Laboratory 06
Level of Aggregation

Case Study 1 – Clothing Company

Let's consider a Clothing case study. The operational database consists of the following tables:

CUSTOMER1

CUSTID	LNAME	FNAME	ADDRESS	PHONE	CITY
107	Smith	John	731 Plenty Road	9231455	Clayton
232	Wong	Franklin	638 Voss Street	9756945	Preston
133	Zelaya	Alicia	3321 Castle Ave	9867055	Balwyn
154	Wallace	Jennifer	291 Berry Street	9234536	Preston
179	Narayan	Ramesh	975 Fire Road	9456738	Carlton
181	Jane	Adam	229 Clayton Road	9543877	Clayton
183	Judy	Backhouse	122 Rose Street	9235345	Caulfield

ORDER1

ORDERID	ORDERDATE	PAYMETHOD	ORDERSOURCE	CUSTID
1057	20/02/06	CARD	WEB SITE	107
1058	03/03/06	CARD	PHONE	232
1059	12/03/06	CHEQUE	WEB SITE	133
1060	20/03/06	CHEQUE	WEB SITE	133
1061	10/04/06	CARD	FAX	179
1062	01/04/06	CARD	FAX	179
1063	07/09/06	CARD	WEB SITE	154
1064	14/07/06	CARD	WEB SITE	154
1065	30/11/06	CARD	PHONE	179
1066	20/01/06	CHEQUE	WEB SITE	179

ORDER_INV1

ORDERID	INVID	ORDERPRICE	QUANTITY
1057	11668	259.99	10
1058	11668	239.99	20
1059	11780	21.99	5
1060	11776	20.99	50
1061	11779	29.95	25
1061	11780	29.95	50
1062	11669	229.99	40
1063	11778	25.95	50
1064	11779	29.95	12
1065	11780	26.95	32
1066	11775	29.95	30

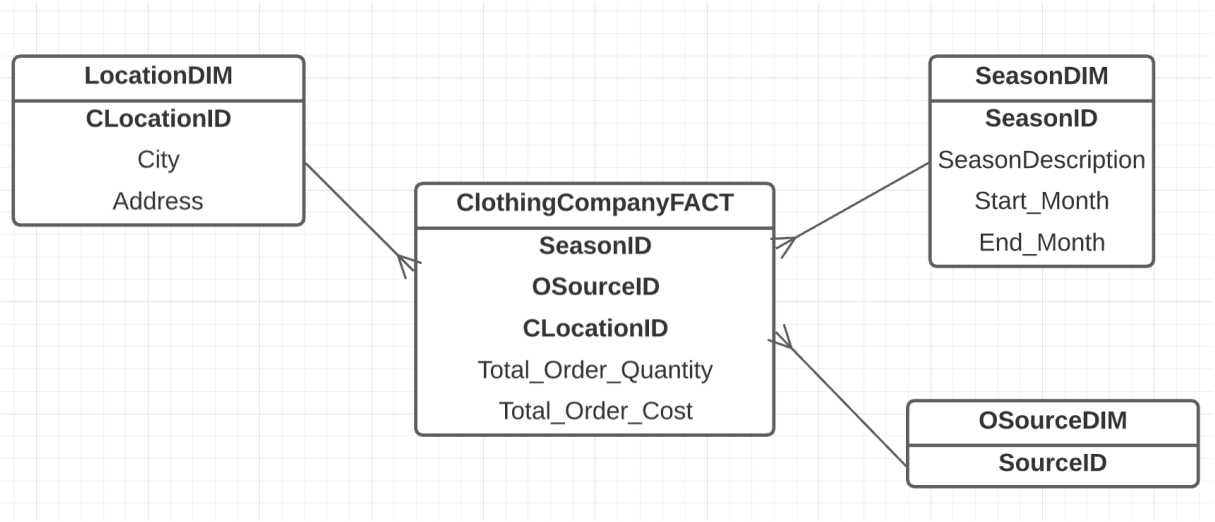
INVENTORY1 (*QOH = Quantity on Hand*)

INVID	QOH	ITEMID	ITEMSIZE	COLOUR
11668	16	786	M	Sienna

11669	12	786	L	Forest
11775	150	894	S	Khaki
11776	147	894	M	Khaki
11777	0	894	L	Khaki
11778	139	894	S	Olive
11779	137	894	M	Olive
11780	115	894	L	Olive

ITEM1			
ITEMID	**CURRENT PRICE**	**ITEMDESC**	**CATEGORY**
894	29.95	Women's Hiking Shorts	Women's Clothing
897	200.95	Women's Fleece Pullovers	Women's Clothing
995	50.00	Children's Beachcomber Sandals	Children's Clothing
559	35.00	Men's Expedition Parka	Men's Clothing
786	259.99	3-Season Jacket	Men's Clothing

A star schema has been created:



The fact measures included in the above star schema are TotalOrderQuantity, which is taken from the Quantity attribute from table O**RDER_INV1**, and TotalOrderCost, which is based on order price * quantity.

The above star schema contains highly aggregated data, and let's assume that this star schema is at level-2 in the data warehouse architecture.

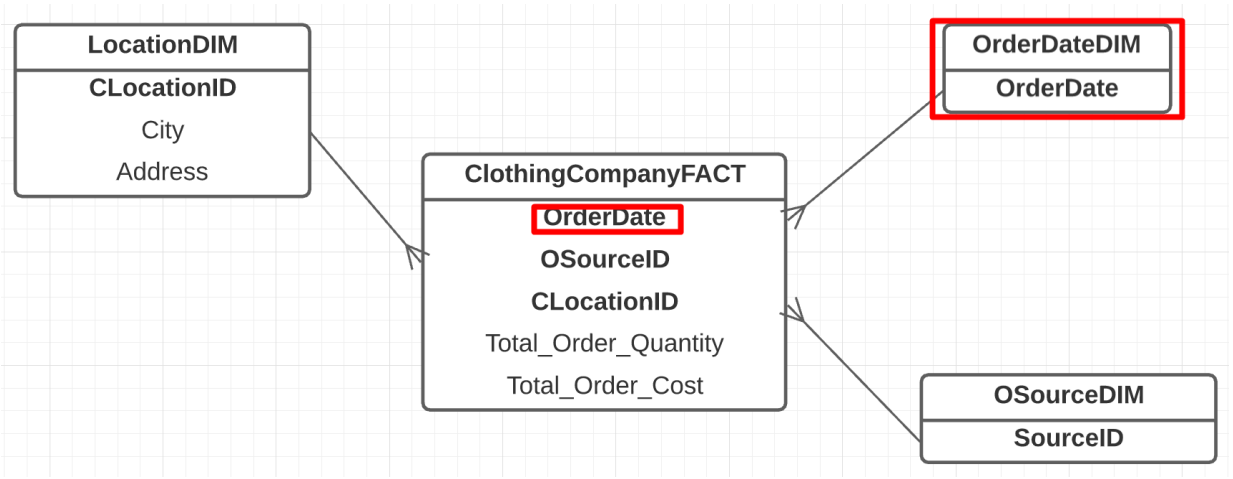
Questions: Draw and implement level-1 and level-0 star schemas for the above clothing data warehouse case study.

The following operational databases have been provided for you:
clothing.Customer1: table that stores information about customer.
clothing.Inventory1: table that stores information about inventory.
clothing.Item1: table that stores information about item.
clothing.Order_Inv1: table that stores information about order and inventory.
clothing.Order1: table that stores information about order.

Sample answer:

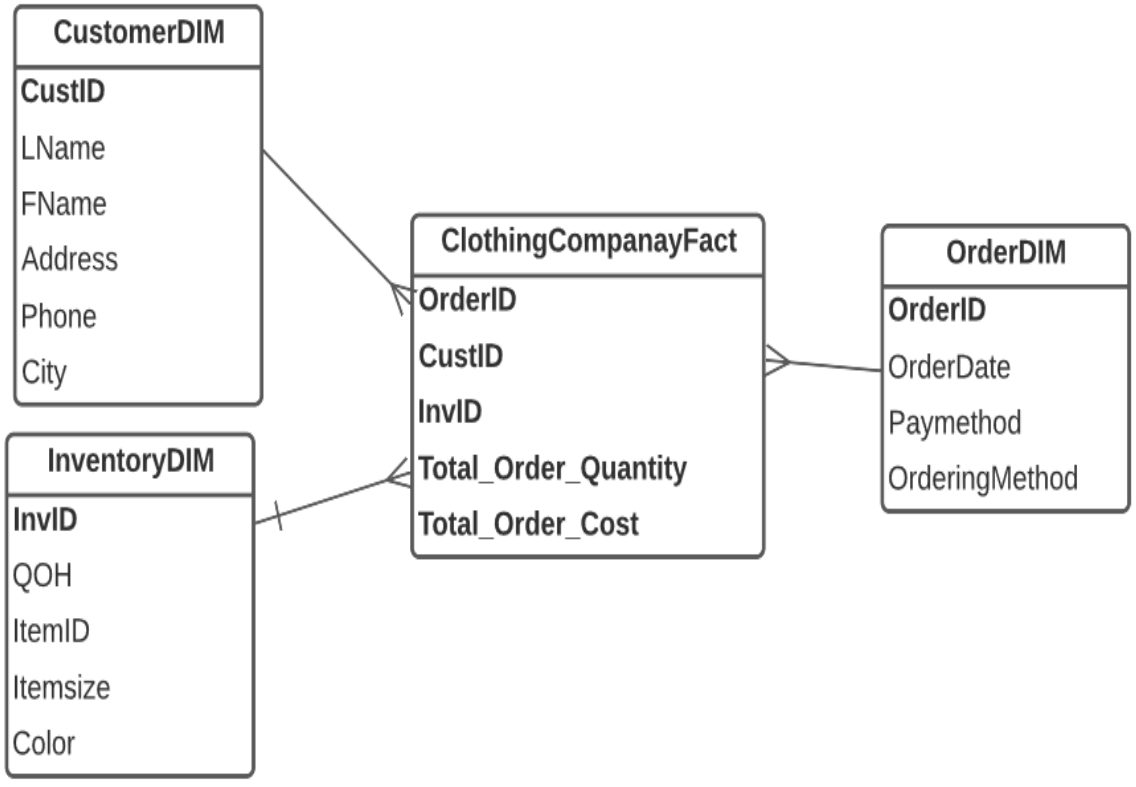
There are two aggregated dimensions: SeasonDIM and LocationDIM.

In Level-1, one of these dimensions is replaced by a lower level of aggregation. For example, SeasonDIM is changed to OrderDateDIM, or LocationDIM is changed to CustomerDIM.
In Level-1, only one of these dimensions needs to be lowered down.



In Level-0, both dimensions need to be lowered down.

Level-0



IMPLEMENTATION
LEVEL 2

```
--LEVEL 2

-- create location dimension
drop table locationdim2 purge;
create table locationdim2 as
select distinct address || ' ' || city as clocationid, city, address
from clothing.customer1;
```

```
-- create season dimension
drop table seasondim2 purge;
create table seasondim2 (
seasonid number(2), seasondescription varchar2(20), start_month varchar2(20), end_month
varchar2(20));
insert into seasondim2 values (1, 'Summer', 'December', 'February');
insert into seasondim2 values (2, 'Autumn', 'March', 'May');
insert into seasondim2 values (3, 'Winter', 'June', 'August');
insert into seasondim2 values (4, 'Spring', 'September', 'November');

select * from seasonDIM2;
-- create source dimension
drop table osourcedim2 purge;
create table osourcedim2
as select distinct ordersource as sourceid
from clothing.order1;

-- create tempfact table
drop table tempclothingcompanyfact purge;
create table tempclothingcompanyfact as
select o.ordersource, o.orderdate, c.address || ' ' || c. city as clocationid,
i.quantity, i.orderprice
from clothing.order1 o, clothing.customer1 c, clothing.order_inv1 i
where o.custid=c.custid and o.orderid=i.orderid;

-- add column and update value
alter table tempclothingcompanyfact add seasonid number(2);

update tempclothingcompanyfact
set seasonid=
case
when to_char(orderdate,'mm') in ('12','01','02') then 1
when to_char(orderdate,'mm') in ('03','04','05') then 2
when to_char(orderdate,'mm') in ('06','07','08') then 3
else 4
end;

-- create clothingcompanyfact
drop table clothingcompanyfact2 purge;

create table clothingcompanyfact2
as
select clocationid, seasonid, ordersource as osourceid,
sum(quantity) as total_order_quantity,
sum(orderprice* quantity) as total_order_cost
from tempclothingcompanyfact
group by seasonid, ordersource, clocationid;
```

LEVEL 1

```
-- create location dimension
drop table locationdim1 purge;

create table locationdim1 as
select distinct address || ' ' || city as clocationid, city, address
from clothing.customer1;

-- create order date dimension
drop table orderdatedim1 purge;

create table orderdatedim1 as
```

```
select distinct orderdate from clothing.order1;

-- create source dimension
drop table osourcedim1 purge;

create table osourcedim1 as
select distinct ordersource as sourceid
from clothing.order1;

-- create clothingcompanyfact
drop table clothingcompanyfact1 purge;

create table clothingcompanyfact1 as
select c.address || ' ' || c. city as clocationid, o.orderdate, o.ordersource as osourceid,
sum(quantity) as total_order_quantity,
sum(orderprice*quantity ) as total_order_cost
from clothing.order1 o, clothing.customer1 c, clothing.order_inv1 i
where o.custid=c.custid and o.orderid=i.orderid
group by c.address || ' ' || c. city, o.orderdate, ordersource;
```

LEVEL 0

```
-- create customer dimension
drop table customerdim0 purge;

create table customerdim0 as
select distinct *
from clothing.customer1;

--create inventory dimension
drop table inventorydim0 purge;

create table inventorydim0
as select distinct *
from clothing.inventory1;

-- create order date dimension
drop table orderdim0 purge;

create table orderdim0 as
select distinct * from clothing.order1;

-- create clothingcompanyfact
drop table clothingcompanyfact0 purge;

create table clothingcompanyfact0 as
select c.custid, o.orderid,i.invid,
sum(i.quantity) as total_order_quantity,
sum(i.orderprice*i.quantity) as total_order_cost
from clothing.order1 o, clothing.customer1 c, clothing.order_inv1 i
where o.custid=c.custid and o.orderid=i.orderid
group by c.custid, o.orderid,i.invid;
```

Case Study 2 – Toll Way

There is a toll way in a metropolitan city (such as CityLink or EastLink in Melbourne, or any similar toll roads in other major cities in the world). This toll way has a number of gates, where the motorist needs to pay. Every time a motorist passes through this toll gate, the registration number of the vehicle, vehicle type (e.g. car, bus, truck, etc), amount paid, and time, are recorded in the operational database.

A data warehouse needs to be built, for analysing the *revenue* from the toll payments. The management would like to drill down this revenue based on the *tollgate* (there is a number of toll gates along the toll way), *day of week* (e.g. weekdays, weekends), and *time period of a day* (e.g. peak hours, non-peak hours, late nights).

You are required to draw **three levels of star schemas** showing three different levels of aggregation for the above data warehouse. Level-0 star schema contains the most detailed data, whereas level-2 star schema is the highly aggregated (e.g. containing highly aggregated data).

Questions:

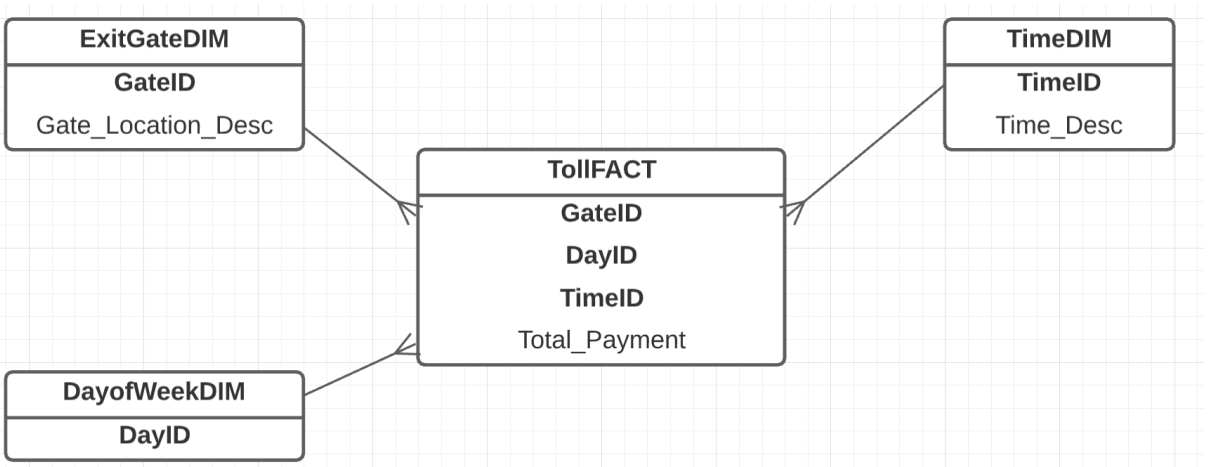
(a) Draw a level-2 star schema.

(b) Draw a level-1 star schema. You may want to add a new dimension, called *vehicle* (e.g. cars, trucks, busses, etc).

(c) Draw a level-0 star schema.

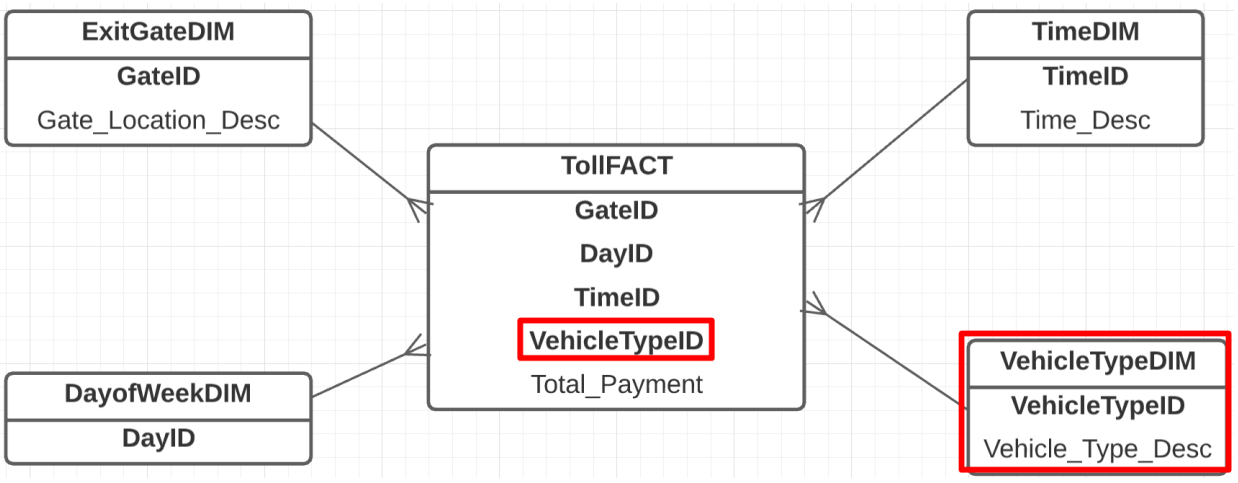
Sample answer:

Level 2 – Highly aggregated. This is the most highly aggregated data where we keep "Total Payment" in the Fact table, and the Dimensions are based on Days of Week (not the actual date), and Time Period (not the actual travel time).



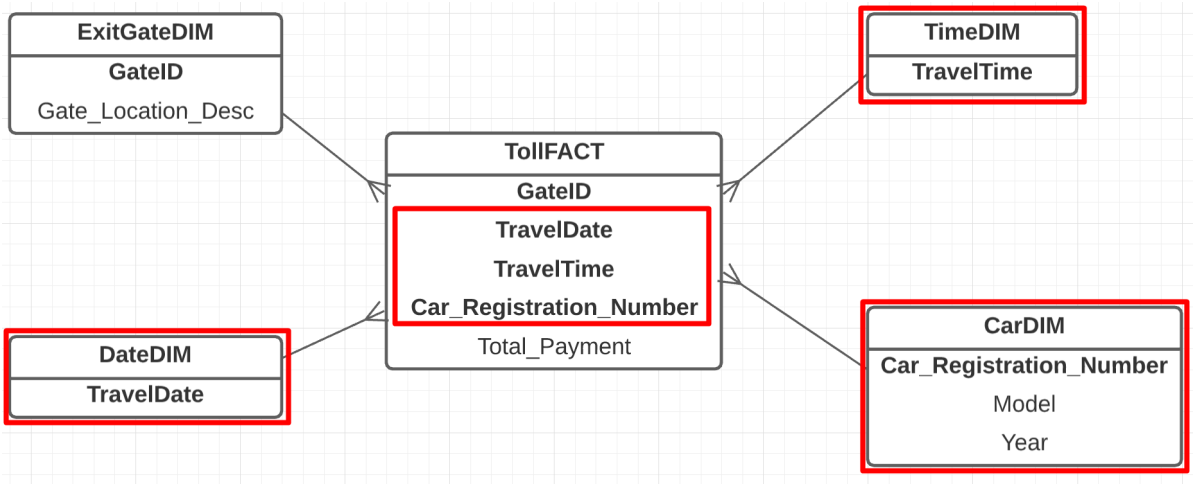
Level 1– We add with an additional dimension, for example the Vehicle Type Dimension. In the Vehicle Type Dimension, we store the vehicle types, such as bus, truck, cars, etc. Note that we do not keep the "Registration Number" of the car. The granularity is at the vehicle type level, not at an individual car level.

So, from level 2 to level 1, the total payment in level 2 is broken down by Vehicle Type. Other dimensions remain unchanged.



Level 0– Detail Level

This is the most detail level as we have Date_DIM and Time_DIM to indicate the actual travel dates and travel time; as well as Car DIM to indicate the actual car (rather than just the type of vehicle)



THE END