

👤 個人總結

以 Databricks Lakehouse、Delta Live Tables 與 Unity Catalog 建立機器學習和 BI 共用的數據底座。熟悉作業效能、資料品質、特徵資料集與 MLflow，能把模型團隊的需求轉成可維運管線。

🏢 工作經歷

Databricks 數據工程師 曜星推薦科技

5/2022 - 至今

- 以 Delta Live Tables 重建推薦特徵管線，和 10 位資料科學家訂定 156 個欄位契約，每日 4.2TB 特徵產出準時率達 99.4%。
- 透過 Photon、Z-Ordering 與自動縮放優化 23 個 PySpark 工作，支援 1,100 萬會員行為回訓，計算成本降低 38%。
- 用 Unity Catalog 設計跨環境權限與血緣，協作資安、MLOps 團隊管理 300 張表，資料申請核准時間由 2 天降至 3 小時。
- 把 MLflow 模型輸入檢查與資料漂移指標接入 Workflows，服務 7 個推薦模型，回訓失敗率由 8.1% 降至 1.6%。

Spark 數據工程師 橙果智慧製造

9/2019 - 4/2022

- 用 Delta Lake 合併 860 台設備的感測資料，與製程工程師校對 72 種訊號，良率分析從隔日更新改為每 15 分鐘。
- 把 Kafka 串流接入 Structured Streaming，協作現場 IT 4 人團隊調整 watermark，異常停機告警提前 12 分鐘觸發。
- 以 DLT expectations 擋下缺值、時間倒退與重複事件，和 QA 團隊建立 44 項品質規則，無效資料進入模型的比例下降 91%。
- 用 Terraform 管理 Databricks workspace 與 cluster policy，支援 6 個工廠複製部署，上線準備時間從 3 週縮至 2 天。

🎓 教育背景

理學碩士

國立陽明交通大學

9/2017 - 6/2019

☆ 技能專長

Databricks

Delta Lake, Delta Live Tables, Unity Catalog, MLflow, Workflows

處理技術

PySpark, Python, SQL, Kafka, NVIDIA RAPIDS

平台與治理

AWS, Terraform, Data Quality, GitHub Actions, Docker

<>專案經驗

工廠預知維護湖倉

將設備感測、維修工單與零件資料建成可供模型與營運查詢的 Lakehouse。

特徵資料集註冊中心

提供模型團隊共用、可追蹤版本的特徵註冊流程。

語言能力
中文
英文

證書獎項
Databricks Certified Data Engineer Professional
Databricks
2024
Databricks Certified Machine Learning Associate
Databricks
2022