

Case Study

Real-Time Voice AI Platform & Managed Telephony Cloud Services

How AceCloud powers Servetel real-time conversational AI at scale
Managed Services Since 3 March 2025 — Ongoing

CLIENT	INDUSTRY	SERVICE
Servetel	Cloud Telephony / SaaS Communications	Managed Voice AI & Cloud Operations

KEY OUTCOMES AT A GLANCE

< 95ms	↑ 4x	99.95%	↓ 50%	Zero
Avg voice AI response latency	Avg voice AI response latency	Platform uptime	MTTR reduction	Model-related regressions

BUSINESS CHALLENGES

The Technical Hurdles Behind Real-Time Voice AI

Servetel, a fast-growing cloud telephony provider, sought to embed real-time conversational AI into its voice call platform to differentiate its offering and deliver intelligent call automation for enterprise customers. The challenges were significant:

- **Ultra-Low Latency Requirement:** Real-time voice AI demands sub-100ms response times. Any latency in infrastructure, session management, or model invocation would perceptibly degrade call quality.

- **High-concurrency scale:** The platform needed to support thousands of simultaneous AI-assisted calls — requiring parallel management of WebSocket connections, session state, and AI inference at scale.
- **Stateful sessions at scale:** Conversation context had to be preserved across a distributed, stateless backend without adding latency — a technically demanding requirement under high concurrency.
- **Mission-critical availability:** The telephony platform operates 24x7 with zero tolerance for downtime, as any degradation directly impacts Servetel's enterprise customers and their end-users.
- **Operational complexity:** Managing a real-time AI voice stack — model updates, infrastructure scaling, incident response — required a dedicated managed services partner. Servetel's team did not have internally.

SOLUTION DELIVERED

Solution Delivered Using AceCloud Managed Services

AceCloud operates Servetel's real-time voice AI platform as a fully managed service — covering real-time infrastructure, AI model operations, session layer management, and 24x7 network operations:

- **Real-time infrastructure operations:** AceCloud manages the Lambda-based streaming backend, API Gateway WebSocket configurations, and auto-scaling policies — ensuring the platform absorbs call volume spikes without manual intervention and maintains sub-100ms processing latency.
- **Session state management:** AceCloud operates the Amazon ElastiCache (Redis) cluster — including sizing, replication, eviction policies, and failover configuration — delivering sub-millisecond state lookups across all concurrent call sessions.
- **Voice AI model operations:** AceCloud manages Servetel's Amazon Nova Sonic integration via Amazon Bedrock, handling configuration, latency tuning, and response quality monitoring. LLM orchestration for complex reasoning tasks (Nova Pro and GPT-class models via Bedrock) is managed by AceCloud with circuit-breaker and fallback configurations to ensure resilience. All model version changes are benchmarked by AceCloud before production promotion.
- **24x7 NOC operations:** AceCloud runs a continuous network operations function monitoring call success rates, AI response latency, session health, and infrastructure metrics. P1 incidents are responded to within 10 minutes, with root-cause analysis and post-mortem documentation provided to Servetel.
- **Security operations:** End-to-end security management — IAM policies, TLS configuration, API Gateway authorization, CloudTrail logging, and periodic vulnerability assessments — is handled by AceCloud.

OUTCOMES

Measurable Impact Delivered by AceCloud

- < 95ms average end-to-end voice AI response latency maintained consistently under production load, meeting Servetel's call quality SLA requirements.
- ↑ 4x increase in concurrent AI-assisted call capacity since AceCloud took over managed operations, enabled by optimized auto-scaling and ElastiCache tuning.
- 99.95% platform uptime achieved over the managed services engagement period, supported by AceCloud's 24x7 NOC and defined incident response procedures.
- ↓ 50% reduction in mean time to resolution (MTTR) for infrastructure incidents following AceCloud's adoption of structured P1/P2/P3 escalation procedures.
- Zero model-related call quality regressions in production — a direct result of AceCloud's mandatory benchmarking process before every model change.

Conclusion

AceCloud's managed services partnership enables Servetel to offer enterprise customers a reliable, low-latency, AI-powered telephony experience — without the burden of operating a complex real-time AI stack internally. With AceCloud managing every layer of the platform around the clock, Servetel's engineering team focuses on building new product capabilities while AceCloud ensures the underlying infrastructure and AI models perform consistently at scale.

About Us

15+	10+	100+
Years of Exp	Data Center	Awards

AceCloud is a leading provider of end-to-end cloud computing solutions to global organizations across industries at scale. It offers a full spectrum of cloud services, including Public Cloud, Application Hosting, AWS Services, Managed Security Services, and Hosted Virtual Desktop Solutions.

AceCloud is a brand of Real Time Data Services (RTDS) group of companies - a leading provider of information technology capabilities specializing in Cloud Computing and Cloud Telephony. RTDS Group has an employee base of 600+ across India, the US, and the UK, supporting over 20,000+ customers and IT infrastructure in 10+ data centers spanning the globe. It offers industry-leading technological solutions that help customers streamline their operations and enhance efficiency.

As an advanced consulting and certified AWS Well-Architected Partner, AceCloud recently achieved the prestigious AWS Storage Competency and is also recognized as an Amazon RDS Service Delivery Partner. These achievements underscore its deep expertise in AWS technologies and commitment to delivering high-quality cloud solutions.



- Public Sector
- Storage Consulting Competency
- Small and Medium Business Services Competency
- Amazon RDS Delivery
- AWS Database Migration Service Delivery
- Amazon CloudFront Delivery
- AWS Well-Architected Partner Program
- Authorized Commercial Reseller



Contact Us

Are you ready to cut down your cloud spending? Contact AceCloud Experts today to get a customized quote.

awspartner@acecloud.ai

Mail to

<https://acecloud.ai/aws/>

Website

Our Location

2069 Oneill Dr, Bethel Park,
Pennsylvania, 15102-6602

2637 E Atlantic Blvd #23519
Pompano Beach, FL 33062,
Florida

Ace Tower, 809-A, Udyog
Vihar, Phase 5, Gurugram
122016, Haryana, India