1. A 3-D view of sales data for company A, according to the dimensions time, item, and location is given in Table 1. The measure displayed is dollars_sold in thousands.
    a. Find the 3-D data cube representation of the data in Table 1.
    b. Show the results of roll-up on location from cities to countries.
    c. Show the results of drill-down on time from quarters to months.
    d. Show the results of slice for time = Q3.
    e. Show the results of dice for location = New York or Chicago and time = Q3 or Q4 and item comp. or phone.

Table 1.

	Location=`Chicago`				Location=`New York`				Location=`Toronto`				Location = `Vancouver`			
	item				item				item				item			
time	Home E.	Comp.	Phone	Sec.	Home E.	Comp.	Phone	Sec.	Home E.	Comp.	Phone	Sec.	Home E.	Comp.	Phone	Sec.
Q1	854	882	89	623	1087	968	38	872	818	746	43	591	605	825	14	400
Q2	943	890	64	698	1130	1024	41	925	894	769	52	682	680	952	31	512
Q3	1032	924	59	789	1034	1048	45	1002	940	795	58	728	812	1023	30	501
Q4	1129	992	63	870	1142	1091	54	984	978	864	59	784	927	1038	38	580

2. Explain the differences and similarities of star and snowflake schemas. Analyze their advantages and disadvantages.
3. Briefly explain the advantageous of vertical data format approach to mine frequent itemsets.
4. Explain the differences and similarities of data cleaning and data transformation.
5. Explain anti-monotone and monotone rule constraints with examples.
6. The following contingency table (Table 2) summarizes supermarket transaction data, where H refers to the transactions containing hot dogs, H' refers to the transactions that do not contain hot dogs, HA refers to the transactions containing hamburgers, and HA' refers to the transactions that do not contain hamburgers.
    a. Suppose that the association rule H--> HA is mined. Given a minimum support threshold of 25% and a minimum confidence threshold of 50%, is this association rule strong? Explain why.
    b. Based on the given data, is the purchase of hot dogs independent of the purchase of hamburgers? If not, what kind of correlation relationship exists between the two?

    Table 2.

	H	H'	Sum_Row
HA	2000	500	2500
HA'	1000	1500	2500
Sum_Column	3000	2000	5000

7. Table 3 is a database containing five transactions. Suppose that min_sup (*s*) = 60% and min_conf (*c*) = 80%. Find all frequent itemsets using Apriori. List all of the strong association rules with support *s* and confidence *c*.

    Table 3.

TID	Items
t1	M, O, N, K, E, Y
t2	D, O, N, K, E, Y
t3	M, A, K, E
t4	M, U, C, K, Y
t5	C, O, O, K, I, E