
25GIIN – Diseño y Desarrollo de Sistemas de Información

Actividad 4 – Portafolio

Gagliardo Miguel Angel – **Grupo 6**

21 de Diciembre de 2024

1. Introducción

- **Explicar la filosofía data-driven.**

La filosofía **data-driven** es un paradigma en el que el aprendizaje se produce mediante la acumulación de datos y la aplicación de métodos para extraer **patrones significativos** de los mismos. **Bajo este paradigma, los datos se convierten en el activo más valioso de una organización.**

En el marco de la investigación basada en datos, se espera que los investigadores utilicen los mismos como punto de partida para la inferencia inductiva, sin depender de preconcepciones teóricas. En el marco de una empresa, **la recopilación y el procesamiento de datos son los que respaldan las decisiones operativas y estratégicas de la organización.**

- **Discutir la importancia creciente de los datos.**

La importancia de los datos ha crecido exponencialmente con la digitalización de casi todos los aspectos de la vida humana y empresarial.

En el entorno empresarial, los datos son **un recurso estratégico** de inmenso valor, ya que las organizaciones que logran capturar, procesar y analizar estos datos tienen una ventaja significativa en términos de competitividad, eficiencia y adaptabilidad. Desde la mejora de procesos internos hasta el diseño de productos más ajustados a las necesidades de los consumidores.

La **filosofía data-driven** es indispensable en este contexto, ya que proporciona un marco claro para aprovechar al máximo estos recursos de datos.

2. Diseño y Modelado de Datos en un Contexto Data-Driven

- **Requerimientos de Datos:** Explica cómo una filosofía data-driven exige una reestructuración en la forma en que las empresas capturan, almacenan y organizan los datos. ¿Qué tecnologías y enfoques son necesarios para manejar grandes volúmenes de datos (Big Data)? ¿Cómo influye esto en el diseño y modelado de las bases de datos, especialmente en términos de normalización y calidad de los datos?

La filosofía data-driven implica una **profunda transformación cultural** en las empresas. Se trata de abrazar a los datos como medio para tomar decisiones informadas y sustentadas. Esto significa que la cultura data-driven se caracteriza por el uso de tecnologías avanzadas para buscar insights, **identificar tendencias y ahorrar tiempo y recursos**.

Algunos de los **desafíos** en un enfoque data-driven incluyen:

- **Escalabilidad:** Dado que los volúmenes de datos crecen de manera exponencial, las soluciones de almacenamiento deben ser capaces de escalar de manera eficiente para manejar datos en tiempo real y en grandes cantidades.

- **Flexibilidad:** Los datos no siempre son estructurados, por lo que es necesario usar tecnologías que permitan almacenar tanto datos estructurados como no estructurados.
- **Calidad de los Datos:** Los datos deben ser precisos, consistentes, completos y oportunos. Los sistemas deben contar con mecanismos para validar la calidad de los datos a lo largo del ciclo de vida.
- **Silos de Datos:** En muchas organizaciones, los diferentes departamentos (como ventas, marketing, operaciones, finanzas, etc.) pueden tener sus propias bases de datos y sistemas, lo que genera silos de información. Esto dificulta la colaboración entre áreas y la toma de decisiones basadas en datos integrados.

Algunas de las **herramientas** utilizadas a la hora de manejar grandes volúmenes de datos son:

- **Data Analytics (análisis de datos):** Se refiere al proceso de inspeccionar, “limpiar” y modelar datos con el objetivo de descubrir información útil, concluir y respaldar decisiones empresariales. A través de técnicas estadísticas y herramientas computacionales, Data Analytics ayuda a las empresas a identificar patrones, tendencias y relaciones en los datos, lo que permite tomar decisiones informadas. Es comúnmente utilizado para la optimización de procesos, la gestión de riesgos o el análisis del comportamiento del cliente.

Un **ejemplo de uso** de data analytics podría ser un comercio minorista podría analizar los datos de ventas para identificar productos más populares y así ajustar su inventario.

- **Data Science (ciencia de datos):** Es un campo interdisciplinario que utiliza métodos, algoritmos y sistemas científicos para extraer conocimiento y predicciones de grandes volúmenes de datos. Data Science involucra áreas como la **estadística, la computación y la programación**, y se centra en la creación de **modelos predictivos** y el **análisis de datos complejos**. Es esencial para descubrir patrones ocultos o para construir modelos predictivos que ayuden a anticipar comportamientos o resultados futuros.

Un **ejemplo de uso** de data science sería cómo una empresa de telecomunicaciones puede usar ciencia de datos para predecir la rotación de clientes (en la industria esto se conoce como “**churn**”) y desarrollar estrategias para retenerlos.

- **Data Mining (minería de datos):** Es el proceso de descubrir **patrones, correlaciones y anomalías en grandes conjuntos de datos** mediante el uso de **técnicas estadísticas** y de **machine learning**. La minería de datos se centra en explorar y extraer información oculta en los datos, lo cual puede resultar en descubrimientos significativos o nuevas oportunidades de negocio. A menudo se utiliza en el análisis de grandes bases de datos para encontrar patrones ocultos que podrían ser relevantes para la toma de decisiones estratégicas.

Un **ejemplo** de data mining podría ser un banco que utiliza esta técnica para identificar patrones fraudulentos en las transacciones de tarjetas de crédito.

- **Data Visualization (visualización de datos):** Se refiere a la **representación gráfica de los datos para facilitar su comprensión y análisis**. Al transformar datos complejos en gráficos, mapas o diagramas, se facilita la identificación de **patrones, tendencias y relaciones** que de otra forma podrían ser difíciles de detectar. Las herramientas de Data Visualization permiten a los usuarios presentar la información de manera clara y accesible, lo que es esencial para la comunicación efectiva de insights dentro de la organización.
- **Integración de Fuentes de Datos:** Discute cómo los sistemas de información deben integrar datos de diferentes áreas y fuentes, y cómo esto impacta en la estructura de la base de datos y las herramientas utilizadas.

Bajo un sistema data-driven los datos provienen de sistemas operacionales, bases de datos internas, aplicaciones de terceros, redes sociales, sensores IoT (Internet de las Cosas), entre otros. La integración eficiente de estos datos es esencial para obtener una visión coherente que permita a la organización tomar decisiones basadas en información unificada.

La **integración de datos** implica diversas etapas para **reunir, combinar y armonizar información de fuentes dispares**. Estas son algunas de las prácticas y tecnologías más comunes utilizadas para integrar datos:

- **ETL (Extracción, Transformación y Carga):**
 - **Extracción:** Los datos se extraen de diversas fuentes, como bases de datos, APIs, archivos planos, o aplicaciones empresariales.
 - **Transformación:** En este paso se procesan y limpian los datos, transformándolos a un formato común o unificado, asegurando su consistencia y calidad. Esto puede incluir operaciones como la normalización de valores, conversión de tipos de datos, eliminación de duplicados y manejo de datos faltantes.
 - **Carga:** Finalmente, los datos transformados se cargan en un almacén de datos centralizado, como un data warehouse o una base de datos en la nube.
- **Data Warehousing:** Un **almacén de datos** es una base de datos optimizada para consultas y análisis, diseñada para integrar datos de diferentes fuentes. Herramientas como **Amazon Redshift** o **Google BigQuery** son ejemplos de plataformas modernas de data warehousing que soportan grandes volúmenes de datos.
- **Middleware y APIs:** El uso de **middleware** (software que actúa como intermediario entre aplicaciones o sistemas) y **APIs** permite integrar datos de manera más ágil y en tiempo real. Las APIs proporcionan una interfaz para la conexión entre sistemas dispares y facilitan la integración continua de datos de diversas aplicaciones o servicios externos.
- **Data Lake:** Un **data lake** es un **repositorio centralizado que permite almacenar grandes volúmenes de datos en su formato original**, sin necesidad de estructurarlos previamente. A diferencia de los data warehouses, que son más estructurados, los data

lake pueden manejar tanto datos estructurados como no estructurados (por ejemplo, audio, imágenes, vídeos, logs de eventos). Herramientas como **Apache Hadoop** y **Azure Data Lake** son populares para almacenar y procesar grandes volúmenes de datos no estructurados.

3. Arquitectura de Sistemas de Información Data-Driven

- **Requerimientos de Infraestructura:** Describe cómo una filosofía data-driven impacta la arquitectura de los sistemas de información, especialmente en términos de escalabilidad y almacenamiento. ¿Qué soluciones tecnológicas (como bases de datos NoSQL, almacenamiento en la nube, etc.) son necesarias para soportar grandes volúmenes de datos?

La filosofía data-driven impone grandes exigencias sobre la **infraestructura de los sistemas de información**. Como antes mencionábamos, se necesita un sistema que sea **escalable** y capaz de manejar **grandes cantidades de datos** de manera **eficiente**. Esto incluye no solo almacenamiento de datos, sino también la capacidad de procesar grandes volúmenes de datos en tiempo real para ofrecer insights inmediatos.

Entre las tecnologías que permiten esta infraestructura destacan (algunas ya mencionadas):

- **Bases de datos NoSQL:** Como se mencionó anteriormente, **las bases de datos NoSQL son clave en un enfoque data-driven** debido a su capacidad para manejar grandes cantidades de datos no estructurados y semi-estructurados.

- **Almacenamiento en la Nube:** Soluciones como **S3** de Amazon Web Services (AWS), **Cloud Storage** de Google Cloud Platform (GCP) y **Blob Storage** de Microsoft Azure permiten la escalabilidad de almacenamiento y procesamiento de datos sin necesidad de infraestructura física.
- **Plataformas de Big Data:** Tecnologías como **Apache Hadoop**, **Apache Spark** y **Apache Kafka** permiten procesar grandes volúmenes de datos distribuidos de manera rápida y efectiva, haciendo frente al desafío de la escalabilidad horizontal.

- **Optimización del Rendimiento:** Explica cómo los sistemas de información deben ser diseñados para permitir consultas rápidas y el procesamiento en tiempo real de datos, y cómo esto afecta la arquitectura técnica y las decisiones de implementación.

Para **optimizar el rendimiento de los sistemas de información data-driven**, se deben considerar aspectos como:

- **Procesamiento en Tiempo Real:** Se requieren arquitecturas basadas en streaming y procesamiento en tiempo real (herramientas como Apache Kafka o Apache Flink) para permitir consultas y análisis de datos instantáneos.
- **Consultas rápidas:** Para soportar consultas rápidas en grandes volúmenes de datos, las arquitecturas deben contar con tecnologías de caching (Redis, Memcached), índices eficientes y optimización en el diseño de consultas.

4. Impacto en la Implementación de Sistemas de Información

- **Fases de Implementación:** Describe cómo la filosofía data-driven cambia la forma en que se implementan los sistemas de información. Además de la construcción del software, ¿qué pasos adicionales se deben tomar para integrar los datos de diversas fuentes y garantizar que el sistema sea efectivo en el soporte de la toma de decisiones?

La implementación de sistemas de información data-driven **requiere un enfoque más iterativo y basado en datos**. Además de la construcción del software, deben tomarse pasos adicionales como:

- **Definición de Objetivos Basados en Datos:** En un enfoque data-driven, la implementación comienza con la identificación de objetivos estratégicos que se logran mediante el uso de datos. Esto asegura que **cada paso esté alineado** con la mejora de la toma de decisiones.
- **Recopilación e Integración de Datos:** Es fundamental integrar datos de diversas fuentes (internas y externas). Los datos deben ser limpiados y transformados para asegurar que sean consistentes y útiles para los análisis.
- **Desarrollo de Infraestructura:** Además de construir el software, se necesita una infraestructura que permita almacenar y procesar grandes volúmenes de datos, utilizando herramientas como **bases de datos en la nube y plataformas de análisis**.

- **Desarrollo Analítico:** Se crean modelos y herramientas para analizar los datos y generar insights útiles. Esto permite a los usuarios visualizar y entender mejor la información para tomar decisiones informadas.
- **Capacitación y Cultura Data-Driven:** Los empleados deberán ser capacitados para **usar los datos de manera efectiva en su trabajo**, promoviendo una cultura organizacional centrada en las **decisiones basadas en datos**.
- **Monitoreo y Ajustes Continuos:** **El sistema debe ser monitoreado constantemente y ajustado** según sea necesario para mejorar la calidad de los datos y la efectividad en la toma de decisiones.

5. Metodologías Ágiles y Desarrollo Iterativo en el Marco Data-Driven

- **Enfoque Basado en Evidencia:** Analiza cómo la filosofía data-driven puede favorecer el uso de metodologías ágiles, como Scrum o Kanban, en el desarrollo de sistemas. Explica cómo los datos reales, obtenidos durante el uso de los sistemas, pueden influir en la toma de decisiones para mejorar continuamente el sistema.

La filosofía data-driven favorece el uso de metodologías ágiles ya que las mismas permiten la flexibilidad de **adaptar el desarrollo a medida que se obtienen datos reales del uso del sistema**. La **retroalimentación** continua de los usuarios y la medición del rendimiento a través

de datos influyen directamente en la priorización de tareas y características a implementar en los sprints siguientes.

Por **ejemplo**, dado que en cada sprint se decide que tareas se van a priorizar en base a los datos correspondientes al equipo, básicamente se retroalimentan las decisiones del mismo y se facilita la priorización de tareas, proyectos y decisiones.

- **Priorización de Características:** Discute cómo las organizaciones pueden priorizar características del sistema basadas en datos reales de usuarios y métricas de rendimiento, en lugar de basarse únicamente en los requisitos iniciales o intuiciones.

La priorización de características en un enfoque data-driven se basa en el **análisis de datos reales**, lo que permite a las organizaciones **tomar decisiones más informadas y ajustadas a las necesidades del usuario y el rendimiento del sistema**. En lugar de depender únicamente de los requisitos iniciales o de la intuición de los stakeholders, las organizaciones pueden utilizar **métricas de uso, feedback de usuarios y análisis de comportamiento** para identificar qué funcionalidades son realmente valiosas.

Por **ejemplo**, si se observa que una característica del sistema es utilizada de manera frecuente por los usuarios, o si los datos muestran que una funcionalidad está impactando positivamente en el rendimiento o en la satisfacción del cliente, esta debería tener una mayor prioridad en el desarrollo o mejora del sistema. Además, los datos pueden revelar áreas de mejora que no se habían identificado previamente, permitiendo a las organizaciones enfocarse en lo que realmente importa para los usuarios y el negocio.

6. Conclusiones

La adopción de una filosofía data-driven transforma la manera en que las organizaciones desarrollan e implementan sistemas de información. Si bien presenta desafíos, como la necesidad de una infraestructura robusta y la integración de múltiples fuentes de datos, también ofrece grandes oportunidades en términos de eficiencia operativa y capacidad para tomar decisiones más informadas y precisas.

Los principales retos incluyen la gestión de la calidad de los datos y la integración de sistemas heterogéneos, pero estos pueden superarse mediante el uso de tecnologías adecuadas y metodologías ágiles. Las empresas deben fomentar una cultura de dato-centrismo, donde todos los niveles de la organización valoren la toma de decisiones basada en datos y estén comprometidos con el proceso de mejora continua.

Finalmente, tal vez algunos de los mejores **ejemplos de empresas (“Big Tech”) conocidas por ser data-driven sean:**

- **Facebook (META):** Tal vez una de las más icónicas de este siglo, Facebook utiliza una cantidad abismal de datos para dirigir “ads” (publicidad) específicos a cada uno de sus clientes. Sólo pensar la inmensa cantidad de datos a través de sus plataformas más conocidas (Facebook, Instagram, Whatsapp), los millones de usuarios que cada una de estas maneja y como en cada una de ellas ofrece publicidad específica a sus gustos o necesidades, nos da a pensar el nivel de procesamiento, cómputo y desarrollo que la empresa posee, así como su estrategia al ofrecer **productos “data-driven”**.

- **Amazon:** La empresa tiene como **cultura interna** el ser “data-driven” a tal punto que **cada decisión debe ser o bien respaldada o bien contrastada con datos**. Esto nos muestra como la compañía que se jacta de ser la más “customer centric” (enfocada en el cliente) del mundo no toma decisiones a la ligera. Desde el board hasta los individuos de cada equipo, todas las decisiones se basan y debaten **siempre con un fuerte sesgo en los datos**.

7. Bibliografía

- [1] Stanford University Encyclopedia of Philosophy. 2020. **Scientific Research and Big Data** – <https://plato.stanford.edu/entries/science-big-data/>
- [2] Mayer-Schönberger, Victor and Kenneth Cukier, 2013, **Big Data: A Revolution that Will Transform How We Live, Work, and Think** – https://youtu.be/bYS_4CWu3y8
- [3] Khaled Gubran Al-Hashedi, Pritheega Magalingam. **Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019** – <https://doi.org/10.1016/j.cosrev.2021.100402>
- [4] He, Y., & Liu, Y. (2020). **Data-driven Decision Making and Business Model Innovation in the Digital Economy**
- [5] Facebook Research. **Data warehousing and analytics infrastructure at Facebook (2019)**. <https://research.facebook.com/publications/data-warehousing-and-analytics-infrastructure-at-facebook/>
- [6] Forbes. **Predicting The Future Of Demand: How Amazon Is Reinventing Forecasting With Machine Learning** (2022) –

<https://www.forbes.com/sites/amazonwebservices/2021/12/03/predicting-the-future-of-demand-how-amazon-is-reinventing-forecasting-with-machine-learning/>

[7] MarketingScopp. **Uncovering Amazon's Big Data Secrets: How the E-Commerce Giant Leverages Data to Dominate Retail** (2024) - <https://www.marketingscoop.com/consumer/how-does-amazon-use-big-data/>