



Modèle de maturité pour l'adoption de l'IA générative sur AWS

AWS Directives prescriptives



AWS Directives prescriptives: Modèle de maturité pour l'adoption de l'IA générative sur AWS

Copyright © 2026 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Les marques et la présentation commerciale d'Amazon ne peuvent être utilisées en relation avec un produit ou un service qui n'est pas d'Amazon, d'une manière susceptible de créer une confusion parmi les clients, ou d'une manière qui dénigre ou discrédite Amazon. Toutes les autres marques commerciales qui ne sont pas la propriété d'Amazon appartiennent à leurs propriétaires respectifs, qui peuvent ou non être affiliés ou connectés à Amazon, ou sponsorisés par Amazon.

Table of Contents

Introduction	1
Public visé	2
Cibler les objectifs commerciaux	2
Vue d'ensemble du modèle	3
Niveaux de maturité	3
Aspects de maturité	8
Les piliers de l'adoption	9
Domaines d'intérêt	9
Principales activités	10
Stratégie de transformation	10
Niveau 1 : Envision	11
Objectif et critères	11
Principales activités	12
Stratégie de transformation	16
Niveau 2 : Expérience	18
Objectif et critères	18
Principales activités	19
Stratégie de transformation	22
Niveau 3 : Lancement	24
Objectif et critères	24
Principales activités	25
Stratégie de transformation	28
Niveau 4 : Échelle	30
Objectif et critères	30
Principales activités	31
Poursuivre le voyage	34
Étapes suivantes	36
Ressources	36
Service AWS documentation	36
AWS Conseils prescriptifs	37
Autres ressources	37
Collaborateurs	38
Conception	38
Révision	38

Rédaction technique	38
Historique du document	39
Glossaire	40
#	40
A	41
B	44
C	46
D	49
E	53
F	56
G	58
H	59
I	61
L	63
M	64
O	69
P	71
Q	74
R	75
S	78
T	82
U	83
V	84
W	85
Z	86

Modèle de maturité pour l'adoption de l'IA générative sur AWS

Amazon Web Services ([contributeurs](#))

Juin 2025 ([historique du document](#))

L'IA générative est un sous-ensemble de modèles d'IA qui ont été entraînés sur de grandes quantités de données et peuvent générer de nouveaux contenus, notamment du texte, des images, de la musique et des vidéos. Les modèles peuvent utiliser des modèles de base préentraînés, des modèles personnalisés et des ensembles de données augmentés ou propriétaires. L'impact de l'IA générative couvre tous les secteurs. Elle peut renforcer la créativité, améliorer la productivité et permettre la mise en place de nouveaux modèles commerciaux. Si votre entreprise souhaite que l'IA générative améliore ses opérations, stimule l'innovation et stimule la croissance de l'entreprise, une approche structurée et progressive est essentielle pour suivre le parcours d'adoption.

Selon un [article du CIO](#), 88 % des pilotes d'IA ne parviennent pas à la production. Cela entraîne ce que l'on appelle la fatigue du pilote. L'article indique que « les entreprises sont simplement fatiguées de consacrer plus de temps, d'argent et d'énergie à soutenir des projets pilotes qui ne passent pas rapidement ou pas du tout à la production ». Cette lassitude peut freiner l'innovation et décourager de nouvelles expérimentations avec l'IA générative. En outre, selon un [McKinsey rapport](#), les entreprises sont confrontées à d'importants défis en matière de qualité des données et d'intégration dans leurs implémentations d'IA.

Ce document de stratégie fournit un cadre structuré pour aider les organisations à mettre en œuvre des solutions d'IA générative. Ce cadre est conçu pour vous aider à faire face aux complexités de l'adoption des technologies et à ne pas négliger les étapes critiques ou les meilleures pratiques. Utilisez les recommandations de ce guide pour comprendre de manière exhaustive votre maturité en matière d'IA générative. En évaluant le niveau de maturité, vous pouvez identifier les domaines prioritaires pour chaque niveau et lancer un parcours d'adoption de l'IA end-to-end générative. Ce cadre explore quatre niveaux de maturité, de la prise de conscience initiale à la transformation à grande échelle. Il décrit les activités clés et les pratiques essentielles pour chaque niveau.

Public visé

Cet article s'adresse aux cadres, aux directeurs des technologies, aux chefs d'entreprise, aux data scientists, aux AI/ML spécialistes de l'IA générative, aux professionnels de l'informatique et aux décideurs qui souhaitent créer de la valeur en adoptant l'IA générative dans leurs organisations.

Cibler les objectifs commerciaux

Grâce à une progression systématique vers les niveaux de maturité de l'IA générative, les entreprises peuvent atteindre les principaux résultats commerciaux suivants :

- Innovation stratégique en matière de processus métier grâce à des cas d'utilisation validés de l'IA générative
- Excellence opérationnelle grâce à des solutions d'IA robustes et prêtes pour la production
- Efficacité à l'échelle de l'entreprise grâce à des composants d'IA standardisés et réutilisables
- Avantage concurrentiel grâce à une transformation stratégique et à des capacités d'IA évolutives

Présentation du modèle de maturité de l'IA générative

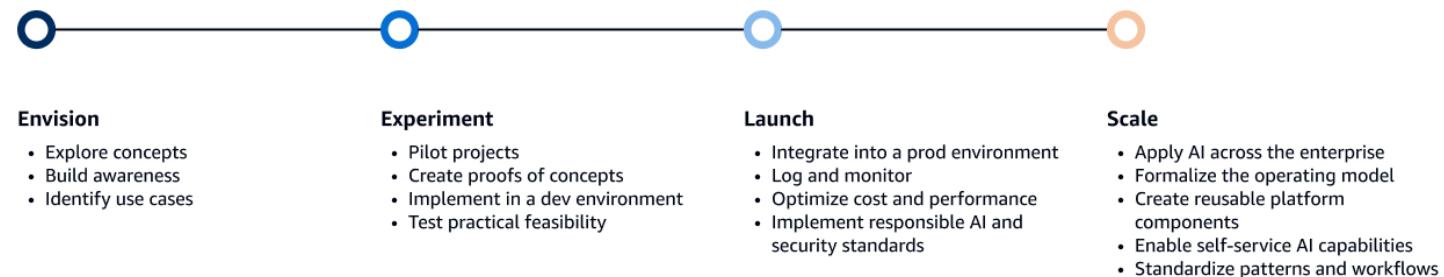
Organisations peuvent utiliser le cadre du modèle de maturité pour intégrer efficacement les fonctionnalités d'IA générative et éviter les pièges et les lacunes courants en matière de mise en œuvre. Grâce à une évaluation détaillée de la maturité, vous pouvez mieux comprendre où en est votre organisation dans son parcours vers l'IA et identifier les domaines spécifiques qui nécessitent une attention particulière. La progression s'étend sur quatre niveaux distincts, en commençant par une compréhension de base et en aboutissant à une transformation organisationnelle complète. Chaque niveau contient des actions ciblées et des directives stratégiques pour garantir le succès.

Cette section contient les rubriques suivantes :

- [Niveaux du modèle de maturité de l'IA générative](#)
- [Aspects de la maturité de l'IA générative](#)

Niveaux du modèle de maturité de l'IA générative

Le modèle de maturité de l'IA générative est structuré en quatre niveaux principaux. Chaque niveau représente les progrès d'une organisation vers l'utilisation des capacités d'IA générative. Ce modèle peut aider les organisations à comprendre leur situation actuelle et à les guider vers les prochaines étapes de leur parcours vers l'IA générative. Le schéma suivant montre les quatre niveaux du modèle de maturité de l'IA générative et les principales activités pour chaque niveau.



Les quatre niveaux du modèle de maturité de l'IA générative sont les suivants :

- [Niveau 1 : Envision](#)
- [Niveau 2 : Expérience](#)
- [Niveau 3 : Lancement](#)
- [Niveau 4 : Échelle](#)

Les labels pour chaque niveau de maturité reflètent l'impact de l'adoption de l'IA générative au sein de l'organisation. Au fur et à mesure que vous identifiez la position de votre organisation à un niveau donné, vous pouvez mieux comprendre les opportunités du prochain niveau de maturité. Les niveaux inférieurs incluent généralement des cas d'utilisation plus tactiques de l'IA générative, tandis que les niveaux supérieurs ont tendance à être de nature plus stratégique et transformatrice.

De nombreuses entreprises constateront que les caractéristiques de plusieurs niveaux de maturité s'appliquent à l'ensemble de leurs équipes et de leurs cas d'utilisation. En effet, aucun niveau n'est intrinsèquement supérieur ou inférieur. Le niveau de maturité approprié dépend des objectifs et de l'état de préparation de l'organisation.

Note

Ce modèle de maturité de l'IA générative ne vise pas à classer une organisation ou ses capacités d'IA générative dans la catégorie des entreprises uniquement débutantes ou transformatrices. Au contraire, chaque aspect de l'adoption de l'IA générative doit être considéré indépendamment. Les caractéristiques de chaque niveau de maturité représentent un continuum au sein de cet aspect spécifique, mais ne sont pas nécessairement corrélées au même niveau pour les autres aspects.

Le tableau suivant donne un aperçu des quatre niveaux.

Catégorie	Niveau 1 : Envision	Niveau 2 : Expérience	Niveau 3 : Lancement	Niveau 4 : Mise à l'échelle
Description	Organisations explorent les concepts de l'IA générative, sensibilisent et identifient les cas d'utilisation potentiels.	Organisations valident le potentiel de l'IA générative par le biais de projets pilotes structurés et de preuves de concept, tout en développant les capacités techniques de base et les	Organisations déploient systématiquement des solutions d'IA générative prêtes pour la production, dotées de mécanismes de gouvernance, de surveillance	Organisations mettent en place des capacités d'IA générative à l'échelle de l'entreprise grâce à des composants réutilisables, à des modèles standardisés et à des plateformes en libre-service

Catégorie	Niveau 1 : Envision	Niveau 2 : Expérience	Niveau 3 : Lancement	Niveau 4 : Mise à l'échelle
		cadres de base pour la mise en œuvre.	et de support robustes afin de fournir une valeur constante et une excellente opérationnelle tout en respectant les normes de sécurité et de conformité.	afin d'accélérer l'adoption tout en maintenant une gouvernance automatisée et en favorisant l'innovation.
Concentrez-vous	Faire connaître et comprendre les technologies d'IA générative, explorer les applications potentielles et identifier les domaines dans lesquels l'IA peut apporter une valeur ajoutée à l'entreprise	Valider les valeurs de l'entreprise par le biais de programmes pilotes structurés et renforcer les compétences de base	Déployez des solutions prêtes pour la production qui offrent une valeur commerciale mesurable grâce à des processus de lancement robustes, à des cadres de gouvernance complets et à un suivi des performances	Créez des composants et des modèles réutilisables qui accélèrent l'adoption de l'IA générative au sein de l'entreprise

Catégorie	Niveau 1 : Envision	Niveau 2 : Expérience	Niveau 3 : Lancement	Niveau 4 : Mise à l'échelle
Critères	<ul style="list-style-type: none"> • Acquérir une compréhension de base des concepts de l'IA générative • Pas de projets formels ni d'allocation de ressources • Prenez conscience des tendances du secteur et des opportunités de valeur 	<ul style="list-style-type: none"> • Exécuter des projets pilotes et des validations de concepts • Formez de petites équipes pour explorer les capacités de l'IA générative • Établir des cadres de base et de gouvernance 	<ul style="list-style-type: none"> • Lancer quelques applications d'IA générative en production • Mettre en œuvre des politiques en matière de risque, de gouvernance et d'IA responsables pour les applications d'IA générative • Mettre en place des équipes opérationnelles et de soutien 	<ul style="list-style-type: none"> • Adopter largement l'IA générative dans les différents départements de l'organisation • Lancer de nombreuses applications d'IA générative en production • Prioriser les investissements dans l'infrastructure et les outils d'IA générative • Formaliser le modèle opérationnel et la matrice responsable, responsable, consultée, informée (RACI)

Catégorie	Niveau 1 : Envision	Niveau 2 : Expérience	Niveau 3 : Lancement	Niveau 4 : Mise à l'échelle
Principales activités	<ul style="list-style-type: none"> Participez à des formations, à des ateliers et à des conférences de sensibilisation à l'IA Communiquez avec des experts et des consultants en matière d'IA Explorez les cas d'utilisation potentiels et les avantages commerciaux Évaluer l'état de préparation culturelle Évaluer la gouvernance de l'IA générative Acquérir des connaissances 	<ul style="list-style-type: none"> Définir et affiner les cas d'utilisation métier pour les projets pilotes Développez des preuves de concepts Évaluer et sélectionner les modèles et outils d'IA générative appropriés Mesurer la concrétisation des avantages commerciaux Renforcer les capacités internes et l'expertise technique 	<ul style="list-style-type: none"> Initialisation d'un modèle d'exploitation Création d'une gouvernance de l'architecture des solutions Créez une stratégie de mise en œuvre prête pour la production Mettre en place des mécanismes de surveillance et de suivi des performances Mettre en œuvre la gestion des risques et de la gouvernance Intégrer un framework de bibliothèque d'infrastructure informatique (ITIL) 	<ul style="list-style-type: none"> Formaliser le modèle de fonctionnement de l'IA générative et la matrice RACI Créez des fonctionnalités et des composants d'IA générative réutilisables Standardisez les modèles de cas d'utilisation de l'IA générative Mettre en place un cadre de développement collaboratif à l'échelle de l'organisation Faites évoluer les capacités de l'IA vers une plateforme de développement interne (IDP) ou un logiciel en tant

Catégorie	Niveau 1 : Envision	Niveau 2 : Expérience	Niveau 3 : Lancement	Niveau 4 : Mise à l'échelle
			<ul style="list-style-type: none"> • Configuration de l'exploitation et de la structure de support 	<ul style="list-style-type: none"> que service (SaaS) • Partager et démocratiser les connaissances

Pour mieux expliquer et comprendre le modèle de maturité, il est important de comprendre comment les entreprises progressent généralement dans leur parcours d'adoption de l'IA générative. Cette progression reflète non seulement la manière dont les organisations utilisent les capacités d'IA générative, mais également ce qui les motive à poursuivre leur adoption. Au début, de nombreux utilisateurs n'avaient peut-être pas du tout formalisé les processus d'IA. Ils considèrent plutôt leurs outils comme un ensemble amélioré de capacités provenant de diverses sources internes. Au fur et à mesure que les entreprises mûrissent, ces fonctionnalités sont gérées et standardisées de manière plus cohérente. Au fur et à mesure que les capacités deviennent plus affinées et plus faciles à découvrir et que les utilisateurs choisissent naturellement d'utiliser les fonctionnalités de l'IA, les organisations abandonnent généralement les motivations externes telles que les mandats ou les incitations. Idéalement, ils commencent même à investir leurs propres efforts dans l'innovation et le développement de l'IA au sens large.

Aspects de la maturité de l'IA générative

L'adoption réussie de l'IA générative nécessite une compréhension globale des multiples dimensions organisationnelles. Cette section examine quatre aspects clés que les organisations doivent prendre en compte et développer tout au long de leur parcours de maturité : les piliers fondamentaux qui soutiennent l'adoption de l'IA, les domaines prioritaires qui orientent les priorités stratégiques, les activités clés qui orientent la mise en œuvre et la stratégie de transformation qui guide l'évolution de la maturité de l'organisation. Ensemble, ces aspects fournissent un cadre complet pour évaluer et faire progresser les capacités d'IA générative. Organisations peuvent utiliser ce cadre pour identifier les lacunes, hiérarchiser les investissements et créer des plans pratiques pour progresser jusqu'aux niveaux de maturité. Chaque aspect a été choisi sur la base d'une vaste expérience de terrain en matière d'adoption de l'IA en entreprise. Ils reflètent les éléments essentiels qui distinguent les implémentations réussies de celles qui échouent.

Cette section contient les rubriques suivantes :

- [Les piliers de l'adoption](#)
- [Domaines d'intérêt](#)
- [Principales activités](#)
- [Stratégie de transformation](#)

Les piliers de l'adoption

Chaque niveau de maturité est évalué en fonction des piliers d'adoption suivants :

- Entreprise — Alignement stratégique et impact mesurable sur les objectifs commerciaux
- Personnel — Développement des talents, renforcement des compétences et collaboration interfonctionnelle
- Gouvernance — Établissement de directives en matière de gestion des risques, de conformité et d'éthique
- Plateforme — Investissement dans une infrastructure et des plateformes évolutives pour les capacités d'IA générative
- Sécurité — Protection des données, de la confidentialité et déploiement de modèles d'IA génératifs
- Opérations — Gestion du cycle de vie des solutions d'IA générative, optimisation des déploiements, mise en œuvre de mécanismes de feedback et surveillance des performances

Ces piliers s'alignent sur le [cadre d'adoption du AWS cloud \(AWS CAF\)](#) et l'étendent pour répondre aux besoins en IA générative. Les recommandations de ce document de stratégie ajoutent des éléments spécifiques à l'IA à chaque pilier, tels que la mise en œuvre éthique de l'IA, la gestion du cycle de vie des modèles et les exigences en matière d'infrastructure d'IA. Cet alignement aide les organisations à utiliser les meilleures pratiques existantes de la AWS CAF tout en relevant les défis uniques liés à l'adoption de l'IA.

Domaines d'intérêt

Les domaines prioritaires pour chaque niveau de maturité aident les organisations à hiérarchiser les activités et les investissements. Les quatre domaines prioritaires sont les suivants :

- Innovation et faisabilité — Exploration et validation de cas d'utilisation innovants de l'IA générative ainsi que de la disponibilité et de la qualité des ensembles de données requis

- Intégration et efficacité — Intégrer l'IA générative dans les processus métier existants
- Évolutivité et optimisation : mise à l'échelle des applications d'IA génératives et amélioration continue des performances
- Transformation et leadership — Utiliser l'IA générative pour susciter des changements stratégiques et acquérir un avantage concurrentiel

Principales activités

Organisations peuvent utiliser les activités clés du modèle de maturité de l'IA générative pour suivre leur parcours et définir et mettre en œuvre avec succès leur stratégie d'IA générative. Les activités vont de l'exploration initiale et de la compréhension des technologies d'IA générative à l'expérimentation de prototypes, à l'intégration de solutions d'IA dans les processus métier, à leur extension à l'échelle de l'organisation, puis à l'établissement d'une gouvernance pour une amélioration continue et une transformation stratégique. Les principales activités entrent dans l'une des catégories suivantes :

- Exploration et sensibilisation — Développez des connaissances de base sur les technologies d'IA générative et identifiez les opportunités stratégiques d'adoption
- Expérimentation et validation — Faciliter et mener des projets pilotes et des prototypes pour évaluer la faisabilité technique et la valeur commerciale
- Intégration et mise en œuvre — Intégrez des fonctionnalités d'IA générative dans les processus métier existants et déployez des solutions dans les environnements de production
- Mise à l'échelle et optimisation — Intégrez les applications d'IA génératives au sein de l'organisation et améliorez continuellement leurs performances et leur efficacité
- Gouvernance et leadership — Établir des cadres et des meilleures pratiques pour gérer les initiatives d'IA générative et les utiliser pour une transformation stratégique

Stratégie de transformation

La stratégie de transformation à chaque niveau vise à guider les organisations par le biais d'améliorations progressives. Cela inclut l'élaboration d'une feuille de route en matière d'IA générative et d'une stratégie de données, l'alignement sur les objectifs commerciaux, l'investissement dans les talents et les outils, et la mise en œuvre de cadres de gouvernance.

Modèle de maturité de l'IA générative de niveau 1 : Envision

Ce niveau fondamental constitue un point de départ essentiel où les organisations explorent les concepts d'IA générative, renforcent leur prise de conscience organisationnelle et identifient les cas d'utilisation potentiels qui correspondent à leurs objectifs commerciaux. En établissant ces bases essentielles, les entreprises peuvent développer une vision claire de leur parcours vers l'IA tout en tenant compte de considérations clés liées aux activités, aux personnes, à la gouvernance, à la plateforme, à la sécurité et aux dimensions opérationnelles.

Cette section comprend les rubriques suivantes:

- [Objectif et critères](#)
- [Principales activités](#)
- [Stratégie de transformation pour passer au niveau supérieur](#)

Objectif et critères

L'objectif à ce niveau est de développer une compréhension et une prise de conscience fondamentales des technologies d'IA générative et des nouvelles tendances du secteur liées à ces technologies. Cela inclut l'évaluation des applications potentielles et l'identification des domaines dans lesquels l'IA générative pourrait bénéficier à l'entreprise. Ce niveau vise à sensibiliser les parties prenantes à l'IA générative et à commencer à explorer des cas d'utilisation et à effectuer une évaluation des risques et de l'état de préparation culturelle.

Les critères pour être à ce niveau sont les suivants :

- L'organisation a démontré une connaissance de base des principes fondamentaux de l'IA générative.
- L'organisation a documenté sa connaissance des applications et des opportunités de l'IA générative dans le secteur.
- L'organisation commence à comprendre son niveau de préparation culturelle à l'égard de l'IA.
- L'organisation a effectué une première exploration des cas d'utilisation et des avantages potentiels.
- L'organisation a examiné à titre préliminaire les exigences en matière de gouvernance et de sécurité.

Principales activités

Le tableau suivant présente les principales activités pour chaque pilier de l'adoption.

Pilier de l'adoption	Activités
Entreprise	<ul style="list-style-type: none">• Découvrez comment l'IA générative peut résoudre des problèmes commerciaux spécifiques.• Associez les premiers cas d'utilisation de l'IA générative aux objectifs commerciaux, tels que l'amélioration de l'engagement des clients ou l'automatisation de la création de contenu.• Identifiez les sources de données à forte valeur par rapport à des cas d'utilisation sélectionnés.
Personnes	<ul style="list-style-type: none">• Organiser des sessions de formation internes et des ateliers de partage des connaissances.• Identifiez les champions de l'IA au sein de l'organisation pour mener l'exploration des opportunités de l'IA générative.• Évaluez la culture de votre entreprise et son état de préparation à la gestion du changement en vue de l'adoption de l'IA générative.

Pilier de l'adoption	Activités	
	<ul style="list-style-type: none"> • Évaluez les lacunes technologiques actuelles de votre organisation et déterminez les investissements nécessaires à l'adoption de l'IA générative. • Concevez des initiatives pédagogiques pour aider les cadres supérieurs à comprendre le potentiel stratégique de l'IA, ses capacités technologiques, son impact commercial transformateur et l'importance des données dans les projets d'IA générative. • Participez à des forums et à des conférences du secteur pour tirer parti des expériences d'autres entreprises en matière d'adoption de l'IA. • Organisez des hackathons internes pour encourager l'expérimentation et favoriser l'innovation. 	

Pilier de l'adoption	Activités	
Gouvernance	<ul style="list-style-type: none"> Explorez les considérations éthiques et réglementaires liées à l'adoption de l'IA générative, telles que la confidentialité et la souveraineté des données. Élaborer un ensemble initial de directives pour une utilisation responsable de l'IA dans l'organisation. 	
Plateforme	<ul style="list-style-type: none"> Découvrez les exigences relatives à l'adoption de l'IA générative afin de s'aligner sur les normes de votre organisation. Explorez les AI/ML modèles et les outils, tels qu'Amazon Bedrock pour accéder aux modèles de base et Amazon SageMaker AI, pour une expérimentation rapide. Évaluez et cataloguez les sources de données internes et externes existantes. Évaluez l'infrastructure et la qualité des données afin de déterminer la faisabilité de l'IA générative et les exigences de mise en œuvre potentielles. 	

Pilier de l'adoption	Activités	
Sécurité	<ul style="list-style-type: none"> • Comprenez les implications en matière de sécurité et les tâches associées à l'adoption de l'IA générative dans l'organisation, telles que : <ul style="list-style-type: none"> • Risques liés à la confidentialité et à la protection des données, notamment l'exposition potentiel le à des informations sensibles par le biais de données de formation, d'instructions et de résultats de modèles • Les défis liés au contrôle d'accès et à l'authentification, qui incluent la complexité de la vérification des utilisateurs et des autorisations basées sur les rôles dans les systèmes d'IA • Modélisez les vulnérabilités de sécurité, notamment la vulnérabilité aux attaques par injection rapide et le potentiel de génération de contenu dangereux ou inapproprié 	

Pilier de l'adoption	Activités
Opérations	<ul style="list-style-type: none"> • Comprenez les défis opérationnels associés à l'adoption de l'IA générative dans l'organisation, tels que : <ul style="list-style-type: none"> • Planifiez les besoins de surveillance des performances de vos solutions d'IA. • Tenez compte des exigences en matière de gouvernance et de gestion des versions. • Comprenez ce qui est requis pour les procédures de réponse aux incidents.

Stratégie de transformation pour passer au niveau supérieur

Pour passer au niveau de maturité suivant, prenez en compte les aspects suivants :

- Mettez en place des équipes interfonctionnelles d'IA générative — Formez des équipes interfonctionnelles d'IA générative dotées de rôles et de responsabilités clairs. Les équipes devraient inclure des représentants informatiques, des représentants des entreprises, des acteurs de la sécurité et de la gouvernance, ainsi que des IA génératives SMEs capables de mener les efforts d'expérimentation. Ce groupe constituera la base d'un centre d'excellence (CoE) défini de manière plus formelle ultérieurement, à mesure que vous intensifiez vos efforts en matière d'IA générative.
- Identifiez et hiérarchisez les cas d'utilisation — Développez une matrice de cas d'utilisation qui vous aide à hiérarchiser les projets en fonction de leur faisabilité, de leur impact commercial et de leur alignement sur les objectifs stratégiques. Pour les preuves de concept (PoCs), créez une courte liste des principaux cas d'utilisation.

- Allouer des ressources pour des projets pilotes — Garantir le budget et le personnel nécessaires à la gestion à petite échelle PoCs.
- Développez des compétences en IA générative : améliorez les compétences du personnel en matière d'outils et de technologies spécifiques, tels qu'[Amazon Bedrock](#), [SageMaker AI](#), [Amazon Q Business](#), [Amazon Q Developer](#), [prompt engineering](#), [Retrieval Augmented Generation \(RAG\)](#), ainsi que l'IA agentique et les flux de travail.
- Gouvernance préliminaire complète — Établissez une gouvernance préliminaire qui oriente l'utilisation de l'IA générative. Il doit couvrir la conformité, la gestion des risques et les considérations éthiques.
- Préparation culturelle — Commencez à planifier la gestion du changement organisationnel en vue de l'adoption de l'IA générative à l'échelle de l'entreprise.
- Identifier les indicateurs de réussite — Pour chaque PoC, définissez les critères de réussite ainsi que les indicateurs commerciaux et techniques.

En prenant ces mesures, les organisations peuvent s'attendre à :

- Acquérez une expérience pratique des technologies d'IA générative.
- Validez la faisabilité et l'impact potentiel de cas d'utilisation spécifiques.
- Développez les capacités et l'expertise internes en matière d'IA générative.
- Identifiez les défis et les risques potentiels associés à l'adoption de l'IA générative.
- Améliorez le niveau de préparation à l'adoption de l'IA générative afin de passer au niveau supérieur.

Modèle de maturité de l'IA générative de niveau 2 : expérience

S'appuyant sur les connaissances fondamentales établies au niveau précédent, le niveau Expérience marque une transition cruciale entre l'exploration théorique et la mise en œuvre pratique des technologies d'IA générative. À ce niveau, les organisations vont au-delà de la compréhension conceptuelle pour s'engager dans des projets PoC pratiques et des programmes pilotes. Ce PoC et ces projets pilotes sont conçus pour valider la valeur commerciale et renforcer les compétences de base. Ce niveau est caractérisé par une expérimentation structurée, au cours de laquelle les organisations forment des équipes dédiées, établissent des cadres de gouvernance et commencent à développer une expertise technique interne. Grâce à des projets pilotes soigneusement contrôlés, les organisations peuvent tester leurs hypothèses sur le potentiel de l'IA générative tout en minimisant les risques et en maximisant les opportunités d'apprentissage. Cela ouvre la voie à une mise en œuvre plus large et à une plus grande échelle des initiatives réussies.

Cette section comprend les rubriques suivantes:

- [Objectif et critères](#)
- [Principales activités](#)
- [Stratégie de transformation pour passer au niveau supérieur](#)

Objectif et critères

À ce niveau, les organisations passent de l'exploration à l'expérimentation pratique du PoC et à des projets pilotes utilisant des technologies d'IA générative. L'accent est mis sur la validation de la valeur commerciale par le biais de programmes pilotes structurés et du renforcement des compétences de base. Ce niveau met l'accent sur l'apprentissage pratique, le renforcement des capacités internes et de l'expertise technique, et l'établissement de cadres de base et de gouvernance.

Les critères pour être à ce niveau sont les suivants :

- L'organisation a des projets pilotes actifs et des preuves de concept en cours.
- Des équipes interfonctionnelles dédiées sont affectées aux initiatives d'IA générative.
- Un programme de formation interne structuré est établi.
- Les organisations ont sélectionné et validé des modèles et des outils d'IA.

- L'organisation a défini sa gouvernance initiale et ses cadres de données.

Principales activités

Le tableau suivant présente les principales activités pour chaque pilier de l'adoption.

Pilier de l'adoption	Activités
Entreprise	<ul style="list-style-type: none"> • Définissez et hiérarchisez les cas d'utilisation stratégiques en fonction de la valeur commerciale et de la faisabilité. • Pour PoCs cela, établissez des indicateurs de réussite et des cadres pour mesurer le retour sur investissement (ROI). • Créez des fiches d'évaluation de la valeur pour chaque PoC. • Limitez le champ d'application PoCs à une échelle gérable avec des indicateurs de réussite clairs. • Pour chaque PoC, mesurez le retour sur investissement et évaluez s'il répond aux critères de réussite.
Personnes	<ul style="list-style-type: none"> • Mettez en œuvre des programmes de formation structurés portant sur l'ingénierie rapide, le RAG et le réglage précis des modèles. • Créez des parcours de certification en IA générative et des cadres de progression de carrière. • Engagez des experts en IA générative et en science des données. • Associez-vous à des spécialistes externes, tels que le AWS Generative AI Innovation Center ou AWS Professional Services, pour

Pilier de l'adoption	Activités
	<p>co-construire un PoC, fournir un soutien et transférer des connaissances.</p> <ul style="list-style-type: none"> • Établissez des parcours de certification en IA et des cadres de progression de carrière.
Gouvernance	<ul style="list-style-type: none"> • Développez des cadres préliminaires qui englobent la gouvernance des données pour l'IA générative, tels que la qualité du contenu utilisé pour la recherche vectorielle. • Établissez des critères d'évaluation du modèle et des contrôles de qualité. • Mettez en place des protocoles d'évaluation des risques pour les projets d'IA générative. • Mettre en place des directives pour une utilisation éthique et responsable de l'IA générative. Formez les développeurs, les data scientists et les spécialistes de l'IA générative pour qu'ils se conforment à ces directives.

Pilier de l'adoption	Activités
Plateforme	<ul style="list-style-type: none"> Configurez l'infrastructure de base pour le PoC, telle qu'une zone d'AWS atterrissage et les autorisations dont les développeurs ont besoin. Configurez un environnement pour l'expérimentation de l'IA générative et le développement de PoC, tel qu'un terrain de jeu Amazon Bedrock, un JupyterLab espace Amazon SageMaker AI ou une instance de bloc-notes. Mettez en œuvre une approche RAG ou un flux de travail agentic que les développeurs peuvent facilement utiliser. Pour une approche RAG, considérez les bases de connaissances Amazon Bedrock, et pour un flux de travail agentique, pensez à Amazon Bedrock Agents. Configurez des frameworks ou des pipelines qui gèrent les invités, les modèles et les évaluations rapides. Ces ressources devraient aider les développeurs à évaluer rapidement les résultats et les performances de l'application PoC. Mettez en œuvre des efforts d'intégration des données à un stade précoce, y compris des pipelines de données structurés et non structurés. Configurez des bases de données vectorielles pour les expériences RAG. Évaluez les modèles de base en fonction du coût, des performances et de la pertinence des cas d'utilisation. Vous pouvez utiliser Amazon Bedrock, Amazon SageMaker AI et Amazon SageMaker AI JumpStart.

Pilier de l'adoption	Activités
Sécurité	<ul style="list-style-type: none"> Mettez en œuvre des contrôles d'accès aux données pour la formation de modèles d'IA générative et assurez-vous qu'ils respectent les exigences de conformité. Amazon Q Business peut simplifier la mise en œuvre de RAG en activant des contrôles précis qui permettent aux charges de travail génératives basées sur l'IA de récupérer uniquement les données auxquelles l'utilisateur est autorisé à accéder. Développez une stratégie pour protéger les informations personnelles identifiables (PII) dans les ensembles de données utilisés pour entraîner des modèles.
Opérations	<ul style="list-style-type: none"> Créez des processus de documentation et de support pour les éléments suivants : <ul style="list-style-type: none"> Implémentations et apprentissages du PoC Configurations de base de la plateforme et contrôles de sécurité Procédures de test et d'évaluation Processus de transfert réussis PoCs qui passent à la production

Stratégie de transformation pour passer au niveau supérieur

Organisations peuvent passer au niveau de maturité suivant en procédant comme suit :

- Créez une infrastructure adaptée à la production pour prendre en charge l'IA générative : Services AWS à utiliser pour implémenter des CI/CD pipelines, des modèles de déploiement standardisés et des mécanismes de mise à l'échelle appropriés pour les déploiements de production.
- Mettre en œuvre la gouvernance — Établissez des cadres de gouvernance adaptés à la production pour gérer l'utilisation continue de l'IA générative et les mises à jour des modèles.

- Mettre en œuvre l'observabilité — Mettez en œuvre des pratiques d'observabilité, de surveillance et de journalisation spécifiquement adaptées aux charges de travail génératives de l'IA. Cela inclut les mesures de performance du modèle, les modèles d'utilisation et l'évaluation de la qualité des réponses.
- Concentrez-vous sur la conformité : assurez-vous de respecter les normes et réglementations du secteur en matière de confidentialité et de sécurité des données.
- Constituez des équipes dédiées à l'IA — Configurez une équipe chargée de créer et de maintenir des voies de production standardisées pour les solutions d'IA générative.
- Mettez en œuvre l'excellence opérationnelle : créez un processus de réponse aux incidents et d'escalade. Établissez des accords de niveau de service (SLAs) et des indicateurs de performance. Mettez en œuvre des stratégies d'optimisation des coûts.

En prenant ces mesures, les organisations peuvent :

- Vérifiez que les applications d'IA générative sont stables, fiables et qu'elles apportent continuellement de la valeur à l'organisation.
- Support à la croissance des solutions d'IA générative à mesure que la demande et l'utilisation augmentent dans les différents services.
- Gérez les risques, maintenez une surveillance et alignez les initiatives d'IA sur les normes réglementaires à mesure qu'elles font partie intégrante des opérations commerciales.
- Assurez une surveillance, une amélioration et un support continu pour les solutions d'IA générative. Cela réduit le recours à des équipes de projet ad hoc ou temporaires.
- Préparer l'organisation à passer de projets isolés à une approche stratégique et cohérente, dans laquelle l'IA devient un élément essentiel des processus métier. L'organisation est prête à prendre de l'ampleur et à élargir son adoption.

Modèle de maturité de l'IA générative de niveau 3 : lancement

À ce niveau, les organisations passent des proof-of-concept initiatives au déploiement méthodique de solutions d'IA générative sélectionnées et éprouvées dans des environnements de production. Ce niveau représente un changement essentiel par rapport à l'expérimentation pour se concentrer sur des protocoles de gouvernance robustes, des systèmes de surveillance en temps réel et des infrastructures de support dédiées. Les entreprises se concentrent sur le lancement de quelques applications destinées à la production qui ont un impact commercial clair. Ce niveau met l'accent sur la rigueur opérationnelle : mise en œuvre de cadres de lancement complets, établissement de directives de gouvernance claires et maintien de normes de sécurité strictes. La publication de solutions d'IA générative fiables qui fournissent des résultats quantifiables prépare l'organisation à une adoption plus large.

Cette section comprend les rubriques suivantes:

- [Objectif et critères](#)
- [Principales activités](#)
- [Stratégie de transformation pour passer au niveau supérieur](#)

Objectif et critères

À ce niveau, les organisations déploient systématiquement des solutions d'IA générative dans les environnements de production et mettent en œuvre des mécanismes de gouvernance, de surveillance et de support robustes. Ces mécanismes offrent une valeur constante et une excellence opérationnelle tout en maintenant les normes de sécurité et de conformité. L'accent passe des applications expérimentales d'IA générative au déploiement de solutions prêtes pour la production qui offrent une valeur commerciale mesurable grâce à des processus de lancement robustes, à des cadres de gouvernance complets et à un suivi systématique des performances. Ce niveau se concentre sur le déploiement d'un certain nombre de solutions d'IA génératives prêtes pour la production qui servent de bases pour les cadres de lancement et les mécanismes de gouvernance.

Les critères pour être à ce niveau sont les suivants :

- Les solutions d'IA générative prêtes pour la production produisent des résultats commerciaux mesurables.

- L'organisation a mis en place des cadres de base en matière de sécurité, de gouvernance et d'IA responsable.
- Des contrôles opérationnels sont établis et incluent des systèmes automatisés de surveillance et d'alerte.
- L'organisation a défini un human-in-the-loop processus pour les décisions relatives à l'IA.
- Pour les équipes interfonctionnelles chargées de l'IA, les rôles préliminaires et les responsabilités opérationnelles ont été définis.

Principales activités

Le tableau suivant présente les principales activités pour chaque pilier de l'adoption.

Pilier de l'adoption	Activités
Entreprise	<ul style="list-style-type: none">• Approuver une première version d'une matrice RACI pour les opérations d'IA générative.• Identifiez les rôles clés nécessaires à l'architecture, au développement et au support de la plateforme.• Mesurez l'efficacité opérationnelle et la valeur commerciale grâce à des tableaux de bord complets.• Suivez et optimisez les coûts opérationnels et l'utilisation des ressources.
Personnes	<ul style="list-style-type: none">• Créez des équipes ou des escouades de plateformes d'IA générative pour l'architecture, le développement et la maintenance.• Mettez en place une structure de soutien à plusieurs niveaux et des programmes de formation toujours disponibles.

Pilier de l'adoption	Activités
Gouvernance	<ul style="list-style-type: none"> • Obtenir des approbations formelles en matière d'architecture auprès d'un comité d'examen de l'architecture d'entreprise. • Établissez un cadre politique responsable en matière d'IA et obtenez l'approbation des parties prenantes. • Créez un comité de supervision interfonctionnel pour les examens de la mise en œuvre de l'IA. • Pour les solutions d'IA générative, maintenez la documentation pour les approbations de gouvernance, les évaluations des risques, les modèles de conception standardisés et les spécifications techniques.
Plateforme	<ul style="list-style-type: none"> • Mettez en œuvre des CI/CD pipelines automatisés pour les solutions d'IA génératives. • Déployez l'infrastructure sous forme de code (IaC) pour gérer les AWS ressources. • Documentez les modèles de conception et les spécifications techniques des solutions d'IA générative. • Conservez les enregistrements CMDB pour les composants de la plateforme d'IA générative.

Pilier de l'adoption	Activités
Sécurité	<ul style="list-style-type: none">• Mettez en œuvre des contrôles de sécurité robustes pour les solutions d'IA générative et leurs pipelines de données.• Mettre en œuvre une politique préliminaire pour une IA responsable.• Optimisez l'infrastructure évolutive pour prendre en charge l'ingestion de données en temps réel, la recherche vectorielle et le réglage précis.• Procéder à des évaluations et à des audits de sécurité réguliers.• Déployez Amazon Bedrock Guardrails pour standardiser les contrôles de sécurité et de confidentialité dans les applications d'IA générative.

Pilier de l'adoption	Activités
Opérations	<ul style="list-style-type: none"> • Établissez des cadres de SLA et des indicateurs de performance. • Surveillez les performances des modèles et les violations des garde-corps. Configurez des alertes. • Créez des tableaux de bord opérationnels dotés de systèmes d'alerte automatisés. • Suivez les processus ITIL pour la gestion du changement et la gestion des actifs. • Création d'un référentiel de connaissances centralisé contenant des manuels opérationnels, des playbooks et des FAQs guides de dépannage. • Établissez des pratiques d'observabilité des données. Suivez le lignage des données, leur provenance et les indicateurs de qualité afin d'identifier les lacunes avant de procéder à une mise à l'échelle. • Établissez des niveaux de support hiérarchisés dotés de voies d'escalade claires. • Mettez en œuvre des évaluations régulières des performances et analysez les commentaires des clients.

Stratégie de transformation pour passer au niveau supérieur

Pour étendre les initiatives d'IA générative, les entreprises doivent :

- Formaliser le modèle opérationnel de l'IA générative : formaliser la matrice RACI au sein de l'organisation.
- Améliorez la plateforme d'IA générative : évaluez les implémentations d'IA générative existantes afin d'identifier les modèles et les composants réutilisables. Évaluez si la pile technologique est

prête à évoluer. Commencez à imaginer et à concevoir une architecture modulaire dotée d'une gestion rapide centralisée, de cadres d'évaluation automatisés et de modèles standardisés pour une mise à l'échelle efficace des solutions d'IA générative.

- Élargissez les cas d'utilisation : intégrez les fonctionnalités d'IA dans plusieurs départements et explorez de nouvelles applications.
- Améliorez l'expérience des développeurs : transformez la plateforme existante en plateforme interne en libre-service. Cette plate-forme est un environnement complet qui fournit des outils, des flux de travail et une gouvernance standardisés pour le développement de l'IA au sein de l'entreprise.
- Partage des connaissances — Établissez des pratiques internes et créez un marché de composants pour partager les actifs d'IA réutilisables entre les équipes. Les pratiques internes sont la stratégie qui consiste à appliquer une approche de développement open source au sein d'une organisation.
- Configurez l'évolutivité opérationnelle : améliorez votre infrastructure de support grâce à une réponse automatisée aux incidents et à une planification des capacités. Cela prépare l'infrastructure à évoluer en vue de l'adoption de l'IA générative à l'échelle de l'entreprise.
- Investissez dans des analyses avancées : utilisez des outils d'analyse avancés dans le cloud, tels qu'[Amazon CloudWatch](#) pour les métriques et [Amazon Quick Suite](#) pour la visualisation, afin d'utiliser l'analyse des données à des fins d'amélioration continue.
- Passez en revue le modèle de gouvernance des données : déterminez si votre modèle de gouvernance des données prend actuellement en charge les fonctionnalités de libre-service tout en maintenant des politiques et des contrôles d'accès standardisés. Une approche trop restrictive ou centralisée peut entraver votre capacité à étendre les initiatives relatives aux données au-delà de l'équipe principale, en particulier au sein de diverses unités commerciales.

En prenant ces mesures, les organisations peuvent :

- Développez les initiatives d'IA générative à l'échelle de l'organisation pour un impact à grande échelle.
- Continuez à améliorer la plateforme tout en identifiant les opportunités d'amélioration de la productivité et de la réutilisabilité.
- Améliorez l'expérience des développeurs et réduisez les charges cognitives.
- Favorisez une culture axée sur les données.
- Attirez les meilleurs talents en positionnant l'organisation comme un leader de l'IA générative.

Modèle de maturité de l'IA générative de niveau 4 : échelle

Le niveau 4 du modèle de maturité de l'IA générative, le niveau Scale, permet de passer de l'excellence opérationnelle à l'innovation évolutive. Organisations commencent à aller au-delà des déploiements de production individuels pour créer un écosystème robuste de composants réutilisables, de modèles standardisés et de flux de travail automatisés. Cet écosystème aide les organisations à accélérer l'adoption de l'IA générative dans plusieurs départements tout en maintenant une gouvernance robuste et une optimisation des coûts. En établissant des architectures évolutives et des fonctionnalités en libre-service, ces niveaux de maturité permettent aux entreprises de déployer efficacement de nombreuses applications d'IA générative, qui favorisent en fin de compte la transformation à l'échelle de l'organisation et l'innovation durable.

Cette section comprend les rubriques suivantes:

- [Objectif et critères](#)
- [Principales activités](#)

Objectif et critères

À ce niveau, les entreprises passent de l'excellence opérationnelle à l'innovation évolutive, en se concentrant sur la création de composants et de modèles réutilisables qui accélèrent l'adoption de l'IA générative au sein de l'entreprise. L'accent passe des déploiements de production individuels au développement de capacités qui permettent des fonctionnalités en libre-service, des modèles standardisés et des flux de travail automatisés tout en optimisant les coûts et en maintenant la gouvernance à grande échelle. Contrairement au niveau 3 qui se concentre sur certaines charges de travail de production, le niveau 4 permet le déploiement rapide d'un grand nombre d'applications d'IA génératives par le biais de composants standardisés et réutilisables, ce qui permet de réaliser des gains d'efficacité et de productivité à l'échelle de l'entreprise.

Les critères pour être à ce niveau sont les suivants :

- De nombreux départements ont adopté une utilisation généralisée de l'IA générative.
- L'organisation a mis en place une infrastructure d'IA générative et un écosystème d'outils à l'échelle de l'entreprise.
- Un modèle d'exploitation et une matrice RACI sont définis et mis en œuvre.

- Une bibliothèque disponible inclut des composants, des modèles et des applications d'IA normalisés et réutilisables. Les fonctionnalités en libre-service rendent la bibliothèque accessible dans l'ensemble de l'organisation.
- Les mécanismes de gouvernance automatisés fonctionnent à l'échelle de l'entreprise.
- L'organisation possède des preuves de pratiques et de résultats durables en matière d'innovation.

Principales activités

Le tableau suivant présente les principales activités pour chaque pilier de l'adoption.

Pilier de l'adoption	Activités
Entreprise	<ul style="list-style-type: none">• Alignez les projets d'IA générative avec les objectifs commerciaux à long terme. Concentrez-vous sur la croissance du chiffre d'affaires, la réduction des coûts et la satisfaction des clients.• Favorisez l'adoption de l'IA générative à l'échelle de l'entreprise grâce à des composants réutilisables et à des modèles standardisés qui apportent de la valeur.• Finalisez le modèle d'exploitation de l'IA générative et la matrice RACI pour les opérations à grande échelle.• Constituez des équipes spécialisées pour l'architecture, le développement et la maintenance des plateformes.• Créez des flux de travail de gouvernance et d'approbation standardisés.• Mettez en œuvre des analyses et une surveillance avancées pour une amélioration continue.• Établissez une approche proactive pour identifier les prochains cas d'utilisation innovants et à forte valeur ajoutée pour l'IA.

Pilier de l'adoption	Activités
	<p>Envisagez les cas d'utilisation internes qui améliorent la productivité et les cas d'utilisation externes axés sur les produits.</p> <ul style="list-style-type: none"> • Évaluez les opportunités d'automatisation des prises de décision complexes • Évaluez les possibilités de personnalisation et d'amélioration du produit
Personnes	<ul style="list-style-type: none"> • Formez le personnel à l'utilisation des outils d'IA générative et favorisez une culture d'apprentissage continu et d'innovation. • Au sein du centre d'excellence, développez des programmes de mentorat qui transfèrent les connaissances des experts en IA générative aux autres membres de l'équipe. • Utilisez un modèle interne ou participatif pour accélérer le développement des composants réutilisables de l'IA générative. • Exécutez des programmes de certification de l'IA par le biais d'un centre d'excellence.
Gouvernance	<ul style="list-style-type: none"> • Établissez des cadres éthiques et de gouvernance de l'IA à l'échelle de l'entreprise qui couvrent l'utilisation des données, l'équité des modèles et la transparence. • Développez les pratiques responsables en matière d'IA grâce à des cadres standardisés et à des garde-fous automatisés. • Établir des directives de contribution et des normes de qualité.

Pilier de l'adoption	Activités
Plateforme	<ul style="list-style-type: none"> Développez des composants d'IA réutilisables, tels que des architectures de microservices et des pipelines automatisés pour évaluer les solutions sous supervision humaine. Créez des modèles de solutions standardisés tels que des implémentations RAG et des flux de travail agentiques. Établissez un plan standardisé à intégrer à des outils tiers, en utilisant les normes du secteur telles que le Model Context Protocol (MCP). Mettez en œuvre des fonctionnalités de libre-service via un portail interne, tel qu'une architecture d'intégration axée sur les API et un marché de composants.
Sécurité	<ul style="list-style-type: none"> Mettez en œuvre des contrôles de sécurité de niveau entreprise et une vérification automatique de la conformité.
Opérations	<ul style="list-style-type: none"> Élaborez un processus et des directives pour soutenir un modèle de développement interne ou participatif. Déployez des cadres d'observabilité complets. Créez des tableaux de bord qui vous aident à surveiller les performances. Mettez en œuvre des systèmes automatisés pour recueillir des commentaires.

Poursuivre le parcours de maturité

Les entreprises qui ont atteint avec succès le niveau 4 du modèle de maturité de l'IA générative peuvent continuer à atteindre des niveaux de sophistication encore plus élevés. Cela nécessite une stratégie globale qui va au-delà de la mise en œuvre technique. Cette progression exige des initiatives stratégiques qui ancrent profondément l'IA générative dans l'ADN de l'organisation, alliant vision organisationnelle, transformation culturelle et excellence technique. Pour dépasser les quatre niveaux de maturité, les entreprises doivent renforcer leurs capacités internes, nouer des partenariats stratégiques et investir dans la recherche et le développement de pointe. Cette stratégie d'avancement globale, associée à un fort accent mis sur le développement des talents, permet aux entreprises d'aller au-delà des opérations à grande échelle pour adopter un leadership transformateur en matière d'IA. Cela permet d'améliorer l'efficacité opérationnelle et d'obtenir des avantages concurrentiels durables.

Envisagez les actions suivantes pour aller au-delà du modèle de maturité :

- **Intégrer l'IA générative dans la vision stratégique de l'organisation** — Positionner l'IA générative comme une composante essentielle de la mission et de la vision de l'entreprise. Assurez-vous d'utiliser ses capacités pour mener des initiatives stratégiques et conserver un avantage concurrentiel.
- **Favorisez une culture d'innovation continue** — Encouragez les employés à explorer de nouvelles applications de l'IA générative et récompensez les expériences qui correspondent aux objectifs de l'entreprise.
- **Collaborez avec des partenaires industriels et universitaires** : établissez des partenariats de recherche et collaborez avec des experts externes pour rester à la pointe de l'innovation en matière d'IA.
- **Investissez dans la recherche et le développement de pointe en matière d'IA générative** — Consacrez des ressources à l'exploration de nouvelles méthodologies, telles que l'IA multimodale et l'apprentissage par renforcement avancé, susceptibles de repousser les limites de l'IA générative.
- **Attirer et fidéliser les meilleurs talents en IA générative** — Concentrez-vous sur la création d'un solide vivier de talents en proposant des incitations attrayantes, des opportunités de développement professionnel et un environnement collaboratif.

En continuant à étendre les solutions d'IA générative à l'ensemble de l'organisation, les entreprises peuvent bénéficier des avantages suivants :

- Grand impact sur l'ensemble des unités commerciales — Les solutions d'IA générative sont intégrées aux opérations quotidiennes de plusieurs départements, ce qui améliore la productivité et l'efficacité.
- Prise de décision améliorée — Grâce aux informations en temps réel et aux capacités prédictives de l'IA générative, les entreprises peuvent prendre des décisions plus rapidement basées sur les données.
- Avantage concurrentiel stratégique — En utilisant l'IA générative à des fins d'innovation et d'optimisation, les entreprises peuvent se différencier de leurs concurrents et créer de nouvelles sources de revenus.
- IA générative mature platform/blueprints et gestion optimisée des ressources : en automatisant les processus et en améliorant la gestion des solutions génératives, vous pouvez réduire les coûts opérationnels et améliorer l'évolutivité.

Étapes suivantes

Le modèle de maturité de l'IA générative fournit une approche structurée permettant aux entreprises de poursuivre leur parcours d'adoption de l'IA générative AWS. La compréhension des différents niveaux de maturité et des différentes activités aide les organisations à évaluer leur niveau de préparation et à prendre des mesures éclairées pour exploiter le plein potentiel de l'IA générative. Ce cadre aide les organisations à développer des stratégies sur mesure qui correspondent à leurs objectifs commerciaux uniques afin que l'IA générative devienne un moteur clé de croissance et d'innovation.

Il est important de reconnaître que l'adoption de l'IA générative n'est pas un one-size-fits-all processus. Le parcours de chaque organisation est unique et est influencé par des facteurs tels que le secteur d'activité, les objectifs commerciaux et les capacités technologiques existantes. Cependant, ce document de stratégie constitue un guide précieux. Il offre un cadre permettant aux organisations d'évaluer leur état de préparation, d'identifier les lacunes et de mettre en œuvre les mesures nécessaires pour utiliser avec succès le potentiel de transformation de l'IA générative.

Alors que les entreprises entament leur parcours d'adoption de l'IA générative, elles doivent rester agiles et adaptables. Réévaluez continuellement votre niveau de maturité et ajustez vos stratégies en conséquence. Le rythme rapide de l'innovation dans le domaine de l'IA nécessite un engagement en faveur de l'apprentissage continu, du développement des compétences et de l'adoption des meilleures pratiques.

En suivant ces conseils et en utilisant les services d' AWS intelligence artificielle et de machine learning, les entreprises peuvent saisir de nouvelles opportunités, gagner en efficacité et obtenir un avantage concurrentiel durable dans un monde de plus en plus axé sur l'IA.

Ressources

Les ressources suivantes peuvent vous aider à en savoir plus sur l'adoption de l'IA générative.

Service AWS documentation

- [Amazon Bedrock](#)
- [Rambardes Amazon Bedrock](#)
- [Amazon Q Business](#)
- [Amazon Q Developer](#)

- [Amazon SageMaker AI](#)

AWS Conseils prescriptifs

- [Accélérer les cycles de développement des logiciels AWS grâce à l'IA générative](#)
- [Évaluation de la charge de travail générative de](#)
- [Récupérez les options et architectures de génération augmentée sur AWS](#)
- [Transformer les modèles opérationnels de développement et de maintenance d'applications AWS grâce à l'IA générative](#)

Autres ressources

- [L'état de l'IA : comment les organisations se recâblent pour générer de la valeur \(McKinsey rapport\)](#)
- [88 % des pilotes d'IA ne parviennent pas à atteindre la phase de production, mais cela n'est pas uniquement dû à l'informatique \(article du CIO\)](#)

Collaborateurs

Conception

- Haofei Feng, consultant principal en livraison, AWS
- Bin Liu, conseiller principal en livraison, AWS
- Chris Dorrington, consultant principal en matière de prestation, AWS
- Melanie Li, architecte de solutions senior, AWS
- Romain Vivier, responsable principal de l'architecture de solutions, AWS
- Sam Edwards, architecte de solutions, AWS
- Xin Chen, consultante principale en livraison, AWS

Révision

- Melchi Salins, architecte de solutions senior, AWS
- Junaid Baba, consultant principal en livraison, AWS

Rédaction technique

- Lilly AbouHarb, rédactrice technique senior, AWS

Historique du document

Le tableau suivant décrit les modifications importantes apportées à ce guide. Pour être averti des mises à jour à venir, abonnez-vous à un [fil RSS](#).

Modification	Description	Date
<u>Publication initiale</u>	—	4 juin 2025

AWS Glossaire des directives prescriptives

Les termes suivants sont couramment utilisés dans les stratégies, les guides et les modèles fournis par les directives AWS prescriptives. Pour suggérer des entrées, veuillez utiliser le lien Faire un commentaire à la fin du glossaire.

Nombres

7 R

Sept politiques de migration courantes pour transférer des applications vers le cloud. Ces politiques s'appuient sur les 5 R identifiés par Gartner en 2011 et sont les suivantes :

- Refactorisation/réarchitecture : transférez une application et modifiez son architecture en tirant pleinement parti des fonctionnalités natives cloud pour améliorer l'agilité, les performances et la capacité de mise à l'échelle. Cela implique généralement le transfert du système d'exploitation et de la base de données. Exemple : migrez votre base de données Oracle sur site vers l'édition compatible avec Amazon Aurora PostgreSQL.
- Replateformer (déplacer et remodeler) : transférez une application vers le cloud et introduisez un certain niveau d'optimisation pour tirer parti des fonctionnalités du cloud. Exemple : migrez votre base de données Oracle sur site vers Amazon Relational Database Service (Amazon RDS) pour Oracle dans le AWS Cloud
- Racheter (rachat) : optez pour un autre produit, généralement en passant d'une licence traditionnelle à un modèle SaaS. Exemple : migrez votre système de gestion de la relation client (CRM) vers Salesforce.com.
- Réhéberger (lift and shift) : transférez une application vers le cloud sans apporter de modifications pour tirer parti des fonctionnalités du cloud. Exemple : migrez votre base de données Oracle locale vers Oracle sur une EC2 instance du AWS Cloud.
- Relocaliser (lift and shift au niveau de l'hyperviseur) : transférez l'infrastructure vers le cloud sans acheter de nouveau matériel, réécrire des applications ou modifier vos opérations existantes. Vous migrez des serveurs d'une plateforme sur site vers un service cloud pour la même plateforme. Exemple : migrer une Microsoft Hyper-V application vers AWS.
- Retenir : conservez les applications dans votre environnement source. Il peut s'agir d'applications nécessitant une refactorisation majeure, que vous souhaitez retarder, et d'applications existantes que vous souhaitez retenir, car rien ne justifie leur migration sur le plan commercial.

- Retirer : mettez hors service ou supprimez les applications dont vous n'avez plus besoin dans votre environnement source.

A

ABAC

Voir contrôle [d'accès basé sur les attributs](#).

services abstraits

Consultez la section [Services gérés](#).

ACIDE

Voir [atomicité, consistance, isolation, durabilité](#).

migration active-active

Méthode de migration de base de données dans laquelle la synchronisation des bases de données source et cible est maintenue (à l'aide d'un outil de réplication bidirectionnelle ou d'opérations d'écriture double), tandis que les deux bases de données gèrent les transactions provenant de la connexion d'applications pendant la migration. Cette méthode prend en charge la migration par petits lots contrôlés au lieu d'exiger un basculement ponctuel. Elle est plus flexible mais demande plus de travail qu'une migration [active-passive](#).

migration active-passive

Méthode de migration de base de données dans laquelle les bases de données source et cible sont synchronisées, mais seule la base de données source gère les transactions liées à la connexion des applications pendant que les données sont répliquées vers la base de données cible. La base de données cible n'accepte aucune transaction pendant la migration.

fonction d'agrégation

Fonction SQL qui agit sur un groupe de lignes et calcule une valeur de retour unique pour le groupe. Des exemples de fonctions d'agrégation incluent SUM et MAX.

AI

Voir [intelligence artificielle](#).

AIOps

Voir les [opérations d'intelligence artificielle](#).

anonymisation

Processus de suppression définitive d'informations personnelles dans un ensemble de données. L'anonymisation peut contribuer à protéger la vie privée. Les données anonymisées ne sont plus considérées comme des données personnelles.

anti-motif

Solution fréquemment utilisée pour un problème récurrent lorsque la solution est contre-productive, inefficace ou moins efficace qu'une alternative.

contrôle des applications

Une approche de sécurité qui permet d'utiliser uniquement des applications approuvées afin de protéger un système contre les logiciels malveillants.

portefeuille d'applications

Ensemble d'informations détaillées sur chaque application utilisée par une organisation, y compris le coût de génération et de maintenance de l'application, ainsi que sa valeur métier. Ces informations sont essentielles pour [le processus de découverte et d'analyse du portefeuille](#) et permettent d'identifier et de prioriser les applications à migrer, à moderniser et à optimiser.

intelligence artificielle (IA)

Domaine de l'informatique consacré à l'utilisation des technologies de calcul pour exécuter des fonctions cognitives généralement associées aux humains, telles que l'apprentissage, la résolution de problèmes et la reconnaissance de modèles. Pour plus d'informations, veuillez consulter [Qu'est-ce que l'intelligence artificielle ?](#)

opérations d'intelligence artificielle (AIOps)

Processus consistant à utiliser des techniques de machine learning pour résoudre les problèmes opérationnels, réduire les incidents opérationnels et les interventions humaines, mais aussi améliorer la qualité du service. Pour plus d'informations sur son AIOps utilisation dans la stratégie de AWS migration, consultez le [guide d'intégration des opérations](#).

chiffrement asymétrique

Algorithme de chiffrement qui utilise une paire de clés, une clé publique pour le chiffrement et une clé privée pour le déchiffrement. Vous pouvez partager la clé publique, car elle n'est pas utilisée pour le déchiffrement, mais l'accès à la clé privée doit être très restreint.

atomicité, cohérence, isolement, durabilité (ACID)

Ensemble de propriétés logicielles garantissant la validité des données et la fiabilité opérationnelle d'une base de données, même en cas d'erreur, de panne de courant ou d'autres problèmes.

contrôle d'accès par attributs (ABAC)

Pratique qui consiste à créer des autorisations détaillées en fonction des attributs de l'utilisateur, tels que le service, le poste et le nom de l'équipe. Pour plus d'informations, consultez [ABAC pour AWS](#) dans la documentation Gestion des identités et des accès AWS (IAM).

source de données faisant autorité

Emplacement où vous stockez la version principale des données, considérée comme la source d'information la plus fiable. Vous pouvez copier les données de la source de données officielle vers d'autres emplacements à des fins de traitement ou de modification des données, par exemple en les anonymisant, en les expurgant ou en les pseudonymisant.

Zone de disponibilité

Un emplacement distinct au sein d'un Région AWS réseau isolé des défaillances dans d'autres zones de disponibilité et fournissant une connectivité réseau peu coûteuse et à faible latence aux autres zones de disponibilité de la même région.

AWS Cadre d'adoption du cloud (AWS CAF)

Un cadre de directives et de meilleures pratiques visant AWS à aider les entreprises à élaborer un plan efficace pour réussir leur migration vers le cloud. AWS La CAF organise ses conseils en six domaines prioritaires appelés perspectives : les affaires, les personnes, la gouvernance, les plateformes, la sécurité et les opérations. Les perspectives d'entreprise, de personnes et de gouvernance mettent l'accent sur les compétences et les processus métier, tandis que les perspectives relatives à la plateforme, à la sécurité et aux opérations se concentrent sur les compétences et les processus techniques. Par exemple, la perspective liée aux personnes cible les parties prenantes qui s'occupent des ressources humaines (RH), des fonctions de dotation en personnel et de la gestion des personnes. Dans cette perspective, la AWS CAF fournit des conseils pour le développement du personnel, la formation et les communications afin de préparer l'organisation à une adoption réussie du cloud. Pour plus d'informations, veuillez consulter le [site Web AWS CAF](#) et le [livre blanc AWS CAF](#).

AWS Cadre de qualification de la charge de travail (AWS WQF)

Outil qui évalue les charges de travail liées à la migration des bases de données, recommande des stratégies de migration et fournit des estimations de travail. AWS Le WQF est inclus avec

AWS Schema Conversion Tool (AWS SCT). Il analyse les schémas de base de données et les objets de code, le code d'application, les dépendances et les caractéristiques de performance, et fournit des rapports d'évaluation.

B

mauvais bot

Un [bot](#) destiné à perturber ou à nuire à des individus ou à des organisations.

BCP

Consultez la section [Planification de la continuité des activités](#).

graphique de comportement

Vue unifiée et interactive des comportements des ressources et des interactions au fil du temps. Vous pouvez utiliser un graphique de comportement avec Amazon Detective pour examiner les tentatives de connexion infructueuses, les appels d'API suspects et les actions similaires. Pour plus d'informations, veuillez consulter [Data in a behavior graph](#) dans la documentation Detective.

système de poids fort

Système qui stocke d'abord l'octet le plus significatif. Voir aussi [endiánité](#).

classification binaire

Processus qui prédit un résultat binaire (l'une des deux classes possibles). Par exemple, votre modèle de machine learning peut avoir besoin de prévoir des problèmes tels que « Cet e-mail est-il du spam ou non ? » ou « Ce produit est-il un livre ou une voiture ? ».

filtre de Bloom

Structure de données probabiliste et efficace en termes de mémoire qui est utilisée pour tester si un élément fait partie d'un ensemble.

déploiement bleu/vert

Stratégie de déploiement dans laquelle vous créez deux environnements distincts mais identiques. Vous exécutez la version actuelle de l'application dans un environnement (bleu) et la nouvelle version de l'application dans l'autre environnement (vert). Cette stratégie vous permet de revenir rapidement en arrière avec un impact minimal.

bot

Application logicielle qui exécute des tâches automatisées sur Internet et simule l'activité ou l'interaction humaine. Certains robots sont utiles ou bénéfiques, comme les robots d'exploration Web qui indexent des informations sur Internet. D'autres robots, appelés « bots malveillants », sont destinés à perturber ou à nuire à des individus ou à des organisations.

botnet

Réseaux de [robots](#) infectés par des [logiciels malveillants](#) et contrôlés par une seule entité, connue sous le nom d'herder ou d'opérateur de bots. Les botnets sont le mécanisme le plus connu pour faire évoluer les bots et leur impact.

branche

Zone contenue d'un référentiel de code. La première branche créée dans un référentiel est la branche principale. Vous pouvez créer une branche à partir d'une branche existante, puis développer des fonctionnalités ou corriger des bogues dans la nouvelle branche. Une branche que vous créez pour générer une fonctionnalité est communément appelée branche de fonctionnalités. Lorsque la fonctionnalité est prête à être publiée, vous fusionnez à nouveau la branche de fonctionnalités dans la branche principale. Pour plus d'informations, consultez [À propos des branches](#) (GitHub documentation).

accès par brise-vitre

Dans des circonstances exceptionnelles et par le biais d'un processus approuvé, il s'agit d'un moyen rapide pour un utilisateur d'accéder à un accès auquel Compte AWS il n'est généralement pas autorisé. Pour plus d'informations, consultez l'indicateur [Implementation break-glass procedures](#) dans le guide Well-Architected AWS .

stratégie existante (brownfield)

L'infrastructure existante de votre environnement. Lorsque vous adoptez une stratégie existante pour une architecture système, vous concevez l'architecture en fonction des contraintes des systèmes et de l'infrastructure actuels. Si vous étendez l'infrastructure existante, vous pouvez combiner des politiques brownfield (existantes) et [greenfield](#) (inédites).

cache de tampon

Zone de mémoire dans laquelle sont stockées les données les plus fréquemment consultées.

capacité métier

Ce que fait une entreprise pour générer de la valeur (par exemple, les ventes, le service client ou le marketing). Les architectures de microservices et les décisions de développement

peuvent être dictées par les capacités métier. Pour plus d'informations, veuillez consulter la section [Organisation en fonction des capacités métier](#) du livre blanc [Exécution de microservices conteneurisés sur AWS](#).

planification de la continuité des activités (BCP)

Plan qui tient compte de l'impact potentiel d'un événement perturbateur, tel qu'une migration à grande échelle, sur les opérations, et qui permet à une entreprise de reprendre ses activités rapidement.

C

CAF

Voir le [cadre d'adoption du AWS cloud](#).

déploiement de Canary

Diffusion lente et progressive d'une version pour les utilisateurs finaux. Lorsque vous êtes sûr, vous déployez la nouvelle version et remplacez la version actuelle dans son intégralité.

CCo E

Voir [le Centre d'excellence du cloud](#).

CDC

Consultez la section [Capture des données de modification](#).

capture des données de modification (CDC)

Processus de suivi des modifications apportées à une source de données, telle qu'une table de base de données, et d'enregistrement des métadonnées relatives à ces modifications. Vous pouvez utiliser la CDC à diverses fins, telles que l'audit ou la réplication des modifications dans un système cible afin de maintenir la synchronisation.

ingénierie du chaos

Introduire intentionnellement des défaillances ou des événements perturbateurs pour tester la résilience d'un système. Vous pouvez utiliser [AWS Fault Injection Service \(AWS FIS\)](#) pour effectuer des expériences qui stressent vos AWS charges de travail et évaluer leur réponse.

CI/CD

Découvrez [l'intégration continue et la livraison continue](#).

classification

Processus de catégorisation qui permet de générer des prédictions. Les modèles de ML pour les problèmes de classification prédisent une valeur discrète. Les valeurs discrètes se distinguent toujours les unes des autres. Par exemple, un modèle peut avoir besoin d'évaluer la présence ou non d'une voiture sur une image.

chiffrement côté client

Chiffrement des données localement, avant que la cible ne les Service AWS reçoive.

Centre d'excellence du cloud (CCoE)

Une équipe multidisciplinaire qui dirige les efforts d'adoption du cloud au sein d'une organisation, notamment en développant les bonnes pratiques en matière de cloud, en mobilisant des ressources, en établissant des délais de migration et en guidant l'organisation dans le cadre de transformations à grande échelle. Pour plus d'informations, consultez les [CCoarticles électroniques](#) du blog sur la stratégie AWS Cloud d'entreprise.

cloud computing

Technologie cloud généralement utilisée pour le stockage de données à distance et la gestion des appareils IoT. Le cloud computing est généralement associé à la technologie [informatique de pointe](#).

modèle d'exploitation du cloud

Dans une organisation informatique, modèle d'exploitation utilisé pour créer, faire évoluer et optimiser un ou plusieurs environnements cloud. Pour plus d'informations, consultez la section [Création de votre modèle d'exploitation cloud](#).

étapes d'adoption du cloud

Les quatre phases que les entreprises traversent généralement lorsqu'elles migrent vers AWS Cloud :

- Projet : exécution de quelques projets liés au cloud à des fins de preuve de concept et d'apprentissage
- Base : réaliser des investissements fondamentaux pour accélérer votre adoption du cloud (par exemple, créer une zone de landing zone, définir un CCo E, établir un modèle opérationnel)
- Migration : migration d'applications individuelles
- Réinvention : optimisation des produits et services et innovation dans le cloud

Ces étapes ont été définies par Stephen Orban dans le billet de blog [The Journey Toward Cloud-First & the Stages of Adoption](#) publié sur le blog AWS Cloud Enterprise Strategy. Pour plus d'informations sur leur lien avec la stratégie de AWS migration, consultez le [guide de préparation à la migration](#).

CMDB

Consultez la base de [données de gestion des configurations](#).

référentiel de code

Emplacement où le code source et d'autres ressources, comme la documentation, les exemples et les scripts, sont stockés et mis à jour par le biais de processus de contrôle de version. Les référentiels cloud courants incluent GitHub ou Bitbucket Cloud. Chaque version du code est appelée branche. Dans une structure de microservice, chaque référentiel est consacré à une seule fonctionnalité. Un seul pipeline CI/CD peut utiliser plusieurs référentiels.

cache passif

Cache tampon vide, mal rempli ou contenant des données obsolètes ou non pertinentes. Cela affecte les performances, car l'instance de base de données doit lire à partir de la mémoire principale ou du disque, ce qui est plus lent que la lecture à partir du cache tampon.

données gelées

Données rarement consultées et généralement historiques. Lorsque vous interrogez ce type de données, les requêtes lentes sont généralement acceptables. Le transfert de ces données vers des niveaux ou classes de stockage moins performants et moins coûteux peut réduire les coûts.

vision par ordinateur (CV)

Domaine de l'[IA](#) qui utilise l'apprentissage automatique pour analyser et extraire des informations à partir de formats visuels tels que des images numériques et des vidéos. Par exemple, Amazon SageMaker AI fournit des algorithmes de traitement d'image pour les CV.

dérive de configuration

Pour une charge de travail, une modification de configuration par rapport à l'état attendu. Cela peut entraîner une non-conformité de la charge de travail, et cela est généralement progressif et involontaire.

base de données de gestion des configurations (CMDB)

Référentiel qui stocke et gère les informations relatives à une base de données et à son environnement informatique, y compris les composants matériels et logiciels ainsi que leurs

configurations. Vous utilisez généralement les données d'une CMDB lors de la phase de découverte et d'analyse du portefeuille de la migration.

pack de conformité

Ensemble de AWS Config règles et d'actions correctives que vous pouvez assembler pour personnaliser vos contrôles de conformité et de sécurité. Vous pouvez déployer un pack de conformité en tant qu'entité unique dans une région Compte AWS et, ou au sein d'une organisation, à l'aide d'un modèle YAML. Pour plus d'informations, consultez la section [Packs de conformité](#) dans la AWS Config documentation.

intégration continue et livraison continue (CI/CD)

Processus d'automatisation des étapes de source, de construction, de test, de préparation et de production du processus de publication du logiciel. CI/CD est communément décrit comme un pipeline. CI/CD peut vous aider à automatiser les processus, à améliorer la productivité, à améliorer la qualité du code et à accélérer les livraisons. Pour plus d'informations, veuillez consulter [Avantages de la livraison continue](#). CD peut également signifier déploiement continu. Pour plus d'informations, veuillez consulter [Livraison continue et déploiement continu](#).

CV

Voir [vision par ordinateur](#).

D

données au repos

Données stationnaires dans votre réseau, telles que les données stockées.

classification des données

Processus permettant d'identifier et de catégoriser les données de votre réseau en fonction de leur sévérité et de leur sensibilité. Il s'agit d'un élément essentiel de toute stratégie de gestion des risques de cybersécurité, car il vous aide à déterminer les contrôles de protection et de conservation appropriés pour les données. La classification des données est une composante du pilier de sécurité du AWS Well-Architected Framework. Pour plus d'informations, veuillez consulter [Classification des données](#).

dérive des données

Une variation significative entre les données de production et les données utilisées pour entraîner un modèle ML, ou une modification significative des données d'entrée au fil du temps. La dérive

des données peut réduire la qualité, la précision et l'équité globales des prédictions des modèles ML.

données en transit

Données qui circulent activement sur votre réseau, par exemple entre les ressources du réseau.

maillage de données

Un cadre architectural qui fournit une propriété des données distribuée et décentralisée avec une gestion et une gouvernance centralisées.

minimisation des données

Le principe de collecte et de traitement des seules données strictement nécessaires. La pratique de la minimisation des données AWS Cloud peut réduire les risques liés à la confidentialité, les coûts et l'empreinte carbone de vos analyses.

périmètre de données

Ensemble de garde-fous préventifs dans votre AWS environnement qui permettent de garantir que seules les identités fiables accèdent aux ressources fiables des réseaux attendus. Pour plus d'informations, voir [Création d'un périmètre de données sur AWS](#).

prétraitement des données

Pour transformer les données brutes en un format facile à analyser par votre modèle de ML. Le prétraitement des données peut impliquer la suppression de certaines colonnes ou lignes et le traitement des valeurs manquantes, incohérentes ou en double.

provenance des données

Le processus de suivi de l'origine et de l'historique des données tout au long de leur cycle de vie, par exemple la manière dont les données ont été générées, transmises et stockées.

sujet des données

Personne dont les données sont collectées et traitées.

entrepôt des données

Un système de gestion des données qui prend en charge les informations commerciales, telles que les analyses. Les entrepôts de données contiennent généralement de grandes quantités de données historiques et sont généralement utilisés pour les requêtes et les analyses.

langage de définition de base de données (DDL)

Instructions ou commandes permettant de créer ou de modifier la structure des tables et des objets dans une base de données.

langage de manipulation de base de données (DML)

Instructions ou commandes permettant de modifier (insérer, mettre à jour et supprimer) des informations dans une base de données.

DDL

Voir [langage de définition de base](#) de données.

ensemble profond

Sert à combiner plusieurs modèles de deep learning à des fins de prédiction. Vous pouvez utiliser des ensembles profonds pour obtenir une prévision plus précise ou pour estimer l'incertitude des prédictions.

deep learning

Un sous-champ de ML qui utilise plusieurs couches de réseaux neuronaux artificiels pour identifier le mappage entre les données d'entrée et les variables cibles d'intérêt.

defense-in-depth

Approche de la sécurité de l'information dans laquelle une série de mécanismes et de contrôles de sécurité sont judicieusement répartis sur l'ensemble d'un réseau informatique afin de protéger la confidentialité, l'intégrité et la disponibilité du réseau et des données qu'il contient. Lorsque vous adoptez cette stratégie AWS, vous ajoutez plusieurs contrôles à différentes couches de la AWS Organizations structure afin de sécuriser les ressources. Par exemple, une defense-in-depth approche peut combiner l'authentification multifactorielle, la segmentation du réseau et le chiffrement.

administrateur délégué

Dans AWS Organizations, un service compatible peut enregistrer un compte AWS membre pour administrer les comptes de l'organisation et gérer les autorisations pour ce service. Ce compte est appelé administrateur délégué pour ce service. Pour plus d'informations et une liste des services compatibles, veuillez consulter la rubrique [Services qui fonctionnent avec AWS Organizations](#) dans la documentation AWS Organizations .

déploiement

Processus de mise à disposition d'une application, de nouvelles fonctionnalités ou de corrections de code dans l'environnement cible. Le déploiement implique la mise en œuvre de modifications dans une base de code, puis la génération et l'exécution de cette base de code dans les environnements de l'application.

environnement de développement

Voir [environnement](#).

contrôle de détection

Contrôle de sécurité conçu pour détecter, journaliser et alerter après la survenue d'un événement. Ces contrôles constituent une deuxième ligne de défense et vous alertent en cas d'événements de sécurité qui ont contourné les contrôles préventifs en place. Pour plus d'informations, veuillez consulter la rubrique [Contrôles de détection](#) dans [Implementing security controls on AWS](#).

cartographie de la chaîne de valeur du développement (DVSM)

Processus utilisé pour identifier et hiérarchiser les contraintes qui nuisent à la rapidité et à la qualité du cycle de vie du développement logiciel. DVSM étend le processus de cartographie de la chaîne de valeur initialement conçu pour les pratiques de production allégée. Il met l'accent sur les étapes et les équipes nécessaires pour créer et transférer de la valeur tout au long du processus de développement logiciel.

jumeau numérique

Représentation virtuelle d'un système réel, tel qu'un bâtiment, une usine, un équipement industriel ou une ligne de production. Les jumeaux numériques prennent en charge la maintenance prédictive, la surveillance à distance et l'optimisation de la production.

tableau des dimensions

Dans un [schéma en étoile](#), table plus petite contenant les attributs de données relatifs aux données quantitatives d'une table de faits. Les attributs des tables de dimensions sont généralement des champs de texte ou des nombres discrets qui se comportent comme du texte. Ces attributs sont couramment utilisés pour la contrainte des requêtes, le filtrage et l'étiquetage des ensembles de résultats.

catastrophe

Un événement qui empêche une charge de travail ou un système d'atteindre ses objectifs commerciaux sur son site de déploiement principal. Ces événements peuvent être des

catastrophes naturelles, des défaillances techniques ou le résultat d'actions humaines, telles qu'une mauvaise configuration involontaire ou une attaque de logiciel malveillant.

reprise après sinistre (DR)

La stratégie et le processus que vous utilisez pour minimiser les temps d'arrêt et les pertes de données causés par un [sinistre](#). Pour plus d'informations, consultez [Disaster Recovery of Workloads on AWS : Recovery in the Cloud in the AWS Well-Architected Framework](#).

DML

Voir [langage de manipulation de base](#) de données.

conception axée sur le domaine

Approche visant à développer un système logiciel complexe en connectant ses composants à des domaines évolutifs, ou objectifs métier essentiels, que sert chaque composant. Ce concept a été introduit par Eric Evans dans son ouvrage Domain-Driven Design: Tackling Complexity in the Heart of Software (Boston : Addison-Wesley Professional, 2003). Pour plus d'informations sur l'utilisation du design piloté par domaine avec le modèle de figuier étrangleur, veuillez consulter [Modernizing legacy Microsoft ASP.NET \(ASMX\) web services incrementally by using containers and Amazon API Gateway](#).

DR

Voir [reprise après sinistre](#).

détection de dérive

Suivi des écarts par rapport à une configuration de référence. Par exemple, vous pouvez l'utiliser AWS CloudFormation pour [détecter la dérive des ressources du système](#) ou AWS Control Tower pour [détecter les modifications de votre zone d'atterrissement](#) susceptibles d'affecter le respect des exigences de gouvernance.

DVSM

Voir la [cartographie de la chaîne de valeur du développement](#).

E

EDA

Voir [analyse exploratoire des données](#).

EDI

Voir échange [de données informatisé](#).

informatique de périphérie

Technologie qui augmente la puissance de calcul des appareils intelligents en périphérie d'un réseau IoT. Comparé au [cloud computing, l'informatique](#) de pointe peut réduire la latence des communications et améliorer le temps de réponse.

échange de données informatisé (EDI)

L'échange automatique de documents commerciaux entre les organisations. Pour plus d'informations, voir [Qu'est-ce que l'échange de données informatisé ?](#)

chiffrement

Processus informatique qui transforme des données en texte clair, lisibles par l'homme, en texte chiffré.

clé de chiffrement

Chaîne cryptographique de bits aléatoires générée par un algorithme cryptographique. La longueur des clés peut varier, et chaque clé est conçue pour être imprévisible et unique.

endianisme

Ordre selon lequel les octets sont stockés dans la mémoire de l'ordinateur. Les systèmes de poids fort stockent d'abord l'octet le plus significatif. Les systèmes de poids faible stockent d'abord l'octet le moins significatif.

point de terminaison

Voir [point de terminaison de service](#).

service de point de terminaison

Service que vous pouvez héberger sur un cloud privé virtuel (VPC) pour le partager avec d'autres utilisateurs. Vous pouvez créer un service de point de terminaison avec AWS PrivateLink et accorder des autorisations à d'autres Comptes AWS ou à Gestion des identités et des accès AWS (IAM) principaux. Ces comptes ou principaux peuvent se connecter à votre service de point de terminaison de manière privée en créant des points de terminaison d'un VPC d'interface.

Pour plus d'informations, veuillez consulter [Création d'un service de point de terminaison](#) dans la documentation Amazon Virtual Private Cloud (Amazon VPC).

planification des ressources d'entreprise (ERP)

Système qui automatise et gère les principaux processus métier (tels que la comptabilité, le [MES](#) et la gestion de projet) pour une entreprise.

chiffrement d'enveloppe

Processus de chiffrement d'une clé de chiffrement à l'aide d'une autre clé de chiffrement. Pour plus d'informations, consultez la section [Chiffrement des enveloppes](#) dans la documentation AWS Key Management Service (AWS KMS).

environment

Instance d'une application en cours d'exécution. Les types d'environnement les plus courants dans le cloud computing sont les suivants :

- Environnement de développement : instance d'une application en cours d'exécution à laquelle seule l'équipe principale chargée de la maintenance de l'application peut accéder. Les environnements de développement sont utilisés pour tester les modifications avant de les promouvoir dans les environnements supérieurs. Ce type d'environnement est parfois appelé environnement de test.
- Environnements inférieurs : tous les environnements de développement d'une application, tels que ceux utilisés pour les générations et les tests initiaux.
- Environnement de production : instance d'une application en cours d'exécution à laquelle les utilisateurs finaux peuvent accéder. Dans un CI/CD pipeline, l'environnement de production est le dernier environnement de déploiement.
- Environnements supérieurs : tous les environnements accessibles aux utilisateurs autres que l'équipe de développement principale. Ils peuvent inclure un environnement de production, des environnements de préproduction et des environnements pour les tests d'acceptation par les utilisateurs.

épopée

Dans les méthodologies agiles, catégories fonctionnelles qui aident à organiser et à prioriser votre travail. Les épopées fournissent une description détaillée des exigences et des tâches d'implémentation. Par exemple, les points forts de la AWS CAF en matière de sécurité incluent la gestion des identités et des accès, les contrôles de détection, la sécurité des infrastructures, la protection des données et la réponse aux incidents. Pour plus d'informations sur les épopées dans la stratégie de migration AWS , veuillez consulter le [guide d'implémentation du programme](#).

ERP

Voir [Planification des ressources d'entreprise](#).

analyse exploratoire des données (EDA)

Processus d'analyse d'un jeu de données pour comprendre ses principales caractéristiques. Vous collectez ou agrégez des données, puis vous effectuez des enquêtes initiales pour trouver des modèles, détecter des anomalies et vérifier les hypothèses. L'EDA est réalisée en calculant des statistiques récapitulatives et en créant des visualisations de données.

F

tableau des faits

La table centrale dans un [schéma en étoile](#). Il stocke des données quantitatives sur les opérations commerciales. Généralement, une table de faits contient deux types de colonnes : celles qui contiennent des mesures et celles qui contiennent une clé étrangère pour une table de dimensions.

échouer rapidement

Une philosophie qui utilise des tests fréquents et progressifs pour réduire le cycle de vie du développement. C'est un élément essentiel d'une approche agile.

limite d'isolation des défauts

Dans le AWS Cloud, une limite telle qu'une zone de disponibilité Région AWS, un plan de contrôle ou un plan de données qui limite l'effet d'une panne et contribue à améliorer la résilience des charges de travail. Pour plus d'informations, consultez la section [Limites d'isolation des AWS pannes](#).

branche de fonctionnalités

Voir [succursale](#).

fonctionnalités

Les données d'entrée que vous utilisez pour faire une prédition. Par exemple, dans un contexte de fabrication, les fonctionnalités peuvent être des images capturées périodiquement à partir de la ligne de fabrication.

importance des fonctionnalités

Le niveau d'importance d'une fonctionnalité pour les prédictions d'un modèle. Il s'exprime généralement sous la forme d'un score numérique qui peut être calculé à l'aide de différentes techniques, telles que la méthode Shapley Additive Explanations (SHAP) et les gradients intégrés. Pour plus d'informations, voir [Interprétabilité du modèle d'apprentissage automatique avec AWS](#).

transformation de fonctionnalité

Optimiser les données pour le processus de ML, notamment en enrichissant les données avec des sources supplémentaires, en mettant à l'échelle les valeurs ou en extrayant plusieurs ensembles d'informations à partir d'un seul champ de données. Cela permet au modèle de ML de tirer parti des données. Par exemple, si vous décomposez la date « 2021-05-27 00:15:37 » en « 2021 », « mai », « jeudi » et « 15 », vous pouvez aider l'algorithme d'apprentissage à apprendre des modèles nuancés associés à différents composants de données.

invitation en quelques coups

Fournir à un [LLM](#) un petit nombre d'exemples illustrant la tâche et le résultat souhaité avant de lui demander d'effectuer une tâche similaire. Cette technique est une application de l'apprentissage contextuel, dans le cadre de laquelle les modèles apprennent à partir d'exemples (prises de vue) intégrés dans des instructions. Les instructions en quelques clics peuvent être efficaces pour les tâches qui nécessitent un formatage, un raisonnement ou des connaissances de domaine spécifiques. Voir également l'[invite Zero-Shot](#).

FGAC

Découvrez le [contrôle d'accès détaillé](#).

contrôle d'accès détaillé (FGAC)

Utilisation de plusieurs conditions pour autoriser ou refuser une demande d'accès.

migration instantanée (flash-cut)

Méthode de migration de base de données qui utilise la réplication continue des données par [le biais de la capture des données de modification](#) afin de migrer les données dans les plus brefs délais, au lieu d'utiliser une approche progressive. L'objectif est de réduire au maximum les temps d'arrêt.

FM

Voir le [modèle de fondation](#).

modèle de fondation (FM)

Un vaste réseau neuronal d'apprentissage profond qui s'est entraîné sur d'énormes ensembles de données généralisées et non étiquetées. FMs sont capables d'effectuer une grande variété de tâches générales, telles que comprendre le langage, générer du texte et des images et converser en langage naturel. Pour plus d'informations, voir [Que sont les modèles de base ?](#)

G

IA générative

Sous-ensemble de modèles d'[IA](#) qui ont été entraînés sur de grandes quantités de données et qui peuvent utiliser une simple invite textuelle pour créer de nouveaux contenus et artefacts, tels que des images, des vidéos, du texte et du son. Pour plus d'informations, consultez [Qu'est-ce que l'IA générative.](#)

blocage géographique

Voir les [restrictions géographiques.](#)

restrictions géographiques (blocage géographique)

Sur Amazon CloudFront, option permettant d'empêcher les utilisateurs de certains pays d'accéder aux distributions de contenu. Vous pouvez utiliser une liste d'autorisation ou une liste de blocage pour spécifier les pays approuvés et interdits. Pour plus d'informations, consultez [la section Restreindre la distribution géographique de votre contenu](#) dans la CloudFront documentation.

Flux de travail Gitflow

Approche dans laquelle les environnements inférieurs et supérieurs utilisent différentes branches dans un référentiel de code source. Le flux de travail Gitflow est considéré comme existant, et le [flux de travail basé sur les troncs](#) est l'approche moderne préférée.

image dorée

Un instantané d'un système ou d'un logiciel utilisé comme modèle pour déployer de nouvelles instances de ce système ou logiciel. Par exemple, dans le secteur de la fabrication, une image dorée peut être utilisée pour fournir des logiciels sur plusieurs appareils et contribue à améliorer la vitesse, l'évolutivité et la productivité des opérations de fabrication des appareils.

stratégie inédite

L'absence d'infrastructures existantes dans un nouvel environnement. Lorsque vous adoptez une stratégie inédite pour une architecture système, vous pouvez sélectionner toutes les nouvelles technologies sans restriction de compatibilité avec l'infrastructure existante, également appelée [brownfield](#). Si vous étendez l'infrastructure existante, vous pouvez combiner des politiques brownfield (existantes) et greenfield (inédites).

barrière de protection

Règle de haut niveau qui permet de régir les ressources, les politiques et la conformité au sein des unités organisationnelles (OUs). Les barrières de protection préventives appliquent des politiques pour garantir l'alignement sur les normes de conformité. Elles sont mises en œuvre à l'aide de politiques de contrôle des services et de limites des autorisations IAM. Les barrières de protection de détection détectent les violations des politiques et les problèmes de conformité, et génèrent des alertes pour y remédier. Ils sont implémentés à l'aide d'Amazon AWS Config AWS Security Hub CSPM GuardDuty AWS Trusted Advisor, d'Amazon Inspector et de AWS Lambda contrôles personnalisés.

H

HA

Découvrez [la haute disponibilité](#).

migration de base de données hétérogène

Migration de votre base de données source vers une base de données cible qui utilise un moteur de base de données différent (par exemple, Oracle vers Amazon Aurora). La migration hétérogène fait généralement partie d'un effort de réarchitecture, et la conversion du schéma peut s'avérer une tâche complexe. [AWS propose AWS SCT](#) qui facilite les conversions de schémas.

haute disponibilité (HA)

Capacité d'une charge de travail à fonctionner en continu, sans intervention, en cas de difficultés ou de catastrophes. Les systèmes HA sont conçus pour basculer automatiquement, fournir constamment des performances de haute qualité et gérer différentes charges et défaillances avec un impact minimal sur les performances.

modernisation des historiens

Approche utilisée pour moderniser et mettre à niveau les systèmes de technologie opérationnelle (OT) afin de mieux répondre aux besoins de l'industrie manufacturière. Un historien est un type de base de données utilisé pour collecter et stocker des données provenant de diverses sources dans une usine.

données de rétention

Partie de données historiques étiquetées qui n'est pas divulguée dans un ensemble de données utilisé pour entraîner un modèle d'[apprentissage automatique](#). Vous pouvez utiliser les données de blocage pour évaluer les performances du modèle en comparant les prévisions du modèle aux données de blocage.

migration de base de données homogène

Migration de votre base de données source vers une base de données cible qui partage le même moteur de base de données (par exemple, Microsoft SQL Server vers Amazon RDS for SQL Server). La migration homogène s'inscrit généralement dans le cadre d'un effort de réhébergement ou de replateforme. Vous pouvez utiliser les utilitaires de base de données natifs pour migrer le schéma.

données chaudes

Données fréquemment consultées, telles que les données en temps réel ou les données translationnelles récentes. Ces données nécessitent généralement un niveau ou une classe de stockage à hautes performances pour fournir des réponses rapides aux requêtes.

correctif

Solution d'urgence à un problème critique dans un environnement de production. En raison de son urgence, un correctif est généralement créé en dehors du flux de travail de DevOps publication habituel.

période de soins intensifs

Immédiatement après le basculement, période pendant laquelle une équipe de migration gère et surveille les applications migrées dans le cloud afin de résoudre les problèmes éventuels. En règle générale, cette période dure de 1 à 4 jours. À la fin de la période de soins intensifs, l'équipe de migration transfère généralement la responsabilité des applications à l'équipe des opérations cloud.

I

laC

Considérez [l'infrastructure comme un code](#).

politique basée sur l'identité

Politique attachée à un ou plusieurs principaux IAM qui définit leurs autorisations au sein de l'AWS Cloud environnement.

application inactive

Application dont l'utilisation moyenne du processeur et de la mémoire se situe entre 5 et 20 % sur une période de 90 jours. Dans un projet de migration, il est courant de retirer ces applications ou de les retenir sur site.

Illo T

Voir [Internet industriel des objets](#).

infrastructure immuable

Modèle qui déploie une nouvelle infrastructure pour les charges de travail de production au lieu de mettre à jour, d'appliquer des correctifs ou de modifier l'infrastructure existante. Les infrastructures immuables sont intrinsèquement plus cohérentes, fiables et prévisibles que les infrastructures [mutables](#). Pour plus d'informations, consultez les meilleures pratiques de [déploiement à l'aide d'une infrastructure immuable](#) dans le AWS Well-Architected Framework.

VPC entrant (d'entrée)

Dans une architecture AWS multi-comptes, un VPC qui accepte, inspecte et achemine les connexions réseau depuis l'extérieur d'une application. L'[architecture AWS de référence de sécurité](#) recommande de configurer votre compte réseau avec les fonctions entrantes, sortantes et d'inspection VPCs afin de protéger l'interface bidirectionnelle entre votre application et l'Internet en général.

migration incrémentielle

Stratégie de basculement dans le cadre de laquelle vous migrez votre application par petites parties au lieu d'effectuer un basculement complet unique. Par exemple, il se peut que vous ne transfériez que quelques microservices ou utilisateurs vers le nouveau système dans un premier temps. Après avoir vérifié que tout fonctionne correctement, vous pouvez transférer

progressivement des microservices ou des utilisateurs supplémentaires jusqu'à ce que vous puissiez mettre hors service votre système hérité. Cette stratégie réduit les risques associés aux migrations de grande ampleur.

Industry 4.0

Un terme introduit par [Klaus Schwab](#) en 2016 pour désigner la modernisation des processus de fabrication grâce aux avancées en matière de connectivité, de données en temps réel, d'automatisation, d'analyse et d'IA/ML.

infrastructure

Ensemble des ressources et des actifs contenus dans l'environnement d'une application.

infrastructure en tant que code (IaC)

Processus de mise en service et de gestion de l'infrastructure d'une application via un ensemble de fichiers de configuration. IaC est conçue pour vous aider à centraliser la gestion de l'infrastructure, à normaliser les ressources et à mettre à l'échelle rapidement afin que les nouveaux environnements soient reproductibles, fiables et cohérents.

Internet industriel des objets (IIoT)

L'utilisation de capteurs et d'appareils connectés à Internet dans les secteurs industriels tels que la fabrication, l'énergie, l'automobile, les soins de santé, les sciences de la vie et l'agriculture.

Pour plus d'informations, voir [Élaboration d'une stratégie de transformation numérique de l'Internet des objets \(IIoT\) industriel](#).

VPC d'inspection

Dans une architecture AWS multi-comptes, un VPC centralisé qui gère les inspections du trafic réseau VPCs entre (identique ou Régions AWS différent), Internet et les réseaux locaux.

L'[architecture AWS de référence de sécurité](#) recommande de configurer votre compte réseau avec les fonctions entrantes, sortantes et d'inspection VPCs afin de protéger l'interface bidirectionnelle entre votre application et l'Internet en général.

Internet des objets (IoT)

Réseau d'objets physiques connectés dotés de capteurs ou de processeurs intégrés qui communiquent avec d'autres appareils et systèmes via Internet ou via un réseau de communication local. Pour plus d'informations, veuillez consulter la section [Qu'est-ce que l'IoT ?](#)

interprétabilité

Caractéristique d'un modèle de machine learning qui décrit dans quelle mesure un être humain peut comprendre comment les prédictions du modèle dépendent de ses entrées. Pour plus d'informations, voir [Interprétabilité du modèle d'apprentissage automatique avec AWS](#).

IoT

Voir [Internet des objets](#).

Bibliothèque d'informations informatiques (ITIL)

Ensemble de bonnes pratiques pour proposer des services informatiques et les aligner sur les exigences métier. L'ITIL constitue la base de l'ITSM.

gestion des services informatiques (ITSM)

Activités associées à la conception, à la mise en œuvre, à la gestion et à la prise en charge de services informatiques d'une organisation. Pour plus d'informations sur l'intégration des opérations cloud aux outils ITSM, veuillez consulter le [guide d'intégration des opérations](#).

ITIL

Consultez la [bibliothèque d'informations informatiques](#).

ITSM

Voir [Gestion des services informatiques](#).

L

contrôle d'accès basé sur des étiquettes (LBAC)

Une implémentation du contrôle d'accès obligatoire (MAC) dans laquelle une valeur d'étiquette de sécurité est explicitement attribuée aux utilisateurs et aux données elles-mêmes. L'intersection entre l'étiquette de sécurité utilisateur et l'étiquette de sécurité des données détermine les lignes et les colonnes visibles par l'utilisateur.

zone de destination

Une zone d'atterrissement est un AWS environnement multi-comptes bien conçu, évolutif et sécurisé. Il s'agit d'un point de départ à partir duquel vos entreprises peuvent rapidement lancer et déployer des charges de travail et des applications en toute confiance dans leur environnement de sécurité et d'infrastructure. Pour plus d'informations sur les zones de destination, veuillez consulter [Setting up a secure and scalable multi-account AWS environment](#).

grand modèle de langage (LLM)

Un modèle d'[intelligence artificielle basé](#) sur le deep learning qui est préentraîné sur une grande quantité de données. Un LLM peut effectuer plusieurs tâches, telles que répondre à des questions, résumer des documents, traduire du texte dans d'autres langues et compléter des phrases. Pour plus d'informations, voir [Que sont LLMs.](#)

migration de grande envergure

Migration de 300 serveurs ou plus.

LBAC

Voir contrôle d'[accès basé sur des étiquettes.](#)

principe de moindre privilège

Bonne pratique de sécurité qui consiste à accorder les autorisations minimales nécessaires à l'exécution d'une tâche. Pour plus d'informations, veuillez consulter la rubrique [Accorder les autorisations de moindre privilège](#) dans la documentation IAM.

lift and shift

Voir [7 Rs.](#)

système de poids faible

Système qui stocke d'abord l'octet le moins significatif. Voir aussi [endianité.](#)

LLM

Voir le [grand modèle de langage.](#)

environnements inférieurs

Voir [environnement.](#)

M

machine learning (ML)

Type d'intelligence artificielle qui utilise des algorithmes et des techniques pour la reconnaissance et l'apprentissage de modèles. Le ML analyse et apprend à partir de données enregistrées, telles que les données de l'Internet des objets (IoT), pour générer un modèle statistique basé sur des modèles. Pour plus d'informations, veuillez consulter [Machine Learning.](#)

branche principale

Voir [succursale](#).

malware

Logiciel conçu pour compromettre la sécurité ou la confidentialité de l'ordinateur. Les logiciels malveillants peuvent perturber les systèmes informatiques, divulguer des informations sensibles ou obtenir un accès non autorisé. Parmi les malwares, on peut citer les virus, les vers, les rançongiciels, les chevaux de Troie, les logiciels espions et les enregistreurs de frappe.

services gérés

Services AWS pour lequel AWS fonctionnent la couche d'infrastructure, le système d'exploitation et les plateformes, et vous accédez aux points de terminaison pour stocker et récupérer des données. Amazon Simple Storage Service (Amazon S3) et Amazon DynamoDB sont des exemples de services gérés. Ils sont également appelés services abstraits.

système d'exécution de la fabrication (MES)

Un système logiciel pour le suivi, la surveillance, la documentation et le contrôle des processus de production qui convertissent les matières premières en produits finis dans l'atelier.

MAP

Voir [Migration Acceleration Program](#).

mécanisme

Processus complet au cours duquel vous créez un outil, favorisez son adoption, puis inspectez les résultats afin de procéder aux ajustements nécessaires. Un mécanisme est un cycle qui se renforce et s'améliore lorsqu'il fonctionne. Pour plus d'informations, voir [Création de mécanismes](#) dans le cadre AWS Well-Architected.

compte membre

Tous, à l' Comptes AWS exception du compte de gestion, qui font partie d'une organisation dans AWS Organizations. Un compte ne peut être membre que d'une seule organisation à la fois.

MAILLES

Voir le [système d'exécution de la fabrication](#).

Transport télémétrique en file d'attente de messages (MQTT)

[Protocole de communication léger machine-to-machine \(M2M\), basé sur le modèle de publication/d'abonnement, pour les appareils IoT aux ressources limitées.](#)

microservice

Un petit service indépendant qui communique via un réseau bien défini APIs et qui est généralement détenu par de petites équipes autonomes. Par exemple, un système d'assurance peut inclure des microservices qui mappent à des capacités métier, telles que les ventes ou le marketing, ou à des sous-domaines, tels que les achats, les réclamations ou l'analytique. Les avantages des microservices incluent l'agilité, la flexibilité de la mise à l'échelle, la facilité de déploiement, la réutilisation du code et la résilience. Pour plus d'informations, consultez la section [Intégration de microservices à l'aide de services AWS sans serveur](#).

architecture de microservices

Approche de création d'une application avec des composants indépendants qui exécutent chaque processus d'application en tant que microservice. Ces microservices communiquent via une interface bien définie en utilisant Lightweight APIs. Chaque microservice de cette architecture peut être mis à jour, déployé et mis à l'échelle pour répondre à la demande de fonctions spécifiques d'une application. Pour plus d'informations, consultez la section [Implémentation de microservices sur AWS](#).

Programme d'accélération des migrations (MAP)

Un AWS programme qui fournit un support de conseil, des formations et des services pour aider les entreprises à établir une base opérationnelle solide pour passer au cloud, et pour aider à compenser le coût initial des migrations. MAP inclut une méthodologie de migration pour exécuter les migrations héritées de manière méthodique, ainsi qu'un ensemble d'outils pour automatiser et accélérer les scénarios de migration courants.

migration à grande échelle

Processus consistant à transférer la majeure partie du portefeuille d'applications vers le cloud par vagues, un plus grand nombre d'applications étant déplacées plus rapidement à chaque vague. Cette phase utilise les bonnes pratiques et les enseignements tirés des phases précédentes pour implémenter une usine de migration d'équipes, d'outils et de processus en vue de rationaliser la migration des charges de travail grâce à l'automatisation et à la livraison agile. Il s'agit de la troisième phase de la [stratégie de migration AWS](#).

usine de migration

Équipes interfonctionnelles qui rationalisent la migration des charges de travail grâce à des approches automatisées et agiles. Les équipes de Migration Factory comprennent généralement des responsables des opérations, des analystes commerciaux et des propriétaires, des ingénieurs de migration, des développeurs et DevOps des professionnels travaillant dans le cadre de sprints.

Entre 20 et 50 % du portefeuille d'applications d'entreprise est constitué de modèles répétés qui peuvent être optimisés par une approche d'usine. Pour plus d'informations, veuillez consulter la rubrique [discussion of migration factories](#) et le [guide Cloud Migration Factory](#) dans cet ensemble de contenus.

métadonnées de migration

Informations relatives à l'application et au serveur nécessaires pour finaliser la migration. Chaque modèle de migration nécessite un ensemble de métadonnées de migration différent. Les exemples de métadonnées de migration incluent le sous-réseau cible, le groupe de sécurité et le AWS compte.

modèle de migration

Tâche de migration reproductible qui détaille la stratégie de migration, la destination de la migration et l'application ou le service de migration utilisé. Exemple : réorganisez la migration vers Amazon EC2 avec le service de migration AWS d'applications.

Évaluation du portefeuille de migration (MPA)

Outil en ligne qui fournit des informations pour valider l'analyse de rentabilisation en faveur de la migration vers le AWS Cloud. La MPA propose une évaluation détaillée du portefeuille (dimensionnement approprié des serveurs, tarification, comparaison du coût total de possession, analyse des coûts de migration), ainsi que la planification de la migration (analyse et collecte des données d'applications, regroupement des applications, priorisation des migrations et planification des vagues). L'[outil MPA](#) (connexion requise) est disponible gratuitement pour tous les AWS consultants et consultants APN Partner.

Évaluation de la préparation à la migration (MRA)

Processus qui consiste à obtenir des informations sur l'état de préparation d'une organisation au cloud, à identifier les forces et les faiblesses et à élaborer un plan d'action pour combler les lacunes identifiées, à l'aide du AWS CAF. Pour plus d'informations, veuillez consulter le [guide de préparation à la migration](#). La MRA est la première phase de la [stratégie de migration AWS](#).

stratégie de migration

L'approche utilisée pour migrer une charge de travail vers le AWS Cloud. Pour plus d'informations, reportez-vous aux [7 R](#) de ce glossaire et à [Mobiliser votre organisation pour accélérer les migrations à grande échelle](#).

ML

Voir [apprentissage automatique](#).

modernisation

Transformation d'une application obsolète (héritée ou monolithique) et de son infrastructure en un système agile, élastique et hautement disponible dans le cloud afin de réduire les coûts, de gagner en efficacité et de tirer parti des innovations. Pour plus d'informations, consultez [la section Stratégie de modernisation des applications dans le AWS Cloud](#).

évaluation de la préparation à la modernisation

Évaluation qui permet de déterminer si les applications d'une organisation sont prêtes à être modernisées, d'identifier les avantages, les risques et les dépendances, et qui détermine dans quelle mesure l'organisation peut prendre en charge l'état futur de ces applications. Le résultat de l'évaluation est un plan de l'architecture cible, une feuille de route détaillant les phases de développement et les étapes du processus de modernisation, ainsi qu'un plan d'action pour combler les lacunes identifiées. Pour plus d'informations, consultez la section [Évaluation de l'état de préparation à la modernisation des applications dans le AWS Cloud](#).

applications monolithiques (monolithes)

Applications qui s'exécutent en tant que service unique avec des processus étroitement couplés. Les applications monolithiques ont plusieurs inconvénients. Si une fonctionnalité de l'application connaît un pic de demande, l'architecture entière doit être mise à l'échelle. L'ajout ou l'amélioration des fonctionnalités d'une application monolithique devient également plus complexe lorsque la base de code s'élargit. Pour résoudre ces problèmes, vous pouvez utiliser une architecture de microservices. Pour plus d'informations, veuillez consulter [Decomposing monoliths into microservices](#).

MPA

Voir [Évaluation du portefeuille de migration](#).

MQTT

Voir [Message Queuing Telemetry Transport](#).

classification multi-classes

Processus qui permet de générer des prédictions pour plusieurs classes (prédiction d'un résultat parmi plus de deux). Par exemple, un modèle de ML peut demander « Ce produit est-il un livre, une voiture ou un téléphone ? » ou « Quelle catégorie de produits intéresse le plus ce client ? ».

infrastructure mutable

Modèle qui met à jour et modifie l'infrastructure existante pour les charges de travail de production. Pour améliorer la cohérence, la fiabilité et la prévisibilité, le AWS Well-Architected Framework recommande l'utilisation [d'une infrastructure immuable comme](#) meilleure pratique.

O

OAC

Voir [Contrôle d'accès à l'origine.](#)

OAI

Voir [l'identité d'accès à l'origine.](#)

OCM

Voir [gestion du changement organisationnel.](#)

migration hors ligne

Méthode de migration dans laquelle la charge de travail source est supprimée au cours du processus de migration. Cette méthode implique un temps d'arrêt prolongé et est généralement utilisée pour de petites charges de travail non critiques.

OI

Consultez la section [Intégration des opérations.](#)

OLA

Voir l'accord [au niveau opérationnel.](#)

migration en ligne

Méthode de migration dans laquelle la charge de travail source est copiée sur le système cible sans être mise hors ligne. Les applications connectées à la charge de travail peuvent continuer à fonctionner pendant la migration. Cette méthode implique un temps d'arrêt nul ou minimal et est généralement utilisée pour les charges de travail de production critiques.

OPC-UA

Voir [Open Process Communications - Architecture unifiée.](#)

Communications par processus ouvert - Architecture unifiée (OPC-UA)

Un protocole de communication machine-to-machine (M2M) pour l'automatisation industrielle. L'OPC-UA fournit une norme d'interopérabilité avec des schémas de cryptage, d'authentification et d'autorisation des données.

accord au niveau opérationnel (OLA)

Accord qui précise ce que les groupes informatiques fonctionnels s'engagent à fournir les uns aux autres, afin de prendre en charge un contrat de niveau de service (SLA).

examen de l'état de préparation opérationnelle (ORR)

Une liste de questions et de bonnes pratiques associées qui vous aident à comprendre, à évaluer, à prévenir ou à réduire l'ampleur des incidents et des défaillances possibles. Pour plus d'informations, voir [Operational Readiness Reviews \(ORR\)](#) dans le AWS Well-Architected Framework.

technologie opérationnelle (OT)

Systèmes matériels et logiciels qui fonctionnent avec l'environnement physique pour contrôler les opérations, les équipements et les infrastructures industriels. Dans le secteur manufacturier, l'intégration des systèmes OT et des technologies de l'information (IT) est au cœur des transformations de [l'industrie 4.0](#).

intégration des opérations (OI)

Processus de modernisation des opérations dans le cloud, qui implique la planification de la préparation, l'automatisation et l'intégration. Pour en savoir plus, veuillez consulter le [guide d'intégration des opérations](#).

journal de suivi d'organisation

Un parcours créé par AWS CloudTrail qui enregistre tous les événements pour tous les membres Comptes AWS d'une organisation dans AWS Organizations. Ce journal de suivi est créé dans chaque Compte AWS qui fait partie de l'organisation et suit l'activité de chaque compte. Pour plus d'informations, consultez [la section Crédit d'un suivi pour une organisation dans la CloudTrail documentation](#).

gestion du changement organisationnel (OCM)

Cadre pour gérer les transformations métier majeures et perturbatrices du point de vue des personnes, de la culture et du leadership. L'OCM aide les organisations à se préparer et à effectuer la transition vers de nouveaux systèmes et de nouvelles politiques en accélérant

l'adoption des changements, en abordant les problèmes de transition et en favorisant des changements culturels et organisationnels. Dans la stratégie de AWS migration, ce cadre est appelé accélération du personnel, en raison de la rapidité du changement requise dans les projets d'adoption du cloud. Pour plus d'informations, veuillez consulter le [guide OCM](#).

contrôle d'accès d'origine (OAC)

Dans CloudFront, une option améliorée pour restreindre l'accès afin de sécuriser votre contenu Amazon Simple Storage Service (Amazon S3). L'OAC prend en charge tous les compartiments S3 dans leur ensemble Régions AWS, le chiffrement côté serveur avec AWS KMS (SSE-KMS) et les requêtes dynamiques PUT adressées au compartiment S3. DELETE

identité d'accès d'origine (OAI)

Dans CloudFront, une option permettant de restreindre l'accès afin de sécuriser votre contenu Amazon S3. Lorsque vous utilisez OAI, il CloudFront crée un principal auprès duquel Amazon S3 peut s'authentifier. Les principaux authentifiés ne peuvent accéder au contenu d'un compartiment S3 que par le biais d'une distribution spécifique CloudFront . Voir également [OAC](#), qui fournit un contrôle d'accès plus précis et amélioré.

ORR

Voir [l'examen de l'état de préparation opérationnelle](#).

DE

Voir [technologie opérationnelle](#).

VPC sortant (de sortie)

Dans une architecture AWS multi-comptes, un VPC qui gère les connexions réseau initiées depuis une application. L'[architecture AWS de référence de sécurité](#) recommande de configurer votre compte réseau avec les fonctions entrantes, sortantes et d'inspection VPCs afin de protéger l'interface bidirectionnelle entre votre application et l'Internet en général.

P

limite des autorisations

Politique de gestion IAM attachée aux principaux IAM pour définir les autorisations maximales que peut avoir l'utilisateur ou le rôle. Pour plus d'informations, veuillez consulter la rubrique [Limites des autorisations](#) dans la documentation IAM.

informations personnelles identifiables (PII)

Informations qui, lorsqu'elles sont consultées directement ou associées à d'autres données connexes, peuvent être utilisées pour déduire raisonnablement l'identité d'une personne. Les exemples d'informations personnelles incluent les noms, les adresses et les coordonnées.

PII

Voir les [informations personnelles identifiables](#).

manuel stratégique

Ensemble d'étapes prédéfinies qui capturent le travail associé aux migrations, comme la fourniture de fonctions d'opérations de base dans le cloud. Un manuel stratégique peut revêtir la forme de scripts, de runbooks automatisés ou d'un résumé des processus ou des étapes nécessaires au fonctionnement de votre environnement modernisé.

PLC

Voir [contrôleur logique programmable](#).

PLM

Consultez la section [Gestion du cycle de vie des](#) produits.

politique

Objet capable de définir les autorisations (voir la [politique basée sur l'identité](#)), de spécifier les conditions d'accès (voir la [politique basée sur les ressources](#)) ou de définir les autorisations maximales pour tous les comptes d'une organisation dans AWS Organizations (voir la politique de contrôle des [services](#)).

persistance polyglotte

Choix indépendant de la technologie de stockage de données d'un microservice en fonction des modèles d'accès aux données et d'autres exigences. Si vos microservices utilisent la même technologie de stockage de données, ils peuvent rencontrer des difficultés d'implémentation ou présenter des performances médiocres. Les microservices sont plus faciles à mettre en œuvre, atteignent de meilleures performances, ainsi qu'une meilleure capacité de mise à l'échelle s'ils utilisent l'entrepôt de données le mieux adapté à leurs besoins. Pour plus d'informations, veuillez consulter [Enabling data persistence in microservices](#).

évaluation du portefeuille

Processus de découverte, d'analyse et de priorisation du portefeuille d'applications afin de planifier la migration. Pour plus d'informations, veuillez consulter [Evaluating migration readiness](#).

predicate

Une condition de requête qui renvoie `true` ou `false`, généralement située dans une `WHERE` clause.

prédictat pushdown

Technique d'optimisation des requêtes de base de données qui filtre les données de la requête avant le transfert. Cela réduit la quantité de données qui doivent être extraites et traitées à partir de la base de données relationnelle et améliore les performances des requêtes.

contrôle préventif

Contrôle de sécurité conçu pour empêcher qu'un événement ne se produise. Ces contrôles constituent une première ligne de défense pour empêcher tout accès non autorisé ou toute modification indésirable de votre réseau. Pour plus d'informations, veuillez consulter [Preventative controls](#) dans [Implementing security controls on AWS](#).

principal

Entité AWS capable d'effectuer des actions et d'accéder aux ressources. Cette entité est généralement un utilisateur root pour un Compte AWS rôle IAM ou un utilisateur. Pour plus d'informations, veuillez consulter la rubrique Principal dans [Termes et concepts relatifs aux rôles](#), dans la documentation IAM.

confidentialité dès la conception

Une approche d'ingénierie système qui prend en compte la confidentialité tout au long du processus de développement.

zones hébergées privées

Conteneur contenant des informations sur la manière dont vous souhaitez qu'Amazon Route 53 réponde aux requêtes DNS pour un domaine et ses sous-domaines au sein d'un ou de plusieurs VPCs domaines. Pour plus d'informations, veuillez consulter [Working with private hosted zones](#) dans la documentation Route 53.

contrôle proactif

[Contrôle de sécurité](#) conçu pour empêcher le déploiement de ressources non conformes. Ces contrôles analysent les ressources avant qu'elles ne soient provisionnées. Si la ressource n'est pas conforme au contrôle, elle n'est pas provisionnée. Pour plus d'informations, consultez le [guide de référence sur les contrôles](#) dans la AWS Control Tower documentation et consultez la section [Contrôles proactifs dans Implémentation](#) des contrôles de sécurité sur AWS.

gestion du cycle de vie des produits (PLM)

Gestion des données et des processus d'un produit tout au long de son cycle de vie, depuis la conception, le développement et le lancement, en passant par la croissance et la maturité, jusqu'au déclin et au retrait.

environnement de production

Voir [environnement](#).

contrôleur logique programmable (PLC)

Dans le secteur manufacturier, un ordinateur hautement fiable et adaptable qui surveille les machines et automatise les processus de fabrication.

chaînage rapide

Utiliser le résultat d'une invite [LLM](#) comme entrée pour l'invite suivante afin de générer de meilleures réponses. Cette technique est utilisée pour décomposer une tâche complexe en sous-tâches ou pour affiner ou développer de manière itérative une réponse préliminaire. Cela permet d'améliorer la précision et la pertinence des réponses d'un modèle et permet d'obtenir des résultats plus précis et personnalisés.

pseudonymisation

Processus de remplacement des identifiants personnels dans un ensemble de données par des valeurs fictives. La pseudonymisation peut contribuer à protéger la vie privée. Les données pseudonymisées sont toujours considérées comme des données personnelles.

publish/subscribe (pub/sub)

Modèle qui permet des communications asynchrones entre les microservices afin d'améliorer l'évolutivité et la réactivité. Par exemple, dans un [MES](#) basé sur des microservices, un microservice peut publier des messages d'événements sur un canal auquel d'autres microservices peuvent s'abonner. Le système peut ajouter de nouveaux microservices sans modifier le service de publication.

Q

plan de requête

Série d'étapes, telles que des instructions, utilisées pour accéder aux données d'un système de base de données relationnelle SQL.

régression du plan de requêtes

Le cas où un optimiseur de service de base de données choisit un plan moins optimal qu'avant une modification donnée de l'environnement de base de données. Cela peut être dû à des changements en termes de statistiques, de contraintes, de paramètres d'environnement, de liaisons de paramètres de requêtes et de mises à jour du moteur de base de données.

R

Matrice RACI

Voir [responsable, responsable, consulté, informé \(RACI\)](#).

CHIFFON

Voir [Retrieval Augmented Generation](#).

rançongiciel

Logiciel malveillant conçu pour bloquer l'accès à un système informatique ou à des données jusqu'à ce qu'un paiement soit effectué.

Matrice RASCI

Voir [responsable, responsable, consulté, informé \(RACI\)](#).

RCAC

Voir [contrôle d'accès aux lignes et aux colonnes](#).

réplica en lecture

Copie d'une base de données utilisée en lecture seule. Vous pouvez acheminer les requêtes vers le réplica de lecture pour réduire la charge sur votre base de données principale.

réarchitecte

Voir [7 Rs.](#)

objectif de point de récupération (RPO)

Durée maximale acceptable depuis le dernier point de récupération des données. Il détermine ce qui est considéré comme étant une perte de données acceptable entre le dernier point de reprise et l'interruption du service.

objectif de temps de récupération (RTO)

Le délai maximum acceptable entre l'interruption du service et le rétablissement du service.

refactoriser

Voir [7 Rs.](#)

Région

Ensemble de AWS ressources dans une zone géographique. Chacun Région AWS est isolé et indépendant des autres pour garantir tolérance aux pannes, stabilité et résilience. Pour plus d'informations, voir [Spécifier ce que Régions AWS votre compte peut utiliser](#).

régression

Technique de ML qui prédit une valeur numérique. Par exemple, pour résoudre le problème « Quel sera le prix de vente de cette maison ? », un modèle de ML pourrait utiliser un modèle de régression linéaire pour prédire le prix de vente d'une maison sur la base de faits connus à son sujet (par exemple, la superficie en mètres carrés).

réhéberger

Voir [7 Rs.](#)

version

Dans un processus de déploiement, action visant à promouvoir les modifications apportées à un environnement de production.

déplacer

Voir [7 Rs.](#)

replateforme

Voir [7 Rs.](#)

rachat

Voir [7 Rs.](#)

résilience

La capacité d'une application à résister aux perturbations ou à s'en remettre. [La haute disponibilité et la reprise après sinistre](#) sont des considérations courantes lors de la planification de la résilience dans le AWS Cloud. Pour plus d'informations, consultez [AWS Cloud Résilience](#).

politique basée sur les ressources

Politique attachée à une ressource, comme un compartiment Amazon S3, un point de terminaison ou une clé de chiffrement. Ce type de politique précise les principaux auxquels l'accès est autorisé, les actions prises en charge et toutes les autres conditions qui doivent être remplies. matrice responsable, redevable, consulté et informé (RACI)

Une matrice qui définit les rôles et les responsabilités de toutes les parties impliquées dans les activités de migration et les opérations cloud. Le nom de la matrice est dérivé des types de responsabilité définis dans la matrice : responsable (R), responsable (A), consulté (C) et informé (I). Le type de support (S) est facultatif. Si vous incluez le support, la matrice est appelée matrice RASCI, et si vous l'excluez, elle est appelée matrice RACI.

contrôle réactif

Contrôle de sécurité conçu pour permettre de remédier aux événements indésirables ou aux écarts par rapport à votre référence de sécurité. Pour plus d'informations, veuillez consulter la rubrique [Responsive controls](#) dans [Implementing security controls on AWS](#).

retain

Voir [7 Rs.](#)

se retirer

Voir [7 Rs.](#)

Génération augmentée de récupération (RAG)

Technologie d'[IA générative](#) dans laquelle un [LLM](#) fait référence à une source de données faisant autorité qui se trouve en dehors de ses sources de données de formation avant de générer une réponse. Par exemple, un modèle RAG peut effectuer une recherche sémantique dans la base de connaissances ou dans les données personnalisées d'une organisation. Pour plus d'informations, voir [Qu'est-ce que RAG ?](#)

rotation

Processus de mise à jour périodique d'un [secret](#) pour empêcher un attaquant d'accéder aux informations d'identification.

contrôle d'accès aux lignes et aux colonnes (RCAC)

Utilisation d'expressions SQL simples et flexibles dotées de règles d'accès définies. Le RCAC comprend des autorisations de ligne et des masques de colonnes.

RPO

Voir l'[objectif du point de récupération](#).

RTO

Voir l'[objectif en matière de temps de rétablissement](#).

runbook

Ensemble de procédures manuelles ou automatisées nécessaires à l'exécution d'une tâche spécifique. Elles visent généralement à rationaliser les opérations ou les procédures répétitives présentant des taux d'erreur élevés.

S

SAML 2.0

Un standard ouvert utilisé par de nombreux fournisseurs d'identité (IdPs). Cette fonctionnalité permet l'authentification unique fédérée (SSO), afin que les utilisateurs puissent se connecter AWS Management Console ou appeler les opérations de l' AWS API sans que vous ayez à créer un utilisateur dans IAM pour tous les membres de votre organisation. Pour plus d'informations sur la fédération SAML 2.0, veuillez consulter [À propos de la fédération SAML 2.0](#) dans la documentation IAM.

SCADA

Voir [Contrôle de supervision et acquisition de données](#).

SCP

Voir la [politique de contrôle des services](#).

secret

Dans AWS Secrets Manager des informations confidentielles ou restreintes, telles qu'un mot de passe ou des informations d'identification utilisateur, que vous stockez sous forme cryptée. Il comprend la valeur secrète et ses métadonnées. La valeur secrète peut être binaire, une chaîne unique ou plusieurs chaînes. Pour plus d'informations, voir [Que contient le secret d'un Secrets Manager ?](#) dans la documentation de Secrets Manager.

sécurité dès la conception

Une approche d'ingénierie système qui prend en compte la sécurité tout au long du processus de développement.

contrôle de sécurité

Barrière de protection technique ou administrative qui empêche, détecte ou réduit la capacité d'un assaillant d'exploiter une vulnérabilité de sécurité. Il existe quatre principaux types de contrôles de sécurité : [préventifs](#), [détectifs](#), [réactifs](#) et [proactifs](#).

renforcement de la sécurité

Processus qui consiste à réduire la surface d'attaque pour la rendre plus résistante aux attaques. Cela peut inclure des actions telles que la suppression de ressources qui ne sont plus requises, la mise en œuvre des bonnes pratiques de sécurité consistant à accorder le moindre privilège ou la désactivation de fonctionnalités inutiles dans les fichiers de configuration.

système de gestion des informations et des événements de sécurité (SIEM)

Outils et services qui associent les systèmes de gestion des informations de sécurité (SIM) et de gestion des événements de sécurité (SEM). Un système SIEM collecte, surveille et analyse les données provenant de serveurs, de réseaux, d'appareils et d'autres sources afin de détecter les menaces et les failles de sécurité, mais aussi de générer des alertes.

automatisation des réponses de sécurité

Action prédéfinie et programmée conçue pour répondre automatiquement à un événement de sécurité ou y remédier. Ces automatisations servent de contrôles de sécurité [détectifs ou réactifs](#) qui vous aident à mettre en œuvre les meilleures pratiques AWS de sécurité. Parmi les actions de réponse automatique, citons la modification d'un groupe de sécurité VPC, l'application de correctifs à une EC2 instance Amazon ou la rotation des informations d'identification.

chiffrement côté serveur

Chiffrement des données à destination, par celui Service AWS qui les reçoit.

Politique de contrôle des services (SCP)

Politique qui fournit un contrôle centralisé des autorisations pour tous les comptes d'une organisation dans AWS Organizations. SCPs définissez des garde-fous ou des limites aux actions qu'un administrateur peut déléguer à des utilisateurs ou à des rôles. Vous pouvez les utiliser SCPs comme listes d'autorisation ou de refus pour spécifier les services ou les actions autorisés ou interdits. Pour plus d'informations, consultez la section [Politiques de contrôle des services](#) dans la AWS Organizations documentation.

point de terminaison du service

URL du point d'entrée pour un Service AWS. Pour vous connecter par programmation au service cible, vous pouvez utiliser un point de terminaison. Pour plus d'informations, veuillez consulter la rubrique [Service AWS endpoints](#) dans Références générales AWS.

contrat de niveau de service (SLA)

Accord qui précise ce qu'une équipe informatique promet de fournir à ses clients, comme le temps de disponibilité et les performances des services.

indicateur de niveau de service (SLI)

Mesure d'un aspect des performances d'un service, tel que son taux d'erreur, sa disponibilité ou son débit.

objectif de niveau de service (SLO)

Mesure cible qui représente l'état d'un service, tel que mesuré par un indicateur de [niveau de service](#).

modèle de responsabilité partagée

Un modèle décrivant la responsabilité que vous partagez en matière AWS de sécurité et de conformité dans le cloud. AWS est responsable de la sécurité du cloud, alors que vous êtes responsable de la sécurité dans le cloud. Pour de plus amples informations, veuillez consulter [Modèle de responsabilité partagée](#).

SIEM

Consultez les [informations de sécurité et le système de gestion des événements](#).

point de défaillance unique (SPOF)

Défaillance d'un seul composant critique d'une application susceptible de perturber le système.

SLA

Voir le contrat [de niveau de service](#).

SLI

Voir l'indicateur de [niveau de service](#).

SLO

Voir l'objectif de [niveau de service](#).

split-and-seed modèle

Modèle permettant de mettre à l'échelle et d'accélérer les projets de modernisation. Au fur et à mesure que les nouvelles fonctionnalités et les nouvelles versions de produits sont définies, l'équipe principale se divise pour créer des équipes de produit. Cela permet de mettre à l'échelle les capacités et les services de votre organisation, d'améliorer la productivité des développeurs et de favoriser une innovation rapide. Pour plus d'informations, consultez la section [Approche progressive de la modernisation des applications dans le AWS Cloud](#)

SPOF

Voir [point de défaillance unique](#).

schéma en étoile

Structure organisationnelle de base de données qui utilise une grande table de faits pour stocker les données transactionnelles ou mesurées et utilise une ou plusieurs tables dimensionnelles plus petites pour stocker les attributs des données. Cette structure est conçue pour être utilisée dans un [entrepôt de données](#) ou à des fins de business intelligence.

modèle de figuier étrangleur

Approche de modernisation des systèmes monolithiques en réécrivant et en remplaçant progressivement les fonctionnalités du système jusqu'à ce que le système hérité puisse être mis hors service. Ce modèle utilise l'analogie d'un figuier de vigne qui se développe dans un arbre existant et qui finit par supplanter son hôte. Le schéma a été [présenté par Martin Fowler](#) comme un moyen de gérer les risques lors de la réécriture de systèmes monolithiques. Pour obtenir un exemple d'application de ce modèle, veuillez consulter [Modernizing legacy Microsoft ASP.NET \(ASMX\) web services incrementally by using containers and Amazon API Gateway](#).

sous-réseau

Plage d'adresses IP dans votre VPC. Un sous-réseau doit se trouver dans une seule zone de disponibilité.

contrôle de supervision et acquisition de données (SCADA)

Dans le secteur manufacturier, un système qui utilise du matériel et des logiciels pour surveiller les actifs physiques et les opérations de production.

chiffrement symétrique

Algorithme de chiffrement qui utilise la même clé pour chiffrer et déchiffrer les données.

tests synthétiques

Tester un système de manière à simuler les interactions des utilisateurs afin de détecter les problèmes potentiels ou de surveiller les performances. Vous pouvez utiliser [Amazon CloudWatch Synthetics](#) pour créer ces tests.

invite du système

Technique permettant de fournir un contexte, des instructions ou des directives à un [LLM](#) afin d'orienter son comportement. Les instructions du système aident à définir le contexte et à établir des règles pour les interactions avec les utilisateurs.

T

balises

Des paires clé-valeur qui agissent comme des métadonnées pour organiser vos AWS ressources. Les balises peuvent vous aider à gérer, identifier, organiser, rechercher et filtrer des ressources. Pour plus d'informations, veuillez consulter la rubrique [Balisage de vos AWS ressources](#).

variable cible

La valeur que vous essayez de prédire dans le cadre du ML supervisé. Elle est également qualifiée de variable de résultat. Par exemple, dans un environnement de fabrication, la variable cible peut être un défaut du produit.

liste de tâches

Outil utilisé pour suivre les progrès dans un runbook. Liste de tâches qui contient une vue d'ensemble du runbook et une liste des tâches générales à effectuer. Pour chaque tâche générale, elle inclut le temps estimé nécessaire, le propriétaire et l'avancement.

environnement de test

Voir [environnement](#).

entraînement

Pour fournir des données à partir desquelles votre modèle de ML peut apprendre. Les données d'entraînement doivent contenir la bonne réponse. L'algorithme d'apprentissage identifie des modèles dans les données d'entraînement, qui mettent en correspondance les attributs des données d'entrée avec la cible (la réponse que vous souhaitez prédire). Il fournit un modèle de ML

qui capture ces modèles. Vous pouvez alors utiliser le modèle de ML pour obtenir des prédictions sur de nouvelles données pour lesquelles vous ne connaissez pas la cible.

passerelle de transit

Un hub de transit réseau que vous pouvez utiliser pour interconnecter vos réseaux VPCs et ceux sur site. Pour plus d'informations, voir [Qu'est-ce qu'une passerelle de transit](#) dans la AWS Transit Gateway documentation.

flux de travail basé sur jonction

Approche selon laquelle les développeurs génèrent et testent des fonctionnalités localement dans une branche de fonctionnalités, puis fusionnent ces modifications dans la branche principale. La branche principale est ensuite intégrée aux environnements de développement, de préproduction et de production, de manière séquentielle.

accès sécurisé

Accorder des autorisations à un service que vous spécifiez pour effectuer des tâches au sein de votre organisation AWS Organizations et dans ses comptes en votre nom. Le service de confiance crée un rôle lié au service dans chaque compte, lorsque ce rôle est nécessaire, pour effectuer des tâches de gestion à votre place. Pour plus d'informations, consultez la section [Utilisation AWS Organizations avec d'autres AWS services](#) dans la AWS Organizations documentation.

réglage

Pour modifier certains aspects de votre processus d'entraînement afin d'améliorer la précision du modèle de ML. Par exemple, vous pouvez entraîner le modèle de ML en générant un ensemble d'étiquetage, en ajoutant des étiquettes, puis en répétant ces étapes plusieurs fois avec différents paramètres pour optimiser le modèle.

équipe de deux pizzas

Une petite DevOps équipe que vous pouvez nourrir avec deux pizzas. Une équipe de deux pizzas garantit les meilleures opportunités de collaboration possible dans le développement de logiciels.

U

incertitude

Un concept qui fait référence à des informations imprécises, incomplètes ou inconnues susceptibles de compromettre la fiabilité des modèles de ML prédictifs. Il existe deux types

d'incertitude : l'incertitude épistémique est causée par des données limitées et incomplètes, alors que l'incertitude aléatoire est causée par le bruit et le caractère aléatoire inhérents aux données. Pour plus d'informations, veuillez consulter le guide [Quantifying uncertainty in deep learning systems](#).

tâches indifférenciées

Également connu sous le nom de « levage de charges lourdes », ce travail est nécessaire pour créer et exploiter une application, mais qui n'apporte pas de valeur directe à l'utilisateur final ni d'avantage concurrentiel. Les exemples de tâches indifférenciées incluent l'approvisionnement, la maintenance et la planification des capacités.

environnements supérieurs

Voir [environnement](#).

V

mise à vide

Opération de maintenance de base de données qui implique un nettoyage après des mises à jour incrémentielles afin de récupérer de l'espace de stockage et d'améliorer les performances.

contrôle de version

Processus et outils permettant de suivre les modifications, telles que les modifications apportées au code source dans un référentiel.

Appairage de VPC

Une connexion entre deux VPCs qui vous permet d'acheminer le trafic en utilisant des adresses IP privées. Pour plus d'informations, veuillez consulter la rubrique [Qu'est-ce que l'appairage de VPC ?](#) dans la documentation Amazon VPC.

vulnérabilités

Défaut logiciel ou matériel qui compromet la sécurité du système.

W

cache actif

Cache tampon qui contient les données actuelles et pertinentes fréquemment consultées.

L'instance de base de données peut lire à partir du cache tampon, ce qui est plus rapide que la lecture à partir de la mémoire principale ou du disque.

données chaudes

Données rarement consultées. Lorsque vous interrogez ce type de données, des requêtes modérément lentes sont généralement acceptables.

fonction de fenêtre

Fonction SQL qui effectue un calcul sur un groupe de lignes liées d'une manière ou d'une autre à l'enregistrement en cours. Les fonctions de fenêtre sont utiles pour traiter des tâches, telles que le calcul d'une moyenne mobile ou l'accès à la valeur des lignes en fonction de la position relative de la ligne en cours.

charge de travail

Ensemble de ressources et de code qui fournit une valeur métier, par exemple une application destinée au client ou un processus de backend.

flux de travail

Groupes fonctionnels d'un projet de migration chargés d'un ensemble de tâches spécifique.

Chaque flux de travail est indépendant, mais prend en charge les autres flux de travail du projet. Par exemple, le flux de travail du portefeuille est chargé de prioriser les applications, de planifier les vagues et de collecter les métadonnées de migration. Le flux de travail du portefeuille fournit ces actifs au flux de travail de migration, qui migre ensuite les serveurs et les applications.

VER

Voir [écrire une fois, lire plusieurs](#).

WQF

Voir le [cadre AWS de qualification de la charge](#) de travail.

écrire une fois, lire plusieurs (WORM)

Modèle de stockage qui écrit les données une seule fois et empêche leur suppression ou leur modification. Les utilisateurs autorisés peuvent lire les données autant de fois que nécessaire,

mais ils ne peuvent pas les modifier. Cette infrastructure de stockage de données est considérée comme [immutable](#).

Z

exploit Zero-Day

Une attaque, généralement un logiciel malveillant, qui tire parti d'une [vulnérabilité de type « jour zéro »](#).

vulnérabilité « jour zéro »

Une faille ou une vulnérabilité non atténuée dans un système de production. Les acteurs malveillants peuvent utiliser ce type de vulnérabilité pour attaquer le système. Les développeurs prennent souvent conscience de la vulnérabilité à la suite de l'attaque.

invite Zero-Shot

Fournir à un [LLM](#) des instructions pour effectuer une tâche, mais aucun exemple (plans) pouvant aider à la guider. Le LLM doit utiliser ses connaissances pré-entraînées pour gérer la tâche. L'efficacité de l'invite zéro dépend de la complexité de la tâche et de la qualité de l'invite. Voir également les instructions [en quelques clics](#).

application zombie

Application dont l'utilisation moyenne du processeur et de la mémoire est inférieure à 5 %. Dans un projet de migration, il est courant de retirer ces applications.

Les traductions sont fournies par des outils de traduction automatique. En cas de conflit entre le contenu d'une traduction et celui de la version originale en anglais, la version anglaise prévaudra.