

5.1 - O que é um banco de dados: diferença de bancos relacionais e não relacionais

00:00:05:29 - 00:00:25:05

Olá pessoal! Bem-vindes a mais um módulo do curso de Análise de Dados. Galera, pra deixar a nossa análise ainda mais completa, nós vamos puxar alguns dados externos sobre os municípios dos participantes para a gente entender se, por exemplo, nos municípios que tem um índice de educação maior, temos mais gestores em dados ou se nos municípios de maior renda os salários são maiores.

00:00:25:24 - 00:01:00:19

Assim, conseguimos entender se a desigualdade salarial é uma coisa regional ou se é maior mesmo nos critérios que já analisamos. Legal, né? Bom, hoje no Brasil nós temos mais de 5000 municípios e vamos pensar que esses índices de renda, educação, saúde, entre outros, sejam calculados anualmente. É bastante informação para armazenar, não é mesmo? Além disso, imagine que, além dessa atualização anual dos índices, também sejam armazenados dados com atualizações mais frequentes, como custos mensais dos municípios, funcionários públicos, eventos regionais, entre outros.

00:01:01:00 - 00:01:31:14

São informações que podem ser atualizadas, apagadas, corrigidas etc. Para isso, é muito comum que a gente utilize uma estrutura de armazenamento de dados mais complexos do que um arquivo Excel ou CSV, que é um banco de dados. Para nós, que somos analistas de dados, é muito comum que a gente consulte esse tipo de estrutura no nosso dia a dia, pois os dados de empresas costumam ser armazenados nelas e não em arquivos comuns. Mas bora lá aprender de fato o que é esse banco de dados.

00:01:32:18 - 00:01:56:11

Imagine que a pesquisa sobre as pessoas que trabalham em dados ganhasse uma nova versão em 2024. Aí a gente ia querer poder comparar as duas versões para ver o que mudou e o que continua igual. Porém, para isso, a gente ia precisar inserir novos dados às pessoas que só responderão em 2024, atualizar registros existentes, por exemplo os erros de formulário agora seriam corrigidos, e excluir registros indesejados.

00:01:56:24 - 00:02:17:25

Só que fazer isso com um monte de planilha fica difícil. É para isso que usamos o banco de dados. Um banco de dados é um sistema projetado para organizar, armazenar e gerenciar dados de maneira eficiente, facilitando o acesso à gestão e a atualização dessas informações. É como um armário digital, onde você guarda informações de modo organizado.

00:02:18:08 - 00:02:45:04

Isso permite que você encontre, use e atualize esses dados facilmente quando precisar. Uma das formas de organizar esse banco é com tudo bem estruturado: cada coisa é uma tabela e as tabelas se relacionam. Nesse caso, chamamos de base de dados relacional. Um exemplo mais simples seria assim: vamos supor que eu, eu aqui no cantinho, na tela, tenha uma loja de CDs e eu preciso de várias informações para tomar decisões.

00:02:45:14 - 00:03:03:26

Por exemplo, será que eu ainda tenho estoque de algum álbum ou preciso de mais? Quem é a funcionária do mês? O que o cliente costuma comprar para eu poder fazer uma oferta personalizada para esse cliente? Então eu quero armazenar informações sobre as funcionárias da minha empresa. Assim, eu criei uma tabela chamada Funcionárias no meu banco de dados.

00:03:04:10 - 00:03:35:16

Essa tabela é a nossa entidade que armazena os dados dos funcionários. Cada entidade é geralmente representada por uma tabela em um banco de dados relacional. Cada linha da tabela representa uma instância ou ocorrência dessa entidade, e cada coluna da tabela representa um atributo ou característica dessa entidade. Um atributo é uma característica específica, uma propriedade que descreve uma entidade. Olhando nossa entidade funcionários, temos atributos como ID_F, nome e salário.

00:03:36:00 - 00:04:03:01

Temos duas ocorrências, que são as funcionárias Stefani e Jéssica. O ID_F, que é o id do meu funcionário, cada funcionário tem um id único, então a Stefani tem um, a Jéssica tem outro. Assim eu posso filtrar pelo id de cada uma. Coloquei esses dados, mas é claro que posso ter muitos outros de cada funcionário, como data de nascimento para oferecer um sanduíche ou um suco de aniversário e tal, podemos ter a quantidade de vendas por mês, a marcação das próximas férias, enfim.

00:04:03:16 - 00:04:41:13

Então, essa tabela eu já consigo gerenciar meus funcionários muito bem. Além dos funcionários, eu tenho uma tabela, uma entidade chamada "álbuns". Aqui na minha tabela eu coloquei atributos como o nome dos álbuns (que chama produto, no caso), o valor desse álbum e o ID_P, que é o id de cada álbum, que é único. Podemos colocar muitas outras colunas com informações desses produtos, como a quantidade dele no estoque e também eu tenho uma tabela aqui dos meus amados clientes, também tem o ID_C, que é o id único para cada cliente, o nome dele e o id do produto que ele adquiriu.

00:04:41:24 - 00:05:05:08

Tendo essa coluna de id de produto, podemos dizer que essa tabela se relaciona com a tabela de álbuns, pois é através do id do produto, o ID_P, nós conseguimos saber qual o nome do produto, qual o nome do álbum na outra tabela. Por exemplo, aqui a gente tem a cliente Ana, a cliente Ana adquiriu o ID_P 2.

00:05:05:25 - 00:05:40:20

Qual que é o ID_P 2? Se a gente for na tabela de álbuns, a gente vê que o ID_P 2 é o Baby Gal, que é o álbum da Gal Costa. Ok? E por fim, a gente tem a tabela de vendas. Observem que, sem saber o contexto, essa tabela parece apenas um tanto de números soltos. Mas nós sabemos que essa tabela representa as vendas da minha loja e que a coluna ID_C serve para que a gente relacione com a tabela de vendas, que é dos meus clientes, com a tabela de álbuns também, aqui do meu ID_P e com a tabela de funcionários com o meu ID_F.

00:05:40:29 - 00:06:03:16

Então aqui nessa tabela de vendas, eu consigo saber qual funcionário vendeu o quê para qual cliente, ok? Então, através da tabela de vendas, a gente consegue relacionar com as outras tabelas e ter informações de qual funcionário realizou as vendas. Eu posso ter outras entidades, ou seja, tabelas para gerenciar melhor o conteúdo da minha loja, como por exemplo, uma tabela de artistas.

00:06:03:16 - 00:06:28:15

Assim, consigo relacionar a tabela de artistas e álbum e saber quantos álbuns algum artista tem e por aí vai. Legal, né? Só com essas tabelas básicas eu já consigo gerenciar minha loja, viram? Esse é um exemplo de estrutura de banco de dados. De modo mais específico, uma base de dados relacional, que é um banco de dados que tem uma estrutura bem certinha, segue uma estrutura tabular, onde os dados são organizados em tabelas com linhas e colunas.

00:06:28:19 - 00:06:58:29

Esse tipo de banco de dados é baseado em relações entre as tabelas. Vamos aprender mais pra frente como fazer essas relações. Geralmente a linguagem utilizada é o SQL para consultar e manipular os dados. Existem alguns sistemas muito conhecidos para gerenciamento de banco de dados relacional, como o My SQL, o Postgre SQL e o Oracle. Agora vamos supor que você, ao invés de uma loja de disco, seja dono de um lojão de 10 reais em que você vende o que aparecer lucrativo e dê para cobrar 10 reais.

00:06:59:14 - 00:07:28:21

Isso inclui roupa, CD, brinquedo, eletrônico, chinelo, etc. Cada um desses itens tem atributos diferentes, certo? Uma roupa tem tamanho, cor, modelo, marca diferente. Um eletrônico tem voltagem, pilha, bateria, coisas assim. Um CD tem artista diferente, álbum, ano e por aí vai. Ao invés de ter uma tabela para cada tipo de produto que tem seus devidos atributos ou de ter uma tabela enorme com todos atributos possíveis,

00:07:29:00 - 00:07:52:11

você gostaria de ter uma tabela flexível que, para cada linha, poderia ter um tipo e só os atributos necessários. Para isso existem os bancos NoSQL, que é flexível e pode lidar com uma variedade de dados não estruturados e sem a necessidade de seguir um esquema fixo. Bom, os bancos de dados não relacionais podem ter estruturas mais flexíveis, como documentos, gráficos, etc.

00:07:52:25 - 00:08:16:14

Os bancos de dados não relacionais são frequentemente escolhidos para cenários de alta escalabilidade, onde é necessário lidar com grandes volumes de dados distribuídos em vários servidores. Diferente do banco relacional, aqui não temos uma linguagem de consulta padrão como o SQL. Cada sistema pode ter sua própria maneira de acessar e manipular os dados. Alguns exemplos são Mongo DB, o Cassandra, Neo4j, cada um com suas particularidades.

00:08:16:17 - 00:08:46:01

A gente tem o Redis também. Então vamos para alguns exemplos de aplicações reais de banco de dados, começando por bancos de dados relacionais. Temos os sistemas de gerenciamento de conteúdos como o WordPress e Drupal, que usam bancos de dados relacionais para armazenar conteúdo de páginas, postagens, usuários e configurações do site. Os sistemas de comércio eletrônico, como Shopify e Magento empregam bancos de dados relacionais para gerenciar inventário de produtos, detalhes do cliente, histórico de pedidos e transações.

00:08:46:03 - 00:09:11:24

Já os bancos não relacionais temos as aplicações de Big Data utilizando bancos de dados para armazenar e processar grandes volumes de dados, como registro de eventos em tempo real, análise de redes da web, etc. Redes sociais como o Facebook e o Twitter (X) fazem uso de bancos de dados não relacionais para lidar com uma grande quantidade de dados gerados por usuários, como postagens, comentários, curtidas e conexões entre usuários.

00:09:12:14 - 00:09:33:24

Muitos jogos online e móveis usam um banco de dados não relacional para armazenar informações de jogadores como perfis de usuário, progresso no jogo de dados de interação em tempo real entre os jogadores. Tem bancos mistos em que os cadastros e informações de progresso dos jogos ficam em relacionais e as interações em tempo real ficam em não relacionais.

00:09:34:13 - 00:09:56:17

Mas bom, uma dúvida que pode surgir é qual a diferença entre um banco de dados relacional para um arquivo CSV Excel, por exemplo. As principais diferenças envolvem estrutura e recursos. Um banco de dados tem uma estrutura mais complexa e oferece recursos avançados para manipular e gerenciar dados, enquanto um arquivo CSV é simplesmente uma lista de valores separados por vírgula.

00:09:57:00 - 00:10:19:28

Outra coisa é a capacidade. Um banco de dados pode lidar com grandes volumes de dados e oferecer consultas complexas, segurança avançada e controle de transações. Um banco de dados pode ter várias tabelas, por exemplo, sendo mais complexo que um CSV. Agora é até possível ter um CSV com muitos dados, mas vai ficando cada vez mais difícil de manipular e exigindo também muito do sistema computacional.

00:10:20:27 - 00:10:48:09

Resumindo, os bancos de dados relacionais são bons para cenários onde a estrutura dos dados é bem definida e os relacionamentos entre eles são importantes. Enquanto os bancos de dados não relacionais são mais adequados para situações onde a flexibilidade e escalabilidade são prioridades, especialmente quando estamos lidando com grandes volumes de dados não estruturados. No nosso curso, como estamos lidando com dados estruturados, vamos nos aprofundar em relacional.

00:10:48:24 - 00:11:01:00

Na próxima aula, vamos explorar mais essa questão do relacionamento das tabelas. Mas vamos deixar alguns materiais complementares para vocês conhecerem mais sobre os não relacionais. E é isso. Até a próxima!