Name: goodreads2

mysql> SHOW TABLES;

+----------------------+

| Tables\_in\_goodreads2 |

+----------------------+

| adds\_to\_wishlist |

| author |

| book |

| genre |

| is\_part\_of |

| is\_similar\_to |

| is\_tagged\_as |

| is\_written\_by |

| series |

| user |

| work |

+----------------------+

11 rows in set (0.00 sec)

## adds\_to\_wishlist

mysql> DESCRIBE adds\_to\_wishlist;

+---------+--------+------+-----+---------+-------+

| Field | Type | Null | Key | Default | Extra |

+---------+--------+------+-----+---------+-------+

| User\_ID | bigint | NO | PRI | NULL | |

| Book\_ID | bigint | NO | PRI | NULL | |

+---------+--------+------+-----+---------+-------+

2 rows in set (0.01 sec)

This is the connection from User to Book. It was not imported, since we are creating the users, and is mostly empty.

## author

mysql> DESCRIBE author;

+-----------------------+--------------+------+-----+---------+-------+

| Field | Type | Null | Key | Default | Extra |

+-----------------------+--------------+------+-----+---------+-------+

| Author\_ID | int | NO | PRI | NULL | |

| Author\_Name | varchar(255) | YES | | NULL | |

| Author\_Average\_Rating | decimal(3,2) | YES | | NULL | |

| Average\_Ratings\_Count | bigint | YES | | NULL | |

+-----------------------+--------------+------+-----+---------+-------+

4 rows in set (0.00 sec)

This was imported from a 26,103 KB CSV (tab-delimited) file. There are 829, 529 rows.

## book

mysql> DESCRIBE book;

+----------------------+----------------+------+-----+---------+-------+

| Field | Type | Null | Key | Default | Extra |

+----------------------+----------------+------+-----+---------+-------+

| Book\_ID | bigint | NO | PRI | NULL | |

| Book\_Title | varchar(255) | YES | | NULL | |

| Book\_Description | varchar(15000) | YES | | NULL | |

| Book\_Average\_Rating | decimal(3,2) | YES | | NULL | |

| Book\_Ratings\_Count | bigint | YES | | NULL | |

| Book\_Format | varchar(255) | YES | | NULL | |

| Book\_Number\_Of\_Pages | bigint | YES | | NULL | |

| Book\_Image | varchar(255) | YES | | NULL | |

| Work\_ID | bigint | YES | | NULL | |

+----------------------+----------------+------+-----+---------+-------+

9 rows in set (0.00 sec)

This was imported from a 1,939,574 KB CSV (tab-delimited) file. There were 2,360,655 rows, but I’ve been working on removing lines that have non-roman letters (e.g., Arabic, Cyrillic, Chinese, etc.) to simplify, so it now has 2,358,140 rows.

## genre

mysql> DESCRIBE genre;

+------------+--------------+------+-----+---------+-------+

| Field | Type | Null | Key | Default | Extra |

+------------+--------------+------+-----+---------+-------+

| Genre\_Name | varchar(255) | NO | PRI | NULL | |

+------------+--------------+------+-----+---------+-------+

1 row in set (0.00 sec)

This was not imported. It has 10 rows, one for each of the genres used in the dataset: “children”, “comics, graphic”, “fantasy, paranormal”, “fiction”, “history, historical fiction, biography”, “mystery, thriller, crime”, “non-fiction”, “poetry”, “romance”, “young-adult”.

## is\_part\_of

mysql> DESCRIBE is\_part\_of;

+-----------+------+------+-----+---------+-------+

| Field | Type | Null | Key | Default | Extra |

+-----------+------+------+-----+---------+-------+

| Book\_ID | int | NO | PRI | NULL | |

| Series\_ID | int | NO | PRI | NULL | |

+-----------+------+------+-----+---------+-------+

2 rows in set (0.00 sec)

This is the connection from Book to Series. This was imported from a 13,371 KB CSV (comma delimited). It has 822,460 rows.

## is\_similar\_to

mysql> DESCRIBE is\_similar\_to;

+-----------+--------+------+-----+---------+-------+

| Field | Type | Null | Key | Default | Extra |

+-----------+--------+------+-----+---------+-------+

| Book\_ID\_1 | bigint | NO | PRI | NULL | |

| Book\_ID\_2 | bigint | NO | PRI | NULL | |

+-----------+--------+------+-----+---------+-------+

2 rows in set (0.00 sec)

This is the connection from Book to another Book. This was imported from a 285,567 KB CSV (comma delimited). It has 16,809,717 rows.

## is\_tagged\_as

mysql> DESCRIBE is\_tagged\_as;

+------------------------------+--------------+------+-----+---------+-------+

| Field | Type | Null | Key | Default | Extra |

+------------------------------+--------------+------+-----+---------+-------+

| Book\_ID | bigint | NO | PRI | NULL | |

| Genre\_Name | varchar(255) | NO | PRI | NULL | |

| Is\_Tagged\_As\_Number\_Of\_Times | bigint | YES | | NULL | |

+------------------------------+--------------+------+-----+---------+-------+

3 rows in set (0.00 sec)

This is the connection from Book to Genre. This was imported from a 139,321 KB CSV (tab delimited). It has 5,042,542 rows.

## is\_written\_by

mysql> DESCRIBE is\_written\_by;

+-----------+--------+------+-----+---------+-------+

| Field | Type | Null | Key | Default | Extra |

+-----------+--------+------+-----+---------+-------+

| Book\_ID | bigint | NO | PRI | NULL | |

| Author\_ID | bigint | NO | PRI | NULL | |

+-----------+--------+------+-----+---------+-------+

2 rows in set (0.00 sec)

This is the connection from Book to Author. This was imported from a 54,317 KB CSV (comma delimited). It has 3,323,621 rows.

## series

mysql> DESCRIBE series;

+---------------------------+---------------+------+-----+---------+-------+

| Field | Type | Null | Key | Default | Extra |

+---------------------------+---------------+------+-----+---------+-------+

| Series\_ID | bigint | NO | PRI | NULL | |

| Series\_Name | varchar(255) | YES | | NULL | |

| Series\_Description | varchar(8000) | YES | | NULL | |

| Series\_Primary\_Work\_Count | bigint | YES | | NULL | |

+---------------------------+---------------+------+-----+---------+-------+

4 rows in set (0.00 sec)

This was imported form a 53,425 KB CSV (tab delimited). It has 400,390 rows.

## user

mysql> DESCRIBE user;

+---------------+--------------+------+-----+---------+----------------+

| Field | Type | Null | Key | Default | Extra |

+---------------+--------------+------+-----+---------+----------------+

| User\_ID | bigint | NO | PRI | NULL | auto\_increment |

| User\_Name | varchar(255) | NO | UNI | NULL | |

| User\_Login | varchar(255) | YES | | NULL | |

| User\_Password | varchar(255) | YES | | NULL | |

+---------------+--------------+------+-----+---------+----------------+

4 rows in set (0.00 sec)

This was not imported since we create the users. It only has a few rows in it.

## work

mysql> DESCRIBE work;

+----------------------------+-------------+------+-----+---------+-------+

| Field | Type | Null | Key | Default | Extra |

+----------------------------+-------------+------+-----+---------+-------+

| Work\_ID | bigint | NO | PRI | NULL | |

| Work\_Ratings\_Sum | bigint | YES | | NULL | |

| Work\_Ratings\_Count | bigint | YES | | NULL | |

| Work\_Original\_Publish\_Date | varchar(10) | YES | | NULL | |

+----------------------------+-------------+------+-----+---------+-------+

4 rows in set (0.00 sec)

This was imported from a 31,383 CSV (tab-delimited). It has 1,521,962 rows.

## Original JSON.GZ files

The JSON files that were parsed to create the CSV files are as follows.

|  |  |
| --- | --- |
| goodreads\_book\_authors.json.gz | 17,457 KB |
| goodreads\_book\_genres\_initial.json.gz | 23,686 KB |
| goodreads\_book\_series.json.gz | 27,612 KB |
| goodreads\_book\_works.json.gz | 73,145 KB |
| goodreads\_books.json.gz | 2,029,883 KB |