



Pomagamy wydobyć wiedzę ukrytą
w Twoich danych.

Enterprise Databots
Projektowanie i zarządzanie
inteligentnymi DataBotami w organizacji

Łukasz Grała
Maciej Rubczyński
Jakub Wawrzyniak

Prowadzący



Łukasz Grala
MVP Data Platform & AI



Maciej Rubczyński
MVP AI



Jakub Wawrzyniak
MVP Data Platform

Sprawy organizacyjne

- Dostęp do środowiska
- Laboratoria – sposób realizacji
- Repozytorium kodu: <https://github.com/MaciejR/DataBotSQLWorkshop>
- Przerwy:
 - 11:00 – 11:15
 - 12:30 – 13:30
 - 15:30 – 15:45
- Feedback

WiFi

network: SQLDay2025

pass: SQLDay*25(!)

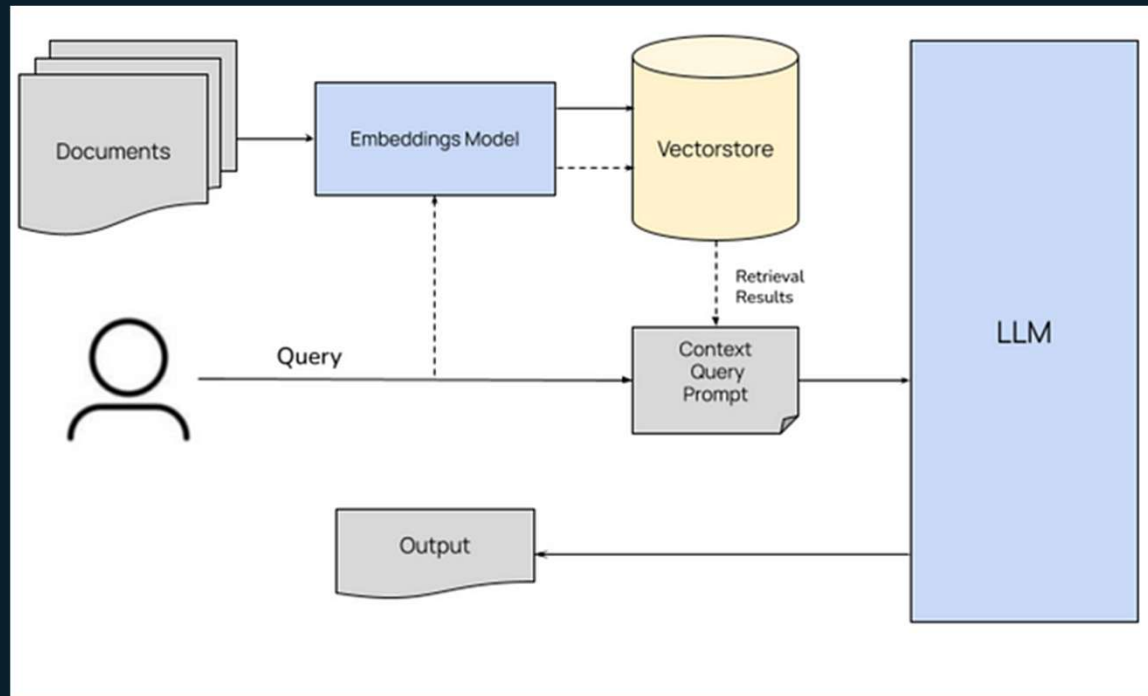
Agenda

- | | |
|---|------------------------|
| 1. Wprowadzenie, teoria i architektura Databota | [8:30 – 10:00] |
| 2. Dostęp do danych organizacyjnych | [10:00 – 11:00] |
| <i>PRZERWA</i> | <i>[11:00 – 11:15]</i> |
| 3. Budowa core Databota | [11:15 – 12:30] |
| <i>PRZERWA NA LUNCH</i> | <i>[12:30 – 13:30]</i> |
| 4. Uczenie i zarządzanie modelem | [13:30 – 14:45] |
| 5. Front-end i interfejs użytkownika | [14:45 – 15:30] |
| <i>PRZERWA</i> | <i>[15:30 – 15:45]</i> |
| 6. Zarządzanie i skalowanie | [15:45 – 16:30] |
| 7. Demo Microsoft Fabric AI Skills | [16:30 – 17:00] |
| 8. Podsumowanie, Q&A | [17:00 – 17:30] |

* Czas przeznaczony na dany moduł jest orientacyjny i może ulec zmianie.

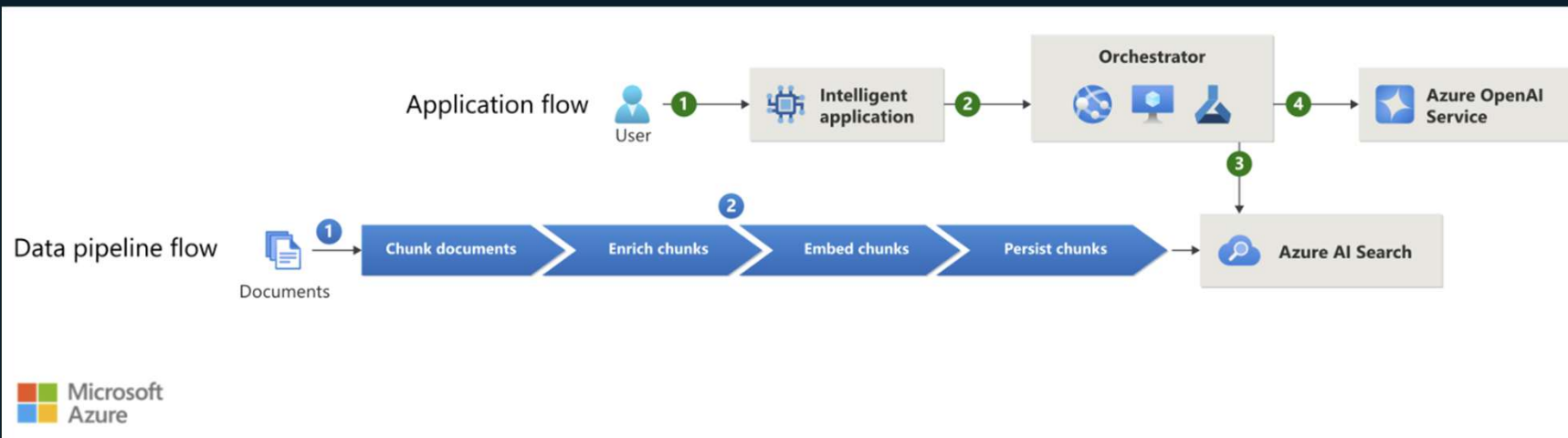
Wprowadzenie, teoria i architektura Databota

Architektura RAG (*Retrieval-Augmented Generation*)

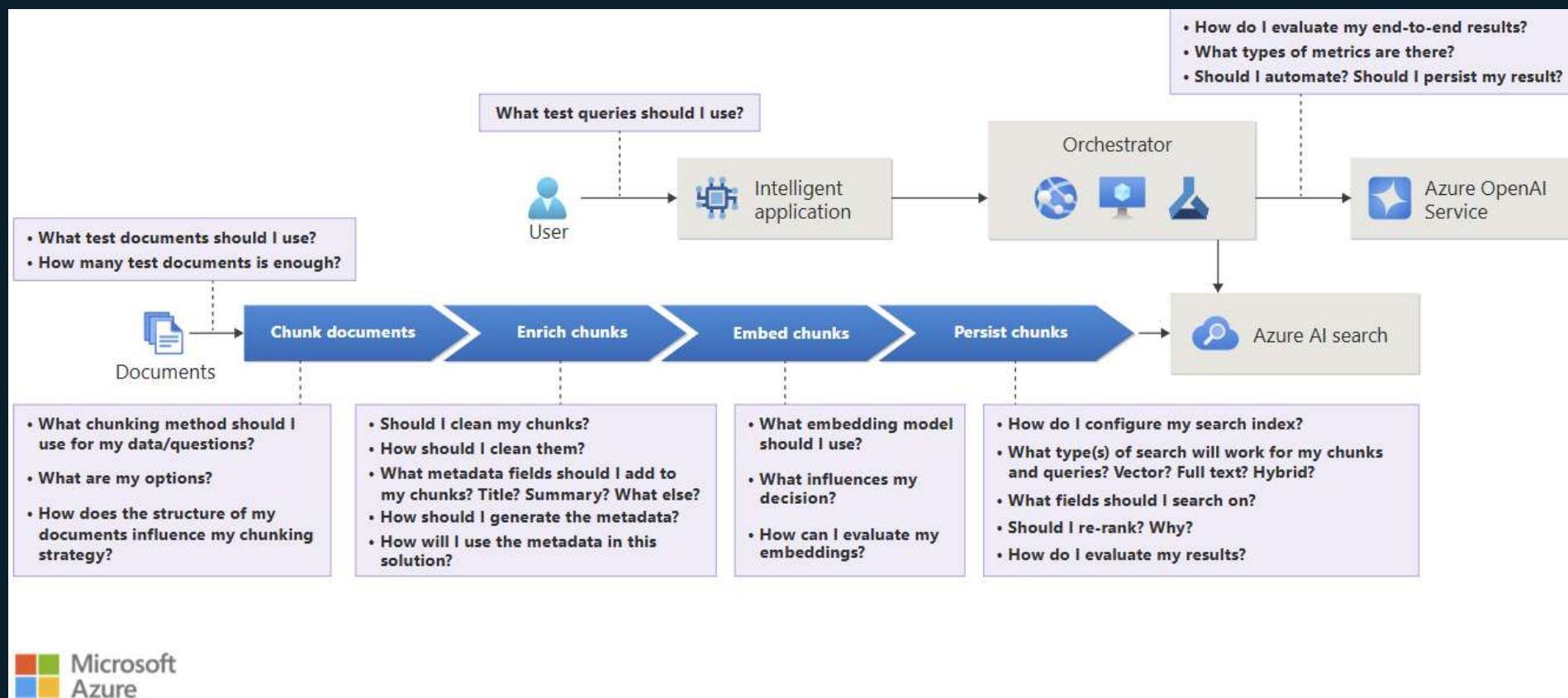


<https://blog.stackademic.com>

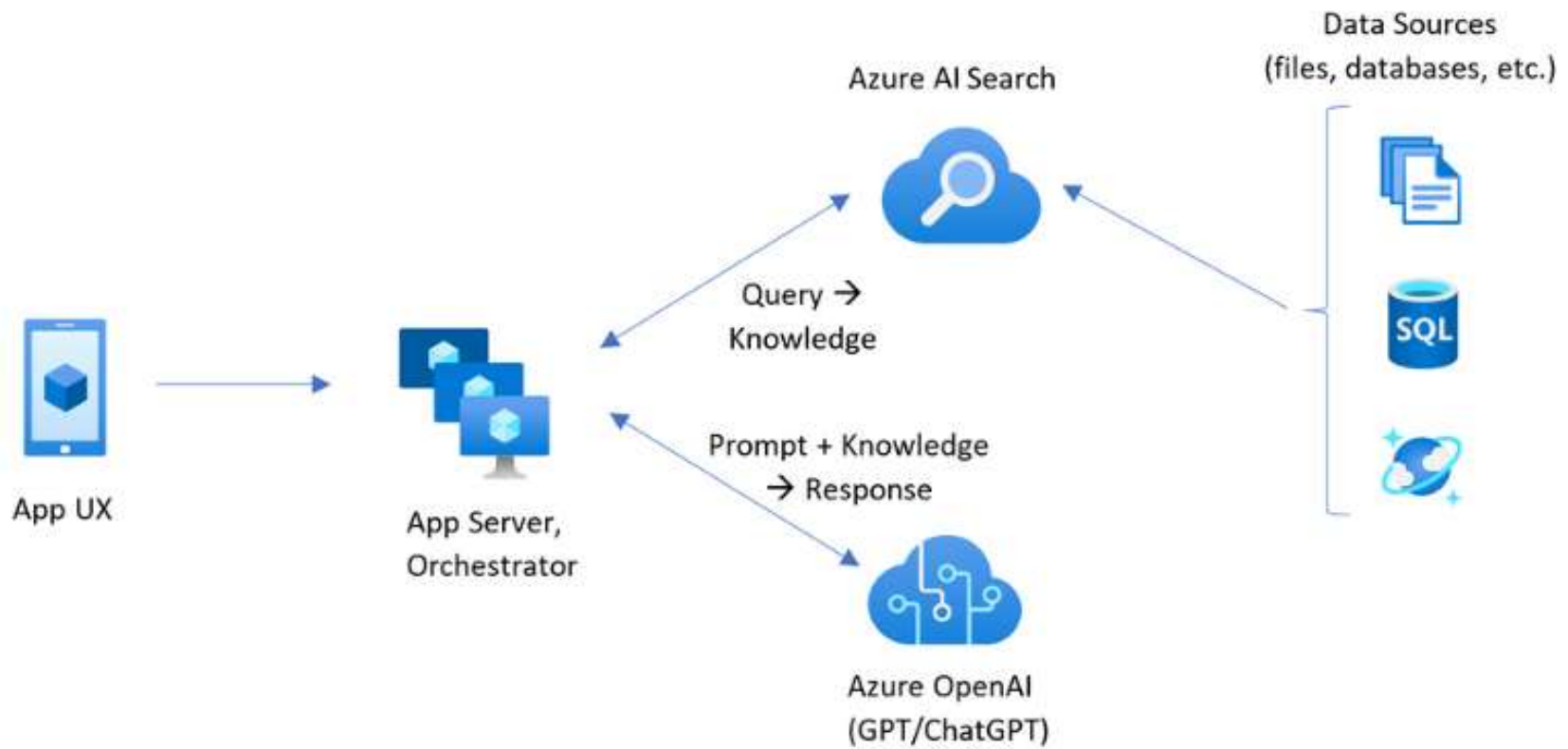
Architektura RAG (*Retrieval-Augmented Generation*)

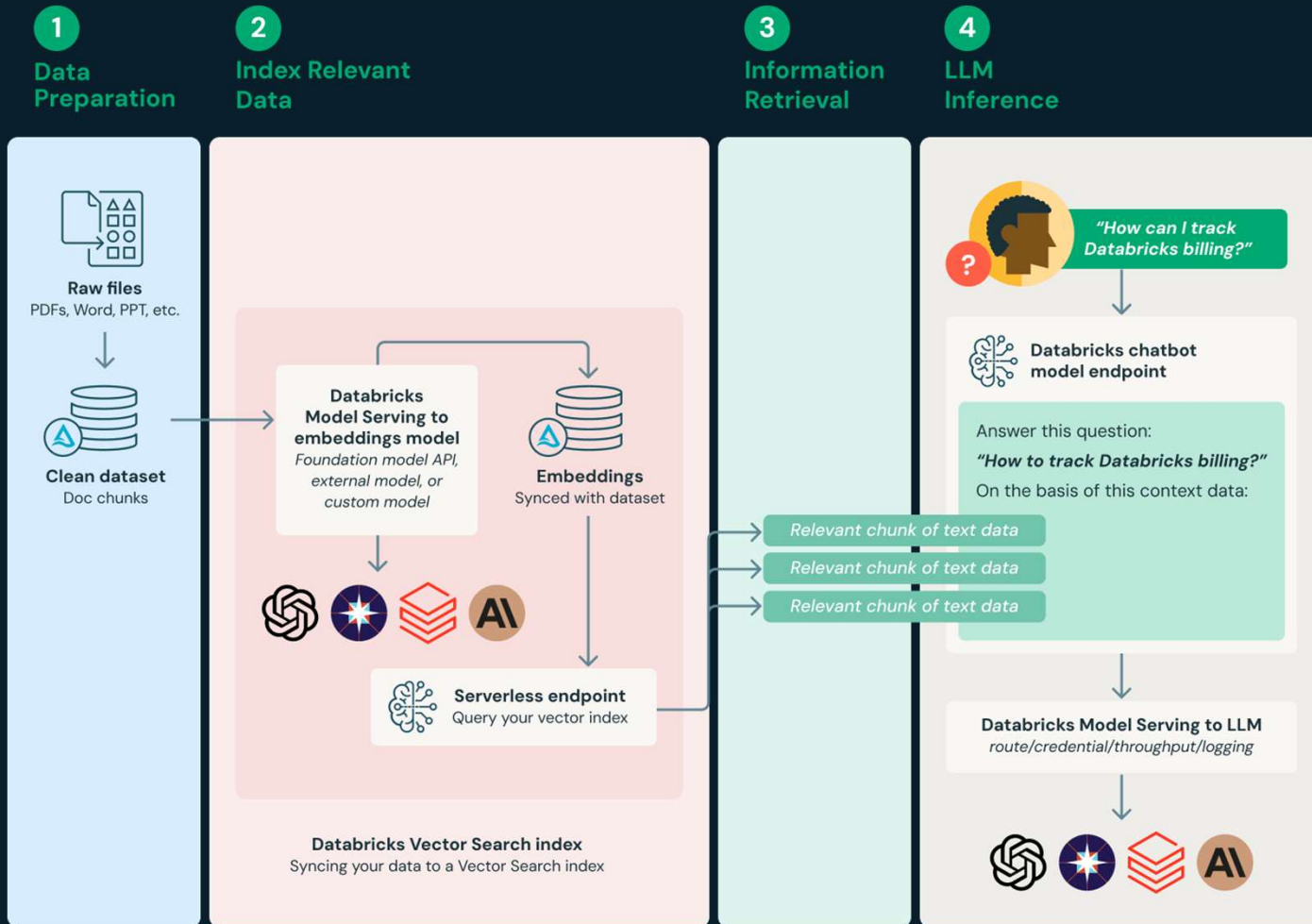


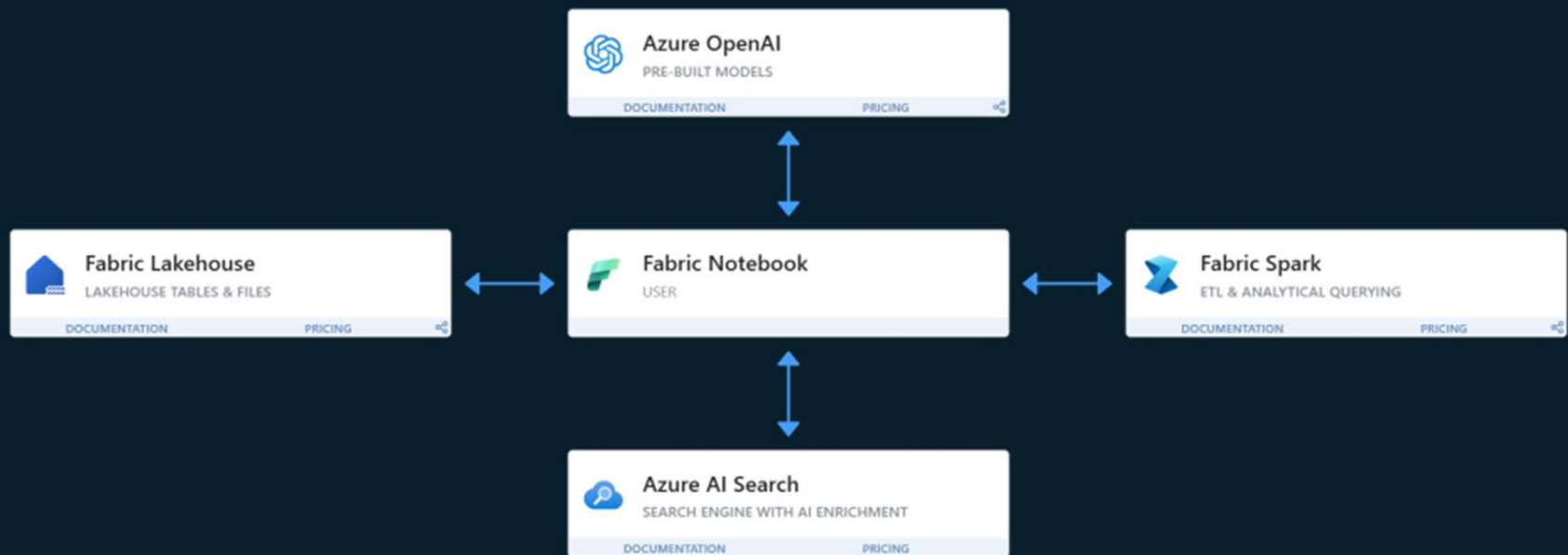
Architektura RAG (Retrieval-Augmented Generation)



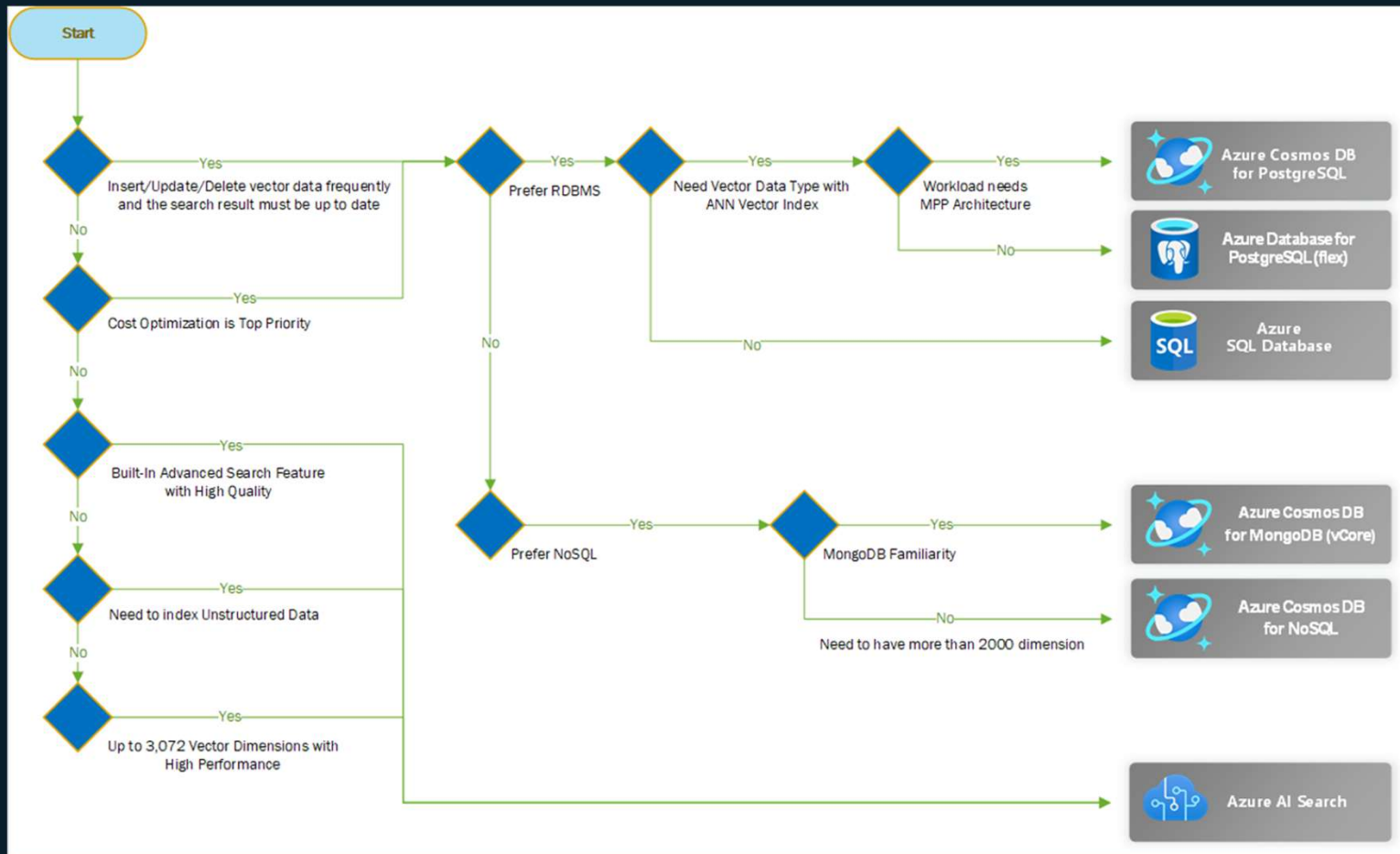
Przegląd architektur referencyjnych







Wymagania do rozważenia



Podstawowe cechy

Capability	Azure Cosmos DB for PostgreSQL	Azure Cosmos DB for NoSQL	Azure Cosmos DB for MongoDB (vCore)	Azure Database for PostgreSQL (Flex)	Azure AI Search	Azure SQL Database
Built-in vector search	Yes ¹	Yes	Yes ²	Yes ¹	Yes ³	Yes ⁴
Vector data type	Yes	Yes	Yes	Yes	Yes	Yes ⁹
Dimension limits ⁶	16,000 ⁷ or 2000	505 ⁸ or 4096	2,000	16,000 ⁷ or 2000	3,072	1998 (preview) ⁵
Multiple vector fields	Yes	Yes	No	Yes	Yes	Yes
Multiple vector indexes	Yes	Yes	No	Yes	Yes	Yes

1. "pgvector" supports vector search, which is the [extension of PostgreSQL](#).

2. [Use vector search on embeddings](#) in Azure Cosmos DB for MongoDB vCore

3. Vectors in Azure AI Search

4. Native vector search is available with Azure SQL Database [Early adopter preview](#) ²

5. Vectors can be stored in a *VARBINARY(8000) column or variable*.

6. Embedding models from OpenAI, 1536 for both text-embedding-ada-002 and text-embedding-3-small, and 3072 for text-embedding-3-large. For [Azure AI Vision multimodal embedding models](#), 1024 for both image and text.

7. Vectors can have up to [16,000 dimensions](#) ². But index using "IVFFlat" and "HNSW" supports vectors with up to 2,000 dimensions.

8. Vectors indexed with the flat index type can be at most 505 dimensions. Vectors indexed with the quantizedFlat or DiskANN index type can be at most 4,096 dimensions.

9. SQL Database [Vector Data Type](#)

Metody przeszukiwania

Search method	Azure Cosmos DB for PostgreSQL	Azure Cosmos DB for NoSQL	Azure Cosmos DB for MongoDB (vCore)	Azure Database for PostgreSQL (Flex)	Azure AI Search	Azure SQL Database
Full text search	Yes ¹	Yes ⁹	Yes ²	Yes ¹	Yes ³	Yes ⁴
Hybrid search	Yes ⁵	Yes ¹⁰	No	Yes ⁵	Yes ⁶	Yes ⁷
Built-in reranking	No	Yes ⁹	No	No	Yes ⁸	No

1. PostgreSQL [Full Text Search](#) ¹
2. [Search and query with text indexes](#) in Azure Cosmos DB for MongoDB vCore
3. [Get started with Full-Text Search](#)
4. [Vector data](#) on SQL Server
5. Not provided as a first-class feature but [sample codes](#) ⁵ are provided.
6. Hybrid search (combination of full text search, vector search, and semantic ranking) is provided as a first-class feature.
7. Hybrid search [example](#) ⁷ for Azure SQL database and SQL Server.
8. Reranking called [Semantic Ranking](#) is a first-class feature for reranking the result of full text search and/or vector search.
9. Cosmos DB NoSQL Full Text Search with full text scoring
10. Cosmos DB NoSQL Hybrid Search

Algorytmy indeksowania

Indexing approach	Azure Cosmos DB for PostgreSQL	Azure Cosmos DB for NoSQL	Azure Cosmos DB for MongoDB (vCore)	Azure Database for PostgreSQL (Flex)	Azure AI Search	Azure SQL Database
DiskANN	Yes	Yes	Yes (preview) ²	Yes ¹	No	No
Exhaustive K-nearest Neighbor (EKN)	Yes	Yes	Yes	Yes	Yes	Yes
Hierarchical Navigable Small World (HNSW)	Yes	No	Yes ²	Yes	Yes	No
IVFflat	Yes	No	Yes	Yes	No	No
Other	-	flat, quantizedFlat ³	Vector field limitation ⁴ Vector index limitation ⁵	-	-	External libraries are available ⁶

1. [DiskANN for Azure Database for PostgreSQL - Flexible Server](#)
2. [Azure Cosmos DB for MongoDB - Vector search overview](#)
3. [Vector indexing policies](#)
4. Only one vector field is available per container.
5. Only one vector index is available per container.
6. Index can be created with the aid of external libraries like [Scikit Learn](#) or [FAISS](#)

Azure AI Foundry

Announcing



Azure AI Foundry



Copilot Studio



Visual Studio



GitHub



Azure AI
Foundry SDK



Model Catalog

Foundational models

Open-source models

Task models

Industry models



Azure
OpenAI Service



Azure
AI Search



Azure AI
Agent Service



Azure AI
Content Safety

Evaluations

Customization

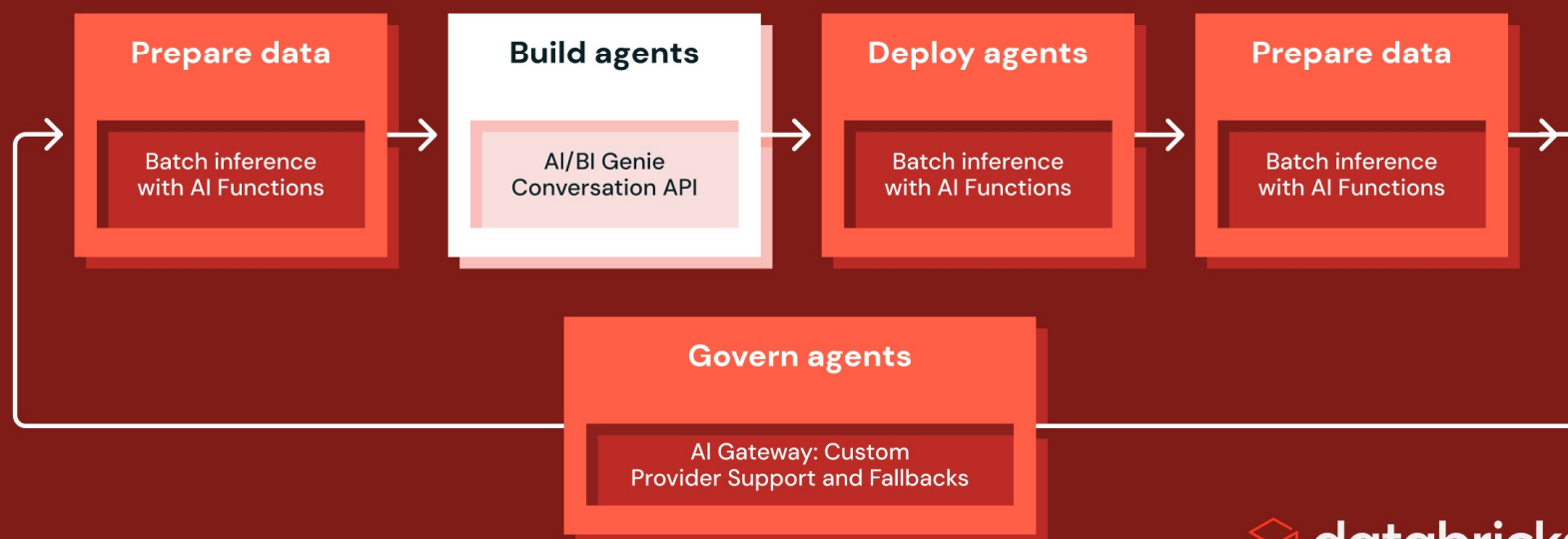
Governance

Monitoring

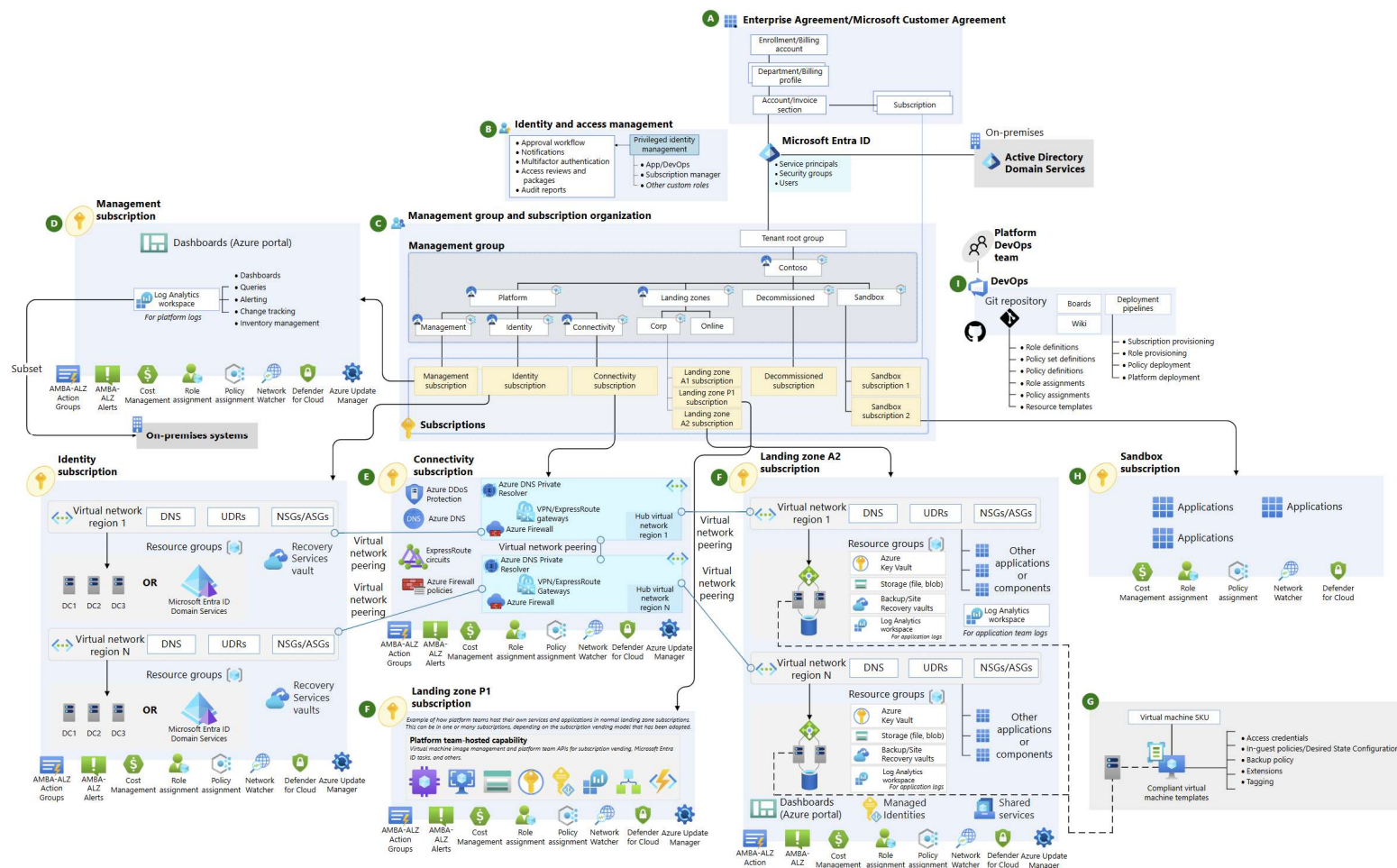
Observability

Databricks Genie

AI AGENTS: BUILD



Landing zone



Lab #1: Budowa pipeline dla danych ustrukturyzowanych i nieustrukturyzowanych

Lab #2: Tworzenie szkieletu Databota z Databricks Genie i Azure OpenAI

Lab #3: Fine-tuning modelu odpowiedzi, rejestracja w AI Foundry

Lab #4: Wdrożenie Databota z uwierzytelnieniem użytkowników

Lab #5: Symulacja zarządzania dostępem do Databota (role, zakresy, ograniczenia)

Demo: Microsoft Fabric AI Skills

TECHNOLOGY
INNOVATION
DATA
KNOWLEDGE

tidk

Pomagamy wydobyć wiedzę ukrytą
w Twoich danych.



Q&A

<https://forms.office.com/e/c7h7THWCF6>