*TP.HCM - Thứ năm 17/11/2016*

**TRƯỜNG ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**

**ĐẠI HỌC KHOA HỌC TỰ NHIÊN**

**-o0o-**

Khoa Công Nghệ Thông Tin

Bộ môn Hệ Thống Thông Tin

**BÁO CÁO**

**TÌM HIỂU AMAZONS3**



****

Lớp **PTTK UDHTTT HD** **- TH2014/01**

Giáo viên hướng dẫn: **Phạm Minh Tú**

**Mục lục**

[I. Dịch vụ AmazonS3 là gì? Ưu điểm để chọn AmazonS3? So sanh với các dịch vụ khác? Nhược điểm? 3](#_Toc499503551)

[II. Cơ bản về hệ thống. 5](#_Toc499503552)

[III. Cách thức hoạt động. 5](#_Toc499503553)

[IV. Cài đặt và thao tác cơ bản. 9](#_Toc499503554)

[V. Các tính năng khác. 18](#_Toc499503555)

[VI. Demo Host và Deploy Static Web on AmazonS3 26](#_Toc499503556)

[VII. Nguồn tham khảo. 26](#_Toc499503557)

# **Dịch vụ AmazonS3 là gì? Ưu điểm để chọn AmazonS3? So sanh với các dịch vụ khác? Nhược điểm?**

**Định nghĩa:**

Dịch vụ Amazon S3,tên đầy đủ là Amazon simple storage Service Là dịch vụ lưu trữ dữ liệu online(cloud) giúp bạn có thể lưu trữ dữ liệu,truy xuất mọi nơi có internet với sự tiện lợi cao và chi phí thấp.Nó cung cấp cho các lập trình viên truy cập vào cùng một cơ sở dữ liệu lưu trữ dữ liệu không tốn kém, đáng tin cậy, nhanh chóng, rẻ tiền mà Amazon sử dụng để điều hành mạng lưới các trang web toàn cầu của riêng mình. Dịch vụ này nhằm mục đích tối đa hóa lợi ích của quy mô và để vượt qua những lợi ích đó cho các lập trình viên.

**Lý do lựa chọn Amazon S3:**

Được phát triển từ Amazon,một trong những công ty lớn trên thế giới nên dịch vụ sử dụng cũng sẽ tin cậy hơn.

**Tính ổn định**: Nó được thiết kế để chịu được các hỏng hóc và phục hồi hệ thống rất nhanh với thời gian tối thiểu. Amazon cung cấp một thỏa thuận cấp dịch vụ (service-level agreement – SLA) để duy trì tính sẵn sàng ở mức 99.99 phần trăm.

**Đơn giản**: Dễ dùng S3 được xây dụng trên các khái niệm đơn giản và cung cấp tính mềm dẻo cao cho việc phát triển các ứng dụng của bạn. Bạn có thể xây dựng các lược đồ lưu trữ phức tạp hơn, nếu cần, bằng cách thêm các hàm vào các thành phần của S3.

**Tính mở rộng**: Thiết kế của S3 cung cấp một cấp độ cao về tính mở rộng và cho phép sự điều chỉnh dễ dàng trong dịch vụ khi lượng truy cập vào ứng dụng web của bạn tăng đột biến với lưu lượng khổng lồ.

**Rất Rẻ**: Chi phí sử dụng S3 rất cạnh tranh với các giải pháp của công ty và cá nhân khác trên thị trường.

**Bất lợi của Amazon S3:**

Khó cấu hình chính xác dịch vụ khi sử dụng vài lần đầu nếu không được hướng dẫn đung.

Phí gia tăng khi kích thước hoặc sự phổ biến của trang web tăng lên.

**So sánh với các dịch vụ khác:**

Đối thủ cạnh tranh chính của Amazon S3 hiện nay là MS Azure của Microsoft và Google Cloud Storage của Google nên sẽ tiến hành so sánh với hai dịch vụ này.

Tiêu chí so sánh:So sánh dựa trên giá của một số lớp dịch vụ mà 3 dịch vụ trên cung cấp.

**HOT**: Đây là dịch vụ cung cấp khả năng truy xuất dữ liệu thường xuyên với tốc độ thực thi cao và bền vững.

**WARM**: Chi phí thấp hơn, nhưng ít bền hơn đối với các dữ liệu lặp lại không quan trọng được truy cập thường xuyên.

**COOL**: Lớp lưu trữ dữ liệu được truy cập ít thường xuyên hơn nhưng cần truy cập nhanh khi cần thiết.

**COLD**: An toàn, bền, và chi phí thấp dịch vụ lưu trữ để lưu trữ dữ liệu.

Bảng giá cho các dịch vụ trên(trên GB) cho 1TB lưu trữ đầu tiên:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Amazon S3** | **Microsoft Azure** | **Google Cloud Storage** |
| HOT | $ 0.023 | $ 0.0368 | $ 0.026 |
| WARM | $ 0.024 | $ 0.024 | $ 0.02 |
| COOL | $ 0.0125 | $ 0.01 | $ 0.01 |
| COLD | $ 0.004 | $ 0.0018 | $ 0.007 |

Bảng giảm giá cho dịch vụ HOT

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Amazon S3** | **Microsoft Azure** | **Google Cloud Storage** |
| 2-50 TB/month | $ 0.0230 | $ 0.0368 | $ 0.026 |
| 50-500 TB/month | $ 0.0220 | $ 0.0354 | $ 0.026 |
| 500+ TB/month | $ 0.0210 | $ 0.0339 | $ 0.026 |

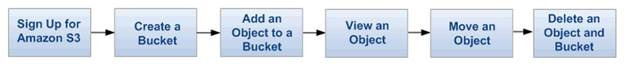
Bảng giảm giá dịch vụ WARM

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Amazon S3** | **Microsoft Azure** | **Google Cloud Storage** |
| 2-50 TB/month | $ 0.0236 | $ 0.0184 | $ 0.020 |
| 50-500 TB/month | $ 0.0232 | $ 0.0177 | $ 0.020 |
| 500-1000 TB/month | $ 0.0228 | $ 0.0170 | $ 0.020 |
| 1000-1500 TB/month | $ 0.0224 | $ 0.0170 | $ 0.020 |
| >5000 TB/month | $ 0.0220 | $ 0.0170 | $ 0.020 |

Các lớp dịch vụ COOL và COLD không giảm giá.

# **Cơ bản về hệ thống.**

Amazon Simple Storage Service (Amazon S3) là vùng lưu trữ trên Internet. Dùng Amazon S3 để lưu trữ và truy xuất bất kỳ dữ liệu tại bất kỳ thời điểm nào, từ bất cứ nơi đâu trên web. Bạn có thể tương tác với Amazon S3 dùng AWS Management Console, giao diện web đơn giản nhưng trực quan.



Những hoạt động cơ bản ta có thể làm thông qua giao diện Amazon S3, tạo bucket, thêm object vào bucket, xem object, di chuyển, xóa object và bucket.

Để hiểu cơ bản về Amazon S3, bạn chỉ cần biết 2 khái niệm là bucket và object. Amazon S3 lưu trữ dữ liệu như các object trong các bucket. Một object gồm 1 file và metadata mô tả cho file đó.

# **Cách thức hoạt động.**

Mỗi đối tượng lưu trữ trong một bucket trên S3 được định danh bởi một khóa duy nhất. Nó cũng giống như khái niệm tên tệp trong một thư mục trên hệ thống tệp của bạn. Tên tệp trong một thư mục trên đĩa cứng của bạn phải là duy nhất. Mỗi đối tượng trong một bucket chỉ có một khóa. Tên của thùng và khóa được sử dụng cùng nhau để cung cấp định danh duy nhất cho mỗi đối tượng lưu trữ trong S3.

**Đối với việc lập trình các ứng dụng web có thể kết nối đến Amazon S3, bạn cần nằm các điều sau:**

Amazon S3 là một dịch vụ REST, ta có thể gửi yêu cầu đến Amazon S3 bằng cách sử dụng API REST hoặc AWS SDK, giúp cho việc lập trình của bạn trở nên đơn giản hơn.

Mọi tương tác với Amazon S3 đều được đánh dấu là “chứng thực” hoặc ”ẩn danh”. Xác thực là một quá trình xác minh danh tính của người yêu cầu cố gắng truy cập vào một sản phẩm của Amazon Web Services (AWS).

Yêu cầu được xác thực phải bao gồm một chữ ký chứng thực người gửi yêu cầu. Giá trị chữ ký, một phần, được tạo ra từ các phím truy cập AWS của người yêu cầu (AWS access key).

1. Khái niệm Bucker và Object

Bucker và object được xem là các tài ngiên mà Amazon S3 cung cấp các API để bạn quản lí chúng..

**Bucker**

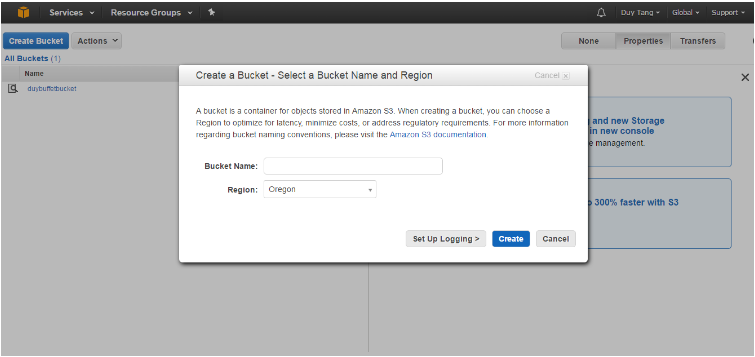
Bạn sẽ tạo một Bucket chính để quản lí(bạn được phép tạ tối đa 100 bucket, mỗi bucket được xác định duy nhất) các thư mục và file lưu trữ (gọi là object

Tên của mỗi bucket đó phải là duy nhất, vì thế lúc tạo bucker, hệ thống sẽ xét tên bucker đó có đã tồn tại chưa.

Và đặt tên nó nên tuân theo luật sau:

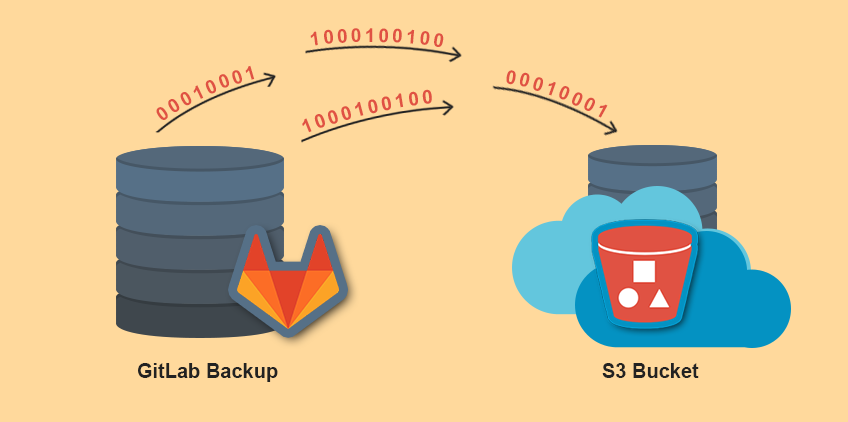
* Bucket có thể chứa các ký tự như chữ cái, hoặc số, dấu gạch dưới(\_\_), dấu chấm, dấu phảy hoặc dấu gạch trên(-)
* Bucket cần phải bắt đầu số hoặc ký tự
* Bucket không quá ngắn cần phải đặt tên trên 3 và không quá dài ký tự nên dưới 255
* Bucket không thể chứa dạng địa chỉ IP, ví dụ 192.168.1.1 sẽ không hợp lệ

Amazon S3 tạo bucket trong một khu vực bạn chỉ định, bạn có thể chọn bất kì khu vực địa lí nào. Việc chọn giúp ta tối ưu hóa độ trễ, giảm thiểu chi phí,…

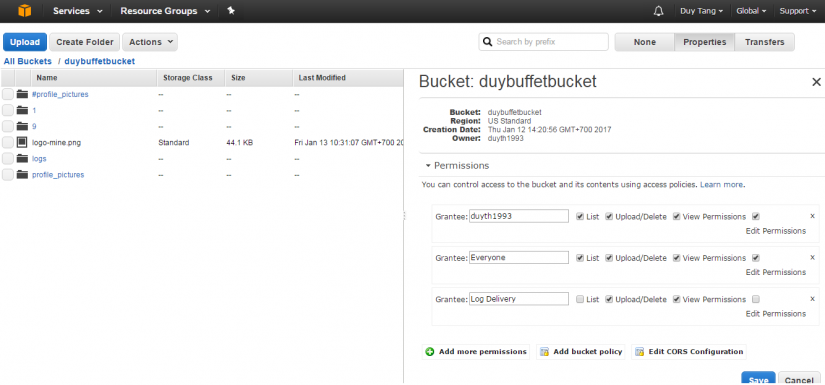
****

Bucker: là nơi chứa các đối tượng lưu trữ trong Amazon S3, có thể hiểu như một folder chứa trong một hệ thống tệp. Để truy cập một bucker, bạn có thể thông qua một URL.

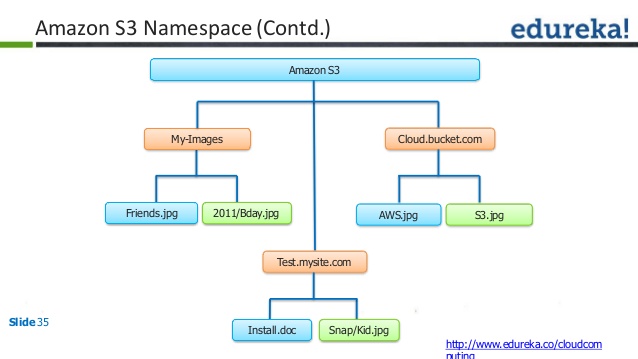
VD: <http://buffet.s3.amazonaws.com> (truy cập bucker tên buffet).

****

Dưới đây là tổng hợp tất cả folder và file trong bucket của ta:

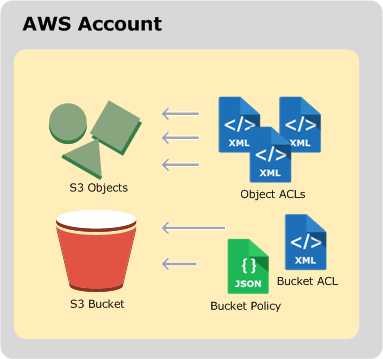


Tóm lại, cấu trúc của hệ thống lưu trữ của Amazon s3 cũng tương tự như các cậy thư mục ta sử dụng lưu dữ liệu trên máy.



**Object**

Object: Các đối được lưu trữ trong các bucker, các đối tượng này có thể là file ảnh, video hay loại khác. Khi thêm tập tin vào Amazon S3, có thể tùy chọn thêm metadata và gán quyền điều khiển việc truy cập vào tập tin.



Không giống như sự giới hạn về số lượng các thùng, số lượng đối tượng không bị hạn chế. Bạn có thể lưu trữ một số lượng vô hạn các đối tượng trong các thùng của bạn, và mỗi đối tượng có thể chứ đến 5GB dữ liệu. Cá dữ liệu trong các object trên S3 của bạn có thể lấy được qua HTTP, HTTPS, hoặc BitTorrent

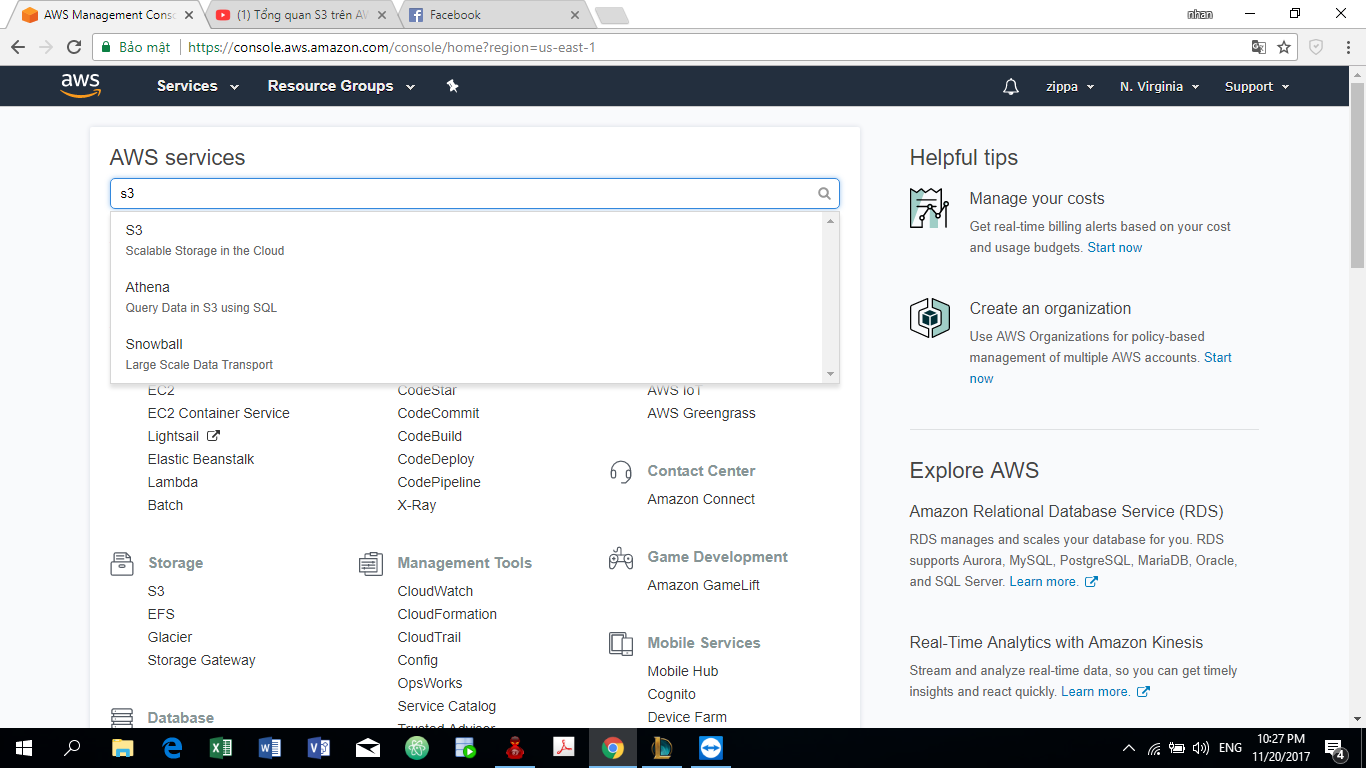
Tất cả các thông tin của một object mà bạn cần biết:

* **Key**: tên mà bạn gán cho object, bạn sẽ lấy đối tượng thông qua tên này.
* **Version ID**: Trong một bucker, ID chính là thông tin giúp nhận diện một object với các object còn lại. ID Version là chuỗi mà Amazon S3 tạo ra khi bạn thêm object vào một bucker.
* **Value**: Nội dung mà object lưu trữ. Đó là một dãy byte có kích thước từ 0 đến 5TB.
* **Metadata**: một tập hợp các tập tên-giá trị mà bạn lưu trữ các thông tin liên quan đến object. Bạn có thể gán một metadata, được gọi là metadata do người dùng định nghĩa đến object của bạn trong Amazon S3. Amazon S3 cũng chỉ định hệ thống siêu dữ liệu cho các đối tượng này, nó sử dụng để quản lý các đối tượng.
* **Subresource**: Amazon S3 sử dụng cơ chế cấp nguồn phụ để lưu trữ thông tin bổ sung cho một object cụ thể. Bởi vì subresource là cấp dưới của một object, nó liên kết đến một object hoặc một bucker khác.
* **Access control information**: bạn có thể kiểm soát truy cập các đối tượng bạn lưu trữ trong Amazin S3.

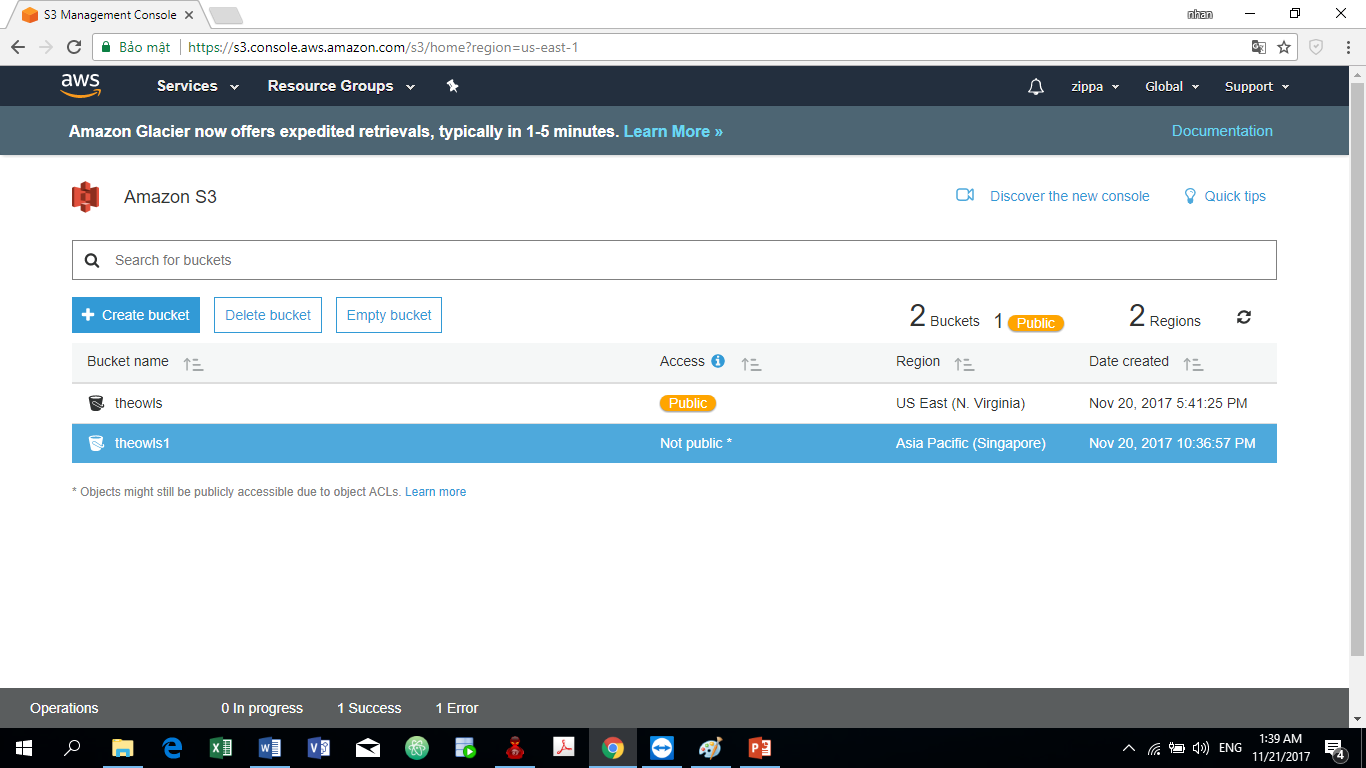
Lưu ý: các tài nguyên lưu trữ trên Amazon S3 của bạn mặc định được bảo mật riêng tư. Nếu bạn muốn chia sẻ các file ảnh hay video lưu trữ trên Amazon S3 đến trang web của mình, bạn cần đặt các đối tượng này công khai hay sử dụng một địa chỉ URL được chỉ định trên trang web của bạn.

# **Cài đặt và thao tác cơ bản.**

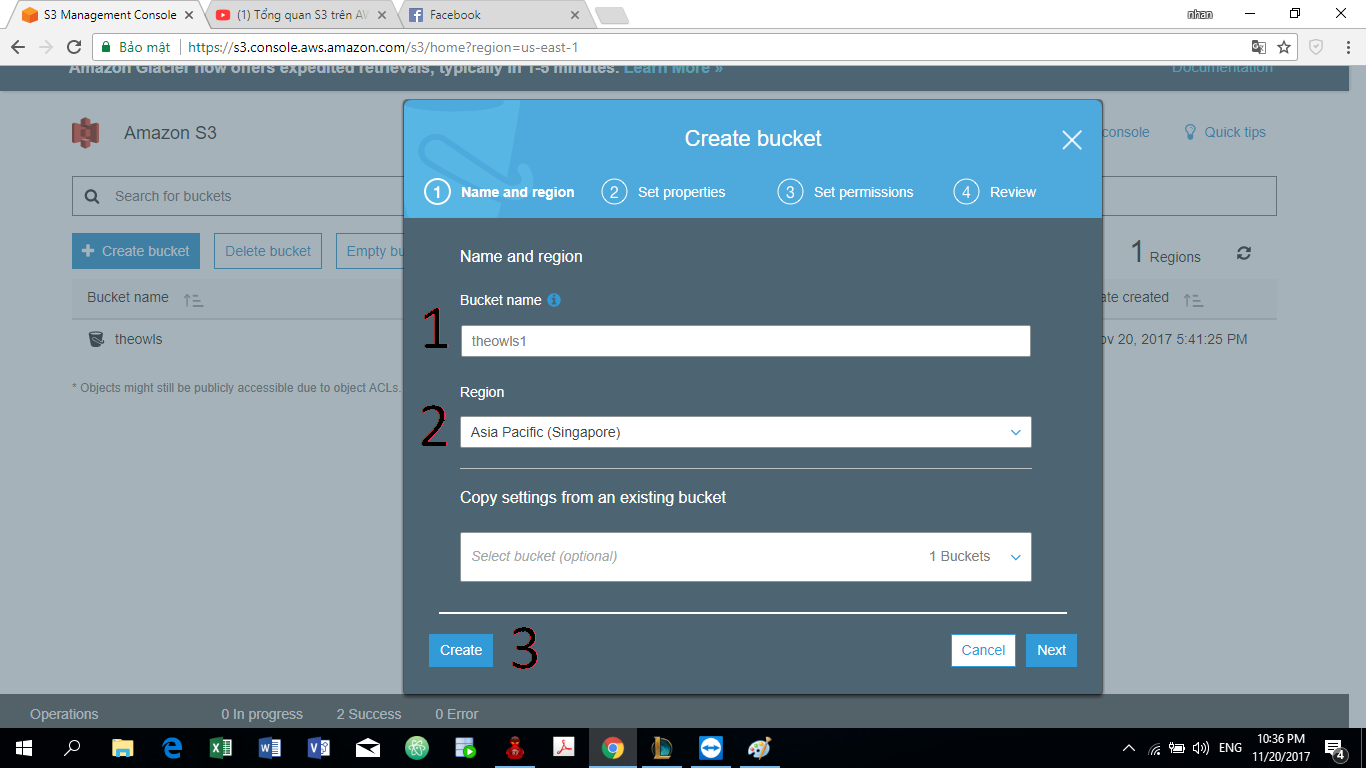
Sau khi đăng nhập vào trang, tìm kiếm services tên s3 ( viết tắt của amazon S3)

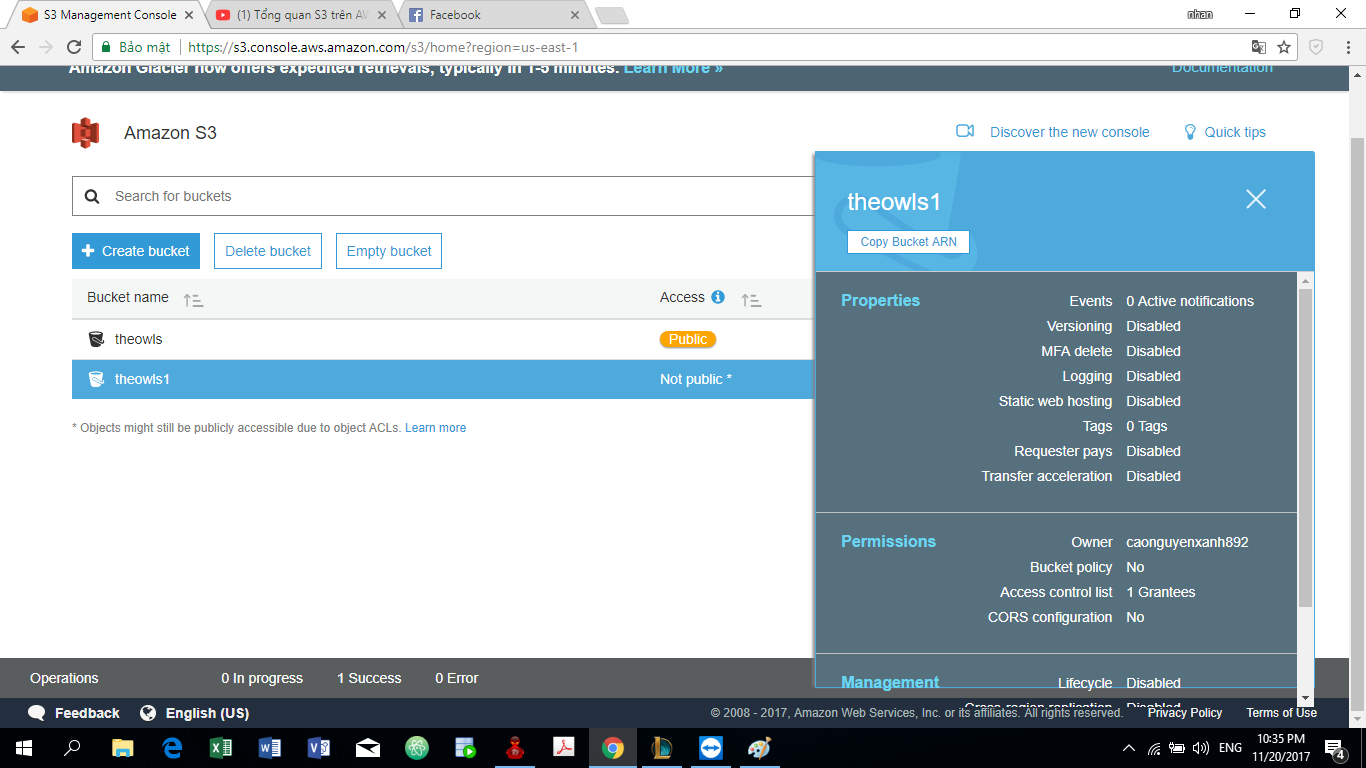


Giao diện tiếp theo sẽ hiển thị danh sách các bucket, nhấn nút **Create bucket** để thêm mới bucket

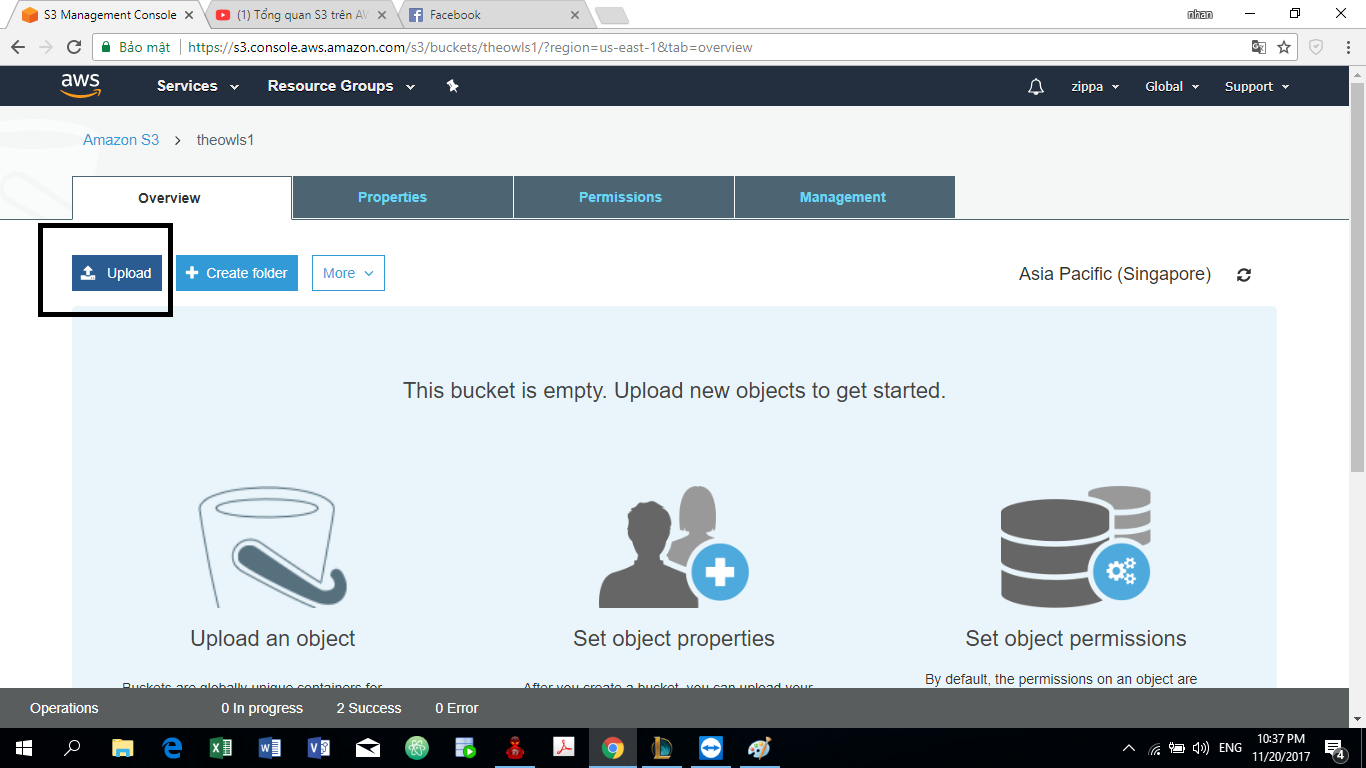


Sau đó điền tên bucket, chọn khu vực, rồi nhấn **Create**

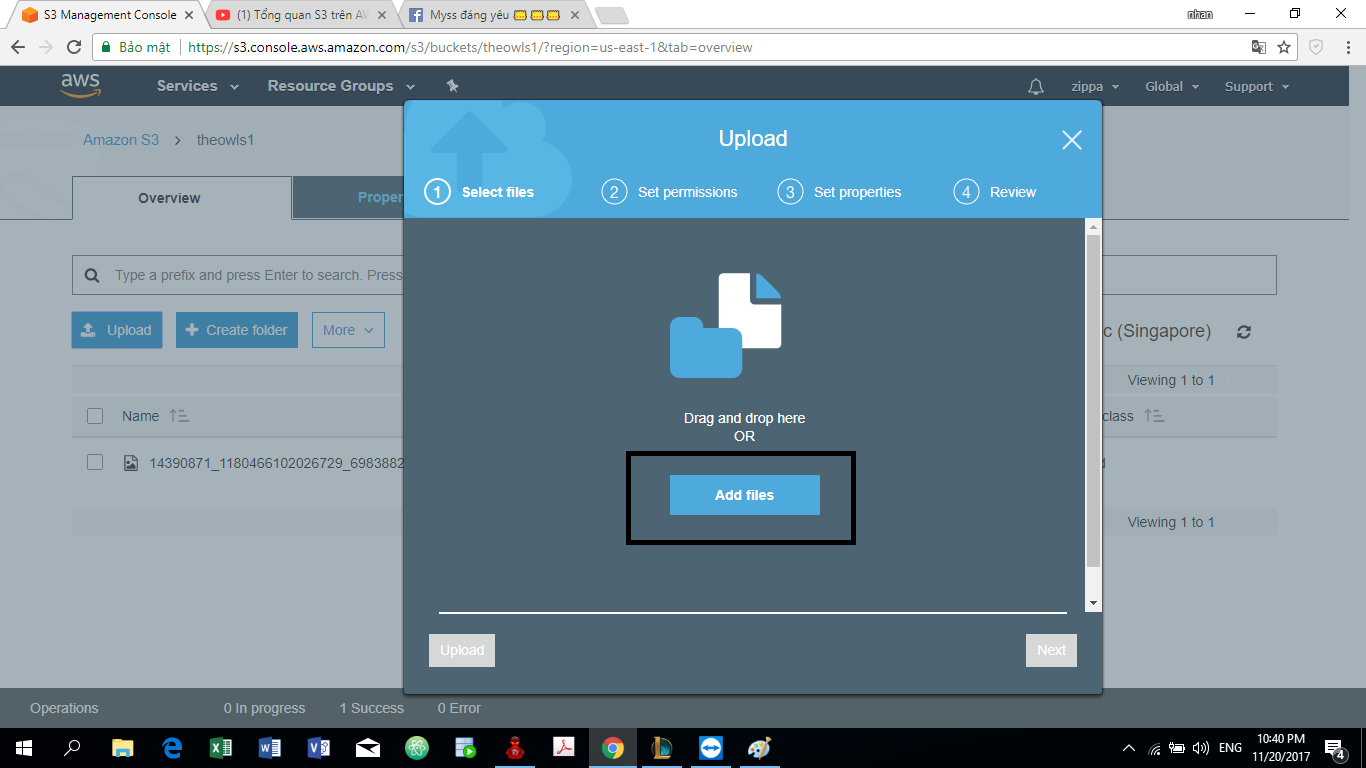


Bucket mới sau khi tạo sẽ có những thông số cơ bản.

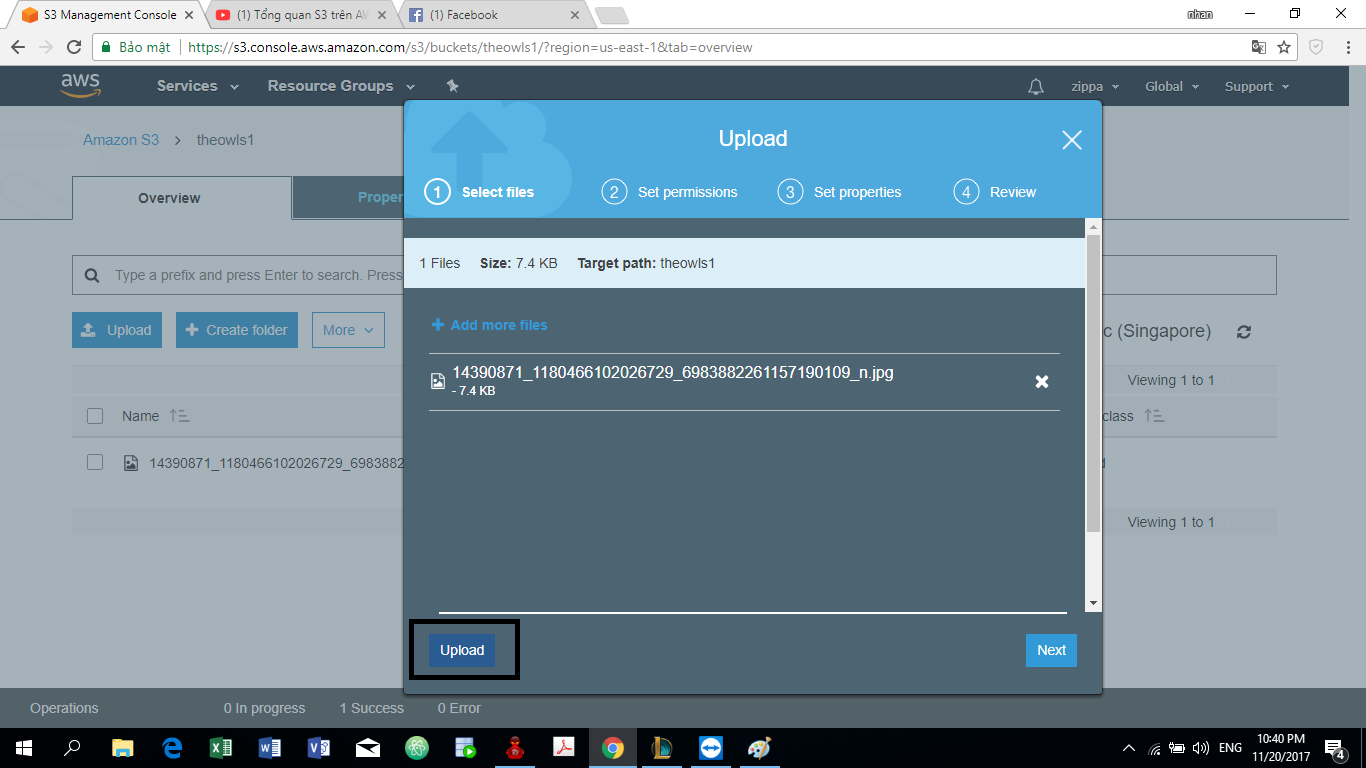
Tiếp đó nhấn vào bucket sẽ hiển thị danh sách các đối tượng đang đươc lưu trữ, nhấn **Upload**



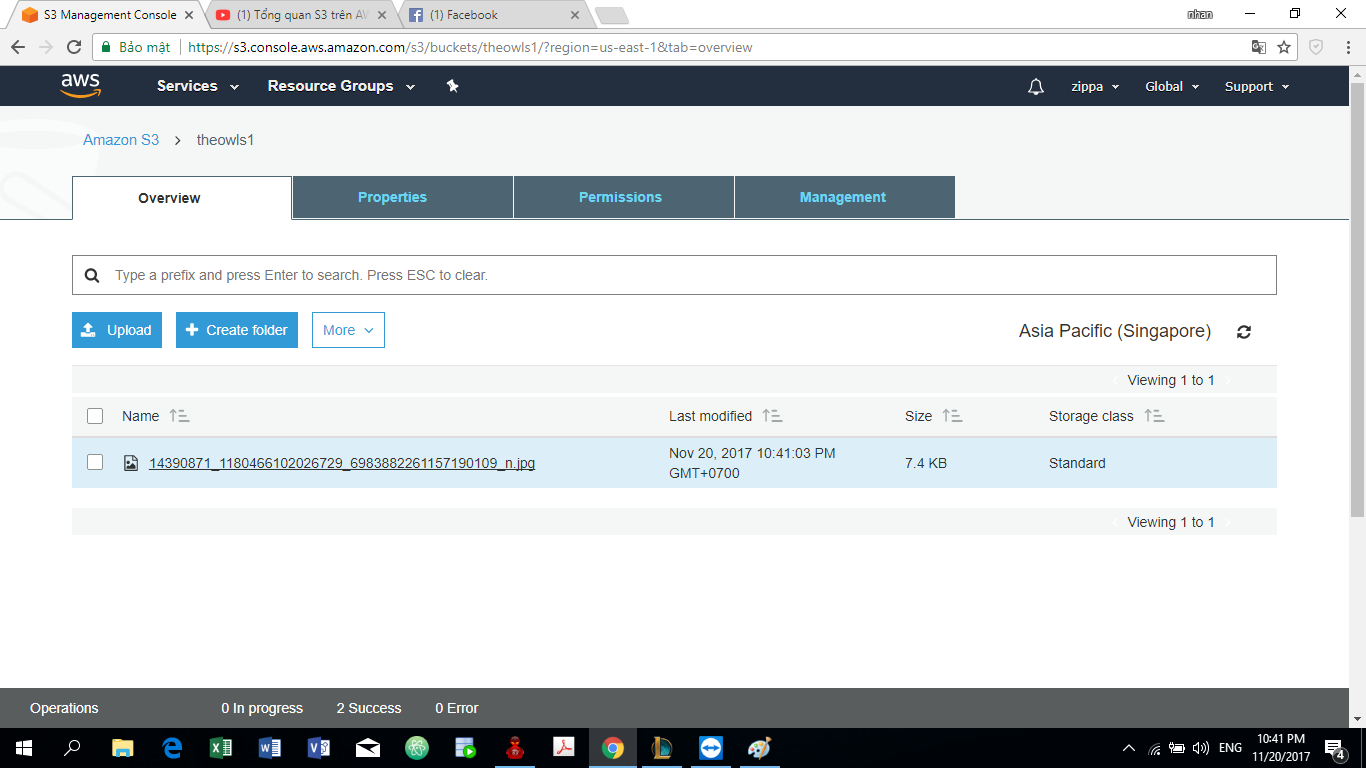
Chọn **add files**



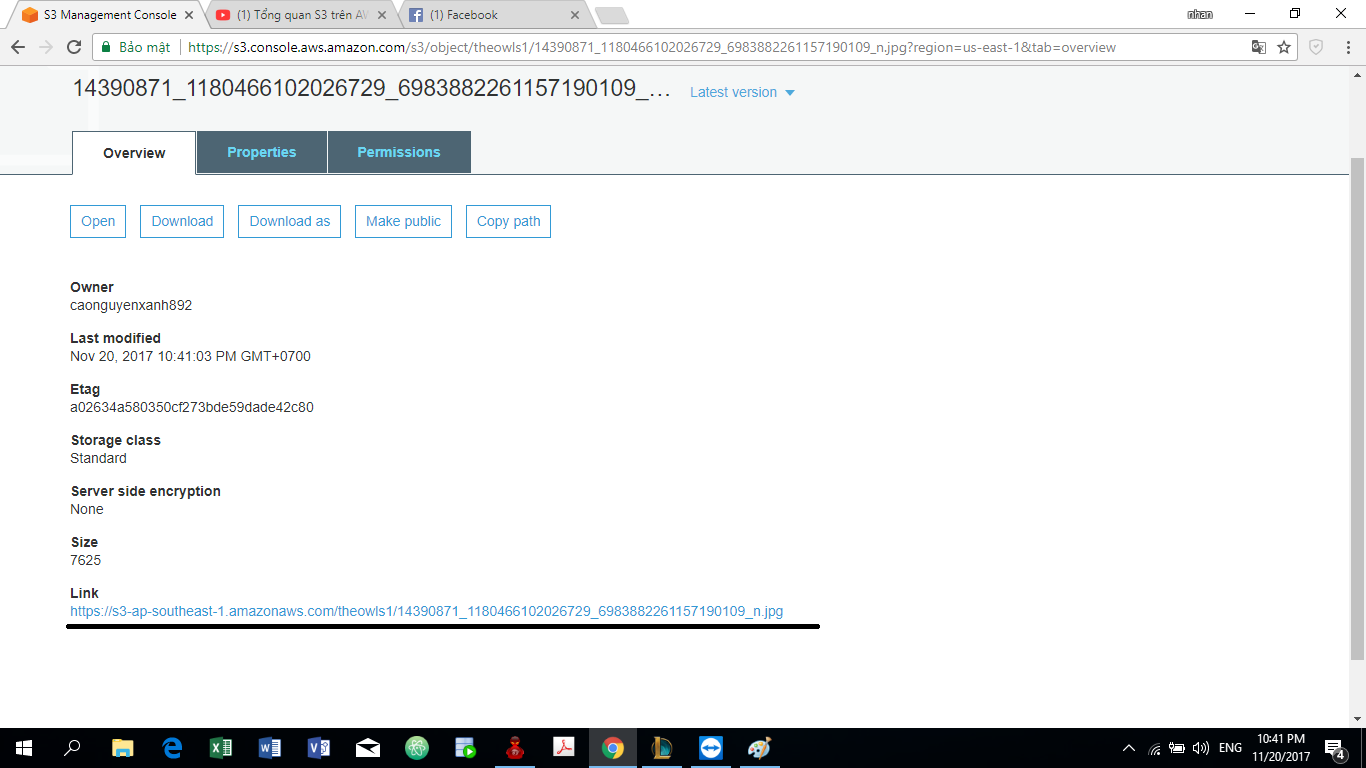
Chọn tệp cần tải lên, sau đó bấm **Upload**

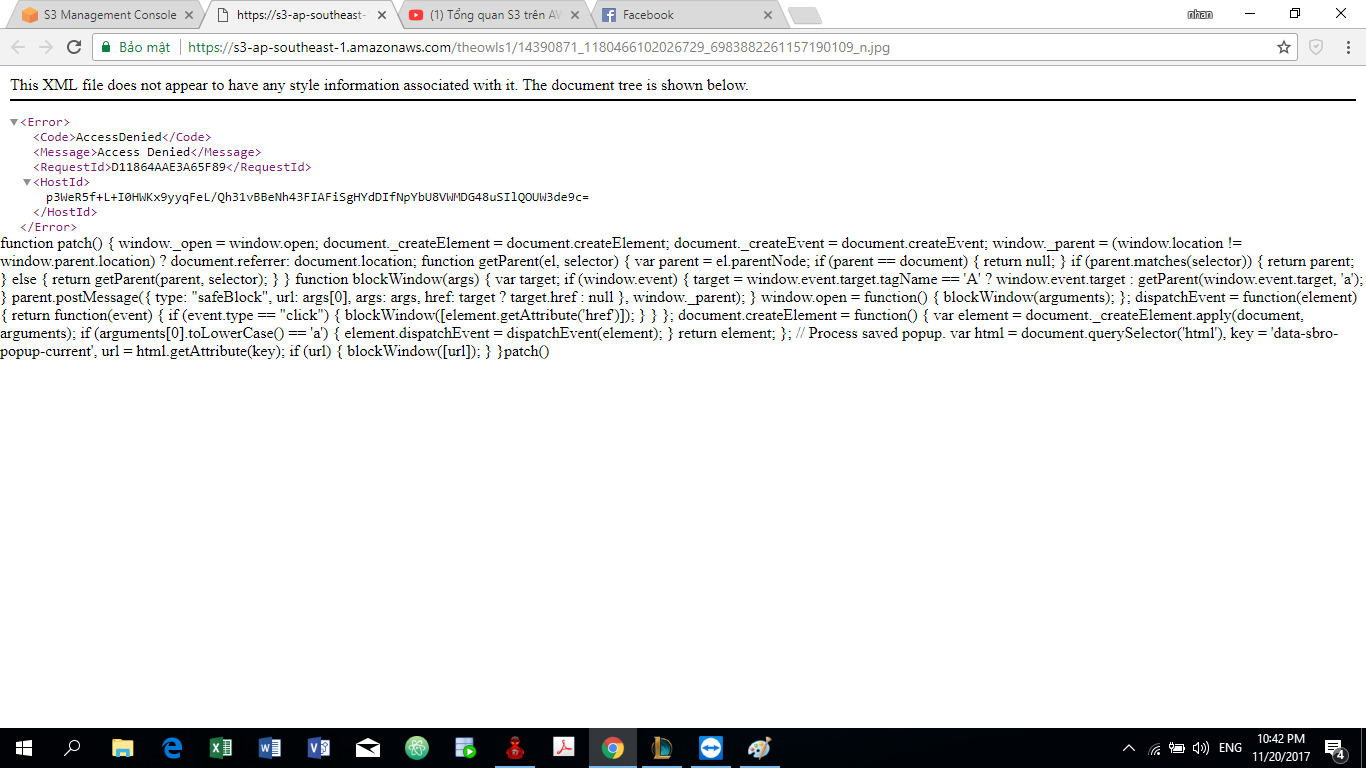


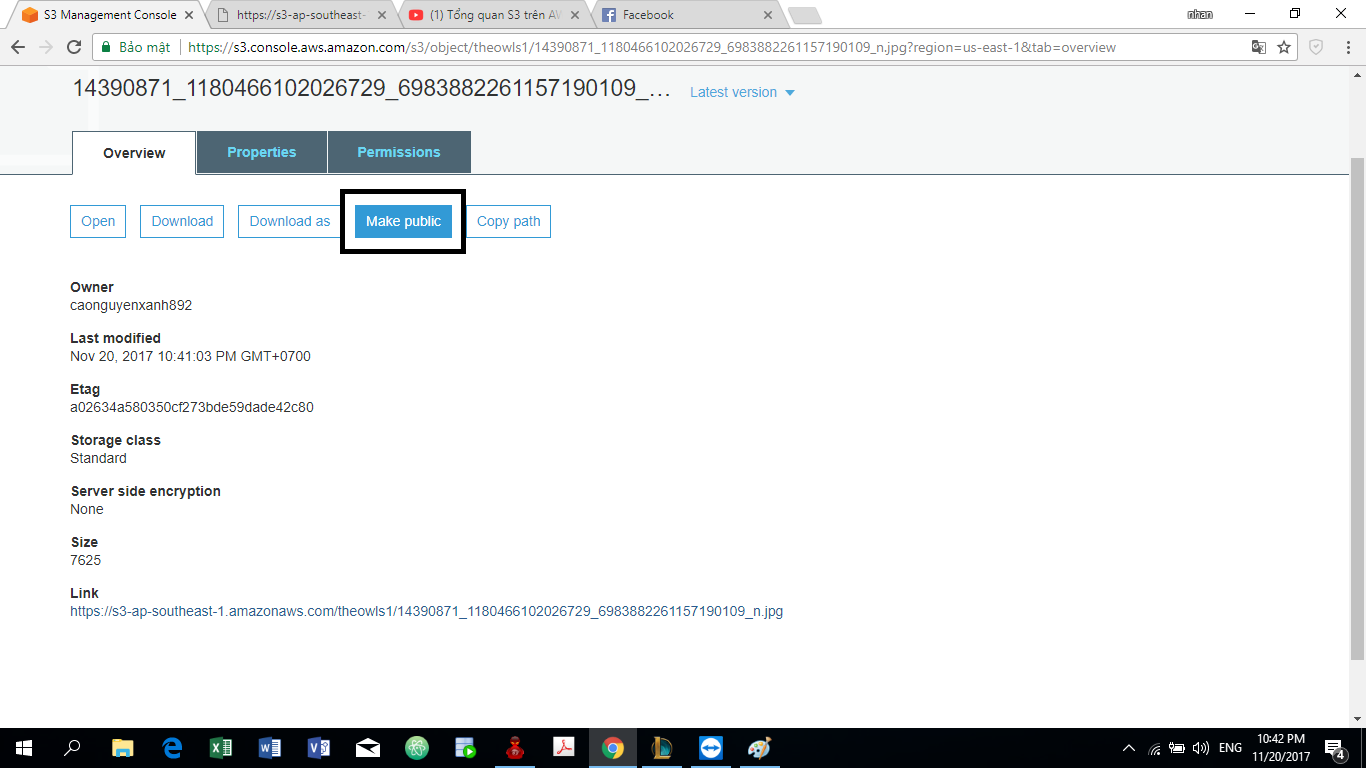
Bấm vào tệp vừa tải lên để xem các thông số cơ bản

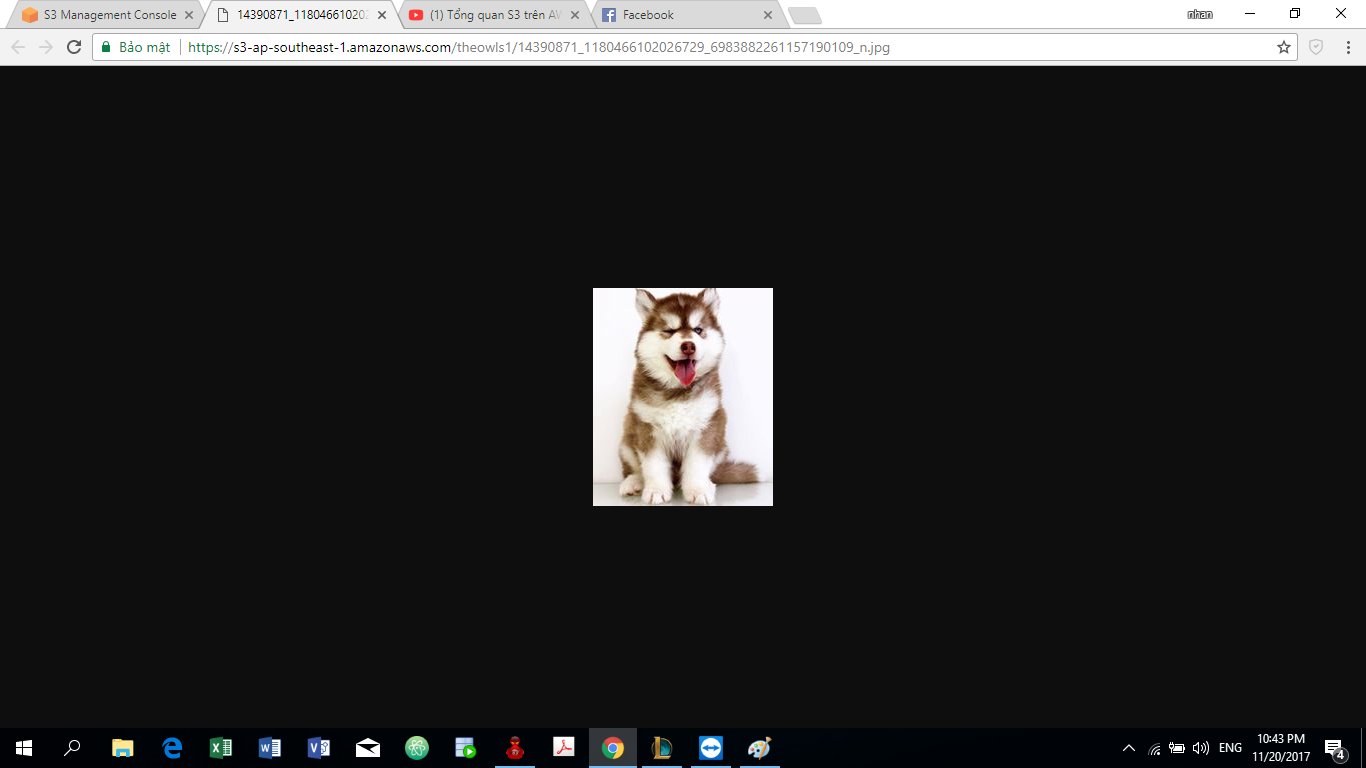


Các thông số của tệp, dưới cùng có một đường link để dẫn tới tệp này trên mạng

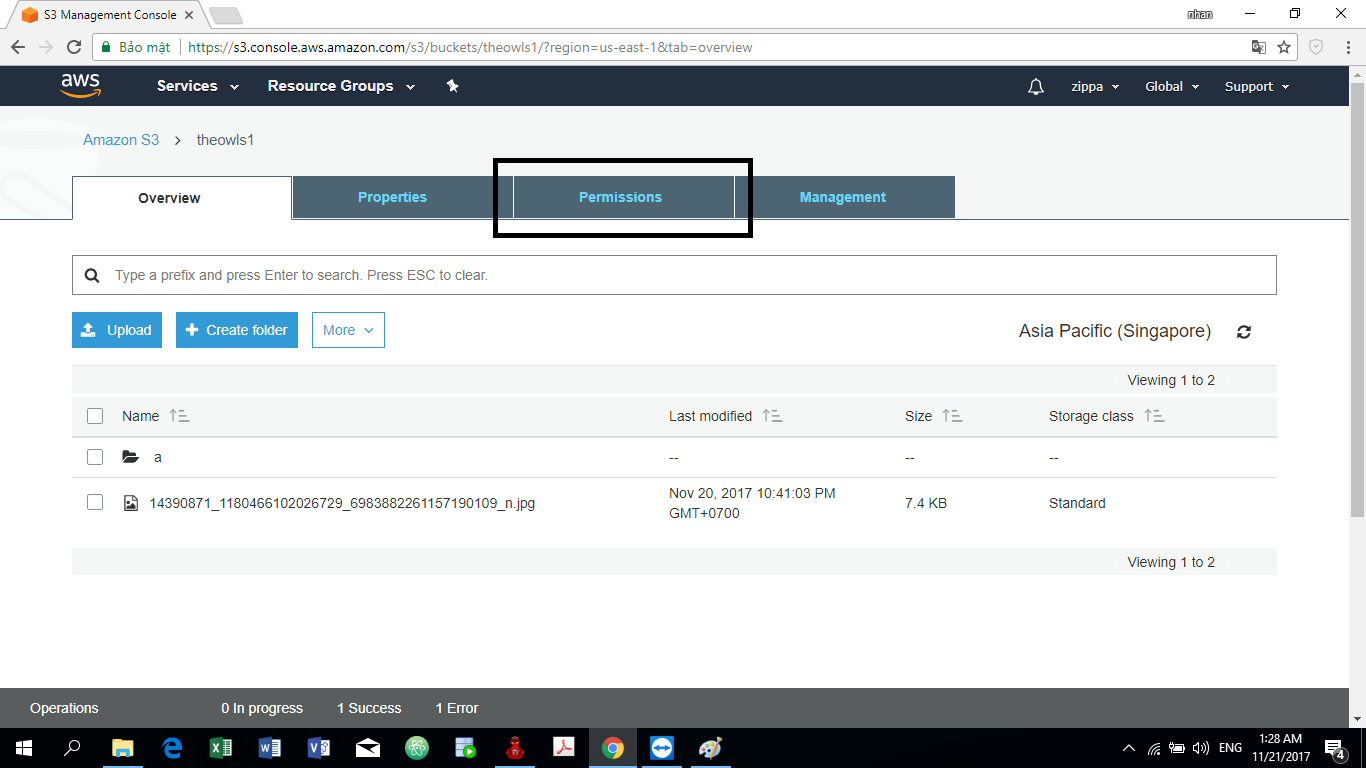


Truy cập vào đường link sẽ hiện rat rang web này vì tệp vừa được load lên chưa đang ở chế độ private, chưa được public ra ngoai.

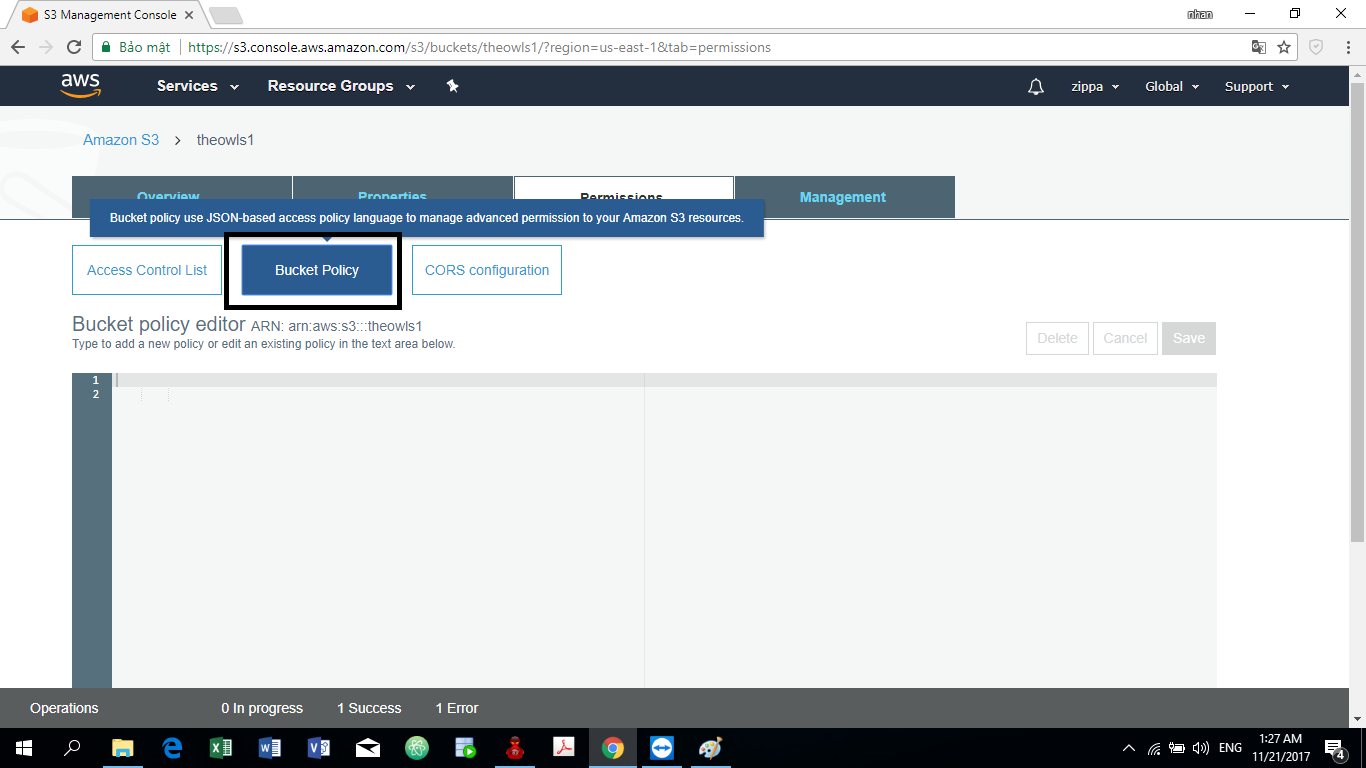
Để public được tệp, ta nhấn vào **Make public**

Sau đó truy cập lại vào đường link và tệp đã được public

Vậy với trường hợp load một lần nhiều tệp, sẽ phải public nhiều lần sẽ là một bất tiện. Vì thế để pulic ngay lập tức tệp vừa mới được load lên, ta cấu hình lại bucket, nhấn **Permissions.**



Nhấn **bucket policy**



Sau đó thêm đoạn code:

{

"Version": "2008-10-17",

"Statement": [

{

"Sid": "AllowPublicRead",

"Effect": "Allow",

"Principal": {

"AWS": "\*"

},

"Action": "s3:GetObject",

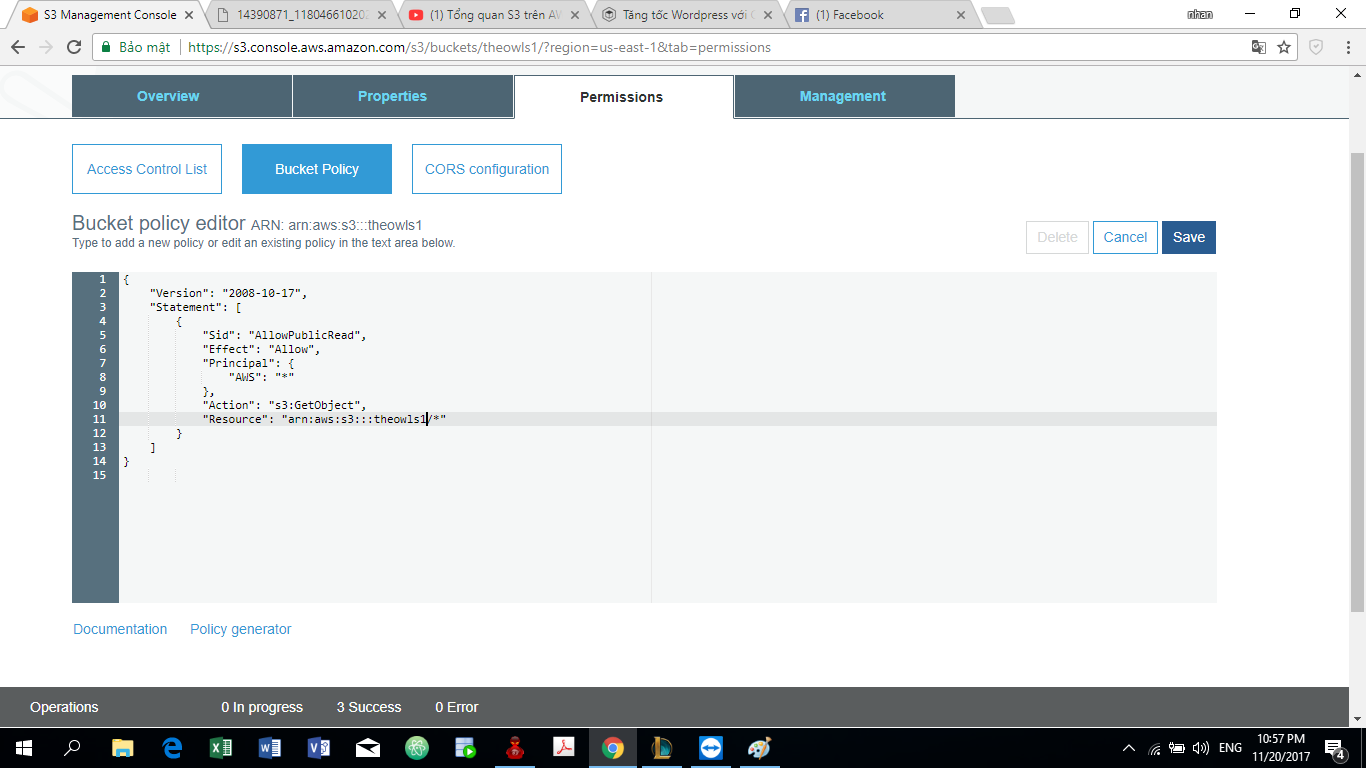
"Resource": "arn:aws:s3:::theowls/\*"

}

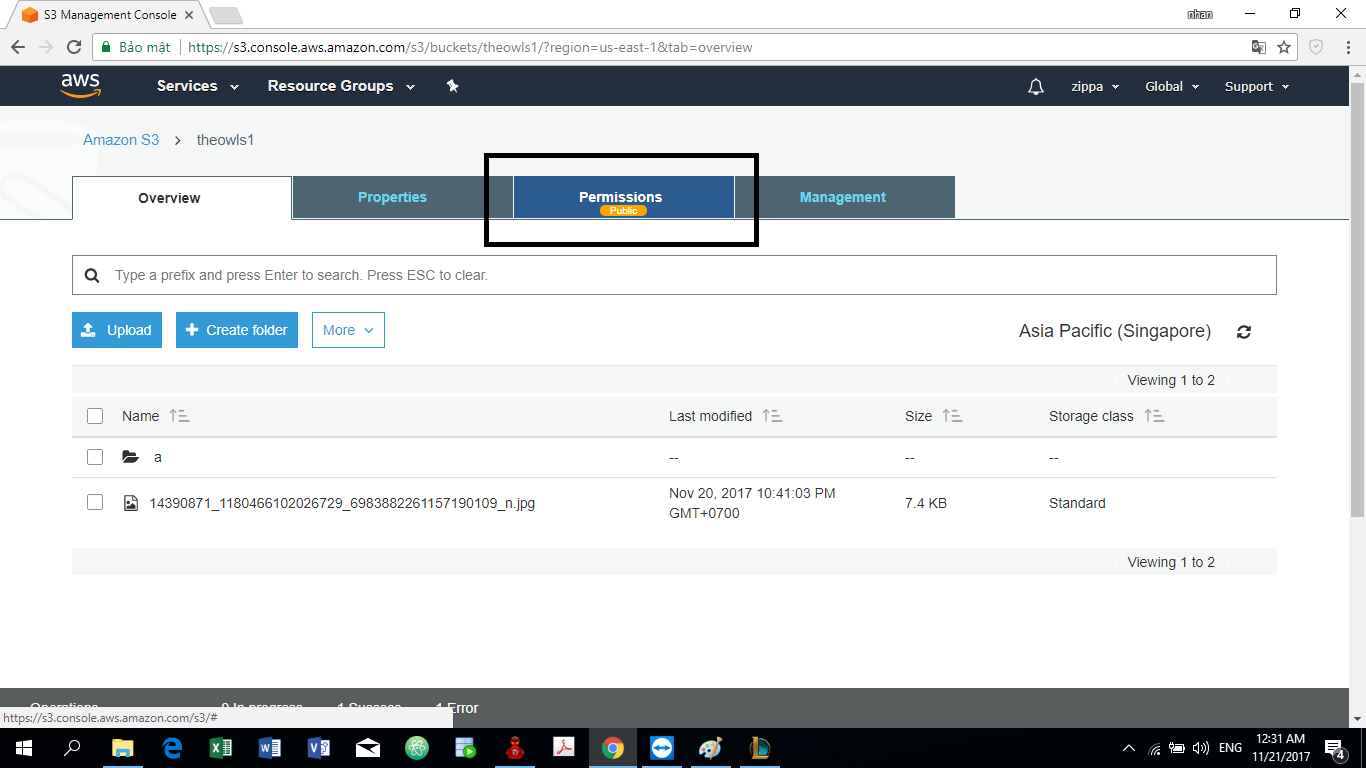
]

}

Và lưu lại



Lưu thành công sẽ hiện chữ Public ở trong ô **Permissions**



Thế là bucket đã được chuyển sang chế độ public, mỗi khi có tệp nào được up lên sẽ tự động chuyển sang chế độ public và nguồn bên ngoài có thể truy cập vào.

# **Các tính năng khác.**

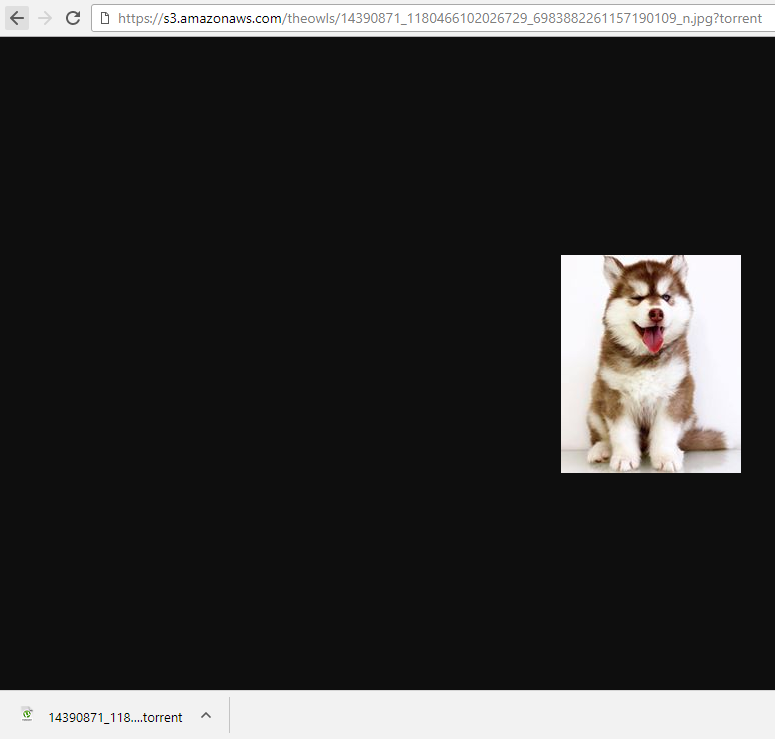
**AmazonS3 with Bittorent**

Bittorent là một phương thức peer-to-peer mở rộng cho dữ liệu phân tán. AmazonS3 có hỗ trợ phương thức Bittorent cho việc truy cập và download dữ liệu trong kho. Nhờ đó bạn có thể giảm chi phí truyền tải dữ liệu và cải thiện tốc độ truy cập.

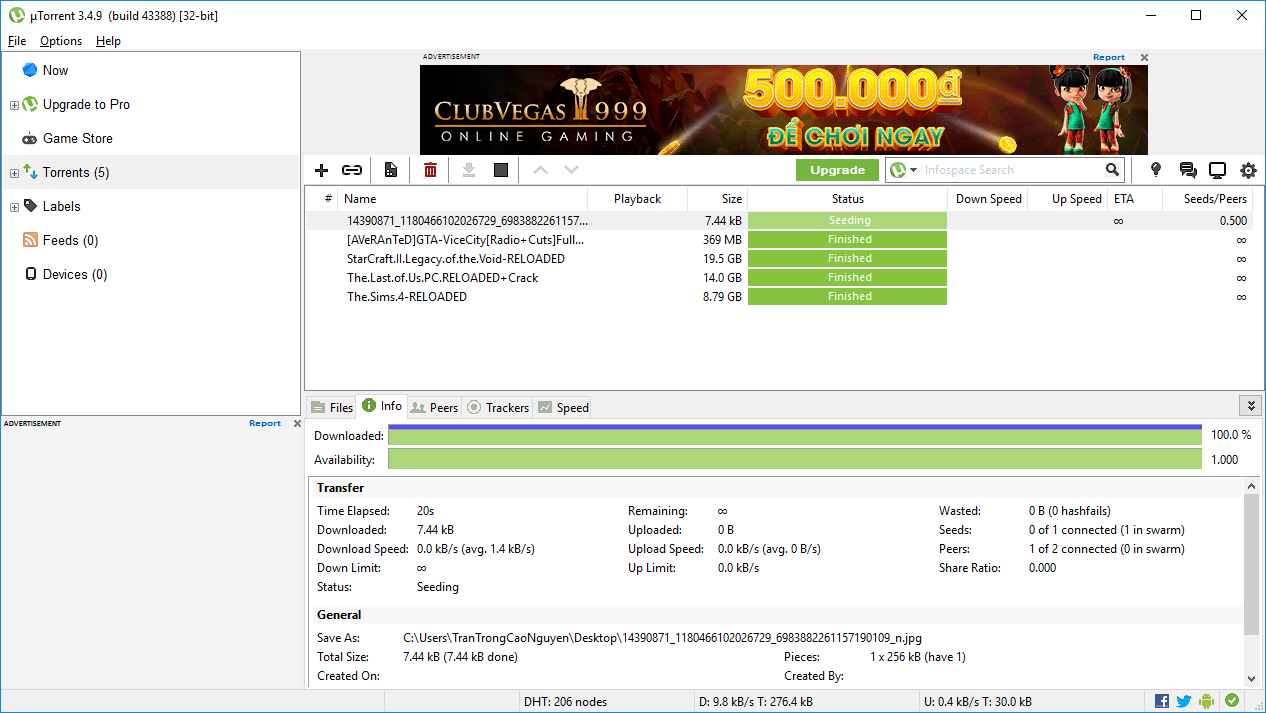
Thường áp dụng trên các bộ dữ liệu lớn và lượng truy cập nhiều.

Cách để có link torrent từ một object trong kho chứa dữ liệu:

https://s3.amazonaws.com/namebucket/nameoject?torrent



Dùng một trình torrent để mở file torrent vừa download. Bạn có thể thực hiện việc tải xuống tệp dữ liệu gốc từ kho dữ liệu của AmazonS3



Dữ liệu truyền tải dùng Bittorrent có thể được tối ưu và truyền tải nhanh, tùy thuộc vào độ phổ biến của dữ liệu truyền tải (có nhiều người tải không). Nếu không phổ biến, chỉ 1 người tải và seed thì tốc độ chuyển dữ liệu cũng giống như tốc độ truyền tải trực tiếp client/server của AmazonS3

Một chút về Torrent…

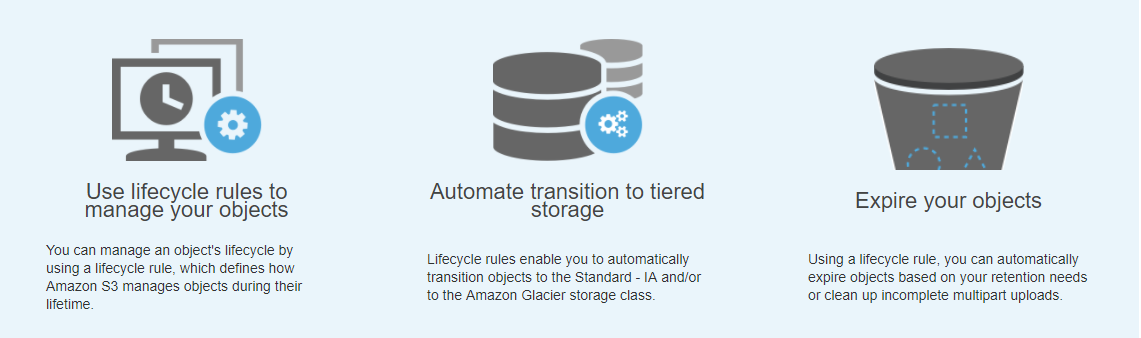
Torrent là một mạng lưới P2P(peer to peer = ngang hàng, tức là nhiều người cùng kết nối trực tiếp với nhau để chia sẻ file). Chuyên được dùng để trao đổi nhưng dữ liệu như Movie, Games, Software, Anime,... và một số File có dung lượng lớn!

Giao thức BitTorrent có thể thay thế cho máy chủ và cung cấp mạng lưới để truyền cac file lớn. Tốt hơn là download 1 file từ một máy chủ (nguồn cấp) duy nhất, BitTorrent cho phép người dùng tham gia vào một cộng đồng mạng để có thể upload/download (trao đổi tệp tin) với nhau.

Nếu như sử dụng phương thức tải truyền thống thì tốc độ tải về tập tin sẽ phụ thuộc vào đường truyền Internet và khả năng đáp ứng của máy chủ tại thời điểm đó. Nếu như có nhiều người cùng truy cập vào máy chủ để tải file thì tốc độ sẽ bị giảm xuống. Còn về phần máy chủ nếu hoạt động quá cao cũng sẽ có nguy cơ bị sập.

Ngược lại, phương thức tải Torrent sẽ có tốc độ download càng nhanh nếu như có càng nhiều người tham gia vào download và chia sẻ lại tập tin mà bạn đang cần tải.

**Object LifeCycle Management**

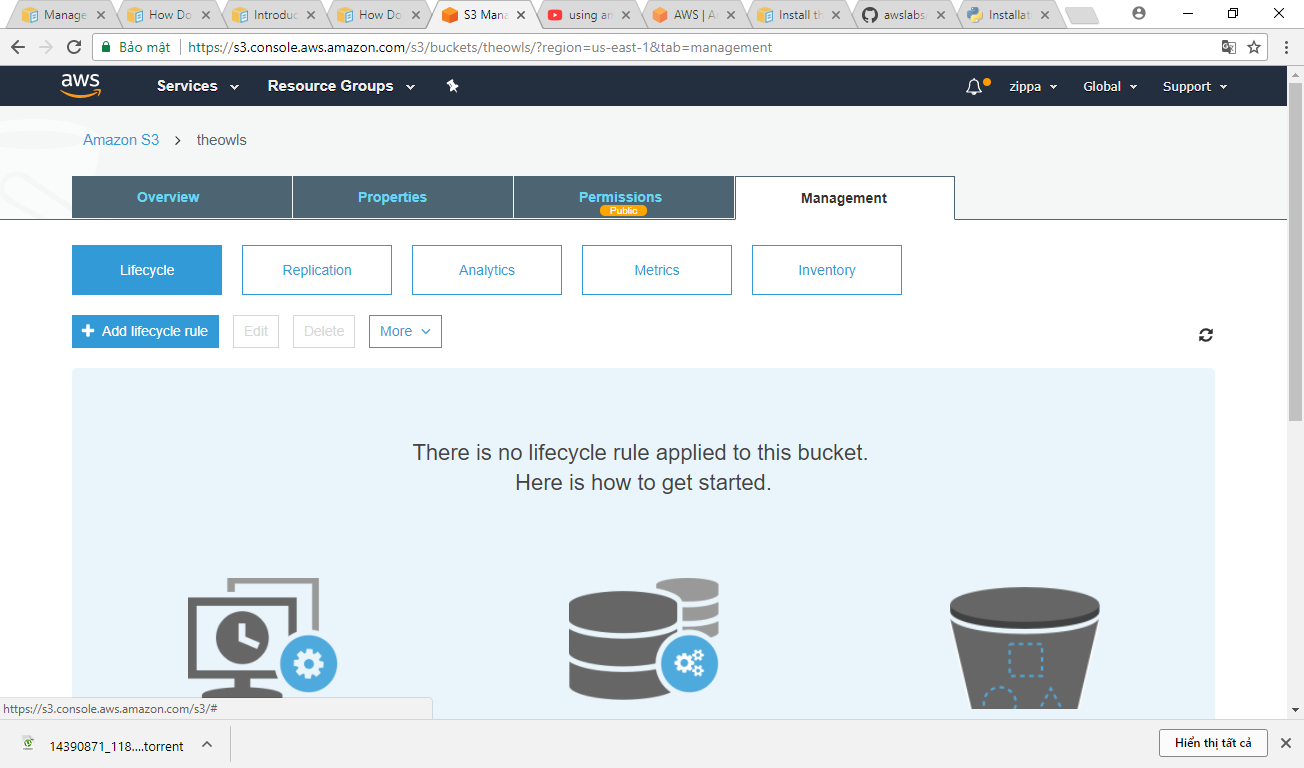


Là một cách để quản lý các Object trong Bucket. Có thể xem là một tập các điều lệ, quy định, hoặc bảng biểu hoạt động cho các Object trong quãng thời gian tồn tài của Object.

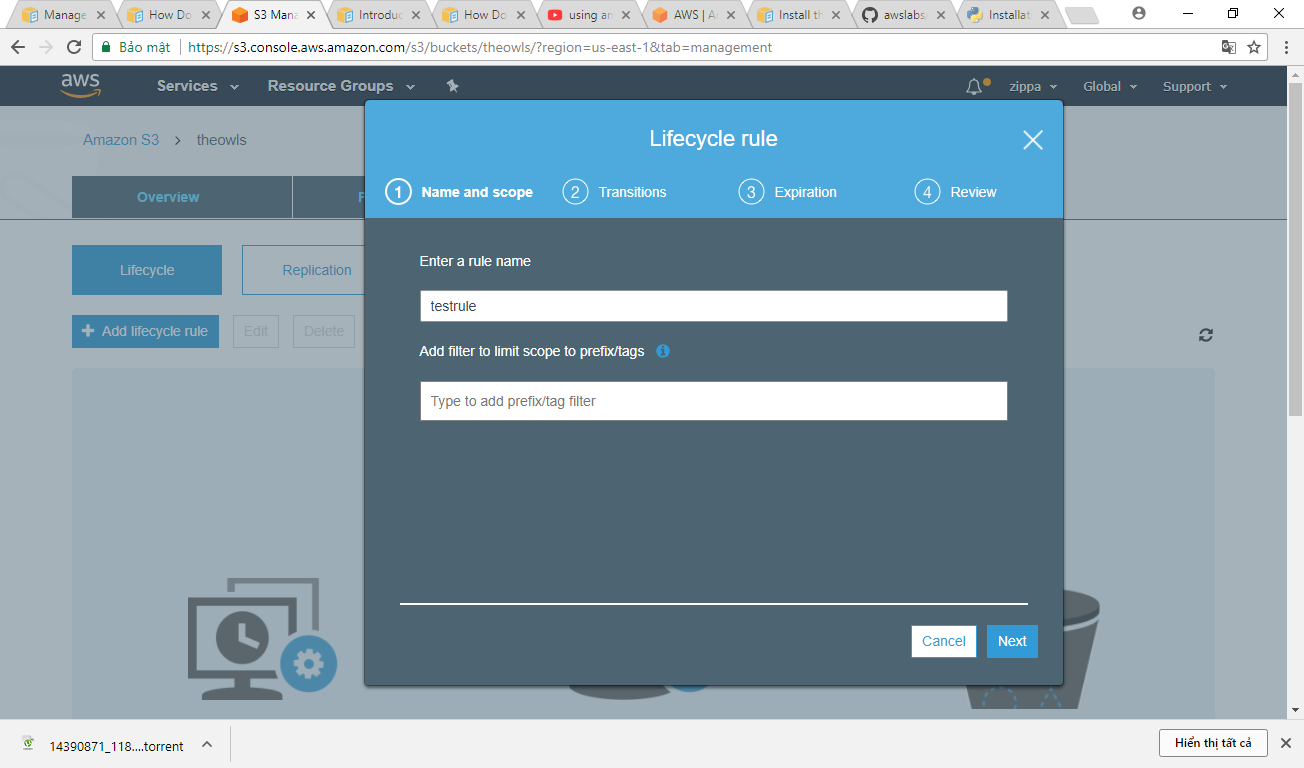
Ví dụ: Sau 6 ngày thì file abc.png trong bucket A phải bị xóa.

Bằng việc áp dụng một số điều luật mà bạn muốn qua cơ chế Manage Lifecycle Configuration được cung cấp trên AmazonS3. AmazonS3 có thể đóng vai trò một admin tự động điều phối hoạt động của kho lưu trữ AmazonS3.

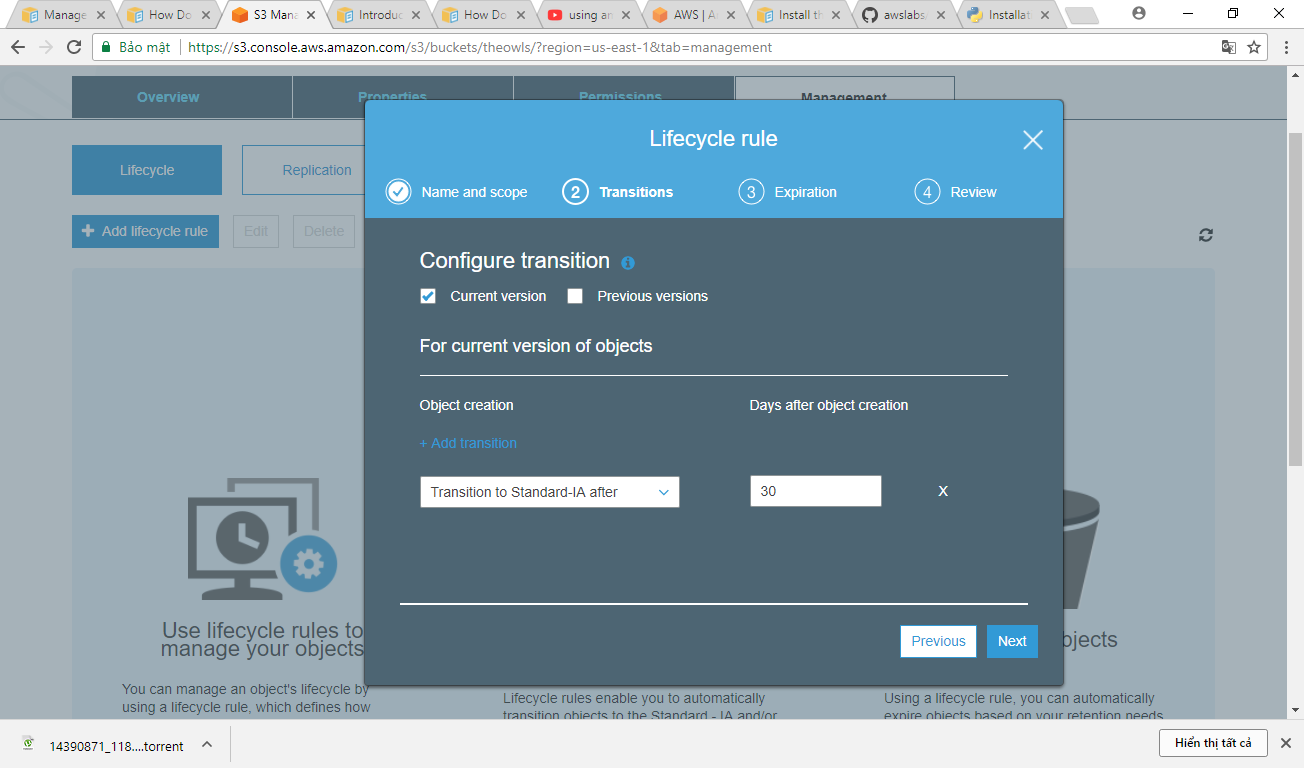
Các bước tạo một Lifecycle Rule (thông qua giao diện):



Chọn tab Management trong thẻ Bucket. Chọn Add lifecycle rule.



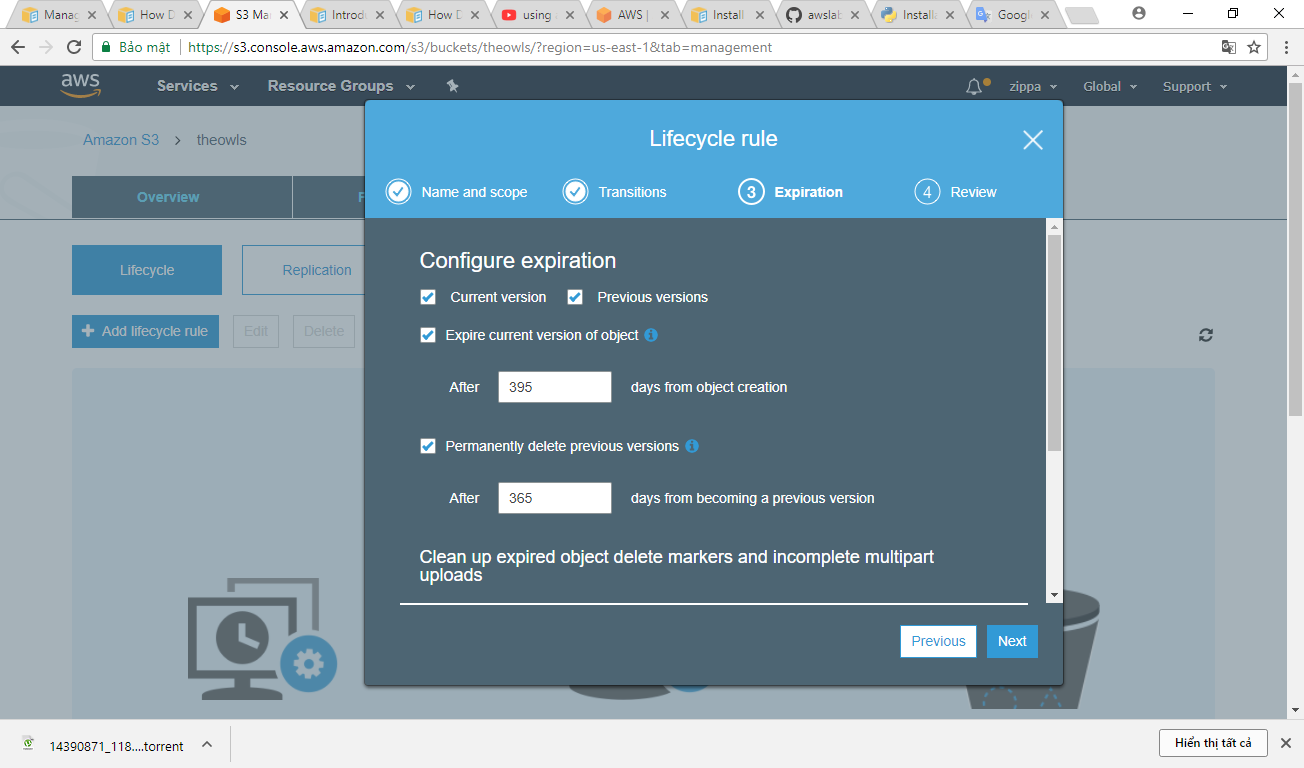
Đặt tên rulename



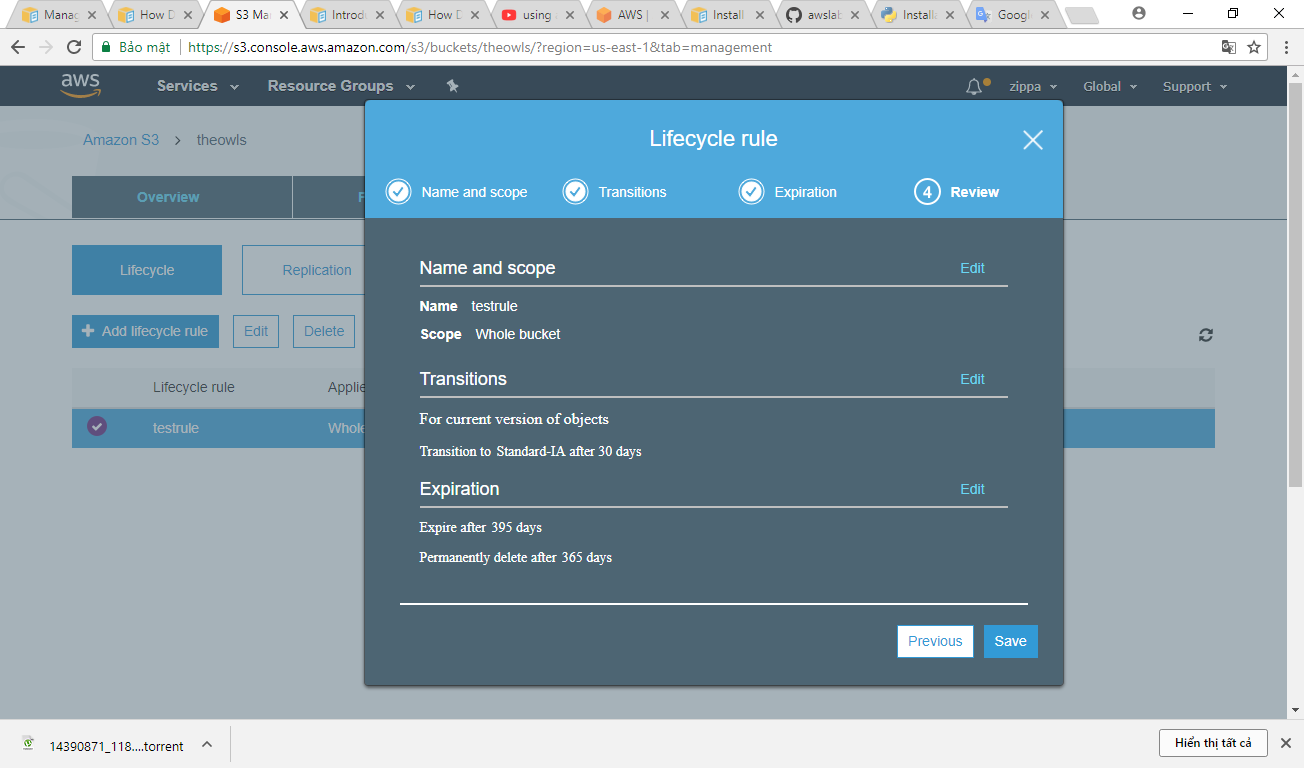
Chọn loại Transition

Current version là những hành động sẽ ảnh hưởng lên quá trình chuyển đổi hệ thống hiện tài. Ngược lại Previous Versions là cho các phiên bản trước.

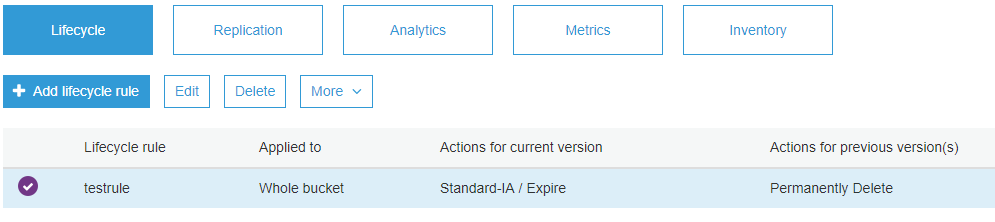
Chọn Add transition. Chọn Transition to Standard-IA after. Days after object creation chọn 30 ngày. Tức 30 ngày sau khi object được khởi tạo thì hành động sẽ có hiệu lực.



Tab Expiration để chọn thời gian ngày xóa Delete Object kể từ lúc khởi tạo.



Mục review để xem thống kê lại các luật mà bạn đã khởi tạo. Chọn mục Save để tiến hành tạo Rule.



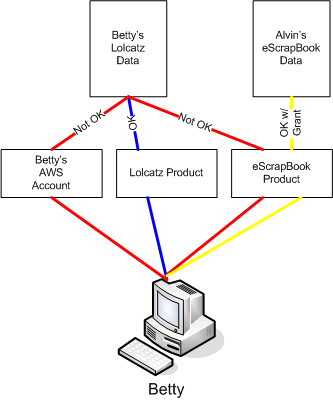
Kết quả tạo rule thành công.

**Amazon DevPay**

Là phương thức thanh toán online đơn giản do Amazon cung cấp xử dụng phương thức thanh toán của Amazon Payment Method. Một dịch vụ quản lý tài khoản người dùng và thanh toán online giúp cho việc thuận tiện cho việc kinh doanh, buôn bán các ứng dụng ở tầng trên nền dịch vụ Web Amazon. Giúp developer cài đặt dễ dàng hơn.

Bạn không cần phải mắc công code quản lý hóa đơn, thanh toán, phương thức chi trả Mà dùng chính phương thức thanh toán của Amazon thông qua dịch vụ tài khoản Amazon của người dùng. Amazon sẽ tự động tính toán thanh toán và chuyển khoản. Giúp cho việc quản lý về khoản này được giảm chi phí khá nhiều.

[**Amazon S3 Customer Data Isolation**](http://docs.aws.amazon.com/AmazonS3/latest/dev/DevPayDataIsolation.html)**:**

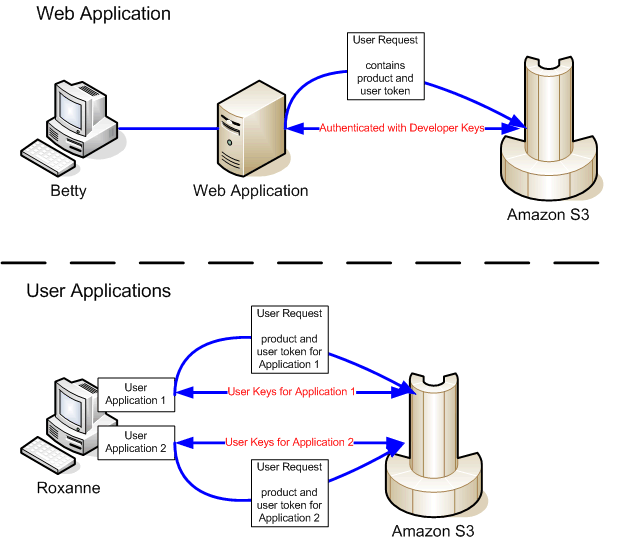


DevPay còn giúp phân vùng và bảo mật dữ liệu. Khách hàng lưu trữ data trên AWS3 thông qua sản phẩm của bạn thì chỉ được truy cập dữ liệu thông qua sản phẩm của bạn. Không thể truy cập qua bất kỳ tài khoản amazon nào hay sản phẩm Amazon nào khác.

Betty có thể truy cập dữ liệu Lolcatz thông qua sản phẩm Lolcatz.

Betty có thể truy cậm dữ liệu eScrapBook của Alvin thông qua sản phẩm eScrapBook nếu được Alvin cấp quyền.

[**Amazon S3 and Amazon DevPay Authentication**](http://docs.aws.amazon.com/AmazonS3/latest/dev/DevPayAuthentication.html):



Phương thức chứng thực giữa Ứng dụng web và ứng dụng người dùng.

Nếu sản phẩm Amazon Devpay là ứng dụng web, thì bạn được lưu trữ Secret Access Key một cách an toàn trên servers của bạn dùng user token để xác dịnh khách hàng cho request yêu cầu đăng thực hiện.

Tuy nhiên nếu ứng dụng AmazonS3 được cài đặt trên máy tính người dùng. Thì ứng dụng của bạn phải sinh ra một Access Key ID và Secret Access Key cho mỗi công việc cài đặt. Và người dùng phải dùng khóa chứng thực này khi giao tiếp với AmazonS3

Amazon dùng token của người dùng và sản phẩm token để xác định phương thức thanh toán lưu lượng sử dụng AmazonS3

Mỗi tháng một lần, Amazon tính toán lưu lượng dữ dữ liệu sử dụng, và thanh toán người dùng dựa trên chính sách bạn đã tạo.

Tiến hành thanh toán và chuyển khoản sau đó.

Phần demo hiện chưa có: do chỉ áp dụng trên tài khoản Amazon lâu năm. Không dành cho tài khoản mới.

# **Demo Host và Deploy Static Web on AmazonS3**

**Hosting**: <https://drive.google.com/file/d/1GZbrPsTH2NHJL_Tanf18BcffLD1h1Xmu/view?usp=sharing>

**Deploy**: <https://drive.google.com/file/d/1ovqoCwFp-gNGBJOB2PxmW6w_QnDaIgF3/view?usp=sharing>

# **Nguồn tham khảo.**

* <https://gsviec.com/blog/amazon-s3-la-gi-va-tai-sao-ban-nen-dung-no>
* <https://dinhnn.com/2014/11/18/bat-dau-voi-amazon-simple-storage-service-amazon-s3/>
* <http://docs.aws.amazon.com/AmazonS3/latest/dev/Welcome.html>

**-Hết-**