SUMATOSOFT

PROJECT TYPE
An AI agent for shipment exception triage and resolution

TECHNOLOGIES
Python, TypeScript, React, Node.js, PostgreSQL, Pgvector, Redis, Apache Kafka, Anthropic Claude, Azure OpenAI, LangGraph, EDIFACT IFTSTA and IFTMIN parsers, Microsoft Graph API, REST and SOAP TMS connectors, OpenTelemetry, Grafana, Docker, Kubernetes

DURATION
7 months

TEAM
1 AI Architect
1 Business Analyst
2 ML Engineers
2 Backend Developers
1 Frontend Developer
1 QA Engineer

sumatosoft.com
info@sumatosoft.com

Industries: Freight forwarding and third-party logistics
Region: Netherlands

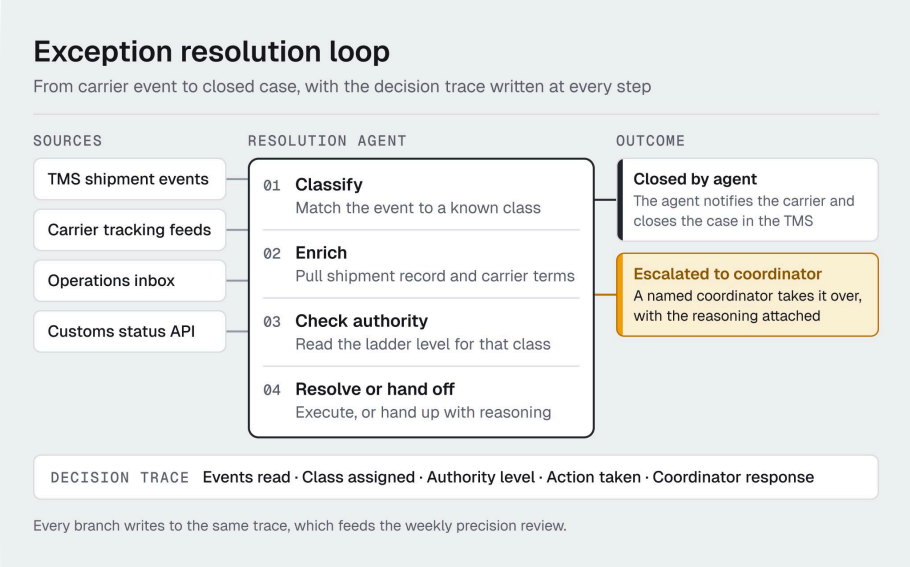
AI agent for freight exception resolution

An event-driven agent that triages shipment exceptions and closes routine cases on its own, while routing customs and detention risks to a named coordinator.

Project background

The Client moves 310,000 shipments a year on road and short-sea lanes across Western Europe. It works from 6 branch offices with 240 staff, and two shared operations inboxes absorb everything the network sends back. Carrier status messages, delay notices, customs updates, and customer chasers all land in the same queue. 9 coordinators work through that queue in arrival order, so a customs hold waits behind a routine address query.

Exception handling ran as manual triage against no written standard. Two coordinators could close the same exception type differently, and neither decision left a record anyone could retrieve. Median time to first action stood at 95 minutes, and the Monday queue reached 280 messages during peak weeks. Detention and demurrage accrued while notices waited their turn, and the Client's largest customers had begun asking for exception reporting it could not produce.



Project Distinctive Features

- An autonomy ladder fixing what the agent may do per exception class, with promotion to a higher level gated on four weeks of clean audit history

SUMATOSOFT

- Resolution rules that read detention clocks from the Client's own carrier agreements instead of generic delay assumptions
- Event-driven triage across EDIFACT status messages, mailbox arrivals, TMS milestone changes, and customs portal updates
- A decision trace on every action, sampled weekly against coordinator verdicts, feeding the threshold tuning cycle

Business challenge

The Client wanted to:

- Triage exceptions the moment they arrive rather than when a coordinator opens the message,
- Close routine classes without a coordinator while keeping customs and damage cases with a named person,
- Read detention clocks from its own carrier contracts,
- Record why every exception closed the way it did,
- Give branch managers exception volume by carrier and by class,
- Keep the existing TMS, shared mailboxes, and carrier EDI connections in place.

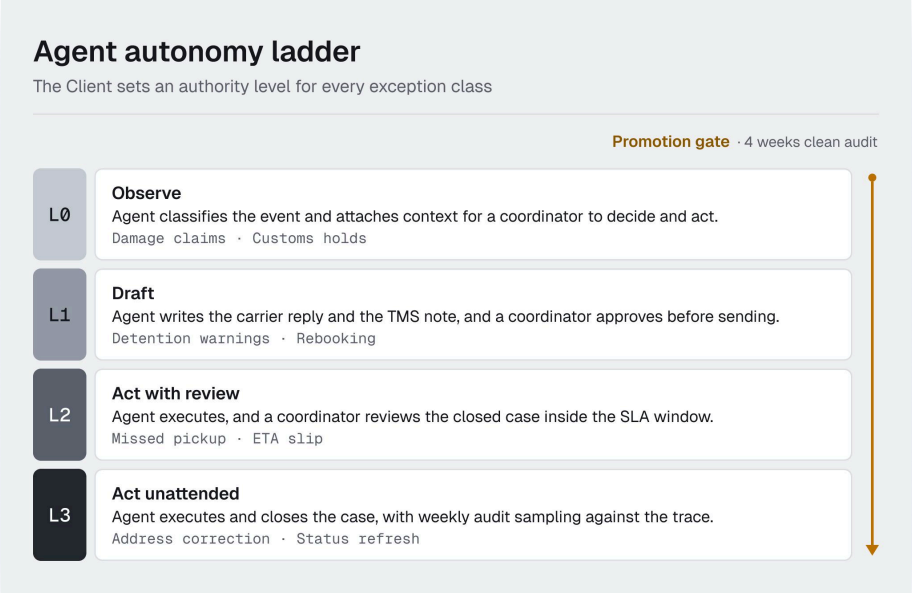
Our solution

SumatoSoft built an exception resolution agent that watches the Client's shipment event stream rather than an inbox queue. Carrier EDIFACT messages, mailbox arrivals through the Microsoft Graph API, TMS milestone changes, and customs portal updates all land on a Kafka topic.

The agent classifies each event into one of 14 classes, then assembles the context that class requires. It calls the TMS for the shipment record and the contract store for the account's carrier terms. Where a hold is in play, it also reads the customs portal, and the history index in pgvector supplies comparable past closures.

Every exception class carries an authority level that the Client's operations lead sets before deployment. At the lowest level, the agent drafts a carrier reply and waits for a coordinator to approve it. At the highest level, it contacts the carrier, updates the TMS record, notifies the consignee, and closes the case unattended. A LangGraph state machine holds the sequence, and each step writes to a decision trace covering the events read, the rule applied, the action taken, and the coordinator response. A class moves up a level only after four weeks of clean audit history. The weekly precision review then samples closed cases against coordinator verdicts to tune the confidence thresholds.

SUMATOSOFT



Customer's benefits

Over the 4-month measurement window, the Client achieved the following:

- Median time to first action down from 95 minutes to 45 seconds across covered exception classes,
- Detention and demurrage charges on covered lanes down 21%, worth EUR 186,000 a year,
- 34% of exception volume routed to a coordinator, with the remainder closed under the ladder,
- Monday peak backlog down from 280 untriaged messages to 95 awaiting a coordinator decision, each already classified and enriched before the desk opens,
- Coordinator hours released per week: 71, redeployed to claims handling and lapsed-account calls,
- Every closed exception carrying an auditable decision trace, against no record before the project.

What's happening with the project right now?

The agent runs across all 6 branch offices on the Client's road and short-sea lanes. The Client has promoted 5 exception classes a level since go-live, and customs holds stay at the observation level by its own decision. SumatoSoft supports the platform and is extending the agent to quote requests. There, the same authority ladder governs which quotes are released without a person and which are routed to the pricing desk.

sumatosoft.com

info@sumatosoft.com

One Boston Place, Suite 2602 Boston, MA 02108, United States