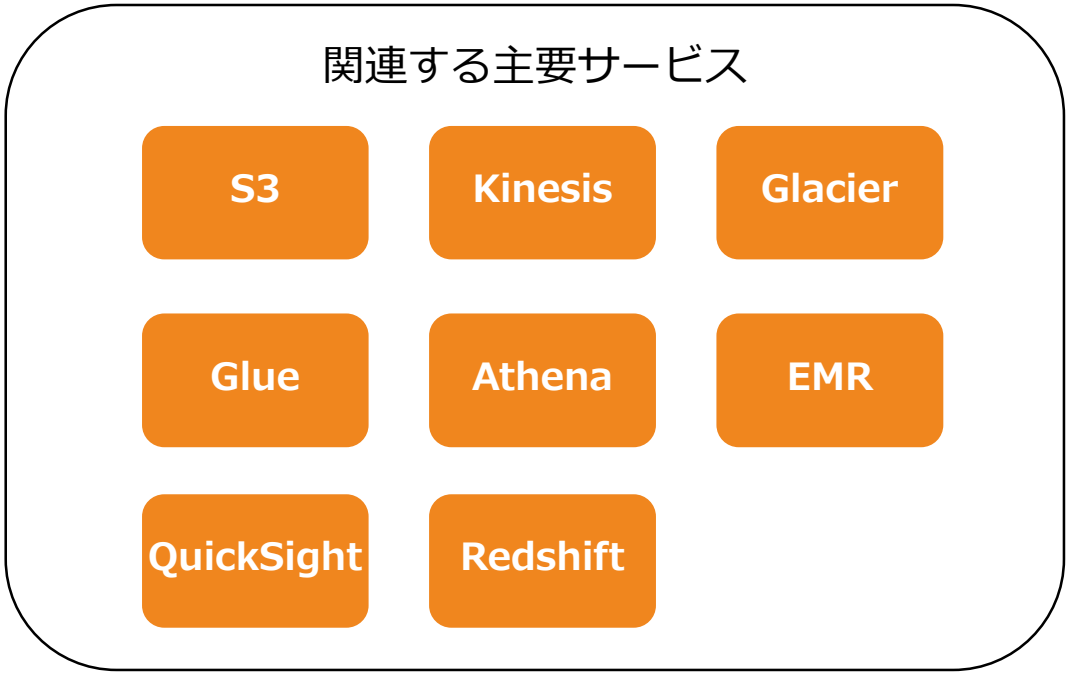
増大するデータ量
への対応

⑦増大するデータ量対応

IoT／ビッグデータなどで絶えず増加するデータの保持を効率的に実施する



データ量への対応

効率的なデータ蓄積とIoTなどの大量のストリームデータ処理や解析の方法が必要不可欠となる

ビッグデータに必要な技術

ビッグデータに対応したデータ蓄積・処理技術が必要不可欠

Volume
大量データ

✓ 大量のデータを効率的に蓄積可能なデータベース技術

Variety
多様なデータ

✓ 多様な形式のデータを蓄積可能なデータベース技術

Velocity
速い処理

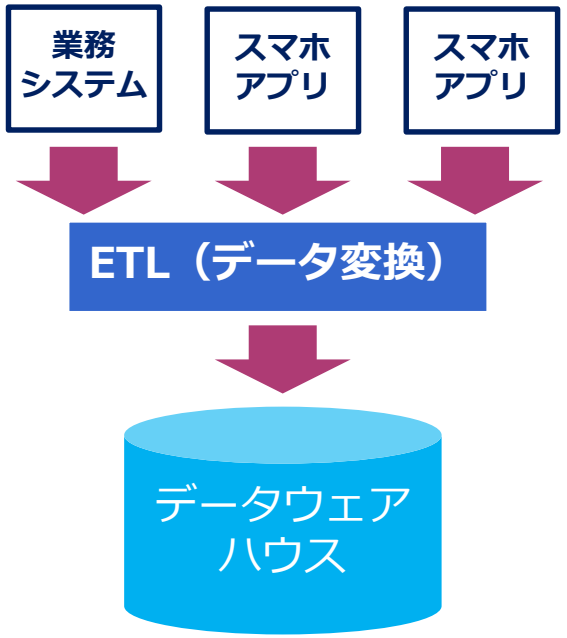
✓ 高速処理が可能なデータ処理ソフトウェア／ハードウェア

データレイクの活用

ビッグデータ活用の中心はデータレイク型データベース

データウェアハウス中心

利用用途に応じたデータを貯めて活用するデータウェアハウス



データレイク中心

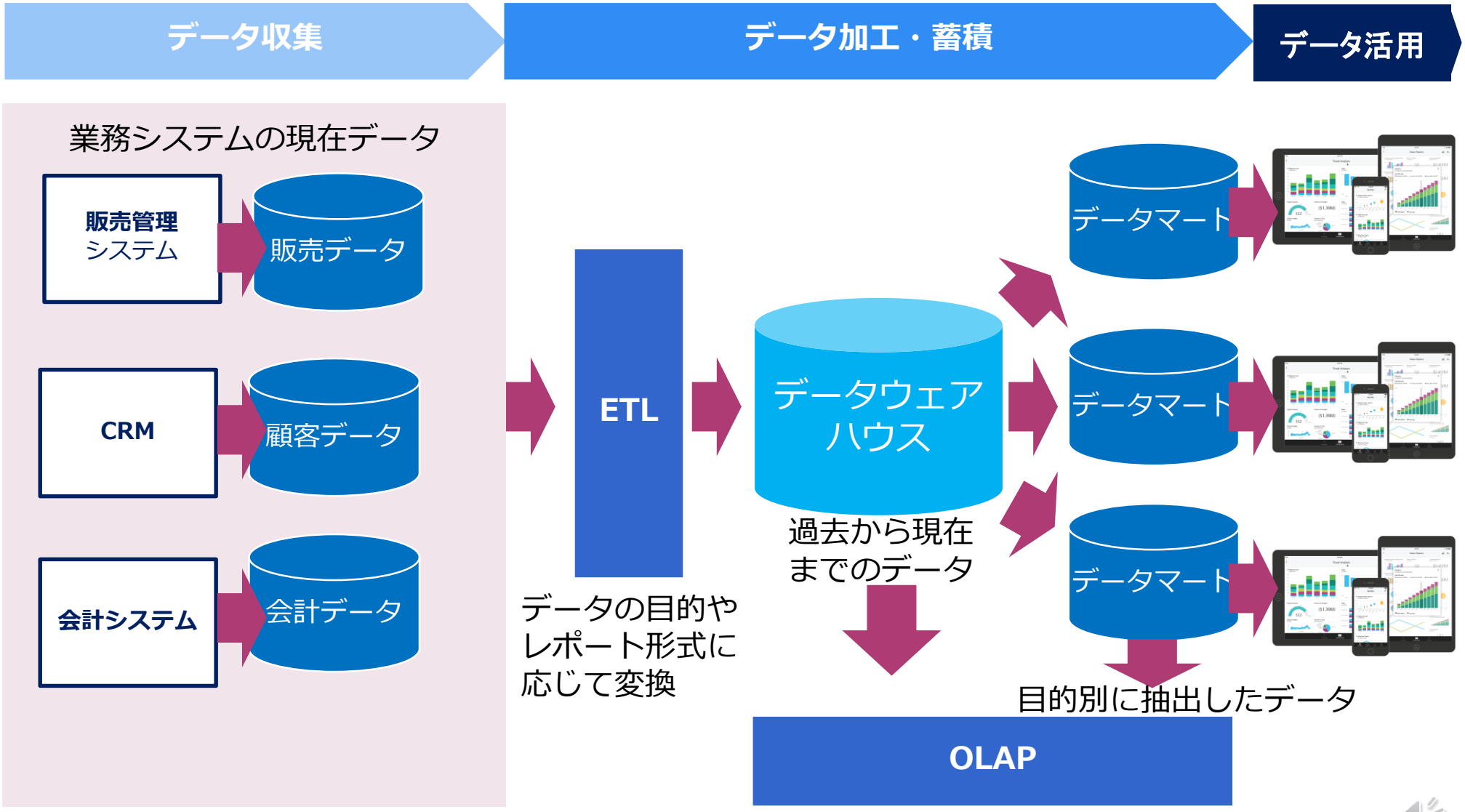
出来る限り生データをほぼ全データ保存するデータレイク

データレイクの活用

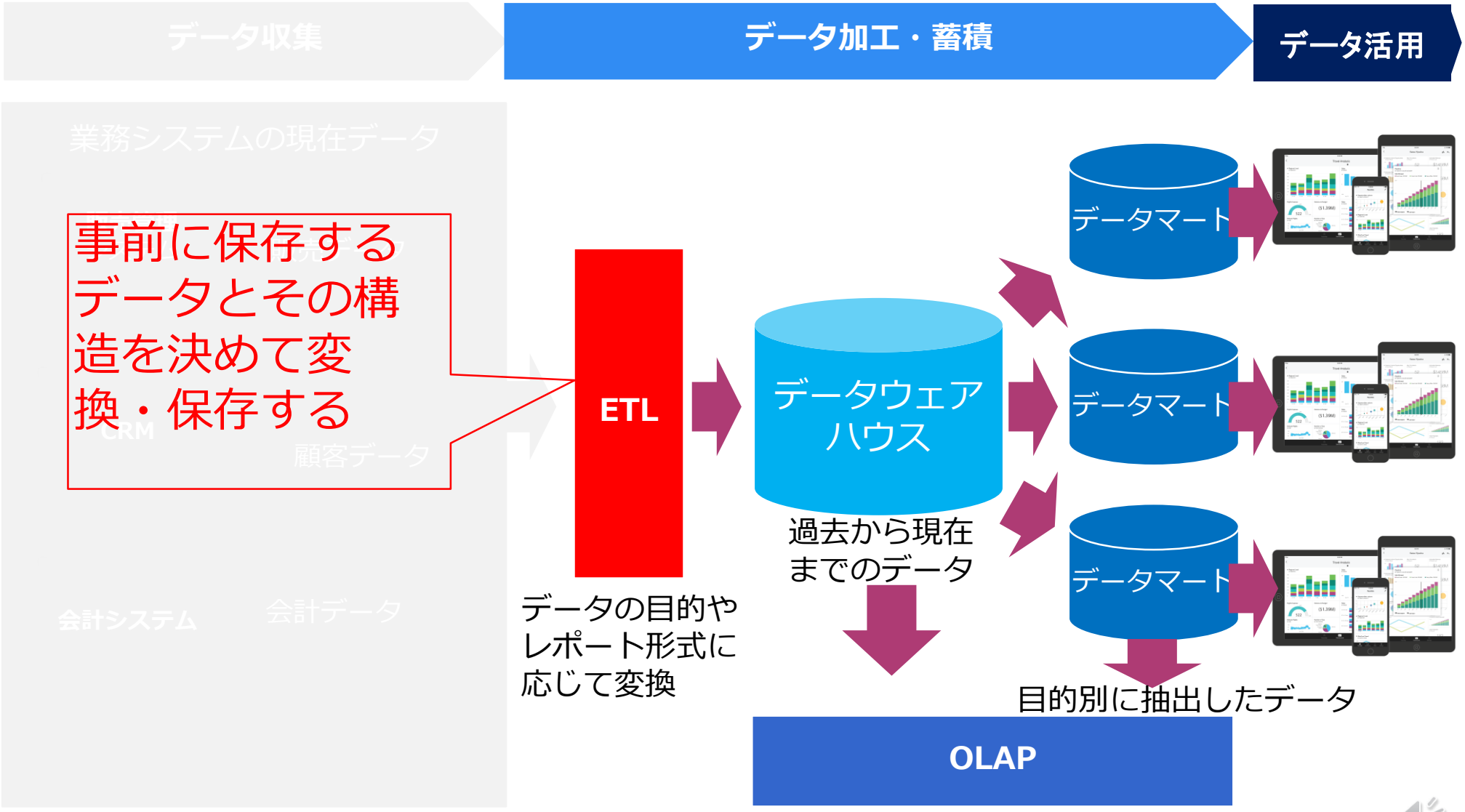
データレイクでは全データを生データのまま保存する

	データウェアハウス中心	データレイク中心
データ収集	✓ 目的別データ ⇒必要なデータのみを抽出・収集 ✓ 構造化データ中心	✓ 生データ＋目的別データ ✓ 構造化／半構造化／非構造化データ
蓄積	✓ 必要なデータのみを抽出・蓄積	✓ 変換しないで生データ形式で保存 ✓ もしくはエッジ処理したデータを保存
処理･加工	✓ 関連するデータ構造（スキーマ）に変換・蓄積 ✓ SQLによる操作	✓ 事前にスキーマ（データ構造）を定義しない ✓ SQL/SAS/MapReduce/R/NoSQLなどで操作
可視化 分析	✓ 利用者がデータ分析／レポート内容などの利用目的を事前に特定し構築	✓ 事前に目的を定義せず、ユーザーがデータ群から新たな価値を抽出しデータを解釈・活用

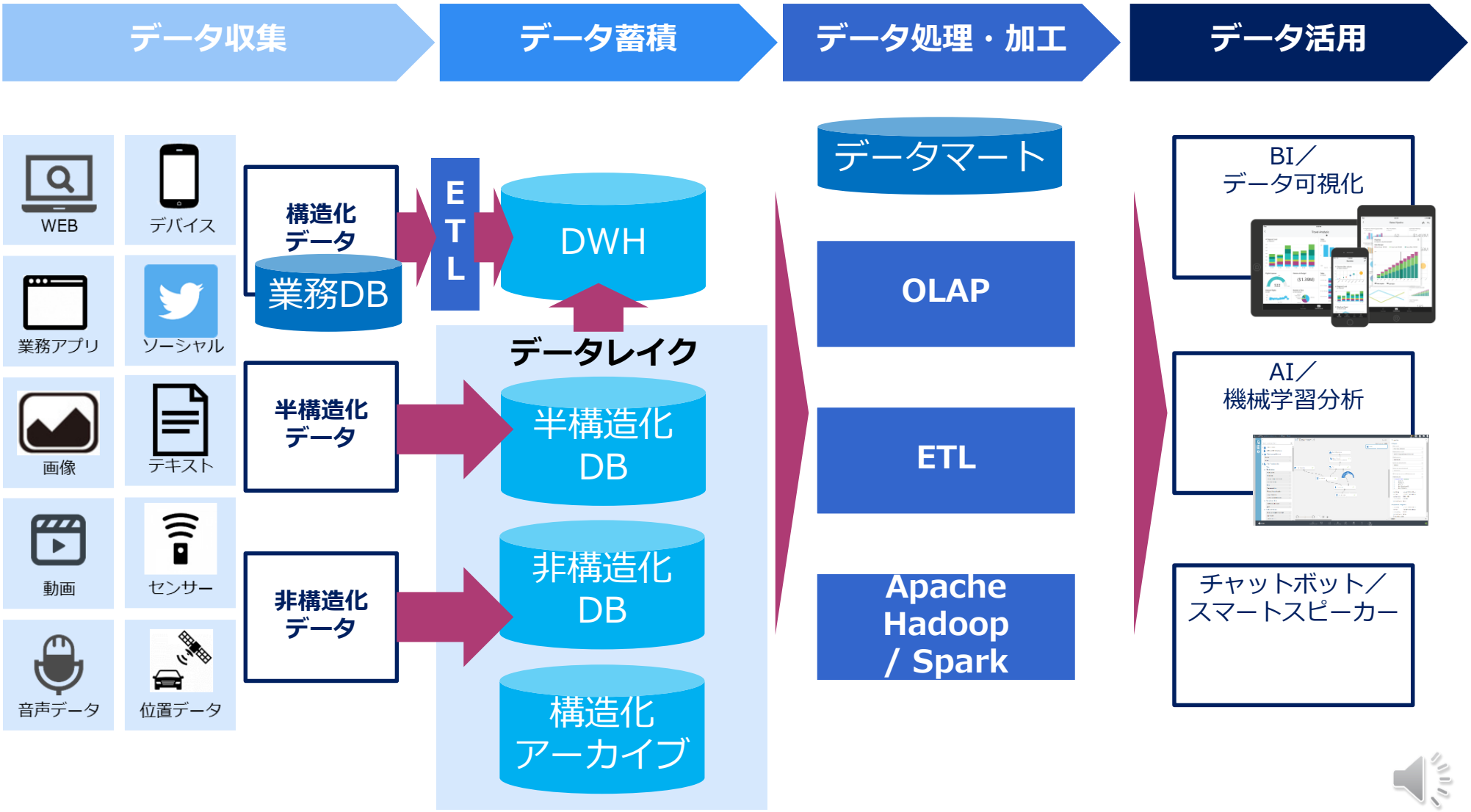
データウェアハウス型のデータ処理基盤



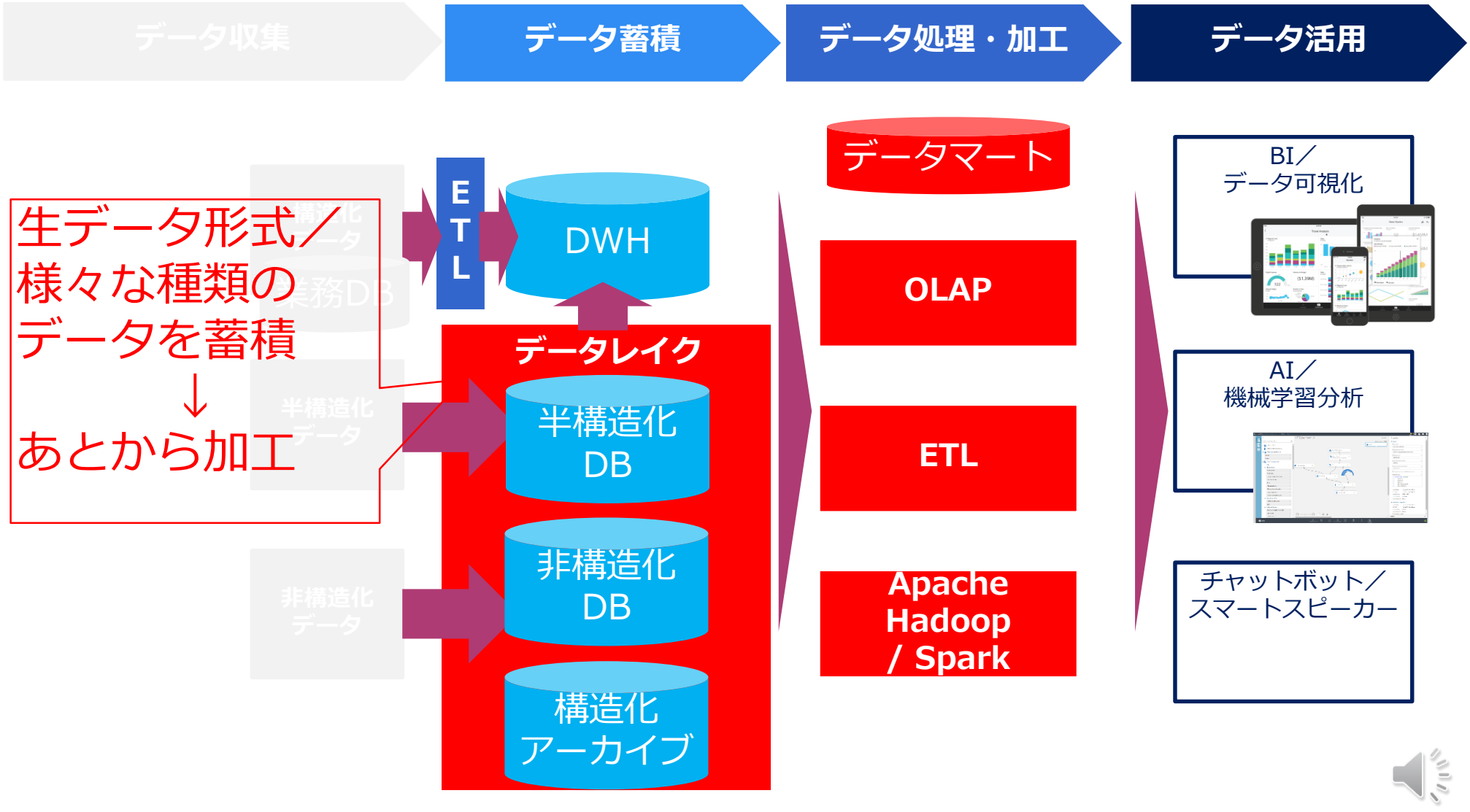
データウェアハウス型のデータ処理基盤



データレイク型のデータ処理基盤



データレイク型のデータ処理基盤



Apacheシリーズ

ビッグデータ分散処理向けの代表的な仕組み（ミドルウェア）がApacheシリーズ

大量データバッチ処理向け

Apache Hadoop

ストリーミング処理向け

Apache Spark

AWSのデータレイク構成