

# AI活用のためのデータモダナイゼーション ～レガシーシステムのデータを活かす実践方法～

株式会社セゾンテクノロジー

## 株式会社 セゾンテクノロジー プロダクトセールスエンジニアリング部

瓦井 美由紀 Miyuki Kawarai



金融系SIerで電力会社の共済システムにて  
SI開発（Webシステム）を担当。HULFT 

某銀行の情報系システムにて  
PJ管理に従事。構築・運用に携わる。  
HULFT 

HULFTやDataSpider、HULFT Squareなど  
プロダクトコンテンツ企画・制作を担当

HULFT 

HULFT  
SQUARE

DataSpider   
Servista

- データモダナイゼーションとは
- データモダナイゼーションにおけるハードル
- 具体的な解決方法
- プロダクト・サービスでできること

# データを「過去の記録」から 「未来を動かす戦略的資産」へと転換すること

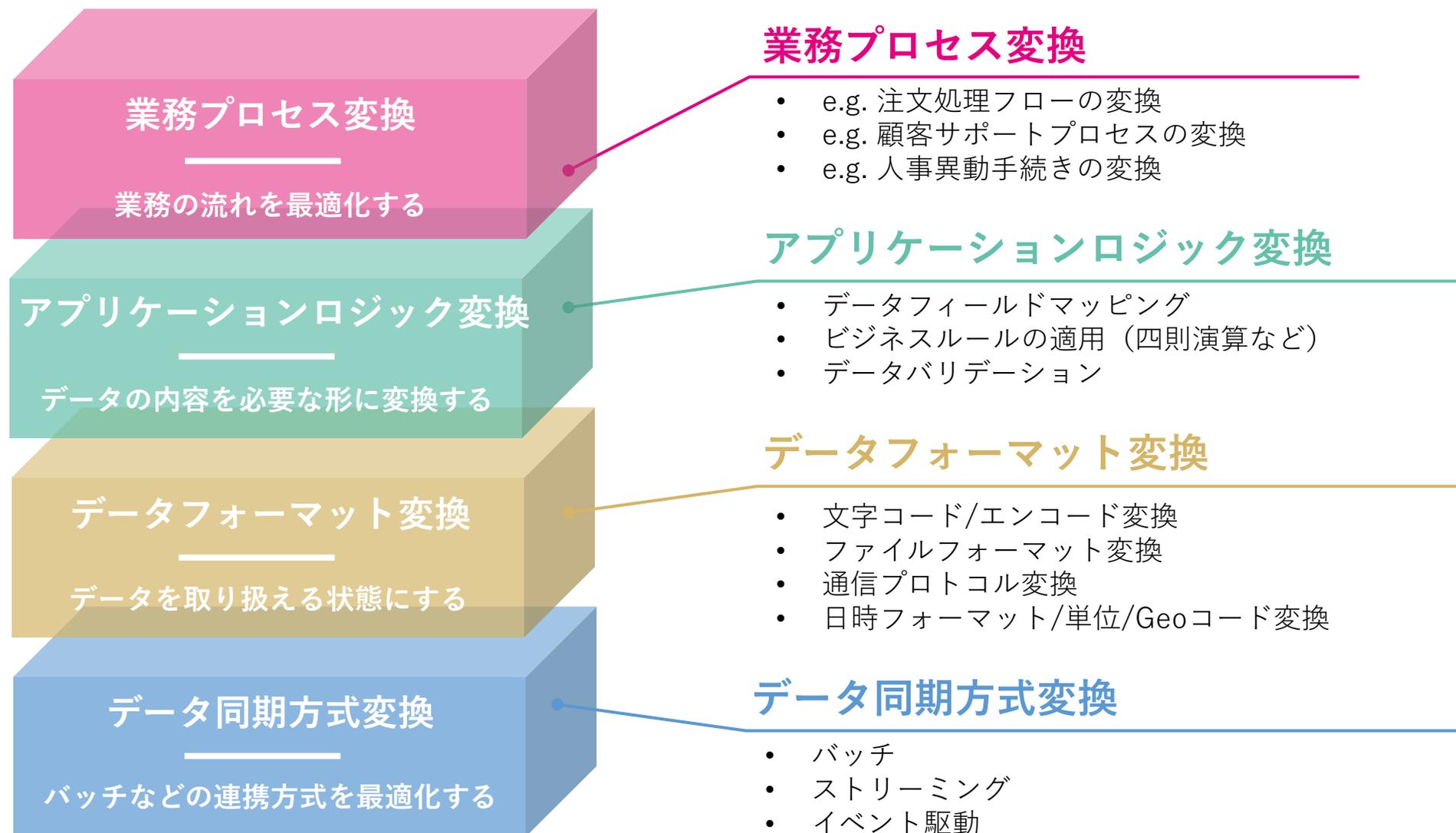
## 1. リアルタイムな現状把握

“ダッシュボードとレポート”  
複雑なデータを直感的に可視化し、  
組織全体で同じ事実に基づく意思決定。  
データの民主化を推進。

## 2. 高度な未来予測

“生成AI・機械学習の組み込み”  
人間が見逃すパターンをデータから抽出し、  
需要予測、行動分析、異常検知など  
未来のビジネス行動を支援。

# データモダナイゼーションの担う部分



# データモダナイゼーションの担う部分

## 業務プロセス変換

業務の流れを最適化する

### 業務プロセス変換

- e.g. 注文処理フローの変換
- e.g. 顧客サポートプロセスの変換
- e.g. 人事異動手続きの変換

## アプリケーションロジック変換

データの内容を必要な形に変換する

### アプリケーションロジック変換

- データフィールドマッピング
- ビジネスルールの適用（四則演算など）
- データバリデーション

## データフォーマット変換

データを取り扱える状態にする

### データフォーマット変換

- 文字コード/エンコード変換
- ファイルフォーマット変換
- 通信プロトコル変換
- 日時フォーマット/単位/Geoコード変換

## データ同期方式変換

バッチなどの連携方式を最適化する

### データ同期方式変換

- バッチ
- ストリーミング
- イベント駆動

# システムモダナイゼーションとの違い

## システムモダナイゼーション



### データは後から使うもの

#### レガシー脱却

- ✓ クラウド移行
- ✓ サーバーレス環境構築

#### モダン技術

- ✓ 生成AI活用
- ✓ ダッシュボード分析

目的 システムを変える（新たに作る）



## データモダナイゼーション



### データは使う前提で整える

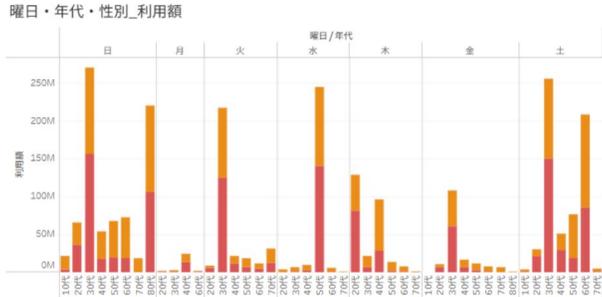
- ✓ 基幹システムは止めなくてもいい
- ✓ 基幹システムの運用も変えなくてもいい
- ✓ 基幹システムのデータを供給する

目的 データの使い方を変える（データ活用）

# データモダナイゼーションを実現している世界

## (活用例) クレジットカード利用情報を分析

クレジットカード利用状況ダッシュボード



## (自社事例) 社内イベント参加者数をグラフ表示

設定一覧:

New Chat +

search

- 本部ごとの社内イベン...
- InfoSquareでいいね数...
- InfoSquareで活動が最...
- 最もいいね数が多い上...
- 当社とまだ取引が無い...
- 社員の人数を教えてく...
- 社内イベントの参加者...
- 当社とまだ取引が無い...
- HULFT認定資格の保有...

質問をどうぞ

本部ごとの社内イベント参加者数をグラフで表示してください。

SQLを自分で書く

SQLを入力

GPTの回答作成過程を表示

実行

質問を入力してください

質問を入力してください

BI×データ分析基盤



集約されたデータをツールで可視化

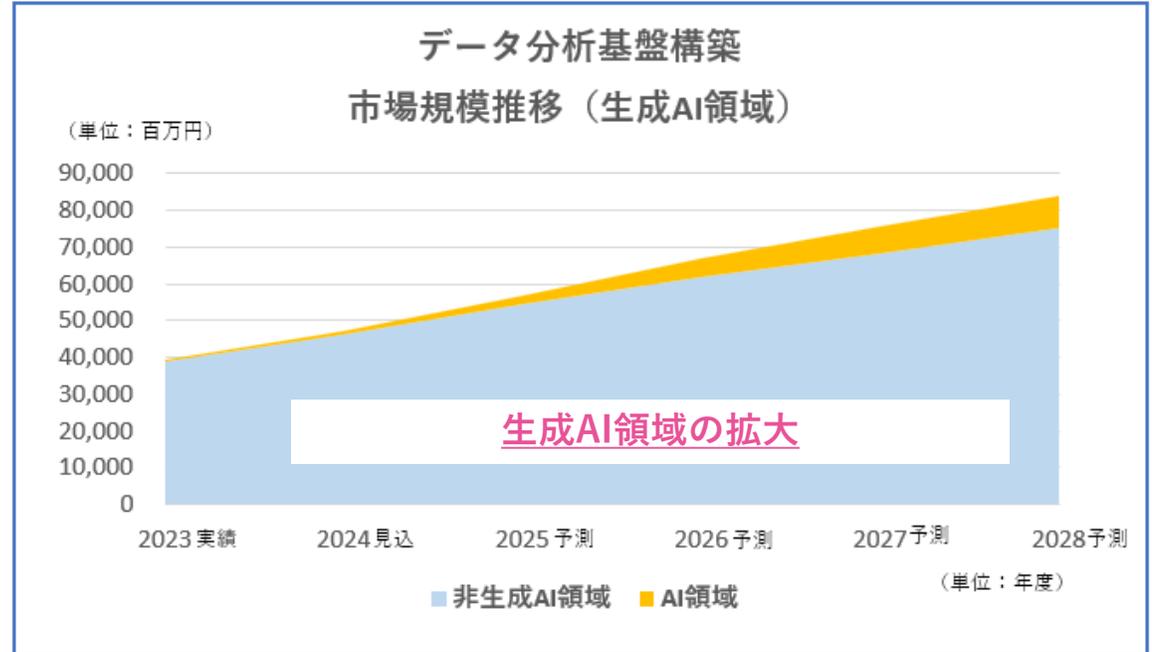
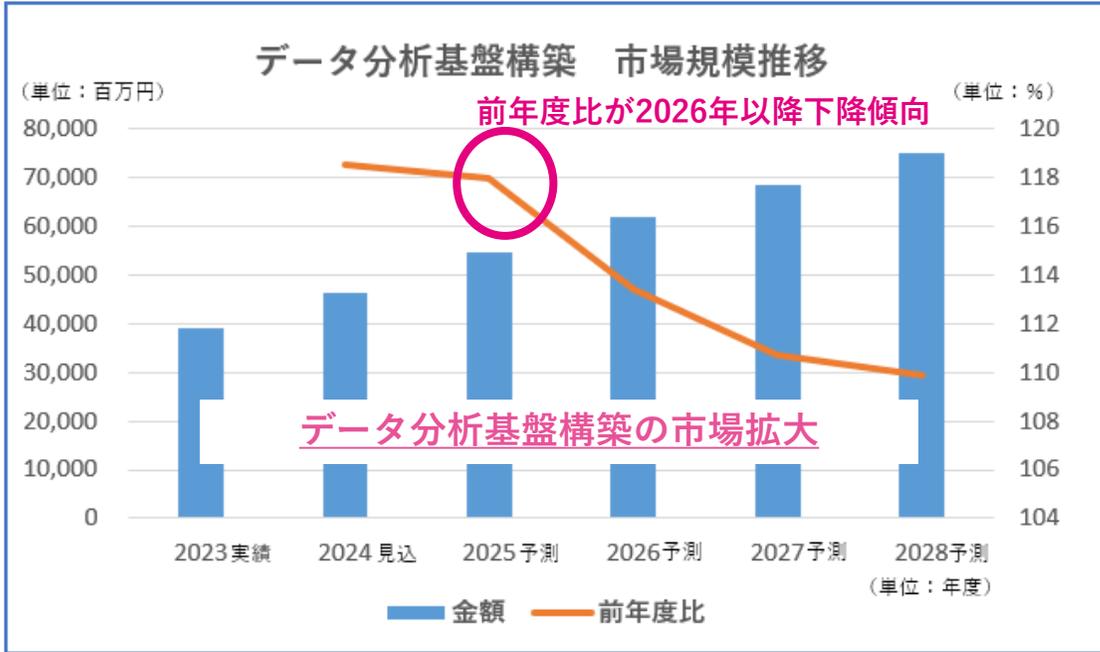
生成AI×データ分析基盤



自然言語でデータを探索し同時に可視化

# データ分析基盤の必要性

データ分析基盤構築の市場規模は拡大しており、生成AI領域の比率も上昇しています。



データ分析基盤構築

年平均成長率 (CAGR) **14.0%**

※CAGR算出の対象期間：2023年度～2028年度予測の6年間

(内) 生成AI領域

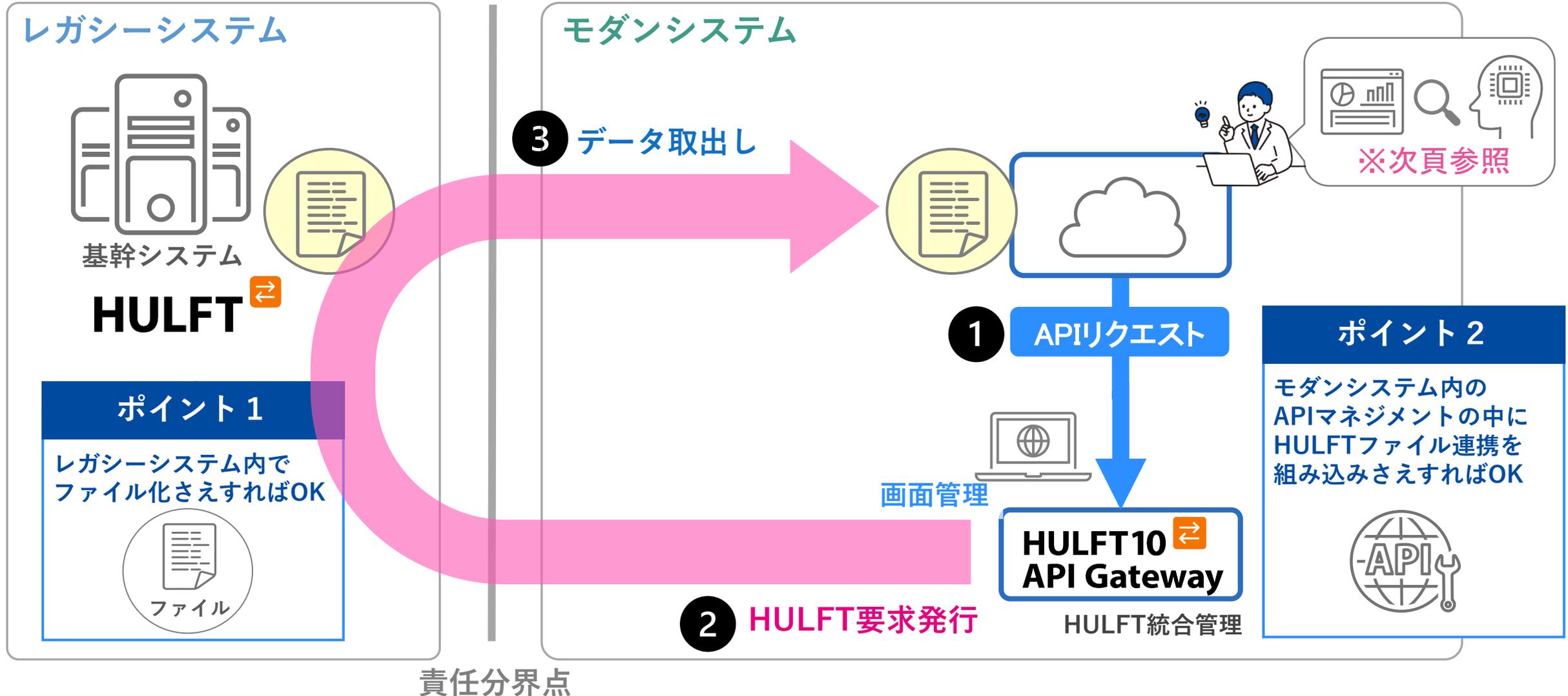
年平均成長率 (CAGR) **114.1%**

※CAGR算出の対象期間：2023年度～2028年度予測の6年間

引用：富士キメラ総研「2025 生成AI/LLMで飛躍するAI市場総調査  
～注目度が高まる生成 AI / LLMビジネス市場の将来展望  
および有望分野/各種トレンドの徹底調査～」  
データ分析基盤構築 2.市場規模推移より

# データモダナイゼーションの第一歩

## モダンシステム側から基幹システムのデータを操作する



# データモダナイゼーションにおけるハードル

# モダナイゼーションできている世界

## ロードマップ：モダンデータスタックからレガシーまで

オンプレミス環境



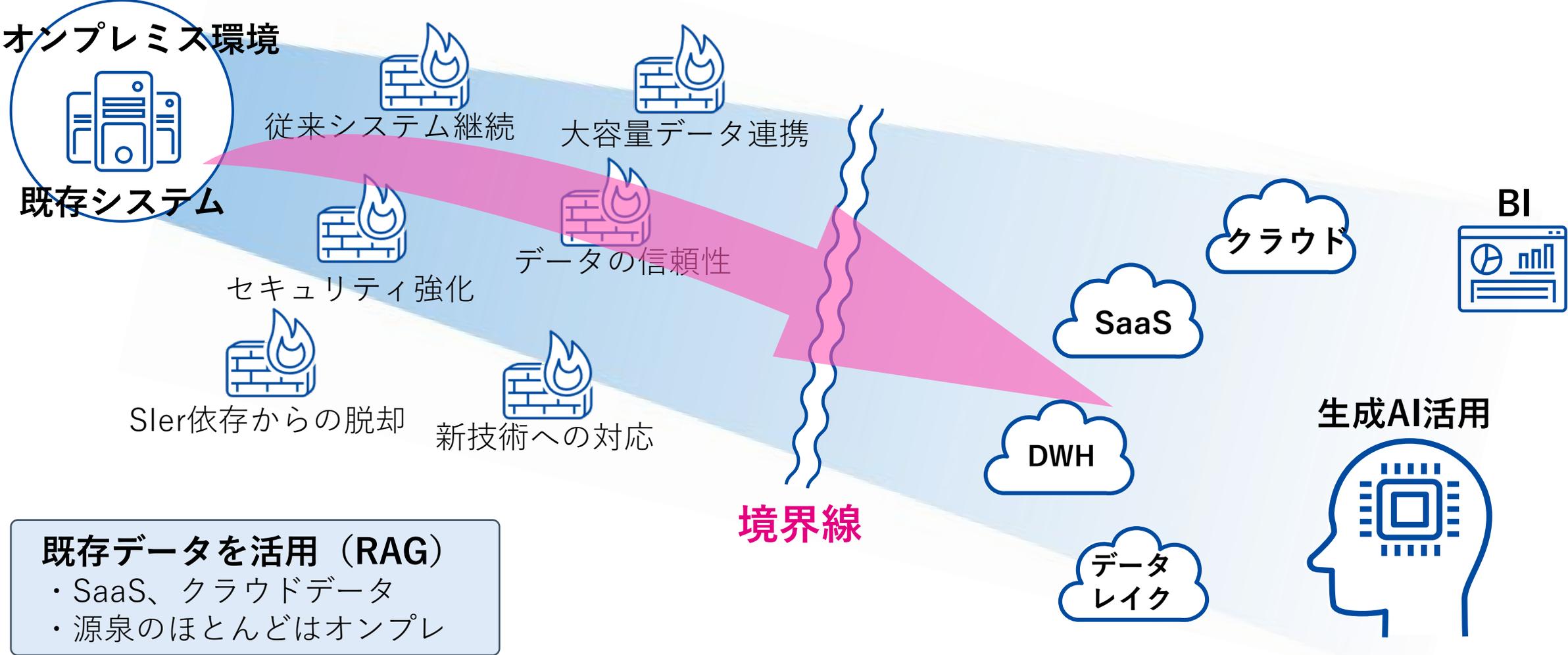
境界線



既存データを活用 (RAG)  
・ SaaS、クラウドデータ  
・ 源泉のほとんどはオンプレ

# モダナイゼーションできている世界

## レガシーシステムのデータを自由に扱えなければならない



# レガシーシステムを取り巻く現状

## ■レガシーシステムで困っていませんか？

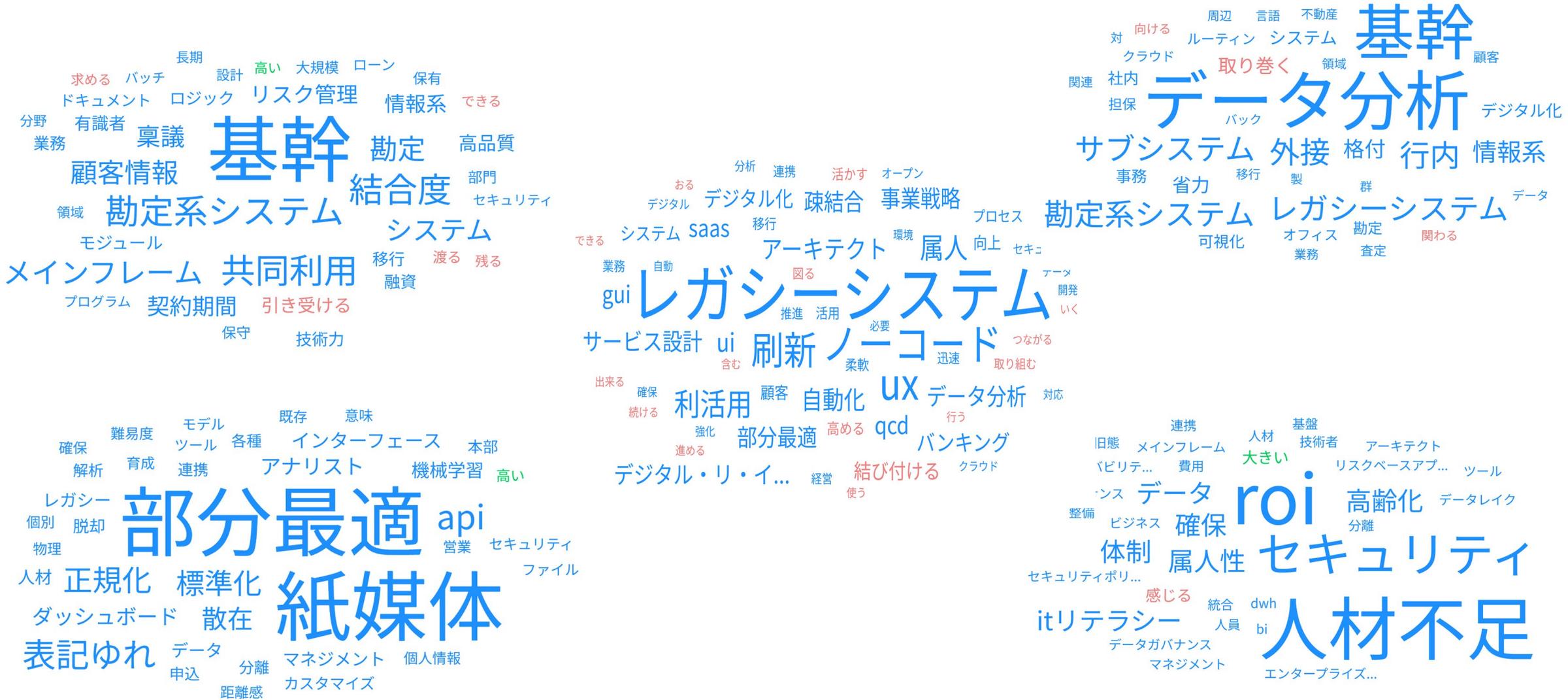
レガシーシステム



基幹システム

- ✓ コストが高い
- ✓ 開発に時間がかかる
- ✓ 人材不足（COBOL開発者不足）
- ✓ DX推進を遅らせている

# アンケート結果



# データモダライゼーションに必要なポイント ～具体的な解決方法～

# データモダナイゼーションに必要な3つのポイント

## 1 データ抽出

- レガシーデータ抽出
- 非構造化データ対応

## 2 データ変換

- 文字コード変換
- 異なるI/F
- 異なるフォーマット

## 3 データ品質

- ファイルの整合性
- データの整合性

# データモダナイゼーションに必要な3つのポイント

## 1 データ抽出

- レガシーデータ抽出
- 非構造化データ対応

## 2 データ変換

- 文字コード変換
- 異なるI/F
- 異なるフォーマット

## 3 データ品質

- ファイルの整合性
- データの整合性

# ①データ抽出（レガシーデータ抽出）

01

## 必要なデータをリアルタイムにオンデマンド連携

責任分界点

レガシーシステム



基幹システム



データ抽出方法の検討から大変で時間がかかる



担当

データ連携開発は不要。  
データ準備のみでいい。

モダンシステム



早急に状況分析したいのにデータがなかなか来ない...



管理者



APIリクエストによる  
レガシーデータのオンデマンド連携



SaaS

法規制等の外部要因



# 1 データ抽出（非構造化データ対応）

02

## 非構造化データの整備により生成AIの回答精度向上



お客様

### 問い合わせ

〇〇製品のベーシックプランの価格を教えてください。

### 回答

〇〇製品のベーシックプランの価格はXX円です。ベーシックプランは～であり、～できます。

価格表(PDF)



価格QA表(CSV)



### 生成AI

LLM



データソース



Vector DB



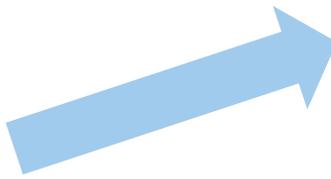
非構造化データを  
構造化データに変換整備

# ①データ抽出（非構造化データ対応）

## ①非構造化データ（PDF形式）の場合

非構造化データ（PDF形式）から構造化データ（CSV形式）に整備することで、生成AIが活用しやすい状態にデータを整備し、生成AIの回答結果精度を向上させる。

CSV	質問	回答	メタデータ
Column1	Column2	Column3	Column4 Column5
172	HULFT Series製品は1つの製品につき何台のOSで使用でき...	HULFT Series製品は、1つの製品を1つのオペレーティングシステム...	{%source%:"price_HULFT10_202412.pdf%" %"page%":3 %"section%":%ライセンス構成について%}"
173	クラスタ環境でHULFTを使用する場合のライセンス要件は...	クラスタ運用・VM・ホットスタンバイ・コールドスタンバイ・マルチ...	{%source%:"price_HULFT10_202412.pdf%" %"page%":3 %"section%":%ライセンス構成について%}"
174	HULFT10のAIX、Linux、Windows対応製品のグレード構成...	AIX、Linux、zOS対応の製品は標準グレードは「Standard」、拡張グ...	{%source%:"price_HULFT10_202412.pdf%" %"page%":3 %"section%":%グレードについて%}"
175	IBMi版のグレード設定はありますか？	IBMi版につきましては、グレード設定はありません。	{%source%:"price_HULFT10_202412.pdf%" %"page%":3 %"section%":%グレードについて%}"
176	待機系ライセンスの価格はどのように設定されていますか？	災害対策（DR）などを目的とした待機系環境に導入するライセンス...	{%source%:"price_HULFT10_202412.pdf%" %"page%":3 %"section%":%待機系ライセンスについて%}"
177	月額利用ライセンスとは何ですか？	特定期間に限りライセンスを、利用用途に応じて月額で利用できるラ...	{%source%:"price_HULFT10_202412.pdf%" %"page%":3 %"section%":%月額利用ライセンスについて%}"
178	移行期間ライセンスの特徴を説明してください。	保有しているHULFT Seriesのライセンスの使用権の数を、特定期間に...	{%source%:"price_HULFT10_202412.pdf%" %"page%":3 %"section%":%移行期間ライセンスについて%}"
179	移行期間ライセンスの購入単位と料金について教えてください。	ヶ月単位で購入できます。料金は各製品の月額利用料金をご確認くだ...	{%source%:"price_HULFT10_202412.pdf%" %"page%":3 %"section%":%移行期間ライセンスについて%}"
180	移行期間ライセンスの開始日と終了日の設定はどうなってい...	開始日は1日固定、終了日は月末固定となります。お申込期間が2ヶ月...	{%source%:"price_HULFT10_202412.pdf%" %"page%":3 %"section%":%移行期間ライセンスについて%}"
181	期間限定ライセンスとは何ですか？	開発用などで特定期間に限りファイル転送を行うことができる、「期...	{%source%:"price_HULFT10_202412.pdf%" %"page%":3 %"section%":%期間限定ライセンスについて%}"
182	期間限定ライセンスの購入単位と料金体系を教えてください。	ヶ月単位で購入できます。料金は各製品の月額利用料金をご確認くだ...	{%source%:"price_HULFT10_202412.pdf%" %"page%":3 %"section%":%期間限定ライセンスについて%}"
183	期間限定ライセンスの開始日と終了日の設定はどうなってい...	開始日は1日固定、終了日は月末固定となります。	{%source%:"price_HULFT10_202412.pdf%" %"page%":3 %"section%":%期間限定ライセンスについて%}"



### PDF

#### ライセンス構成について

- ◆ 基本方針について
  - ・ HULFT Series製品は、1つの製品を1つのオペレーティングシステム上で使用することができます。
  - ・ HULFT Series製品を、クラスタ運用・VM・ホットスタンバイ・コールドスタンバイ・マルチクラスタ・PR/SM・並列シブレックス等の環境でご利用になる場合、稼働オペレーティングシステム分の製品を購入して頂きます。
- ◆ グレードについて
  - ・ HULFT10のAIX、Linux、Windows対応の各製品には標準搭載される機能に応じた2種類の【グレード】があります。AIX、Linux、zOS対応の製品は標準グレードは「Standard」、拡張グレードは「Enterprise」となります。
  - ・ Windows対応の製品は標準グレードは「Client」、拡張グレードは「Server」となります。
  - ・ IBMi版につきましては、グレード設定はありません。

製品名称	機能					
	基本機能 (ファイル転送 関連機能)	クラスタ対応機能	フィルトリガ機能	ネットワークファイル 対応	スケジューラ	圧縮
HULFT10 for Windows-Server	○	○	○	○	○	○
HULFT10 for Windows-Client	○	-	-	-	-	○
HULFT10 for Linux-Enterprise	○	○	○	○	-	○
HULFT10 for Linux-Standard	○	-	-	-	-	○
HULFT10 for AIX-Enterprise	○	○	-	-	-	○
HULFT10 for AIX-Standard	○	-	-	-	-	○
HULFT10 for zOS-Enterprise	○	-	-	-	-	○※
HULFT10 for zOS-Standard	○	-	-	-	-	-

※ zEDCによるDEFLATE圧縮

- ✓ PDFに含まれるテキストや表形式のデータをQA形式に変換
- ✓ メタデータを付与  
(例) ドキュメント名、ページ名、セクション名

# ①データ抽出（非構造化データ対応）

## ①非構造化データ（手書きスキャン）の場合

非構造化データ（手書きスキャン）から構造化データ（CSV形式）に整備することで、生成AIが活用しやすい状態にデータを整備し、生成AIの回答結果精度を向上させる。

第 **テキスト抽出前** 令和 7 年 3 月 31 日  
千代田区長殿 助成金交付申請書

申請者

所在地	(〒160-0053)
会社名	株式会社世存情報
代表者役職・氏名	信自 太郎
電話番号	03-1111-1111

※助成金交付請求時は押印が必要です。(インク浸透印は不可)

下記のとおり千代田区低炭素建築物助成金について、千代田区低炭素建築物助成金交付要綱第5条の規定に基づき、関係書類を添えて申請します。

記

### 1 概要

建物名称	情報本社ビル		
建築場所	千代田区上海3-1-2		
延べ面積	3000㎡	敷地面積	1500㎡
着工予定日	令和7年4月15日	完了予定日	令和8年2月23日
スケルトン部分工事の有無	有	スケルトン部分工事着工予定日	令和7年5月15日
		スケルトン部分工事完了予定日	令和7年10月1日
連絡担当者(施工業者等)	住所 上海市長寧区金鐘路968号 電話 021-1111-1111 氏名 信自 花子		

※スケルトン部分工事有の場合は、スケルトン部分工事の着工予定日及び完了予定日を記入してください。



AI-OCRのように  
帳票レイアウトや各項目  
属性の設定をしなくても  
テキスト抽出可能

**テキスト抽出後**  
令和7年3月31日

### 助成金交付申請書

#### 申請者

- 所在地: 〒160-0053
- 会社名: 株式会社世存情報
- 代表者役職・氏名: 信自 太郎
- 電話番号: 03-1111-1111

#### 1. 概要

- 建物名称: 情報本社ビル
- 建築場所: 千代田区上海3-1-2
- 延べ面積: 3000㎡
- 敷地面積: 1500㎡
- 着工予定日: 令和7年4月15日
- 完了予定日: 令和8年2月23日
- スケルトン部分工事: 有
  - スケルトン部分工事着工予定日: 令和7年5月15日
  - スケルトン部分工事完了予定日: 令和7年10月1日
- 連絡担当者:
  - 住所: 上海市長寧区金鐘路968号
  - 電話: 021-1111-1111
  - 氏名: 信自花子

# データモダナイゼーションに必要な3つのポイント

## 1 データ抽出

- レガシーデータ抽出
- 非構造化データ対応

## 2 データ変換

- 文字コード変換
- 異なるI/F
- 異なるフォーマット

## 3 データ品質

- ファイルの整合性
- データの整合性

## ②データ変換（文字コード変換）

### 01 主に扱っているファイルや文字コードなどを適切に変換し活用

#### レガシーシステム



##### ■ファイル形式

固定長

可変長

VSAM

##### ■文字コード

シングルバイトコード： EBCDIC

マルチバイトコード： IBM / JEF / NEC /  
KEIS / 外字

##### ■文字属性

文字列（シフトコードあり、なし）

数値（パック10進数/サイン付ゾーン10進数/  
サイン無ゾーン10進数）

変換

#### モダンシステム



##### ■ファイル形式

CSV

Excel

JSON

XML

##### ■文字コード

シングルバイトコード： ASCII

マルチバイトコード： Shift-JIS / UTF-8  
/ UTF-16

##### ■文字属性

文字列 数値 日付/時間

##### ■変換

和暦⇔西暦 全角⇔半角統一 四則演算 etc.

## ②データ変換（異なるインターフェイス）

02

### 異なるI/Fを同時に扱い、あらゆるデータを活用

#### ファイル連携

- 接続先のシステムからAPIが提供されていない
- 外部からアクセスができない閉鎖環境

#### プライベート環境



基幹システム



- ✓ 専用線の手間とコスト
- ✓ インターネット通信へのセキュリティ課題

#### API連携

- 「REST API」「JDBC」「FTP」「IMAP/POP3/SMTP」など

REST APIが提供されているサービス



REST API

REST  
コネクター

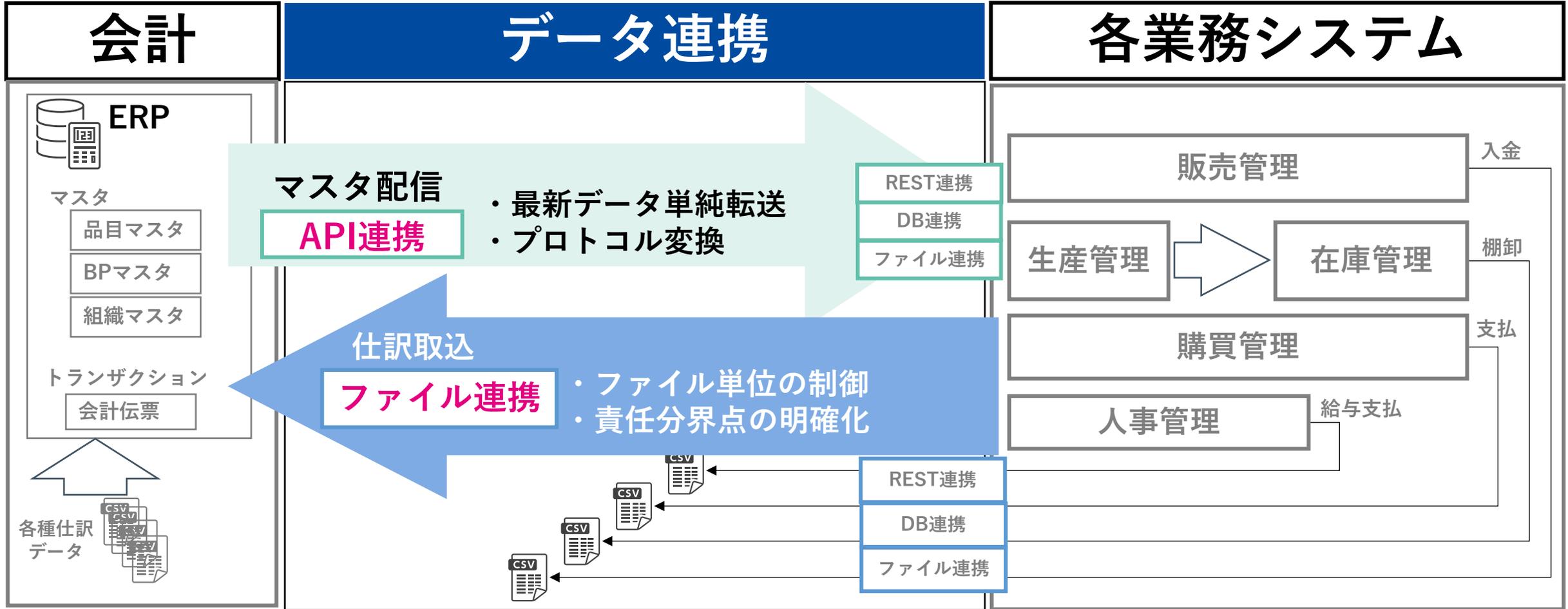
JDBCが提供されているサービス



JDBC

JDBC  
コネクター

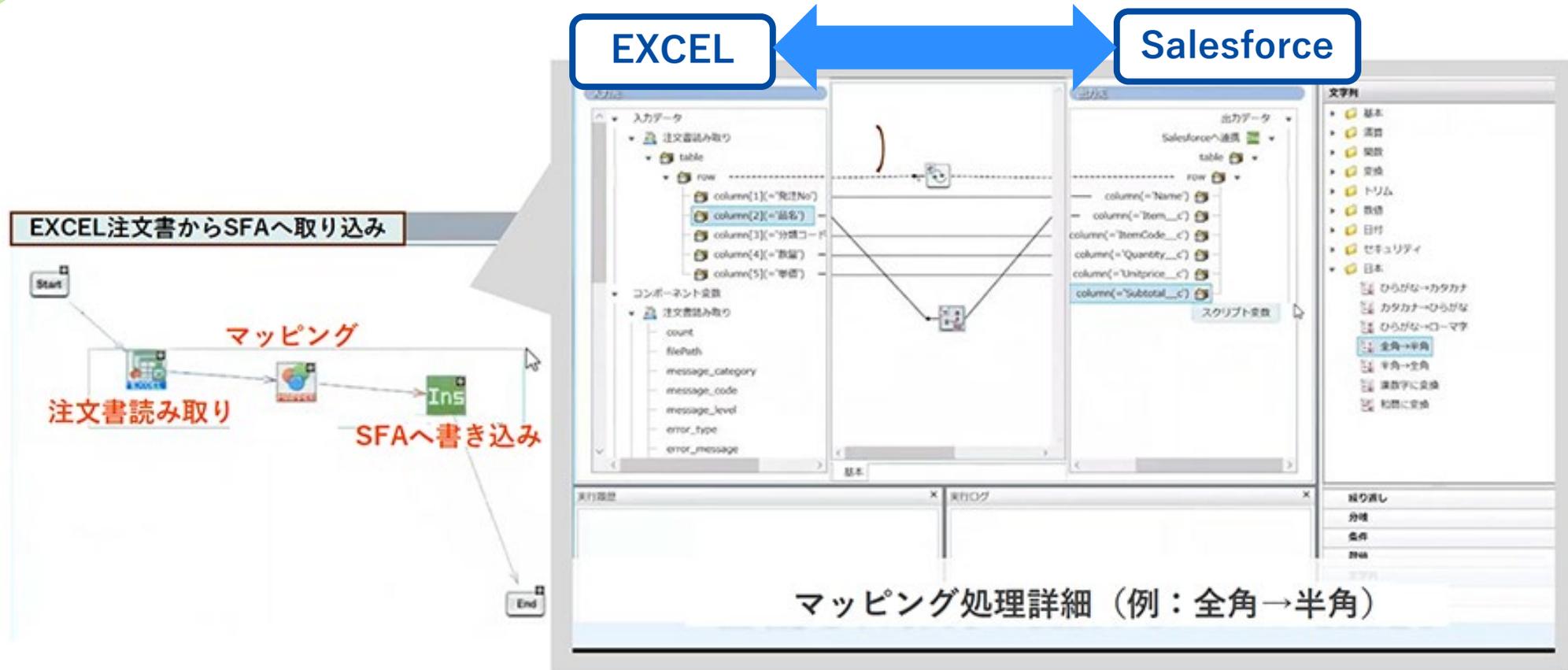
## ②データ変換（異なるインターフェイス）



## 2 データ変換（異なるフォーマット）

03

異なるファイルフォーマットも直感的に扱い活用につなげる



# データモダナイゼーションに必要な3つのポイント

## 1 データ抽出

- レガシーデータ抽出
- 非構造化データ対応

## 2 データ変換

- 文字コード変換
- 異なるI/F
- 異なるフォーマット

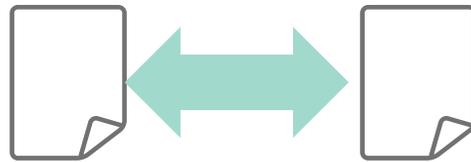
## 3 データ品質

- ファイルの整合性
- データの整合性

### ③データ品質（ファイルの整合性）

## 01 レガシーデータの欠落・改ざんがなくモダンシステムに確実に連携

#### ①連携ファイルの完全性

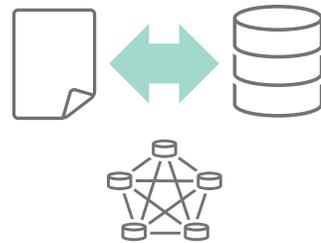


送信元 → 送信先

1. データ件数
2. サイズ
3. ハッシュ値

データ連携間での不整合を検知

#### ②データ要件の整合性



- 文字コード
- 日付/時間
- 表記ゆれ etc..



モダンシステム要件

XX	XX	XX
AAA	1000	0
BBB	2000	1

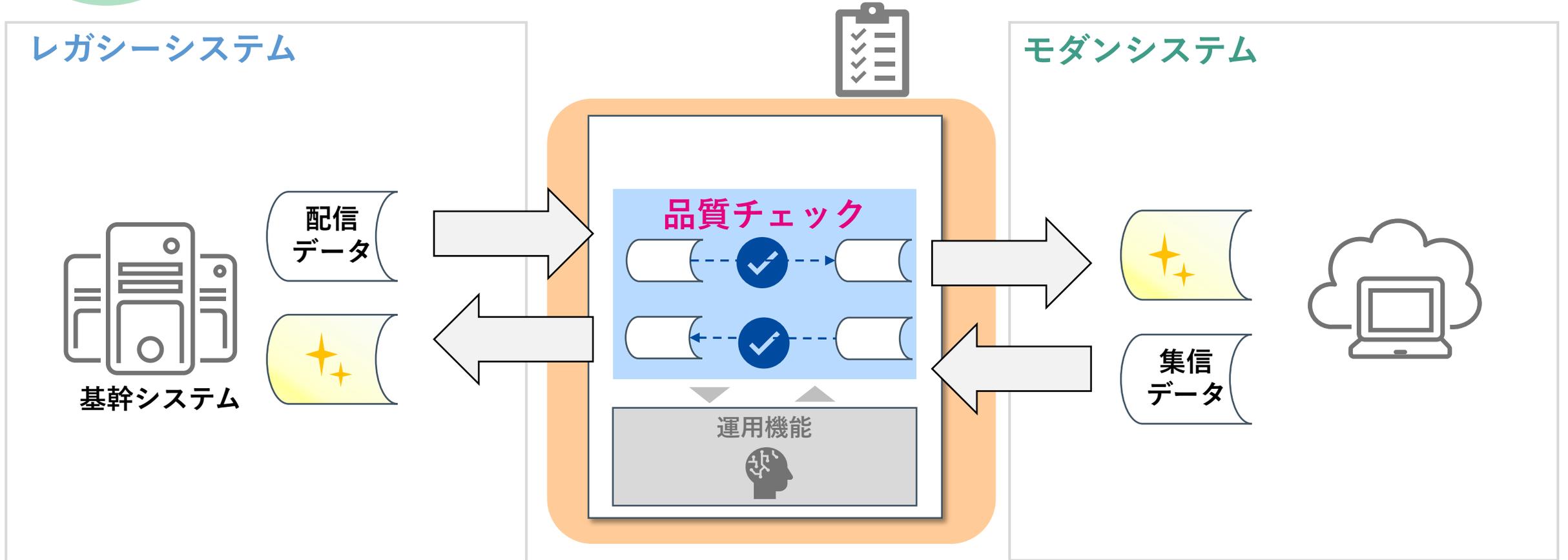
1. ファイル形式変換
2. データ形式変換

データ値の整合性を検知し、適切な値に変換

### ③データ品質（データの整合性）

02

レガシーシステム側と同等の品質チェックをモダン連携前に



## プロダクト・サービスでできること

# データモダナイゼーションに必要な3つのポイント

## 1 データ抽出

- レガシーデータ抽出
- 非構造化データ対応

**HULFT10**   
**HULFT SQUARE**

## 2 データ変換

- 文字コード変換
- 異なるI/F
- 異なるフォーマット

**DataSpider Servista**   
**HULFT SQUARE**  
**DataMagic** 

## 3 データ品質

- ファイルの整合性
- データの整合性

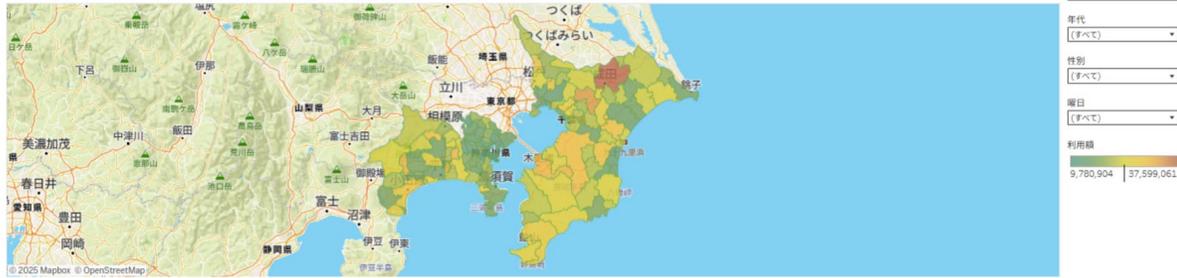
**HULFT10**   
**HULFT SQUARE**

# データモダナイゼーションを実現している世界

## (活用例) クレジットカード利用情報を分析

クレジットカード利用状況ダッシュボード

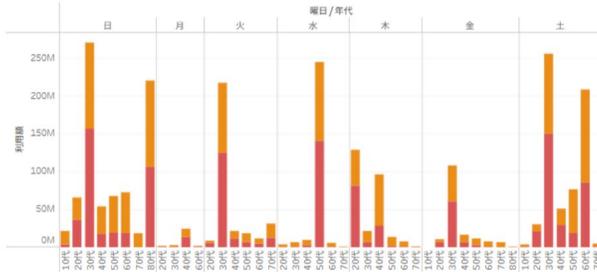
利用額マップ



利用額ヒートマップ



曜日・年代・性別\_利用額



## (自社事例) 社内イベント参加者数をグラフ表示

設定一覧:

New Chat +

search

本部ごとの社内イベン... X

InfoSquareでいいね数... X

InfoSquareで活動が最... X

最もいいね数が多い上... X

当社とまだ取引が無い... X

社員の人数を教えてく... X

社内イベントの参加者... X

当社とまだ取引が無い... X

HULFT認定資格の保有... X

質問をどうぞ

本部ごとの社内イベント参加者数をグラフで表示してください。

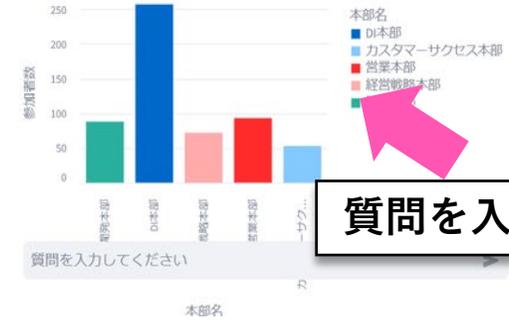
SQLを自分で書く

以下の本部が含まれています：開発本部、DI本部、経営戦略本部、営業本部、カスタマーサクセス本部。参加者数が最も多いのはDI本部で、次に営業本部、開発本部、経営戦略本部、そしてカスタマーサクセス本部の順です。

SQLを入力

GPTの回答作成過程を表示

実行



質問を入力してください

BI×データ分析基盤



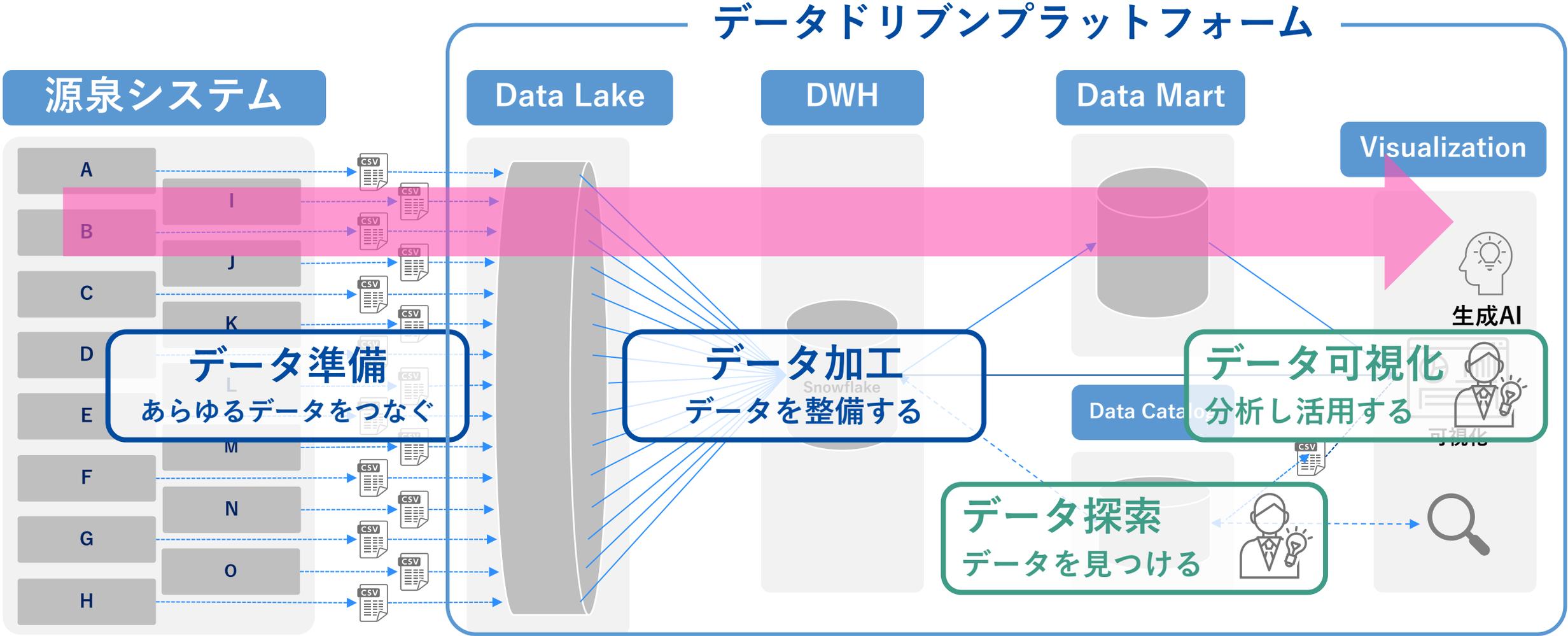
集約されたデータをツールで可視化

生成AI×データ分析基盤

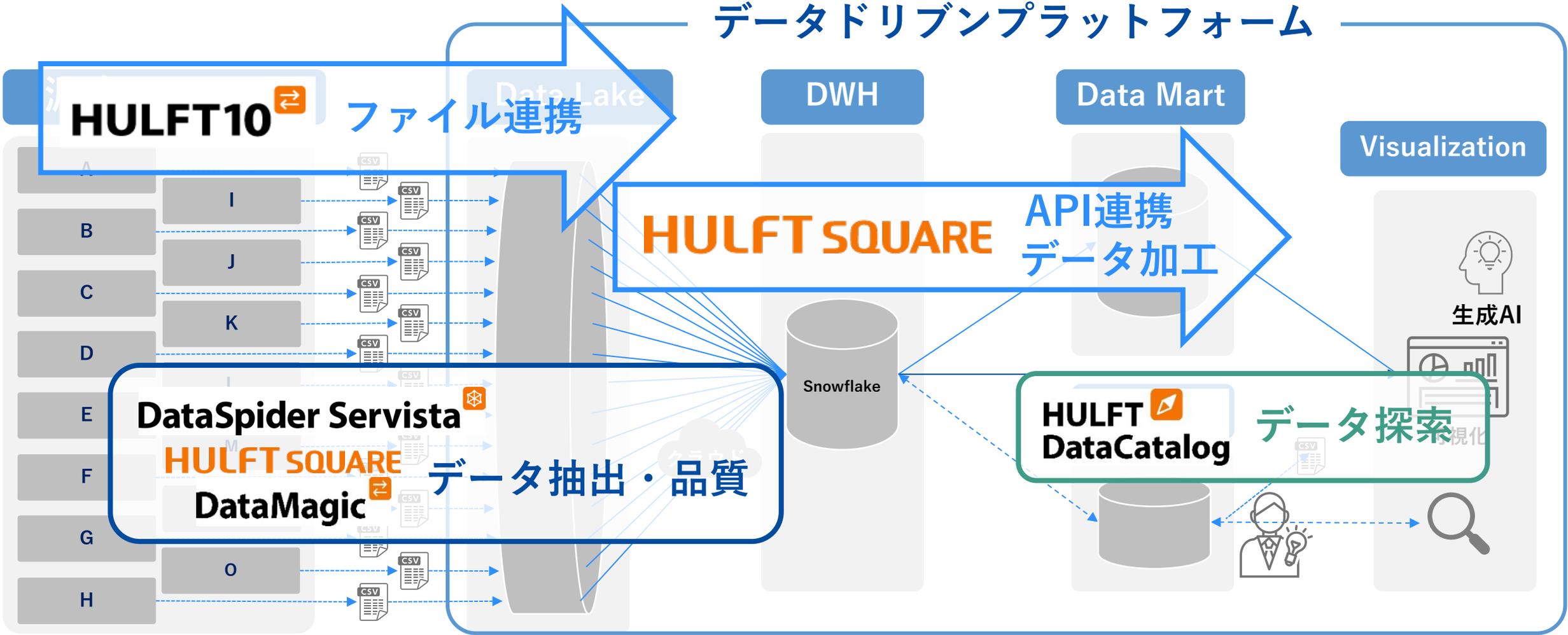


自然言語でデータを探索し同時に可視化

# データドリブンプラットフォームの構成（当社事例）

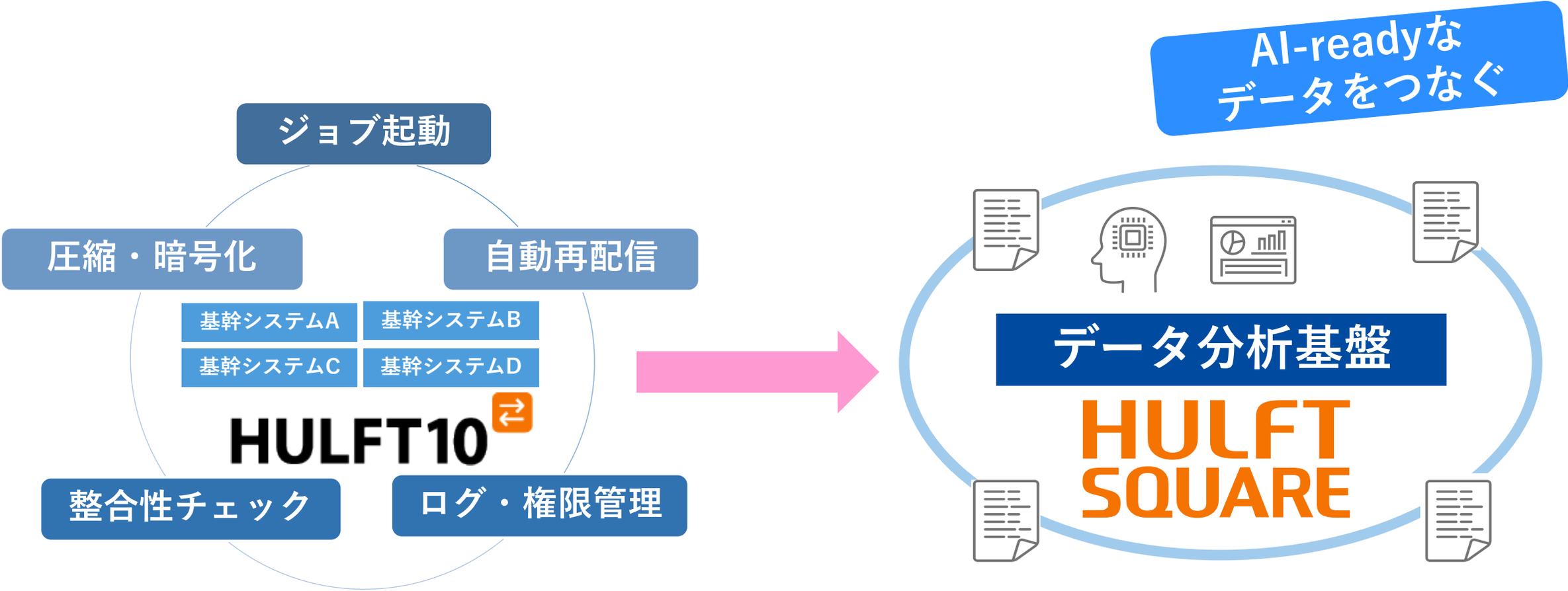


# データドリブンプラットフォームの構成（プロダクトマップ）

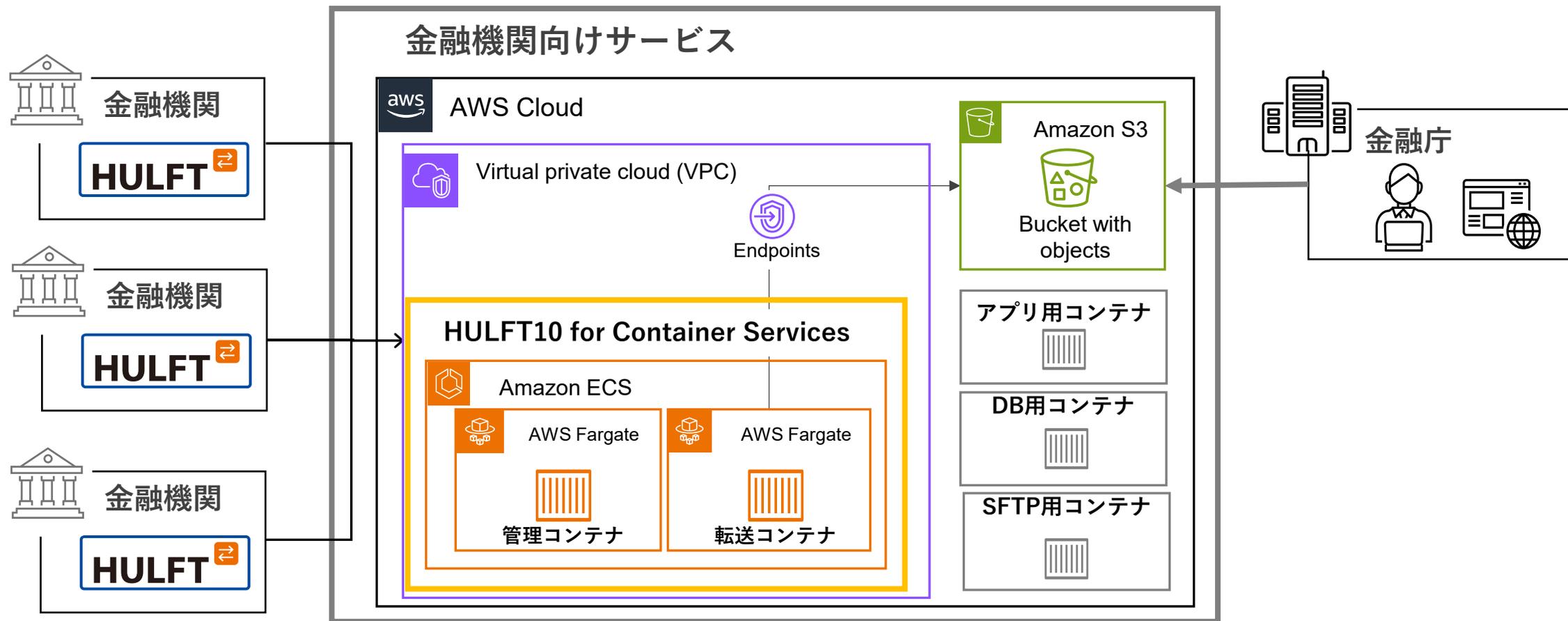


# 確実なデータ連携で共通基盤につなげる

「ファイル連携」に必要な「構築・運用・移行」の重要機能をHULFTは標準搭載。あらゆるデータを統合管理するHULFT Squareで業務運用を支援します。



- 金融機関を対象にした国税庁認定クラウドサービスを提供。
  - ・ 「開発生産性」「拡張性」「災害対策」の観点で、**サーバーレス（Fargate）方式**を選択
  - ・ 「**確実なファイル転送**」「**圧縮**」「**暗号化**」を実現する機能を搭載した**HULFT**を選択



# レガシーシステムのデータ活用

レガシーシステムの価値は変えずにモダン技術とつないでいき、  
長年積み重ねてきたデータを価値に変換して活用し進化していく。



現場で守り続けてきた**信頼性 / 完全性**

現場で蓄積しより良く最適化されてきた**業務ロジックと業務運用**



< 免責条項 >

本資料の内容は、資料作成時点の当社の判断に基づいて作成されているものであり、今後予告なしに変更されることがあります。よって本資料使用の結果生じたいかなる損害についても、当社は一切責任を負いません。また、本資料の無断での複製、転送等を行わないようお願いいたします。なお、本資料に記載されている会社名、製品名は各社の商標または登録商標です。