

Red Hat Enterprise Linux 10

RHEL 10 新時代：從基礎作業系統到智慧平台的進化之路

Part 2

Ato Lin
Sr. Solutions Architecture
Red Hat Taiwan





System Roles 簡化一致性設定

Ato Lin | 林嘉彥

Senior Solution Architect

Red Hat



Red Hat Enterprise (RHEL) System Roles

RHEL System Roles 是 Ansible 角色和模組的集合，提供穩定一致的設定介面，用於自動化 Red Hat Enterprise Linux 的多個版本。

- ▶ 基於 Linux 系統角色上游專案的開發
- ▶ 可透過 Red Hat 軟體管道取得
 - 以 Ansible 集合和 RPM 包的形式分發
- ▶ 由 Red Hat 提供全面支援、維護和更新

哪裡可以取得 System Roles

可從 RPM 套件 以及在 Automation Hub 上取得

從 Automation Hub 取得

- ▶ 透過 ansible-galaxy 安裝

```
# ansible-galaxy collection install redhat.rhel_system_roles
```

- ▶ 安裝後角色可使用 `redhat.rhel_system_roles.<role_name>`

從 RPM 套件安裝

- ▶ `yum install rhel-system-roles`

如何開始使用RHEL System Roles

1. 啟用RHEL

2. 安裝套件

```
sudo yum install ansible rhel-system-roles
```

3. 看一下 [README.md](#)

```
/usr/share/doc/rhel-system-roles/<role_name>/README.md
```

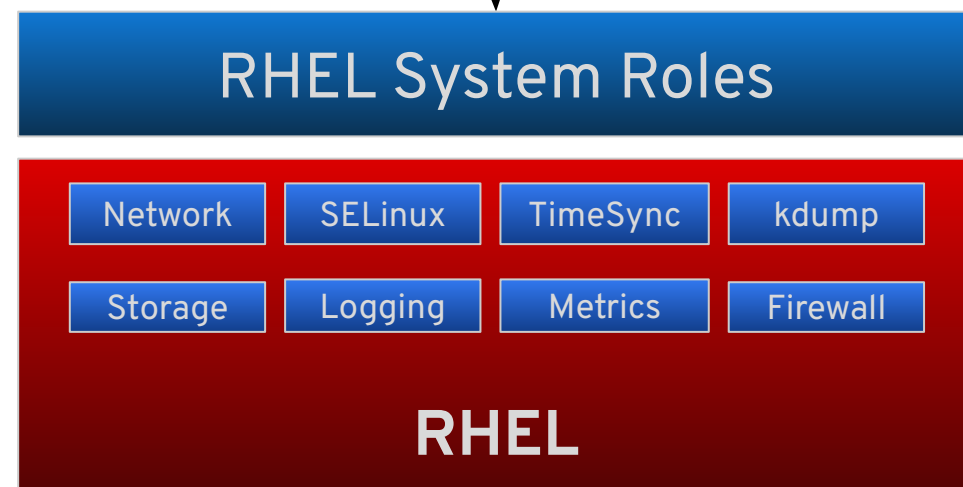
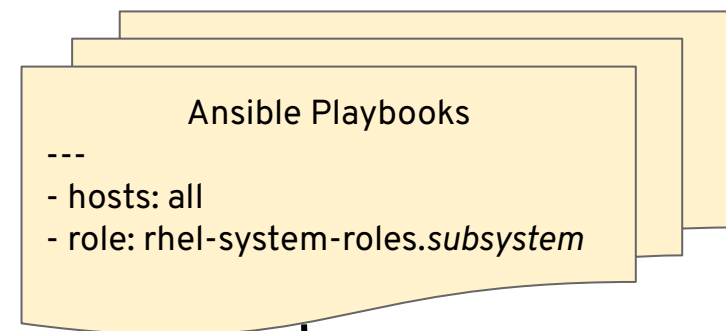
4. 撰寫腳本 -> 設定變數 -> 引用Roles

5. 執行腳本

RHEL System Roles

RHEL System Roles (as of 8.6/9.1)

- ▶ ad_integration
- ▶ certificate
- ▶ cockpit
- ▶ crypto_policies
- ▶ firewall
- ▶ ha_cluster
- ▶ kdump
- ▶ kernel_settings
- ▶ logging
- ▶ metrics
- ▶ nbde_client
- ▶ ndbe_server
- ▶ network
- ▶ podman
- ▶ postfix
- ▶ selinux
- ▶ ssh[d]
- ▶ storage
- ▶ timesync
- ▶ tlog
- ▶ vpn



simple DHCP

```
---
- hosts: rhel10
  vars:
    # network_provider: initscripts # or nm
  network_connections:
    - name: Private_Mgmt
      type: ethernet
      interface_name: eno0
      autoconnect: yes
      ip:
        dhcp4: yes
        auto6: no
  roles:
    - role: rhel-system-roles.network
```

simple STATIC

```
---
- hosts: rhel10
  vars:
    network_connections:
      - name: Pub_Web
        type: ethernet
        autoconnect: yes
        ip:
          auto6: no
          route_metric6: -1
          gateway6: 2001:db8::1
          address:
            - 192.168.99.99/24
            - 2001:db8::80/7
  roles:
    - role: rhel-system-roles.network
```


TimeSync

```
---
- hosts: rhel10
  vars:
    #ntp_implementation: ntp    # or chrony
    ntp_servers:
      - hostname: 0.rhel.pool.ntp.org
        iburst: true
      - hostname: foo.example.org
        pool: true
        minpoll: 6
        maxpoll: 10
        iburst: no
    ptp_domains:
      - interfaces: [ eth0 ]
  roles:
    - role: rhel-system-roles.timesync
```

SELinux

```
---
- hosts: rhel10
  vars:

    # Enable SELinux. Yes, do it, else Dan Walsh cries.
    SELinux_type: targeted
    SELinux_mode: enforcing
    SELinux_change_running: 1

    SELinux_booleans:
      - { name: 'samba_enable_home_dirs', state: 'on' }
      - { name: 'ssh_sysadm_login', state: 'on', persistent: 'yes' }

  roles:
    - role: rhel-system-roles.selinux
```

kdump

```
---  
- hosts: rhel10  
  
vars:  
  core_collector: "makedumpfile -l --message-level 2 -d 31 -c"  
  path: /var/crash  
  system_action: reboot # reboot | halt | poweroff | shell  
  
roles:  
  - role: rhel-system-roles.kdump
```

為什麼需要用到 System Roles 跟 Collections?



使用 Ansible Roles 和 Ansible Content Collections, 可以更快速地開發 playbooks, 並重複利用 Ansible 程式碼。

如果我看得更遠, 那是因為我站在巨人的肩膀上 - 牛頓

System Role Documentation

- ▶ RPM installation
 - roles_directory: /usr/share/ansible/roles
 - The roles directory is not added to the Ansible roles_path in /etc/ansible/ansible.cfg automatically.
- ▶ Documentation: /usr/share/doc/rhel-system-roles
 - https://access.redhat.com/documentation/en-us/red_hat_enterprise_linux/8/html/administration_and_configuration_tasks_using_system_roles_in_rhel/index

Demo : 使用 System Roles 設定防火牆

- ▶ <https://www.redhat.com/en/interactive-labs/configure-firewalls-red-hat-enterprise-linux-system-roles>
- ▶ 安裝 system roles
- ▶ 開啟連接埠，以允許外部連線至埠號 80 上的 httpd 服務。
- ▶ 開啟任意連接埠。
- ▶ 將對連接埠 9999 的 TCP 請求轉送至連接埠 12345。
- ▶ 刪除我們對防火牆規則所做的所有變更。

Demo : 使用 System Roles 建構標準作業環境

- ▶ <https://www.redhat.com/en/interactive-labs/build-standard-operating-environment-system-roles>
- ▶ 安裝 system roles
- ▶ 檢查腳本並使用。
- ▶ 驗證 System Roles 和設定是否已套用。
- ▶ 使用 System Roles 配置和更新系統配置。
- ▶ 更新現有設定。

RHEL System Roles 重點優勢

- ▶ 自動化與一致性

透過Ansible平台進行批量自動化管理, 確保多台伺服器設定標準化, 減少人為操作錯誤。

- ▶ 快速部署

系統角色預設最佳實踐配置, 只需設定關鍵變數, 即可快速完成如網路、時鐘同步、儲存等基礎設施的佈建。

- ▶ 節省維運成本

免去重複手動設定, IT 團隊可將精力集中於更高價值的工作, 節省人力及時間成本。

- ▶ 內建文件與持續更新

每個系統角色皆附詳細README, 清楚列出支援參數、範例與相容性, 且隨Red Hat官方定期更新, 保障安全與功能。

深度整合Ansible 實現自動化操作流程

Ato Lin | 林嘉彥

Senior Solution Architect

Red Hat



IT自動化的不同階段

個人



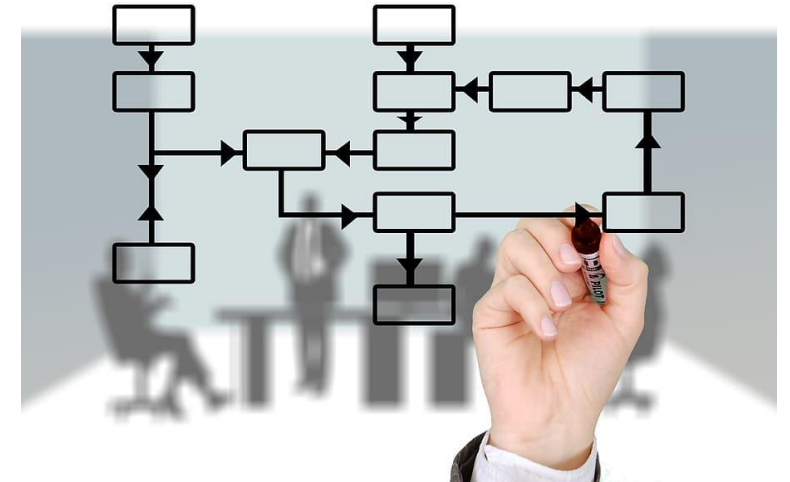
- 各式腳本
- 限於個人
- 欠缺規範
- 難以維護或分享

同一團隊



- 不同腳本或工具
- 限於局部
- 缺乏共同標準, 整合困難
- 資安管理是挑戰

組織共用機制



- 共用自動化機制
- 組織通用或跨部門
- 需要標準一致規範
- 必須有資安管控能力

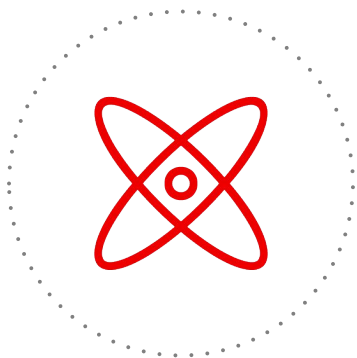
我已經有自動化機制，還需要繼續聽嗎？



好還要更好...

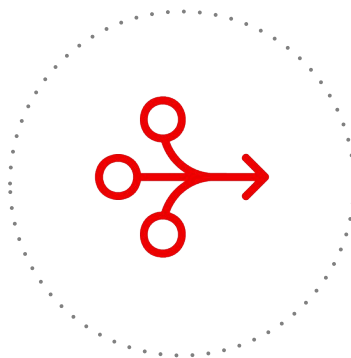
- 去除依賴限制
 - 使用Script, 無一定標準, 可自由發揮, 但是會造成維持、拓展或傳承不易
- 避免技術孤島
 - 使用專屬自動化機制於特定領域, 但跨領域整合不易且自動化範圍有限
- 需要集中化安全管理
 - 需要確保自動化機制本身機密資訊安全性, 以及工作執行的安全控管, 例如基於角色授權、稽核紀錄等

為什麼選擇 Ansible Automation Platform?



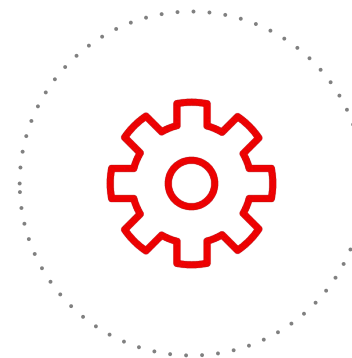
強大的

可用於大型企業環境
編排複雜的自動化流程



簡單的

橫跨多個專業領域
簡單的自動化腳本創建與管理



無代理程式的

輕鬆地整合在任何環境
可橫跨混合雲架構

自動化部署與管理自動化

您完整的IT足跡, 環境與設備支援廣泛, 所以多種IT角色與工作使用場景都可用

Do this...

資源編排

配置管理

應用部署

部署工裝

持續交付

安全與合規

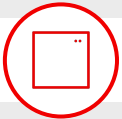
On these...



Firewalls



Load balancers



Applications



Containers



Virtualization platforms



Servers



Clouds



Storage



Network devices



And more ...

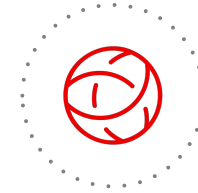
130+
certified platforms



Infrastructure



Cloud



Network



Security



ARISTA



Check Point
SOFTWARE TECHNOLOGIES LTD

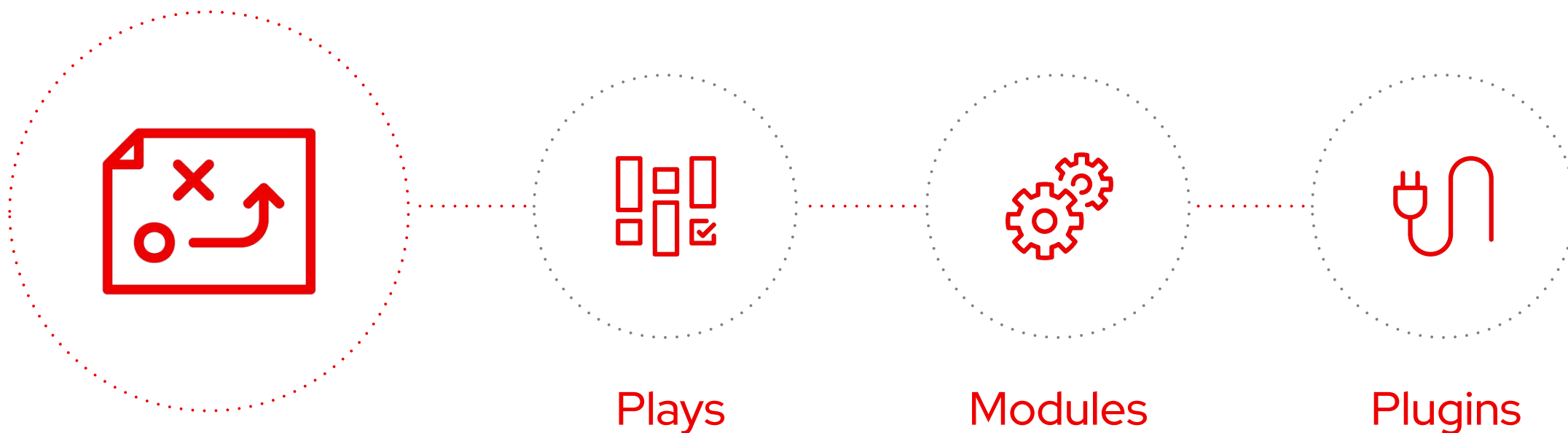


CYBERARK®



FORTINET®

What makes up an Ansible playbook?



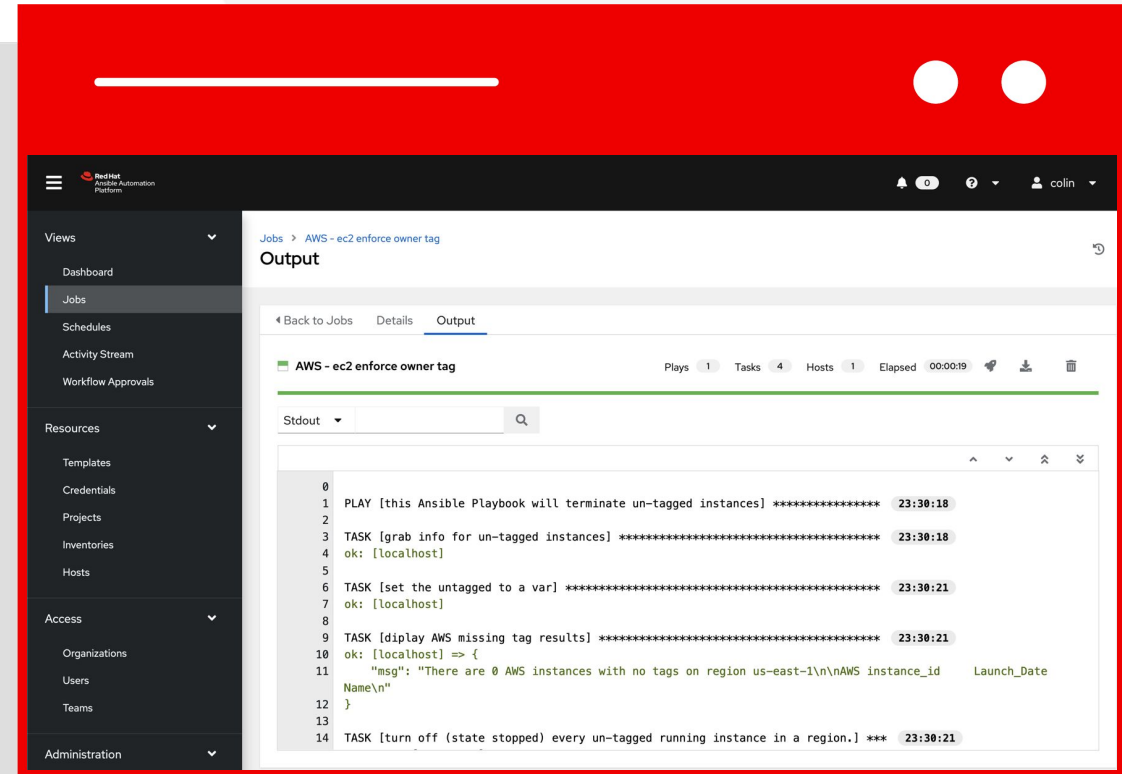


```
---  
- name: install and start apache  
  hosts: web  
  become: yes  
  
  tasks:  
    - name: httpd package is present  
      yum:  
        name: httpd  
        state: latest  
  
    - name: latest index.html file is present  
      template:  
        src: files/index.html  
        dest: /var/www/html/  
  
    - name: httpd is started  
      service:  
        name: httpd  
        state: started
```


A playbook run

Where it all starts

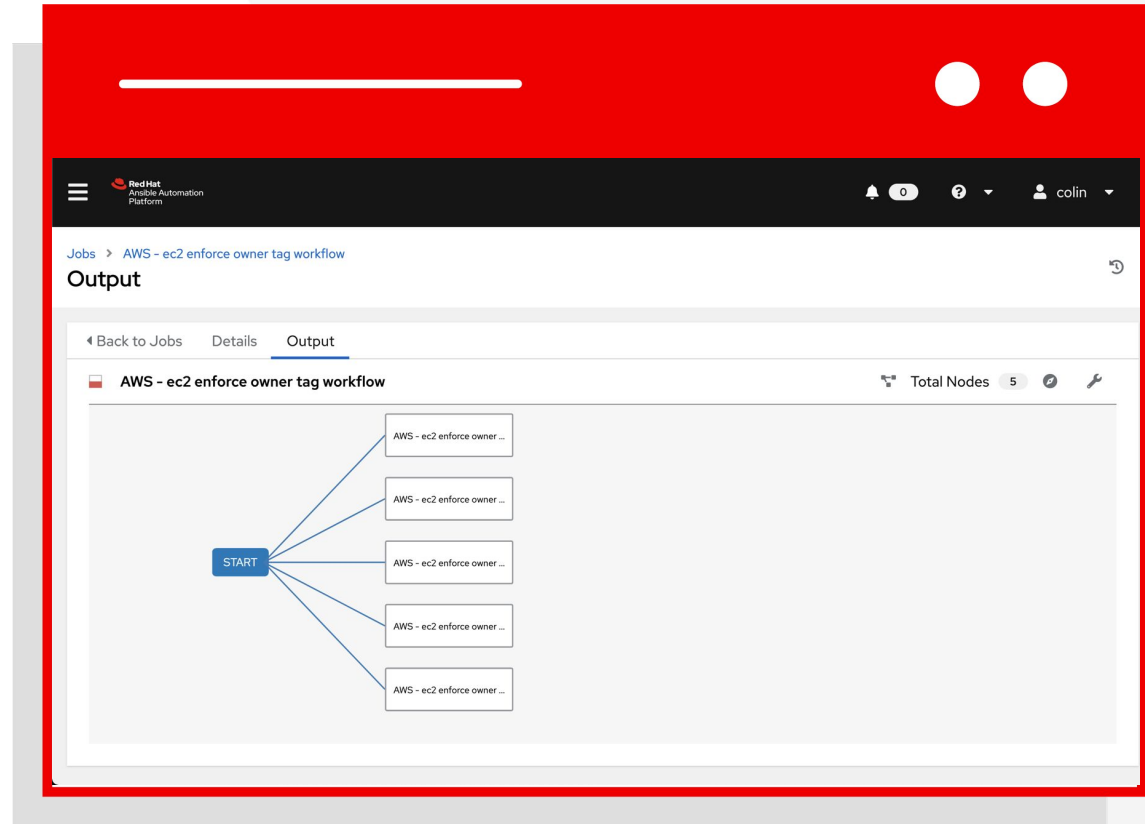
- ▶ A playbook is interpreted and run against one or multiple hosts – task by task. The order of the tasks defines the execution.
- ▶ In each task, the module does the actual work.



Workflows

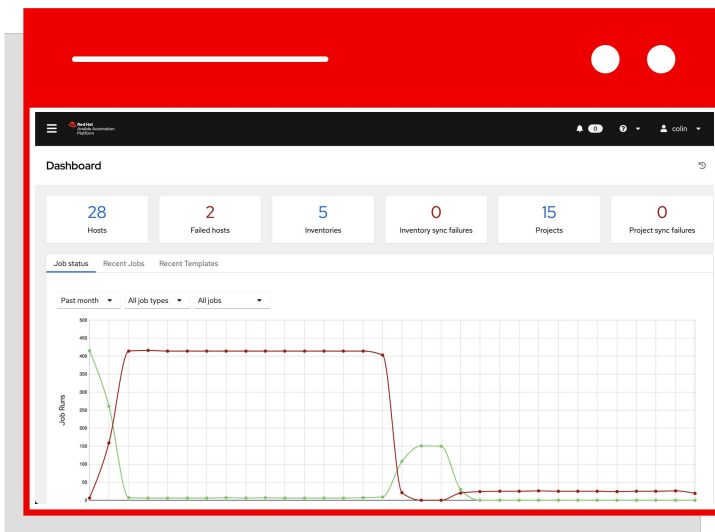
Combine automation to create something bigger

- ▶ Workflows enable the creation of powerful holistic automation, chaining together multiple pieces of automation and events.
- ▶ Simple logic inside these workflows can trigger automation depending on the success or failure of previous steps.

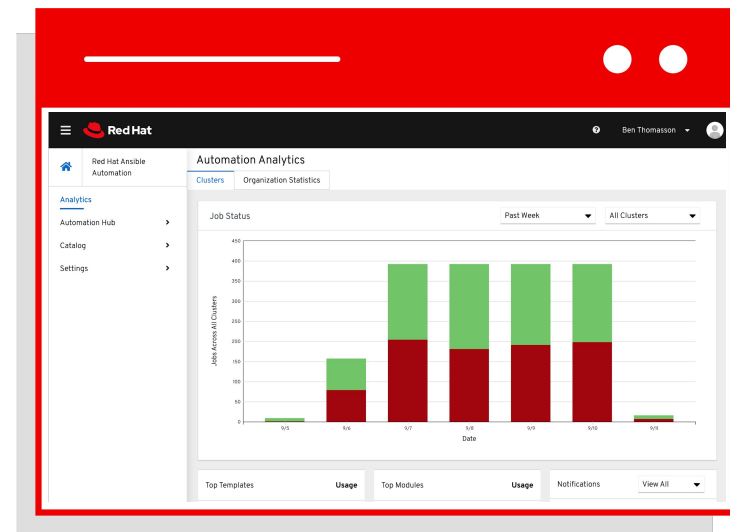


Dashboards

Understand what is going on



Overview of the actual cluster, the jobs happening, the nodes connected, what works and fails right at this moment.

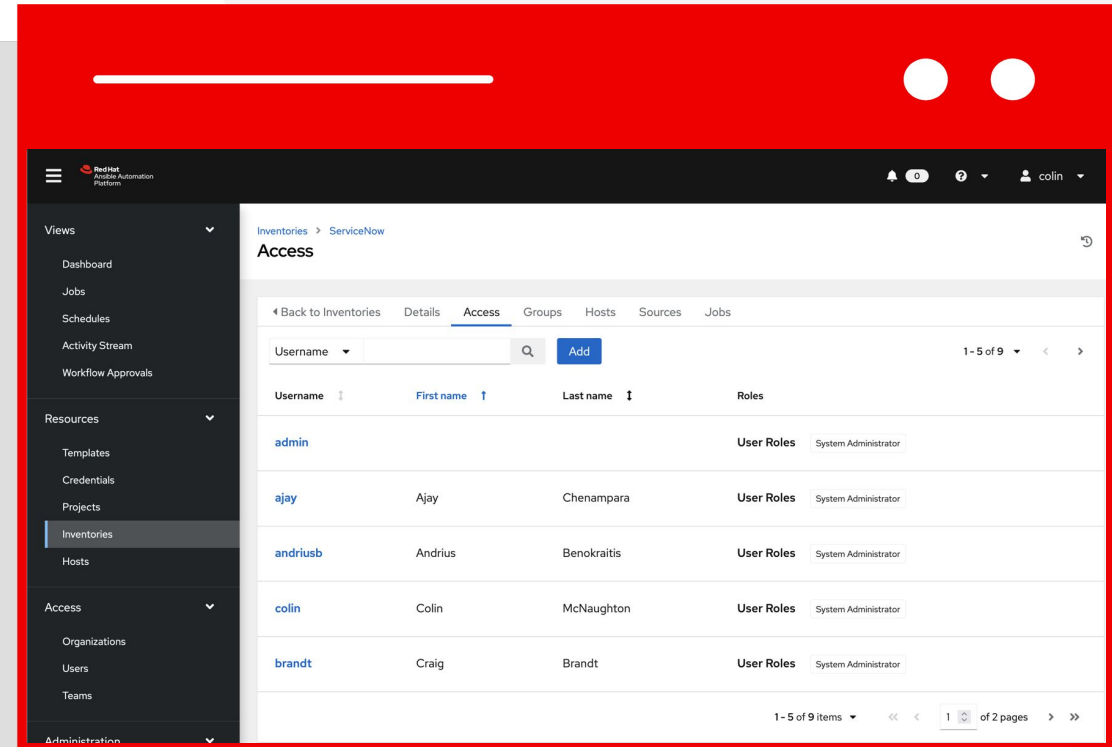


Overview across clusters, better insight into use cases of automation; insight into adoption of automation per organization.

Role-based access control

How to manage access

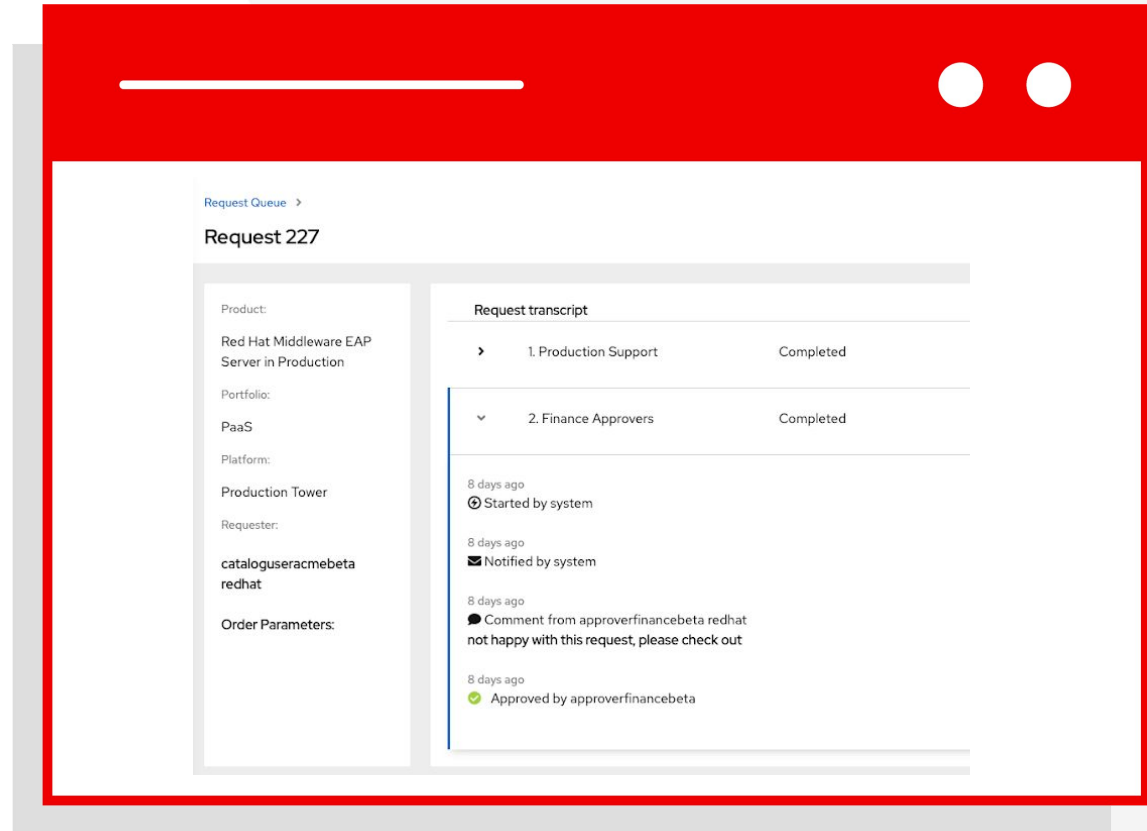
- ▶ Role-based access control system:
Users can be grouped in teams, and roles can be assigned to the teams.
- ▶ Rights to edit or use can be assigned across all objects.
- ▶ All backed by enterprise authentication if needed.



Approval

Govern access

- ▶ Approval processes allow a human interaction to the automation, to add governance like human level administration.
- ▶ Both available at the operational level on the Automation controller UI as well as in the service catalog level.



Ansible 自動化平台可支持的各式場景

一般通用

- 企業控管機制對接
- 作業系統/應用/配置的部署、設定、盤點、比對、調整、補丁、重啟
- 事故偵測自動回應

企業控管機制對接

- 簽核管理
- 事件管理
- 日誌管理
- 組態管理
- 帳號管理

系統/儲存/網路

- 運作狀態偵測與呈現
- VM resize
- ACL/Policy 調整
- 負載平衡流量調整
- 高可用提升
- 組態備份

混合雲

- 實體機安裝作業系統
- VM/K8S/OCP平台創建
- 雲端遷移
- 雲端運維
- 成本管理(ESG)

安全

- 減少人為介入
- 安全加固
- 憑證更新
- 帳號盤點
- 權限配置

DevOps

- API webhook集成
- 開發工具對接
- 應用程式部署
- 配置分發

異地備援與演練

- 啟動順序安排
- 啟動後檢查與回應
- 時間紀錄與報告

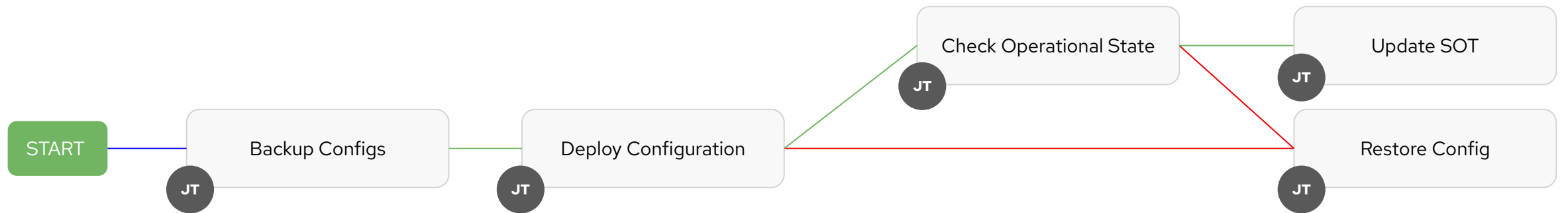
OT/Edge

- 連線管理
- 資訊回傳

Ansible Automation Platform

Using workflows to enhance your automation

WORKFLOW VISUALIZER | Operational State Workflow



CLOSE

SAVE

Ansible 重點優勢

- ▶ 自動化與一致性

透過Ansible平台進行批量自動化管理, 確保多台伺服器設定標準化, 減少人為操作錯誤。

- ▶ 快速部署

系統角色預設最佳實踐配置, 只需設定關鍵變數, 即可快速完成如網路、時鐘同步、儲存等基礎設施的佈建。

- ▶ 節省維運成本

免去重複手動設定, IT 團隊可將精力集中於更高價值的工作, 節省人力及時間成本。

- ▶ 內建文件與持續更新

每個系統角色皆附詳細README, 清楚列出支援參數、範例與相容性, 且隨Red Hat官方定期更新, 保障安全與功能。

A vertical red bar on the left side of the slide contains a complex, stylized graphic. It features a cloud with a keyhole, a database cylinder, a server rack, a computer monitor, and various arrows and symbols like 'X' and 'O' connected by lines, suggesting a technical or security theme.

SELinux 精細化策略應用

Ato Lin | 林嘉彥

Senior Solution Architect

Red Hat



檔案權限 RWX



Security-Enhanced Linux (Selinux)



預設安全性

- 沒有 SELinux, 使用者 Bob 對家目錄擁有完整存取權限, 可以更改其他人的權限。

Bob 也可能在家目錄中執行自訂腳本, 使系統容易受到攻擊。

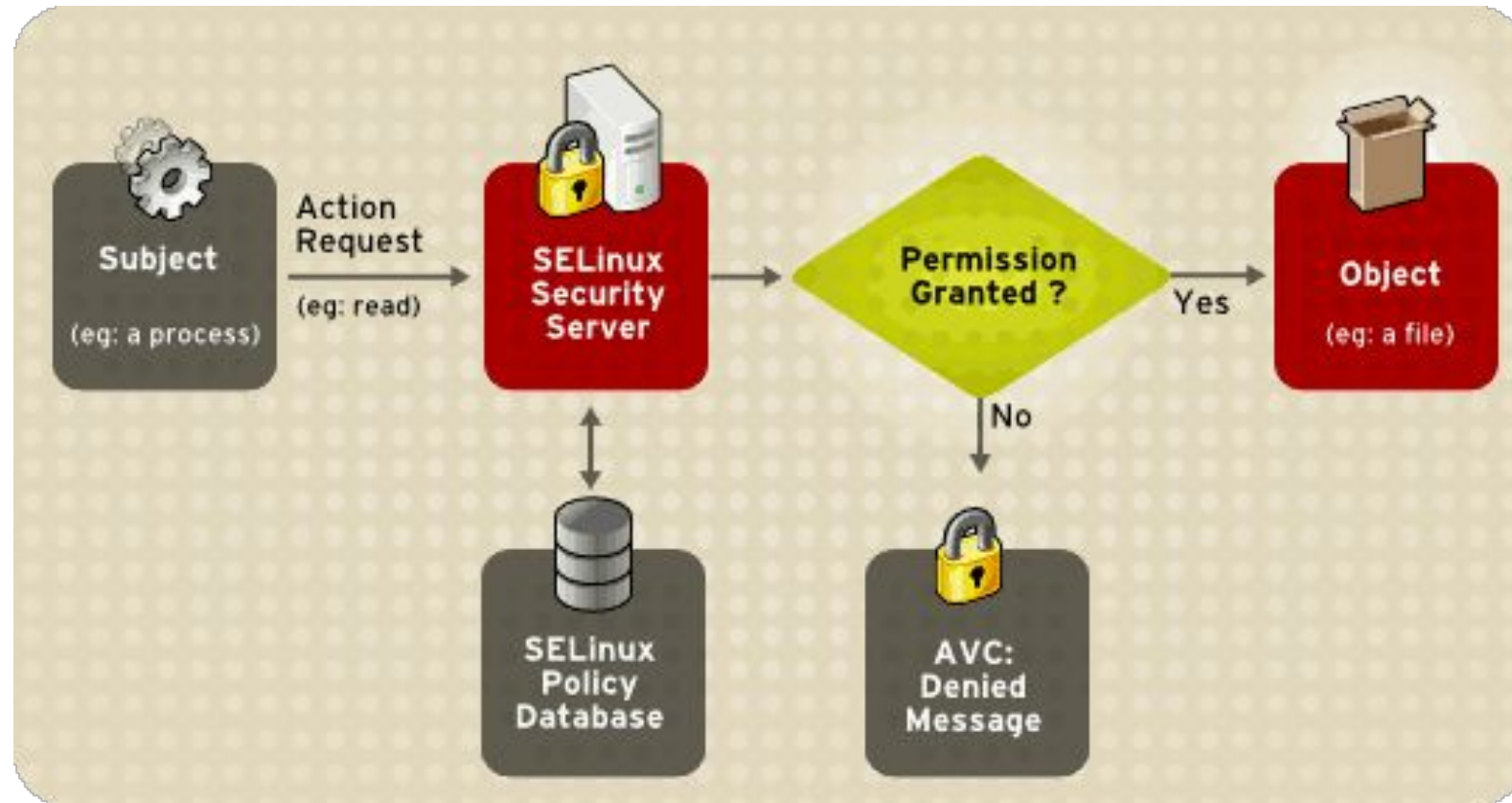
```
user  rwx  ->    目錄
group rwx  ->
other rwx  ->
```

- 有了 SELinux, 我們可以為使用者指定存取控制。
 - 我們可以阻止使用者在家目錄中執行腳本。

什麼是 SELinux

- SELinux 是一個 Linux 核心安全模組，提供支援存取控制安全政策的機制。
- 政策是 SELinux 的核心。政策包含使用者、角色和網域。
 - 政策定義
 - 使用者對角色的存取權
 - 角色對網域的存取權
 - 網域對檔案的存取權
- 不同的 SELinux 模式
 - 強制 (Enforcing): Fedora、CentOS、RHEL 的預設模式
 - 寬容 (Permissive): 允許程序執行的動作，然後記錄下來。有助於監控程式。
 - 停用 (Disabled): 允許動作且不記錄下來。

Security-Enhanced Linux (Selinux)



What is SELinux

每條 SELinux 政策規則都描述了程序Process與系統資源system resource之間的互動:

```
ALLOW apache_process apache_log:FILE READ;
```

允許 Apache 服務程式有讀取其日誌檔案的權限

SELinux 標籤以檔案系統的擴展屬性形式保存:

```
$ ls -Z /etc/passwd  
system_u:object_r:passwd_file_t:s0 /etc/passwd
```

SELinux 指令

- sestatus
- getenforce
- setenforce 1 # 1 means enforcing mode, 0 means permissive
- semodule -l
- semanage boolean -l
- getsebool <entry from semanage file>
- setsebool <entry from semanage file> on // if selinux is turned off for the entry
- /etc/selinux/config 控制 SELinux 的啟用狀態與預設模式，若要永久更改(如開關或從 permissive 改成 enforcing)，需改這個檔案並重開機才會生效

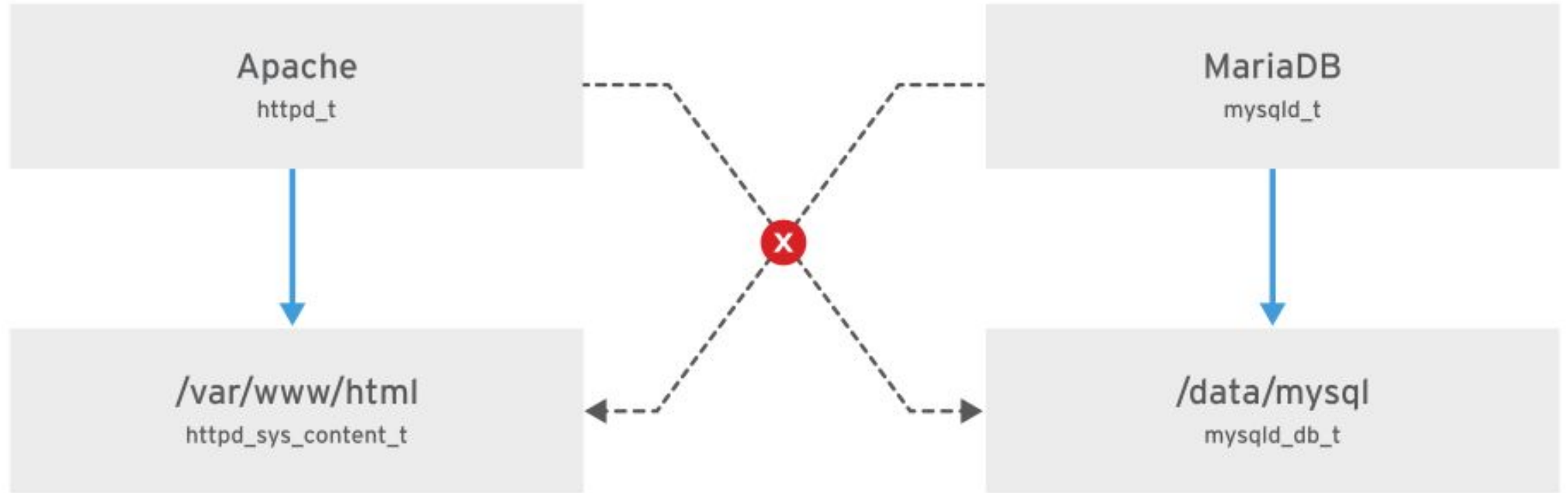
SELinux 的工作原理

- SELinux 是一套規則，決定哪個程序可以存取哪些檔案、目錄和連接埠
- 檔案、程序、目錄、連接埠都有 SELinux 上下文，一個特殊的安全標籤
- 上下文是政策用來決定程序是否可以存取檔案、目錄或連接埠的名稱
- SELinux 標籤有以下四個上下文：
 - User 使用者, role 角色, type 類型, sensitivity 敏感度

`unconfined_u:object_r:httpd_sys_content_t:s0 /var/www/html/file2`

The diagram illustrates the components of the SELinux context string `unconfined_u:object_r:httpd_sys_content_t:s0 /var/www/html/file2`. Each component is connected by a vertical line to its corresponding label below: `unconfined_u` is labeled "SELinux User", `object_r` is labeled "Role", `httpd_sys_content_t` is labeled "Type", `s0` is labeled "Level", and `/var/www/html/file2` is labeled "File".

Type Context



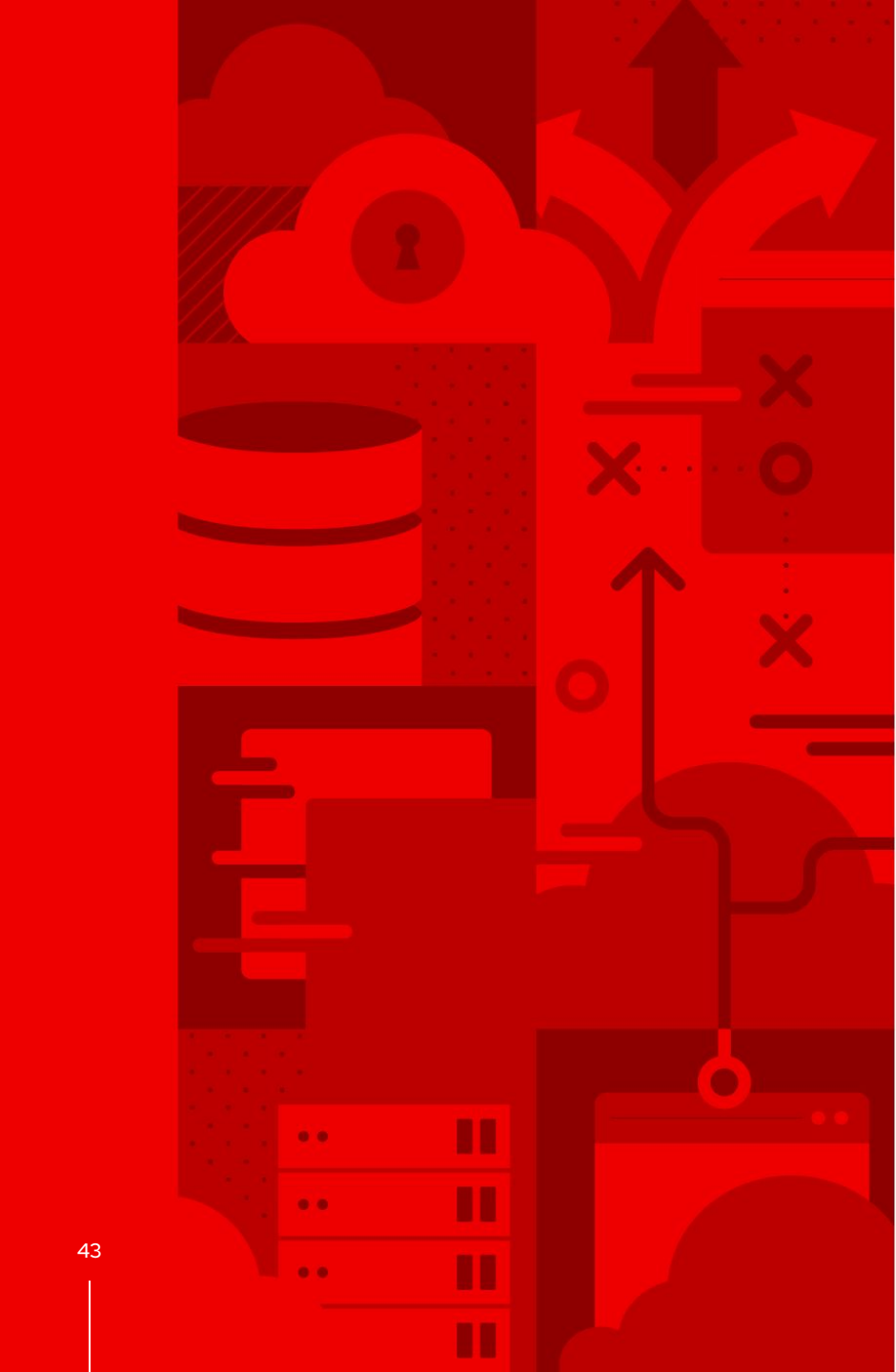
A vertical red bar on the left side of the slide contains a complex, stylized graphic. It includes a cloud with a keyhole, a database cylinder, a server rack, a computer monitor, and various arrows and symbols like 'X' and 'O' connected by lines, suggesting a technical or infrastructure theme.

Image Mode(鏡像模式) 信任鏈建立

Ato Lin | 林嘉彥

Senior Solution Architect

Red Hat



基礎設施和組織複雜性

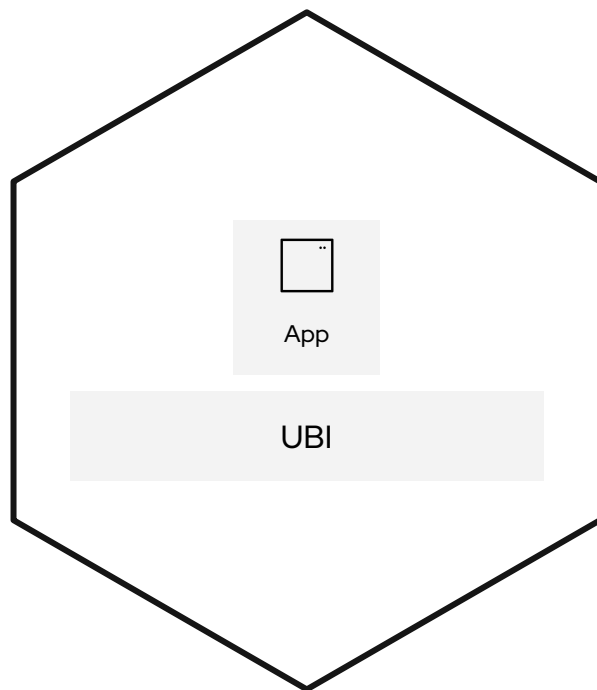
仍然是個問題...

作業系統的常見挑戰



- 不同的平台需要不同的工具、團隊和專業知識
- 測試和驗證非常耗時
- 應用程式支援清單
- 無人負責維護和升級, 且可以沒預算升級
- 不同負責人之間的溝通
- 鏡像、實例和運行時之間的搖擺
- 不變的願望 vs. 可變的現實
- 鏡像清單、版本控制
- 安全不容忽視！

容器徹底改變了應用程式部署

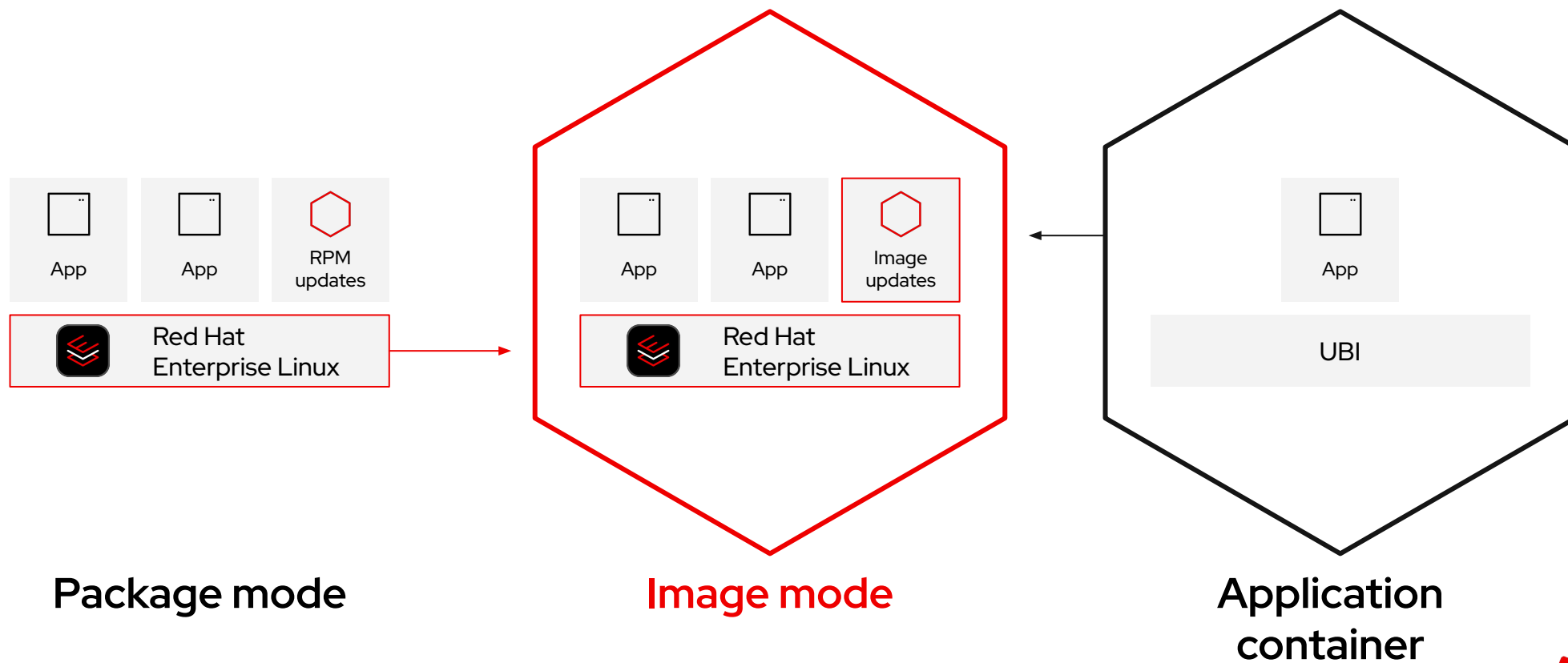


- 透過 OCI 鏡像格式實現標準化打包
- 透過 OCI 註冊表實現標準化交付
- Container file 清晰透明
- 部署可移植性和可預測性
- 豐富的安全、自動化與編排工俱生態系統
- 快速採用且普及

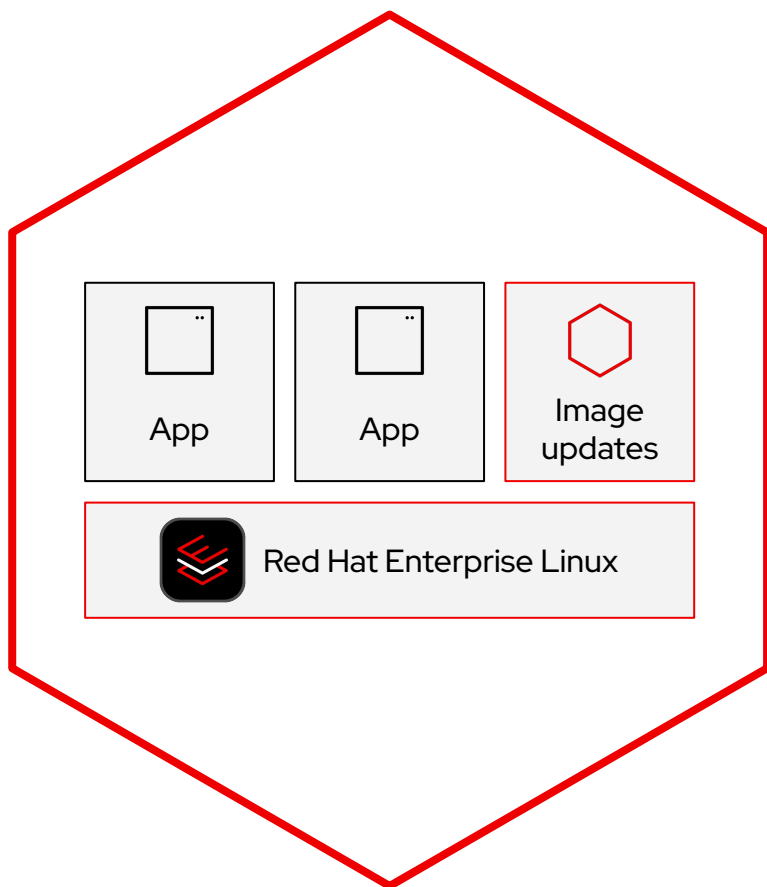
...它們也將成為現代 IT 的語言

RHEL Image Mode 鏡像模式介紹

將 RHEL 的強大功能與容器的優勢結合



利用容器實現標準化和創新



RHEL 的Image Mode鏡像模式是一種使用**容器技術**建置、部署和管理作業系統的簡單且一致的方法。現在，您可以使用與應用程式相同的工具和工作流程來管理作業系統，從而促進跨團隊的通用體驗和語言。

選擇權在你手中!

選擇最適合您組織需求的方法

	Package mode	Image mode
Build	Image builder	Container tools
Update	Packages (dnf)	Images (bootc)
Distribute	rpm repository	Container registry
Change	Run time	Build time
Manage	Red Hat Insights, Satellite, Ansible	
Deploy	Bare metal, VM, cloud, edge	

Image mode for Red Hat Enterprise Linux

簡單。一致。任何地方。

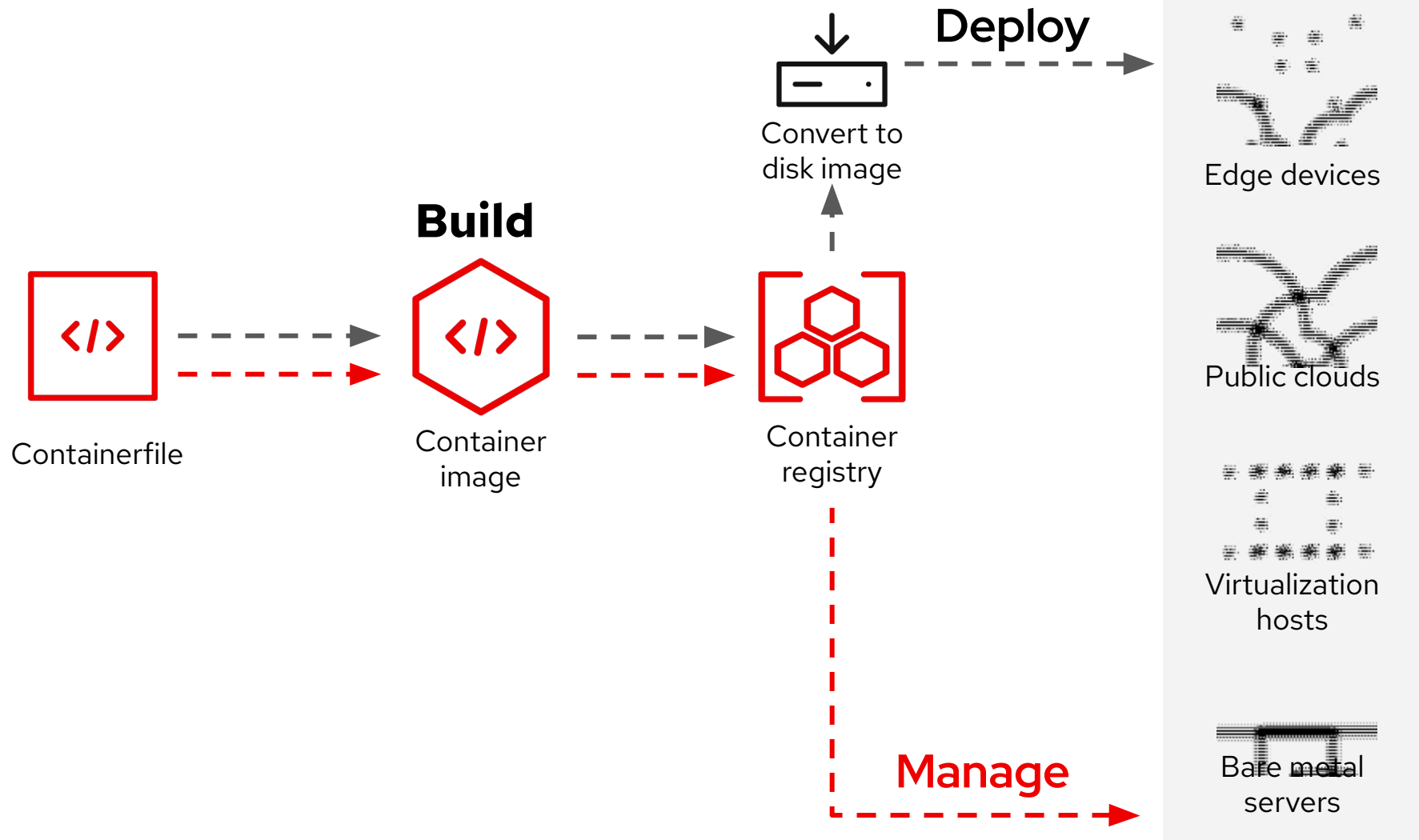


Image mode for RHEL

系統生命週期的容器原生工作流程

```
FROM rhel10/rhel-bootc:latest

RUN dnf install -y [software]
[dependencies] && dnf clean all

ADD [application]
ADD [configuration files]

RUN [config scripts]
```

建構 Build

僅使用 bootc 基礎映像和容器檔案即可定義您的整個系統(作業系統、應用程式和相依性)。利用您現有的容器工具和流水線(pipelines)快速建立和測試鏡像。

部署 Deploy

輕鬆轉換為虛擬機器/雲端鏡像, 透過 RHEL 安裝程式在裸機上部署, 甚至使用 bootc 在現有雲端鏡像上重新安裝。

管理 Manage

專為現代 GitOps 和 CI/CD 工作流程而設計。透過管線全面驅動和自動化系統, 或透過 Red Hat Insights、Satellite 和 Ansible 進行規模控制。

registry.redhat.io/rhel10/rhel-bootc:10.0

從 9.6 和 10.0 及更高版本開始支援 RHEL bootc 映像

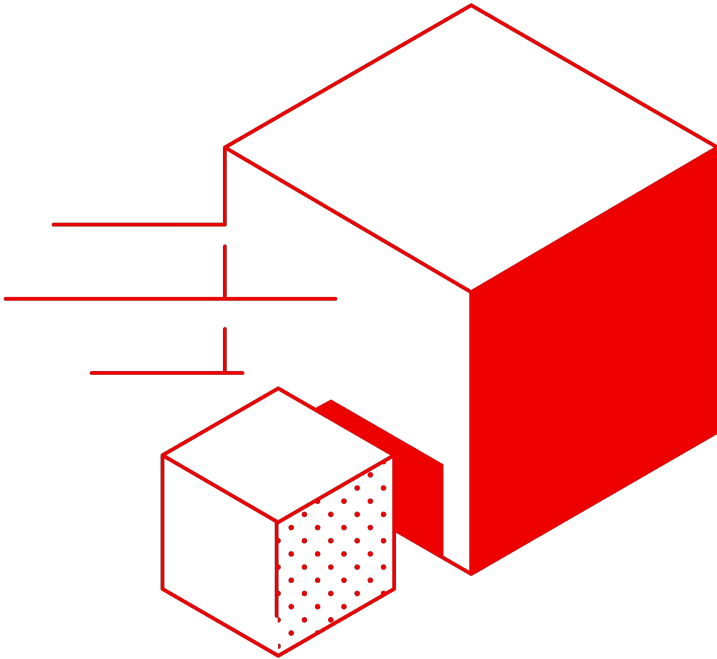


Image Specs:

- ~465 rpms
- ~690M compressed
- ~1.5G on disk

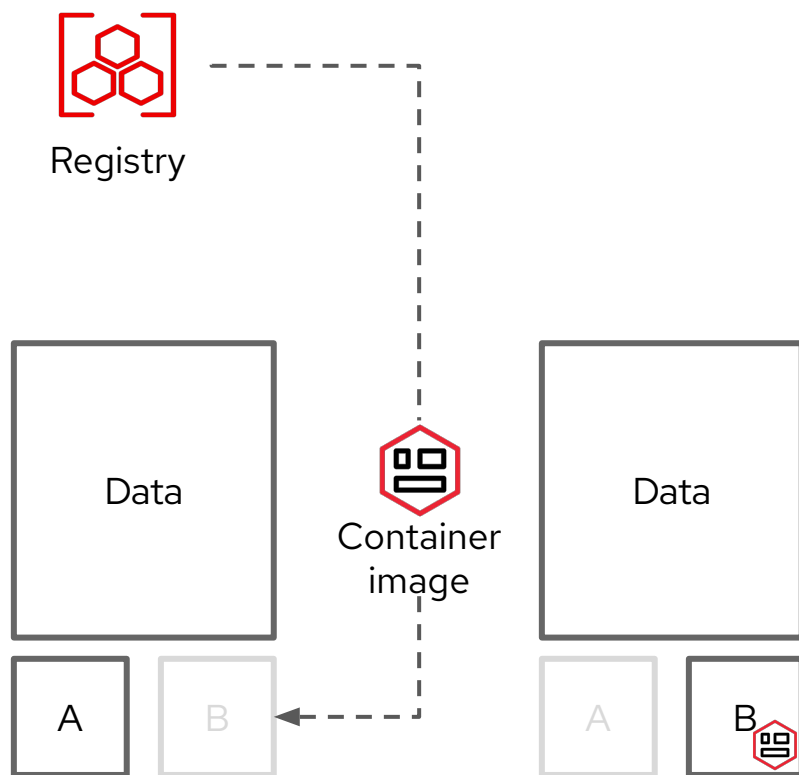
Primary contents:

- systemd, kernel, bootc
- rpm-ostree¹
- linux-firmware
- NetworkManager
- podman
- python
- Misc CLI tools: jq, sos

No cloud-init or virt agents

Bootc: Image-based updates perfected

Immutable by default - secure by design



交易式更新 (A → B model)

Bootc 使用 composefs 和 ostree 將容器映像轉換為主機上的根檔案系統。

升級 Roll forward or 退版 backwards

更新在背景暫存，並在系統重新啟動時套用。交易模型支援回滾，以提供額外的保障。

升級從未如此簡單

雖然存在一些限制，但 bootc 支援在 RHEL 的次要版本 (9.5 → 9.6) 和主要版本 (9.6 → 10.0) 之間切換。

Filesystem Layout

Similar to previous ostree setups - but better!!

建構時

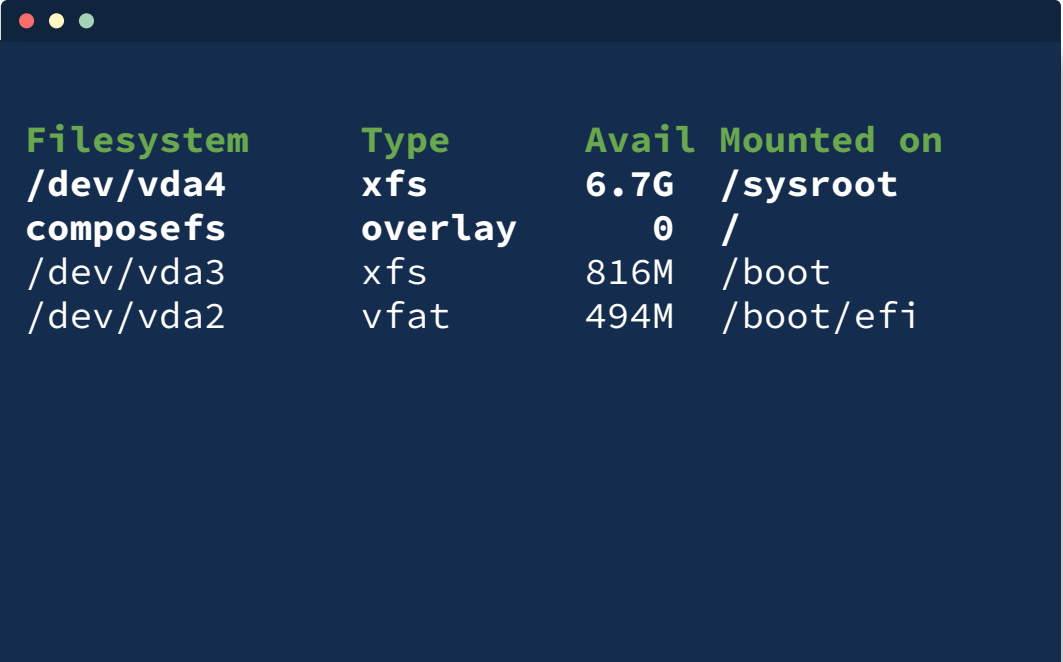
一切都是可寫的 e.g. /usr, /etc, /opt, ...

運行時

所有鏡像檔內容都是唯讀的

/var - 讀寫, 實例持久化。安裝後不更新

/etc - 讀寫, 類似 RHEL CoreOS 的三向合併。
機器本地狀態(主機名稱、靜態 IP)

A terminal window with a dark blue background and light green text. It displays a table of filesystem information. The table has four columns: Filesystem, Type, Avail, and Mounted on. The rows show the following data: /dev/vda4 (xfs, 6.7G, /sysroot), composefs (overlay, 0, /), /dev/vda3 (xfs, 816M, /boot), and /dev/vda2 (vfat, 494M, /boot/efi).

Filesystem	Type	Avail	Mounted on
/dev/vda4	xfs	6.7G	/sysroot
composefs	overlay	0	/
/dev/vda3	xfs	816M	/boot
/dev/vda2	vfat	494M	/boot/efi

應用案例

鏡像模式適合於哪裡？



Use Cases Blog



AI/ML 堆疊



1:1 應用程式/主機



邊緣設備



單獨的容器主機

完美控制核心、GPU
和加速器驅動程式、框
架、運行時等應用依賴
項的版本

將作業系統和應用程
式作為單一單元進行
管理

使用鏡像倉儲和自動
更新輕鬆管理

使用通用工具和流水
線建立容器化應用程
式和託管作業系統

Demo : RHEL Image mode

- ▶ <https://www.redhat.com/en/introduction-to-image-mode-for-red-hat-enterprise-linux-interactive-lab>
- ▶ 從 Containerfile 建構 Image 並上傳鏡像倉儲。
- ▶ 建立一個QCOW2格式的, 虛擬機格式。
- ▶ 運行 Image mode
- ▶ 升級

邁向AI平台新角色 RHEL AI

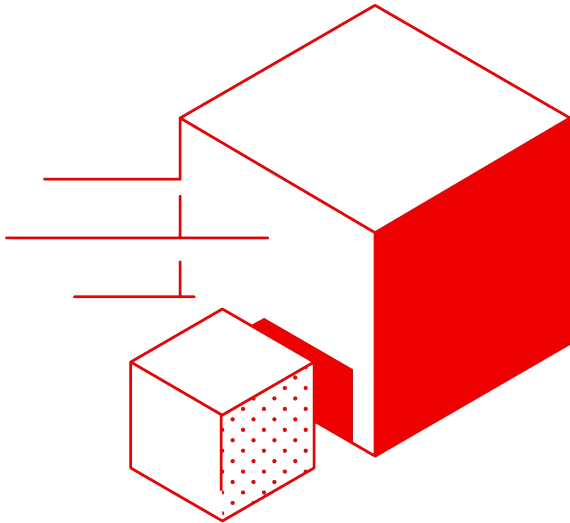
Ato Lin | 林嘉彥

Senior Solution Architect

Red Hat



推理是現實世界價值發生的地方



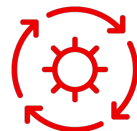
- ▶ 賦能使用者與模型互動的 AI 體驗
- ▶ 可跨硬體、模式和混合雲進行
- ▶ 透過實現預期業務成果，為 AI 計劃創造價值



基礎模型平台

無縫開發、測試和運行企業應用程式的大型語言模型 (LLM)。

Red Hat Enterprise Linux AI 推理伺服器



優化混合雲中的模型推理，以創建更快、更具成本效益的模型部署，並存取預先優化的模型庫。

InstructLab 模型對齊工具



可擴展、經濟高效的解決方案，用於增強 LLM 功能並使 AI 模型開發。

Granite 家族模型



開源授權的 LLM，根據 Apache-2.0 授權，訓練資料集和模型智慧財產賠償完全透明。

最佳化的可啟動模型執行程式



Granite 模型和 InstructLab 工具打包為可啟動的 RHEL 映像，包括 Pytorch/runtimes 和硬體優化 (NVIDIA、Intel 和 AMD)。

Granite 模型

一系列開放、高效、值得信賴的 AI 模型，加速企業 AI 的採用

Open

在 apache 2.0 許可下開源並可在 watsonx.ai、Hugging Face 和其他平台上使用。

Trusted

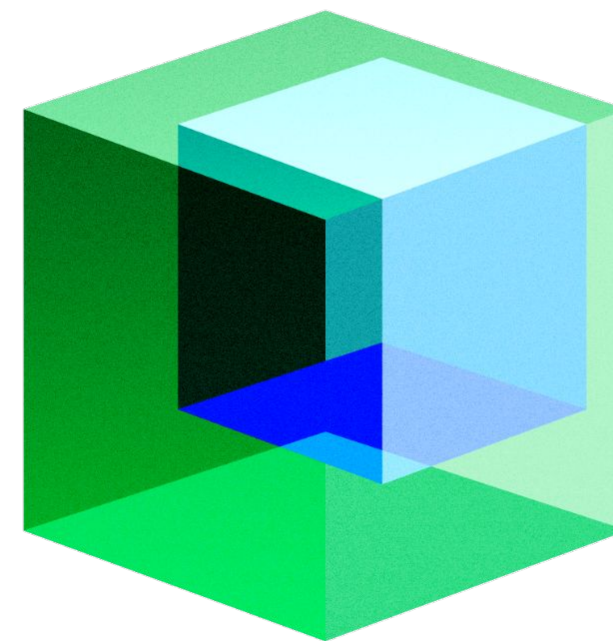
對與企業領域相關的可信賴且受管控的資料進行訓練。模型提供了智慧財產權的保障與支援。

Performant

多種較小、適合用途的模型，提供與類似尺寸模型相當的性能。

Cost-effective

提供更低的推理成本和更低的基礎設施託管成本。



基礎模型包括：程式碼、語言、時間序列、代理、透過 Granite Guardian 配套模型實現的安全性，甚至地理空間數據

InstructLab vs 其他模型對齊方法

RAG

Retrieval Augmented Generation

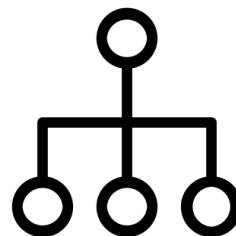


透過從外部來源檢索相關資訊來增強 Gen AI 模型產生的文本，提高模型回應的準確性和深度。

NEW

InstructLab

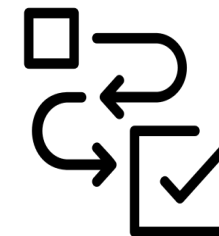
Large-scale Alignment for chatBots



利用分類法指導的合成資料生成流程和多階段調整框架來提升模型效能。

Fine tuning

Fine Tuning

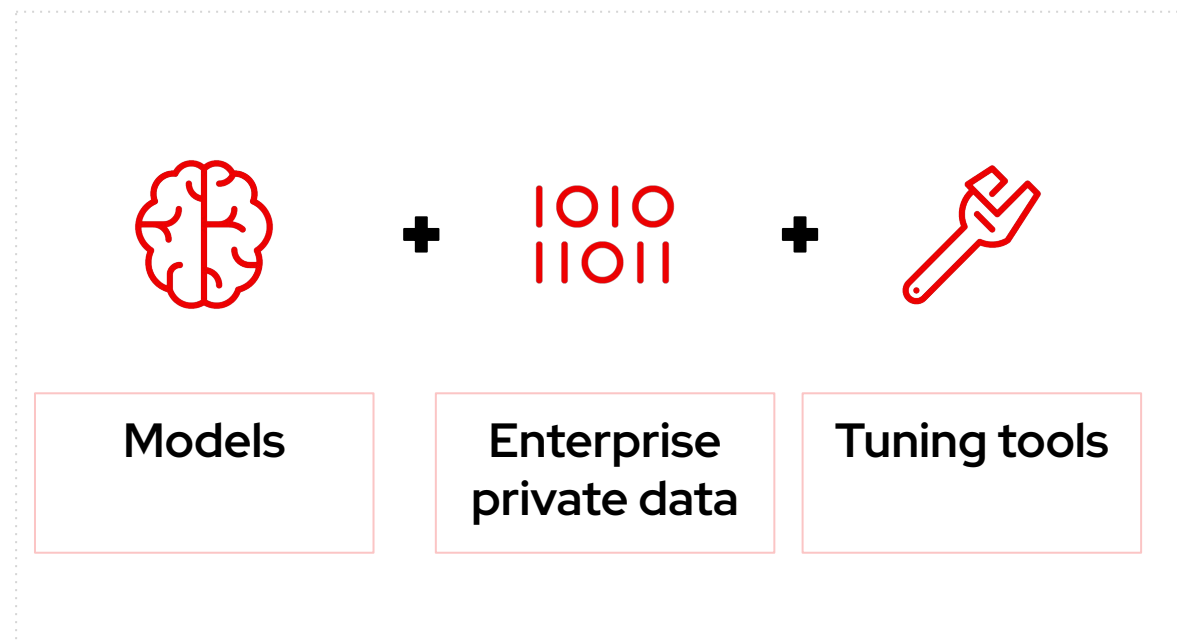


針對特定任務或資料調整預訓練模型，無需完全重新訓練即可提高其在專門應用中的效能和準確性。

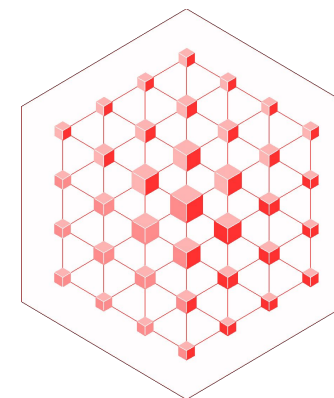
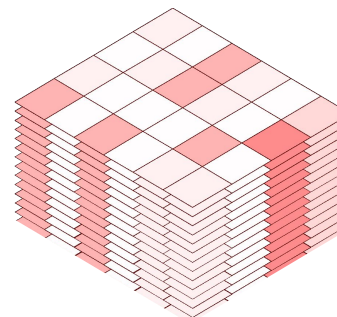
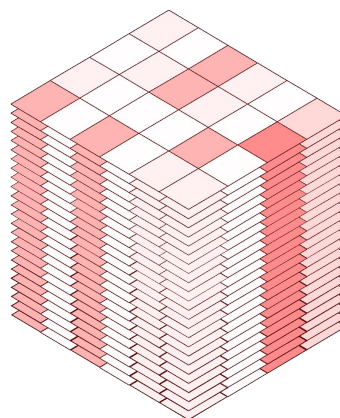
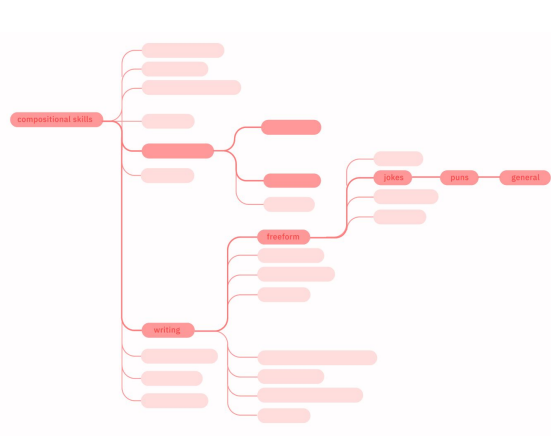
InstructLab 提供更**易於存取的微調** 並**補充 RAG** (RAFT pattern)

使用 **InstructLab** 客製化更小、更適合用途的模型，以建立高效、經濟的解決方案。

InstructLab 方法以極低的成本降低了使用企業私有資料客製化模型的複雜性。



LAB (Large-scale Alignment for ChatBots) Method



分類法 skill & knowledge

分層分類, 分類成技能或知識, 每個新知需提供至少五個問與答

生成綜合數據 + 老師模式

教師模型產生包含整個分類中數百萬個問題和答案的「課程」。

生成綜合數據 + 評論員模式

批評模型過濾問題的正確性和品質。掃描合成資料以尋找違禁資料。

基於學生模型的技能 and 知識培訓

使用新穎的訓練方法對學生模型進行課程訓練。

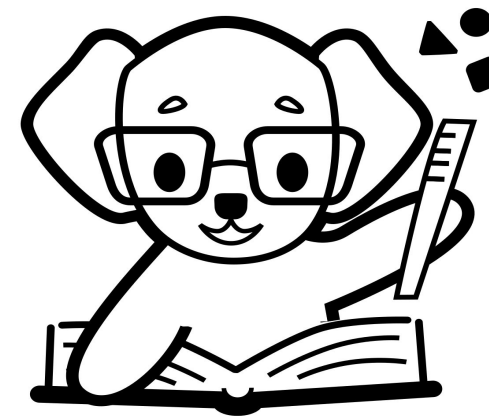
<https://arxiv.org/abs/2403.01081>

InstructLab 主要指令

InstructLab

主要指令

- ilab model download -> 模型下載
- ilab data generate -> 生成學習數據
- ilab model train -> 訓練模型
- ilab serve -> 提供模型服務
- ilab chat -> 使用聊天機器人



Red Hat 生成式 AI 產品



InstructLab

開源工具

透過桌機或筆電規模的資源，搭配小型數據集進行學習與實驗



Laptop / desktop



Red Hat
Enterprise Linux AI

小規模

生產級模型訓練，採用全合成數據生成，並結合教師模型與評論員模型。提供 CLI 工具與模組化組件，實現高效訓練與部署。



Server / VM



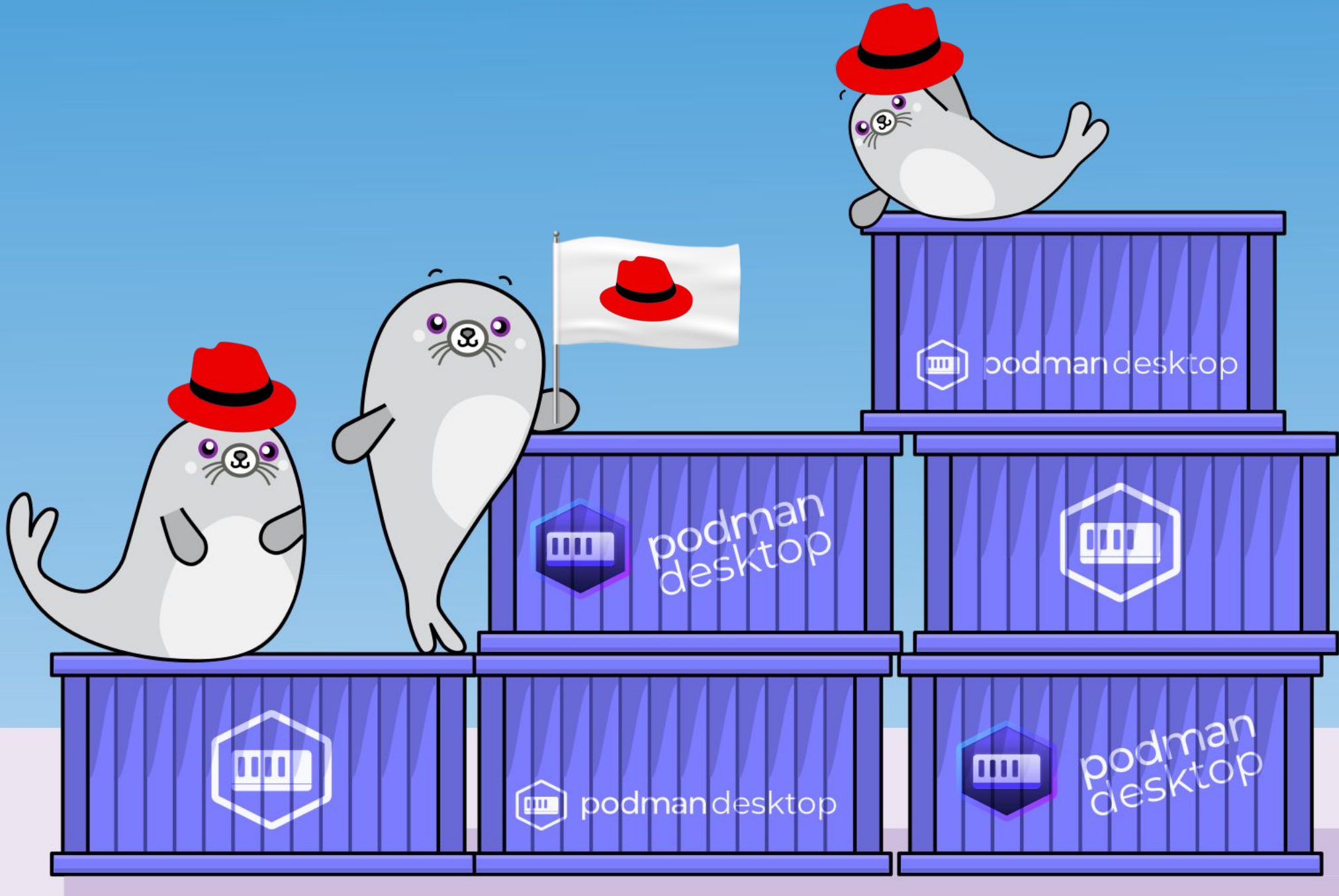
Red Hat
OpenShift AI

大規模

如同 RHEL AI，具備企業級模型訓練能力，利用 Kubernetes 的完整擴展能力、自動化與 MLOps 服務來強化 AI 部署與管理。



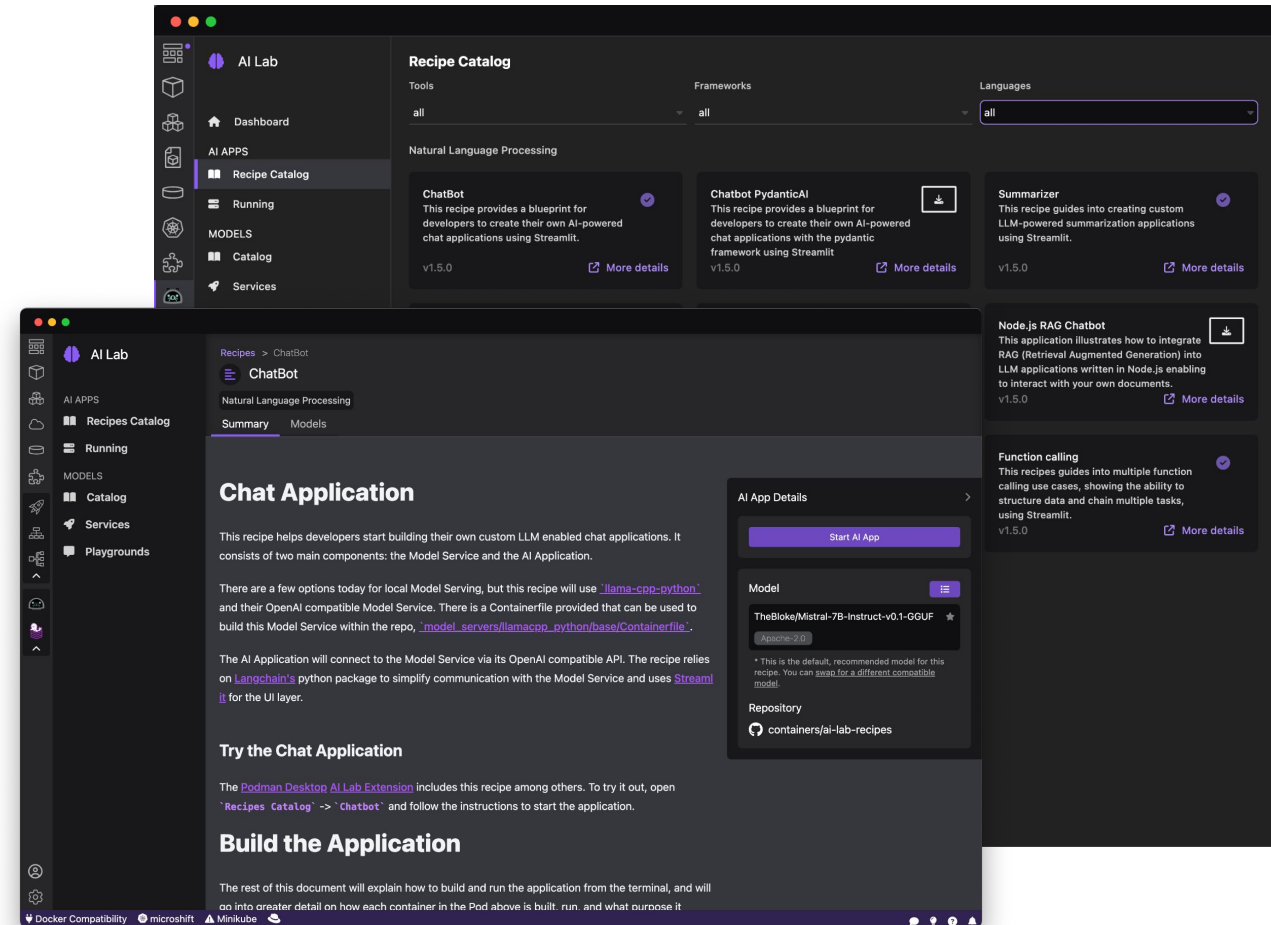
Cluster



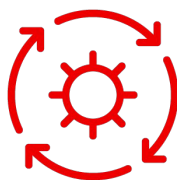
補充: Podman GenAI 菜單

從 GenAI 可以幫助實現的目標中獲得靈感

- ▶ AI 範例的配方和不同類別
- ▶ 示範將LLM整合進應用程式
- ▶ 準備隨時可以部署測試AI應用範例
- ▶ 可嘗試不同的模型
- ▶ 說明如何部署到生產環境, 例如 Kubernetes/OpenShift



Red Hat Enterprise Linux AI 的價值



**增加
效率**

使用由 Red Hat 提供全面支援和保障
的企業級**開源 Granite 模型**



**簡化
經驗**

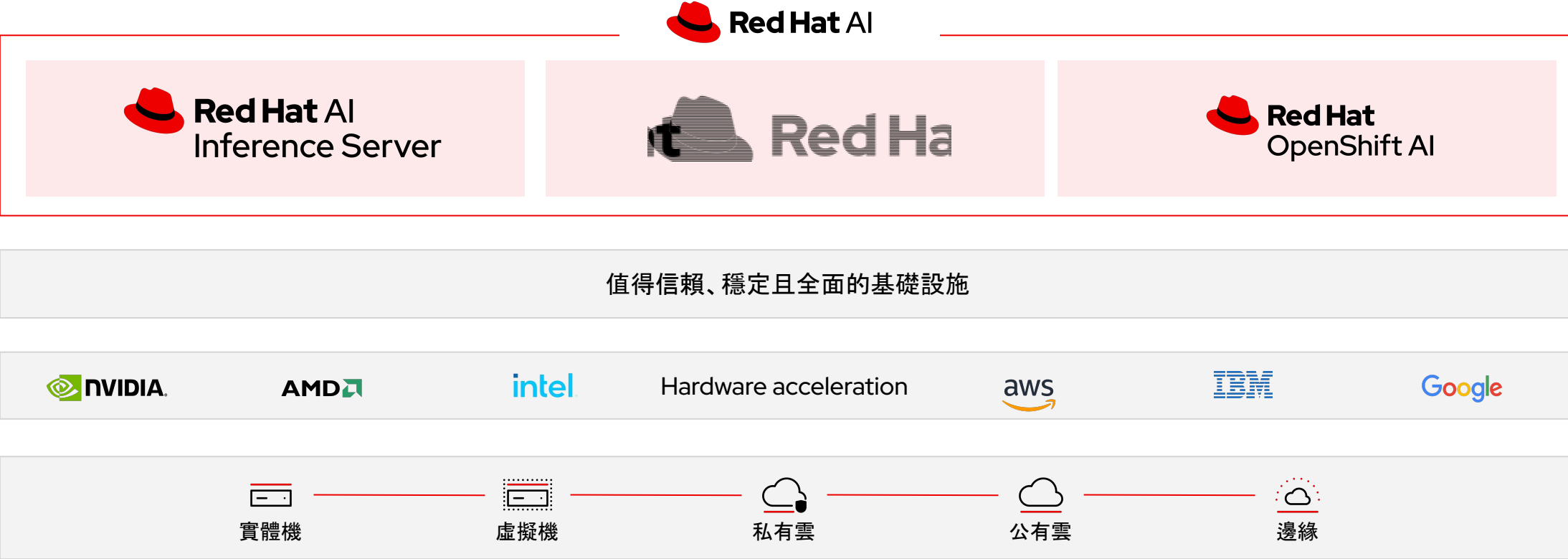
InstructLab 模型對齊工具可協助開發
人員和領域專家更有效地客製化模型



**可靈活部署至
任何地方**

**透過決定在何處訓練和部署 AI 模型
來控制成本**，從而降低風險並創造價
值

生成式與預測式 AI



* NVIDIA GPUs fully supported in Red Hat AI. AMD Instruct \$ Intel Gaudi in Preview, AWS Inferentia/Neuron, Google TPU, IBM AIU are on our roadmap






Thank you

Red Hat is the world's leading provider of enterprise open source software solutions. Award-winning support, training, and consulting services make Red Hat a trusted adviser to the Fortune 500.

 [linkedin.com/company/red-hat](https://www.linkedin.com/company/red-hat)

 [facebook.com/redhatinc](https://www.facebook.com/redhatinc)

 [youtube.com/user/RedHatVideos](https://www.youtube.com/user/RedHatVideos)

 twitter.com/RedHat

