



# EXAMEN

ASIGNATURA: CALIDAD Y PRE-PROCESAMIENTO

NOMBRE DOCENTES: MTRO. JACOBO GERARDO GONZÁLEZ LEON Y DRA. MARIA DEL PILAR ANGELES

NOMBRE DEL ESTUDIANTE

No.DECUENTA: \_\_\_\_\_ Apellido paterno \_\_\_\_\_ apellido materno \_\_\_\_\_ nombre(s) \_\_\_\_\_  
TIPO DE EXAMEN Teoría ☒ PARCIAL NUMERO 1 CALIFICACION: ☐ FINAL

## CALIDAD DE DATOS

1. ¿Qué significa correctes (accuracy)? **(1 punto)**
  - ☐ Cuando una entidad del mundo real es representada solamente una vez
  - ☐ El grado en el que los datos no faltan y son suficientes a lo largo y a lo ancho para la tarea que se desea
  - ☐ Grado de concordancia entre el valor de un dato o colección de datos y la fuente de datos en que éste es correcto
  - ☐ Medición de que tan bien los atributos almacenados en una fuente están llenos con datos reales
  - ☐ La edad promedio de los datos en la fuente de datos.
  - ☐ Medición del número de registros que una fuente de datos almacena. Probabilidad que una entidad del mundo real este representada en la fuente de datos.
2. ¿Qué significa completitud (completeness)? **(1 punto)**
  - ☐ Cuando una entidad del mundo real es representada solamente una vez
  - ☐ El grado en el que los datos no faltan y son suficientes a lo largo y a lo ancho para la tarea que se desea
  - ☐ Grado de concordancia entre el valor de un dato o colección de datos y la fuente de datos en que éste es correcto
  - ☐ Medición de que tan bien los atributos almacenados en una fuente están llenos con datos reales
  - ☐ La edad promedio de los datos en la fuente de datos.
  - ☐ Medición del número de registros que una fuente de datos almacena. Probabilidad que una entidad del mundo real este representada en la fuente de datos.
3. ¿Qué significa densidad (density)? **(1 punto)**
  - ☐ Cuando una entidad del mundo real es representada solamente una vez
  - ☐ El grado en el que los datos no faltan y son suficientes a lo largo y a lo ancho para la tarea que se desea
  - ☐ Grado de concordancia entre el valor de un dato o colección de datos y la fuente de datos en que éste es correcto
  - ☐ Medición de que tan bien los atributos almacenados en una fuente están llenos con datos reales
  - ☐ La edad promedio de los datos en la fuente de datos.
  - ☐ Medición del número de registros que una fuente de datos almacena. Probabilidad que una entidad del mundo real este representada en la fuente de datos.

## INTEGRACIÓN DE DATOS

4. ¿Cuáles son los tipos de heterogeneidad dentro de los componentes del modelo de datos? **(0.25 punto)**
5. ¿Qué es y para qué sirve la correspondencia de esquemas? **(0.5 punto)**
6. ¿Qué es y para qué sirve el mapeo de esquemas? **(0.5 punto)**
7. ¿Por qué la arquitectura de tres niveles no es adecuada para los sistemas federados? **(0.25 punto)**
  - ☐ Porque no soporta las dimensiones de distribución, heterogeneidad y autonomía de los sistemas federados
  - ☐ Porque no soporta las dimensiones de centralización, homogeneidad y autonomía de los sistemas federados
8. ¿Cuáles son los 5 niveles de la Arquitectura propuesta por Sheth y Larson para Sistemas de Bases de datos Federados? **(0.5 punto)**
  - ☐ Esquema local, esquema componente, esquema de exportación, esquema federado y esquema externo
  - ☐ Esquema remoto, esquema distribuido, esquema de exportación, esquema federado y esquema externo
9. Identifica el Esquema Componente **(0.25 punto)**
  - ☐ Se deriva de la traducción de los esquemas locales a un modelo de datos llamado Modelo de Datos Canónico Común (el modelo de datos que se usará como estándar).
  - ☐ Se forma de la integración de varios esquemas de exportación o modelos de datos, conocido también como Esquema Global o Esquema de Importación.
  - ☐ Es un subconjunto de información de un esquema federado que es relevante a una clase o aplicación de usuario.

(0.25 punto)

10. Identifica el Esquema de Exportación

- ( ) Se forma de la integración de varios esquemas de exportación o modelos de datos, conocido también como Esquema Global o Esquema de Importación.
- ( ) Es un subconjunto de información de un esquema federado que es relevante a una clase o aplicación de usuario.
- ( ) Es un subconjunto de un esquema componente que está disponible a la federación. Este puede incluir información de control de acceso para uso de algún usuario de la federación.

PERFILADO DE DATOS

(1 punto)

11. Menciona tres tipos de perfilado de datos provenientes de múltiples fuentes

(1 punto)

12. Menciona los pasos para el proceso de de-duplicación de registros

(0.5 punto)

13. ¿Cómo puedes calcular la oportunidad (timeliness) de los datos?

(2 puntos)

14. Dado el siguiente par de registros realiza la comparación del apellido paterno por el método Q-gram, con  $Q=2$  y coeficiente Overlap e identifica si son correspondientes, no correspondientes o requieren revisión manual con ayuda de los umbrales  $t_u = 6$  y  $t_l = 4$

	Nombre	Ap. Paterno	Ap. Materno	Dia_Fecha_Nac	Nombre calle	#Calle	Colonia	SimSum
A <sub>11</sub> B <sub>11</sub>	0.57	Cortez Perez	0.084	0.9	0.33	0.97	0.047	
A <sub>12</sub> B <sub>12</sub>	0.14	Moncada Montalvo	0.46	1	1	1	0.25	