Evaluación del módulo 5

Consigna del proyecto



Evaluación del módulo

Proyecto: Arquitectura Datos

Situación inicial 📍



Unidad solicitante: Área de Analítica de Datos de una empresa de e-commerce y retail digital.

📌 La empresa ha experimentado un crecimiento acelerado y actualmente gestiona múltiples fuentes de datos (ventas online, comportamiento de usuarios, redes sociales, logística, inventario). La Dirección General solicita al equipo de datos el diseño de una arquitectura moderna y escalable que permita integrar, almacenar, gobernar y preparar estos datos para el análisis, garantizando su calidad y facilitando la toma de decisiones.

El desafío consiste en **construir una arquitectura robusta** que combine distintos niveles de almacenamiento (Data Lake, Data Warehouse, Data Mart), implemente estrategias de control de calidad y contemple un modelo multidimensional para el consumo analítico por parte de distintas áreas del negocio.

Nuestro objetivo 📋



El objetivo principal del proyecto es diseñar, documentar y justificar una arquitectura de datos integral, que contemple la integración de fuentes heterogéneas, el almacenamiento y gobierno de los datos, la aplicación de criterios de calidad y la estructuración de un modelo dimensional preparado para su análisis.

Este diseño debe ser escalable, seguro y alineado con las necesidades analíticas y operativas de la organización.

Requerimientos 🤝

Requerimientos generales del proyecto

La arquitectura de datos a diseñar deberá cumplir con los siguientes requerimientos generales para garantizar su funcionalidad, calidad y alineación con las necesidades del negocio:

Requerimientos generales:

- → Integración de múltiples fuentes de datos estructurados y no estructurados.
- → Aplicación de los principios y buenas prácticas de arquitectura de datos.
- → Incorporación de modelos de almacenamiento diferenciados: Data Lake, Data Warehouse, Data Mart.
- → Implementación de un plan de aseguramiento y monitoreo de la calidad de los datos.
- → Construcción de un modelo multidimensional alineado a un área clave del negocio.

Paso a paso 👣

Este proyecto refiere exclusivamente al **módulo 5**: Fundamentos de arquitectura y modelamiento de datos, y se compone de **4 etapas (lecciones)**, las cuales podrás avanzar de forma progresiva y escalonada con la ayuda de los manuales teóricos y los contenidos desarrollados en las clases en vivo.

Considera invertir **tiempo asincrónico** para el desarrollo de cada etapa a modo de poder finalizar el módulo y realizar la entrega formal de tu propuesta. Cualquier consulta que surja compártela en los espacios sincrónicos para resolver las dudas en equipo.

A continuación encontrarás las consignas y tareas a desarrollar:

- **Lección 1** Arquitectura de datos
 - **Objetivo:** Diseñar un esquema arquitectónico robusto para la integración, almacenamiento y consumo de datos.

Tareas a desarrollar:

- 1. Identificar **al menos tres fuentes de datos** estructuradas y no estructuradas relevantes para la organización.
- Diseñar y documentar una arquitectura de datos escalable y modular, basada en capas: ingesta, integración, almacenamiento, calidad y consumo.
- 3. **Incorporar principios** de gobierno, escalabilidad y flexibilidad.

- 4. Elaborar **un diagrama arquitectónico** que represente las fuentes, las capas y el flujo de datos.
- → **Nota:** Este diseño será la base para definir las zonas de almacenamiento y las estrategias de calidad en las siguientes lecciones.
- Lección 2 Enfoques para almacenamiento y gestión
 - **Objetivo:** Definir y justificar las estrategias de almacenamiento y gobernanza de los datos, alineadas al diseño arquitectónico.

📍 Tareas a desarrollar:

- 1. Analizar el **esquema arquitectónico diseñado** en la Lección 1.
- Definir las zonas de almacenamiento del Data Lake (raw, trusted, curated) y su relación con Data Warehouse y Data Mart.
- 3. **Documentar las tecnologías y servicios sugeridos** para cada zona.
- 4. Establecer **prácticas de gobernanza y gestión de datos** que aseguren trazabilidad, seguridad y disponibilidad.
- → Nota: Este entregable deberá estar alineado al flujo de datos y arquitectura diseñado en la Lección 1 y será la base para definir los controles de calidad en la siguiente etapa.
- Lección 3 Calidad de los datos
 - **Objetivo:** Diseñar un plan de aseguramiento de calidad de los datos, integrado a la arquitectura definida.

Tareas a desarrollar:

- 1. Revisar las **zonas y el flujo arquitectónico definidos** en la Lección 2.
- 2. Definir **los controles, métricas e indicadores de calidad** aplicables en cada etapa del flujo de datos.
- 3. Diseñar un proceso de monitoreo y remediación de la calidad.
- 4. **Documentar la integración del plan de calidad** en la arquitectura general.
- → **Nota**: Este entregable debe tomar como insumo la estructura de almacenamiento definida en la Lección 2 y dejar preparada la

información limpia para la construcción del modelo dimensional en la siguiente etapa.

• Lección 4 – Modelamiento multidimensional

Objetivo: Diseñar un modelo OLAP coherente con la arquitectura de datos y las estrategias de calidad implementadas.

Tareas a desarrollar:

- 1. Seleccionar **un área clave de negocio** y los datos disponibles a partir de las etapas anteriores.
- 2. Diseñar **un modelo dimensional** (estrella, copo de nieve o híbrido) alineado con la arquitectura y las zonas de almacenamiento definidas.
- 3. Elaborar **un diagrama del modelo** incluyendo tablas de hechos, dimensiones y jerarquías.
- 4. **Documentar las decisiones tomadas** en cuanto a normalización/desnormalización y criterios analíticos.
- → **Nota**: Este entregable consolidará el flujo arquitectónico, ya que consumirá los datos procesados, gobernados y asegurados en las etapas anteriores.

¿Qué vamos a validar?

Con el fin de garantizar la calidad y el éxito del proyecto **"Arquitectura de Datos"**, se evaluarán varios aspectos clave del producto. A continuación, se detallan los puntos que vamos a validar:

→ Aspectos técnicos:

- Coherencia en la selección de componentes arquitectónicos.
- Aplicación correcta de modelos de almacenamiento y calidad de datos.
- o Adecuación del modelo multidimensional al caso planteado.

Aspectos estructurales:

- Cumplimiento de los requerimientos generales y específicos.
- Claridad en la documentación y en los diagramas entregados.
- o Integración de los elementos en un flujo arquitectónico consistente.

→ Aspectos de performance:

- o Escalabilidad y seguridad de la arquitectura propuesta.
- o Pertinencia y aplicabilidad del modelo para el análisis de datos.

Referencias 4

Durante la construcción del proyecto "**Arquitectura de Datos**", se pueden consultar diferentes recursos y referencias que faciliten la resolución de las consignas y contribuyan a la calidad del trabajo realizado. A continuación, se enumeran algunos ejemplos de contenido referencial útiles:

- 1. Diseño de arquitectura de bases de datos
- 2. Analytics on AWS
- 3. Repositorios visuales: draw.io, diagrams.net, Lucidchart
- 4. Ejemplos de arquitectura moderna: towardsdatascience.com

Recursos 🎁

- → https://cloud.google.com/solutions/big-data
- → https://docs.aws.amazon.com/wellarchitected/latest/analytics-lens/analyt
- → https://learn.microsoft.com/en-us/azure/well-architected/

Entregables 🔽

Al finalizar el proyecto "**Arquitectura de Datos para Analytics Integrado**", se espera que el equipo de desarrollo entregue los siguientes elementos como evidencia concreta del trabajo realizado:

Lección 1 – Arquitectura de Datos

- **Informe técnico** que incluya la descripción de las fuentes de datos identificadas y el diseño arquitectónico basado en capas.
- **Diagrama arquitectónico** detallado con el flujo de datos y componentes principales.
- Justificación técnica de las decisiones tomadas en el diseño.



Lección 2 - Enfoques para el Almacenamiento y Gestión de los Datos

- **Documento técnico** detallando las zonas de almacenamiento (raw, trusted, curated) y su función.
- Esquema o diagrama complementario que muestre la segmentación de datos y flujos.
- **Justificación técnica** del uso de Data Lake, Data Warehouse y Data Mart según los casos de uso definidos.

Lección 3 – Calidad de los Datos

- Informe técnico con el plan de calidad de datos, incluyendo métricas, indicadores y técnicas aplicadas.
- Esquema gráfico del proceso de control y remediación de calidad de datos.
- **Justificación** de la integración de la calidad de datos en la arquitectura general.

Lección 4 - Modelamiento Multidimensional

- Documento técnico describiendo el modelo dimensional (OLAP) diseñado.
- **Diagrama del modelo dimensional** que incluya tablas de hechos, dimensiones y jerarquías.
- **Justificación técnica** de las decisiones de modelado y su integración con la arquitectura propuesta.

Entrega Final - Proyecto Integrador

- **Documento integrador consolidado** con todos los informes y diagramas elaborados en cada lección.
- **Diagrama final integrador** que incluya arquitectura, almacenamiento, calidad de datos y modelo dimensional.
- **Presentación breve** en formato PDF o PPT resumiendo la propuesta arquitectónica.
- Link a repositorio online (GitHub, Drive u otro) que contenga todos los archivos del proyecto: informes, diagramas y esquemas.

Portafolio

Este proyecto podrá ser incluido en tu portafolio profesional como evidencia concreta de tus habilidades en diseño de arquitecturas de datos. Te sugerimos presentar:



- El diagrama completo de la arquitectura de datos.
 La estrategia de almacenamiento y calidad de datos aplicada.
- El modelo dimensional documentado.
- La documentación técnica consolidada con las decisiones tomadas.
- Consideraciones de **escalabilidad y mantenimiento**.

También es recomendable agregar una breve reflexión personal sobre los desafíos enfrentados y los aprendizajes obtenidos durante el desarrollo del proyecto.

Esta evidencia mostrará tu capacidad para integrar conceptos de arquitectura, gobernanza, calidad y análisis de datos en una solución real.

¡Muchas gracias!

Nos vemos en la próxima lección

