

Sofía Márquez

PySpark Data Engineer

 sofia.márquez@example.com

 +34 612 71 20 45

 Oviedo



Resumen

PySpark Data Engineer especializado en procesamiento de sensores robóticos. Construyo productos de datos fiables con PySpark, Delta Lake, Structured Streaming, Kafka, MLflow y conecto las decisiones técnicas con resultados operativos medibles.



Experiencia laboral

PySpark Data Engineer

Robótica Astur · Oviedo

Jan 2022 - Present

- Diseñé pipelines con PySpark y Delta Lake para procesamiento de sensores robóticos en Robótica Astur, junto a los equipos de producto y plataforma para procesar 3.200 robots conectados; subí la entrega a tiempo del 92% al 99,6%.
- Reconstruí modelos de Structured Streaming y particionado para el proyecto «Feature lake para robots», coordinando un equipo de 9 ingenieros; bajé la consulta mediana de 24 a 4,1 segundos.
- Instrumenté controles de esquema, freshness y lineage en 38 tablas de producción para procesamiento de sensores robóticos, con 14 analistas o equipos de aplicación; reduje los incidentes de datos un 63%.
- Automaticé backfills, alertas y runbooks alrededor de jobs de Kafka en Oviedo, dando soporte a 12 mercados o unidades; ahorré 86 horas de guardia por trimestre.

Data Platform Engineer

Robótica Astur Data Lab · Oviedo

Apr 2019 - Dec 2021

- Diseñé pipelines con PySpark y Delta Lake para procesamiento de sensores robóticos en Robótica Astur, junto a los equipos de producto y plataforma para procesar 3.200 robots conectados; subí la entrega a tiempo del 92% al 99,6%.
- Reconstruí modelos de Structured Streaming y particionado para el proyecto «Feature lake para robots», coordinando un equipo de 6 ingenieros; bajé la consulta mediana de 24 a 4,1 segundos.
- Instrumenté controles de esquema, freshness y lineage en 38 tablas de producción para procesamiento de sensores robóticos, con 5 analistas o equipos de aplicación; reduje los incidentes de datos un 63%.
- Automaticé backfills, alertas y runbooks alrededor de jobs de Kafka en Oviedo, dando soporte a 4 mercados o unidades; ahorré 86 horas de guardia por trimestre.



Proyectos



Desarrollé un producto de datos documentado para procesamiento de sensores robóticos, con transformaciones probadas y un modelo listo para análisis.

PySpark

Delta Lake

Structured Streaming

Kafka



Habilidades

Stack técnico

PySpark

Delta Lake

Structured Streaming

Kafka

MLflow

Entrega

Modelado de datos

Calidad de datos

CI/CD

Observabilidad

Documentación

Colaboración

Talleres de requisitos

Respuesta a incidentes

Mentoría

Entrega Agile



Educacion

Universidad de Oviedo

Sep 2015 - Jun 2019

Grado en Informática en Data engineering y software



Certifications

• **Databricks Associate Developer for Spark**

Consejo Internacional de Certificación

May 2024

• **Data Reliability Workshop**

Data Engineering Guild

Aug 2023



Idiomas

Español • Nativo

Inglés • Profesional