

Comenzado en	Thursday, 12 de November de 2020, 10:00
Estado	Terminados
Finalizado en	Thursday, 12 de November de 2020, 12:00
Tiempo empleado	1 hora 59 mins
Calificación	Sin calificar aún

Pregunta **1**
Completada
Puntaje de 1

- A. Describa brevemente el uso que tienen la matriz de confusión

B. Es posible evaluar la efectividad de un modelo si los falsos negativos sean menores que los falsos positivos
- A. Se usa básicamente para mostrar que tan bueno es un modelo o algoritmo ya que muestra los posibles casos errores y certezas que se pueden dar como falsos negativos y falsos positivos.

B. Creeria que no, porque si tomamos como un Error cualquiera los falsos negativos (si serían errores) son menos que los falsos positivos (no serían errores) entonces en el modelos diríamos que hay muchos errores y esto puede hacer que tomemos una errónea decisión

Pregunta **2**
Completada
Puntaje de 1

- A. Describa con sus propias palabras los beneficios que puede tener un cubo multidimensional en disco sobre un datamart en una base de datos relacional.

B. Indique al menos una herramienta que pueda utilizar para desarrollar cada una de estas opciones.

C. Explique los componentes de tecnología que pueden hacer que un datamart responsa igual o mejor que un cubo multidimensional al resolver una consulta.
- A. Velocidad
Operaciones más simples entre los datos
estructuras que todos pueden entender

B. Oracle Olap 11g, IBM Cognos (Cubos)
SQL Server Integration Services (Datamart)

C. Tener datos esenciales del negocio, agregarle una columna adicional de métrica y crearle un modelo o esquema adecuado .

Pregunta **3**
Completada
Puntaje de 1

- A, Describa con sus propias palabras los 3 principales beneficios que puede tener para una empresa la implementación de Gobierno de datos.
1. Seguridad de datos

2. Privacidad de datos

3. Estandarización de los datos lo que llevaría a una mejor gestión de los mismos

4. Mejor toma de decisiones

Pregunta **4**

Completada

Puntaje de 1

A. Explique cómo puede resolver la construcción de un datamart con las siguientes características:

- Se cuenta con una clasificación de productos que tiene en su nivel más bajo el SKU del producto que es único.
- Los productos están agrupados por una clasificación general según el departamento al que pertenecen y las líneas disponibles en cada departamento.
- Un 60 % de los productos son alimentos para los cuales se cuenta con una clasificación de si requieren refrigeración o no.
- Un 25% de los productos se reclasifican de una línea a otra al menos una vez al año.

Dimensiones:

Lugar (País, Departamento, linea1, linea2, linea3...), Productos (LineaPertenece, CodProducto, nombre ...SKU), Tiempo (Año, mes, día)

Esa seria la estructura que debería de tener el Datamart para así poder ingresar los datos por medio de un ETL

Pregunta **5**
Completada
Puntaje de 1

Para el caso descrito a continuación deberá identificar las dimensiones con sus niveles, métricas y atributos relevantes para construir un modelo estrella que permita resolver las consultas analíticas que se indican. El diseño deberá identificar cada una de las tablas con sus campos que debe contener el datamart.

Una compañía dedicada a la importación, producción y venta internacional de alimentos requiere de un modelo analítico para conocer mejor el comportamiento de sus ventas. la compañía posee 500 productos diferentes entre abarrotes y productos enlatados. Todo sus productos se venden através de tres canales diferentes (ventas en línea, mayoristas y tiendas especializadas). las ventas en línea y las ventas a mayoristas poseen cada una un vendedor mientras que las tiendas minoristas están organizadas por vendedores ruteros y supervisores. Los productos se clasifican por categorías y dentro de cada categoría existen líneas de productos. Se cuenta tanto con marcas internacionales como con marcas propias, cuyos productos. En cada canal se desarrolla un presupuesto por marca para todo el año detallado por mes en unidades y monto en US\$.

Consultas analíticas:

Análisis de las ventas a través de los diferentes canales de distribución con una frecuencia diaria, semanal y mensual.

Análisis de utilidad marginal por producto, línea y categoría. (Utilidad marginal = precio unitario de venta - costo unitario)

Análisis de ventas en unidades de productos por marca identificando las marcas importadas y las producidas localmente..

Análisis de metas de venta en unidades y en US\$ por vendedor, y canal

Dimensiones

1. Lugar

- 1. País
- 2. Departamento
- 3. Municipio

2. Productos

- 1. Cod producto
- 2. linea
- 3. categoria
- 4. precio unitario venta
- 5. costo unitario

3. Canal

- 1. ventas en línea
- 2. mayoristas
- 3. tiendas especializadas
- 4. Presupuesto

4. Personal

- 1. Vendedor rutero
- 2. Vendedor mayorista
- 3. Supervisor

5. Marca

- 1. nombre
- 2. codMarca
- 3. propia
- 4. Internacional

6. Tiempo

- 1. Año
- 2. Mes
- 3. Dia

Pregunta **6**
Completada
Puntaje de 1

A. Describa cuál sería la herramienta o sus características para tener una adecuada implementación de un modelo analítico de indicadores (KPI)

- 1. Indicadores tienen que tener valores de tolerancia e incerteza
- 2. Los indicadores deben tener un sentido
- 3. Tener asociado otro indicador o bien tener una relación
- 4. Ser numéricos

Pregunta **7**
Completada
Puntaje de 1

- A. Con sus propias palabras explique la diferencia entre construir un modelo estrella y construir un modelo copo de nieve (snowflake)
- B. Muestre un ejemplo de un modelo copo de nieve para una dimensión de 3 niveles inventada por usted. Indique claramente la jerarquía de los niveles y todos los campos que pueden existir, detallando el tipo de campo que puede utilizar tanto para la dimensión como para la tabla de hechos con la cual se relacione. No es necesario adicionar métricas a la tabla de hechos (Fact)

A. El modelo estrella su estructura está compuesta por una tabla de hechos y varias tablas de dimensión alrededor de ella y esas no se pueden relacionar entre sí , lo que hace que las extraccion de informacion sea mas rapida. Mientras que el modelo copo de nieve en su estructura cuenta con una tabla de hechos central relacionadas con tablas de dimensiones pero estas tablas de dimensión se pueden descomponer por decir algo en más dimensiones hacia abajo lo que hace que no todas las tablas de dimensión tengan relación con la tabla de hechos central lo que provoca que la extracción de información mas difícil y lenta.

B. modelo para venta de carros

Dimensiones:

Tiempo

- 1. Año
- 2. mes
- 3. dia

Producto

- 1. codigo producto
- 2. nombre
- 3. color

Sub Categoria de Producto (Tipo modelo)

- 1. cod modelo
- 2. nombre modelo

Pregunta **8**
Completada
Puntaje de 15

Usted ha sido contratado por una empresa dedicada a la producción y distribución de alimentos, con operaciones en varios países del mundo. la empresa cuenta con 10 plantas de producción en 5 países y vende sus productos en todo el mundo a partir de oficinas locales en cada país. Actualmente la compañía cuenta con 3 servidores independientes de datawarehouse desarrollados por los departamentos de Producción, Finanzas y Ventas, en los cuales han almacenado la información detallada de producción, distribución y venta de todos sus productos durante los últimos 10 años. La implementación de estos datawarehouse surgió como iniciativa de estas áreas de negocio de manera independiente del área de tecnología. El presidente de la compañía frecuentemente se topa con informes contradictorios proporcionados por las áreas de finanzas, producción y ventas ya que detalles por ejemplo del volumen de unidades vendidas por país no cuadran entre si. para resolver este problema el área de auditoria a realizado una revisión de todos los sistemas transaccionales que alimentan cada uno de los datawarehouse y a encontrado que toda la información es correcta y cuadra a nivel detallado entre todos los departamentos. Los ejecutivos de las diferentes áreas de negocio de la compañía, así como los gerentes regionales se quejan frecuentemente que no cuentan con información correcta, por lo que deben invertir mucho tiempo en llevar registros paralelos en excel para poder tomar sus decisiones.

El director de mercadeo ha plateado que la solución es contratar a un equipo de científicos de datos y a solicitado presupuesto para la construcción de un sistema de Big data con hadoop para hacer un lago de datos y así poder integrar toda la información de los 3 datawarehouse y poder tomar las decisiones correctas rápidamente.

El presidente de la compañía recientemente a contratado un nuevo gerente de tecnología quien ha indicado que el problema es haber dejado a las áreas desarrollar sus propios datawarehouse y ha solicitado reasignar todos los recursos responsables de los datawarehouse de estas áreas (personal, hardware y software) para desarrollar un nuevo modelo analítico definido por el área de tecnología.

Usted esta contratado como experto directamente por la junta directiva de la compañía quienes le han pedido su opinión respecto de la situación y que recomendación puede darles para asegurar que en el mediano plazo puedan llevar la compañía a mejorar su eficiencia operativa en todas las áreas. Detalle claramente el porqué de su opinión.

Creo que lo mejor sería hacer el proceso de ETL y para la información existente y con esa información hacer un cubo para ya asi poder consultar los datos que se deseen. Esta opción sería mejor a volver a crear los data warehouse ya que al final seria misma información la que se le asignaría al nuevo data warehouse, porque a consideración mia la estructura no cambiaría

Pregunta **9**
Completada
Puntaje de 1

- A. Identifique 3 potenciales usos que pueda tener una cadena de farmacias para los modelos predictivos y
- B. qué consideraciones debe cuidar para asegurar el éxito de sus modelos.

- A.
- 1. Tener mas cantidad de medicina para ciertas enfermedades que mas tengan demanda
 - 2. Siempre contar con stock de medicinas mas solicitadas
 - 3. Optimizar dinero en almacenaje de medicinas
 - 4. Optimizar espacio de almacenamiento de medicinas

- B.
- Se debe cuidar la cantidad de errores que puedan dar asi como tambien la cantidad de datos que usen para construirlos

Pregunta **10**
Completada
Puntaje de 1

- A. Explique cuales pueden ser los beneficios de utilizar una herramienta de ETL para la construcción de un datawarehouse.
- B. Proporcione dos ejemplos de herramientas comerciales que puede utilizar para el efecto.
- C. Liste cuales opciones pueden ser válidas si no desea realizar las actividades en la secuencia de E --> T --> L y cuál de ellas podría recomendar si el destino de la información es un appliance analítico.

- A.
- 1. Datos limpios
 - 2. Datos ordenados
 - 3. Datos utiles

- B. Bi Tool, CloverETL, Cognos Decisionstream (IBM)

- C.
- Archivos de texto
- CSV

Pregunta **11**
Completada
Puntaje de 1

- A. Explique con sus palabras las consideraciones que debe tener al diseñar la tabla dimensional de un modelo estrella, para una dimensión de dos niveles, en la cuál se realizan frecuentemente cambios en la descripción que deben quedar registrados en el tiempo.
- B. Realice un ejemplo para una dimensión de 3 niveles inventada por usted, asegúrese de indicar claramente el concepto de cada campo y el tipo de campo a utilizar.
- Debe de tener datos suficientes para luego hacer los reportes

Pregunta **12**
Sin contestar
Puntaje de 1

- A. Proporcione 3 ejemplos de soluciones que puede implementar un supermercado de algoritmos predictivos.
- B. Consideraría válido utilizar machine learning para encontrar la ruta óptima que debe seguir un camión de reparto para minimizar el tiempo de entrega y consumo de combustible? Explique su respuestas.

◀ Tarea Cloudera

Ir a...