

유나 홍

PySpark 데이터 엔지니어

유나.홍@example.com

+82 10 2412 8112

청주



요약

PySpark 데이터 엔지니어로서 로봇 센서 데이터 처리 데이터 제품을 운영해 왔습니다. PySpark, Delta Lake, Structured Streaming, Kafka, MLflow를 활용해 안정적인 파이프라인과 측정 가능한 개선을 연결합니다.



업무 경험

PySpark 데이터 엔지니어

Jan 2022 - 현재의

직지로보틱스 · 청주

- 직지로보틱스의 로봇 센서 데이터 처리를 위해 PySpark-Delta Lake 기반 운영 파이프라인을 설계하고 15명 팀과 204개 소스를 관리해 정시 완료율을 91%에서 99.5%로 높였습니다.
- '로봇 피쳐 스토어'에서 Structured Streaming 파티션과 증분 모델을 개선하고 개발자 8명과 신규 릴리스를 진행해 중앙값 쿼리 시간을 28초에서 4.2초로 줄였습니다.
- 로봇 센서 데이터 처리 테이블 36개의 스키마-신선도 검사를 자동화하고 분석가 8명과 운영 규칙을 합의해 데이터 불일치를 70% 낮췄습니다.
- Kafka 작업의 재처리-알림-런북을 운영 체계로 정비해 6개 사업부와 운영하고 분기 장애 대응 시간을 86시간 절감했습니다.

데이터 플랫폼 엔지니어

Apr 2019 - Dec 2021

직지로보틱스 데이터스튜디오 · 청주

- 직지로보틱스의 로봇 센서 데이터 처리를 위해 PySpark-Delta Lake 기반 이관 파이프라인을 설계하고 15명 팀과 204개 소스를 관리해 정시 완료율을 91%에서 99.5%로 높였습니다.
- '로봇 피쳐 스토어'에서 Structured Streaming 파티션과 증분 모델을 개선하고 개발자 8명과 이관 릴리스를 진행해 중앙값 쿼리 시간을 28초에서 4.2초로 줄였습니다.
- 로봇 센서 데이터 처리 테이블 36개의 스키마-신선도 검사를 자동화하고 분석가 8명과 마이그레이션 규칙을 합의해 데이터 불일치를 70% 낮췄습니다.
- Kafka 작업의 재처리-알림-런북을 전환 체계로 정비해 6개 사업부와 운영하고 분기 장애 대응 시간을 86시간 절감했습니다.



프로젝트

로봇 피쳐 스토어
로봇 센서 데이터 처리 문제를 해결하는 데이터 제품을 설계하고 테스트된 변환과 운영 절차를 구축했습니다.

PySpark

Delta Lake

Structured Streaming

Kafka



기술

기술

PySpark

Delta Lake

Structured Streaming

Kafka

MLflow

품질과 운영

데이터 모델링

데이터 품질

CI/CD

관측성

문서화

협업

요구사항 워크숍

장애 대응

멘토링

애자일 개발



교육

충북대학교

Mar 2015 - Feb 2019

컴퓨터공학 학사 at 데이터 엔지니어링과 분산 시스템



인증

• Databricks Associate Developer for Spark

May 2024

전문 인증기관

• Data Reliability Workshop

Aug 2023

Korea Data Guild



언어

한국어 • 원어민

영어 • 비즈니스