Practice Questions

Question 01:

Artist

AID	Name	City
1	Ali Ahmad	Lahore
2	Alia Butt	Gujranwala
3	Rani Khurram	Faisalabad
4	Hadia Tanzeel	Lahore
5	Tanzeela Akbar	Jhelum
6	Talha Rana	Lahore
7	Zulfiqar Rana	Narowal
8	Sheeraz Ali	Zafarwal
9	Usman Rana	Karachi
10	Wasif Wajid	Islamabad

ArtPiecesDetail

APID	ATID	AID	Release_Date
1	5	1	2018-02-25
2	2	1	2018-03-02
3	3	6	2018-01-05
4	4	4	2017-02-25
9	5	6	2018-01-05
6	1	1	2018-02-25
7	2	10	2017-02-03
8	3	1	2016-12-25
9	4	1	2018-02-25
8	4	3	2018-10-02
3	4	3	2018-03-02

A) Suppose you want to improve the performance of the following queries.Onwhich column you create the index and index type (i.e. Hash index, Bitmap, B++ tree). [2+1+2]

1) Select Name from Artist where city="Lahore"
2) Select AID from Artist where AID > 2
3) Select Name from Artist join ArtPiecesDetail on Artist.AID=ArtPiecesDetail.AID where city="karachi"

Ans: **Column Name** **Index Type**

city ⟶ Hash/B++ tree
AID ⟶ B++ tree
AID, city ⟶ Hash/B++ tree

Q1. *(8 points)* Assume a relation R (A, B, C) is given; R is stored as an ordered file (un-spanned) on non-key field C and contains 100,000 records. Attributes A, B and C need 10 bytes of storage each (i.e. record size= 30), and blocks have a size of 512 Bytes. Each A value occurs at an average 5 times in the database, each B value occurs 50 times in the database, and each C value occurs 5000 times in the database. Assume there is no index structure exists. Estimate the number of block fetches needed to compute the following queries (where C_a, C_b, C_{c1} and C_{c2} are integer constants):

a) SELECT B, C FROM R WHERE A = C_a;
b) SELECT COUNT(*) FROM R WHERE B = C_b;
c) SELECT A, B FROM R WHERE C = C_{c1};
d) SELECT A, B FROM R WHERE C = C_{c1} OR C = C_{c2};

Ans: bfr=512/30=17; b=100,000/17= 5883
a) O(b) = 5883
b) O(b) = 5883
c) O(log(b) + s/bfr - 1) = O(13 + 5000/17 – 1) = O(13 + 295 - 1) = 307
d) 2 * 307 (i.e. same cost as of part-c)= 614

Q3. *(6 points)* Suppose you are building an extensible hash index on a table of 100,000 rows. Key values are 8 bytes, a pointer (block/record) to a row is 8 bytes, and a disk block is 2048 bytes. Assume all keys are distinct.

a) What is the (lowest possible) global depth? Provide valid reasons.
b) What is the average occupancy of a bucket, assuming all buckets have a local depth equal to the global depth from part (a)? Justify your answer.

Ans:
a) Bucket entries will be key/pointer pairs, so 16 bytes each. Floor(2048/16) = 128 entries / bucket. 100,000/128 = at least 782 buckets needed. Since the directory is always a power of 2 size, it will have at least 2^{10} = 1024 entries, so the global depth is 10.

b) If all buckets have local depth equal to global depth, then every pointer in the directory points to a unique bucket. Thus, there are 1024 buckets. 1024 * 128 = capacity of 131,072. 100,000/131,072 ~= 76.3% occupancy.
