Distributed Databases 2018 HW 1

Student ID: **2018280070** Name: **Peter Garamvoelgyi**

**Problem 5.2**

Q1: duration of assignment (ASG.DUR)

– 5 different sites based on ASG.RESP

Q2:

– 2 different sites based on ASG.DUR<20

**Primary horizontal fragmentation:**

Pr = { RESP="Manager", RESP="Consultant", RESP="Engineer",

RESP="Programmer", RESP="Analyst",

DUR<20, DUR>=20}

… i.e. there are 7 predicates, 5 based on RESP, 2 based on DUR. Pr is complete and minimal.

**Minterm predicates:**

|  |  |
| --- | --- |
| m1: RESP="Manager" ∧ DUR<20 | m2: RESP="Manager" ∧ DUR>=20 |
| m3: RESP="Consultant" ∧ DUR<20 | m4: RESP="Consultant" ∧ DUR>=20 |
| m5: RESP="Engineer" ∧ DUR<20 | m6: RESP="Engineer" ∧ DUR>=20 |
| m7: RESP="Programmer" ∧ DUR<20 | m8: RESP="Programmer" ∧ DUR>=20 |
| M9: RESP="Analyst" ∧ DUR<20 | m10: RESP="Analyst" ∧ DUR>=20 |

**The corresponding fragments** (see next page):

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  | | --- | --- | --- | --- | | ASG.m1 |  |  |  | | **ENO** | **PNO** | **RESP** | **DUR** | | E1 | P1 | Manager | 12 | | |  |  |  |  | | --- | --- | --- | --- | | ASG.m2 |  |  |  | | **ENO** | **PNO** | **RESP** | **DUR** | | E5 | P2 | Manager | 24 | | E6 | P4 | Manager | 48 | | E8 | P3 | Manager | 40 | |
| |  |  |  |  | | --- | --- | --- | --- | | ASG.m3 |  |  |  | | **ENO** | **PNO** | **RESP** | **DUR** | | E3 | P3 | Consultant | 10 | | |  |  |  |  | | --- | --- | --- | --- | | ASG.m4 |  |  |  | | **ENO** | **PNO** | **RESP** | **DUR** | |
| |  |  |  |  | | --- | --- | --- | --- | | ASG.m5 |  |  |  | | **ENO** | **PNO** | **RESP** | **DUR** | | |  |  |  |  | | --- | --- | --- | --- | | ASG.m6 |  |  |  | | **ENO** | **PNO** | **RESP** | **DUR** | | E3 | P4 | Engineer | 48 | | E7 | P3 | Engineer | 36 | |
| |  |  |  |  | | --- | --- | --- | --- | | ASG.m7 |  |  |  | | **ENO** | **PNO** | **RESP** | **DUR** | | E4 | P2 | Programmer | 18 | | |  |  |  |  | | --- | --- | --- | --- | | ASG.m8 |  |  |  | | **ENO** | **PNO** | **RESP** | **DUR** | |
| |  |  |  |  | | --- | --- | --- | --- | | ASG.m9 |  |  |  | | **ENO** | **PNO** | **RESP** | **DUR** | | E2 | P2 | Analyst | 6 | | |  |  |  |  | | --- | --- | --- | --- | | ASG.m9 |  |  |  | | **ENO** | **PNO** | **RESP** | **DUR** | | E2 | P1 | Analyst | 24 | |

**Problem 5.8**

First, let us construct the use matrix:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **(use)** | **AEMP.ENO** | **AEMP.ENAME** | **AEMP.TITLE** | **AASG.PNO** | **AASG.RESP** | **AASG.DUR** |
| q1 | 1 | 1 | 1 | 1 | 1 | 1 |
| q2 | 1 | 0 | 0 | 0 | 0 | 1 |

… and the frequency matrix:

|  |  |  |  |
| --- | --- | --- | --- |
| **(freq)** | **S1** | **S2** | **S3** |
| q1 | 10 | 20 | 0 |
| q2 | 0 | 20 | 10 |

Using these, we can calculate the affinity matrix:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **(aff)** | **AEMP.ENO** | **AEMP.ENAME** | **AEMP.TITLE** | **AASG.PNO** | **AASG.RESP** | **AASG.DUR** |
| **AEMP.ENO** | 60 | 30 | 30 | 30 | 30 | 60 |
| **AEMP.ENAME** | 30 | 30 | 30 | 30 | 30 | 30 |
| **AASG.PNO** | 30 | 30 | 30 | 30 | 30 | 30 |
| **AASG.RESP** | 30 | 30 | 30 | 30 | 30 | 30 |
| **AASG.DUR** | 60 | 30 | 30 | 30 | 30 | 60 |

… and the transformed affinity matrix:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **(aff')** | **AEMP.ENO** | **AASG.DUR** | **AEMP.ENAME** | **AEMP.TITLE** | **AASG.PNO** | **AASG.RESP** |
| **AEMP.ENO** | 60 | 60 | 30 | 30 | 30 | 30 |
| **AASG.DUR** | 60 | 60 | 30 | 30 | 30 | 30 |
| **AEMP.ENAME** | 30 | 30 | 30 | 30 | 30 | 30 |
| **AASG.PNO** | 30 | 30 | 30 | 30 | 30 | 30 |
| **AASG.RESP** | 30 | 30 | 30 | 30 | 30 | 30 |

From this we can see that it’s best to have a vertical split where

F1 ~ { EMP.ENO, ASG.DUR }

F2 ~ { EMP.ENAME, EMP.TITLE, ASG.PNO, ASG.RESP }

The resulting fragments on ASG are:

ASGPROJ1 ~ { ENO, PNO, DUR }

ASGPROJ2 ~ { ENO, PNO, RESP }

*(Note that we had to add PNO and ENO, respectively, as these two together form the key.)*

**Problem 5.17**

First, let us perform horizontal fragmentation on ASG and EMP using

PrASG = { DUR=24, DUR≠24 }

PrEMP = { TITLE="Programmer", TITLE≠"Programmer" }

The resulting two fragments are:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  | | --- | --- | --- | --- | | ASG.f1 |  |  |  | | **ENO** | **PNO** | **RESP** | **DUR** | | E2 | P1 | Analyst | 24 | | E5 | P2 | Manager | 24 | | |  |  |  |  | | --- | --- | --- | --- | | ASG.f2 |  |  |  | | **ENO** | **PNO** | **RESP** | **DUR** | | E1 | P1 | Manager | 12 | | E2 | P2 | Analyst | 6 | | E3 | P3 | Consultant | 10 | | E3 | P4 | Engineer | 48 | | E4 | P2 | Programmer | 18 | | E6 | P4 | Manager | 48 | | E7 | P3 | Engineer | 36 | | E8 | P3 | Manager | 40 | |

… and the corresponding derived fragments on EMP:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | EMP.f1 |  |  | | **ENO** | **ENAME** | **TITLE** | | E4 | J. Miller | Programmer | | |  |  |  | | --- | --- | --- | | EMP.f2 |  |  | | **ENO** | **ENAME** | **TITLE** | | E1 | J. Doe | Elect. Eng | | E2 | M. Smith | Syst. Anal | | E3 | A. Lee | Mech. Eng. | | E5 | B. Casey | Syst. Anal | | E6 | L. Chu | Elect. Eng | | E7 | R. Davis | Mech. Eng. | | E8 | J. Jones | Syst. Anal | |

q2 only accesses the attributes ENO and DUR from ASG. Thus, we can use vertical fragmentation on ASG. (This insight is also supported by the affinity matrix shown previously.)

(see next page…)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | ASG.f11 |  |  | | **ENO** | **PNO** | **DUR** | | E2 | P1 | 24 | | E5 | P2 | 24 | | |  |  |  | | --- | --- | --- | | ASG.f12 |  |  | | **ENO** | **PNO** | **RESP** | | E2 | P1 | Analyst | | E5 | P2 | Manager | |
| |  |  |  | | --- | --- | --- | | ASG.f21 |  |  | | **ENO** | **PNO** | **DUR** | | E1 | P1 | 12 | | E2 | P2 | 6 | | E3 | P3 | 10 | | E3 | P4 | 48 | | E4 | P2 | 18 | | E6 | P4 | 48 | | E7 | P3 | 36 | | E8 | P3 | 40 | | |  |  |  | | --- | --- | --- | | ASG.f22 |  |  | | **ENO** | **PNO** | **RESP** | | E1 | P1 | Manager | | E2 | P2 | Analyst | | E3 | P3 | Consultant | | E3 | P4 | Engineer | | E4 | P2 | Programmer | | E6 | P4 | Manager | | E7 | P3 | Engineer | | E8 | P3 | Manager | |

q2 only accesses ASG.f11 and ASG.f21. As site3 only handles this kind of query, it is sufficient to only store these two fragments on this site.

q1 only accesses ASG.f11, ASG.f12 and EMP.f1. As site1 only handles this kind of query, it is sufficient to only store these two fragments on this site.

EMP.f2 and ASG.f22 are not accessed by any of these queries. From what we know, we can store them anywhere, they will not contribute to our cost estimation.

Now, the question is: What should we store on site2?

* As access frequencies of site2 are higher (20), let us start from the assumption that we store everything on this site. (ASG.f11, ASG.f12, ASG.f21, EMP.f1)
* Data transfer is reasonably fast between site2 and site3, but it’s slower between site1 and site2. Thus, let us store a replica of ASG.f11 on site1. (This way, we only need to keep track of two replicas of this fragment.)
* Furthermore, let us also store EMP.f1 and ASG.f12 on site1 and ASG.f21 on site2. This way we have 2-replication for all fragments considered.

