Otimização de Interrogações

Trabalho Prático nº 1

Mestrado Integrado em Engenharia Informática e Computação

Tecnologias de Bases de Dados

Elementos do grupo:
André Pires - 201207106 - ei12058@fe.up.pt
João Bandeira - 201200615 - ei12022@fe.up.pt

Faculdade de Engenharia da Universidade do Porto
Rua Roberto Frias, sn, 4200-465 Porto, Portugal

29 de Março de 2016

Conteúdo

1 Objetivo do trabalho

Analisar os planos de execução de diferentes interrogações SQL a uma BD de teste. Avaliar o impacto da existência de índices e de estatísticas e do recurso a diferentes estratégias de estruturação das perguntas. O relatório a produzir deve, para cada situação, indicar a formulação da pergunta em SQL, a resposta obtida, as condições de execução do teste incluindo duração se relevante, o plano de execução e as conclusões a que chegou.[1]

2 Criação de índices

Para as tabelas "z", foram criados três índices cujo objetivo foi permitir uma maior eficiência de certas interrogações. Os índices criados foram:

- Btree na coluna "disciplina_id"da tabela "zrespostas"

  Foi criado um índice para a coluna relativa à disciplina, pois várias interrogações requerem agregação com a tabela das disciplinas, nesta mesma coluna.

- *Bitmap* na coluna "semestre_id"da tabela "zrespostas"

  Foi criado um índice para a coluna relativa ao semestre, pois várias interrogações esperam resultados de um semestre específico. Foi escolhido um índice *bitmap* devido às vantagens que o mesmo traz, e ao facto da cardinalidade da coluna semestre não ser muito elevada (existem respostas relativas a apenas alguns semestres).

- Btree nas colunas "semestre_id"e "disciplina_id"da tabela "zrespostas"

  Este último índice foi criado para tentar melhorar a eficiência das interrogações que envolvem a agregação de resultados por disciplina, ao mesmo tempo que se filtra um semestre específico.

Para criação dos índices referidos foi utilizado o código SQL abaixo.

```
CREATE INDEX Z_RESPOSTAS_DISCIPLINA_IDX ON zrespostas (disciplina_id);

CREATE INDEX Z_RESPOSTAS_SEMESTRE_IDX ON zrespostas (semestre_id);

CREATE INDEX Z_RESP_DISC_SEMESTRE_IDX ON zrespostas (semestre_id,
    disciplina_id);
```

3 Metodologia

Para cada uma das perguntas propostas no enunciado, o código SQL escrito foi corrido três vezes, sendo que o tempo de execução de cada uma delas foi registado. Sendo assim, os tempos apresentados abaixo correspondem a uma média dos tempos de cada conjunto de três execuções. Relativamente a estes tempos, apesar de serem apresentados para todas as perguntas, por vezes não serão muito significativos, uma vez que após a primeira execução de cada interrogação, frequentemente o servidor retornava resultados num tempo muito inferior, pois utiliza muito provavelmente informação em *cache*.

Foram ainda registados e apresentados abaixo os planos de execução e resultados de todas as interrogações.

4 Respostas às perguntas

4.1 Pergunta 1 - Seleção

4.1.1 Formulação SQL

```
-- a
select count(*)
from xrespostas
where semestre_id = 21;

-- b
select count(*)
from xrespostas
Where disciplina_id = 1237;

-- c
select count(*)
from xrespostas, xsemestre
where xrespostas.semestre_id = xsemestre.semestre_id
and xsemestre.ano_lectivo = '2008/2009'
and xsemestre.semestre = '1S';
```

4.1.2 Tempos

Alínea	X (ms)	Y (ms)	Z (ms)
a	130.67	130.33	29.67
b	132	132.67	40.67
c	309	285.33	42.67

Tabela 1: Resultados pergunta 1

4.1.3 Plano de execução

(x)

OPERATION	OBJECT_NAME	CARDINALITY	COST
SELECT STATEMENT		1	4392
SORT (AGGREGATE)		1	
TABLE ACCESS (FULL)	XRESPOSTAS	338964	4392
Filter Predicates			
SEMESTRE_ID=21			

(y)

OPERATION	OBJECT_NAME	CARDINALITY	COST
SELECT STATEMENT		1	4392
SORT (AGGREGATE)		1	
TABLE ACCESS (FULL)	YRESPOSTAS	371106	4392
Filter Predicates			
SEMESTRE_ID=21			

(z)

OPERATION	OBJECT_NAME	CARDINALITY	COST
SELECT STATEMENT		1	9
SORT (AGGREGATE)		1	
BITMAP CONVERSION (COUNT)		359679	9
BITMAP INDEX (SINGLE VALUE)	Z_RESPOSTAS_SEMES...		
Access Predicates			
SEMESTRE_ID=21			

Figura 1: Plano de execução da alínea 1.a)

(x)

OPERATION	OBJECT_NAME	CARDINALITY	COST
SELECT STATEMENT		1	4397
SORT (AGGREGATE)		1	
TABLE ACCESS (FULL)	XRESPOSTAS	2069	4397
Filter Predicates			
DISCIPLINA_ID=1237			

(y)

OPERATION	OBJECT_NAME	CARDINALITY	COST
SELECT STATEMENT		1	4397
SORT (AGGREGATE)		1	
TABLE ACCESS (FULL)	YRESPOSTAS	2034	4397
Filter Predicates			
DISCIPLINA_ID=1237			

(z)

OPERATION	OBJECT_NAME	CARDINALITY	COST
SELECT STATEMENT		1	7
SORT (AGGREGATE)		1	
INDEX (RANGE SCAN)	Z_RESPOSTAS_DISCI...	2140	7
Access Predicates			
DISCIPLINA_ID=1237			

Figura 2: Plano de execução da alínea 1.b)

(x)

OPERATION	OBJECT_NAME	CARDINALITY	COST
SELECT STATEMENT		1	4402
SORT (AGGREGATE)		1	
HASH JOIN		147180	4402
Access Predicates			
XRESPOSTAS.SEMESTRE_ID=XSEMESTRE.SEMESTRE_ID			
TABLE ACCESS (FULL)	XSEMESTRE	1	3
Filter Predicates			
AND			
XSEMESTRE.ANO_LECTIVO='2008/2009'			
XSEMESTRE.SEMESTRE='1S'			
TABLE ACCESS (FULL)	XRESPOSTAS	2747365	4385

(y)

OPERATION	OBJECT_NAME	CARDINALITY	COST
SELECT STATEMENT		1	4402
SORT (AGGREGATE)		1	
HASH JOIN		228947	4402
Access Predicates			
YRESPOSTAS.SEMESTRE_ID=YSEMESTRE.SEMESTRE_ID			
TABLE ACCESS (FULL)	YSEMESTRE	1	3
Filter Predicates			
AND			
YSEMESTRE.ANO_LECTIVO='2008/2009'			
YSEMESTRE.SEMESTRE='1S'			
TABLE ACCESS (FULL)	YRESPOSTAS	2747365	4385

(z)

OPERATION	OBJECT_NAME	CARDINALITY	COST
SELECT STATEMENT		1	7
SORT (AGGREGATE)		1	
NESTED LOOPS		228947	7
TABLE ACCESS (FULL)	ZSEMESTRE	1	3
Filter Predicates			
AND			
ZSEMESTRE.ANO_LECTIVO='2008/2009'			
ZSEMESTRE.SEMESTRE='1S'			
BITMAP CONVERSION (COUNT)		196240	7
BITMAP INDEX (SINGLE VALUE)	Z_RESPOSTAS_SEMES...		
Access Predicates			
ZRESPOSTAS.SEMESTRE_ID=ZSEMESTRE.SEMESTF			

Figura 3: Plano de execução da alínea 1.c)

4.1.4 Resultados

	a	b	c
Count(*)	351671	6525	351671

Tabela 2: Resultados pergunta 1

4.1.5 Conclusões

Neste caso, nota-se claramente um resultado favorável, em tempo e em custo de execução, com a adição dos índices extra (tabelas *z*). No entanto, os índices "normais"(como *primary* e *foreign keys*) não têm qualquer influência na seleção.

4.2 Pergunta 2 - Agregação

4.2.1 Formulação SQL

```
-- a
select avg(xrespostas.resposta)
from xrespostas, xdisciplina
where xrespostas.disciplina_id = xdisciplina.disciplina_id
and xdisciplina.sigla = 'FPRO';

-- b
select * from (
```

```
  select xdisciplina.disciplina_id, xdisciplina.sigla,
      avg(xrespostas.resposta) nota, count(*) nr
  from xrespostas, xdisciplina
  where xrespostas.semestre_id = 21
  and xrespostas.disciplina_id = xdisciplina.disciplina_id
  group by xdisciplina.disciplina_id, xdisciplina.sigla
  having count(*) > 300
  order by nota desc
)
where rownum <= 1;

-- c
CREATE VIEW x_media_mais_300_semestre_21 AS(
  SELECT disciplina_id, avg(resposta) nota_glob, COUNT(*) nr_respostas
  FROM xrespostas
  WHERE semestre_id = 21
  GROUP BY disciplina_id
  HAVING COUNT(*) > 300
);

CREATE VIEW x_med_pergunta_apreciacao_glob AS(
  SELECT xr.disciplina_id, avg(xr.resposta) nota_perg, COUNT(*)
      nr_respostas
  FROM xrespostas xr, xpergunta xp
  WHERE xr.semestre_id = 21
  AND xr.pergunta_id = xp.pergunta_id
  AND xp.nome LIKE 'Apreciacao Global'
  GROUP BY xr.disciplina_id
);

SELECT d.disciplina_id, d.sigla, med_glob.nota_glob, med_perg.nota_perg
FROM x_media_mais_300_semestre_21 med_glob,
    x_med_pergunta_apreciacao_glob med_perg, xdisciplina d
WHERE med_glob.disciplina_id = med_perg.disciplina_id
AND med_glob.disciplina_id = d.disciplina_id
AND nota_glob + 0.1 < nota_perg;
```

4.2.2 Tempos

Alínea	X (ms)	Y (ms)	Z (ms)
a	299	306	18.33
b	296	262.67	256.67
c	520.67	368.33	318

Tabela 3: Resultados pergunta 2

4.2.3 Plano de execução

(x)

OPERATION	OBJECT_NAME	CARDINALITY	COST
SELECT STATEMENT		1	4430
SORT (AGGREGATE)		1	
HASH JOIN		3151	4430
Access Predicates			
XRESPOSTAS.DISCIPLINA_ID=XDISCIPLINA.DISCIPLINA_II			
TABLE ACCESS (FULL)	XDISCIPLINA	1	10
Filter Predicates			
XDISCIPLINA.SIGLA='FPRO'			
TABLE ACCESS (FULL)	XRESPOSTAS	2747365	4406

(y)

OPERATION	OBJECT_NAME	CARDINALITY	COST
SELECT STATEMENT		1	4430
SORT (AGGREGATE)		1	
HASH JOIN		3151	4430
Access Predicates			
YRESPOSTAS.DISCIPLINA_ID=YDISCIPLINA.DISCIPLINA_II			
TABLE ACCESS (FULL)	YDISCIPLINA	1	10
Filter Predicates			
YDISCIPLINA.SIGLA='FPRO'			
TABLE ACCESS (FULL)	YRESPOSTAS	2747365	4406

(z)

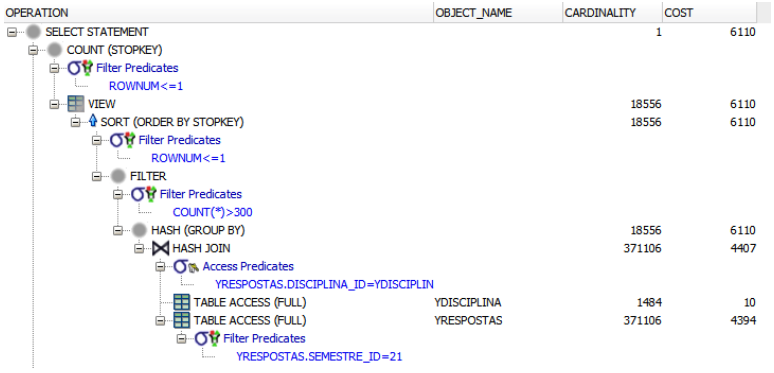
OPERATION	OBJECT_NAME	CARDINALITY	COST
SELECT STATEMENT		1	180
SORT (AGGREGATE)		1	
NESTED LOOPS			
NESTED LOOPS		3151	180
TABLE ACCESS (FULL)	ZDISCIPLINA	1	10
Filter Predicates			
ZDISCIPLINA.SIGLA='FPRO'			
INDEX (RANGE SCAN)	Z_RESPOSTAS_DISCI...	2282	5
Access Predicates			
ZRESPOSTAS.DISCIPLINA_ID=ZDISCIPLINA.DISCI			
TABLE ACCESS (BY INDEX ROWID)	ZRESPOSTAS	2282	170

Figura 4: Plano de execução da alínea 2.a)

(x)

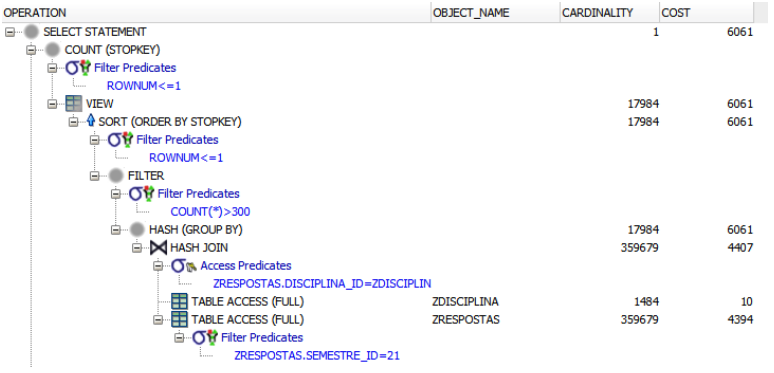
OPERATION	OBJECT_NAME	CARDINALITY	COST
SELECT STATEMENT		1	5963
COUNT (STOPKEY)			
Filter Predicates			
ROWNUM<=1			
VIEW		16949	5963
SORT (ORDER BY STOPKEY)		16949	5963
Filter Predicates			
ROWNUM<=1			
FILTER			
Filter Predicates			
COUNT(*)>300			
HASH (GROUP BY)		16949	5963
HASH JOIN		338964	4406
Access Predicates			
XRESPOSTAS.DISCIPLINA_ID=XDISCIPLIN			
TABLE ACCESS (FULL)	XDISCIPLINA	1484	10
TABLE ACCESS (FULL)	XRESPOSTAS	338964	4394
Filter Predicates			
XRESPOSTAS.SEMESTRE_ID=21			

(y)



OPERATION	OBJECT_NAME	CARDINALITY	COST
SELECT STATEMENT		1	6110
COUNT (STOPKEY)			
Filter Predicates			
ROWNUM<=1			
VIEW		18556	6110
SORT (ORDER BY STOPKEY)		18556	6110
Filter Predicates			
ROWNUM<=1			
FILTER			
Filter Predicates			
COUNT(*)>300			
HASH (GROUP BY)		18556	6110
HASH JOIN		371106	4407
Access Predicates			
YRESPOSTAS.DISCIPLINA_ID=YDISCIPLIN			
TABLE ACCESS (FULL)	YDISCIPLINA	1484	10
TABLE ACCESS (FULL)	YRESPOSTAS	371106	4394
Filter Predicates			
YRESPOSTAS.SEMESTRE_ID=21			

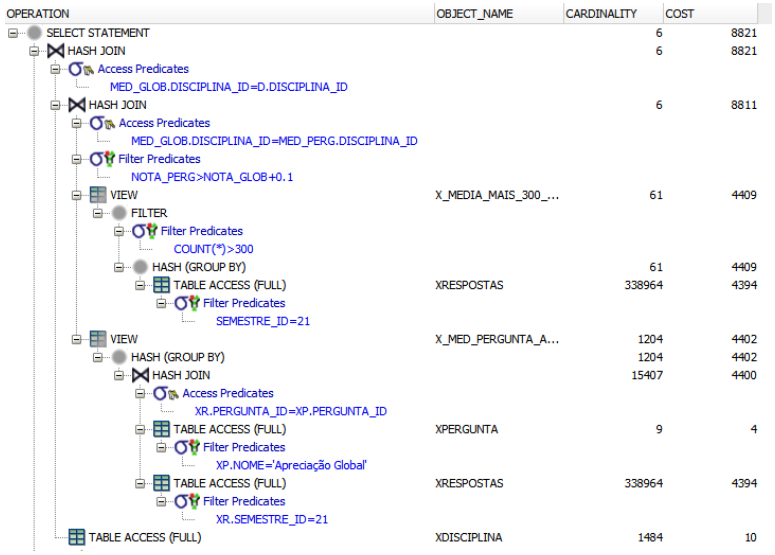
(z)



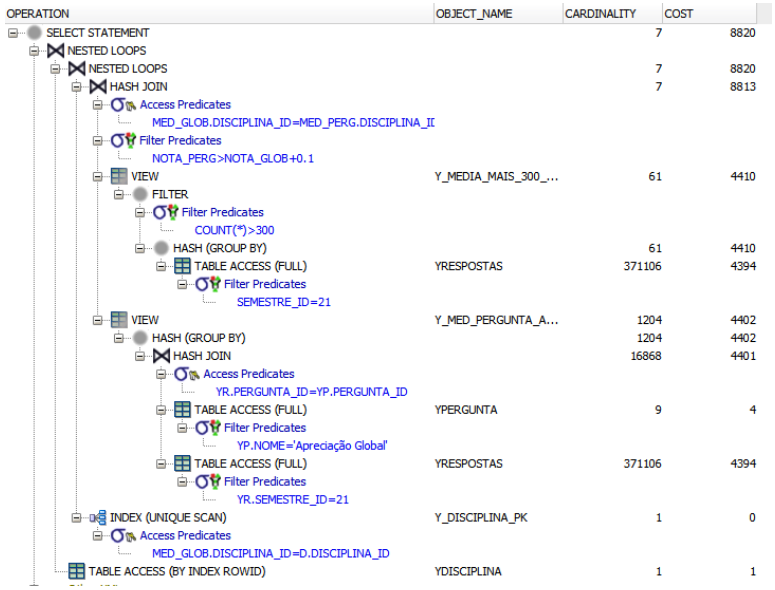
OPERATION	OBJECT_NAME	CARDINALITY	COST
SELECT STATEMENT		1	6061
COUNT (STOPKEY)			
Filter Predicates			
ROWNUM<=1			
VIEW		17984	6061
SORT (ORDER BY STOPKEY)		17984	6061
Filter Predicates			
ROWNUM<=1			
FILTER			
Filter Predicates			
COUNT(*)>300			
HASH (GROUP BY)		17984	6061
HASH JOIN		359679	4407
Access Predicates			
ZRESPOSTAS.DISCIPLINA_ID=ZDISCIPLIN			
TABLE ACCESS (FULL)	ZDISCIPLINA	1484	10
TABLE ACCESS (FULL)	ZRESPOSTAS	359679	4394
Filter Predicates			
ZRESPOSTAS.SEMESTRE_ID=21			

Figura 5: Plano de execução da alínea 2.b)

(x)



(y)



(z)

Figura 6: Plano de execução da alínea 2.c)

4.2.4 Resultados

- **Alínea a:** avg() = 3,979617
- **Alínea b:** disciplina_id = 1949, sigla = SAD, nota = 4,519126
- **Alínea c:**

	DISCIPLINA_ID	SIGLA	NOTA_GLOB	NOTA_PERG
1	1172	MAT1	3,630962343096234309623430962343096234 31	3,7368421052631578947368421052631578947 4
2	1661	TTC	4,056580565805658056580565805658056580 57	4,1666666666666666666666666666666666666 7
3	1470	MICR	4,0863930885529157667386609071274298056 2	4,2
4	1511	QUIM1	3,979899497487437185929648241206030150 75	4,1111111111111111111111111111111111111 1
5	1510	FISI1	3,775773195876288659793814432989690721 65	3,8947368421052631578947368421052631578 9
6	1907	AFHU	4,387096774193548387096774193548387096 77	4,6363636363636363636363636363636363636 4

Figura 7: Resultado 2c)

4.2.5 Conclusões

Relativamente à alínea a), através da análise dos planos de execução pode verificar-se que para os conjuntos de tabelas x e y, as interrogações seguem uma execução semelhante. Apesar das tabelas y terem índices nas colunas que correspondem a chaves primárias, estas não são utilizadas na seleção. Já no terceiro plano de execução, relativo às tabelas z, observa-se que a interrogação tomou partido da existência de um índice na coluna 'disciplina_id', apresentando um custo substancialmente inferior aos das tabelas x e y.

Já quanto à alínea c), é possível verificar diferenças entre todos os planos de execução. Na execução das tabelas y, pode verificar-se que esta tomou partido da existência de um índice na coluna 'disciplina_id' da tabelas de disciplinas (chave primária). No que diz respeito às tabelas z, para além das melhorias que se observou em y em relação a x, a execução da interrogação foi ainda beneficiada pela existência do índice em múltiplas colunas (disciplina_id e semestre_id) na tabela de respostas. uma vez que esta interrogação envolve seleção de tuplos especificando o seu semestre, e agrupando-os por disciplina.

4.3 Pergunta 3 - Negação

4.3.1 Formulação SQL

```sql
-- not in
select *
from xdisciplina
where disciplina_id not in(
  select disciplina_id
  from xrespostas
  group by disciplina_id
);

-- juncao externa e filtragem para nulo
select disc.sigla, disc.nome
from (
  select disciplina_id
  from xrespostas
```

```
  group by disciplina_id
) resp FULL OUTER JOIN xdisciplina disc
ON resp.disciplina_id = disc.disciplina_id
where resp.disciplina_id is null;
```

4.3.2 Tempos

Alínea	X (ms)	Y (ms)	Z (ms)
Baseada em *not in*	1020.67	788	448.67
Baseada em junção externa e filtragem para nulo	931.33	964.33	411

Tabela 4: Resultados pergunta 3, com otimizador = *CHOOSE*

Alínea	X (ms)	Y (ms)	Z (ms)
Baseada em *not in*	-	-	-
Baseada em junção externa e filtragem para nulo	419	419.67	418.33

Tabela 5: Resultados pergunta 3, com otimizador = *RULE*

4.3.3 Plano de execução

(x)

OPERATION	OBJECT_NAME	CARDINALITY	COST
SELECT STATEMENT		280	4528
HASH JOIN (RIGHT ANTI NA)		280	4528
Access Predicates			
DISCIPLINA_ID=DISCIPLINA_ID			
VIEW	VW_NSO_1	1204	4517
HASH (GROUP BY)		1204	4517
TABLE ACCESS (FULL)	XRESPOSTAS	2747365	4390
TABLE ACCESS (FULL)	XDISCIPLINA	1484	10

(y)

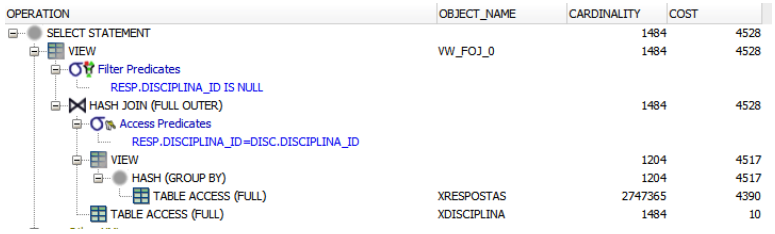
OPERATION	OBJECT_NAME	CARDINALITY	COST
SELECT STATEMENT		280	4528
HASH JOIN (RIGHT ANTI NA)		280	4528
Access Predicates			
DISCIPLINA_ID=DISCIPLINA_ID			
VIEW	VW_NSO_1	1204	4517
HASH (GROUP BY)		1204	4517
TABLE ACCESS (FULL)	YRESPOSTAS	2747365	4390
TABLE ACCESS (FULL)	YDISCIPLINA	1484	10

(z)

OPERATION	OBJECT_NAME	CARDINALITY	COST
SELECT STATEMENT		280	4528
HASH JOIN (RIGHT ANTI NA)		280	4528
Access Predicates			
DISCIPLINA_ID=DISCIPLINA_ID			
VIEW	VW_NSO_1	1204	4517
HASH (GROUP BY)		1204	4517
TABLE ACCESS (FULL)	ZRESPOSTAS	2747365	4390
TABLE ACCESS (FULL)	ZDISCIPLINA	1484	10

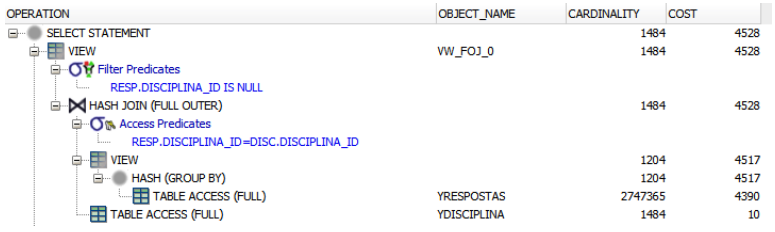
Figura 8: Plano de execução da alínea 3 *not in*, com otimizador *CHOOSE*

(x)



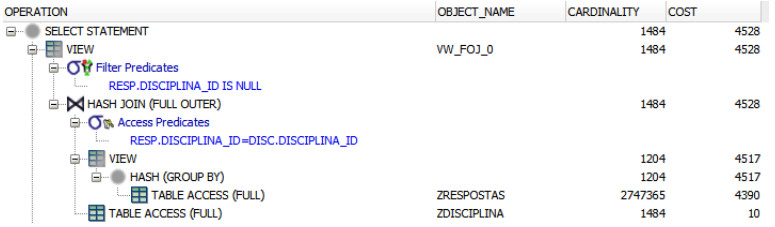
OPERATION	OBJECT_NAME	CARDINALITY	COST
SELECT STATEMENT		1484	4528
VIEW	VW_FOJ_0	1484	4528
Filter Predicates			
RESP.DISCIPLINA_ID IS NULL			
HASH JOIN (FULL OUTER)		1484	4528
Access Predicates			
RESP.DISCIPLINA_ID=DISC.DISCIPLINA_ID			
VIEW		1204	4517
HASH (GROUP BY)		1204	4517
TABLE ACCESS (FULL)	XRESPOSTAS	2747365	4390
TABLE ACCESS (FULL)	XDISCIPLINA	1484	10

(y)



OPERATION	OBJECT_NAME	CARDINALITY	COST
SELECT STATEMENT		1484	4528
VIEW	VW_FOJ_0	1484	4528
Filter Predicates			
RESP.DISCIPLINA_ID IS NULL			
HASH JOIN (FULL OUTER)		1484	4528
Access Predicates			
RESP.DISCIPLINA_ID=DISC.DISCIPLINA_ID			
VIEW		1204	4517
HASH (GROUP BY)		1204	4517
TABLE ACCESS (FULL)	YRESPOSTAS	2747365	4390
TABLE ACCESS (FULL)	YDISCIPLINA	1484	10

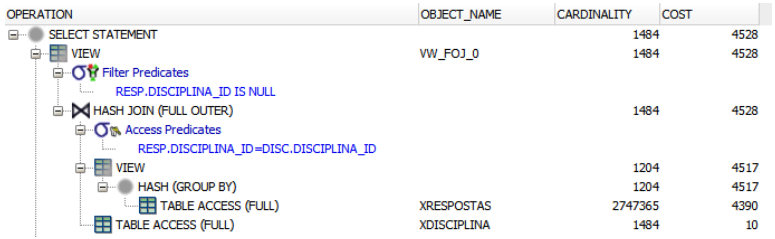
(z)



OPERATION	OBJECT_NAME	CARDINALITY	COST
SELECT STATEMENT		1484	4528
VIEW	VW_FOJ_0	1484	4528
Filter Predicates			
RESP.DISCIPLINA_ID IS NULL			
HASH JOIN (FULL OUTER)		1484	4528
Access Predicates			
RESP.DISCIPLINA_ID=DISC.DISCIPLINA_ID			
VIEW		1204	4517
HASH (GROUP BY)		1204	4517
TABLE ACCESS (FULL)	ZRESPOSTAS	2747365	4390
TABLE ACCESS (FULL)	ZDISCIPLINA	1484	10

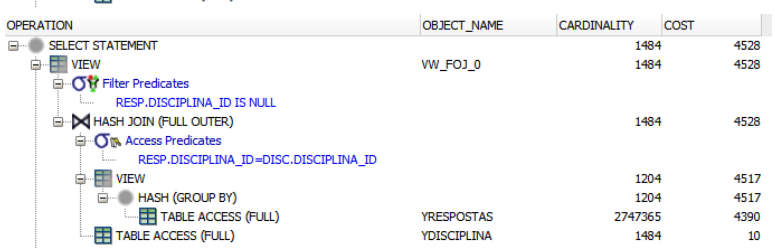
Figura 9: Plano de execução da alínea 3 *junção*, com otimizador *CHOOSE*

(x)



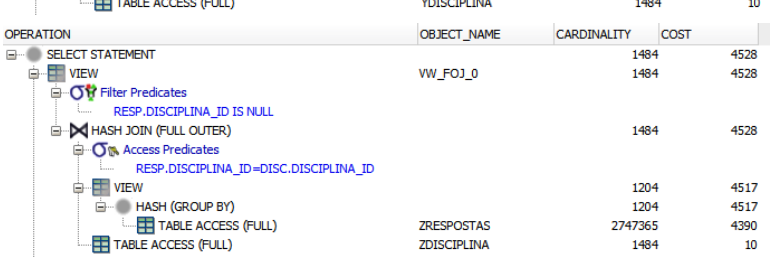
OPERATION	OBJECT_NAME	CARDINALITY	COST
SELECT STATEMENT		1484	4528
VIEW	VW_FOJ_0	1484	4528
Filter Predicates			
RESP.DISCIPLINA_ID IS NULL			
HASH JOIN (FULL OUTER)		1484	4528
Access Predicates			
RESP.DISCIPLINA_ID=DISC.DISCIPLINA_ID			
VIEW		1204	4517
HASH (GROUP BY)		1204	4517
TABLE ACCESS (FULL)	XRESPOSTAS	2747365	4390
TABLE ACCESS (FULL)	XDISCIPLINA	1484	10

(y)



OPERATION	OBJECT_NAME	CARDINALITY	COST
SELECT STATEMENT		1484	4528
VIEW	VW_FOJ_0	1484	4528
Filter Predicates			
RESP.DISCIPLINA_ID IS NULL			
HASH JOIN (FULL OUTER)		1484	4528
Access Predicates			
RESP.DISCIPLINA_ID=DISC.DISCIPLINA_ID			
VIEW		1204	4517
HASH (GROUP BY)		1204	4517
TABLE ACCESS (FULL)	YRESPOSTAS	2747365	4390
TABLE ACCESS (FULL)	YDISCIPLINA	1484	10

(z)



OPERATION	OBJECT_NAME	CARDINALITY	COST
SELECT STATEMENT		1484	4528
VIEW	VW_FOJ_0	1484	4528
Filter Predicates			
RESP.DISCIPLINA_ID IS NULL			
HASH JOIN (FULL OUTER)		1484	4528
Access Predicates			
RESP.DISCIPLINA_ID=DISC.DISCIPLINA_ID			
VIEW		1204	4517
HASH (GROUP BY)		1204	4517
TABLE ACCESS (FULL)	ZRESPOSTAS	2747365	4390
TABLE ACCESS (FULL)	ZDISCIPLINA	1484	10

Figura 10: Plano de execução da alínea 3 *junção*, com otimizador *RULE*

4.3.4 Resultados

	SIGLA	NOME
1	IE	Informação Empresarial
2	ASI	Análise de Sistemas de Informação
3	ETELE	Electrónica das Telecomunicações
4	PSD	Projecto de Sistemas Digitais
5	GA	Geologia Ambiental
6	PSA	Projecto de Sistemas de Automação
7	IMTM	Informática Médica e Telemedicina
8	LAUT	Laboratório de Automação
9	MIAGI	Metodologias de Investigação Aplicadas à Gestão de Informação
10	GI	Gestão de Informação
11	ARI	Armazenamento e Recuperação de Informação
12	SM	Sistemas Multimédia
13	CRC	Computadores e Redes de Comunicação
14	EMBS	Equipamento Médico e Biosensores
15	CR	Corrosão e Revestimentos
16	PSI	Produtos e Serviços de Informação
17	RM	Resistência dos Materiais
18	DRSIG	Detecção Remota e Sistemas de Informação Geográfica
19	BD	Bases de Dados
20	F I	Física I
21	SDO	Sistemas Dinâmicos e Optimização
22	CM	Caracterização de Materiais
23	ME	Métodos Estatísticos

Figura 11: Resultado 3

4.3.5 Conclusões

Relativamente a esta pergunta, começamos por referir que não são apresentados resultados relativamente à execução da técnica "not in"com o otimizador "RULE"pois não se conseguiu que a interrogação executasse em tempo útil (após alguns minutos de espera, não tinham sidos retornados resultados).

Analisando os planos de execução dos restantes casos, começando pelos casos onde foi utilizada a estratégia de junção externa e filtragem para nulo, verificamos que são semelhantes, não tendo a escolha do otimizador feito diferença alguma.

Já quanto à estratégia que utiliza "not in", e relativamente à estratégia anterior, apesar de requerer menos operações durante a execução, resulta num custo igual.

4.4 Pergunta 4 - Quantificação universal

4.4.1 Formulação SQL

```
-- contagem
create view x_total_respostas as (
```

```
  SELECT disciplina_id , count(*) total_respostas
  FROM (
    SELECT disciplina_id, pergunta_id, count(*) c
    FROM xrespostas
    WHERE semestre_id = 21
    GROUP BY disciplina_id, pergunta_id
  )
  GROUP BY disciplina_id
);

create view x_respostas_5 as (
  SELECT disciplina_id , count(*) respostas_5
  FROM (
    SELECT disciplina_id, pergunta_id, count(*)
    FROM xrespostas
    WHERE semestre_id = 21
    AND resposta = 5
    GROUP BY DISCIPLINA_ID, pergunta_id
  )
  GROUP BY disciplina_id
);

SELECT d.disciplina_id, d.sigla, t.total_respostas, r.respostas_5
FROM x_total_respostas t, x_respostas_5 r, xdisciplina d
WHERE t.disciplina_id = r.disciplina_id
AND t.disciplina_id = d.disciplina_id
AND t.total_respostas = r.respostas_5;

-- dupla negacao
create view x_disciplinas_inq_21 as (
  SELECT disciplina_id
  FROM xrespostas
  WHERE semestre_id = 21
  GROUP BY disciplina_id
);

SELECT db.disciplina_id, db.sigla
FROM x_disciplinas_inq_21 da, xdisciplina db
WHERE da.disciplina_id = db.disciplina_id
AND da.disciplina_id NOT IN (
SELECT disciplina_id
FROM xrespostas ra
WHERE semestre_id = 21
AND NOT EXISTS (
  SELECT rb.disciplina_id, rb.pergunta_id
  FROM xrespostas rb
  WHERE rb.disciplina_id = ra.disciplina_id
  AND rb.pergunta_id = ra.pergunta_id
  AND rb.semestre_id = 21
  AND rb.resposta = 5
  GROUP BY disciplina_id, pergunta_id
)
GROUP BY disciplina_id, pergunta_id
);
```

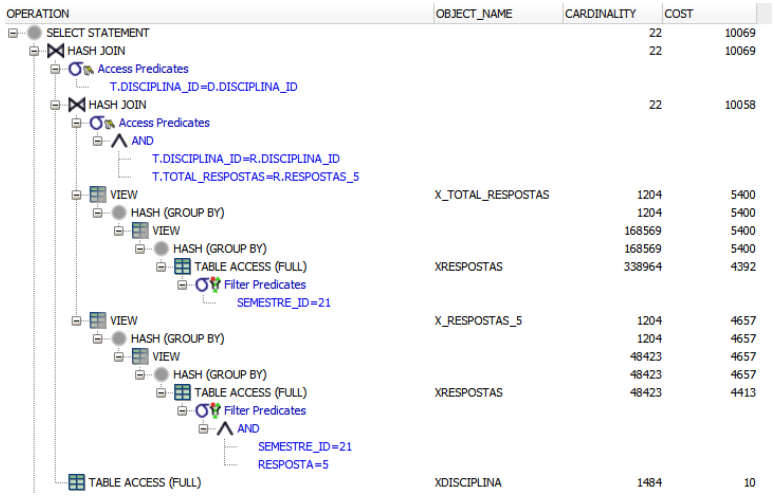
4.4.2 Tempos

Alínea	**X (ms)**	**Y (ms)**	**Z (ms)**
Estratégia de dupla negação	336.67	365.67	294
Estratégia de contagem	496	496	393.33

Tabela 6: Resultados pergunta 4

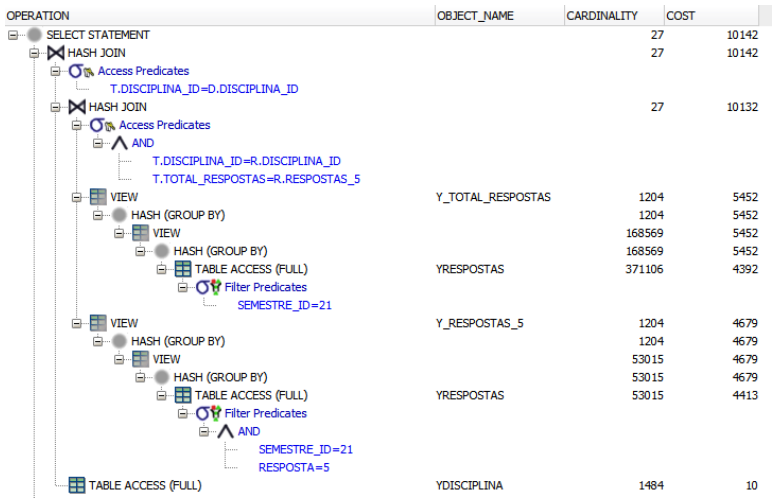
4.4.3 Plano de execução

(x)



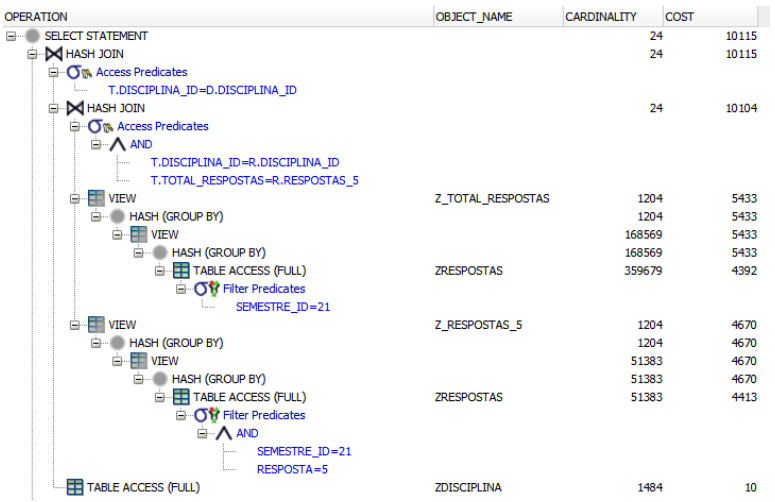
OPERATION	OBJECT_NAME	CARDINALITY	COST
SELECT STATEMENT		22	10069
HASH JOIN		22	10069
Access Predicates			
T.DISCIPLINA_ID=D.DISCIPLINA_ID			
HASH JOIN		22	10058
Access Predicates			
AND			
T.DISCIPLINA_ID=R.DISCIPLINA_ID			
T.TOTAL_RESPOSTAS=R.RESPOSTAS_5			
VIEW	X_TOTAL_RESPOSTAS	1204	5400
HASH (GROUP BY)		1204	5400
VIEW		168569	5400
HASH (GROUP BY)		168569	5400
TABLE ACCESS (FULL)	XRESPOSTAS	338964	4392
Filter Predicates			
SEMESTRE_ID=21			
VIEW	X_RESPOSTAS_5	1204	4657
HASH (GROUP BY)		1204	4657
VIEW		48423	4657
HASH (GROUP BY)		48423	4657
TABLE ACCESS (FULL)	XRESPOSTAS	48423	4413
Filter Predicates			
AND			
SEMESTRE_ID=21			
RESPOSTA=5			
TABLE ACCESS (FULL)	XDISCIPLINA	1484	10

(y)



OPERATION	OBJECT_NAME	CARDINALITY	COST
SELECT STATEMENT		27	10142
HASH JOIN		27	10142
Access Predicates			
T.DISCIPLINA_ID=D.DISCIPLINA_ID			
HASH JOIN		27	10132
Access Predicates			
AND			
T.DISCIPLINA_ID=R.DISCIPLINA_ID			
T.TOTAL_RESPOSTAS=R.RESPOSTAS_5			
VIEW	Y_TOTAL_RESPOSTAS	1204	5452
HASH (GROUP BY)		1204	5452
VIEW		168569	5452
HASH (GROUP BY)		168569	5452
TABLE ACCESS (FULL)	YRESPOSTAS	371106	4392
Filter Predicates			
SEMESTRE_ID=21			
VIEW	Y_RESPOSTAS_5	1204	4679
HASH (GROUP BY)		1204	4679
VIEW		53015	4679
HASH (GROUP BY)		53015	4679
TABLE ACCESS (FULL)	YRESPOSTAS	53015	4413
Filter Predicates			
AND			
SEMESTRE_ID=21			
RESPOSTA=5			
TABLE ACCESS (FULL)	YDISCIPLINA	1484	10

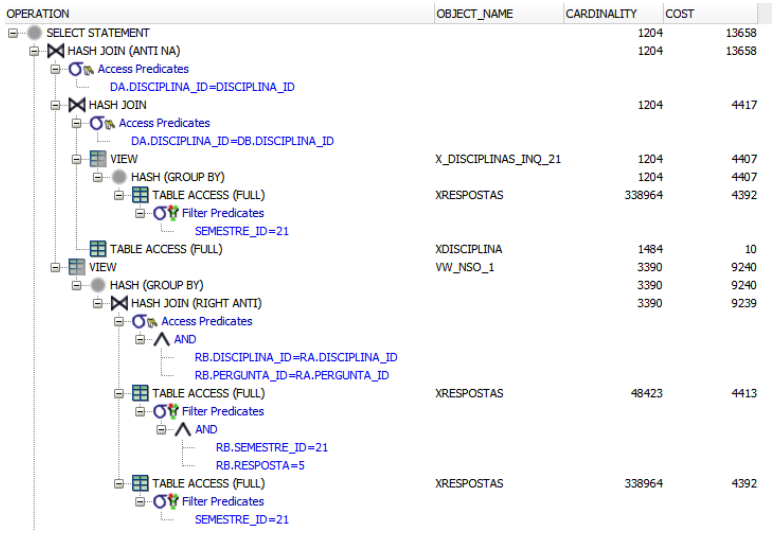
(z)



OPERATION	OBJECT_NAME	CARDINALITY	COST
SELECT STATEMENT		24	10115
HASH JOIN		24	10115
Access Predicates			
T.DISCIPLINA_ID=D.DISCIPLINA_ID			
HASH JOIN		24	10104
Access Predicates			
AND			
T.DISCIPLINA_ID=R.DISCIPLINA_ID			
T.TOTAL_RESPOSTAS=R.RESPOSTAS_5			
VIEW	Z_TOTAL_RESPOSTAS	1204	5433
HASH (GROUP BY)		1204	5433
VIEW		168569	5433
HASH (GROUP BY)		168569	5433
TABLE ACCESS (FULL)	ZRESPOSTAS	359679	4392
Filter Predicates			
SEMESTRE_ID=21			
VIEW	Z_RESPOSTAS_5	1204	4670
HASH (GROUP BY)		1204	4670
VIEW		51383	4670
HASH (GROUP BY)		51383	4670
TABLE ACCESS (FULL)	ZRESPOSTAS	51383	4413
Filter Predicates			
AND			
SEMESTRE_ID=21			
RESPOSTA=5			
TABLE ACCESS (FULL)	ZDISCIPLINA	1484	10

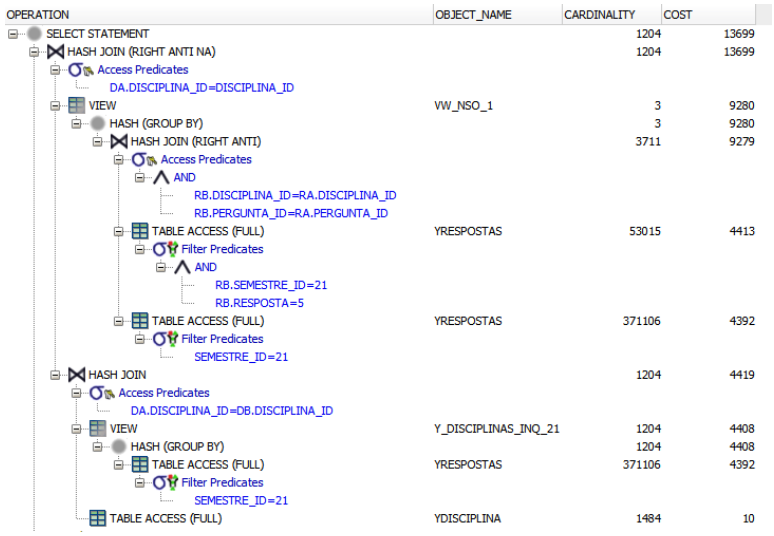
Figura 12: Plano de execução da alínea 4 *contagem*

(x)



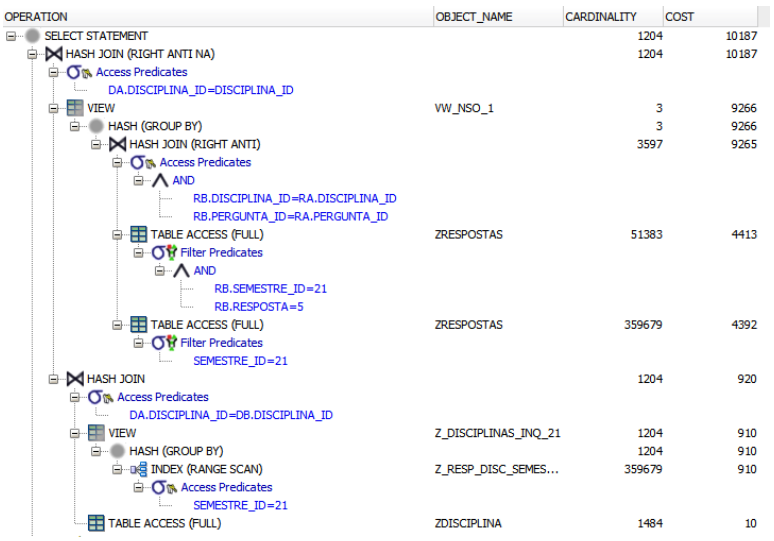
OPERATION	OBJECT_NAME	CARDINALITY	COST
SELECT STATEMENT		1204	13658
HASH JOIN (ANTI NA)		1204	13658
Access Predicates			
DA.DISCIPLINA_ID=DISCIPLINA_ID			
HASH JOIN		1204	4417
Access Predicates			
DA.DISCIPLINA_ID=DB.DISCIPLINA_ID			
VIEW	X_DISCIPLINAS_INQ_21	1204	4407
HASH (GROUP BY)		1204	4407
TABLE ACCESS (FULL)	XRESPOSTAS	338964	4392
Filter Predicates			
SEMESTRE_ID=21			
TABLE ACCESS (FULL)	XDISCIPLINA	1484	10
VIEW	VW_NSO_1	3390	9240
HASH (GROUP BY)		3390	9240
HASH JOIN (RIGHT ANTI)		3390	9239
Access Predicates			
AND			
RB.DISCIPLINA_ID=RA.DISCIPLINA_ID			
RB.PERGUNTA_ID=RA.PERGUNTA_ID			
TABLE ACCESS (FULL)	XRESPOSTAS	48423	4413
Filter Predicates			
AND			
RB.SEMESTRE_ID=21			
RB.RESPOSTA=5			
TABLE ACCESS (FULL)	XRESPOSTAS	338964	4392
Filter Predicates			
SEMESTRE_ID=21			

(y)



OPERATION	OBJECT_NAME	CARDINALITY	COST
SELECT STATEMENT		1204	13699
HASH JOIN (RIGHT ANTI NA)		1204	13699
Access Predicates			
DA.DISCIPLINA_ID=DISCIPLINA_ID			
VIEW	VW_NSO_1	3	9280
HASH (GROUP BY)		3	9280
HASH JOIN (RIGHT ANTI)		3711	9279
Access Predicates			
AND			
RB.DISCIPLINA_ID=RA.DISCIPLINA_ID			
RB.PERGUNTA_ID=RA.PERGUNTA_ID			
TABLE ACCESS (FULL)	YRESPOSTAS	53015	4413
Filter Predicates			
AND			
RB.SEMESTRE_ID=21			
RB.RESPOSTA=5			
TABLE ACCESS (FULL)	YRESPOSTAS	371106	4392
Filter Predicates			
SEMESTRE_ID=21			
HASH JOIN		1204	4419
Access Predicates			
DA.DISCIPLINA_ID=DB.DISCIPLINA_ID			
VIEW	Y_DISCIPLINAS_INQ_21	1204	4408
HASH (GROUP BY)		1204	4408
TABLE ACCESS (FULL)	YRESPOSTAS	371106	4392
Filter Predicates			
SEMESTRE_ID=21			
TABLE ACCESS (FULL)	YDISCIPLINA	1484	10

(z)



OPERATION	OBJECT_NAME	CARDINALITY	COST
SELECT STATEMENT		1204	10187
HASH JOIN (RIGHT ANTI NA)		1204	10187
Access Predicates			
DA.DISCIPLINA_ID=DISCIPLINA_ID			
VIEW	VW_NSO_1	3	9266
HASH (GROUP BY)		3	9266
HASH JOIN (RIGHT ANTI)		3597	9265
Access Predicates			
AND			
RB.DISCIPLINA_ID=RA.DISCIPLINA_ID			
RB.PERGUNTA_ID=RA.PERGUNTA_ID			
TABLE ACCESS (FULL)	ZRESPOSTAS	51383	4413
Filter Predicates			
AND			
RB.SEMESTRE_ID=21			
RB.RESPOSTA=5			
TABLE ACCESS (FULL)	ZRESPOSTAS	359679	4392
Filter Predicates			
SEMESTRE_ID=21			
HASH JOIN		1204	920
Access Predicates			
DA.DISCIPLINA_ID=DB.DISCIPLINA_ID			
VIEW	Z_DISCIPLINAS_INQ_21	1204	910
HASH (GROUP BY)		1204	910
INDEX (RANGE SCAN)	Z_RESP_DISC_SEMES...	359679	910
Access Predicates			
SEMESTRE_ID=21			
TABLE ACCESS (FULL)	ZDISCIPLINA	1484	10

Figura 13: Plano de execução da alínea 4 *dupla negação*

4.4.4 Resultados

	DISCIPLINA_ID	SIGLA
1	1235	ALGE
2	1236	AMAT
3	1237	FPRO
4	1238	IOCO
5	1239	IBDA
6	1245	AEDA
7	1265	AIAD
8	1266	GEMP
9	1267	AMAT1
10	1268	ALGE
11	1269	PROG1
12	1270	LSDI
13	1387	LESO
14	1388	SDIS
15	1393	MCM II
16	1402	EI
17	1403	TC
18	1405	IO
19	1124	GOPE
20	1144	RTSEE
21	1156	TPRE
22	1158	EESO
23	1688	ALGA

Figura 14: Resultado 4

4.4.5 Conclusões

Na execução das interrogações escritas para esta pergunta, verificou-se que existem diferenças na execução de cada uma das estratégias (contagem e dupla negação), na medida em que, para os conjuntos de tabelas x e y, existe uma diferença de custo relativamente significativa entre ambas, com vantagem para a estratégia que envolve contagem.

No que diz respeito ao conjunto de tabelas z, a diferença entre as duas estratégias não foi significativa, e até desprezável, uma vez que na estratégia de dupla negação o sistema conseguiu tomar partido da existência de um índice nas colunas disciplina_id e semestre_id, reduzindo o custo total da interrogação.

Referências

[1] Gabriel David. Otimização de interrogações. Enunciado do projeto, `https://moodle.up.pt/pluginfile.php/102650/mod_resource/content/0/37-TrabOpt3.pdf`, 2016.