}

**Guía1. Definición Proyecto APT**

**Asignatura Capstone**

1. **PARTE I**

|  |
| --- |
| **1. Antecedentes Personales** |
| A continuación, se presenta una tabla en la que debes completar la información solicitada. |

|  |  |
| --- | --- |
| Nombre estudiantes | **Kiara Caycho Daniela Gálvez Sebastian Orellana Ignacio PI** |
| Rut | **18862994-6** |
| Carrera | **Ing Informática** |
| Sede | **Antonio Varas** |

|  |
| --- |
| **2. Descripción Proyecto APT** |
| En la descripción debes señalar brevemente el nombre de tu proyecto APT y las competencias del perfil de egreso que vas a poner en práctica. Si en tu carrera están definidas las áreas de desempeño, también menciona a qué áreas de desempeño está vinculado el proyecto. |

|  |  |
| --- | --- |
| 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)* |

|  |
| --- |
| **3. Fundamentación Proyecto APT** |
| A continuación, se presentan distintos campos que debes completar con la información solicitada. Esta sección busca que describas en detalle tu proyecto y justifiques su relevancia y pertinencia. |

|  |  |
| --- | --- |
| Relevancia del proyecto APT | *Todo proyecto, ya sea una innovación, producto, servicio, etc., pretende dar respuesta a una situación o problemática. Señala qué problema busca solucionar tu proyecto y la relevancia que tiene para el campo laboral de tu carrera. También menciona el contexto en que esta problemática se sitúa (lugar, a quienes impactaría, etc.). Es importante que esta problemática sea relevante en el contexto de la profesión, siendo su resolución un aporte real o simulado a la organización u entorno en el que se sitúa. Algunas preguntas que pueden ayudarte a responder este apartado son:*   * *¿Por qué escogiste este tema?*   *Porque es una problemática real que enfrenta la empresa actual y que nos permitiría poder enfrentar una contingencia actual propia de nuestra mención de carrera utilizando tecnologías y plataformas vigentes.*   * *¿Por qué es relevante este tema para el campo laboral de tu carrera?*   *Porque permite dar un paso en la trasformación digital y hacer uso de los datos y hacer de ellos información accesible para la toma de decisiones, optimizando así el rendimiento empresarial y ofreciendo una ventaja competitiva en un entorno impulsado por datos y herramientas modernas como modelos de ML.*   * *¿Dónde se ubica la situación que vas a abordar?*   *El proyecto será dirigido para HVR con oficinas en Chile, RM, Providencia.*   * *¿Cuáles son las características principales de ese lugar?*   *Es una empresa que ofrece un sistema de ERP bajo las normas y legislaciones actuales de Chile.*   * *¿A quiénes afecta o impacta la situación que vas a abordar? (Ej.: Grupo etario, usuarios de algún servicio, etc.).*   *Afecta a los clientes de diversos rubros y servicios que operan con el sistema HVR a lo largo de Chile, con un enfoque tanto a operarios como a gerencia.*   * *¿Cuál sería el aporte de valor (real o simulado) de tu Proyecto APT para el contexto laboral y/o social en que se situaría?*   *Mejorar la toma de decisiones empresariales mediante la creación de un Data Warehouse optimizado y proyección de datos con ML, lo que permite generar informes y visualizaciones claras y precisas, facilitando el acceso a información crítica.* |
| Descripción del Proyecto APT | *Crear un Data Warehouse alojado en GCP con la carga de información de forma diaria, además de la creación de modelos de ML. Optimizando la creación de informes, y dashboard con información a nivel de usuarios y otra para gerencia.* |
| Pertinencia del proyecto con el perfil de egreso | *Justifica cómo se relaciona tu Proyecto APT con el perfil de egreso de tu carrera y, en particular, con las competencias del perfil de egreso que seleccionaste anteriormente.*  *¿De qué manera se relaciona el Proyecto APT con el perfil de egreso de tu carrera? ¿De qué manera son necesarias las competencias que seleccionaste para resolver la problemática a trabajar?*   * *Diseño y Modelado de Bases de Datos: Fundamental para estructurar el Data Warehouse, asegurando que los datos sean organizados de manera eficiente para su posterior análisis.* * *Procesos ETL (Extract, Transform, Load): Crucial para extraer datos del ERP, transformarlos y cargarlos en el Data Warehouse alojado en GCP, garantizando la calidad y consistencia de la información.* * *Herramientas de BI y Visualización de Datos: Esenciales para la creación de informes y dashboards que presenten la información de manera clara y accesible, adaptada tanto para usuarios finales como para la gerencia.* * *Entrenamiento de modelos ML: Crea valor para obtener valores predictivos que optimizan la toma de decisiones.* * *Seguridad de Datos y Cumplimiento Normativo: Importante para proteger la integridad y confidencialidad de los datos manejados en el Data Warehouse, cumpliendo con las normativas correspondientes.* * *Gestión de Proyectos Ágiles: Necesaria para coordinar el desarrollo del proyecto, asegurando que se cumplan los plazos y se entreguen resultados de valor incremental.* * *Implementación de modelos dimensionales y cubos OLAP siguiendo las prácticas de* ***Ralph Kimball****.* |
| Relación con los intereses profesionales | *Señala cómo se relaciona el Proyecto APT que propones con tus intereses profesionales.*  *¿Cuáles son tus intereses profesionales? ¿Qué aspectos de tus intereses profesionales se ven reflejados en tu Proyecto APT? Realizar este Proyecto APT, ¿de qué manera va a contribuir a tu desarrollo profesional?*  *El proyecto se alinea directamente con nuestra mención de la carrera y con el interés que tenemos en nuestro desarrollo profesional. Ya que podremos desarrollar aspectos como el manejo de datos en infraestructura Cloud, el análisis de ella, utilizar los datos obtenidos para el entrenamiento de modelos de ML y la posterior entrega de información crítica de forma inteligente. Este proyecto nos contribuirá ya que pondremos en práctica las metodologías, técnicas y herramientas aprendidas en nuestra mención e interés profesional.* |
| Factibilidad de desarrollo del Proyecto APT | *Justifica brevemente por qué es posible desarrollar tu proyecto APT. Considera el tiempo y materiales que necesitas para desarrollarlo, así como los posibles factores externos que podrían dificultar y facilitar su desarrollo.*  *¿Por qué crees es posible desarrollar tu Proyecto APT? Para responder esta pregunta debes tener en consideración:*   1. *Duración del semestre*   *Desde el 13 de agosto hasta mediados de Diciembre.*   1. *Horas asignadas a la asignatura*   *3 horas semanales presenciales.*   1. *Materiales requeridos*   *Acceso a la BD de la empresa, Créditos de GCP*   1. *Factores externos que facilitan su desarrollo*   *Apoyo de empresa promotora, tanto logístico como con los datos.*   1. *Factores externos que dificultan su desarrollo y maneras en que podrías solucionarlos*   *Tiempos de respuesta de empresa patrocinadora y cantidad de créditos disponibles.* |

1. **PARTE II**

|  |
| --- |
| **4. Objetivos** |
| En este apartado debes definir objetivos generales y específicos del Proyecto APT. Es importante aclarar que los objetivos se deben plantear en forma clara, concisa y sin dar mayores explicaciones, es decir, deben entenderse por sí solos. Se sugiere redactarlos utilizando un verbo en infinitivo, pues ello obliga a precisar acciones concretas. |

|  |  |
| --- | --- |
| Objetivo general | *Describe el o los objetivos generales de tu trabajo. Estos representan las grandes metas del proyecto que realizarás, de manera que te servirán de guía para que, una vez finalizado todo el proceso, puedas contrastar el resultado con lo planificado y así ver en qué medida fue posible cumplirlo.*  *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 | *Describe los objetivos específicos del proyecto. Estos permiten aterrizar el trabajo y trazar procedimientos concretos a seguir. Se desprenden del objetivo general.*   * *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.* |

|  |
| --- |
| **5. Metodología** |
| En el siguiente apartado deberás describir la metodología, propia de tu disciplina, que utilizarás para resolver el proyecto APT antes descrito, incluyendo las etapas y métodos de trabajo. |

|  |
| --- |
| Descripción de la Metodología |
| *Describe cómo abordarás el problema o situación que se identificó anteriormente, señalando la metodología que se utilizará para cumplir con tu objetivo.*  *Cuando el proyecto a desarrollar es grupal, es necesario incorporar la definición de las funciones, tareas y responsabilidades asociadas a cada integrante del equipo.*  *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:   Planificación y Definición del Proyecto:*   * *Definir los requerimientos y el alcance del proyecto según las historias de usuario.* * *Diseñar el modelo de datos.*   * Desarrollo e Implementación:*   * *Planificación de Sprints.* * *Desarrollo iterativo de modelos de estrellas.* * *Entrenamiento de modelos de ML.* * *Reuniones diarias (Daily).*   * Integración y Pruebas:*   * *Pruebas unitarias y de integración.* * *Revisión y ajuste.*   * Capacitación y Despliegue:*   * *Desarrollar material de capacitación.* * *Entrenamiento de usuario.* |

|  |
| --- |
| **6. Evidencias** |
| A continuación, describe qué evidencias serán evaluadas en el informe de avance y en el informe final de tu proyecto APT. Estas evidencias deben ser acordadas con tu docente. Se entenderá por evidencia los productos que se desarrollen durante el proyecto y cuyo propósito sea visibilizar o documentar cómo se ha implementado el trabajo. |

|  |  |  |  |
| --- | --- | --- | --- |
| **Tipo de evidencia**  **(avance o final)** | **Nombre de la evidencia** | **Descripción** | **Justificación** |
| **avance** | **Historias de usuario** | *Son las necesidades explicadas desde el punto de vista del usuario final, de una manera general e informal sobre la(s) función(es) de un software.* | *El propósito de una historia de usuario es alinear el producto final con las expectativas de los usuarios, asegurando que las funcionalidades del sistema respondan a necesidades reales del negocio y los objetivos del proyecto.* |
| **avance** | **Toma de requerimientos** | *Documento que identifica las necesidades del negocio y las expectativas de los Stakeholders acerca del proyecto.* | *Es fundamental para entender el alcance del proyecto, asegurando que el sistema Data Warehouse responda a las necesidades de los usuarios finales y la gerencia.* |
| **avance** | **Pipeline en Modelo medallón** | *Desarrollo del pipeline de ETL basado en el modelo medallón, cargando los datos desde el ERP a GCP en tres capas: bronce, plata y oro.* | *Permite estructurar el flujo de datos de manera eficiente, asegurando que la calidad de los datos mejore progresivamente y estén listos para análisis avanzado.* |
| **avance** | **Implementación de modelos Estrella** | *Implementación de un esquema de datos en el Data Warehouse basado en el modelo estrella, con tablas de hechos y dimensiones.* | *El modelo estrella es ideal para análisis en BI, ya que optimiza las consultas y facilita la creación de dashboards e informes para usuarios de negocio.* |
| **avance** | **Entrenamiento de modelos de ML** | *Entrenamiento de modelos de Machine Learning con los datos cargados en el Data Warehouse, orientados a generar predicciones y análisis avanzados.* | *Demuestra el uso avanzado de los datos para optimizar la toma de decisiones, destacando la capacidad del sistema para realizar predicciones basadas en datos históricos.* |
| **final** | **Puesta en marcha de modelos ML** | *Despliegue de los modelos de Machine Learning entrenados, listos para su uso en producción y su integración con el sistema Data Warehouse.* | *Refleja el éxito en la integración de herramientas de inteligencia artificial en el proyecto, mejorando la capacidad predictiva y de análisis para la empresa.* |
| **final** | **Carga de Dashboards e informes en Looker** | *Creación y publicación de dashboards interactivos y reportes en Looker Studio, utilizando los datos procesados en el Data Warehouse.* | *Facilita la visualización de información clave para los usuarios finales, permitiendo una toma de decisiones más informada y accesible para la gerencia y otros usuarios* |

|  |
| --- |
| **7. Plan de Trabajo** |
| En la siguiente tabla define la planificación de tu Proyecto APT de acuerdo a lo requerido. |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Plan de Trabajo Proyecto APT** | | | | | | |
| Competencia o unidades de competencias | Nombre de Actividades/Tareas | Descripción Actividades/Tareas | Recursos | Duración de la actividad | Responsable[[1]](#footnote-1) | Observaciones |
| *Dominio de TIC* | *Configuración de infraestructura en GCP* | *Configurar la infraestructura necesaria en GCP para alojar el Data Warehouse y conectar con las fuentes de datos.* | *Acceso a GCP, documentación de GCP, equipo técnico* | *2 semanas* | *Daniela Gálvez*  *Kiara Caycho* | *La disponibilidad de GCP puede ser limitada por problemas de configuración o permisos de acceso.* |
| *Análisis y Diseño de Sistemas* | *Diseño del Data Warehouse* | *Definir el esquema de datos que se implementará en el Data Warehouse, incluyendo la identificación de tablas de hechos y dimensiones.* | *Software de modelado de datos, documentación del ERP* | *3 semanas* | *Sebastian Orellana* | *Las modificaciones en los requerimientos pueden retrasar la entrega del diseño.* |
| *Desarrollo y Mantenimiento de Software* | *Implementación del Pipeline de ETL* | *Desarrollar el pipeline para extraer, transformar y cargar los datos desde el ERP hacia GCP en el modelo medallón.* | *Python/Java, herramientas ETL, GCP* | *4 semanas* | *Kiara Caycho* | *Se necesita asegurar la limpieza y calidad de los datos para evitar problemas en los informes y dashboards finales.* |
| *Gestión de Proyectos Informáticos* | *Planificación y seguimiento del proyecto* | *Gestionar las actividades y tiempos del equipo, coordinando entregables parciales y revisiones con los stakeholders.* | *Herramientas de gestión de proyectos (Trello, Jira, etc.)* | *Todo el proyecto* | *Ignacio Pi* | *La alineación de tiempos y recursos con los Stakeholders es clave para asegurar la entrega dentro del plazo estimado.* |
| *Trabajo en Equipo y Comunicación* | *Revisión y pruebas de prototipos* | *Trabajar con el equipo para revisar los primeros prototipos del Data Warehouse y hacer ajustes según feedback recibido.* | *Equipos de pruebas, datos de prueba* | *2 semanas* | *Ignacio Pi* | *Puede haber dificultades técnicas o de comunicación que afecten la coordinación de las revisiones, lo que podría causar ajustes y retrasos.* |
| *Innovación y Emprendimiento* | *Integración de modelos de ML* | *Desarrollar e integrar modelos de Machine Learning para mejorar el análisis de los datos almacenados en el Data Warehouse .* | *Herramientas de ML (TensorFlow, scikit-learn), GCP* | *5 semanas* | *Sebastian Orellana* | *Asegurar que los modelos sean efectivos y estén alineados con los objetivos del negocio; posibles ajustes en base a resultados o feedback.* |
| *Seguridad Informática* | *Aseguramiento de la calidad y seguridad de datos* | *Implementar medidas de seguridad para proteger los datos en GCP y durante la transmisión desde el ERP.* | *Herramientas de seguridad, GCP* | *2 semanas* | *Daniela Gálvez* | *Es importante asegurar que los datos sensibles estén protegidos durante todo el proceso para evitar posibles fugas o accesos no autorizados.* |

|  |
| --- |
| **8. Carta Gantt** |
| Busca un formato de Carta Gantt que te acomode y organiza en este las actividades planificadas en el punto anterior considerando el periodo asignado para el desarrollo de tu Proyecto APT. Debes mantener la temporalidad del periodo académico en el desarrollo de las tres fases que contempla la Asignatura de Portafolio de Título. |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Actividad** | **Fase 1** | | | | **Fase 2** | | | | | | | | | | | | **Fase 3** | | | |
| **S 1** | **S 2** | **S 3** | **S 4** | **S 5** | **S 6** | **S 7** | **S 8** | **S 9** | **S 10** | **S 11** | **S 12** | **S 13** | **S 14** | **S 15** | **S 16** | | **S 17** | **S 18** |
| **Definición del proyecto** | **X** | **X** | **X** | **X** |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |
| **Historias de usuario y toma de requerimientos** |  |  |  | **X** | **X** |  |  |  |  |  |  |  |  |  |  |  | |  |  |
| **Configuración de infraestructura en GCP** |  |  |  |  | **X** | **X** |  |  |  |  |  |  |  |  |  |  | |  |  |
| **Diseño del Data Warehouse** |  |  |  |  | **X** | **X** | **X** |  |  |  |  |  |  |  |  |  | |  |  |
| **Pipeline en Modelo medallón (ETL)** |  |  |  |  |  |  |  | **X** | **X** | **X** |  |  |  |  |  |  | |  |  |
| **Implementación de modelos Estrella** |  |  |  |  |  |  |  |  | **X** | **X** | **X** |  |  |  |  |  | |  |  |
| **Entrenamiento de modelos de ML** |  |  |  |  |  |  |  |  |  | **X** | **X** | **X** | **X** |  |  |  | |  |  |
| **Puesta en marcha de modelos ML** |  |  |  |  |  |  |  |  |  |  |  |  |  | **X** |  |  | |  |  |
| **Carga de Dashboards e informes en Looker** |  |  |  |  |  |  |  |  |  |  |  |  |  | **X** | **X** |  | |  |  |
| **Revisión de Flujo de datos** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | **X** | |  |  |
| **Revisión de documentación** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | **X** | | **X** |  |
| **Preparación para Defensa** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | **X** | **X** |

1. En caso de que el Proyecto APT sea grupal, en esta columna deben indicar el nombre de los responsables de cada tarea o actividad. Esto posteriormente permitirá diferenciar la evaluación por cada integrante. [↑](#footnote-ref-1)