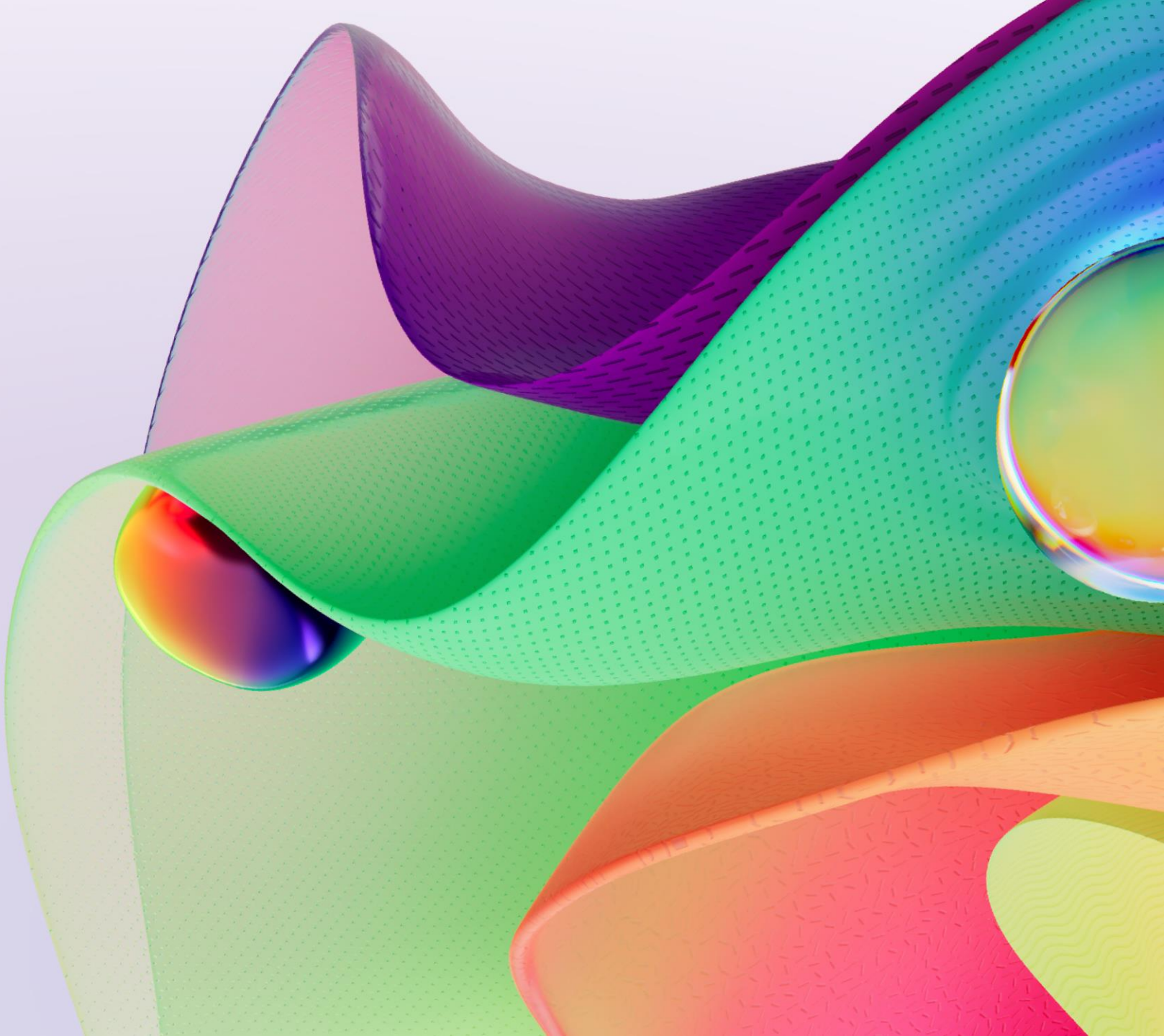
Microsoft

Microsoft AI Tour

In partnership with NVIDIA.



Microsoft

Building AI applications using Azure Cosmos DB and Azure Database for PostgreSQL

Khelan Modi
Product Manager

Agenda

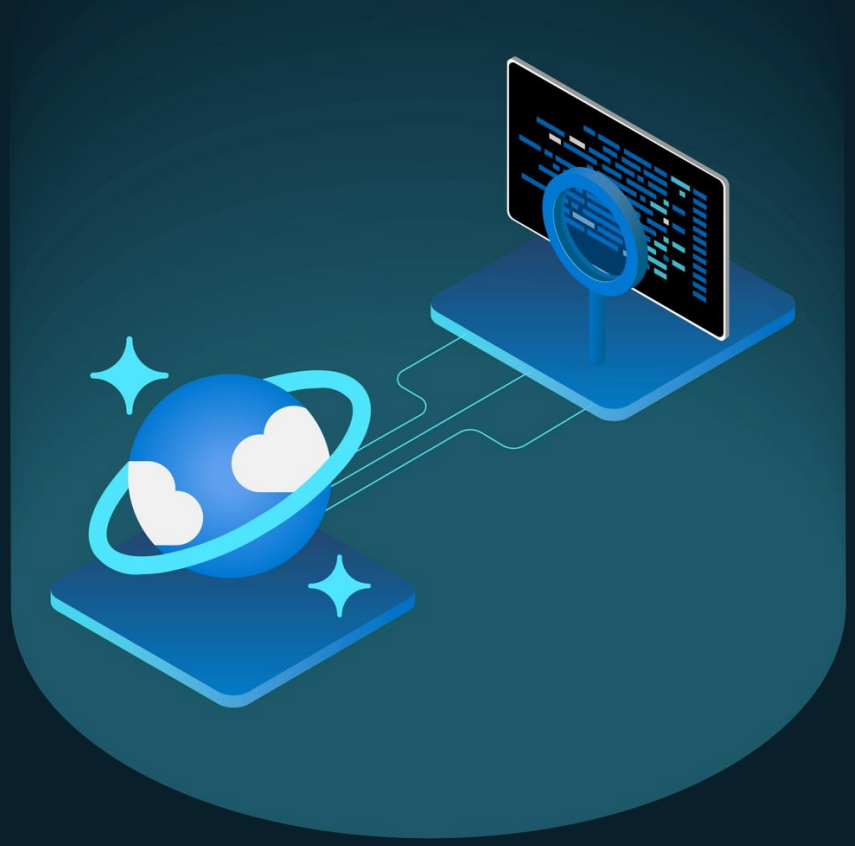
- Problem Statement
- Concepts
- Azure Cosmos DB for MongoDB vCore
- Copilot in Azure Cosmos DB
- Azure Databases for PostgreSQL Flex

Modern, intelligent applications have unique requirements

- Data is highly variable and unstructured
- Variable, high-volume traffic
- Fast, real-time, always-on digital experiences
- Globally-distributed users

AI ready databases in Azure

- All-in-one Solution: Transactional and vector database in ONE!!
- Save cost and complexity
- Real-time AI
- Highest fidelity with Azure Services
- Native Vector search



OpenAI is built on Azure Cosmos DB

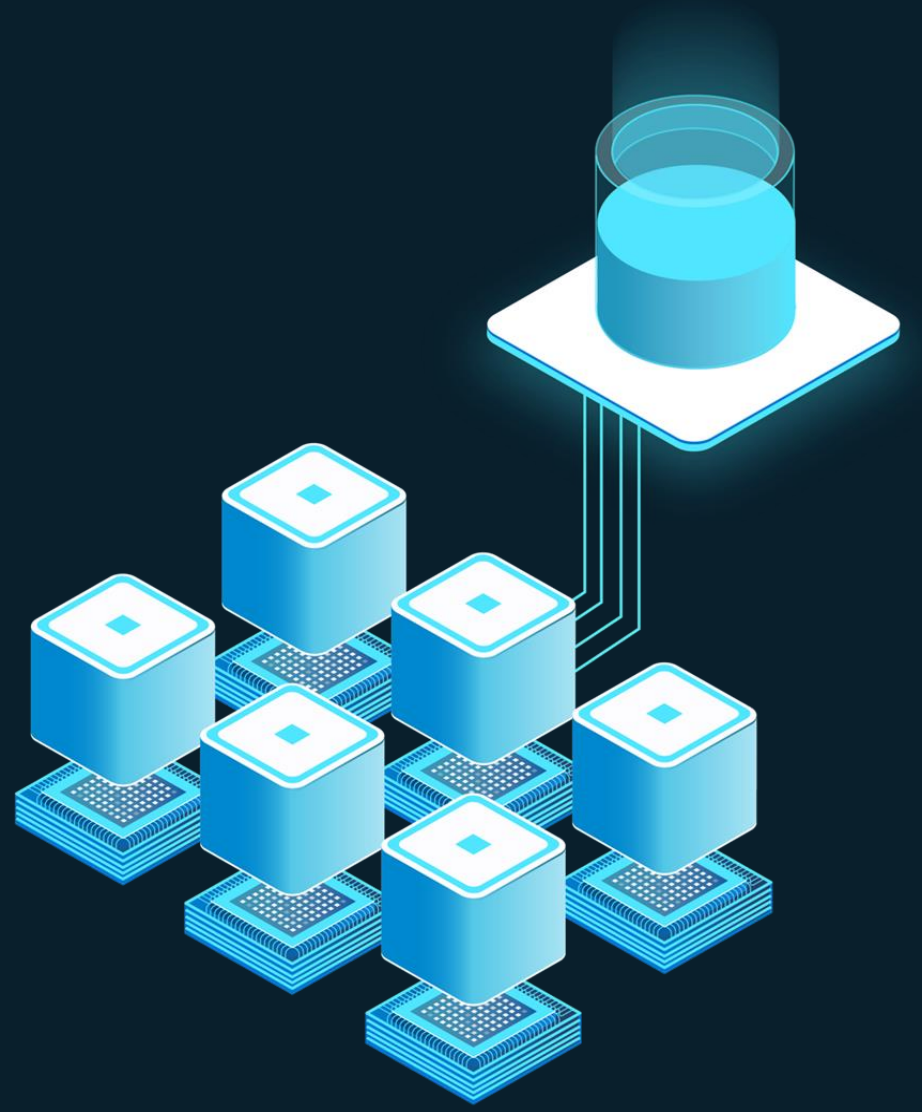
Your AI-powered apps can be too

Concepts

- Retrieval Augmented Generation (RAG)
- Vector Embeddings & Vector Search
- Vector Indexes: IVF & HNSW

Concepts – Retrieval Augmented Generation (RAG)

Retrieval Augmented Generation (RAG) intelligently retrieves a subset of data from data stores to provide specific, contextual knowledge to the large language model to support how it answers a user's prompt.



Concepts – Vector Embeddings

- **Vector embeddings** are compact, semantically-rich representations of any data
- Vectors that are “close” are semantically similar
- Closeness is measured by distance (cosine, dot product, Euclidean, etc.)
- Easy to generate embeddings from your data via APIs (OpenAI, Hugging Face, etc.)

Use cases

Answering Questions

Detecting anomalies

Making personalized recommendations

Searching for similar content

Vector indexes supported by Azure Cosmos DB and Azure Databases for PostgreSQL today

IVF (Inverted File Index)

- Partitions vectors into clusters and assigns each vector to one cluster.
- **Building the index is fast and memory-efficient**
- Requires a separate clustering step before indexing (slow)
- **Tuning parameters is important.** Can be very accurate if configured properly

HNSW (Hierarchical Navigable Small World)

- Builds a multi-layer graph with long and short connections between the vectors.
- **Robust and accurate at scale**
- No-preprocessing step.
- **Can support many inserts/deletes efficiently.**
- **Larger memory footprint**
- It also has many parameters (such as the number of layers and neighbors) that need to be tuned carefully.

Azure Cosmos DB for MongoDB vCore

New Additions

- Free tier w/ 32GB storage
- Burstable SKUs
- New cluster tiers & storage SKUs
- Private link
- Migration from MongoDB

AI Ready

- Native Vector Search, including HNSW
- Plugins: LangChain, Semantic Kernel, and LlamaIndex
- Integration with Azure OpenAI Studio

Learn more: aka.ms/tryvcore

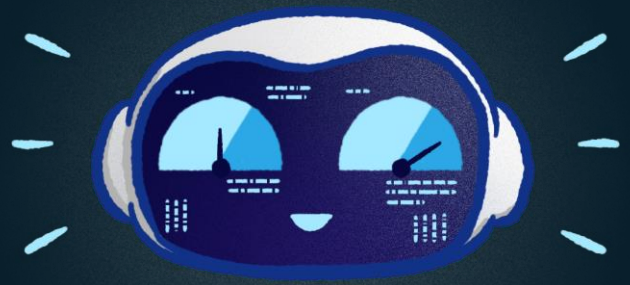
KPMG KymChat

AI agent to streamline KPMG employee operational tasks.

Leveraging Vector Search in Azure Cosmos DB for MongoDB vCore enabled KPMG to provide value to their employees at scale.

Accurate

PCI, a key relevancy metric increased from **50% to 90%+**



Performance

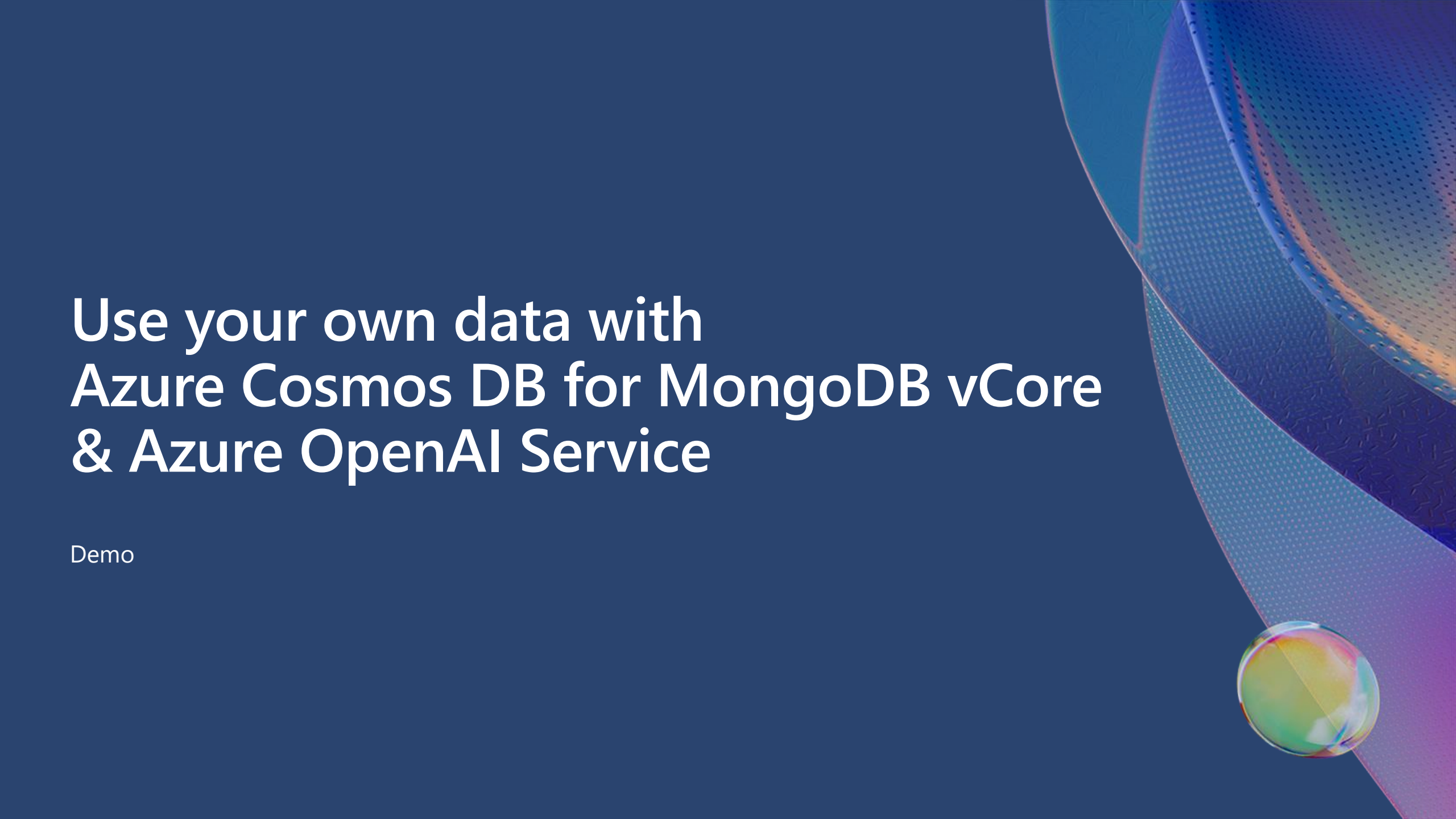
7,000+ employee issuing 120,000+ requests for up to 50% productivity gain

Scalable

Performance improvements enabled rollout to all KPMG member firms

Use your own data with Azure Cosmos DB for MongoDB vCore & Azure OpenAI Service

Demo



Microsoft Copilot for Azure

Enabling natural language queries for Azure Cosmos DB data

Turn your natural language questions into Cosmos DB NoSQL queries

Powered by state-of-the-art Azure OpenAI LLMs

Your data and usage is private and secure

Developed with Microsoft's Responsible AI principles

Copilot for Azure in Cosmos DB

Demo

Azure AI Advantage free offer

Up to $6,000 Azure Cosmos DB free for 90 days[1]

Eligibility: customers using Azure AI Services or GitHub Copilot

Why Azure Cosmos DB for Era of AI

AI ready

Guaranteed performance and scale

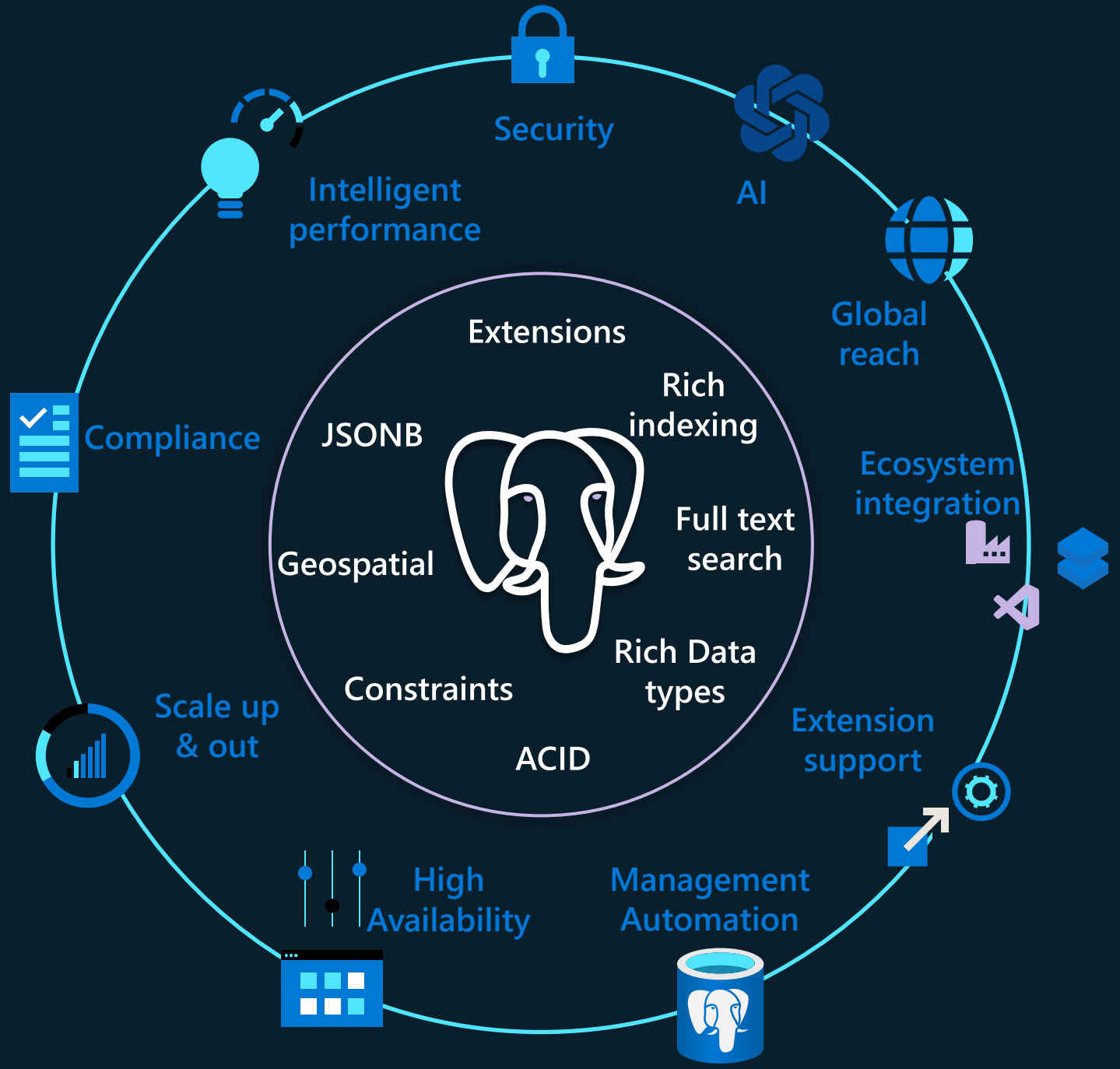
Flexibility and efficiency

Mission critical

Learn more: Aka.ms/AzureAIAdvantageBlog

**Azure AI Advantage Offer entitles customers to up to 40,000 Request Units per second for free for 90 days. This is the equivalent of up to $6,000 in savings.*

Azure Database for PostgreSQL: AI-Ready for Enterprise Applications



Azure Database for PostgreSQL—Intelligent apps

Azure AI extension

SQL Interface to Azure OpenAI

Create embeddings from SQL Statements

SQL interface to Azure AI Language services

Complimentary to vector data type

Vector data type

Pg Vector extension update—0.5.1 GA

Vector data type—natively store embeddings

Vector indexing for performant searches

Efficient similarity searches within the DB

Azure AI extension – Azure Open AI

```
SELECT title FROM recipes
order by
recipe_embedding <#>
azure_openai.create_embeddings
('Deployment', 'high protein
recipes')::vector
```

```
POST {Endpoint}/depolments/{deployment-id}/embeddings
{
  "operationId": "embeddings_create"
  ...
    "requestBody": {
        "input" : "High protein recipes.."
        }
    ....
  }
```

Azure OpenAI

Application

title	cosine_similarity_score
Chicken Bake(Low Calorie)	0.8125699032035647
Oatmeal breakfast bars	0.8116828687002345
Hot Chicken Salad	0.8066745034844333

```
[0.01665339,-0.00458135,-0.00493248,-0.00530701,
-0.00034506, 0.00672823,-0.01681390,-0.02811681,
-0.00896206,-0.02639127,-0.00998534, 0.02411732,
-0.02094715, 0.00103080, 0.00453119, 0.01050033,
 0.02910665, 0.03346729, 0.02641803,-0.03475141,
-0.01490110, 0.00411987,-0.01630561,-0.02398355,
-0.01812477,-0.00369518, 0.02870536,-0.02080001,
 0.04023566,-0.02189686, 0.00957737,-0.00053713,
```

Vector Extension

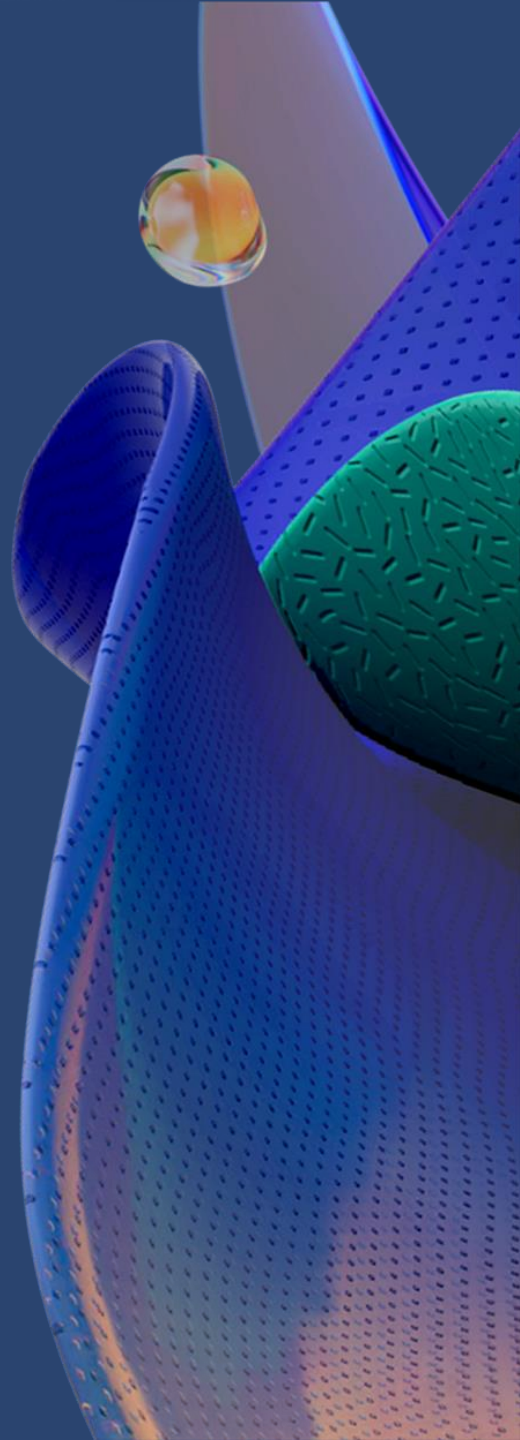
Efficient similarity searches

Vector

Search recipes with Azure Databases for PostgreSQL Flexible server

Demo

Thank you!



Learn More

Azure Cosmos DB for Mongo vCore Free tier:	**Aka.ms/tryvcore**
Vector Search & AI Assistant demo:	**Aka.ms/MongovCoreAzureAIsample**
Microsoft Copilot for Azure in Cosmos DB:	**Aka.ms/CopilotForAzureInAzureCDBBlog**
Azure AI Advantage:	**Aka.ms/AzureAIAdvantageBlog**
Azure Database for PostgreSQL - Flexible Server:	**Aka.ms/azurepgflex**
Sign-up for a Free account:	**Aka.ms/freeazurepostgres**
Azure Database for PostgreSQL Blog:	**Aka.ms/azurepostgresblog**

Microsoft

Get started skilling with AI on Microsoft Learn

Build AI skills, connect with the community, earn Microsoft Credentials, learn from experts, and take the Cloud Skills Challenge.

aka.ms/LearnAtAITour

UPCOMING SESSIONS:

Make your data AI ready with Microsoft Fabric 2:15 PM to 3:00 PM	Paul DeCarlo Level 2 – Room 2014
Get started with data science in Microsoft Fabric 2:15 PM to 3:30 PM	Patrick Chanezon, Graeme Malcolm Level 2 – Room 2006