Data Ferry 製品概要資料

株式会社データビークル
2017年9月

Copyright © 2017 DATAVEHICLE Inc.

データビークルが提供する三つの武器

データ
ハンドリング

ツールと仕組で真の
データ活用を実現する

見つける

Data
Diver

シチズンデータサイエンティストを
短時間で創り出す

- リサーチデザインという新たな概念
- 分析の専門家以上の結果を出せるUI
- 既存のDWH.BIを加速させる機能
- 2015年7月リリース

DATA DIVER

運ぶ

Data
Ferry

既存ITと調和しながら
短時間でデータ加工を実現する

- データ分析専用変換ツール
- プログラム経験ゼロで使えるUI&UX
- エンジン部分は4000社以上で実績ある既存製品を活用
- 2016年9月リリース

DATA FERRY

次の次元へ

Moon
Base
（仮）

データホルダーとユーザのマッチングで
今までになかったデータを入手する

- 分析に必要な社外データをワンストップで提供
- 国内有名データホルダーが参画
- 他社の追従を許さないキーテクノロジーとノウハウ
- 2017年リリース予定

Moon Base

Copyright © 2017 DATAVEHICLE Inc.

データサイエンスとデータビークル

課題：誰とデータサイエンスを実現するか？

構成要素：工数・時間・コスト ・ 分析スキル・データの質 ・ 分析データの多様性

自社で解決

ツールで解決

人材育成で解決

データで解決

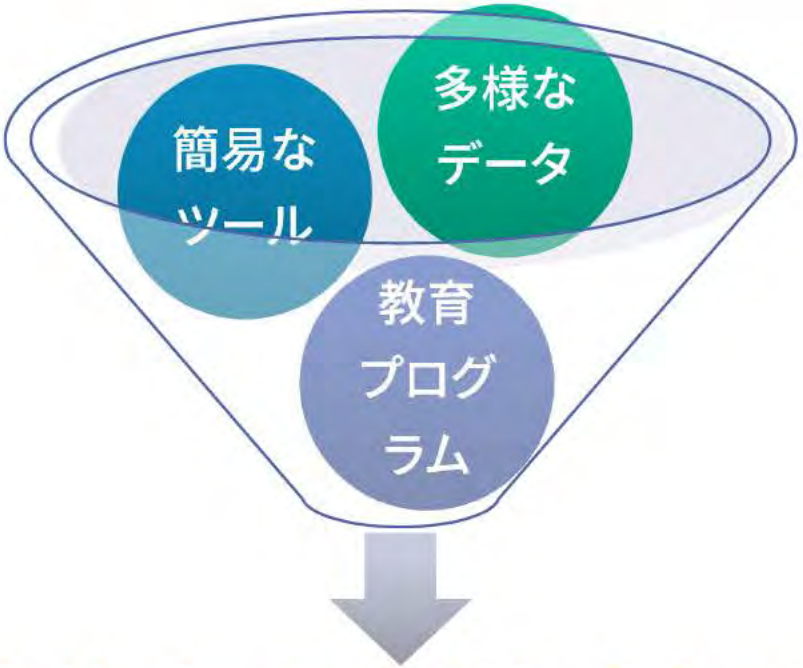
データビークルの支援

外部データサイエンティストに委託

アウトソーシングで解決

Copyright © 2017 DATAVEHICLE Inc.

データビークルが目指す未来



データサイエンスをみんなの手に

Copyright © 2017 DATAVEHICLE Inc.

Data Ferryとは

-データサイエンス「専用」変換ツール

分析に特化した機能を優先して実装

素早く、確実にデータを手元に届けます

-対象ユーザーはIT部門「以外」

設定はすべて「自然言語」、分析チームが自分で使えるツール

ExcelやAccessで行っている作業をよりスムーズに実行

-処理は開発でなく設定で作業完了

プログラムスキルは一切不要、思い通りにデータを「Tune Up」

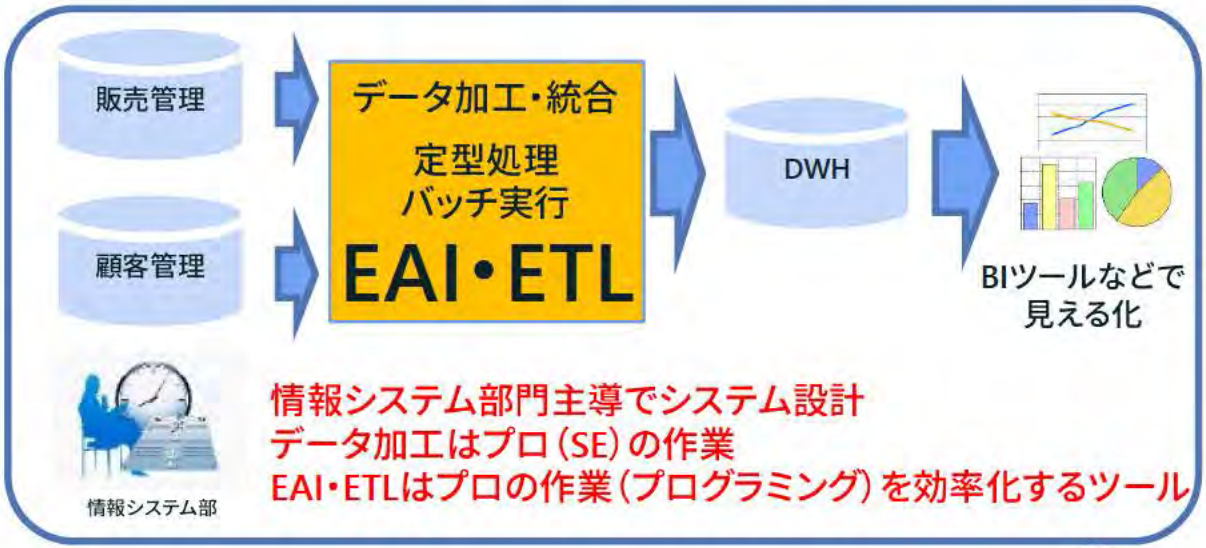
Webブラウザを用意すればウィザード形式の設定で作業完了

Copyright © 2017 DATAVEHICLE Inc.

データ加工・統合を取り巻く環境の変化

・これまで

アウトプットが明確

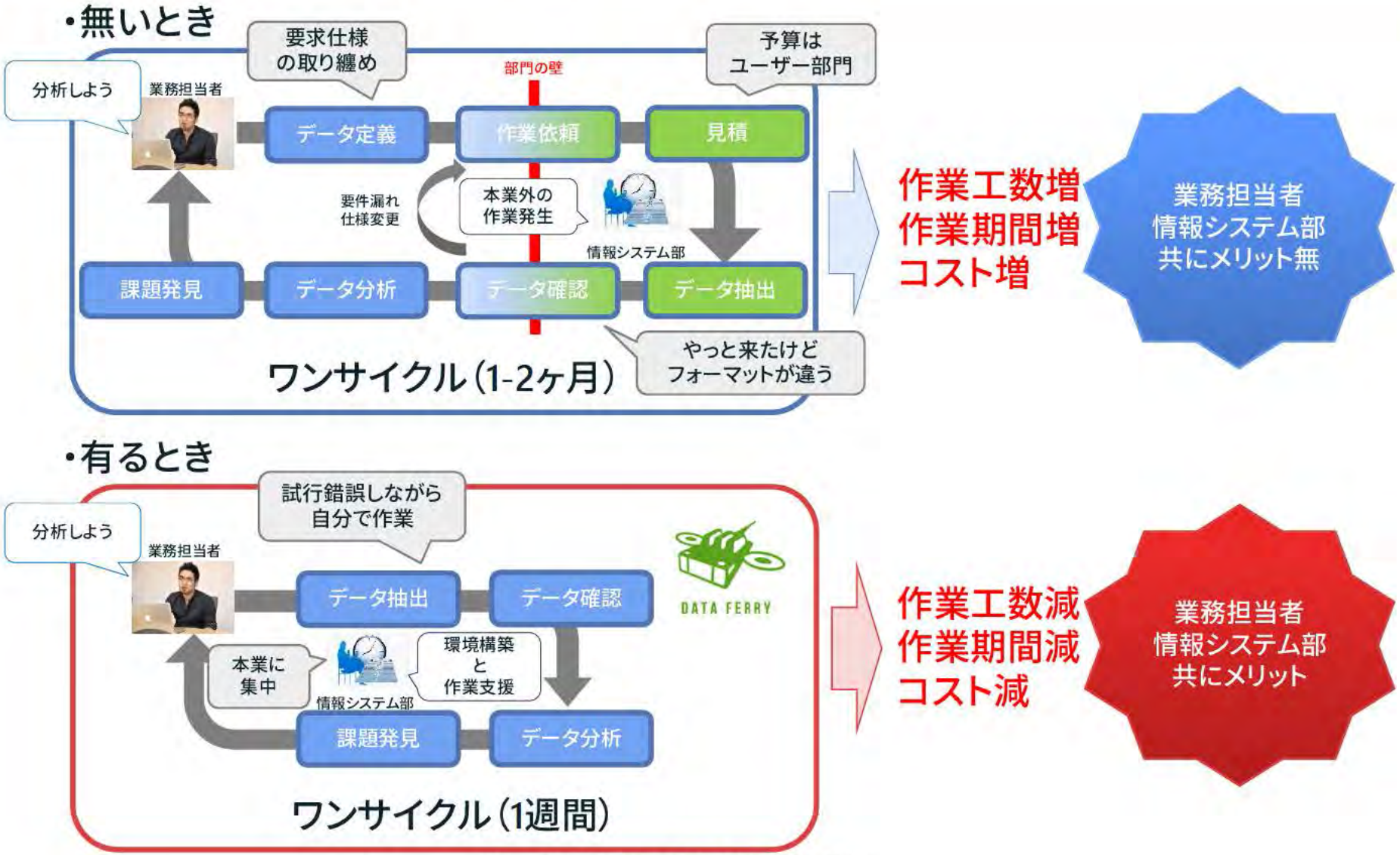


・これから

アウトプットが不明確

Copyright © 2017 DATAVEHICLE Inc.

Data Ferryがもたらす変化



Copyright © 2017 DATAVEHICLE Inc.

Data Ferry 二大特長

1.分析に向かないデータの削除

2.分析に向けたデータの修正・生成

Copyright © 2017 DATAVEHICLE Inc.

分析に向かないデータの削除とは

分析のためのデータ特徴

中身は「数値の大小」「(数十個程度への)分類」

中身に抜け・漏れ・異常値があると使えない

上記2条件に該当しないデータはData Ferryで削除

数値でも分類でもない個人属性

年齢に異常値が設定

・INPUT

顧客ID	名前	県	TEL	性別	年齢	・・
ID001	ヤマサキ	兵庫県	090-0000		999	・・
ID002	フジタ	東京都	03-0000	女	28	・・
ID003	ユノ	大阪府	06-0000	男	30	・・
ID004	ヤマナカ	奈良県	0742-00		32	・・
ID005	イワイ	愛知県	052-000	女		
・	・	・	・			
・	・	・	・			

顧客データ

性別が抜け

数値でも分類でもない個人属性

・OUTPUT

顧客ID	県	性別	年齢	・・
ID002	東京都	女	28	・・
ID003	大阪府	男	30	・・
ID005	愛知県	女	23	・・
・	・	・	・	・・
・	・	・	・	・・

顧客データ

ユニークキーと数値、分類のみのデータ構成
かつ各項目に抜け・漏れ・異常値無し

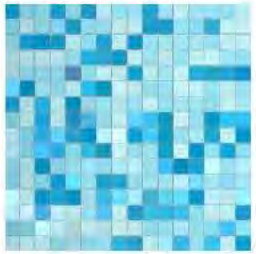
Copyright © 2017 DATAVEHICLE Inc.

分析に向けたデータの修正・生成とは

分析に向けたデータの修正・生成

中身に抜け・漏れがあるデータの修正

業務データに存在しない値（フラグ、計算値）の追加



分析方針に応じたデータ修正・生成を実行

・INPUT

顧客ID	県	性別	生年月日	・・
ID002	東京都	女	1990/5/10	・・
ID003	大阪府	男	1985/3/3	・・
ID005	愛知県		1994/10/5	・・
・	・	・	・	・・
・	・	・	・	・・

顧客データ

購入日	顧客ID	商品ID	数量	単価	・・
2017年4月1日	ID002	A	2	480	・・
2017年5月2日	ID002	B	1	150	・・
2017年4月2日	ID003	A	2	480	・・
・					
・					

販売データ

生年月日から年齢を計算

販売データを確認し分類フラグを追加

・OUTPUT

顧客ID	県	性別	年齢	毎月購入有無	・・
ID002	東京都	女	28	有り	・・
ID003	大阪府	男	30	無し	・・
ID005	愛知県	不明	23	有り	・・
・		・	・	・	・・
・		・	・	・	・・

抜け項目に値を設定

顧客データ

購入日	顧客ID	商品ID	数量	単価	売上合計	・・
2017年4月1日	ID002	A	2	480	960	・・
2017年5月2日	ID002	B	1	150	150	・・
2017年4月2日	ID003			480	960	・・
・	・				・	・・
・	・				・	・・

単価×数量を計算

販売データ

Copyright © 2017 DATAVEHICLE Inc.

Data Ferryで出来ること

対応フォーマット	・CSV/RDB/Excel
フィルタリング	・重複行を削除 ・数値範囲指定 ・文字列指定（完全一致・曖昧） ・複合条件（AND/OR）指定
ファイル結合	・縦結合（ファイル統合）、横結合（レコード統合） ・横結合種の選択（内部結合・外部結合など） ・横結合時のカラム統合（出力項目選択）
集計	・特定の項目でグループ化し、数値項目について SUM/MAX/MIN/AVEなどの集計処理データを出力 ・特定の項目でグループ化し、日付項目について 年別/月別/曜日別などの集計処理データを出力
クレンジング	・ユニーク値付与 ・異常値を排除 ・文字列加工（半角/全角統一、各種置換処理、カラム結合など） ・数値加工（範囲置換、四則演算など） ・日付加工（日付計算、フォーマット整形など）
サンプリング	・特定の割合/件数を指定してランダムでデータを抽出
その他	・文字コード変換 ・データ確認（データセット率の表示など）

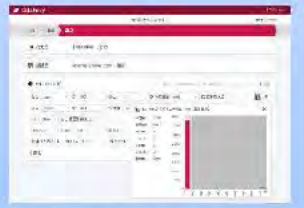
Copyright © 2017 DATAVEHICLE Inc.

Data Ferry利用手順

1.取り込みたい
データを選択し
アップロードします。

2.利用したいデータ
だけ選択します。

3.複数のデータを
一つのデータに
纏めます。



4.データを保存する際に実
行するクレンジング
設定を行います。

5.処理が終了する
まで待ちます。

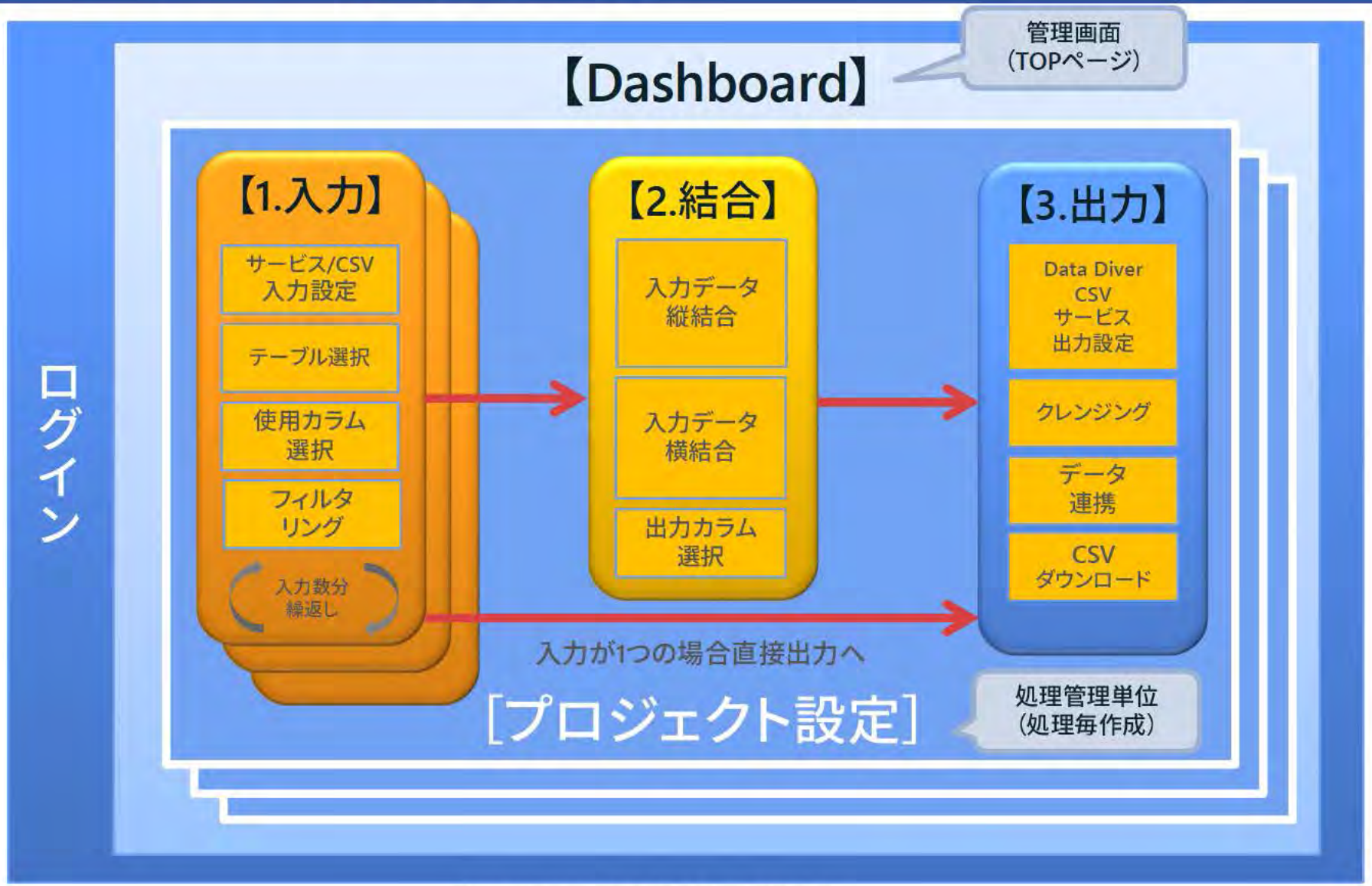
6.実行結果を確認し
データをダウンロードし
ます。

毎週、毎月同じ処理を行いフレッシュなデータを作成することもできます。また、一度行った処理を一部変更し再利用も可能です。

- 何度も試行錯誤しながら分析用データを作成することが可能です。
- 必要なタイミングで分析用データを作成することが可能です。
- データのゆらぎを削減し使える分析用データを作成することが可能です。

Copyright © 2017 DATAVEHICLE Inc.

Data Ferry画面構成



Copyright © 2017 DATAVEHICLE Inc.

画面詳細

1.入力

Data Ferryで加工する入力元を登録します。

Copyright © 2017 DATAVEHICLE Inc.

画面詳細

1.入力
入力元のフィルタリングを設定します。

Copyright © 2017 DATAVEHICLE Inc.

画面詳細

2.結合
入力データの結合設定をします。

Copyright © 2017 DATAVEHICLE Inc.

画面詳細

3.出力
取り込んだデータのクレンジング・出力設定を行います。

Copyright © 2017 DATAVEHICLE Inc.

サービス仕様

- 提供方式
  - クラウドサービス

    月額40万円（税別）最低契約期間6ヶ月
    １契約＝１アカウントでの提供となります（複数アカウントについては要相談）。
    ストレージ容量は初期容量１TB、ご要望に応じて追加可能です（別途有償）。
  - オンプレミス

    ライセンス方式
    サブスクリプション型（年間利用料）200万（税別）
    使用権購入型600万円（税別）※別途年間15％の保守費

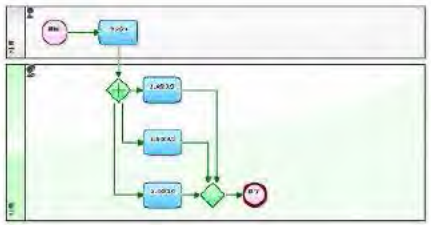
- 稼働環境ブラウザ（最新版に対応）

  Firefox
  Google Chrome
  Safari

Copyright © 2017 DATAVEHICLE Inc.

他製品比較

これまでのデータ連携・統合ツールとは別カテゴリとなります。

	利用形態	変換方法（提供機能）	想定ユーザー
EAI・ETL	・バッチ実行 ・定型処理 	・演算中心 （部品提供） ・処理は開発 	・SIerのSE ・情報システム部門
Data Ferry	・リアルタイム処理 ・非定形処理 Think × Try × Select	・データの意味づけ・分類 （機能提供） ・処理に従って設定のみ 	・業務担当者

Copyright © 2017 DATAVEHICLE Inc.

まとめ

Data Ferryとは

-データサイエンス「専用」変換ツール

分析に特化した機能を優先して実装
素早く、確実にデータを手元に調達

-対象ユーザーはIT部門「以外」

設定はすべて「自然言語」、分析チームが自分で使えるツール
ExcelやAccessで行っている作業をよりスムーズに実行

-処理は開発でなく設定で作業完了

プログラムスキルは一切不要、思い通りにデータを「Tune Up」
Webブラウザを用意すればウィザード形式の設定で作業完了

作業工数減　コスト減　作業期間減

Copyright © 2017 DATAVEHICLE Inc.

DATAVEHICLE

Copyright © 2017 DATAVEHICLE Inc.