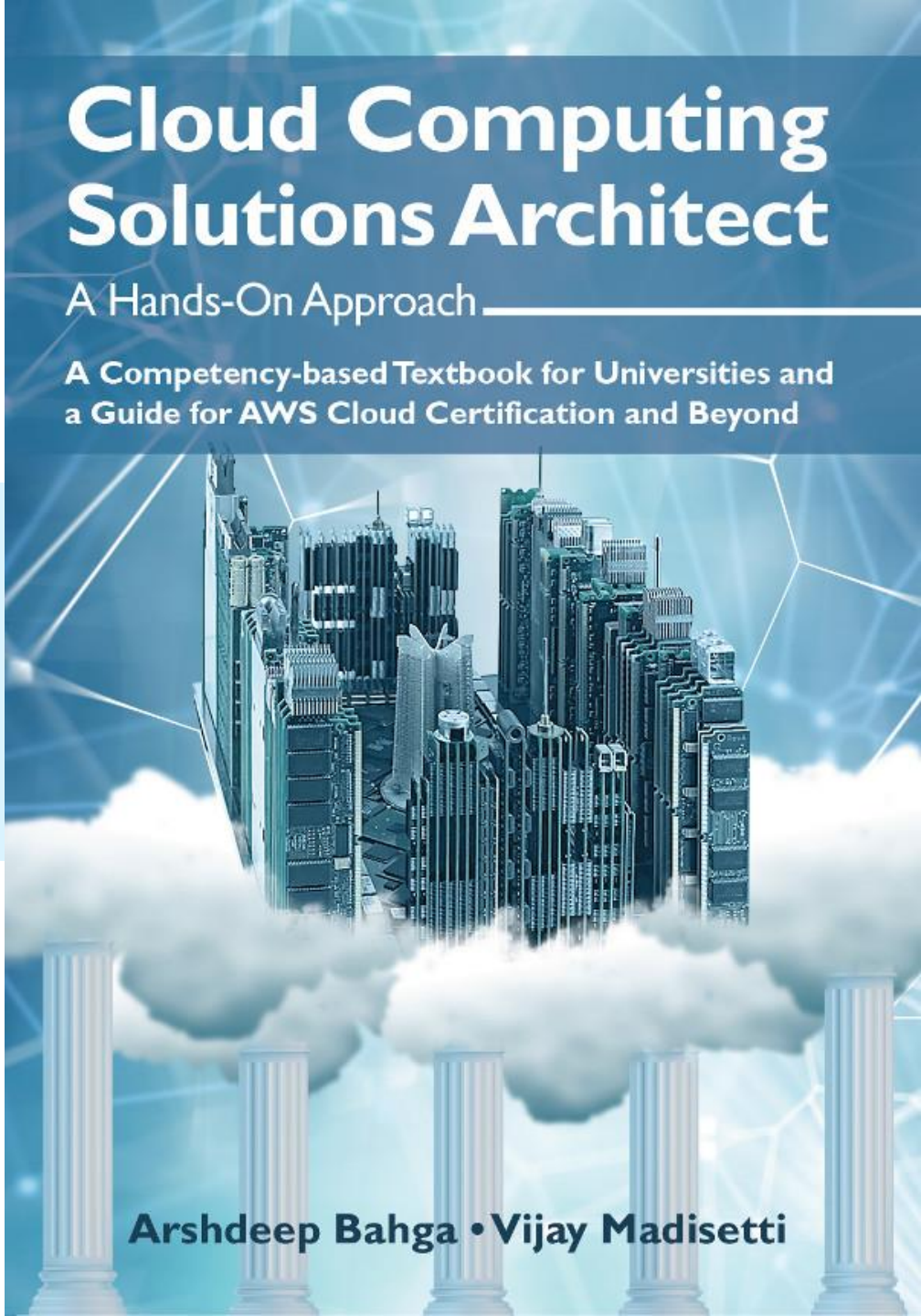
Chapter 19

Applying the Performance Efficiency Pillar



Performance Efficiency Pillar

- The Performance Efficiency pillar includes the ability to use computing resources efficiently to meet system requirements and to maintain that efficiency as demand changes and technologies evolve.
- Within the Performance Efficiency pillar, there are four best practice areas:
  - Selection
  - Review
  - Monitoring
  - Tradeoffs

Design Principles for Performance Efficiency Pillar

- Use managed and hosted services provided by the cloud vendor (such as databases, messaging queues, media transcoders) instead of hosting and managing them by yourself.
- Deploy your system across multiple regions to provide lower latency and a better experience for your users.
- Use serverless architectures where possible to remove the operational burden of managing servers and lower your cloud expenditure.
- Experiment with different options available for compute and storage.
- When selecting a cloud service or available options within a service, consider the requirements of your application such as data access patterns or data storage requirements.

Best Practice Area: Selection

- The Selection best practice area highlights the importance of selecting the optimal solution for a particular system based on application design, usage patterns, data access method (block, file, or object), data access patterns (random or sequential), and requirements for availability, consistency, partition tolerance, latency, durability, scalability and querying.
- To select the best performing architecture, you should understand the range of services and resources available and how to use them to achieve optimal performance.
- To select your compute solution, evaluate the available compute options and available configurations.
- To select your database solution, understand the different characteristics of data.
- To configure your networking solution, you should understand how networking related decisions impact performance.

Pillar IV: Performance Efficiency - Best Practice Area: Selection	
Consideration	**Best practice**
Select the best performing architecture	Understand the available services and resources
	Define a process for architectural choices
	Factor cost or budget into decisions
	Use policies or reference architectures
	Use guidance from AWS or an APN Partner
	Benchmark existing workloads
	Load test your workload
Select your compute solution	Evaluate the available compute options
	Understand the available compute configuration options
	Collect compute-related metrics
	Determine the required configuration by right-sizing
	Use the available elasticity of resources
	Re-evaluate compute needs based on metrics
Select your storage solution	Understand storage characteristics and requirements
	Evaluate available configuration options
	Make decisions based on access patterns and metrics
Select your database solution	Understand data characteristics
	Evaluate the available options
	Collect and record database performance metrics
	Choose data storage based on access patterns
	Optimize data storage based on access patterns and metrics
Configure your networking solution	Understand how networking impacts performance
	Understand available product options
	Evaluate available networking features
	Use minimal network ACLs
	Leverage encryption offloading and load-balancing
	Choose network protocols to improve performance
	Choose location based on network requirements
	Optimize network configuration based on metrics

Best Practice Area: Review

- The Review best practice area highlights the importance of reviewing the available solutions and services as newer technologies and approaches may become available that could improve the performance of your system's performance.
- To evolve your workload to take advantage of new releases, you should keep up-to-date on new resources and services.
- Evaluate new services, design patterns, resource types, and configurations as they become available.

Pillar IV: Performance Efficiency - Best Practice Area: Review	
Consideration	**Best practice**
Evolve your workload to take advantage of new releases	Keep up-to date on new resources and services
	Define a process to improve workload performance
	Evolve workload performance over time

Best Practice Area: Monitoring

- The Monitoring best practice area highlights the importance of monitoring the performance of your system so that you can take timely actions if any performance issues arise.
- To monitor your resources to ensure they are performing as expected, you should record performance-related metrics and analyze the metrics when events or incidents occur.
- Identify the KPIs for your system and use monitoring and alerting systems to address performance issues.

Pillar IV: Performance Efficiency - Best Practice Area: Monitoring	
Consideration	**Best practice**
Monitor your resources to ensure they are performing as expected	Record performance-related metrics
	Analyze metrics when events or incidents occur
	Establish KPIs to measure workload performance
	Use monitoring to generate alarm-based notifications
	Review metrics at regular intervals
	Monitor and alarm proactively

Best Practice Area: Tradeoffs

- The Tradeoffs best practice area highlights the importance of thinking about tradeoffs (such as consistency, durability, and space versus time or latency) to deliver higher performance.
- To use tradeoffs to improve performance, you should identify areas where increasing the performance will have a positive impact on efficiency or customer experience.
- You should understand various design patterns and services and use various performance related strategies such as caching, read-replicas, sharding, and data compression.

Pillar IV: Performance Efficiency - Best Practice Area: Tradeoffs	
Consideration	**Best practice**
Use tradeoffs to improve performance	Understand the areas where performance is most critical
	Learn about design patterns and services
	Identify how tradeoffs impact customers and efficiency
	Measure the impact of performance improvements
	Use various performance-related strategies

Recipe for Performance Efficiency Pillar

- In this recipe, we create a cross-region deployment of the photo gallery application comprising cross-region read replicas with multi-AZ deployments for RDS database instances.
- Operating a read replica in a different region from the master database region improves the disaster recovery of the application.
- In case of a regional disruption, you can promote the read replica to be the new master and keep the application in operation.
- Another benefit of operating a read replica in a different region is that you can scale out the application globally and serve the read queries from an AWS region that is close to the users.
- In this recipe, we use Route 53 for domain registration and DNS routing.