Tiger Analytics

Tiger Analytics

Building the **world's best** AI and Analytics team

Background and Challenges

Tedious Information search

Analysts in the organization needs to go through a lot of documents to find relevant information

Over-dependence on IT team for Data access

Non-technical persons need to depend on the IT team to get data from the RDBMS or they need to learn SQL to get the data access

Summarization

Analysts in the organization needs to summarize lengthy documents in a specific formats, it's quite a tedious and consuming process

Data Analysis

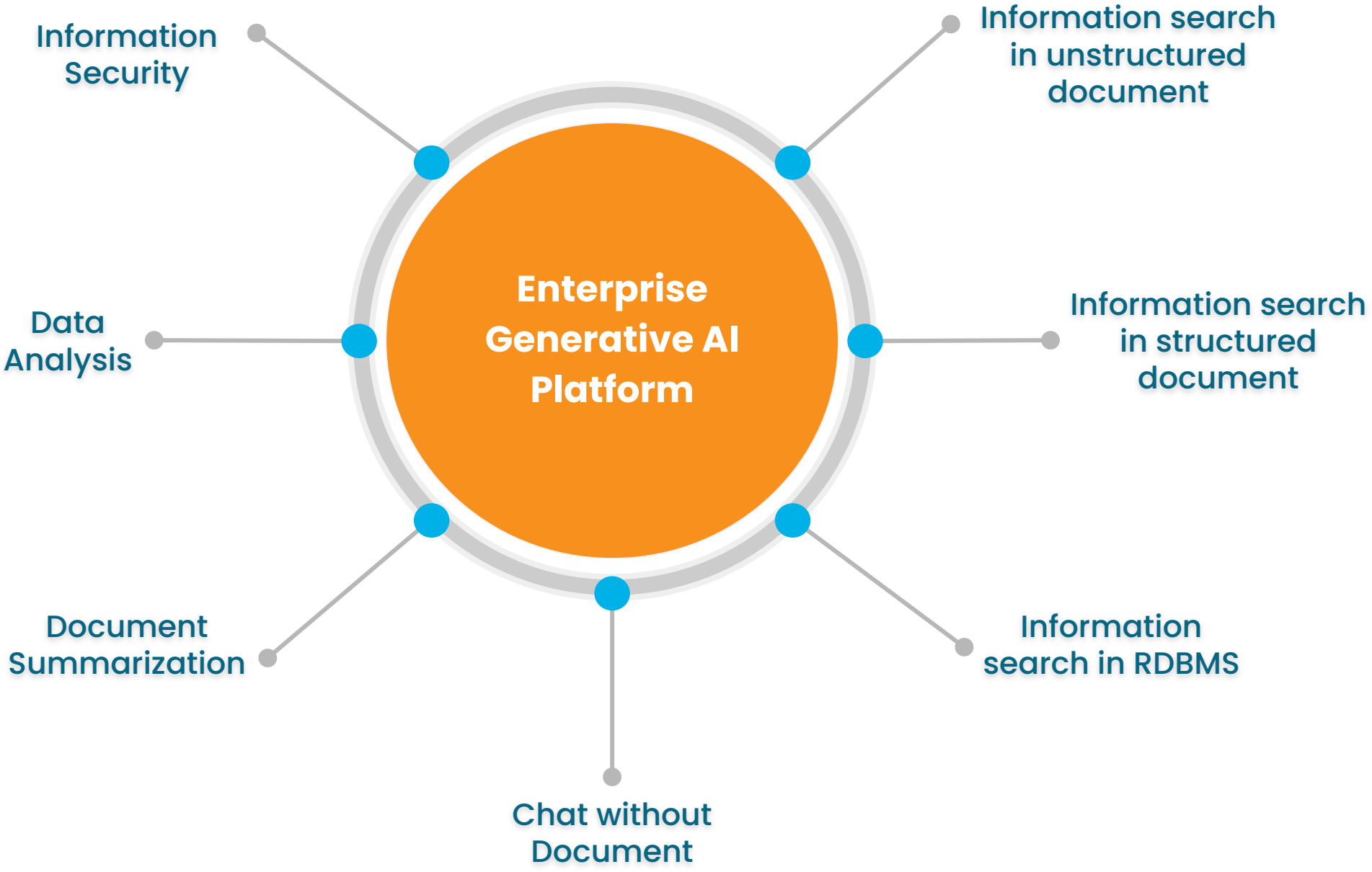
Non-technical persons in the organization need to depend on the organization for data analysis, data visualization. This creates bottlenecks in the decision-making process and hinders data driven decision making

Information Security

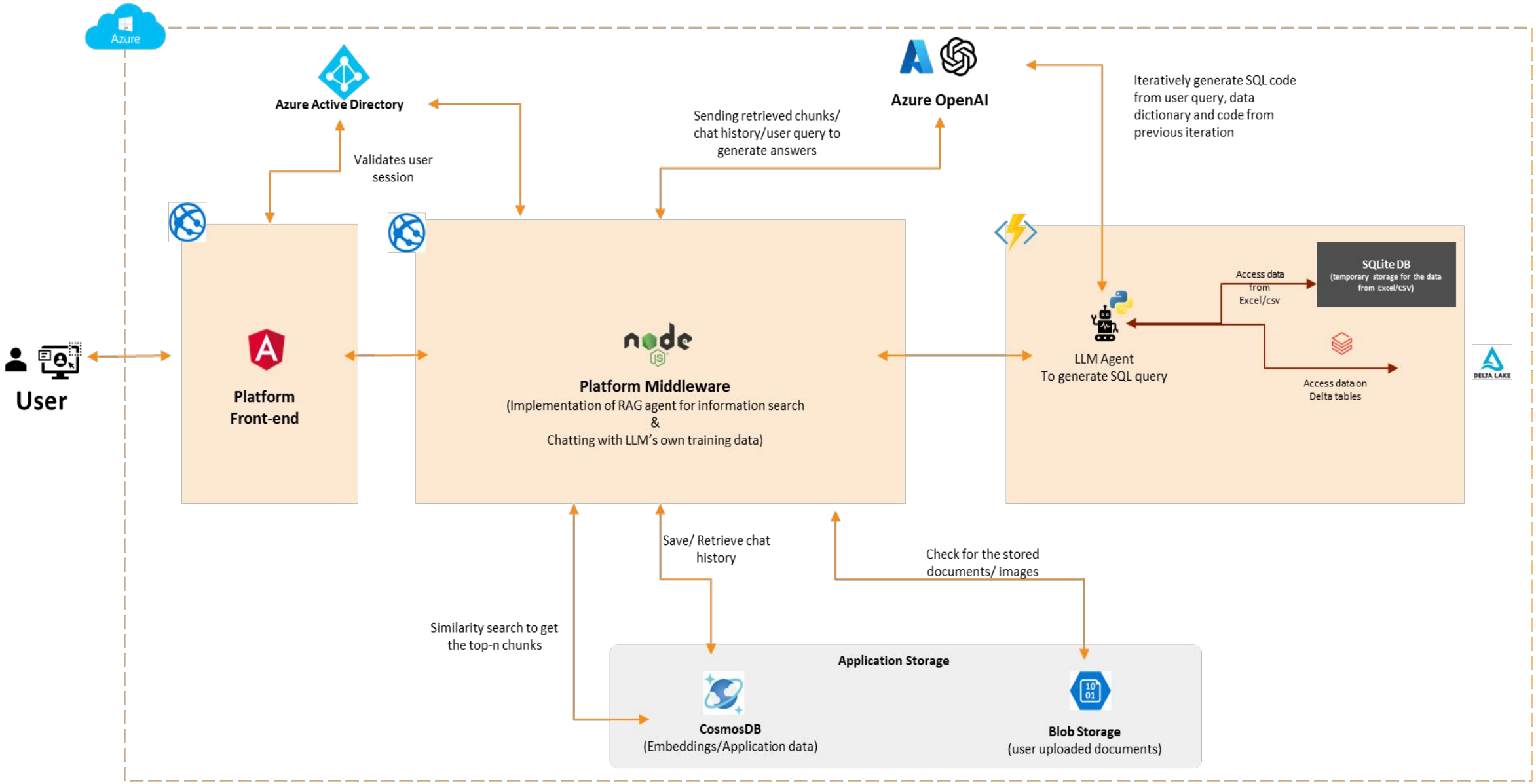
After the release of ChatGPT, employees are using ChatGPT to fetch information from certain text so they are sending these texts to OpenAI server, this might pose a security threat to the organization.

LLMs has revolutionized the way users find information and derive insights from the data. It has the potential to increase employee productivity by accelerating information search, democratizing data access, derive deep insights from data and task automation while ensuring proper information security and access control in an organizational setting.

Solution


Information Security
Information search in unstructured document
Data Analysis
Enterprise Generative AI Platform
Information search in structured document
Document Summarization
Information search in RDBMS
Chat without Document

Enterprise GenAI Platform | Overall Architecture



Azure Cosmos DB | Overcoming challenges of traditional On-Prem DB

- **Complexity in Schema Design**: Our information search platform integrates various search capabilities across diverse sources (unstructured docs, structured docs, images, SQL databases, etc.). The information to be stored in the database varies significantly depending on the source type. Creating a one-size-fits-all schema in this scenario is extremely challenging
- **High Latency:** As a conversational information search platform, latency directly impacts user experience and can hinder adoption. Therefore, taking effective strategies to lower the latency of the application is crucial
- **Scalability:** The application serves all employees of the organization and select partner organizations, presenting scalability challenges. Scaling the on-premise database requires meticulous planning, scheduled downtime, and substantial manual intervention, making it time-consuming and costly
- **Availability & Disaster Recovery:** This production-grade application must maintain availability across different time zones and requires robust disaster recovery measures
- **Information Security:** The information search platform contains sensitive organizational information, it's very important to implement robust security measures and keeping up with the ever-evolving security & compliance requirements can be complex, resource-intensive and time-consuming
- **Accurate Budgeting:** As a novel and innovative application, providing accurate budget estimates for the required infrastructure support (hardware, software licensing, and manpower) is extremely challenging. The budget heavily depends on the platform's adoption rate. If adoption falls short of expectations, a significant capital expenditure on extensive infrastructure could prove futile

Solving the challenges using Azure Cosmos DB

- **Flexible Schema:** Azure Cosmos DB provides a NoSQL database with a MongoDB API, allowing the flexibility to use the same Collection for different types of information search. The schema can evolve dynamically without requiring downtime or migrations
- **Multi-Model Support:** Cosmos DB supports multiple data models, including document, key-value, graph, and column-family. This allows us to store different types of data models in a single database
- **Query Performance:** Cosmos DB offers automatic query indexing and optimization, delivering single-digit millisecond latencies for reads and writes:
  - During write operations, data is split into small slices and written into different shards in parallel
  - For read operations, the wildcard index set by Cosmos DB on the dataset enables remarkably fast data retrieval
- **Scalability:** Cosmos DB provides automated elastic scalability for both storage and throughput. It allows resources to be scaled up and down without any downtime
- **High Availability:** Azure Cosmos DB provides 99.999% availability SLA for read operations and 99.99% for write operations
- Pay-as-you-go: Azure Cosmos DB's pay-as-you-go model allows for more flexible and scalable budgeting, aligning costs with actual usage and reducing the risk of over-investment in underutilized resources
- **Automated Fail-safe Procedure:** Azure Cosmos DB's automated multi-region replication provides failover support in case of regional outage
- **Managed Services:** Azure Cosmos DB abstracts away the complexities of infrastructure management, it does auto backups, automatic software updates, built-in monitoring & alerting etc. This reduces the operational overhead cost of hardware, software and man-power
- **Managed Security:** Azure Cosmos DB provides security features like encryption at rest and in transit along with compliance towards various industry standards (HIPAA, SOC, ISO, etc.) as per requirements