

透過整合 SAS®  
及開源程式推動  
分析創新

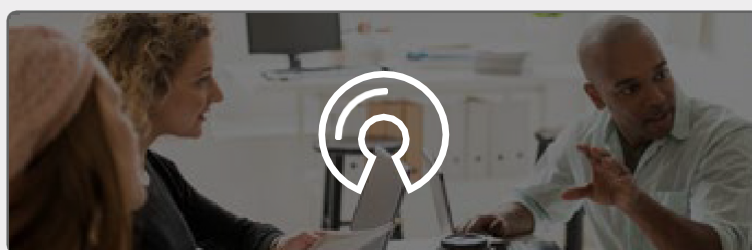
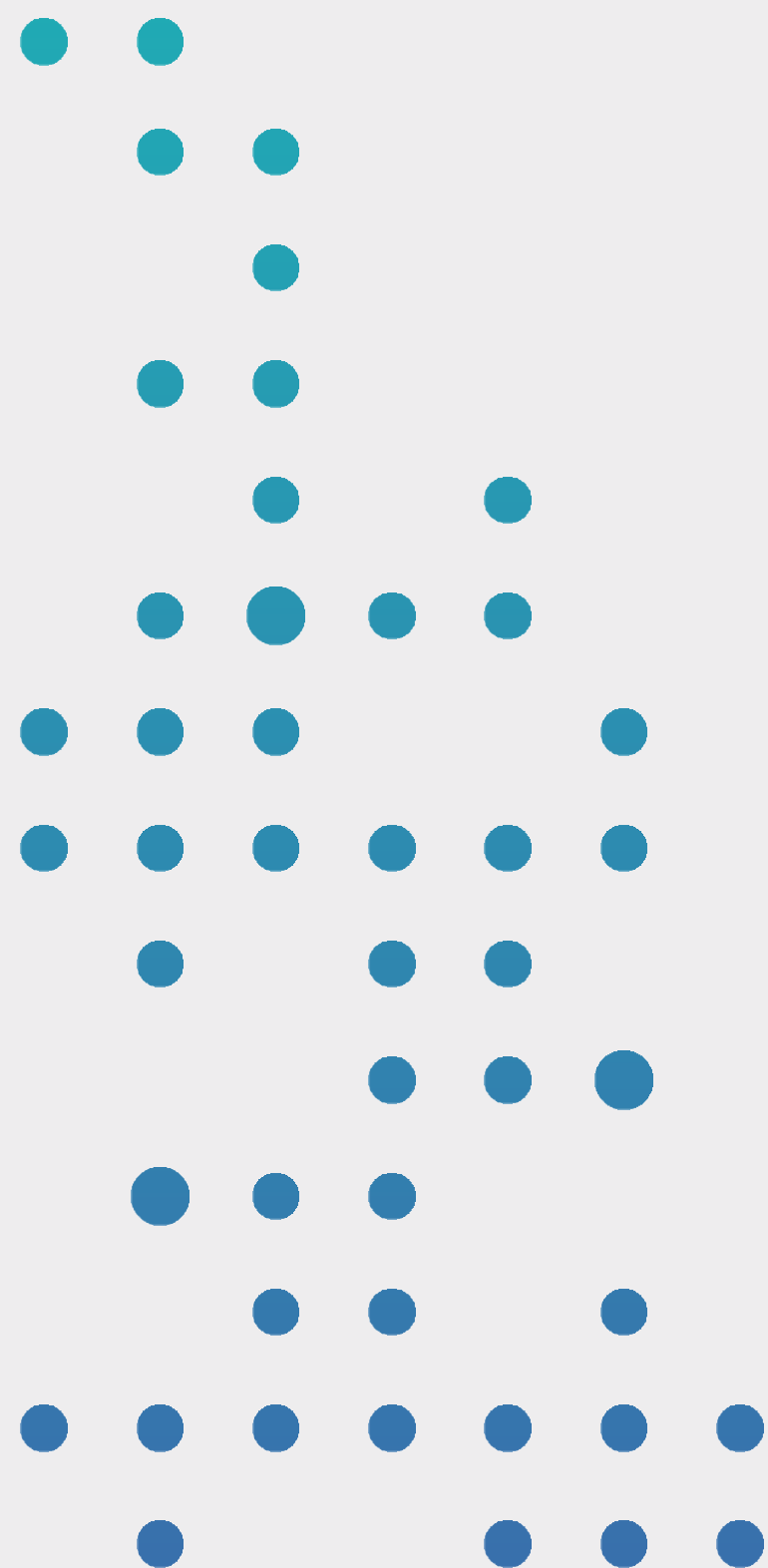




## 本電子書的目的

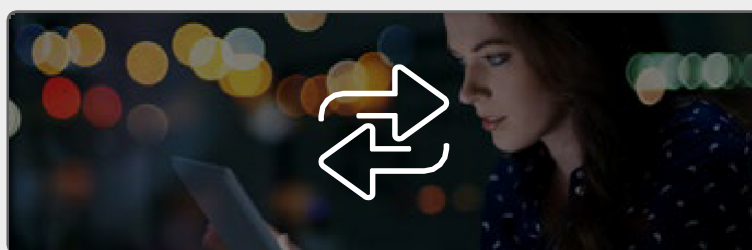
本電子書提供給希望了解如何使用 SAS 和開源程式來推動分析創造價值並做出可靠決策的人士。無論您是對開源程式感興趣的 SAS 使用者，還是想使用 SAS 的開源程式使用者，此電子書都將協助您入門。

# 內容



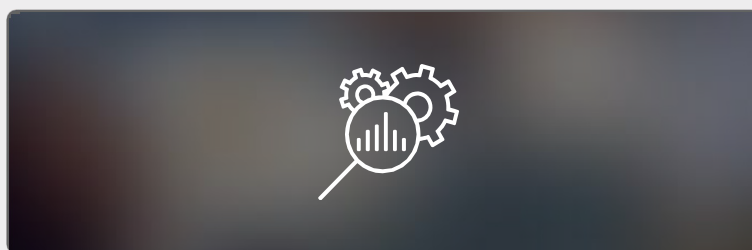
4

是什麼推動了當今的分析創新？



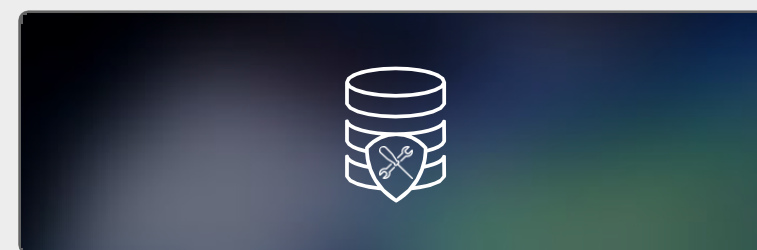
10

跨分析生命週期的整合



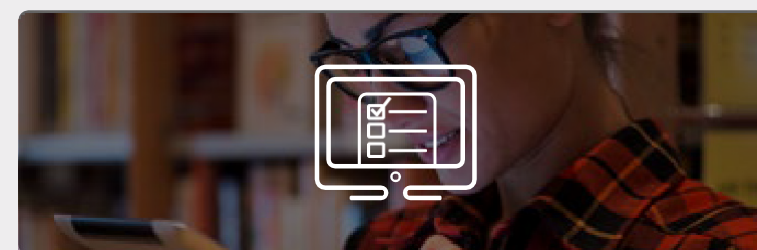
16

模型建立



17

模型管理



19

模型部署



21

後續步驟



是什麼推動了當今的分析創新？



# 當今 AI 和分析的挑戰

影響業務成效、增加分析成本和拖慢創新步伐的四大挑戰。

**資料：**在許多組織中，我們需要業務團隊、分析團隊、應用程式開發人員和 IT 營運團隊之間進行更多協作。這些團隊經常在資料孤島中處理資料，最終導致重複工作、無法整合或錯過挖掘資料中蘊含之價值的機會。

除了孤立的工作之外，資料科學家還面臨著資料量和資料獲取速度不斷增加的問題。並且，人們期望他們回答問題的速度能像以前一樣快，或者比以前更快。

使用正確的資料和正確的技術來確保最佳結果非常重要。

**技術：**我們在大學裡學到的是，資料不會神奇地以這些現成適合分析的格式出現。需要大量的轉換工作來獲取關聯式資料並將其轉換為支援模型開發的格式。

這歸結於您能以多快的速度探索和確定正確的運算方法，並最終訓練一個或多個模型以實現您的分析目標。

好的，您已經建立了一個模型。但在部署並整合到營運環境之前，這對業務成果不會有任何影響。

整個過程花費的時間越長，產生的模型進入營運流程時，面臨資料不再有效的風險就越大。

如果市場條件發生了變化，您能以多快速度確定自己的模型效力是否在降低？您能以多快速度重新訓練？能以多快速度進行重新部署？

資料科學家必須回答所有的問題以及更多其他的問題。

» 當今 AI 和分析的挑戰

» 與開源程式無縫整合，並以此作為成功因素

» 為什麼要使用 AI、分析和資料管理平台？

» 資料科學家將獲得哪些益處

是什麼推動了當今的分析創新？

跨分析生命週期的整合

模型建立

模型管理

模型部署

後續步驟



**流程：**在開發方面，資料科學家通常不願花時間處理諸如資料與模型的文件、可追溯性、版本設定或可解釋性等日常活動。這些日常活動能提供對模型內容的理解和的模型績效的管控。

這些因素增加了模型試驗失敗的可能性，成為一個永遠無法加入營運的模型。

但是，建立一個有效的流程即可確保管理和透明度，從而提高整體效率、生產力，最終提供可檢視的流程以及具可重複利用性等特點。

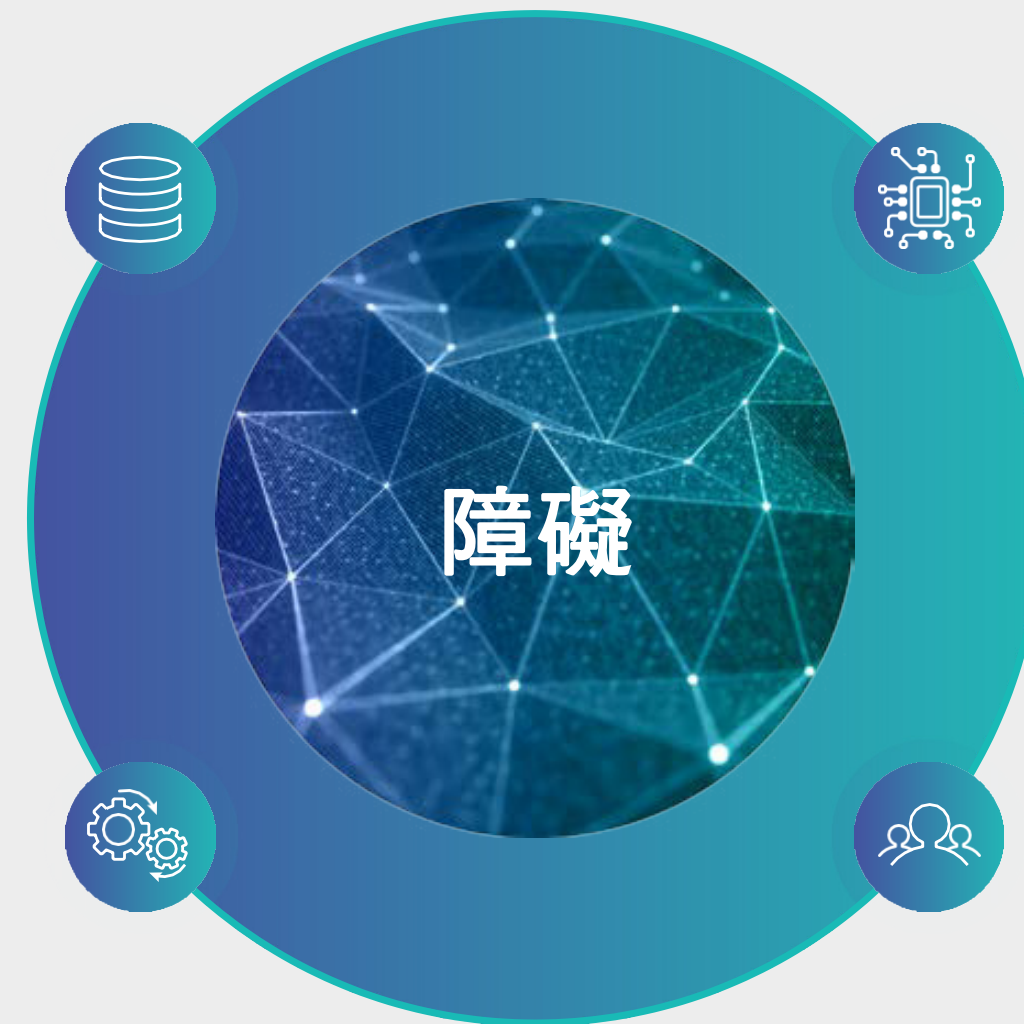
**人員：**只有當您能夠將人員置於合適的位置，並讓他們能夠使用正確的工具來支援創新時，這樣才是一個有效的流程。資料科學家必須能夠以一致的方式迅速適應和更新這些工具，讓他們能夠將更多時間用於運用工具創造價值，而不是將時間花在維護工具上。

**數據**  
孤立的、重複的資料；  
速度提升和  
資料量增加

**流程**  
治理；可重複性  
和透明度

**科技**  
各種工具；  
可擴展性

**人員**  
專業技能；  
聘僱費用



» 當今 AI 和分析的挑戰

» 與開源程式無縫整合，並以此作為成功因素

» 為什麼要使用 AI、分析和資料管理平台？

» 資料科學家將獲得哪些益處



# 與開源程式無縫整合， 並以此作為成功因素

資料科學家有一個充滿活力的生態系統可供選擇。

這可以跨越 SAS、Python 或 R 等語言、整合式開發環境、部署技術、虛擬機器、Kubernetes 等。

然而，這可能會造成選項疲勞，導致不一致的情況，從而難以擴展分析。越來越多的公司意識到，應用其他技術來整合開源程式和其他軟體 (例如 SAS® Viya® 等分析平台) 可能有助於從開源程式中建立協同工作的能力和共用程式。

期望和業務需求



- » 當今 AI 和分析的挑戰
- » 與開源程式無縫整合，並以此作為成功因素
- » 為什麼要使用 AI、分析和資料管理平台？
- » 資料科學家將獲得哪些益處



# 為什麼要使用人工智慧、 分析和資料管理平台？

SAS 資料科學產品營銷經理 Marinela Profi 在她的文章「[Six trends that will influence Data Science practitioners priorities in 2021](#)」(2021 年將影響資料科學從業者優先事項的六大趨勢) 中提到，「持續依賴開源程式並專注於其整合」是 2021 年的趨勢之一。

這些平台的一個關鍵成功因素是不限制資料科學家或 IT 開發人員可以使用的語言（包括開源程式）。他們還需要透過開放的 API 進行整合並確保無限的可擴展性。

組織依賴於開源程式 (如 Python 和 R)。並且它們也被證明是用於特定分析任務的寶貴工具。然而，當談到建立一個長期的分析戰略 - 一個自我維持且無縫整合的生命週期 - 開源程式在協調、整合方面會有相當大的挑戰。

越來越多的公司認識到，利用其他技術來整合開源程式和其他軟體 (例如 Viya 等分析平台) 可能有助於從開源程式建立協同工作的能力和共用程式。

無論您選擇哪種技術，您都必須致力於簡化此流程並實現自動化。您需要為自己的生態系統解決這個問題。如果您能這樣做，定能收穫巨大回報。

» 當今 AI 和分析的挑戰

» 與開源程式無縫整合，並以此作為成功因素

» 為什麼要使用 AI、分析和資料管理平台？

» 資料科學家將獲得哪些益處





# 資料科學家將獲得哪些益處

根據採訪和專家估計，資料科學家在獲得有用資料之前，所花費的時間中有 50% 到 80% 用到了收集和準備不同規則的數字資料這一平凡工作上。更重要的是，他們建立了模型後，部署可能是另一個難題。

平民資料科學家將能更快地將他們的工作帶入商業環境，得到組織的青睞並被認可為創新者。

## 將開源程式與 SAS 結合使用的好處：

- 利用 SAS Viya 的雲原生高效能架構
- 大規模的分散平行處理以實現無限可擴展性
- API 優先的開發策略結合容器化的微服務架構
- 自動化特徵工程與 ML 支援的資料準備
- 將 Python 或 R 直接與 SAS 結合使用，或使用 REST API 將 SAS 整合到應用程式中
- 執行開源模型無需重新編輯程式碼
- SAS 和開源模型的容器部署
- 在批次處理、串流、雲端或邊緣裝置中部署 SAS、R、Python 模型
- 在持續整合和持續部署的管道中自動化地管理程式碼工具
- 複合式 AI 在同一環境中結合不同的 AI 和分析技術 (如影像辨識和最佳化)
- 資料血緣網絡圖以及稽核，您可以查看資料如何在整個系統中移動
- 監控同一儲存庫中的開源程式和 SAS 模型並建立自訂的效能報告
- 商業規則和正式決策整合
- 從程式碼轉到拖拉點選，然後依需求可返回程式碼
- 透過支援自然語言的內建報告確保資料科學專案的可解釋性

» 當今 AI 和分析的挑戰 » 與開源程式無縫整合，並以此作為成功因素

» 為什麼要使用 AI、分析和資料管理平台？

» 資料科學家將獲得哪些益處



# 跨分析生命週期的整合



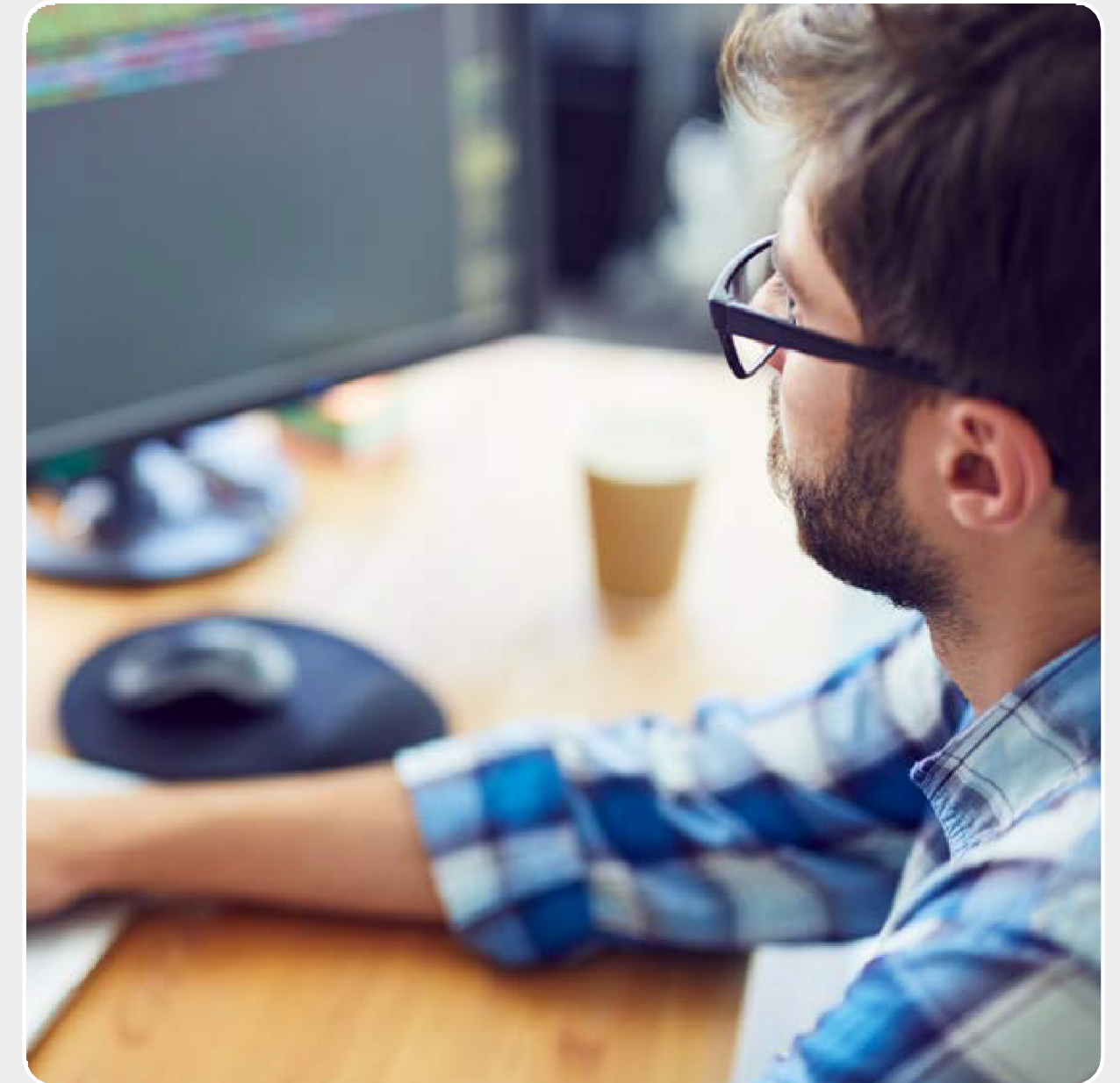


# 跨分析生命週期的整合

由於分析生命週期是一個廣泛的主題，因此在本手冊的其餘部分，我們將重點關注其中一個特定的面相：模型生命週期。

我們發現模型生命週期最有可能實現開源程式技術和 SAS 技術之間的交互和整合，因此將成為我們討論的重點。

模型生命週期是資料科學家和從業者用來建立、管理和部署分析模型以產生業務價值的方法論程序。



» 模型生命週期流程

» 整合建模

» 整合是如何發生的？

» 我如何在模型生命週期中進行整合？

是什麼推動了當今的分析創新？

跨分析生命週期的整合

模型建立

模型管理

模型部署

後續步驟



# 模型生命週期 處理流程

注意：建模是一個迭代過程。任何運行狀況良好的模型生命週期都需要持續的模型建立、測試和監控。



» 模型生命週期流程

» 整合建模

» 整合是如何發生的？

» 我如何在模型生命週期中進行整合？



# 整合建模

整合建模為用戶提供了更多功能和靈活性。但下一個問題是，如何實現整合？

為了回答這個問題，在瞭解整合的過程中涉及的各项技術之前，我們將先研究兩個重要領域。



透過 Viya 支援開放式整合：  
介紹和概述

- 1 整合是如何發生的？
- 2 我如何在模型生命週期中進行整合？

» 模型生命週期流程	» 整合建模	» 整合是如何發生的？	» 我如何在模型生命週期中進行整合？		
是什麼推動了當今的創新式分析？	跨分析生命週期的整合	模型建立	模型管理	模型部署	後續步驟



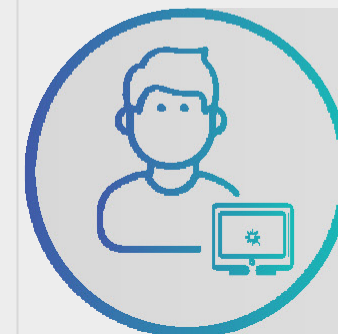
# 整合是如何發生的？

有兩種整合方法：從開源程式到 SAS 和從 SAS 到開源程式。提供兩個選項的好處是，可以在組織中向多種類型的使用者開放與擴展分析功能，無論其技術背景為何。



## 開源程式到 SAS

適用於具有開源程式背景知識、希望直接從開源程式介面探索 SAS 功能的使用者。



## SAS 到開源程式

適用於想要在 SAS 的視覺化介面中同時使用 SAS 和開源程式資產的使用者。



## 價值

無論採用何種方法，使用者都可以共用和協作。

» 模型生命週期流程

» 整合建模

» 整合是如何發生的？

» 我如何在模型生命週期中進行整合？



# 我如何在模型生命週期中進行整合？

透過 API，SAS 可以在模型生命週期的任何階段與開源程式技術整合。您的整合路徑取決於您的整合方式偏好，以及您希望在模型生命週期中的哪個階段進行整合。

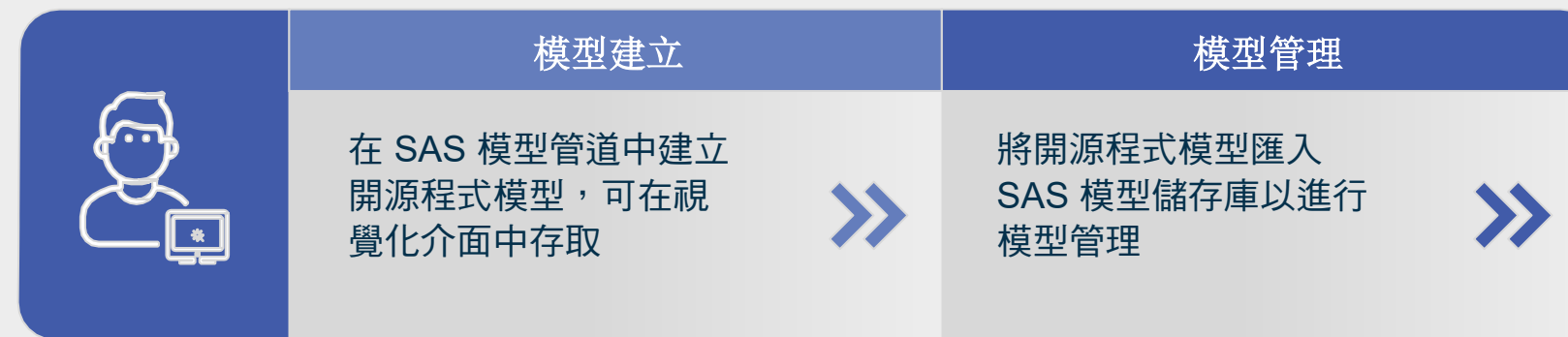
## 開源程式到 SAS

/ 從這裡開始！



## SAS 到開源程式

/ 從這裡開始！



### 模型部署

將模型發佈到一系列 SAS 和開源程式的執行引擎中，以進行批處理、單次使用或即時處理

» 模型生命週期流程	» 整合建模	» 整合是如何發生的？	» 我如何在模型生命週期中進行整合？		
是什麼推動了當今的創新式分析？	跨分析生命週期的整合	模型建立	模型管理	模型部署	後續步驟



# 模型建立



**模型建立 1：**  
透過 SWAT 和 DLPy 實現開源程式到 SAS



**模型建立 2：**  
透過 Model Studio 實現 SAS 到開源程式





# 模型管理

## 開源程式到 SAS



**模型管理 1：**  
透過 SAS Ctl/Pzmm 實現開源程式到 SAS

## SAS 到開源程式



**模型管理 2：**  
透過 Model Manager 實現 SAS 到開源程式

» 我的模型正在執行！接下來該做什麼？



# 我的模型正在執行！ 接下來該做什麼？

構建 Python 或 R 模型後，您可以將這些模型註冊到 SAS Model Manager，以便在將模型發佈到測試或生產環境之前比較、評估和監控模型的效能。

模型管理是模型建立後的重要步驟。使用 SAS Model Manager 的益處包括：

它確保您在任何給定的時間區間內始終執行最佳模型，以最大限度地減少業務的模型衰減影響。

可透過集中式模型儲存庫進行完整的可追溯性和分析管理，並讓模型工作流程版本控制進行更進階的管理。

只需按幾下，即可批量和即時的部署模型，從而更快地實踐業務價值。

*注意：要在 SAS 中管理模型，您需要在您的環境中授權 SAS 模型管理器。*

» 我的模型正在執行！接下來該做什麼？



# 模型部署



# 我的模型準備好了！ 接下來呢？

## 了解開源程式生態系統

現在您的模型可以運用了，是時候進行部署了。  
部署模型僅是意味著讓模型在生產環境中可以使用。

這通常由組織中的 DevOps 或 IT 團隊完成。

在這些影片中，我們將展示使用 SAS 工具的部署方法，結合使用 SAS 和開源程式技術將使生產環境中 SAS 和開源程式模型的部署更加穩健。

模型部署：  
SAS 雲端分析服務



模型部署：  
SAS 微分析服務



模型部署：  
Docker/Kubernetes



模型部署：  
SAS 串流處理



» 我的模型準備好了！  
接下來呢？



# 後續步驟

至此，您應該已經很好地了解了 SAS 與開源程式整合的原因、這樣做的好處以及您可以採用的各種方式。

但是，要進行成功的整合，您需要仔細思考在您的情境中是否需要整合，如果需要，需要整合到何種程度。

以下是一些常見的問題：

您資料的複雜性和/或資料量的增長情況如何？

您的團隊/組織中的技能組合是什麼？

您正在嘗試解決哪些分析問題？它們嚴重嗎？它們複雜嗎？它們急迫嗎？

您的 IT 環境如何運作？您會產生更多的技術債嗎？

如需更多資訊，請造訪 [sas.com/viya](https://sas.com/viya) 頁面。

著作權所有 © 2021, SAS Institute Inc. 保留一切權利。112252\_G182515.0821