

## BIG DATA PROFESSIONAL CERTIFICATE - BDPC™

### Sample Exam V092022

1. ¿Qué tipos de datos podemos encontrar?
  - A. Estructurados, semiestructurados, no estructurados
  - B. Estructurados, superestructurados, no estructurados
  - C. Estructurados, semiestructurados, no estructurados o todos los anteriores
  - D. Ninguna de las anteriores
  
2. ¿Cuáles son las 2Vs de Big Data que se unieron a las 3 principales para formar las 5Vs de Big Data?
  - A. Valor y volumen
  - B. Valor y velocidad
  - C. Valor y veracidad
  - D. Valor y visualización
  
3. ¿A cuál de las Vs de Big Data hace referencia hablar de tipos de datos?
  - A. Velocidad
  - B. Volatilidad
  - C. Visualización
  - D. Variedad
  
4. ¿Cuál de los siguientes enunciados hace referencia a lo que conceptualizaba la ley de Moore?
  - A. El procesamiento computacional aumentaría a mayor velocidad, sería más grande el hardware requerido para su procesamiento y su eficiencia aumentaría con el pasar de tiempo
  - B. El procesamiento computacional aumentaría a mayor velocidad, sería más pequeño el hardware requerido para su procesamiento y su eficiencia aumentaría con el pasar de tiempo
  - C. El procesamiento computacional aumentaría a menor velocidad, sería más pequeño el hardware requerido para su procesamiento y su eficiencia aumentaría con el pasar de tiempo
  - D. Ninguna de las anteriores

5. ¿Cuál de las siguientes afirmaciones es la correcta?

- A. Las bases de datos de código abierto pueden ser bases de datos únicamente SQL
- B. Las bases de datos de código abierto pueden ser bases de datos únicamente NoSQL
- C. Las bases de datos de código abierto no pueden ser bases de datos SQL, ni tampoco NoSQL
- D. Las bases de datos de código abierto pueden ser bases de datos tanto SQL como NoSQL

6. ¿Cuál de los siguientes tipos de bases de datos es el que se enfoca en combinar distintos tipos en un único servidor integrado?

- A. Base de datos en la nube
- B. Bases de datos de autogestión (o autónomas)
- C. Bases de datos distribuidas
- D. Bases de datos multimodelo

7. ¿Cuál de los siguientes modelos de bases de datos es el que incluye nociones de funciones, encapsulación e identidad de objetos?

- A. Modelo de datos relacional
- B. Modelo de datos entidad-relación
- C. Modelo de datos basado en objetos
- D. Ninguna de las anteriores

8. ¿Cuál de las siguientes NO es una fase o etapa del procesamiento de datos?

- A. Recopilación de datos
- B. Preparación de datos
- C. Interpretación o salida de datos
- D. Integración de los datos a la nube

9. ¿Cuál de las siguientes es la explicación que se le puede dar al nivel de incertidumbre como característica dependiente del proceso utilizado, el dominio y la interpretación?

- A. Variabilidad implícita que se deriva de las operaciones de agregación, y que modifica la percepción de precisión del valor puntual
- B. Variabilidad explícita que se deriva de las operaciones de agregación, y que modifica la percepción de precisión del valor puntual
- C. Variabilidad implícita que se deriva de las operaciones de connotación, y que modifica la percepción de precisión del valor puntual
- D. Ninguna de las anteriores

10. ¿A qué hace referencia la frontera más allá de la cual la información pierde sentido y aplicación?
- A. Alcance o ámbito
  - B. Rango de validez
  - C. Connotación
  - D. Ninguna de las anteriores
11. ¿Cuál de las siguientes NO es una característica de los clústeres o sistemas distribuidos?
- A. Independencia de componentes
  - B. Connotación
  - C. Concurrencia
  - D. Transparencia
12. ¿Cuál es la característica de los clústeres de datos o sistemas distribuidos que se da para la escalabilidad de los sistemas de datos distribuidos?
- A. Apertura
  - B. Concurrencia
  - C. Independencia de componentes
  - D. Modularidad
13. ¿A cuál de las siguientes hace referencia la letra "K" dentro de la sigla DIKW?
- A. Data
  - B. Sabiduría
  - C. Información
  - D. Conocimiento
14. El propósito hace referencia a que:
- A. Existe un objetivo asociado a la información que regula la manera de transformar los datos en información
  - B. Existe un objetivo asociado al conocimiento que regula la manera de transformar los datos en información
  - C. Posee una estructura de interrelaciones con otros conocimientos, bases teóricas y experiencias
  - D. Ninguna de las anteriores

15. Es una técnica para agrupar data no etiquetada de acuerdo con sus similitudes o diferencias
- A. Asociación
  - B. Reducción de dimensionalidad
  - C. Clustering
  - D. Ninguna de las anteriores
16. Es un algoritmo para asignar con precisión datos de prueba en categorías específicas
- A. Clasificación
  - B. Regresión
  - C. Clustering
  - D. Ninguna de las anteriores
17. ¿Quiénes se encontraron la World Wide Web y desarrollaron HTML, URL y HTTP?
- A. Tim Berners-Lee y Robert Cailou
  - B. Tim Berners-Lee y Robert Caiou
  - C. Tim Berners-Lee y Robert Kailliw
  - D. Tim Berners-Lee y Robert Cailliau
18. ¿A cuál de las grandes fases de la evolución de Big Data se hace referencia cuando se habla de contenido estructurado basado en SGBD?
- A. Fase 3 de Big Data: 2010-presente
  - B. Fase 1 de Big Data: 1970-2000
  - C. Fase 2 de Big Data: 2000-2010
  - D. Ninguna de las anteriores
19. ¿A qué se refiere la sigla HMI?
- A. Ninguna de las anteriores
  - B. Human-Machine Introduction
  - C. Human-Machine Introspective
  - D. Human-Machine Interface
20. ¿Cuál de las siguientes profesiones es la que Harvard Business Review consideró como la más deseada del siglo XXI en 2012?
- A. Desarrollador Blockchain
  - B. Científico de datos
  - C. Desarrollador de Hadoop
  - D. Ninguna de las anteriores

21. ¿Cuál de los siguientes organismos brinda la siguiente definición sobre gobernanza de datos?

"Es un sistema de derechos de decisión y responsabilidades para los procesos relacionados con la información, ejecutado de acuerdo con modelos acordados que describen quién puede tomar qué acciones con qué información y cuándo, bajo qué circunstancias, utilizando qué métodos."

- A. Data Governance Institute
- B. Data Government Institute
- C. Data Governance Institution
- D. Ninguna de las anteriores

22. Es la actividad del gobierno de datos donde se da el cumplimiento de las regulaciones de la organización tanto gubernamentales como del sector al que pertenece la organización.

- A. Cumplimiento estructural
- B. Cumplimiento legislativo
- C. Cumplimiento normativo
- D. Cumplimiento organizacional

23. ¿Cuáles de las siguientes son actividades del gobierno de datos?

- A. Definir estrategia de información, proyectos de gestión de datos, valoración de activos, promoción y comunicación
- B. Definir estrategia de datos, proyectos de gestión de datos, valoración de activos, promoción y comunicación
- C. Definir estrategia de datos, proyectos de gestión de información, valoración de activos, promoción y comunicación
- D. Ninguna de las anteriores

24. ¿Cuál de las siguientes es una ventaja del gobierno de datos?

- A. Toma de decisiones más acertadas y oportunas
- B. Mejora los controles de costes
- C. Todas las anteriores
- D. Gestión de riesgos con más facilidad

25. ¿A qué hace referencia el data steward?

- A. Al propietario de datos
- B. Al registrador de datos
- C. AL comité directivo
- D. Al administrador de datos

26. ¿En cuál de los siguientes sectores se tiene como un desafío importante el fraude y compliance?

- A. Financiero
- B. Salud
- C. Telecomunicaciones
- D. Gobierno

27. ¿Cuáles son los desafíos de los servicios financieros?

- A. Fraude y compliance, antilavado de criptos, impulsar la innovación, análisis normativo y de cumplimiento financiero
- B. Fraude y compliance, antilavado de dinero, impulsar la tecnología, análisis normativo y de cumplimiento financiero
- C. Fraude y compliance, antilavado de dinero, impulsar la innovación, análisis normativo y de cumplimiento financiero
- D. Fraude y compliance, antilavado de dinero, impulsar la innovación, análisis normativo y de cumplimiento legal

28. El customer lifetime value es un desafío propio del sector:

- A. Retail
- B. Financiero
- C. Salud
- D. Gaming

29. La optimización de producción es un desafío propio del sector:

- A. Retail
- B. Telecomunicaciones
- C. Bienes raíces
- D. Manufactura

30. Los informes Ad Hoc y OLAP forman parte de:

- A. Analítica tradicional
- B. Analítica avanzada
- C. Analítica prescriptiva
- D. Ninguna de las anteriores

31. ¿Cuál de las siguientes NO es una de las mejores prácticas de BI?
- A. Alinear Big Data con objetivos empresariales generales
  - B. Alinear Big Data con objetivos empresariales específicos
  - C. La principal ventaja reside en alinear los datos estructurados y no estructurados
  - D. Planificar el laboratorio de hallazgos en pro del rendimiento
32. ¿Cuál es el método de BI encargado de convertir el análisis de datos en representaciones visuales como tablas, gráficos e histogramas para consumir datos más fácilmente?
- A. Consultas
  - B. Visualización de datos
  - C. Análisis visual
  - D. Métricas de rendimiento y evaluación comparativa
33. ¿Cuál de las siguientes NO es una característica de un almacén de datos?
- A. Diseño de la nube
  - B. Diseño de esquemas
  - C. Modificaciones de datos
  - D. Carga de trabajo
34. ¿A cuál de las fases de la evolución de BI y Analytics se relaciona web intelligence y analytics?
- A. BI&A 4.0: 2020
  - B. BI&A 1.0: 1990
  - C. BI&A 3.0: 2010
  - D. BI&A 2.0: 2000
35. ¿Cuál de las siguientes afirmaciones es la correcta?
- A. La preparación de datos (Data preparation) es la primera fase de la metodología CRISP-DM
  - B. El entendimiento del negocio (Business understanding) es la primera fase de la metodología CRISP-DM
  - C. El modelado (Modeling) es la primera fase de la metodología CRISP-DM
  - D. La evaluación (Evaluation) es la primera fase de la metodología CRISP-DM

36. ¿Cuál de las siguientes afirmaciones NO es la correcta?
- A. El despliegue o implementación (Deployment) es la última de la metodología CRISP-DM
  - B. El entendimiento del negocio (Business understanding) es la primera fase de la metodología CRISP-DM
  - C. El despliegue o implementación (Deployment) es la primera fase de la metodología CRISP-DM
  - D. Ninguna de las anteriores
37. ¿CRISP-DM fue la evolución de cuál metodología?
- A. Agile
  - B. Six Sigma
  - C. SEMMA
  - D. Ninguna de las anteriores
38. ¿Con cuál de las siguientes fases empieza la metodología SEMMA?
- A. Exploración (Explore)
  - B. Muestreo (Sample)
  - C. Modelado (Model)
  - D. Valoración (Assess)
39. Explotar sucesos pasados para poder predecir eventos futuros hace parte del tipo de análisis:
- A. Prescriptivo
  - B. Predictivo
  - C. De diagnóstico
  - D. Descriptivo
40. ¿Cuál NO es una característica del método de agrupamiento del aprendizaje no supervisado?
- A. Aglomerativo
  - B. Probabilístico
  - C. Solapamiento
  - D. Datos etiquetados



41. ¿Cuál tipo de aprendizaje es en el que se desconoce sobre el resultado deseado?
- A. Aprendizaje supervisado
  - B. Aprendizaje automático
  - C. Ninguna de las anteriores
  - D. Aprendizaje no supervisado
42. El \_\_\_\_\_ es la capacidad de encontrar distintas soluciones a los problemas que se afrontan, lo cual requiere de una profunda investigación con todos los recursos que estén a su disposición, pero sin transgredir en ningún momento la gobernanza, privacidad y tratamiento de los datos manejados
- A. Hacking normativo
  - B. Hacking organizacional
  - C. Hacking ético
  - D. Ninguna de las anteriores
43. ¿Cuál es la institución que provee la arquitectura de referencia de Big Data?
- A. National Institute of Stereotypes and Technology
  - B. National Industry of Standards and Technology
  - C. National Institute of Standards and Technology
  - D. National Incorporation of Standards and Technology
44. ¿Cuál de los siguientes se encuentra dentro de la arquitectura de Referencia de Big Data del NIST?
- A. Proveedor del marco Big Data (Big Data framework provider)
  - B. Todas las anteriores
  - C. Proveedor de datos (data provider)
  - D. Consumidor de datos (data consumer)
45. ¿Cuáles son los dos tipos de trabajos que se encuentran en la arquitectura MapReduce?
- A. Mapeadores y aumentadores
  - B. Mapeadores y relativos
  - C. Ninguna de las anteriores
  - D. Mapeadores y reductores

46. ¿Cuál de las siguientes NO es un tipo de base NoSQL?
- A. Valor Creación
  - B. Valor Clave
  - C. Familia de Columnas
  - D. Documentos
47. ¿Cuál de los siguientes funciona como un ecosistema de datos, es gratuito y de código abierto, usado para procesar grandes volúmenes de datos en bloques usando modelos de programación?
- A. Apache Storm
  - B. Apache Hadoop
  - C. Apache Spark
  - D. Apache Kafka
48. ¿Para qué es utilizado el ZooKeeper en Apache Kafka?
- A. Para almacenar información sobre el clúster de Kafka, así como los detalles de los clientes consumidores
  - B. Para almacenar información sobre el clúster de Kafka, así como los detalles de los clientes productores
  - C. Para servir como puente entre productores y consumidores.
  - D. Para componerse de diferentes corredores, temas y sus respectivas particiones
49. ¿Cuál de los siguientes es un tipo de servicio de la nube?
- A. Todas las anteriores
  - B. Nube pública
  - C. Nube privada
  - D. Nube híbrida
50. Dataproc forma parte de:
- A. Google Cloud Platform
  - B. Amazon Web Services
  - C. Microsoft Azure
  - D. Oracle Cloud Infrastructure

RESPUESTAS

- |       |       |
|-------|-------|
| 1. C  | 26. A |
| 2. C  | 27. C |
| 3. D  | 28. A |
| 4. B  | 29. D |
| 5. D  | 30. A |
| 6. D  | 31. A |
| 7. C  | 32. B |
| 8. D  | 33. A |
| 9. A  | 34. D |
| 10. A | 35. B |
| 11. B | 36. C |
| 12. D | 37. C |
| 13. D | 38. B |
| 14. B | 39. B |
| 15. C | 40. D |
| 16. A | 41. D |
| 17. D | 42. C |
| 18. B | 43. C |
| 19. D | 44. B |
| 20. B | 45. D |
| 21. A | 46. A |
| 22. C | 47. B |
| 23. B | 48. A |
| 24. C | 49. A |
| 25. D | 50. A |