

Noora Kallio

PySpark Data Engineer

✉ noora.kallio@example.com ☎ +358 40 712 31 📍 Lappeenranta

Yhteenveto

PySpark Data Engineer, joka keskittyy robottien sensorikäsittelyä. Rakennan luotettavia datatuotteita työkaluilla PySpark, Delta Lake, Structured Streaming, Kafka, MLflow ja yhdistän tekniset valinnat mitattaviin tuloksiin.

Työkokemus

PySpark Data Engineer

Jan 2022 - Nykyinen

Preeria Robotics · Lappeenranta

- Rakensin PySpark- ja Delta Lake-putket robottien sensorikäsittelyä-dataa varten yrityksessä Preeria Robotics nykyisessä roolissa; työskentelin tuote- ja alustatiimien kanssa 3 600 robottiyksikköä mittakaavassa ja nostin ajallaan valmistuvien ajojen osuuden 92 prosentista 99,6 prosenttiin.
- Uudistin Structured Streaming-mallit ja osioinnin projektissa Robottisensorien feature lake tuotantovastuulla; koordinoituani 9-henkistä insinööriä lyhensin mediaanikyselyn ajan 24 sekunnista 4,1 sekuntiin.
- Mittaroin skeema-, tuoreus- ja lineage-tarkistukset 38 tuotantotaululle robottien sensorikäsittelyä-ympäristössä julkaisuprosessin yhteydessä; tein yhteistyötä 14 analyytikon kanssa ja vähensin datapoikkeamia 63 prosenttia.
- Automatisoin backfillit, hälytykset ja runbookit Kafka-ajolle Lappeenrantaissa päivittäistä operointia varten; tuimme 12 markkinaa ja säästimme 86 päivystystuntia neljänneksessä.

Data Platform Engineer

Apr 2019 - Dec 2021

Preeria Robotics Data Studio · Lappeenranta

- Rakensin PySpark- ja Delta Lake-putket robottien sensorikäsittelyä-dataa varten yrityksessä Preeria Robotics aiemmassa roolissa; työskentelin tuote- ja alustatiimien kanssa 3 600 robottiyksikköä mittakaavassa ja nostin ajallaan valmistuvien ajojen osuuden 92 prosentista 99,6 prosenttiin.
- Uudistin Structured Streaming-mallit ja osioinnin projektissa Robottisensorien feature lake siirtymävaiheen vastuulla; koordinoituani 6-henkistä insinööriä lyhensin mediaanikyselyn ajan 24 sekunnista 4,1 sekuntiin.
- Mittaroin skeema-, tuoreus- ja lineage-tarkistukset 38 tuotantotaululle robottien sensorikäsittelyä-ympäristössä ylläpitovaiheen yhteydessä; tein yhteistyötä 5 sovellustiimin kanssa ja vähensin datapoikkeamia 63 prosenttia.
- Automatisoin backfillit, hälytykset ja runbookit Kafka-ajolle Lappeenrantaissa siirrettävää ylläpitoa varten; tuimme 4 liiketoimintayksikköä ja säästimme 86 päivystystuntia neljänneksessä.

Hankkeet

Robottisensorien feature lake	Rakensin dokumentoidun datatuotteen robottien sensorikäsittelyä-tarpeeseen, testatut muunnokset ja analytiikkaa tukevan toimitusmallin.	PySpark	Delta Lake
		Structured Streaming	Kafka

Taidot

Teknologia	PySpark	Delta Lake	Structured Streaming	Kafka	MLflow
Toimitus	Tietomallinnus	Datan laatu	CI/CD	Observability	Dokumentointi
Yhteistyö	Vaatimusworkshopit	Incident response	Mentorointi	Ketterä toimitus	

Koulutus

LUT-yliopisto	Sep 2015 - Jun 2019
Tietojenkäsittelyn kandidaatti at Data engineering ja ohjelmistot	

Sertifikaatit

Databricks Associate Developer for Spark	May 2024
Ammatillinen sertifiointielin	
Data Reliability Workshop	Aug 2023
Data Engineering Guild	

Kielet

Suomi • Äidinkieli	Englanti • Työkieli
--------------------	---------------------