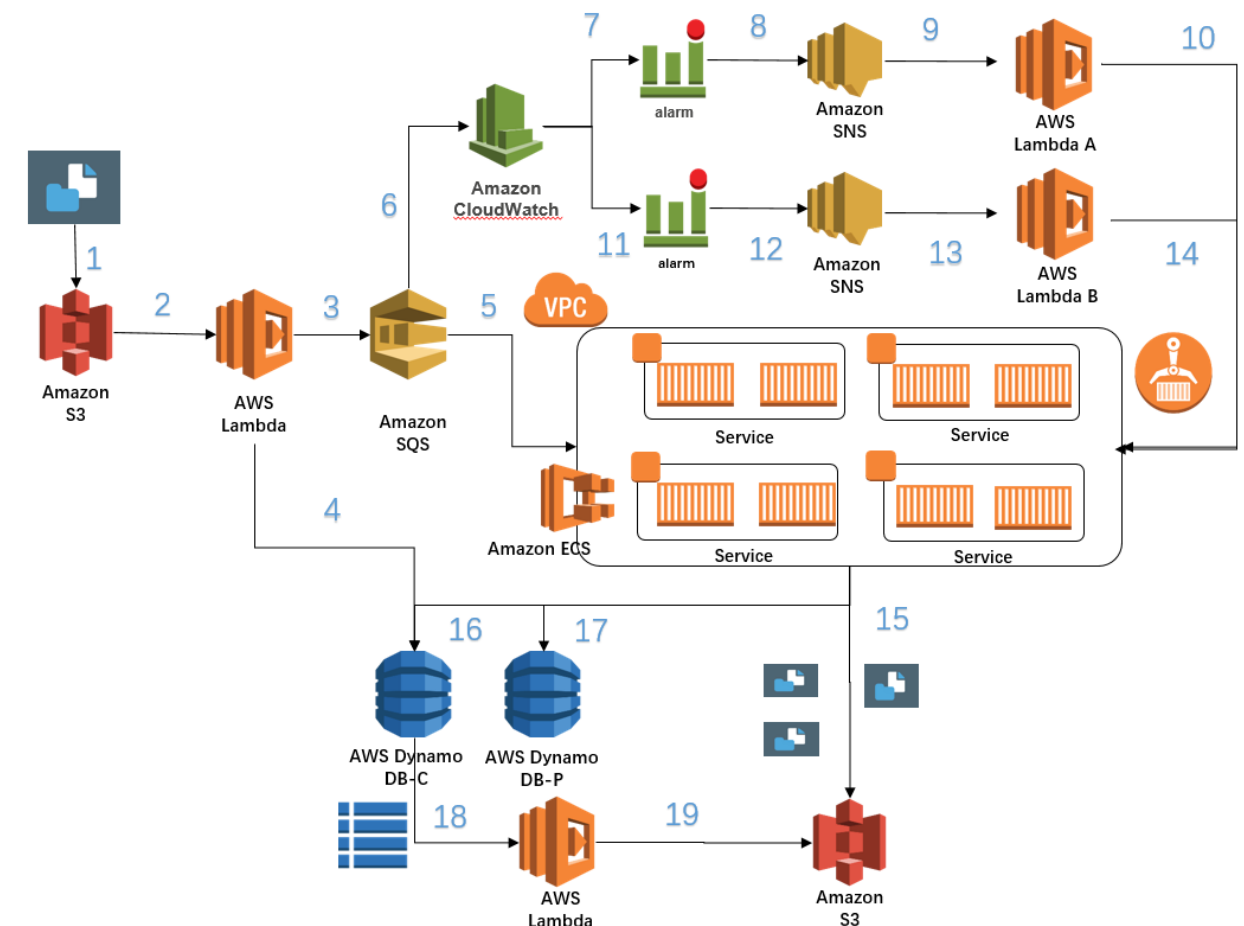


基于 Amazon ECS 的国内外文件同步系统解决方案

很多跨国企业都会有数据同步的需求，这些大型文件主要是存储在 S3 桶上，通过 console 或者命令行的形式从本地上传至 S3 中。本文描述了一种基于容器化的，在 Amazon ECS 平台上搭建的、快速同步的且基于负载进行扩容和收缩的、通用的解决方案。

解决方案概述

该方案如下图所示。



在该方案中，我们会启动以下资源：

1. 一个 global 的 S3 桶：用于 global 作为数据传输的源
2. 4 个 Lambda 函数： 分别用于文件分片，容器的扩张与收缩，发送传输完成信号

3. 1 个 sqs 队列：用于存储传输所用参数
4. 2 个 cloud watch 警报： 监控 sqs 内消息数量，分别用于需要收缩和扩张容器
5. 一个 ECS 集群，一个服务和若干个 task：用于搭载容器
6. 两个 dynamoDB 表： T 和 C。表 T 用于记录传输的 uploadID 和分段，传输情况，表 C 用于记录每个片段的唯一标记值-etag。
7. 具有相应的权限的角色。

在运行 cloudformation 之前需要您准备好：

国内：一个 S3 桶，用户的访问密钥 AccessId 和 AccessKey

国外：一个 VPC 和其中的一个公有子网， EC2 密钥对，用于接收信息的 Email 地址

步骤：

1. 用户在 global bucket 中成功上传文件。
2. 触发 lambda 函数。Lambda 函数会判断新文件的大小判断用哪种方式传输，当小于 5M 的时候会选择直接进行下载上传，当大于 5M 的时候会进行分段上传。
3. 当确定用哪种方式进行传输之后会把相应的信息存储至 sqs 中，包括文件名称，uploadID，分段情况等。
4. 同时 Lambda 会将传输情况备份至 dynamoDB 表 C 中,表 1 用于记载传输情况。
5. 容器运行在 ecs 集群中， SQS 内的信息被 ecs 中的 servces 和 task 消耗，
6. 我们为 SQS 设定了 cloudwatch 监控，用于根据负载实现容器的扩张和缩小。
7. 我们创建了警报 A 用于判断是否需要根据情况扩张和缩小容器的数量，当 sqs 内消息数量较大时会触发警报 A
8. 警报 A 会触发发送消息给 SNS
9. SNS 接到消息之后会触发 LambdaA
10. LambdaA 会调整容器的数量到 10 个，实现容器的扩张
11. 当 sqs 内消息数量较小时会触发警报 B
12. 警报 B 会触发发送消息给 SNS
13. SNS 接到消息之后会触发 LambdaB
14. LambdaB 会调整容器的数量到 1 个，实现容器的收缩
15. 容器会将下载的片段传至国内 S3 桶。
16. 容器更改 dynamo DB 表 C 中的数据，更新已完成的片段数量

17. 容器更改 dynamo DB 表 P，插入新的数据存入 etag 和 upload id partnumber 等信息。
18. 当表 C 中已完成的片段数量和公有片段数量一致的时候会触发 Lambda 函数
19. 函数搜索同一个 uploadID 的所有的片段的 etag 并按 partnumber 排序，发送给国内的 S3 桶。S3 桶在核对所有的 etag 之后进行片段组装，出现在国内的 S3 桶上。

架构部署

1. 登录 AWS 管理控制台并通过以下网址打开 AWS CloudFormation 控制台：
<https://console.aws.amazon.com/cloudformation>。
2. 如果这是新的 AWS CloudFormation 账户，请单击“Create New Stack”。否则，请单击“Create Stack”。
3. 在模板部分，选择指定 Amazon S3 模板 URL
4. 在 Specify Details 部分的 Name 字段中，输入堆栈名称。堆栈名中不得含有空格。
5. 在 Specify Parameters (指定参数) 页上，您将看到模板的 Parameters 部分中的参数。请按实际情况填写。
6. 单击 Next (下一步)。
7. 在这种情况下，我们不会添加任何标签。单击“Next”。作为密钥值对的标签可帮助您识别堆栈。
8. 审查堆栈信息。如果满意该设置，则单击“Create”。