

**SERVICIO NACIONAL DE ADIESTRAMIENTO EN TRABAJO INDUSTRIAL**

1. **INFORMACIÓN GENERAL**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Apellidos y Nombres: | Tumi Roque, Christopher Sandy D’Paris | ID: | 1435031 | |
| Dirección Zonal/CFP: | Arequipa-Puno | | | |
| Carrera: | Ingeniería de software con Inteligencia Artificial | Semestre: | | VI |
| Curso/ Mód. Formativo | INTELIGENCIA DE NEGOCIOS Y DATAWARE | | | |
| Tema del Trabajo: | Implementación de un Sistema de Inteligencia de Negocios | | | |

1. **PLANIFICACIÓN DEL TRABAJO**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **N°** | **ACTIVIDADES/ ENTREGABLES** | **CRONOGRAMA/ FECHA DE ENTREGA** | | | | | | | | | |
|  | Definición y Diseño del Data warehouse | 2 | 8 | / | 1 | 1 | / | 2 | 0 | 2 | 4 |
|  | Proceso ETL | 2 | 8 | / | 1 | 1 | / | 2 | 0 | 2 | 4 |
|  | Modelado de Datos | 2 | 8 | / | 1 | 1 | / | 2 | 0 | 2 | 4 |
|  | Arquitectura BI | 2 | 8 | / | 1 | 1 | / | 2 | 0 | 2 | 4 |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |

|  |  |  |
| --- | --- | --- |
|  | | |
| C:\Unidad_D\Nuevo Logo\SENATI_FF-01 Modificado.jpg | [NOMBRE DEL TRABAJO] | |
| [APELLIDOS Y NOMBRES] | [ESCALA] |

**PROCESO DE EJECUCIÓN**

|  |  |
| --- | --- |
| **OPERACIONES / PASOS /SUBPASOS** |  |
| ***1 Definición y Diseño del Data Warehouse:*** |  |
| **1.1 Definición de un data warehouse y su importancia**  Qué es un data warehouse?  Un data warehouse es un sistema de datos almacenados que centraliza, almacena y organiza grandes volúmenes de información proveniente de diferentes fuentes.  Importancia del Data Warehouse:  \*Toma de decisiones.  \*Análisis y generación de informes.  \*Integración de datos de multiples fuentes. |  |
| **1.2 Diseño del esquema dimensional para ElectroTiendas**  En los esquemas dimensionales, los datos se estructuran en hechos y dimensiones. Se describen las entidades relevantes para el negocio de ElectroTiendas:  **Tablas Hechos:** Esta tabla contiene datos transaccionales numéricos, datos cuantitativos  \*Ventas:  idVenta,fecha,producto,cliente,  vendedor,cantidad,precio unitario,  descuento  **Tablas Dimensiones:** Estas tablas contienen datos cualitativos que complementan a las tablas de hechos.  \*Vendedores: Información de los vendedores (id, nombres, apellidos)  \*Productos: Información del producto (id, nombre, categoría, proveedor)  \*Clientes: Información del cliente (id, nombre, apellidos)  \*Fecha: Información de la fecha (id, año, mes, trimestre, día) |  |
| **2 Proceso ETL:** |  |
| **2.1 Descripción del proceso ETL**  El proceso ETL es el mecanismo por el cual los datos de las fuentes externas se integran en el Data Warehouse.   * **Extracción (Extract)**:   + Los datos provienen de sistemas operacionales (ERP, CRM, bases de datos de ventas e inventarios) y de fuentes externas (proveedores, plataformas de marketing, etc.).   + Ejemplo: Datos de ventas, inventario y proveedores provenientes de diferentes bases de datos y aplicaciones. * **Transformación (Transform)**:   + Los datos se limpian, transforman y estandarizan para asegurar consistencia.   + Operaciones comunes incluyen la normalización de fechas, la conversión de divisas, la integración de registros duplicados y la agregación de datos a nivel de resumen (por ejemplo, total de ventas por sucursal).   + También se puede aplicar enriquecimiento de datos, como la asignación de categorías de productos a partir de códigos de barras. * **Carga (Load)**:   + Los datos transformados se cargan en el Data Warehouse. Dependiendo de la carga, puede ser **inmediata** (real-time) o **por lotes** (batch) según los requerimientos del negocio.   + Se pueden usar herramientas como **Apache NiFi**, **Talend**, **Microsoft SQL Server Integration Services (SSIS)**, o **Apache Airflow** para este proceso. |  |
| **2.2 Identificación de las herramientas y tecnologias para realizar el ETL**  Herramientas y Tecnologías para el ETL:  Herramientas ETL: Apache NiFi, Talend, SSIS, Alteryx.  Almacenamiento: Bases de datos relacionales como SQL Server, PostgreSQL, o plataformas en la nube como Amazon Redshift, Google BigQuery.  Lenguajes de Programación: Python, SQL, y lenguajes de script para transformar y cargar los datos de manera eficiente. |  |
| **3 Modelado de Datos:** |  |
| **3.1 Explicación de la importancia del modelado**  Para la implementación de un Date Warehouse es importante el proceso de modelado de datos, esto nos permitirá organizar los datos de manera que permita consultas rápidas y eficientes. |  |
| **3.2 Técnicas de modelado dimensional**  Técnicas de modelado dimensional:  - Modelo Estrella: este modelo es el mas común y empleado al momento de crear data warehouse, donde la tabla de hechos se conecta a varias tablas dimensionales. Es sencillo y eficiente en las consultas.  -Modelo copo de nieve: Similar al modelo estrella, pero las tablas de dimensiones están normalizadas, lo que genera un ahorro en disco y una desventaja en las consultas. |  |
| **4 Arquitectura BI:** |  |
| **4.1 Diseño de arquitectura BI**  Componentes de la Arquitectura BI:  Data Warehouse: Almacén centralizado donde se almacenan los datos históricos de ventas, productos, clientes, etc.  Herramientas de Visualización: Como Power BI que permiten crear dashboards interactivos y reportes personalizados.  Herramientas de Análisis y Reportes: Se usan para generar reportes automáticos o personalizados, a menudo integrados con herramientas como SQL Server Reporting Services (SSRS) o Power BI. |  |
| **4.2 Identificación de componentes para la toma de decisiones**  Contribución a la Toma de Decisiones:  Los dashboards permitirán monitorear KPIs(Indicadores de rendimiento) como las ventas diarias, los márgenes de ganancia por producto, productos demandados, etc.  Los reportes personalizados permitirán a la gerencia obtener información relevante para cada área de negocio. |  |
| **5 Seguridad y Calidad de Datos:** |  |
| **5.1 Planteamiento de Estrategias de seguridad**  Seguridad de los Datos:  Implementación de políticas de control de acceso: Asegurando que solo los usuarios autorizados puedan acceder o modificar datos sensibles.  Cifrado de datos: Para proteger la información durante la transmisión y almacenamiento. |  |
| **5.2 Descripción de implementación, mecanismos de control de accesos** |  |
| Control de Calidad de los Datos:  Validación de datos: Verificar que los datos cumplen con los requisitos de calidad (por ejemplo, datos no nulos, formatos correctos).  Monitoreo continuo: Implementación de procesos de monitoreo para detectar y corregir problemas de calidad de datos, como duplicados o inconsistencias.  Auditoría de Datos:  Implementación de registros de auditoría para rastrear cambios en los datos y accesos al sistema, utilizando herramientas como Apache Kafka para registro de eventos. |  |
|  |  |
|  |  |

**INSTRUCCIONES:** debes ser lo más explícito posible. Los gráficos ayudan a transmitir mejor las ideas. No olvides los aspectos de calidad, medio ambiente y SHI.

**INSTRUCCIONES: completa la lista de recursos necesarios para la ejecución del trabajo.**

|  |  |  |
| --- | --- | --- |
| **1. MÁQUINAS Y EQUIPOS** | | |
|  | Computadora Personal |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| **3. HERRAMIENTAS E INSTRUMENTOS** | | |
|  | Excel |  |
|  | Sql Server |  |
|  | Power Bi |  |
|  | Browser |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| **5. MATERIALES E INSUMOS** | | |
|  | Acceso a Internet |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

