

Conceitos de Data Warehouse

Conceitos básicos do Amazon Redshift

O que é um data warehouse?

Um data warehouse é um repositório central de informações que podem ser analisadas para tomar decisões mais adequadas. Os dados fluem de sistemas transacionais, [bancos de dados relacionais](#) e de outras fontes para o data warehouse, normalmente com uma cadência regular. Analistas de negócios, engenheiros de dados, cientistas de dados e tomadores de decisões acessam os dados por meio de [ferramentas de inteligência de negócios \(BI\)](#), clientes SQL e outros aplicativos de análise.

Dados e análises se tornaram indispensáveis para que as empresas se mantenham competitivas. Os usuários corporativos contam com relatórios, painéis e análises para extrair insights dos dados, monitorar a performance dos negócios e apoiar a tomada de decisões. Os data warehouses alimentam esses relatórios, painéis e ferramentas de análise armazenando dados de maneira eficiente para minimizar a entrada e saída (E/S) dos dados e fornecer resultados de consulta rapidamente para centenas e milhares de usuários simultaneamente.

Como um data warehouse é arquitetado?

Uma arquitetura de data warehouses é composta de camadas. A camada superior é o cliente de front-end, que apresenta os resultados por meio de ferramentas de relatórios, análises e mineração de dados. A camada intermediária consiste no mecanismo de análises, usado para acessar e analisar os dados. A camada inferior da arquitetura é o servidor de banco de dados, onde os dados são carregados e armazenados. Os dados são armazenados de dois modos diferentes: 1) os dados acessados com frequência são armazenados em armazenamento muito rápido (como unidades SSD) e 2) os dados acessados com pouca frequência são armazenados em um armazenamento de objetos barato, como o Amazon S3. O data warehouse garantirá automaticamente que os dados acessados com frequência sejam movidos para o armazenamento "rápido", para otimizar a velocidade da consulta.

Como funciona um data warehouse?

Um data warehouse pode conter vários bancos de dados. Dentro de cada banco de dados, os dados são organizados em tabelas e colunas. Dentro de cada coluna, você pode definir uma descrição dos dados, como número inteiro, campo de dados ou sequência. As tabelas podem ser organizadas dentro de esquemas, que você pode considerar como pastas. Quando os dados são consumidos, eles são armazenados em várias tabelas descritas pelo esquema. As ferramentas de consulta usam o esquema para determinar as tabelas de dados que serão acessadas e analisadas.

Quais são os benefícios de usar um data warehouse?

Os benefícios de um data warehouse incluem o seguinte:

- Tomada de decisão adequada
- Dados consolidados de várias fontes
- Análise de dados históricos
- Qualidade, consistência e precisão de dados
- Separação do processamento analítico dos bancos de dados transacionais, o que melhora o desempenho dos dois sistemas

Como os data warehouses, os bancos de dados e data lakes funcionam juntos?

Normalmente, as empresas usam uma combinação de banco de dados, data lake e data warehouse para armazenar e analisar dados.

A [arquitetura de lake house](#) do Amazon Redshift facilita essa integração.

À medida que o volume e a variedade de dados aumentam, é vantajoso seguir um ou mais padrões comuns para trabalhar com dados em seu banco de dados, data lake e data warehouse:

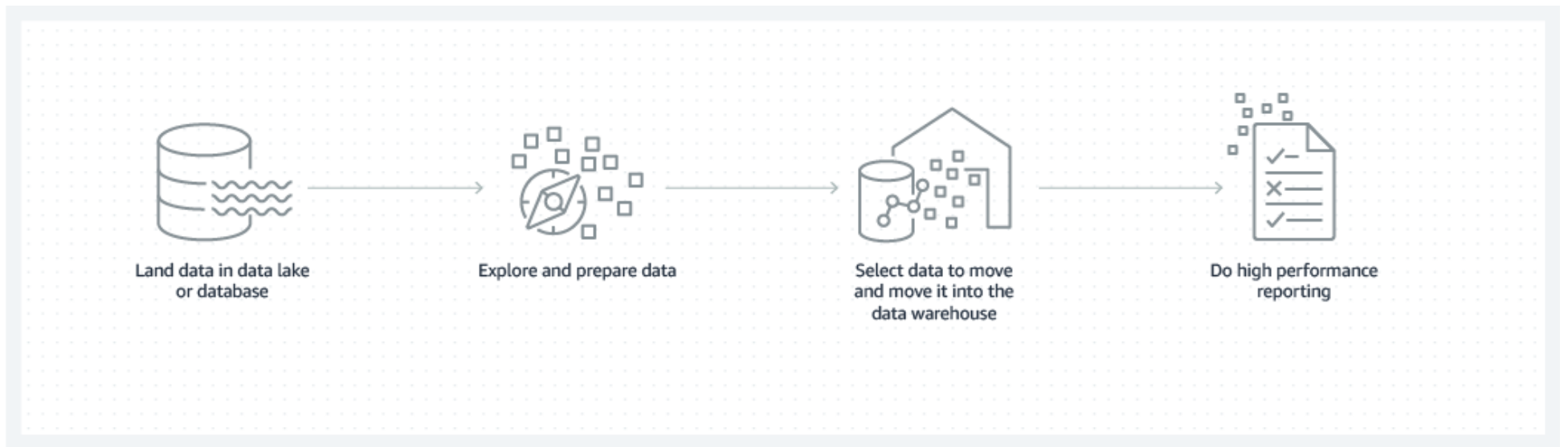


Imagem (acima): armazenar os dados em um banco de dados ou datalake, preparar os dados, mover os dados selecionados para um data warehouse e executar relatórios.



Image (acima): armazenar os dados em uma data warehouse, analisar os dados e compartilhá-los para usar com outros serviços de análise e machine learning.

Um data warehouse é projetado especificamente para análises de dados, que envolvem a leitura de grandes quantidades de dados para compreender relações e tendências entre os dados. Um banco de dados é usado para capturar e armazenar dados, como o registro de detalhes de uma transação.

Ao contrário de um data warehouse, um data lake é um repositório centralizado para todos os dados, incluindo estruturados, semiestruturados e não estruturados. Um data warehouse exige que os dados sejam organizados em um formato tabular, onde o esquema torna-se necessário. O formato tabular é necessário para que o SQL possa ser usado para consultar os dados, mas nem todos os aplicativos exigem que os dados estejam em formato de tabela. Alguns aplicativos, como análise de big data, pesquisa de texto completo e machine learning, podem acessar dados mesmo que sejam "semiestruturados" ou completamente não estruturados.

Comparação entre data warehouses e data lakes

Características	Data warehouse	Data lake
Dados	Dados relacionais de sistemas transacionais, bancos de dados operacionais e aplicativos de linha de negócios	Todos os dados, incluindo estruturados, semiestruturados e não estruturados
Esquema	Geralmente projetado antes da implementação do data warehouse, mas também pode ser gravado no momento da análise (esquema na gravação ou esquema na leitura)	Gravado no momento da análise (esquema na leitura)

Preço/performance	Resultados de consulta mais rápidos, usando armazenamento local	Resultados da consulta cada vez mais rápidos usando armazenamento de baixo custo e desacoplamento de computação e armazenamento
Qualidade dos dados	Dados altamente organizados, que representam a versão central da verdade	Quaisquer dados, organizados ou não (ou seja, dados brutos)
Usuários	Analistas de negócios, cientistas de dados e desenvolvedores de dados	Analistas de negócios (usando dados organizados), cientistas de dados, desenvolvedores de dados, engenheiros de dados e arquitetos de dados
Análises	Geração de relatórios em lote, BI e visualizações	Machine learning, análise exploratória, descoberta de dados, streaming, análise operacional, big data e criação de perfil

Comparação entre data warehouses e bancos de dados

Características	Data warehouse	Banco de dados transacional
Cargas de trabalho adequadas	Análises, relatórios e big data	Processamento de transações

Fonte de dados	Dados coletados e normalizados de diversas fontes	Dados capturados no estado em que se encontram, de uma única fonte, como um sistema transacional
Captura de dados	Operações de gravação em massa, executadas normalmente em uma programação de lotes pré-determinada	Otimizado para operações contínuas de gravação à medida que novos dados são disponibilizados para maximizar o throughput das transações
Normalização de dados	Esquemas desnormalizados, como Star ou Snowflake	Esquemas estáticos altamente normalizados
Armazenamento de dados	Otimizado para simplicidade de acesso e alto desempenho de consultas usando armazenamento colunar	Otimizado para operações de gravação de alto throughput em um único bloco físico orientado a linhas
Acesso aos dados	Otimizado para minimizar a E/S e maximizar o throughput de dados	Grandes volumes de pequenas operações de leitura

Como um data mart se compara a um data warehouse?

Um data mart é um data warehouse que atende às necessidades de uma equipe ou unidade de negócios específica, como finanças, marketing ou vendas. O data mart é menor, mais focado e pode conter resumos de dados para atender melhor à comunidade de

usuários. Um data mart também pode ser uma parte de um data warehouse.

Comparação entre data warehouses e data marts

Características	Data warehouse	Data mart
Escopo	Várias áreas centralizadas e integradas	Uma área específica e descentralizada
Usuários	De toda a organização	Uma única comunidade ou departamento
Fonte de dados	Muitas fontes	Uma ou poucas fontes, ou uma parte dos dados já coletados em um data warehouse
Tamanho	Grande, pode variar de centenas de gigabytes a petabytes	Pequeno, normalmente até algumas dezenas de gigabytes
Projeto	De cima para baixo	De baixo para cima
Detalhes dos dados	Dados completos e detalhados	Pode manter dados resumidos

Como um data warehouse pode ser implantado na AWS?

A AWS permite que você aproveite todos os principais benefícios associados à computação sob demanda: acessando a capacidade aparentemente ilimitada de armazenamento e computação, dimensionando o sistema em paralelo com a crescente quantidade de dados coletados, armazenados e consultados e pagando apenas pelos recursos que você provisiona. A AWS oferece um amplo conjunto de recursos gerenciados integrados de forma transparente entre si, o que permite implantar rapidamente uma solução completa de análises e data warehousing.

A ilustração a seguir mostra as principais etapas de um processo analítico de ponta a ponta, também chamado de pilha. A AWS oferece uma [variedade de serviços gerenciados](#) a cada etapa.



Imagem (acima): A AWS oferece uma variedade de produtos e serviços em cada etapa do processo de análise.

O [Amazon Redshift](#) é o nosso serviço de data warehouse rápido, totalmente gerenciado e econômico. Ele fornece data warehousing em escala de petabyte e análise de data lake em escala de exabyte juntos em um serviço, [pelo qual você paga apenas pelo que usar](#).

Próximas etapas

- Siga este guia detalhado e implante um [data warehouse no Amazon Redshift](#) »
- Comece a usar com o nosso [teste gratuito de dois meses](#) »

Anúncios de produtos

Acompanhe os mais recentes anúncios do AWS re:Invent »



Blog de notícias da AWS

Veja as principais escolhas de Jeff Barr e colegas no re:Invent »



TREINAMENTO DA AWS GRATUITO

Acesse mais de 500 cursos digitais gratuitos em funções, níveis de habilidade e domínios para desenvolver suas habilidades na Nuvem AWS



Faça login no console

Saiba mais sobre a AWS

O que é a AWS?

O que é a computação em nuvem?

Diversidade, equidade e inclusão na AWS

O que é o DevOps?

O que é um contêiner?

O que é um data lake?

Segurança na Nuvem AWS

Novidades

Blogs

Press Releases

Ajuda

Entre em contato conosco

Crie um tíquete de suporte

Central de conhecimento

AWS re:Post

Visão geral do AWS Support

Legal

Carreiras na AWS

Recursos da AWS

Conceitos básicos

Treinamento e certificação

Portfólio de soluções da AWS

Centro de arquitetura

Perguntas frequentes sobre produtos e tópicos técnicos

Relatórios de analistas

Parceiros da AWS

Desenvolvedores na AWS

Centro do desenvolvedor

SDKs e ferramentas

.NET na AWS

Python na AWS

Java na AWS

PHP na AWS

JavaScript na AWS

[Crie uma conta da AWS](#)



A Amazon é uma empresa empregadora orientada pelos fundamentos de igualdade de oportunidades e ações afirmativas, que não faz distinção entre *minorias, mulheres, portadores de deficiência, veteranos, identidade de gênero, orientação sexual nem idade.*

Idioma

[عربي](#) |
[Bahasa Indonesia](#) |
[Deutsch](#) |
[English](#) |
[Español](#) |
[Français](#) |
[Italiano](#) |
[Português](#) |
[Tiếng Việt](#) |
[Türkçe](#) |
[Русский](#) |
[ไทย](#) |
[日本語](#) |
[한국어](#) |
[中文 \(简体\)](#) |
[中文 \(繁體\)](#)

[Privacidade](#)

|

[Termos do site](#)

|

[Preferências de cookies](#)

|

© 2022, Amazon Web Services, Inc. ou suas afiliadas. Todos os direitos reservados.