Bueno comencé creando la estructura de grupos:

1- AWS Cloud

2-Dentro de esta plantee al menos dos AV (usé el Region Group porque no encontré el de AV), la idea de usar 2 garantiza HA.

3-Esas dos AV participan dentro de la misma VPC que nos proporciona un aislamiento seguro, la posibilidad de usar subredes, etc.

4-Una AV se utiliza como el despliegue principal mientras la otra garantiza Alta Disponibilidad mediante Recurrencia.

5-En cada AV habrá una subred Publica para los Servicios Web(Front), una para la Aplicaciónes Web(Back) Y una subred privada con control de acceso para las base de datos tanto relacional(RDS) como no relacional(DynamoDB).\*\*\*

Se garantizan las cargas variables mediante el Auto Scaling, se crearon dos Auto Scalling Group: Uno para los Servicios Web y otro para las Aplicaciones Web.

Junto a esto, ELB se encargará de distribuir el tráfico entre las instancias garantizando escalabilidad y nuevamente disponibilidad.

Se utilizará Route 53 para la gestión del DNS, gestionando solicitudes de manera global distribuida.

Amazon CloudFront será el CDN para distribuir el contenido Front, acelerando la entrega del contenido estático. Junto a este S3 Estándar con sus buckets nos permitirá almacenar cualquiera archivo estático que nos haga falta, utilizamos S3 que si bien es más caro nos ofrece mayor durabilidad y disponibilidad

Por último y un poco a falta de conocimiento sobre cómo funciona en si la parte de Applications decidí que las instancias EC2 del backend mediante Lambda se comuniquen con los microservicios externos y que a su vez lambda tenga acceso a las Base de datos si hicieran falta.

\*\*\*La idea es que existan grupos de seguridad y listas de control de acceso en toda la arquitectura, no sabía bien cómo diagramar eso entonces decidí ponerlo en el acceso a la subred privada siendo una zona crítica.

Otro punto que no sabía bien cómo diagramar sería la parte de ambas bases de datos, pero la idea es que estén conectadas garantizando flexibilidad en la operación de datos, y optimizando rendimiento y escalabilidad