Project Title: Website Traffic Database

Project Objective:

The goal of this database project is to create four relational database systems that represent various analytics and traffic of four websites, which are from hotels Four Seasons, Hilton, Marriott, and Shangri-La, so that the international partners from SQU can write their analysis based on the data collected and stored on the databases. The intended outcomes include an efficient way to manage and access visitors' information. The database is designed to provide easy retrieval and updating information, supporting the operations of the website while offering a platform for reporting and analysis.

Project Scope:

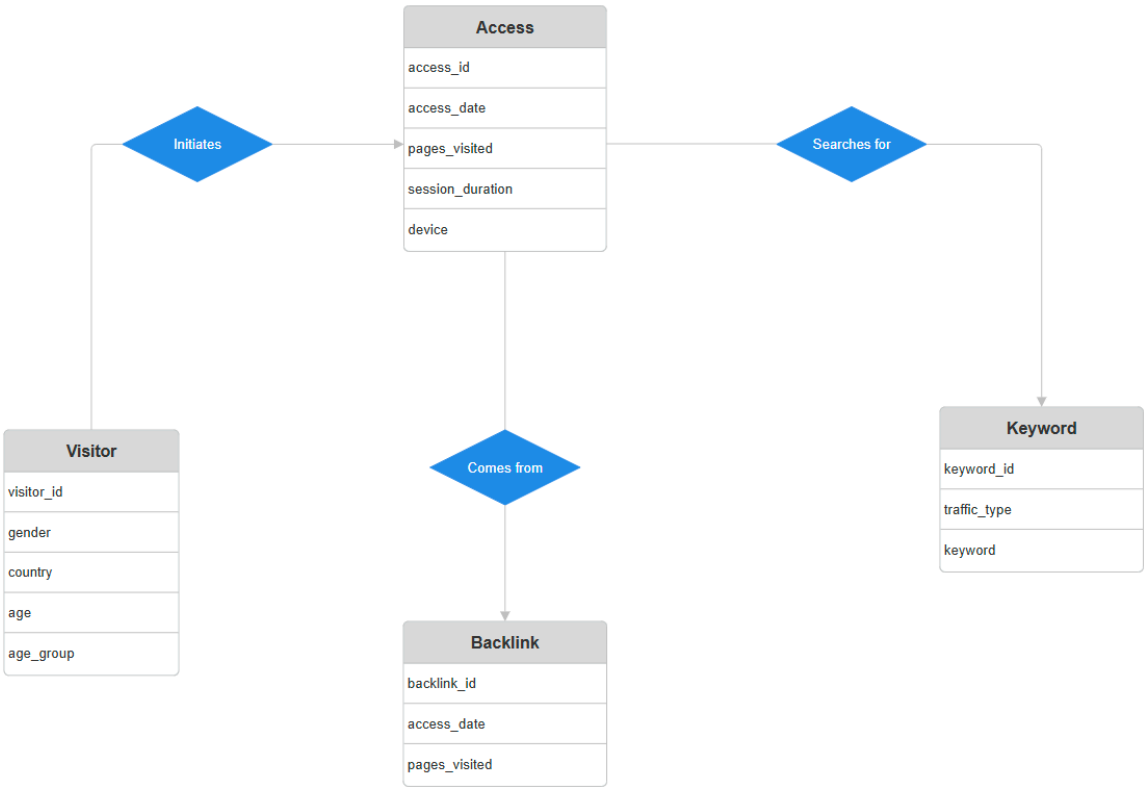
The scope of the project will contain the following aspects:

- Visitor information, including gender and age group
- Access information, including country and date aspects
- The number of pages visited during access
- Backlinks information, the number of pages linking to the website

The project will not cover aspects like clicked buttons, input received, content of pages visited, or any interactions between user and the website. The purpose of this project is to track and analyze the traffic on the website, that being the user information and the access information.

ER Diagram:

This ER Diagram represents the general entities and relationships without considering the companies, all companies will follow the same ER diagram. Below is the ER diagram for a website traffic database:



Relational Model:

Based on the ER diagram the following schemas were developed.

1. The entity sets and their attributes below, with primary keys underlined.

   Access (access_id, access_date, pages_visited, session_duration, device)

   Visitor (id, gender, country, age, age_group)

   Keyword (id, traffic_type, keyword)

   Backlink (backlink_id, backlink_url, source_url)

2. Schemas derived from relationship sets in the ER diagram.

None required

All three diamonds (Visitor–Access, Access–Keyword, Access–Backlink) are 1 : N with “Access” on the “many” side, so we implement them via foreign keys in Access rather than separate join tables.

To overcome the redundancy and follow the normalization principle we reduced the schema to the following final schemas. Since all relationships include only two columns for IDs of the tables, and the column “access_id” is present is almost all relationships, we decided to represent the relationships as foreign keys.

Access (access_id, access_date, pages_visited, session_duration, device, visitor_id → Visitor(visitor_id), keyword_id → Keyword(keyword_id), backlink_id → Backlink(backlink_id)_))

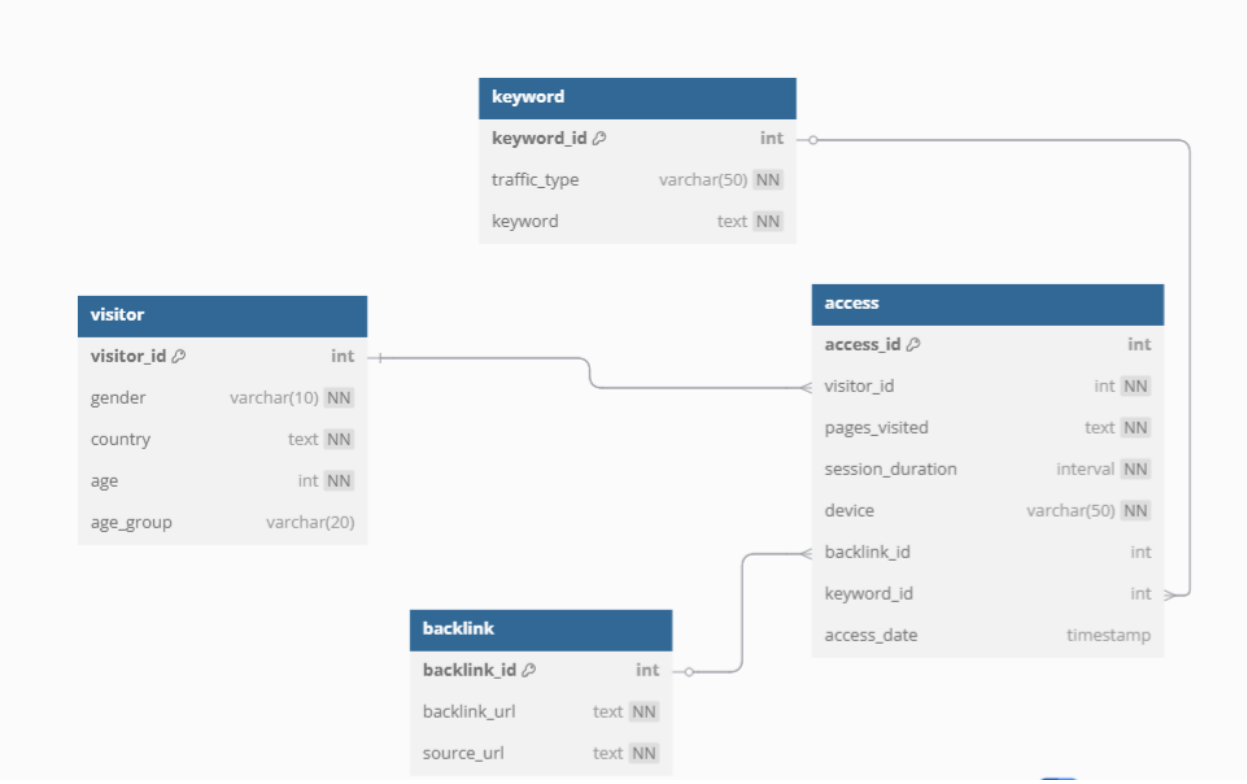
Visitor (id, gender, country, age, age_group)

Keyword (id, traffic_type, keyword)

Backlink (backlink_id, backlink_url, source_url)

Schema Diagram for Website Traffic Database:

This Database Schema represents the general tables and relationships without considering the companies, all companies will follow the exact same schema. Below is schema for a website traffic database:



Dataset Description:

- **Session**: Each session will be identified using a unique ID and will have details such as date, pages visited, session duration, and device.
- **Visitor**: This table will have information about the visitors. Each visitor will have an ID, gender, country, and age group.
- **Backlink**: This table will have information about external websites connected to the website. This information will include backlink ID and URL.
- **Keyword**: This table will present the information about the keyword traffic. It will have a unique ID for each keyword, and traffic type.

Tables Creation and Database Schema

Attached is the SQL file for creating tables for the website traffic database project. Each hotel will have versions of the following tables.

Relation	Attribute	Domain
Access	access_id	SERIAL
	visitor_id	Int

	access_date	TIMESTAMP
	pages_visited	int
	session_duration	Interval
	device	varchar(50)
	backlink_id	int
	Keyword_id	int
Visitor	visitor_id	SERIAL
	gender	varchar (10)
	country	text
	age	int
	age_group	Varchar (20)
Backlink	backlink_id	SERIAL
	backlink_url	text
	source_url	text
Keyword	keyword_id	SERIAL
	traffic_type	varchar (50)
	keyword	text

Data Population:

Attached is the SQL file for populating data for the website traffic database project. For our data, to populate the keyword table and backlink table we used real data collected from websites that give statistics and information about the traffic for a specific website. For the other tables, since this data is confidential, therefore unavailable, we used a python program to generate data and simulate real data, making sure our data matches the statistics found about the websites.

Query Questions:

Below are thirteen questions that answer specific questions according to the requirements given by the international partners to us. Those questions will be applied to all hotels in the attached SQL file Team-3-Query.sql. These questions aim to help international partners to do their analysis.

1. International partners want to identify which keywords drive the most traffic to the website. Find the percentage of total traffic associated with each keyword.

2. International partners want to analyze website engagement trends. Calculate the bounce rate (percentage of visits with only one page visited) per month.
3. International partners want to understand how long visitors stay on the website. Find the average session duration in seconds and display it formatted as HH:MM:SS.
4. International partners want to measure user engagement. Determine the average number of pages visited per session.
5. International partners want to evaluate overall user retention. Calculate the total bounce rate across all website visits.
6. International partners want to analyze the demographic composition of visitors. Find the percentage of visitors by gender.
7. International partners want to understand the age distribution of website visitors. Calculate the percentage of visitors by age group.
8. International partners want to know which countries contribute the most visitors. Determine the percentage of total visitors from each country, sorted in descending order.
9. International partners want to track monthly website traffic trends. Count the total number of accesses per month.
10. International partners want to monitor engagement fluctuations over time. Calculate the monthly bounce rate.
11. International partners want to assess how deeply users interact with the site each month. Determine the average number of pages visited per month.
12. International partners want to know which devices visitors use the most. Count the number of accesses per device type and calculate their respective percentages.
13. International partners want to monitor engagement trends over time. Calculate the average session duration per month.

Data Visualization Using Python:

To visualize the data stored in our website traffic databases, we built a Python-based dashboard using libraries such as **pandas**, **Matplotlib**, **Seaborn**, and **Tkinter**. This dashboard allows users to filter by hotel and view insights through interactive visualizations. Key metrics include age group and gender distribution, device usage, country-wise visitors, keyword traffic types, backlink sources, session trends over time, and raw age histograms.

These visualizations helped international partners better understand user behavior, traffic trends, and engagement across the four hotel websites. The dashboard connects

directly to the PostgreSQL database, runs SQL queries for each hotel, and displays the results in a clean, dark-themed interface for easy analysis and comparison.

Conclusion:

In our project, we faced several challenges while building a database in collaboration with international partners, especially when it came to accessing data due to confidentiality issues. This forced us to come up with a way to simulate the missing data from averages using AI and Python scripts, which was a big learning experience. This gave us experience working with data generation and helped us see how AI can be a valuable tool for these kinds of tasks. Throughout the project, we were able to apply the concepts we had learned in class, particularly in database design, data population, and building meaningful queries.

Using Jupyter Notebooks, we ran Python scripts to manipulate and store data, and later used those scripts to create a visualization dashboard that allowed users to explore key trends in the data across each hotel. This taught us how to turn raw data into insights through visual representation, which added real value to the project. Working with international partners introduced some difficulty in communication between the parties, mainly due to time zone differences. But coordinating these meetings helped us understand the challenges of working with a potential overseas employer.

Overall, this project not only helped us solidify our technical skills but also emphasized the importance of creative problem-solving and effective collaboration when facing real-world obstacles. The experience gave us valuable insights into working with data, managing international teams, and turning class concepts into practical tools.