Even More Architecture Discussion

Event processing in AWS

SQS + Lambda

Lambda service is going to poll from SQS, however, there can be problem with the message if it cannot be processed and go into infinite loops. In that case a dead letter queue can be set up to sent problematic message after certain retries.

SQS FIFO, if a message isn't processed then it is going to be blocking, in that case the dead letter queue can again be used to sent retried message that didn't work into it.

SNS + Lambda

The subscriber will be lambda and messages will be sent to the lambda. If the processing fails on the lambda level then you can set up a SQS dead letter queue for the lambda to sent the message it is unable to process to.

Fan out pattern: Deliver to multiple SQS

Use SNS and make SQS as the subscriber to make it reliable.

S3 events

Use it EventBridge to expand on the capability to react to S3 events.

EventBridge can also intercept API calls, to react to when a specific API occurs.

Caching strategies

Caching at edge: CloudFront, the caching strategies is near the user. However, the cache might not be updated, compared to the backend so you need to set up TTL.

Caching at API Gateway: Caching at API Gateway is doable. It is cached at the region it is deployed in.

Database cache: Redis, DAX, ElastiCache, help save frequently read data so your database isn't overwhelmed.

Lots of way of doing caching in AWS.

Blocking IP address

First line of defense is NACL for the VPC.

Then you have security group for EC2 instances, which only have allow rules. So you just need to allow specific IP range.

Install WAF. To do IP filtering.

If ALB is fronted by CloudFront because CloudFront is outside of VPC, then WAF can do IP filtering on CloudFront and use Geo restriction in CloudFront to restrict traffic that can come from certain countries.

High performance computing on AWS

Cloud is good for HPC, because you can spin up EC2 instances on-demand and just pay for the system you have used.

You can use HPC to do modeling, machine learning, a lot of things!

Services that help to do HPC is:

1. AWS Direct connect: Move data to the cloud via private network.
2. Snowball: Move lots of data to AWS offline.
3. DataSync: Transfer data from on-premise to the cloud.
4. EC2 instances: Can have CPU optimized instances, cost saving via spot instances and spot fleet.
5. Placement group with cluster you can communicate between EC2 instances very fast because the EC2 instances are on the same rack.
6. **Elastic network adapter** to speed up network speed for your EC2 instance
7. **Elastic fabric adapter**, improved ENA for HPC but only works for Linux. Good for inter-node communication.
8. FSx for Lustre: Dedicated for HPC, backed by S3. Only for Linux
9. AWS ParallelCluster: Deploy HPC cluster on AWS. Automate creation of EC2 instances and cluster type. **Use EFA on the clustser is available to improve network performance.**

Highly available EC2 instances

By default EC2 launch in one availability zone. How can we make it more highly available and resist to failure.

We can have a standby EC2 instance in another availability zone. CloudWatch event or health watch to monitor the health of the EC2 and do the failover if the primary EC2 instance failed.

Revision #1
Created 2023-02-28 17:34:52 UTC by Tamarine
Updated 2023-02-28 20:04:29 UTC by Tamarine