

2019年10月17日

## 新システムへのスムーズなデータ移行を支援する 「データ仕様可視化ツール」を提供開始

既存システムのデータベースに格納されている実データを分析・可視化し、スムーズなデータ移行をサポート

株式会社日立製作所(以下、日立)は、データベースの各カラムのデータバリエーションや異常データを分析する「データ仕様可視化ツール」を2019年10月17日から提供開始します。

データ仕様可視化ツールは、データベースの各カラムの分析に、データの形式をパターン化して蓄積した豊富なナレッジ情報を活用したパターンマッチング技術<sup>\*1</sup>と高速な統計解析処理を駆使することで、プロファイリング結果を、分かりやすくかつ詳細にレポートとして出力するツールです。データベースに格納されているデータを分析し、データ形式やそのレコード件数、比率を可視化し、異常データの候補を抽出できます。これにより、新システムへの移行にあたって、現行システムのデータ分析と調査にかかる工数削減や、データ品質の向上を実現するとともに、想定外データ混入による設計手戻りなどデータ移行トラブルの抑制に貢献します。

長年運用されてきたITシステムのデータベースには、データ仕様の変更やパッチなどの保守作業により、初期稼働時の仕様が存在しなかったデータが含まれることが少なくありません。このため、新システムへ移行するマイグレーションの際には、実データを対象にしたデータの分析、調査により現行データ仕様を正確に把握することが重要です。現行データに異常データが含まれている場合は、データ移行設計における考慮やクレンジングなどの対応が不可欠となります。しかし、データ量が数百万・数千万件と大量の場合、SQLや表計算ソフトを利用した分析は極めて困難なものになります。

「データ仕様可視化ツール」は、日立の社内ツールとして多くのプロジェクトに導入され、マイグレーションにおけるデータ移行プログラムの設計検証の効率化やブラックボックス化したデータ仕様の可視化など、さまざまなユースケースで作業の効率化を実現してきました。このたび、「データ仕様可視化ツール」を製品として提供し、お客さまの開発プロジェクトにおいてデータ分析、調査作業の工数削減と分析精度向上の実現を支援します。今後、分析範囲の拡大や分析結果をテストデータの生成に利用する機能などを強化し、さらなるデータ分析の効率化と品質向上に寄与してまいります。

\*1：データベースに格納されているデータとナレッジ情報をマッチングし、データの種類（例：氏名や住所、電話番号など）を判別する技術です。

データ仕様可視化ツールが支援すること

- ✓ 実データの分析や仕様調査にかかる時間を短縮
- ✓ 効率的かつ精度が高いデータ移行設計による移行後のトラブル防止

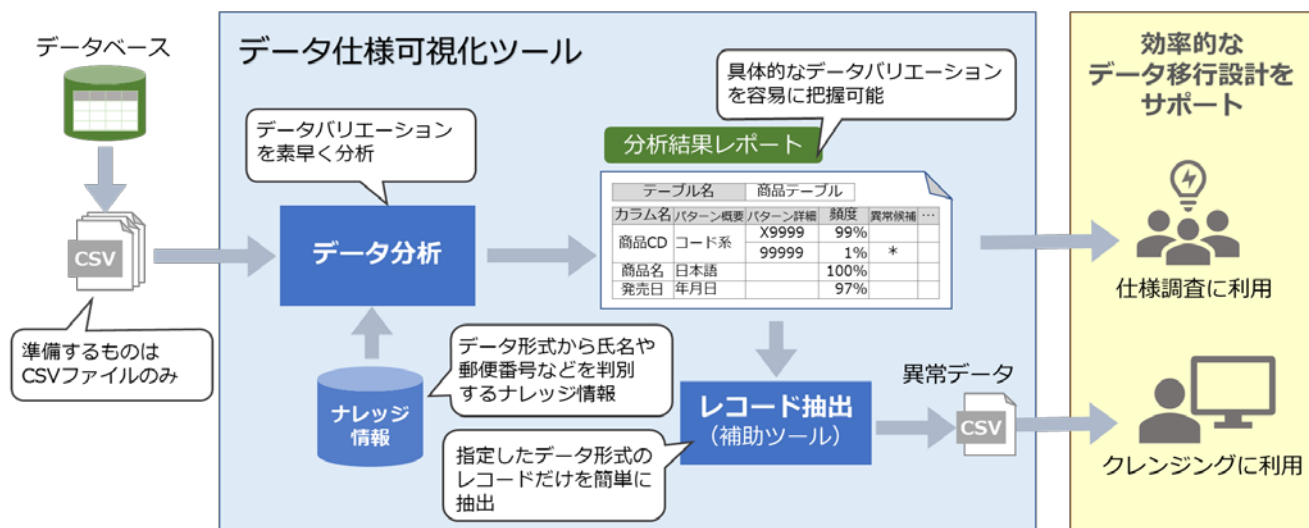


図 1 | データ仕様可視化ツールの活用イメージ

## 「データ仕様可視化ツール」の特長

### (1) 面倒な準備が不要

準備するものは CSV 形式のデータファイルのみで、テーブル定義やデータ仕様などの情報は必要ありません。また、データベースや統計解析に関する高度な知識も不要です。ご利用のデータベースから CSV ファイルをエクスポートすることで、データ分析の準備が完了します。

### (2) 膨大なデータも素早く分析

500 万件を超えるような膨大なデータでも素早く分析します。全データを網羅的に分析し、隠れた少数派データの特徴も顕在化できます。

### (3) 格納されているデータのバリエーションを具体的に出力

データの形式を 150 個以上にパターン化して蓄積したナレッジ情報と統計解析処理によって、各カラムに格納されたデータのバリエーションを、正規表現やデータ形式のパターンなどで、より詳細に区別してレポートに出力します。これにより、データバリエーションを詳細化、具体化できます。

### (4) 異常なデータを検出

分析対象データの中から、異常なデータの候補を推測してレポートに出力するため、より簡単にデータの正当性確認を行うことができます。また、付属のレコード抽出ツールを利用することで、異常データのレコードだけを CSV ファイルに抽出することもできます。これにより、効率的なデータクレンジングが可能となり、正確でスムーズなデータ移行・システム更改を支援します。

## (5) 明確で分かりやすい分析結果

各カラムのデータバリエーションやデータ分布、異常候補、データの例などを Excel ファイルに一覧形式で出力し、確認すべき問題点や検討事項の把握を容易にします。

カラム名	①パターン概要	②パターン簡易	③パターン詳細	④出現頻度	⑤異常候補	⑥出現実数	⑦出現種数	⑧例
商品CD	コード系	XX99999	^[A-Z]{2}[0-9]{5}\$	80%		8,010	8,010	FC05678
		9999999	^[0-9]{7}\$	19%		1,900	1,900	1234567
		XXX9999	^DMY[0-9]{4}\$	1%	*	90	90	DMY0001
商品区分	少データ種		<<出現データ>>1	33%		3,310	1	1
			<<出現データ>>2	33%		3,300	1	2
			<<出現データ>>3	33%		3,300	1	3
			<<出現データ>>9999	1%	*	90	1	9999
商品名	雑多		全角文字 (含: 日本語)	70%		7,000	6,000	G 4 型車輪
			半角英数 (含: 記号)	30%		3,000	2,000	R-XG43K
発売日	年月日		YYYYMMDD(区切文字有)	90%		9,000	1,500	2013/04/20
			<<空>>	10%		1,000	1	

図2 | 分析結果の例 (イメージ)

- ① データの形式を分類して出力します。上記のほかに、数値や人名、地名などにも対応しています。
- ② ①がコード系の場合、桁数の把握が容易な形式でデータ形式を出力します。
- ③ 格納されているデータの形式を出力します。
- ④ 該当するデータの割合を出力します。小数点以下の表示は Excel の書式で変更可能です。(デフォルトは小数点以下第2位まで)
- ⑤ 出現頻度が1%未満の場合に、異常データの候補として\*マークを出力します。
- ⑥ 該当するデータの件数を出力します。
- ⑦ データ件数から重複データを除いた件数を出力します。
- ⑧ 分析対象に含まれる該当データの例を出力します。その他、最小値や最大文字数などの情報も出力します。

## 「データ仕様可視化ツール」の価格

データ仕様可視化ツールは、Essential、Standard、Professionalの3つのEditionをご用意しており、それぞれのEditionで分析できるデータ件数の上限が異なります。売り切りのほかに、サブスクリプションタイプ(年間契約)もご利用できます。

Edition	分析可能データ件数の上限	価格(税別)	
		売り切り	サブスクリプション(年額)
Essential Edition	100万件	100万円	50万円
Standard Edition	500万件	200万円	100万円
Professional Edition	上限なし	400万円	200万円

## 「データ仕様可視化ツール」のウェブサイト

<https://www.hitachi.co.jp/dataviz-tool/>

## 商標に関する表示

- ・ Windows, Excel は、米国 Microsoft Corporation 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。

## お問い合わせ先

株式会社日立製作所 システム&サービスビジネス統括本部 アプリケーションサービス事業部  
事業企画部[担当：酒井、藤野、奥野]

〒212-8567 神奈川県川崎市幸区鹿島田一丁目一番二号

<https://www.hitachi.co.jp/appsvdiv-inq/>

以上