

1. Design Secure Architectures

30–30%

Design secure access to AWS resources

- Usa el usuario raíz solo para las tareas que lo requieren: protegido con MFA resistente al phishing y varios dispositivos, no le crees claves de acceso y usa un correo de grupo. En AWS Organizations puedes centralizar el acceso raíz y eliminar sus credenciales de las cuentas miembro. Para exigir MFA en llamadas a la API, usa la condición `aws:MultiFactorAuthPresent` con credenciales temporales de STS (principal) y usa políticas administradas reutilizables en lugar de políticas insertadas uno a uno. Para que los permisos escalen sin editar políticas, aplica ABAC con etiquetas en principales y recursos. Para delegar la creación de roles, impón un límite de permisos que permita crear solo roles en EC2 a usuarios mediante que fidesite una política de confianza nombre la otra cuenta o principal, y concede a ese principal `sts:AssumeRole`; al cambiar de rol se dejan los permisos propios y se reciben credenciales temporales. A terceros que atienden a muchos clientes exígeles un ID externo para evitar el suplente
- AWS CloudTrail tiene cada tarea de EC2 a un proveedor multicuenta, aprovisiona cuentas con Account Factory y aplica controles a las OU: preventivos (SCP, RCP), detectivos (reglas de AWS Config) y proactivos (enlaces de CloudFormation). Las SCP limitan a los principales de las cuentas miembro, incluido el usuario raíz, y las RCP limitan el acceso a los recursos
- bases de políticas basados en principales (bucket de S3, reglas de SQS, endpoints SNS, acciones IAM, etc.). indican qué principales o servicios pueden usar el recurso. Usa `aws:PrincipalOrgID` para dar acceso a toda una organización y `aws:SourceVpce` para limitar un bucket a un punto de enlace. Entre cuentas hace falta un Allow en ambos lados, y una clave de KMS siempre necesita permiso en su política de clave.

Design secure workloads and applications

- Los grupos de seguridad son el control principal: tienen estado, solo admiten reglas de permitir y pueden referenciar otros grupos. Las ACL de red actúan en la subred, no tienen estado (hay que permitir también las respuestas) y admiten reglas de denegar evaluadas por número. Un gateway NAT público en una subred pública da salida IPv4.
- Las VPC se preparan para si o si permitiendo un trabajo limitado para arrojarse las y los gateways para poder salir de una aplicación. El tipo de subred lo decide el enrutamiento: pública con ruta al gateway de internet, privada sin ella, solo VPN o aislada sin rutas fuera de la VPC. Cada subred vive en una sola AZ, los balanceadores llegan a destinos en todas.
- AWS PA Filtros y los puntos de entrada de tráfico (AWS CloudFront, Amazon EC2 Gateway, NAT ejemplo contra inyección SQL; Shield Standard protege gratis frente a DDoS de red y transporte, y Shield Advanced añade mitigación en capa 7, equipo de respuesta y protección de costes. Secrets Manager sustituye credenciales escritas en el código y las rota. Macie
- Seguridad de VPN confidencialmente local, GuardDuty de detección (aunque es un servicio más de una seguridad de cliente para redundancia) y Client VPN da acceso TLS a usuarios individuales. Direct Connect es un enlace dedicado que evita a los proveedores de internet, con interfaces privadas hacia la VPC y públicas hacia servicios como S3, pero no cifra por defecto: usa VPN de IP privada sobre él o MACsec punto a punto.

Determine appropriate data security controls

- AWS Artifact ofrece informes de cumplimiento de AWS y permite aceptar acuerdos como el BAA para toda una organización. AWS Config registra el historial de configuración y, con paquetes de conformidad, despliega reglas y correcciones como una unidad; CloudTrail registra quién hizo cada acción y Amazon Macie detecta datos confidenciales.
- **Confidencialidad** El cifrado de EBS es un ajuste por región que cifra los volúmenes nuevos. En S3, SSE-KMS con clave administrada por el cliente permite controlar la política y compartir entre cuentas, y la S3 Bucket Key reduce las llamadas a KMS. EFS y RDS solo se cifran al crearse: se replica a un sistema nuevo o se restaura desde una copia cifrada de restauración.
- **Copias** El cifrado de ALB con un certificado de ACM termina TLS en el balanceador; si TLS debe descifrarse en los destinos, se usa un NLB con agente de escucha TCP en el puerto 443. En S3 se rechaza el tráfico sin cifrar con aws:SecureTransport, y en CloudFront se fuerza HTTPS con las políticas de protocolo del visor.
- **Políticas de clave** La política de clave es el control principal de una clave de KMS: si no habilita las políticas de IAM, estas no conceden acceso. Separa administradores (gestionan, no cifran) y usuarios (operaciones criptográficas). El acceso entre cuentas exige permiso en la política de clave y en IAM de la otra cuenta, y kms:ViaService limita el uso a un servicio AWS.
- **AWS Backup** centraliza planes de copia de varios servicios y copia a otras regiones y cuentas de la organización; Vault Lock en modo de cumplimiento impide que nadie, ni el usuario raíz, borre las copias. La replicación de S3 exige control de versiones y copia solo objetos nuevos, salvo con S3 Batch Replication. La recuperación a un momento dado de DynamoDB restaura siempre a una tabla nueva.

26–26%

- El patrón de distribución SNS a SQS entrega a cada cola una copia de cada evento a varias colas que cada consumidor procesa a su ritmo. Step Functions orquesta microservicios con reintentos, historial y esperas de aprobación: los flujos
- El escalado horizontal requiere capas sin estado: las sesiones se guardan fuera, por ejemplo en ElastiCache o Standard admiten devoluciones de llamada largas. El ALB enruta por ruta a un grupo de destino por servicio, y S3 DynamoDB, Aurora Auto Scaling añade y retira réplicas de lectura, no la instancia de escritura. El escalado predictivo puede empujar a SQS para que las colas enriquezcan al menos una vez y las FIFO conservan el orden y lanzan capacidad antes de patrones diarios, y para consumidores de SQS se sigue el trabajo pendiente por instancia, procesan exactamente una vez. El tiempo de espera de visibilidad evita que otro consumidor procesa un mensaje
- Los consumidores que no se ejecutan se quiere empaquetar una aplicación con sus dependencias y escalarla por tareas. Amazon ECS en AWS Fargate evita gestionar servidores: ECS Managed Instances da acceso a GPU u otros recursos
- Los patrones de transferencia de datos entre regiones o con tráfico irregular: S3 invoca Lambda al tipos de instancia con la infraestructura gestionada por AWS. Amazon EKS conserva manifiestos y herramientas para crear objetos, SQS absorbe picos y la concurrencia máxima de la asignación de origen de eventos protege los consumidores sin reconfigurar. Para estado compartido entre tareas se monta Amazon EFS.

Designing multi-serverless and/or multi-region architectures ofrece aprobaciones con tokens de tarea, y DynamoDB

- La integridad de la infraestructura se automatiza con infraestructura como código y controles continuos: la detección bajo demanda no cobra rendimiento sin tráfico.
- Para desviaciones de CloudFormation localiza cambios hechos fuera de la plantilla. AWS Config corrige recursos no conformes con runbooks de Automation, las actualizaciones de instancias sustituyen instancias con una nueva AMI
- RDS Multi-AZ con conmutación automática. Entre regiones, Route 53 con enrutamiento de conmutación por error y los registros de los errores en el registro de eventos y como p99 de TargetResponseTime para objetivos de comprobaciones de estado, las bases de datos globales de Aurora para lecturas locales y recuperación regional, y trabajos pendientes por instancia para escalar consumidores de SQS, canarios de CloudWatch Synthetic
- Las métricas globales de DynamoDB para que los datos estén distribuidos: conexiones de Direct Connect en varias ubicaciones y dispositivos, los dos túneles de cada VPN y un segundo dispositivo de gateway de cliente. RDS para la redundancia de la zona combinando copias y protección: restauración a un momento dado de RDS desde copias para copias inmutables, y S3 Versioning con replicación entre regiones, más Batch Replication para los objetos que ya existían.

3. Design High-Performing Architectures

24-24%

Determine high-performing and/or scalable storage solutions

- El rendimiento del almacenamiento depende del tipo y la configuración: io2 Block Express ofrece latencia constante inferior al milisegundo y la mayor durabilidad de EBS; FSx for Lustre da un sistema de archivos POSIX de alto
- Para crecer sin planificar capacidad conviene elegir servicios elásticos: Amazon S3 guarda cualquier cantidad de rendimiento vinculado a S3; EFS con rendimiento Elastic absorbe picos impredecibles, y en S3, Transfer Acceleration y la carga multiparte aceleran las subidas.

Design high-performing and/or elastic compute solutions

- Desacoplar significa que productores y consumidores no dependan entre sí. Amazon SQS guarda los mensajes locales con caché respaldados por S3.
- Hasta que un consumidor los procesa y borra; el patrón de SNS a varias colas SQS entrega una copia a cada servicio; las colas FIFO mantienen el orden y el procesamiento exactamente una vez por grupo; y las colas de ALB RequestCountPerTarget para solicitudes por instancia, mensajes pendientes por instancia para colas de SQS y mensajes de SQS para el procesamiento en la carga de la instancia para memoria en bases de datos en escalado predictivo para patrones cíclicos. La supervisión detallada y el calentamiento predeterminado de instancias memoria, grupos de ubicación en cluster para HPC con mucho tráfico entre nodos, Amazon ECS con AWS Fargate hacen más fácil la escalabilidad para aplicaciones de alto rendimiento, porque la CPU se asigna en proporción; el almacenamiento de los contenedores sin gestionar servidores, y AWS Batch con instancias de spot para trabajos por lotes que toleran interrupciones al menor coste.

Determine high-performing and/or scalable database solutions

- Los recursos posteriores. AWS Compute Optimizer recomienda tamaños
- Las réplicas de lectura de RDS reciben los cambios de forma asíncrona, atienden lecturas y se pueden promover, de EC2, Auto Scaling, Lambda y otros recursos a partir del uso real.
- Pueden estar en otra región para acercar lecturas y mejorar la recuperación, pero RDS no las escala automáticamente.
- Diseñar la base de datos empieza por el patrón de acceso: RDS Proxy agrupa y reutiliza conexiones ante picos; mente, en Aurora, las réplicas comparten el volumen del cluster, se reparten con el punto de conexión de lector, DMS Schema Conversion y AWS DMS con replicación continua permiten migraciones entre motores con poca
- escala en Amazon Aurora Scaling y el motor de PostgreSQL de Aurora PostgreSQL acepta clientes y T-SQL parada: DAX da lecturas de microsegundos a DynamoDB. Si apenas cambios; y Neptune o ElastiCache cubren de SQL Server con pocos cambios, Backtrack (solo Aurora MySQL) rebobina el mismo cluster en minutos y RDS
- datos de base de datos se puede seguir el modelo de datos y el patrón de acceso: DynamoDB para clave-valor sin Custom da acceso al sistema operativo para Oracle o SQL Server. Para cambiar de motor se convierte el esquema servidor a cualquier escala (con bases globales multiactivas entre regiones). Neptune para datos muy conectados, para AWS Glue la columna en la clave de los datos con AWS Glue para datos de alto rendimiento y actualización
- MemoryDB como base de datos principal en memoria y duradera, y Aurora Serverless para cargas relacionales la caché en cada escritura y un TTL hace caducar las claves. DAX es la caché de DynamoDB compatible con su
- inmutables. En Aurora Global Database solo escribe el cluster principal.

Determine high-performing and/or scalable network and/or edge services

- OS ofrece tipos complejos y replicación, y
- En varios niveles, el balanceador va en subredes públicas y los servidores en subredes privadas de al menos dos zonas. En híbrido, una gateway de Direct Connect lleva una VIF privada a VPC de cualquier región, y una VPN sobre
- Memcached solo tipos simples.
- Para crecer sin rehacer la red se asocian bloques CIDR secundarios que no se solapan (un CIDR existente no se una VIF publica añade IPsec con una red mas constante. En global, Global Accelerator da IP estáticas anycast para puede redimensional) y se planifican los rangos con Amazon VPC IPAM. A gran escala, Transit Gateway, compartida
- los puertos de CloudFront en la red de contenido de Amazon CloudFront dentro de una zona; los de particiones con AWS RAM, sustituye a las mallas de emparejamiento, que no son transativas, y PrivateLink expone servicios a separan grupos de instancias por racks para cargas distribuidas como Cassandra, y los distribuidos aíslan pocas
- Aplicación VPC CloudFront en la red de contenido de Amazon CloudFront dentro de una zona; los de particiones con AWS RAM, sustituye a las mallas de emparejamiento, que no son transativas, y PrivateLink expone servicios a separan grupos de instancias por racks para cargas distribuidas como Cassandra, y los distribuidos aíslan pocas
- Aplicación VPC CloudFront en la red de contenido de Amazon CloudFront dentro de una zona; los de particiones con AWS RAM, sustituye a las mallas de emparejamiento, que no son transativas, y PrivateLink expone servicios a separan grupos de instancias por racks para cargas distribuidas como Cassandra, y los distribuidos aíslan pocas

Determine high-performing and/or scalable ingestion and transfer solutions

- red. Gateway Load Balancer inserta y escala dispositivos virtuales como firewalls de forma transparente.
- formatos columnares como Parquet y las particiones reducen los datos analizados. Lake Formation centraliza
- Kinesis Data Streams recoge datos en tiempo real para varias aplicaciones, y la distribución ramificada mejorada permisos detallados por fila, columna y celda en Athena, Amazon Quick y otros servicios, y SSE-KMS con una
- da a cada una rendimiento de lectura dedicado. Amazon Data Firehose entrega streaming a destinos como S3 sin
- datos administrados y reducidos en tamaño y a los clientes de las claves-S3 con cifrado y validación de
- escribir consumidores y puede convertir JSON a Parquet. Managed Service for Apache Flink procesa secuencias
- integridad. S3 File Gateway ofrece recursos NFS o SMB respaldados por S3 con caché local, y Volume Gateway
- Amazon Quick Start para el Quick Start para crear planes de migración de datos a AWS Serverless. Sin gestionar capacidad
- volumenes ISC SI. Transfer Family da SFTP, FTPS, FTP y AS2 administrados hacia S3 o EFS, y Transfer Acceleration
- en memoria para que las consultas sean rápidas y no se pague cada consulta, y la seguridad de nivel de fila filtra
- datos por usuario. Para registros indexados se usa OpenSearch Dashboards, incluido en cada dominio, y para
- En EMR sobre EC2, los nodos de tareas no guardan HDFS y son los candidatos a instancias de spot, y el escalado
- métricas operativas Amazon Managed Grafana.
- administrado ajusta el tamaño a la carga. AWS Glue ofrece ETL de Spark sin servidor, Managed Service for Apache
- Flink procesa flujos con estado y Lambda sirve para tareas cortas.

24-24%

- El rendimiento del almacenamiento depende del tipo y la configuración: io2 Block Express ofrece latencia constante

- Para crecer sin planificar capacidad conviene elegir servicios elásticos: Amazon S3 guarda cualquier cantidad de rendimiento vinculado a S3; EFS con rendimiento Elastic absorbe picos impredecibles; y en S3, Transfer Acceleration y la carga multiparte aceleran las subidas.

Design high performing and elastic compute solutions Los híbridos S3 File Gateway ofrece recursos NFS y SMB

- Desacoplar significa que productores y consumidores no dependen entre sí: Amazon SQS guarda los mensajes locales con caché respaldados por S3, hasta que un consumidor los procesa y borra; el patrón de SNS a varias colas SQS entrega una copia a cada una.
- La métrica de escalado debe reflejar lo ocupada que está cada instancia y variar en proporción a la capacidad del servicio; las colas FIFO mantienen el orden y el procesamiento exactamente una vez por grupo; y las colas de AlBRequestCountPerTarget para solicitudes por instancia, mensajes pendientes por instancia para colas de SQS y mensajes pendientes por instancia para colas de SNS.
- El escalado predictivo para patrones cíclicos. La supervisión detallada y el calentamiento predeterminado de instancias memoria, grupos de ubicación en cluster para HPC con mucho tráfico entre nodos, Amazon ECS con AWS Fargate hacen más fácil el escalado a las máquinas de procesamiento de extremo, porque la CPU se asigna en proporción; el uso para contenedores sin gestionar servidores, y AWS Batch con instancias de spot para trabajos por lotes que toleran almacenamiento efímero fija el espacio de /tmp; la simultaneidad aprovisionada elimina arranques en frío y la interrupción al menor coste.

Deservir a altas temperaturas database solutions

- Las réplicas de lectura de RDS reciben los cambios de forma asincrónica, atienden lecturas y se pueden promover; de EC2, Auto Scaling, Lambda y otros recursos a partir del uso real.
- Diseñar la base de datos empieza por el patrón de acceso: RDS Proxy agrupa y reutiliza conexiones ante picos; DMS Schema Conversion y AWS DMS con replicación continua permiten migraciones entre motores con poca pérdida.
- Escalar en Amazon Aurora ScalingGroup y promover a ElasticSearch o Aurora PostgreSQL acepta clientes y T-SQL parados: DAX da lecturas de microsegundos a DynamoDB sin apenas cambios; y Neptune o ElastiCache cubren de SQL Server con pocos cambios, Backtrack (solo Aurora MySQL) rebobina el mismo cluster en minutos y RDS Data Proxy conecta datos y acceso seguro a un modelo de datos y el patrón de acceso: DynamoDB para clave-valor sin Custom da acceso al sistema operativo para Oracle o SQL Server. Para cambiar de motor se convierte el esquema servidor a cualquier escala (con tablas globales multiactivas entre regiones), Neptune para datos muy conectados, Carbons para series de tiempo y se conectan a datos con AWS Devisión de datos a través de un canal de comunicación.
- Para DMS Schema Conversion y se conectan a datos con AWS Devisión de datos a través de un canal de comunicación.
- MemoryDB como base de datos principal en memoria y duradera y Aurora Serverless para cargas relacionales la cache en cada escritura y un TTL hace caducar las claves. DAX es la cache de DynamoDB compatible con su imprevisibles. En Aurora Global Database solo escribe el cluster principal.

- **Determina high-performance and/or scalable network algorithms** - SS ofrece tipos complejos y replicación, y

- En varios niveles, el balanceador va en subredes públicas y los servidores en subredes privadas de al menos dos Memcached solo tipos simples.
- En híbrido, una gateway de Direct Connect lleva una VIF privada a VPC de cualquier región, y una VPN sobre Para crecer sin rehacer la red se asocian bloques CIDR secundarios que no se solapan (un CIDR existente no se una VIF publica añade IPsec con una red mas constante. En global, Global Accelerator da IP estaticas anycast para puede redimensionar) y se planifican los rangos con Amazon VPC IPAM. A gran escala, Transit Gateway, compartida
- Los grupos de CloudFront al usar caché de contenido dentro de una zona; los de particiones con AWS RAM, sustituye a las mallas de emparejamiento, que no son transitivas, y PrivateLink expone servicios a separan grupos de instancias por racks para cargas distribuidas como Cassandra, y los distribuidos aíslan pocas
- Autenticar VPC con bases de datos en la capa 7 con HTTP y HTTPS: enruta por ruta o host y puede autenticar instancias críticas. Para acercar recursos a los usuarios: zonas locales cerca de grandes ciudades, Wavelength en usuarios con OIDC o Cognito en agentes HTTPS. Network Load Balancer trabaja en la capa 4 con TCP, UDP y el borde de las redes 5G y Outposts en las instalaciones del cliente.

Definición de high-availability en la nube de datos gestionada por transferWise.com. Gateway Load Balancer inserta

- Un lago de datos en S3 se cataloga con rastreadores de AWS Glue y se consulta sin servidor con Athena; los formatos columnares como Parquet y las particiones reducen los datos analizados. Lake Formation centraliza permisos detallados por fila, columna y celda en Athena, Amazon Quick y otros servicios, y SSE-KMS con una clave administrada por el cliente para la fuente principal y su autenticación hacia las claves-Sx con cifrado y validación de describir consumidores y puede convertir JSON a Parquet. Managed Service for Apache Flink procesa secuencias de integridad. S3 File Gateway ofrece recursos NFS o SMB respaldados por S3 con caché local y Volume Gateway.
 - Amazon Quick Signiant es el único proveedor de almacenamiento de objetos que ofrece video Serverless. Sin gestión de capacidad, volúmenes iSCSI, Transfer Family da SFTP, FTPS, FTP y AS2 administrados hacia S3 o EFS, y Transfer Acceleration en memoria para que las consultas sean rápidas y no se pague cada consulta, y la seguridad de nivel de fila filtra.
 - Amazon Athena multiparte permite hacer consultas operando desde el MS. Serverless hace lo mismo sin clusters que operar. datos por usuario. Para registros indexados se usa OpenSearch Dashboards, incluido en cada dominio, y para métricas operativas Amazon Managed Grafana.
- En EMR sobre EC2, los nodos de tareas no guardan HDFS y son los candidatos a instancias de spot, y el escalado administrado ajusta el tamaño a la carga. AWS Glue ofrece ETL de Spark sin servidor, Managed Service for Apache

Flink procesa flujos con estado y Lambda sirve para tareas cortas.

4. Design Cost-Optimized Architectures

20–20%

Design cost-optimized storage solutions

- S3 cobra cada solicitud, así que agrupar muchos registros pequeños en menos objetos grandes (por ejemplo, con el búfer de Firehose) abarata la ingesta, y al archivar en Glacier conviene empaquetar archivos pequeños por la sobrecarga de metadatos por objeto. Los objetos grandes se suben con carga multiparte, y con pago por solicitante.
- Dimensionar bien el almacenamiento evita pagar capacidad vacía: gp3 aprovisiona IOPS y caudal con independencia del tamaño, st1 sirve para E/S secuencial frecuente y sc1 para datos fríos. Elastic Volumes amplía un volumen de EBS sin detener la instancia, pero no lo reduce. En FSx for ONTAP se aprovisiona SSD para los datos activos y la entrada de datos a S3 desde Internet no se cobra, mientras que Transfer Acceleration añade cargos. Desde una VPC, un punto de enlace de tipo gateway llega a S3 sin NAT y sin coste adicional. DataSync copia desde NFS, SMB, HDFS u otras nubes, puede enviar solo lo modificado y escribir directamente en la clase de almacenamiento de S3 elegida.
- Hay almacenamiento que crece solo (volumen de cluster de Aurora, Amazon EFS, grupo de capacidad de FSx for ONTAP) y almacenamiento aprovisionado que necesita escalado. En RDS se activa el escalado automático de almacenamiento, que nunca supera el umbral máximo fijado; en FSx for Windows, AWS ofrece una plantilla de CloudFormation que amplía la capacidad ante una alarma de espacio libre.
- Las reglas de ciclo de vida de S3 pasan objetos a clases más baratas siguiendo un modelo en cascada y los hacen caducar al final de su vida útil, con filtros por prefijo, etiquetas o tamaño. En buckets con versiones, NoncurrentVersionExpiration limpia las versiones antiguas y AbortIncompleteMultipartUpload elimina partes de

Design cost-optimized compute solutions

- El Application Load Balancer trabaja en capa 7 con HTTP y HTTPS: enruta por ruta, host, cabeceras o consulta y admite destinos Lambda, lo que permite consolidar varias aplicaciones en un balanceador. El Network Load Balancer trabaja en capa 4 (TCP, UDP, TLS) con IP estáticas o elásticas por subred. El Gateway Load Balancer inserta y escala dispositivos virtuales de terceros de forma transparente.
- El escalado horizontal con Auto Scaling añade o quita instancias sin parar el servicio, mientras que el vertical exige detener la instancia para cambiar su tipo. Hay políticas de seguimiento de destino, por pasos, programadas y predictivas, estas para patrones cíclicos. La hibernación guarda la RAM en EBS y no cobra la instancia mientras está detenida; los warm pools mantienen instancias preinicializadas paradas o hibernadas.
- Lambda ejecuta código por eventos sin servidores y se paga por uso, ideal para cargas intermitentes. Fargate ejecuta contenedores de ECS sin administrar instancias, y Fargate Spot abarata tareas tolerantes a interrupciones.
- EC2 da control total: para ahorrar se combinan Spot (trabajo interrumpible), Savings Plans (Compute, flexible entre familias, regiones, Fargate y Lambda) y reservas.
- Producción suele requerir varias zonas de disponibilidad (Auto Scaling repartido, RDS Multi-AZ), pasar a multi-región sin necesidad es un antipatón. Los entornos que no son de producción pueden usar una sola zona y apagarse fuera de horario con Instance Scheduler on AWS o con acciones programadas de Auto Scaling. Las reservas de capacidad bajo demanda garantizan capacidad en una zona sin compromiso de plazo.
- Cada categoría de EC2 equilibra recursos distintos: uso general (equilibrado, con instancias ampliables para CPU baja o moderada), optimizadas para computación (cargas limitadas por CPU como transcodificación o servidores de juego), para memoria (bases de datos en memoria), para almacenamiento (IOPS aleatorias altas en disco local),

Design cost-optimized database solutions

- Las copias automatizadas de RDS y Aurora permiten restaurar a cualquier momento del periodo de retención, y la recuperación a un momento dado de DynamoDB ofrece copias continuas sin programar nada. Para retenciones largas, un plan de AWS Backup define frecuencia y ciclo de vida, y algunos recursos, como DynamoDB con funciones avanzadas, pueden pasar a almacenamiento en frío más barato.
- Elegir motor es elegir compatibilidad y licencias: Aurora es compatible con MySQL y PostgreSQL y evita licencias comerciales, Babelfish permite que aplicaciones de SQL Server usen Aurora PostgreSQL con pocos cambios, y extensiones como PostGIS solo existen en PostgreSQL. En SQL Server y Oracle, el modelo de licencia (incluida o propia) y la edición, como Express o Developer para pruebas, deciden gran parte del coste.
- DynamoDB bajo demanda cobra por solicitud y no exige planificar capacidad, mientras que el modo aprovisionado con capacidad reservada abarata cargas constantes y predecibles. Aurora Serverless ajusta la capacidad a la demanda y puede pausarse a 0 ACU en entornos inactivos. Para bases de datos relacionales, RDS Proxy agrupa cada tipo de datos, tiene su base de datos específica: Timestream para series temporales, Neptune para grafos, conexiones y ElastiCache descarga las lecturas repetidas sin ampliar la instancia.
- Keyspaces para cargas de Cassandra y Redshift para análisis. El almacenamiento por columnas de Redshift, o de Parquet y ORC en S3 consultados con Athena, reduce la E/S en consultas que leen pocas columnas de muchas filas, mientras que el almacenamiento por filas conviene a OLTP.
- Para cambiar de motor, DMS Schema Conversion (o AWS SCT) convierte el esquema y el código, y AWS DMS migra los datos con carga completa y replicación continua. Entre motores iguales, las migraciones homogéneas de DMS usan herramientas nativas sin servidor; de RDS for MySQL a Aurora MySQL se recomienda una réplica de

Design cost-optimized network architectures

- Una puerta de enlace NAT zonal vive en una sola zona: compartir una es más barato pero se pierde internet si cae su zona, y poner una por zona da resiliencia y evita tráfico entre zonas. La NAT regional se extiende sola por las zonas sin subred pública, la NAT privada conecta con otras redes sin internet y, para IPv6, la puerta de enlace Site-to-Site VPN crea túneles IPsec cifrados sobre internet y es la opción económica para poco tráfico o como respaldo de Direct Connect. Direct Connect usa una conexión privada que puede reducir costes y dar un rendimiento más constante con mucho tráfico, pero no cifra por defecto: se combina con VPN o MACsec. Client VPN da acceso remoto cifrado a usuarios individuales.
- La transferencia entre zonas de una región se cobra en ambos sentidos y la transferencia entre regiones se cobra a la salida de la región de origen, así que conviene mantener el tráfico intenso en la misma zona y región. Los puntos de enlace de puerta de enlace para S3 y DynamoDB evitan la NAT sin cargo adicional, la interconexión de VPC es la opción más barata entre pocas VPC y Global Accelerator da IP estáticas anycast para TCP y UDP.
- CloudFront sirve el contenido desde la ubicación perimetral de menor latencia y solo acude al origen cuando no lo tiene en caché, lo que reduce la carga y la transferencia desde el origen. Para subir la proporción de aciertos conviene un max-age largo y una clave de caché solo con los valores que cambian la respuesta; Origin Shield agrupa solicitudes al origen y CloudFront Functions resuelve tareas ligeras en el borde.
- y Cost Explorer desglosa el gasto por grupos de tipo de uso, como transferencia entre zonas o entre regiones, y por cuenta. Las etiquetas de asignación de costes y AWS Budgets reparten y vigilan el gasto por equipo, y Trusted Advisor detecta puertas de enlace NAT inactivas e IP elásticas sin asociar.