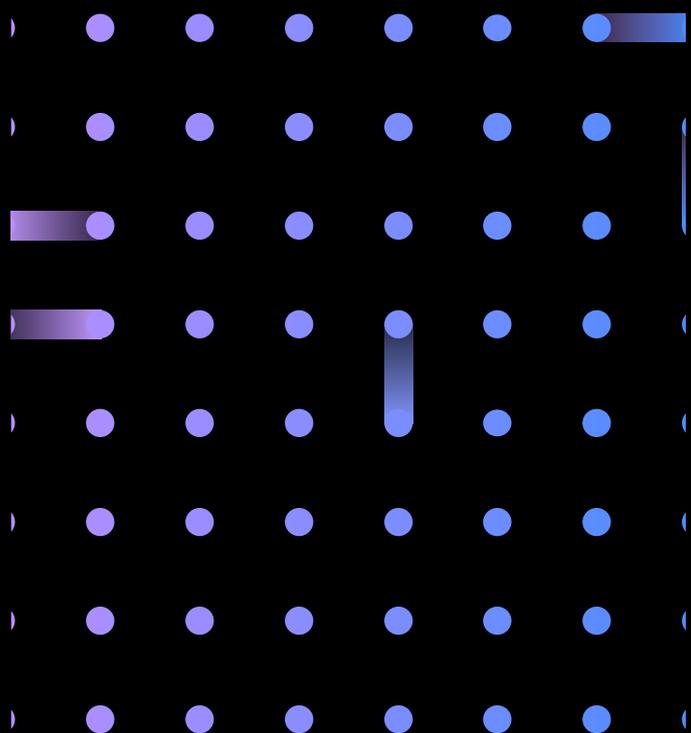


透過 DataOps 快速交付業務 就緒資料

IBM DataOps 方法論和
實踐簡介



內容

簡介	3
定義 DataOps	3
比較目標	4
IBM DataOps 方案	7
以 IBM Cloud Garage 方法論搭配 DataOps	7
DataOps 實踐成功的影響	10
結論	11
附錄:DataOps 測試方案範本	12

內容重點

- DataOps 是人員、流程和技術的和諧組合，為資料使用者快速交付高品質的可信資料。
- DataOps 以自動化為動力，可解決與存取、準備、整合和獲取資料相關的效率不彰問題。
- IBM DataOps 實踐運用絕對自動化，可對下列功能產生顯著且可測量的影響：資料維護服務、中繼資料管理、資料治理、主資料管理，和自助式互動。
- IBM 透過規範性方法論、啟用人工智慧的自動化和 IBM DataOps 卓越中心，為DataOps 實踐提供方向。
- DataOps 實作坊是 DataOps 整合性藍圖的一部分，協助組織和評估他們 DataOps 的成熟度，且可對測試專案執行進行規劃。

簡介

資料是創新和維持競爭優勢的燃料。資料是驅動分析和了解企業趨勢及機會的關鍵元素。以新方式解開資料價值，甚至可加速組織在人工智慧上的發展之途。

然而，當與資料相關的專案無法實現所承諾的投資報酬時，利益相關人會想知道原因。根據 Experian 在 2019 年所做的全球資料管理研究報告，有 89% 的企業都表示苦於管理資料。這些挑戰包括洞察不夠快，還有對基礎資料缺乏信任。¹

要對分析和人工智慧發展出有效的資料策略，了解一個組織的企業目標是一大關鍵。不論要讓哪一種企業模式運作成功，都必須滿足客戶的需求。成功端視能否以一套業務就緒的整合型資料處理流程，讓資料運作完美流暢，對任何一點都能及時提供完整一致的業務觀點。

大家對於想更快速達成結果的期望持續上升不斷。各地的企業都在尋找可以改善營運效率和有效性的方式，以做出最佳決策，尤其是因為組織中存在繁多的孤島。這兩個因素讓企業領導人想在一個單一的架構中，找出可以戰勝他們最大挑戰的方式。

對於追求資料運作轉型升級的組織來說，自動化技術可帶給他們競爭優勢。當可信的業務就緒資料可協助組織激發出具差異性的洞察和營運卓越效果時，資料就變得極具價值。

本白皮書目的在於突顯 DataOps 方法論、實踐和藍圖的效益。

定義 DataOps

資料運作 (DataOps) 是人員、流程和技術的和諧組合，為資料使用者快速交付高品質的可信資料。實踐著重於在組織間達成最佳協作，以達成敏捷、速度，並可大規模驅動新型資料倡議。DataOps 利用自動化的力量，可解決與存取、準備、整合和獲取資料相關的效率不彰問題。

DataOps 可發揮的潛在效益包括在將資訊和資料交付給個人，以及改善流程效率和優化時，所達成的顯著生產力。自動化資料運作包含以資料為優先的人工智慧倡議，此種運作有助於達成下列成果：

- 交付業務就緒的整合型資料，以此種資料大規模推動分析和人工智慧
- 達成營運效率
- 符合資料隱私和法規遵循要求

89%

的企業苦於管理資料問題。¹ 要對分析和人工智慧發展出有效的資料策略，了解一個組織的企業目標是一大關鍵。



DataOps 並非 DevOps

多數的組織在其開發進程中，都有實施某種程度的 DevOps。大家對於 DevOps 實踐廣泛的熟悉度以及在命名慣例上的相似性，已為 DataOps 這套正受矚目的實踐招引來比較。雖然兩者都是可驅動運作最佳實踐的方法論，但在組織內都有各自獨特的地位。

下表為兩項實踐之間的比較，可看出兩者對組織的目標和效益。

比較目標

	DataOps	DevOps
關鍵焦點	可供快速使用的業務就緒高品質的可信資料	應用程式和軟體開發
轉型目標	<ul style="list-style-type: none">– 藉由為所有資料使用者啟用資料存取自助服務，獲取高品質的可信資料，為企業推動持續不斷又快速的創新– 可藉自動化資料治理持續交付資料，同時又能遵循法規– 可藉由監控和優化資料處理流程，為持續學習提供所有資料使用者的反饋迴路	<ul style="list-style-type: none">– 藉由在價值鏈之間啟動協作開發和測試，加速構想的持續創新– 藉由軟體交付過程自動化和消除浪費，以持續實現這些創新，同時又能協助解決法規問題– 可藉由監控和優化以軟體驅動的創新，為持續學習提供所有資料使用者的反饋迴路
效率目標	<ul style="list-style-type: none">– 藉促進資訊系統支援、營運和企業間更緊密的連結，來矯正人員和目標未保持一致之處– 藉由在資料交付循環中導入自動化，加速變更並改善交付品質– 藉由使用可推動優化的結果，來改善對於中繼資料和資料真正價值的洞察	<ul style="list-style-type: none">– 藉促進開發人員、營運和企業間更緊密的連結，來矯正人員和目標未保持一致之處– 藉由在開發循環中導入自動化，加速對變更的實現並可刪除錯誤– 藉由使用可推動優化的客戶意見，來改善對於應用程式真正價值的洞察

DataOps 是人員、流程和技術的和諧組合，為持續保持對 DataOps 實踐的承諾，需要在所有部門間進行密切的協作。此需要對於資料管理實踐和流程的重視程度，以改善分析的速度和準確性。

人員和流程

DataOps 以自動化技術支援具高生產力的團隊，有助於對專案產出和實現所需時間達成有效率的收穫。然而，企業內部需要進化為真正以資料為主的文化，如此才能享受效益。越來越多不同性質的企業都要求也想要藉管理資料來獲取有脈絡的洞察，此刻正是進行下列事項的正確時機：

- 增加在組織內流動的資料品質和速度。
- 取得領導階層的承諾，以在企業間支援和維持一個以資料為驅動的願景。

此種轉型升級始於對企業真正目標的了解。資料如何為影響客戶的決策和服務提供資訊？資料如何協助在市場中維持競爭優勢？資料可幫我們解決的財務優先問題為何？

DataOps 領導人應為所有資料使用者所扮演的角色定義，使文化和 DataOps 實踐可向前邁進。在IT、資料科學和各業務單位中的利益相關人，都要能增加價值以驅動企業成功，這是每一個組織的獨特需求。同時，充分運用現有的資料監管委員會以及從已執行的資料治理方案中學習，有助於建立此種文化和承諾，因為監管單位是支援 DataOps 最為需要的推動力量。

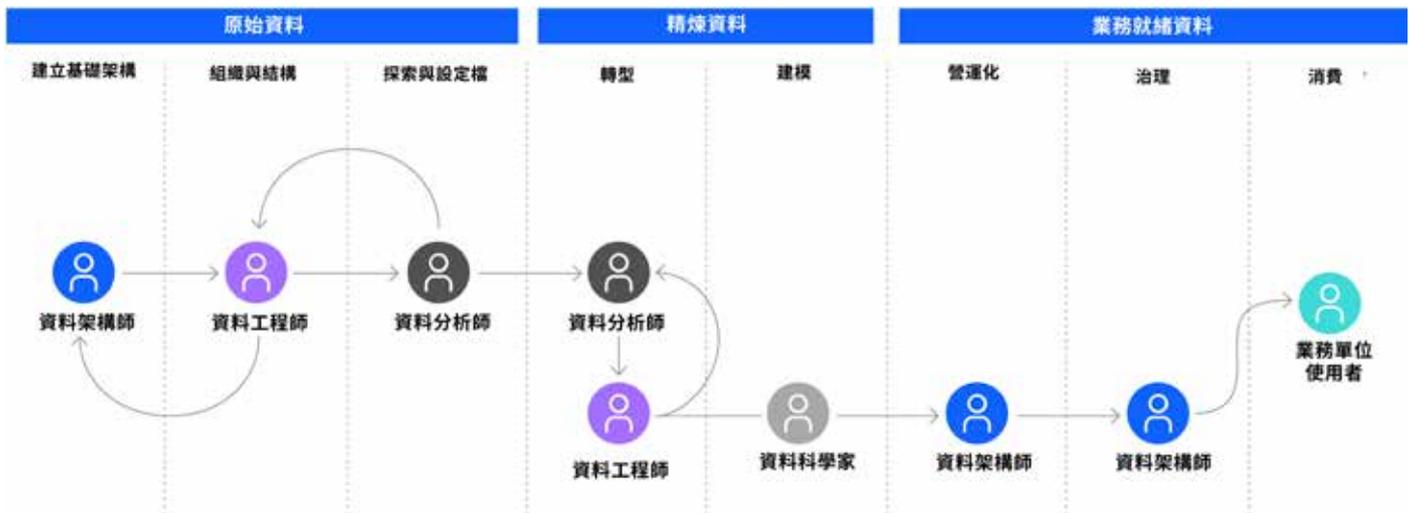


圖 1：依角色的 DataOps 工作流程範例

技術

DataOps 的核心是企業的**資訊架構**。您知道您的資料嗎？您信任您的資料嗎？您可以快速偵測錯誤嗎？您可以在不「破壞」所有的資料處理流程下增加變更嗎？為回答這些問題，第一步即是清點您的**資料治理**和**資料整合**工具及作法。必須使用工具才能支援任何依賴自動化的實踐。

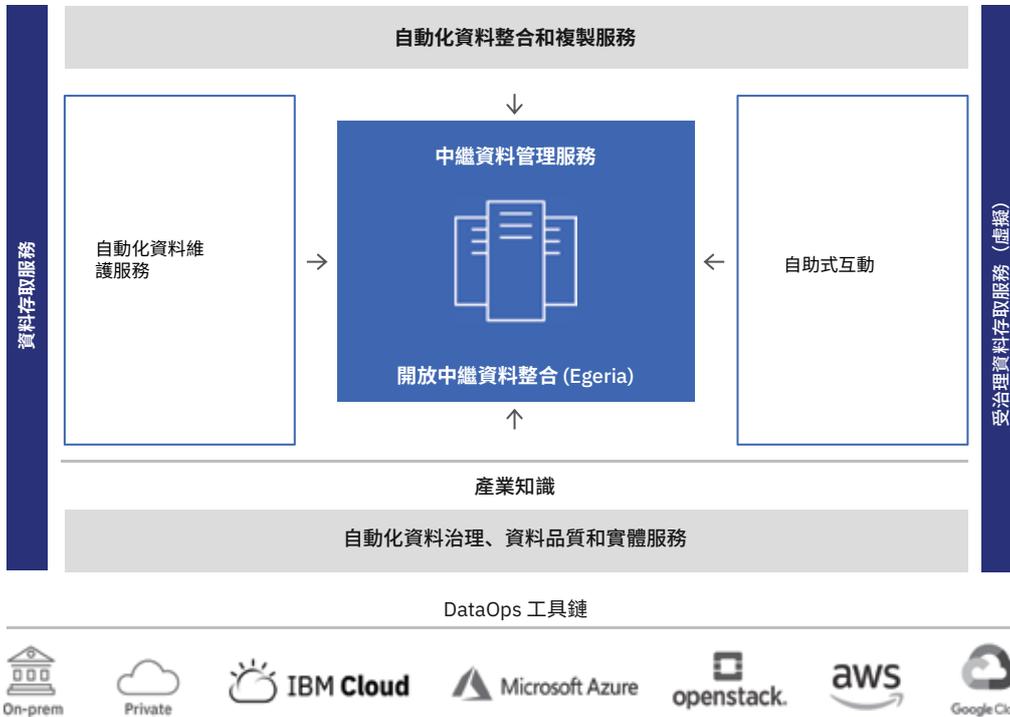
當您在考慮要使用哪些工具在組織間支援 DataOps 實踐時，可以思考自動化如何在這五個關鍵領域中為資料處理流程轉型：

1. 資料維護服務
2. 中繼資料管理
3. 資料治理
4. 主資料管理
5. 自助式互動

DataOps 功能

資料來源

紀錄系統
物聯網
洞察系統
雲端
Hadoop
社群媒體
非結構化
其他外部日誌
Logs



使用者

資料長
監管主管
資料品質分析師
資料籌備員
資料科學家
業務單位使用者
資料工程師
應用程式開發人員
應用程式測試人員

圖 2：檢查 DataOps 支援中的資訊架構

對業務就緒資料的提供包括這些所有層面，DataOps 實踐也必須包含一個可以體現這 5 個層面的完整方式。只專注在資料處理流程其中一個元素，而犧牲其他部分的組織，是不太可能從執行 DataOps 實踐中實現效益的。技術討論和執行不應該將對人員和流程的持續性規劃獨立出來。使用工具有助於支援和維持這套文化。

不同組織公認最顯著的失敗經驗在於他們的資料湖。許多組織會多次嘗試取得技術上的成功，並重振領導團隊來進行所需的**文化變革**。但為何這些資料湖執行措施在過去都失敗了？

這些專案中有多數都僅著重在將未清理和未治理的資料投入資料湖中。這些失敗很可能都是因為無法有效解決人員、流程 and 技術問題而產生的。

底線

當組織還在為有如為資料籌備人員角色定義，或為建立資料驗證規則而奮戰時，DataOps 看起來會挺令人害怕的。但是，DataOps 實踐可為組織在數位轉型倡議過程中所經歷到的失敗，提供一套解決方案。

IBM DataOps 方案

採納 DataOps 的趨勢是真實的。根據最近的一項研究，73% 的公司計畫對 DataOps 進行投資。² IBM 透過規範性方法論、領先科技和 IBM DataOps 卓越中心 (CoE)，為 DataOps 實踐提供方向。在 CoE 中，IBM 專家與各個組織共同合作，訂製出一套基於企業目標的方式，並確認正確的測試專案，為相關利益人創造出價值。

IBM DataOps 藉由提供領先業界的技術，搭配可運用人工智慧的自動化、所融入的治理功能和強有力的知識目錄，在企業中為資料維持品質並可持續操作，協助交付一套業務就緒的資料。此可增加效率、資料品質、資料搜尋能力並貫徹治理規則，以在正確的時間向正確的人從任何一個來源提供自助式資料處理流程。

IBM DataOps 協助組織在優化決策和時間時可展示資料價值，從促進對資料湖的治理，到開發應用程式和確保遵循法規。當組織了解、信任和使用資料從雲端和任何重要環境來驅動價值時，就不難交付高品質的企業資料以啟動人工智慧。

以 IBM Cloud Garage 方法論搭配 DataOps

IBM Cloud Garage 方法論是一個可以讓企業、開發和營運持續設計、交付和驗證新功能的方式。此實踐、架構和工具鏈涵蓋產品的整體生命週期，從初始經由擷取和回應客戶意見再到市場變更。此開放工具鏈架構設計旨在易於將開放來源和具領導性的第三方工具結合 IBM Cloud™ Platform 服務，例如 Continuous Delivery (CD)，成為一組與 DataOps 實踐保持一致的整合性工具鏈。這些模式可做為範本在不同團隊之間共享，以在組織內促進成功採納 DataOps。

為使 DataOps 實踐之執行成功，IBM 已就 DataOps 生命週期識別出六個階段，再加上必要的文化考量。此為根據 IBM 在內部採用 DataOps 的情況，也做為轉型的部分進程。

IBM Cloud Garage 方法論對這六階段的說明為：

- **思考**。概念化、精煉和功能的優先順序
- **準則**。建立、提升、優化和特色功能測試
- **交付**。自動化生產和服務產品交付
- **執行**。服務、選項和需要執行的功能
- **管理**。持續監控、支援和產品復原
- **學習**。持續學習和根據試驗結果的回饋



圖 3. IBM Cloud Garage 方法論六階段

思考：持續評量您的 DataOps 成熟度並與企業目標保持一致

DataOps 可被轉換為現有的組織流程和已建立的流程。DataOps 旨在為眾多現有的人工作業自動化，並使資料處理流程更順暢。不論是展開或維持一個基本 DataOps 實踐，評估團隊快速交付業務就緒資料，以及制定計劃以改善與企業價值保持一致的能力是很重要的。

DataOps 的成功始於資料資產目錄，方式是擷取中繼資料並為資料分類決定策略、為資料品質評估和評分，並充分運用工具來整合資料（並非試算表、部落知識或人工編碼）。當團隊的成熟程度被定義後，目標和目的即應為在 DataOps 間改善其越多層面的功能越好。

DataOps 團隊應著重在所交付的必要資料，是否與其帶給企業的價值保持一致性。詢問問題：如果這些資訊可以更快取得，將可省下多少錢或賺多少錢？

準則：使用版本控制系統——來源控制管理

資料處理流程是一原始碼，負責將原始內容轉換為有用資訊。此一處理流程是資料分析的核心，可端對端自動化，以產生一個可以利用一套可重製形式供使用的原始碼。如與分析關聯的不同檔案、配置和參數在無任何監管控制的情況下，被發佈到組織內的不同地方和環境，將導致部署無法維持一致。一套修訂控制工具，例如 GitHub，可協助儲存和管理所有對編碼和配置的變更。中央化儲存區也可幫助企業每次在不同環境間都擁有一致且可靠的資訊，包括任何可能發生的事或是災難都可達成可靠的復原。修訂控制也可幫助團隊將他們在開發上所投入的努力並行化，且讓他們在交付流程中藉由使用分支和合併來維持敏捷性。

為確認資料分析流程的功能運作正常，其必須經過測試。其可在持續整合和持續開發，搭配參數化的情況下，透過完全自動化的方式來部署和測試。測試輸入、輸出和企業邏輯必須在資料分析流程的每一個階段都進行，並在發表前確保品質一致，確認準確性或可能出現的變異，還有錯誤和警告。人工測試不適用於高績效組織中，因為容易出錯、費時且費力。一套健全的自動化測試套裝是達成 CI/CD 的關鍵要素，且在以需求為主的經濟中極為重要。

交付：DataOps 流程和工作流程自動化——資料技術

要讓 DataOps 方法論成功，自動化很重要，且需要一套運行靈活的資料分析流程。交付可信資料的關鍵要求是一套受治理和保持一致的資料處理流程，其依賴使用中繼資料和資料取樣技術的資料維護、資料吸收、目錄和分類。

一套可重複且健全有力的資料處理流程需要一個可以進行下列活動的機制，以交付受監管的可信資料：

- 保持一致地為資料治理和資料隱私政策進行定義和強制執行。
- 支援有效的資料移動。
- 開始實施矯正或擁有以產業區分的最佳實踐，和有預定義詞彙表的範本。

此流程可在不同平台間保持一致地部署一套受治理的資料處理流程，無需任何原始碼或配置變更，且可交付完全受監管的可信資料。DataOps 也需要以使用適當的矯正工具來支援，以支援例外處理和管理。對這些受治理的資料處理流程來說，為任何變更提供回溯性和可稽核能力是一項事實要求。

IBM 提供新的創新功能，其包含嵌入的機器學習、人工智慧自動化、融入的治理和一套強有力的資料目錄，以在企業間持續執行高品質的資料。DataOps 的效率是依賴資料技術元件的絕對自動化，而這些元件則是用於資料處理流程。

- 包含 IBM Watson® Knowledge Catalog (WKC) 的 IBM Cloud Pak for Data® 運用有效、健全、自動化和具重複性的方式來達成這些要求。
- IBM Cloud Pak for Data Server 可解決資料移動的需要，在資料處理流程中發佈和使用，同時又能協助確保資料品質和政策執行。其可透過有效的來源控制管理進行自動化，並在 CI/CD 處理流程之間有效率地執行。
- 內建在 IBM Watson Knowledge Catalog for IBM Cloud Pak for Data 中的機器學習與自動化流程相得益彰，並可利用為健全處理流程的每一個迭代來優化。
- IBM Cloud™ DevOps Insights 有助於提供營運洞察及資料處理流程視覺化。其可協助執行安全和品質措施，這些措施可持續受監控、偵測任何不預期的變異，並根據絕對自動化和自訂整合，使用 IBM Cloud Pak for Data 來建立營運分析。
- Apache Airflow 和 NiFi 可協助工作流程設計和其和諧的結合。
- 使用 REST 端點絕對自動化及參數化模式，可協助以動態方式選擇特定資料集或一個環境，並在不影響處理流程原始碼的情況下改變行為，符合資料分析專業人員每日的需求。

執行：持續整合和部署

持續整合

資料處理流程工程師或持有人可隨時更新或變更流程，並在修訂控制系統內為流程保有一個，在開發分支或私有分支中的副本。多數工程師可平行工作並同時將變更交付給開發或私有分支，還可讓生產力倍增。當流程變更在分支中完成也測試過後，原始碼可合併至主碼庫或框架，然後交付至生產線中。如被合併的碼無作用，資料處理流程始終可回到流程原始碼原先的工作版本內。進行分支和合併可讓資料分析團隊執行他們自己的測試，進行變更，並在變更集已被證明失敗時，承擔風險、試驗及丟棄。

持續部署

資料分析專業人員需要可提供給處理流程使用的相關資料，此資料有別於原始碼副本和執行這些流程的環境。直接在生產資

料庫或環境上作業並非有效率的方式，且常會造成衝突。為減少衝突和依賴性，資料處理流程需要：

- 有效率的來源控制管理
- 靈活環境部署選項可得性
- 測試資料行為

Jenkins Pipeline是一項可以與 IBM 交付流程相搭配的工具。Red Hat® OpenShift® 提供一個具可重複性、一致性的部署平台，可利用運行時所提供的不同數值，以平台來驗證資料處理流程同時發生的事件。

管理：以一致頻繁的部署來進行

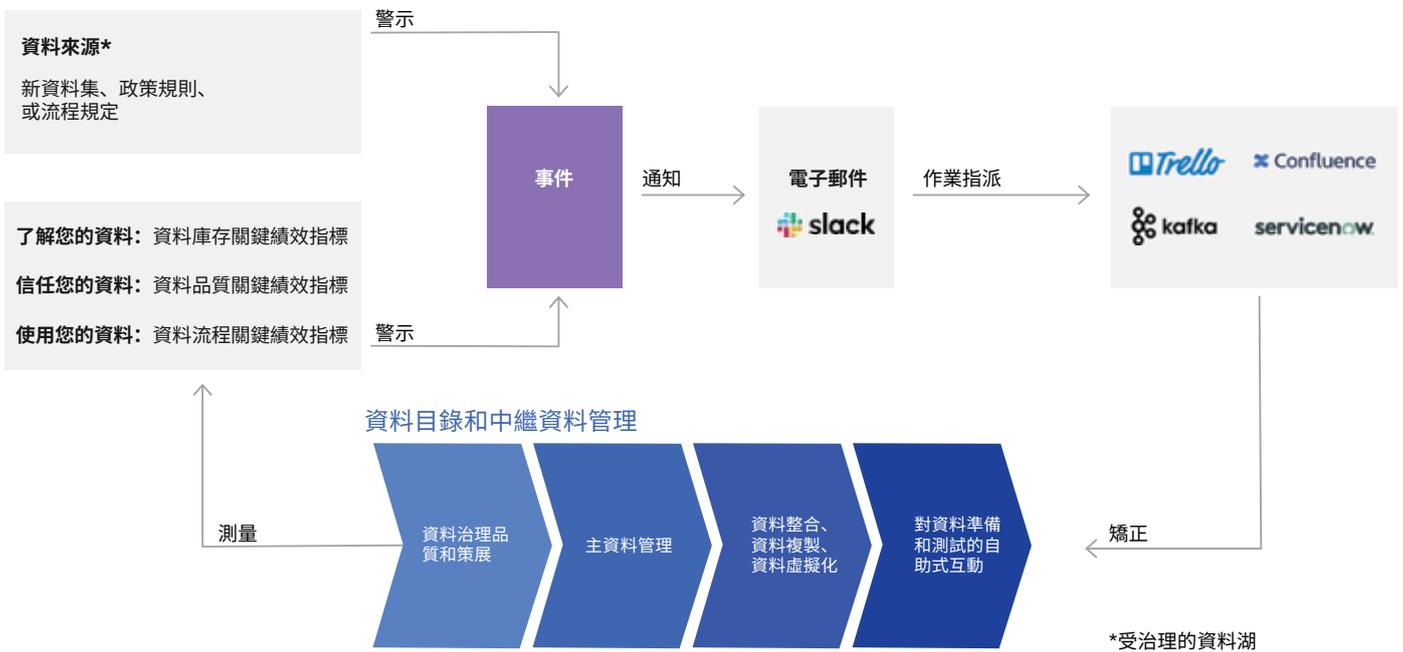
會在生產時破壞現有資料處理流程的變更，資料分析專業人員想要避免為其進行布署。兩道關鍵工作流程可解決此情況：

- 價值流程。利用 進入生產的資料流即刻建立組織價值。
- 創新流程。可開展新分析的未來，其歷經開發並加至生產管路。

兩個流程在生產中交叉，DataOps 組織在此控制交付到產品的資料組合，以及部署新功能，同時還能維持零缺點的品質。資料處理流程品質控制，例如監控資料和新開發流程的統計流程控制，開發團隊部署時無需擔心會破壞生產系統。透過敏捷開發和 DevOps，新分析的速度可最大化且快速。其有助於在將企業需求轉為分析資料，並發佈為可重複和可再使用的生產流程時，將所花費的時間和精力降至最低。

學習：通訊和流程管理

有效率和自動化的通知在 DataOps 方法論中是通訊和矯正流程的核心。當任何原始碼出現變更，或當流程被觸發、失敗、完成或部署時，就會通知。如果失敗時，資訊會與通知一起推送以突顯問題。後矯正流程可自動觸發 來驗證管路，將其部署至下一階段，並以最新資訊和資料品質更新面板。Slack、Apache Kafka、PagerDuty 和 Trello 常見用於促進跨利益相關人之間的通訊、回饋擷取和分享，做為DataOps 工具鏈的一部分。



ata and AI Forum / © 2019 IBM Corporation

圖 4：在受治理資料湖環境中的視覺化通訊和流程管理

DataOps 實踐成功的影響

藉採用一套 DataOps 實作，零售商在其資料管路中獲得改善，兩分鐘內即可讓變更在組織間執行，而之前需要花到三週的時間。結果是零售商可充分運用業務就緒資料，一天內即可對顧客喜好進行分析，而這一個流程在過去需要花費 20 天。更甚者，此套實踐降低回報庫存狀態時所需的時間，只需過去時間的六分之一。

DataOps 實踐成功的標準包含：

1. **建立資料辦公室。** 此流程在提供做為儲存資源的資料時涉及：給予角色範圍一個明確的定義，確認關鍵管理階層的利益相關人，和了解每一個利益相關人在資料處理流程中對於協作營運和文化的承諾。
2. **與企業目標一致。** 為維持競爭力，市場需要能快速回應新機會，而這僅能藉由一個有資訊性和以資料領導的方式來進行。簡言之，除非在企業和資料交付間擁有優異的通訊能力，否則領導階層知道他們的組織不會成功。
3. **擴充資料成功。** 對於每一道以資料領導的倡議，領導人需要確保所生產的資料可持續被使用和再使用，且每一次都可增加價值。此結果僅在中央能分享結果，結果可搜尋得到，且與企業語言相通時才能達成。



結論

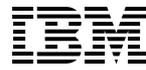
成功部署 DataOps 的組織知道他們可存取何種資料資產，他們也相信資料的意義和其品質，並將資料發揮得淋漓盡致。當業務就緒的可信資料有助於驅動與眾不同的洞察、營運卓越性、協作和競爭優勢時，資料才顯現價值。

建立一套 DataOps 實踐需要：

- 投資在一項測試專案的執行上，以瞭解組織的獨特能力和挑戰
- 運用測試專案的成功來擴展 DataOps 技巧並使組織成長
- 推廣其成功以延攬更多團隊來參與 DataOps 實踐
- 分享所學到的心得並開始建立一所 DataOps 卓越中心

採取下一步並預約您自己的 IBM DataOps Garage 工作研討會，聯絡 dataops@us.ibm.com 以加速通往業務就緒資料之路。

如果組織已經投入精神在資料目錄、資料湖泊或主資料倡議上，他們會發現自己已站穩腳步準備採用 DataOps。如需了解更多有關以市場領先技術支援 DataOps 的資訊，請造訪：ibm.com/DataOps。



© Copyright IBM Corporation 2020
IBM Corporation
New Orchard Road, Armonk, NY 10504
美國印製

2020 年 4 月

IBM、IBM 標誌、ibm.com、IBM Cloud、IBM Cloud Garage、IBM Watson 與 IBM Cloud Pak for Data 是 International Business Machines Corp. 在全球許多司法管轄區註冊的商標。其他產品及服務名稱可能為 IBM 或其他公司的商標。IBM 目前擁有的商標清單請參閱 IBM 網站的著作權與商標資訊，網址為 ibm.com/legal/copytrade.shtml

Microsoft 和 Windows 標誌是 Microsoft Corporation 在美國、其他國家或兩者的商標。

Red Hat and OpenShift 是 Red Hat, Inc. 及其子公司在美國和其他國家的註冊商標。

本文件內容為出版日期時的最新資訊，IBM 得隨時變更。並非所有 IBM 分公司所在國家皆可提供所有供應內容。

所引用的客戶範例僅供說明之用。實際效能結果可能因具體配置和操作條件而異。本文件中的資訊乃是以「現狀」提供，不具任何明示或默示的保證，也不擔保適銷性及任何特定目標的適用性，包括但不限於適銷性及特定目的適用性，以及無侵權的任何保證或條件。IBM 產品依照所提供的相關合約條件享有產品保固。

客戶有責任確保遵守適用的法律與法規。IBM 不提供法律建議或聲明或保證其服務或產品將確保客戶遵守任何法律或法規。

良好的安全措施聲明：IT 系統安全涉及透過預防、檢測和回應企業內外的不當存取行為來保護系統和資訊。不當存取可導致資訊被篡改、破壞、盜用或誤用，或可能導致系統被損壞或誤用，包括利用您的系統攻擊他人。任何 IT 系統或產品均不應被視為完全安全，並且任何單一產品、服務或安全措施均不能完全有效地防止不當使用或存取等行為。IBM 的系統、產品與服務的設計旨在成為合法且全面的安全舉措的一部分，當中必定涉及額外的操作程序，並且可能要求其他系統、產品或服務處於最有效的狀態。IBM 不保證任何系統、產品或服務可免受任何一方的惡意或非法行為的影響，或讓您的企業免受任何一方的惡意或非法行為的影響。

- 1 2019 Global data management research: Taking control in the digital age.” Experian, 2019.
- 2 Jarah Euston, “The DataOps Trend is Real: 73% of Companies Plan to Invest in DataOps to Manage Data Teams in 2018,” Nexla

10028810-TWZH-01

附錄：DataOps 測試方案範本

專案名稱：

日期：

部門或單位：

測試方案主要人員：

姓名	角色	電子郵件	電話

企業其他利益相關人

姓名	角色	電子郵件	電話

問題說明：

根本原因檢查表：

挑戰	是否可行？是/否	其他備註

成功測量標準

開始日期：

衝刺結束日期：

評量：

執行	目前狀態	預期衝刺結果	為達成預期結果的行動步驟
資料資產吸收、自動復原和分類			
資料品質評量 和矯正			
企業術語指派			
資料隱私、法規遵循以及企業政策定義和執行			
資料消費者規定定義要求處理			
資料要求通訊和通知,包括例外和錯誤處理及矯正。			
維護資料發佈至目錄			
資料處理歷程和回報			
協作、回饋與稽核			

執行稽核期間詢問問題範例：

資料資產吸收、自動復原和分類

- 我們會為企業延續完整性而進行大量低延遲的複製嗎？
- 我們會為即時低延遲分析使用先進的串流分析嗎？
- 我們是否可輕鬆連接到任何資料來源,並執行複雜的資料轉型和整合？
- 我們是否可從社群媒體、天氣資料或其他公有雲資料來源提供資料？
- 我們的資料消費者是否可從任何桌上型應用程式,即時存取我們的中繼資料儲存區？
- 我們的資料消費者在復原與工作相關的資料集時,是否可即時存取我們的資料目錄以獲得自助服務來協助復原？
- 我們是否使用資料設定檔工具來了解資料、驗證資料值、列和表之間的關係,並搜尋和分析異常之處？
- 我們的企業規則管理是否與我們的中繼資料管理基礎架構整合在一起？

