

Universidad de Ingeniería y Tecnología

Proyecto de Base de Datos I



BookerFest: Reserva de Eventos en Línea

Docente: Quintanilla Cruz, Rommel

Integrantes

Lazo Mera, Kalos - 202210184

Alferez Vicente, Bladimir - 202210391

Llallire Meza, Mateo Gabriel - 202210187

2023-I

Lima, 7 de mayo de 2023

Índice

1. Requisitos	2
1.1. Introducción	2
1.2. Descripción general del problema/organización/empresa	2
1.3. Necesidad/usos de la base de datos	3
1.4. ¿Cómo resuelven el problema hoy?	3
1.4.1. ¿Cómo se almacenan/procesa los datos hoy?	3
1.4.2. Flujo de datos	3
1.5. Descripción detallada del sistema	4
1.5.1. Objetos de información actuales	4
1.5.2. Características y funcionalidades esperadas	4
1.5.3. Tipos de usuarios existentes/necesarios	5
1.5.4. Tipos de consulta, actualizaciones	5
1.5.5. Tamaño estimado de la base de datos	6
1.6. Objetivos del proyecto	7
1.7. Referencias del proyecto	7
1.8. Eventualidades	7
1.8.1. Problemas que pudieran encontrarse en el proyecto	7
1.8.2. Límites y alcances del proyecto	8
2. Modelo Entidad-Relación	9
2.1. Reglas semánticas	9
2.2. Modelo Entidad-Relación	9
2.3. Especificaciones y consideraciones sobre el modelo	9
3. Modelo Relacional	10
3.1. Especificación	10
3.1.1. Entidades	10
3.1.2. Atributos Multivaluados	10
3.1.3. Entidades Débiles	10
3.1.4. Relaciones Ternarias	10
3.1.5. Relaciones Binarias	10
3.2. Diccionario de Datos	10
3.3. Creación de tablas	10

1 | Requisitos

1.1. Introducción

BookerFest es una empresa creada por un grupo de jóvenes apasionados por la tecnología y las actividades sociales. Desde sus inicios se interesaron en actividades de venta y promoción de entradas para eventos pequeños y medianos. Fue así que gradualmente las ventas aumentaron gracias a la gran demanda del público asistente a eventos en Lima, motivo por el que han pensado en optar por una base de datos para optimizar sus consultas de información, pues actualmente utilizan archivos `.csv` en línea para el almacenamiento de sus datos.

A día de hoy, BookerFest busca convertirse en una plataforma líder en la reserva de eventos en línea. Para ello, tienen pensado expandir sus alcances: capacidad de promoción de eventos en todo el Perú. Para lograrlo, necesitan una plataforma web que tenga una interfaz de usuario intuitiva y la implementación de un sistema de base de datos escalable y seguro, que permita la consulta de grandes volúmenes de información eficientemente.

Frente a esta situación, BookerFest solicita la implementación de una base de datos que permita almacenar información de eventos, información de usuarios registrados en la plataforma y el proceso de compra de entradas: desde la venta hasta la generación del QR's para permitir su acceso. Del mismo modo, la implementación debe ser capaz de permitir realizar búsquedas personalizadas con ciertos filtros (lugar, hora, precio), además de brindar estadísticas en tiempo real de su alcance y ventas.

1.2. Descripción general del problema/organización/empresa

A raíz del crecimiento de usuarios y eventos por parte de la empresa BookerFest, se necesita de una base de datos de alto rendimiento. Pues su problema es que trabajan con archivos en línea del tipo `.csv`, por lo que se debe pensar primero en una correcta **migración** de datos, para no perder su información.

Gracias a que existe gran competitividad en este mercado, se busca que la plataforma web sea lo más estable posible, pues al realizar su lanzamiento ofrecerán promociones imbatibles en pro de fidelizar nuevos clientes. En consecuencia van a manejar gran volumen de peticiones e interacción de usuarios en simultáneo.

1.3. Necesidad/usos de la base de datos

La necesidad de la base de datos nace cuando surge la necesidad de desarrollar una plataforma que reúna tanto a clientes como embajadores de eventos, para que sea fácil la creación de eventos y compra de QR's. No obstante además de necesitar guardar información de clientes de manera automatizada, su principal objetivo es poder acceder a estadísticas personalizadas para cada entidad: Evento, Persona y Embajador.

Estas estadísticas son muy atractivas para los Embajadores de eventos, dado que pueden ver el alcance del público, el rango de edades predominantes: ¿Qué parte de esa muestra realiza la compra de sus QR?, además de los ingresos estimados, tickets disponibles, tendencia de compra, etc. De esta forma los Embajadores pueden tomar decisiones más acertivas, que de no ser por una base de datos que guarde registros de la audiencia y una plataforma que genere estadísticas a partir de la información almacenada, sería imposible.

1.4. ¿Cómo resuelven el problema hoy?

1.4.1. ¿Cómo se almacenan/procesa los datos hoy?

Como se mencionó anteriormente, se utilizan archivos Excel del tipo `.csv`, estos son guardados en la nube, y cuando se necesita interactuar con ellos, se tienen que descargar en una computadora local, y utilizando el lenguaje de Python, en conjunto con la librería `pandas` realizan la filtración de la información e incluso la interpretación gráfica.

Si bien realizan consultas utilizando un pequeño script, se dieron cuenta que si quieren escalar sus registros e incorporar/eliminar funcionalidades, ejecutar el script en servidores web en tiempo real ó consultar usuarios, el proceso es muy tedioso pues este depende de pequeños módulos (archivos) que pueden corromperse y confundirse entre sí (posibilidad de guardar información errónea de un usuario en otra tabla). Por todo eso y más su modelo actual no es eficiente ni escalable.

1.4.2. Flujo de datos

Los datos son generados principalmente por los usuarios y embajadores de eventos. El primero es una entidad que participa de la demanda de actividades sociales, por lo que acude a la búsqueda y análisis de distintas opciones. Para ello necesita de introducir información como su edad, departamento del Perú en el que reside y tipo de eventos, de esa forma se le puede ofrecer información personalizada, aumentando la posibilidad de compra.

Por otro lado los Embajadores de eventos tienen la posibilidad de publicar eventos con ciertos filtros, que incrementan su posibilidad de ser encontrado por algún usuario. Además utilizando datos de búsqueda hechas por usuarios, es posible ver estadísticas de audiencia y de ventas de QR's, lo que permite a los Embajadores optar por diferentes estrategias y medir fácilmente su impacto.

1.5. Descripción detallada del sistema

1.5.1. Objetos de información actuales

Dentro de objetos de información encontramos los archivos de Excel exportados en formato `.csv`, que para simular la migración de datos, serán rellenos con datos aleatorios provenientes de un pequeño script automatizado hecho en Python. De igual forma se cuenta con información externa, puesto que muchos Embajadores publican eventos en Facebook, Instagram, ó redes sociales afines, los utilizaremos para simular eventos activos en nuestra fase Beta de la página web y implementación de base de datos. Para ello se extraeran de distintos recursos en línea los eventos más sonados del Perú: cada uno con sus créditos correspondientes.

Trataremos de enfocarnos más en optimizar el tiempo de respuesta del usuario a peticiones como: ver evento, comprar QR's, etc. Todo esto será posible medirse a partir de diferentes Unit Testing: Load Testing, Stress Testing, etc. Esto nos permitirá ver cómo reacciona nuestra aplicación en escenarios de alta demanda de usuarios para 1.000.000 (un millón) de datos almacenados.

1.5.2. Características y funcionalidades esperadas

El principal objetivo es diseñar una base de datos flexible, que permite la incorporación de nuevas funcionalidades (columnas) a las distintas tablas pre-existentes. Además se necesita que sea posible de manejar un gran volumen de datos, teniendo como referencia 1.000.000 (un millón) de datos almacenados para su primer despliegue a producción.

Otro punto fundamental es el tiempo de respuesta para una petición proveniente del usuario, debe ser rápido. Comúnmente el público que compra entradas a eventos exclusivos (ciertos casos), lo realiza *formando* una cola de espera virtual, donde la velocidad y estabilidad del sistema es fundamental. Una mala optimización puede definir si un cliente consigue entrada o no. En este contexto, si un cliente obtiene una mala experiencia, nuestro modelo de negocio se verá afectado, puesto que el cliente se irá con la competencia.

Otra funcionalidad esperada es la implementación de la conexión entre el registro e inicio de sesión por parte de los usuarios con la creación de información en la base de da-

tos. Este formulario debe pedir información como: nombre, apellidos, edad, DNI y correo electrónico, y una vez creada pasar por un breve cuestionario que permite ofrecer eventos personalizados. Entre las preguntas más básicas, debe estar la locación del usuarios (departamento en el que se encuentra, si no es ingresado, por defecto es todo el Perú), sus eventos objetivos: conciertos, fiestas, discotecas, actividades al aire libre, etc.

1.5.3. Tipos de usuarios existentes/necesarios

BookerFest actualmente es una startup peruana con 3 dueños y un personal en crecimiento. Dentro del área de desarrollo de software cuenta con 1 programador a tiempo parcial, mientras que en lado de marketing cuenta con 2 trabajadores quienes se encargan de promover el uso de su administración de eventos, además de manejar sus redes sociales. Por otro lado cuenta con 2 practicantes de la carrera de Ciencia de Datos, quienes se encarga de manipular los datos, además de realizar proyecciones en cuanto a su audiencia.

Clientes, Estos son la base de la plataforma, pues son quienes ofrecen la demanda a los Embajadores de eventos. Además gracias a proporcionar sus datos, ellos contribuyen a generar estadísticas para mejorar tanto el servicio de la plataforma, como estrategias para aumentar las ventas de eventos. Así mismo deben de tener la funcionalidad de realizar compras en línea a través de distintos medios de pago, para incrementar la posibilidad de compra.

Embajadores Usuarios con un rol que les permite publicar eventos, obtener datos de su audiencia y vender QR's a través de distintos medios de pago. Por otro lado son capaces de editar sus propios eventos, actualizar el precio de sus QR's en tiempo real, así como ofrecer descuentos por un tiempo determinado. También pueden generar códigos de descuentos, para incrementar las posibilidades de venta.

1.5.4. Tipos de consulta, actualizaciones

A modo de preparar nuestra implementación de base de datos, será necesario plantear preguntas (posibles consultas) que servirán para que después de terminar nuestro modelo observemos a juicio crítico si pueden ser respondidas, caso contrario hacer sus modificaciones.

- ¿Cuáles son los filtros disponibles que permiten una búsqueda personalizada?
- ¿Cuáles son los medios de pago disponibles?
- ¿Existe algún impuesto a la hora de comprar los QR's?
- ¿Cómo puedo filtrar el evento más económico cerca de mi ubicación?

- ¿Puedo ver el evento más vendido en las últimas 24 horas?
- ¿Cómo puedo visualizar las estadísticas de alcance de audiencia para las últimas 24 horas?
- ¿Cómo coloco una promoción sobre el precio base de un tipo de entrada durante 24 horas con un máximo stock de 100 entradas?

1.5.5. Tamaño estimado de la base de datos

Se tiene pensado que la base de datos debe tener un tamaño mínimo que permita el registro de 10.000 usuarios, así como la publicación y mantenimiento para aproximadamente 1.000 eventos, donde cada uno de ellos puede estar en constante crecimiento, ya que se pueden cargar flyers que se encuentren en un rango máximo de 1 a 5MB.

<i>Nombre de la tabla</i>	<i>Longitud de atributos (Bytes)</i>	<i>Longitud del registro (Bytes)</i>
Evento	50 + 1 + 100 + 2000 + 8 + 4 + 4 + 20	2187
Usuario	8 + 50 + 50 + 100 +	100
Boleta	4 + 8 + 8 + 4 + 100	124
Local	20 + 4 + 100 + 50	174
Entrada	20 + 50 + 10 + 10	60
Embajador	50 + 50 + 4 + 8 + 4	116
Pago	20 + 8 + 8 + 6	50
Reseña	20 + 100 + 1 + 6 + 20	147
Reseña	20 + 100 + 1 + 6 + 20	147

Cuadro 1: Estimación del tamaño de registros por tabla en Bytes.

<i>Nombre de la tabla</i>	<i>Longitud del registro (Bytes)</i>	<i>Datos estimados</i>	<i>Tamaño estimado (Bytes)</i>
Evento	2.187	10.000	$2.187 \cdot 10^7$
Usuario	100	100.000	$1 \cdot 10^6$
Boleta	124	300.000	$3.72 \cdot 10^7$
Local	174	5.000	$8.7 \cdot 10^5$
Entrada	60	300.000	$1.8 \cdot 10^5$
Embajador	116	1.000	$1.16 \cdot 10^5$
Pago	50	100.000	$5 \cdot 10^6$
Reseña	147	300.000	$4.41 \cdot 10^7$

Cuadro 2: Estimación de tamaño ocupado por tabla en Bytes

Por otro lado, los usuarios pueden subir fotos de perfil para ser fácilmente identificados. Además conectado a cada usuario existe estadísticas de la cantidad de entradas compradas QR's. Las imágenes subidas por usuarios deben tener tamaño máximo de 100KB a

IMB. Para poder poner límites a la creación en tablas, el nombre del usuario tendrá como máximo una longitud de 100 caracteres.

Para el DNI sólo permitiremos 8 dígitos del tipo entero. Mientras que la edad es un numero entero menor a 150 que ocupa 4 Bytes (entero). Para concluir el correo electrónico tendrá 50 caracteres como máximo, por lo que le dedicaremos 50 bytes fijos.

En conclusión por usuario tendremos un aproximado de 150 Bytes para cada uno, lo que permite una escalabilidad ya que no consume tanto recurso. Mientras que los embajadores y en conjunto con los eventos creados obtendremos un aproximado de 5MB para una configuración completa, considerando: descripción del evento, aforo, puntuación del embajador, filtros, localización, precio de QR's, etc.

1.6. Objetivos del proyecto

1. Implementar una base de datos escalable y segura
2. Desarrollar una página web con UI (User Interface) sencilla e intuitiva.
3. Asegurar la correcta implementación de migración de datos de archivos `.csv` a PostgreSQL.
4. Permitir herramientas de análisis estadístico para Embajadores de eventos.

1.7. Referencias del proyecto

Para la realización de este proyecto, hemos realizado una investigación de mercado para conocer a la competencia, para diferenciarnos y fomentar una identidad única. Dicho así las principales plataformas para venta/compra/reserva de eventos en línea son las siguientes:

- Teleticket, <https://teleticket.com.pe/>
- Joinnus, <https://www.joinnus.com/>
- Eventbrite, <https://www.eventbrite.com.pe/>

1.8. Eventualidades

1.8.1. Problemas que pudieran encontrarse en el proyecto

- Si existe una mala implementación en la migración de datos, vamos a perder una muestra representativa y contacto del cliente. Por lo que debemos de pasar este script por copias de los ficheros existentes con distintos Unit Testing.

- Evitar que la página web salga a producción con una mala optimización, pues esa primera impresión afecta en el proceso de fidelización del cliente.
- Es imprescindible que en las estadísticas de embajadores sólo se vea la ubicación aproximada (departamento) y edad de la audiencia, más no los nombres, ni el Documento Nacional de Identidad (DNI).

1.8.2. Límites y alcances del proyecto

El siguiente proyecto busca tener un alcance que abarque a toda la población peruana. Existe mucha atención en la capital aún cuando existen eventos en provincias que son populares, pero sólo por la población local. Entonces además de exponer estas tradiciones apoyamos al crecimiento del turismo en estas provincias gracias a las facilidades que nos brindan las herramientas de tecnología.

Por otro lado existen muchas limitaciones a la hora de desarrollar el sistema, la primera es que debemos trabajar con plataformas como PayU, pues que generan confianza en usuarios, y tenemos que asegurar su correcta implementación. Otra limitante son el espacio actual de nuestro servidores que de momento son locales. Por último nuestro sitio web debe seguir un estudio de diseño UI/UX puesto que accederán personas de distintas partes del Perú, por lo que a su vez debemos asegurar su correcta velocidad.

2 | Modelo Entidad-Relación

2.1. Reglas semánticas

2.2. Modelo Entidad-Relación

2.3. Especificaciones y consideraciones sobre el modelo

3 | Modelo Relacional

3.1. Especificación

3.1.1. Entidades

3.1.2. Atributos Multivaluados

3.1.3. Entidades Débiles

3.1.4. Relaciones Ternarias

3.1.5. Relaciones Binarias

3.2. Diccionario de Datos

3.3. Creación de tablas