

una regresión lineal $h_{\theta} = \theta_0 + \theta_1 X$ es

☐

un algoritmo de aprendizaje paramétrico, inflexible y sin supervisión

☒

un algoritmo de aprendizaje supervisado inflexible y paramétrico

☐

un algoritmo de aprendizaje paramétrico, inflexible y sin supervisión

☐

un algoritmo de aprendizaje supervisado inflexible no paramétrico

un sistema DFS tiene ventajas y desventajas, Estas son

correcta

☐

TOLERANCIA AL FALLO

☒

BAJA CONSISTENCIA

☐

ALTA CONSISTENCIA

☒

ALTA CONCURRENCIA

Las características de un DFS son (señala todas las correctas)

☒

permite la tolerancia frente al fallo

☐

a mayor replicación menor consistencia

Opción omitida: incorrecto

☒

permite la alta concurrencia

☒

divide los archivos en partes homogéneas en distintas máquinas

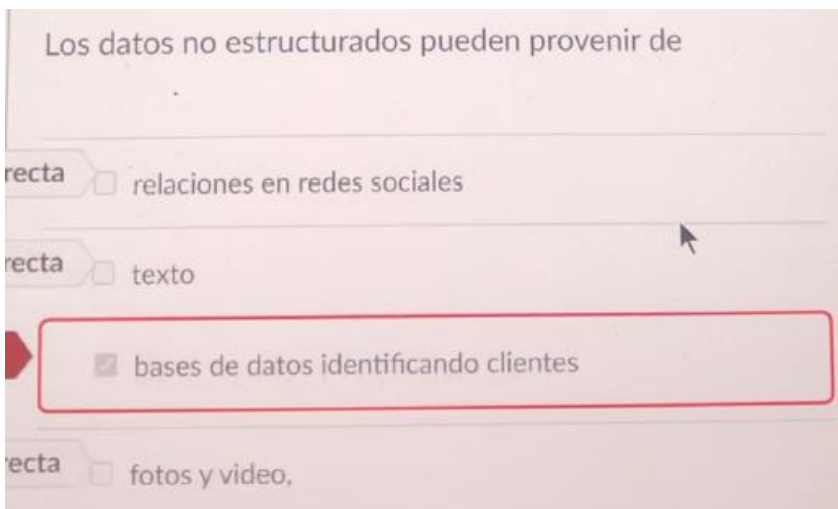
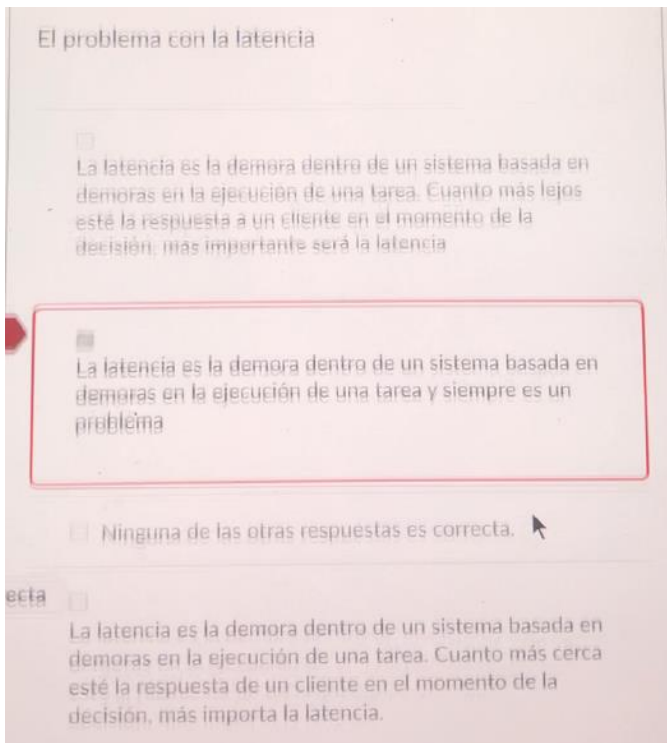
un datamart es

☐ un servicio de coordinación

☐ un ETL

☒ un subconjunto de un almacén de datos orientado a una línea de negocio específica. Los data marts contienen depósitos de datos resumidos recopilados para su análisis en una sección o unidad específica

☐ un servicio de workflow



el big data stack se refiere a



a la arquitectura de la plataforma y programabilidad del big data



A las posibilidades de análisis



a las alternativas de visualización



Ninguna de las otras respuestas es correcta.

Para ejecutar la acción de contar palabras con MapReduce usando hadoop



Ninguna de las anteriores es correcta



Wordcount en la práctica de hadoop no realiza una tarea MapReduce



Primero carga el archivo en el sistema usando HDFS y luego ejecuta la aplicación de recuento de palabras basada en MapReduce sobre él



Simplemente ejecuta MapReduce sobre el archivo

Un sistema que abre la puerta con la huella digital es



a prediction (una predicción)



an inference (una inferencia)



a Recommendation system (sistema de recomendación)



Cognitive system (sistema cognitivo)

La **varianza** se refiere al error en el resultado debido al cambio del **conjunto de entrenamiento**.

El **sesgo** se refiere al error en el resultado debido al cambio en la **forma de la función**

Y en **métodos más flexibles**, la varianza aumentará y el sesgo disminuirá. por lo tanto, en muchas ocasiones, un método simple supera a los más flexibles

La escalabilidad horizontal

☐ Ninguna de las otras respuestas es correcta

☐ Añadiendo más procesadores y RAM a cada equipo y con redes más rápidas

☐ Añadir más procesadores en el mismo equipo e incrementar RAM. Nodos más grandes

☒ implica agregar más máquinas, posiblemente menos potentes, a una red relativamente más lenta. Muchos nodos de productos básicos en un clúster

☐ Primero clusterizar, Luego utilizar un algoritmo supervisado de predicción que determine con precisión las medidas de la prenda personalizada más adecuada y luego clasificar por KNN

recta

☐ Existen varias alternativas si bien podríamos primero clusterizar para encontrar grupos (si bien podría no ser necesario) y dentro de ellos utilizar KNN

☒ Existen varias alternativas si bien podríamos primero clusterizar para encontrar grupos (si bien podría no ser necesario) y dentro de ellos utilizar un algoritmo supervisado de clasificación.

Serialization es

☒ Necesario para la carga y conversión de datos estructurados y no estructurados en Hadoop

cta

☐ La serialización es el proceso de traducir las estructuras de datos o el estado de los objetos en forma binaria o textual para transportar los datos a través de la red o almacenarlos en algún almacenamiento persistente.

Virtualization (Virtualización es)

☐ es un sistema para hacer posible la concurrencia

☐ hace posible asignar tareas a un conjunto de recursos físicos

☐ es solo por tolerancia a fallas

☒ hace posible asignar tareas a un conjunto de recursos virtuales

Las inferencias están relacionadas con



¿Cuál es la relación entre la respuesta y cada predictor?



Qué predictores están asociados con la respuesta

recta



Qué predictores no están asociados con la respuesta

¿qué pasaría probablemente si el parámetro alfa fuera 0,1)



simplemente tardaría más en llegar al error mínimo



que el método convergería muy rápidamente hasta llegar al error mínimo



que el método convergería muy rápidamente hasta llegar al error cero

recta



que el método no convergería. Estos es, el error cada vez sería más grande

number of train linear regression	4	850	1,78	195,50	17,50	306,25
				suma	-69,50	39551,25
					(100,01)	

¿qué pasaría si con los parámetros de inferencia si cambio o añado alguno de esos datos de aprendizaje?



Que la predicción sería peor



que se alcanzaría el mínimo de error antes, si bien sólo, si la muestra es más grande

recta



que los parámetros de inferencia cambiarían, ajustándose mejor a la nueva muestra



que los parámetros de inferencia serían los mismo

K-means y KNN

¿Cuáles de las siguientes afirmaciones son CORRECTAS referidas a las dos metodologías?

☒ KNN es supervisado y K-means no supervisado

☐ ambos tienen el mismo propósito

☐ ambos son no supervisados

☐ El propósito de K-means es clasificar y K-NN es encontrar grupos

☐ El propósito de K-means es encontrar grupos, y K-NN es clasificar un nuevo elemento dado

un data mart

☒ un subconjunto de un almacén de datos orientado a una línea de negocio específica. Los data marts contienen depósitos de datos resumidos recopilados para su análisis en una sección o unidad específica

☐ un ETL

☐ un servicio de coordinación

☐ un servicio de workflow

Las inferencias están relacionadas con

☒ Qué predictores no están asociados con la respuesta

☒ Qué predictores están asociados con la respuesta

☒Cuál es la relación entre la respuesta y cada predictor

El auge de la IA ha tenido lugar porque

☐ ninguna de las otras repuesta es correcta

☐

porque el rendimiento de las redes neuronales no aumenta cuando aumenta la cantidad de datos. Por el contrario, la IA tradicional aumenta el rendimiento hasta cierto volumen de datos.

ecta

☐

porque el rendimiento de las redes neuronales no aumenta cuando aumenta la cantidad de datos. Por el contrario, la IA tradicional no aumenta el rendimiento hasta cierto volumen de datos.

☐

porque el rendimiento de las redes neuronales aumenta cuando aumenta la cantidad de datos. Por el contrario, la IA tradicional no aumenta el rendimiento hasta cierto volumen de datos.

onder

Pregunta 35

0 / 0.1 pts

La Ley de Moore se refiere a la percepción de Moore de que la cantidad de transistores en un microchip se duplica cada _____ (una palabra) años, aunque el costo de las computadoras se reduce a la mitad.

ido

s correctas dos

2

Data Munging es el proceso inicial de refinación de datos en bruto en contenido o formatos más adecuados para el consumo por sistemas y usuarios de corriente. Este concepto incluye (seleccione todo lo que corresponda)

- ☒ enriquecimiento de datos
- ☒ validación de datos
- ☒ transformación de datos e integración
- ☒ explotación de datos

correcta

La persistencia garantiza que los datos almacenados en una base de datos no serán _____ (una palabra en forma verbal adecuada y minúsculas) sin permisos y que estarán disponibles si son importantes para el negocio.

inmutables

correctas

- variados
- modificados
- cambiados
- alterados

A lo largo de los años, el lenguaje de consulta estructurado (SQL) ha evolucionado al mismo ritmo que la tecnología RDBMS y es el mecanismo más utilizado para crear, consultar, mantener y operar bases de datos relacionales. Estas tareas se denominan CRUD (seleccione todas las que correspondan)

- ☒ Update (actualizar)
- ☒ Create (crear)
- ☐ Clean) limpiar)
- ☐ Retrieve (extraer)

correcta

Mientras que RDBMS usa ACID (Atomicidad, Consistencia, Aislamiento, Durabilidad) como mecanismo para garantizar la consistencia de los datos, DBMS no relacional usa _____ (acrónimo en inglés y mayúsculas)

base

correctas base

BASE

Datos en movimiento: analizados a medida que se generan. De lo contrario, los registros se recopilan en una ubicación de búfer (es decir, se crea un lote de registros) y se envían a la base de datos en fragmentos discretos a intervalos periódicos (proceso por lotes_Batch process)

_____: (tres palabras, incluyendo la preposición. puede ser en inglés o español. minúsculas) recopilados antes del análisis

data at rest

correctas datos en reposo

data at rest

Streaming Computing está diseñado para manejar un flujo continuo de una gran cantidad de datos no estructurados. Por el contrario, _____ (utilice las iniciales o el acrónimo en INGLÉS Y MAYÚSCULAS) generalmente trata con algunas variables que deben correlacionarse con un proceso comercial específico.

CEP

correctas CEP

cep

Un _____ (una palabra en inglés y minúsculas) de datos es un tipo de base de datos que integra copias de datos de transacciones de sistemas de origen dispares y los aprovisiona para uso analítico.

datawarehouse

Los datalakes garantizan que todos los datos se almacenen para un uso potencialmente desconocido más adelante: schema on ____; esquema en _____(añadir **una sola palabra. válido en español ó inglés minúsculas**)

read

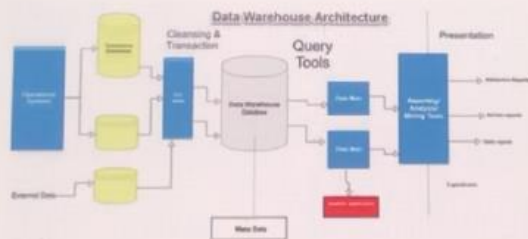
correctas read

lectura

Las tecnologías emergentes de memoria principal no volátil (NVMM), por lo general, ofrecen una memoria de _____ (**una sola palabra, español, minúsculas**) mucho más alta, un costo por bit y un consumo de energía en espera mucho más bajos que la DRAM.

densidad

en la arquitectura abajo indicada, el datawarehouse se usa porque



Incorrecta

para preparar data marts para realizar diferentes análisis para diferentes departamentos en una empresa

☒ tenemos varias bases de datos transaccionales y hay que combinar y preparar para el análisis

En la arquitectura de la figura se combinan (seleccionar todos los que aplican)

Incorrecta

☐ streaming data y data at rest

Incorrecta

☐ datos estructurados y no estructurados

☒ bases de datos distribuidas sobre hadoop y datawarehouses

Incorrecta

☐ datos internos y externos

Si estamos interesados al analizar principalmente en _____ (una palabra, español, minúsculas), entonces los modelos restrictivos son mucho más interpretables

organizar

Incorrectas inferencia

INFERENCIA

AI performance

☐ increases exponentially when the amount of data increases

☒ increases more with large neural networks and a big amount of data

☐ increases when the amount of data increases until a certain point using traditional AI

Opción omitida: incorrecto

☐ AI has really taken off recently due to the rise of neural networks and deep learning

Opción omitida: incorrecto

Valence (select all that apply)

☒ That higher valence results in greater coding, that is greater connectedness

☒ refers to the many types of data that are available

Respuesta seleccionada: incorrecta

☐ Is the fast rate at which data is received and (perhaps) acted on

☒ refers to connectedness

Over the years, the **structured query language (SQL)** has evolved in lock step with RDBMS technology and is the most widely used mechanism for creating, querying, maintaining, and operating relational databases. These tasks are referred to as **CRUD** (select all that apply)

☐ Clean

☒ Retrieve

☒ Update

☒ Create

47. A _____ is a massive storage depository with huge processing power and ability to handle a very large number of concurrences, data management and analytical tasks. In a data lake is not storage into a ware unless they are in use

☒ data lake

50. The accuracy of Y as a prediction for Y depends on two quantities, which we will call

☐ variance and bias

☐ high bias and irreducible error

☐ reducible error and high bias

☐ reducible error and the irreducible error

☒ (sin respuesta)

Respuesta correcta: **reducible error and the irreducible error**

Data munging is the initial process of refining raw data into content or formats better suited for consumption by downstream systems and users. This concept includes (select all that apply)

- ☒ data transformation and integration
- ☒ data validation
- ☒ data enrichment
- ☒ data exploration