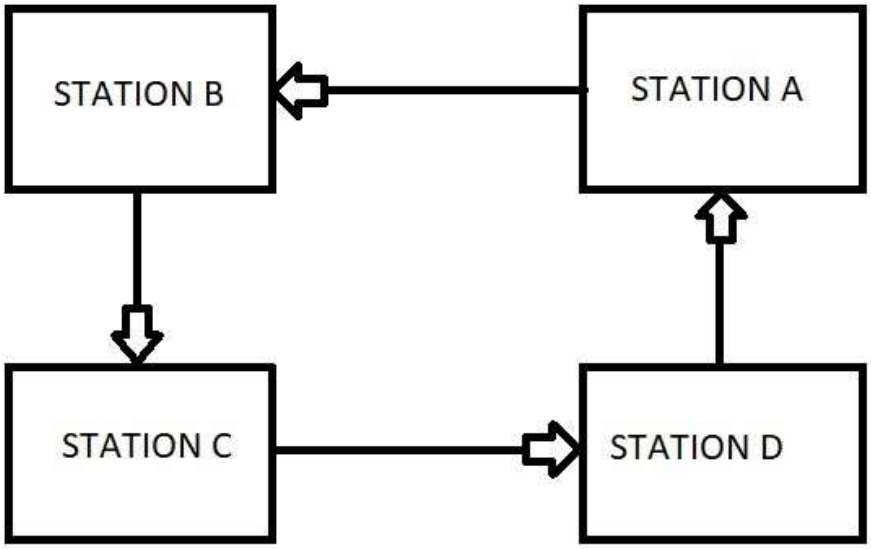
ECE 9014A – GROUP PROJECT

RAILWAY SYSTEM

PART 2: BOOKINGS DATA MART (10%) GROUP – 11

Aman Budhraja, Arpit Gupta, Rushika Jani, Sathyanath Nandakumar,
Srichaitanya Ravupalli

MODEL OF THE TRAIN SYSTEM



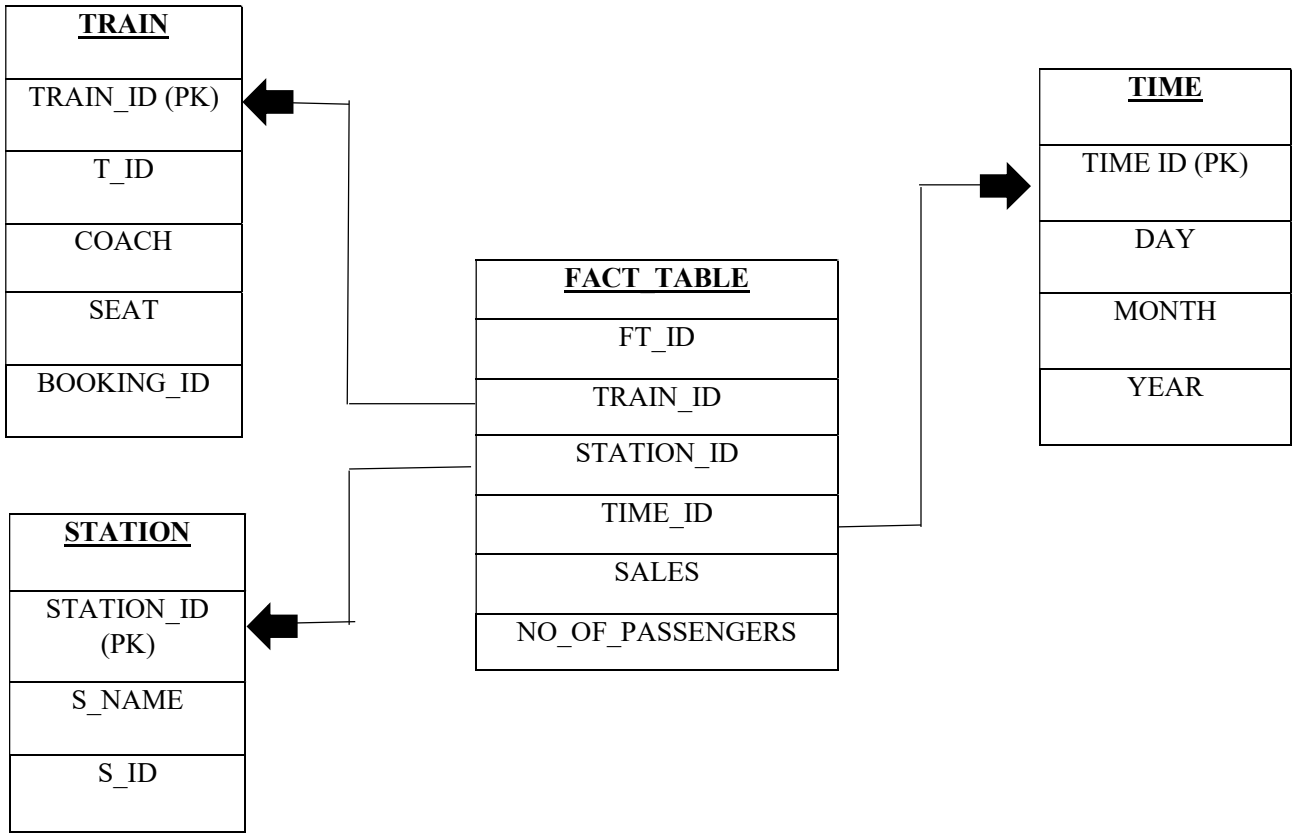
The following are the assumptions we have made:

- The 4 stations taken are connected in a single, unidirectional loop.
- 4 trains run along this loop.
- There is only 1 track between two consecutive stations.
- Train are operational only for the months of November and December, for the years 2018 and 2019.
- No two records have the same date.

Additional attributes taken include:

- Booking_Id: This attribute has been added because we have assumed that none of the columns can be uniquely defined in the passenger booking table. Thus, we assume that no two booking IDs are the same.
- Price: The price per seat, which varies between train and coach.

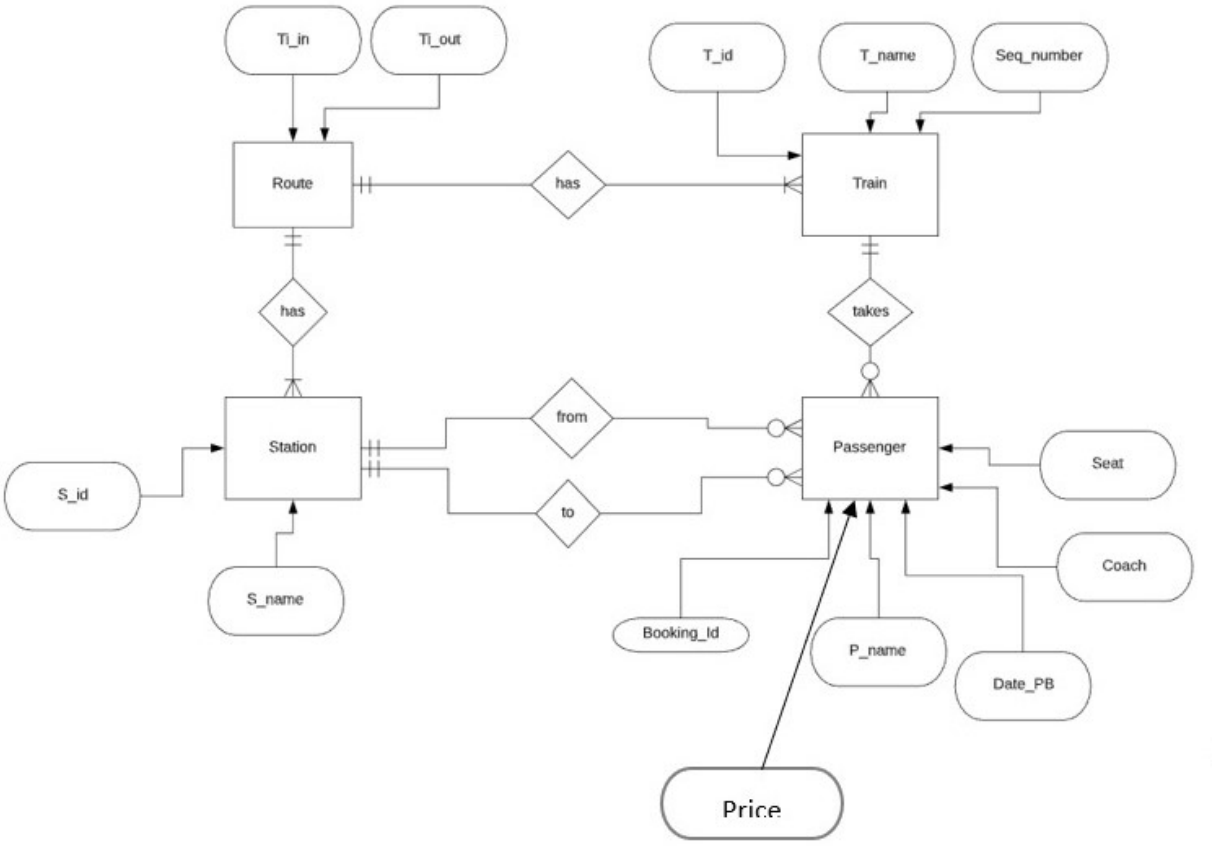
STAR SCHEMA



We have chosen a star schema for the following reasons:

- Our dimensional tables are fully normalized hence we do not need a snowflake schema to further normalize them.

ENTITY RELATIONSHIP DIAGRAM



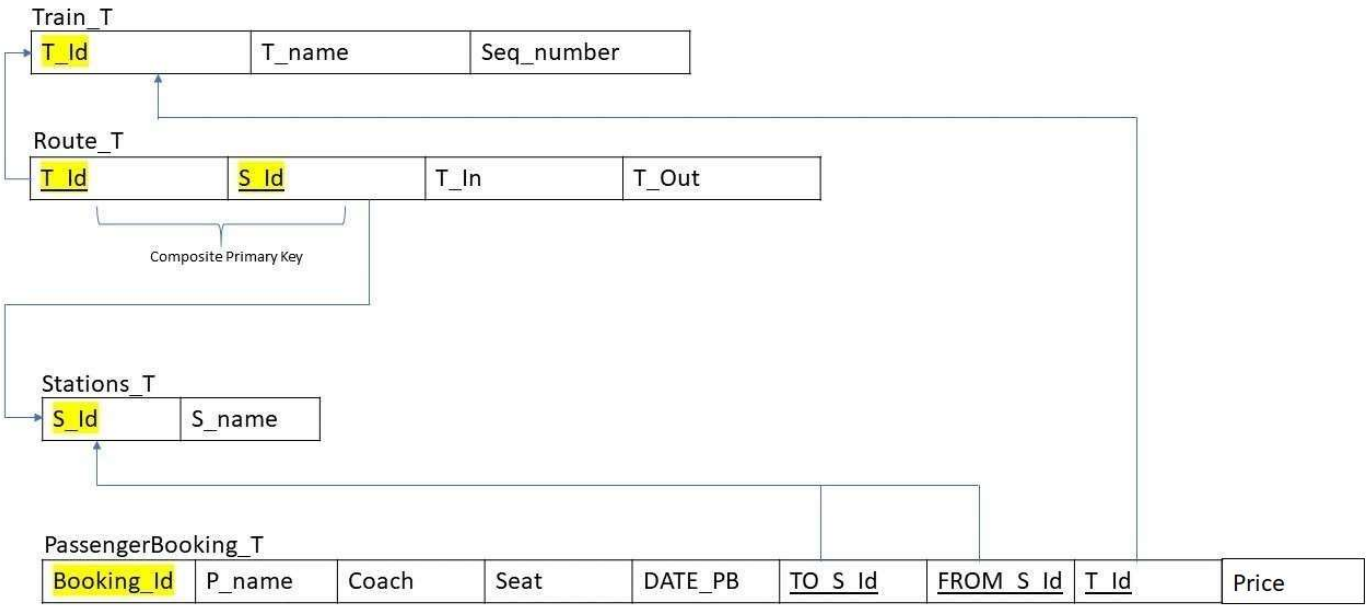
Representation Notes:

- We replaced the relationship between train and station that is train schedule, with an entity named route, since they have a many to many relationships.
- Hence route has the same attributes as the train schedule, that is, time in and time out.

Explanation:

- Every train has one and only one route while every route has at least one or many trains.
- Every station has one and only one route while every route has at least one or many stations.
- Every passenger departs from one and only one station and arrives at one and only one station
- Every station can have zero or many passengers.
- Every passenger booking is for one and only one train.
- Every train can be booked by zero or many passengers.

RELATIONAL MODEL



The relational model is already in third normal form.

Key for the relational model:

- Highlighted value is the primary key.
- Underlined value is the foreign key.
- Highlighted and underlined value is both a primary and a foreign key with respect to its corresponding tables.

SQL SCHEMA

Table definitions scripts:

1. To create table Train_T :

```
create table Train_T
(T_id varchar(12) NOT NULL,
T_name VARCHAR(20),
Seq_number VARCHAR(20),
CONSTRAINT Train_PK PRIMARY KEY (T_id));
```

2. To create table Station_T :

```
create table Station_T
(S_id VARCHAR(12) NOT NULL,
S_name VARCHAR(20),
CONSTRAINT Station_PK PRIMARY KEY (S_id));
```

3. To create table Route_T :

```
create table Route_T
(T_id varchar(12) NOT NULL,
S_id varchar(12) NOT NULL,
T_in TIME,
T_out TIME,
CONSTRAINT Route_PK PRIMARY KEY (T_id,S_id),
CONSTRAINT Route_FK1 FOREIGN KEY (T_id) REFERENCES Train_T(T_id),
CONSTRAINT Route_FK2 FOREIGN KEY (S_id) REFERENCES Station_T(S_id) );
```

4. To create table PassengerBooking_T:

```
create table PassengerBooking_T
(Booking_Id VARCHAR(30) NOT NULL,
P_name VARCHAR(30) NOT NULL,
Coach VARCHAR(3),
Seat int(2),
DATE_PB DATE,
TO_S_Id varchar(12) NOT NULL,
FROM_S_Id varchar(12) NOT NULL,
PB_T_Id varchar(12) NOT NULL,
Price int(8),
CONSTRAINT PassengerBooking_PK PRIMARY KEY (Booking_Id),
CONSTRAINT PassengerBooking_FK1 FOREIGN KEY (PB_T_Id) REFERENCES Train_T(T_id),
CONSTRAINT PassengerBooking_FK2 FOREIGN KEY (TO_S_Id) REFERENCES Station_T(S_id),
CONSTRAINT PassengerBooking_FK3 FOREIGN KEY (FROM_S_Id) REFERENCES Station_T(S_id));
```

5. Inserting values into the tables:

- insert into Train_T values ('T1','ALL_STOP','ABCD'),('T2','EXPRESS_12','ACD'),('T3','EXPRESS_14','AD'),
  ('T4','EXPRESS_90','BA');
- insert into Station_T values ('S_A','TOWN1'),('S_B','TOWN2'),('S_C','TOWN3'),('S_D','TOWN4');
- insert into Route_T values('T1','S_A','06:30:00','06:45:00'),
  ('T1','S_B','06:50:00','07:05:00'),('T1','S_C','07:10:00','07:25:00'),('T1','S_D','07:30:00','07:45:00'),
  ('T2','S_A','07:05:00','07:20:00'),('T2','S_B','07:25:00','07:25:00'),('T2','S_C','07:30:00','07:45:00'),('T2','S_D','07:50:00','08:05:00'),('T3','S_A','08:00:00','08:15:00'),
  ('T3','S_B','08:20:00','08:20:00'),('T3','S_C','08:25:00','08:25:00'),('T3','S_D','08:30:00','08:45:00'),('T4','S_B','09:00:00','09:15:00'),('T4','S_C','09:20:00','09:20:00'),
  ('T4','S_D','09:25:00','09:25:00'),('T4','S_A','09:30:00','09:45:00');
- insert into PassengerBooking_T values ('B001','Arpit','AC1','32','2019-12-10','S_A','S_D','T3','250'),
  ('B002','Mashrukh','SC','01','2018-11-08','S_B','S_A','T4','400'),
  ('B003','Kushwant','AC1','44','2018-12-24','S_B','S_D','T1','500'),
  ('B004','Heena','SC','02','2016-09-11','S_B','S_A','T4','650'),
  ('B005','Sandra','AC6','24','2019-11-08','S_A','S_D','T2','1000');
- insert into PassengerBooking_T values ('B006','Luong','AC1','11','2016-07-23','S_C','S_D','T1','10'),('B007','Tamara','AC2','29','2018-12-16','S_A','S_D','T1','50'),
  ('B008','Thomas','AC1','47','2019-12-25','S_A','S_C','T1','100');

CREATING A DATA WAREHOUSE

1. To create a train table:

```
create table train
(Train_id int(12) NOT NULL auto_increment,
Coach VARCHAR(20),
Seat int(3),
T_id VARCHAR(20),
Booking_id varchar(30),
CONSTRAINT Train_PK PRIMARY KEY (Train_id),
CONSTRAINT Train_FK1 FOREIGN KEY (T_id) REFERENCES project.Train_T(T_id),
CONSTRAINT Train_FK2 FOREIGN KEY (Booking_id) REFERENCES
project.PassengerBooking_T(Booking_Id)
);
```

2. To create a station table:

```
create table station
(Station_id int(12) NOT NULL auto_increment,
S_id varchar(20),
CONSTRAINT Station_PK PRIMARY KEY (Station_id),
CONSTRAINT Station_FK1 FOREIGN KEY (S_id) REFERENCES project.Station_T(S_id)
);
```

3. To create a table tim_dim:

```
create table time_dim
(Time_id int(12) NOT NULL auto_increment,
time_day varchar(10) NOT NULL,
time_month varchar(15),
time_year int(4),
CONSTRAINT Time_PK PRIMARY KEY (Time_id));
```

4. To create a table fact_table:

```
create table fact_table
(ft_id int(20) NOT NULL auto_increment,
train_id int(20) , station_id int(20),
 time_id int(20),
 Num_Passenger int(20),
 sales int(30),
CONSTRAINT fact_PK PRIMARY KEY (ft_id),
CONSTRAINT fact_FK1 FOREIGN KEY (train_id) REFERENCES train(Train_id),
CONSTRAINT fact_FK2 FOREIGN KEY (station_id) REFERENCES station(Station_id),
CONSTRAINT fact_FK3 FOREIGN KEY (time_id) REFERENCES time_dim(Time_id));
```

5. To auto increment values:

```
Alter table station auto_increment = 10;
Alter table train auto_increment = 100;
Alter table time_dim auto_increment = 10000;
```

INSERTING INTO THE DATA WAREHOUSE

1. Inserting into the station table :

```
Insert into station(S_id)
(
   Select  S_id
         From   project.Station_T
) ;
```

2. Inserting into the train table :

```
Insert into train(T_id,Coach,Seat,Booking_id)
(
   Select  a.T_id,b.Coach,b.Seat,b.Booking_Id
   From   project.Train_T a, project.PassengerBooking_T b
      where a.T_id = b.PB_T_Id
) ;
```

3. Inserting into tim_dim table :

```
Insert into time_dim(time_day, time_month, time_year) values ("1","11","2018"),("2","11","2018"),
("3","11","2018"),("4","11","2018"),("5","11","2018"),("6","11","2018"),("7","11","2018"),("8","11","2018"),
("9","11","2018"),("10","11","2018"),("11","11","2018"),("12","11","2018"),("13","11","2018"),("14","11","2018"),
("15","11","2018"),("16","11","2018"),("17","11","2018"),("18","11","2018"),("19","11","2018"),("20","11","2018"),
("21","11","2018"),("22","11","2018"),("23","11","2018"),("24","11","2018"),("25","11","2018"),("26","11","2018"),
("27","11","2018"),("28","11","2018"),("29","11","2018"),("30","11","2018"),("1","12","2018"),("2","12","2018"),
("3","12","2018"),("4","12","2018"),("5","12","2018"),("6","12","2018"),("7","12","2018"),("8","12","2018"),
("9","12","2018"),("10","12","2018"),("11","12","2018"),("12","12","2018"),("13","12","2018"),("14","12","2018"),
("15","12","2018"),("16","12","2018"),("17","12","2018"),("18","12","2018"),("19","12","2018"),("20","12","2018"),
("21","12","2018"),("22","12","2018"),("23","12","2018"),("24","12","2018"),("25","12","2018"),("26","12","2018"),
("27","12","2018"),("28","12","2018"),("29","12","2018"),("30","12","2018"),("31","12","2018");
```

```
Insert into time_dim(time_day, time_month, time_year) values ("1","11","2019"),("2","11","2019"),
("3","11","2019"),("4","11","2019"),("5","11","2019"),("6","11","2019"),("7","11","2019"),("8","11","2019"),
("9","11","2019"),("10","11","2019"),("11","11","2019"),("12","11","2019"),("13","11","2019"),("14","11","2019"),
("15","11","2019"),("16","11","2019"),("17","11","2019"),("18","11","2019"),("19","11","2019"),("20","11","2019"),
("21","11","2019"),("22","11","2019"),("23","11","2019"),("24","11","2019"),("25","11","2019"),
("26","11","2019"),("27","11","2019"),("28","11","2019"),("29","11","2019"),("30","11","2019"),("1","12","2019"),
("2","12","2019"),("3","12","2019"),("4","12","2019"),("5","12","2019"),("6","12","2019"),("7","12","2019"),
("8","12","2019"),("9","12","2019"),("10","12","2019"),("11","12","2019"),("12","12","2019"),("13","12","2019"),
("14","12","2019"),("15","12","2019"),("16","12","2019"),("17","12","2019"),("18","12","2019"),("19","12","2019"),
("20","12","2019"),("21","12","2019"),("22","12","2019"),("23","12","2019"),("24","12","2019"),("25","12","2019"),
("26","12","2019"),("27","12","2019"),("28","12","2019"),("29","12","2019"),("30","12","2019"),("31","12","2019");
```

4. Inserting into the fact table :

```
insert into fact_table (Num_Passenger, sales, train_id, station_id, time_id)
(
    Select count(*), sum(k.Price), a.Train_id, b.Station_id, d.Time_id
    from train a, station b, time_dim d, project.PassengerBooking_T k
    where d.time_day = Day(k.DATE_PB)
    and d.time_month = Month(k.DATE_PB)
    and d.time_year = Year(k.DATE_PB)
    and a.Booking_id = k.Booking_Id
    and a.T_id = k.PB_T_Id
    and b.S_id= k.TO_S_Id
    group by k.DATE_PB
    );
```

OUTPUT

The following rows do not show up because of the chosen time frame (Nov, Dec – 2018,2019)

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell

Booking_Id	P_name	Coach	Seat	DATE_PB	TO_S_Id	FROM_S_Id	PB_T_Id	Price
B001	Arpit	AC1	32	2019-12-10	S_A	S_D	T3	250
B002	Mashrukh	SC	1	2018-11-08	S_B	S_A	T4	400
B003	Kushwant	AC1	44	2018-12-24	S_B	S_D	T1	500
B004	Heena	SC	2	2016-09-11	S_B	S_A	T4	650
B005	Sandra	AC6	24	2019-11-08	S_A	S_D	T2	1000
B006	Luong	AC1	11	2016-07-23	S_C	S_D	T1	10
B007	Tamara	AC2	29	2018-12-16	S_A	S_D	T1	50
B008	Thomas	AC1	47	2019-12-25	S_A	S_C	T1	100
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

1. Output of the fact_table:

Result Grid | Filter Rows: | Edit:

ft_id	train_id	station_id	time_id	Num_Passenger	sales
1	7	11	10007	1	400
2	3	10	10045	1	50
3	1	11	10053	1	500
4	5	10	10068	1	1000
5	6	10	10100	1	250
6	4	10	10115	1	100
NULL	NULL	NULL	NULL	NULL	NULL

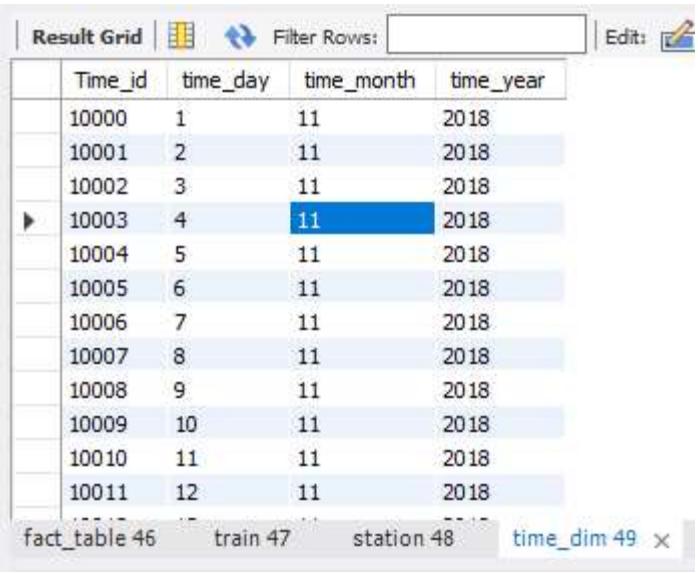
fact_table 45 ×

2. Output of the station table:

Result Grid | Filt

Station_id	S_id
10	S_A
11	S_B
12	S_C
13	S_D
NULL	NULL

3. Output of the time table:

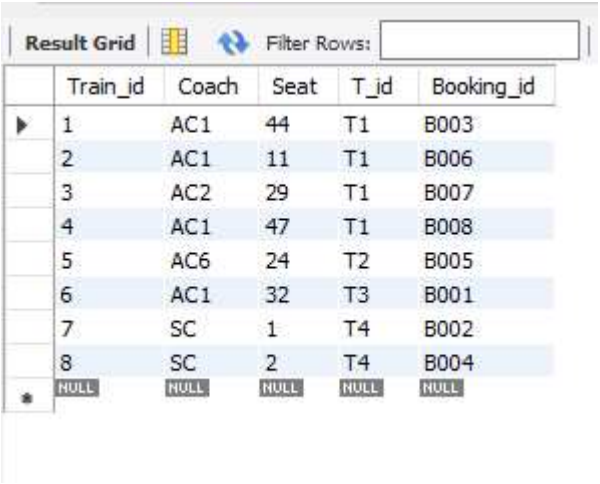


Result Grid | Filter Rows: | Edit:

Time_id	time_day	time_month	time_year
10000	1	11	2018
10001	2	11	2018
10002	3	11	2018
10003	4	11	2018
10004	5	11	2018
10005	6	11	2018
10006	7	11	2018
10007	8	11	2018
10008	9	11	2018
10009	10	11	2018
10010	11	11	2018
10011	12	11	2018

fact_table 46 | train 47 | station 48 | time_dim 49

4. Output of the train table:



Result Grid | Filter Rows:

Train_id	Coach	Seat	T_id	Booking_id
1	AC1	44	T1	B003
2	AC1	11	T1	B006
3	AC2	29	T1	B007
4	AC1	47	T1	B008
5	AC6	24	T2	B005
6	AC1	32	T3	B001
7	SC	1	T4	B002
8	SC	2	T4	B004
NULL	NULL	NULL	NULL	NULL