

TASK 6 – Cost Optimization

Objective:

The primary objective of this task was to design and implement a cloud architecture that maximizes resource efficiency while minimizing expenditure. This involved leveraging the AWS Free Tier, implementing automated resource management, and setting up financial guardrails to prevent unexpected billing.

1. Select Free-Tier Instance Types

AWS provides **750 hours per month** of specific instances for the first 12 months.

- **x86 (Standard):** Use `t2.micro` or `t3.micro` (depending on your region).
- **Arm-based (Bonus):** Check if `t4g.small` is still on a free trial in your region, it's often faster and more efficient than `t2.micro`.
- **Check "Free Tier Eligible":** When creating your **Launch Template**, always look for the "Free tier eligible" label under the AMI and Instance Type selection.

2. Configure a "Zero-Spend" Budget

This is your safety net. AWS will email you the second you drift out of the Free Tier.

1. Go to **Billing and Cost Management > Budgets**.
2. Click **Create budget**.
3. Select **Use a template (simplified)**.
4. Choose the **Zero spend budget** template.
5. Enter your email address.

This will notify you as soon as your spending exceeds **\$0.01**.

3. Minimal Resource Setup

To keep costs at zero, you need to be careful with "hidden" costs like Load Balancers and EBS volumes.

- **EBS Volumes:** Stay under **30 GB** total across all your instances. When you create your Launch Template, set the volume size to **8 GB** or **10 GB** (the minimum needed for Docker and Nginx).

- **Load Balancer (ALB):** Note that the ALB itself is only free for 750 hours/month if you stay under specific "LCU" (Load Balancer Capacity Unit) limits. Don't run multiple ALBs; one is enough for all your testing.
- **Elastic IPs:** Only use **one**. AWS charges you for Elastic IPs that are *not* attached to a running instance.

4. Optimize Auto Scaling for Cost

Your ASG should be "aggressive" about shrinking when traffic is low.

1. **Group Size:** Set **Minimum: 1** and **Desired: 1**. This ensures that during quiet times, you only pay for (or use free credits for) one single instance.
2. **Scale-In Protection:** Ensure this is **disabled** so the ASG is free to terminate instances when they aren't needed.
3. **Cooldown Period:** Set your "Default Cooldown" to **300 seconds** (5 minutes). This prevents the ASG from launching too many instances at once before the first one has even finished loading.

5. Monitor Your Free Tier Usage

AWS has a specific dashboard just for this.

- Go to **Billing and Cost Management > Free Tier**.
- Look at the **Table** view. It shows exactly how many hours of "EC2-instances" or "ELB" you have used this month.