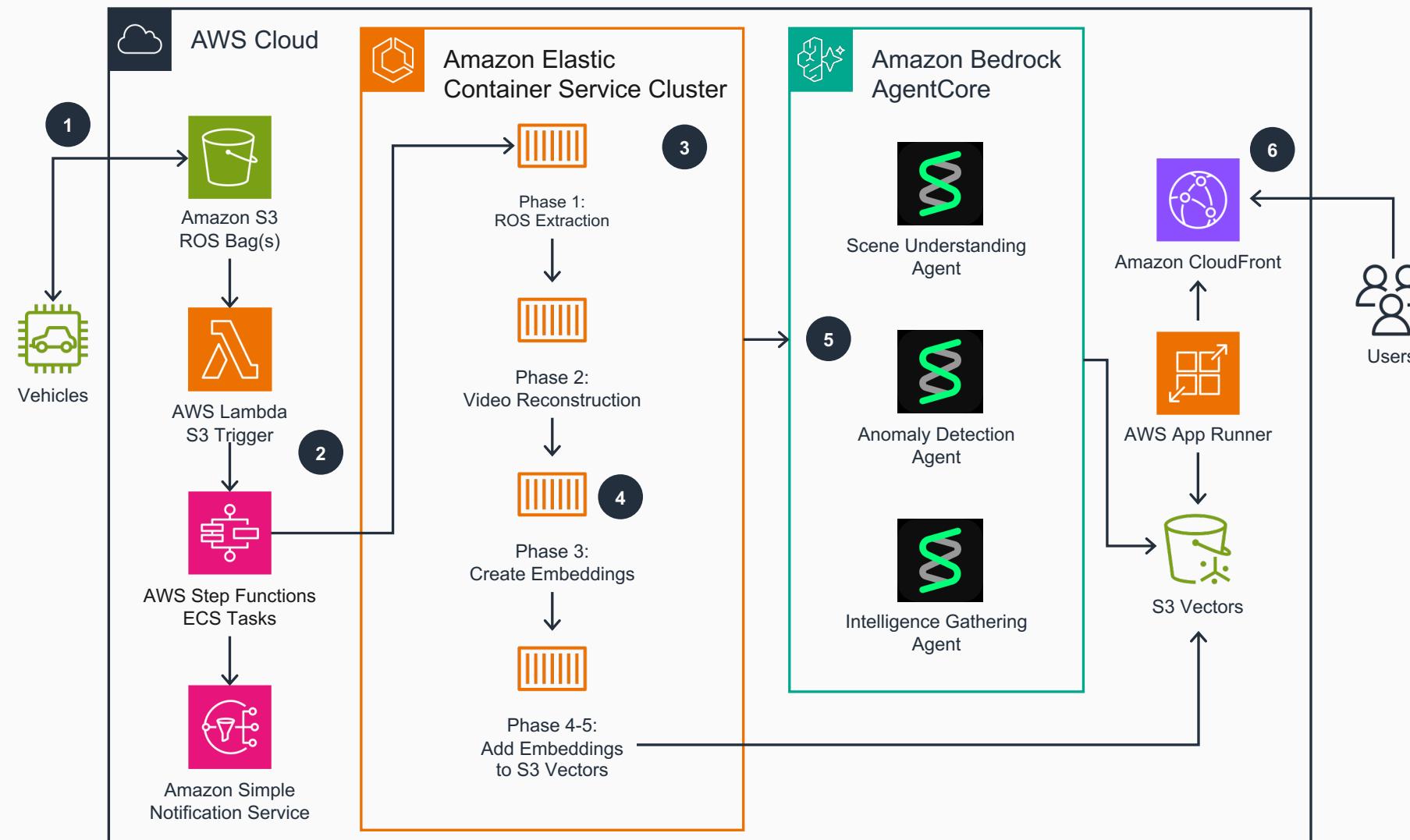


Multi-Agent AI Platform for Vehicle Fleet Data Discovery and Edge Case Classification

An agentic pipeline that ingests raw sensor data, auto-labels it using AI-driven behavioral analysis, and enables semantic search to surface high-value scenarios - including edge cases that weren't predefined. Multiple specialized agents iteratively analyze data using fleet-wide context, reducing blind batch processing and letting teams find what matters without writing rules for it first.



1 Vehicles upload ROS Bag files directly to S3. These files often contain: multi-sensor data streams (camera feeds, lidar, radar), compressed JPEG image sequences, vehicle telemetry and algorithm configuration metadata, and timestamped sensor synchronization data.

2 S3 ObjectCreated events trigger a Lambda function that initiates a Step Functions state machine. The pipeline uses WAIT_FOR_TASK_TOKEN integration for durable, asynchronous processing with error handling and immediate failure routing to SNS/SQS dead letter queues

3 A single ECS cluster manages two Auto Scaling Groups: Graviton for cost-optimized CPU workloads (Phases 1-2, 4-6), and GPU exclusively for transformer model inference in Phase 3. This right-size hardware to workload approach optimizes cost-performance.

4 InternVideo 2.5 and NVIDIA Cosmos Embed models analyze synchronized camera feeds to generate quantified behavioral metrics (following distance, lane positioning, risk scores). Full precision inference ensures accuracy for regulatory validation. Results are converted to 1024-dimensional embeddings and stored in S3 Vectors for semantic search.

5 Three specialized AgentCore agents coordinate via Strands SDK using Coordinator→Workers→Aggregator pattern. Pydantic schemas enforce structured outputs and anti-hallucination validation.

6 A JavaScript frontend provides a semantic search via S3 Vectors, natural language business objective translation, and real-time pipeline status via WebSocket. Fully decoupled from the processing pipeline.



© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Reference Architecture