Aula 6

1

Sistema Gerenciador de Banco de Dados

Prof. Leonel da Rocha

J |___

2

Nessa aula, veremos alguns assuntos extras relacionados a banco de dados

SGBD e o seu universo

Por exemplo os sistemas de banco de dados distribuídos (BDD), que é uma relação de nós ou servidores, onde cada um pode participar na execução de transações que acessam dados em um ou mais nós

SGBD e o seu universo

Trataremos do data Warehouse, que é um repositório central de dados, tendo por objetivo melhorar a análise dos dados

Conversa Inicial

Veremos o que é um data lake, que é um método para organizar grandes volumes de dados altamente diversificados de várias fontes

3 4

SGBD e o seu universo

- Trataremos do NoSQL, que é um padrão de armazenamento de dados alternativo ao SQL, oferecendo uma robustez e escalabilidade melhores
- Estudaremos sobre ciência de dados, ou data science, que trata da coleta de dados e da elaboração de interpretações significativas auxiliando no crescimento dos negócios
- Vem comigo e bons estudos

Banco de dados distribuídos

Banco de dados distribuídos

- Um sistema de banco de dados distribuído (BDD) é composto por uma relação de nós
- O banco de dados é armazenado em diversos computadores, que são os nós
- Esses nós comunicam-se entre si por intermédio de vários meios de comunicação

Banco de dados distribuídos



7

Banco de dados distribuídos

- Uma relação r, possui diversos enfoques para o armazenamento em BDD
- Replicação: o sistema gerencia réplicas idênticas da relação, onde cada réplica é armazenada em locais diferentes
- Fragmentação: a relação é particionada em vários fragmentos armazenados em locais diferentes
- Replicação e fragmentação: a relação é particionada em vários segmentos, e o sistema mantém diversas réplicas de cada fragmento

Banco de dados distribuídos

- Em um BDD, a transparência para leitura é mais fácil de se conseguir e manter do que a transparência para atualização
- O maior problema para a atualização é garantir que todas as réplicas e fragmentos sejam atualizados, após uma alteração em uma das réplicas ou fragmentos
- Ela deve ser replicada para todas as cópias, réplicas e fragmentos, existentes no sistema

9 10

BDD - Vantagens

- Confiabilidade e disponibilidade funcionamento do sistema mesmo com falhas, sem violar a consistência do banco de dados
- Autonomia local um departamento pode controlar seus dados
- Desempenho consultas podem ser subdivididas e realizadas paralelamente
- Disponibilidade uma falha afeta somente um fragmento e não o banco de dados inteiro

BDD – Desvantagens

- Complexidade os problemas são mais complexos do que em bancos de dados centralizados
- Implantação mais cara maior complexidade e uma infraestrutura mais extensa
- Falta de padrões falta de metodologias e ferramentas para converter um SGBD centralizado em SGBD distribuído

11 12

Data Warehouse

Data Warehouse

- Um data Warehouse é um repositório central de dados
- Os dados são gerados a partir das transações, dos bancos de dados relacionais, distribuídos e de diversas fontes para o data Warehouse
- Esses dados são acessados através de ferramentas de inteligência de negócios, ou Business Intelligence (BI), clientes SQL e aplicativos de análise

13 14

Data Warehouse

- O data Warehouse é fonte de consulta para relatórios gerenciais, painéis e ferramentas de análise
- Ele armazena dados de maneira eficiente, minimizando a entrada e saída, disponibilizando os resultados das consultas de uma forma rápida e eficiente
- É projetado especificamente para análises de dados, envolvendo a leitura de grandes volumes de dados

Data Warehouse

- ETL (extraction, transformation, loading) é um tipo de integração de dados, realizado em três etapas: extração, transformação e carregamento
- ETL pode ser utilizado para construir um data Warehouse. Nesse processo, os dados são retirados (extraídos) de um sistema-fonte, convertidos (transformados) em um formato que possa ser analisado e armazenados (carregados)

15 16

Data Warehouse

- A utilização de um data Warehouse apresenta os seguintes pontos positivos:
 - Dados consolidados de várias fontes
 - Análise de dados históricos
 - Qualidade, consistência e precisão de dados
 - Separação do processamento analítico dos bancos de dados transacionais

Data Warehouse

- Desvantagens na utilização de um data Warehouse:
- Perigo de legamarts: criação de data mart (DM) independentes, com foco em um único assunto, que podem não considerar a arquitetura de forma global. Desse modo, os DM's independentes transformam-se em DM legados ou legamarts

(...)

Data Warehouse

(...)

- Desafio de possuir a visão de empreendimento como um todo
- Administrar e coordenar múltiplas equipes e iniciativas
- A maldição de sucesso: atender a demanda de desenvolvimento de um data Warehouse

Data lake

19 20

Data lake

- Um data lake é um lugar para armazenar dados estruturados e não estruturados
- É um método para organizar grandes volumes de dados altamente diversificados e de várias fontes
- Está se tornando cada vez mais importante à medida que os usuários desejam executar exploração e descoberta de dados

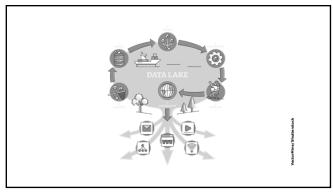
Data lake

A principal diferença entre um data lake e um data Warehouse é que o data lake tende a armazenar dados de uma forma muito rápida, preparando-os posteriormente à medida que os usuários fazem o acesso, ao contrário do data Warehouse, onde os dados são preparados com muito cuidado antes de armazená-los

21 22

Data lake

- Com o data lake, os usuários geralmente utilizam os dados na forma original, sem alterá-los
- Primeiro por razões de desempenho, depois pelo desejo de executar análises avançadas que dependem de dados originalmente detalhados



Data lake

- Uma solução adequada de um Data Lake deve ser capaz de oferecer melhores maneiras de:
 - Utilizar e transformar: mova e converta diferentes tipos e formatos de dados
 - Persistir e acessar: dados seguros, facilmente descobertos, facilmente dimensionados e que sejam acessados conforme necessidades
- Analisar e usar a ciência de dados, descobrindo insights e tendências nos dados

Data lake

- Soluções de data lake são aplicadas em várias atividades:
 - Marketing omnicanal, que é abordagem integrada de uma marca para cada ponto de contato com o cliente em todos os canais
 - Captura de dados através da utilização de um app em tempo real para monitoramento de uso do cliente
 - Apoio na cadeia de suprimentos digitais captados em arquivos de log muito grandes e altamente numerosos provenientes de servidores web

25 26

Data lake

- Criação de novas fontes de dados pela Internet das Coisas
- Captura de dados através de sensores em veículos de transporte, seja trem ou caminhão, para rastreamento da operação, consumo de combustível, rota, entre outros itens
- Grandes quantidades de informações estão vindo desses lugares, e o data lake fornece um repositório para todos esses dados

Data lake

- Outra abordagem para o data lake é a TI fornecer um único grande data lake multilocatário
- Ele pode ser utilizado por vários departamentos, unidades de negócios e programas de tecnologia
- À medida que as pessoas se acostumam com o data lake, elas descobrem como otimizá-lo para diversos usos e operações, análises e até mesmo conformidades

27 28

NoSQL

NoSQL

- NoSQL é um padrão de armazenamento de dados alternativo ao SQL, que oferece uma robustez e escalabilidade melhores
- O termo NoSQL foi primeiramente utilizado em 1998 como o nome de um banco de dados não relacional de código aberto
- Seu autor, Carlo Strozzi, alega que o movimento NoSQL "é completamente distinto do modelo relacional e, portanto, deveria ser mais apropriadamente chamado 'NoREL'"

29 30

NoSQL

- Com o crescimento da internet, novos dados foram surgindo e tratá-los foi se tornando gradualmente mais complexo e sua manutenção cada vez mais cara
- O nome, NoSQL, era uma tentativa de descrever o surgimento de um número crescente de bancos de dados não relacionais e fazia uma referência ao esquema de atribuição de nomes dos bancos de dados relacionais mais populares do mercado, como MySQL, MS SQL e PostgreSQL

NoSQL

- Bancos de dados relacionais escalam, mas quanto maior o tamanho, mais custoso se torna essa escalabilidade, seja pelo custo de novas máquinas, seja pela necessidade de mais especialistas nos bancos de dados utilizados
- Já os não relacionais possibilitam uma escalabilidade mais barata e menos complicada, pois não exigem máquinas superdimensionadas e sua facilidade de manutenção permite que um número menor de profissionais seja necessário para a manutenção do sistema

31 32

NoSQL

Os bancos de dados NoSQL vão ficando cada vez mais populares entre as grandes organizações, pois reúnem as características de poder trabalhar com dados semiestruturados ou não tratados vindos de diversas origens, tais como: arquivos de log, websites, arquivos multimídia, entre outros

NoSQL

A distribuição do NoSQL pode ser feita em escala global, o que se mostra interessante, pois as organizações, para atender seus usuários de forma eficiente, utilizam vários data centers, localizados em diversas partes. Com isso, uma série de questões sobre disponibilidade e performance são levantadas no desenvolvimento de sistemas

33

Ciência de dados

Ciência de dados

- Ciência de dados, ou data science, é o tema do momento entre profissionais da área de dados e organizações que se concentram na coleta de dados e na elaboração de interpretações significativas para auxiliar no crescimento dos negócios
- As informações são um ativo importante para qualquer setor, mas apenas têm valor se forem trabalhadas de forma eficiente

35

Ciência de dados

- As ferramentas tradicionais de negócios não permitem processar uma quantidade volumosa de dados não estruturados
- Por sua vez, a ciência de dados oferece soluções avançadas para analisar grandes volumes de dados provenientes de diferentes tipos de fontes, como transações financeiras, multimídia, marketing, sensores, instrumentos e arquivos de texto

Ciência de dados

É um campo do conhecimento que tem por objetivo explorar dados para aprender sobre um assunto específico e oferece inúmeras possibilidades de aplicação, em particular com inteligência artificial, aprendizado de máquina e processos de análise estatística

37 38

Ciência de dados

- A ciência de dados tem como objetivo principal melhorar os serviços e os produtos das organizações de modo que elas tenham uma vantagem competitiva real
- Exemplos da aplicação de data science: previsão do tempo – dados coletados de satélites, radares, navios e aeronaves, e a partir dos dados coletados, é possível construir modelos para prever o tempo e emitir alertas sobre possíveis calamidades naturais iminentes, com uma precisão muito grande

Ciência de dados

Já empresas ligadas à área da saúde estão utilizando data science para analisar dados de exames clínicos e auxiliar os médicos a fazerem diagnósticos precoces, permitindo que os pacientes sejam tratados com mais eficácia e, muitas vezes, minimizando o custo do tratamento

39 40

Ciência de dados

Companhias de logística utilizam esse campo do conhecimento para analisar o tráfego, condições climáticas, melhores rotas, além de outros fatores para melhorar as condições de entrega, reduzindo dessa maneira os custos operacionais

Ciência de dados

Os principais mecanismos de busca na internet utilizam a ciência de dados em conjunto com o aprendizado de máquina para encontrar o objeto de busca de forma mais precisa em frações de segundos. Essa rapidez nos motores de busca só é possível em virtude da utilização da ciência de dados

41 42

Ciência de dados

O conteúdo de marketing digital, na sua grande maioria, é escolhido por algoritmos que utilizam ciência de dados. Dessa maneira, as empresas obtêm resultados muito satisfatórios e melhores do que o marketing convencional, porque os anúncios são criados conforme o histórico do usuário

Ciência de dados

- Muitas empresas de e-commerce de viagens utilizam a ciência de dados para melhorar os resultados em seus motores de busca, mostrando para o cliente os resultados de pesquisas sobre hotéis ou voos
- Serviços complementares são sugeridos, como reserva de carros, pacotes de passeios, seguro de viagens e entretenimentos no destino