

Sistema y Gestión Middleware ETL

Integrantes: Benjamin Pavez y Jhopcel Rosales

Índice

Índice.....	1
Abstract.....	2
Desarrollo de ingeniería	2
Conclusiones.....	4
Reflexiones	5

Abstract

El presente proyecto, denominado APT, se centra en el desarrollo de un Middleware para la empresa Cencosud, la cual opera en diversas zonas a nivel internacional. El objetivo principal es centralizar los datos de Recursos Humanos (RR.HH.) de Cencosud a nivel internacional, creando un sistema unificado y accesible que permita una gestión eficiente del talento y la toma de decisiones estratégicas basadas en datos. Este Middleware actuará como un puente intermediario que permitirá compartir información entre las diferentes áreas de la empresa, facilitando la inteligencia de negocios en distintos sectores del país. El Middleware gestionará la integridad y calidad de los datos mediante expectativas autogeneradas, y se implementará un proceso diario de extracción, transformación y carga (ETL) de datos hacia un Datawarehouse, que será consumido por aplicaciones de Business Intelligence (BI). Esto permitirá automatizar el proceso ETL para crear una base de datos centralizada y consistente a nivel global, y garantizar la accesibilidad a los datos procesados a través de diferentes plataformas.

Desarrollo de ingeniería

El desarrollo del Middleware se basa en una arquitectura robusta y escalable que garantiza la integración eficiente de datos provenientes de múltiples fuentes. Se implementarán mecanismos avanzados de validación y limpieza de datos para asegurar su calidad e integridad antes de su carga en el Datawarehouse. Además, se utilizarán herramientas de alto calibre para garantizar la escalabilidad y el mantenimiento del sistema a largo plazo, incluyendo tecnologías robustas y escalables, una arquitectura modular y flexible, y mecanismos de seguridad y control de acceso.

Tecnologías Utilizadas

- Python: Lenguaje de programación principal para el desarrollo del proyecto.

- Apache Airflow: Herramienta para la orquestación de tareas en el proceso ETL.
- Great Expectations: Herramienta para gestionar la integridad y calidad de los datos.
- SQLAlchemy: ORM agnóstico para diferentes motores de base de datos.
- Azure: Proveedor de nube para el despliegue de la aplicación y bases de datos.

Proceso ETL

1. Extracción: Se realizará una extracción diaria de datos desde diversas fuentes, incluyendo bases de datos relacionales, archivos planos y APIs. Este proceso garantizará que se recopile la información más actualizada y relevante de todas las áreas de RR.HH. de la empresa.
2. Transformación: Aplicación de lógica de transformación a los datos extraídos para asegurar su calidad e integridad, incluyendo la normalización y enriquecimiento de datos. Este paso es crucial para unificar los datos provenientes de diferentes sistemas y formatos, asegurando que sean consistentes y útiles para el análisis.
3. Carga: Los datos transformados se cargarán masivamente en un Datawarehouse, optimizado para consultas rápidas y eficientes. Este Datawarehouse servirá como la fuente central de verdad para todas las aplicaciones de BI y análisis de datos de la empresa.

Casos de Uso

- Análisis del Comportamiento de Empleados: Utilización de modelos de machine learning para identificar patrones en el desempeño y comportamiento de los empleados, permitiendo a la empresa personalizar programas de desarrollo y retención de talento. Esto ayudará a identificar a los empleados con alto potencial y a diseñar planes de carrera personalizados.

- Predicciones de Rotación de Personal: Implementación de modelos de regresión logística para predecir la rotación de personal, ayudando a la empresa a planificar estrategias de retención y contratación de manera más efectiva. Con esta información, Cencosud podrá anticipar y mitigar los riesgos asociados con la pérdida de talento clave.
- Evaluación del Clima Laboral: Análisis de datos de encuestas y feedback de empleados para evaluar el clima laboral y tomar decisiones informadas sobre mejoras en el ambiente de trabajo. Este análisis permitirá identificar áreas de mejora y diseñar intervenciones que promuevan un ambiente de trabajo positivo y productivo.

Conclusiones

El desarrollo del Middleware permitirá a Cencosud mejorar significativamente la gestión de sus datos de RR.HH., asegurando su integridad y calidad. Esto facilitará la toma de decisiones informadas y estratégicas en los diferentes sectores de la empresa, potenciando la inteligencia de negocios y la competitividad a nivel internacional. El sistema centralizado de datos de RR.HH. no solo mejorará la eficiencia operativa, sino que también permitirá a Cencosud crear programas de desarrollo efectivos que se adapten a las necesidades específicas de cada país, promoviendo así una gestión del talento más coherente y alineada con los objetivos estratégicos de la empresa.

Reflexiones

La implementación de un sistema unificado de datos de RR.HH. tiene el potencial de transformar la manera en que las empresas gestionan su talento y planifican sus estrategias a nivel global. Al centralizar la información, se facilita el acceso a datos precisos y actualizados, lo que es crucial para la toma de decisiones informadas. Esto no solo mejora la eficiencia operativa, sino que también permite a la empresa adaptarse rápidamente a los cambios del mercado y a las necesidades de sus empleados. El uso de tecnologías avanzadas como Apache Airflow para la orquestación de tareas ETL y Great Expectations para la gestión de la calidad de los datos asegura que el sistema sea robusto y confiable. Además, la integración de modelos de machine learning para el análisis de datos de RR.HH. abre nuevas posibilidades para la personalización de programas de desarrollo y la predicción de tendencias, como la rotación de personal. Desde una perspectiva estratégica, este proyecto posiciona a Cencosud como una empresa innovadora y proactiva en la gestión de su capital humano. La capacidad de analizar y predecir comportamientos y necesidades de los empleados permite diseñar políticas y programas que no solo retienen el talento, sino que también lo desarrollan y potencian. Esto es especialmente relevante en un entorno globalizado donde la competencia por el talento es feroz. En conclusión, el desarrollo de este Middleware no solo resuelve un problema técnico, sino que también impulsa a Cencosud hacia una gestión más estratégica y centrada en las personas. La capacidad de tomar decisiones basadas en datos precisos y de alta calidad es un diferenciador clave en el mercado actual, y

este proyecto coloca a Cencosud en una posición ventajosa para enfrentar los desafíos futuros.