AWS Cloud Adoption Framework

Operations Perspective

November 2015

Notices

This document is provided for informational purposes only. It represents AWS's current product offerings and practices as of the date of issue of this document, which are subject to change without notice. Customers are responsible for making their own independent assessment of the information in this document and any use of AWS's products or services, each of which is provided "as is" without warranty of any kind, whether express or implied. This document does not create any warranties, representations, contractual commitments, conditions or assurances from AWS, its affiliates, suppliers or licensors. The responsibilities and liabilities of AWS to its customers are controlled by AWS agreements, and this document is not part of, nor does it modify, any agreement between AWS and its customers.

Contents

Abstract

The Amazon Web Services (AWS) Cloud Adoption Framework (CAF)[1] provides best practices and prescriptive guidance to accelerate an organization's move to cloud computing. The AWS CAF guidance is broken into a number of areas of focus that are relevant to implementing cloud-based IT systems. These focus areas are called *perspectives*. Each perspective is covered in a separate whitepaper. This whitepaper covers the Operations Perspective, which focuses on operating an AWS-enabled IT environment efficiently.

Introduction

Cloud adoption business objectives often include words and phrases such as agility, improved time to market, and cost transparency each targeting real or perceived inefficiencies in traditional IT operations models. Finding the balance between these objectives and appropriate centralized control, governance, and supportability has created a healthy tension between "the business" and IT operations. In most organizations, solving this challenge has been left to the operations team with some advice from their business partners. That advice is often to create a DevOps model or even a DevSecOps model, but overcoming the challenges posed by organizational change, large monolithic architectures, and traditional funding models can drastically slow progress.

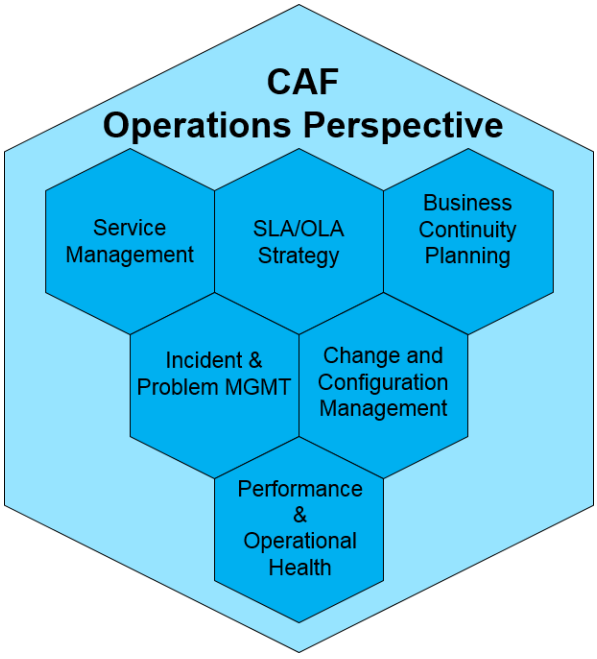

Figure 1: Components of the Operations Perspective

The Operations perspective focuses on techniques and approaches solutions architects and other professionals at AWS have observed and prescribed to change the relationship of the operating model from agility versus control to agility

with control. Agility with control is at the heart of goals and objectives associated with establishing modern development models.

When you adopt cloud capabilities in a service catalog, such as the AWS Service Catalog, you can plan up front and provide operational guardrails that development teams can operate within to provide greater freedom to innovate.

As organizations shift from on premises, to mixed (hybrid) on-premises/cloud-based environments, to fully cloud-based IT environments the operations team needs to respond by making incremental changes to the support model. Shifting to the cloud over time allows the people responsible for operating and supporting the IT environment to grow their skill and experiment with the techniques needed to operate cloud-based environments. Targeting a modern operations model such as DevSecOps (development/security/operations) can create a common goal and shared set of concepts to facilitate and enable cultural change in the organization. Achieving the DevSecOps model will then expand the operations business capability into development teams to create a shared operational responsibility.

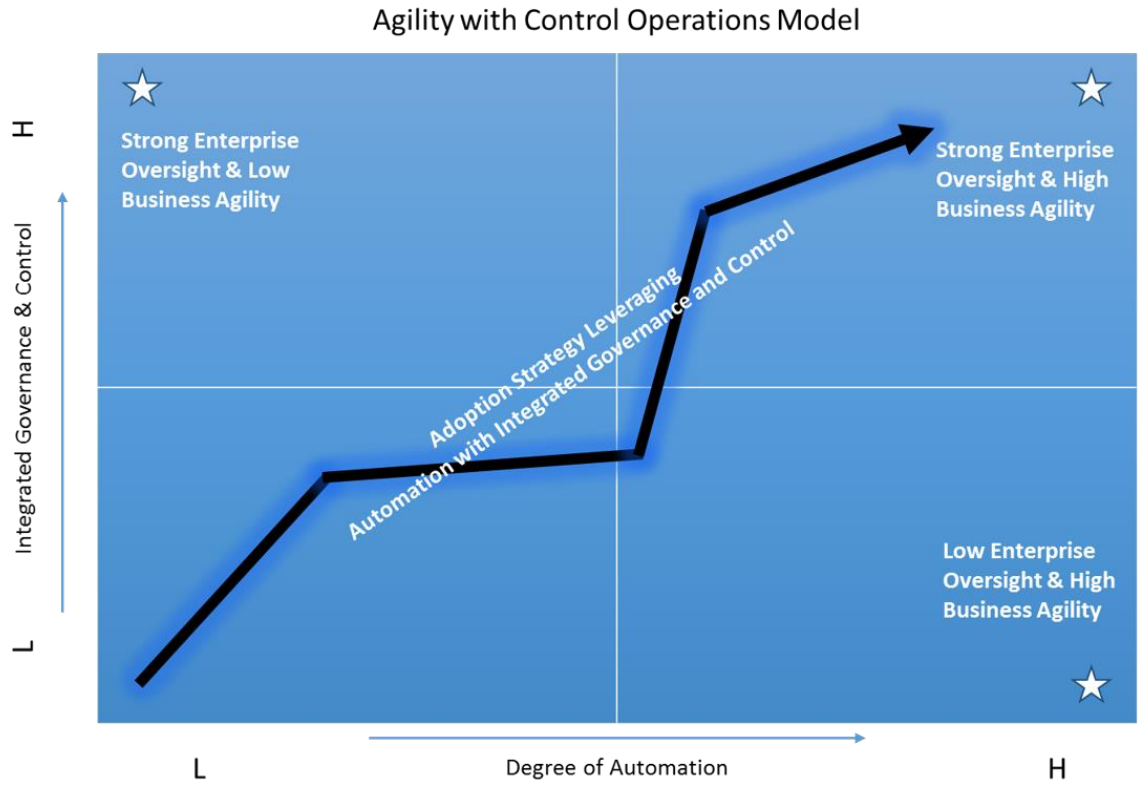

Figure 2: Creating a Balanced Environment of Agility with Control

In any case, whether you operate as traditional, agile, DevOps, or DevSecOps in the cloud you can only achieve efficiency, control, and governance through a highly automated support backbone.

In all models, the goal of operations remains the same with a single twist. The target is still to achieve levels of availability, reliability, and maintainability that are appropriate for the business. The twist is to meet these objectives while providing product development teams and end users as much control over the environment as possible.

- Availability - In on-premises IT environments having multiple data centers is required to guarantee a high level of availability. With the AWS cloud, multi-Availability Zone architectures combined with automation can exceed previous standards at a reduced cost.
- Reliability – In on-premises environments, you need to have multiple systems that are partially utilized to provide redundancy in each portion of the IT environment so that the entire system can recover from a single point of failure. You can use the same approach using AWS cloud assets to provide redundancy. However, you can achieve the same redundancy by using the same number of Amazon EC2 instances, distributed across multiple Availability Zones, using an Active/Active approach, and avoid having partially used assets.
- Maintainability – In an on-premises environment IT operations is responsible for all aspects of providing the service to the end user. With the AWS cloud, IT operations shares responsibility for maintaining the environment with AWS and often the business. This provides IT operations the same capability with less responsibility for maintaining the entire environment, as documented on the AWS Shared Responsibility[2] web page.
- End-User Control – With on-premises environments delivering self-service capabilities that promote end user control of environments requires complex development and tooling architectures. With the AWS cloud, you can use services such as AWS Service Catalog to centrally manage commonly deployed IT services and AWS CloudFormation to create and manage a collection of related AWS resources. Use AWS CodeDeploy to allow business unit IT teams to automate software deployments.

The remainder of this white paper focuses on how to achieve agility with control while evolving the enterprise culture to a modern IT operating model. We focus on each of the following components of the AWS CAF Operations perspective, providing activities (prescriptive guidance) you can use to create actionable plans to move to the cloud and to operate cloud-based solutions. :

- Service Management
- SLA/OLA Strategy
- Business Continuity Planning
- Incident and Problem Management
- Change and Configuration Management
- Performance and Operational Health

Service Management

The service management component of the AWS CAF Operations perspective promotes agility with control in an AWS environment. To be optimal service management must be proactive and supported by automation, as opposed to reactive and supported by manual human intervention. This applies both to deploying resources and automating responses to potential issues by designing for failure. The AWS cloud platform provides comprehensive automation capabilities that can save cost and time, as well as improve service quality. Repetitive manual tasks should be reduced throughout adoption to allow people on your operations teams to focus on value-add work.

To establish an environment of agility with control, operations must create an inventory of assets and prioritize new products and services as needed. This can be viewed as a grid of services and capabilities that have built-in control and are organized into a service catalog, such as the AWS Service Catalog. Development teams can become more agile if the organization makes this catalog available in a self-service model. Development teams can then provision resources as needed, which enables greater experimentation and improves time to market for new business portfolio capabilities. Development teams that automate the integration of governance and control components into services can achieve operational efficiency at higher levels.

Part of the portfolio management's charter is to define the capabilities that make up the service catalog. The operations group owns the execution on the services. Additionally, the development team should view the capabilities in the AWS Service Catalog as software as a service (SaaS). Working with catalog features using a SaaS-first approach can minimize the amount of code that must be maintained and should be a goal for the team managing the service catalog. This will minimize the amount of time spent maintaining existing code and allow more time for the development of new capabilities.

Using this approach development becomes relatively free and open and automation is used to mitigate cost of delivery. Delivery becomes a self-service tool enabling development teams to push their solutions into production. This changes the operations responsibility for delivery to managing delivery.

Additionally, the AWS Marketplace provides commercial off-the-shelf (COTS)-like solutions that can be incorporated and maintained by third-party providers, making it unnecessary to build custom solutions. This can greatly extend the capability of the service catalog with minimal effort. This also means the development teams have more autonomy to select the partners and services involved in delivery of the solution.

The service catalog then grows exponentially while dropping the requirement to host and support the larger amount of code required to provide the functionality. This also allows for strategic sourcing and leveraging best-of-breed solutions without increasing the introduction of risk into the operating environment. Use the procurement process to identify an initial set of vendors available in the AWS Marketplace that can be included in the service catalog. Approve only the set of services the vendor offers that are relevant, a good fit for the organization, and that meet governance and control expectations.

As part of approving AWS services for inclusion in your service catalog, consider identifying AWS technology partners in the AWS Marketplace. AWS conducts reviews to ensure the quality of services offered through the AWS Marketplace. This approach can be used to increase the size of the service catalog while limiting the amount of time resources will need to spend researching and approving vendor services.

Service Terms and Conditions (T&C) can be set and governed by the IT staff initially to provide a robust catalog. New services can then be justified and introduced using a governed process. These guardrails provide the value of procurement without the heavy business cost of procurement for each use. Procurement approves the suppliers initially and then the supplier's product becomes part of the catalog. As part of a set of checks and balances (and to distribute the workload needed to operationalize a service catalog), consider separating the function for approving new services from the function of maintaining the catalog.

Considerations

- **Do** create a culture of dev/sec/ops to instill ownership in all teams that take part in design, delivery, and operation of a solution.
- **Do** design for failure.
- **Do** manage configurations using agents inside of an Amazon Machine Image (AMI) to ensure governance and control.
- **Do** create a service catalog of infrastructure recipes used to quickly design and automate your infrastructure environment.
- **Do** manage bootstrap instances including additional governance and control features.
- **Don't** continue with human configuration of your infrastructure environment. Consider automating your infrastructure deployment process early to remove human error and mitigate concerns with scalability of your infrastructure.

SLA/OLA Strategy

The SLA/OLA strategy component of the AWS CAF Operations perspective encourages a shift from an IT-centric view of IT service delivery to a user-centric view. Service level agreement (SLA) and operational level agreement (OLA) standards are not new for the delivery of IT services. Traditionally these standards are established through a series of negotiations between end users or portfolio owners and the operations team. Expectations for metrics like availability, reliability, and response time have to be balanced with the investment levels that are required to achieve these standards, ensuring that the cost of the solution doesn't outweigh the solution's value.

Well-architected solutions on AWS that use capabilities such as multiple Availability Zones or regions can ease the burden of achieving SLA/OLA standards. Many organizations revisit expectations for their SLAs and OLAs early in their application or service migration projects due to the degree of change in capability.

When you follow best practices for cloud adoption, you will find that setting SLAs and OLAs for traditional things like uptime or availability will become irrelevant over time. Instead, there is a shift from evaluating service and operational levels from an IT-centric view to evaluating them from an end-user experience view. AWS encourages this shift to a user-centric or use- experience level for service and operational level definitions.

You can improve development and business unit agility by reducing the need to spend large amounts of time and money planning how to purchase, configure, and deploy high availability (HA) solutions. Following best practice guidance and reusing effective cloud architecture patterns will make HA solutions readily available to development teams.

In a cloud environment, when you automate failure detection and recovery, your systems can continually fail and recover automatically, and the failure will not be visible or apparent to the users of the IT environment. Additionally, developers can instrument code to represent what an SLA or OLA will be to ensure that developed code will meet the agreed-to end user experience level.

Considerations

- **Do** revisit SLA/OLA standard levels.
- **Do** create and follow best practices and create well-architected design standards for applications at each support tier.
- **Do** use automation to support multiple Availability Zone architectures that enable high availability (HA) applications.
- **Do** move to measuring SLA and OLA expectations from the end user's perspective.
- **Don't** "lift and shift" top-tier architectures that don't make use of advanced cloud architecture.

Business Continuity Planning

The Business Continuity Planning component of the AWS CAF Operations perspective encourages organizations to have processes and procedures in place to maintain business functionality if they are affected by a disaster. From an infrastructure and technology capability perspective, AWS provides solutions that remove cost and complexity from providing business continuity when disaster occurs. By adding additional cloud-based capability in different, geographically separated regions, any disasters that might occur can be more easily mitigated.

The IT organization should ensure that it can continue to operate if it is affected by a disaster. Disaster recovery plans and processes must be in place and tested at the rhythm required by the business to allow IT to recover from a disaster and support business operations. Operations management on the cloud focuses on proactive, end-to-end automated management. Netflix has embraced this approach. The Netflix Simian Army tools, including Chaos Monkey,[3] are an example of just how effective testing for failure can be. The Information Technology Infrastructure Library (ITIL) also provides valuable guidance on IT Service Management (ITSM) under adverse conditions. This guidance is still applicable to the cloud environment.

Depending on the extent of requirements for providing business continuity, the availability provided in a single AWS region might be sufficient to satisfy disaster recovery and business continuity requirements. For more demanding requirements, AWS service features are available that allow customers to replicate data across regions.

AWS recommends frequently shipping backup snapshots to different accounts across different geographical locations at regular intervals. In the event of a catastrophic event, you can then initiate a disaster recovery process.

You can use multiple accounts to limit the impact of compromising the primary account's authentication credentials. Setting up multiple accounts to manage your AWS services will help to enable the use of least privilege permissions and provide additional protection for maintaining business continuity.

When you set up applications as active/active rather than active/passive and launch them in multiple Availability Zones, you build a highly available solution that can withstand disaster. You don't need to require a business continuity switch from a normal operating mode to a recovery mode in a secondary site or region. However, this can be a challenge for teams supporting large legacy monolithic applications. The level of effort and investment required to transform the application architecture from monolithic to loosely-coupled can feel discouraging. However, organizations that make the investment in their tier 1 portfolio are seeing a high return on investment (ROI) through higher quality production landscapes and reduced future investment requirements for enhancements. For additional details, see the AWS Innovation at Scale[4] presentation delivered by James Hamilton at the AWS re:Invent 2014 conference.

An organization that follows these patterns could improve Recovery Time Objectives (RTO) and Recovery Point Objectives (RPO) by using a well-architected solution based on the use of AWS services.

Considerations

- **Do** create best practice well architected design standards for applications at each support tier.
- **Do** use multiple Availability Zone and multiple region architectures to meet safe distance requirements for business continuity planning activities.
- **Do** look at using AWS cloud resources for expanding existing on-premises business continuity plans at a reduced cost.
- **Do** use a multi-account approach to appropriately plan for a response related to the threat of a compromised account.
- **Do** test and validate your design, architecture, and implementation against RPO and RTO requirements with regular business continuity plan testing.
- **Do not** continue with traditional on-premises business continuity planning practices in cloud-only environments.

Incident and Problem Management

The Incident and Problem Management component of the AWS CAF Operations perspective encourages the design of an automated solution to recovery from failure. The goal of incident and problem management is to reduce the time required to resolve incidents and problems and restore normal service operation, thus minimizing the impact on end users.

With cloud-based solutions, proactively delivering resilient solutions designed for failure allows the IT team to focus on managing systemic problems rather than point failures. Setting acceptance criteria focused on self-healing solutions that have been designed for failure and validated (by proactively causing failure scenarios) to ensure proper recovery is critical for incident and problem management.

The shift in approach allows incident and problem response to become an automated portion of a solution instead of a manual process that requires a person to troubleshoot, repair, and document incidents and problems.

With the self-healing nature of cloud-based environments, many incidents and problems that occur will recover so quickly that human interaction will not be necessary as part of the recovery process. Automated ticket logging of problems and incidents should become part of the solution design and development culture. The automated nature of recovery from failure gives incident management teams time to focus on searching for trends rather than responding to point failures.

Considerations

- **Do** design for failure with fault-tolerant architectures that automate incident and problem response.
- **Do** automate ticket creation and logging for existing problem management platforms.
- **Do** create well-architected patterns that improve consistency and fault tolerance in your AWS environments.
- **Do** create a library of infrastructure recipes you can use to quickly design and automate your infrastructure environment.
- **Do** measure and monitor incident and problem response times to ensure improvement over time.
- **Do** identify systemic solutions for recurring incidents and problems.
- **Do not** continue with human error resolution for incidents and problems that happen again and again.

Change and Configuration Management

The Change and Configuration Management component of the AWS CAF Operations perspective encourages the automation of infrastructure and software deployment, including problem identification and mitigation, to significantly increase the velocity of change. As part of an AWS adoption strategy that targets an operating model of agility with control the responsibility structure for change management must be adjusted. When all teams involved in designing, creating, delivering, and operating a solution share responsibility for change, problems can be identified and mitigated in a quicker manner. The organization can move from a rhythm of several releases per year to a rhythm of several releases per day.

To achieve this distributed model, and improve agility throughout the business, all teams that are part of the delivery of applications must improve control through automation. It is important to establish appropriate checks and balances, including those for human interaction, prior to making changes that affect production environments. As the organization's cloud adoption standards mature, these checks and balances can be baked into automation, thus reducing the level of effort required by human interaction.

The AWS services distributed responsibility model provides capabilities for easily managing and monitoring the environment. You can obtain a resource inventory, a configuration history, and configuration change notifications. You can monitor a variety of metrics, collect log files, and set alarms. You can have system-wide visibility into resource utilization, application performance, and operational health. All service delivery management processes should be candidates for automation because this will improve efficiency and accuracy and reduce cost.

Hardware configuration tasks that had been manually completed by systems and network engineers can now be automated through infrastructure as code and unified under the change and configuration umbrella. Operations and maintenance shift from a focus on physical hardware to a focus on a service that is part of a catalog that must be maintained. In the AWS environment, when change is needed an engineer doesn't go to a computer room and physically configure hardware, rather the engineer configures a template that can be used to build an unlimited number of servers, networks, and other cloud resources.

Infrastructure as code provides the foundation for IT infrastructure deployment automation and should be captured in industry-standard notation and stored in configuration management tools, just like code. From there the organization can move into continuous integration / continuous deployment (CI/CD) to combine infrastructure and software deployment automation.

Change management should be focused on moving from a known good state to a new known good state using a repeatable process. Using traditional change management approaches slows the velocity of change, while automation greatly increases that velocity. To support the level of change in cloud-based environments operations change-management teams must shift their practices. They need to move away from holding change management meetings weekly or monthly and move toward adopting software development best practices based on CI/CD techniques. This approach allows for rapid fall back to a previously known good state in the event a new feature capability has an unexpected negative impact to users.

To realize the full value of a cloud environment, change can and should be delivered by the team developing and delivering the solution or product being released. The velocity of changes could shift from occurring weekly or monthly to occurring several times per day. CI/CD techniques are often the final step before achieving a DevOps or DevSecOps culture, and these techniques set the stage to achieve an optimal level of agility with control.

Considerations

- **Do** strive to create a Dev/Sec/Ops culture.
- **Do** distribute change management responsibilities out further into the organization and development teams.
- **Do** enable control and oversight through the use of appropriate AWS services and third-party monitoring, logging, and control tooling.
- **Do** create a library of infrastructure recipes you can use to quickly design and automate your infrastructure environment.
- **Do** apply CI/CD techniques as an early start on moving to a distributed control model; include appropriate source code management of infrastructure as code.
- **Do not** force development teams with modern development procedures into monthly or bi-weekly change control procedures.

Performance and Operational Health

The Performance and Operational Health component of the AWS CAF Operations perspective encourages the use of AWS services that provide the tools needed to monitor the health of cloud assets and ensure that a desired level of performance is being reached. AWS also has tools to help with forensic and root cause analysis on instances that might or might not still be in operation and accessible.

In AWS cloud environments it is easy to configure the aggregation of the information needed to monitor performance and operational health. Then the information can be analyzed using AWS services and integrated with other services that might already be used to monitor performance and health. The AWS environment can be configured to auto scale up or down when specified thresholds are exceeded, and mitigate health issues automatically.

You can use an AWS service to collect and consolidate event logs from instances. Instead of logging on to each instance and reviewing logs that capture events, logs are pushed to a centralized source and can be accessed from there. In the AWS cloud, an “off-box” solution such as this is easily achievable.

Event information can be incorporated into a console or dashboard to provide a consolidated view of near real-time performance and operational health data. (If needed, it’s possible to manually review logs.) This improves an organization’s security posture by minimizing direct access to the instances by multiple people. Using this off-box approach, people have access to centrally-sourced log information they need, but they do not have access to the actual instances.

The cost of storing logs is greatly reduced in the cloud. Organizations achieve greater agility with control by increasing the quantity of data that is logged. This enables application development teams to improve the end user experience by logging more application usage data and then driving enhancement releases based on that data. An organization that leverages business intelligence (BI) tools can simplify and expedite log analysis. Incorporate predictive analytics to shift your approach to performance and operational health from reactive to proactive. Use of BI tools can reduce the amount of time required to analyze application or service health issues.

Considerations

- **Do** make use of AWS services to extend operational and health monitoring to the cloud.
- **Do** integrate AWS monitoring with existing monitoring solutions.
- **Do** aggregate logging and monitoring information across resources by using an off-box approach.
- **Do** evaluate where increased logging on high tier applications can streamline and reduce deployment cycles through increased application intelligence.
- **Do** use a modern business intelligence tool to make aggregation more efficient and effective including predictive analytics.
- **Do not** continue with reduced logging and monitoring using on-premises methods and techniques.
- **Do not** continue with on-box logging if it can be avoided.

CAF Taxonomy and Terms

AWS created the Cloud Adoption Framework (CAF) to capture guidance and best practices from customer engagements. An AWS CAF *Perspective* represents an area of focus relevant to customers implementing cloud-based IT systems in their organizations. For instance, when a customer plans to implement a cloud solution, the Operations Perspective provides guidance on using modern, agile operations models as you move to an AWS-enabled environment.

Each CAF Perspective is made up of components and activities. A *component* is a sub-area of a Perspective that represents a specific aspect that needs attention. An *activity* provides more prescriptive guidance for customers who want to create actionable plans their organization can use to move to the cloud and to operate cloud-based solutions on an ongoing basis.

For example, *Service Management* is one component of the Operations Perspective and proactively designing for failure may be an activity within that component.

You can combine the guidance provided by the Cloud Adoption Framework (CAF) and the Cloud Adoption Methodology (CAM) to use during your journey to the AWS cloud.

Conclusion

Adopting AWS services provides customers an opportunity to have greater visibility into IT costs and to enhance delivery time of IT capability. Accelerated delivery of capability into production requires balancing speed and agility with operational control and stability. Finding the balance between these objectives and appropriate control, governance, and supportability has created a healthy tension between "the business" and IT operations. In most organizations, solving this challenge has been left to the Operations team with some advice from their business partners.

The Operations perspective focuses on techniques and approaches AWS has observed and prescribed to change the relationship from agility versus control to agility with control. The Operations perspective provides guidelines that go to the heart of goals and objectives associated with establishing modern development models.

Notes

[1] https://do.awsstatic.com/whitepapers/aws_cloud_adoption_framework.pdf

[2] https://aws.amazon.com/compliance/shared-responsibility-model/

[3] http://techblog.netflix.com/2011/07/netflix-simian-army.html

[4] https://www.youtube.com/watch?v=JIQETrFC_SQ