Consideraciones de la protección del dato en nube

Abstract: los grandes proveedores de tecnología de nube tienen el objetivo de conseguir generar la confianza suficiente en las funcionalidades que implementan. De hecho, tienen cada vez más presente la protección y la soberanía del dato mediante un aislamiento adecuado con independencia del tratamiento (reposo, tránsito y en uso). En definitiva, el propósito es que los servicios en la nube soporten arquitecturas que equilibren el compromiso entre seguridad, cumplimiento de la legislación, protección del dato, soberanía del dato, funcionalidad, control, innovación, ofreciendo al usuario flexibilidad de elección.

1. PROPUESTA TECNOLÓGICA DE NUBE HIPERESCALA

1.1 Introducción

Los organismos públicos necesitan asegurar un nivel de protección adecuado para sus datos y cargas de trabajo en la nube. Los proveedores de soluciones en la nube (CSP) brindan la posibilidad de que sus clientes utilicen la nube pública y aunque se es consciente de que existen datos que difícilmente abandonarán los centros de datos de cliente, existen otros datos que con las salvaguardas adecuadas pueden contar con los beneficios de usar la nube pública.

En este sentido, algunos proveedores detallan las siguientes categorías de datos:

- **Datos públicos**: son los datos públicamente accesibles. Típicamente pueden almacenarse en infraestructuras de nube hiperescala sin ningún tipo de riesgo, ya que dicha infraestructura de nube proporciona los niveles de seguridad básicos exigidos y los datos no entrañan ningún riesgo.
- **Datos sensibles**: son datos que requieren de mecanismos de protección adicionales a los básicos requeridos para los datos públicos. Estos mecanismos de protección adicionales pueden ir desde cifrado de la información en reposo gestionado por el proveedor o el cliente hasta mecanismos de protección de datos en uso. La premisa es que se sigue pudiendo garantizar dicha protección con infraestructura de nube hiperescala pública.

- **Datos confidenciales**: son aquellos datos que contienen información crítica (incluso información de estado). Estos datos están sujetos a extensas protecciones, incluyendo la completa desconexión de internet.

El reto de los proveedores es soportar arquitecturas que balanceen el compromiso entre la seguridad, cumplimiento, funcionalidad, control, innovación y regulación preservando la flexibilidad de elección. A tal efecto, se busca cubrir las siguientes necesidades desde una nube hiperescala:

- **Ciberseguridad**: maximizando la protección de las cargas de trabajo y tratando la seguridad de forma global e integral, aprovechando las capacidades de IA/ML, volumen global de señales y arquitecturas Zero Trust.
- **Escalabilidad**: permitiendo utilizar lo que necesitas, cuando lo necesitas. Eliminando los costes incurridos al realizar compras para cargas máximas con capacidad sin uso durante largos períodos de tiempo y permitiendo expandir bajo demanda dicha infraestructura durante los picos inesperados.
- **Disponibilidad y resiliencia**: seguridad sabiendo que tus datos, servicios, y cargas de trabajo están replicados en múltiples infraestructuras dentro de la región escogida con los máximos SLA y gracias a sus diseños redundantes ante fallos.
- **Agilidad**: permitiendo reaccionar a las necesidades, levantando nuevas cargas de trabajo en el menor tiempo posible, experimentando y reduciendo los costes de fallo.
- **Innovación**: permitiendo obtener el beneficio de forma inmediata utilizando las nuevas tecnologías y eliminando los tiempos de contratación, instalación y configuración.
- **Gobernanza**: permitiendo dar visibilidad y control de todo el entorno de computación usando las mejores tecnologías y herramientas.
- **Sostenibilidad**: dotando de tecnologías sostenibles que garanticen por un lado la innovación y por otro el mínimo impacto medioambiental con PUE inferiores a los que el mercado demanda con las nuevas generaciones de DC.

1.2 Objetivos

Por lo tanto, una infraestructura de nube hiperescala debe tener como objetivo, **permitir a los clientes disfrutar de la innovación, la agilidad y la escalabilidad de la nube, pero siempre asegurando y preservando la privacidad y soberanía del dato, la flexibilidad de elección de modelo tecnológico, y la complementariedad al modelo de inversiones existente del cliente**.

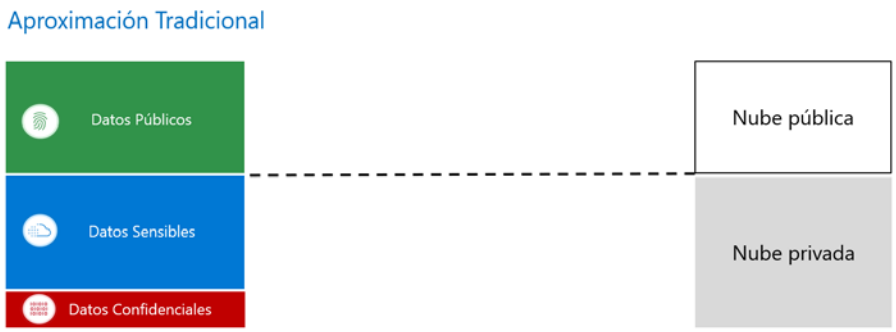
Para ello, el proveedor garantiza la privacidad de los datos de sus clientes a través de cuatro (4) principios de privacidad (el cliente controla sus datos, el cliente escoge

donde almacenar sus datos, el proveedor protege los datos de sus clientes y el proveedor defiende los datos de sus clientes), pero además lo hace a través de una aproximación tecnológica basada en **nube pública cifrada, nube sensible y nube privada.**

1.3 Solución tecnológica

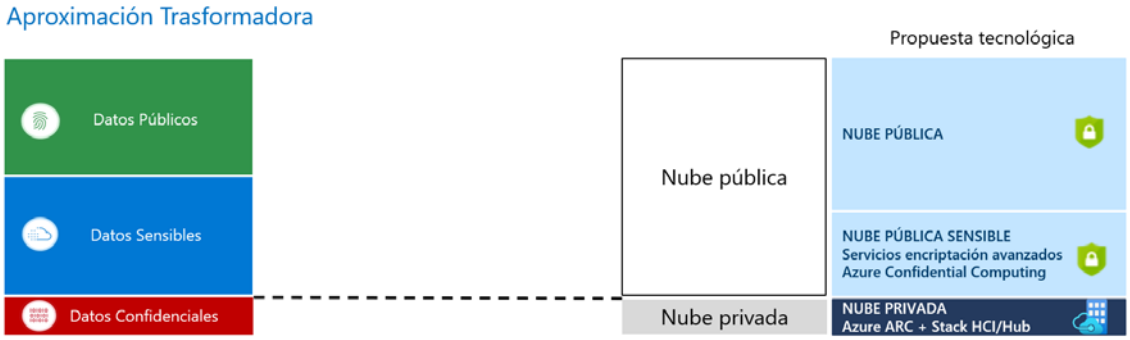
Nota Importante: aunque generalmente sustentadas en una base común, las tecnologías que tratan la protección de los datos en la nube pueden ofrecer variantes, desarrolladas y/o adoptadas por los distintos proveedores. El presente documento contempla exclusivamente la aproximación de Microsoft a dicha problemática.

La aproximación tradicional a los requisitos de seguridad de los datos es como la que se muestra a continuación, una aproximación que establece un límite claro entre lo que es dato público y dato sensible y/o confidencial.



Sin embargo, esta aproximación no permite cubrir los objetivos antes identificados ni las necesidades que se comentan en el punto 2 de este documento. Es por ello por lo que los proveedores han seguido buscando la forma de seguir asegurando y preservando la seguridad, la privacidad y la soberanía del dato mientras se sigue ofreciendo todas las ventajas de una nube hiperescala.

Dicha aproximación se basa en la utilización de tecnologías de protección de datos siguiendo el siguiente modelo:



Como se puede apreciar, en el caso de Microsoft, la tecnología se considera lo suficientemente madura para abordar:

1. Los escenarios de **datos sensibles mediante una propuesta de nube pública** o **nube pública sensible que añade medidas de protección adicionales** a la nube pública.
2. Los escenarios de **datos confidenciales** mediante soluciones de nube privada apoyadas en **Azure Stack HCI/Hub + Azure Arc**.

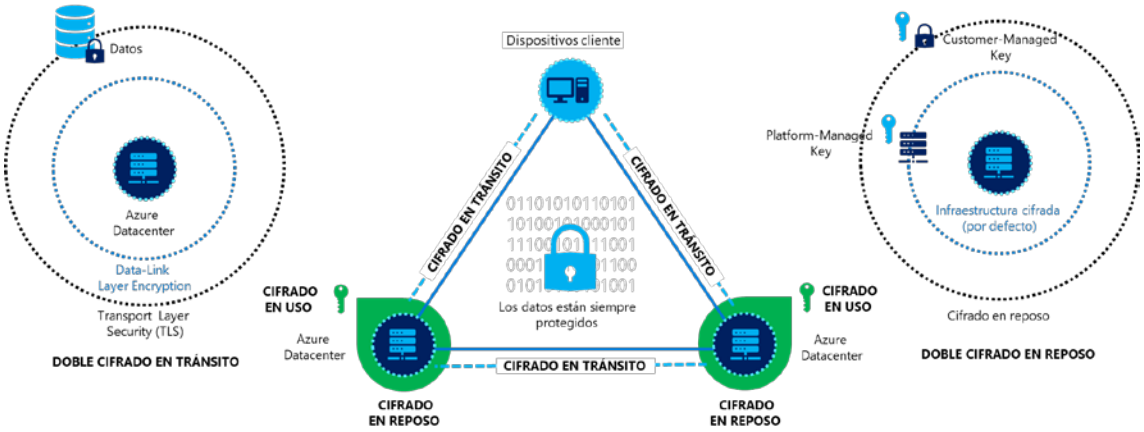
El mapeo tecnológico sería el siguiente:

- **Nube Pública**: utilizando los mecanismos existentes actuales de doble cifrado en tránsito y las capacidades de cifrado en reposo existentes en todos los servicios de Azure.
- **Nube Pública Sensible**: añadiendo a las capacidades de nube pública las herramientas de Azure Key Vault, Managed HSM, Dedicated HSM y para garantizar el cifrado en uso utilizando Azure Confidential Computing.
- **Nube Privada**: añadiendo la posibilidad de escenarios semi-desconectados o totalmente desconectados con nube pública a través de Azure Stack HCI (hiper convergente semi desconectado) y Azure Stack Hub (Azure on-premise totalmente desconectado).

  Adicionalmente cabe la posibilidad de gestionar entornos de nube privada como VMWare a través de Azure Arc desde Azure Stack HCI/Hub.

2. MODELO DE PROTECCIÓN DE DATOS

El cifrado es fundamental para ayudar a garantizar la confidencialidad de las cargas de trabajo en la nube. El proveedor utiliza múltiples métodos, protocolos y algoritmos de cifrado en todos sus productos y servicios tanto para el almacenamiento como en tránsito y en uso.



Además, la administración adecuada de claves es un elemento esencial en los procedimientos recomendados de cifrado. Las claves de cifrado se pueden gestionar de tres (3) maneras: en el caso de tecnología Microsoft por Azure en Azure Key Vault, por el cliente en Key Vault o gestionado y almacenado *on-premise*.

La gestión de claves incluye opciones que van desde claves administradas por servicio del lado del servidor hasta el cifrado del lado cliente en las que los servicios de la nube no tienen acceso a dichas claves y no puede descifrar los datos del cliente.

- Más información: https://aka.ms/AZ-encryption-overview y https://aka.ms/AZ-encryption-best

2.1 Gestión de claves

Antes de profundizar en las tecnologías de cifrado en reposo, en tránsito y en uso es conveniente hacer un pequeño repaso a la gestión de claves.

La administración de claves puede ser realizada **por la nube o por el cliente**, y el cifrado puede ser realizado **en el lado del servidor o en el lado del cliente**.

- **Cifrado del lado del servidor:** hay tres (3) modelos de cifrado del lado del servidor que ofrecen diferentes características de gestión de claves entre los que se pueden elegir según los requisitos de la organización:
  - **Claves administradas por el servicio** usan la nube (por ejemplo, Azure Key Vault[1]) para proporcionar una combinación de control y facilidad de uso unido a unos costes controlados. La nube (Azure) realiza las operaciones de cifrado y descifrado y el proveedor (Microsoft) administra las claves.
  - **Claves administradas por el cliente** permite darle el control sobre las claves al cliente, incluyendo soporte para traer su propia clave BYOK (bring-your-own-key) o generar otras nuevas. La nube realiza las operaciones de cifrado y descifrado. El cliente controla las claves mediante la tecnología de nube (por ejemplo, Azure Key Vault).

    El proceso se detalla en el siguiente enlace: Generación y transferencia de claves protegidas con HSM (BYOK): Azure Key Vault | Microsoft Docs y la lista de HSM compatibles es la siguiente: HSM compatibles.

    - Más información: https://aka.ms/AZ-CMK y https://aka.ms/AZ-HMS-Keys
  - **Claves proporcionadas por el cliente** (CPK) permitiendo almacenar y gestionar claves *on-premise* o en almacenes de claves distintos de la tecnología de nube.
- **Cifrado del lado del cliente**: con el cifrado del lado del cliente, el proveedor (Microsoft) no tiene acceso a las claves de cifrado y no se pueden descifrar los

[1] **Azure Key Vault** es un servicio alojado en la nube que proporciona almacenamiento y administración centralizados de claves criptográficas y otros secretos que se utilizan en las aplicaciones en la nube. Este servicio de Azure permite a los clientes proteger dichas claves criptográficas, certificados y las contraseñas de aplicaciones, y ayuda a proteger los secretos de fugas de información accidentales.

datos. Los clientes cifran los datos y cargan los datos como un blob cifrado. El cliente mantiene un control completo de las claves y mantiene las mismas *on-premise* (o en otras localizaciones) y de esta manera las claves no están disponibles para los servicios de la nube (por ejemplo, Azure).

El modelo de cifrado del lado del cliente[2] se refiere a las claves de cifrado **on-premise gestionadas por el propio cliente**, es decir, el cifrado se realiza fuera del proveedor de recursos o fuera de Azure. Este cifrado es gestionado por una aplicación que se ejecuta en el centro de datos del cliente o por un servicio de aplicación.

- Más información: https://aka.ms/AZ-encryption-overview y https://aka.ms/AZ-client-encrypt

2.2 Cifrado en reposo

Los datos de los clientes se cifran en reposo automáticamente cuando se almacenan en cualquier servicio del proveedor (por ejemplo, Azure Storage).

- Más información: https://aka.ms/AZ-Storage-Encrypt

La nube (por ejemplo, Azure) permite múltiples opciones de cifrado para el sistema operativo y los discos de datos, tanto para instancias de Windows Server como de Linux. Azure Disk Encryption cifra los discos virtuales para infraestructura como servicio (IaaS) de Windows y Linux mediante el uso de BitLocker en el caso de Windows y DM-Crypt en el caso de Linux para proporcionar un cifrado completo tanto para el disco de sistema operativo como para el disco de datos.

- Por ejemplo, en el caso de Azure Disk Encryption se requiere un Azure Key Vault para controlar y administrar las claves y secretos de cara al cifrado del disco. El almacén de claves y las máquinas virtuales deben residir en la misma región y suscripción de Azure.
- BitLocker también cifra las Shielded VM en Windows Server 2016 para garantizar que los administradores del fabricante (administradores de Azure) no puedan acceder a la información dentro de una máquina virtual. La solución Shielded VMs incluye el Host Guardian Service, que se utiliza para la certificación del host de virtualización y el uso de las claves de cifrado.
  - Más información: https://aka.ms/AZ-encryption-vms

2.3 Cifrado en tránsito

La protección de los datos en tránsito es una parte esencial de cualquier estrategia de protección de la información. Las organizaciones que no protegen los datos en tránsito son más susceptibles a ataques *man-in-the-middle* y ataques de sesión. Debido al

[2] Cuando se sigue este modelo de cifrado, el proveedor de recursos de Azure recibe un blob cifrado de datos y Azure no tienen la capacidad de descifrar los datos de ninguna manera o tener acceso a las claves de cifrado

continuo flujo de datos entre diferentes localizaciones, la recomendación general es que los clientes siempre utilicen protocolos de seguridad en la capa de transporte (SSL/TLS) para dicho intercambio.

La inversión de los proveedores en investigación y desarrollo ha dado lugar a grandes avances en el cifrado de datos en tránsito. En el caso de Microsoft, cada servidor de Azure dispone de un *custom silicon* denominado Azure SmartNIC, que está basado en Field Programmable Gate Array (FPGA). Estas FPGA son módulos de hardware programables, que aceleran significativamente el procesamiento de datos, incluido el cifrado de los mismos en tránsito. Esto permite mejorar el rendimiento para todas las cargas de trabajo, reduciendo la latencia.

- Más información: https://aka.ms/AZ-SmartNICs

Además, existen casos en los que los clientes deciden aislar todo el canal de comunicación entre infraestructuras *on-premise* y la nube mediante el uso de una red privada virtual (VPN). En el caso de Microsoft, podemos considerar las siguientes alternativas:

- Para los datos que se mueven entre la infraestructura *on-premise* y Azure, se puede considerar protocolos como **HTTPS** o **VPN**.
- Para organizaciones que necesitan proteger el acceso desde estaciones de trabajo *on-premise* hacia Azure, es posible utilizar Azure Site-to-Site VPN
- Para organizaciones que necesitan proteger el acceso desde una sola estación de trabajo ubicada *on-premise* hacia Azure, es posible utilizar Point-to-Site VPN
- Para las interacciones de los clientes con Azure Storage a través del Portal de Azure, todas las transacciones se producen sobre HTTPS. También se puede interactuar con dicho almacenamiento mediante Storage REST API sobre HTTPS para trabajar con Azure Storage y Azure SQL Database.
- Los conjuntos de datos más grandes se pueden mover a través de un enlace WAN de alta velocidad dedicado, como ExpressRoute. Los clientes que decidan utilizar ExpressRoute también pueden cifrar los datos a nivel de aplicación mediante SSL/TLS u otros protocolos de mayor protección.
    - Más información https://aka.ms/AZ-expressroute-Enc

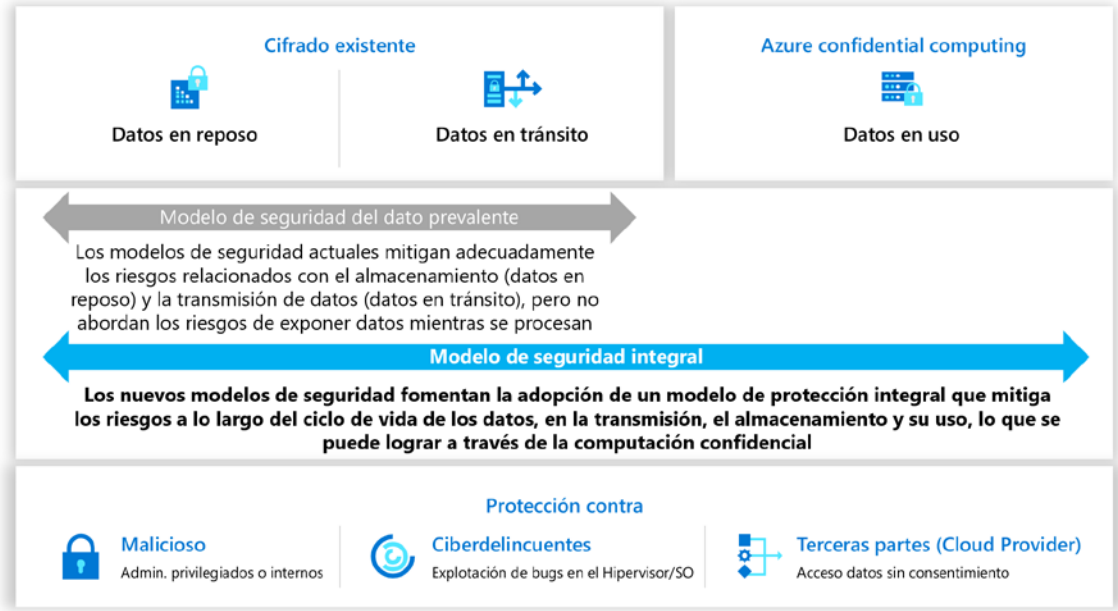
2.4 Cifrado en uso

En estos últimos años, varios fabricantes han creado la Confidential Computing Consortium. Este consorcio reúne a proveedores de hardware, proveedores de nube y desarrolladores de software para acelerar la adopción de tecnologías y estándares relacionadas con los Entornos de Ejecución Confiables o TEE (de sus siglas en inglés Trusted Execution Environment).

CCC es una comunidad de proyectos bajo la Fundación Linux dedicada a definir y acelerar la adopción de la computación confidencial. La computación confidencial protege los datos en uso mediante la ejecución de procesos en un entorno de ejecución de confianza basado en hardware. Estos entornos seguros y aislados evitan el acceso no autorizado o la modificación de aplicaciones y datos mientras están en uso, lo que aumenta las garantías de seguridad para las organizaciones que administran datos sensibles y/o regulados.

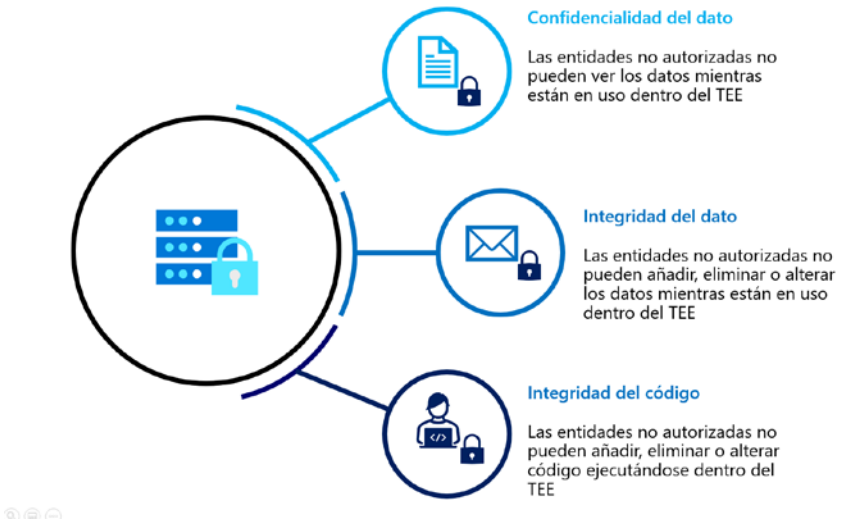
Hoy en día, los datos se cifran en reposo y en tránsito, pero no mientras se usan, es decir cuando están en memoria. Además, la capacidad de proteger los datos y el código mientras están en uso estaba limitada con la infraestructura informática convencional. Las organizaciones que manejan datos sensibles, como información de identificación personal (PII), datos financieros o información de salud, deben aplicar salvaguardas para contrarrestar las amenazas que se dirigen a la confidencialidad e integridad tanto de las aplicaciones como de los datos que están en la memoria de un sistema.

A tal efecto, por ejemplo Microsoft lanzó hace casi ya dos años Azure Confidential Computing que viene a complementar la estrategia de seguridad y privacidad de los datos e iniciar la senda para la construcción de una nube confidencial:



La **computación confidencial es un principio informático que preserva la privacidad** aprovechando **los entornos de ejecución de confianza (TEE) basados en hardware para proteger los datos y el código durante su procesamiento**.

Un TEE es un área segura dentro de un procesador que ejecuta, tras su verificación, un entorno aislado paralelo al sistema operativo principal. A través de este aislamiento a nivel de hardware, el TEE garantiza que los datos y el código cargados en él no pueden ser manipulados por los operadores de nube, administradores maliciosos o ciberdelincuentes que aprovechen vulnerabilidades en el hipervisor.



Con esta tecnología además de aprovechar las nuevas capacidades para seguir manteniendo la privacidad y la seguridad del dato, se pueden abordar escenarios más complejos como por ejemplo beneficiarse del análisis de datos entre múltiples organizaciones o el aprendizaje automático para combinar conjuntos de datos de organizaciones que no confían entre ellas manteniendo los datos privados incluso entre las partes.

- Ejemplo: confidential computing en salud

Desde su concepción, la computación confidencial proporciona los controles técnicos necesarios para aislar los datos de los accesos por parte de los administradores del proveedor, los administradores del cliente o incluso ambos.

En el caso de Microsoft, la propuesta tecnológica para computación confidencial incluye:

- Confianza basada en el hardware para garantizar que los datos están siempre protegidos y *anclados al chip.* La confianza está encapsulada al fabricante del hardware, por lo que incluso los administradores de Microsoft no pueden modificar dichas configuraciones.
- Atestación remota (mediante Azure Attestation) para que los clientes verifiquen la integridad del entorno. Los clientes pueden verificar que tanto el hardware como el software en los que se ejecutan las diferentes cargas de trabajo son versiones aprobadas y protegidas antes de permitirles acceder a los datos.
- Trusted Launch, mecanismo (producto) que garantiza que las máquinas virtuales arrancan con software autorizado y que utiliza la atestación remota para que los clientes puedan verificarla. Está disponible para todas las máquinas virtuales, incluidas las máquinas virtuales del tipo computación confidencial, que traen arranque seguro y vTPM para protegerse contra rootkits, bootkits y firmware malicioso.

- Aislamiento y cifrado de memoria para garantizar que los datos estén protegidos durante el procesamiento de estos. Azure ofrece **aislamiento por máquina virtual, contenedor o aplicación y cifrado basado en hardware** para evitar la visualización no autorizada de datos, incluso en circunstancias donde se tenga acceso físico en el centro de datos.
- Administración de claves segura para garantizar que las claves permanece cifradas durante todo su ciclo de vida y se liberen sólo en el código autorizado.

Más información de los anuncios alrededor de computación confidencial:

- Confidential virtual machines with Intel SGX secure enclaves (preview).
- Confidential VMs with AMD SEV-SNP (preview).
- Trusted Launch with secure boot and vTPMs across all Azure Gen 2 virtual machines, to verify only trusted code runs on a VM.

Servicios confidenciales, herramientas y frameworks:

- Azure confidential ledger actualmente en preview
- SQL Always Encrypted with secure enclaves
- Azure Key Vault Managed HSM
- Microsoft Azure Attestation
- Confidential Consortium Framework (CCF) framework open source para ledgers basados en blockchain
- Open Enclave librería open source SDK para C/C++
- ONNX Runtime framework open source para elaborar modelos de inferencia confidencial de Machine Learning
- Mystikos for porting .Net apps into application enclaves
- Enclave Device Blueprint para computación confidencial en el Edge/IoT
- Confidential containers on AKS
- AKS with confidential computing nodes
- Video demostración completo End-to-end Sensitive Web App with Azure Confidential Services - YouTube