

Relational Database Model in 3NF and

Physical Database Creation

**Contents**

[Relational Data Model 3](#_TOC_250006)

[Assumptions/Notes About Data Entities and Relationships 3](#_TOC_250005)

[Entity-Relationship Diagram 8](#_TOC_250004)

[Physical MySQL Database 9](#_TOC_250003)

[Assumptions/Notes About Data Set 9](#_TOC_250002)

[Screen shot of Physical Database objects 9](#_TOC_250001)

[Data in the Database 9](#_TOC_250000)

# 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.

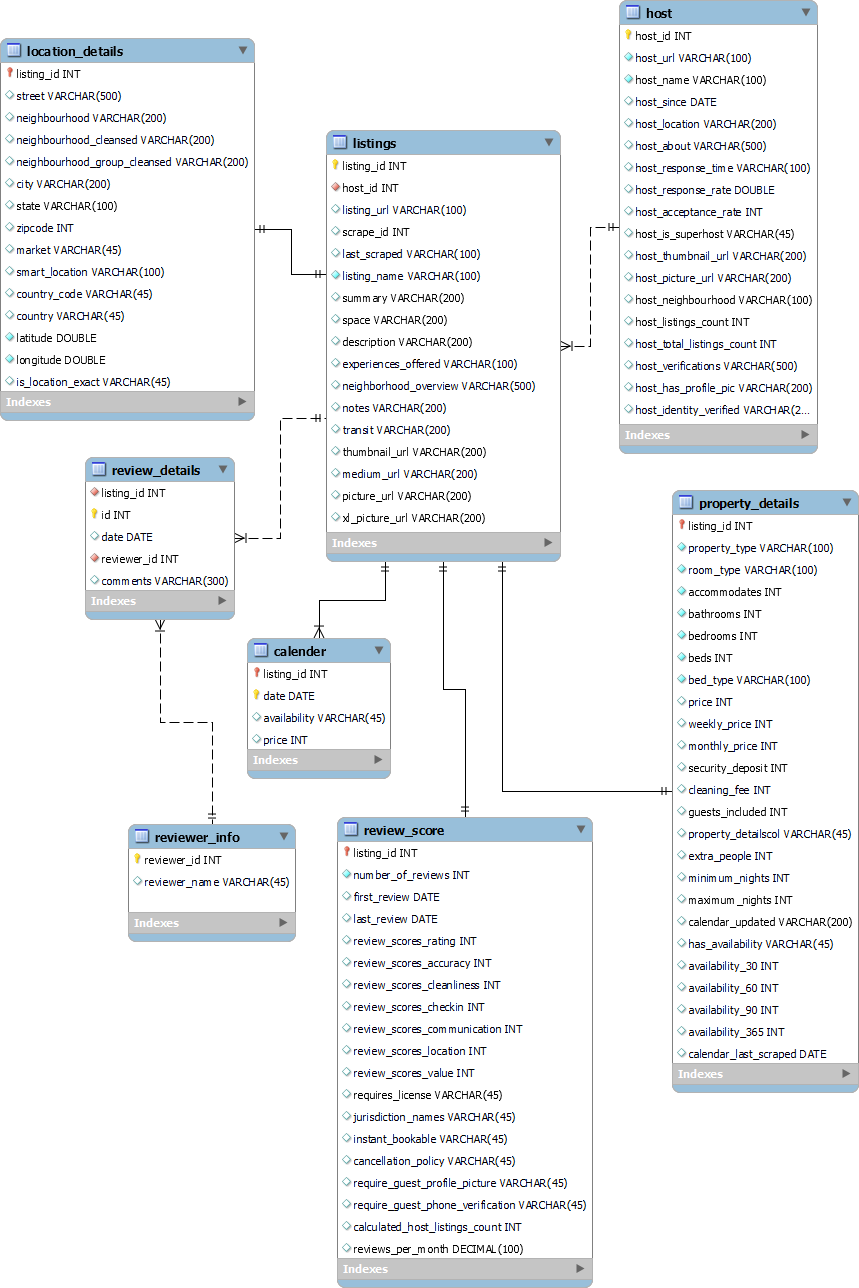
For 3NF, all table should obey the following rules,

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

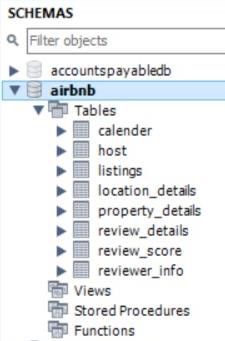


# Physical MySQL Database

## 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 |