



**TIS3351 - ADVANCED DATABASE**

Assignment 1

**Data Warehouse for GrabCar**

SUBMISSION DATE: 22/08/2019

PREPARED BY:

NAME	ID	EMAIL
Baobaid, Husam Omar Ali	1161304070	1161304070@student.mmu.edu.my
Muhammad Wafi bin Othman	1161304576	1161304576@student.mmu.edu.my
Mansoor	1161303225	1161303225@student.mmu.edu.my
Mohab Mohamed	1161302776	1161302776@student.mmu.edu.my

## **TABLE OF CONTENTS**

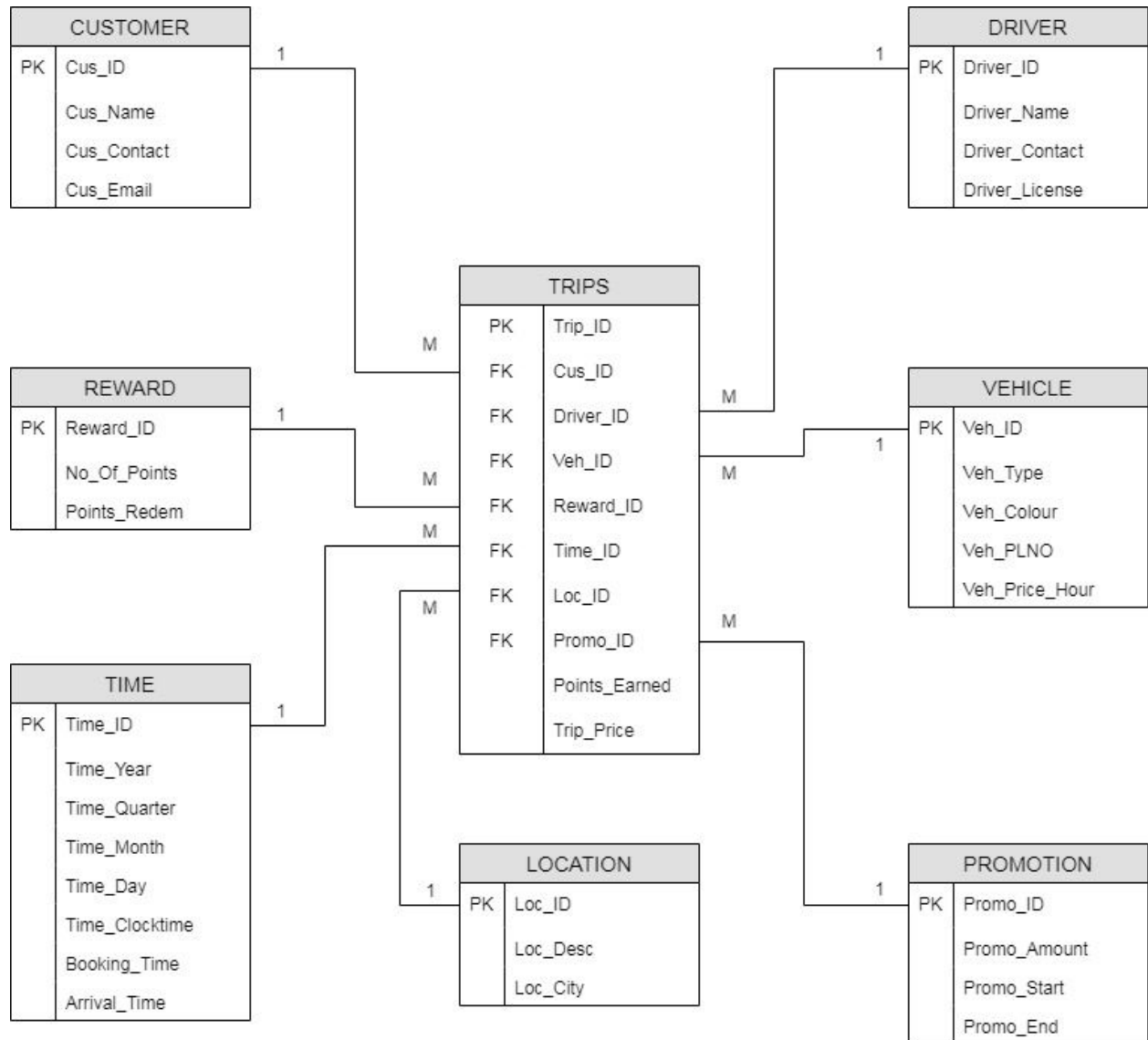
<b>1.Differences of operational database and data warehouse for the company</b>	<b>3</b>
<b>2.Conceptual Schema</b>	<b>5</b>
<b>3.Data Dictionary</b>	<b>6</b>
<b>4.Calculation for Fact Table Size and Storage</b>	<b>9</b>
<b>5.Data Definition Language(DDL)</b>	<b>10</b>
<b>6.Data Manipulation Language(DML)</b>	<b>13</b>

## 1.Differences of operational database and data warehouse for the company

CHARACTERISTIC	OPERATIONAL DATABASE DATA	DATA WAREHOUSE DATA
Integrated	Similar data can have different representations. Trip price may be in tens or hundreds. Year may be identified as dd/mm/yy or dd/mm/yyyy.	Provide a unified view of all data elements. The data found within the data warehouse is integrated. Since it comes from several operational systems, all inconsistencies must be removed. For example: in one department the status of booking maybe labeled as 'canceled' or 'confirmed' and in another department it is labeled as '1' or '2'. To avoid the potential format tangle, the data must conform to a common format acceptable through out the company. Once accomplished, it helps in enhancing decision making and better understanding of the company's operation.
Subject-oriented	Data are stored with a functional or process orientation. Data can be stored for invoices or payments. For example: customer may receive an invoice to his/her email containing booking details.	Subject orientation can be really useful for decision making. From TRIPS fact table, data can be recorded by Location, Driver or Customer. For example: we can view what type of car is favourite to the customers so that we can add more vehicles of that type, or which drivers took the most trips and so on.

Time-variant	Operational systems reflect current values as they support day-to-day operations, not used for decision making, promotions may be applied with the given validity date.	Data warehouse data represents data over a long time which means it stores historical data. It is mainly meant for data mining and forecasting, We can predict which months the cars are booked less so we can release promotions.
Nonvolatile	Is updated frequently, as customers earned points may change whenever they book a car.	Data cannot be changed, the data environment is relatively static, car plate number is never changed

## 2. Conceptual Schema



### 3.Data Dictionary

TABLE NAME	ATTRIBUTE NAME	CONTENTS	TYPE	FORMAT	RANGE	REQUIRED	PK OR FK	FK REFERENCED TABLE
TRIPS	Trip_ID	Trip identification	CHAR(6)	XXXXXX	NA	Y	PK	
	Cus_ID	Customer identification	CHAR(5)	XXXXX	NA	Y	FK	CUSTOMER
	Driver_ID	Driver identification	CHAR(4)	XXXX	NA	Y	FK	DRIVER
	Veh_ID	Vehicle identification	CHAR(4)	XXXX	NA	Y	FK	VEHICLE
	Reward_ID	Reward identification	CHAR(5)	XXXXX	NA	Y	FK	REWARD
	Time_ID	Time identification	CHAR(4)	XXXX	NA	Y	FK	TIME
	Loc_ID	Location identification	CHAR(4)	XXXX	NA	Y	FK	LOCATION
	Promo_ID	Promo identification	INTEGER	#####	0-99999		FK	PROMOTION
	Points_Earned	Points earned per trip	INTEGER	#####	0-99999		FK	REWARD
	Trip_Price	Trip price	DECIMAL(7,2)	#####.##	0.00-99999.99			
CUSTOMER	Cus_ID	Customer identification	CHAR(5)	XXXXX	NA	Y	PK	
	Cus_Name	Customer name	VARCHAR(30)	XXXXXXXX	NA	Y		
	Cus_Contact	Customer phone number	CHAR(11)	999-9999999	NA	Y		
	Cus_Email	Customer Email	VARCHAR(30)	xxxxxxxx	NA	Y		
DRIVER	Driver_ID	Driver identification	CHAR(4)	XXXX	NA	Y	PK	
	Driver_Name	Driver Name	VARCHAR(30)	XXXXXXXX	NA	Y		
	Driver_Contact	Driver phone number	CHAR(11)	999-9999999	NA	Y		

	Driver_License	Driver license	INTEGER	#####	0-999999	Y		
VEHICLE	Veh_ID	Vehicle identification	CHAR(4)	XXXX	NA	Y	PK	
	Veh_Type	Vehicle type	VARCHAR(10)	XXXXXXX	NA	Y		
	Veh_Colour	Vehicle colour	VARCHAR(10)	XXXXXXX	NA	Y		
	Veh_PLNO	Vehicle plate number	VARCHAR(7)	XXXXXXXX	NA	Y		
	Veh_Price_Hour	Vehicle price per hour	DECIMAL(5,2)	#####	0.00-999.99	Y		
PROMOTION	Promo_ID	Promotion Identification	INTEGER	#####	0-99999	Y	PK	
	Promo_amount	Promo amount	DECIMAL(5,2)	#####	0.00-999.99	Y		
	Promo_Start	Promo validity time	DATE	YYYY-MM-DD	NA	Y		
	Promo_End	Promo validity time	DATE	YYYY-MM-DD	NA	Y		
LOCATION	Loc_ID	Location identification	CHAR(4)	XXXX	NA	Y	PK	
	Loc_Desc	Location description	VARCHAR(30)	XXXXXXX	NA	Y		
	Loc_City	City name	VARCHAR(15)	XXXXXXX	NA	Y		
TIME	Time_ID	Time identification	CHAR(4)	XXXX	NA	Y	PK	
	Time_Year	Specified year	INTEGER	####	NA	Y		
	Time_Quarter	Quarter of the year	INTEGER	#	1-4	Y		
	Time_Month	Specified month	INTEGER	##	01-12	Y		
	Time_Day	Day	INTEGER	##	01-31	Y		
	Time_Clocktime	Clock Time	TIME	hh:mm	00:00-23:59	Y		
	Booking_Time	Time of booking	TIMESTAMP	yyyy-mm-dd hh:mm:ss	NA	Y		
	Arrival_Time	Time of arrival	TIMESTAMP	yyyy-mm-dd hh:mm:ss	NA	Y		

REWARD	Reward_ID	Reward identification	CHAR(4)	XXXX	NA	Y	PK	
	No_Of_Points	Points in customer account	INTEGER	# # # # #	0-999999	Y		
	Points_Redem	Points Redeemed	INTEGER	# # # # #	0-999999	Y		



## 4.Calculation for Fact Table Size and Storage

Dimension Tables: 7 Tables

CUSTOMER Table: 22 Rows

DRIVER Table: 20 Rows

VEHICLE Table: 20 Rows

PROMOTION Table: 20 Rows

LOCATION Table: 21 Rows

TIME Table: 20 Rows

REWARD Table: 20 Rows

**Size of Fact table:**  $22 \times 20 \times 20 \times 20 \times 21 \times 20 \times 20 = 1,478,400,000$  Rows

Fact Table (TRIPS):

Number of Attributes: 10 Attributes

Attributes	Data type	Size in bytes	Attributes	Data Type	Size in bytes
Trip_ID	CHAR(6)	6	Loc_ID	CHAR(4)	4
Cus_ID	CHAR(5)	5	Time_ID	CHAR(4)	4
Driver_ID	CHAR(4)	4	Reward_ID	CHAR(4)	4
Veh_ID	CHAR(4)	4	Points_Earned	INT	4
Promo_ID	INT	4	Trip_Price	DECIMAL(7,2)	$(7/2) + 1 = 4$

$$\text{Average field size} = \frac{6 + 5 + 4 + 4 + 4 + 4 + 4 + 4 + 4 + 4}{10} = 4.3 \text{ bytes}$$

$$\text{Size of storage of fact table (in bytes)} = 1,478,400,000 \times 10 \times 4.3 = 6.35712 \times 10^{10} \text{ bytes}$$

63.5712 GB

## 5.Data Definition Language(DDL)

### TRIPS

```
create table TRIPS
(
Trip_ID char(6) not null primary key,
Cus_ID char(5) not null,
Driver_ID char(4) not null,
Veh_ID char(4) not null,
Reward_ID char(4) not null,
Time_ID char(4) not null,
Loc_ID char(4) not null,
Promo_ID int,
Points_Earned int default 0,
Trip_Price decimal(7,2) default 0,
Foreign key (Cus_ID) references Customer,
Foreign key (Driver_ID) references Driver,
Foreign key (Veh_ID) references Vehicle,
Foreign key (Reward_ID) references Reward,
Foreign key (Time_ID) references Time,
Foreign key (Loc_ID) references Location
);
```

## CUSTOMER

```
create table CUSTOMER
(
Cus_ID char(5) not null primary key,
Cus_Name varchar(30) not null,
Cus_Contact char(10) not null,
Cus_Email varchar(30) not null
);
```

## DRIVER

```
create table DRIVER
(
Driver_ID char(4) not null primary key,
Driver_Name varchar(30) not null,
Driver_Contact char(10) not null,
Driver_License int not null
);
```

## VEHICLE

```
create table VEHICLE
(
Veh_ID char(4) not null primary key,
Veh_Type varchar(10) not null,
Veh_Colour varchar(10) not null,
Veh_PLNO varchar(7) not null,
Veh_Price_Hour decimal(5,2) not null
);
```

<p style="text-align: center;"><b>PROMOTION</b></p> <pre> create table PROMOTION ( Promo_ID int not null primary key, Promo_amount decimal(5,2) not null, Promo_Start date not null, Promo_End date not null ); </pre>	<p style="text-align: center;"><b>LOCATION</b></p> <pre> create table LOCATION ( Loc_ID char(4) not null primary key, Loc_Desc varchar(30) not null, Loc_City varchar(15) not null ); </pre>
<p style="text-align: center;"><b>TIME</b></p> <pre> create table TIME ( Time_ID char(4) not null primary key, Time_Year int not null, Time_Quarter int not null, Time_Month int not null, Time_Day int not null, Time_Clocktime time not null, Booking_Time timestamp not null, Arrival_Time timestamp not null ); </pre>	<p style="text-align: center;"><b>REWARD</b></p> <pre> create table REWARD ( Reward_ID char(4) not null primary key, No_Of_Points int not null, Points_Redem int not null ); </pre>

## 6.Data Manipulation Language(DML)

insert into TRIPS (Trip\_ID, Cus\_ID, Driver\_ID, Veh\_ID, Reward\_ID, Time\_ID, Loc\_ID, Promo\_ID) values

```
('TR0001','C0001','DR01','V001','R001','T001','L005',2019),
('TR0002','C0002','DR05','V005','R002','T002','L010',2015),
('TR0003','C0003','DR07','V007','R003','T003','L008',2094),
('TR0004','C0004','DR01','V001','R004','T004','L011',2183),
('TR0005','C0005','DR06','V006','R005','T005','L001',0179),
('TR0006','C0006','DR01','V010','R006','T006','L002',0191),
('TR0007','C0007','DR11','V011','R007','T007','L012',0192),
('TR0008','C0008','DR17','V017','R008','T008','L013',NULL),
('TR0009','C0009','DR19','V019','R009','T009','L002',2011),
('TR0010','C0010','DR19','V019','R010','T010','L007',2181),
('TR0011','C0001','DR15','V015','R011','T011','L006',2077),
('TR0012','C0014','DR14','V014','R012','T012','L010',2185),
('TR0013','C0018','DR03','V003','R013','T013','L020',2196),
('TR0014','C0020','DR02','V002','R014','T014','L019',2195),
('TR0015','C0019','DR01','V001','R015','T015','L005',2180),
('TR0016','C0002','DR08','V008','R016','T016','L015',2079),
('TR0017','C0016','DR19','V019','R017','T017','L009',2184),
('TR0018','C0011','DR09','V009','R018','T018','L004',2194),
('TR0019','C0015','DR16','V016','R019','T019','L003',2195),
('TR0020','C0007','DR13','V013','R020','T020','L012',2188);
```

insert into CUSTOMER values

```
('C0001','Albert Robert','0161652378','albertrobert@gmail.com'),
('C0002','Isaac Smith','0161447222','isaacsmith@yahoo.com'),
('C0003','John Cena','0146576546','johnccena@gmail.com'),
('C0004','Marie Curie','0171262743','mariecurie@hotmail.com'),
('C0005','Thomas Anderson','0152337766','thomasanderson@yahoo.com'),
('C0006','Niaz Parvez','0163827872','niazparvez@gmail.com'),
('C0007','Tanveer Ahmad','0173423414','tanveerahmad@hotmail.com'),
('C0008','Ling Hui','0124241525','linghui@yahoo.com'),
('C0009','Mahedi Munem','0132232421','mahedimunem@gmail.com'),
('C0010','Sasmitta Krishnan','0122421421','sasmittakrishnan@gmail.com'),
('C0011','Muhaimenul Abir','0167623126','muhaimenulabir@yahoo.com'),
```

('C0012','Bulbul Alam','0171242885','bulbulalam@hotmail.com'),  
('C0013','Edward George','0194573572','edwardgeorge@gmail.com'),  
('C0014','Ridwan Mahir','0171335671','ridwanmahir@yahoo.com'),  
('C0015','Joy Das','0186737883','joydas@yahoo.com'),  
('C0016','Bella Bose','0174545366','bellabose@gmail.com'),  
('C0017','Johnny Depp','0187576647','johnnydepp@hotmail.com'),  
('C0018','Adam Levine','0164887589','adamlevine@gmail.com'),  
('C0019','Han Solo','0195425765','hansolo@yahoo.com'),  
('C0020','Ellen Ripley','0157576888','ellenripley@hotmail.com'),  
('C0021','Heath Ledger','0187587698','heathledger@yahoo.com'),  
('C0022','Forrest Gump','0198875765','forrestgump@gmail.com');

insert into DRIVER values

('DR01','Roosevelt Merry','0161236379',675483),  
('DR02','Tamzid Hossain','0161467246',109435),  
('DR03','Michael Stevens','0182624235',654732),  
('DR04','James Bond','0142425267',203946),  
('DR05','Munira Hossain','0181686472',763043),  
('DR06','Ginny Anderson','0185765674',564781),  
('DR07','Mandy Stewart','0197765477',235967),  
('DR08','Philip Knight','0163543425',634570),  
('DR09','John Kok','0142564677',654389),  
('DR10','Ganesh Rao','0187587688',345678),  
('DR11','Yan Gun','0176453442',987654),  
('DR12','Kong Wang','0198656443',320959),  
('DR13','Peter Venkman','0186649787',326874),  
('DR14','Jasmine Sikder','0157799085',302234),  
('DR15','Jahid Hasan','0197856746',352698),  
('DR16','Jewel Hasan','0176342324',5463791),  
('DR17','Stewart Rick','0168753456',357845),  
('DR18','Tom Riddle','0136528744',493544),  
('DR19','Farooq Sheikh','0876635236',445763),  
('DR20','Andrew Hanson','0137545153',437984);

insert into VEHICLE values

('V001','SEDAN','Blue','6738IAL',10.00),  
('V002','MPV','Black','1644SGB',17.00),  
('V003','SEDAN','Grey','1367KIT',10.00),  
('V004','MPV','Red','2457OVF',17.00),  
('V005','SEDAN','Black','4635MEK',10.00),  
('V006','MPV','White','6488HDL',17.00),  
('V007','SUV','Green','4639MDK',20.00),  
('V008','SUV','Blue','4366OMD',20.00),  
('V009','SEDAN','Black','1746KIF',10.00),  
('V010','MPV','White','2487OVF',17.00),  
('V011','MPV','Red','5799MHK',17.00),  
('V012','SEDAN','Blue','4396OMD',10.00),  
('V013','SEDAN','Green','4932MHH',10.00),  
('V014','SUV','White','4629ZDK',20.00),  
('V015','SEDAN','Blue','6148HDL',10.00),  
('V016','MPV','Green','1607KIT',17.00),  
('V017','SUV','Brown','4972MHH',20.00),  
('V018','SUV','Silver','4355MEK',20.00),  
('V019','MPV','Black','9448DON',17.00),  
('V020','SEDAN','White','1664SGB',10.00);

insert into PROMOTION values

(20177,2.00,'2017-07-01','2017-07-31'),  
(20179,2.30,'2017-09-01','2017-09-30'),  
(20183,1.50,'2018-03-01','2018-03-31'),  
(20184,5.00,'2018-04-01','2018-04-30'),  
(20185,3.10,'2018-05-01','2018-05-31'),  
(20188,1.00,'2018-08-01','2018-08-31'),  
(21180,4.50,'2018-10-01','2018-10-31'),  
(21181,2.60,'2018-11-01','2018-11-30'),  
(20191,10.00,'2019-01-01','2019-01-31'),  
(20192,4.70,'2019-02-01','2019-02-28'),  
(20193,3.00,'2019-03-01','2019-03-31'),  
(20194,2.22,'2019-04-01','2019-04-30'),  
(20195,1.60,'2019-05-01','2019-05-31'),  
(20196,2.01,'2019-06-01','2019-06-30'),

(20197,1.40,'2019-07-01','2019-07-31'),  
(20198,3.33,'2019-08-01','2019-08-31'),  
(20199,5.50,'2019-09-01','2019-09-30'),  
(21190,1.78,'2019-10-01','2019-10-31'),  
(21191,4.45,'2019-11-01','2019-11-30'),  
(21192,3.45,'2019-12-01','2019-12-31');

insert into LOCATION values

('L001', 'Serin residency', 'Cyberjaya'),  
( 'L002', 'Parklane OUG', 'Bukit Jalil'),  
( 'L003', 'KLCC', 'Kuala Lumpur'),  
( 'L004', 'MMU', 'Cyberjaya'),  
( 'L005', 'KLIA', 'Sepang'),  
( 'L006', 'Masjid Sultan Salahuddin', 'Shah Alam'),  
( 'L007', 'Pavilion Mall', 'Kuala Lumpur'),  
( 'L008', 'Gurney Plaza', 'Penang'),  
( 'L009', 'Sunway Pyramid', 'Petaling Jaya'),  
( 'L010', 'CIMB Bank', 'Putrajaya'),  
( 'L011', 'IOI City Mall', 'Putrajaya'),  
( 'L012', 'Tune hotel', 'Cyberjaya'),  
( 'L013', 'The Mines', 'Serdang'),  
( 'L014', 'LimKokWing', 'Cyberjaya'),  
( 'L015', 'Merchantrade', 'Dengkil'),  
( 'L016', 'Maybank', 'Cyberjaya'),  
( 'L017', 'MidValley', 'Kuala Lumpur'),  
( 'L018', 'Dpulze', 'Cyberjaya'),  
( 'L019', 'Ministry of Education', 'Putrajaya'),  
( 'L020', 'KPJ Hospital', 'Ampang'),  
( 'L021', 'US Embassy', 'Kuala Lumpur');

insert into TIME values

('T001',2019,3,07,24,'14:45','2019-07-24 14:45:42','2019-07-24 15:45:00'),  
( 'T002',2019,2,05,21,'04:00','2019-05-21 04:00:12','2019-05-21 6:04:10'),  
( 'T003',2019,2,04,06,'17:15','2019-04-06 17:15:56','2019-04-06 17:24:08'),  
( 'T004',2018,1,03,16,'00:50','2018-03-16 00:05:24','2018-03-16 01:45:12'),



('T005',2017,3,09,21,'07:20','2017-09-21 07:20:34','2017-09-21 14:20:00'),  
 ('T006',2019,1,01,01,'09:13','2019-01-01 09:13:23','2019-01-01 09:33:09'),  
 ('T007',2019,1,02,04,'13:07','2019-02-09 13:07:27','2019-02-09 14:55:24'),  
 ('T008',2018,4,12,09,'18:11','2018-12-09 18:11:23','2018-12-09 18:19:43'),  
 ('T009',2019,1,01,07,'06:35','2019-01-07 06:35:43','2019-01-07 08:35:00'),  
 ('T010',2018,4,11,29,'04:02','2018-11-29 04:02:12','2018-11-29 08:01:22'),  
 ('T011',2017,3,07,15,'12:00','2017-07-15 12:00:42','2017-07-15 13:08:43'),  
 ('T012',2018,2,05,14,'21:30','2018-05-14 21:30:11','2018-05-14 23:30:10'),  
 ('T013',2019,2,06,09,'23:55','2019-06-09 23:55:44','2019-06-10 00:54:34'),  
 ('T014',2019,2,05,04,'14:28','2019-05-04 14:28:01','2019-05-04 15:18:13'),  
 ('T015',2018,4,10,22,'11:04','2018-10-22 11:04:04','2018-10-22 11:12:00'),  
 ('T016',2017,3,09,25,'08:05','2017-09-25 08:05:07','2017-09-25 08:10:01'),  
 ('T017',2018,2,04,05,'09:17','2018-04-05 09:17:10','2018-04-05 09:55:59'),  
 ('T018',2019,2,04,27,'10:00','2019-04-27 10:00:21','2019-04-27 14:08:28'),  
 ('T019',2019,2,05,30,'10:45','2019-05-30 10:45:46','2019-05-30 11:45:05'),  
 ('T020',2018,3,08,17,'09:45','2018-08-17 09:45:32','2018-08-17 10:28:19');

insert into REWARD values

('R001',10000,2000),  
 ('R002',1300,100),  
 ('R003',2490,10),  
 ('R004',15000,0),  
 ('R005',0,5000),  
 ('R006',7000,150),  
 ('R007',5555,345),  
 ('R008',1200,500),  
 ('R009',3000,1000),  
 ('R010',1350,50),  
 ('R011',9800,3000),  
 ('R012',1344,20),  
 ('R013',1000,90),  
 ('R014',1600,500),  
 ('R015',2300,900),  
 ('R016',1230,780),  
 ('R017',1340,122),  
 ('R018',5000,6000),  
 ('R019',0,0),  
 ('R020',605,5);