**Tiêu đề: Cách Aqua Security tự động hóa việc điều phối fast clone trên Amazon Aurora ở quy mô lớn**

Đây là bài viết khách mời từ Semyon Mor – Quản lý Kỹ thuật SRE tại Aqua Security, và Daniel Vidal – Kiến trúc sư Đám mây tại Aqua Security, phối hợp cùng AWS.

[Aqua Security](https://www.aquasec.com/) là nhà cung cấp hàng đầu về các giải pháp bảo mật dựa trên đám mây, được các doanh nghiệp toàn cầu tin tưởng để bảo vệ ứng dụng của họ từ giai đoạn phát triển cho đến khi vận hành chính thức. Là một phần trong sứ mệnh cung cấp thông tin bảo mật nhanh chóng, đáng tin cậy và ở quy mô lớn, Aqua phụ thuộc rất nhiều vào các khối công việc phân tích dựa trên dữ liệu – trong đó có những truy vấn chạy lâu dài và các lần xuất truy vấn từ các cụm [Amazon Aurora PostgreSQL-Compatible Edition](https://aws.amazon.com/rds/aurora/postgresql-features/)

Là một tổ chức lấy khách hàng làm trung tâm, Aqua luôn cam kết mang đến các giải pháp sáng tạo nhằm giải quyết những nhu cầu bảo mật không ngừng thay đổi của khách hàng. Giống như nhiều tổ chức khác, Aqua đối mặt với thách thức trong việc xuất và phân tích khối lượng dữ liệu lớn một cách hiệu quả nhằm đáp ứng các yêu cầu kinh doanh. Cụ thể, Aqua cần xuất dữ liệu truy vấn ở quy mô lớn và chia sẻ với khách hàng để phục vụ việc giám sát liên tục và phân tích bảo mật.

Để làm được điều này, họ đã sử dụng hàm [aws\_s3.query\_export\_to\_s3](https://docs.aws.amazon.com/AmazonRDS/latest/AuroraUserGuide/postgresql-s3-export-functions.html#aws_s3.export_query_to_s3) kết hợp cùng AWS Step Functions nhằm tự động hóa quá trình xuất kết quả truy vấn từ các cụm Aurora PostgreSQL của mình sang Amazon S3. (Để xem hướng dẫn chi tiết về quy trình xuất dữ liệu này, bạn có thể tham khảo bài viết **“How Aqua Security exports query data from Amazon Aurora to deliver value to their customers at scale”**.)

Mặc dù giải pháp này đã giải quyết được nhu cầu cung cấp dữ liệu, nhưng nó lại tạo ra một thách thức mới khi một truy vấn xuất dữ liệu dài và phức tạp trên bản sao (replica) cố gắng đọc một trang dữ liệu đã bị writer chỉnh sửa. Nếu trang đó đang chờ được cập nhật mà đồng thời lại bị truy vấn đọc truy cập, thì xung đột truy cập sẽ xảy ra, dẫn đến hiện tượng trễ trong quá trình sao chép (replication lag) và thậm chí có thể gây lỗi (xem phần giải thích chi tiết trong mục tiếp theo).

Để giải quyết vấn đề này, Aqua Security đã áp dụng tính năng [**fast cloning**](https://docs.aws.amazon.com/AmazonRDS/latest/AuroraUserGuide/Aurora.Managing.Clone.html)trong Aurora PostgreSQL (tính năng này cũng có trong Aurora phiên bản tương thích MySQL), một tính năng cho phép sao chép cụm nhanh hơn và tiết kiệm chi phí hơn so với cách [Aurora backup and restore](https://docs.aws.amazon.com/AmazonRDS/latest/AuroraUserGuide/Aurora.Managing.Backups.html). Bằng cách chuyển khối lượng công việc xuất dữ liệu sang các môi trường đã được clone, Aqua đảm bảo rằng hiệu suất và độ khả dụng của các hệ thống quan trọng trong hoạt động của họ không bị ảnh hưởng.

Trong bài viết này, chúng tôi sẽ khám phá cách Aqua Security tự động hóa việc sử dụng các bản sao fast clone để hỗ trợ các hoạt động đọc dữ liệu nặng ở quy mô lớn, đơn giản hóa các luồng xử lý dữ liệu, và duy trì hiệu quả vận hành tổng thể.

**Tổng quan giải pháp**

Đối với các trường hợp sử dụng mà việc xuất truy vấn chỉ diễn ra thỉnh thoảng – ví dụ như một lần mỗi ngày, giống như môi trường của Aqua Security – và khi không cần đảm bảo tính nhất quán mạnh với cơ sở dữ liệu chính, **Aurora fast clones** mang lại một giải pháp thực tiễn và tiết kiệm chi phí.

Vì fast clone hoạt động như các cụm Aurora độc lập, chúng hoàn toàn cách ly với môi trường chính. Điều này đồng nghĩa với việc các truy vấn xuất dữ liệu chạy lâu có thể thực hiện mà không lo bị hủy do giới hạn timeout max\_standby\_streaming\_delay.

Fast clone cũng mang lại hiệu quả cao về chi phí: vì các bản sao dùng chung lớp lưu trữ bên dưới với cụm gốc, nên **không phát sinh thêm chi phí lưu trữ** trừ khi dữ liệu bị chỉnh sửa. Chi phí duy nhất là chi phí của các instance Aurora được dùng để chạy cụm clone, được tính theo **giá tiêu chuẩn trong suốt thời gian cụm clone hoạt động**. Sau khi quá trình xuất dữ liệu hoàn tất, cụm clone có thể được **dừng và xóa an toàn**, từ đó **ngừng phát sinh chi phí tính toán**.

Để tận dụng tối đa fast clone, Aqua Security đã phát triển một **hệ thống điều phối tự động** giúp quản lý linh hoạt vòng đời của các cơ sở dữ liệu được clone. Hệ thống này được xây dựng dựa trên ba nguyên tắc cốt lõi:

* **Clone tự động** – Khi một truy vấn dài hoặc tác vụ xuất dữ liệu được kích hoạt, một bản sao nhanh của cơ sở dữ liệu nguồn sẽ được tạo ra một cách tự động
* **Thực thi cách ly** – Tác vụ sẽ chạy trên bản sao, đảm bảo rằng các bản sao trong môi trường sản xuất không bị ảnh hưởng bởi khóa truy vấn, mức sử dụng bộ nhớ hay tải I/O
* **Dọn dẹp tự động** – Sau khi tác vụ hoàn tất, cụm clone sẽ được xóa để giải phóng tài nguyên và giữ cho môi trường gọn nhẹ

Sơ đồ sau đây minh họa kiến trúc của giải pháp này.

A diagram of a software application

AI-generated content may be incorrect.

**Kiến trúc giải pháp của Aqua bao gồm các thành phần sau:**

**1. Bộ lập lịch công việc (Job Scheduler):**

* Bộ lập lịch kích hoạt luồng công việc (sử dụng [**AWS Step Functions**](https://aws.amazon.com/step-functions/)) dựa trên lịch trình đã định sẵn (dựa theo thời gian) thông qua [**Amazon EventBridge**](https://aws.amazon.com/eventbridge/), hoặc cũng có thể khởi động luồng công việc **theo yêu cầu**, giúp linh hoạt hơn tùy vào nhu cầu vận hành.

**2. Dịch vụ điều phối (Orchestration Service):**

* Được xây dựng sử dụng **AWS SDK** và các **API của Aurora**, được gọi thông qua **AWS Lambda**
* **Khởi tạo fast clone** của Aurora thông qua các thao tác CreateDBCluster và RestoreDBClusterToPointInTime, cho phép tạo nhanh môi trường cơ sở dữ liệu cách ly
* Quản lý cấu hình cụm Aurora được clone: thiết lập quyền truy cập endpoint riêng, áp dụng **security group phù hợp**, và thiết lập các **tham số cơ sở dữ liệu**. Những cấu hình này đảm bảo các bản sao vừa an toàn, vừa tối ưu hiệu năng cho các khối lượng công việc phân tích
* **Giám sát liên tục** trạng thái bản sao (polling) cho đến khi bản sao sẵn sàng để phục vụ các truy vấn phân tích
* **Tự động xóa cụm Aurora được clone** sau khi quá trình phân tích hoàn tất, hoặc sau khoảng thời gian TTL (Time-to-Live) được cấu hình sẵn để tối ưu hóa việc sử dụng tài nguyên và kiểm soát chi phí một cách hiệu quả

**3. Tích hợp với công cụ nội bộ của Aqua:**

* Chuyển hướng các truy vấn phân tích trực tiếp đến các endpoint của cơ sở dữ liệu Aurora được clone, **cách ly các khối lượng truy vấn nặng (đọc nhiều)** khỏi cơ sở dữ liệu sản xuất chính
* Giao tiếp kết quả thực thi công việc với lớp điều phối thông qua các **callback định nghĩa sẵn** (SendTaskSuccess hoặc SendTaskFailure), hỗ trợ việc giám sát, ghi log và xử lý lỗi hiệu quả
* Đảm bảo các công việc phân tích **chỉ đọc dữ liệu từ bản sao Aurora**, trong khi **ghi kết quả và đầu ra phân tích trở lại cụm Aurora chính**, từ đó đảm bảo tính toàn vẹn và nhất quán dữ liệu

**Cách tiếp cận này giúp Aqua tách biệt hoàn toàn các truy vấn đọc lâu khỏi hạ tầng sản xuất chính**, cải thiện cả về **độ ổn định và khả năng mở rộng** theo cách tiết kiệm chi phí.

**Kết quả và tác động đến hoạt động kinh doanh**

Khi áp dụng điều phối fast clone trên Aurora PostgreSQL, **Aqua Security đã đạt được nhiều kết quả ấn tượng**:

* ✅ **Tăng độ tin cậy:** Các truy vấn báo cáo dài và công việc xuất dữ liệu không còn bị lỗi vì giờ đây chạy trên các bản sao Aurora mà **không ảnh hưởng đến môi trường sản xuất**
* 📈 **Khả năng mở rộng:** Hệ thống có thể **mở rộng theo nhu cầu công việc** của Aqua, hỗ trợ các lần xuất dữ liệu một lần, báo cáo phức tạp, và cả thử nghiệm của nhà phát triển mà không gây ảnh hưởng đến hệ thống chính. Trung bình Aqua chạy khoảng **280 bản sao mỗi ngày**
* 💰 **Hiệu quả chi phí:** Bản sao Aurora **tiết kiệm chi phí** vì dùng chung lớp lưu trữ với cụm nguồn (chỉ tốn thêm chi phí lưu trữ nếu dữ liệu bị thay đổi), và chỉ tính phí điện toán khi bản sao hoạt động. **Chi phí tổng cộng mỗi tháng vào khoảng 2.100 USD**, tức là chỉ **khoảng 7,50 USD mỗi bản sao cơ sở dữ liệu**

**Kết luận**

Việc Aqua Security áp dụng điều phối fast clone trên Aurora thể hiện sức mạnh của **kiến trúc lưu trữ Aurora** trong việc giải quyết các thách thức vận hành thực tế. Thông qua tự động hóa quy trình **tạo và xóa cơ sở dữ liệu được clone**, Aqua đã xây dựng một nền tảng **linh hoạt và tiết kiệm chi phí** để hỗ trợ các khối lượng công việc dữ liệu yêu cầu cao — **mà không ảnh hưởng đến hiệu suất hay tính khả dụng** của hệ thống.

Nếu bạn đang chạy các truy vấn báo cáo phức tạp trên Aurora và muốn tránh gây ảnh hưởng đến môi trường sản xuất, hãy cân nhắc sử dụng [**Aurora fast clones**](https://docs.aws.amazon.com/AmazonRDS/latest/AuroraUserGuide/Aurora.Managing.Clone.html), và lấy cảm hứng từ **cách tiếp cận điều phối của Aqua Security**.

Để tìm hiểu thêm về Aqua Security, hãy truy cập [**aquasec.com**](https://aquasec.com).

**Thông tin về tác giả**

****

**Semyon Mor**

Semyon là **Quản lý Kỹ thuật SRE** tại Aqua Security, chịu trách nhiệm cho nền tảng Aqua SaaS. Với hơn **một thập kỷ kinh nghiệm**, anh đã xây dựng và mở rộng hạ tầng đám mây, đồng thời lãnh đạo các nhóm DevOps và SRE trong cả môi trường startup lẫn doanh nghiệp lớn. Công việc của anh tập trung vào việc cung cấp các hệ thống **bảo mật, linh hoạt và có khả năng mở rộng cao**, hỗ trợ giải pháp bảo mật dựa trên đám mây cho một số tổ chức lớn nhất thế giới.

**Daniel Vidal**

Daniel là **Kiến trúc sư Đám mây** tại Aqua Security với hơn **10 năm kinh nghiệm** trong ngành công nghệ, chuyên sâu về **DevOps, SRE và hạ tầng đám mây**. Trong vai trò hiện tại, anh tập trung vào việc xây dựng các môi trường đám mây an toàn, có khả năng mở rộng cao và tư vấn các **thực tiễn tốt nhất trong lĩnh vực bảo mật đám mây** cho ngành an ninh mạng.

**Pini Dibask**

[**Pini**](https://www.linkedin.com/in/pinidibask/)là **Kiến trúc sư Giải pháp Cơ sở dữ liệu Cấp cao** tại AWS với **20 năm kinh nghiệm** làm việc với các công nghệ cơ sở dữ liệu, đặc biệt là cơ sở dữ liệu quan hệ. Trong vai trò của mình, Pini làm việc với các **khách hàng lớn nhất tại Israel** với tư cách là **cố vấn đáng tin cậy** về kiến trúc cơ sở dữ liệu AWS, thực tiễn tốt nhất, và chiến lược di chuyển hệ thống lên đám mây.