

FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

Mestrado Integrado em Engenharia Informática e Computação

Tecnologias de Bases de Dados

Otimização de Queries: Serviço de Ensino

Ana Rita Torres - up201406093@fe.up.pt Catarina Correia - up201405765@fe.up.pt José Oliveira - up201406208@fe.up.pt

9 de abril de 2018

Conteúdo

1	Obj	etivo d	do Projeto	3
2	Cria	ação d	e Índices	4
3	Res	posta	às Perguntas	6
	3.1	Pergu	nta 1 - Seleção	6
		3.1.1	Formulação SQL	6
		3.1.2	Resposta	6
		3.1.3	Plano de Execução	6
		3.1.4	Tempos de Execução	8
	3.2	Pergu	nta 2 - Agregação	8
		3.2.1	Formulação SQL	8
		3.2.2	Resposta	8
		3.2.3	Plano de Execução	8
		3.2.4	Tempos de Execução	9
	3.3	Pergu	nta 3 - A	10
		3.3.1	Formulação SQL	10
		3.3.2	Resposta	11
		3.3.3	Plano de Execução	12
		3.3.4	Tempos de Execução	13
	3.4		nta 3 - B	13
		3.4.1	Formulação SQL	13
		3.4.2	Resposta	15
		3.4.3	Plano de Execução	16
		3.4.4	Tempos de Execução	16
	3.5	Pergui		17
	0.0	3.5.1	Formulação SQL	17
		3.5.2	Resposta	17
		3.5.2	Plano de Execução	17
		3.5.4	Tempos de Execução	19
	3.6		nta 5 - A	19
	5.0	3.6.1	Formulação SQL	19
		3.6.2	Resposta	20
		3.6.2	•	$\frac{20}{20}$
			Plano de Execução	-
	2.7	3.6.4	Tempos de Execução	21 22
	3.7	_	nta 5 - B	
		3.7.1	Formulação SQL	22
		3.7.2	Resposta	22
		3.7.3	Plano de Execução	22
	0.0	3.7.4	Tempos de Execução	23
	3.8	Pergu		24
		3.8.1	Formulação SQL	24
		3.8.2	Resposta	24
		383	Plano de Evecução	24

3.8.4	Tempos de Execução																				26
0.0.1	zempes de zneedgae	•	 •	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	

1 Objetivo do Projeto

No âmbito da unidade curricular de Tecnologias de Bases de Dados, foi proposto ao grupo desenvolver um projeto que permitisse ao grupo analisar diversos planos de execução de SQL no base de dados de teste. Assim como compreender o impacto do uso de índices e das diferentes estratégias de organização de queries.

Para este efeito, o grupo analisou uma base de dados relativa a um serviço de ensino.

2 Criação de Índices

De forma a melhorar a eficiência das *queries* (diminuir o tempo de execução e reduzir o custo associado às mesmas), foram criados os seguintes índices:

• B-Tree na coluna ANOLETIVO na tabela ZTIPOSAULA

```
CREATE INDEX ANO
ON ZTIPOSAULA(ANO_LETIVO);
```

 B-Tree nas colunas CODIGO, ANOLETIVO e PERIODO da tabela ZTI-POSAULA

```
CREATE INDEX CAP
ON ZTIPOSAULA(CODIGO, ANO_LETIVO, PERIODO);
```

• B-Tree na coluna TIPO da tabela ZTIPOSAULA

```
CREATE INDEX T
ON ZTIPOSAULA(TIPO);
```

• B-Tree na coluna ID da tabela ZDSD

```
CREATE INDEX I
ON ZDSD(ID);
```

• B-Tree nas colunas CURSO e DESIGNACAO da tabela ZUCS

```
CREATE INDEX U
ON ZUCS(CURSO, DESIGNACAO);
```

• Bitmap na coluna CURSO da tabela ZUCS

```
CREATE BITMAP INDEX C
ON ZUCS(CURSO);
```

Para além dos índices já existentes nas chaves primárias de cada tabela, o grupo considerou necessário criar os índices acima, com base em múltiplos fatores:

- 1. Em tabelas grandes onde se pretende retornar poucas linhas
- 2. Em colunas usadas para juntar tabelas (JOIN)
- 3. Em colunas usadas para selecionar (WHERE)
- 4. Em colunas com pouca cardinalidade

5. Em colunas em que os valores se repetem muitas vezes

Usou-se índices Bitmap em casos que cumprem pelo menos um dos pontos 4 ou 5 (acima referidos). Usou-se índices B-Tree nos restantes casos.

É importante referir que a escolha destes índices criados justifica-se pelo simples facto de só serem efetuados SELECTs. Se também se fizesse INSERTs, UPDATEs e DELETEs, os índices seriam diferentes.

3 Resposta às Perguntas

3.1 Pergunta 1 - Seleção

3.1.1 Formulação SQL

```
SELECT DISTINCT XUCS.CODIGO, DESIGNACAO, XOCORRENCIAS.ANO_LETIVO, INSCRITOS, TIPO, TURNOS
FROM XUCS
JOIN XOCORRENCIAS
ON XUCS.CODIGO = XOCORRENCIAS.CODIGO
JOIN XTIPOSAULA
ON (XUCS.CODIGO = XTIPOSAULA.CODIGO
AND XTIPOSAULA.ANO_LETIVO = XOCORRENCIAS.ANO_LETIVO
AND XTIPOSAULA.PERIODO = XOCORRENCIAS.PERIODO)
WHERE XUCS.DESIGNACAO = 'Bases de Dados'
AND XUCS.CURSO = 275;
```

3.1.2 Resposta

	∯ CODIGO		CAO			∯ TIPO	⊕ TURNOS
1	EIC3106	Bases de	Dados	2003/2004	92	T	1
2	EIC3111	Bases de	Dados	2005/2006	(null)	T	1
3	EIC3106	Bases de	Dados	2003/2004	92	TP	4
4	EIC3106	Bases de	Dados	2004/2005	114	TP	4
5	EIC3111	Bases de	Dados	2005/2006	(null)	TP	6
6	EIC3106	Bases de	Dados	2004/2005	114	T	1

3.1.3 Plano de Execução

Pelas imagens abaixo pode-se concluir que com a criação de chaves primárias e chaves estrangeiras, há uma grande redução de custos na *query* em Y em relação à *query* em X. Isto deve-se ao facto das colunas que representam chaves primárias serem indexadas.

No entanto há ainda uma melhoria mais significativa ao correr a query em Z. Esta melhoria deve-se aos índices U e CAP como se pode ver pela coluna OPTIONS na imagem abaixo.

```
CREATE INDEX U
ON ZUCS(DESIGNACAO, CURSO);
CREATE INDEX CAP
ON ZTIPOSAULA(CODIGO, ANO_LETIVO, PERIODO);
```

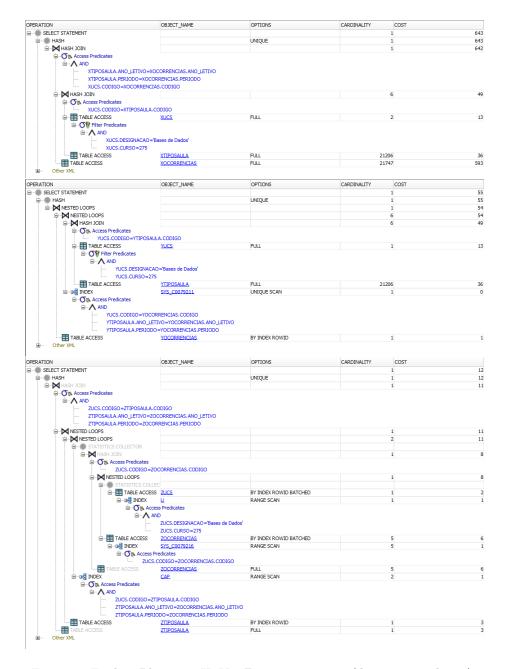


Figura 1: Explain Plan para X, Y e Z, respetivamente (de cima para baixo)

3.1.4 Tempos de Execução

Para além das alterações a nível de custo também se nota uma melhoria nos tempos de X para Y devido à introdução de chaves primárias e estrangeiras, e ainda de Y para Z e como tal de X para Z devido aos índices.

X(s)	Y(s)	Z(s)
0,076	0,043	0,018

3.2 Pergunta 2 - Agregação

3.2.1 Formulação SQL

```
SELECT TIPO, SUM(HORAS_TURNO * TURNOS)

FROM XTIPOSAULA

JOIN XOCORRENCIAS

ON (XOCORRENCIAS.CODIGO = XTIPOSAULA.CODIGO

AND XOCORRENCIAS.ANO_LETIVO = XTIPOSAULA.ANO_LETIVO

AND XOCORRENCIAS.PERIODO = XTIPOSAULA.PERIODO)

JOIN XUCS

ON XTIPOSAULA.CODIGO = XUCS.CODIGO

WHERE XTIPOSAULA.ANO_LETIVO = '2004/2005'

AND XUCS.CURSO = 233

GROUP BY TIPO;
```

3.2.2 Resposta

	∯ TIPO	⊕ SUM(HORAS_TURNO*TURNOS)
1	P	581,5
2	Т	308
3	TP	697,5

3.2.3 Plano de Execução

É possível observar, nas imagens que se seguem, que, com a criação de chaves primárias e chaves estrangeiras, os custos da *query* em Y reduzem-se significativamente quando comparados com a *query* em X. Tal deve-se ao facto das colunas que representam chaves primárias serem colunas indexadas.

No entanto, existe uma melhoria ainda mais significativa quando se executa a query em Z. Esta melhoria deve-se à existência dos índices "ANO" e "C", como é possível observar na coluna OPTIONS das seguintes imagens do Explain Plan.

CREATE INDEX ANO
ON ZTIPOSAULA(ANO_LETIVO);
CREATE BITMAP INDEX C
ON ZUCS(CURSO);



Figura 2: Explain Plan para X, Y e Z, respetivamente (de cima para baixo)

3.2.4 Tempos de Execução

Para além do já observado no *Explain Plan*, pode comprovar-se efetivamente a eficiência das queries nos tempos de execução na seguinte tabela:

X(s)	Y(s)	Z(s)
0,056	0,033	0,021

3.3 Pergunta 3 - A

3.3.1 Formulação SQL

3.3.2 Resposta

	A copico			
	∯ CODIGO	∯ CODIGO	CODIGO	
CODIGO	CI008	EC5287	MEMT135	
MEB205	CI037	MDI1209	MPPAU2218	
MEMT102	MTM110	MDI1108	EEC5272	
MFAMF1108	EQ308	MEMT120	EQ411	
MEMT1000	MDI1100	MPFCA105	MEM180	
MPPAU2215	MEAM1312	MPFCA106	MEA219	
MPPAU1114	MEAM1310	MPFCA100	MEA215	
MEM183	MPPAU2216	EIC5126	MEMT110	
MEA112	MEM157			
MPPAU2219	MEM184	CI004	MEMT106	
MPFCA104	MDI1106	MEA412	MPFCA100	
MPFCA203	MDI1107	GEI512	MPFCA102	
MPFCA205	MDI1205	MPPAU1112		
EEC5022	MPPAU1115	MEM5000	MPFCA202	
EIC5123	MPPAU2217	EQ418	EIC4223	
EIC5125	MMCCE1220	MEMT131	EIC5124	
CI003	MPFCA103	MEM191	CI023	
CI016	MPFCA107	MEA217	CI020	
CI017	EEC2207	MEA320	MTM114	
CI009		EI1107	MEMT105	
CI027	CI002	MPFCA206	MEAM5000	
MTM111	CI011	EIC4220	MEAM1314	
EQ407	MTM115	EIC4221	MPPAU2220	
MEM158	MDI1204	EIC5122	MTM108	
MEM188	MDI1105	CI025	MEEC1053	
MEA319	MEMT2000	MEMT107	MEM163	
MEA415	MEMT100	MEB105	MEM175	
MDI1207	MPPAU1113	MEM179	MEM187	
MPFCA101	EC5280	MEM182	MVC1211	EIC4222
EIC5127	EIC3209	MPFCA204	MEA414	EIC4224
EIC5129	MEM181	EMM528	MDI1206	EIC4225
CI018	MEM189	MEM1205	MDI1208	CI014
CI007	MEA216	MTM104	MEB204	CI019
CI013	MEST210	MDI1103	EC5200	CI038

3.3.3 Plano de Execução

Ao analisar as imagens abaixo apresentadas, pode-se ver que a *query* em Z apresenta resultados de custo melhores que as outras *queries*. A razão pela qual isso acontece foi a criação de índices relativos à tabela TIPOSAULA, sendo eles:

```
CREATE INDEX CAP
ON ZTIPOSAULA(CODIGO, ANO_LETIVO, PERIODO);
CREATE INDEX I
ON ZDSD(ID);
```

Uma diferença ainda mais significativa é verificada entre as *queries* em X e em Y, diferença esta que é consequência do uso de chaves primárias e estrangeiras nas tabelas em Y.





Figura 3: Explain Plan para X, Y e Z, respetivamente (de cima para baixo)

3.3.4 Tempos de Execução

Os tempos de execução apresentados na tabela reafirmam o anteriormente representado nos planos de execução, ou seja, que a query mais eficiente é a em Z e a menos eficiente é a query em X.

X(s)	Y(s)	Z(s)
0.148	0.062	0.035

3.4 Pergunta 3 - B

3.4.1 Formulação SQL

CREATE OR REPLACE VIEW CODIGOS AS
SELECT DISTINCT CODIGO
FROM XTIPOSAULA
JOIN XDSD
ON XTIPOSAULA.ID = XDSD.ID
WHERE ANO_LETIVO = '2003/2004';
SELECT DISTINCT XUCS.CODIGO
FROM XUCS
JOIN XOCORRENCIAS
ON XOCORRENCIAS.CODIGO = XUCS.CODIGO

AND ANO_LETIVO='2003/2004'
LEFT OUTER JOIN CODIGOS
ON CODIGOS.CODIGO=XUCS.CODIGO
WHERE CODIGOS.CODIGO IS NULL;

3.4.2 Resposta

A	A	CODIGO	
∯ CODIGO	∯ CODIGO	EC5287	CODIGO
MEB205	CI008	MDI1209	MEMT135
MEMT102	CI037	MDI1108	MPPAU2218
MFAMF1108	MTM110	MEMT120	EEC5272
MEMT1000	EQ308	MPFCA105	EQ411
MPPAU2215	MDI1100	MPFCA106	MEM180
MPPAU1114	MEAM1312	MPFCA201	MEA219
MEM183	MEAM1310	EIC5126	MEA215
MEA112	MPPAU2216		MEMT110
MPPAU2219	MEM157	CI004	MEMT106
MPFCA104	MEM184	MEA412	MPFCA100
MPFCA203	MDI1106	GEI512	MPFCA102
MPFCA205	MDI1107	MPPAU1112	MPFCA200
EEC5022	MDI1205	MEM5000	MPFCA202
EIC5123	MPPAU1115	EQ418	EIC4223
EIC5125	MPPAU2217	MEMT131	EIC5124
CI003	MMCCE1220	MEM191	CI023
CI016	MPFCA103	MEA217	CI020
CI017	MPFCA107	MEA320	MTM114
CI009	EEC2207	EI1107	MEMT105
CI027	CI002	MPFCA206	MEAM5000
MTM111	CI011	EIC4220	MEAM1314
EQ407	MTM115	EIC4221	MPPAU2220
MEM158	MDI1204	EIC5122	MTM108
MEM188	MDI1105	CI025	MEEC1053
MEA319	MEMT2000	MEMT107	MEM163
MEA415	MEMT100	MEB105	MEM175
MDI1207	MPPAU1113	MEM179	MEM187
MPFCA101	EC5280	MEM182	MVC1211
EIC5127	EIC3209	MPFCA204	MEA414
EIC5129	MEM181	EMM528	MDI1206
CI018	MEM189	MEM1205	MDI1208
CI007	MEA216	MTM1154	MEB204
CI013	MEST210	MDI1103	EC5200

∯ CODIGO	
EIC4222	
EIC4224	
EIC4225	
CI014	
CI019	
CI038	

3.4.3 Plano de Execução

Nesta alínea pode-se observar uma situação muito semelhante à anterior, na medida em que a utilização de índices conduz ao menor custo. E, o uso de chaves primárias e estrangeira, embora não apresente o menor custo já é melhor do que a query em X que não tem qualquer tipo de índices ou chaves associados.

É de notar também que as perguntas 3a e a 3b apresentam o mesmo resultado com queries diferentes e custos com valores bastante próximos.



Figura 4: Explain Plan para X, Y e Z, respetivamente (de cima para baixo)

3.4.4 Tempos de Execução

Novamente, os tempos de execução reforçam o observado no plano de execuções. Comparando com a pergunta 3a verifica-se uma diminuição significativa no tempo de execução na query em X.

X(s)	Y(s)	Z(s)
0.072	0.053	0.035

3.5 Pergunta 4

3.5.1 Formulação SQL

CREATE OR REPLACE VIEW SOMA_HORAS_PROFESSOR_TIPO AS
SELECT XDOCENTES.NR, NOME, TIPO, SUM(HORAS * FATOR) AS SOMA
FROM XDOCENTES
JOIN XDSD
ON XDSD.NR = XDOCENTES.NR
JOIN XTIPOSAULA
ON XTIPOSAULA.ID = XDSD.ID
WHERE XTIPOSAULA.ANO_LETIVO = '2003/2004'
GROUP BY XDOCENTES.NR, NOME, TIPO;

CREATE OR REPLACE VIEW MAX_HORAS_TIPO AS SELECT TIPO, MAX(SOMA) AS MAXIMO FROM SOMA_HORAS_PROFESSOR_TIPO GROUP BY TIPO;

SELECT NR, NOME, MAX_HORAS_TIPO.TIPO, MAX_HORAS_TIPO.MAXIMO AS TOTAL_HORAS FROM SOMA_HORAS_PROFESSOR_TIPO
JOIN MAX_HORAS_TIPO

ON (MAX_HORAS_TIPO.TIPO = SOMA_HORAS_PROFESSOR_TIPO.TIPO AND MAX_HORAS_TIPO.MAXIMO = SOMA_HORAS_PROFESSOR_TIPO.SOMA);

3.5.2 Resposta

	♦ NR ♦ NOME	∯ TIPO	↑ TOTAL_HORAS
1	210006 João Carlos Pascoal de Faria	OT	3,5
2	249564 Cecília do Carmo Ferreira da Silva	TP	26
3	208187 António Almerindo Pinheiro Vieira	P	30
4	207638 Fernando Francisco Machado Veloso Gomes	T	30,67

3.5.3 Plano de Execução

Nas seguintes imagens é possível observar que os *Explain Plans* para X e para Y não diferem. Não há melhoria de custos e as cardinalidades mantêm-se. Isto acontece porque as colunas usadas para junção de tabelas (JOIN) e para seleção (WHERE) não representam chaves primárias (colunas indexadas).

Para Z, já são usadas colunas indexadas para a junção de tabelas (JOIN) e a seleção (WHERE), pelo que já se nota melhorias nos custos da query. Tal

pode ser observado na junção das tabelas ZDSD E ZTIPOSAULA e na seleção "ZTIPOSAULA.ID = '2003/2004'", por exemplo.





Figura 5: Explain Plan para X, Y e Z, respetivamente (de cima para baixo)

3.5.4 Tempos de Execução

Tendo em conta a análise efetuada acima, pode comprovar-se a eficiência da query em Z com os tempos de execução. É possível observar que os tempos de execução em X e em Y são aproximadamente os mesmos, mas que a query em Z é ligeiramente mais rápida.

X(s)	Y(s)	Z(s)
0.074	0.078	0.052

3.6 Pergunta 5 - A

3.6.1 Formulação SQL

DROP INDEX A;
CREATE INDEX A ON XTIPOSAULA (ANO_LETIVO,TIPO);

SELECT XOCORRENCIAS.CODIGO, XOCORRENCIAS.ANO_LETIVO, XOCORRENCIAS.PERIODO, XTIPOSAULA.TURNOS * XTIPOSAULA.HORAS_TURNO AS TOTAL_HORAS FROM XOCORRENCIAS

JOIN XTIPOSAULA

ON XOCORRENCIAS.CODIGO = XTIPOSAULA.CODIGO

AND XOCORRENCIAS.ANO_LETIVO = XTIPOSAULA.ANO_LETIVO

AND XOCORRENCIAS.PERIODO = XTIPOSAULA.PERIODO

WHERE (XTIPOSAULA.ANO_LETIVO = '2002/2003'

OR XTIPOSAULA.ANO_LETIVO = '2003/2004')

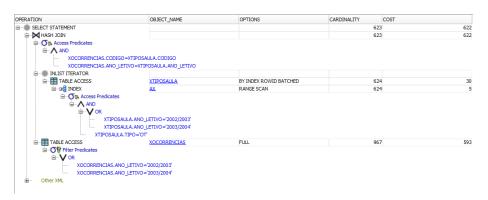
AND XTIPOSAULA.TIPO = 'OT';

3.6.2 Resposta

				★ TOTAL_HORAS
1	EIC5202	2002/2003	25	27
2	EIC5202	2003/2004	25	24

3.6.3 Plano de Execução

Pelas imagens abaixo, pode concluir-se que com a criação de chaves primárias e chaves estrangeiras , há uma grande redução de custos na query em Y e em Z em relação à query em X. Isto deve-se ao facto das colunas que representam chaves primárias serem indexadas. Contudo, entre Y e Z não se vê diferenças nos custos e nas cardinalidades.



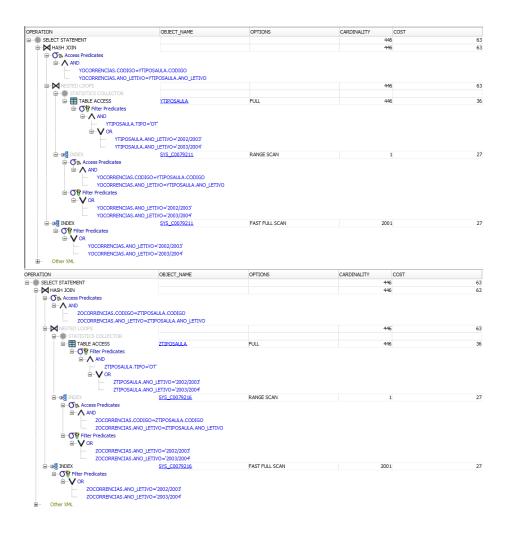


Figura 6: Explain Plan para X, Y e Z, respetivamente (de cima para baixo)

3.6.4 Tempos de Execução

Para além das melhorias observadas nos custos que as queries Y e Z apresentam relativamente a X, também se pode observar nos tempos de execução que são significativamente mais rápidas do a query em X. No entanto, os tempos de execução de Y e Z são muito aproximados.

X(s)	Y(s)	Z(s)
0,062	0,036	0.032

3.7 Pergunta 5 - B

3.7.1 Formulação SQL

```
DROP INDEX B;
CREATE BITMAP INDEX B ON XTIPOSAULA (ANO_LETIVO,TIPO);

SELECT XOCORRENCIAS.CODIGO, XOCORRENCIAS.ANO_LETIVO, XOCORRENCIAS.PERIODO,
XTIPOSAULA.TURNOS * XTIPOSAULA.HORAS_TURNO AS TOTAL_HORAS
FROM XOCORRENCIAS
JOIN XTIPOSAULA
ON XOCORRENCIAS.CODIGO = XTIPOSAULA.CODIGO
AND XOCORRENCIAS.ANO_LETIVO = XTIPOSAULA.ANO_LETIVO
AND XOCORRENCIAS.PERIODO = XTIPOSAULA.PERIODO
WHERE (XTIPOSAULA.ANO_LETIVO = '2002/2003'
OR XTIPOSAULA.ANO_LETIVO = '2003/2004')
AND XTIPOSAULA.TIPO = '0T';
```

3.7.2 Resposta

		\$ ANO_LETIVO	♦ PERIODO	↑ TOTAL_HORAS
1	EIC5202	2002/2003	2S	27
2	EIC5202	2003/2004	25	24

3.7.3 Plano de Execução

Pelas imagens abaixo pode-se concluir que com a criação de chaves primárias e chaves estrangeiras , há uma grande redução de custos no query em Y e em Z em relação à query em X. Isto deve-se ao facto das colunas que representam chaves primárias serem indexadas. Tal como na pergunta anterior, também não existem diferenças entre os custos e as cardinalidades de Y e Z.

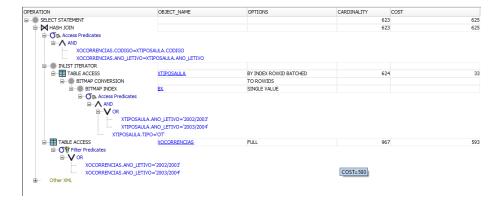




Figura 7: Explain Plan para X, Y e Z, respetivamente (de cima para baixo)

3.7.4 Tempos de Execução

Assim como na pergunta anterior, para além das melhorias observadas nos custos que as *queries* Y e Z apresentam relativamente a X, também se pode observar nos tempos de execução que são significativamente mais rápidas do a *query* em X. No entanto, os tempos de execução de Y e Z são semelhantes.

X(s)	Y(s)	Z(s)
0,037	0,029	0.027

3.8 Pergunta 6

3.8.1 Formulação SQL

```
SELECT DISTINCT XUCS.CURSO
   FROM XUCS
    WHERE
        (EXISTS(
            SELECT XTIPOSAULA.ID
            FROM XTIPOSAULA
            WHERE XTIPOSAULA.TIPO='P'
                AND XUCS.CODIGO=XTIPOSAULA.CODIGO)
        AND EXISTS(
            SELECT XTIPOSAULA.ID
            FROM XTIPOSAULA
            WHERE XTIPOSAULA.TIPO='TP'
                AND XUCS.CODIGO=XTIPOSAULA.CODIGO)
        AND EXISTS(
            SELECT XTIPOSAULA.ID
            FROM XTIPOSAULA
            WHERE XTIPOSAULA.TIPO='T'
                AND XUCS.CODIGO=XTIPOSAULA.CODIGO)
        AND EXISTS(
            SELECT XTIPOSAULA.ID
            FROM XTIPOSAULA
            WHERE XTIPOSAULA.TIPO='L'
                AND XUCS.CODIGO=XTIPOSAULA.CODIGO)
        AND EXISTS(
            SELECT XTIPOSAULA.ID
            FROM XTIPOSAULA
            WHERE XTIPOSAULA.TIPO='OT'
                AND XUCS.CODIGO=XTIPOSAULA.CODIGO)
        );
```

3.8.2 Resposta



3.8.3 Plano de Execução

Analisando os planos de execução que se seguem pode-se verificar que todas as *queries* apresentam o mesmo custo e cardinalidade. O que significa que nesta alínea a criação de chaves e indices não teve qualquer tipo de efeito sobre o custo.



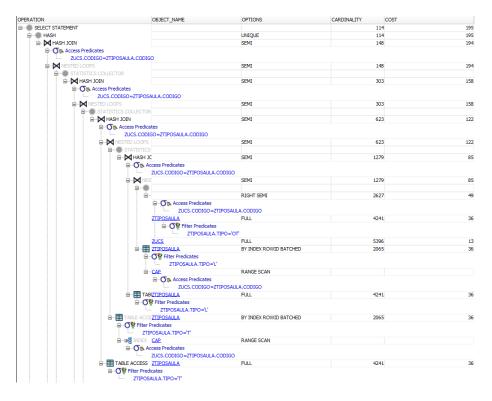


Figura 8: Explain Plan para X, Y e Z, respetivamente (de cima para baixo)

3.8.4 Tempos de Execução

Embora não se verifique qualquer alteração a nível do custo, pode-se verificar que a presença de indices e chaves primárias e estrangeiras afetou o tempo de execução das *queries*. Sendo esta a ordem da mais eficiente para a menos eficiente: Z,Y, X.

X(s)	Y(s)	Z(s)
0,196	0,081	0.035