

AWS Cloud architecture for the Tech Solutions API

When there is heavy traffic, there may be a requirement for additional instances beyond the ones initially planned. This is essential to ensure the application can scale, maintain availability, and deliver optimal performance. To achieve this, the application will be hosted on EC2 instances within an autoscaling group, allowing dynamic scaling to match demand. A load balancer will evenly distribute traffic across these instances. In case any instance encounters health issues, it will be promptly removed from the pool, and a new one will be added using a predefined launch template. This setup ensures the application can effectively handle varying loads and maintain its efficiency. Please see diagram for visual interpretation of above statement:

