

# Tìm hiểu về dịch vụ Recommendation của AWS

Trần Minh Sang – 19110278

Cao Xuân Thắng – 19119222

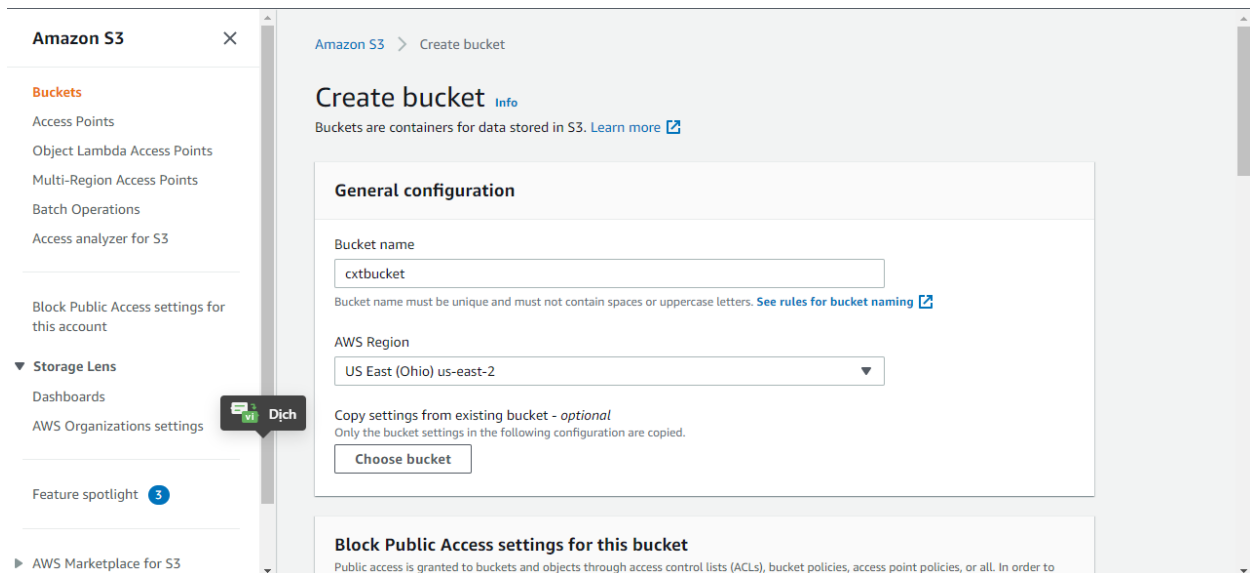
## 1. Mô tả

Amazon Personalize là một dịch vụ máy học giúp các nhà phát triển dễ dàng thêm các đề xuất dành riêng cho khách hàng sử dụng ứng dụng của họ. Đây là demo xây dựng hệ thống đề xuất bằng Amazon Personalize.

## 2. Cách thực hiện

\*Đầu tiên chúng ta phải tạo Buckets tại trang Amazon S3 của mình.

Ta bấm Create bucket -> Nhập Bucket name, Region -> 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.



Amazon S3

Buckets

Access Points

Object Lambda Access Points

Multi-Region Access Points

Batch Operations

Access analyzer for S3

Block Public Access settings for this account

Storage Lens

Dashboards

AWS Organizations settings

Feature spotlight 3

AWS Marketplace for S3

Amazon S3

Buckets

Access Points

Object Lambda Access Points

Multi-Region Access Points

Batch Operations

Access analyzer for S3

Block Public Access settings for this account

Storage Lens

Dashboards

AWS Organizations settings

Feature spotlight 3

AWS Marketplace for S3

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 this bucket or objects within, you can customize the individual settings below to suit your specific storage use cases. [Learn more](#)

☐ 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)

S3 will block public access permissions applied to newly added buckets or objects, and prevent the creation of new public access ACLs for existing buckets and objects. This setting doesn't change any existing permissions that allow public access to S3 resources using ACLs.

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


S3 will ignore all ACLs that grant public access to buckets and objects.

☐ Block public access to buckets and objects granted through new public bucket or access point policies

S3 will block new bucket and access point policies that grant public access to buckets and objects. This setting doesn't change any existing policies that allow public access to S3 resources.

☐ Block public and cross-account access to buckets and objects through any public bucket or access point policies

S3 will ignore public and cross-account access for buckets or access points with policies that grant public access to buckets and objects.

 Turning off block all public access might result in this bucket and the objects within becoming public

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

☒ I acknowledge that the current settings might result in this bucket and the objects within becoming public.

No tags associated with this bucket.

Add tag

Default encryption


Automatically encrypt new objects stored in this bucket. [Learn more](#)

Server-side encryption

☒ Disable

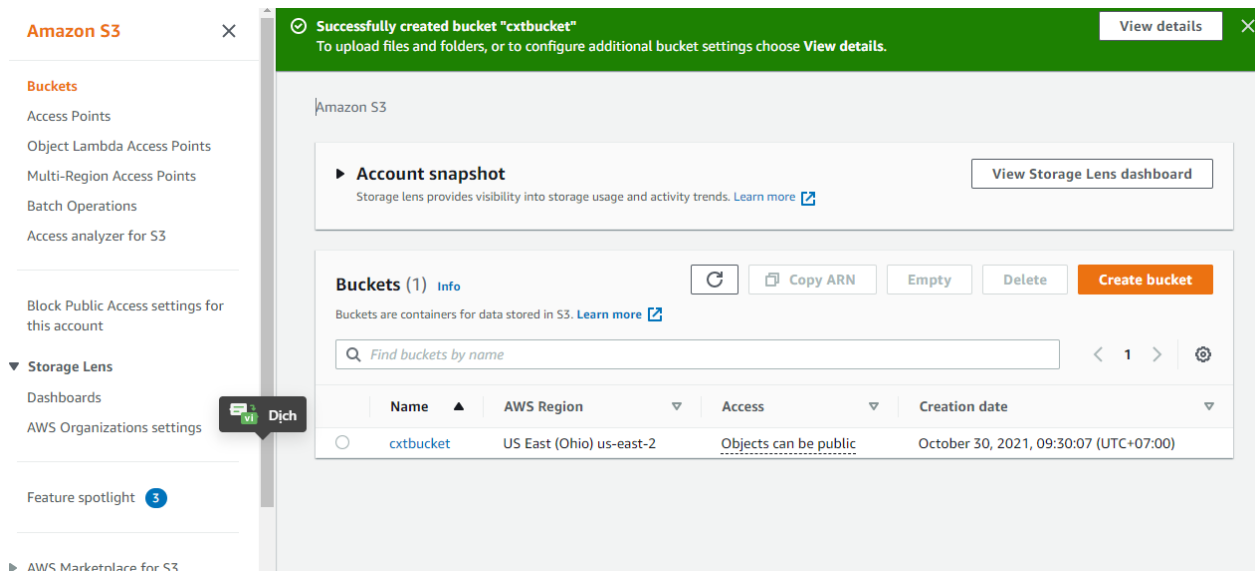
☐ Enable

Advanced settings

 After creating the bucket you can upload files and folders to the bucket, and configure additional bucket settings.

Cancel

Create bucket



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",
      "Principal": "*",
      "Effect": "Allow",
      "Action": [
        "s3:GetObject",
        "s3:ListBucket"
      ],
      "Resource": [
```

```

"arn:aws:s3:::bucket-name",
"arn:aws:s3:::bucket-name/*"
]
}
]
}

```

Amazon S3

Buckets

Access Points
Object Lambda Access Points
Multi-Region Access Points
Batch Operations
Access analyzer for S3

Block Public Access settings for this account

Storage Lens

Dashboards
AWS Organizations settings

Feature spotlight 3

AWS Marketplace for S3

Files and folders (4 Total, 3.1 MB)

Remove Add files Add folder

Find by name

	Name	Folder	Type	Size
<input type="checkbox"/>	links.csv	-	application/vnd.ms-excel	193.3 KB
<input type="checkbox"/>	movies.csv	-	application/vnd.ms-excel	482.8 KB
<input type="checkbox"/>	ratings.csv	-	application/vnd.ms-excel	2.4 MB
<input type="checkbox"/>	tags.csv	-	application/vnd.ms-excel	115.9 KB

Destination

Destination

s3://cxtbucket

Destination details

Bucket settings that impact new objects stored in the specified destination.

Amazon S3

Buckets

Access Points
Object Lambda Access Points
Multi-Region Access Points
Batch Operations
Access analyzer for S3

Block Public Access settings for this account

Storage Lens

Dashboards
AWS Organizations settings

Feature spotlight 3

AWS Marketplace for S3

Successfully edited bucket policy.

Bucket policy

The bucket policy, written in JSON, provides access to the objects stored in the bucket. Bucket policies don't apply to objects owned by other accounts. [Learn more](#)

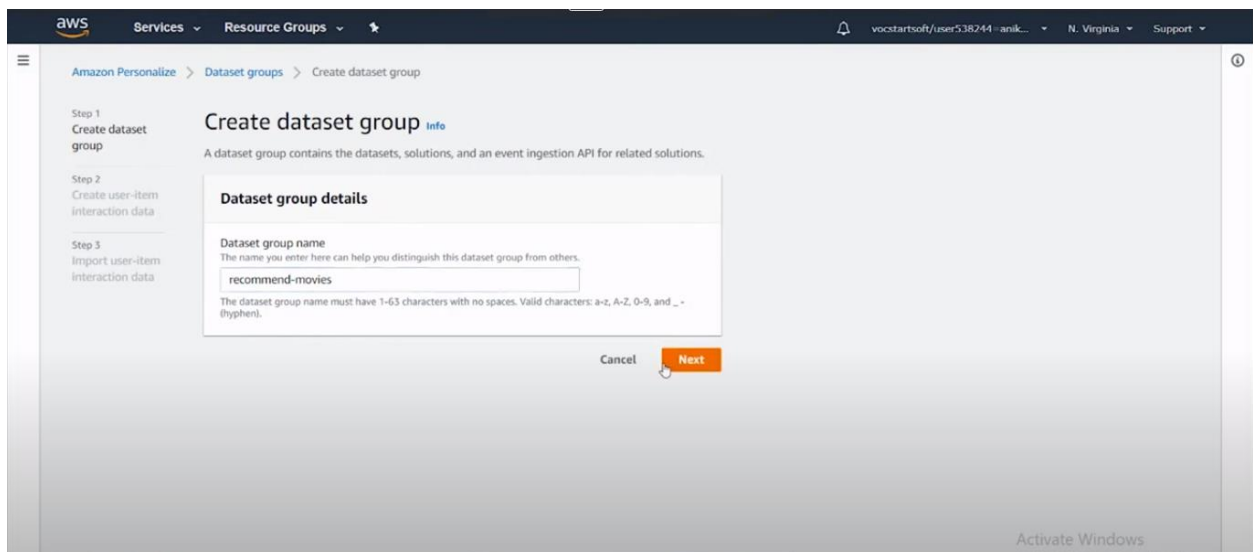
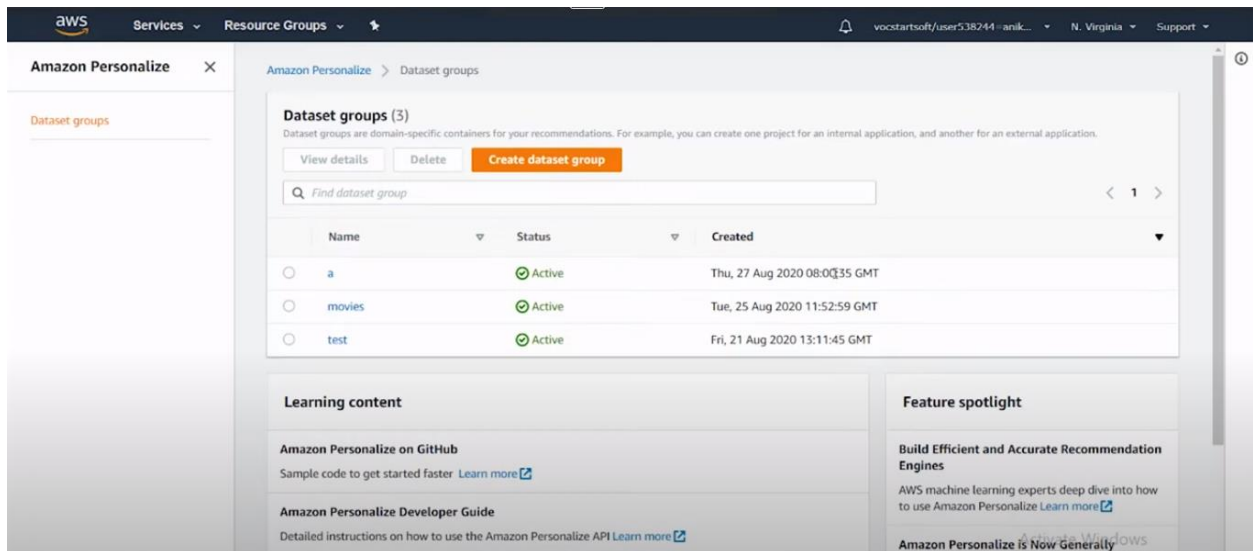
Edit Delete

```
{
  "Version": "2012-10-17",
  "Id": "PersonalizeS3BucketAccessPolicy",
  "Statement": [
    {
      "Sid": "PersonalizeS3BucketAccessPolicy",
      "Effect": "Allow",
      "Principal": "*",
      "Action": [
        "s3:GetObject",
        "s3:ListBucket"
      ],
      "Resource": [
        "arn:aws:s3:::cxtbucket",
        "arn:aws:s3:::cxtbucket/*"
      ]
    }
  ]
}
```

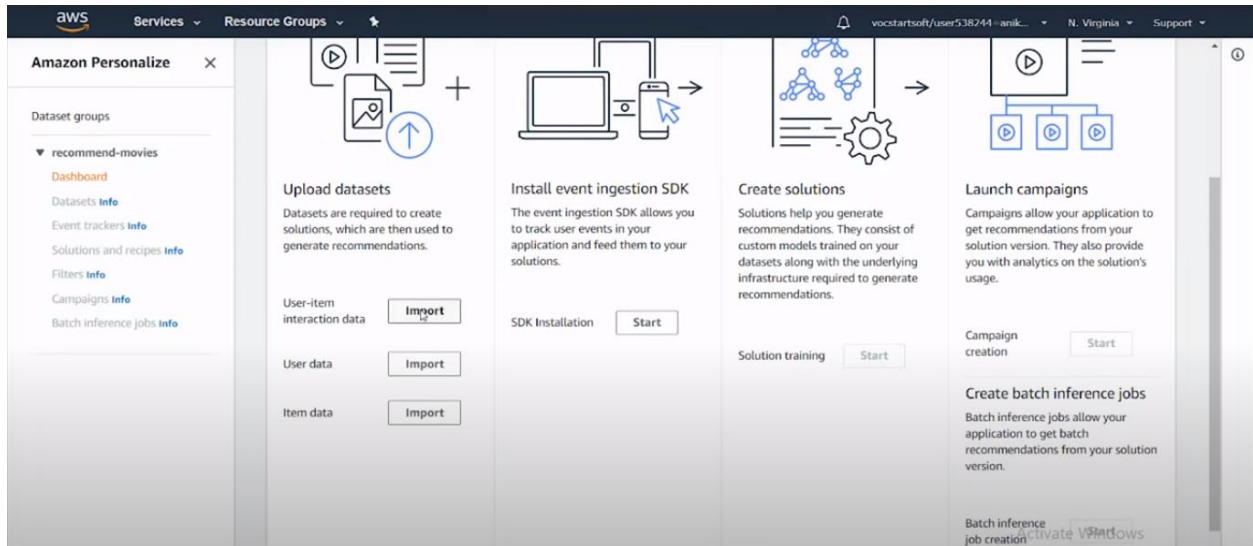
Copy

Ta đã tạo xong Bucket.

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 details' 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 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.

**Schema details**

**Schema selection** Info

☐ Use existing schema  
Choose an existing schema that matches your dataset.

☒ Create new schema  
Create a new schema to match your dataset.

**New schema name**  
The name you enter here appears in the Schema dashboard. It can help you distinguish this schema from others.

user-interaction

The schema name must have 1-63 characters with no spaces. Valid characters: a-z, A-Z, 0-9, and \_ - (hyphen).

**Schema definition** Info  
Ensure your dataset's schema matches the following schema.

```

1 {
2   "type": "record",
3   "name": "Interactions",
4   "namespace": "com.amazonaws.personalize.schema",
5   "fields": [
6     {
7       "name": "USER_ID",
8       "type": "string"
9     },
10    {
11      "name": "ITEM_ID",
12      "type": "string"
13    }
14  ]
15 }
```

**Schema definition**

Amazon Personalize will use this schema to ingest your dataset. Amazon Personalize has some required fields that vary for each dataset type.

Amazon Personalize requires that your schema conform to Avro specifications. The data files that you point to need to be in a CSV format.

Each dataset type has different fields that it requires be present in your dataset. Define **all the fields present in your data files** to ensure Amazon Personalize can process your data. An example schema for a **user-item interaction** dataset could be:

```

{
  "type": "record",
  "name": "Interactions",
  "namespace": "com.amazonaws.pe
  "fields": [

```

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' field contains 's3://bucket/path-to-your-data/'. At the bottom, there are 'Cancel', 'Previous', and 'Start import' buttons.

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

**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.

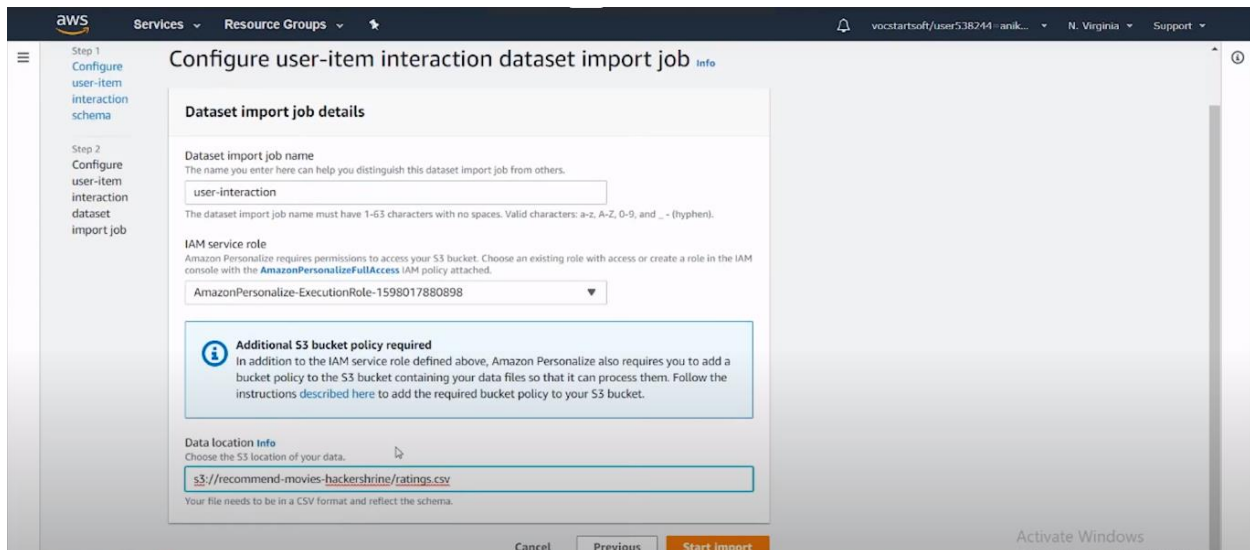
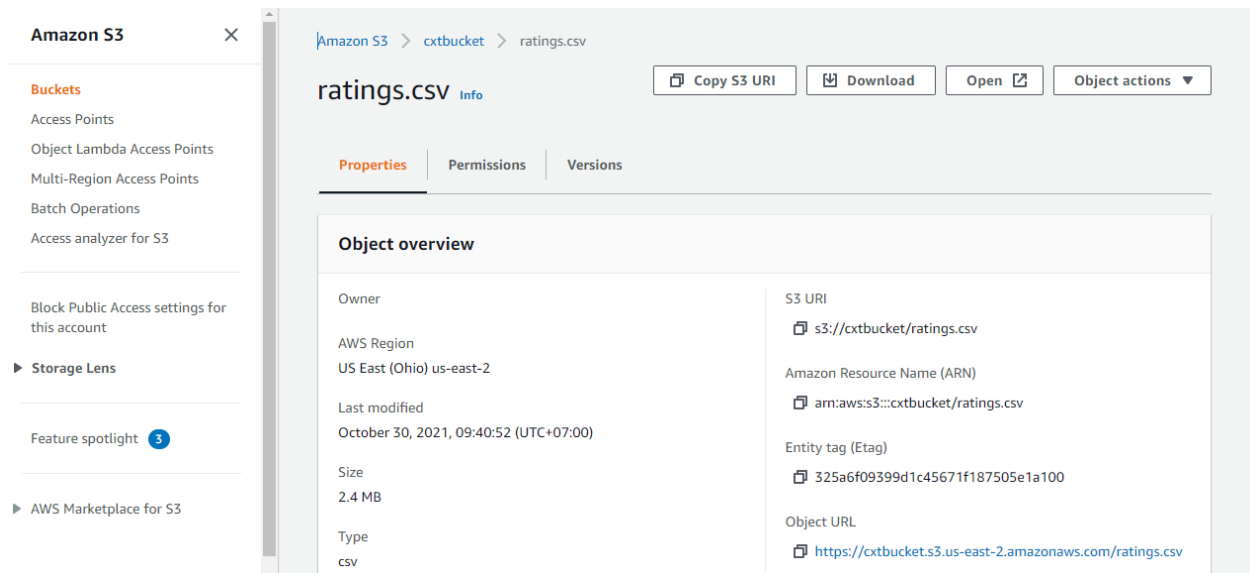
**Data location** Info  
Choose the S3 location of your data.

s3://bucket/path-to-your-data/

Your file needs to be in a CSV format and reflect the schema.

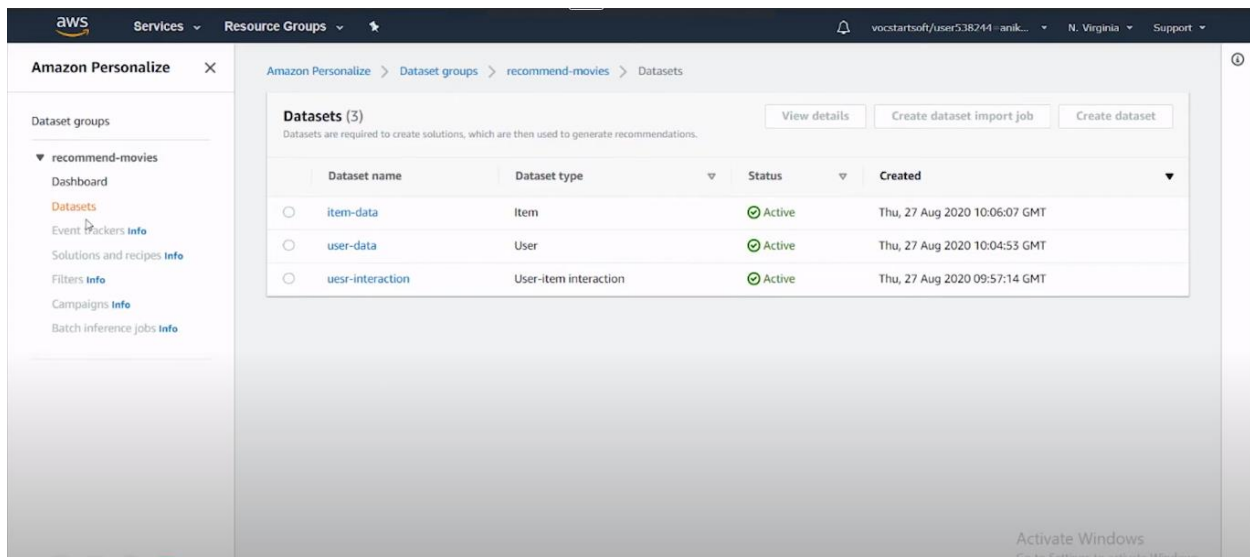
Cancel Previous **Start import**

Copy URL S3 -> Paste đường dẫn vào Data location theo thiết kế dataset.



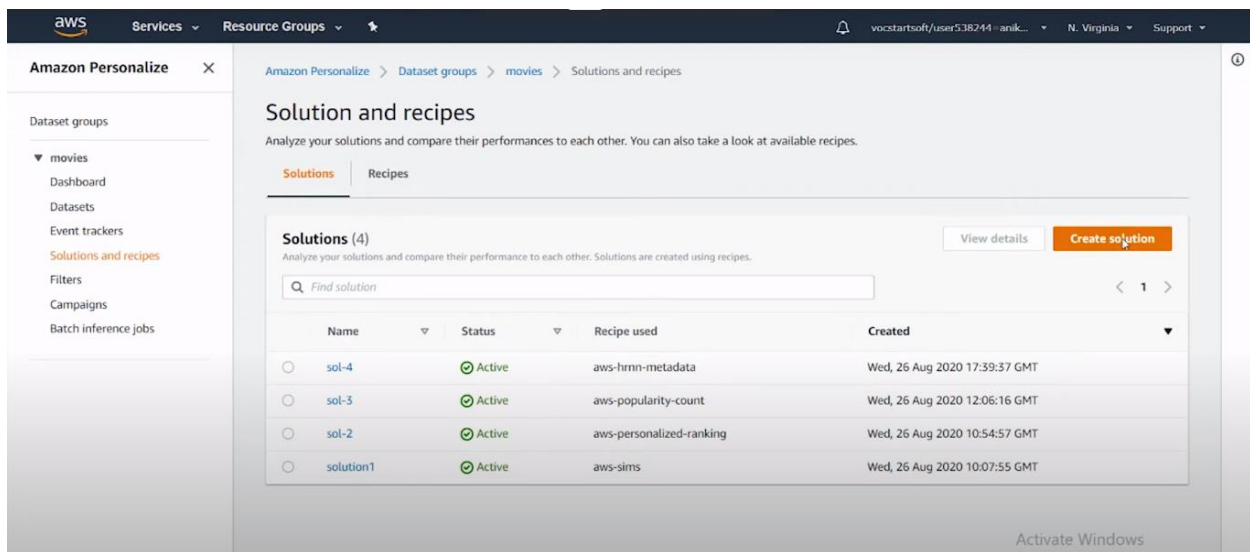
Quay lại bước upload datasets với hai phần còn lại là User data và Item data. Lưu ý ta phải thiết kế Schema definition phù hợp với file mình load lên từ máy.





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 cài thiết lập theo mong muốn tại Advanced configuration -> Bấm Next -> Bấm Finish.



Choose whether you would like Amazon Personalize to find optimal hyperparameters.

▼ **Featurization transformation hyperparameters - optional info**

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

Min	Max
0.00	0.99

▼ **Algorithm hyperparameters - optional info**

**Bpdt**  
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

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

► Solution config

Cancel Previous Finish

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

aws

Services

Resource Groups

vocstartsoft/user538244 - anik...N. VirginiaSupport

Amazon Personalize

Dataset groups

movies

Dashboard

Datasets

Event trackers

Solutions and recipes

Filters

Campaigns

Batch inference jobs

Amazon PersonalizeDataset groupsmoviesCampaigns

Campaigns (4)

Your existing campaigns are listed here. Campaigns allow your application to get recommendations from your solution version.

View detailsDeleteCreate campaign

Find campaign

	Name	Status	Solution used	Created
<input type="radio"/>	campaign-4	Active	sol-3	Thu, 27 Aug 2020 07:37:41 GMT
<input type="radio"/>	campaign-3	Active	sol-4	Thu, 27 Aug 2020 07:18:30 GMT
<input type="radio"/>	campaign-2	Active	sol-2	Wed, 26 Aug 2020 11:40:05 GMT
<input type="radio"/>	campaign	Active	solution1	Wed, 26 Aug 2020 10:45:11 GMT

Activate Windows

aws

Services

Resource Groups

vocstartsoft/user538244 - anik...N. VirginiaSupport

Amazon PersonalizeDataset groupsmoviesCampaignsCreate new campaign

Create new campaign

Campaign details

Campaign name

The text you enter here appears in the Campaign dashboard and detail page. It can help you distinguish this campaign from others.

test-campaign

The campaign name must have 1-63 characters with no spaces. Valid characters: a-z, A-Z, 0-9, and \_ - (hyphen).

Solution

The selected solution is used to generate the recommendations provided in your campaign.

sol-2

Solution version ID

The selected solution version is used to generate the recommendations provided in your campaign.

7c69f5f0  
Wed, 26 Aug 2020 11:21:46 GMT

Minimum provisioned transactions per second [info](#)

The minimum amount of throughput in transactions per second (TPS) that is provisioned for this campaign.

1

Enter a number from 1-500.

CancelCreate campaign

Activate Windows