Relational Database Model in 3NF and Physical Database Creation

By
Bharath Venkatesh Srinivasan

Contents

Relational Data Model

Assumptions/Notes About Data Entities and Relationships

Assumptions helps us to figure out the relations between the tables like 1:1 or 1:M or M:M or M:1, helps in determining the cardinality, modality and the type of participation of the tables in our database. Further it helps in building the reference and relationships between the parent and child tables and building the Normal forms of our database design which helps us to keep track and check our database design's Normal form easily. It also makes the process of building the Entity Relationship diagrams easier.

1. Each listing has single host and each host can have 1:M listings.
2. Each location is for one listing and each listing can have only one location.
3. Each listing has 1:1 property_details since all the property details correspond to unique listing_id.
4. Each review_details corresponds to one listing_id but listing_id has 1:M review_details as each listing can have multiple reviews.
5. Each listing has 1:1 review_score since review_score in listings data shared is for unique listing_id.
6. Each listing can be booked on multiple calendar dates(1:M relationship).
7. Reviewer_name is repeated in review_details table and to avoid redundancy a new table called reviewer_info is created.
8. All listing details except host details can be added in one single table since they are all based on unique listing_id but to enhance understandability and to improve explainability they are divided into 4 tables.
9. Listing_id is considered both Primary key and foreign in location_details, property_details and review_score tables.
10. Identifying relationship is present between listings table and location details, property_details, review_score and calendar tables each since the relationship between these two entities is in such a way that child entity is identified through its association with the parent entity (listings table).
11. Amenities column present in listings data is removed to incorporate 1NF where single cell cannot hold multiple values. This problem can be tackled by creating a new table with unique amenities and then connecting the listing table with the unique amenities table using a linkage table. This linkage table will have amenity IDs corresponding to each listing id.

Entity – Data Attributes

Table Name	Table – Attributes	Details
listings	listing_id host_id listing_url scrape_id last_scraped listing_name summary space description experiences_offered neighborhood_overview notes transit thumbnail_url medium_url picture_url xl_picture_url	This table provides all information about listings/apartments, their address, name, summary, notes , space, description, neighborhood overview, notes, transportation details and some picture of the listings.
location_details	listing_id street neighbourhood neighbourhood_cleansed neighbourhood_group_cleansed city state zipcode market smart_location country_code country latitude longitude is_location_exact	This table has data about the location details of all listings such as Street name, city, state, zip code, neighborhood information, market, smart location ID, country code, country, and latitude & longitude along with set “t” if the location is exact and “f” is the location is not exact.

property_details	listing_id property_type room_type accommodates bathrooms bedrooms beds bed_type price weekly_price monthly_price security_deposit cleaning_fee guests_included property_detailscol extra_people minimum_nights maximum_nights calender_updated has_availability availability_30 availability_60 availability_90 availability_365 calender_last_scraped	This table provides information about all property such as the type of each listing, room details, accommodation, bathrooms, bedrooms, total number beds, bed type, amenities information, their pricing details for daily, weekly, monthly and for annual, extra fee like deposits and the availability details which includes the availability for every 30, 60, 90 and 365 days.
host	host_id host_url host_name host_since host_location host_about host_response_time host_response_rate host_acceptance_rate host_is_superhost host_thumbnail_url host_picture_url host_neighbourhood host_listings_count host_total_listings_count	This table contains information about the unique host ID & their names, host since date , host's state & city along with their details, host response time , host response rate along their profile pictures and details like whether they have a profile picture or their profile is verified or not by having Boolean information like "t" and "f".

	host_verifications host_has_profile_pic host_identity_verified	
calendar	listing_id date availability price	This table has 1million information about calendar date for each listing along with their availability information and the price per day.
review_details	listing_id id date reviewer_id comments	This table will have details of reviewer like review ID their comments and date of comment posted for each listing id.
review_score	listing_id number_of_reviews first_review last_review review_scores_rating review_scores_accuracy review_scores_cleanliness review_scores_checkin review_score_communication review_score_location review_score_value requires_license jurisdiction_names instant_bookable cancellation_policy require_guest_profile_picture require_guest_phone_verification calculated_host_listings_count reviews_per_month	Review_score table contains all information on reviews like no. of reviews, date of reviews, their scores on rating, accuracy, cleanliness, checkin, communication, location & overall valuealong with the inforamtion whether the guest needs to have profile picture and verified phone number.
reviewer_info	reviewer_id reviewer_name	This table provides the information about reviewer id and their name.

Entity – Relationship

Entity – Entity	Relationship	Explanation
listings - location_details	1:1	For each listing id, there is one location details.
listings - host	1:M	One listing id will have many hosts id and vice versa.
listings - property_details	1:1	1 listing can have one property detail and Vice versa.
listings - calender	1:M	One listing id will have multiple date and vice versa
listings - review_details	1:M	One listing id will have multiple reviews and vice versa
listings - review_score	1:1	Each listing will have one unique entry from review_score
reviewer_info - review_details	1:M	one reviewer will have multiple details

Reasoning why the data model is in 3NF.

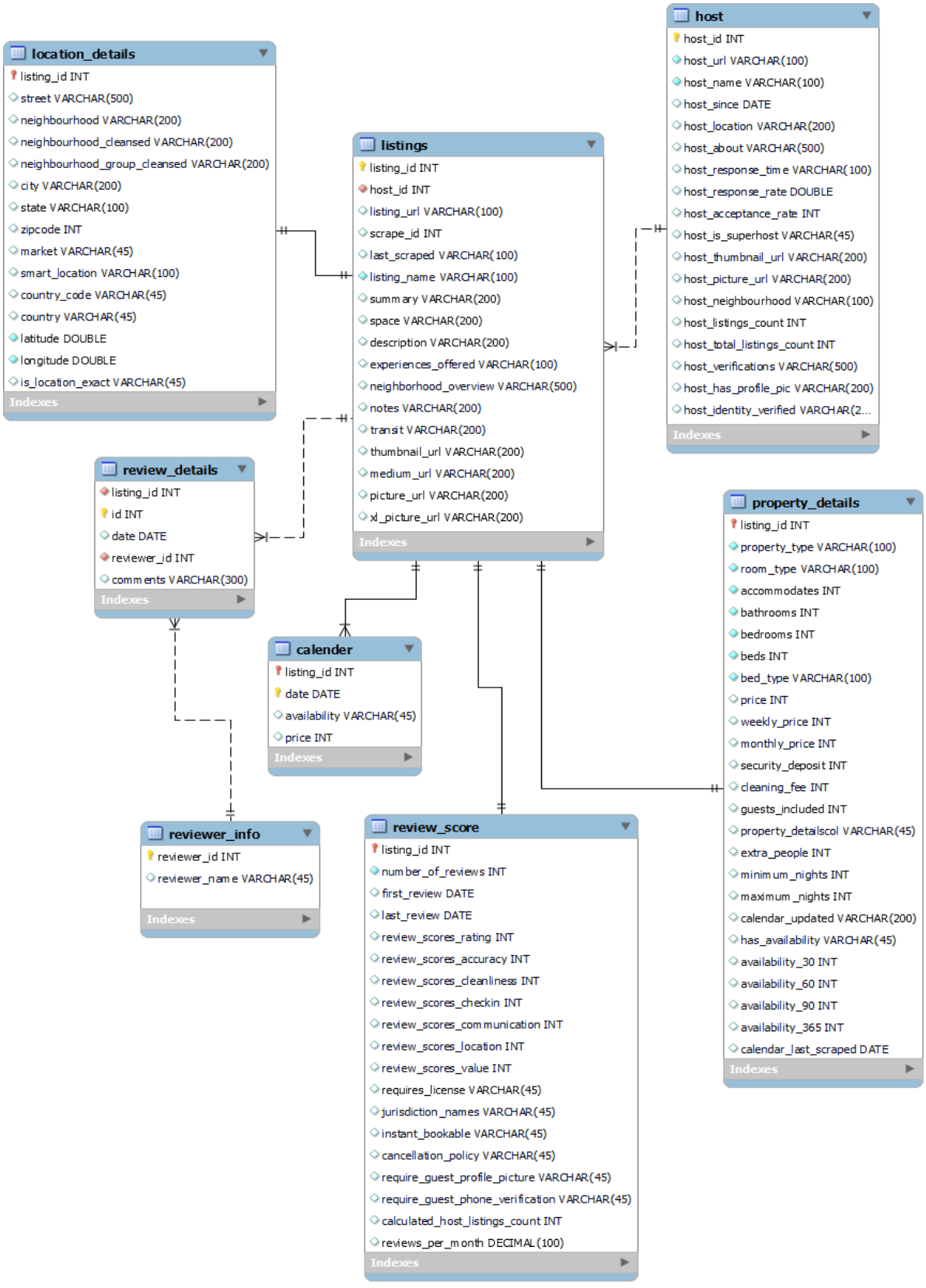
For 3NF, all table should obey the following rules,

1. Must be One-Normal Form (1NF)
   - Establish functional dependency.
   - Every attribute can contain only an atomic value.
   - Every entity has a single identifier that uniquely identifies each instance.
2. Must be Two-Normal Form (2NF)
   - Every non-primary-key attribute is fully functionally dependent on the primary key.
3. Finally, Every non-primary key attribute is functionally dependent only on complete primary key and not on any other non-key attribute.

Reason why the model is 3NF:

In this model, all duplicate columns from table have been removed or moved to separate table. Further, every table has one unique primary key. Also, whichever column had multiple values, example amenities we have created a new table for them and separated them into different columns and linked with main table with the help of connecting table. There is no indirect relationship between values in the same table that causes functional dependency so we can say that is no transitive dependency in our model. Finally, every table has unique subject there is no functional dependencies or no non-primary-key attribute is transitively dependent on the primary key.

Entity-Relationship Diagram

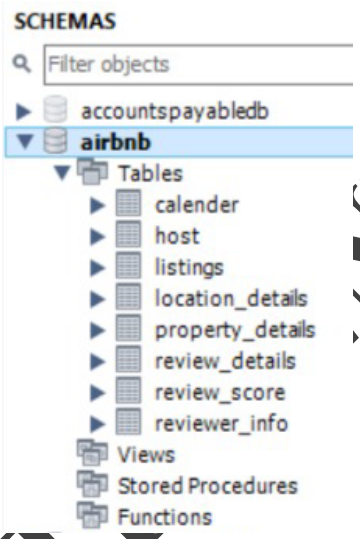


Physical MySQL Database

Assumptions/Notes About Data Set

1. In the reviewer_info, the column reviewer name had some special character which is removed.
2. Under the review_details table, comments column with special characters are replaced with appropriate characters.
3. Pricing information under the property_details tables contained $, which is removed and formatted as regular number.

Screen shot of Physical Database objects



Data in the Database

Table Name	Primary Key	Foreign Key	# Rows in Table
calender	listing_id & date	listing_id	1000
host	host_id		769
listings	listing_id	host_id	1000
location_details	listing_id	listing_id	1000
property_details	listing_id	listing_id	1000
review_details	id	listing_id & reviewer_id	1000
review_score	listing_id	listing_id	1000
reviewer_info	reviewer_id		989