AWS Weekly Roundup: Giảm Giá Amazon S3 Express One Zone, Pixtral Large trên Amazon Bedrock, Amazon Nova Sonic và Nhiều Hơn Nữa (14 tháng 4 năm 2025)

**Tác giả: Elizabeth Fuentes**Ngày: 14 tháng 4 năm 2025  
Danh mục: Amazon Aurora, Amazon Bedrock, Amazon Bedrock Agents, Amazon Bedrock Guardrails, Amazon Bedrock Prompt Management, Amazon CloudWatch, Amazon Nova, Amazon Personalize, Amazon SageMaker, Announcements, AWS Step Functions, Launch, News, Storage

# Giới Thiệu

Mùa AWS Summit 2025 đã được khởi động trong tuần này, bắt đầu với Paris Summit. Những sự kiện miễn phí này tập hợp cộng đồng điện toán đám mây toàn cầu để học hỏi và hợp tác. AWS Community Day Romania, được tổ chức vào ngày 11 tháng 4, đã thể hiện cách cộng đồng địa phương tạo ra cơ hội cho sự phát triển tập thể và hòa nhập.



# Các Ra Mắt Tuần Trước

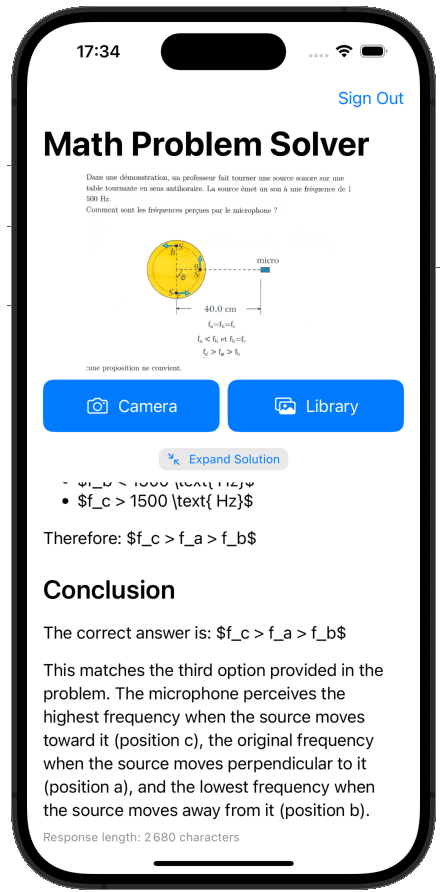
## Thông Báo Giảm Giá Lên Đến 85% cho Amazon S3 Express One Zone

S3 Express One Zone, một lớp lưu trữ hiệu suất cao, hiện đã giảm giá lưu trữ 31%, giá yêu cầu PUT 55% và giá yêu cầu GET 85%. Ngoài ra, S3 Express One Zone đã giảm phí per-GB cho việc tải lên và truy xuất dữ liệu 60%. Những phí này hiện áp dụng cho tất cả byte được truyền thay vì chỉ các phần của yêu cầu lớn hơn 512 KB.  
  
Dưới đây là bảng giảm giá trong AWS Region US East (N. Virginia):

|  |  |  |  |
| --- | --- | --- | --- |
| Giá | Trước đây | Mới | Giảm giá |
| Lưu trữ (per GB-Tháng) | $0.16 | $0.11 | 31% |
| Ghi (PUT requests) | $0.0025 per 1,000 requests lên đến 512 KB | $0.00113 per 1,000 requests | 55% |
| Đọc (GET requests) | $0.0002 per 1,000 requests lên đến 512 KB | $0.00003 per 1,000 requests | 85% |
| Tải lên dữ liệu (per GB) | $0.008 | $0.0032 | 60% |
| Truy xuất dữ liệu (per GB) | $0.0015 | $0.0006 | 60% |

## AWS Thông Báo Mô Hình Pixtral Large 25.02 trong Amazon Bedrock Serverless

Pixtral Large 25.02, được phát triển bởi Mistral AI, kết hợp khả năng hiểu biết thị giác và ngôn ngữ tiên tiến, tự hào với cửa sổ ngữ cảnh 128K và khả năng đa ngôn ngữ. Thiết kế tập trung vào agent này đơn giản hóa việc tích hợp với các hệ thống hiện có. Việc tuân thủ prompt cải thiện độ tin cậy khi làm việc với các ứng dụng Retrieval Augmented Generation (RAG) và các tình huống ngữ cảnh lớn.



## Giới Thiệu Amazon Nova Sonic: Cuộc Trò Chuyện Bằng Giọng Nói Giống Con Người cho Ứng Dụng AI Tạo Sinh

Amazon Nova Sonic, bổ sung mới nhất cho gia đình các mô hình nền tảng (FMs) Amazon Nova, có sẵn trong Amazon Bedrock để tạo ra các cuộc trò chuyện bằng giọng nói giống con người cho các ứng dụng. Nó thống nhất xử lý giọng nói và văn bản thành một mô hình, giảm độ phức tạp và tăng cường tương tác tự nhiên. Bắt đầu ngay hôm nay với kho lưu trữ cookbook mô hình Amazon Nova.

## Amazon Bedrock Guardrails Tăng Cường An Toàn Ứng Dụng AI Tạo Sinh với Các Khả Năng Mới

Amazon Bedrock Guardrails giới thiệu các khả năng mới để tăng cường an toàn ứng dụng AI tạo sinh, bao gồm phát hiện độc tính đa phương thức, bảo vệ Thông Tin Nhận Dạng Cá Nhân (PII) được cải thiện, thực thi chính sách AWS Identity and Access Management (AWS IAM), ứng dụng guardrail có chọn lọc và chế độ giám sát cho phân tích trước triển khai.

## AWS App Studio Giới Thiệu Catalog Giải Pháp Được Xây Dựng Sẵn và Import/Export Liên Instance

Đây là catalog giải pháp được xây dựng sẵn với các ứng dụng và mẫu sẵn sàng sử dụng cùng chức năng Import và Export liên instance. Những tính năng này giúp bạn hợp lý hóa các ứng dụng phát triển, giảm thời gian thiết lập xuống dưới 15 phút.

## Amazon Nova Reel 1.1: Có Tính Năng Video Multi-shot Lên Đến 2 Phút

Amazon Nova Reel 1.1 tăng cường tạo video thông qua Amazon Bedrock với hỗ trợ video multi-shot 2 phút. Bây giờ bạn có thể tạo nội dung bằng cách sử dụng prompt đơn cho tạo tự động hoặc prompt tùy chỉnh cho từng cảnh riêng lẻ, cung cấp các tùy chọn linh hoạt cho việc tạo nội dung marketing và truyền thông xã hội.

## AWS IAM Identity Center Hiện Cung Cấp Thông Báo Lỗi Được Cải Thiện và AWS CloudTrail Logging cho Các Vấn Đề Provisioning

AWS Identity and Access Management (IAM) Identity Center đã nâng cao dịch vụ của mình với thông báo lỗi được cải thiện và khả năng AWS CloudTrail logging. Những cập nhật này giúp người dùng khắc phục sự cố đồng bộ hóa tốt hơn khi quản lý danh tính lực lượng lao động trên các tài khoản và ứng dụng AWS, đồng thời cho phép giám sát và kiểm toán tự động các vấn đề provisioning.

# Các Bài Viết Blog AWS Khác

## Giảm Chi Phí Đào Tạo ML với Amazon SageMaker HyperPod

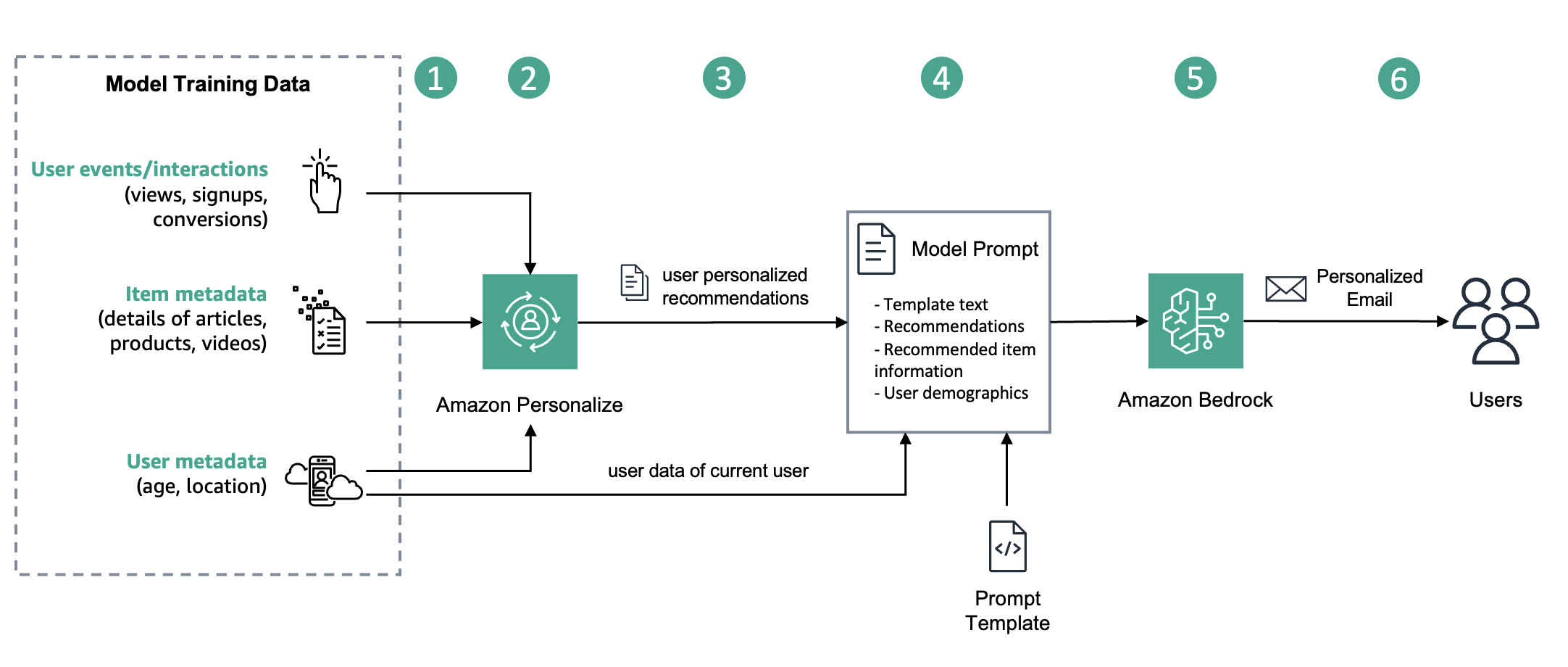
Amazon SageMaker HyperPod giải quyết các lỗi phần cứng trong đào tạo mô hình Machine Learning (ML) quy mô lớn bằng cách tự động phát hiện và thay thế các instance bị lỗi. Giải pháp này giảm thời gian ngừng hoạt động từ 280 xuống 40 phút mỗi lỗi, có khả năng tiết kiệm 32% thời gian đào tạo cho các cluster lớn. Đối với công việc đào tạo 10 triệu GPU-giờ, điều này có nghĩa là tiết kiệm chi phí 25,6 triệu đô la.

## Tùy Chỉnh Mô Hình, RAG, hoặc Cả Hai: Nghiên Cứu Trường Hợp với Amazon Nova

Một nghiên cứu so sánh tùy chỉnh mô hình với fine-tuning và các phương pháp Retrieval Augmented Generation (RAG) với các mô hình Amazon Nova. Những phát hiện chính cho thấy việc kết hợp cả hai phương pháp mang lại kết quả tốt nhất: RAG hoạt động tốt cho dữ liệu động và thông tin miền, trong khi fine-tuning xuất sắc trong các tác vụ chuyên biệt và giảm độ trễ.

## Tạo Giao Tiếp Được Cá Nhân Hóa với Amazon Personalize và Amazon Bedrock

Amazon Personalize và Amazon Bedrock làm việc cùng nhau để tạo ra các email marketing được cá nhân hóa. Tìm hiểu cách tạo giao tiếp người dùng được cá nhân hóa bằng cách kết hợp Amazon Personalize cho đề xuất phim với Amazon Bedrock để tạo nội dung email phù hợp dựa trên sở thích và nhân khẩu học của người dùng.



# Bài Viết Yêu Thích Từ Cộng Đồng

## Xây Dựng Hệ Thống RAG cho Tìm Kiếm và Phân Tích Nội Dung Video

Trong blog này, tôi sẽ chỉ cho bạn cách xây dựng hệ thống RAG làm cho nội dung video có thể tìm kiếm và phân tích được. Mở khóa nội dung video chưa bao giờ quan trọng hơn trong bối cảnh kỹ thuật số ngày nay.

## Xây Dựng Ứng Dụng GenAI Serverless Nhanh Hơn với Amazon Q Developer CLI Agent

Amazon Q Developer CLI Agent cho phép phát triển ứng dụng GenAI serverless nhanh chóng. Với một prompt, nó tạo ra mã cơ sở hạ tầng, hàm Lambda và tích hợp với Claude 3 Haiku trên Amazon Bedrock.

## AI Speech-to-Speech: Từ Dr. Sbaitso đến Amazon Nova Sonic

Sự tiến hóa của AI speech-to-speech, từ Dr. Sbaitso (những năm 1990) đến Amazon Nova Sonic. Dịch vụ AWS mới cho phép các cuộc trò chuyện hai chiều thời gian thực thông qua Amazon Bedrock cho các ứng dụng tự nhiên hơn.

# Các Sự Kiện AWS Sắp Tới

Kiểm tra lịch của bạn và đăng ký cho các sự kiện AWS sắp tới:  
  
• AWS GenAI Lofts — GenAI Lofts có sẵn trên khắp thế giới, cung cấp không gian hợp tác và trải nghiệm nhập vai cho các startup và nhà phát triển.  
  
• AWS Summits — Tham gia các sự kiện trực tuyến và trực tiếp miễn phí tập hợp cộng đồng điện toán đám mây để kết nối, hợp tác và học về AWS. Đăng ký tại thành phố gần nhất: Amsterdam (16 tháng 4), London (30 tháng 4) và Poland (5 tháng 5).  
  
• AWS re:Inforce — AWS re:Inforce (16-18 tháng 6) tại Philadelphia, PA, là sự kiện học tập hàng năm của chúng tôi dành riêng cho mọi thứ về bảo mật đám mây AWS.  
  
• AWS Community Days — Tham gia các hội nghị do cộng đồng dẫn dắt với các cuộc thảo luận kỹ thuật, workshop và phòng thí nghiệm thực hành.

# Kết Luận

Tạo AWS Builder ID của bạn và đặt trước alias của bạn. Builder ID là thông tin đăng nhập phổ quát cung cấp cho bạn quyền truy cập—ngoài AWS Management Console—vào các công cụ và tài nguyên AWS, bao gồm hơn 600 khóa đào tạo miễn phí, tính năng cộng đồng và các công cụ phát triển như Amazon Q Developer.  
  
Đó là tất cả cho tuần này. Hãy theo dõi Weekly Roundup tuần tới!  
  
— Eli  
  
Cảm ơn Andra Somesan cho ảnh AWS Community Romania và Thembile Martis cho ảnh AWS Paris Summit.  
  
Bài viết này là một phần của loạt Weekly Roundup của chúng tôi. Hãy quay lại mỗi tuần để có bản tóm tắt nhanh về tin tức và thông báo thú vị từ AWS!

# Thông Tin Tác Giả



Elizabeth Fuentes  
  
Sứ mệnh của tôi là phân tích các khái niệm phức tạp thành những giải thích dễ hiểu, truyền cảm hứng cho các nhà phát triển liên tục mở rộng kỹ năng và kiến thức của họ. Thông qua các hội nghị, hướng dẫn và tài nguyên trực tuyến, tôi chia sẻ chuyên môn của mình với cộng đồng nhà phát triển toàn cầu, cung cấp cho họ các công cụ và sự tự tin để đạt được tiềm năng đầy đủ của họ. Với cách tiếp cận thực hành và cam kết đơn giản hóa sự phức tạp, tôi cố gắng trở thành chất xúc tác cho sự phát triển và học tập trong thế giới công nghệ AWS.