

Nóra Papp

Databricks Data Engineer

nóra.papp@example.com +36 30 707 31 Kecskemét

Összegzés

Databricks Data Engineer, aki ipari lakehouse és ML-jellemzők területén épít megbízható adatplatformokat. Databricks, Delta Lake, DLT, PySpark, Unity Catalog használatával a technikai döntéseket mérhető üzleti eredményekké alakítom.

Munkatapasztalat

- Databricks Data Engineer

GépMérnök

1/2022 - Jelen

 - Felépítettem a Databricks és Delta Lake alapú adatfolyamokat a ipari lakehouse és ML-jellemzők kezelésére a GépMérnök jelenlegi csapatában; a termék- és platformmérnökökkel 5,2 TB szenzoradat naponta léptékben dolgoztam, így a határidőre elkészült futások aránya 92%-ról 99,6%-ra nőtt.
 - Újraterveztem a DLT modelleket és particionálást a „Karbantartási lakehouse” 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 ipari lakehouse és ML-jellemzők környezetében; 14 elemzővel egyeztetve 63%-kal kevesebb adateltérést mértünk.
 - Automatizáltam a backfill-folyamatokat, riasztásokat és runbookokat a PySpark futásaihoz Kecskemét csapatában; 12 piacot támogattunk, és negyedévente 86 ügyeleti órát takarítottunk meg.
- Analytics Engineer

GépMérnök Adatstúdió

4/2019 - 12/2021

 - Felépíttem a Databricks és Delta Lake alapú adatfolyamokat a ipari lakehouse és ML-jellemzők kezelésére a GépMérnök korábbi csapatában; a termék- és platformmérnökökkel 5,2 TB szenzoradat naponta léptékben dolgoztam, így a határidőre elkészült futások aránya 92%-ról 99,6%-ra nőtt.
 - Újraterveztem a DLT modelleket és particionálást a „Karbantartási lakehouse” 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 ipari lakehouse és ML-jellemzők környezetében; 5 alkalmazáscsapattal egyeztetve 63%-kal kevesebb adateltérést mértünk.
 - Automatizáltam a backfill-folyamatokat, riasztásokat és runbookokat a PySpark futásaihoz Kecskemét csapatában; 4 üzleti egységet támogattunk, és negyedévente 86 ügyeleti órát takarítottunk meg.

Oktatás

Mérnökinformatikus MSc
Neumann János Egyetem
9/2015 - 6/2019

Készségek

Technológiák

Databricks, Delta Lake, DLT, PySpark, Unity Catalog

Szállítás

Adatmodellezés, Adatminőség, CI/CD, Megfigyelhetőség, Dokumentáció

Együttműködés

Követelmény-workshop, Incidenskezelés, Mentorálás, Agilis szállítás

Nyelvek Magyar Anyanyelvi
Angol Munkanyelv

Databricks Certified
Data Engineer
Professional

Szakmai tanúsító szervezet
2024

Tanúsítványok

Data Reliability
Workshop

Data Engineering Guild
2023

Projektek

Karbantartási lakehouse
Dokumentált adatterméket építettem a ipari lakehouse és ML-jellemzők igényeire, tesztelt

Databricks
Delta Lake
DLT

transzformációkkal és elemzést támogató
szállítási modellel.

PySpark