**Recursos de armazenamento**

Introdução

O armazenamento eficaz de dados é crucial para que as organizações gerenciem e utilizem informações com eficiência, especialmente no caso da inteligência artificial (IA), onde as estratégias de armazenamento de dados devem atender a requisitos específicos. De bancos de dados estruturados a arquivos não estruturados, a forma como os dados são armazenados pode impactar significativamente a acessibilidade, a segurança, a escalabilidade e o desempenho de modelos e cargas de trabalho de IA. Desenvolver uma estratégia robusta de armazenamento de dados é essencial para garantir que os dados sejam armazenados e organizados com segurança para oferecer fácil recuperação, processamento em alta velocidade, análise avançada e conformidade com os requisitos regulatórios. Você aprenderá sobre diversas estratégias e opções de armazenamento personalizadas para IA, protocolos de transporte de armazenamento e armazenamento definido por software que garantirão uma base de dados sólida para suas iniciativas, cargas de trabalho e clusters de IA.

Estratégia de Armazenamento de Dados

Uma estratégia de armazenamento de dados é um plano para gerenciar e usar dados. Envolve a escolha de opções de armazenamento, a organização dos dados e a garantia de backups e recuperação. O objetivo é manter os dados seguros, acessíveis e econômicos.

Visão geral do armazenamento de dados



Interface gráfica do usuário, Aplicativo

O conteúdo gerado por IA pode estar incorreto.

A seguir estão os benefícios da estratégia de armazenamento de dados:

* Dados mais seguros
* Conformidade garantida
* Aplicações com melhor desempenho
* Custos mais baixos
* Operações mais eficientes
* Utilização otimizada de recursos
* Mais flexível
* Escalabilidade mais fácil

Os principais componentes do armazenamento de dados são mídias de armazenamento, sistemas de armazenamento, tecnologias de armazenamento e gerenciamento de dados.

A tabela a seguir lista os principais componentes do armazenamento de dados com exemplos.

| Componente | Descrição | Exemplos |
| --- | --- | --- |
| Mídia de armazenamento | Dispositivos físicos usados ​​para armazenar dados | HDDs, SSDs, discos ópticos, unidade de fita |
| Sistemas de armazenamento | Métodos para organizar e acessar recursos de armazenamento | DAS, NAS, SAN, armazenamento em nuvem |
| Tecnologias de armazenamento | Tecnologias que melhoram o desempenho e a flexibilidade do armazenamento | SDS, RAID, armazenamento Flash, armazenamento híbrido |
| Gestão de dados | Sistemas e estruturas para gerenciamento de armazenamento e recuperação de dados | Sistemas de arquivos (NTFS, ext4), DBMS (MySQL, PostgreSQL), virtualização de armazenamento |

**Tipos de estratégia de armazenamento de dados**

Uma estratégia de armazenamento de dados para clusters de IA é um plano detalhado que descreve como os dados serão armazenados, gerenciados e acessados ​​para dar suporte às cargas de trabalho de IA e aprendizado de máquina (ML). Considerando as demandas significativas de dados para tarefas de IA e ML, uma estratégia de armazenamento robusta é essencial para desempenho, escalabilidade e eficiência ideais. As principais considerações e componentes que geralmente compõem uma estratégia de armazenamento de dados para clusters de IA podem ser divididos em dois grupos: considerações de desempenho e escalabilidade e considerações de gerenciamento e segurança.

As considerações sobre desempenho e escalabilidade concentram-se na otimização do desempenho, da escalabilidade e da acessibilidade do sistema de armazenamento para atender às altas demandas das cargas de trabalho de IA e ML. As considerações incluem o seguinte:

* **Desempenho e baixa latência**
  1. Unidades de estado sólido (SSDs) de alta velocidade, soluções de armazenamento NVMe (Non-Volatile Memory Express) e protocolos de armazenamento otimizados, como Remote Direct Memory Access (RDMA) sobre Ethernet convergente (RoCE), alcançam alto rendimento e baixa latência.
* **Escalabilidade**
  1. Sistemas de arquivos distribuídos (como HDFS) ou soluções de armazenamento baseadas em nuvem (por exemplo, Amazon S3, Google Cloud Storage) oferecem suporte à expansão de cargas de trabalho de IA sem sacrificar o desempenho.
* **Acessibilidade**
  1. Soluções de armazenamento compartilhado (como armazenamento conectado à rede [NAS] ou SAN) para garantir que todos os nós dentro do cluster de IA possam acessar rapidamente os dados necessários.
* **Sistemas de arquivos paralelos**
  1. Sistemas de arquivos paralelos permitem acesso e gravação simultâneos de dados por vários processos, o que aumenta a produtividade e minimiza gargalos.
* **Camadas e cache de dados**
  1. A hierarquização de dados otimiza custos e desempenho movendo dados acessados ​​com pouca frequência para um armazenamento mais lento e retendo dados acessados ​​com frequência em armazenamento de alta velocidade.
  2. O armazenamento em cache de dados usados ​​com frequência melhora o desempenho geral do sistema.
* **Compatibilidade com estruturas e ferramentas de IA**
  1. Garanta que a solução de armazenamento se integre perfeitamente com estruturas de IA como TensorFlow e PyTorch e ferramentas como Kubeflow e MLflow para otimizar a eficiência do fluxo de trabalho.
* **Custo** e **eficiência**
  1. Equilibre custo e desempenho com uma combinação de soluções de armazenamento no local e na nuvem; SSDs locais de alta velocidade para dados acessados ​​com frequência e armazenamento em nuvem econômico para dados de arquivo.

As considerações de gerenciamento e segurança abordam os aspectos de gerenciamento, segurança e confiabilidade do sistema de armazenamento para garantir a proteção de dados e o manuseio eficaz. As considerações incluem o seguinte:

* **Segurança e conformidade**
  1. Implemente criptografia, controle de acesso e auditorias regulares para proteger dados confidenciais e cumprir com os requisitos regulatórios.
* **Durabilidade e confiabilidade dos dados**
  1. Soluções de armazenamento redundantes, backups regulares e sistemas tolerantes a falhas protegem dados valiosos e evitam perdas por falhas de hardware.
* **Armazenamento definido por software (SDS)**
  1. As soluções SDS para flexibilidade e gerenciamento simplificado abstraem o hardware de armazenamento e o apresentam como um conjunto unificado de recursos, que permitem o gerenciamento e a automação por meio de software.

Cada grupo aborda aspectos diferentes, mas complementares, de uma estratégia abrangente de armazenamento de dados, garantindo que as necessidades de desempenho e os requisitos de gerenciamento sejam atendidos.

Na prática, uma estratégia de armazenamento de dados para clusters de IA normalmente envolve uma combinação de armazenamento local de alto desempenho para lidar com cargas de trabalho ativas, armazenamento em nuvem para escalabilidade e flexibilidade e armazenamento em arquivo para retenção de dados a longo prazo. Essa abordagem permite que as organizações equilibrem a necessidade de velocidade e capacidade de resposta com a capacidade de escalar e gerenciar custos de forma eficaz.

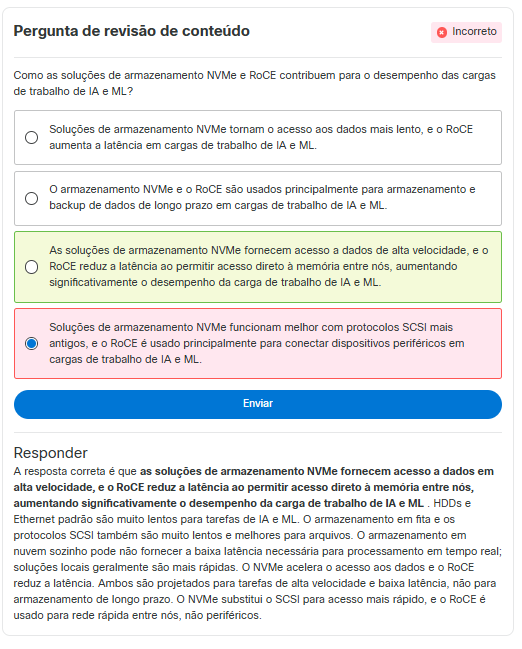
O armazenamento local fornece o acesso de baixa latência necessário para o processamento de dados em tempo real e tarefas de computação de alto desempenho. O armazenamento em nuvem oferece a escalabilidade necessária para acomodar volumes crescentes de dados e flexibilidade no gerenciamento de dados em diferentes localizações geográficas. O armazenamento em arquivo retém com segurança dados históricos que podem ser acessados ​​com pouca frequência, mas ainda são valiosos para conformidade, pesquisa ou análises futuras.

A estratégia específica deve ser adaptada aos requisitos específicos da sua organização, incluindo a natureza das cargas de trabalho de IA, restrições orçamentárias e objetivos de gerenciamento de dados a longo prazo. Considere fatores como padrões de acesso aos dados, a importância da durabilidade dos dados e a integração com os sistemas existentes. Ao utilizar essa abordagem de armazenamento em várias camadas, as organizações podem garantir que sua infraestrutura de IA seja robusta e adaptável para atender às necessidades atuais e futuras de gerenciamento de dados.

O diagrama a seguir fornece uma visão geral estruturada dos principais componentes de uma estratégia de armazenamento de dados e ilustra como eles se inter-relacionam para atender às necessidades de clusters e cargas de trabalho de IA.

Diagrama

O conteúdo gerado por IA pode estar incorreto.



**Gerenciamento de Estratégia de Armazenamento de Dados**

Sistemas de armazenamento de dados não são fáceis de gerenciar. Os dados devem ser de fácil acesso, mas armazenados com segurança para proteção contra erros humanos, ataques maliciosos e falhas de máquina.

A lista a seguir inclui algumas boas práticas que ajudam você a gerenciar sua estratégia de armazenamento de dados com mais eficiência:

* **Aplicativos de gerenciamento de recursos de armazenamento (SRM) :** os SRMs são especialmente úteis em ambientes grandes e de vários fornecedores, para garantir que os componentes estejam operando com eficiência e que a capacidade não seja desperdiçada.
* **Consolidação de sistemas:** Gerenciar o armazenamento se torna mais complexo à medida que mais sistemas de armazenamento são adicionados para atender às crescentes necessidades de capacidade. Em vez de substituir os sistemas existentes por outros maiores, geralmente é mais fácil adicionar novas unidades. No entanto, gerenciar muitos sistemas separados, especialmente sistemas NAS que lidam com um volume crescente de dados de arquivos, pode ser difícil. Consolidar vários sistemas de armazenamento em uma unidade maior simplifica o gerenciamento, mas pode exigir investimentos em ferramentas ou serviços para mover os dados, especialmente se os sistemas forem de fornecedores diferentes.
* **Armazenamento em camadas:** O armazenamento em camadas envolve a categorização dos dados com base em sua importância para a empresa e o armazenamento em tipos apropriados de armazenamento. Dados menos críticos podem residir em sistemas de disco mais lentos e baratos, enquanto dados cruciais e acessados ​​com frequência são armazenados em SSDs rápidos. Essa abordagem simplifica o gerenciamento de dados e garante que diferentes tipos de dados recebam níveis adequados de proteção.
* **Implantação estratégica de SSDs:** Um dos desafios mais complexos no gerenciamento de armazenamento é otimizar um conjunto de armazenamento para obter o desempenho máximo. Tradicionalmente, técnicas como o short-stroking eram usadas para aumentar o desempenho de sistemas baseados em disco, concentrando-se nas partes mais rápidas dos discos rígidos. No entanto, esse método era complexo de gerenciar e ineficiente em termos de espaço em disco. Uma solução mais eficaz hoje em dia é integrar SSDs estrategicamente ao conjunto para aplicações de alto desempenho. Embora os SSDs sejam mais caros do que os discos tradicionais, sua implantação direcionada pode otimizar o gerenciamento de desempenho e potencialmente reduzir custos ao longo do tempo.
* **Armazenamento em nuvem híbrida:** O uso do armazenamento em nuvem ajuda a reduzir a necessidade de armazenamento local e facilita o gerenciamento. Os dados armazenados na nuvem exigem supervisão mínima e praticamente nenhum esforço de gerenciamento. Ao transferir dados menos utilizados e menos críticos para a nuvem, os administradores de armazenamento podem se concentrar em dados de maior prioridade que precisam de acesso rápido e soluções de armazenamento robustas.
* **Opte por sistemas de armazenamento escaláveis:** Escolher um sistema de armazenamento escalável pode simplificar o gerenciamento do seu ambiente de armazenamento. Ao contrário dos sistemas escaláveis, que não permitem adicionar controladores à medida que a capacidade aumenta, os arrays escaláveis ​​permitem que os administradores adicionem capacidade e controladores. Esse recurso ajuda a manter os níveis de desempenho à medida que as necessidades de armazenamento da empresa aumentam.
* **Maximize a eficiência da utilização de máquinas virtuais:** a virtualização de servidores permite a criação fácil de máquinas virtuais (VMs), mas essa situação pode levar a muitas VMs não utilizadas ou abandonadas, também conhecidas como VMs órfãs. Essas VMs ainda ocupam espaço de armazenamento e aumentam os esforços de gerenciamento, especialmente para tarefas de proteção de dados.

Para resolver esse problema, os administradores podem usar ferramentas como o vCenter para VMware ou scripts para Hyper-V e outros ambientes de servidores virtuais para identificar e remover VMs órfãs com eficiência. Essa abordagem ajuda a otimizar as operações do data center e reduz o uso desnecessário de armazenamento.

Interface gráfica do usuário, Texto, Aplicativo, Email

O conteúdo gerado por IA pode estar incorreto.

Fibre Channel e FCoE

Fibre Channel é um protocolo de rede de alta velocidade usado em SANs para transferência de dados entre computadores e dispositivos de armazenamento. SANs baseadas em Fibre Channel dependem de hardware especializado e dedicado e criam uma infraestrutura separada da LAN. O Fibre Channel over Ethernet (FCoE) permite a comunicação via Protocolo Fibre Channel em redes Ethernet, combinando a confiabilidade do Fibre Channel com a flexibilidade da Ethernet.

A figura a seguir mostra uma SAN usando o Protocolo Fibre Channel. Os dois sistemas de armazenamento na parte superior se conectam a vários switches Fibre Channel no meio, que por sua vez se conectam a vários servidores na parte inferior. Essa topologia permite que os servidores acessem o armazenamento compartilhado em uma rede de alta velocidade.

Diagrama

O conteúdo gerado por IA pode estar incorreto.

**Protocolo Fibre Channel**

O Fibre Channel foi criado devido a limitações no transporte de armazenamento, como o barramento paralelo SCSI (Small Computer Systems Interface). O Protocolo Fibre Channel permite a multiplexação de múltiplos comandos de armazenamento (como SCSI) em paralelo e conecta um servidor a vários dispositivos de armazenamento em grandes redes e distâncias geográficas. Esses dispositivos de armazenamento aparecem para o sistema operacional do servidor como discos conectados localmente.

Uma SAN é uma rede dedicada para acesso a armazenamento em nível de bloco. Ela consiste em dispositivos de comunicação especializados que executam o Protocolo Fibre Channel, como matrizes de armazenamento Fibre Channel e servidores com adaptadores de barramento de host (HBAs). Os principais participantes de uma SAN são servidores (iniciadores) e sistemas de armazenamento (destinos). O HBA do servidor envia comandos de armazenamento pela SAN Fibre Channel. Os destinos de armazenamento então processam esses comandos.

O Fibre Channel Protocol oferece suporte a protocolos de camada superior, como SCSI, IBM's Fiber Connection (FICON), ATM e IP por meio de IP sobre Fibre Channel (IPFC), e é usado em configurações Just a Bunch of Disks (JBOD) que conectam discos em uma topologia de loop arbitrada.

Ao transferir dados em blocos, também conhecido como transferência de dados baseada em blocos, é importante criar uma infraestrutura de comunicação segura que abranja tanto o protocolo de comunicação a ser utilizado quanto o hardware necessário para suportá-lo. O Protocolo Fibre Channel cria um ambiente seguro e confiável ao implementar um controle rigoroso na conectividade entre os dispositivos, o que envolve a atribuição de diferentes funções de porta (discutidas no próximo parágrafo). Outro mecanismo é o controle da transmissão exercido pelo lado receptor com a ajuda de buffers de memória de hardware.

O Protocolo Fibre Channel utiliza diferentes tipos de portas para conectar dispositivos e gerenciar o tráfego dentro de uma SAN. Os principais tipos de portas incluem portas de nó (N), portas de malha (F) e portas de expansão (E). A porta de expansão de tronco (TE) é uma opção especializada da porta E que conecta switches para permitir a expansão da SAN com maior largura de banda por meio da agregação de links.

A Porta F conecta um único servidor ou dispositivo de armazenamento a um switch em uma SAN. A Porta TE conecta switches, expande a SAN e suporta maior largura de banda usando agregação de links. A Porta N é um ponto final na rede Fibre Channel. A Porta E conecta dois switches Fibre Channel. O HBA é um componente de hardware que conecta um sistema host (como um servidor) a uma rede ou dispositivo de armazenamento.

A figura a seguir mostra um iniciador Fibre Channel conectado a uma SAN Fibre Channel baseada no Cisco Multilayer Director Switch (MDS) e posteriormente conectado ao sistema de armazenamento Fibre Channel.

Diagrama

O conteúdo gerado por IA pode estar incorreto.

Como resultado, o sistema operacional do servidor percebe os números de unidade lógica (LUNs) do conjunto de armazenamento como se estivessem conectados diretamente ao servidor. LUNs são identificadores exclusivos para dispositivos de armazenamento ou partições individuais dentro de um conjunto de armazenamento. Esses LUNs são visíveis no destino Fibre Channel através da rede Fibre Channel e simulam uma conexão direta por meio de um barramento SCSI paralelo dentro da arquitetura do servidor.

O Fibre Channel SAN oferece conectividade e armazenamento de data center de alto desempenho para recuperação de desastres e soluções de continuidade de negócios.

O armazenamento Fibre Channel também oferece armazenamento em nuvem que é usado nas seguintes tarefas:

* Proteção de dados
* Backup
* Restauração e replicações no nível do bloco
* Serviços de arquivamento ou preservação de dados
* Sistemas de desduplicação
* Bibliotecas de fitas virtuais (VTLs)

Às vezes, um dispositivo front-end NAS é conectado ao subsistema de armazenamento back-end SAN Fibre Channel para formar um híbrido de armazenamento NAS e SAN.

Em redes de data center Cisco, você pode implantar o iniciador Fibre Channel em servidores blade Cisco UCS Série B ou servidores rack Cisco UCS Série C. Esses servidores se conectam através da SAN Fibre Channel ao subsistema de armazenamento de destino Fibre Channel. O subsistema de armazenamento de destino pode ser de fornecedores terceirizados de armazenamento Fibre Channel, como Dell EMC, NetApp, IBM, Hitachi, HP, Pure Storage, Red Hat e outros.

**Canal de fibra sobre Ethernet**

Em um data center típico, você teria duas redes separadas: uma para Ethernet e outra para Fibre Channel. Para suportar vários tipos de redes, a abordagem tradicional utiliza módulos de interface redundantes separados para cada rede: placas de interface de rede Ethernet (NICs) e HBAs Fibre Channel em seus servidores, além de pares redundantes de switches em cada camada da arquitetura de rede. O uso de infraestruturas paralelas aumenta os custos de capital, dificulta a gestão do data center e reduz a flexibilidade dos negócios.

A consolidação do tráfego de E/S no data center reúne as redes Fibre Channel e Ethernet em uma única infraestrutura integrada. A fusão dos dispositivos que gerenciam o tráfego Fibre Channel e Ethernet proporciona economia em termos de aquisição de equipamentos e custos de espaço, energia e refrigeração. Um pilar importante dessa abordagem consolidada é o FCoE.

A figura a seguir mostra três topologias de rede: SAN, Ethernet e infraestrutura convergente (FCoE). À esquerda, a infraestrutura SAN possui duas redes de armazenamento separadas, cada uma com seus próprios cabos. No meio, a infraestrutura Ethernet processa o tráfego LAN separadamente da SAN. À direita, a infraestrutura convergente (FCoE) combina SAN e Ethernet em uma única rede; a Ethernet transporta o tráfego LAN e SAN, o que torna o sistema mais simples e eficiente.

Diagrama

O conteúdo gerado por IA pode estar incorreto.

**Observação**

*A infraestrutura convergente combina computação, armazenamento, rede e virtualização em um sistema unificado, o que simplifica o gerenciamento e melhora a eficiência.*

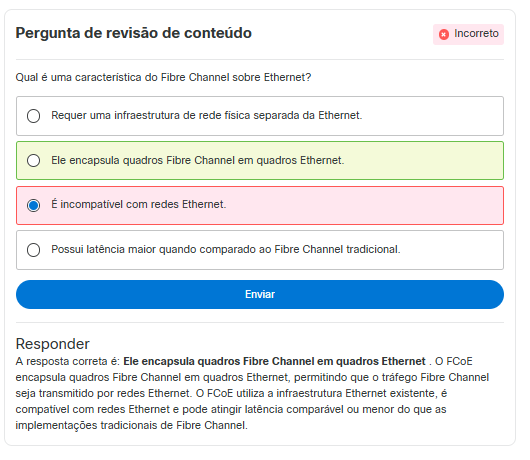
O FCoE é uma solução ponta a ponta (inicializando o conjunto de armazenamento do servidor para o destino) que permite a comunicação via Protocolo Fibre Channel sobre Ethernet convergente. O FCoE cria uma única rede LAN-SAN convergente que mantém os recursos nativos do Fibre Channel. O principal benefício é a infraestrutura consolidada e centralizada.

Adaptadores de rede convergentes (CNAs) são componentes de infraestrutura convergente que combinam a funcionalidade de NICs Ethernet virtuais, HBAs Fibre Channel virtuais, switches FCoE e subsistemas de armazenamento com controladores de armazenamento FCoE unificados.

A tecnologia FCoE encapsula quadros do Protocolo Fibre Channel em quadros Ethernet, permitindo o transporte de tráfego Fibre Channel por redes Ethernet padrão. Essa convergência simplifica o cabeamento e reduz o número de adaptadores e switches de rede necessários.

O subsistema de armazenamento baseado em FCoE pode ser conectado aos módulos de switch Cisco MDS série 9000, como os seguintes:

* Cisco MDS série 9500 com módulo FCoE de 10 Gbps e 8 portas
* Cisco MDS série 9700 com módulo FCoE de 10 Gbps e 48 portas
* Switch de malha Cisco MDS 9220i
* Switch Cisco Nexus 93108TC-FX
* Interconexão de malha Cisco



* NVMe e NVMe sobre Fabrics
* NVMe e NVMe sobre Fabrics (NVMe-oF) são essenciais para o armazenamento de dados de IA e oferecem baixa latência, alta taxa de transferência e escalabilidade essenciais para lidar com grandes conjuntos de dados e processamento rápido de dados. O NVMe garante acesso rápido a SSDs, enquanto o NVMe-oF estende esses benefícios às fabrics de rede para permitir acesso e gerenciamento remotos eficientes em ambientes distribuídos. Vamos analisar cada uma dessas soluções em detalhes.

Diagrama

O conteúdo gerado por IA pode estar incorreto.

NVMe

NVMe é um protocolo de armazenamento otimizado para SSDs modernos e tecnologias de memória não volátil. Ao contrário de protocolos mais antigos, como Serial Advanced Technology Attachment (SATA) e Serial-Attached SCSI (SAS), projetados para discos rígidos tradicionais, o NVMe aproveita ao máximo a baixa latência, a alta velocidade e o paralelismo oferecidos pela memória flash NAND. Posicionado diretamente no barramento Peripheral Component Interconnect Express (PCIe), o NVMe elimina os gargalos dos protocolos legados e permite taxas de transferência de dados mais rápidas, latência reduzida e processamento de dados mais eficiente.

Linha do tempo

O conteúdo gerado por IA pode estar incorreto.

O NVMe é cada vez mais crítico em cargas de trabalho de IA e ML devido às suas capacidades de transferência de dados em alta velocidade, baixa latência e capacidade de lidar com grandes volumes de dados de forma eficiente.

Os principais recursos e vantagens do NVMe em IA são os seguintes:

* **Alto rendimento e baixa latência:** o NVMe permite o processamento rápido de dados para cargas de trabalho de IA e minimiza gargalos de armazenamento durante tarefas como o treinamento de redes neurais.
* **Paralelismo** : projetado para CPUs modernas, o NVMe lida com várias solicitações de E/S simultaneamente, o que é essencial para processar grandes conjuntos de dados em aplicativos de IA.
* **Escalabilidade** : o NVMe escala eficientemente com conjuntos de dados de IA crescentes, suporta múltiplas unidades sem degradar o desempenho e é crucial para sistemas de IA de alta potência.
* **Processamento de dados em tempo real** : as altas velocidades de leitura e gravação do NVMe oferecem suporte à inferência de IA em tempo real e ao treinamento de modelos.
* **Otimização do data center** : o NVMe reduz os requisitos de energia, calor e espaço e, portanto, é adequado para operações de IA em larga escala em data centers modernos.
* **Computação de ponta** : a velocidade e a confiabilidade do NVMe são essenciais para tarefas de IA em tempo real na ponta, como em veículos autônomos e dispositivos de Internet das Coisas (IoT).

A figura a seguir descreve casos de uso em que o NVMe é usado em IA.

Diagrama, Esquemático

O conteúdo gerado por IA pode estar incorreto.

NVMe sobre Tecidos

NVMe-oF é uma extensão do protocolo NVMe projetada para permitir acesso de alto desempenho a dispositivos de armazenamento NVMe em uma rede. O NVMe foi originalmente desenvolvido para se conectar diretamente a dispositivos de armazenamento de alta velocidade, como SSDs, usando o barramento PCIe. O NVMe-oF permite que benefícios como baixa latência e alta taxa de transferência sejam estendidos a vários tipos de malhas de rede, o que permite acesso remoto ao armazenamento com degradação mínima do desempenho.

As características do NVMe-oF são as seguintes:

* **Objetivo:** O NVMe-oF estende o desempenho do NVMe ao armazenamento em rede e permite acesso de alta velocidade em ambientes distribuídos.
* **Protocolos de transporte**
  1. **NVMe sobre Fibre Channel (FC-NVMe):** Este protocolo oferece alta confiabilidade e desempenho em redes Fibre Channel.
  2. **RDMA:** Este protocolo fornece baixa latência com acesso direto à memória via RoCE, Internet Wide Area RDMA (iWARP) e InfiniBand.
  3. **NVMe sobre TCP:** este protocolo oferece implantação flexível por Ethernet padrão com benefícios de desempenho.
* **Desempenho e latência:** o NVMe-oF mantém baixa latência e alto rendimento, o que é adequado para computação de alto desempenho e IA.
* **Escalabilidade e flexibilidade:** o NVMe-oF oferece suporte a pools de armazenamento grandes e escaláveis ​​para data centers modernos.
* **Casos de uso:** NVMe-oF é ideal para análises em tempo real, bancos de dados, IA e cargas de trabalho de ML.

O NVMe-oF funciona com um controlador e uma opção como Fibre Channel ou InfiniBand para transportar dados.

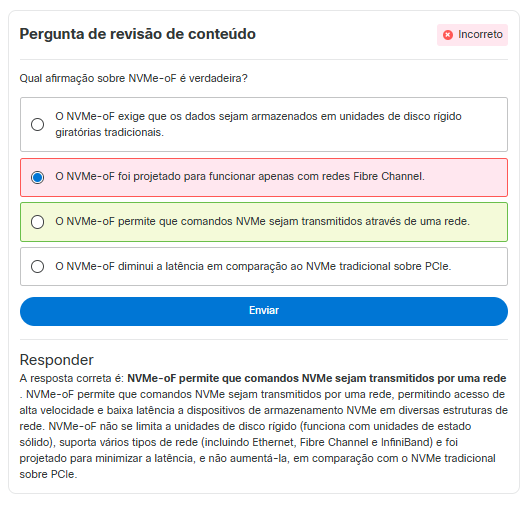
Interface gráfica do usuário

O conteúdo gerado por IA pode estar incorreto.

NVMe sobre Fabrics vs. NVMe

A tabela a seguir compara NVMe e NVMe-oF e destaca suas diferenças em design de protocolo, características de desempenho, escalabilidade, protocolos de transporte, casos de uso, latência, taxa de transferência e flexibilidade.

| Aspecto | NVMe | NVMe-oF |
| --- | --- | --- |
| Protocolo | Projetado para acesso de armazenamento local (SSDs) | Estende o protocolo NVMe sobre estruturas de rede |
| Desempenho | Latência muito baixa, alto rendimento | Mantém baixa latência; a taxa de transferência depende da rede |
| Escalabilidade | Suporta filas de comando paralelas | Escalável dentro do servidor ou sistema; menos escalável na rede |
| Protocolos de transporte | PCI | Canal de Fibra, RDMA (RoCE, iWARP, IB), TCP/IP |
| Caso de uso | Armazenamento local em um servidor | Acesso remoto ao armazenamento NVMe por meio de redes |
| Latência | Extremamente baixo | Baixo, comparado aos protocolos tradicionais, maior que NVMe |
| Taxa de transferência | Alto | Alto, mas dependente da infraestrutura de rede |
| Flexibilidade | Limitado ao ambiente local | Permite armazenamento compartilhado em ambientes distribuídos |



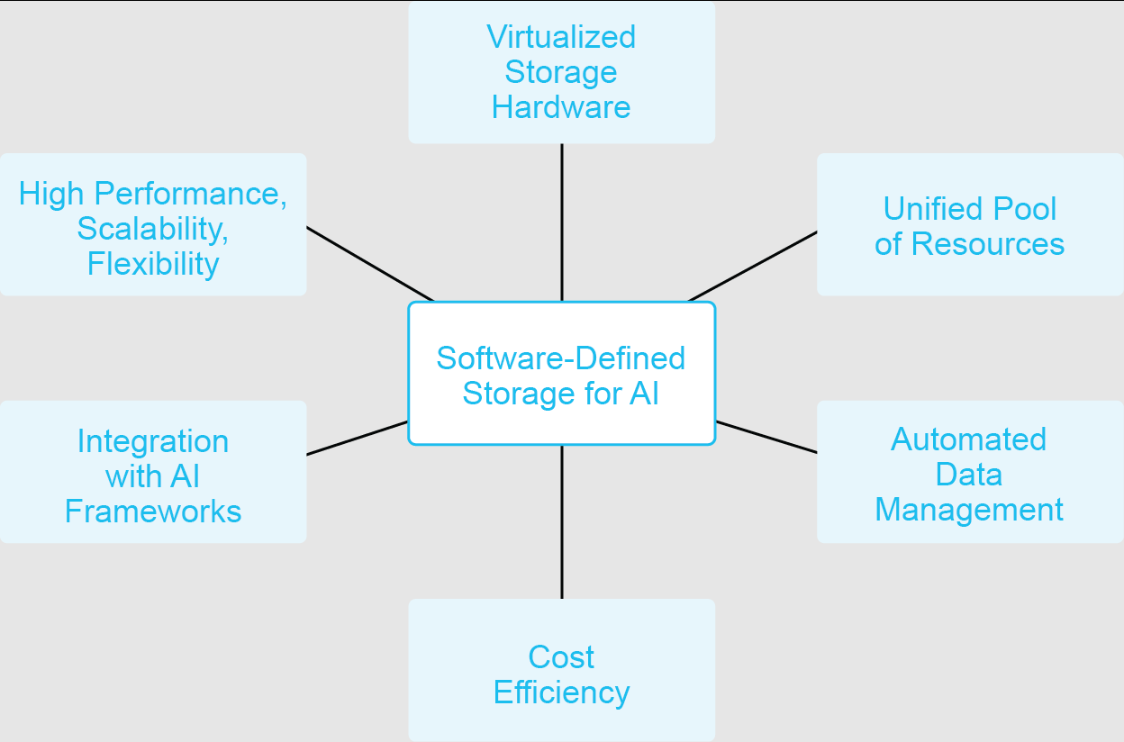
**Armazenamento definido por software**

O armazenamento definido por software é um tipo de virtualização de armazenamento que introduz a simplicidade e a flexibilidade do gerenciamento de armazenamento de dados. Ele abstrai e virtualiza o hardware de armazenamento físico em um conjunto unificado de recursos gerenciados por software. Ele atinge esse objetivo adicionando uma camada de software que opera independentemente do hardware de armazenamento físico, como discos e SSDs. Essa camada de software oferece mais liberdade e controle sobre o armazenamento e o acesso aos dados, tornando o gerenciamento de armazenamento mais eficiente e adaptável.

Essa abstração é crucial para cargas de trabalho de IA, que exigem acesso rápido a grandes conjuntos de dados e capacidade de escalabilidade perfeita. O armazenamento definido por software oferece gerenciamento automatizado de dados, alto desempenho e adaptabilidade a diferentes estruturas e ferramentas, incluindo ferramentas de IA, o que o torna uma solução ideal para as demandas dinâmicas das aplicações.

Os benefícios de usar armazenamento definido por software com cargas de trabalho de IA são os seguintes:

* **Hardware de armazenamento virtualizado:** dispositivos de armazenamento físico são abstraídos.
* **Pool unificado de recursos:** os recursos de armazenamento são consolidados.
* **Gerenciamento automatizado de dados:** o tratamento e a otimização de dados são controlados por software.
* **Integração com estruturas de IA:** o armazenamento definido por software é compatível com ferramentas e plataformas de IA.
* **Alto desempenho:** as velocidades de acesso e processamento de dados são rápidas.
* **Escalabilidade:** você pode expandir a capacidade de armazenamento conforme necessário.
* **Flexibilidade:** o armazenamento definido por software se adapta às mudanças nas cargas de trabalho e nos requisitos.
* **Eficiência de custos:** o armazenamento definido por software otimiza o uso do armazenamento e reduz a dependência de hardware.



A imagem a seguir mostra um ambiente de armazenamento definido por software e ilustra sua flexibilidade e integração para gerenciar o armazenamento em vários ambientes. Dois servidores com dispositivos de armazenamento locais (discos rígidos e SSDs) conectam-se a uma camada central de armazenamento definido por software. Essa camada gerencia e agrupa recursos de armazenamento de servidores e de uma SAN física centralizada. Acima da camada de armazenamento definido por software, VMs e contêineres acessam os recursos de armazenamento combinados.

Diagrama

O conteúdo gerado por IA pode estar incorreto.

Interface gráfica do usuário, Texto, Aplicativo, Email

O conteúdo gerado por IA pode estar incorreto.

**Armazenamento tradicional vs. armazenamento definido por software**

Os sistemas de armazenamento tradicionais dependem de abordagens definidas por hardware para conectar o gerenciamento de dados e as operações de armazenamento a dispositivos de armazenamento físico, como HDDs e SSDs. Esses sistemas normalmente exigem gerenciamento dedicado para cada componente de hardware, o que limita a flexibilidade e a escalabilidade.

Em contraste, o armazenamento definido por software representa uma abordagem moderna que desvincula o gerenciamento de armazenamento do hardware físico. Ele introduz uma camada de software que abstrai e controla os recursos de armazenamento em diversos tipos de hardware. O armazenamento definido por software oferece maior flexibilidade, simplifica o gerenciamento de armazenamento e permite a alocação dinâmica de recursos com base nas necessidades dos aplicativos e nos requisitos de negócios.

O armazenamento tradicional é rígido e centrado em hardware. O armazenamento definido por software proporciona agilidade e eficiência por meio de seu controle baseado em software e aprimora a escalabilidade e a adaptabilidade em ambientes de TI modernos.

A lista a seguir inclui as vantagens do armazenamento definido por software em relação às soluções de armazenamento tradicionais:

* **Economia de custos** : Ao contrário das configurações tradicionais de NAS ou SAN, que limitam você a pacotes de hardware e software de fornecedores específicos, o armazenamento definido por software permite que você utilize o hardware de armazenamento e servidor existente. Essa flexibilidade elimina a necessidade de sistemas proprietários caros e ajuda a reduzir os gastos gerais de TI.
* **Escalabilidade** : O armazenamento definido por software abstrai os recursos de armazenamento do hardware, o que facilita a expansão da capacidade de armazenamento conforme necessário. Essa flexibilidade permite que as organizações se adaptem rapidamente às mudanças nas demandas sem a complexidade de integrar novos recursos de hardware.
* **Agilidade** : No acelerado ambiente de negócios atual, a agilidade é crucial. O armazenamento definido por software oferece a flexibilidade necessária para provisionar e otimizar recursos de armazenamento dinamicamente com base nos requisitos atuais. Essa adaptabilidade garante que as organizações respondam rapidamente às novas tendências e necessidades de negócios.
* **Gerenciamento unificado** : Ambientes de armazenamento tradicionais frequentemente resultam em sistemas díspares e difíceis de gerenciar com eficiência. O armazenamento definido por software integra e consolida o gerenciamento de armazenamento em uma plataforma unificada definida por software. Essa abordagem otimizada aumenta a eficiência operacional, fornecendo controle centralizado e visibilidade em toda a infraestrutura de armazenamento.

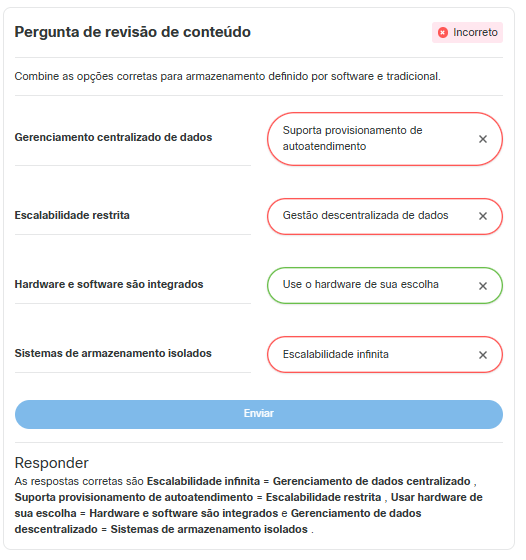
A tabela a seguir resume as principais diferenças entre o armazenamento tradicional e o definido por software.

| Armazenamento tradicional | Armazenamento definido por software |
| --- | --- |
| Hardware e software são integrados | Solução baseada em software |
| Gerenciamento centralizado de dados | Gestão descentralizada de dados |
| Escalabilidade restrita | Escalabilidade infinita |
| Opções de hardware inflexíveis | Use qualquer hardware que você escolher |
| Sistemas de armazenamento isolados | Arquitetura de escalonamento unificada |
| Sem provisionamento de autoatendimento | Suporta provisionamento de autoatendimento |

A figura a seguir mostra como o armazenamento é gerenciado antes e depois do uso do armazenamento definido por software. Sem o armazenamento definido por software, cada aplicativo fica vinculado a um sistema de armazenamento separado. Essa situação causa ineficiências, incluindo alguns sistemas com capacidade pela metade, enquanto outros com capacidade quase completa. Com o armazenamento definido por software, todo o armazenamento é combinado em um pool central, o que permite o uso equilibrado e eficiente do armazenamento em todos os aplicativos.

Calendário

O conteúdo gerado por IA pode estar incorreto.



Tipos de armazenamento definidos por software

O armazenamento definido por software inclui uma gama de tecnologias que desvinculam o gerenciamento e os serviços de armazenamento do hardware subjacente. Para aplicações de IA, sistemas de arquivos escaláveis ​​e armazenamento de objetos são geralmente os mais adequados devido à sua capacidade de lidar com eficiência com grandes conjuntos de dados distribuídos. A infraestrutura hiperconvergente também é vantajosa, especialmente em ambientes de ponta ou locais, e fornece uma plataforma unificada que pode suportar demandas intensivas de IA.

Interface gráfica do usuário, Texto

O conteúdo gerado por IA pode estar incorreto.

Interface gráfica do usuário, Texto, Aplicativo

O conteúdo gerado por IA pode estar incorreto.

Interface gráfica do usuário, Texto, Aplicativo, Email

O conteúdo gerado por IA pode estar incorreto.

Interface gráfica do usuário, Texto

O conteúdo gerado por IA pode estar incorreto.

Interface gráfica do usuário, Texto, Aplicativo

O conteúdo gerado por IA pode estar incorreto.

Interface gráfica do usuário, Texto

O conteúdo gerado por IA pode estar incorreto.

Interface gráfica do usuário, Texto, Aplicativo

O conteúdo gerado por IA pode estar incorreto.

Interface gráfica do usuário, Texto

O conteúdo gerado por IA pode estar incorreto.

Cada solução de armazenamento definida por software oferece pontos fortes distintos, é adaptada para casos de uso específicos e fornece flexibilidade e escalabilidade em diversos ambientes de TI.

Interface gráfica do usuário, Texto, Aplicativo

O conteúdo gerado por IA pode estar incorreto.

Resumo

Parabéns por concluir o treinamento *de Recursos de Armazenamento* !

Este treinamento abordou as principais estratégias de armazenamento de dados e melhores práticas que são essenciais para o gerenciamento eficiente de informações, especialmente para IA, que exige soluções de armazenamento especializadas.

O evento explorou o armazenamento definido por software, que abstrai o hardware físico em um pool de recursos unificado para simplificar o gerenciamento e aumentar a flexibilidade. Também discutiu o Fibre Channel para conexões SAN de alta velocidade e baixa latência e seu aprimoramento, o FCoE, que reduz custos enquanto mantém o desempenho.

Para acesso de dados em alta velocidade, o NVMe melhora o desempenho do SSD com menor latência e maior rendimento, enquanto o NVMe-oF estende essa eficiência pelas redes para melhorar a escalabilidade e o desempenho.

Agora que você concluiu o treinamento, reflita sobre as seguintes perguntas:

* Como o uso de NVMe e NVMe-oF melhora o desempenho de cargas de trabalho de IA que exigem processamento de dados em tempo real?
* Como a combinação de armazenamento definido por software com FCoE pode melhorar o gerenciamento de armazenamento para aplicativos de IA?