



Treasure Academy

トレジャーアカデミー

トレーニングと認定資格制度のご紹介

2つのプログラムおよび認定資格制度を提供

これからCDPに関わる方

レギュラープログラム

- CDPプロジェクトリーダー養成コース
 - CDP活用、データ利活用を進めるために必要な知識、およびCDPプロジェクトの推進ノウハウを学ぶ
- CDPエンジニア養成コース
 - CDPの活用を進めるためにエンジニアが必要な知識を学ぶ

レギュラープログラム修了認定

- コース終了後の修了試験で合格者を認定
 - TD Certified Project Leader Course
 - 修了証を発行
- コース終了後の修了試験で合格者を認定
 - TD Certified Engineer Course
 - 修了証を発行

CDPの経験をお持ちの方

アドバンスプログラム

- CDPを実装/活用する方を育成
 - CDPの知識を持つエンジニアやマーケターを対象としたコースで、CDPの実装方法や活用方法を演習やデモ環境を通して実践的に学ぶ

アドバンス認定

- 認定試験を用意
 - 合格者には認定証を発行



レギュラープログラム

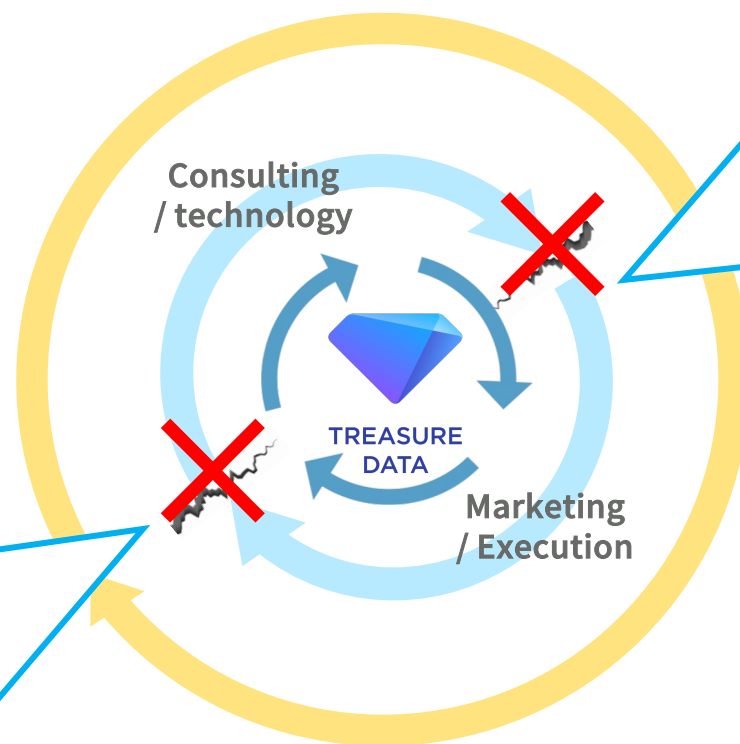


背景：データ利活用を実現しようとする、よく陥る状況

よくある企業の課題。双方のスキルや理解、言語に乖離があり、推進できない

ビジネス・マーケ視点

仮説ありきの施策立案中心のため、テクノロジーに関するリテラシーが追いつかない…。



テクノロジー視点

各種ツールのソリューションに関する理解は深いが、何のために導入するのかビジョンを描けない…。



CDP Project Leader 養成コース

- Treasure Data CDPの活用、データ利活用を進めるために必要な知識、およびプロジェクト設計、推進方法を学ぶ
- 本コースを通じて社内でTreasure Data CDPのProject Leaderとなれる人材を養成



CDP Engineer養成コース

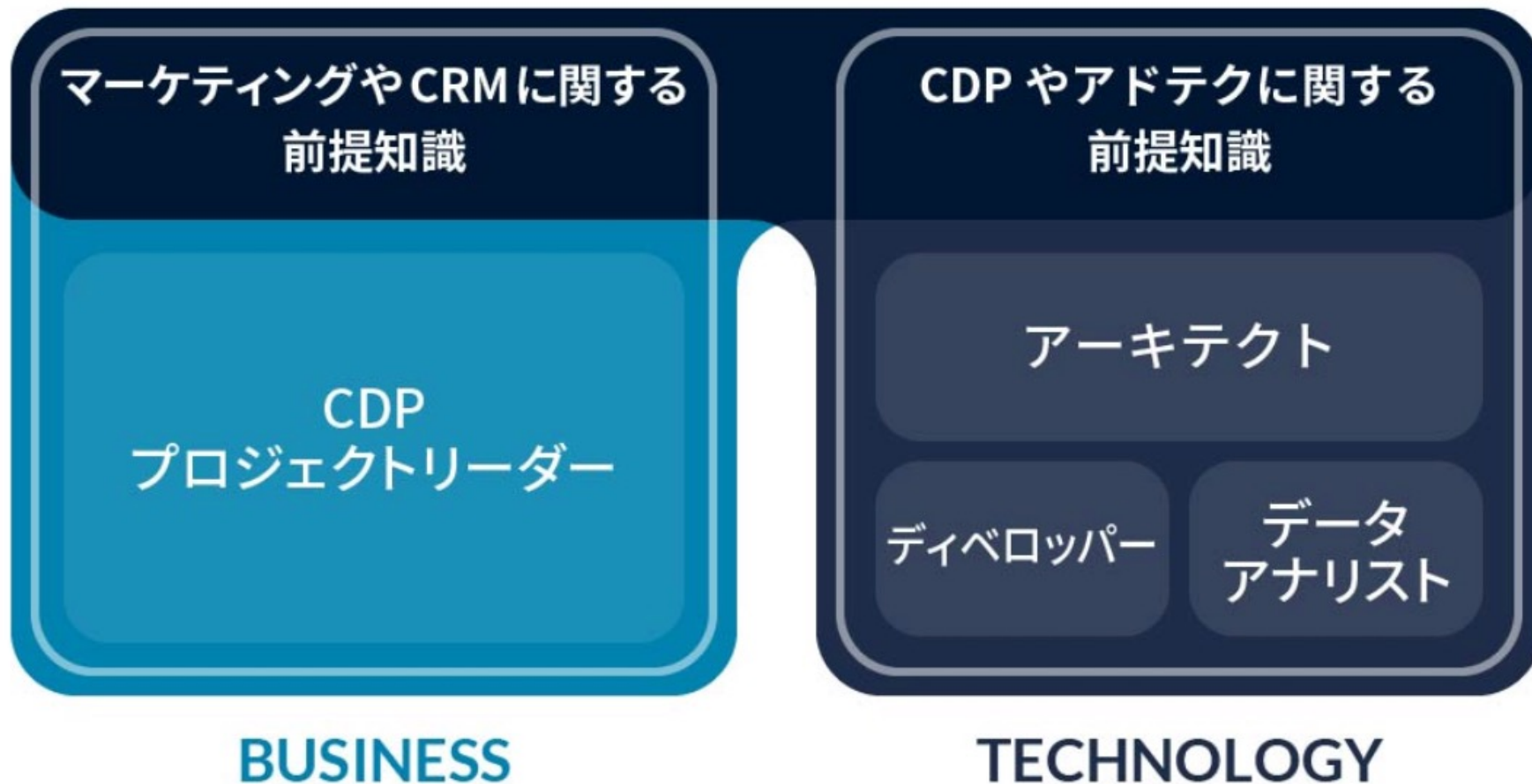
- Treasure Data CDPの活用、データ利活用を進めるためにエンジニアとして必要な知識を学ぶ
- Treasure Data CDPの特徴、各機能に精通し、施策を実現するために必要な知識を習得

レギュラープログラムの目指す姿

技術のわかるプロマネ／ビジネス理解のあるエンジニア、
二種類のπ字型人材を揃えることがCDPプロジェクト成功の要件

ビジネスチームのカバレッジ

エンジニアチームのカバレッジ



レギュラープログラムの全体像

手法だけでなく、実際のケースをもとに実務プロセスに使えるカリキュラム構成

- ・ CDP Project Leader 養成コース & CDP Engineer 養成コース



CDP Project Leader養成コース

CDPプロジェクトの目的/戦略設計からデータ活用までの実務プロセス、CDPを活用するための手法、およびプライバシー保護に関連する法的な知識を習得



Treasure Data CDPの構築やデータ活用のために、エンジニアとして必要な手法を講義とハンズオン形式の演習を通して習得



データ利活用およびCDPの活用を実現するケイパビリティの“インハウス化”を徹底サポート

ハンズオンセッションの充実

- 各コースの内容自体を詳細化していることに加えて、参加者は実際にサンプルデータを使いながら、CDPを操作するハンズオンを実施
- 少人数制を採用し、講師が一人ひとり質問に対して対応する時間を確保
- 学習初期のつまづきを最小化し、社内データ活用の早期立ち上げに貢献
 - データインポート、アグリゲーション、データエクスポート等、実務に沿ったコンテンツを用意

PLのノウハウを体系的に提供

- これまで形式化されてこなかった、CDP導入やデータマネジメント系プロジェクトのノウハウを体系化
- プロジェクトを成功に導く設計やステークホルダーマネジメントから、CDP実装時の注意点、そしてユースケースまでを一気通貫で指導し、生きたノウハウを身につける
- 技術面での知識を実務に使える形に理解・昇華させるためのプログラム

プライバシー保護への注力

- データ利活用時に今や避けては通れない、プライバシー保護の問題を徹底解説し、企業のリスクマネジメントへ貢献
- 大手法律事務所と提携することで、体系的かつ正確な法律知識をマーケティング実務者にも身につけてもらう
- GDPRや個人情報保護法改正、Cookie規制等、重要なトピックを網羅

Project Leader / Engineer 養成コースのコンテンツ内訳（１）

No	科目	講義時間の目安(h)	Engineer Course	Project Leader Course	科目修了試験	備考
1	はじめてのCDP	2.0	任意	任意	なし	動画のみ
2	Basic	3.5	必須	必須	あり	
3	Administrator	3.0	必須	－	あり	
4	SQL 基礎1	3.0	必須	必須	あり	動画あり
5	SQL 基礎2	3.0	必須	－	あり	
6	SQL 基礎3	3.0	必須	－	あり	
7	Treasure Workflow/Custom Scripts概論	3.0	必須	－	あり	
8	Treasure Workflow演習	2.0	必須	－	あり	演習解説（講義は動画）
9	Custom Scripts演習	3.0	必須	－	あり	
10	Audience Studio for Marketers	4.0	必須	必須	あり	
11	Audience Studio for Engineers	4.0	必須	－	なし	
12	TD AutoML	3.0	任意	任意	なし	
13	Project Leader Session 1: 重要要素と課題の把握・理解	1.5	聴講可	必須	あり	講義は動画
14	Project Leader Session 2: 構想・要求定義	1.0	聴講可	必須	あり	講義は動画
15	Project Leader Session 3: CDP基盤設計パート	1.0	聴講可	必須	あり	講義は動画
16	Project Leader Session 4: CDP基盤実装パート	1.0	聴講可	必須	あり	講義は動画
17	Project Leader Session 5: 施策活用パート（可視化・分析）	1.0	聴講可	必須	あり	講義は動画
18	Project Leader Session 6: 施策活用パート（セグメント基盤）	1.0	聴講可	必須	あり	講義は動画

100%

[illegible]



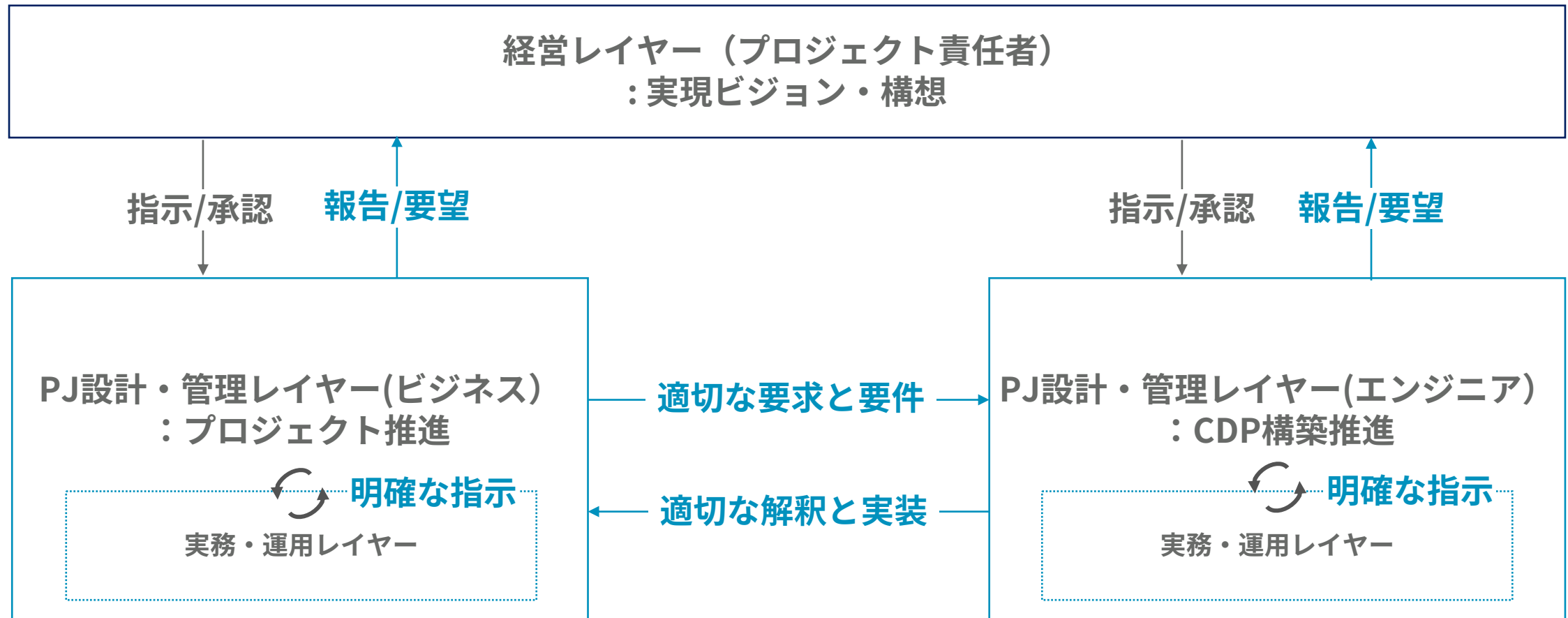
実務プロセス別プログラム概要 (CDP Project Leader養成コース)



CDP Project Leader 養成コースの対象者とゴール

ビジネス/エンジニア領域のPJ設計・管理レイヤーが主な対象

ビジネス/エンジニア双方の視点で、プロジェクト推進に必要な情報を理解し、全体コントロールを図る



Project Leader / Engineer 共通コース



プロジェクトリーダー

エンジニア

以下の3コースは、Project LeaderとEngineerの共通コースです

※各コースの詳細はエンジニアコースのスライドをご覧ください

- Basic

- Treasure Data CDPの全体像を理解し、使い始められるようになります

- SQL基礎編1

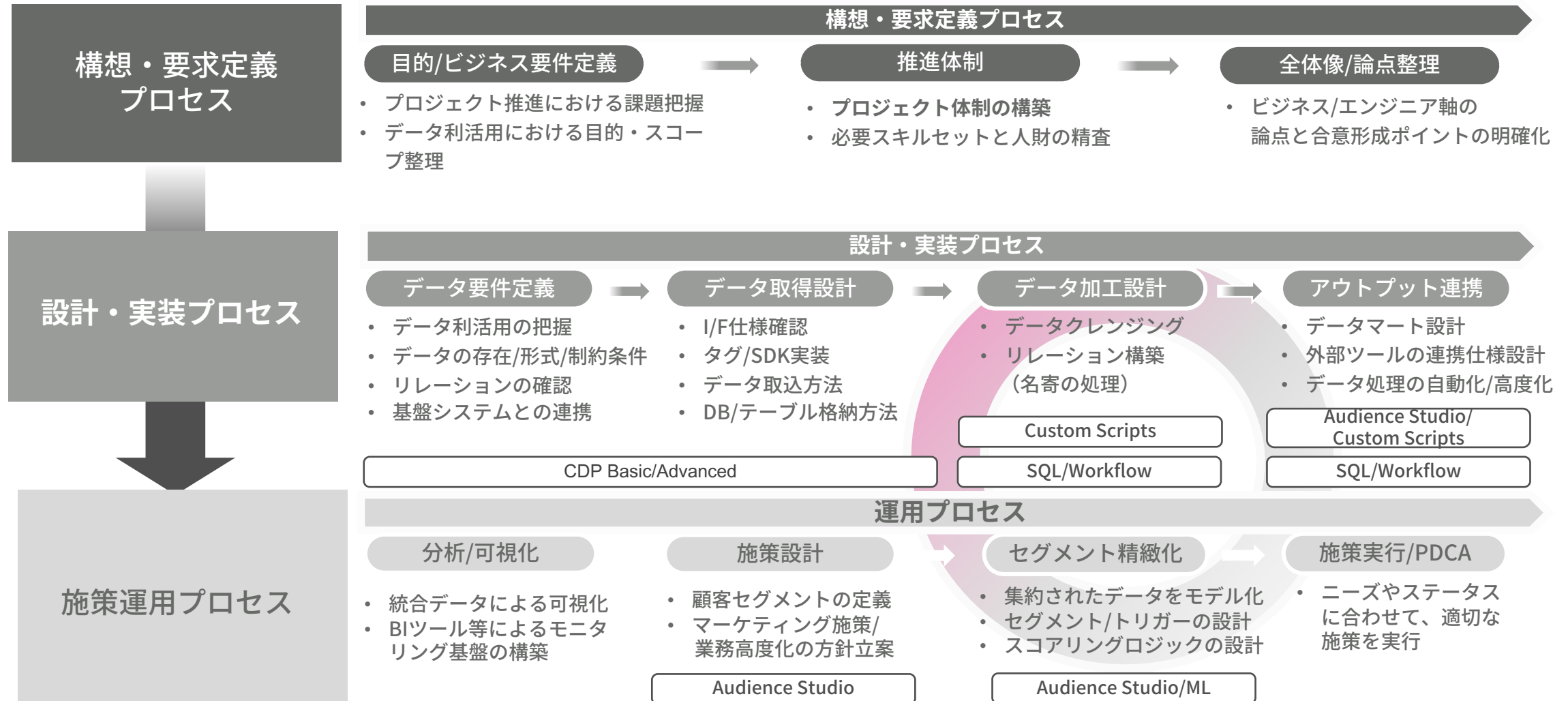
- Treasure Data CDPに保存されたデータにアクセスするためには、データアクセスの標準であるSQLを利用します。このコースではSQLの基礎を学ぶことができます

- Audience Studio

- Treasure Data CDPに蓄積された顧客データを活用するためのツール、オーディエンススタジオについて学びます

CDP Project Leader養成コースの全体像（２）

Project Leader Session 1-6



CDP Project Leader養成コースの全体像（3）

プロジェクトリーダー

エンジニア（聴講可）

Project Leader Session 1-6

構想・要求定義
プロセス

はじめに：イントロダクション

推進体制

全体像/論点整理

Session1：重要要素と課題の把握・理解

Session2：CDP構想・要求定義

設計・実装プロセス

Session3：CDP基盤設計パート

Session4：CDP基盤実装パート

施策運用プロセス

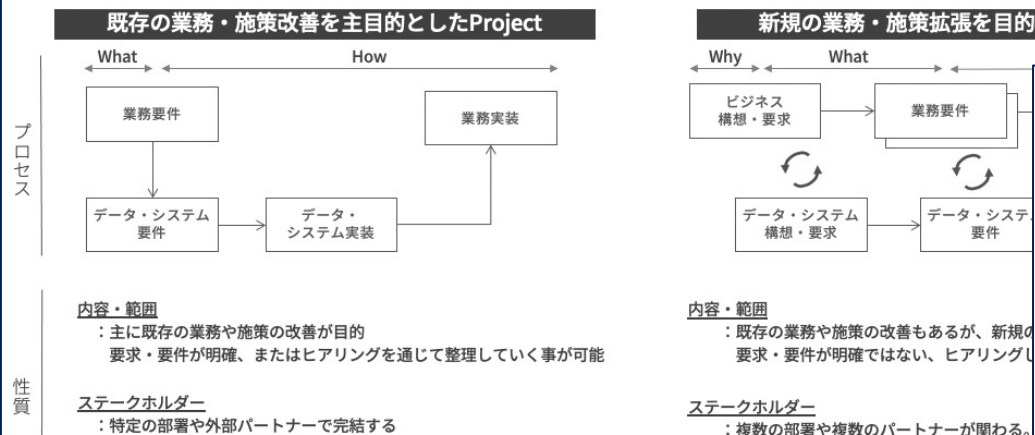
Session5：施策活用パート（可視化・分析）

Session4：施策活用パート（セグメント基盤）

CDP Project Leader Session 1-6 の資料の例

重要要素・課題①PJ設計・推進 | PJ性質の理解不足

- 既存の改善とともに、新規の拡張を目的としているものが多く存在し、プロジェクト性質が異なる
- 今までのプロジェクト設計・推進を行っても、上記の理解がないと上手く進まない



データ要件定義：なぜIDの統合が重要か

- システム/ツールごとにIDが異なり「ひとり」を捉えられない現状を解決するためのTreedata
- 特に「Named」と「Anonymous」との転換ポイントにおけるID統合が重要

ID統合前のコミュニケーション



(参考) 構築スケジュールイメージ

- 大枠の構築スケジュールイメージ

区分	工程	1月			2月			3月			4月			5月			6月		
		上	中	下	上	中	下	上	中	下	上	中	下	上	中	下	上	中	下
マイルストーン	ビジネスイベント																		
	会議体																		
CDP基盤構築	設計																		
	データ・システム把握																		
	データ取得																		
	データクレンジング・加工																		
ダッシュボード構築	設計																		
	データ連携																		
	画面設計・構築																		
	検証・フィードバック																		
活用	本番化																		



リーガルプログラム概要



リーガルプログラム 講師紹介

TMI総合法律事務所 パートナー 大井 哲也 弁護士



2001年弁護士登録。IPO、企業間紛争。クラウド、インターネット・インフラ／コンテンツ、SNS、アプリ・システム開発、アドテクノロジー、ビッグデータアナリティクス、IoT、AI、サイバー・セキュリティの各産業分野における実務を専門とし、ISMS認証機関公平性委員会委員長、社団法人クラウド利用促進機構（CUPA）法律アドバイザー、経済産業省の情報セキュリティに関するタスクフォース委員を歴任する。

【専門領域】

クラウド

- EU一般データ保護規則（GDPR）の適合性評価およびGDPR対応・実装支援
- クラウド導入に伴う個人情報保護法（日本）の適合性評価
- 世界各国グローバルでの個人情報保護法・個人情報保護自主規制・ガイドライン等の適合性評価
- クラウド導入のための安全管理基準適合評価
- クラウド利用規約/SLAのレビュー

サイバー・セキュリティ

- 情報漏えいインシデントのフォレンジック調査・第三者調査委員会の組成
- PCI-DSS、ISMS等のセキュリティ基準適合性評価
- 個人情報保護法対応（日本及びアジア、アフリカ、南米、中東の各地域）
- EU一般データ保護規則対応

アドテク

- DMP（Data management platform）事業における「匿名加工手続き」の個人情報保護法の適法性評価
- 匿名加工情報の再識別化リスク評価
- ユーザを代理したDMP導入のための適法性評価

IoT、AI

- 自動運転車、コネクティッド・カーに関する以下の法的対応
- センシングにより収集されたデータの著作権、データ提供契約作成
- IoTデバイスの脆弱性とIoTデバイス提供事業者の責任
- AIシステムの瑕疵と法的責任

リーガルレクチャー編（Engineer養成コースの方も視聴可）

プロジェクトリーダー

エンジニア（視聴可）

プライバシーやビッグデータ関連法の要点をマーケティング実務者向けに徹底解説

①基礎編

- ビッグデータの利活用事例
- データ入手・加工・第三者提供までのライフサイクル
- 個人情報保護法の解説
- 近時のビッグデータ活用の炎上事例

②応用編

- データエクステンジにまつわる法律問題
- 改正個人情報保護法上のCookie規制
- GDPRなど外国の個人情報保護規制
との比較
- アドテク業界の自主規制
(JIAAガイドライン)

受講者参加型のトレーニングを介して、自社に役立つ知識を体得

③ データマッピング

- データの収集ルートを洗い出そう
- 収集データの保管場所・データ構成を洗い出そう
- データの利活用の方法（アウトプット）を洗い出そう
- 更なるデータの利活用方法を考えてみよう
- データマッピングを踏まえ、発生し得るリスクを分析してみよう

④ プライバシーポリシー作成

- 架空のプライバシーポリシーを参考に、顧客の目線に立って、自分のデータがどのように利用されているかを考えてみよう
- 顧客の納得を得るために、何を説明すべきかを考えてみよう
- 顧客からどのように同意を取得すべきかを考えてみよう
- 顧客が不安を持つ要因を、近時の炎上事例から学ぼう

データマッピングを元に、データライフサイクルや炎上事例、Cookie/端末識別子等の取り扱い、改正個人情報保護法/JIAAガイドライン、プライバシーポリシー作成時の注意点など、プライバシーに関するトピックを網羅的に解説

実践編②: プライバシーポリシーを作成してみよう(事前課題)



- 顧客の目線に立って、自分のデータがどのように利用されているかを考えてみよう

★ 自社のプライバシーポリシー・クッキーポリシーを読み込んでみてください。

- 顧客の納得を得るために、何を説明すべきかを考えてみよう

★ 顧客目線に立った時に、自分が入力したデータや、アプリ・ウェブから自動的に収集されるデータがどのように料理され、提供されるか、自社のプライバシーポリシー・クッキーポリシーから読み取れるか検討してみてください。

- 顧客からどのように同意

★ 長文のプライバシーポリシーを読み、企業からどのようなポイントを説明しては

- 顧客が不安を持つ要因

★ 別添の、近時の炎上事例に関する



データマッピング

Treasure Data CDP



プライバシーポリシー (仮想事例)

個人情報の取り扱いについて (プライバシーポリシー)

はじめに

当社は、お客様のプライバシーを尊重し、お客様の個人データを保護することをお約束します。このプライバシーポリシー（以下「本プライバシーポリシー」といいます。）は、当社がお客様の個人データをどのように管理しているかを説明し、お客様のプライバシーに関する権利についてお知らせするものです。

なお、本プライバシーポリシーで使用される用語の意味については、本プライバシーポリシーの末尾の用語集をご参照ください。

<目次>

- 1. 重要情報及び当社について
- 2. 取得するデータ
- 3. 個人データの取得方法
- 4. 個人データの利用目的
- 5. お客様の個人データの開示
- 6. 越境移転
- 7. データ・セキュリティ
- 8. データの保有期間
- 9. お客様の法的権利
- 10. カメラ画像の取得等
- 11. 用語集

1. 重要情報及び当社について

本プライバシーポリシーの目的

本プライバシーポリシーは、お客様が本アプリ、当社のウェブサイト又は当社のサービスを利用する際に、当社がどのようにお客様の個人データを取得し、これを取り扱うかをお客様にお知らせすることを目的としています。

なお、本アプリは子どもを対象とするものではなく、**子どもに関するデータを意図的に取得するものではありません。**しかし、仮に子どもに関するデータを取得した場合、適用される法令に従って対応いたします。



手法別プログラム概要 (CDP Engineer養成コース)

Treasure Data CDP内でのデータ活用フロー

データを活用して〇〇するために、

データ収集

統合・分析・セグメンテーション

施策への連携

データ活用手順

データ種別



①

どのデータを使うかデータソースを選定

②

各ソースからデータを収集

③

データを使える形に加工・集計

④

ユーザーの塊
=マスターセグメントを作成

⑤

特定ユーザー
=セグメントを抽出

⑥

施策・分析に連携



サイト



広告



MA



プッシュ
通知

Raw
Data

各データソースにて発生した、イベント毎の時系列のデータ

Pre-processed
data

集計/結合を行った加工済みデータ
・マスタ(IDリスト)
・属性
・行動

Master
Segment
Data

Pre-processed Dataを結合し作成したユーザーの母集団
・会員
・アノニマス

Segment
Data

Master Segmentデータから様々な条件によって抽出したデータ

CDP Engineer養成コースの各コースの位置づけ

データを活用して〇〇するために、

データ収集

統合・分析・セグメンテーション

施策への連携

データ活用手順



①

どのデータを使うかデータソースを選定

②

各ソースからデータを収集

③

データを使える形に加工・集計

④

ユーザーの塊
=マスターセグメントを作成

⑤

特定ユーザー
=セグメントを抽出

⑥

施策・分析に連携



サイト



広告



MA



プッシュ通知

Basic / Administrator

Workflow / Custom Scripts

Audience Studio(Engineer)

SQL

Audience Studio(Marketer)

データ種別

Treasure Data CDPを導入するための概要、機能、設定などの基礎知識を学習

概要

対象者：

Treasure Data CDP未経験者向け

ゴール：

CDPの基本的な概念を理解し、運用を考慮しながら基本的な作業が実施できるようになる

概要：

- CDPとは何か
 - データフロー概要
- Treasure Data CDPへのアクセス
- データベースとテーブル
 - データベースのアクセス権限
 - テーブルスキーマ
- データインポート方法
 - Data Connector
- データ加工と分析、およびエクスポート
 - SQLによるデータの加工と分析
 - Result Export

コンテンツ例

arm TREASURE DATA

Treasure Dataとは

特徴
活用ケース
データ分析基盤
Customer Data Platform

利用アクセス方法

Web コンソール
パスワード認証によるログイン
Google SSO認証によるログイン
Google SSOのトラブルシューティング
新規ユーザーの招待について
ユーザーの種類
CLI
CLIの活用用途

データベースとテーブル

データベースの作成
テーブルの作成
データベースとテーブルの命名について
命名規則
命名ガイドライン
デスクリプションについて

データベースへのアクセス権限について

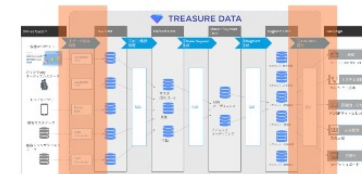
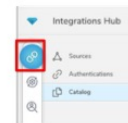
データベースアクセス権限の種類
FULL/READ ONLY/WRITE ONLY/NOTHING
他のユーザーへのアクセス許可方法
Policy Based Permissionの有効の場合
Policy Based Permissionの無効の場合

テーブルスキーマ

スキーマについて
カラム名
カラム名規則
デスクリプション
カラムの別名、物理カラム名と論理カラム名
データタイプ
Int型
String型

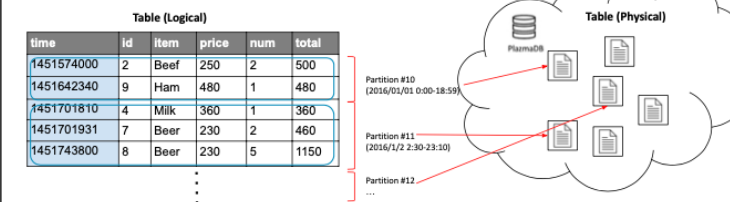
Integrations Hub

- データ収集やデータのエクスポート時の設定に使用する
- さまざまなデータソースとの接続設定やインポート設定をGUIで行う



トレジャーデータの特徴 - Time Index - (1)

- timeカラムによるTime Index
 - テーブルのtimeカラムに格納された「時間でのパーティショニング機能」により、時間に基づいてデータを効率的にアクセスできる構造
 - timeカラムにはUNIX時間（UTCで1970年1月1日午前0時0分0秒からの形式的な経過秒数）を格納



Treasure Data CDPを導入後、安定した運用を行うためのノウハウを習得

概要

対象者：

CDPの運用管理に関わる方

ゴール：

CDP環境の設計・運用責任者として、
管理手法・セキュリティを理解する

概要：

- Treasure Data CDPの管理
- ユーザ管理
 - ユーザ権限
 - 管理者機能
 - その他のユーザ管理
- データセキュリティ管理
 - データ管理ポリシー
 - ハッシュ・暗号化
 - プライベートコネクト
- Audit Log
- システムリソース管理
 - リソースモニタリング
 - リソースプール

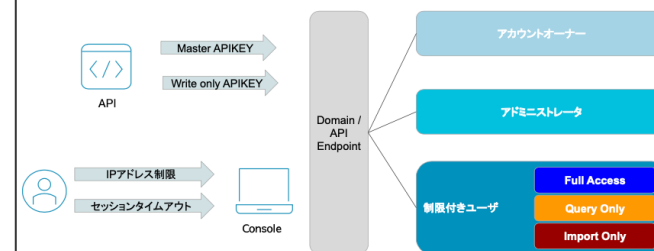
コンテンツ例

運用管理に必要な設定や権限管理、安定的な稼働を実現するために必要なリソース管理やチューニング手法について体系的な学習を行う

arm TREASURE DATA

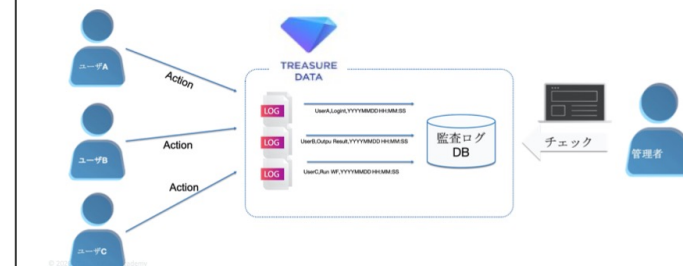
- ◻ Treasure CDPの管理について
 - 管理の目的と必要性
 - データセキュリティ
- ユーザ管理
 - ユーザの種類
 - アカウントオーナー
 - アドミニストレータ
 - 制限付きユーザ
 - 個別およびメーリングリストアカウントの利用
 - 個別およびメー...
 - パスワード認証でのユーザ追加
 - GoogleSSO認証
 - Google SSO認証でのユーザ追加
 - Active Directory/SAML SSO 認証
 - Active Directory / SAML SSO 認証でのユーザ追加
 - パスワードポリシー設定 (DONE)
 - パスワードポリシーのベストプラクティス
 - 複雑なパスワードよりも、長いパスワードを使用する
 - パスワードマネージャを使用する
 - ログインおよびセッションセキュリティポリシーの設定
 - IPホワイトリストの設定
 - デフォルトIPホワイトリスト
 - ユーザIPホワイトリスト
 - IPホワイトリストの使い分け
 - ユーザの作成
 - ユーザの削除
 - 削除ユーザのAPIキーの引き継ぎについて
 - 削除ユーザのデータベース、Authenticationsのオーナー
 - 削除ユーザのスケジュールクエリ
 - 削除ユーザの登録ワークフロー
 - ユーザ権限の設定
 - データベース権限設定
 - ワークフロー権限設定
 - Audience Studio権限設定
 - ポリシーベース権限管理
 - Policy Based Permissionとは

ユーザ管理とアクセスコントロール



監査ログ (Audit Log) によるアクセスの記録

アカウントに紐づく全てのユーザのアクティビティの詳細な監査証跡をCDP内に保管し、確認するための機能



SQL基礎編 1～3 コースのスコープ

プロジェクトリーダー※

エンジニア

セグメント作成の出発点になるSQLを実践を交えながら理解する

概要

対象者：

Treasure Data CDPに関わる全ての方がSQLの知識をSQL基礎編1、2、3の3ステップに分けて学習

SQL基礎編1

対象: 初学者:エンジニア/Project Leader

- ゴール：CDP上のデータをSQLで扱えるようになる
- SELECT、WHERE、JOIN、GROUP BYによるデータ集計

SQL基礎編2,3

対象: エンジニア

- ゴール：テキストを参考に自分で学習が進めるようになる
- Treasure Data独自関数の理解
- WINDOW関数、正規表現などを用いる高度なクエリ

コース内容

- 実践的なデータセットを豊富に用意
- 文字列関数・正規表現・数学関数などのリファレンスの活用
- 重要な Treasure Dataの DATE/TIME 関数の徹底理解

Lesson 2. データセット紹介

アクセスログ

sample_accesslog: TD Webサイト

Treasure JS SDK によって収集されるアクセスログを Web ログ分析のためのサンプルデータとして提供します。項目と内容については次のようになっています。

カラム名	概要	サンプル値
time	アクセス時のタイムスタンプ	1466409507
td_client_id	ユーザー識別子	5276cd9-6991-47bc-b9b7-ccb6ac100920
td_title	ページタイトル	Treasure Data - データ分析をクラウドで、シンプルに。
td_browser	ブラウザ種別	Safari
td_host	ホスト	www.treasuredata.com
td_path	パス	/jp/
td_url	URL	https://www.treasuredata.com/jp/
td_referrer	リファラ	http://search.yahoo.co.jp/search?e=
td_ip	IPアドレス	124.32.162.10
td_os	OS	Mac OS X
td_language	言語	jp-jp

今後のレッスンを進めていく上で、いくつか注意点を挙げておきます。

- 「time」は timestamp 型と言われ、直感的な日時表現では無いこと
- 「td_client_id」はユーザーを区別するための識別ID
- 「td_url」は「td_host」や「td_path」に比べ、余計な文字列（パラメータなど）がついて雑多な文字列になること
- 「td_referrer」も雑多な文字列になるが、自社サイト外からの流入を知る意味で重要なカラムであること

基礎2-3-3.

TD_TIME_ADD: 親しみやすい時間の加減算

- time カラムの値が「2019-12-12 09:30:13」の unix_timestamp の場合

```
SELECT
  TD_TIME_FORMAT( time, 'yyyy-MM-dd HH:mm:ss', 'JST'),
  TD_TIME_FORMAT( TD_TIME_ADD( time, '1d', 'JST'), 'yyyy-MM-dd HH:mm:ss', 'JST'), --1日後
  TD_TIME_FORMAT( TD_TIME_ADD( time, '-1d', 'JST'), 'yyyy-MM-dd HH:mm:ss', 'JST'), --1日前
  TD_TIME_FORMAT( TD_TIME_ADD( time, '1h', 'JST'), 'yyyy-MM-dd HH:mm:ss', 'JST'), --1時間後
  TD_TIME_FORMAT( TD_TIME_ADD( time, '-1h', 'JST'), 'yyyy-MM-dd HH:mm:ss', 'JST'), --1時間前
  TD_TIME_FORMAT( TD_TIME_ADD( time, '30m', 'JST'), 'yyyy-MM-dd HH:mm:ss', 'JST'), --30分後
  TD_TIME_FORMAT( TD_TIME_ADD( time, '-30m', 'JST'), 'yyyy-MM-dd HH:mm:ss', 'JST'), --30分前
  TD_TIME_FORMAT( TD_TIME_ADD( time, '10s', 'JST'), 'yyyy-MM-dd HH:mm:ss', 'JST'), --10秒後
  TD_TIME_FORMAT( TD_TIME_ADD( time, '-10s', 'JST'), 'yyyy-MM-dd HH:mm:ss', 'JST'), --10秒前
FROM ( VALUES 1576110613 ) AS t
```

Window 関数の肝

1. PARTITION BY: 区間分け (GROUP BY と類似)

2. ORDER BY: 並び替え

3. RANGE/ROWS BETWEEN

: 範囲指定

		1. PARTITION BY user_id		2. ORDER BY id		3. RANGE/ROWS BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW	
user_id	id	LAG(order, 1)	sales	SUM(sales)			
a	1	NULL	100	100			
a	2	1	200	300			
a	3	2	300	600			
a	4	3	200	800			
b	1	NULL	1000	1000			
b	2	1	500	1500			
c	1	NULL	300	300			
c	2	1	300	600			

Workflow / Custom Scripts 概論のスコープ

Treasure Workflowを用いたデータ加工、処理の定期実行や高度な自動化手法、およびPythonを用いた高度なWorkflow (Custom Scripts) の用途、利用方法を学習する

エンジニア

概要

対象者：

Treasure Data CDPに関わる全てのエンジニア

ゴール

Treasure WorkflowとCustom Scriptsのコンセプトを理解し、利用できるようになる

概要：

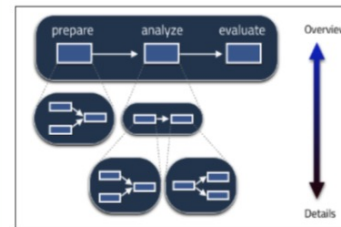
- Treasure Workflowの概要
- Treasure Workflowの記法
- ハンズオン演習
- Custom Scriptsの概要
- 基本的なCustom Scriptsの使い方
- ハンズオン演習

コンテンツ例

Treasure Workflowのユースケース1

定期的、定型的なSQL実行

- ・ データロードから利用可能なデータまでの一連の処理の自動化
- ・ 複数のワークフローをつなげて処理することが可能
- ・ トレジャーデータ自体もワークフローを最大限に活用 (例: Audience Studio)



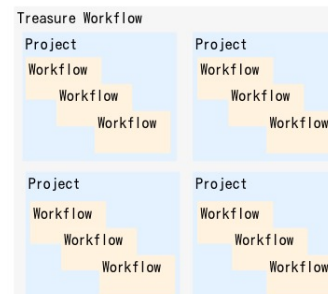
プロジェクト、ワークフロー(P8)

プロジェクト

- ・ Treasure Workflowの実行に必要なファイルをまとめた概念
- ・ ひとつのプロジェクトに複数のワークフローを配置できる

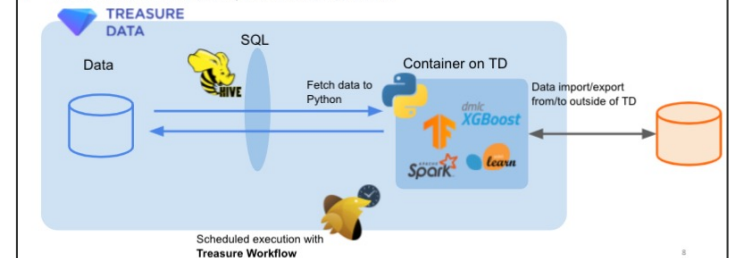
ワークフロー

- ・ Treasure Workflow上で動作させる一連の処理のこと、またその処理を定義したファイルのこと
- ・ 定義ファイルは .dig という拡張子で保存するため、Digファイルとも呼ばれる



Custom Scriptsとは

- ・ TDのホストされた環境でPythonを実行可能
- ・ Treasure Workflowを使いPythonを定期実行できる

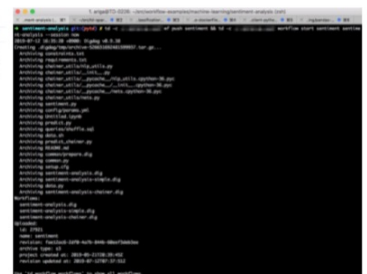


Custom Scriptsの使い方

TD Console (Webブラウザ)



TD Toolbelt (ターミナル)



Treasure Workflowを本番環境で利用するために必要となる実践的な知識を習得する

概要

対象者：
概論コースを受講済みで、実際に
Workflowを記述、デバッグ、運用する
エンジニア

ゴール：
Workflow / Custom Scripts概論を理解したエンジニアが、実環境で利用する
様々なWorkflowを開発できるようになる

概要：

- Treasure Workflowとは（概論の復習）
- Treasure Workflowの構成要素
- Workflowベストプラクティス 1、2
- ハンズオン演習

コンテンツ例

TDのテーブルを読み込むワークフローをつくる

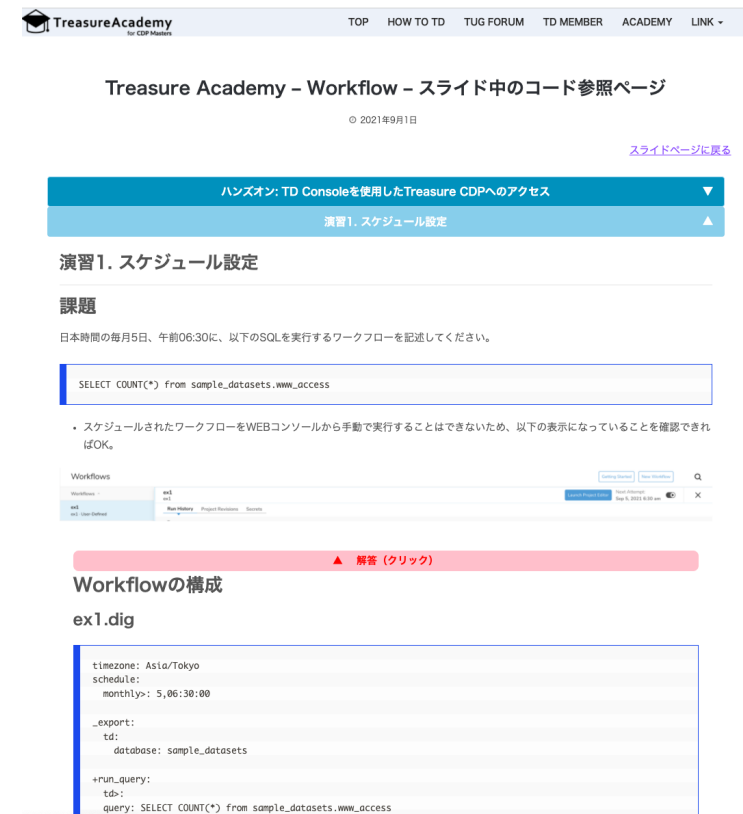
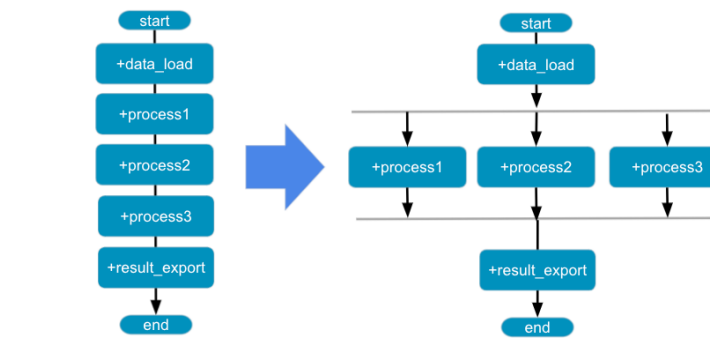
Custom Scriptを使ってデータを読み込み、ログにprintするタスクを作成しよう

以下のようなワークフローと、方向性のPythonスクリプトを作成するのが目標
pandas-df.dig → py_scripts/examples.py

```
+read_into_df:
  py: py_scripts/examples/read_td_table
  database_name: sample_datasets
  table_name: pandas-df.dig
doc:
  1. ファイル名、関数名を py> オペレータで指定
  2. TD_API_KEY: ${secret:td.apikey}
  3. TD_API_SERVER: "https://api.treasuredata.com"

+read_td_table(database_name, table_name):
  import os
  apikey = os.environ["TD_API_KEY"]
  apiserver = os.environ["TD_API_SERVER"]
  # TDのデータを読み込みprintするための関数
  some_func_to_read_td(database_name, table_name)
```

parallel を利用した並列処理



Workflow 演習のスコープ

Treasure Workflowを本番環境で利用するために必要となる実践的な知識を習得する
説明部分は動画で提供し、講義ではハンズオン演習の解説を行う

エンジニア

概要

対象者：
概論コースを受講済みで、実際に
Workflowを記述、デバッグ、運用する
エンジニア

ゴール：
Workflow / Custom Scripts概論を理解したエンジニアが、実環境で利用する
様々なWorkflowを開発できるようになる

- 概要：
- Treasure Workflowとは（概論の復習）
 - Treasure Workflowの構成要素
 - Workflowベストプラクティス 1、2
 - ハンズオン演習

コンテンツ例

TDのテーブルを読み込むワークフローをつくる

Custom Scriptを使ってデータを読み込み、ログにprintするタスクを作成しよう

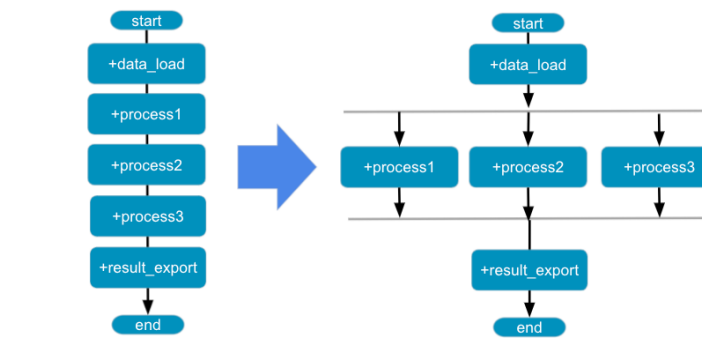
以下のようなワークフローと、方向性のPythonスクリプトを作成するのが目標
pandas-df.dig → py_scripts/examples.py

```
+read_into_df:
py> py_scripts/examples.read_td_table
database_name: sample_datasets
table_name: pd
doc:
  ファイル名、関数名を py> オペレータで指定
_env:
  TD_API_KEY: ${secret:td.apikey}
  TD_API_SERVER: "https://api.treasuredata.com"

import os
read_td_table(database_name, table_name):
  apikey = os.environ["TD_API_KEY"]
  apiserver = os.environ["TD_API_SERVER"]
  # TDのデータを読み込みprintするための関数
  some_func_to_read_td(database_name, table_name)
```

51 © 2020 Treasure Academy

_parallel: を利用した並列処理



Treasure Academy - Workflow - スライド中のコード参照ページ

© 2021年9月1日

[スライドページに戻る](#)

ハンズオン: TD Consoleを使用したTreasure CDPへのアクセス

演習1. スケジュール設定

演習1. スケジュール設定

課題

日本時間の毎月5日、午前06:30に、以下のSQLを実行するワークフローを記述してください。

```
SELECT COUNT(*) from sample_datasets.www_access
```

・スケジュールされたワークフローをWEBコンソールから手動で実行することはできないため、以下の表示になっていることを確認できればOK。

Workflows	edit	run
td1	Run Workflow	Project Workflow
td2	Run Workflow	Project Workflow

▲ 解答 (クリック)

Workflowの構成

ex1.dig

```
timezone: Asia/Tokyo
schedule:
  monthly>: 5,06:30:00

_export:
  td:
    database: sample_datasets

+run_query:
  td:
    query: SELECT COUNT(*) from sample_datasets.www_access
```

Custom Scriptsを本番環境で利用するために必要となる実践的な知識を習得する

概要

対象者：
概論コースを受講済みで、実際に
Custom Scriptsを記述、デバッグ、運
用するエンジニア

ゴール：
Workflow / Custom Scripts概論を理
解したエンジニアが、実環境で利用す
る様々なCustom Scriptsを開発できる
ようになる

- 概要：
- Custom Scriptsとは（概論の復習）
 - pytdとは
 - ハンズオン演習

コンテンツ例

py>オペレータの意味

```
py>: py_scripts.examples.print_arg
```

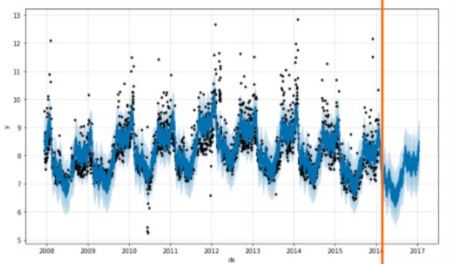
ディレクトリ名 ファイル名 関数名
 examples.py

```
simple.dig
+simple_with_arg:
py>: py_scripts.examples.print_arg
msg: "Hello World"
docker:
image: "digdag/digdag-python:3.9"
```

```
py_scripts/examples.py
def print_arg(msg):
print(f"Message is {msg}")
```

Prophetの予測の例

黒丸が実測値、青が予測値、水色が不確実性の幅



周期性を考慮した時系列を
予測している



Audience Studio (for Marketer/Engineer) コースのスコープ

Audience Studioの機能全般を理解し、利用者としての使い方および
Audience Studioを利用できる環境を構築するエンジニア向けの講義を提供

プロジェクトリーダー

エンジニア

概要

対象者:

Treasure Data CDPを利用するエンジニア、プロジェクトリーダー、マーケター、その他全ての方

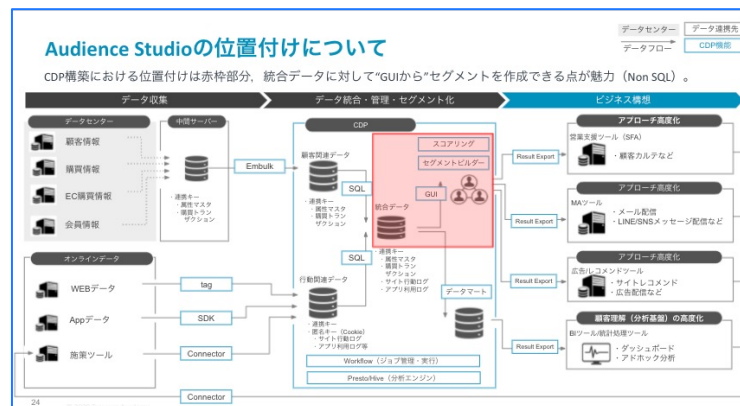
ゴール:

Audience Studioの役割を理解し、Audience Studioを用いて何ができるのかを理解、実践できるようになる

概要

- CDPにおける立ち位置の理解
- Audience Studioの機能紹介
- ハンズオン演習

コンテンツ例

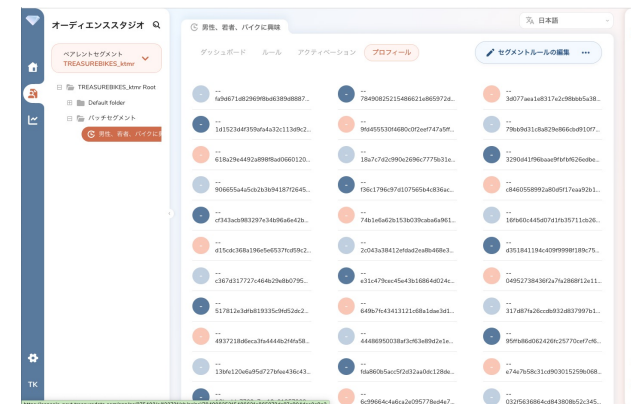
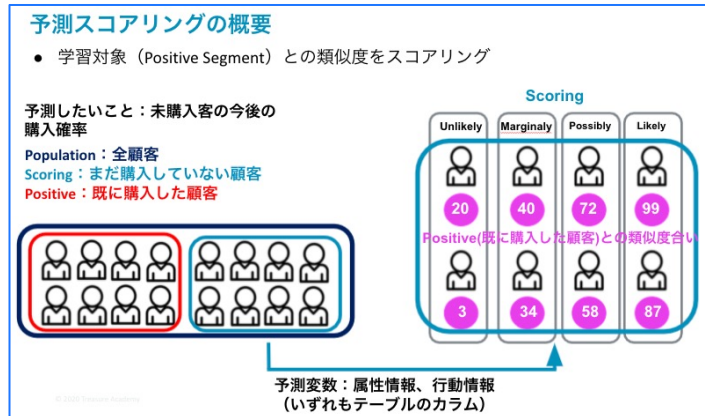


「マスターセグメント」：1つのユーザーリストと、それに紐づく複数の属性テーブル・行動テーブルとの紐付き関係を定義したもの



なぜマスターセグメントを作成するのか？

- データ統合の結果は、必ずしも一枚のテーブルに集約して保持されるものではない、以下の構成
 - IDのマッピングテーブル（次ページ）
 - 前処理・整形された各種テーブル（3テーブル）
- データ統合の結果、データを処理する際に「容易にデータを繋ぐことができる」状態になる



Treasure Dataが提供するAutoMLの機能を理解する

概要

対象者:
Treasure Data CDPを利用するエンジニア、プロジェクトリーダー、マーケター、その他全ての方

ゴール:
Treasure Data AutoMLの役割を理解し、TD AutoMLを用いて何ができるのかを理解、実践できるようになる

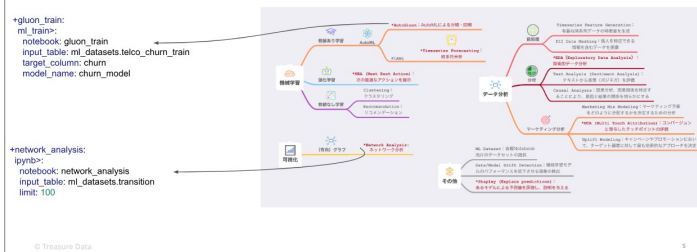
概要

- TD AutoMLとは何か
- Notebookについて
- 各Notebookの機能

コンテンツ例

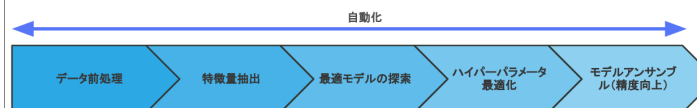
TD AutoML とは？

Workflow で専用のオペレータと簡単な記述で、AutoML などの機械学習にとどまらず、可視化や分析などの様々なソリューションが利用できる仕組み
ソリューションごとに専用 Notebook が用意され、今後も Notebook が増えることで、できることが増えていく



AutoGluon とは？ AutoML とは？

AWS が開発している AutoML (Automated Machine Learning) のツールキット。データ前処理、ハイパーパラメータチューニング、モデルアンサンブルまでの全てを自動化。AutoML は、機械学習の専門家でなくても、機械学習のモデルを作成できるようにすることを目的としている。



- AutoGluonの適用範囲 (TD AutoMLではテーブルデータのみ対象)
 - テーブルデータ: 解約予測 (分類)、スパム判定 (分類)、売上予測 (回帰)、スコア予測 (回帰)
 - 画像分類: 商品の分類、異常画像の判定
 - 物体検出: 写真中の人・動物・車の分類
 - 文章分類: レビューの分類、フィードバックの分類

1. TD AutoML の記述方法

TD AutoML 用の WF オペレータを使用し、必要なパラメータとその値を定義するだけ
SQLやPythonのコードらしいコードの記述は基本必要ない

```
+train_task:  
ml_train>
```

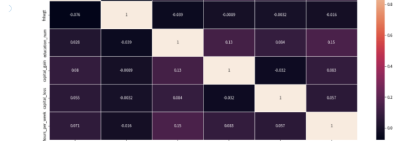
```
notebook: gluon_train  
input_table: input_db.train_data  
target_column: churned  
model_name: churn_model
```

AutoGluon automl Notebook を使用する場合の例

Basic EDA

```
SELECT * FROM train_data LIMIT 1000  
--> 1000 rows
```

```
SELECT * FROM train_data LIMIT 1000  
--> 1000 rows
```



相関関係の可視化(ヒートマップ)

コース終了認定証の発行

コース終了後、修了テスト合格者へ発行（Partner向けには、パートナー制度と連動）



有効期限: 2 年間



アドバンスプログラム

レギュラープログラム修了レベルのユーザーを対象とするトレーニングと認定資格を提供

背景

CDPの知識を持つエンジニア、プロジェクトリーダーの育成が進んできたが、具体的な内容を求める声が増加

➡CDPの知識は得たが、具体的に「CDPの環境の**実装**」の方法を学びたいエンジニアの増加

➡CDP導入後、蓄積された顧客データをフルに「**活用**」するための体系的スキルが必要



施策

対応するアドバンストレーニングと認定資格を提供

トレーニング



シナリオに沿って行うトレーニングと演習

- CDP実装
- オーディエンススタジオ実装/活用

認定資格



「ハンズオン」形式の認定資格試験による、実践的なスキルを証明する認定資格を提供

アドバンスプログラム（1）

目的に特化したトレーニングコースをご用意



エンジニア

CDP Implementation （CDP実装）コース

- 顧客データをCDPに収集し、加工、施策連携するために必要となる、CDP実装の知識、手法を豊富なハンズオンを通して学習する



エンジニア

利用者

Audience Studio コース

- CDPの利用者がオーディエンススタジオを使いこなし、CDPに格納された顧客データを最大限活用するためのノウハウを提供。また、エンジニア向けにオーディエンススタジオが利用するDBやワークフローについての詳細動画も提供



エンジニア

TD AutoMLコース (2023年秋頃提供予定)

- Treasure Data AutoMLの各機能を理解し、ハンズオン演習を通してそれぞれのノートブックが使いこなせるようになる。また、ノートブックに出力される学習、予測フェーズや分析結果の読み方を学ぶ

アドバンスプログラム（２）

目的に特化したトレーニングコースをご用意



エンジニア

利用者

Journey Orchestration コース (2023年秋頃提供予定)

- マーケターでも直感的に利用できるJourney Orchestrationを用いカスタマージャーニーにおける施策を行う方法を学習する
- オーディエンススタジオの知識を持っていることを推奨



エンジニア

ID Unification コース (2023年秋頃提供予定)

- CDPに収集したさまざまなデータソースを用いてID統合を行うために、Treasure Dataが提供するID Unification機能の使い方を学習する
- 事前に基本的なTD Workflowの知識があることを推奨

アドバンスプログラムは3つのトレーニングコースを用意。それぞれ異なる認定資格を提供

- **CDP（データプラットフォーム）サイドのコース（3種類）**

- 1. CDP Implementation Course（通称：CDP 実装コース）[エンジニア向け]
 - このコースでは、CDPの実装を目的とし、さまざまなデータソースからのデータ取得、名寄せやクレンジングなどのデータ加工、連携先への出力が実装できるようになる人材育成を目指す
- 2. ID Unification Course [エンジニア向け]
 - CDPに収集したさまざまなデータソースを用いてID統合を行うために、Treasure Dataが提供する ID Unification機能の使い方を学習する
- 3. TD AutoML Course [エンジニア向け]
 - Treasure Data AutoMLの各機能を理解し、ハンズオン演習を通してそれぞれのノートブックが使用できるようになる

- **CDP（オーディエンススタジオ）サイドのコース（2種類）**

- 4. Audience Studio Course [エンジニア・利用者向け]
 - オーディエンススタジオの全ての機能の基礎知識と活用方法を知り、CDP（オーディエンススタジオ）とマーケティングを最大限活用できる人材育成を目指す
 - また、オプションではあるがオーディエンススタジオのバックエンド（データ、Workflow、挙動）を理解するための動画を提供
- 5. Journey Orchestration Course [エンジニア・利用者向け]
 - マーケターでも直感的に利用できるJourney Orchestrationを用いカスタマージャーニーにおける施策を行う方法を学習する

ハンズオン形式の試験により「実践力」を証明する認定資格



Certified CDP Specialist:

- Treasure Data CDPを用いたハンズオン形式の認定資格試験
- 与えられた課題を期間内に実施する
- 認定資格の取得者は、Treasure Data CDPの知識に加えて実践的なスキルを持っていることが証明される

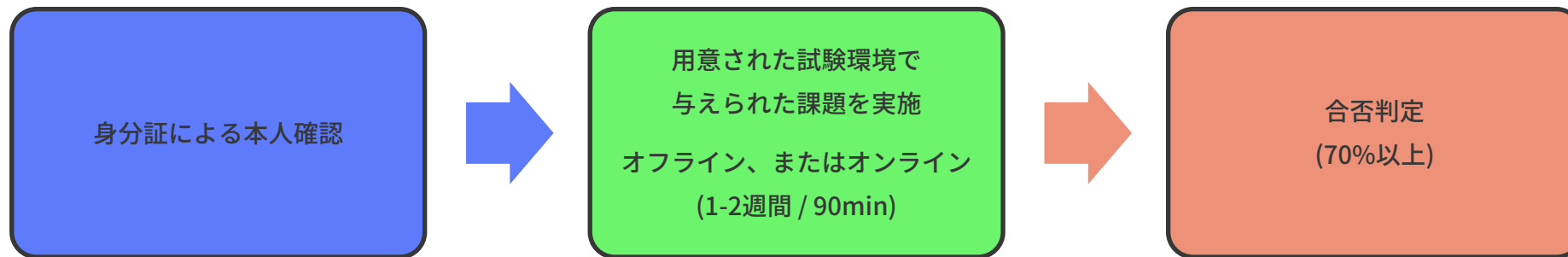
※認定資格の有効期限は2年間

注) 試験の手順、試験時間などは
認定資格試験ごとに異なります

- ハンズオン形式の試験

- エンジニアは、実際に「手を動かせる」ことを評価する、ハンズオン形式の認定資格試験
- ビジネス開発者は、データ活用できることを評価するハンズオンとヒアリング形式の認定資格試験

- ハンズオン試験のイメージ



課題例

SFTPサーバから顧客データをCDPに取り込む

環境についての情報： SFTP server: sftp.example.com、DB name: l0_xxx, Table name: yyy

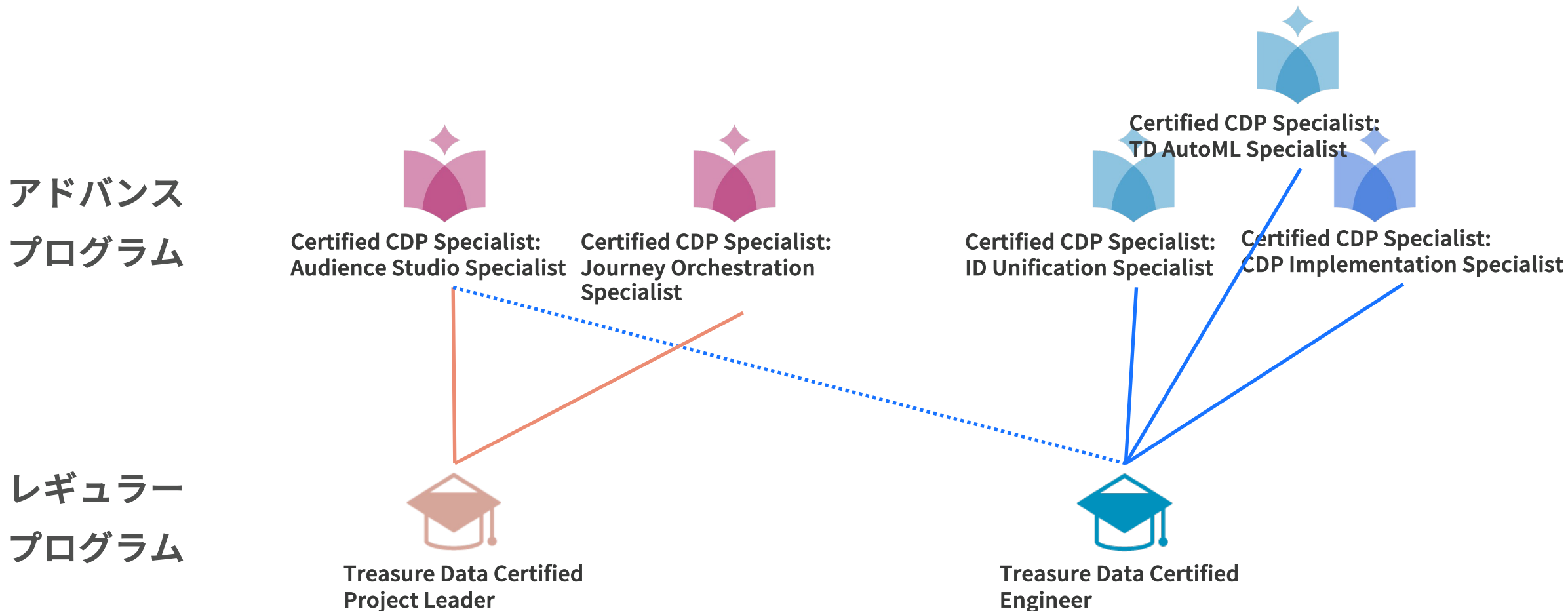
Treasure Workflowを使用して、データ収集から加工、出力までを自動化する

ペアレントセグメント（マスターセグメント）の構築

指定したルールでセグメントの作成

アドバンスプログラムの認定資格体系

アドバンスプログラムの各コースの認定資格はレギュラープログラムの上に位置づけられ、アドバンスプログラムの認定資格取得者は「Certified CDP Specialist ***」として認定される



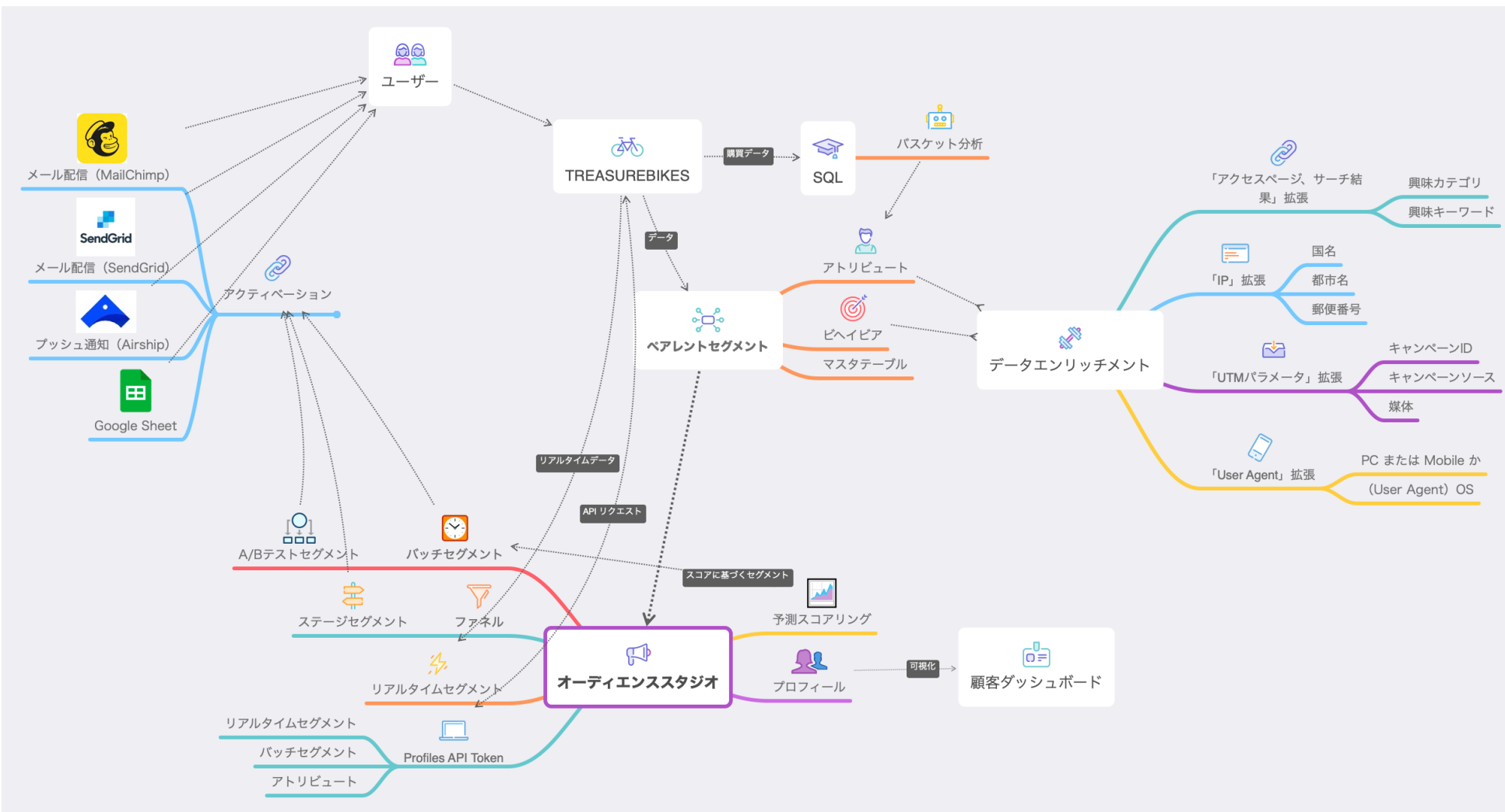
※認定資格の有効期限は2年間



Audience Studio コース紹介

オーディエンススタジオの様々な機能を習得する

オーディエンススタジオには様々な機能があり、TREASUREBIKES という実サイトのデータをもとに実践ベースで習得を行う



TREASUREBIKES ウェブサイト - 使用するデータセット

<https://treasurebikes.com/>

自転車本体・グッズの販売や宣伝を行う擬似 Web+EC サイト



BICYCLES

HELMETS

ACCESSORIES

BLOG

REGISTER



Ride Bikes.
Feel Good.

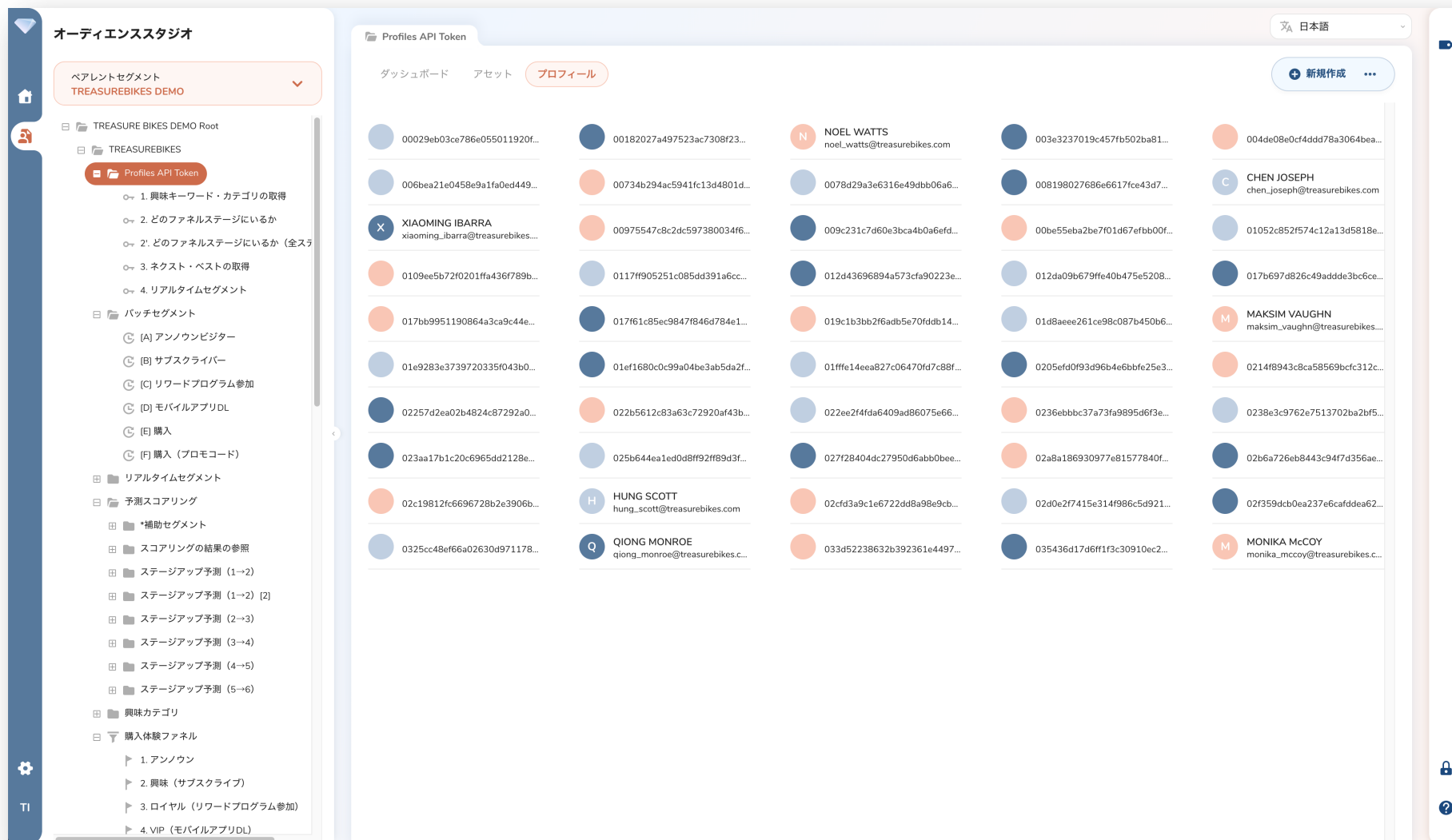
LEARN MORE

BIKES



オーディエンススタジオ アカウント

各自オーディエンススタジオのアカウントが与えられ、自身のアカウントにデモ（お手本）アカウントと同じオーディエンススタジオの構築を行う「実践ベース」の講義を展開



The screenshot displays the Audience Studio interface. On the left is a sidebar menu with a home icon, a user icon, and a settings icon. The menu is organized into sections: 'ペARENTセグメント' (Parent Segments) with a dropdown for 'TREASUREBIKES DEMO'; 'TREASURE BIKES DEMO Root'; 'TREASUREBIKES'; 'Profiles API Token' (highlighted in orange); a list of tasks (1. 興味キーワード・カテゴリの取得, 2. どのファネルステージにいるか, 2'. どのファネルステージにいるか (全ス), 3. ネクスト・ベストの取得, 4. リアルタイムセグメント); 'バッチセグメント' (Batch Segments) with sub-items (A) アンノウンビジター, (B) サブスクライバー, (C) リワードプログラム参加, (D) モバイルアプリDL, (E) 購入, (F) 購入 (プロモコード); 'リアルタイムセグメント'; '予測スコアリング' (Predictive Scoring) with a sub-item '補助セグメント'; 'スコアリングの結果の参照'; 'ステージアップ予測' (1→2), (1→2) [2], (2→3), (3→4), (4→5), (5→6); '興味カテゴリ'; and '購入体験ファネル' (Purchase Experience Funnel) with sub-items 1. アンノウン, 2. 興味 (サブスクライブ), 3. ロイヤル (リワードプログラム参加), and 4. VIP (モバイルアプリDL). The main content area is titled 'Profiles API Token' and has tabs for 'ダッシュボード', 'アセット', and 'プロフィール' (highlighted in orange). It shows a list of profiles with columns for a profile ID, a name, an email address, and a profile ID. The profiles listed include NOEL WATTS, XIAOMING IBARRA, HUNG SCOTT, and QIONG MONROE. The interface also includes a language selector (日本語) and a '新規作成' (New Creation) button.

本プログラムでは、講義は「オリエンテーション」+「本講義」の2日間、または3日間構成

事前課題を本講義までにこなしておくことが前提となる

エンジニアに必要なマスターセグメントの作成、バックエンドDBやワークフローについては動画で講義を提供

試験の受験は任意

オリエンテーション [3h]

実施内容

- ・ アドバンスプログラム紹介
- ・ 「オーディエンススタジオ 活用方法」での講義
- ・ 事前課題の説明



ポイント

- ・ オーディエンススタジオの活用例を知り、それを実現するための各機能の理解のモチベーションとする

事前課題

実施内容

- ・ 自身のアカウントに各機能（の一部）を実装



ポイント

- ・ 各機能ごとに「スライド」+「ビデオ」を参照しながら、自身のアカウントにデモアカウントと同じ機能を実装していく

本講義 [3h x 2 or 6h]

実施内容

- ・ オーディエンススタジオの各機能の説明や使い所を解説



ポイント

- ・ 事前課題で行ってもらったことの解説

認定試験 [オフライン]

実施内容

- ・ オーディエンススタジオのオフラインハンズオン試験（約2週間）



ポイント

- ・ 指定した時間内で実装・記述問題を解いてもらう

2週間

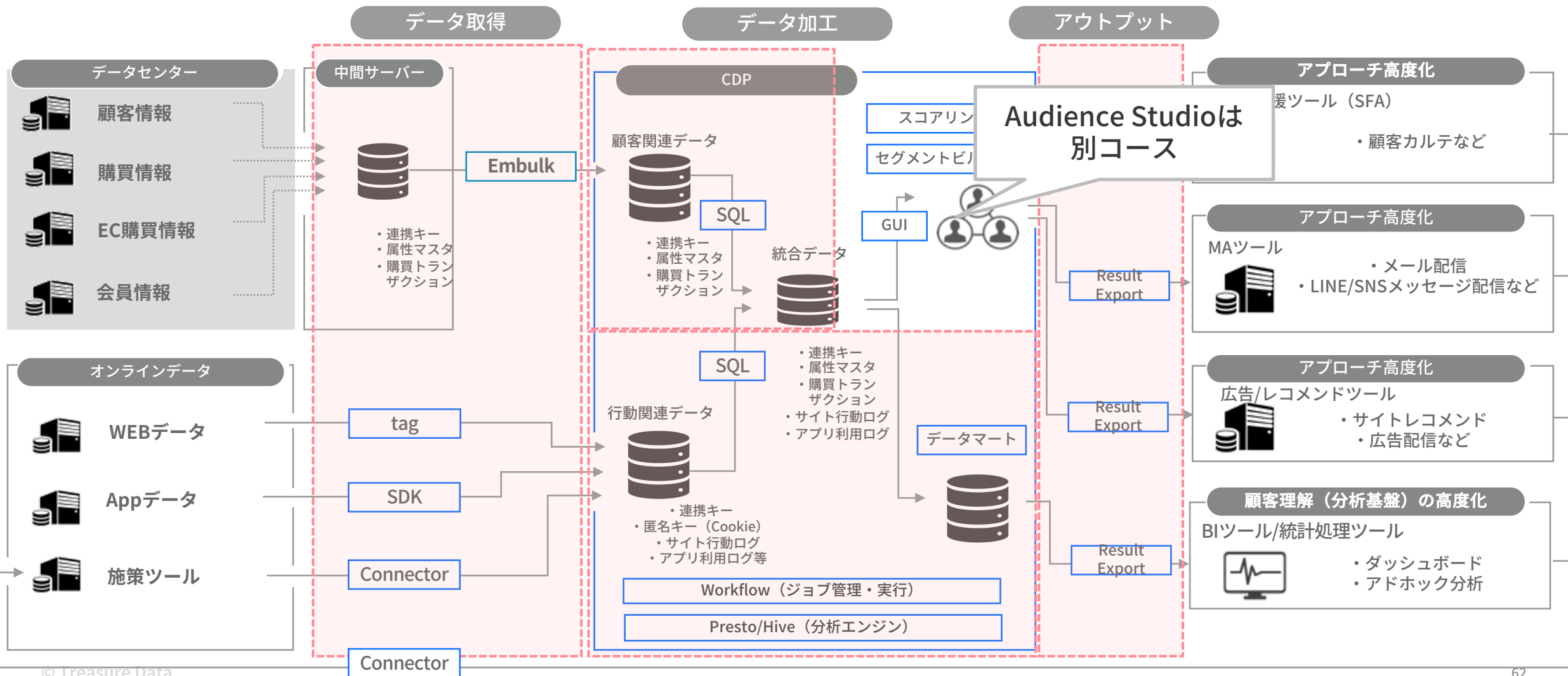
2週間



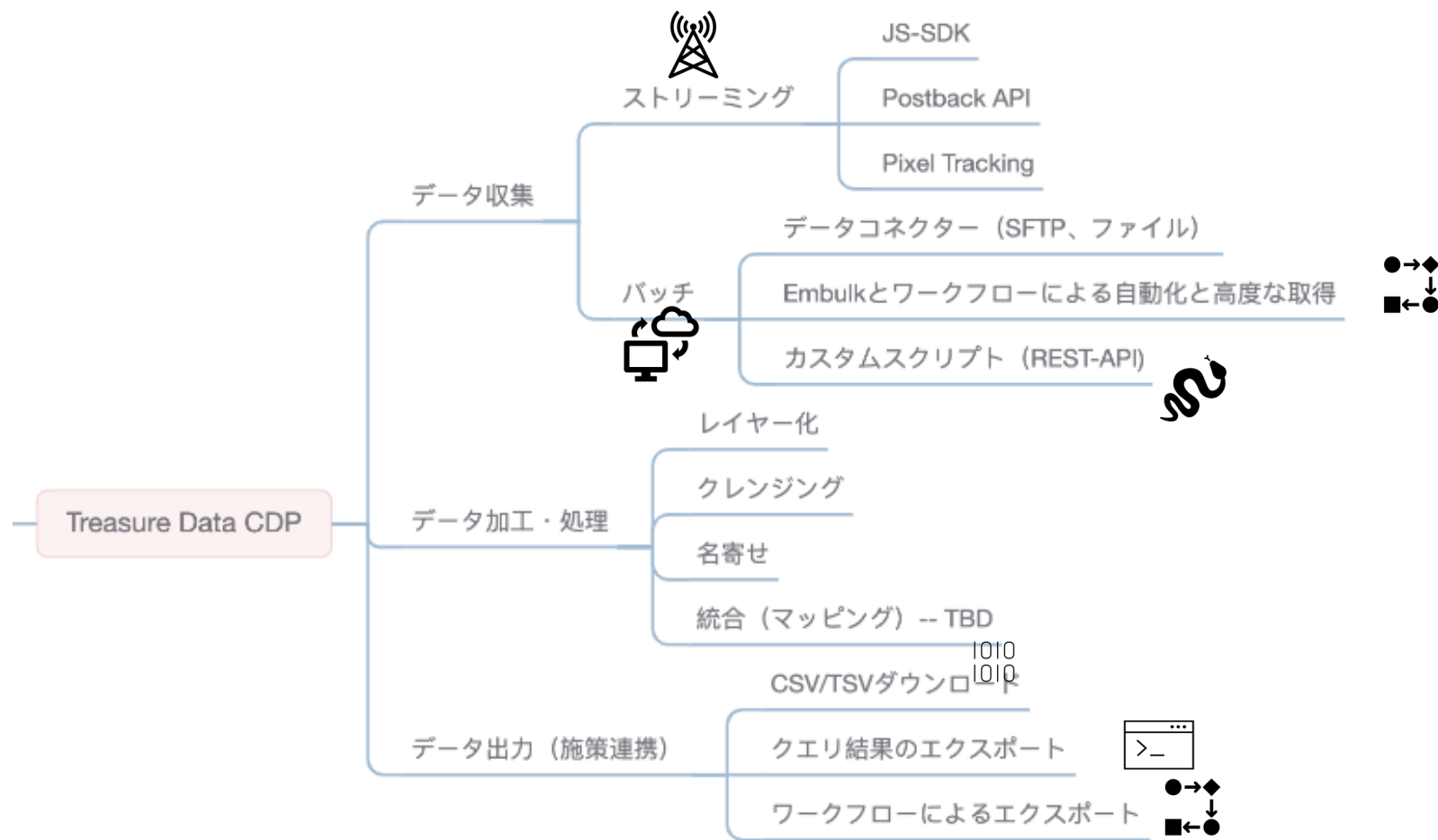
CDP実装コース

CDP基盤構築に必要なスキルを獲得する

ハンズオン演習を中心とした実践形式のトレーニングでCDP基盤構築に必要なスキルを習得する



CDPのデータフローの3つのフェーズ、「データ収集」「データ加工」「データ出力」を網羅



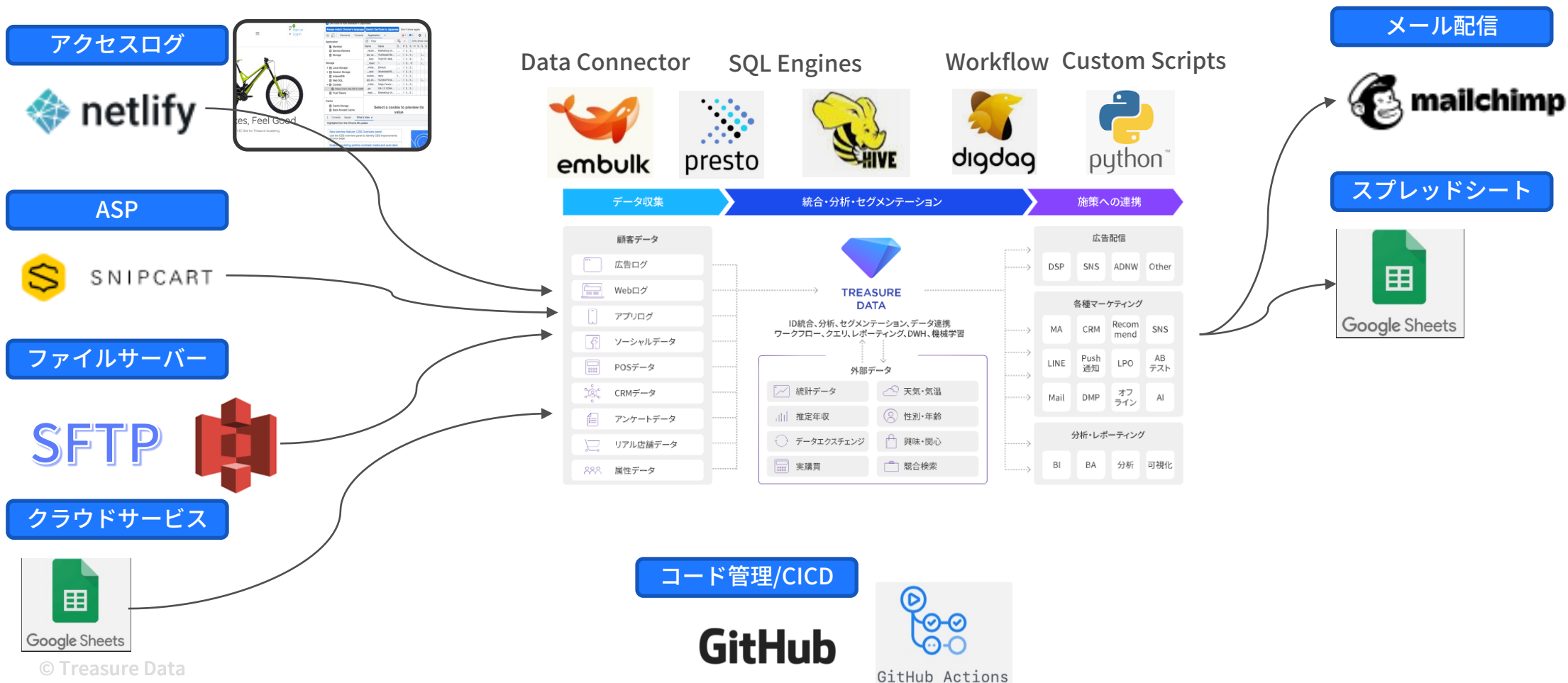
CDP実装コース - 演習で構築する環境

本コースで構築するCDP環境（データソース、データ加工/処理で使用するコンポーネント、アウトプット連携先）

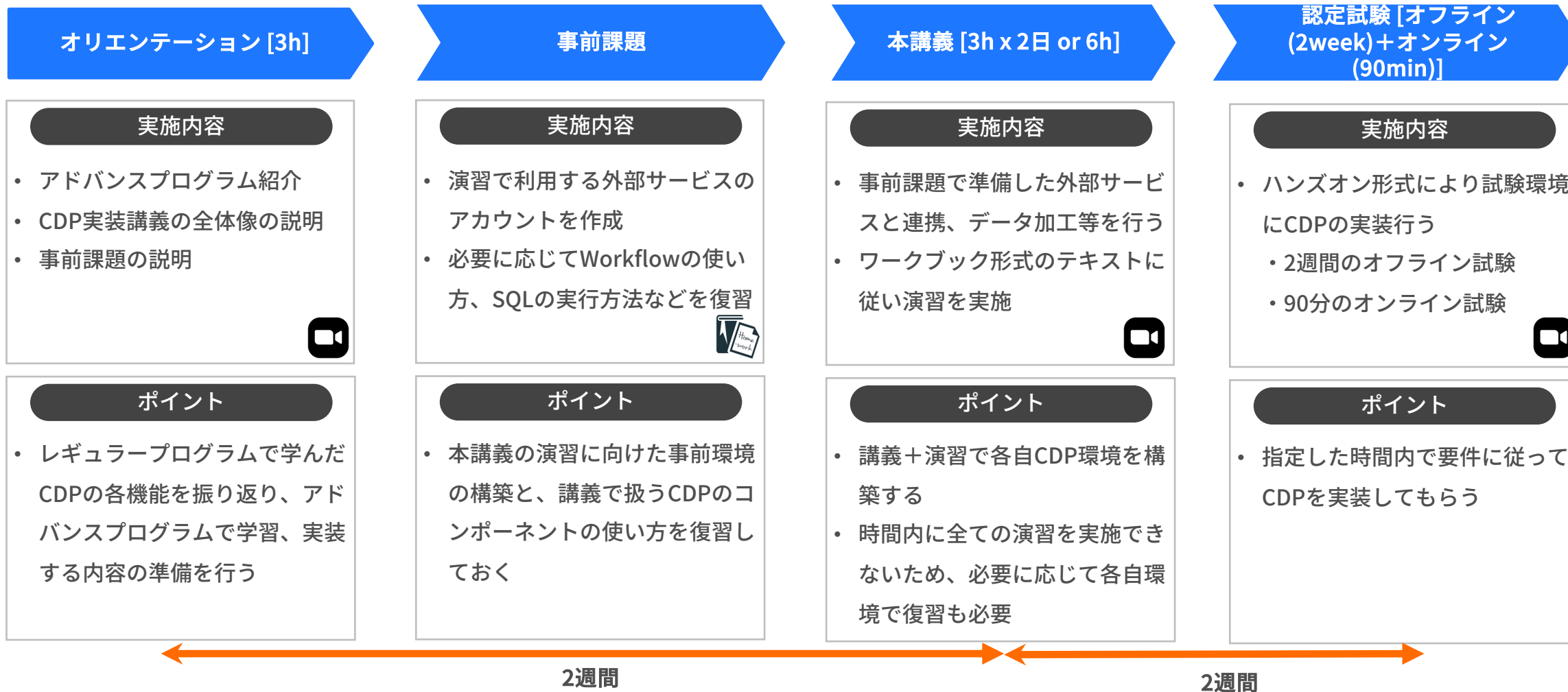
データソース

CDP基盤

アウトプット連携先



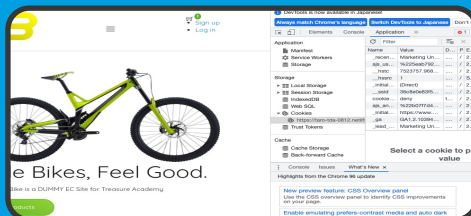
本プログラムでは講義は「オリエンテーション」+「本講義」の2日間または3日間構成
事前課題を本講義までにこなしておくことが前提となる





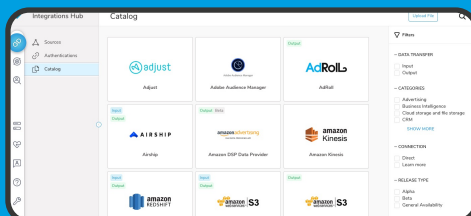
「CDP実装コース テキストブック」

- 自己学習にも利用できる、ガイド付きの演習を含むワークブック形式のテキストブック



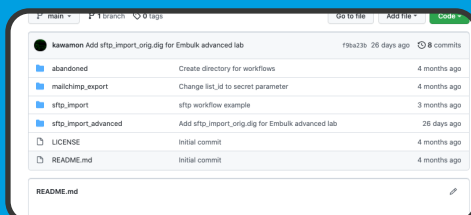
「演習用のデモウェブサイト環境」

- 受講者は「Audience Studio」コースで利用するTREASURE BIKESライクな演習用ウェブサイトを構築
- ウェブサイトのコードはGitHubで提供され、全ての受講者がJS-SDKを実装する演習で利用



「デモ用CDPアカウント」

- 受講者毎に提供される、演習で利用するTreasure Data CDPのデモアカウント



「演習用サンプルコード」

- GitHubで提供される演習用のサンプルコード



トレーニングご契約後のフロー



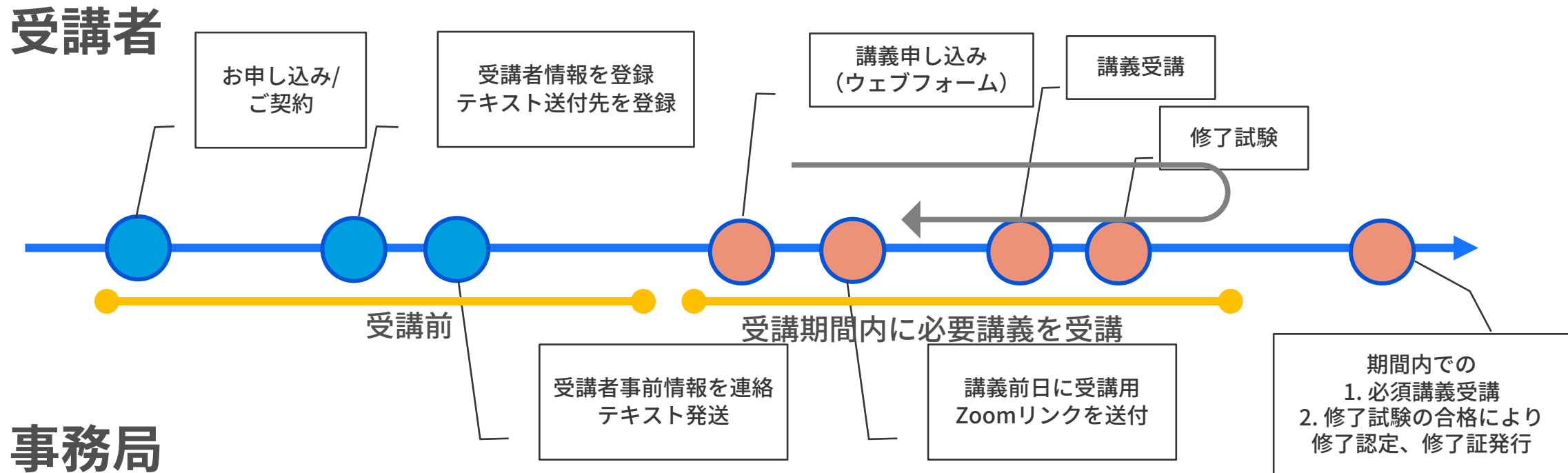
トレーニングの受講には以下の環境が必要となります

- 講義は原則としてオンライン(Zoom)で実施します。以下の条件を満たす環境が必要です
 - PC
 - Zoomにアクセスできる（カメラ、マイクは任意）
 - 演習環境にアクセスできる（ウェブブラウザからトレジャーデータのコンソール画面にログインできる）
 - 受講者ポータルサイト（ウェブサイト）から以下にアクセスできる
 - Googleスライド（ポータルサイトで講義資料の閲覧に使用します）
 - Googleフォーム（ポータルサイトから講義申込、修了試験、アンケート等で使用します）
 - 講義中の質問掲示板 (Slide.com)（※必須ではありません）

注）受講者ポータルサイト、演習環境への接続情報はトレーニング受講申し込み後に提供します。セキュリティ上Googleスライドにアクセスできない場合は代替手段を提供していますが制約が生じる場合があります

ご契約後のフロー

トレーニングのご契約後のフローは次のとおりです



- トレーニングチケットおよび購入したトレーニング、認定資格試験の受験期限は購入日から**原則3ヶ月**となります。期間終了後の修了認定は行なっておりません。
- 各プログラムは内容を変更、追加することがございますが、本資料内で示したゴール、目的を達成できるようにする点は変更ございません。
- トレーニング開催場所はオンラインが中心となります。
- ある程度まとまった数の受講者が同時に受講を検討している場合は一社向けの開催も提供可能です。詳細は担当営業までお問い合わせください。
- トレーニングマテリアルの権利は全てトレジャーデータへ帰属します。

