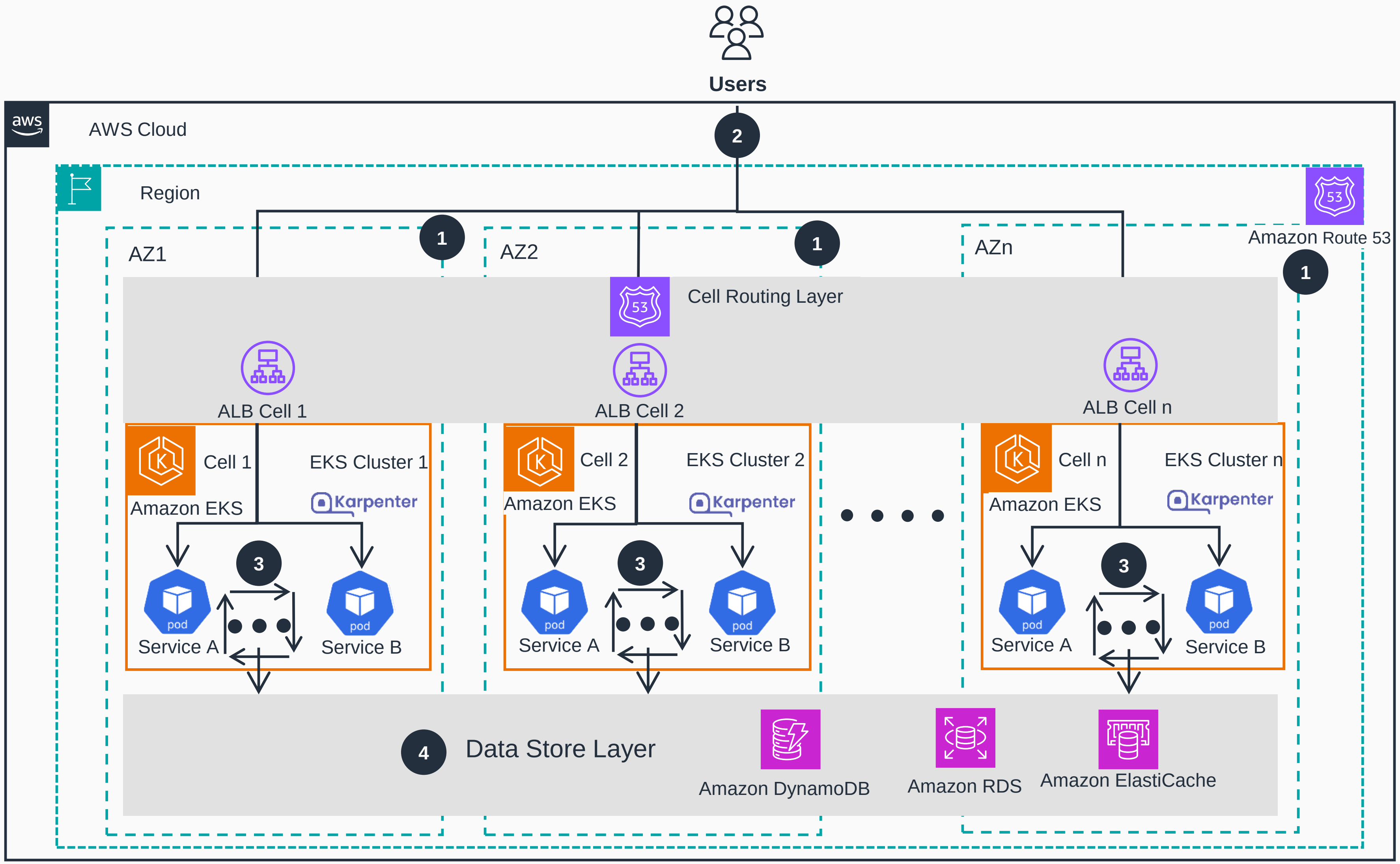
Guidance for a Cell-Based Architecture for Amazon EKS

This architecture diagram shows how you can use a cell-based architecture to improve resiliency and reduce data transfer costs for Amazon EKS workloads. Dedicated Load Balancer per cell ensures session affinity is confined to a single cell; minimizing cross-AZ dependencies



1. A cell consists of an **Amazon Elastic Kubernetes Service** (**Amazon EKS**) cluster having its compute nodes (workloads) and dedicated **Application Load Balancers(ALB)** deployed within a single Availability Zone (AZ). These cells are independent replicas of the application and create a fault isolation boundary to limit the scope of impact. There can be multiple cells per AZ, and multiple cells can be deployed across multiple AZs to provide high availability and resiliency against AZ failures.

2. Clients are routed towards **Amazon EKS** workloads within each cell by a cell-routing layer, which consists of **Amazon Route 53** weighted routing records, and **Amazon Route 53 Application Recovery Controller** to provide readiness checks, routing control, and zonal shifts capability. An **application load balancer** balances the traffic to the Kubernetes resources within each cell.

3. Once the request reaches a cell, all subsequent internal communications among the Kubernetes (k8s) workloads stays within the cell. This prevents cross-cell dependency, making each cell **statically stable** and more resilient. Additionally, with minimal inter-AZ communication, there are no inter-AZ data transfer costs for chatty workloads, as traffic never leaves the AZ boundary. **Amazon EKS** workloads utilize Karpenter for compute autoscaling needs.

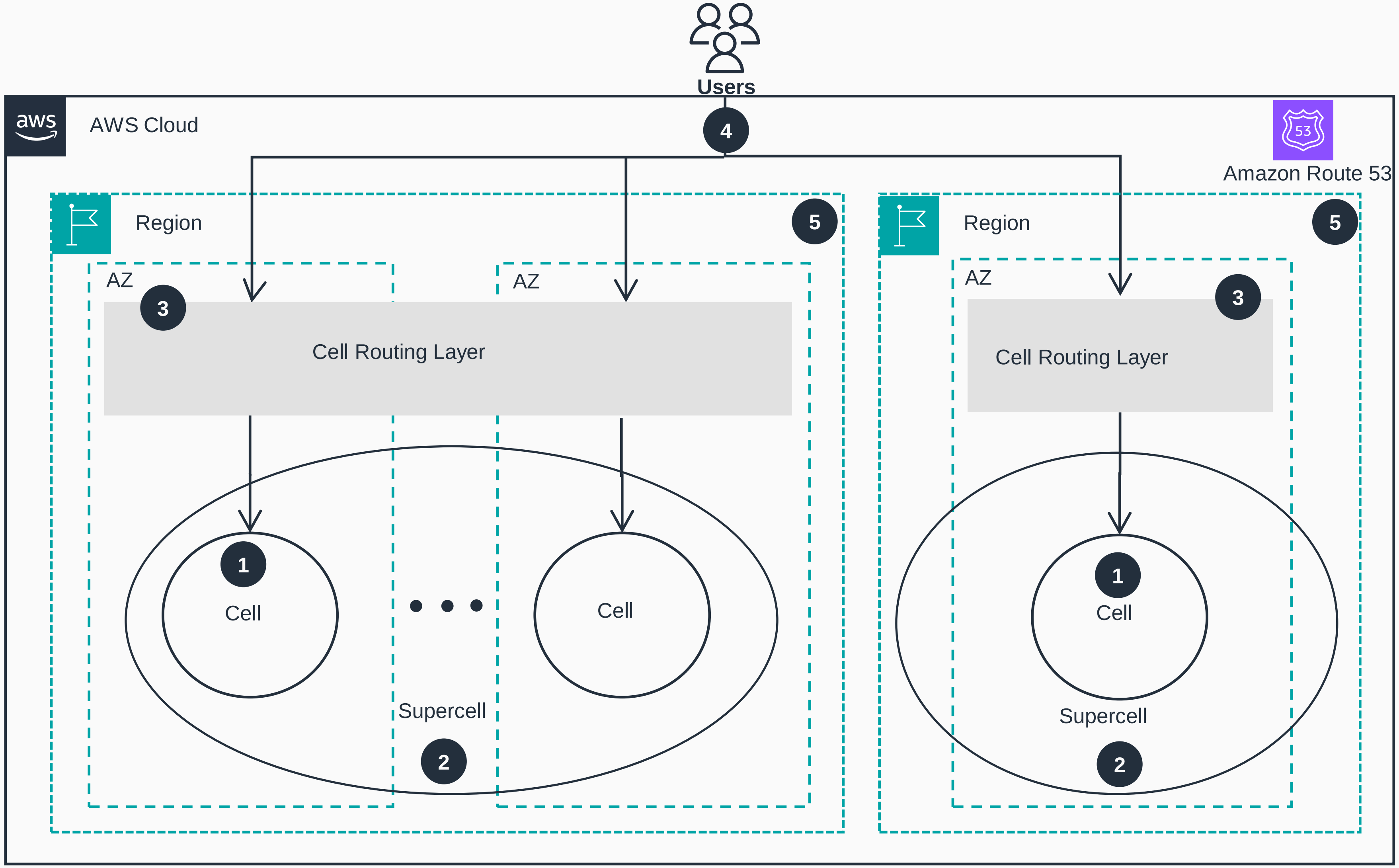
4. **Amazon EKS** workloads that require access to data persistence can continue to use other data store services managed by AWS, like **Amazon Relational Database Service (Amazon RDS)**, **Amazon DynamoDB**, and **Amazon ElastiCache**, which span across multiple AZs for high availability.

Reviewed for technical accuracy August 13, 2025

AWS Reference Architecture

Guidance for a Cell-Based Architecture for Amazon EKS

This architecture diagram shows how multiple cells are aggregated to create a supercell. It also outlines how those supercells are routed.



1 A cell consists of an **Amazon Elastic Kubernetes Service** (**Amazon EKS**) cluster having its compute nodes (workloads) and dedicated **Application Load Balancers(ALB)** deployed within a single Availability Zone (AZ). These cells are independent replicas of the application and create a fault isolation boundary to limit the scope of impact. There can be multiple cells per AZ, and multiple cells can be deployed across multiple AZs to provide high availability and resiliency against AZ failures.

2 An aggregation of multiple cells within a Region is called a supercell.

3 **Amazon EKS** workloads in each AWS Region, or supercell, use **ELB** to load balance the traffic to **Amazon EKS** workloads within each cell.

4 Clients are routed to a supercell using the **Route 53** weighted routing policy and also use the **Route 53** Application Recovery Controller to provide routing control and zonal shift capabilities.

5 Multiple supercells can be deployed across AWS Regions for disaster recovery, or to satisfy data residency requirements.

Reviewed for technical accuracy August 13, 2025

AWS Reference Architecture