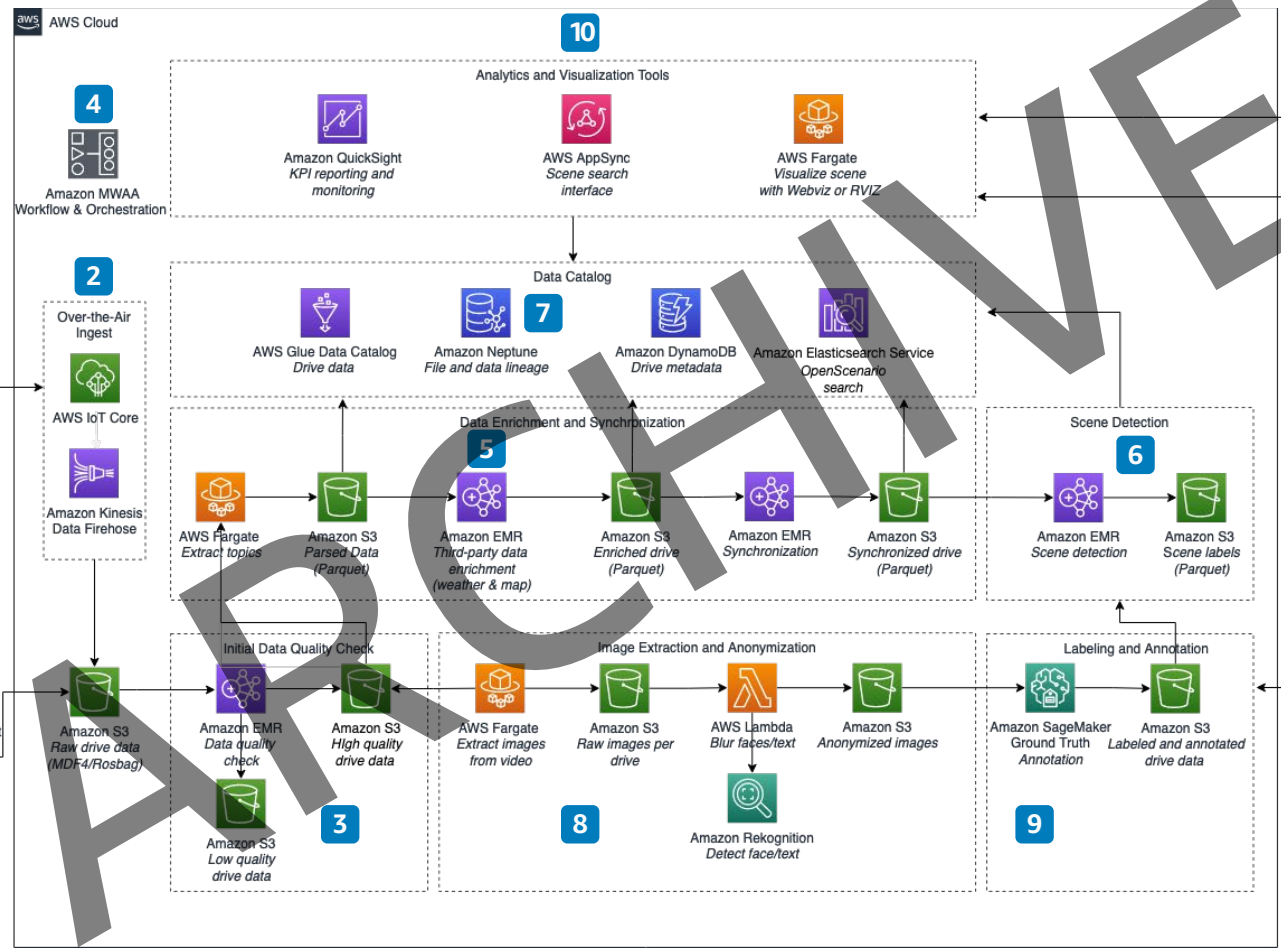# Autonomous Driving Data Lake

This diagram has been archived. For the latest in AWS Architecture Center diagrams, go to
https://aws.amazon.com/architecture/

AWS Reference Architecture

1. **AWS Outposts** or third-party hardware processes raw drive data from autonomous fleet for validation/quality checks.

2. Ingest vehicle telemetry data in near real time using **AWS IoT Core** and **Amazon Kinesis Data Firehose**.

3. Remove and transform low quality data.

4. Schedule and manage the extract, transform, load (ETL) jobs and pipelines using **Amazon Managed Workflows for Apache Airflow (Amazon MWAA)**.

5. Enrich data with weather conditions based on GPS location and timestamp. Then, synchronize data for post processing pipelines.

6. Detect scenes using **Amazon EMR** in ASAM OpenSCENARIO into **Amazon DynamoDB** and **Amazon Elasticsearch Service**.

7. Store data lineage in **Amazon Neptune** and catalog data using **AWS Glue Data Catalog**.

8. Detect and blur faces and text using **AWS Lambda** and **Amazon Rekognition**.

9. Perform automated labeling on raw data or anonymized data using **Amazon SageMaker Ground Truth** and/or third-party labeling tools/systems.

10. Provide an advanced analytics and visualization toolchain including search function for particular scenarios using **AWS AppSync**, **Amazon QuickSight** (KPI reporting and monitoring), and Webviz, RVIZ, or other tooling for visualization.