

Kata Fekete

PySpark Data Engineer

 kata.fekete@example.com

 +36 30 712 31

 Szombathely



Összegzés

PySpark Data Engineer, aki robotikai szenzorfeldolgozás területén épít megbízható adatplatformokat. PySpark, Delta Lake, Structured Streaming, Kafka, MLflow használatával a technikai döntéseket mérhető üzleti eredményekké alakítom.



Munkatapasztalat

PySpark Data Engineer

PannonRobot · Szombathely

Jan 2022 - Jelen

- Felépítettem a PySpark és Delta Lake alapú adatfolyamokat a robotikai szenzorfeldolgozás kezelésére a PannonRobot jelenlegi csapatában; a termék- és platformmérnökökkel 3600 robotegység léptékben dolgoztam, így a határidőre elkészült futások aránya 92%-ról 99,6%-ra nőtt.
- Újraterveztem a Structured Streaming modelleket és particionálást a „Robotszenzor feature lake” projektben éles üzemeltetési felelősséggel; 9 fős mérnöki csapattal együttműködve a medián lekérdezési idő 24 másodpercről 4,1 másodpercre csökkent.
- Mérhetővé tettem a séma-, frissesség- és lineage-ellenőrzéseket 38 éles táblán a robotikai szenzorfeldolgozás környezetében; 14 elemzővel egyeztetve 63%-kal kevesebb adateltérést mértünk.
- Automatizáltam a backfill-folyamatokat, riasztásokat és runbookokat a Kafka futásaihoz Szombathely csapatában; 12 piacot támogattunk, és negyedévente 86 ügyeleti órát takarítottunk meg.

Data Platform Engineer

PannonRobot Adatstúdió · Szombathely

Apr 2019 - Dec 2021

- Felépítettem a PySpark és Delta Lake alapú adatfolyamokat a robotikai szenzorfeldolgozás kezelésére a PannonRobot korábbi csapatában; a termék- és platformmérnökökkel 3600 robotegység léptékben dolgoztam, így a határidőre elkészült futások aránya 92%-ról 99,6%-ra nőtt.
- Újraterveztem a Structured Streaming modelleket és particionálást a „Robotszenzor feature lake” projektben migrációs felelősséggel; 6 fős mérnöki csapattal együttműködve a medián lekérdezési idő 24 másodpercről 4,1 másodpercre csökkent.
- Mérhetővé tettem a séma-, frissesség- és lineage-ellenőrzéseket 38 éles táblán a robotikai szenzorfeldolgozás környezetében; 5 alkalmazáscsapattal egyeztetve 63%-kal kevesebb adateltérést mértünk.
- Automatizáltam a backfill-folyamatokat, riasztásokat és runbookokat a Kafka futásaihoz Szombathely csapatában; 4 üzleti egységet támogattunk, és negyedévente 86 ügyeleti órát takarítottunk meg.



Projektek

Robotszenzor
feature lake

Dokumentált adatterméket építettem a robotikai szenzorfeldolgozás igényeire, tesztelt transzformációkkal és elemzést támogató szállítási modellel.

PySpark

Delta Lake

Structured Streaming

Kafka



Készségek

Technológiák

PySpark

Delta Lake

Structured Streaming

Kafka

MLflow

Szállítás

Adatmodellezés

Adatminőség

CI/CD

Megfigyelhetőség

Dokumentáció

Együttműködés

Követelmény-workshop

Incidenskezelés

Mentorálás

Agilis szállítás



Oktatás

Széchenyi István Egyetem

Programtervező informatikus BSc at Adatmérnökség és szoftverrendszerek

Sep 2015 - Jun 2019



Tanúsítványok

- Databricks Associate Developer for Spark**
Szakmai tanúsító szervezet

May 2024

- Data Reliability Workshop**
Data Engineering Guild

Aug 2023



Nyelvek

Magyar • Anyanyelvi

Angol • Munkanyelv