

MICROSOFT
HACKATHON INNOVATION CHALLENGE
OCTOBER 2024
CÓDIGO FACILITO





DataVerse

DataVerse: Interactive & Efficient Chat

Powered by  Microsoft Azure

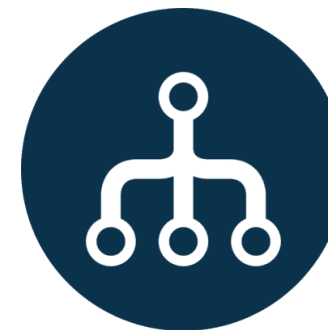
<https://youtu.be/nuuE4h0uz3k>

Descripción del proyecto

- DataVerse es una solución de chat interactiva e intuitiva diseñada específicamente para que los profesionales de RR.HH. accedan y analicen datos de empleados de múltiples fuentes. Con solo hacer preguntas en lenguaje cotidiano, los usuarios de RR.HH. pueden recuperar información relevante sobre su fuerza laboral, lo que facilita la toma de decisiones informadas rápidamente.
- DataVerse elimina la necesidad de conocimientos técnicos, lo que permite a los equipos de RR.HH. interactuar con datos complejos sin problemas a través de una interfaz conversacional.



Transparencia en costos y uso:
monitorear la utilización del sistema para
brindar visibilidad sobre los costos de las
consultas y el uso de los recursos, lo que
garantiza operaciones eficientes y
rentables.



Acceso unificado a datos de múltiples
fuentes: integrar información de varias
bases de datos, agilizando el acceso a
diversos datos de empleados en una
única interfaz cohesiva.



Potenciar a RR.HH. con acceso en
lenguaje natural: permitir que los
usuarios no técnicos obtengan
información esencial sobre la fuerza
laboral a través de consultas en
lenguaje natural, eliminando la
necesidad de conocimientos de SQL.



Objetivos



Microsoft



códigofacilito



Procesamiento de consultas conversacionales: acceda fácilmente a los datos de los empleados escribiendo preguntas como "Mostrar asistencia para octubre" o "Enumerar empleados inscritos en capacitación".



Integración de datos inteligente: extrae y agrega información de varias fuentes en una plataforma fácil de usar.

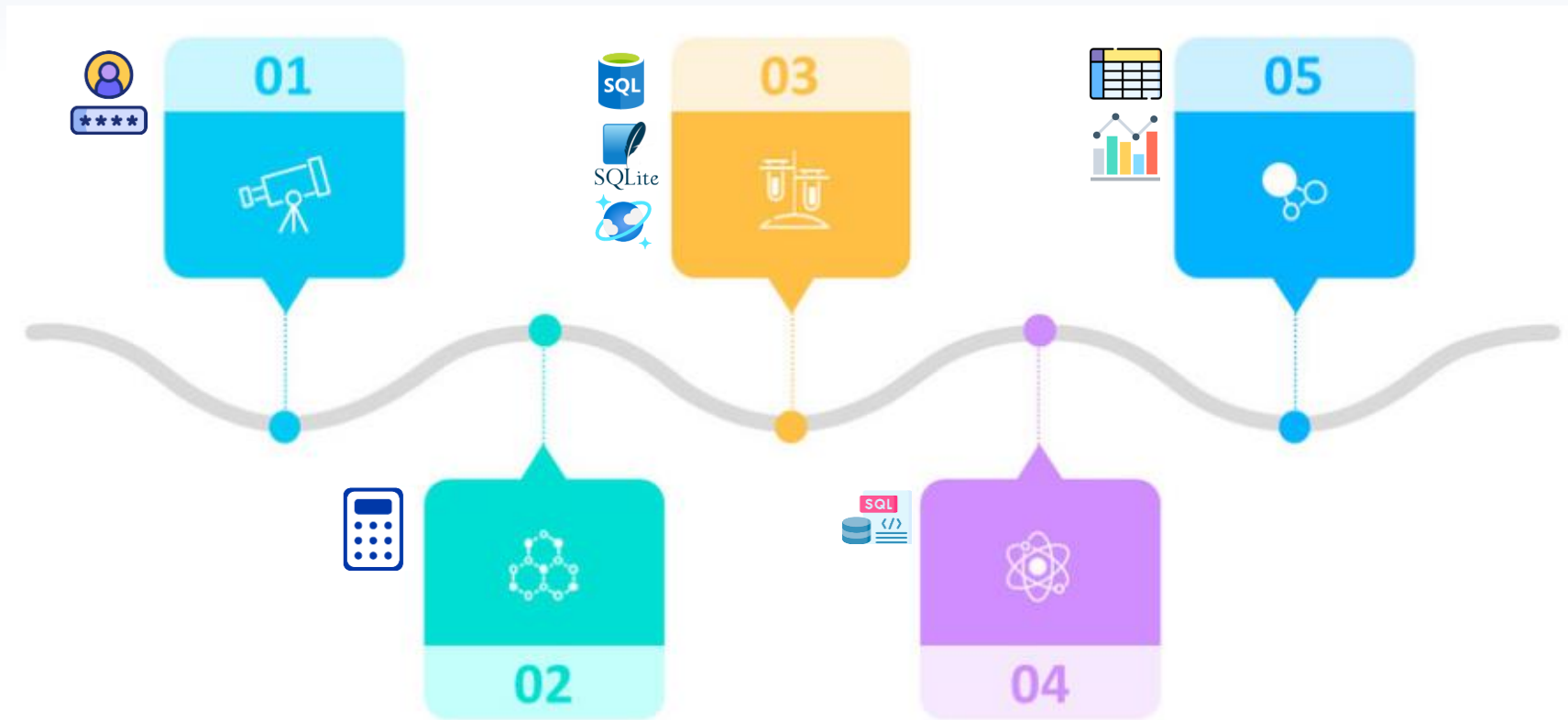


Monitoreo de uso: realiza un seguimiento de los costos de recuperación de datos, lo que ayuda a los usuarios a comprender y administrar el uso de manera efectiva.

1. Iniciar sesión y obtener tokens disponibles de acuerdo al especialista de RR.HH.

3. Identificar la intención de la consulta y determinar el contexto de datos para generar la consulta SQL.

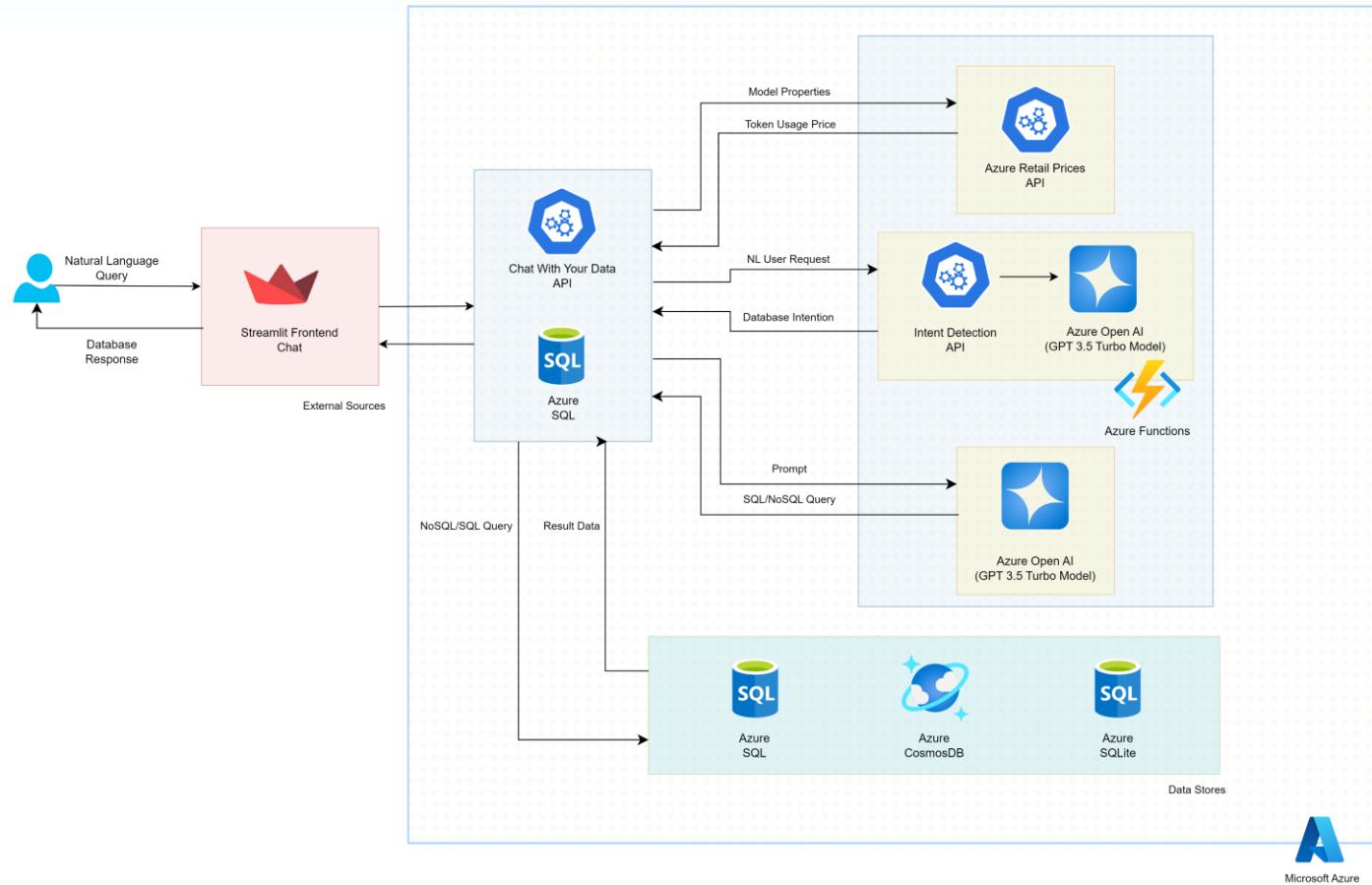
5. Visualizar los registros a nivel de tabla en el chat interactivo así como el histórico de consumo realizado.



2. Obtener el consumo de tokens de la consulta generada en lenguaje cotidiano así como su valor estimado de acuerdo a la calculadora de Azure.

4. Ejecutar la consulta SQL en el contexto de datos y retornar la información de los registros.
Totalizar tokens de entrada y salida para descontar al usuario.

Arquitectura del proyecto



El modelo C4 para visualizar la arquitectura de software.

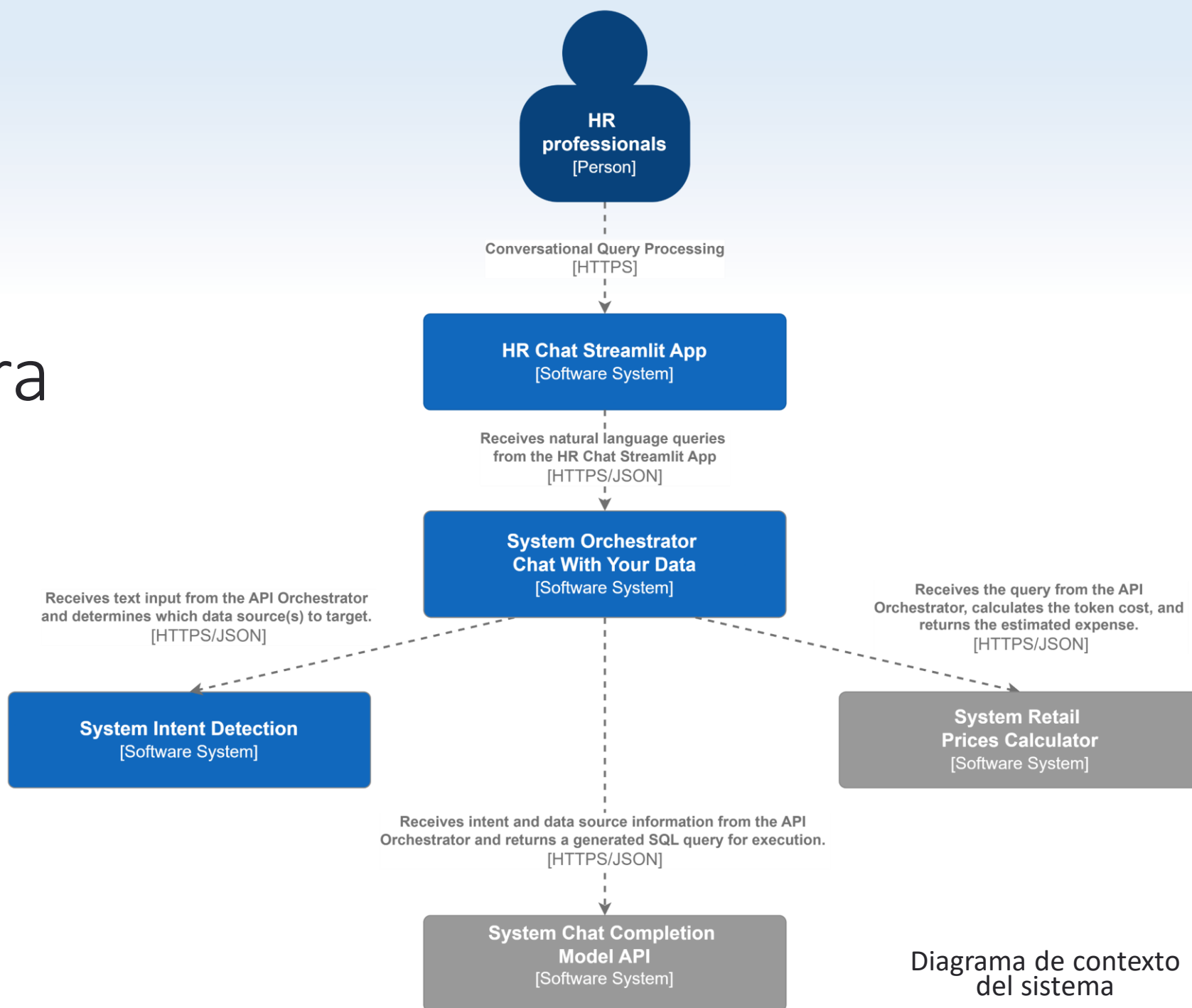


Diagrama de contexto del sistema

Diagrama de contenedor del sistema

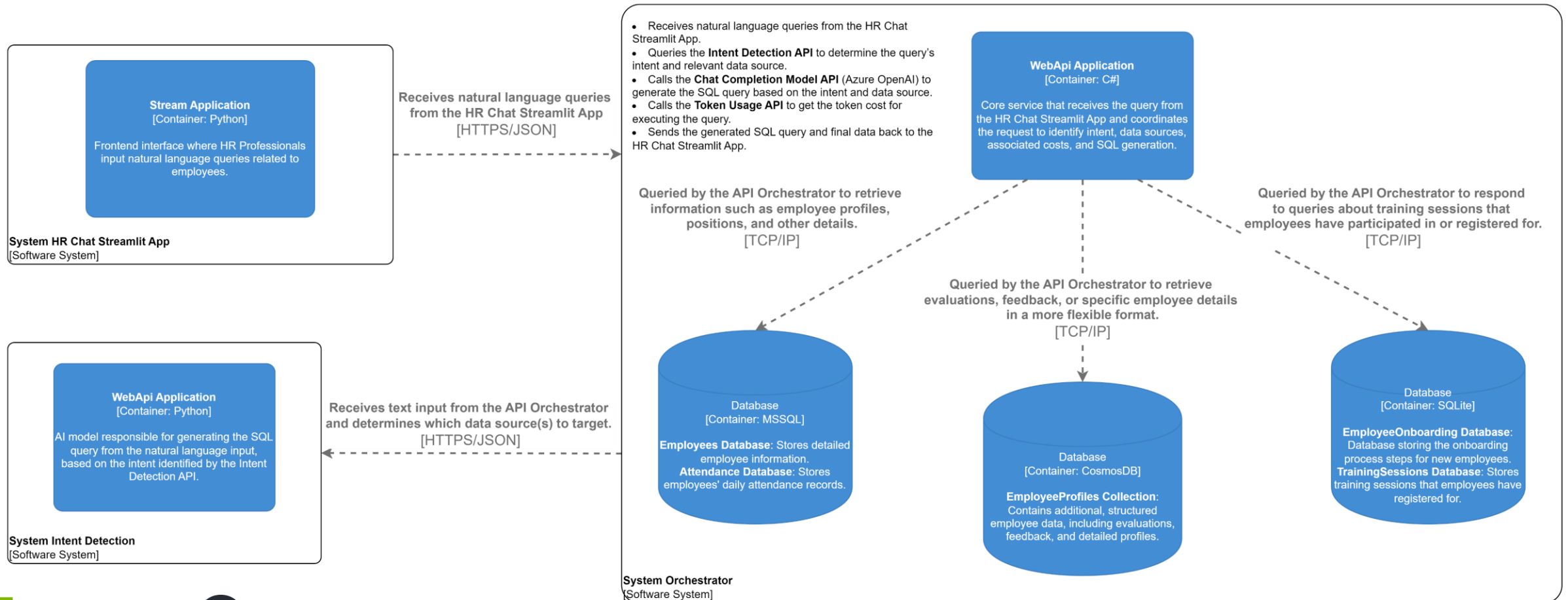
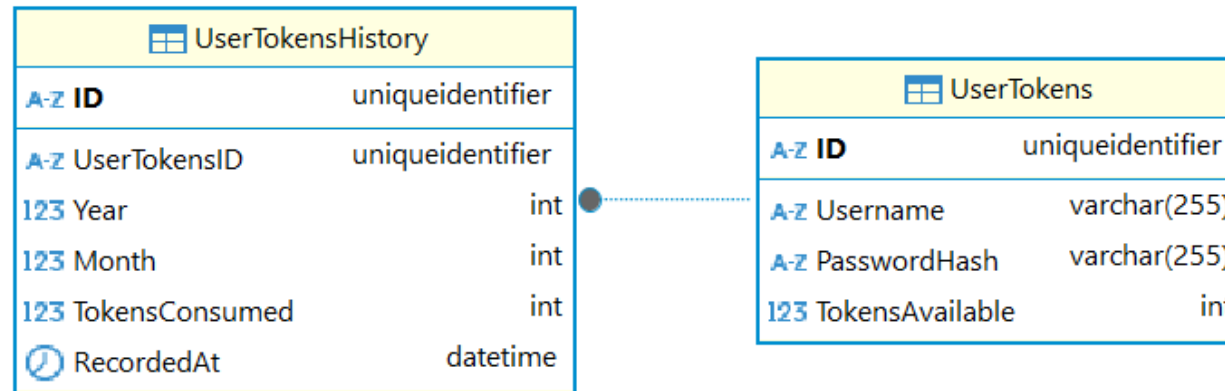


Diagrama de base de datos

Esta base de datos consta de dos tablas principales diseñadas para gestionar los tokens de autenticación de usuarios y rastrear su uso a lo largo del tiempo.



Registra el historial de consumo de tokens de cada usuario mensualmente, incluyendo detalles sobre cuándo y cuántos tokens se consumieron.

Almacena información de los usuarios y registra el número de tokens disponibles para cada usuario.



Attendance		
123 AttendanceID	int	
123 EmployeeID	int	
🕒 Date	date	
🕒 CheckIn	time	
🕒 CheckOut	time	

Almacena los registros diarios de asistencia de los empleados.

Employees		
123 EmployeeID	int	
A-Z DUI	varchar(10)	
A-Z FullName	varchar(255)	
A-Z Position	varchar(100)	
A-Z Department	varchar(100)	
🕒 HireDate	date	
123 Salary	decimal(10,2)	

Almacena información detallada de los empleados.



TrainingSessions		
123 SessionID	int	
123 EmployeeID	int	
A-Z SessionTitle	varchar(100)	
🕒 Date	date	

Almacena los pasos del proceso de incorporación para nuevos empleados.

EmployeeOnboarding		
123 EmployeeID	int	
A-Z Step	varchar(50)	
123 Completed	bit	
🕒 CompletionDate	date	

Almacena las sesiones de capacitación en las que los empleados se han registrado.



Contiene datos estructurados adicionales de los empleados, incluyendo evaluaciones, comentarios y perfiles detallados.

```
{
  "type": "object",
  "properties": {
    "employee_id": { "type": "string" },
    "full_name": { "type": "string" },
    "contact_info": {
      "type": "object",
      "properties": {
        "email": { "type": "string" },
        "phone": { "type": "string" }
      }
    },
    "position": { "type": "string" },
    "skills": {
      "type": "array",
      "items": { "type": "string" }
    },
    "certifications": {
      "type": "array",
      "items": {
        "type": "object",
        "properties": {
          "name": { "type": "string" },
          "issued_date": { "type": "string" }
        }
      }
    },
    "feedback": {
      "type": "array",
      "items": {
        "type": "object",
        "properties": {
          "date": { "type": "string" },
          "reviewer": { "type": "string" },
          "comments": { "type": "string" }
        }
      }
    },
    "DUI": { "type": "string" },
    "id": { "type": "string" }
  },
  "required": ["employee_id", "full_name", "contact_info", "position", "DUI", "id"]
}
```



Microsoft



códigofacilito

DataVerse IA responsable



Fairness

Equidad: Garantizar que el sistema proporcione respuestas y acceso equitativo a todos los usuarios, independientemente de sus consultas o contexto. Implementar controles de uso de tokens y asignación de tokens iniciales igualitarios para todos los usuarios podría ayudar a evitar discriminación en el acceso a datos.



Reliability & Safety

Confiabilidad y seguridad: Asegurar que los datos, especialmente información personal y confidencial, estén protegidos. Configurar una autenticación segura y encriptación en el almacenamiento de contraseñas y tokens de usuario es fundamental para evitar accesos no autorizados y proteger los datos sensibles.



Privacy & Security

Privacidad y seguridad de datos: Mantener la privacidad de los datos gestionando la consulta de información sensible y aplicando el principio de “mínima exposición” para solo permitir a los usuarios acceder a la información necesaria. El almacenamiento seguro de información personal debe adherirse a normativas de privacidad como GDPR o HIPAA, según corresponda.



Inclusiveness

Inclusión: Diseñar el sistema para admitir múltiples idiomas y consultas diversas en lenguaje natural, proporcionando una interfaz inclusiva que permita a usuarios de distintos niveles de habilidad tecnológica interactuar con el sistema de manera eficaz. Esto se puede lograr optimizando la comprensión de consultas en lenguaje natural para garantizar que todos los usuarios puedan expresar sus preguntas de manera accesible.



Microsoft



códigofacilito

DataVerse IA responsable

Transparencia: Asegurar que los usuarios comprendan cómo el sistema genera respuestas y consultas SQL. Proporcionar un resumen que explique la transformación de la consulta en lenguaje natural a SQL y detallar el consumo de tokens por cada consulta contribuye a la transparencia del sistema, permitiendo a los usuarios tomar decisiones informadas sobre su uso.

 Transparency Accountability

Responsabilidad: Implementar auditorías y registros en el sistema para monitorear y corregir posibles errores o sesgos en el modelo. Mantener un historial detallado de las consultas generadas y su consumo de tokens permite al equipo responsable revisar y ajustar el sistema según sea necesario para mantener la precisión y responsabilidad en la IA del proyecto.

Hackers



Daniela Vallejo
Python developer



Jonathan Narváez
Python developer



David Santafe
Backend .NET developer

Advisors



Carla Mamani
*MSFT MVP y Especialista de
Azure en Código Facilito*



Tomás Pucutay
*Data Science e IA Código
Facilito*

