陳家豪 | 資深數據工程師

中國香港，香港島，香港，中環皇后大道中 99 號
+852 9123 4567 • kaho.chan@example.com • https://kahochan.dev
@kahochan-data • @kahochan-data

簡介

資深數據工程師，擁有超過八年設計與建置現代數據平台嘅經驗，專注於即時數據管道、數據倉儲同雲端數據架構。擅長將複雜業務需求轉化為可擴展、可靠且易維護嘅數據解決方案。

- 精通 Python、SQL、Spark 及 dbt，熟悉 Kafka、Airflow 同 Snowflake 等現代數據工具鏈
- 曾領導跨職能團隊完成多個大型數據遷移及實時分析項目，平均延遲降低 40%
- 重視數據品質、可觀測性同成本效益，致力推動數據治理最佳實踐

教育背景

香港科技大學 **https://www.ust.hk**

學士，計算機科學，成績：3.7 2012 年 9 月–2016 年 6 月

- 主修計算機科學，副修統計學，畢業成績優異
- 畢業論文探討分散式資料處理框架於金融風險分析之應用
- 曾獲院長嘉許名單及多項獎學金

課程：資料結構與演算法、資料庫系統、分散式系統、大數據分析、雲端運算、機器學習導論

工作經歷

智慧數據科技有限公司 **https://www.smartdata-tech.hk**

高級數據工程師 2021 年 3 月至今

- 帶領四人數據工程團隊，設計及維護每日處理超過 5 TB 數據嘅雲端數據平台
- 主導由傳統數據倉儲遷移至 Snowflake 嘅項目，查詢效能提升 3 倍，年度成本節省約 25%
- 建立基於 Airflow 同 dbt 嘅 ELT 管道，實現數據血緣追蹤同自動化測試，數據事故減少 60%
- 與數據科學團隊緊密合作，交付實時推薦系統所需嘅特徵管道，模型上線時間由數週縮短至數日

關鍵字：數據架構、Snowflake、Airflow、dbt、團隊領導

金融雲端服務有限公司 **https://www.fincloud.hk**

數據工程師 2017 年 7 月–2021 年 2 月

- 負責建置及優化銀行交易數據嘅批處理管道，每日處理超過 2 億筆記錄
- 使用 Spark 及 Kafka 開發近實時風險監控數據流，將欺詐偵測延遲由小時級降至分鐘級
- 設計分區及壓縮策略，使數據湖儲存成本降低 35%，同時維持查詢效能
- 參與數據治理委員會，制定數據品質指標及元數據管理標準

關鍵字：Spark、Kafka、數據湖、風險監控、數據治理

初創分析有限公司 **https://www.startup-analytics.hk**

數據分析師 2016 年 8 月–2017 年 6 月

- 為零售客戶建立銷售及庫存儀表板，支援每日營運決策
- 以 Python 及 SQL 自動化報表流程，每週節省約 15 小時人手工作
- 協助設計 A/B 測試框架，分析促銷活動成效並提出優化建議

關鍵字：Python、SQL、數據視覺化、A/B 測試

語言

粵語：母語或雙語水平

英語：完全專業水平 **關鍵字**：IELTS 8.0

普通話：完全專業水平

專業技能

數據工程：專家 — **關鍵字**：Python、SQL、Spark、Kafka、Airflow、dbt

雲端平台：高級 — **關鍵字**：AWS、GCP、Snowflake、BigQuery、Docker、Kubernetes

數據治理：高級 — **關鍵字**：數據品質、元數據管理、數據血緣、隱私合規、成本優化

榮譽

年度最佳數據項目

智慧數據科技有限公司 — 2022 年 12 月

表揚在 Snowflake 遷移項目中展現卓越技術領導力，顯著提升平台效能並降低營運成本。

院長嘉許名單

香港科技大學 — 2015 年 6 月

因學術成績優異而獲頒，每學年僅頒授予成績排名前 10% 的學生。

證書

AWS Certified Data Analytics –Specialty — **https://aws.amazon.com/certification/**

Amazon Web Services — 2022 年 8 月

Google Professional Data Engineer — **https://cloud.google.com/certification/data-engineer**

Google Cloud — 2020 年 11 月

出版刊物

香港科技大學計算機科學系 — **https://hdl.handle.net/123456789/1234**

分散式資料處理框架於金融風險分析之應用 — 2016 年 6 月

- 探討 Apache Spark 及 Flink 於即時風險計算嘅效能比較
- 提出基於微批次嘅優化策略，將延遲降低約 30%

項目

實時欺詐偵測管道 — **https://github.com/kahochan-data/fraud-pipeline**

基於 Kafka、Spark Structured Streaming 及 Redis 嘅實時欺詐偵測系統 — 2023 年 1 月–2023 年 8 月

- 設計及實作每秒處理 5,000 筆交易嘅串流管道，端到端延遲低於 2 秒
- 整合機器學習模型評分服務，實現自動化風險決策
- 建立監控及告警機制，確保管道可用性達 99.9%

關鍵字：Kafka、Spark、Redis、實時數據

數據品質自動化框架 — **https://github.com/kahochan-data/dq-framework**

以 dbt 及 Great Expectations 為核心嘅數據品質測試框架 — 2022 年 3 月–2022 年 12 月

- 提供宣告式數據品質規則，支援欄位級及表級驗證
- 自動產生數據品質報告並整合至 CI/CD 流程
- 於三個業務團隊推行，數據事故平均修復時間縮短 50%

關鍵字：dbt、Great Expectations、CI/CD、數據品質

興趣愛好

運動：龍舟、遠足、游泳

音樂：結他、爵士樂

志願服務

香港數據工程師社群 — **https://www.hkde.community**

導師及講者 — 2019 年 9 月至今

- 定期舉辦數據工程工作坊，分享雲端數據架構及管道設計經驗
- 指導初入行工程師，協助他們建立實戰項目及職涯規劃
- 籌辦年度數據馬拉松，吸引超過 200 名參加者