

* Elastic Load Balancer 란?

- AWS 의 L4/ L7 로드밸런싱(서버 부하분산) 서비스
- 외부 혹은 내부에서 들어오는 클라이언트의 요청을 ELB 에 연결된 가용영역에 있는 EC2 로 고르게 나누어 전달하고 등록된 EC2 들을 상태 검사하는 서비스
- 외부/내부 트래픽은 ELB 의 DNS 주소를 목적으로 하여 들어오며 이 트래픽을 받은 ELB 는 가용영역에 등록된 EC2 로 전달
- 보안그룹 적용
- ELB 또한 VPC 내 존재하며 공인 IP 혹은 공인/사설 IP 를 모두 가질 수 있음
 - 인터넷 연결 option : 공인/사설 IP 생성
 - 내부 option : 사설 IP 생성
- 기본적으로 Port 를 이용하여 로드밸런싱
(예를 들어 HTTP 요청을 로드밸런싱하는 경우, 그 ELB 는 HTTP 요청만을 전달하고 다른 포트 요청은 전달하지 않음)
- ELB 는 '리스너'를 거느리며 리스너 하나당 요청을 받아들이는 Port 하나씩을 지정함
 - 예를 들어, 80 Port 요청을 받아 들이는 리스너, 443 Port 요청을 받아들이는 리스너를 모두 등록할 수 있음
- 상태 검사에 성공한 EC2 만이 요청 전달 대상이 되며 상태 검사에 실패한 EC2 는 요청 전달 대상에 제외됨
- SSL 인증서를 등록하여 HTTPS 암호화/복호화 통신을 대신하는 SSL Offload 가능
- ELB 는 세 가지 로드밸런서로 구성됨
 - ALB(Application Load Balancer)
 - NLB(Network Load Balancer)
 - CLB(Classic Load Balancer)

* Target Group(대상그룹)

- ELB 에 등록된 EC2 인스턴스들의 그룹
- 대상 그룹에 등록된 EC2, EC2 상태, 상태 검사 방법 등을 정의
- 상태 검사 방법에는 HTTP, HTTPS, TCP 등이 있음
- 방법뿐만 아니라 상태 검사에 필요한 시간도 정의 가능
- ALB 와 NLB 는 대상그룹을 지정하며, CLB 는 로드밸런서에 직접 등록

* ALB(Application Load Balancer)

- L7 로드밸런서
- HTTP/HTTPS Header 정보를 기반으로 요청을 전달하는 로드밸런서
 - 즉 리스너가 HTTP/HTTPS 를 전문적으로 다룸
- 요청에 따라 고정 페이지를 반환하거나 다른 경로로 리다이렉트
- HTTP Header / HTTP 요청 메서드 / Host Header / Source IP 등을 이용하여 요청 전달 가능
 - 예를 들어 HTTP Header 내 User-agent에는 Host의 OS 정보가 있는데 Android일 경우, ELB 리스너가 Android만 사용하는 대상 그룹으로 전달 가능
- SSL 인증서를 이용하여 SSL Offload 지원
- 후술할 교차영역 로드밸런싱 지원
- X-forwarded-for를 지원하여 EC2로 하여금 클라이언트의 IP를 확인할 수 있도록 함
- ALB가 소유한 공인 IP가 유동적으로 변경됨

* NLB(Network Load Balancer)

- L4 로드밸런서
- TCP / UDP를 기반으로 요청을 전달하는 로드밸런서
- 프로토콜, 원본 IP 주소, 원본 포트, 대상 IP 주소, 대상 포트, TCP 시퀀스 번호를 이용하여 요청 전달
- TCP, UDP, TCP/UDP로 리스너를 설정 가능
- TCP Header를 조작하여 전달하는 것이 아닌 그저 TCP의 기본인 3-way handshake만을 대신하여 진행
- Client IP를 변경하지 않고 서버에게 전달
- NLB의 IP가 ALB와 달리 변경되지 않고 고정됨
- 교차 영역 로드밸런싱 지원

* CLB(Classic Load Balancer)

- L4/L7을 모두 지원하는 '과거의' 로드밸런서
- EC2-Classic을 사용하는 경우 CLB를 사용하면 됨

- TCP / HTTP / HTTPS(SSL)을 지원
- X-forwarded-for 지원
- 교차영역 로드밸런싱 지원

* Sticky Session

- 세션 고정 기능
- 하나의 요청이 특정 EC2 로 들어와 세션이 생성되었다가 종료되고, 들어왔던 요청이 다시 들어올 경우 이미 연결되었던 특정 EC2 로 전달하는 기능

* 교차영역 로드밸런싱

- 기본적으로 ELB 는 대상그룹에 등록된 EC2 별로 적절히 부하분산을 하는 것이 아닌 가용영역별로 비율을 나눔
 - 즉 교차영역이 2 개라면 각 가용영역에 50%씩 전달
- 이렇게 될 경우, A 가용영역에 EC2 가 2 개이고, B 가용영역에 EC2 가 8 개일 경우 A/B 가용영역에 50%씩 전달되므로 A 가용영역의 EC2 의 부하가 심해짐
- 이를 방지하기 위해 교차영역 로드밸런싱을 활성화하면 가용영역이 아닌 EC2 갯수를 기반으로 계산하여 전달하므로 부하가 고르게 나누어짐

* X-Forwarded-for

- Client 의 IP 를 담고 있는 HTTP Header
- EC2 내 서비스가 Client 의 IP 를 확인할 필요가 있는 경우 ELB 는 HTTP Header 에 X-Forwarded-for 를 장착하여 전달

* 연결 드레이닝

- Autoscaling 과 ELB 를 함께 사용시 필요에 따라 EC2 가 없어질 경우, 해당 EC2 에 요청이 들어와있다면 사용중인 세션이 피해를 볼 수 있음
- 삭제되기 직전에 유효 세션이 작업을 마칠 때까지 기다리는 기능