

Fujitsu Software

ビッグ データ インテグレーション サーバー

Big Data Integration Server

ご紹介

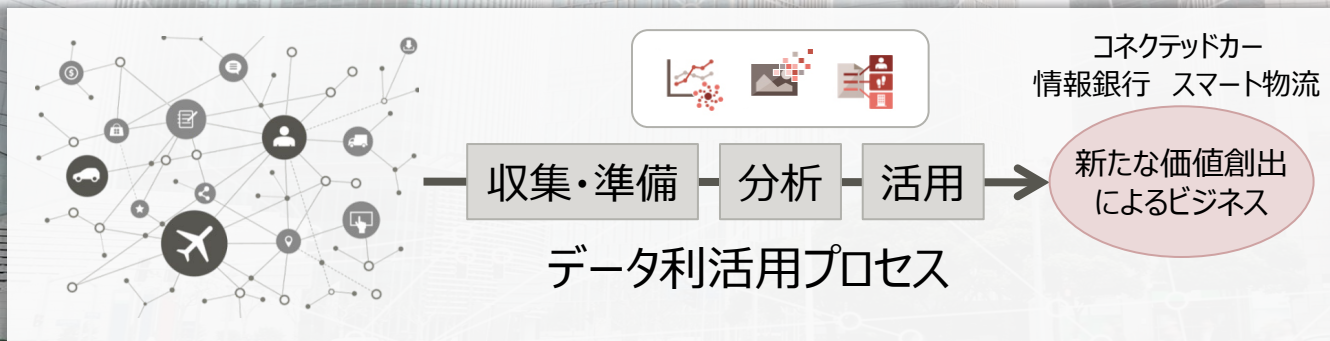
2024年4月

富士通株式会社

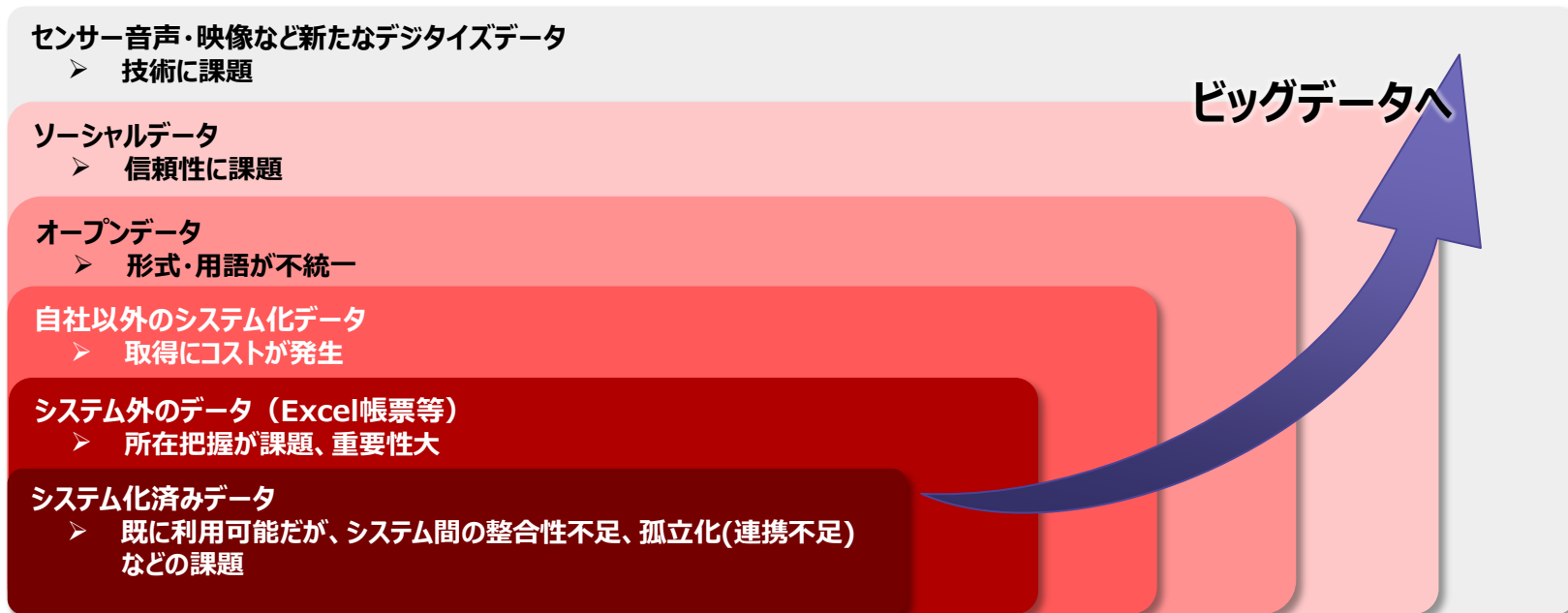


データ利活用ビジネスの到来

- 人やあらゆるモノがネットワークで繋がり、膨大なデータの収集・蓄積および分析により、新たな価値を創出
- 個人情報・現場のデータなどを有効かつ安全に活用した業務改善・ビジネス化が求められている

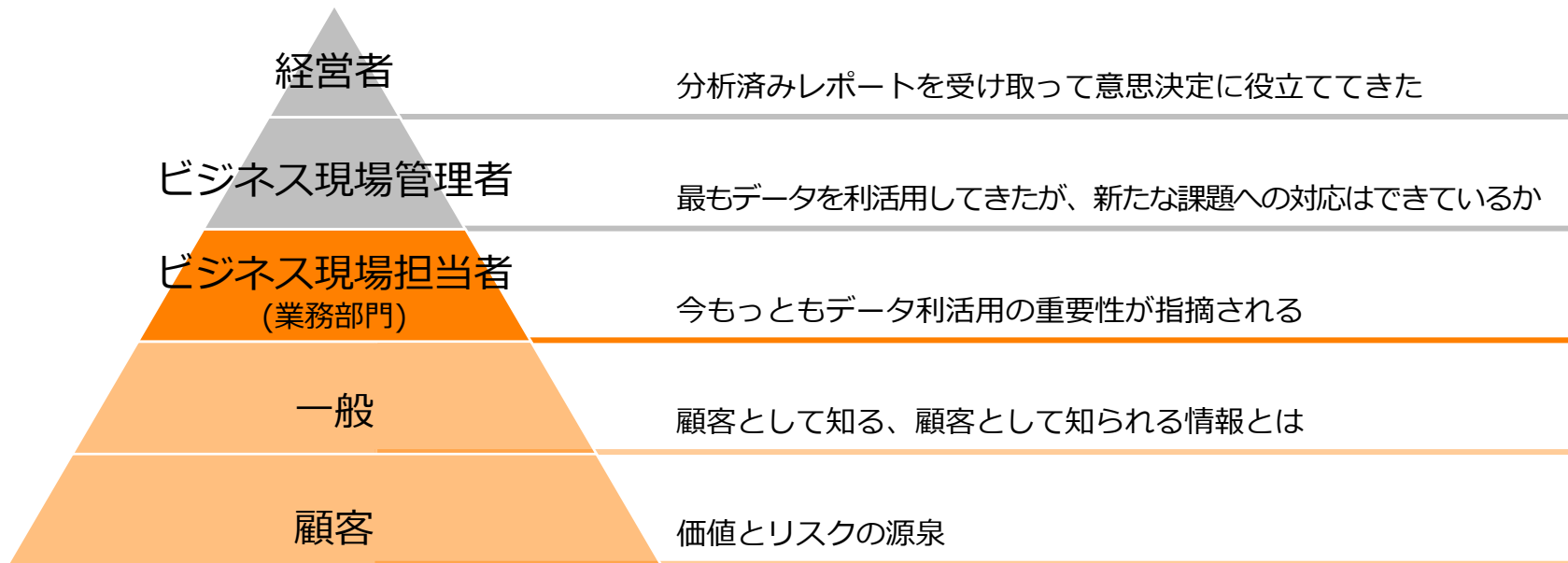


- 活用したいデータが多種・多様化しているだけでなく、**膨大なデータからデータを取り出し活用できる形に加工する**必要が高まっている



● データの利用者の多様化

- 従来の業務データを活用した経営層のKPI分析から、業務部門でデータ活用による「予測・分析」をベースとした業務の効率化・価値創出が必要



データの民主化：誰もがデータにアクセスしやすく、必要な時にデータ活用できるようにすること

データ利活用シーンの変化

- 業務部門が自らデータ分析し、意思決定・アクションにつなげる

従来のBI

- ✓ 記述型分析(Query&Report)
- ✓ バッチ型
- ✓ システム開発

ビジネス
インテリジェンス



基幹系システム(既存)

構造化データ

新たなデータ利活用

- ✓ 未来予測型
- ✓ リアルタイム
- ✓ アジャイル

① ビジネスプロセスとの融合

② セルフサービス型のデータ利活用

③ 多種多様・大量データの統合管理



基幹系システム(既存)

構造化データ

明細データ



ログ/SNS

非構造化データ

リアルタイム



センサー/デバイス

ストリームデータ

データ利活用の進め方のポイント

- 部門導入で実践と効果を創出しながらデータ利活用のノウハウを蓄積、データ利活用の対象範囲を拡大して本格的なデータ連携基盤の整備を進める

● 社外情報含めたデータ利活用を推進

ノウハウを蓄積しながらデータの収集領域を社外まで拡張

● データ利活用領域を社内全体に拡張

スモールスタートで得られた成果を基に社内全体まで段階的に拡張

● 社内の一部でデータ利活用を実践

データ分析やモニタリングなどのデータ利活用を実現し、成果を創出

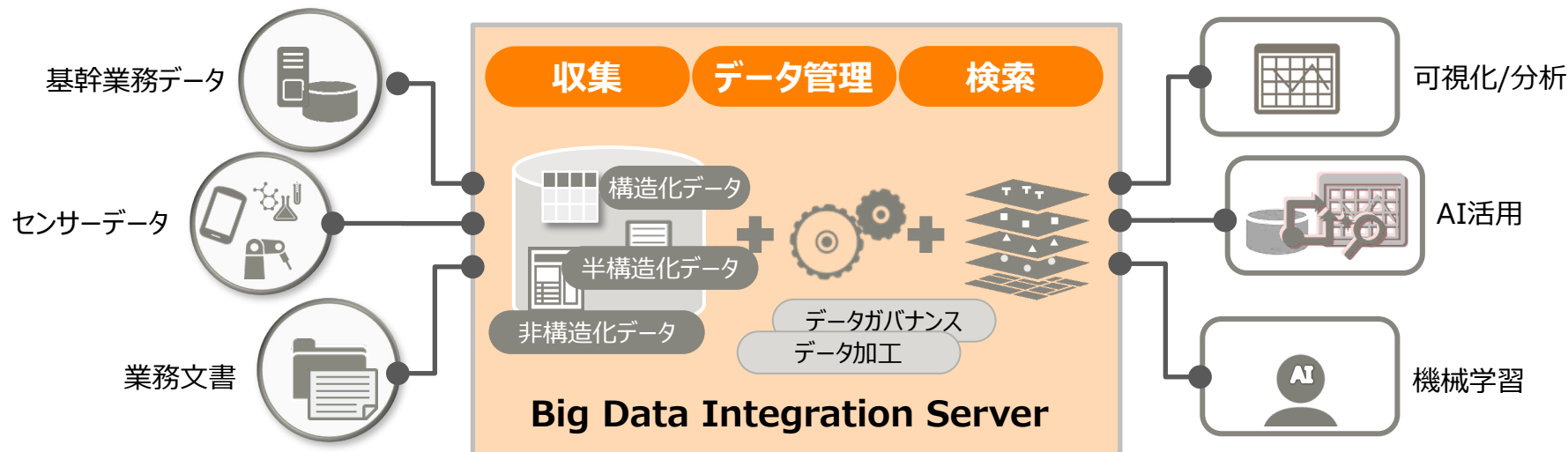
分析からスモールスタートし、多種多様なデータを横断的に活用できることが重要

Big Data Integration Serverとは

Big Data Integration Serverとは

多種・大量データのタイムリーな活用を実現するデータ統合基盤

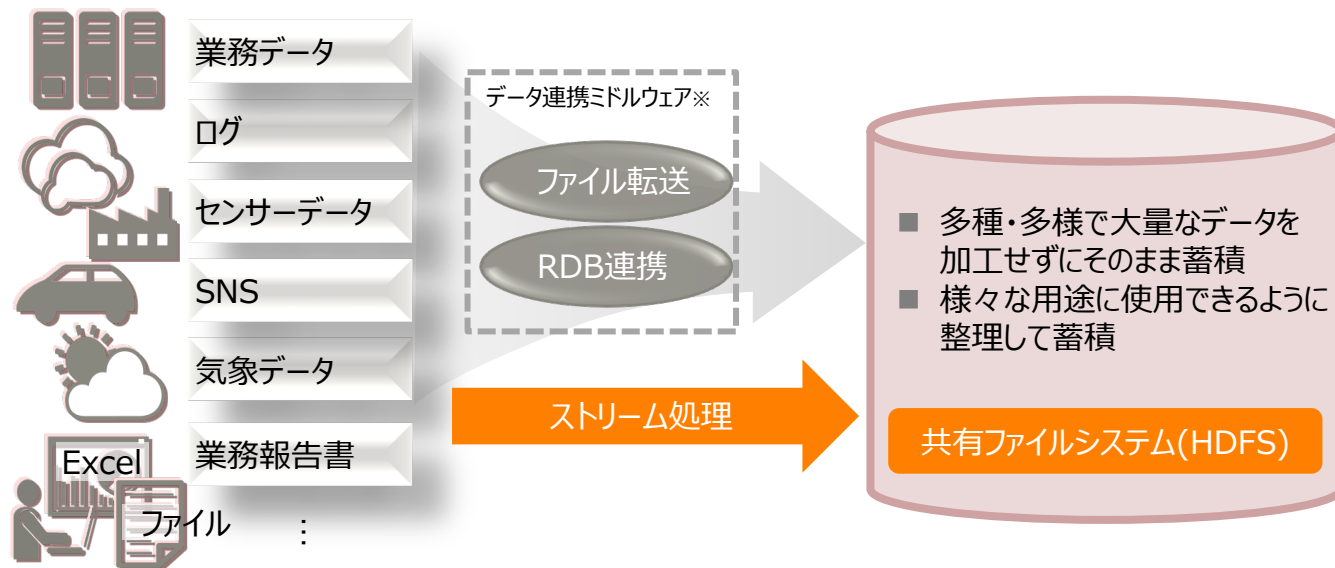
- **収集** : データ構造に関わらず、複雑な設計なしに収集
- **データ管理** : データ資産の管理を統制することで、データを安全に活用
- **検索** : インデックス不要で、どの項目でも安定した速度で検索



さまざまなデータを容易に収集

データ構造に関わらず、複雑な設計なしに収集

- 定型バッチの業務データに加えて、ストリームデータもリアルタイムに収集
- ログやSNSなどデータ構造に関わらずに蓄積

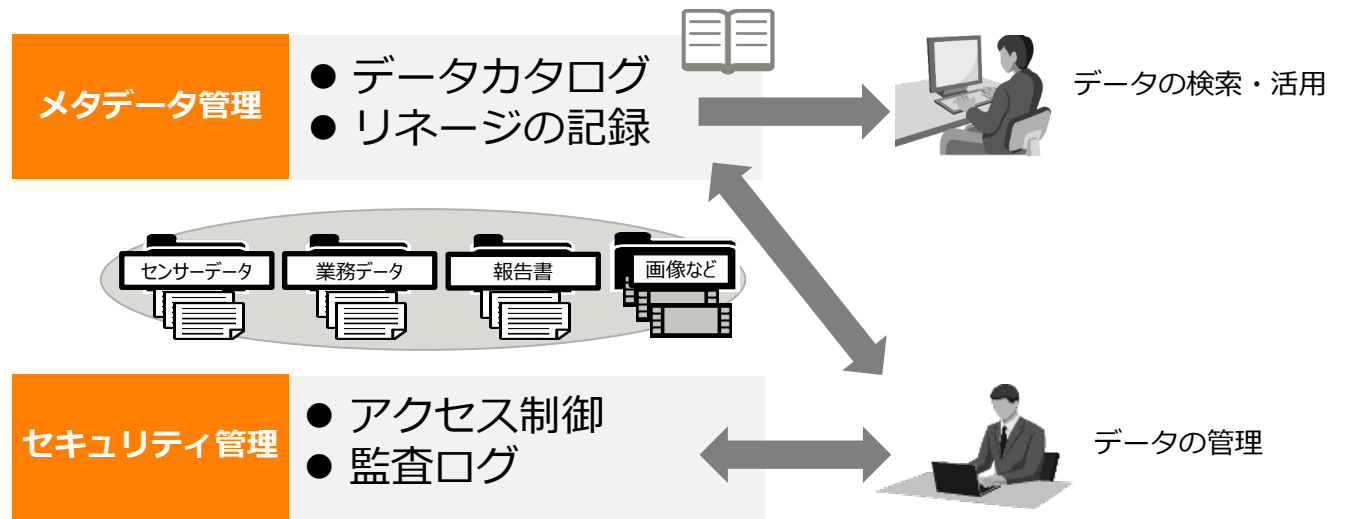


※データ連携ミドルウェア：Interstage Information Integratorを利用

データガバナンス

データ資産の管理を統制することで、高品質なデータを安全に活用

- データの来歴を自動的に記録し、情報源が最新かつ信頼できるか確認
- アクセスポリシーによるアクセス制御と監査ログ出力により、データを安全に活用



データ構造を意識せずにデータ活用の目的に合わせて高速に加工

- 抽出/連結/集計/ソートを自在に組み合わせ、定義ベースで簡単に指定
- 集計/ソートに特化した独自技術を搭載し、大量データの加工時間を短縮
- Apache Hadoop/Sparkによる並列分散処理でスループット性能を向上

抽出、連結、集計、ソートの
4機能を必要に合わせて自在に組み合わせ



- ▶ 抽出 : 条件抽出、条件仕分け
- ▶ 連結 : マッチング、マージ、コンバート
- ▶ 集計 : 集計、グルーピング
- ▶ ソート : 並び替え、キー仕分け

抽出 連結 集計 ソート

並列分散処理



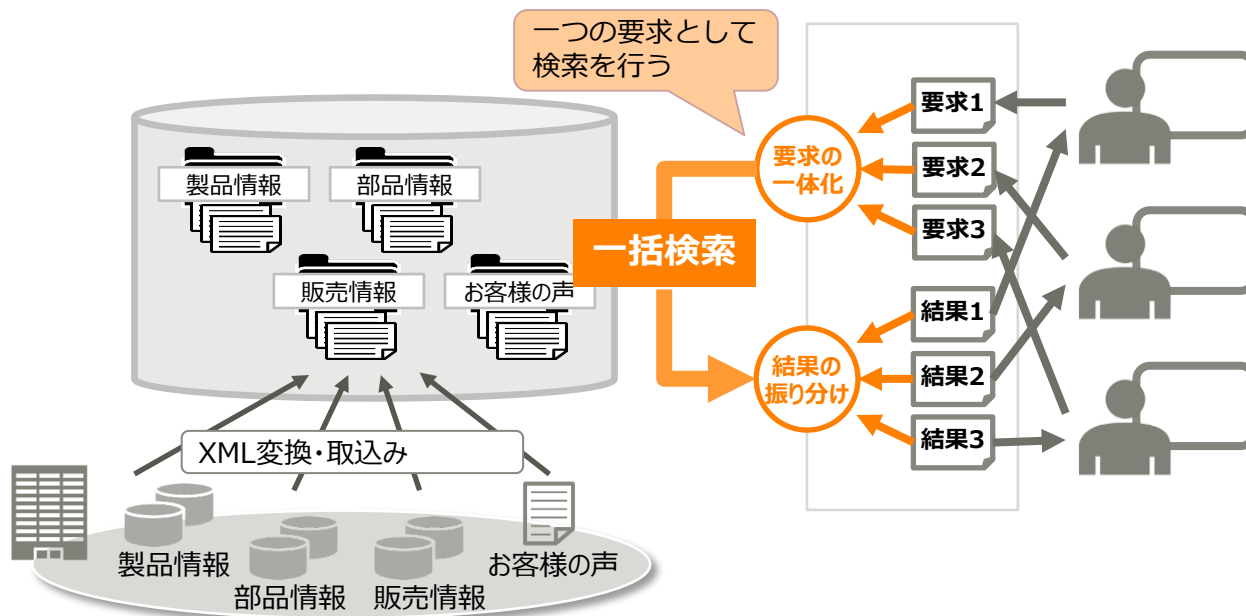
複雑な加工条件でも
入力データをワンリードで処理

- 高速パターンマッチング技術
- ハイトラフィック技術

複雑な条件でも安定した検索

インデックス不要で、どの項目でも安定した速度で検索

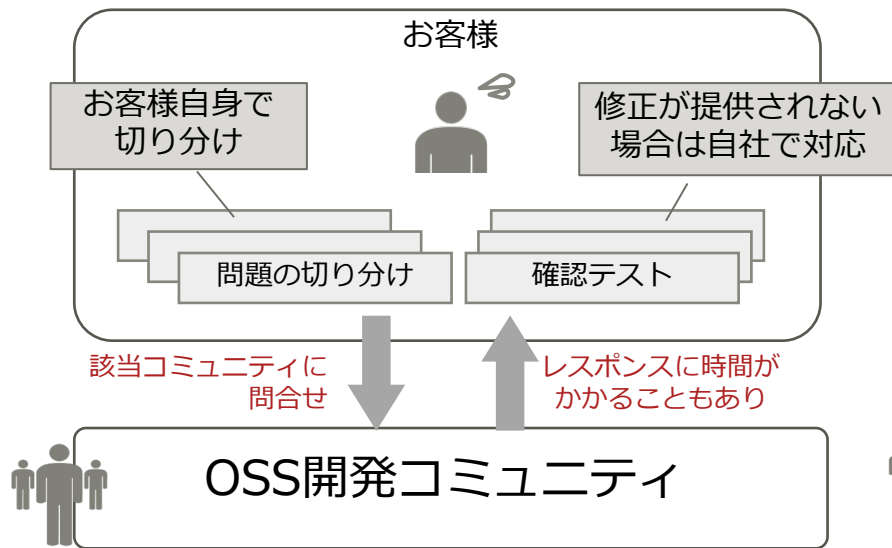
- 事前スキーマ定義が不要。インデックスレスで全項目・全文を自在に検索
- 複数の検索要求を一体化することで、アクセス集中時のレスポンスを平準化



お客様の“あんしん”運用を支える富士通のワンストップサポートをご提供

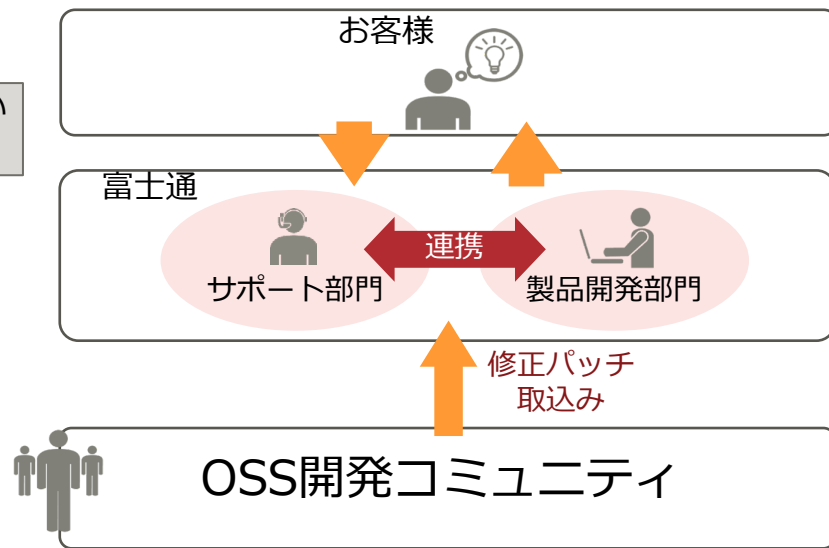
OSSコミュニティ版の場合

- 複数のコミュニティと連携しながら、自身で切り分けやテストを実施、またはサポートベンダーと契約



富士通ミドルウェアの場合

- OSSのトラブルも含めて富士通がサポート



多様なプラットフォームに柔軟に対応

マルチベンダークラウドに対応、どのクラウド環境でも同一のシステム構成を実現

- マルチベンダークラウドでの動作をサポートします
 - 対象クラウドサービス : Microsoft Azure、Amazon Web Serviceなど
- IaaS環境に構築することで、どのクラウド環境でも共通のインターフェースをご提供します
- オンプレミスからのクラウドリフト時も同一のシステム構成で構築が可能です



FUJITSU Hybrid IT Service
FJcloud-O



AWS



Microsoft Azure

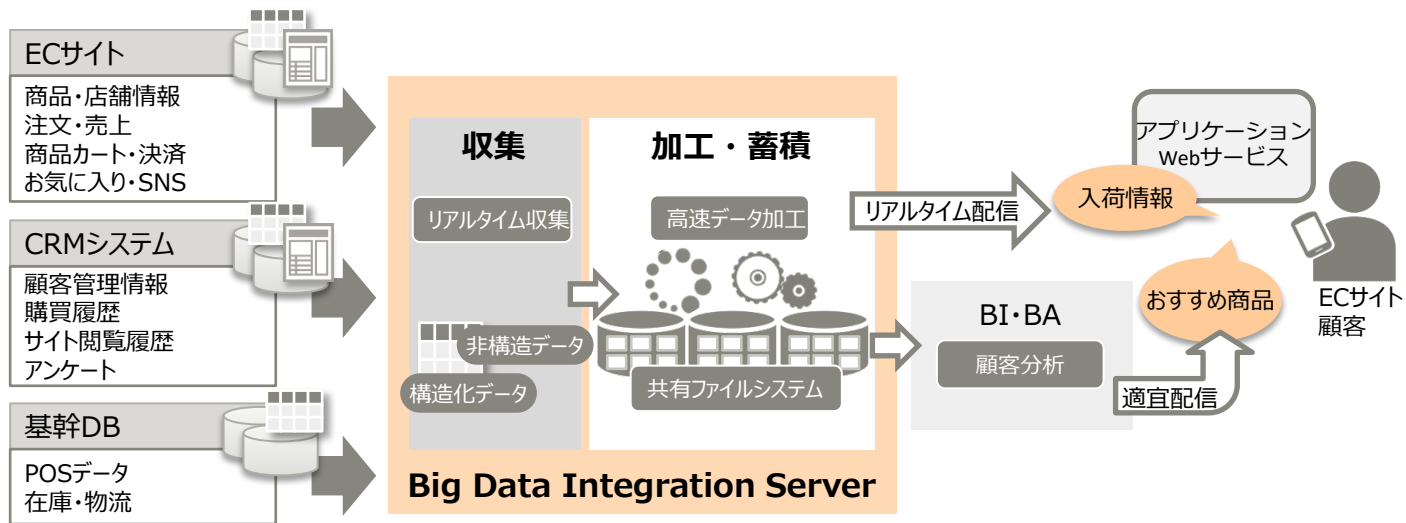


オンプレミス

活用シナリオ

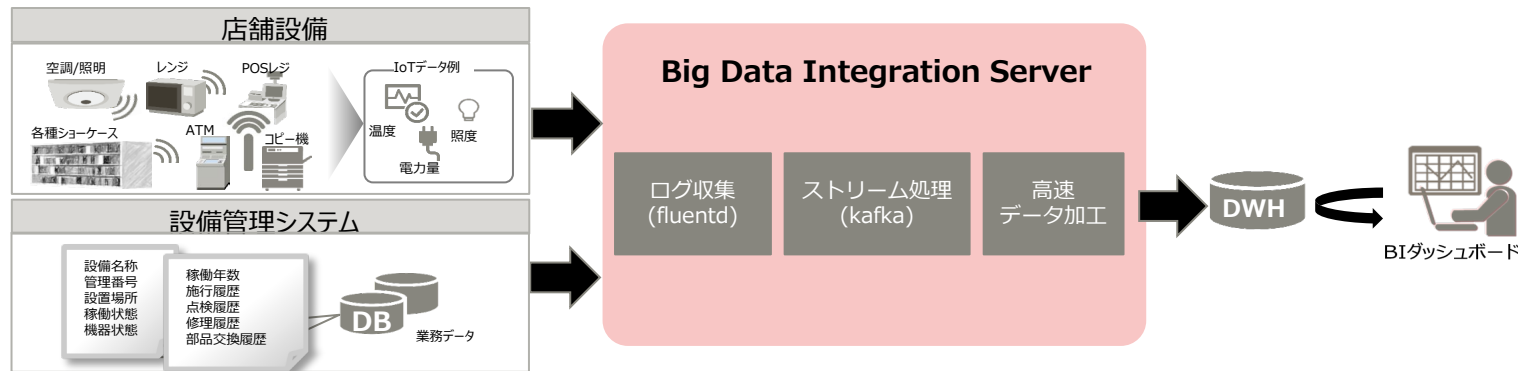
お客様ごとに関心が高い商品情報をリアルタイムに配信

- 商品・在庫・顧客情報(構造化データ)とアンケート結果(非構造化データ)を収集/一元管理
- お気に入り登録している商品が入荷されたら在庫情報をリアルタイム配信
- ECサイト・CRMシステム・基幹DBで管理しているデータを横断的に活用した顧客分析



IoTデータを用いて設備の稼働状況を監視、
本部による省エネ対策や設備保守の効率化を実現

店舗設備から発生するセンサーデータと基幹システムで管理される業務データを統合し、リアルタイムにデータを収集、加工/蓄積、可視化



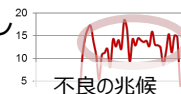
最適な省エネ対策

- ✓ 電力、温度/湿度をリアルタイム監視
- ✓ 温対法の報告業務の軽減、省エネ計画策定



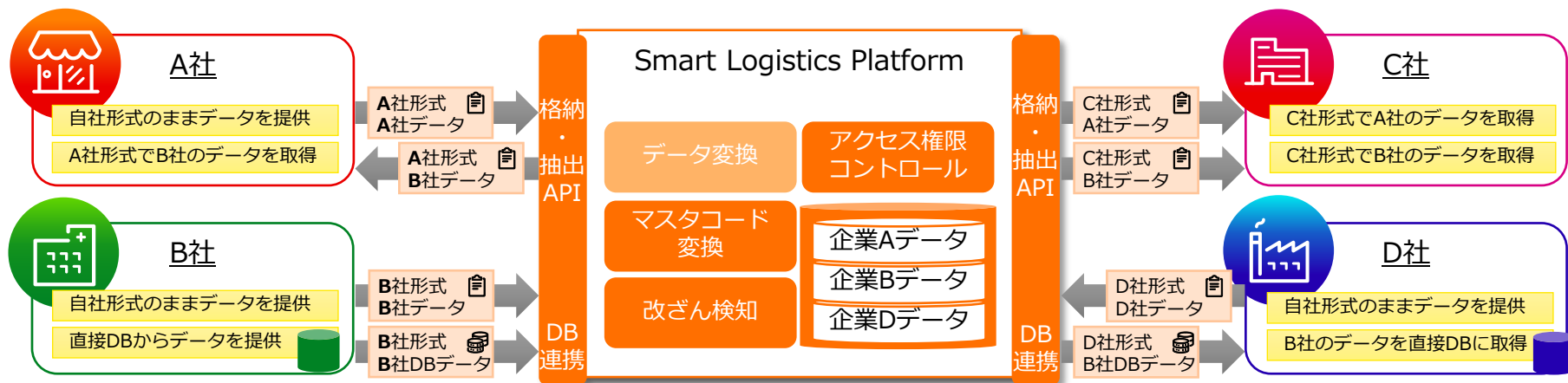
設備の故障予測

- ✓ 修理/交換が必要な設備の洗い出し
- ✓ 故障によるダウンタイムの回避



SIPスマート物流サービス基盤（Smart Logistics Platform）の コア機能群を提供

- データ変換：他の企業のデータを自社のデータ形式で取得できるため、社内システムでそのまま活用
- データ共有：必要なタイミングでほかの企業のデータを取得できる。暗号化や改ざん検知機能で安全に共有
- DB連携：データベースから差分データを定期的にデータ基盤へ提供／データ基盤からデータベースへ取り込み



■ : Big Data Integration Server提供機能

■ : ACMS Apex提供機能

企業の壁を越えたデータ交換が可能な基盤の実現

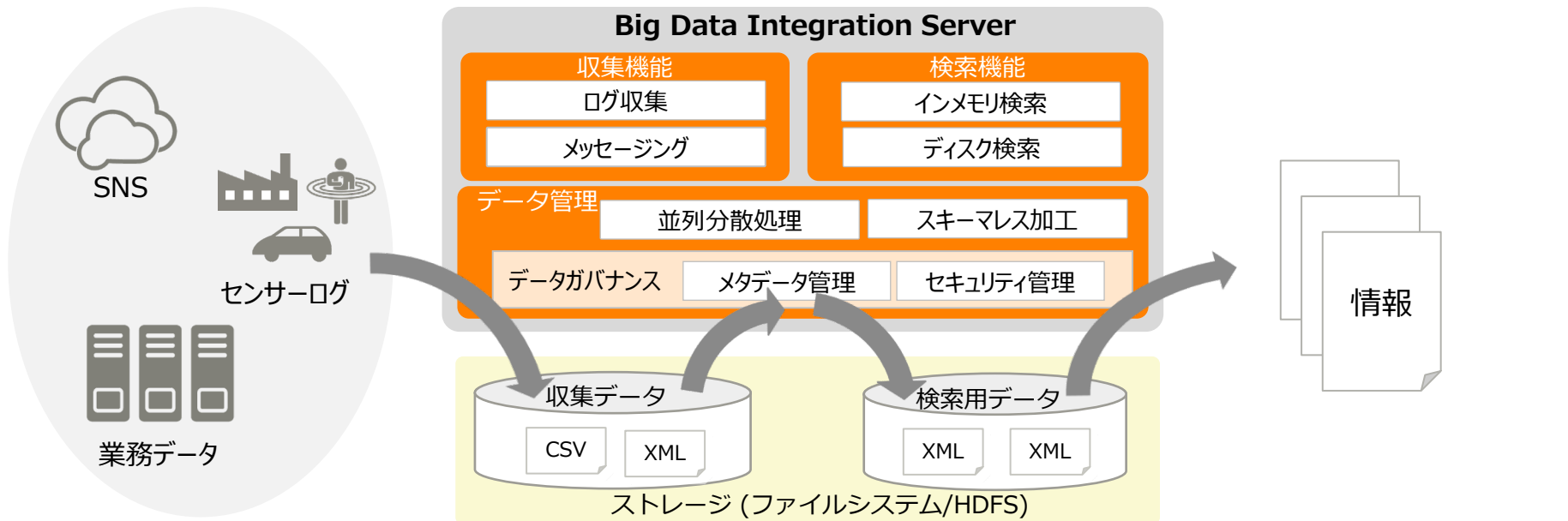
企業間でデータをやり取りする仕組みを短期で実現

格納抽出API	データ交換する各企業は、共通の格納API、抽出APIを使用してデータ基盤へアクセス
データ変換定義	データ変換では、定義ベースで共通形式と自社形式の変換を作成 作成したデータ変換定義は基盤内で一元管理され、共有・再利用することが可能
マスタコード 変換支援	コード変換定義作成支援機能により、各企業のマスタコードと共通のマスタコードの変換を支援 コード対応表の自動生成を支援、自動生成時に提示される候補からGUIで選択して定義
アクセス権限 コントロール	データ提供者がアクセス権限コントロールの定義を作成することができ、利用者の属性・データ に応じたアクセス権限を設定可能
改ざん検知	蓄積されているデータが改ざんされた場合、その改ざんを検知しエラー通知によって対処が可能
DB連携	事前設定のみで自社のDBから自動的にデータ基盤にデータを提供（前回からの差分データ） 同様にデータ基盤から差分のデータを自動的に自社のDBに取り込んで活用 業務DBの負荷に応じてスループットを自動で最適化し業務アプリケーションの性能影響を抑止

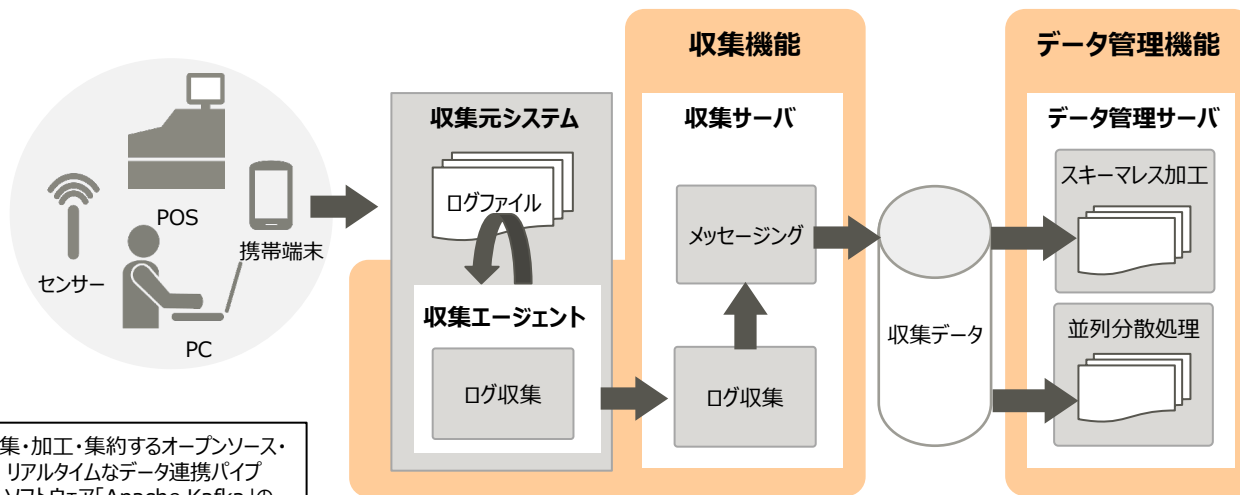
機能のご紹介

機能概要

- **収集機能** : 収集した各種データ(CSV/XML)を、XML形式で統合・管理
- **データ管理機能** : メタデータやデータリネージの利用およびセキュリティポリシーの設定によりデータ資産の管理を統制
- **検索機能** : 独自の検索技術により、マートを作成しなくてもデータを高速抽出



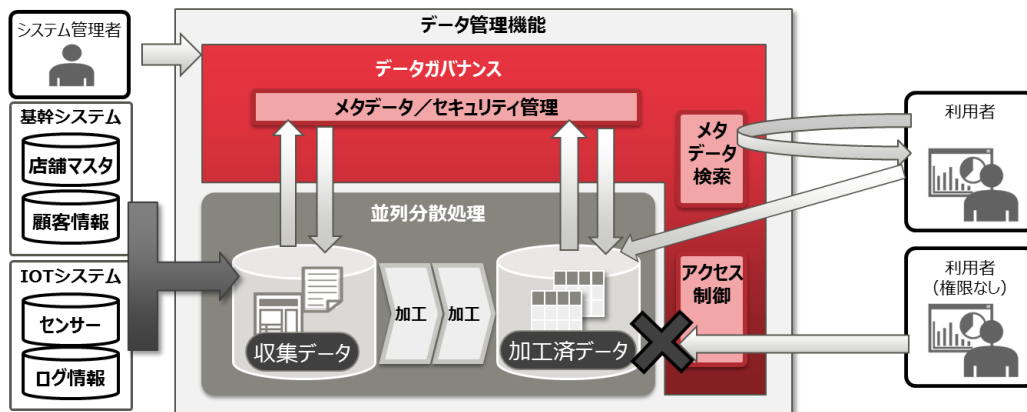
- 様々なデバイス、アプリケーション、ミドルウェアから出力されるログを収集
 - ログ収集：収集したデータのバッファリングや送信に失敗したデータの再送により、ログファイルなどストリームデータ*の収集をより簡単・安全に実現
 - メッセージング：データを一時的に保持し、データ入出力を非同期化することで、全体のレスポンス・スループット性能を最適化



本機能では、散在するログを収集・加工・集約するオープンソース・ソフトウェア「Fluentd」、および、リアルタイムなデータ連携パイプラインを実現するオープンソース・ソフトウェア「Apache Kafka」の機能を提供しています。

*：継続的に発生、かつタイムスタンプを含むデータ

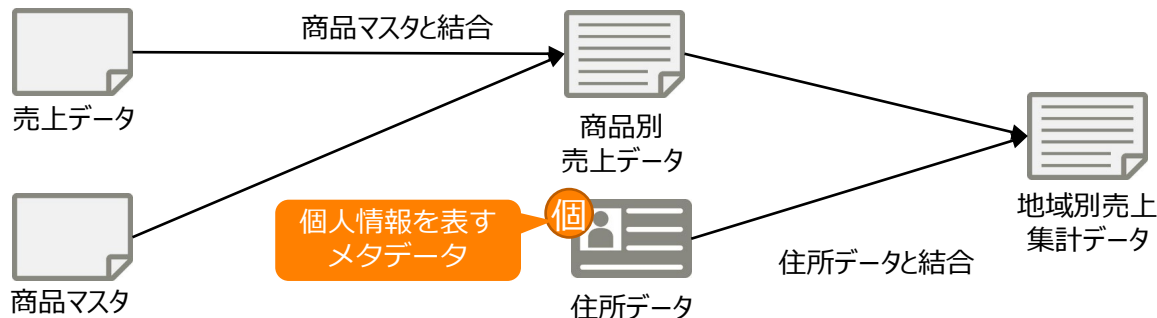
- データの来歴管理によるアクセシビリティ向上、およびセキュリティポリシーの定義による強固なセキュリティを実現
 - メタデータ管理
 - ✓ 利用者が必要なデータをすぐに見つけやすいよう、データ管理者がスキーマ情報や業務上の意味などの情報(メタデータ)を収集したデータに付与
 - ✓ 収集データからどのような加工・変換されてデータが作成されたのか(データリネージ)を記録
 - セキュリティ管理
 - ✓ アクセスポリシーによるアクセス制御と監査ログの出力により、機密情報を含むデータでも安全に活用



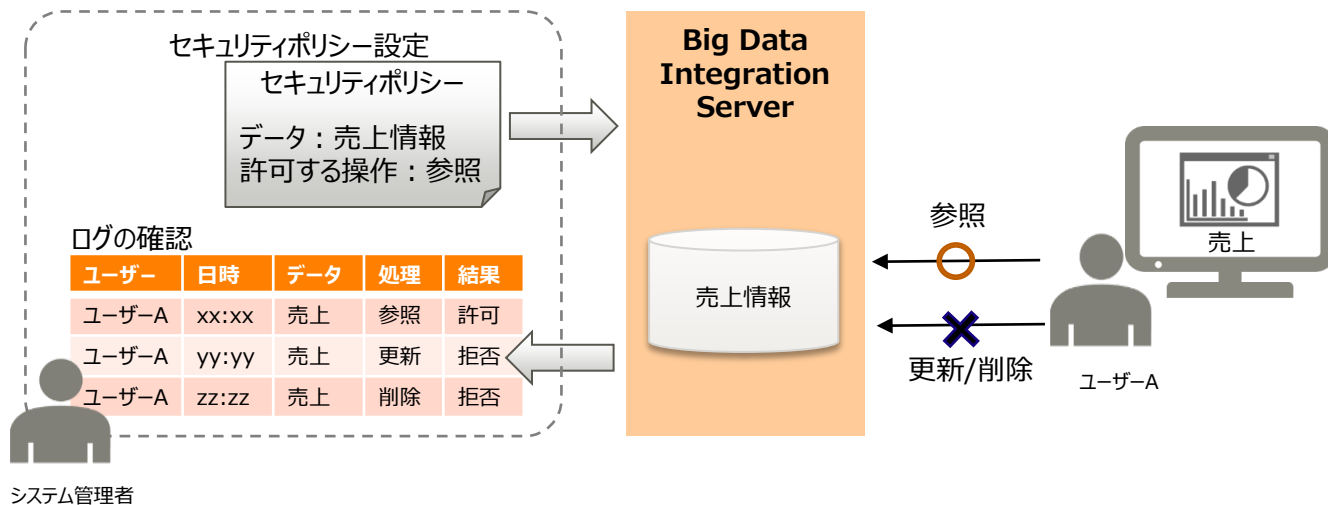
- Apache Atlasを利用しメタデータやデータリネージを蓄積
 - 蓄積されているデータに対し、データの種類やフリーワードで検索
 - メタデータの検索やデータリネージを確認することで、目的に合ったデータかどうかを判断
 - システム管理者/利用者が構造化/非構造化データに関わらずメタデータを付与

例：地域別売上集計データのデータリネージ

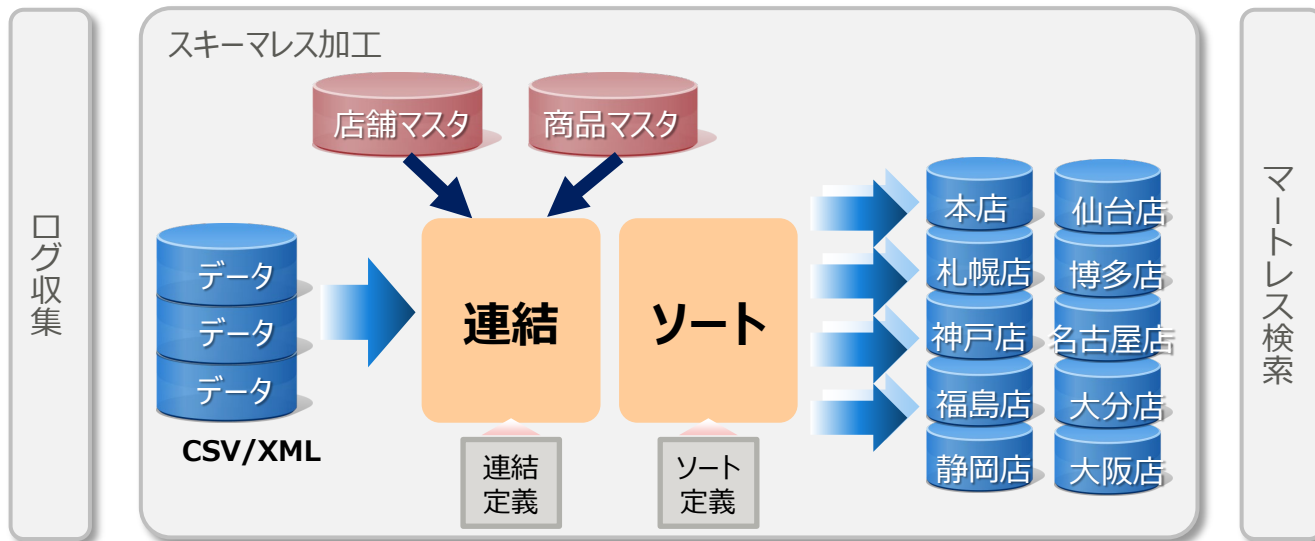
- 地域別売上集計データがどのようなデータから作成されているか分かる
- 結合する住所データに個人情報を表すメタデータが付与することで、地域別売上集計データに対する個人情報の有無が簡単に分かる



- Apache Rangerを利用しセキュリティポリシーの設定や監査ログを蓄積
 - データごとにアクセスできる人や操作の可否を設定し、公開範囲を限定
 - いつ、誰が、どのデータにアクセスしたかをログに蓄積し、不適切なアクセスを確認



- CSV/XML形式の構造の異なるデータでも、そのまま処理可能
 - マスタ連結時に事前の並び替え不要
 - データの構造(スキーマ)を意識せずに、入出力ファイルと該当キーの定義で一括ソート
- 複雑な加工条件でも入力データをワンリードで処理するため、安定した処理性能

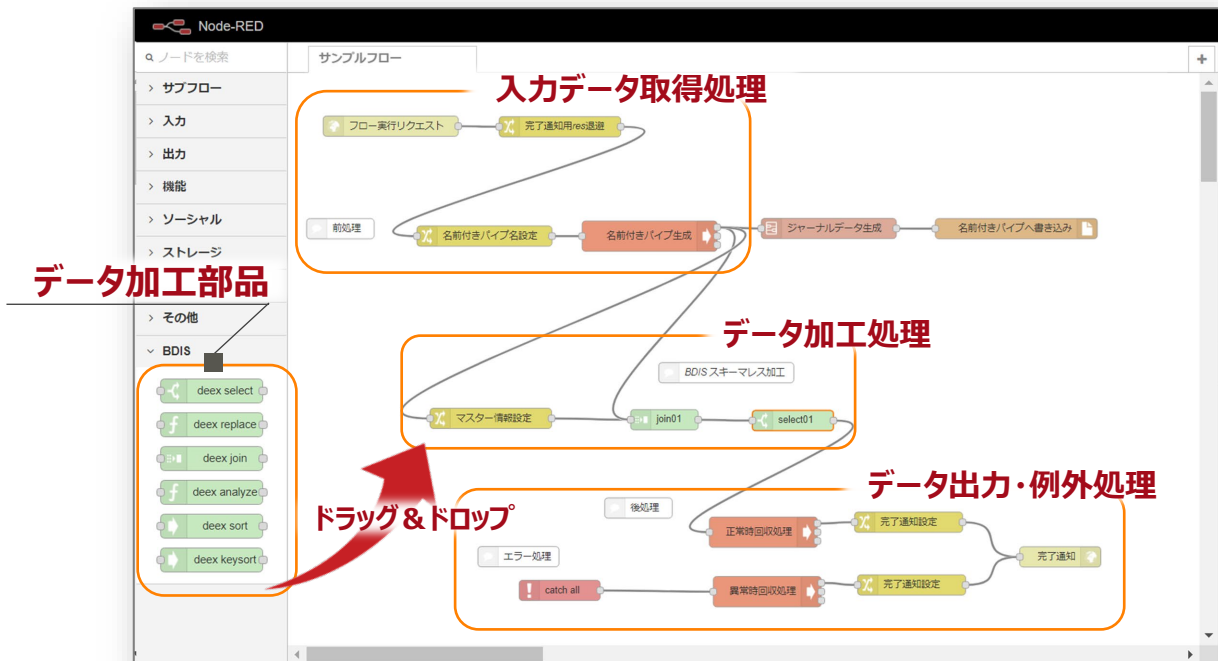


スキーマレス加工 (Node-RED連携)

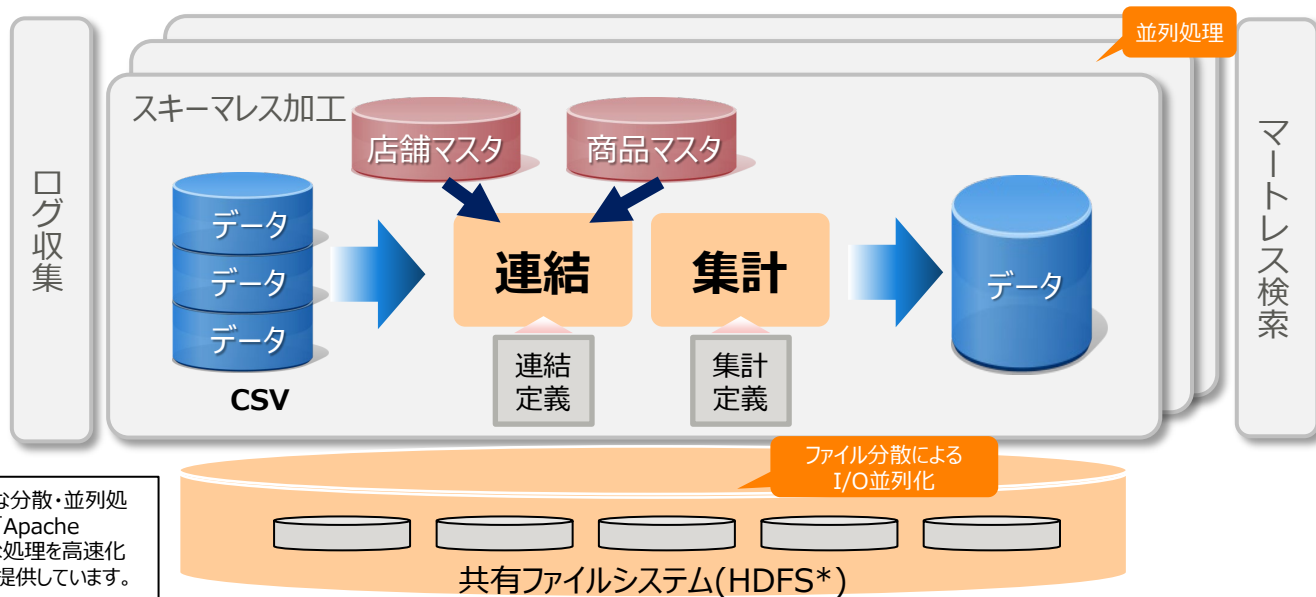
- オープンソース・ソフトウェアのNode-RED*1と連携

- 直感的なGUI上でスキーマレス加工の定義(抽出/連結/集計/ソート)を作成
- Big Data Integration Serverとしてデータ加工部品を提供

*1 : Node-REDは別途用意していただく必要があります



- データ加工をApache Hadoop/Spark上で動作させ、スループットを向上
 - CPUリソースを効率よく利用できる並列分散制御により処理時間を短縮
 - HDFSのファイル分散により、ファイルアクセス集中によるI/O競合を解消
 - SQLライクなクエリ言語で、並列分散処理を実行(Apache Hive)

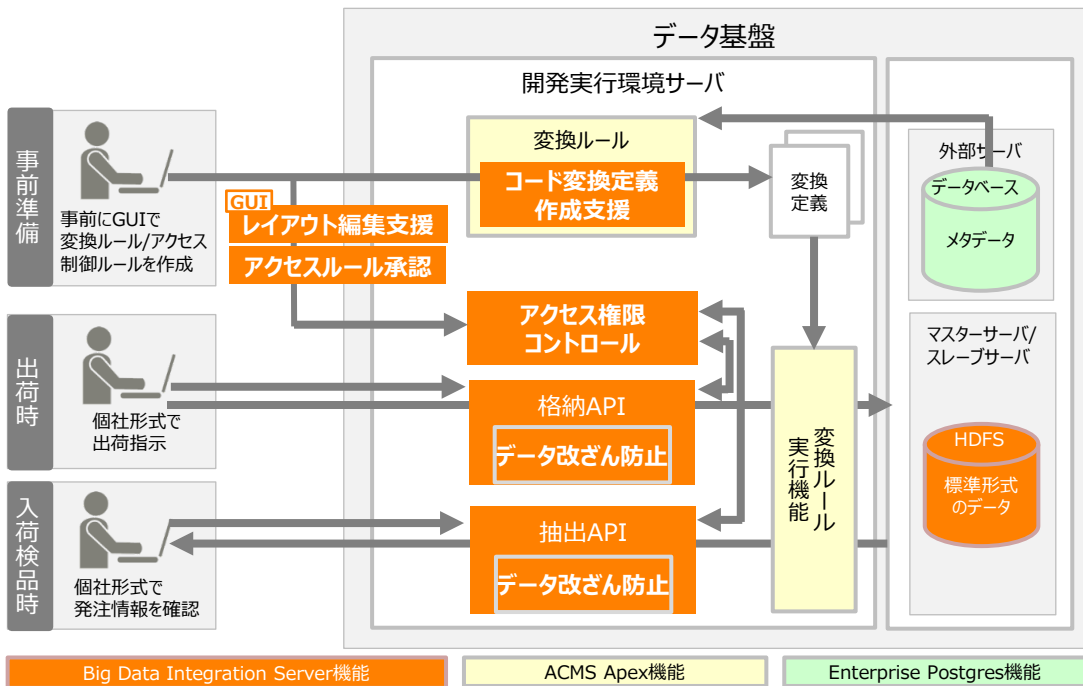


本機能では、大量データの効率的な分散・並列処理を行うオープンソース・ソフトウェア「Apache Hadoop」、およびメモリ上で複雑な処理を高速化できる「Apache Spark」の機能を提供しています。

* : Hadoopが利用する分散ファイルシステム

EDI (企業間取引データ交換)連携*

● 物流・商流分野でのデータ連携の使いやすさを向上します

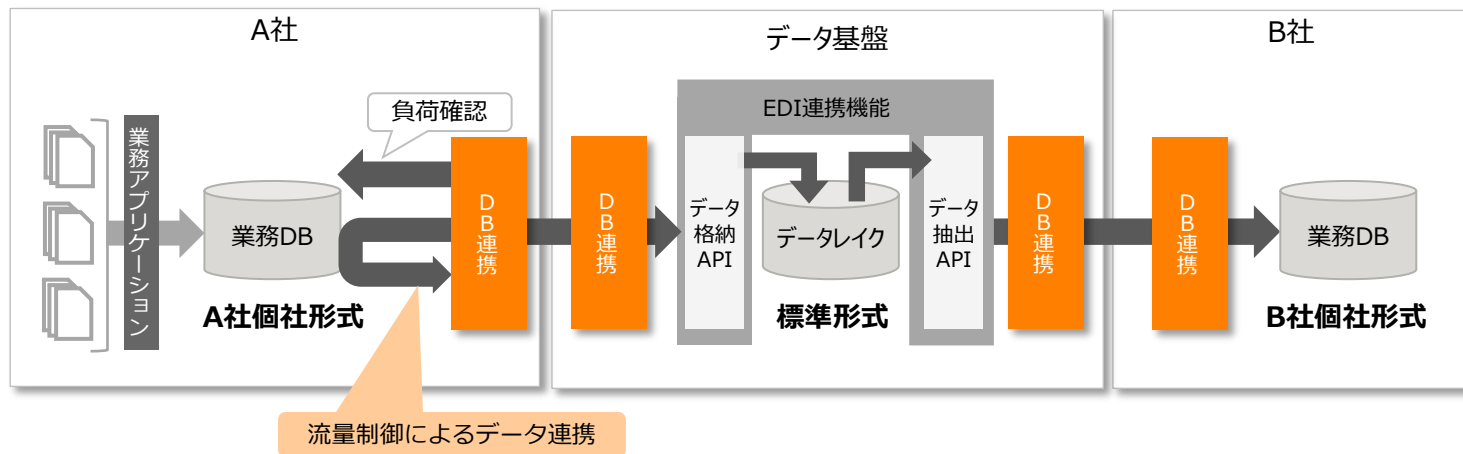


- 個社形式コードと標準形式コードの名寄せなどにより、個社形式のデータを標準化するための変換ルールの作成を支援します
- GUIによる効率的な構築操作で、データ変換ルールやアクセスルールの作成を支援します
- 作成した変換ルールをデータの格納/取だし時に適用することで、企業間の取引データの交換を効率化します
- 事前に設定したアクセスルールに応じて、ユーザー・属性ごとのアクセスや、表形式データに対するカラム単位のアクセスを制御できます
- データ改ざん防止機能によりデータレイクに格納したデータの真正性を保証します

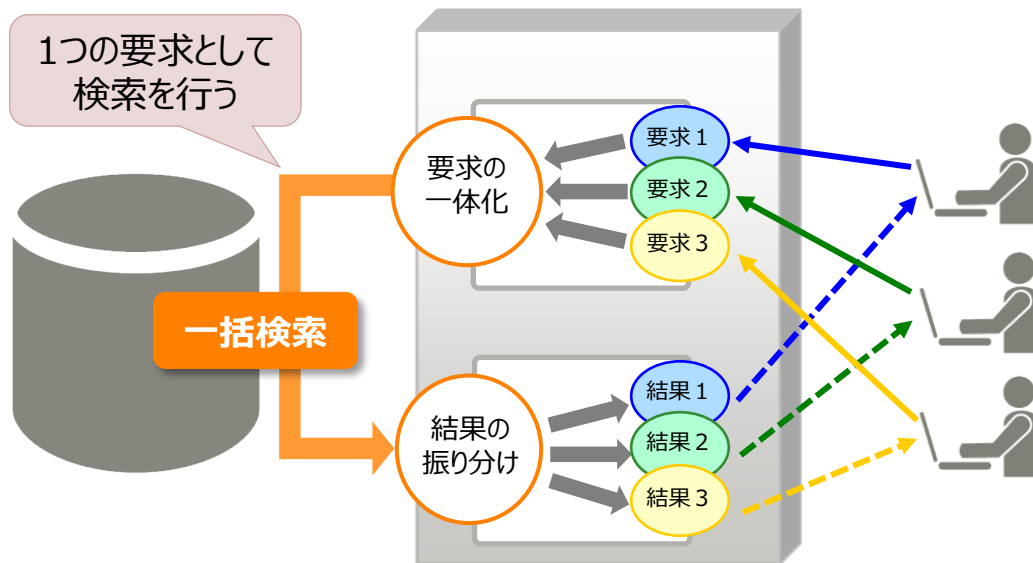
* : 本機能は、ACMS ApexおよびEnterprise Postgresとの連携によりご利用いただけます

● 業務DBの負荷に応じてスループットを自動で最適化し業務アプリケーションの性能影響を抑止

- 事前設定のみでA社の業務DBで発生した差分データを個社形式のままデータ格納APIに渡し、データ抽出APIによって取り出したB社の個社形式での差分データをB社の業務DBに反映
- データ抽出時のレスポンスからデータベースの負荷を推定し、データ連携のタイミングや量を自動で調整(流量制御)

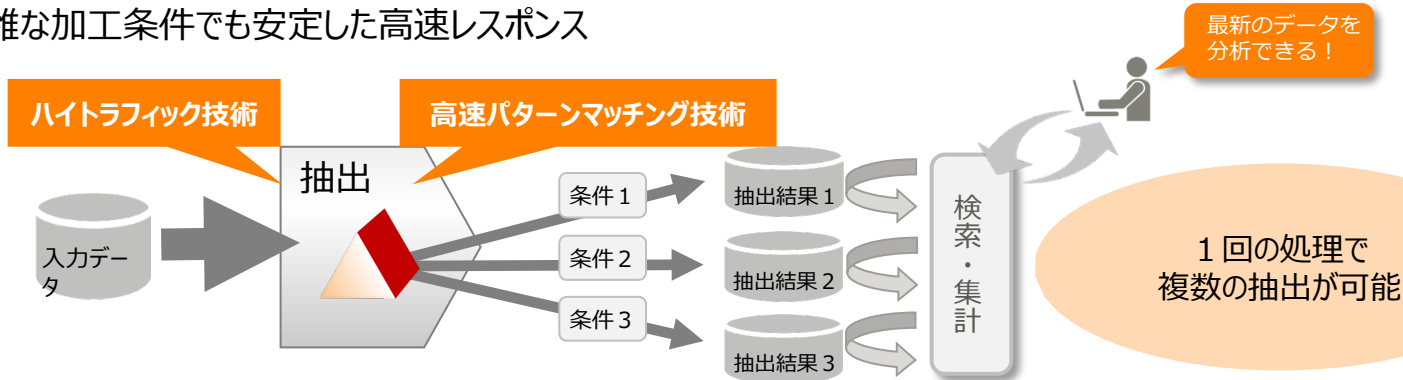


- 富士通独自のハイトラフィック技術により、複数の検索条件を取りまとめて一括検索
- アクセスが集中しても安定した検索性能を維持

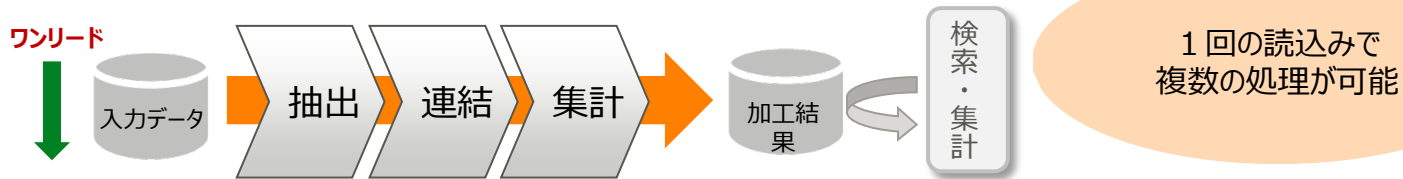


- ディスク上の大量データから、目的の情報を高速に取出し
- コマンド・定義体ベースで、効率的にデータを取り出すアプリケーションを作成

✓ 複雑な加工条件でも安定した高速レスポンス



✓ ワンリードによるデータ加工の高速化



機能名		機能概要
収集機能	ログ収集	様々なデバイス、アプリケーション、ミドルウェアから出力されるログを効率的に収集します。
	メッセージング	ログ収集で外部から集めてきたデータを一時的にメモリに保持（キューイング）し、スキーマレス加工などを使用したデータ加工が入力データとして利用できるようにします。
データ管理機能	スキーマレス加工	データの構造（スキーマ）を意識せず、そのままの形式で収集したデータから、使いたい項目だけを抽出・加工することで、検索用のデータの準備が容易にできます。
	並列分散処理	大量のデータや繰り返しの多い複雑な処理を、数十～数千台のサーバで分散処理することによって、短時間で処理できます。
	データガバナンス	収集・加工したデータにメタデータを付与し、データのアクセシビリティを向上できます。また、データリネージやセキュリティ機能を備えており、組織のデータ管理ルールに沿ってデータを管理することができます。
EDI連携機能	強化	物流・商流分野でのデータ連携の使いやすさを向上させるための機能です。
DB連携機能	強化	データ基盤と連携するデータベースの負荷を抑えながらデータ管理の安全性を高める機能です。
検索機能	インメモリ検索 強化	サマリデータや直近データなどの重要なデータをインメモリに格納し、インメモリ検索API（JavaまたはC）を使用することで、格納データの絞り込み、ソート/集計を高速にできます。
	ディスク検索	ディスク上の大量データのすべてをキーワードで検索できます。また、絞り込み、ソート/集計、項目編集、結合が可能なたため、大量データを要約（サマリ）して抽出もできます。

強化 : V1a (V1.6.0) 強化機能

製品情報

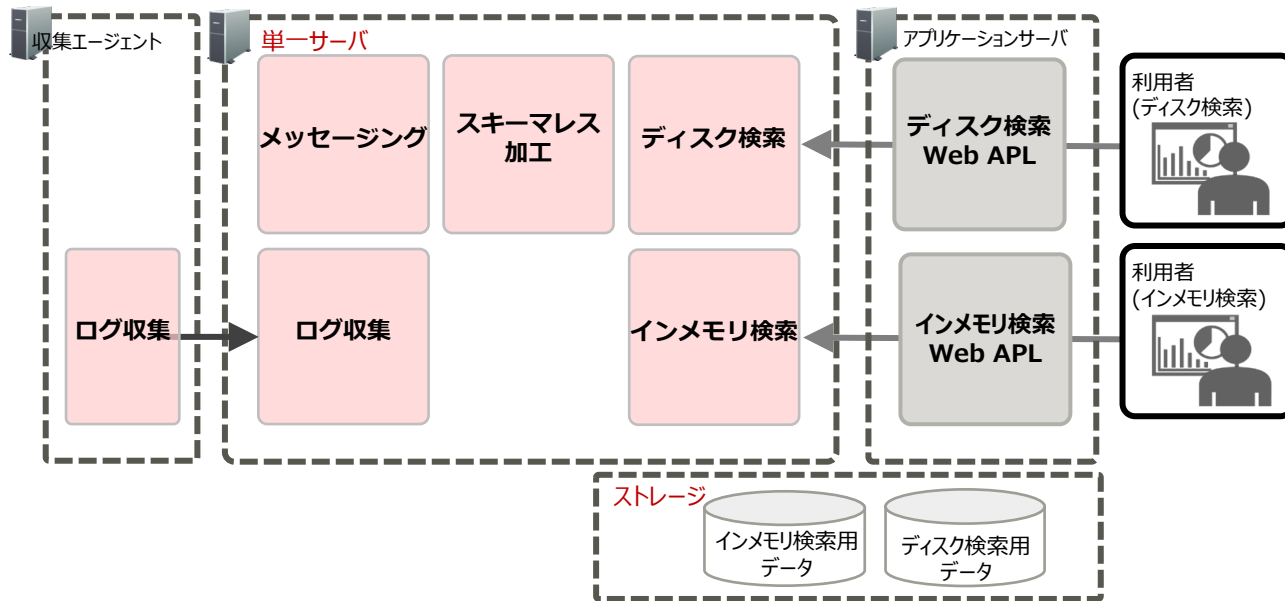
システム構成

ライセンス

同梱OSS最新化の考え方

動作環境

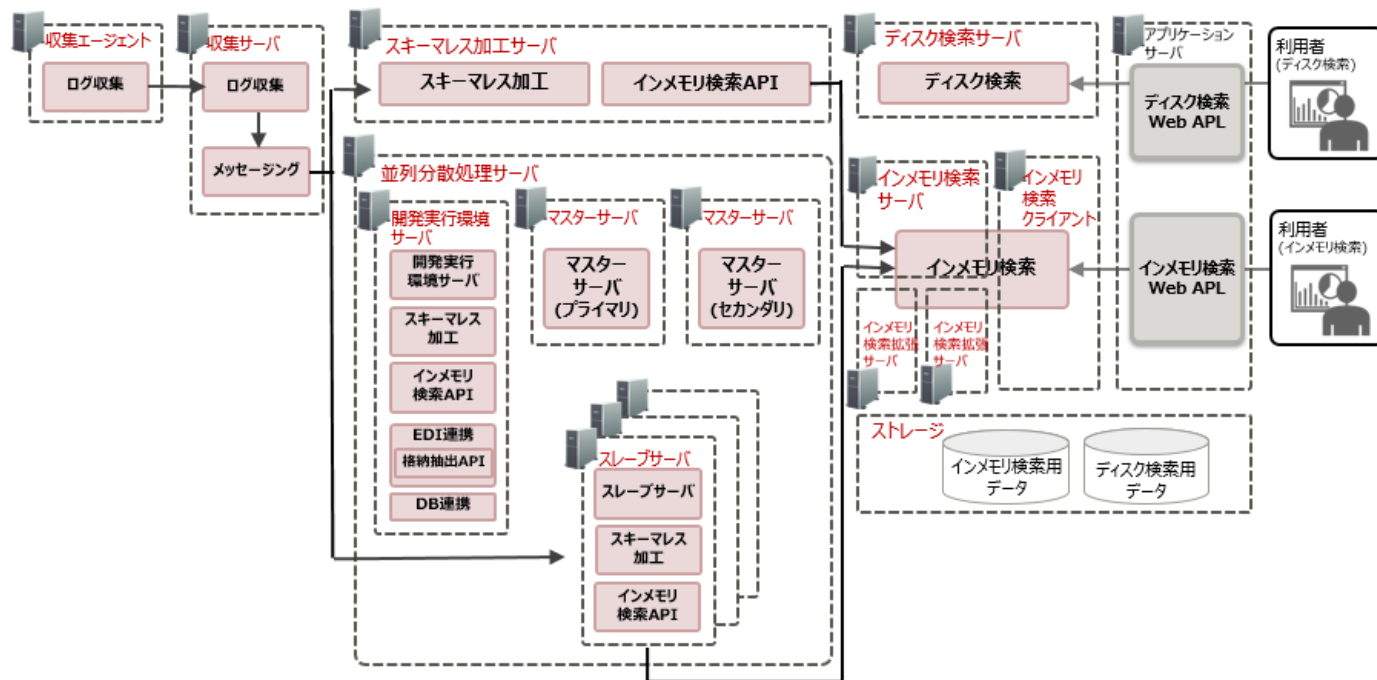
システム構成：単一サーバ構成



注) 単一サーバ構成では、並列分散処理はできません。

システム構成：マルチサーバ構成

- ログ収集、スキーマレス加工、マートレス検索でサーバを分離し、並列処理します。
- 並列分散処理でのスレーブサーバ増設により、スキーマレス加工をさらに高速化します。



システム構成：サーバ概要

サーバ種別		説明	
収集機能	収集サーバ	ログ収集を配置するサーバです。	
	収集エージェント	ログ収集を配置するサーバです。	
データ管理機能	スキーマレス加工サーバ	データ管理機能を配置するサーバです。	
	並列分散処理 (注1)	マスターサーバ	大量のデータファイルをブロックに分けた上でファイル化(分散ファイルシステム)し、そのファイル名や保管場所を一元管理するサーバです。プライマリとセカンダリの2台構成が基本です。
		スレーブサーバ	マスターサーバによってブロック化されたデータファイルを並列に分散処理するサーバです。複数台用意します。
		開発実行環境サーバ (注2)	並列分散のアプリケーションを実行するサーバです。
検索機能	インメモリ検索サーバ	検索機能のうち、インメモリ検索を配置するサーバです。	
	インメモリ検索拡張サーバ	インメモリ検索の性能拡張をする場合に配置するサーバです。	
	ディスク検索サーバ	検索機能のうち、ディスク検索を配置するサーバです。	
インメモリ検索クライアント		インメモリ検索のうち、検索要求の受付け、検索ユーザー認証・許可、同時実行制御、検索サーバへの検索実行依頼、検索結果返却を受け持つコンピュータです。	
ストレージ		検索サーバの検索対象データを格納する記憶装置です。	

注1) データガバナンスも配置されます。

注2) EDI連携機能およびDB連携機能も配置されます。

● ライセンス形態と課金対象

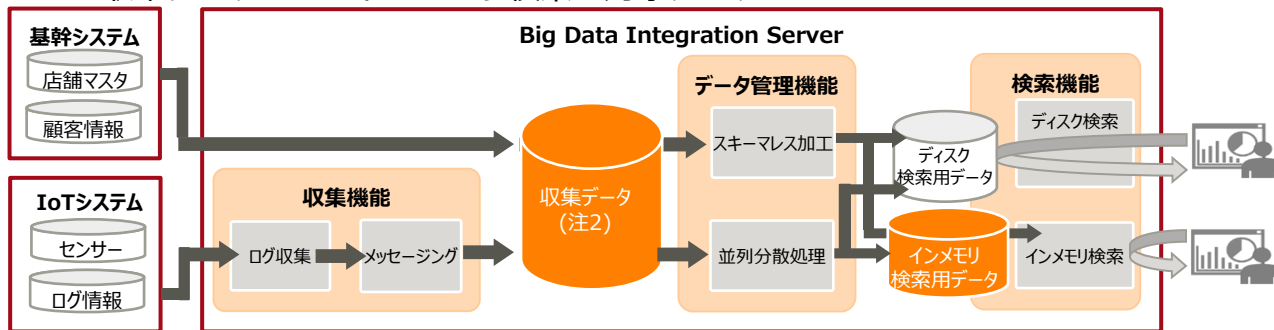
- ライセンス形態：サブスクリプションライセンス/サポート（使用权とサポートを組み合わせた商品(注1)）

サービス商品		データ容量(TB)	月額支払い価格(円)
標準プラン	Big Data Integration Server 標準プラン 基本 1テラバイト for Linux (SL&S)	1	916,740
	Big Data Integration Server 標準プラン 追加 1テラバイト for Linux (SL&S)	1	320,870
大容量プラン	Big Data Integration Server 大容量プラン 基本 10テラバイト for Linux (SL&S)	10	2,750,000
	Big Data Integration Server 大容量プラン 追加 1テラバイト for Linux (SL&S)	1	229,240

注1：詳細については、お問合せください。

備考. 商品の導入にあたり、最低 1 本のメディアパックが必要となります(別売：1万円/枚)。メディアパックは媒体（DVD等）のみの提供で、使用权はありません。

- 課金対象：収集データ量 + インメモリ検索の対象データ量



注2) スキーマレス加工処理および並列分散処理の対象となる入力データを指します。

：課金対象データ

本製品は、バージョンアップ/レベルアップで同梱のオープンソース・ソフトウェア(OSS)を最新化し、新機能や障害修正を提供します。そのため、定期的に製品の最新バージョン/レベルに入れ替えることを推奨します。

本製品の機能に活用しているOSSは以下の通りです。

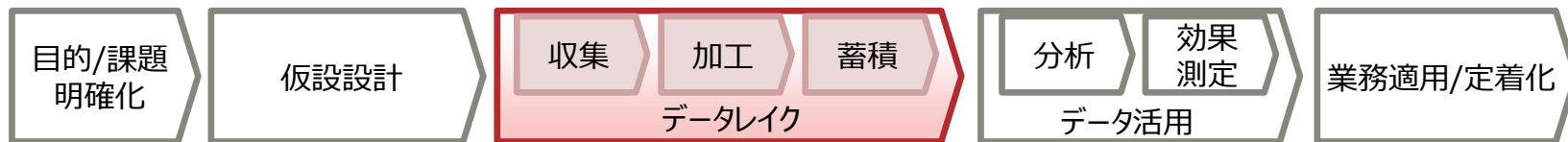
機能名	OSS名
ログ収集	Fluentd
メッセージング	Apache Kafka、Apache ZooKeeper
並列分散処理	Apache Hadoop、Apache Spark、Apache Hive
データガバナンス	Apache Atlas、Apache Ranger
EDI連携	NextCloud、Open Policy Agent、XmlNotepad

サーバ種別		説明	
収集機能	収集サーバ	Red Hat Enterprise Linux 8.2以降 (for Intel64) Red Hat Enterprise Linux 9.0以降 (for Intel64)	
	収集エージェント		
データ管理機能	スキーマレス加工サーバ		
	並列分散処理		マスターサーバ
			スレーブサーバ
	開発実行環境サーバ		
検索機能	インメモリ検索サーバ		
	インメモリ検索拡張サーバ		
	ディスク検索サーバ		
インメモリ検索クライアント			
ストレージ			

関連サービス

ミドルウェア開発/実務経験で培ってきた技術力・現場力で
Big Data Integration Serverを適用するサービスをご提供

データ利活用プロセス



データ利活用スペシャリストサービス

サービス体系

Big Data Integration Server 環境設計支援

Big Data Integration Server 環境構築支援

- Oracleは、Oracle Corporationおよびその子会社、関連会社の米国およびその他の国における登録商標です。
- Linux®は、米国およびその他の国におけるLinus Torvaldsの登録商標です。
- Red Hat、RPMおよびRed Hatをベースとしたすべての商標とロゴは、Red Hat, Inc.の米国およびその他の国における登録商標または商標です。
- Apache Atlas、Apache Hadoop、Apache Hive、Apache Kafka、Apache Ranger、Apache Spark、Apache Tomcat、Apache ZooKeeper、HDFSは、The Apache Software Foundationの米国およびその他の国における商標または登録商標です。
- Fluentdは、トレジャーデータ株式会社の登録商標です。
- ACMS Apexは、株式会社データ・アプリケーションの日本および海外での商標ならびに登録商標です。
- Pythonは、Python Software Foundationの登録商標です。
- Interstageは、富士通株式会社の登録商標です。
- 記載されている製品名、会社名などの固有名詞は、各社の登録商標または商標です。
- その他、本資料に記載されているシステム名、製品名などは、必ずしも商標表示を付記していません。

Thank you

