

Tarea de Diseño e Implementación

Arquitecturas de Flujo de Datos - ETL

A. Fecha de entrega

La tarea se hace en equipo y la fecha de entrega es el jueves 30 de marzo de 2023.

B. Descripción general

El objetivo de esta tarea es experimentar con las arquitecturas de Flujo de Datos, a través del patrón ETL, y obtener una mejor comprensión de estas. El dominio de aplicación de esta tarea es el de sistemas de integración de información.

Las lecturas que debes realizar para esta tarea son:

- Data Pipeline Design Patterns (<https://www.eckerson.com/articles/data-pipeline-design-patterns>)
- C4 - context, containers, components and clases.pdf
- Patrón ETL.pdf

C. Forma de evaluación

Se evaluarán dos actividades:

- Diseño e implementación de la nueva versión del sistema: 5 puntos.
La implementación se debe hacer en equipo y se refiere a un mantenimiento evolutivo del sistema considerando un conjunto de requerimientos.

- Reporte individual: 5 puntos.

Cada integrante debe contestar un conjunto de preguntas sobre el sistema y entregarlas en un reporte de forma individual. Aunque las preguntas son sobre el sistema que han desarrollado en equipo, el reporte debe reflejar tu autonomía. La forma de redacción de las respuestas debe demostrar tus capacidades para analizar/resolver problemas, exponer argumentos sólidos y presentar puntos de vista basados en información específica.

Para la evaluación de estos rubros se tomará en cuenta:

- La consistencia entre el diseño y la implementación.
- El grado en que las soluciones respetan el patrón arquitectónico.
- El uso de principios de diseño, buenas prácticas de programación y documentación.
- El funcionamiento de los requerimientos implementados.
- La calidad de las respuestas a las preguntas.
- La calidad del reporte (ej. formato, ausencia de errores de gramática y ortografía, etc.).
- La entrega en tiempo y forma de los artefactos solicitados.

D. Descripción del sistema

Esto se explica en la lectura "Patrón ETL.pdf"

E. Puesta en marcha del sistema

Esto se explica en la lectura "Patrón ETL.pdf"

F. Preguntas y actividades a desarrollar

1. De acuerdo con el diagrama mostrado en la Figura 8.1 de la lectura "Patrón ETL.pdf" y lo descrito en la lectura "Data Pipeline Design Patterns", identifica los elementos del Sistema Janus que implementan elementos del patrón ETL. Para esto, puedes usar una tabla como esta.

| Nombre del elemento en el patrón arquitectónico | Responsabilidad | Nombre del elemento en en la figura 8.5 |
|---|-----------------|---|
| Fuente de datos | | |
| Extractor | | |
| Transformador | | |
| Cargador | | |
| Destino de datos | | |

2. Una decisión de diseño importante en la arquitectura de este sistema es el uso del Patrón ETL (el cual se implementó a través de la tecnología Luigi). Comenta cómo esta decisión de diseño contribuye a soportar el objetivo de negocio y los requerimientos de arquitectura. Para esto, puedes usar una tabla como esta agregando todos los requerimientos de arquitectura especificados para el sistema. Justifica tu respuesta indicando el por qué y el cómo son soportados.

| Requerimiento de arquitectura | | ¿Es soportado? Sí/No | ¿Cómo es soportado? | ¿Contribuye a lograr el objetivo de negocio? Sí/No |
|--|--|-------------------------|---------------------|---|
| Requerimientos funcionales de alto nivel | Como gerente de ventas necesito visualizar un reporte de ventas de un periodo específico para presentar el reporte financiero a los accionistas de la empresa. | | | |
| | Como gerente de almacén necesito visualizar un reporte de los productos más vendidos en un periodo específico para mantener un stock suficiente. | | | |
| | ... | | | |
| Requerimientos de atributos de calidad | Flexibilidad: el sistema deberá procesar distintos formatos de archivos y permitir la inclusión de nuevos... | | | |
| | Escalabilidad: el sistema | | | |

| | | | | |
|------------------------|--|--|--|--|
| | deberá soportar el procesamiento de datos de ventas mensuales de hasta 500 sucursales. | | | |
| | ... | | | |
| Restricciones técnicas | Se deberá utilizar Python como lenguaje de programación, ya que este es el que ... | | | |
| | ... | | | |

3. Aunque estos atributos de calidad no forman parte de los requerimientos de arquitectura del sistema; escribe la definición del atributo de calidad y comenta cómo son estos soportados en el sistema actual. Para esto, puedes usar una tabla como esta. Justifica tu respuesta indicando el por qué y el cómo son o no soportados.

| Atributo de calidad | Definición | ¿Es soportado? Sí/No | ¿Cómo es soportado? ¿Por qué no? |
|----------------------------|------------|-------------------------|----------------------------------|
| Integridad | | | |
| Interoperabilidad | | | |
| Disponibilidad | | | |
| Reutilización | | | |
| Facilidad de mantenimiento | | | |
| Facilidad de prueba | | | |

4. Considerando la estructuración actual del sistema, ¿a qué componentes se les tendrían que hacer cambios si se cambia el manejador de la base de datos? Especifica también qué cambios se deberían hacer. Puedes usar una tabla como esta para contestar esta pregunta.

| Nombre del componente | Responsabilidad | Qué cambios se deberían hacer |
|-----------------------|-----------------|-------------------------------|
| | | |
| | | |

5. Considerando la estructuración actual del sistema, y lo descrito en la lectura Data Pipeline Design Patterns, imagina que dentro de 1 año ya se han abierto más sucursales y que cada una de estas tiene un volumen muy alto de ventas y canales de venta. Por esta razón, los gerentes deben analizar más aspectos sobre la información. Por ello, han solicitado la implementación de nuevos reportes, sobre un subconjunto de sucursales y los patrones de compra de sus clientes por el canal de ventas on-line (ej. domicilio de esos clientes, monto de compras por mes, tipo de artículos comprados, mecanismos de pago, etc.) para ayudar a definir estrategias de marketing para los clientes de ese canal. Los requerimientos arquitecturales son:

Como gerente de ventas necesito visualizar un reporte de los patrones de compra de los clientes del canal online de una sucursal dentro de un periodo específico para poder definir estrategias de marketing para esos clientes.

Seguridad: la confidencialidad de los datos personales de los clientes debe ser asegurada.

Desempeño: los reportes deben generarse en no más de 3 segundos.

¿Qué cambios propondrías? ¿Cuáles serían los argumentos para las decisiones de diseño que tomarías para atender esos cambios? ¿Qué tradeoffs identificas las decisiones de diseño que tomarías?

6. Considerando la estructuración actual del sistema, lo descrito en la lectura Data Pipeline Design Patterns, y la

situación planteada en la pregunta anterior. Si alguien te sugiriera que debes modificar el sistema usando el Patrón Change Data Capture (CDC) ¿Qué pensarías de esto? Justifica tu respuesta indicando el por qué sí o el por qué sí no la sugerencia es correcta.

7. Consideren que la versión actual del sistema debe ser sujeta a un mantenimiento evolutivo considerando los siguientes requerimientos.

Requerimientos funcionales

- a) Como gerente de ventas necesito visualizar un reporte de ventas, para un periodo específico, para presentar el reporte financiero a los accionistas de la empresa.
- b) Como gerente de almacén necesito visualizar un reporte, para un periodo específico, de los productos más vendidos para mantener un stock suficiente.
- c) Como operador de logística necesito visualizar un reporte, para un periodo específico, de la cantidad de productos vendidos para ayudar a mantener un periodo de entrega de productos a los clientes no mayor a 40 días.
- d) Como CSO de la empresa necesito visualizar un reporte, para un periodo específico, de los indicadores de ventas de todas las sucursales.

Requerimientos de atributos de calidad

- a) Flexibilidad: el sistema deberá procesar datos de ventas en distintos formatos y permitir la inclusión de nuevos. Se debe incluir en el sistema datos de américa latina que fueron provistos en el archivo en formato .txt
8. Actualicen el diagrama de las Figuras 8.5, 8.6 y 8.7 para documentar la nueva versión del sistema. Incluirlos en este punto y describirlos brevemente.

9. Indica la url del repositorio de código desde donde podemos descargar la nueva versión del sistema.
10. Bonus: generen la documentación del sistema en backstage (<https://backstage.io/>). Indica la url de dicha documentación.