

DATAFLUCT cloud terminal. × 東芝 GridDB Cloudで実現

店舗単位の来客数予測を 最適化する 最新機械学習アプローチ

スーパーマーケット関係者向け

コンビニ関係者向け

ドラッグストア関係者向け

株式会社DATAFLUCT

東芝デジタルソリューションズ株式会社





Index

03 | 本資料の趣旨

04 | データ活用、進んでいますか？

05 | 店舗の特性に合わせたデータ活用方法

06 | DATAFLUCT cloud terminal. × GridDB Cloud

09 | ケーススタディ

18 | DATAFLUCT cloud terminal.

21 | GridDB Cloud

23 | 最後に / お問い合わせ

24 | 会社概要

本資料の趣旨

本資料は、現在スーパーマーケットやコンビニエンスストア、ドラッグストアなどの小売店の店舗運営に携わっており、下記のような悩みや課題をもつ方々へ向けたものです。

ご覧いただいた後に、それらを解決に導く糸口が見つかれば幸いです。

予測したよりも来客数が少なく、

売上目標額に満たなかった。

前任店舗で培った予測の経験則が

ほとんど役に立たず困っている。

機械学習での予測を取り入れたが、

地域ごとの実情が反映されないため

店舗運営に効果が出ない。

勘からデータへ

新型コロナウイルスの影響によって消費者の行動は大きく変わり、これまでの常識が通用しない時代を迎えました。変化し続ける社会情勢や消費者行動を正しくとらえる手段として、データを活用した意思決定の重要性が高まっています。本資料では、勘や経験に頼らないデータドリブン型の店舗運営のソリューションを紹介します。

店舗ごとに役立つ データが必要

消費者の購買行動は、年齢や性別、店舗の立地、競合店舗、メディア情報など、様々な要因によって変化します。ある特定の店舗での傾向が他の店舗に当てはまらないケースも多くあります。

つまり、店舗ごとに来客数予測や購買行動分析を行うには、それぞれの特性に応じたデータ分析が必要なのです。



データ活用、進んでいますか？

IT化は進展、データ活用は不十分

皆様が運営に携わっているスーパーやコンビニ、ドラッグストアなどの小売業の店舗において、データ活用は十分に進んでいるのでしょうか？

近年、ID-POSの導入や自動発注システム、電子チラシなどが普及し、店舗経営におけるIT化は進展しているといえるでしょう。しかし、IT化によって得られ、蓄積されたデータを店舗経営に活用できているかという点とそうではありません。

データ活用の難しさ

データを店舗経営に活かせていない主な理由として、下記が考えられます。

- 蓄積しているデータの分析が十分ではない。
- 収集すべきデータが多岐にわたり、異なるデータ構造やデータ処理要求に対応できていない。
- 汎用的に作られたシステムでの予測モデルのため、精度が不十分。
- 店舗ごとに最適化された予測モデルを作るために必要なデータサイエンティストが不足。
- モデルの作成にかかる費用が高額で、投資対効果に合わない。

データを活かさない店舗が陥る状況

店舗経営において重要な“店舗ごとの来客数予測”や“購買行動の分析”が不十分な場合、商品の発注過多による廃棄ロスや発注過小による機会損失が発生する可能性が著しく高まります。そのような事態に陥るのを防ごうと売れ行き商品ばかりを陳列していると、店舗から個性が失われ、次第に顧客が離れていきかねません。



店舗の特性に合わせたデータ活用方法

来客数予測と顧客の購買行動に影響を及ぼす要素

データ分析のなかでも、来客数予測や顧客の購買行動の分析はなぜ難しいのでしょうか？理由のひとつは、店舗周辺環境の特性が異なるためです。東京23区内の駅近店舗と地方の店舗では、顧客の来店頻度も来店目的も大きく変わります。

もうひとつは、思わぬ事象の発生によって顧客の需要が変わるためです。たとえば、競合店舗がセールやキャンペーンを開始した場合やニュース・SNSの発信によって新たなトレンドが発生した場合、災害が発生した場合などに顧客の行動は変化し、店舗に影響を及ぼします。

まず着手すべきは「店舗の特性に合わせたデータ収集と活用の仕組みづくり」

思わぬ事象に対しデータを活用して立ち向かおうとしても、その仕組みづくりは複雑で、容易に始められるものではありません。まずは、店舗の特性に合わせたデータ収集と活用の仕組みづくりに着手しましょう。そこから始めることが、“店舗A向けに作成したモデルが店舗Bでは十分な精度が出ない”や“全体最適化したモデルを個別の店舗に適用しても十分な精度が出ない”という課題を解決する近道となります。

機械学習によるモデル構築の困難を乗り越えるための武器

仕組みづくりに挑む皆様には、実現のための武器が必要です。その候補となるのが、機械学習モデルを構築する費用の高さ・難しさ・スキル不足といった悩みを自動機械学習（AutoML）で解決するサービス『DATAFLUCT cloud terminal.』と、多様なデータ構造に対応し、大容量・リアルタイム処理が強みのデータベースサービス『GridDB Cloud』を組み合わせたソリューションです。

DATAFLUCT cloud terminal. × GridDB Cloud

■概要

DATAFLUCT cloud terminal.

マルチクラウド環境で最適なAutoMLモデルを簡単・スピーディーに構築できる機械学習プラットフォーム。

簡単に高精度の機械学習モデルをつくれるAutoML

マルチクラウド対応

簡単デプロイ機能
(オンライン処理、バッチ処理)

信頼性の高いプラットフォーム

月額固定のシンプルな料金メニュー

詳細は [p.18](#) へ

AutoML

高度な機械学習モデルを、専門家なしに簡単に構築できる。

時系列データベース

様々なリアルタイムデータを処理・蓄積できる。

- 店舗ごとに最適な予測モデルを構築できるサービス。
- DATAFLUCTと東芝デジタルソリューションズのもつサービスを組み合わせて共同開発。

TOSHIBA GridDB 時系列データベースサービス 『GridDB Cloud』

人流データや気象データなど大量のデータを蓄えつつ、オンデマンド分析を可能にするクラウドデータベースサービス。

時々刻々発生するデータを効率よく蓄え、分析する時系列データモデル

大量のデータに対応可能な高い処理能力と拡張性

多くの社会インフラシステムで使用された実績が示す高い信頼性

運用を気にしないですむクラウドマネージドサービス

詳細は [p.21](#) へ

DATAFLUCT cloud terminal. × GridDB Cloud

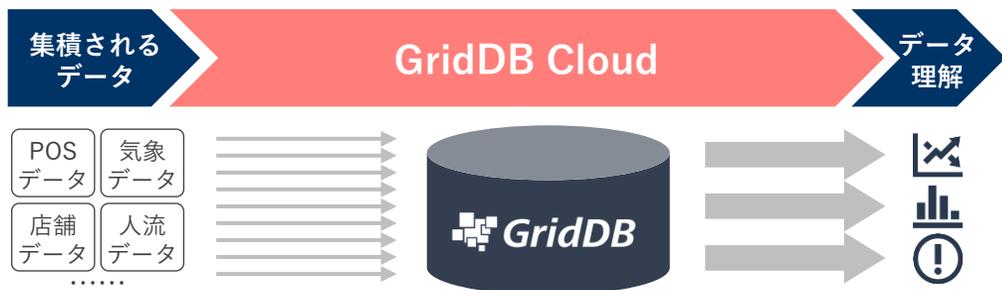
■各ソリューションの強み ～従来のデータ基盤・機械学習との違い～

データ基盤

従来のフロー



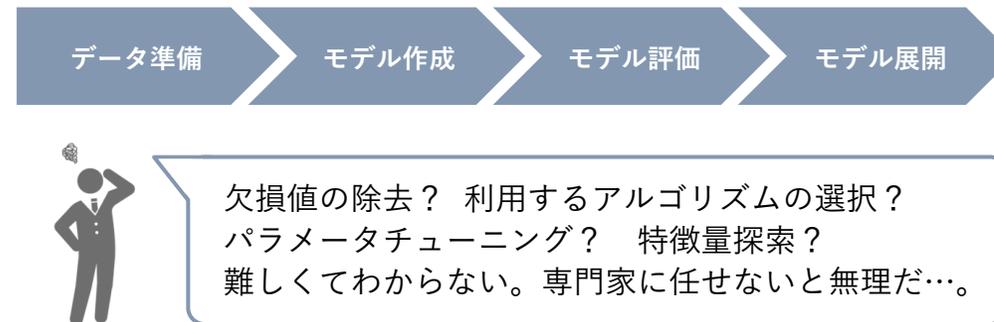
GridDB Cloudが提供するデータ基盤の場合



オンデマンドの分析が可能に。
しかも、クラウドマネージドサービスだから運用が簡単。

機械学習

従来のフロー



DATAFLUCT cloud terminal.が提供するAutoMLの場合

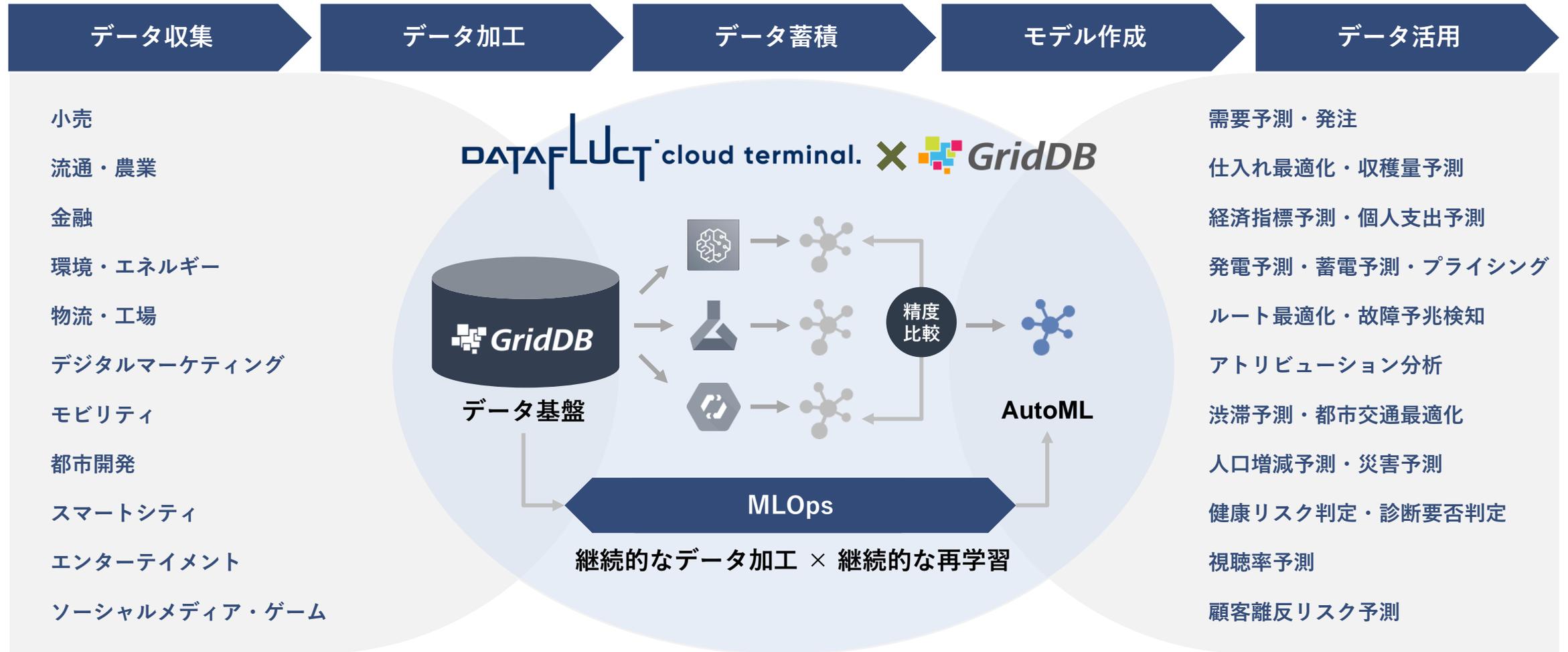


専門家でも従来は1ヶ月程度掛かる作業を、
機械学習に関する深い知識がなくても数時間で可能に。

※1 ODS (Operational Data Store) : 収集データを一時的に保持するデータベース
※2 ETL (Extract/Transform/Load) : データの整形、統合を専門に行うツール
※3 DWH (Data Warehouse) : さまざまな情報を時系列に整理して保管するデータベース

DATAFLUCT cloud terminal. × GridDB Cloud

■概要



A person wearing a dark blue suit jacket and a red and white striped tie is sitting at a wooden desk, reading an open book with a red cover. The book is open to two pages of text. A white lightbulb icon with radiating lines is superimposed on the right side of the image, symbolizing an idea or insight. A yellow rectangular box with rounded corners is positioned in the center of the image, containing the Japanese text 'ケーススタディ' (Case Study).

ケーススタディ



ケーススタディ

ここからは、『DATAFLUCT cloud terminal.×GridDB Cloud』を活用すると店舗運営にどのような効果を与えるかについて、想定事例を交えてご紹介します。

悩みを抱えている人

データ カツヒロ
出板 克洋さん (42歳)

- 現在、大手食品スーパーマーケットの本部で発注業務を担当。
- 食品廃棄物の発生量を数年以内に半減させるプロジェクトにアサイン中。各店舗のID-POSデータを分析し、日々対策を検討している。
- その一方で、惣菜市場の成長に合わせた売上向上施策の立案も期待されており、両立に頭を悩ませている。



出板さんが抱えている悩み

食品廃棄物の発生を抑制するために各店舗を分析した結果、売り上げを伸ばしている店舗は食品廃棄量が増加、売り上げが伸び悩んでいる店舗は取り扱い量自体を抑制していることが分かりました。この状況を招いている背景のひとつは、材料の発注量や調理現場のシフト計画、調理のタイミングなどの意思決定が、店長の経験頼りになっている点です。顧客の世帯構成や立地条件などが店舗ごとに異なるため、売れ行きの傾向に差異が表れており、店舗それぞれが最適な意思決定をしなければ目標の達成は難しいようです。

課題解決を目指すなかで、出板さんは機械学習を用いて店舗ごとの時間帯別来客数を予測し、結果をリアルタイムの意思判断に活用するダッシュボードの開発を思いつきました。しかし、店舗ごとに最適な予測モデルを構築するには想定以上の時間とコストが掛かるとわかり、躊躇していました。



ID-POSのみで、来客数予測モデルの作成に挑戦

STEP 1 データ準備

「待っていても、始まらない」と悟った出板さんは、すぐに入手可能なID-POSのデータを使い、来客数の予測モデルを作成してみることにしました。特性の異なる以下の店舗を対象に、ID-POSのデータから時間帯別の来客数情報を抽出します。

■ 店舗データ（店舗ごとの特性情報を含むデータ）

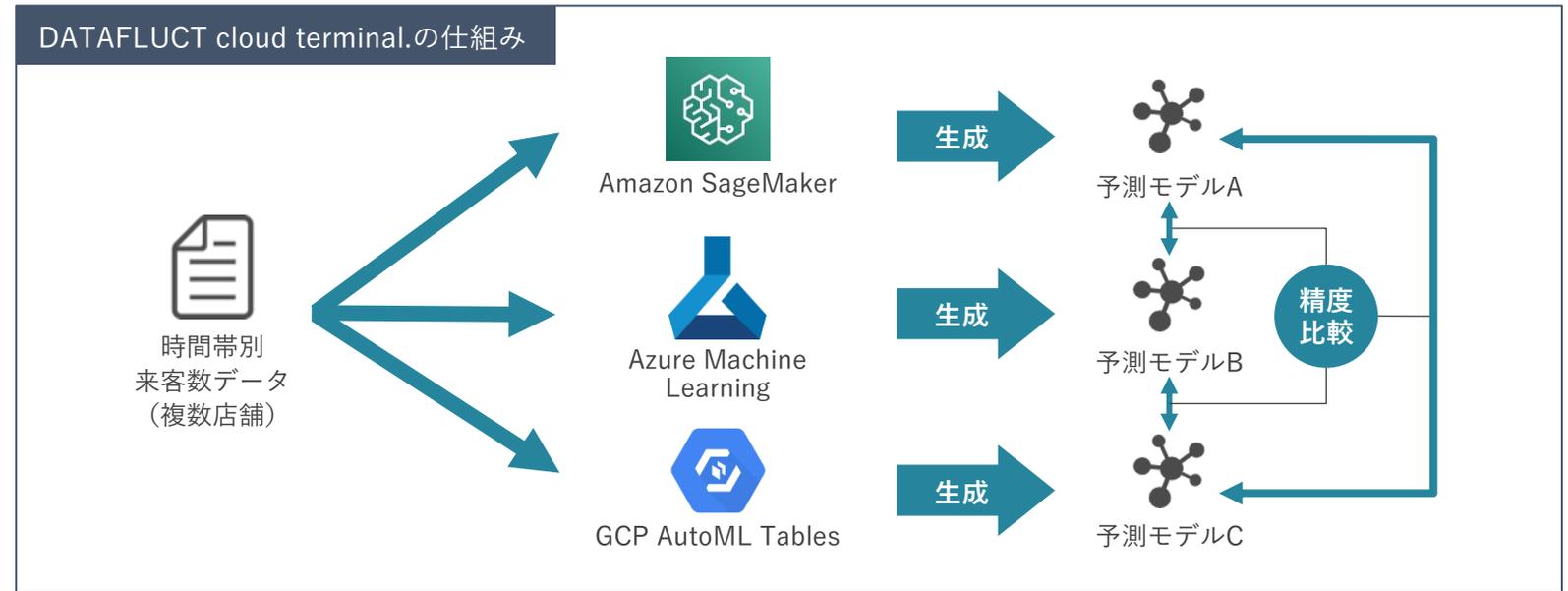
店舗	気象観測の位置	常住人口	昼間人口比率	駅からの距離	客タイプ	店舗面積
店舗1	東京	多い	増	普通	近所の住民、高級志向	中
店舗2	東京	普通	増	近い	若い主婦、学生	小
店舗3	東京	普通	減	普通	主婦	中
店舗4	東京	少ない	減	遠い	近所のお年寄り、買い溜め	大
店舗5	大阪	普通	減	普通	主婦	中

■ ID-POS データ（2021年2月の1ヶ月分のデータを時間帯別に抽出）

date	time	customerID	storeID	itemCategory	itemName	price
2021/2/16	9	294	1001	6	非食品3	200
2021/2/16	9	294	1001	3	惣菜4	360
2021/2/16	9	294	1001	5	一般食品5	600
2021/2/16	9	294	1001	5	一般食品5	600
2021/2/16	9	294	1001	4	日配品4	130

STEP 2 モデル構築

次に、STEP1で用意したデータセットを使い、AutoMLで予測モデルを作成します。調査中にみつけたAutoMLサービス『DATAFLUCT cloud terminal.』にデータセットをアップロードし、予測したいターゲット列（来客数）を選択するだけで、AutoMLの予測モデルの生成が完了しました。



早速、生成したモデルを使って“時間帯別の来客数を予測した結果”と“実際の来客数データ”を比較してみた結果、残念ながら人の経験則と比べても見劣りするレベルの精度（予測誤差：±30%～50%）となってしまいました。なぜ、このような結果になってしまったのでしょうか？



STEP 3 モデル評価

ID-POSデータのみを使用して生成した予測モデルの予測精度が低かったことについて考察すると、以下のようなことが分かりました。

- 1日単位・1週間単位で分析した場合、予測誤差は±10%程度だった。
- 時間帯別の来客数は平日でも曜日によって異なる上、店舗によっても傾向が異なる。
- 人の店舗前の通行量は、天候によって変動する。

また、データサイエンティストからは、“特定の1ヶ月だけのデータを使用したこと”と“データ量と特性が異なる店舗データ混ぜてしまったこと”によって、**教師データに分散と偏りが生じている可能性**があると助言をもらいました。

- 東京の店舗のデータ量が多すぎるため、大阪の店舗の予測が影響を受けている。
- 同じ年代の顧客でも駅近の店舗の方が人数が多いため、郊外の店舗の予測に悪影響が出ている。
- 2021年2月のデータで作成したモデルは、季節性の違いやコロナ禍などの社会情勢の違いによって、3ヶ月後、半年後、1年後には更に精度が下がる可能性が考えられる。

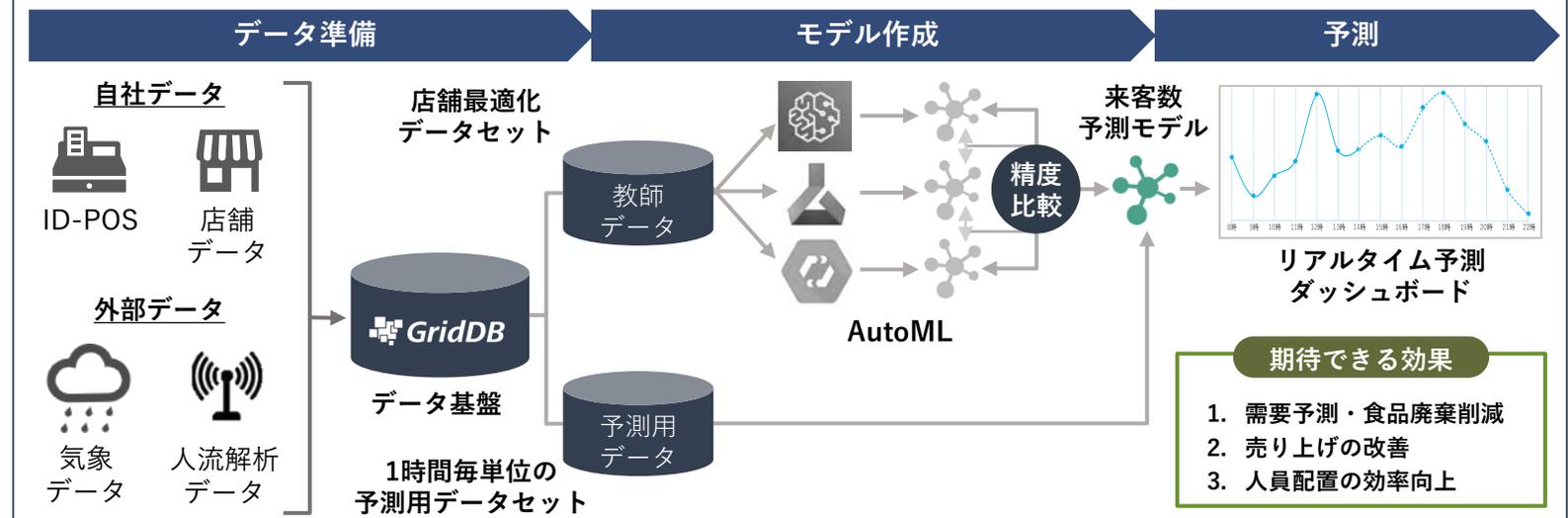
予測モデルの作成には、**分散と偏りのないデータが必要**です。上記の評価を踏まえ、機械学習の予測精度を高く維持し続けるために必要なものは、「**自社のデータと外部のデータ**」、そして「**季節性・社会性を常時反映するための継続的なデータの収集・加工・蓄積できるデータ基盤**」であることがわかりました。

DATAFLUCT cloud terminal. × GridDB Cloudで来客数予測モデルの作成に挑戦

精度の高い予測モデルを作成できるデータ基盤について調査した出板さんは、リアルタイムデータに強く、多様なデータを扱える『GridDB Cloud』と先述した『DATAFLUCT cloud terminal.』を組み合わせたソリューションの存在を知り、自社に導入することにしました。導入時の主な流れは下記の通りでした。

- 各店舗に導入しているID-POSデータを集約するデータ基盤としてGridDB Cloudを用意。
- 外部要因を考慮するために外部データ（気象データ、人流データ）をリアルタイムに収集・蓄積・加工。
- 収集した店舗別データと外部データを基に、店舗ごとに最適な予測モデルを構築。

ソリューションのフロー

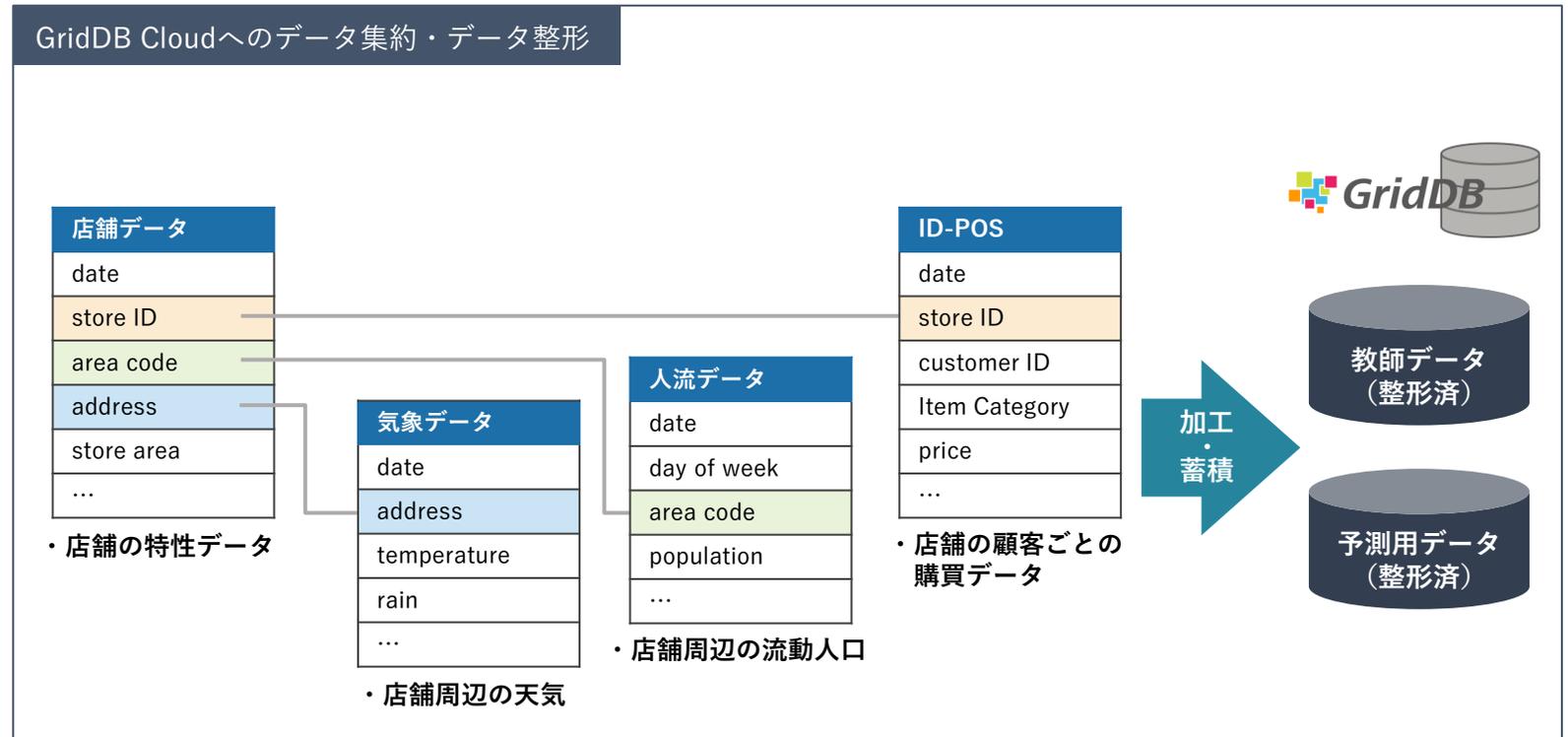


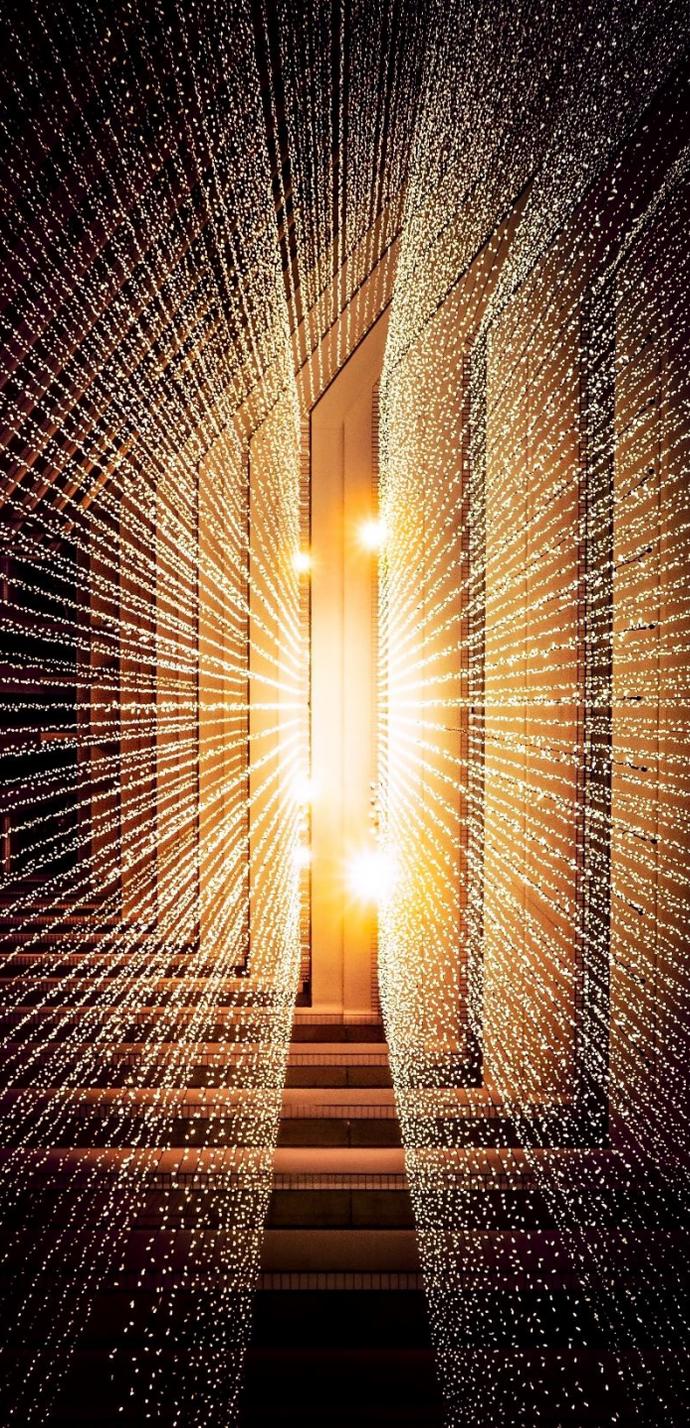
STEP 1 データ準備

店舗ごとの来客数を予測するために下記のデータを用意しました。

自社データ	ID-POS、店舗データ	外部データ	気象データ、人流解析データ
-------	--------------	-------	---------------

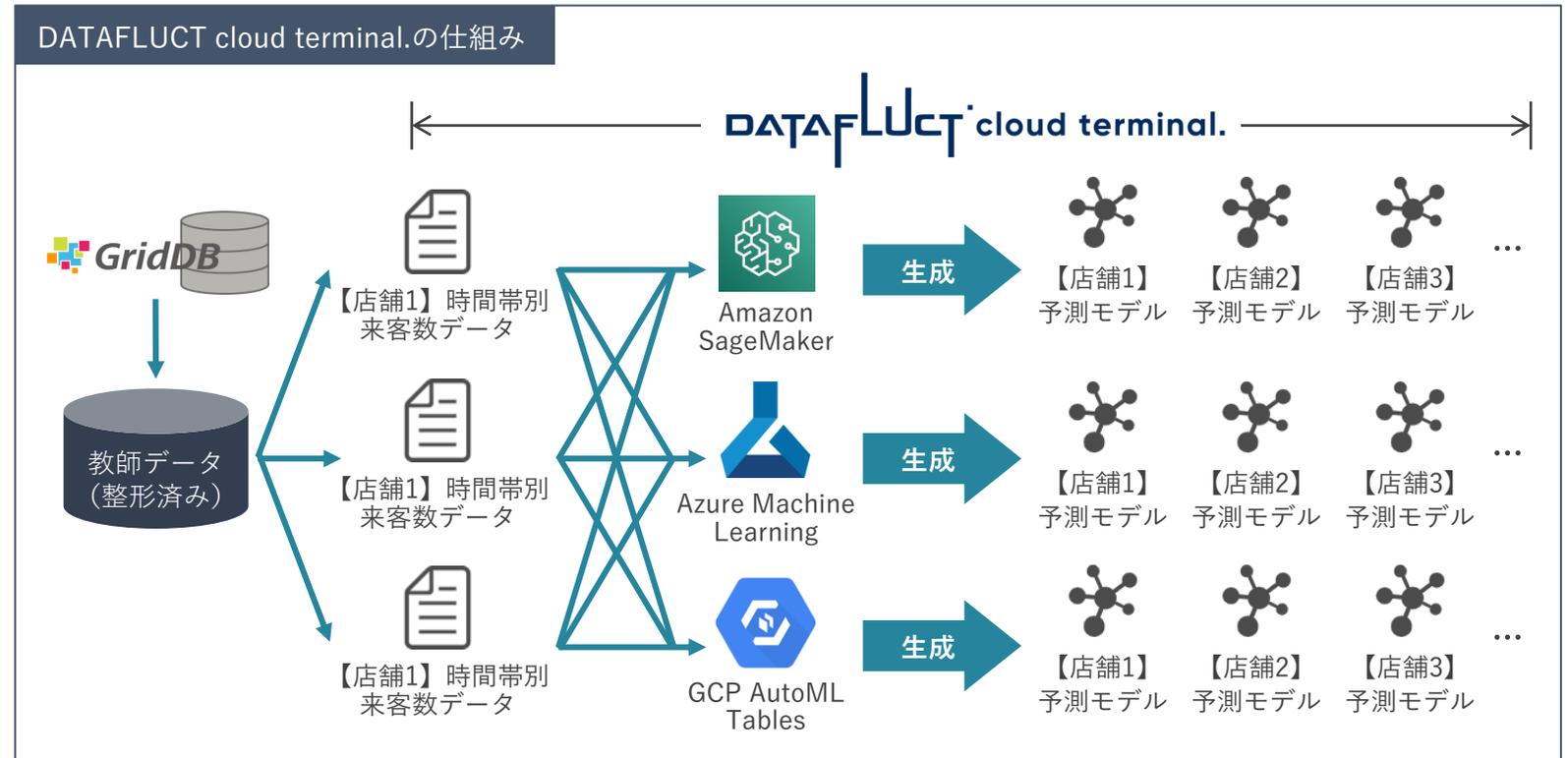
データ構造や更新頻度が異なるデータを全てGridDB Cloudに集約し、店舗ごとに最適なモデルを構築できるようにデータを整形します。





STEP 2 モデル構築

次に、GridDB Cloud上に用意した店舗ごとに整形済のデータセットを使い、AutoMLで予測モデルを作成します。

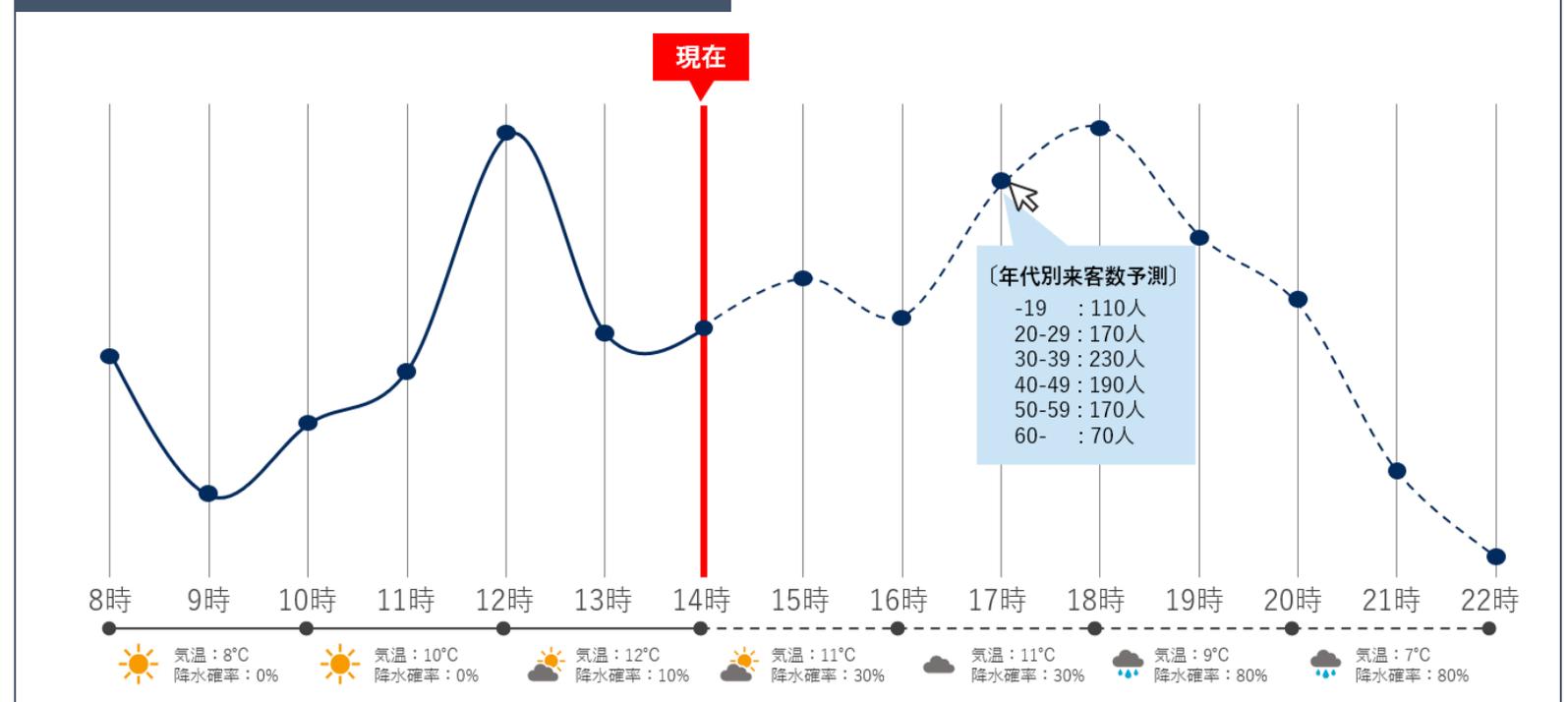


作成するモデル数は増えましたが、マルチクラウドAutoMLによって、各クラウドが作成したモデルから最も精度の高いものを人手をかけずに簡単に構築することができました。

STEP 3 モデル評価・活用

外部データを取り込み、店舗ごとにデータ整形・モデル生成を実施することで、時間帯別の予測精度は【予測誤差：±5%~15%】まで向上しました。最終的に当初の構想どおりに年代別来客数の予測結果をリアルタイムに表示するダッシュボードを用意した結果、惣菜品の調理タイミングや棚出し計画のリアルタイムな意思決定が改善され、食品廃棄物の発生抑制、売上向上の両立を実現しました。

ダッシュボードのイメージ図（年代別来客数予測）



DATAFLUCT cloud terminal.

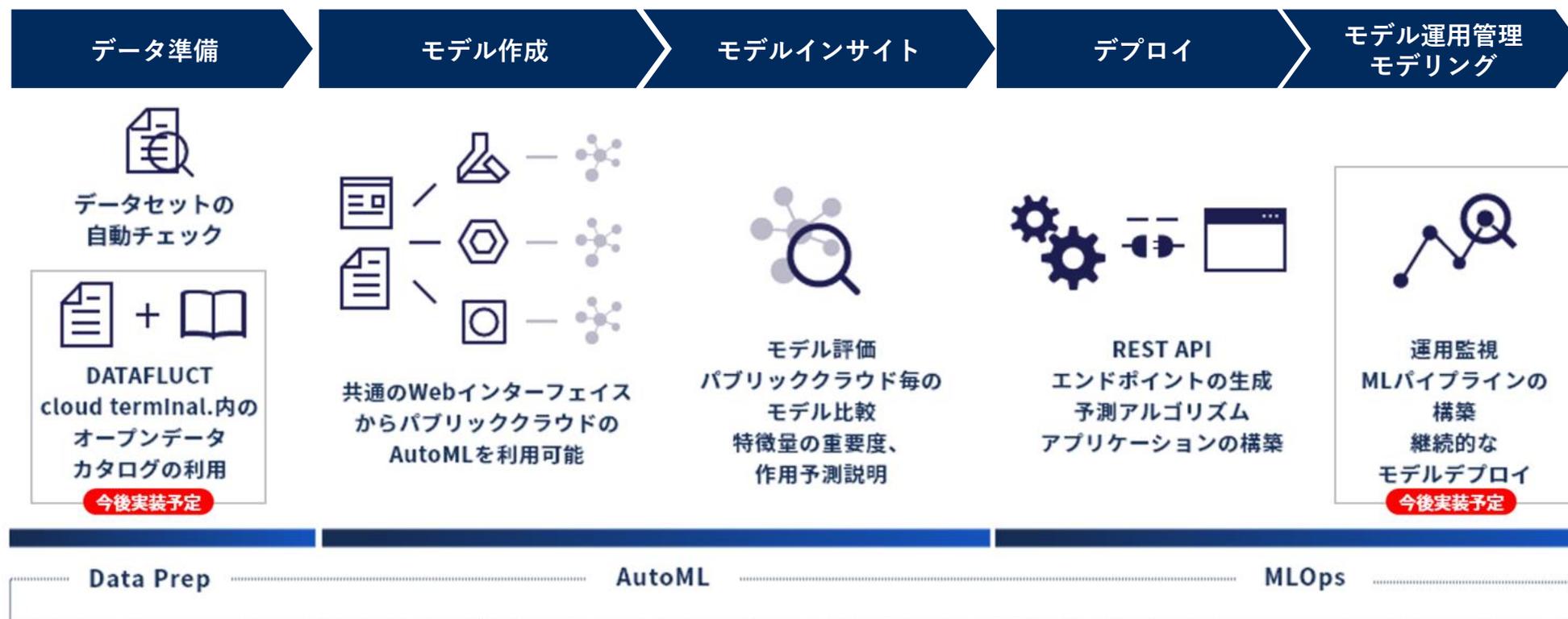
マルチクラウド環境で最適なAutoMLモデルを
簡単・スピーディーに構築できる機械学習プラットフォーム



さらに詳しい情報は[こちら](#)

DATAFLUCT cloud terminal.

データセットを用意するだけで、AWS, Azure, GCP の3クラウドのAutoMLを同時実行・比較し、最も精度の高いモデルを作成、予測結果をすぐに得られる



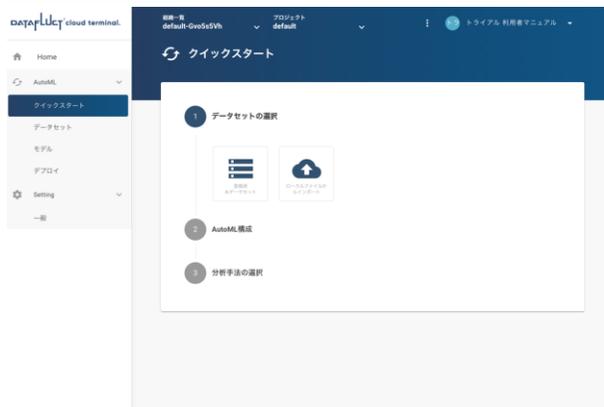
さらに詳しい情報は[こちら](#)

次ページ | マルチクラウドでのAutoMLが、3STEPで簡単に

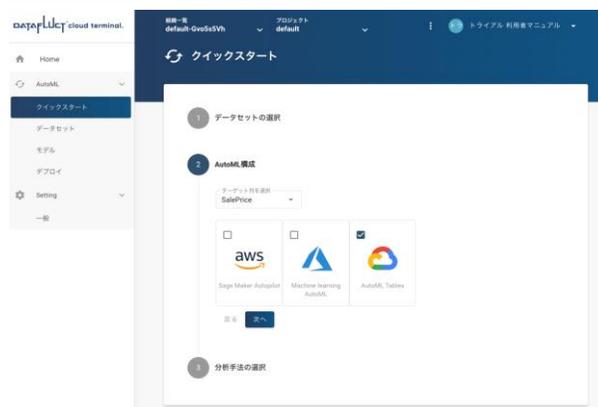
DATAFLUCT cloud terminal.

マルチクラウドでのAutoMLを、3STEPで簡単に実行

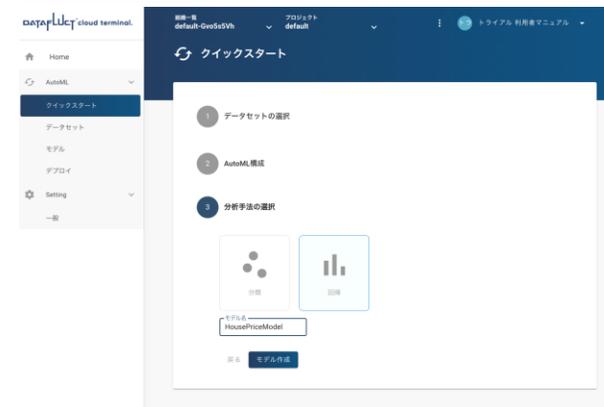
【STEP1】 データセットの用意



【STEP2】 予測対象・クラウド選択



【STEP3】 分析手法の選択

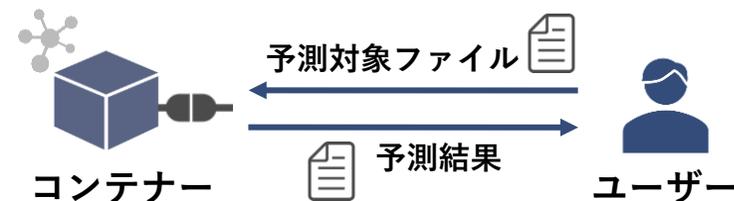


展開

オンライン予測：アプリケーションに組み込む



バッチ予測：複数データの一括予測



さらに詳しい情報は [こちら](#)

次ページ | クラウドデータサービス『GridDB Cloud』

GridDB Cloud

簡単かつスマートに環境設定・運用監視・メンテナンスを実現する
クラウドデータベース サービス

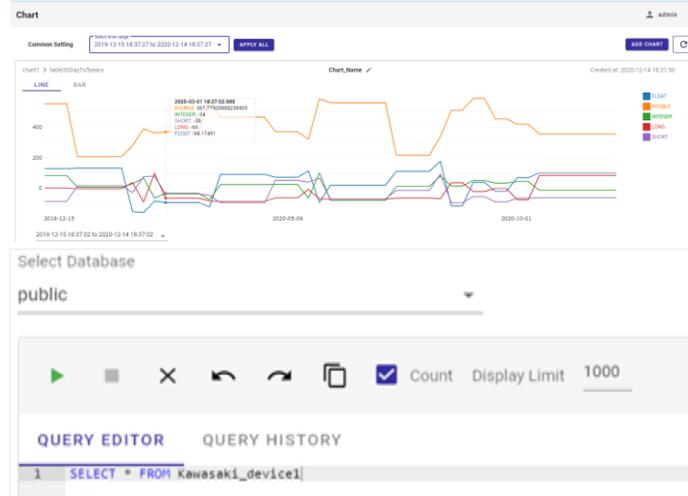
GridDBの特徴を継承しつつ、
マネージドサービスを実現



GridDBの構築や運用・監視などは不要。また、アプリケーションは従来のGridDBと同じAPIを使用できるため、今までの開発手法を継承でき、技術習得のコストが抑えられます。

さらに詳しい情報は[こちら](#)

格納されたデータの可視化や
SQL操作など、充実した機能



アプリケーション開発者やデータベース運用者は、データの収集状況の確認やデータの検索が可能です。アプリケーションのデバッグや、欠損データの検出などに役立てることができます。

用途に合わせた
ラインナップを用意

スタンダード
Standard

月額料金300,000円～

- 4 vCPU
- 16 GB メモリ
- 1 TB ストレージ (オプションで追加可能)
- 1ノードまたは3ノード構成を用 (オプションでノード追加可)

プロフェッショナル
Professional

月額料金600,000円～

- 8 vCPU
- 32 GB メモリ
- 4 TB ストレージ (オプションで追加可能)
- 1ノードまたは3ノード構成を用 (オプションでノード追加可)

エンタープライズ
Enterprise

月額料金900,000円～

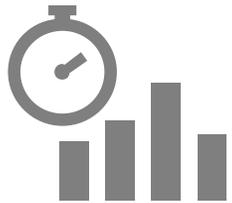
- 16 vCPU
- 64 GB メモリ
- 8 TB ストレージ (オプションで追加可能)
- 1ノードまたは3ノード構成を用意 (オプションでノード追加可)

スタンダード、プロフェッショナル、エンタープライズの3つのモデルを用意。データの増加によって性能のアップが必要になったとき、モデルの変更やサーバの追加で簡単に拡張できます。

[参考] GridDB

時々刻々と発生する膨大な時系列データの蓄積とリアルタイムな活用をスムーズに実現し、
ビジネスを大きく成長させるために最適な次世代データベース

IoT時系列
データ指向



高頻度で膨大な時系列データを効率よくリアルタイム処理・管理する時系列データ指向。

ペタバイト級の
高い処理能力



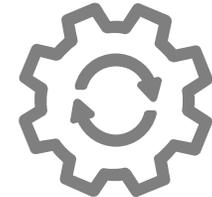
ペタバイト規模に対応可能なさまざまなアーキテクチャ上の工夫を組み込み、高い処理能力を実現。

高い信頼性と
柔軟な拡張性



万一の障害やサーバ増設においてもノンストップ運用を実現する、高い信頼性と柔軟な拡張性。

開発の俊敏性と
使いやすさ



NoSQLインターフェースだけではなくSQLインターフェースを用意し、開発の俊敏性と使いやすさを実現。

さらに詳しい情報は[こちら](#)

次ページ | データドリブンで意思決定できる環境・文化へ

最後に

変化し続ける社会情勢や顧客の購買行動を捉え、競合優位性を強化するためには、データを集約・分析し、データドリブンで意思決定ができる環境・文化をつくる必要があります。

東芝デジタルソリューションズが培ってきた信頼性の高いデータ基盤とDATAFLUCTが提供するデータ分析・機械学習サービスの組み合わせは、データ基盤作りから高度なデータ活用までのライフサイクル全体をサポートできる強力なソリューションです。データ活用による様々な課題解決に挑戦する皆さまを、全力で支援します。



お問い合わせ

株式会社DATAFLUCT

Mail : info@datafluct.com

フォームから問い合わせる

東芝デジタルソリューションズ株式会社

Mail : griddb@ml.toshiba.co.jp

フォームから問い合わせる

※ Amazon Web Services、AWS、および"Powered by Amazon Web Services"ロゴは米国およびその他の諸国における、Amazon.com, Inc. またはその関連会社の商標です。

※ Microsoft、Azure、および"Azure"ロゴは、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。

※ Google Cloud Platform (GCP)、および"Google Cloud Platform (GCP)"ロゴはGoogle LLC の登録商標または商標です。

会社概要



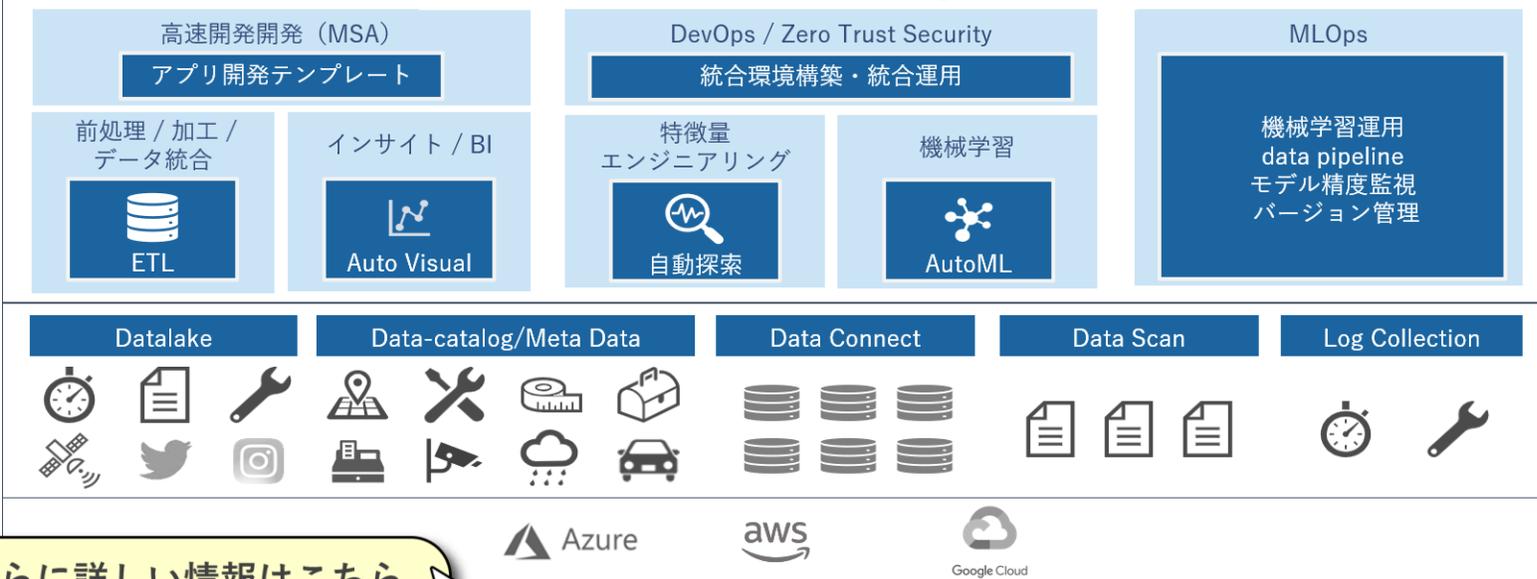
DATAFLUCTは、「データを商いに」というビジョンのもと、データから新たなビジネスを創出し企業と社会に価値を生み出すデータサイエンス・スタートアップスタジオです。衛星画像からPOSデータまで幅広い分析実績をもち、技術やデータに拘

らず業界をこえたアルゴリズム構築が可能です。多彩な自社サービス開発の知見を生かしたDX支援のほか、データ活用でSDGsに貢献することを目指し、ビジネスと社会貢献を両立させる新規事業開発にも取り組んでいます。

Data Science Startup Studio

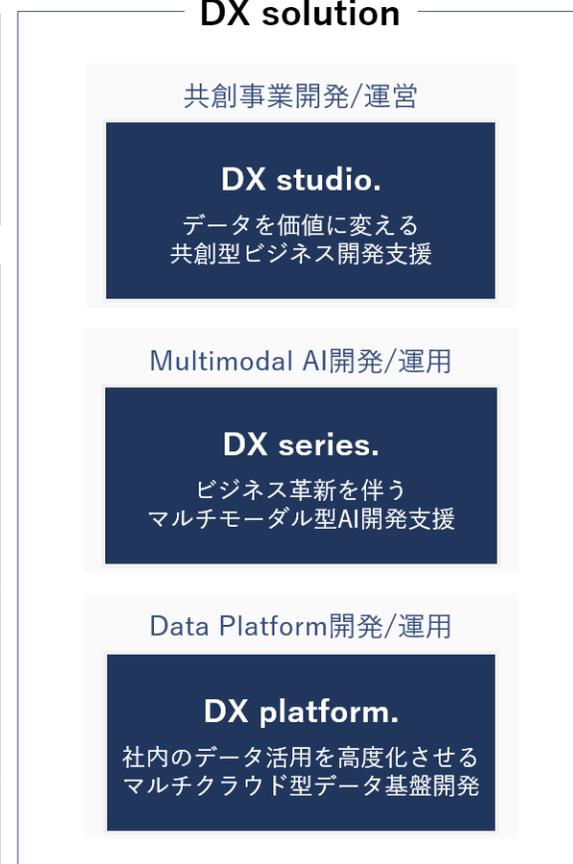


Multimodal Machine Learning Platform



さらに詳しい情報は[こちら](#)

DX solution



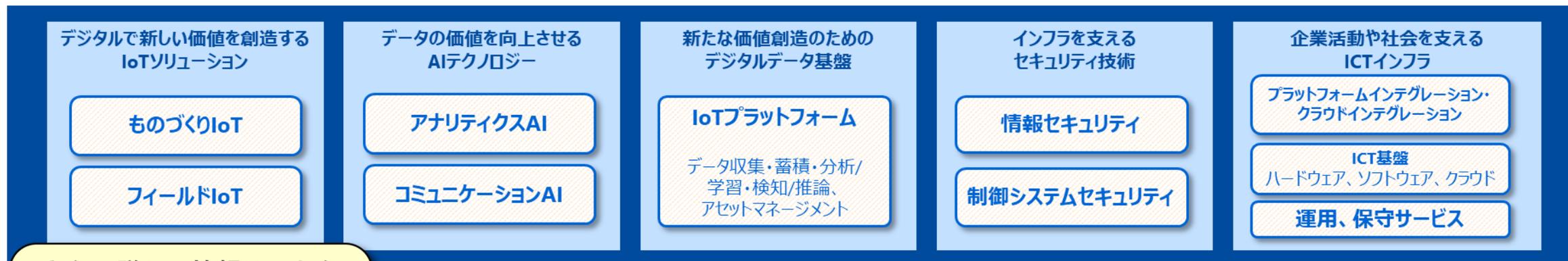
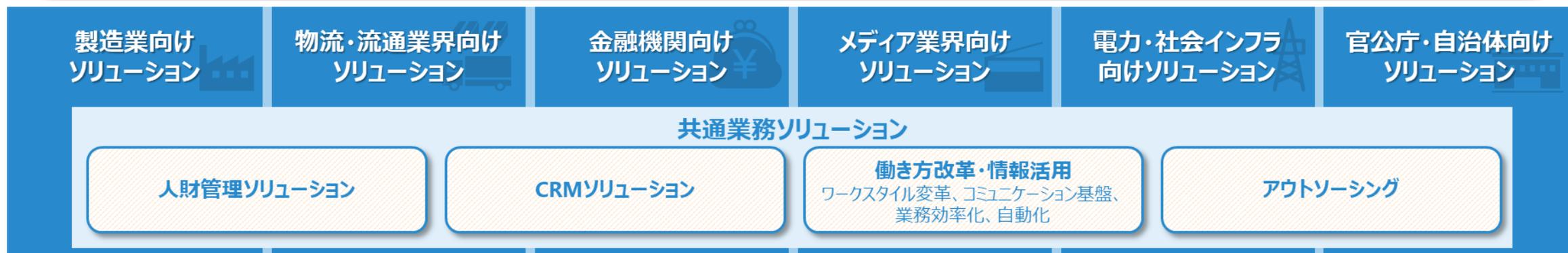
会社概要

TOSHIBA 東芝デジタルソリューションズ株式会社

東芝グループは、製造業として長年にわたり培ってきた幅広い事業領域の知見や実績と情報処理やデジタル・AI技術の強みを融合し、サイバーフィジカルシステムテクノロジー企業として成長することを目指します。
東芝デジタルソリューションズ株式会社は、



デジタルソリューション事業領域の中核企業としてデジタルトランスフォーメーションを推進し、各業種向けシステムインテグレーション、共通業務ソリューションおよびAI・IoTを活用したサービスソリューションを提供します。



さらに詳しい情報はこちら