

七海 藤田

Databricksデータエンジニア



七海.藤田@example.com



+81 90 707 31



京都

概要

Databricksデータエンジニアとして製造lakehouseとML特徴量のデータ基盤を担当。Databricks、Delta Lake、DLT、PySpark、Unity Catalog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

- Databricksデータエンジニア** 京匠ファクトリー 1/2022 - 現在
 - 京匠ファクトリーの製造lakehouseとML特徴量向けにDatabricks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1日5.2TBのセンサーデータ規模で運用し、定刻完了率を92%から99.6%へ改善。
 - 「設備保全lakehouse」でDLT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。
 - 製造lakehouseとML特徴量の本番テーブル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。
 - PySparkジョブのbackfill、アラート、runbookを自動化し、12市場と「設備保全lakehouse」を支援。オンコール対応を四半期86時間削減。
- アナリティクスエンジニア** 京匠ファクトリー データスタジオ 4/2019 - 12/2021
 - 京匠ファクトリーの製造lakehouseとML特徴量向けにDatabricks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1日5.2TBのセンサーデータ規模で運用し、定刻完了率を92%から99.6%へ改善。
 - 「設備保全lakehouse」でDLT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。
 - 製造lakehouseとML特徴量の本番テーブル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。
 - PySparkジョブのbackfill、アラート、runbookを自動化し、4事業部と「設備保全lakehouse」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

製造lakehouseとML特徴量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。

Databricks
Delta Lake
DLT
PySpark

学歴

情報科学修士

京都大学

9/2015 - 6/2019

スキル

技術

Databricks, Delta Lake, DLT, PySpark, Unity Catalog

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

言語
日本語 ネイティブ
英語 ビジネスレベル

Databricks Certified Data Engineer Professional

専門認定機関
2024

Data Reliability Workshop

Data Engineering Guild
2023