



RENAN HENRIQUE GOMES DAMAZIO ASSUNÇÃO RA 21038114
ALEX ARANTES GONÇALVES RA 21011214

RELATÓRIO 5

Tópicos Emergentes em Bancos de Dados

QUESTÃO 1

A partir de inserção do RA foi gerada a seguinte Matriz de Incidência:

Transação	Atributos					Estatísticas		
	Att1	Att2	Att3	Att4	Att5	Número de Tuplas aproximado (A)	Frequência de Uso por período (B)	Custo (A * B)
T1	1	0	1	1	1	336	25	8400
T2	0	1	0	1	0	496	50	24800
T3	0	1	0	1	0	448	30	13440
T4	1	1	0	1	0	448	20	8960
T5	1	0	1	1	0	512	20	10240
T6	1	1	0	1	0	496	20	9920
T7	1	0	1	1	0	448	20	8960

Tabela 1 – Matriz de Incidência

A partir da matriz de incidência foi calculada a matriz de afinidade para os atributos

Atributo	Att1	Att2	Att3	Att4	Att5
Att1	46480	18880	27600	46480	8400
Att2	18880	57120	0	57120	0
Att3	27600	0	27600	27600	8400
Att4	46480	57120	27600	84720	8400
Att5	8400	0	8400	8400	8400

Tabela 2 – Matriz de Afinidade

Utilizando como métrica o somatório das linhas e colunas, a matriz de afinidade foi clusterizada em 3, gerando a seguinte matriz:

Atributo	Att4	Att1	Att2	Att3	Att5
Att4	84720	46480	57120	27600	8400
Att1	46480	46480	18880	27600	8400
Att2	57120	18880	57120	0	0
Att3	27600	27600	0	27600	8400
Att5	8400	8400	0	8400	8400

Tabela 3 – Matriz de afinidade Após Clustering

Para composição do Cenário Alternativo 1 foi considerada a divisão da Relação inicial em duas relações, conforme a tabela 4:

Rel 1	PK	Att4	Att1	Att2
Rel 2	PK	Att3	Att5	

Tabela 4 – Cenário alternativo 1

Em seguida foi calculado o total de bytes irrelevantes para o Cenário Atual, exposto na tabela 5:

Tran	Att1 (50b)	Att2 (40b)	Att3 (35b)	Att4 (200b)	Att5 (90b)	Bytes Irrelevantes (A)	Número de Tuplas aproximado (B)	Frequência de Uso por período (C)	Total Bytes Irrelevantes (A*B*C)
T1	0	40	0	0	0	40	336	25	336000
T2	50	0	35	0	90	175	496	50	4340000
T3	50	0	35	0	90	175	448	30	2352000
T4	0	0	35	0	90	125	448	20	1120000
T5	0	40	0	0	90	130	512	20	1331200
T6	0	0	35	0	90	125	496	20	1240000
T7	0	40	0	0	90	130	448	20	1164800
									11884000

Tabela 5 – Total Bytes Irrelevantes Cenário Atual

Da mesma forma foi calculado o total de bytes irrelevantes para o Cenário Alternativo 1:

Tran.	Att1 (50b)	Att2 (40b)	Att3 (35b)	Att4 (200b)	Att5 (90b)	Bytes Irrelevantes (A)	Número de Tuplas aproximado (B)	Frequência de Uso por período (C)	Total Bytes Irrelevantes (A*B*C)
T1	0	40	0	0	0	40	336	25	336000
T2	50	0	0	0	0	50	496	50	1240000
T3	50	0	0	0	0	50	448	30	672000
T4	0	0	0	0	0	0	448	20	0
T5	0	40	0	0	0	40	512	20	409600
T6	0	0	0	0	0	0	496	20	0
T7	0	40	0	0	0	40	448	20	358400
									3016000

Tabela 6 – Total bytes Irrelevantes Cenário Alternativo 1

- Cálculos de Blocagem

R1 e R2 possuem 400.000 tuplas cada

Para R1 temos:

A tupla apresenta 290 bytes (em média)

O fator de blocagem é 14 ((4.096-24) / 290)

Total de pages requeridas é 28.572 ((400.000)/14)

Espaço aproximado de: 1429 * 8192 = 106 Mbytes

Para R2 temos:

A tupla apresenta 125 bytes (em média)

O fator de blocagem é 32 $((4096-24)/125)$

Total de pages requeridas é 12.500 $((400.000)/32)$

Espaço aproximado de: $12.500 * 4.096 = 51,2$ Mbytes

- Cálculos de Join

Considerando o índice nas primary key e foreign key de R1 e R2

Considerando o join entre todas as tuplas:

Número de pages lidas: $12.500 + 28.572 = 41.072$

Custo: $41.072 * 4.096 = 168,2$ Mbytes

Considerando o join entre duas tuplas:

Número de pages lidas: $1 + 1 = 2$

Custo: $2 * 4.096 = 8.192$ bytes ou 8,2 Kbytes

T						Bytes Irrel. (A)	N de Tuplas aprox. (B)	Blocos	Freq de Uso por período (C)	Total Bytes Irrel. (A*B*C)	Custo do Join (D * 4096)
	Att1 50b	Att2 40b	Att3 35b	Att4 200b	Att5 90b			Join Rel2 Melhor Caso (D)			
T1	0	40	0	0	0	40	336	10,5	25	336000	43008
T2	50	0	0	0	0	50	496	0	50	1240000	0
T3	50	0	0	0	0	50	448	0	30	672000	0
T4	0	0	0	0	0	0	448	0	20	0	0
T5	0	40	0	0	0	40	512	16	20	409600	65536
T6	0	0	0	0	0	0	496	0	20	0	0
T7	0	40	0	0	0	40	448	14	20	358400	57344
										3016000	165888
										Total	3181888

Tabela 7 – Total bytes Irrelevantes Cenário Alternativo 1 + Custo Join

- Conclusões:

- I. Comparando o número de bytes perdidos no Cenário Atual (11.884.000) com o custo do Cenário Alternativo 1 (3.181.888) fica claro que para este caso é aconselhável a utilização do particionamento vertical;
- II. Ressalta-se que o custo do join para o Cenário Alternativo 1 baseia-se no caso mais otimista onde todas as tuplas encontram-se salvas de forma consecutiva nos blocos.

QUESTÃO 2

Os atributos nrosequencia e codemp foram retirados da análise por serem parte da PK composta da relação depte e logo devem estar presentes em ambas as relações

A partir de inserção das transações foi gerada a seguinte Matriz de Incidência:

Trans.	Att3	Att4	Att5	Att6	Att7	Att8	Att9	Número de Tuplas aproximado (A)	Frequência de Uso por período (B)	Custo (A * B)
T1	1	0	0	0	0	0	0	1000	490	490000
T2	1	1	1	1	1	1	1	1	8	8
T3	1	1	0	0	0	0	0	1	125	125
T4	0	0	0	0	0	0	0	1	2	2
T5	0	0	0	0	1	0	0	1	89	89
T6	0	1	0	1	0	0	0	6389	347	2216983

Tabela 1 – Matriz de Incidência

A partir da matriz de incidência foi calculada a matriz de afinidade para os atributos

Atributo	Att3	Att4	Att5	Att6	Att7	Att8	Att9
Att3	490133	133	8	8	8	8	8
Att4	133	2217116	8	2216991	8	8	8
Att5	8	8	8	8	8	8	8
Att6	8	2216991	8	2216991	8	8	8
Att7	8	8	8	8	97	8	8
Att8	8	8	8	8	8	8	8
Att9	8	8	8	8	8	8	8

Tabela 2 – Matriz de Afinidade

Utilizando como métrica o somatório das linhas e colunas, a matriz de afinidade foi clusterizada em 2, gerando a seguinte matriz:

Atributo	Att3	Att4	Att6	Att5	Att7	Att8	Att9
Att3	490133	133	8	8	8	8	8
Att4	133	2217116	2216991	8	8	8	8
Att6	8	2216991	2216991	8	8	8	8
Att5	8	8	8	8	8	8	8
Att7	8	8	8	8	97	8	8
Att8	8	8	8	8	8	8	8
Att9	8	8	8	8	8	8	8

Tabela 3 – Matriz de afinidade Após Clustering

Tópicos Emergentes em Bancos de Dados

Para composição do Cenário Alternativo 1 foi considerada a divisão da Relação inicial em duas relações, conforme a tabela 4:

Rel 1	PK1	PK2	Att3	Att4	Att6	
Rel 2	PK1	PK2	Att5	Att7	Att8	Att9

Tabela 4 – Cenário alternativo 1

Em seguida foi calculado o total de bytes irrelevantes para o Cenário Atual, exposto na tabela 5:

Tran	Att3 100b	Att4 25b	Att5 8b	Att6 1b	Att7 30b	Att8 30b	Att9 30b	Bytes Irrelevantes (A)	Número de Tuplas aproximado (B)	Frequência de Uso por período (C)	Total Bytes Irrelevantes (A*B*C)
T1	0	25	8	1	30	30	30	124	1000	490	60760000
T2	0	0	0	0	0	0	0	0	1	8	0
T3	0	0	8	1	30	30	30	99	1	125	12375
T4	100	25	8	1	30	30	30	224	1	2	448
T5	100	25	8	1	0	30	30	194	1	89	17266
T6	100	0	8	0	0	30	30	168	6389	347	372453144
											433243233

Tabela 5 – Total Bytes Irrelevantes Cenário Atual

Da mesma forma foi calculado o total de bytes irrelevantes para o Cenário Alternativo 1:

Tran	Att3 100b	Att4 25b	Att5 8b	Att6 1b	Att7 30b	Att8 30b	Att9 30b	Bytes Irrelevantes (A)	Número de Tuplas aproximado (B)	Frequência de Uso por período (C)	Total Bytes Irrelevantes (A*B*C)
T1	0	25	0	1	0	0	0	26	1000	490	12740000
T2	0	0	0	0	0	0	0	0	1	8	0
T3	0	0	0	1	0	0	0	1	1	125	125
T4	100	25	8	1	30	30	30	224	1	2	448
T5	0	0	8	0	0	30	30	68	1	89	6052
T6	100	0	0	0	0	0	0	100	6389	347	221698300
											234444925

Tabela 6 – Total bytes Irrelevantes Cenário Alternativo 1

• Cálculos de Blocagem

R1 e R2 possuem 28.000 tuplas cada

Para R1 temos:

A tupla apresenta 126 bytes (em média)

O fator de blocagem é 68 $((8.192-24) / 126)$

Total de pages requeridas é 412 $((28.000)/68)$

Espaço aproximado de: $1429 * 8192 = 3,4$ Mbytes

Para R2 temos:

A tupla apresenta 98 bytes (em média)

O fator de blocagem é 83 $((8.192-24)/98)$

Total de pages requeridas é 374 $((28.000)/ 32)$

Espaço aproximado de: $374 * 8192 = 3,0$ Mbytes

- Cálculos de Join

Considerando o índice nas primary key e foreign key de R1 e R2

Considerando o join entre todas as tuplas:

Número de pages lidas: $412 + 374 = 786$

Custo: $786 * 4.096 = 6,4$ Mbytes

Considerando o join entre duas tuplas:

Número de pages lidas: $1 + 1 = 2$

Custo: $2 * 8.192 = 16.384$ bytes ou 16,3 Kbytes

Tran	Att3 100b	Att 4 25b	Att 5 8b	Att 6 1b	Att 7 30b	Att 8 30b	Att 9 30b	Byte s Irrel. (A)	Nro Tupla s aprox. (B)	Freq. Uso por períod o (C)	Bloco s Join (D)	Total Bytes Irrel. (A*B*C)	Custo Join Bytes D * 8192
T1	0	25	0	1	0	0	0	26	1000	490	0	12740000	0
T2	0	0	0	0	0	0	0	0	1	8	8	0	65536
T3	0	0	0	1	0	0	0	1	1	125	0	125	0
T4	100	25	8	1	30	30	30	224	1	2	2	448	16384
T5	0	0	8	0	0	30	30	68	1	89	0	6052	0
T6	100	0	0	0	0	0	0	100	6389	347	26719	221698300	218882048
												234444925	218963968
												Total	453.408.893

Tabela 7 – Total bytes Irrelevantes Cenário Alternativo 1 + Custo Join

- Conclusões:

- Comparando o número de bytes perdidos no Cenário Atual (**433.243.233**) com o custo do Cenário Alternativo 1 (**453.408.893**) fica claro que, por muito pouco, para este caso não é aconselhável a utilização do particionamento vertical;
- O fato de T6 não utilizar o Att3 foi o que mais pesou para o custo do join, visto que esse atributo é o maior, com 100 bytes, a transição retorna 6389 tuplas e é utilizada com alta frequência, 347 vezes.

- III. Ressalta-se que o custo do join para o Cenário Alternativo 1 baseia-se no caso mais otimista onde todas as tuplas encontram-se salvas de forma consecutiva nos blocos.