

# 1人でも使えるSageMaker

クックパッド株式会社 染谷悠一郎

2019/03/12 第3回 Amazon SageMaker事例祭り + 体験ハンズオン

# 自己紹介

# 染谷 悠一郎 [ayemos]

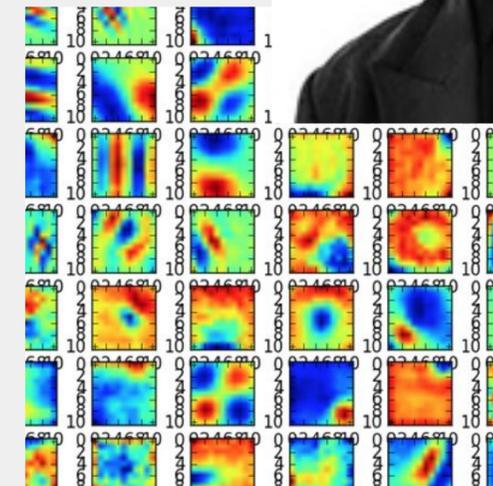
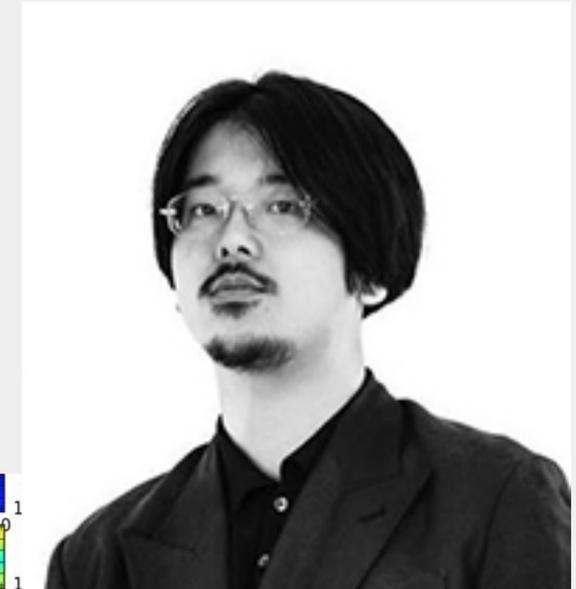
# 東工大 計算工学専攻 修士

# クックパッド株式会社 2016年新卒入社

# 研究開発部 エンジニア

# 機械学習基盤、画像認識系の研究開発

# <https://speakerdeck.com/ayemos>



**aws**  **CERTIFIED**

 Solutions Architect - Associate

 Developer - Associate



300万 レシピ

料理名・食材名 目的・用途 レシピ検索

クリスマスケーキ クリスマス おにぎり ホットサンド お弁当

レシピ関連サービス

- みんなのレシピ
- プロのレシピ
- 献立
- 料理動画
- 話題のキッチン

プレミアムサービス

- 人気順検索
- レシピランキング
- 殿堂入りレシピ
- 専門家厳選レシピ
- プレミアム献立 [もっと見る](#)



11月27日の おすすめ **ねぎと大葉の肉巻き＊バター醤油風味＊**  
Blue-Island

ニュース

**サクっもち食感♪片栗粉ドーナツ**



# クックパッドと機械学習

# 300万品以上のレシピデータ

# タイトル、手順、材料、分量、各種写真

# レシピ以外のデータ

# レシピを元に料理を作ったユーザーによる報告(つくれぽ)

# 検索、その他行動ログ

# 加えて海外も

16/04/17

[x]



スペインのジャガイモオムレツの伝統レシピ

by 火水流整体術院

おいしく作れました！

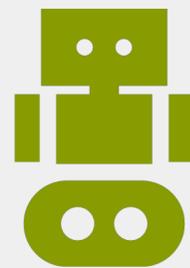
# クックパッドと機械学習

- # 2016年7月に研究開発部が発足
- # 初期メンバー3名、現在は10~15名の体制で機械学習に取り組む
- # <https://techlife.cookpad.com>
- # レシピの画像検索に必要な技術: <https://techlife.cookpad.com/entry/2018/08/28/162000>

# 料理きろく



スマートフォンの写真の内、  
料理写真を自動的に収集/記録



畳み込みニューラルネットワーク  
による料理画像の自動認識

# 料理きろく



27万人以上のユーザーが利用

3000万枚以上の料理写真を記録

(2018年12月10日現在)

# その他研究開発

## # 料理写真の類似画像検索

<https://techlife.cookpad.com/entry/2018/08/28/162000>

## # ユーザーからのご意見の自動タグ付け

<https://techlife.cookpad.com/entry/2018/08/08/170000>

## # 材料名の正規化 ([シヨージュ, おしょうゆ] => 醤油)

<https://techlife.cookpad.com/entry/2017/10/30/080102>

## # その他色々

# クックパッドとSageMaker

- # 現状の学習環境はGPUインスタンスのオンデマンド利用
- # 学習前後もインスタンスが立ち上がっており、ロスタイムが大きい
- # 学習環境としてSageMaker Training Job環境の利用を検討中

# クックパッドとSageMaker

- # 現状の学習環境はGPUインスタンスのオンデマンド利用
- # 学習前後もインスタンスが立ち上がっており、ロスタイムが大きい
- # 学習環境としてSageMaker Training Job環境の利用を検討中

# 目次

# コンテナを用いた機械学習の利点とSageMakerについて

# babysage の紹介

# 「今日から一人でSageMakerのTraining Jobを使って  
独自モデルを学習したい人向け」のお助けツール

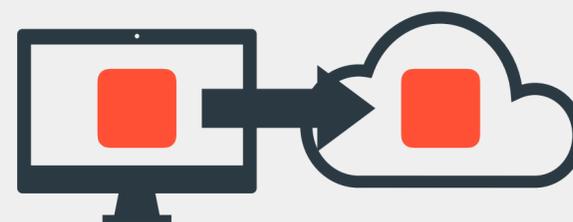
# コンテナと機械学習

# 『学習に必要な環境を揃えたdocker imageを作り、  
コンテナとして学習を実行し、  
結果(モデル)をコンテナから取り出す』

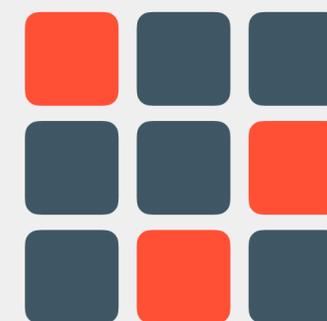
# 以下のような利点がある



環境の再現性の向上



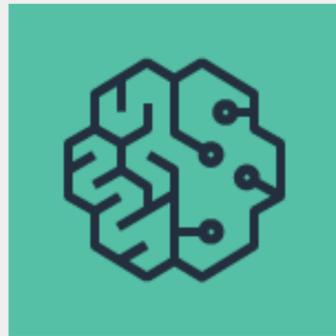
可搬性の向上



分散実行

# コンテナと機械学習

# コンテナ + 機械学習の例



# SageMakerとは

# Notebook, Training Job, Endpoint

# 今回はTraining Job、特に独自モデルを学習するケース  
について話す

# Training Jobとは

# Training Jobとは

- # 機械学習向けに拡張されたコンテナ実行環境
- # + 学習/定義済みモデルなど、AWSが用意したリソースを活用できる
- # GPUを搭載したP3 インスタンスなど実行環境を選べる
- # 基本は、『docker imageを指定した環境で実行できる仕組み』
- # AWS Batch +  $\alpha$  と見ることも出来る

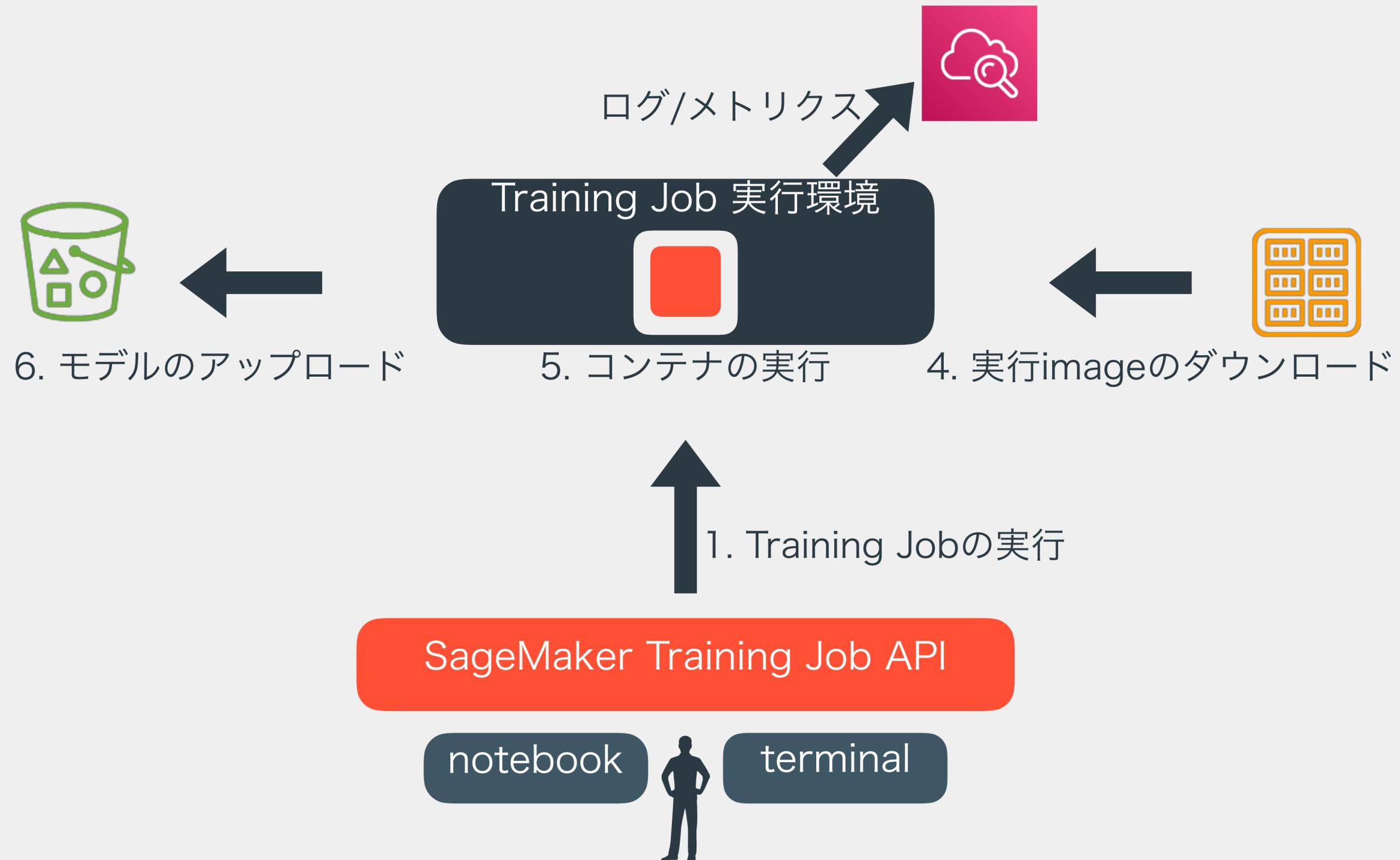
# Training Jobとは



# Training Jobとは



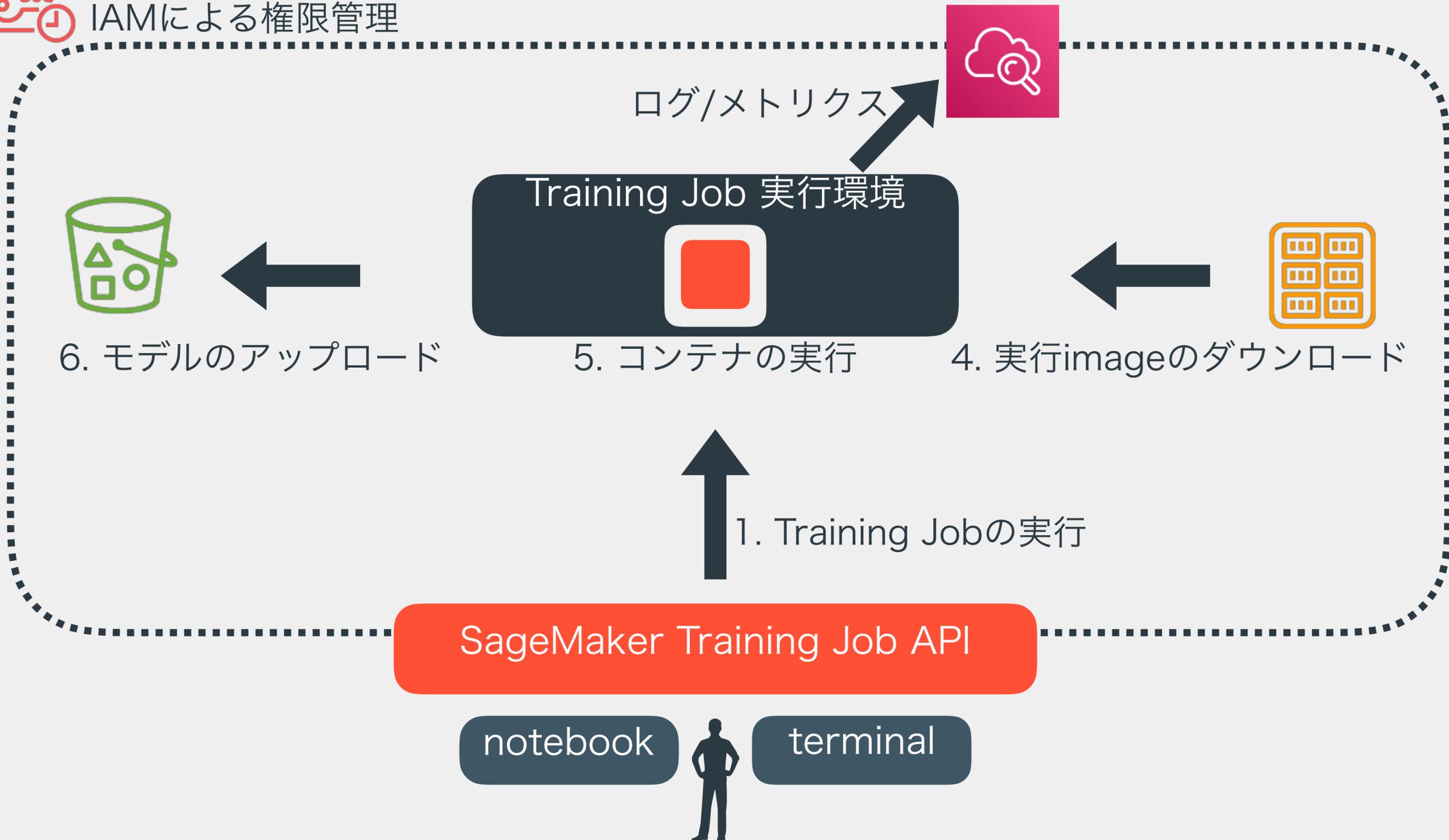
# Training Jobとは



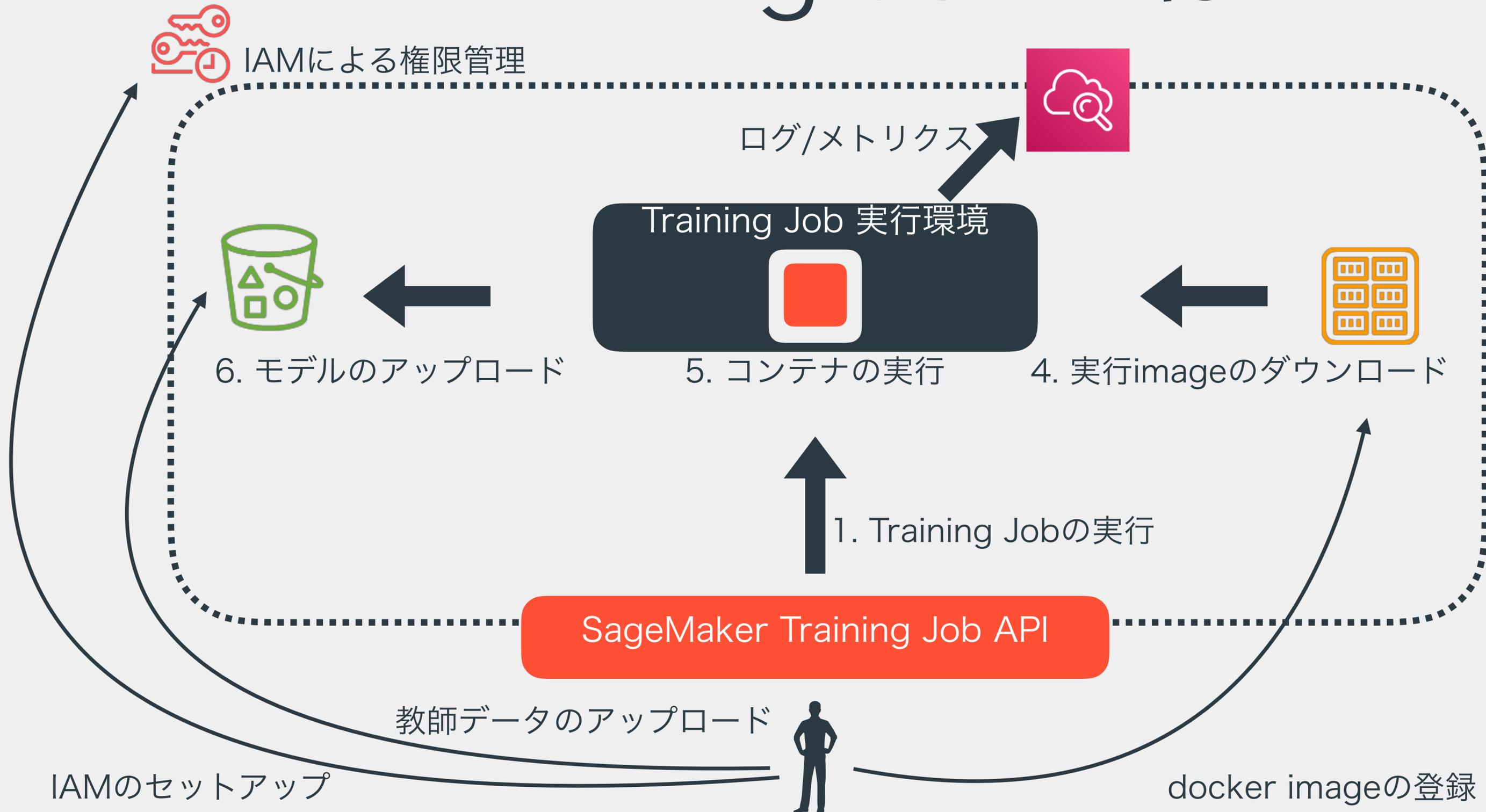
# Training Jobとは



IAMによる権限管理



# Training Jobとは



# 意外と険しいコンテナへの道

- # コンテナの利点は活かしたい、SageMakerを使いたい
- # 一方で独自モデルの学習を行うには覚えることも多い
- # AWSの他のリソースの使い方等 例: ECR, S3, CloudWatch

# 意外と険しいコンテナへの道

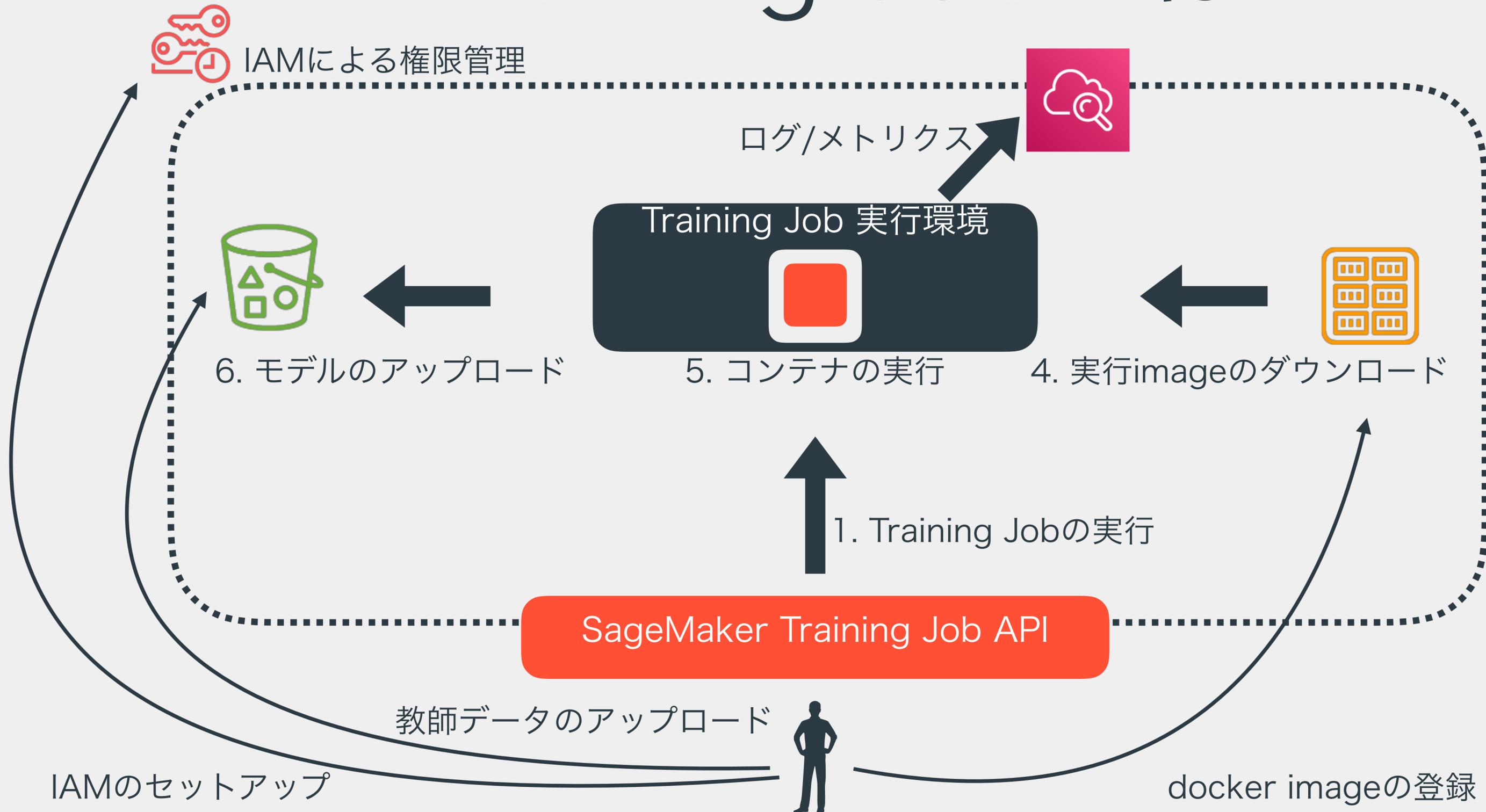
- # コンテナの利点は活かしたい、SageMakerを使いたい
- # 一方で独自モデルの学習を行うには覚えることも多い
- # AWSの他のリソースの使い方等 例: ECR, S3, CloudWatch
- # これらは他の用途にも使える汎用なツール
- # SageMakerから利用する限りでは全ての機能/APIを覚える必要はない

# babysage

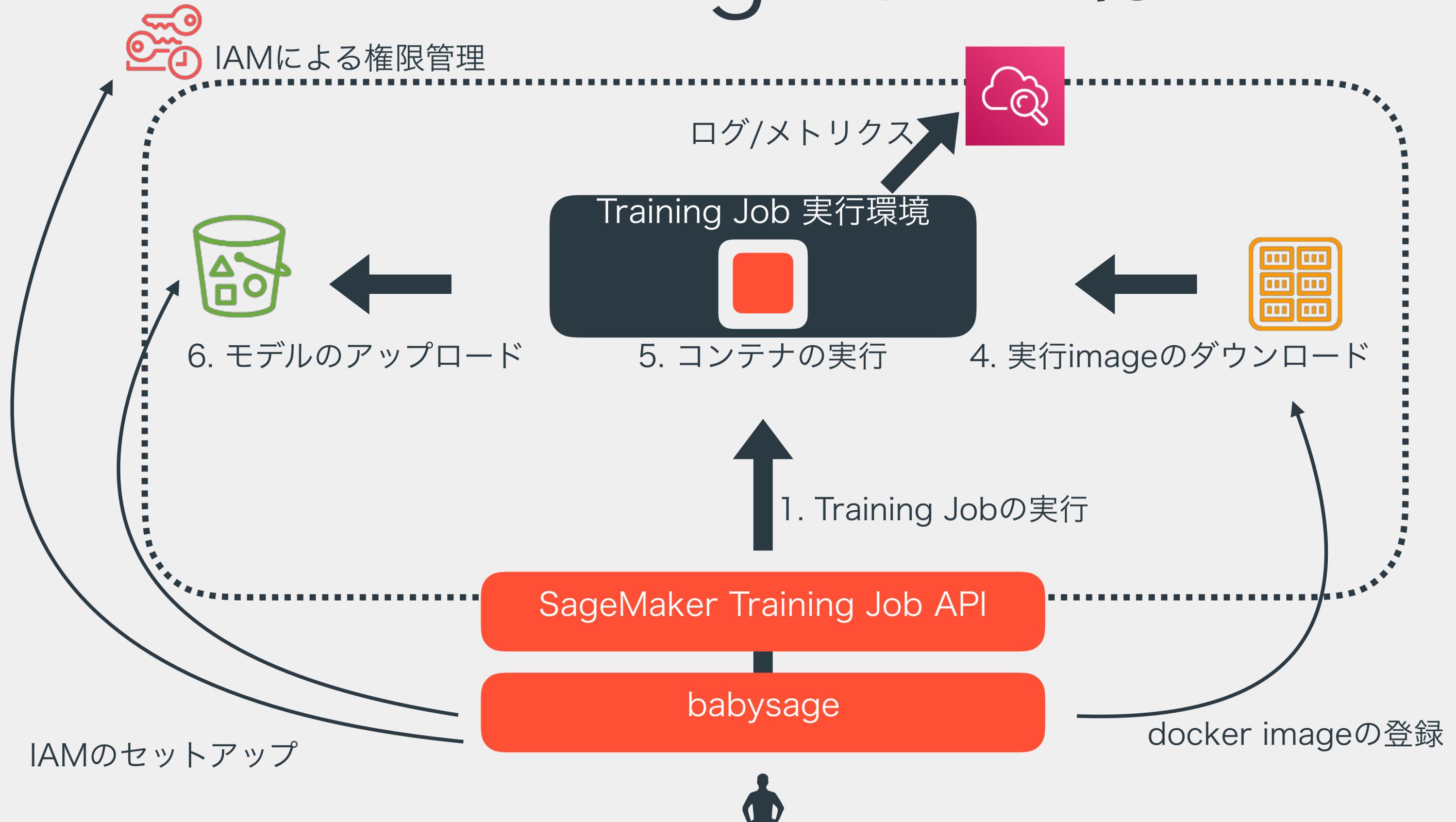
- # SageMaker (Training Job)を使うときの補助輪となるツール
- # 実験 (experiment) のテンプレート出力
- # ECR, S3などの設定サポート
- # ローカル, Training Jobでの実行をシームレスに行う
- # その他 (ログ、CPU/GPU監視等)  
(CloudWatchの利用サポート)

デモ

# Training Jobとは



# Training Jobとは



# babysage

# SageMaker (Training Job)を使うときの補助輪となるツール

# <https://github.com/ayemos/babysage>

# 近日公開予定！🙏👷🙏

# まとめ

- # SageMakerのTraining Jobは機械学習用コンテナ実行環境
- # 豊富なマシンリソースの利用、複数実験の並列実行などの利点
- # 一方で利用するまでに覚えることも少なくない
- # babysageを使って最初の障壁を突破しよう