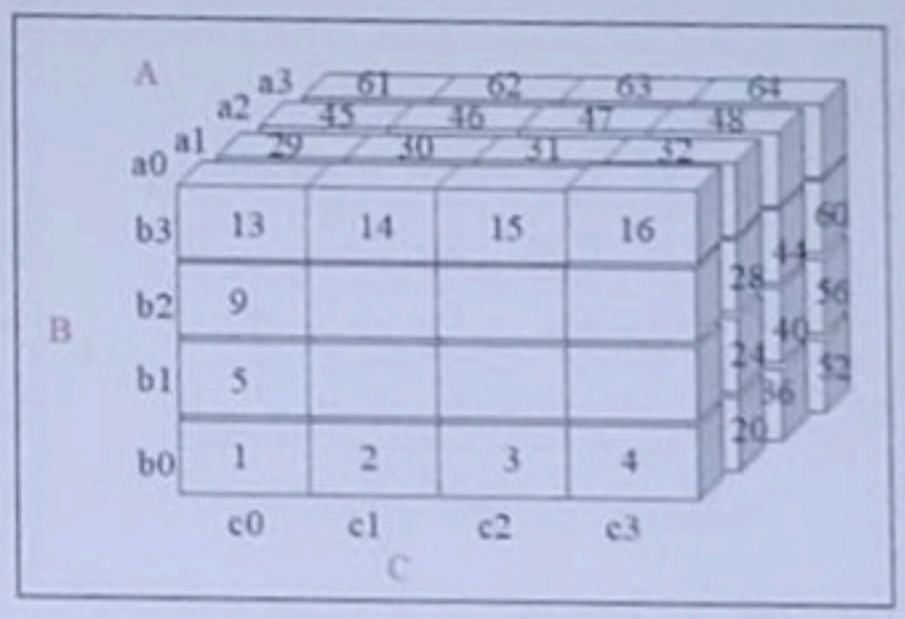
Question No. 1. Consider the following data cube where size(A) = 2000, size(B) = 200 and size(C) = 20. The size of chunk of A, B and C are 500, 50 and 5 respectively. Calculate the number of operations required for the traversing order for {1,5,9,13,2,6,10,…60,64}.



If the ordering is 1,5,9,13,2,6,10,…60,64 memory requirement is: B → C → A (BC → BA → CA

(full data size for BC) 200*20
+ (full data size of B and chunk of A for BA) 200*500
+ (chunks of C&A for CA) 5*500
= 4,000 + 10,0000 + 2,500 = 106,500

Question No. 2. Consider the following base table of Saudi Airline. Analyze the **Query: "Asia" and "Retailer"** by SQL with bitmap indexing technique. Write the SQL query for "Asia" and "Retail"

Indexing OLAP Data: Bitmap Index

- Index on a particular column
- Each value in the column has a bit vector: bit-op is fast
- The length of the bit vector: # of records in the base table
- The *i*-th bit is set if the *i*-th row of the base table has the value for the indexed column

Base table

Cust	Region	Type
C1	Asia	Retail
C2	Europe	Dealer
C3	Asia	Dealer
C4	America	Retail
C5	Europe	Dealer

Index on Region

RecID	Asia	Europe	America
1	1	0	0
2	0	1	0
3	1	0	0
4	0	0	1
5	0	1	0

Index on Type

RecID	Retail	Dealer
1	1	0
2	0	1
3	0	1
4	1	0
5	0	1

Base Table: Saudi Airline

Cust	Region	Type
C1	Asia	Retailer
C2	Eurpoe	Dealer
C3	Asia	Dealer
C4	America	Retailer
C5	Europe	Dealer

SELECT Cust
FROM Base Table
WHERE Region="Asia" and Type="Retail"

Question No. 3. Suppose a data cube for eXtra Electronics is created as

Define cube sales [year, product, location]: sum (total sales)

Dimension hierarchies used are
item < brand < supplier
street < city < state < country < continent

Cuboids are:
cuboid 1: {brand, city, year}
cuboid 2: {supplier, country, year}
cuboid 3: {supplier, state, year}
cuboid 4: {brand, state} where year = 2020

Analyze each of four cuboids in term of required number of computations for the following query.

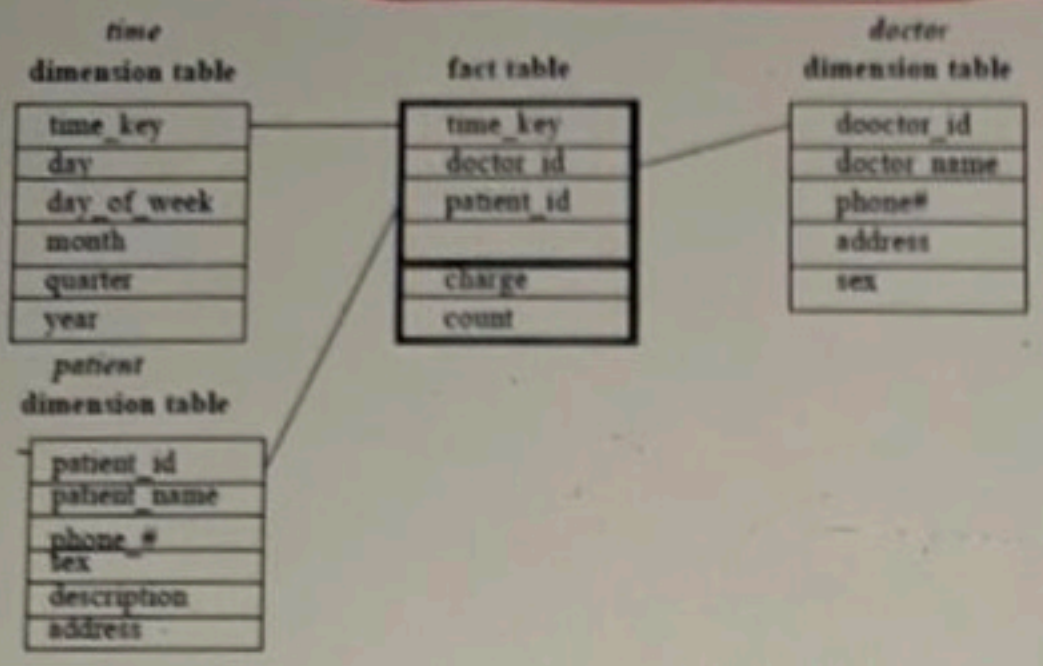
Query {supplier, state} with year = 2020?

cuboid 1 costs the most since brand and city are at lower level than supplier and state
cuboid 2 cannot be used since state ≠ country, include all years
cuboid 3 is better than cuboid 4, if there are few year values associated with brands and several brands for each supplier)
cuboid 4 is better than cuboid 3, if efficient indices of brand to supplier are available.

Question No. 4. Suppose that a data warehouse consists of the three dimensions (*time*, *doctor*, and *patient)*, and the two measures (*count* and *charge)*, where *charge* is the fee that a doctor charges a patient for a visit, and count represent number of patients seen by the doctor.

Dimensional attributes are:
Time (time_key, day, month, quartor, year)
Doctor (doctor_id, doctor_name, specialty, phone#, address)
Patients (patient_id, patient_name, disease, gender, critical)

(a) Draw a schema diagram for the above data warehouse.



(b) Write an SQL query assuming the data is stored in a relational database with the schema *fee* (*day*, *month*, *year*, *doctor*, *patient*, *count*, *charge*) to list the total fee collected by each doctor in 2010 .

SESECT doctor, SUM (charge)
FROM fee
WHERE year = 2010
GROUP BY doctor

Question No. 5. Consider the following four states (S1, S2, S3,S4), three decisions (Gold, Bond, Stock) with mapping table for payoff and corresponding utility values. Find the following under the "Expected Utility Criterion"?:

Pay-0ff	900	600	450	400	300	-150	-200	-400	-500
Utility	100	97	94	80	70	50	45	30	0

Payoff Table				
Alternative	S1	S2	S3	S4
Gold	900	400	-500	300
Bond	450	300	-400	-150
Stock	300	300	-150	450
Prob.	0.4	0.3	0.2	0.1

Utility Table					
Alternative	S1	S2	S3	S4	Expected Utility
Gold	100	80	0	70	71
Bond	94	70	30	50	69.6
Stock	70	70	50	94	68.4

EV(Stock) = 100*0.4+80*0.3+0+70*0.1= 71
The optimal decision is Stock, its expected utility is 71 and worth of its utility is (400-300)/(80-70) = 10 .
So, for EU 71 it will be 300+1(10) = 310

Question No. 6. Data series for the period 1 to 7 is given in the following. The initial forecast at period 3 is given as 272. Find the forecast for the period 1 to 10 using exponential smoothing technique applied to stationary time series for $\alpha = 0.2$. Show all calculations. $F_4 = \alpha Y_3 + (1 - \alpha) F_3$

Period	Series	Forecast
1	272	N/A
2	280	N/A
3	395	0.2(395)+0.8(272) = 272.0
4	438	296.6
5	431	324.9
6	446	346.1
7	354	366.1
8		363.66
9		363.66
10		363.66

Question No. 7. Consider the following forecasting technique applied to stationary time series. Find forecast and errors applied to stationary time series for 4-period weighted moving average technique with probabilities P (0.4, 0.3, 0.2, 0.1). Show all calculation.

Time	1	2	3	4	5	6	7
Data Series	200	220	180	160	210	230	205
4-period weighted moving average					182	190	205
Error for 4-Period WMA					28	40	0

(a) Performance measure using **MSE** for 4-period WMA technique is ___794.66__

(b) Performance measure using **MAPE** for 4-period WMA technique is __0.102____

Question No. 8. Find the Expected Value of Sample Information (EVSI) for the following payoff table? Given information are: prior probability, posterior probability for positive forecast, posterior probability for negative forecast, $\sum$ Joint Probability (+) = 0.59, $\sum$ Joint Probability (-) = 0.41.
EV{Gold, Bond, Stock, IPO} = {182.5, 80, 82.5, 185}

Decision	States of Nature						
Alternatives	Large Rise	Small Rise	No change	Small Fall	Large Fall	Forca(+)	Forca(-)
Gold	500	250	100	-200	-600	281.01	39.26
Bond	-100	100	200	300	0	51.05	122.09
Stock	200	150	150	-200	-150	124.89	20.86
IPO	400	200	150	-100	-150	241.77	102.45
Prior Prob.	0.35	0.25	0.15	0.2	0.05		
Posterior Prob. (+)	0.443	0.295	0.127	0.135	0.000		
Posterior Prob. (-)	0.215	0.184	0.184	0.294	0.123		

ERSI=	0.59*281+0.41*122.1 = 216.25
EVSI=	216.25 – 185 = 31.25