



Chat with your data to enrich online business

Team

Gabriela Moya
Amparo Iglesias Gordillo
Ana Graciela Perez Romero
Luis Gerardo Ramos Martinez

Promotor

★ Codigo Facilito

Innovation Challenge October 2024

Executive Challenge

Chat with your data" token utilization and cost monitoring

EN/ES

Content

| | |
|---|----------|
| Technical Report | 3 |
| Project Objective | 3 |
| Problem Addressed by the Project | 3 |
| Main Problems | 3 |
| Project Goals | 4 |
| System Description and Features | 4 |
| MVP (Minimum Viable Product) | 5 |
| Technologies and Services Used | 5 |
| Solution Architecture | 5 |
| Functional Requirements | 6 |
| Non-Functional Requirements | 6 |
| Learnings | 7 |
| Informe Técnico | 8 |
| Objetivo del Proyecto | 8 |
| Problemática que Resuelve el Proyecto | 8 |
| Principales Problemas | 8 |
| Metas del Proyecto | 9 |
| Descripción del Sistema y Funcionalidades | 9 |
| MVP (Producto Mínimo Viable) | 10 |
| Tecnologías y Servicios Utilizados | 10 |
| Arquitectura de la Solución | 10 |
| Diagrama de Arquitectura General | 10 |
| Requerimientos Funcionales | 11 |
| Requerimientos No Funcionales | 11 |
| Aprendizajes | 12 |

Technical Report

QueryWise

Data Query Platform with Natural Processing and Integrated Consumption Visualization

Project Objective

The primary objective is to develop a web application with a backend provisioned in Azure, using Microsoft authentication, enabling users to access uploaded Excel data, process it with artificial intelligence, and perform natural language queries. This platform will be implemented as a highly scalable solution, utilizing Azure infrastructure and Microsoft components to create a robust system that facilitates the interpretation of large volumes of data. The solution is designed to optimize administrative management and resource consumption control in companies, providing advanced and efficient decision-making tools. QueryWise provides advanced and efficient decision-making tools.

Problem Addressed by the Project

In many companies, especially small and medium-sized businesses (SMEs), the management of critical information is still done using traditional formats such as Excel. Excel has significant limitations when working with large volumes of data and lacks advanced analytical tools that allow for the visualization of patterns, trends, and metrics in real-time.

Main Problems

1. Difficulty in analyzing large volumes of data.
2. Limitations for advanced natural language queries.
3. Lack of real-time resource consumption monitoring.
4. Limited accessibility to advanced analysis tools.

Project Goals

1. Facilitate the analysis and use of large volumes of data for companies that still manage information in Excel.
2. Provide an intuitive tool to monitor and control costs and consumption associated with resource usage in real-time.
3. Offer managers and administrators a platform that allows them to quickly and effectively visualize business activities.
4. Enable small and medium-sized businesses to benefit from an accessible and cost-effective tool to improve the management and control of IT resource costs.
5. In the future, expand the accepted data sources, including PDFs and external databases, to enhance the tool's scalability and functionality.

System Description and Features

1. **Login:** Authentication through Microsoft Entra accounts to ensure users belong to a specific group.
2. **Excel File Upload:** Administrators can upload files that are validated and processed.
3. **Data Cleaning and Processing (ETL):** Use of Azure Data Factory to normalize data, with alerts to the user if the format is incorrect.
4. **Natural Language to SQL Querying:** Users make natural language queries using Azure OpenAI and Microsoft AI Services that translate the queries into SQL.
5. **Resource Consumption Visualization:** Real-time visualization of resources using Power BI.
6. **Permission and Resource Management:** The administrator assigns permissions and resource limits, which can be edited at any time. Additionally, the system allows each user's resource limit to be reset at defined intervals.
7. **Admin Control Panel:** Includes a detailed view of each user's resource consumption, automatic alerts if a user approaches their limit, and a general breakdown of associated costs.

MVP (Minimum Viable Product)

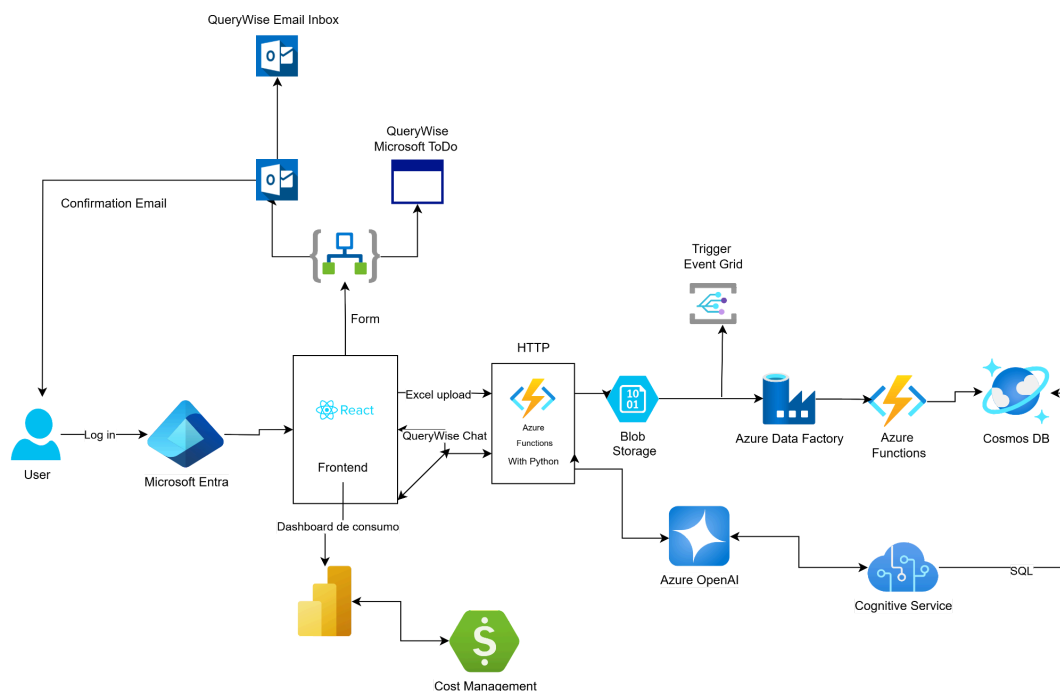
The minimal functional version of the system will include:

- Microsoft account login.
- Excel file upload and validation.
- Basic data cleaning and structuring.
- Data storage in a database.
- Natural language query function.
- Consumption visualization via Power BI.

Technologies and Services Used

- **Frontend:** React, Next.js.
- **Cost Visualization and Analysis:** Embedded Power BI, Azure Cost Management.
- **Data Processing:** Azure Data Factory, Azure Blob Storage.
- **Data Storage and Querying:** Cosmos DB, Microsoft SQL.
- **Authentication and Security:** Microsoft Entra, Azure Managed Identity.
- **Artificial Intelligence and Natural Language Processing:** Azure OpenAI, Microsoft AI Services.
- **Form Handling:** Logic App, Microsoft Outlook, Microsoft To Do.

Solution Architecture



Functional Requirements

1. **Authentication:** Use of Microsoft Entra for secure login.
2. **File Upload:** Functionality for uploading Excel files and storing them in Azure Blob.
3. **Natural Language Processing:** Azure OpenAI converts natural language queries to SQL.
4. **Query and Visualization:** Processed data is presented in charts and tables through embedded Power BI.
5. **Alerts and Notifications:** Automatic alerts when usage limits or high costs are reached.

Non-Functional Requirements

1. **Scalability:** Capability to expand Azure infrastructure to handle a larger volume of data and users.
2. **Security:** Data protection through Microsoft Entra and Azure Managed Identity, with encryption of data at rest and in transit.
3. **Availability:** Use of Azure cloud services to ensure high availability.
4. **Performance:** Optimization in data processing and querying.
5. **Usability:** Intuitive interface so that users without technical knowledge can query data.
6. **Maintenance and Updates:** Automation of updates in Power BI and cost reports.
7. **Fault Tolerance:** Use of Azure cloud services with replication and redundancy configurations to ensure fault tolerance.

Learnings

1. **Authentication and Security:** Integration with Microsoft Entra and Azure Managed Identity ensures effective user and permission control.
2. **Report Automation:** Power BI facilitates automatic information updates and real-time decision-making.
3. **Natural Language Processing:** Using Azure OpenAI enables the transformation of natural language queries into SQL, improving accessibility.
4. **Cost Optimization:** Azure Cost Management and Power BI provide a detailed and transparent view of resource-related costs.

Informe Técnico

QueryWise

Plataforma de Consulta de Datos con Procesamiento Natural y Visualización de Consumo Integrado

Objetivo del Proyecto

El objetivo principal es desarrollar una aplicación web con Backend aprovisionado desde azure, con autenticación mediante Microsoft, que permita a los usuarios acceder a datos cargados en formato Excel, procesarlos mediante inteligencia artificial y realizar consultas en lenguaje natural. Esta plataforma se implementará como una solución altamente escalable, utilizando la infraestructura de Azure y componentes de Microsoft para crear un sistema robusto que facilite la interpretación de grandes volúmenes de datos. La solución está diseñada para optimizar la gestión administrativa y el control de consumo de recursos en empresas, ofreciendo herramientas avanzadas y eficientes para la toma de decisiones. QueryWise proporciona herramientas avanzadas y eficientes para la toma de decisiones.

Problemática que Resuelve el Proyecto

En muchas empresas, especialmente en pequeñas y medianas (pymes), la gestión de información crítica se sigue realizando en formatos tradicionales como Excel. Excel presenta importantes limitaciones al trabajar con grandes volúmenes de datos y carece de herramientas de análisis avanzado que permitan visualizar patrones, tendencias y métricas en tiempo real.

Principales Problemas

1. Dificultad para analizar grandes volúmenes de datos
2. Limitación para consultas avanzadas en lenguaje natural
3. Falta de supervisión en tiempo real del Consumo de recursos
4. Accesibilidad limitada a herramientas de análisis avanzado

Metas del Proyecto

1. Facilitar el análisis y uso de grandes volúmenes de datos a empresas que aún gestionan información en Excel.
2. Proveer una herramienta intuitiva para supervisar y controlar los costos y consumos asociados al uso de recursos en tiempo real.
3. Brindar a los gerentes y administradores una plataforma que les permita visualizar actividades empresariales rápidamente y de manera eficaz.
4. Permitir que pequeñas y medianas empresas puedan aprovechar una herramienta accesible y rentable para mejorar la gestión y el control de costos de recursos informáticos.
5. En el futuro, se busca ampliar las fuentes de datos aceptadas, incluyendo PDF y bases de datos externas, con el fin de mejorar la escalabilidad y funcionalidad de la herramienta.

Descripción del Sistema y Funcionalidades

1. **Inicio de Sesión:** Autenticación mediante cuentas de Microsoft Entra, para asegurar la pertenencia de los usuarios al grupo.
2. **Subida de Archivos Excel:** Los administradores pueden cargar archivos que se validan y procesan.
3. **Limpieza y Procesamiento de Datos (ETL):** Uso de Azure Data Factory para normalizar los datos, con alertas al usuario en caso de formato incorrecto.
4. **Consulta en Lenguaje Natural a SQL:** Los usuarios consultan en lenguaje natural gracias a Azure OpenAI y Microsoft AI Services que traducen las consultas en SQL.
5. **Visualización de Consumo de Recursos:** Visualización en tiempo real de recursos mediante Power BI.
6. **Gestión de Permisos y Recursos:** El administrador asigna permisos y límites de recursos, los cuales pueden ser editados en cualquier momento. Adicionalmente, el sistema permite restaurar el límite de recursos de cada usuario a intervalos definidos.

7. **Panel de Control del Administrador:** Se incluye una vista detallada del consumo de recursos de cada usuario, alertas automáticas si un usuario se acerca a su límite, y un desglose general de costos asociados.

MVP (Producto Mínimo Viable)

La versión mínima funcional del sistema incluirá:

- Inicio de sesión con cuentas de Microsoft.
- Subida y validación de archivos Excel.
- Limpieza y estructuración básica de datos.
- Guardar los datos en Base de datos
- Función de consulta en lenguaje natural.
- Visualización del consumo mediante Power BI.

Tecnologías y Servicios Utilizados

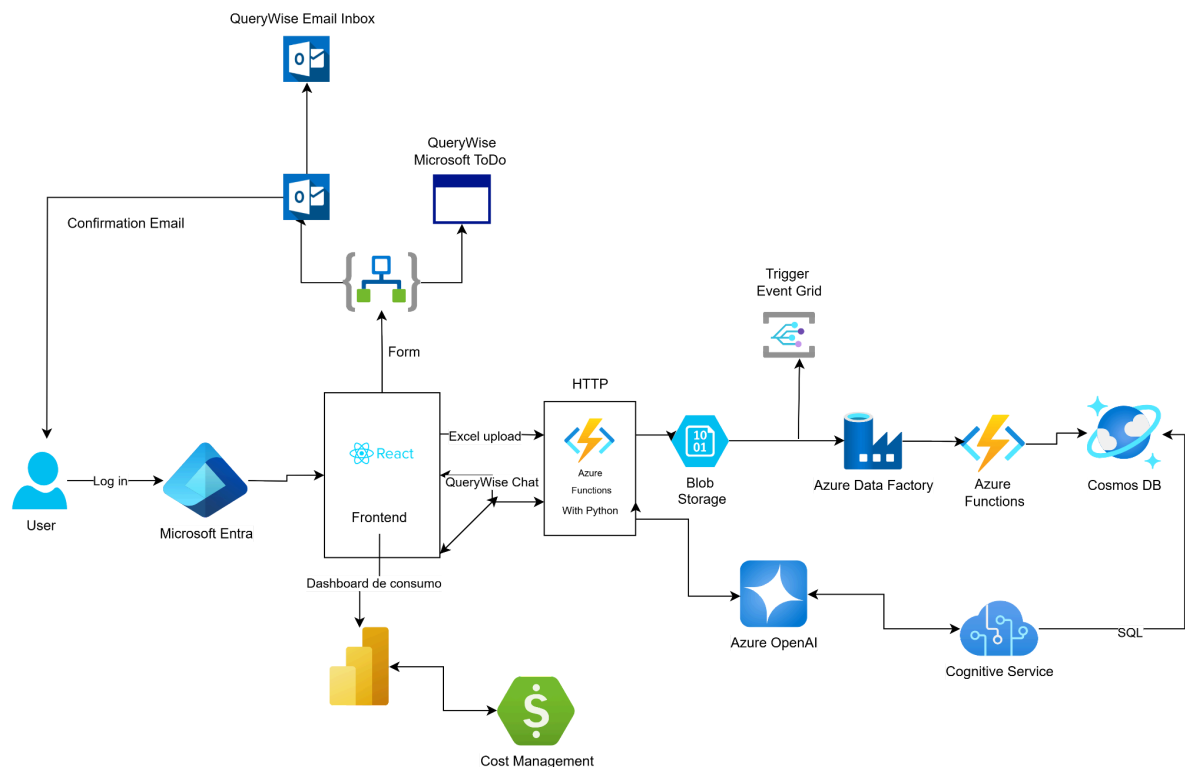
- **Frontend:** React, Next.js.
- **Visualización y Análisis de Costos:** Power BI embebido, Azure Cost Management.
- **Procesamiento de Datos:** Azure Data Factory, Azure Blob Storage.
- **Almacenamiento y Consulta de Datos:** Cosmos DB, Microsoft SQL.
- **Autenticación y Seguridad:** Microsoft Entra, Azure Managed Identity.
- **Inteligencia Artificial y Procesamiento Natural:** Azure Open AI, Microsoft AI Services.
- **Formulario:** Logic app, Microsoft Outlook, Microsoft To Do

Arquitectura de la Solución

Diagrama de Arquitectura General

Flujo de Datos: Desde la carga de datos en Excel hasta la consulta de datos procesados.

Interacción con Usuarios: Diagrama que muestra la autenticación, consulta y visualización.



Requerimientos Funcionales

1. **Autenticación:** Uso de Microsoft Entra para el inicio de sesión seguro.
2. **Carga de Archivos:** Funcionalidad de carga de archivos Excel y almacenamiento en Azure Blob.
3. **Procesamiento de Lenguaje Natural:** Azure OpenAI convierte consultas en lenguaje natural a SQL.
4. **Consulta y Visualización:** Los datos procesados se presentan en gráficos y tablas a través de Power BI embebido.
5. **Alertas y Notificaciones:** Alerta automática al alcanzar límites de uso o costos elevados.

Requerimientos No Funcionales

1. **Escalabilidad:** Capacidad de ampliar la infraestructura en Azure para manejar un mayor volumen de datos y usuarios.
2. **Seguridad:** Protección de datos a través de Microsoft Entra y Azure Managed Identity. Cifrado de datos en reposo y en tránsito.
3. **Disponibilidad:** Uso de servicios en la nube de Azure para garantizar alta disponibilidad.

4. **Rendimiento:** Optimización en el procesamiento y consulta de datos.
5. **Usabilidad:** Interfaz intuitiva para que usuarios sin conocimientos técnicos puedan consultar datos.
6. **Mantenimiento y Actualización:** Automatización de actualizaciones en Power BI y reportes de costos.
7. **Tolerancia a fallos:** Uso de los servicios en la nube de Azure con configuraciones para replicas y redundancia para asegurar la tolerancia a fallos.

Aprendizajes

1. **Autenticación y Seguridad:** La integración con Microsoft Entra y Azure Managed Identity asegura un control efectivo de usuarios y permisos.
2. **Automatización de Reportes:** Power BI facilita la actualización automática de información y la toma de decisiones en tiempo real.
3. **Procesamiento de Lenguaje Natural:** El uso de Azure OpenAI permite transformar consultas en lenguaje natural en SQL, mejorando la accesibilidad.
4. **Optimización de Costos:** Azure Cost Management y Power BI ofrecen una visión detallada y transparente de los costos asociados a los recursos.