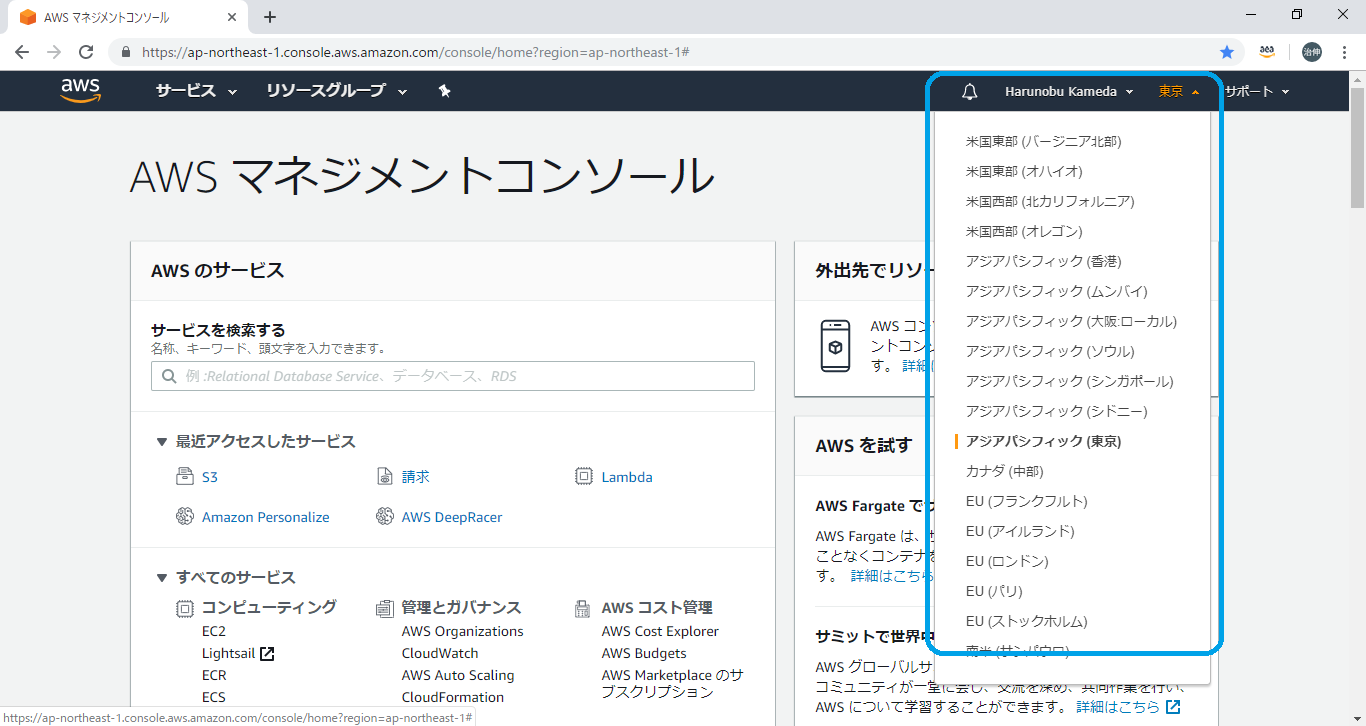
Amazon Personalize ハンズオン (No.Codes)

はじめに：Amazon PersonalizeはAWSが提供している過去の購買履歴などから個人化レコメンデーションを行う機械学習サービスです。データの投入や学習などで待ち時間が発生するため、Amazon Forecastハンズオンとの同時並行をお勧めします。

1. ハンズオンを行うリージョンを画面右上から選択。リージョンは講師の指示に従ってください。



1. 映画の視聴履歴ファイルをダウンロード

https://github.com/harunobukameda/Amazon-Personalize-Handson/blob/master/ratings.csv

ユーザー610人による9700本の映画視聴履歴とその評価が10万件以上ファイルに含まれています。

USER\_ID : ユーザーの識別ID（商用環境では別途ユーザーマスタと連結してください）

ITEM\_ID：映画のID（商用環境では別途アイテムマスタと連結してください）

TIMESTAMP：映画批評を入力された時間です。（このハンズオンでは用いませんが、Amazon Personalizeの実行では必須なカラムです）

Amazon Personalizeはこれらのデータをもとに、批評の相関図を作成しレコメンドを行います。従来これらのデータ分析は、Amazon EMRなどを用いたヒューリステック分析が必要でしたが、Amazon Personalizeではこれらが自動化されることが特徴です。

例）User (1) が 映画A、B、Cを視聴した場合、User(2)がBを視聴した場合、AとCがレコメンドされる。レコメンド以外にも、データの相関関係を見出す用途で様々な用途が今後想定されます。また、本シナリオでは使用しませんが、CSVファイルにはRating（視聴済映画に対するユーザーからの評価）も準備されており、ユーザー評価に基づいたさらに細かいレコメンドがAmazon Personalizeでは可能です。

３．Amazon S3 バケットに先ほどDLしたratings.csvをアップロードします。S3バケットは既存のものを作成してもよいですし、新規で作成しても問題ありません。S3バケットは以下のバケットポリシーを設定してください。すでにバケットポリシーが設定されている既存S３バケットでの作業はお勧めしていません。

{

"Version": "2012-10-17",

"Id": "PersonalizeS3BucketAccessPolicy",

"Statement": [

{

"Sid": "PersonalizeS3BucketAccessPolicy",

"Effect": "Allow",

"Principal": {

"Service": "personalize.amazonaws.com"

},

"Action": [

"s3:GetObject",

"s3:ListBucket"

],

"Resource": [

"arn:aws:s3:::*bucket-name*",

"arn:aws:s3:::*bucket-name*/\*"

]

}

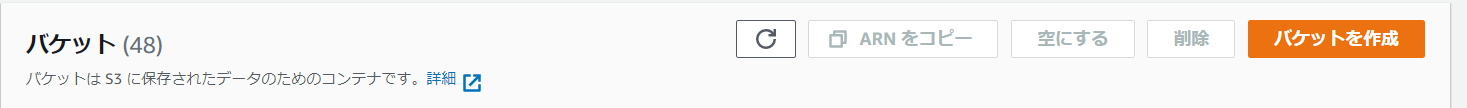
]

}

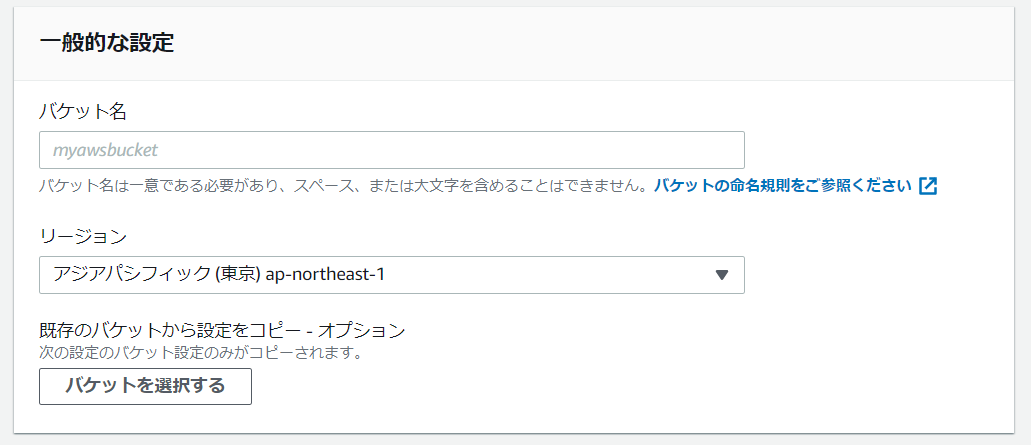
*bucket-name　はご自身で作業するバケットの名前を入れてください。*

S3バケットを新規に作成する場合は、全てデフォルトで作成してください。

【バケットを作成する】のボタンを押します。



バケット名は全AWSユーザーの全リージョンで一意である必要がありますので、ある程度長い名称にします。リージョンは、講師から指示されたリージョンを指定してください。



バケットが作成されたらバケット名をダブルクリックして詳細画面へ遷移し、【アクセス許可】タブを押します。

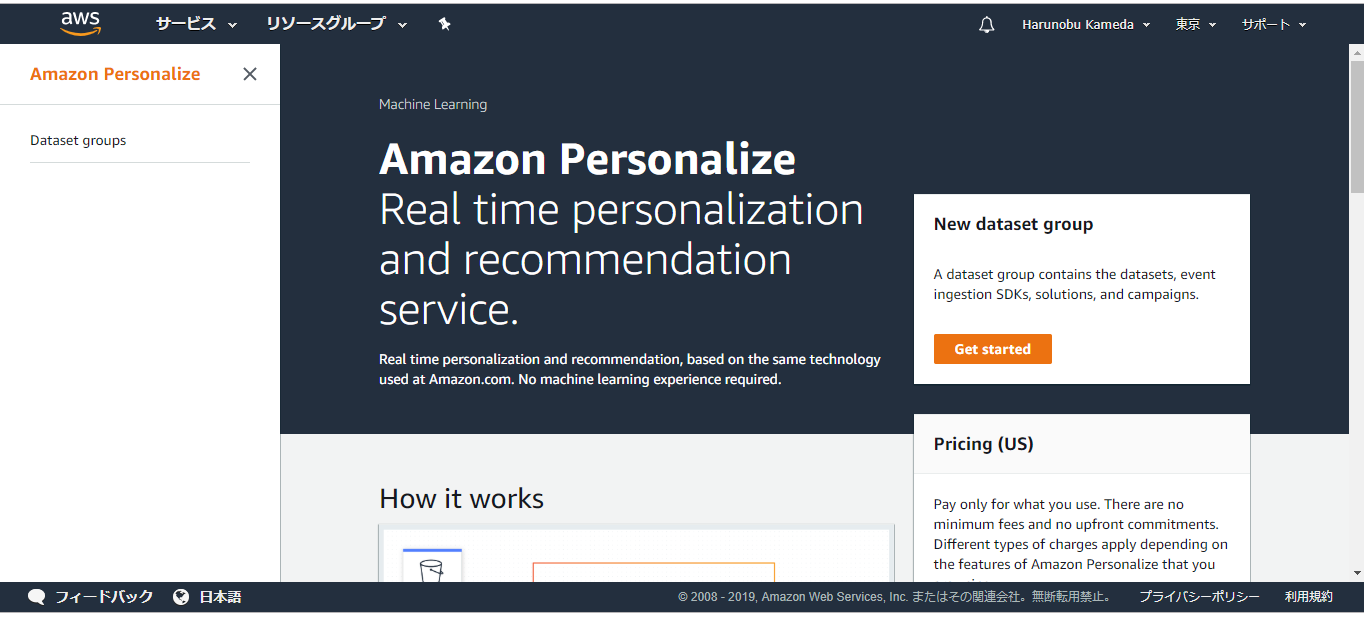


バケットポリシーの【編集する】をボタンを押し、上記のポリシー内容をコピペし、【変更の保存】を押します。

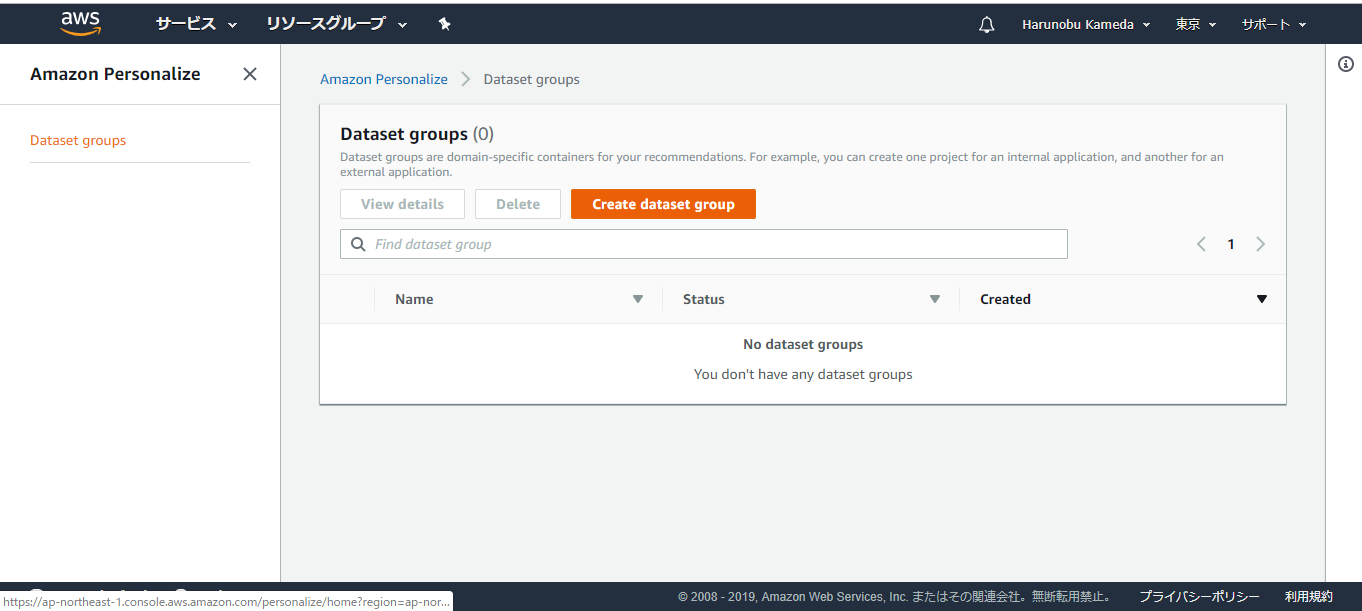


４．Amazon Personalizeのマネージメントコンソール（<https://console.aws.amazon.com/personalize/> ）にアクセスします。

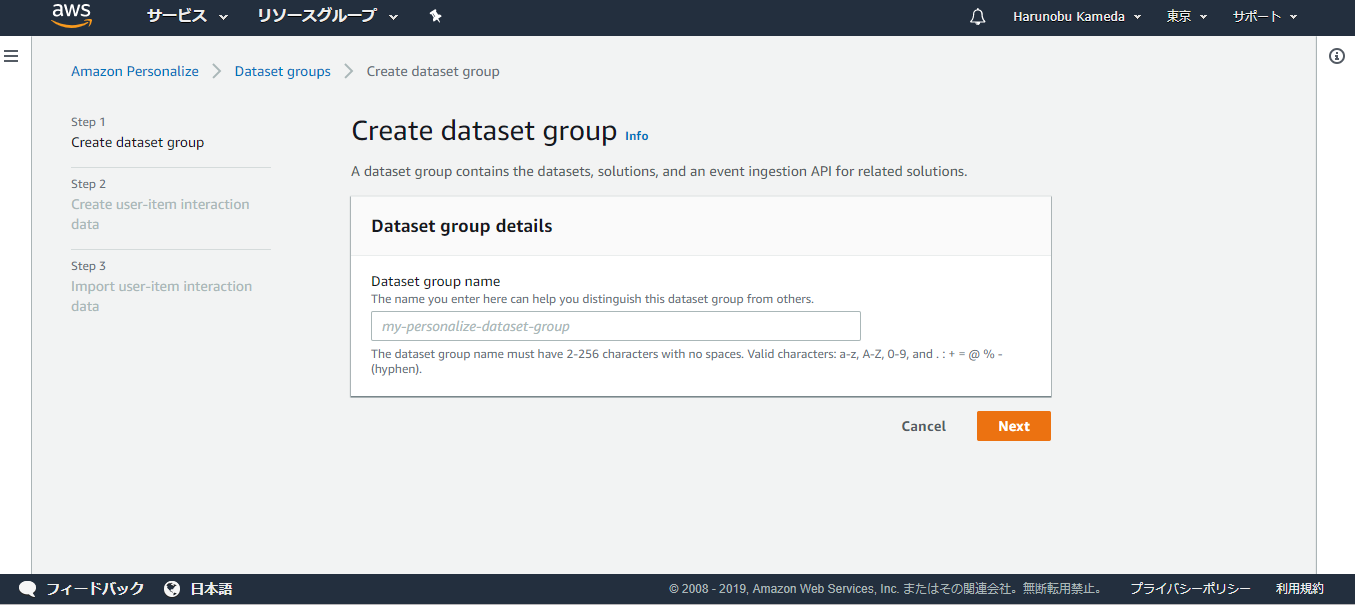
＃先ほど指定したリージョンになっていることを改めて確認します。



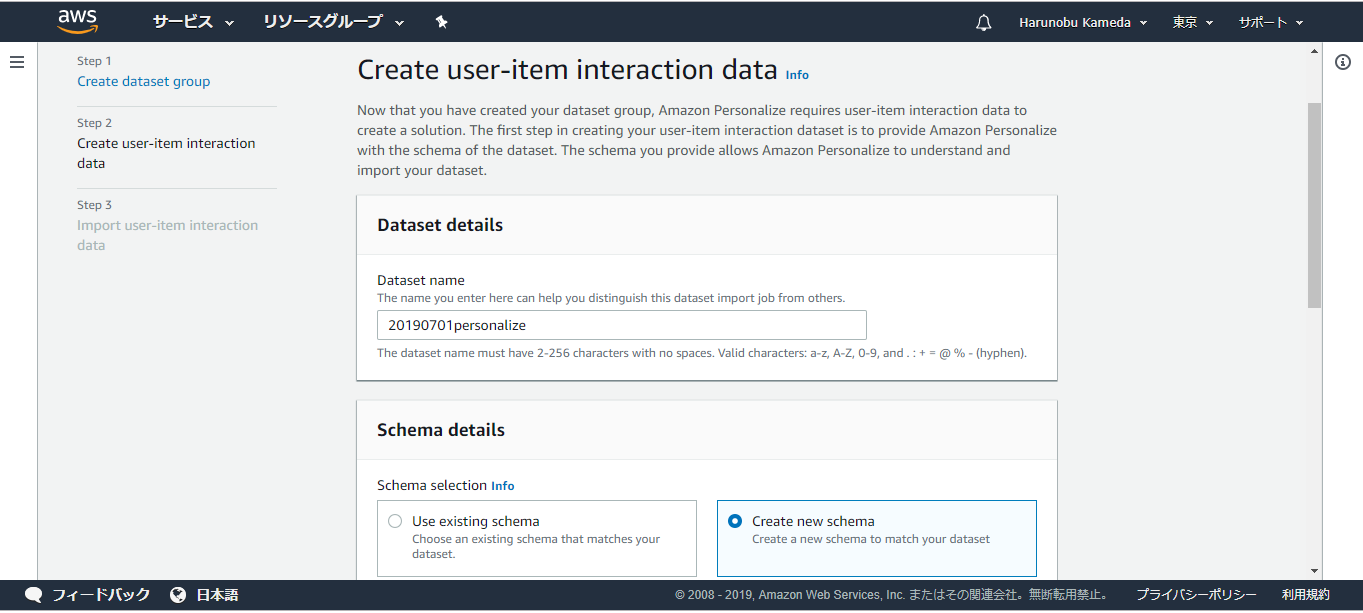
５．画面左の [Dataset groups]を選択し、[Cerate dataset group]を押してください。



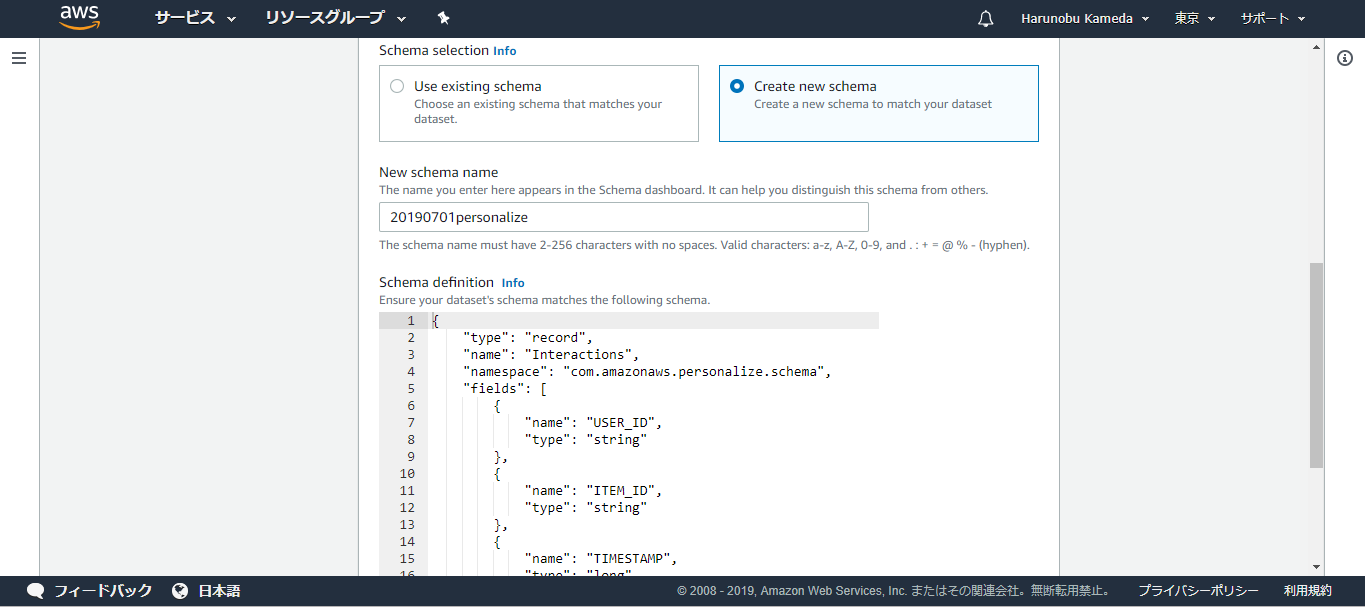
６．作成するDataset Groupに任意の名前を付けてください。



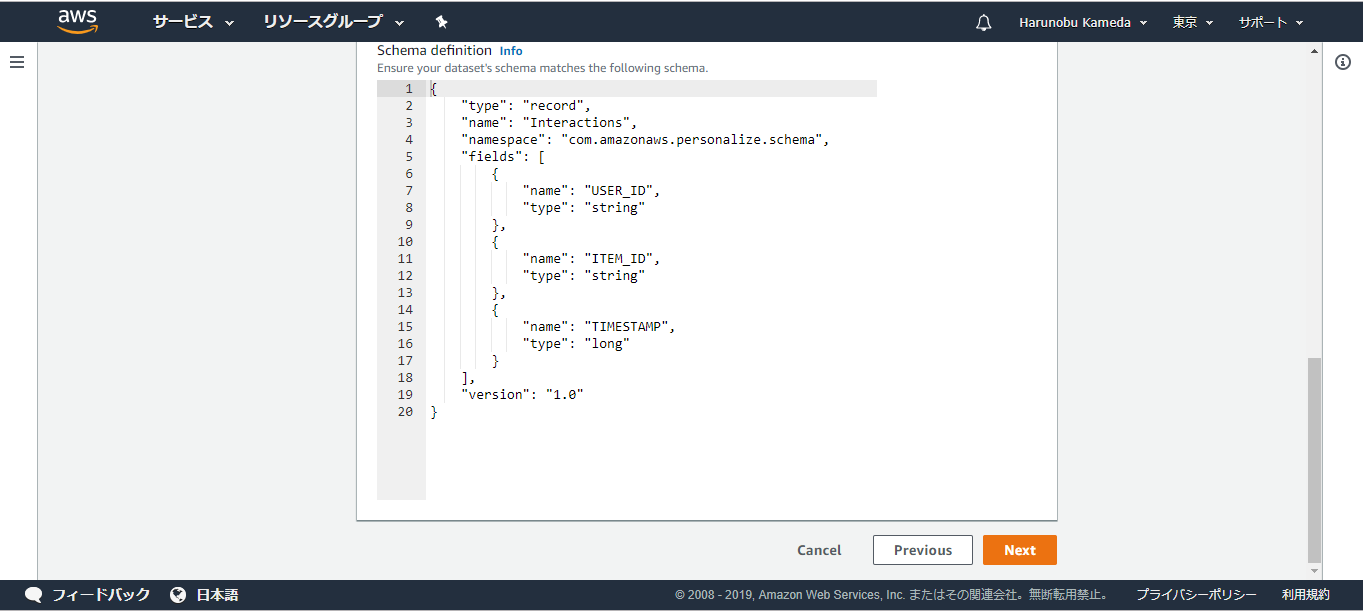
７．作成するDatasetに名前を付けます



８．[Create new schema]を選択し、Schemaに任意の名前を設定します。

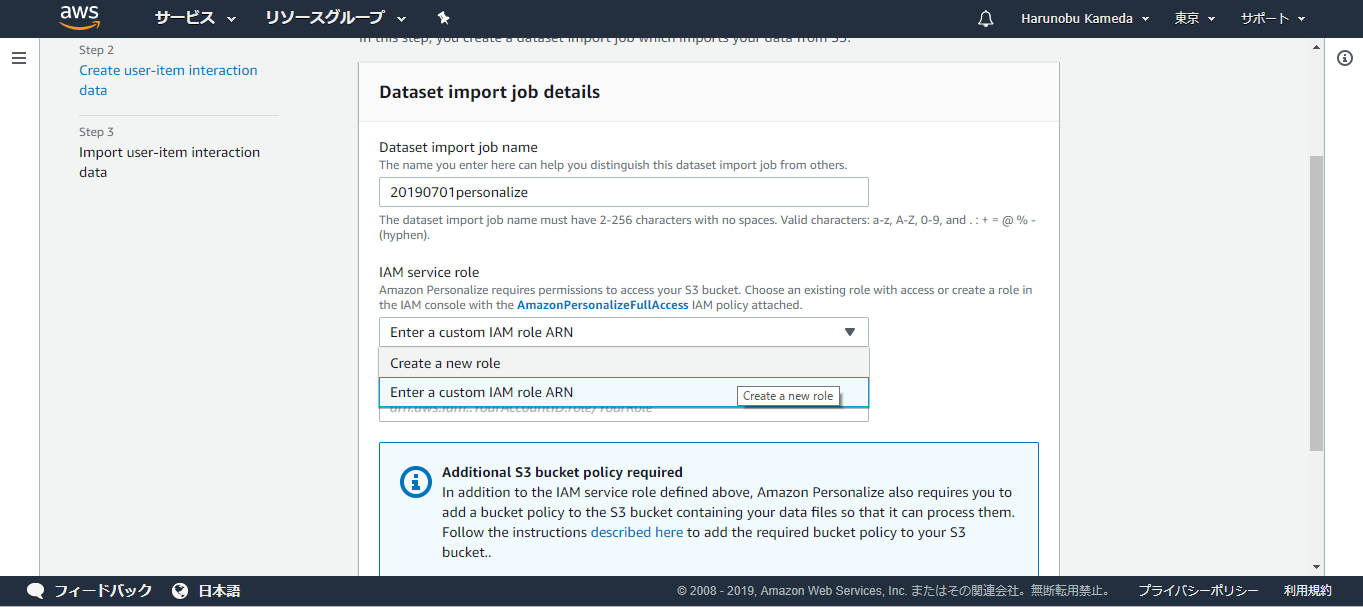


９．[Next]を押します

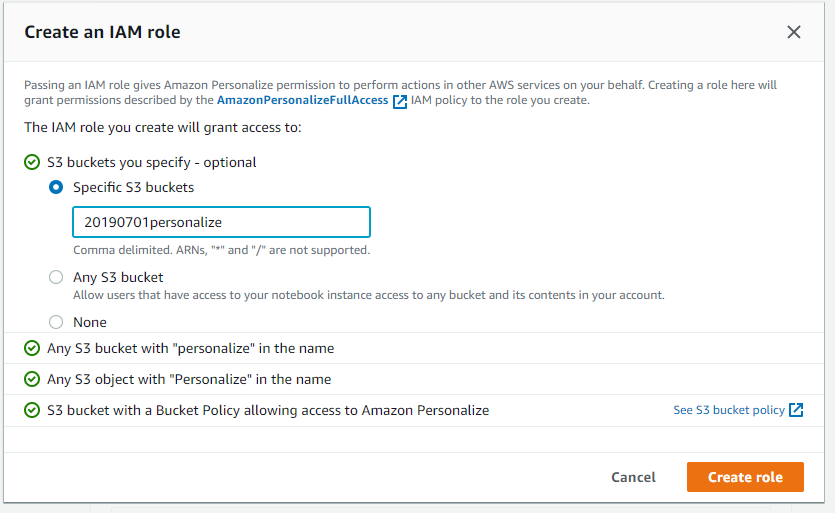


１０．[user-item interaction data] (先ほどのratings.csv)をインポートするjobの起動を行います。[Dataset import job name]に名前を入力します。

[IAM service role]で[Create a new role]を選択します。

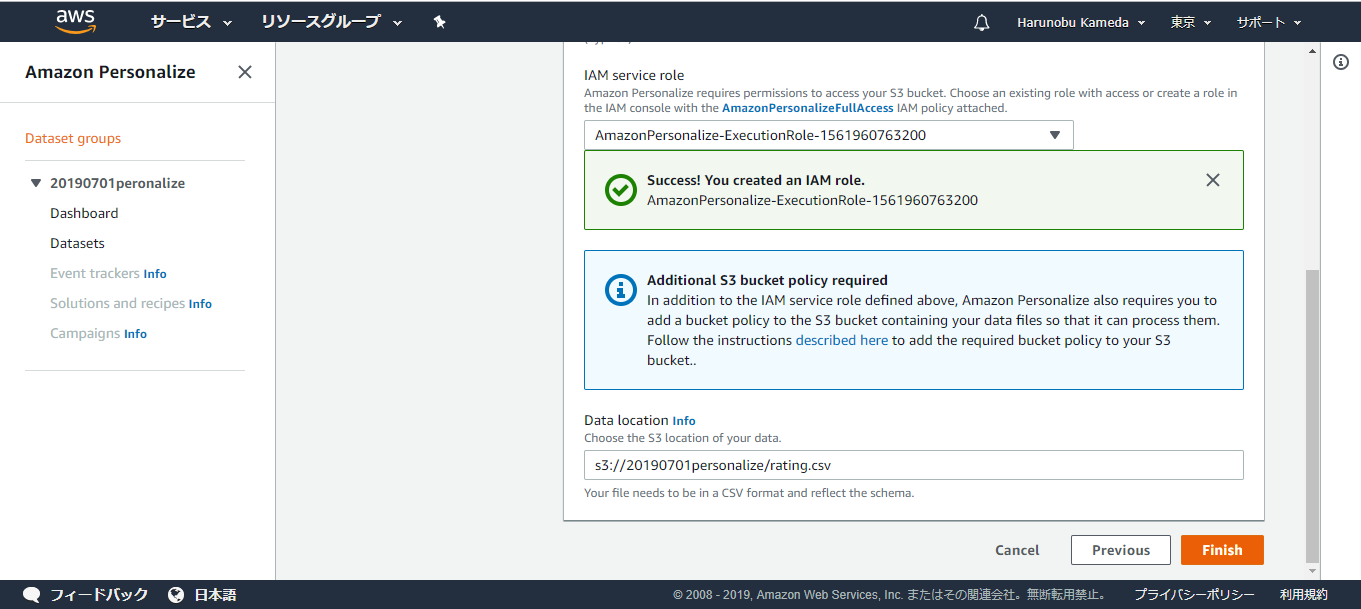


１１．先ほど作成したS3バケットの名前を[Specify S3 buckets]に入力して[Create role]を押します。



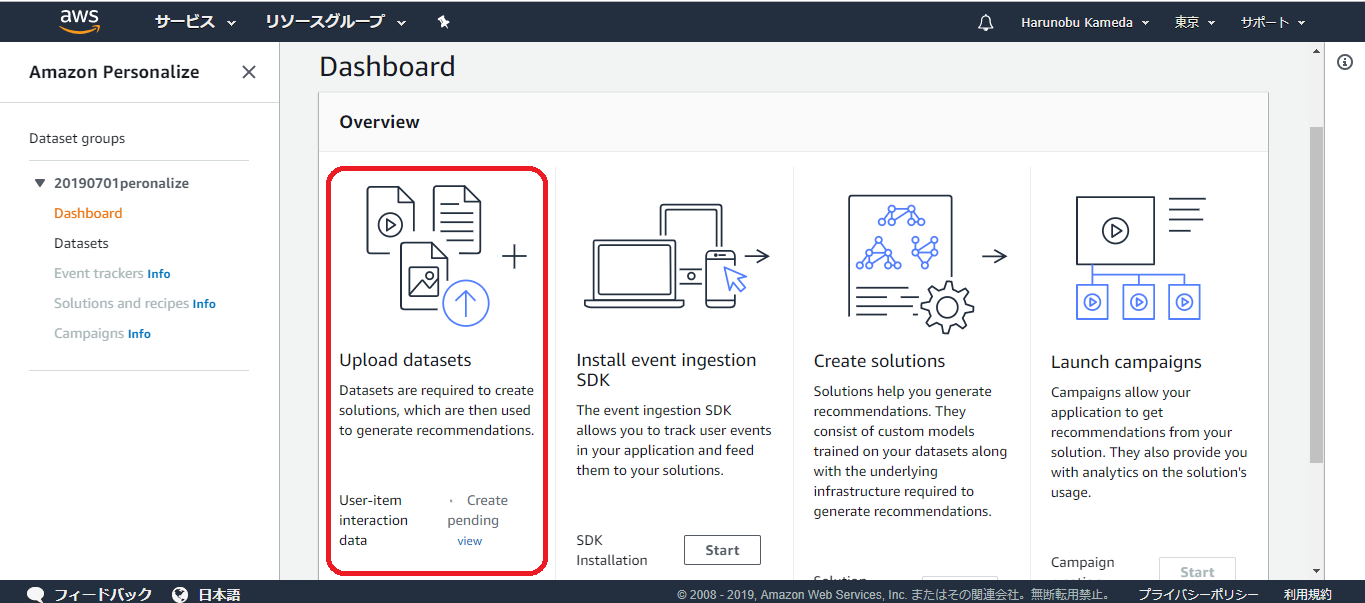
１２．[Success! You created an IAM role.]が表示されていることを確認します。

[Data location]にs3://<bucketname>/ratings.csvを入力します。作業が完了したら[Finish]ボタンをおします。



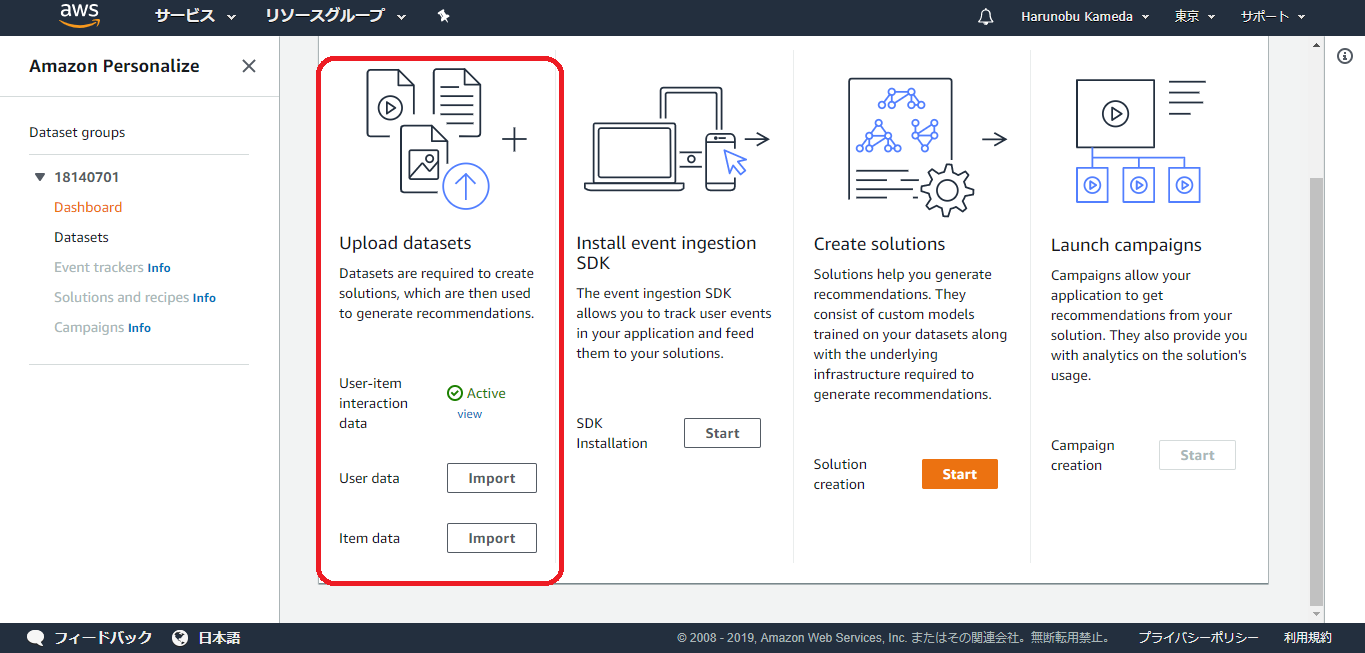
１３．ステータスが[Create Pending] or [Create in Progress]になりますので少し待ちます。

Finishボタンを押しても画面が遷移しない場合、ブラウザ上部の見えないところでエラーが出ていますので、右側のブラウザスクロールバーで一番上まで戻りエラーメッセージを確認してください。（ほとんどのケースにおいて、S3のパス、もしくはファイル名のタイプミスです）

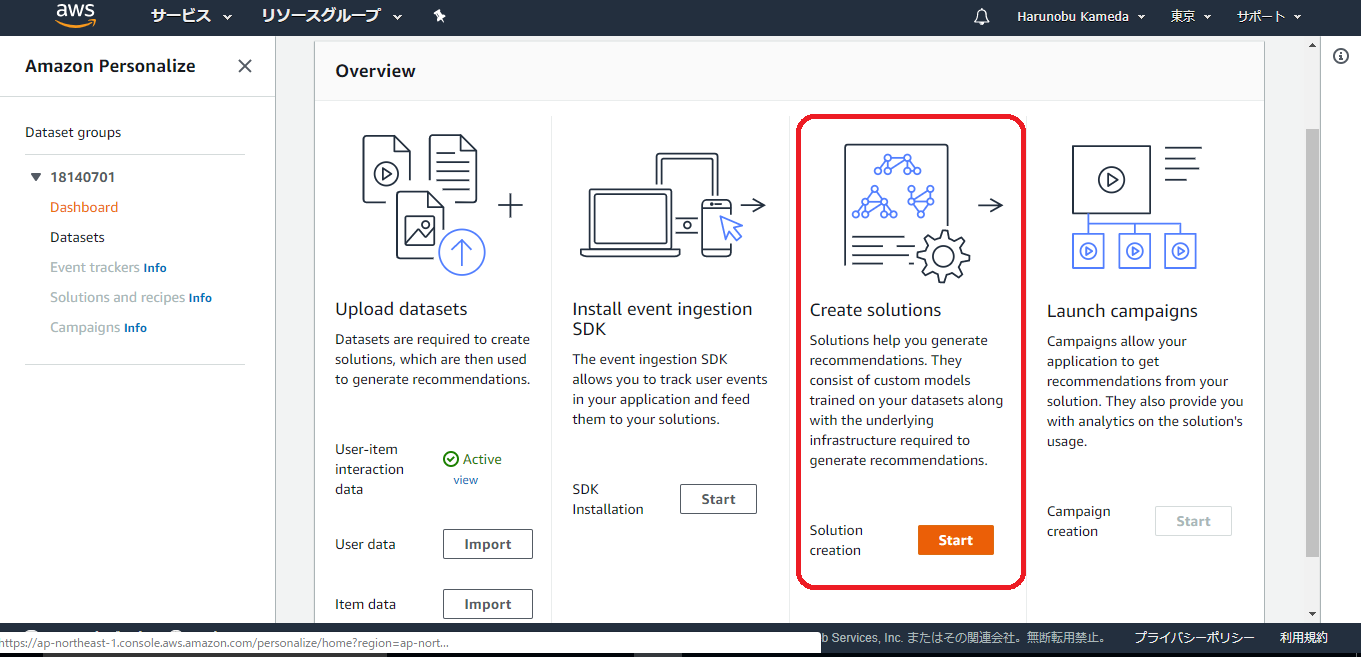


１４．取り込んだデータ(ratings.csv)の読み込みに少し時間がかかります。（20分ぐらい）。待ち時間の間に、Forecastのハンズオンを続行します。

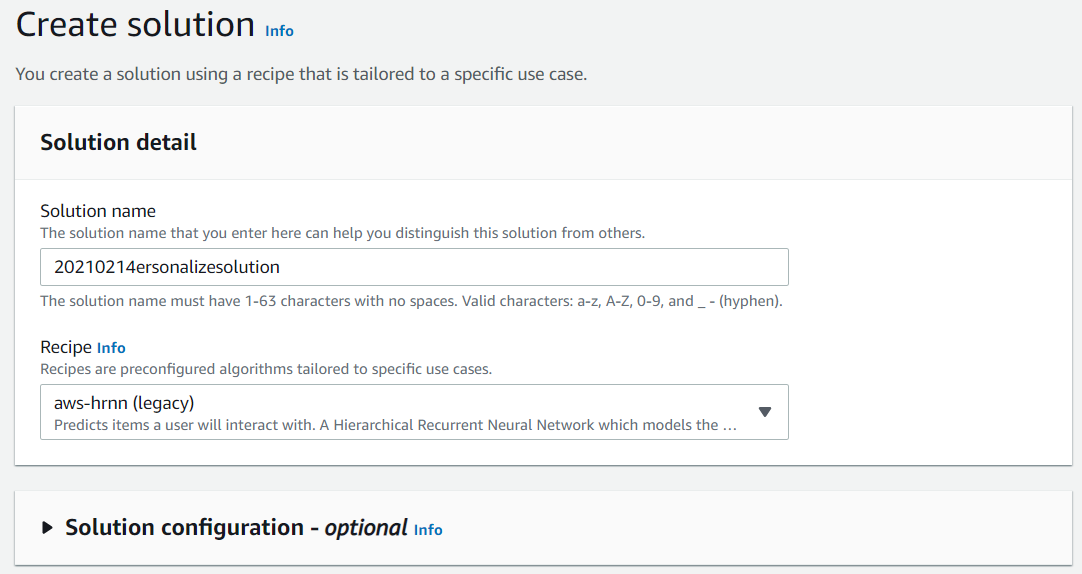
[Active]と表示されれば完了です。

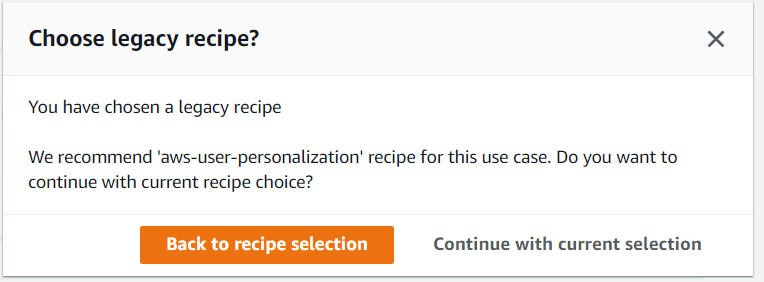


１５．[Create Solutions]の[Start]を押します。読み込んだdatasetsをもとに学習を開始します。

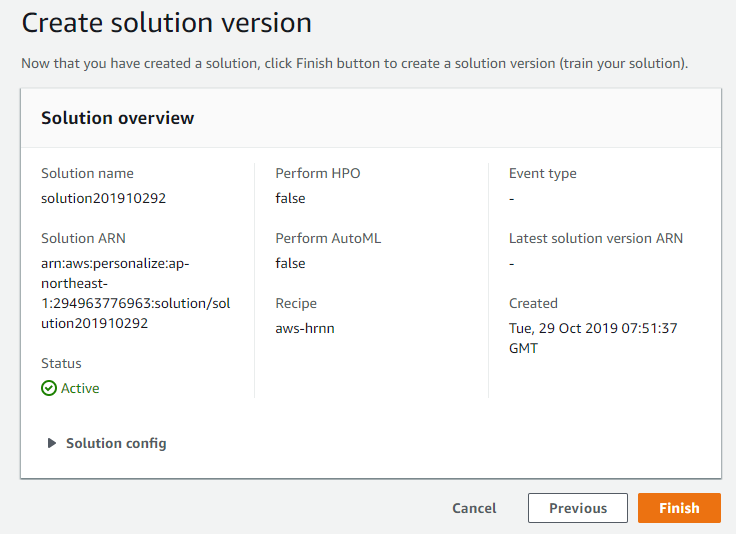


１６．[Solution Name]を入力し、[Manual]を選択します。ドロップダウンリストから[aws-hrnn]を選択し以下の画面のようになれば、[Next]を押します。次に出るダイアログで[Continue with current selection]を押します。

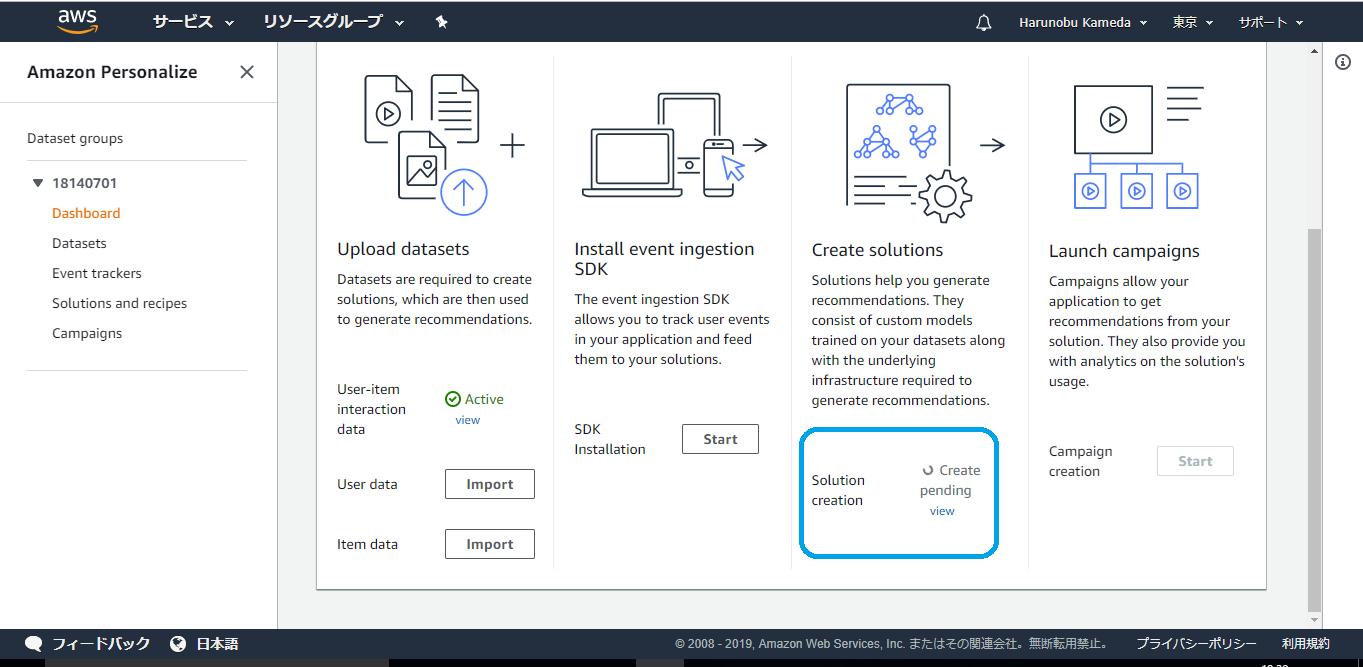




１７．以下の画面で[Finish]を押すと学習が開始されます。



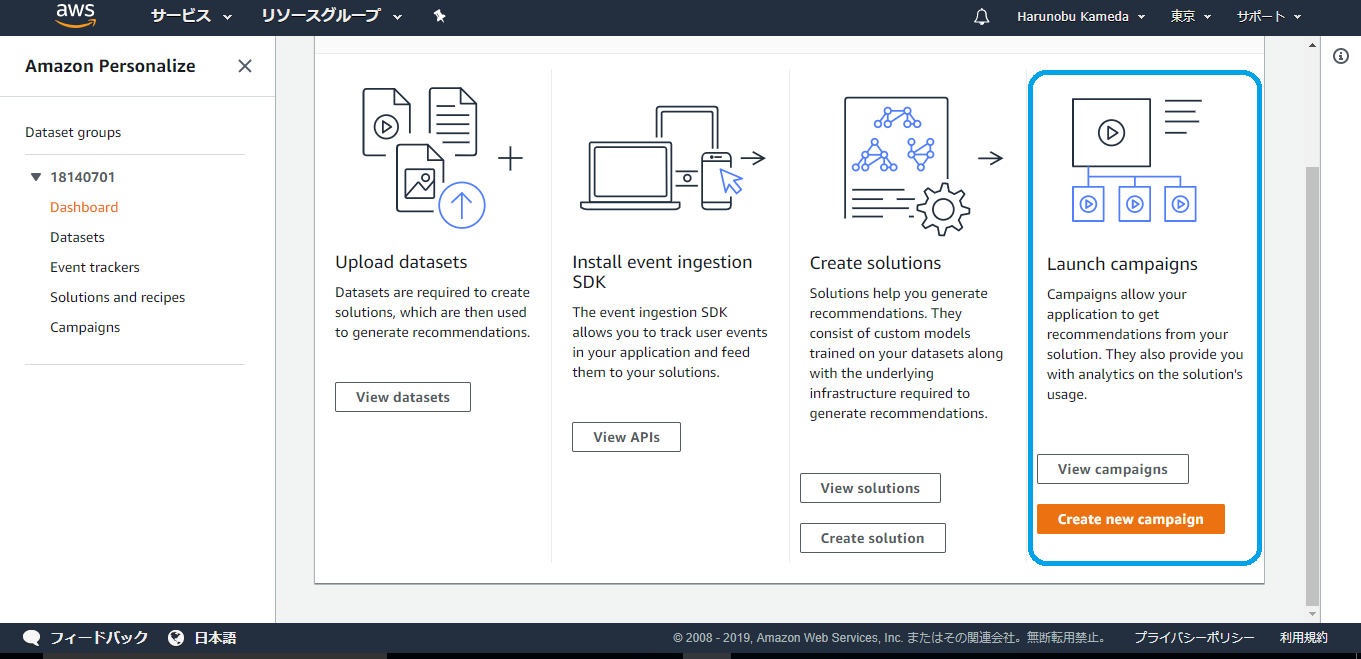
１８．学習は４０分ぐらいかかります。



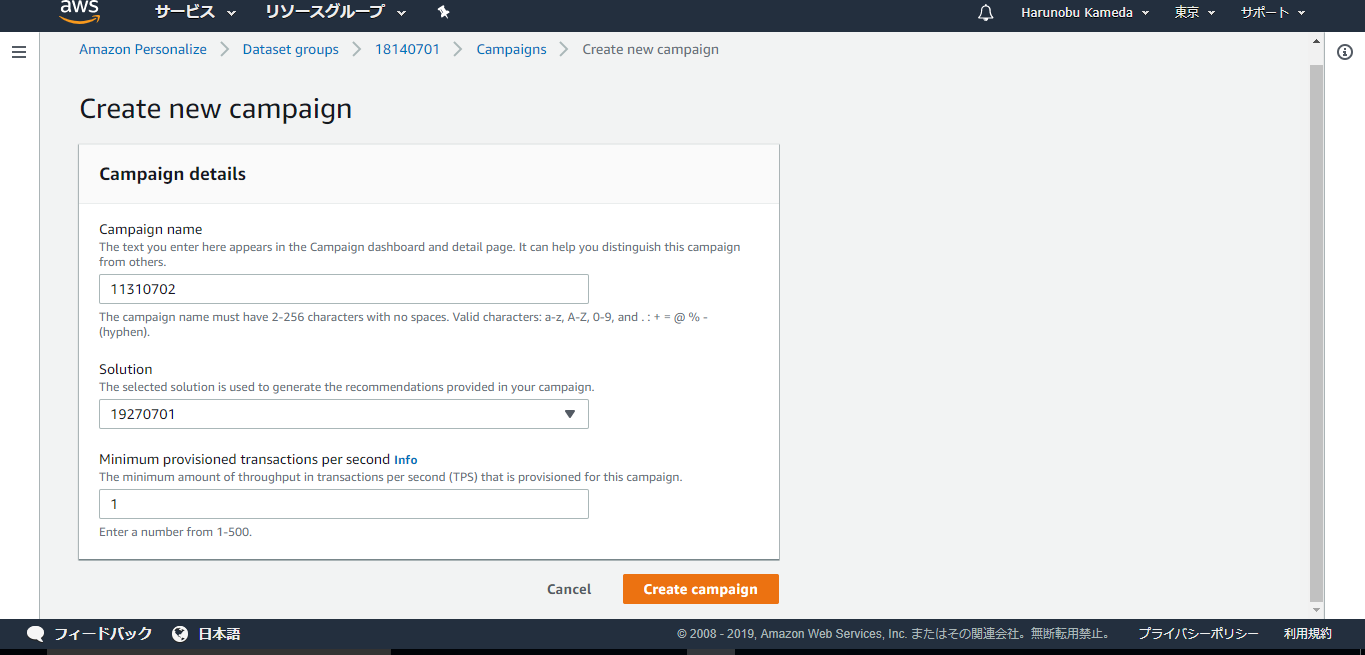
待ち時間の間にForecastのハンズオンを続行します。

AWS CLIが使える方は待ち時間に余裕があれば巻末の手順を参考にAWS CLIの実行環境の準備をしてみてください。

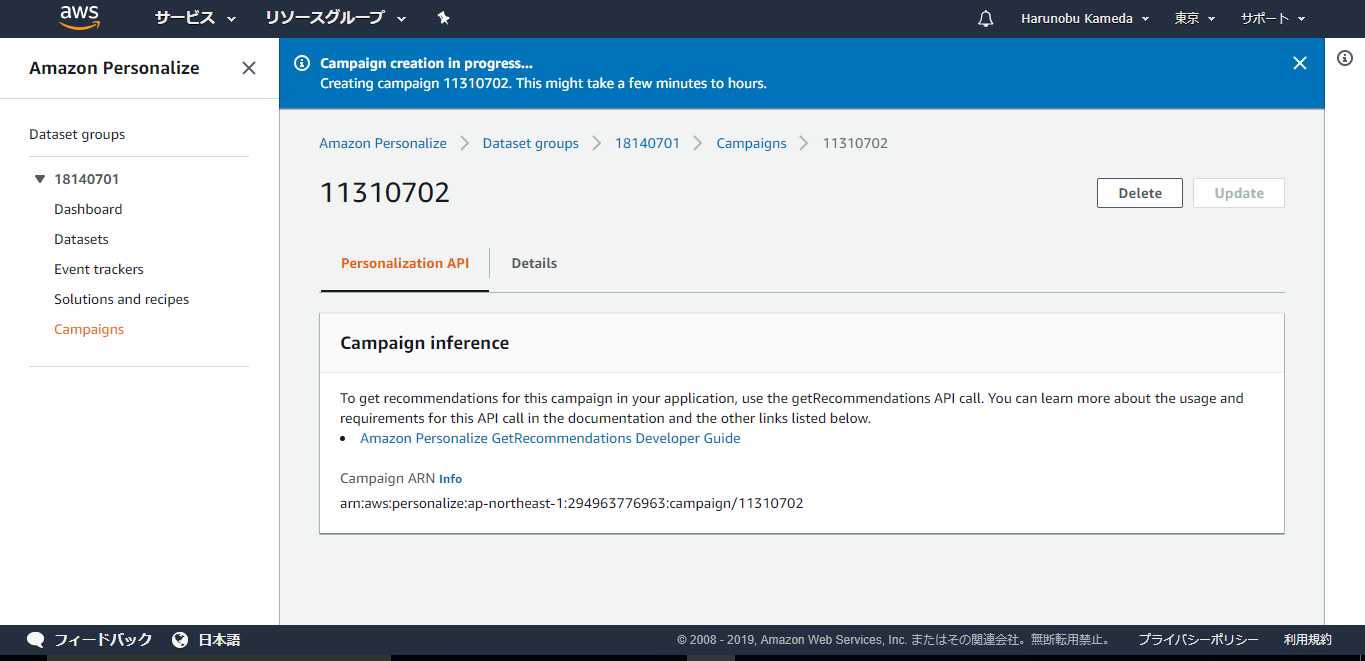
１９．学習が完了すると、[Create New Campaign]のボタンを押せるようになります。Campaignは一般的な機械学習用語でいうと、推論モデルのデプロイになります。



２０．[Campaign Name]に任意の名前を入れます。[Solution]は先ほど作成したSolutionの名前をドロップダウンから選択します。[Minimum provisioned transactions per second]はデフォルト状態のまま[Create Campaign]を押します。



２１．推論環境の構築が終わるまで待ちます。



待ち時間の間にForecastのハンズオンを続行します。

２２．作成が完了すると[Test Campaign Result]が表示されますので、1から601までの任意のユーザーIDを入力し、[Get Recommendations]を押します。ratings.csvをもとに、入力されたユーザーの映画視聴履歴と相関のある別ユーザーの映画視聴履歴からレコメンドが出力されます。

例えば商用環境で使う場合、映画はIDで出力されますので以下のデータをもとに補正を行うことでユーザーに映画情報を出力することができます。

<https://grouplens.org/datasets/movielens/>

<http://files.grouplens.org/datasets/movielens/ml-latest-small.zip>

３．おつかれさまでした！

削除は作成と逆の順番で消す必要があります。

Campaigns→Solutions→Datasets→Dataset groups

S3バケット

その他：時間に余裕があればAWS CLI経由での操作に挑戦してみてください。

[GetRecommendations]を用いて以下のようにレコメンドを出力できます。

aws personalize-runtime get-recommendations \

--campaign-arn arn:aws:personalize:us-west-2:*acct-id*:campaign/MovieRecommendationCampaign \

--user-id 123

以下のようにJSONでレコメンドが出力されます。

{

"itemList": [

{

"itemId": "14"

},

{

"itemId": "15"

},

{

"itemId": "275"

},

{

"itemId": "283"

},

{

"itemId": "273"

},

...

]

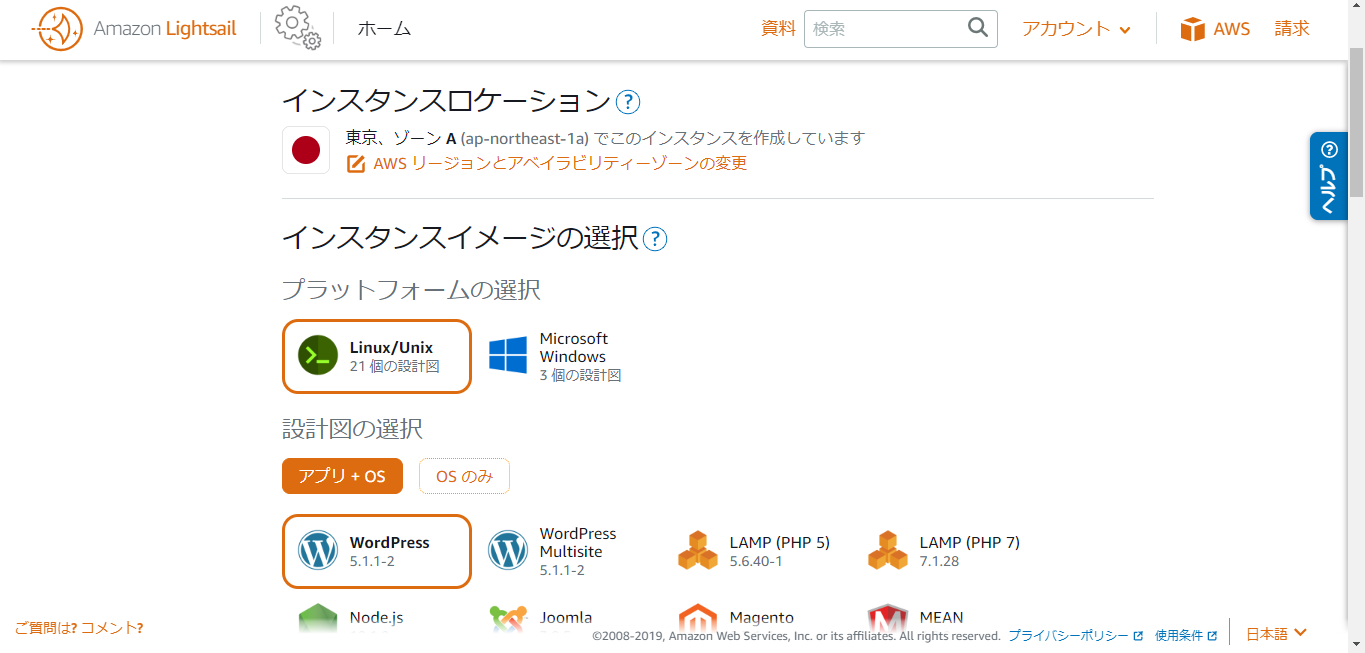
}

一番簡単なAWS CLI環境の構築はAmazon Lightsailです。

1. Amazon Lightsailの管理画面で[インスタンスの作成]を押します



２．[OSのみ]を選択します



1. Amazon Linuxが選択されていることを確認し、残りは全てデフォルト値で[インタンスの作成]を押します。
2. インスタンスが作成中となりますので少しまちます



1. 実行中になったら起動が完了ですので、コンソールボタンを押します。



５．[sudo pip install -U awscli]でCLIをアップグレードします。

６．[aws configure]でデフォルトリージョンを作業中のリージョンに指定します。（東京であれば[ap-northeast-1]）

CLIの実行権限([AmazonPersonalizeFullAccess]と[AmazonPersonalizeFullAccess]を合わせて設定します。)この際設定するIAMロールは[AmazonPersonalizeFullAccess]を保持している必要があります。

[Default output format]はデフォルトのまま[None]を設定します。

7. aws personalize-runtime get-recommendations --campaign-arn arn:<personalize campaignのarn> --user-id 123

のように入力し実行します。

以下のようなエラーがでた場合、IAMロールの設定ミスが発生しています。

An error occurred (AccessDeniedException) when calling the GetRecommendations operation: User: arn:aws:sts::320976636470:assumed-role/AmazonLightsailInstanceRole/i-0fa8803af5e4611ee is not authorized to access this resource

8. 成功した場合以下が表示されます。

