

1. Um analista de dados precisa processar três conjuntos de informações: registros de vendas em planilha Excel, transcrições de atendimento ao cliente e fotografias de produtos cadastrados no sistema. Ao planejar o fluxo de processamento, ele percebe que esses conjuntos exigem abordagens técnicas distintas. O que justifica essa diferença de tratamento?

- (a) Registros em planilha sempre exigem técnicas mais avançadas por conterem maior volume de informações.
- (b) Transcrições e fotografias possuem campos predefinidos e estruturados, dificultando a consulta por sistemas relacionais.
- (c) Registros em planilha seguem estrutura padronizada, enquanto transcrições e fotografias não possuem formato fixo nem organização predefinida.
- (d) Fotografias e transcrições são sempre geradas por sistemas automatizados, ao contrário dos registros em planilha.
- (e) Registros em planilha dependem somente de inteligência artificial para serem processados, enquanto os demais tipos não.

2. Bases governamentais com registros incompletos ou desatualizados dificultam a formulação de políticas públicas eficazes. Para garantir a qualidade dessas bases, diferentes processos precisam atuar de forma complementar. Avalie as afirmações a seguir.

- I. A atualização contínua contribui para a confiabilidade das informações governamentais.
- II. A coleta padronizada contribui para a consistência dos registros.
- III. A integração entre bases reduz riscos de interpretações equivocadas.
- IV. A limpeza de dados, por si só, elimina a necessidade de qualquer etapa posterior de validação.

É correto o que se afirma em

- (a) II, III e IV, apenas.
- (b) I, II e III, apenas.
- (c) I, III e IV, apenas.
- (d) II e III, apenas.
- (e) I e II, apenas.

3. Em uma pesquisa sobre competências digitais no setor público, servidores com formação em leitura de indicadores e interpretação estatística conseguiram extrair conclusões a partir de painéis de dados. Já aqueles sem essa formação apenas visualizaram os gráficos sem compreender as implicações. Qual competência explica a diferença de desempenho entre os dois grupos?

- (a) Domínio de ferramentas tecnológicas avançadas para visualização e geração de relatórios.
- (b) Acesso a sistemas de dados mais completos e atualizados do que os colegas.
- (c) Letramento em dados, que permite interpretar indicadores e extrair sentido das informações apresentadas.
- (d) Experiência prévia em gestão pública, que facilita a leitura de contextos organizacionais.
- (e) Capacidade de programação, necessária para manipular os painéis de dados diretamente.

4. Um modelo de rede neural apresentou acurácia de 98% nos dados de treinamento, mas apenas 72% em dados novos. A análise indicou que o modelo memorizou ruídos e padrões irrelevantes da amostra original, perdendo a capacidade de generalizar. Qual conceito de aprendizado de máquina descreve esse comportamento?

- (a) *Overfitting*, pois o modelo memorizou padrões específicos do treinamento e perdeu capacidade de generalização.
- (b) *Underfitting*, pois o modelo memorizou padrões específicos do treinamento e perdeu capacidade de generalização.
- (c) *Overfitting*, pois o modelo não conseguiu aprender nenhum padrão relevante dos dados de treinamento.
- (d) *Underfitting*, pois o modelo não conseguiu capturar a estrutura subjacente dos dados de treinamento.
- (e) Regularização, pois o modelo aplicou penalidades excessivas durante o treinamento.

5. Desde os anos 1990, o volume de dados armazenados por organizações cresceu de terabytes para exabytes. Esse fenômeno, chamado Big Data, envolve múltiplas dimensões além do volume. Avalie as afirmações a seguir.

- I. Big Data descreve o crescimento exponencial de dados estruturados e não estruturados.
- II. Big Data envolve volumes que excedem a capacidade de ferramentas convencionais de processamento em tempo tolerável.
- III. Big Data se limita ao volume massivo de dados, sem considerar velocidade ou variedade.
- IV. Big Data é um fenômeno surgido apenas com o advento das tecnologias de armazenamento em exabytes, não havendo registros anteriores do conceito.

É correto o que se afirma em

- (a) II, III e IV, apenas.
- (b) I, III e IV, apenas.
- (c) I e III, apenas.
- (d) II e IV, apenas.
- (e) I e II, apenas.

6. Machine learning tradicional inclui técnicas como modelos lineares e árvores de decisão, que aprendem padrões a partir de dados sem regras explicitamente programadas. Deep learning, por sua vez, utiliza redes neurais com múltiplas camadas para aprender representações progressivamente mais abstratas dos dados brutos. Do ponto de vista conceitual e estrutural, o que fundamentalmente distingue deep learning das demais técnicas de machine learning?

- (a) Aprende padrões sem necessidade de dados rotulados, dispensando supervisão humana.
- (b) Exige que o programador defina manualmente as regras de classificação dos dados.
- (c) Apresenta desempenho superior ao ML tradicional em qualquer tipo de tarefa ou volume de dados.
- (d) Aprende representações hierárquicas em múltiplas camadas de abstração, diretamente dos dados brutos.
- (e) Depende de engenharia manual de atributos para extrair características relevantes dos dados.

7. A escolha entre machine learning (ML) tradicional e deep learning (DL) depende de critérios como volume de dados, tipo de tarefa e condições de validação. Qual afirmação sobre esses critérios está INCORRETA?

- (a) A divisão dos dados em conjuntos de treino, validação e teste é condição indispensável para resultados confiáveis.
- (b) Tarefas com grande volume de dados brutos, como reconhecimento de imagem, tendem a favorecer DL.
- (c) Modelos de ML tradicional são, em geral, mais simples de interpretar e treinar do que modelos de DL.
- (d) Tarefas que envolvem representações hierárquicas, como processamento de texto e imagem, tendem a favorecer DL.
- (e) Em cenários com poucos dados e possibilidade de engenharia de atributos, DL tende a superar ML tradicional

8. Sistemas de inteligência artificial tradicionais realizam tarefas como classificação e previsão a partir de padrões em dados existentes. Uma subcategoria específica vai além disso ao produzir conteúdo inédito — como texto, imagens e áudio — a partir de redes neurais profundas. Qual característica melhor define a IA Generativa em relação aos demais sistemas de IA?

- (a) Capacidade de classificar grandes volumes de dados com maior precisão do que modelos tradicionais.
- (b) Capacidade de aprender padrões sem necessidade de dados rotulados ou supervisão humana.
- (c) Capacidade de produzir conteúdo original, como texto, imagens ou áudio, de forma autônoma.
- (d) Capacidade de realizar previsões em tempo real com base em dados históricos estruturados.
- (e) Capacidade de integrar múltiplas fontes de dados para gerar relatórios analíticos automatizados.

9. A integração de dados de múltiplas fontes, aliada a técnicas analíticas e boas práticas de governança, pode transformar a forma como organizações tomam decisões. Qual é a principal implicação desse processo para a tomada de decisão nas organizações?

- (a) Elimina a necessidade de estruturas de governança, pois os dados integrados já garantem decisões corretas por si mesmos.
- (b) Aumenta a capacidade de tomar decisões embasadas em evidências, mas exige competências analíticas e estruturas de governança adequadas.
- (c) Reduz a autonomia dos gestores, pois as decisões passam a ser determinadas automaticamente pelos sistemas analíticos.
- (d) Garante decisões rápidas e sem riscos, desde que os dados sejam coletados de fontes internas à organização.
- (e) Melhora a qualidade das decisões operacionais, mas é irrelevante para decisões de nível estratégico.

10. Na hierarquia clássica dado → informação → conhecimento, cada nível possui características próprias. Assinale a alternativa que apresenta a sequência correta dessas definições, na ordem mencionada.

- (a) Registros brutos sem contexto / resultado da organização e interpretação dos dados / aplicação prática de informações para resolver problemas.
- (b) Resultado da interpretação de informações / registros brutos organizados / conjunto de experiências acumuladas sem aplicação prática.
- (c) Informação contextualizada / dado organizado em padrões / experiência adquirida sem necessidade de aplicação.
- (d) Registros brutos sem contexto / padrões identificados a partir do conhecimento / aplicação estratégica de dados.
- (e) Conjunto de fatos interpretados / registros sem valor analítico / experiências práticas não sistematizadas.

11. Analise as afirmativas a seguir acerca das etapas do processo de ETL (Extract, Transform, Load) no contexto do Pentaho Data Integration (PDI) e assinale a alternativa correta:

I. A etapa "Extract" é responsável pela coleta de dados nas fontes de origem, podendo abranger fontes heterogêneas como bancos de dados relacionais, arquivos de texto e planilhas eletrônicas, sem aplicar regras de negócio sobre os dados coletados.

II. Na etapa "Extract", são aplicadas as regras de negócio definidas pela organização, como padronização de formatos, eliminação de duplicatas e cálculo de campos derivados, antes de os dados serem transferidos ao destino.

III. A etapa "Load" tem como finalidade transferir os dados já transformados para o sistema de destino — como um data warehouse ou banco de dados analítico —, encerrando o ciclo do processo ETL.

- (a) Apenas a afirmativa I está correta.
- (b) Apenas as afirmativas I e III estão corretas.
- (c) Apenas as afirmativas II e III estão corretas.
- (d) As afirmativas I, II e III estão corretas.
- (e) Apenas a afirmativa II está correta.

11. São afirmações corretas acerca dos componentes de entrada e processamento disponíveis nas Transformações do Pentaho Data Integration (PDI), EXCETO:

- (a) O componente "Text File Input" permite configurar livremente o caractere delimitador de campos (vírgula, ponto e vírgula, tabulação, barra vertical, etc.), o formato de data e o tipo de dado de cada campo lido do arquivo.
- (b) O componente "Microsoft Excel Input" é compatível com planilhas nos formatos .xls e .xlsx, possibilitando a seleção da aba (sheet) e do intervalo de células a serem processados.
- (c) O componente "Table Input" permite executar consultas SQL personalizadas — incluindo filtros, junções e subconsultas — para extração de dados de qualquer banco de dados relacional que disponha de driver JDBC configurado no PDI.

(d) O componente "Dummy" realiza a validação automática do tipo de dado dos campos que fluem através dele, interrompendo a execução da Transformação com uma mensagem de erro quando identifica incompatibilidade de tipo entre os campos do fluxo.

(e) Em uma única Transformação do PDI, é possível utilizar simultaneamente "Text File Input", "Microsoft Excel Input" e "Table Input" para ler dados de fontes distintas, combinando os fluxos resultantes por meio de steps adequados.

13. Analise as afirmativas a seguir acerca dos componentes "Merge Join", "Filter Rows" e "Sort Rows" do Pentaho Data Integration (PDI) e assinale a alternativa correta:

I. O "Merge Join" exige que os dois fluxos de dados a serem unidos estejam previamente ordenados pelos campos-chave de junção; na prática, é usual a utilização do "Sort Rows" antes de cada stream de entrada do "Merge Join" para garantir essa pré-condição.

II. O "Filter Rows" divide o fluxo em dois caminhos distintos — TRUE e FALSE — de acordo com uma condição lógica avaliada registro a registro; quando o caminho FALSE não está conectado a nenhum step subsequente, os registros que não satisfazem a condição são silenciosamente descartados, sem geração de erro de execução.

III. O componente "Sort Rows", além de ordenar os registros do fluxo, elimina automaticamente os registros duplicados, funcionando de forma equivalente à cláusula DISTINCT da linguagem SQL.

- (a)** Apenas as afirmativas I e II estão corretas.
- (b)** Apenas as afirmativas I e III estão corretas.
- (c)** Apenas as afirmativas II e III estão corretas.
- (d)** As afirmativas I, II e III estão corretas.
- (e)** Apenas a afirmativa I está correta.

14. São afirmações corretas sobre o conceito e o funcionamento de Transformações no Pentaho Data Integration (PDI), EXCETO:

- (a)** Uma Transformação no PDI é composta por steps (etapas de processamento) interligados por hops (conexões direcionais), sendo cada step responsável por uma operação específica sobre o fluxo de dados.
- (b)** Os steps de uma Transformação operam em modelo de pipeline: assim que um step produz registros, estes são imediatamente transmitidos ao step seguinte, sem aguardar o processamento completo do step anterior.
- (c)** O utilitário de linha de comando Kitchen, integrante do PDI, é responsável pela execução de Transformações (.ktr) em modo batch, sem a necessidade de abertura da interface Spoon.
- (d)** Transformações são salvas em arquivos com extensão .ktr e podem conter múltiplos steps de entrada, etapas de transformação e steps de saída, desde que o fluxo de dados entre eles esteja definido por hops.
- (e)** Uma única Transformação pode realizar operações de leitura de dados de fontes distintas, aplicação de regras de transformação e escrita dos resultados em sistemas de destino, constituindo a unidade básica de implementação de lógica ETL no PDI.

15. Um desenvolvedor PDI está configurando uma Transformação para processar um arquivo com os seguintes campos:

- **nome_completo:** texto com letras e espaços (ex.: "Maria da Silva");
- **codigo_servidor:** identificador numérico inteiro, sem casas decimais (ex.: 123456);
- **salario_bruto:** valor monetário com casas decimais (ex.: 4523,75);
- **data_admissao:** data no formato DD/MM/AAAA (ex.: 15/03/2010);
- **ultimo_acesso:** data e hora no formato DD/MM/AAAA HH:MM:SS (ex.: 24/05/2026 08:42:10).

A atribuição de tipos de dados PDI mais adequada para esses cinco campos, na ordem em que foram apresentados, é:

- (a) String, Number, Number, Timestamp, Date
- (b) String, Integer, Number, Date, Timestamp
- (c) String, Integer, BigNumber, Date, Date
- (d) String, Integer, Number, Timestamp, Timestamp
- (e) String, Number, Integer, Date, Timestamp

16. Acerca dos componentes "Select Values", "String Operations" e "String Cut" disponíveis nas Transformações do PDI, assinale a alternativa INCORRETA:

- (a) No componente "Select Values", a conversão de tipos de dados dos campos — como transformar um campo do tipo String em Integer ou Number — é realizada na aba "Select & Alter", onde o desenvolvedor especifica o novo tipo e o formato de conversão de cada campo.
- (b) O componente "String Operations" permite converter campos de texto para maiúsculas (UPPER CASE) ou minúsculas (lower case), remover espaços em branco nas extremidades do valor (trim) e inverter a sequência de caracteres do campo (reverse), entre outras manipulações textuais.
- (c) O componente "String Cut" extrai uma substring de um campo de texto com base em posições de início e fim configuradas pelo desenvolvedor, gerando um novo campo com a porção recortada do valor original.
- (d) O componente "Select Values" oferece três abas funcionais: "Select & Alter" (para selecionar, renomear e reordenar campos), "Remove" (para excluir campos do fluxo) e "Meta-data" (para converter tipos e ajustar comprimento e precisão dos campos).
- (e) O cálculo do comprimento (número de caracteres) de um campo de texto não é realizado pelo "String Operations"; essa operação é executada pelo componente "Calculator", por meio da função "Length of field A".

17. Um analista de dados precisa implementar, em uma única Transformação PDI, as seguintes operações, nesta ordem:

- 1) Calcular um novo campo "valor_liquido", obtido pela subtração do campo "desconto" do campo "valor_bruto";
- 2) Converter o campo "nome_cliente" para letras maiúsculas;
- 3) Exportar o resultado final para um arquivo CSV com campos separados por ponto e vírgula.

Assinale a alternativa que apresenta a sequência correta de componentes PDI para implementar esse fluxo:

- (a) String Operations → Calculator → Text File Output
- (b) Text File Output → Calculator → String Operations
- (c) Calculator → Text File Output → String Operations
- (d) Calculator → String Operations → Text File Output
- (e) String Operations → Text File Output → Calculator

18. Analise as afirmativas a seguir acerca do componente "Concat Fields" disponível nas Transformações do Pentaho Data Integration (PDI) e assinale a alternativa correta:

- I. O "Concat Fields" mantém no fluxo de dados os campos originais utilizados na concatenação, adicionando o campo resultante da união como um novo campo, sem remover os campos de origem.
- II. O separador entre os valores concatenados é obrigatório no "Concat Fields"; caso o desenvolvedor não especifique um caractere separador, o componente gera um erro de configuração e impede a execução da Transformação.
- III. O "Concat Fields" converte automaticamente para sua representação textual os campos de tipos não textuais (como Number, Integer, Date e Timestamp) antes de realizar a concatenação, sendo possível utilizá-lo com campos de qualquer tipo de dado suportado pelo PDI.

- (a) Apenas a afirmativa I está correta.
- (b) Apenas as afirmativas I e II estão corretas.
- (c) Apenas as afirmativas I e III estão corretas.
- (d) As afirmativas I, II e III estão corretas.
- (e) Apenas a afirmativa III está correta.

19. Analise as afirmativas a seguir sobre as Transformações do Pentaho Data Integration (PDI) e assinale a alternativa correta:

I. Uma única Transformação do PDI pode ler dados de múltiplas fontes heterogêneas de forma simultânea — como arquivos de texto (CSV/TXT), planilhas Excel e bancos de dados relacionais —, utilizando os componentes de entrada específicos para cada tipo de fonte.

II. O componente "Merge Join" é indicado para realizar a junção de dois fluxos com base em um campo-chave de correspondência (junção relacional), enquanto o "Append Streams" é indicado para a concatenação sequencial de dois fluxos com a mesma estrutura de campos, sem exigir campo-chave.

III. O componente "Text File Input" detecta e converte automaticamente os tipos de dados dos campos lidos (como Integer, Number, Date e Timestamp) com base no conteúdo do arquivo, dispensando qualquer configuração manual de tipo pelo desenvolvedor.

- (a) Apenas a afirmativa I está correta.
- (b) Apenas as afirmativas I e II estão corretas.
- (c) Apenas as afirmativas II e III estão corretas.
- (d) As afirmativas I, II e III estão corretas.
- (e) Apenas a afirmativa III está correta.

20. Um desenvolvedor PDI precisa calcular, para cada registro de um fluxo de dados, a diferença em dias entre dois campos do tipo Date presentes no fluxo — "data_conclusão" e "data_inicio" — e armazenar o resultado em um novo campo denominado "prazo_dias". Assinale a alternativa que identifica corretamente o componente PDI indicado para essa finalidade e apresenta a justificativa adequada:

- (a) O componente "Calculator", pois disponibiliza funções predefinidas para operações com campos do tipo Date e Timestamp — incluindo o cálculo da diferença entre duas datas em dias —, sem necessidade de escrita de código, por meio de configuração visual no PDI.
- (b) O componente "String Operations", pois permite realizar cálculos sobre campos de data por meio da função de diferença temporal configurável em suas propriedades.
- (c) O componente "Select Values", aba "Meta-data", pois é nessa aba que se configuram operações aritméticas sobre campos do tipo Date para geração de campos derivados.
- (d) O componente "Filter Rows", pois permite calcular expressões numéricas — incluindo diferenças entre datas — como parte da avaliação das condições de filtragem configuradas.
- (e) O componente "Replace in String", pois suporta a manipulação de campos do tipo Date por meio de operações de substituição configuradas com expressões regulares.