**Data cleaning, creating new tables, querying**

Please do this part on MySQL or SQL Server or any type of SQL  
-------------------------------------------------------------------------------------  
Please find attached csv files for 3 tables  
  
ims\_sales\_order\_item  
ims\_sales\_order\_item\_status  
ims\_sales\_order\_item\_status\_history  
  
  
ims\_sales\_order\_item - each row in this table represents one item.  
Primary key:  
id\_sales\_order\_item  
Foreign keys:  
fk\_sales\_order  
fk\_sales\_order\_item\_status  
  
ims\_sales\_order\_item\_status - this table provides all the possible status that an item can have.  
Primary key:  
id\_sales\_order\_item\_status  
Foreign keys:  
fk\_oms\_function  
  
ims\_sales\_order\_item\_status\_history - this table provides all the status that each item has gone through. The created\_at field provides the timestamp at which the item when through that particular status.  
Primary key:  
id\_sales\_order\_item\_status\_history  
Foreign keys:  
fk\_sales\_order\_item  
fk\_sales\_order\_item\_status  
  
Foreign key - Primary key relationships are as below:  
fk\_sales\_order\_item\_status from ims\_sales\_order\_item is linked to id\_sales\_order\_item\_status from ims\_sales\_order\_item\_status  
fk\_sales\_order\_item\_status from ims\_sales\_order\_item\_status\_history is linked to id\_sales\_order\_item\_status from ims\_sales\_order\_item\_status  
fk\_sales\_order\_item from ims\_sales\_order\_item\_status\_history is linked to id\_sales\_order\_item from ims\_sales\_order\_item  
  
  
Question 1. Write a SQL query to display the unit\_price, name, SKU, shipping\_type of items with id\_sales\_order\_item = 229884  
  
Question 2. Write a SQL query to display the id\_sales\_order\_item, name, SKU, sales\_order\_item\_status for all the items with shipping\_type = warehouse and status = delivered  
  
Question 3. Write a SQL query to display the total unit\_price of all items which have shipping\_type = cross\_docking  
  
Question 4. Write a SQL query to display the id\_sales\_order\_item, name, SKU for all the items that went to status = picklisted on anytime during the day on 11/6/2013  
  
Question 5. Write a SQL query to display the total count of items for each of the 3 shipping types - marketplace, cross\_docking, warehouse. The count should include only items which have a unit\_price between 50 and 100, and also went to status = picklisted on anytime during the day on 11/6/2013  
  
Question 6. Write a SQL query to display the fk\_sales\_order, id\_sales\_order\_item, name, SKU, sales\_order\_item\_status for all the items which have status before last status  
  
(example for question 6: Item 1 have status a,b,c,d,e,f,g. The last status of the item 1 is g and the status before last status is f )

***Hệ CSDL: Hive SQL***

***Giao diện tương tác: DBeaver***

***Phương án đã thực hiện:***

* ***Create table từ các file csv lên 1 folder được quy hoạch trên DB***
* ***Query từ bảng đã tạo***

**Question 1:**

**select** id\_sales\_order\_item,

unit\_price,

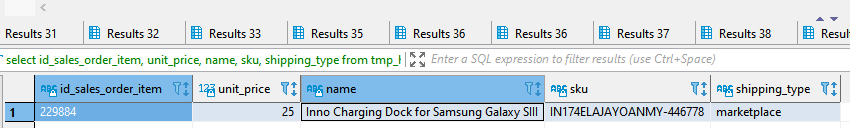
name,

sku,

shipping\_type

**from** tmp\_hanhpn7\_sales\_order\_item\_3

**where** id\_sales\_order\_item = 229884



**Question 2:**

**select** a.\*, b.status

**from**

(**select** id\_sales\_order\_item,

name,

SKU,

fk\_sales\_order\_item\_status,

shipping\_type

**from**

tmp\_hanhpn7\_sales\_order\_item\_3) a

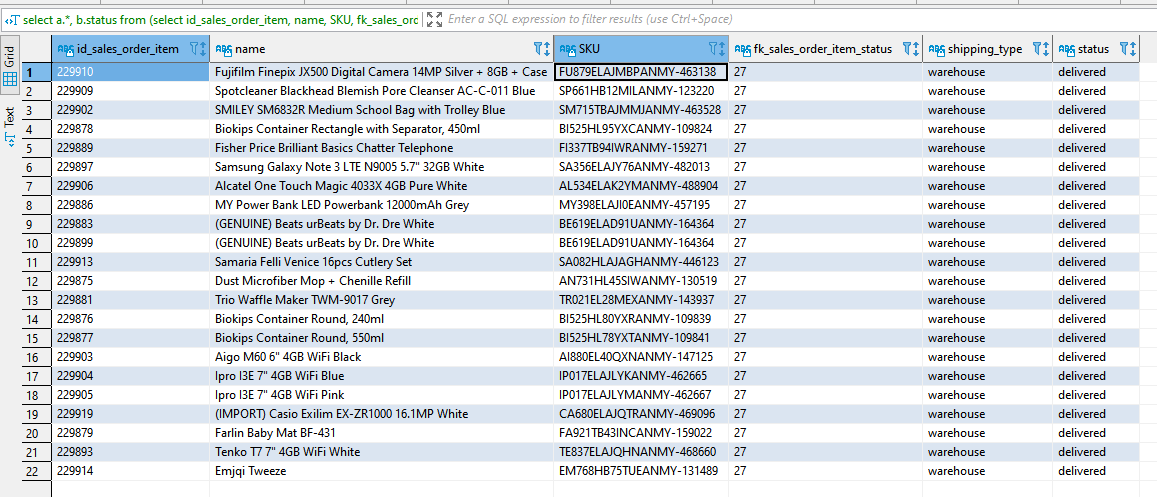
**left** **join**

(**select** id\_sales\_order\_item\_status, status

**from** tmp\_hanhpn7\_sales\_order\_item\_status\_3) b

**on** a.fk\_sales\_order\_item\_status = b.id\_sales\_order\_item\_status

**where** a.shipping\_type = 'warehouse' **and** b.status = 'delivered'



**Question 3:**

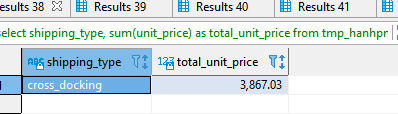
**select** shipping\_type,

**sum**(unit\_price) **as** total\_unit\_price

**from** tmp\_hanhpn7\_sales\_order\_item\_3

**where** shipping\_type = 'cross\_docking'

**group** **by** 1;



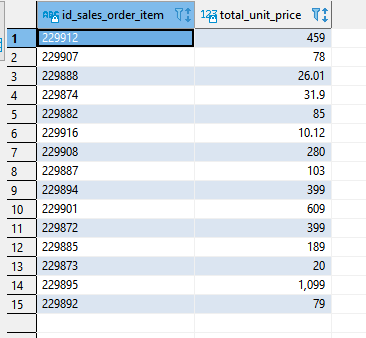
**select** id\_sales\_order\_item,

**sum**(unit\_price) **as** total\_unit\_price

**from** tmp\_hanhpn7\_sales\_order\_item\_3

**where** shipping\_type = 'cross\_docking'

**group** **by** 1;



**Question 4:**

**select** item\_info.id\_sales\_order\_item,

item\_info.name,

item\_info.SKU,

item\_list.id\_sales\_order\_item\_status,

item\_list.status,

item\_list.updated\_at

**from**

(**select** b.fk\_sales\_order\_item,

a.id\_sales\_order\_item\_status,

a.status,

a.updated\_at

**from**

(**select** id\_sales\_order\_item\_status, status, updated\_at

**from** tmp\_hanhpn7\_sales\_order\_item\_status\_3

**where** status = 'picklisted' **and** updated\_at = 20131106) a

**left** **join**

(**select** fk\_sales\_order\_item,

fk\_sales\_order\_item\_status

**from** tmp\_hanhpn7\_sales\_order\_item\_history\_3) b

**on** a.id\_sales\_order\_item\_status = b.fk\_sales\_order\_item\_status) item\_list

**left** **join**

(**select**

id\_sales\_order\_item,

name,

SKU

**from** tmp\_hanhpn7\_sales\_order\_item\_3) item\_info

**on** item\_list.fk\_sales\_order\_item = item\_info.id\_sales\_order\_item



**Question 5:**

**select** **count** (picklisted\_item.fk\_sales\_order\_item),

**count** (**distinct** picklisted\_item.fk\_sales\_order\_item)

**from**

(**select** b.fk\_sales\_order\_item,

a.id\_sales\_order\_item\_status

**from**

(**select** id\_sales\_order\_item\_status

**from** tmp\_hanhpn7\_sales\_order\_item\_status\_3

**where** status = 'picklisted' **and** updated\_at = 20131106) a

**left** **join**

(**select** fk\_sales\_order\_item,

fk\_sales\_order\_item\_status

**from** tmp\_hanhpn7\_sales\_order\_item\_history\_3) b

**on** a.id\_sales\_order\_item\_status = b.fk\_sales\_order\_item\_status) picklisted\_item

**left** **join**

(**select**

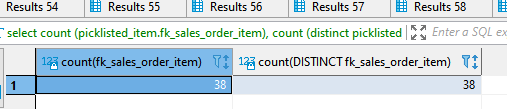
id\_sales\_order\_item,

unit\_price

**from** tmp\_hanhpn7\_sales\_order\_item\_3

**where** unit\_price >= 50 **and** unit\_price <= 100) price\_item

**on** picklisted\_item.fk\_sales\_order\_item = price\_item.id\_sales\_order\_item



**Question 6**

----list of items which have status before last status

**select** all\_item\_list.fk\_sales\_order\_item,

all\_item\_list.fk\_sales\_order\_item\_status,

last\_status\_list.id\_last\_status

**from**

(**select**

fk\_sales\_order\_item,

fk\_sales\_order\_item\_status

**from** tmp\_hanhpn7\_sales\_order\_item\_history\_3) all\_item\_list

**left** **join**

(**select**

fk\_sales\_order\_item,

**max**(fk\_sales\_order\_item\_status) **as** id\_last\_status

**from** tmp\_hanhpn7\_sales\_order\_item\_history\_4

**group** **by** fk\_sales\_order\_item) **as** last\_status\_list

**on** all\_item\_list.fk\_sales\_order\_item = last\_status\_list.fk\_sales\_order\_item

**where** all\_item\_list.fk\_sales\_order\_item\_status < last\_status\_list.id\_last\_status

---display fk\_sales\_order, id\_sales\_order\_item, name, SKU, sales\_order\_item\_status

**select** b.id\_sales\_order\_item,

b.fk\_sales\_order,

b.name,

b.SKU,

a.fk\_sales\_order\_item\_status

**from**

(**select** all\_item\_list.fk\_sales\_order\_item,

all\_item\_list.fk\_sales\_order\_item\_status,

last\_status\_list.last\_status

**from**

(**select**

fk\_sales\_order\_item,

fk\_sales\_order\_item\_status

**from** tmp\_hanhpn7\_sales\_order\_item\_history\_4) all\_item\_list

**left** **join**

(**select**

fk\_sales\_order\_item,

**max**(fk\_sales\_order\_item\_status) **as** last\_status

**from** tmp\_hanhpn7\_sales\_order\_item\_history\_3

**group** **by** fk\_sales\_order\_item) **as** last\_status\_list

**on** all\_item\_list.fk\_sales\_order\_item = last\_status\_list.fk\_sales\_order\_item

**where** all\_item\_list.fk\_sales\_order\_item\_status < last\_status\_list.last\_status) a

**left** **join**

(**select** fk\_sales\_order, id\_sales\_order\_item, name, SKU

**from** tmp\_hanhpn7\_sales\_order\_item\_3) b

**on** a.fk\_sales\_order\_item = b.id\_sales\_order\_item

