

## Exercise 1: Schema Fragmentation (3 Questions)

### Question 1

Given the following database table at a TV channel, the list of queries issued by two applications on this table, and their frequencies, answer the given questions.

PK	player	goals	fouls	viewer satisfaction
1	Beckham	1	10	80
2	Beckham	2	15	95
3	Zidane	1	8	70
4	Zidane	2	12	65
5	Zidane	8	20	75
6	Rooney	1	8	60
7	Klose	1	6	70
8	Klose	4	18	75

Application 1	Application 2
<b>Q1</b> : 70 times monthly <i>fouls &lt;= 10 AND goals &gt;=1</i>	<b>Q1</b> : 100 times monthly <i>goals &gt;=4</i>
<b>Q2</b> : 100 times monthly <i>player = Beckham</i>	<b>Q2</b> : 100 times monthly <i>player = Beckham</i>
<b>Q3</b> : 60 times monthly <i>fouls &lt;= 8</i>	<b>Q3</b> : 100 times monthly <i>NOT player=Beckham AND fouls &gt;=8 AND fouls&lt;=12</i>
<b>Q4</b> : 30 times monthly <i>goals &gt;=2</i>	<b>Q4</b> : 100 times monthly <i>player = Klose AND fouls &lt;=10</i>
<b>Q5</b> : 30 times monthly <i>player = Zidane AND fouls &gt; 20</i>	<b>Q5</b> : 20 times monthly <i>goals &lt;= 1</i>

The following table presents the primary key (PK) of the accessed rows with their monthly access frequencies per query, per application:

	Application 1 Row_PK[,Row_PK...] (Monthly Access)	Application 2 Row_PK[,Row_PK...] (Monthly Access)
Q1	1,3,6,7 (70)	5,8 (100)
Q2	1,2(100)	1,2 (100)
Q3	3,6,7(60)	3,4,6 (100)
Q4	(empty)	(empty)
Q5	(empty)	(empty)

Answer the following questions:

- Fill the empty cells in the above table.
- Create a table of the tuple access frequencies per application.
- Determine the horizontal fragmentation using the Minfrag algorithm and using the tuple access frequency table.

## Question 2

Three police stations ( $P_1$ ,  $P_2$ , and  $P_3$ ) access the following *Police* database where *Number* is the primary key:

Number	Car_type	Owner_name	Year	City
VD111222	Trabant	John Smith	1976	Ecublens
ZH222111	Isuzu	Smit Johnson	1993	Zurich
...	...	...	...	...
GE333111	Smart	Peter Smithson	2001	Geneva

$P_1$ ,  $P_2$ , and  $P_3$  issue the following queries:

**Q1:** SELECT COUNT(year) FROM Police WHERE year < 1976

**Q2:** SELECT Car\_type FROM Police WHERE year > 2000

**Q3:** SELECT COUNT(city) FROM Police WHERE city LIKE "Zurich"

**Q4:** SELECT Owner\_name FROM Police WHERE city LIKE "Ecublens"

$P_1$  issues Q1 once per day and Q2 15 times per day.

$P_2$  issues Q2 5 times, Q3 once, and Q4 10 times per day.

$P_3$  issues Q2 and Q4 10 times per day each.

We would like to vertically fragment the *Police* database to improve the query performance.

- a) Determine the attribute affinity matrix for Q1 to Q4 with the given access patterns.
- b) Use the BEA algorithm to determine the best split of the database into two fragments.

### **Question 3 (\*\* difficult \*\*)**

Given the following relational table *Person* with two sites (*Site1* and *Site2*) and the following queries on the table:

**Person table:**

ID	Name	Age	Country	Sport
1	Ken	23	CH	Golf
2	Alice	36	US	Tennis
3	Peter	31	A	Tennis
4	Dan	27	D	Tennis
5	Kerry	51	CH	Bike
6	Julia	42	CH	Golf

Queries from Site 1:

**Q1:** SELECT \* FROM Person WHERE Age <=40 AND NOT Country=D

**Q2:** SELECT \* FROM Person WHERE NOT Sport=Tennis

**Q3:** SELECT \* FROM Person WHERE Country=D

Queries from Site 2:

**Q4:** SELECT \* FROM Person WHERE Sport=Tennis

**Q5:** SELECT \* FROM Person WHERE Age>40 OR Country=D

**Q6:** SELECT \* FROM Person WHERE Age<=40 AND NOT Sport=Tennis

Given that queries Q3 and Q4 are issued **10** times a week and the horizontal fragmentation algorithm resulted in the following two fragments (F1 and F2, represented with IDs of the member tuples):

F1= {2,3,5,6}, F2={1,4},

Determine the number of times the other queries are issued per week?