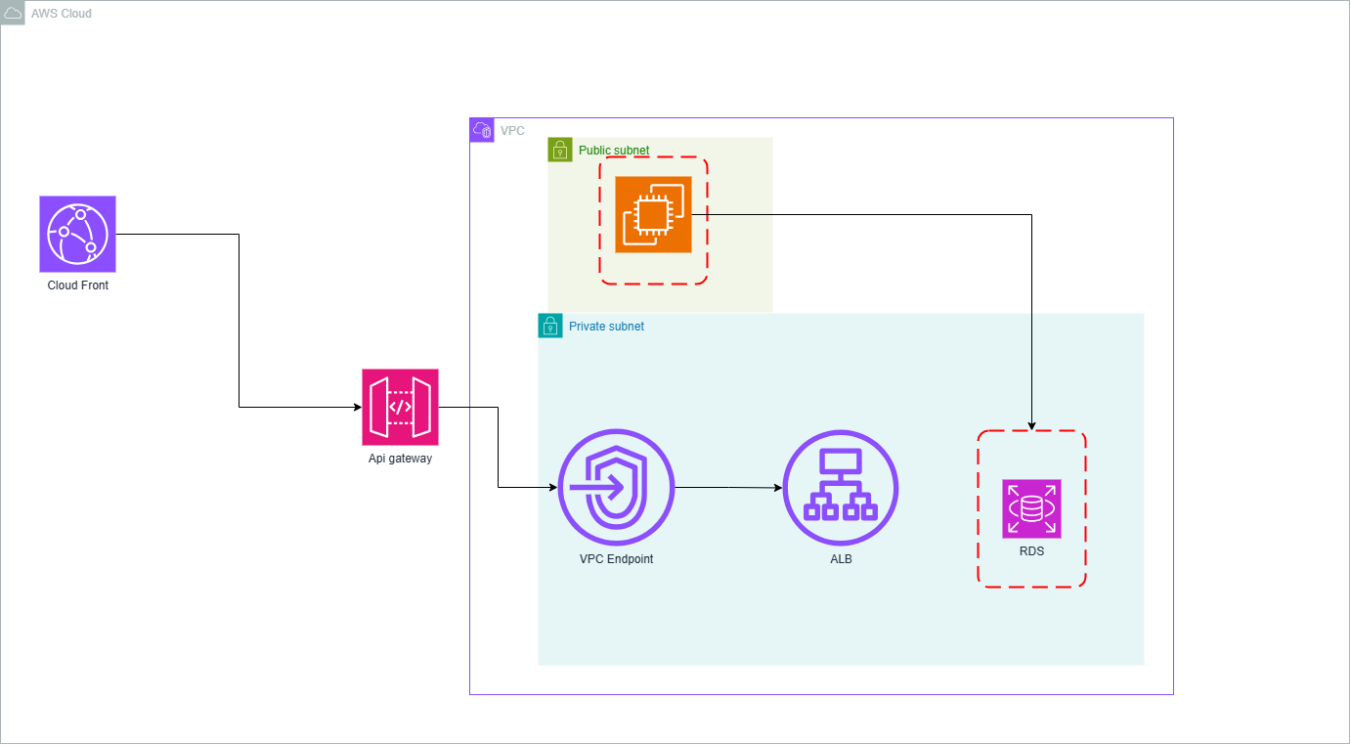
Documentation

Architecture



* 1. The design choices to ensure performance for users in low-bandwidth areas.

1. To use a Cloudfront to provide caché and low latency in communication
   1. The approach to secure the ML processing layer while keeping it accessible to the app.

A very typical deployment architecture for smaller startups is to have an API Gateway at the front, which passes on requests to an ELB, which in turn distributes them to a bunch of EC2 instances. ELBs and EC2s are typically inside a VPC. For an API Gateway to use an ELB as the HTTP endpoint for integration, the ELB needs to be exposed to the internet. This implies that an API request can theoretically be made directly to the ELB, bypassing all the rules configured on the API Gateway. To prevent the previous issue, You can allow only requests originating from the API Gateway on your app by configuring client certificates, but that does not prevent flooding style DDOS attacks on the ELB (and the EC2s). An ideal solution is to create a link between the API Gateway and the ELB which is NOT exposed to the internet.