

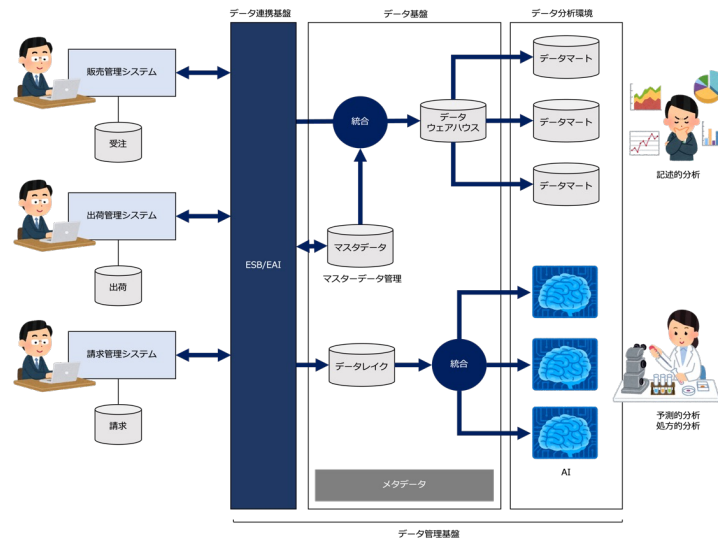
データ管理基盤

- データ管理基盤とは
- リレーショナルデータベース
- データウェアハウス
- データレイク
- NoSQL
- マスターデータ管理
- データ統合
- データ連携
- データ管理基盤の設計

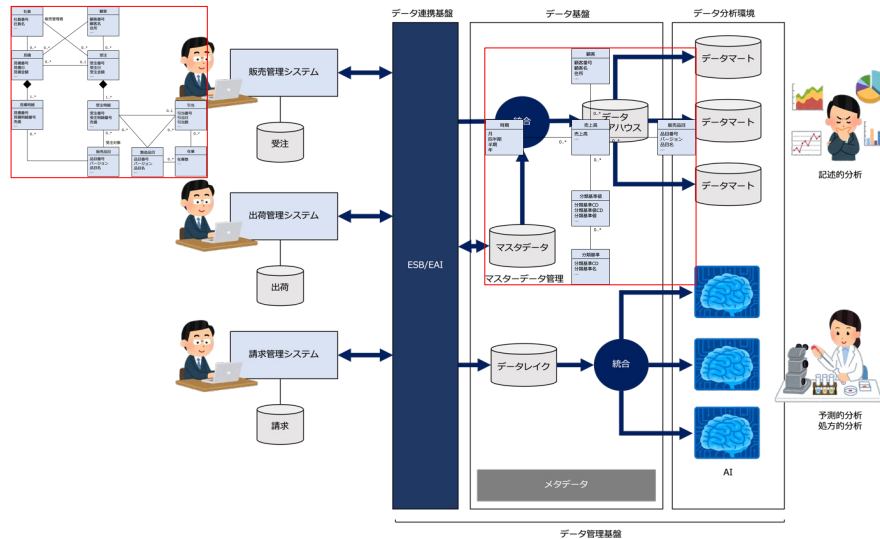
データ管理基盤

- データ管理基盤とは
- リレーショナルデータベース
- データウェアハウス
- データレイク
- NoSQL
- マスターデータ管理
- データ統合
- データ連携
- データ管理基盤の設計

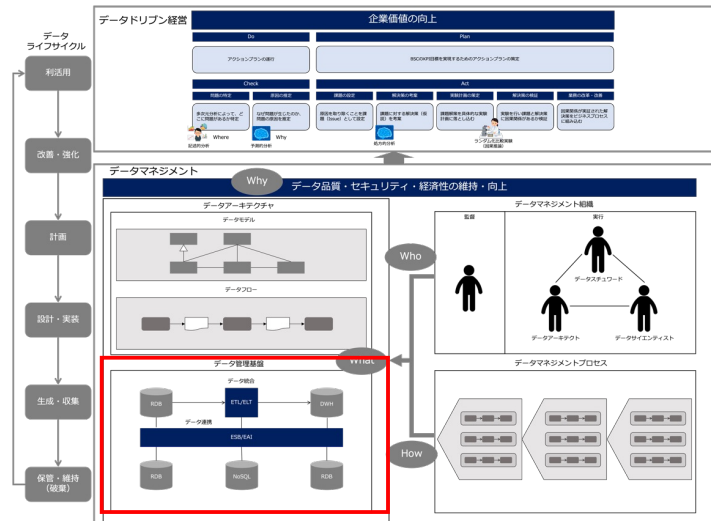
データ管理基盤とは



データ管理基盤とは



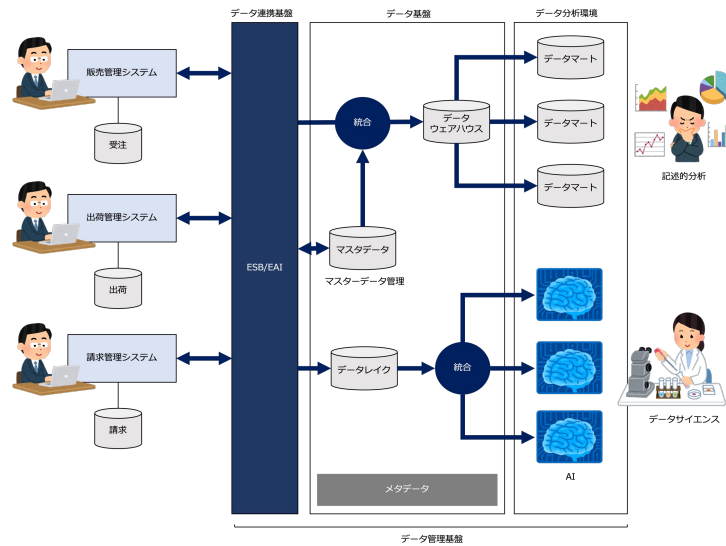
データ管理基盤とは



論理レベルのデータアーキテクチャの設計

	設計レベル	戦略レベル	実例レベル
経営方針	・経営理念 ・事業ドメイン ・事業パーパス ・事業成長モデル ・戦略マップ	・事業ビジョン ・事業単位 ・事業戦略	・事業実績
BA	【静的モデル】 ・組織（ジョブ）構成 ・資産（タイプ）構成、場所（タイプ）構成 【動的モデル】 ・ビジネスプロセス構成 ・業務フロー	【静的モデル】 ・組織（部門）構成 ・資産（種類）構成、場所（種類）構成 【動的モデル】 ・アクションプラン	【静的モデル】 ・資産（実例）構成、場所（実例）構成 【動的モデル】 ・ワークフロー
DA	【静的モデル】 ・エンタープライズデータモデル（概念レベル） 【動的モデル】 ・データフロー（概念レベル）	【静的モデル】 ・エンタープライズデータモデル（論理レベル） 【動的モデル】 ・データフロー（論理レベル）	【静的モデル】 ・エンタープライズデータモデル（物理レベル） 【動的モデル】 ・データフロー（物理レベル）
AA	【静的モデル】 ・アプリケーション構成（概念レベル） 【動的モデル】 ・アプリケーション連携モデル（概念レベル）	【静的モデル】 ・アプリケーション構成（論理レベル） 【動的モデル】 ・アプリケーション連携モデル（論理レベル）	【静的モデル】 ・アプリケーション構成（実装レベル） 【動的モデル】 ・アプリケーション連携モデル（実装レベル）
TA	・IT基盤構成（方針レベル） ・コミュニケーション基盤構成（方針レベル） ・データ管理基盤構成（方針レベル） ・アプリケーション基盤構成（方針レベル） ・BPM基盤構成（方針レベル）	・IT基盤構成（製品レベル） ・コミュニケーション基盤構成（製品レベル） ・データ管理基盤構成（製品レベル） ・アプリケーション基盤構成（製品レベル） ・BPM基盤構成（製品レベル）	・IT基盤構成（実装レベル） ・コミュニケーション基盤構成（実装レベル） ・データ管理基盤構成（実装レベル） ・アプリケーション基盤構成（実装レベル） ・BPM基盤構成（実装レベル）

データ管理基盤とは



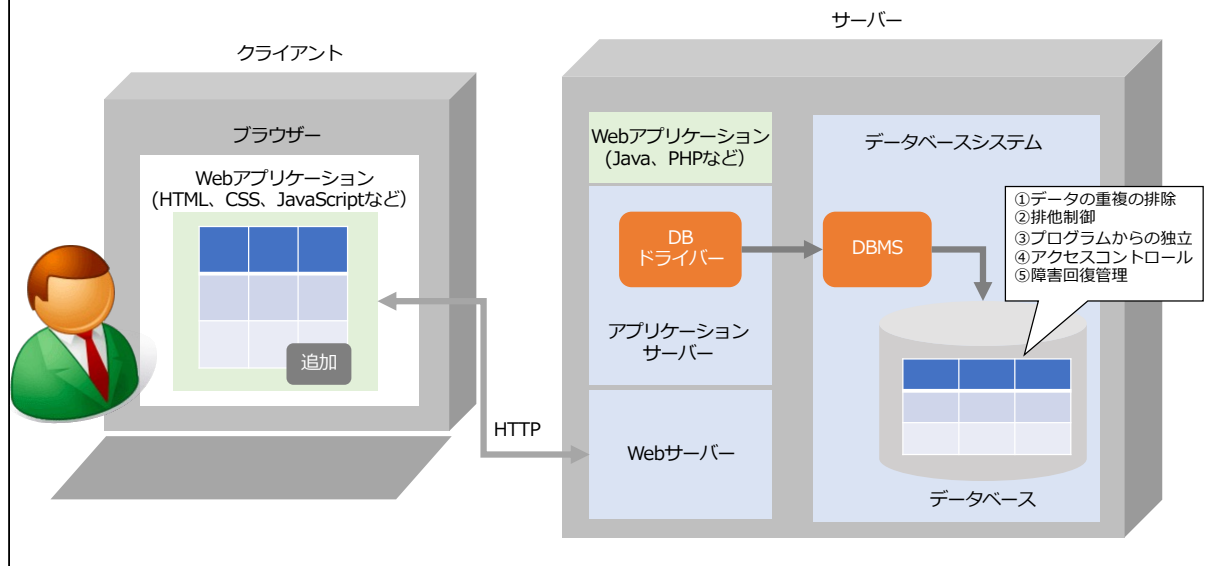
データ管理基盤

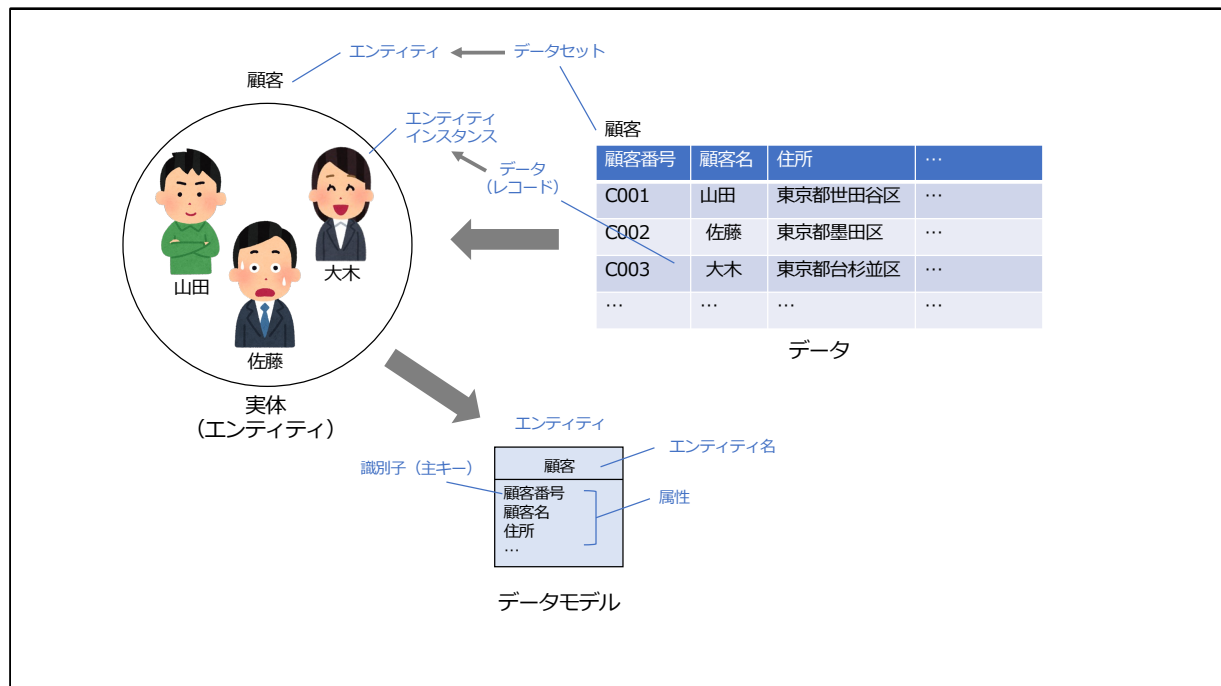
- データ管理基盤とは
- リレーショナルデータベース
- データウェアハウス
- データレイク
- NoSQL
- マスターデータ管理
- データ統合
- データ連携
- データ管理基盤の設計

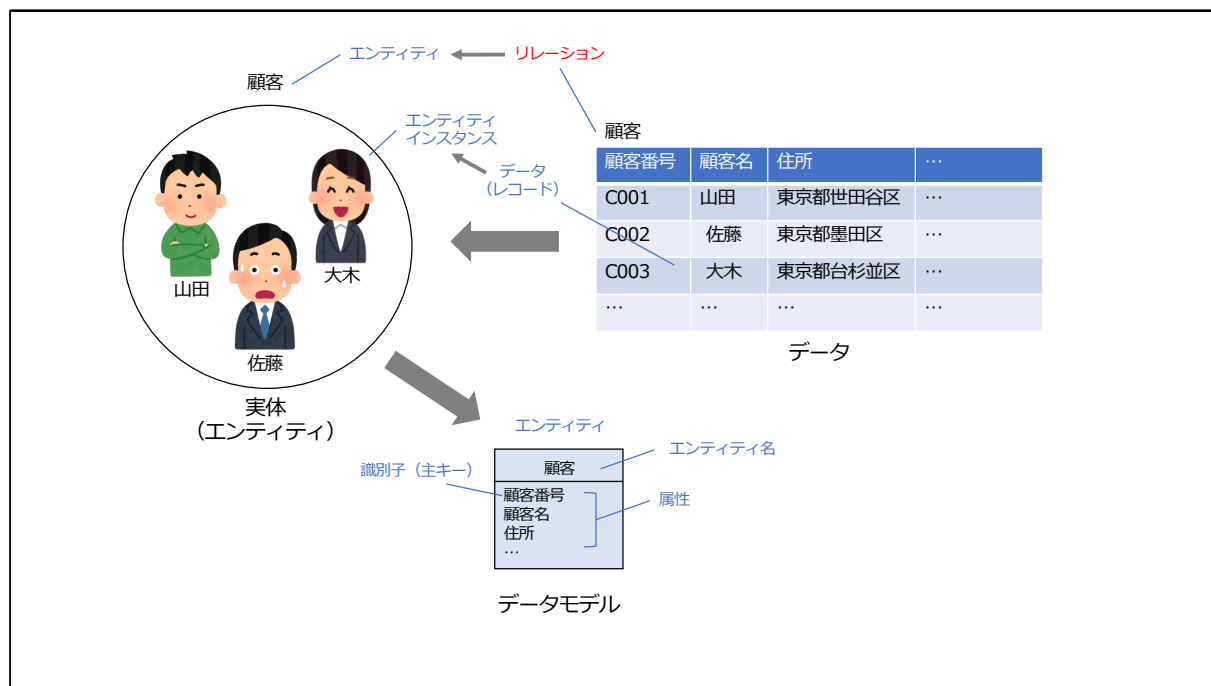
データ管理基盤

- データ管理基盤とは
- リレーショナルデータベース
- データウェアハウス
- データレイク
- NoSQL
- マスターデータ管理
- データ統合
- データ連携
- データ管理基盤の設計

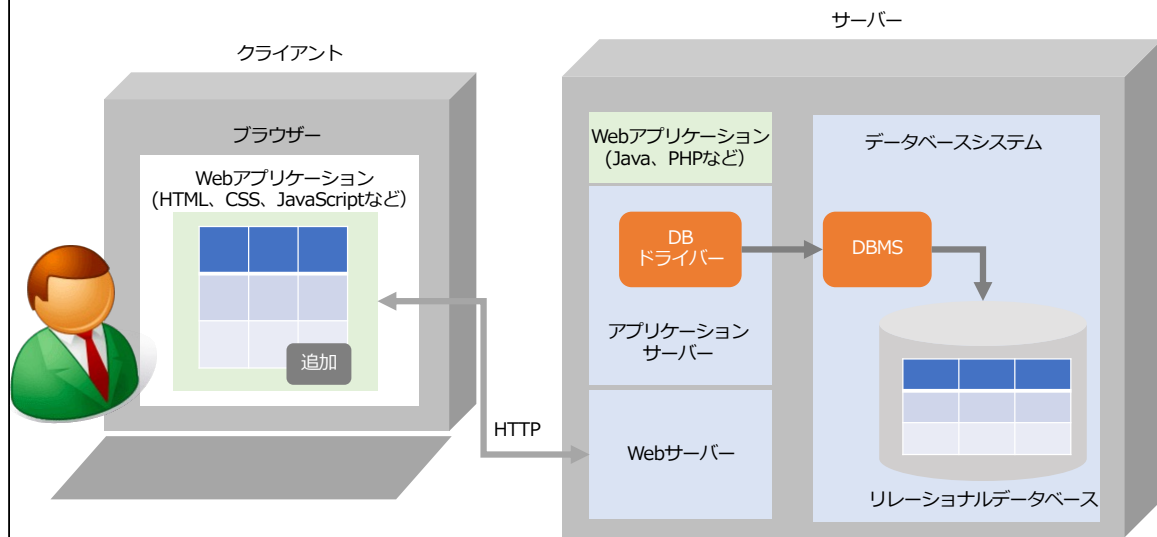
データベース



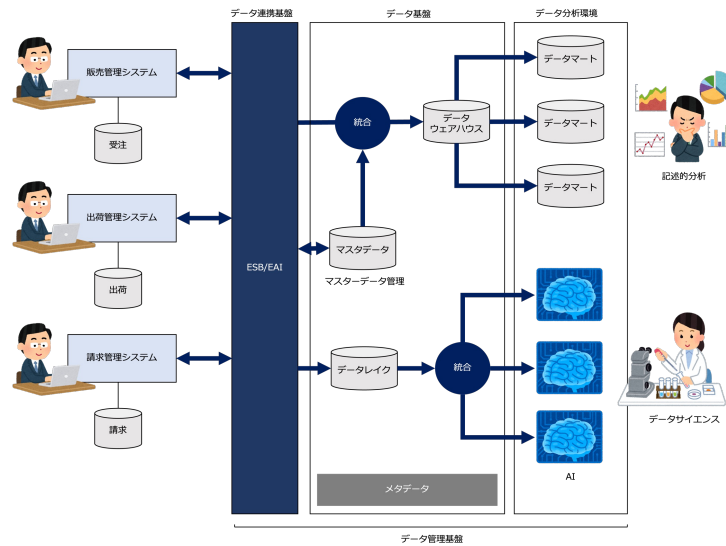




リレーショナルデータベース



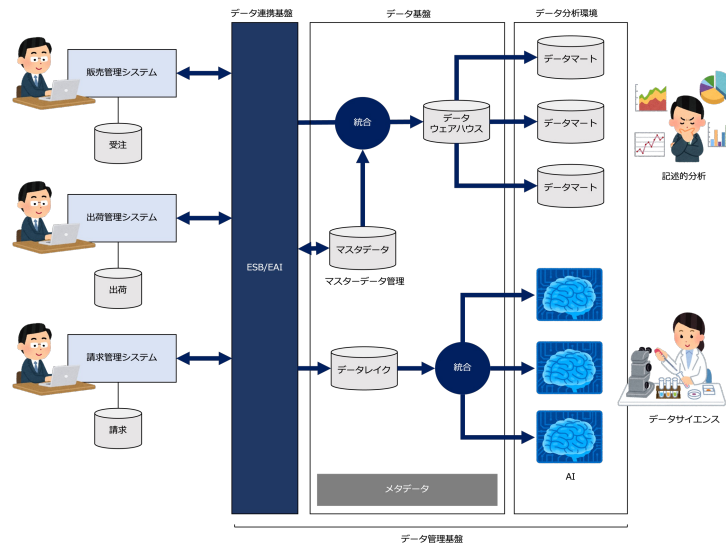
データ管理基盤



データ管理基盤

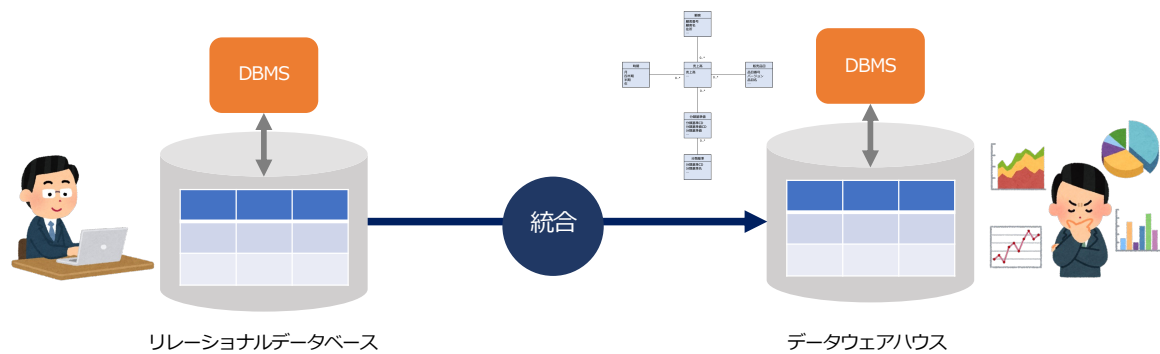
- データ管理基盤とは
- リレーショナルデータベース
- データウェアハウス
- データレイク
- NoSQL
- マスターデータ管理
- データ統合
- データ連携
- データ管理基盤の設計

データ管理基盤



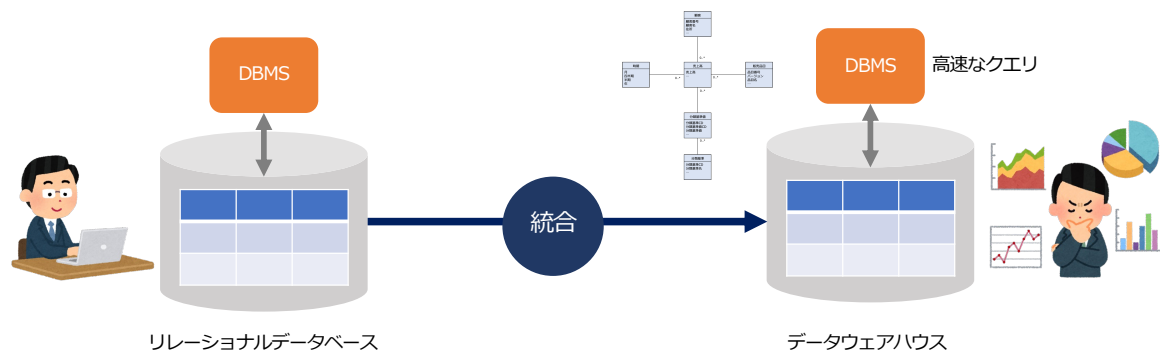
データウェアハウス

大量のデータを収集・統合し、組織全体での分析や意思決定支援に使用するためのデータベースシステム。



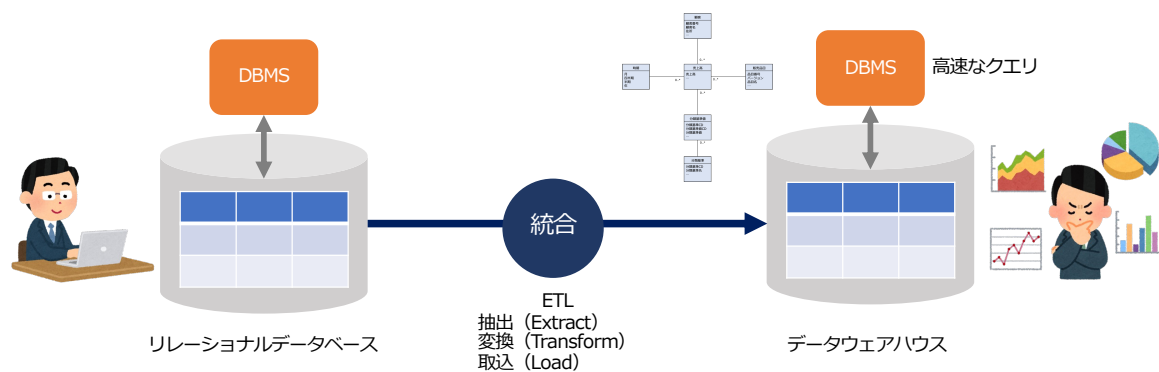
データウェアハウス

大量のデータを収集・統合し、組織全体での分析や意思決定支援に使用するためのデータベースシステム。

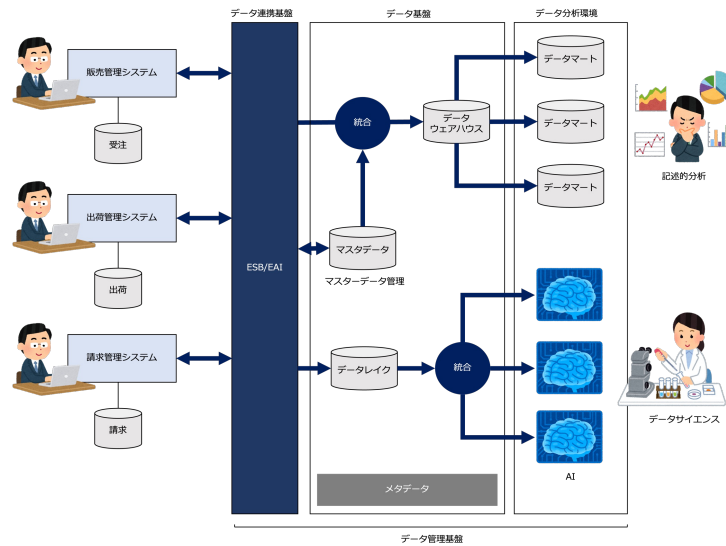


データウェアハウス

大量のデータを収集・統合し、組織全体での分析や意思決定支援に使用するためのデータベースシステム。



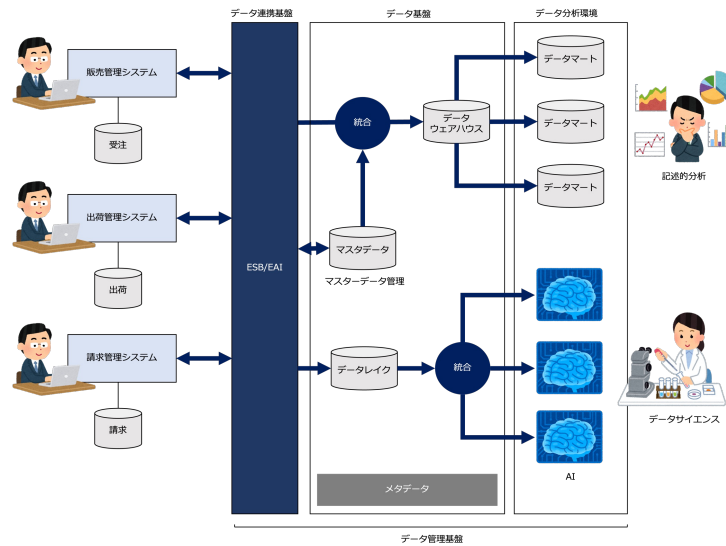
データ管理基盤



データ管理基盤

- データ管理基盤とは
- リレーショナルデータベース
- データウェアハウス
- データレイク
- NoSQL
- マスターデータ管理
- データ統合
- データ連携
- データ管理基盤の設計

データ管理基盤

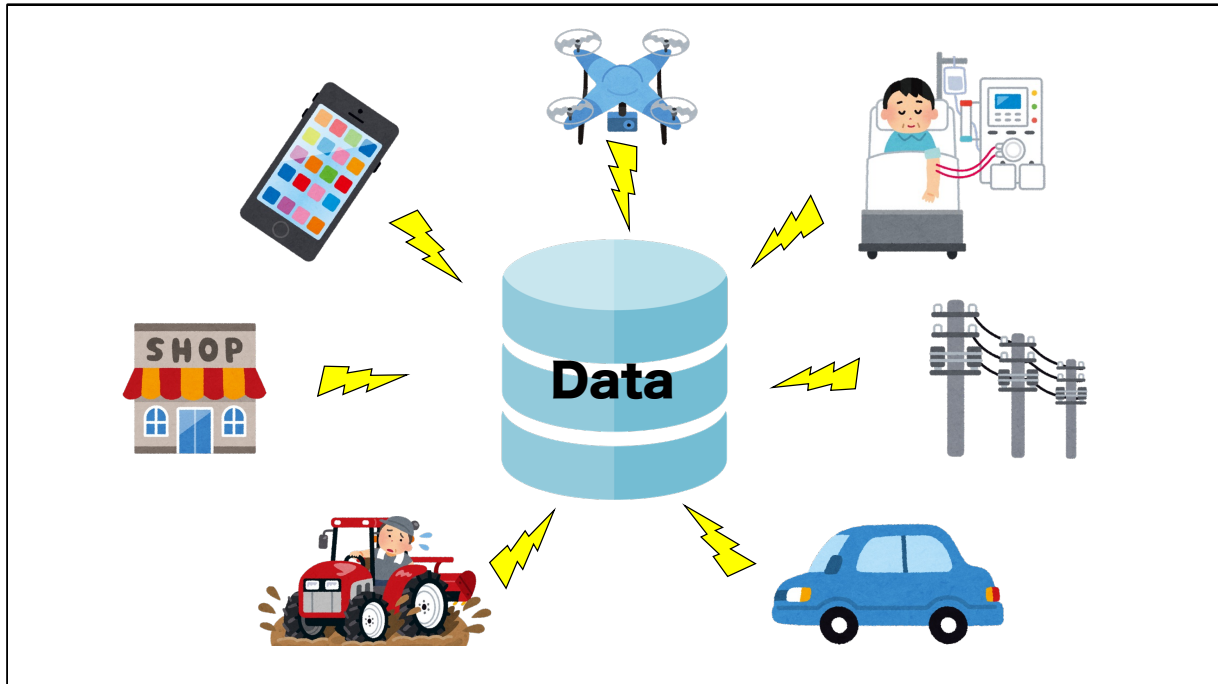


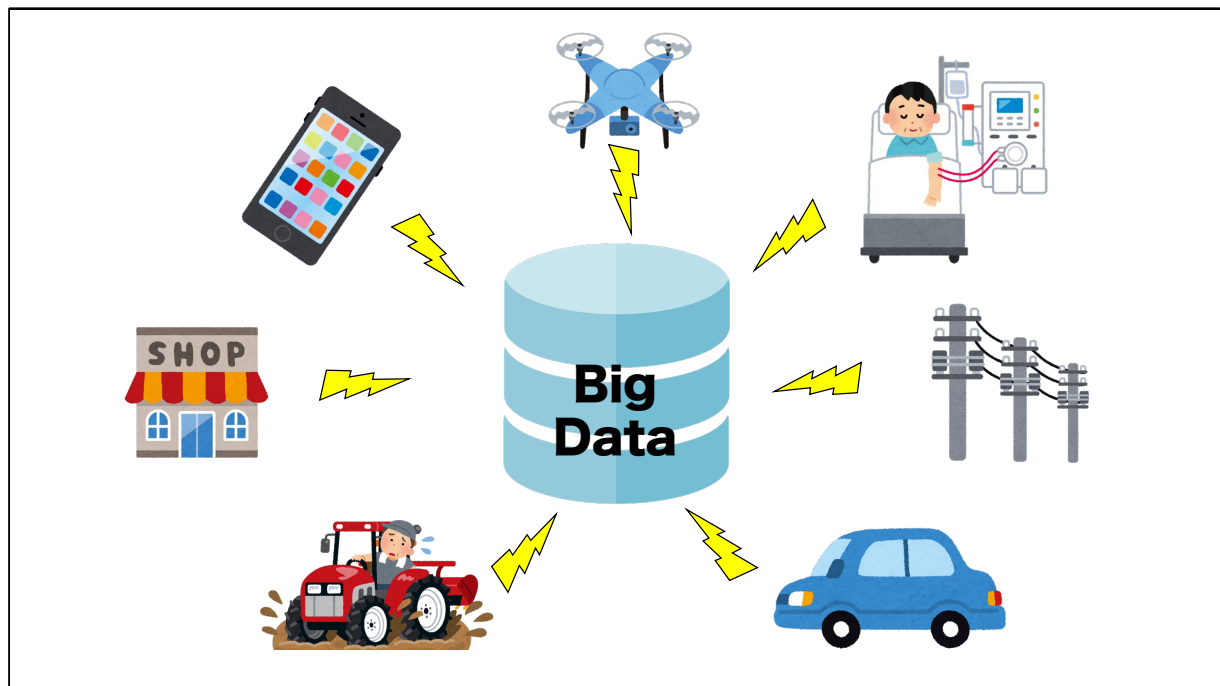
データレイク

さまざまな種類や構造を持った膨大な量のデータをインGESTし保存し評価し分析できる環境。

- データサイエンティストがデータをマイニングし分析するための環境となる
- データ変換が必要であっても最小限ですむような、生データの中央記憶装置となる
- 詳細な履歴データを持つデータウェアハウスの代替ストレージとなる

データマネジメント知識体系 第二版 (DMBOK2)



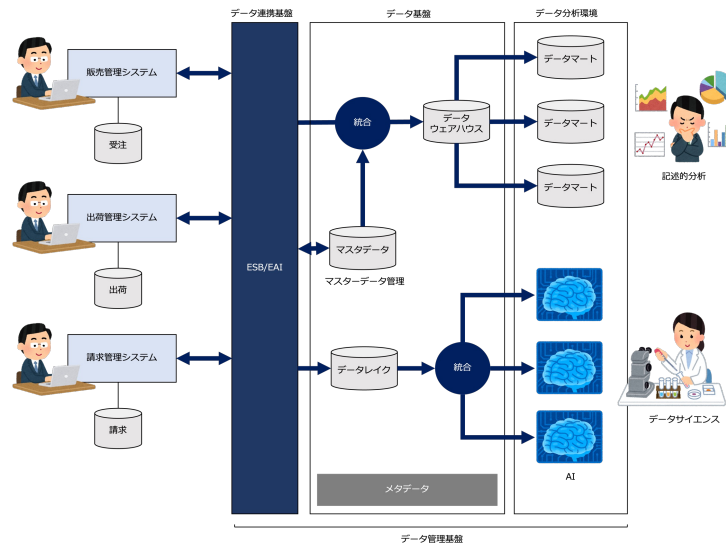


Big Dataの特徴

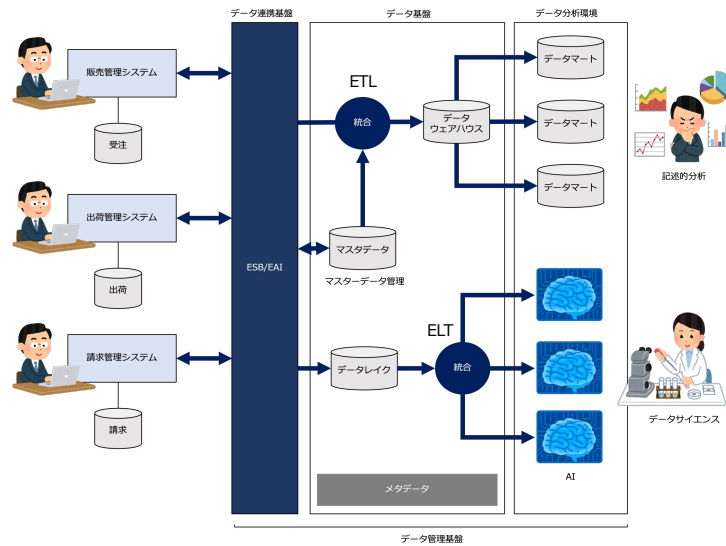
- **Volume**
データ量の多さ。
- **Variety**
データ形式の多様さ。
- **Velocity**
データの処理の速さ。



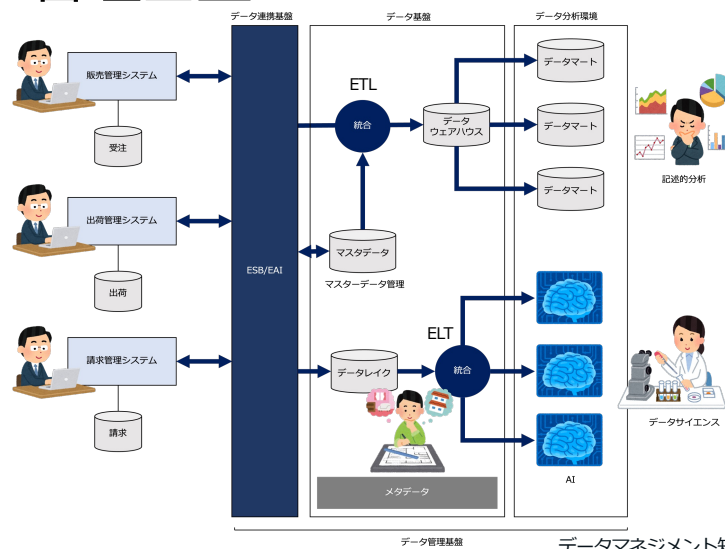
データ管理基盤



データ管理基盤



データ管理基盤



記述的分析

データレイク (湖) の危険性は、それがすぐにデータスワンプ (沼) になってしまうことである。つまり乱雑で、汚れていて、一貫性がないものになってしまう。データレイクに含まれているデータを台帳化するには、データがインGESTされる時にメタデータを管理すべきである。

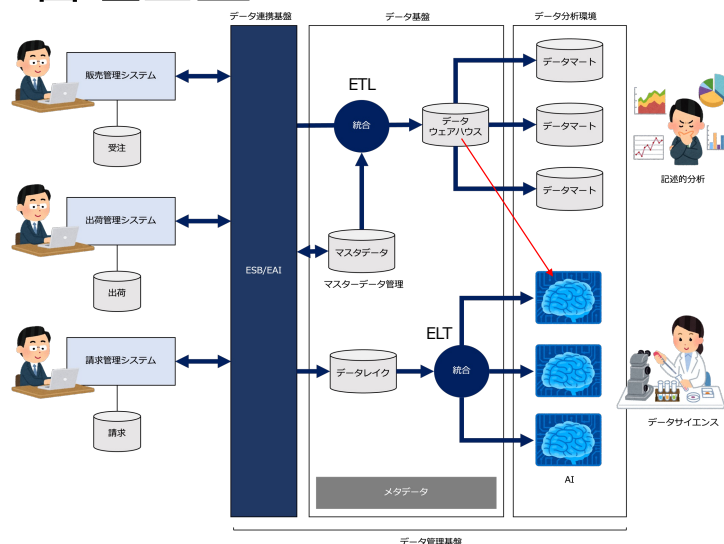


データサイエンス

データ管理基盤

データマネジメント知識体系 第二版 (DMBOK2)

データ管理基盤



ここでは、多次元分析のような記述的分析をする場合は、DWHを使い、さまざまなデータをAIがマイニングする必要があるデータサイエンスにはDLを活用するという例を示しています。
もちろん、データウェアハウスの構造化データをAIが活用することも可能です。

データ管理基盤

- データ管理基盤とは
- リレーショナルデータベース
- データウェアハウス
- データレイク
- NoSQL
- マスターデータ管理
- データ統合
- データ連携
- データ管理基盤の設計

NoSQLとは

リレーショナルデータベースの表形式のように厳密に構造化できないが、IoTからのデータなど、様々な種類の値に対応できるデータ項目の関係を構造化したい場合、NoSQLデータベースを検討します。

リレーショナルデータベースやデータウェアハウスが表形式で構造化されたデータを扱い、データレイクがさらに非構造化データを扱うのに対し、NoSQLは、半構造化データを扱うデータベースです。

リレーショナルデータベースでは、データの問合せにSQL（Structured Query Language：構造化照会言語）という言葉を使います。

NoSQLはNot only SQLの略で、データはSQL以外でも扱えることを強調した表現です。

NoSQLの種類

- **ドキュメント指向データベース (Document Stores)**

MongoDBなどが代表的。JSONやBSONなどのドキュメント形式でデータを格納し、柔軟なスキーマを提供します。

- **キー・バリュー型データベース (Key-Value Stores)**

RedisやAmazon DynamoDBなどが代表的。キーと値の単純なペアでデータを格納し、高速な読み書きを提供します。

- **ワイドカラムストア (Wide-Column Stores)**

Apache CassandraやHBaseなどが代表的。行ごとに複数の列族 (Column Family) を持つことができ、スケーラビリティが高いです。

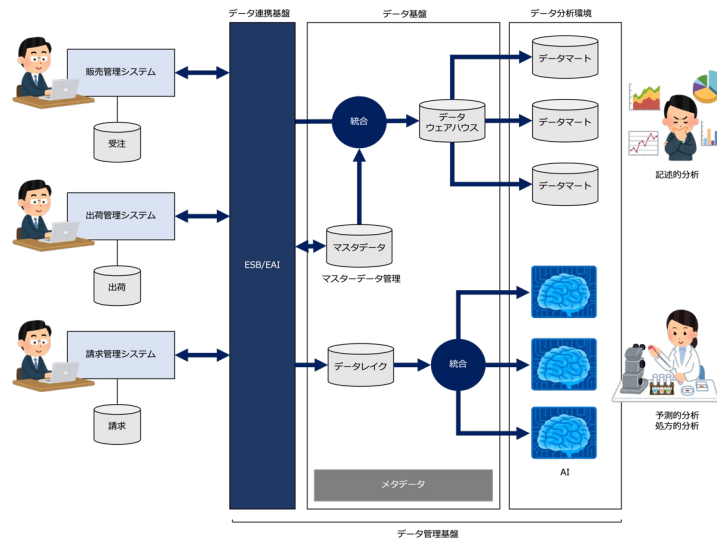
- **グラフデータベース (Graph Databases)**

Neo4jやAmazon Neptuneなどが代表的。データの関係性をグラフ構造で表現し、関連するデータを効率的にクエリできます。

データ管理基盤

- データ管理基盤とは
- リレーショナルデータベース
- データウェアハウス
- データレイク
- NoSQL
- マスターデータ管理
- データ統合
- データ連携
- データ管理基盤の設計

データ管理基盤とは



マスターデータ管理とは

マスターデータ管理（Master Data Management、MDM）は、組織が所有するさまざまなデータのうち、特に重要で共有されるマスターデータを一元管理し、一貫性を保ちながら利用するための手法やプロセスを指します。

マスターデータの種類

- **パーティマスターデータ**

パーティマスターデータのパーティは、Party、人の集まりを表しています。なので、顧客データや、社員データ、取引先のデータが該当します。

- **財務マスターデータ**

事業単位、コストセンター、プロフィットセンター、勘定科目、予算、財務予測、プロジェクトなど財務に関するデータのことです。

- **法令マスターデータ**

契約や規則などの法的事項に関するデータが含まれています。

- **製品マスターデータ**

社内の製品やサービス、業界全体（競合他社含む）の製品やサービスのデータのことです。
製品や部品の品目を表す品目マスターも製品マスターとして考えることができます。

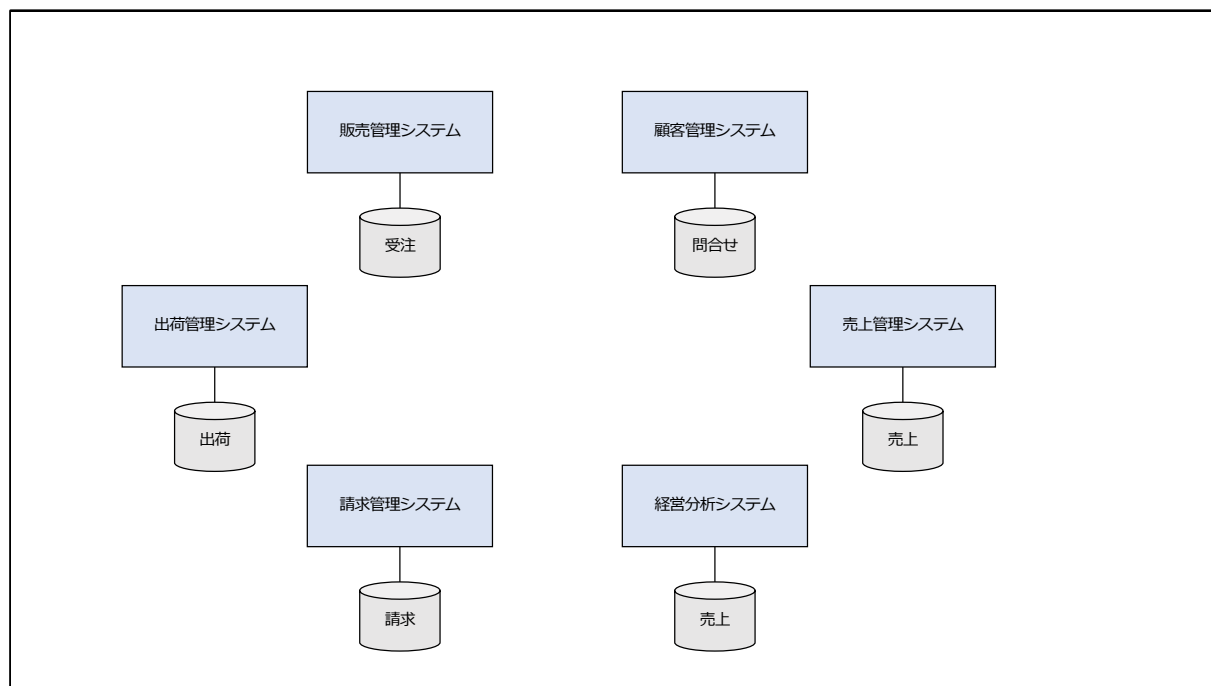
- **ロケーションマスターデータ**

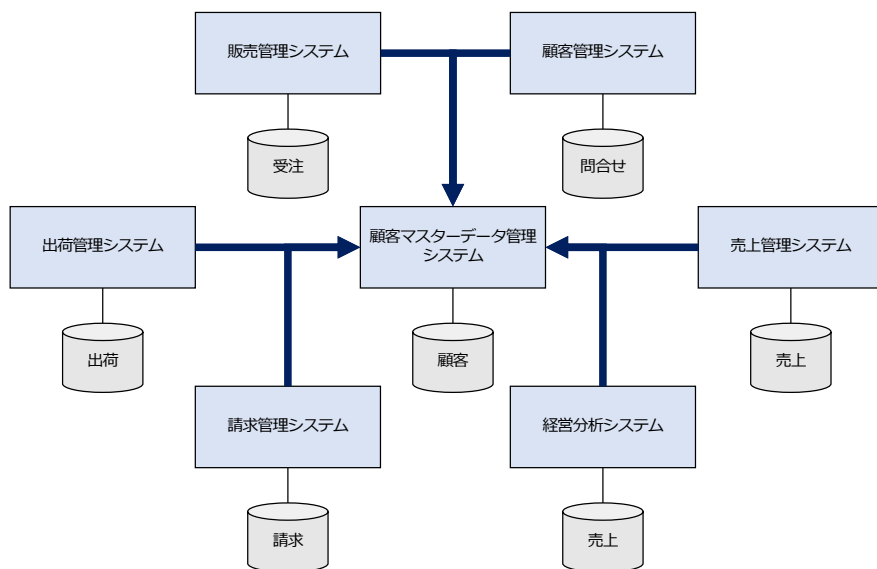
拠点に関するデータのこと、地理情報を追跡して共有し、それに基づいて階層関係や担当地域を生成する機能を持ちます。

- **業種マスターデータ**

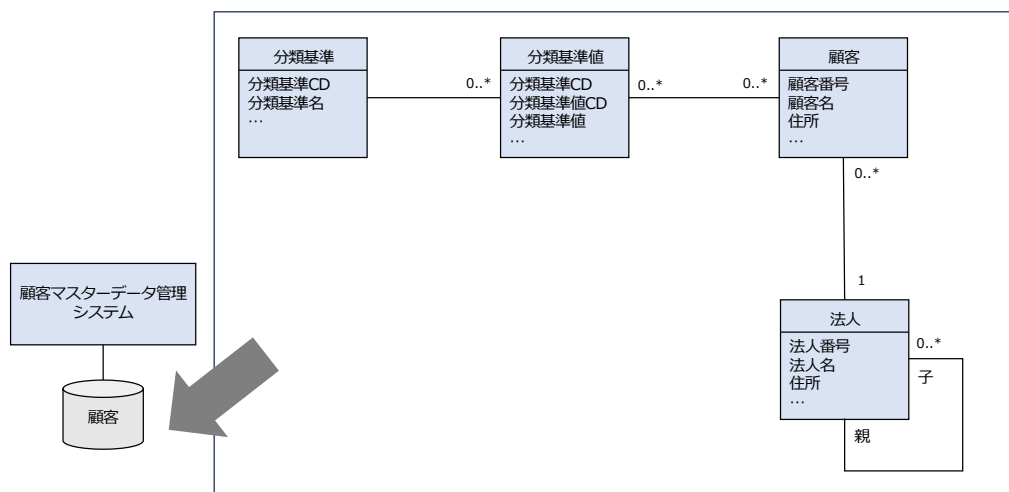
業界で共有できるデータのことです。例えば、米国医師会の処方医師データベースなど、外部から購入できるマスターデータがあります。

データマネジメント知識体系 第二版（DMBOK2）





顧客マスターデータ



マスターデータ管理の仕組（アーキテクチャ）

- レジストリ
マスターデータ管理システムは、さまざまなシステムのマスターデータを指すインデックス（レジストリ）になる。
- トランザクションハブ
マスターデータ管理システムは、マスターデータのハブとなる。
マスターデータは、マスターデータ管理システムのみに存在し、他のシステムには存在しない。
- 統合アプローチ
レジストリとトランザクションハブのハイブリット。
各システムにあるマスターデータが、マスターデータ管理システムで統合され、他のシステムから利用される。

マスターデータ管理の仕組（アーキテクチャ）

- **レジストリ**

マスターデータ管理システムは、さまざまなシステムのマスターデータを指すインデックス（レジストリ）になる。

- **トランザクションハブ**

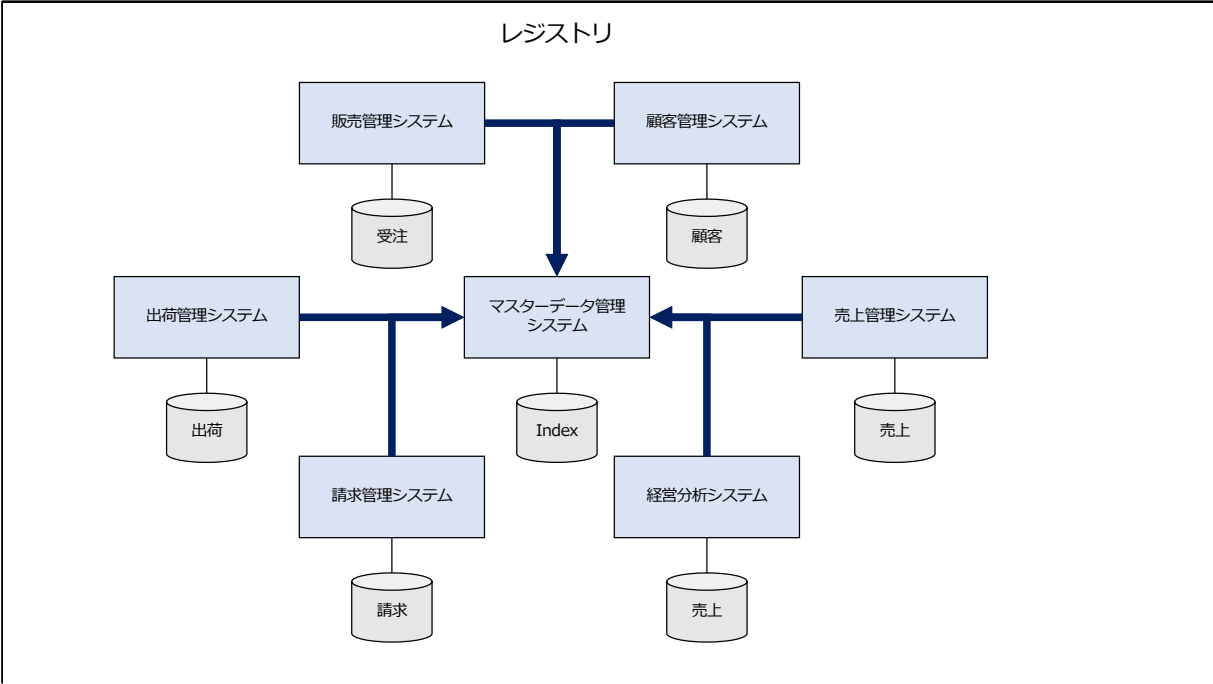
マスターデータ管理システムは、マスターデータのハブとなる。

マスターデータは、マスターデータ管理システムのみに存在し、他のシステムには存在しない。

- **統合アプローチ**

レジストリとトランザクションハブのハイブリット。

各システムにあるマスターデータが、マスターデータ管理システムで統合され、他のシステムから利用される。



マスターデータ管理の仕組（アーキテクチャ）

- レジストリ

マスターデータ管理システムは、さまざまなシステムのマスターデータを指すインデックス（レジストリ）になる。

- **トランザクションハブ**

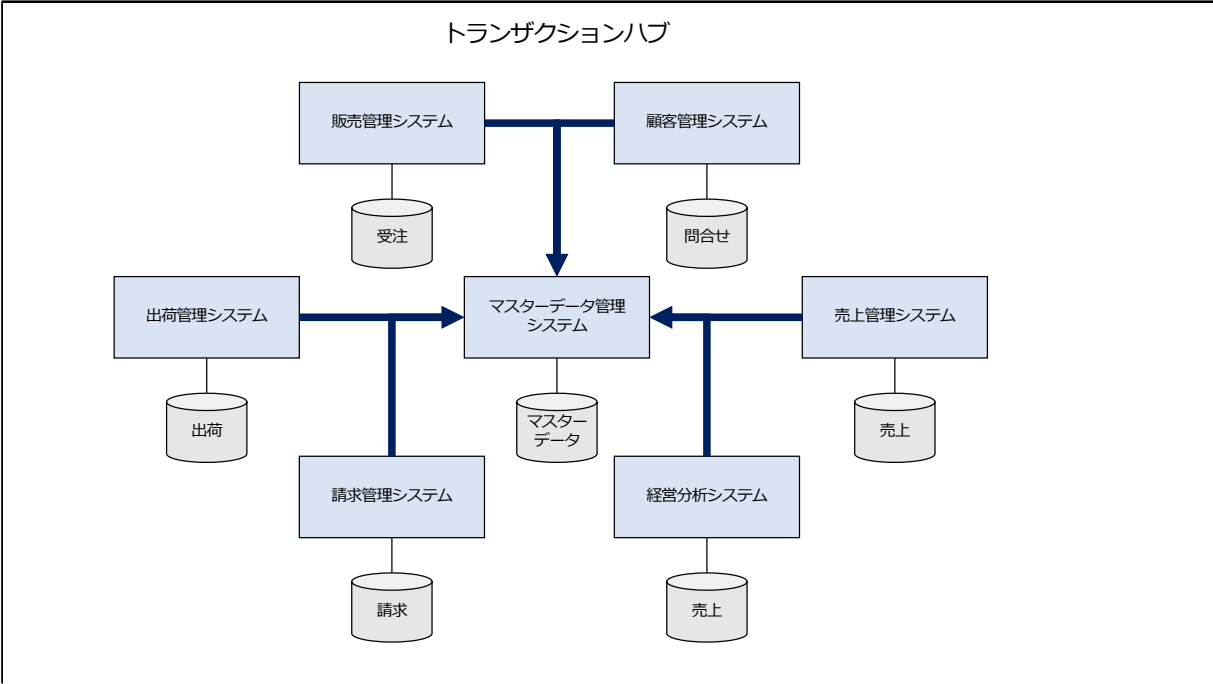
マスターデータ管理システムは、マスターデータのハブとなる。

マスターデータは、マスターデータ管理システムのみに存在し、他のシステムには存在しない。

- 統合アプローチ

レジストリとトランザクションハブのハイブリット。

各システムにあるマスターデータが、マスターデータ管理システムで統合され、他のシステムから利用される。



マスターデータ管理の仕組（アーキテクチャ）

- レジストリ

マスターデータ管理システムは、さまざまなシステムのマスターデータを指すインデックス（レジストリ）になる。

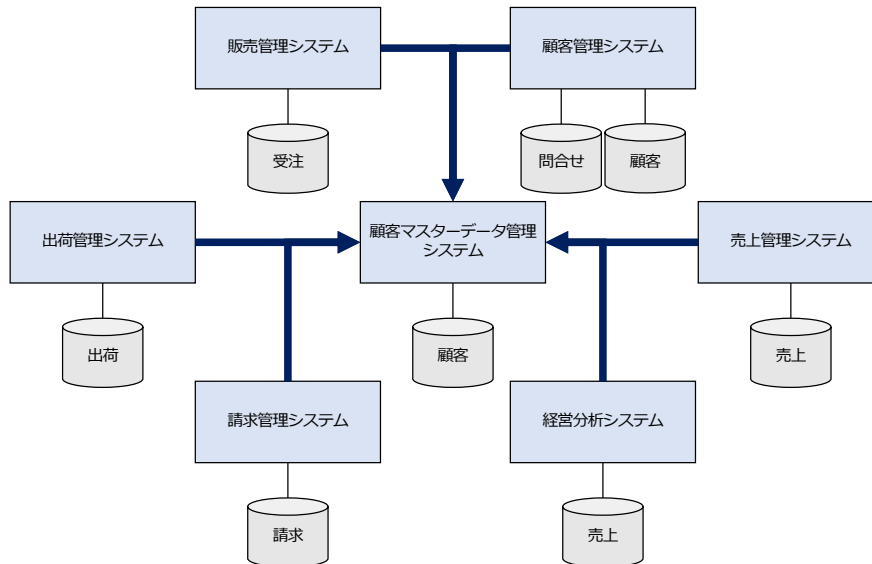
- トランザクションハブ

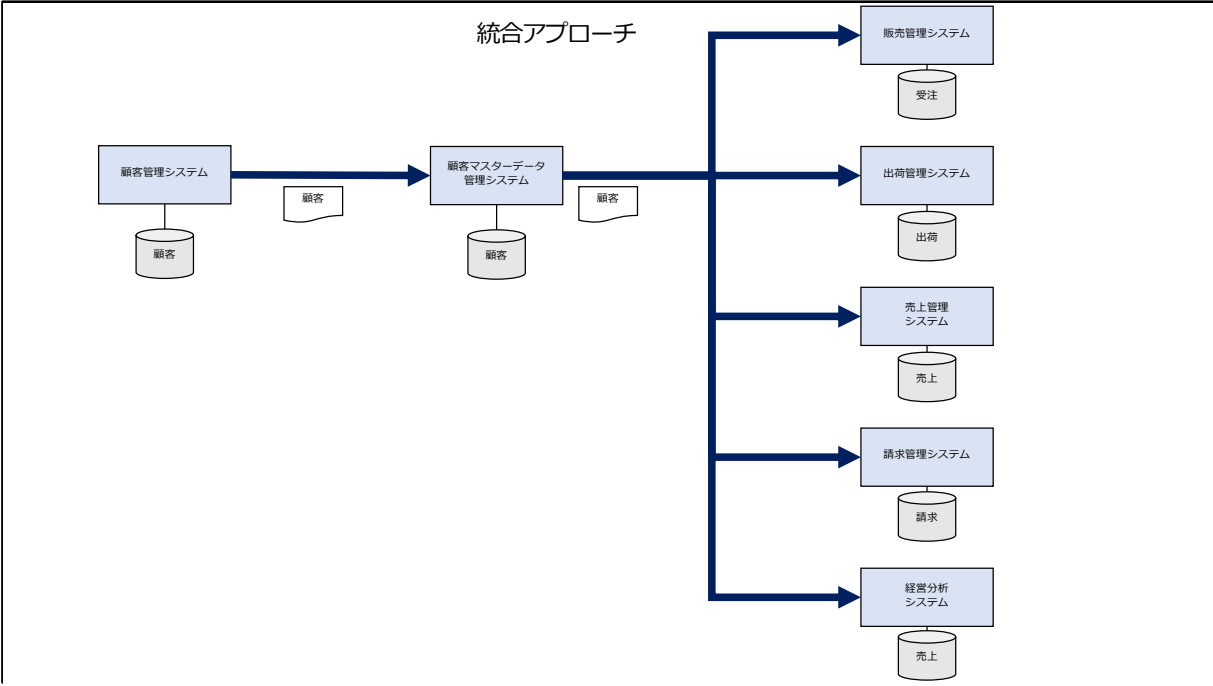
マスターデータ管理システムは、マスターデータのハブとなる。
マスターデータは、マスターデータ管理システムのみに存在し、他のシステムには存在しない。

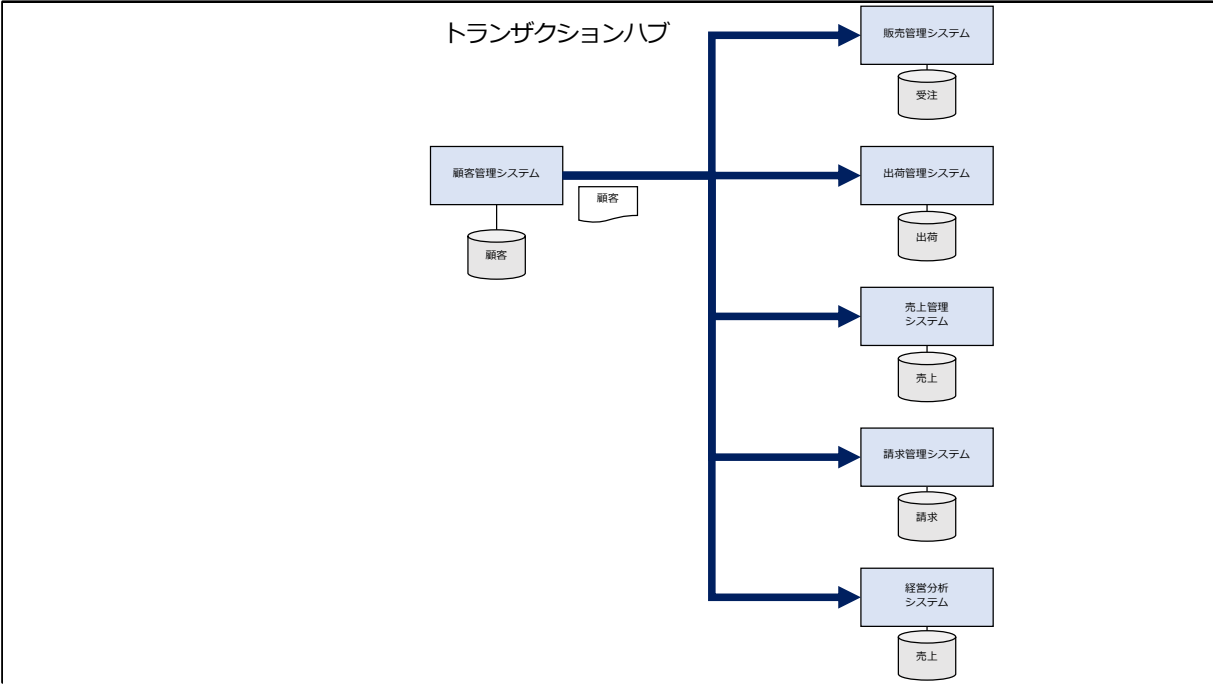
- **統合アプローチ**

レジストリとトランザクションハブのハイブリット。
各システムにあるマスターデータが、マスターデータ管理システムで統合され、他のシステムから利用される。

統合アプローチ



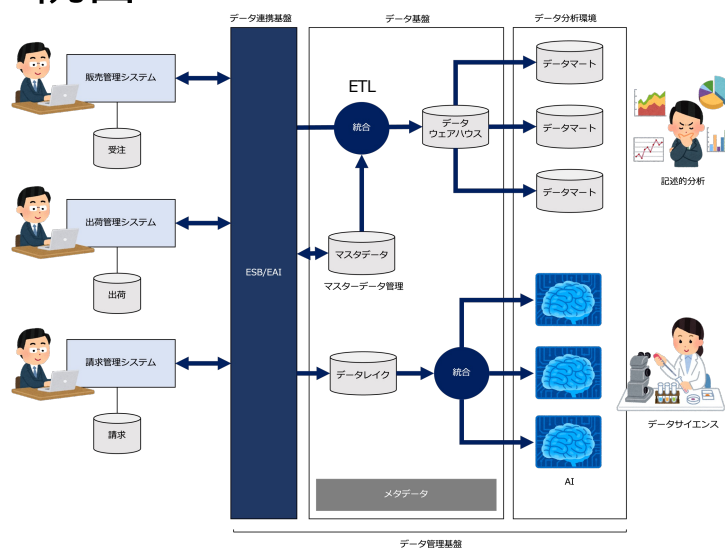




データ管理基盤

- データ管理基盤とは
- リレーショナルデータベース
- データウェアハウス
- データレイク
- NoSQL
- マスターデータ管理
- データ統合
- データ連携
- データ管理基盤の設計

データ統合



変換の例

- **フォーマットの変換**

例えばEBCDICからASCIIなど技術的なフォーマットを変換します。

- **構造の変換**

例えば正規化データを非正規化データにするなどデータの構造を変換します。

- **意味の変換**

例えばソースデータの性別コードの値0,1,2を、ターゲットデータの性別コードの値FEMALE、MALE、NOT PROVIDEDに変えるなど、データの値を変換します。

- **重複排除**

データをユニークにする必要がある場合、重複データを排除します。

- **並べ替え**

データの順序を変更します。

統合の方法

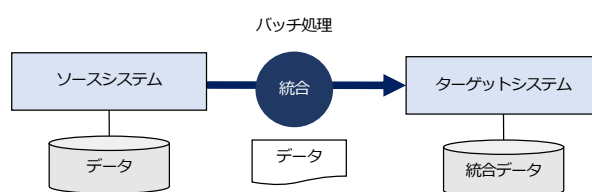
ソースシステムでデータが生成されてからターゲットシステムでデータが利用可能になるまでの時間差をレイテンシといますが、データ統合を考える場合、レイテンシが高い場合、低い場合、非常に低い場合で、統合の方法を分ける必要があります。

通常、レイテンシが高い場合はバッチ、低い場合はイベント駆動、非常に低い場合はリアルタイムという方法を選択します。

- バッチ
- イベント駆動
- リアルタイム

バッチ

データウェアハウスやデータマート上の分析やレポート作成に必要なデータのように、レイテンシが低い場合、定期的に一定量のデータをまとめて処理するバッチ処理がとられます。



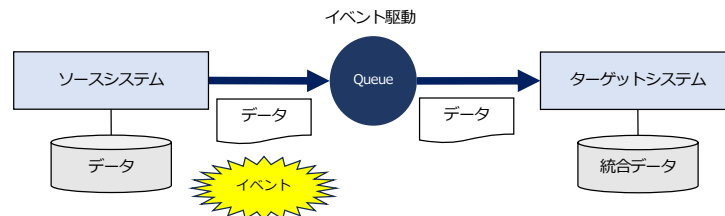
データマネジメント知識体系 第二版 (DMBOK2)

イベント駆動

ソースシステム側に起きたイベントをターゲットシステム側に通知することで、ターゲットシステム側の処理が駆動される形態です。

この場合、ソースシステム側は、ターゲットシステム側の更新を待たずに処理を続行することができます（非同期処理）。

非同期処理によって、ターゲットシステムが利用できない場合でも、ソースシステムが止まったり利用できなくなることはなく、システム間を疎結合にすることができます。

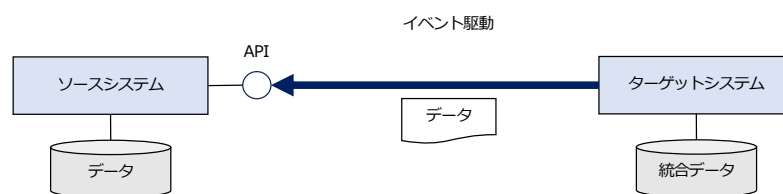


データマネジメント知識体系 第二版 (DMBOK2)

リアルタイム

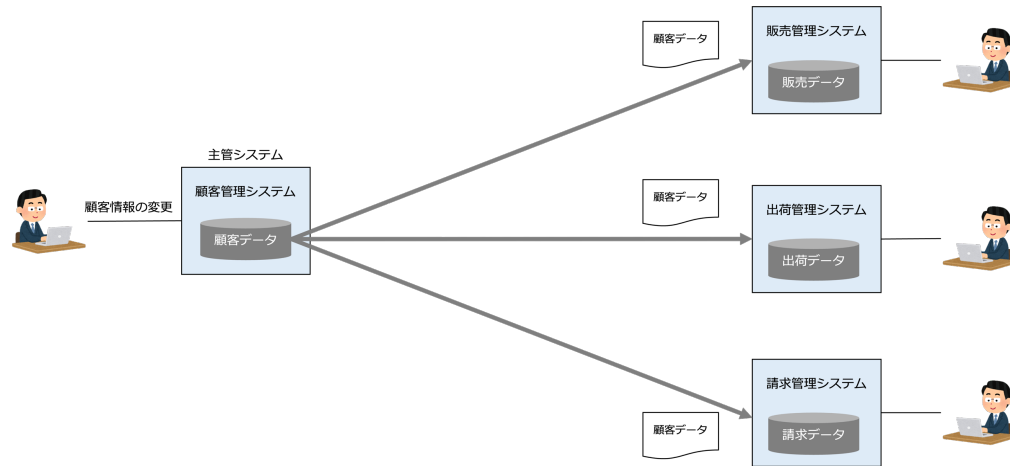
例えば、システム間でトランザクション整合性を保持する必要があるなど、ソースシステムとターゲットシステムのデータ間の時間の遅れや相違が許容されない場合は、リアルタイムにデータを同期させる必要があります（同期処理）。

同期処理の場合、ソースシステムが利用できない場合、ターゲットシステムが利用できる可能性があり、システム間を密結合にします。

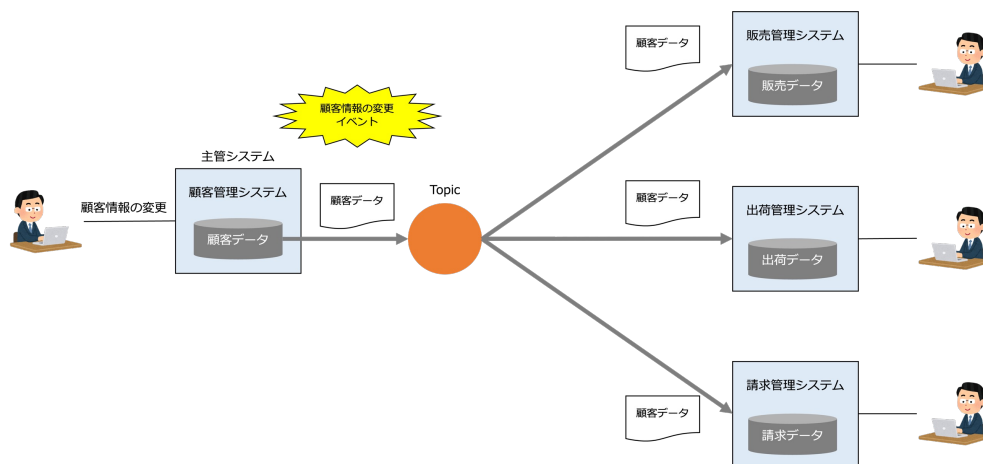


データマネジメント知識体系 第二版 (DMBOK2)

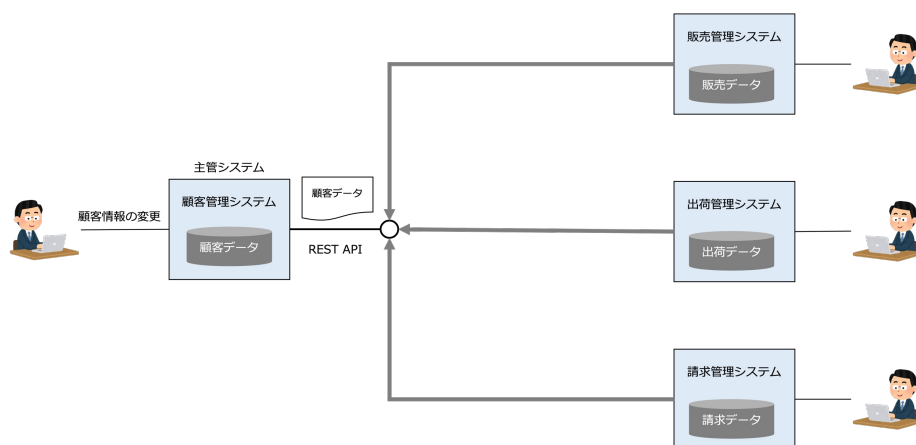
顧客管理システムをトランザクションハブにした顧客データの共有例



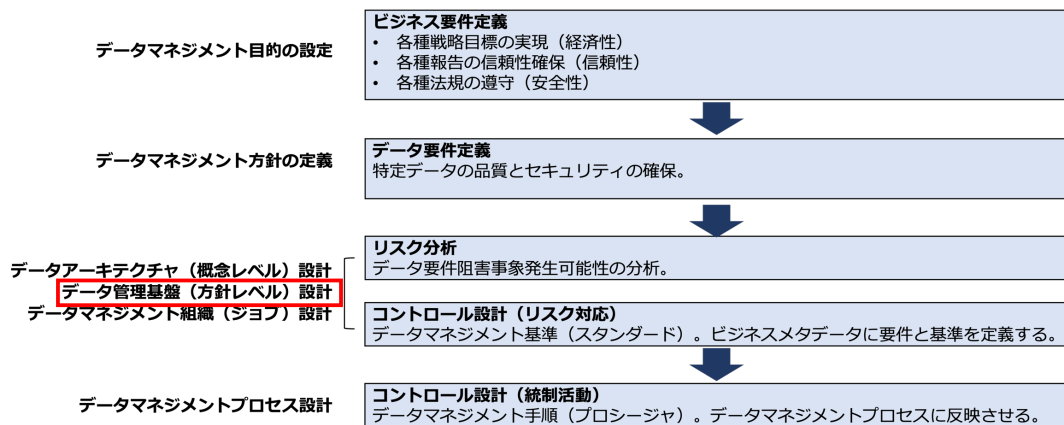
イベント駆動で顧客データを共有する例



リアルタイムで顧客データを共有する例



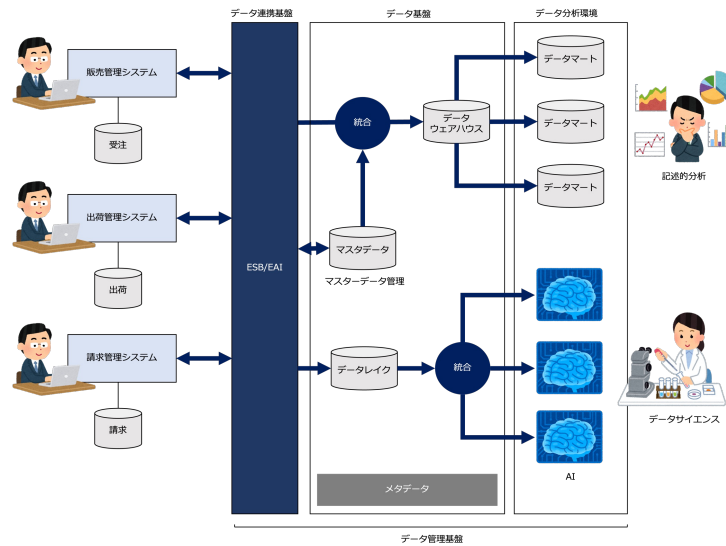
データマネジメントモデル設計プロセス



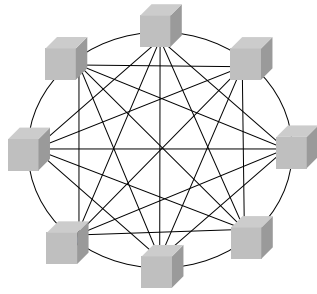
データ管理基盤

- データ管理基盤とは
- リレーショナルデータベース
- データウェアハウス
- データレイク
- NoSQL
- マスターデータ管理
- データ統合
- データ連携
- データ管理基盤の設計

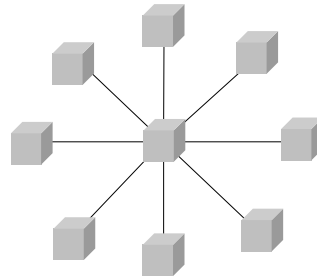
データ連携



データ連携モデル



ポイント・ツー・ポイント



ハブ&スポーク

データ連携アーキテクチャ

- **EAI (Enterprise Application Integration)**

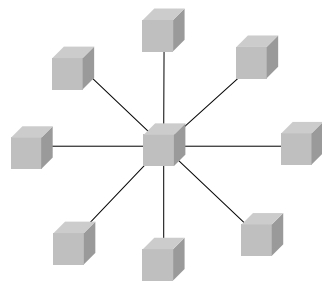
企業内における多種多様なコンピュータシステム群や各種ビジネスパッケージ群を有機的に連携する方式や技術のことです。

- **ESB (Enterprise Service Bus)**

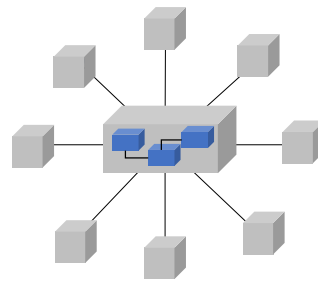
企業内の異なるソフトウェアアプリケーションやサービス間での通信とデータ転送を管理・仲介するための方式や技術のことです。

データマネジメント知識体系 第二版 (DMBOK2)

EAIとESB

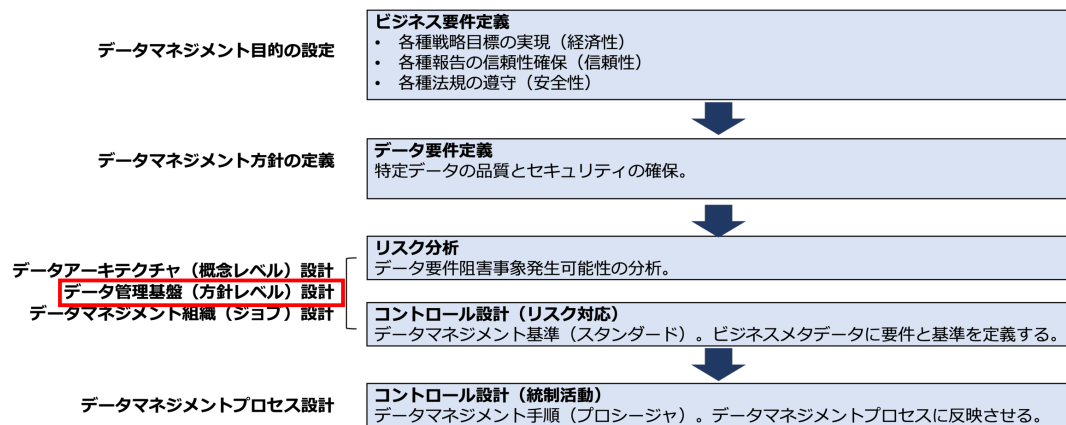


EAI



ESB

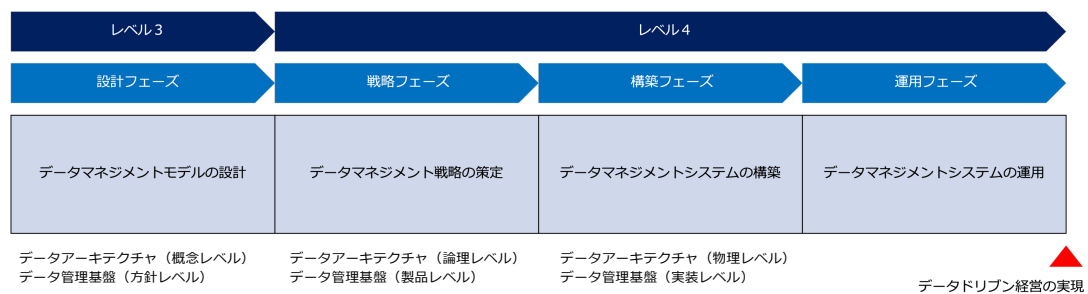
データマネジメントモデル設計プロセス



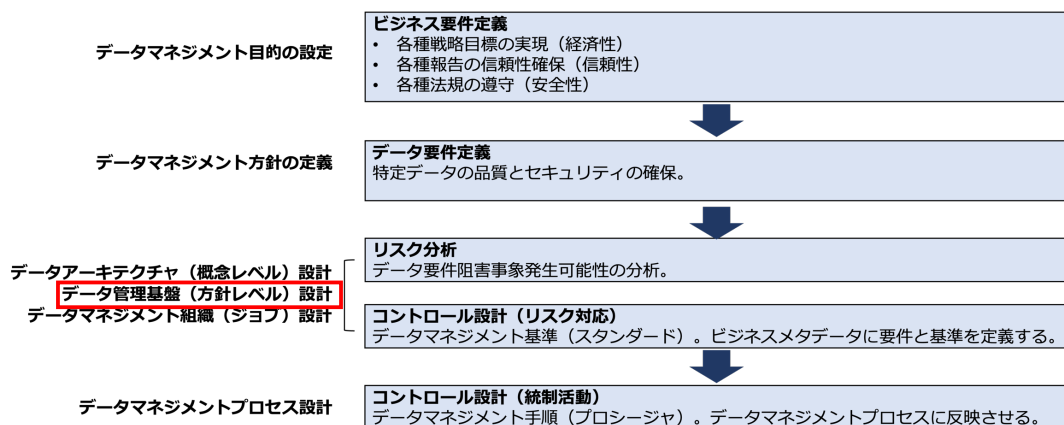
データ管理基盤

- データ管理基盤とは
- リレーショナルデータベース
- データウェアハウス
- データレイク
- NoSQL
- マスターデータ管理
- データ統合
- データ連携
- データ管理基盤の設計

データマネジメント導入プロセス



データマネジメントモデル設計プロセス



データ管理基盤（方針レベル）の設計

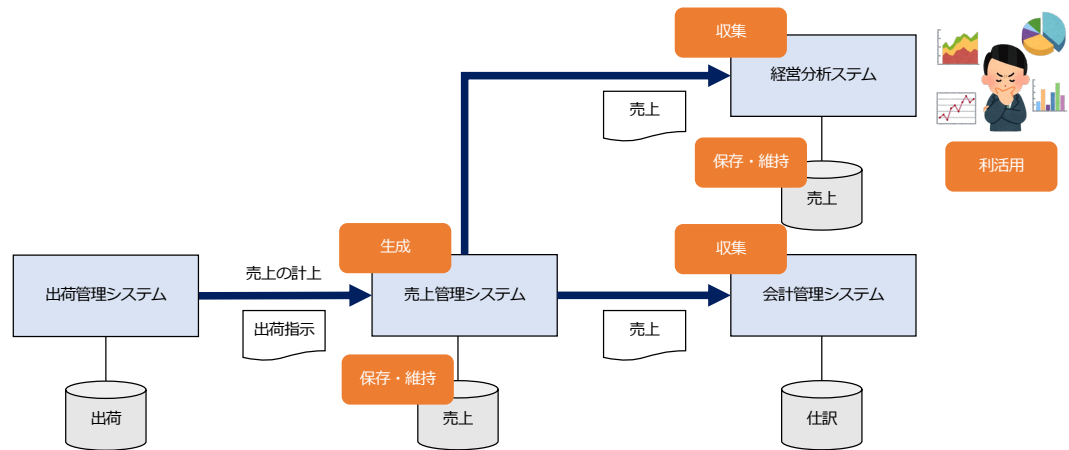
コントロール設計（リスク対応）



データマネジメント基準に準ずるデータ管理基盤の方針を設計する

- データ基盤の方針設計
 - マスターデータ管理の方針設計
 - データウェアハウスの方針設計
 - データレイクの方針設計
 - NoSQLの方針設計
 - データ統合の方針設計
- データ連携基盤の方針設計
- データ分析環境の方針設計

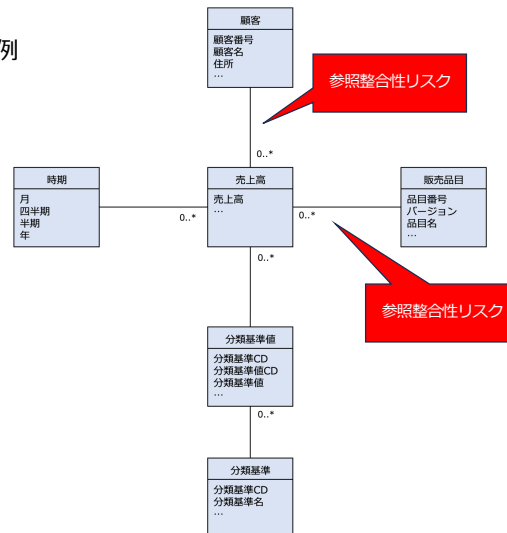
売上データのライフサイクル



売上データの全体概念データフローの例

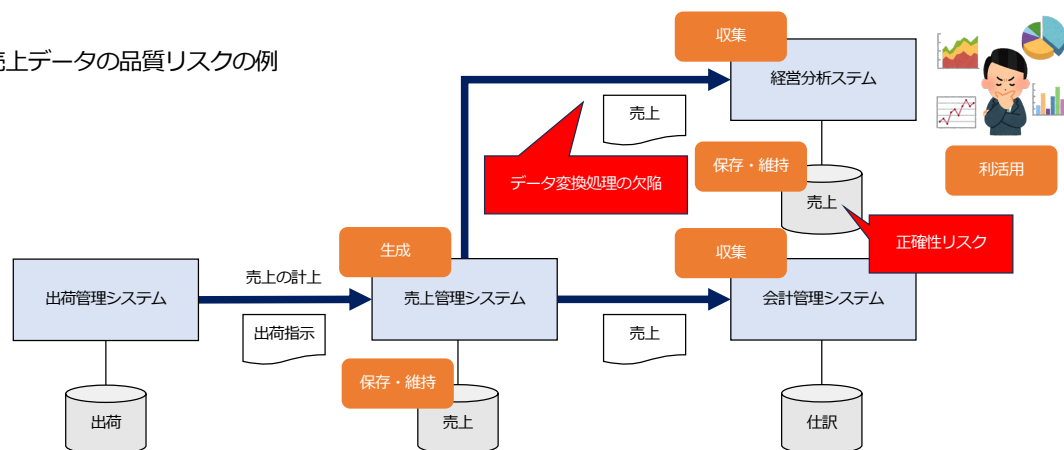
データ品質のリスク分析

売上データの品質リスクの例



データ品質のリスク分析

売上データの品質リスクの例



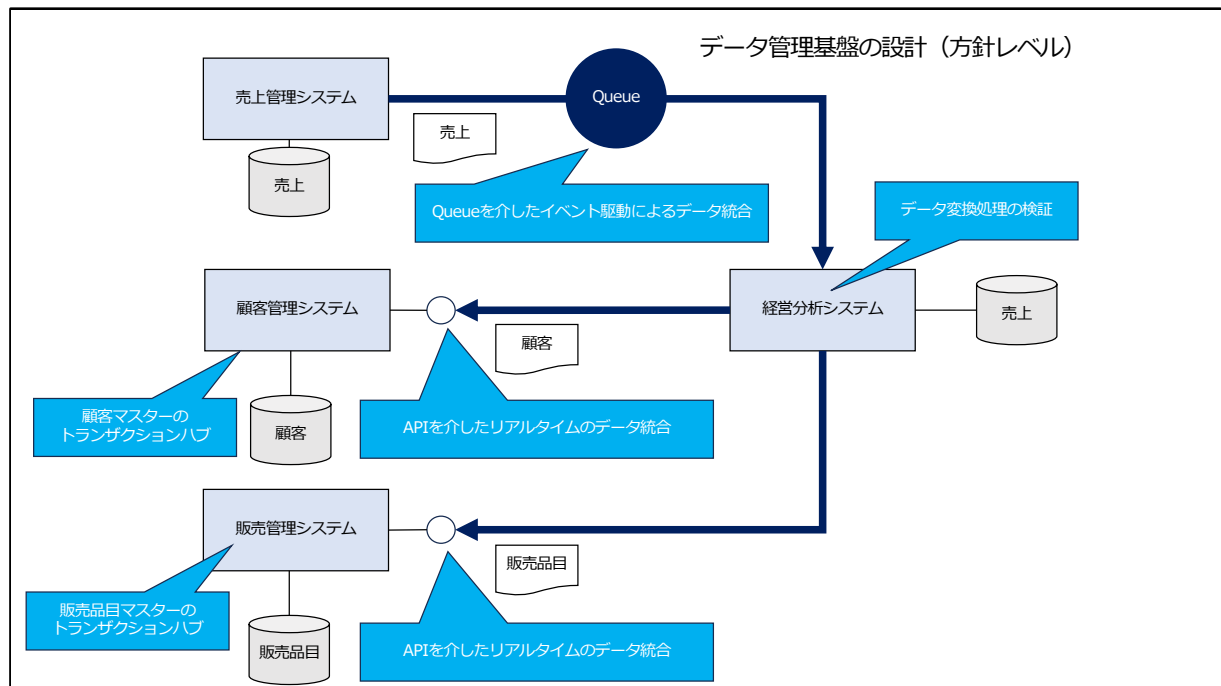
売上データの全体概念データフローの例

データ品質のリスク対応

売上データの品質リスク対応の例
ビジネスメタデータの定義

エンティティ番号	エンティティ名	説明	経済価値	システム	データリネージュ
E001	売上	売上のデータ。	戦略目標XXの価値を測るために必要なデータ。	経営分析システム	売上管理システムの売上データ、顧客管理システムの顧客データ、販売管理システムの販売品目データを売上の多次元分析用に変換。

エンティティ番号	データ品質要件				
	データ品質評価指標	ビジネスルール	品質測定方法	品質評価基準	品質評価結果
E001	参照整合性	売上は顧客の顧客番号の値と同じ外部キーを持つこと（参照整合性制約）。	条件を満たすレコードの数をレコードの総数で除算し100を掛けてパーセントで表す。	許容できない結果 ルールの遵守が100%未満。	—
	参照整合性	売上は販売品目の品目番号の値と同じ外部キーを持つこと（参照整合性制約）。	条件を満たすレコードの数をレコードの総数で除算し100を掛けてパーセントで表す。	許容できない結果 ルールの遵守が100%未満。	—
	正確性	売上管理システムの売上データ、顧客管理システムの顧客データ、販売管理システムの販売品目データが正しく売上の多次元分析ができる形に変換されていること。	条件を満たすレコードの数をレコードの総数で除算し100を掛けてパーセントで表す。	許容できない結果 ルールの遵守が100%未満。	—

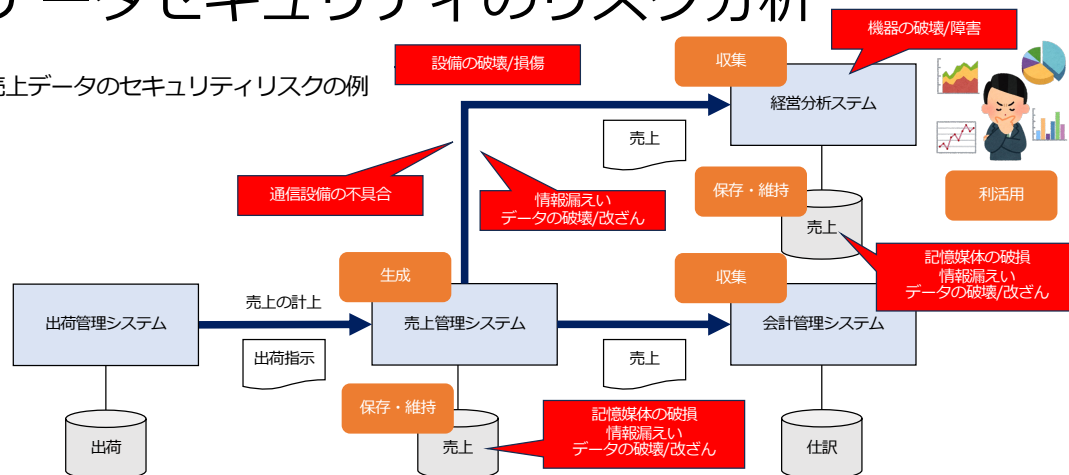


データ管理基盤の設計（方針レベル）

- **顧客管理システムを顧客マスターのトランザクションハブにする**
顧客データを一元管理することで顧客データの一意性、完全性を上げることができる。
よって、売上と顧客の間の参照整合性リスクを低減することができる。
- **販売管理システムを販売品目マスターのトランザクションハブにする**
販売品目データを一元管理することで販売品目データの一意性、完全性を上げることができる。
よって、売上と販売品目の間の参照整合性リスクを低減することができる。
- **売上データを統合するときデータ変換処理を検証する**
データ変換処理の欠陥リスクを低減することができる。

データセキュリティのリスク分析

売上データのセキュリティリスクの例



売上データの全体概念データフローの例

データセキュリティのリスク対応

売上データのセキュリティリスク対応の例
ビジネスメタデータの定義

エンティティ番号	エンティティ名	説明	経済価値	システム
E001	売上	売上のデータ。	戦略目標XXの価値を測るために必要なデータ。	売上管理システム

エンティティ番号	データセキュリティ要件				
	機密レベル	規制対象カテゴリ	セキュリティリスク	リスク対応	セキュリティ監査結果
E001	制限付機密	財務上のセンシティブデータ	情報の漏洩	認証、認可、データの分離、暗号化、マスキング	—
			データの破壊/改ざん	認証、認可、データの分離、暗号化	—
			記憶媒体の破損	バックアップと復旧	—
			通信設備の不具合	通信設備の冗長化	—
			機器の破壊/障害	機器の冗長化	—
			設備の破壊/損傷	緊急事態計画の策定と実施	—

データ管理基盤の設計・演習

データセキュリティのリスク分析の結果を見ると、経営分析システムが保管する売上データの情報漏洩、データの破壊/改ざんリスクがあります。

また、そのリスクに対して、データを分離するという対応基準があります。

売上データの情報漏洩、データの破壊/改ざんリスクを低減するためにデータ管理基盤をどう設計すればよいか考えてください。

なお、経営分析システムが保管する売上データはデータウェアハウスとして管理されています。

データ管理基盤の設計・演習・解答例

データセキュリティのリスク分析の結果を見ると、経営分析システムが保管する売上データの情報漏洩、データの破壊/改ざんリスクがあります。

また、そのリスクに対して、データを分離するという対応基準があります。

売上データの情報漏洩、データの破壊/改ざんリスクを低減するためにデータ管理基盤をどう設計すればよいか考えてください。

なお、経営分析システムが保管する売上データはデータウェアハウスとして管理されています。

データ分析環境としてデータマートを設け、売上データを参照できる役割や部門のみ多次元分析ができるようにする。