

AWS

S U M M I T

Machine Learning on AWS

アマゾンウェブサービス株式会社
ソリューションアーキテクト 志村 誠
2017.06.01



志村 誠 (Makoto Shimura)



所属:

アマゾンウェブサービスジャパン株式会社

業務:

ソリューションアーキテクト
(データサイエンス領域)

経歴:

Hadoop ログ解析基盤の開発

データ分析

データマネジメントや組織のデータ活用

Target / Key takeaway

機械学習を使ってビジネスを加速させたい方

機械学習サービスを AWS 上で構築したい方

AWS 上で機械学習のサービスをどう構築するか

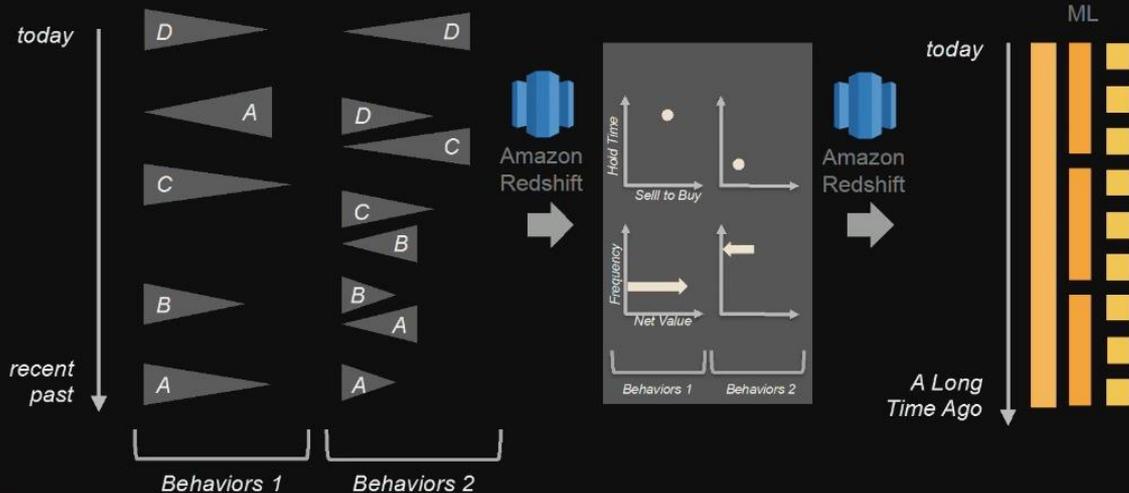
問題ごとに AWS サービスをどう選択するか

業種業態問わずさまざまなビジネスで
機械学習が活用されるようになってきている

Howard Hughes Corporation 利用者の物件購入性向を予測するモデルを構築

Amazon Redshift + Amazon Machine Learning ...better together

Time-series difference analysis example

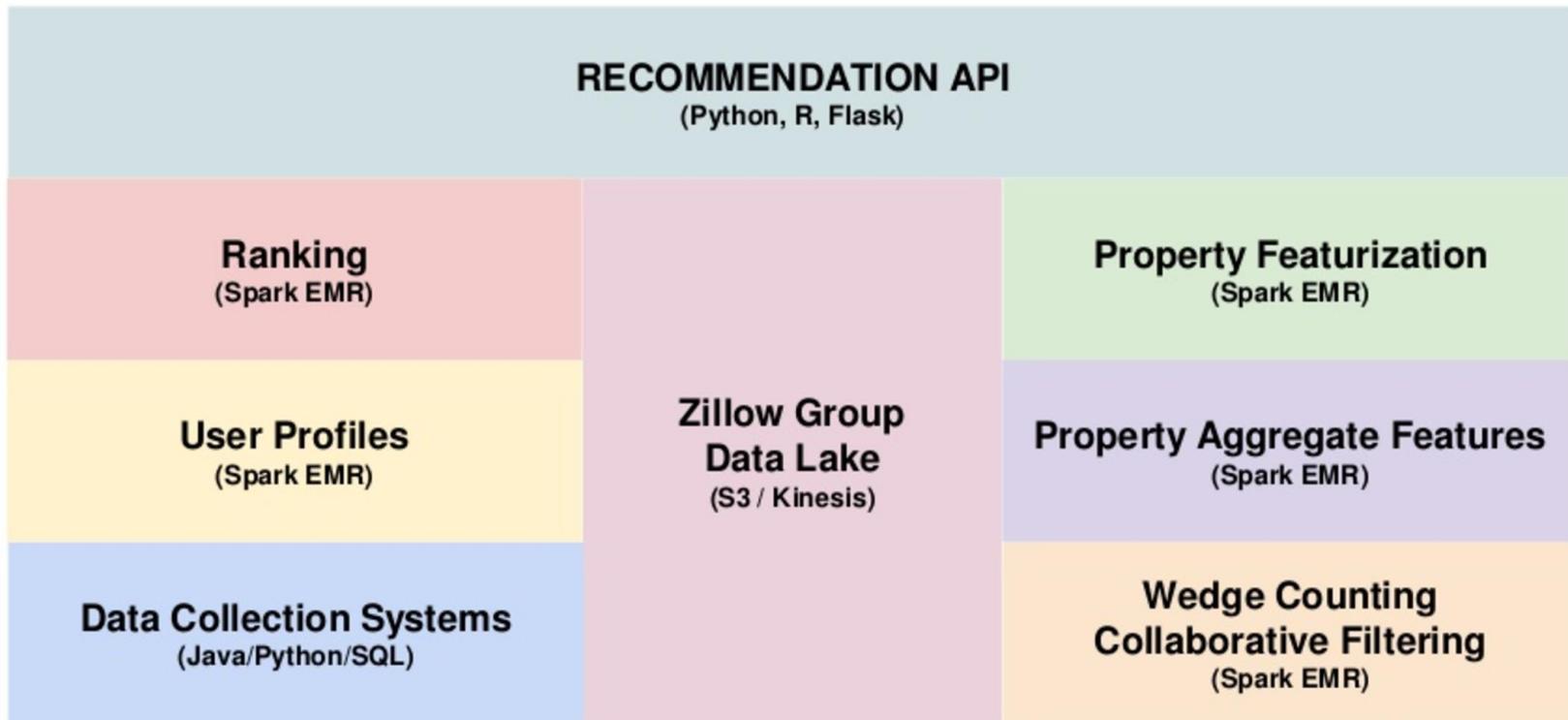


The image shows three screenshots of the Amazon ML Lead Scoring interface:

- 1. Train the Model:** Shows the process of training a model. It includes a 'Train Model' section with 'ML Training Inputs' (Candidate Qualities, Transaction History, 3rd Party Data) and 'General Predictions' (Total Amount of Future Real Estate Purchases). A 'Generic' model is selected.
- 2. Use the Model:** Shows the model being used to predict the rank of candidates. It includes a 'Predicted Rank of Candidates' section with a list of candidates and their predicted ranks (e.g., 5%, 60%, 10%, 0%).
- 3. Review / Refine Model Performance:** Shows the model's performance metrics and a distribution graph of predicted ranks.

Zillow Group

Apache Spark on EMR による各種レコメンデーション



AWS を活用して、機械学習を用いた
サービスを展開するにはどうすればよいか

本セッションのアジェンダ

解くべき問題を明確にする

AWS の機械学習サービス

AWS を活用したサービスの構築

典型的なユースケース

解くべき問題を明確にする

解決したいビジネス課題から出発する



まずは自分たちのビジネスを振り返る



ニーズがあるが実現できていない部分に注目



機械学習から出発するのではなく

あくまで課題を解決するためのツールと考える

機械学習とはなにか

特定事象のデータを学習し、モデルを獲得
それに基づいて判断や予測を実施する

機械学習の一般的な処理の流れ

Training

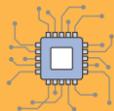
Prediction

機械学習の一般的な処理の流れ

Training



大量の
データ



学習
アルゴ
リズム



機械
学習
モデル

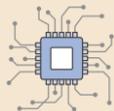
Prediction

機械学習の一般的な処理の流れ

Training



大量の
データ



学習
アルゴ
リズム



機械
学習
モデル

Prediction



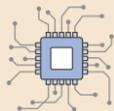
機械
学習
モデル

機械学習の一般的な処理の流れ

Training



大量の
データ



学習
アルゴ
リズム



機械
学習
モデル

Prediction



入力
データ



機械
学習
モデル



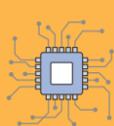
予測
判断
結果

機械学習の一般的な処理の流れ

Training



大量の
データ



学習
アルゴ
リズム



機械
学習
モデル

Prediction



入力
データ



機械
学習
モデル

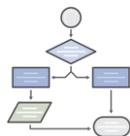


予測
判断
結果

ビジネス課題で機械学習を活用するためのポイント



(大量の)良質なデータでモデルの精度が向上



判断や予測を自動化することが可能

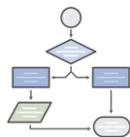


大規模に展開するほど(単位)コストが下がる

ビジネス課題で機械学習を活用するためのポイント



良質なデータが継続的に手に入るか？



自動化する価値のあるクリティカルな予測か？



費用対効果に見合うか？

機械学習に向いている典型的なビジネス課題

レコメンド

人間の手では不可能な量の商品 / ユーザーに対して個別のおすすめを提供したい

異常検知

24時間365日人間が監視するのはコストが非常に高い異常が起きたらすぐに検知して対策をとりたい

画像認識

大量の画像をチェックするのは不可能自動的に顔の位置や不適切な画像か等を判定したい

クラスタリング

ユーザをいくつかのセグメントに分類してセグメント別のマーケティング施策を実施したい

AWS の機械学習サービス

4 層から構成される AWS の機械学習サービス



4 層から構成される AWS の機械学習サービス



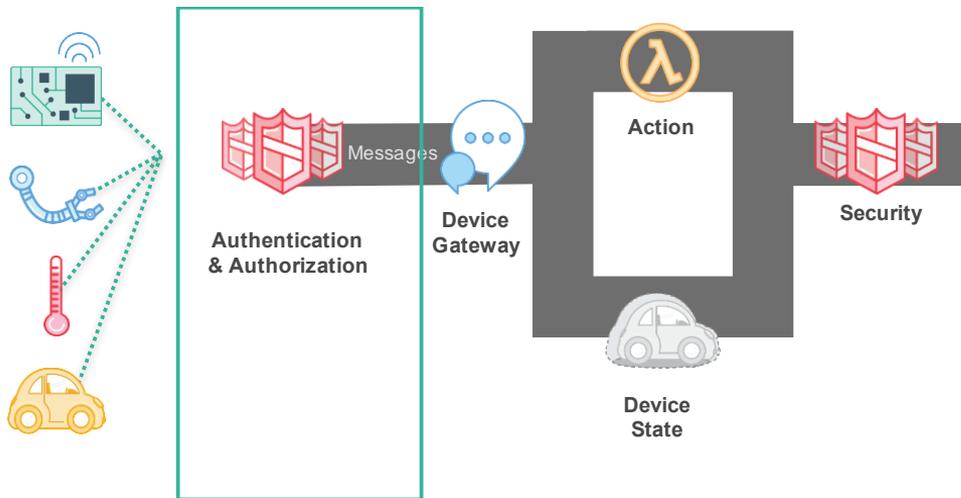
Deep Learningに最適なP2インスタンス

GPU アクセラレータ NVIDIA Tesla K80 を最大 16 個搭載可能
大量の GPU を使うことで、ディープラーニングのモデル構築にかかる時間を大幅に短縮可能

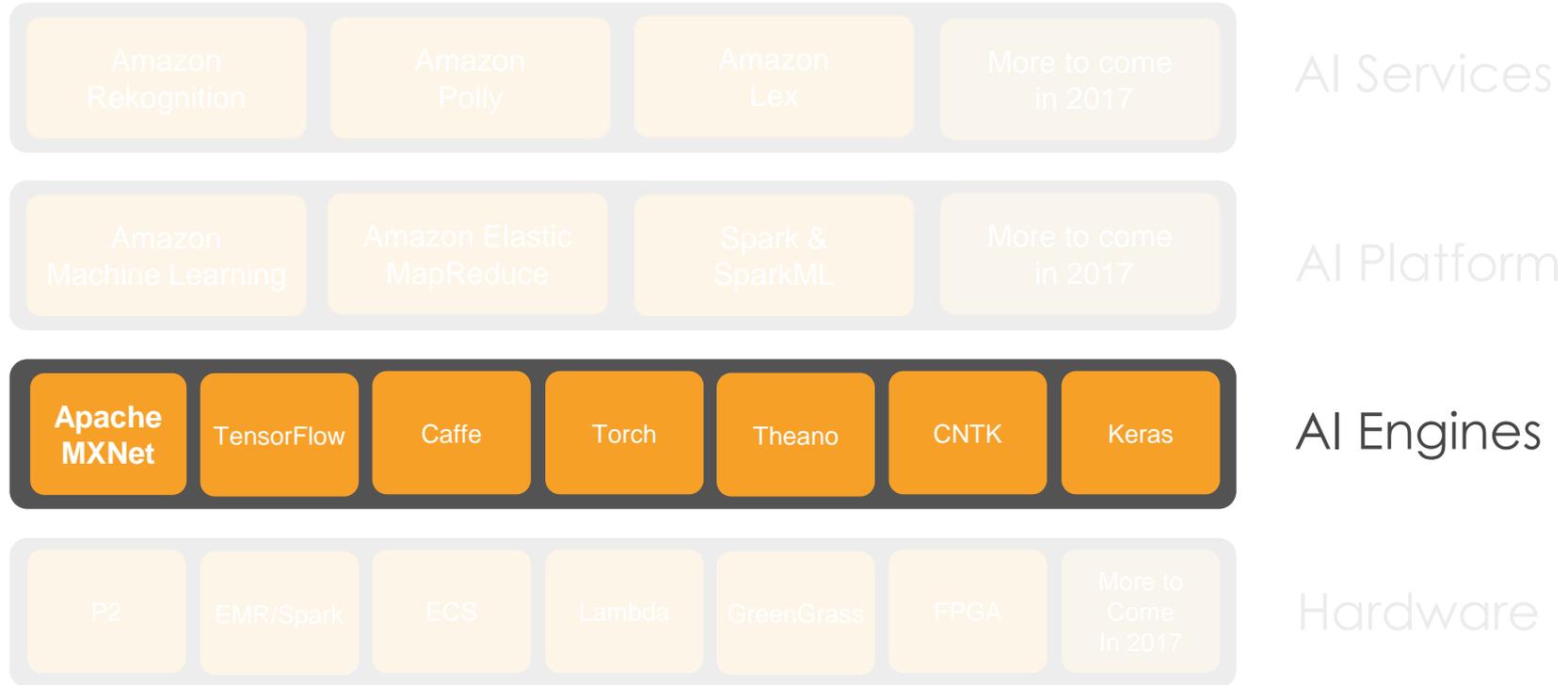
Instance Name	GPU Count	Memory	GPU Memory	Network Performance
P2.xlarge	1	61GiB	12 GiB	High
P2.8xlarge	8	488GiB	96 GiB	10 Gigabit
P2.16xlarge	16	732GiB	192 GiB	20 Gigabit

Greengrass によるエッジコンピューティング

オフラインでニアリアルタイムのローカルイベント処理
AWS Lambda を利用したプログラミングモデル
オンラインになったときにクラウドと同期



4 層から構成される AWS の機械学習サービス



Deep Learning AMI で各種パッケージを手軽に利用

OS はAmazon Linux と
Ubuntu の2種類

各種フレームワークが
プリインストール*

MXNet v0.9.3 tag

Tensorflow v1.0.1 tag

Theano rel-0.8.2 tag

Caffe rc5 tag

Caffe2 v0.6.0 tag

CNTK v2.0 beta 12.0 tag

Torch master branch

Keras 2.0.2 tag

amazon
webservices

Deep Learning AMI Amazon Linux Version

Sold by: Amazon Web Services

The Deep Learning AMI is an Amazon Linux image supported and maintained by Amazon Web Services for use on Amazon Elastic Compute Cloud (Amazon EC2). It is designed to provide a stable, secure, and high performance execution environment for deep learning applications running on Amazon EC2. It includes popular deep learning frameworks, including MXNet, Caffe, Tensorflow, Theano, CNTK and Torch as well as packages that enable easy integration with AWS, including launch configuration tools and many popular AWS libraries and tools. It also includes the Anaconda Data Science Platform for Python2... [Read more](#)

Customer Rating	★★★★☆ (11 Customer Reviews)
Latest Version	2.0 (Other available versions)
Operating System	Linux/Unix, Amazon Linux 2015.05.04
Delivery Method	64-bit Amazon Machine Image (AMI) (Read more)
Support	See details below
AWS Services Required	Amazon EC2, Amazon EBS

Highlights

- New in Version 2.0 - CPU Instance Type Support, MXNet built with MKL support.
- 6 Deep Learning Frameworks - contains the most popular Deep Learning Frameworks (MXNet, Caffe, Tensorflow, Theano, Torch and CNTK) all prebuilt and pre-installed.
- Pre-installed components to speed productivity include Nvidia drivers, CUDA, cuDNN, Anaconda, Python2 and Python3.

Product Description

The Deep Learning AMI is an Amazon Linux image supported and maintained by Amazon

Continue You will have an opportunity to review your order before launching or being charged.

Pricing Information

Use the Region dropdown selector to see software and infrastructure pricing information for the chosen AWS region.

For Region

US East (N. Virginia)

Free Tier Eligible EC2 charges for Micro instances are free for up to **750 hours** a month if you qualify for the **AWS Free Tier**.

Additional Taxes May Apply

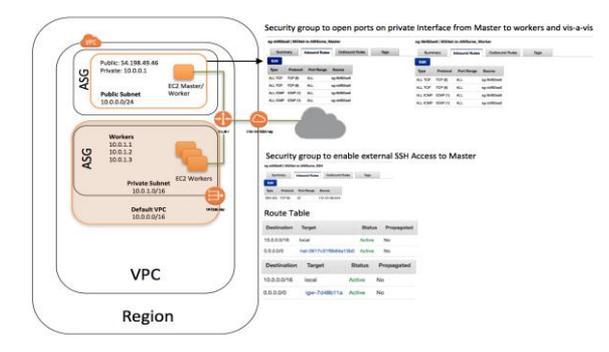
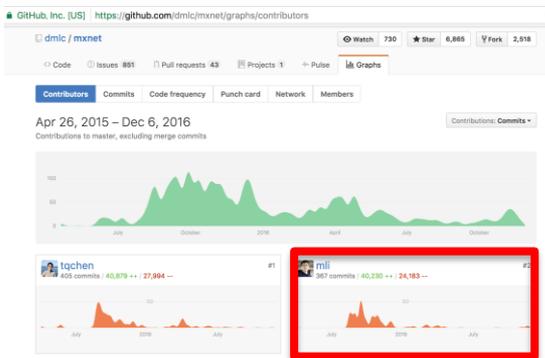
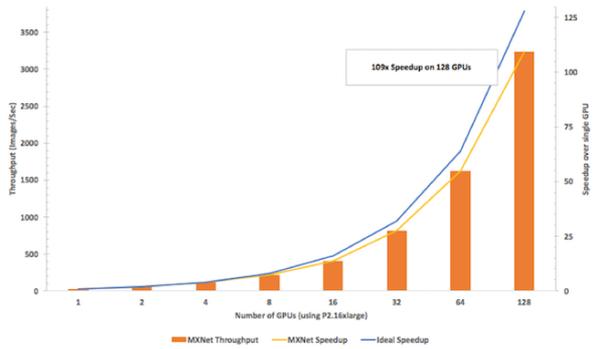
Pricing Details

Software pricing is based on your chosen options, such as subscription term and AWS region. Infrastructure prices are estimates only. Final prices will be calculated according to actual usage and reflected on your monthly report.

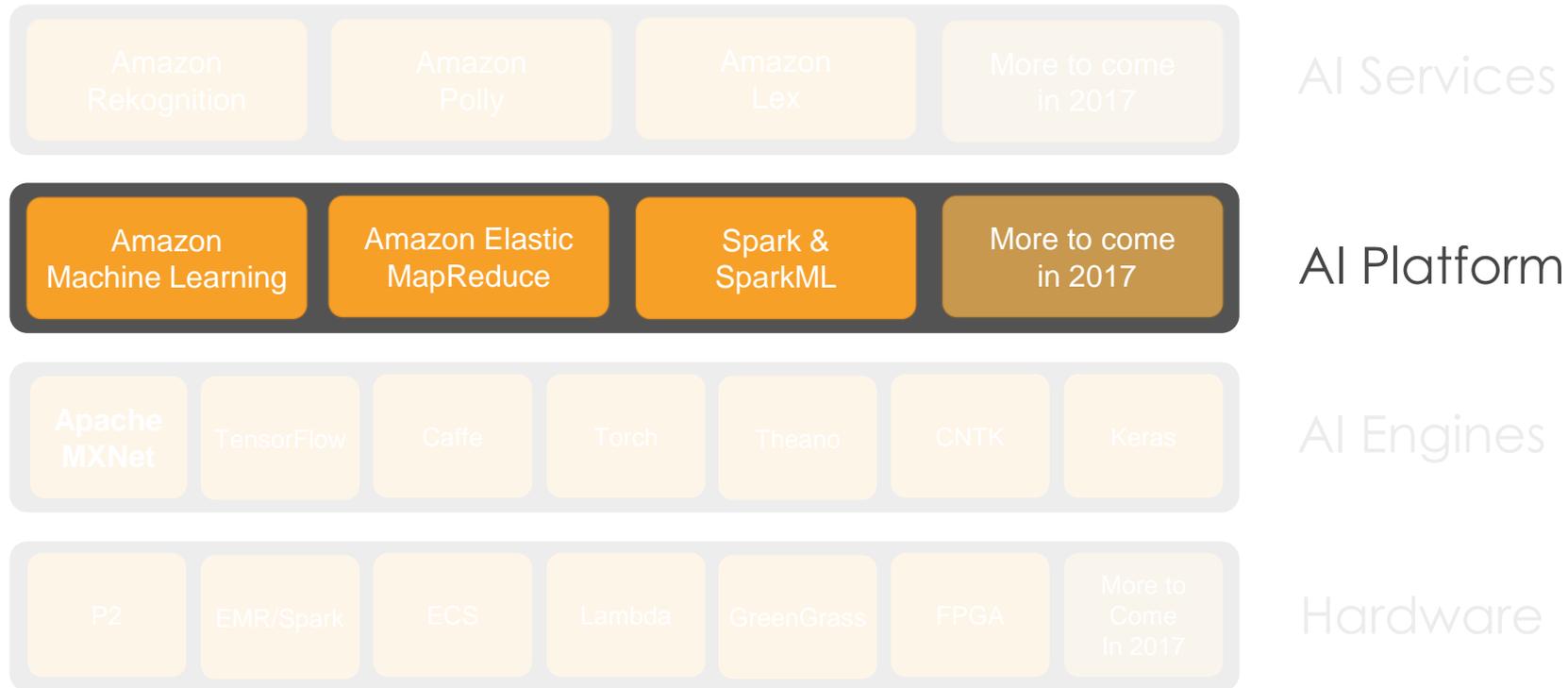
AWS は MXNet を全面的にサポート

マルチ GPU でのスケーラビリティに優れている
AWS 社員もコミッターとして活躍

CloudFormation で MXNet クラスタを構築するテンプレートを
提供



4 層から構成される AWS の機械学習サービス



Amazon machine Learning

スケーラビリティ:

システムの拡張や運用を考える必要がない

パッケージ化:

テストデータ分割や評価手法, 予測 API の提供などが組み込まれている

アプリケーション



Amazon Redshift



Redshift上の取引データを使ってAmazon MLで不正取引の抽出を行い、S3に出力される結果をRedshiftに戻して利用。

2クラス分類

ロジスティック回帰

多クラス分類

多項ロジスティック回帰

回帰分析

線形回帰

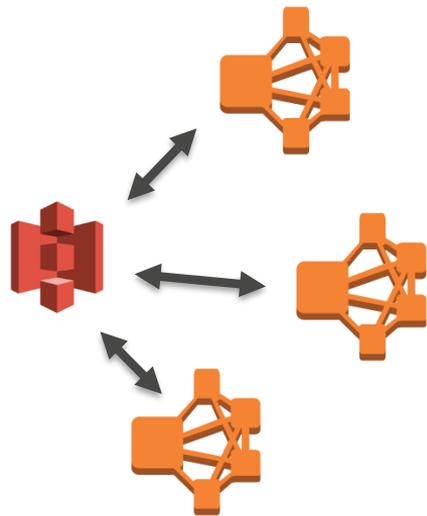
Amazon EMR (Elastic MapReduce)

フルマネージドな Hadoop を提供しており、
クラスターの構築、構成変更、破棄まで管理可能

HDFS ではなく S3 上のデータを直接読み書きすることで、**コンピュートとストレージを分離**可能

ワークロードに応じたクラスターを複数立ち上げて、
同一データを処理可能

Spark MLlib や Mahout でスケーラブル機械学習



4 層から構成される AWS の機械学習サービス

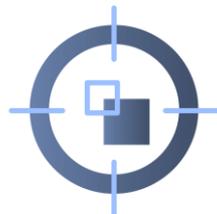


AI Services



Polly

文章をリアルな音声
に変換するサービス



Rekognition

画像分析機能をアプリケーション
に簡単に追加できるようにする
サービス



Lex

声やテキストを使用した会話
型インターフェイスを様々な
アプリケーションに構築する
ためのサービス

これらは、fine-tuning された AWS が提供するディープラーニングモデルを
API で簡単に利用できるマネージドサービスとなります

そのほかのサービスも合わせてシステムを構築

機械学習用途にも利用できるサービス



Amazon Kinesis Analytics

異常検知アルゴリズムをそなえた RUNDOM_CUT_FOREST 関数



Amazon Elasticsearch Service

more like this のように、検索機能を機械学習用途で利用可能

機械学習のサイクルを構成するサービス



Amazon Data Pipeline

スケジュールを組んで EMR のジョブを実行



Amazon API Gateway

機械学習の結果を提供するための API を作成

AWS を活用したサービスの構築

サービス選択における指針

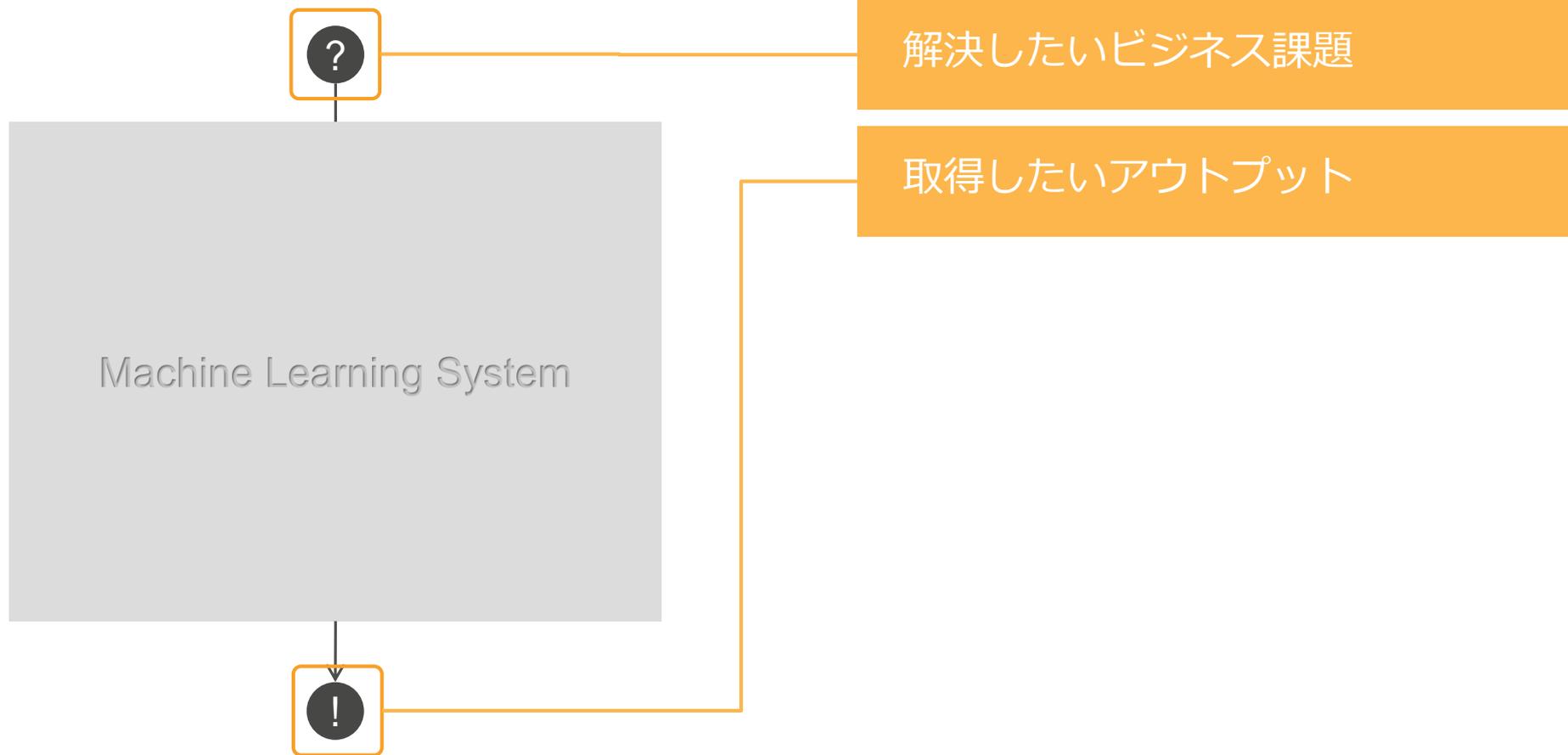
Right tools for the job

機能的特性と性能的特性の両面から、最適なサービスを見極める

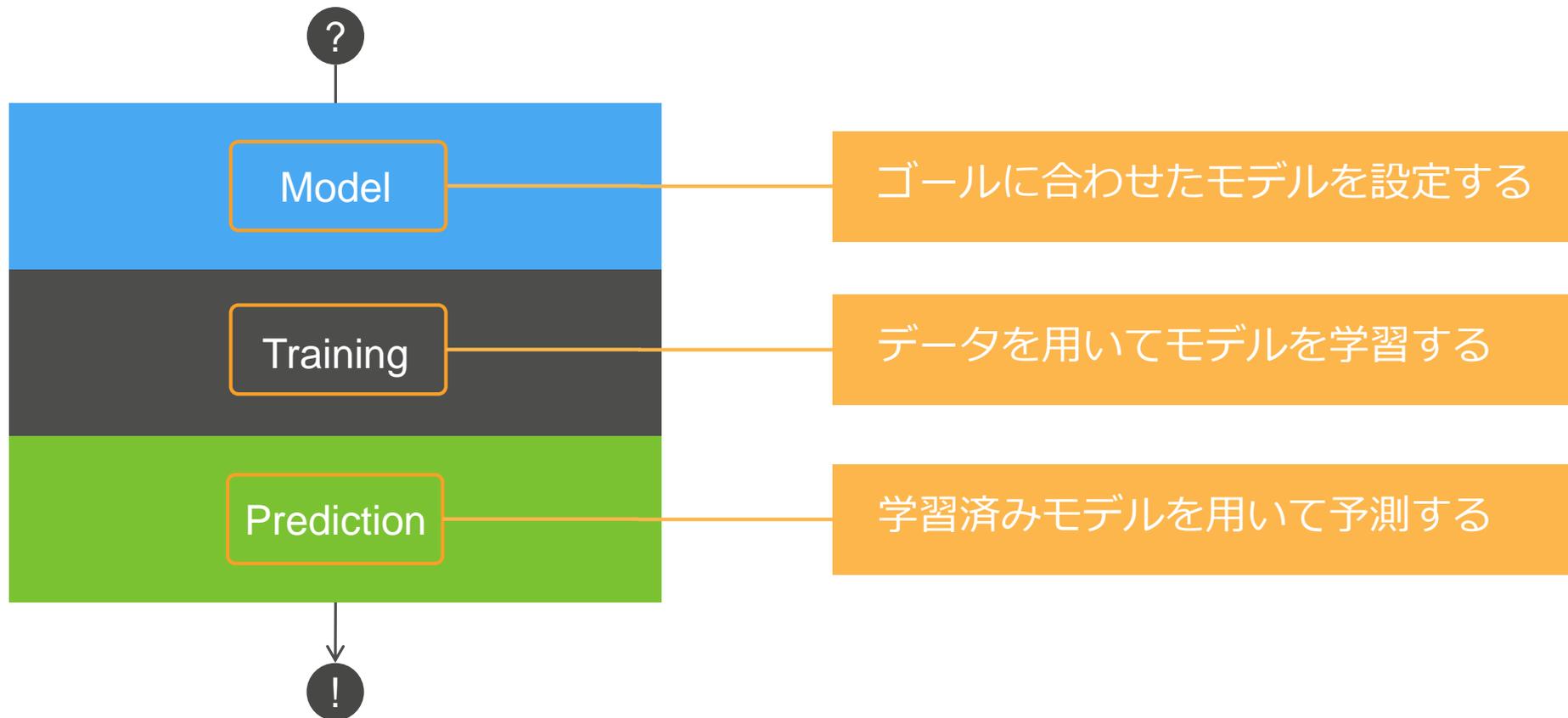
Managed services first

インフラの運用を AWS に任せて、アルゴリズムやアプリケーションに集中

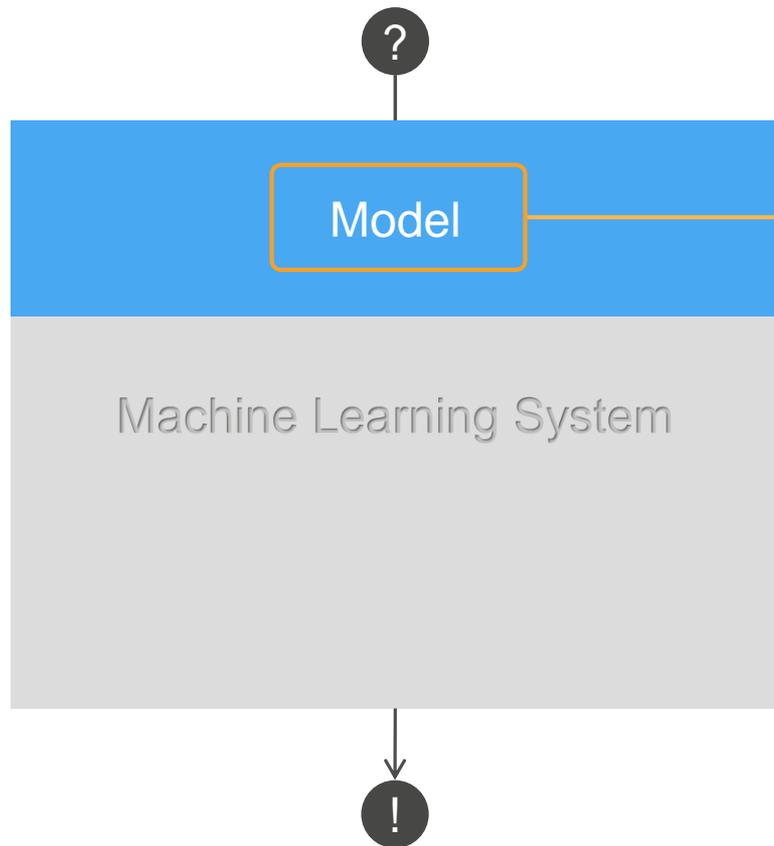
ゴールを明確にする



機械学習における 3 つのステージ



適切なモデルを選択する

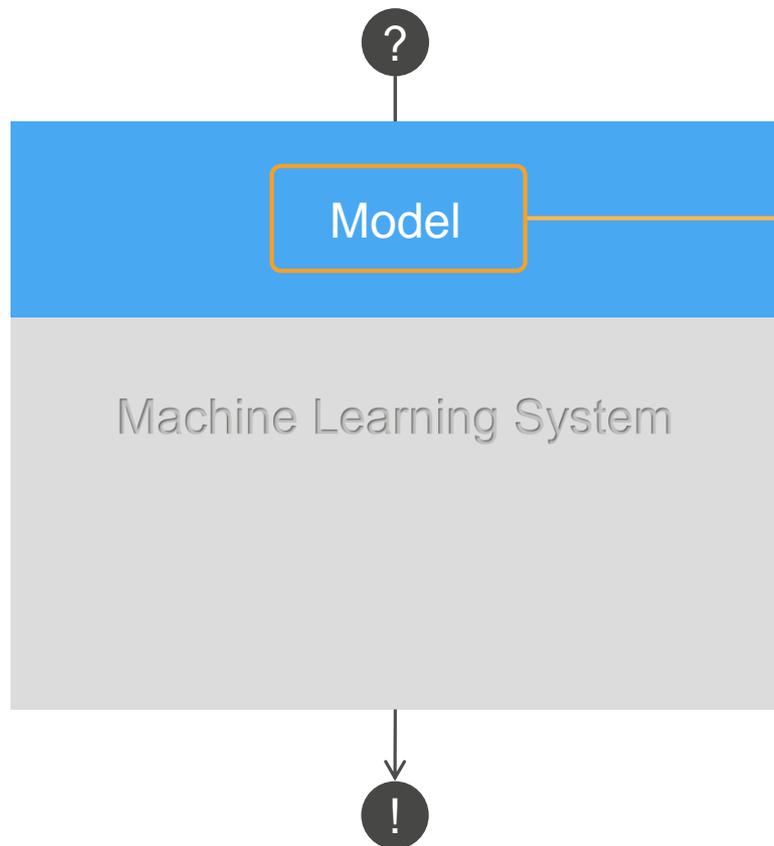


レコメンド

異常検知

分類問題

適切なモデルを選択する



レコメンド

協調フィルタリング

コンテンツベース

ルールベース



Spark MLlib or Mahout
on Amazon EMR

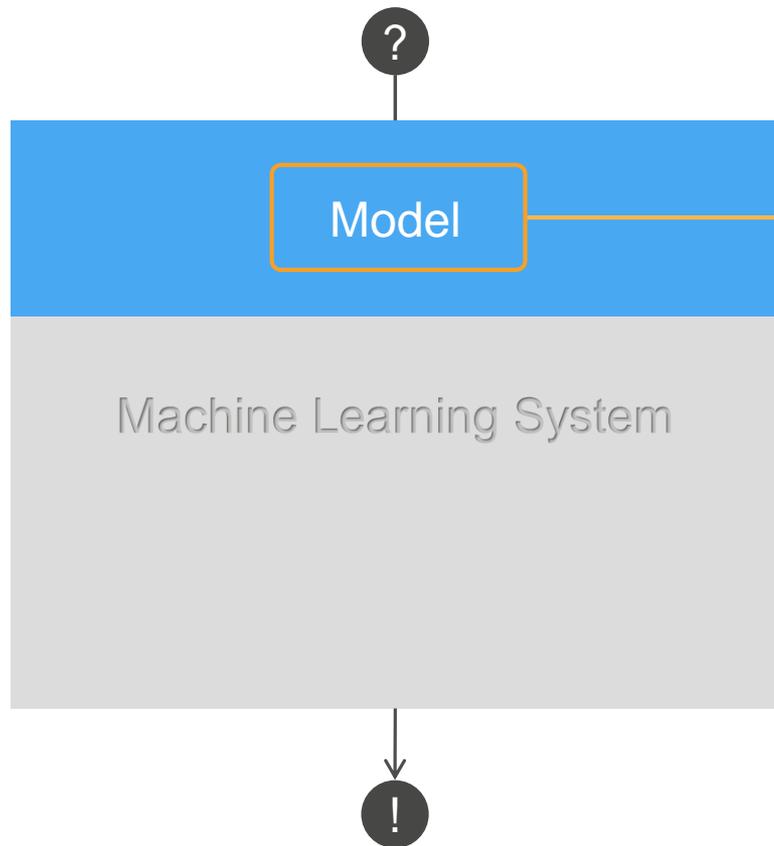


Amazon Elasticsearch Service



Amazon EC2

適切なモデルを選択する



異常検知

K-means

ChangeFinder

Random Cut Forest



Spark streaming
on Amazon EMR

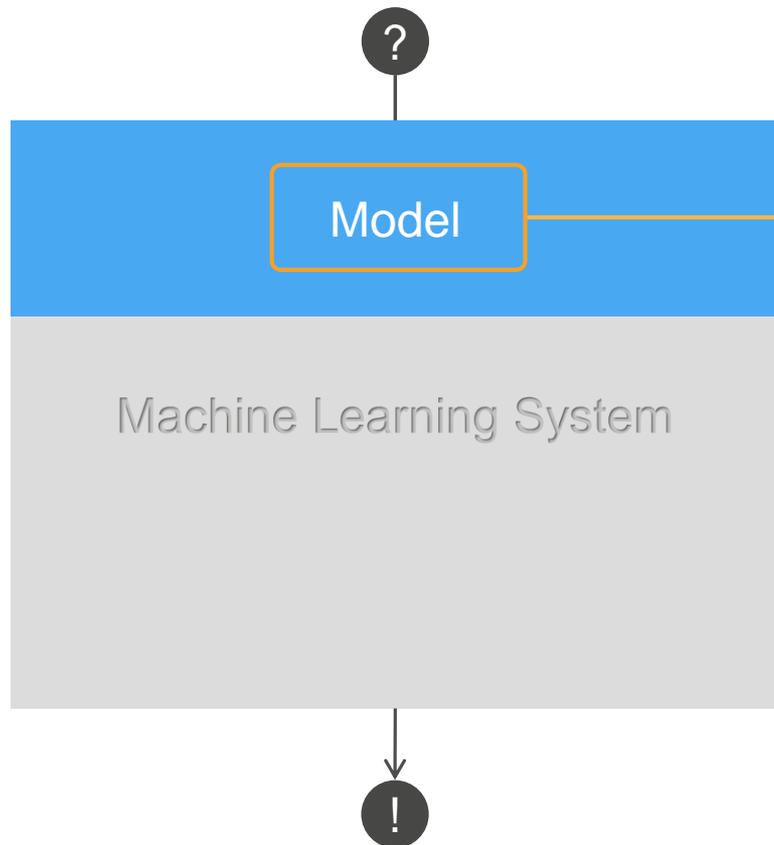


Storm on Amazon EC2



Amazon Kinesis Analytics

適切なモデルを選択する



分類問題

Logistic Regression

Deep Learning

SCW



Amazon Machine Learning



Spark on Amazon EMR



Amazon Rekognition



Amazon Polly

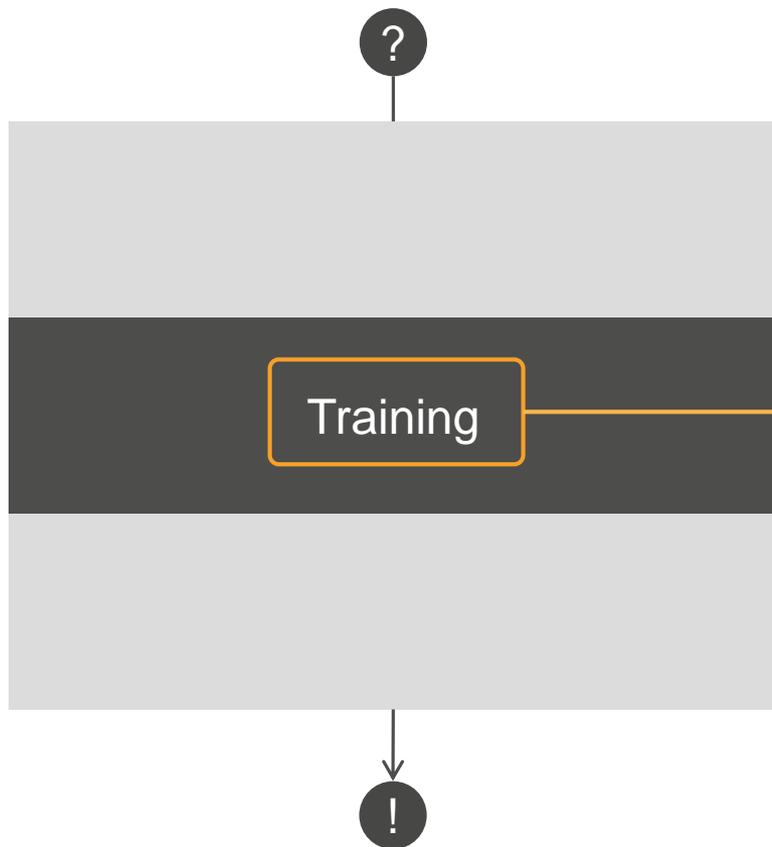


MXNet on Amazon EC2



Hivemall on Amazon EMR

データからモデルを学習する



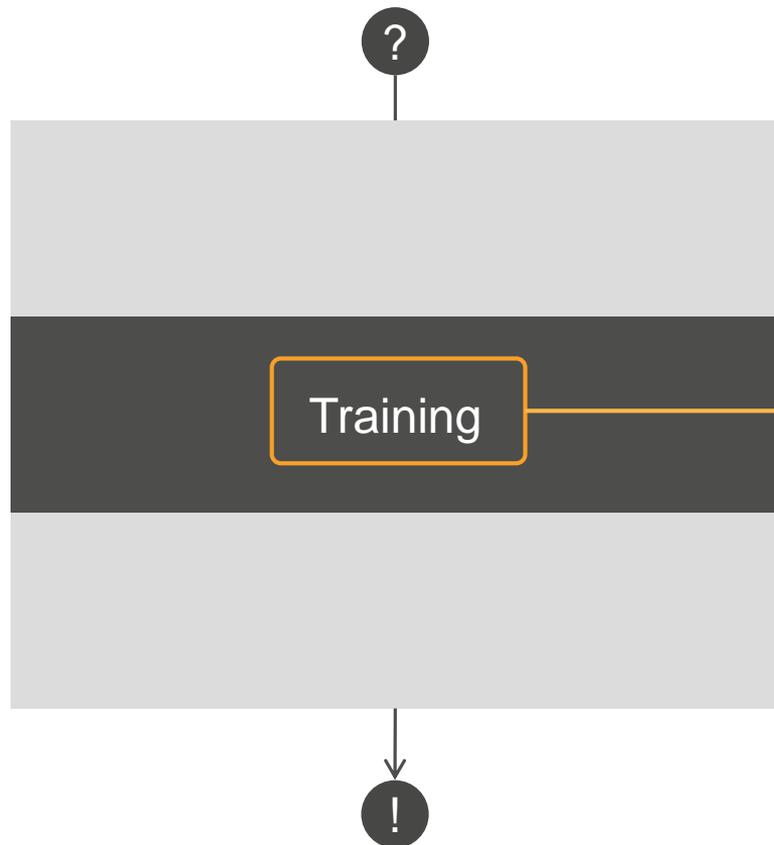
データサイズ

データの場所

モデルの更新頻度

ハードウェア特性

データからモデルを学習する



データサイズ

単一のインスタンスで扱えるサイズ

クラスタで処理する必要があるサイズ



Amazon EC2

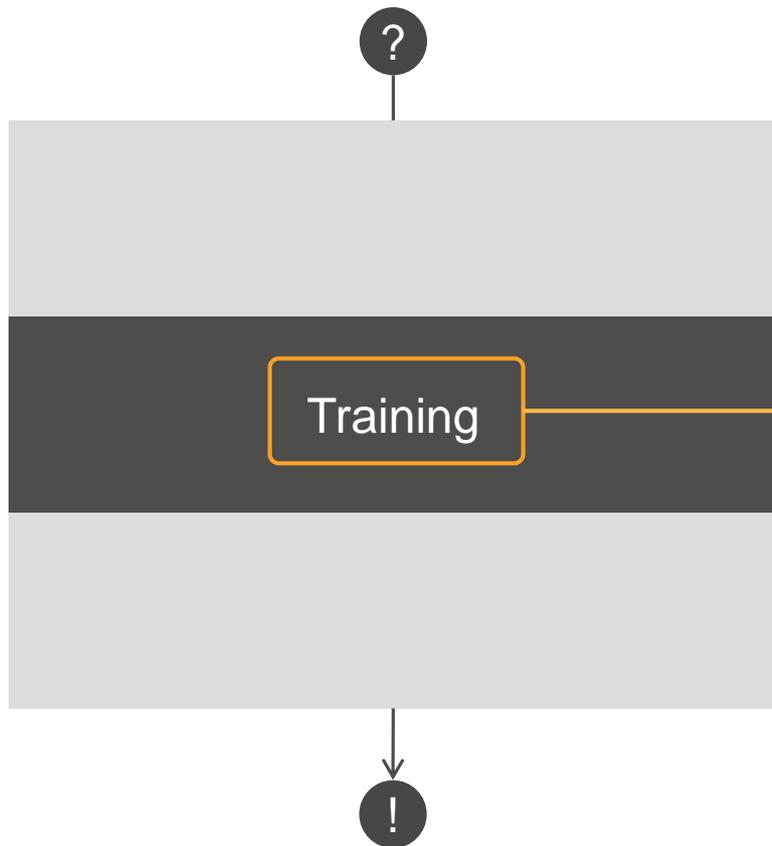


Spark on Amazon EMR



Amazon Machine Learning

データからモデルを学習する



データの場所

ストレージに保存されている

DB に保存されている

ストリームで流れてくる



Amazon S3

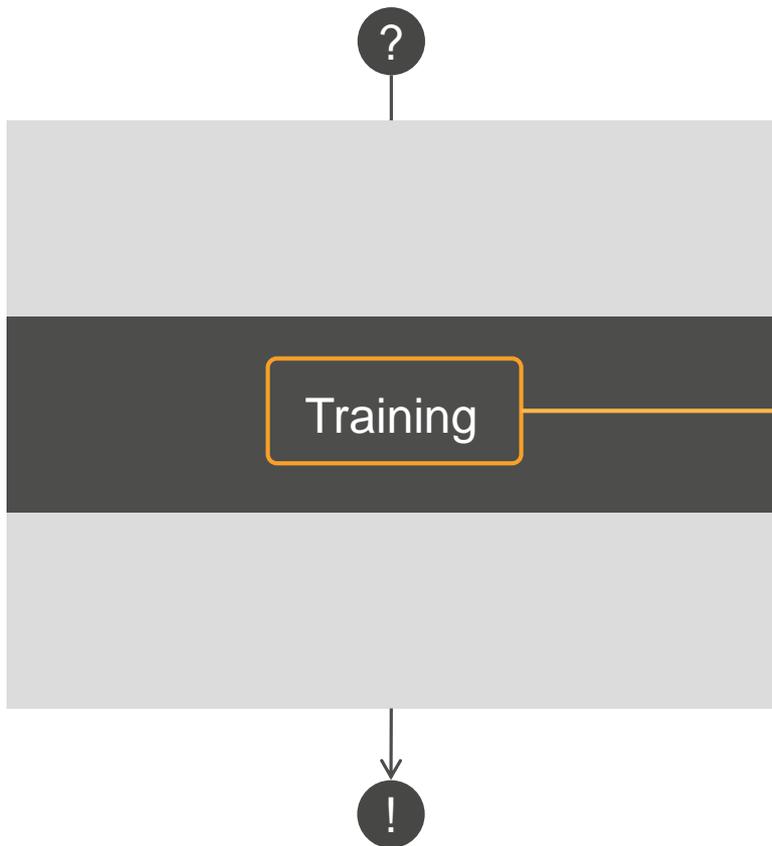


Amazon RDS



Amazon Kinesis Stream

データからモデルを学習する



モデルの更新頻度

バッチで定期的に更新

オンラインで更新し続ける

更新の必要が（ほとんど）ない

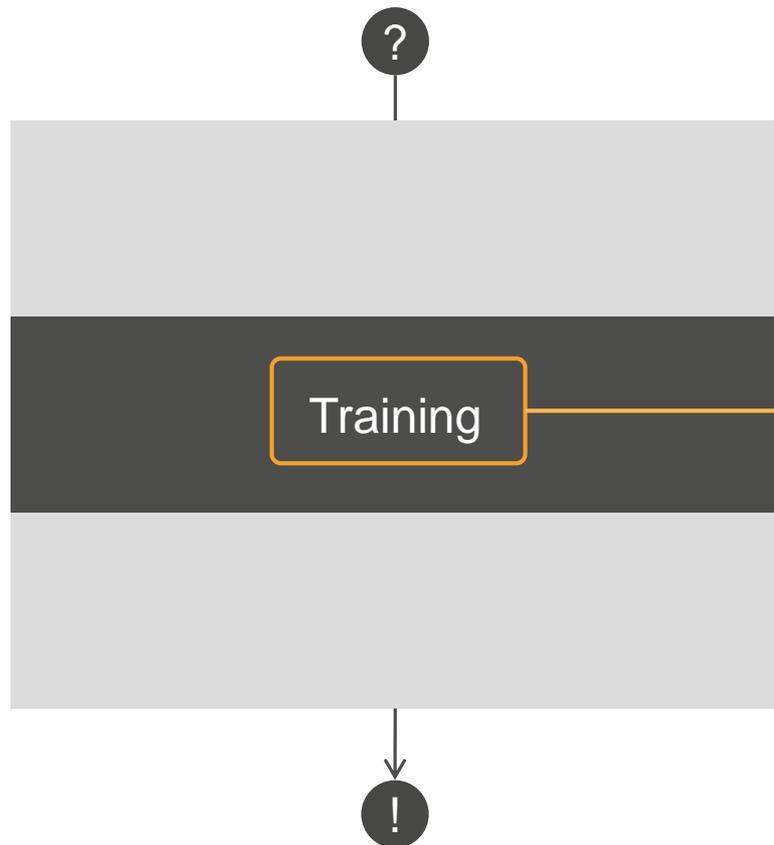


Amazon Data Pipeline



Spark streaming
on Amazon EMR

データからモデルを学習する



ハードウェア特性

CPU のみ

CPU + GPU

メモリ最適化



C4 instance on Amazon EC2

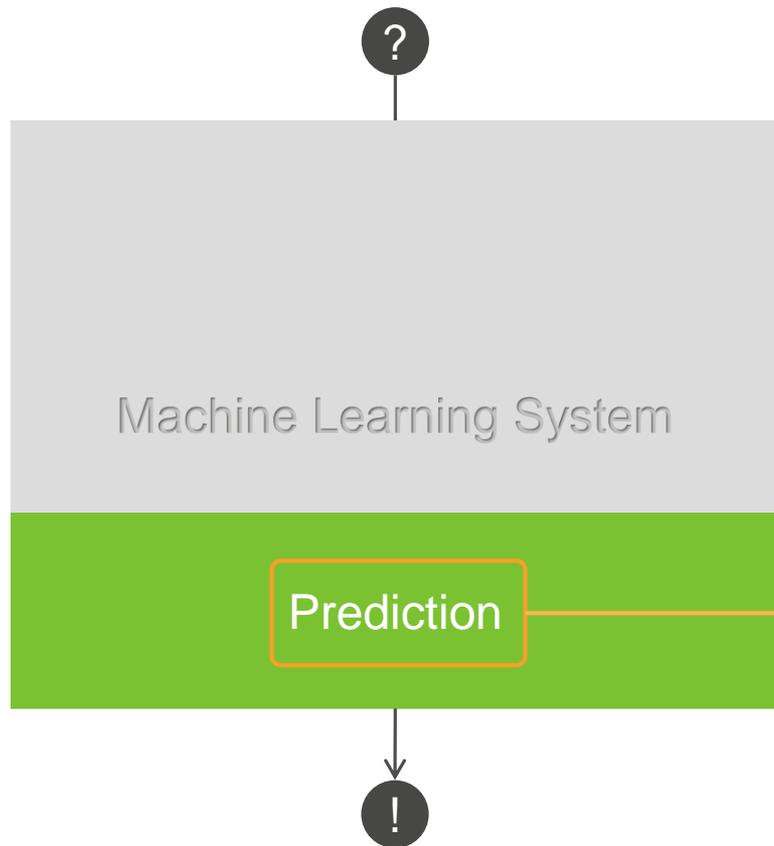


P2 instance on Amazon EC2



X1 instance on Amazon EC2

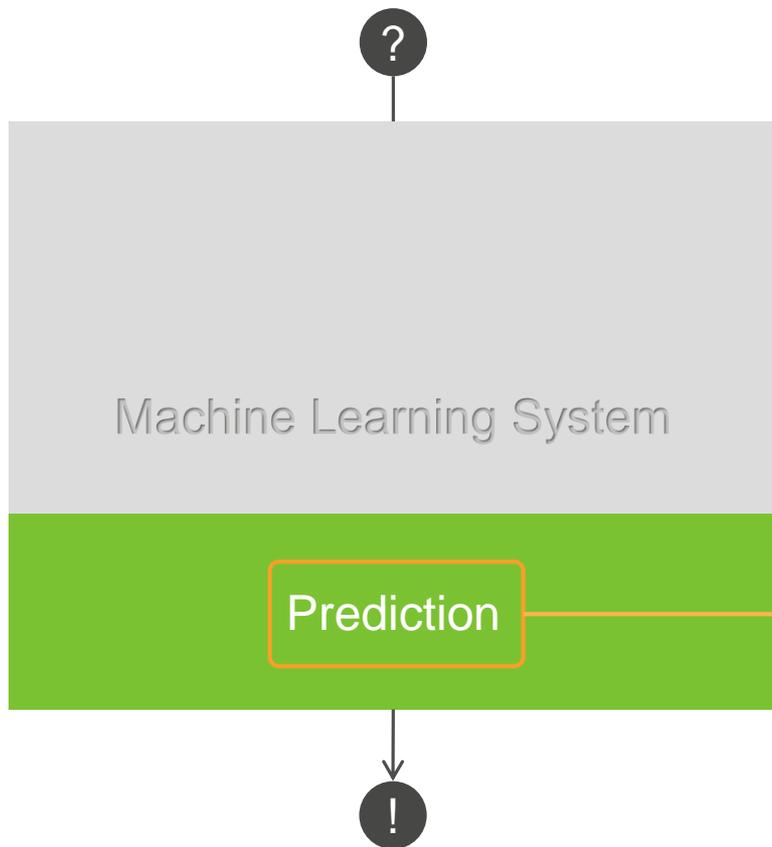
学習したモデルで予測する



モデルサイズ

提供形態

学習したモデルで予測する



モデルサイズ

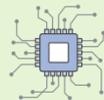
単一のインスタンスで扱える

モバイルデバイスに組み込める

DB に格納する必要がある



Amazon EC2

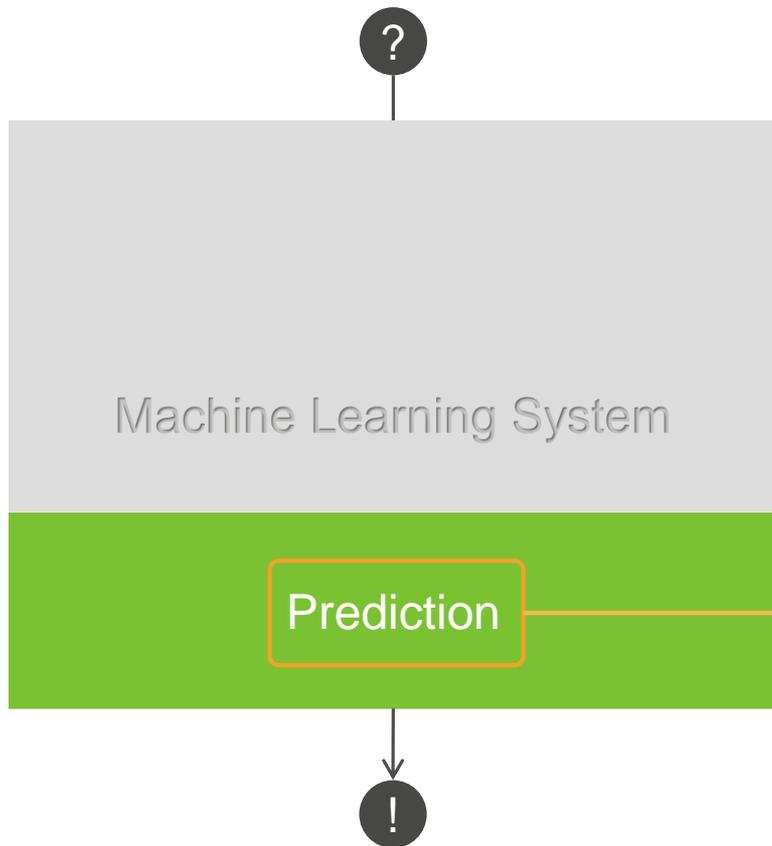


AWS Greengrass (Preview)



Amazon Dynamo DB

学習したモデルで予測する



提供形態

API を提供

ストリームデータを逐次処理

デバイス等に組み込む



Amazon Machine Learning



Amazon Rekognition



Amazon Polly



Amazon API Gateway



Amazon Kinesis Analytics



Spark Streaming on Amazon EMR



AWS Greengrass (Preview)

典型的なユースケース

ユースケース 1

ECサイトにおける商品のレコメンド

あなたはハンドメイドの雑貨をあつかう EC ショップを運営しています。1回の購入あたりの売り上げが伸び悩んでおり、商品ページでほかのおすすめ商品を表示することで、合わせ買いを促進したいと思っています。毎日新しい商品が入荷してくるので、できるだけ早くおすすめに反映させたいです。またそれとは別に、セール商品は手動でおすすめできるようにもしたいと考えています。

ユースケース 1

ECサイトにおける商品のレコメンド

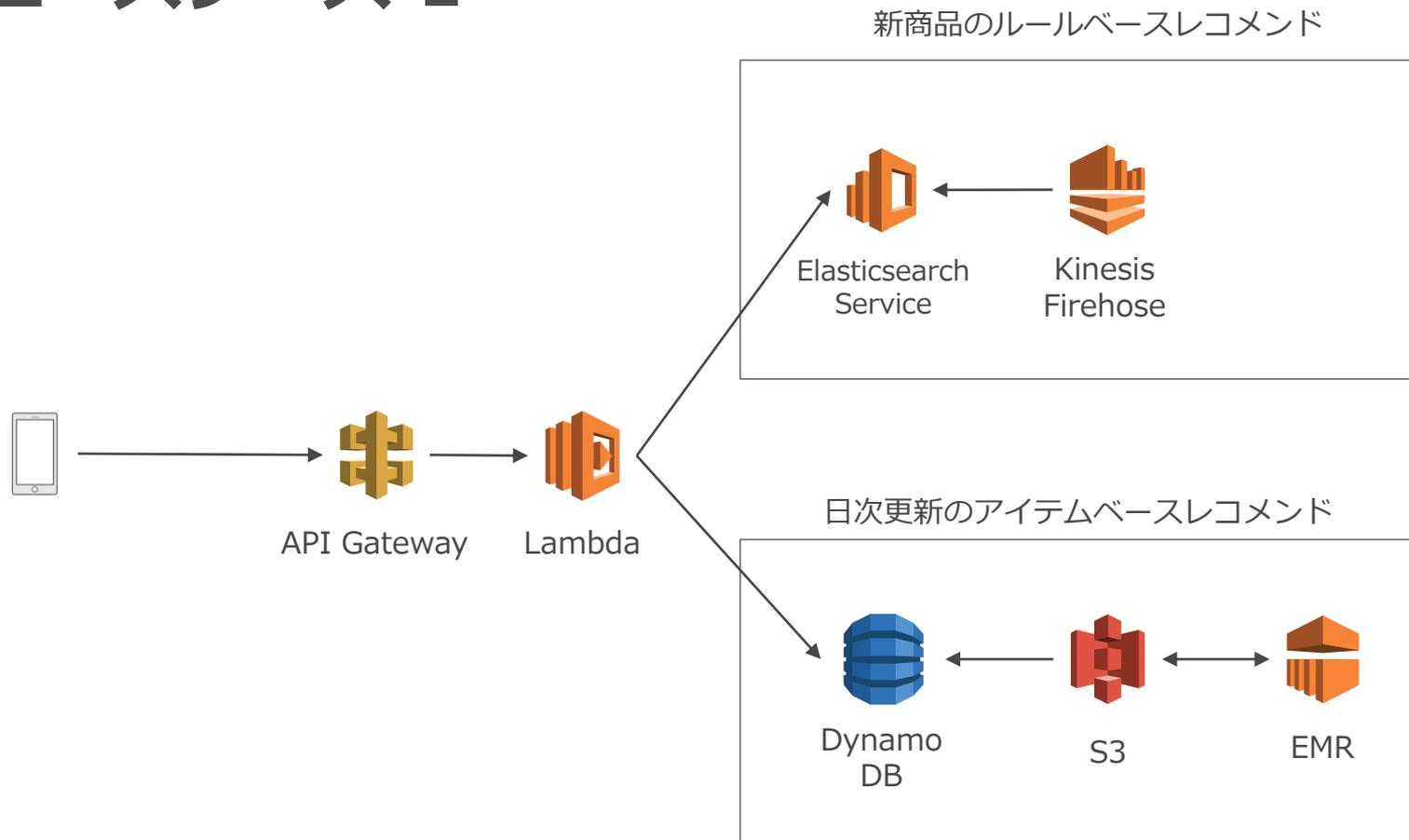
アイテムベースのレコメンドと、ルールベースのレコメンドの 2 種類

ストレージに溜まった巨大なログデータと、DBにあるマスタデータ

アイテムベースのレコメンド内容は日次で更新

スマホアプリから結果を API 経由で取得して表示

ユースケース 1



ユースケース 2

画像 SNS における同一人物の顔判定

あなたは画像ベースの SNS を運営しています。ユーザーがアップロードした写真から顔部分を抽出して人物判定を行い、同一人物が写っている場合、そのユーザーを「知り合いかも？」とおすすめる機能を追加しようとしています。

これにより、ユーザー同士のつながりが強化され、サービスへのコミットメントも高まるのではないかと考えています。

ユースケース 2

画像 SNS における同一人物の顔判定

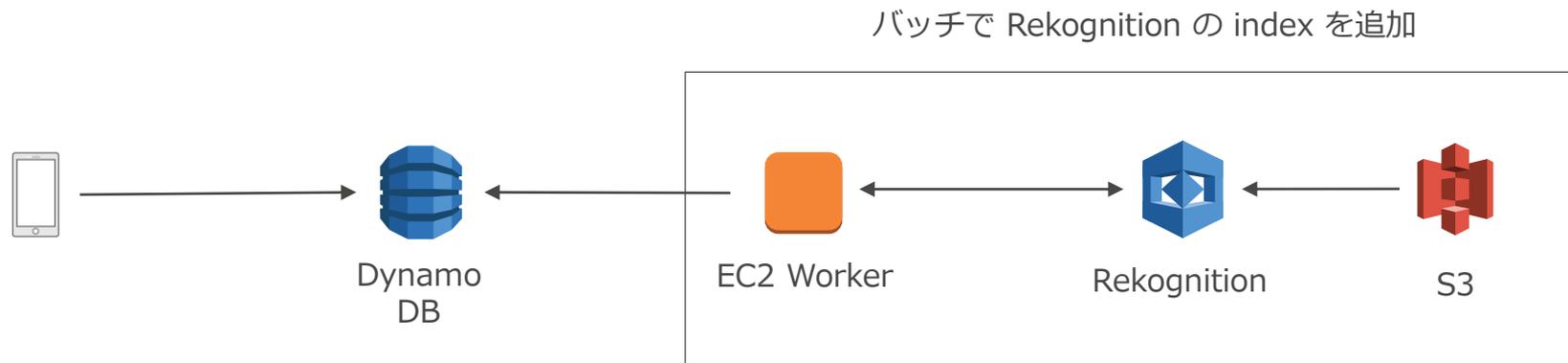
複数の画像から顔を抽出して、同一人物の判定

S3 上の大量の画像データ

モデルの更新頻度は週次-日次

スマホアプリから結果を API 経由で取得して表示

ユースケース 2



ユースケース 3

プラントのセンサデータにおける異常検知

あなたの会社では、常時稼働させているプラントにおいて、複数のセンサから常にデータを取得し続けています。

現在は 3 交代制で作業員が監視する仕組みで、非常に負荷が高い状況ですが、センサデータから異常検知を行う仕組みを作ることで、作業員の負荷を下げたり、判断ミスを減らしたりすることを考えています。

ユースケース 3

プラントのセンサデータにおける異常検知

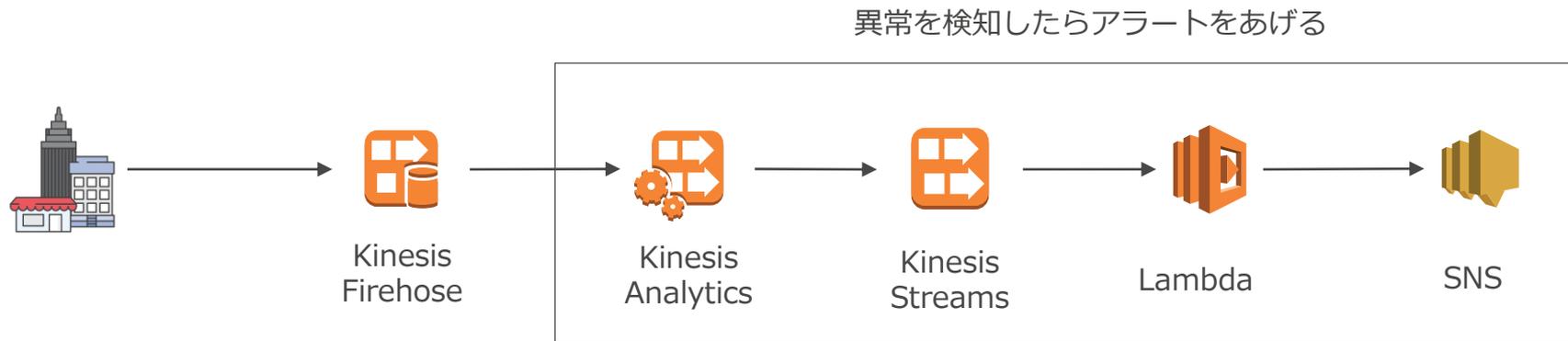
数百のセンサが5秒間隔で JSON データを送信

ストリーム処理

異常検知

アラートをあげる

ユースケース 3



まとめ

まとめ

解決したいビジネス課題から出発する
その中でニーズはあるが実現できていない部分に注目
機械学習はあくまでツールであり、目的ではない

システム要件を定める
要件に即した AWS サービスを選択する
ユースケースを参考に

本セッションのFeedbackをお願いします

受付でお配りしたアンケートに本セッションの満足度やご感想などをご記入ください。アンケートをご提出いただきました方には、もれなく**素敵なAWSオリジナルグッズ**をプレゼントさせていただきます



アンケートは受付、パミール3FのEXPO展示会場内にて回収させていただきます

AWS ソリューション Day 2017- Database Day ~すでに始まっている！「クラウドへのデータベース移行」と 「データレイクを軸としたビッグデータ活用」~

📦 Database Dayとは？

ユーザー企業/パートナー/AWSによる導入事例や活用動向また技術情報をご紹介するIT部門（エンジニア・管理者など）向けのカンファレンス

📦 開催日時・会場

- 2017年7月5日(水) 10:00~17:30（9:30開場予定）
- 大崎ブライトコアホール（JR大崎駅より徒歩5分）

📦 セッション

①基調講演 ②ブレイクアウトセッション - 2トラック構成

トラック1：データベース移行（事例セッションあり）

トラック2：データレイク（JAWSUG-BigData支部 事例セッションあり）

📦 お申込み

<https://aws.amazon.com/jp/about-aws/events/2017/solutiondays20170705/>

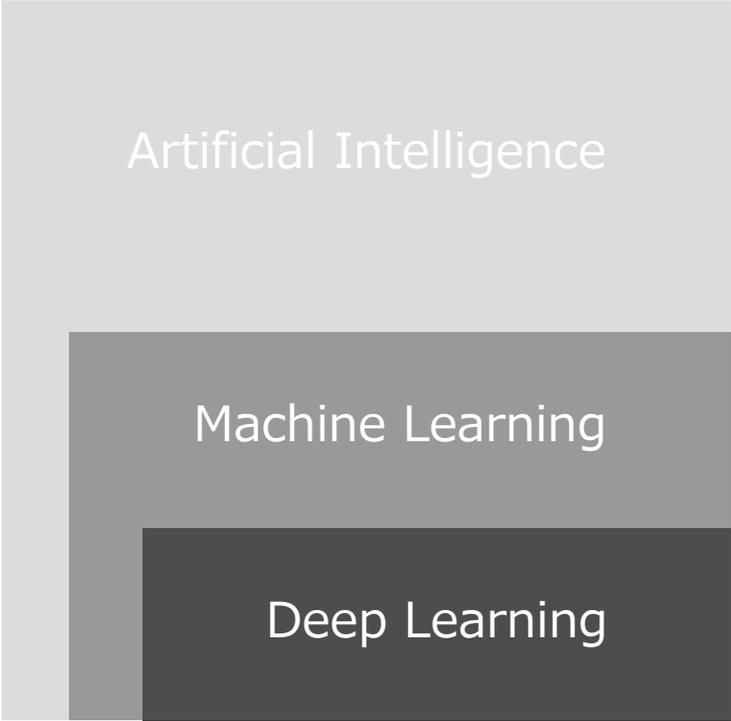
AWS

S U M M I T



Appendix

機械学習とはなにか

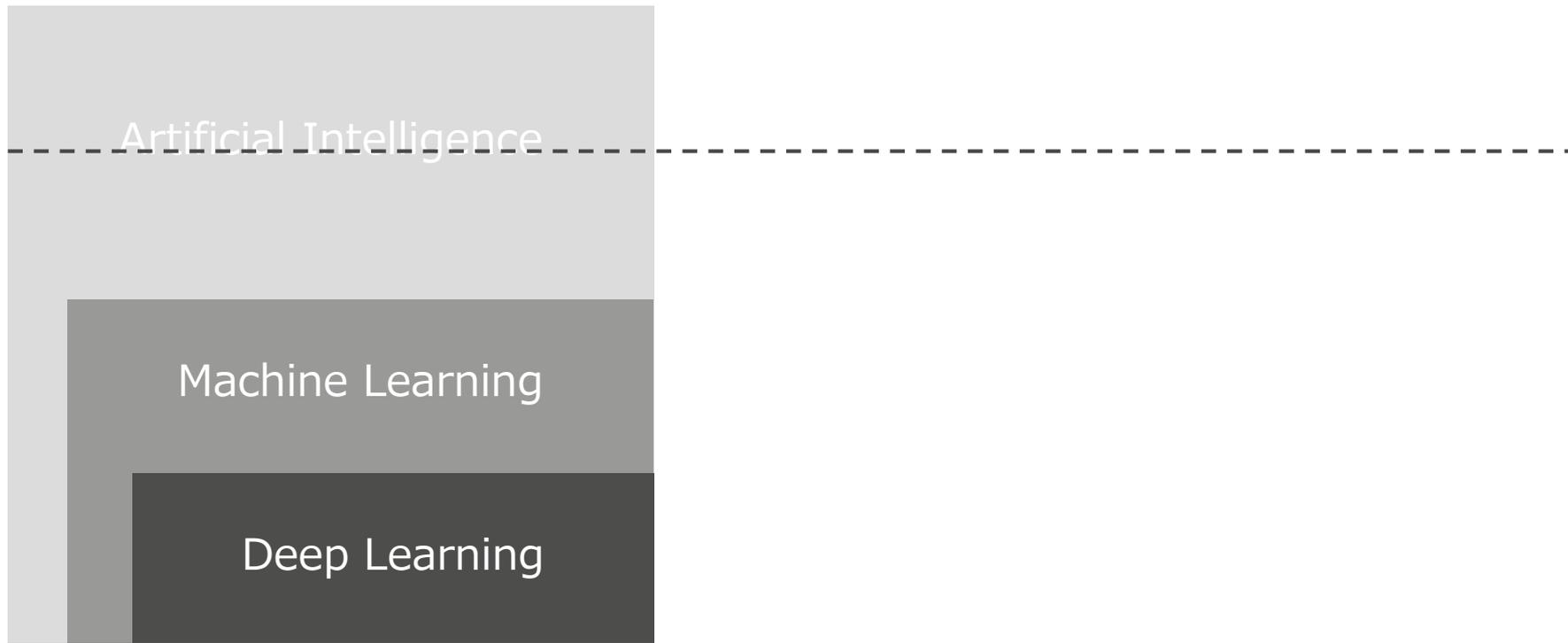


Artificial Intelligence

Machine Learning

Deep Learning

機械学習とはなにか



機械学習とはなにか

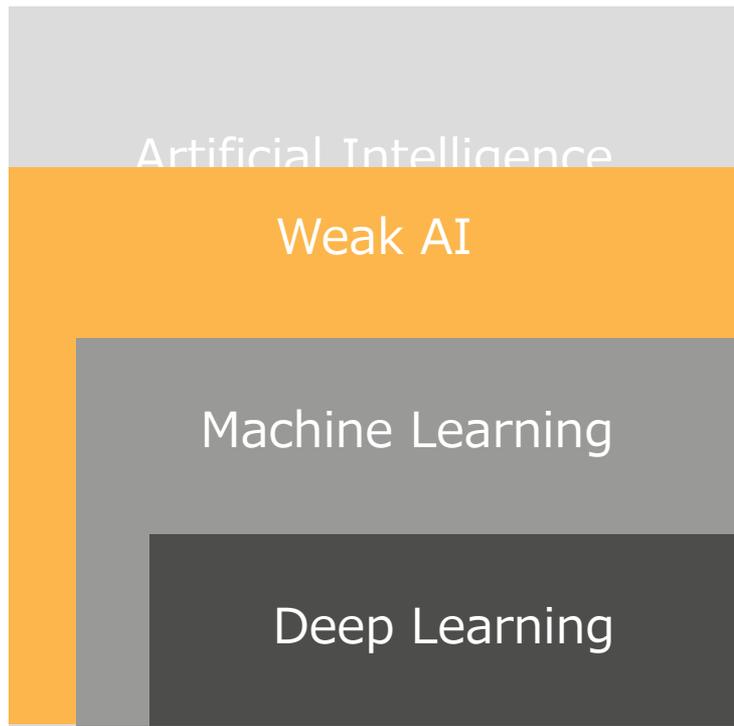
Strong AI

汎用人工知能
人間のよう“知的”な知能

Machine Learning

Deep Learning

機械学習とはなにか



特化型人工知能

汎用人工知能のための要素技術
人間側でいろいろお膳立てが必要
“知的”ではない

機械学習とはなにか

Artificial Intelligence

Machine Learning

Deep Learning

機械学習

特定事象のデータを学習してモデルを獲得
それに基づいて判断や予測を実施

機械学習とはなにか

Artificial Intelligence

Machine Learning

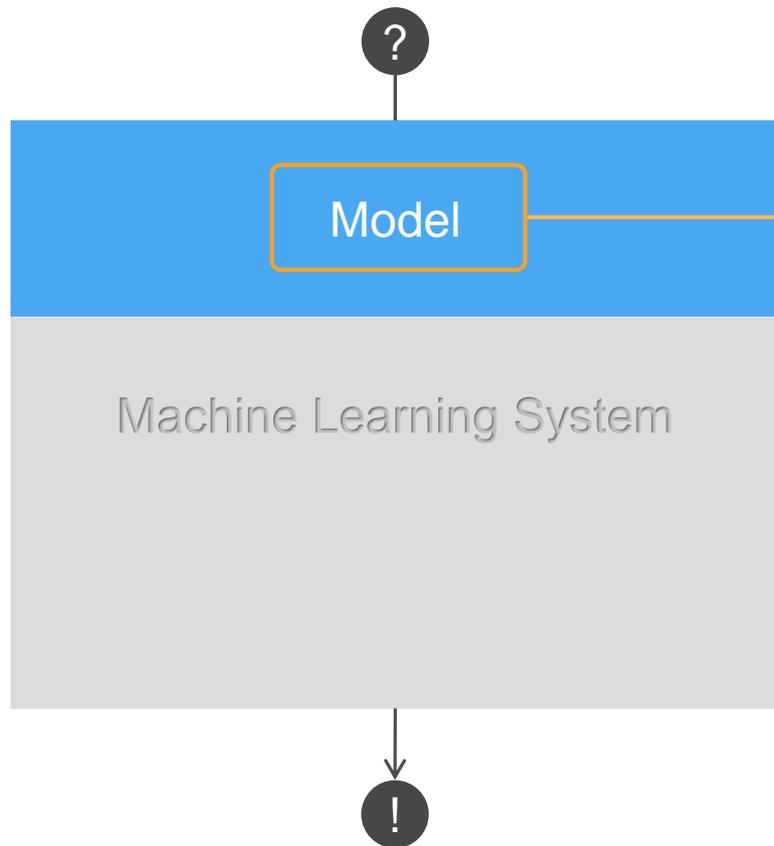
Deep Learning

ディープラーニング
機械学習の手法のひとつ

機械学習の一般的な分類

Machine Learning		教師あり学習	機械に学習させたい事象の (正しい) 結果を与えて学習させる
		教師なし学習	与えたデータから その分布の特徴を捉える学習を行う
		強化学習	正解ではなく 行動に対する評価を与えて学習する

ユースケース 1



モデル選定

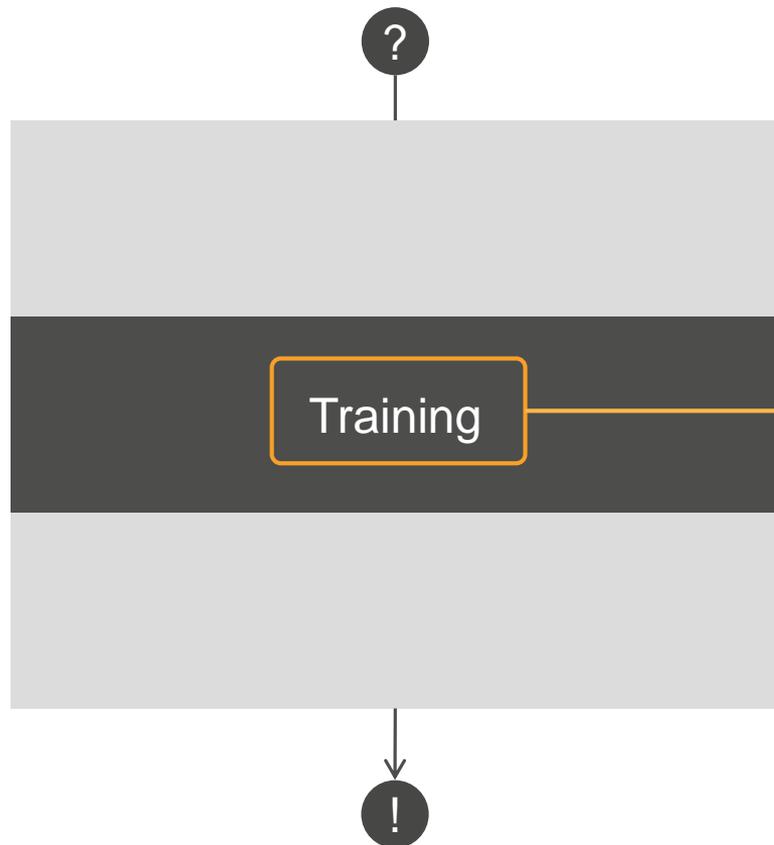


協調フィルタリング

コンテンツベース

Spark MLlib の ALS を使って
協調フィルタリングのモデルを更新
また新しい商品が追加されたら、
レコードを Elasticsearch Service に
追加して、more like this 検索

ユースケース 1



学習システム



クラスタで処理する必要があるサイズ

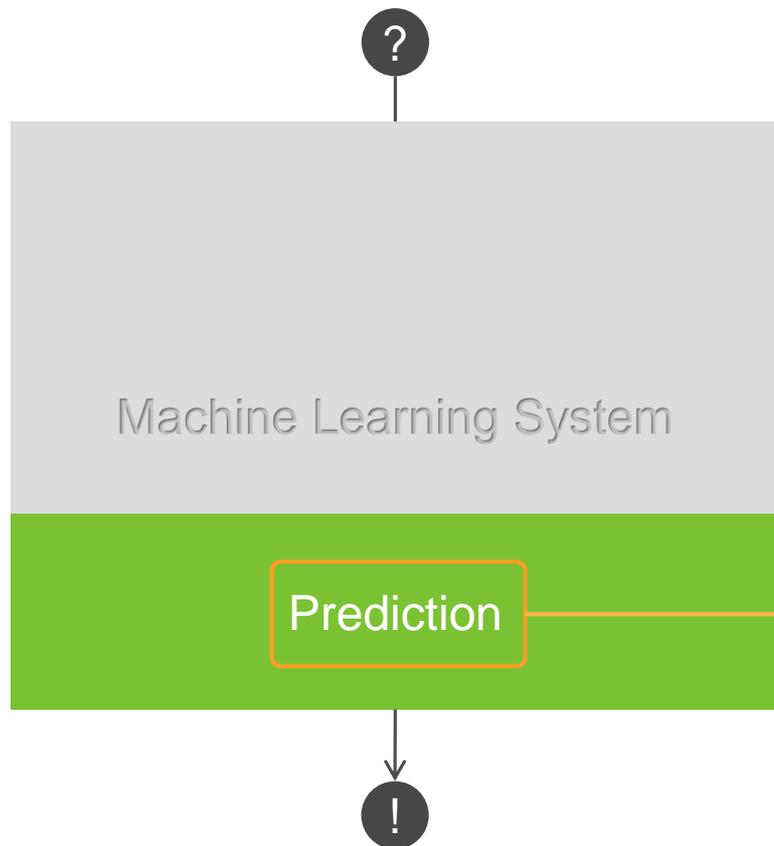
ストレージ保存 + バッチ更新

ストリーム + オンライン更新

S3 上のログデータを Spark on EMR
で読みこんで日次バッチを実行

Kinesis Firehose から直接データを
挿入してオンラインでデータ更新

ユースケース 1



予測システム



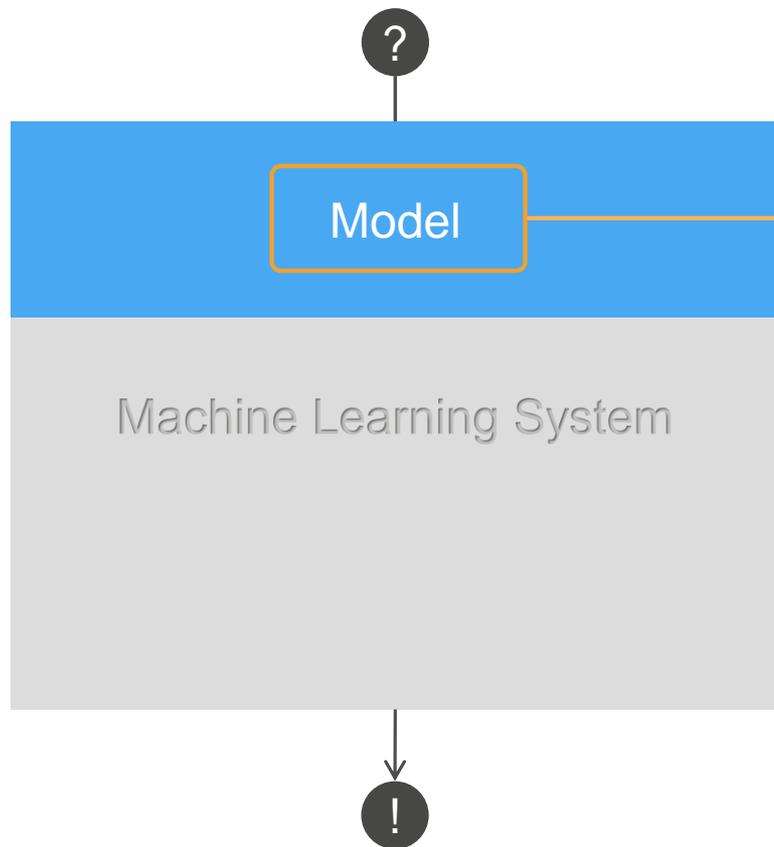
DB に格納する必要がある

API で提供

協調フィルタリングの結果は
Dynamo DB に格納

API Gateway から lambda で
Dynamo DB および Elasticsearch
Service 両方から結果を取得

ユースケース 2



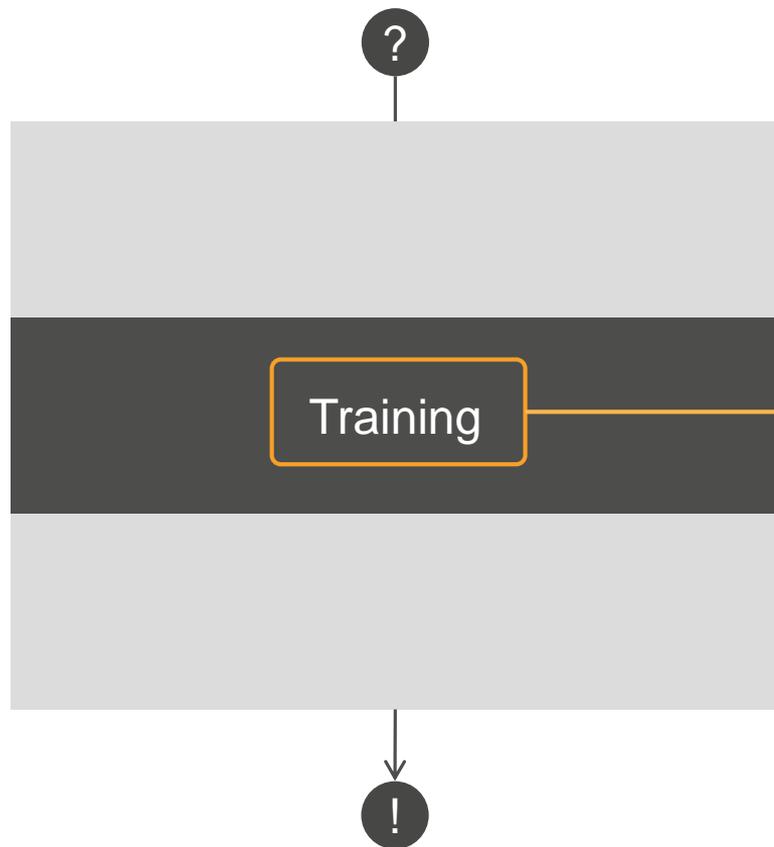
モデル選定



Deep Learning

Rekognition で顔認識と
類似度判定をおこなう

ユースケース 2



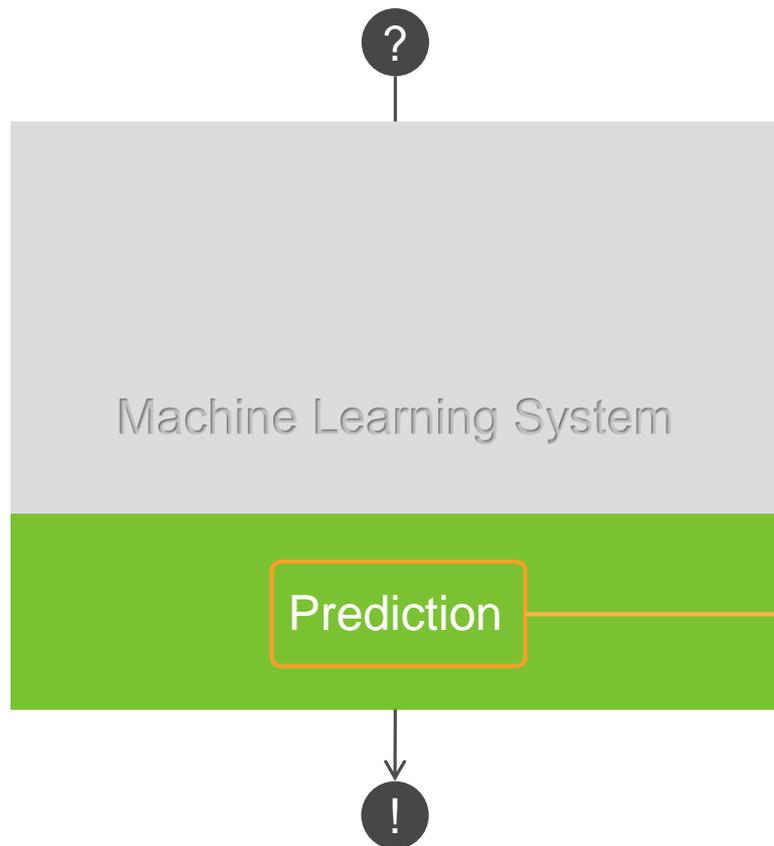
学習システム



ストレージ保存 + バッチ更新

S3 上の画像データを Rekognition の indexFaces API 経由で認識させて index を作成する

ユースケース 2



予測システム

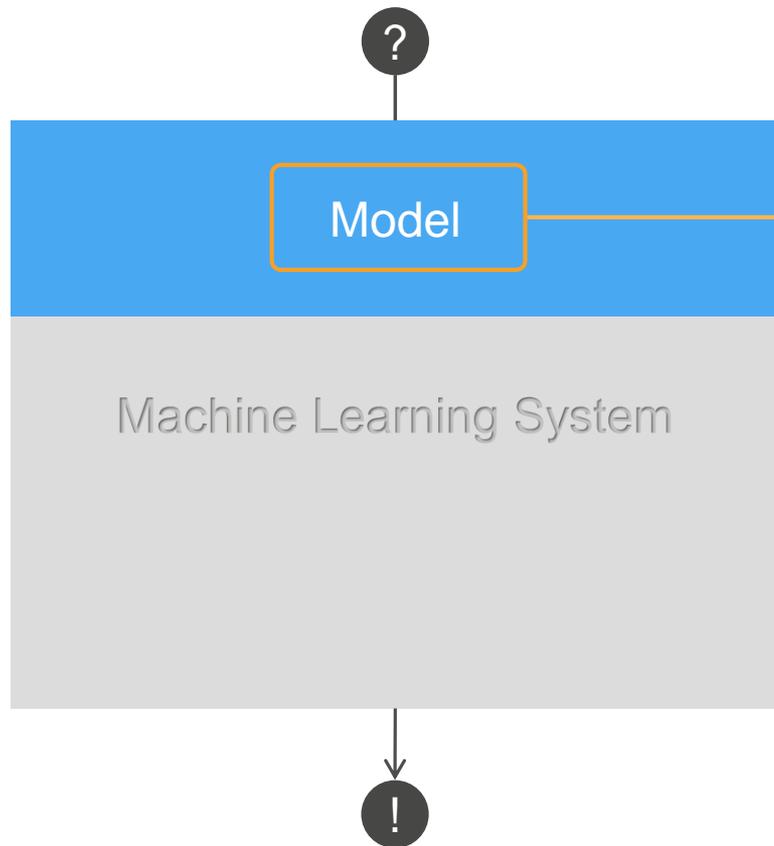


DB に格納する必要がある

API で提供

ユーザーごとの「知り合いかも？」
リストは Dynamo DB に格納して
API 経由で提供

ユースケース 3



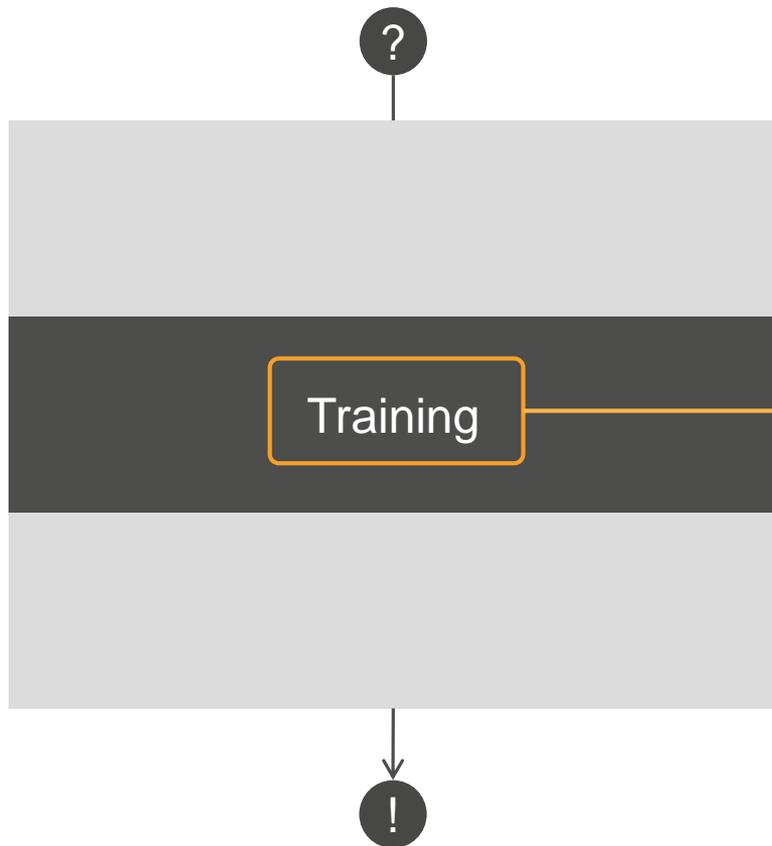
モデル選定



Random Cut Forest

Kinesis Analytics で変化点を検知
データの種類や問題によっては、
Spark Streaming 等を使う場合も

ユースケース 3



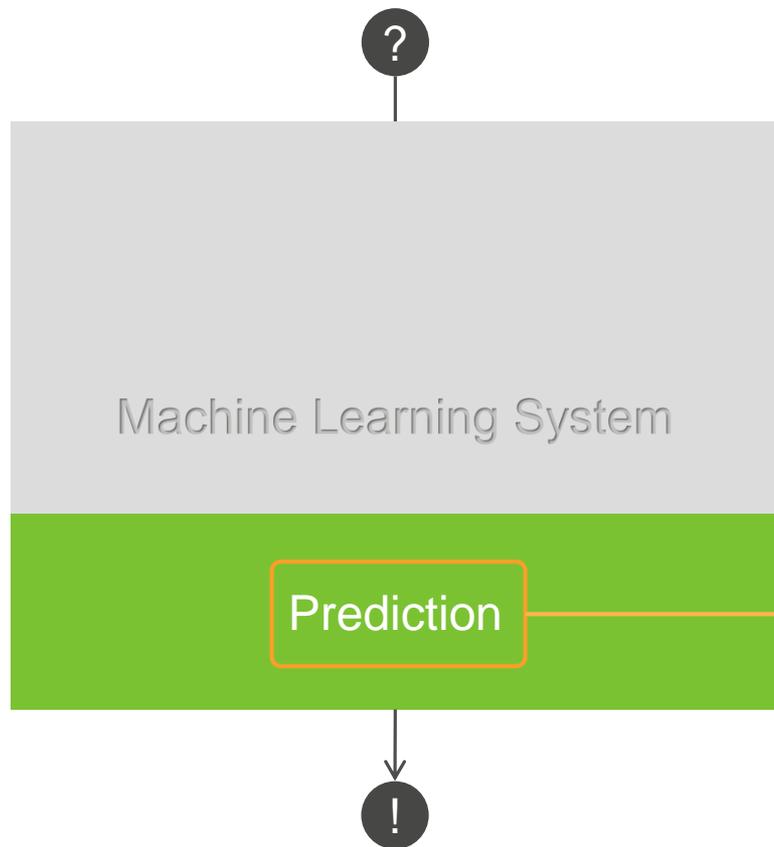
学習システム



ストリーム + オンライン更新

Kinesis Stream 経由で取得したデータを Kinesis Analytics でストリームで処理

ユースケース 3



予測システム



アラートをあげる

Kinesis Analytcs の分析結果を
Lambda 経由で SNS に送り、
アラートをあげる