Building an industrial Internet of Things (IIoT) digital transformation strategy

AWS Prescriptive Guidance

Copyright © 2024 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

AWS Prescriptive Guidance: Building an industrial Internet of Things (IIoT) digital transformation strategy

Copyright © 2024 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon's trademarks and trade dress may not be used in connection with any product or service that is not Amazon's, in any manner that is likely to cause confusion among customers, or in any manner that disparages or discredits Amazon. All other trademarks not owned by Amazon are the property of their respective owners, who may or may not be affiliated with, connected to, or sponsored by Amazon.

Table of Contents

Building an industrial Internet of Things (IIoT) digital transformation strategy

Salih Bakir, Vladi Salomon, and Asim Kumar Sasmal, Amazon Web Services (AWS)

June 2022 (document history)

The *industrial Internet of Things* (IIoT) refers to the use of internet-connected sensors and devices in the industrial sectors, such as manufacturing, energy, automotive, healthcare, life sciences, and agriculture. IIoT allows telemetry data collection from equipment, machines, and field devices in an operational environment. These environments are commonly subject to industrial regulations; compute, networking, and power constraints; and inclement conditions, and all of these challenges contribute to the complexity of designing an IIoT solution.

You can use IIoT data and other enterprise, IT, and operation technology (OT) data to create additional business value, such as optimizing operations, improving productivity, and increasing availability. According to The age of analytics: Competing in a data-driven world (McKinsey Global Institute study), manufacturers can use IIoT data to reduce product development costs by up to 50%, reduce operating costs by up to 25%, and increase gross margins by up to 33%. Therefore, many organizations in the industrial sectors are starting their digital transformation journey to solve business problems by using IIoT.

Targeted business outcomes

This guide helps you create a customized roadmap for your journey, from identifying your IIoT business objectives to realizing them. This guide is based on the experiences of the Amazon Web Services (AWS) Professional Services team partnering with customers to help with their IIoT digital transformation journey. Using the phased approach described in this document, you can:

- Identify business objectives, measurable key performance indicators (KPIs), and prioritized use cases.
- Evaluate your current systems and technologies, assess the skills of your team, and identify gaps.
- Implement a repeatable and reusable blueprint for rapid deployment at scale. A blueprint is an end-to-end IIoT system reference architecture you adopt on your digital transformation journey.
- Prepare your organization for continuous innovation.

Intended audience

This guide is for IT and business executives, program and project managers, architects, product owners, and decision makers in OT, such as heads of operations, plant managers, operations managers.

Whether you are at the beginning or in the middle of your IIoT digital transformation journey, you can use the phased approach described in this guide to either build a customized plan for your journey or identify any gaps in your current plan.

Phases of an IIoT digital transformation journey

AWS Professional Services uses a phased approach to build and realize a plan for an IIoT digital transformation journey:

- Phase 1: Identifying business objectives – Clearly identify and prioritize the business opportunities or problems to address. These are the main driver and the foundation for the overall process. Your business objectives should be ambitious yet achievable, with top-down shared goals for the organization. A typical IIoT digital transformation journey is more than a single project. We recommend a holistic approach for success by thinking big, starting small, and scaling fast.
- Phase 2: Evaluating your current state – Assess your current enterprise, IT, and OT systems, and assess the skill sets of your current team to identify any gaps. Invest in technology, training, or resources for the areas where you need long-term success. For success, ensure that you have strategic alignment between your IT and OT teams.
- Phase 3: Defining a blueprint – Define your blueprint, which is your target reference architecture. It should be repeatable and reusable so that you can rapidly deploy it at scale across industrial sites. This blueprint is the foundation of your journey, and it helps you realize your business objectives by using a think big, start small, and scale fast approach.
- Phase 4: Enabling continuous innovation – When your blueprint is operational, use the collected data to identify opportunities for continuous improvement and refinement. Continue to explore solutions to maximize insights from existing and new data.

Before you begin

It is important to have executive-level commitment to the long-term investment of an IIoT digital transformation journey. Executive sponsors must be aligned to a sustainable strategy and have patience for achieving the target outcomes. According to The age of analytics: Competing in a data-driven world (McKinsey Global Institute study), "Less than one-third of all respondents say their organizations have engaged a chief digital officer to support their transformations. But those that do are 1.6 times more likely than others to report a successful digital transformation." Therefore, prior to starting your journey, make sure the executive team understands and is aligned on the investment strategy, budget, and timeline. Confirm that all of the business stakeholders across IT and OT are committed.

Phase 1: Identifying business objectives

The business case for the digital investment report (Econsultancy website) states that "The lack of a clear longer-term business case and ROI, an absence of board-level understanding and sponsorship, and the perception of digital marcoms as tactical rather than strategic are all viewed as significant challenges to securing the right levels of investment for a digital strategy by at least three-quarters of responding companies." Therefore, it is important to identify and prioritize business opportunities based on measurable KPIs.

Note that business cases might differ for different industrial sectors, such as manufacturing, energy, automotive, healthcare, life sciences, agriculture. For an example of the business impacts to manufacturing, see the Achieved impact from *Successful Digital Industrial Solutions* section of the Hitachi Vantara Solution Brief.

In this phase, you do the following:

1. Identify business challenges
2. Identify measurable KPIs
3. Identify business objectives
4. Identify use cases

The outcome of this phase is that all stakeholders are aligned on the target objectives, understand expectations, and know how success will be measured.

Identify business challenges

The first step in defining the business objectives is making a list of the current business challenges that you want to solve and the new business challenges you might encounter by implementing an IIoT digital transformation solution in your environment.

The following are some common business challenges for manufacturing and industrial companies that are in the early phases of their IIoT digital transformation journey:

- Making legacy industrial machines and equipment smart
- Extracting trapped production data for new insights
- Reduced productivity and increased downtime due to chaotic operations and slow processes for root-cause analysis

- Asset management challenges due to data silos and lack of digital tracking of assets
- Lack of near real-time monitoring at different levels of operations, such as monitoring overall equipment effectiveness (OEE), throughput, and cycle time at the plant, line, and machine level

Identify measurable KPIs

Based on the identified business challenges, you can start asking *How would I measure a successful solution to this problem?* Answering this question helps you take data-driven approach to evaluating the success of the solution.

Determine the KPIs that you will use to measure the success of your journey, and make sure that they are measurable. The following are example KPIs for that apply to a variety of industrial sectors:

- % improvement for overall equipment effectiveness (OEE) or similar KPIs
- % reduction in operational cost
- % reduction in storage and compute cost for cloud, compared to on-premises
- % reduction in unplanned downtime due to proactive monitoring and maintenance
- % accuracy in demand forecasting and inventory management
- % reduction in latency observed by business users for business intelligence (BI) reporting
- % reduction in time to make the historical data available for advanced analytics, such as machine learning
- % reduction in time to scale compute and storage
- % increase in system uptime
- % increase in productivity
- % reduction in downtime

Identify business objectives

Now that you have identified the business challenges you want to solve and decided how to measure success, you can now define your business objectives. These objectives help you answer the questions *Why is this problem worth solving?* and *Who benefits from solving this problem?* You decide on a data-driven strategy for measuring success, such as comparing current state KPIs with the target state KPIs for the particular business objective.

For each metric or KPI you want to use, rephrase it as a business objective with a target measurable value. For example, if your business challenge is *product1 is frequently out of stock due to a manual detection process* and your metric is *% reduction in latency to detect the problem*, the business objective might be *95% reduction in latency to identify possible out-of-stock situations for product1*.

Prioritize your business objectives so that the team has a clear understanding of how to prioritize resource allocation.

Identify use cases

After you have defined your business objectives, you can now focus on the use cases. *Use cases* define the exact interactions end-users have with the system, and you use them to determine how to automatically create the expected business outcome. Use cases act as the main requirements when building your blueprint.

Each use case should consist of four key elements:

- One or more end-user personas who interact with the system
- Goals for each persona
- System actions that you want to implement, as experienced by the persona
- The expected result of the system actions, as experienced by the persona

Using the example business objective *95% reduction in latency to identify possible out-of-stock situations for product1*, the following is an example use case for this objective:

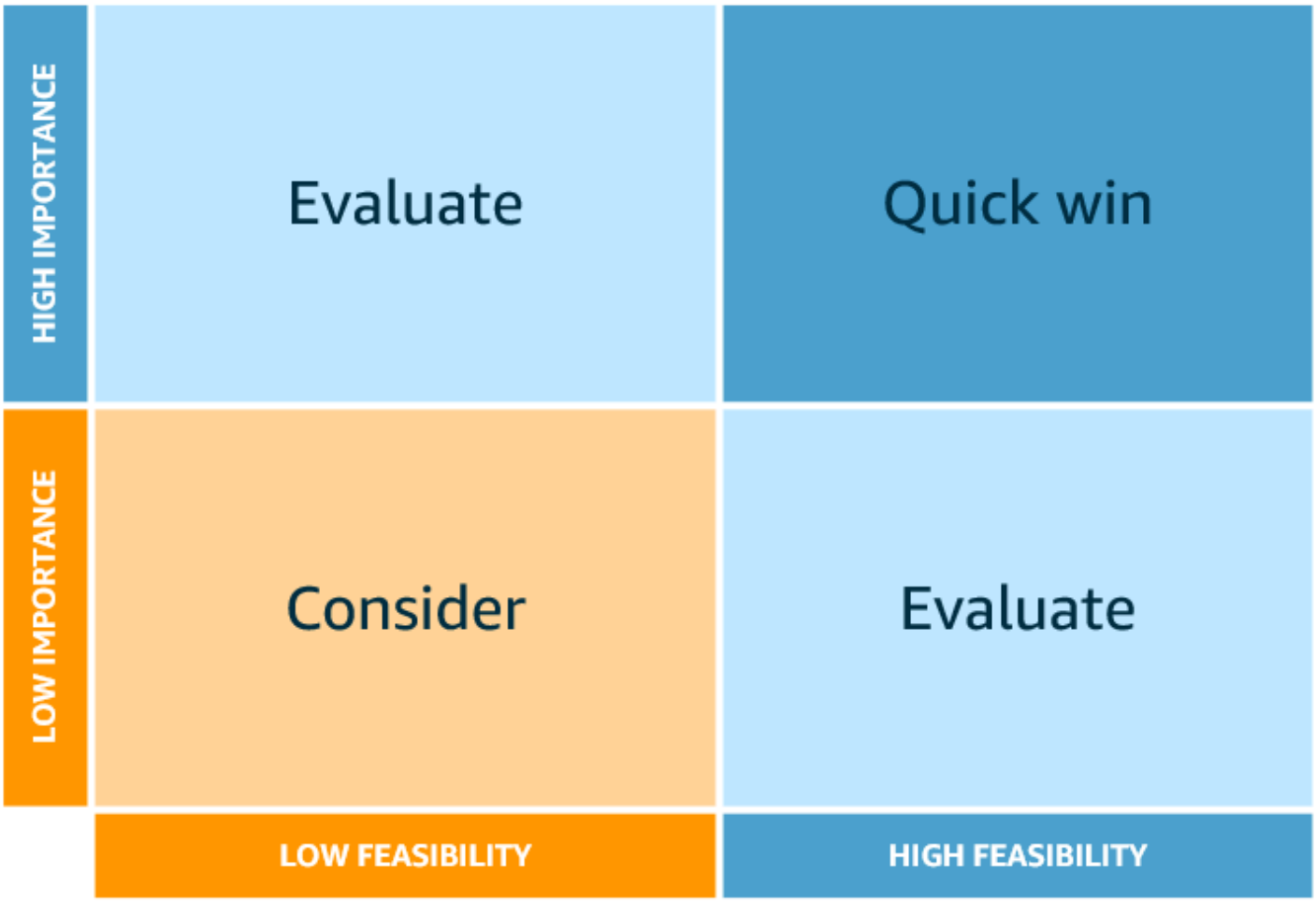
- **Persona** – Business analyst
- **Goal** – Use trend analysis to estimate out-of-stock situations for product1 within minutes
- **Action** – Use a BI reporting tool to generate a report that shows the trend
- **Result** – New solution should have 95% reduced latency to identify out-of-stock situations, compared to the previous manual process

After you have the list of use cases, evaluate them with the stakeholders based on the importance and feasibility of each use case. *Importance* is the value you expect to gain from the use case, such as its return on investment (ROI), and *feasibility* is the ease of implementation. Create a table like the following, and ask stakeholders to vote on the importance and feasibility of each use case. For

example, in the following table, use case 1 received 4 votes for high importance and 3 votes for low importance. The majority vote indicates this use case has a high importance.

	Importance		Feasibility		Majority vote	
	High	**Low**	**High**	**Low**	**Importance**	**Feasibility**
Use case 1	4	3	5	2	High	High
Use case 2	5	2	1	6	High	Low
Use case 3	1	6	3	4	Low	Low

Next, you use the voting results to prioritize the use cases. Use cases with two high ratings are considered *quick wins*. Put use case with one high rating and one low rating in the *evaluate* category, and put use cases with two low ratings in the *consider* category. The following table shows a quadrant chart you can use to visualize this categorization.



Prioritize use cases that are quick wins, and make sure that you consider dependencies. As you complete your journey, you start with the quick wins and as you progress, you can add use cases in the evaluate and consider categories, based on your budget and schedule.

Phase 2: Evaluating your current state

A complete IIoT digital transformation journey encompasses not only the IIoT-specific devices and strategy but also a holistic consideration of how those IIoT assets integrate with your IT and OT infrastructure and operations personas. Your infrastructure might be *on-premises* (local), or it might be *hybrid* of both on-premises and cloud infrastructure. Migrating your infrastructure to the cloud allows you to take full advantage of cloud-native features to improve agility, performance, and scalability.

In this phase, you do the following:

AWS Professional Services uses a set of prescriptive offerings that are tried-and-tested with other customers. We can help you assess your current state and build a phased roadmap for your target state.

Focus on people and culture

One of the key challenges is not having the proper team skill sets to enable and sustain the IIoT digital transformation. You need to upskill your team, create new roles, and hire new talent to drive success. Unlocking success in digital transformations (McKinsey & Company survey report) states that "The digital transformation success is more than three times likelier when respondents say their organizations have invested the right amount in digital talent." We recommend you consider the following topics to assess the skill sets of your current team and take actions accordingly:

- Cloud technology stack expertise
- Primary technical skill sets such as:
  - IIoT
  - Machine learning (ML)
  - Data analytics tools and methods
  - Data lakes
  - Online analytical processing (OLAP) and online transaction processing (OLTP) systems, such as SQL/NoSQL databases and data warehouses
  - Business intelligence tools
  - Real-time monitoring tools
  - Web application development, including frontend and backend
  - Operating systems, such as Linux
  - Programming languages, such as Java, Python, JavaScript
- The resources to build software products and solutions, including:
  - Business analyst
  - Product owner

  - Project manager
  - UX/UI designer
  - Software architect
  - Data architect
  - IoT architect
  - Software developer
  - Software testing and automation engineer
  - Development Operations (DevOps) engineer
  - Data scientist
  - OT subject matter experts (SMEs), such as processing engineers, production engineers, plant managers, and line managers
- The team is sized and structured according to agile principle and practices
- Partners for long-term and short-term acceleration and training

Another important point is having an innovative culture to embrace digital transformation and drive it. Because even if you have the correct strategy, processes, and tools in place, if your organizational culture does not encourage innovation and adoption, the digital transformation is less likely to be successful. Consider some of the following strategies to encourage adoption of the digital transformation in your organization:

- Having a North Star vision, values, and principles (for more information, see North Star vision)
- Having senior leadership support
- Having a roadmap that minimizes disruptions to operations
- Promoting an entrepreneurial mindset and accepting failures
- Having data-driven, customer-focused goals
- Adopting agile process and tools
- Recognizing individuals who advocate for the digital transformation, and providing them opportunities to lead or participate in the initiative
- Involving employees in the initiatives
- Providing more autonomy and flexibility for the teams
- Promoting teamwork, communication, and transparency
- Having strong and fast feedback mechanisms

Discover your current systems and technology stack

The technical capabilities of your existing systems define the scope of the future system architecture. Therefore, you need to discover your IT and OT infrastructure to understand its current technical capabilities.

Consider the following to assess the current edge infrastructure capabilities:

- Current edge architecture
- Existing IoT or IIoT systems or solutions and their capabilities
- Current data analytics and machine learning use cases, such as descriptive analytics, predictive analytics, anomaly detection, predictive and preventive maintenance, near real-time dashboard, and BI reporting
- Scale of existing solutions and future requirements
- Data sources and their capabilities for ingesting data, including:
  - Devices or tools, such as sensors, actuators, programmable logic controllers (PLCs), gateways, and OPC Unified Architecture (OPC UA) servers
  - Supported protocols for those device and tools, such as Modbus, BACnet, MQTT, and OPC UA
  - Data specifications, such as telemetry frequency, size of typical message, format, and volume
- Network infrastructure for clear isolation between OT and IT network
- Network connectivity, such as Ethernet, Wi-Fi, LoRaWAN, and 5G
- Existing historians and data storage systems
- Existing cloud connectivity options

Consider the following to assess the current cloud infrastructure capabilities:

- Current cloud architecture
- Data lakes
- Data analytics
- Data transformation
- Data service layer
- Data monitoring and BI
- Machine learning

- Web applications

Review other key considerations

In addition to the infrastructure considerations, there are also security, compliance, risk management, governance, and operational factors that you need to account for when assessing your current state. Assess the following topics in depth to address some of these considerations:

- Information security strategy that assesses and mitigates threats.
- High-availability requirements, such as recovery time objectives (RTOs) and recovery point objectives (RPOs) for the system.
- Data governance and access control.
- Identity and access management for the system.
- Data retention policies.
- Data classification and sensitivity.
- Data encryption, at-rest and in-transit.
- Compliance and regulatory requirements for processing and storing sensitive data are critical. This includes regulations such as General Data Protection Regulation (GDPR), personally identifiable information (PII), and Health Insurance Portability and Accountability Act (HIPAA).
- Service-level agreements (SLAs) for downstream data consumption and applications.
- Business risk management.
- Asset and device lifecycle management.

Phase 3: Defining a blueprint

Based on the evaluation of your current state in the previous phase, you can start building your blueprint. A blueprint is an end-to-end IIoT system reference architecture you adopt on your digital transformation journey. It serves as the foundation of your IIoT digitalization journey and helps you realize your business objectives. A blueprint:

- Is guided by your North Star vision
- Adheres to the core tenets of a successful solution framework
- Consists of repeatable and reusable building blocks

Sometimes, you might need a quick proof of concept to demonstrate value and feasibility for certain parts of the blueprint.

North Star vision

Your blueprint should be guided by your North Star vision, which is a clear, concise, and long-term goal that provides direction for making business decisions. If you don't have a North Star vision, think big when creating one. This vision generally takes 3–5 years to realize. To achieve this vision, starting small and scaling fast are the keys to success.

Core tenets of a successful solution framework

To create a unified IT and OT data backbone in your blueprint, you need a functional architecture. Based on our experiences, we've identified the following three core tenets of the solution framework:

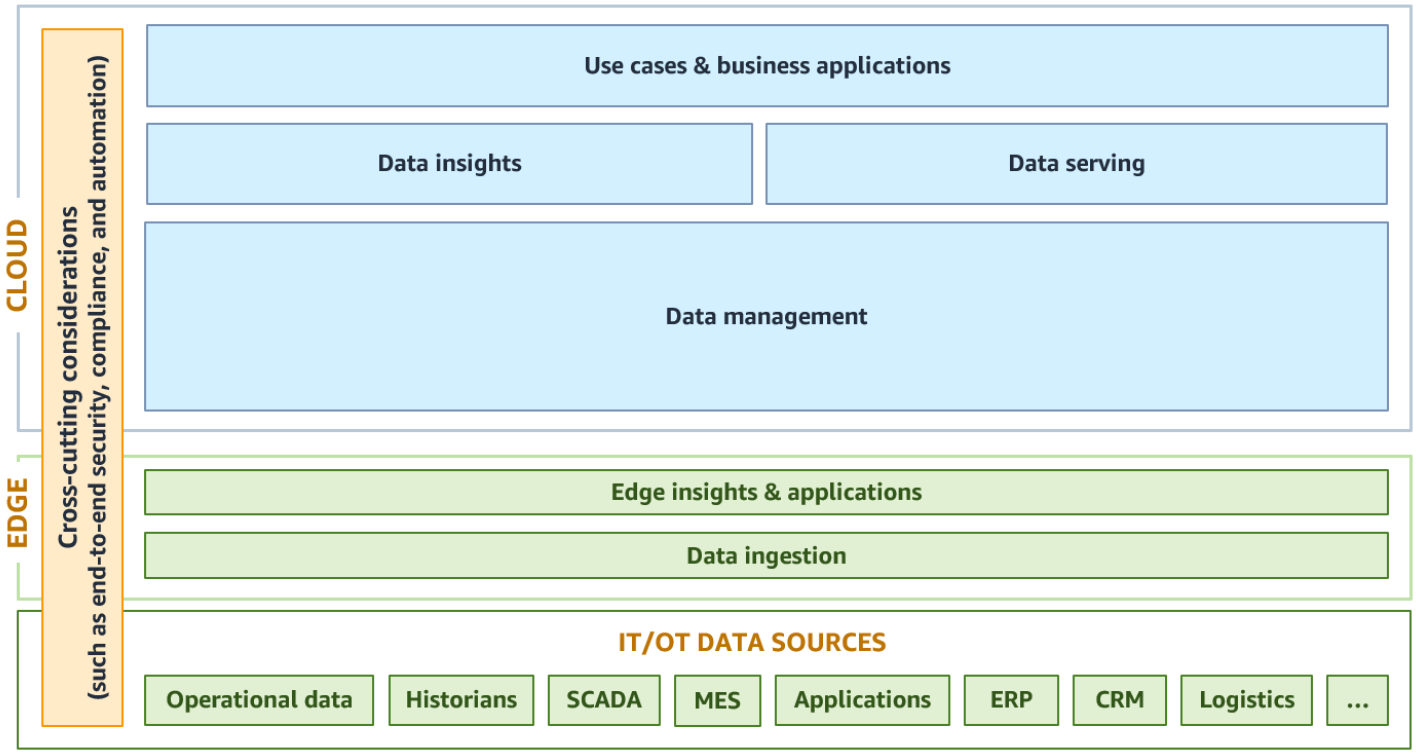
- Maximize insights
  - Democratizing access to data provides diverse insights and drives business value, such as SKU margin optimization.
  - Performing descriptive analytics on real-time or historical operational data helps you monitor KPIs, identify trends, identify potential areas of improvement, and take actions.
  - Performing diagnostic analytics on data helps you identify the root cause of operational events.
  - Performing predictive analytics on data helps you forecast future events in your business and operations.
  - Performing prescriptive analytics on data suggests multiple solutions for solving a given problem, based on descriptive and predictive analytics results.
- Minimize technical debt
  - Integrating seamlessly with the key existing IT/OT systems eliminates temporary solutions.
  - Automating the deployment pipeline removes manual process from your operations.
  - Standardizing tools prevents proliferation of tools and bespoke applications.
  - Using centralized management services to deploy standardized configurations across the environment, preventing the use of non-standard and potentially problematic configurations at the local site.

- Creating patterns for updating and deploying infrastructure automatically or with minimal intervention for repeatable tasks. Examples include updating operating systems, periodically rotating device certificates, installing patches, or scaling data storage.
- Designing and implementing repeatable and reusable patterns for rapid production deployment across sites at scale.

- Modular and future-proof blueprint
  - Designing for interoperability with existing IT/OT systems and infrastructures.
  - Designing for modularity, which helps you start small and scale fast, iteratively add new components, and select the best option for your use case.
  - Designing for flexibility with existing (*brownfield*) and new (*greenfield*) infrastructures.

Repeatable and reusable building blocks

The *building blocks* of an IIoT digital transformation journey are the various functional layers, considerations, and use cases that comprise the blueprint. The following image shows the high-level repeatable and reusable functional building blocks of a blueprint.



The following are the layers of a blueprint:

- **Data ingestion** – This edge layer collects data from various sources in your on-premises infrastructure or cloud environment. Typical IT/OT data sources might include telemetry data from supervisory control and data acquisition (SCADA) systems, distributed control systems (DCS), PLCs, secondary sensors, manufacturing execution systems (MES), software as a service (SaaS) and legacy applications, enterprise resource planning (ERP) systems, customer relationship management (CRM) systems, various supply chain systems, and data historians.
- **Edge insights and applications** – Depending on your use cases, you might want to deploy this edge layer. This layer is used to address any low latency and data residency requirements for your architecture, support production continuation when disconnected from the cloud, and enable innovation at the edge.
- **Data management** – This layer is responsible for various aspects of typical data management functions, such as:
  - Building and managing semantic data models (SDMs) for IT/OT resources for governance. Adding contexts to the machine data by using a semantic data model helps with downstream analytics for process and machine modeling.
  - Storing the data collected in the data ingestion layer. Use the data stored in this layer for processing and providing local insights, and for providing store-and-forward functionality when disconnected from the cloud.
  - Processing the data in the cloud to meet various consumption needs for end users, such as data integration, data normalization, data enrichment, data quality, data discovery, data catalog, and search.
  - Enabling a flexible data consumption service for external consumers to provide business insights.
- **Data insights** – This cloud layer is used for business insights that range from simple, such as near real-time KPI dashboards, to advanced, such as predictive maintenance, demand forecasting, and inventory management that uses the flexible data consumption service from the data management layer.
- **Data serving** – This cloud layer is used to democratize access to the data for various end users, such as various OT personas, data scientists, data engineers, and data analysts. This layer seamlessly serves data to other enterprise systems and third-party solutions to enable use cases and business applications.
- **Use cases and business applications** – This is the top layer of the architecture. This cloud layer contains the business applications and tools that address your business use cases. As needed, the applications and tools in this layer access the data and insights in the supporting layers.

- **Cross-cutting considerations** – This layer contains key non-functional requirements that apply to the data sources, edge, and cloud. This layer includes must-have elements, such as end-to-end security, configuration management, logging, compliance, and regulatory requirements. This layer helps you securely and efficiently operate your architecture, providing opportunities to enhance performance, reduce costs, or use automations that enable rapid deployment at scale across sites.

To create this unified data solution, we recommend using a unified functional architecture similar to the one presented. This holistic approach helps you think big, start small, and scale fast. Rather than taking on the entire digital transformation journey at once and making the journey impossibly difficult, you keep iterating on smaller deliverables that help you achieve your business outcomes. You might already have some of these building blocks in place today, and if so, you can reuse them.

AWS IDP solution offering

AWS Professional Services uses a tried-and-tested approach, AWS Industrial Data Platform (IDP), to discover, design, and implement a flexible and extensible unified data solution for Industry 4.0 (also known as smart manufacturing, smart factory, or smart industrial) success. The AWS IDP addresses a catalog of common use cases, such as:

- Operational and actionable KPIs for production and asset optimization, including overall equipment effectiveness (OEE), throughput, yield, and cycle time
- Automated quality and defect management solutions for predictive quality
- Predictive maintenance that reduces downtime and catastrophic equipment failures
- Energy optimization and carbon footprint reduction for sustainable manufacturing
- Supply chain optimization, including inventory management, demand forecasting, and track and trace

Your blueprint architecture might vary based on your use cases, your current state assessment, and the identified gaps. For more information about the relevant AWS services that you can use in your blueprint, see the AWS Industrial Data Platform (IDP) reference architecture.

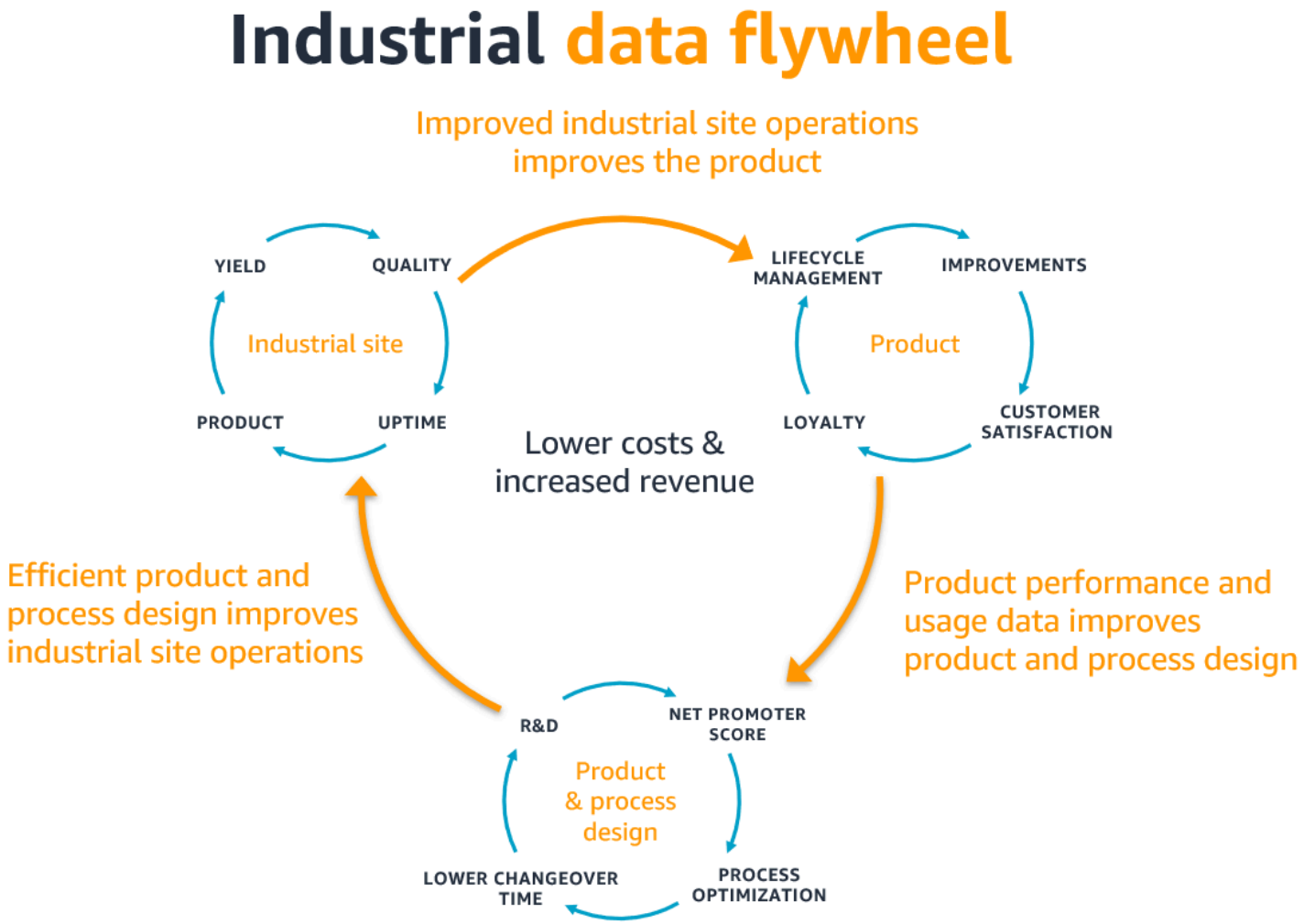
Phase 4: Enabling continuous innovation

We recommend you consider your IIoT digital transformation initiative as a journey and not a single project. After defining business KPIs, enabling the organization, establishing the necessary

skill sets, and building the blueprint; you increase the pace of innovation and explore more opportunities to build and scale throughout your business. When you have the blueprint running, you have a more transparent and data-driven business that you can monitor and track extensively.

You can use the IIoT data to identify challenges and opportunities. Furthermore, it is vital to always start with a business objective that is measurable to drive innovation. Then you can extend the blueprint to support it with a use case. It is also important to evaluate the current business model and evolve it to unlock new business opportunities so that you can be more competitive. For instance, you can shift from selling products to selling value-added services. Because IIoT capabilities provide extensive control of your products, you can extend or limit their features based on customers' requests with more competitive prices.

Finally, AWS Professional Services can help you build your own data flywheel by defining main elements to enable a virtuous cycle for innovation, just like the Amazon flywheel (YouTube video). The following image is an example of a flywheel.



Conclusion and next steps

The latest industrial revolution, the IIoT digital transformation, is unlocking new possibilities for businesses of all sizes and in every industry. Using IIoT data helps solve business problems and create new opportunities. Businesses that embrace this IIoT journey are quickly realizing the benefits of improved insights and productivity.

However, the biggest challenges in any IIoT digital transformation is not knowing where to start and not defining solid business objectives. By defining your business objectives, the KPIs you use to measure success, and your North Star vision blueprint, you can adopt a systematic approach to your journey. You assess the current situation to understand any gaps, building the blueprint to close those gaps, and then focus on continuous innovation.

In this guide, we explained a phased IIoT digital transformation journey. In the experience of AWS Professional Services, this phased approach has been effective and successful. If you want help from AWS Professional Services to accelerate your journey, complete the contact form.

When required for your success, AWS Professional Services also works jointly with our broad AWS Partner Network. For more information, see AWS IoT Competency Partners and AWS Marketplace.

For more information about how AWS can help you achieve your operational objectives, see Resources.

Resources

Customer case studies

- Yara and AWS to Digitalize Crop Nutrition Production System
- Volkswagen Group on AWS
- Coca-Cola İçecek Improves Operational Performance
- How Genie (a Terex brand) improved paint quality

AWS resources

- AWS Industrial Internet of Things
- AWS Industrial Data Platform (IDP)
- AWS for Industry
- AWS Migration Acceleration Program (MAP)
- AWS Cloud Adoption Framework (AWS CAF)
- Optimizing industrial operations with digital twins (AWS Online Tech Talks video)
- Ten security golden rules for Industrial IoT solutions (AWS blog post)

Whitepapers

- Manufacturing Transformation: Journey to the Cloud (AWS whitepaper)
- Enabling Manufacturing Innovation Through the Use of Cloud (IDC whitepaper)
- Mastering the Industrial Internet of Things (IIoT) (Roland Berger whitepaper)

Document history

The following table describes significant changes to this guide. If you want to be notified about future updates, you can subscribe to an RSS feed.

Change	Description	Date
Initial publication	—	June 20, 2022

AWS Prescriptive Guidance glossary

The following are commonly used terms in strategies, guides, and patterns provided by AWS Prescriptive Guidance. To suggest entries, please use the **Provide feedback** link at the end of the glossary.

Numbers

7 Rs

Seven common migration strategies for moving applications to the cloud. These strategies build upon the 5 Rs that Gartner identified in 2011 and consist of the following:

- Refactor/re-architect – Move an application and modify its architecture by taking full advantage of cloud-native features to improve agility, performance, and scalability. This typically involves porting the operating system and database. Example: Migrate your on-premises Oracle database to the Amazon Aurora PostgreSQL-Compatible Edition.
- Replatform (lift and reshape) – Move an application to the cloud, and introduce some level of optimization to take advantage of cloud capabilities. Example: Migrate your on-premises Oracle database to Amazon Relational Database Service (Amazon RDS) for Oracle in the AWS Cloud.
- Repurchase (drop and shop) – Switch to a different product, typically by moving from a traditional license to a SaaS model. Example: Migrate your customer relationship management (CRM) system to Salesforce.com.
- Rehost (lift and shift) – Move an application to the cloud without making any changes to take advantage of cloud capabilities. Example: Migrate your on-premises Oracle database to Oracle on an EC2 instance in the AWS Cloud.
- Relocate (hypervisor-level lift and shift) – Move infrastructure to the cloud without purchasing new hardware, rewriting applications, or modifying your existing operations. This migration scenario is specific to VMware Cloud on AWS, which supports virtual machine (VM) compatibility and workload portability between your on-premises environment and AWS. You can use the VMware Cloud Foundation technologies from your on-premises data centers when you migrate your infrastructure to VMware Cloud on AWS. Example: Relocate the hypervisor hosting your Oracle database to VMware Cloud on AWS.
- Retain (revisit) – Keep applications in your source environment. These might include applications that require major refactoring, and you want to postpone that work until a later

time, and legacy applications that you want to retain, because there's no business justification for migrating them.

- Retire – Decommission or remove applications that are no longer needed in your source environment.

A

ABAC

See attribute-based access control.

abstracted services

See managed services.

ACID

See atomicity, consistency, isolation, durability.

active-active migration

A database migration method in which the source and target databases are kept in sync (by using a bidirectional replication tool or dual write operations), and both databases handle transactions from connecting applications during migration. This method supports migration in small, controlled batches instead of requiring a one-time cutover. It's more flexible but requires more work than active-passive migration.

active-passive migration

A database migration method in which in which the source and target databases are kept in sync, but only the source database handles transactions from connecting applications while data is replicated to the target database. The target database doesn't accept any transactions during migration.

aggregate function

A SQL function that operates on a group of rows and calculates a single return value for the group. Examples of aggregate functions include `SUM` and `MAX`.

AI

See artificial intelligence.

AIOps

See artificial intelligence operations.

anonymization

The process of permanently deleting personal information in a dataset. Anonymization can help protect personal privacy. Anonymized data is no longer considered to be personal data.

anti-pattern

A frequently used solution for a recurring issue where the solution is counter-productive, ineffective, or less effective than an alternative.

application control

A security approach that allows the use of only approved applications in order to help protect a system from malware.

application portfolio

A collection of detailed information about each application used by an organization, including the cost to build and maintain the application, and its business value. This information is key to the portfolio discovery and analysis process and helps identify and prioritize the applications to be migrated, modernized, and optimized.

artificial intelligence (AI)

The field of computer science that is dedicated to using computing technologies to perform cognitive functions that are typically associated with humans, such as learning, solving problems, and recognizing patterns. For more information, see What is Artificial Intelligence?

artificial intelligence operations (AIOps)

The process of using machine learning techniques to solve operational problems, reduce operational incidents and human intervention, and increase service quality. For more information about how AIOps is used in the AWS migration strategy, see the operations integration guide.

asymmetric encryption

An encryption algorithm that uses a pair of keys, a public key for encryption and a private key for decryption. You can share the public key because it isn't used for decryption, but access to the private key should be highly restricted.

atomicity, consistency, isolation, durability (ACID)

A set of software properties that guarantee the data validity and operational reliability of a database, even in the case of errors, power failures, or other problems.

attribute-based access control (ABAC)

The practice of creating fine-grained permissions based on user attributes, such as department, job role, and team name. For more information, see ABAC for AWS in the AWS Identity and Access Management (IAM) documentation.

authoritative data source

A location where you store the primary version of data, which is considered to be the most reliable source of information. You can copy data from the authoritative data source to other locations for the purposes of processing or modifying the data, such as anonymizing, redacting, or pseudonymizing it.

Availability Zone

A distinct location within an AWS Region that is insulated from failures in other Availability Zones and provides inexpensive, low-latency network connectivity to other Availability Zones in the same Region.

AWS Cloud Adoption Framework (AWS CAF)

A framework of guidelines and best practices from AWS to help organizations develop an efficient and effective plan to move successfully to the cloud. AWS CAF organizes guidance into six focus areas called perspectives: business, people, governance, platform, security, and operations. The business, people, and governance perspectives focus on business skills and processes; the platform, security, and operations perspectives focus on technical skills and processes. For example, the people perspective targets stakeholders who handle human resources (HR), staffing functions, and people management. For this perspective, AWS CAF provides guidance for people development, training, and communications to help ready the organization for successful cloud adoption. For more information, see the AWS CAF website and the AWS CAF whitepaper.

AWS Workload Qualification Framework (AWS WQF)

A tool that evaluates database migration workloads, recommends migration strategies, and provides work estimates. AWS WQF is included with AWS Schema Conversion Tool (AWS SCT). It analyzes database schemas and code objects, application code, dependencies, and performance characteristics, and provides assessment reports.

B

BCP

See business continuity planning.

behavior graph

A unified, interactive view of resource behavior and interactions over time. You can use a behavior graph with Amazon Detective to examine failed logon attempts, suspicious API calls, and similar actions. For more information, see Data in a behavior graph in the Detective documentation.

big-endian system

A system that stores the most significant byte first. See also endianness.

binary classification

A process that predicts a binary outcome (one of two possible classes). For example, your ML model might need to predict problems such as “Is this email spam or not spam?" or "Is this product a book or a car?"

bloom filter

A probabilistic, memory-efficient data structure that is used to test whether an element is a member of a set.

branch

A contained area of a code repository. The first branch created in a repository is the *main branch*. You can create a new branch from an existing branch, and you can then develop features or fix bugs in the new branch. A branch you create to build a feature is commonly referred to as a *feature branch*. When the feature is ready for release, you merge the feature branch back into the main branch. For more information, see About branches (GitHub documentation).

break-glass access

In exceptional circumstances and through an approved process, a quick means for a user to gain access to an AWS account that they don't typically have permissions to access. For more information, see the Implement break-glass procedures indicator in the AWS Well-Architected guidance.

brownfield strategy

The existing infrastructure in your environment. When adopting a brownfield strategy for a system architecture, you design the architecture around the constraints of the current systems and infrastructure. If you are expanding the existing infrastructure, you might blend brownfield and greenfield strategies.

buffer cache

The memory area where the most frequently accessed data is stored.

business capability

What a business does to generate value (for example, sales, customer service, or marketing). Microservices architectures and development decisions can be driven by business capabilities. For more information, see the Organized around business capabilities section of the Running containerized microservices on AWS whitepaper.

business continuity planning (BCP)

A plan that addresses the potential impact of a disruptive event, such as a large-scale migration, on operations and enables a business to resume operations quickly.

C

CAF

See AWS Cloud Adoption Framework.

CCoE

See Cloud Center of Excellence.

CDC

See change data capture.

change data capture (CDC)

The process of tracking changes to a data source, such as a database table, and recording metadata about the change. You can use CDC for various purposes, such as auditing or replicating changes in a target system to maintain synchronization.

chaos engineering

Intentionally introducing failures or disruptive events to test a system's resilience. You can use AWS Fault Injection Service (AWS FIS) to perform experiments that stress your AWS workloads and evaluate their response.

CI/CD

See continuous integration and continuous delivery.

classification

A categorization process that helps generate predictions. ML models for classification problems predict a discrete value. Discrete values are always distinct from one another. For example, a model might need to evaluate whether or not there is a car in an image.

client-side encryption

Encryption of data locally, before the target AWS service receives it.

Cloud Center of Excellence (CCoE)

A multi-disciplinary team that drives cloud adoption efforts across an organization, including developing cloud best practices, mobilizing resources, establishing migration timelines, and leading the organization through large-scale transformations. For more information, see the CCoE posts on the AWS Cloud Enterprise Strategy Blog.

cloud computing

The cloud technology that is typically used for remote data storage and IoT device management. Cloud computing is commonly connected to edge computing technology.

cloud operating model

In an IT organization, the operating model that is used to build, mature, and optimize one or more cloud environments. For more information, see Building your Cloud Operating Model.

cloud stages of adoption

The four phases that organizations typically go through when they migrate to the AWS Cloud:

- Project – Running a few cloud-related projects for proof of concept and learning purposes
- Foundation – Making foundational investments to scale your cloud adoption (e.g., creating a landing zone, defining a CCoE, establishing an operations model)
- Migration – Migrating individual applications
- Re-invention – Optimizing products and services, and innovating in the cloud

These stages were defined by Stephen Orban in the blog post The Journey Toward Cloud-First & the Stages of Adoption on the AWS Cloud Enterprise Strategy blog. For information about how they relate to the AWS migration strategy, see the migration readiness guide.

CMDB

See configuration management database.

code repository

A location where source code and other assets, such as documentation, samples, and scripts, are stored and updated through version control processes. Common cloud repositories include GitHub or AWS CodeCommit. Each version of the code is called a *branch*. In a microservice structure, each repository is devoted to a single piece of functionality. A single CI/CD pipeline can use multiple repositories.

cold cache

A buffer cache that is empty, not well populated, or contains stale or irrelevant data. This affects performance because the database instance must read from the main memory or disk, which is slower than reading from the buffer cache.

cold data

Data that is rarely accessed and is typically historical. When querying this kind of data, slow queries are typically acceptable. Moving this data to lower-performing and less expensive storage tiers or classes can reduce costs.

computer vision

A field of AI used by machines to identify people, places, and things in images with accuracy at or above human levels. Often built with deep learning models, it automates extraction, analysis, classification, and understanding of useful information from a single image or a sequence of images.

configuration management database (CMDB)

A repository that stores and manages information about a database and its IT environment, including both hardware and software components and their configurations. You typically use data from a CMDB in the portfolio discovery and analysis stage of migration.

conformance pack

A collection of AWS Config rules and remediation actions that you can assemble to customize your compliance and security checks. You can deploy a conformance pack as a single entity in

an AWS account and Region, or across an organization, by using a YAML template. For more information, see Conformance packs in the AWS Config documentation.

continuous integration and continuous delivery (CI/CD)

The process of automating the source, build, test, staging, and production stages of the software release process. CI/CD is commonly described as a pipeline. CI/CD can help you automate processes, improve productivity, improve code quality, and deliver faster. For more information, see Benefits of continuous delivery. CD can also stand for *continuous deployment*. For more information, see Continuous Delivery vs. Continuous Deployment.

D

data at rest

Data that is stationary in your network, such as data that is in storage.

data classification

A process for identifying and categorizing the data in your network based on its criticality and sensitivity. It is a critical component of any cybersecurity risk management strategy because it helps you determine the appropriate protection and retention controls for the data. Data classification is a component of the security pillar in the AWS Well-Architected Framework. For more information, see Data classification.

data drift

A meaningful variation between the production data and the data that was used to train an ML model, or a meaningful change in the input data over time. Data drift can reduce the overall quality, accuracy, and fairness in ML model predictions.

data in transit

Data that is actively moving through your network, such as between network resources.

data minimization

The principle of collecting and processing only the data that is strictly necessary. Practicing data minimization in the AWS Cloud can reduce privacy risks, costs, and your analytics carbon footprint.

data perimeter

A set of preventive guardrails in your AWS environment that help make sure that only trusted identities are accessing trusted resources from expected networks. For more information, see Building a data perimeter on AWS.

data preprocessing

To transform raw data into a format that is easily parsed by your ML model. Preprocessing data can mean removing certain columns or rows and addressing missing, inconsistent, or duplicate values.

data provenance

The process of tracking the origin and history of data throughout its lifecycle, such as how the data was generated, transmitted, and stored.

data subject

An individual whose data is being collected and processed.

data warehouse

A data management system that supports business intelligence, such as analytics. Data warehouses commonly contain large amounts of historical data, and they are typically used for queries and analysis.

database definition language (DDL)

Statements or commands for creating or modifying the structure of tables and objects in a database.

database manipulation language (DML)

Statements or commands for modifying (inserting, updating, and deleting) information in a database.

DDL

See database definition language.

deep ensemble

To combine multiple deep learning models for prediction. You can use deep ensembles to obtain a more accurate prediction or for estimating uncertainty in predictions.

deep learning

An ML subfield that uses multiple layers of artificial neural networks to identify mapping between input data and target variables of interest.

defense-in-depth

An information security approach in which a series of security mechanisms and controls are thoughtfully layered throughout a computer network to protect the confidentiality, integrity, and availability of the network and the data within. When you adopt this strategy on AWS, you add multiple controls at different layers of the AWS Organizations structure to help secure resources. For example, a defense-in-depth approach might combine multi-factor authentication, network segmentation, and encryption.

delegated administrator

In AWS Organizations, a compatible service can register an AWS member account to administer the organization's accounts and manage permissions for that service. This account is called the *delegated administrator* for that service. For more information and a list of compatible services, see Services that work with AWS Organizations in the AWS Organizations documentation.

deployment

The process of making an application, new features, or code fixes available in the target environment. Deployment involves implementing changes in a code base and then building and running that code base in the application's environments.

development environment

See environment.

detective control

A security control that is designed to detect, log, and alert after an event has occurred. These controls are a second line of defense, alerting you to security events that bypassed the preventative controls in place. For more information, see Detective controls in *Implementing security controls on AWS*.

development value stream mapping (DVSM)

A process used to identify and prioritize constraints that adversely affect speed and quality in a software development lifecycle. DVSM extends the value stream mapping process originally

designed for lean manufacturing practices. It focuses on the steps and teams required to create and move value through the software development process.

digital twin

A virtual representation of a real-world system, such as a building, factory, industrial equipment, or production line. Digital twins support predictive maintenance, remote monitoring, and production optimization.

dimension table

In a star schema, a smaller table that contains data attributes about quantitative data in a fact table. Dimension table attributes are typically text fields or discrete numbers that behave like text. These attributes are commonly used for query constraining, filtering, and result set labeling.

disaster

An event that prevents a workload or system from fulfilling its business objectives in its primary deployed location. These events can be natural disasters, technical failures, or the result of human actions, such as unintentional misconfiguration or a malware attack.

disaster recovery (DR)

The strategy and process you use to minimize downtime and data loss caused by a disaster. For more information, see Disaster Recovery of Workloads on AWS: Recovery in the Cloud in the AWS Well-Architected Framework.

DML

See database manipulation language.

domain-driven design

An approach to developing a complex software system by connecting its components to evolving domains, or core business goals, that each component serves. This concept was introduced by Eric Evans in his book, *Domain-Driven Design: Tackling Complexity in the Heart of Software* (Boston: Addison-Wesley Professional, 2003). For information about how you can use domain-driven design with the strangler fig pattern, see Modernizing legacy Microsoft ASP.NET (ASMX) web services incrementally by using containers and Amazon API Gateway.

DR

See disaster recovery.

drift detection

Tracking deviations from a baselined configuration. For example, you can use AWS CloudFormation to detect drift in system resources, or you can use AWS Control Tower to detect changes in your landing zone that might affect compliance with governance requirements.

DVSM

See development value stream mapping.

E

EDA

See exploratory data analysis.

edge computing

The technology that increases the computing power for smart devices at the edges of an IoT network. When compared with cloud computing, edge computing can reduce communication latency and improve response time.

encryption

A computing process that transforms plaintext data, which is human-readable, into ciphertext.

encryption key

A cryptographic string of randomized bits that is generated by an encryption algorithm. Keys can vary in length, and each key is designed to be unpredictable and unique.

endianness

The order in which bytes are stored in computer memory. Big-endian systems store the most significant byte first. Little-endian systems store the least significant byte first.

endpoint

See service endpoint.

endpoint service

A service that you can host in a virtual private cloud (VPC) to share with other users. You can create an endpoint service with AWS PrivateLink and grant permissions to other AWS accounts

or to AWS Identity and Access Management (IAM) principals. These accounts or principals can connect to your endpoint service privately by creating interface VPC endpoints. For more information, see Create an endpoint service in the Amazon Virtual Private Cloud (Amazon VPC) documentation.

envelope encryption

The process of encrypting an encryption key with another encryption key. For more information, see Envelope encryption in the AWS Key Management Service (AWS KMS) documentation.

environment

An instance of a running application. The following are common types of environments in cloud computing:

- development environment – An instance of a running application that is available only to the core team responsible for maintaining the application. Development environments are used to test changes before promoting them to upper environments. This type of environment is sometimes referred to as a *test environment*.
- lower environments – All development environments for an application, such as those used for initial builds and tests.
- production environment – An instance of a running application that end users can access. In a CI/CD pipeline, the production environment is the last deployment environment.
- upper environments – All environments that can be accessed by users other than the core development team. This can include a production environment, preproduction environments, and environments for user acceptance testing.

epic

In agile methodologies, functional categories that help organize and prioritize your work. Epics provide a high-level description of requirements and implementation tasks. For example, AWS CAF security epics include identity and access management, detective controls, infrastructure security, data protection, and incident response. For more information about epics in the AWS migration strategy, see the program implementation guide.

exploratory data analysis (EDA)

The process of analyzing a dataset to understand its main characteristics. You collect or aggregate data and then perform initial investigations to find patterns, detect anomalies,

and check assumptions. EDA is performed by calculating summary statistics and creating data visualizations.

F

fact table

The central table in a star schema. It stores quantitative data about business operations. Typically, a fact table contains two types of columns: those that contain measures and those that contain a foreign key to a dimension table.

fail fast

A philosophy that uses frequent and incremental testing to reduce the development lifecycle. It is a critical part of an agile approach.

fault isolation boundary

In the AWS Cloud, a boundary such as an Availability Zone, AWS Region, control plane, or data plane that limits the effect of a failure and helps improve the resilience of workloads. For more information, see AWS Fault Isolation Boundaries.

feature branch

See branch.

features

The input data that you use to make a prediction. For example, in a manufacturing context, features could be images that are periodically captured from the manufacturing line.

feature importance

How significant a feature is for a model's predictions. This is usually expressed as a numerical score that can be calculated through various techniques, such as Shapley Additive Explanations (SHAP) and integrated gradients. For more information, see Machine learning model interpretability with :AWS.

feature transformation

To optimize data for the ML process, including enriching data with additional sources, scaling values, or extracting multiple sets of information from a single data field. This enables the ML

model to benefit from the data. For example, if you break down the “2021-05-27 00:15:37” date into “2021”, “May”, “Thu”, and “15”, you can help the learning algorithm learn nuanced patterns associated with different data components.

FGAC

See fine-grained access control.

fine-grained access control (FGAC)

The use of multiple conditions to allow or deny an access request.

flash-cut migration

A database migration method that uses continuous data replication through change data capture to migrate data in the shortest time possible, instead of using a phased approach. The objective is to keep downtime to a minimum.

G

geo blocking

See geographic restrictions.

geographic restrictions (geo blocking)

In Amazon CloudFront, an option to prevent users in specific countries from accessing content distributions. You can use an allow list or block list to specify approved and banned countries. For more information, see Restricting the geographic distribution of your content in the CloudFront documentation.

Gitflow workflow

An approach in which lower and upper environments use different branches in a source code repository. The Gitflow workflow is considered legacy, and the trunk-based workflow is the modern, preferred approach.

greenfield strategy

The absence of existing infrastructure in a new environment. When adopting a greenfield strategy for a system architecture, you can select all new technologies without the restriction of compatibility with existing infrastructure, also known as brownfield. If you are expanding the existing infrastructure, you might blend brownfield and greenfield strategies.

guardrail

A high-level rule that helps govern resources, policies, and compliance across organizational units (OUs). *Preventive guardrails* enforce policies to ensure alignment to compliance standards. They are implemented by using service control policies and IAM permissions boundaries. *Detective guardrails* detect policy violations and compliance issues, and generate alerts for remediation. They are implemented by using AWS Config, AWS Security Hub, Amazon GuardDuty, AWS Trusted Advisor, Amazon Inspector, and custom AWS Lambda checks.

H

HA

See high availability.

heterogeneous database migration

Migrating your source database to a target database that uses a different database engine (for example, Oracle to Amazon Aurora). Heterogeneous migration is typically part of a re-architecting effort, and converting the schema can be a complex task. AWS provides AWS SCT that helps with schema conversions.

high availability (HA)

The ability of a workload to operate continuously, without intervention, in the event of challenges or disasters. HA systems are designed to automatically fail over, consistently deliver high-quality performance, and handle different loads and failures with minimal performance impact.

historian modernization

An approach used to modernize and upgrade operational technology (OT) systems to better serve the needs of the manufacturing industry. A *historian* is a type of database that is used to collect and store data from various sources in a factory.

homogeneous database migration

Migrating your source database to a target database that shares the same database engine (for example, Microsoft SQL Server to Amazon RDS for SQL Server). Homogeneous migration is typically part of a rehosting or replatforming effort. You can use native database utilities to migrate the schema.

hot data

Data that is frequently accessed, such as real-time data or recent translational data. This data typically requires a high-performance storage tier or class to provide fast query responses.

hotfix

An urgent fix for a critical issue in a production environment. Due to its urgency, a hotfix is usually made outside of the typical DevOps release workflow.

hypercare period

Immediately following cutover, the period of time when a migration team manages and monitors the migrated applications in the cloud in order to address any issues. Typically, this period is 1–4 days in length. At the end of the hypercare period, the migration team typically transfers responsibility for the applications to the cloud operations team.

I

IaC

See infrastructure as code.

identity-based policy

A policy attached to one or more IAM principals that defines their permissions within the AWS Cloud environment.

idle application

An application that has an average CPU and memory usage between 5 and 20 percent over a period of 90 days. In a migration project, it is common to retire these applications or retain them on premises.

IIoT

See industrial Internet of Things.

immutable infrastructure

A model that deploys new infrastructure for production workloads instead of updating, patching, or modifying the existing infrastructure. Immutable infrastructures are inherently

more consistent, reliable, and predictable than mutable infrastructure. For more information, see the Deploy using immutable infrastructure best practice in the AWS Well-Architected Framework.

inbound (ingress) VPC

In an AWS multi-account architecture, a VPC that accepts, inspects, and routes network connections from outside an application. The AWS Security Reference Architecture recommends setting up your Network account with inbound, outbound, and inspection VPCs to protect the two-way interface between your application and the broader internet.

incremental migration

A cutover strategy in which you migrate your application in small parts instead of performing a single, full cutover. For example, you might move only a few microservices or users to the new system initially. After you verify that everything is working properly, you can incrementally move additional microservices or users until you can decommission your legacy system. This strategy reduces the risks associated with large migrations.

infrastructure

All of the resources and assets contained within an application's environment.

infrastructure as code (IaC)

The process of provisioning and managing an application's infrastructure through a set of configuration files. IaC is designed to help you centralize infrastructure management, standardize resources, and scale quickly so that new environments are repeatable, reliable, and consistent.

industrial Internet of Things (IIoT)

The use of internet-connected sensors and devices in the industrial sectors, such as manufacturing, energy, automotive, healthcare, life sciences, and agriculture. For more information, see Building an industrial Internet of Things (IIoT) digital transformation strategy.

inspection VPC

In an AWS multi-account architecture, a centralized VPC that manages inspections of network traffic between VPCs (in the same or different AWS Regions), the internet, and on-premises networks. The AWS Security Reference Architecture recommends setting up your Network account with inbound, outbound, and inspection VPCs to protect the two-way interface between your application and the broader internet.

Internet of Things (IoT)

The network of connected physical objects with embedded sensors or processors that communicate with other devices and systems through the internet or over a local communication network. For more information, see What is IoT?

interpretability

A characteristic of a machine learning model that describes the degree to which a human can understand how the model's predictions depend on its inputs. For more information, see Machine learning model interpretability with AWS.

IoT

See Internet of Things.

IT information library (ITIL)

A set of best practices for delivering IT services and aligning these services with business requirements. ITIL provides the foundation for ITSM.

IT service management (ITSM)

Activities associated with designing, implementing, managing, and supporting IT services for an organization. For information about integrating cloud operations with ITSM tools, see the operations integration guide.

ITIL

See IT information library.

ITSM

See IT service management.

L

label-based access control (LBAC)

An implementation of mandatory access control (MAC) where the users and the data itself are each explicitly assigned a security label value. The intersection between the user security label and data security label determines which rows and columns can be seen by the user.

landing zone

A landing zone is a well-architected, multi-account AWS environment that is scalable and secure. This is a starting point from which your organizations can quickly launch and deploy workloads and applications with confidence in their security and infrastructure environment. For more information about landing zones, see Setting up a secure and scalable multi-account AWS environment.

large migration

A migration of 300 or more servers.

LBAC

See label-based access control.

least privilege

The security best practice of granting the minimum permissions required to perform a task. For more information, see Apply least-privilege permissions in the IAM documentation.

lift and shift

See 7 Rs.

little-endian system

A system that stores the least significant byte first. See also endianness.

lower environments

See environment.

M

machine learning (ML)

A type of artificial intelligence that uses algorithms and techniques for pattern recognition and learning. ML analyzes and learns from recorded data, such as Internet of Things (IoT) data, to generate a statistical model based on patterns. For more information, see Machine Learning.

main branch

See branch.

managed services

AWS services for which AWS operates the infrastructure layer, the operating system, and platforms, and you access the endpoints to store and retrieve data. Amazon Simple Storage Service (Amazon S3) and Amazon DynamoDB are examples of managed services. These are also known as *abstracted services*.

MAP

See Migration Acceleration Program.

mechanism

A complete process in which you create a tool, drive adoption of the tool, and then inspect the results in order to make adjustments. A mechanism is a cycle that reinforces and improves itself as it operates. For more information, see Building mechanisms in the AWS Well-Architected Framework.

member account

All AWS accounts other than the management account that are part of an organization in AWS Organizations. An account can be a member of only one organization at a time.

microservice

A small, independent service that communicates over well-defined APIs and is typically owned by small, self-contained teams. For example, an insurance system might include microservices that map to business capabilities, such as sales or marketing, or subdomains, such as purchasing, claims, or analytics. The benefits of microservices include agility, flexible scaling, easy deployment, reusable code, and resilience. For more information, see Integrating microservices by using AWS serverless services.

microservices architecture

An approach to building an application with independent components that run each application process as a microservice. These microservices communicate through a well-defined interface by using lightweight APIs. Each microservice in this architecture can be updated, deployed, and scaled to meet demand for specific functions of an application. For more information, see Implementing microservices on AWS.

Migration Acceleration Program (MAP)

An AWS program that provides consulting support, training, and services to help organizations build a strong operational foundation for moving to the cloud, and to help offset the initial

cost of migrations. MAP includes a migration methodology for executing legacy migrations in a methodical way and a set of tools to automate and accelerate common migration scenarios.

migration at scale

The process of moving the majority of the application portfolio to the cloud in waves, with more applications moved at a faster rate in each wave. This phase uses the best practices and lessons learned from the earlier phases to implement a *migration factory* of teams, tools, and processes to streamline the migration of workloads through automation and agile delivery. This is the third phase of the AWS migration strategy.

migration factory

Cross-functional teams that streamline the migration of workloads through automated, agile approaches. Migration factory teams typically include operations, business analysts and owners, migration engineers, developers, and DevOps professionals working in sprints. Between 20 and 50 percent of an enterprise application portfolio consists of repeated patterns that can be optimized by a factory approach. For more information, see the discussion of migration factories and the Cloud Migration Factory guide in this content set.

migration metadata

The information about the application and server that is needed to complete the migration. Each migration pattern requires a different set of migration metadata. Examples of migration metadata include the target subnet, security group, and AWS account.

migration pattern

A repeatable migration task that details the migration strategy, the migration destination, and the migration application or service used. Example: Rehost migration to Amazon EC2 with AWS Application Migration Service.

Migration Portfolio Assessment (MPA)

An online tool that provides information for validating the business case for migrating to the AWS Cloud. MPA provides detailed portfolio assessment (server right-sizing, pricing, TCO comparisons, migration cost analysis) as well as migration planning (application data analysis and data collection, application grouping, migration prioritization, and wave planning). The MPA tool (requires login) is available free of charge to all AWS consultants and APN Partner consultants.

Migration Readiness Assessment (MRA)

The process of gaining insights about an organization's cloud readiness status, identifying strengths and weaknesses, and building an action plan to close identified gaps, using the AWS CAF. For more information, see the migration readiness guide. MRA is the first phase of the AWS migration strategy.

migration strategy

The approach used to migrate a workload to the AWS Cloud. For more information, see the 7 Rs entry in this glossary and see Mobilize your organization to accelerate large-scale migrations.

ML

See machine learning.

MPA

See Migration Portfolio Assessment.

modernization

Transforming an outdated (legacy or monolithic) application and its infrastructure into an agile, elastic, and highly available system in the cloud to reduce costs, gain efficiencies, and take advantage of innovations. For more information, see Strategy for modernizing applications in the AWS Cloud.

modernization readiness assessment

An evaluation that helps determine the modernization readiness of an organization's applications; identifies benefits, risks, and dependencies; and determines how well the organization can support the future state of those applications. The outcome of the assessment is a blueprint of the target architecture, a roadmap that details development phases and milestones for the modernization process, and an action plan for addressing identified gaps. For more information, see Evaluating modernization readiness for applications in the AWS Cloud.

monolithic applications (monoliths)

Applications that run as a single service with tightly coupled processes. Monolithic applications have several drawbacks. If one application feature experiences a spike in demand, the entire architecture must be scaled. Adding or improving a monolithic application's features also becomes more complex when the code base grows. To address these issues, you can use a microservices architecture. For more information, see Decomposing monoliths into microservices.

multiclass classification

A process that helps generate predictions for multiple classes (predicting one of more than two outcomes). For example, an ML model might ask "Is this product a book, car, or phone?" or "Which product category is most interesting to this customer?"

mutable infrastructure

A model that updates and modifies the existing infrastructure for production workloads. For improved consistency, reliability, and predictability, the AWS Well-Architected Framework recommends the use of immutable infrastructure as a best practice.

O

OAC

See origin access control.

OAI

See origin access identity.

OCM

See organizational change management.

offline migration

A migration method in which the source workload is taken down during the migration process. This method involves extended downtime and is typically used for small, non-critical workloads.

OI

See operations integration.

OLA

See operational-level agreement.

online migration

A migration method in which the source workload is copied to the target system without being taken offline. Applications that are connected to the workload can continue to function during the migration. This method involves zero to minimal downtime and is typically used for critical production workloads.

operational-level agreement (OLA)

An agreement that clarifies what functional IT groups promise to deliver to each other, to support a service-level agreement (SLA).

operational readiness review (ORR)

A checklist of questions and associated best practices that help you understand, evaluate, prevent, or reduce the scope of incidents and possible failures. For more information, see Operational Readiness Reviews (ORR) in the AWS Well-Architected Framework.

operations integration (OI)

The process of modernizing operations in the cloud, which involves readiness planning, automation, and integration. For more information, see the operations integration guide.

organization trail

A trail that's created by AWS CloudTrail that logs all events for all AWS accounts in an organization in AWS Organizations. This trail is created in each AWS account that's part of the organization and tracks the activity in each account. For more information, see Creating a trail for an organization in the CloudTrail documentation.

organizational change management (OCM)

A framework for managing major, disruptive business transformations from a people, culture, and leadership perspective. OCM helps organizations prepare for, and transition to, new systems and strategies by accelerating change adoption, addressing transitional issues, and driving cultural and organizational changes. In the AWS migration strategy, this framework is called *people acceleration*, because of the speed of change required in cloud adoption projects. For more information, see the OCM guide.

origin access control (OAC)

In CloudFront, an enhanced option for restricting access to secure your Amazon Simple Storage Service (Amazon S3) content. OAC supports all S3 buckets in all AWS Regions, server-side encryption with AWS KMS (SSE-KMS), and dynamic `PUT` and `DELETE` requests to the S3 bucket.

origin access identity (OAI)

In CloudFront, an option for restricting access to secure your Amazon S3 content. When you use OAI, CloudFront creates a principal that Amazon S3 can authenticate with. Authenticated principals can access content in an S3 bucket only through a specific CloudFront distribution. See also OAC, which provides more granular and enhanced access control.

ORR

See operational readiness review.

outbound (egress) VPC

In an AWS multi-account architecture, a VPC that handles network connections that are initiated from within an application. The AWS Security Reference Architecture recommends setting up your Network account with inbound, outbound, and inspection VPCs to protect the two-way interface between your application and the broader internet.

P

permissions boundary

An IAM management policy that is attached to IAM principals to set the maximum permissions that the user or role can have. For more information, see Permissions boundaries in the IAM documentation.

personally identifiable information (PII)

Information that, when viewed directly or paired with other related data, can be used to reasonably infer the identity of an individual. Examples of PII include names, addresses, and contact information.

PII

See personally identifiable information.

playbook

A set of predefined steps that capture the work associated with migrations, such as delivering core operations functions in the cloud. A playbook can take the form of scripts, automated runbooks, or a summary of processes or steps required to operate your modernized environment.

policy

An object that can define permissions (see identity-based policy), specify access conditions (see resource-based policy), or define the maximum permissions for all accounts in an organization in AWS Organizations (see service control policy).

polyglot persistence

Independently choosing a microservice's data storage technology based on data access patterns and other requirements. If your microservices have the same data storage technology, they can encounter implementation challenges or experience poor performance. Microservices are more easily implemented and achieve better performance and scalability if they use the data store best adapted to their requirements. For more information, see Enabling data persistence in microservices.

portfolio assessment

A process of discovering, analyzing, and prioritizing the application portfolio in order to plan the migration. For more information, see Evaluating migration readiness.

predicate

A query condition that returns `true` or `false`, commonly located in a `WHERE` clause.

predicate pushdown

A database query optimization technique that filters the data in the query before transfer. This reduces the amount of data that must be retrieved and processed from the relational database, and it improves query performance.

preventative control

A security control that is designed to prevent an event from occurring. These controls are a first line of defense to help prevent unauthorized access or unwanted changes to your network. For more information, see Preventative controls in *Implementing security controls on AWS*.

principal

An entity in AWS that can perform actions and access resources. This entity is typically a root user for an AWS account, an IAM role, or a user. For more information, see *Principal* in Roles terms and concepts in the IAM documentation.

Privacy by Design

An approach in system engineering that takes privacy into account throughout the whole engineering process.

private hosted zones

A container that holds information about how you want Amazon Route 53 to respond to DNS queries for a domain and its subdomains within one or more VPCs. For more information, see Working with private hosted zones in the Route 53 documentation.

proactive control

A security control designed to prevent the deployment of noncompliant resources. These controls scan resources before they are provisioned. If the resource is not compliant with the control, then it isn't provisioned. For more information, see the Controls reference guide in the AWS Control Tower documentation and see Proactive controls in *Implementing security controls on AWS*.

production environment

See environment.

pseudonymization

The process of replacing personal identifiers in a dataset with placeholder values. Pseudonymization can help protect personal privacy. Pseudonymized data is still considered to be personal data.

Q

query plan

A series of steps, like instructions, that are used to access the data in a SQL relational database system.

query plan regression

When a database service optimizer chooses a less optimal plan than it did before a given change to the database environment. This can be caused by changes to statistics, constraints, environment settings, query parameter bindings, and updates to the database engine.

R

RACI matrix

See responsible, accountable, consulted, informed (RACI).

ransomware

A malicious software that is designed to block access to a computer system or data until a payment is made.

RASCI matrix

See responsible, accountable, consulted, informed (RACI).

RCAC

See row and column access control.

read replica

A copy of a database that's used for read-only purposes. You can route queries to the read replica to reduce the load on your primary database.

re-architect

See 7 Rs.

recovery point objective (RPO)

The maximum acceptable amount of time since the last data recovery point. This determines what is considered an acceptable loss of data between the last recovery point and the interruption of service.

recovery time objective (RTO)

The maximum acceptable delay between the interruption of service and restoration of service.

refactor

See 7 Rs.

Region

A collection of AWS resources in a geographic area. Each AWS Region is isolated and independent of the others to provide fault tolerance, stability, and resilience. For more information, see Managing AWS Regions in *AWS General Reference*.

regression

An ML technique that predicts a numeric value. For example, to solve the problem of "What price will this house sell for?" an ML model could use a linear regression model to predict a house's sale price based on known facts about the house (for example, the square footage).

rehost

See 7 Rs.

release

In a deployment process, the act of promoting changes to a production environment.

relocate

See 7 Rs.

replatform

See 7 Rs.

repurchase

See 7 Rs.

resource-based policy

A policy attached to a resource, such as an Amazon S3 bucket, an endpoint, or an encryption key. This type of policy specifies which principals are allowed access, supported actions, and any other conditions that must be met.

responsible, accountable, consulted, informed (RACI) matrix

A matrix that defines the roles and responsibilities for all parties involved in migration activities and cloud operations. The matrix name is derived from the responsibility types defined in the matrix: responsible (R), accountable (A), consulted (C), and informed (I). The support (S) type is optional. If you include support, the matrix is called a *RASCI matrix*, and if you exclude it, it's called a *RACI matrix*.

responsive control

A security control that is designed to drive remediation of adverse events or deviations from your security baseline. For more information, see Responsive controls in *Implementing security controls on AWS*.

retain

See 7 Rs.

retire

See 7 Rs.

rotation

The process of periodically updating a secret to make it more difficult for an attacker to access the credentials.

row and column access control (RCAC)

The use of basic, flexible SQL expressions that have defined access rules. RCAC consists of row permissions and column masks.

RPO

See recovery point objective.

RTO

See recovery time objective.

runbook

A set of manual or automated procedures required to perform a specific task. These are typically built to streamline repetitive operations or procedures with high error rates.

S

SAML 2.0

An open standard that many identity providers (IdPs) use. This feature enables federated single sign-on (SSO), so users can log into the AWS Management Console or call the AWS API operations without you having to create user in IAM for everyone in your organization. For more information about SAML 2.0-based federation, see About SAML 2.0-based federation in the IAM documentation.

SCP

See service control policy.

secret

In AWS Secrets Manager, confidential or restricted information, such as a password or user credentials, that you store in encrypted form. It consists of the secret value and its metadata. The secret value can be binary, a single string, or multiple strings. For more information, see Secret in the Secrets Manager documentation.

security control

A technical or administrative guardrail that prevents, detects, or reduces the ability of a threat actor to exploit a security vulnerability. There are four primary types of security controls: preventative, detective, responsive, and proactive.

security hardening

The process of reducing the attack surface to make it more resistant to attacks. This can include actions such as removing resources that are no longer needed, implementing the security best practice of granting least privilege, or deactivating unnecessary features in configuration files.

security information and event management (SIEM) system

Tools and services that combine security information management (SIM) and security event management (SEM) systems. A SIEM system collects, monitors, and analyzes data from servers, networks, devices, and other sources to detect threats and security breaches, and to generate alerts.

security response automation

A predefined and programmed action that is designed to automatically respond to or remediate a security event. These automations serve as detective or responsive security controls that help you implement AWS security best practices. Examples of automated response actions include modifying a VPC security group, patching an Amazon EC2 instance, or rotating credentials.

server-side encryption

Encryption of data at its destination, by the AWS service that receives it.

service control policy (SCP)

A policy that provides centralized control over permissions for all accounts in an organization in AWS Organizations. SCPs define guardrails or set limits on actions that an administrator can delegate to users or roles. You can use SCPs as allow lists or deny lists, to specify which services or actions are permitted or prohibited. For more information, see Service control policies in the AWS Organizations documentation.

service endpoint

The URL of the entry point for an AWS service. You can use the endpoint to connect programmatically to the target service. For more information, see AWS service endpoints in *AWS General Reference*.

service-level agreement (SLA)

An agreement that clarifies what an IT team promises to deliver to their customers, such as service uptime and performance.

service-level indicator (SLI)

A measurement of a performance aspect of a service, such as its error rate, availability, or throughput.

service-level objective (SLO)

A target metric that represents the health of a service, as measured by a service-level indicator.

shared responsibility model

A model describing the responsibility you share with AWS for cloud security and compliance. AWS is responsible for security *of* the cloud, whereas you are responsible for security *in* the cloud. For more information, see Shared responsibility model.

SIEM

See security information and event management system.

single point of failure (SPOF)

A failure in a single, critical component of an application that can disrupt the system.

SLA

See service-level agreement.

SLI

See service-level indicator.

SLO

See service-level objective.

split-and-seed model

A pattern for scaling and accelerating modernization projects. As new features and product releases are defined, the core team splits up to create new product teams. This helps scale your organization's capabilities and services, improves developer productivity, and supports rapid

innovation. For more information, see Phased approach to modernizing applications in the AWS Cloud.

SPOF

See single point of failure.

star schema

A database organizational structure that uses one large fact table to store transactional or measured data and uses one or more smaller dimensional tables to store data attributes. This structure is designed for use in a data warehouse or for business intelligence purposes.

strangler fig pattern

An approach to modernizing monolithic systems by incrementally rewriting and replacing system functionality until the legacy system can be decommissioned. This pattern uses the analogy of a fig vine that grows into an established tree and eventually overcomes and replaces its host. The pattern was introduced by Martin Fowler as a way to manage risk when rewriting monolithic systems. For an example of how to apply this pattern, see Modernizing legacy Microsoft ASP.NET (ASMX) web services incrementally by using containers and Amazon API Gateway.

subnet

A range of IP addresses in your VPC. A subnet must reside in a single Availability Zone.

symmetric encryption

An encryption algorithm that uses the same key to encrypt and decrypt the data.

synthetic testing

Testing a system in a way that simulates user interactions to detect potential issues or to monitor performance. You can use Amazon CloudWatch Synthetics to create these tests.

T

tags

Key-value pairs that act as metadata for organizing your AWS resources. Tags can help you manage, identify, organize, search for, and filter resources. For more information, see Tagging your AWS resources.

target variable

The value that you are trying to predict in supervised ML. This is also referred to as an *outcome variable*. For example, in a manufacturing setting the target variable could be a product defect.

task list

A tool that is used to track progress through a runbook. A task list contains an overview of the runbook and a list of general tasks to be completed. For each general task, it includes the estimated amount of time required, the owner, and the progress.

test environment

See environment.

training

To provide data for your ML model to learn from. The training data must contain the correct answer. The learning algorithm finds patterns in the training data that map the input data attributes to the target (the answer that you want to predict). It outputs an ML model that captures these patterns. You can then use the ML model to make predictions on new data for which you don't know the target.

transit gateway

A network transit hub that you can use to interconnect your VPCs and on-premises networks. For more information, see What is a transit gateway in the AWS Transit Gateway documentation.

trunk-based workflow

An approach in which developers build and test features locally in a feature branch and then merge those changes into the main branch. The main branch is then built to the development, preproduction, and production environments, sequentially.

trusted access

Granting permissions to a service that you specify to perform tasks in your organization in AWS Organizations and in its accounts on your behalf. The trusted service creates a service-linked role in each account, when that role is needed, to perform management tasks for you. For more information, see Using AWS Organizations with other AWS services in the AWS Organizations documentation.

tuning

To change aspects of your training process to improve the ML model's accuracy. For example, you can train the ML model by generating a labeling set, adding labels, and then repeating these steps several times under different settings to optimize the model.

two-pizza team

A small DevOps team that you can feed with two pizzas. A two-pizza team size ensures the best possible opportunity for collaboration in software development.

U

uncertainty

A concept that refers to imprecise, incomplete, or unknown information that can undermine the reliability of predictive ML models. There are two types of uncertainty: *Epistemic uncertainty* is caused by limited, incomplete data, whereas *aleatoric uncertainty* is caused by the noise and randomness inherent in the data. For more information, see the Quantifying uncertainty in deep learning systems guide.

undifferentiated tasks

Also known as *heavy lifting*, work that is necessary to create and operate an application but that doesn't provide direct value to the end user or provide competitive advantage. Examples of undifferentiated tasks include procurement, maintenance, and capacity planning.

upper environments

See environment.

V

vacuuming

A database maintenance operation that involves cleaning up after incremental updates to reclaim storage and improve performance.

version control

Processes and tools that track changes, such as changes to source code in a repository.

VPC peering

A connection between two VPCs that allows you to route traffic by using private IP addresses. For more information, see What is VPC peering in the Amazon VPC documentation.

vulnerability

A software or hardware flaw that compromises the security of the system.

W

warm cache

A buffer cache that contains current, relevant data that is frequently accessed. The database instance can read from the buffer cache, which is faster than reading from the main memory or disk.

warm data

Data that is infrequently accessed. When querying this kind of data, moderately slow queries are typically acceptable.

window function

A SQL function that performs a calculation on a group of rows that relate in some way to the current record. Window functions are useful for processing tasks, such as calculating a moving average or accessing the value of rows based on the relative position of the current row.

workload

A collection of resources and code that delivers business value, such as a customer-facing application or backend process.

workstream

Functional groups in a migration project that are responsible for a specific set of tasks. Each workstream is independent but supports the other workstreams in the project. For example, the portfolio workstream is responsible for prioritizing applications, wave planning, and collecting migration metadata. The portfolio workstream delivers these assets to the migration workstream, which then migrates the servers and applications.

WORM

See write once, read many.

WQF

See AWS Workload Qualification Framework.

write once, read many (WORM)

A storage model that writes data a single time and prevents the data from being deleted or modified. Authorized users can read the data as many times as needed, but they cannot change it. This data storage infrastructure is considered immutable.

Z

zero-day exploit

An attack, typically malware, that takes advantage of a zero-day vulnerability.

zero-day vulnerability

An unmitigated flaw or vulnerability in a production system. Threat actors can use this type of vulnerability to attack the system. Developers frequently become aware of the vulnerability as a result of the attack.

zombie application

An application that has an average CPU and memory usage below 5 percent. In a migration project, it is common to retire these applications.