

Elasticsearch の連携 による生成 AI の最適 化と最先端オブザーバ ビリティ・セキュリティ



Tokyo

Proprietary



コガ アーロン

Elasticsearch株式会社

Senior Solutions Architect



Agenda

01 Elasticsearch による
グラウンディング

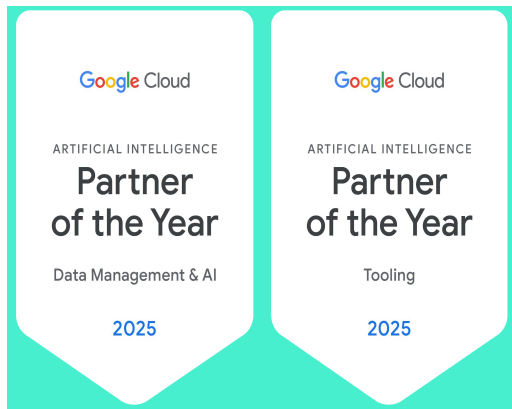
02 LLM・RAG
オブザーバビリティ

03 AI セキュリティ
RAGを活躍した機能

Elastic について



- 2012年設立
- IPO 2018, NYSE: “ESTC”
- 20,000以上の企業での採用実績
- 世界の#1データベース検索エンジン ([DB-Engines](#))
- Developer Community



01. Elasticsearch による グラウンディング

なぜグラウンディングが必要？

LLM は
プライベート データを
持っていない

公開されている情報のみを根拠としていない



グラウンディングにより
社内文書・お客様データ において
フィットした、パーソナライズされた
回答ができる

LLM は
データの更新に追いつかない

データの更新がある度に再学習させる必要があるが、学習は複雑でリソースがかかる



グラウンディングにより
リアルタイムのデータ更新が可能になり、最新情報に対して回答を生成

LLM は
Hallucination してしまふ

どうしても回答するように正しくなくとも回答を作ってしまう



グラウンディングにより
正しい情報に対して回答を生成

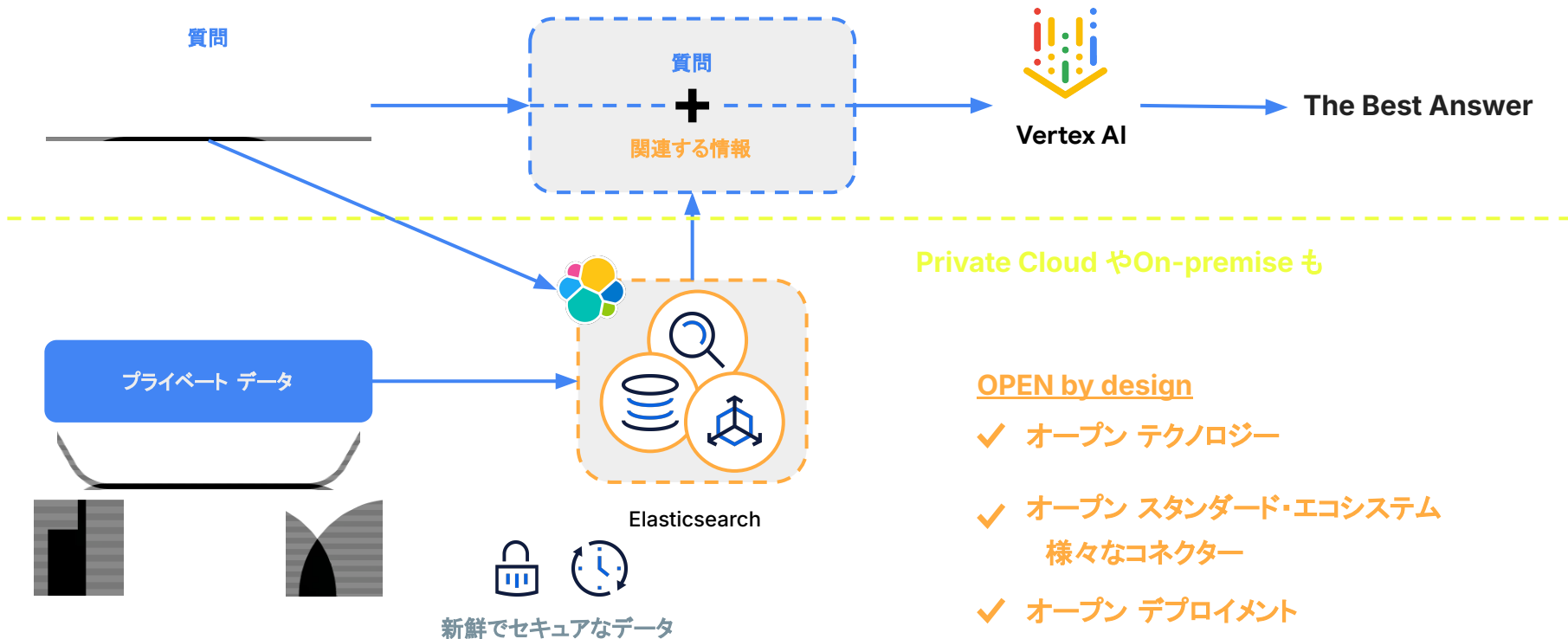
GenAI には
大体の人が慣れていない

一般の人がテキストのキーワード検索を長くやってきたのでサーチバーに慣れている



グラウンディングは
なじみのあるサーチバー のエクスペリエンスを提供しつつ、GenAI の世代へ誘導

Elasticsearch によるグラウンディング



OPEN by design

- ✓ オープン テクノロジー
- ✓ オープン スタンダード・エコシステム
様々なコネクタ
- ✓ オープン デプロイメント

Elasticsearch を選択

根拠づけのカスタマイズ

グラウンディングは、モデル出力を検証可能な情報源と紐付けます。これは、精度と信頼性が重視される場合に有用です。[グラウンディングの詳細](#)

Select a grounding source

- ☐  **RAG Engine**
独自のデータと DIY コンポーネントでグラウンディングできます
- ☐  **Vertex AI Search**
Google が管理する検索エンジンを使用して、独自のデータでグラウンディングできます

☒  **Elasticsearch**
Elasticsearch を使用してグラウンディング

Elasticsearch によるグラウンディング

データをグラウンディングするには、[データのインデックスを作成して（必要に応じてベクトル化も行い）](#) [Elasticsearch](#) に投入します。投入は Elastic Cloud で管理するアトモ、セルフマネ

問題のプロンプト

比較

保存

Clear

コードで構築

ドキュメント

ヒント

を理解し、カスタマイズされたレスポンスを提供するためのコンテキストをモデルに提供します

Prompt Samples

 **Company chatbot**

Create a chatbot for customers with basic company information.

 **Image Editing**

Edit an image of croissants

[プロンプト ギャラリーを参照](#)

プロンプトを記述するか、ボタンを使用して作成します [テキスト作成のヒント](#)

モデル

gemini-2.5-flash

モデルの思考方法をより細かく制御できる標準的な推論モデル

[Switch model](#)

出力

☐ 構造化出力

思考予算
自動

ツール

☐ グラウンディング: Google

Google 検索または Google マップによるグラウンディング。

☐ グラウンディング: 未指定

RAG Engine、Vertex AI Search、Elasticsearch を使用してグラウンディングします。

Elastic Cloud サーバレス



セルフマネージド



Elastic Cloud ホステッド



Elastic Cloud サーバレス

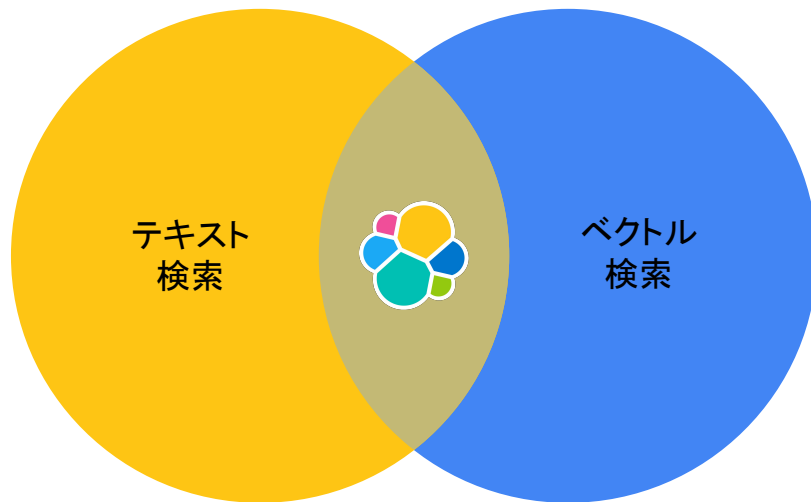
Elastic as a service

- 構築を加速
- One-Click デプロイ
- オートスケーリング
- 利用量に対する課金

* 日本リージョンは対応予定

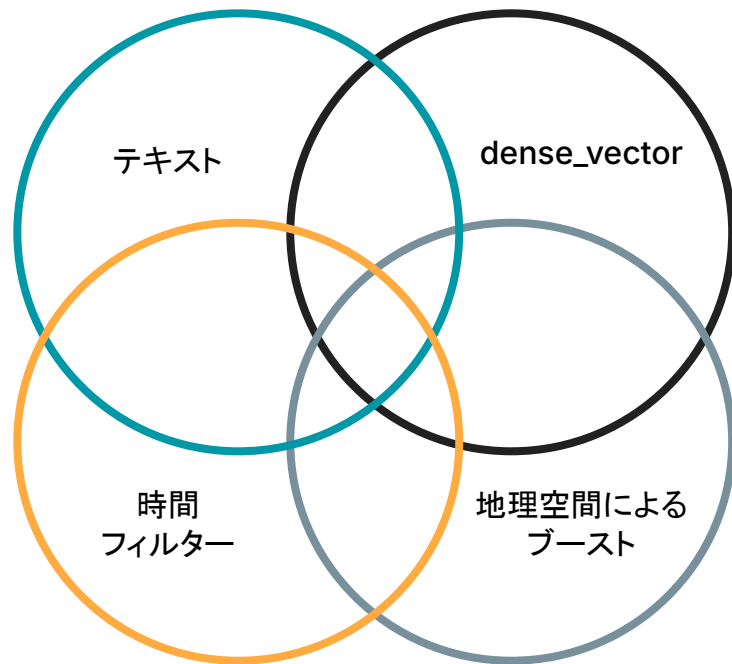
ハイブリッド検索におけるグラウンディング

- キーワード
- 商品ID
- 名前
- 場所
- SKUs
- 類語



- ← 意味
- ← 画像
- ← 意味的類似性

ハイブリッド検索: レストラン検索の例



レストランアプリ

- 「美味しいタイ料理が食べたい」
→ VectorDB: 「選択肢はたくさんあります。」
- 「0.5km以上歩きたくないけど」
→ VectorDB: 「地理空間は何？」
✓ Elastic: 「承知しました、こちらです。」
- 「提案してもらったのは数年前のレビューだけど、最近新しいシェフが採用されたようだから最新のレビューは？」
→ VectorDB: 「時間って？」
✓ Elastic: 「承知しました、こちらです。」

ハイブリッド検索: レストラン検索の例

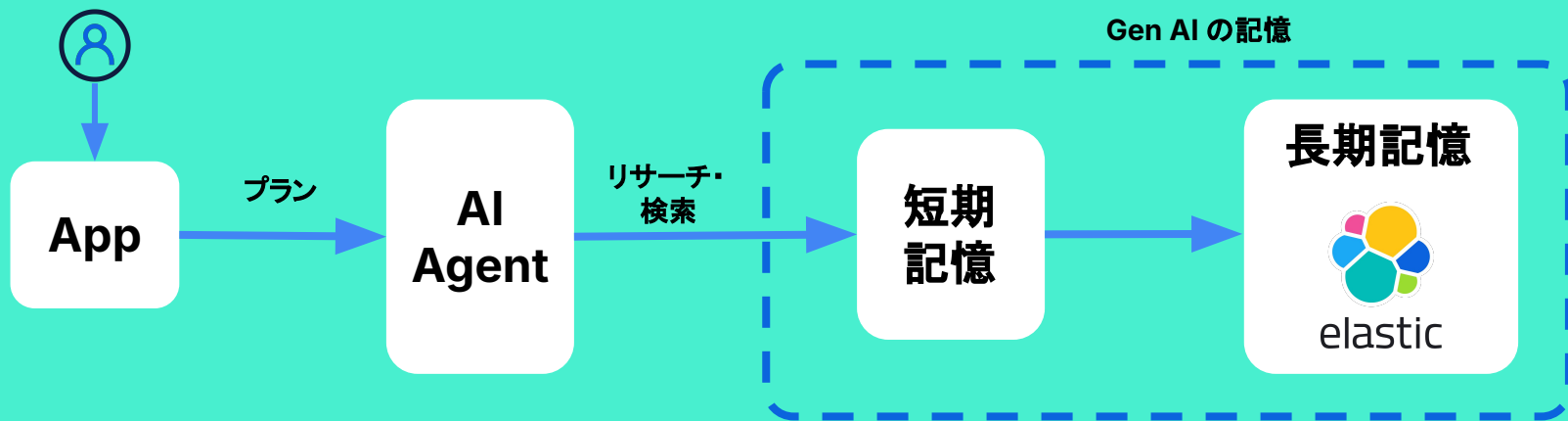


レストランアプリ

- 「美味しいタイ料理が食べたい」
→ VectorDB: 「選択肢はたくさんあります。」
- 「0.5km以上歩きたくないけど」
→ VectorDB: 「地理空間は何？」
✓ Elastic: 「承知しました、こちらです。」
- 「提案してもらったのは数年前のレビューだけど、最近新しいシェフが採用されたようだから最新のレビューは？」
→ VectorDB: 「時間って？」
✓ Elastic: 「承知しました、こちらです。」

期待の結果: 「美味しいタイ料理」+ 「高レビュー」+ 「近い」+ 「最新のレビュー」

AIエージェントの場合



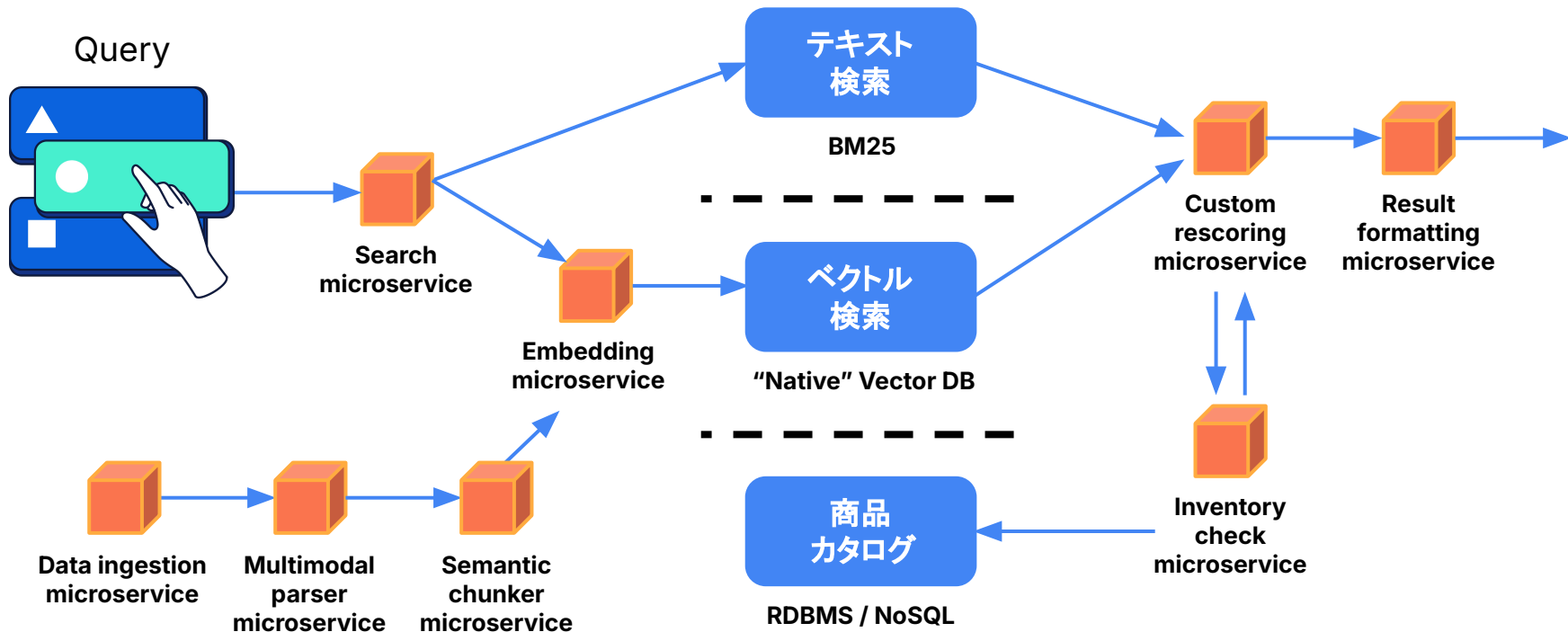
AI Agent の場合AI 主導になる
→ 関連性の高い、正しい情報が重要

昔の検索: とてもシンプルでした

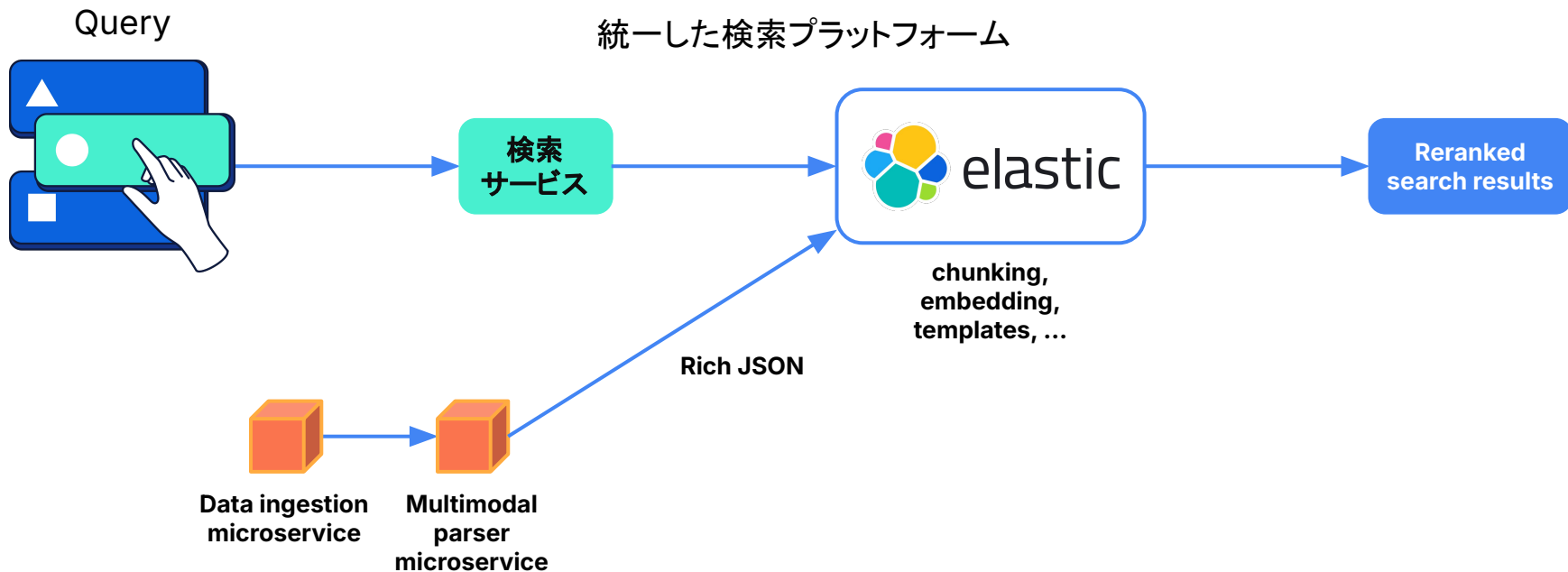
```
GET /my_index/_search
{
  "query": {
    "match": {
      "description": "running shoes"
    }
  }
}
```

複雑になってしまった

複数のデータプラットフォーム

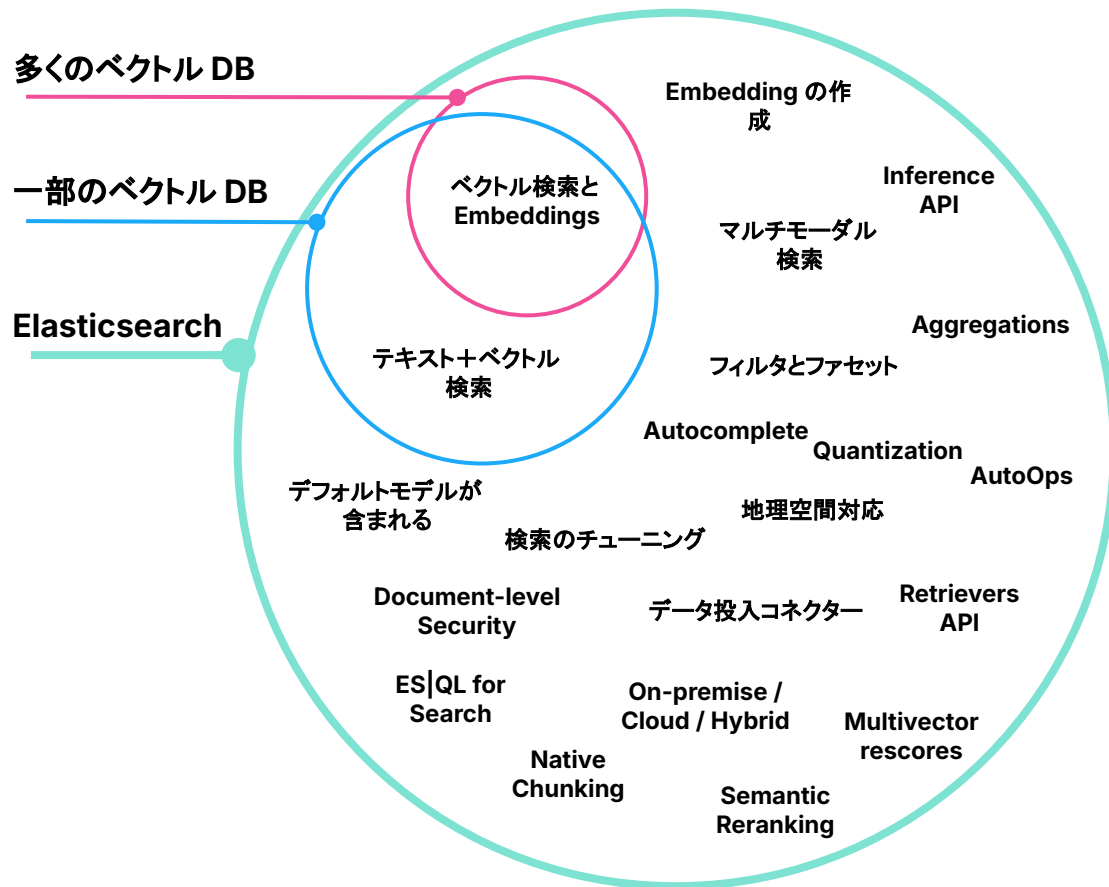


Elasticsearch のご利用においてまたシンプルに



Elasticsearch は常にエン
ハンスをしております。

弊社は Search AI の
プラットフォームです。



Demo: 公園案内のボット

データソース

- 神奈川県オープンデータ

データ一覧

- 公園情報のCSV
- 公園のイメージ

- MLITの公園GISデータ

<https://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-P13.html>

公園

ID	地方公共団体名	名称	所在地	提供開始年	面積	公園種別
2654	神奈川県	境川遊水地公園	神奈川県横浜市泉区下飯田町5-5	2007	261292	総合公園
2664	神奈川県	三ツ池公園	神奈川県横浜市鶴見区三ツ池公園1-1	1957	296800	総合公園
2666	神奈川県	保土ヶ谷公園	神奈川県横浜市保土ヶ谷区花見台4-2	1957	340000	運動公園

トイレ	多目的トイレ	水飲み・手洗い	スプリング遊具・スイング遊具	砂場	複合遊具	健康器具	ブランコ	すべり台	園(風致)
1	1	1	0	0	0	0	0	0	園(風致)
1	1	1	1	1	1	0	1	1	園(カン)
1	1	1	1	1	1	1	1	1	園
1	1	1	1	0	1	0	1	1	園(カン)

備考

河川と公園が連携するという整備手法をとった、初めての県立都市公園です。

日本の「さくら名所100選」に選ばれている公園で、78品種およそ1600本のさくらが来園者に親しまれています。

神奈川県内で初めて造られた運動公園です。

横浜市の市街地にありながら、豊かな里山の自然をそのままに残した公園です。

多摩丘陵の懐かしくも美しい自然を今に伝える公園です。

相模湖に面した豊かな水景色を望む一等地に位置している公園です。

広々とした芝生広場やメタセコイアの並木に囲まれたフランス風庭園などで構成される公園です。

城山ガムの両岸と、戦国時代の山城、津久井城跡を利用した公園です。

横須賀港を望む高台に位置する史跡（三浦按針の墓（安針塚））や風景で名高い公園で、桜の名所としての優れた景観により「かなが」三浦半島の東端に位置する自然に恵まれた公園で、博物館や美術館をはじめ、多彩な施設を持ち、ボランティアによる活動も盛んです。

都心からも程近く、恵まれた眺望は、湘南を代表する人気スポットです。

「湘南」という地名から、多くの方が思い浮かべるイメージを凝縮した公園です

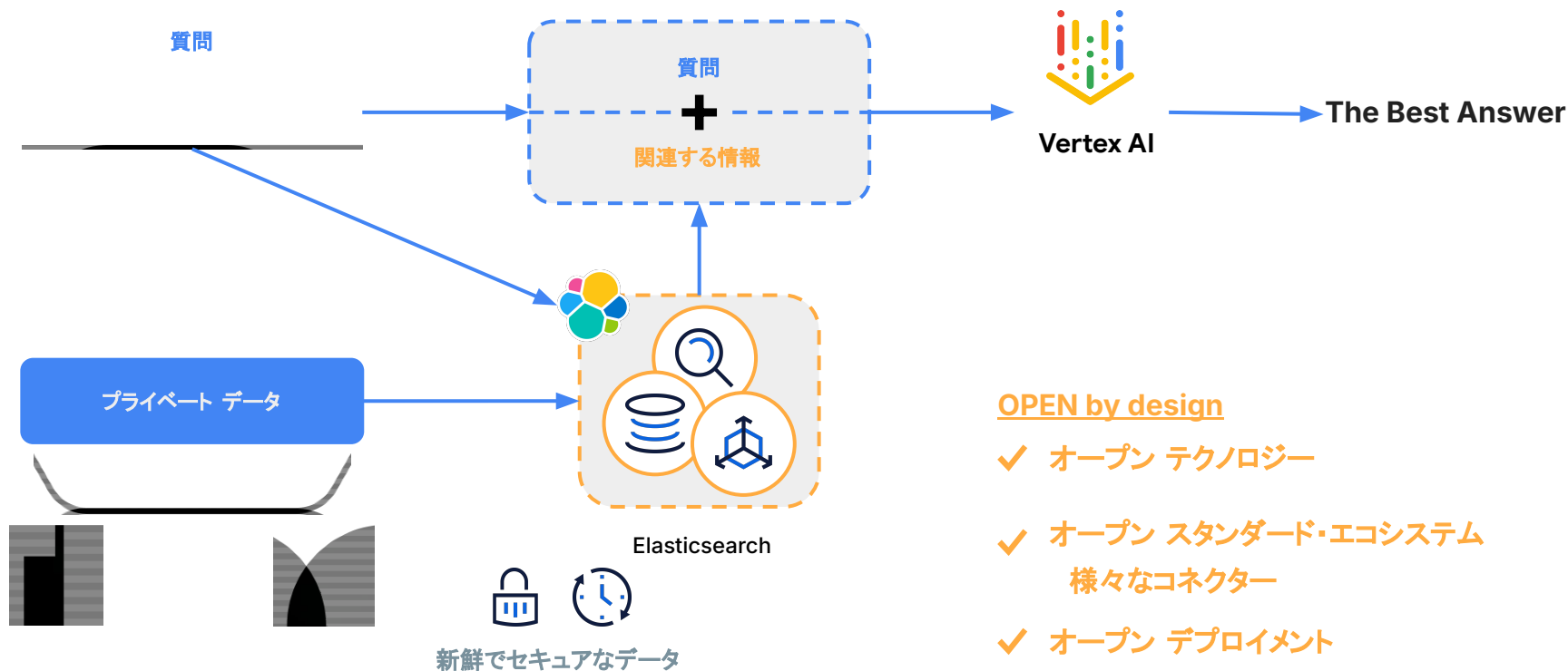
人が集い、遊びながら、里山の自然や季節の花とのふれあいを体験できる公園です。

少年野球場や運動広場、遊具広場のある公園です

里山の風景を残しながら造られた公園です。



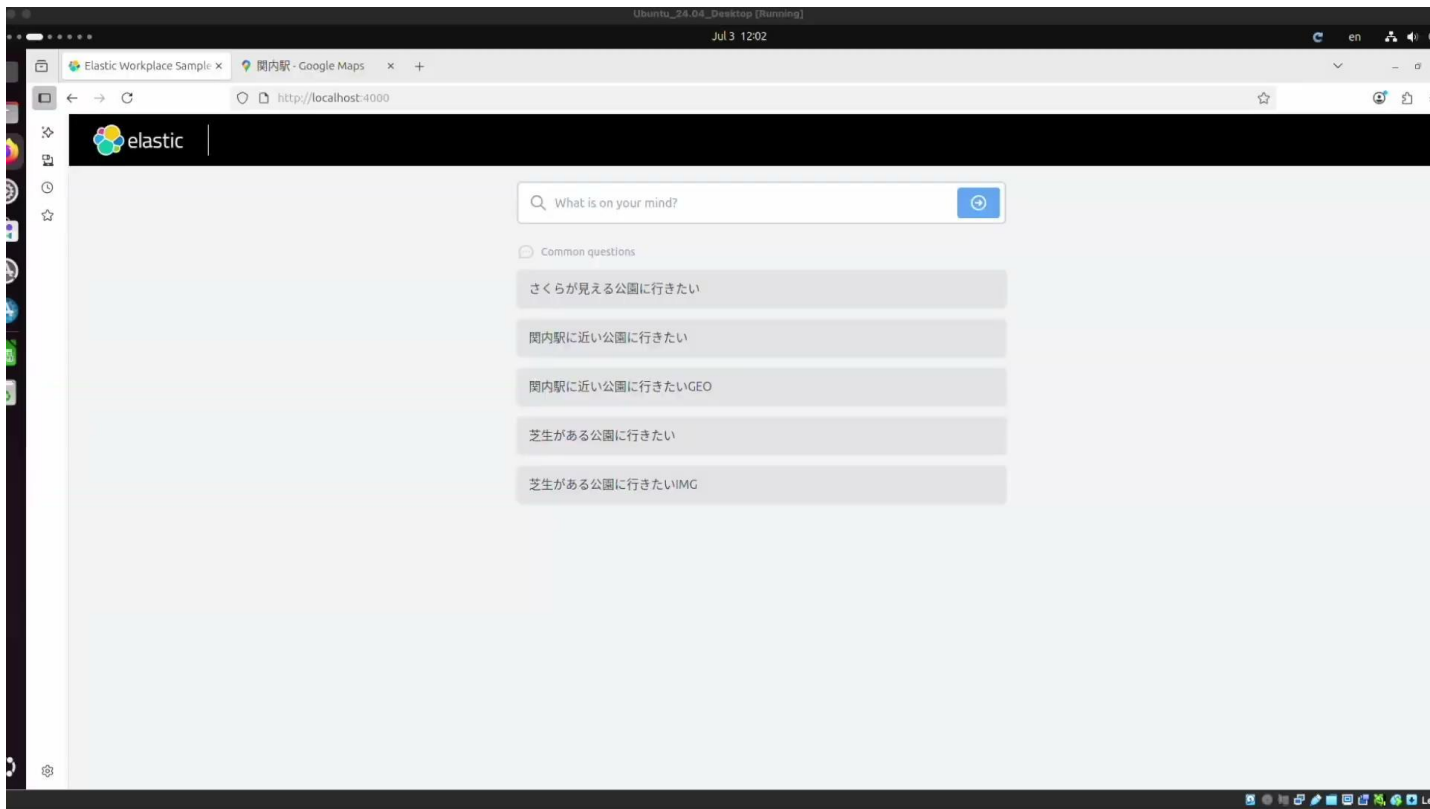
Elasticsearch にデータを導入



OPEN by design

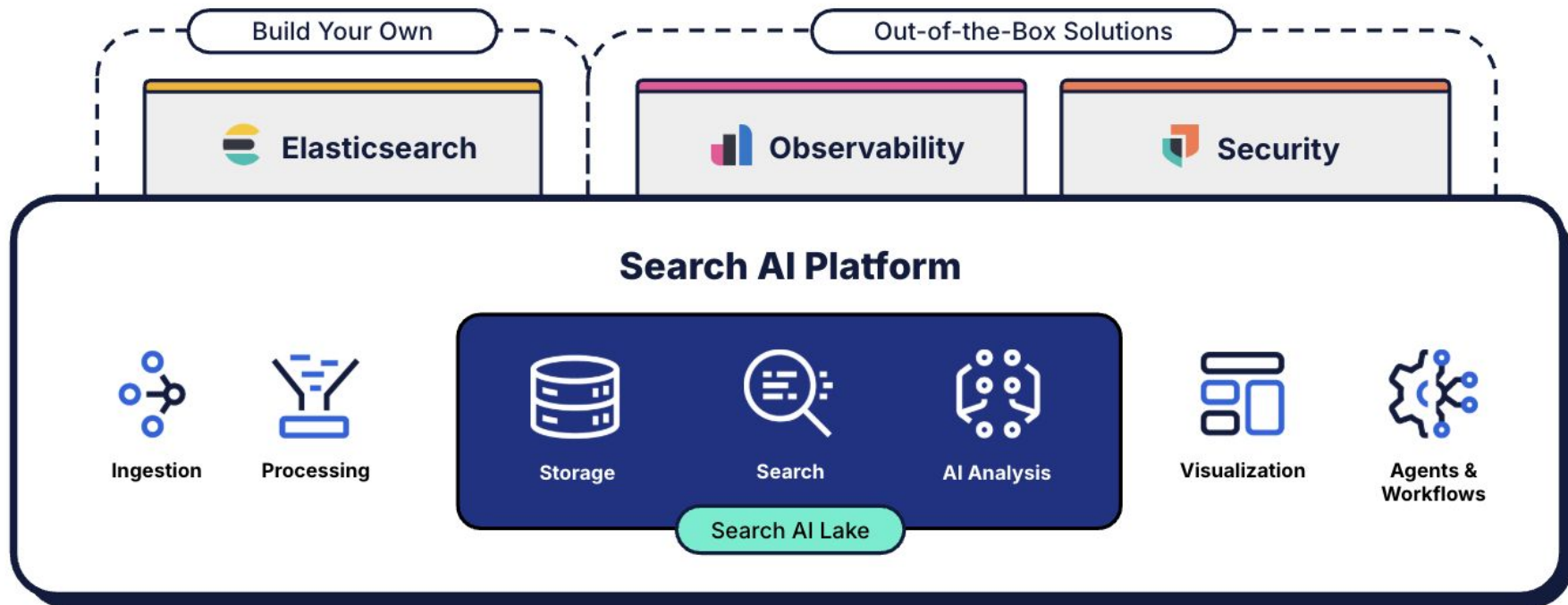
- ✓ オープン テクノロジー
- ✓ オープン スタンダード・エコシステム
様々なコネクタ
- ✓ オープン デプロイメント

DEMO:公園案内のボット



02. LLM・RAG のオブザーバビリティ

The Search AI Company



Vertex AI の LLM オブザーバビリティ

Metrics

- Prediction
- パフォーマンス
- リソース
- Overview



DEMO: RAG アプリのオブザーバビリティ

The screenshot shows the Elastic Spaces management interface in a web browser. The browser's address bar displays the URL: `search-gcp.kb.asia-northeast1.gcp.cloud.es.io/app/management/kibana/spaces/`. The Elastic logo and a search bar are visible at the top of the page. The left sidebar contains a navigation menu with categories like Management, Security, and Observability. The main content area is titled "Spaces" and includes a description: "Organize Kibana into spaces with dedicated navigation, privileges and objects." Below this is a search bar and a table of spaces. The table has columns for Space, Description, Solution view, and Action. One space is listed: "Default" (marked as "current") with the description "This is your default space!". The table also shows "Rows per page: 10" and a pagination control showing "1" of 1 page.

Spaces - Elastic

search-gcp.kb.asia-northeast1.gcp.cloud.es.io/app/management/kibana/spaces/

elastic Find apps, content, and more. Setup guides

Stack Management Spaces

Spaces

Create space

Organize Kibana into spaces with dedicated navigation, privileges and objects.

Search

Space	Description	Solution view	Action
Default current	This is your default space!	Classic	

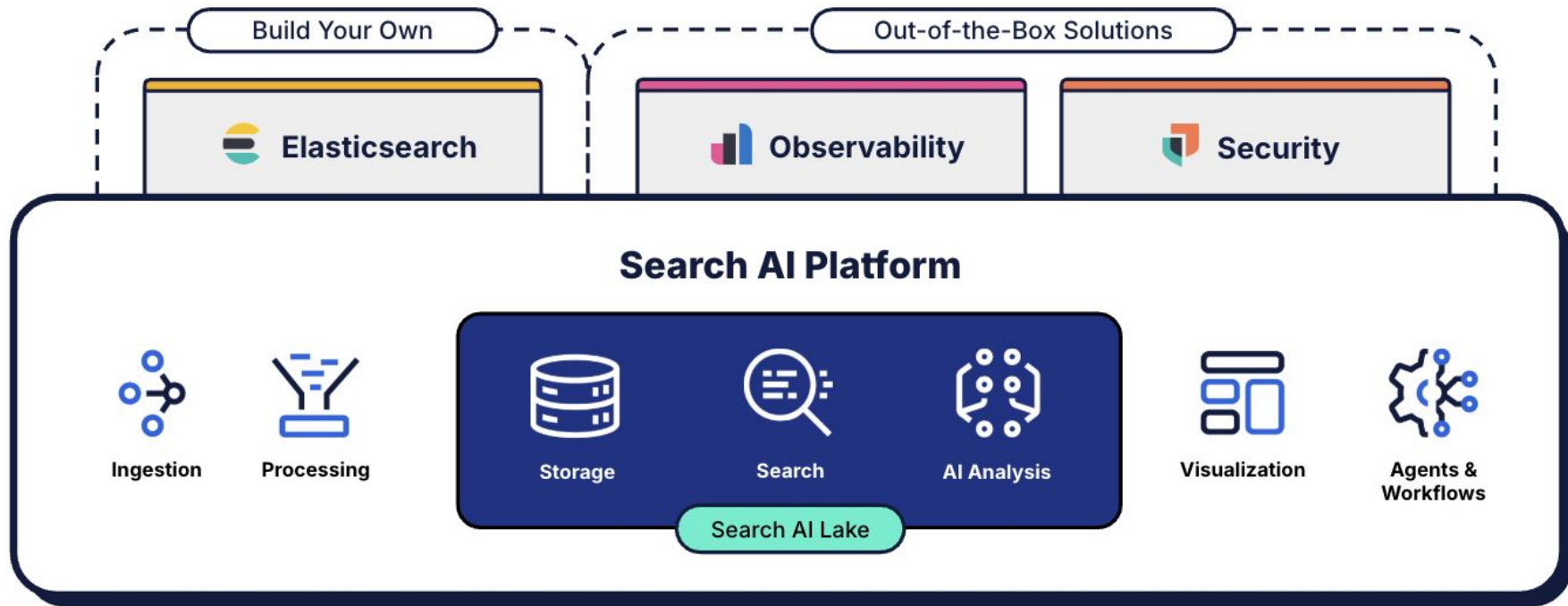
Rows per page: 10 1

LLM オブザーバビリティによる最適化

- LLM のパフォーマンス
- ユーセージ・コストの管理
- アプリのログ・トレース
 - トラブルシューティング
 - アプリの向上

03. AI セキュリティ

The Search AI Company



高度な脅威防御のための実用的なコンテキストの提供

Protect, Investigate & Respond

Bring your data



Extended protection

Optional / Included

Leverage your security protection tools



Native protection

Optional / Included

Block threats with Elastic native protection



AI-driven Security Analytics

Extended protection

Native protection

Threat Research Powered by Elastic Security Labs

DEMO: AI セキュリティ

Security

Discover
Dashboards
Rules
Alerts
Attack discovery
Findings
Cases
Investigations
Intelligence
Explore
Assets

How's the navigation working for you? Missing anything?
Let us know

Get started
Developer tools

Deployment / Alerts

ML job settings Add integrations Data view Update available

Filter your data using KQL syntax Last 1 day Refresh

Alerts

Assignees Manage rules

Status open 1 Severity User Host

Summary Trend Counts Treemap

Severity levels

Levels	Count
Critical	2
High	4
Low	102

108 alerts

Alerts by name

Rule name	Count
Account or Group Discovery via Built-In T...	50
Linux System Information Discovery	45
System Network Connections Discovery	3
Document behaves abnormally	2

Top alerts by

host.name

host.name	Count
srv-nix-es-rsa	92.6%
srv-win-s1-rsa	2.8%

108 alerts 12 groups Group alerts by: Rule name



ご清聴ありがとうございました。