Data Warehousing and Business Intelligence

Assignment 01
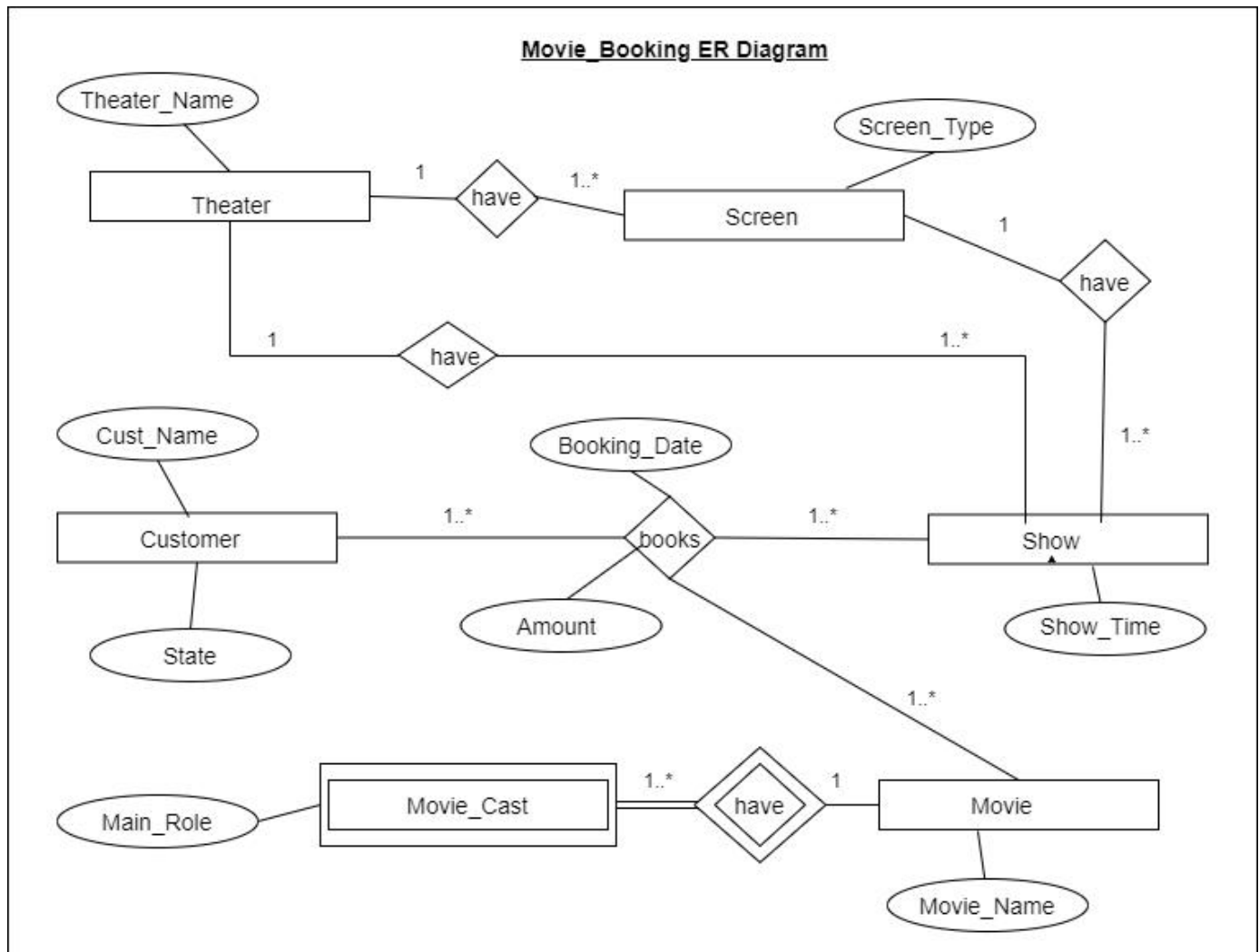
Batch : Year 3 Semester 1 Weekend Malabe (Y3S1.04)

	Student Registration Number	Student Name
1	IT19109886	N.K.GALAPPATHTHI

SLIIT
Discover Your Future

BSc (Hons) in Information Technology
Specialization in Data Science
Assignment

IT3021 – DWBI   **Year 3, Semester 1, 2021**

1) DataSet :

❖ Movie_Booking is a data set that includes booking details about customers for seats in movie theaters from 2019 in US. This data has taken out from a online movie booking platform that provides booking for limited number of theaters in a State. Mainly this data set had 2 CSV files about bookings and related data. For the purpose of this project I have distributed this data along 2 files(CSV and text) and 5 tables.
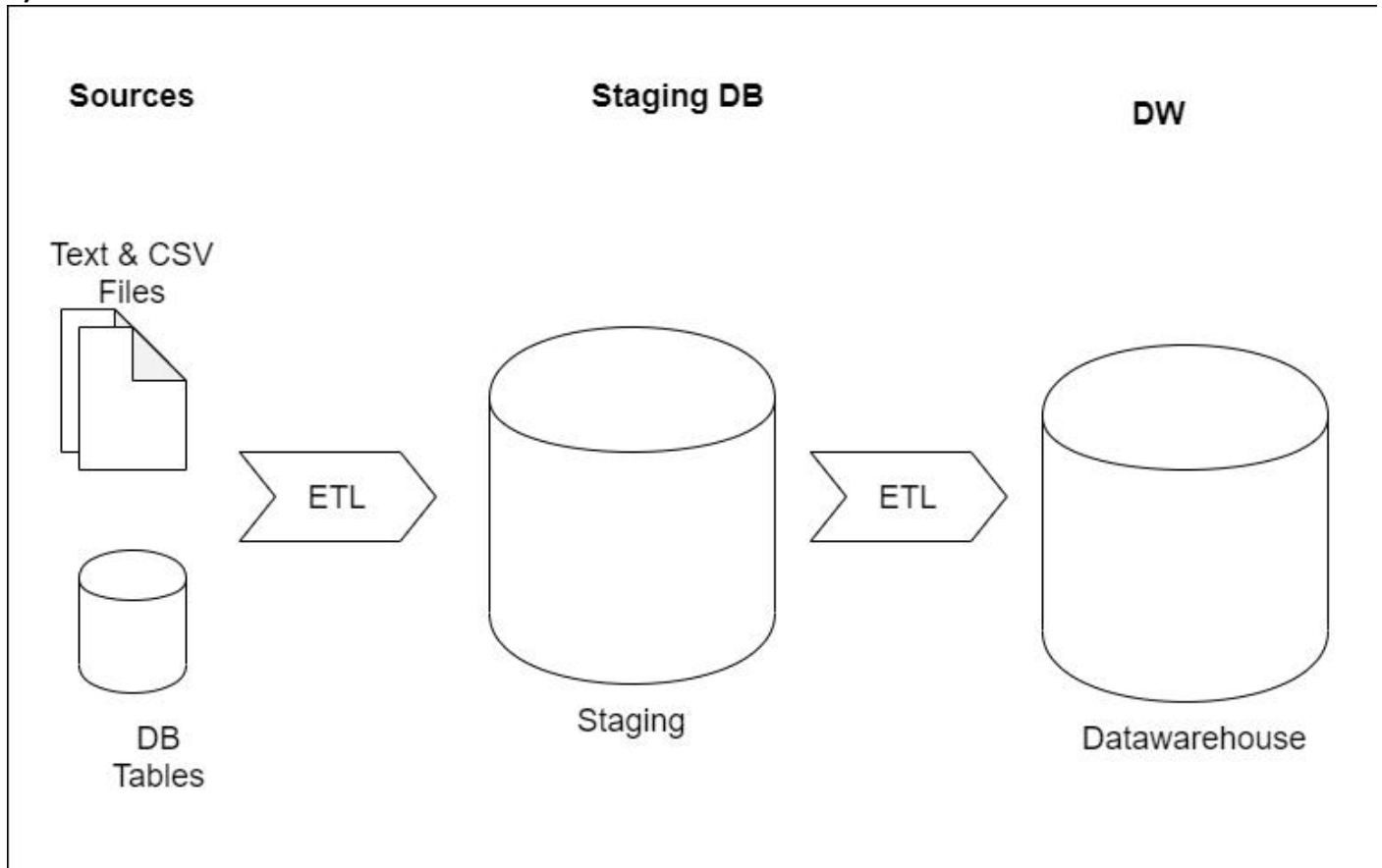
ER DIAGRAM

Movie_Booking ER Diagram

* Only few attributes are displayed in above diagram

2) . Data Sources :

❖ Files
◆ Customer data - CSV file (Customer.csv)
◆ Movie Cast data - Text file (Movie_cast.txt)

❖ Tables
◆ Movie data - DB Table
◆ Theater data - DB table
◆ Screens data - DB table
◆ Shows data - DB table
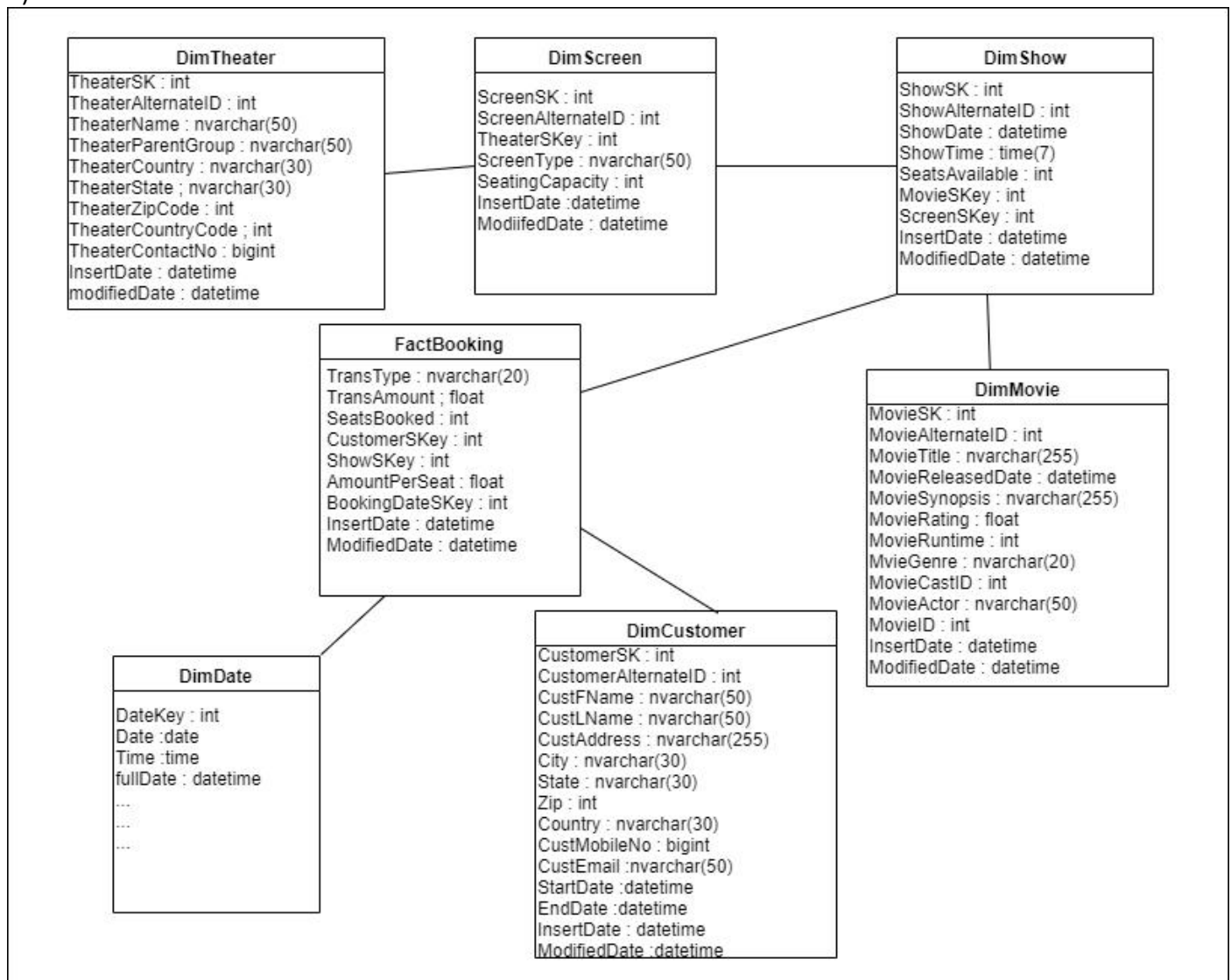◆ Booking data - DB table

3)

❖ Text and CSV files includes source data in file format. Database tables includes data in table format. Both these data sources are source data for staging layer.

❖ ETL(Extract Transform Load) process handled by the SSIS . In this Process 90% work is just Extract

SLIIT
Discover Your Future

BSc (Hons) in Information Technology
Specialization in Data Science
Assignment

IT3021 – DWBI **Year 3, Semester 1, 2021**

and loading. Transformations are minimum. Data extract from all the data sources and load them into tables under one database. Data sources could be any type like APIs, Documents(XML,Json) , tables, files.

❖ Staging component is a temporary database. Updated or new inserted data is stored in this database. Data coming from source type stored in staging db.

❖ Most Transformations are happening in this ETL process. Data from staging are the source for this process. These source table data extracted by ETL process and transform them into dimension tables, load them in Data warehouse.

❖ Final processed data stored in this component.

4)

Snow flake Dimension model.

6 - Dimensions
● Customer Dimension (DimCustomer)
● Movie Dimension (DimMovie)
● Date Dimension (DimDate)
● Show Dimension (DimShow)
● Screen Dimension (DimScreen)
● Theater Dimension (DimTheater)

1 - Fact table
● FactBooking Fact table

Assumptions : Duplicates and Deletes are not happen in data warehouse.
Source records (In source databases) are not being deleted.
Last Execution date is stored separately.
SQL Server Agent handling the periodic execution of ETL processes using last execution date.

5)

A) ETL process 1 : (From Sources to Staging)

1) Load Theater data to staging
i. Extract data from Theater table
ii. Load data to Theater Table

2) Load Screen data to staging
i. Extract data from Screen table
ii. Load data to Screen table

3) Load Movie data to Staging
i. Extract data from Movie Table
ii. Load data to Movie table

4) Load Show data to staging
i. Extract data from Show table
ii. Load data to Show table

5) Load MovieCast to staging
i. Extract data from MovieCast file
ii. Load data to MovieCast table

6) Load Customer to staging
   i. Extract data from Customer text file
   ii. Load data to Customer table

7) Load Booking data to staging
   i. Extract data from Booking table
   ii. Load data to Booking table

B) ETL process 2 : (From Staging to Data warehouse)

1) Transform and Load Theater data
   i. Extract data from TheaterStaging table
   ii. Load data to DimTheater dimension

2) Transform and Load Screen data
   i. Extract data from ScreenStaging table
   ii. Theater surrogate key look up in Theater Dimension
   iii. Load data to DimScreen dimension

3) Transform and Load Movie Data
   i. Extract data from MovieStaging table
   ii. Sort data according to Movie ID
   iii. Extract data from MovieCastStaging table (parallel with step i)
   iv. Sort data according to FK_MovieID
   v. Merge join MovieStaging table and MovieCastStaging table
   vi. Load data to DimMovie dimension

4) Transform and Load Show data
   i. Extract data from ShowStaging table
   ii. Movie surrogate key look up
   iii. Screen surrogate key look up
   iv. Load data to DimShow dimension

5) Transform and Load Customer data
   i. Extract data from CustomerStaging table
   ii. Derived columns InserDate and ModifiedDate
   iii. Apply Slowly Changing Dimension component
   iv. Union All data outputs
   v. Load data to DimCustomer dimension

6) Transform and Load Booking fact
   i. Extract data from BookingStaging table
   ii. BookingDate column data type conversion to datetime

iii.   BookinDate surrogate key look up in DimDate dimension
iv.   CustomerID surrogate key look up DimCustomer dimension
v.   ShowID surrogate key look up in DimShow dimension
vi.   Derived columns InsertDate and ModifiedDate
vii.   Load data to BookingFact