Universidade de Brasília

Faculdade do Gama

Sistemas de Banco de Dados 2

Estudando *Tuning* em Banco de Dados Relacional

(Oportunidade Letiva de Discentes Experientes)

Na disciplina de Sistemas de Banco de Dados 2 (SBD2), oferecida no curso de Engenharia de Software da UnB, alguns estudantes mais experientes nesta disciplina (monitores) compartilharam seus estudos no tema de *Tuning* em Banco de Dados Relacionais nas Consultas SQL, lecionando parte de uma aula regular com a turma de SBD2 atual.

Nesta atividade foram explicados conceitos relevantes ao tema e demonstradas operações simples de serem realizadas pelo Workbench MySQL. Assim, está sendo apresentado abaixo como exemplo e modelo a consulta trabalhada em sala de aula envolvendo uma base de dados utilizada pela turma em exercícios solicitado pelo docente em atividade anterior (base disponível na Área de Compartilhamento - **/aulas/basesDados/projetoBaseDados_Jogos_2020.zip**).

Em seguida, são apresentadas três demandas solicitadas por seu cliente no qual você deverá elaborar uma consulta SQL que resolva cada problema demandado e depois apresentar o resultado da explicação fornecida pelo MySQL (*explain*) sobre sua consulta proposta inicialmente.

Após a explicação de sua solução (*Result Grid*) você deverá solicitar a análise do MySQL (*analyze*) para otimizar a sua proposta inicial, sendo ao final apresentado o resultado e a consulta que atenda a demanda otimizada.

Dessa forma, a apresentação realizada em sala de aula está detalhda a seguir com a indicação de primeira demanda (1) do cliente e você deverá elaborar as outras TRÊS propostas de soluções seguindo EXATAMENTE o mesmo padrão realizado para atender a primeira demanda.

1) Primeira Demanda

Elaborar uma consulta que recupere informações sobre os jogos que possuem vendas na América do Norte superiores a 1 milhão de unidades (note que todos os armazenamentos nesta base estão em unidade de milhão). A consulta deve retornar o nome do jogo, o gênero e as vendas nas regiões Norte-Americana, Europeia e Japonesa, ordenadas de forma decrescentes pela venda na América do Norte.

Proposta de solução inicial:

```
SELECT
    g.name AS game_name,
    ge.description AS genre_description,
    g.na_sales,
    g.eu_sales,
    g.jp_sales
FROM
    GAME g
JOIN
    GENRE ge ON g.id_genre = ge.id_genre
WHERE
    g.na_sales > 1
ORDER BY
    g.na_sales DESC;
```

Na consulta acima foi inicialmente definido os atributos que seriam projetados (apresentados) para o usuário que gerou a demanda como o resultado solicitado para a equipe de **TI** (ou de BD em específico). Em seguida, foi identificada a tabela principal para esta consulta (**GAME**). Uma junção (JOIN) seria necessária entre as tabelas **GAME** e **GENRE** para satisfazer os dados que seriam projetados por esta consulta, sendo implementada a condição para uma junção baseada no atributo **id_genre** presente nas duas tabelas (junção natural nesta consulta porque o projeto desta base de dados implementa esta restrição estabelecendo o relacionamento entre estas duas tabelas).

Na cláusula WHERE desta consulta foi elaborada a condição para a filtragem das vendas acima de um milhão, enquanto a ordenação final solicitada pelo

cliente foi atendida na cláusula ORDER BY desta consulta. Essa ordenação deveria ser decrescente sobre as vendas na América do Norte (**na_sales**).

Solicitando explicação sobre a solução inicial (*explain*):

```
EXPLAIN SELECT
    g.name AS game_name,
    ge.description AS genre_description,
    g.na_sales,
    g.eu_sales,
    g.jp_sales
FROM
    GAME g
JOIN
    GENRE ge ON g.id_genre = ge.id_genre
WHERE
    g.na_sales > 1
ORDER BY
    g.na_sales DESC;
```

Analisando os resultados apresentados pelo MySQL Workbench no *Result Grid* são destacados alguns itens relacionados a consulta proposta inicialmente.

id	select_type	table	partitions	type	possible_keys	key	key_len	ref	rows	filtered	Extra
1	SIMPLE	ge	NULL	ALL	PRIMARY	NULL	NULL	NULL	12	100.00	Using temporary; Using filesort
1	SIMPLE	g	NULL	ref	GAME_GENRE_FK	GAME_GENRE_FK	5	t2_jogos.ge.id_genre	990	33.33	Using where

Interpretando o *Result Grid* do *EXPLAIN*:

Foi possível observar que na tabela **GENRE** (ge) a coluna ***type*** apresenta a expressão **ALL**, que indica que será realizada uma varredura completa na tabela (*full table scan*). Esse tipo de operação não é desejada por ser menos eficiente, especialmente em tabelas com grande volume de dados. Por outro lado, na

tabela **GAME** (g), o filtro é aplicado utilizando o índice disponível pela restrição implementada pela chave estrangeira GAME_GENRE_FK, o que torna o processo mais eficiente em comparação com a varredura completa.

Além disso, na coluna ***Extra*** do *Result Grid* são disponibilizadas informações adicionais sobre o processamento da consulta em cada tabela. Para a tabela GENRE o sistema está utilizando tabela temporária (*Using temporary*) que corresponde a tabela intermediária gerada durante o processamento da consulta, consumindo espaço nos recursos de memória do computador servidor do banco de dados. Na primeira indicação da ***Extra*** ainda está sendo informado que a realização da ordenação solicitada está sendo feita manualmente (expressão *Using filesort*), o que seria ineficiente em consultas envolvendo grandes volumes de dados. Já na tabela GAME a expressão *Using where* esclarecer que a condição da cláusula WHERE estará sendo aplicada como filtro desejado.

Solicitando a análise da solução inicial (*analyze*):

```
EXPLAIN ANALYZE SELECT
        g.name AS game_name,
        ge.description AS genre_description,
        g.na_sales,
        g.eu_sales,
        g.jp_sales
FROM
        GAME g
JOIN
        GENRE ge ON g.id_genre = ge.id_genre
WHERE
        g.na_sales > 1
ORDER BY
        g.na_sales DESC;
```

Interpretando o resultado do *ANALYZE*:

Após a execução da análise sob a consulta inicial o resultado produzido está sendo apresentado a seguir, a fim de abordar os principais esclarecimentos relacionados a primeira análise realizada em uma consulta SQL.

-> Sort: g.na_sales DESC (actual time=60.2..61.2 rows=757 loops=1)

-> Stream results (**cost=2063** rows=3961) (actual time=0.599..58.7 rows=757 loops=1)

-> Nested loop inner join (**cost=2063** rows=3961) (actual time=0.589..56.3 rows=757 loops=1)

-> Table scan on ge (**cost=1.45** rows=12) (actual time=0.0354..0.0707 rows=12 loops=1)

-> Filter: (g.na_sales > 1.00) (**cost=75.5** rows=330) (actual time=0.289..4.48 rows=63.1 loops=12)

-> Index lookup on g using GAME_GENRE_FK (id_genre=ge.id_genre) (**cost=75.5** rows=990) (actual time=0.188..2.97 rows=1004 loops=12)

É importante averiguar o resultado de baixo para cima para acompanhar a sequência do processamento e os seus resultados de apurações intermediária. Assim, a primeira linha do resultado das análises corresponde as apurações finais.

O resultado apresentado acima corresponde ao plano de execução que segue uma sequência de operações realizadas pelo MySQL para atender a consulta proposta inicialmente. As operações mais abaixo nesta sequência são executadas primeiro, e seus resultados são usados em operações mais altas (ou acima nesta sequência).

É interessante também acompanhar os custos (*cost*) que representam o consumo de recursos previstos pelo otimizador do SGBD (Sistema Gerenciador de Banco de Dados) que são necessários para executar cada operação importante para a execução completa da consulta SQL.

No plano de execução é indicado que o índice proveniente da GAME_GENRE_FK será utilizado apenas para a junção entre **GAME** e **GENRE**, enquanto o filtro da cláusula WHERE (g.**na_sales** > 1) fará que o otimizador precise efetuar um processamento extra para examinar as linhas correspondentes para a junção e aplicar o filtro apenas depois dessa junção. Além disso, a tabela **GENRE** realiza um *full scan* devido à ausência de um índice que permita a busca direta pelos valores necessários durante a junção com a tabela **GAME**. Isso força o otimizador a procurar em todas as tuplas da tabela para atender a condição da junção (**g.id_genre** = **ge.id_genre**).

Diante dessas informações do plano de execução será criado um índice adicional para tabela **GAME** para o atributo **na_sales** utilizado na cláusula WHERE e no ORDER BY da consulta.

CREATE INDEX na_sales_IDX ON GAME (na_sales);

Depois da criação deste novo objeto no SGBD MySQL está sendo executada a solicitação de explicação novamente na mesma consulta (execute novamente o *EXPLAIN* somente, conforme demonstrado anteriormente).

Apresentando o *Result Grid* da nova execução do *EXPLAIN*:

id	select_type	table	partitions	type	possible_keys	key	key_len	ref	rows	filtered	Extra
1	SIMPLE	ge	NULL	eq_ref	PRIMARY	PRIMARY	4	NULL	1	100.00	NULL
1	SIMPLE	g	NULL	range	GAME_GENRE_FK, na_sales_idx	na_sales_idx	3	NULL	757	100.00	Using index condition; Using where; Backward index scan

Apresentando o resultado da nova execução do *ANALYZE*:

-> Nested loop inner join (**cost=606** rows=757) (actual time=0.54..21.5 rows=757 loops=1)

-> Filter: (g.id_genre is not null) (**cost=341** rows=757) (actual time=0.512..7.96 rows=757 loops=1)

-> Index range scan on g using idxNa_sales over (1.00 < na_sales) (reverse), with index condition: (g.na_sales > 1.00) (**cost=341** rows=757) (actual time=0.502..4.95 rows=757 loops=1)

-> Single-row index lookup on ge using PRIMARY (id_genre=g.id_genre) (**cost=0.25** rows=1) (actual time=0.0063..0.00821 rows=1 loops=757)

Após a criação do índice na_sales_idx na tabela GAME, foram obtidas melhores nos custos (cost) do plano de execução da consulta. Esse índice adicional permitiu filtrar diretamente as tuplas da tabela GAME que atendiam a condição da cláusula WHERE, eliminando o custo do table scan na tabela GENRE , além de reduzir o número de tuplas processadas nas etapas subsequentes.

O custo da junção (*Nested loop inner join)* diminui de 2063 para 606 e o tempo real reduziu de 56,3ms para 21,5ms. Isso ocorreu principalmente porque a quantidade de tuplas da GAME que foram recuperadas pela cláusula WHERE eram menores (apenas as que atendiam a condição de na_sales > 1), estabelecendo um número de combinações a serem avaliadas na junção menores que a situação anterior. Assim, o otimizador trabalha com uma quantidade de dados menor, o que reduz o tempo total para se processar a junção.

No novo plano na etapa *Single-row index lookup on ge using* PRIMARY é possível notar que agora a junção realiza uma busca indexada na tabela **GENRE**, para cada tupla da tabela **GAME**. Isso é possível porque a filtragem eficiente em **GAME** reduz o número de buscas necessárias, tornando mais eficiente sua execução.

Dessa forma, a consulta inicialmente proposta foi mantida em sua escrita, mas um novo objeto no SGBD foi criado para tornar o desempenho dessa consulta SQL mais eficiente.

2) Segunda Demanda

Liste o nome dos jogos e o nome de seus respectivos publicadores, para os jogos que possuem a pontuação de usuário (*user_score*) igual a 9 (nove).

Proposta de solução inicial (elabore a consulta que resolva esta demanda):

SELECT ...

Solicitando explicação sobre a solução inicial (*explain*):

EXPLAIN SELECT

Apresentando a tabela e interpretando o *Result Grid* do *EXPLAIN*:

Solicitando a análise da solução inicial (*analyze*):

EXPLAIN ANALYZE SELECT

Apresentando e interpretando o resultado do *ANALYZE*:

-> Sort...

Explicar os ajustes realizados para melhoria da consulta inicial da tarefa 2:

==> É importante esclarecer que além da explicação cada comando SQL usado ou alteração da consulta SQL proposta inicialmente também deverá fazer parte deste item da evolução da consulta SQL em questão que você propôs para esta demanda.

Apresentando o *Result Grid* da nova execução do *EXPLAIN*:

Apresentando o resultado da nova execução do *ANALYZE*:

Apresentação da análise final a partir da consulta inicial proposta e quais são os principais indicativos de que a consulta ficou mais eficiente:

3) Terceira Demanda

Liste o nome dos jogos, o ano de lançamento e a descrição do gênero, considerando apenas os jogos que possuem um gênero associado, que foram lançados entre os anos de 2001 e 2012 e cujo nome começa com a letra "M".

Proposta de solução inicial (elabore a consulta que resolva esta nova demanda):

```
SELECT ...
```

Solicitando explicação sobre a solução inicial (*explain*):

```
EXPLAIN SELECT
```

Apresentando a tabela e interpretando o *Result Grid* do *EXPLAIN*:

Solicitando a análise da solução inicial (*analyze*):

```
EXPLAIN ANALYZE SELECT
```

Apresentando e interpretando o resultado do *ANALYZE*:

-> Sort...

Explicar os ajustes realizados para melhoria da consulta inicial da tarefa 3:

==> É importante esclarecer que além da explicação cada comando SQL usado ou alteração da consulta SQL proposta inicialmente também deverá fazer parte deste item da evolução da consulta SQL em questão que você propôs para esta demanda.

Apresentando o *Result Grid* da nova execução do *EXPLAIN*:

Apresentando o resultado da nova execução do *ANALYZE*:

Apresentação da análise final a partir da consulta inicial proposta e quais são os principais indicativos de que a consulta ficou mais eficiente:

4) Quarta Demanda

Listar o máximo do valor de venda de um jogo no Japão por categoria em que a nota dos críticos (*critic_score*) seja maior que 9 e tenha mais de 50 avaliações de críticos (*critic_count*).

Proposta de solução inicial (elabore a consulta que resolva esta nova demanda):

```
SELECT ...
```

Solicitando explicação sobre a solução inicial (*explain*):

```
EXPLAIN SELECT
```

Apresentando a tabela e interpretando o *Result Grid* do *EXPLAIN*:

Solicitando a análise da solução inicial (*analyze*):

```
EXPLAIN ANALYZE SELECT
```

Apresentando e interpretando o resultado do *ANALYZE*:

-> Sort...

Explicar os ajustes realizados para melhoria da consulta inicial da tarefa 4:

==> É importante esclarecer que além da explicação cada comando SQL usado ou alteração da consulta SQL proposta inicialmente também deverá fazer parte deste item da evolução da consulta SQL em questão que você propôs para esta demanda.

Apresentando o *Result Grid* da nova execução do *EXPLAIN*:

Apresentando o resultado da nova execução do *ANALYZE*:

Apresentação da análise final a partir da consulta inicial proposta e quais são os principais indicativos de que a consulta ficou mais eficiente: