承翰 蔡

大數據工程師

chenghan.tsai@example.com

0968-731-402

台北市內湖區

> *"以 Hadoop、Spark 與 Kafka 處理高吞吐資料，熟悉分散式計算、資源佈署與數據可靠性。曾服務零售與電信場景，能將 PB 級資料治理拆成工程里程碑，並用業務指標驗證平台投資。"*

教育背景

國立臺灣科技大學

工程學士 · 資訊工程學系

2014.09 - 2018.06

技能專長

分散式平台

Hadoop　Spark　Kafka　Hive　YARN

程式語言

Scala　Java　Python　SQL　Shell

維運工具

Kubernetes　Prometheus　Grafana　Airflow　Linux

語言能力

中文	母語
英文	閱讀與技術溝通

證書獎項

Cloudera Certified Data Engineer

Cloudera

2021.06

Linux Foundation Kubernetes

工作經歷

大數據工程師

萬象電信數據中心　2020.10 - 至今

用 Kafka 與 Spark Structured Streaming 接收 2.8TB 日流量，和網路優化團隊共同定義尖峰指標，告警提前 18 分鐘發出。

調整 YARN queue、Spark shuffle 與 Hive 分區，支援 120 位分析使用者查詢 9PB 歷史資料，月度報表完成時間縮短 63%。

以 Airflow 編排 64 個跨叢集工作，和 SRE 共同建立 checkpoint 與重試策略，月度成功率從 96.2% 提升到 99.7%。

導入 Prometheus、Grafana 與容量預估模型，管理 36 台工作節點，硬體擴充採購提前 2 個月完成並避免尖峰降級。

分散式數據工程師

城北零售集團　2018.07 - 2020.09

以 Spark SQL 整合 4,200 家門市銷售與會員資料，和商品經理校正促銷口徑，日營運盤點提前 90 分鐘產出。

重寫 Hive 表的 bucketing 與壓縮策略，協作 7 位 BI 開發者處理 1.3PB 歷史資料，查詢掃描量降低 47%。

建立 Kafka topic 命名、保留與 ACL 規範，和應用團隊管理 180 條資料流，遺失訊息事件季度減少 82%。

用 Shell 與 Python 自動化叢集健康檢查，將凌晨人工檢查 3 小時改為 20 分鐘每日報告。

Linux Foundation
2022.03

專案經驗

電信網路品質資料湖
把基地台、客服與維修資料整合成可追溯的網路品質資料湖。
Kafka　Spark　Hive　HDFS

促銷彈性計算框架
依活動期間動態分配 Spark 資源，減少平日閒置成本。
YARN　Spark　Airflow　Grafana