



---

# AWS Summit

Tokyo

---



---

# 新サービス解説セッション

## Amazon Elastic File System と Amazon Machine Learning

アマゾン データ サービス ジャパン株式会社  
ソリューションアーキテクト  
辻 義一 ・ 今井 雄太



## ■ Gold Sponsors



Empowered by Innovation



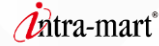
## ■ Global Sponsors



## ■ Silver Sponsors



## ■ Bronze Sponsors



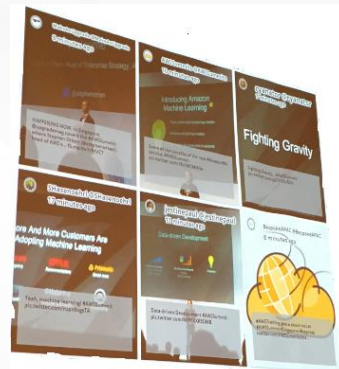
## ■ Global Tech Sponsors



## ■ Logo Sponsors



ハッシュタグ **#AWSummit**  
で、皆さんのツイートが展示エリア  
の大画面に表示されます



公式アカウント **@awscloud\_jp**  
をフォローすると、ロゴ入り  
コースターをプレゼント



【コースター配布場所】

メイン展示会場、メイン会場1F受付、デベロッパーカンファレンス会場





# Amazon Elastic File System (EFS)



# 自己紹介

辻 義一（つじ よしかず）

📦 西日本担当  
ソリューションアーキテクト

📦 簡単な経歴

- 大阪生まれの大阪育ち
- 独立系SIerでインフラエンジニア



# AWS ストレージのポートフォリオ

## Amazon S3

- オブジェクトストレージ: データはオブジェクトの集合 (バケット)
- HTTPベースのAPI経由でアクセス

## Amazon EBS

- SANのようなブロックストレージ: ディスクボリュームとして扱う
- 1台のEC2インスタンスから低遅延でアクセス

## Amazon Glacier

- アーカイブストレージ: オブジェクトをまとめたボードとして管理
- 低コストなストレージで不定期にアクセス

*New*

Amazon Elastic File System (EFS)



# NFSでファイルアクセスできるストレージ

## 「NAS」のようなサービス

- NFSを使ってネットワーク経由でアクセス
  - NFSv4を利用
  - Linuxからマウントして利用できる
- EFSへのアクセスはファイル単位
  - EBSはブロック単位

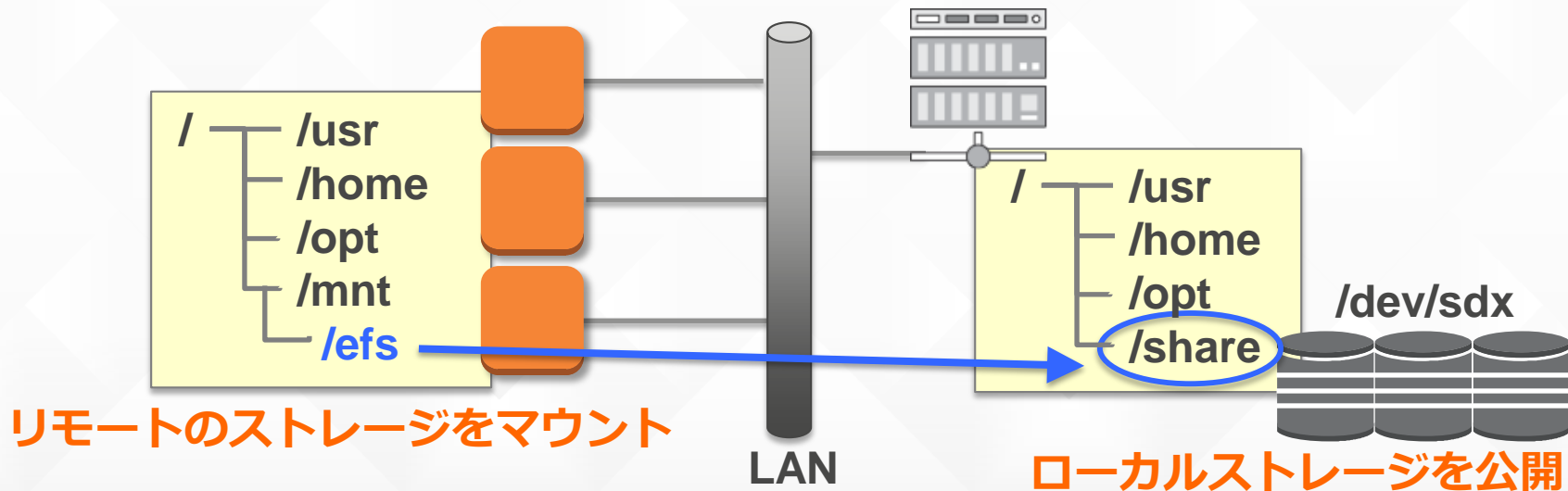


# NFS (Network File System) とは

Unix / Linux でネットワーク経由でストレージを共有するための標準的な仕組みとプロトコル

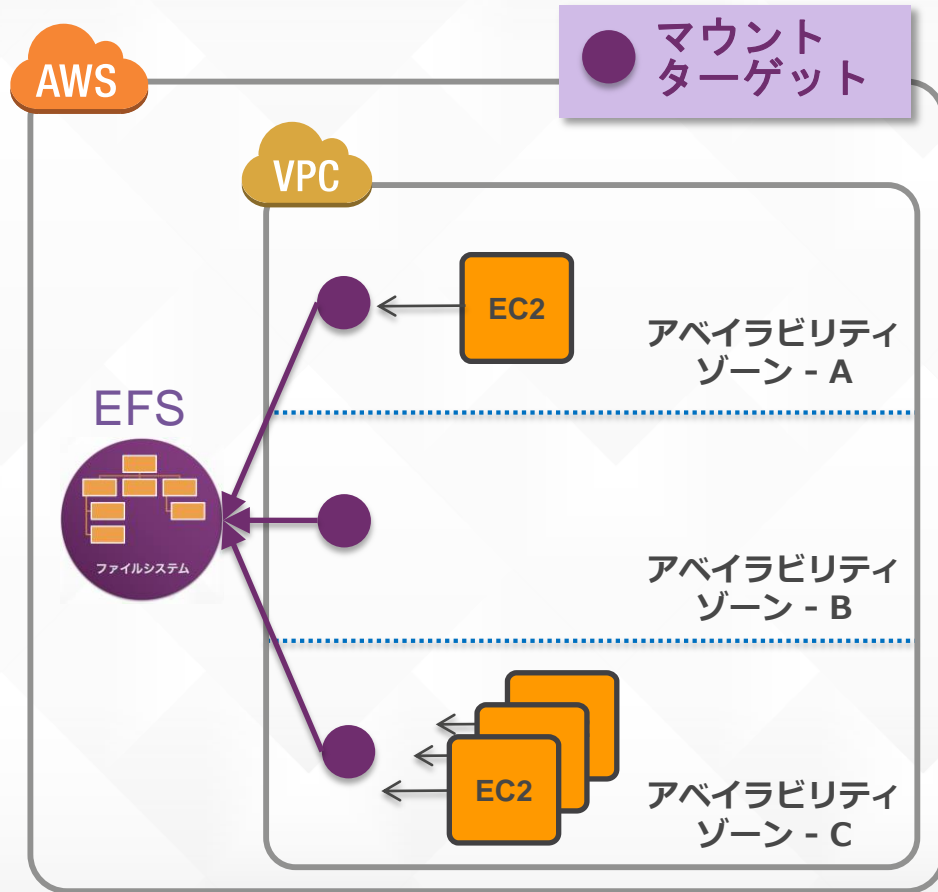
ストレージを利用するサーバ群

ストレージを提供するサーバ



# EFSにはマウントターゲットを通じてアクセス

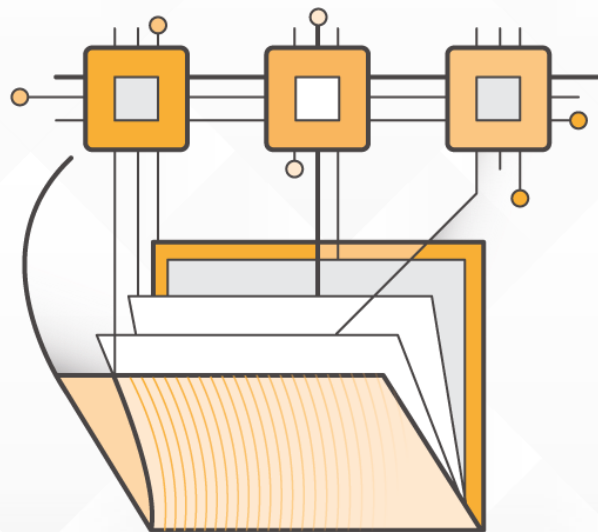
- VPC内のマウントターゲットがNFSの接続先
- 各EC2インスタンスから、同じAZにあるマウントターゲットに接続



# 幅広いユースケースで利用できる

例えば

- コンテンツのレポジトリ
  - AutoScaling するサーバ群で、ユーザがアップロードしたデータを全サーバで共有
- ビックデータ/HPC
  - 大量のサーバに分散して分析する際に分析データの共有



# EFS の特徴

1

シンプル

2

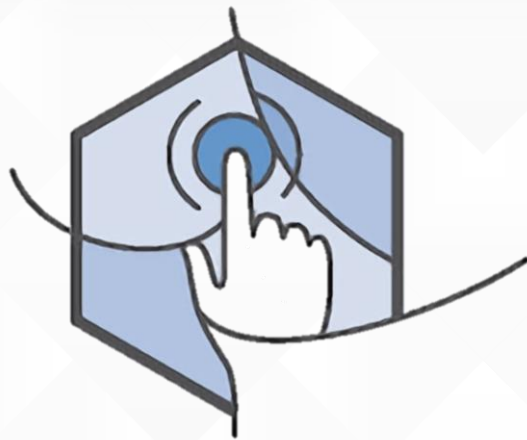
柔軟  
スケーラブル

3

高耐久性  
高可用性

# 1 シンプル

- フルマネージド型
  - サーバ、ディスクなどの管理は不要
  - 数秒で作成できる
- NFSなので既存のツールやアプリとシームレスに連携
  - 標準のOS APIから利用できる
  - 標準的なファイルシステムの仕組み
- 保存された合計容量だけの課金  
⇒ 簡単に料金予測しやすい



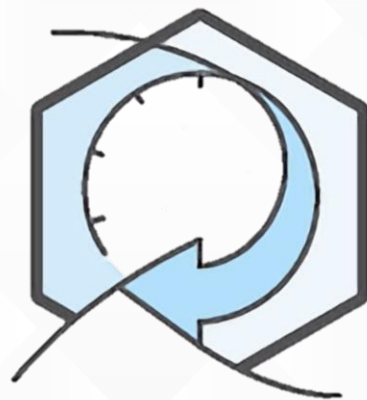
## 2 柔軟・スケーラブル

- 容量は自動的に拡張・縮小
  - ファイルの追加、削除に連動しペタバイトまで
  - 保存している容量だけが課金
- 性能は容量に応じてスケール
  - スループットは容量に応じて性能アップ
  - 小容量時に有利なクレジット制でバーストも
  - SSDベースでファイル操作の応答は1桁ミリ秒程度
- 数千のNFS同時接続をサポート



### 3 高耐久性・高可用性

- 複数AZに複製されて保存される
- 複数のAZから同時に読み書き可
  - 複数のAZにある複数のEC2インスタンスから読み書きできる
  - 書き込み完了直後には他でも反映される







---

# 利用のイメージ



# Elastic File System

Amazon Elastic File System provides file storage for use with your EC2 instances.

[Create File System](#)

[Getting Started Guide](#)



## Create

Create an EFS file system to store your files in the Amazon cloud. A file system grows and shrinks automatically with the files you put in, and you only pay for what you use.



## Access

Write files to and read files from your EFS file system via the NFSv4 protocol. Any number of EC2 instances can work with your file system at the same time, and your instances can be in multiple Availability Zones in a Region.



## Manage

You can easily administer your file system, and you can view and alert on key metrics using CloudWatch.



# Create File System

## Step 1: Configure File System Access

Step 2: Add Tags

Step 3: Review and Create

## Configure File System Access

An EFS file system is accessed by EC2 instances running inside one of your VPCs. Instances connect to a file system via a network interface called a mount target. Each mount target has an IP address, which we assign automatically or you can specify.

VPC:  ⓘ

## Create Mount Targets

Instances connect to a file system via mount targets you create. We recommend creating a mount target in each of your VPC's availability zones so that EC2 instances across your VPC can access the file system.

	Availability Zone	Subnet ⓘ	IP Address ⓘ	Security Group ⓘ
<input checked="" type="checkbox"/>	us-west-2a	<input type="text" value="subnet-f3db6d96"/>	Automatic ✎	<input type="text" value="x sg-23f78a46 - default"/>
<input checked="" type="checkbox"/>	us-west-2b	<input type="text" value="subnet-6b8e5a1c"/>	Automatic ✎	<input type="text" value="x sg-23f78a46 - default"/>
<input checked="" type="checkbox"/>	us-west-2c	<input type="text" value="subnet-3585706c"/>	Automatic ✎	<input type="text" value="x sg-23f78a46 - default"/>

Cancel

Next Step



# Create File System

[Step 1: Configure File System Access](#)

**Step 2: Add Tags**

[Step 3: Review and Create](#)

## Add Tags

You can add tags to describe your file system. A tag consists of a case-sensitive *key* : *value* pair. For example, you can define a tag with *key* : *value* pair *Corporate Department* : *Sales and Marketing*. At a minimum, we recommend a "Name" tag.

Key	Value	Remove
<input type="text" value="Name"/>	<input type="text" value="My first EFS file system"/>	<input type="button" value="✕"/>
<input type="text" value="Add New Key"/>	<input type="text" value="Add New Value"/>	

[Cancel](#)

[Previous](#)

[Next Step](#)



# Create File System

[Step 1: Configure File System Access](#)

[Step 2: Add Tags](#)

**Step 3: Review and Create**

## Review and Create

Review the configuration below before proceeding to create your file system.

### File System Access

VPC	Availability Zone	Subnet	IP Address	Security Group
vpc-5ea27c3b	us-west-2a	subnet-f3db6d96	Automatic	sg-23f78a46 - default
	us-west-2b	subnet-6b8e5a1c	Automatic	sg-23f78a46 - default
	us-west-2c	subnet-3585706c	Automatic	sg-23f78a46 - default

### Tags

Name: My first EFS file system

Cancel

Previous

Create File System

fs-1daa4eb4	0.0 GiB	3	Apr 2, 2015 1:11:34 AM -0400
My first EFS file system fs-d694707f	0.0 GiB	3	Apr 3, 2015 8:00:38 PM -0400

### Other Details

**Owner ID** 000614220061

**Life Cycle State** Available

### Tags

[Manage Tags](#)

**Name:** My first EFS file system

### File System Access

[Manage File System Access](#)

[DNS Names](#)

[EC2 mount instructions](#)

### Mount Targets

VPC	Availability Zone	Subnet	IP Address	Mount Target ID	Network Interface ID	Security Groups	Life Cycle State
vpc-5ea27c3b	us-west-2c	subnet-3585706c	10.0.2.139	fsmt-62a442cb	eni-98a600c0	default - sg-23f78a46	Available
	us-west-2a	subnet-f3db6d96	10.0.0.165	fsmt-63a442ca	eni-73303917	default - sg-23f78a46	Available
	us-west-2b	subnet-6b8e5a1c	10.0.1.187	fsmt-65a442cc	eni-5185d027	default - sg-23f78a46	Available

```
$ sudo mount -t nfs4 us-west-2a.fs-  
d694707f.efs.uswest-2.amazonaws.com:/ /mnt/efs
```

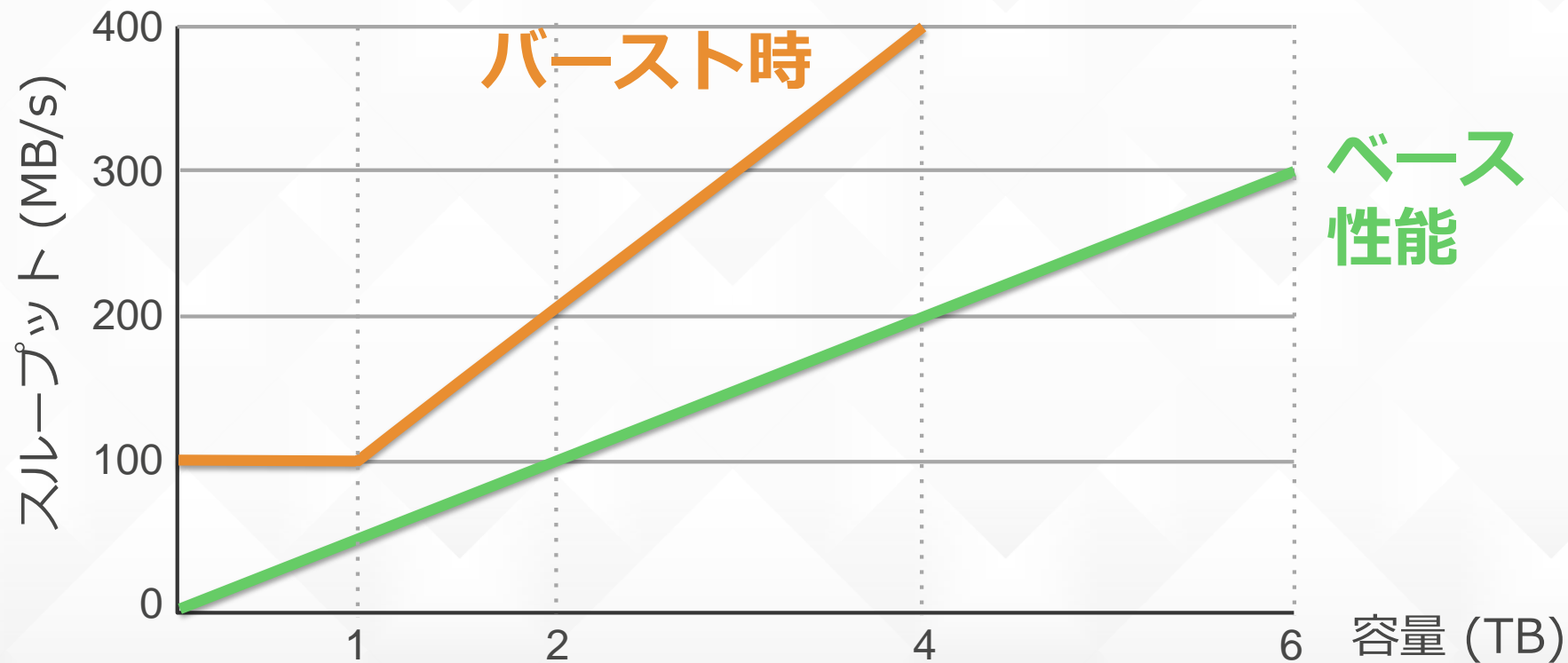


# Deep Dive



# パフォーマンス

- 1TB以上では毎日の12時間、倍にバースト可



# 料金

- EFS 料金: \$0.30/GB・月 (USリージョン)
- EFS は利用したストレージ容量のみのお支払い
  - コミットメントなし、前払いなし
  - 事前プロビジョニングは不要
  - 容量以外の課金項目なし

# セキュリティ

- EFS APIの保護は、IAMで
- NFS アクセスの保護は、マウントポイント毎に設定できるセキュリティグループで
- ファイルへのアクセス保護は、パーミッション、オーナーで

```
drwxr-xr-x  4 root    root    4096 Feb  5 22:37 .
dr-xr-xr-x 25 root    root    4096 Feb  5 22:20 ..
drwxr-xr-x  2 mike    mike    4096 Feb  4 01:18 mike
```

# その他 現時点での仕様や制限

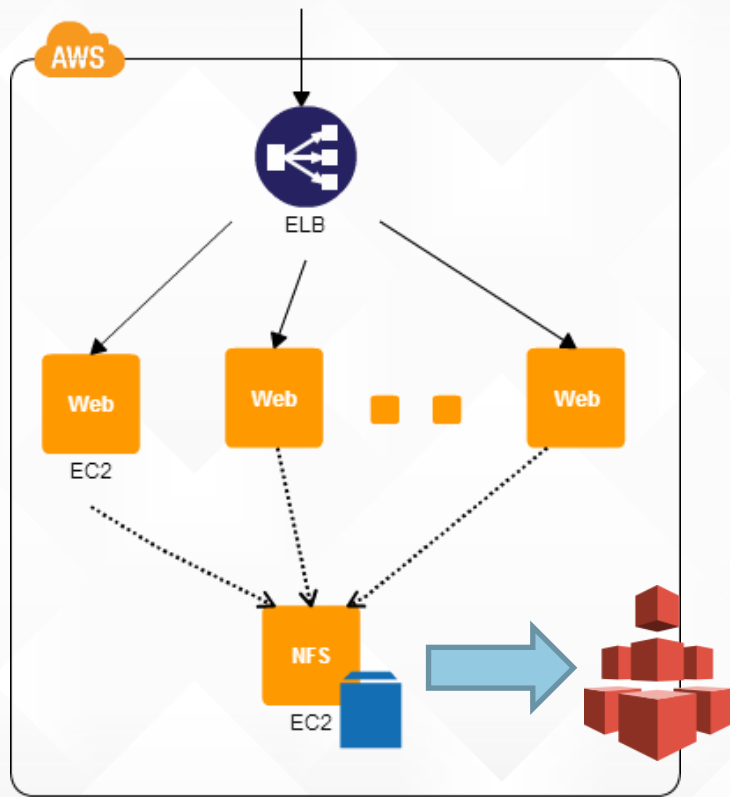
*Limited Preview*

- プロトコル：NFSv4.0 (TCP 2049ポート)
  - ※NFSv4.1 や CIFSのサポートなし
- 推奨クライアント：現行 Linux NFSv4 クライアント
- NFS v4.0の機能であっても未サポートの機能
  - 強制ロック
  - Share Deny (Windowsで使われている)
  - ACL
  - Kerberos認証 など
- 同一VPC内からのみアクセス可

# EBS vs EFS vs S3

		EBS (SSD)	EFS	S3
プロトコル		SCSI	NFS	HTTP API (Webサーバにも)
既存アプリなどの親和性		○	○	△ API対応要
課金	単位	容量 + (性能)	容量	容量 + リクエスト数
	容量単価	\$ 0.12 /GB	\$ 0.30 /GB	\$ 0.03 /GB
耐久性		1 AZ内で複製	複数AZで複製	3箇所以上で複製
性能		GBあたり 3 iops	GBあたり 50 KB/s	スケーラブル

# CDP : NFS Sharingパターンが変わる



## • 元々の注意点

- NFSサーバの管理が必要となる。
- EC2インスタンスが多くなるとNFSアクセスのパフォーマンスの考慮が必要となる。
- NFSサーバがSPOFになるのを防ぐためには、GlusterFSなどのソリューションを検討する。

EC2を使ったNFSサーバのパフォーマンス課題や構築・運用の手間を解消

# 最後に… 提供スケジュール

- Limited Preview での提供を**先週から順次開始**
  - US-West (Oregon) リージョンを対象  
<http://aws.amazon.com/efs/>
- リリースは年内の予定
  - 以下のリージョンを追加予定
    - US-East (Northern Virginia)
    - EU (Ireland)





# Amazon Machine Learning



# 自己紹介

今井雄太 (いまいゆうた)

[yuimai@amazon.co.jp](mailto:yuimai@amazon.co.jp)

- ソリューションアーキテクト
- ゲーム、広告系のお客様を主に担当
- 好きなAWSのサービス：
  - Amazon Elastic MapReduce



# アジェンダ

1. 機械学習とは？
2. Amazon Machine Learning
3. アーキテクチャへの組み込み



# Part1: 機械学習とは？

# 3種類のデータ駆動型アプリケーション



遡及的分析とレポ  
ーティング

Amazon Redshift,  
Amazon RDS  
Amazon S3  
Amazon EMR



即時の判断  
リアルタイム処理と  
ダッシュボード

Amazon Kinesis  
Amazon EC2  
AWS Lambda



予測

スマートアプリ  
ケーションを作  
成可能にする

Amazon ML

# 機械学習の例

## - Amazon Machine Learningで実現可能な例-

- このメールはスパムメールか？
  - 過去のメールアーカイブをもとにYes/Noを予測する
- この商品は本、日用品、食品のいずれなのか？
  - 多くの商品データをもとにその商品のカテゴリを予測する
- 明日の売上はどのくらいになるか？
  - 過去の売上データなどをもとに明日の売上を予測する

# 機械学習の例

- このメールはスパムメールか？
  - 過去のメールアーカイブをもとにYes/Noを予測する

## 教師データ

送信者domain	サーバーIP	送信時刻	言語	Spam?
A	123.123.123.123	...	JP	Y
B	111.111.111.111	...	EN	N
B	111.111.111.111	...	EN	N
D	123.456.789.012	...	FR	Y

別の方法でSpam判定済みの教師データ(過去のデータ)をもとに

## 予測対象データ

送信者domain	サーバーIP	送信時刻	言語	Spam?
B	123.123.123.123	...	JP	
A	123.456.789.012	...	FR	

予測対象データ(新しいデータ)のSpam判定をする

# 機械学習の例

- この商品は本、日用品、食品のいずれなのか？
  - 多くの商品データをもとにその商品のカテゴリを予測する

## 教師データ

商品名	価格	大きさ	メーカー	カテゴリ
A	123	...	AA	Book
B	456	...	BB	Book
C	100	...	CC	Food
D	500	...	DD	Grocery

大量の商品のデータをもとに

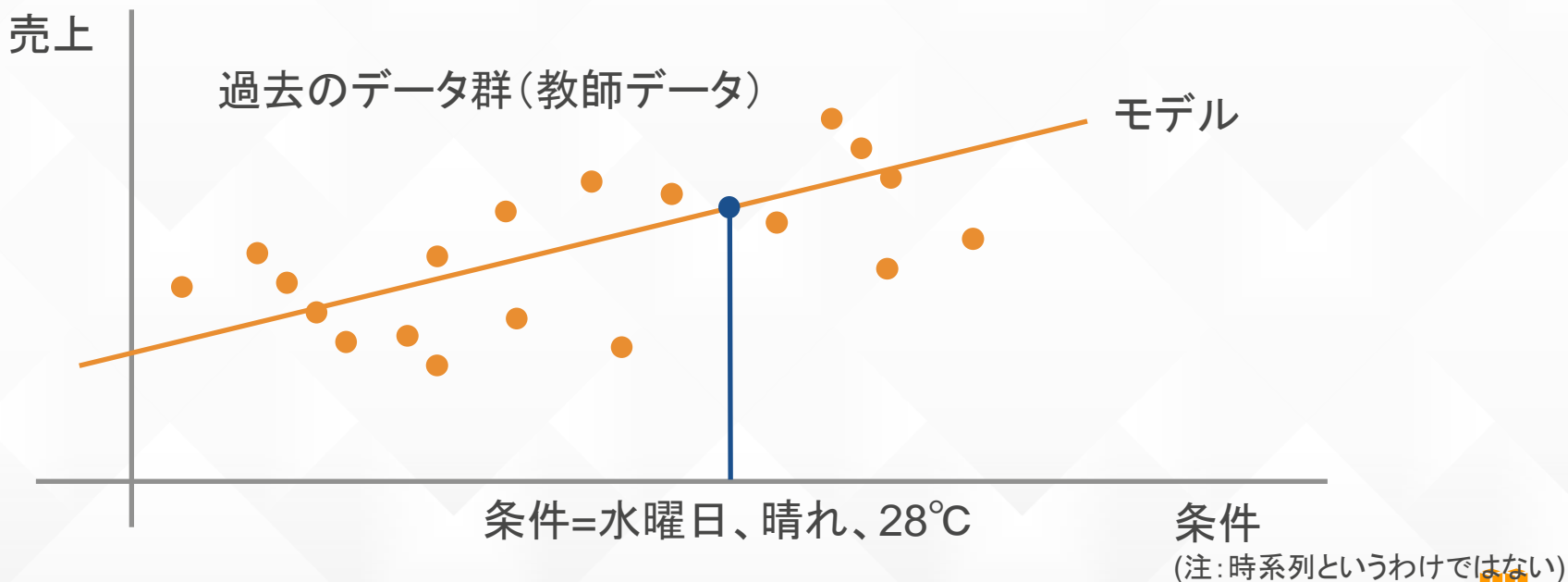
## 予測対象データ

商品名	価格	大きさ	メーカー	カテゴリ
E	500	...	EE	
F	600	...	FF	

未知の商品のカテゴリを予測する

# 機械学習の例

- 明日の売上はどのくらいになるか？
  - 過去の売上データなどをもとに明日の売上を予測する





# その他の例 ...

## 詐欺の検知

不正取引の検知、不正クリック検知、スパムeメールのフィルタリング、疑わしいレビューのマーキング ...

## パーソナライゼーション

コンテンツのレコメンデーション、予測的なコンテンツロード、ユーザエクスペリエンスの改善 ...

## ターゲットマーケティング

オファーとお客様のマッチング、マーケティングキャンペーンの選択、クロスセリングやアップセリング ...

## コンテンツ分類

ドキュメントのカテゴリ分類、履歴書と採用マネージャのマッチング ...

## 変動予測

サービスを使うのを止めそうなお客様の検知、無料ユーザからアップグレードのオファー ...

## カスタマーサポート

お客様からのメールの適切な転送先推測、ソーシャルメディアリスニング ...

# スマートアプリケーション(機械学習を有効活用したアプリケーション)が続々と登場しないのはなぜか?

1. 機械学習の専門家が**少ない**
2. 機械学習の仕組みを作り、スケールさせることは**技術的に困難**
3. モデルとアプリケーションのギャップを縮めるには、**長い時間と高い費用**が必要になる

# スマートアプリケーションを作るには

## 専門家

データサイエンティスト  
の人数は限られる

外注するのは高くつく

## 技術

多くの選択肢があるが  
決定的なものが無い

使いこなし、スケールさ  
せることが困難

カスタムソリューションを  
作成するために多くの細  
かい作業が毎回必要に  
なる

## 使いやすさ

複雑で間違いを起こし  
やすいワークフロー

特殊なプラットフォーム  
とAPI

モデルライフサイクル管  
理を無駄に再発明

# スマートアプリケーションを作るには

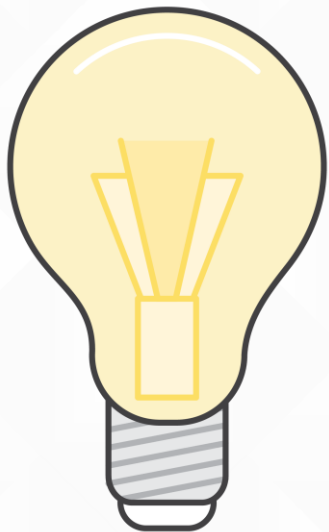
- 機械学習に強くて
- RやPython, 場合によってはHadoopやSparkに  
明るくて
- 特定のビジネス分野の経験が深い

こういう人を採用するのは難しい！



そこで . . .

# Amazon Machine Learningの登場



使いやすく、マネージド。開発者のために作られた機械学習サービス



---

# Part2: Amazon Machine Learning

# Machine Learning as a Service



## Amazonが提供するアルゴリズム

- 利用者は自分でアルゴリズムの実装や詳細なチューニングを行う必要がない

## パッケージサービスとしての提供

- 必要なワークフローが予め提供されている

## スケーラビリティ

- 利用者はシステムの拡張やその運用についても考える必要がない





# Amazon Machine Learningでできること

# 取り扱える予測モデルとアルゴリズム

## 二項分類

ロジスティック回帰

このメールはスパム？ Yes？ No？

## 多クラス分類

多項式ロジスティック回帰

これは本？ 車？ 食べ物？

## 回帰分析

線形回帰

あしたは水曜日。在庫はいくつくらい必要？

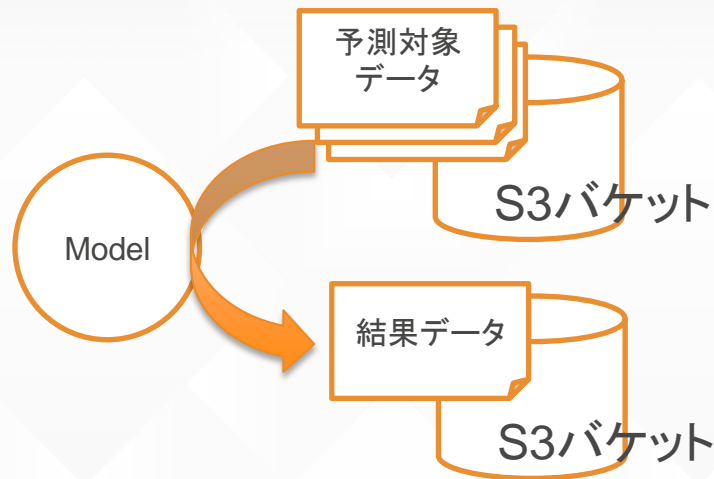


これらのモデルを使って . . .

# 予測手法

## バッチ予測

S3等にアップロードされた予測対象データに対してまとめて予測を実施



## リアルタイム予測

データ 1 件ずつAPIを使って予測を実施する

```
ml = Aws::MachineLearning::Client.new
record = { attr_A: 'foo', attr_B: 'bar', ... }
result = ml.predict(
  ml_model_id: MODELID,
  record: record,
  predict_endpoint: ENDPOINT
)
```



---

# Amazon Machine Learningの使い方

# 4つのステップ

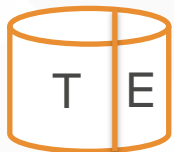
1. 教師用/評価用データを準備
2. モデルを作成(学習、トレーニング)
3. モデルの品質評価
4. 実際の予測の実施

# 1. 教師用/評価用データを準備

## Data Sourceの作成

S3、Amazon Redshift、RDS上のMySQLに格納されたデータを指定し、教師データ/評価用データとして利用する

s3://SOURCEDATA



S3、Redshift、RDSが  
利用可能

デフォルトの設定を使うと、自動的に7割を教師データ(T)、3割を評価用データ(E)に分割して管理してくれる

## 2. 教師データからモデルを作成

教師データを元にAmazon Machine Learningが自動的にモデルを選択してくれる。例えば予測対象のカラムが二値型であれば二項分類が自動的に選択される。

### 二項分類

ロジスティック回帰

### 多クラス分類

多項式ロジスティック回帰

### 回帰分析

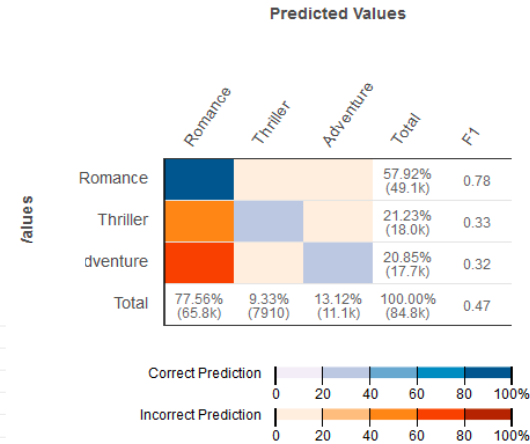
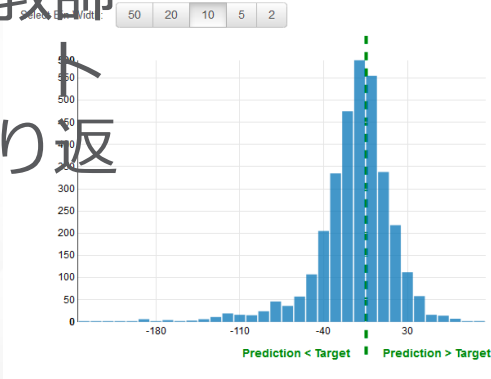
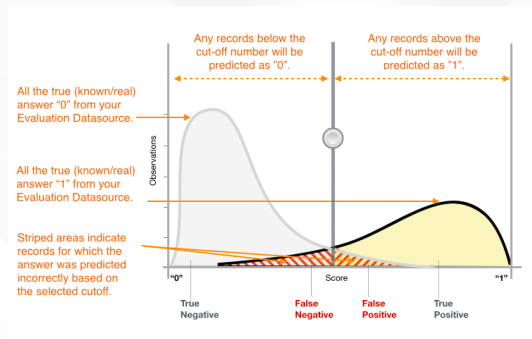
線形回帰



# 3. モデルの品質評価

作成したモデルに対して評価  
(評価用のデータを流してみ  
て予測の精度を測ること)を  
実施する。

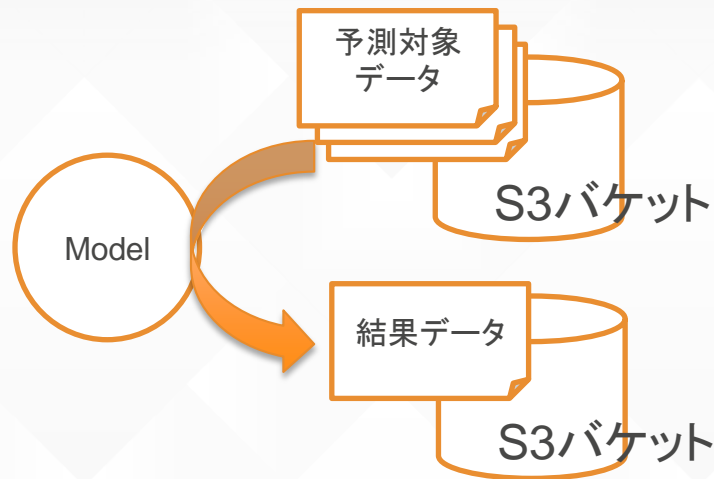
精度に満足できない場合、  
データのETLや量を精査し、  
レーニングと品質評価を繰  
り返す。



# 4. 実際の予測の実施

## バッチ予測

S3(S3、Redshift、RDS) 等にアップロードされた予測対象データに対してまとめて予測を実施



## リアルタイム予測

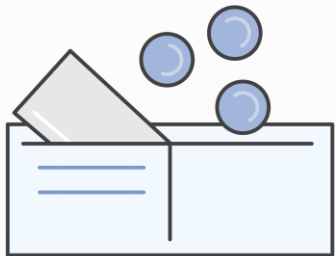
データ 1 件ずつAPIを使って予測を実施する

```
ml = Aws::MachineLearning::Client.new
record = { attr_A: 'foo', attr_B: 'bar', ... }
result = ml.predict(
  ml_model_id: MODELID,
  record: record,
  predict_endpoint: ENDPOINT
)
```



# 料金

# 使った分だけ、安価な支払い



データ分析、モデルトレーニング、評価：  
**\$0.42/インスタンス時**

バッチ予測: **\$0.10/1000**

リアルタイム予測: **\$0.10/1000**

+ 1時間毎のキャパシティリザーベーションチャージ(モデル  
サイズ10MBあたり\$0.001)



# リージョン

# リージョン

- 現在のところus-east-1のみ
- ただし他のリージョンのS3も利用可能



# ユースケースを考えてみる

# 機械学習を活用するために

- 機械学習をうまく活用するためには、自分たちが解決したい問題を予測モデルに対してうまく落としこんでやることが非常に重要
- ここから何枚かのスライドでは、その落とし込みの例について考えてみる



# 広告の不正クリック検出

- 教師データ
  - 過去のクリックログ。不正かそうでないかを別の手法で判断してフラグがついたもの。
- 問題の分類
  - 二項分類
- 予測
  - クリックログを入力し、ロガー行ごとに不正クリックかそうでないかを判定

# デモグラ推定

- 教師データ
  - デモグラがわかっているユーザーの行動ログ
- 問題の分類
  - 多クラス分類
- 予測
  - ユーザーを行動ログからデモグラを判定する

# デモグラに基づいたレコメンデーション

- 教師データ
  - 購入履歴にデモグラ情報がマッピングされたもの
- 問題の分類
  - 多クラス分類
- 予測
  - ユーザーのデモグラを入力し、このユーザーはF1なので商品カテゴリ●●、みたいな出力

# 顔写真から特定の人物かどうかを判定する

- 教師データ
  - 顔写真をグレースケールにしたもののビットマップ
- 問題の分類
  - 二項分類
- 予測
  - 顔写真を入力し、●●さんである/違うを判定する

# 問題をどうモデルに落としこむかが非常に大事

- 教師データの量
- 教師データの形式、入力値
- どのモデルを使うのか
- 何の値を予測/判定するのか



---

## Part3: アーキテクチャへの組み込み

# EMRを使用したバッチ予測



# Amazon Redshiftを使ったバッチ予測





# インタラクティブアプリケーション用のリアルタイム予測



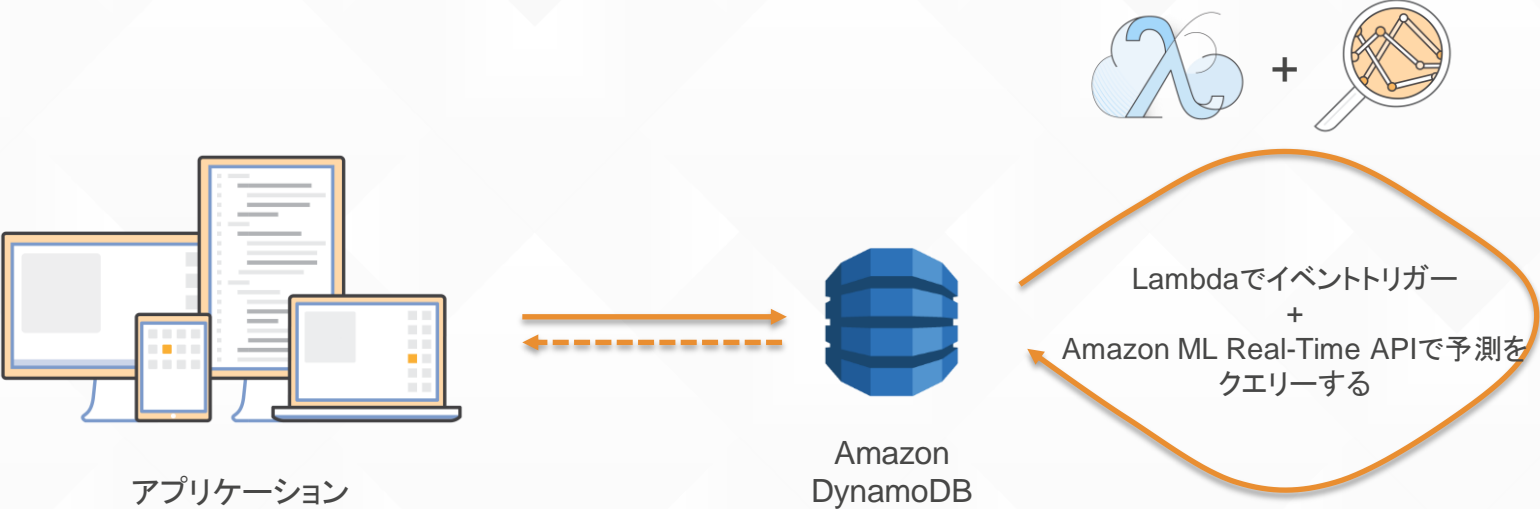
アプリケーション



Amazon ML Real-Time API  
で予測をクエリーする



# 既存データフローに予測を追加する





さいごに

# まとめ: Amazon Machine Learning

- 機械学習の導入を非常に容易にしてくれる。機械学習やそのためのシステムについての専門家がいなくても可能。
- S3やRedshiftにデータがあればすぐにでも使い始めることができる。

溜めていたデータの評価や、その先のビジネスへの活用を簡単に始めることができる。



Thank You