



大規模データ分析の新潮流 ～In-Database Analyticsが実現する 圧倒的なデータ分析パフォーマンス～

2015/11/20
株式会社NTTデータ
技術開発本部サービスイノベーションセンター
データウェアハウス／ビジネスインテリジェンス・ラボ®

NTT DATA

1. NTTデータのビッグデータに対する取り組みについて
2. In-Database Analytics開発



1. NTTデータのビッグデータに対する取り組みについて

お客様

分析
コンサルティング

ソリューション
提供

システム開発

R&D

NTTデータ数理システム

- ✓ Original Data Analysis Packages
- ✓ > 80 Data Scientists
- ✓ > 200 Analysis Use Cases

NTT DATA

NTT DATA Mathematical Systems Inc.

ビッグデータビジネス推進室

- ✓ Cloud Based Solutions : BizXaaS BA
- ✓ Telco., Insurance, Banking, Retail, Manufacture...

技術開発本部

- ✓ As COE of Business Intelligence in NTT DATA Group
- ✓ Evaluation of State-of-the-art
- ✓ Promotion of technical development and demonstration

NTTデータのビッグデータ活用
=「ビジネスインテリジェンス」および「ビジネスアナリティクス」を
ベースとしたアプローチ

1. NTTデータのビッグデータに対する取り組みについて ビッグデータのとらえ方



*2020年時点のGartner等による推定値

お客様のシステムへの期待の変化

従来

現在

分析目的

見える化

予測
(意思決定)

分析対象

全体傾向

個人単位

分析処理

バッチ

リアルタイム

アルゴリズム/システムの
ブラッシュアップ主体

SIベンダー
(受託)

システム自身
(自動学習)

プロアクティブ型BI

コンテキスト・アウェアネス型

予兆発見型

異常検出型

外れ値検出型

不正検出型

WHAT-IF型BI

予測・制御型

リスク・シミュレーション型

収益シミュレーション型

リスクヘッジ型

最適化型

マーチャンダイジング型

集計分析型BI

評価・要因分析型

プロセス・トレース型

発見型BI

ターゲティング型

与信管理型

プロアクティブ型BI

コンテキスト・アウェアネス型

予兆発見型

WHAT-IF型BI

予測・制御型

リスク・シミュレーション型

収益シミュレーション型



BICLAVIS[®]

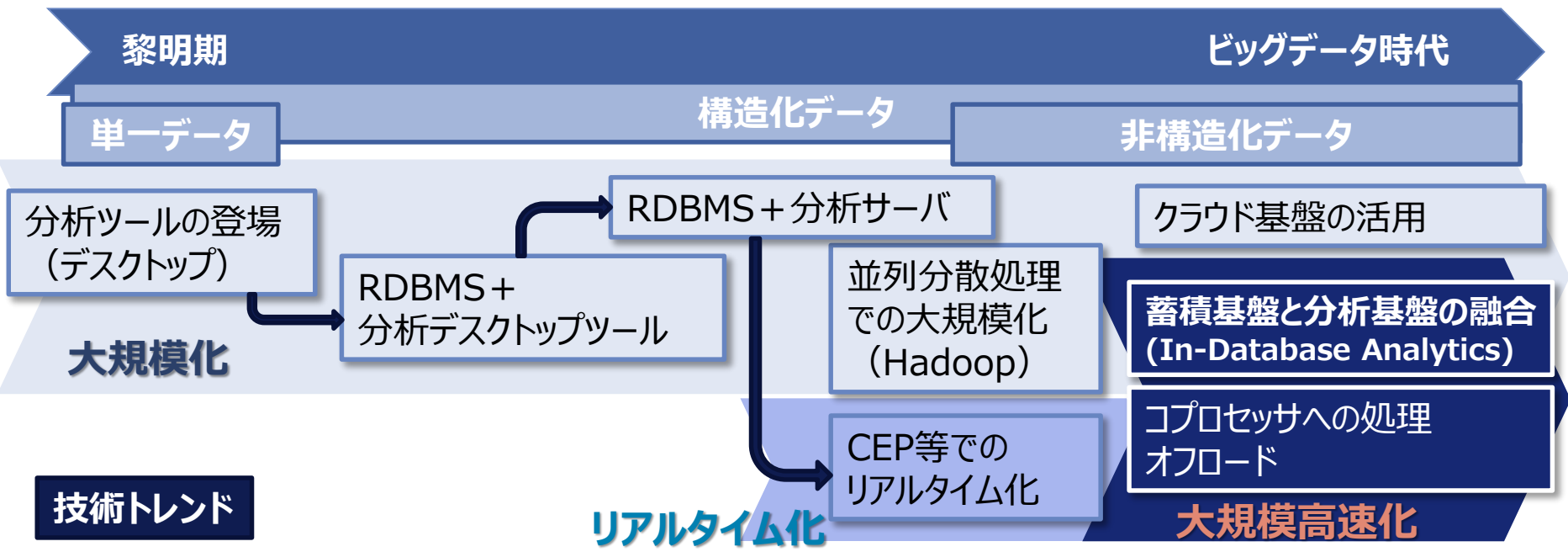
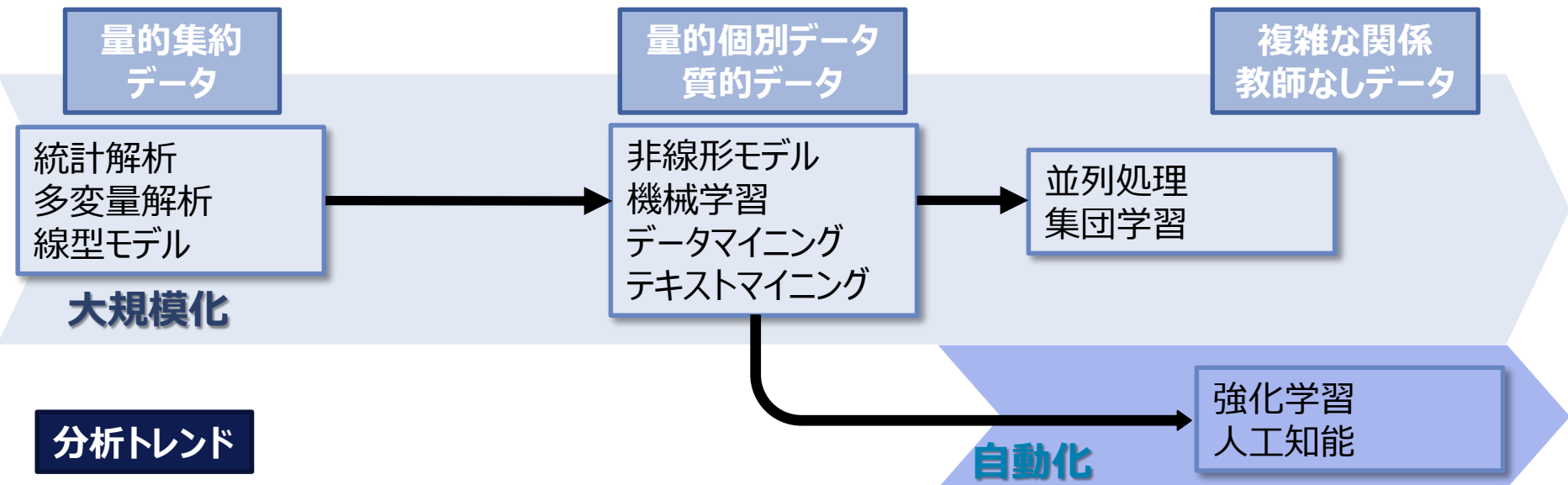
集計分析型BI

発見型BI

目的に応じたシナリオを組み合わせることで
ビッグデータ活用を高度化する

分析シナリオ類型		概要
① 予兆発見型		行動変化や状態変化の監視による予兆の発見
② 異常検出型	不正検出型	不正・異常の定義と合致／類似する行動・状態の検出
	外れ値検出型	標準的な行動・状態の定義と逸脱の検出
③ 予測・制御型	収益シミュレーション型	業務改善による増収効果の試算
	リスク・シミュレーション型	業務のモデル化と不確実要素によるリスクの試算
	リスク・ヘッジ型	業務のモデル化と最適化手法を用いた意思決定策の提示
	最適化型	業務のモデル化とリスク分散手法を用いたリスク低減策の提示
④ ターゲティング型		見込み顧客など重点アプローチすべきターゲットの抽出
⑤ 与信管理型		顧客・企業の滞納・倒産リスクの試算
⑥ 評価・要因分析型		さまざまな対象の比較評価と改善要因の特定
⑦ マーチャンダイジング型		さまざまな視点での売れ筋ランクの作成と品揃えの決定
⑧ コンテキスト・アウェアネス型		行動履歴・嗜好の分析から一步先回りしたサービスの提示
⑨ プロセス・トレース型		成長・発展プロセスの抽出と促進・阻害の特定

1. NTTデータのビッグデータに対する取り組みについて データ分析および技術の変遷





2. In-Database Analytics開発

データ分析に関する要素技術は出そろってきている感はあるが、 ビッグデータ分析プラットフォームは発展途上状態

状況1

(Volume : 量の問題)

- 大量データをNWを介してDWHから分析処理に受け渡すのは非現実的

状況2

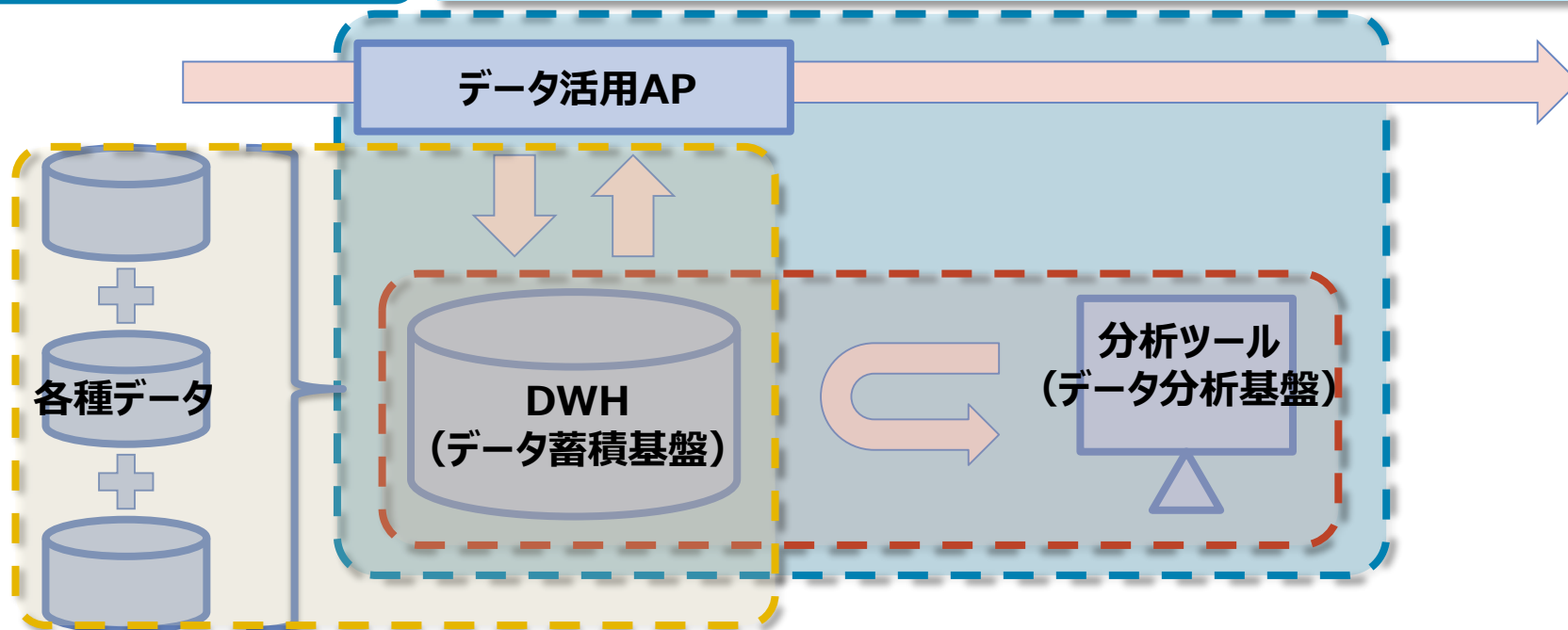
(Variety : 多様性の問題)

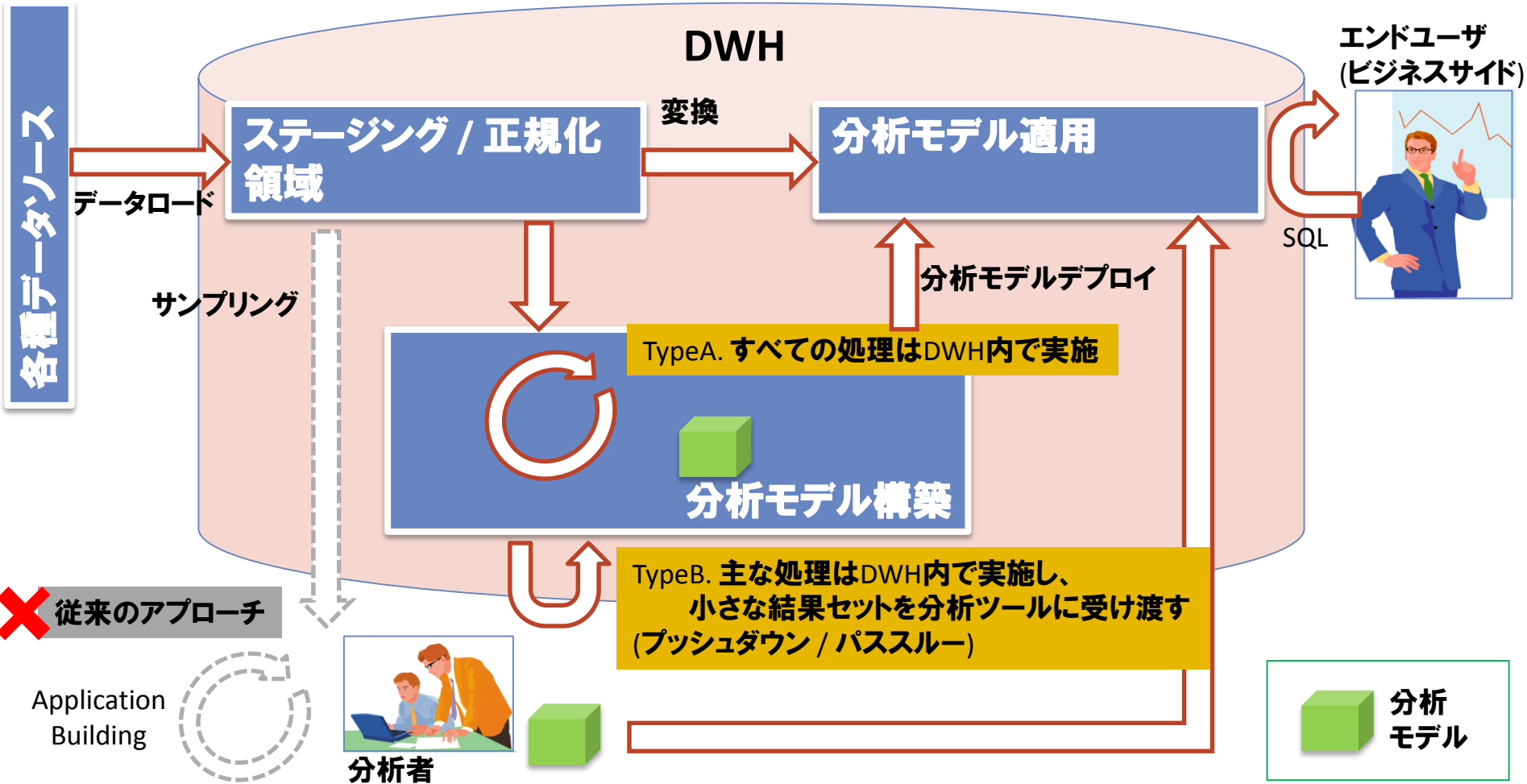
- 企業横断的な複合データで分析を行うため共通DB (統合DWH) が必要

状況3

(Velocity : 速度の問題)

- 発生し続けるデータを即座に分析する必要がある
- ビジネススピードの速度に合わせて分析モデルを更新する必要がある





In-Database Analytics = DWHからのデータ抽出を行うことなく、
高速に大量データに対するデータ分析の実現

課題

To-Be (ビッグデータ対応)

目的

状況1
(Volume)

統合DWHにてデータ分析処理を
実施することによる
Single Version of Truthの実現

データの
コピーを
減らす

データ
ガバナンスの
確立

状況2
(Variety)

大量のデータに対して
統合蓄積基盤 (DWH) 内で
高速にデータ分析処理を実現

データの
移動を
減らす

分析チームの
生産性向上

分析スコアリングの
パフォーマンス向上

状況3
(Velocity)

構築された分析モデル (数式) を
スコアリングAP (Java、SQL等) に
組み替える手間を省く

QCDの
確保

分析モデルの
ポータビリティ向上

ビッグデータに対する高速なデータ分析処理の実現
= 蓄積基盤 (DWH) 上での並列分散処理

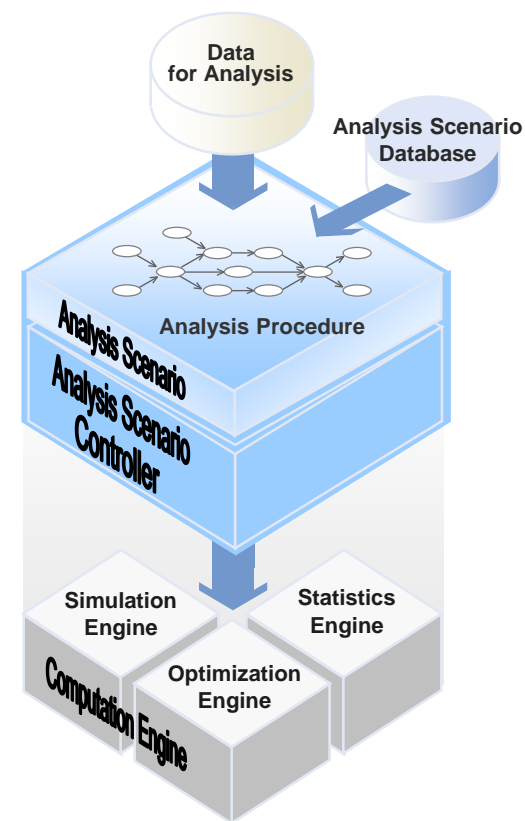
ビッグデータ時代の分析処理（大量、高速）を実行するためのコアエンジン

優位性

- NTTデータ独自の分析方法論「**BICLAVIS®**」に対応したアルゴリズムの搭載
→分析目的が明確になれば分析シナリオが決まり、分析アルゴリズムも定まる
- **NTTデータ数理システム**のノウハウや資産を活かした短期・高品質の開発
→分析精度の確保、大量データのハンドリング、計算量低減の工夫

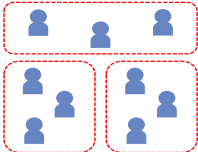
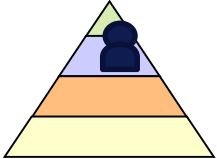
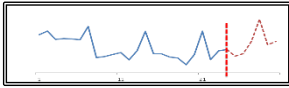
チャレンジ

- **マルチプラットフォーム対応**
- 特定の製品の非依存
- **並列分散処理への対応**
→多くの分析ツールはシングルスレッド処理
- **モデルの適用をリアルタイムに**
→CEPへの分析モデル組み込み



BICLAVIS®

BICLAVIS シナリオ	説明
④ターゲティング型	ターゲット層発見の起点となるセグメント分割 ターゲット顧客発見のためのスコアリング
③予測・制御型	過去時系列から将来推移の予測
⑧コンテキスト・ アウェアネス型	商品などのアイテムレコメンド

<p>[セグメント分割]</p> <ul style="list-style-type: none"> ● K-Means ● Matrix Decomposition etc. 	<p>[スコアリング]</p> <ul style="list-style-type: none"> ● Random Forest ● Neural Network etc. 	<p>[需要予測]</p> <ul style="list-style-type: none"> ● Linear Regression ● ARIMA etc. 	<p>[レコメンデーション]</p> <ul style="list-style-type: none"> ● Collaborative Filtering ● Matrix Decomposition etc.
---	--	---	--

赤字：FY2014実装

処理のタイプ^o

- 1つの処理を細かなサブタスクに分割し並列処理できるアルゴリズム

- 1つの処理はそれ以上細分化は出来ないが、同様の処理を異なる入力データ・パラメータで同時に実行させ、出力結果をマージする

アルゴリズム例

- K-Means
etc.

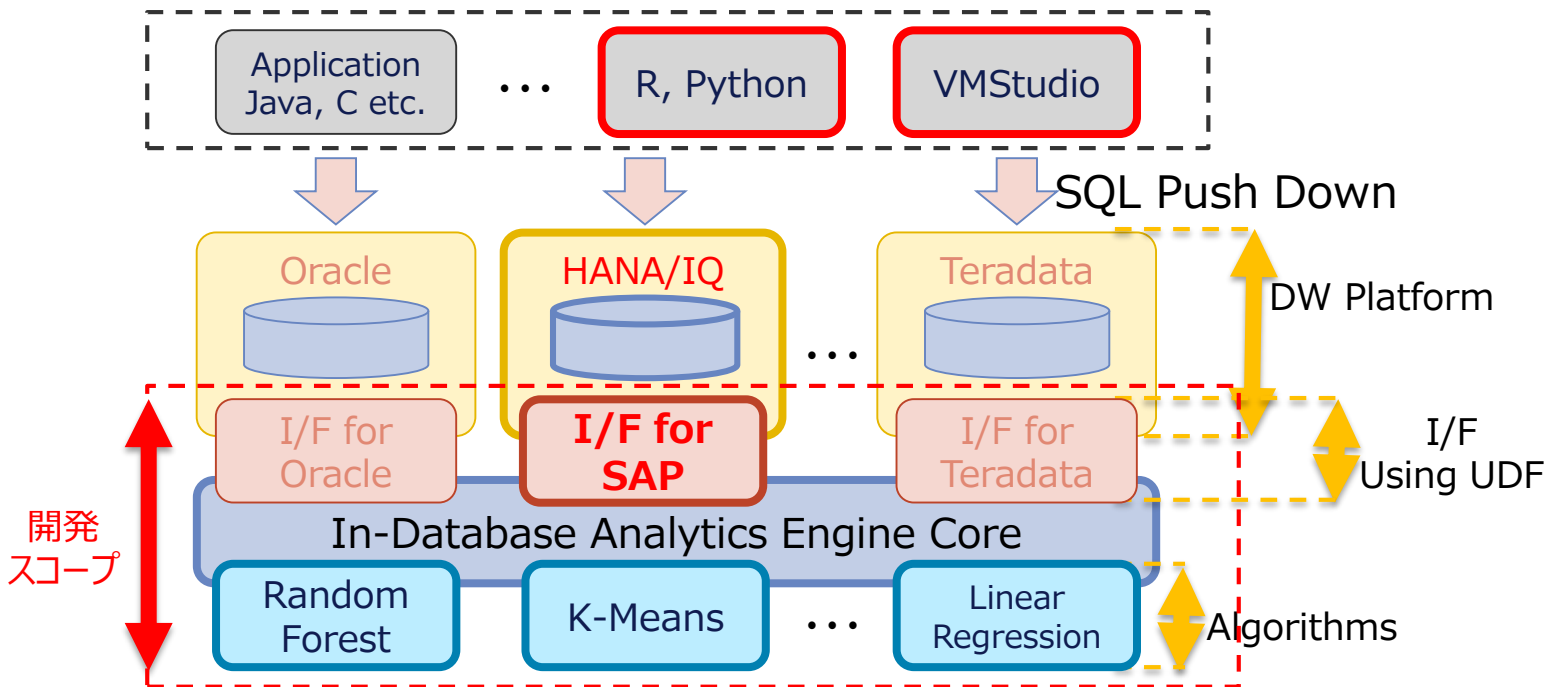
- Linear Regression
- Random Forest
etc.

効果

- 処理速度向上
- 大量データ処理

- 処理速度向上
- 分析精度向上

お客様とのPoCを通じて成長・発展させていく



プラットフォーム
アルゴリズム
分析モデル形式
UI

	FY2014	FY2015※
プラットフォーム	SAP	Oracle
アルゴリズム	3種類	ニーズに応じて拡張
分析モデル形式	オリジナル	PMML対応
UI	VMStudio、Python	ニーズに応じて拡張

※FY2015以降の方針については予告なく変わることがあります
※現在性能評価中

- In-Database Analyticsは各RDBMSが備える、独自のユーザ定義関数を実装するための機能であるUDF(User Defined Function)を用いて、C/C++で実装
- 分析アルゴリズム処理はSQL処理パートとUDF処理パートに分けられる

学習データ情報

分析パラメータ

実行パラメータ

SQL処理パート

前処理

モデル構築SQL入力

変換マップ作成
カテゴリ値↓数値変換処理
カテゴリ値↓数値

サンプリング

縦横変換等

学習データ

UDF処理パート

分析モデル構築

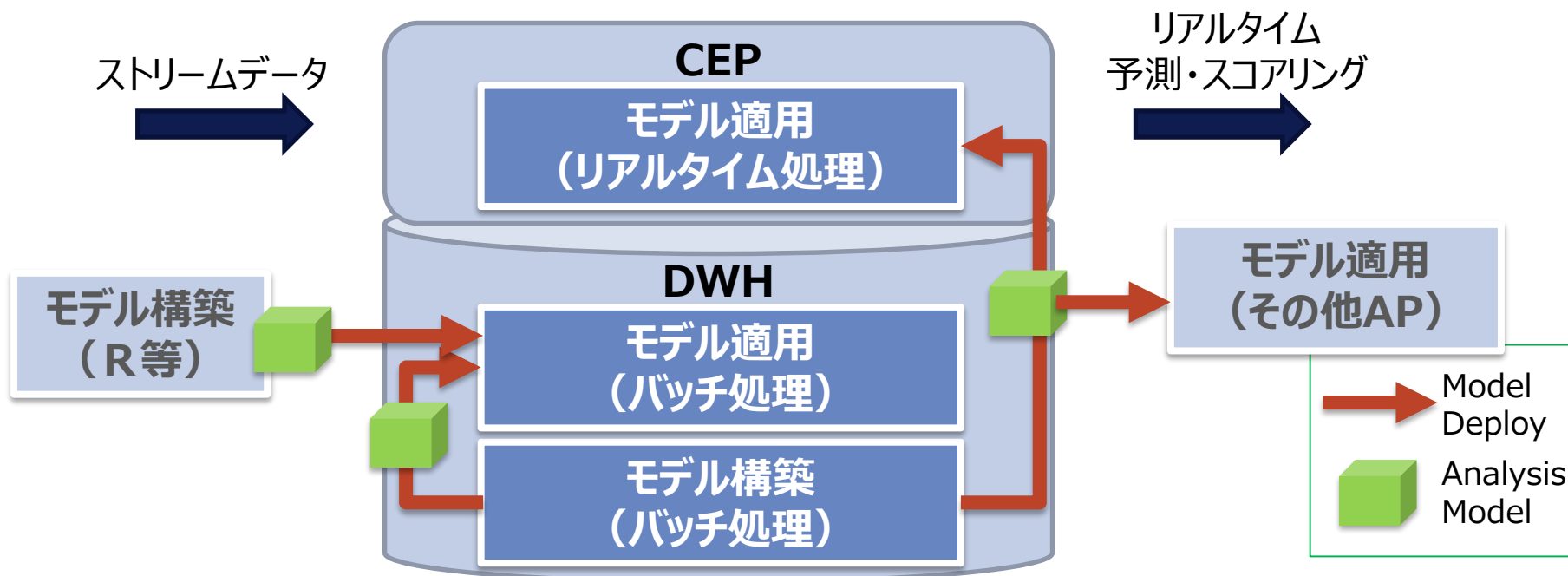
モデル構築
SQL入力Map
分析モデル構築分析モデル
(一次)Reduce
分析モデル統合モデル構築
SQL出力

分析モデル

データ分析における
これまでの課題

- 分析モデルの適用には別途アプリケーション化が必要
→QCDの低下につながる
- どの製品でも実現できていないため、個別開発が必要

分析モデルにPMML等のポータビリティを持たせる

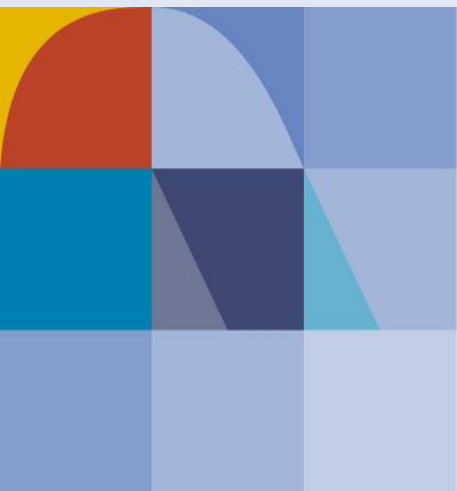


DWH内で作成した分析モデルをCEPに適用することで、
ストリームデータに対してリアルタイムに予測・スコアリングが可能となる

NTT DATA In-Database Analytics

- 蓄積された大量データに対するデータ分析を高速に実行するための分析技術
- これまでのデータ分析では扱えなかったデータ量进行处理することで新たな価値を創出
- NTTデータ数理システムの信頼性の高い分析アルゴリズムを搭載
- NTTデータのデータ分析方法論「BICLAVIS」をベースとしたアルゴリズムの選定

**企業内DWHの大量データを高速に分析を行うことで
新しいビジネス価値の創出をご支援します**



NTT DATA

Global IT Innovator

※記載されている会社名、商品名、サービス名は各社の登録商標または商標です。