

DataMesh架構應用實作 在 K8S 環境下進行讀寫分離和快取管理

Fred Chien, Principle Architect, Brobridge Taiwan

Oct 18, 2022

AGENDA

01

舊系統面臨的挑戰

02

舊資料系統的擴充強化方法

03

以 Data Mesh 思維擴充改造資料系統





舊系統面臨的挑戰

LEGACY SYSTEM'S CHALLENGES

新數位時代 的需求

快

慢

傳統 面臨的挑戰

- 快速作業部署
- 快速掌握狀況
- 快速搶佔市場

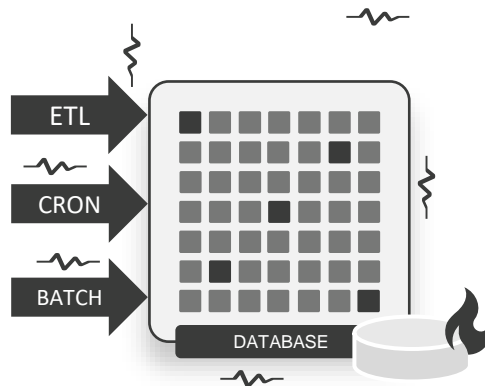
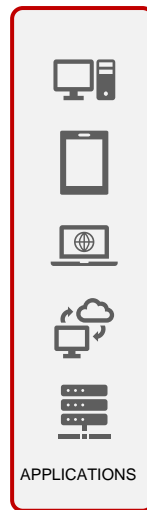
快

捷

- 敏捷調整資源
- 敏捷部署更新
- 敏捷開發新品

- 精準分析戰情
- 精準預估趨勢
- 精準做出決策

準



傳統企業常見的資料問題

COMMON PROBLEMS IN TRADITIONAL ENTERPRISE



缺乏彈性 伸縮能力

資料供應跟不上業務發展需求，應用服務擴展時資料庫無法跟著擴展。



資料時效 不夠即時

眾多 ETL Sprawl 排程作業依存關係嚴重彼此干擾，資料時效性差。



欠缺讀寫 分離機制

老舊系統沒能實作讀寫分離，嚴重衝擊核心系統效能，無法有效支撐重要業務系統。



資料無法 共享統一

傳統數據平台僅解決基礎資料整合和共用問題，應用系統之間仍然煙囪林立，資料價值不彰。



資料產品 開發耗時

資料來源與供應對接耗時費力，異質資料型態複雜無法迅速建立即時資料管線。



資料議題成為數位轉型旅程的隱憂

DATA FRAGMENTATION IS THE BIGGEST WORRY IN THE JOURNEY OF DIGITAL TRANSFORMATION

業務體驗差

分散割裂的資料導致業務流程改善異常困難，造成客戶體驗差，業務擴展不易。

阻礙商業決策

資料的可信度和質量下降，嚴重阻礙了資料的價值呈現，無法協助精準決策。

重複的成本投入

現存的資料不易重用，導致 IT 重複性投資以及應用程式基礎投入增加。

減緩業務創新的進程

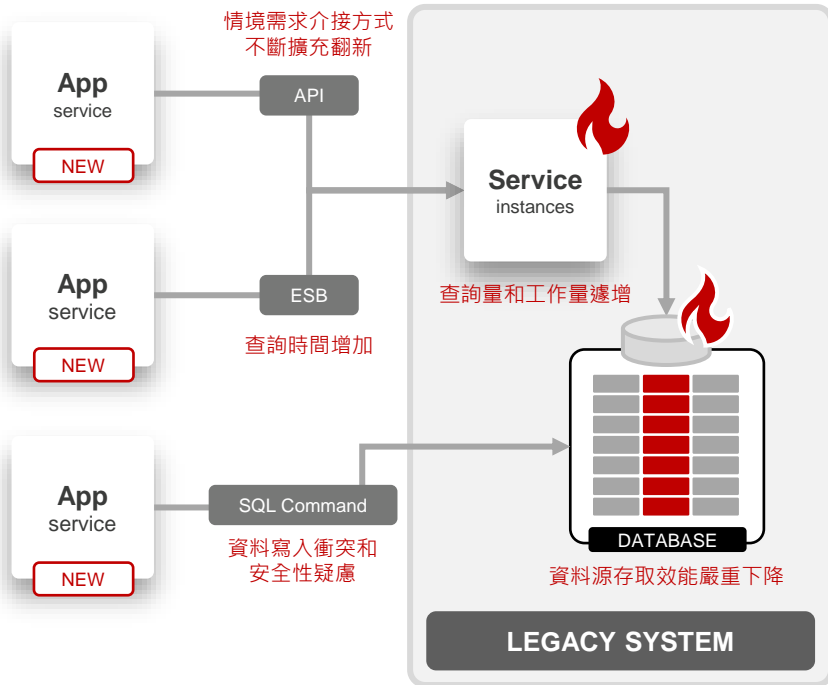
重複低效的 ETL 作業耗費了大量的人力與工時，導致業務建設投資成效不彰。



企業當務之急
是需要一套通
透且即時資料
產品的發佈治
理、能無痛且
高度整合既有
數據系統和工
具的解決方案。

舊系統遭遇的主要困境

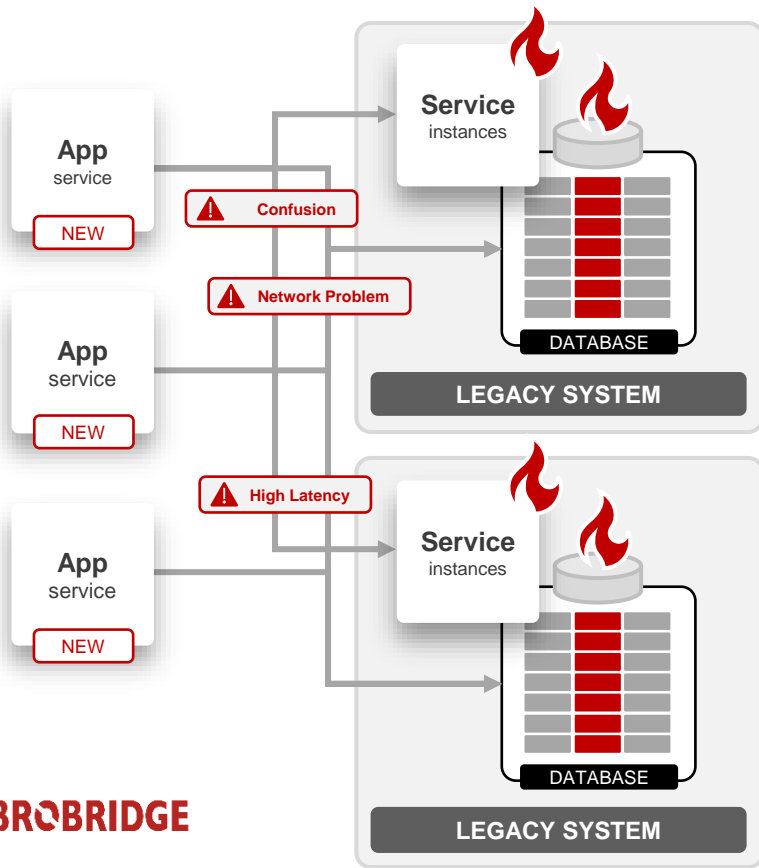
LEGACY SYSTEM 'S CHALLENGES



- ❑ 新業務系統效能表現不佳
- ❑ 資料庫查詢效能跟不上 新應用的擴展
- ❑ 應對新應用的 系統變更工作遽增
- ❑ 舊系統 服務壓力增加
- ❑ 資料交換的 分工權責問題
- ❑ 新應用 嚴重影響舊系統業務
- ❑ 系統擴充、改造有巨大包袱
- ❑ 需要 靠不即時的排程避開尖峰查詢時間

跨多個舊系統的資料源整合問題

DATA INTEGRATION WITH MULTIPLE LEGACY SYSTEMS



資料串接和稽核管理議題

- ❑ 管線 串接和部署方法混亂不統一
- ❑ 系統失效的 權責歸屬難以監控稽核

效能和橫向擴展問題

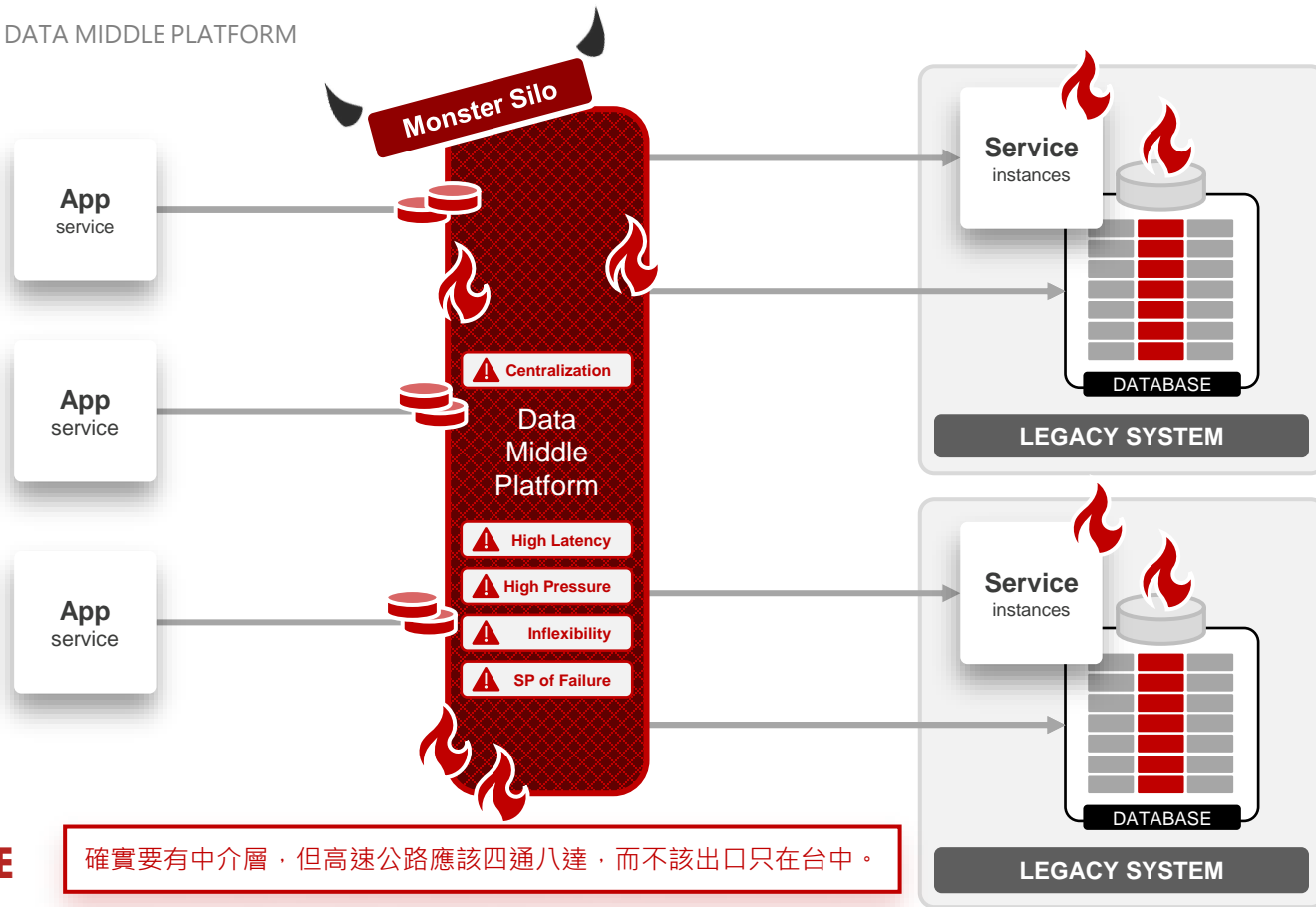
- ❑ 無法支撐大併發 查詢量
- ❑ 可能有 資料死鎖問題
- ❑ 查詢會造成 瞬間系統負載高

資料發佈和派送效率不彰

- ❑ 跨系統關聯性 查詢效能差
- ❑ 新應用 效能表現不彰
- ❑ 資料 即時性差
- ❑ 資料處理的 排程工作更難進行

近年來對「數據中臺」一詞的錯誤解讀所導致的痛苦

MISUNDERSTAND DATA MIDDLE PLATFORM





舊資料系統的擴充強化

ENHANCE LEGACY DATA SYSTEM



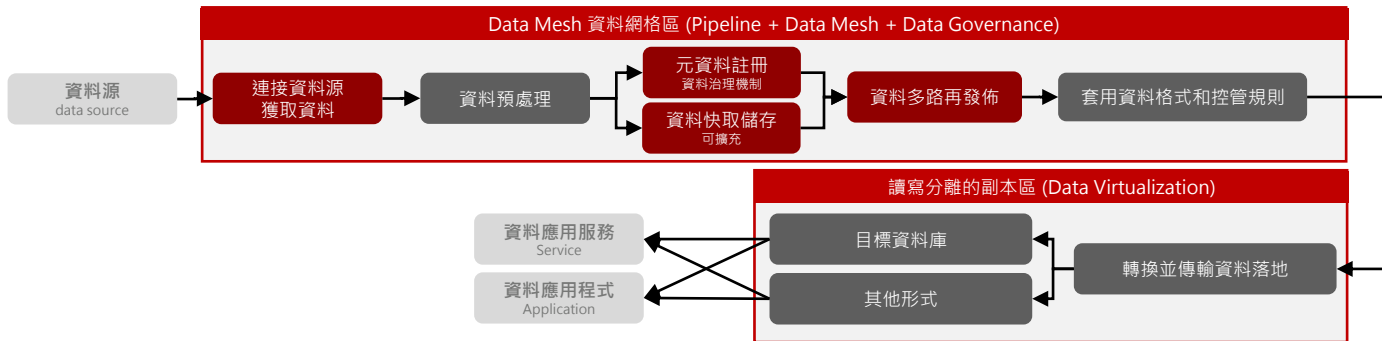
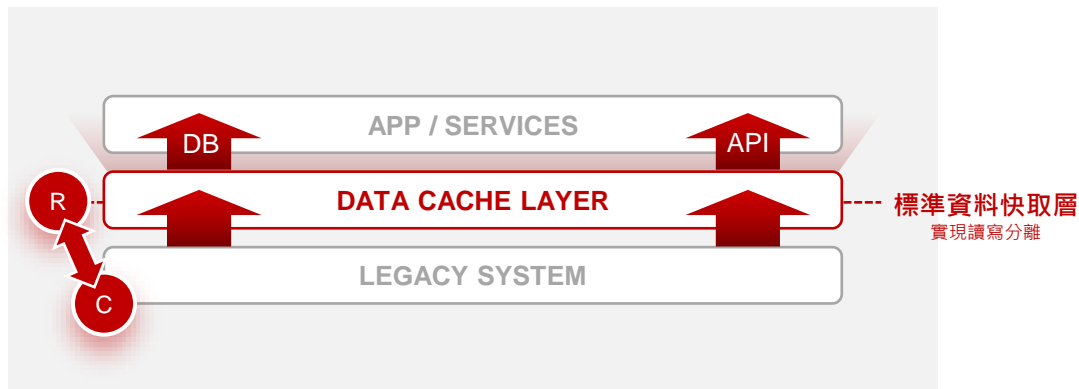
舊系統的擴充強化方向和方法論

CONCEPT FOR LEGACY SYSTEM ENHANCEMENT

數據中臺起手式：先把讀寫分離和資料快取做好。

怕做錯就從「快取(Cache)」無痛做起，進可攻退可守。

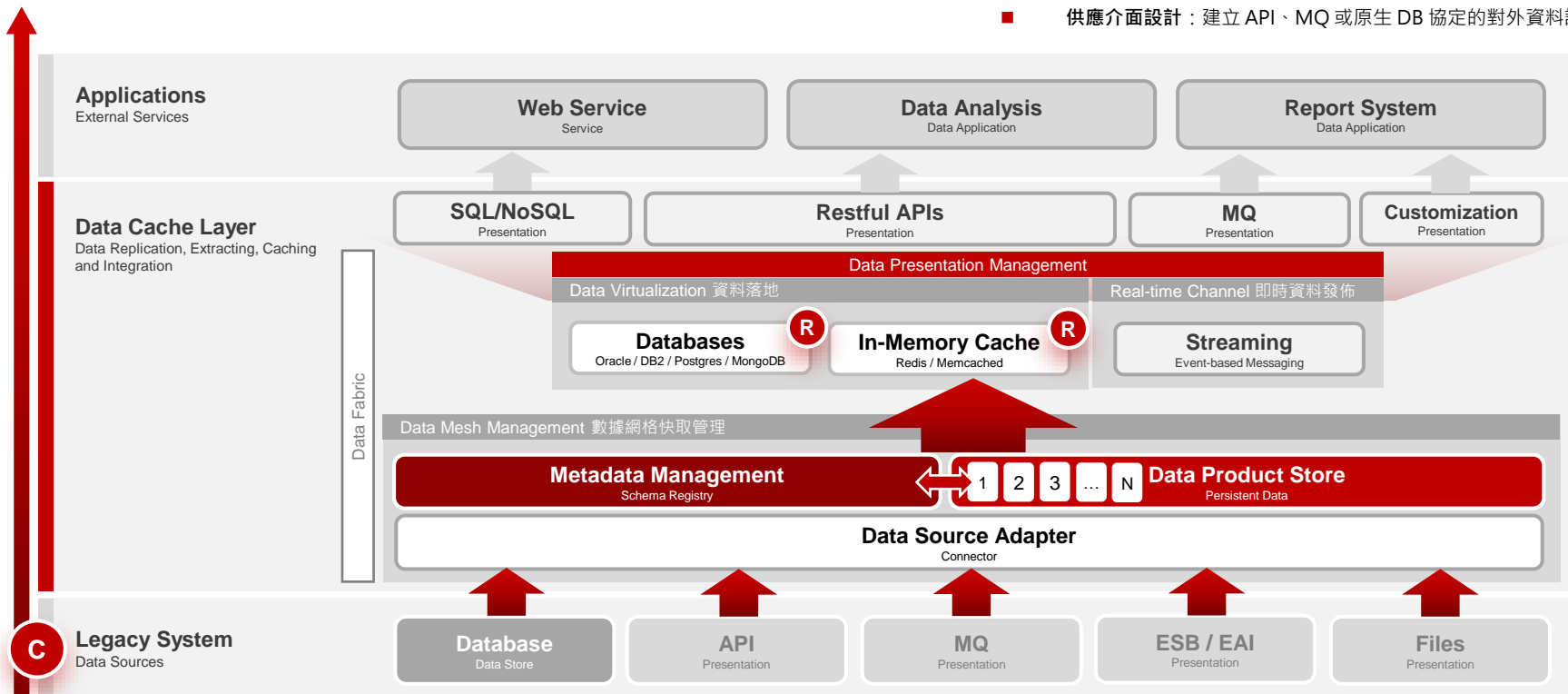
- ✓ 避免改動舊系統既有架構和功能 為前提
- ✓ 建構快取層 以讀寫分離機制擴大資料查詢效能
- ✓ 避免資料查詢工作衝擊舊系統
- ✓ 查詢吞吐量 能隨著需求量的增加擴充
- ✓ 盡可能 提供即時的資料
- ✓ 一次資料源串接提取，可建構多個快取副本
- ✓ 能滿足不同應用所需的資料查詢方法和習慣
 - API
 - SQL/NoSQL Command
 - Files
 - MQ
 - CDC



數據中台資料快取元件架構說明

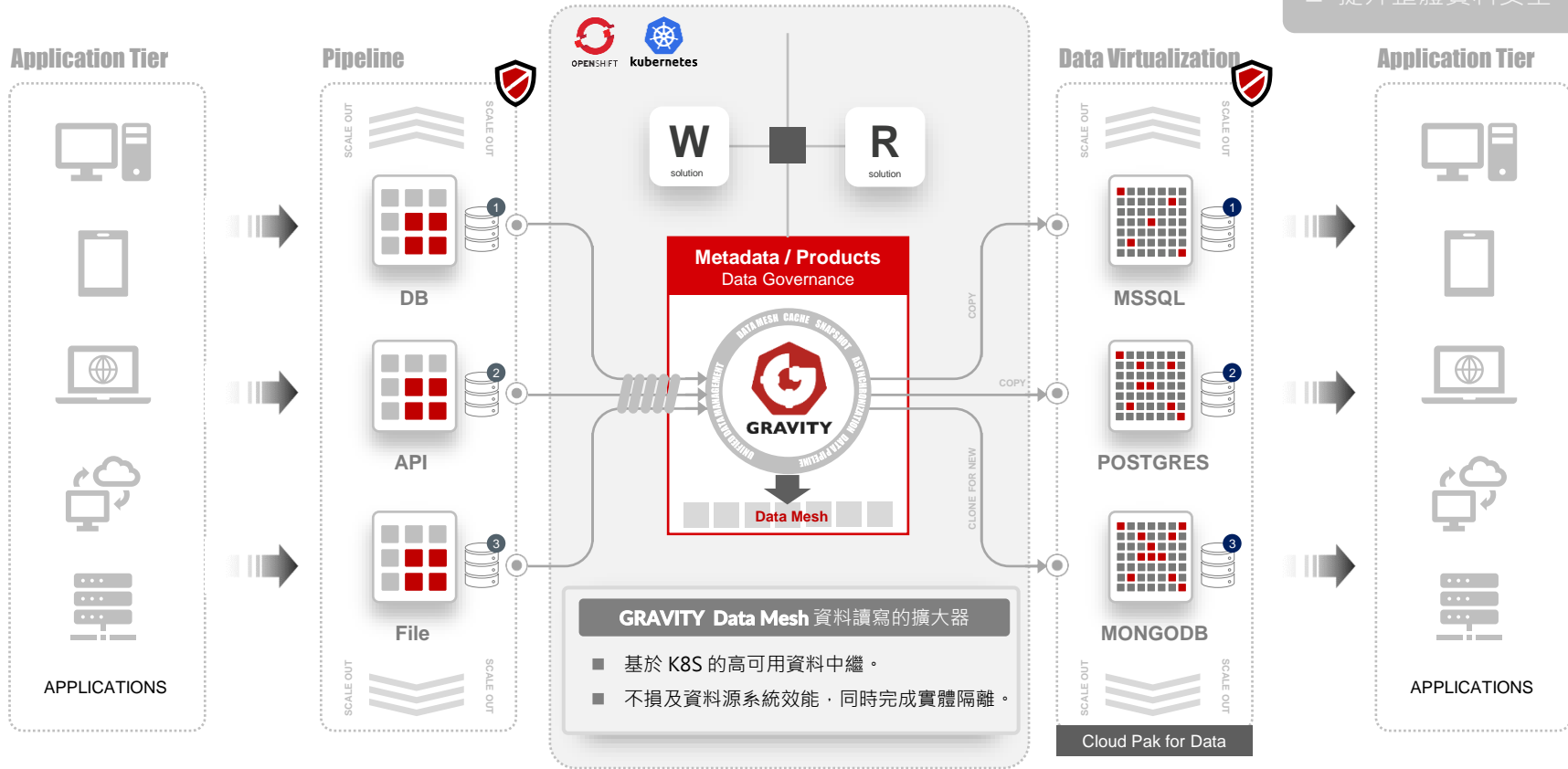
Pipeline + Data Mesh + Data Governance

- 資料源定義：以舊系統為資料源
- 攝取資料：透過資料事件（即時）或主動批次（排程）處理
- 預處理：針對資料進行預處理分類快取後並落地
- 供應介面設計：建立 API、MQ 或原生 DB 協定的對外資料讀取層



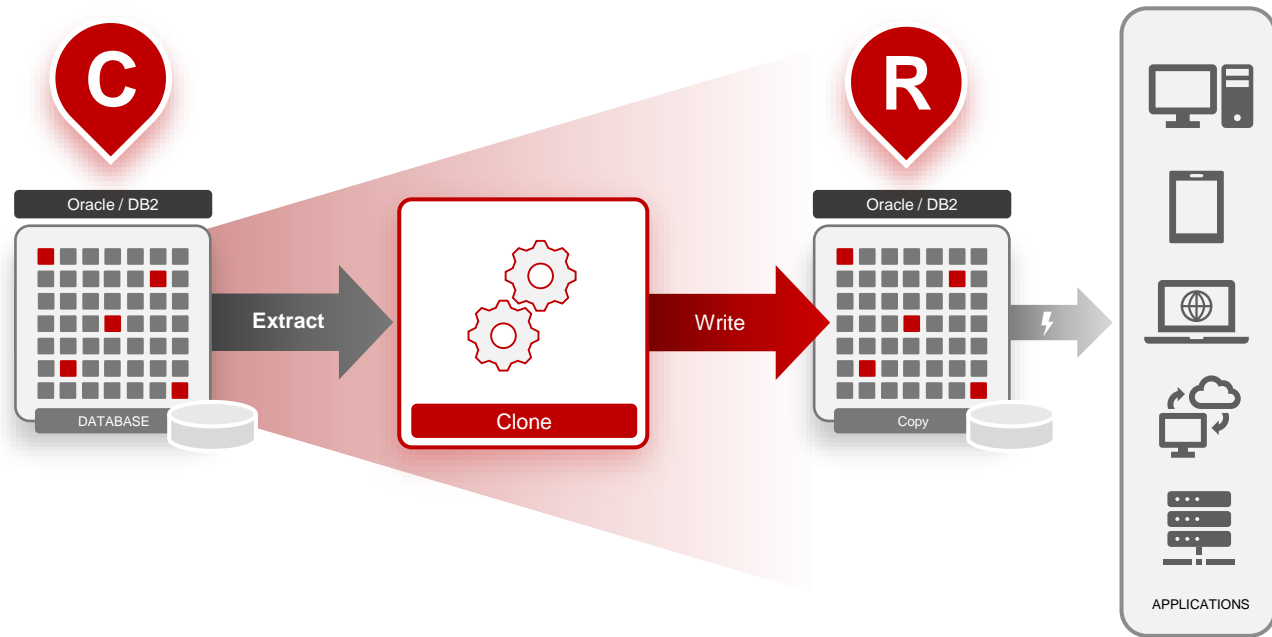
資料中繼、隔離以及彈性擴展資料供應能力

- 擴大既有投資效益
- 提升服務可用性
- 提升整體資料安全



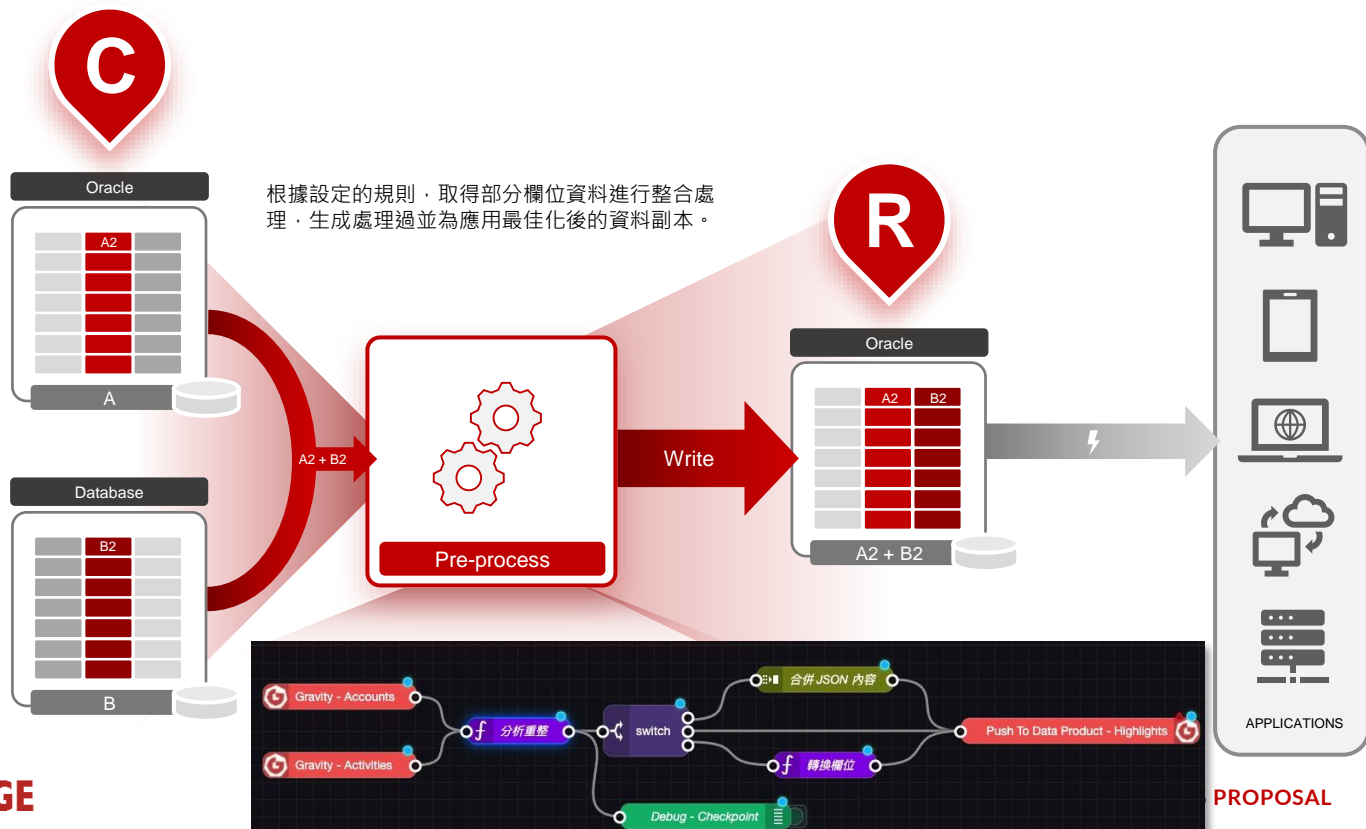
讀寫分離快取場景 1：單一目標對抄

CLONE DATA



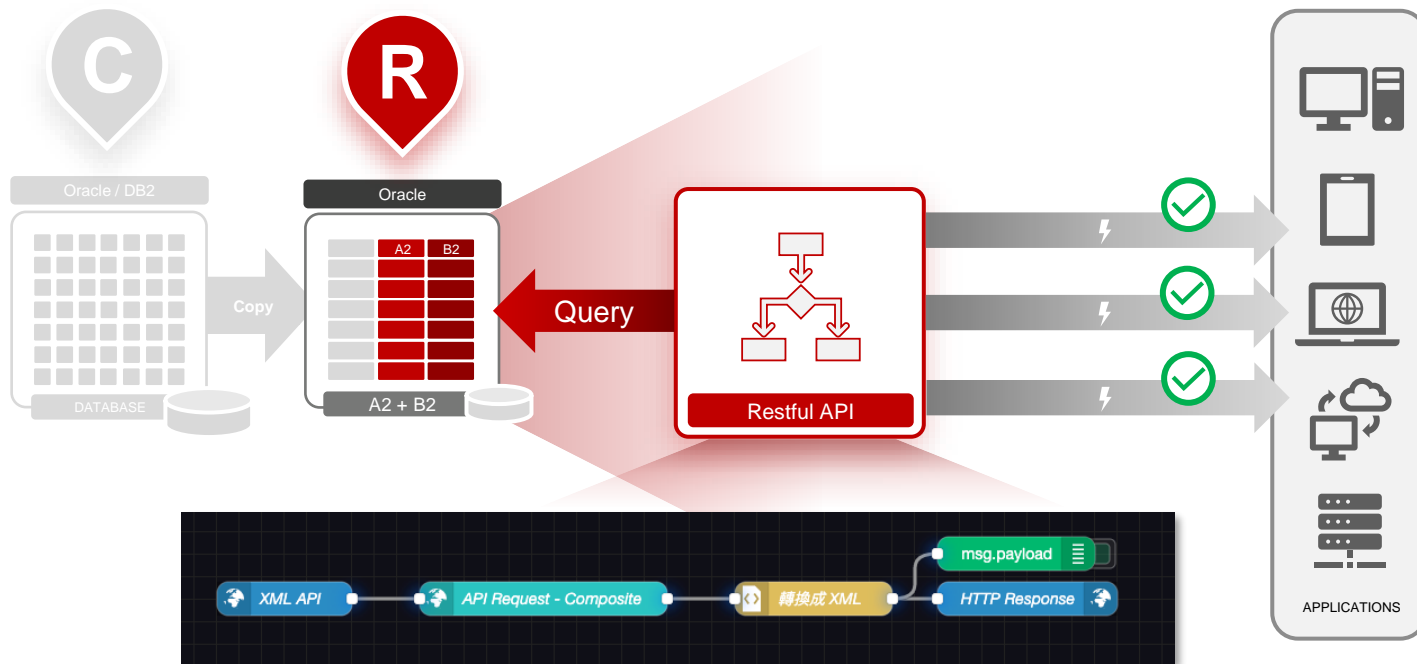
讀寫分離快取場景 2：即時資料過濾、拆分、整合後進行快取

DATA PRE-PROCESS AND CACHING



讀寫分離快取場景 3：快速客製化打造對外資料存取 API

CUSTOMIZED API FOR DATA ACCESS



你講的我都聽過 N 遍了

別廢話！

直接告訴我如何避免 Monster Silo ！



3



以 DATA MESH 思維擴充改造資料系統

POSSIBILITY OF DATA SYSTEM ENHANCEMENT WITH DATA MESH

回顧資料應用效率問題和改善路線

讀寫分離解決方案優缺點比較



優 架構單純直觀

缺 對於資料應用來說，資料過於原始，處理工作複雜度高。資料源工作頻繁且忙碌，查詢效能不彰，擴展能力不佳。

舊式快取方法：引入預處理和快取架構



優 定期預處理資料，提升資料查詢效率。

缺 資料結果更新速度慢，影響後續業務產能。密集資料處理集中在特定時間，會衝擊資料源系統效能。

每份快取副本：都需要處理一次串接資料源的問題



優 資料結果即時性高，資料查詢效率也高。

缺 資料處理過程產生的結果，難以保存管理並重複利用。無法保持即時性時，又同時滿足整批資料需求的應用。

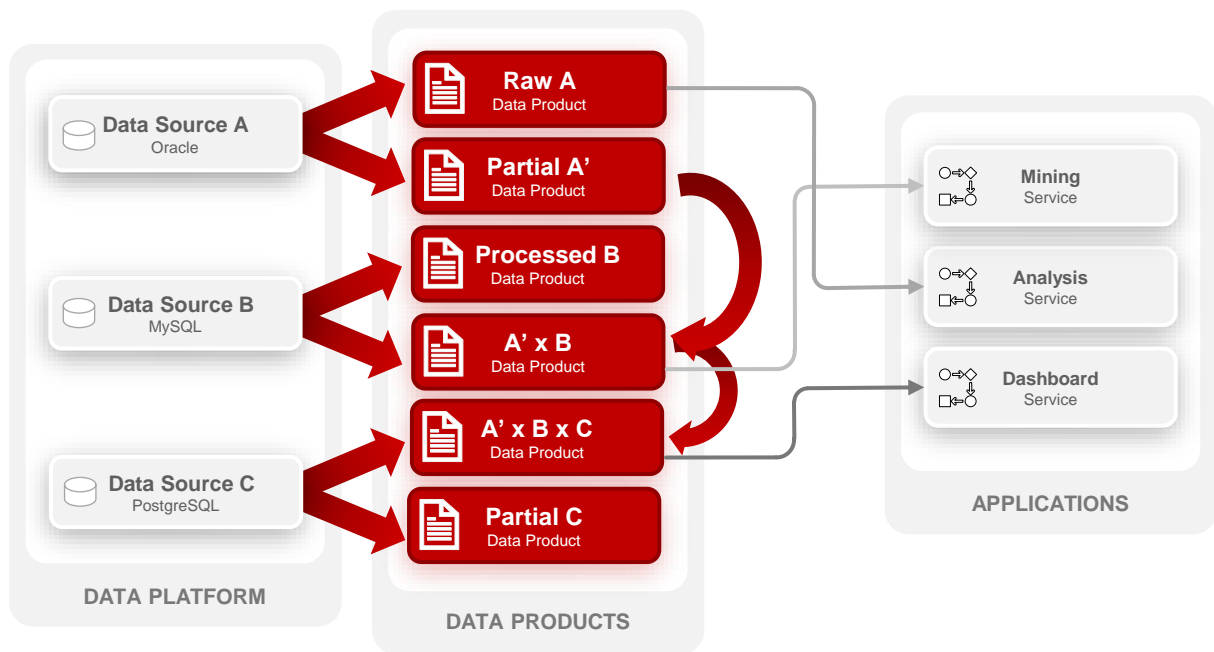
引入數據網格資料快取架構



- ✓ 資料即時性高
- ✓ 資料查詢效率高
- ✓ 資料重用性高
- ✓ 能同時滿足即時和整批資料查詢之需求
- ✓ 有標準資料管理和授權機制
- ✓ 橫向擴展能力佳
- ✓ 可供應給更多應用程式

導入 DATA MESH 實現資料產品管理平台改造資料平台

將資料即時預處理、邏輯運算、快取後，發佈到管理平台上提供應用程式訂閱拉取使用



- **近乎即時預處理所有資料**
無需分批、排程處理資料，滿足即時需求
- **流水線設計，提高資料重複利用率**
成果可共用，避免重複的邏輯運算和處理工作
- **節省大量終端應用的資料查詢時間**
資料已預處理，因此最終的查詢複雜度低
- **資料源沒有任何壓力**
資料源只有在資料變更時拋出事件的工作
- **應用程式數量，不影響資料源系統**
應用程式從管理平台取得資料，不碰觸源頭
- **架構容易橫向擴展**
增加管理平台的運算和儲存資源就可提高吞吐量

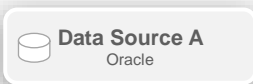
資料即服務的三大要素：運算單元、資料產品、資料庫即服務

實現多元通透的資料產品發佈治理 · 破除資料孤島問題

Business Unit A

Export partial data as data product to data catalog

DATA DOMAIN A



Provision



Provision



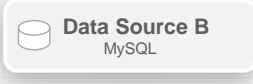
Real-time Streaming

A 地

Business Unit B

Provision raw data and process data from data source.

DATA DOMAIN B



Provision



Real-time Streaming

B 區



Provision



Real-time Streaming

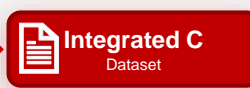
Business Unit C

Integrate data from other domain with local data.

DATA DOMAIN C

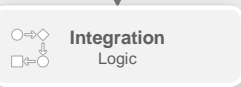


Provision



Real-time Streaming

C 叢集

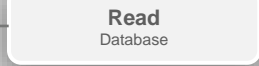
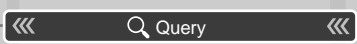
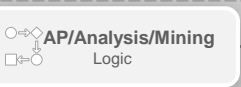


Data User

Requires data across domains.

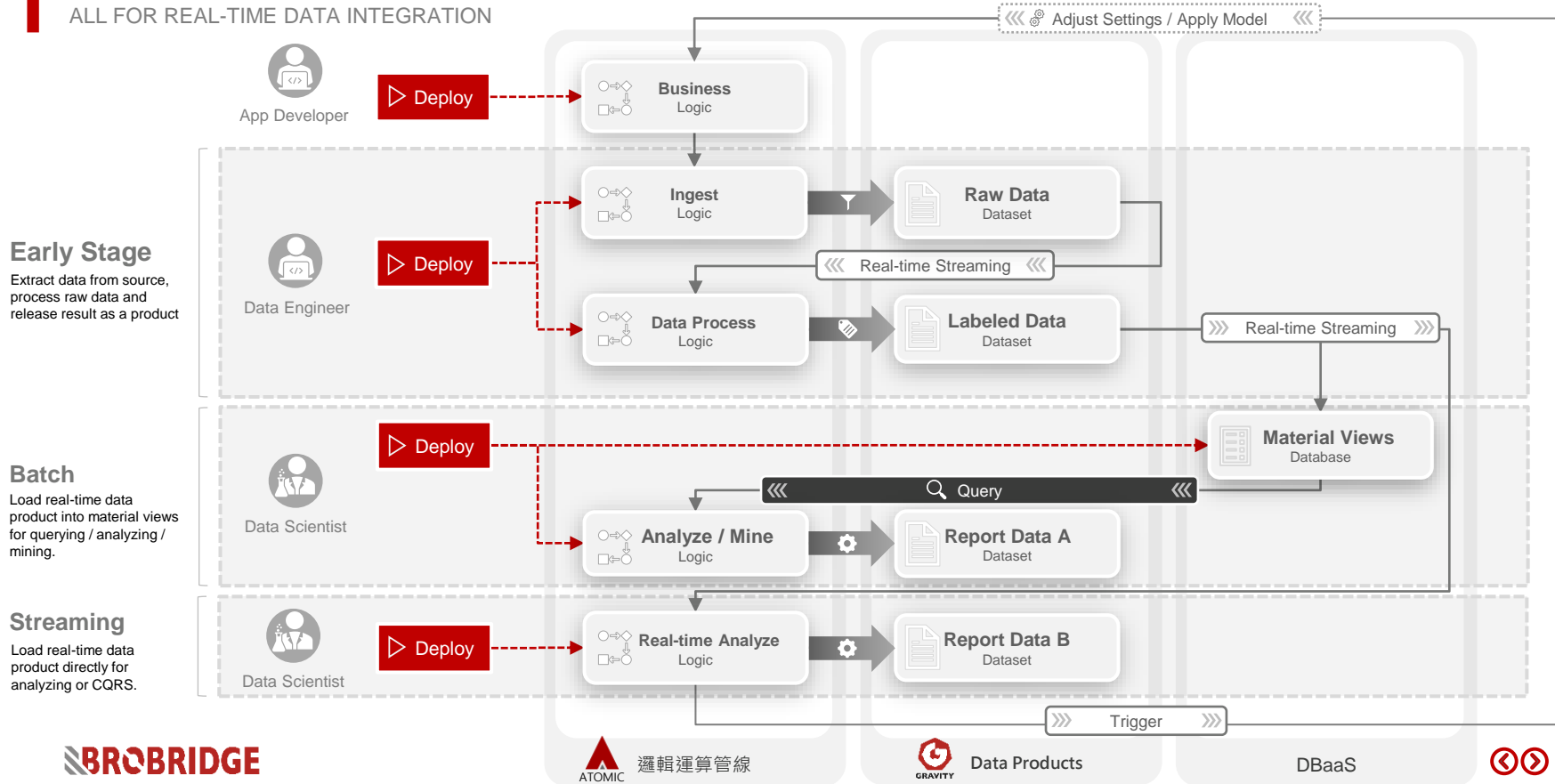


Results



必須同時滿足即時串流與批次的數據作業需求

ALL FOR REAL-TIME DATA INTEGRATION



可各自獨立運作的資料處理流程與系統參考架構設計

DATA FLOW AND SYSTEM ARCHITECTURE

客製化資料收集

因應不同資料的匯入方法，可客製化不同的擷取機制。

資料產品上架

將資料彙整成統一形式，上架成為「隨選即用」的資料產品。

即時的多元查詢副本

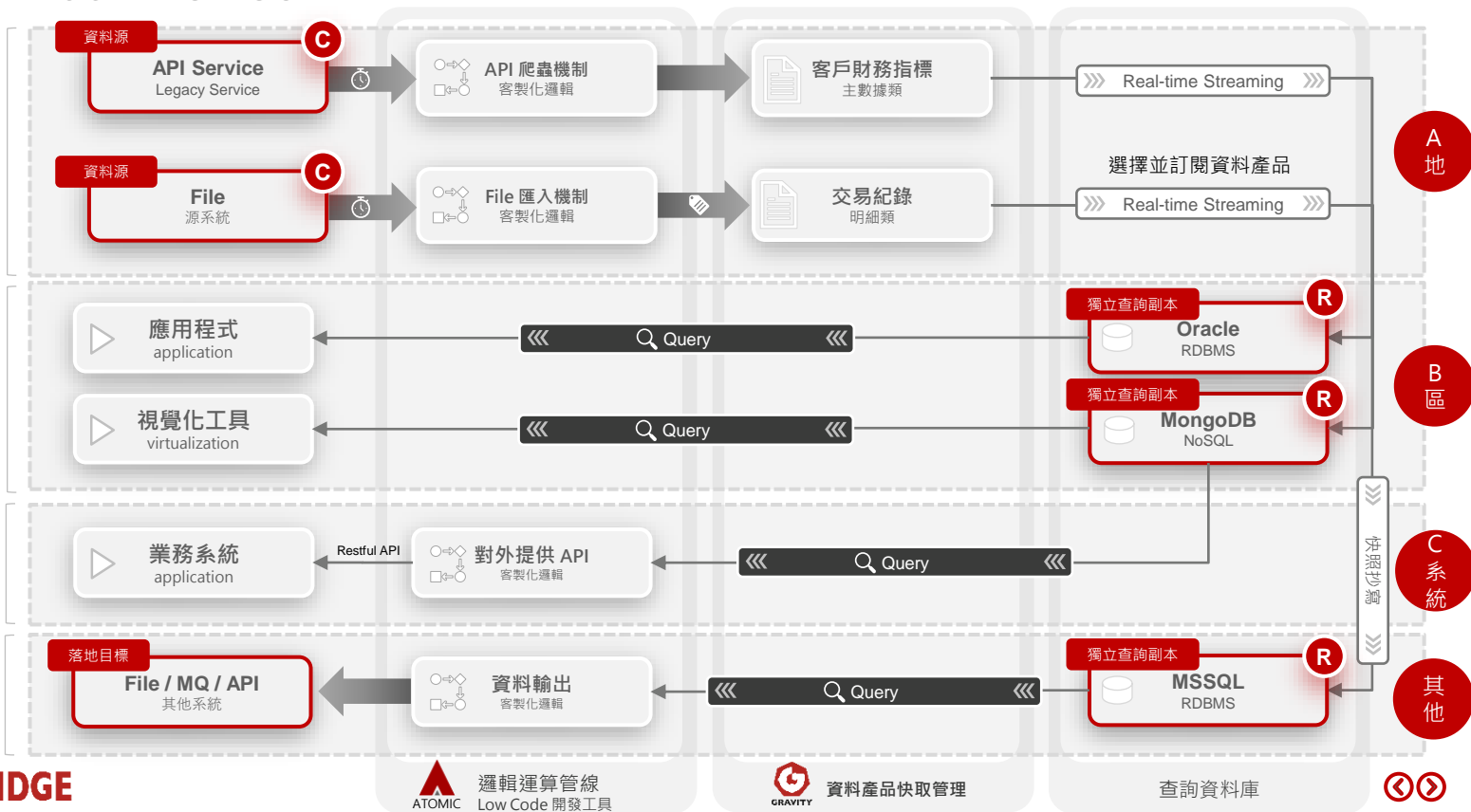
可因應大量查詢需求，建立多個擁有即時資料的資料副本。並依據應用程式所需「支援多種異質資料庫」。

可客製化 API 介面

可客製化 API 並提供給外部應用程式使用。

資料輸出可以客製化

可以依據需求，客製化將資料以任何形態輸出。

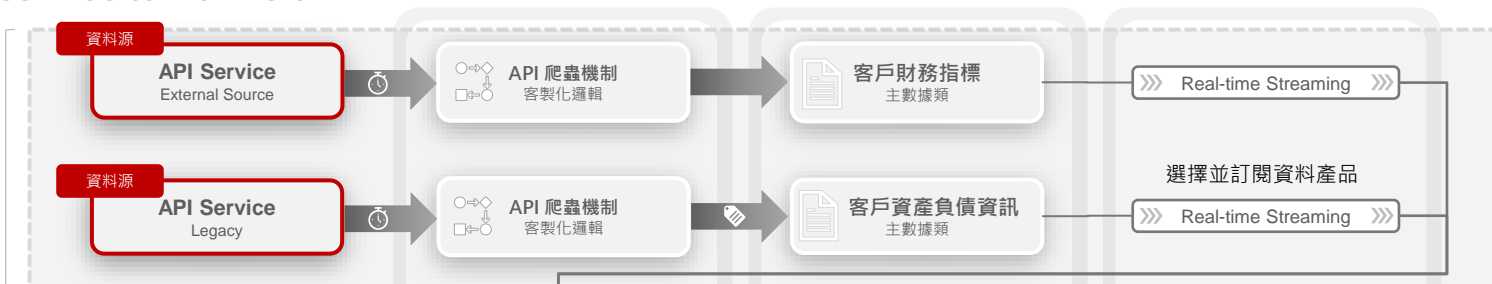


彈性的資料產品加工、分析產線之架構設計

DATA PRODUCT PROCESS ARCHITECTURE

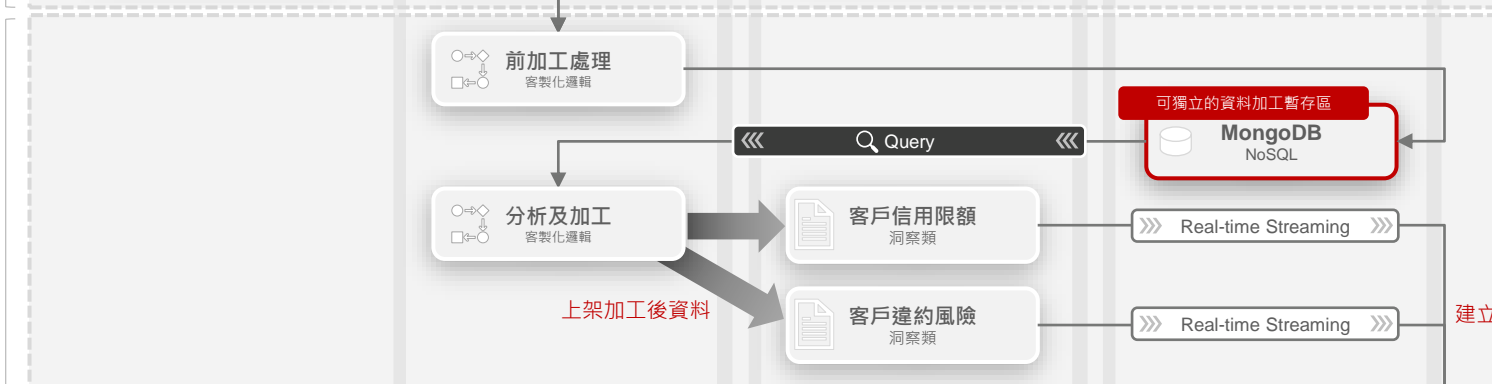
原始資料匯入

以客製化邏輯進行多資料匯入，並將資料歸檔、歸類上架。



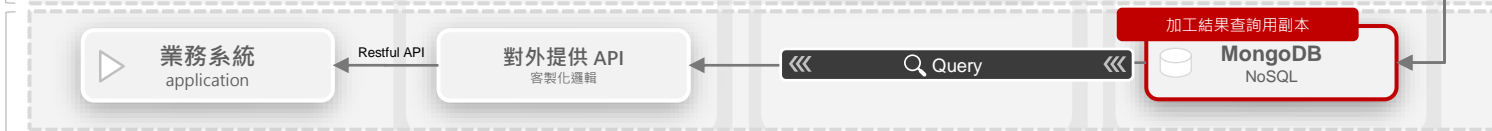
多層資料加工產出

獨立的資料加工區，可設計多層邏輯彈性進行即時加工和分析，並將加工結果上架，提供其他應用邏輯「重複使用」。



仍然有獨立的查詢副本

仍然可以建立獨立的查詢副本，以即時接收最新的分析加工結果，大量查詢也不影響加工作業效能。

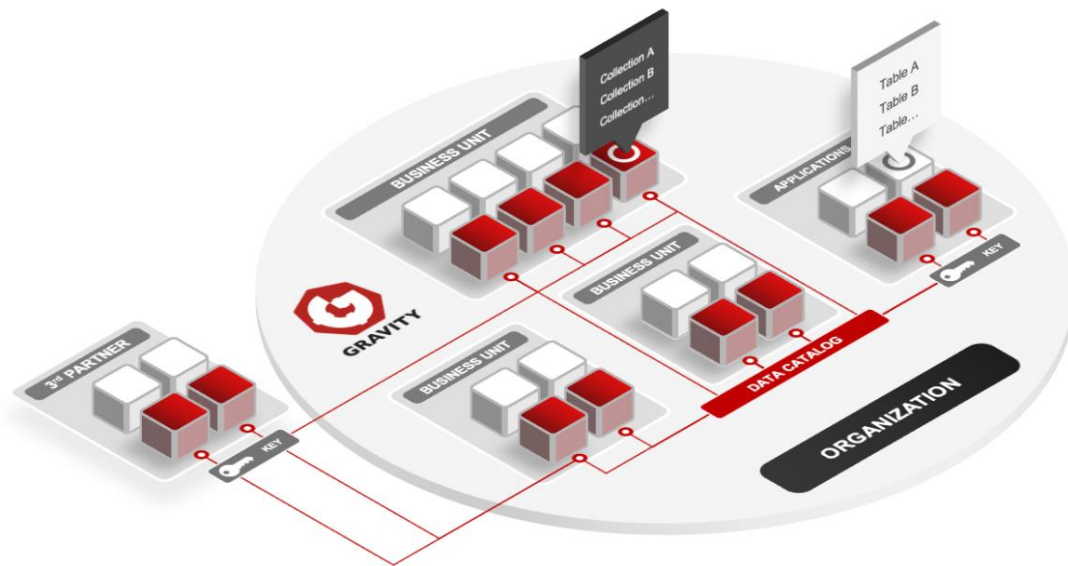


利用數據中台解決現代化數據發佈、轉運傳遞議題

DATA MESH GOVERNANCE

- 滿足企業級管理需求
- 實現資料自助服務
- 提升資料合規性
- 充分發揮投資效益最大化

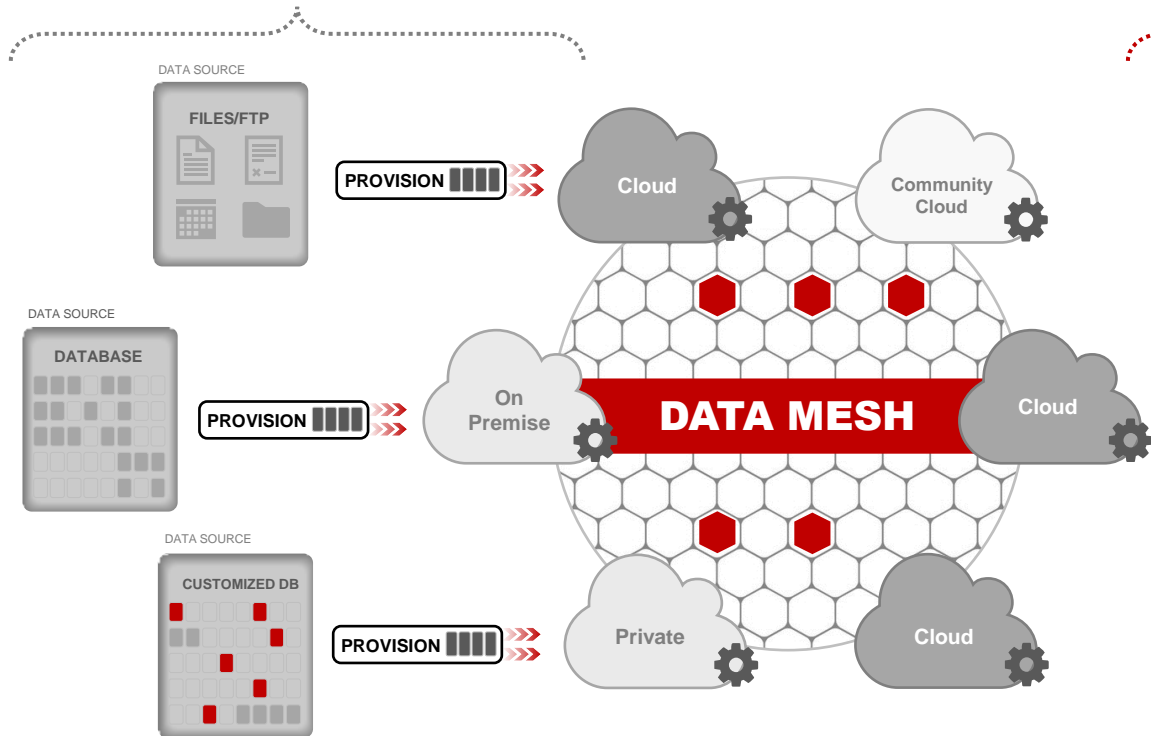
- 透過數據網格的 Data Catalog 機制，資料網路的訂閱者，可以無需考慮原始資料所在，**隨選訂閱所需要的資料**。
- 資料的交換 **不再需要大量的串接源頭客製化**，也可以滿足大量的資料取用需求，輕易的開放近乎即時的資料供外部系統、第三方系統所使用。
- **資料轉運倉機制**，可以事先透過 Data Fabric 網路，發佈到離應用服務最近 Data Mesh 發佈節點，就近進行快取副本建構。



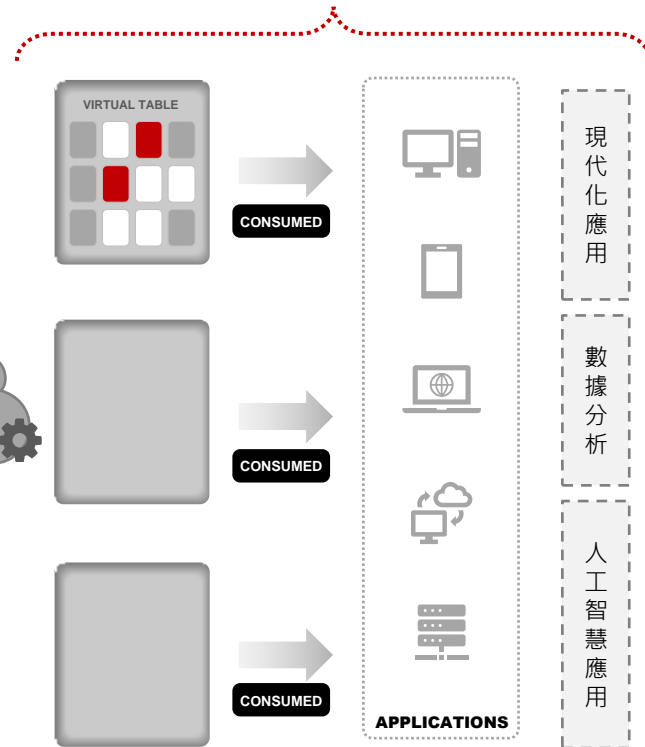
真正的中台核心關鍵是實現「數據物流管理」，而不是實現巨大倉庫和拓寬馬路
專心在設計與建構貨車調度、集貨倉、轉運倉、衛星倉、收發貨窗口



煙囪式(Silo based)數據架構



分散式(Distributed)數據架構



Thanks!

按讚 + 追蹤 臉書專頁
不定期技術文章分享

關於寬橋有限公司

Brobridge (“寬橋”) 泛指寬橋有限公司 (一家根據台灣法律組成的有限公司) 。請參閱 www.brobridge.com 中有關寬橋有限公司的詳細描述。寬橋有限公司為各行各業之上市及非上市客戶提供資訊技術諮詢、應用程式開發、IT基礎架構設計、顧問服務及系統維運服務。寬橋有限公司網羅業界最專業人士，致力於追求卓越，樹立典範。

關於寬橋有限公司

寬橋 (Brobridge) 係指寬橋有限公司 (Brobridge Co., Ltd) 。寬橋有限公司以卓越的客戶服務、優秀的人才、完善的訓練及嚴謹的開發工藝於業界享有良好聲譽。透過寬橋有限公司之資源，提供客戶資訊系統全方位的服務，包括虛擬桌面系統、多雲整合、IT基礎架構規劃設計、先進前沿程式開發等。

本出版物係依一般性資訊編寫而成，僅供讀者參考之用。寬橋有限公司不因本出版物而被視為對任何人提供專業意見或服務。對信賴本出版物而導致損失之任何人，寬橋有限公司之任一個體均不對其損失負任何責任。

