合规保留：侧重于符合法律、法规或外部合规要求的保留策略。
治理保留：侧重于**公司内部**的数据管理和生命周期管理策略，以确保数据得到适当的管理和使用。

S3 生命周期管理：自动将对象从一个存储类转移到另一个存储类（如从标准存储类转到低频访问存储类或归档存储类），基于预设的时间规则，目的是优化存储成本。
S3 智能分层：自动根据数据访问频率将对象在不同的存储类之间切换（如在标准存储类和低频访问存储类之间），无需用户定义转移规则，**适用于访问模式不确定的场景**。

Elastic Beanstalk = 部署 Web 应用
ElastiCache = 让数据库访问更快

对比项	**Elastic Beanstalk**	ElastiCache
主要用途	让开发者快速部署和管理 Web 应用	缓存加速，提高应用性能
适用场景	自动化部署 Web 应用（支持 **Java、Python、Node.js** 等）	降低数据库查询压力、加速 Web 应用
底层架构	运行在 EC2 实例上(全托管 PaaS)	运行在 AWS 内存存储（全托管缓存）
支持的服务	绑定 EC2、RDS、ALB 等	提供 Redis、Memcached 两种缓存
特点	让开发者专注于代码，基础设施由 AWS 管理	通过缓存减少数据库负载，提高响应速度

Redis = 功能丰富、持久化、高可用、支持有序集合 → 适合**排行榜**、会话管理、实时分析
Memcached = 轻量级、分布式、仅 key-value 存储 → 适合高吞吐、临时缓存

✅**ElastiCache** 适用于 数据库查询优化，存储的是 **业务数据**，比如用户的会话数据、**排行榜**分数、临时计算结果等。
✅ CloudFront 适用于 Web 内容加速，存储的是 网页、图片、视频、API 响应，减少服务器负载。

如果你的业务需要：
- 提升数据库查询速度（如 Redis/Memcached）→ ElastiCache
- 提升静态资源/视频/API 响应速度（如全球加速）→ CloudFront

Athena 中使用标准 SQL 查询分析 S3 存储桶中的 CloudFront 日志
Amazon Athena + S3（适用于历史数据分析）

Amazon Rekognition 是一项提供图像和视频对象和场景检测的服务。

Amazon Kinesis Data Streams 可以实时收集和处理大量数据流。
AWS Lambda 函数可以由 Kinesis Data Firehose 触发，以便在数据流式传输时转换数据。
Amazon Kinesis Data Analytics 主要用于**实时流式数据分析**

利用了 **S3** 和 **Lambda** 的**可扩展性和按需付费定价**，适合处理大量数据，并适应需求的快速增长

结合使用 AWS Lambda、Amazon S3 和 Amazon SQS 提供了一个无服务器解决方案，可以自动扩展以处理变化的上传量，无需任何手动干预。

DataSync = 数据搬家（大规模传输数据到 AWS）。
AWS DataSync 主要用于批量、**定期**地将大量数据从本地存储传输到 S3（或其他 AWS 存储服务）。
它需要配置同步任务、计划执行，并且**不是作为一个持续可挂载的网络共享提供服务**

AWS DMS 主要设计用于数据库迁移，可能并不像传输大量 JSON 文件那样高效
AWS Storage Gateway 是一个**混合云存储**服务，帮助本地环境无缝连接 AWS 云存储。

📌 核心作用：
- 让本地应用可以像访问本地存储一样访问 AWS 存储（S3、EBS、Glacier）。
- 适用于企业备份、归档、混合存储扩展、数据库快照等。
- 提供缓存、存储扩展、本地备份、低延迟访问等功能。

📌 网关主要模式（3 种）：

1. ***文件网关***（*File Gateway*）
   把 S3 当作共享网络存储（NFS、SMB），适合文件存储、备份等。
   例如，企业 NAS 设备可以用它来自动将数据存入 S3。

2. ***卷网关***（*Volume Gateway*）
   让本地服务器像在用 本地磁盘 一样使用 AWS 作为后端存储（支持 iSCSI）。

缓存模式（Cached Volumes）：大部分数据存 AWS，**常用数据缓存本地**，节省存储成本，不适合维护对所有数据的本地访问。

存储模式（Stored Volumes）：数据先存本地，并**异步**地将数据备份到 AWS，确保在数据安全传输和存储在云中的同时保持了本地访问。

3. 磁带网关（Tape Gateway）

让传统磁带备份系统无缝使用 AWS Glacier 归档存储，替代实体磁带。

适用于长期数据存档（如金融、医疗机构的合规要求）。

Storage Gateway = 云存储桥梁（长期混合云存储，让本地应用无感知地使用 AWS 存储）。
文件网关 = 共享文件存储
卷网关 = iSCSI 磁盘存储
磁带网关 = 归档存储（磁带模拟）

NAT 网关用于允许私有子网中的实例连接到 internet，同时阻止来自 internet 的入站流量到达它们

Amazon Event Bridge 是一种无服务器事件总线服务，用于从应用程序、SaaS 应用程序和 AWS 服务中摄取数据，并将这些事件路由到目标服务。它使得构建松耦合的**事件驱动**架构更加容易。

AWS Step Functions 是一种可视化**工作流**服务，用于**协调多个 AWS 服务**以构建和运行弹性应用程序。它提供了设计和运行工作流的能力，这些工作流可以调用 AWS Lambda 函数来处理各个步骤。

在设计事件驱动架构时，EventBridge 通常用于事件的捕获和路由，而 Step Functions 负责处理复杂的业务逻辑和工作流管理。因此，二者可以结合使用，以实现一个高效、可扩展的事件驱动架构。

AppFlow
通过 API 从 **Salesforce**、Google Analytics、ServiceNow 等 SaaS 提取数据
在 SaaS 应用和 AWS 之间构建安全的数据流
结构化数据（如数据库表、客户信息、交易数据）**传输**

当企业需要 安全、**API 级数据传输**，并且希望在 Salesforce 和 S3 之间进行高效、可管理的数据同步时，**Amazon AppFlow** 是最佳选择，而 DataSync 主要适用于本地存储迁移，不适合 SaaS 数据集成。

要求提高 UDP 应用的可用性和性能，用 NLB，

NLB + Global Accelerator 组合，确保最低延迟和跨区域流量优化 Global Accelerator 支持 TCP 和 UDP
如果是 HTTP/HTTPS 应用（比如 **Web 服务器、API** ），那就要用 **ALB**。
ALB 支持 **HTTP** 健康检查

CloudFront 主要用于静态/动态**内容分发**（如游戏补丁、图片、视频），HTTP/HTTPS 流量
不适合用于 UDP、实时的游戏数据传输。
静态内容：固定的，不会根据每次请求发生变化的内容，比如图片、CSS、JavaScript 文件和视频等。
Lambda@Edge + CloudFront 是最佳选择，低延迟 + 无服务器 + 自动缩放，适合动态图片处理。

为了设计一个架构，将**多个 AWS Lambda 函数组合**成响应迅速的无服务器应用程序，同时以最小的运营开销，解决方案架构师应该使用 **AWS 步骤函数**

RDS Proxy 设计用于**管理数据库连接**，并且可以处理来自无服务器应用程序的大量连接。

AWS KMS 策略主要用于管理和控制对 KMS 密钥的访问，而**不直接对存储的数据进行加密**。
为确保所有写入 EBS 卷的数据在静止状态下加密，最佳实践是在创建 EBS 卷时启用加密。

Aurora Serverless：
是无服务器架构，不需要预先配置固定的计算资源
根据应用负载自动启动、缩放和暂停，适合间歇性或负载波动大的工作负载。
Amazon RDS for SQL：
需要预先选择和配置固定的**数据库实例**类型。扩展通常依赖于**手动调整实例大小**或使用只读副本，适用于持续稳定的工作负载。

Amazon SES，专注于**电子邮件传递**：SES 是专门用于发送电子邮件的服务，适合发送事务性邮件（如密码重置、订单确认）和营销邮件。
高交付率和专业功能：SES 提供对发送域的验证、退信（Bounce）和投诉处理、发送配额管理等功能，确保邮件高效、可靠地送达收件人邮箱。
成本效益高：配置简单，按使用量计费，运营开销较低，非常适合解决电子邮件传递问题。
Amazon SNS
基于发布/订阅模式的通知服务：SNS 用于将消息分发给多个订阅端点（如 SMS、移动推送、HTTP/S 终端等），适合广播通知、报警或跨系统传递信息
邮件发送仅为附加功能：虽然 SNS 可以通过电子邮件发送通知，但它**不具备 SES 那样的高**

级邮件传递管理功能（例如退信处理、送达率优化等），更适合发送简单通知，而非需要高可靠性和交付率的电子邮件。
总结：
在减少电子邮件传递问题和降低运营开销的场景下，Amazon SES 更适合，因为它专门为电子邮件设计，具备完善的交付和管理功能；而 SNS 更适用于多通道消息广播，不具备专门针对邮件的优化机制

GP3 和 IO2 是两种不同类型的 EBS（Elastic Block Store）卷：
GP3（General Purpose SSD）
用途：通用型 SSD 卷，适用于大多数工作负载，包括开发/测试环境、虚拟桌面、低延迟交互式应用程序等。
性能：提供一致的低延迟和高吞吐量。默认情况下，**GP3 卷提供 3000 IOPS** 和 125 MB/s 的吞吐量，可以根据需要增加。
成本：通常**比 IO2 更便宜**，适合需要高性能但不需要极高 IOPS 的应用。

IO2（Provisioned IOPS SSD）：
用途：专为需要**高 IOPS** 和低延迟的关键任务应用程序设计，如数据库工作负载（例如，SQL、NoSQL 数据库）。
性能：提供高达 **64,000 IOPS** 和 1,000 MB/s 的吞吐量，具有更高的耐久性（99.999%的可用性）。
成本：通常**比 GP3 更贵**，但提供更高的性能和可靠性，适合需要极高 IOPS 和低延迟的应用。
GP3 是 GP2 的增强版（更便宜，性能可调）。
IO2 是 io1 的升级版（更高 IOPS，更耐久）。

高可用性 → 选择 Multi-AZ
减少主实例负载 → 选择读取副本（Read Replica） 用于处理读流量，减少主实例负载，降低成本
降低成本 → 创建一个读取副本

SageMaker → 机器学习（ML）训练 & 部署，训练 AI 模型
Marketplace → AWS 上的“应用商店”，购买第三方软件
OpenSearch → 搜索 & 日志分析（类似 Elasticsearch）

AWS Glue 是一个完全托管的 ETL 服务，可以自动处理.csv 文件到 Parquet 格式的转换，并通过 AWS Lambda 函数触发 S3 PUT 事件

Global Accelerator 适用于动态、低延迟应用，能自动路由流量到健康端点，保证游戏玩家流畅体验。

Amazon Machine Image（AMI）是 EC2 实例的模板，包含操作系统、应用程序、配置等内容。通过 AMI, 用户可以**快速启动新的 EC2 实例**, 而不需要每次手动安装操作系统和软件。

Amazon RDS 中的 Multi-AZ 部署通过自动将数据库复制到不同的可用区的备用实例，提供了**高容错**和**灾难恢复**能力。这个过程确保备用实例始终与主实例保持最新，允许在发生故障时具有非常低的 RPO 和最小的数据丢失。
AWS Backup 允许将备份复制到另一个区域，提供**灾难恢复**能力

Amazon EFS 完全托管的弹性文件存储服务，支持 NFS（Network File System）协议，是用于基于 Linux 的应用程序的文件系统
Amazon FSx for Windows File Server 是一个完全托管的服务，提供熟悉且一致的文件系统体验，与 Windows 应用程序兼容

IAM 角色（AWS Identity and Access Management Role） 是一种 AWS 访问权限的身份，用于允许 AWS 资源或用户在不使用固定凭证（如密码或密钥）的情况下访问 AWS 服务

同一 AWS 环境中的 AWS 服务之间的私有访问 VPC 端点

在高吞吐量 & 低延迟的场景（如 AWS 高性能计算、数据库、高速缓存），会选择 SSD （固态硬盘

AWS Snowcone，旨在将大量数据导入 AWS，但每个设备限 **8TB**
DataSync 它专为处理大规模数据传输而设计，并可以安排定期运行，是一个数据传输服务，可以自动在本地存储和 AWS 服务之间移动数据。它可以用于将数据迁移到 Amazon FSx for Windows File Server，并且它允许带宽限制来控制数据传输速率，最小化对其他部门网络使用的影響。DataSync 还可以在要求的 5 天时间框架内执行迁移。

SFTP（Secure File Transfer Protocol，安全文件传输协议） 服务器是一种用于安全地传输文件的服务器，它基于 SSH（Secure Shell） 协议，提供加密的文件传输，比传统的 FTP（File Transfer Protocol）更安全。

Amazon Inspector 是一项**安全评估服务**，可以自动识别在 EC2 实例上运行的应用程序中的漏洞和安全问题。通过部署 Inspector 代理，公司可以执行**安全评估**并接收详细**报告**

Amazon GuardDuty 监控和检测 AWS 环境中的潜在威胁，自动识别威胁并发出警报
Amazon Detective 主要用于 调查和分析安全事件，它的重点是 调查已发生的安全事件，基于 AWS CloudTrail、VPC Flow Logs 和 Amazon GuardDuty 等数据源进行分析

Kinesis Data Streams 是一个实时数据流 处理平台，用于收集、处理和分析来自各种来源（如网站点击流、社交媒体、传感器等）的数据流。它允许你从流中实时读取数据并进行处理。支持高吞吐量和低延迟，适合需要实时处理的应用
Kinesis Data **Firehose** 是一个 完全托管的服务，用于将**流数据传输**到 AWS 数据存储服务（如 Amazon S3、Amazon Redshift、Amazon Elasticsearch Service 等）中，而无需开发自定义应用程序。自动处理吞吐量，适合高吞吐量但不需要即时响应的场景

如果 VPC 里的 EC2 实例是 IPv6，而你 希望它们可以访问互联网但不被外部访问，NAT 网关不支持 **IPv6**，所以你必须使用**互联网网关**（IGW 或 EIGW）。

VGW(Virtual Private Gateway)是 AWS 端的 VPN 入口，主要用于 AWS 和本地数据中心建立 Site-to-Site VPN
它需要和 Customer Gateway（CGW，本地端的 VPN 设备）配合使用
如果对带宽和稳定性要求更高，可以选择 AWS Direct Connect（DX）

高频错误

1. VPC & 网络相关

VPC 端点（Gateway vs. Interface）
✅ 网关端点（Gateway Endpoint） → S3、DynamoDB，不走公网，免费。
✅ 接口端点（Interface Endpoint） → 适用于其他 AWS 服务（如 SSM、EC2 API），创建 ENI，可能有额外数据处理费。

✔ 考点陷阱：
• S3 访问 → 选 网关端点，因为它不收费。
• 如果 VPC 需要访问 EC2 API 或 SSM → 选 接口端点。
• 私有子网访问 S3 → VPC 网关端点。

❌ 错题点：私有子网如何访问互联网

✅ NAT 网关 → 让私有子网的实例能访问公网（但公网不能访问它们）。
✅ VPC 端点 → 让私有子网访问 S3/DynamoDB 不走公网。

❌ Internet Gateway → 仅适用于公有子网，不能让私有子网访问公网。

✔ 考点陷阱：
• 如果问题问的是「访问 S3」，而不是访问互联网 → 选 VPC 网关端点
• 如果私有子网的 EC2 需要更新软件包访问公网 → 选 NAT 网关

2. 数据库 & 缓存
❌ 错题点：DAX vs. ElastiCache for Redis
错因：DAX 只适用于 DynamoDB，而 Redis 更通用
✅ DAX（DynamoDB Accelerator）→ 适用于 DynamoDB，无需修改应用，减少查询延迟。
✅ ElastiCache for Redis → 适用于所有数据库（如 RDS、DynamoDB）**手动管理**缓存逻辑。

✔ 考点陷阱：
• 如果题目中提到「DynamoDB 高速缓存」→ 选 DAX，因为它专门为 DynamoDB 设计。
• 如果应用需要手动管理缓存（如 RDS 缓存查询） → 选 Redis。
Redis vs. Memcached
• 选 Redis → 需要持久化、多种数据结构。
• 选 Memcached → 简单 Key-Value 缓存，不需要持久化。

3. 安全 & 访问控制
❌ 错题点：IAM 组 vs. IAM 角色
错因：错误使用 IAM 角色，而不是 IAM 组
✅ IAM 组 → 适用于一组用户，例如开发人员组、管理员组。
✅ IAM 角色 → **适用于 EC2、Lambda 等 AWS 资源访问** AWS 服务，或 跨账户访问。

✔ 考点陷阱：
• 如果题目问「管理多个用户权限」 → 选 IAM 组
• 如果 EC2 实例要访问 S3，应该用 IAM 角色（EC2 实例角色）。
• 跨账户访问 AWS 资源 → 选 IAM 角色

4. S3 & 存储
❌ 错题点：S3 版本控制 vs. S3 对象锁定
✅ S3 版本控制 → 保存多个版本，但用户仍可以删除整个对象。
✅ **S3 对象锁定 → 防止文件删除/修改**（适用于合规要求）。

✔ 考点陷阱：

• 如果题目问「防止误删」 → 选 S3 版本控制。
• 如果题目问「防止修改或删除（合规要求）」 → 选 S3 对象锁定。

❌ 错题点：S3 静态网站托管 vs. IAM 只读权限
✅ S3 静态网站托管 → 适用于公开访问的静态内容（如 HTML、CSS）。
✅ IAM 只读权限 → 适用于受控访问、无法让 S3 作为网站公开访问。

✔ 考点陷阱：
• 如果题目说「所有人都能访问」 → 选 S3 静态网站托管。
• 如果题目说「限制某些用户访问 S3」 → 选 **IAM 策略**。

5. 数据迁移 & 传输

❌ 错题点：AWS DataSync vs. AWS S3 sync
✅ **AWS DataSync** → 大规模、自动化、增量数据迁移，支持 **EFS/NFS/SMB**，加速传输。
✅ AWS CLI S3sync → 同步 S3，适用于 手动、小规模同步，但不适合长期自动化迁移。

✔ 考点陷阱：
• 如果题目说「**定期**、自动化、大规模迁移」 → 选 DataSync。
• 如果题目说「一次性、小规模同步」 → 选 s3 sync。
• 如果数据来源是 EFS/NFS/SMB → 只能用 DataSync，s3 sync 只能同步 S3。

📌 1. NAT 网关 vs. 互联网网关（IGW）
易错点：
• NAT 网关只用于私有子网的出站流量。
• IGW 支持双向流量。

📌 2. S3 智能分层 vs. S3 生命周期
S3 生命周期 = S3 智能分层是不同的。
🚀 记忆比喻：
• S3 智能分层 → 自动温控空调，根据温度变化自动调整功率。
• S3 生命周期 → 定时开关空调，按照设定时间切换模式。

SCP 服务控制策略→ “公司级规定”： **root 和 IAM 用户**都必须遵守。
IAM 组 → “部门级权限”：只能限制普通 IAM 用户，无法限制 root 账户
AWS Service Control Policies (SCP)，管理和控制组织内的 AWS 账户

S3 生命周期 vs. S3 智能分层

- S3 生命周期：规则驱动，定期把数据迁移到不同存储层（标准 → IA → Glacier）。
- S3 智能分层：自动检测访问模式，动态调整存储层，无需用户手动设置规则。

DynamoDB 备份 vs. Streams

DynamoDB 持续备份 & **点播恢复**（PITR）：自动存储数据到 S3，可回溯最多 35 天。

DynamoDB Streams：仅记录数据变更（插入、更新、删除），最多保留 24 小时，不存 S3。

比喻：PITR = 时光机（可以回到过去），Streams = 直播回放（只记录最新变化）。

CloudWatch vs. EventBridge

- CloudWatch：监控指标（CPU、内存、日志）。
- EventBridge：事件驱动架构，处理 AWS 资源的变更（如 EC2 启动、S3 上传）。

7. DynamoDB vs. RDS

- DynamoDB：NoSQL，适合**高并发、低延迟**（游戏、IoT）。
- RDS：SQL 关系型数据库（支持 MySQL、PostgreSQL）。

8. ECS vs. EKS

- ECS（Fargate 选项）：AWS 托管的容器编排，简单易用。
- EKS：Kubernetes 兼容，适合复杂微服务架构。需要更多管理

缓存卷网关，只缓存频繁访问的数据，不适合维护对所有数据的本地访问。

存储卷网关保留了本地的全部数据集，并**异步**地将数据**备份**到 AWS，确保在数据安全传输和存储在云中的同时保持了本地访问。

在高吞吐量 & 低延迟的场景（如 AWS 高性能计算、数据库、高速缓存），会选择 SSD （固态硬盘

SFTP（Secure File Transfer Protocol，安全文件传输协议） 服务器是一种用于安全地传输文件的服务器，它基于 SSH（Secure Shell） 协议，提供加密的文件传输，比传统的 FTP 更安全。

🟢 1. 设计具有弹性和高可用性的架构（Resilient Architectures）

✅ 负载均衡（ELB）

- ALB（应用型，支持 HTTP/HTTPS）

• NLB（网络型，适合低延迟、高吞吐量）

✅ 自动伸缩（Auto Scaling）

• 目标追踪（Target Tracking）：按 CPU 利用率或请求数自动调整

• 步进策略（Step Scaling）：根据指标变化增减实例

• 计划扩展（Scheduled Scaling）：预定时间扩展

✅ 高可用性（HA）架构

• 跨 AZ 部署（EC2、RDS Multi-AZ、ELB）

• 跨区域部署（Global Accelerator、CloudFront）

• 数据库备份与恢复（RDS Read Replica vs. Multi-AZ）

✅ CloudFront & Global Accelerator

• CloudFront：CDN，加速静态 & 动态内容

• Global Accelerator：跨区域加速，提高 TCP/UDP 连接性能

✅ EFS vs. FSx

• EFS（Linux 共享存储）：支持 NFS，适合可扩展的应用

• FSx for Windows：适用于 Windows 文件共享（支持 SMB）

• FSx for Lustre：适用于高性能计算（HPC）

🟡 2. 设计高效性能架构（High-Performing Architectures）

✅ 计算服务选择

• EC2 实例类型：计算优化（C5）、内存优化（R5）、存储优化（I3）

• Lambda：适用于无服务器计算，支持 Event-driven 事件触发

✅ 存储选择

• S3 存储类：Standard、IA、Glacier、Glacier Deep Archive

• S3 点播恢复（Restore from Glacier）：Expedited（1-5 分钟）、Standard（3-5 小时）、Bulk（5-12 小时）

• S3 事件通知：可触发 Lambda、SNS、SQS

✅ 数据库架构

• RDS（关系型数据库）

• Multi-AZ（高可用性），Read Replica（读扩展）

• Aurora：支持 Auto Scaling，自动修复

• DynamoDB（NoSQL）

• 适用于高吞吐量、低延迟场景

• DAX（DynamoDB Accelerator）：缓存加速，降低延迟

• 幂等性：防止重复写入，通常使用条件写入或去重 Token

🟠 3. 设计安全架构（Secure Applications & Architectures）

✅ VPC 设计

- 子网（Public Subnet vs. Private Subnet）
- NAT 网关 vs. Internet 网关
- NAT 网关：让私有子网访问公网
- Internet 网关：让公有子网访问公网

✅ 身份和访问管理（IAM）

- IAM 角色：用于 AWS 服务（EC2、Lambda 等）访问资源
- IAM 用户：给具体的用户或应用程序
- 服务控制策略（SCP）

  适用于 组织（Organizations） 级别权限管理

  用于限制子账户的权限（比如禁止创建特定资源）

✅ 网络安全

- 安全组（SG）：状态性（允许的流量自动返回）
- NACL：无状态（进出规则需单独配置）

✅ 堡垒主机 vs. AWS Systems Manager

- 堡垒主机（Bastion Host）：作为跳板机管理私有服务器
- AWS Systems Manager Session Manager：无需公网 IP，安全**管理 EC2**

🔵 4. 设计成本优化架构（Cost-Optimized Architectures）

✅ EC2 计费模式

- 按需实例（On-Demand）：适用于短期任务
- 预留实例（Reserved Instances, RI）：适用于长期稳定负载（可省 75%）
- Spot 实例：适用于容错性任务（便宜但可能被回收）

✅ 存储成本优化

- S3 生命周期管理：自动转移到 IA、Glacier
- EBS 快照 vs. 备份

  EBS 快照：增量备份，适用于短期恢复

  AWS Backup：长期存储策略，可跨区域复制

✅ 数据迁移

- AWS DataSync vs. AWS DMS
- DataSync：用于**文件级**迁移（如 NFS、SMB 到 S3、EFS）
- DMS（Database Migration Service）：用于**数据库**迁移（支持异构迁移）

✅ 离线数据迁移

- AWS Snowball Edge：50TB~80TB 设备，可带计算能力
- AWS Snowmobile：100PB 级别的大规模迁移
- AWS Snowcone，旨在将大量数据导入 AWS，但每个设备限 **8TB**

Lambda 基于请求数量和代码执行时间收费。

AWS Transfer Family vs. Amazon S3 File Gateway

AWS Transfer Family

用途：提供基于 SFTP/FTPS/FTP 的文件传输服务

适用于需要安全地传输文件到 AWS 的场景，例如合作伙伴或企业内部通过 SFTP/FTPS/FTP 访问 S3。

解决的问题是：企业现有系统依赖传统 FTP 服务器，而 AWS 提供了托管方案，无须自行维护 FTP 服务器。

考试重点：

- 协议支持：支持 SFTP、FTPS、FTP
- 数据存储：将上传的文件直接存储到 S3 或 EFS。
- 适用场景：需要兼容传统 FTP 方式传输数据到 AWS，例如 供应链、金融、医疗等行业。
- 身份验证：支持 AWS IAM、Active Directory、本地用户认证。
- 不支持本地缓存，仅作为数据上传/下载的桥梁。

Amazon S3 File Gateway

用途：在本地提供 S3 兼容的文件存储访问

适用于 本地应用需要像 NAS 一样访问 S3，但应用本身不支持 S3 API。

解决的问题是：企业已有本地应用使用 SMB/NFS 访问存储，但希望数据长期存储在 S3，减少本地存储成本。

考试重点：

- 协议支持：支持 SMB/NFS（NAS 方式）
- 数据存储：本地缓存 + S3 （读取频繁的数据会缓存，加快访问速度）。
- 适用场景：本地应用或服务器无缝访问 S3，例如 备份、文件共享、数据归档。
- 支持本地缓存，加快本地数据读取速度，减少 AWS 访问延迟。

两者的核心区别是：

- Transfer Family 解决的是 传统 FTP 迁移到 AWS 的问题。

- S3 File Gateway 解决的是 本地存储无缝扩展到 S3 的问题。

区域级别文件存储

Amazon FSx VS Amazon EFS

✅ **Amazon FSx**（适用于 Windows/Lustre）

Amazon FSx 提供了完全托管的高性能文件系统，支持两种模式：

FSx for Windows File Server（基于 Windows Server）

FSx for Lustre（适用于高性能计算 HPC）

- 支持**多区域复制**（**Multi-AZ 部署**），不支持跨区域自动访问。
- 适合 Windows 文件服务器（FSx for Windows）或高性能计算（FSx for Lustre）。
- 可以**手动配置跨区域复制**，但主要用于**灾难恢复**（DR）

Amazon EFS（适用于 **Linux**）

- 单区域存储，支持多可用区（Multi-AZ）部署。
- 可用于多个 EC2 实例，支持共享访问，适用于 Web 服务器集群、数据分析等场景。
- 可以通过 VPN 或 Direct Connect 挂载到本地数据中心，但不能跨区域共享。

📌 为什么 Amazon FSx 不能做真正的多区域部署？

1. FSx 可以复制数据到另一个区域，但不能直接在多个区域访问。
2. EFS 也无法跨区域访问，但在同一区域内提供高可用性和自动扩展。。

结论：

- 如果是 Linux 环境，需要多个实例访问同一个文件系统，用 EFS。
- 如果是 Windows 或高性能计算环境，或者有灾难恢复需求，FSx 更合适，但它不是跨区域自动共享的解决方案。

适用场景	推荐服务
本地文件访问 S3	S3 File Gateway
Windows 共享存储（SMB）	FSx for Windows
高性能计算、AI 训练（HPC、ML）	FSx for Lustre
多个 EC2 实例共享 NFS 存储	EFS
跨 AZ 高可用文件系统	EFS
数据存储在 S3，但希望本地访问更快	S3 File Gateway

跨区域复制

AWS DataSync：是一个通用的数据传输服务，支持多种存储类型和数据传输场景，适用于数据迁移、备份、归档和灾难恢复。

Amazon RDS 跨区域复制：专门用于关系数据库的跨区域复制，提供针对关系数据库的优化功能，如只读副本和自动故障转移。

Amazon DynamoDB 跨区域复制：专门用于 NoSQL 数据库的跨区域复制，提供**低延迟**的全球访问和多区域写操作，适用于高性能和可扩展性的分布式应用。

✅ ElastiCache = 数据库缓存（加速数据库查询，减少数据库负载）
✅ CloudFront = CDN 缓存（加速 Web 内容分发，减少 Web 服务器负载）

如果你的业务需要：

- 提升数据库查询速度（如 Redis/Memcached）→ ElastiCache
- 提升静态资源/视频/API 响应速度（如全球加速）→ CloudFront

特性	AWS Transfer Family	AWS S3 File Gateway
主要用途	SFTP/FTPS/FTP 访问 S3	本地服务器通过 NFS/SMB 访问 S3
目标用户	外部用户 / 合作伙伴	本地应用 / 内部服务器
协议支持	SFTP, FTPS, FTP	NFS, SMB
是否缓存数据	❌ 不缓存，直接存入 S3	✅ 提供本地缓存，加速访问
适合的场景	B2B 文件共享、外部数据上传	本地服务器访问 S3，混合云存储

如果你是数据分析师，不想写代码，只想拖拽清理数据 → 用 **DataBrew**
如果你是数据工程师，需要运行 Spark 任务处理 **TB 级**数据 → 用 EMR Serverless

特性	**Amazon Aurora**	Amazon Athena
数据库类型	关系型数据库（RDBMS）	无服务器查询引擎（Serverless SQL）
数据存储	存储在 Aurora 的数据库实例中	直接查询 S3 上的数据（如 CSV、Parquet、JSON）
查询方式	结构化查询（SQL，支持事务）	只支持 只读 SQL 查询（不支持事务）
性能优化	适合**高并发事务处理**（OLTP），支持索引	适合大规模数据分析（OLAP），自动优化查询
扩展性	存储可扩展到 128TB，计算需手动	完全无服务器，查询按需扩

	扩展（Aurora Serverless 可自动扩展）	展计算资源
费用模式	按实例计费（或 Serverless 按用量）	按查询扫描的数据量计费（$5/TB）
适用场景	适用于事务型应用（如网站、应用数据库）	适用于大数据分析（如日志分析、BI）

✅ 使用身份策略（Identity-based policy）

- 当你希望在 IAM 级别管理 EFS 访问权限时。
- 适用于公司内部的 IAM 用户、组、角色。
- 适用于精细化权限控制，如只允许某些角色挂载 EFS，但不允许删除。

✅ 使用资源策略（Resource-based policy）

- 当你希望跨 AWS 账户共享 EFS 访问权限时。
- 当你希望直接在 EFS 资源级别管理权限时。
- 需要明确拒绝（Deny）某些用户或账户访问 EFS 时。

你的需求	选择
需要迁移 50 TB – 1 PB 数据	✅ Snowball Edge
需要 在设备上进行计算处理（如 EC2、Lambda）	✅ Snowball Edge
需要 迁移 10 PB 以上数据	✅ Snowmobile
需要 一次性迁移整个数据中心	✅ Snowmobile

关键点	AWS DataSync	AWS DMS
用途	文件存储迁移（**文件**、对象存储）	数据库迁移
支持的数据类型	文件、对象存储（如 NFS、SMB、S3、EFS）	结构化数据库（如 MySQL、PostgreSQL、Oracle）
源端	服务器（NFS/SMB）、EFS、S3、FSx	关系型数据库（RDS、本地数据库）
目标端	S3、EFS、FSx	关系型数据库（RDS、DynamoDB、Redshift）
迁移方式	批量同步（定期同步大量文件）	增量同步（持续复制数据库变更）
适用场景	大规模文件迁移（如本地存储到 S3）	数据库升级、云端迁移、灾难恢复
实时性	**延迟较高**（适用于定期同步）	支持低延迟增量复制（CDC）

✅ FSx for Windows File Server 完全兼容 SMB 和 Windows 应用，无需更改应用程序，适合企业文件共享。
✅ FSx for Lustre 主要用于 HPC 和 Linux，不支持 SMB，不适用于 Windows 服务器文件存储。
✅ AWS Storage Gateway（卷网关）是**块存储**（iSCSI），不支持 SMB 共享，不适用于 Windows 文件系统。

✅ 如果需要 Windows 文件服务器（SMB） → 选 Amazon FSx for Windows File Server
✅ 如果需要 Windows & Linux 混合存储（SMB + NFS） → 选 Amazon FSx for NetApp ONTAP
✅ 如果需要 Linux 共享存储（NFS） → 选 Amazon EFS
✅ 如果需要**本地文件共享并同步到 AWS** → 选 AWS Storage Gateway（文件网关）

Transfer Family 适用于**需要安全**、**标准化文件传输**以及与**外部合作伙伴或内部系统进行集成**的场景
DataSync 更适合大规模、自动化的数据迁移任务，如在本地数据中心与 AWS 存储服务之间的周期性数据同步。
因此，当安全传输（特别是使用 AS2 协议）和集成定制身份验证是关键需求时，AWS Transfer Family 是更合适的选择，而 DataSync 的设计目标并不涵盖这些场景。

DataSync
迁移本地存储（NAS、NFS、SMB）到 S3、EFS、FSx
在 AWS 区域间同步大文件数据
处理非结构化数据（如日志、视频、图片等）

VPC 端点让你的服务器访问 AWS 内部服务时，不用经过公网，提高安全性和速度。

- 网关端点 适用于 S3 和 DynamoDB，直接通过 VPC 路由表 指定访问路径，不需要额外的网络配置。
- 接口端点 适用于 其他 AWS 服务（如 EC2 API、SNS、Secrets Manager），通过 ENI（弹性网络接口）连接，需要分配私有 IP，并可能需要配置安全组，收费

如果你使用 DynamoDB 并且想降低查询延迟，DAX 是更好的选择，因为它与 DynamoDB

API 兼容，且全托管。
如果你需要更通用的缓存（支持 RDS、S3、复杂数据结构、持久化），选择 ElastiCache for Redis。

IAM 角色不适合直接管理长期用户权限
•IAM 角色是 被"假设"使用的（Assumed），不适用于长期权限分配。
•角色需要用户 手动切换 或使用 STS（安全令牌服务）来临时获取权限，增加了复杂性。
如果开发者每天都要访问 AWS 资源，使用 **IAM 组**比要求他们每天切换 IAM 角色更方便。

使用 S3 存储桶策略 来允许公共访问
IAM 策略只能用于 AWS 内部的身份访问。

• 静态网站 适合 简单、展示型内容（如博客、官网、文档站），可以托管在 Amazon S3 上，访问速度快，成本低。
• 动态网站 适合 需要交互和个性化的内容（如电商、社交媒体、后台管理系统），需要运行后端服务器，支持数据库。

✅ AWS AppConfig 适合：
• 需要 动态更新应用配置，而不想重新部署代码。
• 需要 逐步发布配置（灰度发布），避免影响所有用户。
• 需要 自动回滚，减少错误配置带来的风险。

✅ 如果只是存储静态参数，可以用 SSM Parameter Store。
✅ 如果存储的是机密信息（如 API 密钥），用 AWS Secrets Manager。

写入时缓存策略，是正确的解决方案。这种策略确保任何添加到或更新到数据库中的数据立即反映在缓存中，维护缓存和数据库之间的数据一致性。

选择方案	适用场景
AWS DataSync	大规模、定期、自动化 数据迁移 & 备份
AWS CLI s3 sync	小规模、一次性 文件同步（手动运行）

1. 如果你要实时处理数据（低延迟，毫秒级别）→ Kinesis Data Streams
• 适合金融交易监测、点击流分析、实时推荐系统等。
2. 如果你要存储 & 批量传输数据（自动化，无需代码）→ Kinesis Data Firehose
• 适合日志存储、数据湖建设、数据仓库（S3、Redshift）等。

Kinesis Data Firehose 侧重于实时数据流的传输和处理，而 AWS DataSync 则侧重于高效的批量数据传输和同步

•多个 VPC 需要互通，考虑 **AWS Transit Gateway**（类似地铁枢纽）
•**AWS Transfer Family** 提供了一个 SFTP 的托管服务，可以与 AmazonS3 集成，允许供应商继续使用他们的 SFTP 客户端将文件直接传输到 S3
•**Amazon S3 File Gateway** 适用于文件级存储，而不是像 iSCSI 这样的块存储。
EBS 存储与备份到 Amazon S3 不提供所需的低延迟访问。存储卷将所有数据存储在本地，这并没有减少对本地服务器的依赖。
•**Amazon API Gateway**，是将传入的请求路由到 AmazonEKS 集群中适当微服务的最具成本效益的解决方案。API Gateway 是一个完全托管的服务，可以处理 HTTP/S 请求路由到正确的微服务端点，为管理 API 流量提供可扩展和高效的方法。

常是:
✅ **RDS Multi-AZ（自动故障转移）**
⛔ 不要选 Read Replica，因为它是用于读扩展，而不是**高可用（HA）或灾难恢复（DR）。**

Control Tower 基于 Organizations，但自动化了账户创建和治理，适合企业标准化管理。Organizations 则适合更灵活的自定义管理。

⭐AWSKMS 适合用于加密 EBS 卷和 Aurora 数据库的静止数据，
ACM 证书适合加密 ALB 传输中的数据

AWS Config 确实可以跟踪资源配置，但它主要用于审核和合规性检查，而不是专门针对 S3 版本控制的识别

S3 多区域访问端点用于加速跨区域访问，而不是用于识别版本控制状态

预签名 URL 提供了对 **S3 存储桶**中特定对象的时间限制访问，而无需向外部顾问提供对整个存储桶的访问或需要长期凭据

Amazon EFS 旨在提供多个 EC2 实例可以同时访问的共享文件系统。通过将 EFS 挂载到每个实例，网站资产在所有实例中保持一致，延迟最小。

AWS Secrets Manager 可以**存储密码**等敏感信息，还可以自动**管理**它们的**生命周期**。

AWS KMS（Key Management Service）的主要功能是提供**加密**服务

读取副本（Read Replica）是为了提高数据库的**可扩展性和可用性**，特别是在处理大量只读请求时。读取副本的重点通常在于**优化读取**操作
多区域部署，提高数据库的**高可用性，**灾难恢复

多区域部署的 Amazon RDS 适用于需要 跨区域高可用性 和 灾难恢复 的应用场景，尤其适合全球分布式应用，提供跨区域冗余。
Amazon RDS 副本 主要用于提高 读取性能 和 扩展性，适用于 高负载的读操作，但并不涉及跨区域容错，适合在单区域或多个可用区内进行负载分担。

是指将 Amazon RDS 数据库实例的主数据库和备份数据库放置在不同的可用区内。AWS 会自动管理主数据库和备份数据库之间的数据同步 自动故障转移

跨两个可用区 通常指的是将资源分布在不同的可用区，但并不特指 RDS 的特定高可用性配置。它的目的是提高资源的可用性和容错能力，通常用于提高 读写操作的分布 或 负载均衡。
手动故障转移 数据异步

Amazon ElastiCache，特别是使用 Redis 或 Memcached，可以用作**分布式缓存**来存储会话数据

AWS Glue DataBrew 用途：数据清洗和准备

EMR Serverless
适用于 大规模数据处理，但需要使用 Spark、Hive、Presto 这些 编码 方式编写 ETL 任务。

Amazon Redshift 是 专为**大规模数据分析（OLAP）** 设计的，具有 MPP 能力。它可以将查询任务 并行分配到多个节点，非常适合处理海量数据的分析、聚合和计算工作负载

1. Amazon Aurora MySQL
✅ **高性能、自动扩展、高可用的托管数据库**，兼容 MySQL，但底层采用 分布式存储架构。
- 性能：比 RDS for MySQL 快 3-5 倍。
- 存储架构：数据自动分片存储在 6 个不同的 AZ（可用区）。
- 自动扩展：存储可 自动扩展（10GB 起步，最高 128TB）。
- 高可用性：即使主节点故障，Aurora 也能快速故障转移（Aurora 复制延迟 <100ms）。
- 费用：按使用量计费（存储 + 计算分开计费）。

2. Amazon RDS for MySQL

✅ 托管 MySQL 数据库，但底层仍然基于 传统 MySQL 架构。

- 性能：相比 Aurora 较慢，适用于中小型应用。
- 存储架构：数据存储在 单个 AZ 或 跨多个 AZ（手动设置）。
- 手动扩展：存储和计算需要手动扩展（最高 64TB）。
- 高可用性：需要 手动启用多 AZ 复制，否则主节点故障时可能需要较长恢复时间。
- 费用：通常比 Aurora 便宜，但需要更主动的管理。

✅ Aurora MySQL 和 RDS for MySQL 都支持特定时间点恢复（PITR），但方式略有不同：

1. Aurora MySQL

- 自动持续备份，最多可恢复至 35 天内的任意时间点。
- 通过 Aurora Backtrack，可以在数秒内将数据库恢复到 秒级精度的过去状态（最多 72 小时）。

2. RDS for MySQL

- 只支持 从最新的自动快照恢复，或者 恢复到 5 分钟粒度的过去时间点（最长 35 天）。
- 需要新建 RDS 实例 来恢复，不如 Aurora 快速。

✅ 结论：Aurora MySQL 恢复特定时间点更快、更灵活，PITR 的时间粒度也更高。

Redis 提供了高可用性和故障转移能力，确保在 Web 服务器故障时会话状态不会丢失

✅ SSD 适合高性能需求的场景

- 操作系统盘（Windows、Linux 启动盘）
- 高性能数据库（低延迟、高速读写）
- 游戏存储（加快加载速度）
- 视频编辑（4K/8K 视频渲染）

✅ HDD 适合大容量存储、低成本需求的场景

- 数据归档、冷存储（例如电影、照片、备份）
- 服务器日志存储
- 监控录像存储
- 企业级大容量存储（如 NAS、数据中心）

✅ 如果追求性能 → 选 SSD（快、低功耗、无噪音）

✅ 如果追求容量和成本 → 选 HDD（便宜、适合存档）

AWS 不支持跨多个可用区的子网