INNOVATE

MODERN APPLICATIONS EDITION

OCTOBER 27, 2021

アプリケーション開発者向け
機械学習周辺でのサーバーレス利用方法

下川 賢介
アマゾン ウェブ サービス ジャパン株式会社
シニア サーバーレス ソリューションアーキテクト

画面に映る資料の撮影などによる本セッション資料の転用を禁止しております

aws

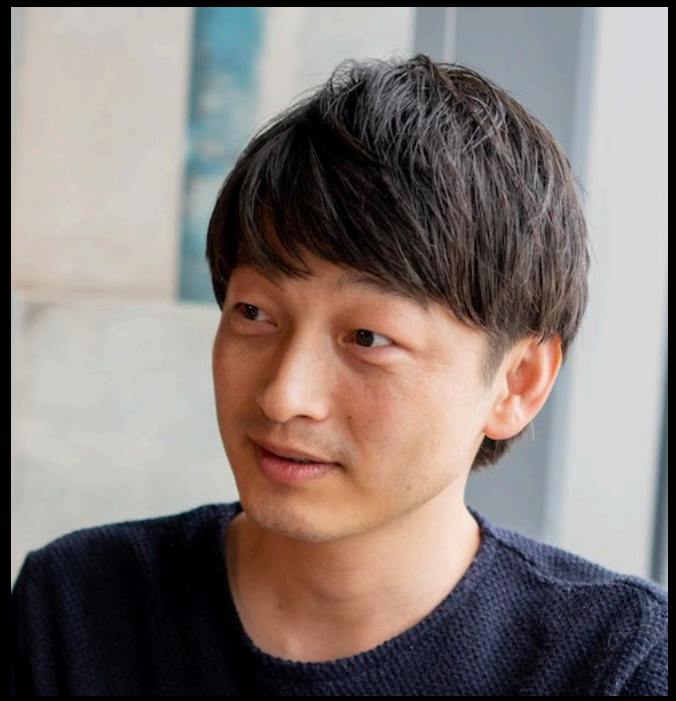
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

下川 賢介

アマゾン ウェブ サービス ジャパン株式会社
技術統括本部 レディネスソリューション本部
シニア サーバーレス ソリューションアーキテクト

AWS のサーバーレスサービスを担当

好きな AWS サービス: AWS Lambda



_kensh

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

このセッションでお話しすること

- アプリケーション開発者と機械学習
  - 機械学習のスキル不要で手軽に使えるAI サービス
  - サーバーレスで ML Ops
  - サーバーレスで 推論エンドポイント構築
  - 推論 Lambda関数 のチューニング
  - まとめ

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS のミッション

全てのデベロッパーの方々の手に機械学習を

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

あらゆる規模や業界のお客様が AWS 上で機械学習を実行しています

数万ものお客様が機械学習のワークロードに AWS を選択

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

アプリケーション開発者と機械学習

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

アプリケーション開発者のマインド

ビジネスにとって機械学習は効果的か？

どこに機械学習を適用すれば良いかを知るには、
プロダクトの本質的な価値を考える必要がある。

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

アプリケーション開発者と機械学習

データを元に継続的に改善していくプロダクト設計が重要

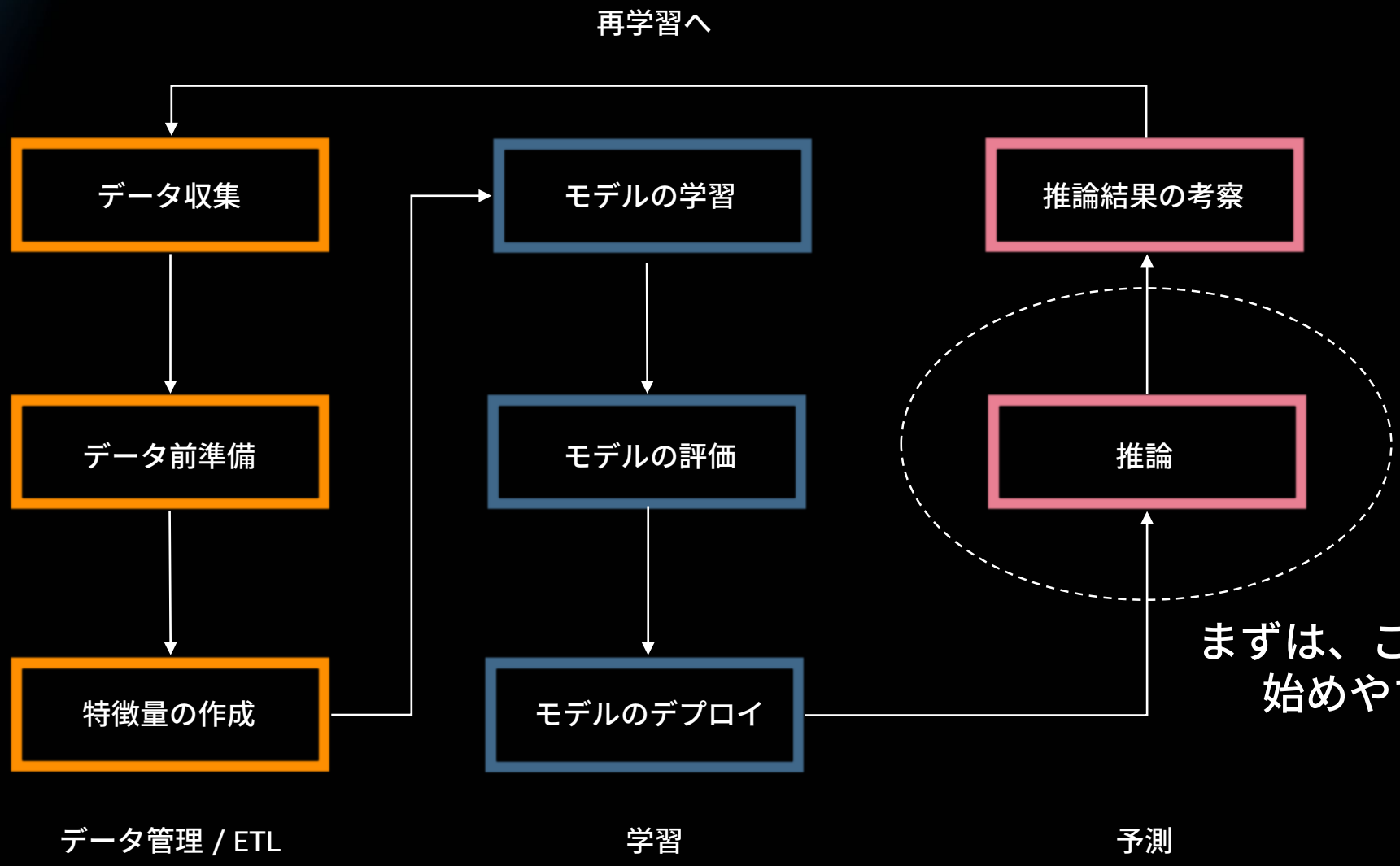
- 機械学習は改善手法のひとつ
- 改善はイノベーションに繋がり、ビジネスを加速させる

プロダクトだけでなく、企業活動全体に広げられると良い

- サービス運用 / 障害対応
- 営業活動
- 採用 / 評価 etc...

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

機械学習の流れ



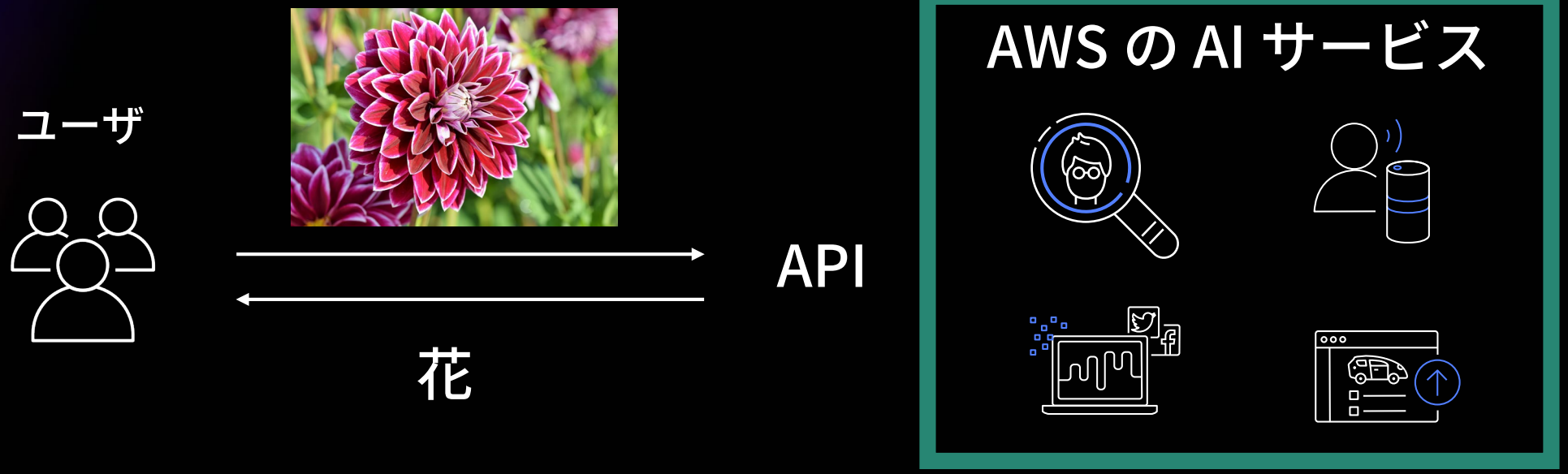
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

機械学習のスキル不要で
手軽に使えるAI サービス

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

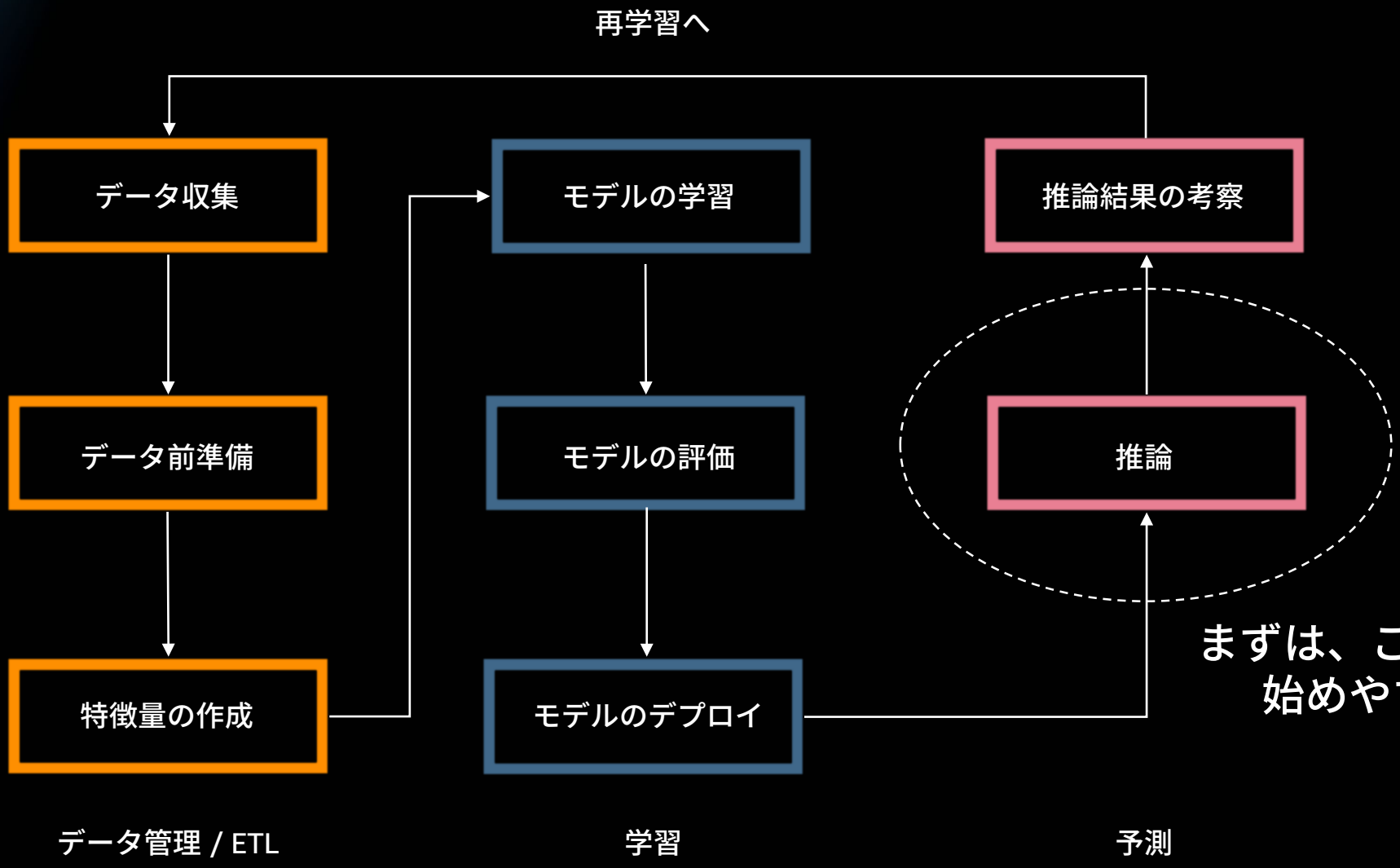
AI サービス

- データを用意するだけで、API から機械学習を利用できる
- 利用する機械学習は、 AWS によって最適な実装がされている



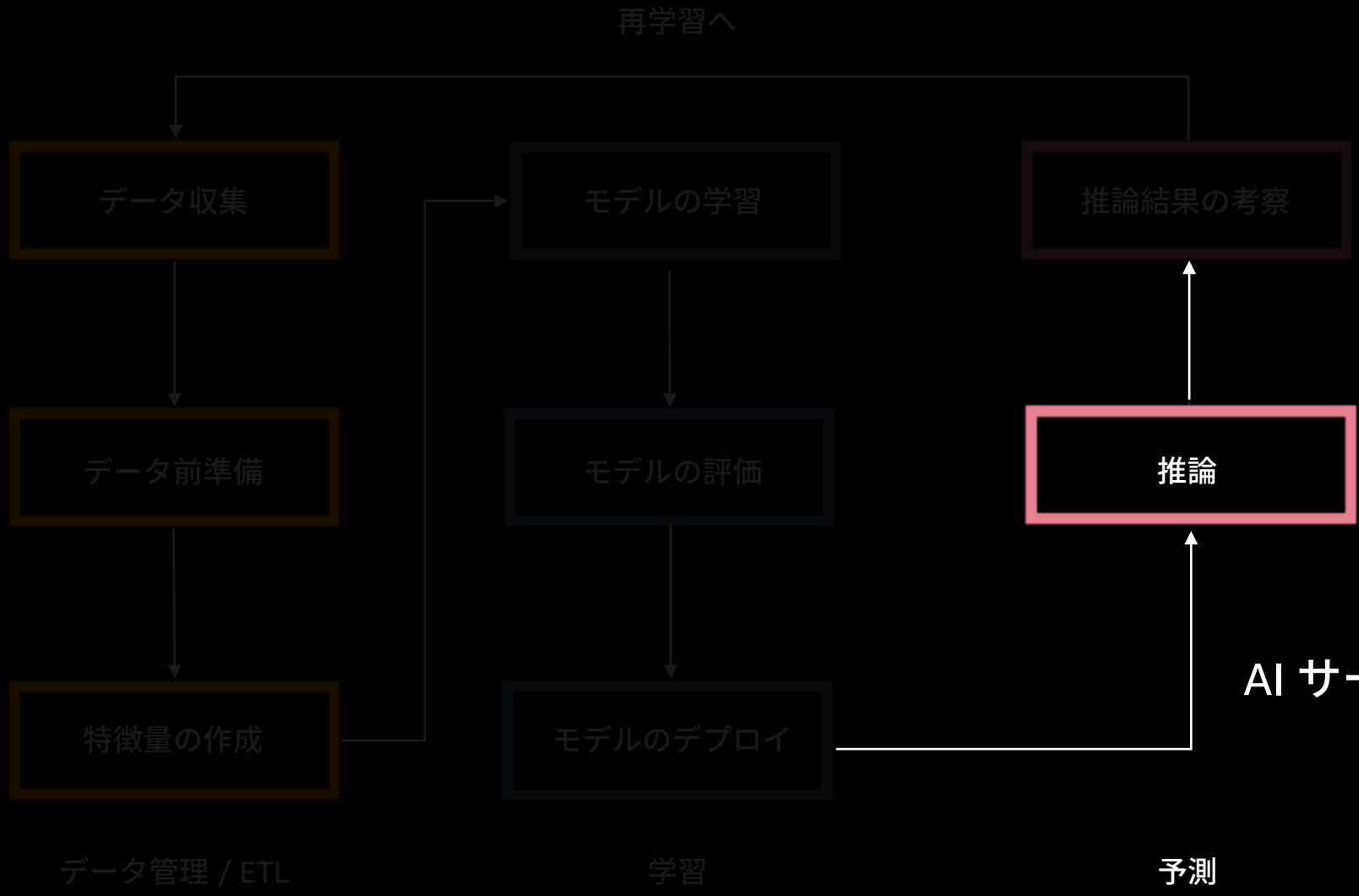
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

機械学習の流れ



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

機械学習の流れ



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

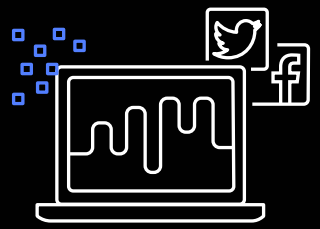
API で手軽に使える機械学習

AI サービスを使えば、機械学習の深いスキルなしに 機械学習をアプリケーションに組み込める

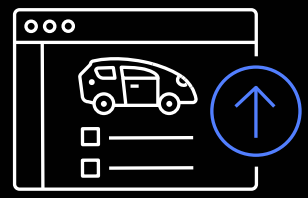
VISION	SPEECH		TEXT			SEARCH	CHATBOTS	PERSONALIZATION	FORECASTING	FRAUD	CONTACT CENTERS
Amazon Rekognition	Amazon Polly	Amazon Transcribe +Medical	Amazon Comprehend +Medical	Amazon Translate	Amazon Textract	Amazon Kendra	Amazon Lex	Amazon Personalize	Amazon Forecast	Amazon Fraud Detector	Contact Lens Voice ID For Amazon Connect

画像・映像系

自然言語系



時系列予測

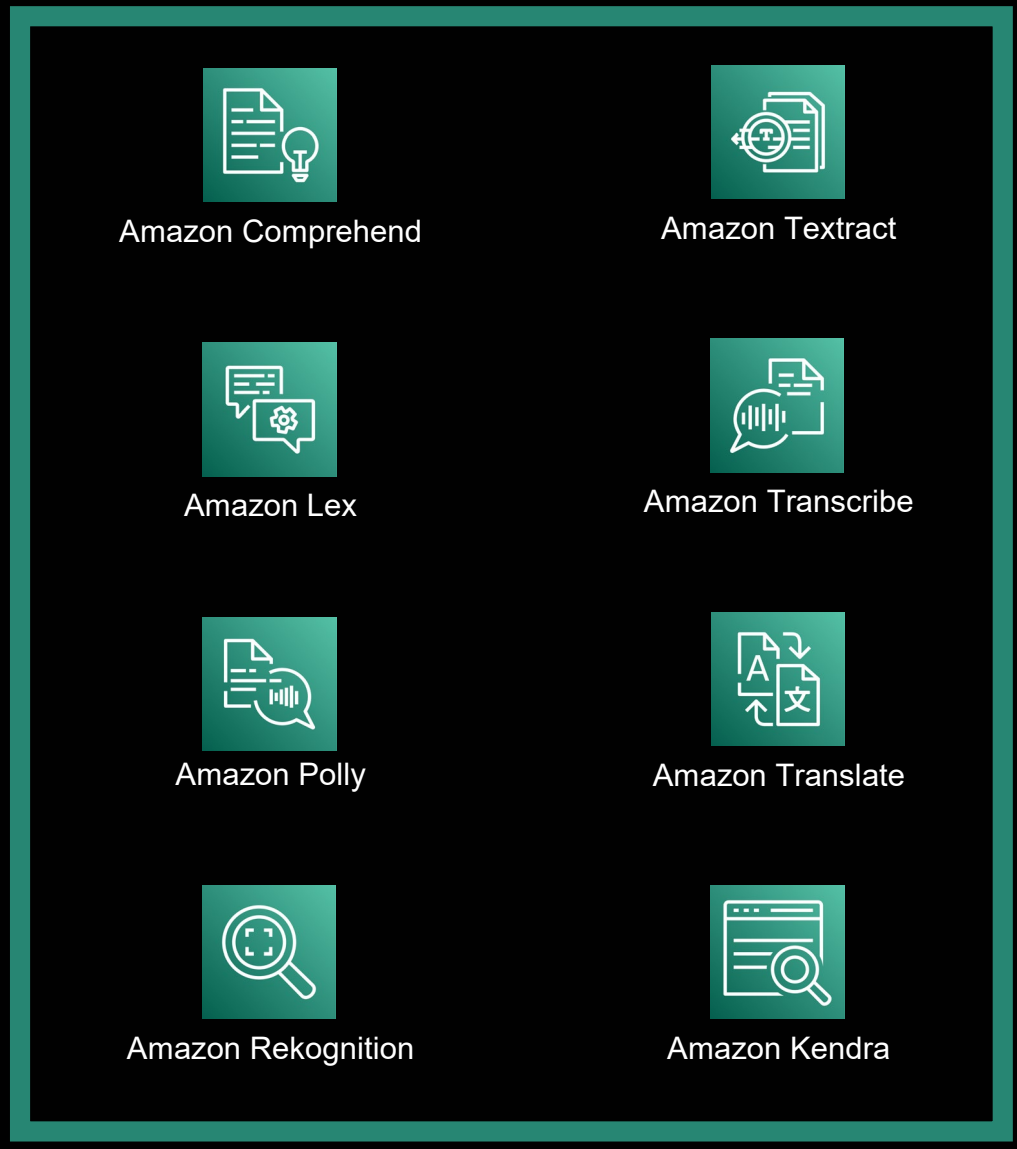


レコメンド

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

サーバーレスサービスと組み合わせる



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

機械学習 API を用いた、画像と動画分析

Amazon Rekognition を使用すると、数分以内に数百万の画像や動画を分析するなど、
手動では実行できないタスクに取り組むことができます

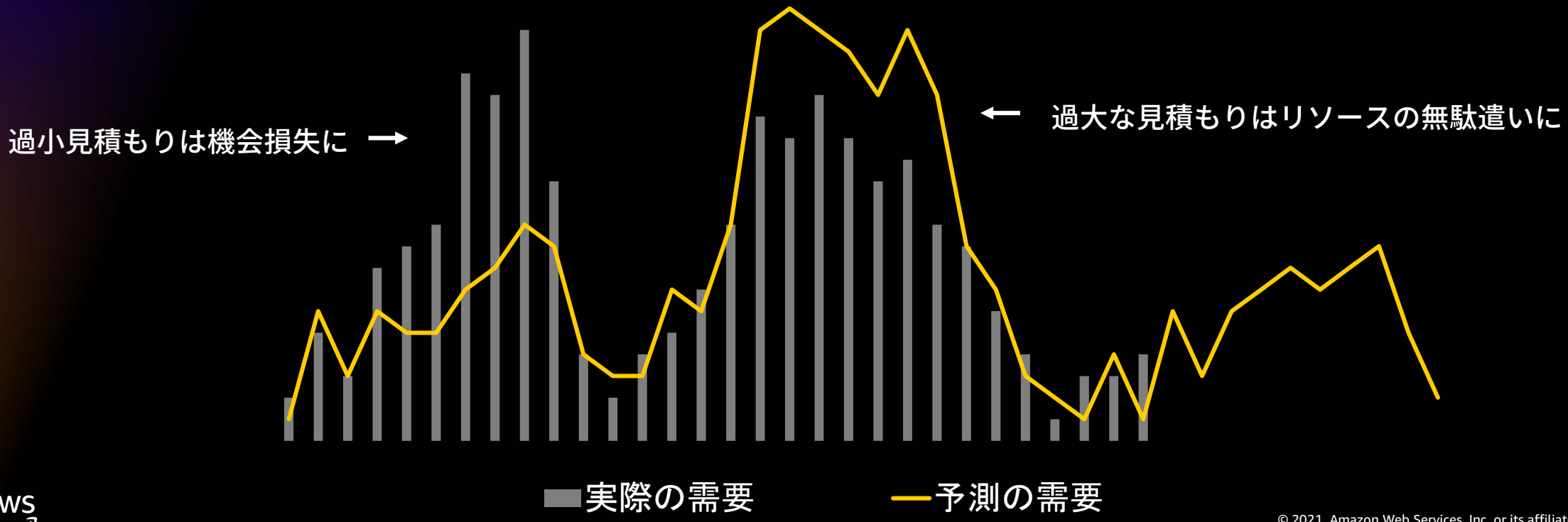
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

機械学習 API を用いた、商品の需要予測

Amazon Forecast は、機械学習を使用して精度の高い需要予測が可能

商品の需要

実際の需要 vs 予測の需要 ($ Millions)



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

機械学習 API を用いた、音声 テキスト相互変換

Amazon Transcribe

音声からテキストへ変換
(日本語含む31種の言語対応)

Amazon Polly

テキストから音声へ変換

Amazon Transcribe

Amazon Polly

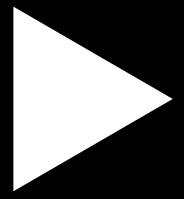
Welcome to Tokyo.

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

機械学習 API を用いた、言語翻訳

Amazon Translate は、高速で高品質な言語翻訳を提供する機械翻訳サービス

Amazon Translate is a neural machine translation service that delivers fast, high-quality, and affordable language translation. Neural machine translation is a form of language translation automation that uses deep learning models to deliver more accurate and more natural sounding translation than traditional statistical and rule-based translation algorithms. Amazon Translate allows you to localize content - such as websites and applications - for international users, and to easily translate large volumes of text efficiently.



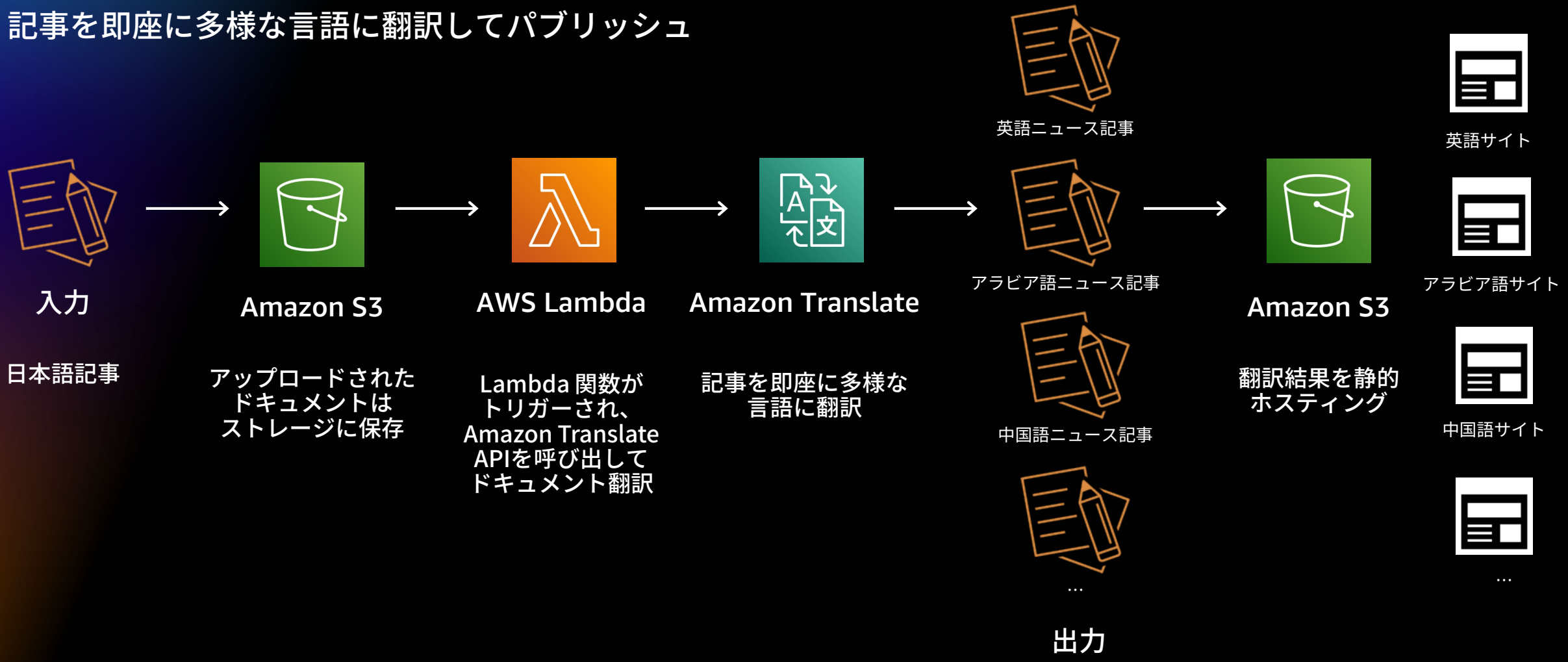
Amazon Translate は、高速、高品質、低コストの言語翻訳を提供するニューラル機械翻訳サービスです。ニューラル機械翻訳は、ディープラーニングモデルを使用して、従来の統計やルールベースの翻訳アルゴリズムよりも正確で自然な翻訳を提供する言語翻訳自動化の一形態です。Amazon Translate では、ウェブサイトやアプリケーションなどのコンテンツを海外ユーザー向けにローカライズし、大量のテキストを効率的に簡単に翻訳できます。

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon Translate による言語翻訳

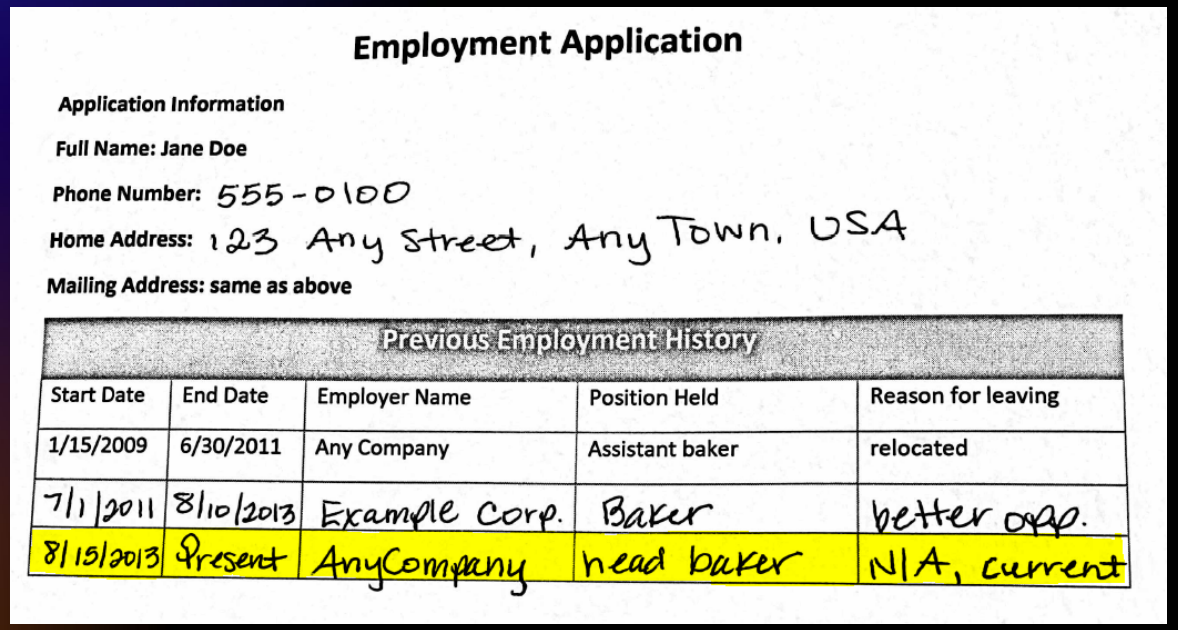
記事を即座に多様な言語に翻訳してパブリッシュ



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

機械学習 API を用いた、手書き文字の抽出

Amazon Textract は、スキャンされたドキュメントから
テキスト、手書きの文字、データを自動抽出



Employment Application

Application Information

Full Name: Jane Doe

Phone Number: 555-0100

Home Address: 123 Any Street, Any Town, USA

Mailing Address: same as above

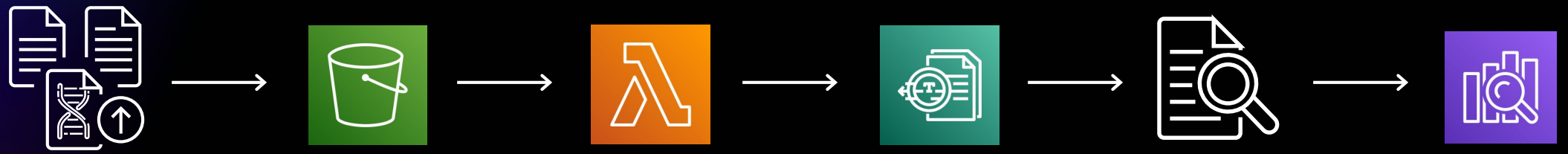
Previous Employment History

Start Date	End Date	Employer Name	Position Held	Reason for leaving
1/15/2009	6/30/2011	Any Company	Assistant baker	relocated
7/1/2011	8/10/2013	Example Corp.	Baker	better opp.
8/15/2013	Present	AnyCompany	head baker	N/A, current


```
{
    "DocumentMetadata": {
        "Pages": 1
    },
    "Blocks": [{
        "BlockType": "WORD",
        "Confidence": 1.0,
        "Text": "SomeText",
        "TextType": "HANDWRITING" // TextType.HANDWRITING
        "RowIndex": 0,
        "ColumnIndex": 0,
        "RowSpan": 1,
        "ColumnSpan": 1,
        "Geometry": {
            "BoundingBox": {
                // BoundingBox definition
            },
            "Polygon": {
                // Polygon definition
            }
        },
        "Id": "SomeId",
        "Relationships": [
            // Relationship
        ],
        "EntityTypes": [
            "VALUE" // EntityType.VALUE
        ],
        "SelectionStatus": "SELECTED",
        "Page": 1
        "Text": "Hello"
    }],
    ...
    }
}
```


© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon Textract による文字抽出と検索



入力

納税申告書、クレジット明細、医療メモなどのアップロードされたドキュメント画像

Amazon S3

アップロードされたドキュメントはストレージに保存

AWS Lambda

Lambda 関数がトリガーされ、Amazon Textract API を呼び出してドキュメント分析が開始

Amazon Textract

キーと値のペアやテーブルなどのテキストを自動的に抽出

出力

抽出データを任意のドキュメント管理システムに統合

Amazon OpenSearch Service *

抽出されたデータと信頼スコアが、ドキュメント検索を可能にするためにインデックス付与される

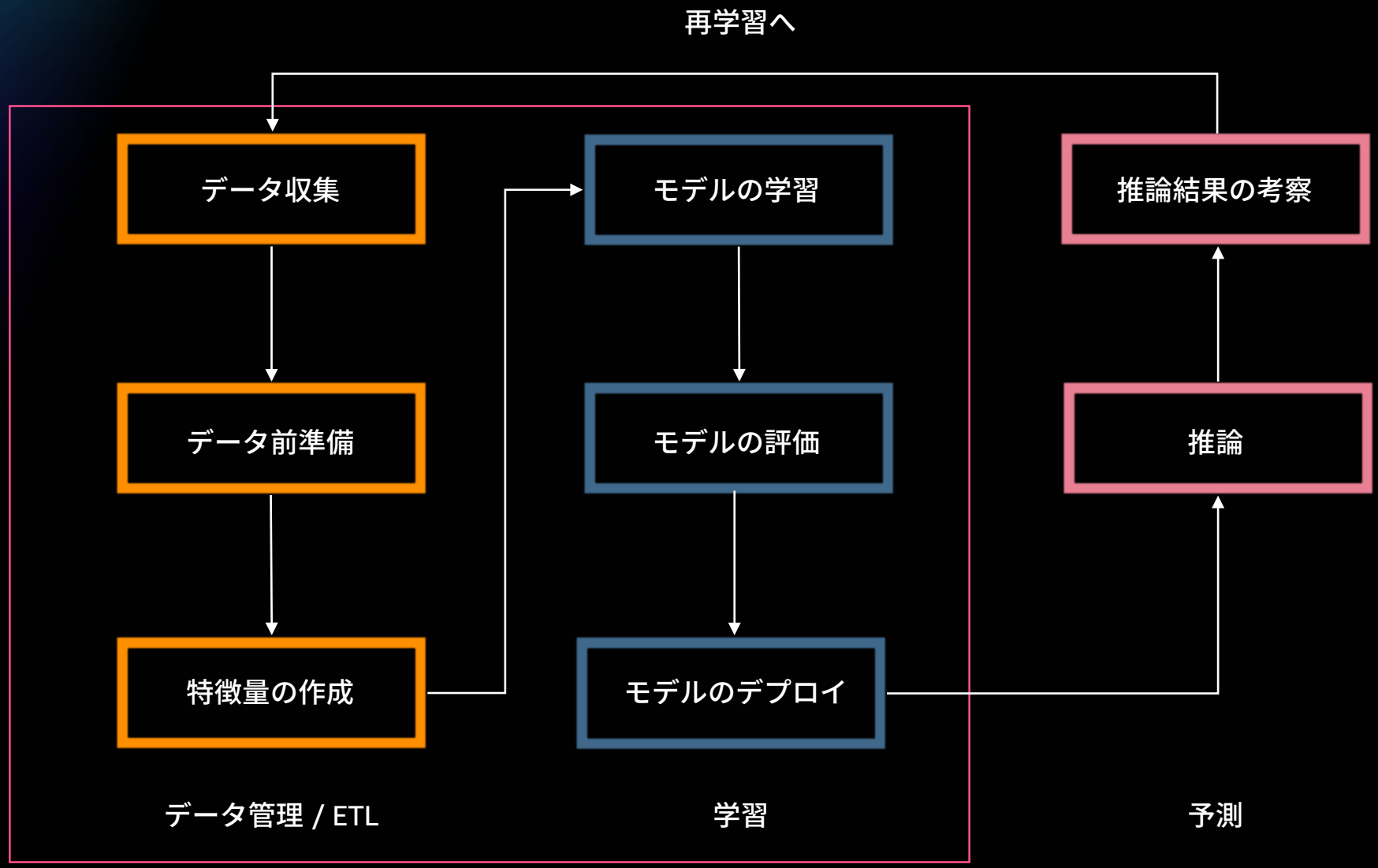
* Amazon Elasticsearch Service の後継サービス

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

サーバーレスで ML Ops

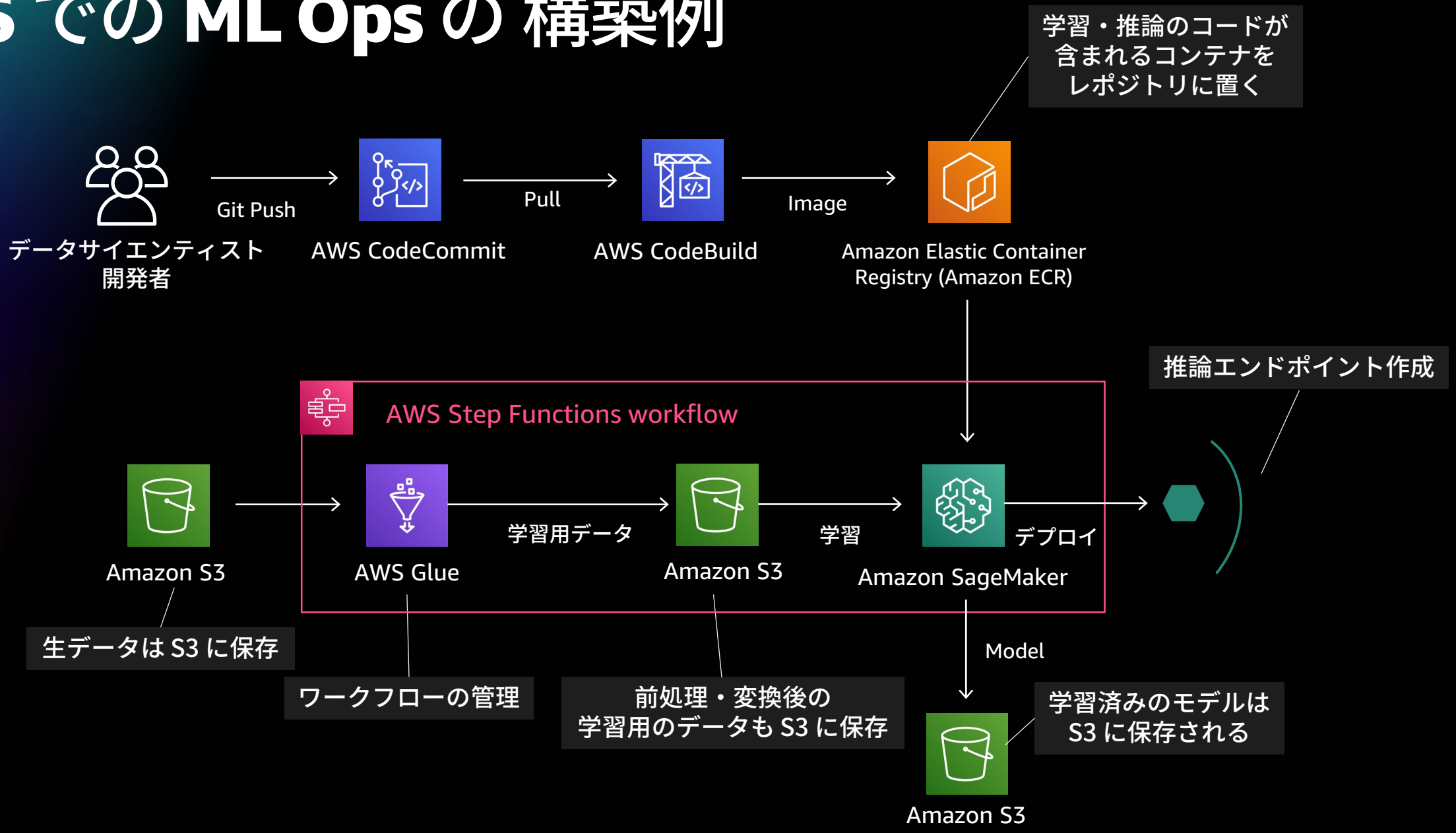
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

機械学習の流れ



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

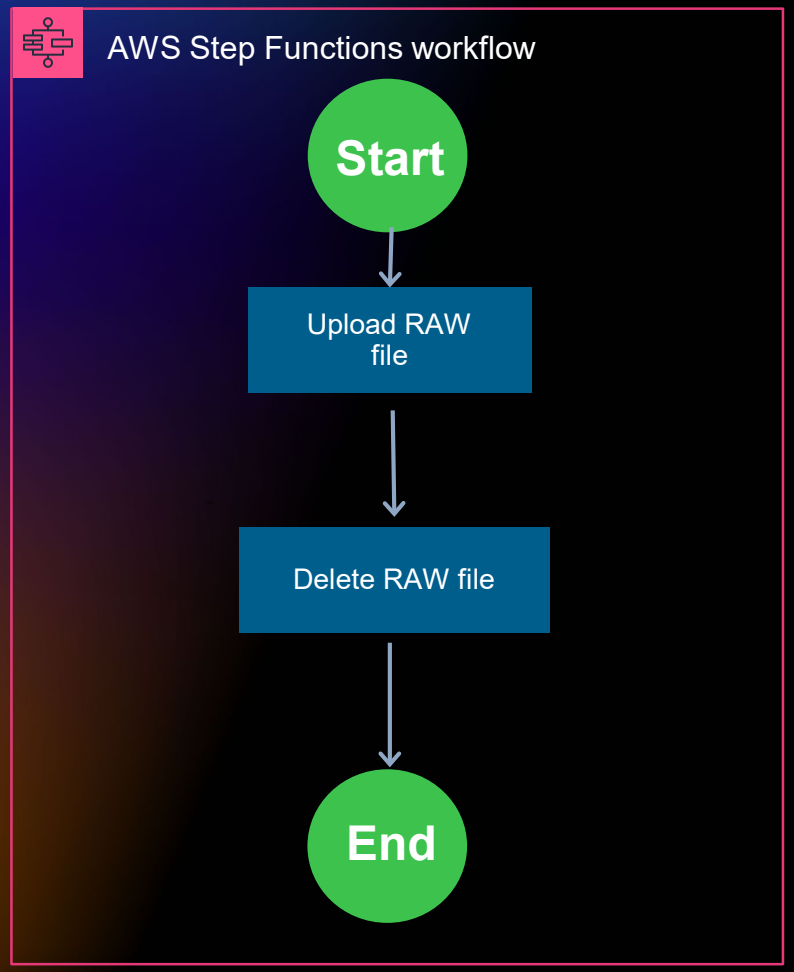
AWS での ML Ops の 構築例



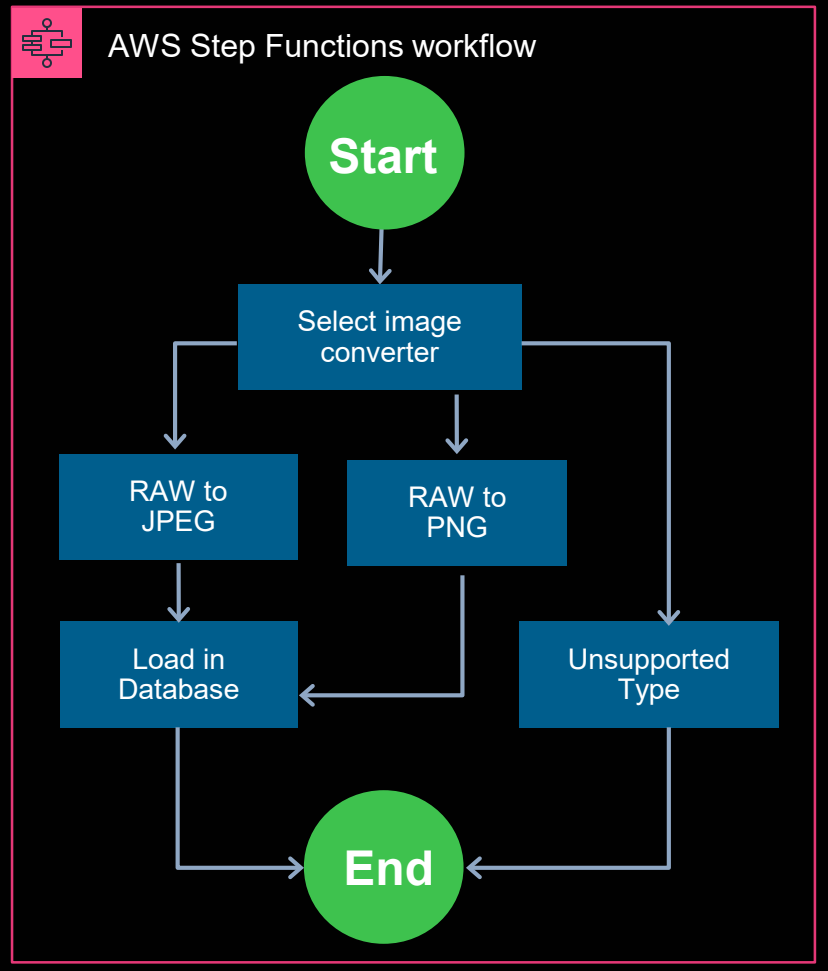
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

ML Ops で必要になるワークフローロジック

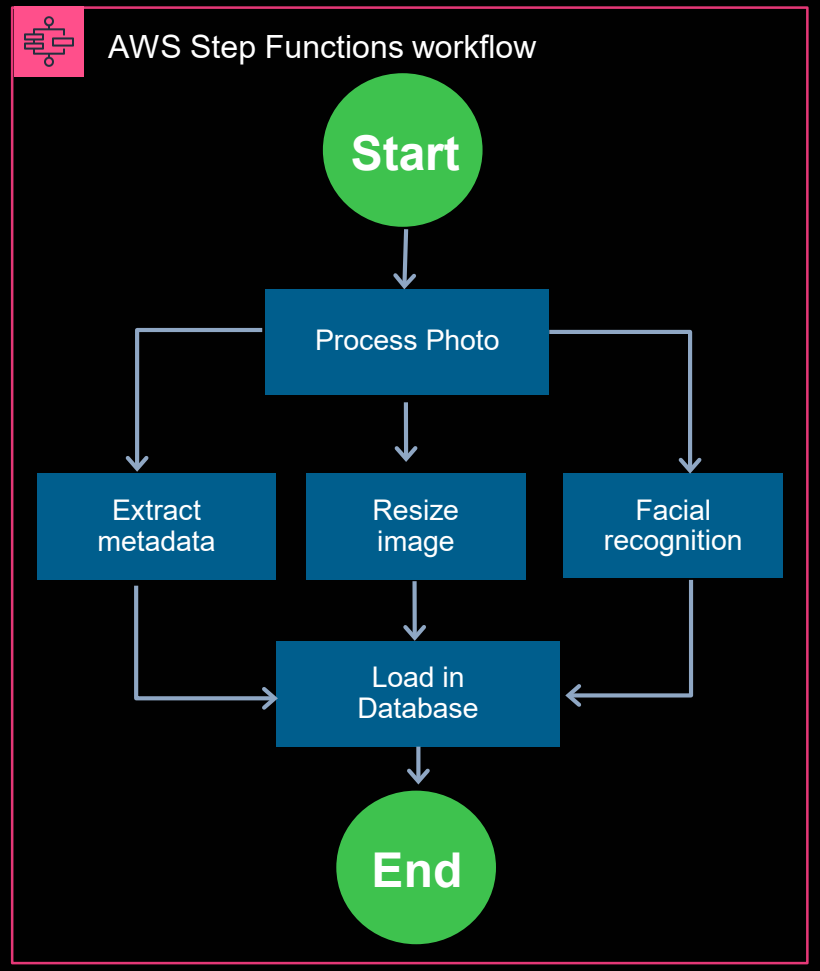
Sequence



Choice



Parallel



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

ワークフローから直接 AWS サービスを利用

Compute

AWS Lambda

AWS Batch

AWS Fargate

Amazon Elastic Container Service (Amazon ECS)

Amazon Elastic Kubernetes Service (Amazon EKS)

Application Integration

Amazon Simple Notification Service (Amazon SNS)

Amazon Simple Queue Service (Amazon SQS)

AWS Step Functions

Amazon EventBridge

Data Science

Amazon EMR

Amazon SageMaker

Amazon Glue

Amazon Athena

Database

Amazon DynamoDB

CI/CD

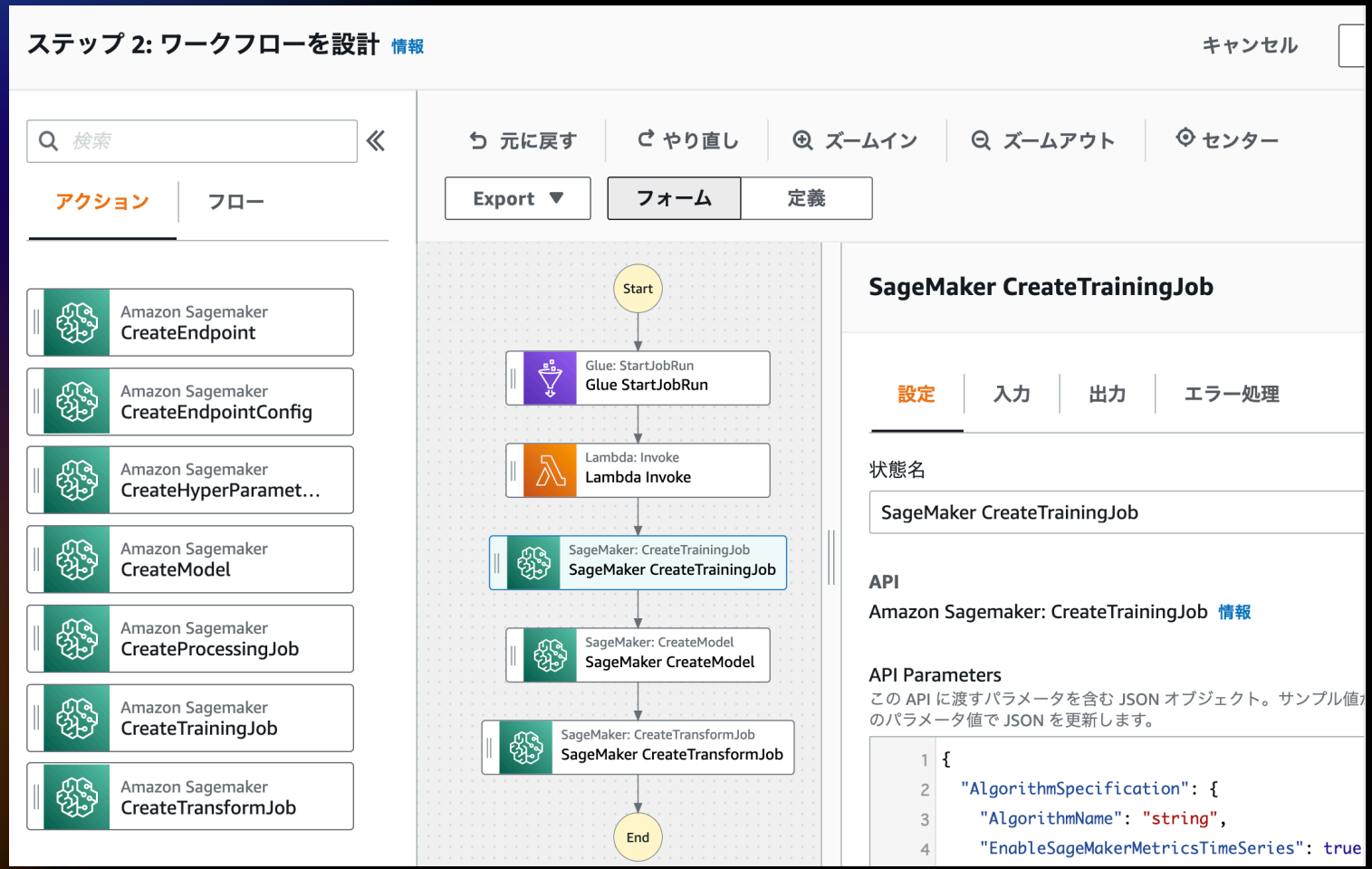
AWS CodeBuild

APIs

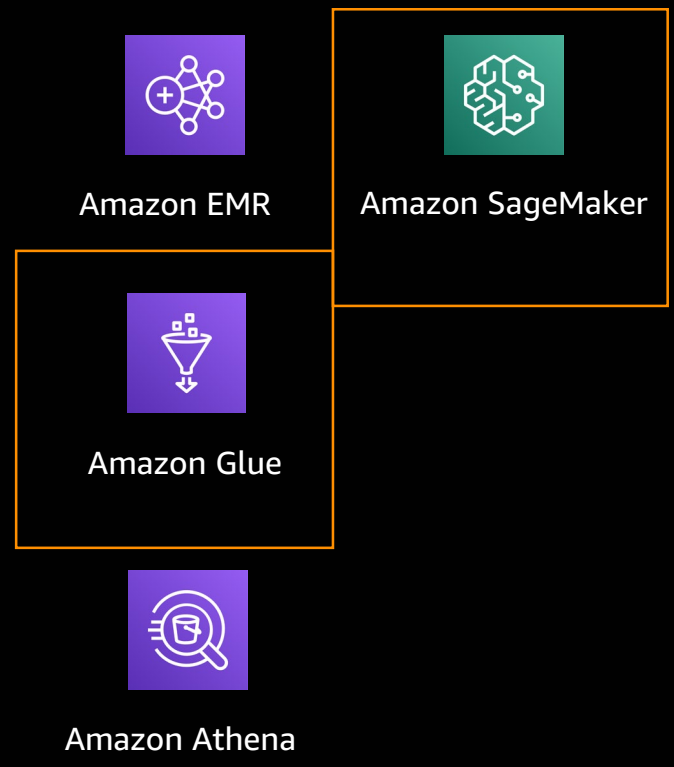
Amazon API Gateway

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Step Functions Workflow Studio による ワークフローの構築



Workflow Studio は、ビジュアルツールを使用してワークフローをより迅速に開発



https://aws.amazon.com/jp/blogs/news/new-aws-step-functions-workflow-studio-a-low-code-visual-tool-for-building-state-machines/

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Step Functions Data Science SDK によるワークフローの構築

AWS Step Functions Data Science SDK により、Pythonで前処理 - 学習 - デプロイ のワークフローを構築

Use Step Functions to run training in SageMaker

The `PyTorch` class allows us to run our training function as a training job on SageMaker. We need to configure it with our training script, an IAM role, the number of training instances, the training instance type, and hyperparameters. In this case we are going to run our training job on 2 `ml.c4.xlarge` instances. But this example can be ran on one or multiple, cpu or gpu instances (full list of available instances). The hyperparameters parameter is a dict of values that will be passed to your training script -- you can see how to access these values in the `mnist.py` script above.

In [ ]:
```
from sagemaker.pytorch import PyTorch

estimator = PyTorch(
    entry_point="mnist.py",
    role=sagemaker_execution_role,
    framework_version="1.8.1",
    train_instance_count=2,
    train_instance_type="ml.c4.xlarge",
    hyperparameters={"epochs": 6, "backend": "gloo"},
)
```

https://aws.amazon.com/jp/about-aws/whats-new/2019/11/introducing-aws-step-functions-data-science-sdk-amazon-sagemaker/

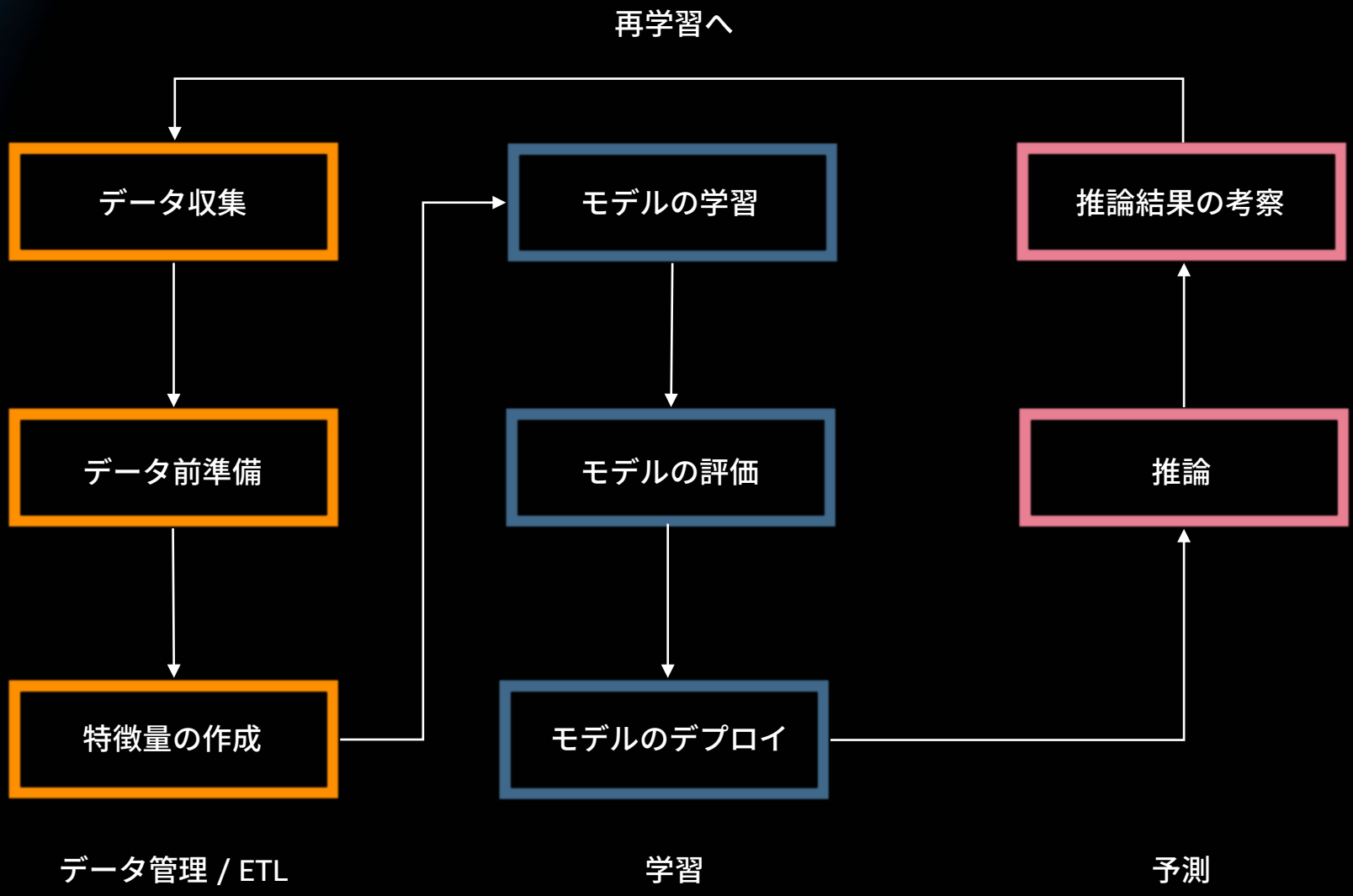
https://github.com/aws/amazon-sagemaker-examples/

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

サーバーレスで 推論エンドポイント構築

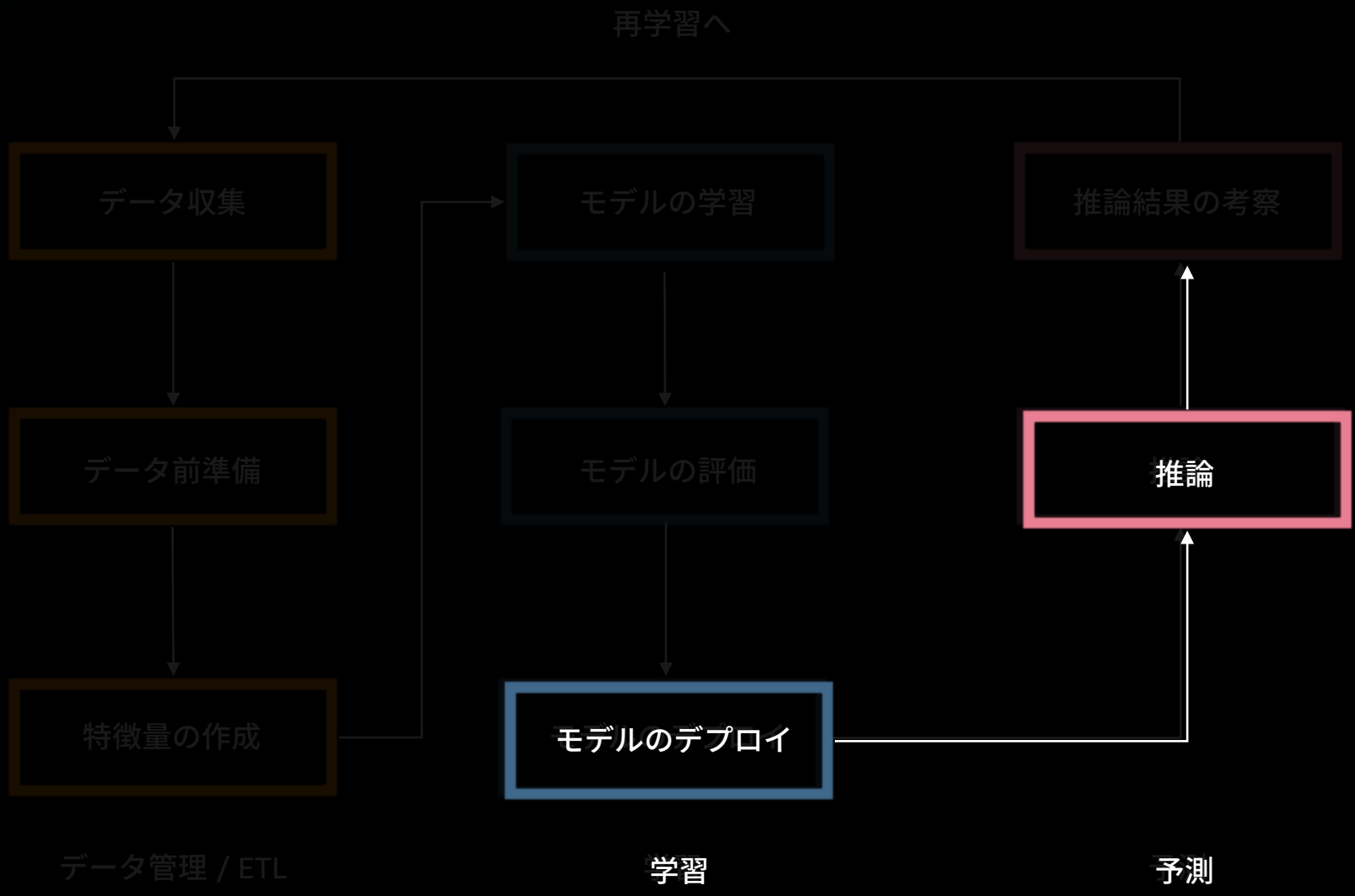
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

機械学習の流れ



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

機械学習の流れ



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

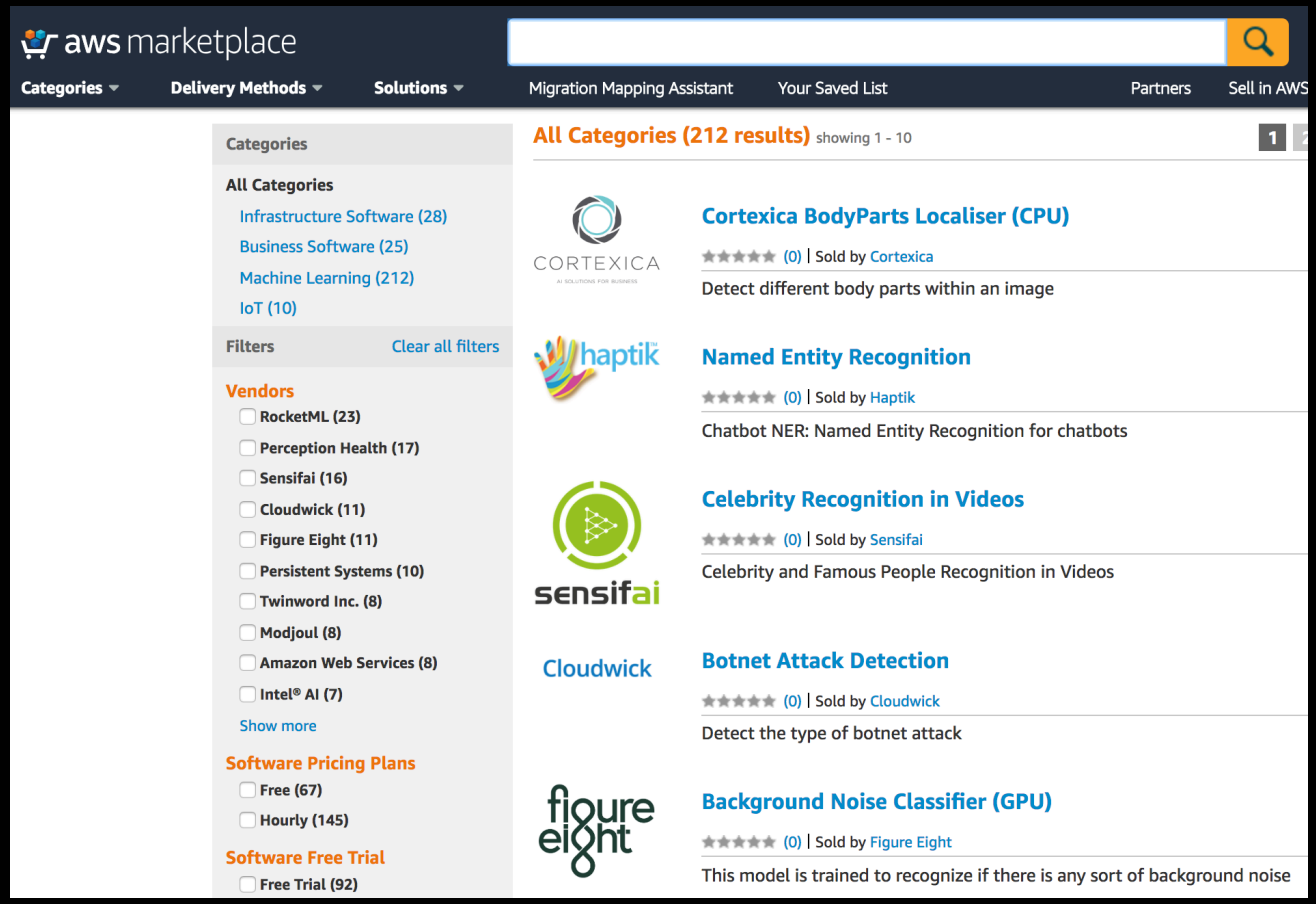
機械学習モデル構築の選択肢

- AWS Marketplace でモデルを購入する
  - SageMaker で利用できる 学習可能なモデル、推論用のエンドポイントを時間単位で購入・利用できる
- 自前の学習用スクリプトを使う
  - TensorFlow, MXNet, PyTorch, Chainer, scikit-learn の 学習・推論用コンテナを提供
- Amazon SageMaker ビルトインアルゴリズムを使う
  - XGBoost，Factorization Machine, Object Detection, Semantic Segmentation など

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Marketplace から機械学習モデルを購入する

SageMaker上で使う機械学習モデルの
売買が可能。小売、メディア向けなど
200 以上のアルゴリズムが公開済み



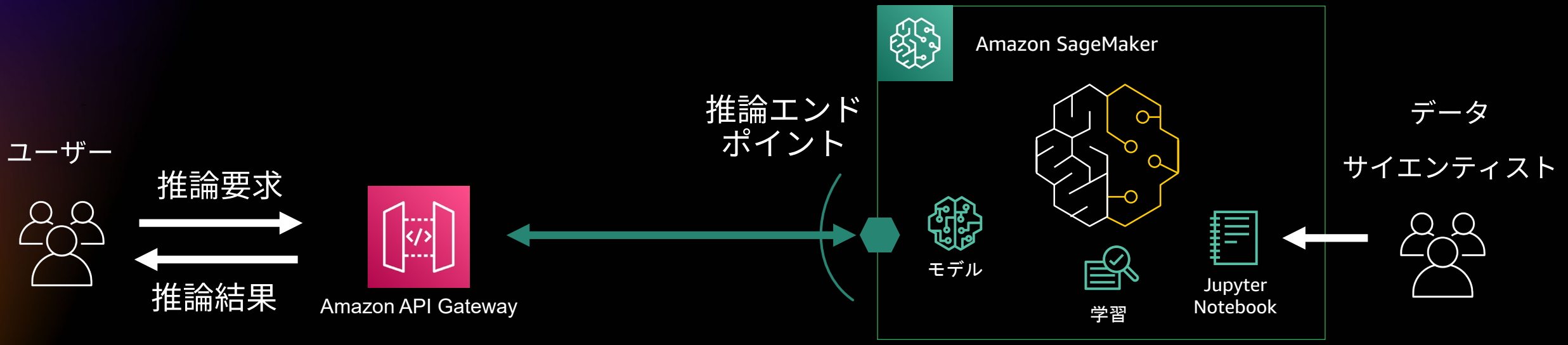
<u>アルゴリズム購入者：</u>
Amazon SageMaker で学習ジョブおよび、
推論エンドポイント（バッチ推論ジョブも）

<u>アルゴリズム販売者：</u>
モデルの中身を秘匿してモデルの出品が可能

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon SageMaker による推論 API 構築

- ニーズにあわせて独自の機械学習サービスを実装する場合 Amazon SageMaker がサービス実装に必要な環境を提供



https://aws.amazon.com/blogs/machine-learning/creating-a-machine-learning-powered-rest-api-with-amazon-api-gateway-mapping-templates-and-amazon-sagemaker/

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

機械学習でのコスト

ディープラーニングを用いたアプリケーションでは、
推論によって、計算コストの90%が発生。

https://aws.amazon.com/blogs/machine-learning/model-serving-with-amazon-elastic-inference/

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

推論エンドポイントのコスト効率について

- 推論エンドポイントを常時起動する場合のコストが気になる

- なぜ、AWS Lambda がコスト効率が良いか？
  - 自動的にゼロへのスケール
  - 1 ms 単位の課金
  - サーバーのパッチ当てなどの管理はクラウドにおまかせ

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

機械学習フレームワークのインストールサイズ

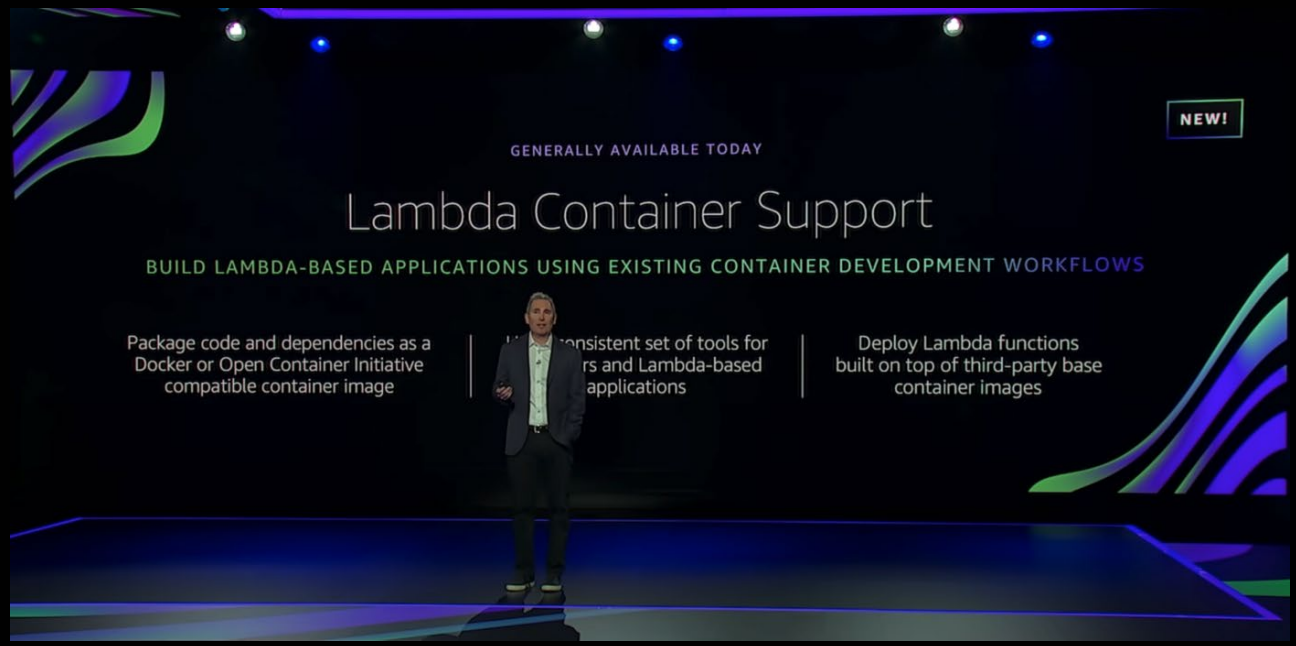
機械学習フレームワーク	インストールサイズ
Tensorflow	1.02 GB
Pytorch	559 MB
Scikit-learn	268 MB
XgBoost	238 MB

Zip形式の Lambda 関数のアーティファクト最大サイズは、250 MB

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Lambda のコンテナサポート

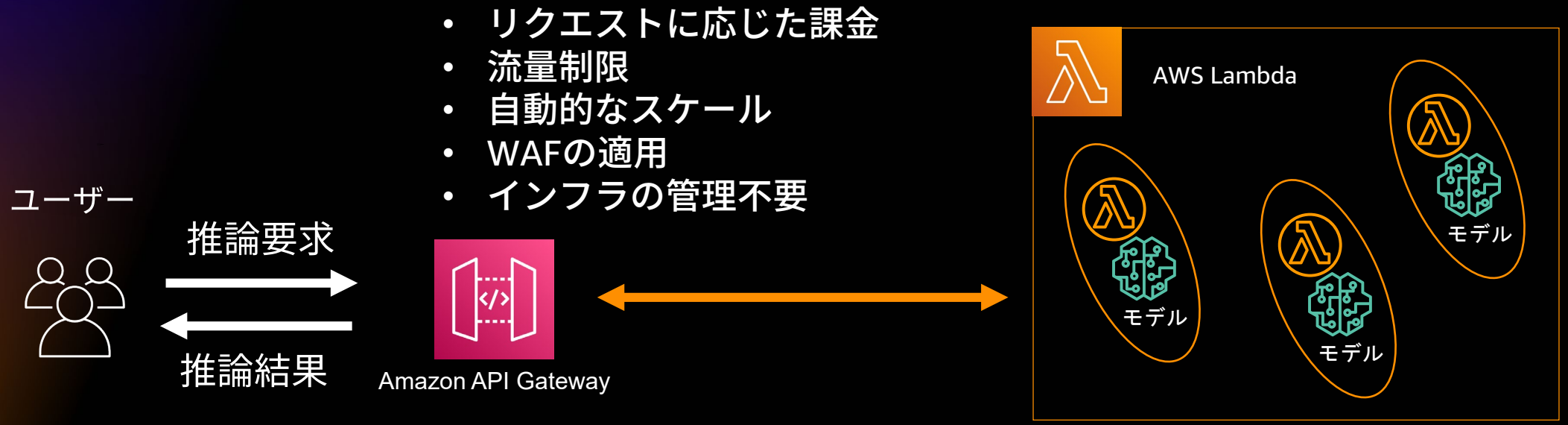
- 10 GB デプロイパッケージ (最大、コンテナイメージのみ)
- 10 GB メモリ(最大)
- 6 vCPU(最大)



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Lambda による推論 API 構築

- Amazon SageMaker で学習したモデルやAWS Marketplaceで購入したモデルを用いる
- AWS Lambda のアーティファクト内にモデルを入れ込む



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

推論 Lambda関数 のチューニング

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Lambda の AVX2 拡張命令 (x86) サポート

- 計算集約型関数のパフォーマンス向上:
  - 機械学習の推論
  - マルチメディア処理
  - HPC
  - 金融モデル計算

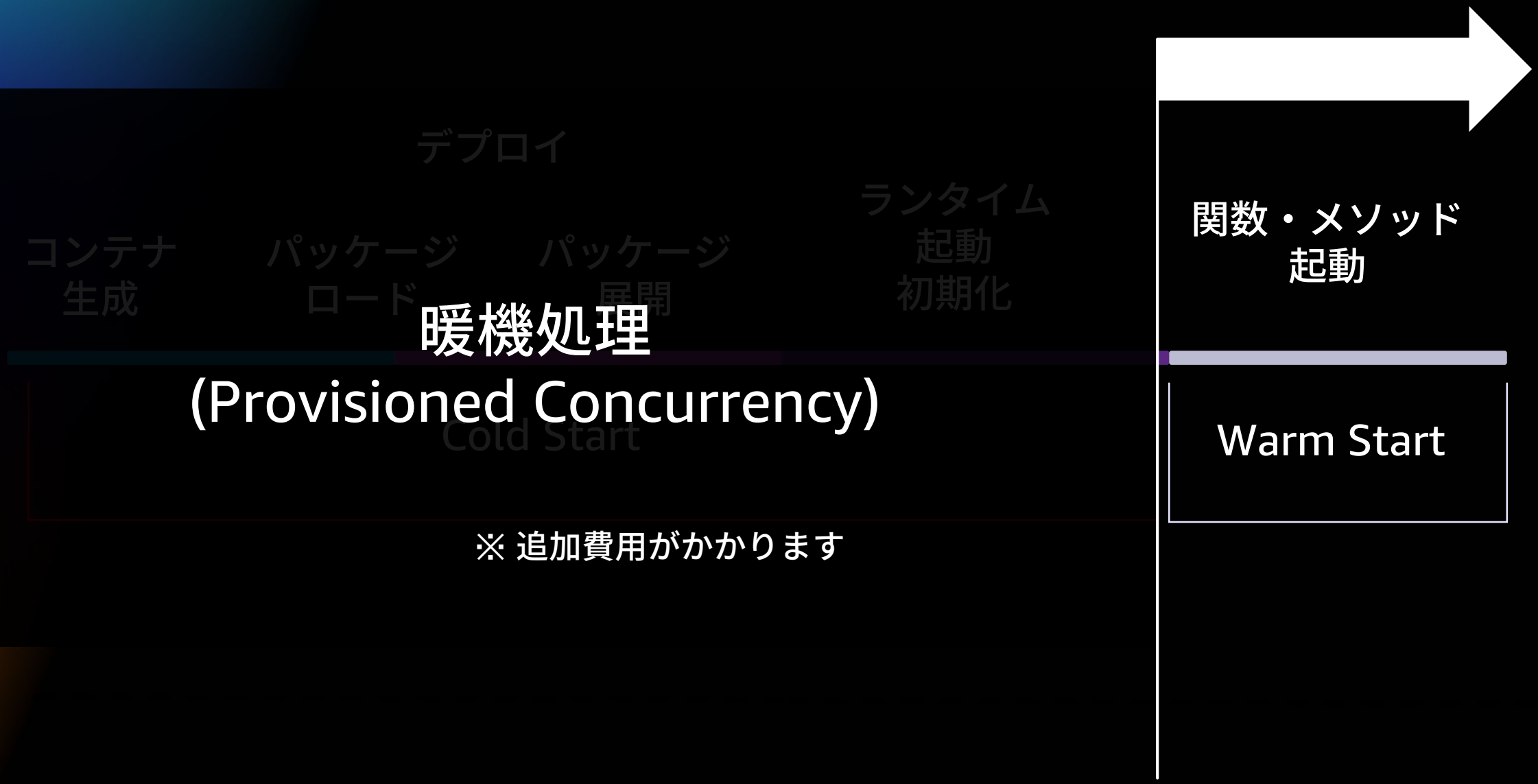
- AVX2 拡張命令セット
  - 利用するためには自身のコードやライブラリが AVX2 命令セットに最適化されている必要があるため注意

https://unsplash.com/photos/IMXhx6qhvf0
. Photo credit: Daniel Seßler.

Filter	Standard	With AVX2	Performance Improvement
1. Bilinear	105 ms	71 ms	32%
2. Bicubic	122 ms	72 ms	40%
3. Lanczos	136 ms	77 ms	43%

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Lambda のコールドスタート抑制

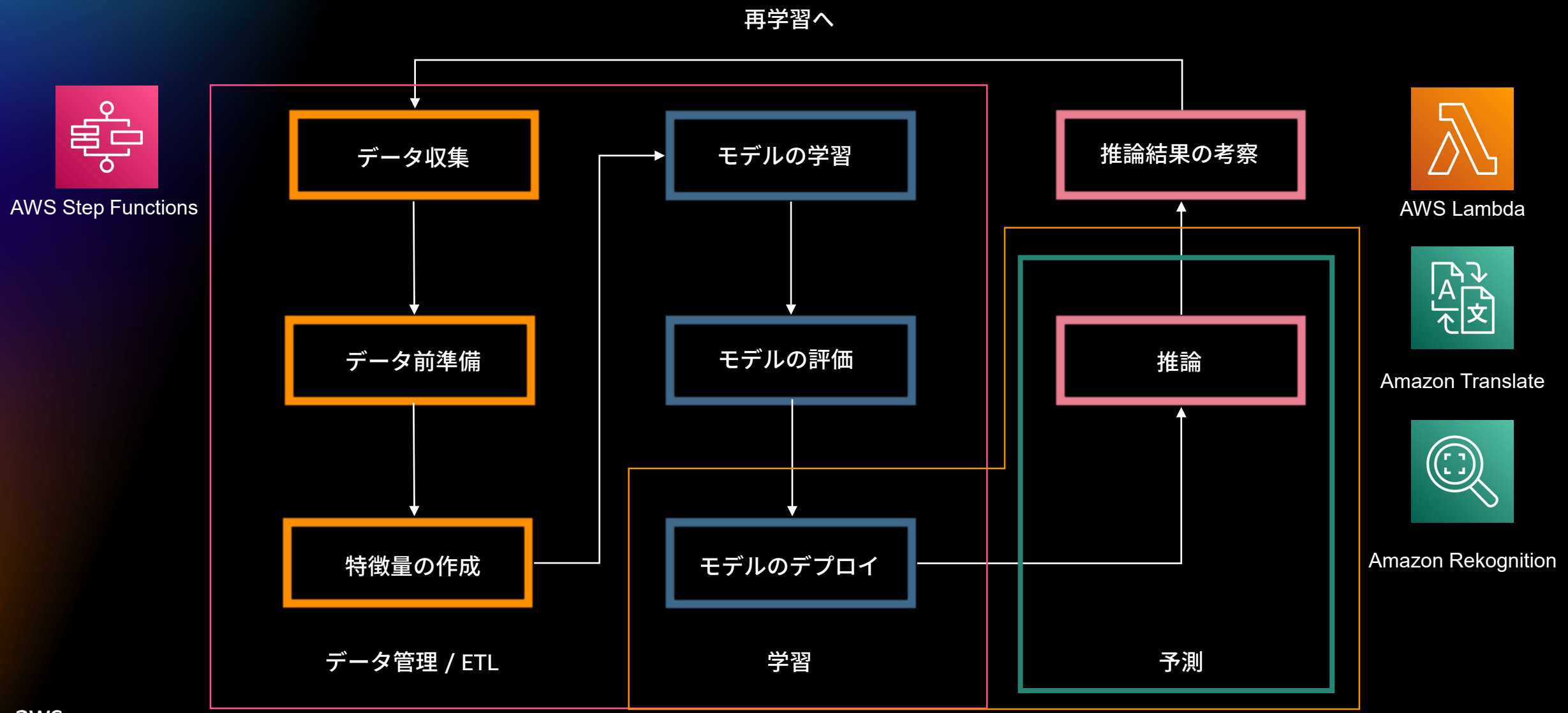


© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

まとめ

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

機械学習の流れ



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

まとめ

- AI サービスで、AWS の機械学習を簡単に試すことが可能
- ビジネス課題から出発し、AWS の様々なサービスを組み合わせて、顧客の価値に貢献する
- サーバーレスサービスを機械学習サービスと組み合わせるとスケーラブルで高可用なサービスを構築できる
- プロダクトを改めて考え、価値を生む箇所に機械学習を

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

ご清聴ありがとうございました

Kensuke Shimokawa

_kensh

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Modern Applications Resource Hub にアクセス

豊富な日本語ガイドで皆様のモダナイゼーションの促進をサポート

- AWS でモダンアプリケーションを構築する
- モバイルアプリ、ウェブアプリを迅速に構築するには
- AWS のコンテナサービスでモダナイゼーションを促進
- サーバーレステクノロジーによる総所有コスト（TCO)
- モダン Dev + Ops モデルの導入

その他にも ユースケースやアナリストレポートを豊富にラインアップ

https://bit.ly/3oVdKPV

Resource Hub はこちら »

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.