

Procesamiento de datos masivos

03MBID

Tema 1: Computación Ubicua – *Cloud Computing*

Parte II



Unificación de recursos distribuidos geográficamente PARA LA VENTA

> ¿Qué es?

- ★ Permite ofrecer servicios de *hosting* en Internet
- ★ Provee acceso fácil y escalable de recursos de computación y servicios de IT
- ★ Es un modelo de entrega y consumo de servicios en el que existe **un consumidor** y **un prestador**; el consumidor busca recursos que no posee y el prestador se los proporciona según algunos términos y condiciones.
- ★ Esta relación, entre el que quiere un recurso para hacer su actividad y el que se lo presta, es un acceso directo a las capacidades de consumir, escalabilidad, pago por uso, transparencia en las condiciones de uso, de términos y condiciones y de esta forma poder acceder a ellas de manera muy rápida

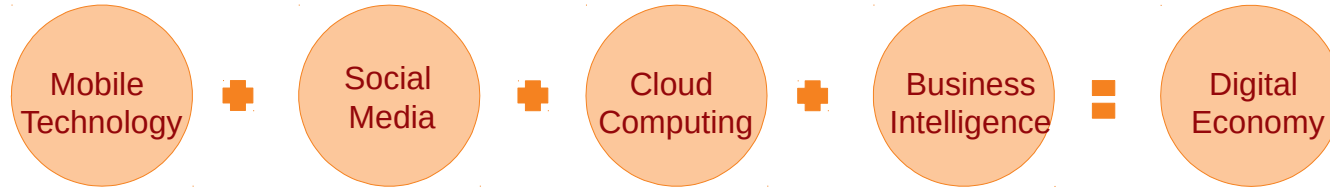
> ¿Qué es?

- ★ Es un paradigma que posibilita el **acceso ubicuo bajo demanda** a servicios TIC accesibles a través de Internet
 - Computación como servicio
 - Externalización
 - Economía de escala
 - Permite centrarse en la lógica de negocio y no en la infraestructura informática
- ★ El término **cloud** (nube) se refiere a la forma de representar la red (Internet) en los diagramas y es una abstracción de las complejidades de su infraestructura



> ¿Qué es?

★ Es una de las piezas claves de la nueva economía digital



- Transformación digital de industrias tradicionales
- Fomento de *start-ups* tecnológicas
 - ✓ Democratización en el acceso a recursos TIC
- Base en el desarrollo de nuevas tecnología
 - ✓ Big Data, IoT, Machine Learning, ...

> ¿Qué es?

★ Supone un cambio en la forma en que los ISPs gestionan y ofrecen recursos

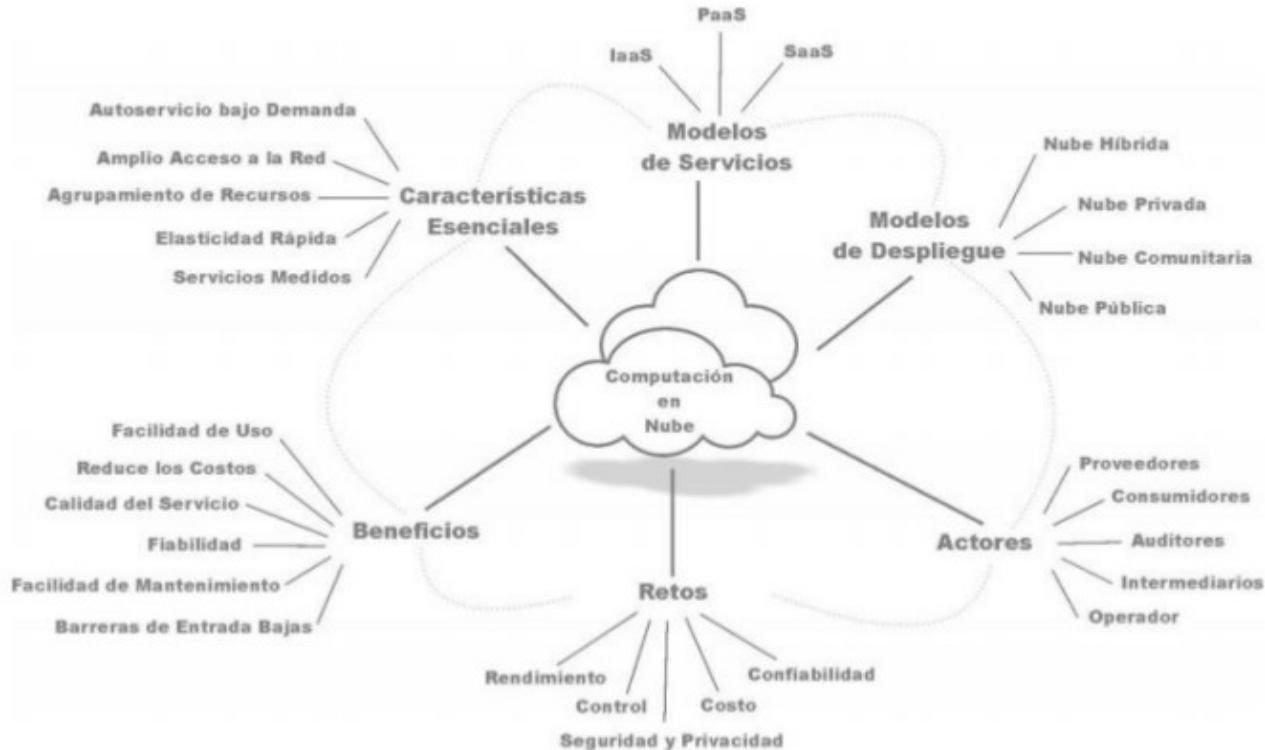
- **Plataformas flexibles y “elásticas”**

- ✓ Adaptación a la demanda real
- ✓ Soporta mejor los picos de demanda
- ✓ Escalabilidad
- ✓ Tolerancia a fallas

- Se usan como base las **tecnologías de virtualización**

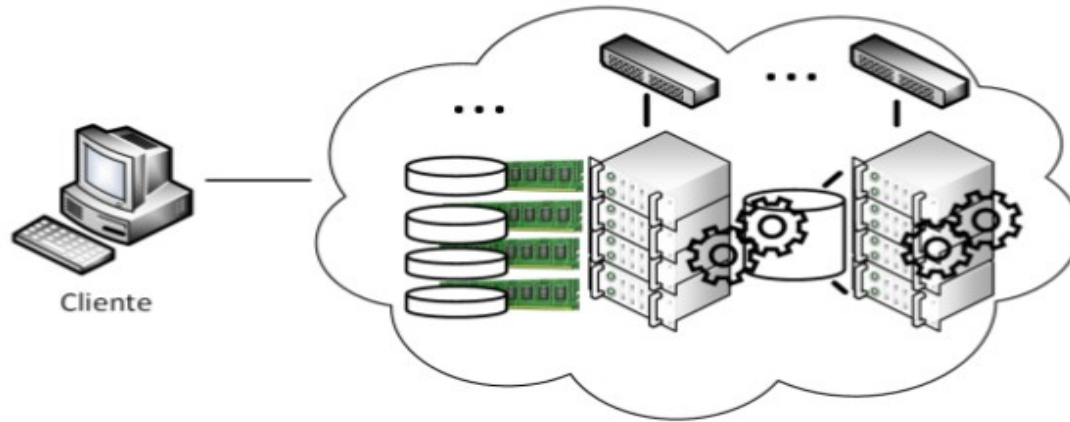
- ✓ Los recursos se virtualizan bajo la forma de máquinas virtuales (*virtual machines* – VMs)
- ✓ Las VMs se despliegan dinámicamente bajo demanda

- Los recursos virtualizados son **expuestos como servicios** a través de **interfaces estandarizadas**

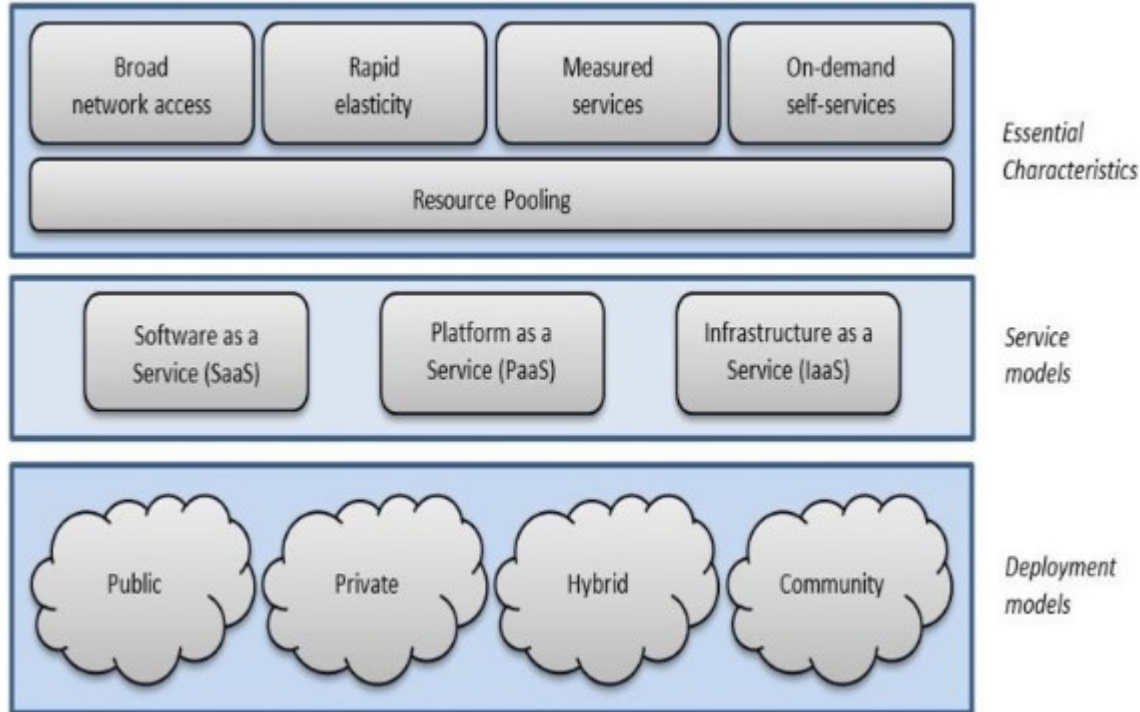


> ¿Qué es?

NIST: “cloud computing is a model for enabling **ubiquitous, convenient, on-demand** network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications and services) that can be **rapidly** provisioned and released with **minimal management** effort or service provider interaction”



> ¿Qué es?



El modelo cloud del NIST

★ Conocido popularmente como 5-3-4

- 5 características fundamentales
- 3 modelos de servicio
- 4 modelos de despliegue

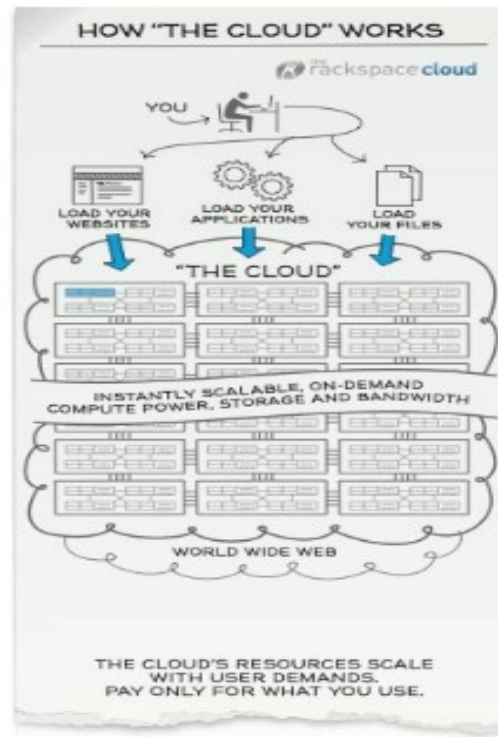
★ Usado como referencia también a nivel europeo

CARACTERÍSTICAS ESENCIALES

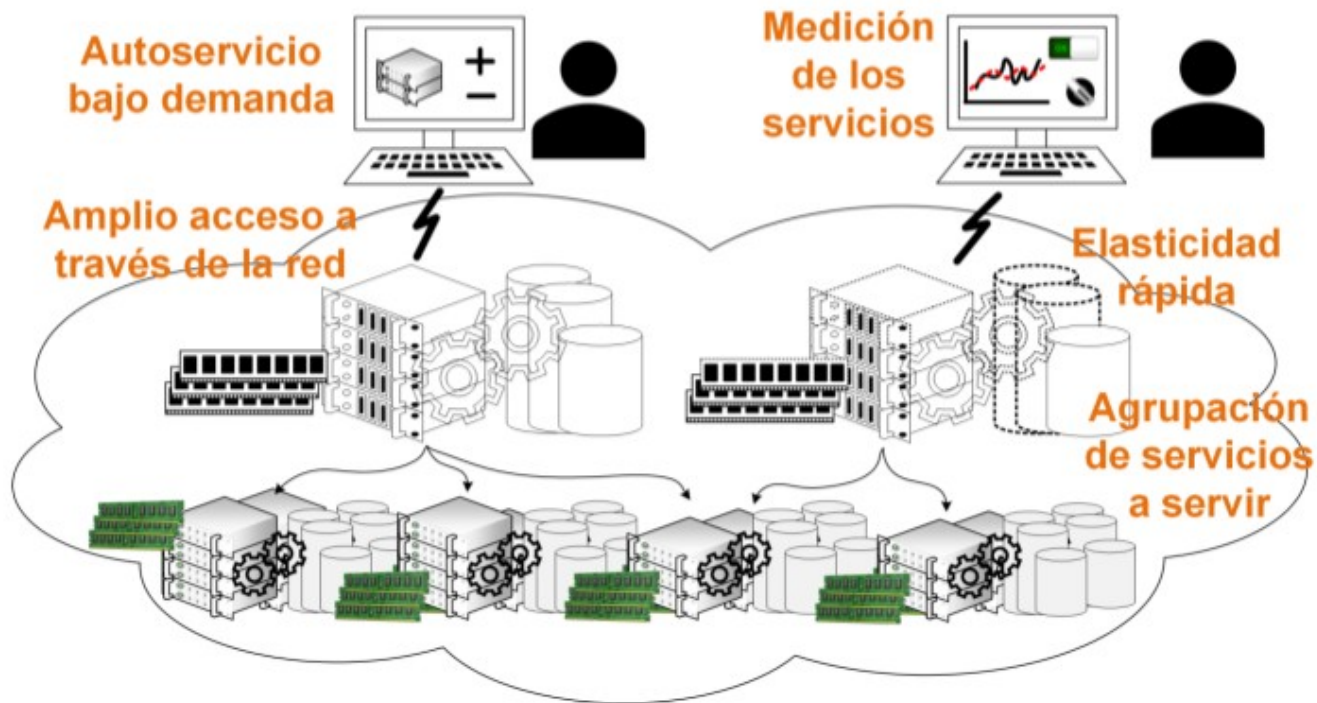


> Supone una nueva forma de proveer y consumir capacidades TIC

- ★ Autoservicio bajo demanda
 - Consume el servicio cuando quieras (*when*)
- ★ Amplio acceso a la red
 - Consume el servicio desde cualquier lugar (*where*)
- ★ Agrupación de recursos (*pool* y virtualización)
 - Capacidades como servicio (*aaS), *pool* de infraestructuras, plataformas virtuales y aplicaciones (*how*)
- ★ Escalabilidad y elasticidad rápida
 - *Pool* de recursos compartidos para ofrecer escalabilidad horizontal (*how*)
- ★ Servicios medidos
 - Pago por lo que se consume, cuando se consume (*how much*)



Características esenciales



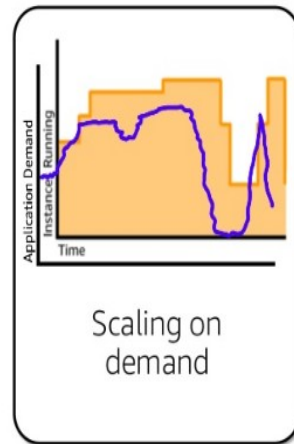
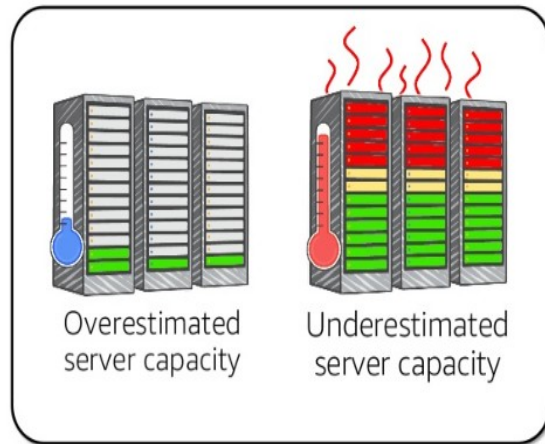
> Autoservicio bajo demanda

★ Infraestructura local

- Problema: ¿cómo dimensionar la capacidad de los servidores para ajustarse a la demanda?

★ Aprovisionamiento bajo demanda

- Consume lo que se necesita
- Pago por lo que se consume



© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

BENEFICIOS



★ Integración probada de servicios de red

- La tecnología *cloud computing* se puede integrar con mucha mayor facilidad y rapidez con el resto de las aplicaciones empresariales

★ Prestación de servicios a nivel mundial

- Infraestructuras que proporcionan mayor capacidad de adaptación, recuperación completa de pérdida de datos (con copias de seguridad) y reducción al mínimo de las tiempos de inactividad
- Ante eventualidades de fallas, el *cloud* re-asigna nuevas VM para las aplicaciones

★ Implementación más rápida y con menos riesgos

- Se comienza a trabajar más rápido y no es necesario una gran inversión

★ Escalamiento dinámico

- Permite *scaling out* y *scaling back* de acuerdo a los requerimientos cambiantes de las aplicaciones (*elastic scaling*)

★ Actualizaciones automáticas

- Que no afectan negativamente a los recursos de TIC
- Ahorro de tiempo y recursos para personalizar e integrar actualizaciones

★ Alta disponibilidad y durabilidad

- El *cloud* ofrece una plataforma para las aplicaciones con acceso y almacenamiento confiable
- Asegura alta disponibilidad de los recursos computacionales

★ Contribuye al uso eficiente de la energía

- Las organizaciones no requieren mantener *datacenters* que consumen gran energía

- ★ La centralización de las aplicaciones y el almacenamiento de los datos origina una interdependencia de los proveedores de servicios.
- ★ La disponibilidad de los servicios está sujeta a la disponibilidad de acceso a Internet.
- ★ Los datos “sensibles” del negocio no residen en las instalaciones de las empresas, lo que podría generar un contexto de vulnerabilidad para la sustracción o robo de información.
- ★ La confiabilidad de los servicios depende de la “salud” tecnológica y financiera de los proveedores de servicios en la nube.
- ★ Escalabilidad a largo plazo
 - A medida que más usuarios comparten la infraestructura de la nube, la sobrecarga en los servidores de los proveedores aumentará
 - Si el proveedor no dispone de un esquema de crecimiento óptimo puede llevar a degradaciones en el servicio o altos niveles de *jitter*.

- ★ Los servicios del *cloud* son más complejos que los servicios tradicionales
- ★ Privacidad
 - La información queda expuesta a terceros
- ★ Seguridad
 - La información de la empresa debe recorrer nodos para llegar a su destino, cada uno de ellos (y sus canales) son un foco de inseguridad
 - Si se utilizan protocolos seguros, como HTTPS, la velocidad total disminuye debido a la sobrecarga que éstos requieren
- ★ El manejo de estándares para el despliegue de nubes aún no está del todo establecido
- ★ Se requieren fuertes fundamentos de “buenas prácticas” en el desarrollo de software, arquitectura de software y gestión de servicios.
- ★ Los *clouds* están conformados por grandes *datacenters* que consumen gran energía

MODELOS DE SERVICIOS



> Los servicios se modelan en varias categorías

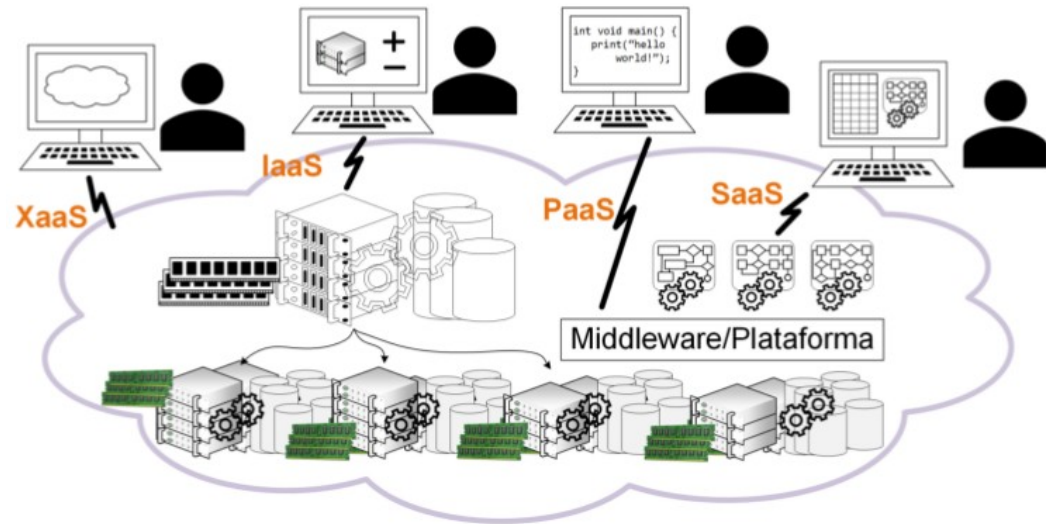
★ Infrastructure-as-a-Service (IaaS)

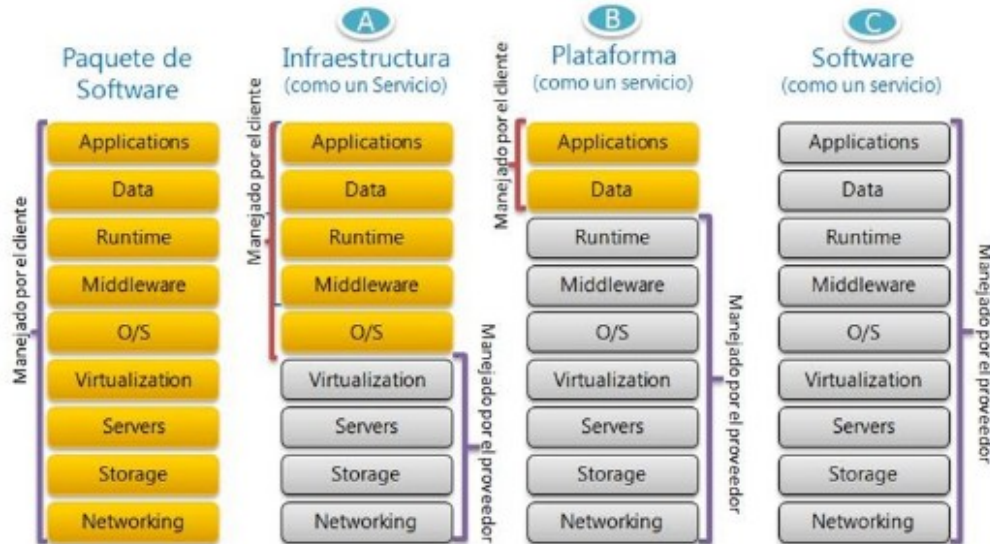
★ Platform-as-a-Service (PaaS)

★ Software-as-a-Service (SaaS)

★ Data-as-a-Service (DaaS)

★ ...







> Infrastructure as a Service (IaaS)

- ★ El consumidor aprovisiona recursos de computación (e.g., CPU, almacenamiento, red) en los que ejecuta su software (incluidas aplicaciones y sistemas operativos).
- ★ El consumidor no controla la infraestructura *cloud* subyacente pero si los sistemas operativos, el almacenamiento y las aplicaciones desplegadas y, posiblemente, tiene un control limitado sobre la red de comunicaciones (e.g., *firewalls*)



> Infrastructure as a Service (IaaS)

- ★ Los proveedores son los propietarios de la infraestructura de hardware y se la ofrecen como servicio a los usuarios a través de interfaces que les permiten gestionarla
- ★ Algunas de las funciones de gestión típicas para el usuario son:
 - Gestión dinámica de recursos
 - ✓ Inicio/parada de VMs, creación/eliminación de volúmenes, ...
 - Funciones de automatización
 - ✓ Balance automático de carga, reglas de auto-escalado horizontal, ...
 - Monitorización de recursos
 - ✓ CPU, memoria, ancho de banda, ...
 - Gestión de consumos y costos

> Platform as a Service (PaaS)

- ★ El consumidor despliega aplicaciones tanto propias como adquiridas, desarrolladas usando entornos de programación soportados por el proveedor, en su infraestructura *cloud*.
- ★ El consumidor no controla la infraestructura *cloud* subyacente pero si las aplicaciones desplegadas y, posiblemente, la configuración del entorno de despliegue

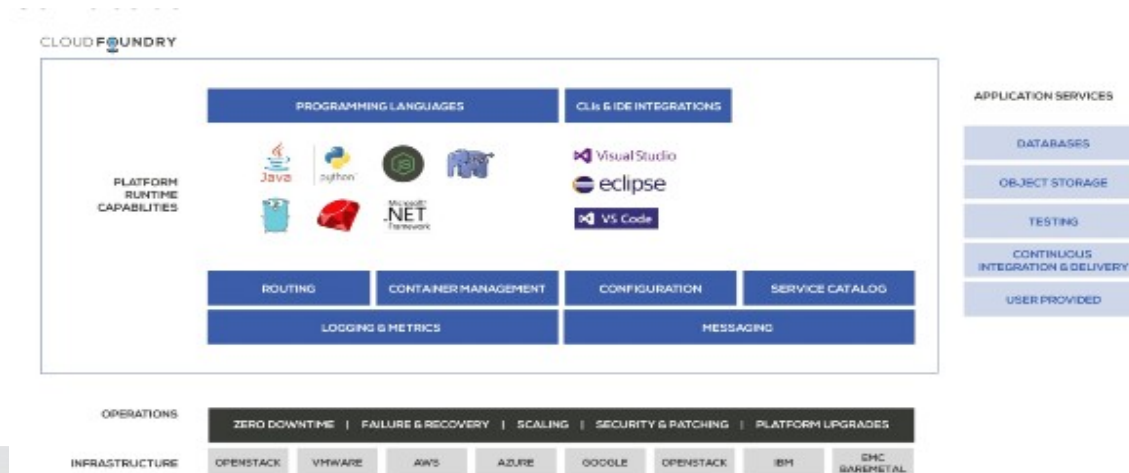


Proveedores PaaS: <https://www.g2crowd.com/categories/cloud-platform-as-a-service-paas>

> Platform as a Service (PaaS)

★ Las plataformas más completas soportan:

- ✓ Diferentes lenguajes, entornos de desarrollo, servidores de aplicaciones, SGBD y almacenamiento NoSQL, proveedores IaaS
- ✓ Gestión automatizada del ciclo de vida completo de las aplicaciones
- ✓ Servicios de aplicación: integración, orquestación, configuración, monitorización, ...



> Software as a Service (SaaS)

- ★ El consumidor utiliza las aplicaciones del proveedor que son ejecutadas en la infraestructura *cloud*.
- ★ El consumidor no controla ni la infraestructura *cloud* subyacente ni las capacidades de las aplicaciones, pero posiblemente si puede controlar la configuración personal.

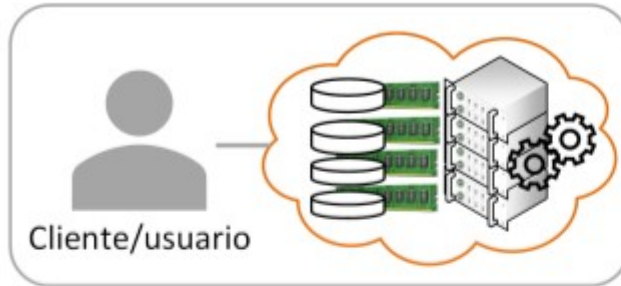


Directorio SaaS: <http://www.portalsaas.com/>

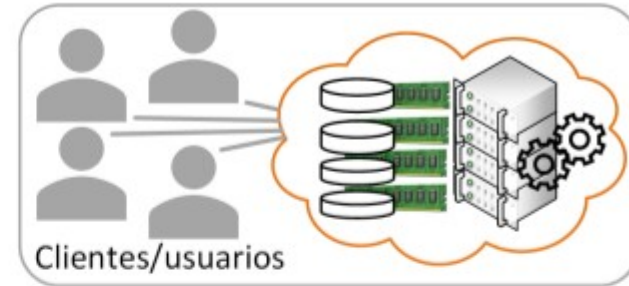
MODELOS DE DESPLIEGUE



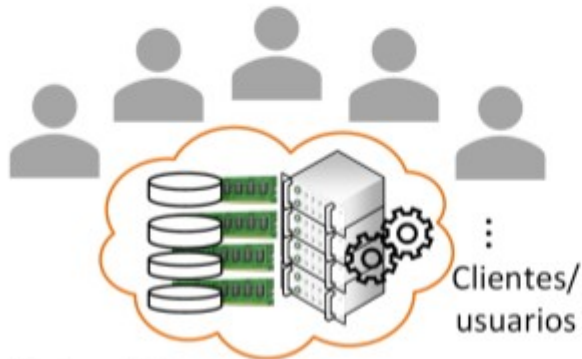
Modelos de despliegue



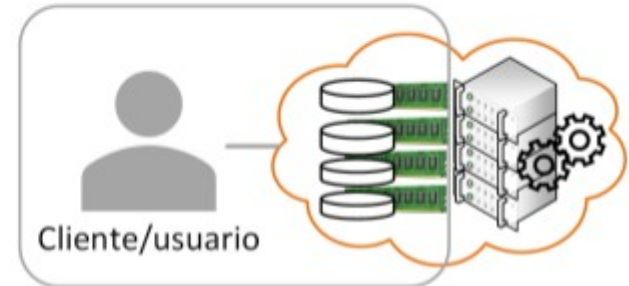
Nube privada



Nube comunitaria



Nube pública



Nube híbrida

★ Cloud privado

- De uso exclusivo para una empresa
- Puede ser de su propiedad o alquilado

★ Cloud comunitario

- Infraestructura compartida por varias organizaciones
- Da soporte a una comunidad específica

★ Cloud público

- Propiedad privada
- Vende servicios al público o a empresas en general

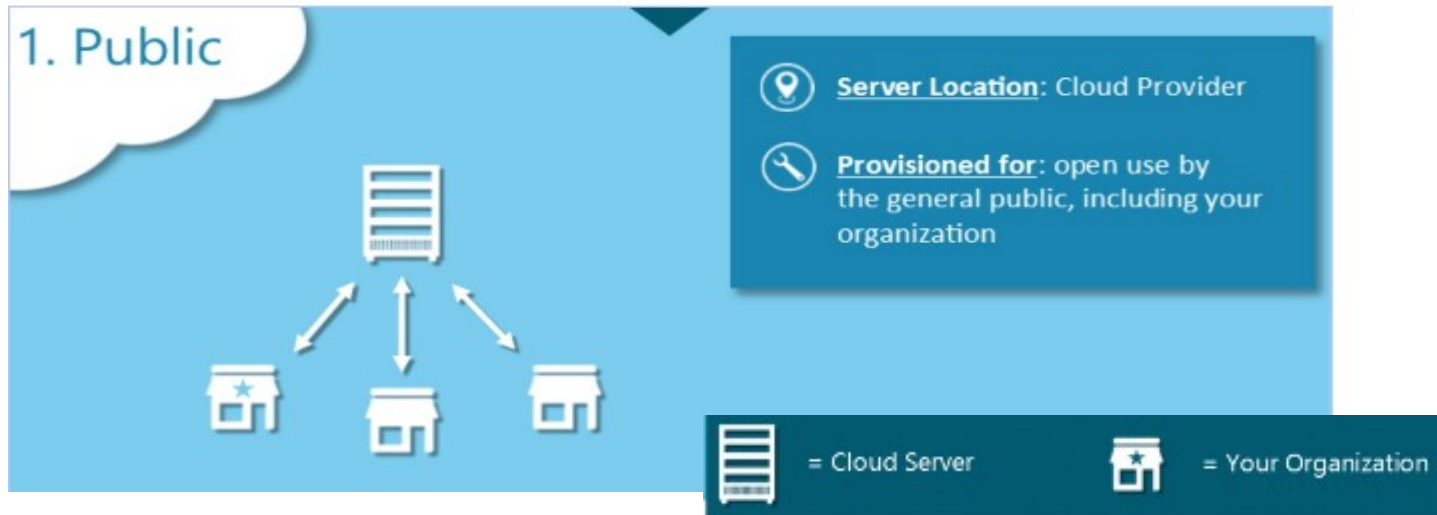
★ Cloud híbrido

- Compuesto por dos o más *clouds*.



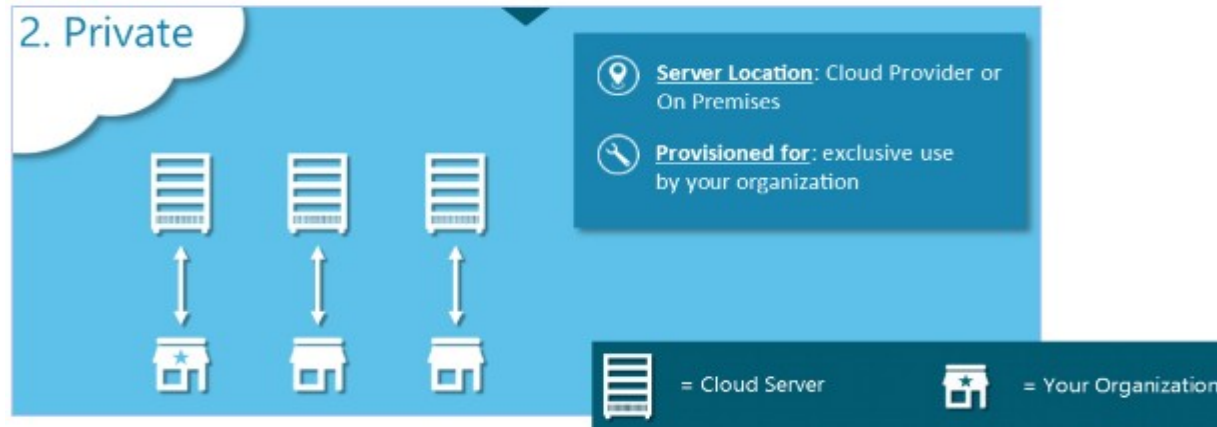
> Cloud público

- ★ Infraestructura para uso del público o empresas en general
- ★ Propiedad del proveedor (e.g., empresa, universidad, organismo estatal)
- ★ Tienen una interfaz para la gestión interna de la infraestructura virtualizada y exponen otra externa para la gestión de los recursos de los clientes

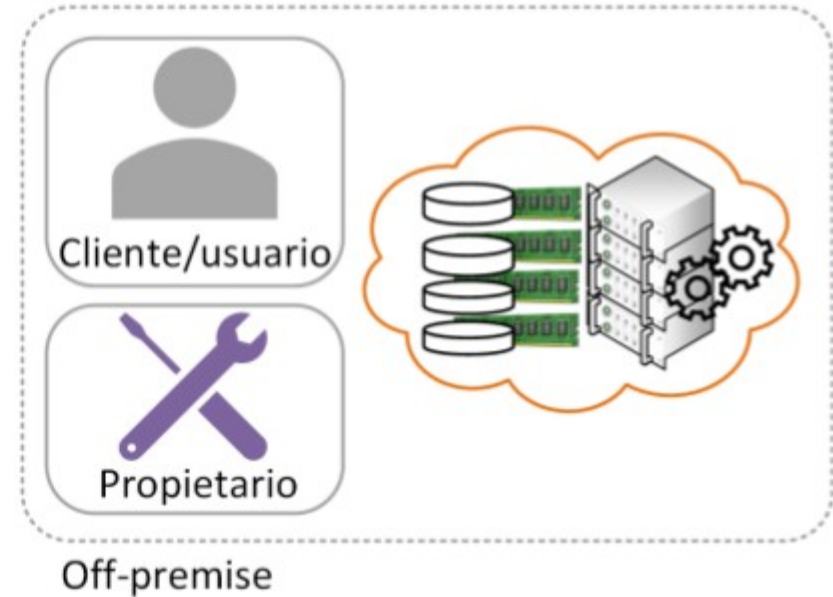
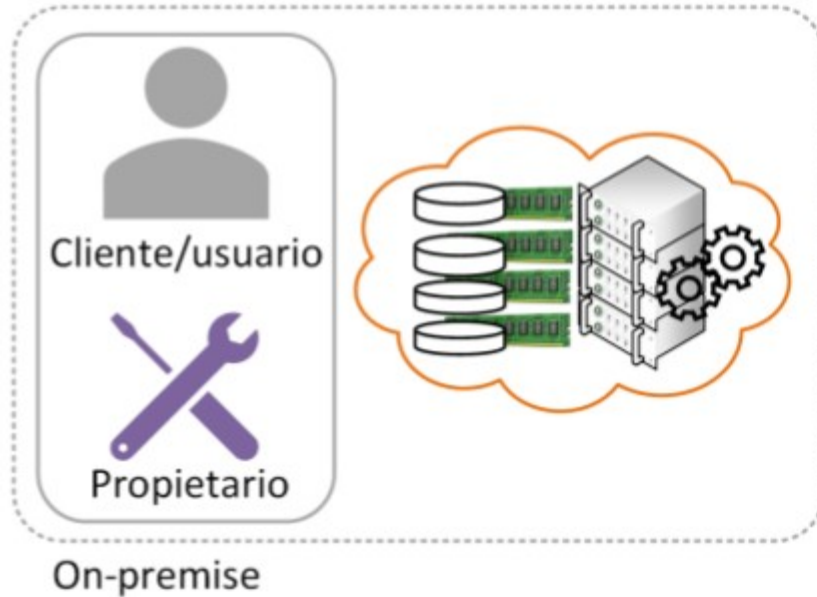


> Cloud privado

- ★ Infraestructura para uso de una organización con múltiples clientes (e.g., departamentos de una empresa).
- ★ Puede ser de su propiedad (*on-premise*), alquilado (*off-premise*) o una combinación de ambas.
- ★ Permite una gestión flexible y ágil de los recursos de la organización.
- ★ Los principales proveedores públicos de IaaS comenzaron creando *clouds* privados para mejorar la gestión de sus *datacenters*.

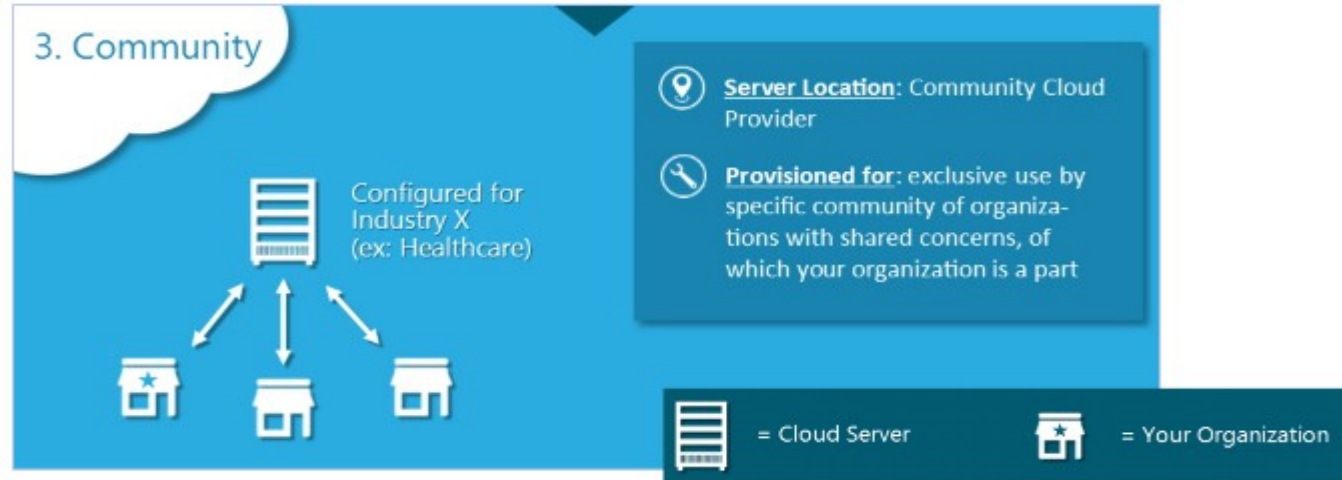


> Cloud privado



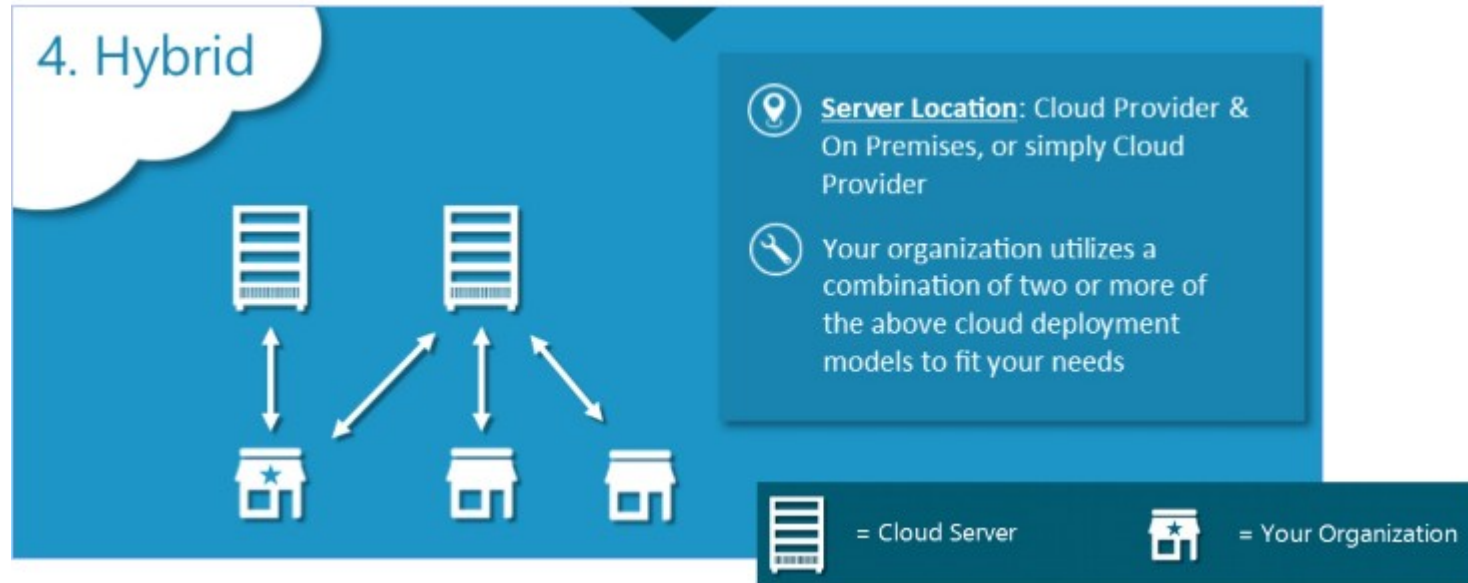
> Cloud comunitario

- ★ Infraestructura para uso de una comunidad de organizaciones que comparten intereses
 - Cada componente conserva su autonomía
 - La relación entre componentes se hace mediante tecnologías (propietarias o estandarizadas) que permiten la portabilidad de los datos y aplicaciones
- ★ Puede ser propiedad de una o varias de las organizaciones, ser alquilada o una combinación de ambas.



> Cloud híbrido

- ★ Combinación de *cloud* privado y público
- ★ Permite gestionar los picos de carga obteniendo recursos del *cloud* público
- ★ El uso del *cloud* público es totalmente transparente para los usuarios del *cloud* privado.



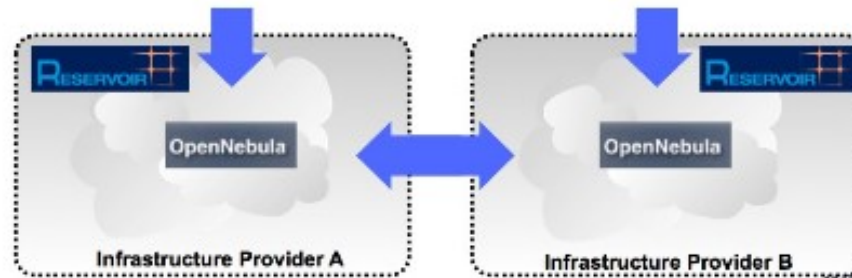
> Cloud híbrido

- ★ **Ejemplo 1:** Una organización puede almacenar datos sensibles de los clientes en una aplicación en una nube privada *on-premise*, pero interconecta esa aplicación con una aplicación de gestión de cuentas provista como servicio en una nube pública. El *cloud* híbrido extiende las capacidades de la empresa de ofrecer servicios de negocios específicos a través de la integración de servicios de nubes públicas externas.
- ★ **Ejemplo 2:** Una organización usa recursos de cómputo de una nube pública para satisfacer necesidades de capacidades temporales que no puede satisfacer con la nube privada. En este caso, el *cloud* híbrido permite escalar a través de *clouds* (*cloud bursting*).

> Otros modelo de clouds

★ Federación de clouds

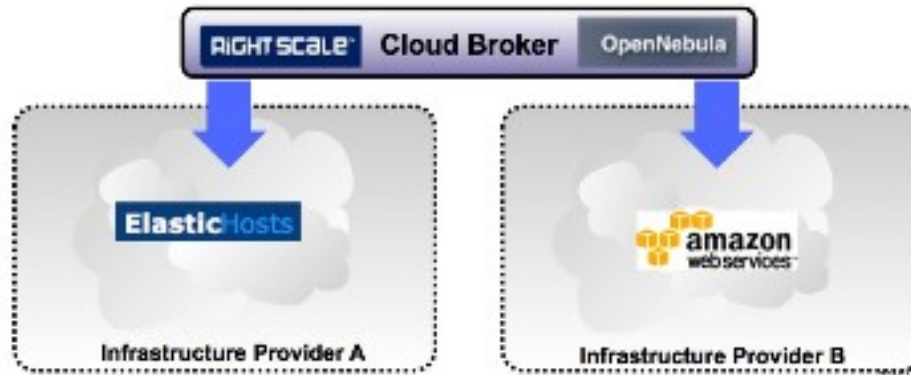
- Extensión del modelo híbrido
- Requieren mecanismos (estándares) que garanticen la interoperabilidad
- Mecanismo usado por el proveedor de la nube, en el que renta recursos, aplicaciones o servicios de varios vendedores de servicios en nube para alcanzar necesidades exponenciales de los clientes.
- La interacción inter-proveedores es realizada por el proveedor
- La alta disponibilidad es mejor porque los datos se replican en múltiples sitios



> Otros modelo de clouds

★ *Clouds brokering*

- Paradigma “*Clouds of Clouds*”
- Provee servicios de intermediación, agregación y arbitraje entre un conjunto de varios proveedores
- Provee la habilidad de seleccionar entre varios *clouds* públicos y de mover instancias de aplicaciones o datos entre *clouds* públicos y privados.



RETOS





Virtualización:

- Rendimiento
- Control de recursos

Contabilidad:

- Control de costo



QoS:

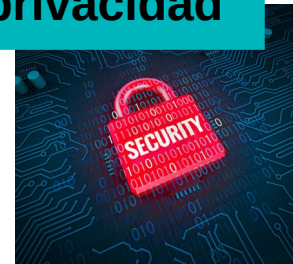
- Rendimiento
- Confiabilidad



Servicios Web



Seguridad y privacidad



★ Calidad de servicio

- La ausencia de QoS puede generar que las organizaciones declinen del uso del *cloud*.
- El compartimiento de recursos y datos complejos requieren de suficiente ancho de banda sin generar más costos.
- Debe ofrecer operaciones libres de fallas → Un *cloud* puede experimentar fallas regulares.

★ Seguridad y privacidad

- De primordial importancia, dado que los datos se comparten en el *cloud* y están susceptibles a ataques cibernéticos.

★ Virtualización

- Los accesos simultáneos de múltiples clientes a los servicios del *cloud* demandan más protección.
- Control de recursos para optimizar el desempeño de las aplicaciones, pero a la vez mejorar la utilización de los recursos y reducir el consumo de energía.

★ Servicios web

- Ofrecer interfaces sencillas que escondan la heterogeneidad de hardware y software en el *cloud*

★ Contabilidad

- Se requiere control de costos para no desbordar a las organizaciones con incrementos de pagos

Retos: estandarización

Project Name	URL	Focus
Cloud Audit	www.cloudaudit.org	Open, extensible, and secure interface, namespace, and methodology for cloud providers and their consumers.
Cloud Computing Interoperability Forum	www.riverbed.com	Common, agreed-on framework for cloud platforms to exchange information in a unified manner.
Cloud Security Alliance	www.cloudsecurityalliance.org	Recommended practices for cloud computing security.
Cloud Standards Customer Council	www.cloudstandardscustomerCouncil.org	Standards, security, and interoperability issues related to migration to the cloud.
Cloud Storage Initiative	www.snia.org/cloud	Adoption of cloud storage as a new delivery model (Data-Storage-as-a-Service). Initiative sponsored by the Storage Networking Industry Association (SNIA), the creator and promoter of the Cloud Data Management Interface (CDMI). SNIA includes members from Oracle, NetApp, and EMC.
Delta Cloud	www.incubator.apache.org/	Abstraction layer for dealing with differences among IaaS providers. API based on representational state transfer (REST). It has libraries for seven providers including Amazon EC2, Eucalyptus, and Rackspace.
Distributed Management Task Force (DMTF)	www.dmtf.org/standards/cloud	Management interoperability for cloud systems, developer of the Open Virtualization Framework (OVF).
IEEE P2301, Guide for Cloud Portability and Interoperability Profiles	www.standards.ieee.org/develop/project/2301.html	Standards-based options for application interfaces, portability interfaces, management interfaces, interoperability interfaces, file formats, and operation conventions.
IEEE P2302, Draft Standard for Intercloud Interoperability and Federation.	www.standards.ieee.org/develop/project/2302.html	Protocols for exchanging data, programmatic queries, functions and governance for cloud sharing data or functions.

https://cloud-standards.org/index_title_main_page/



Retos: estandarización

Project Name	URL	Focus
OASIS Identity in the Cloud (IDCloud)	www.oasis-open.org/committees/tc_home.php?wg_abb=rev=id-cloud	Performs risk analysis on collected use cases, also develops guidelines for reducing vulnerabilities.
Open Cloud Computing Interface	www.occi-wg.org	REST-based interfaces for management of cloud resources such as computing, storage, and bandwidth.
Open Cloud Consortium	www.opencloudconsortium.org	Frameworks for interoperability between clouds and the operation of the Open Cloud Testbed.
Open Data Center Alliance	www.opendatacenteralliance.org	Unified customer vision for long-term data center requirements, developing usage models for cloud vendors.
OpenStack	www.openstack.org	Open source software for running private clouds, founded by Rackspace and NASA.
Standards Acceleration to Jumpstart Adoption of Cloud Computing	https://www.nist.gov/itl	Cloud standards are obtained by providing use cases that can be supported on cloud systems. Use cases should show a set of documented and public cloud system specifications.
The Open Group Cloud Work Group	https://collaboration.opengroup.org/cloudcomputing/	Other cloud standards organizations and this cloud work group together to tell enterprises how to implement cloud services in their companies.



PROVEEDORES



> Ejemplos de proveedores



> En el mercado...



> En el mercado...

Figure 1. Magic Quadrant for Cloud Infrastructure and Platform Services



Figure 1: Magic Quadrant for Cloud Infrastructure and Platform Services



Source: Gartner (July 2021)



<https://aws.amazon.com/es/resources/analyst-reports/gartner-mq-cips-2021/?trk=AR-Card>

> AWS en el mundo

★ 26 regiones, 84 zonas, 210 ubicaciones de borde de red y 245 países y territorios



> Microsoft Azure en el mundo

★ más de 60 regiones y 170 puntos de presencia de red



* Three Azure Government region locations undisclosed

<https://azure.microsoft.com/en-gb/global-infrastructure/geographies/#overview>

> Google cloud en el mundo

★ 28 regiones, 85 zonas, 146 ubicaciones de borde de red y más de 200 países y territorios



<https://cloud.google.com/about/locations>

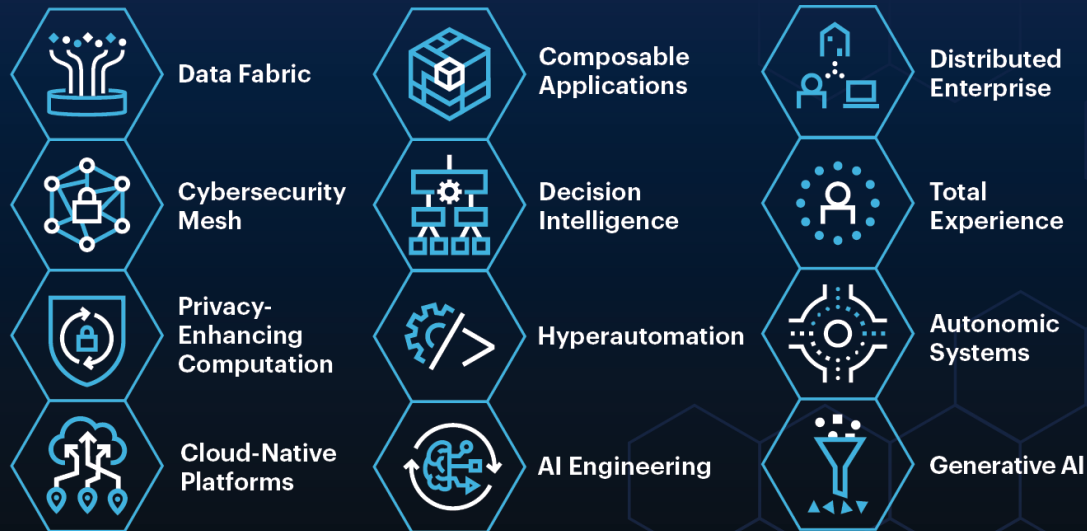
TENDENCIAS

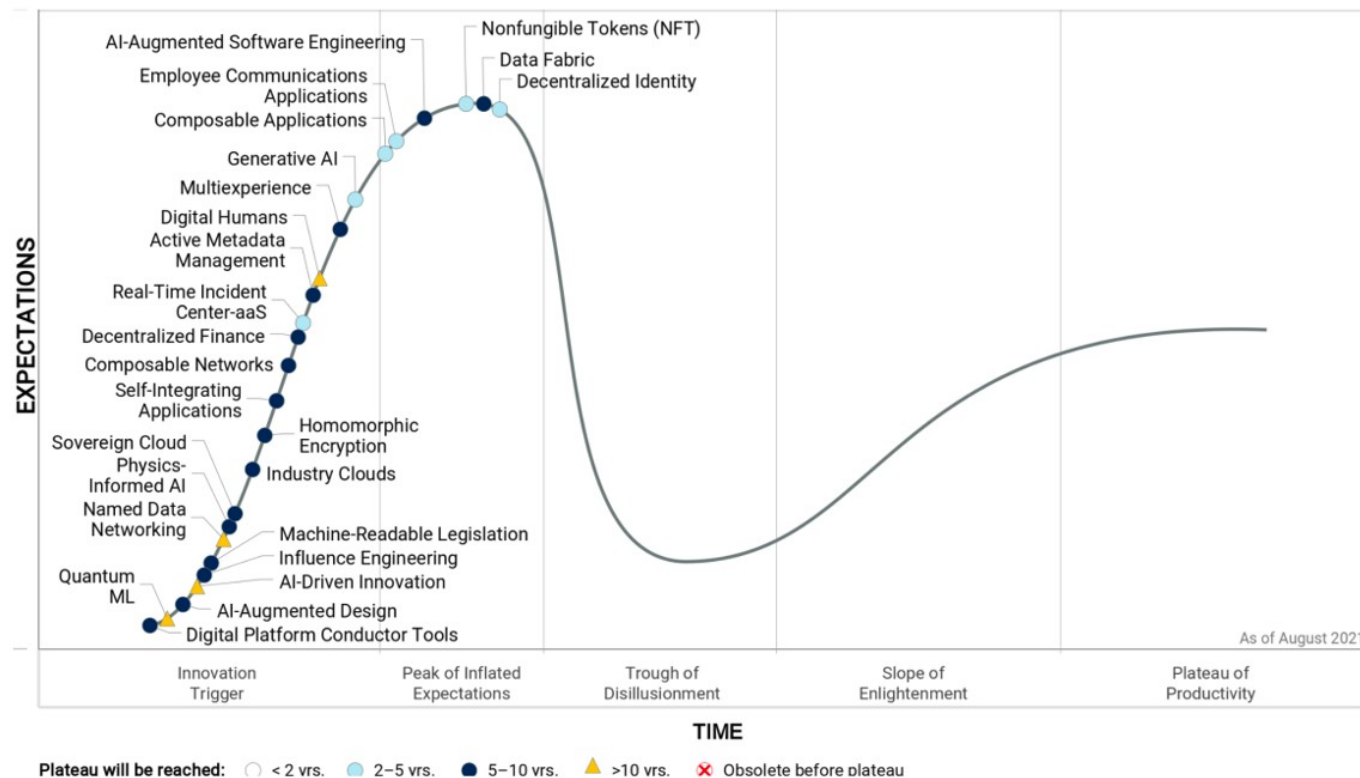


★ Gartner Top Strategic Technology Trends for 2021

- Tendencia 4: Distributed cloud
 - ✓ Nubes que están físicamente distribuidas por muchos sitios
 - ✓ Más cerca del consumidor de servicios
 - ✓ Menos latencias
 - ✓ Según Gartner es el futuro de la nube
- Tendencia 5: Anywhere operations
 - ✓ Que los negocios sigan funcionando en cualquier sitio
 - ✓ Trabajar en remoto
- Tendencia 8: AI engineering
 - ✓ AI + DevOps

Top Strategic Technology Trends for 2022



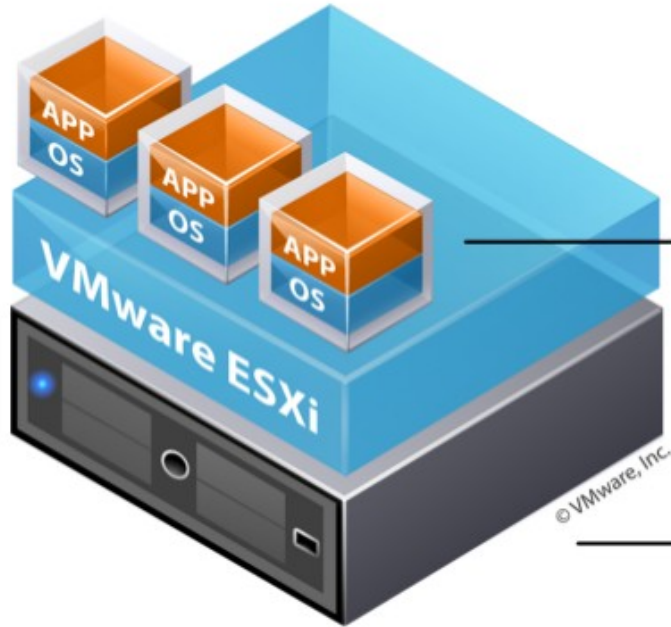


Source: Gartner (August 2021)

747576

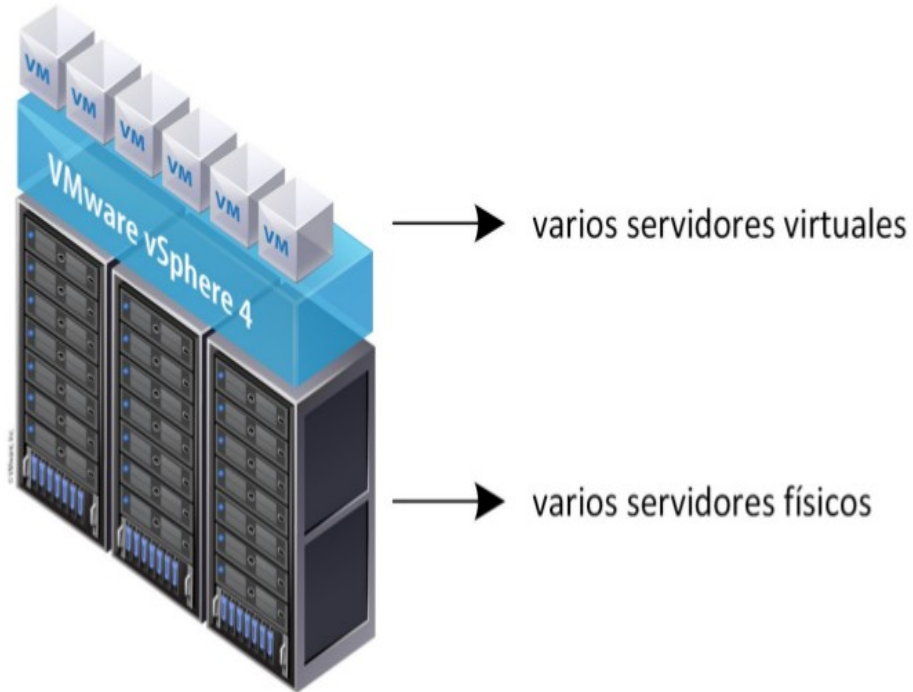
INFRAESTRUCTURA CLOUD





3 servidores virtuales

1 servidor físico







BIG DATA EN CLOUD



> DATABRICKS



★ Plataforma *cloud* de análisis de datos

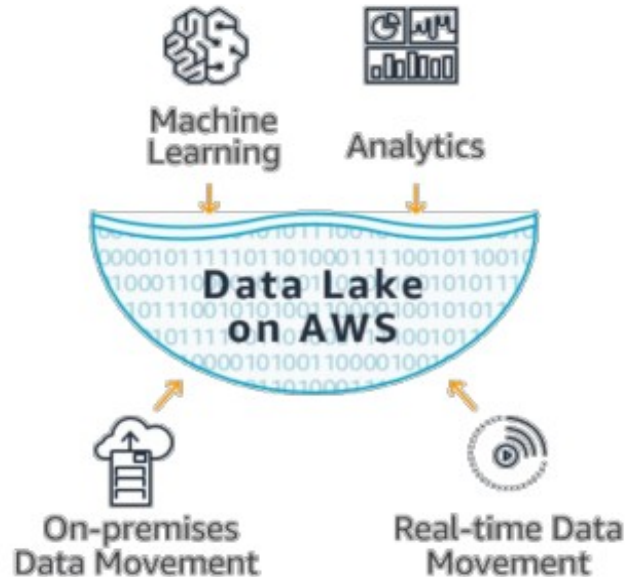
- Big Data: Spark integrado con otras herramientas
- Almacenamiento de datos
- *Machine Learning*: despliegue de modelos
- Colaborativo: cuadernos

★ Desplegado e integrado con la nube: Azure, AWS, GCP, ...

> AWS BIG DATA

Data Lakes and Analytics on AWS

To build your data lakes and analytics solution, AWS provides the most comprehensive set of services to move, store and analyze your data.



Data Movement

Import your data from on-premises, and in real-time.

Data Lake

Store any type of data securely, from gigabytes to exabytes.

Analytics

Analyze your data with a broad selection of analytic tools and engines.

Machine Learning

Forecast future outcomes, and prescribe actions.

> AWS BIG DATA

Data Lake

Once data is ready for the cloud, AWS makes it easy to store data in any format, securely, and at massive scale with Amazon S3 and Amazon Glacier. To make it easy for end users to discover the relevant data to use in their analysis, AWS Glue automatically creates a single catalog that is searchable, and queryable by users.

Object Storage

Amazon S3

Amazon S3 is secure, highly scalable, durable object storage with millisecond latency for data access. S3 is built to store any type of data from anywhere – web sites and mobile apps, corporate applications, and data from IoT sensors or devices. It is built to store and retrieve any amount of data, with unmatched availability, and built from the ground up to deliver 99.999999999% (11 nines) of durability. S3 provides comprehensive security and compliance capabilities that meet even the most stringent regulatory requirements.

Backup and Archive

Amazon Glacier

Amazon Glacier is secure, durable, and extremely low cost storage for long-term backup and archive that can access data in minutes. It is designed to deliver 99.999999999% durability (11 nines), and provides comprehensive security and compliance capabilities that can help meet even the most stringent regulatory requirements. Customers can store data for as little as \$0.004 per gigabyte per month, a significant savings compared to on-premises solutions.

Data Catalog

AWS Glue

AWS Glue is a fully managed service that provides a data catalog to make data in the data lake discoverable, and has the ability to do extract, transform, and load (ETL) to prepare data for analysis. The data catalog is automatically created as a persistent metadata store for all data assets, making all of the data searchable, and queryable in a single view.

> AWS BIG DATA

Analytics

AWS provides the broadest, and most cost-effective set of analytic services that run on the data lake. Each analytic service is purpose-built for a wide range of analytics use cases such as interactive analysis, big data processing using Hadoop and Spark, data warehousing, real-time analytics, operational analytics, dashboards, and visualizations.

Interactive Analytics

Amazon Athena

For interactive analysis, [Amazon Athena](#) makes it easy to analyze data directly in S3 and Glacier using standard SQL queries. Athena is serverless, so there is no infrastructure to setup or manage. You can start querying data instantly, get results in seconds and pay only for the queries you run. Simply point to your data in Amazon S3, define the schema, and start querying using standard SQL. Most results are delivered within seconds.

Real-Time Analytics

Amazon Kinesis

For real-time analytics, [Amazon Kinesis](#) makes it easy to collect, process and analyze streaming data such as IoT telemetry data, application logs, and website clickstreams. This enable you to process, and analyze data as it arrives in your data lake, and respond in real-time instead of having to wait until all your data is collected before the processing can begin.

Big Data Processing

Amazon EMR

For big data processing using the Hadoop and Spark frameworks, [Amazon EMR](#) provides a managed service that makes it easy, fast and cost-effective to process vast amounts data. Amazon EMR supports 19 different open-source projects including [Hadoop](#), [Spark](#), [HBase](#), [Presto](#), and more. Each project is updated in EMR within 30 days of a version release, ensuring you have the latest and greatest from the community.

Operational Analytics

Amazon Elasticsearch Service

For operational analytics such as application monitoring, log analytics and clickstream analytics, [Amazon Elasticsearch Service](#) allows you to search, explore, filter, aggregate, and visualize your data in near real-time. Amazon Elasticsearch Service delivers Elasticsearch's easy-to-use APIs and real-time analytics capabilities alongside the availability, scalability, and security that production workloads require.

Data Warehousing

Amazon Redshift

For data warehousing, [Amazon Redshift](#) provides the ability to run complex, analytic queries against petabytes of structured data, and includes [Redshift Spectrum](#) that runs SQL queries directly against Exabytes of structured or unstructured data in S3 without the need for unnecessary data movement. Amazon Redshift is less than a tenth of the cost of traditional solutions. Start small for just \$0.25 per hour, and scale out to petabytes of data for \$1,000 per terabyte per year.

Dashboards and Visualizations

Amazon QuickSight

For dashboards and visualizations, [Amazon QuickSight](#) provides you a fast, cloud-powered business analytics service, that that makes it easy to build stunning visualizations and rich dashboards that can be accessed from any browser or mobile device.

> AWS BIG DATA

Machine Learning

For predictive analytics use cases, AWS provides a broad set of machine learning services, and tools that run on your data lake on AWS. Our services come from the knowledge and capability we've built up at Amazon, where ML has powered Amazon.com's recommendation engines, supply chain, forecasting, fulfillment centers, and capacity planning.

Frameworks and Interfaces

For expert machine learning practitioners and data scientists, AWS provides [AWS Deep Learning AMIs](#) that make it easy to build deep learning models, and build clusters with ML and DL optimized GPU instances. AWS supports all the major machine learning frameworks, including TensorFlow, Caffe2, and Apache MXNet, so that you can bring or develop any model you choose. These capabilities provide unmatched power, speed, and efficiency that deep learning and machine learning workloads require.

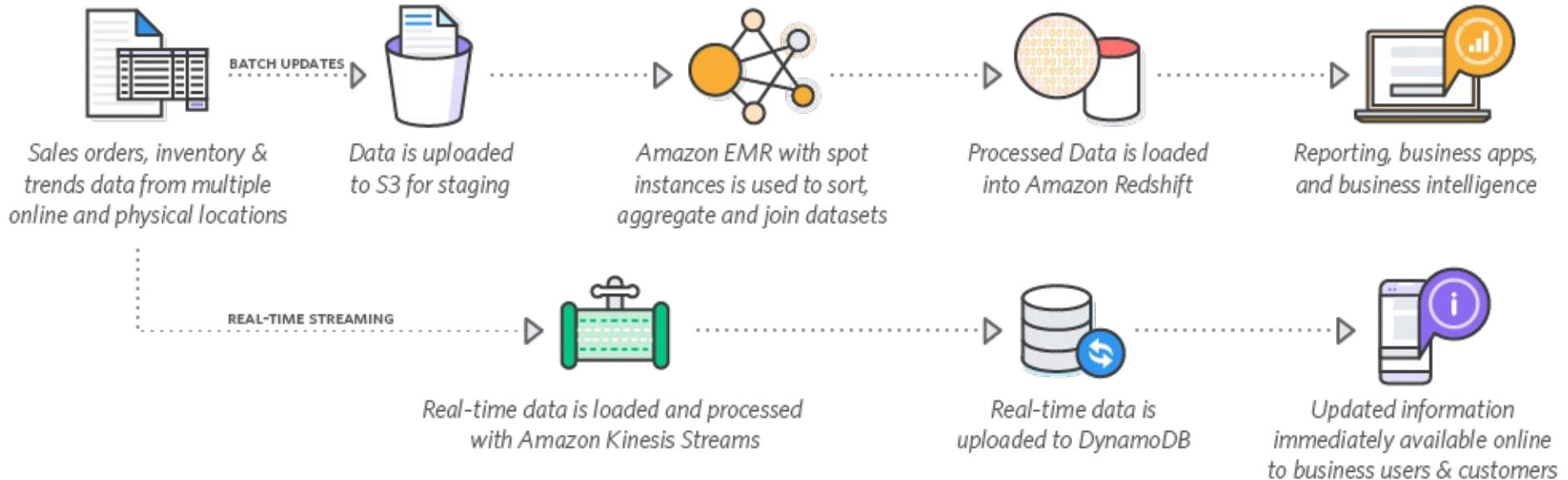
Platform Services

For developers who want to get deep with ML, [Amazon SageMaker](#) is a platform service that makes the entire process of building, training, and deploying ML models easy by providing everything you need to connect to your training data, select, and optimize the best algorithm and framework, and deploy your model on auto-scaling clusters of Amazon EC2. SageMaker also includes hosted Jupyter notebooks that make it is easy to explore, and visualize your training data stored in Amazon S3.

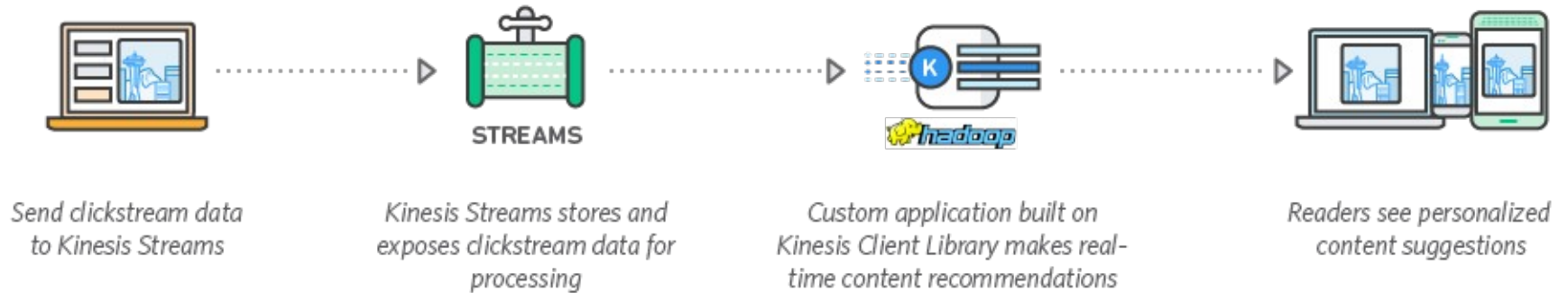
Application Services

For developers who want to plug-in pre-built AI functionality into their apps, AWS provides solution-oriented APIs for computer vision, and natural language processing. These application services lets developers add intelligence to their applications without developing and training their own models.

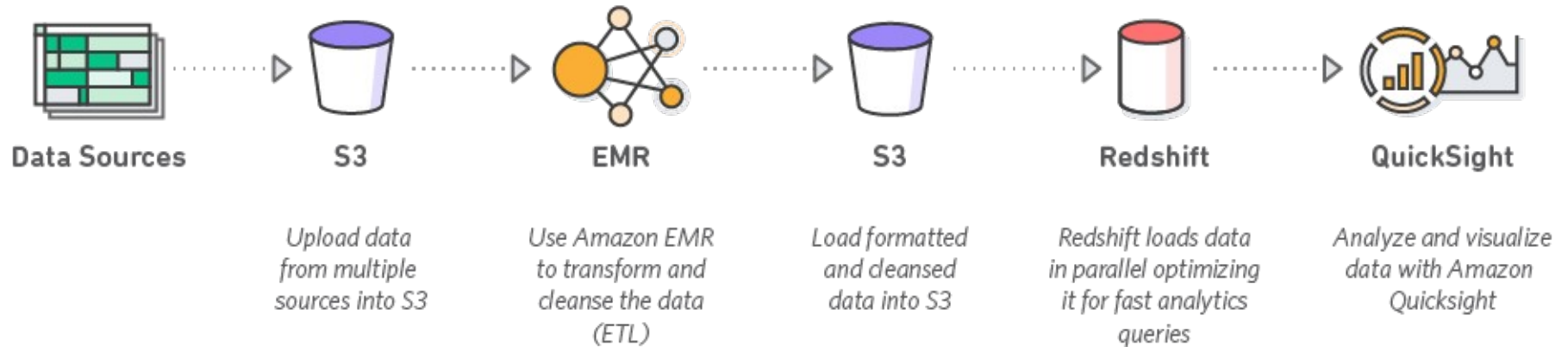
> AWS BIG DATA



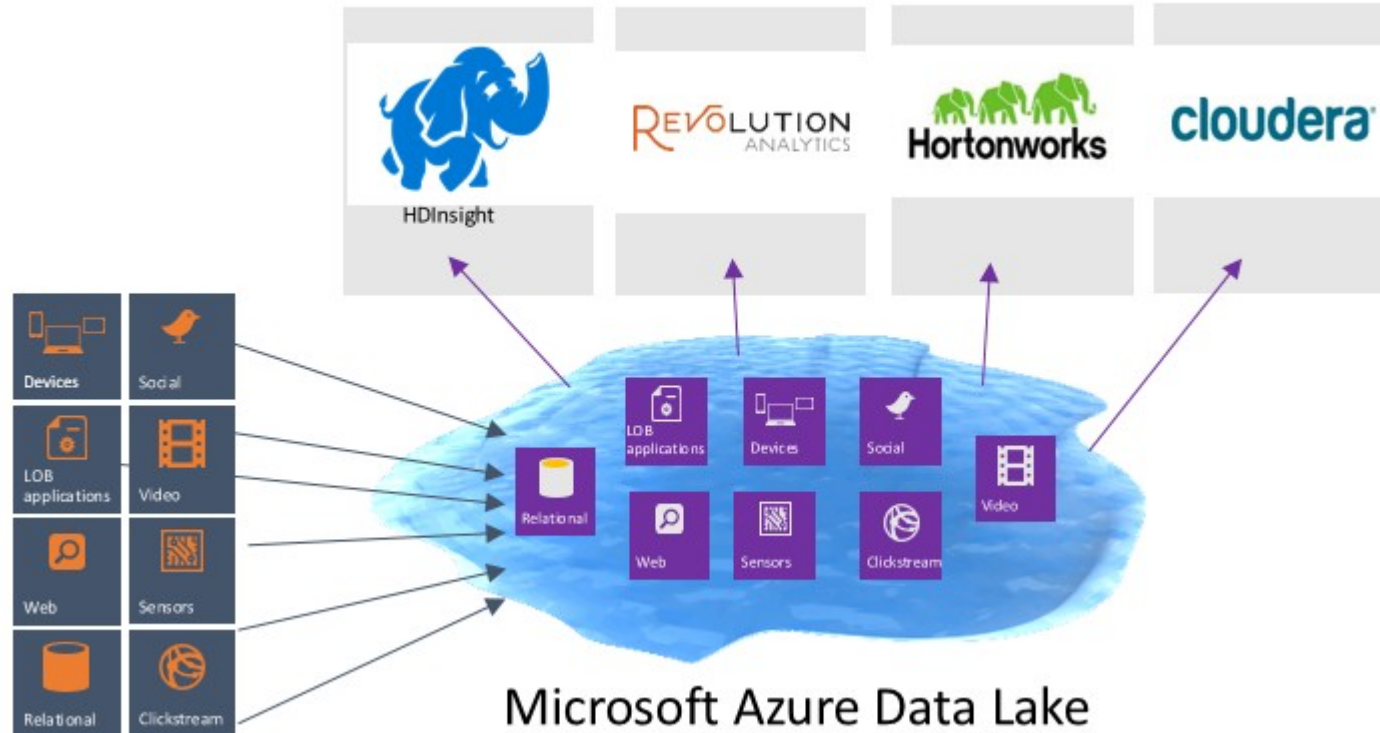
> AWS BIG DATA



> AWS BIG DATA



> Microsoft Azure BIG DATA



> Microsoft Azure BIG DATA

★ ¿Qué incluye HDInsight?



Hadoop



Spark



Kafka



HBase



Hive LLAP



Storm



Machine Learning

★ Y muchos frameworks/tecnologías más:

- Tez
- Zookeeper
- Oozie
- Pig
- ...

> Microsoft Azure BIG DATA

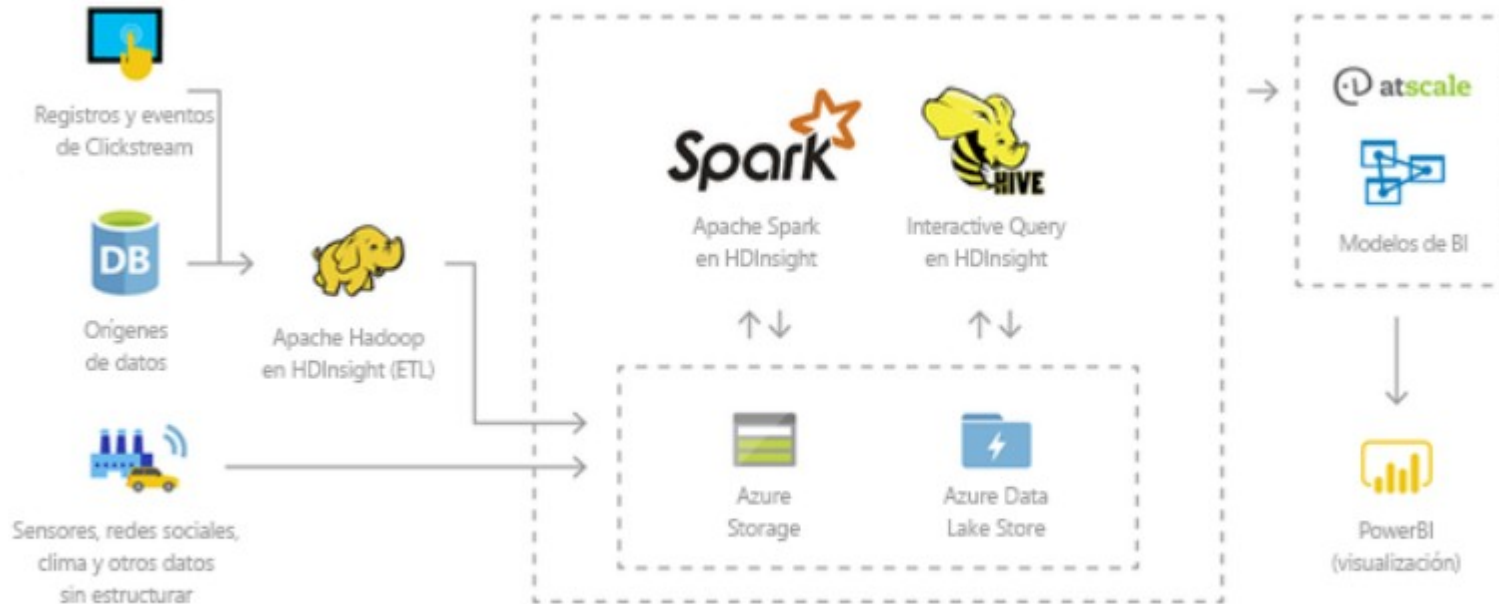


Imagen obtenida de: <https://docs.microsoft.com/es-es/azure/hdinsight/hadoop/apache-hadoop-introduction>

> Microsoft Azure BIG DATA



> Microsoft Azure BIG DATA



Imagen obtenida de: <https://docs.microsoft.com/es-es/azure/hdinsight/hadoop/apache-hadoop-introduction>

Gracias