A 3D perspective view of a blue grid with glowing lines and rectangular blocks, creating a digital or data visualization background.

Building an AI-Driven Supplemental Reporting Tool for Conversational Queries Based on Internal Documents

Tiger Analytics partnered with the client to build the solution by training models on financial documents. This approach provided a solid foundation for developing an AI-driven supplemental reporting tool tailored to handling conversational queries and generating insights from internal documents.

The Background

Our client is a Real Estate Investment Trust (REIT), with services in warehouse leasing, property management, & logistics facilities. The Data & Analytics division of the customer engaged Tiger Analytics to build the solution by training models on financial documents like 10K, 10 Q, supplemental reports.

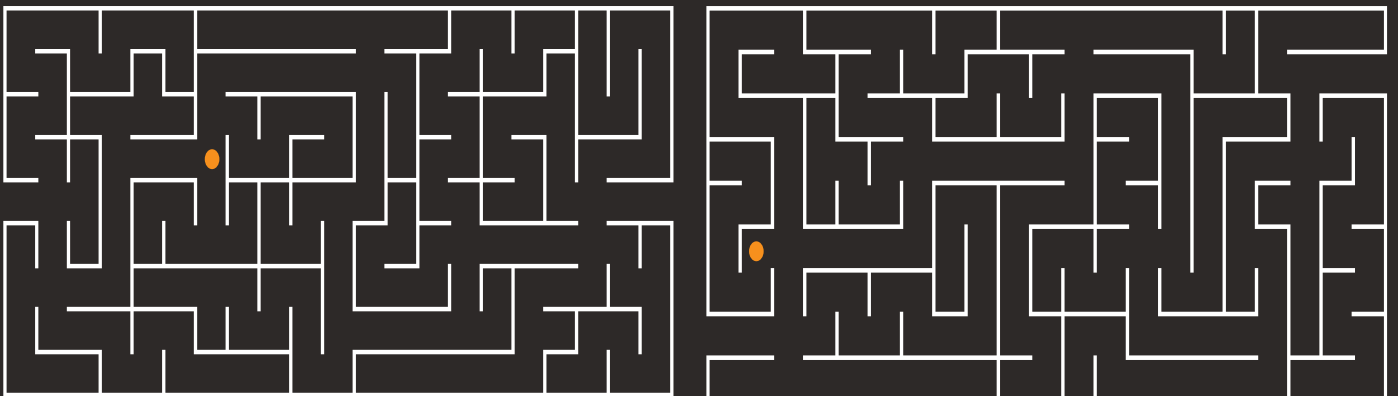
Key Challenges

/ Implementation Challenges:

- **Manual Extraction from Multiple Files:** The need to manually select specific pages from various documents to create a consistent data extraction pipeline, a process that has to be repeated for every new document.

/ Operational Challenges:

- **Handling Comparison Queries:** Addressing complex questions that require comparing data across multiple supplemental documents.
- **Extracting Data from Image-based Charts:** Many charts embedded as images in presentations cannot be automatically extracted, necessitating manual data entry.



Our Solutions

/ Data Extraction Techniques:

- AWS Textract Utilization: Employed AWS Textract to automate the extraction of text and tables from the supplemental PDFs.
- Manual Data Conversion: Converted data from charts embedded as images into tabular form manually to ensure all relevant information is captured.
- Setting up Pipeline : From reading the data to extracting and pushing to the vector database is done in SageMaker.

/ Data Processing:

- Chunking and Embedding Storage: The extracted data is segmented into smaller, manageable chunks, and the embeddings for these chunks are stored in the Pinecone vector index for efficient retrieval and querying.

/ Application development and deployment:

- RAG Model pipeline: Built a Retrieval-Augmented Generation (RAG) model using supplemental PDFs, including 10-K and 10-Q documents, for each quarter starting from Q1-2021.
- Development is done in lambda function and invoked through API gateway.
- Application UI Deployment is done in EKS.

Tech Stack

/ AWS Lambda

/ Amazon S3

/ AWS Textract

/ Pinecone Vector Database

/ Amazon EKS

/ AWS SageMaker

/ API gateway

Value Delivered

/ Enhanced Data Accessibility: Streamlined Access to Key Information in Supplemental Reports.

/ Improved Query Handling: Ability to Answer Complex Questions Across Multiple Quarters.

/ Improved Decision Making: Ability to make decision based on enterprise data.

/ Operational Efficiency: Automation of Data Extraction and Pipeline Creation.

/ Accuracy in Data Interpretation: Reliable Extraction and Transformation of Tabular Data from PDFs and Images.

The top half of the image features a complex, abstract pattern of overlapping, rounded geometric shapes in various shades of blue and purple. The shapes are layered, creating a sense of depth and movement. The colors range from light, airy blues to deep, rich purples and dark blues. The overall effect is modern and tech-oriented.

About Tiger Analytics

Tiger Analytics is a global leader in AI and analytics, helping Fortune 1000 companies solve their toughest challenges. We offer full-stack AI and analytics services & solutions to help businesses achieve real outcomes and value at scale. We are on a mission to push the boundaries of what AI and analytics can do to help enterprises navigate uncertainty and move forward decisively. Our purpose is to **provide certainty to shape a better tomorrow.**

Being a recipient of multiple industry awards and recognitions, we have 4000+ technologists and consultants, working from multiple cities in 5 continents.

www.tigeranalytics.com