



AULA ATIVIDADE ALUNO

Curso:
CST em Análise e
Desenvolvimento de Sistemas

Disciplina: Programação e Desenvolvimento de Banco de Dados

Teleaula: 03

Título: Data Warehouses

Prezado(a) aluno(a),

A aula atividade tem a finalidade de promover o autoestudo das competências e conteúdos relacionados à Programação em Banco de Dados. Ela terá a duração de 1 hora.

Nesta aula, vamos entender um pouco mais sobre um tema muito relevante relacionado à banco de dados: os Data Warehouses. Para entender mais, divida a turma em grupos (ou individualmente) e cada grupo (aluno) escolhe um dos temas (partes) do artigo disponível:

1. O Que É um Data Warehouse?
2. Benefícios de um Data Warehouse
3. Arquitetura de Data Warehouse
4. A Evolução dos Data Warehouses
5. Planejamento de um Data Warehouse

Em seguida, transmitam aos demais grupos os principais pontos do tópico que estudaram! Pensem também nas aplicações de um data warehouse, assim como esta, para a classificação de *fakenews*. Você pode conferir o artigo em: <https://sol.sbc.org.br/index.php/erbd/article/view/8486> (Acesso: 24 Jul. 2020)

Conte sempre com a mediação do seu tutor!

Bons estudos!

Artigo: O que É um Data Warehouse?

“Um data warehouse é um tipo de sistema de gerenciamento de dados projetado para ativar e fornecer suporte às atividades de business intelligence (BI), especialmente a análise avançada. Os data warehouses destinam-se exclusivamente a realizar consultas e análises avançadas e geralmente contêm grandes quantidades de dados históricos. Os dados em um data warehouse geralmente são derivados de uma ampla variedade de fontes, como arquivos de log de aplicativos e aplicativos de transações.

Um data warehouse centraliza e consolida grandes quantidades de dados de várias fontes. Seus recursos analíticos permitem que as organizações obtenham informações de negócios úteis de seus dados para melhorar a tomada de decisões. Com o tempo, cria-se um registro histórico que pode ser inestimável para cientistas de dados e analistas de negócios. Devido a esses recursos, um data warehouse pode ser considerado a “única fonte verdadeira” da organização.

Um data warehouse típico geralmente inclui os seguintes elementos:

- Um banco de dados relacional para armazenar e gerenciar dados
- Uma solução de extração, carregamento e transformação (ELT) para preparar os dados para análise
- Análise estatística, relatórios e recursos de mineração de dados
- Ferramentas de análise de clientes para visualizar e apresentar dados aos usuários de negócios
- Outras aplicações analíticas mais sofisticadas que geram informações úteis por meio de aplicação de algoritmos de machine learning e inteligência artificial (IA)

Benefícios de um Data Warehouse

Os data warehouses oferecem o benefício único e abrangente de permitir que as organizações analisem grandes quantidades de dados variantes e extraiam um valor significativo, além de manter um registro histórico.

Quatro características únicas (descritas pelo cientista da computação William Inmon, que é considerado o pai do data warehouse) permitem que os data warehouses forneçam esse benefício abrangente. De acordo com essa definição, os data warehouses são

- Orientados para o assunto. Eles podem analisar dados sobre um determinado assunto ou área funcional (como vendas).
- Integrados. Os data warehouses criam consistência entre diferentes tipos de dados de fontes distintas.
- Não volátil. Quando os dados estão em um data warehouse, eles são estáveis e não mudam.
- Variáveis de acordo com o tempo. A análise de data warehouse analisa as mudanças ao longo do tempo.

Um data warehouse bem projetado realizará consultas muito rapidamente, fornecerá alta taxa de transferência de dados e dará flexibilidade suficiente para os usuários finais “dividirem e organizarem” ou reduzirem o volume de dados para um exame mais detalhado a fim de atender a uma variedade de demandas, seja em um nível alto ou em um nível muito bom e detalhado. O data



warehouse serve como a base funcional para ambientes de BI de middleware que fornecem aos usuários finais relatórios, painéis e outras interfaces.

Arquitetura de Data Warehouse

A arquitetura de um data warehouse é determinada pelas necessidades específicas da organização. Arquiteturas comuns incluem

- **Simples.** Todos os data warehouses compartilham um design básico no qual metadados, dados de resumo e dados brutos são armazenados no repositório central do warehouse. O repositório é alimentado por fontes de dados em uma extremidade e acessado por usuários finais para análise, relatório e mineração na outra extremidade.
- **Simples com uma área de preparo.** Os dados operacionais devem ser limpos e processados antes de serem colocados no warehouse. Embora isso possa ser feito de forma programática, muitos data warehouses adicionam uma área de preparo de dados antes que eles entrem no warehouse, para simplificar a preparação dos dados.
- **Hub e spoke.** A adição de data marts entre o repositório central e os usuários finais permite que uma organização personalize seu data warehouse para atender a várias linhas de negócios. Quando os dados estiverem prontos para uso, eles serão movidos para o data mart apropriado.
- **Áreas restritas.** As áreas restritas são áreas privadas, seguras e protegidas que permitem às empresas explorar rápida e informalmente novos conjuntos de dados ou formas de analisar dados sem ter que obedecer ou cumprir as regras formais e o protocolo do data warehouse.

A Evolução dos Data Warehouses — da Análise de Dados à IA e ao Machine Learning

Quando os data warehouses entraram em cena pela primeira vez no final dos anos 1980, o objetivo era ajudar o fluxo de dados dos sistemas operacionais em sistemas de suporte à decisão (DSSs). Esses data warehouses iniciais exigiam uma quantidade enorme de redundância. A maioria das organizações tinha vários ambientes DSS que atendiam a seus vários usuários. Embora os ambientes de DSS usassem muitos dos mesmos dados, a coleta, a limpeza e a integração dos dados eram frequentemente replicadas em cada ambiente.

À medida que os data warehouses se tornaram mais eficientes, eles evoluíram do armazenamento de informações que forneciam suporte a plataformas tradicionais de BI a amplas infraestruturas



analíticas compatíveis com uma grande variedade de aplicativos, como análise avançada operacional e gerenciamento de desempenho.

As iterações de data warehouse progrediram ao longo do tempo para fornecer um valor adicional incremental à empresa.

Etapas	Capacidade	Valor do Negócio
1	Relatório transacional	Fornecer informações relacionais para criar instantâneos de desempenho de negócios
2	Divisão e organização, consulta ad hoc, ferramentas de BI	Expandir recursos para informações mais detalhadas e análises mais robustas
3	Previsão do desempenho no futuro (mineração de dados)	Desenvolver visualizações e business intelligence voltado para o futuro
4	Análise tática (espacial, estatística)	Oferecer cenários "hipotéticos" para fundamentar decisões práticas baseadas em análises mais abrangentes
5	Armazena muitos meses ou anos de dados	Armazena dados por apenas semanas ou meses

O suporte a cada uma dessas cinco etapas exigiu uma variedade crescente de conjuntos de dados. As três últimas etapas em particular criam o imperativo para uma gama ainda maior de recursos de dados e análise avançada.

Atualmente, a IA e o machine learning estão transformando quase todos os setores, serviços e ativos empresariais, e os data warehouses não são exceção. A expansão do big data e a aplicação de novas tecnologias digitais estão impulsionando a mudança nos requisitos e recursos de data warehouse.

O data warehouse de dados autônomo é a etapa mais recente nessa evolução, oferecendo às empresas a capacidade de extrair ainda mais valor de seus dados enquanto reduz os custos e melhora a confiabilidade e o desempenho do data warehouse.

Planejamento de um Data Warehouse

Quando uma organização se propõe a projetar um data warehouse, ela deve começar definindo seus requisitos de negócios específicos, concordando com o escopo e elaborando um projeto conceitual. A organização pode então criar o design lógico e físico para o data warehouse. O design lógico envolve as relações entre os objetos e o design físico envolve a melhor maneira de armazenar e recuperar os objetos. O design físico também incorpora os processos de transporte, backup e recuperação.

Todo design de data warehouse de dados deve abordar o seguinte:



- Conteúdo de dados específicos
- Relacionamentos dentro e entre grupos de dados
- O ambiente de sistemas que fornecerá suporte ao data warehouse.
- Os tipos de transformações de dados necessários
- Frequência de atualização de dados

Um fator primário no design são as necessidades dos usuários finais. A maioria dos usuários finais está interessada em realizar análises e analisar dados agregados, em vez de transações individuais. No entanto, muitas vezes os usuários finais realmente não sabem o que querem até que surja uma necessidade específica. Assim, o processo de planejamento deve incluir exploração suficiente para antecipar as necessidades. Finalmente, o design do data warehouse deve permitir espaço para expansão e evolução para acompanhar as necessidades em evolução dos usuários finais.

Fonte: Oracle. **O que É um Data Warehouse?**. 2020. Disponível em:
<<https://www.oracle.com/br/database/what-is-a-data-warehouse/>>. Acesso: 24 jul. 2020.

Preparando-se Para a Próxima Teleaula

1. Acessem a webaula correspondente à próxima aula.
2. Leiam a seção do livro didático correspondente à próxima aula.
3. Resolvam as questões básicas correspondentes à próxima aula.
4. Estudem as atividades preparatórias, técnicas e recursos que serão utilizados na próxima aula, por exemplo, um tutorial eletrônico ou um mapa conceitual.
5. A resolver as atividades de aprofundamento da aula atual no pós-aula.
6. A resolver as atividades complementares no pós-aula.

Bons estudos!

