

Aula 5

Sistema Gerenciador de Banco de Dados

Prof. Leonel da Rocha

1

Conversa Inicial

2

Performance de um SGBD

- Nesta aula trataremos assuntos importantes para o planejamento e a criação de um banco de dados que impacta positivamente a performance de utilização
- Estudaremos a normalização de dados, que trata da organização de um banco de dados visando reduzir a redundância de dados e aumentar a integridade

3

- Trataremos do controle de concorrência que organiza a utilização simultânea por usuários em um mesmo objeto, prevenindo inconsistências nos dados
- Otimização de consultas será outro tema abordado, que tratará como o acesso aos dados deve ser implementado, privilegiando a rapidez e a qualidade dos resultados obtidos

4

- Veremos as estatísticas de consultas, que é uma ferramenta que permite verificar como a consulta está se comportando em relação a sua performance
- Utilização de índices, que são itens importantes na classificação de dados e na melhora de velocidade de acesso a eles

5

Normalização de dados

6

- Normalização de banco de dados é um conjunto de regras que tem por objetivo organizar um projeto de banco de dados para reduzir a redundância e aumentar a integridade e o desempenho
- Para normalizar um banco de dados, serão examinados os atributos de uma entidade e as relações entre as entidades, para evitar anomalias observadas na inclusão, exclusão e alteração dos dados

7

- Para aplicar a normalização é necessário avaliar as entidades e seus atributos tendo como base cinco regras de normalização, chamadas de formas normais
- Pensando na transformação das entidades e seus atributos, em tabelas e colunas, a normalização corresponde a um conjunto de regras de simplificação e adequação dessas tabelas

8

- Veremos a seguir as principais características das três primeiras formas normais
- A normalização nada mais é do que a aplicação de determinadas condições a tabelas de um banco de dados
- Para a maioria dos projetos implementados, podemos considerar que as bases de dados estão normalizadas se aderirem à terceira forma normal

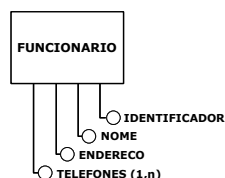
9

- Para dar início ao processo, precisamos definir todos os atributos relacionados a uma entidade principal, sendo que um deles deve ser um atributo identificador
- Depois desta parte inicial, partimos para a análise do documento de acordo com as formas normais descritas a seguir

10

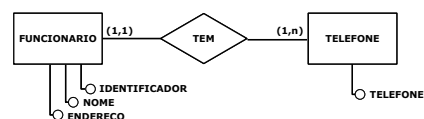
Normalização de dados – 1FN

- Primeira forma normal ou 1FN: nesta forma os atributos precisam ser atômicos, ou seja, as tabelas não podem ter valores repetidos nem atributos que possuam mais de um valor



11

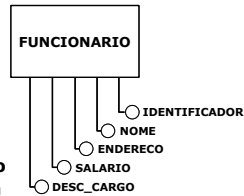
- Para normalizar, será preciso identificar o atributo identificador e o atributo que possui dados repetidos e removê-los da entidade FUNCIONARIO. Devemos então criar uma outra entidade com o atributo em questão, não nos esquecendo de fazer uma relação entre as duas entidades



12

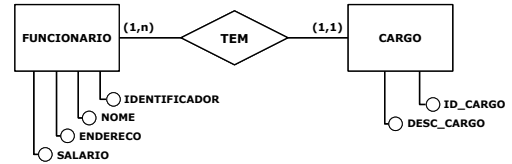
Normalização de dados – 2FN

- Lembrando que para estar na 2FN é preciso estar também na 1FN. 2FN define que os atributos normais, ou seja, os que não são identificadores, deverão depender unicamente do atributo identificador da entidade



13

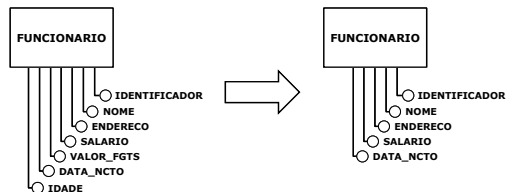
- Para normalizar, é necessário identificar os dados não dependentes da chave primária e removê-los, criando uma nova entidade com os dados em questão



14

Normalização de dados – 3FN

- Assim como para estar na 2FN é preciso estar na 1FN, para estar na 3FN é preciso estar na 2FN
- A 3FN define que todos os atributos dessa entidade devem ser funcionalmente independentes uns dos outros, ao mesmo tempo que devem ser dependentes exclusivamente da chave primária da tabela
- A 3FN foi projetada para melhorar o desempenho de processamento dos bancos de dados e minimizar os custos de armazenamento



15

16

Controle de concorrência

- Um banco de dados, quando é utilizado por mais de um usuário, terá que administrar a concorrência entre as informações que estão sendo acessadas pelos usuários
- A concorrência se dá quando, em uma transação, usuários tentam acessar a mesma informação simultaneamente

17

18

- Transações são todas as operações executadas entre o início e o fim de uma operação de manutenção de dados – insert, delete, update e select
- Para gerenciar as transações é necessário conhecer as propriedades comumente chamadas de ACID (acrônimo de Atomicidade, Consistência, Isolamento e Durabilidade)

19

- A Atomicidade é o princípio de que uma transação é uma unidade de processamento atômica, ou seja, a transação deve ser realizada por completo ou ela não deve ser realizada de forma alguma
- A preservação da Consistência permite que uma transação seja executada do início ao fim, sem que haja a interferência de outras transações durante sua execução

20

- O Isolamento, por sua vez, garante que uma transação seja isolada de outras transações, mesmo tendo várias transações sendo executadas simultaneamente
- A Durabilidade ou Permanência é a garantia de que as mudanças que ocorreram no banco de dados, ao término de uma transação finalizada com sucesso, persistam no banco

21

- Para implementar o controle de concorrência podemos utilizar a técnica de bloqueio em duas fases, que é baseada no bloqueio de itens de dados – chamamos de bloqueio uma variável que fica atrelada ao item de dados
 - Este bloqueio é binário, possuindo dois estados: bloqueado ou não bloqueado
- Esse tipo de procedimento permite que o item de dado só esteja acessível para uma transação se a variável não estiver bloqueada

22

Otimização de consultas

23

- Para a otimização de um SGBD é necessário identificar e eliminar os possíveis problemas de desempenho existentes em todos os níveis do sistema, incluindo as consultas
- Também é necessário otimizar as configurações do SGBD, do sistema operacional e do hardware

24

- Durante o projeto da base de dados é preciso conhecer os tipos de consultas que serão mais comuns no sistema e criar a base de dados de uma maneira que o processo de extração de informações seja otimizado
- Monitorar as consultas já implementadas para descobrir se existe lentidão e, se necessário, tratá-las

25

- Uma vez eliminados os problemas relativos às consultas SQL, será possível modificar as configurações do MySQL de forma a fazer um uso mais apropriado dos recursos disponíveis do sistema operacional e do hardware
- Para este processo é preciso conhecer como o MySQL funciona em relação à utilização de memória e disco, além dos seus principais parâmetros de configuração relacionados ao desempenho

26

- O MySQL possui um comando chamado slow log, que registra todas as consultas em que o tempo de execução seja maior que o parâmetro long-query-time, que por padrão é 10 segundos
- Também é possível configurar este log para registrar as consultas que não utilizam índices ou que realizam consultas com SELECT*

27

- Para finalizar, precisamos monitorar o desempenho do sistema operacional onde o MySQL está instalado
- No sistema operacional podemos utilizar recursos mais apropriados para o banco, tais como sistema de arquivos mais eficientes, processos e threads nativas, além da própria escolha de um sistema operacional mais apropriado ao MySQL

28

Estatísticas de consulta

29

- Para otimizar o desempenho de um SGBD é necessário realizar as rotinas de forma eficiente e eficaz, principalmente em grandes coleções de dados
- A eficiência se refere à execução de consultas e alterações realizadas nos dados armazenados
- A eficácia se refere ao bom funcionamento e atendimento às necessidades do indivíduo que utiliza um SGBD

30

- Para um bom desempenho das consultas de um banco de dados é importante que o projeto dele seja bem desenvolvido
- Após a sua implementação é preciso realizar um acompanhamento operacional para averiguar o desempenho – ajustes devem ser implementados, caso necessário

31

- Database tuning ou simplesmente tuning é o processo de revisão da base de dados por meio da coleta de estatísticas sobre os padrões de seleção
- A utilização dos recursos, bem como o processamento interno do SGBD, pode ser monitorado para revelar gargalos, tais como a disputa pelos mesmos dados por vários usuários ou processos de forma simultânea

32

- Os objetivos do tuning são:
 - Fazer com que as aplicações sejam executadas da forma mais rápida possível
 - Baixar o tempo de resposta das consultas e transações
 - Otimizar o desempenho geral das transações
- A revisão das decisões de projeto na fase tuning consiste em um ajuste do projeto
- As informações de entrada para o processo de tuning incluem estatísticas das consultas

33

- É possível concluir que no momento de escolher um SGBD para uma aplicação é necessário também saber quais recursos serão consumidos e o que o SGBD disponibiliza para otimizar o desempenho
- As estatísticas de consulta servem para mapear de maneira simples e direta os recursos utilizados, seja do sistema operacional, do hardware ou do SGBD

34

Otimizando um DB no MySQL utilizando índices

35

- Índices em um SGBD são importantíssimos quando o assunto é a otimização de desempenho
- É através dos índices que os dados são arranjados em uma tabela e sua recuperação é muito mais rápida

36

- Para criar um índice é necessário conhecer o sistema que está acessando o banco de dados e quais são as consultas que ele faz
- Sem esse conhecimento é impossível melhorar o desempenho, e ainda se corre o risco de ter uma piora no acesso aos dados

37

- Um SGBD acessa dados de dois modos:
 - Table scan, que varre todas as páginas começando do início da tabela e extraindo o pedido da query
 - Através de índices, que podem ser clusterizados e não clusterizados

38

- Em um índice clusterizado os dados são fisicamente ordenados de forma ascendente, por isso é possível ter apenas um índice clusterizado por tabela
- A ordenação física dos dados assim como a do índice é a mesma
- O índice cluster é utilizado em colunas muito acessadas por range de valores ou acessadas por ordenações

39

- Quando o índice é não clusterizado a construção é feita no heap (amontoado) – o SQL usa as linhas como identificadores de índices (a linha identifica o local da informação de armazenamento de dados)
- Chama-se heap porque na sua tradução significa amontoado, uma vez que os dados não têm uma ordem lógica e são gravados nas páginas que têm espaço disponível

40

▪ Exemplo de índices cluster e não cluster

ID	NOME	PONTEIRO	DATA_NCTO	SALARIO	ADMISSAO
1	Leonelson Rocha	2	17/07/1976	5.000,00	11/01/2000
2	Ana Silva	1	28/08/1974	8.000,00	25/02/2005
3	Mara Souza	3	21/11/1991	7.000,00	30/04/2010

Cluster

Não cluster
(Coluna oculta para o usuário)

41