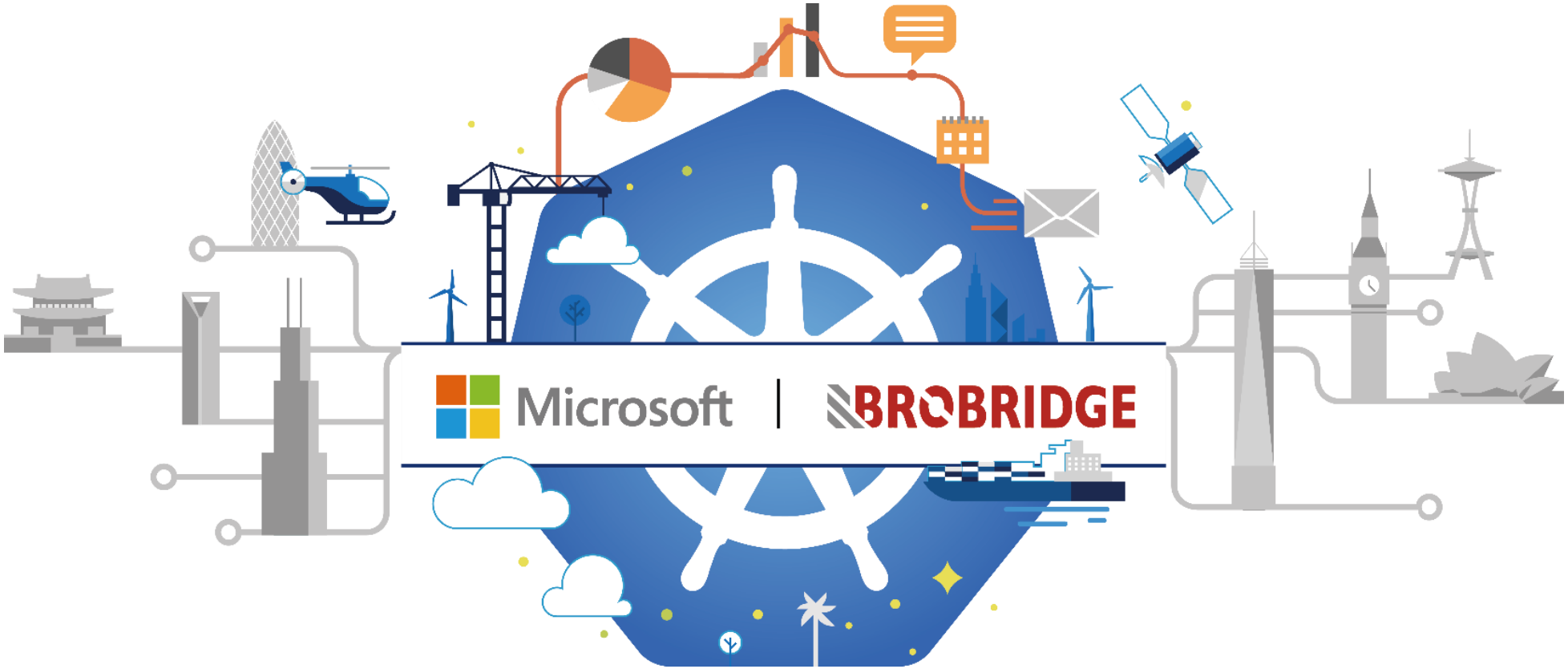


微服務容器時代來臨，傳統難纏的資料議題如何解決？

講者：Kenny Chen(Netman)　　日期：2021.12.22
Thomas Huang

PRESENTATION AGENDA

簡報大綱

Afternoon 16:20 — 數位新時代遭遇的資料問題

GRAVITY數據中台解決方案

AZURE平台應用場景

Afternoon 16:40 — Azure Arc-enabled data services

01

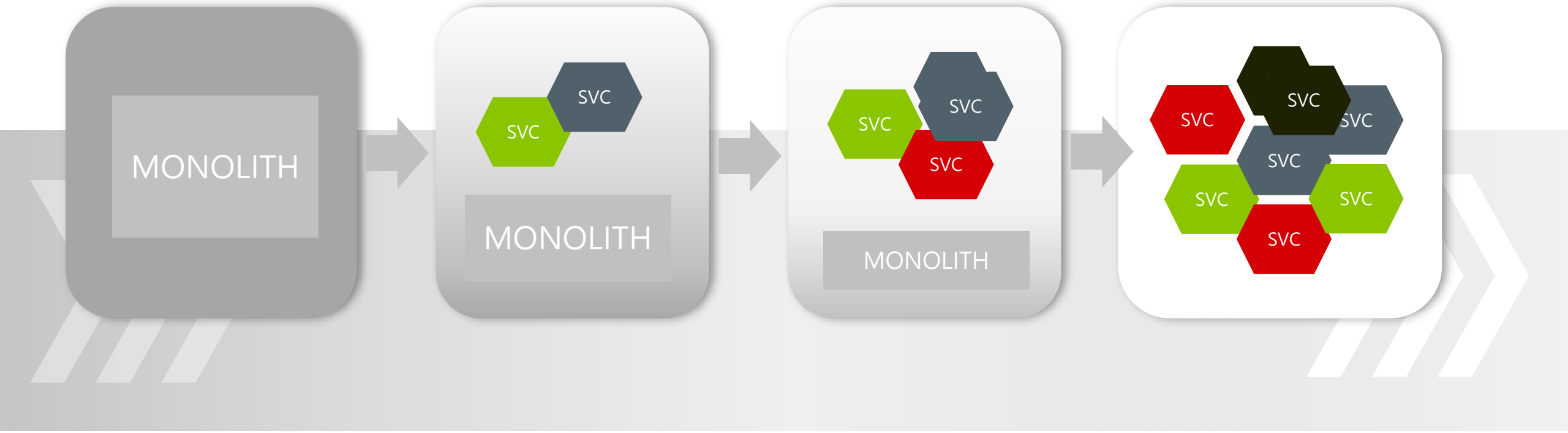
數位新時代遭遇的資料問題

DIGITAL TRANSFORMATION'S CHALLENGE

微服務已成趨勢

MICROSERVICE TRANSFORMATION

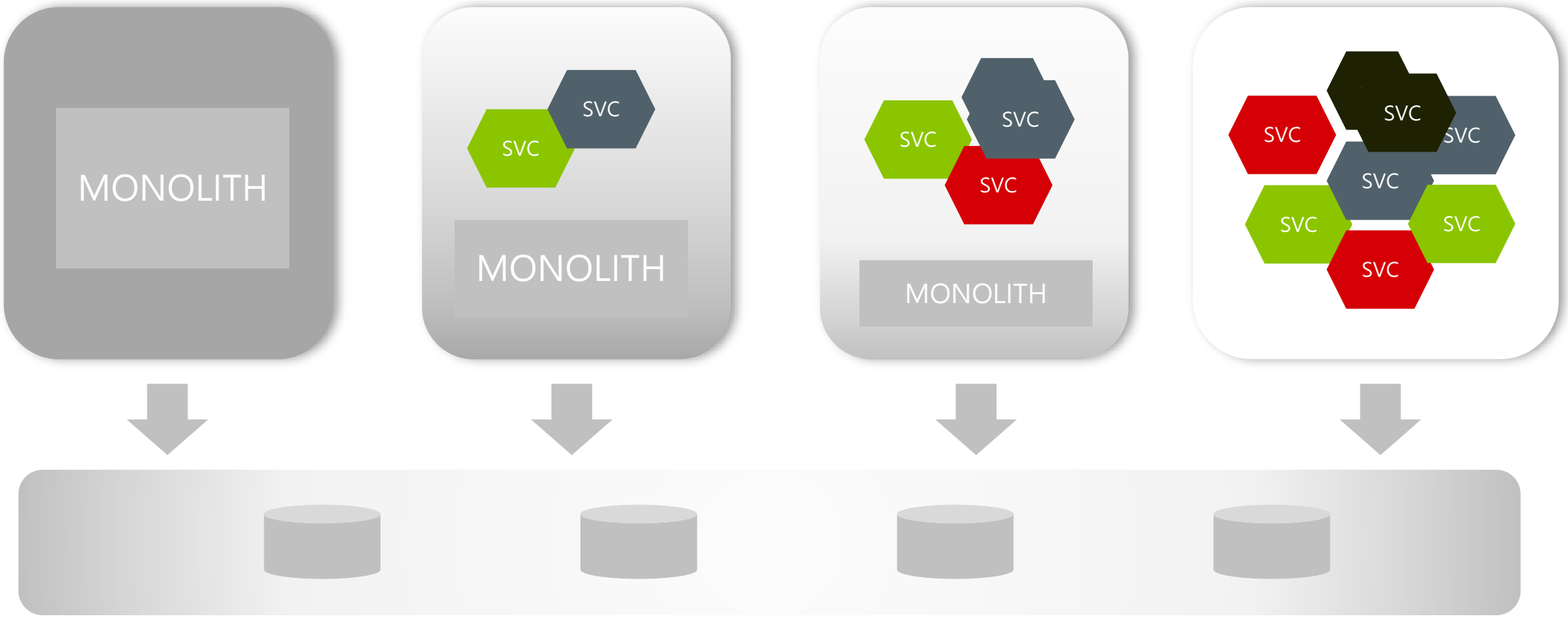
因為雲端技術的逐漸普及，應用服務正大量從單體式架構轉型為微服務架構...

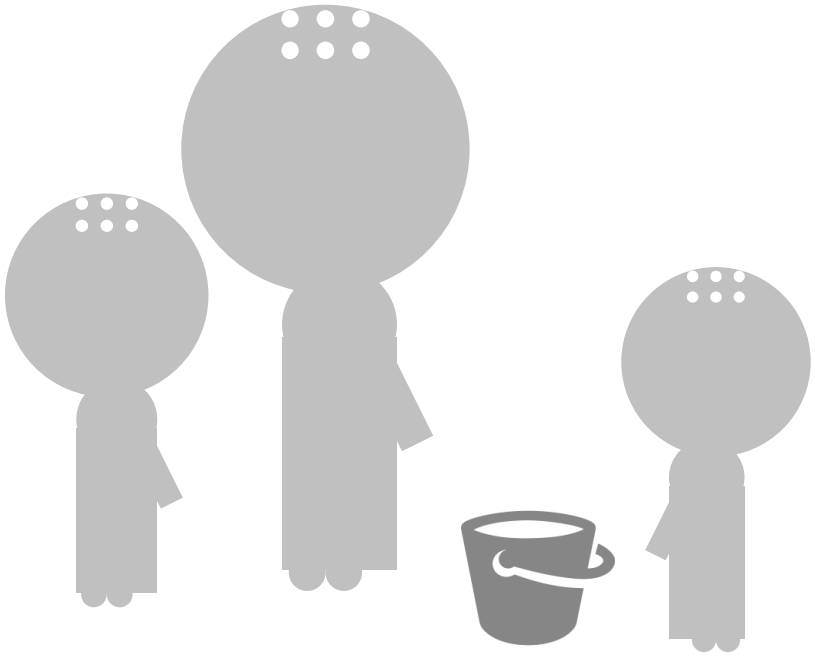


微服務資料供應問題

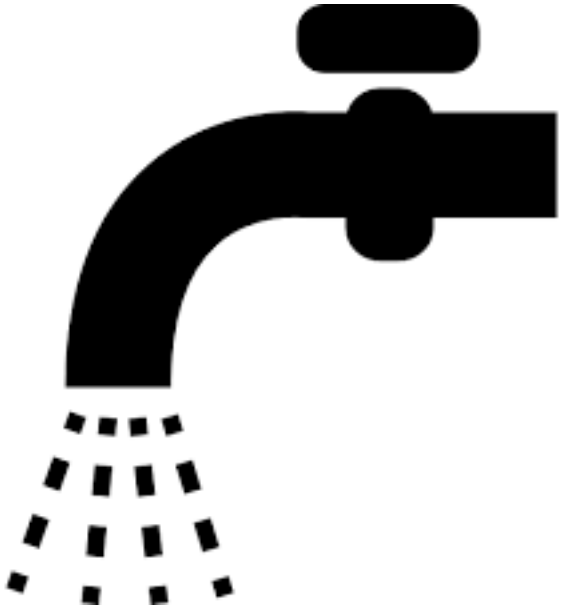
MICROSERVICE DATA SUPPLY ISSUES

然而，傳統的資料供應卻遲遲未能追上微服務的發展需求...



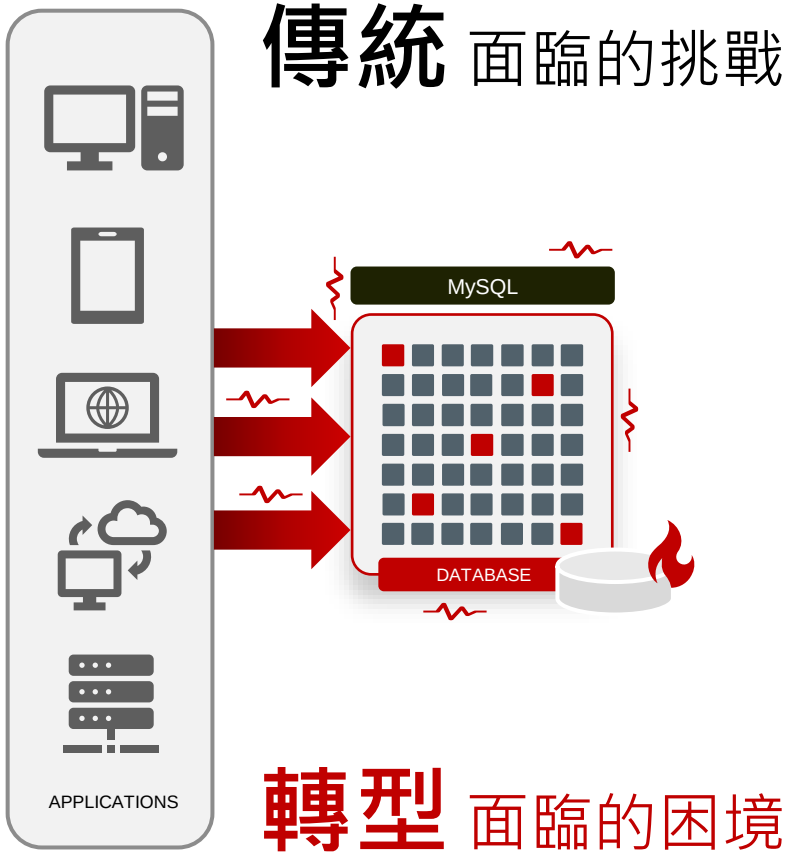


V.S.



數位轉型所面臨的挑戰

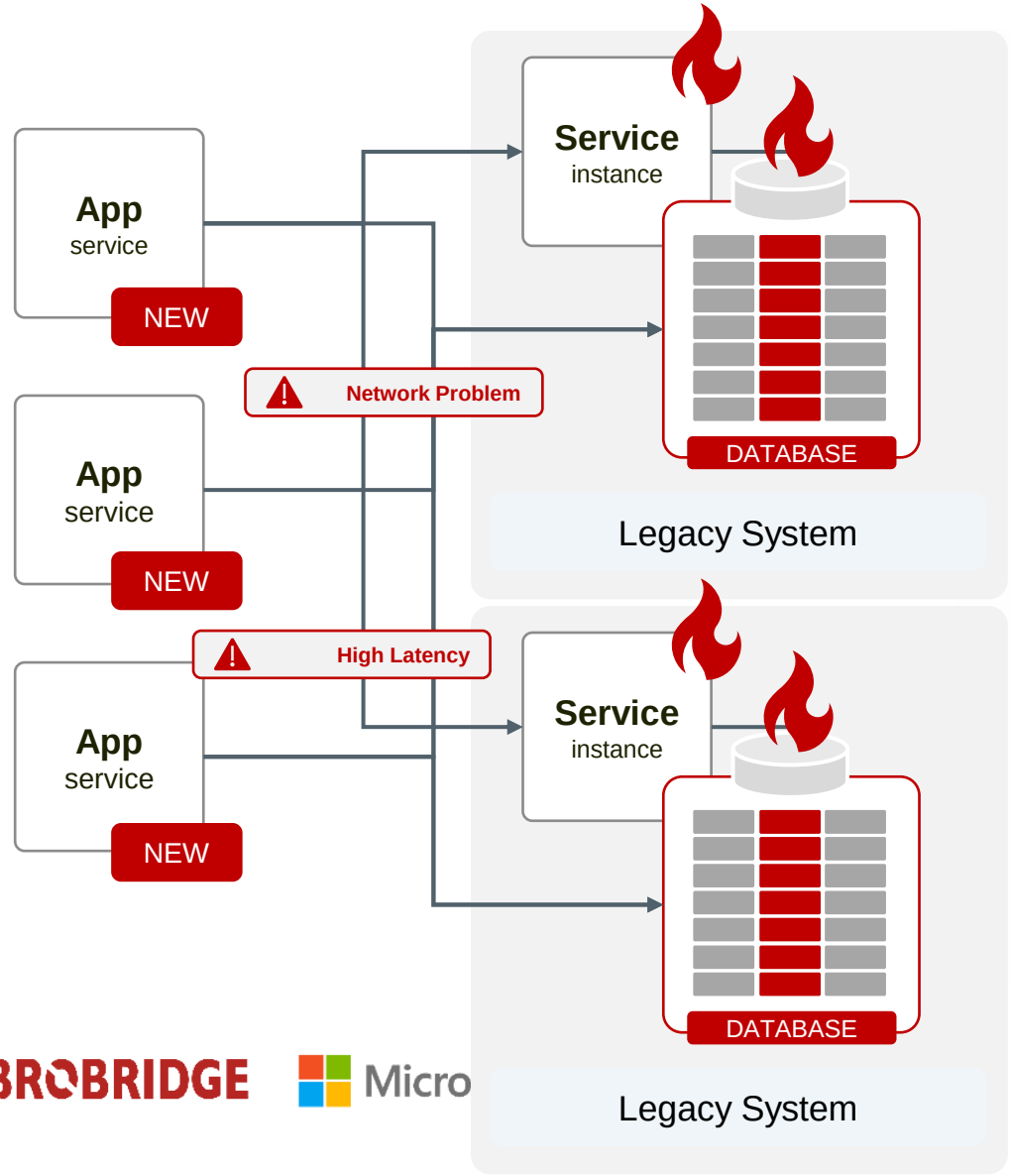
CHALLENGE IN DIGITAL TRANSFORMATION



1. 舊系統過於集中管理，系統LOADING過大。
2. 新服務持續上線，導致原始 DB 承載過高？
3. 隨著組織擴大，資料如何控管？如何交換？
4. 跨系統/部門的資料取得無法自動化、延遲時間過長？
5. 有讀寫分離的需求、但卻會傷及原始 DB 的效能？
6. 有多個資料源的資料整合需求，現行的方案速度過慢、網路 Traffic 大。
7. AI 資料科學家花費半天的時間只為了取得一次分析所需要的 Data？
8. 引進新型態的資料庫（如 MongoDB）如何維持與現有系統的資料一致性？
9. 現行的資料庫如何上雲？ 如何實現混合雲的架構？

跨多個舊系統的整合問題

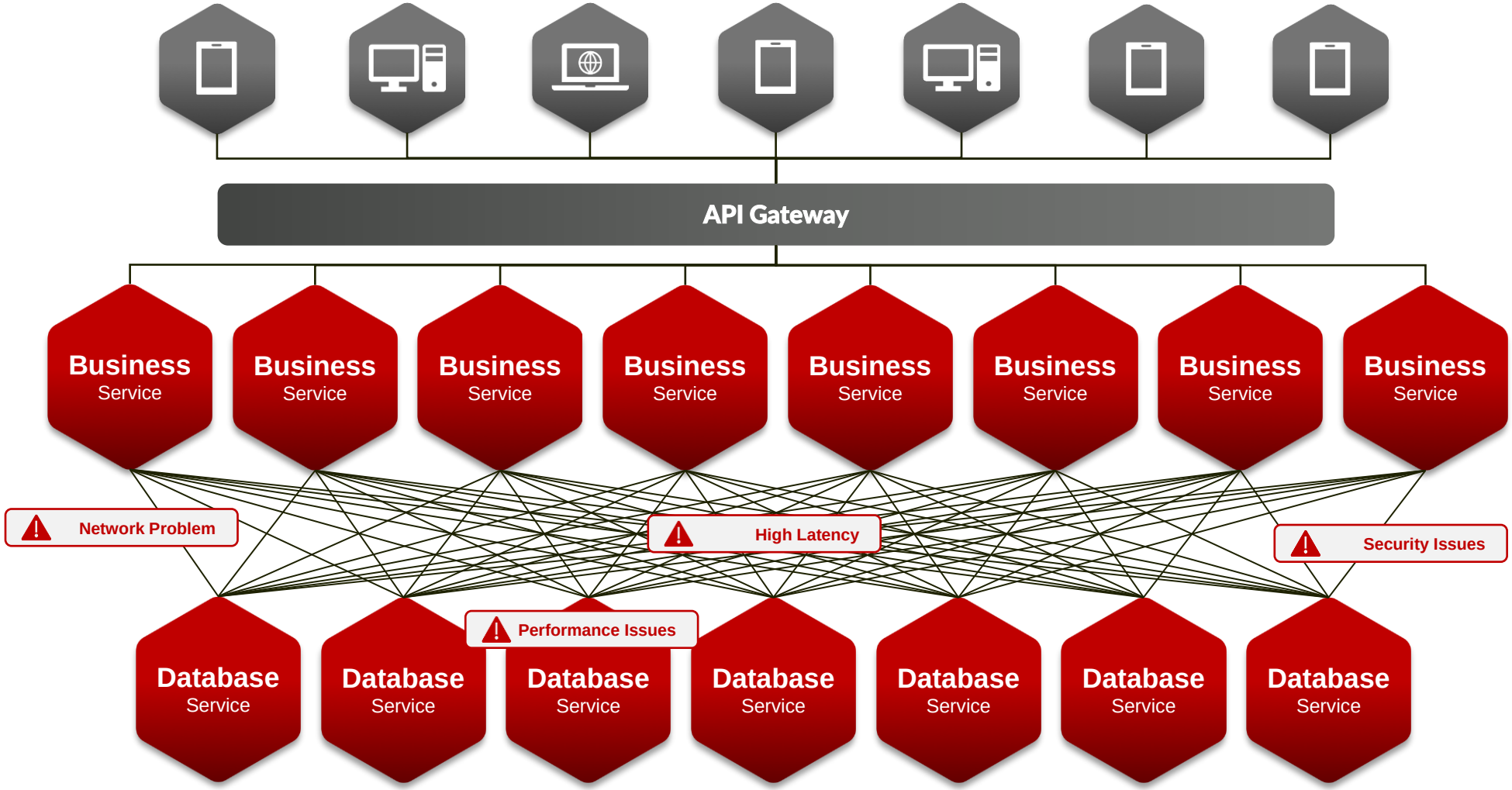
DATA INTEGRATION WITH MULTIPLE LEGACY SYSTEMS



- 無法支撐大併發量
- 跨系統關聯性查詢效能差
- 即時性差
- 可能有資料鎖死的問題
- 瞬間系統負載高
- 新應用效能不彰

混亂不堪的碎片化資料困境

AS-IS



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。

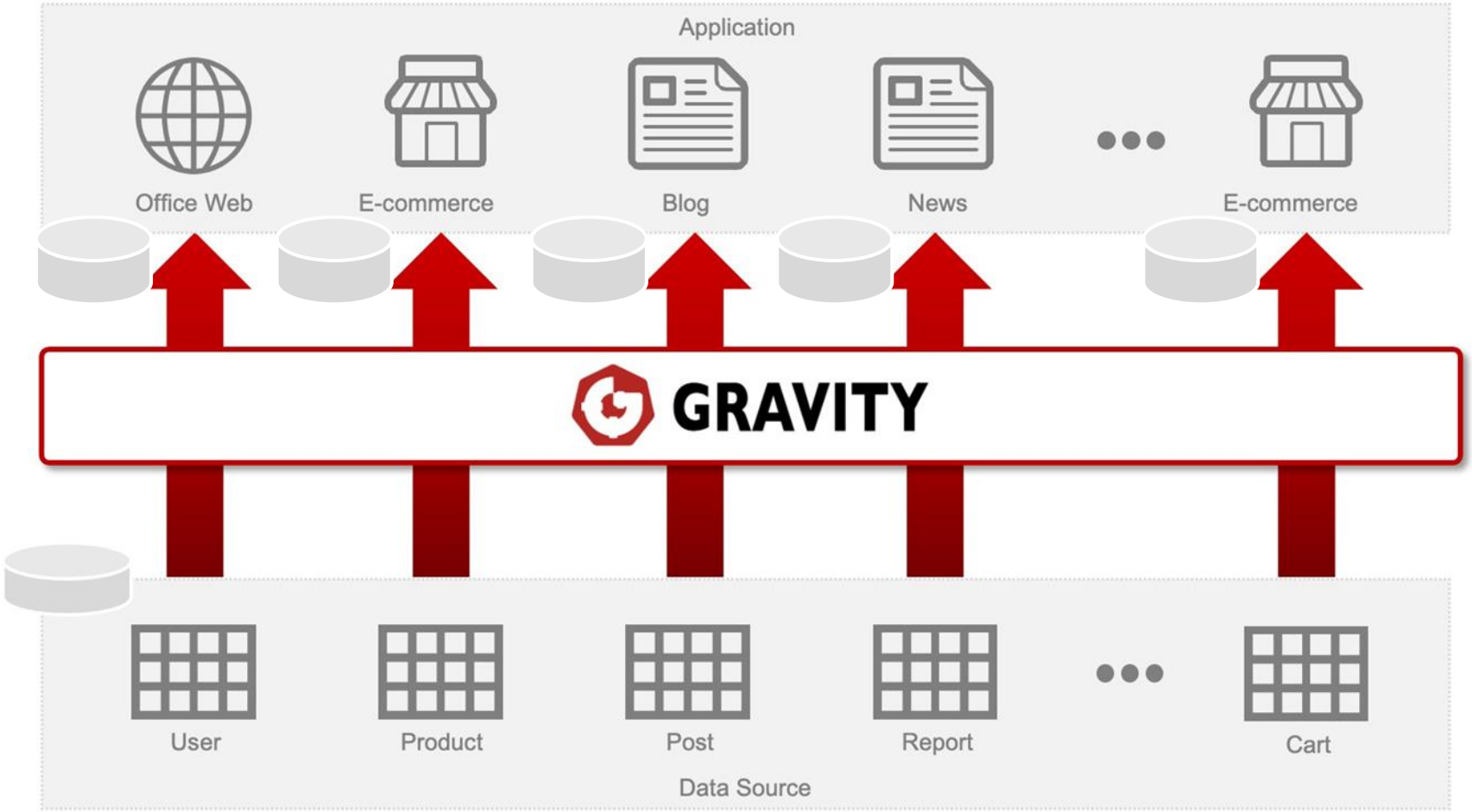
02

GRAVITY數據中台解決方案

BROBRIDGE SOLUTION

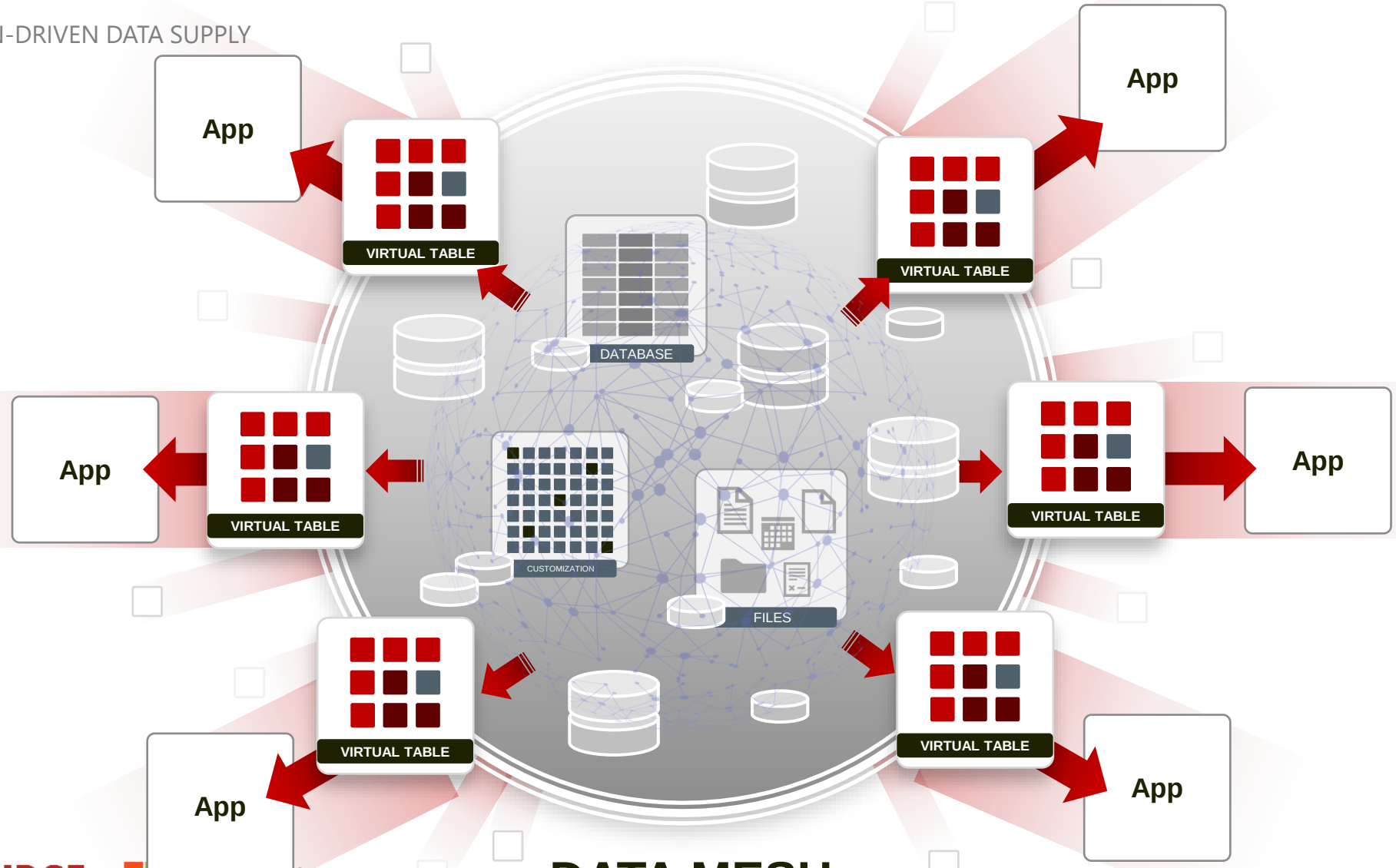
導入數據平台解決資料供應

IMPORT DATA PLATFORM TO SOLVE DATA SUPPLY



應用驅動數據供應需要進行分散式資料整合

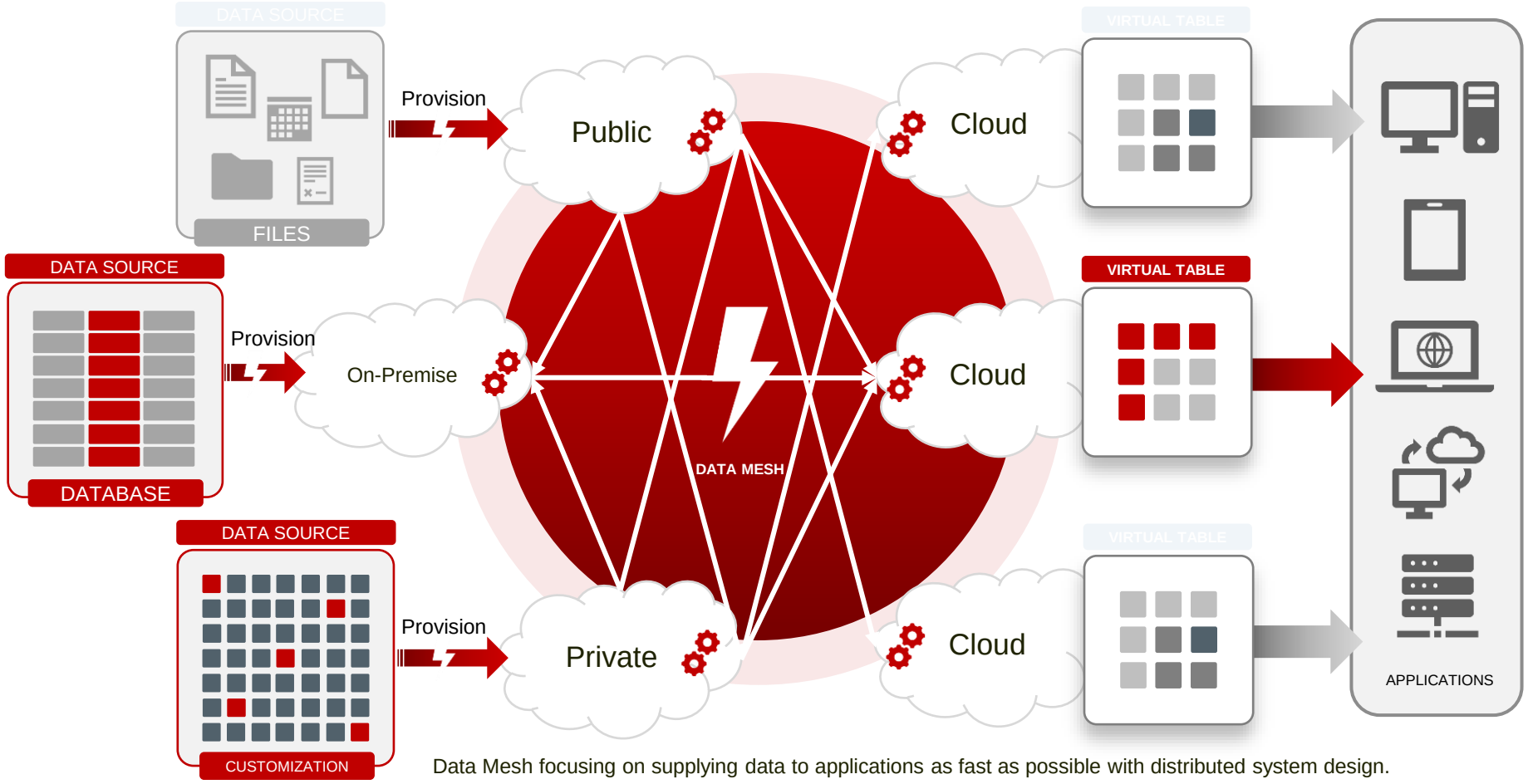
APPLICATION-DRIVEN DATA SUPPLY



DATA MESH

以 DATA MESH 概念實現的分散式資料管理架構

DISTRIBUTED DATA MANAGEMENT WITH DATA MESH



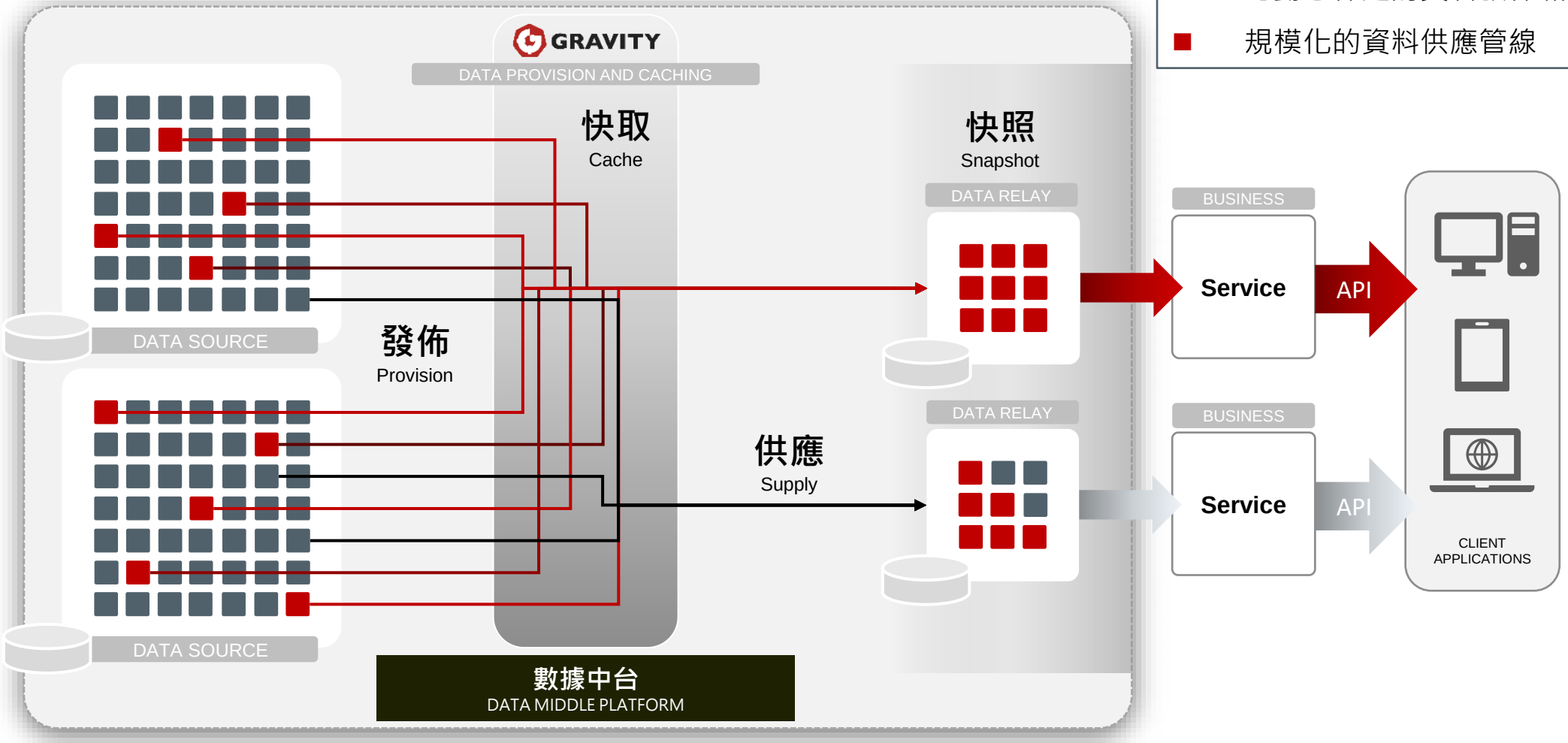
Data Mesh focusing on supplying data to applications as fast as possible with distributed system design.

跨系統的即時資料調度和供給需求

SCALABLE DATA PROVISION AND CACHING PIPELINE FOR BUSINESS

落實目標 (Objective)

- 支撐並供應各應用所需之資料
- 可動態佈建的資料發佈和快取網路
- 規模化的資料供應管線

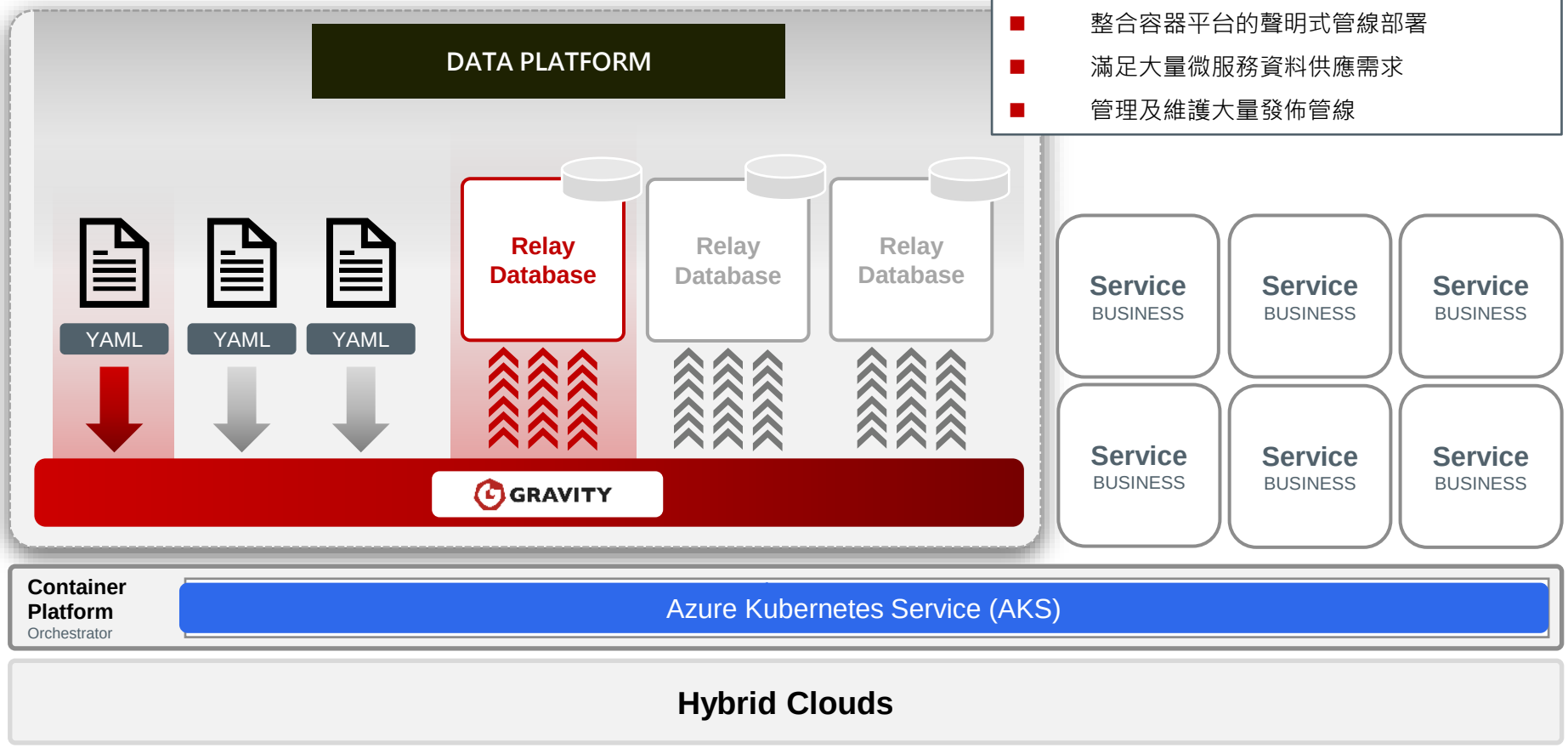


容器化平台 AKS CONTAINER PLATFORM

YAML WAY TO DEPLOY DATA NODE AND PIPELINE INFRSTRUCTURE

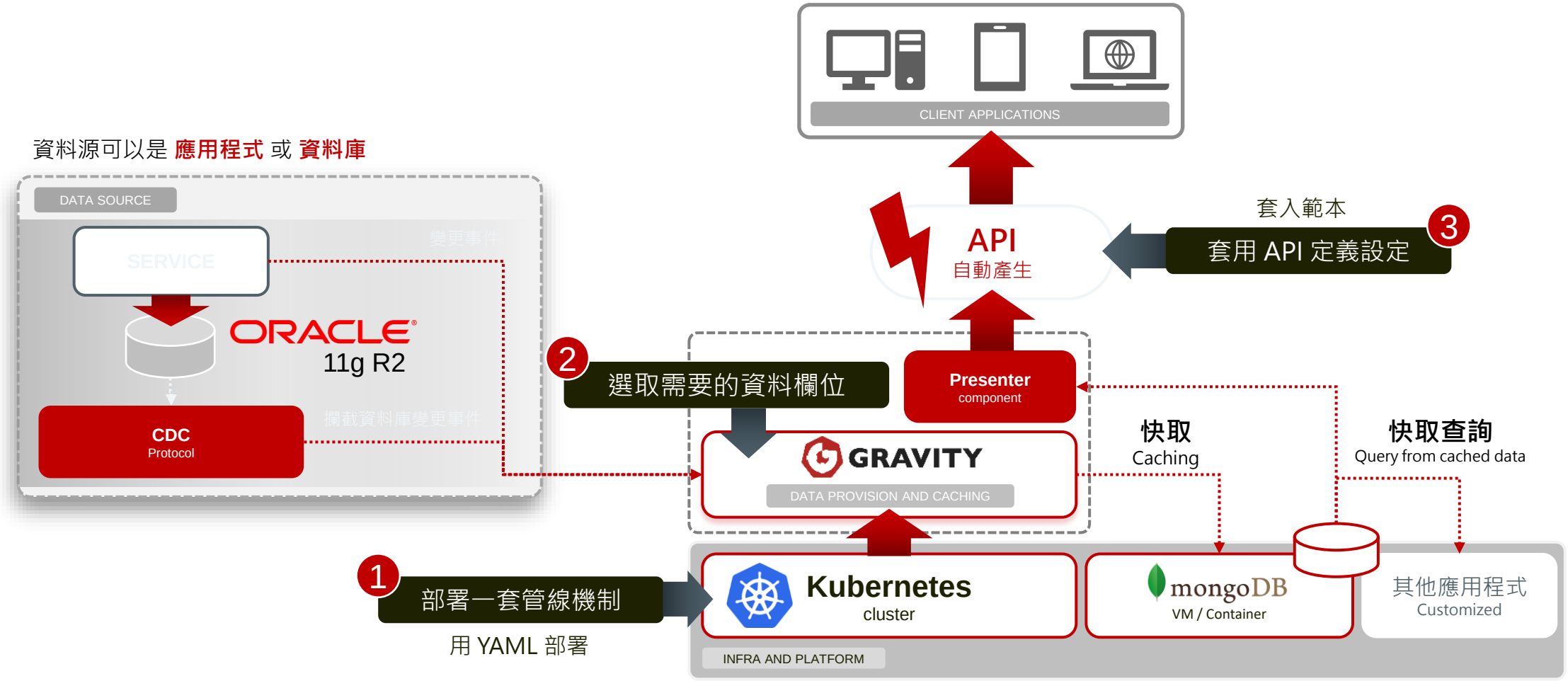
落實目標 (Objective)

- 整合容器平台的聲明式管線部署
- 滿足大量微服務資料供應需求
- 管理及維護大量發佈管線



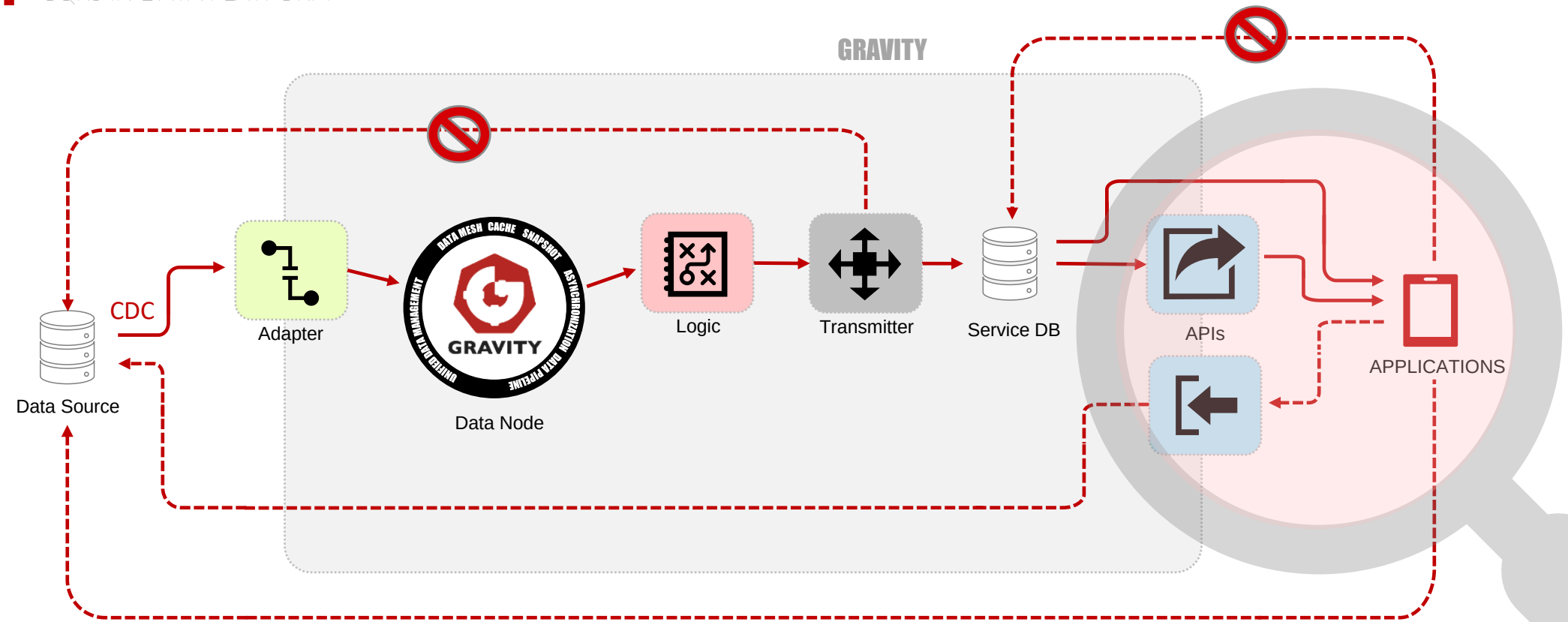
從事件發佈資料、快取、產生虛擬表單並快速產生 API

一行程式都不用寫，數分鐘內建立好整個架構，從 Oracle 到 MongoDB 並產生可用的資料 API



透過數據中台實現讀寫分離(CQRS)

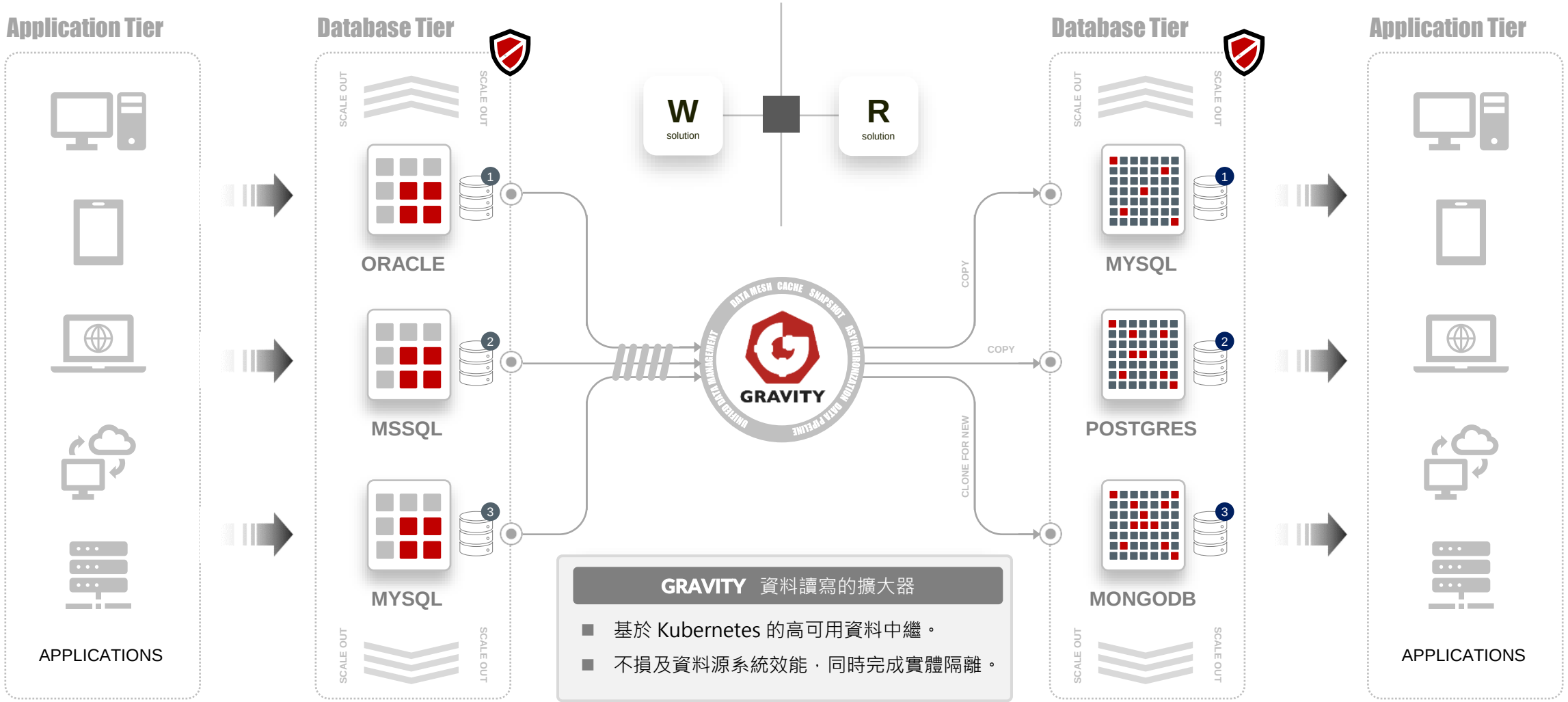
CQRS IN DATA PLATFORM



READ: → WRITE: ---→ FORBIDDEN: 🚫

對應用程式來說，只需要透過 API 直接跟中台讀寫資料就足夠了...

GRAVITY - 資料讀寫的擴大器



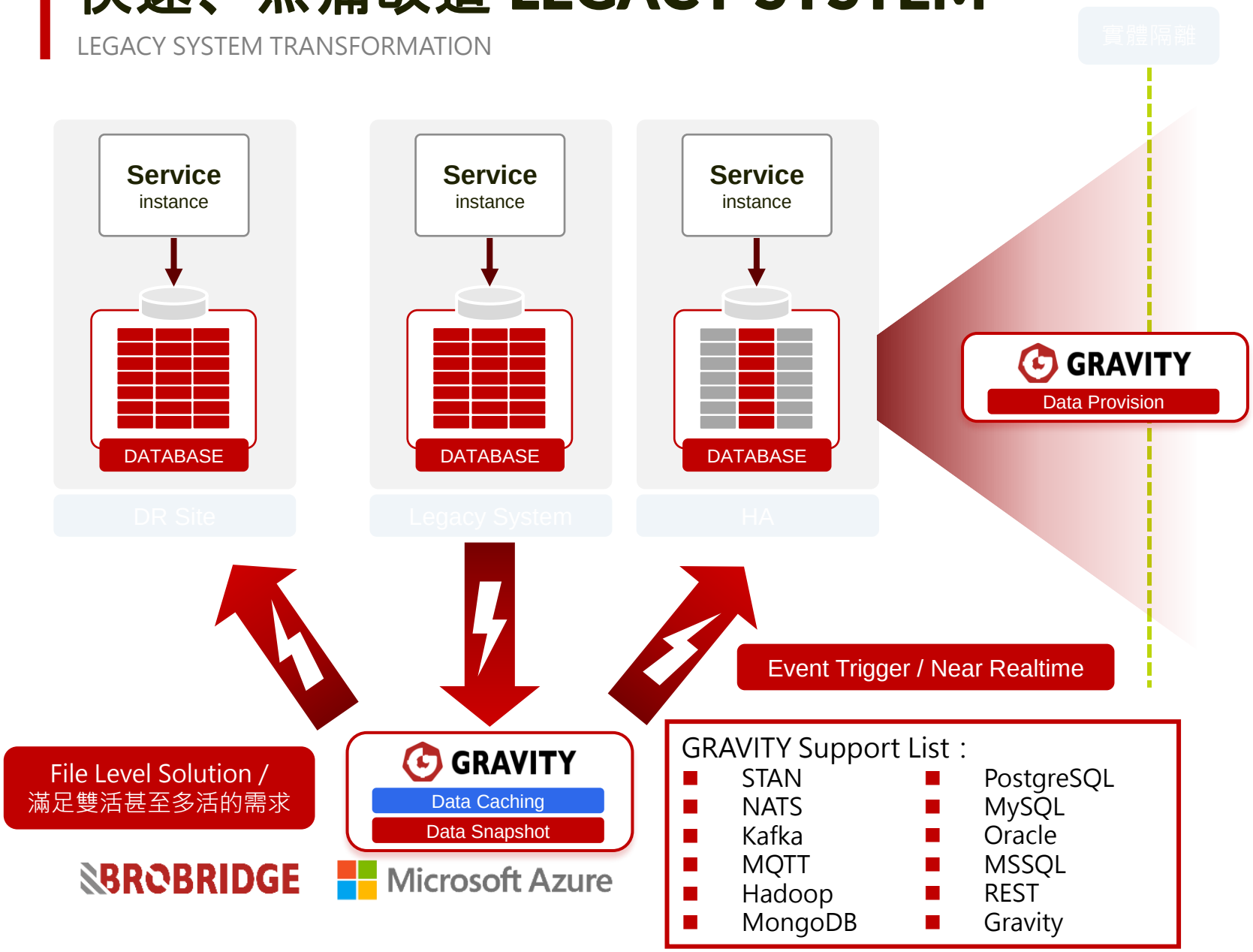
GRAVITY 資料讀寫的擴大器

- 基於 Kubernetes 的高可用資料中繼。
- 不損及資料源系統效能，同時完成實體隔離。

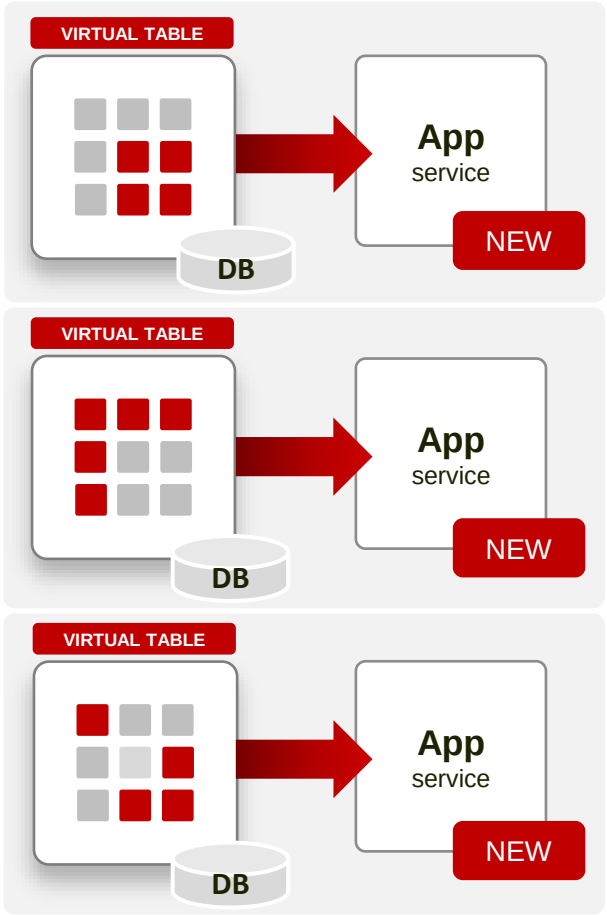
BROBRIDGE Microsoft Azure

快速、無痛改造 LEGACY SYSTEM

LEGACY SYSTEM TRANSFORMATION



擴大舊系統資料供給能力 /
顆粒度小至欄位等級且支援異質資料庫

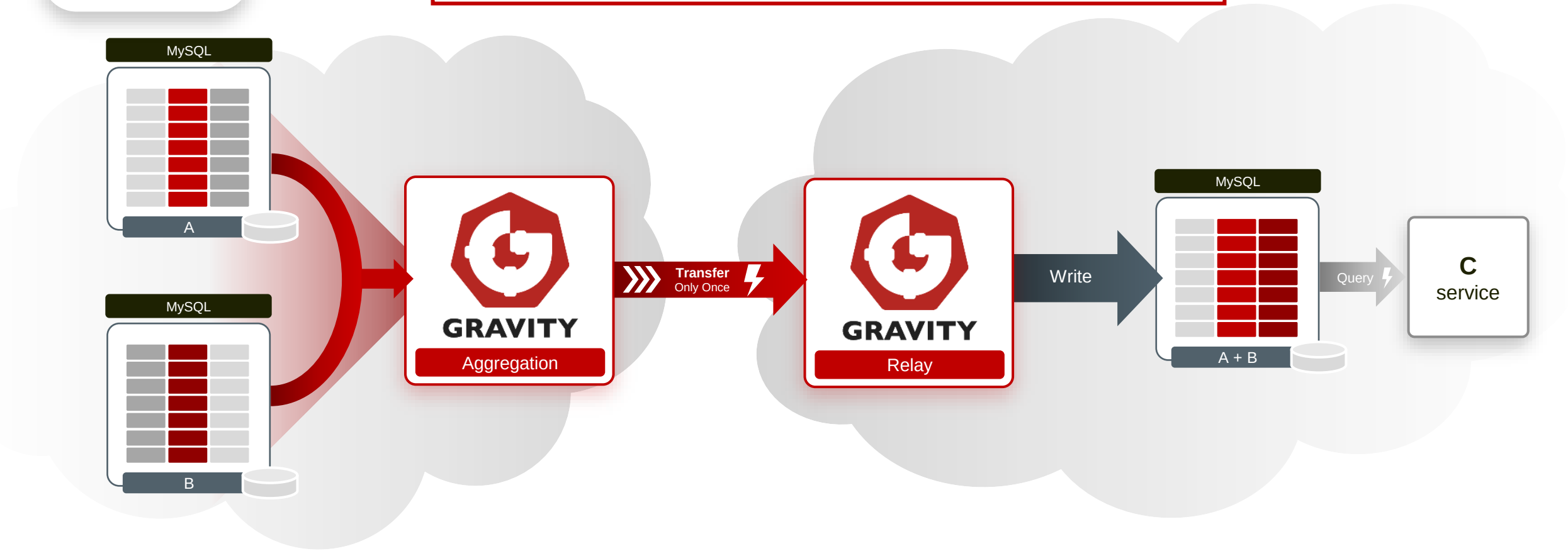


External Systems / Microservice

Hybrid cloud solution

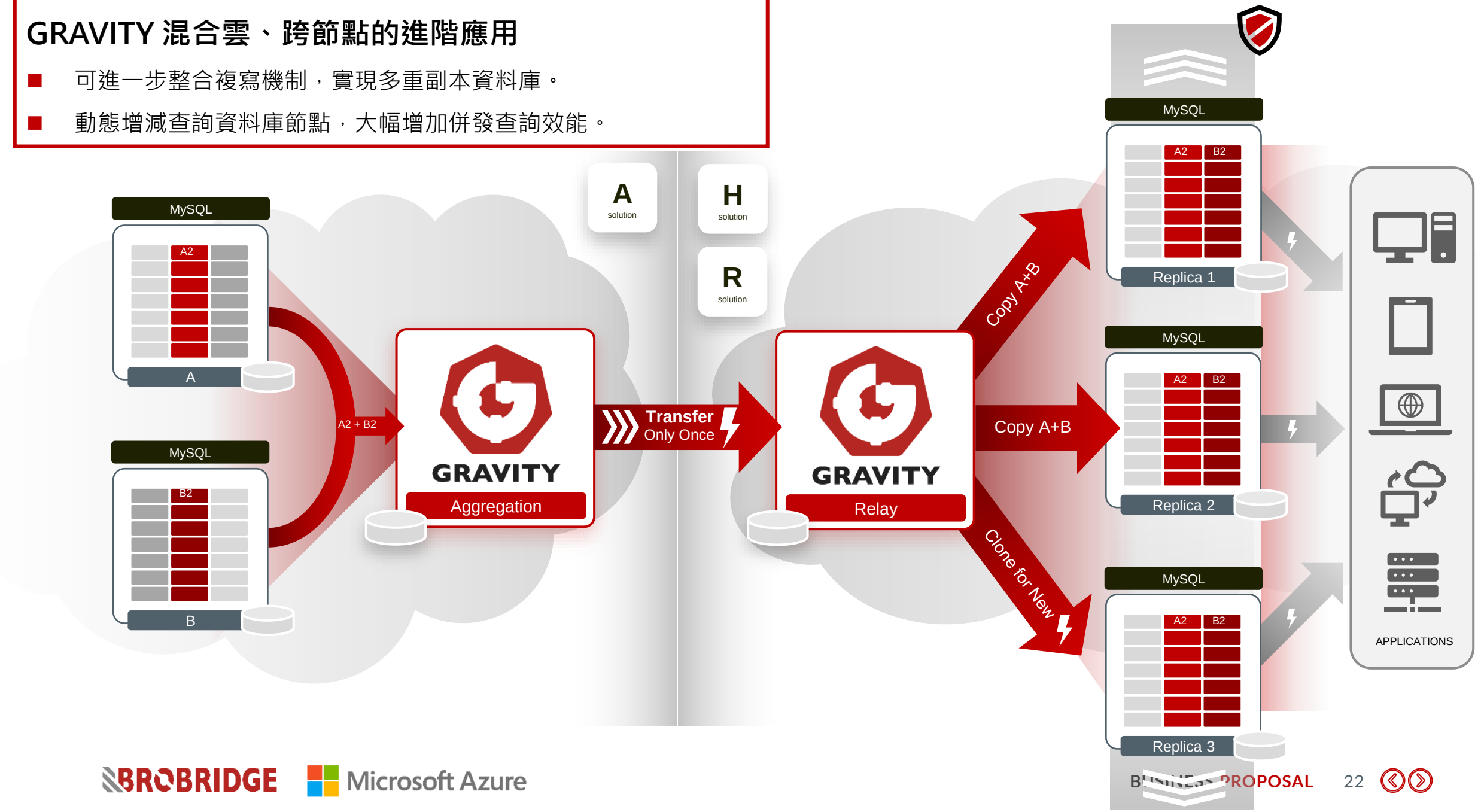
資料上雲就交給 GRAVITY 吧！

- 實現跨雲、跨資料庫的資料關聯查詢
- 資料變更時只發佈一次，大量減少網路頻寬壓力
- 資料中繼機制，讓資料處理無需跨雲、跨網路從源頭拉取



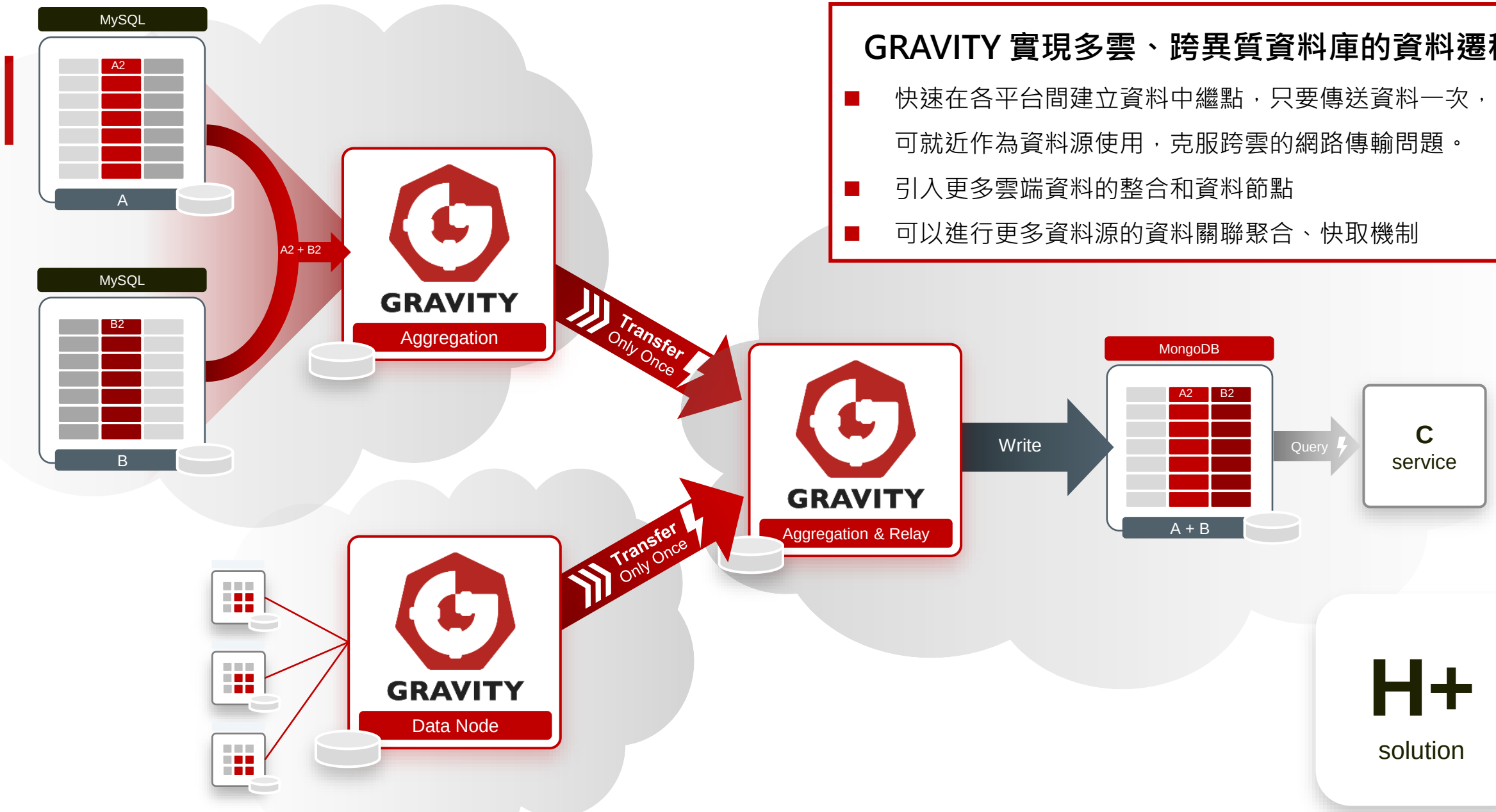
GRAVITY 混合雲、跨節點的進階應用

- 可進一步整合複寫機制，實現多重副本資料庫。
- 動態增減查詢資料庫節點，大幅增加併發查詢效能。



GRAVITY 實現多雲、跨異質資料庫的資料遷移

- 快速在各平台間建立資料中繼點，只要傳送資料一次，即可就近作為資料源使用，克服跨雲的網路傳輸問題。
- 引入更多雲端資料的整合和資料節點
- 可以進行更多資料源的資料關聯聚合、快取機制



03

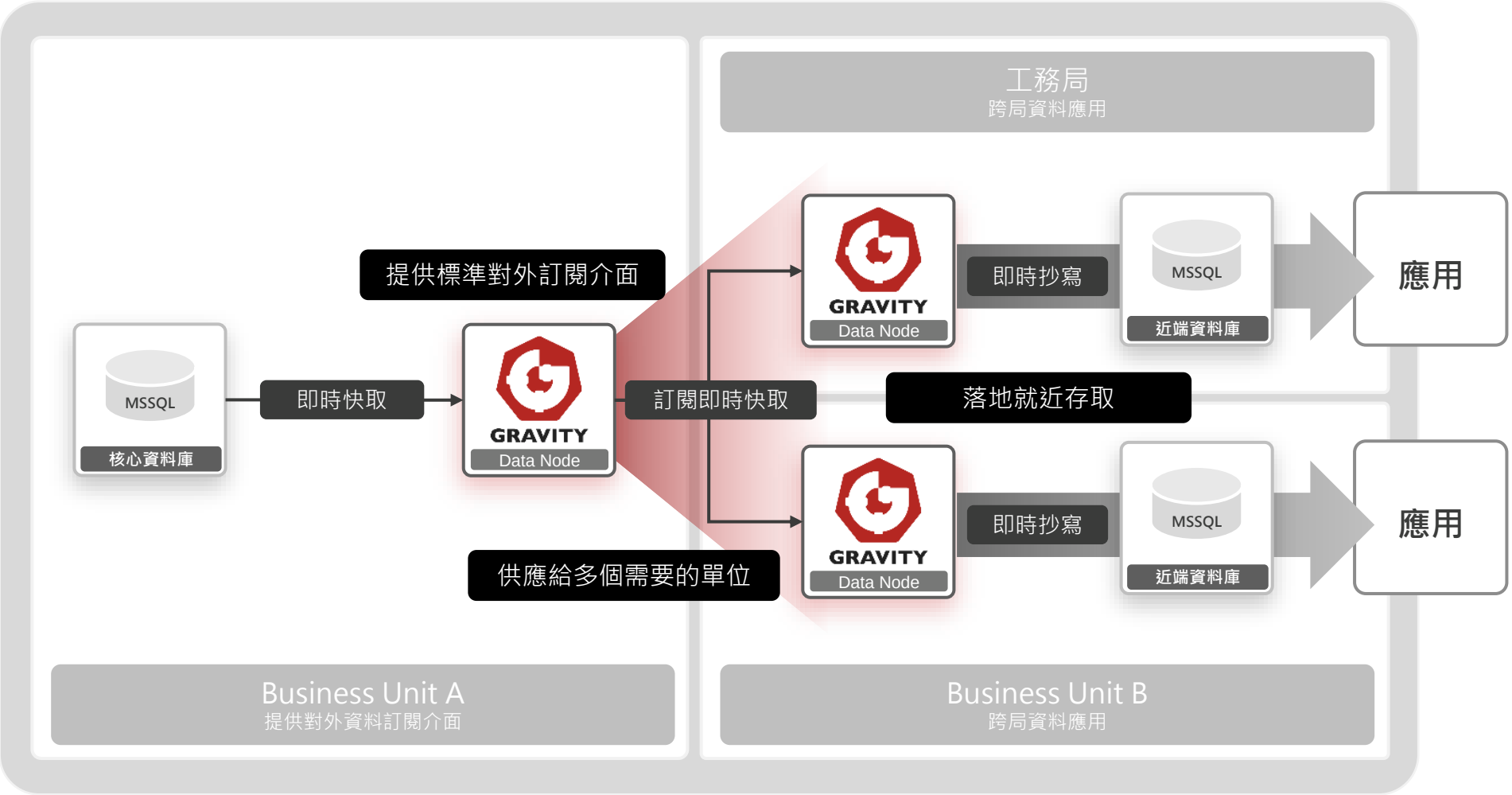
AZURE平台應用場景

USE CASES

情境：跨部門/組織數位資料治理

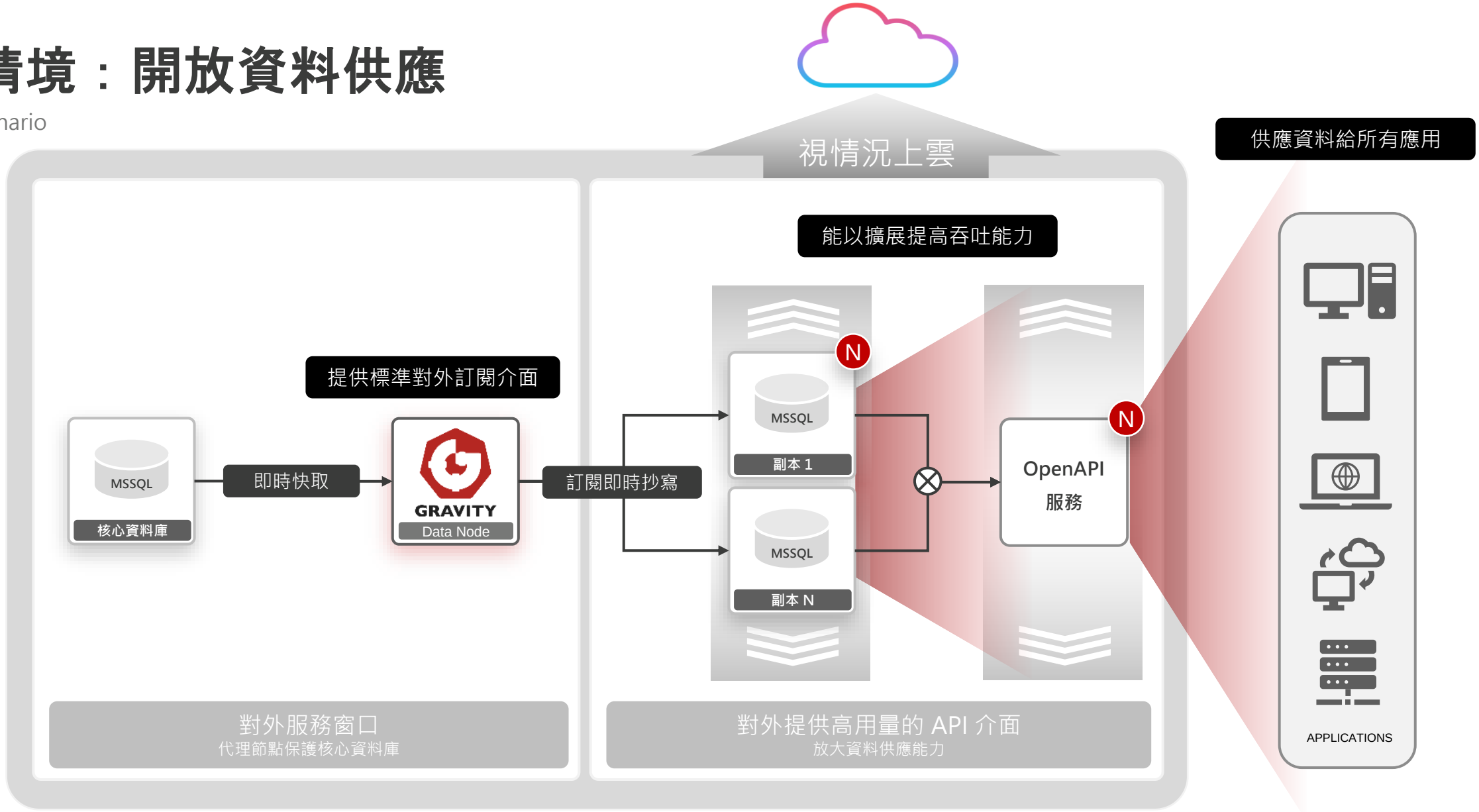
Scenario

資料審核>抄寫>訂閱



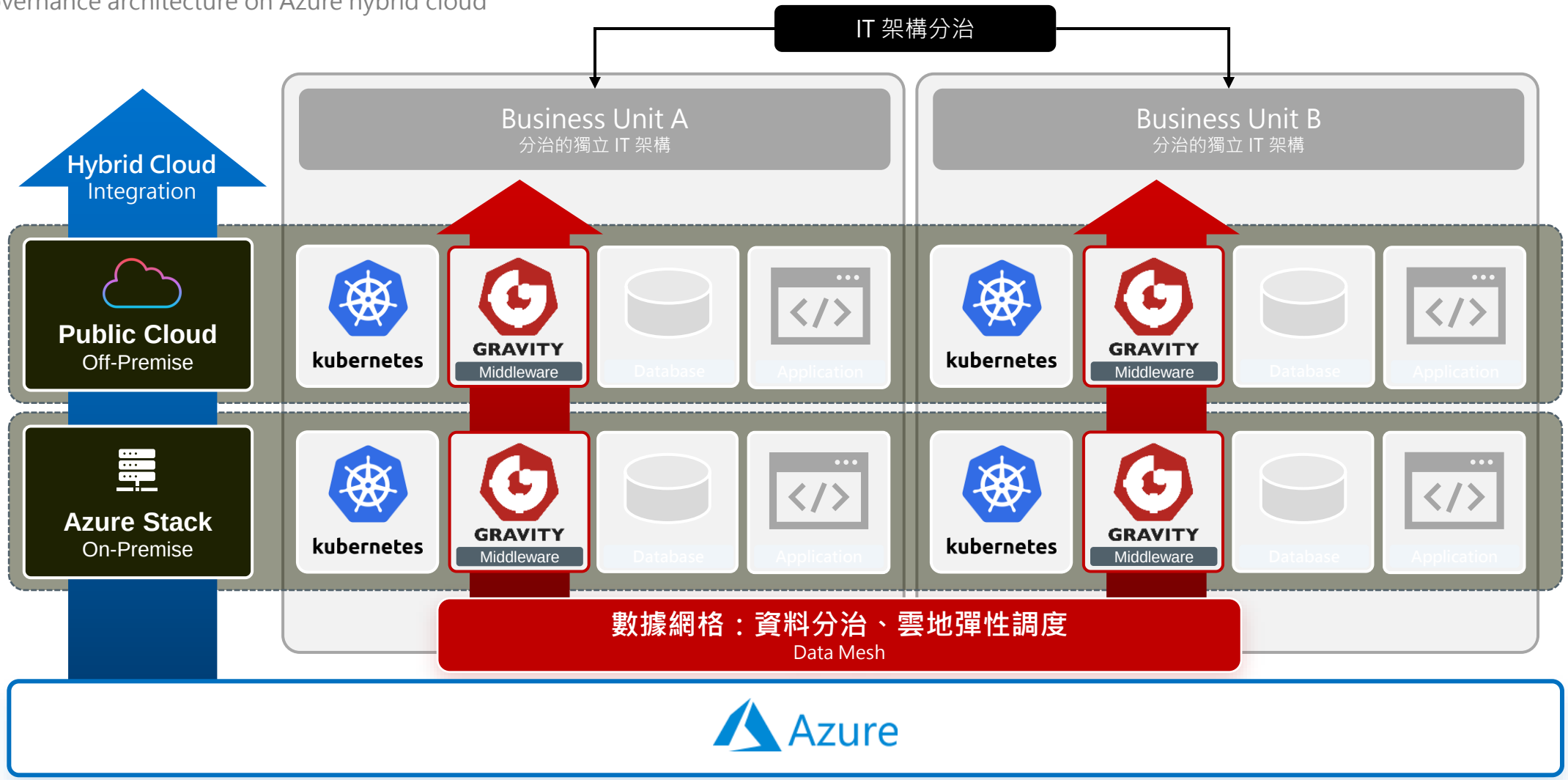
情境：開放資料供應

Scenario



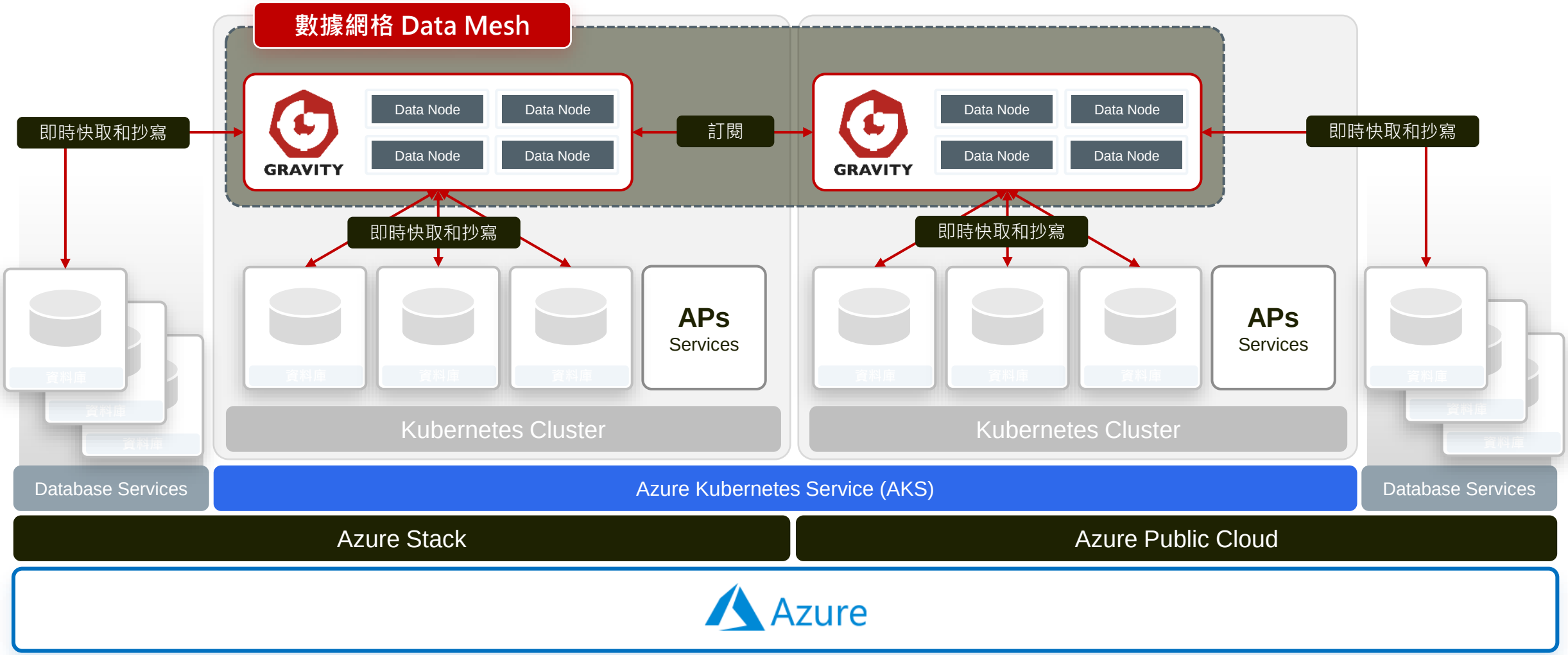
AZURE 數位治理混合雲架構

Governance architecture on Azure hybrid cloud



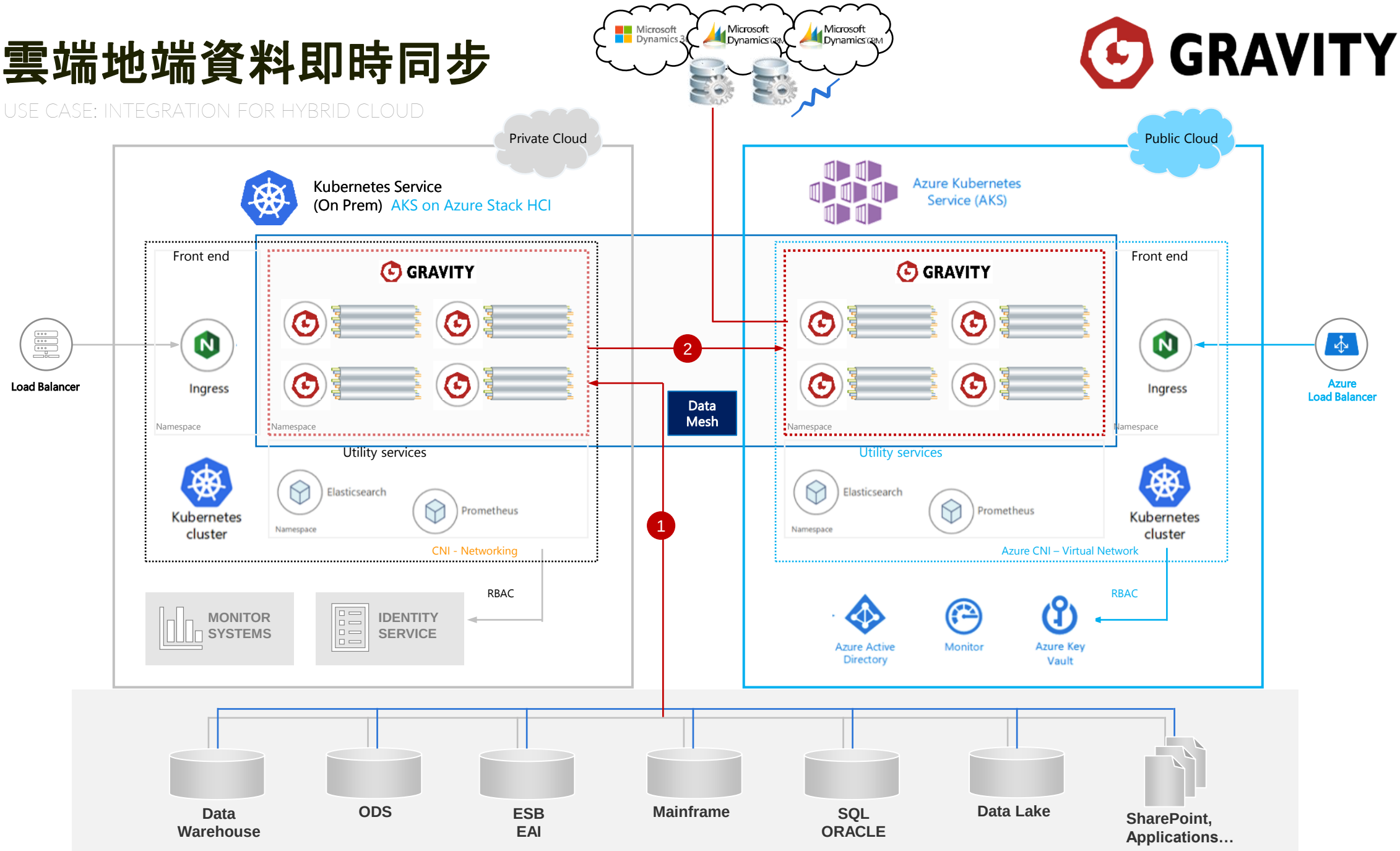
AZURE 數據網格架構

GRAVITY on Azure Platform



雲端地端資料即時同步

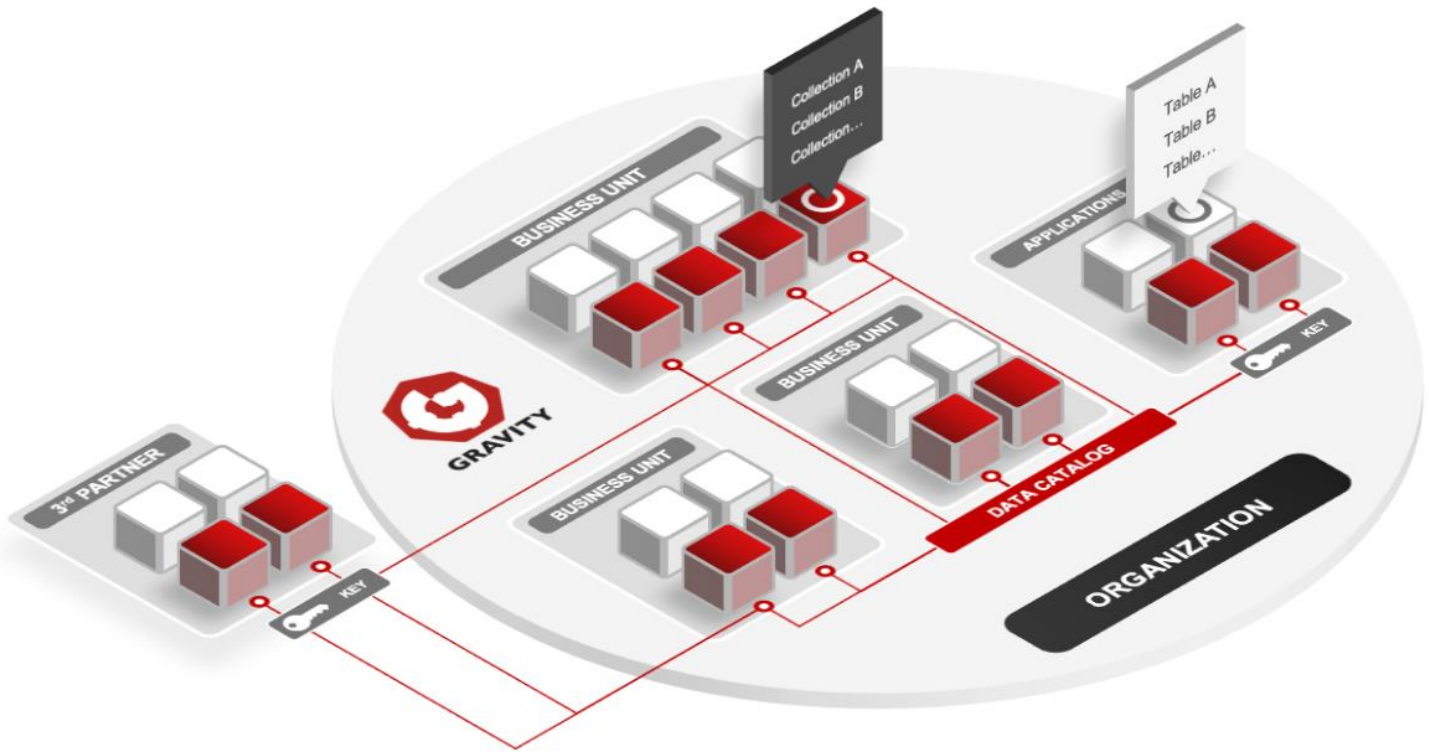
USE CASE: INTEGRATION FOR HYBRID CLOUD



利用數據中台解決現代化數據治理議題

DATA MESH GOVERNANCE

- 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。
- 使用金鑰進行資料的加密授權、存取驗證，確保資料的安全性。
- 資料的交換不再需要大量的客製化，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。
- 資料的流向都有紀錄，除了滿足監控與稽核的需求，也可應用於系統改善調整的參考，協助企業實現 DataOps 目標。



Application
Business logic and algorithm to analyze and solve real world problems based on specific dataset.

- Business Service
- Data Visualization
- Analysis
- Business Intelligence
- AI / ML
- Report

- Tableau
- Denodo
- Microsoft Power BI
- Oracle BI

Data Integration
According to requirements of specific business domain to integrate data for data-intensive applications.

- Data Cache
- Data Mart
- Materialized Views
- ETL / ELT
- Data Warehouse
- Data Fabric

- Informatica
- MicroStrategy
- Greenplum
- Materialize
- Confluent / Kafka
- Oracle DataGuard

Data Mesh
Data Infrastructure for distributed data management across clouds and systems. For the domain-based approach, providing standard interface to integration tools to ingest data from sources efficiently and securely.

- Decentralized Architecture
- Data As a Product
- Data Federation
- Data Delivery
- Data Discovery / Catalog
- Access Control
- Centralized Control
- Data Governance
- Domain-based Management

GRAVITY

Hybrid Clouds
On-premise / Off-premise

- Private Cloud (On-premise)
- Public Cloud (Off-premise)

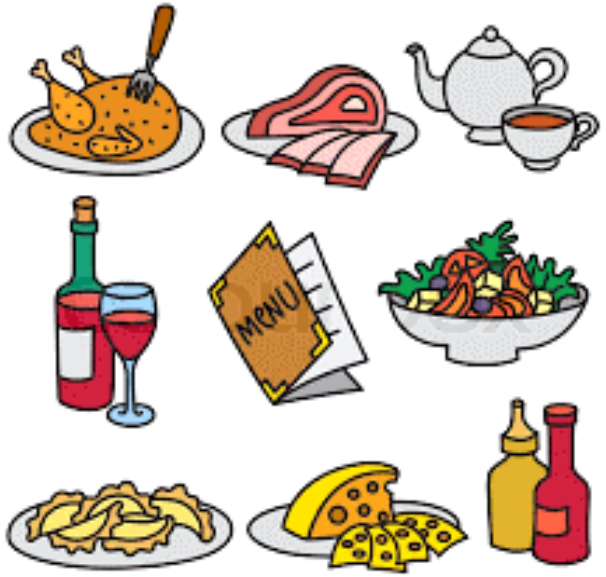
- Host / VM / Container
- Google / Azure / AWS

BROBRIDGE Microsoft Azure

+

V.S.

+

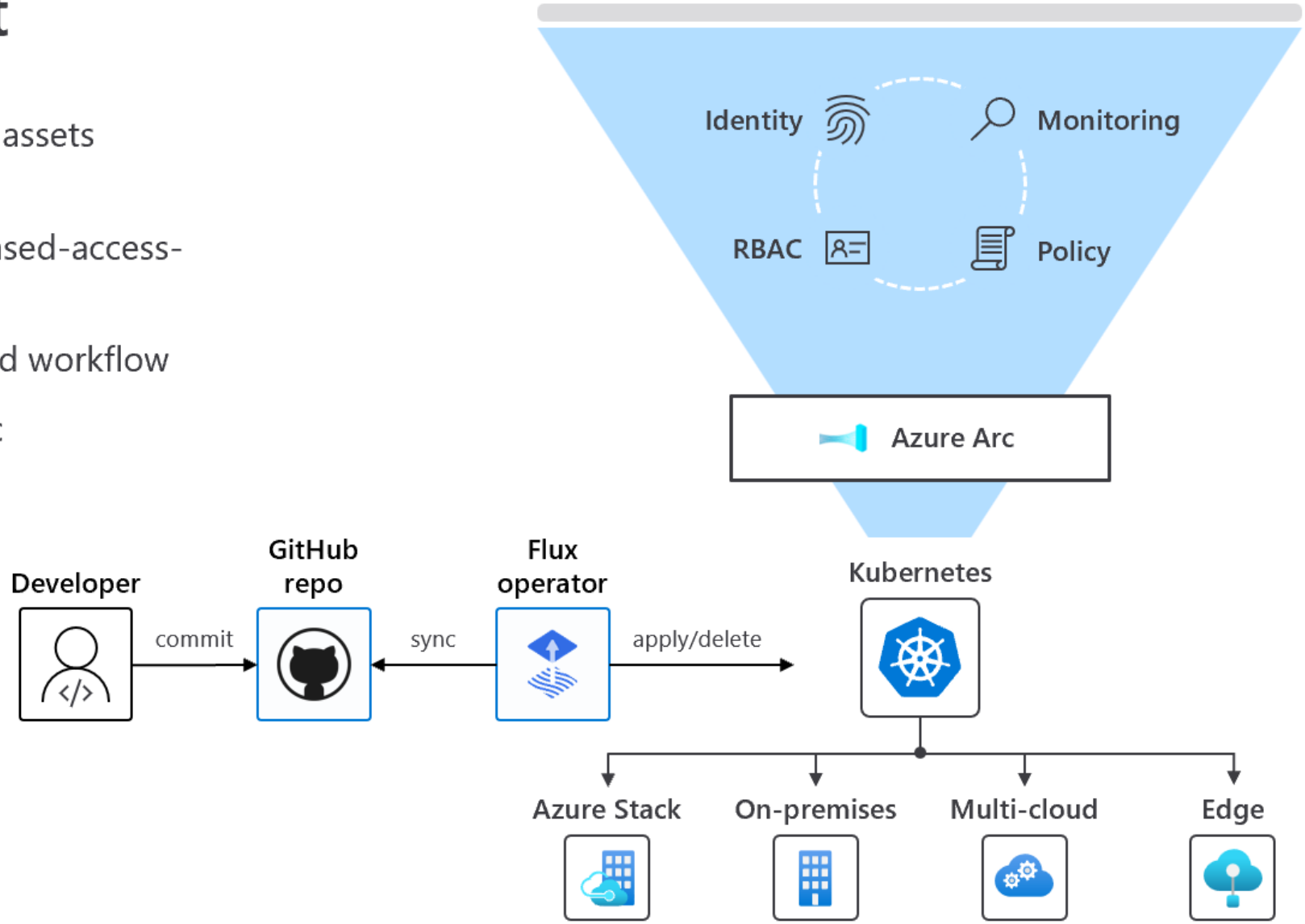


BROBRIDGE Microsoft Azure

Enable hybrid computing with Azure Arc-enabled data services

Unified management

- Central inventory and monitoring of assets running anywhere
- Consistently apply policies & role-based-access-controls (RBAC)
- Deploy resources using GitOps-based workflow
- Use Flux operator for automatic sync

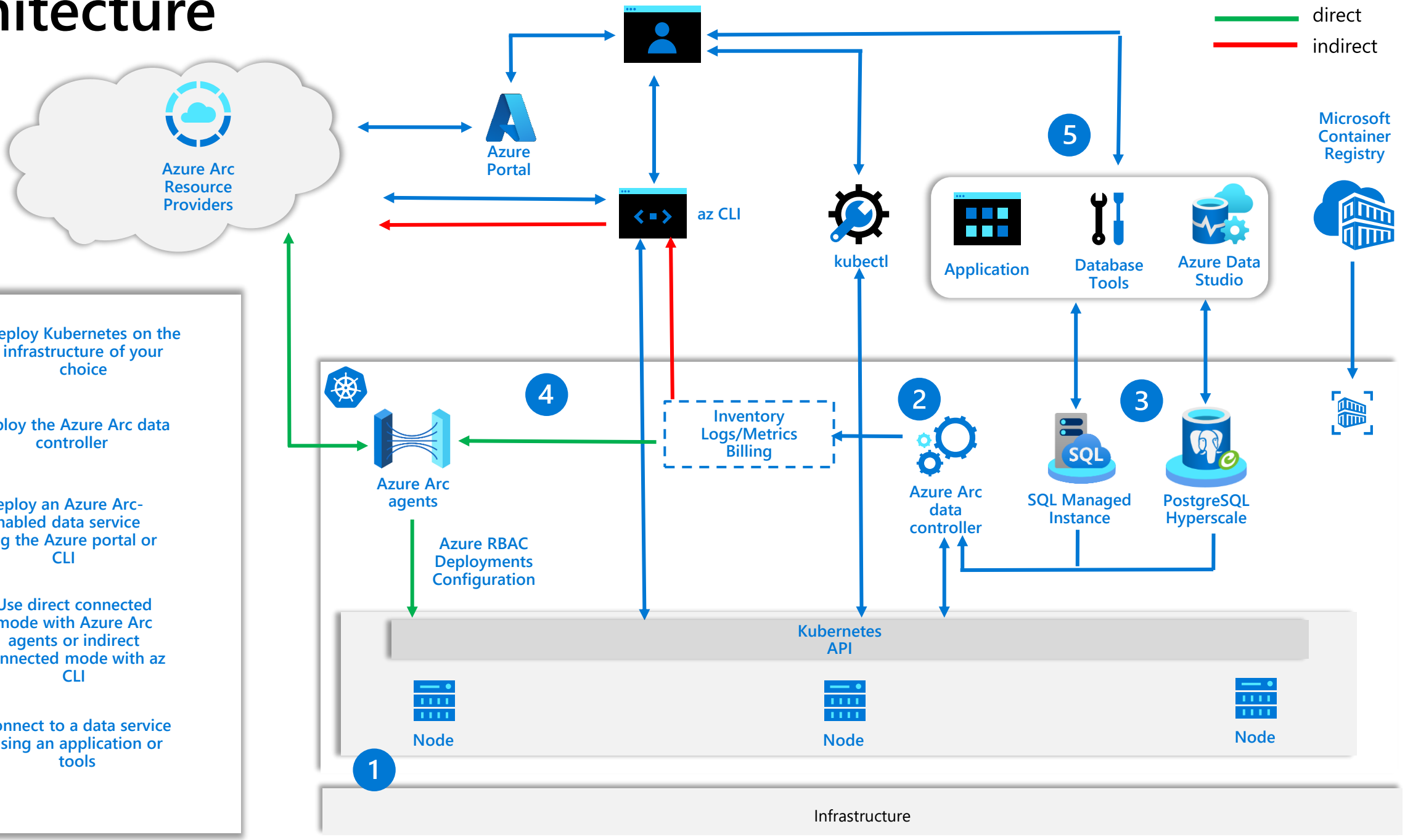


Enable hybrid computing with Azure Arc-enabled data services

Azure Arc-enabled data services are a set of software to provide hybrid data capabilities and managed services using the Kubernetes platform.

- Azure Arc-enabled SQL Managed Instance
- Azure Arc-enabled PostgreSQL Hyperscale (Preview)

Architecture



1. Deploy Kubernetes on the infrastructure of your choice
2. Deploy the Azure Arc data controller
3. Deploy an Azure Arc-enabled data service using the Azure portal or CLI
4. Use direct connected mode with Azure Arc agents or indirect connected mode with az CLI
5. Connect to a data service using an application or tools

Azure Arc Interfaces

- Azure portal
- az CLI (+arcdata extension)
- kubectl
- SQL Server Management Studio (SSMS)
- Azure Data Studio
- SQL command line tools

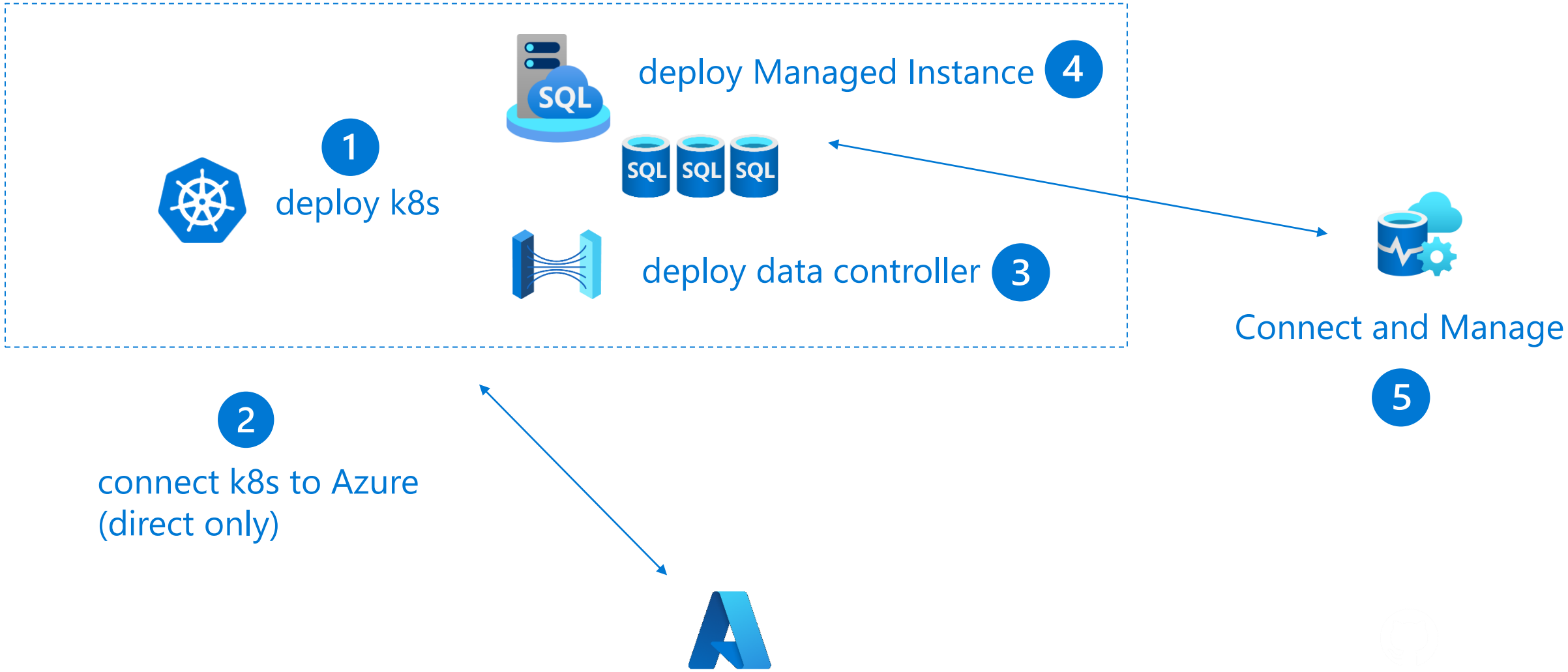
Deploy a managed database service with Azure Arc-enabled SQL Managed Instance

Comparison with Azure SQL

Azure SQL is a family of database services deployed in the Azure infrastructure:

- SQL Server in Azure Virtual Machine
- Azure SQL Managed Instance
- Azure SQL Database
- Azure Arc-enabled SQL Managed Instance

How do I get started?



Comparison of features

To understand the capabilities of Azure Arc-enabled SQL Managed Instance, it's important to look at its core features, deployment options, and integration with Azure.

- Core features
- Service tiers
- Azure integration

Service Tiers

Area	Business Critical (preview)	General Purpose (GA)
Feature set	Same as SQL Server Enterprise Edition	Same as SQL Server Standard Edition
CPU limit/instance	Unlimited	24 cores
Memory limit/instance	Unlimited	128 GB
High availability	Availability group	Single instance w/ Kubernetes redeploy + shared storage
Read scale out	Availability group	None
Azure Hybrid Benefit rates for IP component of price	1:1 Enterprise Edition 4:1 Standard Edition	1:4 Enterprise Edition 1:1 Standard Edition
Dev/Test pricing	No cost	No cost

Comparison of management capabilities

Since Azure Arc-enabled SQL Managed instance is a managed service, it's useful to compare management capabilities with Azure SQL Managed instance such as a versionless engine, deployment and configuration, monitoring, high availability, and disaster recovery.

- Versionless
- Deployment and configuration
- Monitoring
- High availability
- Disaster recovery

Microsoft

Demo

Explore Azure Arc-enabled SQL Managed Instance

Understand Azure Arc-enabled PostgreSQL Hyperscale (Preview)

Understand Azure Arc-enabled PostgreSQL Hyperscale (Preview)

PostgreSQL is a mature, open-source Relational Database Management System (RDBMS).

The nodes store data in tables in one of three modes:

- Distributed tables.
- Reference tables.
- Local tables

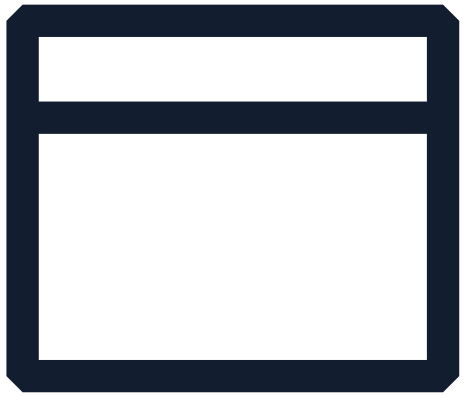
Application type considerations

The following types of applications benefit from a scale-out PostgreSQL Hyperscale deployment:

- Organizations that host multiple applications that are often dedicated to a single client or location work well with a scale-out PostgreSQL Hyperscale.
- Applications with many tables and schemas that service On-Line Transaction Processing (OLTP) or On-Line Analytical Processing (OLAP) workloads benefit from a scale-out architecture.
- Real-time applications that use a smaller number of tables with aggregated analytic queries.
- Applications that ingest a large set of mostly immutable data.

Comparison of management capabilities

There are two primary interfaces for monitoring and managing an Azure Arc-enabled data services PostgreSQL Hyperscale deployment: Command-Line Interfaces (CLI) and Graphical User Interfaces (GUI).



Tools and utilities

In general, you can use graphical tools for discovery and exploration, and command-line tools for standardization and automation.

- The PostgreSQL *pgcli* utility
- The PostgreSQL *psql* utility
- The PostgreSQL *pgAdmin* tool
- The Azure portal interface
- Azure Data Studio
- The Microsoft Azure *az* command utility
- Kubernetes native tools
- Azure Arc-enabled PostgreSQL Hyperscale interfaces
- Standard API calls to documented Microsoft Azure and Microsoft Azure Arc-enabled PostgreSQL

Microsoft

Demo

Explore Azure Arc-enabled PostgreSQL Hyperscale

Knowledge Check

Test your knowledge

Question 1

Which key component of Kubernetes does Azure Arc-enabled data services use?

A. Kubernetes API

B. az CLI

C. kubectl

D. Azure Arc Agents

Question 1

Which key component of Kubernetes does Azure Arc-enabled data services use?

A. Kubernetes API

B. az CLI

C. kubectl

D. Azure Arc Agents

Question 2

Which service tier provides high availability with replicas using Always On Availability Groups?

A. General Purpose

B. Enterprise

C. Business Critical

Question 2

Which service tier provides high availability with replicas using Always On Availability Groups?

A. General Purpose

B. Enterprise

C. Business Critical

Question 3

You can view billing, metrics, and logs for Azure Arc-enabled SQL Managed Instance in the Azure portal using indirect connected mode

A. True

B. False

Question 3

You can view billing, metrics, and logs for Azure Arc-enabled SQL Managed Instance in the Azure portal using indirect connected mode

A. True

B. False

Question 4

Which of the following would be reasons to choose Azure Arc-enabled data services PostgreSQL Hyperscale over the Azure PostgreSQL Hyperscale Service?

A. The data must stay on-premises

B. The system should be deployed from the Azure portal

C. The system must work with a graphical user interface (GUI)

Question 4

Which of the following would be reasons to choose Azure Arc-enabled data services PostgreSQL Hyperscale over the Azure PostgreSQL Hyperscale Service?

A. The data must stay on-premises

B. The system should be deployed from the Azure portal

C. The system must work with a graphical user interface (GUI)

Resources

Go through the learning module

aka.ms/learnazurearcdata

Get started fast with Jumpstart

aka.ms/azurearcjumpstart

Get all the information you need

aka.ms/azurearcdata

Microsoft

© Copyright Microsoft Corporation. All rights reserved.