Well-Architected Pillars – Guidance Template

**Guidance Name: Generating Product Descriptions with Amazon Bedrock**

**Solution Domain name: Retail**

**Guidance Owner/Team: Doug Tiffan (dwtiffan), Nipun Chagari (chagarin), Dhaval Shah (shadhav), Marshall Bunch (marshbun), Kevin Bell (kebel)**

*Template Instructions*

Guidance assets should help users align to the [AWS Well-Architected Framework](https://aws.amazon.com/architecture/well-architected/?wa-lens-whitepapers.sort-by=item.additionalFields.sortDate&wa-lens-whitepapers.sort-order=desc&wa-guidance-whitepapers.sort-by=item.additionalFields.sortDate&wa-guidance-whitepapers.sort-order=desc). You must explain how your Guidance architecture helps users implement the best practices of **each** of the six Well-Architected pillars: Operational Excellence, Security, Reliability, Performance Efficiency, Cost Optimization, and Sustainability.

For each pillar box on the following pages, respond to these three questions:

* **What** service(s)\* are you using?
* **How** do these services help the user?
* **Why** are you using these services?

*\*In your responses, you may include information about services, service features, protocols, or other specific architectural components.*

Once you submit this document, the Solutions Library Technical Writers will edit your responses to create one cohesive paragraph per pillar. We have provided example responses in each pillar box.

As you write, consider the following guidelines:

* Your response to each pillar should be *unique* to each pillar. For example, your response for performance efficiency, cost optimization, sustainability should not all be about scalability.
* Avoid generalizations that could be true of multiple architectures. For example, “This Guidance uses serverless servers, which optimizes resources through dynamic scaling” could be true of many architectures. As such, it is not a helpful statement for users if they are trying to learn about your specific architecture. Instead, explain how this is applicable to your specific architecture or focus on features specific to the architecture.
* Do *not* copy and paste information from existing or past Guidance architectures.

|  |
| --- |
| **Operational Excellence**  How does your Guidance help users align to AWS best practices for operational excellence? The [operational excellence](https://docs.aws.amazon.com/wellarchitected/latest/operational-excellence-pillar/welcome.html) pillar focuses on running and monitoring systems effectively, gaining insights into operations, and continually improving processes and procedures to improve business value. Key topics include automating changes, responding to events, and defining standards to manage daily operations.  *If you need help brainstorming, consider the following concepts:*   * *How do you instrument the Guidance to understand its state and achievement of business outcomes? (*[*Prepare*](https://docs.aws.amazon.com/wellarchitected/latest/operational-excellence-pillar/prepare.html)*)* * *How did you integrate and deploy changes with the Guidance? (*[*Prepare*](https://docs.aws.amazon.com/wellarchitected/latest/operational-excellence-pillar/prepare.html)*)* * *How do you safely operate the Guidance and respond to incidents and events? (*[*Operate*](https://docs.aws.amazon.com/wellarchitected/latest/operational-excellence-pillar/operate.html)*)* * *How do you implement feedback loops within the Guidance? (*[*Evolve*](https://docs.aws.amazon.com/wellarchitected/latest/operational-excellence-pillar/evolve.html)*)*   Operational Excellence Example   * **What** service(s) are you are using to enhance operational excellence? AWS X-Ray, AWS Lambda, AWS Step Functions, and Amazon API Gateway * **How** do these services help the user with operational excellence? AWS X-Ray traces AWS Lambda functions, AWS Step Functions, and Amazon API Gateway. This helps you visualize the components of the state machine and analyze user requests as they travel through your API Gateway APIs to the underlying services. * **Why** are you using these services to support operational excellence? Visualizing and analyzing these components using AWS X-Ray helps you identify performance bottlenecks and troubleshoot requests.   **Final response after the Solutions Library Tech Writers edit this pillar:** AWS X-Ray traces AWS Lambda functions, AWS Step Functions, and Amazon API Gateway. This helps you visualize the components of the state machine and analyze user requests as they travel through your Amazon API Gateway APIs to the underlying services, so you can identify performance bottlenecks and troubleshoot requests that resulted in an error. |
| Response  What service(s) are you are using to enhance operational excellence? Amazon Bedrock, AWS Cloud Development Kit (CDK), AWS Lambda, Amazon API Gateway, and Amazon CloudWatch  How do these services help the user with operational excellence? The CDK codifies all cloud resources through an infrastructure-as-code (IaC) approach. Bedrock and Lambda both provide serverless compute, without any requirement to upgrade or patch VM images or OS versions. Lambda’s out-of-the-box integration with Amazon CloudWatch logs ensures that application logs are stored and searchable without any additional infrastructure required. Amazon API Gateway provides many capabilities to facilitate hosting a production-grade API, such as granular rate-limiting controls to ensure a consistent quality of service for all users.  Why are you using these services to support operational excellence? Using an IaC approach helps ensure that all changes to the environment are controlled and logged, preventing untested changes from making it into production and ensuring that any operator can readily determine the current state of the production system. Leveraging serverless infrastructure removes the need to patch and maintain underlying hosts. Logging and monitoring infrastructure provided by CloudWatch provides observability to help troubleshoot issues or make improvements to the system. |

|  |
| --- |
| **Security**  How does your Guidance help users align to AWS best practices for security? The [security pillar](https://docs.aws.amazon.com/wellarchitected/latest/security-pillar/welcome.html) focuses on protecting information and systems. It follows best practices for protecting data, systems, and assets in a way that improves security posture. Key topics include confidentiality and integrity of data, managing user permissions, and establishing controls to detect security events.  *If you need help brainstorming, consider the following concepts:*   * *What design decision have been factored for secure Authentication and Authorisation mechanism, for people and machine access in this Guidance? (*[*Identity and Access Management*](https://docs.aws.amazon.com/wellarchitected/latest/security-pillar/identity-and-access-management.html)*)* * *How do you protect resources in this Guidance? (*[*Infrastructure Protection*](https://docs.aws.amazon.com/wellarchitected/latest/security-pillar/infrastructure-protection.html)*)* * *How do you protect data in this Guidance? (*[*Data Protection*](https://docs.aws.amazon.com/wellarchitected/latest/security-pillar/data-protection.html)*)*   Security Example   * **What** service(s) are you are using to enhance security? Amazon CloudFront, AWS Shield Standard, AWS Identity and Access Management * **How** do these services help the user with security? Amazon CloudFront improves website security with traffic encryption and access controls and integrates with AWS Shield Standard. All AWS Identity and Access Management (IAM) policies have been scoped down to the minimum permissions required for the service to function properly. * **Why** are you using these services to support security? AWS Shield Standard defends against distributed denial-of-service (DDoS) attacks at no additional charge. By scoping IAM policies to the minimum permissions required, you limit unauthorized access to resources.   **Final response after the Solutions Library Tech Writers edit this pillar:** Amazon CloudFront improves website security with traffic encryption and access controls and can use AWS Shield Standard to defend against distributed denial-of-service (DDoS) attacks at no additional charge. All AWS Identity and Access Management (IAM) policies have been scoped down to the minimum permissions required for the service to function properly. |
| Response  What services are you are using to enhance security? Amazon API Gateway, AWS Identity and Access Management (IAM)  How do these services help the user with security? IAM’s integration with AWS Lambda is used to allow application code running in Lambda to authenticate with other services like Amazon Rekognition and Amazon Bedrock, all without requiring long lived credentials to be stored anywhere. Amazon API Gateway provides a robust outer authentication boundary, without requiring any custom authentication logic. API Gateway supports multiple authentication mechanisms, including AWS SigV4 auth or Amazon Cognito, allowing users to select the best authentication mechanism for their environment and access patterns.  Why are you using these services to support security? Using battle-tested authentication and authorization solutions like API Gateway ensures that AWS customers do not introduce vulnerabilities through custom auth logic. Avoiding long-lived credentials helps reduce the scope of any possible exposure of secrets. |

|  |
| --- |
| **Reliability**  How does your Guidance help users align to AWS best practices for reliability? The [reliability pillar](https://docs.aws.amazon.com/wellarchitected/latest/reliability-pillar/welcome.html) focuses on workloads performing their intended functions correctly and consistently and how to recover quickly from failure to meet demands. Key topics include distributed system design, recovery planning, and adapting to changing requirements.  *If you need help brainstorming, consider the following concepts:*   * *How does this Guidance implement a highly available network topology? (*[*Foundation*](https://docs.aws.amazon.com/wellarchitected/latest/reliability-pillar/foundations.html)*)* * *Are there any limits or constraints that may affect reliability, that the implementer needs to be aware of? (*[*Foundation*](https://docs.aws.amazon.com/wellarchitected/latest/reliability-pillar/foundations.html)*)* * *How does this Guidance implement a reliable application-level architecture? (for example: loosely coupled dependencies, throttling, retry limits, stateless compute) (*[*Architecture*](https://docs.aws.amazon.com/wellarchitected/latest/reliability-pillar/workload-architecture.html)*)* * *How does this Guidance implement logs and metrics and send notifications when thresholds are crossed or significant events occur? (*[*Change Management*](https://docs.aws.amazon.com/wellarchitected/latest/reliability-pillar/change-management.html)*)* * *How does this Guidance adapt to changes imposed on it, such as changes in demand? (*[*Change Management*](https://docs.aws.amazon.com/wellarchitected/latest/reliability-pillar/change-management.html)*)* * *How does this Guidance implement required changes such as deployments and configuration changes? (*[*Change Management*](https://docs.aws.amazon.com/wellarchitected/latest/reliability-pillar/change-management.html)*)* * *How does this Guidance implement data backup and recovery? (*[*Failure Management*](https://docs.aws.amazon.com/wellarchitected/latest/reliability-pillar/failure-management.html)*)* * *How does this Guidance implement resilience to failures? (*[*Failure Management*](https://docs.aws.amazon.com/wellarchitected/latest/reliability-pillar/failure-management.html)*)* * *Does this Guidance enable testing of reliability? (*[*Failure Management*](https://docs.aws.amazon.com/wellarchitected/latest/reliability-pillar/failure-management.html)*)* * *Does this Guidance enable recovery from disaster events? (*[*Failure Management*](https://docs.aws.amazon.com/wellarchitected/latest/reliability-pillar/failure-management.html)*)*   Reliability Example   * **What** service(s) are you are using to enhance reliability? Elastic Load Balancing, Amazon EC2 * **How** do these services help the user with reliability? Elastic Load Balancing (ELB) routes traffic requests from the store’s mobile application to healthy Amazon Elastic Compute Cloud (Amazon EC2) instances. * **Why** are you using these services to support reliability? ELB allows for synchronous loose coupling so that traffic won’t be directed to any Amazon EC2 instances that may be overloaded. This reduces the chance of application failure, meaning that buyers can browse the mobile application without encountering downtime errors.   **Final response after the Solutions Library Tech Writers edit this pillar:** Elastic Load Balancing (ELB) routes traffic requests from the store’s mobile application to healthy Amazon Elastic Compute Cloud (Amazon EC2) instances only so that traffic isn’t directed to instances on the brink of overload. This reduces the chance of application failure, meaning that buyers can browse the mobile storefront without encountering downtime errors. |
| Response  What services are you are using to enhance reliability? Amazon Rekognition, Amazon Bedrock, AWS Lambda, Amazon API Gateway  How do these services help the user with reliability? Serverless offerings like Rekognition, Bedrock, Lambda and API Gateway are all deployed across multiple Availability Zones by default.  Why are you using these services to support reliability? Using regionally-redundant, managed services without any single-points-of-failure provides customers with a high degree of redundancy without requiring any extra work to configure auto-scaling or recovery processes. |
| **Performance Efficiency**  How does your Guidance help users align to AWS best practices for performance efficiency? The [performance efficiency](https://docs.aws.amazon.com/wellarchitected/latest/performance-efficiency-pillar/welcome.html) pillar focuses on structured and streamlined allocation of IT and computing resources. Key topics include selecting resource types and sizes optimized for workload requirements, monitoring performance, and maintaining efficiency as business needs evolve.  *If you need help brainstorming, consider the following concepts:*   * *Why did you select those services, are they purpose built for your use-case? (*[*Selection*](https://docs.aws.amazon.com/wellarchitected/latest/performance-efficiency-pillar/selection.html)*)* * *How can the user experiment with this Guidance and optimize it based on their data? (*[*Review*](https://docs.aws.amazon.com/wellarchitected/latest/performance-efficiency-pillar/review.html)*)* * *How can the location of this Guidance be selected to decrease latency and improve performance? (*[*Network Selection*](https://docs.aws.amazon.com/wellarchitected/latest/performance-efficiency-pillar/network-architecture-selection.html)*)* * *How can this Guidance meet the workload requirements of scaling, traffic patterns, data access patterns? (*[*Performance Architecture*](https://docs.aws.amazon.com/wellarchitected/latest/performance-efficiency-pillar/performance-architecture-selection.html)*)*   Performance Efficiency Example   * **What** service(s) are you are using to enhance performance efficiency? Amazon GameLift * **How** do these services help the user with performance efficiency? Amazon GameLift allows direct client to server communication to optimize near real time performance. * **Why** are you using these services to support performance efficiency? This architecture allows developers to host game servers using Amazon GameLift across multiple AWS Regions, reducing the latency between the game client and the server.   **Final response after the Solutions Library Tech Writers edit this pillar:** Amazon GameLift allows direct client to server communication to optimize near real time performance. The architecture allows developers to host game servers across multiple AWS Regions, reducing the latency between the game client and the server, even when there is an increase in game play. |
| Response  What service(s) are you are using to enhance performance efficiency? Amazon Bedrock, Amazon Rekognition  How do these services help the user with performance efficiency? Especially for AI and ML workloads, using managed services provides excellent performance for inference without needing to test many hardware configurations and optimize models to take advantage of certain chip architectures.  Why are you using these services to support performance efficiency? Amazon Bedrock and Amazon Rekognition make it easy to get great performance with AI/ML models without investing heavily in specialized hardware. |

|  |
| --- |
| **Cost Optimization**  How does your Guidance help users align to AWS best practices for cost optimization? The [cost optimization](https://docs.aws.amazon.com/wellarchitected/latest/cost-optimization-pillar/welcome.html) pillar focuses on avoiding unnecessary costs, building and operating cost-aware workloads, and minimizing costs while maximizing return on investment. Key topics include understanding spending over time and controlling fund allocation, selecting resources of the right type and quantity, and scaling to meet business needs without overspending.  *If you need help brainstorming, consider the following concepts:*   * *How do you evaluate cost when you select services? (*[*Cost Effective Resources*](https://docs.aws.amazon.com/wellarchitected/latest/cost-optimization-pillar/cost-effective-resources.html)*)* * *How do you plan for data transfer charges? (*[*Cost Effective Resources*](https://docs.aws.amazon.com/wellarchitected/latest/cost-optimization-pillar/cost-effective-resources.html)*)* * *How do you use pricing models to reduce cost? (*[*Cost Effective Resources*](https://docs.aws.amazon.com/wellarchitected/latest/cost-optimization-pillar/cost-effective-resources.html)*)* * *How does this Guidance scales to continually match the demand and ensure that only the minimum resources required? (*[*Manage Demand and Supply Resources*](https://docs.aws.amazon.com/wellarchitected/latest/cost-optimization-pillar/manage-demand-and-supply-resources.html)*)*   Cost Optimization Example   * **What** service(s) are you are using to enhance cost optimization? AWS Glue, Amazon DynamoDB * **How** do these services help the user with cost optimization? AWS Glue jobs for ETL are batched in the largest size allowed. The DynamoDB Time to Live (TTL) feature deletes the item from your table without consuming any write throughput at a customer-defined interval. * **Why** are you using these services to support cost optimization? By batching AWS Glue jobs for ETL in the largest size, you minimize transaction costs. Because the DynamoDB TTL feature does not consume write throughput, you do not pay for additional storage.   **Final response after the Solutions Library Tech Writers edit this pillar:** AWS Glue jobs for ETL are batched in the largest size allowed to minimize transaction costs. The DynamoDB Time to Live (TTL) feature deletes the item from your table without consuming any write throughput at a customer-defined interval so that you do not pay for additional storage. |
| Response  What service(s) are you are using to enhance cost optimization? Amazon Bedrock, Amazon Rekognition, AWS Lambda  How do these services help the user with cost optimization? Compared to compute infrastructure that must be provisioned in advance, serverless options like Bedrock, Rekognition, and Lambda can be invoked completely on-demand, without any dedicated hosts running when not needed.  Why are you using these services to support cost optimization? Using completely serverless options allows the workload to scale up and down completely dynamically, without any charges accruing during periods of disuses, while also being able to scale to handle spikes in traffic. |

|  |
| --- |
| **Sustainability**  How does your Guidance help users align to AWS best practices for sustainability? The [sustainability pillar](https://docs.aws.amazon.com/wellarchitected/latest/sustainability-pillar/sustainability-pillar.html) focuses on minimizing the environmental impacts of running cloud workloads. This requires understanding the impacts of the services used, quantifying impacts through the entire workload lifecycle, and applying design principles and best practices to reduce these impacts. Key topics include a shared responsibility model for sustainability, understanding impact, and maximizing utilization to minimize required resources and reduce downstream impacts.  *If you need help brainstorming, consider the following concepts:*   * *How does this Guidance scales to continually match the load and ensure that only the minimum resources required? (*[*User Behavior Patterns*](https://docs.aws.amazon.com/wellarchitected/latest/sustainability-pillar/user-behavior-patterns.html)*)* * *How does this Guidance implement architecture patterns for maintaining consistent high utilization of deployed resources? (*[*Software and Architecture Patterns*](https://docs.aws.amazon.com/wellarchitected/latest/sustainability-pillar/software-and-architecture-patterns.html)*)* * *How does this Guidance use technologies that support data access and storage patterns? (*[*Data Patterns*](https://docs.aws.amazon.com/wellarchitected/latest/sustainability-pillar/data-patterns.html)*)* * *How does this Guidance minimize the amount of hardware needed to provision? (*[*Hardware Patterns*](https://docs.aws.amazon.com/wellarchitected/latest/sustainability-pillar/hardware-patterns.html)*)*   Sustainability Example   * **What** service(s) are you are using to enhance sustainability? Amazon FSx * **How** do these services help the user with sustainability? Amazon FSx automatically scans your file system to find and reduce redundant data through the data deduplication feature. Once enabled, this feature also compresses data by default to reduce data storage. * **Why** are you using these services to support sustainability? The Amazon FSx deduplication feature automates the removal of unneeded data to reduce storage resources, effectively minimizing your workload environmental impact.   **Final response after the Solutions Library Tech Writers edit this pillar:** The Amazon FSx data deduplication feature automatically scans your large datasets to find and reduce redundant data. Once enabled, this feature also compresses data by default to reduce data storage. By automating the removal of unneeded data, you reduce storage resources and effectively minimize your workload environmental impact. |
| Response  What service(s) are you are using to enhance sustainability? Amazon Bedrock, Amazon Rekognition, AWS Lambda  How do these services help the user with sustainability? In the same way that serverless options help reduce cost by eliminating wasteful overprovisioning of compute resources, they reduce the power consumed and environmental impact of the workload as well.  Why are you using these services to support sustainability? Using fully managed, serverless services helps minimize wasted compute resources and thus unnecessary carbon emissions as well. |