AWS re:Invent

NOV. 29 – DEC. 3, 2021 | LAS VEGAS, NV

AUT306

AV data lake reference architecture

Junjie Tang
Principal Consultant, Big Data
Amazon Web Services

Hendrik Schöneberg
Senior Consultant, Big Data
Amazon Web Services

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AV data lake and assets

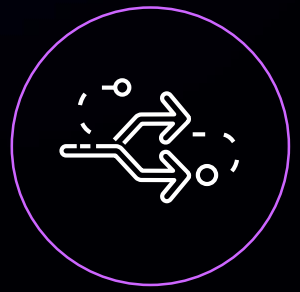
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Autonomous driving: Challenges and pain points

50 CAR FLEET, DRIVING 6 H/DAY GENERATES MORE THAN 2 PB/DAY



Big data repository for more than 2 PB/day of sensor data and drive logs need to be ingested, encoded, stored for AV development and archived for regulatory compliance (ISO 26262)



Enable diverse processing pipelines to, for example, annotate the ingested data with labels and metadata, or identify the 5%–10% of frames that are relevant for your AV use case



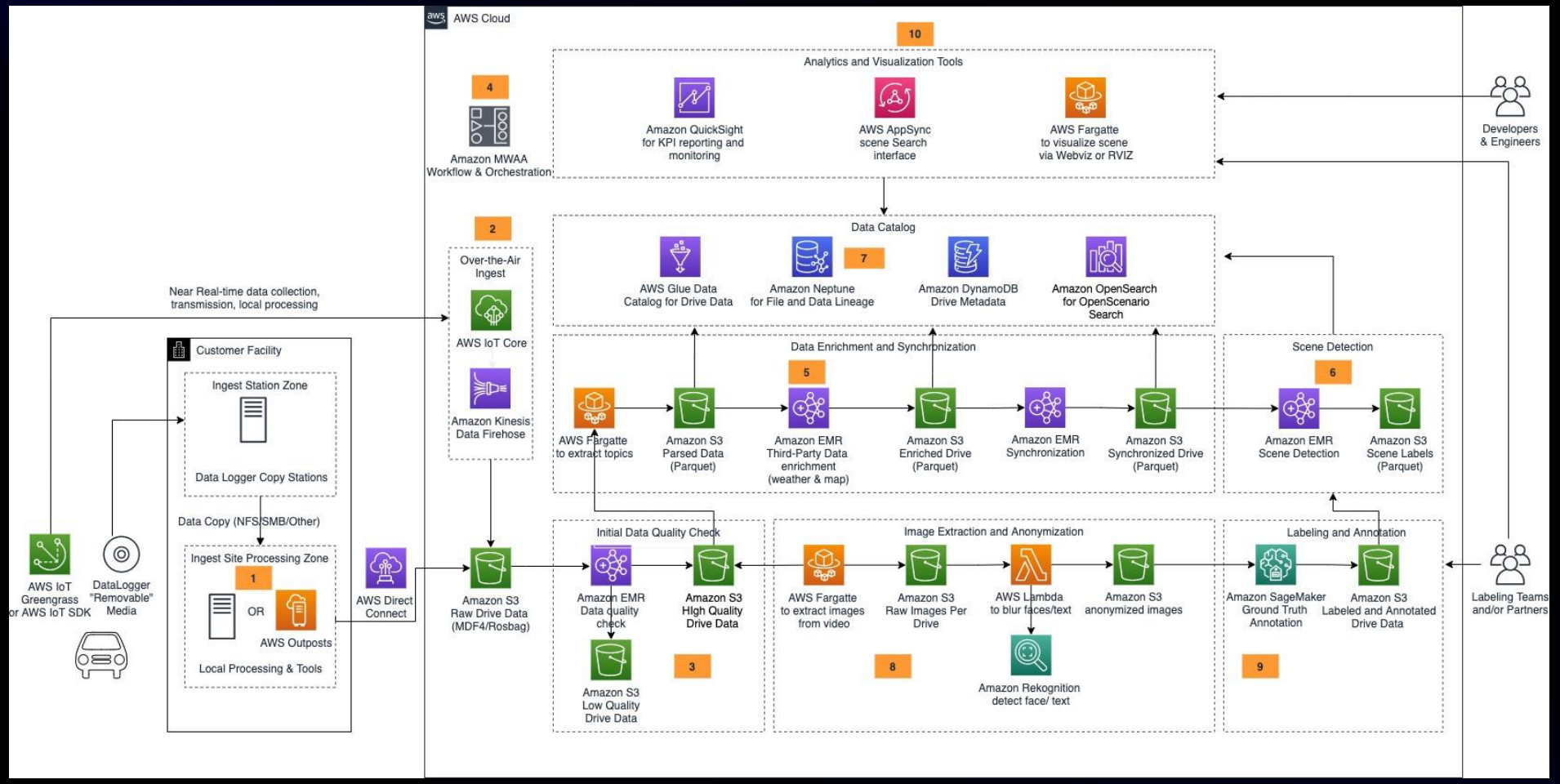
Search, analysis, and visualization of the big data repository by developers, data scientists, researchers, and even legal departments. Enable efficient ML pipelines with parallel experiment and model training



Enable simulation of hundreds of millions of miles in highly distributed AV simulation workloads. Test against 10,000s of hours of sensor data, repeat daily.

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AV Data Lake Reference Architecture 2.0



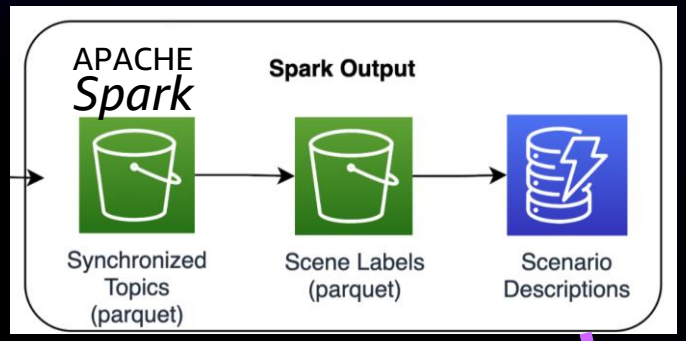
1. Edge-based processing and ingest of test drive logs
2. Ingest vehicle telemetry data
3. Remove and transform low-quality data
4. Schedule and manage pipelines
5. Enrich and synchronize data
6. Extract Scenarios
7. Track data lineage
8. Detect and blur faces and text
9. Labeling and Annotation
10. Advanced analytics and visualization

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What are your challenges?

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

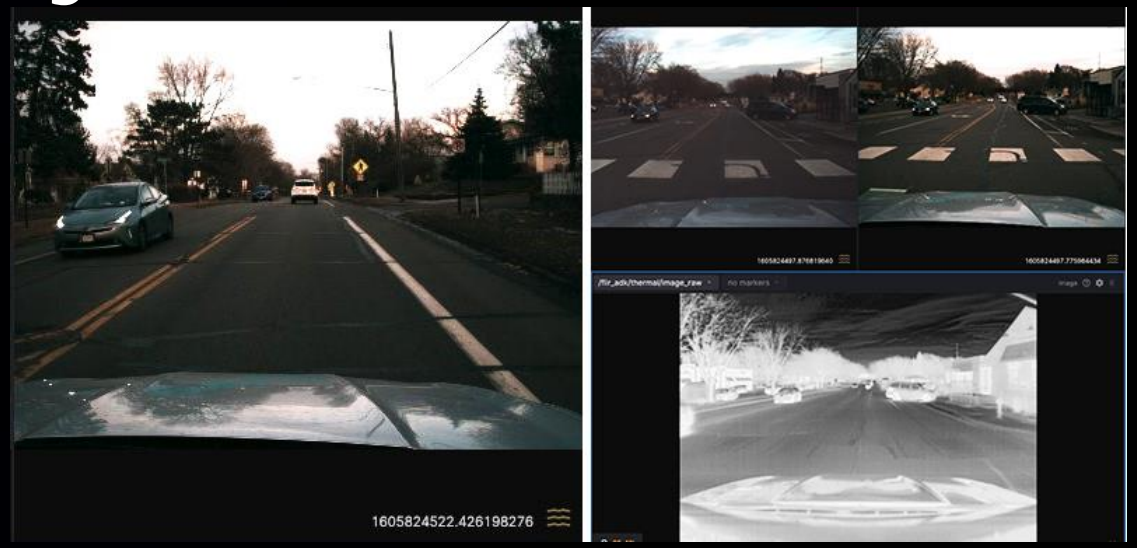
ROS bag ingestion, topic synchronization and scenario detection



Person-in-lane scenario

Summarize all cases where a person

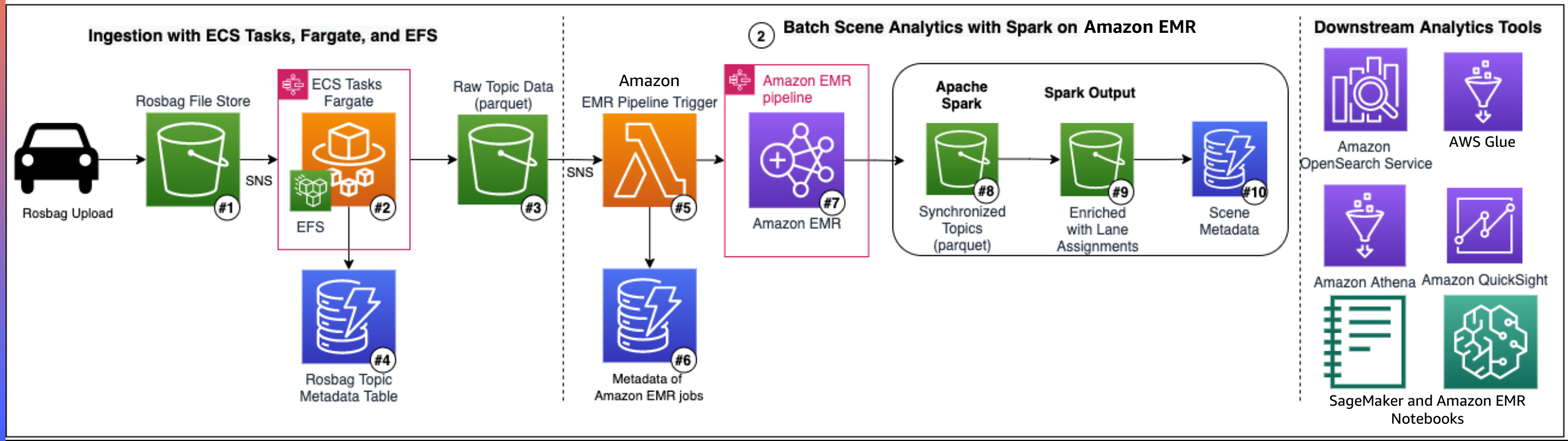
may be in the street




```
bag_file String : _2020-11-19-16-21-22_4
bag_file_bucket String : rosbag-bag-file-ingest-537686541335
bag_file_prefix String : urban_ros_data/_2020-11-19-16-21-22_4.bag
end_time Number : 1605824497.1
num_people_in_scene_start Number : 6
scene_id String : _2020-11-19-16-21-22_4_PersonInLane_1.6058244969E9
scene_length Number : 0.19999980926513672
start_time Number : 1605824496.9
topics_analyzed String : rgb_right_detections_only_clean,post_process_lane_points_rgb_front_right_clean
```

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

ROS bag ingestion, topic synchronization and scenario detection

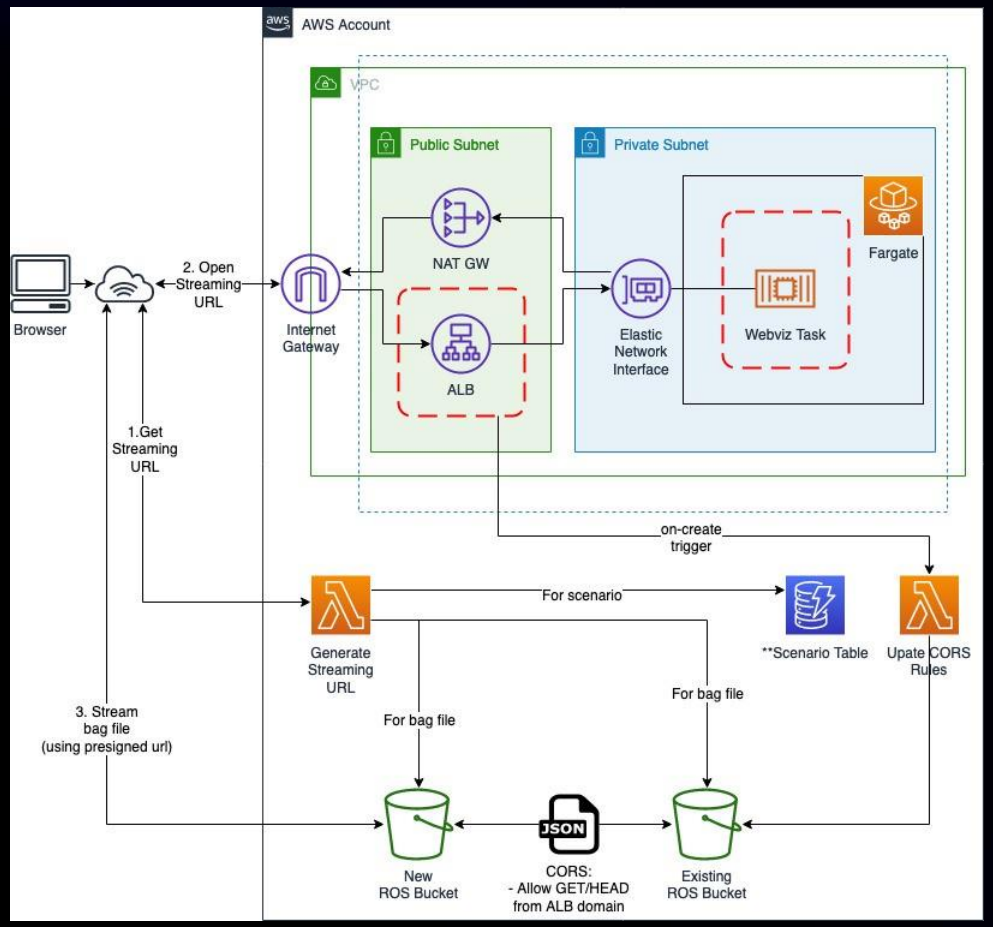


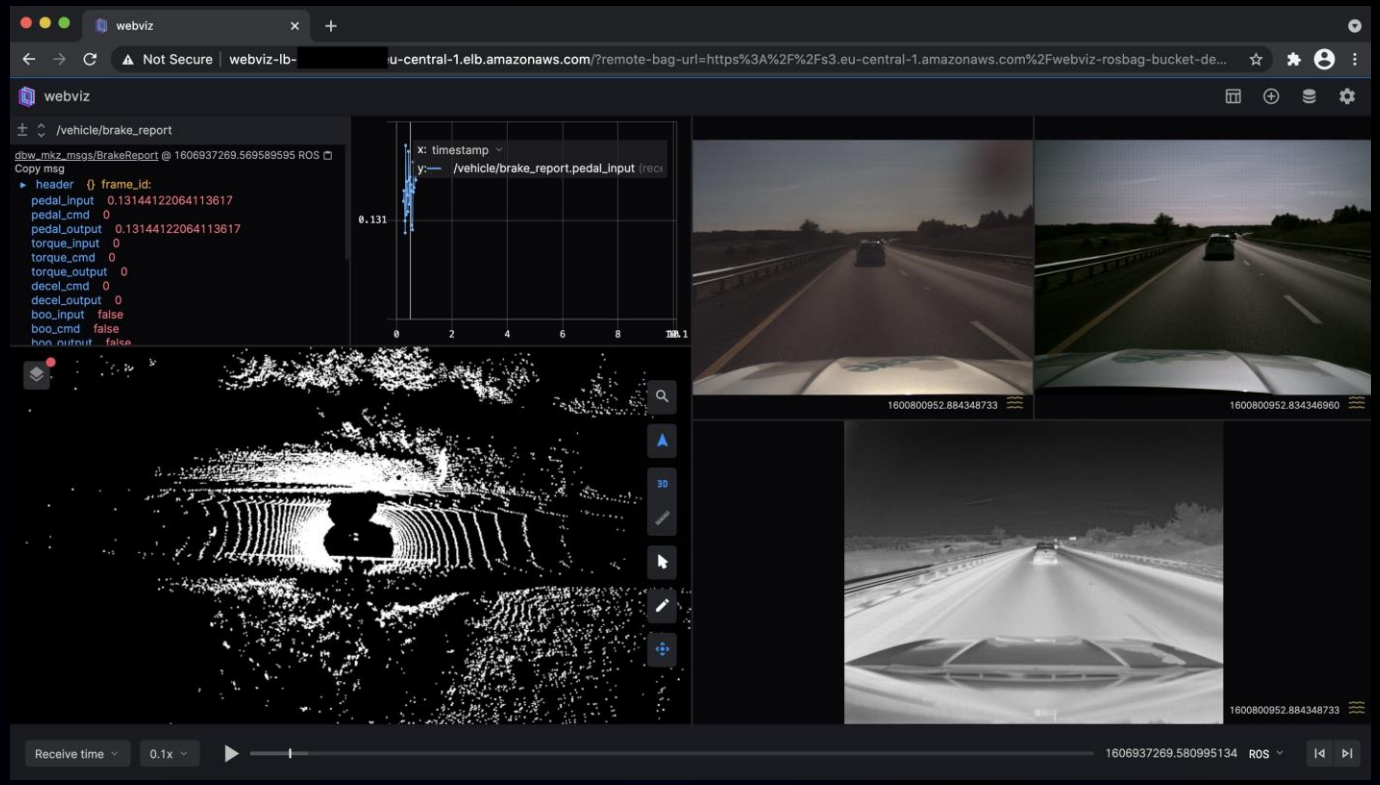
Scene analytics:

From raw Rosbag files to searchable scenario labels and descriptions

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

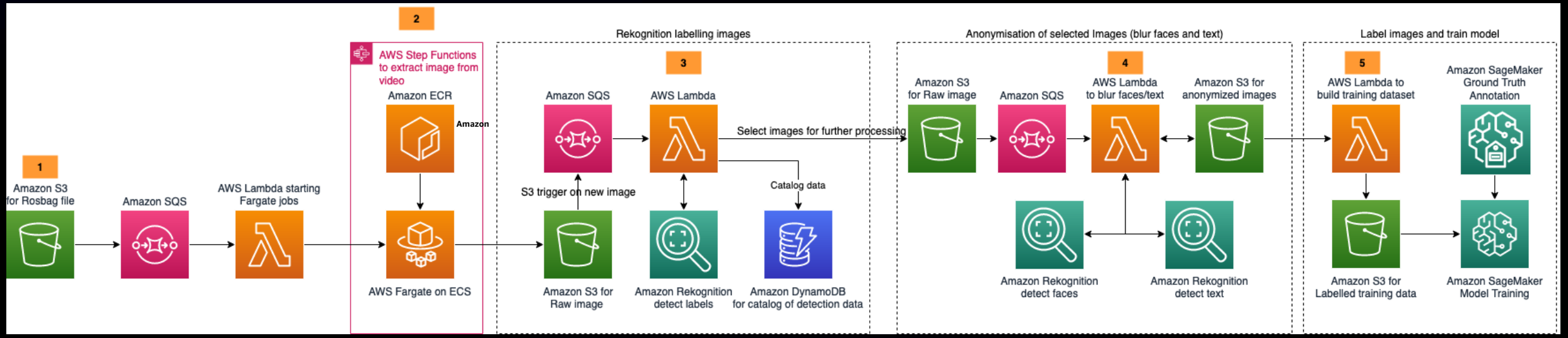
ROS bag file visualization





© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

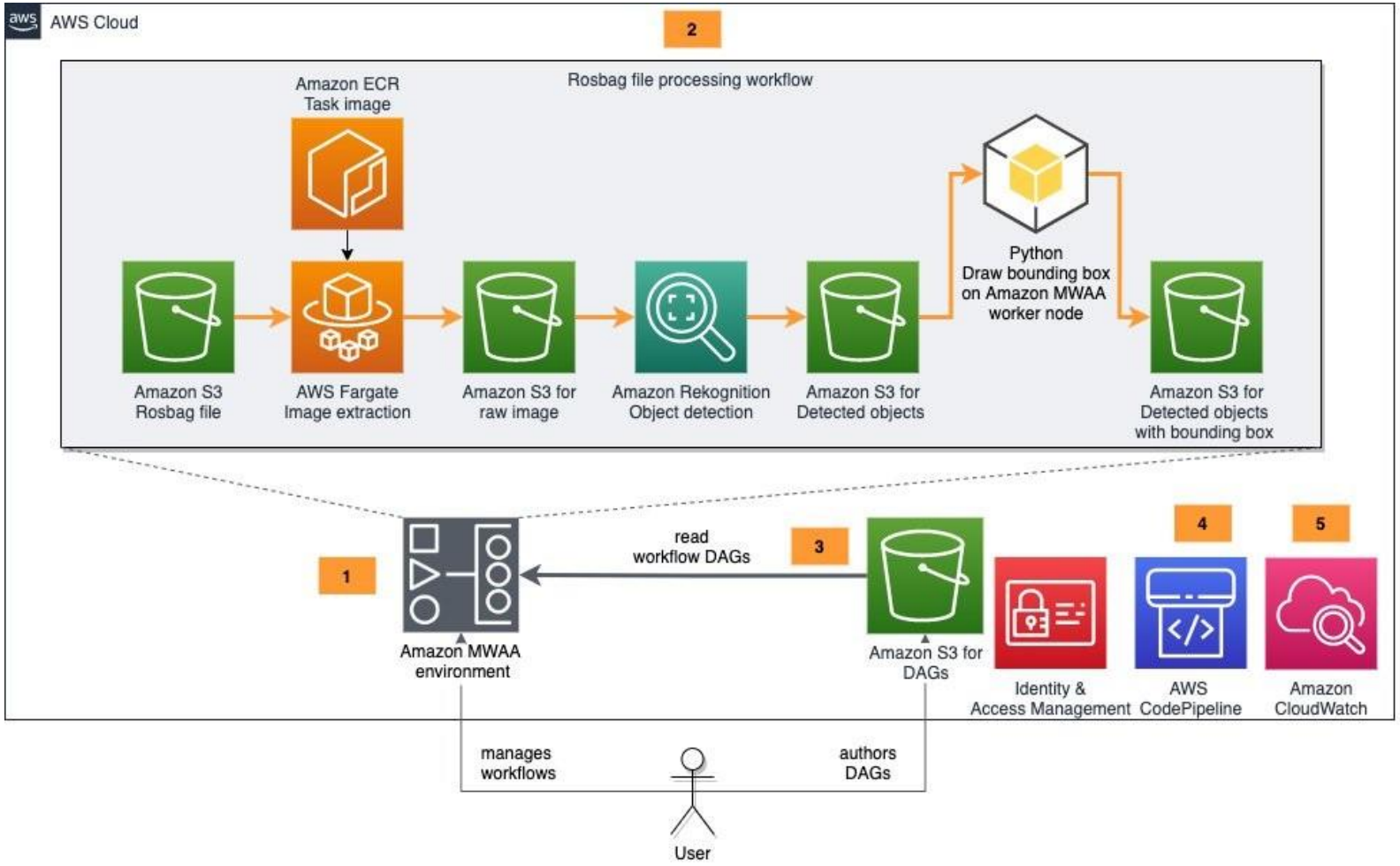
Automated image processing and model training pipeline for autonomous driving



1	Rosbag file contains raw video and sensor data
2	AWS Step Functions to extract image from video and store them in PNG format
3	AWS Lambda to detect labels using Amazon Rekognition, store the labels in Amazon DynamoDB and select images for further processing
4	AWS Lambda to detect faces and text using Amazon Rekognition, blur them and store the anonymized images in Amazon S3
5	AWS Lambda to build the training dataset for Amazon SageMaker Ground Truth and Model Training in Amazon SageMaker (custom perception model)

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Automated image processing and model training pipeline for autonomous driving

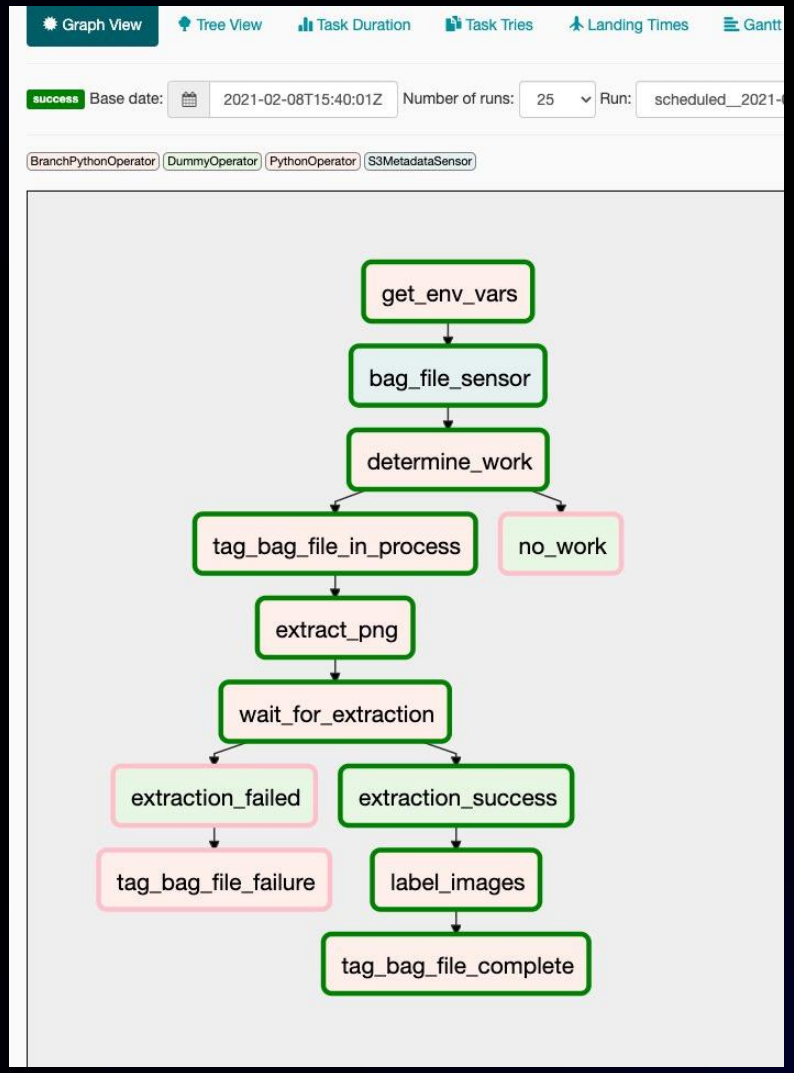


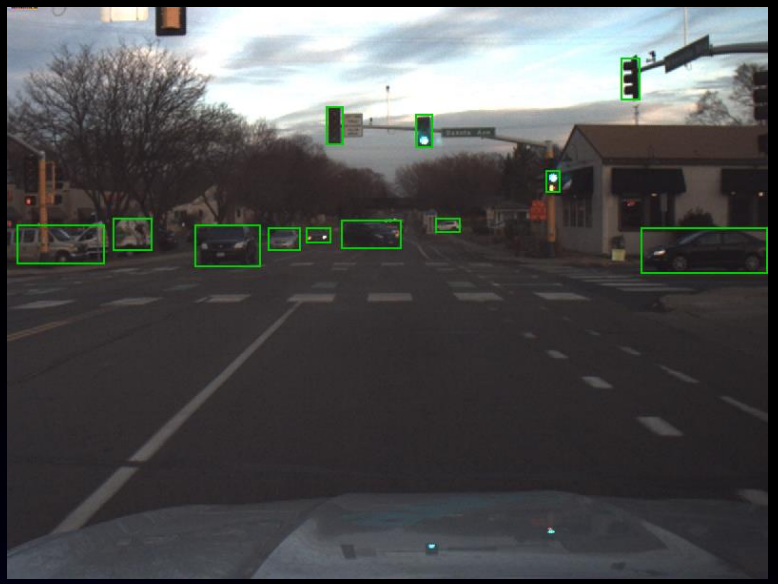
1. Amazon MWAA to run the Rosbag file processing workflow
2. Rosbag file processing workflow including image extraction, object detection and bounding box visualization
3. Predefined DAGs
4. AWS CodePipeline for auotomated CI/CD pipeline
5. Amazon CloudWatch for monitoring and operational dashboard

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Automated image processing and model training pipeline for autonomous driving







© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Recap & next steps

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Resources

Field Notes: Building an Automated Image Processing and Model Training Pipeline for Autonomous Driving

https://aws.amazon.com/blogs/architecture/field-notes-building-an-automated-image-processing-and-model-training-pipeline-for-autonomous-driving/

Field Notes: Deploying Autonomous Driving and ADAS Workloads at Scale with Amazon Managed Workflows for Apache Airflow:

https://aws.amazon.com/blogs/architecture/field-notes-deploying-autonomous-driving-and-adas-workloads-at-scale-with-amazon-managed-workflows-for-apache-airflow/

Field Notes: Deploy and Visualize ROS Bag Data on AWS using rviz and Webviz for Autonomous Driving

https://aws.amazon.com/blogs/architecture/field-notes-deploy-and-visualize-ros-bag-data-on-aws-using-rviz-and-webviz-for-autonomous-driving/

Field Notes: Building an automated scene detection pipeline for Autonomous Driving – ADAS Workflow

https://aws.amazon.com/blogs/architecture/field-notes-building-an-automated-scene-detection-pipeline-for-autonomous-driving/

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.