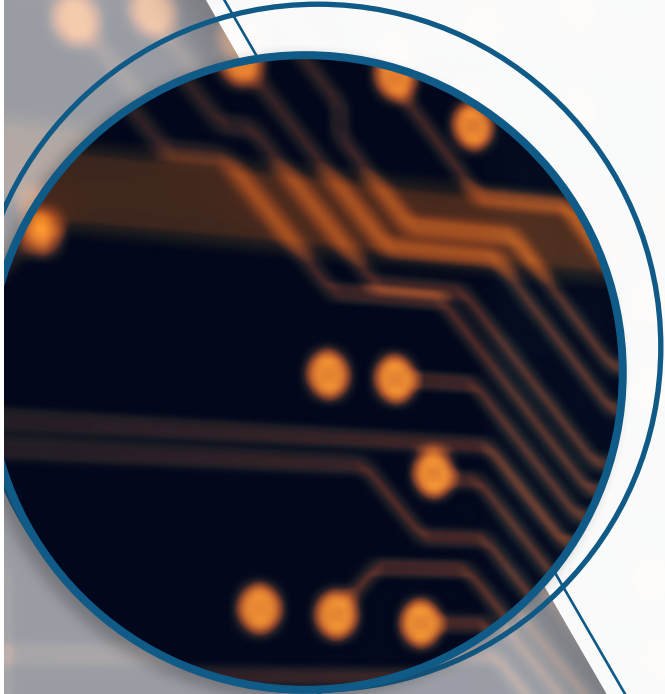




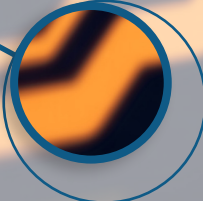
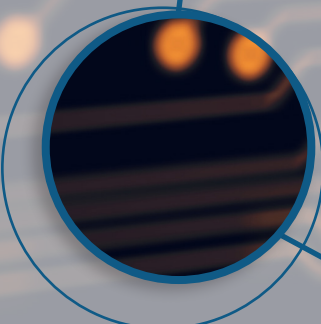
Research®  
Now a Part of

PATHFINDER REPORT

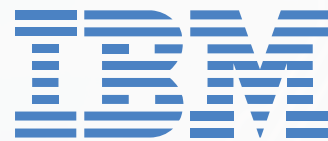
S&P Global Market Intelligence



# 透過資料管理加速 AI ; 借助 AI 加速資料管理



委任公司



2019 年 7 月

## 關於此文章

Pathfinder (探路者) 文章引導決策者解決圍繞特定技術或商業案例的議題，探索採用技術的商業價值，並提供決策過程中應考量的範圍和具體的後續步驟建議。

### 關於作者



#### **MATT ASLETT (麥特·阿斯萊特)**

RESEARCH 副總裁

Matt Aslett 是 Research 的副總裁，負責 451 Research 的資料、AI 和分析管道—包含維運和分析資料庫、Hadoop、網格/快取、串流處理、資料整合、資料治理、資料管理以及資料科學和分析、機器學習和AI。Matt 目前主要專注的領域包含分散式資料管理、資料目錄、商業智慧和分析、資料科學管理以及企業知識圖譜。



#### **JAMES CURTIS (詹姆斯·柯蒂斯)**

資料、AI 與分析資深分析師

James Curtis 是 451 Research 資料、AI 與分析管道的資深分析師。他在 BI 報表和分析領域擁有豐富的經驗，目前致力於包含 Hadoop、NoSQL 以及相關的分析 and 維運資料庫技術。

# 摘要報告

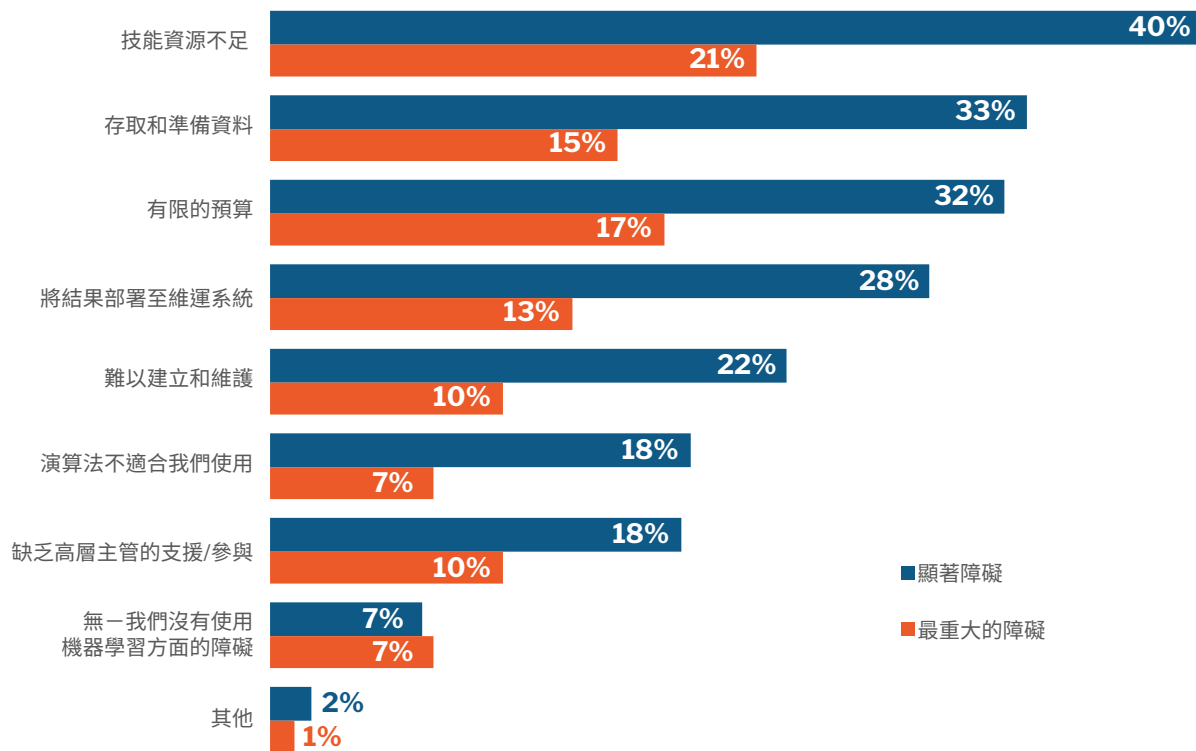
人工智慧 (AI)，包含機器學習和深度學習，將成為史上最具有變革性的技術之一，無論我們是否察覺，它都會對我們生活的多數層面造成影響。這些技術的應用可能會重塑人們的工作、學習、旅行、治理、消費和休閒的方式。

成功的 AI 取決於許多因素，包括大量的資料集、必要的演算法、具有適當技能的資料科學專家以及適當的計算資源。後者不僅指物理和虛擬伺服器基礎架構，還有旨在支援高效能資料處理和分析的資料管理和資料庫軟體。

資料管理是機器學習專案的關鍵啟動推力，因為它有助於克服像存取和準備資料的挑戰，這些挑戰是成功的主要障礙。451 Research 企業之聲 (Voice of the Enterprise) 的結果：與直接參與 AI 和 ML (機器學習) 專案的人員所作的《AI 和機器學習問卷調查》顯示了這一點：33% 的受訪者提出存取和準備資料是使用機器學習的障礙，而 15% 的受訪者提出這是最重要的障礙。

圖 1: 使用機器學習的障礙

資料來源: 451 Research 企業之聲 (Voice of the Enterprise): AI 與機器學習 2H 2018



% 的受訪者 (n=745, 738)

PATHFINDER | 透過資料管理加速 AI; 借助 AI 加速資料管理

雖然資料管理可以改善 AI 應用程式的開發，但 AI 可以用來改善資料擷取和查詢效能等領域的資料管理，使資料工程師能夠加速資料管理和分析專案，並使資料庫管理者 (DBA) 專注在影響更多的任務上。

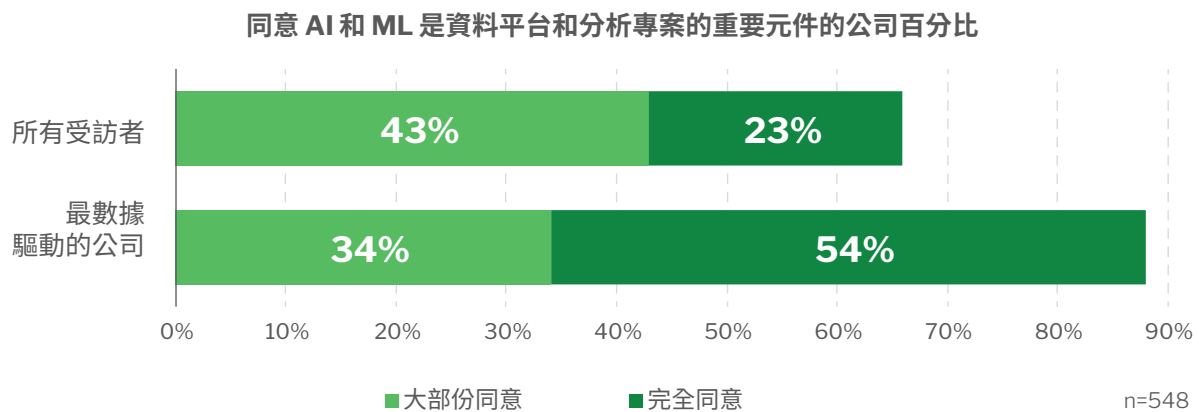
為了充分運用 AI 的所有優點，企業必須從資料層級開始在多個層級嵌入 AI，以確保 AI 能夠在整個資料管理週期中發揮作用—從擷取到策劃和探索，以及驅動為利用資料而構建的應用程式。

451 Research 企業之聲 (Voice of the Enterprise) 的結果：與負責資料平台和分析專案的人員所作的資料平台和分析調查，顯示了企業將 AI 和 ML 視為其資料和分析項目的關鍵層面的程度。三分之二的受訪者同意 AI 和 ML 是他們的資料平台和分析專案的重要元件，但在最以數據的驅動公司 (例如，幾乎所有策略決定都是基於數據) 中，此數字上升到 88%。

圖 2: AI 和機器學習對資料平台和分析專案的重要性

資料來源: 451 Research 企業之聲 (Voice of the Enterprise) : 資料和分析, 1H 2019

在資料和資料管理系統方面，資料架構師和資料庫管理者 (DBA) 在諸多領域面臨了挑戰：他們必須提高維運效率，同時為更多類型的資料消費者提供資料存取權限—包括資料和商業分析師、高級主管、開發人員和資料科學家們。為此，資料庫管理者需要高效運行並能夠產生準確結果的資料管理系統。



此外，企業需要資料科學家可以輕鬆存取資料以構建支援 AI 的應用程式。隨著基礎資料平台的發展來提供 AI 專案更好的支援，這些系統需要提供多種語言的支援，例如 Python、GO、JSON 和 Jupyter notebooks，以加速基於 AI 的應用程式開發和混合式資料模型的構建。

# 資料管理與 AI 的協同效應

資料管理系統和 AI 是相互增效的。當 AI 嵌入整個資料管理系統中時，它就具有能提高資料庫查詢準確度和效能的潛力，並能優化系統資源、減輕 DBA 的負擔，同時改善資料科學家和開發人員的資料存取。

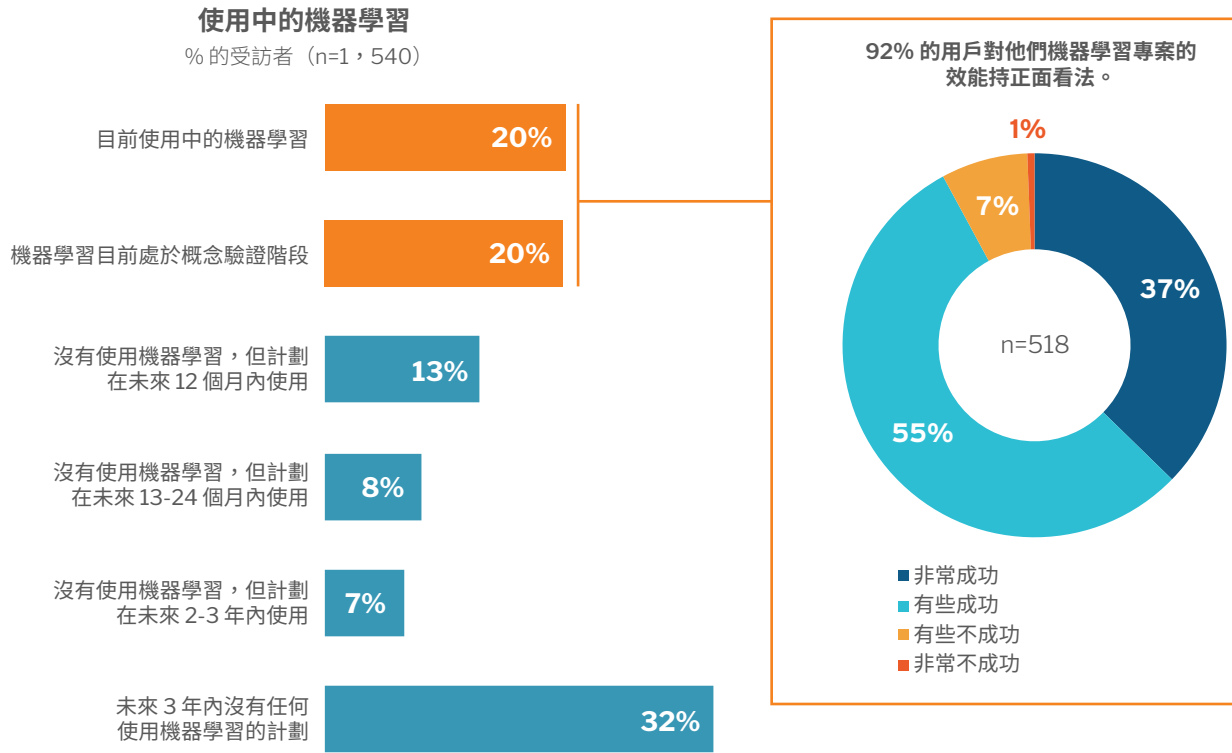
企業在尋求改善資料管理系統的運營效率，同時讓各種資料使用者擁有更多的資料存取權時（尤其是資料分析師和資料科學家）面臨到的挑戰與日俱增。一種解決方式是採用內建 AI 功能的單點解決方案產品。但是，這可能只能解決資料管理環境的一個層面。將資料從一種產品複製和移動到另一種產品方式也可能會增加營運費用。另一種方法是將 AI 嵌入整個環境中，尤其是在所有資料使用者用為共同基礎的資料層。

當 AI 被嵌入資料層時，它會在底層資料管理系統和 AI 應用程式的開發之間建立可能會影響整個資料生命週期的協同關係。換句話說，資料庫管理者和架構師更能夠管理和監督資料，進而產生構建應用程式和服務的資料科學家和開發人員可以更有效存取的高品質資料。

451 Research 企業之聲 (Voice of the Enterprise) 中的資料：《AI 和機器學習問卷調查》顯示，企業採用的機器學習與日俱增。超過三分之二的企業已經採用機器學習，或計劃在未來兩到三年內採用機器學習。也許更明顯的是，在 40% 表示已經使用機器學習或處於概念驗證階段的受訪者中，有 92% 對他們機器學習專案的效能持正面看法，其中 37% 認為他們非常成功，而 55% 認為他們有些成功。

圖 3: 企業對機器學習的採用和態度

資料來源: 451 Research 企業之聲 (Voice of the Enterprise) : AI 與機器學習 2H 2018



人工智慧在多個行業中有許多使用案例。在高層級，這些使用案例可以被認為是頻譜的一部分：在頻譜的一端，部署 AI 以自動執行高度複雜的任務，而在另一端則使用 AI 將可預測和重複的任務自動化。在中層級，它可以用來擴增介於兩者之間做所有事情的人類。



圖 4: 使用 AI 案例的頻譜

資料來源: 451 Research



此頻譜可以用於一窺 AI 應用於資料管理的多種方式,或是資料管理應用於 AI 的方式。雖然這是 AI 和資料管理協同效應方式的高層級視圖,但每種方式都有多個案例,將於以下詳細說明。

圖 5: 與資料管理相關的 AI 使用案例頻譜

資料來源: 451 Research



PATHFINDER | 透過資料管理加速 AI; 借助 AI 加速資料管理

## 將重複的任務自動化

如上所述，資料擷取和準備是 AI 資料管道的基本層面。它們也是最費力和耗時的一些任務，所以成為適合自動化的主要對象。機器學習經由自動辨識和標記資料以減少手動資料準備的需求，在提高管道此階段的效率方面扮演著關鍵的角色。

雖然資料庫和優良的資料管理是加速 AI 的重要因素，但使用 AI 來改善資料管理也具有潛在的優勢，特別在將重複性和稍有預測性的任務自動化方面。聚焦的範疇之一是提高資料庫的效率——特別是查詢的優化。有兩個方面可以加強查詢流程：效能和準確度。但這兩者都不可被認為是單調日常任務——實際上，它們都需要資料工程師和 DBA 很高的精確度——但它們是可能透過自動化加速的高度重複和手動性的任務。

在效能方面，企業經常在確保資料庫系統有效運行時面臨很大的挑戰。使系統超負荷、消耗過多資源或影響其他運行中作業的查詢不僅影響效能，並且還需要花費人為資源來修正。AI 可以透過基於資源消耗來幫助自動化查詢管理，從而提供更穩定可靠、可以優先排序查詢、減少手動治理和資料庫監視的系統。

關於查詢優化，結合 AI 的資料庫可以大幅提高與已安排的查詢相關的估計基數準確度，並建立查詢計劃以產生最高的效率。將機器學習應用於查詢優化過程，能夠產生因不斷反饋所帶來的持續基數估計改善。總體推動力是透過基於 AI 方法以更優化的方式執行查詢，企業不僅可以減少產生洞察所需的時間，並且可以做出更多更有信心的商業決策。

AI 在其他領域可以將日常和可預測的任務自動化，並且這些領域可以與擴增人類的任務重疊。然而，差異是在於人類的參與程度。查詢優化（如上所述）幾乎沒有人為的參與。可能被認為具有減少人工干預優點的其他焦點領域包括：應用安全更新以防禦外部攻擊、管理資料庫正常運行時間和可用性與預期基礎結構故障，以及諸如監視和調整以提高效能的資料庫管理。

再次強調，我們的目標不是杜絕人為參與，而是要利用自動化的優勢，使資料庫管理者和資料工程師能夠專注於能產生更大影響的任務。



## 對介於兩者之間的一切增益

資料庫層級的 AI 可以在兩個關鍵領域擴增人類，以提高商業的整體維運效率並加快商業決策速度：授權商業分析師並不斷開展資料庫管理者的角色。關於前者，歷史上關於分析專案的主要挑戰之一是透過使更廣泛的人們做出分析驅動的決策來使商業智慧「民主化」。

如此看來，451 Research 估計，目前全球有一到二百萬個資料科學家，約有5百萬到1千萬個商業智慧「超級用戶」，5千萬到6千5百萬個商業智慧「臨時用戶」和2億到2.5億個知識工作者。

圖 6：資料科學家和業務單位智慧使用者/消費者的潛在數量

資料來源：451 Research

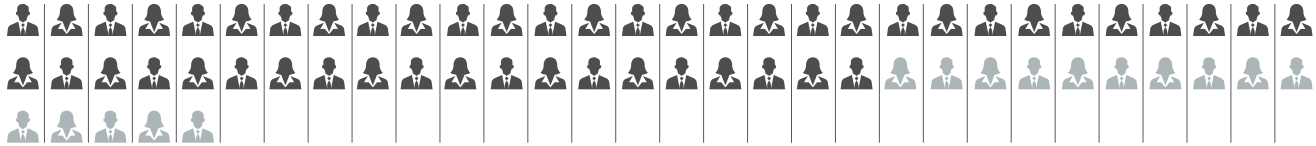
資料科學家：1-2 百萬



BI 「超級用戶」：5百萬 - 1 千萬



BI 「臨時用戶」：50千萬 - 6 千 5 百萬



知識工作者：2 - 2.5 億



儘管資料科學家扮演著無可否認的重要角色，但他們相對的匱乏代表聘雇和留住他們的成本高昂。從提供有效結果所需的資料驅動決策的深度和廣度來看，能確保從資料科學專案產生的智慧到達商業智慧用戶和知識工作者手中，無疑地能帶來更多益處。

旨在補充資料科學的資料庫和資料管理軟體構成了可以加速基於 AI 的應用程式開發，支援開發更多應用程式並使機器學習模型的輸出可更快速地提供給各領域的專家和商業決策者。例如，從 SQL 調用機器學習模型作為用戶定義的功能的能力可以幫助維運機器學習專案，將機器學習模型的使用對資料工程師和業務分析師透明化。

資料處理的自動化和加速也對資料工程師和 DBA 產生影響。資料庫管理的自動化是持續的過程，但是至少已經可以將部分重複性任務自動化，包括：資料庫部署和修補；效能調試；備份，高度可用性和災難復原；資料庫優化；查詢優化；和架構變更。

雖然冗餘仍待解決，這對資料工程師/DBA 所扮演的角色有著影響。我們在此看到了與無人駕駛汽車發展類似的例子。儘管關於自動駕駛汽車的議題很多，但對於現今（以及可預見的未來）主流的使用案例而言，現實更接近於駕駛員輔助以及部分或條件自動化。

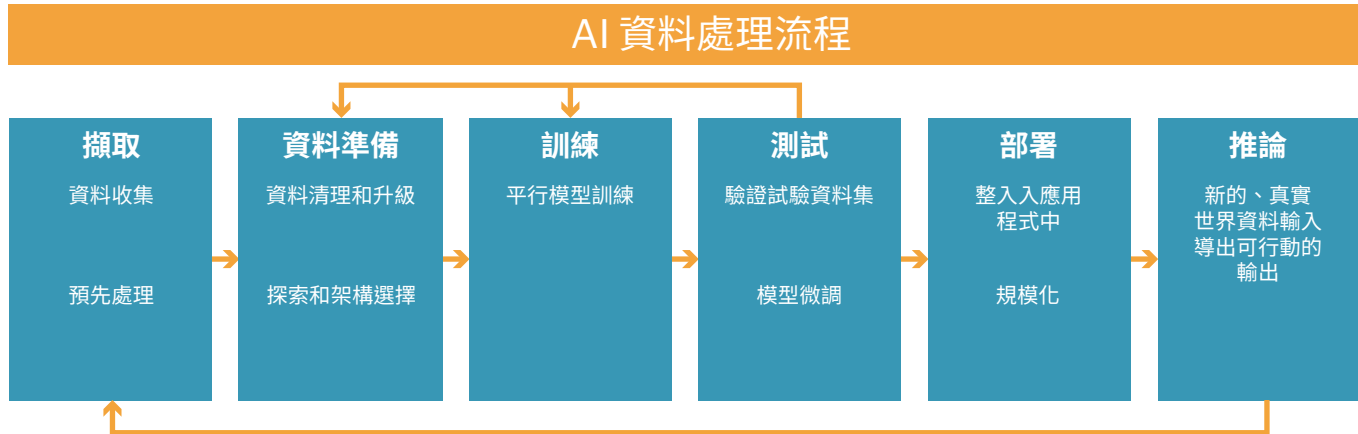
同樣地，自動化可以擴增 DBA，而不是即將取代 DBA。具體而言，有一些影響較大的任務目前無法自動化，例如架構規劃和資料建模；資料安全和生命週期管理；專用於應用程式的調整；服務水準管理；測試資料管理；和資料庫分區化。

## 將複雜的任務自動化

如上所述，451 Research 的調查結果顯示，存取和準備資料是採用機器學習的最顯著的障礙之一。此外，資料管理很明顯地在基於 AI 的應用程式開發和部署中扮演著基本的角色。誠如 AI 資料處理流程的說明（見圖 7）所示，AI 應用程式的開發始於資料的擷取和準備。因此，加速資料探索的資料庫和資料管理產品可以為更高效地開發 AI 應用程式奠定基礎。

圖 7: AI 資料處理流程

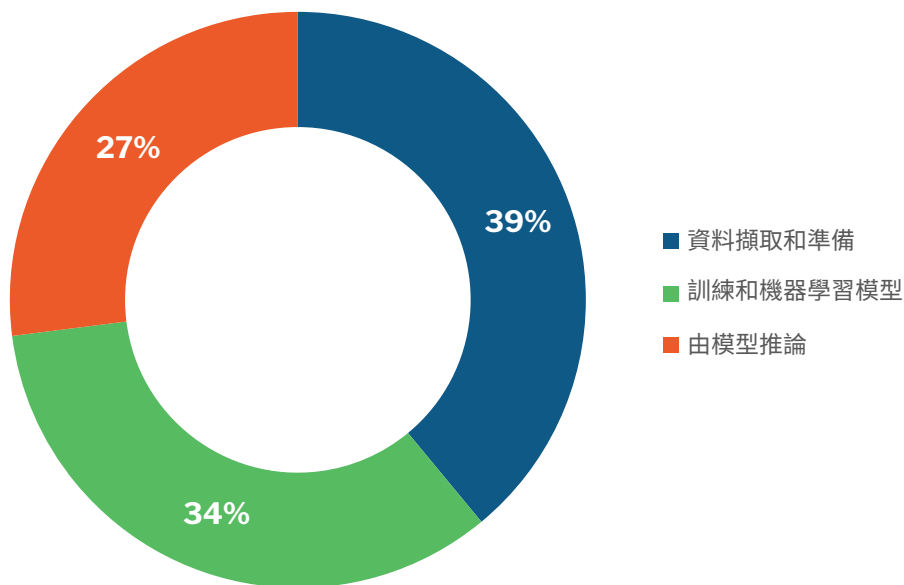
資料來源: 451 Research



資料研究和準備階段的重要不容忽視，因為 451 Research 企業之聲 (Voice of the Enterprise) 有 39% 的受訪者表示：《AI 和機器學習調查》顯示，他們認為這一階段的 AI 流程對他們基層基礎設施的要求最高，相較之下，訓練為 34%、推論為 27%。因此，透過選擇和設定旨在支援更有效的資料擷取和準備的資料庫，資料架構師可以幫助加速 AI 應用程式的開發。

圖 8: AI 流程中最艱鉅的階段

資料來源: 451 Research 企業之聲 (Voice of the Enterprise) : AI 與機器學習 2H 2018



PATHFINDER | 透過資料管理加速 AI; 借助 AI 加速資料管理

但是，正如 AI 資料處理流程(圖 7)所示，資料擷取和資料準備不限於開發過程開端，而是貫穿了訓練、測試和推論階段。支援多種資料類型的多模型資料庫可以降低整合來自多個資料平台的資料的需求，而資料虛擬化功能可以透過允許存取和查詢資料所在的位置來幫助減少資料擷取的營運費用。

此外，與提供直觀(和擴增的)資料探查功能的資料管理工具緊密整合的資料庫可以幫助降低資料準備的花費，並可能在各方面加速 AI 生命週期。使用 AI 的資料庫在減少開發和訓練以及測試、部署和推論的時間方面也扮演著重要角色，例如，透過整合像 Python、GO、JSON 和 Jupyter notebooks 的開發語言和架構。

使用 AI 的資料庫也可以包含直接嵌入資料庫中的機器學習功能。優勢包含了，透過避免用於機器學習開發的獨立環境，需要管理的系統更少，並且可以存取完整資料，而無需從將資料從資料庫中取出並移至分開的系統。另一個相關的優勢是減少了資料移動的費用。

儘管市場上各種資料庫供應商的策略和方式各異，但原則上是相同的：在過去於不同環境中進行的兩個不同活動之間提高效率。

## 結論

在高層級中，支援 AI 的資料庫透過兩種方式與 AI 結合：透過嵌入 AI 來改善資料管理效能，及使用 AI 功能來加速機器學習模型的開發和 AI 應用程式的部署。將 AI 嵌入資料管理系統功能的一部分，可以包含內建功能，例如，可以啟用查詢優化和工作負載管理。但是，AI 也可以與安全、可用、規模化、調整、基礎結構故障、備份和復原以及由 DBA 管理的許多其他任務融合。

然而，資料工程師和 DBA 不一定會受 AI 功能的威脅，主要因為 AI 具有處理許多日常或重複性行政事務的能力。在這層意義上，資料庫管理者和架構師可以專注在 AI 還尚未成熟、影響較大的任務上，像是體系結構規劃、資料建模、生命週期管理、專用於應用程式的調整、服務層級管理、測試資料管理和資料庫分區。重要的一點是，儘管許多資料庫管理任務可以使用 AI 自動執行，但仍有許多任務需要人類專業知識。


AI 功能也可以用作資料管理流程的一個環節來加速資料存取、機器學習模型的開發以及 AI 應用程式的部署。這種方式有其優勢，因為它可以扮演幾個不同的角色。例如，資料和商業分析人員可能沒有 SQL 技能以有效查詢資料，因此使用 AI 的功能可以協助以自然語言搜尋資料，將查詢語法抽離用戶。

此外，內建的 AI 功能可以協助應用程式開發人員，特別是在資料庫系統允許與開發人員語言、工具和架構整合的情況下。對於資料科學家而言，AI 功能可能包含了機器學習演算法和功能，可用於更深入理解、準備和評估資料。

儘管完全支持 AI 的資料庫的概念仍處於早期階段，但仍有許多優勢等待著企業勇於部署這些系統，而不僅是啟用資料庫、系統效能、資料虛擬化和減輕的管理負擔。真正的優勢可能是這些協同效應的加乘結果，例如，與那些在 AI 方面落後的企業比較，具有明顯的競爭優勢。

## 建議

- 儘早確認 AI 的潛在使用案例，尤其是明顯為手動操作的任務。沒有投資 AI 應用程式開發的企業可能遠落後於那些已經在早期計劃中取得成功的企業而望塵莫及。
- 企業必須在資料層級嵌入 AI 來確保他們正在部署旨在支援 AI 功能的資料庫，以確保 AI 能夠在整個資料管理週期中發揮作用——從擷取到策劃和探索，以及驅動為利用資料而構建的應用程式。
- 資料架構師應確保他們提供的資料基礎架構能夠透過直接支援機器學習工具和架構，以及加速完整的 AI 資料處理流程來支援 AI 應用程式的快速開發。
- 企業應該探索自動化對資料工程師和 DBA 角色的潛在影響，確定可以加以自動化的重複性任務，以及確定可以從人類專業知識中受益的更具影響力的任務。
- 企業需要確保來自多個利益相關者和用戶的認同。雖然許多組織對 AI 專案的想法持開放態度，但許多組織缺乏基本知識和背景而無法真正理解管理這些類型的系統的真正意義。一些簡單的成功專案往往在組織內部走了一大段路，以確保利益相關者同意。



IBM Hybrid Data Management 是 AI 優化的資料基礎，它完全整合了機器學習和分析功能，以為各種規模和各個行業的企業驅動洞察。它包括維運資料庫、資料倉儲、資料湖和快速資料，以及內部部署、IBM Cloud Pak for Data 和公有雲部署的選項。

使用 AI 支援的 IBM Hybrid Data Management 策略，組織可以透過多種形式將資料釋放到完全彈性的資料管理服務範圍中，同時還能保持一致且簡單的體驗，提高運營效率並大幅降低維運成本。使用 AI 的資料庫能夠在整體上運用來自所有來源的資料，同時以自助服務的方式為組織提供豐富的客戶觀點。

欲瞭解更多資訊，請見：

*Modernizing your Information Architecture with Ai*  
(運用 AI 將資訊架構現代化) (網路研討會)

[www.ibm.com/db2](http://www.ibm.com/db2)

內容提供者



PATHFINDER | 透過資料管理加速 AI；借助 AI 加速資料管理



Now a Part of

**S&P Global** Market Intelligence

由 IBM 委任

14





**451 Research®**

Now a Part of

**S&P Global** Market Intelligence

## 關於 451 Research

451 Research 是一家領導業界的資訊技術研究與諮詢公司，致力於技術創新和市場干擾。透過組合研究和資料、諮詢和進入市場服務 (go-to-market services) 以及現場活動，超過 100 位分析師和顧問為全球 1,000 多家客戶組織提供了重要的洞察。451 Research 成立於 2000 年，是標普全球市場情報 (S&P Global Market Intelligence) 的一部分。

© 2020 年標普全球市場情報 (S&P Global Market Intelligence)。保留所有權利。未經標普全球市場情報事先書面許可，嚴禁以任何形式全部或部分複製和分發本刊物。內部和外部有關分發的使用條款應受您與 451 Research 和/或其附屬公司的《服務協議》中規定的條款約束。本文內容包含的資訊是從相信為可靠的來源取得。451 Research 和標普全球市場情報對這類資訊的準確性、完整性或充分性不作任何保證。儘管 451 Research 可能會討論與資訊技術商業有關的法律問題，但 451 Research 不提供法律諮詢或服務，因此不應將其研究詮釋或作為法律諮詢或服務。

此文章的內容僅用於教育目的。標普全球市場情報不為任何公司、技術、產品、服務或解決方案作做廣告。標普全球市場情報對本文所包含的資訊或其解釋的錯誤、遺漏或不足之處不承擔責任。讀者對選擇這些資料以達到其預期結果負全責。本文所表達的觀點若有變更，恕不另行通知。



### NEW YORK

55 Water Street  
New York, NY 10041  
+1 212 505 3030



### 舊金山

One California Street,  
31st Floor  
San Francisco, CA 94111  
+1 212 505 3030



### 倫敦

20 Canada Square  
Canary Wharf  
London E14 5LH, UK  
+44 (0) 203 929 5700



### 波士頓

75-101 Federal Street  
Boston, MA 02110  
+1 617 598 7200