

Materiais:

Q: Qual o link para os materiais do professor Herwin?

R: O material de apoio para atividades sobre Bancos de Dados de chave-valor consta no ícone de folha a4 ao lado do título da disciplina, os arquivos das práticas do professor Herwin estão em uma página do Dropbox, é possível fazer o download e executar. Os arquivos solicitados:

- 1.Exemplo de Chaves de Acesso: NoSQL_BD_ChaveValor_AccessKeys.xlsx
- 2.Comandos do AWS CLI: NoSQL_BD_ChaveValor_AWS_CLI.pdf
- 3.Dados do Exercício: NoSQL_BD_ChaveValor_Dados.txt
- 4.Manual do PartiQL: NoSQL_BD_ChaveValor_PartiQL.pdf
- 5.Script em Linguagem R para o Exercício: NoSQL_BD_ChaveValor_Scripts.R
- 6.Slides das Aulas Teóricas: NoSQL_BD_ChaveValor_Slides.pdf
- 7.Vídeo da Aula Prática: NoSQL_BD_ChaveValor_Pratica.mp4

Encontram-se na folha a4 - Aula 1 - Material Complementar (BDChave-Valor).

Os arquivos das práticas sobre Bancos de Dados orientado a Grafos do professor Herwin estão em uma página do Dropbox, você poderá fazer o download e executar. Os arquivos solicitados encontram-se na folha a4 - Aula 2 - BDGrafo.

O material sobre banco de dados orientado a Documentos, ainda não foi disponibilizado pelo professor Herwin por uma falta de atenção. Estamos em contato com ele para poder disponibilizar o mais rápido possível.

Ferramentas:

Estou encontrando um erro ao rodar o código do "shortest path". Neo4j retorna o erro "There is no procedure with the name gds.alpha.kShortestPaths.stream registered for this database instance."

Este erro se dá pois ao fazer a instalação do Neo4j a Local DBMS padrão, ele cria sempre a versão mais nova baixada. É necessário então setar a versão utilizada pelo professor Herwin (4.2.1).

Conceitos e Exercícios:

Q: Supondo que, ao passarmos do modelo conceitual para o modelo lógico, optamos por gerar uma tabela própria para um relacionamento; esta nova tabela automaticamente terá pelo menos duas colunas chave estrangeira se referindo a chave primária da tabela "referidora" e a chave primária da tabela referida.

Estas duas colunas chaves estrangeiras da nova tabela são necessariamente chave primária desta nova tabela?

R: Assumindo que o relacionamento seja muitos-para-muitos (também chamado de N-N), a técnica usual de mapeamento é a geração de uma tabela para esse relacionamento contendo duas chaves estrangeiras, uma para cada entidade envolvida no relacionamento.

Suponha que tenhamos o exemplo da relação "atuação" (N-N) entre uma entidade "engenheiro" e uma entidade "projeto". Agora vamos supor que temos {e1, e2, e3} como engenheiros e {p1, p2} como projetos. Os seguintes pares podem representar a relação "atuação": (e1, p1), (e1, p2), (e2, p1). Observe que precisamos que cada par seja único (por exemplo, o engenheiro e1 não pode aparecer duas vezes atuando no projeto p1), logo a chave primária será uma chave composta pelas colunas de ambas as chaves estrangeiras que referenciam o engenheiro e o projeto.

Se assumirmos que seja um relacionamento um-para-muitos (1-N), a opção de usar uma tabela própria deve ser analisada caso a caso e verificar o comportamento das chaves. Suponha que tenhamos o exemplo da relação "financia" entre uma entidade "financeira" e várias entidades "venda". Agora vamos supor que tenhamos {f1, f2} como financeiras e {v1, v2, v3, v4} como vendas. Os seguintes pares podem representar a relação "financia": (f1, v1), (f1, v2), (f2, v3).

Observe que também neste caso cada venda somente apareceria no máximo uma única vez no relacionamento e, portanto, poderia ser utilizado somente a chave estrangeira para a entidade "venda" como uma chave primária conforme sua sugestão.

Q: Referente à tabela de pré-requisitos, (slide 62 da apresentação de apoio).

As colunas `cod_depto` e `num_disc`, que são chaves estrangeiras e referenciam colunas homônimas na tabela `disciplina`, formam juntas a chave primária da própria tabela `prereq`.

Entretanto, como pode-se esperar que uma disciplina tenha mais de um pré-requisito, de fato observamos ocorrências duplas, por ex., de INF01 e 109 (bem como INF01, 108).

Isto não é uma quebra de integridade de chave? Uma vez que há mais de uma linha com os mesmos valores nas colunas chave-primária da tabela `prereq`.

R: A imagem do slide 62 deveria estar pintada de vermelho todas as quatro colunas da tabela "prereq" de modo a indicar que a chave primária seria composta das quatro colunas. Logo, esta observação de que apenas as colunas "cod_depto" e "num_disc" são insuficientes como chave primária está corretíssima.

Q: Ao marcar com o X os dados que foram excluídos, eles deixam de ser acessados, ou eu poderia utilizá-los como dados históricos?

Por exemplo em uma mudança de endereço, é marcado com o X o antigo endereço, mas se quisermos saber todos os endereços de um cliente é possível extrair essa informação baseado no histórico de alterações neste campo?

R: No Cassandra não tem como consultar um dado deletado. Ele cria o *tombstone* que indica que o dado foi deletado, depois de um tempo ele faz compactação e deleta fisicamente para manter a base mais enxuta. Nesse meio tempo em que não ocorre a deleção fisicamente, ele mantém todos os logs de comandos, mas na busca ele retorna sempre a versão mais recente.

Teria que verificar se tem como desabilitar a compactação, mas acabaria perdendo performance até ficar uma base muito inchada. Para utilizar como histórico, seria melhor a criação de um campo de *deleted* (*boolean*).

Q: Qual a relação possível em um grafo do tipo multigrafo?

R: O professor Herwin é claro em sua explicação durante a parte 01 da aula 02, ao minuto 21 quando explica os tipos de grafos. Quando dizemos multigrafo nos referimos aos casos em que o vértice pode ter mais de 1 saída ou entrada. Quando nos referimos a grafo, aí sim entram os autos relacionamentos. A mesma questão também se encontra elucidada no livro da disciplina, na página 16, parte 1 da aula 2

Q: Fiquei com uma dúvida referente a chave artificial que o professor explicou na última parte da aula 04. Quando crio uma chave artificial eu deixo de utilizar a composição de chaves que haviam dispostos no modelo anterior com um conjunto de colunas que mantinha a integridade tanto dos dados quanto das associações entre as tabelas. Não perco esta referência? Ou seja, não corro o risco de ter conteúdo duplicado nas outras colunas?

R: Como o professor Júlio indicou, não há nenhum demérito em ter as chaves primárias compostas por várias partes (colunas) mas, para simplificar a leitura das chaves e relacionamentos podemos utilizar\criar uma coluna "extra" para ser apenas a chave primária que geralmente é um identificador único. Mas note que ele indica que muitas vezes sabe que as demais colunas também permanecem candidatas e que de acordo com o negócio, ele sabe que a composição permanece única. Essa chave primária artificial é utilizada para colocar de forma genérica a chave primária de uma entidade, pois uma chave natural pode se tornar muito complexa, impactando em termos de performance no banco de dados. Mas usando um atributo artificial na chave primária podemos vir a repetir várias vezes as mesmas colunas, mas não seria possível ter o registro (linha) todo repetido, pois essa chave primária será auto-incremental.

Q: Gostaria de saber mais sobre como se enquadra o "CockroachDB", já que ele é tanto um NoSQL como também um SQL transacional.

R: O CockroachDB é um banco de dados SQL que pode escalar horizontalmente. Se você precisa para suportar maiores volumes, você pode simplesmente adicionar mais máquinas. Com CockroachDB, os servidores replicam automaticamente e reequilibrar-se, que mantém o seu sistema disponível à medida que cresce e evolui. Ele busca oferecer o mesmo desempenho escalável dos bancos NoSQL, para cargas de trabalho OLTP read-write, mantendo as garantias das propriedades ACID em suas transações.

Q: Gostaria de saber se o Neo4j é adequado para modelos em redes de Petri. No caso das redes de Petri é preciso que se insira uma ficha (valor da entidade e uma transição), os arcos podem conter valor. Acha que é possível?

R: Sim, o Neo4J suporta em sua linguagem de modelagem e consulta a estrutura de grafos associada a uma Rede de Petri. Claro que estamos falando no modelo estático, sem a "execução" da rede, o que seria feito externamente ao banco de dados. Aqui há algumas referências que mencionam o uso do Neo4J para armazenamento de dados associados a Redes de Petri: <https://www.hindawi.com/journals/misy/2021/5988742/>
https://mountainscholar.org/bitstream/handle/10217/195323/Basnet_colostate_0053N_15363.pdf

Q: É melhor a criação de índices simples ou compostos? Qual tem melhor performance no sgbd? Ex: Tenho uma tabela PESSOA (codigo, email, nome). Nesse caso sendo todos os 3 campos índices. É melhor fazer um índice composto com os 3 campos ou fazer 3 índices separados, ou tanto faz?

R: O índice irá ajudar a engine do BD a dividir os dados para procurar o mais rápido possível. Caso haja um índice muito específico ele irá percorrer toda a base de dados, implicando no tempo e latência de processamento.

Q: Sobre grafos, qual seria a diferença do short path para o minimum spanning tree?

R: Um spanning tree de um grafo não direcionado G é um subgrafo conectado que cobre todos os nós do grafo com o menor número possível de arestas. Em geral, um grafo pode ter mais de uma árvore geradora. Se o grafo for ponderado por arestas, podemos definir o peso de uma árvore geradora como a soma dos pesos de todas as suas arestas. Uma minimum spanning tree é uma árvore geradora cujo peso é o menor entre todas as árvores geradoras possíveis. No problema de shortest path tree, começamos com um nó de origem s. Para qualquer outro nó v no grafo G, o caminho mais curto entre s e v é um caminho tal que o peso total das arestas ao longo desse caminho é minimizado. Portanto, o objetivo do problema da árvore do caminho mais

curto é encontrar uma árvore geradora tal que o caminho do nó fonte s para qualquer outro nó v seja o mais curto em G.

Q: Sobre a ótica da LGPD, qual tem sido a melhor forma de captar dados (variados) sendo que a Legislação que protege os dados pessoais diz que esta captação deve ser expressa por parte daquele que cede as informações pessoais, ou seja, devemos indicar expressamente a um órgão regulador (como a ANPD) em uma eventual auditoria como obtivemos aqueles dados pessoais (identificado ou identificável)?

R: Sob a ótica da LGPD, qualquer sistema de informação que armazene tipos de dados associados a informações pessoais deverá ter na sua interface gráfica de usuário a notificação expressa de que o usuário concorda com os termos de coleta de informação. Isso é usualmente feito através de uma caixa de seleção. Claro que nesses casos, as normas devem estar disponíveis para o acesso pelo usuário.

Q: No minuto 26.17 o professor Herwin comenta sobre gerenciadores NoSQL e as propriedades ACID. Pelo que há na documentacao, existem gerenciadores noSQL como o MongoDB que tem ACID. Segue o link: <https://www.mongodb.com/basics/acid-transactions>

R: O intuito do prof. Herwin era ser um pouco mais abrangente sobre o assunto. Como podes ver o MongoDB possui essa possibilidade mas como a própria documentação indica, de 80 a 90% dos casos não devemos usar o ACID para BD orientada a dados: "We estimate that 80%-90% of applications that leverage the document model will not need to utilize transactions in MongoDB. For the other 10%-20% of applications, multi-document transactions are available." De qualquer forma, obrigado pelo alerta e pelo compartilhamento deste conhecimento.

**FAQ gerado com base em comentários até o dia 30/11/2022.*