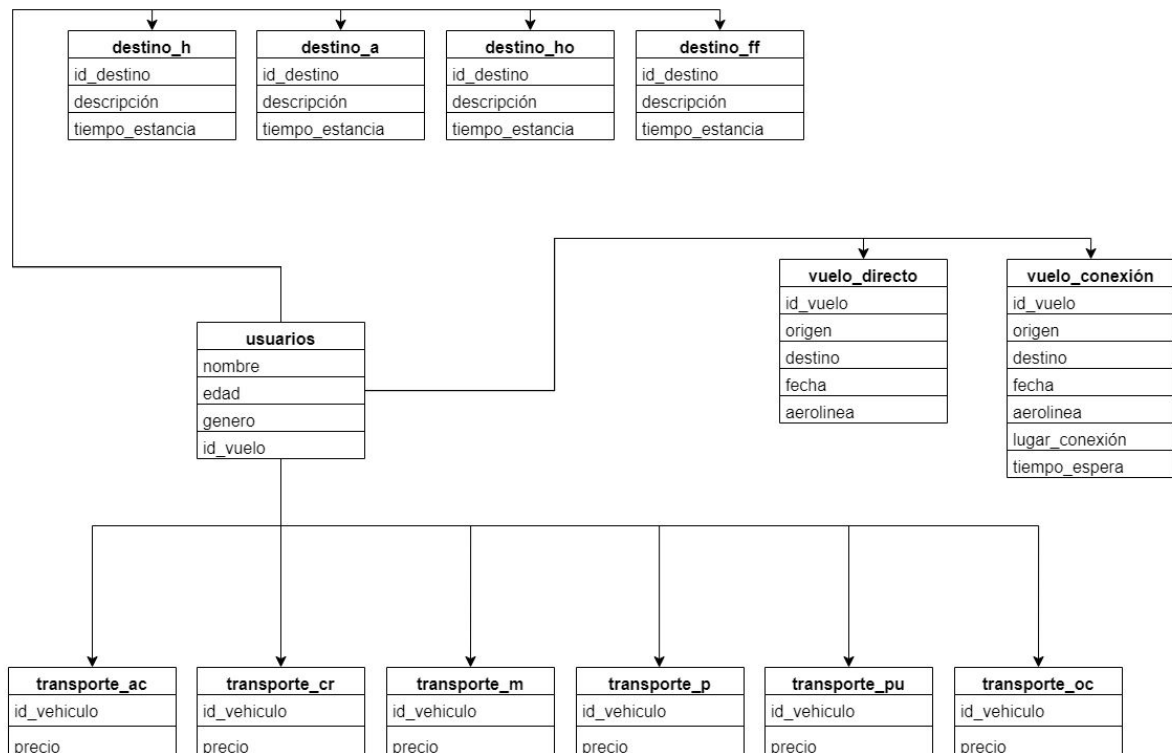


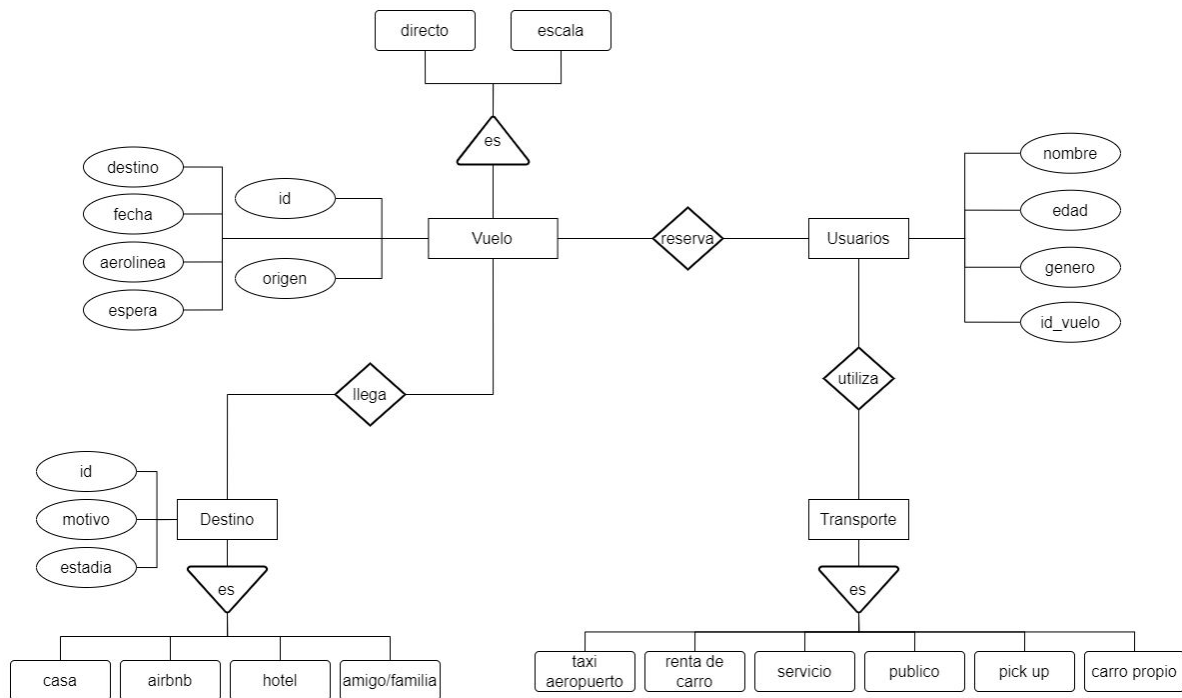
- Modelo adicional: Modelo que determine picos en afluencia de pasajeros en los aeropuertos para valorar la opción de introducir descuentos o distintas estrategias comerciales.

MODELOS
En qué momento es mejor generar ofertas/promociones de viaje



En esta ocasión, se propone un diseño general que daría soporte a los tres modelos propuestos:





Herramientas a utilizar

Una vez que se cuenta con los datos anteriores, evidentemente deben ser manipulados para lograr la interpretación que se mencionaba con anterioridad. Para ello, se hará uso práctica de herramientas revisadas en clase, así como el ejercicio de múltiples temas y conocimientos adquiridos en las mismas.

A continuación se detalla la colección de herramientas que se utilizarán. Algunas de ellas aún no se han revisado dentro del contenido del curso, sin embargo, se sabe de antemano que se hará uso de las mismas:

Cassandra: Se utilizarán las ventajas que ofrece como herramienta de almacenamiento columnar para desarrollar una estructura más flexible pero haciendo uso de las bondades que ofrecen las tablas tradicionales. También, la rapidez que proporciona un buen diseño de tablas en conjunto de una correcta definición de llaves tanto de partición como de agrupamiento, hará la retribución de datos un proceso sumamente eficiente.

MongoDB: Su modelo en base a documentos será el de mayor utilidad (más no el único) para interconectar los resultados de las consultas hacia el lenguaje de programación **Python**. Esto permitirá hacer un mejor manejo de los resultados y proveer una predicción más precisa, además de brindar una espectro extendido de posibilidades en cuanto a construcción y personalización refiere. Adicionalmente, la capacidad de operaciones de agregación y la idealización de relaciones entre colecciones, harán de esta herramienta algo sumamente parecido a lo que se podrían encontrar en una base de datos relacional, lo cual será de suma ayuda para reportes numéricos que se recabarán de los datos. Esto aplica para cualquier modelo.

Neo4j: Su modelo de almacenamiento por medio de grafos hará posible encontrar relaciones dentro de la información utilizando el lenguaje de consultas nativo llamado **Cypher** en tiempos y complejidades de consulta menores. Idealmente, como desarrolladores, las capacidades visuales de la herramienta serán de gran ayuda para corroborar que los datos que se están extrapolando y su interpretación, son correctos, lo cual podrá ser utilizado para cualquier modelo y posteriormente ser adaptado para soportar fuertemente a uno solo.

Adicional a lo descrito y como se ha venido trabajando en sesiones anteriores, el uso de **Docker** existe como una gran posibilidad, ya que el uso de esta herramienta permite una amplia gama de acciones colaborativas que acelerarían el desarrollo del proyecto en curso, además de brindar la infraestructura necesaria en una fracción del tiempo que tomaría si no se utilizara.