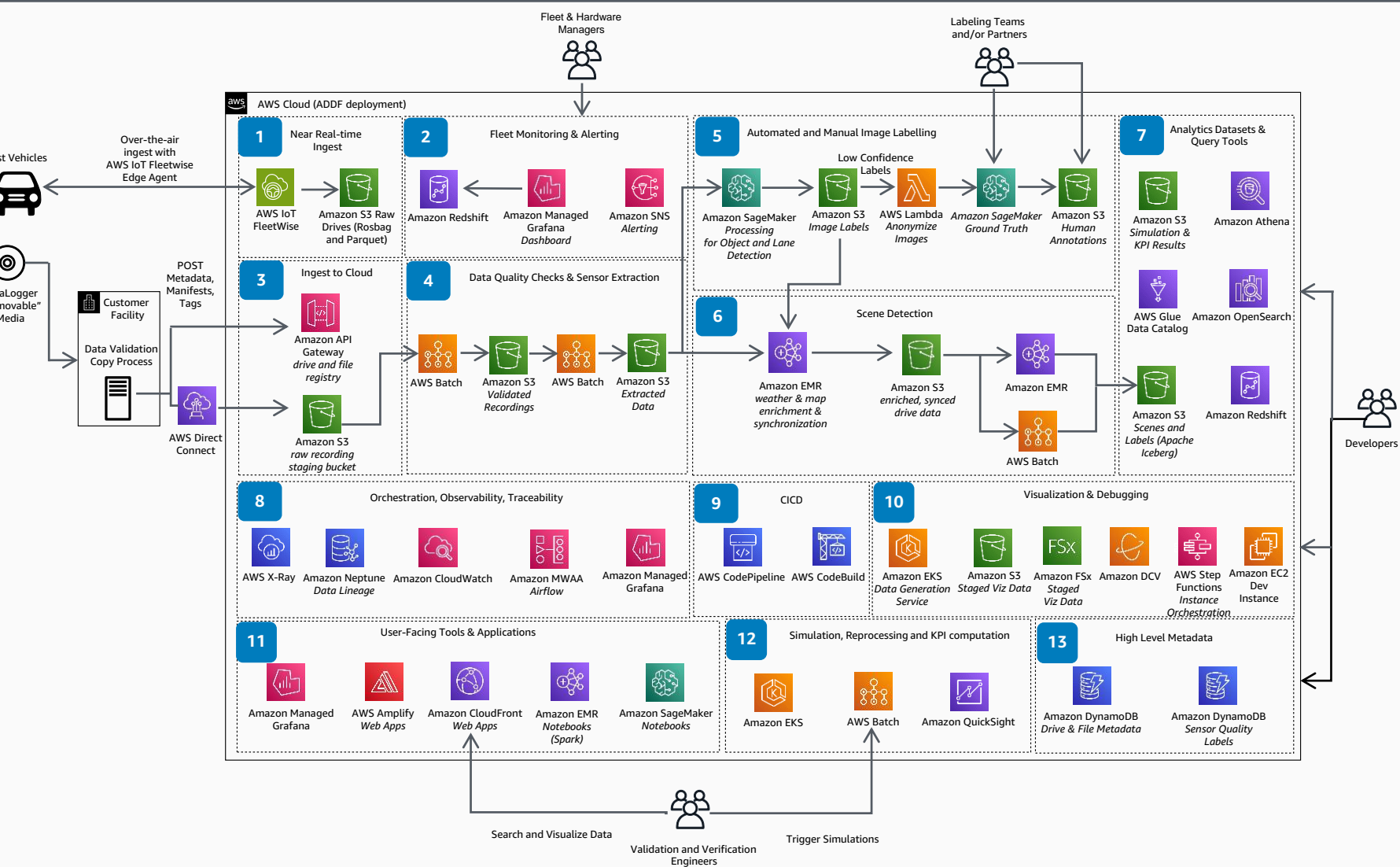


Guidance for Autonomous Driving Data Framework on AWS

Handling data at scale with the Autonomous Driving Data Framework (ADDF)

ADDF is a ready-to-use, open-source framework for advanced driver assistance systems (ADAS) workloads.



- 1 Near real-time ingestion of sensor data, data modeling, indexing, and enrichment through **AWS IoT FleetWise** and data stored in **Amazon Simple Storage Service (Amazon S3)**.
- 2 Near real-time fleet monitoring and alerting with **Amazon Redshift**, **Amazon Managed Grafana**, and **Amazon Simple Notification Service (Amazon SNS)**.
- 3 Bulk upload of recording data from copy stations through **AWS Direct Connect**, ingestion validation and registry with **Amazon API Gateway**, and a raw recording staging bucket with **Amazon S3**.
- 4 Initial data quality checks and data extraction with containers running on **AWS Batch**. Processed data is stored in **Amazon S3**.
- 5 Images are annotated with machine learning models to detect objects and road lanes. Low confidence predictions are set aside for manual annotation. Bounding boxes are used for blurring faces and license plates. **Amazon SageMaker Ground Truth** is used for labeling.
- 6 Use **Amazon EMR** in combination with **Amazon S3** and **AWS Batch** to enrich sensor data with localized weather and map matching info. It also combines image annotations, and sensor data to detect various scenes like traffic intersections or people and objects in the street.
- 7 AWS analytics toolchain manages parquet datasets and schema evolution with Apache Iceberg, **AWS Glue Data Catalog**, querying tools such as **Amazon Athena**, **Amazon Redshift**, and **OpenSearch**.
- 8 Data pipeline orchestration with **Amazon Managed Workflows for Apache Airflow (Amazon MWAA)** observability of distributed workloads with **Amazon Managed Grafana**, **Amazon CloudWatch**, and **AWS X-Ray**. **Amazon Neptune** is used for data lineage.
- 9 Build, test, and deploy using GitOps on **AWS CodePipeline** and **AWS CodeBuild**.
- 10 Host high-performance, on-demand visualization applications on **Amazon Elastic Kubernetes Service (Amazon EKS)** for engineers. Developer instances use **Amazon Elastic Compute Cloud (Amazon EC2)** and **Amazon DCV** to stage and share files with **Amazon FSx for Lustre** or **Amazon S3**. Use **AWS Step Functions** for instance orchestration.
- 11 User-facing tools like Python and Spark Notebook infrastructure use **Amazon EMR** and **SageMaker**. Custom dashboards can be configured with **Grafana**, and web applications are built and hosted on **AWS Amplify** and **Amazon CloudFront**.
- 12 Scalable simulation and KPI calculation modules use **Amazon EKS** or **AWS Batch**. **Amazon QuickSight** is used to analyze KPIs and simulation results.
- 13 Drive and file-level metadata can be stored and queried with **Amazon DynamoDB** at scale for pipeline traceability and to store metadata, manifests, markers, and tags.