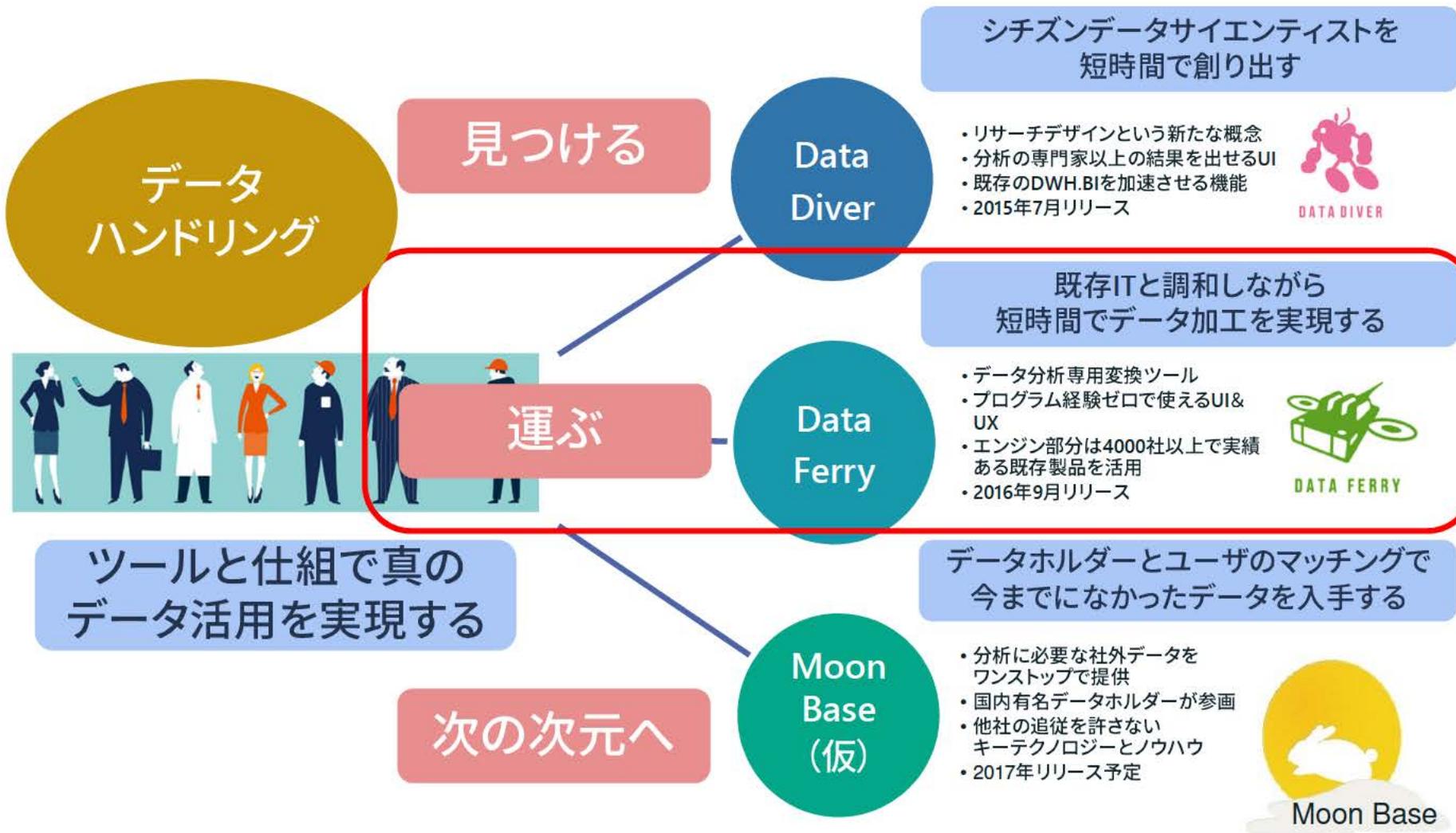




# Data Ferry 製品概要資料

株式会社データビークル  
2017年9月

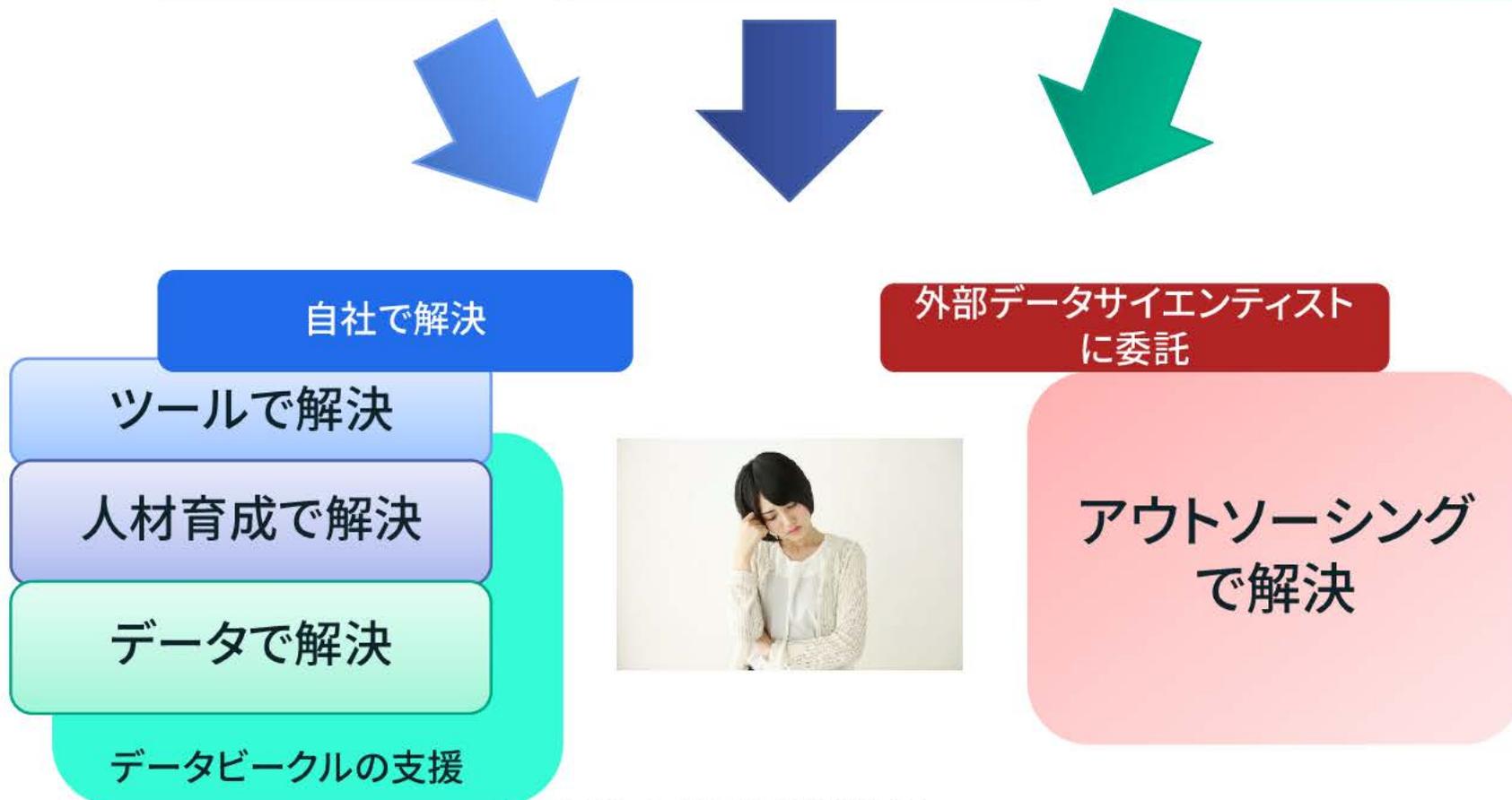
# データビークルが提供する三つの武器



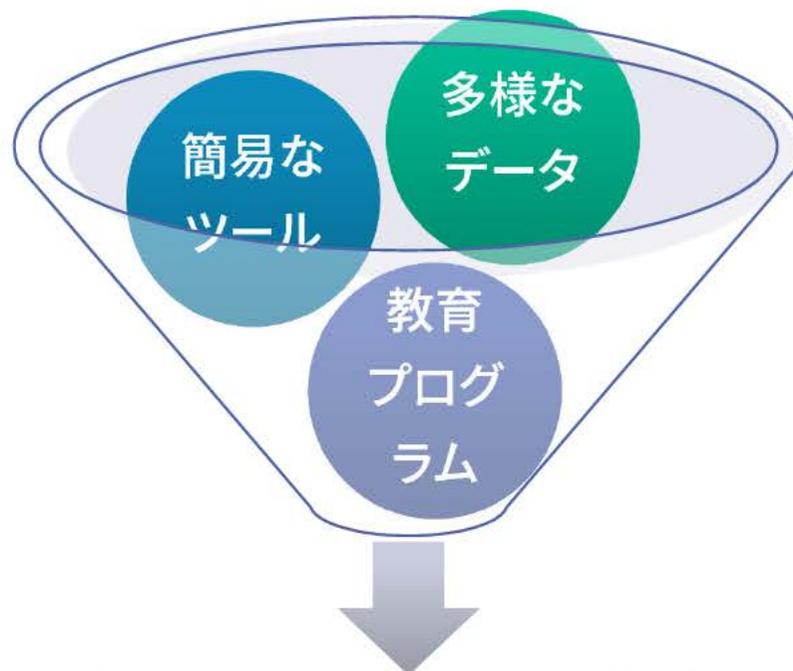
# データサイエンスとデータビークル

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



# 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



# 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

# 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

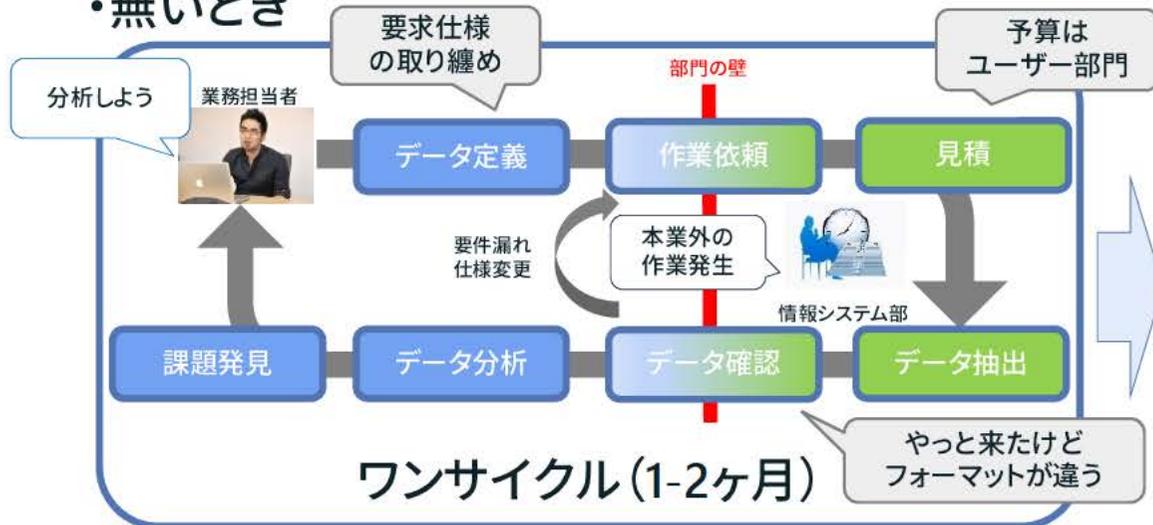


アウトプットが不明確



# Data Ferryがもたらす変化

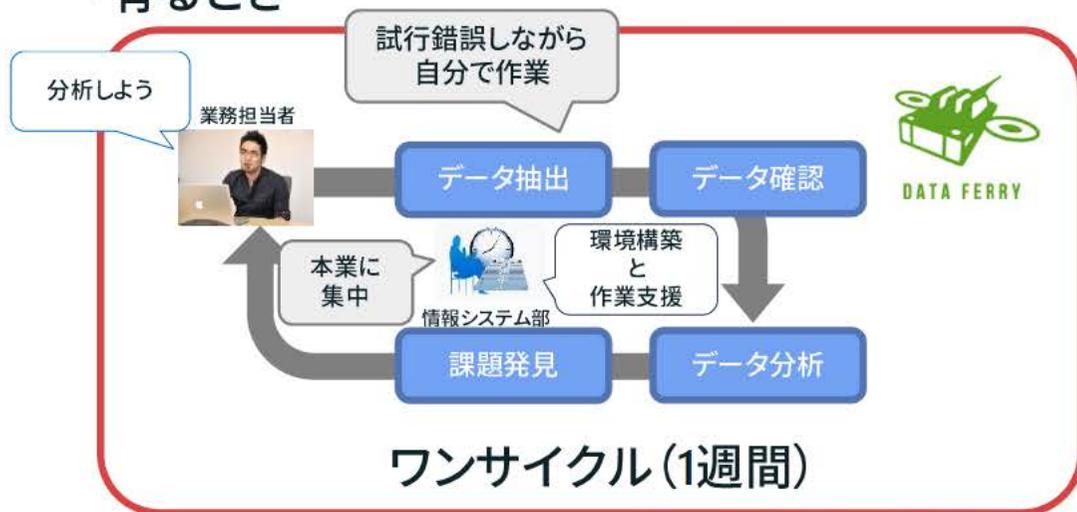
## ・無いとき



作業工数増  
作業期間増  
コスト増

業務担当者  
情報システム部  
共にメリット無

## ・有るとき



作業工数減  
作業期間減  
コスト減

業務担当者  
情報システム部  
共にメリット

# 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

# 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



### •INPUT

顧客ID	名前	県	TEL	性別	年齢	..
D001	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	..
D004	ヤマナカ	奈良県	0742-00		32	..
ID005	イワイ	愛知県	052-000	女		
.	.	.	.			
.	.	.	.			

数値でも分類でもない個人属性

年齢に異常値が設定

性別が抜け

顧客データ

数値でも分類でもない個人属性



DATA FERRY



### •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

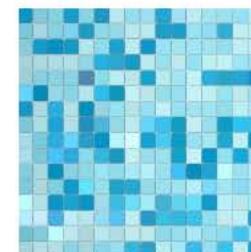
# 分析に向けたデータの修正・生成とは

## 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



DATA FERRY



### •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

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

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	A	2	480	960	..
.	.	.	.	.	.	..
.	.	.	.	.	.	..

単価×数量  
を計算

販売データ

# 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)

# 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。

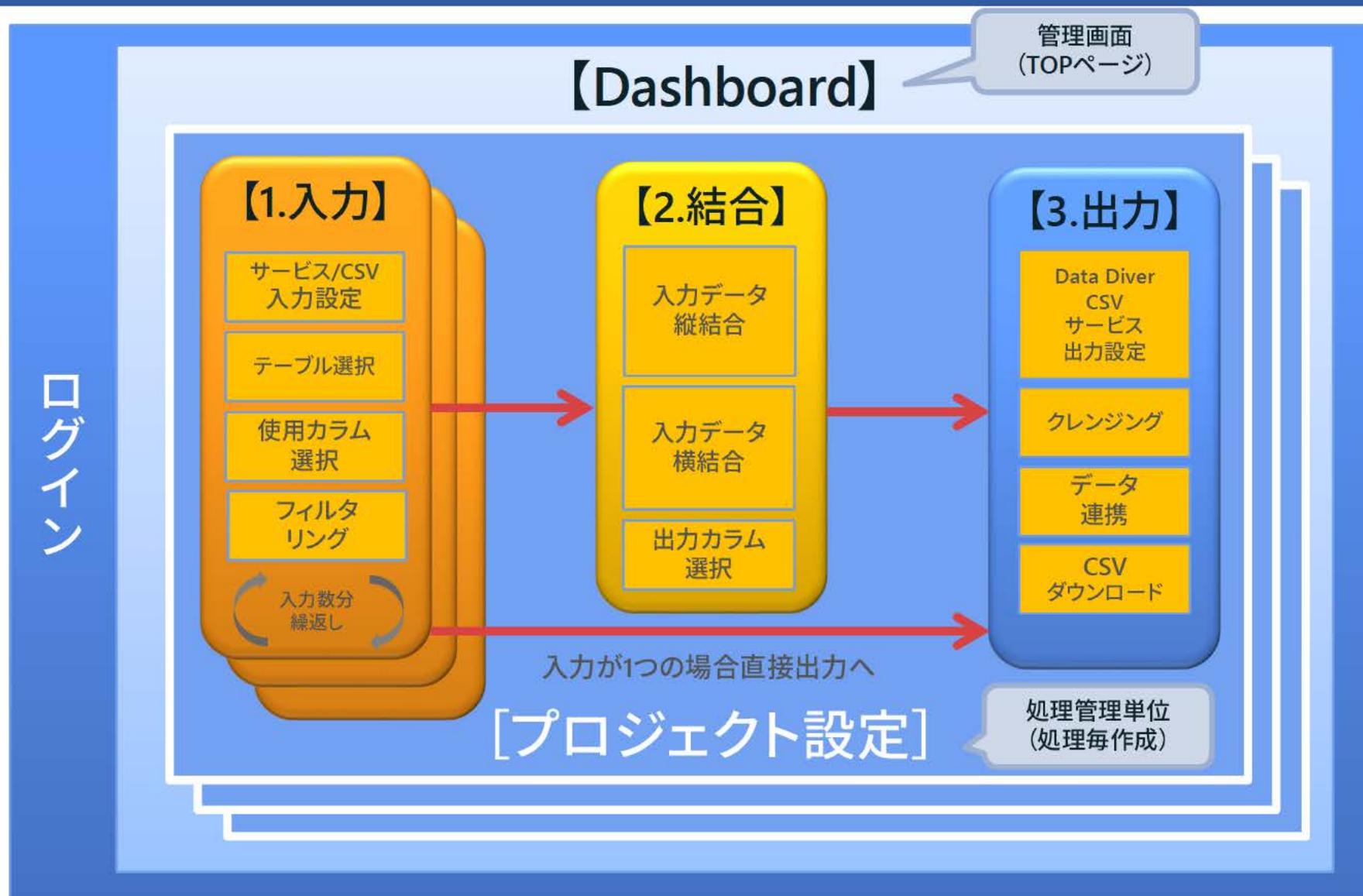


毎週、毎月同じ処理を行いフレッシュなデータを作成することもできます。また、一度行った処理を一部変更し再利用も可能です。



- 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。

# Data Ferry画面構成



# 画面詳細

## 1.入力

Data Ferryで加工する入力元を登録します。

1-1.入力設定  
ファイルやサービスを  
選択します。

2-2.使用カラム選択  
入力データの使用する  
カラムを選択します。

Column Name	Data Set Rate	Type	Check
会員ID	100.00%	文字列	<input checked="" type="checkbox"/>
購入日	100.00%	文字列	<input checked="" type="checkbox"/>
アイテムID	100.00%	文字列	<input checked="" type="checkbox"/>
購入額	100.00%	文字列	<input checked="" type="checkbox"/>
割引コード	68.98%	文字列	<input checked="" type="checkbox"/>
割引額	68.98%	数値	<input checked="" type="checkbox"/>

# 画面詳細

## 1.入力

入力元のフィルタリングを設定します。

The screenshot displays the 'Data Ferry' application interface. The top navigation bar includes the 'Data Ferry' logo, a user profile 'Tatsuya', and a '新規プロジェクト' (New Project) button. The main content area is divided into two panels. The left panel, titled '入力' (Input), shows the configuration for the data source: '入力元選択' (Input Source Selection) set to 'データベース' (Database), '接続先情報入力' (Connection Information Input) set to 'MyDatabase', and 'データ選択' (Data Selection) set to 'example\_service.User'. Below this is the 'フィルタリング' (Filtering) section, which includes a '+ 条件追加' (Add Condition) button and a range filter for 'age' (10 to 100). The right panel shows a data preview table for 'User (2523件中100件を表示)' (User (2523 items, 100 items displayed)). The table has columns for 'ferry\_id', 'shop\_id', 'name', 'postcode', 'address1', 'tel', and 'email'. A callout bubble points to the table with the text '入力データのプレビューが可能' (Data preview possible). Another callout bubble points to the filtering section with the text '1-3.フィルタリング 選択したいレコード条件を設定します。' (1-3. Filtering: Set conditions for the records you want to select).

ferry_id	shop_id	name	postcode	address1	tel	email
0	141	ヒロム屋x号店	111-2222	東京都新宿区	03-xxxx-xxxx	
1	141	ヒロム屋x号店	111-2222	東京都新宿区	03-xxxx-xxxx	
2	141	ヒロム屋x号店	111-2222	東京都新宿区	03-xxxx-xxxx	
3	141	ヒロム屋x号店	111-2222	東京都新宿区	03-xxxx-xxxx	
4	141	ヒロム屋x号店	111-2222	東京都新宿区	02-xxxx-xxxx	info@xxx.com
5	141	ヒロム屋x号店	111-2222	東京都新宿区	03-xxxx-xxxx	info@xxx.com
6	141	ヒロム屋x号店	111-2222	東京都新宿区	03-xxxx-xxxx	info@xxx.com
7	141	ヒロム屋x号店	111-2222	東京都新宿区	03-xxxx-xxxx	info@xxx.com
8	141	ヒロム屋x号店	111-2222	東京都新宿区	03-xxxx-xxxx	info@xxx.com
9	141	ヒロム屋x号店	111-2222	東京都新宿区	03-xxxx-xxxx	info@xxx.com
10	141	ヒロム屋x号店	111-2222	東京都新宿区	03-xxxx-xxxx	info@xxx.com
11	141	ヒロム屋x号店	111-2222	東京都新宿区	03-xxxx-xxxx	info@xxx.com
12	141	ヒロム屋x号店	111-2222	東京都新宿区	03-xxxx-xxxx	info@xxx.com
13	141	ヒロム屋x号店	111-2222	東京都新宿区	03-xxxx-xxxx	info@xxx.com
14	141	ヒロム屋x号店	111-2222	東京都新宿区	03-xxxx-xxxx	info@xxx.com

# 画面詳細

## 2.結合 入力データの結合設定をします。

2-1.結合方向選択  
入力データの結合  
方向を縦・横から  
選択します。

2.2.出力カラム選択  
結合方向に基づき出  
力カラムの設定を行  
います。

Table1	A	B	C
+	+	+	+
Table2	A'	B'	C'

Table1	A	C
+	+	+
Table2	A	B

Output	A"	B"	C"
--------	----	----	----

Table1	A	C
+	+	+
Table2	A	B

Output	A	B	C
--------	---	---	---

User	created_at	id	age	name	sex
example@dev.com					

User1	user_id	age	name	tel	gender	mail
user@dev.com						

User2	uid	age	name	sex	tel	mail	post_code
customer@dev.com							

User3	uid	age	name	sex	tel	mail	post_code
201-61106@example.com							

Output	created_at	id	age	name	sex	tel	post_code	gender
--------	------------	----	-----	------	-----	-----	-----------	--------

# 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。

3-1.出力設定  
ファイルやサービス  
を選択します。

3-2.クレンジング設定  
クレンジングが必要となる  
出力カラム毎にクレンジング  
設定を行います。

# 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)

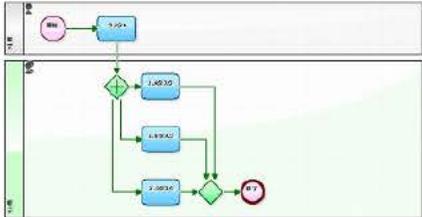
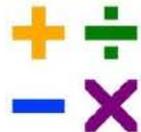
- Firefox

- Google Chrome

- Safari

# 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	<ul style="list-style-type: none"> <li>• バッチ実行</li> <li>• 定型処理</li> </ul> 	<ul style="list-style-type: none"> <li>• 演算中心 (部品提供)   </li> <li>• 処理は開発   </li> </ul>	<ul style="list-style-type: none"> <li>• SierのSE</li> <li>• 情報システム部門</li> </ul> 
Data Ferry	<ul style="list-style-type: none"> <li>• リアルタイム処理</li> <li>• 非定型処理</li> </ul> 	<ul style="list-style-type: none"> <li>• データの意味づけ・分類 (機能提供)</li> <li>• 処理に従って設定のみ</li> </ul> 	<ul style="list-style-type: none"> <li>• 業務担当者</li> </ul> 

# 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



**DATA VEHICLE**