Elasticsearch AI 與 ML 的說明與運用

講師:陳俊佑
集先鋒科技有限公司

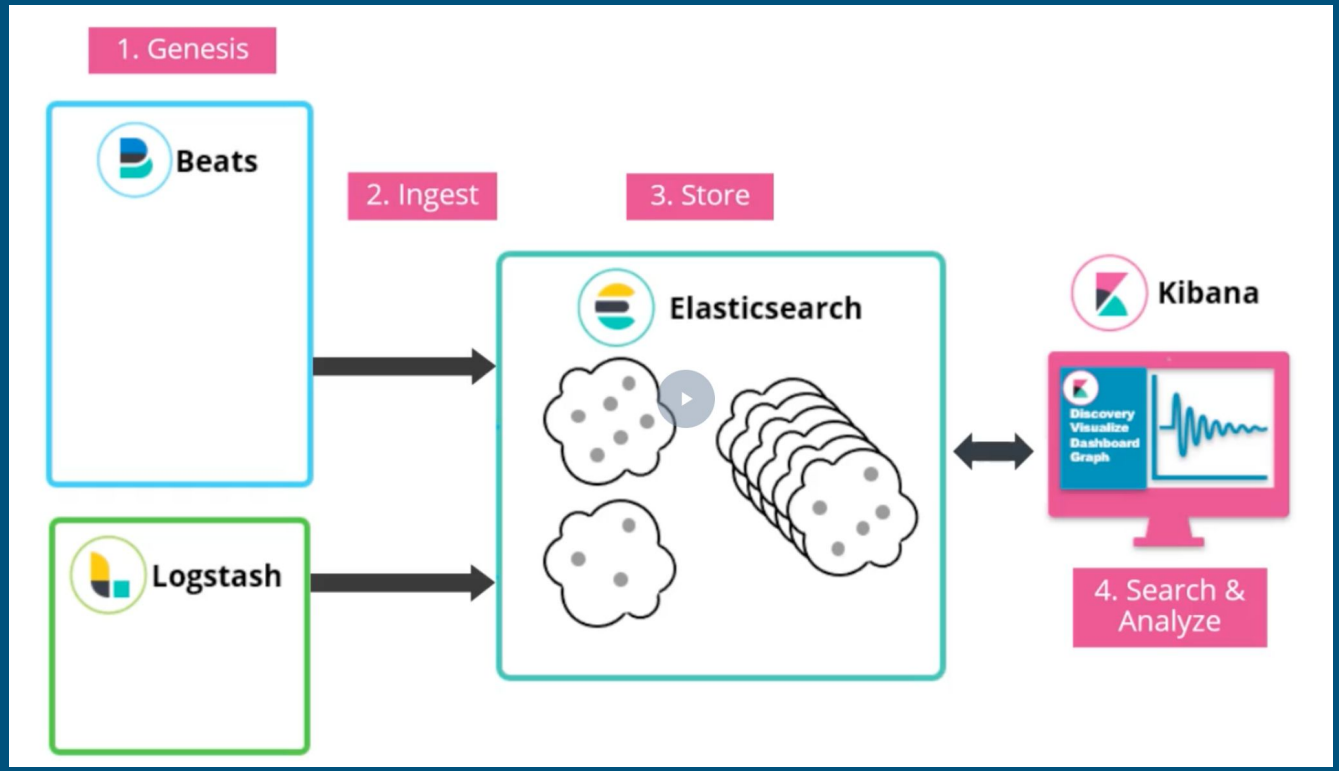
Elasticsearch 基礎說明

Elasticsearch 是什麼?

ElasticSearch是一款非常強大的、基於Lucene的開源搜尋及分析引擎，可以幫助你從海量資料中，快速找到相關的資料資訊

分散式儲存	高拓展性	高安全性	高效檢索
多語種查詢	多租戶技術	Schema Free	RESTful API

什麼是 ELK?


1. Genesis
Beats
Logstash
2. Ingest
3. Store
Elasticsearch
Kibana
Discovery
Visualize
Dashboard
Graph
4. Search &
Analyze

Elasticsearch 排名

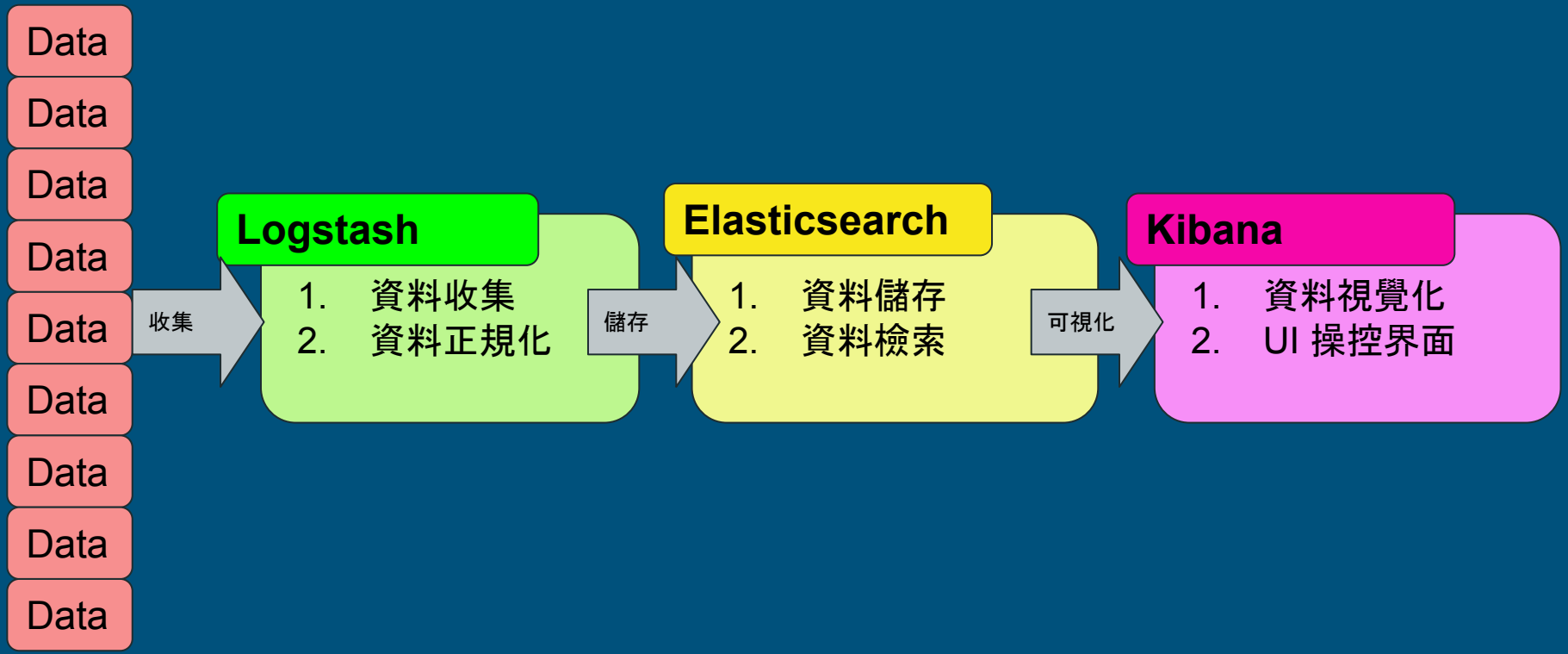
include secondary database models — 21 systems in ranking, December 2020

Rank Dec 2020	Rank Nov 2020	Rank Dec 2019	DBMS	Database Model	Score Dec 2020	Score Nov 2020	Score Dec 2019
1.	1.	1.	Elasticsearch	Search engine, Multi-model	152.49	+0.94	+2.24
2.	2.	2.	Splunk	Search engine	87.00	-2.71	-3.53
3.	3.	3.	Solr	Search engine	51.24	-0.57	-5.98
4.	4.	4.	MarkLogic	Multi-model	10.94	-0.16	-1.53
5.	5.	↑ 8.	Algolia	Search engine	7.83	+0.45	+3.15
6.	6.	6.	Microsoft Azure Search	Search engine	6.85	+0.05	+0.56
7.	7.	↓ 5.	Sphinx	Search engine	6.32	-0.03	-0.30
8.	8.	↓ 7.	ArangoDB	Multi-model	5.51	+0.14	+0.64
9.	9.	9.	Amazon CloudSearch	Search engine	3.06	+0.29	-0.15
10.	10.	↑ 11.	Virtuoso	Multi-model	2.58	+0.05	-0.05
11.	11.	↑ 12.	Xapian	Search engine	1.01	+0.17	+0.25
12.	12.	↑ 13.	CrateDB	Multi-model	0.90	+0.08	+0.28
13.	13.		Alibaba Cloud Log Service	Search engine	0.40	-0.02	
14.	14.	14.	SearchBlox	Search engine	0.40	+0.01	+0.10
15.	15.	↑ 16.	Manticore Search	Search engine	0.07	-0.03	+0.01
16.	16.	↑ 20.	Weaviate	Search engine	0.05	-0.03	+0.05
17.	↑ 18.	17.	Exorbyte	Search engine	0.03	-0.02	-0.03
18.	↑ 19.	↑ 19.	FinchDB	Multi-model	0.03	0.00	+0.00
19.	↓ 17.	↓ 15.	searchxml	Multi-model	0.01	-0.05	-0.12
20.	20.	↑ 21.	Indica	Search engine	0.00	±0.00	±0.00
20.	20.		Rizhiyi	Search engine, Multi-model	0.00	±0.00	

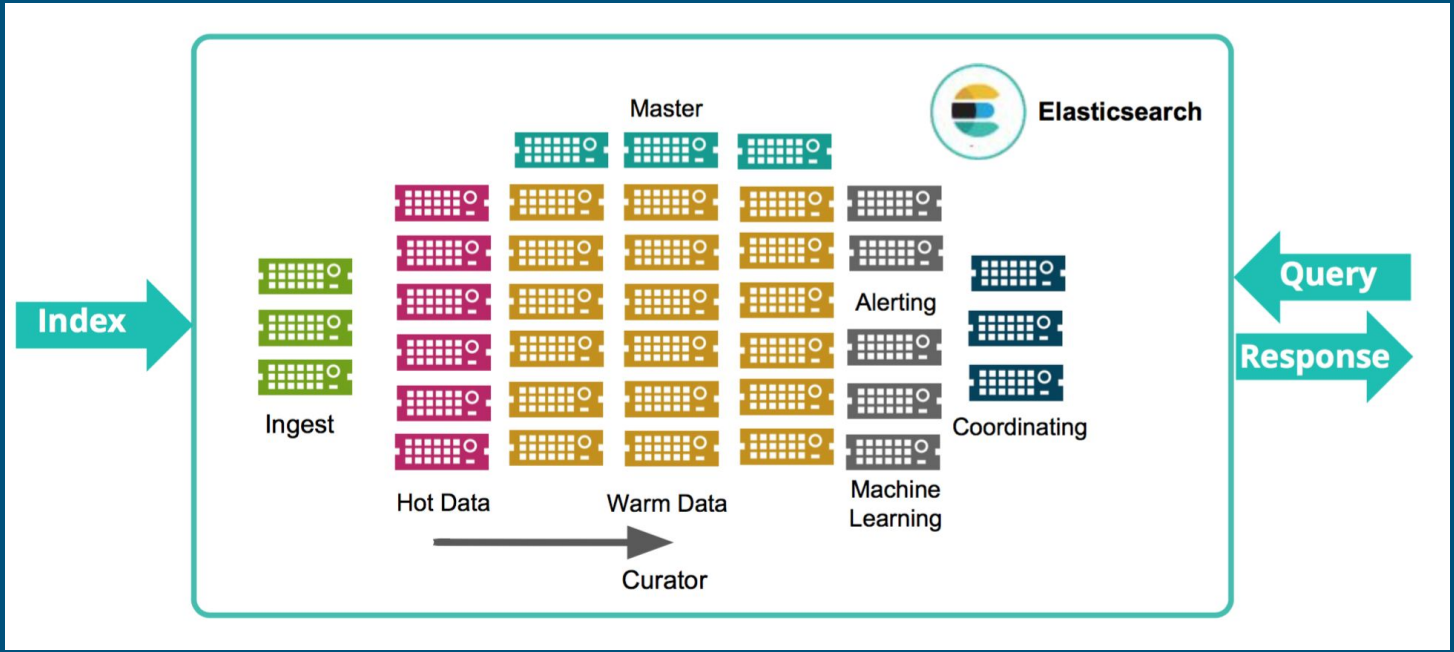
include secondary database models — 26 systems in ranking, August 2024

Rank Aug 2024	Rank Jul 2024	Rank Aug 2023	DBMS	Database Model	Score Aug 2024	Score Jul 2024	Score Aug 2023
1.	1.	1.	Elasticsearch	Search engine, Multi-model	129.83	-0.99	-10.09
2.	2.	2.	Splunk	Search engine	96.10	+3.18	+7.12
3.	3.	3.	Apache Solr	Search engine, Multi-model	36.30	-2.58	-11.51
4.	4.	4.	OpenSearch	Search engine, Multi-model	16.47	-0.17	+4.03
5.	5.	↑ 8.	Sphinx	Search engine	5.97	+0.01	-0.10
6.	↑ 7.	6.	Algolia	Search engine	5.77	+0.20	-1.45
7.	↓ 6.	7.	Microsoft Azure AI Search	Search engine, Multi-model	5.73	+0.05	-0.38
8.	8.	↓ 5.	MarkLogic	Multi-model	4.41	-0.02	-4.03
9.	9.	9.	Virtuoso	Multi-model	3.85	-0.12	-0.94
10.	10.	10.	ArangoDB	Multi-model	3.40	-0.02	-0.97
11.	11.	11.	Coveo	Search engine	2.05	+0.07	-0.50
12.	12.	12.	Amazon CloudSearch	Search engine	1.85	+0.06	-0.32
13.	13.	13.	Meilisearch	Search engine	0.99	+0.02	-0.40
14.	↑ 16.	↑ 17.	Typesense	Search engine	0.78	+0.08	+0.02
15.	↓ 14.	↓ 14.	Xapian	Search engine	0.73	-0.02	-0.47
16.	↓ 15.	↓ 15.	CrateDB	Multi-model	0.69	-0.03	-0.20
17.	17.	↑ 18.	Vespa	Multi-model	0.59	-0.05	+0.03
18.	18.	↑ 20.	Alibaba Cloud Log Service	Search engine	0.42	+0.05	+0.01
19.	↑ 20.	19.	Marqo	Search engine	0.35	+0.05	-0.07
20.	↓ 19.	↓ 16.	SearchBlox	Search engine	0.32	-0.05	-0.47

常規 Elasticsearch 資料處理流程

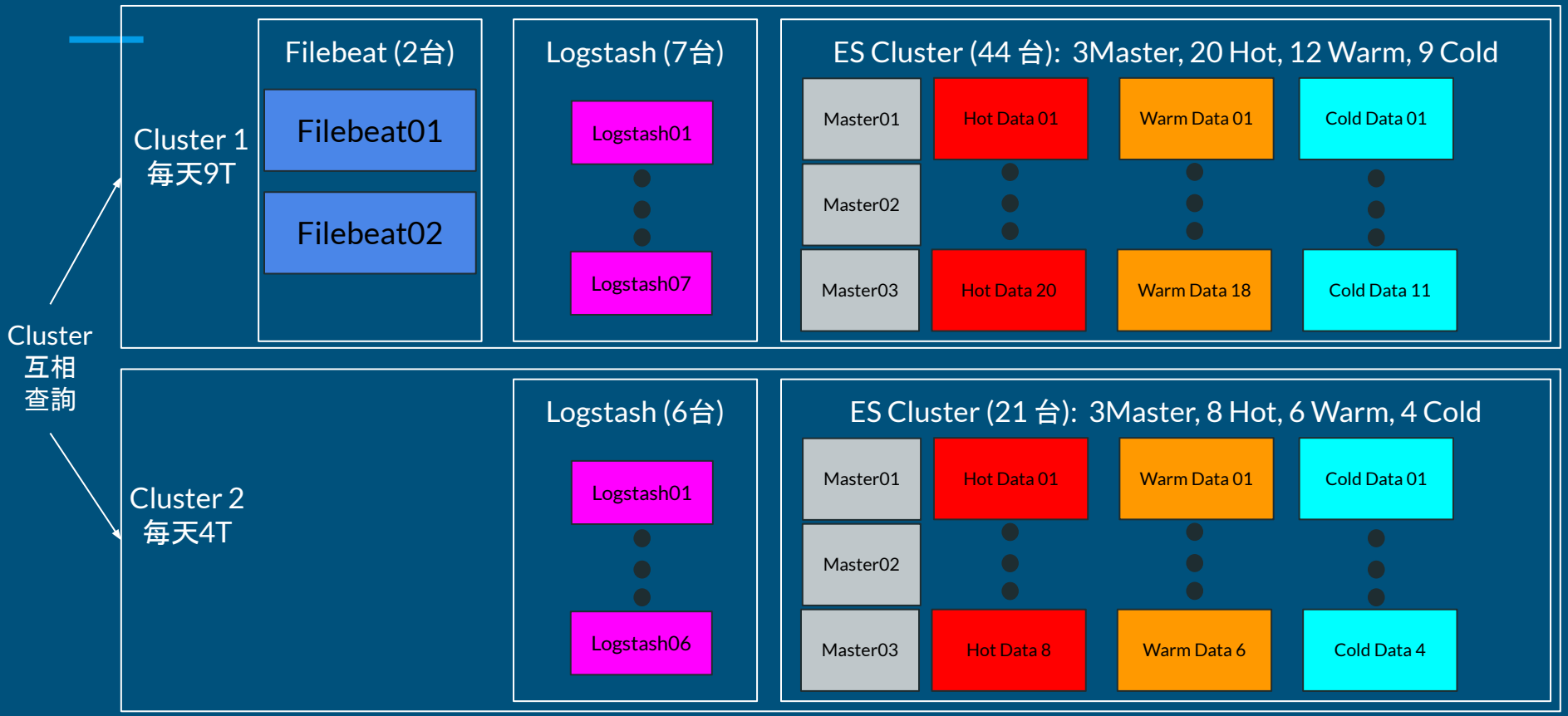


Elasticsearch 集群架構



Elasticsearch 集群架構-範例 每天13T的系統架構

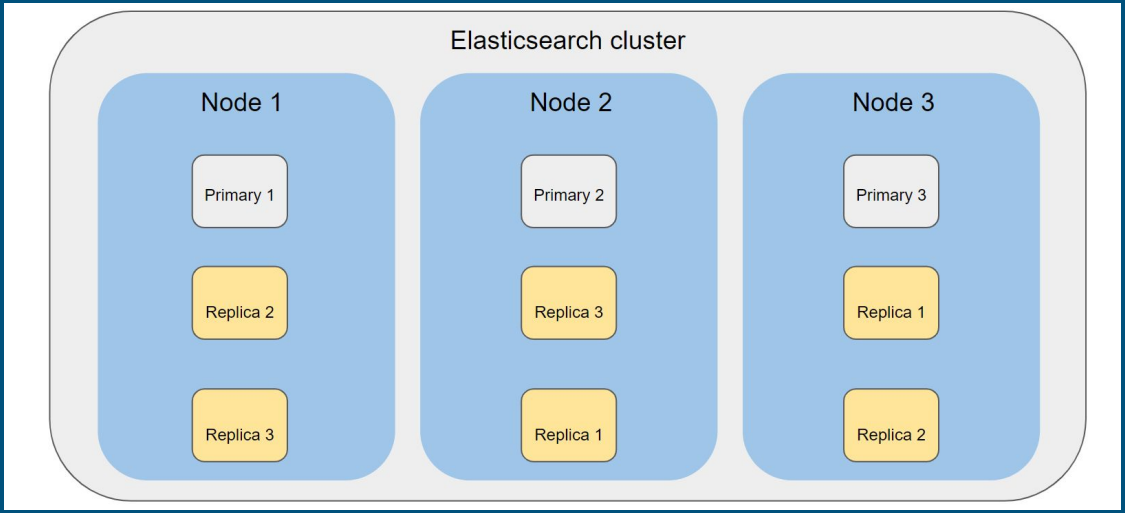
Hot 保留30分钟，Warm 保留3天，Cold 保留30天



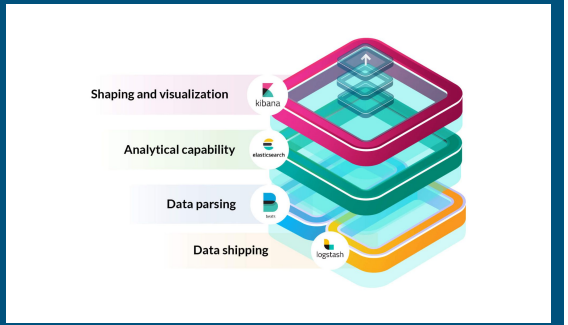
Elasticsearch 分片與副本

1. 分散式儲存採用了多副本備份機制，其中之一不可用時，可用其他代替
2. primary shard 遺失時，replica shard 就可以被 promote 成 primary shard 來保持資料完整性

範例: 1 Cluster，3 Node，3 Primary Shard，2 Replica



Elastic stack 中的 beats 大家族



Beats 系列

全品類採集器，搞定所有數據類型。

Filebeat
日誌文件
→

Metricbeat
指標
→

Packetbeat
網絡數據
→

Winlogbeat
Windows 事件日誌
→

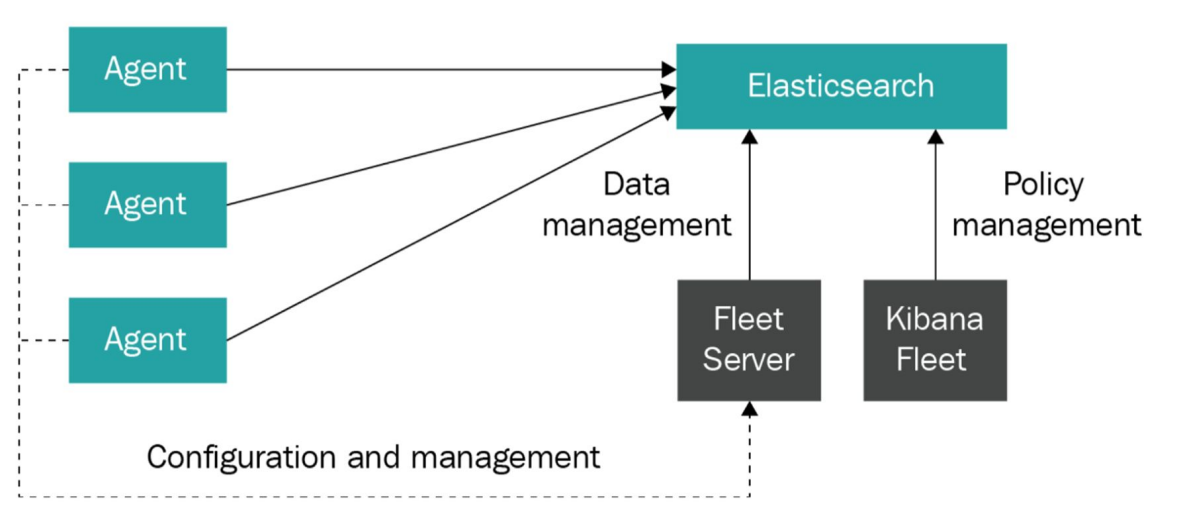
Auditbeat
審計數據
→

Heartbeat
運行時間監控
→

Functionbeat
無需服務器的採集器
→

Beats 是一個免費且開放的平台，集合了多種單一用途數據採集器。它們從成百上千或成千上萬台機器和系統向 Logstash 或 Elasticsearch 發送數據

通過 Fleet 輕鬆整合資料



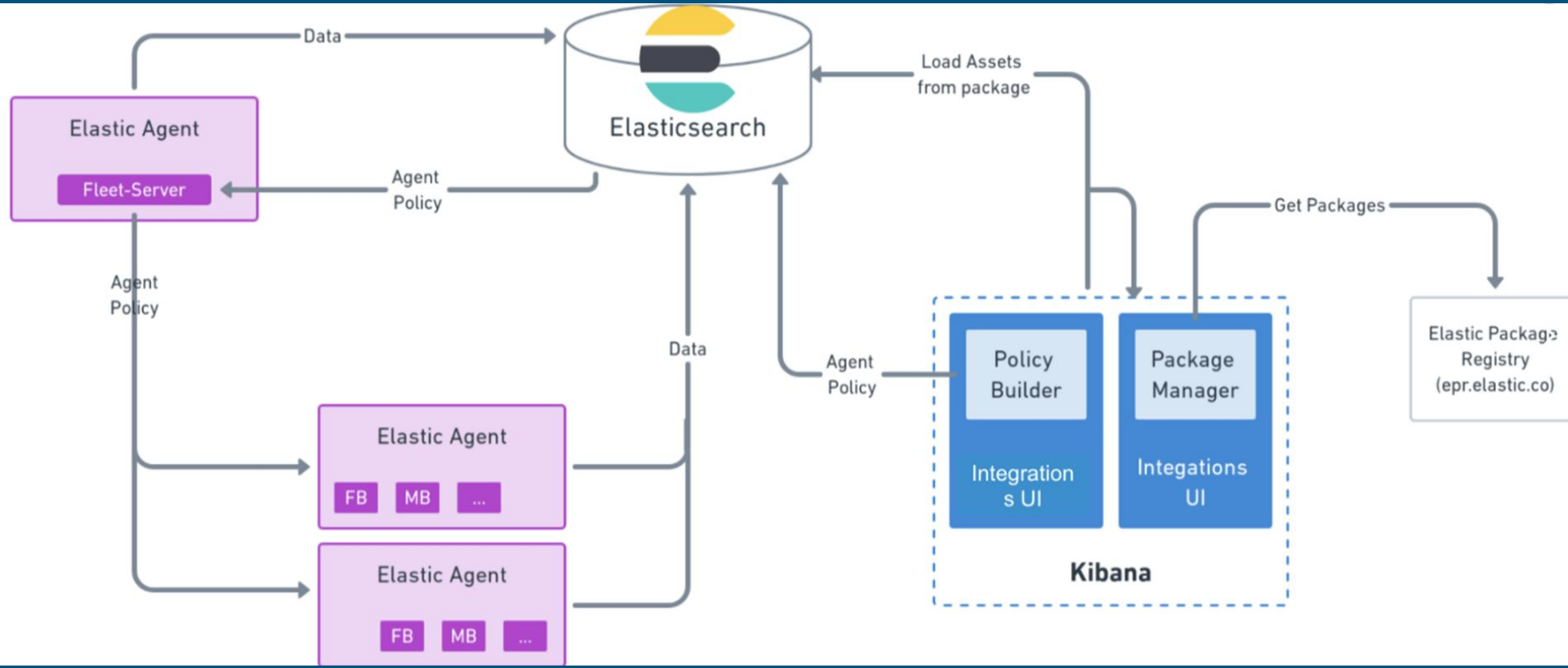
Fleet: Agent 的中控平台，負責管理與查看現有環境中的 Elastic Agent 配置，專用的 Elastic Agent 通信主機運行

Fleet UI: Kibana 上新的頁面，供使用者載入和配置代理、管理載入資料以及管理整個環境中的代理

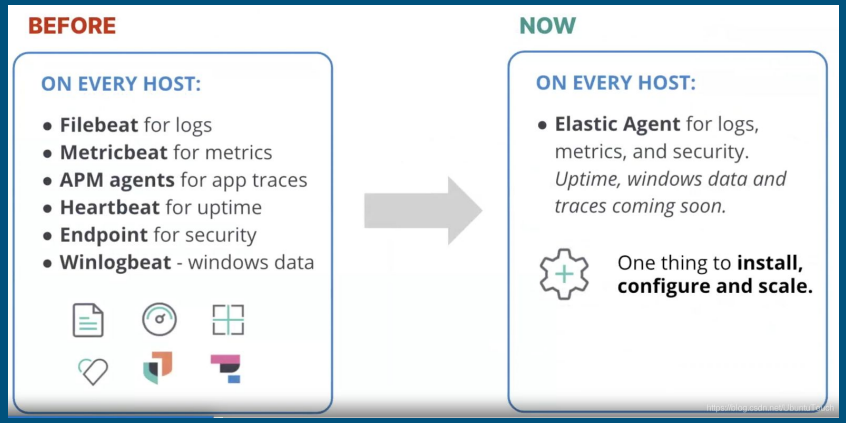
Fleet Server: 後端元件，環境中的彈性代理可以連接到該元件，以檢索代理策略、更新和管理命令。

Elastic Agent: 將資料直接發送到目標 Elasticsearch 集群

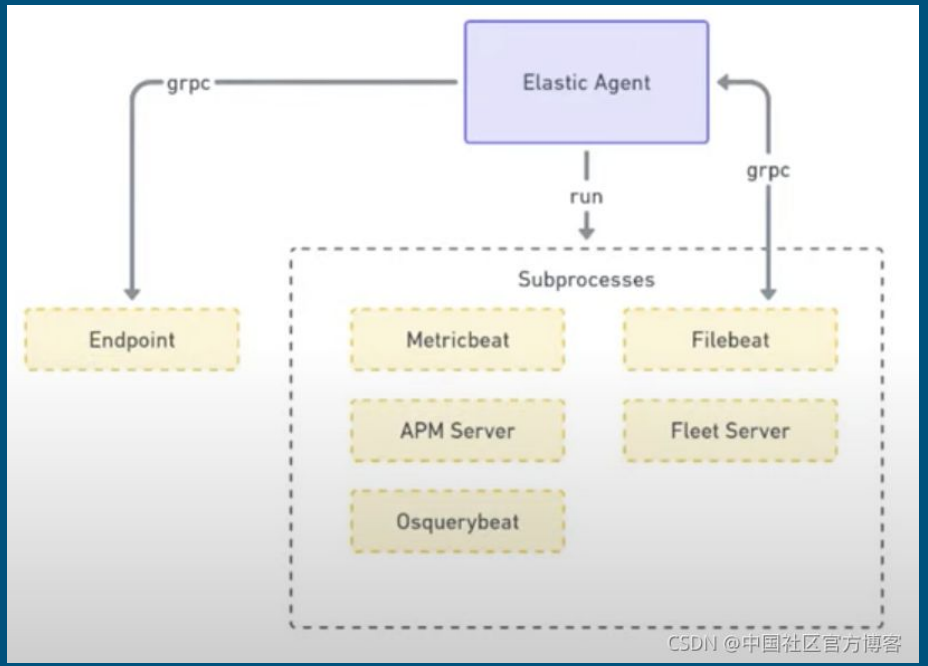
Fleet & Elastic Agent 組件說明



Fleet&Elastic Agent 與 Beat 的區別



1. beat現在被統稱為 Elastic Agents
2. Elastic Agent 本質為 beat 的主管



Fleet & Elastic Agent 的優點

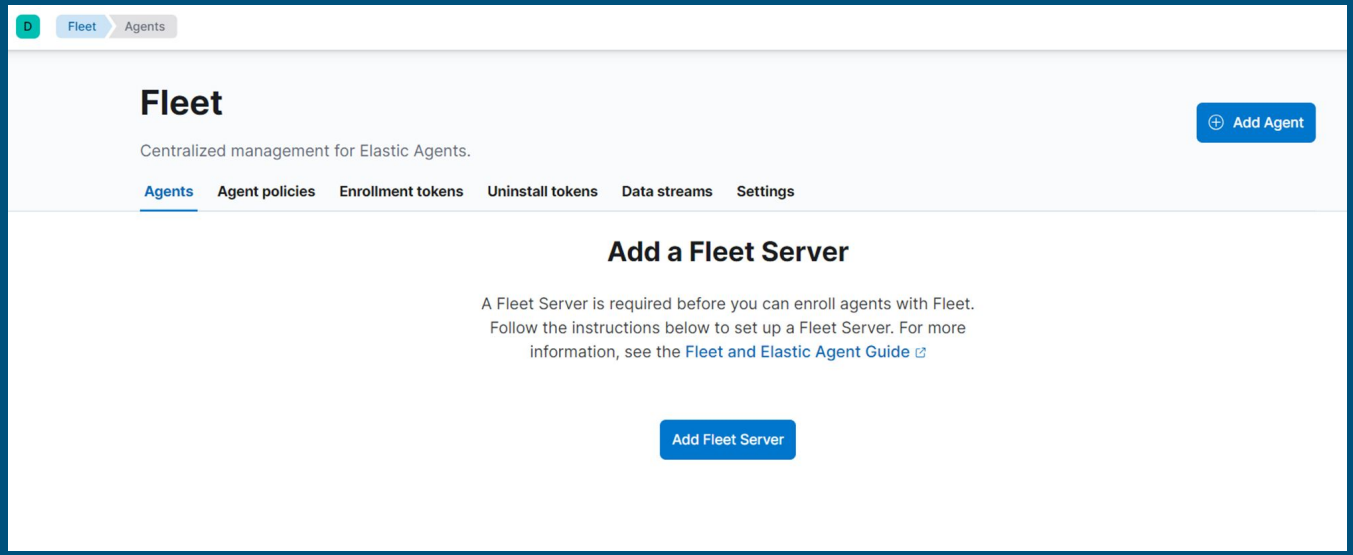
大量部署: 隨著原本 Elastic Agents 的數量增加, 越來越需要一個中控平台對 Elastic Agents 進行統一管理與配置

安全性: 相對 beat 所需的權限更少

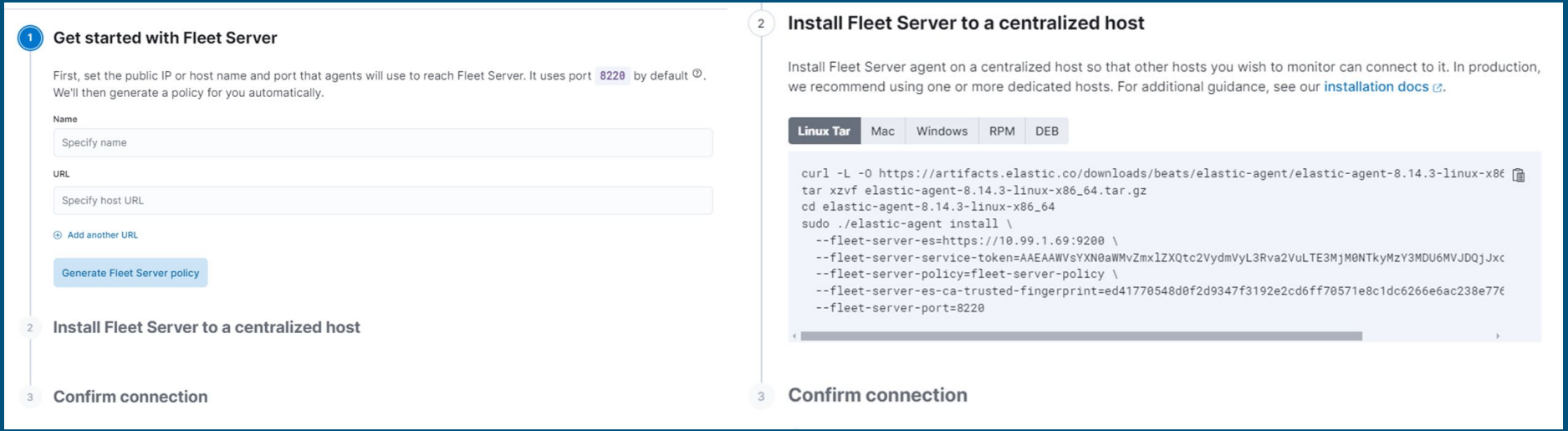
靈活性: 部署方便快捷

簡便性: 有著大量已經預先處理好的模版，目前以有超過300種資料的模版，收取特定資料時可直接以 elastic common schema (ecs) 格式進行儲存

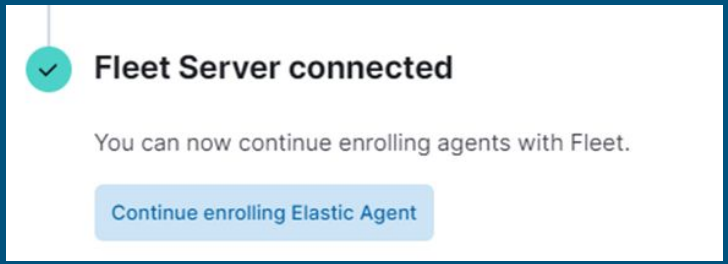
Fleet 安裝


D
Fleet
Agents
Fleet
Centralized management for Elastic Agents.
Add Agent
Agents
Agent policies
Enrollment tokens
Uninstall tokens
Data streams
Settings
Add a Fleet Server
A Fleet Server is required before you can enroll agents with Fleet.
Follow the instructions below to set up a Fleet Server. For more
information, see the Fleet and Elastic Agent Guide
Add Fleet Server

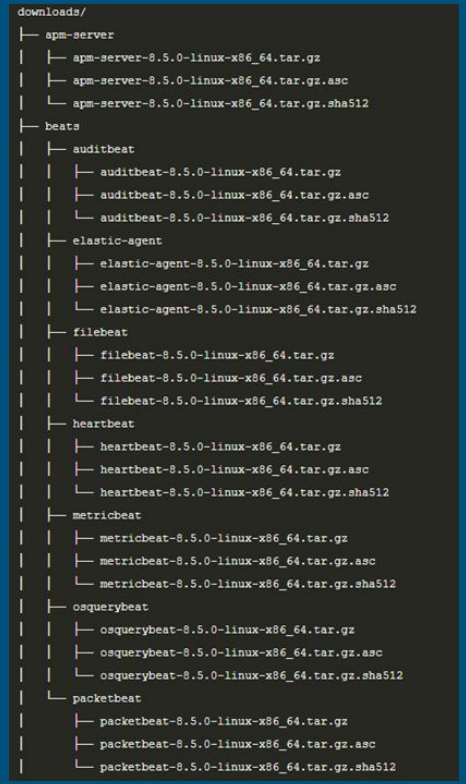
Fleet 安裝



Url 必須使用 https

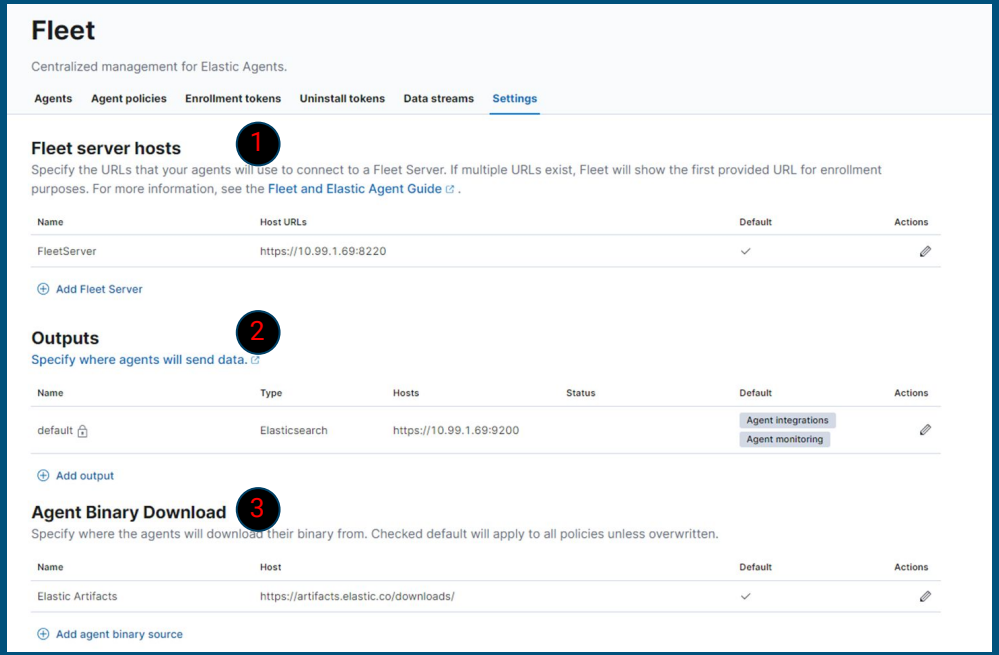


Fleet 離線安裝與 Beats



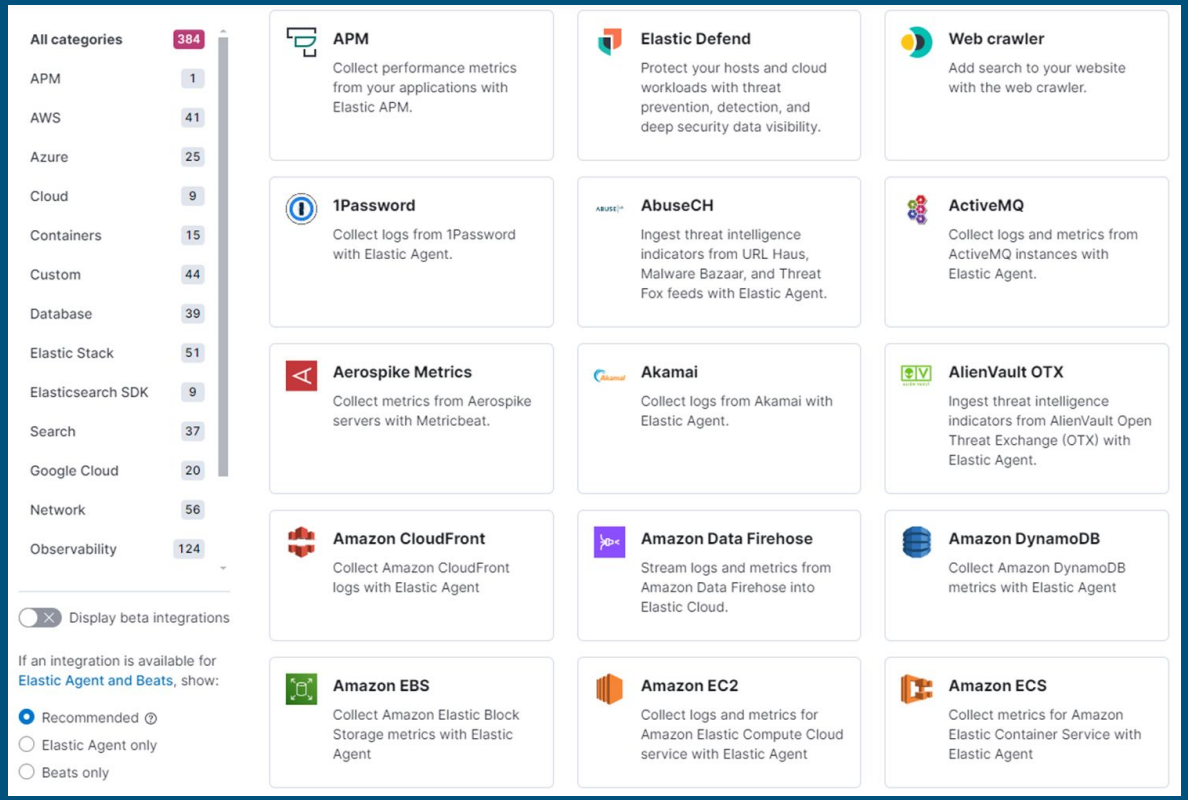
1. 離線安裝 Fleet Server 時需將所有相關組件提前下載
2. 內容包含全部的 beats

檢查 Fleet Settings



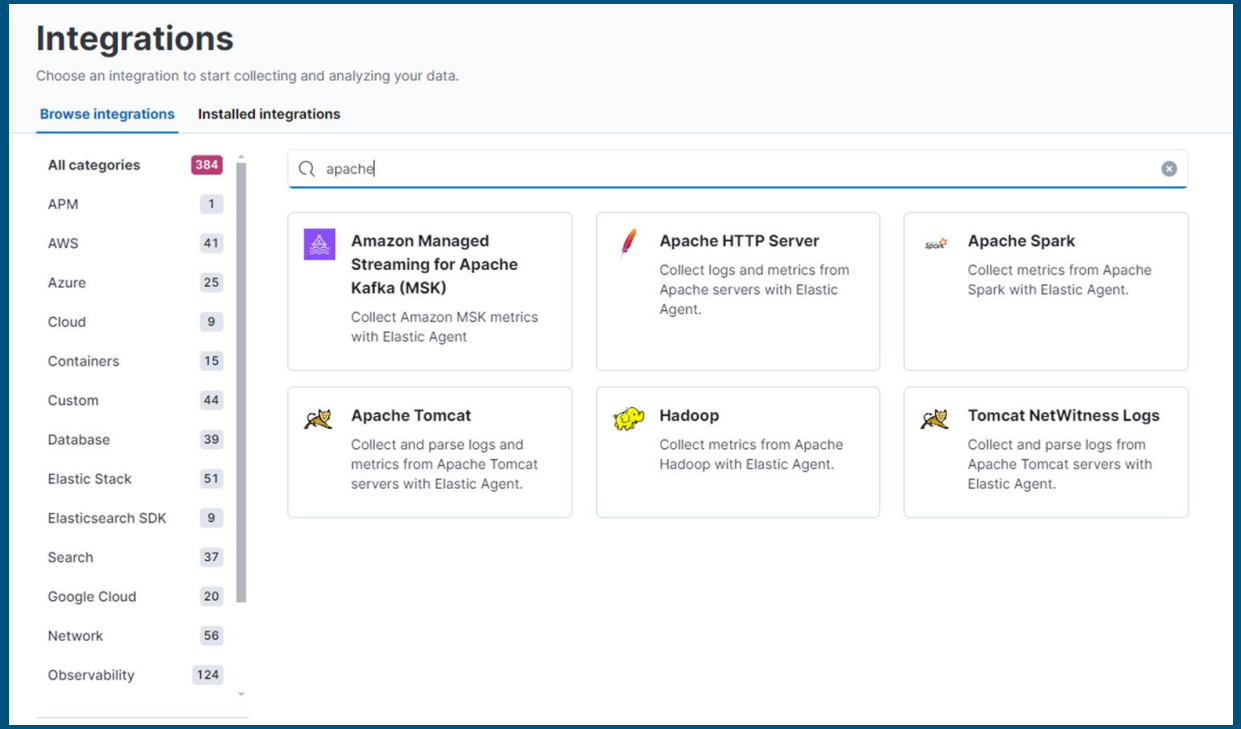
1. Fleet Server 的 host
2. 輸出的 Elasticsearch
3. 下載路徑 (非離線下載不需要)

種類繁多的 Elastic Agents



1. 通過 intergration 界面輕鬆選擇查看所有的可採集資料
2. 目前 Elastic Agent 已對384種資料進行預處理

通過 Intergration 輕鬆整合資料 - Apache 為例

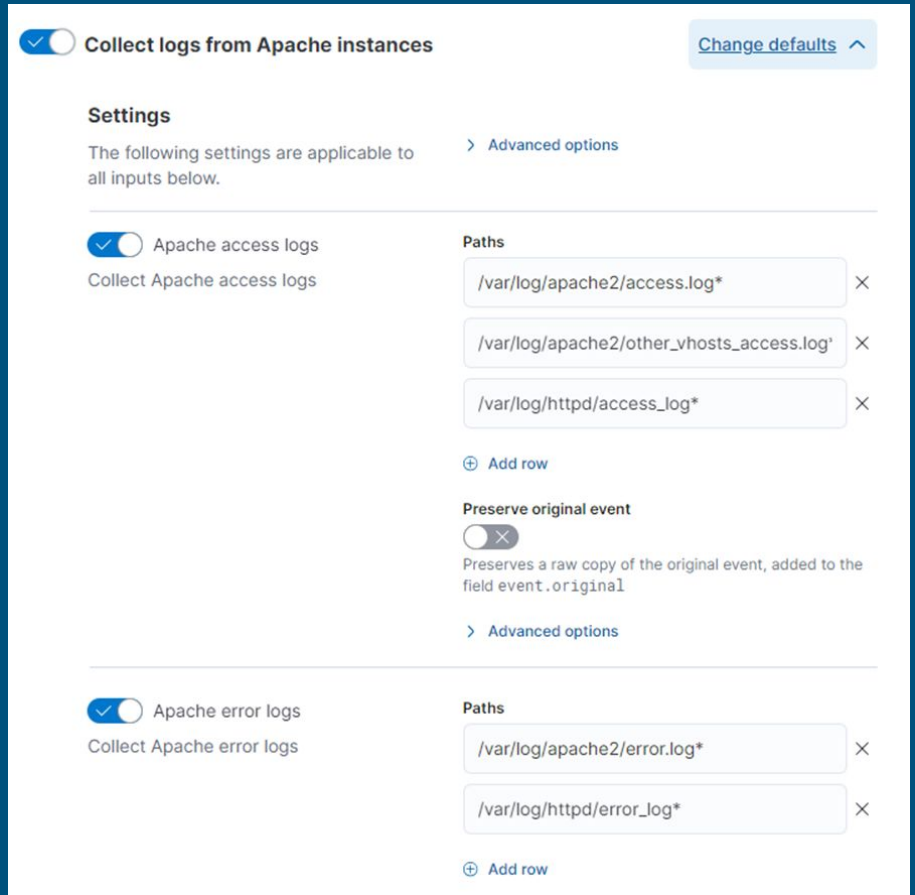


1. 選擇 Apache HTTP Server 來抓取 Apache 日誌

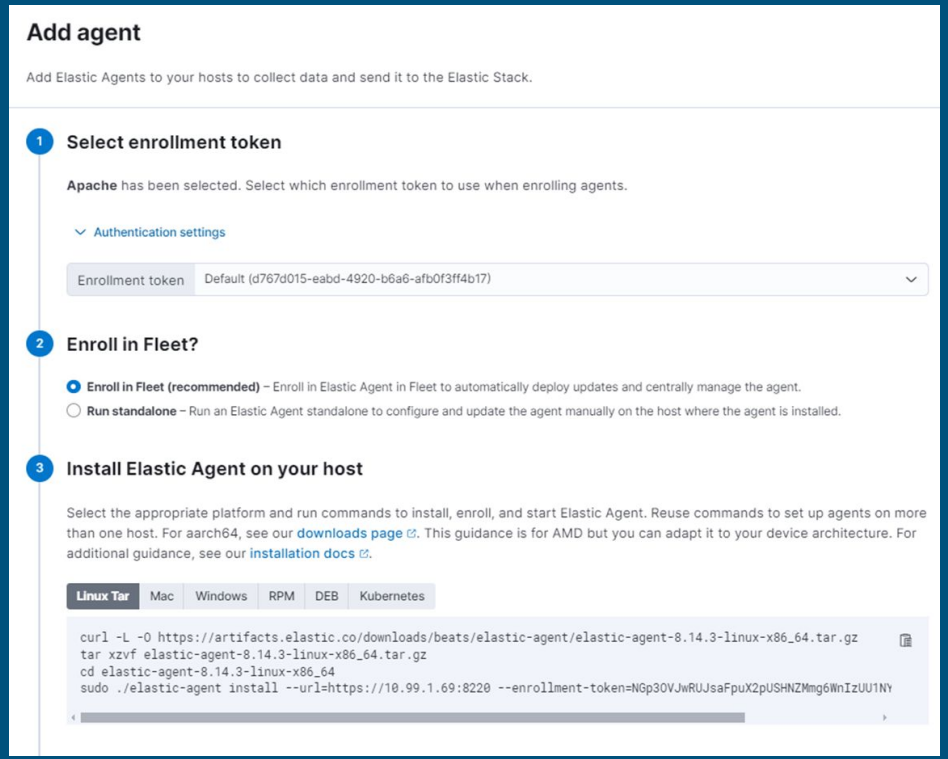
修改設定

針對 Apache Log 僅需修改文件路徑即可

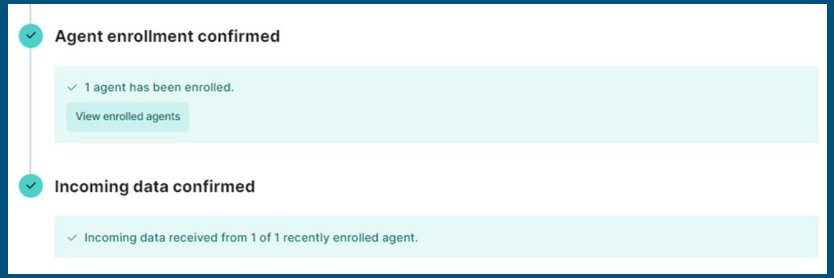
可針對 access 和 error log 進行拆解



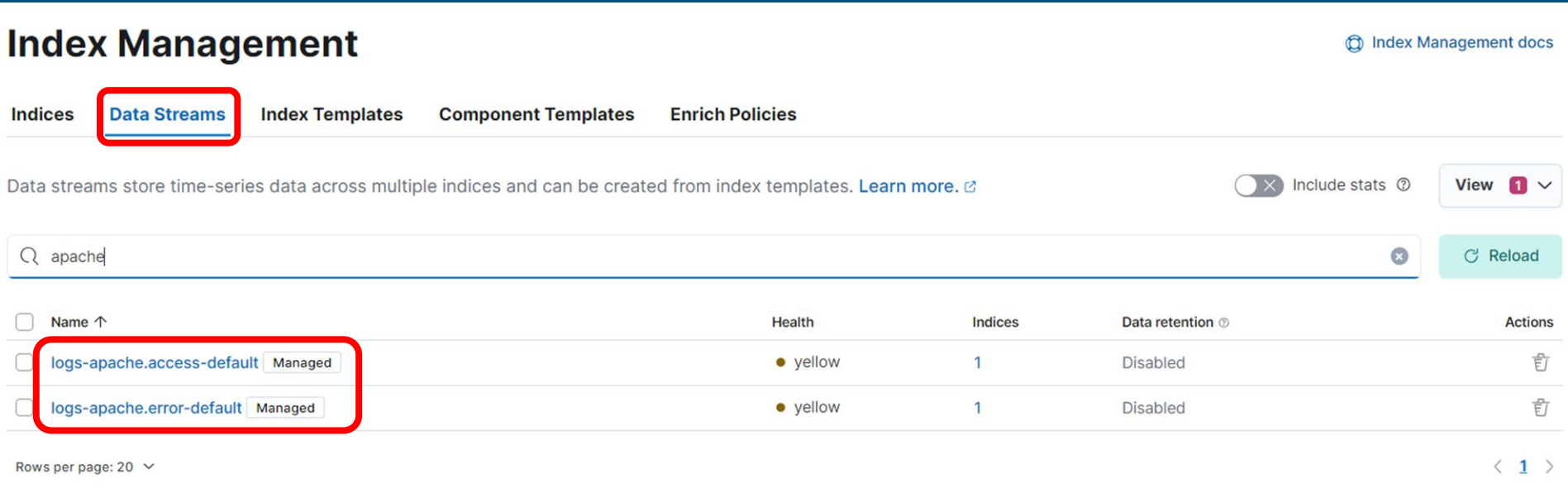
Elastic Agent 安裝



1. 按照步驟安裝即可
2. 最後一步安裝時，如無憑證可加入 --insecure

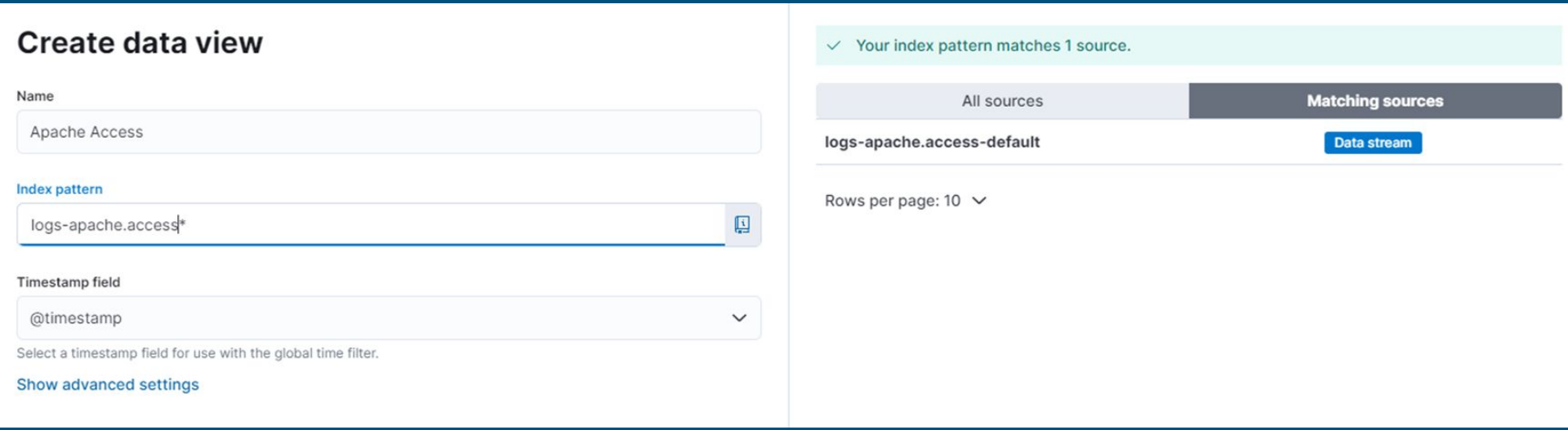


資料寫入查看-data stream



通過 agent 直接寫入的資料並不顯示在普通的 Indices 中，而是以 data stream 的方式儲存

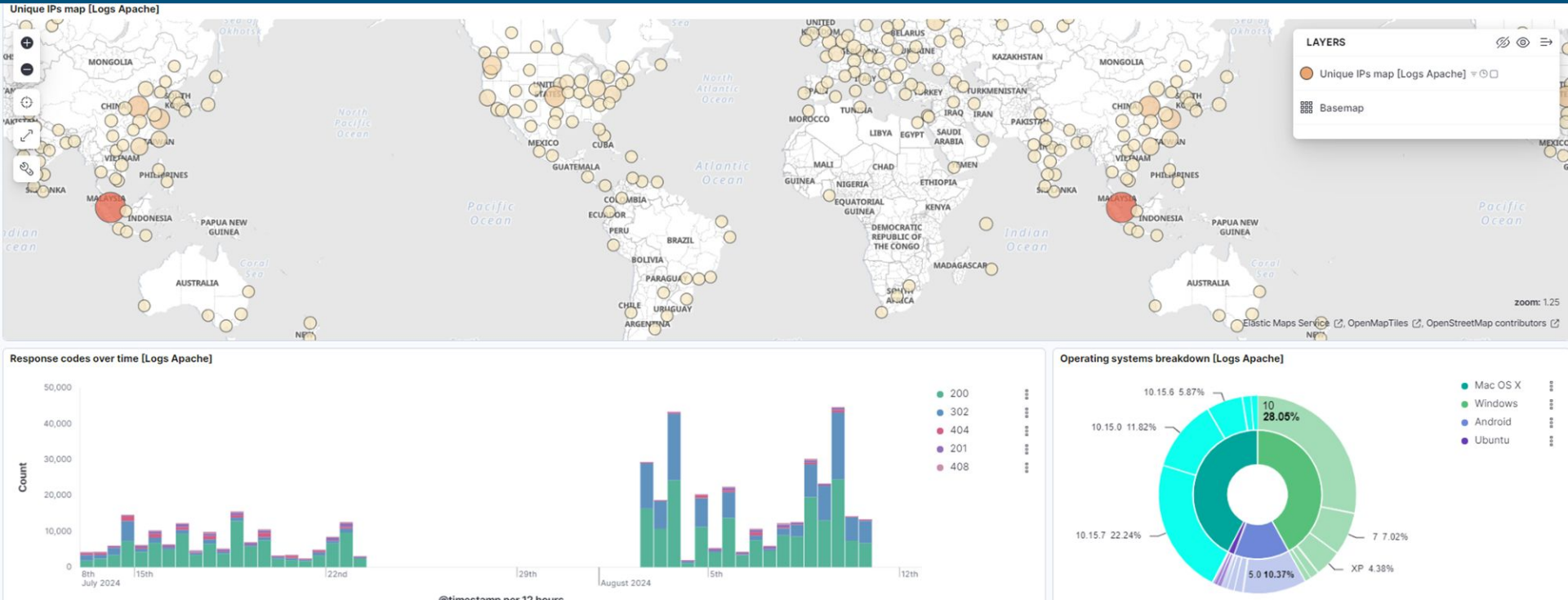
通過製作 Data View 查看 Apache 資料



1. 新資料必須通過建設 Data view 才可查看
2. 以Agent寫入的資料皆以 logs- 開頭

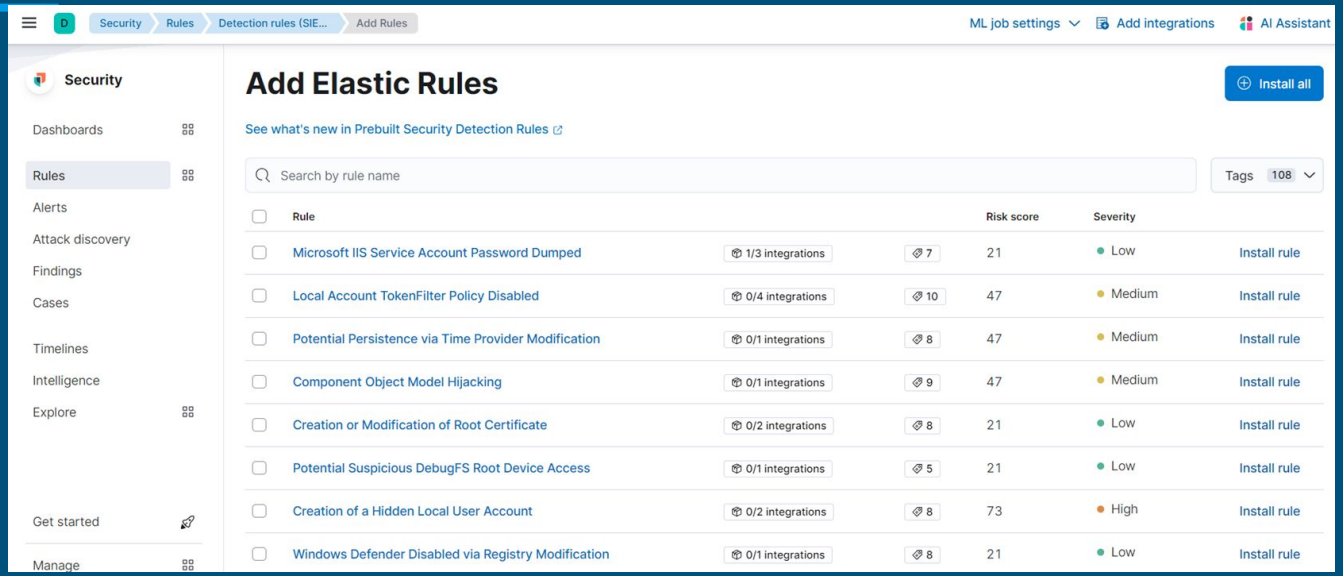
查看 Apache Default 儀表板

通過 Dashboards → [Logs Apache] Access and error logs 可查看默認儀表板



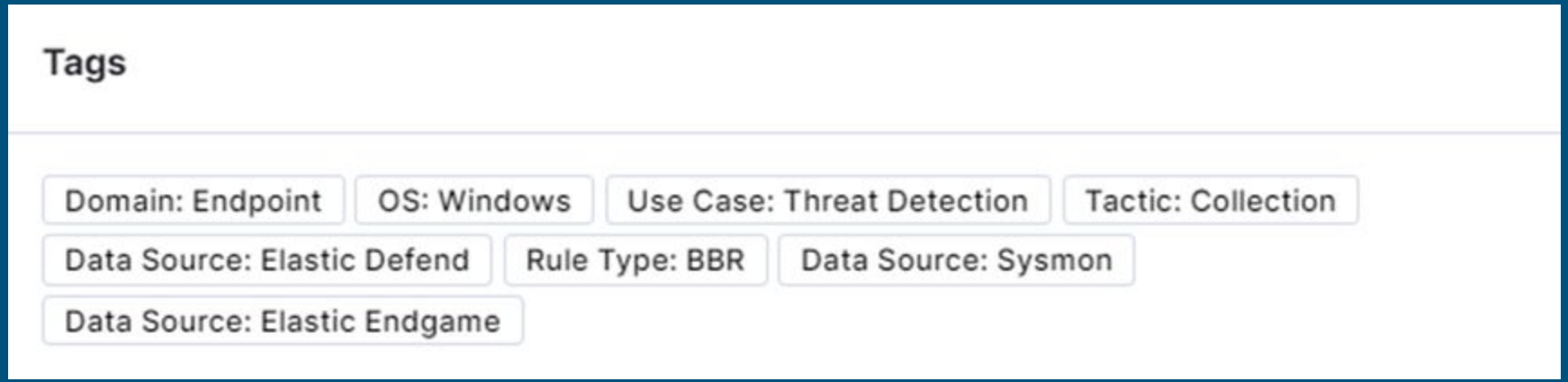
Elasticsearch Security 説明

大量的 Pre-built in rule



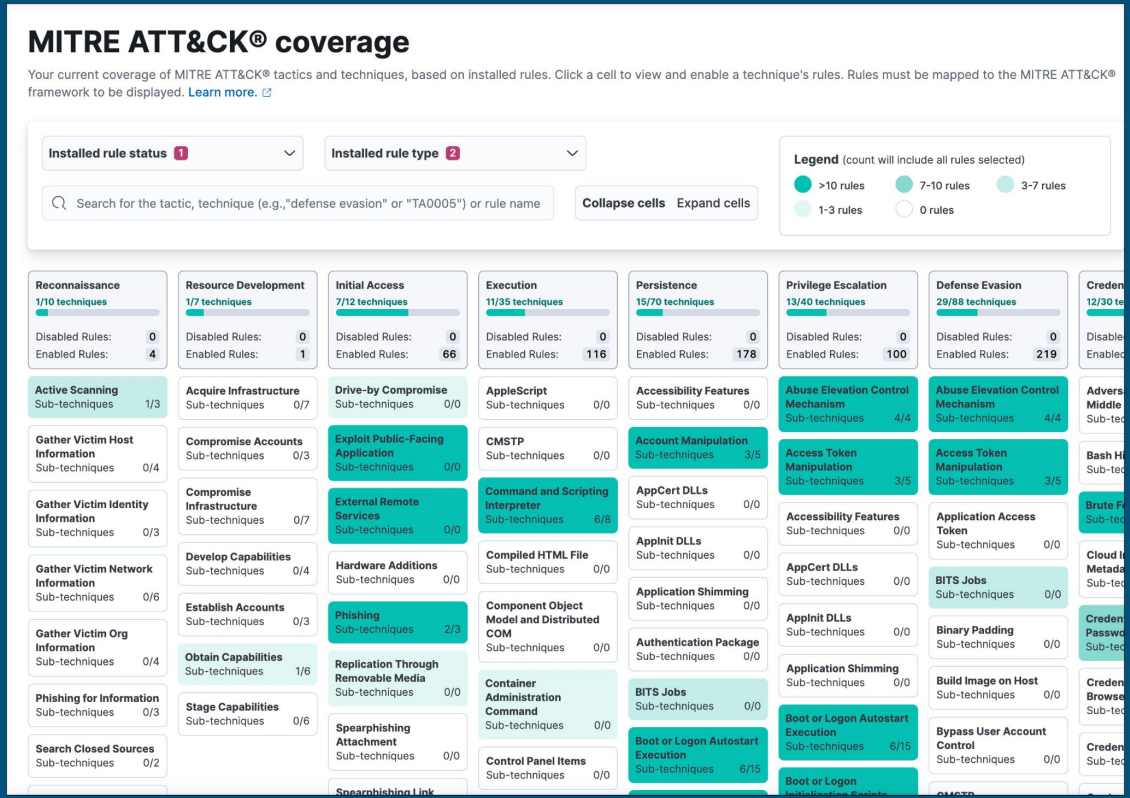
1. 超過1000條的 Pre-built rule，點擊 install rule 即可安裝
2. 需要安裝相應的 intergration 進行配合
3. Pre-built in rule 的安裝不需要開啟 License

可通過查看 Tags 確認所屬 Mitre ATT&CK 類型



Tactic 便為 Mitre ATT&CK 類型

Mitre ATT&CK 頁面 Sample



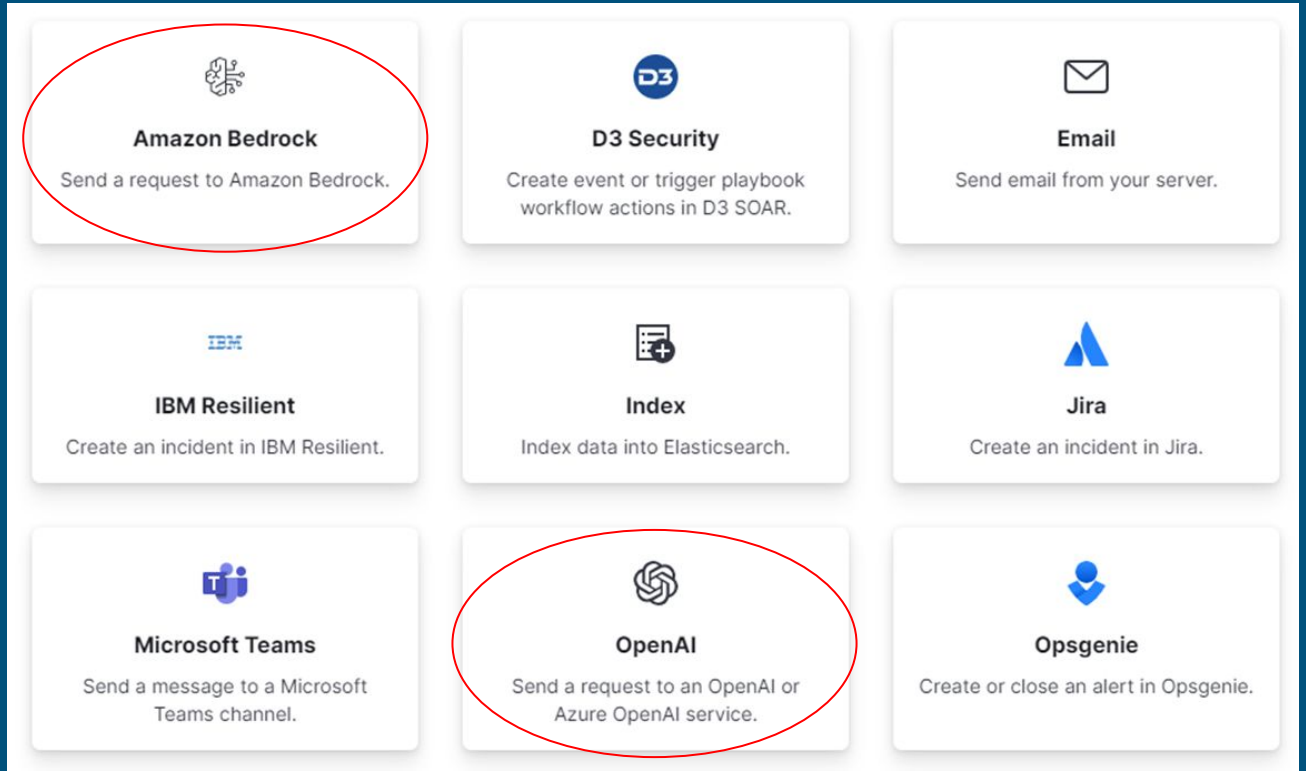
整合生成式AI輕鬆處理解資料

Elasticsearch 整合 AI 計費模式

	Free and open - Basic [1,2]	Platinum	Enterprise
DATA EXPLORATION & VISUALIZATION			
ELASTIC OBSERVABILITY			
Observability overview	✓	✓	✓
User Experience overview	✓	✓	✓
Curated ad hoc data exploration	✓	✓	✓
Service Level Objectives (SLOs)	—	✓	✓
Kibana alerting and actions[5]	✓	✓	✓
Elastic AI Assistant	—	—	✓

1. 要串聯第三方 AI 程式必須購買 Enterprise 版本
2. 第三方程式使用的 Token 需額外付費

Elasticsearch 串聯第三方 LLM



當前可用第三方 LLM 有 Amazon Bedrock 與 OpenAI

生成式 AI 當前面臨問題

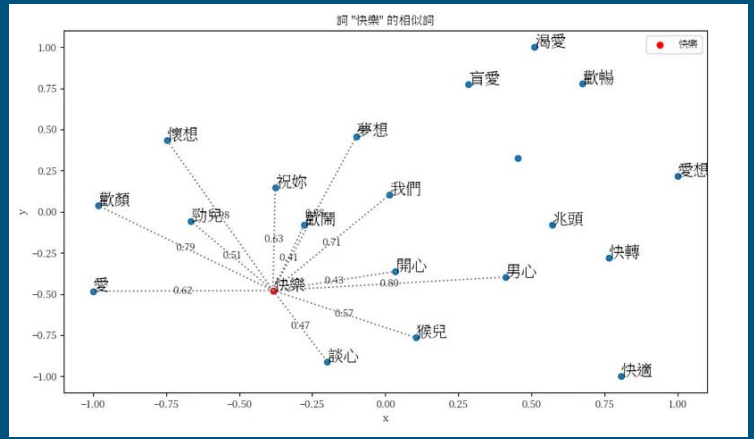
1. 資料訓練截止到 2021年9月，造成資料可能過時或不準確
2. 對特定領域內容知識不足
3. 當問題不夠准確時會提供不正確答案

Elasticsearch 串聯生成式 AI 的優勢

1. Elasticsearch 是一個可大量儲存資料、文本、向量的資料庫。可以非常快速的提供準確的查詢結果
2. 提供混合搜索能力 (BM25 & kNN)
   a. BM25 為 傳統文本搜索
   b. kNN (k-Nearest Neighbor) 為可供 AI 使用的向量搜索 (vector search)
3. 提供豐富的 API 及資料串聯能力

簡單的向量化數據庫說明

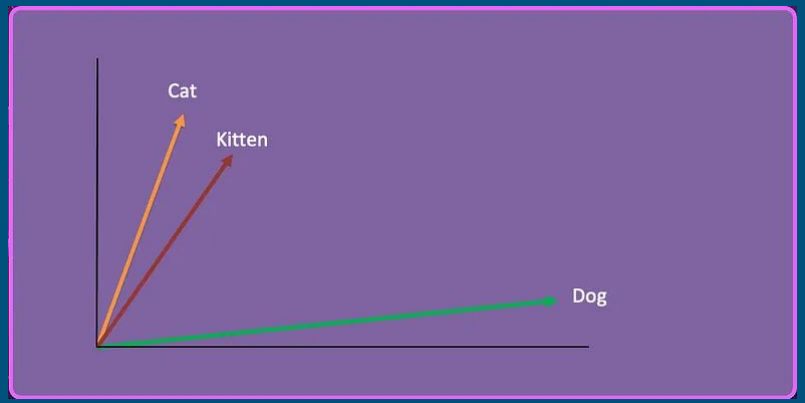
1. 數據 (文本、圖像、影音)會被轉換為向量
   a. 例如 “Hello World!” 會被轉換為 [72, 101, 108, 108, 111, 32, 87, 111, 114, 108, 100, 33]
2. 資料可以 類似 Cosine similarity 的方式查詢相似性
3. 優點:
   a. 可以通過關鍵字或語義 查詢查找相關文檔
   b. 可以查找複雜的資料 (圖像, 影音)
   c. 自然語言處理 (NLP) 的必要前置條件
   d. 不需像傳統 DB 一樣手動定義同義字等



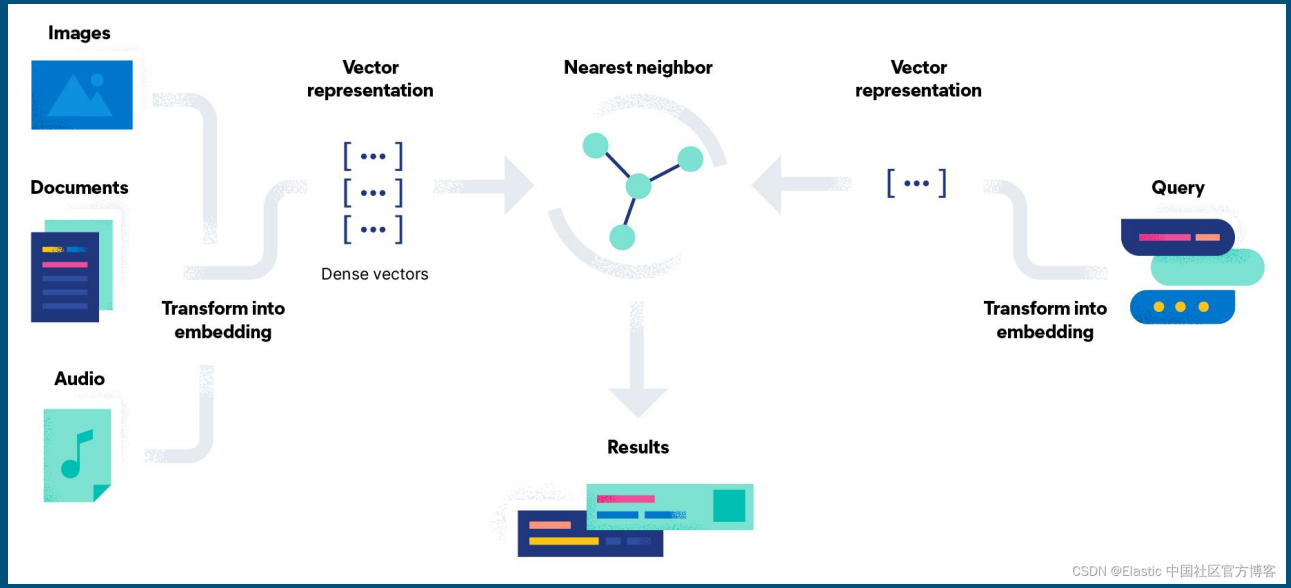
簡單的向量搜索 Sample

1. “Cat” 會被轉換為 [0.49671415, -0.1382643 , 0.64768854, 1.52302986, -0.23415337, -0.23413696, 1.57921282, …]
2. “Dog” 會被轉換為 [1.69052570, -0.46593737, 0.03282016, 0.40751628, -0.78892303, 0.00206557, -0.00089039, …]
3. “Kitten” 會被轉換為 [-0.05196425, -0.11119605, 1.0417968, -1.25673929, 0.74538768, -1.71105376, -0.20586438, …]
4. 從結果來看 Cat 與 Kitten 的 coscine distance

更接近所以 Kitten 和 Cat 相似

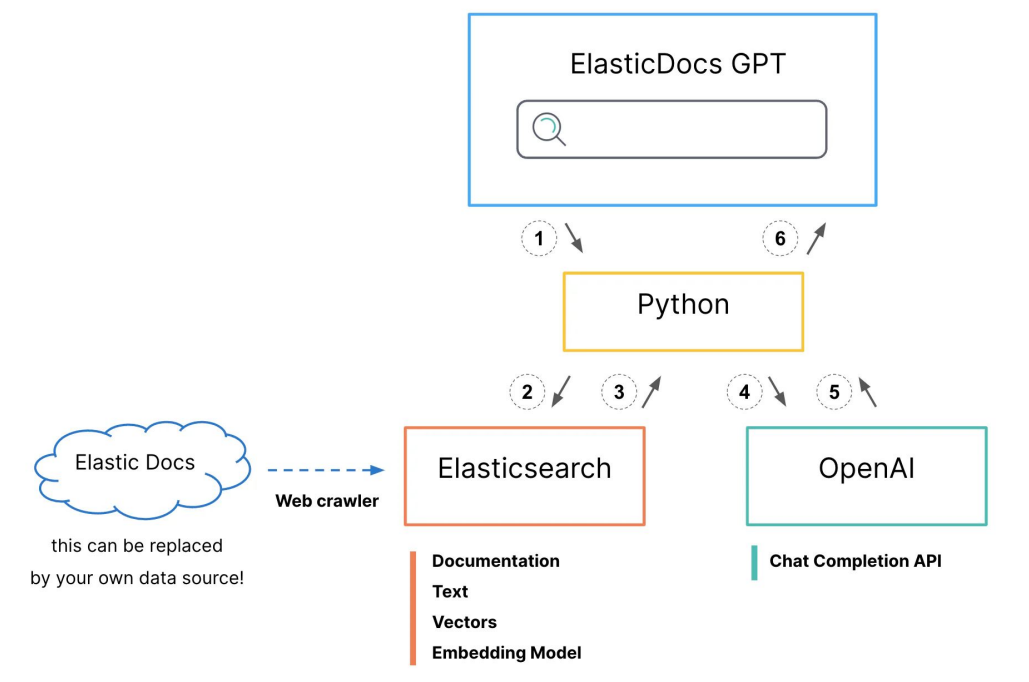


Elasticsearch 簡單的向量搜索說明



通過儲存時將資料轉換為陣列後儲存如Elasticsearch，用戶查詢時將查詢語句也轉換為陣列，在通過 kNN 查詢後返回最相似的結果

Elasticsearch 與 第三方 AI 串聯方法



1. 用戶通過主頁面輸入問題給後台 Python
2. 通過生成 BM25 與 kNN 搜索查詢 Elasticsearch
3. 將最終唯一結果返回給 Python 頁面
4. 調用 OpenAI API 重新匯整問題。要求 OpenAI 回答問題是必須採用 Elasticsearch 的查詢結果
5. 查詢結果返回給 Python 程式
6. Python 將 OpenAI 結果與 Elasitcsearch 查詢結果一併返回

Retrieval-Augmented Generation (RAG)- 檢索增強生成

大型語言模型優點:

1. 可以理解人類上下文和意圖
2. 可與人類進行互動
3. 多語種支援

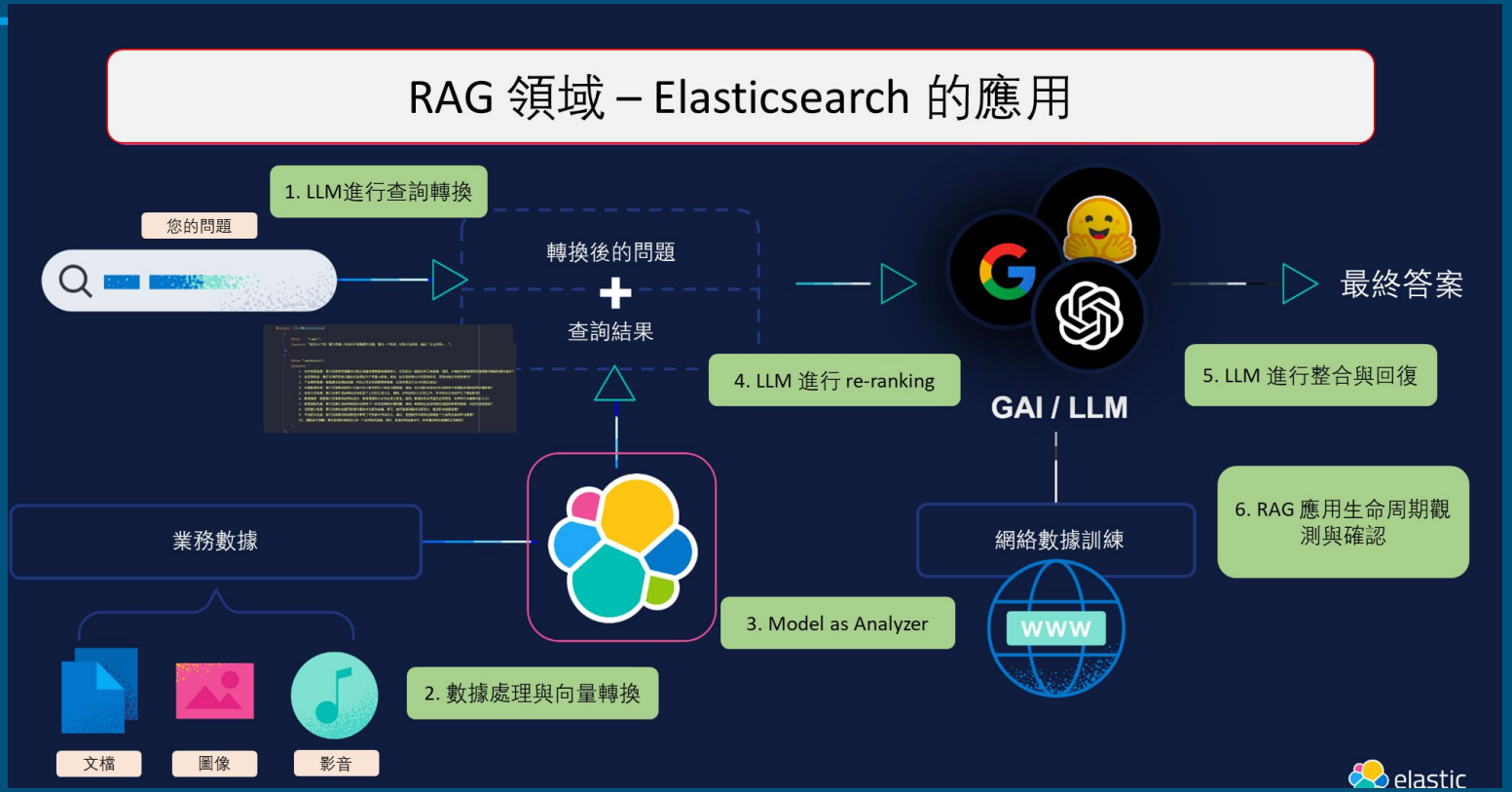
大型語言模型缺點:

1. 無法持續更新資料，不會因為資料更新而重新訓練Model
2. 因缺乏特定行業資料，對特定行業認知有限，可能回答錯誤內容
3. 對問題的輸入方式很敏感，同樣的問題，不同的輸入方式可能有不同的答案
4. 消耗大量資源無法大量遠程部署

Retrieval-Augmented Generation (RAG)- 檢索增強生成

1. 結合 “搜尋檢索” 和 “生成能力” 的自然語言處理架構。
2. 執行步驟:
   a. 從外部資料庫搜索相關信息
   b. 整合相關信息後生成問題後交給 大語言模型 (LLM)
   c. 由大語言模型進行回復完成特定 NLP 任務
3. RAG 基本架構:
   a. Retriever (檢索器): 負責提供儲存與資訊 - Elasticsearch
   b. Generator (生成器): 語言生成模型 - ChatGPT
   c. Integration and Adjustment (整合與調整): 負責將問題發送至檢索器，後基於特定資訊生成問題後交給生成器，最終將結果返回 - Python

Elasticsearch 與 生成式 AI 的整合



Elasticearch 整合生成式 AI 是否會造成隱私洩露?

1. 大語言模型必須有對外
2. 資料會經過 Elasticsearch 加工後送出，且僅送出最相關的1筆資料，不是直接寄送給 ChatGPT
3. 機密資料或欄位可預先處理，僅送出特定欄位

Elasticsearch 與常見 RAG 系統比對

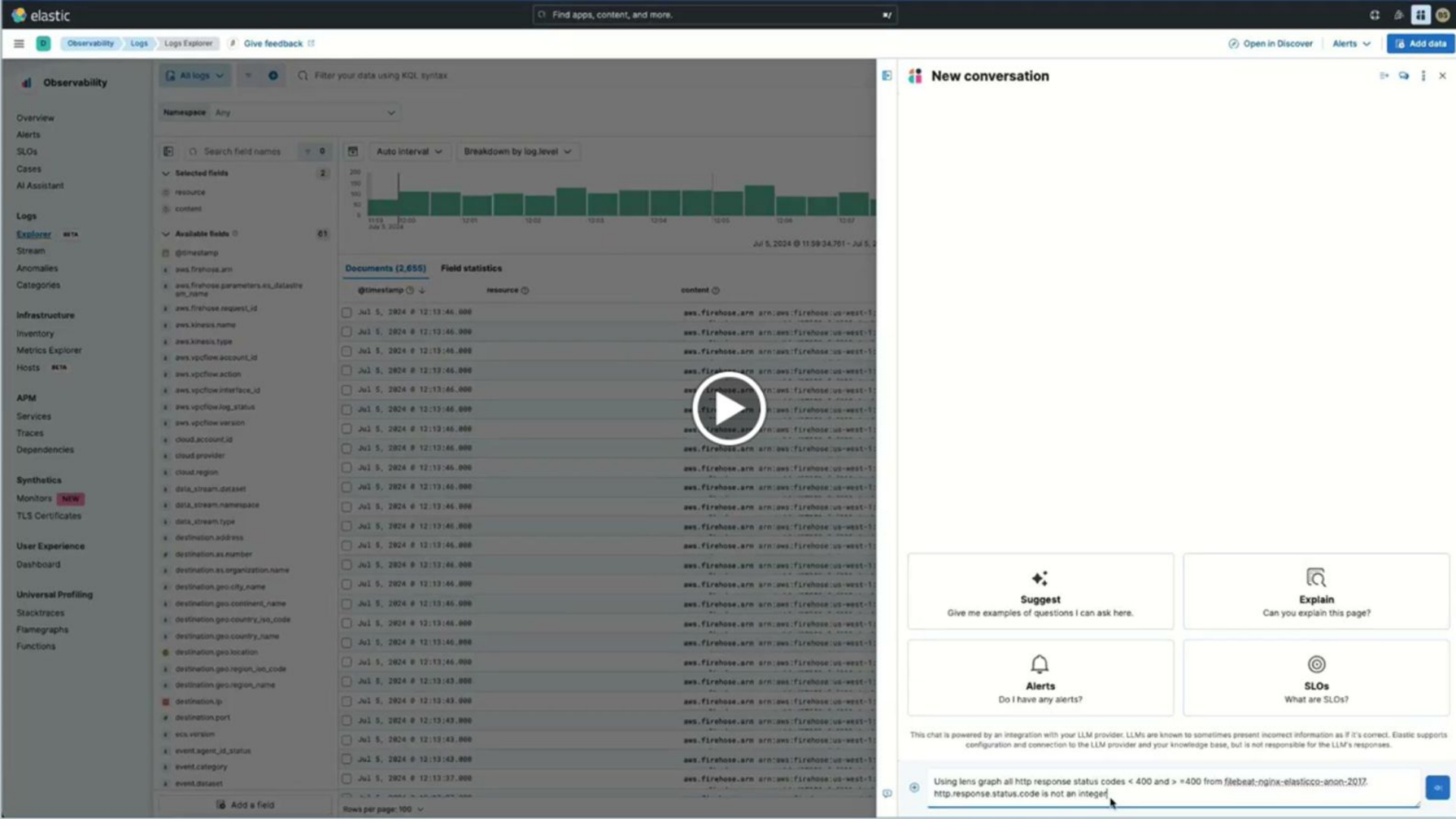
其他 RAG 系統常見問題:

1. 高度依賴檢索能力
2. 計算資源需求高
3. 長文本檢索困難效率低下

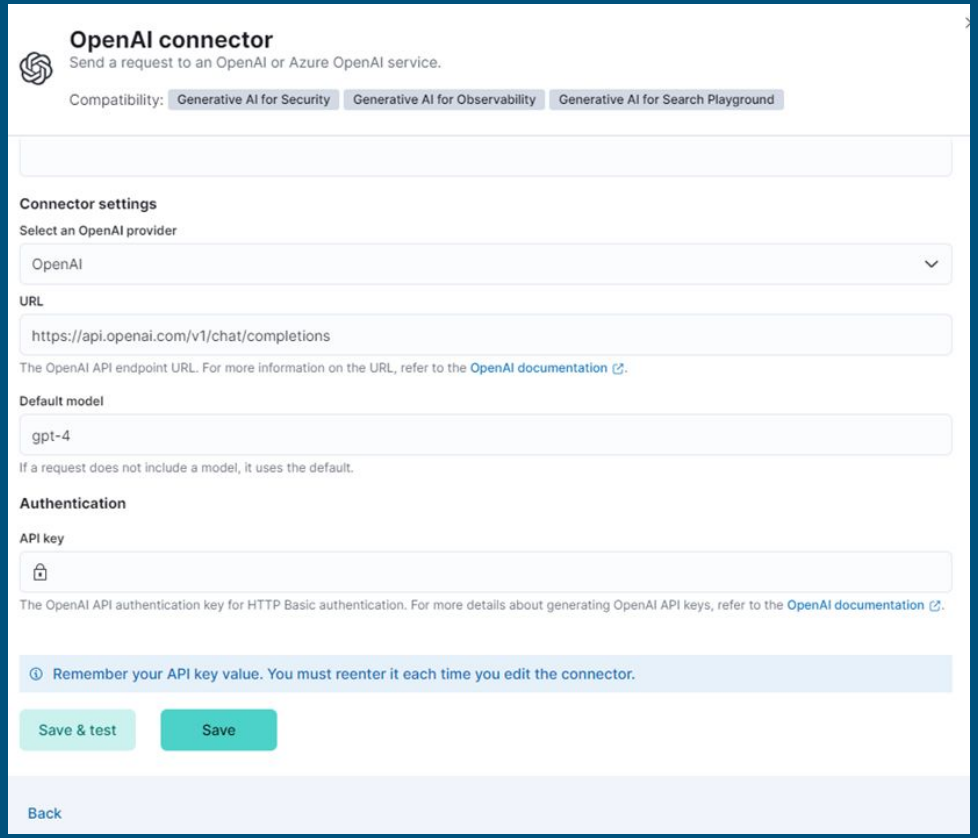
Elasticsearch 優勢:

1. 檢索效率高效
2. 長期優化檢索算法，資源耗費較少

Elasticsearch 結合生成式 AI 結果展示 - Nginx


elastic
Find apps, content, and more.
Observability
Logs
Logs Explorer
Give feedback
Open in Discover
Alerts
Add data
Observability
Overview
Alerts
SLOs
Cases
AI Assistant
Logs
Explorer BETA
Stream
Anomalies
Categories
Infrastructure
Inventory
Metrics Explorer
Hosts BETA
APM
Services
Traces
Dependencies
Synthetics
Monitors NEW
TLS Certificates
User Experience
Dashboard
Universal Profiling
Stacktraces
Flamegraphs
Functions
All logs
Filter your data using KQL syntax
Namespace Any
Search field names
Selected fields 2
resource
content
Available fields 61
@timestamp
Auto interval
Breakdown by log.level
Jul 5, 2024 @ 11:59:34.761 - Jul 5, 2
Documents (2,655)
Field statistics
@timestamp
resource
content
Add a field
Rows per page: 100
New conversation
Suggest
Give me examples of questions I can ask here.
Explain
Can you explain this page?
Alerts
Do I have any alerts?
SLOs
What are SLOs?
This chat is powered by an integration with your LLM provider. LLMs are known to sometimes present incorrect information as if it's correct. Elastic supports configuration and connection to the LLM provider and your knowledge base, but is not responsible for the LLM's responses.
Using lens graph all http response status codes < 400 and > =400 from filebeat-nginx-elasticco-anon-2017.
http.response.status.code is not an integer

OpenAI connector



輸入 API key 即可串聯至 OpenAI

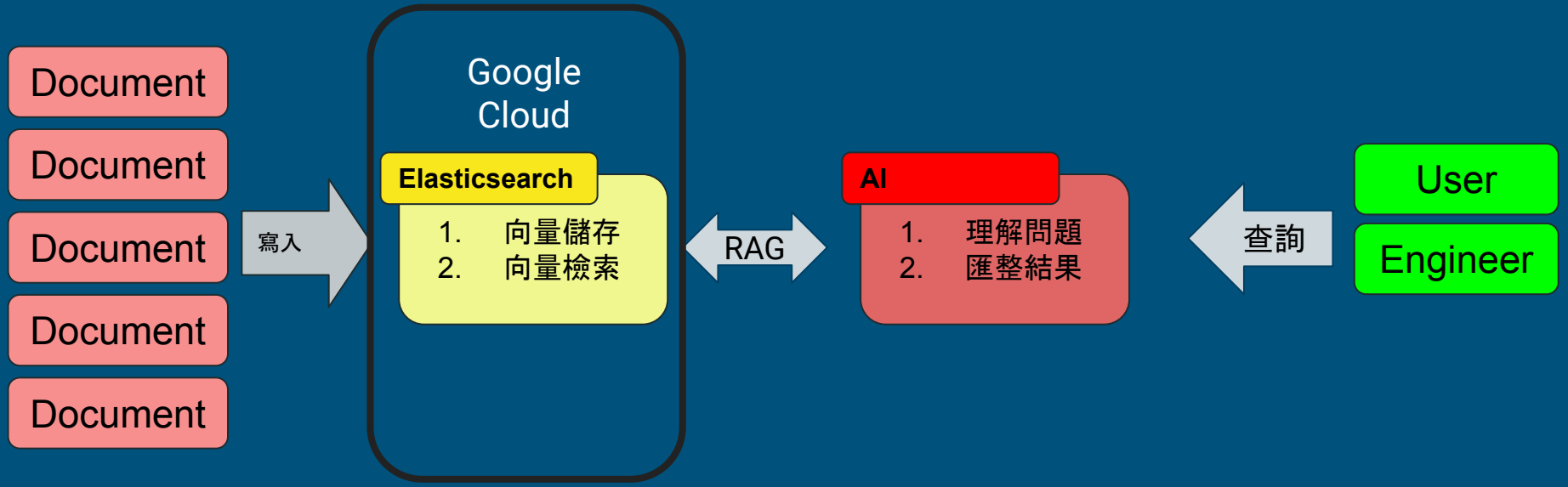
Elasticsearch AI 案例1 - Cisco

Cisco 面臨問題:

1. 歷史資料技術文件太多，且放置位置過多，搜索困難且耗時
2. 工程師要尋找文件尋要從上百萬文件中撈取，耗時且效果差
3. 維護人員處理維護案件時，無法及時的取得常見問題處理方式
4. 客戶無法從官網順利找到自己希望的文件

Elasticsearch AI 案例1 - Cisco

處理方式:



Elasticsearch AI 案例1 - Cisco

結果:

1. 工程師回應時間加快 73%
2. 新系統解決了90% 的 支援請求
3. 每月減少 5000 工程師工時

Elasticsearch AI 官方案例2 - Stack overflow

Stack overflow 遭遇問題:

1. 查詢時間過長: 隨著時間不斷快速增長，預期每項科技產品超過 6000萬個問題
2. 順利找出有價值的答案: 問題中有大量重複問題且價值有限
3. 結果讀取緩慢: 查詢頻率不斷提升，導致查詢時間緩慢
4. 希望導入 AI 查詢，但是資料過大查詢緩慢且準確率僅有 42%

Elasticsearch AI 案例2 - Stack overflow

Stack overflow 基於 AI 解決問題:

1. 建立高可信的用戶資料庫: 優先提供高可信度用戶的回復 (Stack overflow 自行解決)
2. 縮短查詢時間和頻率: 通過 AI 優化查詢問題，使查詢更准確答案更符合需求，降低重複查詢頻率

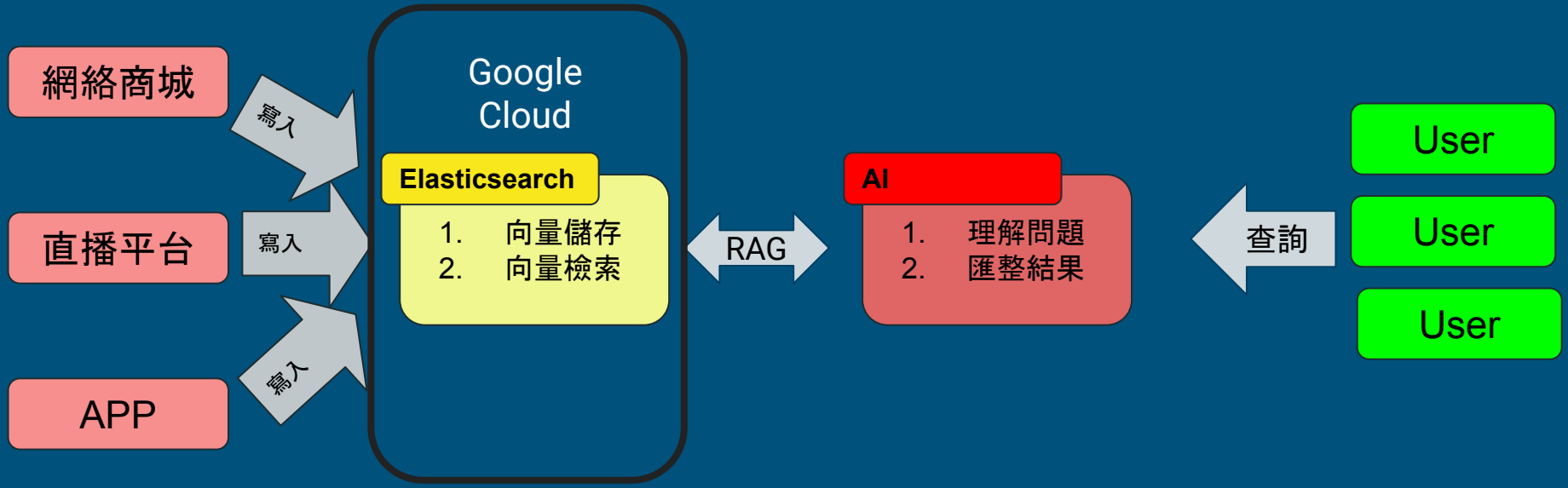
Elasticsearch AI 案例3 - HSE retail

HSE 遭遇問題:

1. 客戶查詢條件不準確
2. 有3個通路: 網絡商城, 直播平台, APP 導致資料無法順利整合
3. 查詢結果不佳, 客戶反饋差, 購買意願低

Elasticsearch AI 案例1 - Cisco

處理方式:



Elasticsearch AI 案例3 - HSE retail

HSE 基於 Elasticsearch AI 解決問題:

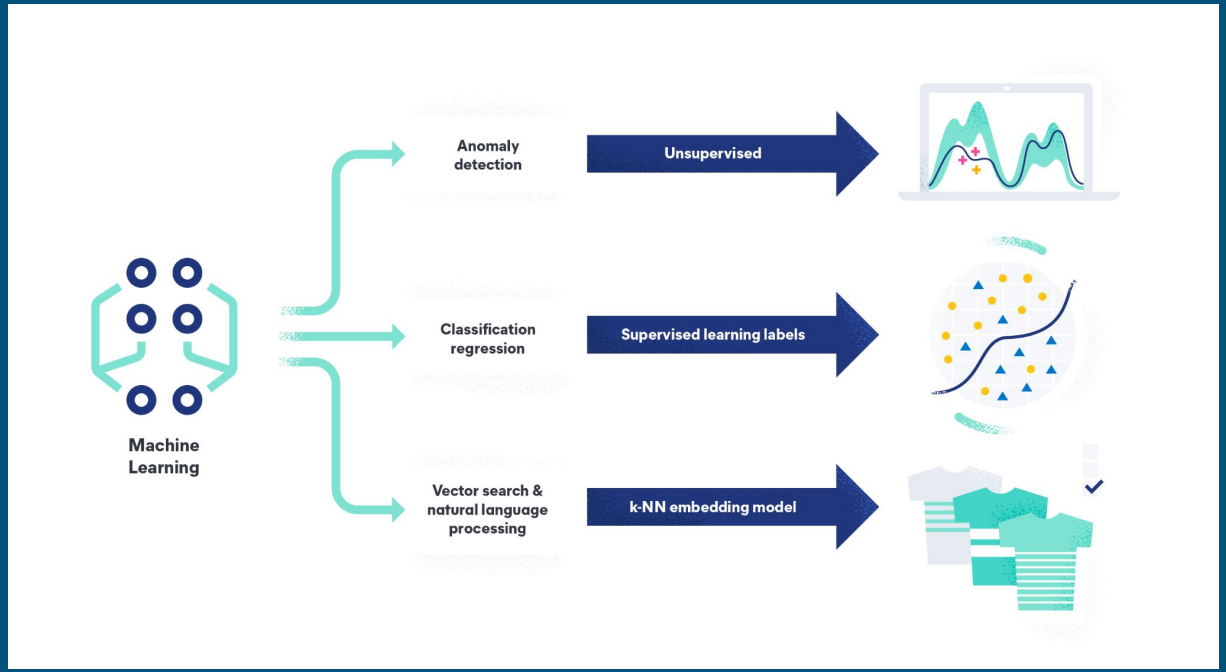
1. 使用 Vector search 使客戶更容易查詢到希望的商品，使點擊購買率增加 4%
2. 減少42%設備的維護時間
3. 減少客戶查詢時的等待時間，客戶滿意度提升 8%

使用 Machine Learning 進行異常偵測

Elasticsearch Machine Learning 優勢

1. 可儲存大量資料: 大量資料是 Machine learning 的前提，而 elasticsearch 正好是儲存大資料的專家
2. 高靈活性: 可簡單擴容，即使資料越來越大，也可輕鬆應對
3. 查詢快速: 通過倒排索引可快速查詢
4. 方便後續排查: Machine learning 後的結果依然需要人為確認，Kibana 可以很簡單的排查問題

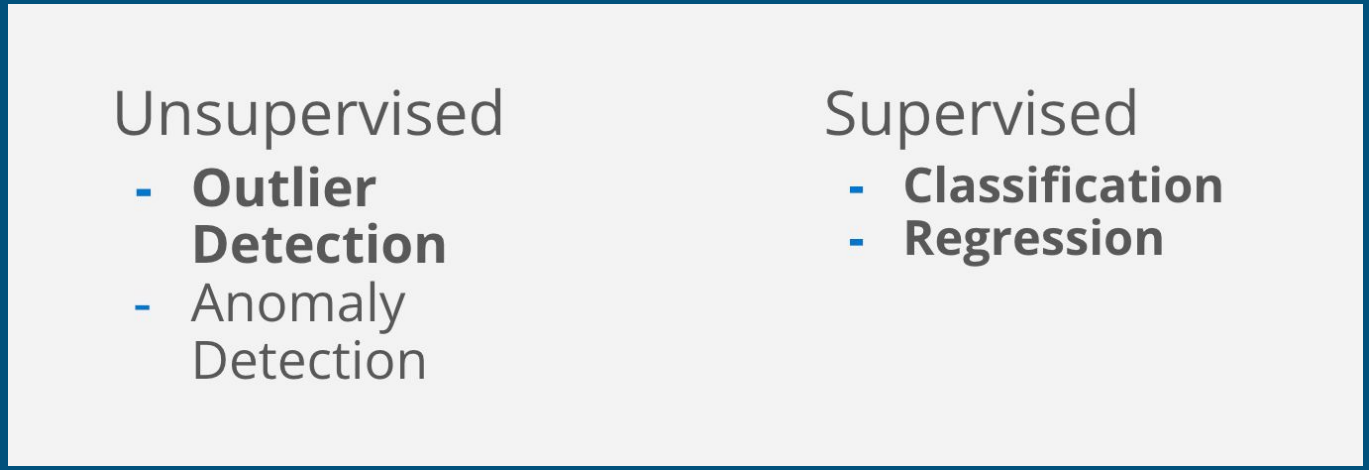
Elasticsearch Machine Learning 分類



Elasticsearch 的 ML 分為3部分:

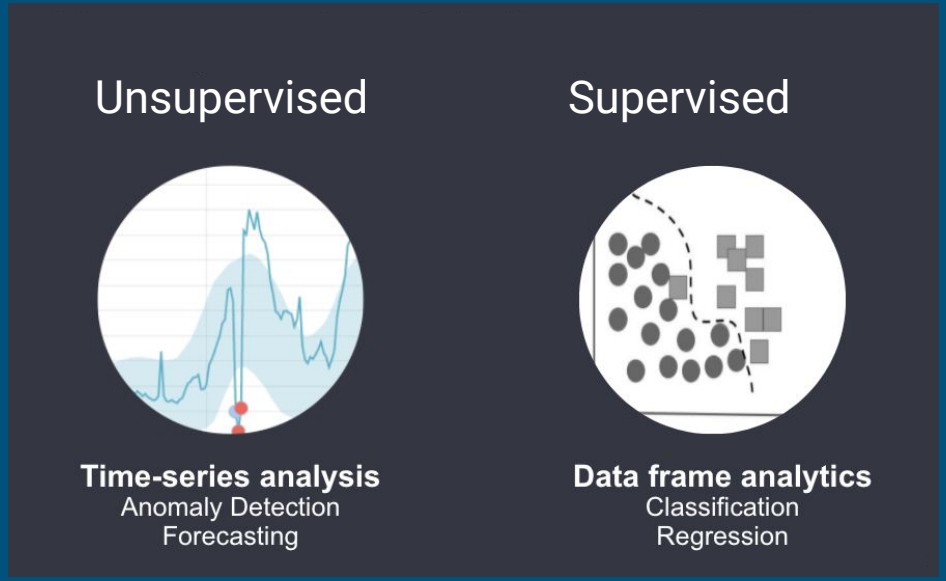
1. 異常偵測
2. 分類預測
3. 向量搜索

資安常用 Machine Learning



1. 資安中僅有 Unsupervised 和 Supervised learning 常用
2. Unsupervised Model 不需額外加工因此絕大部分客戶都使用這個
3. Supervised Model 需要將資料加工成 Panel Data 樣式所以使用率較低

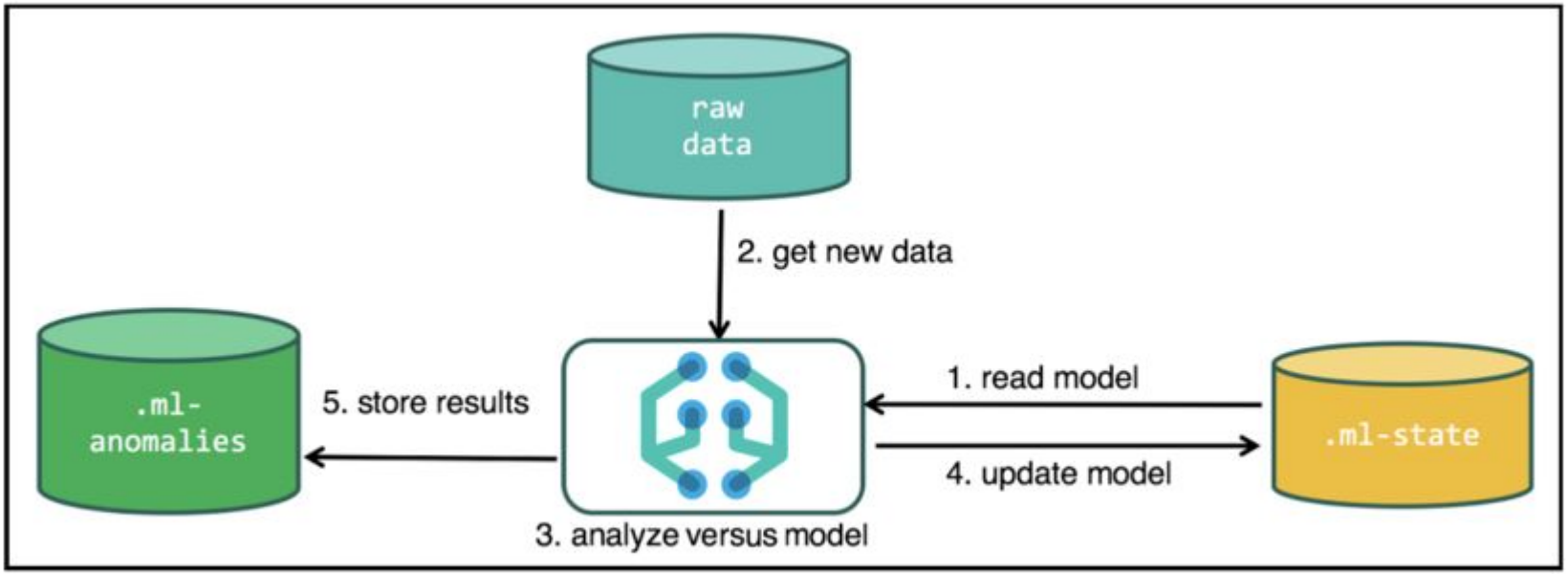
機器學習分類



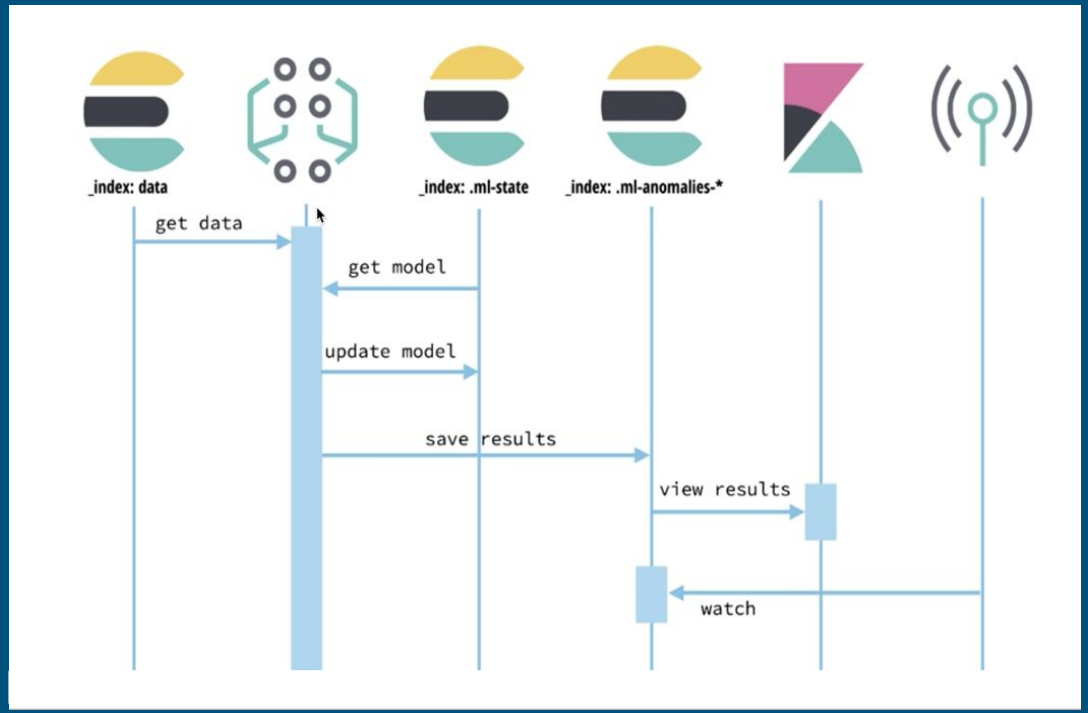
Unsupervised machine learning 作用於 Time series data 資料

Supervised machine learning 作用於 Panel data 資料

Elasticsearch 機械學習說明



Elasticsearch 機械學習說明

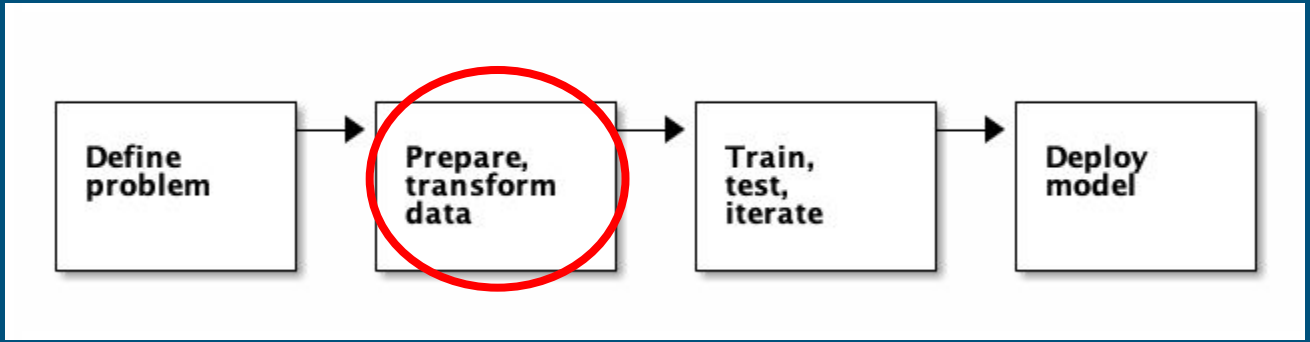


.ml-stata: 負責儲存模型

.ml-anomalies: 負責儲存異常

data: 為原始資料

監督式機械學習過程



1. 資料需要經過額外處理才可以使用，基本不能以 raw data 進行學習，且 Model 需要反復測試
2. 需將 Timeseries 資料整理轉換為 Panel Data 格式

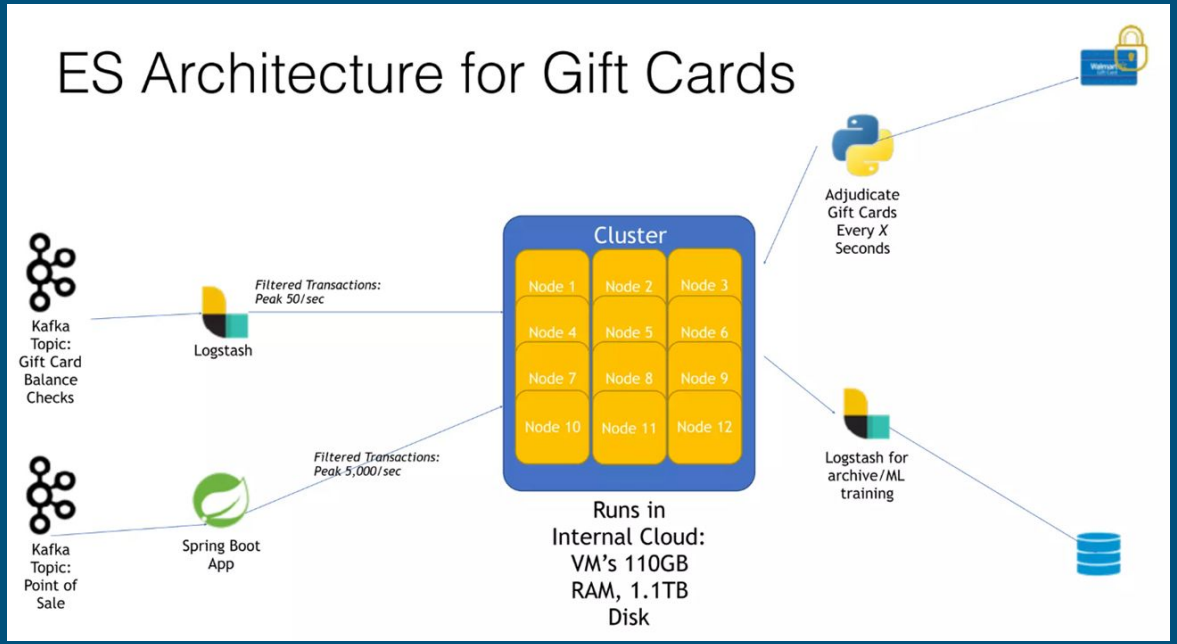
Elasticsearch ML 案例1 - 沃爾瑪

沃爾瑪面臨問題:

電信詐騙誘導受害者通過沃爾瑪購買點數卡轉移現金，避開銀行和監管平台

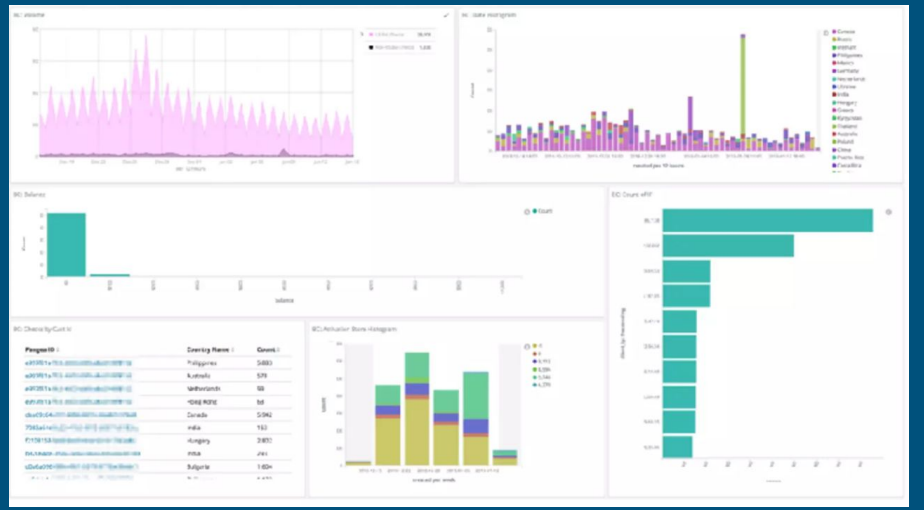
Elasticsearch ML 案例1 - 沃爾瑪

Elasticsearch 架構



Elasticsearch ML 案例1 - 沃爾瑪

通過導入信用卡資料, IP, 銷售金額的 AP 日誌後, 同時進行人工與 ML 作業, 保證用戶不被電信詐騙騙走財產



Elasticsearch ML 案例2 - 歐洲電信警察

面臨問題:

1. 需收集超過250套 IT 系統，超過3500 台 Endpoint 的資安日誌記錄，資料太大太多且增長迅速
2. 黑客行為不斷增加，對警力的需求不斷上升

Elasticsearch ML 案例2 - 歐洲電信警察

結果:

1. Elasticsearch 收取了上一個 SIEM 10倍量的資安資料
2. Endpoint 從 3500 台增加到 35000 台
3. 通過 Elasticsearch 異常偵測自動判斷黑客攻擊, 降低警務勞力

Elasticsearch ML 案例2 - 歐洲電信警察

結果:

1. Elasticsearch 收取了上一個 SIEM 10倍量的資安資料
2. Endpoint 從 3500 台增加到 35000 台
3. 通過 Elasticsearch 異常偵測自動判斷黑客攻擊, 降低警務勞力

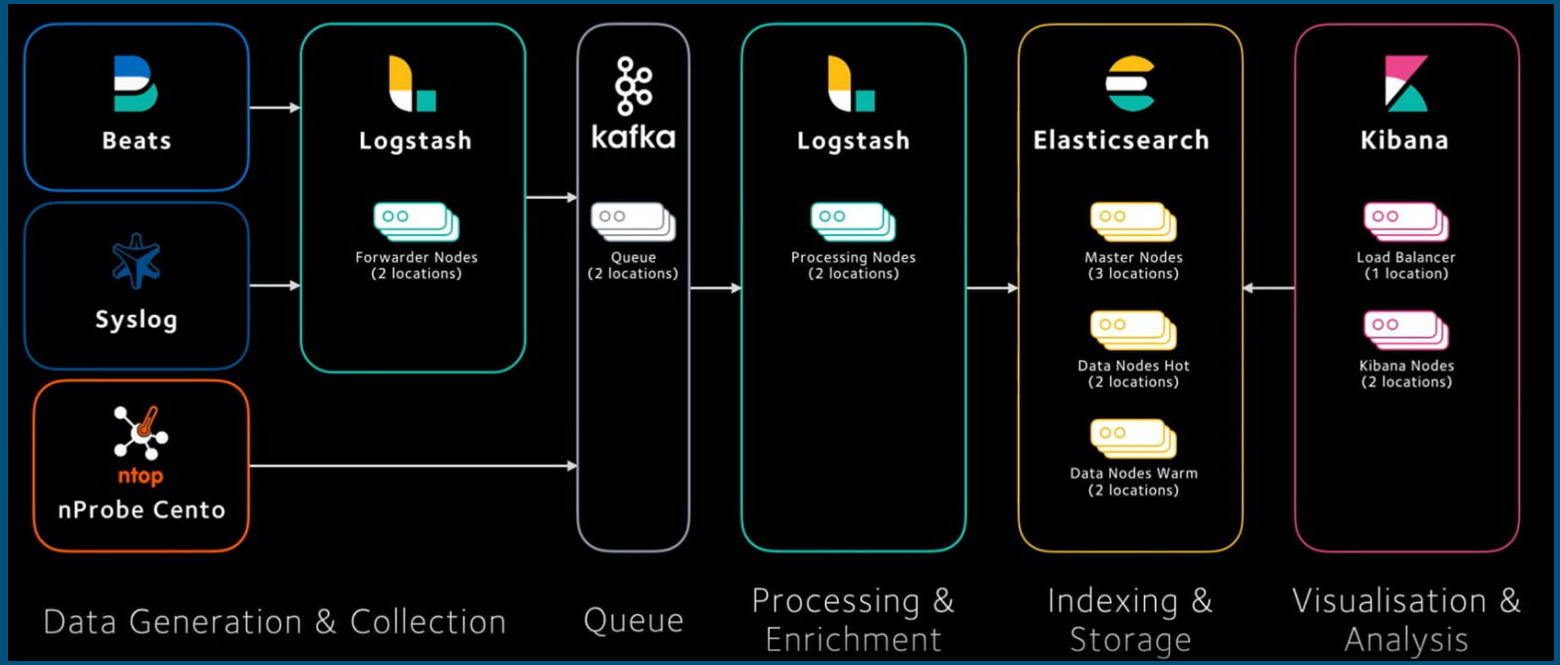
Elasticsearch ML 案例3 - 哈佛大學開發新一代 SIEM

欄位統一:

Codec nProbe Cento	Logstash Module	FileBeat Module	Elastic Common Schema
netflow.ipv4_src_addr	netflow.src_addr	netflow.source_ipv4_address	source.ip
netflow.l4_src_port	netflow.src_port	netflow.tcp_source_port netflow.udp_source_port	source.port
netflow.in_pkts	netflow.packets	netflow.packet_total_count [and many others]	source.packets network.packets

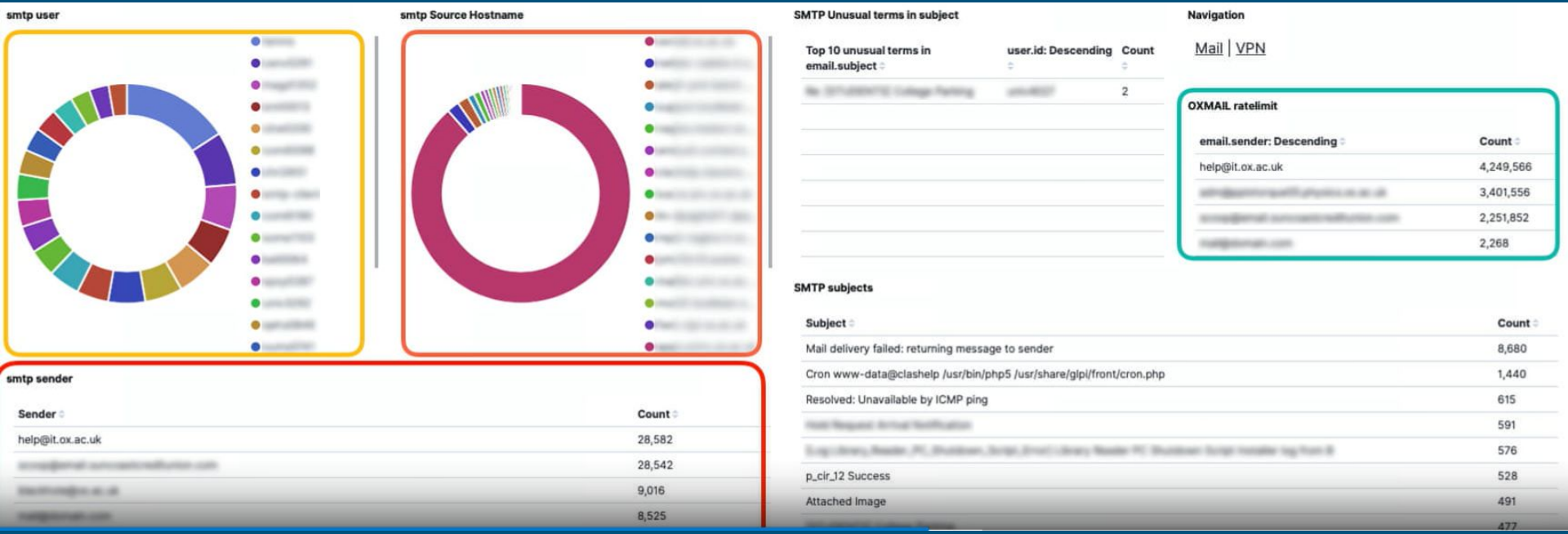
Elasticsearch ML 案例3 - 哈佛大學開發新一代 SIEM

系統架構:



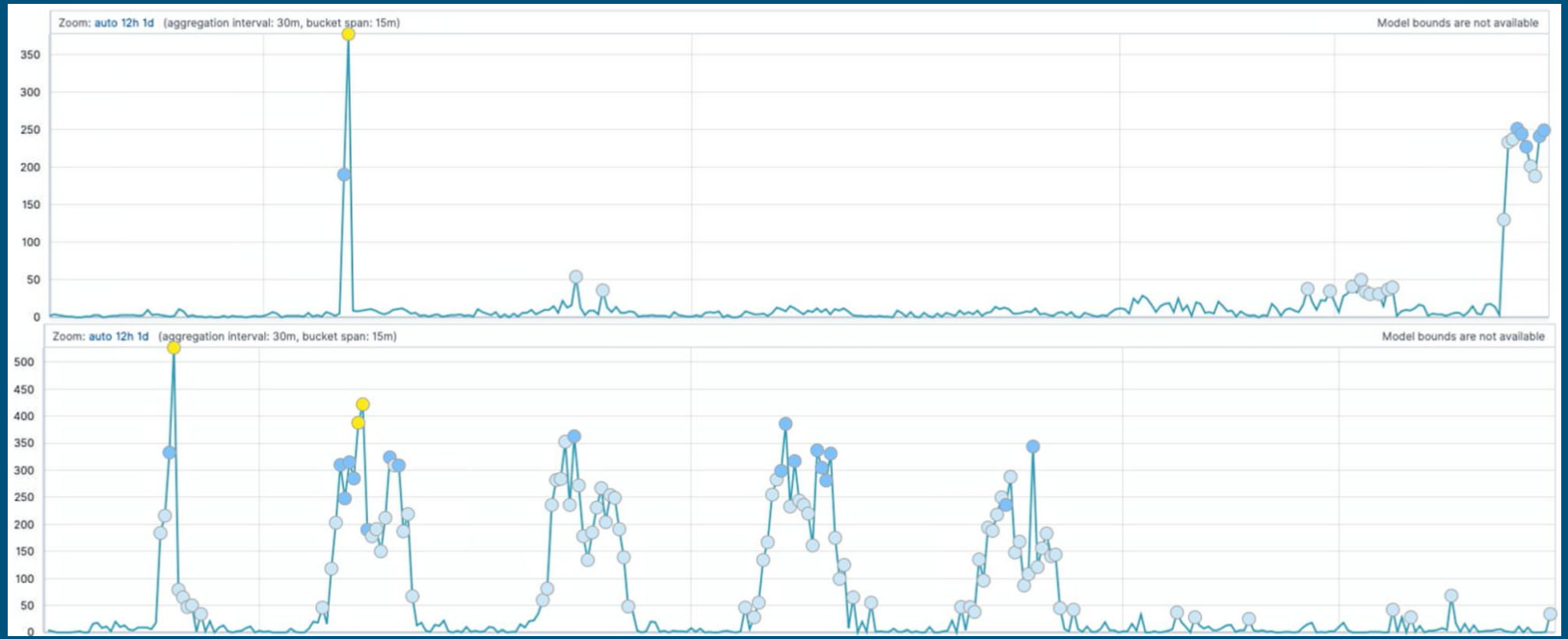
Elasticsearch ML 案例3 - 哈佛大學開發新一代 SIEM

Dashboard 展示:



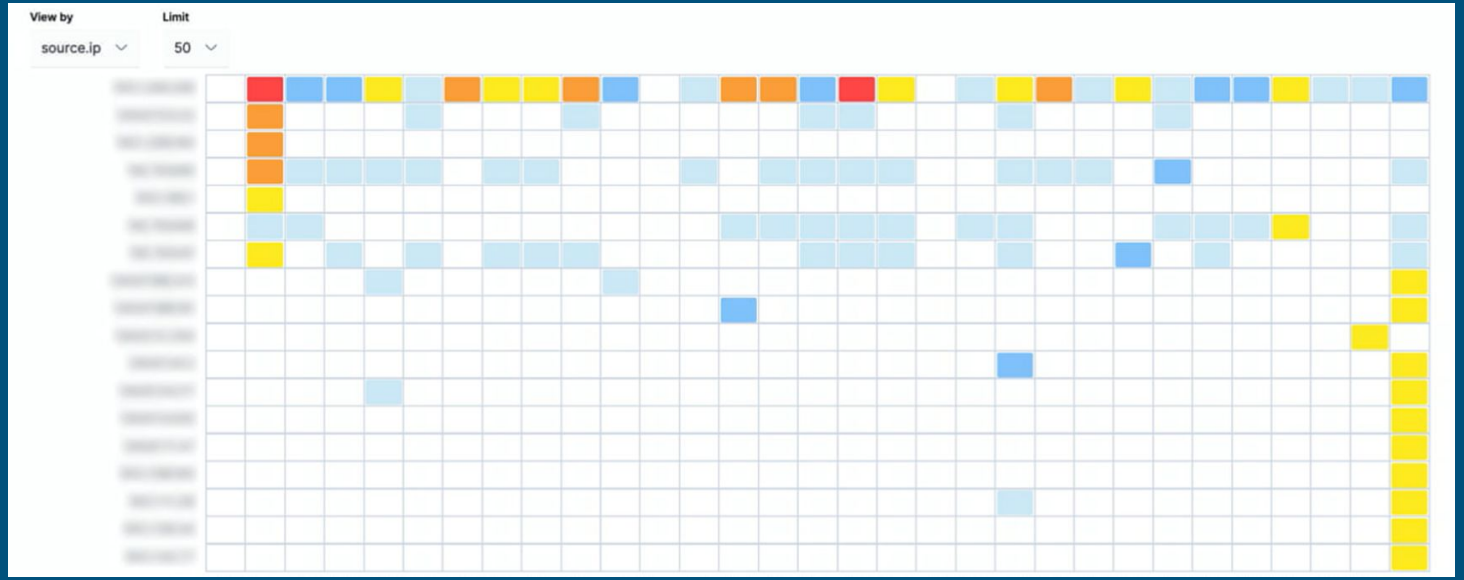
Elasticsearch ML 案例3 - 哈佛大學開發新一代 SIEM

ML 發現展示:



Elasticsearch ML 案例3 - 哈佛大學開發新一代 SIEM

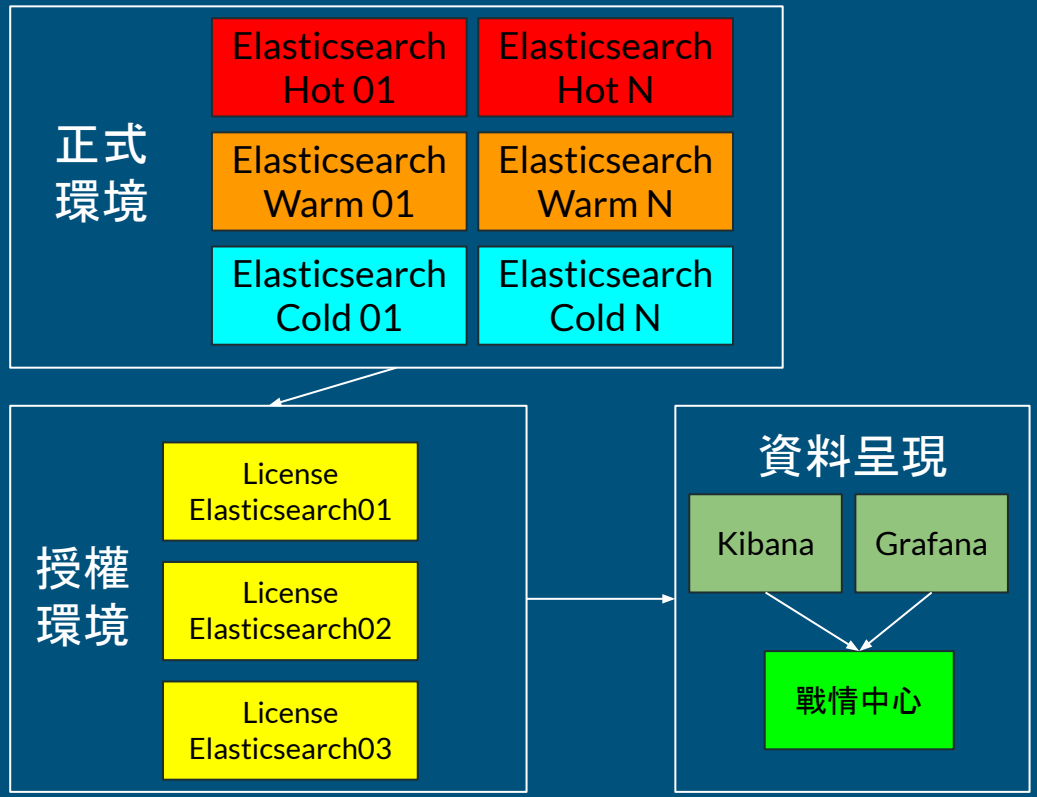
ML 發現展示:



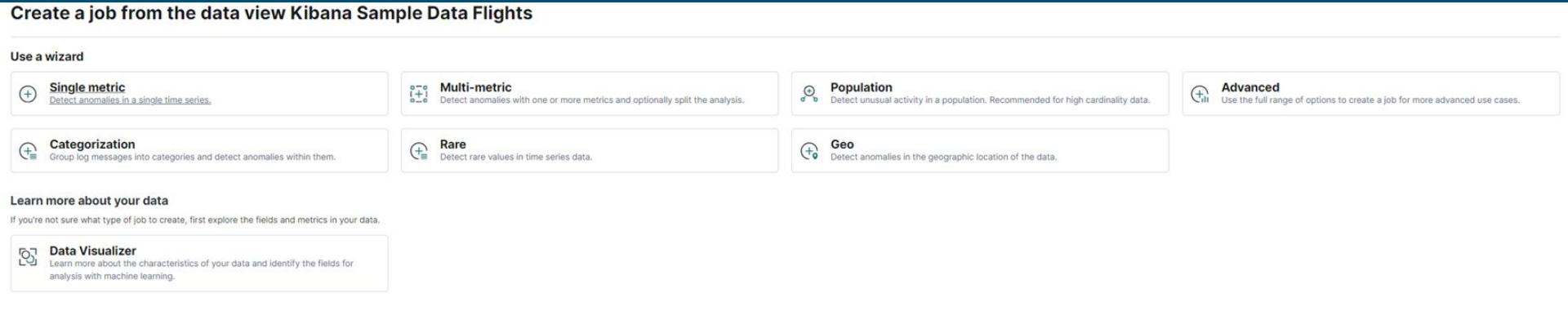
Elasticsearch ML 案例4 - 台灣中 X

系統規劃架構:

因預算有限，將大部分資料儲存在正式環境中，定時將處理後的資料寫入授權環境進行 Machine Learning 作業



Elasticsearch ML Unsupervise Model 類型總覽



Anomaly Detection:

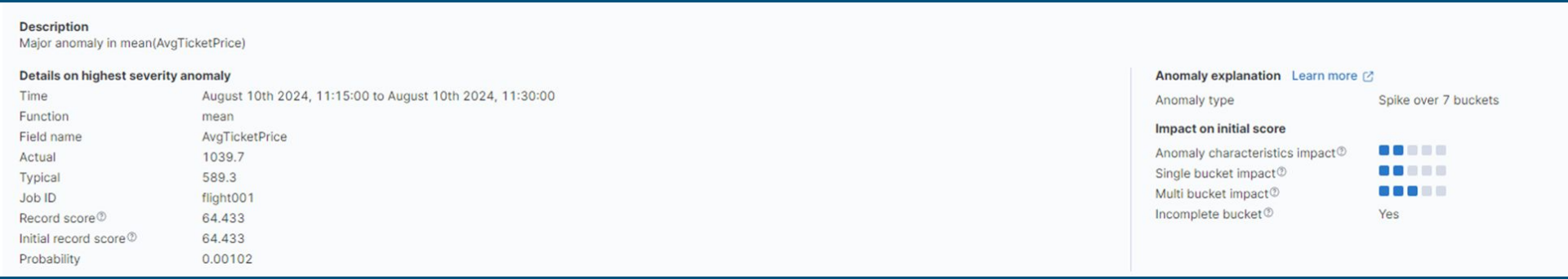
1. Single metric: 單一維度基於時間的異常偵測
2. Multi metric: 多維度基於時間的異常偵測
3. Advanced: Single 和 Multi 的結合
4. Population: 單一維度基於同欄位其他值的異常偵測
5. Categorization: 單一維度基於分組的異常偵測，分組不能太多，一般用於系統 AP 日誌
6. Rare: 異常值出現頻率
7. Geo: 基於地理位置的異常偵測

Elasticsearch ML 實際操作 - Unsupervised

Single metrics: Averaged ticket price

Elasticsearch ML 實際操作 - Unsupervised

Single metrics: Averaged ticket price



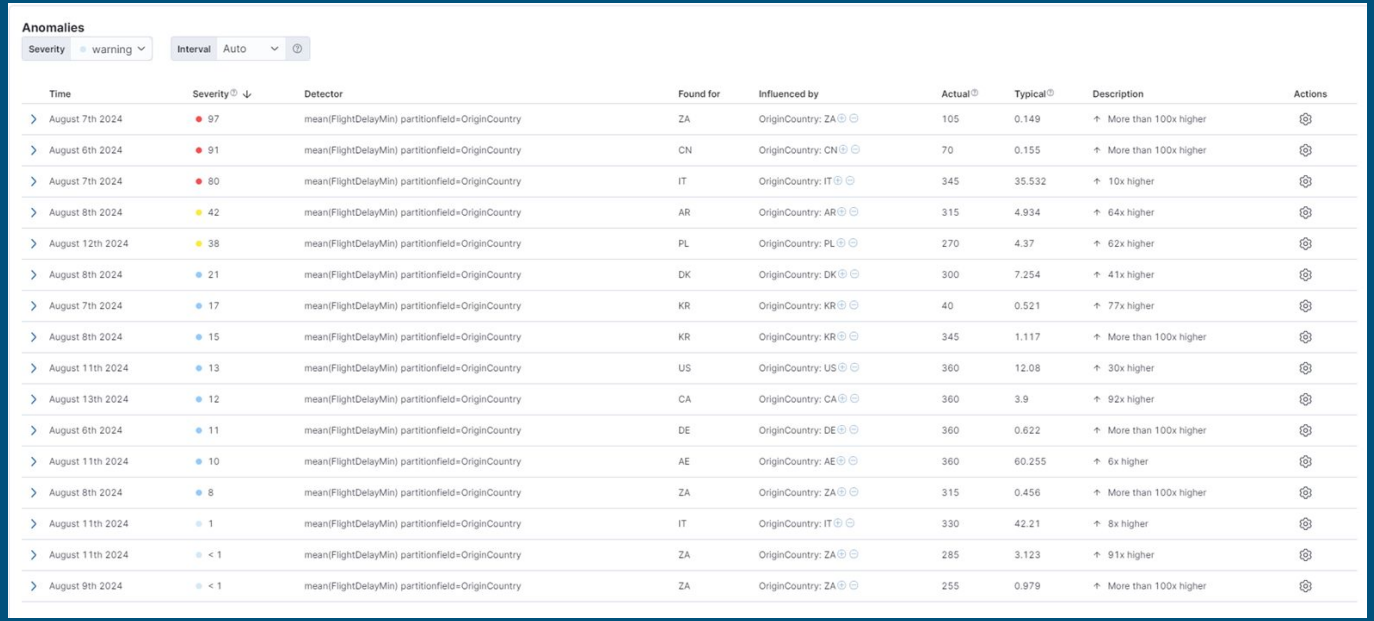
結果說明: 在 8月10日 11:15:00 - 8月10日 11:30:00 間平均票價超過正常值兩倍

Elasticsearch ML 實際操作 - Unsupervised

Mutli metrics: 通過分析 delayed flight minute 找出延遲異常的來源國家

Elasticsearch ML 實際操作 - Unsupervised

Mutli metrics: 通過分析 delayed flight minute 找出延遲異常的來源國家

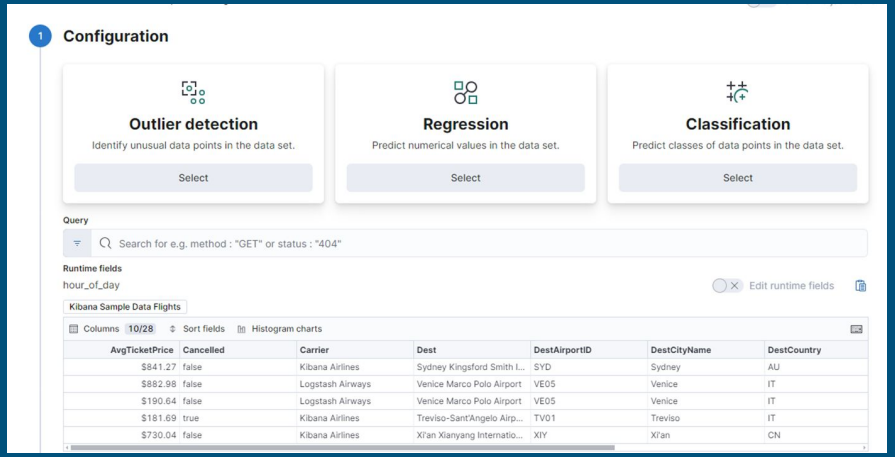


Anomalies

Severity warning Interval Auto

Time	Severity	Detector	Found for	Influenced by	Actual	Typical	Description	Actions
August 7th 2024	97	mean(FlightDelayMin) partitionfield=OriginCountry	ZA	OriginCountry: ZA	105	0.149	More than 100x higher	
August 6th 2024	91	mean(FlightDelayMin) partitionfield=OriginCountry	CN	OriginCountry: CN	70	0.155	More than 100x higher	
August 7th 2024	80	mean(FlightDelayMin) partitionfield=OriginCountry	IT	OriginCountry: IT	345	35.532	10x higher	
August 8th 2024	42	mean(FlightDelayMin) partitionfield=OriginCountry	AR	OriginCountry: AR	315	4.934	64x higher	
August 12th 2024	38	mean(FlightDelayMin) partitionfield=OriginCountry	PL	OriginCountry: PL	270	4.37	62x higher	
August 8th 2024	21	mean(FlightDelayMin) partitionfield=OriginCountry	DK	OriginCountry: DK	300	7.254	41x higher	
August 7th 2024	17	mean(FlightDelayMin) partitionfield=OriginCountry	KR	OriginCountry: KR	40	0.521	77x higher	
August 8th 2024	15	mean(FlightDelayMin) partitionfield=OriginCountry	KR	OriginCountry: KR	345	1.117	More than 100x higher	
August 11th 2024	13	mean(FlightDelayMin) partitionfield=OriginCountry	US	OriginCountry: US	360	12.08	30x higher	
August 13th 2024	12	mean(FlightDelayMin) partitionfield=OriginCountry	CA	OriginCountry: CA	360	3.9	92x higher	
August 6th 2024	11	mean(FlightDelayMin) partitionfield=OriginCountry	DE	OriginCountry: DE	360	0.622	More than 100x higher	
August 11th 2024	10	mean(FlightDelayMin) partitionfield=OriginCountry	AE	OriginCountry: AE	360	60.255	6x higher	
August 8th 2024	8	mean(FlightDelayMin) partitionfield=OriginCountry	ZA	OriginCountry: ZA	315	0.456	More than 100x higher	
August 11th 2024	1	mean(FlightDelayMin) partitionfield=OriginCountry	IT	OriginCountry: IT	330	42.21	8x higher	
August 11th 2024	< 1	mean(FlightDelayMin) partitionfield=OriginCountry	ZA	OriginCountry: ZA	285	3.123	91x higher	
August 9th 2024	< 1	mean(FlightDelayMin) partitionfield=OriginCountry	ZA	OriginCountry: ZA	255	0.979	More than 100x higher	

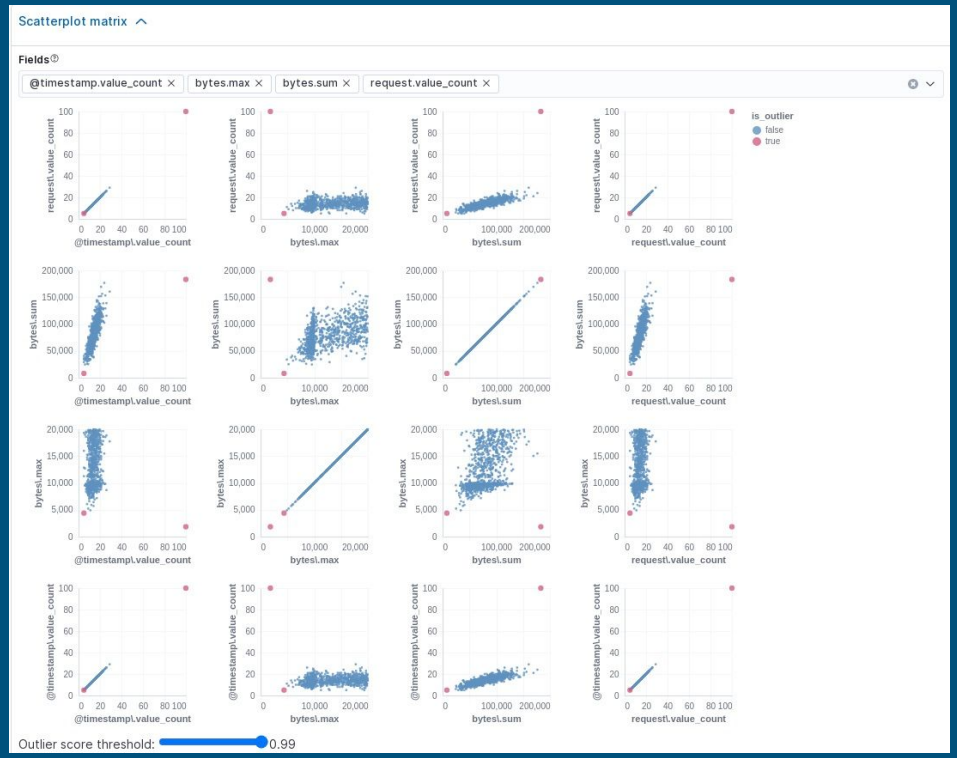
Elasticsearch ML Supervise Model 類型總覽

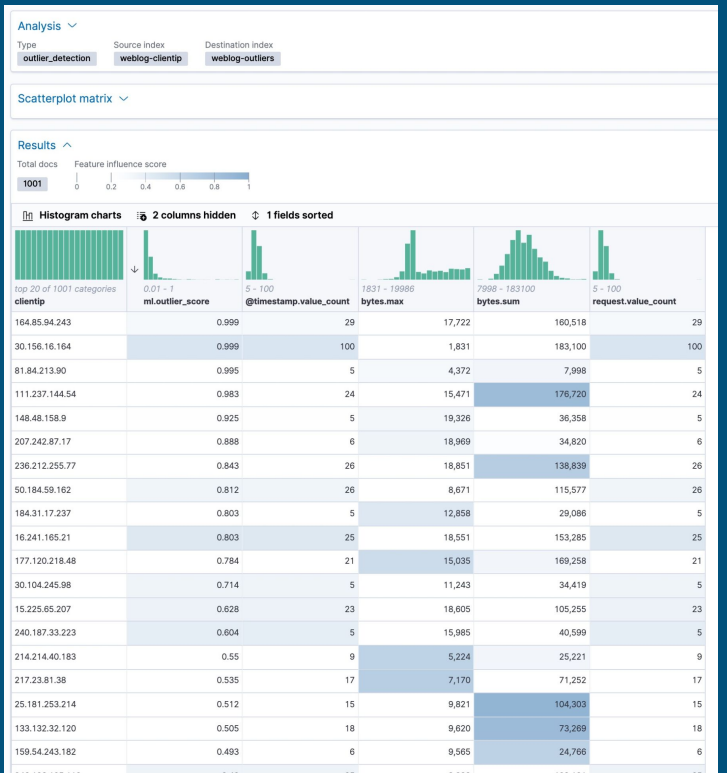


Data Frame Analytics:

1. Outlier detection: 異常政策
2. Regression: 預測
3. Classification: 資料分組

Outlier - Supervised 結果展示





Elasticsearch ML 實際操作 - Supervised 結果展示