FIT3003 Major Assignment - Sem 2 2023

Tutor: Yiwei Zhong
Tutorial: 06_OnCampus

Campus: Clayton

Group Members:
Rian Barrett, Student ID: 31165176
Jasper Wan, Student ID: 33157502

□ A signed coversheet

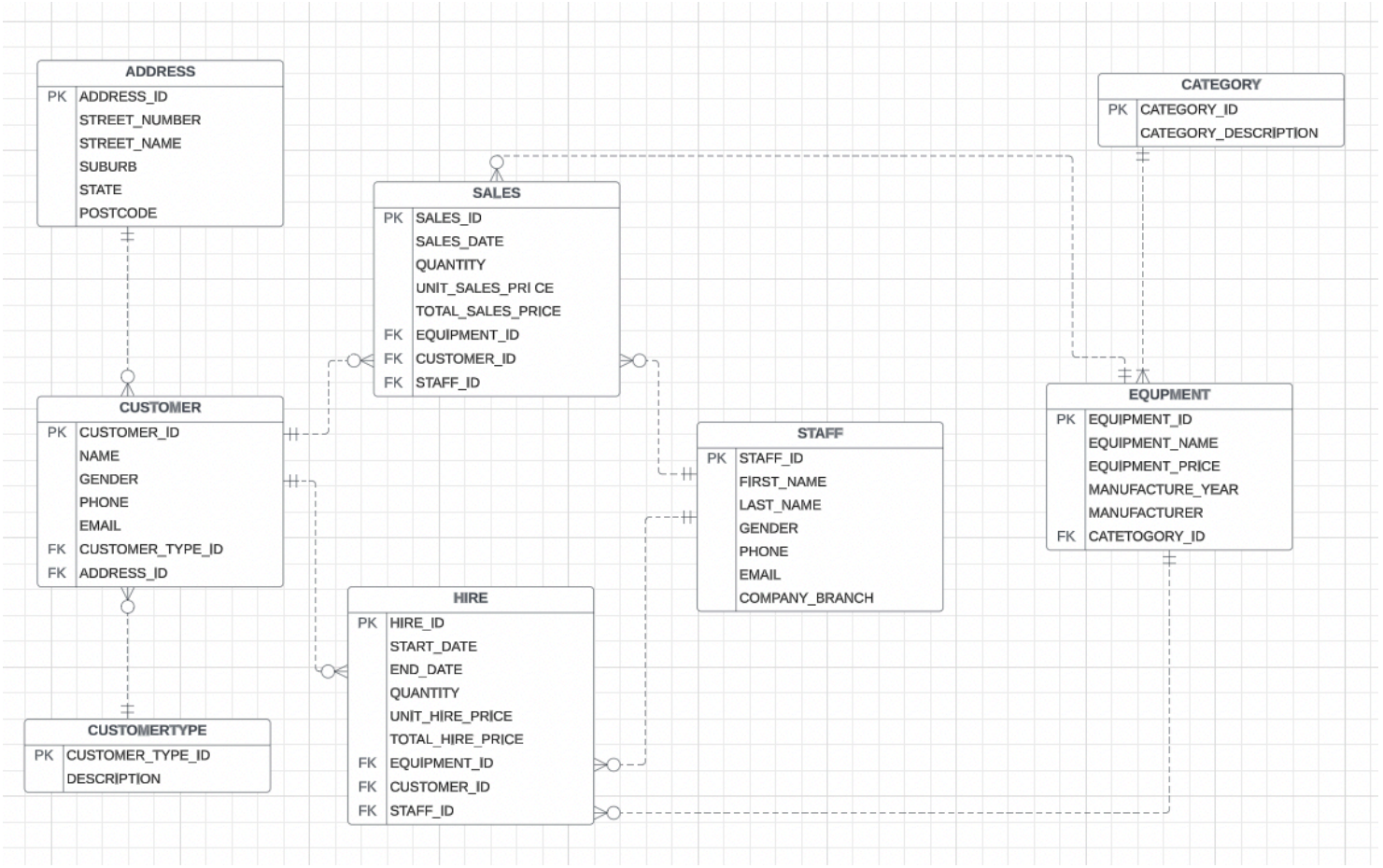
□ A contribution declaration form:
Each student must state the parts of the assignment that they completed. An example is as follows:

Percentage of contribution:
1. Name: Adam, ID: 210008, Contribution: 60%
2. Name: Ben, ID: 230933, Contribution: 40%
List of parts that each student completed:
1. Adam: list the parts that Adam did
2. Ben: list the parts that Ben did

Table of Contents

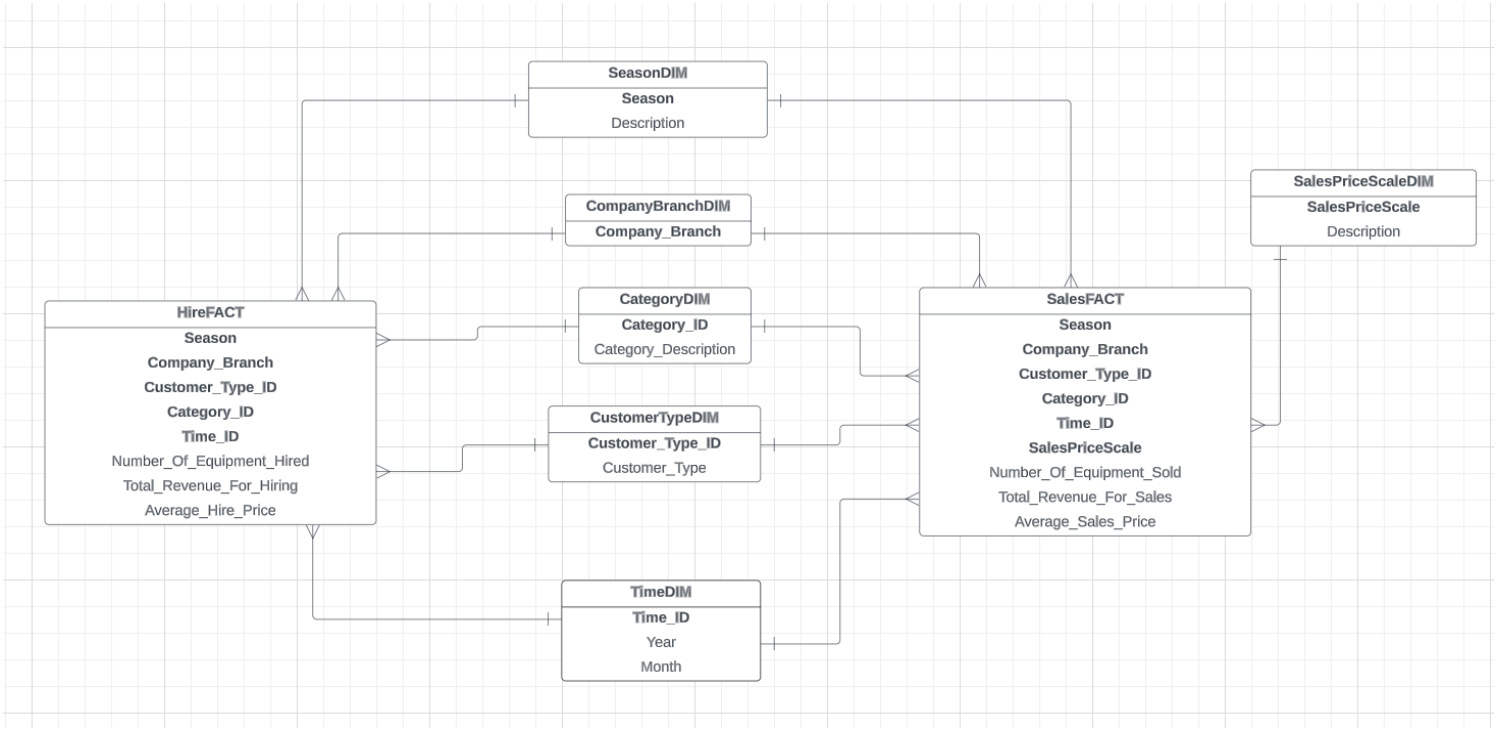
Task C.1 (outputs a, b, c, d)

a) The E/R diagram of the operational database.

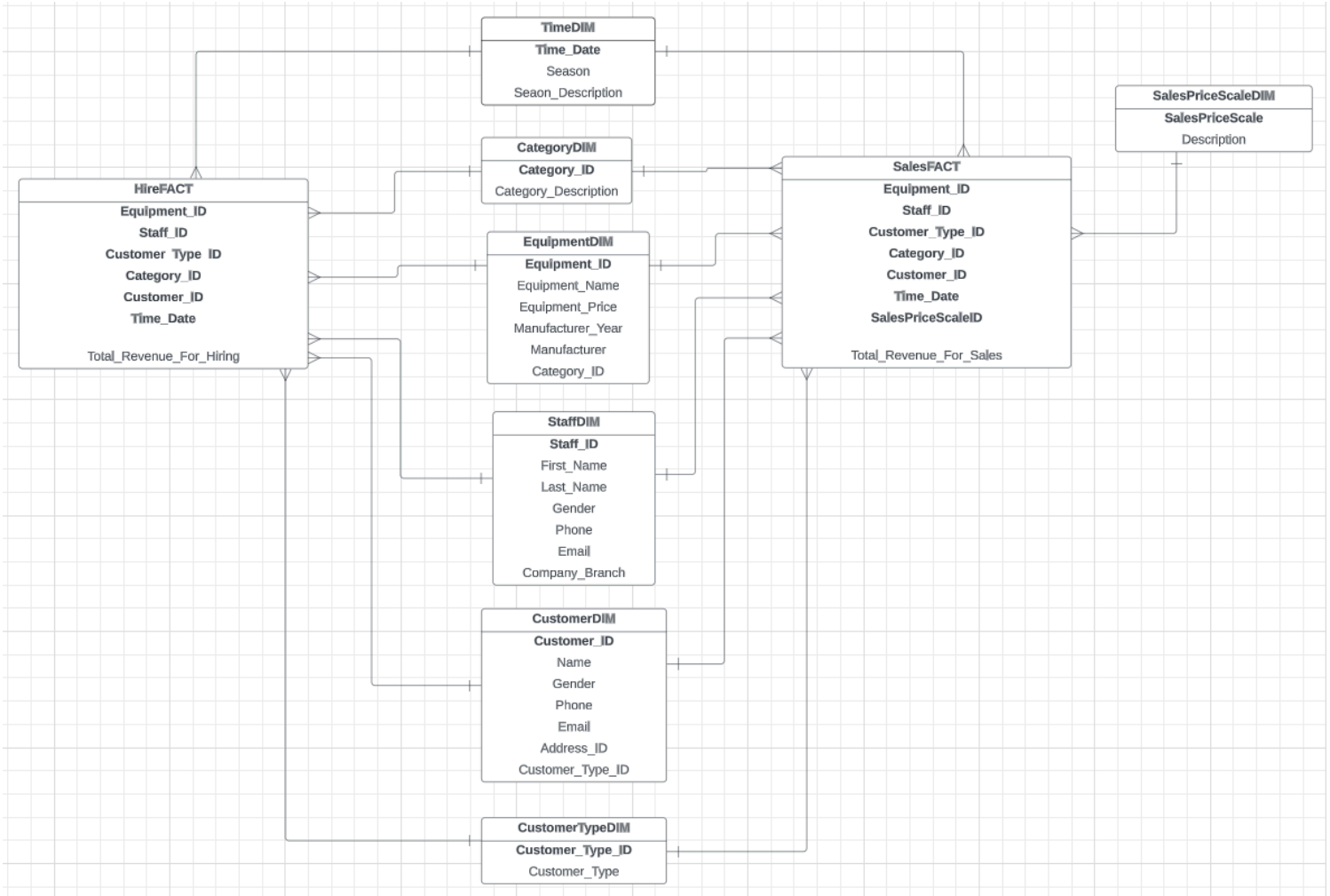


b) Two versions of star/snowflake schema diagrams,

Star schema Version 1



Star schema Version 2



c) An explanation of the difference among SCD types 1, 2, 3, 4, and 6. Explain the reasons for the choice of SCD type(s) for any temporal dimensions in your star schema, if there are any,

An explanation of the difference among SCD types 1, 2, 3, 4, and 6.

The difference between each SCD is whether they track the history of changes and how they track the changes. For SCD type 1, the history of changes is not recorded rather only the latest value is recorded by overwriting the old value. For SCD type 2, the history of changes is recorded and tracked in the main dimension through new records that are continually added to the dimension. For SCD type 3, the entire history of the changes is not maintained but rather it only keeps the last two changes without the need for separate identifiers. For SCD type 4, the history of changes is recorded and tracked in a new dimension specifically used to maintain the history of attribute value changes. For SCD type 6, the history of changes is recorded in the main original dimension table without the need for separate identifiers

Explain the reasons for the choice of SCD type(s) for any temporal dimensions in your star schema, if there are any,

A temporal dimension was not used in our star schema as the from our analysis of the assignment brief and the requested queries to identify the fact measures and dimensions strongly suggested creating new dimensions that tracks the changes in price for equipment in terms of both hiring and sales was not requested for analysis. If the creation of new dimensions that tracks the changes in price for equipment in terms of both hiring and sales was necessary, these dimension would track the change in the price by implementing SCD type 4 because it can maintain the entire history with minimal confusion by having a EquipmentDIM and two dimensions each for hire and sales that track the history of the changes in price for the equipment.

d) An explanation of the difference among the two versions of star/snowflake schema.

The difference between the two versions of the star schemas are the levels of granularity. In the V1 star schema there is a lower level of granularity and higher level of aggregation. While in the V2 star schema there is a higher level of granularity and lower level of aggregation.
In the V1 star schema there are three fact measures involving Number of Equipment, Average, and Total Revenue, all of which are aggregated. In the V2 star schema there is only one fact measure and the fact measures involving Number of Equipment, Average are removed through increasing the level of granularity.

The steps taken to lower the level of aggregation was to add a new dimension such as EquipmentDIM which broke down the fact measure into more details on each record of the new dimension. Secondly existing dimensions such as TimeDIM were replaced with a higher granularity dimension TimeDIM_V2. The difference between the new TimeDIM and the V1

TimeDIM was the granularity at the level of date rather than month. By incorporating the following techniques to lower the level of aggregation the V2 star schema has the highest level of granularity where no aggregation exists.

Task C.2 (outputs a, b)

a) SQL statements to explore the operational database, and SQL statements of the data cleaning,

```
-- Preliminary Data checks
--Number of records in each table
select count(*)  from MonEquip.ADDRESS; --150
select count(*)  from MonEquip.CATEGORY; --15
select count(*)  from MonEquip.CUSTOMER; --153
select count(*)  from MonEquip.CUSTOMER_TYPE;  --2
select count(*)  from MonEquip.EQUIPMENT; --158
select count(*) from MonEquip.HIRE; --304
select count(*)  from MonEquip.SALES; --151
select count(*)  from MonEquip.STAFF; --50

-- Checking if data is in the right table
select *
from MonEquip.sales
where equipment_id not in
   (select equipment_id
   from MonEquip.equipment);

select *
from MonEquip.sales
where customer_id not in
   (select customer_id
   from MonEquip.customer);

select *
from MonEquip.sales
where staff_id not in
   (select staff_id
   from MonEquip.staff);

--Check for simple duplicates

select START_DATE, END_DATE, EQUIPMENT_ID, QUANTITY, UNIT_HIRE_PRICE,
TOTAL_HIRE_PRICE, CUSTOMER_ID, STAFF_ID, count(*)
from MonEquip.hire
group by START_DATE, END_DATE, EQUIPMENT_ID, QUANTITY, UNIT_HIRE_PRICE,
TOTAL_HIRE_PRICE, CUSTOMER_ID, STAFF_ID
```

```
having count(*) > 1;

select SALES_DATE, EQUIPMENT_ID, QUANTITY, UNIT_SALES_PRICE,
TOTAL_SALES_PRICE, CUSTOMER_ID, STAFF_ID, count(*)
from MonEquip.sales
group by SALES_DATE, EQUIPMENT_ID, QUANTITY, UNIT_SALES_PRICE,
TOTAL_SALES_PRICE, CUSTOMER_ID, STAFF_ID
having count(*) > 1;

select sales_id, count(*)
from MonEquip.sales
group by sales_id
having count(*) > 1;

select hire_id, count(*)
from MonEquip.hire
group by hire_id
having count(*) > 1;

select staff_id, count(*)
from MonEquip.staff
group by staff_id
having count(*) > 1;

select EQUIPMENT_ID, count(*)
from MonEquip.equipment
group by EQUIPMENT_ID
having count(*) > 1;

-- DATA ERRORS

-- Duplication Problem

--DATA ERROR: DUPLICATION customer with 4 counts
select customer_id, count(*)
from MonEquip.customer
group by customer_id
having count(*) > 1;

-- Incorrect Values

-- DATA ERROR: INCORRECT VALUE negative TOTAL_HIRE_PRICE in HIRE
select *
from MonEquip.hire
where total_hire_price < 0;

-- DATA ERROR: INCORRECT VALUE negative QUANTITY in SALES
select *
```

```
from MonEquip.sales
where quantity < 0;
--unitsalesprice * quantity != total sales price
select *
from MonEquip.sales
where unit_sales_price * quantity != total_sales_price;

-- DATA ERROR: INCORRECT VALUE START_DATE is set as date after the END_DATE
select *
from MonEquip.hire
where start_date > end_date;

-- DATA ERROR: INCORRECT VALUE START_DATE and END_DATE is set as date in
2090 which has not happened yet.
select *
from MonEquip.hire
where start_date > to_date('202101','YYYYMM');

-- DATA ERROR: INCORRECT VALUE for TOTAL_HIRE_PRICE. TOTAL_HIRE_PRICE !=
(End Date - Start Date) * UnitHirePrice * Quantity and is calculated incorrectly
select *
from MonEquip.hire
where (End_Date - Start_Date) * Unit_Hire_Price * Quantity != total_hire_price;

-- Null Value Problems
--DATA ERROR: Null Category_Description in CATEGORY
select * from MonEquip.equipment e join MonEquip.category c on e.category_id =
c.category_id
where e.category_id = 15;

select *
from MonEquip.category
where category_description = 'null';


-- Date Cleaning
-- Duplication Problem
-- CLEAN DUPLICATION ERROR in CUSTOMER TABLE
create table Cleaned_CUSTOMER as
select distinct *
from  MonEquip.customer;

-- Check data is cleaned
select customer_id, count(*)
from Cleaned_CUSTOMER
group by customer_id
having count(*) > 1;
```

```
select * from  MonEquip.customer
where customer_id = 52;

-- Incorrect Values

-- CLEANED DATA ERROR: INCORRECT VALUE negative QUANTITY in SALES
select *
from MonEquip.sales
where quantity < 0;
--unitsalesprice * quantity != total sales price
select *
from MonEquip.sales
where unit_sales_price * quantity != total_sales_price;

create table Cleaned_SALES as
select *
from  MonEquip.SALES;

UPDATE Cleaned_SALES
SET QUANTITY = 4
WHERE SALES_ID = 151;

select *
from Cleaned_SALES
where  SALES_ID = 151;

-- CLEANED DATA ERROR: INCORRECT VALUE START_DATE is set as date after the
END_DATE
select *
from MonEquip.hire
where start_date > end_date;

create table Cleaned_HIRE as
select *
from  MonEquip.HIRE;

UPDATE Cleaned_HIRE
SET START_DATE = TO_DATE('2020/10/17', 'YYYY/MM/DD'), END_DATE =
TO_DATE('2020/12/05', 'YYYY/MM/DD')
WHERE HIRE_ID = 302;

select *
from Cleaned_HIRE
where  HIRE_ID = 302;

-- CLEANED DATA ERROR: INCORRECT VALUE START_DATE and END_DATE is set as
date in 2090 which has not happened yet.
select *
```

```
from MonEquip.hire
where start_date > to_date('202101','YYYYMM');

DELETE FROM Cleaned_HIRE WHERE  HIRE_ID = 303;

select *
from Cleaned_HIRE
where  HIRE_ID = 303;

-- CLEANED DATA ERROR: INCORRECT VALUE for TOTAL_HIRE_PRICE.
TOTAL_HIRE_PRICE != (End Date - Start Date) * UnitHirePrice * Quantity and is calculated
incorrectly
select *
from MonEquip.hire
where (End_Date - Start_Date) * Unit_Hire_Price * Quantity != total_hire_price;

SELECT * FROM CLEANED_HIRE;

-- Total hire price is calculated as (End Date - Start Date) * UnitHirePrice * Quantity
UPDATE Cleaned_HIRE
SET total_hire_price = (End_Date - Start_Date) * Unit_Hire_Price * Quantity
WHERE START_DATE != END_DATE;

--  If the customer returns the equipment within the same day, they only need to pay for 50%
of the unit hire price.
UPDATE Cleaned_HIRE
SET total_hire_price =  (QUANTITY*Unit_Hire_Price)/2
WHERE START_DATE = END_DATE;

SELECT * FROM monequip.hire;

SELECT * FROM CLEANED_HIRE;

-- CLEANED DATA ERROR: INCORRECT VALUE negative TOTAL_HIRE_PRICE in HIRE
select *
from MonEquip.hire
where total_hire_price < 0;

select *
from Cleaned_HIRE
where  HIRE_ID = 304;

-- Null Value Problems
--DATA ERROR: Null Category_Description in CATEGORY
select * from MonEquip.equipment e join MonEquip.category c on e.category_id =
c.category_id
where e.category_id = 15;
```

```
select *
from MonEquip.category
where category_description = 'null';

select *
from MonEquip.category;
```

b) Screenshot of data before and after data cleaning

Duplication Problems

DATA ERROR: Duplicate entries of customer with ID 52

BEFORE

```
53  --customer with 4 counts
54
55  select customer_id, count(*)
56  from MonEquip.customer
57  group by customer_id
58  having count(*) > 1;
59
60  select * from  MonEquip.customer
61  where customer_id = 52;
```

	CUSTOMER_ID	COUNT(*)
1	52	4

	CUSTOMER_ID	CUSTOMER_TYPE_ID	NAME	GENDER	ADDRESS_ID	PHONE	EMAIL
1	52	2	Abbie Maddie	Male	52	904 627 9038	amaddie1f@columbia.edu
2	52	2	Abbie Maddie	Male	52	904 627 9038	amaddie1f@columbia.edu
3	52	2	Abbie Maddie	Male	52	904 627 9038	amaddie1f@columbia.edu
4	52	2	Abbie Maddie	Male	52	904 627 9038	amaddie1f@columbia.edu

AFTER

Duplicate customer data has been removed

	CUSTOM...	COUNT(*)

Incorrect Values

DATA ERROR: INCORRECT VALUE negative QUANTITY in SALES

BEFORE

	SALES_ID	SALES_DATE	EQUIPMENT_ID	QUANTITY	UNIT_SALES_PRICE	TOTAL_SALES_PRICE
1	151	15/DEC/20	20	-3	45500	182000

AFTER: Quantity is corrected to 4

	SALES_ID	SALES_DATE	EQUIPMENT_ID	QUANTITY	UNIT_SALES_PRICE	TOTAL_SALES_PRICE
1	151	15/DEC/20	20	4	45500	182000

DATA ERROR: INCORRECT VALUE START_DATE is set as date after the END_DATE

BEFORE

	HIRE_ID	START_DATE	END_DATE	EQUIPMENT_ID	QUANTITY	UNIT_HIRE_PRICE	TOTAL_HIRE_PRICE	CUSTOMER_ID	STAFF_ID
1	302	05/DEC/20	17/OCT/20	21	2	100	200	111	123

AFTER: Corrected by swapping the dates around

	HIRE_ID	START_DATE	END_DATE	EQUIPMENT_ID	QUANTITY	UNIT_HIRE_PRICE	TOTAL_HIRE_PRICE	CUSTOMER_ID
1	302	17/OCT/20	05/DEC/20	21	2	100	200	111

DATA ERROR: Incorrect value negative total hire price

BEFORE

Query Result | Query Result 2 | Query Result 4 | Query Result 5 | Query Result 6 | Query Result 7 | ...

SQL | All Rows Fetched: 2 in 0.008 seconds

	HIRE_ID	START_DATE	END_DATE	EQUIPMENT_ID	QUANTITY	UNIT_HIRE_PRICE	TOTAL_HIRE_PRICE	CUSTOMER_ID	STA
1	303	25/JAN/90	27/DEC/99	43	3	50	-150	53	
2	304	08/DEC/20	08/DEC/20	114	1	350	-1	34	

AFTER Price updated

	HIRE_ID	START_DATE	END_DATE	EQUIPMENT_ID	QUANTITY	UNIT_HIRE_PRICE	TOTAL_HIRE_PRICE	CUSTOMER_ID	STAFF_ID
1	304	08/DEC/20	08/DEC/20	114	1	350	175	34	85

DATA ERROR: INCORRECT VALUE START_DATE and END_DATE is set as date in 2090 which has not happened yet.

BEFORE

	HIRE_ID	START_DATE	END_DATE	EQUIPMENT_ID	QUANTITY	UNIT_HIRE_PRICE	TOTAL_HIRE_PRICE	CUSTOMER_ID	STAFF_ID
1	303	25/JAN/90	27/DEC/99	43	3	50	-150	53	223

AFTER Data entry deleted due to too many errors

HIRE_ID	START_D...	END_DATE	EQUIPME...	QUANTITY	UNIT_HIR...	TOTAL_H...	CUSTOM

DATA ERROR: INCORRECT VALUE for TOTAL_HIRE_PRICE. TOTAL_HIRE_PRICE != (End Date - Start Date) * UnitHirePrice * Quantity and is calculated incorrectly

BEFORE

	HIRE_ID	START_...	END_DATE	EQUIPMENT_ID	QUANTITY	UNIT_HIRE_PRICE	TOTAL_HIRE_PRICE	CUSTOMER_ID	STAFF_ID
1	1	11/MAY/18	14/MAY/18	135	3	80	240	77	2
2	2	17/MAY/18	20/MAY/18	49	2	660	1320	70	38
3	3	18/MAY/18	19/MAY/18	117	1	150	150	58	17
4	4	21/MAY/18	25/MAY/18	36	1	540	540	124	41
5	5	21/MAY/18	23/MAY/18	37	2	600	1200	87	35
6	6	22/MAY/18	26/MAY/18	53	2	500	1000	15	31
7	7	24/MAY/18	28/MAY/18	73	3	140	420	77	10
8	8	25/MAY/18	28/MAY/18	135	3	80	240	140	40
9	9	28/MAY/18	28/MAY/18	127	2	170	170	8	46
10	10	29/MAY/18	29/MAY/18	86	2	200	200	5	28
11	11	12/JUN/18	14/JUN/18	71	2	450	900	145	21
12	12	15/JUN/18	15/JUN/18	23	2	360	360	74	43
13	13	24/JUN/18	25/JUN/18	61	3	150	450	114	10
14	14	25/JUN/18	25/JUN/18	85	3	210	315	110	44
15	15	30/JUN/18	01/JUL/18	9	1	360	360	47	8
16	16	05/JUL/18	05/JUL/18	74	1	280	140	129	7

AFTER updated with correct formula

	HIR...	START_DATE	END_DATE	EQUIPMENT_ID	QUANTITY	UNIT_HIRE_PRICE	TOTAL_HIRE_PRICE	CUSTOMER_ID	STAFF_ID
1	1	11/MAY/18	14/MAY/18	135	3	80	720	77	2
2	2	17/MAY/18	20/MAY/18	49	2	660	3960	70	38
3	3	18/MAY/18	19/MAY/18	117	1	150	150	58	17
4	4	21/MAY/18	25/MAY/18	36	1	540	2160	124	41
5	5	21/MAY/18	23/MAY/18	37	2	600	2400	87	35
6	6	22/MAY/18	26/MAY/18	53	2	500	4000	15	31
7	7	24/MAY/18	28/MAY/18	73	3	140	1680	77	10
8	8	25/MAY/18	28/MAY/18	135	3	80	720	140	40
9	9	28/MAY/18	28/MAY/18	127	2	170	170	8	46
10	10	29/MAY/18	29/MAY/18	86	2	200	200	5	28
11	11	12/JUN/18	14/JUN/18	71	2	450	1800	145	21
12	12	15/JUN/18	15/JUN/18	23	2	360	360	74	43
13	13	24/JUN/18	25/JUN/18	61	3	150	450	114	10
14	14	25/JUN/18	25/JUN/18	85	3	210	315	110	44
15	15	30/JUN/18	01/JUL/18	9	1	360	360	47	8
16	16	05/JUL/18	05/JUL/18	74	1	280	140	129	7

Null Value Problems

DATA ERROR: Null Category_Description in CATEGORY

BEFORE

	CATEGORY_ID	CATEGORY_DESCRIPTION
1	15	null

AFTER

Not fixed as it does not majorly affect the data warehousing analysis and requires client info.

If you have done the data cleaning process, explain the strategies you used in this process.

The strategy taken for the data cleaning process involved, first observing the number of records in the operational database to see whether there were any inconsistencies amongst the number of records. Then specific data errors were identified through probing for potential issues that can occur such as incorrect negative values, null values, etc.

Task C.3 (outputs a, b, c)

a) SQL statements (e.g. create table, insert into, etc) to create the star/snowflake schema Version-1

```
-- Create CustomerTypeDIM by Direct Copy
DROP TABLE CustomerTypeDIM CASCADE CONSTRAINTS PURGE;
create table CustomerTypeDIM as
select * from MonEquip.CUSTOMER_TYPE;

select * from CustomerTypeDIM;

-- Create CategoryDIM by Direct Copy
DROP TABLE CategoryDIM CASCADE CONSTRAINTS PURGE;
create table CategoryDIM as
select * from MonEquip.CATEGORY;

-- Create TimeDIM using Temp
DROP TABLE TimeDIM CASCADE CONSTRAINTS PURGE;

-- But you have to get it from both tables
-- The operational database records the transaction from April 2018 to December 2020

select * from MonEquip.SALES;
select * from MonEquip.HIRE;

DROP TABLE TimeDimSalesTemp CASCADE CONSTRAINTS PURGE;

create table TimeDimSalesTemp as
SELECT DISTINCT to_char(SALES_DATE, 'YYYYMM') AS Time_ID,
to_char(SALES_DATE, 'MM') as Time_Month,
to_char(SALES_DATE, 'YYYY') as Time_Year
```

```
from MonEquip.SALES;

select * from TimeDimSalesTemp;

create table TimeDimHireTemp as
SELECT DISTINCT to_char(START_DATE, 'YYYYMM') AS Time_ID,
to_char(START_DATE, 'MM') as Time_Month,
to_char(START_DATE, 'YYYY') as Time_Year
from MonEquip.HIRE;

select * from TimeDimHireTemp;


create table TimeDim as
SELECT DISTINCT Time_ID, Time_Month, Time_Year
from (
SELECT Time_ID, Time_Month, Time_Year from TimeDimSalesTemp
   union all
SELECT Time_ID, Time_Month, Time_Year from TimeDimHireTemp
);

select * from TimeDim;

-- Create SeasonDIM
-- [Australian Season: Summer, Winter, Autumn, Spring

DROP TABLE SeasonDIM CASCADE CONSTRAINTS PURGE;
--Summer (December, January, February)
--Autumn (March, April, May)
--Winter (June, July, August)
--Spring (September, October, November)
create table SeasonDIM
(Season VARCHAR2(6),
Description varchar2(20));

insert into SeasonDIM values ('Summer', 'Dec-Feb');
insert into SeasonDIM values ('Autumn', 'Mar-May');
insert into SeasonDIM values ('Winter', 'Jun-Aug');
insert into SeasonDIM values ('Spring', 'Sep-Nov');

-- Create CompanyBranchDIM

DROP TABLE Company_BranchDIM CASCADE CONSTRAINTS PURGE;

select distinct Company_Branch from MonEquip.Staff;

create table Company_BranchDIM as
select distinct Company_Branch from MonEquip.Staff;
```

```
select * from Company_BranchDIM;

-- Create SalesPriceScaleDIM
-- Sales price scale: low sales <$5,000; medium sales between $5,000 and $10,000; high
sales > $10,000
DROP TABLE SalesPriceScaleDIM CASCADE CONSTRAINTS PURGE;

create table SalesPriceScaleDIM
(SalesPriceScale VARCHAR2(6),
Description varchar2(30));

insert into SalesPriceScaleDIM values ('Low', '< $5,000');
insert into SalesPriceScaleDIM values ('Medium', 'between $5,000 and $10,000');
insert into SalesPriceScaleDIM values ('High', '> $10,000');

select * from SalesPriceScaleDIM;

-- Create HireFACT_V1 and SalesFACT_V1 using TempFacts
-- Create HireFACT_V1
-- HireFact attributes
--Season, Company_Branch, Customer_Type_ID, Category_ID, Time_ID

select * from MonEquip.EQUIPMENT; -- all them are distinct
select * from MonEquip.HIRE;
select * from MonEquip.SALES;


DROP TABLE HireTempFact_V1 CASCADE CONSTRAINTS PURGE;

create table HireTempFact_V1 as
select
to_char(H.START_DATE, 'YYYYMM') AS Time_ID,
to_char(H.START_DATE, 'MM') as Month, -- Need for Season
S.COMPANY_BRANCH,
C.CUSTOMER_TYPE_ID,
E.CATEGORY_ID,
H.START_DATE,
H.END_DATE,
H.QUANTITY, -- for NUMBER_OF_EQUIPMENT
H.UNIT_HIRE_PRICE,
H.TOTAL_HIRE_PRICE, -- for Total_Revenue
H.EQUIPMENT_ID
from MonEquip.HIRE H, MonEquip.CUSTOMER C, MonEquip.EQUIPMENT E,
MonEquip.STAFF S
where H.EQUIPMENT_ID = E.EQUIPMENT_ID AND
H.STAFF_ID = S.STAFF_ID AND
H.CUSTOMER_ID = C.CUSTOMER_ID;
```

```
SELECT * FROM HireTempFact_V1;

alter table HireTempFact_V1 add
(Season VARCHAR2(6));
update HireTempFact_V1
set Season = 'Summer'
where Month >= '12'
OR Month <= '02';

update HireTempFact_V1
set Season = 'Autumn'
where Month >= '03'
and Month <= '05';

update HireTempFact_V1
set Season = 'Winter'
where Month >= '06'
and Month <= '08';

update HireTempFact_V1
set Season = 'Spring'
where Month >= '09'
and Month <= '11';


SELECT * FROM HireTempFact_V1;

DROP TABLE HireFact_V1 CASCADE CONSTRAINTS PURGE;

create table HireFact_V1 as
select
Season,
Company_Branch,
Customer_Type_ID,
Category_ID,
Time_ID,
sum(QUANTITY) AS NUMBER_OF_EQUIPMENT_HIRED,
sum(TOTAL_HIRE_PRICE) AS TOTAL_REVENUE_FOR_HIRING,
AVG(UNIT_HIRE_PRICE) AS AVERAGE_HIRE_PRICE
from HireTempFact_V1
group by
Season,
Company_Branch,
Customer_Type_ID,
Category_ID,
Time_ID,
EQUIPMENT_ID;
```

```
SELECT * FROM HireFact_V1;

-- Create SalesFACT_V1
-- SalesFact attributes
--Season, Company_Branch, Customer_Type_ID, Category_ID, Time_ID, SalesPriceScale
-- Sales price scale: low sales <$5,000; medium sales between $5,000 and $10,000; high
sales > $10,000

select * from MonEquip.SALES;

DROP TABLE SalesTempFact_V1 CASCADE CONSTRAINTS PURGE;

create table SalesTempFact_V1 as
select
to_char(SA.SALES_DATE, 'YYYYMM') AS Time_ID,
to_char(SA.SALES_DATE, 'MM') as Month, -- Need for Season
S.COMPANY_BRANCH,
C.CUSTOMER_TYPE_ID,
E.CATEGORY_ID,
SA.SALES_DATE,
SA.QUANTITY, -- for NUMBER_OF_EQUIPMENT
SA.UNIT_SALES_PRICE,
SA.TOTAL_SALES_PRICE, -- for Total_Revenue
SA.EQUIPMENT_ID
from MonEquip.SALES SA, MonEquip.CUSTOMER C, MonEquip.EQUIPMENT E,
MonEquip.STAFF S
where SA.EQUIPMENT_ID = E.EQUIPMENT_ID AND
SA.STAFF_ID = S.STAFF_ID AND
SA.CUSTOMER_ID = C.CUSTOMER_ID;

SELECT * FROM SalesTempFact_V1;

alter table SalesTempFact_V1 add
(Season VARCHAR2(6));
update SalesTempFact_V1
set Season = 'Summer'
where Month >= '12'
OR Month <= '02';

update SalesTempFact_V1
set Season = 'Autumn'
where Month >= '03'
and Month <= '05';

update SalesTempFact_V1
set Season = 'Winter'
where Month >= '06'
```

```
and Month <= '08';

update SalesTempFact_V1
set Season = 'Spring'
where Month >= '09'
and Month <= '11';


SELECT * FROM SalesTempFact_V1;

-- Sales price scale: low sales <$5,000; medium sales between $5,000 and $10,000; high
sales > $10,000
alter table SalesTempFact_V1 add
(SalesPriceScale VARCHAR2(6));

update SalesTempFact_V1
set SalesPriceScale = 'Low'
where UNIT_SALES_PRICE < 5000;

update SalesTempFact_V1
set SalesPriceScale = 'Medium'
where UNIT_SALES_PRICE >= 5000
and UNIT_SALES_PRICE <= 10000;

update SalesTempFact_V1
set SalesPriceScale = 'High'
where UNIT_SALES_PRICE > 10000;

SELECT * FROM SalesTempFact_V1;

DROP TABLE SalesFact_V1 CASCADE CONSTRAINTS PURGE;

--Season, Company_Branch, Customer_Type_ID, Category_ID, Time_ID, SalesPriceScale

create table SalesFact_V1 as
select
Season,
Company_Branch,
Customer_Type_ID,
Category_ID,
Time_ID,
SalesPriceScale,
sum(QUANTITY) AS NUMBER_OF_EQUIPMENT_SOLD,
sum(TOTAL_SALES_PRICE) AS TOTAL_REVENUE_FOR_SALES,
AVG(UNIT_SALES_PRICE) AS AVERAGE_SALES_PRICE
from SalesTempFact_V1
group by
Season,
```

```
Company_Branch,
Customer_Type_ID,
Category_ID,
Time_ID,
EQUIPMENT_ID,
SalesPriceScale;
```

b) SQL statements (e.g. create table, insert into, etc) to create the star/snowflake schema Version-2

```
-- Create CustomerTypeDIM by Direct Copy
DROP TABLE CustomerTypeDIM CASCADE CONSTRAINTS PURGE;
create table CustomerTypeDIM as
select * from MonEquip.CUSTOMER_TYPE;

create table CustomerDIM as
select * from MonEquip.CUSTOMER;

create table StaffDIM as
select * from MonEquip.STAFF;

create table EquipmentDIM as
select * from MonEquip.STAFF;

create table CategoryDIM as
select * from MonEquip.CATEGORY;

create table SalesPriceScaleDIM
(SalesPriceScale VARCHAR2(6),
Description varchar2(30));

insert into SalesPriceScaleDIM values ('Low', '< $5,000');
insert into SalesPriceScaleDIM values ('Medium', 'between $5,000 and $10,000');
insert into SalesPriceScaleDIM values ('High', '> $10,000');

select * from SalesPriceScaleDIM;

DROP TABLE TimeDim_V2_Temp CASCADE CONSTRAINTS PURGE;

create table TimeDim_V2_Temp as
SELECT DISTINCT Time_ID, Time_Date, Month
from (
SELECT to_char(SALES_DATE, 'YYYYMMDD') AS Time_ID, SALES_DATE as Time_Date,
to_char(SALES_DATE, 'MM') as Month from MonEquip.SALES
   union all
```

```
SELECT to_char(START_DATE, 'YYYYMMDD') AS Time_ID, START_DATE as Time_Date,
to_char(START_DATE, 'MM') as Month from MonEquip.HIRE
);

select * from TimeDim_V2_Temp;

alter table TimeDim_V2_Temp add
(Season VARCHAR2(6),
Season_Description varchar2(20));

update TimeDim_V2_Temp
set Season = 'Summer', Season_Description = 'Dec-Feb'
where Month >= '12'
OR Month <= '02';

update TimeDim_V2_Temp
set Season = 'Autumn' , Season_Description = 'Mar-May'
where Month >= '03'
and Month <= '05';

update TimeDim_V2_Temp
set Season = 'Winter', Season_Description = 'Jun-Aug'
where Month >= '06'
and Month <= '08';

update TimeDim_V2_Temp
set Season = 'Spring', Season_Description = 'Sep-Nov'
where Month >= '09'
and Month <= '11';

DROP TABLE TimeDim_V2 CASCADE CONSTRAINTS PURGE;

create table TimeDim_V2 as
SELECT Time_ID, Time_Date, Season, Season_Description
From TimeDim_V2_Temp;

select * from TimeDim_V2;

--HireFact_V2

DROP TABLE HireFact_V2 CASCADE CONSTRAINTS PURGE;

create table HireFact_V2 as
select
to_char(H.START_DATE, 'YYYYMMDD') AS Time_ID,
H.STAFF_ID,
C.CUSTOMER_TYPE_ID,
E.CATEGORY_ID,
```

```
H.EQUIPMENT_ID,
H.CUSTOMER_ID,
H.TOTAL_HIRE_PRICE -- for Total_Revenue
from MonEquip.HIRE H, MonEquip.CUSTOMER C, MonEquip.EQUIPMENT E,
MonEquip.STAFF S
where H.EQUIPMENT_ID = E.EQUIPMENT_ID AND
H.STAFF_ID = S.STAFF_ID AND
H.CUSTOMER_ID = C.CUSTOMER_ID;


SELECT * FROM HireFact_V2;

select * from monequip.hire;


--SalesFact_V2

create table SalesFact_V2 as
select
to_char(SA.SALES_DATE, 'YYYYMMDD') AS Time_ID,
SA.STAFF_ID,
C.CUSTOMER_TYPE_ID,
E.CATEGORY_ID,
SA.EQUIPMENT_ID,
SA.CUSTOMER_ID,
SA.TOTAL_SALES_PRICE -- for Total_Revenue
from MonEquip.SALES SA, MonEquip.CUSTOMER C, MonEquip.EQUIPMENT E,
MonEquip.STAFF S
where SA.EQUIPMENT_ID = E.EQUIPMENT_ID AND
SA.STAFF_ID = S.STAFF_ID AND
SA.CUSTOMER_ID = C.CUSTOMER_ID;
```

c) Screenshots of the implementation and the tables that you have created; this includes the contents of each table that you have created. If the table is very big, you can only show the first part of the data. Note: The SQL statements for both levels of star schema must be presented in the PDF file

star/snowflake schema Version-1

SeasonDIM

	SEASON	DESCRIPTION
1	Summer	Dec-Feb
2	Autumn	Mar-May
3	Winter	Jun-Aug
4	Spring	Sep-Nov

Company_BranchDIM

	COMPANY_BRANCH
1	Pakenham
2	Richmond
3	Caulfield
4	Clayton
5	Docklands
6	Parkville
7	Toorak
8	Eltham
9	Dandenong
10	Chadstone
11	Geelong
12	Hughesdale
13	Prahran
14	Cheltenham
15	Fitzroy

CustomerTypeDIM

	CUSTOMER_TYPE_ID	DESCRIPTION
1	1	Individual
2	2	Business

CategoryDIM

	CATEGORY_ID	CATEGORY_DESCRIPTION
1	1	Access
2	2	Air Compressor
3	3	Compaction
4	4	Concrete
5	5	Earthmoving
6	6	Generators
7	7	Landscaping
8	8	Lighting
9	9	Plumbing
10	10	Rail
11	11	Safety
12	12	Site Equipment
13	13	Trailers
14	14	Vehicles
15	15	null

TimeDIM_V1

	TIME_ID	TIME_MONTH	TIME_YEAR
1	201810	10	2018
2	201906	06	2019
3	202009	09	2020
4	202010	10	2020
5	201806	06	2018
6	201901	01	2019
7	202004	04	2020
8	201811	11	2018
9	201902	02	2019
10	202001	01	2020
11	202008	08	2020
12	202011	11	2020
13	201905	05	2019
14	201908	08	2019
15	201909	09	2019
16	201910	10	2019

SalesPriceScaleDIM

	SALESPRICESCALE	DESCRIPTION
1	Low	< $5,000
2	Medium	between $5,000 and $10,000
3	High	> $10,000

SalesFact_V1

	SEASON	COMPANY_BRANCH	CUSTOMER_TYPE_ID	CATEGORY_ID	TIME_ID	SALESPRICESCALE	NUMBER_OF_EQUIPMENT_SOLD	TOTAL_REVENUE_FOR_SALES	AVERAGE_SALES_PRICE
1	Spring	Richmond	1	6	202011	High	3	78000	26000
2	Winter	Toorak	1	5	201908	High	2	48000	24000
3	Winter	Parkville	2	2	202007	High	2	132000	66000
4	Winter	Chadstone	2	13	202008	High	4	48000	12000
5	Spring	Docklands	2	7	202009	High	2	32000	16000
6	Spring	Toorak	1	9	201810	High	4	43200	10800
7	Winter	Parkville	1	11	202008	High	3	39000	13000
8	Summer	Docklands	2	11	201901	High	2	43200	21600
9	Summer	Docklands	2	9	202002	Medium	2	11200	5600
10	Summer	Fitzroy	2	10	201901	High	3	46800	15600
11	Spring	Clayton	2	13	201810	High	1	14000	14000
12	Spring	Eltham	1	11	202009	High	4	52000	13000
13	Winter	Clayton	2	3	201907	High	4	52800	13200
14	Spring	Parkville	2	7	202009	High	3	117000	39000
15	Spring	Dandenong	2	10	201811	High	4	96000	24000
16	Spring	Caulfield	2	7	202011	High	1	16000	16000

HireFact_V1

	SEA...	COMPANY_BRANCH	CUSTOMER_TYPE_ID	CATEGORY_ID	TIME_ID	NUMBER_OF_EQUIPMENT_HIRED	TOTAL_REVENUE_FOR_HIRING	AVERAGE_HIRE_PRICE
1	Winter	Toorak	1	6	201806	2	900	450
2	Winter	Fitzroy	1	13	201807	2	200	100
3	Winter	Parkville	1	13	201808	3	600	200
4	Winter	Hughesdale	2	6	201808	2	300	300
5	Winter	Caulfield	2	12	201808	1	420	420
6	Spring	Dandenong	1	9	201809	1	80	160
7	Spring	Clayton	1	11	201809	1	100	100
8	Spring	Caulfield	1	1	201811	3	1200	400
9	Spring	Clayton	1	8	201811	3	900	300
10	Summer	Chadstone	2	10	201812	3	390	130
11	Summer	Pakenham	1	8	201812	3	1110	370
12	Autumn	Richmond	1	10	201903	1	240	240
13	Autumn	Prahran	1	5	201904	2	500	500
14	Autumn	Prahran	2	7	201905	1	225	450
15	Winter	Pakenham	1	8	201906	1	320	320
16	Winter	Toorak	2	11	201906	3	90	30

star/snowflake schema Version-2

CustomerDIM

	CUSTOMER_ID	CUSTOMER_TYPE_ID	NAME	GENDER	ADDRESS_ID	PHONE	EMAIL
1	1	1	Regina Isaacson	Female	1	601 627 5878	risaacson0@tamu.edu
2	2	2	Jaime Whate	Male	2	318 998 0883	jwhate1@ucoz.ru
3	3	1	Thaine Hirche	Male	3	276 571 7986	thirche2@reference.com
4	4	1	Deirdre Reddington	Female	4	585 183 1946	dreddington3@cloudflare.com
5	5	1	Domenic Kirrens	Male	5	798 585 9171	dkirrens4@virginia.edu
6	6	1	Kerk Petera	Male	6	856 940 2206	kpetera5@fastcompany.com
7	7	1	Pammie Futter	Female	7	891 227 4556	pfutter6@woothemes.com
8	8	2	Blaire Christopherson	Female	8	872 144 2174	bchristopherson7@photobucket.com
9	9	1	Gaye Kemmis	Female	9	746 484 4734	gkemmis8@vimeo.com
10	10	2	Cherise Alessandretti	Female	10	501 251 3910	calessandretti9@auda.org.au
11	11	2	Kimmi Deeks	Female	11	128 972 8249	kdeeksa@who.int
12	12	1	Leticia Braiden	Female	12	367 506 7975	lbraidenb@dailymail.co.uk
13	13	2	Orel Greschik	Female	13	866 848 2152	ogreschikc@facebook.com
14	14	1	Saw Gulliver	Male	14	387 132 6717	sgulliverd@paypal.com
15	15	1	Francesco Della	Male	15	447 322 8294	fdellae@icio.us
16	16	2	Edi Larrosa	Female	16	230 157 1885	elarrosaf@360.cn

CustomerTypeDIM

	CUSTOMER_TYPE_ID	DESCRIPTION
1	1	Individual
2	2	Business

StaffDIM

	STAFF_ID	FIRST_NAME	LAST_NAME	GENDER	PHONE	EMAIL	COMPANY_BRANCH
1	1	Carleen	Razzell	Female	323 545 5764	carleen.razzell@monequip.com.au	Caulfield
2	2	Ailee	Paxeford	Female	987 455 1555	ailee.paxeford@monequip.com.au	Hughesdale
3	3	Elissa	Danovich	Female	286 378 7209	elissa.danovich@monequip.com.au	Clayton
4	4	Sonnnie	Chestnutt	Female	245 231 1339	sonnnie.chestnutt@monequip.com.au	Toorak
5	5	Mariska	Holtum	Female	262 960 8943	mariska.holtum@monequip.com.au	Clayton
6	6	Egbert	Earl	Male	290 507 8778	egbert.earl@monequip.com.au	Eltham
7	7	Marylinda	Chanders	Female	398 888 9947	marylinda.chanders@monequip.com.au	Chadstone
8	8	Marcella	Diggons	Female	395 748 7317	marcella.diggons@monequip.com.au	Docklands
9	9	Bethina	Gateman	Female	891 703 6967	bethina.gateman@monequip.com.au	Parkville
10	10	Felecia	Stobbart	Female	735 724 1655	felecia.stobbart@monequip.com.au	Caulfield
11	11	Gratia	MacAlinden	Female	986 594 1206	gratia.macalinden@monequip.com.au	Pakenham
12	12	Arleen	Addison	Female	827 178 5759	arleen.addison@monequip.com.au	Clayton
13	13	Ike	Chadbourne	Male	927 633 9154	ike.chadbourne@monequip.com.au	Pakenham
14	14	Dawn	Vaadeland	Female	643 505 2513	dawn.vaadeland@monequip.com.au	Dandenong
15	15	Fergus	Colvill	Male	477 108 6942	fergus.colvill@monequip.com.au	Richmond
16	16	Daffi	Sapp	Female	896 791 6575	daffi.sapp@monequip.com.au	Clayton

EquipmentDIM

	STAFF_ID	FIRST_NAME	LAST_NAME	GENDER	PHONE	EMAIL	COMPANY_BRANCH
1	1	Carleen	Razzell	Female	323 545 5764	carleen.razzell@monequip.com.au	Caulfield
2	2	Ailee	Paxeford	Female	987 455 1555	ailee.paxeford@monequip.com.au	Hughesdale
3	3	Elissa	Danovich	Female	286 378 7209	elissa.danovich@monequip.com.au	Clayton
4	4	Sonnnie	Chestnutt	Female	245 231 1339	sonnnie.chestnutt@monequip.com.au	Toorak
5	5	Mariska	Holtum	Female	262 960 8943	mariska.holtum@monequip.com.au	Clayton
6	6	Egbert	Earl	Male	290 507 8778	egbert.earl@monequip.com.au	Eltham
7	7	Marylinda	Chanders	Female	398 888 9947	marylinda.chanders@monequip.com.au	Chadstone
8	8	Marcella	Diggons	Female	395 748 7317	marcella.diggons@monequip.com.au	Docklands
9	9	Bethina	Gateman	Female	891 703 6967	bethina.gateman@monequip.com.au	Parkville
10	10	Felecia	Stobbart	Female	735 724 1655	felecia.stobbart@monequip.com.au	Caulfield
11	11	Gratia	MacAlinden	Female	986 594 1206	gratia.macalinden@monequip.com.au	Pakenham
12	12	Arleen	Addison	Female	827 178 5759	arleen.addison@monequip.com.au	Clayton
13	13	Ike	Chadbourne	Male	927 633 9154	ike.chadbourne@monequip.com.au	Pakenham
14	14	Dawn	Vaadeland	Female	643 505 2513	dawn.vaadeland@monequip.com.au	Dandenong
15	15	Fergus	Colvill	Male	477 108 6942	fergus.colvill@monequip.com.au	Richmond
16	16	Daffi	Sapp	Female	896 791 6575	daffi.sapp@monequip.com.au	Clayton

CategoryDIM

	CATEGORY_ID	CATEGORY_DESCRIPTION
1	1	Access
2	2	Air Compressor
3	3	Compaction
4	4	Concrete
5	5	Earthmoving
6	6	Generators
7	7	Landscaping
8	8	Lighting
9	9	Plumbing
10	10	Rail
11	11	Safety
12	12	Site Equipment
13	13	Trailers
14	14	Vehicles
15	15	null

TimeDIM_V2

	TIME_ID	TIME_DATE	SEASON	SEASON_DESCRIPTION
1	20181117	17/NOV/18	Spring	Sep-Nov
2	20190210	10/FEB/19	Summer	Dec-Feb
3	20190218	18/FEB/19	Summer	Dec-Feb
4	20190221	21/FEB/19	Summer	Dec-Feb
5	20190302	02/MAR/19	Autumn	Mar-May
6	20190605	05/JUN/19	Winter	Jun-Aug
7	20190805	05/AUG/19	Winter	Jun-Aug
8	20191115	15/NOV/19	Spring	Sep-Nov
9	20200126	26/JAN/20	Summer	Dec-Feb
10	20200517	17/MAY/20	Autumn	Mar-May
11	20200711	11/JUL/20	Winter	Jun-Aug
12	20200920	20/SEP/20	Spring	Sep-Nov
13	20201120	20/NOV/20	Spring	Sep-Nov
14	20201205	05/DEC/20	Summer	Dec-Feb
15	20180521	21/MAY/18	Autumn	Mar-May
16	20180612	12/JUN/18	Winter	Jun-Aug

SalesPriceScaleDIM

	SALESPRICESCALE	DESCRIPTION
1	Low	< $5,000
2	Medium	between $5,000 and $10,000
3	High	> $10,000

SalesFact_V2

	TIME_ID	STAFF_ID	CUSTOMER_TYPE_ID	CATEGORY_ID	EQUIPMENT_ID	CUSTOMER_ID	TOTAL_SALES_PRICE
1	20201105	31	1	6	62	1	78000
2	20180731	26	1	11	123	1	11200
3	20201215	37	2	2	20	2	182000
4	20201215	37	2	2	20	2	182000
5	20201215	19	1	8	91	4	192000
6	20200517	10	1	2	18	5	130000
7	20180710	39	1	5	58	6	162000
8	20190828	4	1	5	52	7	48000
9	20200612	10	2	9	102	8	138600
10	20180510	39	1	7	77	9	83200
11	20180709	23	2	5	48	10	338000
12	20190805	17	2	3	34	19	18000
13	20181225	20	2	11	120	19	36000
14	20200722	9	2	2	25	21	132000
15	20200510	49	2	5	50	24	54000
16	20200821	34	2	13	144	24	48000

HireFact_V2

	TIME_ID	STAFF_ID	CUSTOMER_TYPE_ID	CATEGORY_ID	EQUIPMENT_ID	CUSTOMER_ID	TOTAL_HIRE_PRICE
1	20180511	2	1	12	135	77	240
2	20180517	38	2	5	49	70	1320
3	20180518	17	2	10	117	58	150
4	20180521	41	2	3	36	124	540
5	20180521	35	2	3	37	87	1200
6	20180522	31	1	5	53	15	1000
7	20180524	10	1	7	73	77	420
8	20180525	40	2	12	135	140	240
9	20180528	46	2	11	127	8	170
10	20180529	28	1	8	86	5	200
11	20180612	21	1	6	71	145	900
12	20180615	43	1	2	23	74	360
13	20180624	10	1	6	61	114	450
14	20180625	44	1	8	85	110	315
15	20180630	8	2	1	9	47	360
16	20180705	7	1	7	74	129	140