

novo programa, o usuário deve executar o aplicativo de instalação no computador, em qualquer pasta ou diretório desejado.

Alguns atalhos:

Tecla Windows + Tab	Chama a Visão de Tarefas
Tecla Windows + A	Chama a Central de Ações
Tecla Windows + D	Mostra o Desktop. (minimiza as janelas de TODOS os monitores)
Tecla Windows + M	Minimiza todas as janelas do monitor atual.
Tecla Windows + E	Abre o Explorador de Arquivos
Tecla Windows + F	Abre o Hub de Comentários
Tecla Windows + H	Chama o recurso Compartilhar
Tecla Windows + I	Chama o recurso Configurações
Tecla Windows + K	Chama o recurso Conectar
Tecla Windows + L	Bloqueia o computador ou troca de usuário.
Tecla Windows + P	Abre a janela para conexão com projetor
Tecla Windows + R	Abre a caixa de diálogos "executar".
Tecla Windows + S	Chama a pesquisa da Cortana
Tecla Windows + X	Abre o mesmo menu que é aberto com o clique com o botão direito sobre o Botão Iniciar, disponibilizando recursos avançados como o Gerenciador de Tarefas, Gerenciador de Dispositivos, dentre outros.

11.1 NOÇÕES DE SISTEMA OPERACIONAL (AMBIENTE LINUX)

1- O que é?

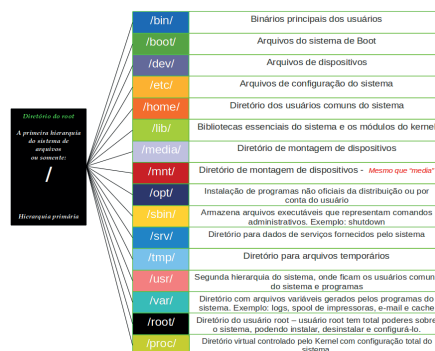
A narrativa da origem do Linux remonta à junção de dois elementos: Linus + Unix. Linus Torvalds, o visionário por trás do Linux, combinou seu nome com o sistema operacional Unix.

Em termos claros:

- **Linux:** Refere-se exclusivamente ao kernel, o núcleo do sistema operacional, que possui acesso privilegiado ao hardware.
- **GNU/Linux:** Indica um sistema operacional completo construído com base no Linux.

Quanto à GPL (GNU General Public License): Esta é a Licença Pública Geral GNU, que acompanha os pacotes distribuídos pelo Projeto GNU. Sendo a licença padrão adotada pelo Linux, ela impede a integração do software em um sistema proprietário e assegura os direitos autorais.

Exemplo: Um grupo de desenvolvedores que estão trabalhando em um sistema de gerenciamento de documentos chamado "DocFlow". Eles optam por licenciar o software sob os termos da GPL para promover a ideia de código aberto e garantir que as liberdades dos usuários sejam preservadas. Os termos da GPL significam que qualquer pessoa pode acessar, modificar e redistribuir o código-fonte do DocFlow. Se um desenvolvedor externo, chamado Maria, quiser melhorar uma funcionalidade ou corrigir um bug no DocFlow, ela pode fazer isso livremente e até distribuir sua versão aprimorada. No entanto, há uma condição: Maria também deve disponibilizar o código-fonte de suas modificações sob os mesmos termos da GPL. Isso garante que as melhorias feitas por ela estejam disponíveis para a comunidade, promovendo a colaboração e o aprimoramento contínuo do software. Outro desenvolvedor, João, que precisa incorporar parte do código do DocFlow em seu próprio projeto, também pode fazê-lo sob a GPL, desde que o projeto resultante seja distribuído como software de código aberto.



11.2 COMANDOS LINUX MAIS IMPORTANTES:

Linux	Exemplo
cd	cd /home/usuario1
ls	ls -al
cp	cp /home/pasta/* /tmp
mv	mv teste.txt casa.txt
mkdir	mkdir Teste1
rm/rmdir	rm -Rf /home/pasta

11.3 PERMISSÕES DE ARQUIVO NO SISTEMA LINUX:

1- O que é?

As permissões de arquivo no sistema Linux são elementos cruciais para controlar o acesso e a segurança dos arquivos e diretórios. Cada arquivo ou diretório possui atributos de permissão que definem quem pode realizar quais operações sobre esses objetos.

2- Como funciona?

Existem três categorias principais de permissões: **proprietário**, **grupo** e **outros**. O proprietário é o usuário que criou o arquivo, o grupo refere-se a um conjunto de usuários com permissões semelhantes, e outros incluem qualquer usuário que não seja o proprietário ou membro do grupo.

As permissões são geralmente representadas por três conjuntos de caracteres: r (leitura), w (escrita) e x (execução), correspondendo ao proprietário, grupo e outros. Além disso, há permissões especiais, como o conjunto de bits SUID (**Set User ID**) e SGID (**Set Group ID**), que concedem temporariamente privilégios de usuário ou grupo durante a execução de um programa.

Os comandos **chmod (change mode)** e **chown (change owner)** são utilizados para modificar permissões e proprietários de arquivos, respectivamente. É fundamental entender e gerenciar adequadamente as permissões de arquivo para garantir a integridade e a segurança do sistema Linux.

12. COMPUTAÇÃO NA NUVEM (CLOUD COMPUTING)

1- O que é?



Questões

Cloud Computing, também conhecida como Computação na Nuvem, abrange uma ampla gama de serviços de hardware e software entregues pela internet. Essas soluções baseadas na nuvem são notáveis por sua notável elasticidade. **Exemplos: Amazon Web Services (AWS); Microsoft Azure; Google Cloud Platform (GCP); Software as a Service (SaaS)**

2- Como funciona?

Dessa forma, é viável “alugar” servidores completos, tanto para processamento de dados e aplicações quanto conjuntos completos com hardware e software integrados, prontos para uso.

As soluções de nuvem são particularmente adequadas para lidar com Big Data, pois esses conjuntos de dados exibem propriedades como volume e velocidade. Em outras palavras, o volume de dados flutua constantemente, tornando imprevisível a demanda por espaço de armazenamento ou capacidade de processamento. Devido à natureza sob demanda desses serviços em nuvem, os custos geralmente são mais baixos para os usuários.

12.1 TIPOS DE NUVENS EM COMPUTAÇÃO

1- O que é?

As **nuvens públicas** são ambientes gerenciados por terceiros, onde várias aplicações compartilham espaço de armazenamento. Apesar de inicialmente parecer ineficiente, se implementadas com foco em desempenho e segurança, as nuvens públicas mantêm transparência quanto à execução simultânea de outras aplicações.

Um benefício notável é a escalabilidade dessas nuvens, evitando a necessidade de aquisição de equipamentos temporários e transferindo riscos de infraestrutura para os provedores. Algumas porções podem ser designadas exclusivamente para um usuário, criando um data center privado virtual com maior visibilidade.

Já as **nuvens privadas** são dedicadas a um único usuário, geralmente uma empresa. Ao contrário de um data center privado virtual, o usuário possui controle total sobre a infraestrutura, exigindo aquisição de novos equipamentos para expansão. São ideais para aplicações permanentes com requisitos específicos.

2- Como funciona?

As **nuvens híbridas** combinam modelos públicos e privados, permitindo que uma nuvem privada amplie seus recursos a par-

tir de uma nuvem pública. Essa flexibilidade mantém níveis de serviço, mesmo com flutuações rápidas na demanda. A complexidade reside na distribuição de aplicações entre ambas as nuvens.

O modelo de nuvens comunitárias envolve a infraestrutura compartilhada por organizações com interesses ou preocupações semelhantes. Podem ser gerenciadas local ou remotamente, proporcionando economia de custos. Esse compartilhamento torna-se economicamente atrativo, permitindo que várias organizações participem e usufruam dos benefícios conjuntos.

13. SISTEMAS DE INFORMAÇÃO

1- O que é?

13.1 TEORIA GERAL DE SISTEMAS

A Teoria Geral de Sistemas (TGS) é um conjunto de princípios que busca compreender a inter-relação entre elementos de um sistema, independentemente de sua natureza específica.

13.2 SISTEMAS DE INFORMAÇÃO

Os sistemas de informação são elementos essenciais para a gestão eficiente e estratégica da informação dentro de organizações. Esses temas são frequentes em concursos e envolvem diversos conceitos-chave:

13.2.1 COMPONENTES DE UM SISTEMA DE INFORMAÇÃO:

Hardware: Equipamentos físicos como computadores, servidores e dispositivos de rede.

Software: Aplicações e programas que permitem o processamento de dados.

Dados: Elementos brutos que, quando processados, tornam-se informações.

Procedimentos: Instruções e diretrizes para operar e manter o sistema.

Pessoas: Usuários e profissionais envolvidos na operação e gestão do sistema.

13.2.2 TIPOS DE SISTEMAS DE INFORMAÇÃO:

Sistemas Transacionais: Registram transações diárias.

Sistemas de Apoio à Decisão: Auxiliam na tomada de decisões gerenciais.

Sistemas de Informação Gerencial (SIG): Fornecem relatórios e análises.

Sistemas Especialistas: Baseados em conhecimento especializado.

Sistemas de Informação Executiva (SIE): Direcionados a executivos.

13.2.3 CICLO DE VIDA DE UM SISTEMA DE INFORMAÇÃO:

Planejamento: Identificação de necessidades e objetivos.

Análise: Entendimento dos requisitos e processos.

Projeto: Desenvolvimento da estrutura e arquitetura do sistema.

Implementação: Construção e instalação do sistema.

Manutenção e Avaliação: Atualização e avaliação contínua.

13.2.4 DESAFIOS EM SISTEMAS DE INFORMAÇÃO:

Segurança da Informação: Proteção contra ameaças e acessos não autorizados.

Integração de Tecnologias: Coexistência de sistemas diferentes.

Atualização Tecnológica: Acompanhamento das inovações.

13.2.5 TENDÊNCIAS E INOVAÇÕES EM SISTEMAS DE INFORMAÇÃO:

Computação em Nuvem: Armazenamento e processamento online.

Big Data: Análise de grandes volumes de dados.

Inteligência Artificial: Uso de algoritmos avançados.

13.3 FASES E ETAPAS DE SISTEMAS DE INFORMAÇÃO

1- O que é?

Identificação de Necessidades e Oportunidades: Nesta fase inicial, são identificadas as demandas e oportunidades que justificam a criação ou atualização de um sistema de informação. Envolve a compreensão das necessidades dos usuários e a análise de como a tecnologia pode atender a essas demandas.

Análise do Sistema Atual: Antes de projetar um novo sistema, é essencial entender o funcionamento do sistema existente. Realiza-se uma análise aprofundada dos processos, dados e requisitos do sistema atual.

Planejamento do Sistema: Nesta etapa, são definidos os objetivos, escopo e recursos necessários para o novo sistema. Inclui a elaboração de um plano estratégico, orçamento e cronograma para o projeto.

Projeto do Sistema: O projeto transforma as informações reunidas na fase de análise em um modelo concreto para o sistema. Define a arquitetura, estrutura de



Questões

dados, interfaces e algoritmos necessários para a implementação.

Implementação do Sistema: É a fase em que o projeto é transformado em um sistema funcional. Inclui codificação, testes unitários e integração de módulos.

Avaliação e Testes: O sistema é submetido a testes rigorosos para garantir que atenda aos requisitos e funcione sem erros. Inclui testes de aceitação do usuário e correções de possíveis defeitos.

Implantação do Sistema: O sistema é instalado e começa a operar em ambiente de produção. Pode ser feita de forma gradual ou total, dependendo do escopo do sistema.

Manutenção e Aprimoramento: Após a implementação, o sistema requer manutenção contínua para corrigir falhas, ajustar funcionalidades e incorporar melhorias. Pode envolver atualizações regulares e suporte técnico aos usuários.

Desativação do Sistema (se necessário): Caso o sistema se torne obsoleto ou não atenda mais às necessidades da organização, pode ser desativado. Envolve o arquivamento de dados e a transição para um novo sistema, se aplicável.

14. TEORIA DA INFORMAÇÃO

1- O que é?

É um campo essencial na ciência da computação e sistemas de informação, frequentemente abordado em concursos. Os pontos fundamentais desse tema:

2- Como funciona?

Entropia: É uma medida da incerteza em um conjunto de informações. Quanto maior a entropia, maior a incerteza. A entropia máxima ocorre quando todos os eventos são igualmente prováveis.

Bit como Unidade de Informação: O bit é a unidade básica de informação. Ele representa um dígito binário, podendo ser 0 ou 1. A quantidade de bits necessários para representar uma mensagem depende da entropia associada a ela.

Codificação de Huffman: É uma técnica de compressão que atribui códigos mais curtos às mensagens mais frequentes e códigos mais longos às menos frequentes. Isso visa reduzir a quantidade de bits necessários para representar a informação.

Teorema da Codificação de Canal: Estabelece limites teóricos para a taxa na qual a

informação pode ser transmitida com erro em um canal de comunicação. Indica a importância da redundância na transmissão de dados.

Capacidade do Canal: Refere-se à quantidade máxima de informação que um canal pode transmitir de forma confiável em um determinado período. A capacidade é influenciada pela largura de banda e pela relação sinal-ruído.

Teorema de Shannon: Desenvolvido por Claude Shannon, é um dos pilares da Teoria da Informação. Ele estabelece limites fundamentais para a compressão de dados e a transmissão de informações, destacando a relação entre entropia e compressão eficiente.

Redundância e Eficiência de Codificação: A redundância é a repetição de informações. Embora possa aumentar a confiabilidade, a eficiência de codificação busca minimizar a redundância para otimizar a transmissão e o armazenamento de dados.

15. BANCO DE DADOS

1- O que é?

Bancos de dados são conjuntos de dados inter-relacionados, sendo que os tradicionais armazenam informações simples, como textos e números, em estruturas fixas e predefinidas.

Em contraste, os bancos de dados modernos permitem o armazenamento de recursos multimídia e outros tipos de dados não estruturados. Em sua essência, um banco de dados é uma coleção de dados relacionados, e essa definição engloba praticamente qualquer conjunto de dados.

2- Como funciona?

Características fundamentais de um banco de dados:

Representação do Mundo Real: Um banco de dados é utilizado para representar aspectos do mundo real, como dados sobre produtos, serviços em estabelecimentos comerciais ou informações relacionadas ao funcionamento de órgãos públicos. Esse cenário do mundo real é denominado de **minimundo** ou **universo de discurso**.

Significado Inerente: Um banco de dados contém dados com significado inerente. Essa característica decorre da representação do mundo real, conferindo um significado implícito às estruturas. Diferentemente, um conjunto aleatório de dados não se qualifica como um banco de dados.

Propósito Específico: Um banco de dados é criado com um propósito específico, atendendo às necessidades de um conjunto predefinido de usuários e aplicações. Geralmente, é destinado a permitir que pessoas e sistemas acessem, leiam ou modifiquem os dados.

O acesso aos bancos de dados não ocorre diretamente; em vez disso, toda a interação é intermediada por um **Sistema de Gerenciamento de Bancos de Dados (SGBD)**.

1- O que é?

Um Sistema de Gerenciamento de Banco de Dados (SGBD) **é um conjunto complexo de programas usado para criar e gerenciar bancos de dados**. Ele oferece ferramentas para gerenciar dados, metadados e fornecer interfaces de acesso para usuários e programas de aplicação acessarem os dados nos bancos de dados. **Exemplo: Imagine uma empresa de comércio eletrônico chamada "TechMart". O TechMart utiliza um SGBD para gerenciar seu extenso inventário de produtos, informações de clientes e histórico de transações. O SGBD da empresa permite aos funcionários adicionar novos produtos ao inventário, atualizar informações de preços e estoque, e rastrear as transações de compra dos clientes. Para facilitar o acesso, o SGBD oferece interfaces intuitivas, permitindo que os funcionários consultem rapidamente as informações necessárias. Além disso, o SGBD é responsável por garantir a integridade dos dados, evitando duplicatas e garantindo que as informações estejam consistentes em todo o sistema. Ele também fornece recursos de segurança, controlando o acesso aos dados com base nas permissões atribuídas a cada usuário.**

2- Como funciona?

Os programas de aplicação enviam consultas em uma linguagem de programação específica para o SGBD. Este, por sua vez, busca os dados solicitados nos bancos de dados e os retorna no formato desejado. Usuários avançados também podem interagir diretamente com o SGBD, enviando consultas em linguagem de programação e visualizando os resultados em uma interface gráfica do SGBD.

Os SGBDs desempenham várias tarefas, como definir, construir, manipular e compartilhar bancos de dados. Esses sistemas são compostos por softwares complexos com diversas funcionalidades.

A função de definição de dados envolve especificar estruturas de armazenamento,



Questões

tipos de dados e restrições no modelo. Já a função de construção refere-se à criação do banco de dados, uma estrutura virtual de armazenamento controlada pelo SGBD. A função de manipulação inclui a recuperação e alteração dos dados e das definições das estruturas.

O Sistema de Bancos de Dados (SBD) é a plataforma que efetivamente nos possibilita trabalhar com Bancos de Dados (BDs). Ao adquirir uma solução de banco de dados, como Oracle ou SQL Server da Microsoft, estamos adquirindo um Sistema de Gerenciamento de Banco de Dados (SGBD), essencial para criar e gerenciar BDs. O SGBD é uma parte integrante do SBD.

Dica da prof: é relevante observar que, frequentemente, as bancas utilizam o termo “banco de dados” em sentido amplo.

Transações referem-se a operações que envolvem um ou mais acessos ao banco de dados, seja para leitura ou gravação de dados. Assim, qualquer ação realizada por usuários ou programas de aplicação para interagir com o BD envolverá uma ou mais transações. Essas transações podem variar em complexidade, desde simples leituras até operações complexas envolvendo diversas leituras, gravações e cálculos.

Nos **Sistemas de Gerenciamento de Bancos de Dados** (SGBDs) relacionais, é comum utilizar uma técnica denominada normalização, que busca reduzir a quantidade de redundâncias no modelo. Contudo, a adoção de um maior ou menor nível de redundância controlada pode depender dos objetivos específicos do sistema.

A redundância controlada de dados implica que o sistema tem conhecimento de múltiplas representações de um objeto no banco de dados. O SGBD é responsável por gerenciar essas diferentes representações, garantindo a sincronização entre elas. Em outras palavras, **o SGBD assegura que um mesmo dado redundante tenha o mesmo valor em todas as suas representações, evitando discrepâncias.**

16. DOCUMENTAÇÃO E PROTOTIPAÇÃO:

1- O que é?

A documentação em projetos de sistemas de informação é crucial para garantir a compreensão, manutenção e evolução do sistema ao longo do tempo. Inclui manuais, diagramas e descrições que elucidam o funcionamento do sistema. A prototipação é uma abordagem iterativa que visa criar versões simplificadas do sistema para validar requisitos e obter feedback dos usuários antes da implementação final. Ambos os as-

pectos são essenciais para o sucesso de projetos de desenvolvimento de sistemas. **Exemplo: uma equipe de desenvolvimento esteja trabalhando em um sistema de gerenciamento de tarefas para uma empresa. Durante a fase inicial, os analistas de sistemas elaboram documentos detalhados, incluindo requisitos funcionais e não funcionais, fluxos de processos e modelos de dados. Essa documentação é compartilhada com os stakeholders, incluindo gerentes de projeto, usuários finais e desenvolvedores. Ela serve como referência crucial para todos os envolvidos, garantindo uma compreensão clara dos objetivos, funcionalidades e restrições do sistema. Em paralelo, a equipe opta por uma abordagem de prototipação. Eles desenvolvem uma versão simplificada do sistema, incluindo interfaces básicas e funcionalidades principais. Essa versão inicial é compartilhada com os usuários para obtenção de feedback. Os usuários interagem com o protótipo, identificam pontos fortes e fraquezas, sugerem melhorias e validam se as funcionalidades atendem às suas necessidades. Esse feedback é fundamental para ajustar o design e os requisitos antes da implementação completa do sistema. Ao longo do ciclo de desenvolvimento, a documentação é continuamente atualizada para refletir as mudanças e evoluções do sistema. Isso assegura que todos os membros da equipe e stakeholders estejam alinhados com as decisões tomadas e compreendam a lógica por trás do desenvolvimento.**

2- Como funciona?

No projeto conceitual, o foco está na criação de uma representação dos aspectos do mundo real que serão armazenados no banco de dados, considerando requisitos, regras de negócio e estruturas desejadas pelos usuários. Este modelo não serve como guia direto para a implementação no banco de dados; em vez disso, é uma referência para a etapa subsequente de modelagem, chamada de modelo lógico. A tradução do modelo conceitual para o modelo lógico possibilita a efetiva implementação do banco de dados.

Um modelo conceitual típico é o modelo **Entidade-Relacionamento (ER)**, inicialmente desenvolvido por Peter Chen em 1976. Este modelo representa o projeto do banco de dados com foco em entidades, relacionamentos e atributos. Utiliza-se o modelo ER para criar diagramas de entidade-relacionamento (DER), os quais são essenciais para visualizar a estrutura do banco de dados.

A independência lógica de dados refere-se à capacidade de alterar o modelo lógico sem afetar o modelo conceitual. Isso permite

adaptações no modelo lógico sem a necessidade de modificar o modelo conceitual. O mesmo princípio se aplica à independência física de dados, possibilitando alterações no modelo físico sem impactar o modelo conceitual ou lógico. Essa flexibilidade proporciona eficiência e adaptabilidade durante o desenvolvimento e evolução do banco de dados.

De maneira formal, as entidades no modelo Entidade-Relacionamento (ER) são denominadas conjuntos de entidades.

Em termos mais precisos, **uma “entidade” seria considerada uma instância específica da entidade.** Por outro lado, o termo “conjunto de entidades” seria mais apropriado para se referir ao retângulo no modelo ER. Portanto, ao se deparar com diferentes bancas, é importante estar ciente dessa variação de terminologia.

16.1 MODELO ENTIDADE-RELACIONAMENTO (ER) ENTIDADES

1- O que é?

Representam objetos ou conceitos do mundo real que se pretende retratar. São simbolizadas por retângulos. Sendo entidades fracas aquelas que carecem de significado próprio, ou seja, dependem de uma entidade forte.

2- Como funciona?

O vínculo entre a entidade fraca e a entidade forte relacionada é denominado **relacionamento identificador ou de dependência**. Entidades associativas surgem quando é necessário estabelecer uma relação entre uma entidade e um relacionamento.

• Atributos:

Descrevem as características das entidades (ou dos relacionamentos).

São representados por elipses.

Atributo chave, também conhecido como identificador, identifica exclusivamente uma entidade.

Atributos simples ou compostos, sendo o segundo capaz de ser dividido em várias partes.

Atributos multivalorados podem assumir múltiplos valores para uma única ocorrência de entidade.

Atributos obrigatórios devem ter um valor para todas as



Questões

ocorrências de uma entidade, enquanto os opcionais podem ser omitidos.

Atributos derivados obtêm seu valor por meio do valor de outro atributo.

•Relacionamentos:

Representam as conexões entre as entidades.

Grau: indica quantos conjuntos de entidades participam do relacionamento (1 a N).

Cardinalidade: expressa o valor máximo de ocorrências de uma entidade associadas a uma ocorrência da entidade relacionada. **Como exemplo: 1:1; 1:N e M:N (ou N:N).**

Participação (ou cardinalidade mínima): estabelece a quantidade mínima de ocorrências da entidade que participam do relacionamento com uma ocorrência da outra entidade. Dessa maneira: 0 (parcial); 1 (total). Existem diferentes notações para representar cardinalidade e participação.

•Autorrelacionamento:

Também chamado de recursivo, representa a relação entre uma entidade e ela mesma.

Sempre de grau 1. A mesma entidade pode desempenhar papéis distintos no relacionamento.

16.2 DADOS ESTRUTURADOS E NÃO ESTRUTURADOS

1- O que é?

Dados Estruturados: Possuem formato organizado e predefinido. Geralmente, estão em tabelas ou bancos de dados relacionais, seguindo um modelo específico.

Dados Não Estruturados: Não seguem um formato predefinido e não se encaixam facilmente em tabelas. Exemplos incluem texto livre, imagens, vídeos e áudio.

Exemplos: De dados estruturados é o banco de dados de uma empresa com informações sobre clientes, produtos e transações. Enquanto os não estruturados são os posts em redes sociais, vídeos no YouTube, e-mails com conteúdo livre.

17. MINERAÇÃO DE DADOS

1- O que é?

É um conjunto de técnicas projetadas para apoiar a tomada de decisões nas organizações. É importante esclarecer que a mineração de dados não se configura como um sistema de armazenamento de dados ou uma variante de banco de dados.

A mineração de dados integra-se ao processo mais amplo de Descoberta de Conhecimento em Bases de Dados (KDD), o qual representa um procedimento que permite extrair informações e conhecimentos a partir do processamento, enriquecimento e interpretação de dados brutos. **Exemplo: Uma loja online querendo entender melhor como seus clientes compram para melhorar as sugestões de produtos. Eles aplicam a Descoberta de Conhecimento em Bancos de Dados (KDD) da seguinte maneira: Primeiro, eles coletam dados sobre compras, preferências e comportamentos dos clientes. Em seguida, organizam e limpam esses dados. Usando algoritmos, identificam padrões, como produtos frequentemente comprados juntos ou hábitos de compra sazonais. Avaliam a importância desses padrões para os objetivos da empresa e interpretam os resultados. Por fim, aplicam esses insights ajustando as recomendações de produtos e estratégias de marketing. Assim, a KDD ajuda a loja online a usar os dados de forma inteligente, proporcionando sugestões de produtos mais personalizadas e aprimorando a experiência de compra dos clientes.**

2- Como funciona?

Existem etapas preliminares em que os dados originais são extraídos, manipulados e transformados antes de chegar à fase central de mineração. Esses processos são frequentes nos sistemas analíticos, pois os dados provenientes de diversas fontes nem sempre estão no formato e qualidade ideais para a extração eficaz de informações.

A mineração de dados proporciona insights nos dados corporativos que muitas vezes não podem ser alcançados por meio do OLAP, ao identificar padrões ocultos e relações em grandes conjuntos de dados, inferindo regras que possibilitam a previsão de comportamentos futuros.

A mineração de dados faz amplo uso de técnicas avançadas de computação, incluindo aprendizado de máquina, uma forma de inteligência artificial. Um aspecto significativo é que essa análise pode ser conduzida em diversos conjuntos de dados, abrangendo desde dados estruturados, como os presentes em um data warehouse ou banco de dados convencional, até dados não estruturados, como aqueles encontrados em abundância na internet, como postagens em redes sociais.

Aplicações:

A mineração de dados possui uma ampla gama de aplicações em diversos setores de negócios e na sociedade como um todo. Difícilmente se encontrará uma área ou ce-

nário em que a mineração não seja aplicável, dado que não se trata de um processo rígido com fases estritamente definidas.

Existem algumas concepções equivocadas sobre a mineração de dados que são amplamente difundidas, mas não refletem a realidade. Podemos destacar quatro equívocos interessantes que podem ser alvo de perguntas em sua prova.

Falácia 1: Existem ferramentas de mineração de dados que podem ser aplicadas aos conjuntos de dados, e elas encontrarão soluções para os nossos problemas.

Realidade: Em primeiro lugar, não existe uma ferramenta de mineração completamente automática. Portanto, não é possível simplesmente executar a ferramenta e aguardar que ela resolva os problemas. A mineração de dados é um processo que demanda a intervenção e avaliação humana em diversas etapas, inclusive na interpretação dos resultados.

Falácia 2: O processo de mineração é autônomo, precisando de pouca ou nenhuma supervisão.

Realidade: Apesar do uso de técnicas de inteligência artificial, a supervisão e ação humana são necessárias durante todo o processo.

Falácia 3: Soluções de mineração de dados se pagam muito rapidamente.

Realidade: Embora a mineração de dados possa agregar valor aos negócios, também pode ser um processo complexo, envolvendo gastos elevados com pessoal e equipamentos. A rapidez com que o processo gera retorno financeiro dependerá de cada caso específico.

Falácia 4: Sistemas de mineração de dados são intuitivos e fáceis de usar.

Realidade: A mineração de dados engloba técnicas avançadas de computação, juntamente com conceitos de estatística e matemática. Nesse sentido, as pessoas envolvidas no processo precisam ser altamente qualificadas e possuir um profundo conhecimento no campo de negócios em que a mineração será aplicada.

É relevante destacar que os Data Warehouses que passam por um tratamento adequado tornam-se excelentes fontes para a mineração de dados, uma vez que seus dados já foram submetidos a esse pré-processamento. **Exemplo: Imagine uma**



Questões

empresa de varejo que mantém um Data Warehouse para armazenar dados de vendas, estoque e comportamento do cliente. Antes de serem armazenados no Data Warehouse, esses dados passam por um tratamento cuidadoso, incluindo limpeza, integração e organização. Com o tratamento adequado, o Data Warehouse torna-se uma fonte confiável e consistente de informações. Agora, a equipe de análise de dados pode aplicar técnicas avançadas de mineração para identificar padrões de compra, tendências sazonais e preferências dos clientes. Por exemplo, ao analisar os dados do Data Warehouse, a equipe pode descobrir que certos produtos têm maior demanda durante determinadas épocas do ano. Isso permite que a empresa ajuste estratégias de marketing e gerencie o estoque de forma mais eficiente, otimizando as operações.

17.1 LIMPEZA DE DADOS

1- O que é?

Para compreender a limpeza de dados, é necessário primeiro entender o conceito de dados “sujos”.

2- Como funciona?

Dados considerados inadequados podem se apresentar de diversas formas. A limpeza de dados deve abordar valores ausentes, como os representados por NULL ou o termo “N/A”. Os dados utilizados na mineração podem originar-se de várias fontes, cada uma representando dados ausentes de maneiras distintas, como NULL, “N/A”, “Não aplicável”, “Desconhecido”, entre outros.

17.2 TRATAMENTO DE DADOS AUSENTES

1- O que é?

Valores ausentes em bancos de dados geralmente prejudicam a capacidade de análise. Embora seja ideal que todos os campos contenham valores, isso raramente ocorre. Assim, a substituição de dados ausentes por valores apropriados é frequentemente realizada, processo conhecido como **imputação de dados**.

2- Como funciona?

A imputação de dados requer cuidado, e a escolha do valor a ser inserido no campo ausente depende da aplicação. Pode-se utilizar observações estatísticas, como médias ou modas, para determinar o valor a ser inserido. Outras opções incluem a geração de valores aleatórios ou a substituição por um valor constante. Caso não haja um substituto adequado, a opção de deixar o valor em

branco e ignorar os registros nas análises também é válida. A escolha da técnica mais apropriada depende do contexto específico da mineração de dados.

17.3 DEDUPLICAÇÃO

Para além dos possíveis equívocos e inconsistências nos dados, ao extrair informações de diversas fontes, é possível que algumas delas apresentem duplicidades entre si. **Exemplo: Suponha que você tenha uma planilha com informações de clientes e alguns deles aparecem mais de uma vez devido a inserções duplicadas ou erros de entrada. A deduplicação identifica que “João” aparecia duas vezes com o mesmo e-mail e removeu a entrada duplicada, deixando apenas uma instância na lista. Isso resulta em uma lista mais limpa e precisa, evitando redundâncias e facilitando a gestão dos dados.**

17.4 REMOÇÃO DE DADOS IRRELEVANTES

Além disso, alguns dados podem não ser duplicados nem conter erros, mas simplesmente não serem pertinentes para a análise em questão. Outliers (anomalias).

17.5 INTEGRAÇÃO DE DADOS

1- O que é?

Conforme mencionado anteriormente, os dados para mineração podem provir de diversas fontes distintas, muitas vezes heterogêneas em suas características. No processo de mineração, podemos lidar com dados provenientes de bancos de dados relacionais, multidimensionais (OLAP), arquivos no formato CSV ou XLS, entre outras possibilidades infinitas.

Quando os dados são provenientes de diferentes origens, podem surgir desafios de integração devido à variedade dessas fontes. Um problema comum é a existência de múltiplos formatos para representar os mesmos dados. Portanto, é crucial estabelecer um novo padrão para essas chaves no conjunto de dados a ser minerado, assegurando que o campo chave continue identificando de maneira única cada registro.

Exemplo: Suponha que uma equipe de análise de uma empresa de varejo esteja buscando entender melhor o comportamento de compra dos clientes. Eles decidem realizar uma análise de dados para identificar padrões de compra e preferências dos clientes.

Para isso, a equipe coleta dados de várias fontes diferentes:

Banco de Dados Relacional: Eles acessam o banco de dados da empresa, que contém

informações sobre transações de vendas, registros de clientes e produtos.

Banco de Dados Multidimensional (OLAP): Eles consultam o banco de dados OLAP, que oferece uma visão analítica dos dados, permitindo análises mais complexas e detalhadas.

Arquivos CSV e XLS: Além disso, eles importam arquivos CSV e XLS contendo dados de pesquisas de satisfação do cliente e feedbacks de compras online.

Fontes Externas: A equipe também utiliza dados provenientes de fontes externas, como redes sociais e sites de análise de mercado, para complementar a análise.

17.6 TRANSFORMAÇÃO DE DADOS

1- O que é?

Além da limpeza, a mineração de dados muitas vezes requer transformações no conjunto de dados.

Normalização: É importante destacar que essa normalização difere do processo aplicado em bancos de dados relacionais. Aqui, estamos nos referindo a uma normalização matemática, uma operação que distribui os dados proporcionalmente para que todos se encontrem em um intervalo específico.

Suavização: Tem como objetivo eliminar ruídos dos dados, ou seja, valores ocasionais que se afastam de um padrão desejado. A aplicação de técnicas de suavização facilita a identificação de tendências temporais e processos que ocorrem de maneira cíclica. Algumas abordagens consideram a suavização como uma forma de limpeza de dados, uma vez que contribui para a remoção de anomalias (outliers) no conjunto de dados.

Agregação: A agregação de conjuntos de dados envolve o aumento da granularidade, isto é, a redução do nível de detalhe dos dados. A agregação é realizada para diminuir a quantidade de registros no conjunto de dados, resultando em economia de espaço de armazenamento e possível agilidade no processo de mineração.

Generalização ou Discretização: Similar à agregação, a generalização reduz a quantidade de registros, mas com a peculiaridade de transformar conjuntos de valores em categorias hierárquicas. Esse processo proporciona uma representação mais abstrata dos dados.

17.6 REDUÇÃO DE DADOS

1- O que é?

As técnicas de redução de dados buscam diminuir o



Questões

tamanho dos conjuntos de dados, preservando a integridade do conjunto original.

2- Como funciona?

Outras técnicas incluem a redução de dimensionalidade (eliminação de atributos e dimensões irrelevantes ou redundantes) e a compressão de dados (semelhante a “zipar” um arquivo de computador). Além disso, existem abordagens mais complexas, como a redução de numerosidade, que busca representar o conjunto de dados de forma mais compacta, utilizando métodos classificados como paramétricos e não paramétricos.

As tarefas de classificação e clusterização em mineração de dados geralmente estão associadas a tipos específicos de aprendizado. A clusterização está relacionada ao aprendizado não supervisionado, pois não há conjunto de treinamento nem grupos pré definidos, enquanto a classificação está relacionada a um processo de aprendizado supervisionado.

A identificação de outliers pode ser realizada com técnicas de aprendizado. Existem três tipos de aprendizado de máquina que podem ser usados para essa tarefa:

Tipo 1: Semelhante à clusterização não supervisionada, o sistema agrupa dados sem conhecimento prévio das categorias e indica valores que não se encaixam em nenhuma delas como potenciais outliers.

Tipo 2: Semelhante à classificação supervisionada, o usuário classifica dados no conjunto de treinamento como normais/anômalos, e o sistema aprende a classificar as próximas entradas.

Tipo 3: Uma abordagem semi-supervisionada em que apenas dados considerados normais são classificados previamente, e o sistema aprende a identificar outliers como aqueles que não se encaixam nesse padrão.

17.7. BIG DATA

1- O que é?

Termo empregado para descrever vastos conjuntos de dados que apresentam desafios consideráveis para processamento em bancos de dados e programas convencionais.

2- Como funciona?

Dessa forma, as ferramentas e sistemas associados ao big data são projetados para lidar com esses extensos conjuntos de dados, que geralmente são predomi-

nantemente não estruturados, permitindo a extração de valor a partir dessa abundância de informações. É válido observar que os dados não se restringem apenas ao formato não estruturado, incluindo também dados estruturados.

Os 5 Vs - Volume refere-se à quantidade massiva de dados produzidos. O conceito fundamental de big data está intrinsecamente vinculado a conjuntos de dados consideráveis.

Velocidade diz respeito à rapidez com que esses dados são gerados e atualizados.

Variedade é um conceito relacionado às diversas estruturas nas quais os dados são encontrados atualmente. O big data abarca dados estruturados, semiestruturados e não estruturados em uma ampla gama de tipos e formatos.

Veracidade evidencia uma preocupação aprimorada com a confiabilidade e organização dos dados.

Valor representa uma das características mais cruciais e de fácil compreensão do big data.

Quando falamos de big data, mencionamos comumente quatro tipos distintos de análise:

Preditiva: Tem como objetivo prever comportamentos futuros e tendências com base nos dados conhecidos.

Prescritiva: Similar à análise preditiva, porém, foca em compreender os efeitos dos eventos futuros. Enquanto a análise preditiva pode detectar um possível declínio no mercado, a análise prescritiva busca prever o impacto desse declínio nas vendas da empresa.

Descritiva: Concentra-se no presente, descrevendo as características dos dados e eventos atuais para embasar decisões de efeitos imediatos.

Diagnóstica: Busca compreender as relações de causa e efeito entre eventos.

Big Data possui uma ampla variedade de aplicações em diversos setores da sociedade. **Exemplo: Os governos podem empregar essas técnicas para otimizar os serviços públicos, como na análise de movimentações nos mercados financeiros para regulação e fiscalização, identificação de fraudes e pesquisas no sistema público de saúde para uma alocação mais eficiente de recursos e melhoria no aten-**

dimento à população, entre outras possibilidades.

No setor privado, é amplamente utilizado em estratégias de marketing, especialmente na coleta e processamento de dados provenientes de redes sociais, que representam um vasto campo de informações não estruturadas. No âmbito educacional, existem aplicações que medem o desempenho dos estudantes em determinadas regiões, correlacionando esses dados com diversos outros fatores do ambiente escolar para aprimorar a qualidade do ensino.

Na área da saúde, já existem diversas aplicações, como aquelas que coletam extensos conjuntos de dados sobre a evolução do estado de pacientes com doenças crônicas. Essas informações são fundamentais para aprofundar o estudo e o tratamento dessas enfermidades.

A arquitetura de Big Data foi concebida para lidar com a ingestão, processamento e análise de dados que são muito grandes ou complexos para os sistemas de banco de dados convencionais. O ponto em que as organizações entram no domínio do Big Data pode variar, dependendo das habilidades dos usuários e de suas ferramentas.

18. LINGUAGEM PYTHON

1- O que é?

Python destaca-se como uma linguagem de alto nível, caracterizada por código conciso, facilidade de leitura e proximidade com a linguagem humana em comparação com outras linguagens.

2- Como funciona?

O código-fonte Python, identificado por arquivos com extensão .py, é interpretado por. Isso significa que suas instruções são executadas interativamente, uma a uma, por um interpretador. Durante a execução, o programa Python é automaticamente convertido em bytecode, um formato de arquivo com extensão .pyc ou .pyo, para ser interpretado por uma máquina virtual.

As principais características do Python incluem:

Sintaxe Clara e Legível: Python utiliza indentação para definir blocos de código, ao contrário de chaves usadas em muitas outras linguagens. Isso torna o código Python muito limpo e fácil de entender.

Tipagem Dinâmica: Python é uma linguagem de tipagem



Questões