

Capítulo 1

Introdução ao Armazenamento e Gerenciamento de Informações

A informação é cada vez mais importante no nosso cotidiano. Vivemos hoje em um mundo de comandos e demandas, precisando de informações a todo momento. Acessamos a Internet todos os dias para fazer consultas, participar de redes sociais, enviar e receber e-mails, compartilhar imagens e vídeos, além de outras atividades. Devido ao número crescente de dispositivos que geram conteúdo, a quantidade de informações produzidas por indivíduos é maior do que a criada por empresas. As informações originadas por indivíduos se tornam mais valiosas quando compartilhadas com outras pessoas. Ao serem produzidas, elas permanecem localmente

em dispositivos como telefones celulares, câmeras e laptops. Para serem compartilhadas, precisam ser enviadas através de redes a data centers. É interessante observar que, embora a maioria dos dados seja criada por indivíduos, ela é armazenada e gerenciada por uma quantidade relativamente pequena de organizações. A Figura 1-1 mostra esse círculo virtuoso das informações.

A importância, a dependência e o volume de informações no mundo corporativo também continuam crescendo muito. As empresas dependem de acesso rápido e confiável a informações cruciais para o seu sucesso. Alguns dos aplicativos comerciais que processam dados incluem reservas de passagens aéreas, sistemas de contas telefônicas, comércio eletrônico, caixas eletrônicos de bancos 24 horas, projetos de produtos, gerenciamento de estoques, arquivamento de correio eletrônico, portais Web, registros médicos de pacientes, cartões de crédito, ciências da vida e mercados de capitais globais.

À medida que as informações têm uma importância cada vez maior para as empresas, aumentam os desafios relacionados à proteção e ao gerenciamento de dados. O volume de dados gerenciado pelas empresas as fez criar estratégias de

CONCEITOS-CHAVE

Dados e informações

Dados estruturados e não estruturados

Tecnologia e arquitetura de armazenamento

Principais elementos de um data center

Gerenciamento da informação

Gerenciamento do ciclo de vida da informação

classificação de acordo com o valor dos dados e também regras para gerenciá-los durante seu ciclo de vida. Essas estratégias trazem vantagens operacionais e reguladoras em nível empresarial e também vantagens gerenciais em nível operacional para a organização.

Os data centers agora veem o armazenamento de informações como um de seus elementos principais, junto aos aplicativos, bancos de dados, sistemas operacionais e redes. A tecnologia de armazenamento continua a evoluir, com avanços técnicos que oferecem níveis cada vez mais altos de disponibilidade, segurança, escalabilidade, desempenho, integridade, capacidade e gerenciabilidade.

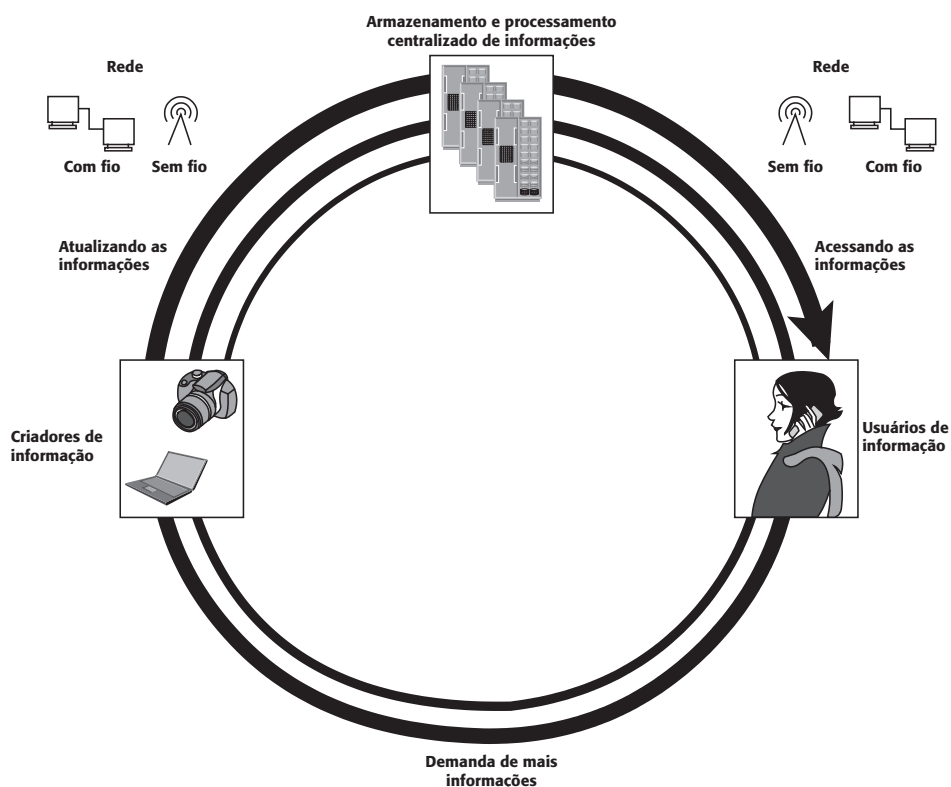


Figura 1-1 O círculo virtuoso da informação.

Este capítulo descreve a evolução da arquitetura de armazenamento de informações dos modelos simples de ligação direta às complexas topologias em rede. Ele introduz a estratégia de gerenciamento do ciclo de vida da informação (ILM, Information Lifecycle Management), que alinha a infraestrutura da tecnologia de informação (TI) às prioridades da empresa.

1.1 O armazenamento de informações

As empresas usam dados para obter informações cruciais para as suas operações diárias. O armazenamento é um repositório que permite aos usuários guardar e buscar esses dados digitais.

1.1.1 Dados

Dados são um conjunto de fatos em estado bruto a partir dos quais conclusões podem ser tiradas. Cartas escritas à mão, livros impressos, fotos de família, filmes em fita de vídeo, cópias assinadas de papéis de hipotecas, livro-razão de um banco e cadernetas bancárias do titular de uma conta são exemplos de dados.

Antes do advento dos computadores, os métodos e procedimentos adotados para a criação e compartilhamento de dados eram limitados a menos formas, como papel e filme. Atualmente, os mesmos dados podem ser convertidos para meios mais convenientes, como uma mensagem de correio eletrônico, um livro eletrônico, uma imagem na forma de bitmaps ou um filme digital. Esses dados podem ser gerados em um computador e armazenados em fluxos de 0s e 1s, conforme mostrado na Figura 1-2. Dados nesse formato são chamados de *dados digitais* e são acessíveis pelo usuário apenas após serem processados por um computador.

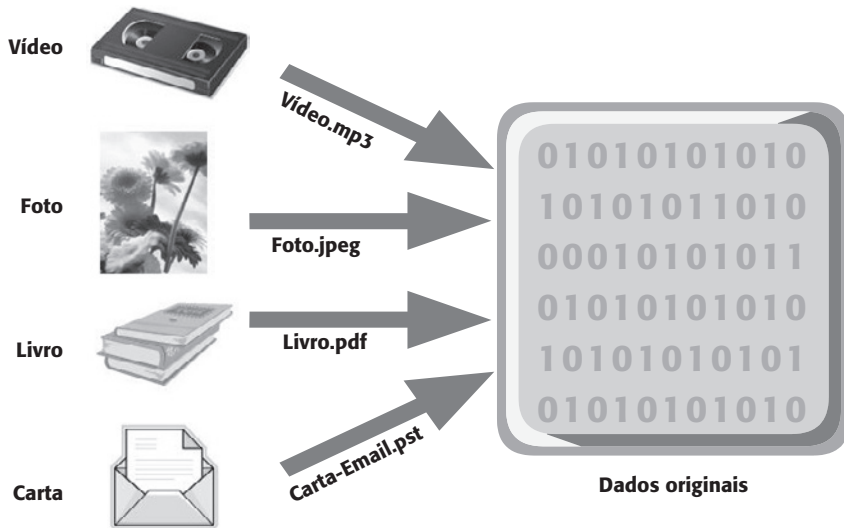


Figura 1-2 Dados digitais.

Com o avanço nas tecnologias computacionais e de comunicações, o volume de geração e compartilhamento de dados cresceu exponencialmente. A seguir,

vemos uma lista de alguns dos fatores que contribuíram para o aumento dos dados digitais:

- **Ampliação da capacidade de processamento de dados:** Os computadores atuais proporcionam um aumento significativo na capacidade de processamento e de armazenamento. Isso permite a conversão de diversos tipos de conteúdo e mídias dos seus formatos convencionais para formatos digitais.
- **Menor custo do armazenamento digital:** Os avanços tecnológicos e a diminuição do custo dos dispositivos de armazenamento permitiram soluções de baixo custo e incentivaram o desenvolvimento de dispositivos de armazenamento de dados mais baratos. Esse benefício nos custos aumentou a taxa de geração e armazenamento de dados.
- **Tecnologias de comunicação mais rápidas e acessíveis:** A velocidade de compartilhamento de dados digitais é agora muito maior do que a das abordagens tradicionais. Uma carta escrita à mão pode levar uma semana para chegar ao destinatário, já uma mensagem de e-mail demora apenas alguns segundos para chegar ao destino.

Formas baratas e fáceis de criar, coletar e armazenar todos os tipos de dados, junto a necessidades individuais e empresariais cada vez maiores, levaram ao crescimento acelerado do volume de dados, o que foi popularmente chamado de *explosão de dados*. Dados têm diferentes objetivos e importância, de modo que tanto os indivíduos quanto as empresas contribuíram em diversas proporções para essa explosão.

A importância dos dados varia com o tempo. A maior parte dos dados criados tem importância em curto prazo, mas se torna menos valiosa com o passar do tempo. Isso determina os tipos usados de soluções de armazenamento de dados, que são estocados em diversos tipos de dispositivos de armazenamento, como discos rígidos, CDs, DVDs ou drives de memória flash USB (Universal Serial Bus).

EXEMPLO DE PESQUISA E DADOS EMPRESARIAIS



- **Sismologia:** Envolve a coleta de dados relacionados a diversas fontes e parâmetros de terremotos, além de outros dados relevantes que precisam ser processados para derivar informações significativas.
- **Dados de produtos:** Incluem dados relacionados a diversos aspectos de um produto, como o seu estoque, descrição, preço, disponibilidade e vendas.
- **Dados de clientes:** Combinação de dados relacionados aos clientes de uma empresa, como detalhes de pedidos, endereços para remessa e históricos de compras.
- **Dados médicos:** Dados relacionados ao sistema de saúde, como históricos de pacientes, imagens radiológicas, detalhes de medicações e outros tratamentos, além de informações sobre seguros.

As empresas geram grandes quantidades de dados e depois extraem deles informações significativas para obter benefícios econômicos. Portanto, precisam guardar e assegurar a disponibilidade desses dados por um longo período. Além

disso, eles podem variar quanto à sua importância e talvez requeiram gerenciamento especial. Por exemplo, requisitos legais e regulatórios exigem que os bancos mantenham informações sobre contas de seus clientes de forma precisa e segura. Algumas empresas lidam com os dados de milhões de clientes e garantem a segurança e integridade dos mesmos por um longo período. Isso requer dispositivos com alta capacidade de armazenamento de dados e com recursos ampliados de segurança que possam mantê-los por um bom tempo.

1.1.2 Tipos de dados

Os dados podem ser classificados como estruturados ou não estruturados (veja a Figura 1-3), com base em como eles são armazenados e gerenciados. Dados estruturados são organizados em linhas e colunas em um formato definido de forma rígida, de modo que os aplicativos possam recuperá-los e processá-los com eficiência. Dados estruturados geralmente são armazenados com o uso de um sistema de gerenciamento de bancos de dados (SGBD).

Os dados são considerados como não estruturados se os seus elementos não puderem ser armazenados como linhas e colunas, sendo, portanto, difíceis de consultar e recuperar através de aplicativos empresariais. Por exemplo, contatos de clientes podem ser armazenados de diversas formas, como em notas autoadesivas, mensagens de e-mail, cartões de apresentação, ou até mesmo em arquivos de formato digital como .doc, .txt e .pdf. Devido à sua natureza não estruturada, é difícil pesquisar por meio de um aplicativo de gerenciamento de relacionamento com os clientes. Dados não estruturados podem não ter os componentes necessários para uma identificação específica de algum tipo de processamento ou interpretação. As empresas se preocupam principalmente com o gerenciamento de dados não estruturados, pois mais de 80% dos dados corporativos não são estruturados e requerem maior espaço e mais gerenciamento.

1.1.3 Informações

Dados, sejam eles estruturados ou não, não satisfazem os objetivos de indivíduos ou empresas a menos que sejam apresentados de uma forma que tenha algum significado. As empresas precisam analisar esses dados para que eles tenham valor. *Informação* é a inteligência e o conhecimento derivados dos dados.

As empresas analisam dados brutos para identificar tendências significativas. Com base nessas tendências, podem planejar ou modificar suas estratégias. Por exemplo, um varejista identifica os produtos e marcas preferidos pelos clientes analisando seus padrões de compra e mantendo um estoque desses produtos.

A análise eficaz de dados não apenas estende seus benefícios para as atividades de negócio existentes, mas também cria o potencial para novas oportunidades de negócios usando as informações de forma criativa. Um portal de empregos é um exemplo. Para atingir um leque maior de prováveis empregadores, quem procura um emprego coloca seu currículo em diversos sites que ofereçam dispositivos de procura de emprego. Estes sites guardam os currículos e os disponibilizam em locais acessíveis a possíveis empregadores. Além disso, as empresas anunciam cargos disponíveis em sites de procura de empregos. Softwares de oferta de empregos associam as palavras-chave de

currículos com as de ofertas de empregos. Dessa forma, o mecanismo de busca usa os dados e os transforma em informação para empregadores e pessoas que procuram emprego.

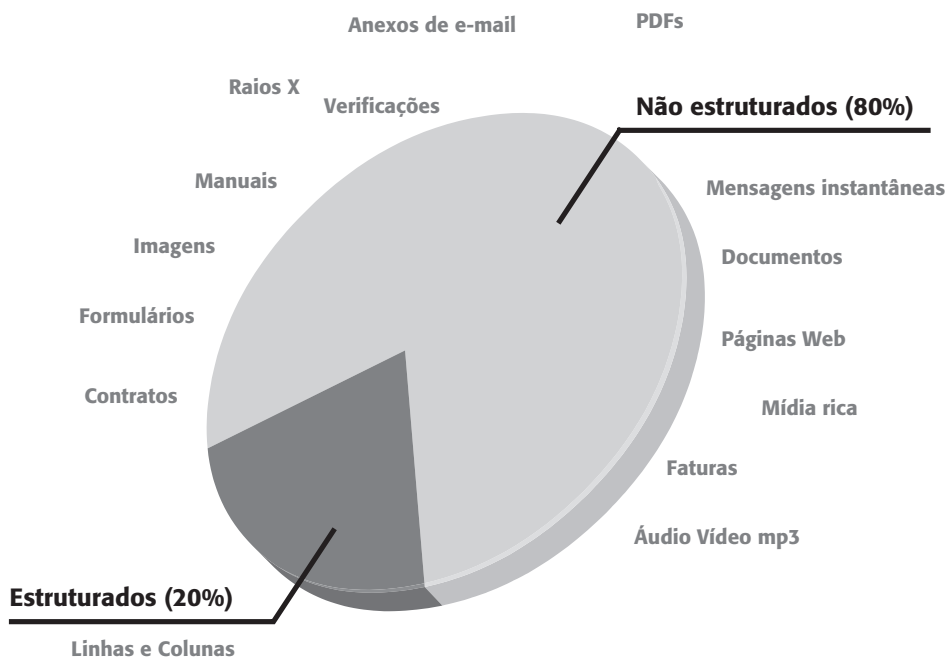


Figura 1-3 Tipos de dados.

Por serem cruciais para o sucesso de uma empresa, há uma preocupação constante com a disponibilidade e proteção das informações. Obrigações legais, regulatórias e contratuais relacionadas à disponibilidade e proteção dos dados só aumentam estas preocupações. Interrupções em indústrias-chave, como as de serviços financeiros, telecomunicações, manufatura, varejista e de energia, custam milhões de dólares por hora.

1.1.4 Armazenamento

Dados criados por indivíduos ou empresas devem ser armazenados a fim de que sejam facilmente acessados para posterior processamento. Em um ambiente computacional, dispositivos projetados para armazenar dados são chamados de *dispositivos de armazenamento*. O tipo de armazenamento usado varia de acordo com os tipos de dados e a velocidade em que são criados e usados. Dispositivos como a memória de um telefone celular ou câmera digital, DVDs, CD-ROMs e discos rígidos em computadores pessoais são exemplos de dispositivos de armazenamento.

As empresas possuem diversas opções disponíveis para armazenar dados, incluindo discos rígidos internos, fitas e disk arrays externos.

1.2 A evolução da tecnologia e arquitetura de armazenamento

Historicamente, as organizações possuem computadores centralizados (mainframes) e dispositivos de armazenamento de informações (rolos de fita e blocos de discos) nos seus data centers. A evolução dos sistemas abertos e sua viabilidade financeira, além da facilidade de implantação que eles oferecem, possibilitaram que departamentos/unidades das empresas tivessem seus próprios servidores e armazenamento. Em implementações iniciais de sistemas abertos, o armazenamento geralmente era interno no servidor.

A proliferação de servidores departamentais nas empresas resultou em ilhas de informações desprotegidas, não gerenciadas e fragmentadas, além de custos operacionais cada vez mais altos. Originalmente, havia políticas e processos muito limitados para o gerenciamento desses servidores e dos dados envolvidos. Para superar essas dificuldades, a tecnologia de armazenamento evoluiu de armazenamento interno não inteligente para armazenamento em rede inteligente (veja a Figura 1-4). Os destaques na evolução desta tecnologia incluem:

- **RAID (*Redundant Array of Independent Disks*):** Tecnologia desenvolvida para direcionar os requisitos de custo, desempenho e disponibilidade de dados. Ela continua a se desenvolver e é usada em todas as arquiteturas de armazenamento, como DAS, SAN, etc.
- **DAS (*Direct-Attached storage*):** Tipo de armazenamento que se conecta diretamente ao servidor (host) ou a um grupo de servidores em um cluster. O armazenamento pode ser interno ou externo ao servidor. O DAS externo diminui os problemas de capacidade limitada de armazenamento interno.
- **SAN (*Storage Area Network*):** Rede FC (*Fibre Channel*) dedicada e de alto desempenho para facilitar a comunicação em *nível de blocos* entre os servidores e o armazenamento. Este é particionado e atribuído a um servidor para acessar os dados. A SAN oferece vantagens de escalabilidade, disponibilidade, desempenho e custo se comparado com servidores DAS.
- **NAS (*Network-Attached Storage*):** Armazenamento dedicado a aplicativos de file serving. Diferentemente de uma SAN, conecta-se a uma rede de comunicações existente (LAN) e fornece acesso a arquivos para clientes heterogêneos. Por ser construído propositalmente para aplicativos de servidor de arquivos, oferece maior escalabilidade, disponibilidade, desempenho e vantagens de custo comparado a servidores de arquivos de propósito geral.
- **IP-SAN (*Internet Protocol SAN*):** Uma das evoluções mais recentes na arquitetura de armazenamento, o IP-SAN é uma convergência de tecnologias usadas em SAN e NAS. O IP-SAN fornece comunicação em nível de blocos através de uma rede local (LAN) ou de longa distância (WAN), resultando em uma maior consolidação e disponibilidade dos dados.

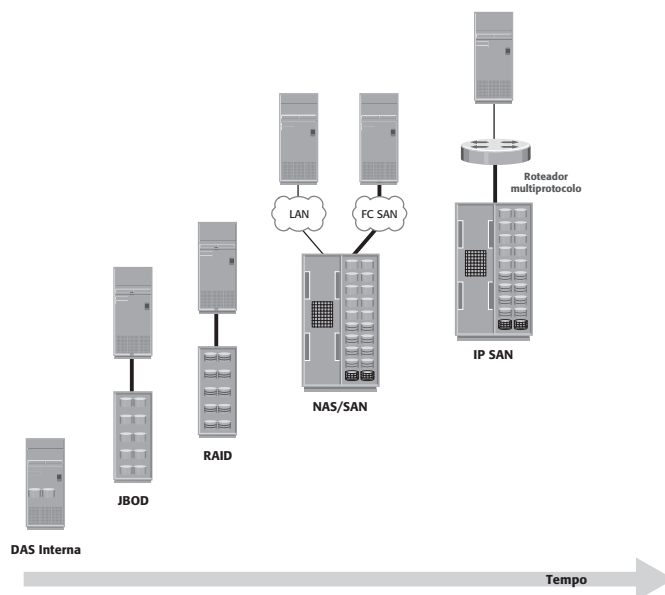


Figura 1-4 A evolução das arquiteturas de armazenamento.

A tecnologia e a arquitetura de armazenamento continuam a evoluir, o que permite às organizações consolidar, proteger, otimizar e alavancar seus dados para obter maior retorno em ativos de informação.

1.3 Estrutura do data center

As organizações mantêm centrais de dados a fim de fornecer processamento centralizado dos dados para toda a empresa. Centrais de dados armazenam e gerenciam grandes quantidades de dados de missão crítica. Sua infraestrutura inclui computadores, sistemas de armazenamento, dispositivos de rede, fontes reservas de alimentação dedicadas e controles ambientais (como condicionadores de ar e sistemas anti-incêndio).

Organizações de grande porte muitas vezes conservam mais de uma central de dados para distribuir as cargas de processamento dos mesmos e fornecer cópias de segurança no evento de um desastre. Os requisitos de armazenamento de dados são satisfeitos por uma combinação de diversas arquiteturas de armazenamento.

1.3.1 Elementos principais

Cinco elementos principais são essenciais para a funcionalidade básica de um data center:

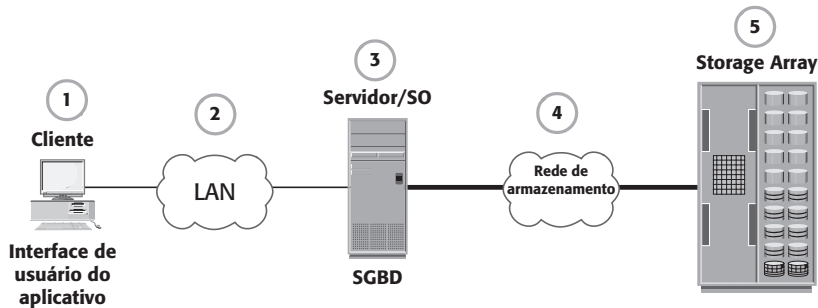
- **Aplicativo:** Trata-se de um programa de computador que fornece a lógica das operações computacionais. Os aplicativos, como um sistema de pro-

cessamento de ordem, podem ser colocados em camadas em um banco de dados, que, por sua vez, usa os serviços de um sistema operacional para executar operações de leitura/gravação em dispositivos de armazenamento.

- **Banco de dados:** De modo geral, um sistema de gerenciamento de bancos de dados (SGBD) fornece uma forma estruturada de armazenamento de dados na forma de tabelas lógicas que são inter-relacionadas. Um SGBD otimiza o armazenamento e a recuperação dos dados.
- **Servidor e sistema operacional:** Caracteriza-se por ser plataforma computacional que executa aplicativos e bancos de dados.
- **Rede:** Refere-se a um caminho de dados que facilita a comunicação entre os clientes e os servidores ou entre os servidores e o armazenamento.
- **Storage array:** Trata-se de um dispositivo que armazena dados de forma persistente para uso posterior.

Esses elementos principais geralmente são visualizados e gerenciados como entidades separadas, mas todos devem trabalhar juntos para atender aos requisitos de processamento de dados.

A Figura 1-5 mostra um exemplo de sistema de processamento de pedidos que envolve os cinco elementos básicos de um data center e ilustra a sua funcionalidade em um processo de negócio.



- 1 Um cliente faz um pedido através da IUA do software de processamento de pedidos, localizado no computador do cliente.
- 2 O cliente se conecta com o servidor através da LAN e acessa o SGBD localizado no servidor para atualizar as informações relevantes como o nome do cliente, endereço, forma de pagamento, produtos pedidos e quantidade encomendada.
- 3 O SGBD usa o sistema operacional do servidor para ler e gravar estes dados no banco de dados localizado nos discos físicos do storage array.
- 4 A rede de armazenamento estabelece a comunicação entre o servidor e o storage array e transporta os comandos de leitura ou gravação entre eles.
- 5 O storage array, após receber os comandos de leitura e gravação, executa as operações necessárias para armazenar os dados em discos físicos.

Figura 1-5 Exemplo de sistema de processamento de pedidos.

1.3.2 Requisitos-chave para elementos de centrais de dados

A operação ininterrupta do data center é crucial para a sobrevivência e o sucesso de uma empresa. É necessário que exista uma infraestrutura confiável para garantir que os dados estejam sempre acessíveis. Embora os requisitos, mostrados na Figura 1-6, sejam aplicáveis a todos os elementos da infraestrutura do data center, nosso foco são os sistemas de armazenamento. As diversas tecnologias e soluções para atender a esses requisitos são explicadas neste livro.

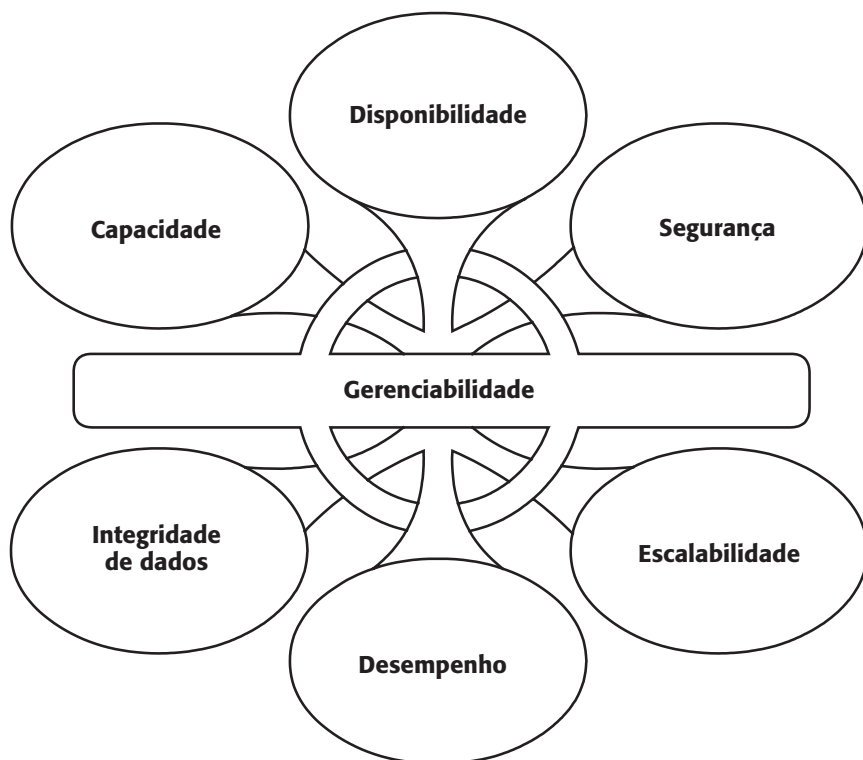


Figura 1-6 Características-chave dos elementos dos data centers.

- **Disponibilidade:** Todos os elementos dos data centers devem ser projetados para garantir a acessibilidade. A empresa poderá sofrer um forte efeito negativo caso os usuários tenham dificuldade em acessar os dados.
- **Segurança:** Devem ser estabelecidas políticas, procedimentos e integração apropriados dos elementos principais do data center que impedirão o acesso não autorizado às informações. Além das medidas de segurança de acesso dos clientes, mecanismos específicos devem permitir aos servidores acessar apenas seus recursos alocados em storage arrays.
- **Escalabilidade:** As operações do data center devem alocar recursos para processamento ou armazenamento adicionais sob demanda, sem interrom-

per as operações de negócio. O crescimento da empresa muitas vezes requer a instalação de mais servidores, novos aplicativos e bancos de dados extras. A solução de armazenamento deve crescer junto com a empresa.

- **Desempenho:** Todos os elementos principais do data center devem fornecer serviço e desempenho otimizados a todas as solicitações de processamento em alta velocidade. A infraestrutura deve ser capaz de atender aos requisitos de desempenho.
- **Integridade de dados:** Refere-se a mecanismos como códigos de correção de erros e bits de paridade que asseguram que os dados sejam gravados no disco exatamente como foram recebidos. Qualquer variação durante a sua recuperação indica que estão corrompidos, podendo afetar as operações da organização.
- **Capacidade:** Operações nos data centers requerem recursos adequados para armazenar e processar grandes quantidades de dados de forma eficiente. Quando os requisitos de capacidade aumentam, a central de dados deve ser capaz de fornecer capacidade adicional sem interromper a disponibilidade ou, pelo menos, com uma interrupção mínima. A capacidade pode ser gerenciada realocando recursos existentes, em vez de adicionar novos recursos.
- **Gerenciabilidade:** Um data center deve executar todas as operações e atividades da maneira mais eficiente. Gerenciabilidade pode ser obtida com a automação e redução de intervenção humana (manual) nas tarefas comuns.

1.3.3 Gerenciando a infraestrutura de armazenamento

Gerenciar um data center moderno e complexo envolve muitas tarefas. As atividades-chave do gerenciamento abrangem:

- *Monitoramento* é a coleta contínua de informações e a revisão da infraestrutura inteira do data center. Os aspectos de um data center monitorado incluem segurança, desempenho, acessibilidade e capacidade.
- *Geração de relatórios* é feita periodicamente com base no desempenho, na capacidade e na utilização dos recursos. Tarefas relacionadas a relatórios auxiliam a estabelecer as justificativas do negócio e o retorno dos custos associados às operações do data center.
- *Provisionamento* é o processo de fornecer hardware, software e outros recursos necessários para o funcionamento de um data center. As atividades de provisionamento incluem o planejamento de capacidade e recursos. O *planejamento de capacidade* assegura que as necessidades futuras do usuário e do aplicativo sejam satisfeitas da forma mais controlada e eficaz quanto aos custos. O *planejamento de recursos* é o processo de avaliação e identificação dos recursos necessários, como pessoal, instalações (local) e tecnologia. Este planejamento garante que os recursos adequados estejam disponíveis para atender aos requisitos dos usuários e dos aplicativos.

Por exemplo, a utilização da capacidade de armazenamento alocado de um aplicativo pode ser monitorada. Assim que ela atingir um valor crítico, a capa-

cidade adicional de armazenamento poderá ser provisionada para o aplicativo. Se a utilização da capacidade de armazenamento for monitorada e informada apropriadamente, o crescimento do negócio pode ser percebido e futuros requisitos de capacidade podem ser previstos. Isso auxilia uma política pró-ativa de gerenciamento de dados.

1.4 Desafios-chave no gerenciamento de informações

Para planejar uma política eficaz de gerenciamento, as empresas precisam levar em consideração os seguintes desafios-chave apresentados pelo gerenciamento de informações:

- **O universo digital em crescimento explosivo:** A taxa de crescimento da quantidade de informações está se expandindo exponencialmente. A duplicação de dados para assegurar alta disponibilidade e readaptação também contribuiu para o grande aumento no crescimento do volume de informações.
- **O aumento na dependência das informações:** O uso estratégico das informações desempenha um papel importante na determinação do sucesso de uma empresa e fornece vantagem competitiva no mercado.
- **O valor inconstante das informações:** Informações que são valiosas hoje podem se tornar menos importantes no futuro. O valor das informações, muitas vezes, varia com o decorrer do tempo.

O estabelecimento de uma política que satisfaça a esses desafios envolve a compreensão do valor das informações durante seu ciclo de vida.

1.5 O ciclo de vida da informação

O *ciclo de vida da informação* é a “mudança no valor da informação” com o decorrer do tempo. Quando os dados são criados, muitas vezes possuem seu valor mais alto e são usados com frequência. À medida que o tempo passa, são acessados com menor regularidade e têm menos valor para a organização. Entender o ciclo de vida da informação ajuda a implantar a infraestrutura de armazenamento apropriada, de acordo com as mudanças no valor dessa informação.

Por exemplo, em um aplicativo de pedidos de venda, o valor das informações muda do momento em que o pedido é feito até a garantia expirar (veja a Figura 1-7). O valor das informações é mais alto quando uma empresa recebe um novo pedido de venda e o processa para entregar o produto. Após o pedido ter sido satisfeito, os dados do cliente ou do pedido não precisam mais ficar disponíveis para acesso em tempo real. A empresa pode transferir esses dados para o armazenamento secundário mais barato com requisitos menores de acessibilidade e disponibilidade, a menos ou até que uma solicitação de garantia ou outro evento gere sua necessidade. Depois de a garantia expirar, a empresa pode arquivar ou descartar os dados, obtendo espaço para outras informações de maior valor.

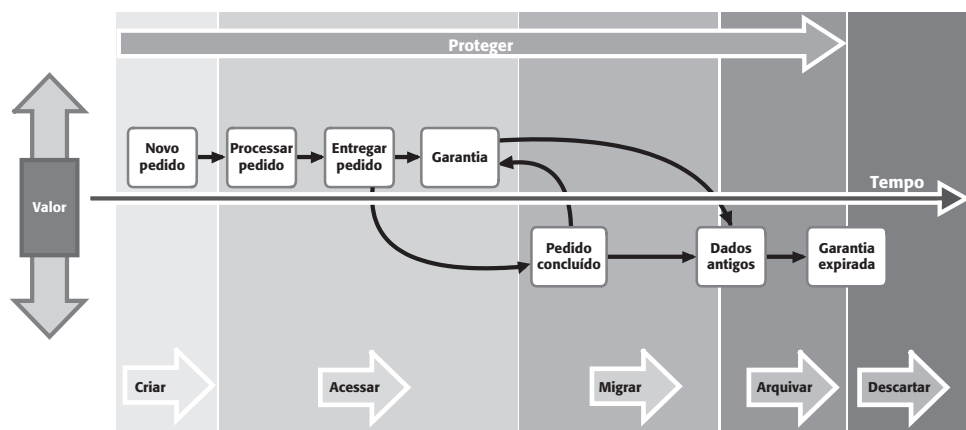


Figura 1-7 Alterando o valor das informações da ordem de vendas.

1.5.1 Gerenciamento do ciclo de vida da informação

As empresas modernas precisam que seus dados estejam protegidos e disponíveis em tempo integral. Os data centers podem conseguir isso com o uso otimizado e apropriado da infraestrutura de armazenamento. Uma política eficaz de gerenciamento de informações é necessária para dar suporte a esta infraestrutura e potencializar seus benefícios.

O *gerenciamento do ciclo de vida da informação (ILM, Information Lifecycle Management)* é uma estratégia pró-ativa que permite a uma organização de TI gerenciar de modo eficaz os dados por todo o seu ciclo de vida, baseada em políticas de negócio preestabelecidas. Uma estratégia de ILM deve incluir as seguintes características:

- **Centrada no negócio:** Deve estar integrada com os processos, aplicativos e iniciativas-chave do negócio para poder lidar tanto com o crescimento atual quanto com o futuro das informações.
- **Gerenciada centralmente:** Todas as informações de um negócio precisam estar sob a supervisão da estratégia de ILM.
- **Baseada em políticas:** A implementação do ILM não deve ficar restrita a alguns departamentos. O ILM deve ser implantado como uma política e abranger todos os aplicativos, processos e recursos da empresa.
- **Heterogênea:** Uma estratégia de ILM precisa levar em consideração todos os tipos de plataformas de armazenamento e sistemas operacionais.
- **Otimizada:** Devido à variação no valor das informações, uma estratégia de ILM deve considerar os diferentes requisitos de armazenamento e alocar recursos de armazenamento baseada no valor das informações para a empresa.

ARMAZENAMENTO HIERÁRQUICO

O armazenamento hierárquico é uma abordagem para definir diversos níveis de armazenamento com o objetivo de reduzir os custos totais. Cada camada possui diferentes níveis de proteção, desempenho, frequência de acesso aos dados e outras considerações. As informações são armazenadas e movidas entre as diferentes camadas com base no seu valor com o decorrer do tempo. Por exemplo, informações de missão crítica e mais acessadas podem ser armazenadas na Camada 1, que consiste em um meio de alto desempenho com o maior nível de proteção. Dados com acesso intermediário em termos de frequência e outros dados importantes são armazenados na Camada 2, que pode ser um meio não tão oneroso, com desempenho e proteção moderados. Informações raramente acessadas ou específicas a eventos podem ser armazenadas em camadas inferiores.

1.5.2 Implementação de ILM

O processo de desenvolvimento de uma estratégia de ILM inclui quatro atividades – classificação, implementação, gerenciamento e organização:

- *Classificação* dos dados e aplicativos com base nas regras e políticas do negócio para permitir o tratamento diferenciado das informações.
- *Implementação* de políticas através do uso de ferramentas de gerenciamento de informações, começando pela criação de dados e terminando com sua eliminação.
- *Gerenciamento* do ambiente usando ferramentas integradas para reduzir a complexidade operacional.
- *Organização* dos recursos de armazenamento hierárquico para alinhar os recursos com as classes de dados e armazenar as informações no tipo correto de infraestrutura com base no seu valor atual.

A implementação do ILM por uma empresa é um processo contínuo. A Figura 1-8 ilustra um mapa com três etapas para ILMs que abrangem toda a empresa.

As Etapas 1 e 2 têm como objetivo implementar o ILM de uma forma limitada a partir de alguns aplicativos críticos para a empresa. Na Etapa 1, o objetivo é implantar um ambiente de rede de armazenamento. Arquiteturas de armazenamento oferecem níveis variáveis de proteção e desempenho e isso serve como base para o futuro gerenciamento das informações baseado em políticas nas Etapas 2 e 3. O valor das plataformas de armazenamento hierárquico pode ser explorado pela alocação de recursos de armazenamento apropriados para os aplicativos com base no valor das informações processadas.

A Etapa 2 leva o ILM ao próximo nível, com a aplicação ou classificação detalhada dos dados e a conexão da infraestrutura de armazenamento com as políticas de negócio. Estas classificações e as políticas resultantes podem ser executadas automaticamente, por meio de ferramentas para um ou mais aplicativos, resultando em um melhor gerenciamento e alocação otimizada dos recursos de armazenamento.

A Etapa 3 da implementação é automatizar mais os aplicativos ou a classificação de dados e atividades de gerenciamento de políticas para atingir um conjunto maior de aplicativos corporativos.

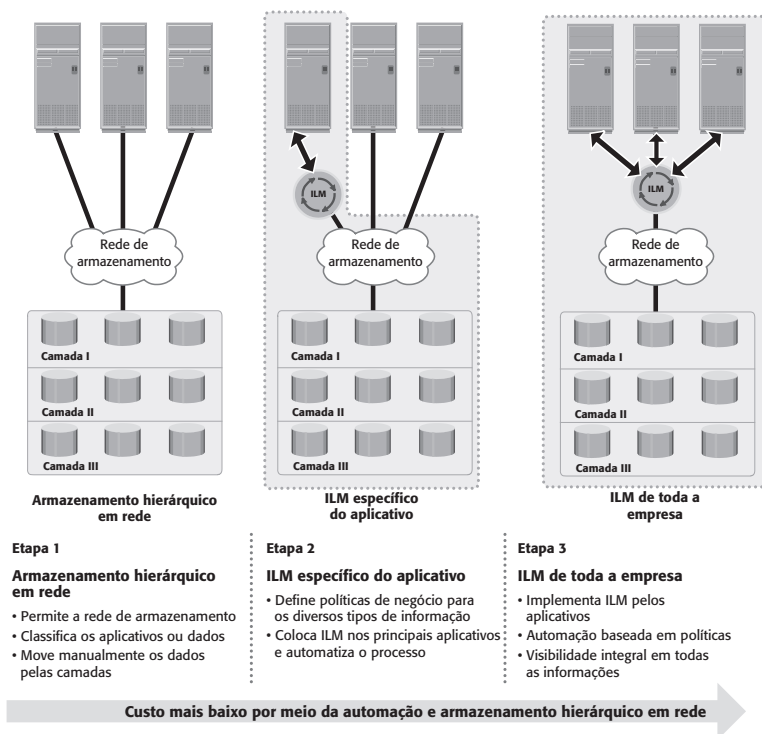


Figura 1-8 Implementação de ILM.

1.5.3 Vantagens do ILM

Implementar uma estratégia de ILM apresenta as seguintes vantagens que abordam diretamente os desafios do gerenciamento de informações:

- *Utilização avançada* pelo uso de plataformas de armazenamento hierárquico e maior visibilidade de todas as informações da empresa.
- *Gerenciamento simplificado* pela integração de etapas do processo e interfaces com ferramentas individuais e pelo aumento na automação.
- *Maior variedade de opções* para cópias de segurança e recuperação com o objetivo de equilibrar a necessidade de continuidade do negócio.
- *Compatibilidade na manutenção*, sabendo-se quais dados precisam ser protegidos por qual período de tempo.

- *Menor custo total de propriedade (TCO, Total Cost of Ownership)* alinhando os custos de infraestrutura e gerenciamento com o valor da informação. Como consequência, não há desperdício de recursos ou complexidade no gerenciamento de dados de baixo valor com o custo de dados de alto valor.

Resumo

Este capítulo descreveu a importância de dados, informação e infraestrutura de armazenamento. A satisfação dos requisitos atuais de armazenamento começa com a compreensão dos tipos de dados, seu valor e requisitos-chave de gerenciamento de um sistema de armazenamento.

Este capítulo também enfatizou a importância da estratégia de ILM, que tem sido adotada pelas empresas para gerenciar as informações de maneira eficaz por todo o ambiente corporativo. O ILM permite que as empresas ganhem vantagem competitiva por meio da classificação, proteção e potencialização das informações.

A evolução das arquiteturas de armazenamento e os elementos principais de um data center relacionados neste capítulo oferecem a fundamentação sobre o armazenamento de informações. O próximo capítulo, por sua vez, explica o ambiente desse sistema de armazenamento.

EXERCÍCIOS

1. Um hospital usa um aplicativo que armazena dados de raios X de pacientes na forma de objetos binários grandes em um banco de dados Oracle. O aplicativo está hospedado em um servidor UNIX e a equipe do hospital acessa os registros de raios X através de um backbone. Um storage array fornece armazenamento para o servidor UNIX, tendo 6 terabytes de capacidade utilizável.
 - Explique os elementos básicos do data center. Quais são os desafios comuns que a equipe de gerenciamento de armazenamento pode encontrar para satisfazer as demandas em nível de serviço da equipe do hospital?
 - Descreva como o valor desses dados de pacientes pode mudar com o decorrer do tempo.
2. Um departamento de projetos de engenharia de uma empresa de grande porte mantém mais de 600 mil desenhos de engenharia que seus projetistas acessam e reutilizam em projetos atuais, modificando-os ou atualizando-os conforme necessário. A equipe de projeto quer acesso instantâneo aos desenhos para seus projetos atuais, mas está restrita no momento por uma infraestrutura que não consegue ser escalada para satisfazer os requisitos de tempo de resposta. A equipe classificou os desenhos como “mais acessados”, “frequentemente acessados”, “ocasionalmente acessados” e “arquivo”.
 - Sugira uma estratégia para o departamento de projetos que otimize a infraestrutura de armazenamento usando o ILM.
 - Explique como você usará o “armazenamento hierárquico” baseado em frequência de acesso.
 - Detalhe os componentes de hardware e software que você precisará para implementar a sua estratégia.
3. O departamento de marketing de uma empresa de médio porte está em expansão. Novos contratados recebem acesso pela rede aos arquivos do departamento. A TI deu ao marketing um drive em rede na LAN, mas ela atinge sua capacidade a cada três semanas. A capacidade atual é de 500 gigabytes (e aumentando), com centenas de arquivos. Os usuários estão reclamando do tempo de resposta e da capacidade da LAN. Como gerente de TI, o que você poderia recomendar para melhorar a situação?
4. Uma empresa de grande porte está analisando uma infraestrutura de armazenamento que seja escalável e forneça alta disponibilidade. A empresa também precisa de desempenho para seus aplicativos de missão crítica. Que topologia de armazenamento você recomendaria (SAN, NAS, IP SAN) e por quê?