Instructions:

Part 1: Designing Cloud Infrastructure

- **Task:**
  - Design a cloud infrastructure for a scalable web application.
  - Include components like compute instances, storage, and network configurations.
  - Use AWS EC2, S3, and VPC to build the basic architecture.

Parte 1: Diseño de la Infraestructura en la Nube

El diseño propuesto utiliza **Amazon VPC** para el aislamiento de red, **Amazon EC2** para la computación y **Amazon S3** para el almacenamiento de objetos estáticos.

Componente	Servicio AWS	Configuración Clave
Red	**Amazon VPC**	Una VPC con **subredes públicas** (para el Balanceador de Carga y EC2 de "saltos") y **subredes privadas** (para EC2 de la aplicación y Base de Datos). Se usa un **Internet Gateway (IGW)** en la VPC y **Network Address Translation (NAT) Gateway** en la subred pública para que las instancias privadas accedan a internet.
Cómputo	**Amazon EC2**	Instancias de aplicación desplegadas en las **subredes privadas**. Se usan **Grupos de Auto Escalado (ASG)** para mantener la disponibilidad.
Almacenamiento Estático	**Amazon S3**	**Buckets de S3** para almacenar contenido estático (imágenes, CSS, JS). Se utiliza **CloudFront** (no obligatorio en la tarea, pero muy recomendado) para distribución de contenido y reducir la carga de EC2.
Balanceo de Carga	**Elastic Load Balancing (ELB)**	Un **Application Load Balancer (ALB)** en la subred pública para distribuir el tráfico a las instancias EC2 en las subredes privadas.
Base de Datos (implícito)	**Amazon RDS** (o similar)	Desplegada en la **subred privada** para asegurar que no sea accesible directamente desde Internet.

Part 2: IAM Configuration

- **Task:**
  - Define IAM roles and policies for different components of the architecture, such as developers, admins, and application servers.
  - Ensure that each role adheres to the principle of least privilege.

Parte 2: Configuración de IAM (Principio de Mínimo Privilegio)

Definiremos **Roles de IAM** para los servicios y **Usuarios/Grupos de IAM** para el personal, siguiendo el principio de mínimo privilegio.

1. Roles de IAM para Servicios

Entidad	Rol de IAM Asociado	Permisos Clave (Mínimo Privilegio)
Instancias EC2 (Servidor Web)	`EC2AppRole`	`s3:GetObject` para leer contenido estático del bucket S3. `logs:CreateLogStream`, `logs:PutLogEvents` para enviar logs a CloudWatch.
Desarrolladores (Acceso a la Consola)	`DeveloperGroup`	`ec2:DescribeInstances` (Solo lectura). `s3:PutObject`, `s3:GetObject` (Solo en el bucket de desarrollo/staging). **Acceso de solo lectura a CloudWatch**.
Administradores (Acceso a la Consola)	`AdminGroup`	`ec2:Start`, `ec2:Stop`, `ec2:Terminate` (Solo para instancias específicas/etiquetadas). `iam:ReadOnlyAccess` (Para auditoría). **Control total sobre los servicios de red (VPC, ELB)**.
Auto Scaling Group (ASG)	`ASGServiceRole`	`ec2:CreateTags`, `ec2:RunInstances`, `ec2:TerminateInstances` (Necesario para gestionar el ciclo de vida de las instancias).

2. Ejemplo de Política para el Principio de Mínimo Privilegio

Para el `EC2AppRole`, la política debe ser muy restrictiva.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "ReadS3",
            "Effect": "Allow",
            "Action": "s3:GetObject",
            "Resource": "arn:aws:s3:::nombre-del-bucket-estatico/*"
        },
        {
            "Sid": "WriteCloudWatchLogs",
            "Effect": "Allow",
            "Action": [
                "logs:CreateLogStream",
                "logs:PutLogEvents"
            ],
            "Resource": "arn:aws:logs:region:account-id:log-group:app-logs:*"
        }
    ]
}
```

Part 3: Resource Management Strategy

- **Task:**
  - Develop a strategy for managing resources that includes auto-scaling, load balancing, and cost optimization using AWS Auto Scaling, ELB, and AWS Budgets.

Parte 3: Estrategia de Gestión de Recursos

El objetivo es lograr **escalabilidad, disponibilidad y optimización de costes**.

1. Escalabilidad y Disponibilidad

- **Auto Scaling (AWS Auto Scaling y EC2 ASG):**
  - **Grupo de Auto Escalado (ASG):** Configurado para abarcar **múltiples Zonas de Disponibilidad (AZ)** dentro de la VPC. Esto garantiza que si una AZ falla, la aplicación permanezca en línea.
  - **Métricas de Escalado:** Utilizar políticas de escalado basadas en métricas de **CPU Utilization** (superior al 60% para escalar, inferior al 30% para reducir).
  - **Costo:** Permite pagar solo por la capacidad que realmente se utiliza, escalando hacia abajo en momentos de baja demanda.
- **Balanceo de Carga (Elastic Load Balancing - ALB):**
  - Distribuye automáticamente el tráfico entrante entre todas las instancias activas del ASG, mejorando la **tolerancia a fallos** y el rendimiento.

2. Optimización de Costes

- **Tipo de Instancia:** Usar la familia de instancias **T3 o T4g (Graviton)** (si la aplicación es compatible) para cargas de trabajo de uso de CPU variable.
- **Capacidad de Reserva:**
  - Usar **Savings Plans** o **Instancias Reservadas (RI)** para la capacidad de base (la carga mínima esperada) para obtener un gran descuento.
  - Usar **Instancias bajo demanda** para la capacidad pico que requiere el escalado automático.
- **Almacenamiento:**
  - Usar **Clases de Almacenamiento de S3** (por ejemplo, S3 Standard-Infrequent Access) para contenido estático que se accede menos a menudo.
- **Presupuestos (AWS Budgets):**
  - Configurar **AWS Budgets** con alertas para ser notificado cuando los costes de EC2, S3 o ELB superen un umbral predefinido (por ejemplo, el 80% del presupuesto mensual esperado), lo que permite una acción correctiva temprana.

Part 4: Theoretical Implementation

Using the AWS services identified, outline the architecture for the web application. Describe how each component interacts with others, focusing on the flow of data and control between services. This description should detail the role of each service in the architecture, ensuring a clear understanding of their interactions and dependencies.

Parte 4: Implementación Teórica y Flujo de Datos

Arquitectura: Una arquitectura de **tres capas** (presentación/web, lógica de la aplicación y datos), con la capa de aplicación y datos alojadas en subredes privadas.

Flujo de Datos y Control

1. **Entrada de Tráfico:** Un usuario accede al **Application Load Balancer (ALB)**, ubicado en las **Subredes Públicas**.
2. **Ruta del Tráfico:** El **ALB** recibe la solicitud y la reenvía a una de las instancias **EC2** registradas y saludables, ubicadas en las **Subredes Privadas**. El **Grupo de Seguridad (Security Group)** del ALB permite el tráfico de entrada (por ejemplo, HTTP/80 o HTTPS/443).
3. **Lógica de la Aplicación (EC2):**
    - El servidor web (ej., Nginx/Apache) o la aplicación se ejecuta en la instancia **EC2**.
    - Si la aplicación necesita datos, se conecta a la base de datos **RDS** a través de la red interna de las Subredes Privadas.
    - Si la aplicación necesita contenido estático (imágenes, archivos de configuración), la instancia **EC2** asume el `EC2AppRole` y realiza una solicitud segura de `s3:GetObject` al bucket **S3**.
4. **Respuesta al Usuario:** La instancia EC2 genera la respuesta (incluyendo el contenido estático de S3) y la envía de vuelta al **ALB**. El ALB devuelve la respuesta al usuario.
5. **Monitoreo y Escalado: CloudWatch** monitorea la utilización de la CPU de las instancias EC2. Si supera el umbral, activa la política de escalado del **ASG**, que lanza una nueva instancia EC2 en una Subred Privada para manejar la carga.

Part 5: Discussion and Evaluation

- **Discussion Points:**
  - Explain the choice of services and how they interact to provide a resilient and secure infrastructure.
  - Discuss how the designed IAM policies contribute to overall security.
  - Review the resource management strategy to ensure it meets the scalability and cost-efficiency needs.

Parte 5: Discusión y Evaluación

1. Elección de Servicios y Resiliencia

- **VPC, Subredes y AZs:** El uso de **VPC** con subredes públicas y privadas, distribuidas en **múltiples Zonas de Disponibilidad (AZs)**, es fundamental para la **resiliencia**. Si una AZ se cae, las otras continúan operando.
- **ALB y ASG:** El **ALB** distribuye la carga y verifica el estado de las instancias, lo que contribuye a la **alta disponibilidad**. El **ASG** garantiza que siempre haya un número mínimo de instancias operativas y ajusta la capacidad según la demanda, asegurando la **escalabilidad**.
- **EC2 y S3: EC2** proporciona el motor de la aplicación, mientras que **S3** ofrece almacenamiento de objetos altamente duradero (99.999999999% de durabilidad) y escalable para contenido estático, desacoplando la capa de presentación de la capa de aplicación.

2. Contribución de las Políticas IAM a la Seguridad

Las políticas de **IAM** refuerzan la seguridad mediante el **Principio de Mínimo Privilegio (PoLP)**:

- **Reducción de la Superficie de Ataque:** Al dar a cada componente (instancia EC2, usuario, ASG) solo los permisos *necesarios* para realizar su trabajo, se reduce significativamente el daño potencial si un componente o una credencial se ven comprometidos. Por ejemplo, si una instancia EC2 es vulnerada, el atacante solo tiene los permisos de `EC2AppRole` (leer S3, escribir logs), lo que le impide modificar la infraestructura (como crear nuevas VPCs o eliminar bases de datos).
- **Aislamiento:** La separación de permisos entre `DeveloperGroup` y `AdminGroup` asegura que los desarrolladores no puedan realizar cambios críticos en la infraestructura de producción.

3. Evaluación de la Estrategia de Gestión de Recursos

La estrategia propuesta cumple con las necesidades de escalabilidad y rentabilidad:

Necesidad	Estrategia Utilizada	Cumplimiento
Escalabilidad	**Auto Scaling Group (ASG)** basado en la utilización de la CPU.	**Alto.** La arquitectura se adapta automáticamente a los picos de tráfico.
Disponibilidad	**ALB** y despliegue del **ASG** en **Múltiples AZs**.	**Alto.** El fallo de una instancia o de una AZ no interrumpe el servicio.
Coste-eficiencia	**Savings Plans/RI** para la carga base y **On-Demand/Spot** para picos; **AWS Budgets**.	**Medio-Alto.** Los Budgets alertan sobre desviaciones, y las RI/Savings Plans reducen el coste fijo de la capacidad mínima.

La estrategia es sólida, ya que equilibra el rendimiento y la disponibilidad (ASG, ALB) con la optimización de costes (Savings Plans, Budgets).