Streaming / Production Architecture for Utopia Video Cameras

Overview

This architecture is designed for a video analytics pipeline where camera events are ingested, deduplicated, enriched, and aggregated to support operational dashboards and downstream machine learning workloads.

The goal is to ensure data quality, fault tolerance, and low-latency insights.

Bronze Layer — Raw Streaming Ingestion

- Ingest raw camera detection events from Kafka (or similar).
- Preserve duplicates and original schema.
- Schema enforcement for corrupted records.
- Key fields: detection_oid, timestamp, camera_id, bbox, confidence.

Silver Layer — Deduplication + Enrichment

- Apply watermarking (e.g., 10 minutes) to handle late events.
- Use dropDuplicates("detection_oid") to remove camera-side retransmits.
- Broadcast join with static dimension table B:
- location metadata
- camera orientation & type
- operational meta (owner, priority zones)

Gold Layer — Aggregation & Serving

- Compute top-N alerts per location (sliding window).
- Generate event counts, confidence distribution, anomaly metrics.
- Store for BI dashboards (Delta + Presto/Trino).

Serving Pattern

- Dashboard always queries Gold, ensuring trusted and enriched data.
- ML models can optionally use Silver for fresher events.

Operational Features

- Checkpointing + exactly-once sinks
- Data quality metrics (schema violations, duplicate ratio)
- Alerting pipeline hooked to Silver