}

**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 estudiante | **Kiara Caycho Daniela Galvez Sebastian Orellana Ignacio PI** |
| Rut | **14.750.456-k**  **20.140.132-1**  **18.862.994-6**  **19.134.443-k** |
| Carrera | **Computer Engineering** |
| 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 | *Implementation of Data Smart in ERP system* |
| Área (s) de desempeño(s) | *Big Data, Business Intelligence and Machine Learning.* |
| Competencias | * Database design and modeling to structure the DatasMart. * Systems integration ensures an efficient and secure flow of information. * ETL processes for extracting, transforming and loading data * Proficient in BI tools such as Looker Studio for the creation of dashboards and reports. * Project management with agile and Kimball methodology * Data security * ML model training |

|  |
| --- |
| **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? ¿Por qué es relevante este tema para el campo laboral de tu carrera?*   *Because it is a real problem faced by the current company and it would allow us to face a current contingency specific to our career mention using current technologies and platforms.*   * *¿Por qué es relevante este tema para el campo laboral de tu carrera?*   *Because it allows you to step into digital transformation and make use of data and make it accessible information for decision making, thus optimizing business performance and providing a competitive advantage in a data-driven environment and modern tools such as ML models.*   * *¿Dónde se ubica la situación que vas a abordar? (Ej.: País, región, comuna o institución) ¿Cuáles son las características principales de ese lugar?*   *The project will be managed for HVR with offices in Chile, RM, Providencia*   * *¿A quiénes afecta o impacta la situación que vas a abordar? (Ej.: Grupo etario, usuarios de algún servicio, etc.).*   *It would affect customers in various industries and services that operate with the HVR system throughout Chile, with a focus on both operators and management.*   * *¿Cuál sería el aporte de valor (real o simulado) de tu Proyecto APT para el contexto laboral y/o social en que se situaría?*   *Improve business decision making by creating an optimized DataMart and data projection with ML, enabling clear and accurate reporting and visualizations, facilitating access to critical information.* |
| Descripción del Proyecto APT | *Create a DataSmart hosted in GCP with the loading of information on a daily basis, in addition to the creation of ML models. Optimizing the creation of reports and dashboard with information for users and management.* |
| 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?*   * *Database Design and Modeling: Fundamental for structuring DataSmart, ensuring that data is efficiently organized for further analysis.* * *ETL (Extract, Transform, Load) processes: Crucial for extracting data from the ERP, transforming it and loading it into the DataSmart hosted in GCP, ensuring quality and consistency of information.* * *BI and Data Visualization Tools: Essential for creating reports and dashboards that present information in a clear and accessible way, tailored for both end users and management.* * *ML model training: Creates value for predictive values that optimize decision making.* * *Data Security and Regulatory Compliance: Important to protect the integrity and confidentiality of the data handled in DataSmart, complying with the corresponding regulations.* * *Agile Project Management: Necessary to coordinate project development, ensuring that deadlines are met and incremental value results are delivered.* |
| 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?*  *The project is directly aligned with our career mention and with the interest we have in our professional development. Since we will be able to develop aspects such as data management in cloud infrastructure, the analysis of it, use the data obtained for the training of ML models and the subsequent delivery of critical information in an intelligent way. This project will contribute to us as we will put into practice the methodologies, techniques and tools learned in our mention and professional interest.* |
| 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*   *From August 13 to mid-December.*   1. *Horas asignadas a la asignatura*   *3 hours per week in person.*   1. *Materiales requeridos*   *Access to the company's DB, GCP Credits*   1. *Factores externos que facilitan su desarrollo*   *Support from the promoter company, both logistically and with the data*   1. *Factores externos que dificultan su desarrollo y maneras en que podrías solucionarlos*   *External factors that hinder your development and ways in which you could address them* |

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.*  *Design and develop a DataSmart hosted in GCP to optimize the daily load of information, train ML models to obtain predictive values and the subsequent generation of reports and 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.*   * *Implementing data integration from ERP to GCP* * *Designing a pipeline for data cleansing and data transformation* * *Designing and implementing a Datasmart in GCP* * *Training ML models.* * *Establish data security mechanisms* * *Create cross-cutting reports and dashboards for the system.* |

|  |
| --- |
| **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.*  *A combined methodology between an agile development for iterative deliveries and a Kimball methodology will be used for the creation of the Datasmart. According to this, it will be developed as follows:   • Project Planning and Definition:*  *Define the requirements and scope of the project.*  *Design the data model.*  *• Development and Implementation:*  *Sprint planning.*  *Iterative development of star models.*  *ML model training*  *Daily meetings.*  *• Integration and Testing:*  *Unit and integration testing.*  *Review and adjustment.*  *• Training and Deployment:*  *Develop training materials.*  *User training.* |

|  |
| --- |
| **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** |
| **advance** | **Intake of requirements** | *Document that identifies business needs and project expectations.* | *It is critical to understand the scope of the project, ensuring that the DataSmart system meets the needs of end users and management.* |
| **advance** | **Pipeline in Medallion Model** | *Development of the ETL pipeline based on the medallion model, loading data from ERP to GCP in three layers: bronze, silver and gold.* | *It allows you to structure the data flow efficiently, ensuring that the quality of the data improves progressively and is ready for advanced analysis.* |
| **advance** | **Implementation of Star models** | *Implementation of a data schema in DataSmart based on the star model, with fact and dimension tables.* | *The star model is ideal for BI analysis, as it optimizes queries and facilitates the creation of dashboards and reports for business users.* |
| **advance** | **ML model training** | *Training of Machine Learning models with the data loaded in the DataSmart, oriented to generate predictions and advanced analysis.* | *Demonstrates the advanced use of data to optimize decision making, highlighting the system's ability to make predictions based on historical data.* |
| **end** | **Commissioning of ML models** | *Deployment of trained Machine Learning models ready for production use and integration with the DataSmart system.* | *Reflects success in integrating artificial intelligence tools into the project, improving predictive and analytical capabilities for the company.* |
| **end** | **Loading Dashboards and reports in Looker** | *Creation and publication of interactive dashboards and reports in Looker Studio, using the data processed in DataSmart.* | *Facilitates visualization of key information for end users, enabling more informed and accessible decision making for management and other users* |

|  |
| --- |
| **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 |
| *ICT proficiency* | *Infrastructure configuration in GCP* | *Set up the necessary infrastructure in GCP to host DataSmart and connect to data sources.* | *Access to GCP, GCP documentation, technical equipment* | *2 weeks* | *Daniela Galvez*  *Kiara Caycho* | *GCP availability may be limited by configuration problems or access permissions.* |
| *Systems Analysis and Design* | *DataSmart Design* | *Define the data schema to be implemented in DataSmart, including identification of fact and dimension tables.* | *Data modeling software, ERP documentation* | *3 weeks* | *Sebastian orellana* | *Changes in requirements may delay the delivery of the design.* |
| *Software Development and Maintenance* | *ETL Pipeline Implementation* | *Develop the pipeline to extract, transform and load data from ERP to GCP in the medallion model.* | *Python/Java, ETL tools, GCP* | *4 weeks* | *Kiara Caycho* | *It is necessary to ensure the cleanliness and quality of the data to avoid problems in the final reports and dashboards.* |
| *IT Project Management* | *Project planning and monitoring* | *Manage team activities and timelines, coordinating deliverables and stakeholder reviews.* | *Project management tools (Trello, Jira, etc.)* | *The entire project* | *Ignacio Pi* | *Alignment of time and resources with stakeholders is key to ensure delivery within the estimated timeframe.* |
| *Teamwork and Communication* | *Prototype review and testing* | *Work with the team to review early DataSmart prototypes and make adjustments based on feedback received.* | *Test equipment, test data* | *2 weeks* | *Ignacio Pi* | *There may be technical or communication difficulties affecting the coordination of reviews, which could cause adjustments and delays.* |
| *Innovation and Entrepreneurship* | *Integration of ML models* | *Develop and integrate Machine Learning models to improve the analysis of the data stored in the DataSmart.* | *ML tools (TensorFlow, scikit-learn), GCP* | *5 weeks* | *Sebastian orellana* | *Ensure models are effective and aligned with business objectives; possible adjustments based on results or feedback.* |
| *Computer Security* | *Data security and quality assurance* | *Implement security measures to protect data in GCP and during transmission from the ERP.* | *Security tools, GCP* | *2 weeks* | *Daniela Galvez* | *It is important to ensure that sensitive data is protected throughout the process to avoid possible leakage or unauthorized access.* |

|  |
| --- |
| **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** |
| *Project* **definition** | **X** | **X** | **X** | **X** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Intake of requirements** |  |  |  |  | **X** |  |  |  |  |  |  |  |  |  |  |  |  |  | |
| **Infrastructure configuration in GCP** |  |  |  |  | **X** | **X** |  |  |  |  |  |  |  |  |  |  |  |  | |
| **DataSmart Design** |  |  |  |  | **X** | **X** | **X** |  |  |  |  |  |  |  |  |  |  |  | |
| **Pipeline in Medallion Model (ETL)** |  |  |  |  |  |  |  | **X** | **X** | **X** |  |  |  |  |  |  |  |  | |
| **Implementation of Star models** |  |  |  |  |  |  |  |  | **X** | **X** | **X** |  |  |  |  |  |  |  |
| **ML model training** |  |  |  |  |  |  |  |  |  | **X** | **X** | **X** | **X** |  |  |  |  |  |
| **Commissioning of ML models** |  |  |  |  |  |  |  |  |  |  |  |  |  | **X** |  |  |  |  |
| **Loading Dashboards and reports in Looker** |  |  |  |  |  |  |  |  |  |  |  |  |  | **X** | **X** |  |  |  | |
| **Data Flow Review** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | **X** |  |  | |
| **Documentation review** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | **X** | **X** |  | |
| **Defense Preparation** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | **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)