

## HVR x Fivetran

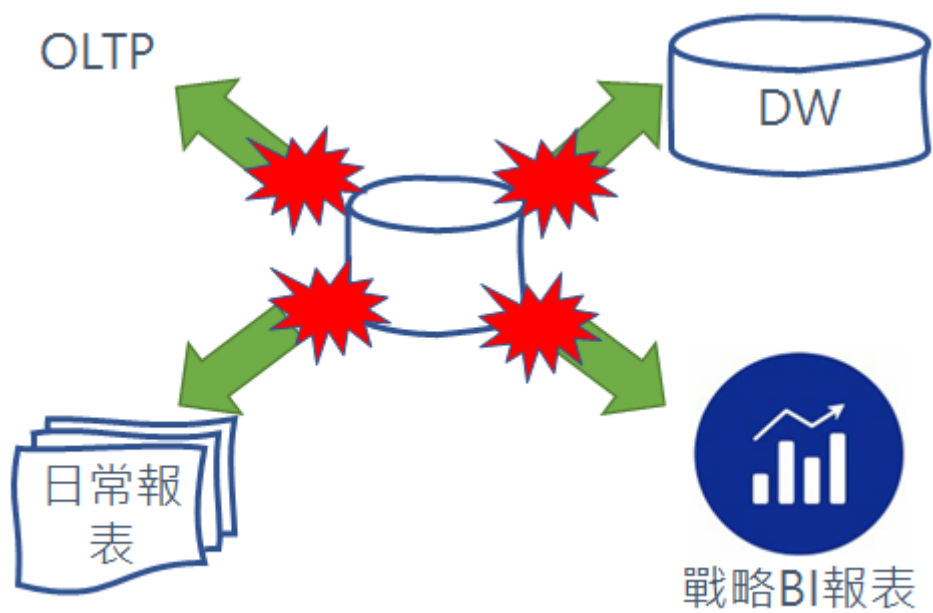
### 解決資料整合挑戰

重磅消息！就在 2021 年 9 月底 Fivetran 宣布收購 HVR。Fivetran 專注於將資料移動到資料倉儲的提取、轉換和載入 ( ETL ) 流程，其優勢之一是能連接各種軟體即服務 ( SaaS ) 的應用程式和基於雲端 ( cloud-based ) 的分析平台。HVR 在基於日誌的變更資料擷取 ( log-based change data capture ) 方面表現出色，這意味著它不會批量複製整個資料庫，而是監視資料庫更新並僅擷取新訊息，並接近即時地複製資料。這樣一個結合互補技術的聯盟，將解決全球最嚴峻的資料整合挑戰，但在我們期待 Fivetran 與 HVR 結合後會迸出什麼火花之前，先來複習一下 HVR 以及了解為何會有資料複製需求。

- **為何會有資料複製需求？**

- 一、效能問題

讓我們用一張圖來說明目前線上系統資料庫所面臨的窘境，即時的交易、日常報表、BI 報表、ETL 轉檔，彼此資源的競爭造成效能下降。OLTP 與 OLAP 的配置該如何取捨，往往讓 DBA 傷透腦筋。



## 二、沒有停機時間

面對全球性的競爭，企業的服務幾乎是 7\*24 不間斷，資料庫停機時間日漸減少，不管是更新修補程式、版本升級或是平台搬遷所被允許的停機時間很少，甚至沒有時間。

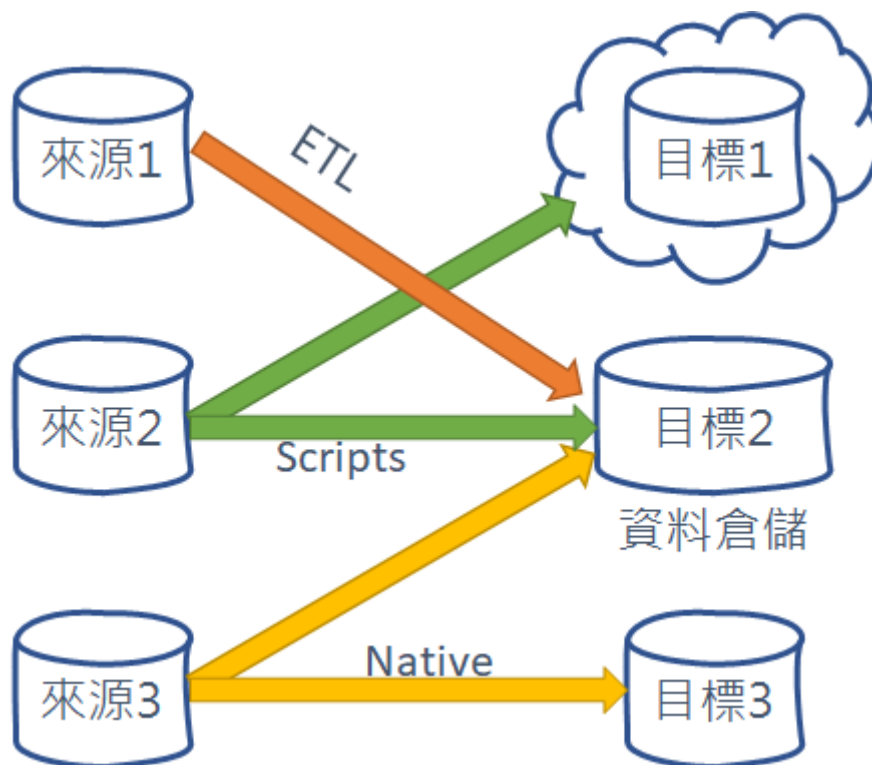
而且這其中還包含一個風險，修補/升級/搬遷是張單程票，不成功便成仁，缺少第二份資料的保障。



資料庫修補或升級或搬遷

### 三、資料整合、資料倉儲、商業智能

整合資料必須面臨許多挑戰，需要面臨不同來源與不同目標的資料，需要處理資料類型的轉換，需要即時的更新資料以供分析，需要不影響線上資料庫效能等等。



#### ● HVR

##### 一、基於日誌的 CDC































變更資料擷取 (CDC) 是 HVR 即時複製資料變更的方式。顧名思義，CDC 能識別資料變更，然後可以將增量變更與另一個系統同步。CDC 有多種形式，包括使用觸發器和讀取日誌的，HVR 兩種類型都支援。但是基於日誌的 CDC 更勝一籌，因為它可以應用於所有可能的情境，其好處如下：

- 不會佔用資料庫資源，對於來源系統幾乎沒有影響。
- 性能高，可以滿足極高交易量的系統的複製需求。
- 支援更多的資料操作，例如 truncate table，以及其它 DDL 操作。
- 可以設定過濾條件。
- 支援從過去某一時間點開始擷取。

## 二、支援異質平台

支援廣泛的異質平台間資料複製，包含資料庫、檔案、本地及雲端資料的雙向和多向複製。

- 來源端列表

Database		Data Lake/Files		SAP
 Azure SQL	 Amazon Aurora	 Amazon S3	 Azure DLS Gen2	 SAP ECC
 PostgreSQL	 Azure PostgreSQL	 Azure Blob Storage	 HDFS	 SAP ECC
 Microsoft SQL Server	 Azure SQL Managed Instance	 Google Cloud Storage	 SharePoint	 SAP HANA
 MariaDB	 IBM DB2 iSeries	 XML	 CSV	
 MySQL	 IBM DB2 z/OS			
 Google Cloud SQL	 IBM DB2 LUW			
 Oracle	 Oracle Exadata			
 Ingres	 SAP HANA			
 SAP ASE				
SaaS		Mainframe		
				 Salesforce
				 IBM DB2 z/OS

## ○ 目標端列表

Database	Data Lake/Files	SAP
Amazon RDS Amazon Aurora Azure SQL Azure PostgreSQL Azure Synapse Analytics Azure SQL Managed Instance Microsoft SQL Server IBM DB2 iSeries PostgreSQL IBM DB2 z/OS Google Cloud SQL IBM DB2 LUW Google BigQuery Ingres MariaDB MySQL Oracle Oracle Exadata Oracle Autonomous Data Warehouse Snowflake SAP HANA SAP ASE Greenplum Actian Vector SingleStore Teradata Tibero Yellowbrick	Amazon S3 Azure Blob Storage Google Cloud Storage Apache Hive Avro JSON XML Streaming Apache Kafka Amazon Kinesis NoSQL Apache HBase Couchbase Azure DLS Gen2 Databricks Delta Lake HDFS SharePoint CSV Parquet	SAP HANA  <b>SaaS</b> Salesforce  <b>Mainframe</b> IBM DB2 z/OS

## 三、靈活、多樣性的複製拓撲

HVR 支持跨多個異質系統的資料複製。以下是基本的複製拓撲，允許您建構任何類型的複製情境以滿足您的業務需求。

### ○ 單向（一對一）



此拓撲是從來源位置到目標位置的單向資料複製。經常出現在當客戶

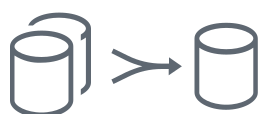
希望分散報表負載，或提供資料湖、資料倉儲資料的情境上。

○ 廣播（一對多）



此拓撲的資料從一個來源位置分佈到多個目標位置。例如，總公司的資料發佈到各個子公司。

○ 合併/整合（多對一）



此拓撲將多個來源位置的資料整合到一個目標位置。多對一是資料倉儲或資料湖整合專案的拓撲結構，或者用於多個分散式系統，將數個

資料子集（例如分公司資料）都輸入中央資料庫進行分析和報告。

○ 串聯



此拓撲將資料推送到目標位置後，該目標也成為資料來

源，又將資料分發到多個目標位置。這通常用於將資料複

製到資料倉儲和建構單獨的資料超市。

○ 雙向（Active/Active）



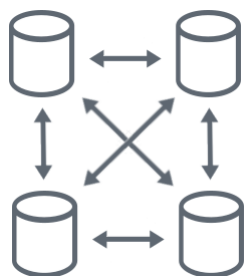
雙向拓撲的資料在兩個方向上複製，應用程式能在兩邊修改資料。它

也被稱為 Active/Active 方案，以保持兩個系統同步，從而能夠在系

統之間共享負載以及它們的高可用性。HVR 提供內建的 loopback 檢測機制來保護系統

免受無限迴圈問題，以及碰撞檢測和解決機制。

## ○ 多向



多向 Active/Active 複製涉及兩個以上保持同步的位置。這是地理分散式系統的典型場景，對任何節點所做的任何更改都將傳播到該網絡中的所有其他節點。這種類型的複製通常會帶來許多額外的挑戰：網絡延遲、頻寬可靠性。HVR 提供了應對所有這些挑戰的技術，包括

自動恢復、透過資料壓縮來最佳化網絡。

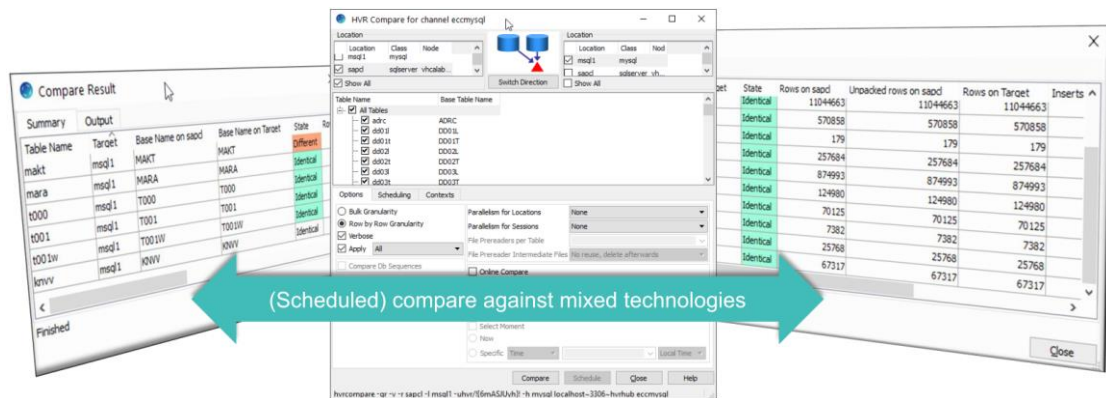
## 四、資料驗證與修復

「Trust is good, control is better.」據說這句話出自於列寧。

一旦我們建置了複製系統，是否只能被動的信任系統沒有問題？可以主動的驗證資料正確性嗎？又或者如果某人修改了目標端的資料，我們如何得知？又要如何修復？

HVR 提供了比較資料的功能，它不僅比較表格結構，還比較表格本身中的資料。您可以將單個來源位置與多個目標位置進行比較，也可以比較整個資料庫（所有表格）或僅比較特定表格。更令人驚艷的是，HVR 還支援比較異質 DBMS 中的資料，即使來源和目標資料型態可能不同，並且具有不同的字元編碼。

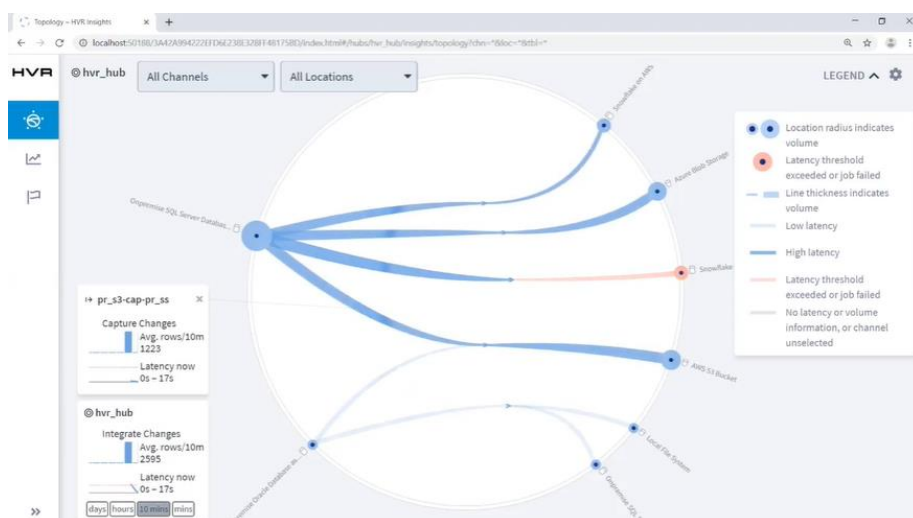
而比較後，可以僅修復不一致的資料，無須重置整個資料庫或整個表格。



## 五、監控與報表

HVR 提供視覺化的拓撲圖，清楚地顯示資料流向、每個節點複製的表格數量、複製延遲時間。讓管理複雜的資料流變得直覺化，能快速地識別阻塞點、最佳化資料，確保傳輸滿足 SLA。

此外，即時的監控 HVR 流程，能透過電子郵件、SNMP、Slack 等方式自動警報與通知，並且與第三方企業監控解決方案整合。



April 22 M-Power eNew

本篇文章版權為倍力資訊股份有限公司所有，未經書面同意，嚴禁複製、轉載



## 六、HVR 6 新功能

2021 年 12 月全新發佈的 HVR 6 提供了嶄新的功能：

- 更容易使用的網頁使用者介面。
- 提供 REST API 呼叫。
- 能解開並提取 SAP ECC 的特殊 ( Cluster、Pool、Long text ) 表格。
- 支援 SAP ASE 交易日誌擷取。

根據市場預測，大數據將繼續增長。根據多項研究和預測，到 2025 年，其全球市場規模將達到驚人的 2500 億美元。Gartner 研究也指出，在 2022 年有超過 50% 的主要業務系統，是依靠即時分析做出業務決策，資料價值將隨著時間的流逝而逝去，即時資料才能保有其競爭優勢。

總而言之，企業必須保持專注並通過採用新穎的解決方案繼續數位化轉型，並繼續改進

他們處理資料的方式，以免落後。而 HVR 會是一個明智的選擇。

◎如您有任何問題，歡迎隨時與我們聯繫：

資料整合處

處長 廖志華

02-27316868 分機 810 / charlieliao@mpinfo.com.tw

資深系統顧問 梁元瑞

02-27316868 分機 814 / yuanruiliang@mpinfo.com.tw