

基礎設施事件準備

AWS 指導方針與最佳實務

2017 年 7 月



©2017，Amazon Web Services, Inc. 或其附屬公司。保留所有權利。

聲明

本文件僅供參考。文件內容為 AWS 在發佈當日最新的產品內容及實務，如有變更，恕不另行通知。客戶需自行獨立評估本文件資訊，任何 AWS 產品或服務皆以「現狀」提供，不包含任何明示或暗示之保證。本文不代表 AWS、其附屬公司、供應商或授權人提供之任何保證、表示、契約承諾、條件或擔保。AWS 對其客戶的責任與義務應由 AWS 協議管轄，本文並非 AWS 與其客戶之間的任何協議的一部分，也並非上述協議的修改。

內容

簡介	1
基礎設施事件準備規劃	1
什麼是規劃的基礎設施事件？	1
在規劃的基礎設施事件期間會發生什麼事？	2
設計原則	3
離散式工作負載	3
自動化	6
多樣性/彈性	7
成本優化	10
事件管理流程	11
基礎設施事件排程	11
規劃和準備	11
營運準備 (事件天數)	18
後續事件活動	20
結論	22
作者群	22
深入閱讀	22
附錄	23
詳細的架構審閱檢查清單	23

摘要

本白皮書為想要設計和佈建雲端應用程式，以處理規劃的擴展活動（如：產品發佈或季節性緩慢與動態流量峰值）的客戶說明部署在 **Amazon Web Services (AWS)** 上的工作負載，其指導方針與最佳實務。我們說明一般設計原則，並提供特定的最佳實務和基礎設施事件規劃的跨多概念指導。接著，我們會將說明營運準備考量和實務以及事後活動。

簡介

基礎設施事件準備的內容是與對會影響您業務的預期和重大事件做出設計與準備。這些事件是確保在所有情況和流量模式的變化下，公司的 **Web** 服務仍能維持可靠、可回應和高度容錯的關鍵。這類事件可以包括新領域的擴展、新產品或功能的啟動、季節性事件，或重大業務公告或行銷事件。

未正確規劃的基礎設施事件，會對貴公司的商業信譽、持續性或財務產生負面影響。基礎設施事件失敗的形式可能為：意外服務故障、負載相關效能緩降、網路延遲、儲存容量限制、系統限制 (如：**API** 呼叫率)、有限的可用 **IP** 地址數量、因監測不足而導致對應用程式堆疊元件的行為了解不足、第三方服務或元件未設定用於擴展的意外依賴性，或其他不可預見的錯誤狀況。

若要降低重要事件期間的意外故障風險，公司應該在規劃和準備、訓練員工以及設計和歸檔相關程序上投資時間和資源。為特定的雲端型應用，或一組應用程式的基礎設施事件規劃所做的投資量，是以系統的複雜性和全球觸角程度所決定。無論公司雲端服務的範圍或呈現的複雜性如何，在本白皮書中所提供的設計原則和最佳實務指導皆相同。

使用 **Amazon Web Services (AWS)**，貴公司可以動態、靈活的按用量付費方式，為規劃好的擴展事件作出準備，以擴展其基礎設施。**Amazon** 提供豐富、彈性的各項選擇，以及可程式化產品與服務，讓貴公司能獲得一個具備高度安全性、可靠的高速基礎設施；**Amazon** 使用該基礎設施來執行自身的全球網路，並讓貴公司靈活調整以回應自身快速變化的業務需求。

本白皮書概述最佳實務和設計原則，可協助您進行基礎設施事件的規劃和執行，並說明如何使用 **AWS** 服務，以確保您的應用程式能因應您的業務需求，向上並向外擴展。

基礎設施事件準備規劃

本章節說明規劃的基礎設施事件的組成要素，和通常會在這類事件中發生的活動類型。

什麼是規劃的基礎設施事件？

規劃的基礎設施事件為業務導向、預期與排定的事件時段，在這個時段中業務關鍵在於維持高回應、高擴展和可容錯的 **Web** 服務。可透過市場行銷活動、與公

司業務相關的新聞事件、產品發佈、業務地區擴張或任何可為公司的 Web 型應用程式與其下基礎設施帶來額外流量的類似活動來推動此要求。

在規劃的基礎設施事件期間會發生什麼事？

在大多數規劃的基礎設施事件中，其主要的考量在於能夠為您的 Web 基礎設施增加容量，以滿足更高的流量需求。在具備實體運算、儲存和聯網資源的傳統現場部署環境中，公司的 IT 部門必須根據他們對於理論最大峰值的最佳估計，佈建額外的容量。這導致了供應能力不足的風險，並且由於 Web 伺服器過載、回應時間慢以及其他執行時間錯誤，公司的業務就會遭受損失。

在 AWS 雲端中，基礎設施是可程式化和彈性的。這表示它能因應即時需求快速佈建。也表示它能夠以自動化、智慧和動態方式配置以回應系統指標 - 在必要時增加或縮小資源 (例如 Web 伺服器叢集、佈建的輸送量、儲存容量、可用的運算核心、串流所需的分片數量等)。

此外，許多 AWS 服務是完全受管的。這些服務包括儲存、資料庫、分析、應用程式和部署服務。這表示，AWS 客戶不需要擔心針對高流量事件配置這些服務會很複雜。AWS 全受管服務，是專為可擴展性和高可用性設計。

一般而言，在準備規劃的基礎設施事件時，AWS 客戶會執行系統審查，以評估他們的應用程式架構和操作準備，主要考量可擴展性和容錯能力這兩項。考慮流量估算值並與正常業務活動績效進行比較，以及確定容量指標與所需額外容量的估算值。識別並處理任何潛在的瓶頸和第三方上游和下游依賴性。如果已規劃的事件包含業務區域的擴展或新交易對象，則也應考量地理區域。擴展至其他 AWS 區域或可用性區域需在已規劃事件前執行。也會檢視客戶的 AWS 動態系統設定，例如 Auto Scaling、負載平衡、地理式路由、高可用性和容錯移轉措施，以確保這些功能已配置，可正確地處理預期的磁碟區和交易率增加。如 AWS 資源限制和內容交付網路 (CDN) 原始伺服器位置這類靜態設定，也會視需要加以考量和修改。

也將審查並適時強化監控和通知機制，以便在發生事件時以及規劃事件完成的後續分析中提供即時透明度。

在已規劃事件作業期間，AWS 客戶也可以在發生任何需要故障排除或即時支援時 (例如，伺服器發生故障時) 開啟 AWS 的支援案例。訂閱 AWS 企業支援計劃的客戶擁有立即與支援工程師交談的彈性，並且若需要迅速回應，則可提高為關鍵嚴重案例。

在事件之後，AWS 資源的設計可自動縮減至符合流量等級的適當等級，或隨著事件要求持續擴展。

設計原則

在開始運用任何雲端應用程式堆疊或工作負載時，規劃事件的準備要從良好的設計開始。

離散式工作負載

一個良好的設計是規劃事件工作負載在一般和逐漸遞升流量等級兩種情況下有效管理的根本。請確定在設計離散式與獨立功能的資源群組在一開始即以特定商業應用程式或產品為中心。本章節說明設計目標的各個面向。

標籤

標籤是用於標記和組織資源。在規劃的基礎設施事件期間，它們是管理基礎設施資源的必要元件。在 AWS，標籤是客戶管理的鍵值標籤，應用於個別的受管資源 (例如：負載平衡器或 Amazon Elastic Compute Cloud (EC2) 執行個體)。透過參照已附加於 AWS 資源的良好定義標籤，您可以輕鬆地識別整體基礎設施中哪一個資源包含您的規劃事件工作負載。然後，使用此資訊就可以做好分析工作，做好準備。標籤也可以用來分配成本。

您可以使用標籤來組織如 EC2 執行個體、Amazon Machine Image (AMI) 影像、負載平衡器、安全群組、Amazon Relational Database Service (RDS) 資源、Amazon Virtual Private Cloud (VPC) 資源、Amazon Route 53 運作狀態檢查和 Amazon Simple Storage Service (S3) 儲存區。

有關有效標籤策略的詳細資訊，請參閱 [AWS 標籤策略](#)。¹

如需範例說明如何建立和管理標籤，並將其放置在資源群組的詳細資訊，請參閱 [AWS 資源群組和標籤](#)。²

鬆耦合

規劃雲端架構時，您應該設計應用程式堆疊的每個元件，好讓各個元件能盡可能地獨立作業。這讓雲端工作負載具有彈性和可擴展性這兩項優勢。

您可以藉著輸入和輸出 (例如：RESTful API) 用的良好定義界面，將每個元件設計為黑色方塊，以降低在雲端應用程式堆疊元件之間的依賴性。如果元件非為應

用程式，而是包含應用程式的服務，則稱為*微型服務架構*。對於應用程式元件之間的通訊與合作，您可以使用事件導向的通知機制，如元件間的 AWS 的訊息佇列，如圖 1 所示。

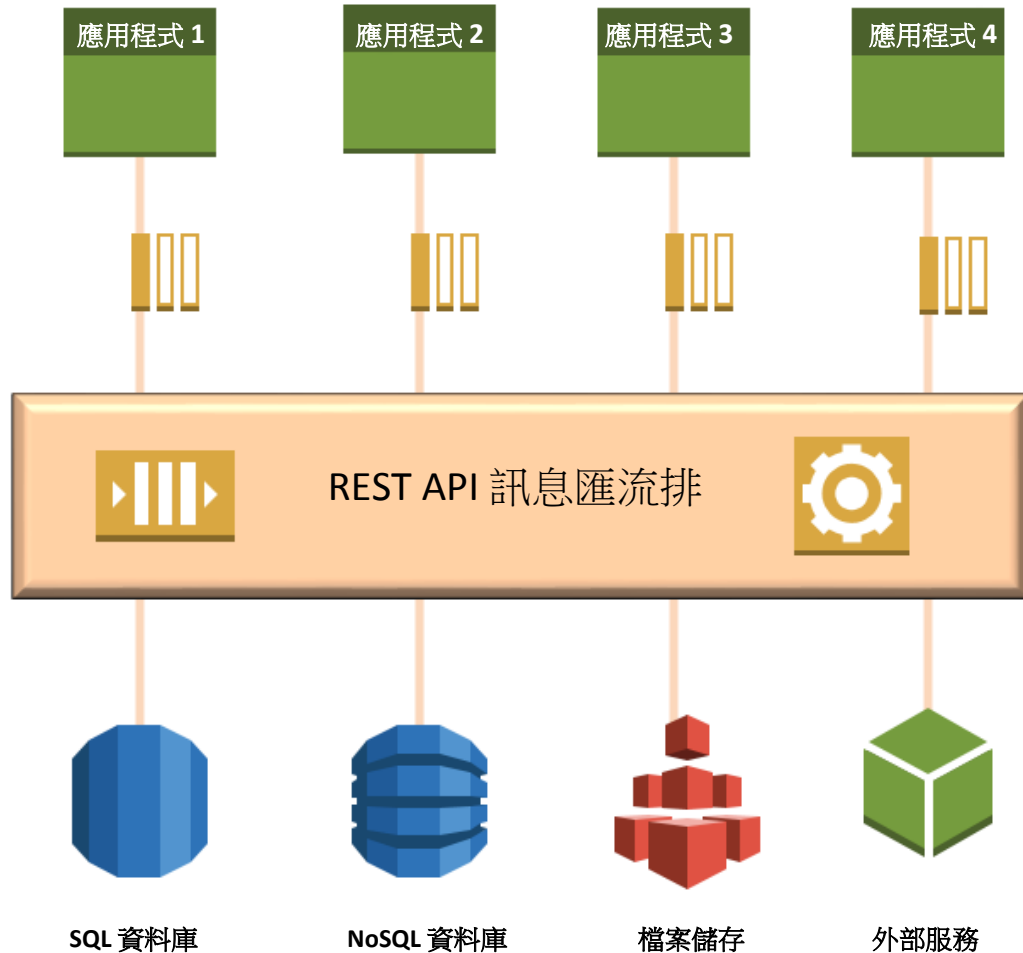


圖 1: 使用 RESTful 界面和訊息佇列的鬆耦合

使用如上列機制，則一個元件變更或故障時，導致其他元件發生同樣狀況的機會將降低很多。例如，如果多層應用程式堆疊中的伺服器無法回應時，可設計鬆散耦合的應用程式略過未回應的層級或切換到降級模式替代交易。

使用中繼訊息佇列的鬆散耦合應用程式元件，也可以更輕鬆地專為非同步整合做出設計。由於應用程式的元件並不是採用直接點對點通訊，但是使用中繼和持久性簡訊層（例如：Amazon Simple Queue Service (SQS) 佇列或如 Amazon Kinesis Streams 這類串流數據機制），因此它們能應付當下游元件處理傳入佇列

時，一個元件內的突增活動。或若出現元件故障，訊息會保留在佇列或串流中，直到故障的元件修復為止。

有關 AWS 提供的訊息佇列和通知服務的詳細資訊，請參閱 [Amazon Simple Queue Service](#)。³

服務，而非伺服器

受管服務和服務終端讓您無需擔心安全性或存取、備份或恢復、修補程式管理或變更控制、設定監控或報告，也無需管理許多傳統系統管理的細節。這些雲端資源可為高可用性與彈性使用多可用性區域 (在某些案例中，多地區) 配置預先佈建。它們通常可以在不需停機的情況下擴展或縮減，而且您可以透過 AWS 管理主控台或 API/CLI 呼叫來即時配置它們。

受管服務和服務終端可使用如：關聯式與 NoSQL 資料庫系統、資料倉庫、事件通知、物件和檔案儲存、即時串流、大數據分析、機器學習、搜尋、轉碼和其他許多能力，以強化客戶的應用程式堆疊。一個終端就是一個 URL，也是 AWS 服務的入口。例如：<https://dynamodb.us-west-2.amazonaws.com> 是 Amazon DynamoDB 服務的入口。

使用受管服務及其服務終端，您就能利用生產就緒型資源的能力，作為在規劃的基礎事件期間所增加的磁碟區、覆蓋區以及交易費率設計解決方案的一部分。您不需要佈建並管理與受管服務相同功能的自身伺服器。

有關 AWS 服務終端的詳細資訊，請參閱 [AWS 區域和終端](#)。⁴另請參閱 [Amazon EMR](#)⁵、[Amazon RDS](#)⁶ 和 [Amazon ECS](#)⁷ 這些擁有終端受管服務的範例。

無伺服器架構

另一個能有效處理在規劃的基礎設施事件期間，對動態變動處理負載的回應需求的策略為利用 AWS Lambda。Lambda 是一種事件導向、無伺服器的運算平台。它是一種為回應事件 (通過通知) 執行 Python、Node.js 或 Java 程式碼動態調用的服務，並自動管理該程碼指定的運算資源。Lambda 不需要預先佈建 Amazon EC2 運算資源。Amazon Simple Notification Service (Amazon SNS) 可設定為觸發 Lambda 函數。有關 Amazon SNS 的詳細資訊，請參閱 [Amazon 推送通知服務](#)。⁸

Lambda 無伺服器函數，可以執程式碼以存取或調用其他 AWS 服務，例如：資料庫操作、資料轉換、物件或檔案擷取，或甚至是擴展操作以回應外部事件或

內部系統負載指標。AWS Lambda 也可以自行產生新的通知或事件，甚至啟動其他 Lambda 函數。

AWS Lambda 可在規劃的基礎設施事件期間，提供精細控制擴展操作的能力。例如，Lambda 可用於擴展 Auto Scaling 操作的功能來執行動作，例如通知第三方系統它們也需要擴展，或新增額外的網路界面，以連接至佈建它們的新執行個體上。請參閱[以 Auto Scaling 生命週期掛鉤使用 AWS Lambda](#)⁹，以查看有關如何使用 Lambda 自訂擴展操作的範例。

有關 AWS Lambda 的詳細資訊，請參閱[AWS Lambda 是什麼？](#)¹⁰

自動化

Auto Scaling

基礎設施事件規劃的關鍵元件為 Auto Scaling。能夠根據預先定義的條件自動擴展或縮小應用程式的容量，以協助規劃的基礎設施事件中所發生的流量模式和磁碟區波動維持應用程式的可用性。

AWS 在多個資源提供 Auto Scaling 功能，包括 EC2 執行個體、資料庫容量、容器等。

Auto Scaling 可用來擴展執行個體的分組，例如包含以雲端為基礎應用程式的伺服器組，好讓它們能依據特定的標準自動擴展。Auto Scaling 也可用於維護固定數量的執行個體，就算執行個體運作不良也一樣。自動擴展和維持執行個數目是 Auto Scaling 服務的核心功能。

Auto Scaling 會在群組中的執行個體執行週期性的運作狀態檢查，以維持其數量。如果執行個體運作狀態不佳，則群組會終止這個執行個體，並啟動另一個執行個體來取代它。

Auto Scaling 政策可用來自動增加或減少運轉中 EC2 執行個體在一組伺服器的數目，以滿足不斷變動的狀況。當擴展政策開始生效，Auto Scaling 群組會調整所需的容量，並在已知並可預測的流量起伏時，視需要或動態或排定時間啟動或終止執行個體。

重新啟動和復原

在任何規劃基礎設施事件中，最重要的設計要素為擁有程序和自動化，以處理組成的執行個體或伺服器並能夠即時復原或重新啟動它們。

當基礎硬體的系统狀態檢查不通過時，EC2 執行個體可設定為自動復原。執行個體將重新啟動（若必要則在新的硬體上），但會保留執行個體 ID、IP 地址、彈性 IP 地址、Amazon Elastic Block Store (EBS) 磁碟區附件和其他設定詳細資訊。有關自動復原 EC2 執行個體的詳細資訊，請參閱 [Amazon EC2 自動復原](#)。¹¹

組態管理/協調流程

強大、可靠且回應力佳的規劃基礎設施事件策略，是對個別資源狀態管理和應用程式堆疊部署進行組態管理和協調工具的整合。

組態管理工具通常處理伺服器執行個體的佈建和組態、負載平衡器、Auto Scaling、個別應用程式部署和應用程式運作狀態監控。它們也提供與其他服務整合的能力，例如資料庫、儲存磁碟區和快取層。

協調工具、組態管理上層抽象層提供指定這些不同資源關係的方法，讓客戶能夠如統一的雲端應用程式基礎設施一般佈建和管理多個資源，而無需擔心資源的依賴性。

由於這些工具如同程式碼一般定義和說明個別的資源以及它們的關係，這個程式碼可由版本控制、促進能夠復原到之前版本或試用新程式碼分支的能力，以作為測試和開發之用。還可以定義為基礎設施事件優化的協調和配置，然後在此類事件之後復原為標準配置。

Amazon Web Services 建議使用以下工具，以達到硬體即程式碼的部署和協調：

- **AWS Config with Config Rules** 或 AWS Config 合作夥伴，以提供詳細的、視覺化的和可搜尋的 AWS 資源、組態歷史記錄和資源組態合規性的清單。
- **AWS CloudFormation** 或第三方 AWS 資源協調工具，以管理 AWS 資源佈建、更新和終止。
- **AWS OpsWorks、Elastic Beanstalk**，或第三方伺服器組態管理工具，以管理作業系統 (OS) 和應用程式組態的變更。

有關如管理程式碼般管理硬體，請參閱[基礎設施組態管理](#)。¹²

多樣性/彈性

移除單一故障點和瓶頸

在規劃基礎設施事件時，您應該就任何單點故障 (SPOF) 或效能瓶頸來分析應用程式堆疊。是否有任何伺服器的單一執行個體、資料磁碟區、資料庫、NAT 閘道或如負載平衡器，可能導致整個應用程式或其中大部分在故障時停止運作？

其次，隨著雲端應用程式在流量上或交易磁碟區的擴展，會不會有一部分基礎設施受到實體限制或管束，像是隨著數據流路徑成長的資料磁碟區網路頻寬或 CPU 處理週期？

一旦發現，這些風險可以通過各種方式得到緩解。

防故障設計

如上所述，使用鬆耦合和訊息佇列的 RESTful 界面是因應個體資源故障或流量或交易磁碟區波動的一種不錯的靈活性策略。靈活設計的另一面向為盡可能配置應用程式元件為無狀態。

無狀態應用程式在交易之前不需要知識，且對其他應用程式元件的依賴性相當鬆散。它們不儲存任何工作階段資訊。無狀態應用程式可作為集區或叢集成員水平擴展，因為集區或叢集內的任何請求皆可由任何執行個體處理。您可以使用 **Auto Scaling** 和運作狀態檢查標準，視需要新增更多資源，以程式設計方式處理波動運算、容量和輸送量需求。一旦應用程式設定為無狀態，它可能會重新架構為無伺服器架構，使用 **Lambda** 函數取代 **EC2** 執行個體。**Lambda** 函數也有內建的動態擴展功能。

在諸如 **Web** 伺服器之類的應用程式資源無法避免有關交易的狀態數據的情況下，您應該考慮設計應用程式，好讓有狀態的應用程式部分與伺服器本身分離。例如，**HTTP Cookie** 或同等狀態的數據可存放在資料庫 (如：**DynamoDB** 或 **S3** 儲存貯體或 **EBS** 磁碟區)。

如果您有一個複雜的多步驟工作流程，需要跟踪工作流程中每個步驟的當前狀態，則可以使用 **Amazon Simple Workflow Service (SWF)** 集中儲存執行歷史記錄，並使這些工作負載成為無狀態。

另一個彈性方法則是採用分散式處理。對於單個計算資源無法滿足及時處理大量數據需求的情況下，您可以設計工作負載，使任務和數據分割成更小的分段，並在一組運算資源並行執行。分散式處理是無狀態的，因為處理中的分區數據和任務的獨立節點可能會失敗。在這種情況下，在另一個分散式處理叢集上的節點自動重新啟動失敗的任務，將會由分散式處理排程引擎自動處理。

AWS 提供各種分散式資料處理引擎 (如 Amazon EMR、Amazon Athena 和 Amazon Machine Learning)，每種引擎都是受管的服務，提供終端並讓您無需面對包含修補、維修、擴展、容錯移轉等種種複雜事宜。

對於即時串流資料的處理，Amazon Kinesis Streams 可以將資料分為多個分片，由多個該資料的消費者處理，例如 Lambda 函數或 EC2 執行個體。

有關這些工作負載類型的詳細資訊，請參閱 [AWS 上的大數據分析選項](#)。¹³

多個可用區域和區域

AWS 服務位於全球多個地點。這些位置由區域和可用區域構成。「區域」是單獨的地理區域。每個區域都有多個單獨的據點，其稱為可用區域。AWS 讓客戶能夠將資源 (如：執行個體和資料) 存放在多個位置。

您應該設計您的應用程式，以讓它們可以分散到多個可用區域和區域。結合在可用區域和區域之間的分配和複製資源，您應該使用負載平衡和容錯移轉機制來設計應用程式，以便在發生故障時，應用程式堆疊會自動將數據流和流量重新導向這些替代位置。

負載平衡

使用 Elastic Load Balancing 服務 (ELB)、應用程式伺服器組就可附加到負載平衡器上，且可跨越多個可用區域。當位於負載平衡器後面的特定可用區域中的 EC2 執行個體運作狀態檢查失敗時，負載平衡器會停止向這些節點發送流量。在結合 Auto Scaling 使用時，正常運作的節點會自動與其他可用區域重新平衡，無需手動干預。

您還可以使用 Amazon Route 53 以及以延遲為基礎的 DNS 路由演算法，跨區域使負載平衡。如需詳細資訊，請參閱[延遲為本的路由](#)。¹⁴

卸載策略

雲端基礎設施中的卸載概念包含重新導向或代理其他位置的伺服器流量，以減輕主要系統的壓力。在某些情況下，卸載策略可觸發活動，您可以選擇卸下特定串流的流量或降低您應用程式的功能，以減輕處理負載並能夠提供至少傳入請求的子集。

卸載可使用許多技術。以延遲為基礎的 DNS 路由就是其中一種方法。另一種方法是使用快取。使用內存快取層 (如 Amazon ElastiCache) 可以在應用程式附近

進行快取。或者您可以使用使用者節點附近的快取層，使用如 **Amazon CloudFront** 這類全球內容分發網路。

有關 **ElastiCache** 和 **CloudFront** 的詳細資訊，請參閱 [ElastiCache](#)¹⁵ 和 [Amazon CloudFront CDN 入門](#)。¹⁶

成本優化

預留 vs 競價型 vs 隨需執行個體

根據系統指標和其他性能和健康檢查標準，與雲端動態佈建資源的能力緊密相連，是控制雲端資源佈建成本的能力。通過 **Auto Scaling**，資源利用可以與實際的處理和儲存需求緊密配合，最大限度地減少浪費的費用和未充分利用的資源。

雲端成本管控的另一個面向是能夠從隨需執行個體、預留執行個體 (RIs) 或競價型執行個體中擇一選用。您還可以為 **DynamoDB** 保留容量能力。

使用隨需執行個體，您只需支付所使用的 **EC2** 執行個體。隨需執行個體讓您無須長期投入只要支付運算容量即可。

Amazon EC2 預留執行個體提供相較於隨需執行個體定價更多的折扣 (最多 75%)，並在用於特定可用區域中時，提供容量保留。但是，除了保留的可用性和折扣，在預留執行個體和隨需執行個體間的功能並沒有差異。

競價型執行個體可讓您對備用的 **Amazon EC2** 運算容量進行競價。競價型執行個體通常會以低於隨需執行個體定價的折扣價提供，這可大幅降低以雲端為基礎的應用程式的成本。

在設計適用於雲端時，一些使用案例是更適合使用競價型執行個體。例如，由於競價型執行個體的競價價格超出您的出價時可隨時停止使用，因此您應該考慮僅在相對較無狀態和水平擴展的應用程式堆疊上，執行競價型執行個體。對於狀態應用程式或昂貴的處理負載，預留執行個體或隨需執行個體可能會是更好的選擇。對於關鍵任務應用程式的容量限制更不用說，預留執行個體就是最佳選擇。

如需詳細資訊，請參閱[預留執行個體](#)¹⁷ 和[競價型執行個體](#)¹⁸。

事件管理流程

基礎設施事件的規劃是包含應用程式開發人員、管理員和企業股東的團體活動。在基礎設施事件前幾週，您應該建立一個包括擁有和操作 Web 服務關鍵基礎設施元件技術人員的定期會議。

基礎設施事件排程

基礎設施事件的規劃應在事件開始前數週開始。規劃的事件生命週期的典型時間表如圖 2 所示。

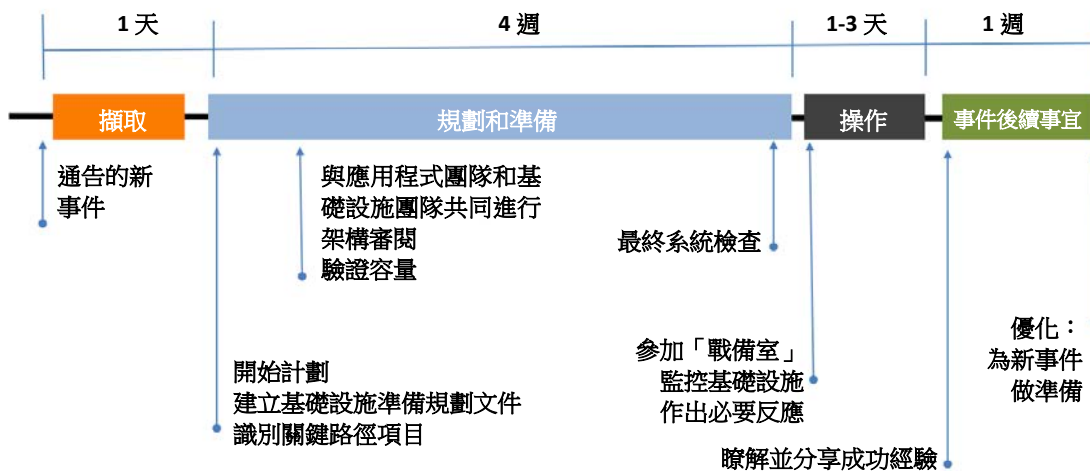


圖 2：典型的基礎設施事件時間表

規劃和準備

排程

在基礎設施事件前數週，我們建議以下活動排程：

第 1 週：

- 將團隊命名以推動基礎設施事件的規劃和工程。
- 在利害關係人之間舉辦會議以了解事件參數（擴展、持續時間、時間、地理覆蓋區域、受影響的工作負載）和成功條件。
- 接觸任何下游或上游合作夥伴和供應商。

第 2 至 3 週：

- 審閱架構，並視需要進行調整。
- 進行營運檢閱；視需要進行調整。
- 依照本文件與最末頁參照中所敘述的最佳實務進行。
- 識別風險和開發緩解計劃。
- 開發規劃事件執行手冊。

第 4 週：

- 查看所有需要依據預期負載擴展的雲端供應商服務。
- 視需要檢查服務限制並提高限制。
- 設定監控儀表板和定義發出警報的閾值。

架構審閱

您為基礎設施事件準備的重要部分是會經驗流量暴漲的應用程式堆疊架構審查。審查的目的為驗證並找出潛在風險領域，不論是應用程式的擴展性或可靠性，並在事件發生前找出優化的機會。

AWS 為企業支援客戶提供框架，以檢視客戶在 5 個設計支柱四周的應用程式堆疊。設計的 5 個支柱為：安全性、可靠性、效能效率、成本優化和卓越營運，如下所述。

表 1: 完整架構的應用程式支柱

支柱名稱	支柱定義	相關利益領域
安全性	能夠保護資訊、系統和資產，同時透過風險評估和緩解策略交付企業價值。	身分管理、加密、監控、記錄、金鑰管理、專用執行個體、合規、管控
可靠性	系統從基礎設施或服務故障中恢復的能力，動態獲取計算資源以滿足需求，並減輕諸如錯誤配置或瞬時網路問題之類的中斷。	服務限制、多個可用區域和區域、可擴展性、運作狀態檢查/監控、備份/DR Networking、自動化自我修復
表現效率	有效利用計算資源來滿足系統要求，以及隨著需求變化和技術的發展而保持效率。	正確的 AWS 服務、資源使用、儲存架構、快取、延遲要求

支柱名稱	支柱定義	相關利益領域
成本優化	能夠避免或消除不需要的成本或次優的資源。	競價型執行個體/預留執行個體、環境調校、服務選項、磁碟區調校、客戶管理、合併帳單，解除委任資源
卓越營運	執行和監控系統以提供業務價值並持續改進支持流程和程序的能力。	Runbooks、Playbooks、RCAs CI/CD、Game Days、基礎設施即程式碼、RCAs

本白皮書的附錄中提供可用於查看基於 AWS 的應用程式堆疊的架構審查項目詳細清單。

操作檢閱

除了更重視應用程式的設計元件的架構檢閱之外，您還應該審查您的雲端操作與管理實務，以評估如何處理雲端工作負載的管理。檢閱的目的是為了識別營運漏洞和問題並在事件前採取動作以最小化問題。

AWS 提供給企叢支援客戶的雲端營運檢閱是適用於基礎設施事件準備的寶貴工具。檢閱著重於評估以下範圍：

- 準備 – 您必須擁有正確的組織結構、程序和技術。您應該對管理您應用程式堆疊的員工明確地定義其角色和責任。程序應事先定義以符合事件。程序應該盡可能地自動化。
- 監控 – 有效監控應用程式執行的方式。監控非常重要，可讓異常變成問題前偵測到，並提供機會，將負面事件所帶來的影響降至最低。
- 營運 – 營運活動需要盡可能地利用自動化即時且可靠的完成，同時還需要處理需要向上呈報的意外營運事件。
- 優化 – 使用集合的指標、營運趨勢和心得改善以擷取和報告，進行事件後續分析，以在未來事件發生期間擁有改進的機會。優化加上準備建立起回饋迴圈，以解決操作問題並避免它們再度發生。

了解 AWS 服務限制

在規劃基礎設施事件期間，避免超過服務限制十分重要，可由雲端供應商在擴展應用程式或工作負載時強制實施。

雲端服務供應商通常會在您可以使用的不同資源上有限制。這些通常是根據每個帳戶和每個區域為準強制實施。受影響的資源包含執行個體、磁碟區、串流、無伺服器呼叫、快照、VPC 數量、安全規則等。它們旨在作為一種安全措施，防止失竊的程式碼或流氓行為者試圖濫用資源，並作為控制來幫助最小化計費風險。

當您在雲端擴展足跡時，一些服務限制會隨著時間的推移而自動增加，儘管大部分服務需要通過打開支持案例來請求提升限制。雖然一些服務限制可透過支援案例增加，其他服務的限制卻無法變更。

AWS 為企業和業務部門的客戶提供了 **Trusted Advisor**，它提供了一個檢查限制的儀表板，使客戶能夠主動管理所有的服務限制。

有關各種 AWS 服務以及如何查看這些資料的詳細資訊，請參閱 [AWS 服務限制](#)¹⁹ 和 [Trusted Advisor](#)。²⁰

了解模式

基準

在基礎設施事件開始之前，您應該記錄關鍵指標的「回復運作狀態良好」值。這有助於您確定在事件完成/結束後應用程式/服務何時可安全地返回到正常級別。例如，通過負載平衡器識別正常的事務速率是每秒 2,500 個請求，這有助於確定事件後何時開始逐步關閉過程是安全的。

資料流和依賴性

了解數據流通過應用程式的各個元件能如何幫助您識別潛在的瓶頸和依賴性。應用程式層或元件在數據流中作為數據的客戶是否能正確調整大小，並且，若應用程式堆疊或元件是否可正確自動擴展？若元件發生故障，數據是否可排序到元件復原為止？任何下游或上游資料供應商或消費者是否可擴展以回應您的事件？

相稱

另一個檢閱基礎設施事件預備的考量為不同應用程式堆疊時所需要的相稱擴展。相稱不一定為一對一。例如，負載平衡器每秒的交易增加十倍可能需要儲存容量或串流分片數量或資料庫讀取和寫入操作數量增加二十倍，因為其計算處理可能在前向應用程式中進行。

通訊計劃

在進行事件前，您應該開發通訊計畫。收集內部利益攸關方和支持小組的名單，並確定在各種場景中的各個階段應該聯繫的人員，例如事件開始、事件期間、事件結束、事件後分析、在故障排除情況時的緊急聯繫人、聯繫人等。

要聯絡的人士與團體包含如下：

- 利害關係人
- 營運經理

- 開發人員
- 支援團隊
- 雲端服務供應商團隊
- 網路營運中心 (NOC) 團隊

當您收集內部聯繫人清單時，您還應該開發涉及應用程式持續實時提交的外部利益相關者的聯繫人清單。這些利益相關者包括支援堆疊關鍵元件的合作夥伴和供應商、提供外部服務、數據饋送、認證服務的下游和上游供應商等。

這個外部聯絡清單應該也包含下列項目：

- 基礎設施託管供應商
- 電訊廠商
- 即時資料串流合作夥伴
- 公關行銷聯絡人
- 廣告合作夥伴
- 涉及服務工程的技術顧問

詢問從每個供應商的下列資訊：

- 在事件發生期間的即時連絡窗口
- 關鍵支援聯絡人和緩步升級流程
- 名稱、電話號碼和電子郵件位址
- 驗證即時技術聯絡人為可用

訂閱企業支持的 AWS 客戶還擁有分配給其帳戶的技術客戶經理 (TAM)，可以協調和驗證並了解事件且預備好支援的專業 AWS 支援人員。TAM 會在事件期間隨時預備上陣、出席戰備室且在需要時也可推動支援往上呈報。

NOC 準備

在事件，您應該指示您的營運和/或開發人員團隊來建立即時指標儀表板，監控在事件發生時 Web 服務在生產時每個關鍵元件的情況。理想情況下，儀表板應該每分鐘或在事件期間任何適當和有效的間隔時間，自動顯示更新的指標。

考慮監控以下元件：

- 每個伺服器的資源使用 (CPU、磁碟和記憶體的使用)
- Web 服務回應時間
- Web 流量指標 (使用者、頁面檢視、工作階段)
- 每個訪客區域 (全球客戶工作階段) 的 Web 流量
- 資料庫伺服器使用
- 行銷流程轉換渠道，例如轉換率和轉換與失敗率
- 應用程式錯誤日誌
- Canary 監控

Amazon CloudWatch 提供了一種方法，可以使用 CloudWatch 自訂儀表板將 AWS 資源中的大部分指標匯總到單一窗格中。此外，CloudWatch 還提供了將自訂指標匯入 CloudWatch 的功能，無論 AWS 是否尚未自動提供該指標皆同。有關 AWS 監視工具和功能的詳細資訊，請參閱本文的監控部分。

執行手冊的準備

您應該為基礎設施事件開發執行手冊。*執行手冊*是一本操作性的手冊，操作員在事件期間將執行的程序和操作的彙編。事件執行手冊可能由現有用於常態操作和異常狀況處理的執行手冊發展而成。一般而言，執行手冊包含程序的開始、停止、督導和偵錯系統。它還應說明處理異常事件和意外情況的程序。

執行手冊應該包含以下部分：

- **事件詳細資訊：**簡短描述該事件的成功標準、媒體涵蓋範圍、事件日期以及客戶端和 AWS 的主要利害關係人聯絡詳細資訊。
- **AWS 服務清單：**一一枚舉在事件期間使用的所有 AWS 服務。此外，這些服務、影響區域和帳戶 ID 的預期負載。
- **架構和應用程式審查：**記錄負載測試結果、任何基礎設施和應用設計中的壓力點、工作負載的彈性測量、單點故障和潛在瓶頸。
- **操作審查：**突出監控設置、正常運作標準、通知機制和服務恢復流程。
- **準備檢查清單：**包括諸如服務限制檢查、預加載應用程式堆疊元件 (如負載平衡器)、預配置資源 (如串流分片、DynamoDB 分區、S3 分區等) 的注意事項。如需詳細資訊，請參閱本白皮書附錄中的「詳細的架構審閱檢查清單」。

監控

監控計劃

資料庫、應用程式和作業系統監控對確保成功事件而言十分重要。應設定全方位的監控系統，讓您可以立即有效偵測和回應基礎設施事件期間的嚴重事件。整體而言，有效的監控策略可確保監控工具會根據應用程式的業務重要性，配備適當的層級。有效的事件管理策略將結合 AWS 和客戶監控數據與其事件管理工具和流程。實施監控計劃，共同從所有 AWS 解決方案部門收集監控數據，這將有助於在發生複雜的故障時進行廣範圍的偵錯。

監控計劃應解決下列問題：

- 必須為事件設定哪些監控工具和儀表板？
- 監控目標和允許的閾值是什麼？哪些事件會觸發動作？
- 哪些資源和來自資源的指標會被監控，及其輪詢頻率？
- 誰將執行監控任務？就位的監控提醒為何？誰將會收到提醒？
- 為普遍和預期的故障設定了何種補救計劃？異常事件如何處理？
- 發生任何故障時的向上呈報流程？

以下 AWS 監控工具可作為此策略的一部分使用：

- **Amazon CloudWatch**：對 AWS 儀表板指標、監控、提醒和自動佈建立即可用的解決方案。
- **Amazon CloudWatch 自訂指標**：用於作業系統、應用程式和業務指標收集。Amazon CloudWatch API 允許收集幾乎任何類型的自訂指標。
- **Amazon EC2 執行個體的運作狀態**：用於查看狀態檢查和根據狀態排程您執行個體的事件，例如自動重新啟動或重新啟動執行個體。
- **Amazon SNS**：用於設定、操作和傳送事件導向的通知。
- **AWS X-Ray**：藉由分析跨系統元件的資料流，對分散式應用程式和微服務架構進行偵錯和分析。
- **Amazon Elasticsearch 服務**：用於集中日誌收集和即時日誌分析。用於進行快速、啟發式問題偵測。
- **第三方工具**：用於即時分析和全堆疊監控與可見性。
- **標準作業系統監控工具**：用於作業系統層級的監控。

有關 AWS 監控工具的詳細資訊，請參閱[自動和手動監控](#)。²¹另請參閱[使用 Amazon CloudWatch 儀表板](#)²²和[發佈自訂指標](#)。²³

通知

您的基礎設施事件設計中的關鍵操作元素為警示與通知的配置，以整合監控解決方案。這些警報和通知可以與 AWS Lambda 等服務一起使用，以觸發基於提醒的操作。自動化對操作事件的回應是以最大回應能力實現緩解，轉返和恢復的關鍵因素。

還應該使用工具來集中監控工作量，並根據與關鍵操作指標相關的可用日誌和指標創建適當的警報和通知。這包括用於超出異常情況的警報和通知，以及服務或元件故障。理想情況下，當低性能閾值被越過或發生故障時，系統已被架構為自動自我修復或回應於此類通知和警報而進行縮放。

如前所述，AWS 提供服務 (Amazon SQS 和 Amazon SNS)，以確保規劃外的營運事件可獲得適當的警示與通知作為回應，並可啟用自動化回應。

營運準備 (事件天數)

計劃執行

在活動當天，參與基礎設施事件的核心團隊應該在現場電話會議監控現場儀表板。執行手冊應該完全開發並可用。確保通訊計劃得到所有支援人員和利益相關者的明確界定並知情，且製定了應急計劃。

戰備室

事件期間，開放以下與會者使用現場會議橋接器：

- 主要負責的應用程式和營運團隊
- 營運團隊領導階層
- 來自外部合作夥伴，直接涉及技術交付的技術資源
- 業務利害關係人

在大多數事件中，這個會議橋接器的對話應該最小化。如果發生不良的操作事件，可對事件作出反應的關鍵人員已經進入房內，準備採取行動和諮詢。

領導階層報告

在事件發生期間，每小時傳送電子郵件到關鍵利害關係人。此更新應該包含下列項目：

- 狀態摘要：綠色 (追蹤中)、黃色 (遭遇問題)、紅色 (重大問題)
- 關鍵指標更新
- 遭遇問題、補救計劃狀況、預計補救時間
- 戰備室電話會議橋分器的電話號碼 (萬一有人想要加入的情況)

在會議結束後，最終摘要電子郵件應依照相似的格式傳送。

緊急計畫

事件準備過程中的每一步都應該有一個已經在測試環境中驗證的相應轉返計劃。

當您在組成轉返計劃時，請考慮以下問題：

- 在事件期間會發生的最糟情況是什麼？
- 什麼類型的事件會造成負面的公共關係影響？
- 活動期間哪些第三方元件和服務可能會故障？
- 應監控哪一項指標才能了解已發生不良的情況？
- 每個可能情況下的轉返計劃為何？
- 轉返程序需要多長時間？什麼是可接受的復原點目標 (RPO) 和復原時間目標 (RTO)？(請參閱[使用 AWS 進行災難復原](#)，以詳細了解這些概念)。²⁴

考慮以下的轉返類型：

- **藍/綠部署**：如果推出新的生產應用程式或環境，請保持前期生產於線上建立，並可快速切換。
- **暖色試用**：在第二個區域啟動可視需要快速擴展的最小環境。在主區域故障的情況下，可在備份區域中快速擴展，並將流量切換到第二個區域。
- **維護模式錯誤頁面**：在 Web 服務的每一層檢查錯誤頁面設施和觸發器。準備根據需要向這些錯誤頁面注入更具體的消息。

測試和記錄每個可能的故障情況的轉返計劃。

後續事件活動

事件後續分析

事件後續分析經常被忽視，因為客戶通常會急於恢復正常運作。但是，我們建議您對任何基礎設施事件管理生命週期要求進行事件後續分析。事件後續分析可讓您與所涉及各團隊合作，並找出可能需要進一步優化的領域，例如：操作程序、實作詳細資訊、容錯移轉與復原流程等。特別若應用程式堆疊在事件中發生中斷時，更是有其必要。事件後續分析也有助於提供文件，若有任何需要開發根本原因分析 (RCA) 文件的需求。

降級流程

立即遵照基礎設施事件的結論，開始降級流程。在此期間，建議您持續監控相關的應用程式和服務，以確保流量已將回復到正常的生產層級。使用在準備階段創建的任何健康儀表板來驗證流量和交易率的正常化。一些事件的降級期可能為線性和直接降級，而其他事件可能會經歷流量不均或逐漸減少。有些流量模式可能會持續存在。例如，從流量復元通常需要直接的降級流程，而諸如應用程式部署或擴展到新的地理區域之類的東西可能具有長期的影響，需要您仔細監視新的流量模式，並對永久應用程式堆疊部分進行額外的監控。

在某個時間點的事件完成之後，您必須決定何時才能安全地結束事件管理操作。請參閱以前記錄的關鍵指標的「正常」值，以幫助確定何時宣布事件已完成或結束。我們建議將緩降活動分為兩個分支，其可能有不同的時間表。第一個分支重點關注事件的運營管理，例如向內部和外部利益相關者和合作夥伴傳送通訊，以及重新設置服務限制。第二分支重點放在下調的技術方面，如縮小程序、環境健康狀況的驗證以及確定架構變更是否應恢復或落實的標準。

與每個分支相關聯的時間表可以根據事件的性質、關鍵指標和客戶舒適度而變化。我們概述了下表中與每個分支相關聯的一些常見任務，以幫助您確定事件的適當時間管理。

表 2：營運降級任務

任務	說明
通訊	通知內部和外部利害關係人事件已結束。結束的通訊應符合事件完成的定義。使用「傳回運作狀態良好的」指標，以確定其為結束通訊的適當時機。或者，您可以依層結束通訊。例如，您可以結束戰備室橋接器，但保留事件提升程序完整性，以備萬一出現後續事件故障時使用。

任務	說明
服務限制/成本控制	雖然在事件發生後可能會保留提高的服務限制，但請注意，服務限制也被用作安全網。服務限制通過防止多餘的服務使用（如：受到損害的帳戶或配置錯誤的自動化），保護您和您的成本。
報告和分析	事件指標的數據收集和整理，以及顯示模式、趨勢、問題領域、成功程序、特設程序、事件時間表以及成功標準是否得到滿足的分析性敘述，均應在通信計劃中制定並分發給所有已確定的內部方。還應制定詳細的成本分析，以顯示支援事件的營運費用。
優化任務	企業組織隨著時間的推移不斷發展，不斷改進業務。操作優化需要不斷收集指標、操作趨勢和從事件中吸取的經驗教訓，以發現改進的機會。優化與準備結合起來形成一個回饋迴圈，以解決運營問題，防止其重複發生。

表 3：技術降級任務

任務	說明
服務限制/成本控制	雖然在事件發生後可能會保持高度的服務限制，但請注意，服務限制也有助於成為安全網。服務限制通過防止多餘的服務使用來保護您的運營和運營成本，無論是由於受到破壞的帳戶所引起的惡意活動莖幹化或配置錯誤的自動化。
縮減程序	在準備階段還原擴展的資源。這些項目對於您的架構是獨一無二的，但以下為常見範例： EC2/RDS 執行個體大小 Auto Scaling 組態 預留容量 佈建 IOPS
良好運作環境的驗證	與基準指標進行比較，並檢查生產情況，以驗證事件發生後和縮小程序完成後，受影響的系統報告正常行為。
架構變更部署	根據事件的性質和操作指標的觀察，為事件準備做出的一些改變可能值得保留。例如，擴展到新的地理區域可能需要永久增加該區域的資源，或者提高某些服務限制或配置參數，例如數據塊中的分區數或卷中的 PIOPS 流中的分片數量可能是應該堅持的性能調校措施。

優化

也許基礎設施事件管理的最重要的組成部分是事件後續分析和確定所觀察到的運營和架構挑戰以及改進的機會。基礎設施事件很少的一次性事件。它們可能是季節性的或與新版本的應用程式重合，或者它們可能成為公司成長的一部分，因為它擴展到新的市場和領域。因此，每個基礎設施事件都是為下一個機會更有效地觀察、改進和準備的機會。

結論

AWS 以彈性和可編程產品和服務的形式提供構建塊，貴公司可以組裝以支持幾乎任何規模的工作量。通過 AWS 基礎設施事件指導方針和最佳實踐，加上我們完整的高可用性服務，貴公司可以為主要業務活動進行設計和準備，並確保擴展需求能夠順利和動態地滿足，確保快速回應和全球性業務覆蓋。

作者群

協力完成這份文件的個人與組織如下：

- Presley Acuna，AWS 企業支援經理
- Kurt Gray，AWS 全球解決方案架構師
- Michael Bozek，AWS 資深技術客戶經理
- Rován Omar，AWS 技術客戶經理
- Will Badr，AWS 技術客戶經理
- Eric Blankenship，AWS 資深技術客戶經理
- Greg Bur，AWS 技術客戶經理
- Billing Hesse，WS 資深技術客戶經理
- Hasan Khan，AWS 資深技術客戶經理
- Varun Bakshi，AWS 資深技術客戶經理

深入閱讀

如需更多閱讀操作和架構最佳實務的資料，請參閱 [AWS 操作檢查清單](#)。²⁵我們建議讀者檢閱 [AWS 良好架構框架](#) ²⁶，以獲得評估他們的雲端應用程式交付堆疊的結構性方法。AWS 提供基礎設施事件管理 (IEM) 做為提供給想要更直接與 AWS 技術客戶經理與支援工程師在其設計、規劃和事件期間操作有接觸的客戶更高階的支援。有關 AWS IEM 高級支援服務的詳細資訊，請參閱[基礎設施事件管理](#)。²⁷

附錄

詳細的架構審閱檢查清單

是 - 否 - 無	安全性
☐	我們依據 AWS 安全性最佳實務，輪換 AWS 識別與存取管理 (IAM) 存取金鑰和使用者密碼，並最長三個月認證一次應用程式所涉及的資源。我們在每個帳戶中皆應用密碼政策，且使用硬體或虛擬多重驗證 (MFA) 裝置。
☐	我們有內部安全流程和控制，用於控制使用 IAM 獨特的、基於角色的、最少權限訪問 AWS API。
☐	我們已移除任何保密或敏感資訊，包括內嵌公有/私有執行個體關鍵配對，並審查從任何自訂 Amazon Machine Image (AMI) 的所有 SSH 授權金鑰檔。
☐	我們使用適用於 EC2 執行個體的 IAM 角色以方便使用，而不是嵌入任何認證於 AMI 內部。
☐	我們通過創建 IAM 管理角色並限制來自其他功能角色的 IAM 操作，來將 IAM 管理權限與常規使用者權限隔離開來。
☐	我們在 Windows 或 Linux 執行個體的 EC2 執行個體上應用最新的安全修補程式。我們使用作業系統存取控制，包括 Amazon EC2 安全群組規則、VPC 網路存取控制清單、作業系統強化、以主機為基礎的防火牆、入侵偵測/防護、監控軟體設定和主機清單。
☐	我們確保來自或至公司的 AWS 以及使用加密通訊協定傳輸公司環境的網路連線。
☐	我們應用集中化日誌和審計管理解決方案，來識別和分析任何異常訪問模式或對環境的任何惡意攻擊。
☐	我們的安全事件和事件管理、校正與回報流程已就位。
☐	我們確保在任一安全群組，沒有能不受限制存取 AWS 資源的群組。
☐	我們在前端連線上 (用戶端到負載平衡器) 使用安全的通訊協定 (HTTPS 或 SSL)、最新的安全政策和加密通訊協定。在用戶端和負載平衡器之間的要求經過加密，使其更加安全。
☐	我們設定 Amazon Route 53 MX 資源記錄集，以擁有包含相應送件人政策框架 (SPF) 值的 TXT 資源記錄集，指定授權為我們網域傳送電子郵件的伺服器。

是 - 否 - 無	可靠性
☐	我們部署在 Auto Scaling 群組的 EC2 執行個體群組上部署我們的應用程式，以確保根據預先定義的擴展計劃自動進行水平擴展。 進一步了解。
☐	我們在我們的 Auto Scaling 群組配置上使用 Elastic Load Balancing 運作狀態檢查，以確保 Auto Scaling 群組在其下健康的 EC2 執行個體上運作。(僅在您在 Auto Scaling 群組中使用負載平衡器時適用)。
☐	我們在跨多個可用區域部署應用程式的關鍵元件會適當地複製區域間的數據。我們測試這些元件的故障會如何影響使用 Elastic Load Balancing、Amazon Route 53 或任何適合的第三方工具的應用程式可用性。

是 - 否 - 無	可靠性
□-□-□	在資料庫層，我們在多個可用區域部署 Amazon RDS 執行個體以藉由同步複寫至不同可用區域中的備用執行個體，來增強資料庫的可用性。
□-□-□	我們已定義在任何停止運轉或效率降低的情況下，程序會自動或手動容錯移轉。
□-□-□	我們使用 CNAME 記錄以將 DNS 名稱映入我們的服務。我們不使用 A 記錄。
□-□-□	我們為 Amazon Route 53 記錄集設定較低的存活時間 (TTL) 值。這可避免在 DNS 解析器在重選流量路徑而請求更新 DNS 記錄時造成的延遲。(例如：當 DNS 容錯移轉偵測並回應其中一個終端節點的故障時)。
□-□-□	我們至少有兩種 VPN 通道設定，以在停止運轉或 AWS 端點裝置計劃性維修時，提供備份功能。
□-□-□	我們使用 AWS Direct Connect，且隨時擁有兩個 Direct Connect 連線設定，以在裝置不可使用的情况下，提供備份功能。連線配置於不同的 Direct Connect 位置，以便在一個位置不可使用時提供備份功能。 我們也設定連接到我們的虛擬私有閘道，以便讓多個虛擬界面配置於多個 Direct Connect 連線和位置。
□-□-□	我們使用 Windows 執行個體，並確保我們使用的是最新的 PV 驅動程式。PV 驅動程式有助於優化驅動程式效能並將執行時間問題和安全風險最小化。我們也確定 EC2Config 代理程式在 Windows 執行個體上所執行的版本為最新版本。
□-□-□	我們為 Amazon Elastic Block Store (EBS) 磁碟區拍攝快照，以確保萬一發生任何故障時，可進行 point-in-time 恢復。
□-□-□	我們在適當的作業系統和應用程式/資料庫的資料上，使用單獨的 Amazon EBS 磁碟區。
□-□-□	我們在任何 Linux 執行個體上套用最新的核心、軟體和驅動程式修補程式。

是 - 否 - 無	表現效率
□-□-□	在上機前，我們為 AWS 託管的應用程式元件進行包括效能測試的全測。我們也執行負載測試，以確保我們使用了適當的 EC2 執行個體大小、IOPS、RDS 資料庫執行個體大小等數量。
□-□-□	我們針對服務限制執行用量檢查報告，並確保目前的 AWS 服務使用量等於或小於服務限制的 80%。 進一步了解
□-□-□	我們使用內容交付/分發網路 (CDN)，以利用我們的應用程式 (Amazon CloudFront) 快取，並作為內容優化傳送與內容自動分發至使用者最近地點的一種方式。
□-□-□	我們了解一些 Amazon CloudFront 收到動態的 HTTP 請求標頭 (使用者代理程式、日期等) 可能會因為降低快取命中率和增加原始伺服器負載而影響效能。 進一步了解
□-□-□	我們確保 EC2 執行個體的最大輸送量大於所附加的 EBS 磁碟區彙總的最大輸送量。我們也使用 EBS 優化執行個體與 PIOPS EBS 磁碟區，以獲得預期的磁碟區效能。
□-□-□	我們確保解決方案設計在基礎設施沒有瓶頸或在資料庫或應用程式設計沒有壓力點。
□-□-□	我們在應用程式資源部署監控並使用 Amazon CloudWatch 或第三方合作夥伴工具，依據任何效能漏洞設定警報。

是 - 否 - 無	表現效率
□-□-□	我們的設計可避免在附加於我們應用程式執行個體的任何安全群組中使用大量規則。在安全群組中的大量規則可能會降低效能。

是 - 否 - 無	成本優化
□-□-□	我們了解基礎設施事件可能包含一些在事件後需要清除的過度佈建容量，以避免任何不必要的成本。
□-□-□	我們在所有的基礎設施元件上，包括 EC2 執行個體大小、RDS 資料庫執行個體大小、快取叢集節點大小和數字、Redshift 叢集節點和 EBS 磁碟區大小，都使用正確的大小。
□-□-□	在方便時，我們會使用競價型執行個體。競價型執行個體是擁有彈性開始與結束時間工作負載的理想選擇。競價型執行個體的典型使用案例：批次處理、報告產出和高效能運算工作負載。
□-□-□	我們有可預測的應用程式容量最低需求，且會利用預留執行個體。預留執行個體可讓您預留 Amazon EC2 的運算容量，以換取相較於隨需執行個體定價更大的小時費率折扣。

Notes

- ¹ <https://aws.amazon.com/answers/account-management/aws-tagging-strategies/>
- ² <https://aws.amazon.com/blogs/aws/resource-groups-and-tagging/>
- ³ <https://aws.amazon.com/sqs/>
- ⁴ <http://docs.aws.amazon.com/general/latest/gr/rande.html>
- ⁵ <https://aws.amazon.com/emr/>
- ⁶ <https://aws.amazon.com/rds/>
- ⁷ <https://aws.amazon.com/ecs/>
- ⁸ <https://aws.amazon.com/sns/>
- ⁹ <https://aws.amazon.com/blogs/compute/using-aws-lambda-with-auto-scaling-lifecycle-hooks/>
- ¹⁰ <http://docs.aws.amazon.com/lambda/latest/dg/welcome.html>
- ¹¹ <https://aws.amazon.com/blogs/aws/new-auto-recovery-for-amazon-ec2/>

- 12 <https://aws.amazon.com/answers/configuration-management/aws-infrastructure-configuration-management/>
- 13 https://d0.awsstatic.com/whitepapers/Big_Data_Analytics_Options_on_AWS%20.pdf
- 14 <http://docs.aws.amazon.com/Route53/latest/DeveloperGuide/routing-policy.html#routing-policy-latency>
- 15 <https://aws.amazon.com/elasticache/>
- 16 <https://aws.amazon.com/cloudfront/>
- 17 <http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/concepts-on-demand-reserved-instances.html>
- 18 <http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-spot-instances.html>
- 19 https://docs.aws.amazon.com/general/latest/gr/aws_service_limits.html
- 20 <https://aws.amazon.com/about-aws/whats-new/2014/07/31/aws-trusted-advisor-security-and-service-limits-checks-now-free/>
- 21 http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/monitoring_automated_manual.html
- 22 http://docs.aws.amazon.com/AmazonCloudWatch/latest/monitoring/CloudWatch_Dashboards.html
- 23 <http://docs.aws.amazon.com/AmazonCloudWatch/latest/monitoring/publishingMetrics.html>
- 24 <https://aws.amazon.com/blogs/aws/new-whitepaper-use-aws-for-disaster-recovery/>
- 25 http://media.amazonwebservices.com/AWS_Operational_Checklists.pdf
- 26 http://d0.awsstatic.com/whitepapers/architecture/AWS_Well-Architected_Framework.pdf
- 27 <https://aws.amazon.com/premiumsupport/iem/>