الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي والبحث العلمي
جامعة سعيدة د. مولاي الطاهر
كلية التكنولوجيا
قسم: الإعلام الآلي

Mémoire de Master

Spécialité : Sécurité Informatique et Cryptographie

Thème

La Sécurité de données dans un environnement cloud public Amazon Web Services AWS

Présenté par :

Nadjia KHATIR

Fatima HAMRI

Dirigé par :

Soumia DIB

Année universitaire 2022-2023

Abstract :

In the contemporary world, the Cloud assumes a significant role in the digital transformation and the evolution of IT infrastructures. It holds immense relevance for both service companies and industries. The primary objective of these cloud-based infrastructures is to increase the provision of On-Demand Services and offering rapidly response to users' needs and requirements.

The services provided by these infrastructures ensure a secure information exchange, and our project's objective is to design a cloud architecture that prioritizes security.

We began by examining the fundamental principles of the Cloud computing and the various mechanisms for protecting data. Following that, our focus shifted towards Amazon Web Services (AWS) platform, utilizing its services and compliance standards as the foundation for developing a secure architecture model. Our approach involved implementing DevOps tools to deploy this model effectively.

Keywords: Public Cloud, AWS, IAM, VPC, EC2, SG, ALB, WAF, TERRAFORM, DevOps, Cloud9, CI/CD, GitHub, Jenkins.

Résumé :

Dans notre monde contemporain, le cloud occupe une place importante dans la transformation numérique et l'évolution des infrastructures informatiques. Il représente un domaine extrêmement pertinent pour les entreprises de services et les industries. En effet, l'objectif de ces infrastructures basées sur le Cloud est d'améliorer l'offre de service à la demande pour répondre rapidement aux besoins et exigences des utilisateurs.

Ces infrastructures reposent sur des modules et services qui assurent l'échange d'informations et la communication d'une façon fiable est sécurisé. L'objectif de notre projet est de proposer une architecture sécurisée en cloud.

Dans un premier temps, nous avons examiné les principes fondamentaux du cloud computing et les divers mécanismes de protection des données dans ce domaine. Ensuite, notre attention s'est portée sur la plateforme AWS (Amazon Web Services), en nous appuyant sur ses services et les normes de conformité, dans le but d'établir un modèle d'architecture sécurisée AWS et le déployer via les outils de DevOps.

Mots clés: Cloud Public, AWS, IAM, VPC, EC2, SG, ALB, WAF, TERRAFORM, DevOps, Cloud9, CI/CD, GitHub, Jenkins.

الملخص :

يحتل الكلاود في عالمنا المعاصر، مكانة هامة لاسيما في ظل التحول الرقمي وتطور البنية التحتية لتكنولوجيا المعلومات. فقد اصبحت تمثل مجالا ذا صلة كبيرة بشركات الخدمات والصناعات. ويهدف هذا النموذج القائم على الكلاود الى تحسين عرض الخدمات عند الطلب للرد بسرعة وفعالية على احتياجات ومتطلبات المستخدمين .

يعتمد هذا الهيكل على وحدات وخدمات تضمن تبادل المعلومات والاتصال بطريقة موثوقة و آمنة. أما بخصوص مشروعنا فإنه يهدف الى تقديم نموذج آمن في الكلاود.

حيث بدأنا بدراسة المبادئ الاساسية لحوسبة الكلاود والآليات المختلفة لحماية البيانات في هذا المجال. ثم تطرقنا الى منصة خدمات الويب أمازون مستفيدين من خدماتها ومعايير المطابقة المرجعية في الأمن السبراني من أجل إعداد نموذج آمن في مجال الكلاود باستعمال الوسائل المتاحة.

Mots clés: Cloud Public, AWS, IAM, VPC, EC2, SG, ALB, WAF, TERRAFORM, DevOps, Cloud9, CI/CD, GitHub, Jenkins.

Remerciement :

La réalisation de ce mémoire a été possible grâce au concours de plusieurs personnes à qui nous voudrions témoigner toute notre gratitude pour leur participation et leur soutien tout au long de ce parcours de cinq ans d'étude.

Nous voudrions tout d'abord adresser tous nos remerciements à notre directrice de ce mémoire Madame Soumia DIB. Nous souhaitons remercier également tous les professeurs du département informatique de l'université Tahar MOULAY qui ont contribué à alimenter nos réflexions et nous avoir aidé pour mieux comprendre notre domaine durant notre parcours de licence et Master.

Nos remerciements iront également vers tous ceux qui ont accepté avec bienveillance de participer au jury de ce mémoire. Nous tenons à remercier en particulier toutes les personnes qui nous avaient aidé de près ou de loin, qui nous ont fait découvrir le sujet et aidé à faire avancer notre projet. Nous remercions spécialement nos chers parents et familles et amis pour leurs soutiens.

Table des matières

Chapitre 1: Introduction au Cloud Computing

Chapitre 2: La sécurité dans le Cloud Computing

Chapitre 3: Cloud public et AWS comme solution cible

Chapitre 4: Implémentation

Liste des figures

Introduction générale

De nos jours, les solutions cloud prennent de plus en plus d'importance dans l'activité quotidienne des entreprises et des investisseurs de services en raison de leur pertinence dans les contextes opérationnels tels que la mobilité des employés, le télétravail et l'accessibilité aux ressources informatiques de manière flexible, à tout moment et depuis n'importe quel type d'appareil.

La compétition entre les Startup, les SSII (Société de Services en Ingénierie Informatique) et les ESN (Entreprise de Services du Numérique) pour aller plus vite et offrir des services et des applications informatiques à moindre coût avec une capacité de traitement de l'information à la demande souligne l'importance croissante des enjeux de sécurité des données dans les infrastructures virtualisées.

Notre travail vise principalement à mener une étude approfondie sur les aspects et les mécanismes de sécurité des données et d'infrastructure dans les environnements de cloud public. En utilisant les outils fournis par les fournisseurs de technologies cloud et leurs partenaires tiers, tout en respectant les normes de sécurité cloud et en adoptant les meilleures pratiques.

Notre projet est structuré de la manière suivante : dans la première section, nous introduirons le Cloud Computing, en abordant ses caractéristiques ainsi que ses différents types de déploiement et de services. La deuxième partie se concentrera sur la sécurité dans le cloud. La troisième section sera dédiée à la solution AWS (Amazon Web Services) de cloud public, où nous expliquerons nos motivations de choix et mettrons en évidence ses principaux services contribuant à la sécurité. Enfin, nous finirons par une phase d'implémentation de notre proposition d'architecture sécurisée d'une infrastructure AWS, en adoptant une approche DevOps.

Chapitre 1

Introduction au Cloud Computing

1.1 Introduction:

Le Cloud Computing (informatique en nuage) est une technologie qui a révolutionné la manière dont les entreprises et les individus accèdent et gèrent leurs données et leurs applications.

Ce chapitre vise à fournir une compréhension du Cloud Computing et comment il fonctionne ainsi que ces caractéristiques permettant de répondre aux besoins du client ou de l'entreprise afin d'améliorer l'efficacité des processus métier. Nous allons présenter les modèles de déploiement et de service du Cloud Computing et l'avantages de son utilisation. Nous allons également citer les défis et les obstacles liés à l'adoption du Cloud Computing.

1.2 Historique:

Certains pensent que le cloud computing est tout nouveau, c'est certainement vrai à cause de son niveau d'excitation qu'il a atteint est vraiment nouveau et que l'exposition qu'il a eu est vraiment toute nouvelle. Mais l'idée de Cloud Computing, qui consiste à partager des ressources qui se trouvent ailleurs, n'est pas si nouvelle que cela.

Comment a-t-on appelé le cloud computing dans le passé?

On parlait alors de l'informatique centralisée, d'informatique en grille, d'informatique distribuée, d'informatique à la demande, d'hébergement, de fournisseurs de services d'application ASP (application service providers). C'était très important dans les années 90. Toutes ces choses sont essentiellement de cloud computing sous une forme ou une autre [1].

Le début de la chronologie de Cloud Computing, c'était l'informatique centrale avec les gros ordinateurs centraux d'IBM dans les années 1950 et 1960. Il s'agissait d'ordinateurs centralisés, d'ordinateurs massifs avec des terminaux muets, des terminaux muets à écran vert qui leur étaient attachés. Ils pouvaient être locaux ou connectés à une ligne terrestre dans un endroit éloigné, mais ils étaient tous connectés au même ordinateur centralisé et cet ordinateur central centralisé est un pool partagé de ressources qui, dans de nombreux cas, se trouve quelque part hors site.

Après, entre 1960-1980, l'internet a connu de grands progrès. Il y a d'abord eu ce qu'on a appelé ARPANET, un projet du ministère de la défense visant à relier les collèges, les universités et les établissements militaires entre eux, puis ce protocole appelé TCP/IP, qui est devenu la norme pour l'internet et qui a rendu la communication beaucoup plus facile. Et enfin, dans les années 80, CompuServe a connecté son réseau de 500 000 utilisateurs directement à l'internet, ce qui a constitué le plus grand saut pour la base d'utilisateurs de l'internet.

Puis en 1970, la première machine virtuelle a été exécutée, pas dans le cloud mais sur des ordinateurs centraux, découpant essentiellement les ressources de ces ordinateurs centraux pour faire d'autres choses, mais c'était en quelque sorte la première forme de virtualisation et le cloud computing a besoin de la virtualisation. Il a également besoin d'Internet et de pools de ressources partagées. Tous ces éléments ont donc constitué des avancées importantes dans ce qui est devenu plus tard le cloud computing.

Un autre progrès important a eu lieu dans les années 1990, lorsque les réseaux privés virtuels VPNs sont devenus disponibles par l'intermédiaire des fournisseurs de télécommunications; un réseau privé virtuel ou un VPN est donc un tunnel ou une connexion sécurisée à travers l'internet (réseau public). Virtualisant essentiellement un réseau et le rendant sécurisé de sorte qu'un ordinateur puisse communiquer avec un autre ordinateur sans que d'autres ordinateurs puissent voir le trafic. Les VPNs sont fortement utilisés dans le cloud computing encore aujourd'hui.

Puis, dans les années 1990, de nombreux progrès ont été réalisés dans le domaine de cloud computing et le terme a commencé à prendre de l'ampleur. Il a commencé à être utilisé dans les discours d'éminents technologues.

C'est en 1999 que le premier logiciel en tant que service s'est véritablement fait connaître et est devenu extrêmement populaire avec le lancement de SalesForce.com, une application de gestion de la relation client CRM (customer relationship management). Traditionnellement, les CRM n'étaient disponibles que sous forme d'applications locales. SalesForce.com a simplifié l'accès aux CRM. Sur Internet, vous pouvez vous abonner à leur CRM et vous pouvez simplement payer en fonction du nombre d'utilisateurs, que votre entreprise utilisait. C'est ainsi qu'est né le premier modèle de réussite de la révolution SaaS (Software-as-a-Service), qui a véritablement pris son essor.

C'est à partir de là qu'Amazon.com a été lancé en 2006. Ils ont lancé EC2, qui signifie Elastic Compute Cloud (cloud de calcul élastique), une infrastructure-as-a-service, en 2006 sous forme de version bêta. EC2 est devenu le premier modèle de réussite pour toutes les autres solutions d'infrastructure-as-a-service.

Ensuite, Google Docs a été lancé en 2006 (Microsoft Office dans le cloud), il fournit des outils de présentation entièrement dans le cloud et les données sont également stockées dans le cloud.

En 2008, Microsoft Azure a été lancé. C'est avec Azure IaaS que Microsoft s'est vraiment lancé dans le cloud computing.

En 2013, Google Compute Engine ou GCE a été lancé. Il s'agit de la solution d'infrastructure en tant que service de Google. En 2016 et 2017, le chiffre d'affaires de AWS EC2 s'élève à 12,22 milliards de dollars et celui de SalesForce.com est estimé à plus de 8 milliards de dollars en 2017. Depuis le lancement de SalesForce.com en 1999 jusqu'à aujourd'hui, le cloud computing existe depuis longtemps et a pris un essor considérable. Evidemment, les end-users et les entreprises adoptent le cloud computing.

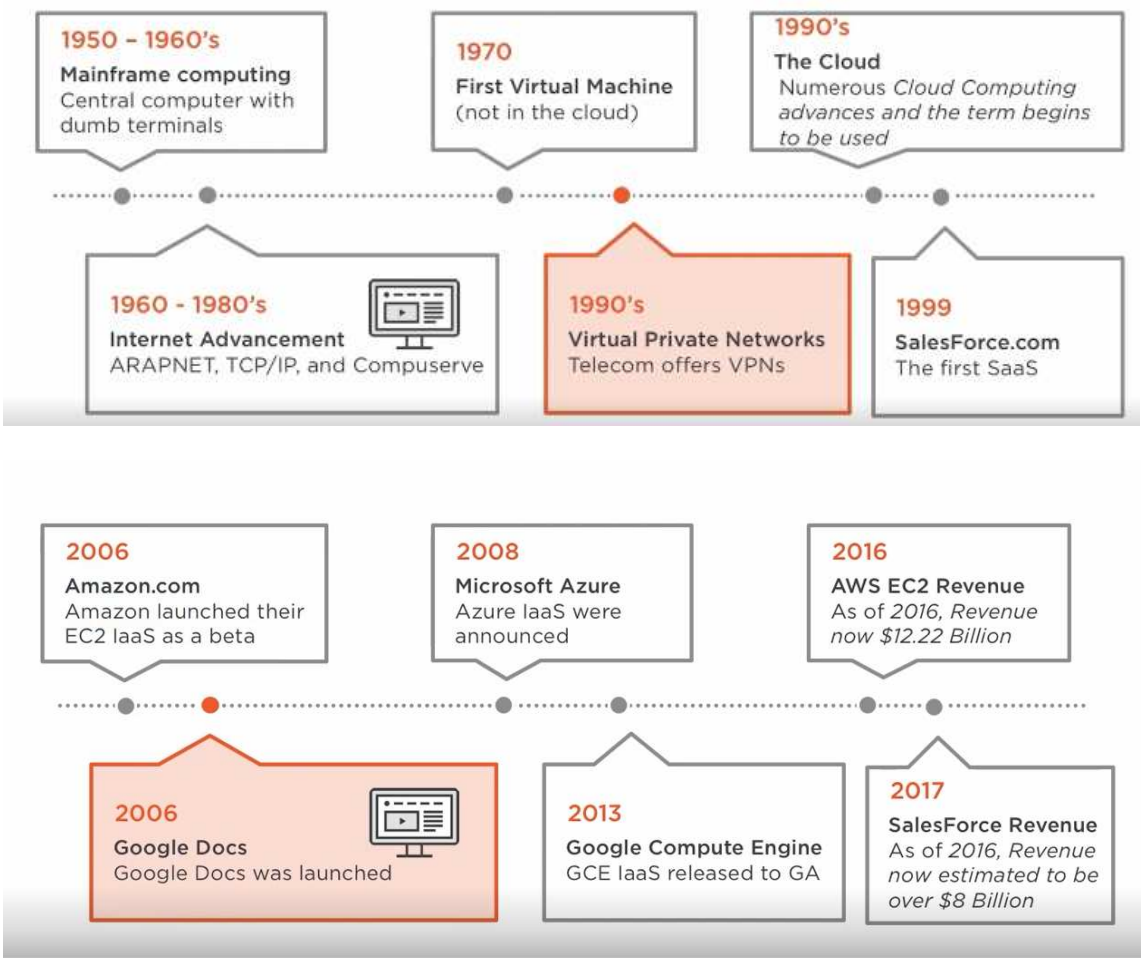


Figure 1: Historique du Cloud Computing (Source: Ryan Kroonenberg, Cloud Guru).

1.3 Définitions:

Il est difficile d'attribuer une définition précise au Cloud, du fait de son évolution progressive avec l'avènement de plusieurs technologies et de concepts ainsi que de son utilisation très vaste (diversité et la richesse).

La notion de Cloud fait référence à un nuage, tel que l'on a l'habitude de l'utiliser dans les schémas techniques pour représenter Internet. C'est une métaphore qui s'agit d'une abstraction de l'internet et de l'infrastructure qui le sous-jacente.

Le terme “Computing", désigne les services informatiques fournis par un ordinateur suffisamment puissant pour offrir une gamme de fonctionnalités, de ressources et de stockage.

Le “Cloud computing”, est considéré comme la fourniture de services informatiques mesurables à la demande sur Internet.

La combinaison des deux termes peut être comprise comme l'utilisation d'un ordinateur suffisamment puissant pour fournir des services via internet.

Le cloud computing a été défini comme:

“Ressources informatiques à la demande, fournies par l'intermédiaire de l'Internet.”

Selon le géant de la virtualisation (VMware) :

Le cloud computing est une approche de l'informatique qui s'appuie sur la mise en commun efficace d'une infrastructure virtuelle autogérée et disponible à la demande.”

Selon l’Organisation internationale de normalisation (ISO/IEC) :

“Le cloud computing est un paradigme permettant l'accès en réseau à un pool partagé et élastique de ressources physiques ou virtuelles partageables, avec un auto-approvisionnement et une administration à la demande.”

Selon, le National Institute of Standards and Technology (NIST), une agence du gouvernement américain qui a donné une brève définition qui reprend les principes de base du Cloud. Il l’a défini comme suit:

"Le cloud computing est un modèle qui permet un accès pratique et à la demande à un réseau partagé de ressources informatiques configurables (par exemple, des réseaux, des serveurs, des stockages, des applications et des services) qui peuvent être rapidement provisionnées et libérées avec un minimum d’effort de gestion ou d'interaction avec le fournisseur de services."

Cette définition souligne les caractéristiques clés du cloud computing, notamment la disponibilité à la demande des ressources informatiques, la possibilité de les provisionner rapidement et de les libérer facilement, ainsi que la configuration flexible des ressources en fonction des besoins de l'utilisateur.

Le NIST a défini un modèle de Cloud Computing comportant cinq caractéristiques essentielles, trois modèles de services et quatre modèles de déploiement [2], comme présenté dans la figure 2 suivante.

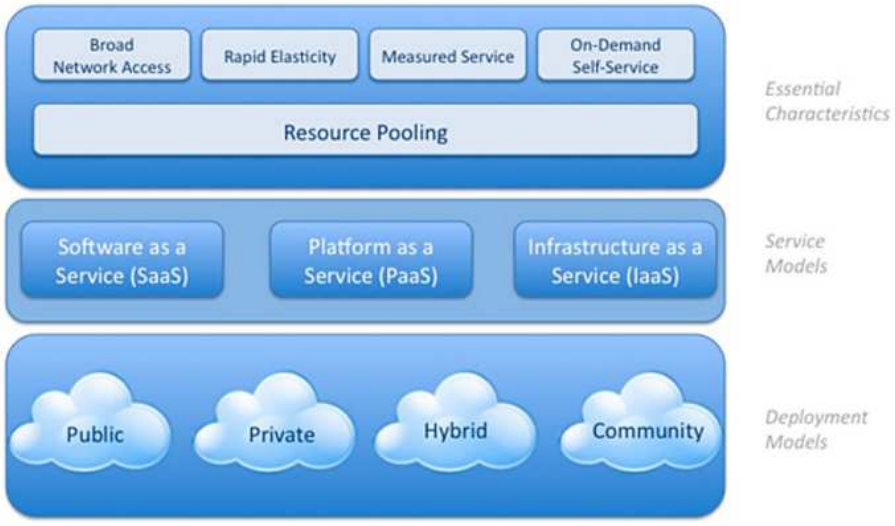


Figure 2: Modèle visuel du NIST pour le Cloud Computing (source : www.researchgate.net).

1.4 Comment fonctionne le cloud computing?

Les entreprises avaient pris pour habitude d'utiliser les ordinateurs disponibles dans leurs locaux pour héberger un grand nombre d'éléments, comme les données. Le cloud computing permet de faire de même, mais à distance, c'est-à-dire sans avoir à investir dans un serveur ou des applications, etc. La seule exigence est de disposer d'ordinateurs et d'une bonne connexion internet.

En ayant recours au cloud computing, les entreprises utilisent un serveur informatique qu'elles louent à un prestataire. L'avantage est qu'elles ont accès à une multitude de services sans avoir à prendre en charge des actions qui demandent des compétences avancées en informatique (sauvegarde des données, mises à jour des logiciels, entretien de la base de données… etc). Ces actions sont assurées par le prestataire.

L'accès au service se fait par une application disponible sur les ordinateurs de l'entreprise. De leur côté, les données et applications ne se trouvent pas sur les ordinateurs locaux, mais bel et bien sur le cloud. Le cloud, aussi appelé nuage en français, est un ensemble de serveurs distants connectés entre eux par le biais de liaisons internet ultra performantes.

Aujourd'hui, les entreprises pour qui la notion de cloud computing semble complexe et qui ne disposent pas d'un service informatique avec des professionnels dédiés, peuvent migrer vers le cloud en sollicitant les compétences de professionnels du secteur comme Premaccess.

1.5 Caractéristiques:

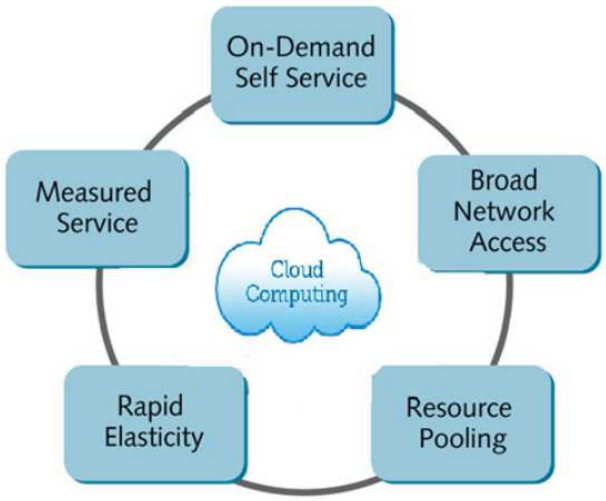


Figure 3: Caractéristiques du Cloud Computing (source : www.semanticscholar.org).

Comme présenté dans la figure 3, le cloud computing possède cinq caractéristiques essentielles:

Libre-service à la demande (On-demand Self-services): Les utilisateurs peuvent accéder aux ressources informatiques (serveur, réseau, capacité de stockage, environnement d'exécution, application et performances de calcul…ETC) via une interface en libre-service d'une manière automatisée, et c'est l'utilisateur qui met en place et gère la configuration à distance sans nécessité d'interaction humaine avec chaque fournisseur de services et au besoin. Ce qui leur permet de provisionner et de libérer rapidement les ressources selon leurs besoins.

Large accès réseau (Broad Network Access): est une connexion basée sur le réseau entre le consommateur de services en nuage et le fournisseur de services cloud**,** les ressources sont disponibles depuis Internet et accessibles via des mécanismes standards, par des équipements traditionnels et hétérogènes, légers ou lourds (ordinateurs portables, tablettes, téléphones…etc).

Mise en commun de ressources (Resource Pooling): Les ressources informatiques sont partagées entre de multiples utilisateurs et clients selon un modèle de corésidence, avec des ressources dynamiquement allouées selon les demandes. Généralement, le client ne connaît pas la localisation exacte des ressources allouées, bien qu'il puisse choisir l'emplacement géographique et spécifier cette localisation (comme le pays, le continent ou le centre de données)

Élasticité rapide (Rapid Elasticity) : les ressources peuvent être provisionnées et libérées de manière élastique, croissante ou décroissante avec une adaptation rapide en fonction de la demande. Pour le client, les ressources sont illimitées et disponibles à tout moment.

Service mesuré (Measured Service): Les ressources allouées sont mesurées et contrôlées automatiquement et facturées à l'utilisateur (facturation et suivi) en fonction de la quantité de ressources utilisées, et par des moyens de mesure à certains niveaux d'abstraction appropriés au type de service (stockage, temps de calcul, bande passante…etc). Afin d'offrir de la transparence pour le fournisseur et le consommateur de service.

1.6 Modèle de déploiements:

Selon le National Institute of Standards and Technology (NIST), il existe quatre modèles de déploiement du cloud computing :

- **Cloud public (Public Cloud):** ce service vendu sur demande est de son côté fourni par un prestataire qui le gère et le maintient et accessible par tous via Internet. Les clients paient uniquement ce qu'ils utilisent. Les ressources informatiques, telles que les serveurs, les applications et les espaces de stockage, sont partagées entre plusieurs clients. Les fournisseurs de cloud public les plus connus sont Amazon Web Services, Microsoft Azure, IBM et Google Compute Engine.

- **Cloud privé (Private Cloud):** il s'agit d'une infrastructure réservée à l'usage exclusif d'une seule et même entreprise ou organisation (non multi-tenant). Elle peut être hébergée en interne ou en externe, dans les locaux de l'entreprise ou chez un fournisseur de services cloud tiers et gérée au sein de l'entreprise. Le cloud privé offre un niveau plus élevé de sécurité et de contrôle que le cloud public.

- **Cloud communautaire (Community Cloud):** il s'agit d'une infrastructure mise à disposition pour une utilisation exclusive par une communauté spécifique de consommateurs ayant des préoccupations communes en termes de sécurité, de conformité et de réglementation. Elle peut être gérée par un membre de la

communauté ou par un tiers. Ces organisations peuvent être du même secteur ou de secteurs différents.

- **Cloud hybride (Hybrid Cloud):** il s'agit d'un croisement entre le cloud privé et le cloud public. Il permet à une entreprise d'utiliser le cloud public pour les tâches non critiques et le cloud privé pour les données sensibles ou confidentielles. Ce type de cloud fournit une combinaison flexible de services de cloud computing en réduisant les coûts et en offrant un niveau de sécurité élevé pour les données sensibles.

Il est important de noter que ces modèles peuvent varier en fonction des fournisseurs de services cloud, des technologies utilisées et des besoins des entreprises.

1.7 Modèle de services:

Selon le National Institute of Standards and Technology (NIST), il existe trois modèles de service cloud [3]:

Infrastructure en tant que Service (Infrastructure-as-a-Service): Le CSP (Cloud Service Provider) fournit une infrastructure virtualisée, qui peut être adaptée aux besoins des clients. Les services IaaS offrent des ressources informatiques, telles que des serveurs, du stockage et des réseaux, à la demande via Internet. Les clients ont le contrôle total sur les ressources qui peuvent les utiliser pour exécuter leurs propres applications et systèmes d'exploitation, et gérer les configurations de leur infrastructure selon leurs besoins.

Le fournisseur de services en nuage gère les installations, le matériel, l'interface avec le matériel par le biais d'un certain type d'abstraction et il s'occupe de la connectivité du réseau. Enfin, il y a les APIs, qui deviennent l'interface entre le consommateur et les zones du fournisseur de services cloud comme le montre la figure 4.

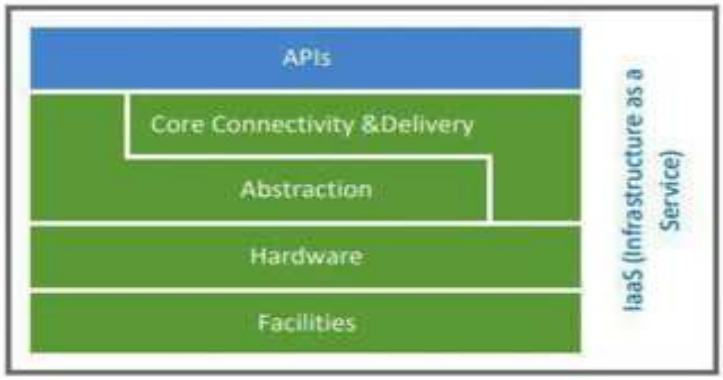


Figure 4: Le modèle Infrastructure as a Service (source : Kevin Henry, Cloud Data Security for CCSP)

Dans ce cas, comme le montre la figure 5, le fournisseur de services cloud a le contrôle total sur le matériel. Il a le contrôle administratif de l'hyperviseur, mais nous pouvons voir que le consommateur de services cloud doit également être en mesure de s'interfacer avec lui. Le contrôle total de l'intergiciel d'application et du système d'exploitation invité est maintenant passé du fournisseur de services cloud au consommateur de services cloud.

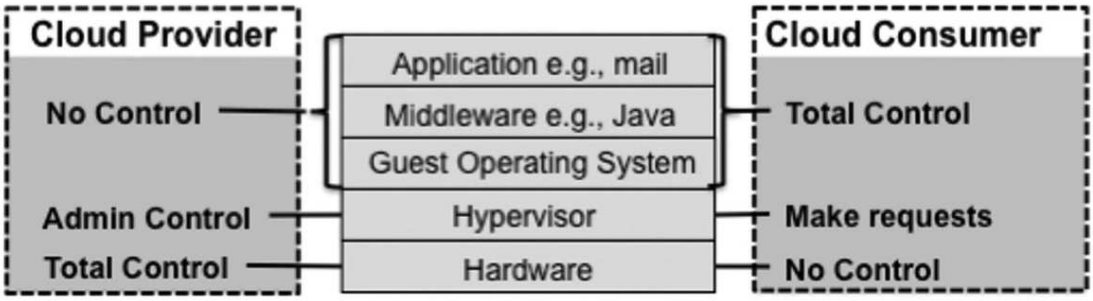


Figure 5: Le contrôle de IaaS (source : Kevin Henry, Cloud Data Security for CCSP).

Plateforme en tant que Service (Platform-as-a-Service): Ce modèle fournit une plateforme de développement pour les applications. Les FSCs mettent à disposition aux clients des plateformes et des outils de développement et d'exécution d'applications via Internet, y compris des bibliothèques, des frameworks, des langages de programmation, des outils de gestion de base de données etc, permettant au client de développer, déployer et gérer des applications sur ces plateformes sans avoir à se soucier de l'infrastructure sous-jacente (qui est gérée par le FSC).

La figure 6 montre que la responsabilité du FSC est l'élément de service d'infrastructure, mais en plus de cela, l'intégration de l'intergiciel nécessaire pour que l'application que le consommateur a achetée fonctionne.

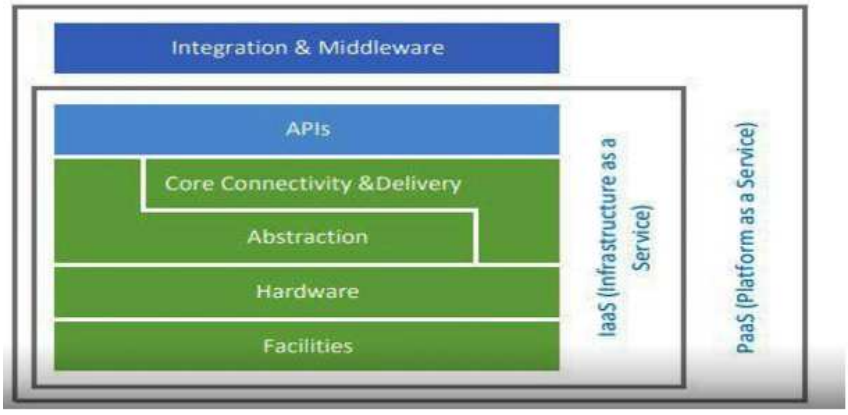


Figure 6: Le modèle Plateforme as a Service (source : Kevin Henry, Cloud Data Security for CCSP).

Dans ce cas, comme la figure 7 montre que le FSC n'a aucun contrôle sur l'application. Celle-ci est gérée par le consommateur, qui se charge des correctifs. Il administre toujours le contrôle de son intergiciel et le contrôle total de l'installation physique. Le consommateur de services cloud exerce un contrôle réel sur l'application (installation, correction, contrôle d'accès, etc.). Il doit également gérer les interfaces entre son programme et l'infrastructure sous-jacente.

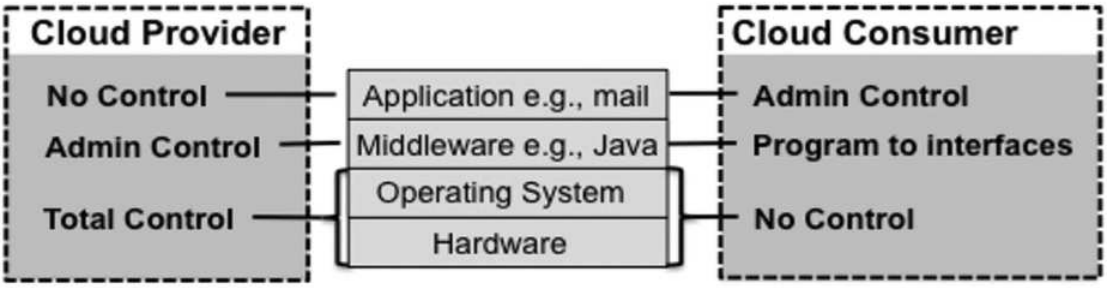


Figure 7: Le contrôle de PaaS (source : Kevin Henry, Cloud Data Security for CCSP).

Logiciel en tant que service (Software-as-a-Service): Le consommateur utilise des applications et des logiciels directement sur le cloud accessibles via Internet, qui sont fournies et exécutées par le FSC qui en gère. Les clients utilisent les applications sur le cloud via un navigateur ou une API afin de les utiliser sans avoir besoin de les installer sur leur propre ordinateur ou serveur. Ces logiciels comprennent les services de messagerie électronique, les applications de gestion de la relation client CRM et de gestion de ressources humaines GRM, les applications de gestion de projet, les outils de collaboration, etc.

Les domaines qui relèvent de la responsabilité du FSC dans la figure 8, sont la manière dont ils interfacent avec l'utilisateur, et la plateforme utilisée pour cela, les APIs pour accéder à des choses telles que leur base de données sous l'application et ils gèrent les données, les métadonnées et le contenu qui est disponible. On peut voir qu'il fonctionne sur une plateforme en tant que service qui, en fait, fonctionne également sur une infrastructure en tant que service.

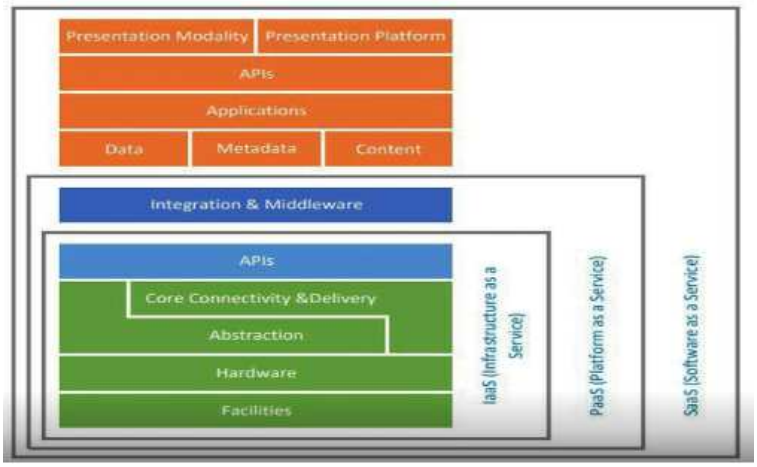


Figure 8: Le modèle Software a Service (source : Kevin Henry, Cloud Data Security for CCSP).

La répartition des responsabilités dans le logiciel en tant que service est telle que nous pouvons la voir dans la figure 9. Le FSC gère l'application, la corrige, etc. Il a également le contrôle total et la propriété de l'intergiciel, du système d'exploitation, du matériel et même de l'installation. Le consommateur n'a aucun contrôle sur l'endroit où ses données sont stockées ou traitées et n'a qu'un contrôle administratif très limité comme, par exemple, l'accès des utilisateurs.

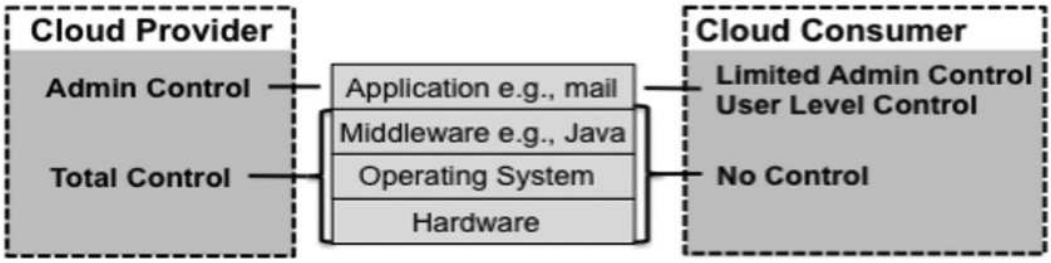


Figure 9: Le contrôle de SaaS (source : Kevin Henry, Cloud Data Security for CCSP).

En résumé, ces trois modèles de service cloud offrent différents niveaux d'abstraction pour les clients qui souhaitent utiliser des services cloud en fonction de leurs besoins spécifiques en matière d'infrastructure, de plate-forme ou d'applications.

1.8 Orchestration des services:

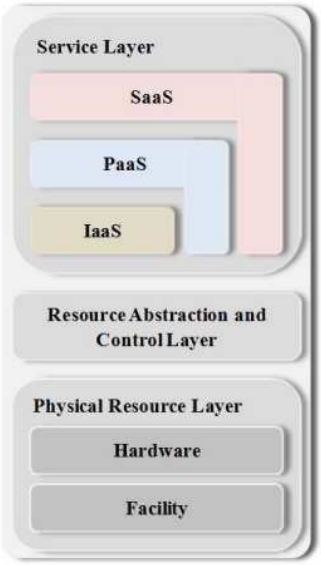


Figure 10: Orchestration des services (Cloud provider).
(source : Kevin Henry, Cloud Data Security for CCSP)

L'avantage de la mise en œuvre du cloud computing réside en grande partie dans l'arrangement, la coordination et la gestion de tous les composants du système utilisés par le fournisseur du cloud pour fournir des services au consommateur.
Un modèle à trois niveaux est utilisé dans cette représentation, représentant le regroupement de trois types de composants de système que les fournisseurs de services cloud doivent composer pour fournir leurs services.

Dans le modèle illustré à la figure 10, la couche supérieure est la couche de service **(Service Layer)**, où les FSCs définissent les interfaces permettant aux consommateurs d'accéder aux services cloud. Les interfaces permettant aux consommateurs d'accéder aux services cloud.

Les interfaces d'accès de chacun des trois modèles de service sont fournis dans cette couche. Il est possible, mais pas nécessaire, que des applications SaaS soient construites au-dessus des composants PaaS et que les composants PaaS soient construits au-dessus des composants IaaS. Les relations de dépendance optionnelle entre les composants SaaS, PaaS et IaaS sont représentées graphiquement par des composants empilés les uns sur les autres. Tandis que l'angle des composants indique que chacun des composants de service peut être autonome. Par exemple, une application SaaS peut être mise en œuvre et hébergée sur des machines virtuelles d'un IaaS, ou être mise en œuvre directement sur les ressources du cloud sans utiliser de machines virtuelles IaaS.

La couche intermédiaire du modèle est la couche d'abstraction et de contrôle des ressources **(Resource Abstraction and Control Layer)**. Cette couche contient les composants système que les fournisseurs de cloud utilisent pour fournir et gérer l'accès aux ressources physiques par le biais d'une abstraction logicielle. Parmi les exemples de composants d'abstraction des ressources, on peut citer les éléments logiciels tels que les hyperviseurs, les machines virtuelles, le stockage de données virtuelles et d'autres abstractions de ressources informatiques. L'abstraction de ressources doit garantir une utilisation efficace, sûre et fiable des ressources physiques sous-jacentes. Bien que la technologie des machines virtuelles soit couramment utilisée à ce niveau, d'autres moyens de fournir les abstractions logicielles nécessaires sont également possibles.

L'aspect contrôle de cette couche se réfère aux composants logiciels responsables de l'allocation des ressources, du contrôle d'accès et de la surveillance. Il s'agit d'un logiciel qui relie les nombreuses ressources physiques et leurs abstractions logicielles pour permettre la mise en commun des ressources, l'allocation dynamique et le service mesuré. Différents logiciels du cloud, logiciels libres et propriétaires sont des exemples de ce type d'intergiciel.

La couche la plus basse de la pile est la couche des ressources physiques **(Physical Resource Layer)**, qui comprend les ressources matérielles, telles que les ordinateurs (CPU et mémoire), les réseaux (routeurs, pare-feu, commutateurs, liens et interfaces réseau), les composants de stockage (disques durs) et d'autres éléments physiques de l'infrastructure informatique. Il comprend également les ressources des installations, telles que le chauffage, la ventilation et la climatisation (HVAC), l'alimentation électrique, les communications et d'autres aspects de l'installation physique.

Conformément aux conventions d'architecture des systèmes, le positionnement horizontal, c'est-à-dire la stratification, dans un modèle représente les relations de dépendance, les composants de la couche supérieure dépendent des composants adjacents de la couche inférieure pour fonctionner. La couche d'abstraction et de contrôle des ressources expose les ressources virtuelles du cloud au-dessus de la couche des ressources physiques et prend en charge la couche des services où les interfaces des services cloud sont exposées aux consommateurs du cloud, alors que ces derniers n'ont pas d'accès direct aux ressources physiques.

1.9 Avantages:

Dans un monde hyperconnecté et ou l'informatique est un service d'un nombre d'entreprises toujours plus important, l'objectif du cloud computing est clair est de simplifier le travail de ceux qui sont amenés à travailler sur des ordinateurs au quotidien au sein de leur société. L'utilisation du cloud computing porte de plusieurs avantages:

- Avoir recours au cloud permet aux petites entreprises de lancer un service pour très peu d'investissement en software et aucun en hardware. Une solution qui reste moins coûteuse que faire évoluer le système sans investir.

- Le cloud computing permet de faire des économies d'échelle, en n'ayant pas besoin d'investir dans les infrastructures et dans l'entretien du système de sauvegarde des données, des serveurs et applications, etc.

- Il met à disposition des entreprises, des services parfois très coûteux, le tout à moindre prix en bénéficiant de l'évolution de ces mêmes services.

- Avec la mutualisation des ressources, le cloud offre des capacités illimitées en ce qui concerne la bande passante et le stockage.

- Le cloud computing offre une sécurité optimisée en ce qui concerne le stockage des données avec le chiffrement des données, la surveillance logicielle et la sécurisation des lieux de stockage (Centres de données).

- La responsabilité de gestion de systèmes et la maintenance des serveurs est prise en charge par le fournisseur du service cloud, ce qui réduit les tâches de l'utilisateur.

- Les mises à jour logicielles sont automatiques, et les utilisateurs n'ont pas besoin d'acheter un logiciel ou une licence.

- Il assure la qualité de service QoS; l'équilibrage de la charge, la sauvegarde et la récupération des données.

- Informatique verte (Green Computing); les infrastructures gérées en interne sont souvent sous-utilisées, alors que l'infrastructure d'un Cloud mutualise l'ensemble de ressources pour un grand nombre de consommateurs. Elle permet alors de minimiser le nombre des équipements et d'augmenter le taux d'utilisation.

1.10 Risques:

Le Cloud Computing n'a pas que des avantages, mais il présente certains obstacles et risques, notamment:

- **Sécurité:** le stockage de données dans le cloud peut être vulnérable aux cyberattaques, aux violations de données et aux fuites d'informations sensibles. Il est important de mettre en place des mesures de sécurité appropriées pour protéger les données dans le cloud.

- **Disponibilité:** la disponibilité du cloud dépend de la qualité et de la fiabilité de la connexion Internet. Des pannes de réseau peuvent rendre le cloud inaccessible et perturber les activités de l'entreprise.
- **Confidentialité:** le cloud peut soulever des préoccupations en matière de confidentialité, car les données sont stockées sur des serveurs appartenant à des tiers. Il est important de s'assurer que les fournisseurs de services cloud respectent les normes de confidentialité et de sécurité.
- **Interopérabilité:** le passage d'un fournisseur de cloud à un autre peut être difficile en raison de la difficulté à transférer les données entre différents fournisseurs. Il est donc important de bien choisir son fournisseur de cloud dès le début.
- **Conformité:** Les entreprises peuvent être tenues de respecter des réglementations strictes en matière de protection des données, de confidentialité, de conformité fiscale, etc. Il est donc important de choisir des fournisseurs de cloud qui répondent à ces exigences réglementaires.
- **Coûts:** les coûts liés au cloud peuvent être difficiles à estimer, car ils dépendent de nombreux facteurs, tels que la taille de l'entreprise, les besoins en matière de stockage et les options de service. Les coûts peuvent également augmenter avec l'utilisation accrue du cloud.
- **Dépendance:** les entreprises peuvent devenir dépendantes du cloud pour leurs activités critiques. Si le fournisseur de cloud rencontre des problèmes ou décide de fermer ses portes, cela peut entraîner des perturbations importantes pour l'entreprise. Il est donc important d'avoir des plans de continuité des activités en place.

Même si le cloud computing présente de nombreux avantages, il est important de prendre en compte les obstacles et les risques associés lors de la prise de décisions concernant la migration vers le cloud afin de les gérer de manière efficace.

1.11 Challenges de recherche dans l'environnement Cloud:

Bien que certaines des caractéristiques essentielles du Cloud Computing ont été réalisées par des efforts commerciaux et universitaires, de nombreux problèmes existants n'ont pas été pleinement pris en compte, et d'autres nouveaux défis continuent d'émerger [4].

Implémentation des politiques de sécurité: L'implémentation des politiques de sécurité est la tâche la plus difficile pour les fournisseurs, ils comprennent des politiques de gestion, des politiques réglementaires qui sont liées à la conformité aux normes, des politiques informatives qui éduquent les parties prenantes internes et externes de l'entreprise, etc. [5].

Gestion de Virtualisation: La virtualisation décrit une technologie dans laquelle les applications, les systèmes d'exploitation ou le stockage de données sont séparés du matériel.

Le logiciel émule le matériel à partir du système d'exploitation (Linux, Windows, etc.) ou un système d'exploitation dédié tel que VMware. La virtualisation est désormais un élément essentiel dans le Cloud, elle constitue la colonne vertébrale de l'infrastructure en tant que service (IaaS), ce qui soulève la question sur les risques de sécurité liés à la virtualisation. [5].

Sécurité de données et confidentialité: Dans le Cloud Computing, les données doivent être transférées entre les dispositifs de l'utilisateur et les Datacenter des fournisseurs de services de Cloud Computing, ce qui les rendra cible facile pour les pirates. La sécurité des données et la confidentialité doivent être garanties, que ce soit sur le réseau ou encore dans les Datacenter de Cloud où elles seront stockées [3].

1.12 Conclusion:

Dans ce chapitre nous avons présenté les notions fondamentales du Cloud Computing, ses évolutions, son fonctionnement et ses caractéristiques. Nous avons étudié également les quatres modèles de déploiement du cloud computing et les trois services principaux, sur lesquels il repose: applicatif, plateforme, infrastructure, et qui ont donné naissance aux fameux SaaS/PaaS/IaaS. Enfin nous avons présenté les différents avantages et inconvénient du Cloud Computing, et les challenges de recherche dans ce domaine.

Dans le chapitre suivant, nous présenterons la sécurité dans le Cloud.

Chapitre 2

La Sécurité dans le Cloud

2.1 Introduction:

Comme nous l'avons vu dans le premier chapitre, le cloud computing est une révolution et une approche qui permet de réduire les coûts et de simplifier la gestion des applications informatiques ainsi que la gestion d'infrastructure. Cependant, la sécurité est le challenge primordial dans le cloud. De ce fait, il est essentiel d'adopter des mesures de sécurité appropriées lors de son utilisation.

2.2 Conditions d'analyse dans le cloud:

Avant de procéder à une explication approfondie des diverses méthodes visant à assurer la protection des données lors de leur envoi vers un service cloud, il est essentiel de commencer par établir les conditions dans lesquelles un tel service est évalué, notamment son architecture, les attaquants potentiels et les propriétés de sécurité impliquées.

2.2.1 Architecture du cloud:

Dans cette présentation, nous considérons l'architecture cloud suivante [7]: d'un côté, nous avons un utilisateur nommé Bob, qui fait partie d'une entreprise donnée. De l'autre côté, nous avons un service cloud que Bob utilise pour stocker des données ou effectuer des traitements. Pour ce faire, Bob peut interagir directement avec le service cloud ou faire appel à une entité intermédiaire, telle qu'un tiers de confiance (Trusted Third Party en anglais, TTP) ou une passerelle de sécurité cloud (Cloud Security Gateway en anglais, CSG).

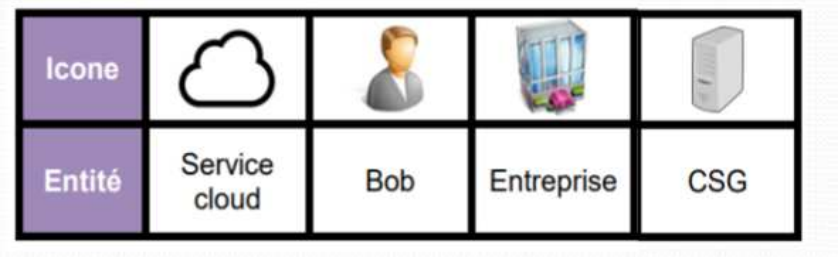


Figure 11: les différentes entités dans le Cloud (Source : TANIA MARTIN, Samls).

2.2.2 Attaquants:

Dans cette publication, nous abordons un attaquant nommé Oscar, qui possède les capacités suivantes:

- d'intercepter les canaux de communication entre les différentes entités de l'architecture.
- d'attaquer le service cloud et d'accéder ainsi à toutes les données stockées chez ce dernier.

2.2.3 Propriétés de sécurité:

Pendant de nombreuses années, la **CID** a joué un rôle essentiel dans la définition de la sécurité de l'information. Il est important d'appliquer ces concepts fondamentaux à toutes les informations et aux systèmes d'information, également dans le cloud.

- **Confidentialité:** Il s'agit de protéger les informations contre toute divulgation non autorisée, qu'il s'agisse de la propriété intellectuelle, des informations personnellement identifiables (PII) ou des données sensibles de l'entreprise. Cependant, assurer cette confidentialité devient beaucoup plus complexe dans le cloud en raison de la dépendance envers le fournisseur de services cloud.

- **Intégrité:** Il s'agit de protéger les informations contre toute modification non autorisée et de garantir leur traitement, leur stockage et leur utilisation corrects. Cependant, dans le contexte du cloud, de nouvelles couches sont introduites dans le modèle des systèmes, ce qui crée de nouvelles surfaces d'attaque ou des points potentiels de compromis.

- **Disponibilité:** L'objectif est de s'assurer que les systèmes et les données sont accessibles au moment voulu. La disponibilité est une mesure de criticité, car le cloud représente une dépendance pour l'utilisateur (consommateur) et la disponibilité est une préoccupation critique.

2.3 Les concepts de la sécurité des données dans le cloud:

2.3.1 Concepts de la Sécurité dans le cloud:

La sécurité des données cloud est cruciale car les données sont un actif précieux pour une organisation. Il est essentiel de mesurer la criticité des données en fonction de leur importance pour les opérations commerciales. Pour les protéger, il est nécessaire d'identifier leur emplacement, leurs utilisations et leurs formes.

Il est également important de prendre en compte toutes les formes de données, qu'elles soient électroniques, papier, verbales ou vidéo. La désignation d'un propriétaire des données est essentielle, conformément aux exigences légales. Ce propriétaire doit être un cadre supérieur avec le pouvoir budgétaire et la capacité d'accepter la responsabilité de la protection des données pour toute l'organisation.

Rôle du Data Owner:

- Assurer que les données bénéficient du niveau de protection approprié de manière cohérente dans toute l'organisation.

- Parfois appelé dans le cloud le contrôleur des données ou le processeur des données [8].

Il est important de veiller à ce que les données bénéficient d'un niveau de protection approprié, évitant à la fois une surprotection excessive et des vulnérabilités. La cohérence de la protection des données lors de leur circulation au sein de l'organisation est essentielle, même lorsqu'elles passent d'un service à un autre et d'un format à un autre. Dans le contexte du cloud, le contrôleur ou le processeur des données est responsable de la protection des données au sein de l'environnement du fournisseur de services cloud, mais le consommateur demeure ultimement responsable. Cette question soulève la notion de responsabilité.

Responsabilité:

Le propriétaire des données est responsable et tenu de rendre des comptes pour la protection des données, même si les données sont traitées par un tiers, même à plusieurs niveaux au sein de l'opération de traitement [8].

En établissant la responsabilité, le propriétaire des données est tenu de rendre des comptes et devient responsable de la protection des données. Ils doivent définir les mesures appropriées de protection des données et veiller à leur mise en œuvre conformément aux mandats et directives qui leur sont donnés.

Ce sont les deux termes de "prudence" et "diligence raisonnable" [8] :

- La **prudence** consiste à s'assurer que les politiques sont en place pour protéger les données.
- Et la **diligence raisonnable** consiste à faire un suivi et à s'assurer que ces politiques sont suivies.

Même lorsque nos données sont traitées par un tiers, comme un fournisseur de services cloud, il est essentiel d'exiger des niveaux de protection appropriés pour ces données, même sur les infrastructures cloud. Cette exigence devient encore plus complexe lorsque le fournisseur de cloud engage des sous-traitants supplémentaires. Ainsi, la nécessité de protéger les données peut s'étendre sur plusieurs niveaux au sein d'une organisation.

Autres rôles et responsabilités liés aux données (Custodian, processor, subject, user):

Les rôles et responsabilités liés aux données, tels que:

- Le gardien **(CUSTODIAN)**, sont liés à la protection des données. Le gardien est la personne qui a la garde d'une donnée à un moment donné. Par exemple, un utilisateur

au sein d'une organisation peut avoir la responsabilité temporaire d'un enregistrement client. Même s'il n'est pas propriétaire des données, il doit les manipuler conformément aux politiques et procédures établies. Cela s'applique également aux administrateurs système qui effectuent des sauvegardes de données dont ils ne sont ni les utilisateurs ni les propriétaires, mais qui restent néanmoins les gardiens de ces données et doivent s'assurer de leur protection adéquate.

- Dans le contexte du cloud, le processeur de données **(PROCESSOR)** est la personne chargée de traiter les données au nom d'un utilisateur du cloud, et il est également responsable de la protection appropriée de ces données.

- Il y a aussi la personne à qui appartiennent les données, que nous pouvons appeler le sujet **(SUBJECT)**. En tant que client d'une organisation, si je leur confie mes données personnelles, je reste le sujet et j'ai donc certaines exigences et obligations pour m'assurer que les données que je leur fournis sont utilisées de manière appropriée et en toute transparence. Dans certains cas, je peux même demander à être informé des données qu'ils détiennent à mon sujet.

- Utilisateur **(USER)**, il est responsable de l'utilisation des données et doit veiller à respecter les procédures de manipulation en fonction de la classification des données.

- Les **administrateurs** sont un groupe spécial qui détient souvent des privilèges élevés. Ils ont la capacité de faire et de voir tout, même s'ils ne sont pas les mieux rémunérés de l'organisation. Leur accès privilégié nécessite une communication claire sur les actions appropriées à prendre concernant les données, que ce soit pour le personnel interne ou celui travaillant pour un fournisseur de services cloud. Il est crucial de définir les limites d'accès pour éviter toute divulgation non autorisée. Cela concerne également le personnel d'exploitation des services, y compris ceux travaillant à distance, afin d'éviter tout abus d'accès.

Les administrateurs internes, tels que les administrateurs réseau, système et de bases de données, ont également des accès et des privilèges élevés et doivent connaître les actions appropriées à entreprendre concernant les données. Lorsqu'une personne se voit accorder un accès, elle assume également une responsabilité quant à son utilisation. Il est donc essentiel de collaborer avec les différents administrateurs pour s'assurer qu'ils comprennent les actions appropriées à prendre concernant les données. Dans la plupart des cas, ils agissent en tant que gardiens des données.

Responsabilités en matière de sécurité:

Un graphique important ici est cet exemple de responsabilités en matière de sécurité [9]. Nous constatons que lorsque nous traitons avec le cloud, nous avons une différenciation de qui est responsable de quoi.

Nous pouvons voir à travers la ligne du haut que la gouvernance des données et la gestion des droits sont toujours conservées par le client, indépendamment du modèle de déploiement du logiciel, de la plateforme, de l'infrastructure ou même de l'utilisation sur site en tant que service.

L'idée est que le consommateur ou le client du cloud doit veiller à la protection des données. Ce n'est que lorsque nous abordons les domaines de l'identité et de l'infrastructure des répertoires que nous commençons à voir soit des responsabilités partagées, soit éventuellement que le système d'exploitation, par exemple dans un modèle SaaS, relève uniquement de la responsabilité du fournisseur de cloud.

Figure 12: les responsabilités liées à la sécurité
(Source: Kevin Henry, Cloud Data Security for CCSP).

2.3.2 Cycle de vie des données dans le cloud:

Cela a été défini par l'Alliance pour la sécurité dans le cloud (Cloud Security Alliance) comme un moyen de garantir que nous comprenons que les données doivent être protégées tout au long du cycle de vie, de leur création à leur suppression. Le cycle de vie est défini comme suit : création, stockage, utilisation, partage, archivage, puis suppression [10].

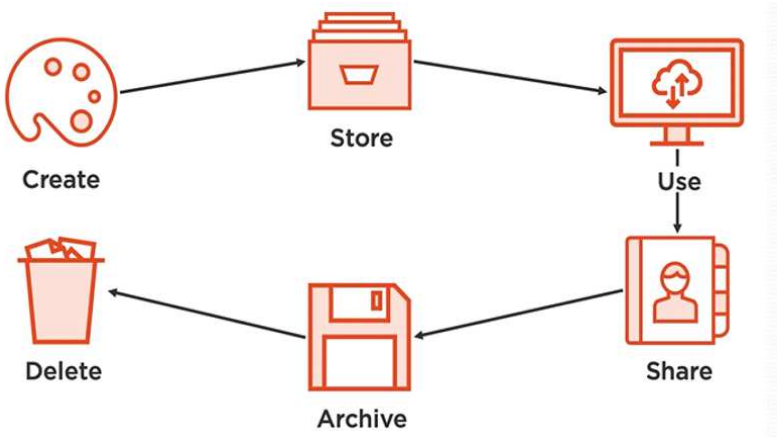


Figure 13: Cycle de vie des données dans le Cloud.
(Source: Kevin Henry, Cloud Data Security for CCSP).

Au cours de la phase de **création**, nous obtenons initialement les données provenant d'une source telle qu'un client, un partenaire commercial ou un autre système. À ce stade, nous entamons déjà le processus d'identification et de catégorisation des données. Une fois les données reçues, nous les étiquetons de manière adéquate, et nous pouvons même prendre des mesures pour garantir une réception sécurisée des données, par exemple, en utilisant le chiffrement.

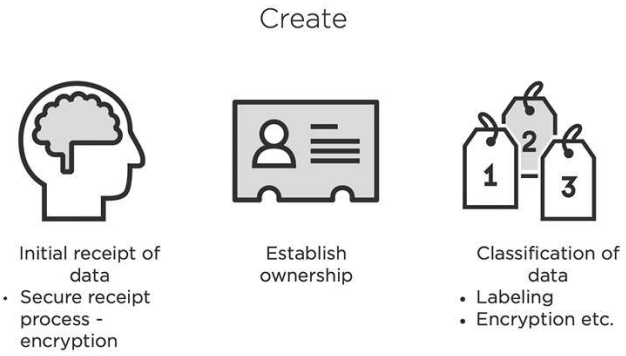


Figure 14: Phase de création des données.
(Source: Kevin Henry, Cloud Data Security for CCSP).

En général, l'étape de stockage (**STORE)** se déroule simultanément avec la création (souvent en parallèle). Nous recevons les données, les intégrons dans notre système et les conservons. Il est alors important de suivre les procédures de traitement des données associées à leur classification. Dès le début, nous procédons à leur étiquetage et spécifions les mesures de protection nécessaires au sein de l'organisation.

La prochaine étape du cycle de vie des données est leur utilisation (**Use**). Pendant cette phase, nous devons protéger les données, former nos utilisateurs sur leur manipulation et leur partage, et parfois utiliser des techniques d'obscurcissement ou de dissimulation. L'objectif est

de montrer aux utilisateurs uniquement les parties des données auxquelles ils ont accès, que ce soit en masquant certaines colonnes ou en utilisant d'autres méthodes d'obscurcissement.

Nous utilisons le chiffrement, le masquage et d'autres techniques pour dissimuler les données. Le chiffrement rend les données invisibles aux personnes non autorisées, le masquage empêche l'espionnage visuel des informations sensibles. De plus, nous utilisons des méthodes d'obscurcissement pour cacher les données et les rendre indéchiffrables. Nous veillons également à supprimer les informations personnelles identifiables afin de préserver l'anonymat des données. Cependant, nous devons rester vigilants face aux risques de l'agrégation de données et de la désanonymisation. Certaines questions démographiques peuvent permettre d'identifier spécifiquement les personnes, ce qui nécessite une protection adéquate.

Nous utilisons également des systèmes de prévention de la perte de données ou de fuite de données (DLP) pour protéger les données contre toute divulgation non autorisée. De plus, nous utilisons des outils tels que la gestion des droits numériques ou la gestion des droits d'information pour protéger les données lorsqu'elles sortent de l'organisation.

Pendant la phase de partage des données (**Share**), nous devons veiller à ce que seuls les utilisateurs autorisés puissent accéder et effectuer des actions autorisées sur ces données. Pour cela, nous utilisons des mesures de sécurité telles que l'authentification multifactorielle (MFA) pour vérifier l'identité des utilisateurs. De plus, nous appliquons des principes de contrôle d'accès, tels que le privilège minimal et le besoin de savoir, afin de limiter l'accès aux informations strictement nécessaires.

La prochaine étape concerne le principe du privilège minimal est utilisé pour déterminer les actions autorisées pour les utilisateurs ayant accès aux données. Nous devons nous assurer que seules les personnes autorisées à connaître ces informations peuvent les consulter ou les modifier. De plus, dans un environnement mondial tel que celui des services cloud, nous devons prendre en compte les lois relatives à la propriété intellectuelle et à l'emplacement où les données sont hébergées. Les réglementations sur la cryptographie dans certains pays peuvent également influencer les algorithmes utilisés. Il est donc crucial de respecter ces aspects lors de la gestion et du partage des données.

L'étape suivante du cycle de vie des données est l'archivage (**Archive**). Nous stockons généralement les données à long terme sur des supports tels que des disques optiques ou des bandes magnétiques. Il est important de choisir le bon support et de s'assurer que les données sont stockées correctement et dans un format durable. Nous devons également conserver les clés de chiffrement et les algorithmes nécessaires pour accéder aux archives. La protection physique des archives est essentielle pour éviter les pertes dues à des catastrophes naturelles

ou des conditions environnementales défavorables. La séparation géographique peut être utilisée pour garantir que les archives sont à l'abri des mêmes incidents que le centre de données principal.

La dernière étape du cycle de vie des données consiste à supprimer ou détruire les données (**Delete**). La destruction sécurisée des données est souvent utilisée pour garantir que les données sont définitivement supprimées. Il est important de pouvoir prouver de manière certaine que les données ont été détruites, afin de satisfaire les auditeurs et leur scepticisme professionnel. Les données doivent être conservées aussi longtemps que requis par la loi ou tant que l'entreprise en a besoin. Il est donc essentiel d'établir des politiques claires sur la durée de conservation des données et le processus de destruction à la fin de cette période.

En résumé, il est essentiel de protéger les données de manière cohérente et appropriée tout au long de leur cycle de vie. La classification des données peut changer, avec une possible réduction de niveau à la fin de leur durée de vie. La responsabilité de déterminer les procédures de manipulation appropriées incombe au propriétaire des données. Il est important de garantir la conformité aux réglementations et la disponibilité des données pour soutenir les opérations commerciales selon les besoins.

2.3.3 Cloud data classification:

La classification des données est une étape essentielle de la gestion des risques en cybersécurité. Elle consiste à identifier les types de données traitées et stockées dans un système d'information d'une organisation. Cette classification tient compte de la sensibilité des données et de l'impact potentiel en cas de compromission, de perte ou d'utilisation abusive.

Pour une gestion efficace des risques, il est recommandé de classer les données en se basant sur leur utilisation contextuelle et de créer un schéma de catégorisation prenant en compte l'impact significatif sur les opérations de l'organisation. La classification des données englobe la confidentialité, l'intégrité et la disponibilité, en utilisant une approche globale comprenant la taxonomie, les schémas et la catégorisation.

Classification de l'information:

Une information peut être catégorisée sous trois grandes classes [11][12]:

- **Information très sensible (Top secret):** C'est le niveau de classification le plus élevé pour les informations. La divulgation publique de ces informations pourrait causer des dommages exceptionnellement graves.

- **Information sensible (secret):** La divulgation publique de ces informations pourrait sérieusement nuire à l'entreprise.

- **Information ordinaire:** La divulgation de ces informations n'aurait aucun impact sur le fonctionnement de l'entreprise.

Bases de classification:

La classification des données repose sur les besoins commerciaux, où certaines informations doivent être classifiées afin de ne pas être divulguées à des personnes non autorisées, car elles sont cruciales pour les opérations commerciales. Cela peut inclure des éléments tels que des projets en cours ou des plans marketing confidentiels. De plus, la classification des données est également guidée par les lois et réglementations qui définissent les données à protéger. En général, la protection des données est divisée en deux catégories: **sensibles** et **critiques**, en fonction de leur niveau de sensibilité et de leur importance pour l'organisation.

- Les **données sensibles** sont des données qui pourraient nuire à une organisation ou à un individu si ces données étaient modifiées ou divulguées de manière inappropriée. La **sensibilité** de ces données correspond donc au degré de **confidentialité et d'intégrité** requis pour les protéger.

- La **criticité** concerne le degré de nécessité pour l'entreprise d'avoir accès à ces informations afin d'assurer le bon déroulement et fonctionnement de ses activités. Elle est souvent associée à la **disponibilité** des données, c'est-à-dire à la capacité de les rendre accessibles lorsque nécessaire.

Les organisations varient, les départements diffèrent, les systèmes sont diversifiés. Il est donc possible qu'une information soit très critique dans un système donné, tandis que cette même information ne soit que facultative dans un autre. Toutefois, il est essentiel de garantir une protection adéquate de ces données, de sorte que lorsqu'elles nécessitent un niveau élevé de protection dans un système, cette protection soit également assurée de manière cohérente dans les autres systèmes.

La valeur de la classification des données:

La classification des données est utilisée depuis longtemps pour aider les organisations à déterminer les mesures de protection appropriées pour les données sensibles ou critiques, qu'elles soient traitées ou stockées sur site ou dans le cloud. Elle permet d'évaluer la sensibilité et l'impact commercial des données, et de prendre en compte les risques associés à différents types de données.

Des organisations de normalisation telles que l'ISO et le NIST recommandent des schémas de classification pour une gestion efficace des informations en fonction du niveau de risque et de criticité. Cependant, une sur-classification des données peut entraîner des dépenses inutiles et limiter l'utilisation commerciale en imposant des exigences de conformité excessives. Il est donc important d'éviter de classer de manière trop générale des ensembles de données disparates au même niveau de sensibilité.

Niveaux de classification des données:

Le débat sur le nombre de niveaux de classification des données dans une organisation est fréquent et doit être décidé par le propriétaire des données. Chaque niveau doit être distinct, avec des exigences de manipulation clairement définies. Des directives précises doivent spécifier les critères de chaque niveau. Dans certains cas, il est préférable d'opter pour un niveau de classification supérieur par défaut. Cependant, cela dépend du type d'organisation, des lois en vigueur et de la nature des données.

Processus de classification des données:

Les clients sont souvent à la recherche de recommandations claires pour établir des politiques de classification des données. Ces étapes sont utiles tant pendant la phase de développement que lors de la réévaluation des données pour garantir leur classification appropriée et les protections correspondantes. Les étapes suivantes présentent une approche progressive basée sur des directives internationales pour aider les clients à élaborer leurs politiques de classification des données.

1. **Établir un catalogue de données (Establish a data catalog):** l'établissement d'un catalogue de données implique de réaliser un inventaire des différents types de données utilisés dans l'organisation, en tenant compte des réglementations et politiques de conformité. Une fois l'inventaire terminé, les données sont regroupées selon les niveaux de classification adoptés par l'organisation. Le catalogue de données AWS Glue permet de stocker, annoter et partager les métadonnées dans le cloud AWS, tout en offrant des fonctionnalités d'audit et de gouvernance complètes, notamment le suivi des modifications de schéma et les contrôles d'accès aux données.

2. **Évaluer les fonctions critiques de l'entreprise et réaliser une évaluation d'impact (Assessing business critical functions and conduct an impact assessment):** l'évaluation des fonctions essentielles de l'entreprise et la réalisation d'une évaluation d'impact sont des aspects importants pour déterminer le niveau de sécurité approprié des ensembles de données. Il est crucial de comprendre la criticité et l'importance de ces données pour l'activité de l'entreprise. Une fois les fonctions essentielles de

l'entreprise évaluées, les clients peuvent procéder à une évaluation d'impact pour chaque type de données.

3. **Étiquetage de l'information (Labeling Information):** Il est essentiel d'effectuer une évaluation d'assurance qualité pour étiqueter correctement les actifs et les ensembles de données dans leurs catégories de classification respective. Des étiquettes supplémentaires peuvent être nécessaires pour différencier certains sous-types de données en fonction des exigences de confidentialité ou de conformité en matière de sécurité. Les services tels que Amazon SageMaker et AWS Glue sont utiles pour obtenir des informations et faciliter les activités d'étiquetage des données.

4. **Gestion des actifs (Handling of assets):** Lorsqu'un niveau de classification est attribué à des ensembles de données, ils doivent être traités conformément aux lignes directrices appropriées pour ce niveau, incluant des contrôles de sécurité spécifiques. Il est important de documenter ces procédures de traitement et de les adapter en fonction des évolutions technologiques. Les considérations spécifiques relatives à la gestion des données sont fournies dans la section suivante du document sur la mise en œuvre des schémas de classification des données.

5. **Surveillance continue (Continuous monitoring):** Il est essentiel de maintenir une surveillance continue de la sécurité, de l'utilisation et des modèles d'accès des systèmes et des données. Cette surveillance peut être effectuée à l'aide de processus automatisés ou manuels, afin de détecter les menaces externes, de garantir le bon fonctionnement du système, de mettre à jour les systèmes et de suivre les changements dans l'environnement.

2.3.4 La protection des données:

La protection des données commence par l'élaboration d'une politique. Cette politique dicte le comportement des individus et est également appliquée via les contrats que nous avons avec un fournisseur de services cloud. La politique précise également qui est responsable et qui est le propriétaire des données. Elle impose le respect de certaines exigences et procédures de manipulation.

L'objectif de la politique est d'attribuer l'autorité à la fonction. Elle démontre que l'organisation reconnaît l'importance d'un élément et se conforme donc aux lois et réglementations en vigueur. Elle exige également la conformité aux procédures et, en ce qui

concerne les données, elle indique souvent la période de conservation en fonction des besoins commerciaux et légaux identifiés précédemment.

2.4 Technologies de sécurité des données:

2.4.1 Technologies de sécurité des données:

La protection des données sur le cloud nécessite une compréhension du stockage et du traitement des données. Le contrôle d'accès est essentiel pour garantir que seules les personnes autorisées peuvent effectuer des actions autorisées sur les données.

Le contrôle d'accès doit prendre en compte des aspects temporels, tels que les heures d'accès, ainsi que des attributs spécifiques aux données, tels que les rôles des utilisateurs.

Menaces sur les données:

Les données sont exposées à différentes menaces lors du stockage et de la transmission. Ces menaces incluent la divulgation non autorisée, où des personnes peuvent accéder ou lire des données auxquelles elles ne devraient pas avoir accès. Cela peut se produire lors du stockage lorsque des personnes obtiennent un accès non autorisé, ou lors de la transmission lorsque des attaques de l'homme du milieu permettent d'intercepter les données en cours de transfert. Ces menaces présentent le risque que les données soient altérées, supprimées ou perdues.

Figure 15: Menaces sur les données. (Source: Kevin Henry, Cloud Data Security for CCSP).

2.4.2 La protection des données dans le cloud:

Principes de la sécurité dans le cloud:

- Les données sont l'un des actifs les plus importants pour presque toutes les organisations aujourd'hui. Elles sont essentielles pour les opérations commerciales.
- Le principe fondamental de la sécurité des données et donc de leur protection est d'établir la propriété des données.

- Les données doivent être protégées tout au long de leur cycle de vie. Lorsque nous traitons avec un cloud, cela devient plus complexe que lorsque nous nous occupons uniquement de nos propres données.

Menaces et sécurité des données:

Les menaces pesant sur les données et la sécurité sont principalement liées à trois aspects essentiels: garantir la disponibilité des données lorsque nécessaire, les traiter correctement conformément à leur classification et aux règles définies par le propriétaire des données, et préserver l'intégrité de nos données ainsi que la manière dont nous les traitons.

Stockage:

Lorsque nous protégeons les données et leur stockage, nous prenons des mesures pour assurer leur sécurité. Cela comprend la création de copies de sauvegarde afin de garantir la disponibilité des données en cas de défaillance matérielle.

De plus, nous utilisons le chiffrement pour protéger les données stockées, empêchant ainsi toute personne non autorisée d'accéder aux informations sensibles, même en cas d'accès physique au matériel. L'un des points importants à considérer est de savoir qui détient la clé de chiffrement.

- Selon différents cas, le fournisseur de services cloud peut avoir les clés, notamment lors de la transmission ou dans le cadre du logiciel en tant que service ou Software as a Service.
- Toutefois, dans le cas du Platform as a Service ou de l'Infrastructure as a Service, le consommateur cloud peut être le seul à détenir la clé. Une mesure de protection des données importante est l'utilisation du hachage pour garantir l'intégrité des données.

Le hachage permet de détecter facilement toute altération, même minime, des données. En fin de compte, il est essentiel de mettre en place un contrôle d'accès pour garantir que seules les personnes autorisées peuvent effectuer des actions autorisées.

Il y a également les technologies de prévention de perte de données, parfois appelées prévention de fuite de données.

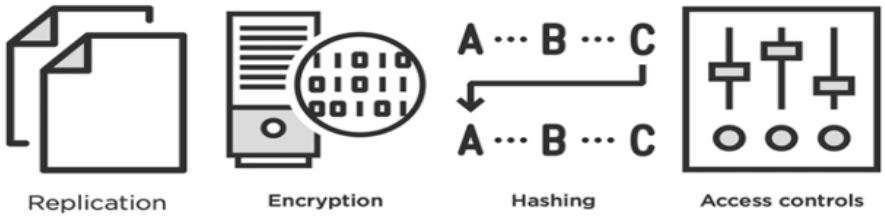


Figure 16: Protection des données et leur stockage.
(Source: Kevin Henry, Cloud Data Security for CCSP).

DLP (Data Loss/or Leakage Prevention DLP):

L'objectif de la prévention de la perte ou de la fuite de données est de protéger les informations confidentielles contre leur divulgation en dehors de l'organisation. Par exemple, si quelqu'un tente d'envoyer un e-mail contenant des données sensibles, le système bloque cette action. Cependant, un aspect moins connu de la DLP concerne également la protection contre la divulgation interne, en veillant à ce que les informations ne soient pas accessibles aux personnes au sein de l'organisation qui n'ont pas besoin d'y avoir accès [13].

DRM/ IRM:

DRM/ IRM "Protection of data that does go outside of the organization [14]"

Le DRM (gestion des droits numériques) ou l'IRM (gestion des droits d'information) vise à protéger les données qui sont envoyées à l'extérieur de l'organisation. Alors que la DLP vise à empêcher la fuite de données, le DRM s'applique lorsque nous devons partager des données tout en souhaitant les protéger.

L'objectif du DRM et de l'IRM est de garantir un certain niveau de protection pour les données partagées avec des tiers extérieurs, en évitant des actions non autorisées telles que la copie ou le stockage après expiration.

2.4.3 Chiffrement et gestion des clés:

Le chiffrement est l'un des outils les plus importants et précieux dont nous disposons pour protéger les données. En tant que professionnels de la sécurité cloud, nous devons prendre en compte divers aspects liés au chiffrement et à la gestion des clés.

Protection des données par le chiffrement:

Le chiffrement offre les avantages de la protection des données:

- Confidentialité.
- Intégrité.
- Control d'accès.
- Authentification.
- Non-répudiation.

Le chiffrement des données présente de nombreux avantages essentiels, tels que la confidentialité des données lors de leur transmission et de leur stockage, ainsi que l'intégrité

des données. Cela garantit également que l'accès à des informations sensibles est limité pour éviter toute violation de la loi. Même si une personne parvient à accéder à un fichier chiffré, elle ne pourra pas accéder aux informations qu'il contient. De plus, le chiffrement asymétrique permet de prouver l'identité des interlocuteurs, notamment grâce à l'utilisation de signatures numériques, assurant ainsi l'authentification et la non-répudiation des communications.

Avec le chiffrement asymétrique, nous pouvons nous assurer que seul le destinataire prévu peut ouvrir le message en utilisant sa clé publique, et nous pouvons prouver notre identité en tant qu'expéditeur en utilisant notre propre clé privée.

Qui détient les clés?

Un des facteurs qui influence le chiffrement dans le cloud est la possession des clés. Qui détient les clés?

- S'il s'agit d'un logiciel en tant que service (**Software as a Service**), le défi auquel nous sommes confrontés est que le fournisseur de logiciel en tant que service détient toutes les clés. Ils gèrent le chiffrement jusqu'à l'utilisateur final via le réseau. Ils gèrent également le chiffrement des données stockées sur leurs systèmes. Et bien que cela puisse être deux types de chiffrement différents, il ya des cas où les fournisseurs de services cloud utilisaient la même clé de chiffrement pour les données stockées de plusieurs clients. Lorsque nous effectuons un chiffrement lors de la transmission, nous utilisons souvent une combinaison de chiffrement symétrique et asymétrique. Et lors du chiffrement lors du stockage, il s'agit presque toujours d'un chiffrement symétrique [15].

- Dans le modèle de **Plateforme en tant que Service** (PaaS), la gestion des clés de chiffrement peut être effectuée par le fournisseur, le consommateur ou les deux parties. Le consommateur de cloud peut contrôler les clés utilisées pour chiffrer les données lors de la transmission et du stockage. Cependant, il y a des cas où le fournisseur de services cloud peut également avoir accès aux clés et potentiellement visualiser les données du consommateur.

  Pour le chiffrement dans une plateforme PaaS, on utilise généralement un chiffrement symétrique, avec une combinaison de chiffrement hybride pour la transmission et du chiffrement symétrique pour le stockage. Il est également possible de réaliser du chiffrement au niveau de l'application, en particulier pour des données sensibles telles que les numéros de cartes de paiement. Dans le contexte du chiffrement au sein d'une base de données, il est possible de chiffrer l'ensemble de la base de données ou des champs individuels, selon les paramètres de déploiement et l'accès aux clés. Il est

également important de noter que le fournisseur de services cloud peut utiliser le chiffrement natif de la base de données, ce qui lui donnerait également accès aux clés.

- Dans le cas de l'**Infrastructure en tant que Service** (IaaS), le consommateur dU cloud a le contrôle total sur les clés de chiffrement car il gère les applications et les plateformes. Le fournisseur de services cloud ne fournit que les infrastructures nécessaires. Le chiffrement peut être réalisé à la fois par le consommateur et par le fournisseur sur le réseau. Le consommateur est responsable de la majeure partie du chiffrement lors de la transmission des données, tandis que le fournisseur peut également participer dans certains cas. Pour le stockage, le consommateur dU cloud utilise un chiffrement symétrique pour avoir un contrôle complet sur la sécurité des données.

2.4.4 Hachage:

Le hachage est une fonction qui crée une empreinte numérique (hash) unique à partir d'un message donné. Contrairement au chiffrement, le hachage est irréversible et ne peut pas être utilisé pour récupérer le message original. Son principal objectif est d'assurer l'intégrité des données et la protection des informations sensibles telles que les mots de passe.

Les fonctions de hachage sont conçues pour être rapides, c'est-à-dire que le calcul de l'empreinte d'une donnée ne doit prendre qu'un temps négligeable. Elles doivent également éviter autant que possible les collisions [16], c'est-à-dire que deux empreintes identiques alors que les données sont différentes.

Le hachage garantit que même la plus petite modification dans le message d'origine entraîne une différence significative dans l'empreinte. Ainsi, il est utilisé pour vérifier l'authenticité des données sans révéler le contenu réel.

2.4.5 Chiffrement classique:

Cette section aborde de manière exhaustive le chiffrement classique [17], tant le chiffrement symétrique tel que l'algorithme AES (Advanced Encryption Standard), que le chiffrement asymétrique tel que le cryptosystème RSA (Rivest, Shamir et Adleman).

La configuration envisagée prend également en considération la possibilité d'appliquer une fonction de hachage ou une signature numérique sur les données, en plus du chiffrement.

Il est essentiel de distinguer deux cas spécifiques d'utilisation du chiffrement:

- **Côté serveur.**
- **Côté client.**

Le chiffrement côté serveur: est la méthode de chiffrement la plus couramment utilisée dans les services cloud. Comme le montre l'illustration suivante, l'utilisateur Bob envoie ses données en clair au service cloud pour les stocker. C'est le service cloud qui se charge de chiffrer les données avant leur stockage. Il est également possible que le service cloud applique une fonction de hachage ou une signature numérique aux données à stocker.

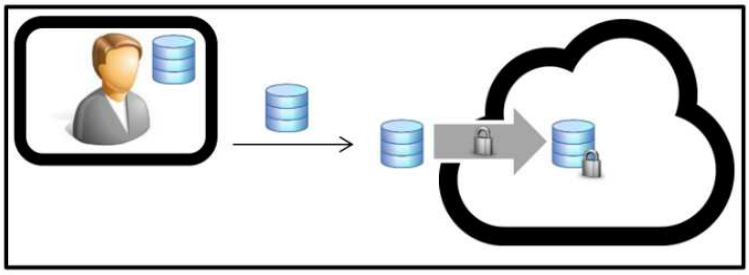


Figure 17: Chiffrement côté serveur.
(Source: TANIA MARTIN, Samls).

Dans cette configuration, la gestion des clés utilisées pour le chiffrement doit obligatoirement être effectuée côté serveur, ce qui présente deux scénarios distincts:

- Stockage non sécurisé des clés.
- Stockage sécurisé des clés.

1. **Stockage non sécurisé des clés:** il représente la situation la plus critique : les clés sont stockées de manière non sécurisée au même endroit que les données chiffrées de Bob. Si Oscar parvient à obtenir les données chiffrées, il est très probable qu'il puisse également mettre la main sur les clés, lui permettant ainsi de retrouver les données en clair de Bob.

Dans ce cas:

- La **confidentialité** des données par rapport à Oscar n'est pas garantie.
- De plus, l'**intégrité** des données d'Oscar n'est plus assurée si les clés utilisées pour la fonction de hachage ou de signature numérique ne sont pas également protégées.

2. Stockage sécurisé des clés: il sert à stocker les clés de manière sécurisée, qu'elles soient situées au même endroit que les données chiffrées de Bob ou non. Une méthode courante consiste à utiliser un **HSM** (Hardware Security Module), qui est un dispositif électronique considéré comme inviolable, permettant de générer, stocker et protéger les clés, ainsi que de réaliser des opérations cryptographiques.

Les HSM respectent des normes de sécurité élevées, offrant ainsi une solide garantie de sécurité. Une autre option est d'utiliser un keystore logiciel chiffré, qui stocke les clés secrètes et est protégé par un mot de passe ou d'autre mécanisme cryptographique.

Dans ce cas: même si Oscar parvient à obtenir les données chiffrées de Bob ainsi que le HSM ou le keystore, il lui sera impossible de récupérer les clés et évidemment les données en clair de Bob et donc:

- La **confidentialité** des données par rapport à Oscar est garantie.
- Également, l'**intégrité** des données d'Oscar est ainsi assurée si les clés utilisées pour la fonction de hachage ou de signature numérique sont protégées.

B. Chiffrement côté client:

Cette autre manière de chiffrement, bien qu'elle permette à l'utilisateur d'avoir un contrôle total sur ses données, est moins répandue. Comme le montre la figure suivante, l'utilisateur Bob chiffre ses données, puis les envoie au service cloud afin de les stocker. Bob peut aussi appliquer une fonction de hachage ou une signature numérique à ses données avant de les transmettre au service cloud.

Dans ce cas de configuration, la gestion des clés cryptographiques doit être réalisée côté client afin que l'utilisateur (client) conserve un contrôle maximal sur ses données. Le service cloud n'a donc connaissance ni des clés ni des données en clair (**Même si Oscar ne peut pas connaître les clés mais il faut les protéger localement**).

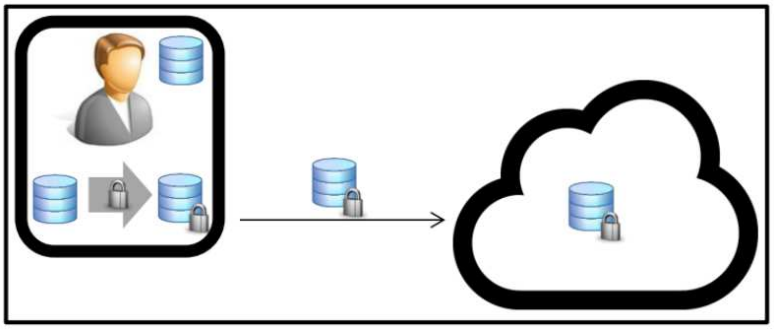


Figure 18: Chiffrement côté client (Source: TANIA MARTIN, Samls).

Et dans ce cas, la **confidentialité** et l'**intégrité** par rapport au Service Cloud et Oscar sont garanties.

2.4.6 Cloud Security Gateway:

Une autre technique pour protéger les données stockées dans le cloud est l'utilisation d'une Cloud Security Gateway (CSG).

Une CSG (Cloud Security Gateway) est un dispositif situé entre un utilisateur et un service cloud, souvent au sein de l'entreprise de l'utilisateur dans un contexte professionnel. Elle agit en tant qu'intermédiaire de confiance du point de vue de la sécurité, généralement sous la forme d'un proxy, qui chiffre et déchiffre les données qui transitent à travers elle.

Dans l'illustration ci-dessous [18], son utilisation est relativement simple: l'utilisateur Bob envoie ses données en clair à la CSG, qui chiffre ensuite ces données avant de les transmettre au service cloud.

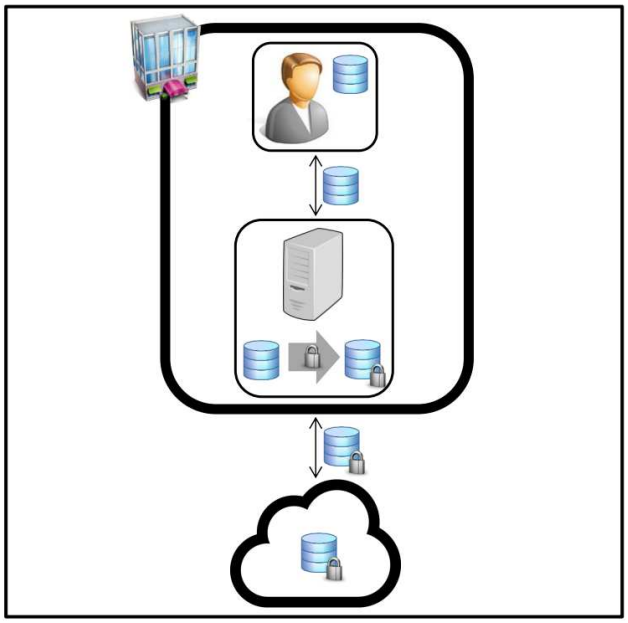


Figure 19: Cloud Security Gateway (Source: TANIA MARTIN, Samls).

Dans cette configuration, la gestion des clés cryptographiques est effectuée du côté client, c'est-à-dire au niveau de la passerelle de sécurité cloud (CSG), afin que l'entreprise conserve un contrôle maximal sur les données de Bob. Ainsi, le service cloud n'a connaissance ni des clés, ni des données en clair.

2.4.7 Gestion des événements de données dans le cloud:

Lorsqu'il s'agit de protéger nos données dans le cloud, nous devons surveiller la conformité [19]. C'est pourquoi nous devons également examiner la gestion des événements de données dans le cloud.

Gestion d'évènements (Event Management):

Lorsqu'il s'agit de protéger nos données dans le cloud, il est essentiel de surveiller la conformité. De nombreuses organisations ne sont pas conscientes du nombre de déploiements

dans le cloud qu'elles utilisent jusqu'à ce qu'elles mettent en place un **CASB** (**C**loud **A**ccess **S**ecurity **B**roker), ce qui leur permet de réaliser l'ampleur des données stockées dans différents emplacements du cloud.

- La première étape consiste à établir un contrat précisant les mesures de protection des données convenues entre le consommateur et le fournisseur de services cloud.
- Ensuite, il faut surveiller la conformité en examinant régulièrement les journaux d'accès pour détecter d'éventuels accès inappropriés ou abusifs. Les contrôles d'accès doivent être soigneusement révisés pour supprimer les privilèges non nécessaires.
- Il est également important de s'assurer que le chiffrement est appliqué correctement, afin d'éviter des situations où le trafic réseau n'est pas réellement chiffré.
- Enfin, il est recommandé de détruire de manière sécurisée le matériel obsolète pour empêcher toute récupération non autorisée des données.

Sources d'événements:

Parmi les nombreux événements que nous cherchons à identifier, nous examinons notamment l'adresse IP d'origine. Nous cherchons à savoir d'où ils proviennent dans le monde et tentons d'identifier les personnes impliquées. S'agit-il d'une entité autorisée ou d'une entité suspecte cherchant à accéder à nos systèmes ? La géolocalisation est également utilisée pour déterminer la position des individus au sein de l'organisation. Par exemple, l'accès d'une personne peut varier en fonction de sa présence physique au bureau. Si elle travaille à distance ou depuis son domicile, ses droits d'accès peuvent différer.

Log Review:

Lorsque nous effectuons une révision des logs, l'un des problèmes majeurs réside dans leur volume excessif. Nous disposons de quantités considérables de données de journalisation et il est tout simplement impossible de les examiner toutes manuellement faute de temps. En outre, la conservation des journaux engendre des coûts et a un impact sur les performances de nos réseaux et systèmes. Toutefois, il est essentiel de constituer une équipe compétente disposant des outils nécessaires pour analyser les journaux et identifier les éléments pertinents.

Enquête dans le cloud:

Lors d'une enquête dans le cloud, la responsabilité varie en fonction du modèle de déploiement.

- Dans le cas du logiciel en tant que service (SaaS), la majeure partie de la responsabilité incombe au fournisseur de services cloud, à l'exception du contrôle d'accès. En tant que consommateur de services cloud, nous déterminons qui a accès aux données, mais le fournisseur gère les journaux et les données.
- Avec la plateforme en tant que service (PaaS), la responsabilité est partagée, où nous gérons le contrôle d'accès et certains journaux, tandis que le fournisseur gère d'autres aspects tels que le journal de base de données.
- Enfin, avec l'infrastructure en tant que service (IaaS), la responsabilité principale revient au consommateur de services cloud, à l'exception de l'accès physique et de la sécurité matérielle pris en charge par le fournisseur.

2.5 Conclusion:

Dans ce chapitre, nous avons présenté les différents aspects et concepts liés à la sécurité dans le cloud, y compris les différentes techniques de protection des données.

Le chapitre suivant concernera la solution de cloud public AWS.

Chapitre 3

Cloud public et AWS comme solution cible

3.1 Introduction:

Dans ce chapitre, nous allons présenter la solution cloud public AWS et nos motivations de choix ainsi que ses mécanismes avancés de protection de données tels que l'IAM, Security Group, VPC et WAF.

3.2 Pourquoi le cloud public?

Parce qu'il est plus facile de mettre en œuvre une architecture complète, sans surcoûts (éviter le support matériels, la gestion des applications, des serveurs, de la disponibilité de services, aucune location de data center, la rapidité de recouvrement de service en cas de problème ou de crash).

La facilité de réaliser des prototypes rapides, évolutifs, services à la demande (si nous organisons un événement en une seule journée, nous pouvons créer un service pour une durée d'utilisation limitée).

3.3 Définition d'Amazon Web Services AWS:

AWS est l'acronyme de Amazon Web Services, est la plateforme cloud la plus complète et la plus largement utilisée à l'échelle mondiale. Elle propose une gamme étendue de plus de 200 services complets hébergés dans des centres de données répartis à travers le monde. Des millions de clients, parmi lesquels des startups innovantes, de grandes entreprises et des agences gouvernementales de premier plan, font confiance à AWS pour réduire leurs coûts, accroître leur agilité et stimuler leur innovation. En optant pour AWS, ces clients bénéficient d'une infrastructure cloud de confiance, de fonctionnalités avancées et d'une vaste communauté d'utilisateurs, leur permettant ainsi de rester compétitifs sur le marché mondial.

3.4 Pourquoi AWS?

- Parce qu'il fournit un grand nombre de cours gratuits et instruits, du matériel, des services, la plupart des entreprises adoptent aujourd'hui AWS comme une solution (90% de 100 entreprises fortunées).

- AWS a été nommé leader en matière d'infrastructure cloud et de services de plateforme pour la **12ème année consécutive** par Gartner, comme illustré dans la figure [20]. Ce rapport Magic Quadrant (MQ) fournit aux acheteurs de services cloud un aperçu du fournisseur d'après les critères de recherche de Gartner. Parmi les huit fournisseurs évalués, AWS se classe à la première place dans la catégorie Capacité d'exécution [20].

- Pour cette raison AWS est le leader. Ce qui valide l'étendue des capacités des services cloud d' AWS qui établissent la norme pour l'innovation des services cloud.

*Industry analyst firm **Gartner** has published its annual report evaluating cloud AI developer services, **the 2022 Magic Quadrant** for Cloud AI Developer Services (CAIDS). Amazon Web Services (AWS) was once again named a Leader and placed highest among 13 recognized vendors for "Ability to Execute."*

Figure 20: Magic quadrant pour les services d'infrastructures et de plateforme Cloud. (Source: Gartner).

3.5 Les principaux services d'AWS:

Amazon Web Services propose une vaste gamme de services cloud pour répondre à divers besoins. Voici quelques-uns des principaux services offerts par AWS:

3.5.1 Identity Access Management IAM:

3.5.1.1 Définition:

AWS a défini l'IAM comme suit: AWS Identity and Access Management (IAM) est un service web qui vous aide à contrôler en toute sécurité l'accès aux ressources AWS. Avec IAM, vous pouvez gérer de manière centralisée les autorisations qui contrôlent les ressources

AWS auxquelles les utilisateurs peuvent accéder. L'IAM permet de contrôler qui est authentifié (connecté) et autorisé (dispose de permissions) à utiliser les ressources.

3.5.1.2 Rôle:

La gestion des accès aux identités (IAM) est un service crucial qui vous permet de gérer les utilisateurs et leur niveau d'accès à la console AWS.

L'IAM vous permet donc de créer des utilisateurs, des groupes, des autorisations et des rôles, essentiellement pour limiter l'accès aux différentes parties de la plateforme AWS.

3.5.1.3 Caractéristiques d'IAM:

IAM offre les fonctionnalités suivantes pour la gestion des identités et des accès:

- Contrôle centralisé de votre compte AWS: IAM vous permet de gérer de manière centralisée les utilisateurs, les groupes et les rôles pour accéder à votre compte AWS.

- Accès partagé à votre compte AWS: Vous pouvez accorder des autorisations d'accès spécifiques à d'autres utilisateurs ou comptes AWS, permettant ainsi une collaboration sécurisée.

- Permissions granulaires: IAM vous permet de définir des autorisations d'accès précises en utilisant des stratégies qui spécifient les actions autorisées sur les services AWS.

- Fédération d'identité: IAM prend en charge la fédération d'identité, ce qui vous permet d'intégrer des sources d'identité externes telles que Active Directory, Facebook, LinkedIn, etc.

- Authentification multi-facteurs (MFA): IAM prend en charge l'authentification multi-facteurs pour renforcer la sécurité de l'accès à votre compte AWS.

- Accès temporaire: IAM permet de générer des jetons d'accès temporaires pour les utilisateurs, les appareils et les services, offrant ainsi un accès sécurisé pour des périodes déterminées.

- Rotation des mots de passe: IAM facilite la mise en place d'une politique de rotation régulière des mots de passe pour renforcer la sécurité.

- Intégration avec les services AWS: IAM s'intègre avec de nombreux services AWS, ce qui vous permet de contrôler finement les autorisations pour chaque service.

- Conformité PCI DSS: IAM est conforme aux exigences de sécurité de la norme de sécurité des données du secteur des cartes de paiement (PCI DSS).

3.5.1.4 Terminologie d'IAM:

Utilisateurs (Users): Un utilisateur IAM est une identité avec des informations d'identification à long terme qui est utilisée pour interagir avec AWS dans un compte. Ce sont des utilisateurs finaux (End-users) qui peuvent être personnes, employés d'une organisation, consommateurs d'applications… etc.

Groupes (Groups): un ensemble d'utilisateurs IAM. Les groupes permettent de définir les autorisations pour un ensemble d'utilisateurs dont chaque utilisateur hérite des autorisations du groupe.

Politiques (Policies): Une politique est un objet dans AWS qui définit les permissions. Elles sont constituées de documents, appelés documents de politique (Policy Documents); ces documents sont dans un format appelé JSON et ils donnent des autorisations sur ce qu'un utilisateur/groupe/rôle est capable de faire (les politiques sont donc les autorisations).

Rôles (Roles): Un rôle IAM est une identité qui peuvent être créer et qui dispose de permissions spécifiques avec des informations d'identification valables pour de courtes durées. Les rôles peuvent être assumés par des entités en qui vous avez confiance.
L'utilisateur crée des rôles et les attribuez ensuite aux ressources AWS. Le rôle est donc un moyen de permettre à une partie d'aws de faire quelque chose avec une autre partie. Ainsi que pouvoir donner à une VM dans aws la capacité d'écrire des fichiers sur S3, qui est un type de stockage au sein d'aws.

3.5.1.5 Fonctionnement d'IAM:

L'IAM fournit l'infrastructure nécessaire pour contrôler l'authentification et l'autorisation de votre compte AWS [21]. L'infrastructure IAM est illustrée par le diagramme dans la figure 21:

Tout d'abord, un utilisateur humain ou une application utilise ses identifiants de connexion pour s'authentifier auprès d'AWS. L'authentification est assurée par la correspondance entre les informations d'identification et un principal (un utilisateur IAM, un utilisateur fédéré, un rôle IAM ou une application) auquel le compte AWS fait confiance.

Ensuite, une demande est faite pour accorder au principal l'accès aux ressources. L'accès est accordé en réponse à une demande d'autorisation. Par exemple, lorsqu'un

utilisateur connecte pour la première fois à la console et qu'il se trouve sur la page d'accueil de la console, il n'accède pas à un service spécifique.

Lorsqu'il sélectionne un service, la demande d'autorisation est envoyée à ce service, qui vérifie si son identité figure sur la liste des utilisateurs autorisés, quelles politiques sont appliquées pour contrôler le niveau d'accès accordé et toutes les autres politiques éventuellement en vigueur. Les demandes d'autorisation peuvent être effectuées par des mandants de votre compte AWS ou d'un autre compte AWS en qu'il a confiance.

Une fois autorisé, le principal peut prendre des mesures ou effectuer des opérations sur les ressources de son compte AWS. Par exemple, le principal peut lancer une nouvelle instance Amazon Elastic Compute Cloud, modifier l'appartenance à un groupe IAM ou supprimer des buckets Amazon Simple Storage Service.

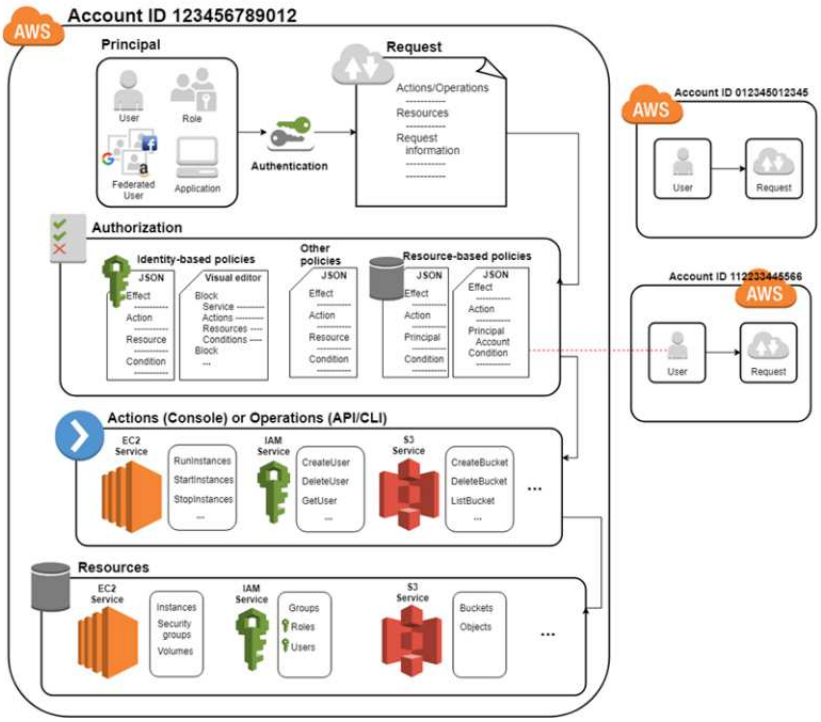


Figure 21: Le fonctionnement d'IAM (Source : aws docs).

3.5.2 Elastic Compute Cloud EC2.

3.5.2.1 Définition:

Amazon Elastic Compute Cloud (EC2) est un service web qui offre une capacité de calcul évolutive et redimensionnable dans le cloud. Il permet de réduire considérablement le temps nécessaire pour obtenir et démarrer de nouvelles instances de serveur et d'ajuster

rapidement la capacité en fonction de l'évolution des besoins informatiques. En permettant d'augmenter ou de réduire la capacité afin de répondre efficacement aux besoins de manière flexible, en quelques minutes.

3.5.2.2 Type d'instances EC2:

Il existe plusieurs types d'instance que l'utilisateur peut l'utiliser selon son besoin comme le montre la figure en dessous:

F: c'est pour Field Programmatic Gate Array (FPGA).
I: pour IOPS (Input/Output operations per second).
G: pour Graphics.
H: pour High Disk Throughout(débit élevé de disque).
T: pour cheap general purpose (e.g: T2 Micro).
D: pour Density (densité).
R: pour RAM.
M: pour Main choice, choix principal pour les applications générales.
C: pour Compute (un rapport élevé entre le calcul et la mémoire).
P: pour Pics, est utilisé pour les traitements graphiques et autres.
X: pour Extreme Memory (bon pour les applications full in-memory).
Z: pour zippy, Extreme Memory & CPU (une grande capacité de calcul et de mémoire à la fois).
A: pour Arm processor (Arm-based workloads), bon pour les charges de travail évolutives, supporté par arm).
U: pour Bare Metal.

Family	Speciality	Use case
F1	Field Programmable Gate Array	Genomics research, financial analytics, real-time video processing, big data etc
I3	High Speed Storage	NoSQL DBs, Data Warehousing etc
G3	Graphics Intensive	Video Encoding/ 3D Application Streaming
H1	High Disk Throughput	MapReduce-based workloads, distributed file systems such as HDFS and MapR-FS
T3	Lowest Cost, General Purpose	Web Servers/Small DBs
D2	Dense Storage	Fileservers/Data Warehousing/Hadoop
R5	Memory Optimized	Memory Intensive Apps/DBs
M5	General Purpose	Application Servers
C5	Compute Optimized	CPU Intensive Apps/DBs
P3	Graphics/General Purpose GPU	Machine Learning, Bit Coin Mining etc
X1	Memory Optimized	SAP HANA/Apache Spark etc
Z1D	High compute capacity and a high memory footprint.	Ideal for electronic design automation (EDA) and certain relational database workloads with high per-core licensing costs.
A1	Arm-based workloads	Scale-out workloads such as web servers
U-6tb1	Bare Metal	Bare metal capabilities that eliminate virtualization overhead

Figure 22: les différents types d'instance (Source : Ryan Kroonenberg, cloud guru).

3.5.2.3 Cycle de vie d'une instance:

Une instance Amazon EC2 traverse divers états depuis son lancement jusqu'à sa désactivation. Le schéma ci-dessous illustre les transitions entre ces états.

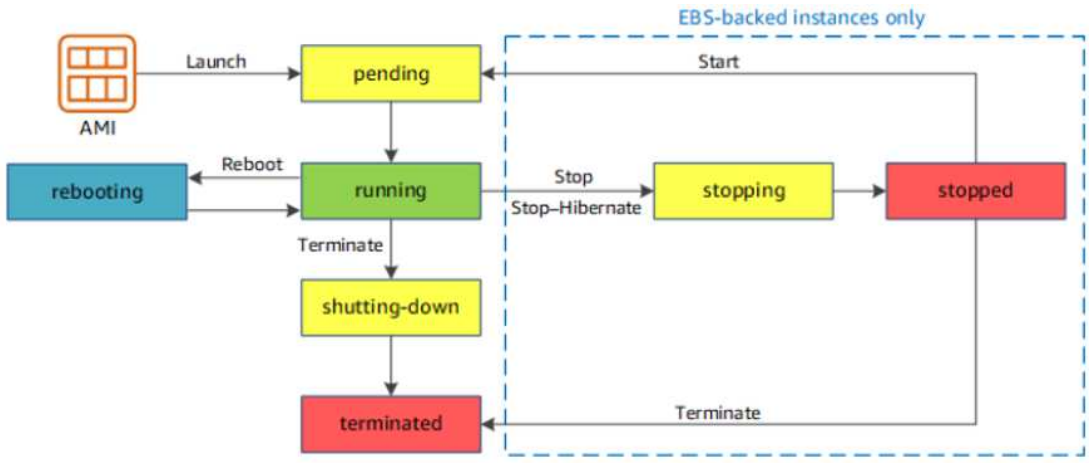


Figure 23: Cycle de vie d'une instance EC2 (Source : aws docs).

- **Pending:** L'instance se prépare à entrer dans l'état "running". Lorsqu'une instance est lancée pour la première fois ou redémarrée après avoir été dans l'état "stopped", elle passe à l'état "pending". L'utilisation de l'instance dans cet état n'est pas facturée.
- **Running:** L'instance est active (en cours d'exécution) et prête à être utilisée et son utilisation est donc facturée.
- **Stopping:** L'instance se prépare à être arrêtée ou à entrer en mode d'arrêt-veille prolongée. Des frais sont facturés lors de la préparation à la mise en veille prolongée de l'instance, mais aucune facturation est appliquée lors de la préparation à l'arrêt de l'instance.
- **Stopped:** L'instance est en état d'arrêt et n'est pas utilisable. Elle peut être démarrée sans cesse. Et dans ce cas aucune facturation n'est effectuée.
- **Shutting-down:** L'instance est en cours de préparation pour être supprimée et donc elle n'est pas facturée.
- **Terminated:** L'instance a été supprimée de manière permanente et ne peut plus être démarrée. Son utilisation n'est pas facturée mais Les instances réservées qui ont été résiliées continuent d'être facturées jusqu'à la fin de leur période de validité, conformément à l'option de paiement sélectionnée.

3.5.2.4 Définition de Elastic Block Store (EBS):

Amazon Elastic Block Store (EBS) offre des volumes de stockage persistants en blocs pour une utilisation avec les instances Amazon EC2 dans le cloud AWS. Chaque volume EBS est automatiquement répliqué dans la zone de disponibilité pour assurer une haute disponibilité et durabilité, protégeant ainsi contre les défaillances des composants.

3.5.2.5 Groupes de sécurité (Security groups):

Tout le trafic entrant est bloqué par défaut ainsi que tout le trafic sortant est autorisé et les modifications apportées aux groupes de sécurité prennent effet immédiatement. Un groupe de sécurité peut contenir un nombre illimité d'instances EC2.

Les groupes de sécurité sont des groupes d'état (STATEFUL) [22]; si l'utilisateur veut créer une règle d'entrée permettant au trafic d'entrer, ce trafic est automatiquement autorisé à sortir. Il ne peut pas bloquer des adresses IP spécifiques à l'aide des groupes de sécurité, mais plutôt utiliser les listes de contrôle d'accès au réseau.

3.5.2.6 Utilisation des rôles IAM avec EC2:

- Les rôles offrent une sécurité accrue par rapport au stockage de vos clés d'accès et clés secrètes sur des instances EC2 individuelles.

- De plus, les rôles sont plus faciles à gérer. Ils peuvent être attribués à une instance EC2 après sa création à l'aide de la console ou de la ligne de commande.

- Un autre avantage des rôles est leur universalité, ce qui signifie la possibilité de les utiliser dans n'importe quelle région.

5.3 Simple Storage Service S3.

3.5.3.1 Introduction:

Les clients de toutes tailles et de tous secteurs peuvent stocker et protéger n'importe quelle quantité de données pour tous les cas d'utilisation, tels que les lacs de données, les applications cloud-natives et les applications mobiles. Grâce à des classes de stockage rentables et à des fonctions de gestion conviviales, que l'utilisateur peut optimiser les coûts, organiser les données et configurer des contrôles d'accès précis pour répondre aux exigences spécifiques de l'entreprise, de l'organisation et de la conformité.

3.5.3.2 Définition:

AWS S3 (Amazon S3) est un service de stockage de données dans le cloud proposé par Amazon Web Services (AWS) qui est évolutif et économique et qui offre une grande capacité de stockage, une durabilité élevée ainsi qu'une disponibilité garantie et des options de classe de stockage pour répondre aux besoins de stockage de données de différentes entreprises et organisations. Permettant de stocker et de récupérer facilement des données à partir de n'importe ou et quelle que soit la quantité de données.

3.5.3.3 Les garanties de S3:

S3 bénéficie des garanties suivantes de la part d'Amazon [23]:

- Il est construit pour une disponibilité de 99,99 % pour la plateforme S3.
- Amazon garantit une disponibilité de 99,9 %.
- Amazon garantit une durabilité de 99,999999999% pour les informations S3 (les données de l'utilisateur ou ces objets ne peuvent pas être perdus).

3.5.3.4 Les principes de base de S3:

Les principes de base de S3 sont les suivants:

- S3 est basé sur des objets, c'est-à-dire qu'il permet de télécharger des fichiers.
- Les fichiers peuvent avoir une taille comprise entre 0 octet et 5 To.
- Le stockage est illimité.
- Les fichiers sont stockés dans des "Buckets".

- S3 est un espace de noms universel. En d'autres termes, les noms doivent être uniques au niveau mondial.
- S3 est basé sur les objets en considérant les objets comme des fichiers.
- Les objets se composent des éléments suivants:
    - Clé (il s'agit simplement du nom de l'objet).
    - Valeur (il s'agit simplement des données, constituées d'une séquence d'octets).
    - ID de version (important pour la gestion des versions).
    - Metadata (données sur les donnéesstockées).
    - Subresources (Listes de contrôle d'accès, Torrent).

3.5.3.5 Les classes de stockage S3:

AWS (Amazon Web Services) propose plusieurs classes de stockage pour Amazon S3 (Simple Storage Service), qui sont conçues pour répondre aux différents besoins de stockage de données. Les classes de stockage disponibles dans Amazon S3 sont les suivantes:

- **S3 Standard:** Cette classe de stockage est conçue pour stocker des données fréquemment utilisées et nécessitant une récupération rapide. Les données sont stockées dans plusieurs emplacements pour garantir la durabilité (99,999999999%) et la disponibilité (99,99%). S3 Standard prend également en charge la réplication croisée entre régions.

- **S3 - Infrequent Access IA (accès peu fréquent):** cette classe de stockage est conçue pour stocker des données auxquelles on accède moins fréquemment, mais qui nécessitent un accès rapide en cas de besoin. Le tarif est moins élevé que pour le S3, mais des frais d'extraction sont facturés (frais supplémentaires pour accéder aux données).

- **S3 Intelligent-Tiering:** Ce type de stockage est conçu pour les données dont l'accès varie de fréquent à rare. Les données sont automatiquement déplacées entre les classes de stockage (vers le niveau d'accès le plus rentable) en fonction de leur fréquence d'accès pour optimiser les coûts de stockage.

- **S3 One Zone - IA:** Ce type de stockage est conçu pour les cas où vous souhaitez une option moins coûteuse pour les données rarement consultées, où vous n'avez pas besoin de la résilience des données de plusieurs zones de disponibilité (les données sont stockées dans une seule zone de disponibilité, par opposition à plusieurs zones de disponibilité pour le stockage standard ou IA).

- **S3 Glacier:** S3 Glacier est une classe de stockage sécurisée, durable et peu coûteuse pour stocker des données qui ne sont pas souvent accédées, de manière fiable

n'importe et de n'importe quelle quantité à des coûts compétitifs ou inférieurs à ceux des solutions sur site. Les délais de récupération sont configurables de quelques minutes à quelques heures.

- **S3 Glacier Deep Archive :** est conçue pour stocker des données qui ne sont pas souvent utilisées et qui doivent être conservées à long terme. La GDA est la classe de stockage la moins chère d'Amazon S3, où un délai de récupération de 12 heures est acceptable (Récupération de données dans un délai de 12 heures ou plus), ce qui réduit encore plus les coûts de stockage par rapport à S3 Glacier.

Il est important de choisir la classe de stockage appropriée pour vos données afin d'optimiser les coûts de stockage et de répondre aux exigences de durabilité et de disponibilité.

3.5.3.6 Accélération du transfert S3:

Amazon S3 “Transfer Acceleration” permet des transferts rapides, faciles et sécurisés de fichiers sur de longues distances entre les end users et un bucket S3.

L'accélération du transfert tire parti des sites périphériques d'Amazon CloudFront répartis dans le monde entier. Lorsque les données arrivent à un site périphérique, elles sont acheminées vers Amazon S3 via un chemin réseau optimisé.

3.5.3.7 Sécurité et cryptage S3:

Par défaut, tous les buckets nouvellement créés sont PRIVÉS. L’utilisateur peut ainsi configurer le contrôle d'accès aux buckets à l'aide de:

- Politiques pour les buckets (Bucket Policies).
- Listes de contrôle d'accès (Access Control Lists).

Les buckets S3 peuvent être configurés pour créer des logs d'accès (Access logs) qui enregistrent toutes les demandes faites au bucket S3. Ce journal peut être envoyé à un autre bucket et même à un autre bucket dans un autre compte. Le cryptage en transit est assuré par SSL/TLS.

3.5.3.8 Caractéristiques et avantages:

AWS représente les avantages et les caractéristiques de Simple Storage Service comme suit:

- **Performance et durabilité des données (Data performance and durability):** Évolution des ressources de stockage pour répondre aux fluctuations de la demande, sans investissement initial ni cycle d'approvisionnement en ressources. Les données

sont disponibles en cas de besoin et protégées contre les défaillances, les erreurs et les menaces.

- **Sécurité, conformité et audit (Security, compliance, and auditing):**
  Sécuriser les données contre tout accès non autorisé grâce à des fonctions de cryptage et de gestion de l'accès. S3 gère également les programmes de conformité et prend en charge de nombreuses fonctionnalités d'audit.

- **Options de stockage flexibles (Flexible storage options):**
  Réaliser des économies sans sacrifier les performances. Stocker les données dans une large gamme de classes de stockage rentables qui prennent en charge différents niveaux d'accès aux données et sont conçues pour des cas d'utilisation spécifiques.

- **Contrôle granulaire des données (Granular data control):**
  Classifier, gérer et créer des rapports sur les données à l'aide de diverses fonctions de gestion du stockage. Enregistrer les activités, définisser des alertes et automatiser les flux de travail sans gérer d'infrastructure supplémentaire.

3.5.4 Virtual Private Cloud:

3.5.4.1 Définition de VPC:

AWS a définit le service VPC comme suit: “Un VPC est une portion isolée du Cloud AWS peuplée d'objets AWS, tels que les instances Amazon EC2.”

Amazon Virtual Private Cloud (AmazonVPC) vous permet de provisionner une section logiquement isolée du Cloud Amazon Web Services (AWS) où l’utilisateur peut lancer des ressources AWS dans un réseau virtuel qu’il définit. Il a un contrôle total sur l’environnement de réseau virtuel, y compris la sélection de la plage d'adresses IP, la création de sous-réseaux et la configuration des tables de routage et de passerelles réseau (Network gateways) [24].

3.5.4.2 Diagramme VPC:

La figure suivante décrit le diagramme d'un VPC. Cette ligne rouge représente la région, c'est donc us-east-1 et à l'intérieur de cette région on a un VPC et à l'extérieur de VPC nous avons deux façons de se connecter via une passerelle internet (**Internet Gateway**) et une passerelle privée virtuelle (**Virtual Private Gateway**), ces deux connexions vont au routeur dans ce VPC.

Le routeur dirige ensuite le trafic vers les différentes tables de routage, et ces tables de routage dirigent ensuite le trafic vers le réseau ACL (**Network ACL**) qui est la première ligne de défense, et qui agit également comme un **FW**.

Ensuite, nous avons les groupes de sécurité qui sont des groupes d'état et qui agissent comme une ligne de défense complète contre nos instances EC2. Et ici, on a deux sous-réseaux différents:

- Le premier sous-réseau est public (**Public subnet**) ou le trafic internet est donc accessible pour toutes les instances EC2 dans ce Public subnet.
- Le deuxième sous-réseau est privé (**Private subnet**) qui signifie simplement que nos instances EC2 ne peuvent pas accéder à l'internet.

Ensuite, Vous pouvez vous connecter à ces instances EC2 en passant par le public subnet et en utilisant ssh directement de la première instance à la deuxième instance, c'est ainsi que vous y accéderez.

Le VPC est donc un ensemble d'Internet gateways avec, des tables de routage, des ACL de réseau, des groupes de sécurité, des instances EC2 ainsi que des sous-réseaux publics et privés.

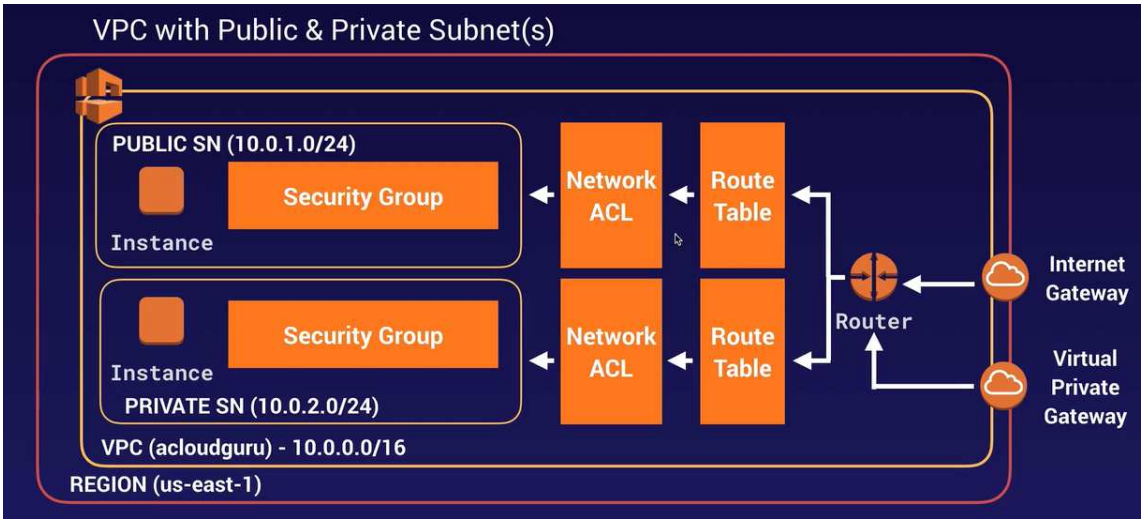


Figure 24: Diagramme VPC (Source : Ryan Kroonenberg, Cloud Guru).

3.5.4.3 Caractéristiques des VPCs:

- Il permet de lancer des instances dans le subnet selon le choix.
- Il est possible d'attribuer des plages d'adresses IP (IP address range) personnalisées à chaque subnet.

- Il permet la configuration des tables de routage entre les subnets.
- La création d'une passerelle internet et son attachement au VPC sont réalisables.
- Il permet de bénéficier d'un meilleur contrôle de la sécurité des ressources AWS.
- Il permet de gérer la sécurité des instances à travers les groupes de sécurité d'instance (Instance Security Groups).
- Les listes de contrôle d'accès aux sous-réseaux ou Access Control Lists (ACLs) pour gérer l'accès au niveau du subnets.

3.5.4.4 Différence entre VPC par défaut et VPC personnalisé:

- Le VPC par défaut est convivial et permet un déploiement immédiat d'instances.
- Tous les subnets du VPC par défaut sont connectés à Internet (disposent d'une route vers Internet).
- Chaque instance EC2 possède une adresse IP publique et une adresse IP privée.

3.5.4.5 VPC Peering:

- Le peering VPC permet la connexion de deux VPC en utilisant des adresses IP privées via une route réseau directe.
- Les instances se comportent comme si elles étaient sur le même réseau privé lorsqu'elles sont en peering.
- Vous pouvez effectuer un peering VPC avec d'autres comptes AWS et d'autres VPC du même compte.
- Le peering VPC se fait selon une configuration en étoile, où un VPC central est en peering avec quatre autres VPC.
- Le peering VPC **n'est pas transitif**, c'est-à-dire qu'il n'est pas possible d'étendre le peering au-delà des VPC impliqués directement.

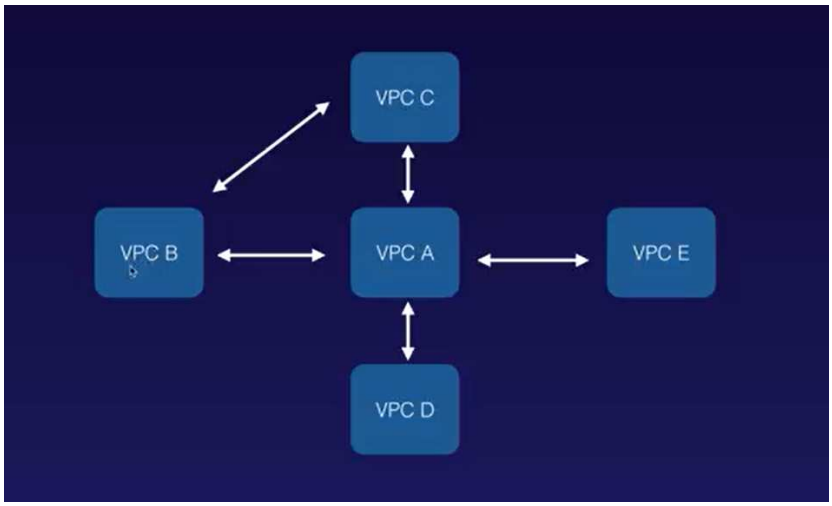


Figure 25: VPC Peering (Source : Ryan Kroonenberg, Cloud Guru).

3.5.5 AWS Web Application Firewall WAF:

3.5.5.1 Définition d'AWS WAF:

AWS WAF est un service de pare-feu d'application web proposé par Amazon, qui surveille les requêtes HTTP(S) adressées aux ressources d'application web protégées. Il permet de gérer l'accès au contenu en fonction de critères que l'utilisateur définit, tels que les adresses IP source des requêtes ou les valeurs des chaînes de requête. En fonction de ces critères, si la requête est autorisée, le service associé à votre ressource protégée répondra à la requête en permettant l'accès et en envoyant la réponse attendue. Sinon, il répond avec un code d'état HTTP 403 (**FORBIDDEN ACCESS**) ou en envoyant une réponse personnalisée.

Les ressources que AWS WAF peut protéger comprennent Amazon API Gateway APIS, les distributions d'Amazon CloudFront, Application Load Balancers et AWS AppSync.

3.5.5.1.1 Amazon CloudFront:

Le CloudFront est un service de réseau de diffusion de contenu (CDN) proposé par Amazon Web Services (AWS). Il permet d'améliorer la performance, la disponibilité et la sécurité de la distribution de contenu web en fournissant une diffusion mondiale rapide et sécurisée des données. CloudFront agit comme un intermédiaire entre les utilisateurs finaux et les serveurs d'origine, en plaçant le contenu statique et dynamique dans des emplacements géographiquement dispersés appelés "Edge Locations".

CloudFront peut être utilisé pour distribuer différents types de contenu, tels que:

- Des fichiers statiques.
- Des vidéos en streaming.
- Des applications web.
- Des API et des jeux en ligne… etc.

Il est hautement évolutif, fiable et s'intègre facilement avec d'autres services AWS, offrant ainsi une solution sécurisée et complète pour la distribution de contenu web à grande échelle.

3.5.5.1.2 Application Load Balancer (ALB):

ALB est un service fourni par Amazon Web Services (AWS) qui distribue le trafic entrant sur plusieurs cibles, telles que les instances EC2, les conteneurs et les adresses IP, dans une région spécifique. Fonctionnant au niveau de la couche application (couche 7) du modèle OSI, l'ALB achemine intelligemment les demandes en fonction du contenu de l'application.

Il agit comme un proxy inverse, appliquant des règles et des politiques pour transmettre les demandes aux cibles appropriées. L'ALB offre des fonctionnalités avancées telles que le routage basé sur le chemin, le routage basé sur l'hôte et l'intégration avec d'autres services AWS.

Avec ALB, l'utilisateur obtient une haute disponibilité et une grande évolutivité pour les applications web en répartissant le trafic sur plusieurs cibles, ce qui améliore les performances globales et la tolérance aux pannes. Il effectue automatiquement les Health checks sur les cibles et ne dirige le trafic que vers des instances saines (healthy instances), garantissant ainsi que les demandes sont acheminées vers des ressources fonctionnelles.

Il prend également en charge la terminaison SSL/TLS et fournit des journaux d'accès (access logs) pour la surveillance.

Globalement, l'ALB offre une solution efficace et flexible pour améliorer la répartition du trafic et les performances des applications web.

3.5.5.1.3 Amazon API Gateway:

Amazon API Gateway est un service de gestion et de création de passerelles d'API entièrement géré fourni par Amazon Web Services (AWS). Il permet aux développeurs de créer, déployer et gérer des API robustes pour leurs applications.

L'API Gateway agit comme un point d'entrée pour les clients souhaitant accéder aux services et données via une API. Il fournit une interface permettant de définir les endpoints de l'API, les opérations disponibles, les paramètres, les modèles de données et les politiques de sécurité.

Les fonctionnalités clés de l'API Gateway incluent:

- **Création d'API RESTful ou WebSocket:** l'utilisateur peut créer des APIs RESTful traditionnelles ou des APIs en temps réel basées sur le protocole WebSocket pour prendre en charge différents types d'applications.

- **Gestion des versions et du déploiement:** L'API Gateway facilite la gestion des versions des APIs et permet de déployer rapidement des mises à jour sans perturber les clients existants.

- **Sécurité et autorisation:** l'utilisateur peut protéger ses APIs en définissant des politiques de sécurité telles que l'authentification, l'autorisation basée sur les rôles IAM (Identity and Access Management) d'AWS, ou l'intégration avec d'autres services de gestion des identités.

- **Surveillance et journalisation:** avoir accès à des métriques et des journaux détaillés pour suivre les performances des APIs, identifier les problèmes et prendre des mesures correctives.

- **Intégration avec d'autres services AWS**: L'API Gateway s'intègre facilement avec d'autres services AWS tels que Lambda, DynamoDB, S3 et bien d'autres encore, pour créer des applications sans serveur et des architectures évolutives.

- Utilisation des modèles AWS CloudFormation pour permettre la création d'API.

- La possibilité d'utiliser des noms de domaine personnalisés est prise en charge.

- Il est possible d'intégrer AWS WAF pour protéger les APIs contre les attaques web courantes.

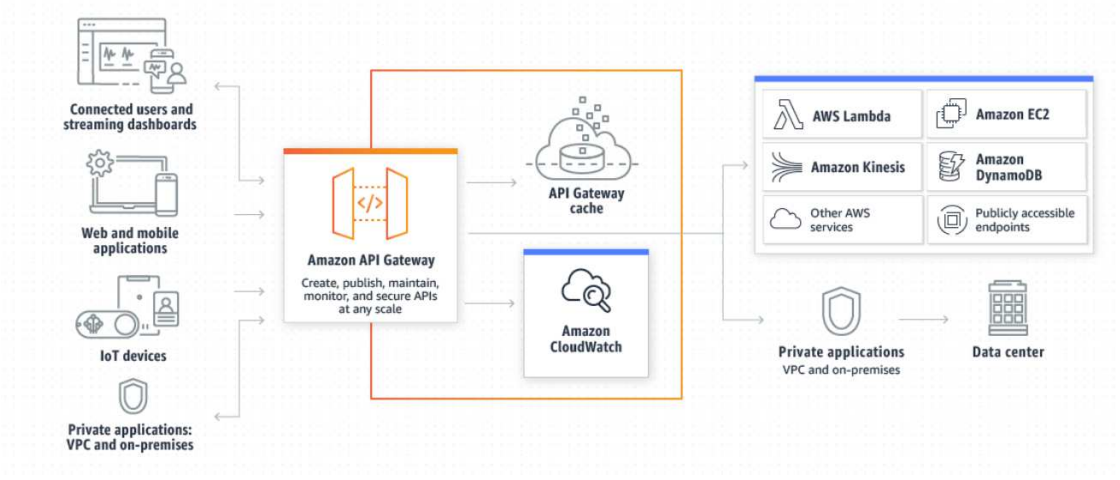


Figure 26: Architecture d'Amazon API Gateway (Source: aws docs).

3.5.5.1.4 AWS AppSync:

AWS AppSync est un service de développement d'applications sans serveur qui facilite la création d'applications évolutives et de temps réel, grâce à un point de terminaison unique pour interroger, mettre à jour ou publier des données en toute sécurité. Il permet aux développeurs de créer des API GraphQL (un langage de requête pour les API) pour leurs applications et de les connecter à diverses sources de données, telles que des bases de données, des services AWS et des sources personnalisées.

Il gère automatiquement la mise en cache des données et la distribution de celles-ci aux utilisateurs finaux, ce qui améliore les performances de l'application et réduit la latence.

Le service offre également des fonctionnalités de sécurité avancées, telles que la gestion fine des autorisations basée sur les rôles IAM (Identity and Access Management) d'AWS, permettant de contrôler l'accès aux données et aux fonctionnalités de l'application.

- En utilisant AWS AppSync, les développeurs frontend peuvent créer des API GraphQL qui leur permettent d'interroger plusieurs bases de données, microservices et API à partir d'un seul point d'accès GraphQL.
- Les développeurs frontend peuvent également créer des API Pub/Sub qui leur permettent de diffuser en temps réel des mises à jour de données aux clients d'API qui sont abonnés via des connexions WebSockets sans nécessiter de serveur.

3.5.5.2 Fonctionnement d'AWS WAF:

AWS WAF est utilisé pour gérer la façon dont les ressources protégées réagissent aux requêtes Web HTTP/HTTPS. Pour ce faire, il faut configurer une liste de contrôle d'accès Web (ACL) et l'associer à une ou plusieurs ressources d'applications Web souhaitant les protéger. Les ressources associées redirigent ensuite les requêtes entrantes vers AWS WAF pour être inspectées par l'ACL Web [25].

Figure 27: Fonctionnement d'AWS WAF (Source : aws docs).

Web Control Access Lists (ACLs), ou Les listes de contrôle d'accès au web (Web ACL) sont utilisées pour surveiller les requêtes HTTP(S) ciblant les ressources AWS (Amazon API Gateway APIS, les distributions d'Amazon CloudFront et Application Load Balancers), améliorant ainsi leur protection contre les attaques potentielles.

Dans chaque ACL Web, l'utilisateur a la possibilité d'inclure des règles et des groupes de règles qui définissent les actions appropriées pour les requêtes entrantes. Ces règles et groupes de règles établissent des conditions telles que les adresses IP d'origine et Pays d'origine de la demande les modèles de regex à analyser (Correspondance de chaîne ou expression régulière dans une partie de la demande). En fonction de ces conditions, l'ACL Web détermine s'il convient d'autoriser ou de refuser chaque demande. En réponse à la demande, la ressource

AWS protégée fournit le contenu demandé (pour les demandes autorisées) ou renvoie un code d'état HTTP 403 (pour les demandes bloquées).

Rules (Règles): les règles dans AWS WAF consistent en une déclaration qui spécifie les conditions d'inspection ainsi que l'action à prendre lorsque les requêtes Web correspondent à ces critères. Lorsqu'une requête Web correspond aux critères, on parle de correspondance. Et les actions à effectuer lors dans ce cas sont:

- Bloquer la requête.
- Autoriser la requête pour accéder au contenu ou à la ressource (traitement réponse).
- Compter les requêtes correspondantes.
- Effectuer des vérifications de robots en lançant des puzzles CAPTCHA.

Il est important de noter qu'une règle n'est pas une ressource AWS WAF à part entière, elle existe uniquement dans le contexte d'une ACL Web ou d'un groupe de règles.

Rule Groups (Les groupes de règles): un groupe de règles est une ressource AWS WAF qui peut être utilisée pour regrouper et organiser DES règles spécifiques. Les groupes de règles permettent de définir des règles directement dans une ACL Web ou dans des groupes de règles réutilisables. L'utilisateur a la possibilité d'utiliser des groupes de règles gérés fournis par AWS et des vendeurs de AWS Marketplace, ainsi que de créer des propres groupes de règles personnalisés.

3.5.5.3 Caractéristiques et avantages:

Amazon Web Service représente les principaux caractéristiques et avantages suivantes:

- **Agile protection against web attacks**:

  AWS WAF offre une protection agile contre les attaques web. Les règles de protection peuvent être propagées et mises à jour en moins d'une minute, ce qui permet une réaction rapide en cas d'attaque ou de problèmes de sécurité. Le WAF prend en charge des centaines de règles qui permettent d'inspecter différentes parties des requêtes web, tout en minimisant l'impact sur la latence du trafic entrant.

- **Improved web traffic visibility:**

  AWS WAF améliore la visibilité du trafic web en offrant une vue en temps quasi réel du trafic. Il permet de pouvoir exploiter cette visibilité pour créer de nouvelles règles ou des alertes dans Amazon CloudWatch. De plus, AWS WAF propose une fonctionnalité de journalisation complète, ce qui permet de capturer toutes les données d'en-tête de chaque requête Web inspectée. Et donc utiliser ces journaux pour automatiser la sécurité, effectuer des analyses ou des audits.

- **Save time with managed rules:**

  Grâce aux règles gérées pour AWS WAF, l'utilisateur peut protéger rapidement ses applications web ou ses APIs contre les menaces courantes. Les règles gérées sont automatiquement mises à jour et disponibles auprès d'AWS ou de fournisseurs sur AWS Marketplace. Cela qui permet de bénéficier d'une protection instantanée sans avoir à configurer manuellement chaque règle.

- **Ease of deployment and maintenance:**

  Le déploiement et la maintenance d'AWS WAF sont faciles. Il assure la protection des applications déployées sur Amazon CloudFront, Application Load Balancer ou Amazon API Gateway sans qu'il soit nécessaire de déployer un logiciel supplémentaire, de configurer le DNS ou de gérer les certificats SSL/TLS.

3.6 Conclusion:

Ce chapitre présente un bref aperçu des aspects et des composants les plus couramment utilisés pour assurer la sécurité dans l'environnement AWS. Nous avons donné leurs définitions et caractéristiques ainsi que leurs fonctionnements. De plus, nous mentionnons les différents services que nous utiliserons dans nos USE CASES lors de la phase d'implémentation que nous verrons dans le chapitre 4.

Chapitre 4
Implémentation

4.1 Introduction:

L'objectif de ce chapitre consiste à introduire les techniques courantes pratiques de sécurité de données dans l'environnement cloud public aws, comprendre leurs comporants ainsi que les enjeux de sécurité d'une part, et de proposer une architecture qui fait face aux problématiques et failles de sécurité d'autre part.

Tout d'abord, nous avons construit un plan de projet avec une application de cartographie mentale en ligne MindMeister afin de visualiser, partager et présenter nos idées au long de notre projet.

Figure 28: Plan de projet.

4.2 Environnement et outils:

Afin de mener à terme notre proposition d'architecture sécurisé cités dans le section architecture, nous avons utilisé les outils ci-dessous:

Local devices & tools:

- Machine local fonctionnant sous OS Windows.
- VMware Workstation.
- VM Linux Ubuntu pour tourner Terraform 1.4.6, AWS CLI en local et open SSL
- vscode avec extension terraform 1.4, remote SSH et AWS CLI
- Mobaxterm, Winscp et Putty, keepass.
- Connection Internet 4G Algérie Télécom.

Remote cloud & tools:

- Console aws
- Cloud9 with Terraform 1.4.6, AWS CLI et remote SSH

- Draw.io,
- Terraform.io
- Github

4.3 Architecture proposée :

La première étape est de comprendre les composants d'AWS liés à l'aspect sécurité. L'idée est d'installer une application web avec DB (front end, backend), et d'utiliser une architecture de sécurité tierce:

1. Gérer l'authentification avec les services IAM et configurer la MFA en tant qu'un premier niveau de protection de notre environnement aws.
2. Créer un VPC avec des Subnets Privé et public pour la ségrégation des Zones et les utiliser pour protéger les serveurs en zone privé (2ème niveau de protection).
3. Mettre en place un ALB avec un Listener en https dans une zone publique et autoriser le flux uniquement depuis le ALB vers nos serveurs applicatifs (3ème niveau de protection).
4. Ajouter la security group pour la protection des ressources cloud (4ème niveau de protection).
5. Tester quelques attaques et essayer d'y remédier avec la solution de protection applicative WAF proposée par AWS (5ème niveau de protection).

La deuxième étape : est de familiariser avec le monde DevOps qui combine les pratiques de développement logiciel et d'exploitation des opérations de sécurité dans AWS. L'objectif est d'accélérer la livraison de déploiement en visant l'automatisation en utilisant l'infrastructure en tant que code (IaC) avec Terraform depuis Cloud9, local machine ou VMware ubuntu.

Le schéma de la figure 29 illustre notre proposition d'architecture full Cloud AWS qui nous aidera à comprendre les composons participantes de sécurité dans le cloud.

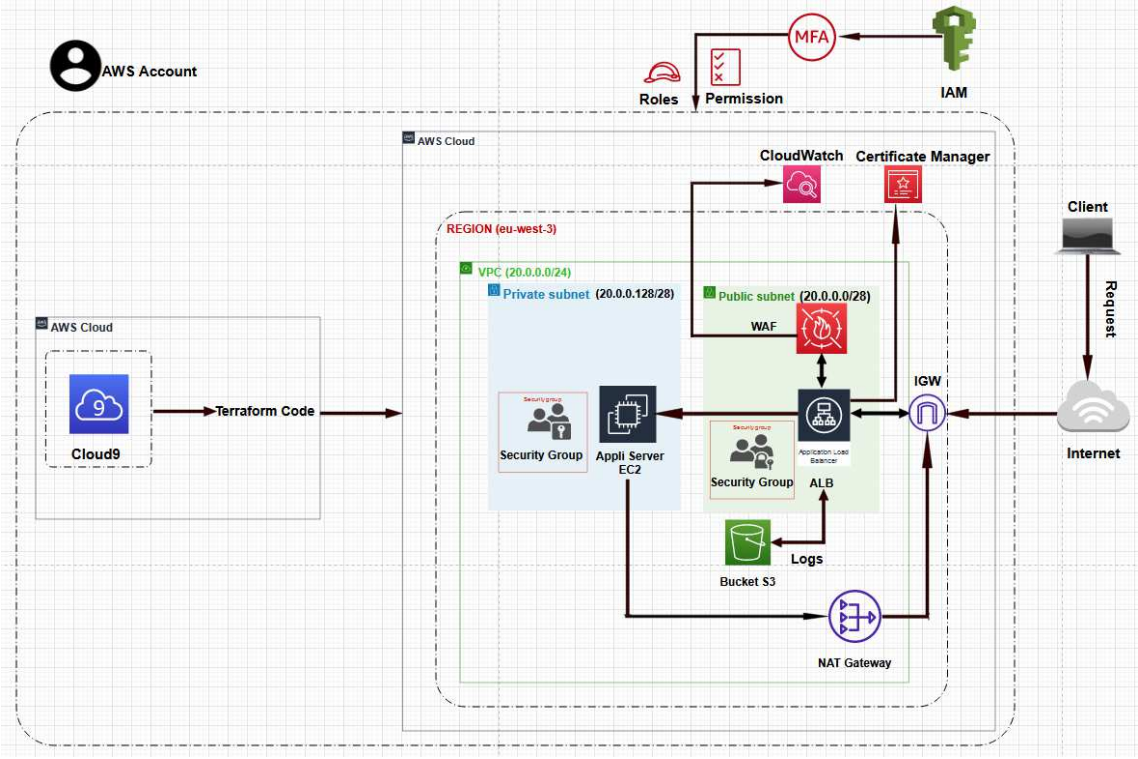


Figure 29: Architecture Proposée.

4.4 Attaques:

Nous avons utilisé le serveur DVWA (**D**amn **V**ulnerable **W**eb **A**pplication) sur nos instances EC2, qui est une application web permettant aux professionnels de la sécurité et aux développeurs de tester et de pratiquer certaines des vulnérabilités Web les plus courantes dans un environnement légal avec une interface simple et directe. Afin de tester les vulnérabilités (attaques) et d'y remédier.

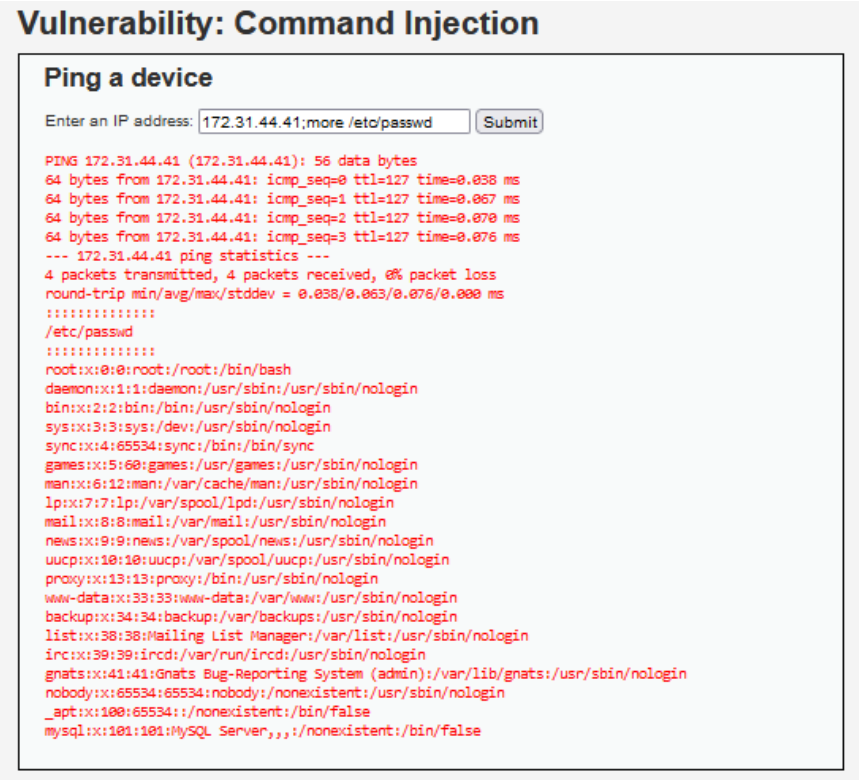
Nous nous sommes concentrés sur les aspects de sécurité AWS plutôt que sur le développement d'une application web d'où la raison d'utilisation de DVWA.

4.4.1 Type d'attaques testées:

Vulnerability Command Injection: exécuter des commandes systèmes sur le serveur hébergeant l'application, l'attaquant peut insérer des caractères spéciaux ou des séquences de commande qui seront interprétés et exécutés par le système d'exploitation.

L'attaque par injection de commande peut avoir des conséquences graves. Un attaquant peut exécuter des commandes arbitraires sur le serveur, ce qui lui donne un contrôle total sur le système. Cela peut entraîner la divulgation d'informations sensibles, la modification ou la suppression de données, voire la compromission complète du serveur.

La capture ci-dessous montre bien comment un attaquant peut lire le contenu de **/etc/passwd** à distance avec la remote command injection.

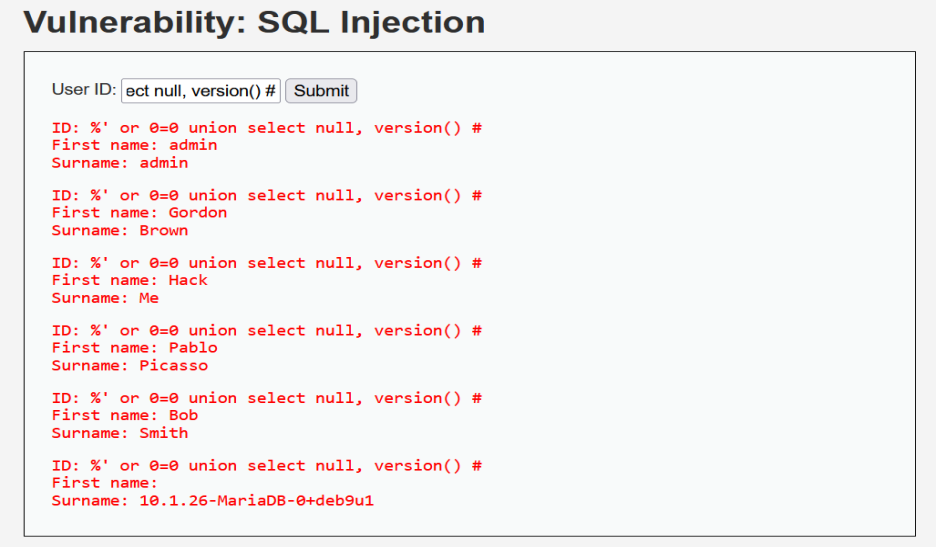


<u>Vulnerability Injection SQL</u>: insérer du code SQL malveillant dans une requête SQL afin de compromettre le serveur et de manipuler ou d'accéder à des données non autorisées dans une base de données.

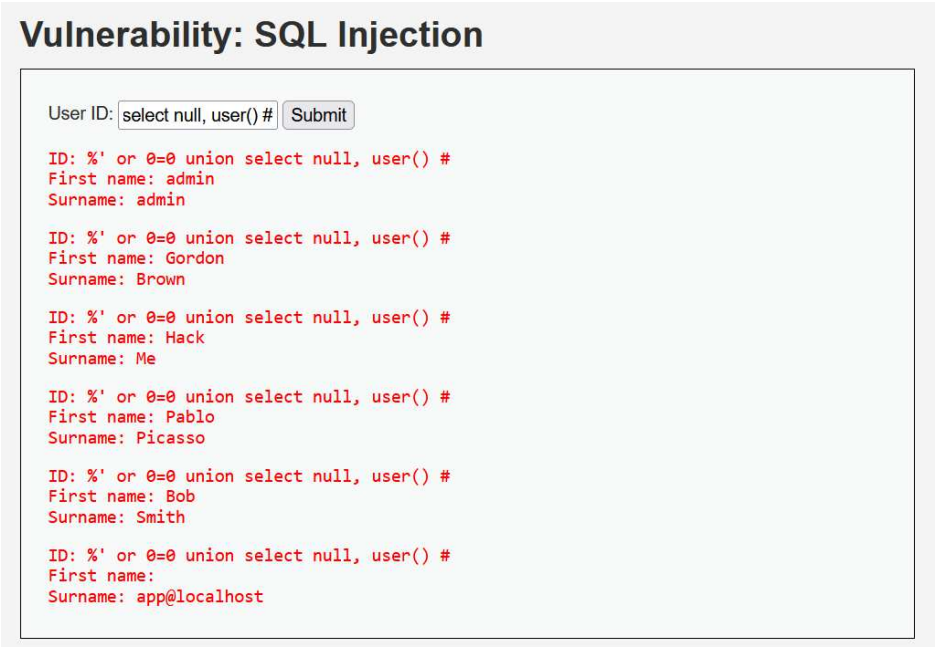
Elle est utilisée pour contourner les mécanismes de sécurité et exécuter des commandes SQL non autorisées.

Les captures suivantes illustrent comment un attaquant peut accéder à la BDD et lire son contenu à distance avec SQL injection query:

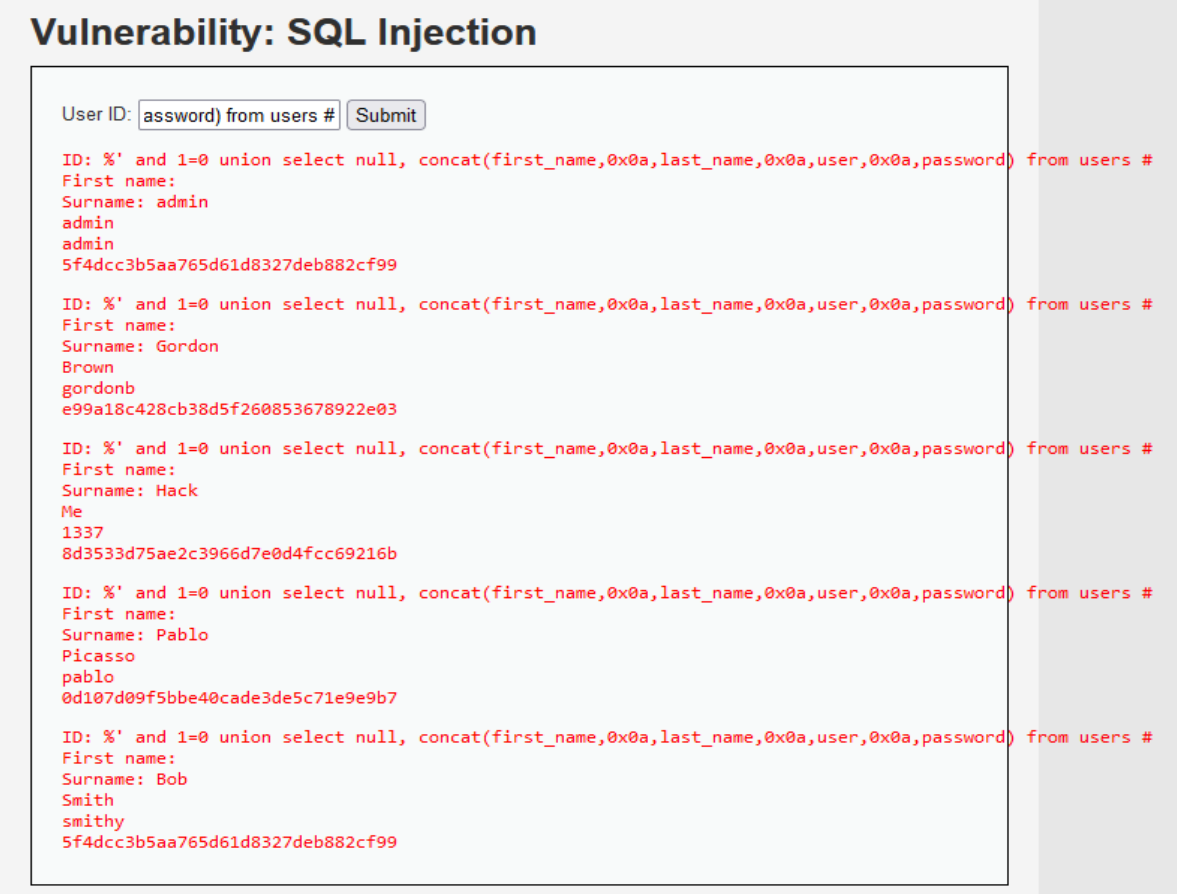
1. **%' or 0=0 union select null, version() #**, pour connaître la version de la base de données sur laquelle s'exécute l'application DVWA.



2. **%' or 0=0 union select null, user() #,** pour afficher l'utilisateur de la base de données qui a exécuté le code PHP alimentant la base de données.



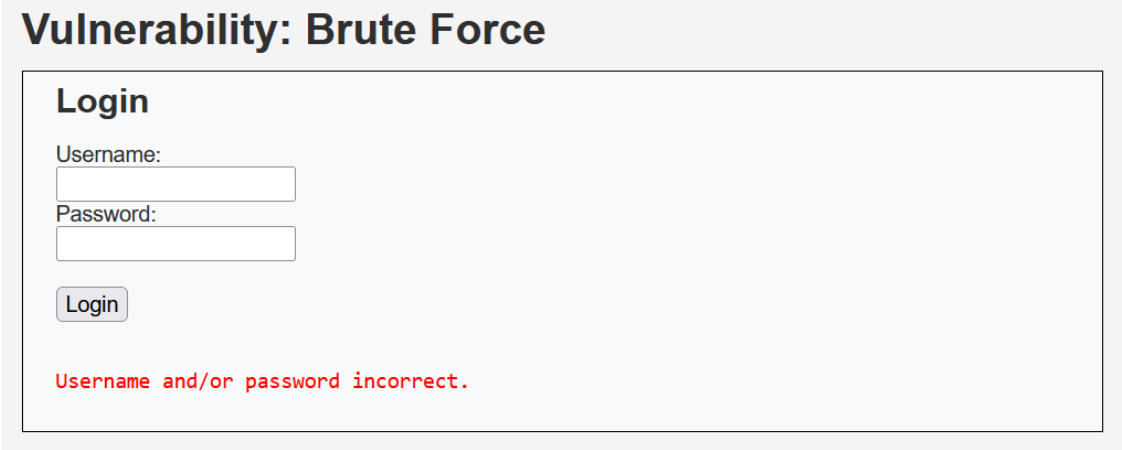
3. **%'and 1=0 union select null, concat(first_name,0x0a,last_name,0x0a,user,0x0a,password) from users #,** pour afficher toutes les informations d'authentification nécessaires présentes dans les colonnes telles qu'elles sont stockées dans l'information_schema.

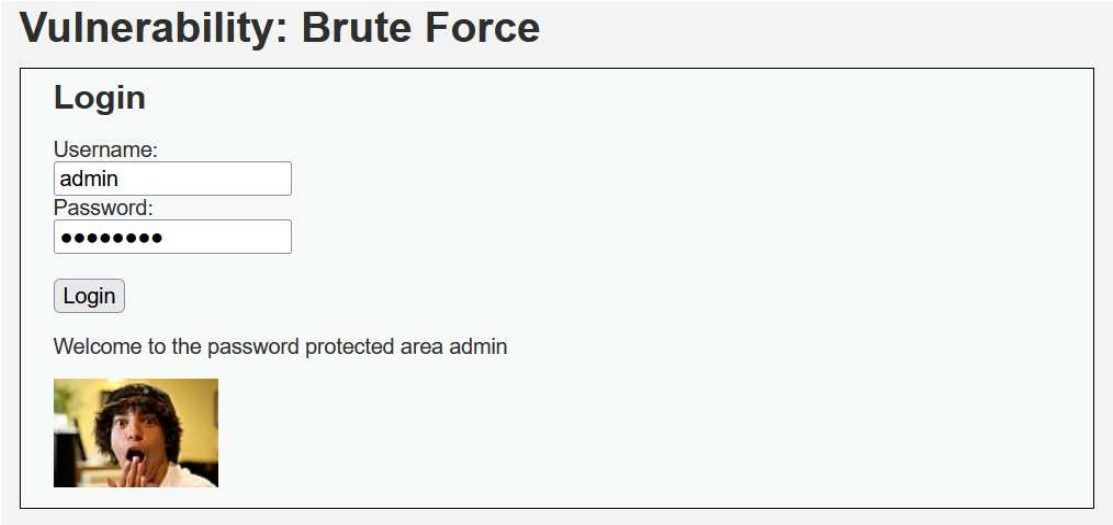


Vulnerability Brute Force: tenter de deviner les mots de passe en envoyant de manière répétée différentes combinaisons possibles de mots de passe jusqu'à ce qu'à trouver celui qui fonctionne.

Elle permet aux utilisateurs de comprendre les risques associés à l'utilisation de mots de passe faibles et l'importance d'implémenter des mesures de sécurité robustes, telles que des politiques de mots de passe forts et des mécanismes de verrouillage de compte après un certain nombre de tentatives infructueuses.

Les captures suivantes illustrent comment un attaquant peut accéder au compte admin et casser son mot de passe après plusieurs tentatives avec le Brute Force.



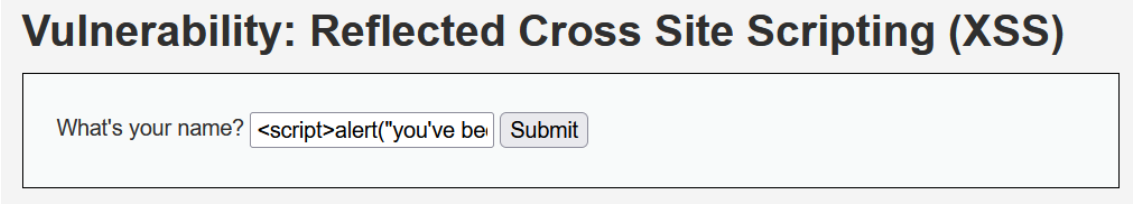


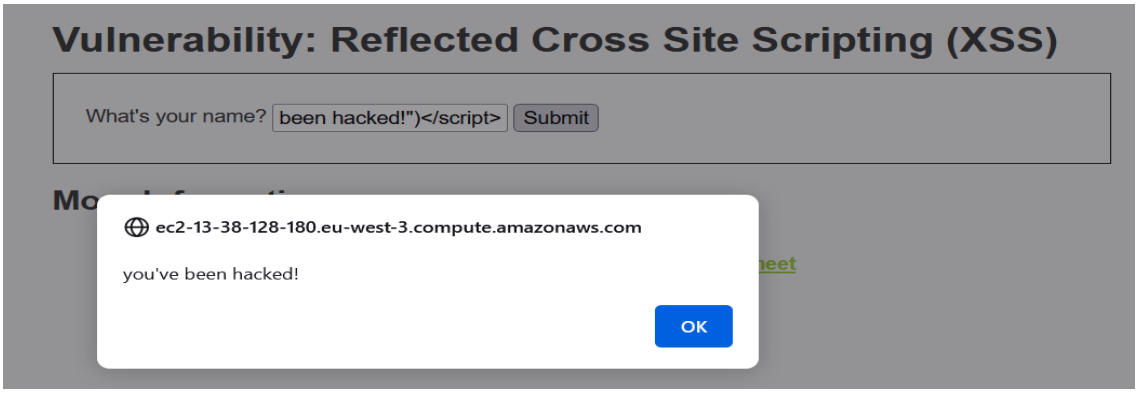
Vulnerability XSS Reflected: injecter et exécuter du code malveillant dans le navigateur d'un utilisateur, l'attaquant peut donc [26]:

- Effectuer toutes les actions possibles dans l'application que l'utilisateur peut effectuer.
- Visualiser toutes les informations accessibles à l'utilisateur.
- Modifier toutes les informations modifiables par l'utilisateur.
- Initier des interactions avec d'autres utilisateurs de l'application, y compris des attaques malveillantes, qui sembleront provenir de l'utilisateur initial victime.

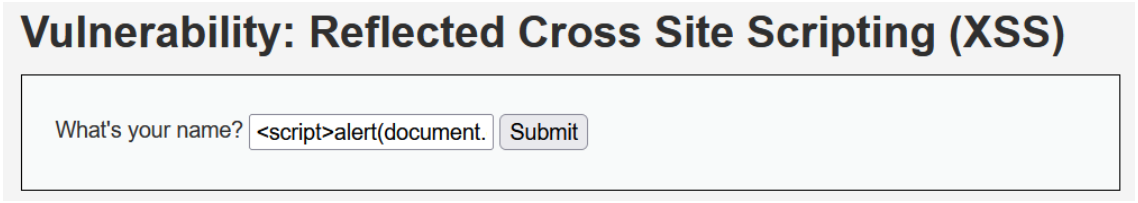
Les captures suivantes montrent comment un attaquant peut injecter et exécuter des scripts à distance avec Reflected Cross Site Scripting:

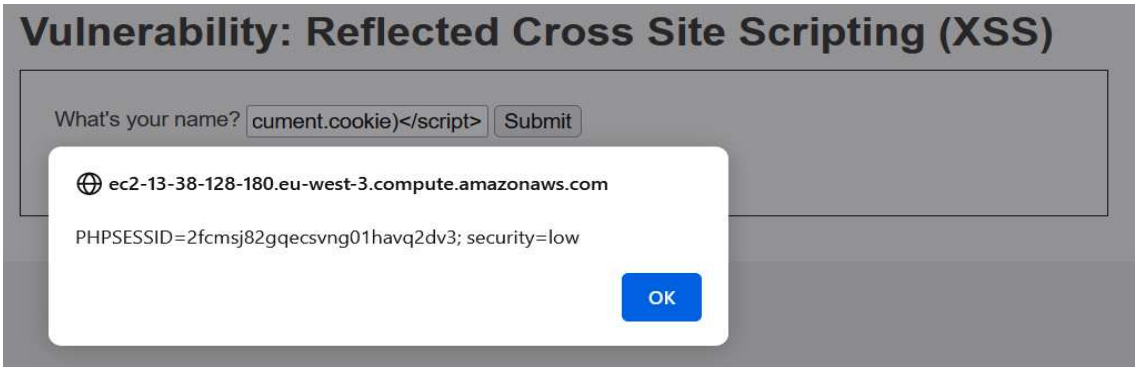
1. **<script>alert("you've been hacked!")</script>**





2. **<script>alert(document.cookie)</script>**



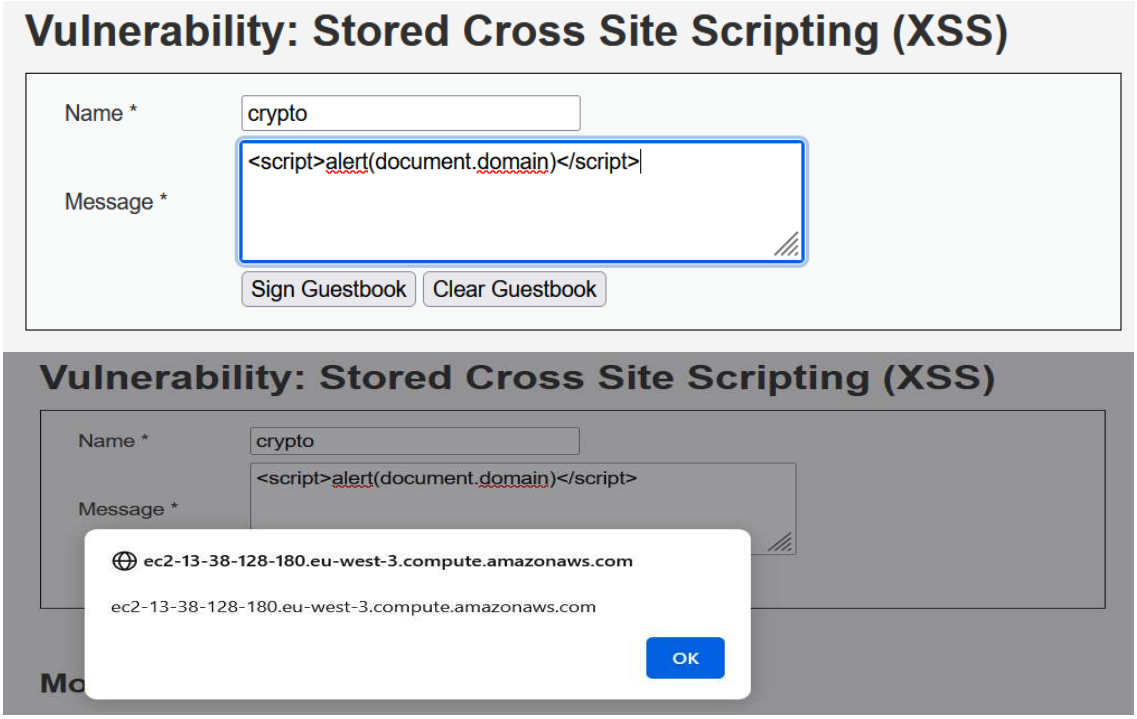


Vulnerability XSS Stored: se produit lorsqu'une application récupère des données d'une source non fiable et les inclut ultérieurement dans ses réponses HTTP de manière non sécurisée.
Les attaques XSS stockées ont un impact important sur les utilisateurs:

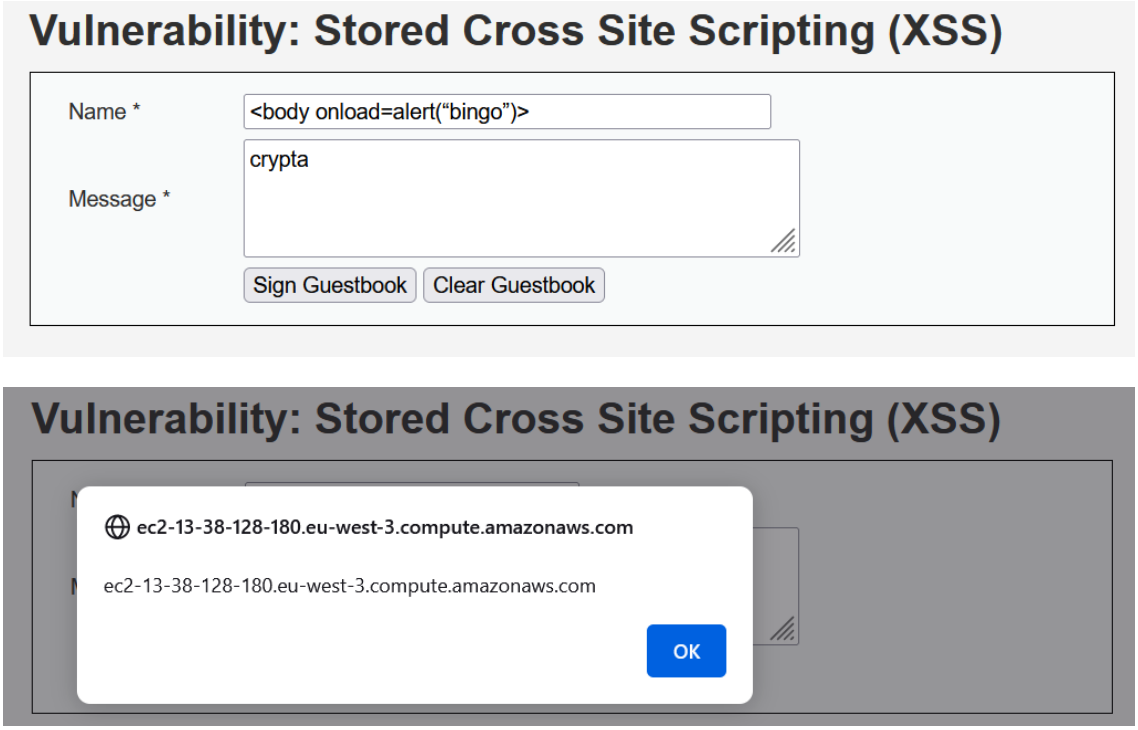
- Si un attaquant peut contrôler un script exécuté dans le navigateur de la victime, il peut compromettre complètement cette personne. Contrairement aux attaques XSS réfléchies, les attaques XSS stockées sont autonomes et ne dépendent pas d'une manipulation externe des utilisateurs.
- L'attaquant insère le script malveillant directement dans l'application et attend que les utilisateurs le rencontrent. Cela peut être particulièrement préoccupant lorsque la vulnérabilité XSS n'affecte que les utilisateurs connectés à l'application, car ils sont garantis d'être compromis lorsqu'ils rencontrent l'exploitation.

Les captures suivantes montrent comment un attaquant peut injecter et exécuter des scripts à distance avec Stored Cross Site Scripting:

1. **<script>alert(document.domain)</script>**



2. **<body onload=alert("bingo")>**



4.5 Protection WAF

Après avoir faire le tour sur les différentes mesures de sécurité d'AWS, telles que la segmentation des subnets et des zones, le VPC, les ACL, la Security Group, l'ALB, la protection SSL, nous allons maintenant nous concentrer sur la protection des contenus grâce au module WAF.

Notre policy WAF propose différents types de règles pour protéger l'application contre les attaques simulées auparavant en basant sur des critères d'inspection définit préalablement selon notre besoin et notre contrainte de coûts.

Nous avons utilisé deux type de règle, les managed rules et les own rules.

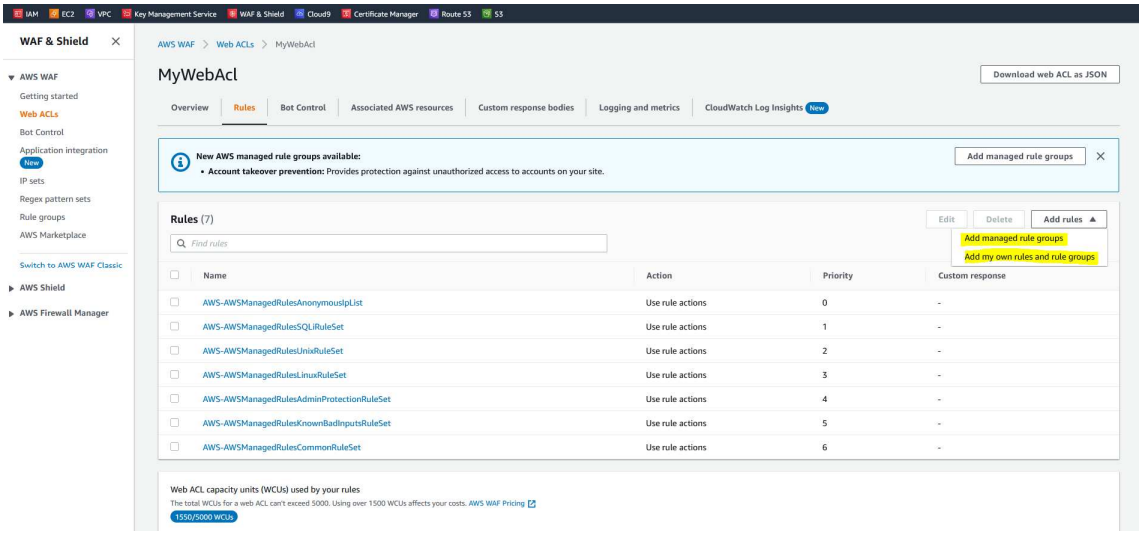


Figure 30: Règles WAF.

Managed rules : nous avons utilisé les règles « Free » préconfigurées fournies par AWS qui sont conçues pour protéger notre serveur web contre des attaques courantes, telles que les injections SQL, les attaques par force brute, les attaques de scripts entre sites (XSS),…elles sont constamment mises à jour par AWS pour inclure de nouvelles règles de protection contre les dernières menaces et vulnérabilités connues.

La figure suivante montre la liste des règles managées activées pour notre politique de sécurité WAF.

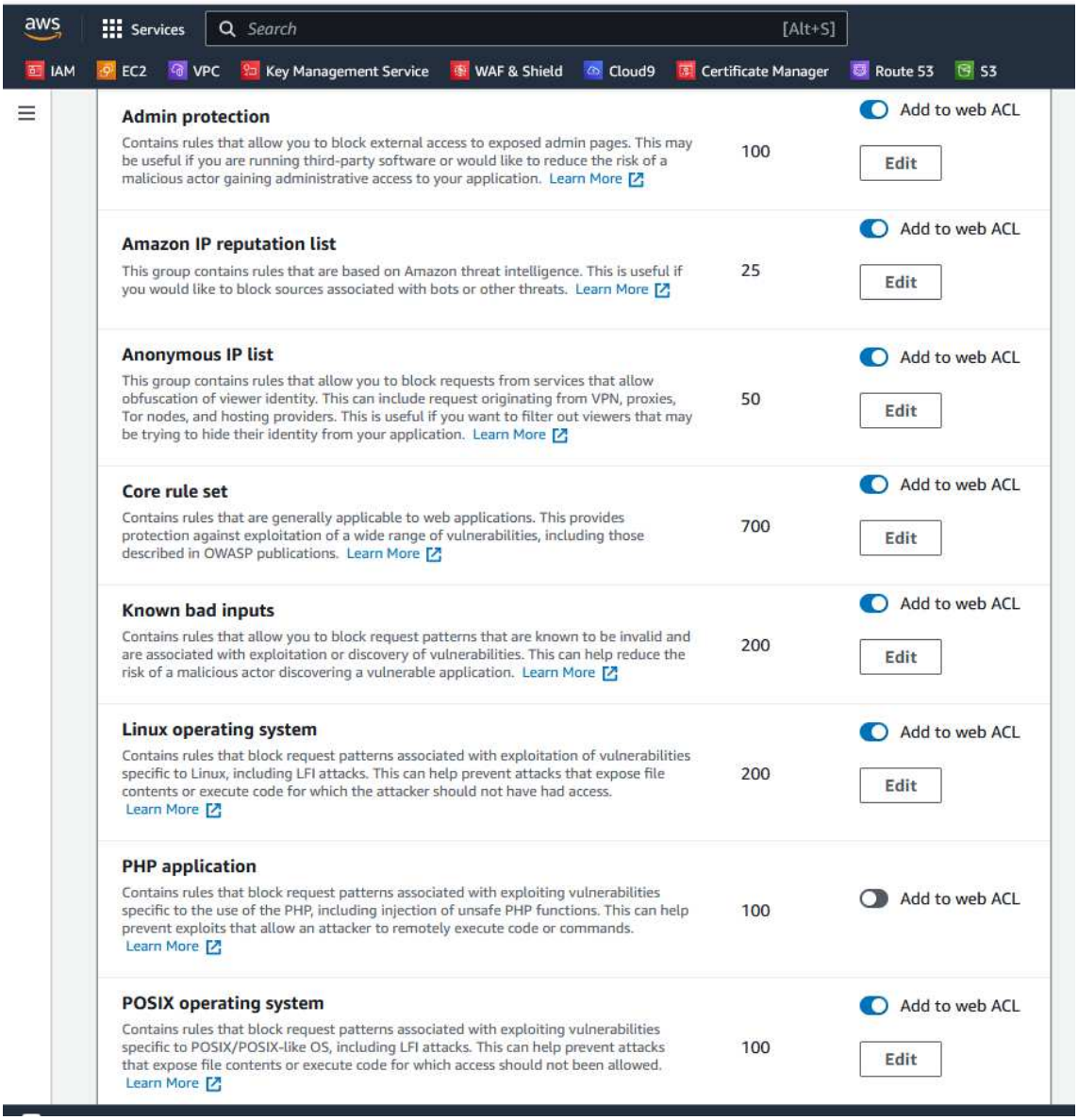


Figure 31: Managed rules.

Own rules: nous avons utilisé aussi nos propres règles personnalisées en basant sur des critères spécifiques pour protéger notre serveur d'application contre les vulnérabilités testées. Cela nous donnera un contrôle plus granulaire de notre politique de sécurité en fonction de nos besoins.

Les figures suivantes montre la liste des règles activées pour notre politique de sécurité WAF.

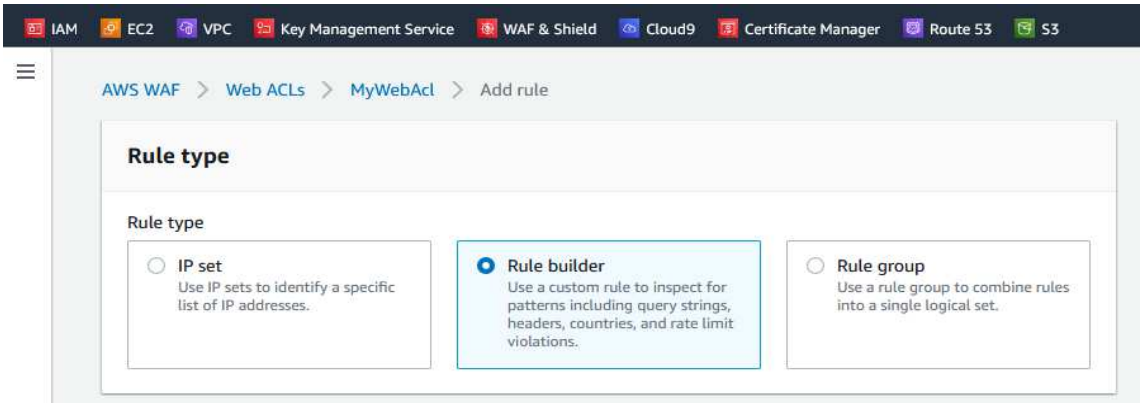


Figure 32: Own rules.

IP set Policy : Autoriser uniquement les requêtes en prévenance des subents de confiance.

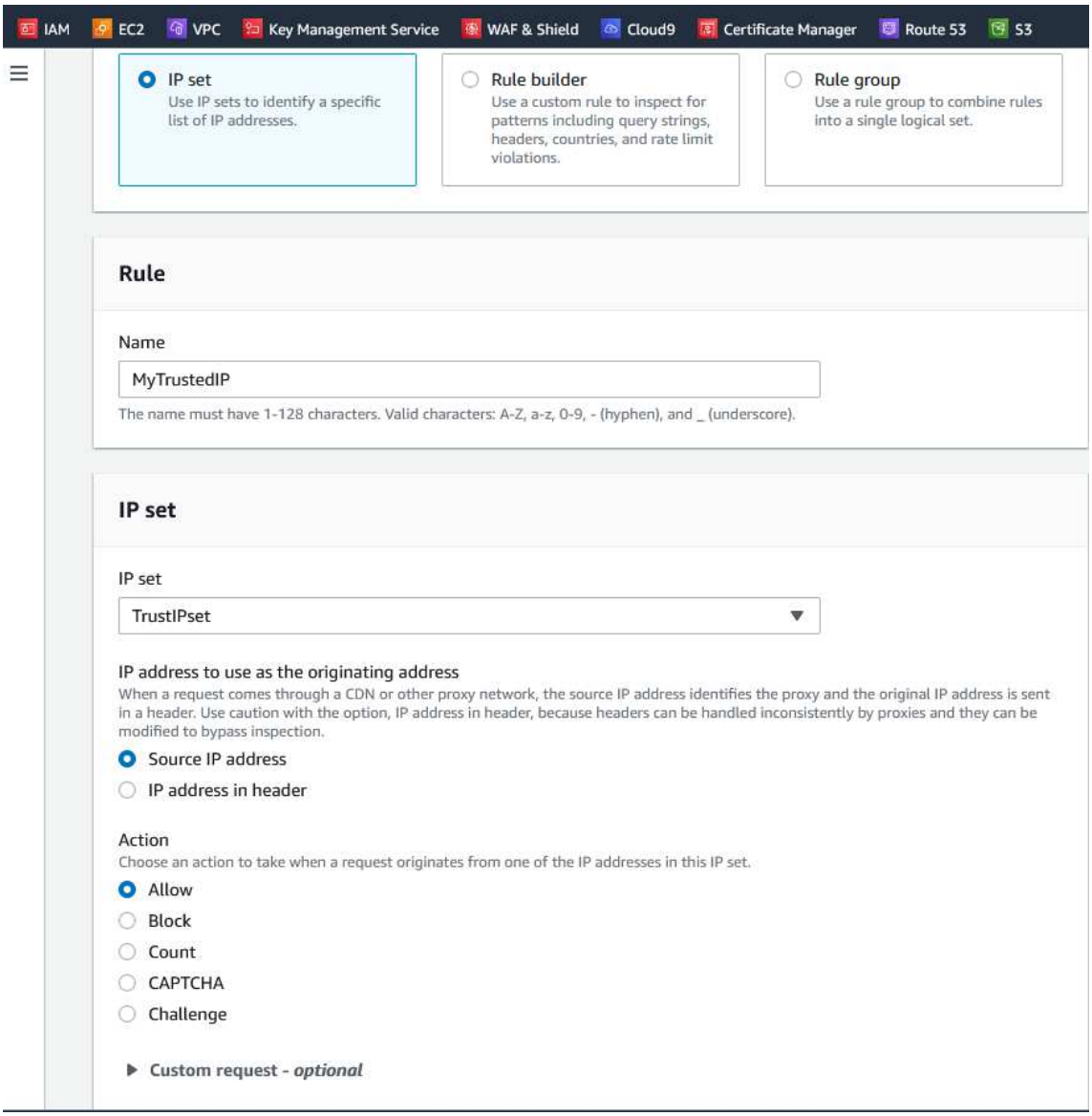


Figure 33: Own rules – IP Set.

Rule builder Policy : Autoriser uniquement les requêtes conformes aux conditions et aux critères définis telles que les en-têtes HTTP, les paramètres/méthodes de requête, les chaînes de caractères,…

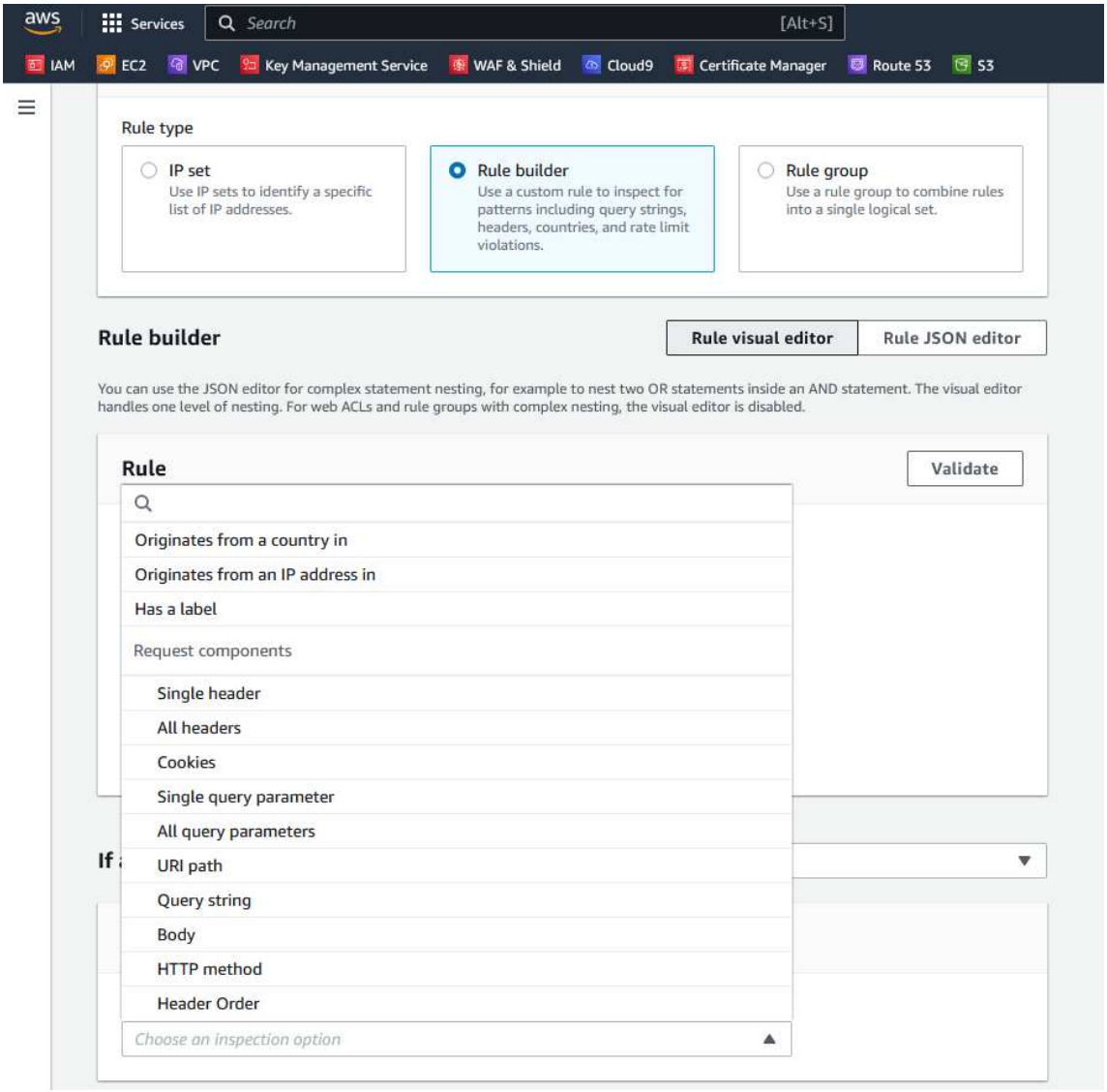
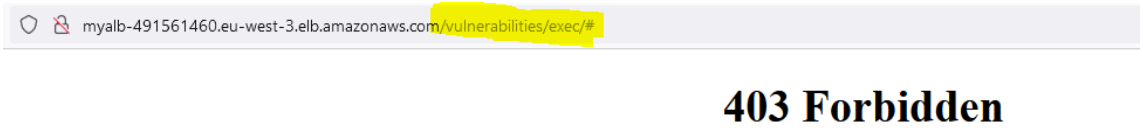


Figure 34: Own rules – Rule builder.

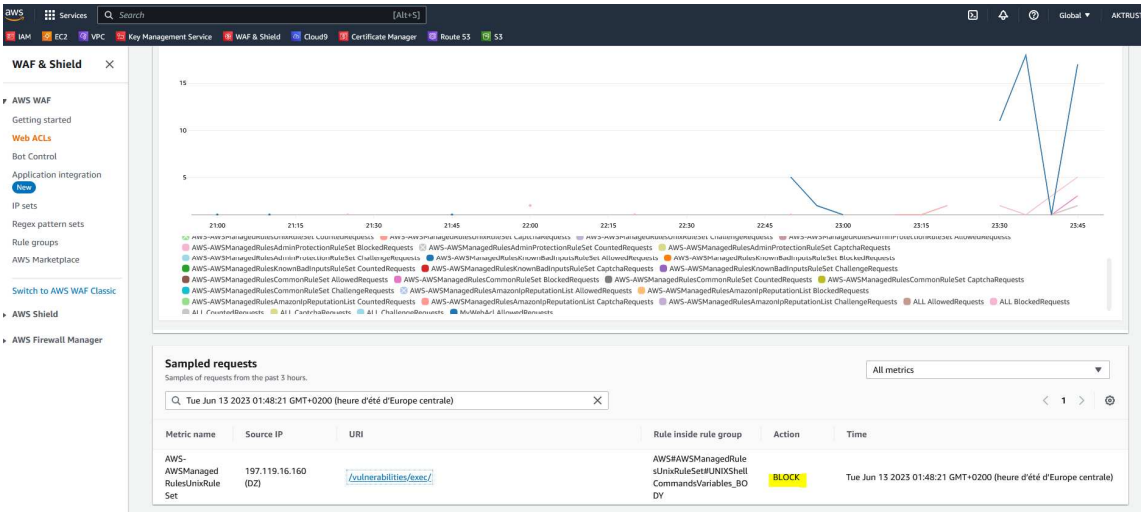
5.4.1 Attaques testées avec WAF:

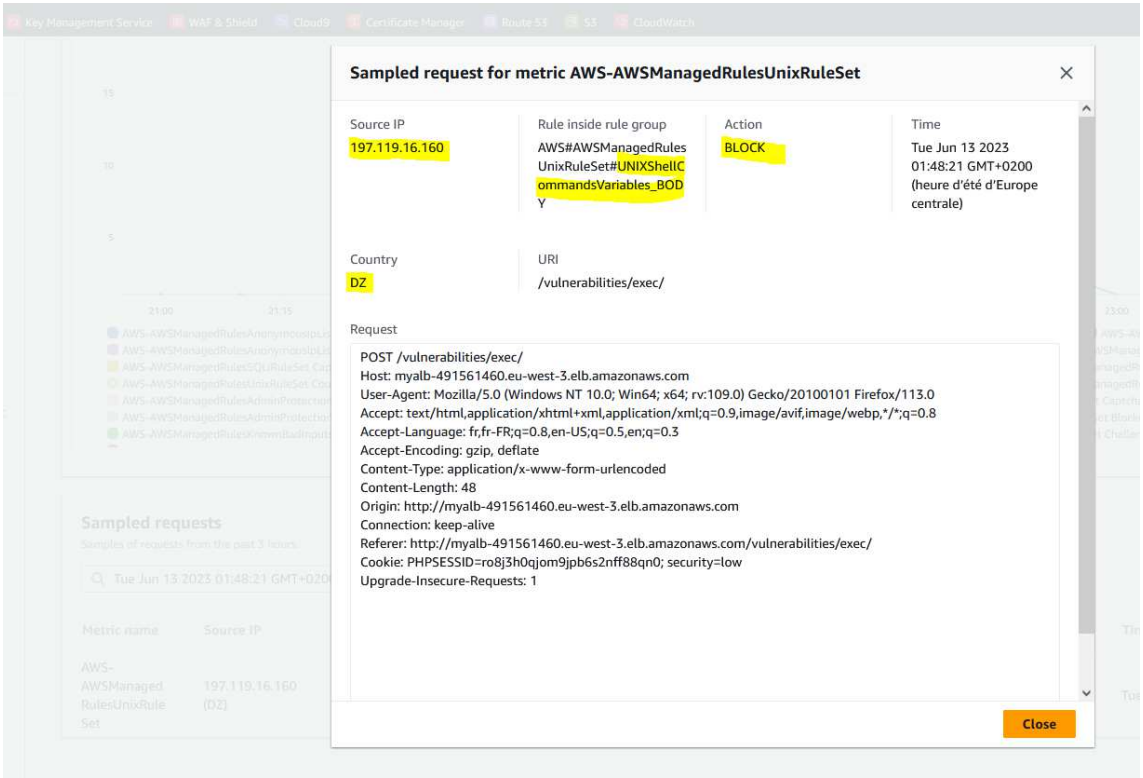
Vulnerability Command Injection: La capture ci-dessous montre bien que le WAF a bien bloqué l'attaquant d'accéder au contenu de **/etc/passwd** à distance avec la remote command injection.



403 Forbidden

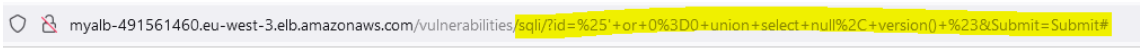
Les captures suivantes montre les logs de blocage au niveau de WAF collecter par CloudWatch et quelque information sur la req IP source, Country, URI et la rule name qui a bloqué le flux.





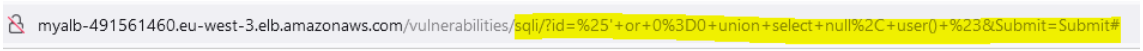
<u>Vulnerability Injection SQL :</u> Les captures suivantes illustrent comment le waf bloque les attaques SQL injection query:

%' or 0=0 union select null, version() #



403 Forbidden

%' or 0=0 union select null, user() #

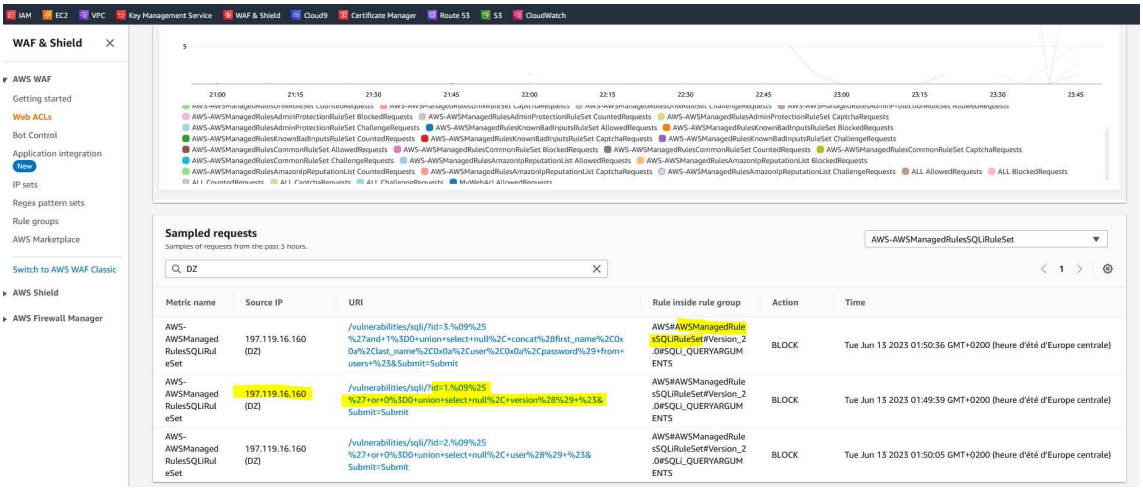


403 Forbidden

%'and 1=0 union select null, concat(first_name,0x0a,last_name,0x0a,user,0x0a,password) from users #,

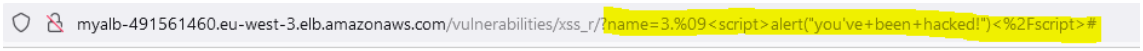
403 Forbidden

Le capture suivante montre les logs de blocage SQL injection sur CloudWatch



Vulnerability XSS Reflected: Les captures suivantes montrent bien que notre policy waf bloque l'exécution des scripts à distance avec Reflected Cross Site Scripting:

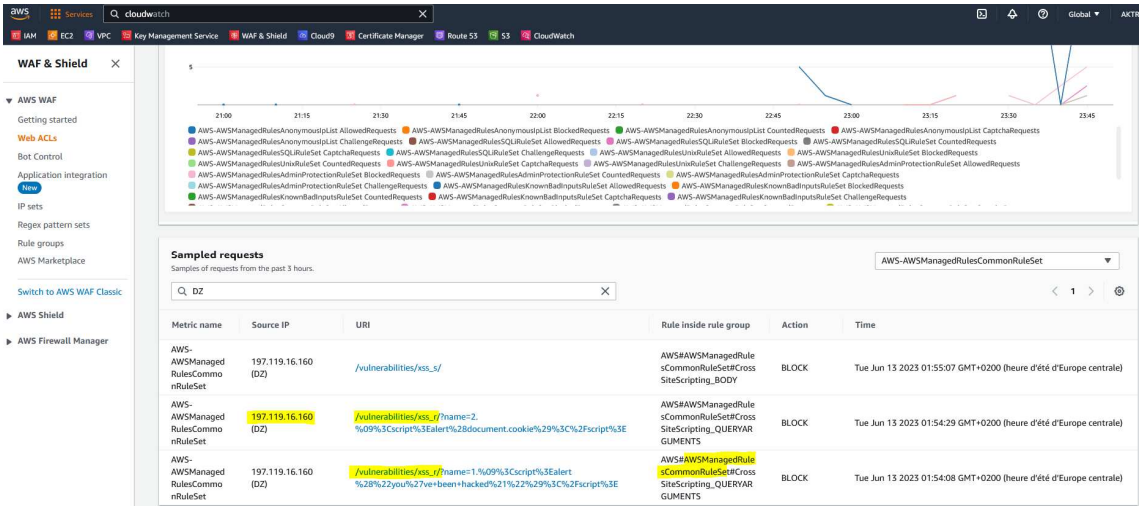
<script>alert("you've been hacked!")</script>



403 Forbidden

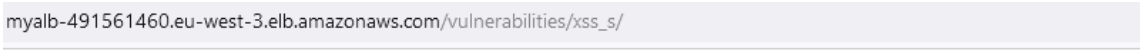
<script>alert(document.cookie)</script>

403 Forbidden

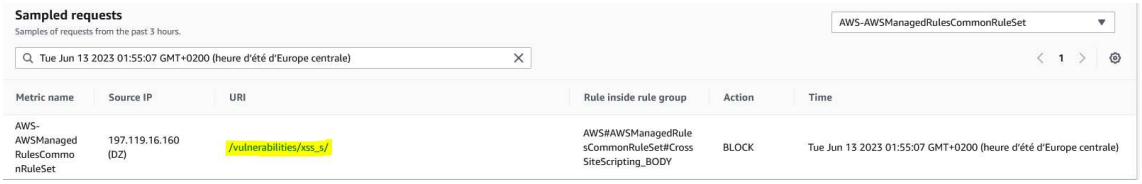


Vulnerability XSS Stored: Les captures suivantes montrent comment le WAF l'exécution des scripts à distance avec Stored Cross Site Scripting:

<script>alert(document.domain)</script>
<body onload=alert("bingo")>



403 Forbidden



4.6 DevOps

En utilisant DevOps AWS, les équipes de développement et d'exploitation peuvent automatiser et rationaliser les tâches liées à la création, au déploiement et à la gestion d'applications dans le cloud AWS. Cela inclut l'utilisation de l'infrastructure en tant que code, l'intégration continue et le déploiement continu, la gestion de la configuration, la surveillance et la journalisation, la scalabilité et la haute disponibilité, ainsi que la sécurité.

Dans notre projet nous avons visé l'aspect collaboration et automatisation afin d'accélérer le déploiement de notre architecture et d'optimiser les coûts d'infrastructure.

4.6.1 AWS Cloud9:

AWS Cloud9 est un environnement de développement intégré (IDE) basé sur le cloud pour développer, collaborer, exécuter et déboguer du code depuis les navigateurs. AWS Cloud9 dispose des éditeurs de code puissants, des débogueurs intégrés et des terminaux préconfigurés avec AWS CLI que nous pouvons le démarrer en quelques minutes, sans avoir à perdre du temps à installer des applications locales ou à configurer des machines de développement.

Avantages:

- Programmer facilement en utilisant simplement un navigateur.
- Collaborer en temps réel lors de la programmation (Lors de la procédure de partage en fournissant le **nom d'utilisateur AWS Identity and Access Management** (IAM) ainsi que les **niveaux d'accès** désirés).
- Créer des applications sans serveur de manière facile.
- Accéder directement à AWS via le terminal.
- Lancer rapidement des projets.

Fonctionnement d'AWS Cloud9:

Le schéma suivant présente un aperçu général du fonctionnement d'AWS Cloud9.
L'IDE AWS Cloud9 s'exécute dans un navigateur web sur l'ordinateur local, pour interagir avec l'environnement AWS Cloud9. Une ressource AWS (instance EC2, serveur web, …) se connecte à cet environnement. Enfin, le travail est stocké dans un référentiel AWS CodeCommit ou un autre tiers.

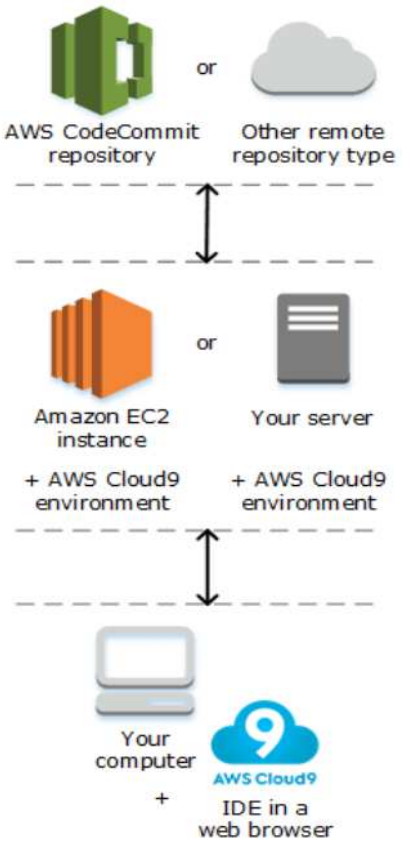


Figure 35: Fonctionnement d'AWS Cloud9 (Source: aws docs)

Donc, nous avons utilisé la solution AWS Cloud9 comme une machine virtuelle de DevOps qui offre plusieurs avantages par rapport à un ordinateur en termes de sécurité, de collaboration et de proximité avec les données.

Tout d'abord, il offre des permissions flexibles, permettant d'attribuer des autorisations basées sur des rôles, facilitant ainsi la communication avec les API et les autres services sans gérer les clés d'accès. Cela résout de nombreux problèmes de sécurité.

Deuxièmement, le trafic réseau est plus rapide car nous travaillons directement à l'intérieur de l'environnement où les données sont situées. Contrairement à une connexion Wi-Fi instable sur l'ordinateur, en utilisant une instance Cloud9, nous communiquons avec les données à l'endroit où elles se trouvent, garantissant une meilleure expérience.

De plus, Cloud9 offre une intégration profonde avec d'autres services cloud et les outils en ligne de commande.

Bonnes pratiques de sécurité pour AWS Cloud9:

Voici quelques bonnes pratiques de sécurité à suivre pour AWS Cloud 9 [27]:

- Stocker le code de manière sécurisée en utilisant un système de contrôle des versions tel que AWS CodeCommit.
- Lors de la configuration des environnements de développement EC2 sur AWS Cloud9, utiliser et configurer des volumes chiffrés Amazon Elastic Block Store.
- Utiliser des identités pour contrôler l'accès au compte AWS Cloud9 dans le cadre des environnements EC2.
- Pour les environnements de développement AWS Cloud9 partagés, suivre les bonnes pratiques de sécurité correspondantes.

4.7 Automatisation:

Terraform HashiCorp:

Terraform HashiCorp est un outil populaire d'infrastructure en tant que code (IaC) permettant d'automatiser l'infrastructure de n'importe quel cloud. Dans notre travail, nous avons utilisé les déploiements terraform pour mettre en place notre infrastructure sur AWS, qui se compose de:

- Une instance AWS Cloud9 en tant qu'IDE basé sur le web.
- Un VPC avec des Public Subnet et Private Subnet.
- Une instance EC2 pour notre serveur d'application dans une zone protégée par le VPC (Private Subnet).
- Une ALB comme ressource protégée par le WAF en frontal de notre application, dans une zone publique.
- Un WAF qui protège notre serveur d'application.
- Des groupes de sécurité pour gérer le flux entrant et sortant.
- Une backet S3 pour le stockage des logs.
- Une NAT Gateway pour que notre instance dans le subnet privé puisse se connecter à Internet (mais elle ne peut pas recevoir un flux entrant depuis Internet).

AWS Cloud9 est notre point central pour déployer du code Terraform, et il s'intègre bien avec les modules Terraform.

La figure ci-dessous montre les solutions testées de déploiements massif de l'infrastructures AWS avec Terraform depuis Cloud9, VMware workstation (VM d'administration ubuntu) ou vsCode sur un PC.

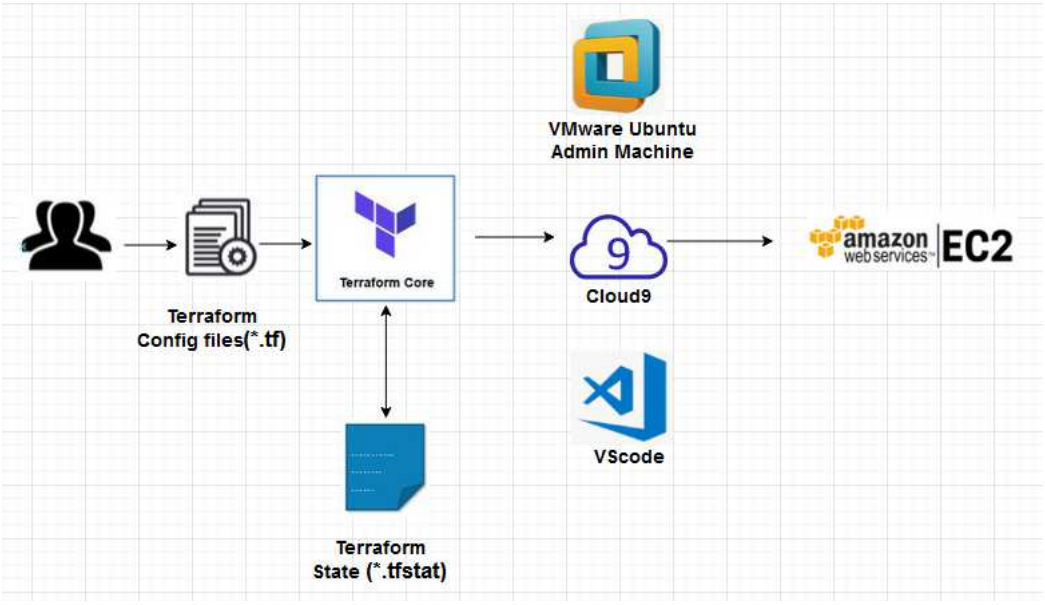
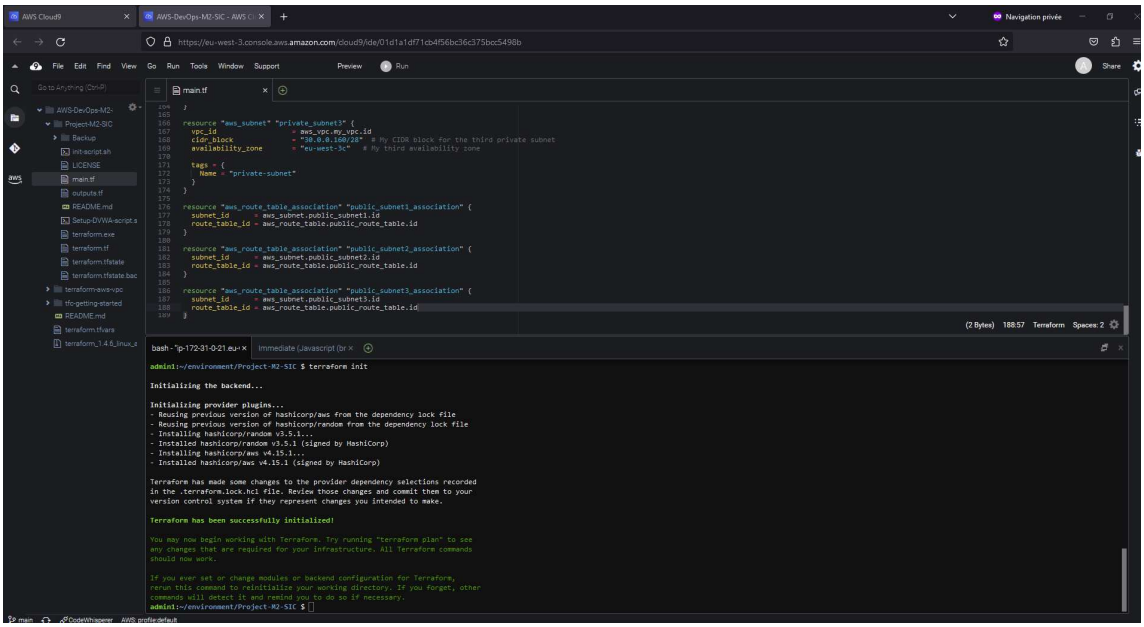


Figure 36: Intégration Terraform avec Cloud9.

Ci-dessous les captures d'écrans du processus de déploiement via Terraform.

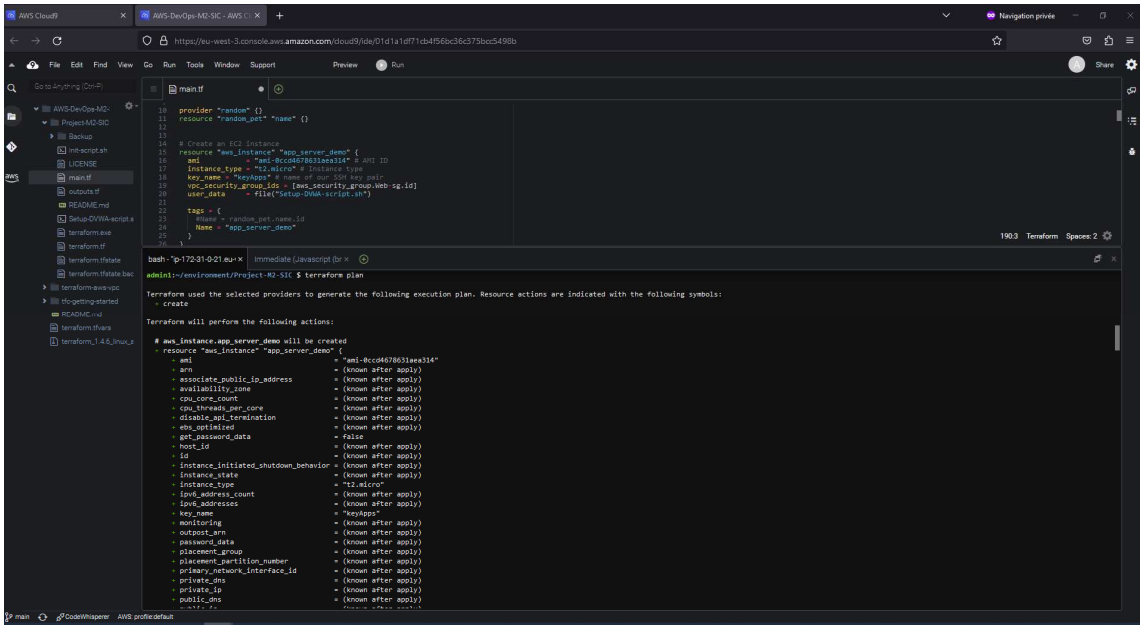
Terraform init : Initialiser le projet Terraform dans notre workplace Projet-M2-SIC.

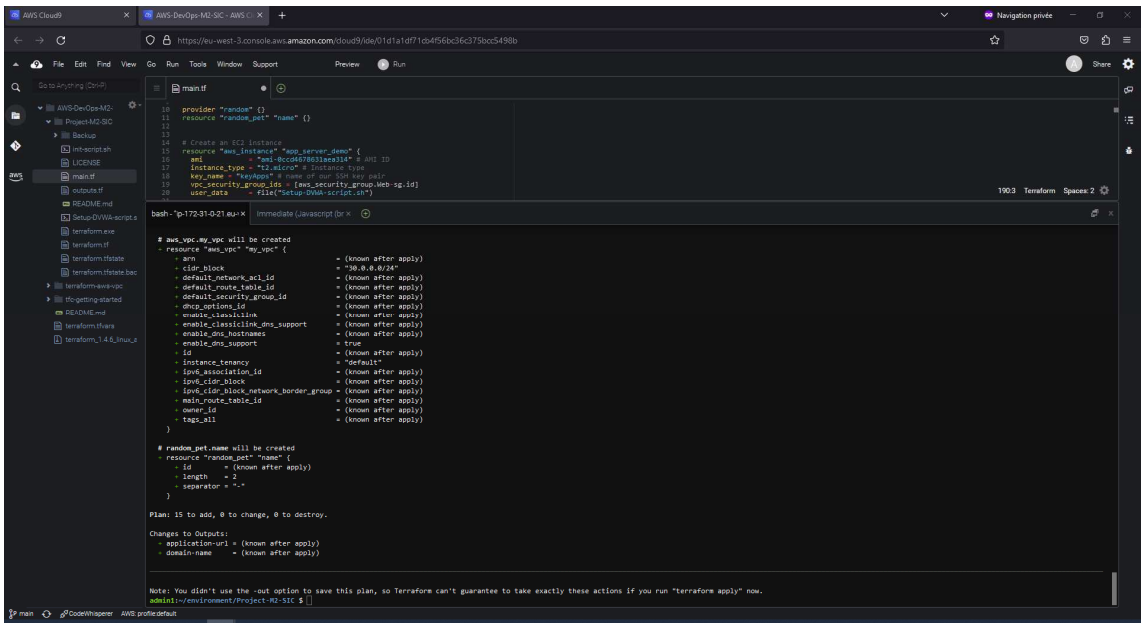


Terraform a été initialisé avec succès, donc nous pouvons lancer Terraform Plan**.**

Terraform Plan:

- Créer un plan d'exécution des modifications de l'infrastructure
- Analyser les fichiers de configuration Terraform
- Comparer l'état actuel de l'infrastructure avec la configuration proposée.
- Afficher les actions prévues, telles que la création, la mise à jour ou la suppression de ressources, ainsi que les éventuelles erreurs ou conflits. Cela permet aux utilisateurs de visualiser les changements qui seront appliqués avant de les exécuter avec la commande "terraform apply".

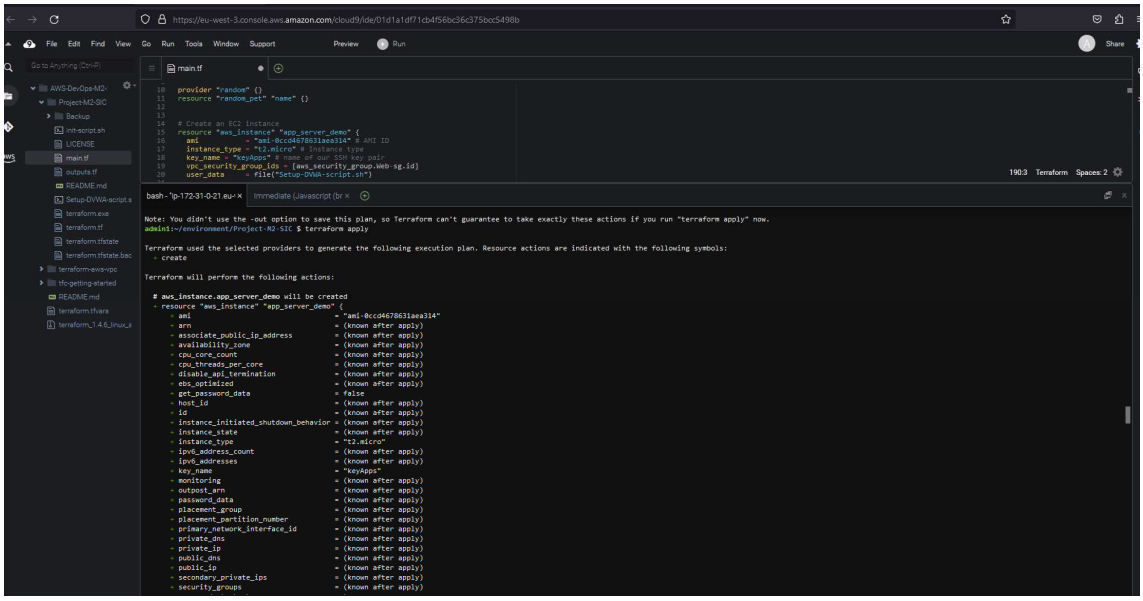


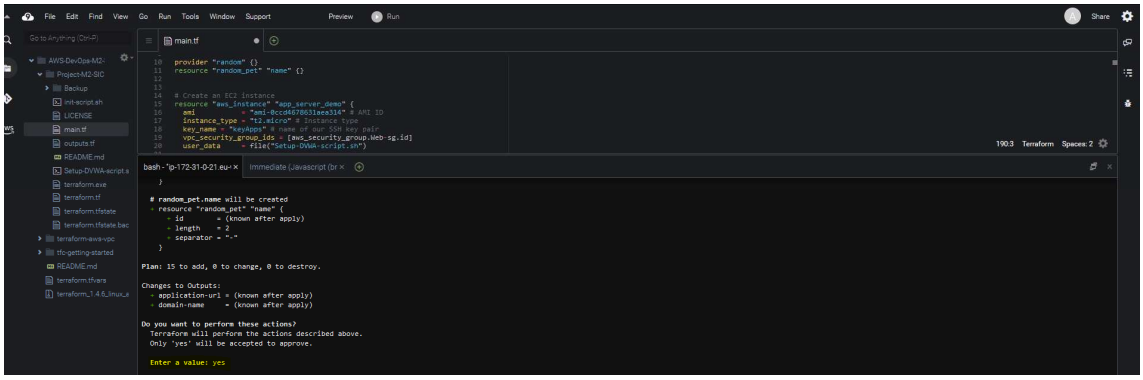


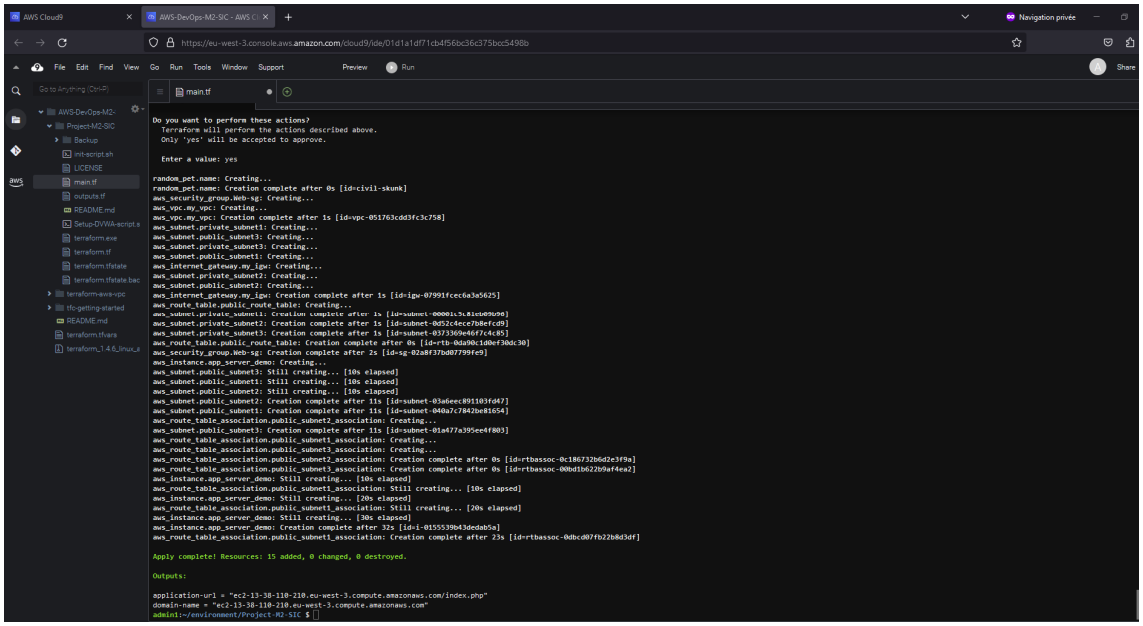
Terraform apply :

- Analyser les fichiers de configuration, y compris tous les fichiers .tf
- Déterminer les actions à prendre pour aligner l'état actuel de l'infrastructure avec l'état souhaité défini dans les fichiers de configuration.
- Re-afficher le plan d'exécution détaillant les modifications apportées à l'infrastructure.
- Demander la confirmation du plan en saisissant yes pour permettre à Terraform d'appliquer les modifications proposées.
- Exécuter les actions requises pour mettre en place ou mettre à jour les ressources conformément au plan.

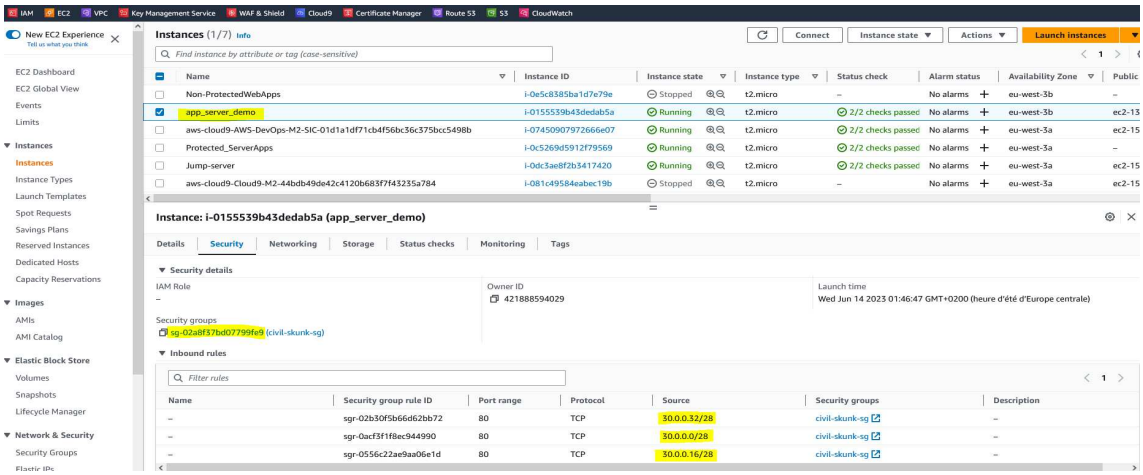
L'exécution de `terraform apply` est un processus potentiellement destructif, car cela peut créer, modifier ou supprimer des ressources d'infrastructure. Par conséquent, il est important de comprendre les conséquences des modifications prévues avant de confirmer l'application.

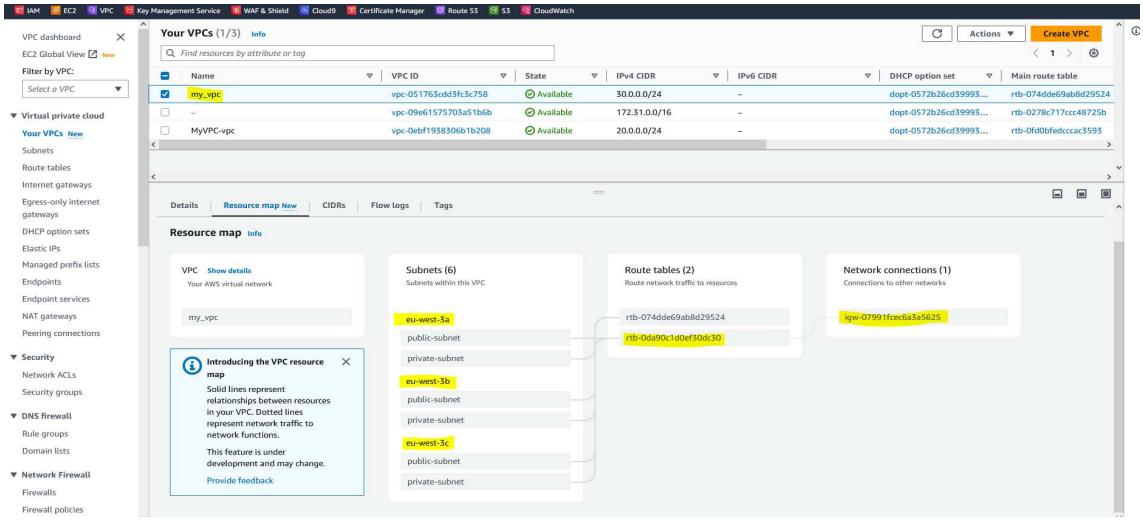


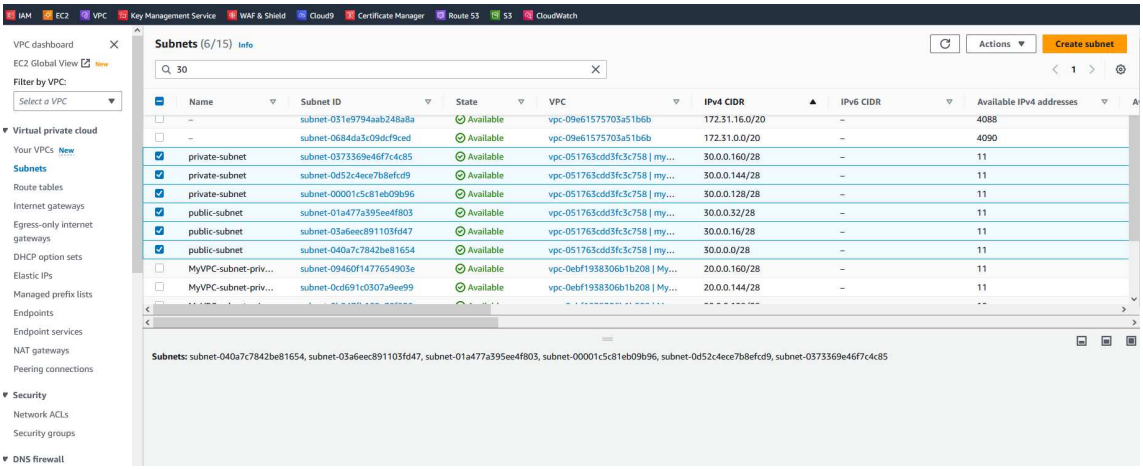




Les captures ci-dessous montrent bien que l'infrastructure est bien créée par Terraform.

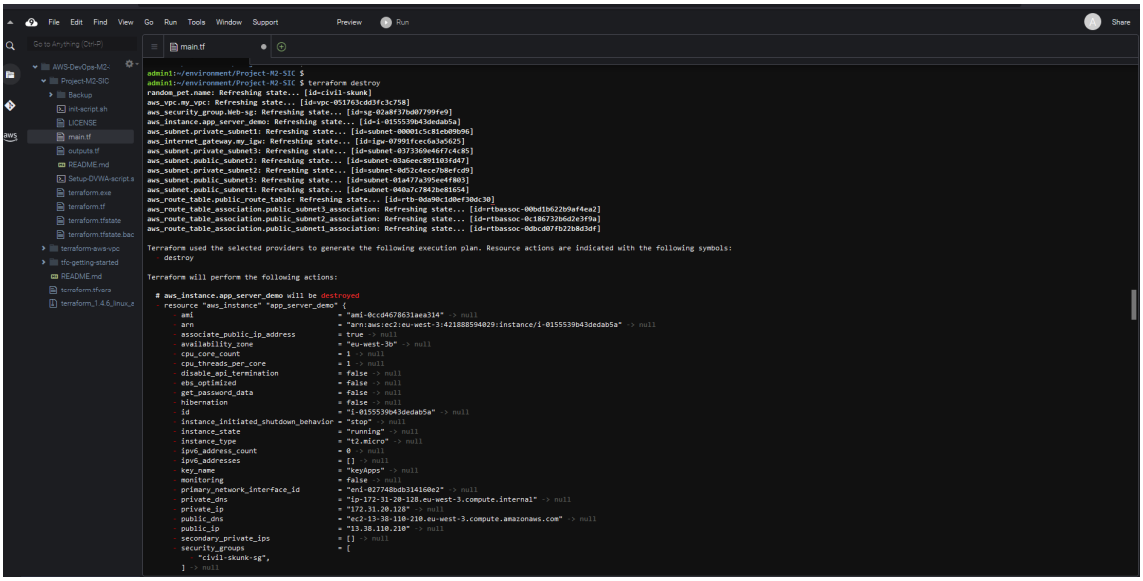


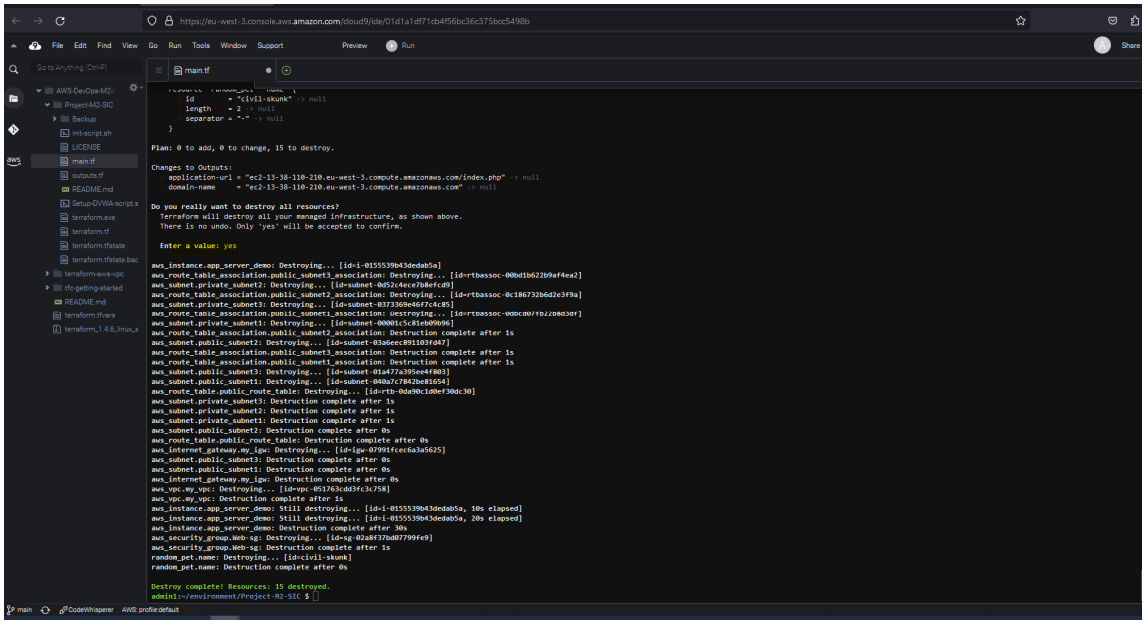




Terraform destroy : supprimer toutes les ressources d'infrastructure créées à l'aide de Terraform. Lors de l'exécution de la commande destroy, terraform:

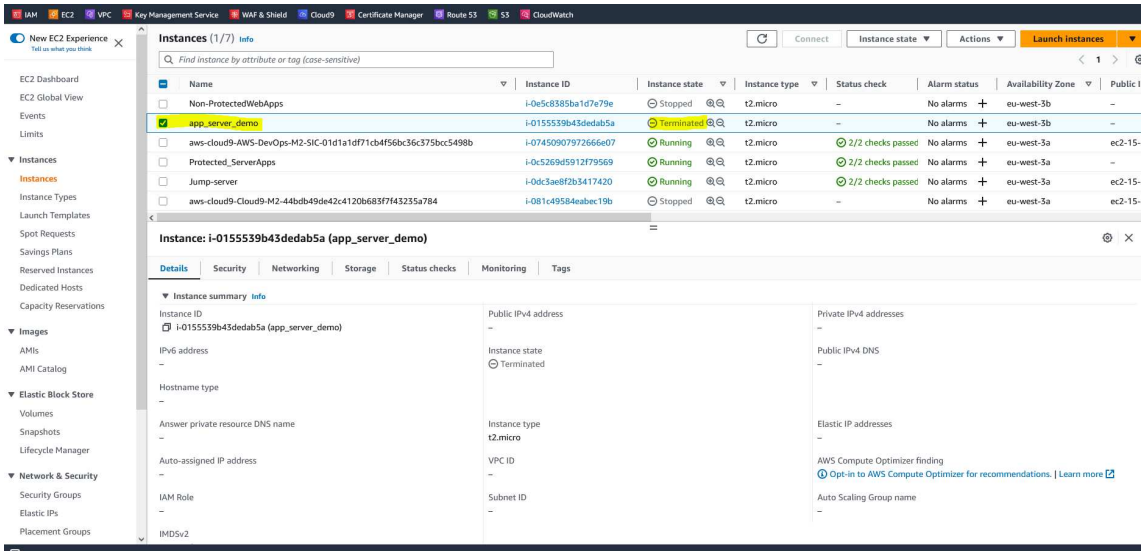
- Analyse l'état actuel de l'infrastructure,
- Identifie les ressources qui ont été créées à l'aide de Terraform,
- Supprime en inversant les actions précédemment exécutées par terraform apply.





Les captures ci-dessous montrent bien que l'infrastructure a été bien détruite par Terraform destroy.

L'instance ec2 app_server_demo passe en état Terminated et elle disparaitra ultérieurement.



4.8 Conclusion:

Dans ce chapitre nous avons proposé et déployé une architecture avec plusieurs niveaux de sécurité, la testé et l'automatiser. Ce travail nous a permis de rentrer dans le domaine d'engineering sécu, de DevOps et d'enrichir nos connaissances afin de progresser dans nos futures aventures professionnelles.

Conclusion générale

La sécurité dans le cloud est devenue une préoccupation majeure des entreprises et d'organisations qui adoptent des solutions basées sur le cloud. Les avantages du cloud, tels que la flexibilité, l'évolutivité et la facilité d'accès, sont incontestables, mais ils sont accompagnés de défis en matière de sécurité des données. Il est primordial de mettre en place des mesures robustes pour protéger les informations sensibles et prévenir les violations de sécurité. Cela nécessite une combinaison de bonnes pratiques de sécurité, de technologies de pointe et d'une collaboration étroite entre les acteurs du domaine et les clients finaux.

Notre objectif a été de mettre en place une infrastructure cloud sécurisée qui répond aux enjeux majeurs et aux principaux défis liés aux menaces et à la perte de données.

Dans un premier temps, nous avons présenté les principales clés, les spécificités et les challenges liés aux cloud Computing. En outre, nous avons également présenté la solution cloud public AWS en mettant l'accent sur l'acquisition des connaissances fondamentales nécessaires pour comprendre les mécanismes avancés de sécurisation des données dans les infrastructures cloud AWS.

En outre, nous avons présenté en détail la solution de cloud public AWS, en mettant l'accent sur l'acquisition des connaissances fondamentales nécessaires pour comprendre les mécanismes avancés de sécurisation des données dans les infrastructures cloud AWS.

En analysant l'architecture mise en place, les mesures concrètes prises et les tests effectués, il a été conclu que la solution AWS offre des services qui garantissent un niveau de sécurité très élevé, avec plusieurs couches de protection.

Ainsi, il est essentiel de mettre en place un cahier des charges précis lors de l'élaboration des appels d'offres pour les solutions cloud, en accordant une attention particulière aux aspects de sécurité. Il est également recommandé de procéder à des phases de workshop pour garantir la pertinence de notre choix de technologie cloud.

A l'avenir nous souhaitons de s'orienter vers le DevOps et la CyberSécurité dans le domaine du cloud, en raison du potentiel et des opportunités qu'offre ce marché dans les pays émergents.

Références bibliographiques:

[1]: Fundamentals of Cloud Computing, Cloud Computing-The basics, History of Cloud Computing, David Davis.

[2]: The NIST Definition of Cloud. Peter Mell, Timothy Grance. 2011.

[3]: Cloud Concepts, Architecture & Design for CCSP, Cloud Service Models,Kevin Henry.

[4]: Cloud Computing. Yumin Danny Z, et al. 2002.

[5] Optimization of Security as an Enabler for Cloud Services and Applications.Deshpande, Varun M, et al. 2018

[7]: TANIA MARTIN, Cryptographie & Cloud Computing - ÉTAT DE L'ART, Samls, 10/2015, Conditions d'analyse dans le cloud, Architecture de Cloud.

[8]: Kevin Henry, Cloud Data Security for CCSP, Cloud Data Security Concepts, Data owner role.

[9]: Kevin Henry, Cloud Data Security for CCSP, Cloud Data Security Concepts,Security Responsibilities.

[10]: Kevin Henry, Cloud Data Security for CCSP, Cloud Data Security Concepts, The cloud data lifecycle.

[11]: Microsoft Trustworthy Computing, Data classification for cloud readiness, 2014.

[12]: Abdelkader YOUSSEFI, Université Mohammed V de Rabat– UM5R, SECURISATION IP SEC DES ECHANGES DE DONNEES POUR UN ENVIRONNEMENT CLOUD COMPUTING: ARCHITECTURES ET PARAMETRAGE, Classification de données et paramétrage du protocole IP Sec, Classification de l'information.

[13]: Kevin Henry, Cloud Data Security for CCSP, Data Security Technologies, Cloud data protection, DLP.

[14]: Kevin Henry, Cloud Data Security for CCSP, Data Security Technologies, Cloud data protection, DRM/IRM.

[15]: Kevin Henry, Cloud Data Security for CCSP, Data Security Technologies, Encryption & Key Management, Who holds the key.

[16]: Abdelkader YOUSSEFI, SECURISATION IP SEC DES ECHANGES DE DONNEES POUR UN ENVIRONNEMENT CLOUD COMPUTING: ARCHITECTURES ET PARAMETRAGE, Université Mohammed V de Rabat– UM5R, LA CRYPTOGRAPHIE : PRINCIPES, TECHNIQUES ET RÉGLEMENTATION, les fonctions de hachages.

[17]: TANIA MARTIN, Cryptographie & Cloud Computing - ÉTAT DE L'ART, Samls, 10/2015, Stockage protégé dans le cloud,Chiffrement classique.

[18]: TANIA MARTIN, Cryptographie & Cloud Computing - ÉTAT DE L'ART, Samls, 10/2015, Stockage protégé dans le cloud, Cloud Security Gateway.

[19]: Kevin Henry, Cloud Data Security for CCSP, Data Security Technologies, Cloud Data Event Management.

[20]: AWS, Magic Quadrant 2022 pour l'infrastructure cloud et les services de plateforme par Gartner, Rapport d'analyse.

[21]: AWS Identity and Access Management User Guide, HOW IAM WORKS?

[22]: Ryan Kroonenberg, AWS community hero & Alexa champion, Founder of a Cloud GURU, EC2 101 Elastic Compute Cloud, EC2 Definition 2019.

[23]: Ryan Kroonenburg, AWS COMMUNITY HERO & ALEXA CHAMPION, Founder of CLOUD GURU, Identity access management S3, S3 guarantees, 2019.

[24]: Ryan Kroonenburg, Introduction to VPCs, AWS Community here & Alexa Champion, Founder of a Cloud GURU 2019.

[25]: AWS website, documentation, AWS WAF, Fonctionnement d'AWS WAF.

[26]: Port Swigger web site, Cross Site Scripting (Reflected), impacts of reflected XSS attacks.

[27]: AWS web site, documentation, cloud9, Bonnes pratiques de sécurité pour AWS Cloud9.