

彩花 清水

PySparkデータエンジニア

✉ 彩花.清水@example.com

+81 90 712 31

📍 静岡

概要

PySparkデータエンジニアとしてロボットセンサーデータ処理のデータ基盤を担当。PySpark、Delta Lake、Structured Streaming、Kafka、MLflowを使い、設計判断を運用の安定性と測定可能な成果につなげます。

職歴

PySparkデータエンジニア

Jan 2022 - 現在

駿河ロボティクス・静岡

- 駿河ロボティクスのロボットセンサーデータ処理向けにPySparkとDelta Lakeの本番データパイプラインを構築。プロダクト・基盤チームと3,600台のロボット規模で運用し、定刻完了率を92%から99.6%へ改善。
- 「ロボット特徴量ストア」でStructured Streamingモデルとパーティション設計を刷新し、9名のエンジニアと移行を推進。クエリの中央値を24秒から4.1秒へ短縮。
- ロボットセンサーデータ処理の本番テーブル38個にスキーマ・鮮度・lineageの検証を計測。14名のアナリストと合意したルールでデータ不整合を63%削減。
- Kafkaジョブのbackfill、アラート、runbookを自動化し、12市場と「ロボット特徴量ストア」を支援。オンコール対応を四半期86時間削減。

データ基盤エンジニア

Apr 2019 - Dec 2021

駿河ロボティクス データスタジオ・静岡

- 駿河ロボティクスのロボットセンサーデータ処理向けにPySparkとDelta Lakeの移行データパイプラインを構築。プロダクト・基盤チームと3,600台のロボット規模で運用し、定刻完了率を92%から99.6%へ改善。
- 「ロボット特徴量ストア」でStructured Streamingモデルとパーティション設計を刷新し、6名のエンジニアと移行を推進。クエリの中央値を24秒から4.1秒へ短縮。
- ロボットセンサーデータ処理の本番テーブル38個にスキーマ・鮮度・lineageの検証を計測。5つのアプリチームと合意したルールでデータ不整合を63%削減。
- Kafkaジョブのbackfill、アラート、runbookを自動化し、4事業部と「ロボット特徴量ストア」を支援。オンコール対応を四半期86時間削減。

プロジェクト

ロボット特徴量ストア

ロボットセンサーデータ処理の課題に合わせたデータプロダクトを設計し、テスト済みの変換処理と分析向けの運用手順を整備。

PySpark

Delta Lake

Structured Streaming

Kafka

スキル

技術

PySpark

Delta Lake

Structured Streaming

Kafka

MLflow

デリバリー

データモデリング

データ品質

CI/CD

可観測性

ドキュメント

協働

要件ワークショップ

インシデント対応

メンタリング

アジャイル開発



学歴

静岡大学

Sep 2015 - Jun 2019

情報工学学士 at データエンジニアリングとソフトウェアシステム



資格・認定

- Databricks Associate Developer for Spark
専門認定機関

May 2024

- Data Reliability Workshop
Data Engineering Guild

Aug 2023



言語

日本語 • ネイティブ

英語 • ビジネスレベル