

Sofie Kjær

PySpark Data Engineer

✉ sofie.kjær@example.com

☎ +45 20 412 71

📍 Silkeborg

Resumé

PySpark Data Engineer med fokus på sensorbehandling til robotter. Bygger stabile dataprodukter med PySpark, Delta Lake, Structured Streaming, Kafka, MLflow og forbinder tekniske valg med målbare driftsresultater.

Erhvervserfaring

PySpark Data Engineer

Jan 2022 - Nuværende

Prærie Robotics · Silkeborg

- Designede PySpark- og Delta Lake-pipelines til sensorbehandling til robotter hos Prærie Robotics; samarbejdede med produkt- og platformteams om 3.200 robotenheder, så leveringen til tiden steg fra 92 % til 99,6 %.
- Ombyggede Structured Streaming-modeller og partitionering til projektet Feature-lake til robotsensorer; koordinerede et 9-personers ingeniørteam og reducerede medianforespørgslen fra 24 til 4,1 sekunder.
- Instrumenterede skema-, friskheds- og lineage-kontroller på 38 produktionstabeller for sensorbehandling til robotter; arbejdede med 14 analytikere, hvilket sænkede datahændelser med 63 %.
- Automatiserede backfills, alarmer og runbooks omkring Kafka-jobs i Silkeborg; understøttede 12 markeder og sparede 86 vagtimer pr. kvartal.

Data Platform Engineer

Apr 2019 - Dec 2021

Prærie Robotics Data Studio · Silkeborg

- Designede PySpark- og Delta Lake-pipelines til sensorbehandling til robotter hos Prærie Robotics; samarbejdede med produkt- og platformteams om 3.200 robotenheder, så leveringen til tiden steg fra 92 % til 99,6 %.
- Ombyggede Structured Streaming-modeller og partitionering til projektet Feature-lake til robotsensorer; koordinerede et 6-personers ingeniørteam og reducerede medianforespørgslen fra 24 til 4,1 sekunder.
- Instrumenterede skema-, friskheds- og lineage-kontroller på 38 produktionstabeller for sensorbehandling til robotter; arbejdede med 5 applikationsteams, hvilket sænkede datahændelser med 63 %.
- Automatiserede backfills, alarmer og runbooks omkring Kafka-jobs i Silkeborg; understøttede 4 forretningsenheder og sparede 86 vagtimer pr. kvartal.

Projekter

Feature-lake til robotsensorer Udviklede et dokumenteret dataprodukt til sensorbehandling til robotter med testede transformationer og en analyseklar leveringsmodel.

PySpark Delta Lake
Structured Streaming
Kafka

Færdigheder

Teknisk stack

- PySpark
- Delta Lake
- Structured Streaming
- Kafka
- MLflow

Levering

- Datamodellering
- Datakvalitet
- CI/CD
- Observability
- Dokumentation

Samarbejde

- Kravworkshops
- Incident response
- Mentoring
- Agile levering

Uddannelse

Professionshøjskolen Absalon Sep 2015 - Jun 2019
Bachelor i datalogi at Data engineering og software

Certificeringer

- Databricks Associate Developer for Spark**
Fagligt certificeringsråd

May 2024
- Data Reliability Workshop**
Data Engineering Guild

Aug 2023

Sprog

- Dansk • Modersmål
- Engelsk • Professionelt arbejdssprog