DIAS: Data-Intensive Applications and Systems Laboratory
School of Computer and Communication Sciences
Ecole Polytechnique Fédérale de Lausanne
Building BC, Station 14
CH-1015 Lausanne
URL: http://dias.epfl.ch/

Databases Project – Spring 2019

Team No: 32

Names: Sophie Ammann, Samuel Chassot and Daniel Filipe Nunes Silva

Contents

DIAS: Data-Intensive Applications and Systems Laboratory
School of Computer and Communication Sciences
Ecole Polytechnique Fédérale de Lausanne
Building BC, Station 14
CH-1015 Lausanne
URL: http://dias.epfl.ch/

Deliverable 1

Assumptions

We use MySQL syntax for this project.
Listings all have a bed_type, property_type, room_type, cancellation_policy, a host and a neighborhood.
A review must be written by a reviewer about a given listing.
A calendar must have a day and a listing, otherwise it should not exist.
A host must be in a neighborhood, which must be itself in a city, which must be itself in a country.

Entity Relationship Schema

Schema

See *ER-Schema.png*

Description

Most of the attributes related to a listing are grouped in one single table *Listing*.
The *Amenity*, Bed_*type*, *Property_type*, *Room_Type*, *Cancellation_policy*, *Country*, *City*, *Day* and *Host_verification* have been normalized to reduce redundancy.
Foreign key relationships are hardly connected to preserve integrity even if it implies to drop a lot of data after deletions (*ON DELETE CASCADE*). For example, if we delete a *Country*, we will delete every *City* in this *Country* and therefore delete every *Neighborhood* in the *Cities* and so forth.

Relational Schema

ER schema to Relational schema

The translation of our ER schema to our Relation schema was more or less straightforward. Nevertheless, the type were not always obvious. We also decided to add a unique *id* in our tables.

DDL

See *Create.sql*

General Comments

For this first work, we thought it was important to work the three together to understand the database correctly. We designed the basis of the ER model, and modified it until the three of us were satisfied. Then we split the work (SQL commands, report, creation of ER model).

DIAS: Data-Intensive Applications and Systems Laboratory
School of Computer and Communication Sciences
Ecole Polytechnique Fédérale de Lausanne
Building BC, Station 14
CH-1015 Lausanne
URL: http://dias.epfl.ch/

Deliverable 2

Assumptions

We assume that all listings' city corresponds to the filename in which they are.
We put new ids for all listings and hosts to not have duplicates between cities.
We remove reviewers if they have the same id but not same name to keep only one.
We remove comments in reviews that consist of only one character (often a quote or a comma).
After parsing and cleaning all the data, we end up with simple csv files, whose name corresponds to our tables' names and columns are the same as specified in our schema.

Data Loading

Query Implementation

<For each query>

Query a:

Description of logic:

<What does the query do and how do I decide to solve it>

SQL statement

<The SQL statement>

Interface

Design logic Description

<Describe the general logic of your design as well as the technology you decided to use>

Screenshots

<Provide some initial screen shots of your interface>

General Comments

We worked all together to make the suggested changes after Milestone 1. Then, Sophie worked on the queries as well as on redesigning the new ER schema, Samuel worked on parsing and cleaning the data and Daniel worked on the interface.