**1. Introducción**

El área comercial de la empresa requiere recibir mensualmente los reportes de ventas generados por los vendedores. Estos reportes deben ser organizados, estandarizados, limpiados y parametrizados antes de ser almacenados en un repositorio central, desde donde los analistas puedan consultarlos y aprovechar los datos de manera eficiente.

**1.1 Problema a solucionar**

Actualmente, el proceso de recepción y procesamiento de los reportes es ineficiente, ya que se realiza de forma manual utilizando archivos de Excel. Los reportes llegan en distintos formatos, como Excel, PDF, CSV e incluso imágenes, lo que complica su estandarización. Esta diversidad de formatos genera un cuello de botella al momento de consolidar la información en un único archivo Excel.

Además, el volumen de datos ha crecido hasta el punto en que Excel se ha vuelto insostenible para su manejo. Las consultas (queries) se ejecutan lentamente, y no existe un dashboard centralizado que permita visualizar la información de forma eficiente. Esto obliga a los usuarios a abrir múltiples archivos, lo que incrementa el riesgo de modificar accidentalmente los datos y generar errores que afectan al resto del equipo.

**1.2 Solución propuesta**

Se propone automatizar la recepción de correos electrónicos que contienen los reportes, con el fin de estandarizar todos los archivos en un único formato (CSV). Estos archivos serán preparados para su procesamiento en una base de datos SQL, lo que permitirá una gestión más eficiente de la información. Posteriormente, los datos serán almacenados en un repositorio central y se desarrollará un dashboard en Power BI, facilitando así la consulta y el análisis por parte del equipo comercial.

**2. Arquitectura de la Solución**

La solución será desarrollada por un equipo de tres personas que trabajarán de forma remota. Para garantizar una colaboración eficiente, se utilizarán herramientas modernas de control de versiones y almacenamiento en la nube, permitiendo una integración fluida entre los distintos componentes del sistema.

**2.1 Flujo General del Proceso**

1. **Recepción de Reportes**  
   Un script en **Python** se conectará al servicio de **correo de Microsoft (Outlook/Exchange)** para identificar y descargar automáticamente los correos que contienen los reportes de ventas adjuntos.
2. **Extracción y Homogeneización de Datos**  
   Los archivos adjuntos, que pueden estar en formatos como Excel, PDF, CSV o imagen, serán procesados mediante librerías de Python como pandas, openpyxl, pdfplumber, tabula y pytesseract. El objetivo es extraer los datos relevantes y convertirlos a un formato **CSV estandarizado**.
3. **Centralización en la Nube**  
   Los archivos CSV generados serán almacenados en un servicio de almacenamiento en la nube como **Google Drive**, **SharePoint** o **OneDrive**, permitiendo el acceso compartido y seguro entre los miembros del equipo y otras áreas interesadas.
4. **Carga y Procesamiento en Base de Datos**  
   Los datos estandarizados serán cargados en una base de datos **MariaDB**, donde podrán ser procesados y consultados de forma eficiente. Posteriormente, los resultados procesados también serán exportados y almacenados en la nube, sirviendo como fuente de datos para la visualización.
5. **Visualización en Power BI**  
   **Power BI** se conectará a los archivos de datos procesados almacenados en la nube para construir un **dashboard interactivo**, que permitirá a los analistas del área comercial consultar la información de forma centralizada, segura y sin necesidad de manipular archivos manualmente.
6. **Colaboración y Control de Versiones**  
   Todo el código fuente, scripts de automatización y documentación del proyecto serán gestionados mediante **Git** y alojados en **GitHub**, lo que facilitará el trabajo colaborativo, el seguimiento de cambios y la integración continua entre los miembros del equipo.

**2.2 Beneficios de la Solución**

* **Automatización del flujo de trabajo**, reduciendo errores humanos y tiempos de procesamiento.
* **Estandarización de formatos**, facilitando el análisis y la comparación de datos.
* **Centralización en la nube**, permitiendo acceso remoto, seguro y compartido.
* **Procesamiento eficiente**, gracias al uso de MariaDB como motor de base de datos.
* **Visualización dinámica**, mediante dashboards intuitivos y actualizados en tiempo real.
* **Colaboración efectiva**, con Git y GitHub como herramientas clave para el trabajo en equipo.
* **Escalabilidad y flexibilidad**, permitiendo incorporar nuevos formatos, fuentes de datos o miembros al equipo sin afectar la operación.

**3. Objetivos del Proyecto**

**3.1 Objetivo General**

Desarrollar una solución automatizada y colaborativa que permita la recolección, estandarización, almacenamiento, procesamiento y visualización de reportes de ventas, optimizando el flujo de trabajo del área comercial y facilitando la toma de decisiones basada en datos.

**3.2 Objetivos Específicos**

* Automatizar la descarga de reportes de ventas desde correos electrónicos corporativos.
* Estandarizar los datos provenientes de múltiples formatos (Excel, PDF, CSV, imágenes) en archivos CSV uniformes.
* Centralizar los archivos estandarizados en un servicio de almacenamiento en la nube para facilitar el acceso remoto y compartido.
* Cargar y procesar los datos en una base de datos MariaDB para mejorar la eficiencia en las consultas.
* Exportar los datos procesados a la nube como fuente de datos para visualización.
* Diseñar un dashboard interactivo en Power BI que permita consultar la información de forma clara, segura y centralizada.
* Implementar un flujo de trabajo colaborativo utilizando Git y GitHub para el control de versiones y la coordinación entre los miembros del equipo.

**4. Alcance del Proyecto**

**4.1 Alcance Funcional**

* Automatización de la recepción de correos electrónicos con reportes adjuntos.
* Extracción de datos desde archivos en distintos formatos mediante scripts en Python.
* Conversión y limpieza de datos para su estandarización en formato CSV.
* Almacenamiento temporal y final de los archivos en servicios de nube (Google Drive, SharePoint o OneDrive).
* Carga de datos en una base de datos MariaDB para su procesamiento.
* Exportación de resultados procesados a la nube.
* Visualización de los datos mediante dashboards en Power BI.
* Gestión del código fuente y documentación del proyecto en GitHub.

**4.2 Alcance No Funcional**

* El sistema será desarrollado y mantenido por un equipo remoto de tres personas.
* Se garantizará la seguridad y privacidad de los datos mediante el uso de servicios en la nube con acceso controlado.
* La solución será escalable, permitiendo la incorporación de nuevos formatos de entrada o fuentes de datos en el futuro.
* Se priorizará la eficiencia en el procesamiento y la facilidad de uso para los analistas del área comercial.

Tareas para designar en Jira

| **Fase** | **Tarea** | **Duración (semanas)** | **Responsable** |
| --- | --- | --- | --- |
| Análisis | Recolección de Requisitos | 1 | Persona 1 |
| Análisis | Análisis de Requisitos | 1 | Persona 2 |
| Desarrollo | Desarrollo de Script de Recepción de Correos | 2 | Persona 1 |
| Desarrollo | Desarrollo de Script de Extracción y Homogeneización | 3 | Persona 2 |
| Desarrollo | Desarrollo de Carga en Base de Datos | 2 | Persona 3 |
| Desarrollo | Desarrollo de Exportación y Almacenamiento en la Nube | 2 | Persona 1 |
| Pruebas | Pruebas Unitarias | 2 | Persona 2 |
| Pruebas | Pruebas de Integración | 2 | Persona 3 |
| Implementación | Despliegue en Producción | 1 | Persona 1 |
| Implementación | Capacitación y Documentación | 1 | Persona 2 |