**PERF 1. How do you select the best performing architecture?**

Often, multiple approaches are required for optimal performance across a workload. Well-architected systems use multiple solutions and features to improve performance.

[Introducing The Amazon Builders’ Library (DOP328)](https://www.youtube.com/watch?v=sKRdemSirDM&ref=wellarchitected)

Afbeelding met tekst, schermafbeelding, computer, computer

Automatisch gegenereerde beschrijving

**Understand the available services and resources**Learn about and understand the wide range of services and resources available in the cloud. Identify the relevant services and configuration options for your [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html), and understand how to achieve optimal [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html).

**Define a process for architectural choices**Use internal experience and knowledge of the cloud, or external resources such as published use cases, relevant documentation, or whitepapers to define a process to choose resources and services. You should define a process that encourages experimentation and benchmarking with the services that could be used in your [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html).

**Factor cost requirements into decisions**[Workloads](https://wa.aws.amazon.com/wat.concept.workload.en.html) often have [cost](https://wa.aws.amazon.com/wat.pillar.costOptimization.en.html) requirements for operation. Use internal [cost](https://wa.aws.amazon.com/wat.pillar.costOptimization.en.html) controls to select resource types and sizes based on predicted resource need.

**Use policies or reference architectures**Maximize [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html) and efficiency by evaluating internal policies and existing reference [architectures](https://wa.aws.amazon.com/wat.concept.architecture.en.html) and using your analysis to select services and configurations for your [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html).

**Use guidance from your cloud provider or an appropriate partner**Use cloud company resources, such as solutions architects, professional services, or an appropriate partner to guide your decisions. These resources can help review and improve your [architecture](https://wa.aws.amazon.com/wat.concept.architecture.en.html) for optimal [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html).

**Benchmark existing workloads**Benchmark the [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html) of an existing [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html) to understand how it performs on the cloud. Use the data collected from benchmarks to drive architectural decisions.

**Load test your workload**Deploy your latest [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html) [architecture](https://wa.aws.amazon.com/wat.concept.architecture.en.html) on the cloud using different resource types and sizes. Monitor the deployment to capture [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html) metrics that identify bottlenecks or excess capacity. Use this [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html) information to design or improve your [architecture](https://wa.aws.amazon.com/wat.concept.architecture.en.html) and resource selection.

**PERF 2. How do you select your compute solution?**

The optimal compute solution for a workload varies based on application design, usage patterns, and configuration settings. Architectures can use different compute solutions for various components and enable different features to improve performance. Selecting the wrong compute solution for an architecture can lead to lower performance efficiency.

[Amazon EC2 foundations (CMP211-R2)](https://www.youtube.com/watch?v=kMMybKqC2Y0&ref=wellarchitected)  
 [Powering next-gen Amazon EC2: Deep dive into the Nitro system](https://www.youtube.com/watch?v=rUY-00yFlE4&ref=wellarchitected)  
 [Deliver high performance ML inference with AWS Inferentia (CMP324-R1)](https://www.youtube.com/watch?v=17r1EapAxpk&ref=wellarchitected)  
 [Optimize performance and cost for your AWS compute (CMP323-R1)](https://www.youtube.com/watch?v=zt6jYJLK8sg&ref=wellarchitected)  
 [Better, faster, cheaper compute: Cost-optimizing Amazon EC2 (CMP202-R1)](https://www.youtube.com/watch?v=_dvh4P2FVbw&ref=wellarchitected)  
 [Cloud Compute with AWS](https://aws.amazon.com/products/compute/?ref=wellarchitected)  
 [EC2 Instance Types](https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/instance-types.html?ref=wellarchitected)  
 [Processor State Control for Your EC2 Instance](https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/processor_state_control.html?ref=wellarchitected)  
 [EKS Containers: EKS Worker Nodes](https://docs.aws.amazon.com/eks/latest/userguide/worker.html?ref=wellarchitected)  
 [ECS Containers: Amazon ECS Container Instances](http://docs.aws.amazon.com/AmazonECS/latest/developerguide/ECS_instances.html?ref=wellarchitected)  
 [Functions: Lambda Function Configuration](https://docs.aws.amazon.com/lambda/latest/dg/best-practices.html?ref=wellarchitected#function-configuration)

Afbeelding met tekst, schermafbeelding, computer, computer

Automatisch gegenereerde beschrijving

**Evaluate the available compute options**Understand the [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html) characteristics of the compute-related options available to you. Know how instances, containers, and functions work, and what advantages, or disadvantages, they bring to your [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html).

**Understand the available compute configuration options**Understand how various options complement your [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html), and which configuration options are best for your system. Examples of these options include instance family, sizes, features (GPU, I/O), function sizes, container instances, and single versus multi-tenancy.

**Collect compute-related metrics**One of the best ways to understand how your compute systems are performing is to record and track the true utilization of various resources. This data can be used to make more accurate determinations about resource requirements.

**Determine the required configuration by right-sizing**Analyze the various [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html) characteristics of your [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html) and how these characteristics relate to [memory](https://wa.aws.amazon.com/wat.concept.memory.en.html), network, and CPU usage. Use this data to choose resources that best match your [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html)'s profile. For example, a [memory](https://wa.aws.amazon.com/wat.concept.memory.en.html)-intensive [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html), such as a database, could be served best by the r-family of instances. However, a bursting [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html) can benefit more from an elastic container system.

**Use the available elasticity of resources**The cloud provides the flexibility to expand or reduce your resources dynamically through a variety of mechanisms to meet changes in demand. Combined with compute-related metrics, a [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html) can automatically respond to changes and utilize the optimal set of resources to achieve its goal.

**Re-evaluate compute needs based on metrics**Use system-level metrics to identify the behavior and requirements of your [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html) over time. Evaluate your [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html)'s needs by comparing the available resources with these requirements and make changes to your compute environment to best match your [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html)'s profile. For example, over time a system might be observed to be more [memory](https://wa.aws.amazon.com/wat.concept.memory.en.html)-intensive than initially thought, so moving to a different instance family or size could improve both [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html) and efficiency.

**PERF 3. How do you select your storage solution?**

The optimal storage solution for a system varies based on the kind of access method (block, file, or object), patterns of access (random or sequential), required throughput, frequency of access (online, offline, archival), frequency of update (WORM, dynamic), and availability and durability constraints. Well-architected systems use multiple storage solutions and enable different features to improve performance and use resources efficiently.

[Cloud Storage with AWS](https://aws.amazon.com/products/storage/?ref=wellarchitected)  
 [Deep dive on Amazon EBS (STG303-R1)](https://www.youtube.com/watch?v=wsMWANWNoqQ&ref=wellarchitected)  
 [Optimize your storage performance with Amazon S3 (STG343)](https://www.youtube.com/watch?v=54AhwfME6wI&ref=wellarchitected)  
 [Amazon EC2 Storage](http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/Storage.html?ref=wellarchitected)  
 [Amazon EBS Volume Types](http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSVolumeTypes.html?ref=wellarchitected)  
 [EBS I/O Characteristics](http://docs.aws.amazon.com/AWSEC2/latest/WindowsGuide/ebs-io-characteristics.html?ref=wellarchitected)  
 [Amazon S3: Request Rate and Performance Considerations](http://docs.aws.amazon.com/AmazonS3/latest/dev/request-rate-perf-considerations.html?ref=wellarchitected)  
 [Amazon Glacier: Amazon Glacier Documentation](http://docs.aws.amazon.com/amazonglacier/latest/dev/introduction.html?ref=wellarchitected)  
 [Amazon EFS: Amazon EFS Performance](http://docs.aws.amazon.com/efs/latest/ug/performance.html?ref=wellarchitected)  
 [Amazon FSx for Lustre Performance](https://docs.aws.amazon.com/fsx/latest/LustreGuide/performance.html?ref=wellarchitected)  
 [Amazon FSx for Windows File Server Performance](https://docs.aws.amazon.com/fsx/latest/WindowsGuide/performance.html?ref=wellarchitected)

Afbeelding met tekst, schermafbeelding, computer, computer

Automatisch gegenereerde beschrijving

**Understand storage characteristics and requirements**Understand the different characteristics (for example, shareable, file size, [cache](https://wa.aws.amazon.com/wat.concept.cache.en.html) size, access patterns, [latency](https://wa.aws.amazon.com/wat.concept.latency.en.html), throughput, and persistence of data) that are required to select the services that best fit your [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html), such as object storage, block storage, file storage, or instance storage.

**Evaluate available configuration options**Evaluate the various characteristics and configuration options and how they relate to storage. Understand where and how to use [provisioned IOPS](https://wa.aws.amazon.com/wat.concept.piops.en.html), [SSDs](https://wa.aws.amazon.com/wat.concept.ssd.en.html), magnetic storage, object storage, archival storage, or ephemeral storage to optimize storage space and [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html) for your [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html).

**Make decisions based on access patterns and metrics**Choose storage systems based on your [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html)'s access patterns and configure them by determining how the [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html) accesses data. Increase storage efficiency by choosing object storage over block storage. Configure the storage options you choose to match your data access patterns.

**PERF 4. How do you select your database solution?**

The optimal database solution for a system varies based on requirements for availability, consistency, partition tolerance, latency, durability, scalability, and query capability. Many systems use different database solutions for various subsystems and enable different features to improve performance. Selecting the wrong database solution and features for a system can lead to lower performance efficiency.

[AWS purpose-built databases (DAT209-L)](https://www.youtube.com/watch?v=q81TVuV5u28&ref=wellarchitected)  
 [Amazon Aurora storage demystified: How it all works (DAT309-R)](https://www.youtube.com/watch?v=uaQEGLKtw54&ref=wellarchitected)  
 [Amazon DynamoDB deep dive: Advanced design patterns (DAT403-R1)](https://www.youtube.com/watch?v=6yqfmXiZTlM&ref=wellarchitected)  
 [Cloud Databases with AWS](https://aws.amazon.com/products/databases/?ref=wellarchitected)  
 [AWS Database Caching](https://aws.amazon.com/caching/database-caching/?ref=wellarchitected)  
 [Amazon DynamoDB Accelerator](https://aws.amazon.com/dynamodb/dax/?ref=wellarchitected)  
 [Amazon Aurora best practices](http://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/Aurora.BestPractices.html?ref=wellarchitected)  
 [Amazon Redshift performance](http://docs.aws.amazon.com/redshift/latest/dg/c_challenges_achieving_high_performance_queries.html?ref=wellarchitected)  
 [Amazon Athena top 10 performance tips](https://aws.amazon.com/blogs/big-data/top-10-performance-tuning-tips-for-amazon-athena/?ref=wellarchitected)  
 [Amazon Redshift Spectrum best practices](https://aws.amazon.com/blogs/big-data/10-best-practices-for-amazon-redshift-spectrum/?ref=wellarchitected)  
 [Amazon DynamoDB best practices](http://docs.aws.amazon.com/amazondynamodb/latest/developerguide/BestPractices.html?ref=wellarchitected)

Afbeelding met tekst, schermafbeelding, computer, computer

Automatisch gegenereerde beschrijving

**Understand data characteristics**Understand the different characteristics of data in your [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html). Determine if the [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html) requires transactions, how it interacts with data, and what its [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html) demands are. Use this data to select the best performing database approach for your [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html) (for example, [relational databases](https://wa.aws.amazon.com/wat.concept.relational.en.html), [NoSQL](https://wa.aws.amazon.com/wat.concept.nosql.en.html) Key-value, document, wide column, graph, time series, or [in-memory](https://wa.aws.amazon.com/wat.concept.in-memory.en.html) storage).

**Evaluate the available options**Evaluate the services and storage options that are available as part of the selection process for your [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html)'s storage mechanisms. Understand how, and when, to use a given service or system for data storage. Learn about available configuration options that can optimize database [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html) or efficiency, such as [provisioned IOPs](https://wa.aws.amazon.com/wat.concept.piops.en.html), [memory](https://wa.aws.amazon.com/wat.concept.memory.en.html) and compute resources, and caching.

**Collect and record database performance metrics**Use tools, libraries, and systems that record [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html) measurements related to database [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html). For example, measure transactions per second, slow queries, or system [latency](https://wa.aws.amazon.com/wat.concept.latency.en.html) introduced when accessing the database. Use this data to understand the [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html) of your database systems.

**Choose data storage based on access patterns**Use the access patterns of the [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html) to decide which services and technologies to use. For example, utilize a [relational database](https://wa.aws.amazon.com/wat.concept.relational.en.html) for [workloads](https://wa.aws.amazon.com/wat.concept.workload.en.html) that require transactions, or a key-value store that provides higher throughput but is eventually consistent where applicable.

**Optimize data storage based on access patterns and metrics**Use [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html) characteristics and access patterns that optimize how data is stored or queried to achieve the best possible [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html). Measure how optimizations such as indexing, [key distribution](https://wa.aws.amazon.com/wat.concept.key-distribution.en.html), [data warehouse](https://wa.aws.amazon.com/wat.concept.data-warehouse.en.html) design, or caching strategies impact system [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html) or overall efficiency.

**PERF 5. How do you configure your networking solution?**

The optimal network solution for a workload varies based on latency, throughput requirements, jitter, and bandwidth. Physical constraints, such as user or on-premises resources, determine location options. These constraints can be offset with edge locations or resource placement.

[Connectivity to AWS and hybrid AWS network architectures (NET317-R1)](https://www.youtube.com/watch?v=eqW6CPb58gs&ref=wellarchitected)  
 [Optimizing Network Performance for Amazon EC2 Instances (CMP308-R1)](https://www.youtube.com/watch?v=DWiwuYtIgu0&ref=wellarchitected)  
 [Networking Products with AWS](https://aws.amazon.com/products/networking/?ref=wellarchitected)  
 [Transitioning to Latency-Based Routing in Amazon Route 53](https://docs.aws.amazon.com/Route53/latest/DeveloperGuide/TutorialTransitionToLBR.html?ref=wellarchitected)  
 [Amazon EBS - Optimized Instances](https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ebs-optimized.html?ref=wellarchitected)  
 [EC2 Enhanced Networking on Linux](https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/enhanced-networking.html?ref=wellarchitected)  
 [EC2 Enhanced Networking on Windows](https://docs.aws.amazon.com/AWSEC2/latest/WindowsGuide/enhanced-networking.html?ref=wellarchitected)  
 [EC2 Placement Groups](https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/placement-groups.html?ref=wellarchitected)  
 [Enabling Enhanced Networking with the Elastic Network Adapter (ENA) on Linux Instances](https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/enhanced-networking-ena.html?ref=wellarchitected)  
 [Transit Gateway](https://docs.aws.amazon.com/vpc/latest/tgw?ref=wellarchitected)  
 [VPC Endpoints](https://docs.aws.amazon.com/vpc/latest/userguide/vpc-endpoints.html?ref=wellarchitected)  
 [VPC Flow Logs](https://docs.aws.amazon.com/vpc/latest/userguide/flow-logs.html?ref=wellarchitected)  
 [Application Load Balancer](https://docs.aws.amazon.com/elasticloadbalancing/latest/application/introduction.html?ref=wellarchitected)  
 [Network Load Balancer](https://docs.aws.amazon.com/elasticloadbalancing/latest/network/introduction.html?ref=wellarchitected)

Afbeelding met tekst, schermafbeelding, computer

Automatisch gegenereerde beschrijving

**Understand how networking impacts performance**Analyze and understand how network-related decisions impact [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html) [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html). For example, network [latency](https://wa.aws.amazon.com/wat.concept.latency.en.html) often impacts the user experience, and using the wrong protocols can starve network capacity through excessive overhead.

**Evaluate available networking features**Evaluate networking features in the cloud that may increase [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html). Measure the impact of these features through testing, metrics, and analysis. For example, take advantage of network-level features that are available to reduce [latency](https://wa.aws.amazon.com/wat.concept.latency.en.html), network distance, or jitter.

**Choose appropriately sized dedicated connectivity or VPN for hybrid workloads**When there is a requirement for on-premise communication, ensure that you have adequate bandwidth for [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html) [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html). Based on bandwidth requirements, a single dedicated connection or a single VPN might not be enough, and you must enable traffic load balancing across multiple connections.

**Leverage load-balancing and encryption offloading**Distribute traffic across multiple resources or services to allow your [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html) to take advantage of the [elasticity](https://wa.aws.amazon.com/wat.concept.elasticity.en.html) that the cloud provides. You can also use load balancing for offloading encryption termination to improve [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html) and to manage and route traffic effectively.

**Choose network protocols to improve performance**Make decisions about protocols for communication between systems and networks based on the impact to the [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html)’s [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html).

**Choose your workload’s location based on network requirements**Use the cloud location options available to reduce network [latency](https://wa.aws.amazon.com/wat.concept.latency.en.html) or improve throughput. Utilize [AWS Regions](https://wa.aws.amazon.com/wat.concept.region.en.html), [Availability Zones](https://wa.aws.amazon.com/wat.concept.az.en.html), [placement groups](https://wa.aws.amazon.com/wat.concept.placement-group.en.html), and [edge locations](https://wa.aws.amazon.com/wat.concept.edge-location.en.html) such as Outposts, Local Regions, and Wavelength, to reduce network [latency](https://wa.aws.amazon.com/wat.concept.latency.en.html) or improve throughput.

**Optimize network configuration based on metrics**Use collected and analyzed data to make informed decisions about optimizing your network configuration. Measure the impact of those changes and use the impact measurements to make future decisions.

**PERF 6. How do you evolve your workload to take advantage of new releases?**

When architecting workloads, there are finite options that you can choose from. However, over time, new technologies and approaches become available that could improve the performance of your workload.

[What's New](https://aws.amazon.com/new/?ref=wellarchitected)  
 [AWS Blog](https://aws.amazon.com/blogs/?ref=wellarchitected)  
 [Amazon Web Services YouTube Channel](https://www.youtube.com/channel/UCd6MoB9NC6uYN2grvUNT-Zg?ref=wellarchitected)  
 [AWS Online Tech Talks YouTube Channel](https://www.youtube.com/user/AWSwebinars?ref=wellarchitected)  
 [AWS Events YouTube Channel](https://www.youtube.com/channel/UCdoadna9HFHsxXWhafhNvKw?ref=wellarchitected)

Afbeelding met tekst, schermafbeelding, computer, computer

Automatisch gegenereerde beschrijving

**Stay up-to-date on new resources and services**Evaluate ways to improve [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html) as new services, design patterns, and product offerings become available. Determine which of these could improve [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html) or increase the efficiency of the [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html) through ad-hoc evaluation, internal discussion, or external analysis.

**Define a process to improve workload performance**Define a process to evaluate new services, design patterns, resource types, and configurations as they become available. For example, run existing [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html) tests on new instance offerings to determine their potential to improve your [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html).

**Evolve workload performance over time**As an organization, use the information gathered through the evaluation process to actively drive adoption of new services or resources when they become available.

**PERF 7. How do you monitor your resources to ensure they are performing?**

System performance can degrade over time. Monitor system performance to identify degradation and remediate internal or external factors, such as the operating system or application load.

[Cut through the chaos: Gain operational visibility and insight (MGT301-R1)](https://www.youtube.com/watch?v=nLYGbotqHd0&ref=wellarchitected)  
 [X-Ray Documentation](https://docs.aws.amazon.com/xray/latest/devguide/aws-xray.html?ref=wellarchitected)  
 [CloudWatch Documentation](http://docs.aws.amazon.com/AmazonCloudWatch/latest/monitoring/WhatIsCloudWatch.html?ref=wellarchitected)  
 [Monitoring, Logging, and Performance APN Partners](https://aws.amazon.com/devops/partner-solutions/?ref=wellarchitected#_Monitoring.2C_Logging.2C_and_Performance)

Afbeelding met tekst, schermafbeelding, computer, computer

Automatisch gegenereerde beschrijving

**Record performance-related metrics**Use a monitoring and observability service to record [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html)-related metrics. For example, record database transactions, slow queries, I/O [latency](https://wa.aws.amazon.com/wat.concept.latency.en.html), HTTP request throughput, service [latency](https://wa.aws.amazon.com/wat.concept.latency.en.html), or other key data.

**Analyze metrics when events or incidents occur**In response to (or during) an [event](https://wa.aws.amazon.com/wat.concept.event.en.html) or [incident](https://wa.aws.amazon.com/wat.concept.incident.en.html), use monitoring dashboards or reports to understand and diagnose the impact. These views provide insight into which portions of the [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html) are not performing as expected.

**Establish Key Performance Indicators (KPIs) to measure workload performance**Identify the KPIs that indicate whether the [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html) is performing as intended. For example, an API-based [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html) might use overall response [latency](https://wa.aws.amazon.com/wat.concept.latency.en.html) as an indication of overall [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html), and an e-commerce site might choose to use the number of purchases as its KPI.

**Use monitoring to generate alarm-based notifications**Using the [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html)-related key [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html) indicators (KPIs) that you defined, use a monitoring system that generates alarms automatically when these measurements are outside expected boundaries.

**Review metrics at regular intervals**As routine maintenance, or in response to [events](https://wa.aws.amazon.com/wat.concept.event.en.html) or [incidents](https://wa.aws.amazon.com/wat.concept.incident.en.html), review which metrics are collected. Use these reviews to identify which metrics were key in addressing issues and which additional metrics, if they were being tracked, would help to identify, address, or prevent issues.

**Monitor and alarm proactively**Use key [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html) indicators (KPIs), combined with monitoring and alerting systems, to proactively address [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html)-related issues. Use alarms to trigger automated actions to remediate issues where possible. Escalate the alarm to those able to respond if automated response is not possible. For example, you may have a system that can predict expected key [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html) indicators (KPI) values and alarm when they breach certain thresholds, or a tool that can automatically halt or roll back deployments if KPIs are outside of expected values.

**PERF 8. How do you use tradeoffs to improve performance?**

When architecting solutions, determining tradeoffs enables you to select an optimal approach. Often you can improve performance by trading consistency, durability, and space for time and latency.

[Introducing The Amazon Builders’ Library (DOP328)](https://www.youtube.com/watch?v=sKRdemSirDM&ref=wellarchitected)  
 [Amazon Builders’ Library](https://aws.amazon.com/builders-library?ref=wellarchitected)  
 [Best Practices for Implementing Amazon ElastiCache](http://docs.aws.amazon.com/AmazonElastiCache/latest/UserGuide/BestPractices.html?ref=wellarchitected)

Afbeelding met tekst, schermafbeelding, computer, computer

Automatisch gegenereerde beschrijving

**Understand the areas where performance is most critical**Understand and identify areas where increasing the [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html) of your [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html) will have a positive impact on efficiency or customer experience. For example, a website that has a large amount of customer interaction can benefit from using edge services to move content delivery closer to customers.

**Learn about design patterns and services**Research and understand the various design patterns and services that help improve [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html) [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html). As part of the analysis, identify what you could trade to achieve higher [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html). For example, using a [cache](https://wa.aws.amazon.com/wat.concept.cache.en.html) service can help to reduce the load placed on database systems; however, it requires some engineering to implement safe caching or possible introduction of eventual [consistency](https://wa.aws.amazon.com/wat.concept.consistency.en.html) in some areas.

**Identify how tradeoffs impact customers and efficiency**When evaluating [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html)-related improvements, determine which choices will impact your customers and [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html) efficiency. For example, if using a key-value data store increases system [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html), it is important to evaluate how the eventually consistent nature of it will impact customers.

**Measure the impact of performance improvements**As changes are made to improve [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html), evaluate the collected metrics and data. Use this information to determine impact that the [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html) improvement had on the [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html), the [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html)’s [components](https://wa.aws.amazon.com/wat.concept.component.en.html), and your customers. This measurement helps you understand the improvements that result from the tradeoff, and helps you determine if any negative side-effects were introduced.

**Use various performance-related strategies**Where applicable, utilize multiple strategies to improve [performance](https://wa.aws.amazon.com/wat.pillar.performance.en.html). For example, using strategies like caching data to prevent excessive network or database calls, using read-replicas for database engines to improve read rates, sharding or compressing data where possible to reduce data volumes, and buffering and streaming of results as they are available to avoid blocking.