Information Disclosure Classification: PUBLIC



AWS Summit Japan 2026

TIER IV AWSで実現する自動運転のデータ基盤とMLOps

2026 / 06 / 26

Speaker



Riki Ota

ML&DataOpsグループ
グループリーダー

経歴

~ 2022.9　ソフトバンク株式会社

MaaSアプリの開発、自動運転運用のためのPoCアプリの開発

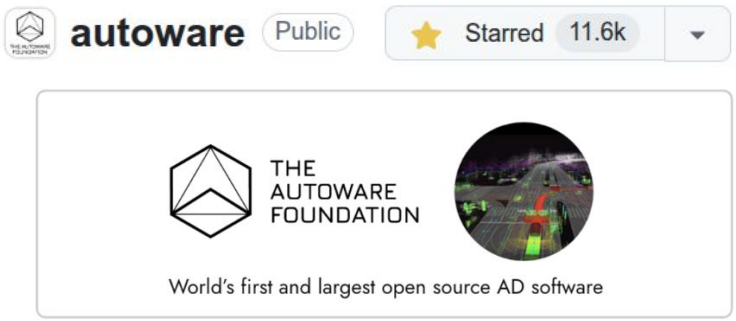
2022.10~　株式会社ティアフォー

自動運転のデータ基盤、MLOps基盤の開発・運用

自動運転の民主化

自動運転OSS Autoware

世界初の自動運転OSSであるAutowareの開発をリード。世界115以上の企業・大学・研究機関が参画する Autoware Foundation を通じて、自動運転のエコシステム構築を推進。



https://github.com/autowarefoundation/autoware

自動運転レベル4の実現

自動運転レベル4の認可および特定自動運行許可を取得して一般公道での運行を実現

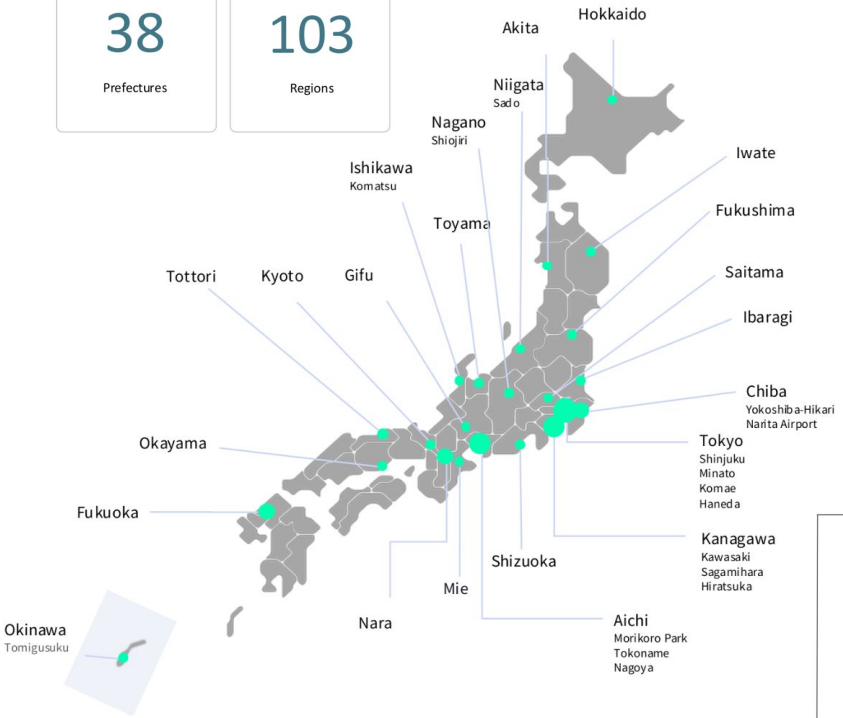


長野県塩尻市で自動運転レベル4の運行許可を取得

自動運転の走行地域・車種

38
Prefectures

103
Regions

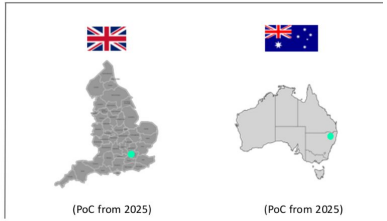






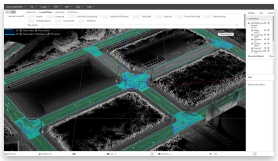


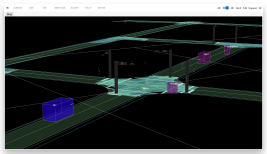


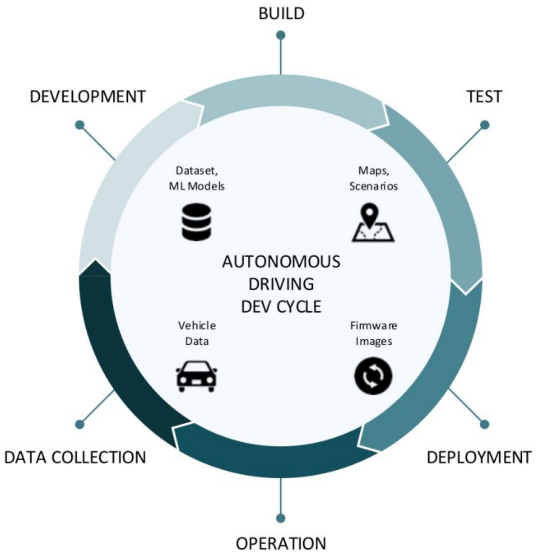


自動運転DevOps基盤 - Web.Auto

CI/CD PIPELINE







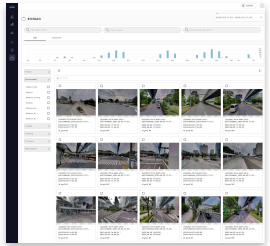
SIMULATION





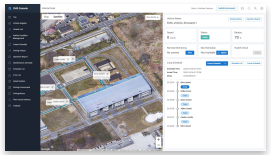


DATA PLATFORM/ML PIPELINE





FLEET MANAGEMENT



REMOTE OPERATION





AI開発とMLOps

—

01

自動運転の機械学習モデル

Modular AIによる自動運転レベル4の実現と End-to-End AIによる次世代アーキテクチャへの取り組み

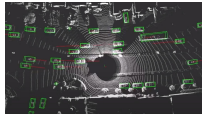
Perception AI

Dynamic Object

2D Object Detection

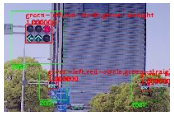


3D Object Detection



Traffic Light

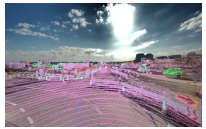
Fine Detection

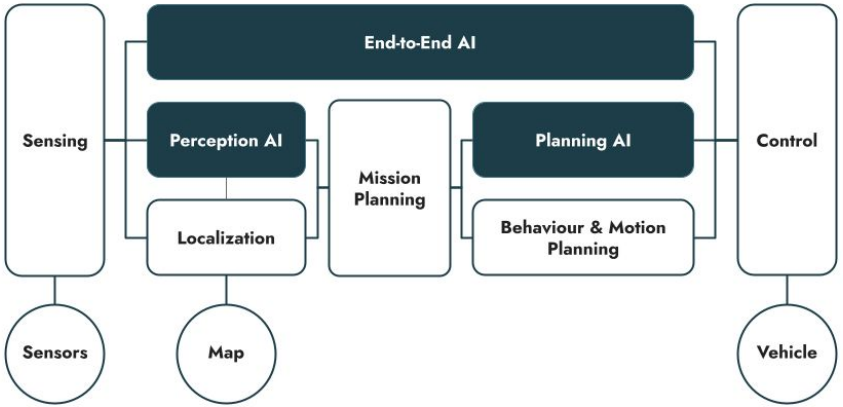


Classification

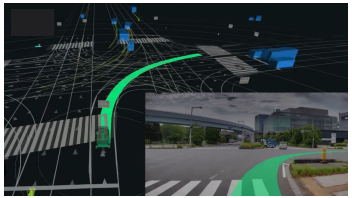


3D Semantic Segmentation

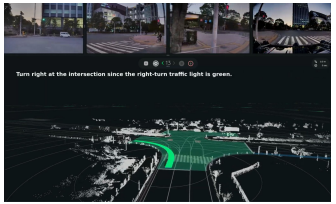




Planning AI



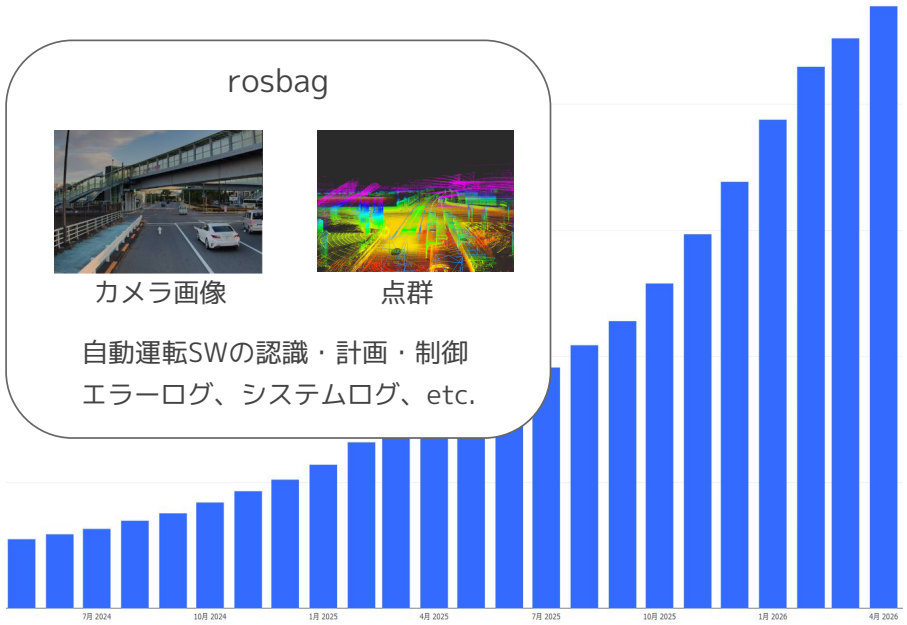
End-to-End AI



自動運転走行データの課題

ペタバイト級の走行データを学習・評価に活用するためのデータ基盤が必要

月次データ増加量: 約300TB



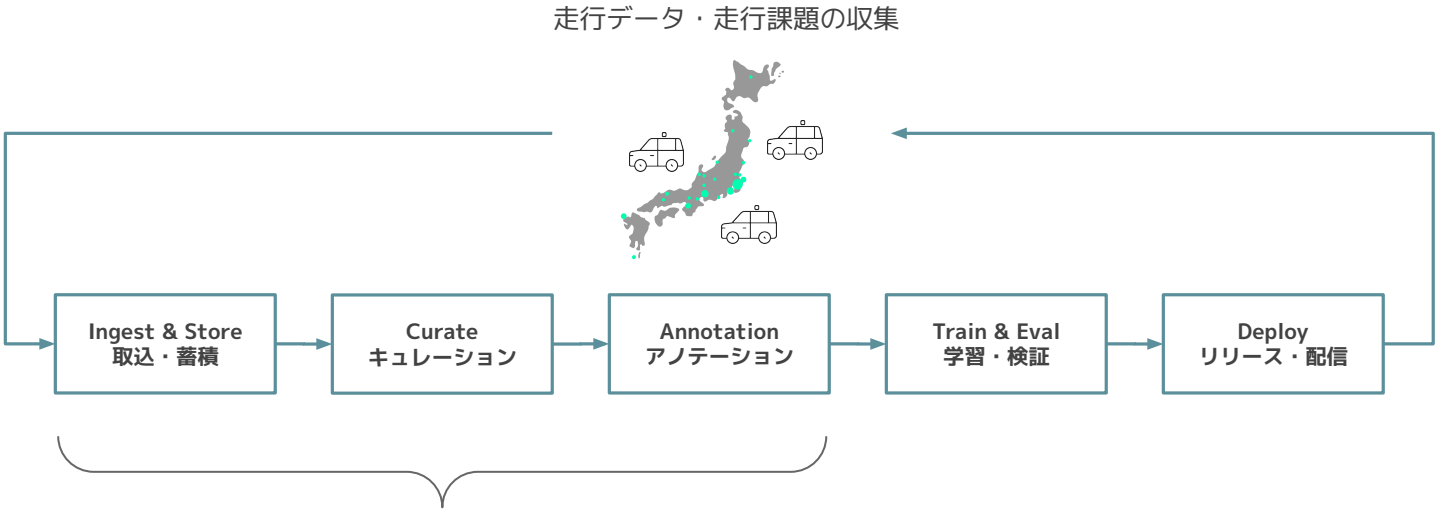
課題

- 増加し続ける走行データとストレージコスト
- 走行データのシーンや内容を把握できない
- 学習・評価に必要なシーンを見つけられない

大規模な走行データを可視化・分析し、
容易に検索できるデータ基盤の構築

自動運転のMLOps

AI開発を加速させるMLOpsワークフローの構築



- 多様な走行データの取り込み
- 大規模な非構造化データの選定
- 効率的な学習データの作成

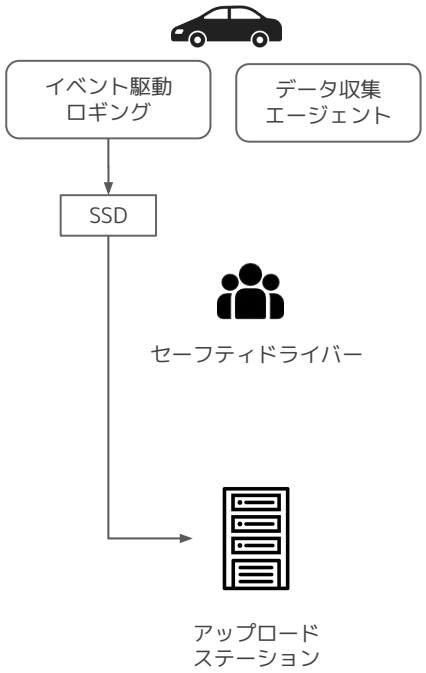


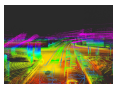
データパイプライン

—

02

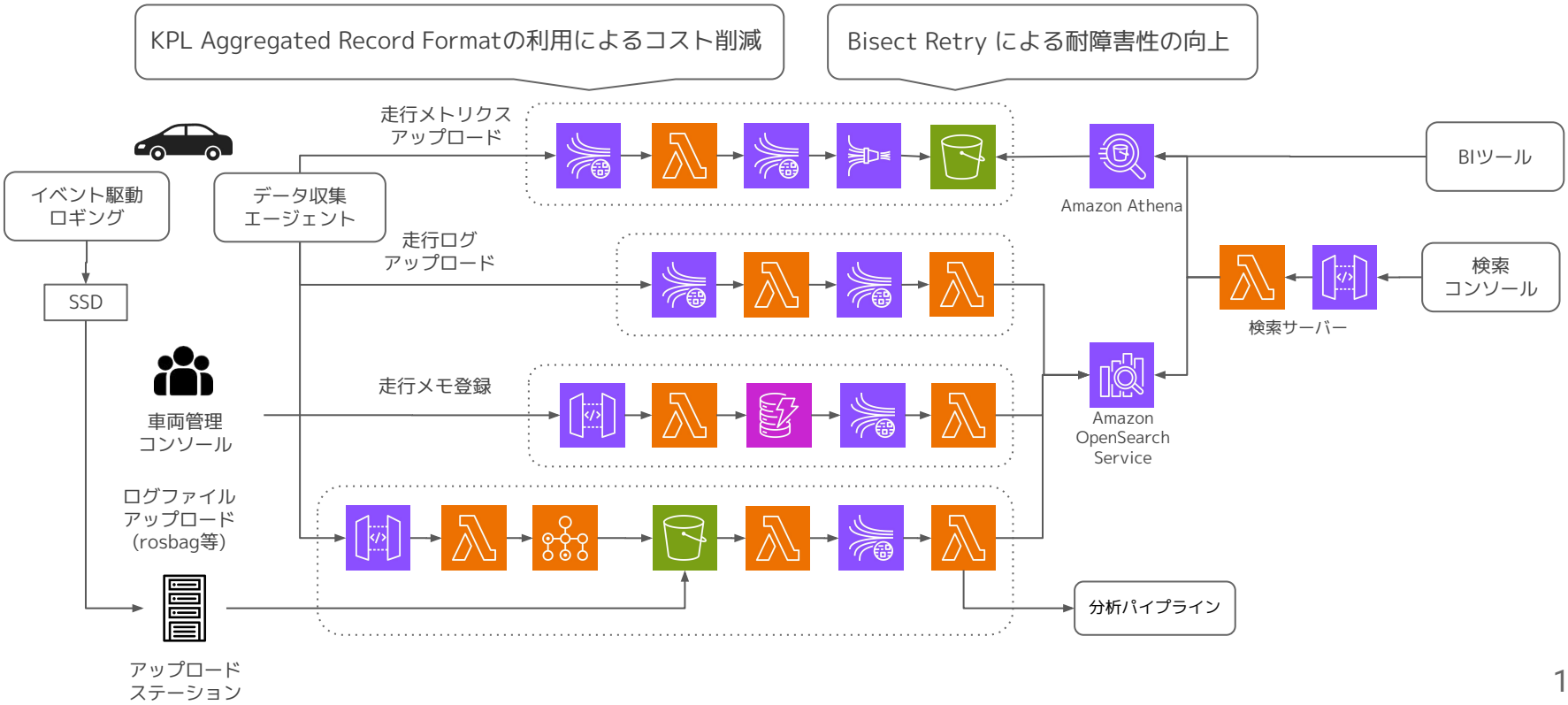
走行データの種別と収集方法



データ種別	収集方法
走行メトリクス CPU利用率、オーバーライドイベントなど **走行ログ** システムログ、エラーログなど	ニアリアルタイムな解析が必要となるため、各ECU上のデータ収集エージェントが集約して車両からアップロード 利用サービス Amazon Kinesis Data Streams
走行メモ セーフティードライバーが検知したイベント	車両管理コンソールからユーザーが直接登録 利用サービス Amazon API Gateway
ログファイル センサーデータを含むrosbagなど rosbag  カメラ画像  点群	毎分あたり10GB前後のデータになるためSSDをアップロードステーションに直接接続してアップロード 利用サービス Amazon S3

検索基盤への走行データの集約

走行データを一元管理するデータパイプラインを構築

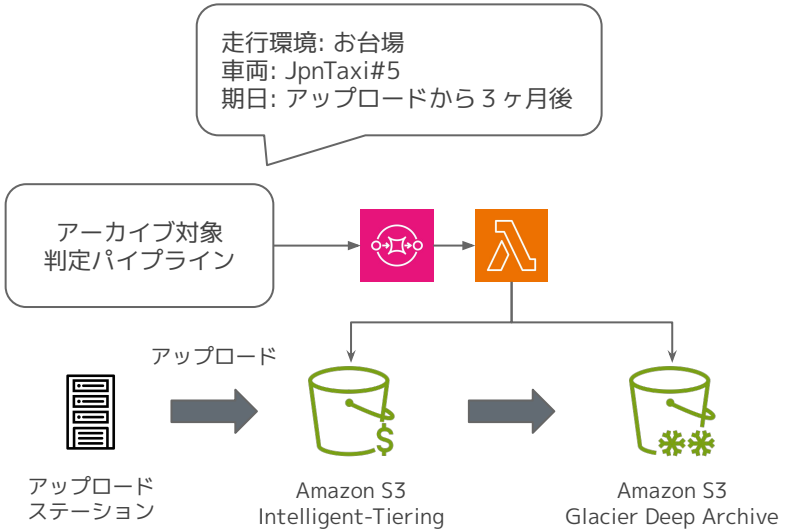


走行データの運用最適化の取り組み

ログファイルの条件付き自動アーカイブ

特定条件に合致したログファイルのストレージクラスの移行

- 車両や走行環境に合わせた柔軟な条件設定
- 解析対象のデータや学習に利用されたデータの除外



走行ログの自動アーカイブ

走行ログをJSONLにまとめてS3に恒久保存し、OpenSearchは一定期間後に自動削除

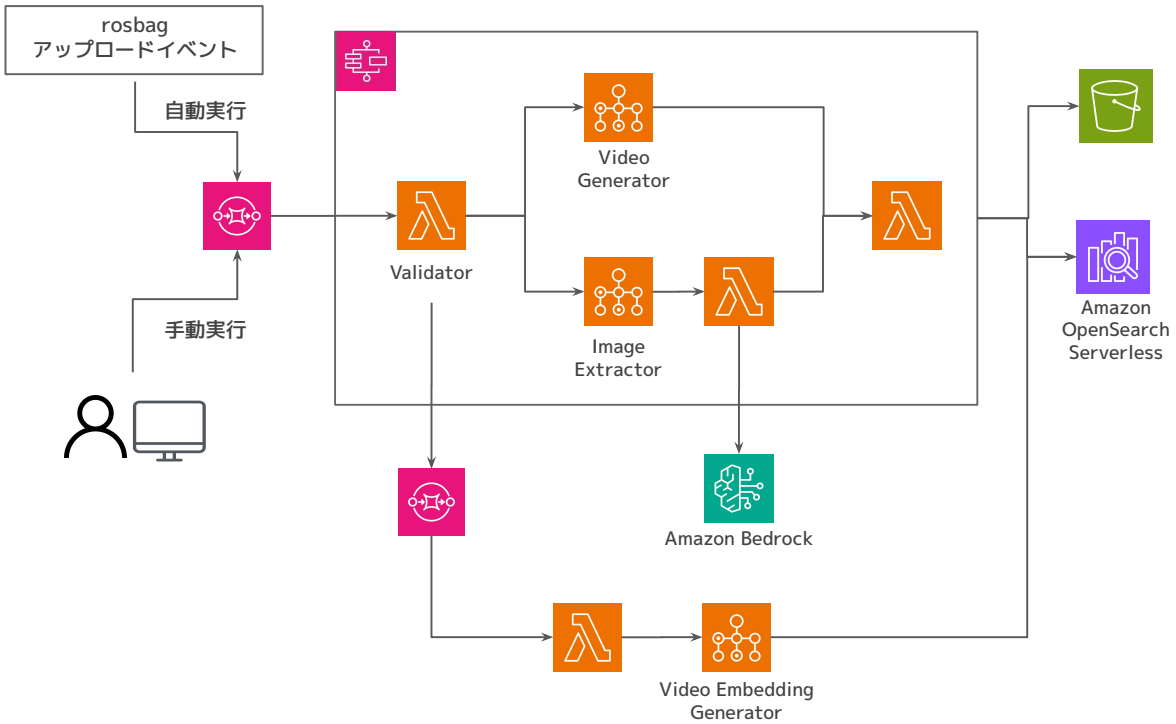
- OpenSearch Compute Unit (OCU) の増加を抑制
- 最新のデータの検索性を担保しつつ、過去のデータも検索可能な状態を維持



走行データの可視化と分析

03

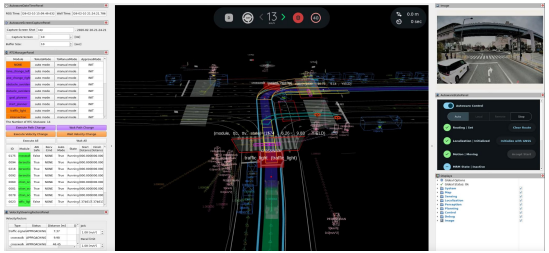
走行データの可視化・分析パイプライン



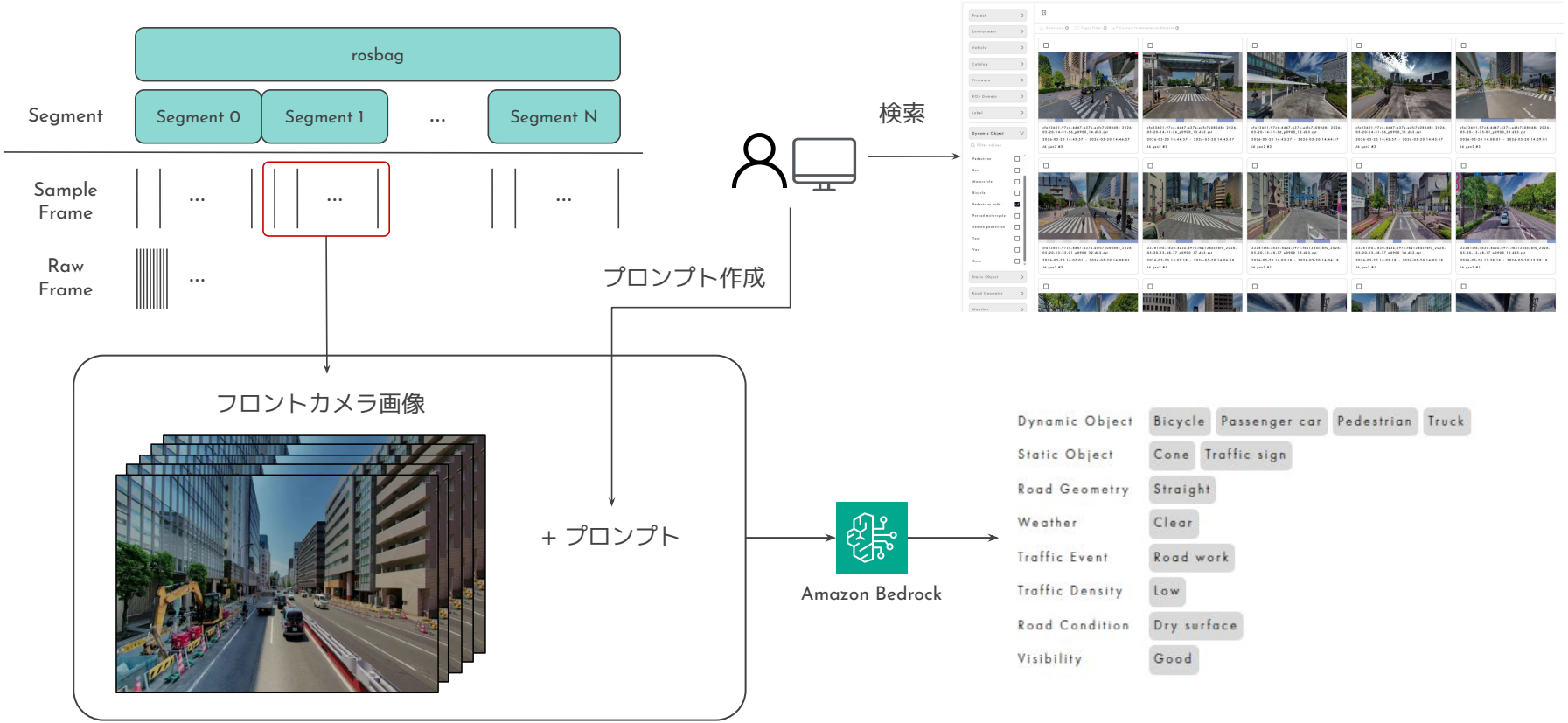
3D物体認識動画



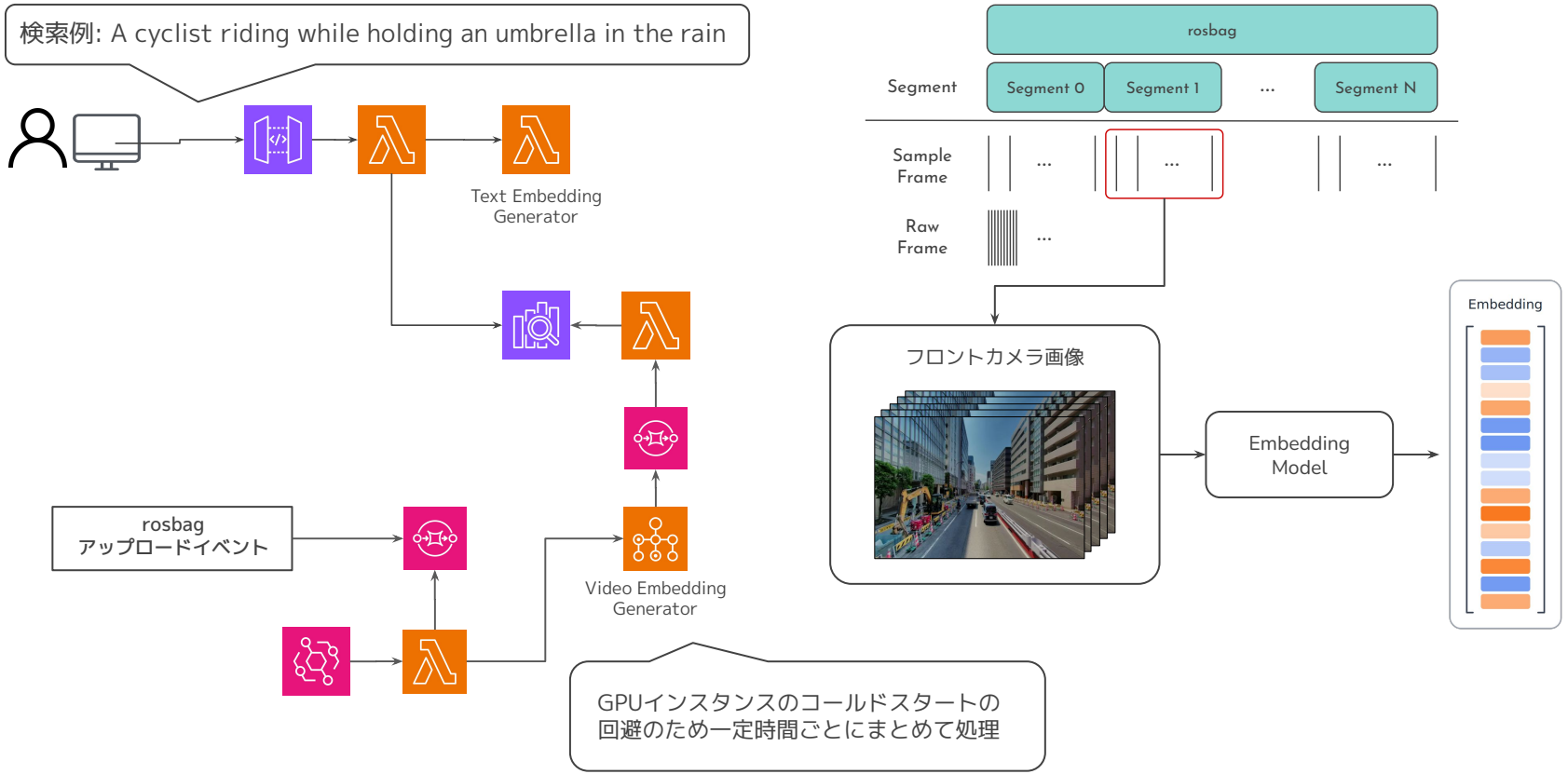
経路計画動画



Amazon Bedrockによるシーンタグの生成



Video Embeddingを利用したセマンティック検索





学習データセットの作成

04

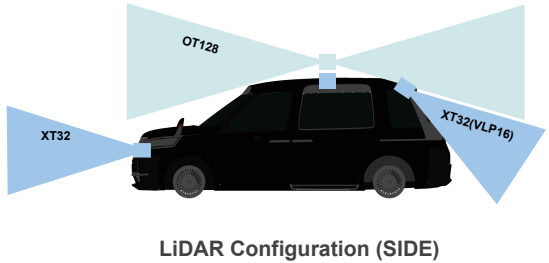
学習・評価データセットのフォーマット

T4 format

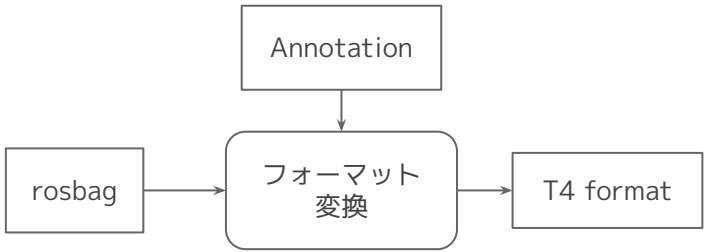
- nuScenesのフォーマットを拡張したデータフォーマット

rosbag → T4 format 変換

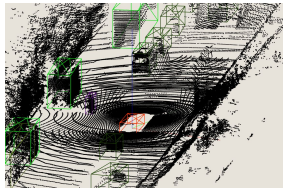
- 自動運転ソフトウェアを利用した点群の生成
- 結合された点群とカメラの画像取得時刻の調整
- アノテーションツールで読み込み可能なフォーマットへの変換



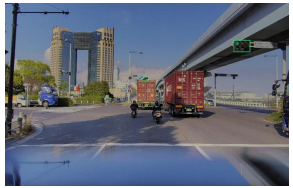
LiDAR Configuration (SIDE)



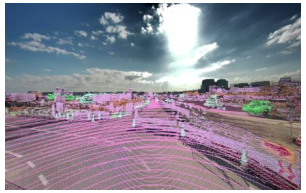
3D



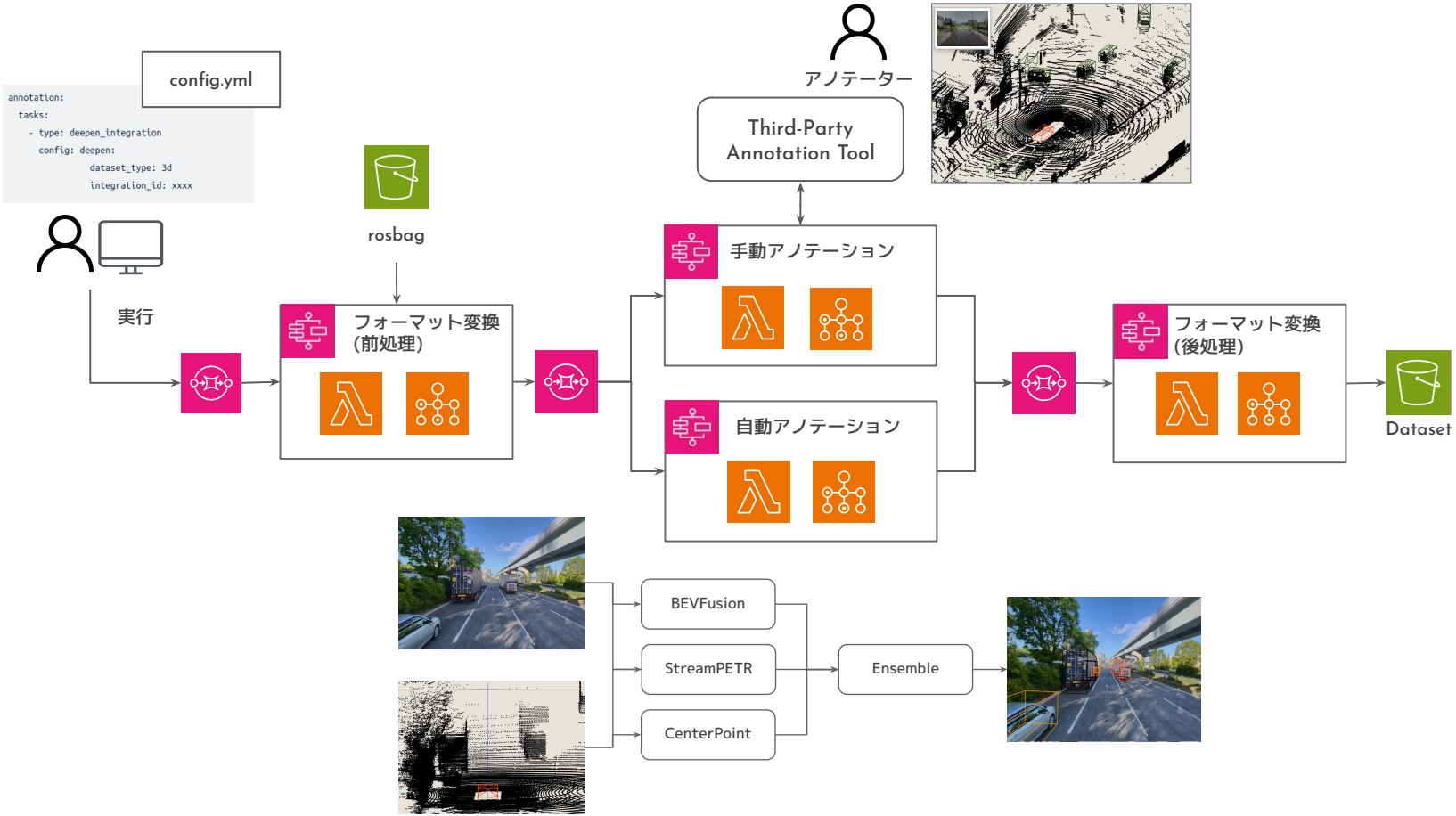
2D



Semantic Segmentation



データセット作成パイプライン

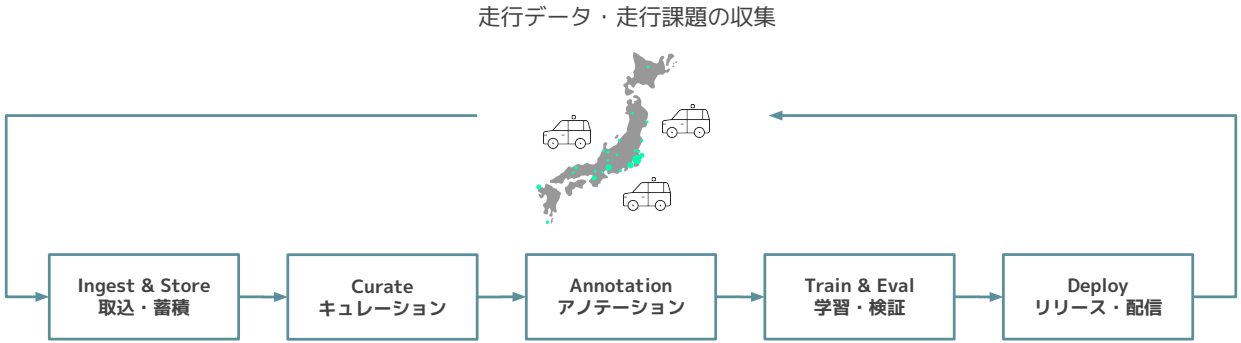




まとめ

05

まとめ



Ingest & Store

- 各データ種別に合わせたサーバーレスなデータパイプラインによる検索基盤
- 走行データの自動アーカイブ機能によるストレージコストの最適化

Curation

- 非構造化データの可視化および分析のパイプラインを構築
- VLMによるタグ付けやVideo Embeddingによる特定シーンの検索

Annotation

- 車両・用途ごとのデータ変換およびアノテーションツールの自動連携
- 自動アノテーションによるアノテーションコストの削減

次世代のEnd-to-End AI自動運転へ

データ基盤・MLOps基盤を通して次世代のEnd-to-End AI開発を加速



CONTACT US

https://tier4.jp/

Thanks Again !

「この成果は、国立研究開発法人新エネルギー・産業技術総合開発機構（ＮＥＤＯ）の助成事業
(JPNP21027)の結果得られたものです。」
Green Innovation Fund Projects / Development of In-vehicle Computing and Simulation Technology for Energy Saving in Electric Vehicles