pip install pymongo[srv] pip install PyMySQL

Fetch data from MongoDb

In [1]:
```
from pymongo import MongoClient
import pandas as pd
from sqlalchemy import create_engine
from random import randint
import pymysql
```

In [2]:
```
client = MongoClient('localhost',27017)
```

In [3]:
```
client
```

Out[3]:
```
MongoClient(host=['localhost:27017'], document_class=dict, tz_aware=False, connect=True)
```

In [4]:
```
db = client['TheatreDb']
```

In [5]:
```
engine = create_engine("mysql+pymysql://{user}:{pw}@localhost/{db}".format(user="root",pw="password",db="theatre"))
print(engine)
```

```
Engine(mysql+pymysql://root:***@localhost/theatre)
```

Getting data form indivisual collections in TheatreDb to populate respective tables.

In [ ]:
```
```

In [6]:
```
users = db['users']

users_cursor = users.find()
users_df = pd.DataFrame(list(users_cursor))

users_df.columns
users_df = users_df[['_id', 'full_name', 'email', 'phone_number', 'address', 'gender', 'role']]

users_df = users_df.rename(columns = {'_id':'user_id'})

#display(users_df)
users_df.to_sql('user', con = engine, if_exists = 'append', chunksize = 1000, index= False)
```

Out[6]:
```
297
```

In [7]:
```
movies = db['movies']

movies_cursor = movies.find()
movies_df = pd.DataFrame(list(movies_cursor))

movies_df.columns
movies_df = movies_df[['_id', 'title', 'overview', 'duration','released_date', 'original_language','age_type']]

movies_df = movies_df.rename(columns = {'_id':'Movie_id','title':'Title','overview':'Overview','duration':'Duration',
                                        'released_date':'Release_date'})


movies_df.to_sql('movie', con = engine, if_exists = 'append', chunksize = 1000,index= False)
```

Out[7]:
```
330
```

In [8]:
```
theatres = db['theatres']

theatres_cursor = theatres.find()
theatres_df = pd.DataFrame(list(theatres_cursor))

theatres_df.columns
theatres_df = theatres_df[['_id', 'name', 'address','phone_number', 'email', 'opening_hours']]

theatres_df = theatres_df.rename(columns = {'_id':'Theatre_id','name':'Theatre_name','address':'Theatre_Address'
                                            ,'phone_number':'contact_number'})

theatres_df.to_sql('theatre', con = engine, if_exists = 'append', chunksize = 1000, index= False)
```

Out[8]:
```
5
```

In [9]:
```
shows = db['show_times']

shows_cursor = shows.find()
shows_df = pd.DataFrame(list(shows_cursor))

shows_df.columns
shows_df = shows_df[['_id', 'movie', 'theatre','start_time', 'end_time', 'room']]

shows_df = shows_df.rename(columns = {'_id':'Showtime_id','movie':'Movie_id','theatre':'Theatre_id','start_time':'Start_
                                      'end_time':'End_time'})

shows_df.to_sql('showtime', con = engine, if_exists = 'append', chunksize = 1000, index= False)
```

Out[9]: 1722

In [10]:
```
#payment table population
payments = db['reservations']

payments_cursor = payments.find()
payments_df = pd.DataFrame(list(payments_cursor))

payments_df.columns
payments_df = payments_df[['payment_intent_id']]
payments_df['card_number'] = [randint(1000000000,9999999999) for i in payments_df.index]

payments_df = payments_df.rename(columns = {'payment_intent_id':'Payment_id'})
#display(payments_df)
payments_df.to_sql('payment', con = engine, if_exists = 'append', chunksize = 1000, index= False)
```

Out[10]: 74

In [11]:
```
bookings = db['reservations']

bookings_cursor = bookings.find()
bookings_df = pd.DataFrame(list(bookings_cursor))

bookings_df.columns
bookings_df = bookings_df[['_id', 'user', 'show_time', 'total_price', 'payment_intent_id']]

bookings_df = bookings_df.rename(columns = {'_id':'Booking_id','user':'user_id',
                                            'show_time':'Showtime_id','payment_intent_id':'Payment_id'})
#display(bookings_df)
bookings_df.to_sql('booking', con = engine, if_exists = 'append', chunksize = 1000, index= False)
```

Out[11]: 74

In [ ]:
```

```

In [12]:
```
ratings = db['comments']

ratings_cursor = ratings.find()
ratings_df = pd.DataFrame(list(ratings_cursor))

ratings_df.columns
ratings_df = ratings_df[['_id', 'user','movie', 'content', 'rate_star']]

ratings_df = ratings_df.rename(columns = {'_id':'Rating_id','user':'user_id','movie':'Movie_id','content':'comment'})

ratings_df.to_sql('rating', con = engine, if_exists = 'append', chunksize = 1000, index= False)
```

Out[12]: 8471

In [ ]:
```

```

In [13]:
```
genres = db['categories']

genres_cursor = genres.find()
genres_df = pd.DataFrame(list(genres_cursor))

genres_df.columns
genres_df = genres_df[['_id', 'name']]

genres_df = genres_df.rename(columns = {'_id':'Genre_id','name':'Name'})

genres_df.to_sql('genre', con = engine, if_exists = 'append', chunksize = 1000, index= False)
```

Out[13]: 19

In [ ]:
```

```

In [14]:
```
genres_relation = db['movie_category']

genres_relation_cursor = genres_relation.find()
genres_relation_df = pd.DataFrame(list(genres_relation_cursor))

genres_relation_df.columns
genres_relation_df = genres_relation_df[['movie_id','category_id']]


genres_relation_df = genres_relation_df.rename(columns = {'movie_id':'Movie_id','category_id':'Genre_id'})
genres_relation_df = genres_relation_df.dropna()

genres_relation_df = movies_df.merge(genres_relation_df, on='Movie_id', indicator=True)
genres_relation_df = genres_relation_df[['Movie_id','Genre_id']]
#display(genres_relation_df)

genres_relation_df.to_sql('movie_genre', con = engine, if_exists = 'append', chunksize = 1000, index= False)
```

Out[14]: 536

In [15]:
```
casts = db['people']

casts_cursor = casts.find()
casts_df = pd.DataFrame(list(casts_cursor))

casts_df.columns
casts_df = casts_df[['_id', 'full_name']]
casts_df = casts_df.rename(columns = {'_id':'Cast_id'})

casts_df.to_sql('movie_cast', con = engine, if_exists = 'append', chunksize = 1000, index= False)
```

Out[15]: 4721

In [16]:
```
directors = db['movies']

directors_cursor = directors.find()
directors_df = pd.DataFrame(list(directors_cursor))

directors_df.columns
directors_df = directors_df[['_id', 'directors']]
directors_df = directors_df.explode('directors')

directors_df = directors_df.rename(columns = {'_id':'Movie_id','directors':'Director_id'})
directors_df = directors_df.dropna()
#display(directors_df)
directors_df.to_sql('movie_director', con = engine, if_exists = 'append', chunksize = 1000, index= False)
```

Out[16]: 338

In [17]:
```
actors = db['movies']

actors_cursor = actors.find()
actors_df = pd.DataFrame(list(actors_cursor))

actors_df.columns
actors_df = actors_df[['_id', 'actors']]
actors_df = actors_df.explode('actors')


actors_df = actors_df.rename(columns = {'_id':'Movie_id','actors':'Actor_id'})
actors_df = actors_df.dropna()

actors_df.to_sql('movie_actor', con = engine, if_exists = 'append', chunksize = 1000, index= False)
```

Out[17]: 2202

In [19]:
```
theatre_room = db['theatres']

theatre_room_cursor = theatre_room.find()
theatre_room_df = pd.DataFrame(list(theatre_room_cursor))

theatre_room_df.columns
theatre_room_df = theatre_room_df[['_id','rooms']]
theatre_room_df = theatre_room_df.explode('rooms')
theatre_room_df['room'] = theatre_room_df["rooms"].str.split()
theatre_room_df['room_type'] = theatre_room_df["room"].str[0]
theatre_room_df['room_no'] = theatre_room_df["room"].str[1]

theatre_room_df = theatre_room_df[['_id','room_no','room_type']]
theatre_room_df.loc[theatre_room_df['room_type']=='3D','room_no'] = 8

theatre_room_df = theatre_room_df.rename(columns = {'_id':'Theatre_id','room_no':'room_number'})

#theatre_room_df = theatre_room_df.drop_duplicates(subset=['Theatre_id'])
#display(theatre_room_df)

theatre_room_df.to_sql('theatre_room', con = engine, if_exists = 'append', chunksize = 1000, index= False)
```

Out[19]: 18

In [20]:
```
seats = db['seats']

seats_cursor = seats.find()
seats_df = pd.DataFrame(list(seats_cursor))

seats_df.columns
seats_df = seats_df[['_id', 'theatre', 'room', 'row','column']]

seats_df = seats_df.rename(columns = {'_id':'Seat_id','theatre':'Theatre_id','row':'row_','column':'column_'})

seats_df.to_sql('seats', con = engine, if_exists = 'append', chunksize = 1000, index= False)
```

Out[20]: 1761

In [ ]:
```

```

In [21]:
```
tickets = db['tickets']

tickets_cursor = tickets.find()
tickets_df = pd.DataFrame(list(tickets_cursor))

tickets_df.columns
tickets_df = tickets_df[['_id','seat', 'show_time', 'reservation', 'price']]

# ab = tickets_df.groupby(['reservation'])
# print(ab.first())

tickets_df = tickets_df.rename(columns = {'_id':'Ticket_id','seat':'Seat_id','show_time':'Showtime_id',
                                          'reservation':'Booking_id','price':'Price'
                                         })

tickets_df = seats_df.merge(tickets_df, on='Seat_id', indicator=True)
tickets_df = tickets_df[['Ticket_id','Seat_id','Showtime_id','Booking_id','Price']]


tickets_df = shows_df.merge(tickets_df, on='Showtime_id', indicator=True)
tickets_df = tickets_df[['Ticket_id','Seat_id','Showtime_id','Booking_id','Price']]

#display(tickets_df)

tickets_df.to_sql('ticket', con = engine, if_exists = 'append', chunksize = 1000, index= False)
```

Out[21]: 120462

In [ ]:
```

```