



CASE STUDY

AI Voice Assistant for Real Estate Agent Support Using Amazon Bedrock and Amazon Nova Sonic

Mactores built a low-latency, cost-efficient AI voice assistant for eXp Realty using Amazon Bedrock and Amazon Nova Sonic to automate repetitive agent support calls. The solution enabled real-time voice interactions, significantly reduced per-call costs, and delivered 24/7 instant responses for a globally distributed agent workforce.

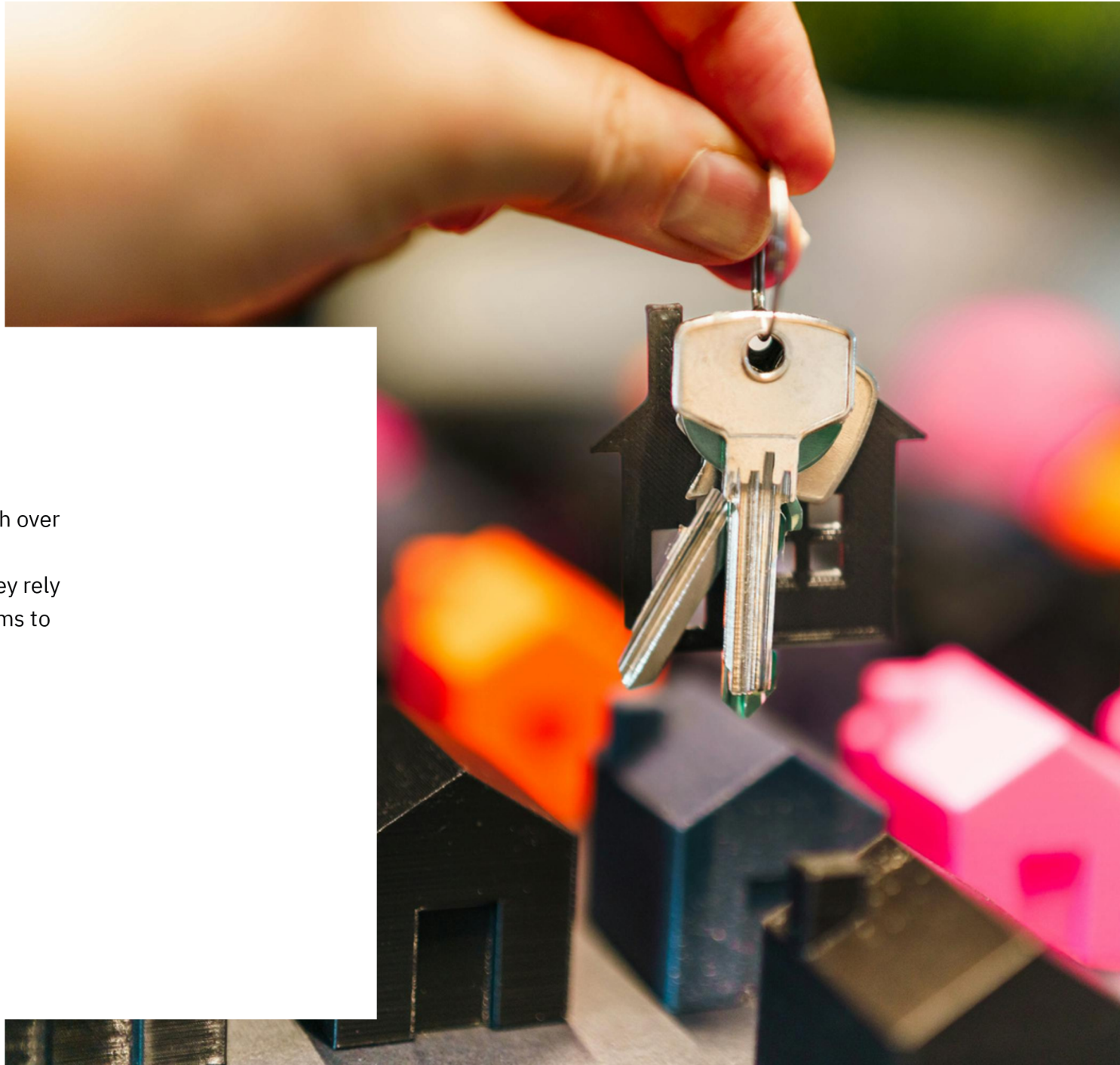
Case Study Summary

- Built a real-time AI voice assistant for large-scale agent support using Amazon Bedrock and Amazon Bedrock AgentCore to orchestrate conversational reasoning and task execution.
- Leveraged Amazon Nova Sonic on Amazon Bedrock to enable native speech-to-speech interactions, eliminating traditional STT/TTS pipelines and achieving ~300ms end-to-end latency.
- Implemented serverless request orchestration with AWS Lambda, enabling scalable, event-driven handling of concurrent voice interactions.
- Integrated historical call recordings and structured knowledge assets stored in Amazon S3 to provide context-aware, accurate responses.
- Delivered full operational observability using Amazon CloudWatch for latency tracking, invocation monitoring, and error detection.

About The Customer



eXp Realty is a cloud-native real estate brokerage with over 85,000 agents maintaining a global presence. Unlike traditional brokerages, eXp has no physical office. They rely entirely on digital platforms and virtual support systems to serve their agent network at scale.



Customer Situation

eXp Realty faced a growing operational challenge as its agent base scaled rapidly. Hundreds of agents regularly contacted support teams with repetitive questions related to onboarding, policies, commissions, tools, and workflows. Human-led call centers struggled with delayed response times, rising operational costs, and limited availability outside standard business hours. The organization needed a solution that could deliver instant, accurate, and conversational voice responses that mimic human conversations while significantly reducing support costs.

Traditional AI voice solutions based on modular pipelines, speech-to-text, LLM reasoning, and text-to-speech introduced additional latency, architectural complexity, and higher costs. eXp required a more streamlined, scalable approach that could operate reliably at enterprise scale.



Our Approach

Mactores designed a unified, serverless AI voice assistant using Amazon Bedrock as the core platform for inference, security, and cost governance. After evaluating multiple voice and language model options, Amazon Nova Sonic was selected for its native speech-to-speech capabilities, enabling natural voice interactions without separate STT and TTS components and significantly improving conversational fluidity.

The assistant was implemented as an agentic system using Amazon Bedrock AgentCore, orchestrating conversational reasoning, retrieval, and response generation. A Retrieval-Augmented Generation (RAG) architecture enabled the agent to access over two million tokens of knowledge from eXp’s CODA workspace and historical call data, ensuring grounded and accurate responses.

AWS Lambda provided elastic, event-driven execution to support high concurrency. Historical support call recordings and curated knowledge content were stored in Amazon S3. Amazon CloudWatch provided full observability into latency, invocation metrics, and error rates, ensuring performance reliability and production readiness while minimizing operational overhead.

Business Outcomes

The AI voice assistant modernized eXp Realty’s agent support by delivering instant, human-like responses at a fraction of traditional costs. Per-call expenses dropped from **\$50** to approximately **\$2**, while response times improved from minutes to near real-time. The solution enabled **24/7** global support without added headcount, improved agent satisfaction through natural conversations, and scaled seamlessly to support a growing, distributed workforce—creating a high-ROI, sustainable support model.

Technical Outcomes

The solution delivered a production-ready AI voice assistant using Amazon Nova Sonic on Amazon Bedrock, achieving **~300ms** speech-to-speech latency while eliminating separate STT/TTS pipelines. Amazon Bedrock AgentCore orchestrated conversational reasoning and retrieval, AWS Lambda enabled elastic execution, Amazon S3 grounded responses in historical data, and Amazon CloudWatch ensured operational observability and reliability at scale.

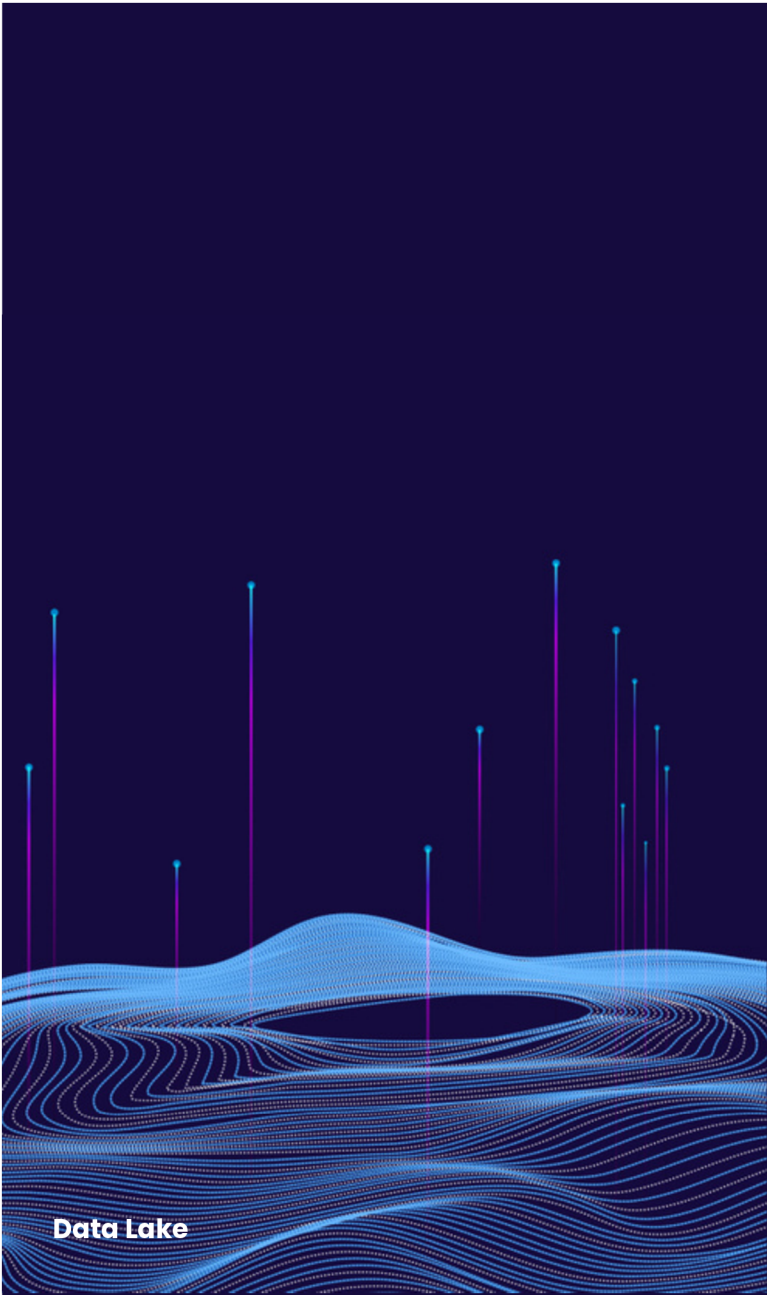


Getting Started

Mactores began with discovery sessions to understand eXp's support workflows, call patterns, and scalability requirements. Multiple voice and LLM options were benchmarked for latency, accuracy, and cost. Amazon Nova Sonic emerged as the optimal choice due to its real-time performance and simplified architecture.

The team designed and implemented a working prototype within **6 weeks**, validating business value through live demos and cost modeling. Mactores supported end-to-end delivery, from architecture design and model selection to demo execution and performance validation, laying the foundation for production-scale deployment.

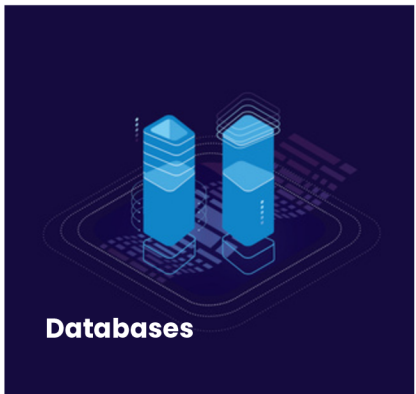
Our
Solutions



Data Lake



Data Lake



Databases



Data Ops



Generative AI

Our **Process**

Digital transformation via assessments, migration or modernization

We work alongside your tech team to assess and strategize what you need and how to implement the right data solutions on time, on budget and with c-suite buy in.



Assess

- Discovery Automation
- Future State Assessment
- GAP Analysis
- End State
- Road Map
- TCO



Migrate

- Strategy
- Execution
- Migrate
- Migration Acceleration



Modernize

- StrategyFuture State
- Design
- Build
- Automate

AWS Validated Competencies



AWS Validated Service Deliveries



www.mactores.com/lets-talk

Mactores Cognition Inc

2018 156th Ave NE

Bellevue, WA 98007