

Oana Stancu

Ingineră de date PySpark

oana.stancu@example.ro

+40 721 4321772

Târgu Mureș

Rezumat profesional

Ingineră PySpark în produse medicale digitale. Pregătește măsurători și rezultate de laborator pentru studii, respectând pseudonimizarea și reproducibilitatea cohortelor.

Experiență profesională

Ingineră de date PySpark

Senzor MedTech · Târgu Mureș

Jan 2022 - Present

- A repartitionat 2,8 miliarde de măsurători Delta Lake după pacient și lună; cu 5 cercetători a redus pregătirea cohortelor de la 2 ore la 27 de minute.
- A adăugat Structured Streaming pentru 74 de aparate de monitorizare; alertele validate au ajuns la laborator în 3 minute, nu după închiderea turei.
- A verificat 46 de reguli de pseudonimizare înainte de exportul cohortelor; echipa de conformitate a aprobat toate cele 18 livrări pentru studiu.
- A optimizat shuffle-ul pentru 12 joburi PySpark folosite de 3 echipe; rata joburilor care depășeau fereastra de procesare a coborât de la 17% la 2%.

Inginer de platformă de date

Analitica Clinică Mureș · Târgu Mureș

Apr 2018 - Dec 2021

- A curățat în Spark 31 de milioane de rezultate de laborator cu 4 biostatisticieni; probele duplicate din raport au scăzut cu 48%.
- A conectat 8 formate HL7 la modelul comun al clinicii; maparea noilor analize a scăzut de la 3 zile la 5 ore.
- A versionat codul și cohortele cu MLflow pentru 9 experimente; cercetătorii au reprodus toate cele 9 rezultate în auditul trimestrial.
- A pus teste de valori imposibile și unități pe 1,2 milioane de rânduri; erorile de conversie au fost prinse înainte de export în 7 săptămâni consecutive.

Proiecte

Pregătirea cohortelor pentru studii

Produs de date pentru măsurători medicale anonimizate și teste de laborator, cu surse controlate, transformări versionate și reguli de calitate verificabile.

PySpark

Delta Lake

Structured Streaming

Kafka

Competențe

Platforme și limbaje

PySpark

Delta Lake

Structured Streaming

Kafka

MLflow

Calitate și operare

Modelare de date

Testare

CI/CD

Monitorizare

Documentare

Colaborare

Clarificarea cerințelor

Revizuire de cod

Răspuns la incidente

Lucru între echipe



Studii

Universitatea de Medicină, Farmacie, Științe și Tehnologie George Emil Palade

Master în bioinformatică la măsurători medicale anonimizate și teste de laborator

Oct 2014 - Jun 2018



Certificări

● **Databricks Certified Associate Developer for Apache Spark**

Organizația emitentă

Mar 2025

● **Atelier de calitate a datelor: măsurători medicale anonimizate și teste de laborator**

Comunitatea Inginerilor de Date din România

Oct 2023



Limbi

Română • Nativ

Engleză • Avansat