OpsVoice Assistant

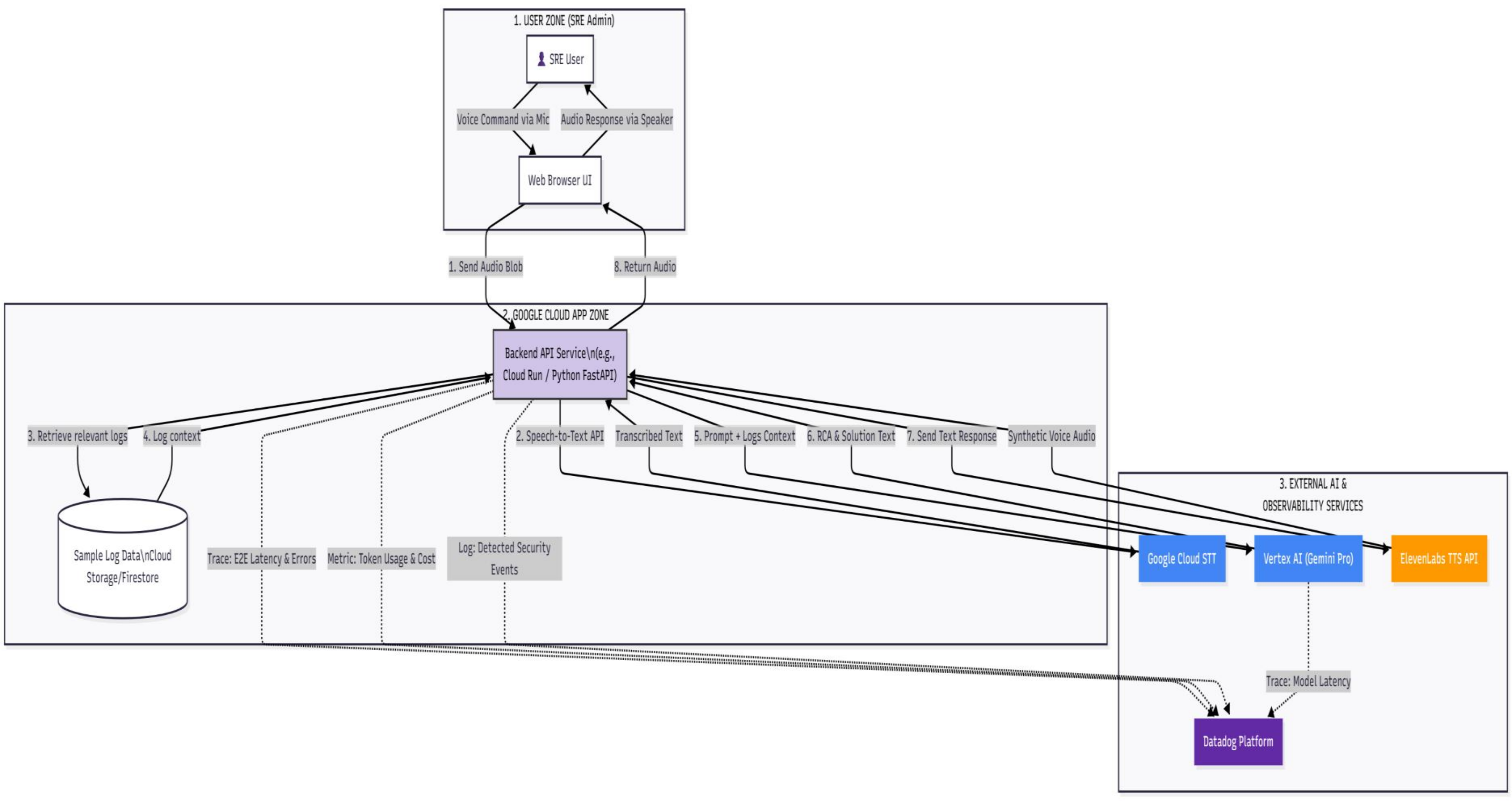
VOICE-ENABLED OBSERVABILITY FOR AI LATENCY NIRAJ – SENIOR SOFTWARE ENGINEER & PRODUCT BUILDER

PROBLEM

- AI backends suffer from hidden latency blind spots
- Traditional monitoring tools miss AI-layer delays
- Slow responses → poor user retention & lost revenue

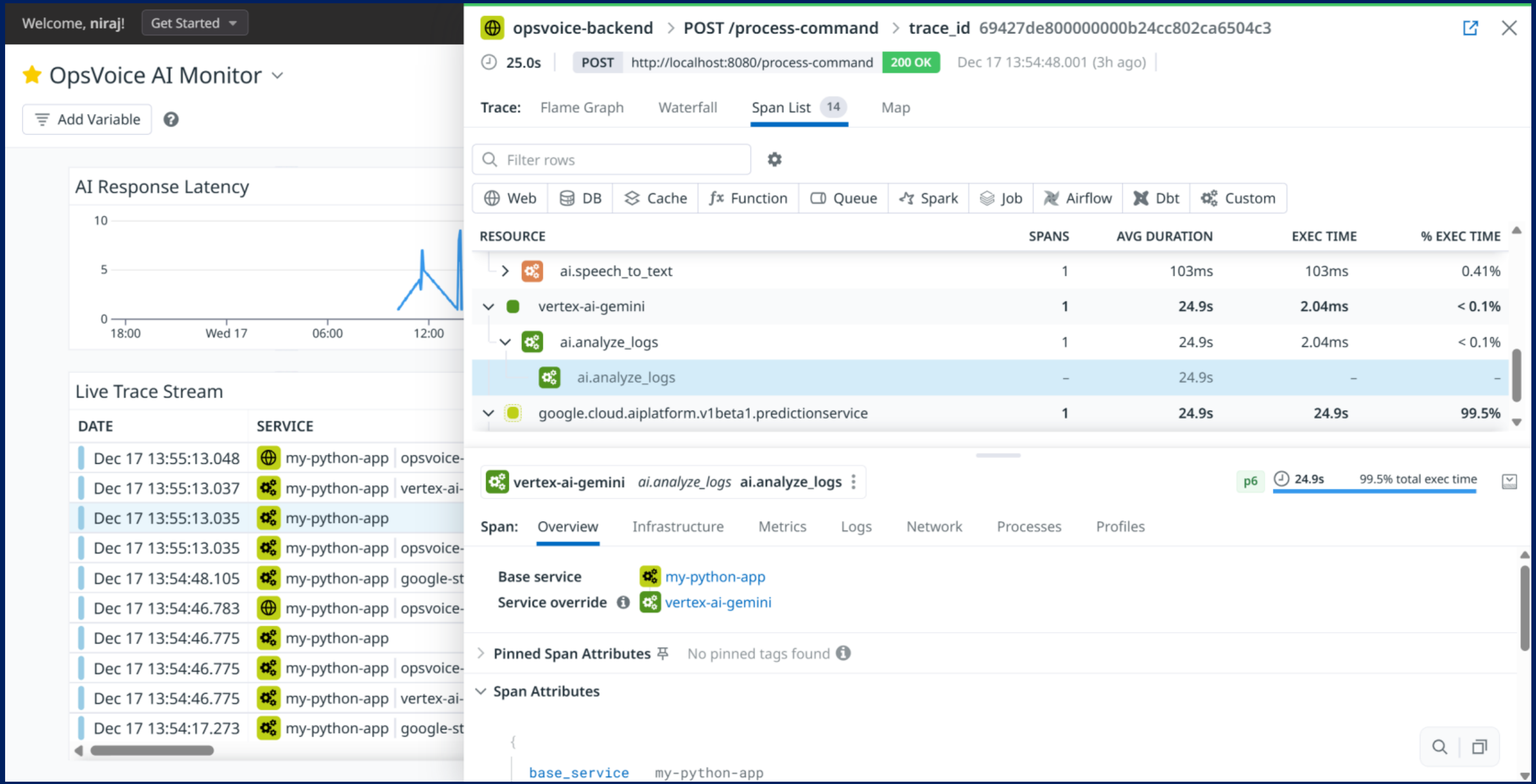
SOLUTION

- •CLOUD RUN DEPLOYMENT FOR SCALABLE, SERVERLESS INFRA
- •FLASK BACKEND WITH DATADOG TRACING & LOGGING
- •VERTEX AI GEMINI INTEGRATION FOR REAL LLM ANALYSIS
- •VOICE-ENABLED QUERIES FOR REAL-TIME OBSERVABILITY



DEMO VALUE

- DETECTED A 33-SECOND LATENCY ANOMALY IN GEMINI INFERENCE
- PINPOINTED ROOT CAUSE AT AI LAYER VS NETWORK LAYER
- DASHBOARDS SHOW LATENCY, ERROR RATES, RESOLUTION TIME
- JUDGE-FRIENDLY, REPRODUCIBLE DEMO ASSETS



★ OpsVoice AI Monitor

AI Response Latency

Live Trace Stream

DATE	SERVICE
Dec 17 13:55:13.048	my-python-app | opsvoice-
Dec 17 13:55:13.037	my-python-app | vertex-ai-
Dec 17 13:55:13.035	my-python-app
Dec 17 13:55:13.035	my-python-app | opsvoice-
Dec 17 13:54:48.105	my-python-app | google-st
Dec 17 13:54:46.783	my-python-app | opsvoice-
Dec 17 13:54:46.775	my-python-app
Dec 17 13:54:46.775	my-python-app | opsvoice-
Dec 17 13:54:46.775	my-python-app | vertex-ai-
Dec 17 13:54:17.273	my-python-app | google-st

opsvoice-backend > POST /process-command > trace_id 69427de800000000b24cc802ca6504c3

25.0s | POST http://localhost:8080/process-command 200 OK Dec 17 13:54:48.001 (3h ago) |

RESOURCE	SPANS	AVG DURATION	EXEC TIME	% EXEC TIME
ai.speech_to_text	1	103ms	103ms	0.41%
vertex-ai-gemini	1	24.9s	2.04ms	< 0.1%
ai.analyze_logs	1	24.9s	2.04ms	< 0.1%
ai.analyze_logs	–	24.9s	–	–
google.cloud.aiplatform.v1beta1.predictionservice	1	24.9s	24.9s	99.5%

vertex-ai-gemini *ai.analyze_logs* ai.analyze_logs

p6 24.9s 99.5% total exec time

Base service my-python-app

Service override vertex-ai-gemini

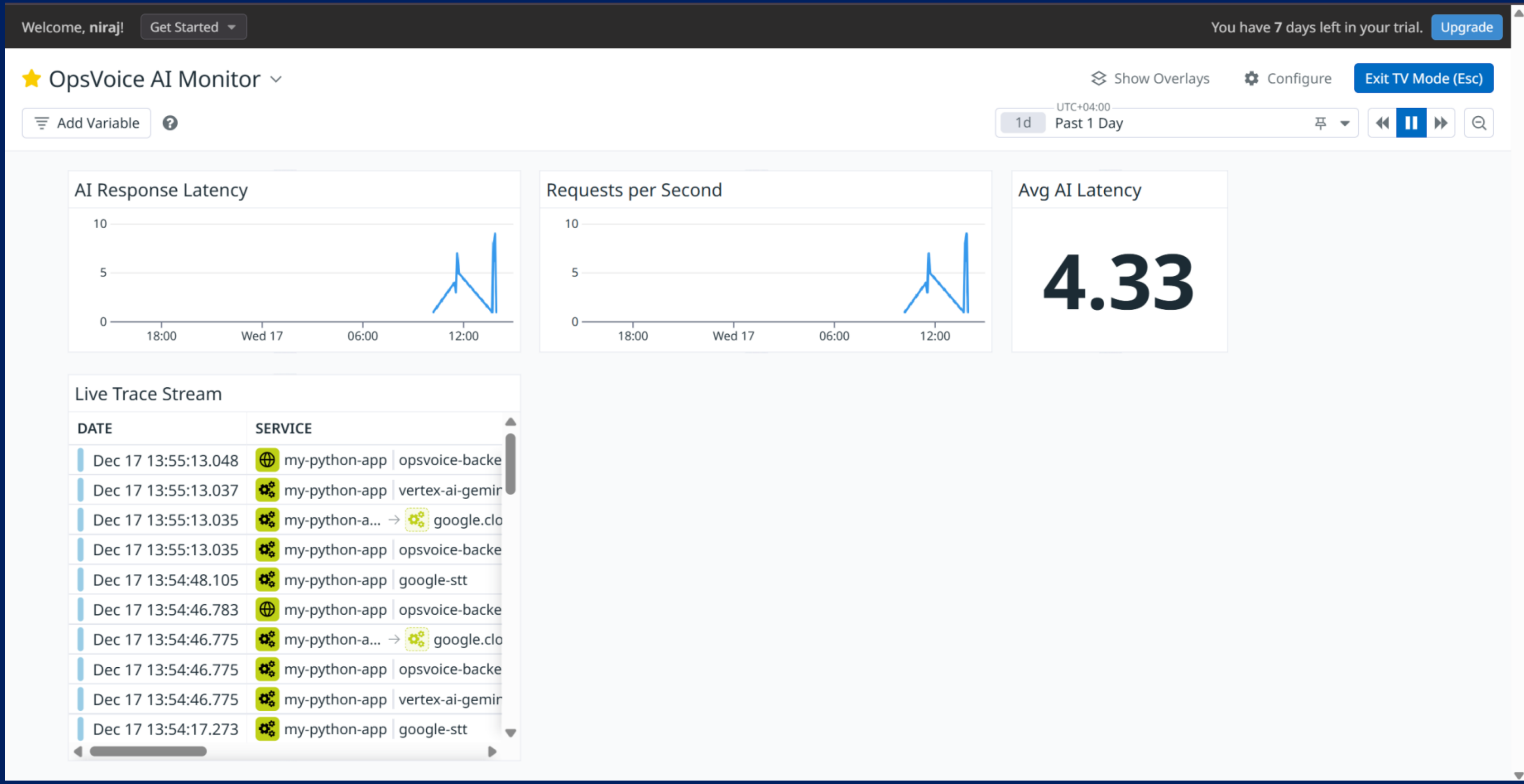
Pinned Span Attributes No pinned tags found

Span Attributes

```
{
  base_service   my-python-app
```

Market Opportunity

- AI OBSERVABILITY MARKET: $20B+ BY 2028
- ENTERPRISES ADOPTING LLMS AT 40% CAGR
- OPSVOICE BRIDGES THE CRITICAL AI-LAYER VISIBILITY GAP



DATE	SERVICE
Dec 17 13:55:13.048	my-python-app | opsvoice-backe
Dec 17 13:55:13.037	my-python-app | vertex-ai-gemir
Dec 17 13:55:13.035	my-python-a... → google.clo
Dec 17 13:55:13.035	my-python-app | opsvoice-backe
Dec 17 13:54:48.105	my-python-app | google-stt
Dec 17 13:54:46.783	my-python-app | opsvoice-backe
Dec 17 13:54:46.775	my-python-a... → google.clo
Dec 17 13:54:46.775	my-python-app | opsvoice-backe
Dec 17 13:54:46.775	my-python-app | vertex-ai-gemir
Dec 17 13:54:17.273	my-python-app | google-stt

BUSINESS MODEL

- SAAS SUBSCRIPTION FOR ENTERPRISE SRE TEAMS
- USAGE-BASED PRICING TIED TO TOKENS PER SECOND
- FUTURE: COST-PER-QUERY ANALYTICS FOR BUDGET CONTROL

COMPETITIVE EDGE

- COMPETITORS: DATADOG, NEW RELIC, ELASTIC
- DIFFERENTIATOR: AI-NATIVE OBSERVABILITY + VOICE INTERFACE
- FASTER SETUP, HACKATHON-PROVEN, REPRODUCIBLE DASHBOARDS

TEAM

- NIRAJ – BACKEND & OBSERVABILITY EXPERT (PYTHON, FLASK, DATADOG, CLOUD RUN)
- HACKATHON-PROVEN DELIVERY MINDSET
- FUTURE HIRES: SALES, DEVREL, PRODUCT

ASK

- SEEKING $500K SEED FUNDING
- FUNDS FOR INFRA SCALING, GTM, AND ENTERPRISE PILOTS
- GOAL: ONBOARD 10 ENTERPRISE CUSTOMERS IN 18 MONTHS

VISION

“STOP GUESSING. START OBSERVING AI PERFORMANCE IN REAL TIME.”
CONTACT: NIKY.SWAY@GMAIL.COM