Dec 25

Read Time: 10 Mins.

TECH mahindra

Case Study

Leading European Telco Accelerates Data Engineering with Agentic AI Solution

Overview

A telecommunications major needed a solution that could counter the delays and inconsistencies in delivery due to high volumes of daily pipeline requests consuming significant engineering capacity, slowing discovery, extending cycle times, and making onboarding inefficient.

We, at Tech Mahindra, implemented an AI-driven data engineering solution powered by Azure's generative AI capabilities to address these roadblocks. The solution automates code and pipeline generation, accelerating delivery and reducing overall engineering effort.

Client Background and Challenges

A regional telecommunications pioneer based in Europe, the client is known for its innovation and delivers a broad portfolio of digital services, including telephony, internet, and network-based ICT solutions across multiple brands.

The company is a key contributor to the economy it operates in and aims to create a trusted, secure connected world.

Key Challenges Included

- Siloed documentation and standards leading to slow discovery, inconsistent delivery
- High engineering effort for data pipelines, resulting in increased cycle time per request
- Need for standardization and automation to scale pipeline delivery
- High skill dependency resulting in delivery risk and slower onboarding

Business Drivers

- Approx – 300+ pipeline queries daily required engineers to manually handle them with majority of the time spent in DB/file pipelines and doc hunting (knowledge is fragmented across Confluence pages and standards) → slower delivery and inconsistency
- Need a self-service agent that automates schema detection, YAML authoring, SQL/PySpark generation, and enforces EDH standards

Solution and Approach

We built and deployed an Agentic AI-based autonomous pipeline builder across DevOps, enabling seamless delivery from UAT to production. The scalable, intelligence-led solution allows users to:

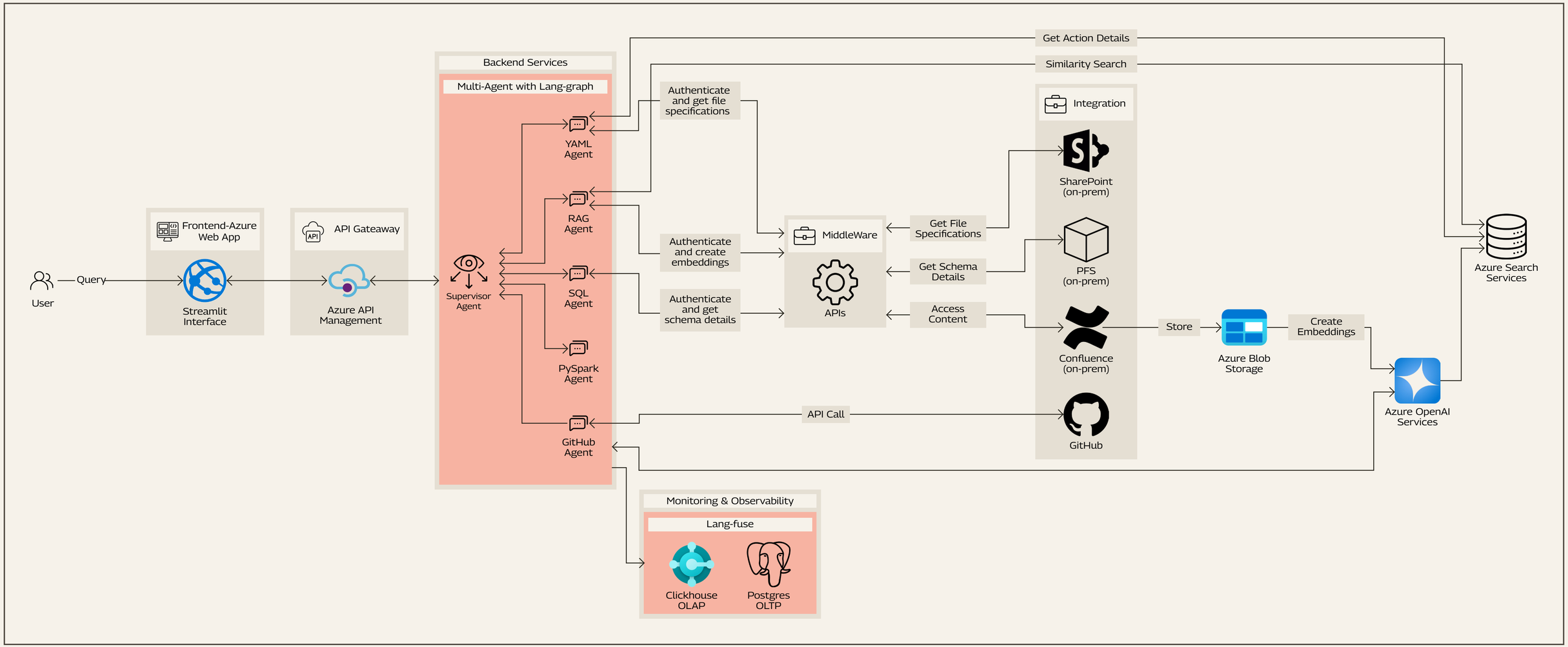
- Generate code and configurations (YAML/SQL/PySpark) from natural-language pipeline requests
- Use RAG-based Q&A for EDH guidance and patterns
- Auto-derive and validate parameters (mandatory/derived/default) with approval workflow
- Run pre- and post-validation pipelines

To deliver high reliability and performance, the solution integrates multiple Azure services, including Azure Machine Learning Workspace, Azure Search Service, Azure OpenAI Service, Azure Container Registry, and Azure App Service.

Key components include:

- **Azure OpenAI GPT models** for code and schema generation
- **Azure AI Search** for RAG
- **LangGraph agentic framework**
- **Streamlit UI** for request submission
- **GitHub** for automated commits
- **Confluence** for knowledge base retrieval
- **PFS** for configuration details
- **Collibra** for table and schema metadata
- **Langfuse** for tracing, evaluations, and metrics

This architecture streamlines user interactions, reduces manual engineering effort, and ensures consistent, high-quality responses with real-time accuracy.



Technical Architecture

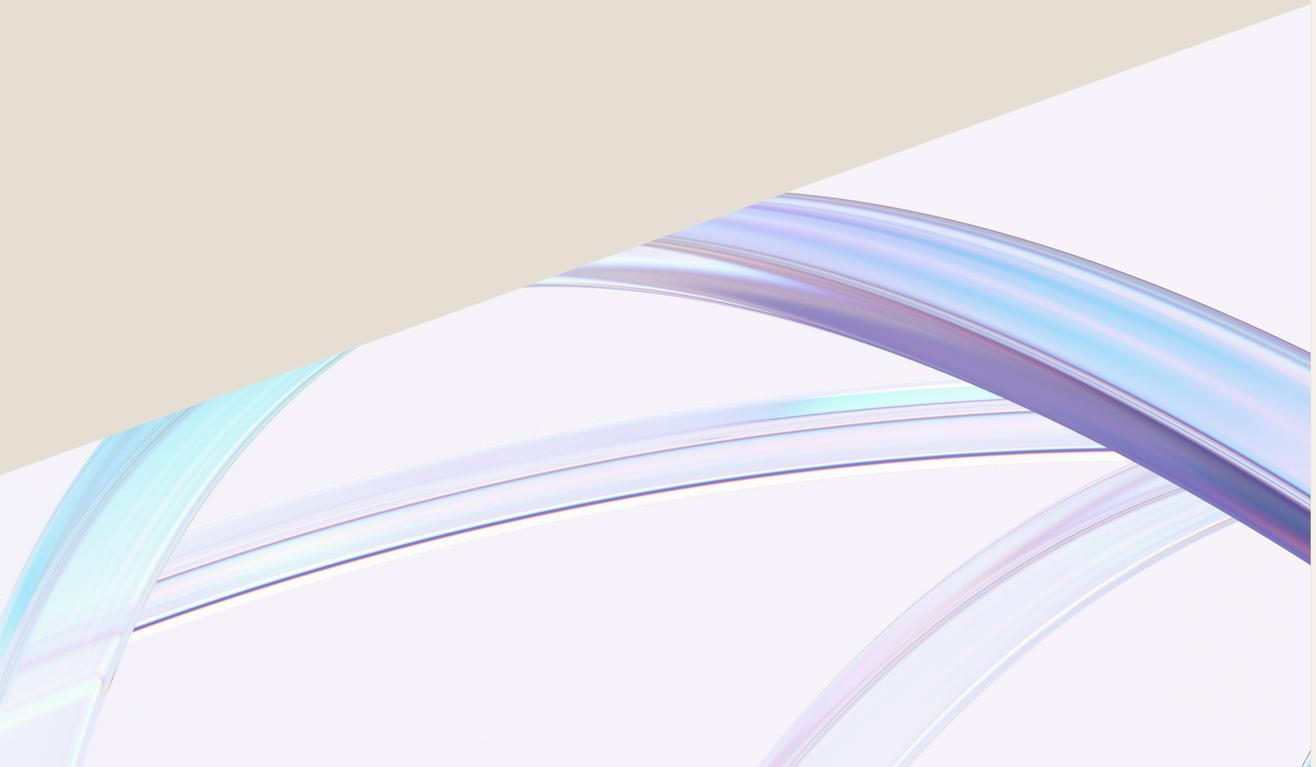
- **Azure Search Service:** Confluence documents chunked, indexed, and stored for semantic vector retrieval
- **Azure OpenAI:**
  - Text-Embedding-3-large for embeddings
  - GPT-4o for intent recognition, schema creation, YAML/PySpark generation
- **Azure Compute Instance:** Environment for containerised services
- **Docker + ACR:** Container images built and published to ACR
- **App Service:** Production deployment for the core service and Streamlit UI

Roles and Responsibilities

1. **Developers**
   Schema, logic, and code generation
2. **AI Engineers**
   Prompt engineering for SQL/YAML/PySpark
3. **Data Scientists**
   Validate generated schemas and transformations
4. **DevOps**
   Containerisation, CI/CD, monitoring

Compliance and Governance

- GDPR alignment and data residency
- Key Vault usage and Defender for Cloud image scanning
- Backup policies for Azure SQL; budget alerts via cost management



Documentation and Handover

Delivered a comprehensive Data Engineering Agent user manual covering:

- RAG usage and pipeline request templates
- Parameter lists (mandatory/derived/default)
- YAML/SQL/PySpark generation and approval workflows
- Troubleshooting scenarios (metadata gaps, FTP checks)
- Escalation paths

Business Outcomes and Impact

The AI pipeline automation has demonstrated improvements to the client's data engineering operations. By automating, it delivers measurable business value that includes:

Upto 50%
faster pipeline development

Shorter wait times for pipeline automation,
improving SLA performance and reducing engineer workload

3-5 hours
saved per request with use of agentic AI systems driving
operational efficiency and cost savings across the team

About Tech Mahindra

Tech Mahindra (NSE: TECHM) offers technology consulting and digital solutions to global enterprises across industries, enabling transformative scale at unparalleled speed. With 152,000+ professionals across 90+ countries helping 1100+ clients, Tech Mahindra provides a full spectrum of services including consult-ing, information technology, enterprise applications, business process services, engineering services, network services, customer experience & design, Al & analytics, and cloud & infrastructure services. It is the first Indian company in the world to have been awarded the Sustainable Markets Initiative's Terra Carta Seal, which recognises global companies that are actively leading the charge to create a climate and nature-positive future. Tech Mahindra is part of the Mahindra Group, founded in 1945, one of the largest and most admired multinational federation of companies. For more information on how TechM can partner with you to meet your Scale at Speed™ imperatives, please visit https://www.techmahindra.com/.

www.techmahindra.com
www.twitter.com/tech_mahindra
www.linkedin.com/company/tech-mahindra