**Guía 3. Informe final Proyecto APT**

**Asignatura Capstone**

|  |
| --- |
| **1. Informe final Proyecto APT** |
| El objetivo de este informe es que describas los aspectos más relevantes de tu Proyecto APT. Es importante que fundamentes las decisiones que tuviste que tomar a lo largo del proceso.  A continuación, encontrarás distintos campos que deberás completar con la información solicitada, los que dan cuenta del resumen de tu proyecto APT y sus principales resultados. |

|  |  |
| --- | --- |
| Nombre del proyecto | *Implementación de Data Warehouse en sistema ERP* |
| Área (s) de desempeño(s) | *Big Data, Inteligencia de negocios y Machine Learning.* |
| Competencias | * *Diseño y modelado de bases de datos para estructurar el Data Warehouse.* * *Integración de sistemas aseguran un flujo eficiente y seguro de la información.* * *Procesos ETL para extraer, transformar y cargar datos* * *Dominio herramientas de BI como Looker Studio para la creación de dashboards e informes* * *Gestión de proyectos con metodología ágil y Kimball* * *Seguridad de datos* * *Entrenamiento de modelos de ML (regresión lineal, regresión logística, random forest, boosted trees)* |

|  |  |
| --- | --- |
| **Contenidos del informe final** | |
| 1. Relevancia del proyecto APT | *El proyecto busca optimizar la toma de decisiones empresariales mediante la creación de un Data Warehouse eficiente y la implementación de modelos de Machine Learning (ML), orientados a transformar datos en información accesible y útil. Esto es especialmente relevante en el contexto de la transformación digital, ya que permite a las empresas, particularmente en un entorno impulsado por datos como el actual, obtener una ventaja competitiva. La situación se ubica en la empresa HVR, con sede en Providencia, Santiago de Chile, y afecta tanto a operarios como a la gerencia de clientes de diferentes rubros que utilizan el sistema HVR en todo el país. El valor aportado por este proyecto consiste en una mejora real en el rendimiento y accesibilidad de los datos, posibilitando informes y visualizaciones claras que permiten tomar decisiones informadas y estratégicas. En el contexto de la profesión, el proyecto ilustra cómo la integración de un Data Warehouse y el uso de ML contribuyen directamente a la optimización del rendimiento empresarial y fortalecen las capacidades de análisis en un mercado altamente competitivo.* |
| 2. Objetivos | * *Objetivo General:*   + *Diseñar y desarrollar un Data Warehouse alojado en GCP que optimice la carga diaria de información, entrenar modelos de ML para obtención de valores predictivos y la posterior generación de informes y dashboards* * *Objetivos específicos:*    + *Implementar la integración de datos desde el ERP hacia GCP*   + *Diseñar un pipeline para la limpieza y transformación de datos*   + *Diseñar e implementar un Data Warehouse en GCP*   + *Entrenar modelos ML (regresión lineal, regresión logística, random forest, boosted trees) para proyección de datos. Se escogerá el modelo que mejor se adapte a las necesidades del negocio.*   + *Establecer mecanismos de seguridad de los datos*   + *Crear informes y dashboard transversales para el sistema.* |
| 3. Metodología | *Se utilizará una metodología combinada entre un desarrollo ágil para entregas iterativas y una metodología Kimball para la creación del Data Warehouse. Según esto se desarrollará de la siguiente forma:  1. Planificación y Definición del Proyecto*   * *Objetivo: Alinear los requerimientos con el diseño de datos, asegurando un inicio claro y estructurado.* * *Actividades: Definición de requerimientos, alcance y diseño del modelo de datos basados en historias de usuario.*   *2. Desarrollo e Implementación*   * *Objetivo: Crear entregas incrementales y optimizadas, iterando sobre modelos y datos.* * *Actividades: Planificación de Sprints, desarrollo de modelos de estrellas, entrenamiento de ML y dailies.*   *3. Integración y Pruebas*   * *Objetivo: Garantizar la calidad del sistema y precisión en los datos, asegurando su funcionamiento completo.* * *Actividades: Pruebas unitarias e integración para ajustar componentes y revisar cumplimiento de requerimientos.*   *4. Capacitación y Despliegue*   * *Objetivo: Asegurar una transición fluida al usuario final, promoviendo un uso efectivo del sistema.* * *Actividades: Creación de materiales de capacitación y entrenamiento a usuarios para el uso del sistema.* |
| 4. Desarrollo | ***Descripción de las etapas o actividades del Proyecto APT:***  *El desarrollo del proyecto Incluye las siguientes etapas:*   * ***Configuración inicial del entorno:*** *Instalamos y configuramos la infraestructura en GCP, asegurando los permisos necesarios para el funcionamiento del sistema.* * ***Desarrollo de procesos ETL:*** *Implementamos la extracción de datos desde el ERP hacia la capa de almacenamiento inicial (capa bronze) en GCP y configuramos las transformaciones hacia la capa silver.* * ***Diseño del Data Warehouse:*** *Modelamos el esquema estrella, incluyendo tablas de hechos y dimensiones, optimizadas para consultas rápidas.* * ***Entrenamiento de modelos ML:*** *Comenzamos el entrenamiento de modelos predictivos utilizando los datos transformados.* * ***Desarrollo de dashboards:*** *Se preparan dashboards iniciales en Looker Studio para visualizar los datos procesados según objetivo de negocio.*   ***Dificultades y facilitadores en el desarrollo del Proyecto APT:***  ***Facilitadores:***   * *Contar con el apoyo logístico de la empresa patrocinadora para acceder a los datos y definir los requerimientos del sistema.* * *La documentación oficial de GCP fue clave para resolver dudas técnicas durante la configuración.* * *El trabajo en equipo permitió avanzar de manera organizada y eficiente en las etapas del proyecto.*   ***Dificultades:***   * *Problemas iniciales con la configuración de permisos en GCP que retrasaron el avance del proyecto.* * *La limpieza y calidad de los datos resultaron ser más complejas de lo esperado debido a inconsistencias en las fuentes.* * *Los tiempos de respuesta de los stakeholders fueron más lentos de lo previsto, afectando algunas decisiones clave.*   ***Ajustes realizados:*** *Para abordar estas dificultades:*   * *Reasignamos tiempos en el cronograma para resolver problemas técnicos con GCP.* * *Diseñamos procesos adicionales de validación y limpieza de datos para garantizar su calidad.* |
| 5. Evidencias | * *Adjunta evidencias que permitan dar cuenta del desarrollo del Proyecto APT y sus resultados finales.*   *¿Qué evidencias pueden servir para que los demás puedan visualizar y entender las distintas etapas de tu Proyecto APT y el resultado final?*   * [**Documento de Historias de usuario**](https://docs.google.com/spreadsheets/d/1mYHkbl4eTBivwpUys9payER6KD-f7_ddYFtaPFfWhqs/edit?usp=sharing) * [**Asignación de roles y perfiles**](https://docs.google.com/document/d/1HvY40efkcte_pLuLR5d9iRGqWi3GJ5brW7jsAzGnO3I/edit?usp=sharing) * [**Evidencias de sistema de aplicacion**](https://docs.google.com/document/d/1T-jyHQbKZODOXiUrCQaXj0rKAowscSWTbae8vnNI7sw/edit?usp=sharing) * [**Diagrama de Pipeline en Modelo medallón**](https://drive.google.com/file/d/1owNL1TPppAsBEOwc24iM3mHkYH0G7gMp/view?usp=sharing) |
| 6. Intereses y proyecciones profesionales | *El Proyecto APT ha sido fundamental para reforzar y ampliar nuestro interés profesional en áreas relacionadas con el manejo de datos, infraestructura en la nube y machine learning. Gracias a este proyecto, logramos aplicar de manera práctica conocimientos adquiridos en la carrera, como la integración de datos, el diseño de un Data Warehouse y el desarrollo de modelos predictivos. Esto nos permitió entender mejor cómo estas tecnologías pueden generar valor en entornos empresariales reales. Esperamos que al finalizar el proyecto, nuestros intereses profesionales se mantengan alineados con los planteados inicialmente, pero con una visión más clara y concreta de su aplicación práctica en el campo laboral.*   * *Proyecciones laborales a partir del Proyecto APT:*   *Nos gustaría seguir profundizando en áreas como la analítica avanzada, el machine learning y la inteligencia de negocios, particularmente en la optimización de procesos empresariales mediante el uso de datos. Después de finalizar este proyecto, nos proyectamos como profesionales capaces de liderar iniciativas en transformación digital, integrando soluciones en la nube y analítica avanzada para tomar decisiones estratégicas basadas en datos.* |