cloudGeometry

Case Study: Streamlining Whistleblowing Compliance and Actionability with Generative AI-enabled Platform

EXECUTIVE SUMMARY

FaceUp, a secure and anonymous whistleblowing platform, partnered with CloudGeometry, an AWS Advanced Consulting Partner, to leverage the power of Generative AI and enhance its services. This collaboration utilized AWS cloud technologies to design and implement an AI-driven solution capable of automating critical processes, including report summarization, generating follow-up questions, and providing actionable recommendations for administrators. The Proof of Concept (PoC) achieved remarkable success, streamlining operations, improving scalability, and significantly enhancing user experience while adhering to stringent privacy and data security regulations.

CloudGeometry played a crucial role as an AWS Advanced Consulting Partner by aligning FaceUp's business needs with the most appropriate AWS services, ensuring the solution was efficient, scalable, and cost-effective. The partnership highlighted the transformative potential of cloud-based AI technologies for solving complex real-world challenges.

INTRODUCTION

FaceUp is a cutting-edge platform designed to promote transparency and accountability within organizations. By offering a secure and anonymous channel for reporting unethical or illegal activities, the platform empowers employees and students to address issues such as bullying, harassment, and other forms of misconduct effectively.

As the platform's user base expanded, customers encountered significant challenges in scaling their operations. Manual report processing and a lack of clarity in some of the reports (with no option for iteration as the reporting is done in an anonymized mode) became major bottlenecks, leading to delays and inefficiencies. These limitations highlighted the need for a solution capable of managing increasing data volumes without sacrificing speed, accuracy, or security while simultaneously creating a competitive edge for FaceUp.

To address these challenges, FaceUp partnered with CloudGeometry to harness the potential of Generative AI. By integrating advanced machine learning models hosted on AWS, the collaboration aimed to automate repetitive tasks, enhance decision-making processes, and deliver an improved experience for both administrators and whistleblowers.

NEED FOR AI IN WHISTLEBLOWING PLATFORMS

Traditional whistleblowing systems are labor-intensive, requiring administrators to manually sift through large volumes of reports, identify actionable insights, and generate follow-up questions. These processes are time-consuming and prone to human error, particularly as the scale of operations increases.

The integration of Generative AI addresses these inefficiencies by automating key tasks:

- **Automated Report Summarization**: Quickly extracting and presenting key points from lengthy reports to administrators, reducing the time required for initial assessments.
- **Follow-Up Question Generation:** Suggesting contextually relevant questions to gather additional information from whistleblowers, ensuring thorough case resolution.
- **Actionable Recommendations:** Providing administrators with specific next steps, tailored to the nature of each case, based on historical data and best practices.

This not only enhances operational efficiency but also ensures a consistent and high-quality user experience for all stakeholders involved.

SYSTEM ARCHITECTURE OVERVIEW

The PoC leveraged a combination of AWS services and cloud-native design principles to create a robust, scalable, and secure architecture.

Data Storage and Indexing

- **Amazon S3:** Used as the primary storage layer for whistleblowing reports, summaries, and recommendations. The service ensured durability and scalability for storing large datasets.
- *Amazon OpenSearch:* Enabled efficient vector storage and retrieval, crucial for implementing the RAG (Retrieval-Augmented Generation) architecture. This allowed the AI system to quickly access relevant documents and provide contextually accurate responses.

AI Model Integration

AWS Bedrock: Hosted the Claude 3 Haiku model, which powered the Generative AI capabilities of the solution. Bedrock ensured the flexibility to test, deploy, and scale the AI model without the need for extensive infrastructure management.

Retrieval-Augmented Generation (RAG): Combined generative AI with vector search to enhance the relevance and accuracy of AI-generated responses.

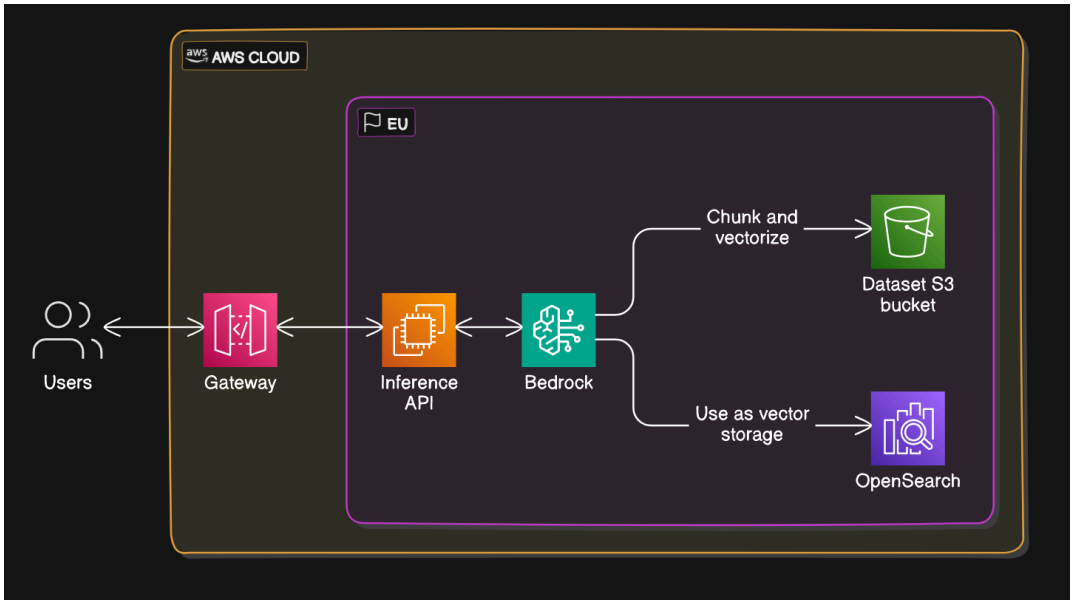

FIGURE 1: General architecture with Amazon Bedrock

Deployment Framework

AWS Lambda: Managed data ingestion and indexing pipelines, ensuring real-time processing and minimal latency.

Amazon ECS (Elastic Container Service): Deployed the application on GPU-enabled instances, providing the computational power needed for processing complex AI workloads.

IMPLEMENTATION DETAILS

Data Collection and Preparation

- Reports from FaceUp's existing database were collected, anonymized, and categorized. This ensured compliance with regulations like GDPR and CCPA while maintaining the quality of training data for AI models.
- Additional datasets, including best practices and regulatory documents, were indexed to enhance the knowledge base used by the RAG architecture.

AI Model Deployment

- AWS Bedrock's Claude 3 Haiku model was configured with optimized prompts to generate high-quality outputs. The model was trained to handle domain-specific tasks, such as generating follow-up questions and actionable recommendations.

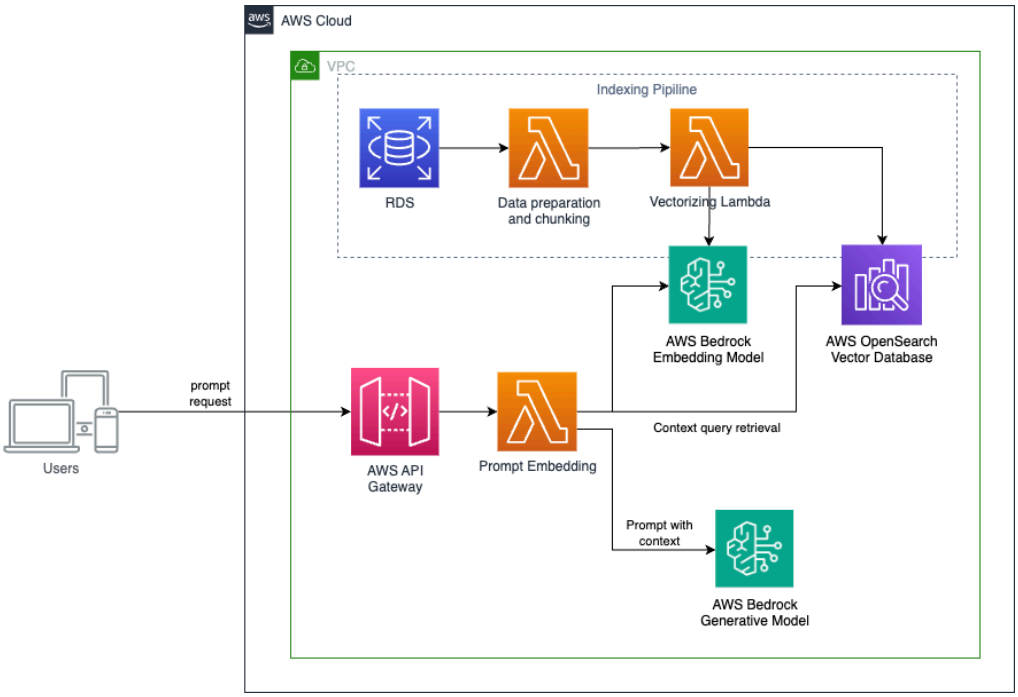

FIGURE 2: AI Model Deployment

RAG Architecture Setup

- By combining document retrieval with AI inference, the RAG architecture enabled the AI system to dynamically update its knowledge base without retraining, reducing costs and improving scalability.

Deployment

- The application was deployed using AWS SAM, ensuring a seamless setup process. ECS clusters with auto-scaling capabilities were used to handle varying workloads efficiently.

RESULTS AND FINDINGS

The PoC achieved significant milestones:

- **Performance**: Follow-up questions were generated within ~3 seconds, meeting the desired response time for real-time applications.
- **Scalability**: The solution demonstrated the ability to handle up to 1,500 reports concurrently without degradation in performance.
- **Cost Efficiency**: Total costs for processing 1,500 reports were estimated at $1,687, showcasing the financial viability of the solution.

CHALLENGES AND SOLUTIONS

1. **Privacy and Compliance**
    - Solution: Implemented end-to-end encryption and strict anonymization protocols to safeguard sensitive data.
2. **Multi-Language Support**
    - Solution: Tested the multilingual capabilities of Claude 3 Haiku, with plans to extend support to additional languages post-PoC.
3. **Scalability and Latency**
    - Solution: Deployed AWS resources in multiple regions to reduce latency and ensure a consistent experience for global users.

BENEFITS OF THE SYSTEM

- **Efficiency Gains**: Reduced administrator workload by automating repetitive tasks, with a ~30% improvement in report processing time.
- **Enhanced User Experience**: AI-assisted insights provided actionable, contextually relevant information to administrators.
- **Operational Scalability**: The cloud-native design ensured the platform could scale to meet growing user demands seamlessly.

FUTURE DIRECTIONS

1. **Multi-Language Support**
    - Extend the Claude 3 Haiku model implementation to support additional languages like Czech, German, and French.
2. **Advanced Data Analysis**
    - Develop capabilities to identify trends, root causes, and patterns across historical reports for improved decision-making.
3. **Improved User Interfaces**
    - Introduce a feedback loop for administrators to continuously refine
    - AI-generated outputs.

CONCLUSION

The collaboration between FaceUp, CloudGeometry, and AWS demonstrated the immense potential of Generative AI in transforming whistleblowing platforms. By automating critical processes, the PoC enhanced operational efficiency, reduced costs, and delivered a superior user experience. With AWS services providing the foundation, the solution is poised for further innovation, ensuring scalability, compliance, and excellence in whistleblowing practices.

