Sofia Lindholm

PySpark Data Engineer

sofia.lindholm@example.com +1 555 0122 Minneapolis, MN

Summary

PySpark Data Engineer focused on sensor processing for robotics. Builds reliable data products with PySpark, Delta Lake, Structured Streaming, Kafka, MLflow, connects engineering decisions to measurable operations, and communicates clearly with cross-functional teams.

Work Experience

PySpark Data Engineer

Jan 2022 - Present

Prairie Robotics · Minneapolis, MN

• Designed PySpark and Delta Lake pipelines for sensor processing for robotics at Prairie Robotics; while partnering with product, platform, and finance leads, processing 3,200 robot units, and raised on-time delivery from 92% to 99.6%.
• Reworked Structured Streaming models and partitioning for the Robot Sensor Feature Lake warehouse; coordinated a 9-person engineering group, cutting median query time from 24 seconds to 4.1 seconds.
• Instrumented schema, freshness, and lineage checks across 38 production tables for sensor processing for robotics; worked with 14 analysts, reducing data incidents by 63%.
• Automated backfills, alerts, and runbooks around Kafka jobs in Minneapolis, MN; supported 12 markets and saved 86 on-call hours per quarter.

Data Platform Engineer

Apr 2019 - Dec 2021

Prairie Robotics Data Studio · Minneapolis, MN

• Designed PySpark and Delta Lake pipelines for sensor processing for robotics at Prairie Robotics; with analysts, SREs, and domain owners across the delivery squad, processing 3,200 robot units, and raised on-time delivery from 92% to 99.6%.
• Reworked Structured Streaming models and partitioning for the Robot Sensor Feature Lake warehouse; coordinated a 6-person engineering group, cutting median query time from 24 seconds to 4.1 seconds.
• Instrumented schema, freshness, and lineage checks across 38 production tables for sensor processing for robotics; worked with 5 application teams, reducing data incidents by 63%.
• Automated backfills, alerts, and runbooks around Kafka jobs in Minneapolis, MN; supported 4 business units and saved 86 on-call hours per quarter.

Projects

Robot Sensor Feature Lake

Built a governed data product for sensor processing for robotics, with documented ownership, tested transformations, and an analyst-ready serving layer.

PySpark Delta Lake Structured Streaming Kafka

Skills

Core stack

PySpark · Delta Lake · Structured Streaming · Kafka · MLflow

Delivery

Data modeling · Data quality · CI/CD · Observability · Documentation

Collaboration

Requirements workshops · Incident response · Mentoring · Agile delivery

Education

University of Minnesota — Sep 2015 - Jun 2019

Bachelor of Science at Computer Science and Data Systems

Certifications

Databricks Associate Developer for Spark — May 2024

Professional Training Board

Data Reliability Workshop — Aug 2023

Data Engineering Guild

Languages

English • Native

Swedish • Professional working proficiency