

CONCURSO PÚBLICO  
Edital: 2021

**TARDE**



## CADERNO DE PROVAS OBJETIVAS

### LEIA COM ATENÇÃO AS INSTRUÇÕES ABAIXO.

- 1** Ao receber este caderno de provas, confira inicialmente se os dados transcritos acima estão corretos e se estão corretamente registrados na sua **Folha de Respostas**. Confira também seus dados em cada página numerada deste caderno de provas (**desconsidere estas instruções, caso se trate de caderno de provas reserva**). Em seguida, verifique se ele contém a quantidade de itens indicada em sua **Folha de Respostas**, correspondentes às provas objetivas. Caso o caderno esteja incompleto, tenha qualquer defeito e(ou) apresente divergência quanto aos seus dados, solicite, de imediato, ao(à) aplicador(a) de provas mais próximo(a) que tome as providências necessárias.
- 2** Durante a realização das provas, não se comunique com outros(as) candidatos(as) nem se levante sem autorização de um(a) dos(as) aplicadores(as) de provas.
- 3** Na duração das provas, está incluído o tempo destinado à identificação — que será feita no decorrer das provas — e ao preenchimento da **Folha de Respostas**.
- 4** Ao terminar as provas, chame o(a) aplicador(a) de provas mais próximo(a), devolva-lhe a sua **Folha de Respostas** e deixe o local de provas.
- 5** Nenhuma folha deste caderno pode ser destacada, exceto a folha que contém o documento **Folha de Respostas**, cujo cabeçalho será destacado pelo(a) chefe de sala ao final das provas, para fins de desidentificação.
- 6** A desobediência a qualquer uma das determinações constantes em edital, no presente caderno ou na **Folha de Respostas** implicará a anulação das suas provas.

### OBSERVAÇÕES:

- Não serão conhecidos recursos em desacordo com o estabelecido em edital.
- É permitida a reprodução deste material apenas para fins didáticos, desde que citada a fonte.

### INFORMAÇÕES ADICIONAIS:

0(XX) 61 3448-0100  
www.cebraspe.org.br  
sac@cebraspe.org.br



- Cada um dos itens das provas objetivas está vinculado ao comando que imediatamente o antecede. De acordo com o comando a que cada um deles esteja vinculado, marque, na **Folha de Respostas**, para cada item: o campo designado com o código **C**, caso julgue o item **CERTO**; ou o campo designado com o código **E**, caso julgue o item **ERRADO**. A ausência de marcação ou a marcação de ambos os campos não serão apenadas, ou seja, não receberão pontuação negativa. Para as devidas marcações, use a **Folha de Respostas**, único documento válido para a correção das suas provas objetivas.
- Nos itens que avaliarem conhecimentos de informática e(ou) tecnologia da informação, a menos que seja explicitamente informado o contrário, considere que todos os programas mencionados estão em configuração-padrão e que não há restrições de proteção, de funcionamento e de uso em relação aos programas, arquivos, diretórios, recursos e equipamentos mencionados.
- Eventuais espaços livres — identificados ou não pela expressão “**Espaço livre**” — que constarem deste caderno de provas poderão ser utilizados para rascunho.

## -- CONHECIMENTOS BÁSICOS --

### Língua Portuguesa

#### Texto CB1A1-I

Não estamos opondo máquinas a ecologia, como se as máquinas fossem aquelas coisas que só servem para violentar a Mãe Natureza e violar a harmonia entre o ser humano e a natureza — uma imagem atribuída à tecnologia desde o fim do século XVIII. Também não estamos seguindo a hipótese de Gaia de que a Terra é um único superorganismo ou uma coletividade de organismos. Em vez disso, gostaria de propor uma reflexão sobre a ecologia das máquinas. Para dar início a essa ecologia das máquinas, precisamos primeiro voltar ao conceito de ecologia. Seu fundamento está na diversidade, já que é apenas com biodiversidade (ou multiespécies que incluam todas as formas de organismos, até mesmo bactérias) que os sistemas ecológicos podem ser conceitualizados. A fim de discutir uma ecologia de máquinas, precisaremos de uma noção diferente e em paralelo com a de biodiversidade — uma noção a que chamamos tecnodiversidade. A biodiversidade é o correlato da tecnodiversidade, uma vez que sem esta só testemunharemos o desaparecimento de espécies diante de uma racionalidade homogênea. Tomemos como exemplo os pesticidas, que são feitos para matar certa espécie de insetos independentemente de sua localização geográfica, precisamente porque são baseados em análises químicas e biológicas. Sabemos, no entanto, que o uso de um mesmo pesticida pode levar a diversas consequências desastrosas em biomas diferentes. Antes da invenção dessas substâncias, empregavam-se diferentes técnicas para combater os insetos que ameaçavam as colheitas dos produtos agrícolas — recursos naturais encontrados na região, por exemplo. Ou seja, havia uma tecnodiversidade antes do emprego de pesticidas como solução universal. Os pesticidas aparentam ser mais eficientes a curto prazo, mas hoje é fato bastante consolidado que estávamos o tempo todo olhando para os nossos pés quando pensávamos em um futuro longínquo. Podemos dizer que a tecnodiversidade é, em essência, uma questão de localidade. Localidade não significa necessariamente etnocentrismo ou nacionalismo, mas é aquilo que nos força a repensar o processo de modernização e de globalização e que nos permite refletir sobre a possibilidade de reposicionar as tecnologias modernas.

Considerando as ideias, os sentidos e os aspectos linguísticos do texto CB1A1-I, julgue os itens a seguir.

- 1 Quanto à tipologia, o texto é predominantemente expositivo.
- 2 Conclui-se das ideias do texto que o autor concorda com a imagem que se tinha da tecnologia no fim do século XVIII.
- 3 Consoante as ideias do texto, a proposta de uma ecologia das máquinas baseada na noção de tecnodiversidade leva ao reconhecimento de múltiplas racionalidades.
- 4 O conceito de tecnodiversidade apresentado no texto visa a situar as tecnologias modernas em termos de localidade.
- 5 O texto recomenda a infalibilidade dos pesticidas como solução universal para ameaças à produção agrícola.
- 6 O texto estende a noção de risco de extinção do campo biológico ao campo tecnológico.
- 7 No trecho “gostaria de propor uma reflexão sobre a ecologia das máquinas” (terceiro período), a forma verbal “gostaria” expressa uma incerteza.
- 8 No trecho “Para dar início a essa ecologia das máquinas” (quarto período), o acréscimo do sinal indicativo de crase no vocábulo “a” manteria a correção gramatical do texto.
- 9 No trecho “precisamos primeiro voltar ao conceito de ecologia” (quarto período), o vocábulo “primeiro” estabelece concordância com o termo “conceito”.
- 10 No sexto período do texto, a locução “A fim de” introduz uma oração que expressa finalidade.

Ainda com relação aos sentidos e aos aspectos linguísticos do texto CB1A1-I, julgue os seguintes itens.

- 11 No trecho “é apenas com biodiversidade (ou multiespécies que incluam todas as formas de organismos, até mesmo bactérias) que os sistemas ecológicos podem ser conceitualizados”, a oração “que incluam todas as formas de organismos” restringe o sentido do termo “multiespécies”, enquanto todo o segmento entre parênteses especifica a ideia de “biodiversidade” veiculada no texto.
- 12 A forma pronominal “Seu”, na oração “Seu fundamento está na diversidade” (quinto período), tem como referente o termo “ecologia das máquinas”, apresentado no período imediatamente anterior.
- 13 No trecho “A biodiversidade é o correlato da tecnodiversidade”, a palavra “correlato” tem o mesmo sentido de **análogo**.
- 14 A coerência e a correção gramatical do texto seriam mantidas caso o período “Os pesticidas aparentam ser mais eficientes a curto prazo, mas hoje é fato bastante consolidado que estávamos o tempo todo olhando para os nossos pés quando pensávamos em um futuro longínquo” fosse reescrito da seguinte maneira: Mesmo que os pesticidas pareçam mais eficientes a curto prazo, hoje é fato bastante consolidado que estávamos o tempo todo olhando para os nossos pés quando pensávamos em um futuro longínquo.

- 15 No período “A biodiversidade é o correlato da tecnodiversidade, uma vez que sem esta só testemunharemos o desaparecimento de espécies diante de uma racionalidade homogênea”, a forma pronominal “esta” retoma “biodiversidade”.
- 16 No trecho “Tomemos como exemplo os pesticidas, que são feitos para matar certa espécie de insetos”, a retirada da vírgula alteraria os sentidos originais do texto.
- 17 No período “A biodiversidade é o correlato da tecnodiversidade, uma vez que sem esta só testemunharemos o desaparecimento de espécies diante de uma racionalidade homogênea”, a substituição da expressão “uma vez que” por **contanto que** preservaria os sentidos originais do texto.
- 18 No trecho “Antes da invenção dessas substâncias, empregavam-se diferentes técnicas para combater os insetos”, a substituição da forma “empregavam-se” pela locução **foram empregadas** alteraria os sentidos originais do texto.
- 19 Infere-se da expressão “Em vez disso” (terceiro período) que o autor, a partir desse momento, inicia uma argumentação contrária ao posicionamento defendido por ele nos dois primeiros períodos do texto.
- 20 No trecho “Localidade não significa necessariamente etnocentrismo ou nacionalismo, mas é aquilo que nos força a repensar o processo de modernização”, a forma verbal “é” concorda com o termo “Localidade”.

## Língua Inglesa

The pandemic has exposed a contradiction in modern society. We are more connected than ever – proven by the speed and truly global spread of the virus; nevertheless, we are also deeply divided.

Today’s connectedness, enforced by social distancing, has pushed the role of technology in every aspect of our lives to a new level. The tech industry is now in a challenging space: trapped between the increasing politicization of technology, the still superficial social debate around its impact and the slow adaptation of appropriate legislation. Right now, the industry needs to enable and push forward this ethical debate, recognizing its responsibility in this historic moment.

The first step is to end the myth of tech neutrality. We need to acknowledge that technology is caught between private economic interests and its social value. Some tech business models, for example, prioritize commodification of data or advertising revenue over the social value of their services — the sharing of knowledge, connecting people, etc.

The social impact of tech is deeper than we usually perceive and most of the time invisible to us. Digital architecture determines, increasingly, social interactions in our digital society. An example of this is the invisible influence of search engine and social media algorithms, the results of which shape our understanding of the world.

In this moment of crisis, technology can play a fundamental role in overcoming the economic recession and reducing the systemic injustices that have been exposed.

**Social exclusion - the next pandemic.**

Concerning the ideas stated in the text and the words used in it, judge the following items.

- 21 The word “nevertheless” introduces an idea of condition in the first paragraph.
- 22 The sentence “the industry needs to enable and push forward this ethical debate” (in the last sentence of the second paragraph) can be correctly rephrased as **the industry ought to facilitate and advance forward this ethical debate**, without changing the meaning of the text.
- 23 Understanding that digital technology does not play an impartial role in the tense dynamics between private and social spheres is an important stage in the debate in the text.
- 24 The words “trapped” (in the second sentence of the second paragraph) and “caught” (in the second sentence of the third paragraph) can be considered synonyms and are used with similar meanings in the text.
- 25 According to the text, even when we do not notice, the mechanisms that organize the digital society shape the way people relate online.

## Estatuto Social do SERPRO

Com base nas disposições do Estatuto Social do SERPRO, julgue os itens a seguir.

- 26 Em razão de os serviços prestados pelo SERPRO envolverem informações relacionadas à segurança nacional, é vedada ao SERPRO a celebração de contratos, convênios ou parcerias com empresas estrangeiras.
- 27 Para fins de composição dos recursos financeiros do SERPRO, não se admitem outras rendas ou receitas que não as taxativamente previstas no referido estatuto social.
- 28 A deliberação da Assembleia Geral acerca de eventual transformação, incorporação, fusão ou cisão do SERPRO deve ocorrer de forma prévia à decisão do Ministro de Estado da Economia.
- 29 Embora os cargos de titulares das unidades organizacionais do SERPRO sejam, em regra, privativos de empregados integrantes do seu quadro de pessoal, o referido estatuto social admite exceções quanto à ocupação de alguns desses cargos.
- 30 A deliberação da Assembleia Geral e do Conselho de Administração a respeito da indicação de conselheiros e diretores subordina-se à decisão do Comitê de Elegibilidade.

## Raciocínio Lógico

Julgue os itens seguintes, relativos a raciocínio lógico.

- 31 Suponha que seja válida a seguinte proposição:  
P: Se Aldo não tem um número de CPF, então ele não pode viajar nem pode abrir uma empresa.  
Nessa situação, se Aldo pode abrir uma empresa, então ele já obteve o seu número de CPF.
- 32 A negação da afirmação “Todos os amigos de Bianca são naturais de São Paulo” é: “Nenhum amigo de Bianca é natural de São Paulo”.

Suponha que sejam gerados 5 números válidos de CPF para serem atribuídos a 5 indivíduos distintos. Com base nessas informações, julgue os itens seguintes.

- 33 A quantidade de formas de se fazer a atribuição desses CPFs a esses indivíduos é maior que 100.
- 34 Suponha que, logo após a atribuição dos CPFs aos indivíduos, são escolhidos aleatoriamente 2 desses CPFs e separados 3 desses indivíduos. Nessa situação, a probabilidade de pelo menos um dos CPFs escolhidos pertencer a um dos indivíduos separados é igual a  $3/5$ .
- 35 Suponha que  $a$  seja o último dígito de um dos CPFs gerados, que  $b$  seja o último dígito de outro desses CPFs e que  $a$  e  $b$  sejam números ímpares consecutivos. Nessa situação,  $a + b$  é múltiplo de 4.

### Legislação acerca de privacidade e proteção de dados pessoais

Com base na Lei n.º 13.709/2018 — Lei Geral de Proteção de Dados Pessoais (LGPD) —, julgue os itens a seguir.

- 36 A anonimização impossibilita que um dado seja associado, direta ou indiretamente, a um indivíduo.
- 37 O tratamento dos dados regulados deve atender ao princípio da adequação, o qual limita o tratamento ao mínimo necessário para a atividade.
- 38 O tratamento de dados pessoais poderá ser realizado a pedido do próprio titular dos dados quando for necessário para a execução de contrato do qual ele seja parte.
- 39 Consentimento é a manifestação do titular — pessoa natural ou jurídica — sobre o tratamento de seus dados para uma finalidade específica.
- 40 Independentemente do modo como tenham sido coletados, todos os dados pessoais que estejam sob o controle dos operadores devem ser abrangidos por uma governança de privacidade.

De acordo com a Lei n.º 12.527/2011 — Lei de Acesso à Informação (LAI) —, julgue os próximos itens.

- 41 A formulação de pedido de acesso à informação é direito exclusivo de órgãos de defesa do consumidor ou de empresas ou órgãos de comunicação.
- 42 Uma informação que coloque em risco a soberania nacional pode ser classificada como reservada, secreta ou ultrassecreta.
- 43 O acesso às informações solicitadas deverá ser concedido em até cinco dias úteis, a contar da data do pedido, prazo que, a interesse do órgão ou da entidade pública, poderá ser prorrogado por mais dois dias úteis.
- 44 É considerada conduta ilícita e passível de responsabilização a recusa, por agente público ou militar, de fornecimento das informações solicitadas.

Acerca de privacidade e proteção de dados pessoais, julgue os itens subsequentes.

- 45 É indispensável o consentimento do titular ao uso dos seus dados pessoais em pesquisas estatísticas que necessitem de tais informações, mesmo que as pesquisas sejam de evidente interesse público ou geral.
- 46 Para fins de aplicação da LGPD, dado pessoal é o que permite identificar ou tornar identificável, de forma inequívoca, um indivíduo.
- 47 Pessoa física que violar a LAI ao usar informações às quais ela teve acesso em razão de vínculo com o poder público poderá ser penalizada com declaração de inidoneidade para licitar ou contratar com a administração pública.
- 48 Informações classificadas como ultrassecretas deverão ficar com acesso restrito por vinte e cinco anos, contados a partir de sua produção.
- 49 Em caso de infração à LGPD cometida por agente de tratamento de dados, um dos critérios para a aplicação da sanção administrativa ao infrator é a sua condição econômica.
- 50 O tratamento de dados pessoais previsto na LGPD poderá ser feito quando necessário para o atendimento dos interesses legítimos do controlador, exceto nas situações em que prevalecerem direitos e liberdades fundamentais do titular que exijam a proteção dos dados pessoais.

Espaço livre

**-- CONHECIMENTOS ESPECÍFICOS --**

Sobre as técnicas de classificação em aprendizado de máquina, julgue os itens a seguir.

- 51 As árvores de decisão apresentam significativa independência dos dados de treinamento, o que garante forte estabilidade dos caminhos em caso de mudanças nos dados.
- 52 Devido ao fato de pressupor independência entre atributos, o algoritmo *Naive Bayes* é capaz de realizar, com precisão, o treinamento de um modelo com uma quantidade reduzida de amostras.
- 53 Em razão de seu bom desempenho, o algoritmo SVM (*support vector machines*) é invariante à escala dimensional dos conjuntos de dados, o que torna dispensável a padronização e o pré-processamento dos dados.

No que se refere a técnicas de regressão, técnicas de agrupamento e aprendizado profundo, julgue os itens seguintes.

- 54 Nos agrupamentos hierárquicos, um dendrograma é uma árvore que controla quando os *clusters* são criados e que determina qual é a métrica das distâncias.
- 55 Uma máquina de Boltzmann instanciada em um determinado contexto identifica os seus objetos e as suas características e classificações bem como reforça os padrões aprendidos e monitora os padrões desconhecidos.
- 56 O modelo de regressão linear é suscetível à multicolinearidade, a qual diz respeito ao fato de o erro da predição permanecer estável, dentro de um intervalo de confiança aceitável, à medida que os valores de entrada mudam.

Dados ausentes são muito comuns em aplicações que envolvem preenchimentos de formulários. Para o seu tratamento estatístico, encontra-se, na literatura, uma taxonomia para o mecanismo gerador de dados ausentes que os classifica em tipos como MAR (*missing at random*), MCAR (*missing completely at random*) e MNAR (*missing not at random*).

Considerando essas informações, julgue os itens subsequentes, relativos ao tratamento de dados ausentes.

- 57 Geralmente, os dados ausentes do tipo MNAR são ignoráveis por não haver relação entre o processo gerador de dados ausentes e os parâmetros que são objeto de estudo estatístico.
- 58 Os dados ausentes do tipo MAR são aqueles que dependem de fatores — ou variáveis — observáveis, como, por exemplo, o horário de preenchimento de um formulário.
- 59 Para o tratamento estatístico de valores ausentes do tipo MCAR, é suficiente preencher com zeros as lacunas existentes em um conjunto de dados.

Determinado parâmetro  $\beta$  será estimado recursivamente com a ajuda de um método de otimização matemática com base em uma função objetivo  $g(\beta)$ . Para essa estimação, a base de dados de treinamento consistirá de  $n$  observações.

Considerando essa situação hipotética, julgue os itens que se seguem.

- 60 O gradiente descendente em lote é um método probabilístico de otimização no qual, para cada iteração, encontram-se  $L \times n$  observações geradas mediante amostragem (com reposição) da base de dados de treinamento (em que  $L$  representa o número de lotes, com  $L > 1$ ).
- 61 O método do gradiente descendente é equivalente ao método de Newton-Raphson, no qual o incremento, para a estimação do parâmetro  $\beta$ , depende da primeira e da segunda derivada da função objetivo  $g(\beta)$ .
- 62 Entre as condições ideais relativas à função objetivo  $g(\beta)$  para a aplicação do método do gradiente descendente incluem-se convexidade, continuidade e diferenciabilidade.
- 63 Em cada iteração na estimação do parâmetro  $\beta$ , o método do gradiente descendente requer  $n$  observações da base de treinamento, ao passo que o método do gradiente descendente estocástico utiliza uma observação selecionada aleatoriamente dessa base de treinamento.

A respeito da linguagem de programação Python, julgue os itens a seguir.

- 64 As tuplas, embora sejam semelhantes às listas, estão limitadas a, no máximo, cinco níveis.
- 65 Listas são coleções alteráveis de qualquer tipo de objeto — como, por exemplo, outras listas — capazes de gerar um efeito *top-down* sem limite de níveis.

Sobre a linguagem de programação JAVA, julgue os próximos itens.

- 66 A compilação de um programa JAVA para ambiente Windows gera um programa com extensão EXE, o qual é executado pelo sistema operacional.
- 67 O polimorfismo ocorre quando a mesma operação é construída em uma mesma classe ou quando o método da subclasse sobrepõe-se ao método da superclasse.

A respeito de *frameworks*, julgue os itens subsequentes.

- 68 TensorFlow é utilizado com Java para implementar espaços de colaboração, unindo os desenvolvedores de um time em um mesmo ambiente virtual.
- 69 O PySpark funciona como uma *interface* entre Python e JAVA para a construção de *pipelines* de dados.

Considerando que o número  $X$  de erros registrados em determinado tipo de código computacional siga uma distribuição binomial com média igual a 4 e variância igual a 3, julgue os itens subsequentes.

- 70 É impossível haver registros de 18 erros nesse tipo de código computacional.
- 71 A quantidade padronizada  $Z = \frac{X-4}{\sqrt{3}}$  segue distribuição normal padrão.
- 72 A mediana de  $X$  é igual a 4.
- 73  $P(X = 0) = 3/4$ .
- 74 O coeficiente de variação da distribuição de erros  $X$  é igual a 3.

Em uma organização, foram elencadas as necessidades a seguir.

- I Aplicação de modelo de gerenciamento com base no Six Sigma, de modo a gerenciar as etapas da Mineração de Dados do setor responsável.
- II Implantação do processo Sequenciar Atividades, no desenvolvimento do cronograma do projeto, com base no PMBOK.
- III Implantação de abordagem ágil em relação à criação do *backlog* do projeto — lista ordenada de todo o trabalho, apresentado em forma de história.

Tendo como referência essa situação hipotética e a gestão de projetos em ciência de dados, julgue os itens a seguir.

- 75 Quando da implantação do *backlog*, descrito em III, não é necessária a criação de todas as histórias para todo o projeto antes do início do trabalho, e sim apenas o suficiente para entender os detalhes da primeira entrega.
- 76 Para o atendimento à necessidade I, deve-se implantar a CRISP-DM, cujas etapas são Entendimento do Negócio, Entendimento dos Dados, Preparação dos Dados, Modelagem, Avaliação e Aplicação.
- 77 Para a implantação do processo descrito em II, pode-se utilizar o método de diagrama de precedência, quando da definição da sequência lógica do trabalho e dos relacionamentos entre as atividades, tendo como saída a Estrutura Analítica do Projeto (EAP).

Com relação aos fundamentos do *big data*, julgue os itens que se seguem.

- 78 No que se refere aos três Vs do *big data*, o termo volume refere-se a dados que, atualmente, não são estruturados nem armazenados em tabelas relacionais, o que torna sua análise mais complexa.
- 79 *Big data* caracteriza-se, principalmente, por volume, variedade e velocidade, o que se justifica devido ao fato de os dados serem provenientes de sistemas estruturados, que são maioria, e de sistemas não estruturados, os quais, embora ainda sejam minoria, vêm, ao longo dos anos, crescendo consideravelmente.

Considerando o projeto Apache Hadoop, julgue os itens subsequentes.

- 80 Ao serem armazenados no HDFS (*Hadoop Distributed File System*), os dados do Hadoop são divididos em blocos e distribuídos em discos distintos de um mesmo servidor, o que acelera o seu processamento, já que são pesquisados de forma simultânea, e não de forma sequencial.
- 81 As formas de execução do Hadoop podem ser modo local, modo pseudodistribuído ou modo completamente distribuído, e a especificação do seu modo de execução é definida na configuração dos arquivos *core-site.xml*, *hdfs-site.xml* e *mapred-site.xml*.
- 82 O subprojeto Sqoop atua na camada funcional de *data warehouse* e *queries* do Hadoop.
- 83 Um projeto relacionado ao Hadoop e mantido pela Apache é o Hive, que é uma camada de *data warehouse* que roda em cima do Hadoop e que utiliza uma linguagem similar à SQL, denominada Hive SQL.

A respeito de conceitos de NoSQL orientado a grafos, orientado a documentos e orientado a colunas, julgue os itens a seguir.

- 84 *Cypher query language*, linguagem de consulta de gráfico do Neo4j que possibilita aos usuários o armazenamento e a recuperação de dados do banco de dados de gráficos, permite declarar o que se deseja selecionar, inserir, atualizar ou excluir desses dados de gráficos.
- 85 Uma coleção e um documento, no MongoDB, são equivalentes à tabela e à linha, no Modelo Relacional de Dados.
- 86 A compactação, recurso para otimizar espaço de armazenamento, é um processo pelo qual o HBase se utiliza das probabilidades da ocorrência de símbolos e de palavras em um conjunto de dados, para determinar quantos *bits* serão utilizados para cada símbolo.

Com relação a conceitos de sistemas de arquivos distribuídos e sistemas de indexação, julgue os itens subsequentes.

- 87 HDFS (*Hadoop Distributed File System*), dispositivo de tecnologia de memória, é um tipo de arquivo de dispositivo constante dos sistemas operacionais abertos que interage com a memória *flash*.
- 88 Cada instância de *elasticsearch* é um nó; uma coleção de vários nós que podem trabalhar em harmonia formam um *cluster elasticsearch*.

Acerca dos conceitos de ingestão de dados, julgue os itens subsequentes.

- 89 O principal objetivo da ingestão de dados é organizar, armazenar e recuperar dados em tabelas relacionadas a entidades do mundo real.
- 90 No ambiente de ingestão de dados, CDC refere-se ao controle distribuído de computação de dados coletados remotamente para o ambiente de análise.
- 91 Quando a ingestão de dados ocorre em lotes, esses dados são movidos em intervalos programados, abordagem que é benéfica para os processos repetíveis.

Julgue os itens subsequentes relativos a conceitos de ingestão de dados estruturados, semiestruturados, não estruturados e em *streaming*.

- 92 Os dados não estruturados constam de bancos de dados relacionais que são eminentemente pesquisáveis tanto por meio de consultas realizadas por humanos quanto por meio daquelas realizadas por algoritmos que usam tipos de dados e nomes de campos, como alfabéticos ou numéricos, moeda ou data.
- 93 Os dados de *streaming* referem-se a dados que são gerados continuamente, geralmente em grandes volumes e em alta velocidade, por meio de diferentes fontes.

Sobre processamento de dados, julgue os itens que se seguem.

- 94 MapReduce divide o conjunto de dados de entrada em blocos independentes que são processados pelas tarefas de mapa de uma maneira completamente paralela. Essa estrutura classifica as saídas dos mapas, as quais são, então, inseridas nas tarefas de redução.
- 95 Processamento em lote envolve entrada, processo e saída contínua de dados, em que a chave é o tempo de resposta.
- 96 O processamento em tempo real é feito à medida que os dados são inseridos, portanto, é necessário um fluxo contínuo de dados de entrada para fornecer uma saída contínua.

Julgue os próximos itens, relativos à tecnologia de *big data* e ao Hadoop.

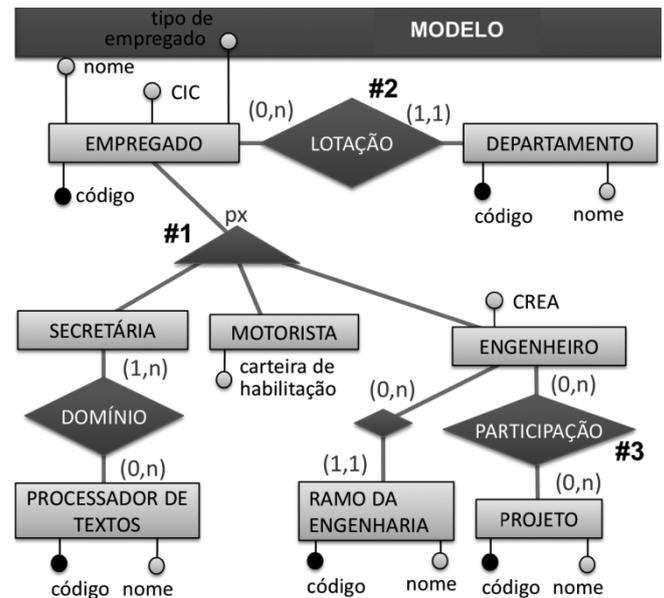
- 97 O Hadoop consiste em um único produto, ou seja, um *software* monolítico, que possibilita análise de *logs* e outros dados da Web.
- 98 Uma das principais características de *big data* é que seu custo de armazenamento de dados é relativamente baixo se comparado a um *data warehouse*.
- 99 Apesar de ser uma tecnologia de código aberto disponibilizada pela ASF (Apache Software Foundation), o Hadoop também é oferecido por distribuidores comerciais, de maneira que fornecedores oferecem distribuições específicas que incluem não só ferramentas administrativas adicionais, mas também suporte técnico.

Acerca de *dashboards*, julgue os itens subsequentes.

- 100 Relatórios do tipo *dashboard* visam apresentar uma visão integrada do sucesso numa organização, incluindo desempenhos financeiros e perspectivas sobre clientes, processos comerciais e aprendizado e crescimento.
- 101 Um *dashboard* bem projetado caracteriza-se por apresentar um conjunto robusto de codificação customizada para sua implementação, utilização e manutenção.
- 102 O conceito de *dashboards* não é novo e, atualmente, eles são considerados onipresentes, sendo sua característica mais distintiva as três seguintes camadas de informação: monitoramento, análise e gerenciamento.

Julgue os próximos itens, relativos à qualidade de dados.

- 103 O gerenciamento de qualidade de dados inclui a definição de padrões e métricas sobre os dados, porém dispensa o gerenciamento do ciclo de vida desses dados.
- 104 A acurácia na qualidade de dados está diretamente associada à de duplicação, uma vez que indica que há exclusividade da fonte de dados e de suas entidades, de forma a garantir a precisão dos dados na vida real.
- 105 Executar o *profiling* dos dados é similar a fazer uma fotografia inicial do estado dos dados em um domínio ou em uma área de assunto da organização.
- 106 Definir processos de qualidade para modelos de dados implica analisar as regras de negócios fundamentais dos processos, bem como observar a qualidade dos dados, com a finalidade de garantir a conformidade da regra com o processo.
- 107 Uma boa prática em relação à qualidade de dados é verificar a precisão dos dados, validando-se se estão corretamente representados, e a sua consistência, avaliando-se se há integridade cruzada entre duas ou mais fontes que armazenem o mesmo dado.

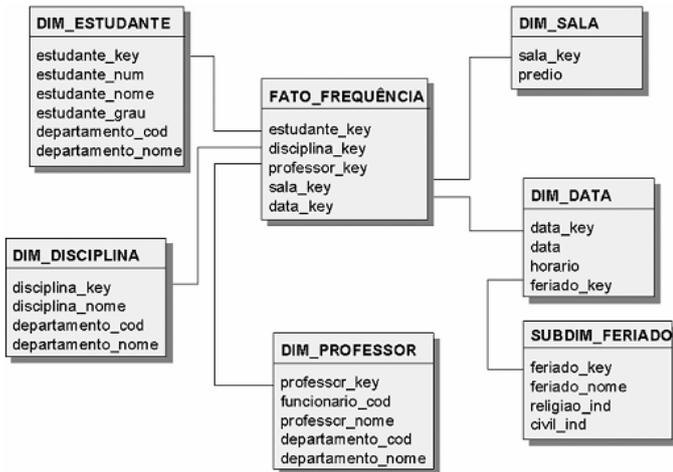


Cada uma das próximas assertivas se refere à transição do modelo precedente para o modelo relacional.

- I No modelo resultante, a relação #1 deverá ser transformada em três tabelas associativas que liguem EMPREGADO a cada uma de suas especializações.
- II No modelo resultante, a relação #2 indica que, na transformação, deverá haver uma fusão entre as entidades EMPREGADO e DEPARTAMENTO, pois cardinalidades (1,1) devem ser normalizadas dessa forma.
- III No modelo resultante, é opcional a criação de uma tabela para suprir a regra de cardinalidade da relação #3, pois PARTICIPAÇÃO não possui chave própria.

A respeito dessas assertivas, julgue os itens a seguir.

- 108 A assertiva II é errada e a III é correta, pois, em ambos os casos, aplica-se a regra de normalização da primeira forma normal (1.ª FN).
- 109 A assertiva I está correta, pois modelos relacionais não possuem herança.



Internet: <sandrorrguimaraes.blogspot.com> (com adaptações).

Tendo como referência inicial esse modelo, e considerando que, para representar a quantidade de faltas em valores inteiros, seja inserida em **FATO\_FREQUÊNCIA** a métrica *qtd\_faltas*, em que se possam realizar *drill up* e a soma de seus valores ao longo do tempo, julgue os itens a seguir, à luz dos conceitos afetos à modelagem dimensional.

- 110 A métrica *qtd\_faltas* é aditiva.
- 111 Todos os elementos descritos em **FATO\_FREQUÊNCIA** fazem parte de uma única chave primária composta, sendo cada elemento uma chave estrangeira oriunda de cada dimensão a ela vinculada.

Julgue os itens seguintes, referentes a conceitos e especificidades de MDM (*master data management*).

- 112 Como o MDM envolve controle sobre os valores e identificadores de dados mestres, é correto afirmar que um de seus objetivos é reduzir o risco de identificadores ambíguos.
- 113 Um dado mestre está no estado conhecido como *golden record* quando se encontra no seu estado mais puro, completo e confiável possível.
- 114 A fim de gerenciar a qualidade de dados, o MDM se serve de ferramentas de validação dos dados que ajudam na visualização de todo o fluxo de gestão dos dados mestres, o que torna possível, de maneira rápida, a identificação de quaisquer desvios em relação à política de dados da empresa.
- 115 Com a utilização de metadados, é possível, ao nível de elemento e do conjunto de dados, capturar classificações de segurança e sensibilidade regulatória e, assim, criar uma abordagem para gerenciar dados confidenciais.
- 116 A grande diferença entre algoritmos *fuzzy matching* e *stemming* reside no fato de que, enquanto o primeiro combina palavras com a mesma raiz linguística, o segundo trabalha com semelhanças de ortografia.
- 117 A descrição da arquitetura de dados de uma organização inclui uma coleção integrada de documentos de projeto mestre em diferentes níveis de abstração, excluindo-se dessa coleção os padrões que determinam o modo como os dados são coletados, armazenados, organizados, usados e removidos.

Em uma organização, foram elencadas as seguintes necessidades:

- I identificar os *data stewards* como parte da governança de dados e do planejamento da gestão dos dados;
- II implementar a gestão de qualidade de dados e a gestão de arquitetura de dados;
- III gerenciar fatores ambientais tais como pessoas e tecnologias afetas à governança de dados.

A partir da situação hipotética apresentada, julgue os itens a seguir, relativos à governança de dados.

- 118 Atender a necessidade I é importante, pois os *data stewards* são responsáveis, dentro da área de negócios, pelo controle e uso dos dados, assim como pela gestão dos metadados, que são elementos que ajudam a entender os objetos, a sua composição e seu relacionamento.
- 119 A implementação da gestão da necessidade III deve incluir a cultura organizacional e excluir os processos, uma vez que estes últimos não são abrangidos pelo DMBOK.
- 120 A necessidade II descreve duas áreas de conhecimento do DMBOK que, uma vez implementadas, ajudam a organização a entender as necessidades de informações da empresa, assim como a avaliar a adequação dos dados ao seu uso pretendido.

Espaço livre