

# Tìm hiểu về dịch vụ Recommendation của AWS

Trần Minh Sang – 19110278

Cao Xuân Thắng – 19119222

## 1. Mô tả

Đây là demo xây dựng hệ thống đề xuất bằng Amazon Personalize. Sử dụng data đề xuất là đề xuất cho giáo dục mà phát triển.

Một số thuật toán được sử dụng trong Personalization:

- User-Personalization recipe

Gồm các thuộc tính sau:

- + Name – aws-user-personalization
- + Recipe Amazon Resource Name (ARN) – arn:aws:personalize:::recipe/aws-user-personalization
- + Algorithm ARN – arn:aws:personalize:::algorithm/aws-user-personalization

- Popularity-Count recipe

Gồm các thuộc tính sau:

- + Name – aws-popularity-count
- + Recipe ARN – arn:aws:personalize:::recipe/aws-popularity-count
- + Algorithm ARN – arn:aws:personalize:::algorithm/aws-popularity-count
- + Feature transformation ARN – arn:aws:personalize:::feature-transformation/sims
- + Recipe type – USER\_PERSONALIZATION

- Personalized-Ranking recipe

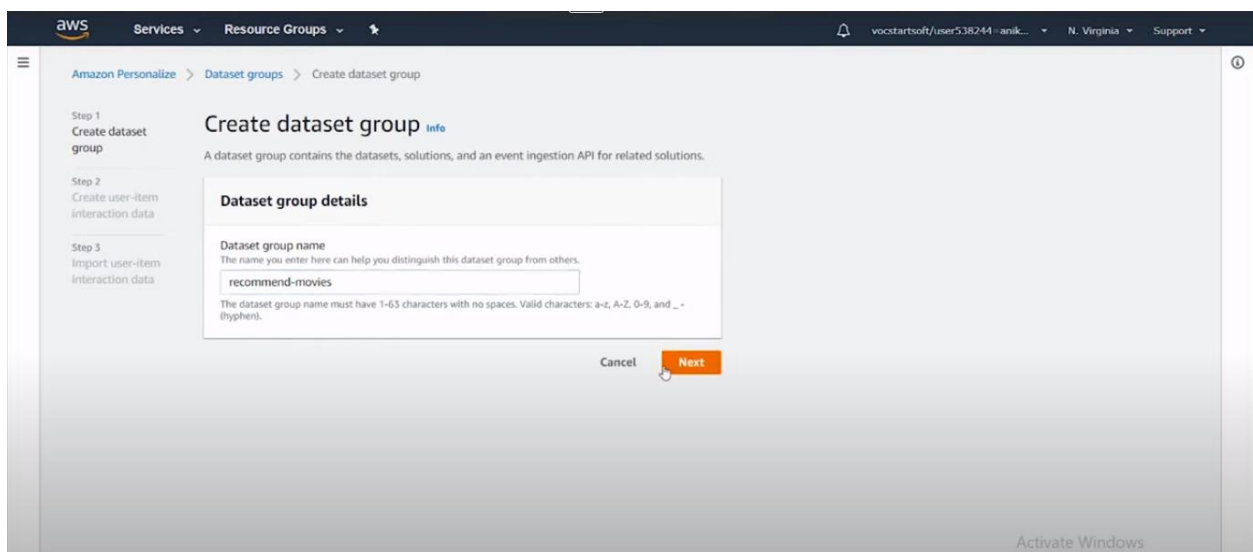
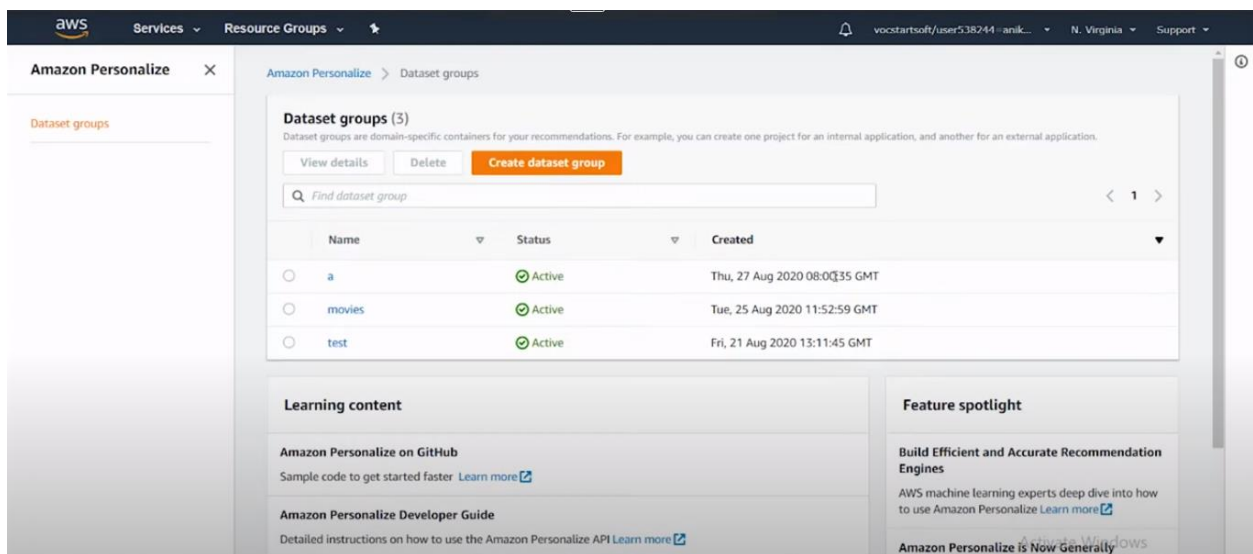
Gồm các thuộc tính sau:

- + Name – aws-personalized-ranking
- + Recipe Amazon Resource Name (ARN) – arn:aws:personalize:::recipe/aws-personalized-ranking

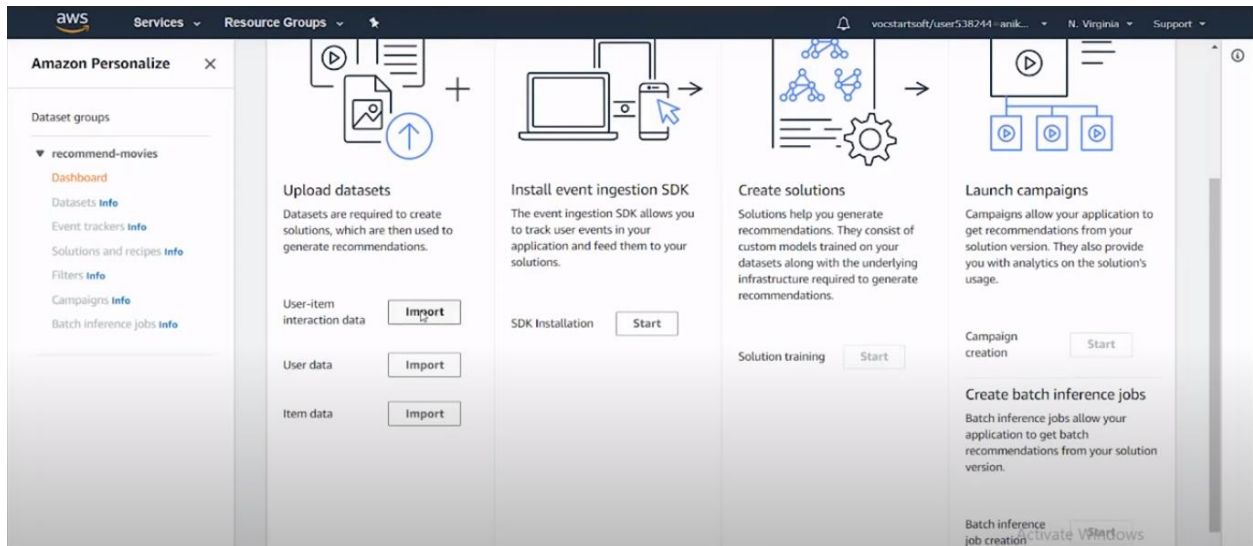
- + Algorithm ARN – `arn:aws:personalize:::algorithm/aws-personalized-ranking`
- + Feature transformation ARN – `arn:aws:personalize:::feature-transformation/JSON-percentile-filtering`
- + Recipe type – `PERSONALIZED_RANKING`

## 2. Cách thực hiện

Đầu tiên chúng ta cần tạo một tài khoản AWS và đăng nhập vào AWS. Sau đó bạn truy cập vào Amazon Personalize của mình -> Bấm vào Create dataset group -> Nhập tên cho Dataset group name -> Bấm Next.



Quay lại trang Amazon Personalize của mình, bấm vào Dataset group vừa tạo -> Trong mục Dashboard -> Tại phần User-item interaction data -> Bấm Import.



Nhập tên cho Dataset name -> Chọn vào Create new schema -> Nhập tên cho New schema name. Bởi vì data có dạng .csv nên bạn phải thay đổi Schema definition với định dạng:

```
{
  "type": "record",
  "name": "Interaction",
  "namespace": "com.amazonaws.personalize.schema",
  "fields":
  [
    {
      "name": "Ten_cot_1",
      "type": "Kieu_du_lieu_cot_1",
    },
    {
      "name": "Ten_cot_2",
      "type": "Kieu_du_lieu_cot_2",
    }
  ]
}
```

“version”: “1.0”

}

The screenshot shows the 'Schema definition' page in the AWS console. The 'Schema selection' section has 'Create new schema' selected. The 'New schema name' field contains 'user-interaction'. The 'Schema definition' section shows a JSON schema for a record type named 'Interactions' with fields 'USER\_ID' and 'ITEM\_ID'. A 'Schema definition' sidebar on the right explains that Amazon Personalize uses this schema to ingest data and lists required fields for each dataset type.

Sau khi hoàn thành ta bấm Next -> Nhập tên cho Dataset import job name.

The screenshot shows the 'Configure user-item interaction dataset import job' page. The 'Dataset import job name' field contains 'user-interaction'. The 'IAM service role' dropdown is set to 'AmazonPersonalize-ExecutionRole-1598017880898'. A blue box indicates that an additional S3 bucket policy is required. The 'Data location info' section shows the S3 path 's3://bucket/path-to-your-data/'. At the bottom, there are 'Cancel', 'Previous', and 'Start Import' buttons.

\*Trước khi nhập đường dẫn đến Data location chúng ta phải tạo Buckets tại trang Amazon S3 của mình.

Ta bấm Create bucket -> Nhập Bucket name, Region -> Bấm Next -> Bấm Next -> Bỏ chọn Block all public access -> Phần Manage system permissions chọn Grant Amazon S3 Log Delivery group with access to this bucket -> Bấm Next -> Bấm Create bucket.

aws

Services

Resource Groups

notifications

voctartsoft/user538244-anik...

Global

Support

Amazon S3

Buckets

Batch operations

Access analyzer for S3

Block public access (account settings)

Feature spotlight 2

Create bucket

1 Name and region

2 Configure options

3 Set permissions

4 Review

Name and region

Bucket name recommend-movies-hackerrhine

Region US East (N. Virginia)

Copy settings from an existing bucket Select bucket (optional)4 Buckets

Create

Cancel

Next

aws

Services

Resource Groups

notifications

voctartsoft/user538244-anik...

Global

Support

Amazon S3

Buckets

Batch operations

Access analyzer for S3

Block public access (account settings)

Feature spotlight 2

Create bucket

1 Name and region

2 Configure options

3 Set permissions

4 Review

Properties

Versioning Keep all versions of an object in the same bucket. Learn more

Server access logging Log requests for access to your bucket. Learn more

Tags You can use tags to track project costs. Learn more

Object-level logging Record object-level API activity using AWS CloudTrail for an additional cost. See CloudTrail pricing or learn more

Default encryption Automatically encrypt objects when they are stored in S3. Learn more

Previous

Next

aws

Services

Resource Groups

notifications

voctartsoft/user538244-anik...

Global

Support

Amazon S3

Buckets

Batch operations

Access analyzer for S3

Block public access (account settings)

Feature spotlight 2

Create bucket

1 Name and region

2 Configure options

3 Set permissions

4 Review

Note: You can grant access to specific users after you create the bucket.

Block public access (bucket settings)

Public access is granted to buckets and objects through access control lists (ACLs), bucket policies, access point policies, or all. In order to ensure that public access to all your S3 buckets and objects is blocked, turn on Block all public access. These settings apply only to this bucket and its access points. AWS recommends that you turn on Block all public access, but before applying any of these settings, ensure that your applications will work correctly without public access. If you require some level of public access to your buckets or objects within, you can customize the individual settings below to suit your specific storage use cases. Learn more

Disabling Block all public access may result in this bucket and the objects within becoming public

AWS recommends that you block all public access to your bucket, unless public access is required for specific and verified use cases such as static website hosting.

acknowledge that the current settings may result in this bucket and the objects within becoming public

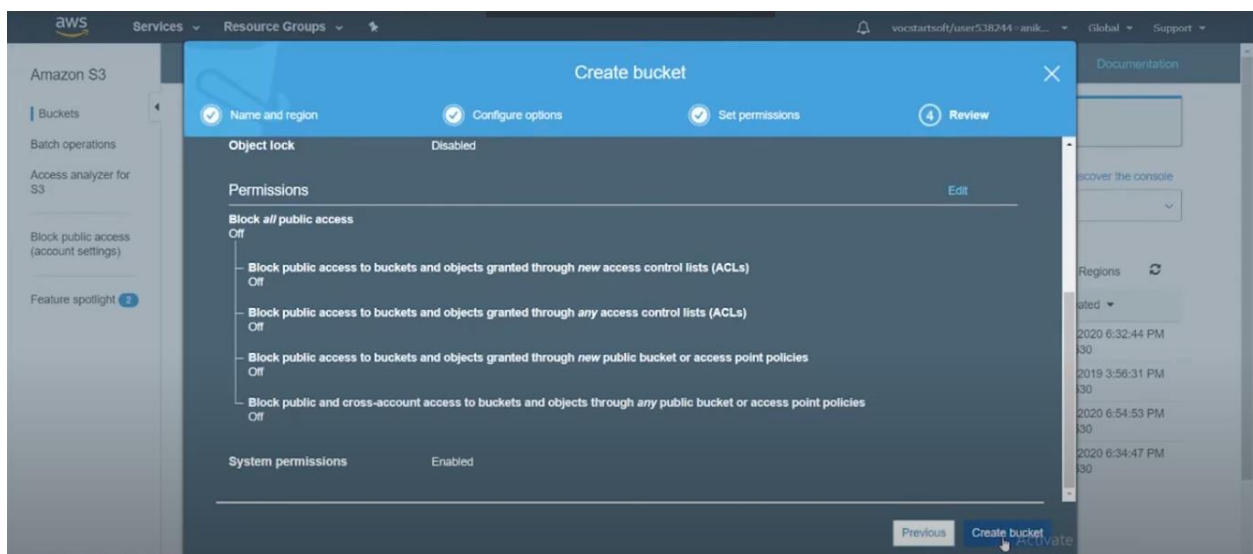
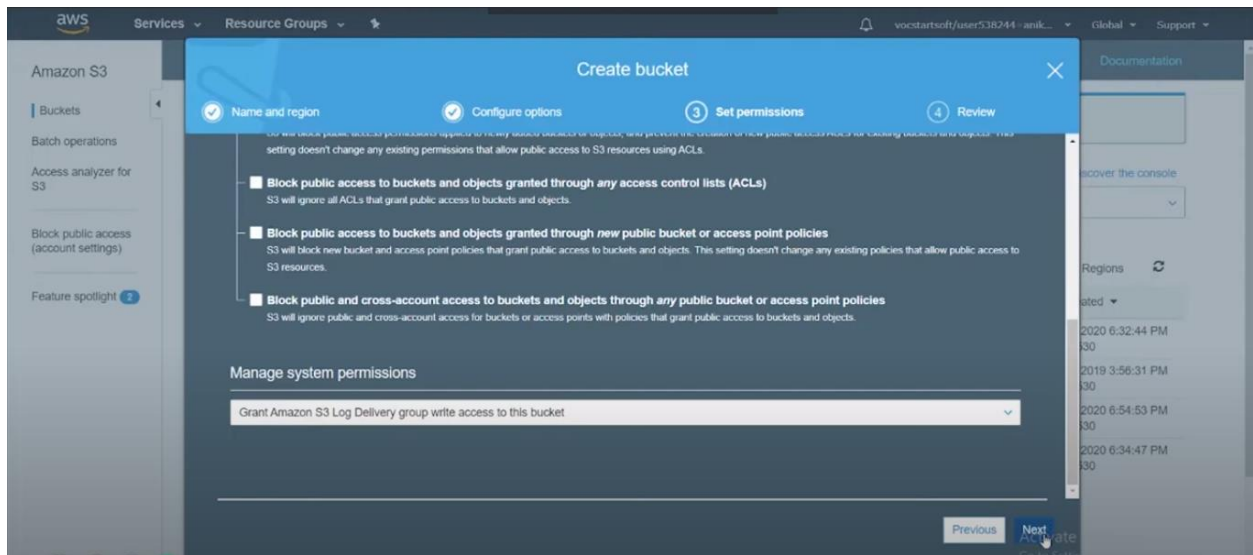
Block all public access

Turning this setting on is the same as turning on all four settings below. Each of the following settings are independent of one another.

Block public access to buckets and objects granted through new access control lists (ACLs)

Previous

Next



Ta đã có Bucket. Chọn Bucket vừa tạo -> Tải dữ liệu phù hợp lên từ máy (.csv) -> Trong mục Permissions -> Ta vào Bucket Policy nhập vào phần Bucket policy editor đoạn code:

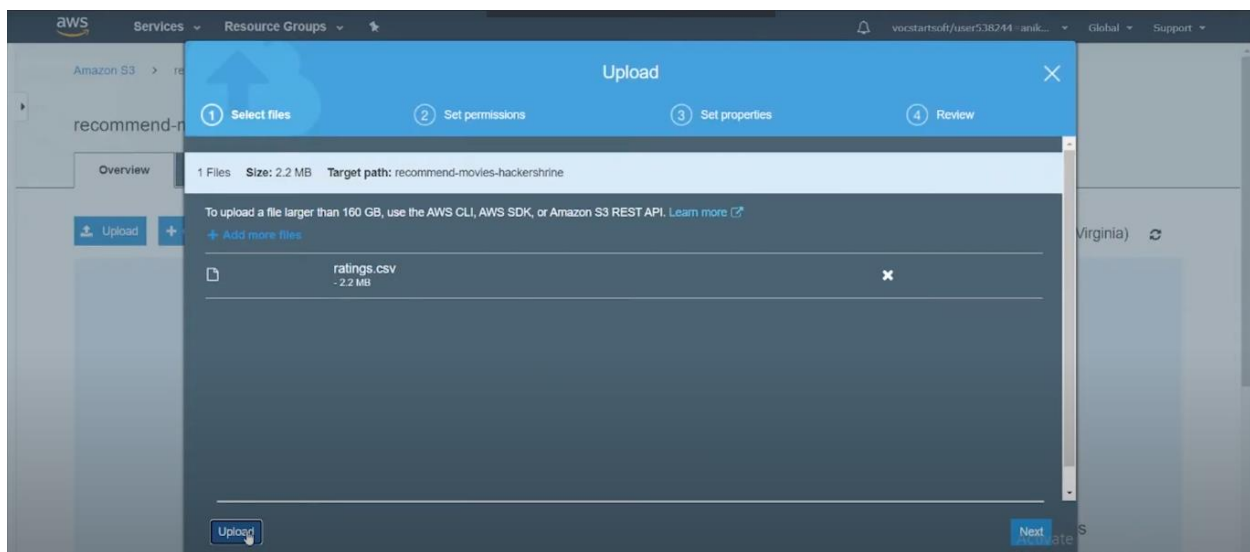
Bucket Policy json:

```
{
  "Version": "2012-10-17",
  "Id": "PersonalizeS3BucketAccessPolicy",
  "Statement": [
    {
      "Sid": "PersonalizeS3BucketAccessPolicy",
```

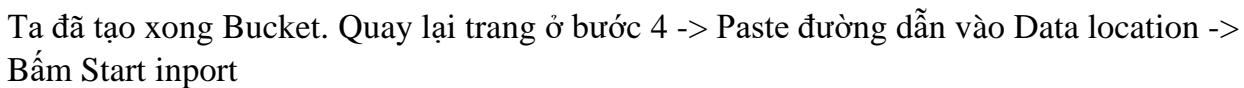
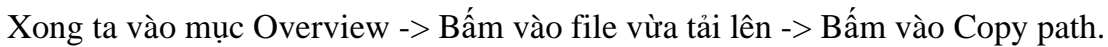
```

    "Effect": "Allow",
    "Action":
    [
        "s3:GetObject",
        "s3:ListBucket"
    ],
    "Resource":
    [
        "arn:aws:s3:::bucket-name",
        "arn:aws:s3:::bucket-name/*"
    ]
}
]
}
}

```









**Configure user-item interaction dataset import job**

**Dataset import job details**

**Dataset import job name**  
The name you enter here can help you distinguish this dataset import job from others.  
user-interaction  
The dataset import job name must have 1-63 characters with no spaces. Valid characters: a-z, A-Z, 0-9, and \_ - (hyphen).

**IAM service role**  
Amazon Personalize requires permissions to access your S3 bucket. Choose an existing role with access or create a role in the IAM console with the [AmazonPersonalizeFullAccess](#) IAM policy attached.  
AmazonPersonalize-ExecutionRole-1598017880898

**Additional S3 bucket policy required**  
In addition to the IAM service role defined above, Amazon Personalize also requires you to add a bucket policy to the S3 bucket containing your data files so that it can process them. Follow the instructions [described here](#) to add the required bucket policy to your S3 bucket.

**Data location info**  
Choose the S3 location of your data.  
s3://recommend-movies-hackershrine/ratings.csv  
Your file needs to be in a CSV format and reflect the schema.

Cancel Previous **Start import**

Quay lại bước 3-4-6-7-8 (không cần tạo thêm Bucket nữa) với hai phần còn lại là User data và Item data. Lưu ý ở bước 4 ta phải thiết kế Schema definition phù hợp với file mình load lên từ máy.

**Amazon Personalize**

Amazon Personalize > Dataset groups > recommend-movies > Datasets

**Datasets (3)**  
Datasets are required to create solutions, which are then used to generate recommendations.

	Dataset name	Dataset type	Status	Created
<input type="radio"/>	item-data	Item	Active	Thu, 27 Aug 2020 10:06:07 GMT
<input type="radio"/>	user-data	User	Active	Thu, 27 Aug 2020 10:04:53 GMT
<input type="radio"/>	user-interaction	User-item interaction	Active	Thu, 27 Aug 2020 09:57:14 GMT

View details Create dataset import job Create dataset

Đây là kết quả

Tại mục Solutions and recipes -> Bấm vào Create solution -> Nhập Solution name -> Chọn thuật toán bạn muốn tại Recipe -> Thay đổi cái thiết lập theo mong muốn tại Advanced configuration -> Bấm Next -> Bấm Finish.

Services

Resource Groups

aws

Services

Resource Groups

notifications

vocstartsoft/user538244 - anik...

N. Virginia

Support

Amazon Personalize

Dataset groups

movies

Solutions and recipes

Solution and recipes

Analyze your solutions and compare their performances to each other. You can also take a look at available recipes.

Solutions

Recipes

Solutions (4)

Analyze your solutions and compare their performance to each other. Solutions are created using recipes.

Find solution

< 1 >

	Name	Status	Recipe used	Created
<input type="radio"/>	sol-4	Active	aws-hrnn-metadata	Wed, 26 Aug 2020 17:39:37 GMT
<input type="radio"/>	sol-3	Active	aws-popularity-count	Wed, 26 Aug 2020 12:06:16 GMT
<input type="radio"/>	sol-2	Active	aws-personalized-ranking	Wed, 26 Aug 2020 10:54:57 GMT
<input type="radio"/>	solution1	Active	aws-sims	Wed, 26 Aug 2020 10:07:55 GMT

Activate Windows

Services

Resource Groups

aws

Services

Resource Groups

notifications

vocstartsoft/user538244 - anik...

N. Virginia

Support

Amazon Personalize

Dataset groups

movies

Solutions and recipes

Featurization transformation hyperparameters - optional info

User history length percentile

Choose minimum and maximum length (relative) of user history to use for modeling.

Min

0.00

Max

0.99

Algorithm hyperparameters - optional info

Bpitt

Choose how many past events to consider when modeling for long term effects of these events on rewards. Larger value signify longer sequences of past events are considered to predict rewards.

32

Choose a value between 2 and 32.

Hidden dimension

Choose the number of hidden variables used in the model. Specify a greater number of hidden dimensions when your interactions dataset includes more complicated patterns.

149

Choose a value between 32 and 256.

Recency mask

Choose whether the model should consider the latest popularity trends in your dataset. To train a model that places more weight on recent events, set recency\_mask to true. To train a model that equally weighs all past interactions, set recency\_mask to false.

☒ true

☐ false

Cancel

Next

Activate Windows

Services

Resource Groups

aws

Services

Resource Groups

notifications

vocstartsoft/user538244 - anik...

N. Virginia

Support

Amazon Personalize

Dataset groups

movies

Solutions and recipes

test-solution was successfully created.

Amazon Personalize

Dataset groups

movies

Solutions

Create solution

Create solution version

Now that you have created a solution, click Finish button to create a solution version (train your solution).

Solution overview

Solution name	Perform HPO	Event type
test-solution	false	-
Solution ARN	Perform AutoML	Latest solution version ARN
arn:aws:personalize:us-east-1:766953567828:solution/test-solution	false	-
Status	Recipe	Created
Active	aws-hrnn	Thu, 27 Aug 2020 10:08:33 GMT

Solution config

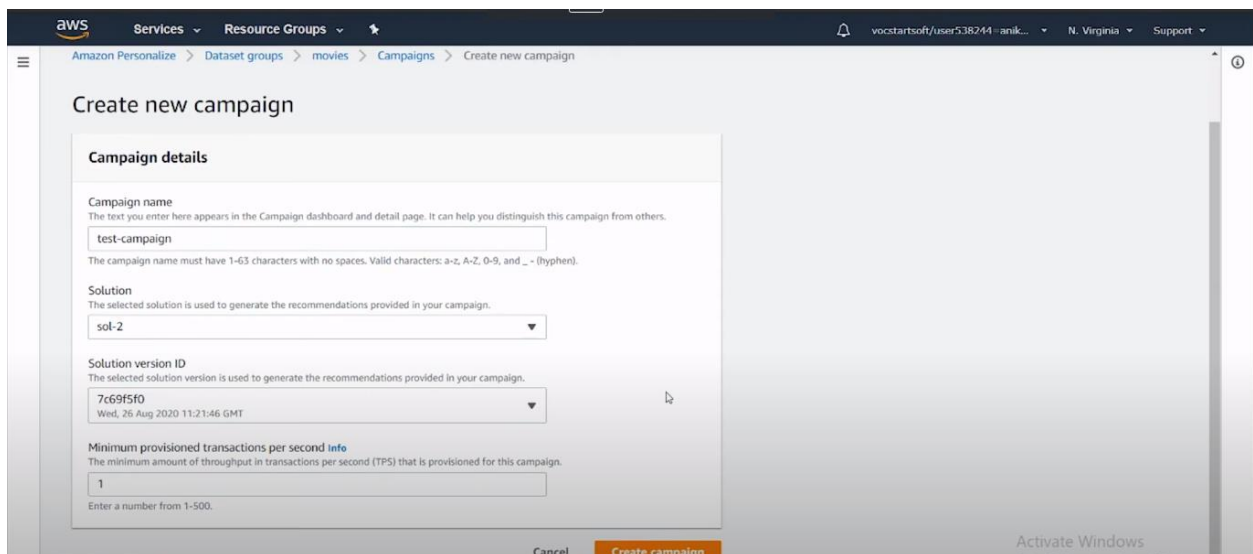
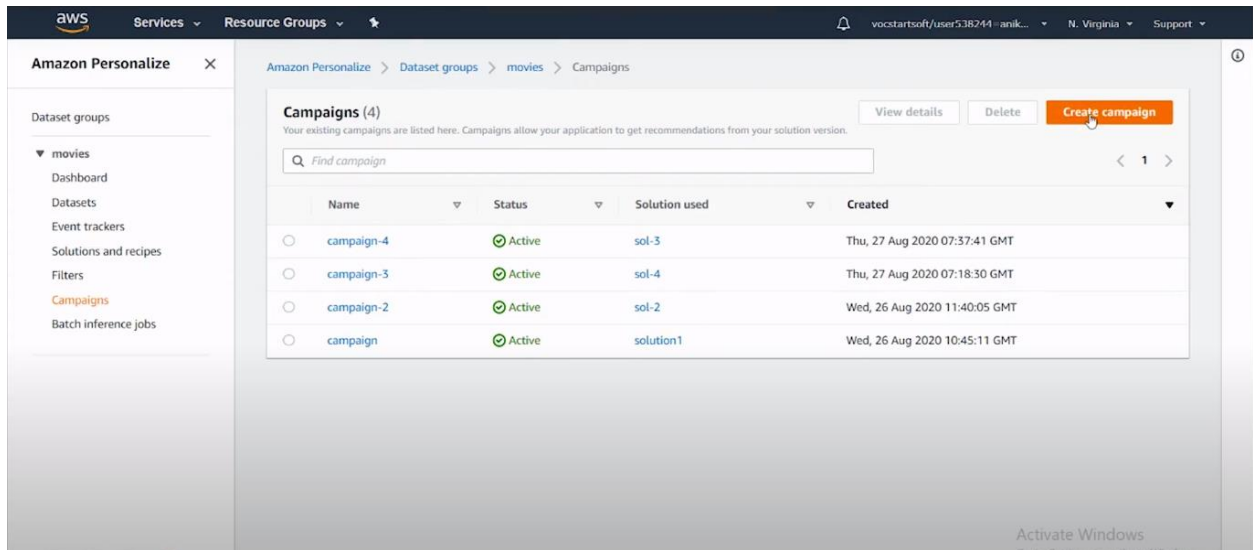
Cancel

Previous

Finish

Activate Windows

Tại mục Campaigns -> Bấm vào Create campaign -> Nhập Campaign name, Solution, Solution ID -> Bấm Create campaign.



VD:

- “Recommend\_movie” đề xuất bộ phim phù hợp với bản thân mình nhất dựa theo các yếu tố như: thể loại, thời lượng, điểm đánh giá, số lượng người coi.

- Thết kê Dataset group như sau:

User-item interaction data: ratings.csv.

User data: tags.csv.

Item data: movies.csv, links.csv.

