

El cloud basado en hechos reales

**SYSTRAN ha  
especializado  
la traducción  
automática  
online mediante  
la inteligencia  
artificial**





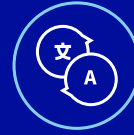
**+ de 30 000**

horas acumuladas  
de cálculo con GPU  
NVIDIA Tesla V100



**50 000**

modelos intermedios  
almacenados en Object Storage  
con un volumen de 30 TB



**5000 millones**

palabras diarias  
de capacidad  
de traducción

## El contexto

SYSTRAN ofrece a sus clientes soluciones profesionales de traducción automática. Con más de 140 pares de idiomas disponibles, SYSTRAN personaliza sus servicios para adaptarlos al contexto de cada cliente. Numerosas empresas internacionales, organismos públicos y agencias de traducción recurren a ellos.

Desde su creación hace más de 50 años, SYSTRAN ha sido siempre pionero en el tratamiento automatizado de la lengua. A finales de 2016, volvió a tomar la delantera lanzando el primer motor profesional de traducción neuronal. Este motor aprovecha los últimos avances que ofrecen las tecnologías de deep learning para mejorar la calidad de la traducción automática.

La empresa, que confió en OVHcloud desde el primer momento, se asoció en 2018 con el proveedor cloud para desarrollar una solución llamada SYSTRAN Marketplace. Esta plataforma comunitaria permite ofrecer los mejores modelos de traducción del mercado, entrenados por expertos multilingües de diversos ámbitos. Los modelos, disponibles a través de herramientas de traducción profesional tanto en la nube como on-premises, están integrados en el sistema de información del cliente.

Para hacer frente a este desafío, SYSTRAN eligió un enfoque comunitario basado en cuatro pilares: la tecnología, los datos, el conocimiento humano y la infraestructura. Su objetivo era ofrecer una solución abierta, responsable, dimensionada para internet y con alta disponibilidad.

# El desafío

Desde 2016, el mundo de la traducción automática se ha transformado por completo. La traducción neuronal —un enfoque procedente de la investigación en inteligencia artificial y, en particular, del deep learning— se ha impuesto como el estándar, sustituyendo a la traducción estadística, basada principalmente en el big data y en la representación de las reglas que rigen las lenguas por parte de expertos.

Esta transición ha ido acompañada de profundos cambios. En el aspecto tecnológico, los algoritmos necesarios no dejan de evolucionar, y provienen directamente de los grandes laboratorios de investigación privados y públicos. Gracias al enfoque neuronal, se ha desarrollado e impuesto una corriente open source que ha facilitado una progresión científica reproducible y un desarrollo industrial casi instantáneo.

Aunque la cantidad de datos necesarios es menor que antes, ahora la calidad de los mismos es fundamental, ya que los modelos neuronales intentarán interpretar cualquier «ruido» como una regla de la lengua. Con el big data, se corre el riesgo de perder de vista que la información utilizada para entrenar los modelos de traducción es producida por traductores humanos y que, aunque esos datos estén publicados en internet, eso no significa que no estén sujetos a derechos de autor. La calidad de un modelo es consecuencia directa de la inversión realizada en esos mismos datos, lo que exige una perfecta trazabilidad. Sin este rigor, sería peligroso confiar en modelos de traducción que podrían estar sesgados por los datos de origen.

El conocimiento humano, que en la época estadística quedó relegado a un segundo plano, recupera así su hegemonía. Aunque los algoritmos son extremadamente potentes, necesitan la supervisión de expertos lingüistas, así como de especialistas de diversos ámbitos.



Por último, el enfoque neuronal ha hecho que los requisitos que exigen las infraestructuras de cómputo cambien radicalmente. Al igual que para cualquier otro algoritmo de deep learning, es necesario utilizar tarjetas gráficas (GPU) durante la fase de entrenamiento de los modelos. En cambio, para la inferencia —el uso de modelos en producción— los algoritmos requieren servidores optimizados para el cálculo y relativamente poca memoria en comparación con las generaciones anteriores. Asimismo, debido al endurecimiento de la reglamentación en materia de protección de los derechos de los usuarios, es necesario prestar especial atención a las infraestructuras que alojan servicios que puedan traducir datos confidenciales.

Más allá de la aparente simplicidad de cada uno de estos cambios —a menudo ilustrada mediante demostraciones de rendimiento en casos de uso extremadamente limitados— es necesario realizar cambios de calado que permitan obtener una cadena de producción a gran escala responsable, transparente y capaz de ofrecer la mejor calidad a todos los sectores. Este enfoque parte de un principio fundamental: reconocer el conocimiento de los distintos actores involucrados y hacer que colaboren para alcanzar la excelencia.

SYSTRAN, por su parte, ha invertido en el open source cofundando en 2016 OpenNMT, un framework de algoritmos de traducción neuronal. Esta tecnología, que a día de hoy es la más popular y activa en su sector, es utilizada por miles de investigadores y fabricantes, que la nutren día a día con sus contribuciones. Gracias a este software puntero, el equipo de I+D de SYSTRAN ha desarrollado soluciones completas de traducción pensadas para los usuarios finales.

Por último, la empresa ha creado un marketplace con distintos servicios que permite a la comunidad de expertos producir y compartir modelos de alta calidad, recibiendo una remuneración directa por sus contribuciones. Para construir esta plataforma era necesario disponer de una infraestructura flexible, robusta y adaptable que ofreciese la potencia de cálculo necesaria para entrenar motores neuronales. El entorno también debía ser escalable para desplegar los modelos en producción, ajustarse a las fluctuaciones de la demanda y respetar el espíritu responsable de este enfoque comunitario... y todo ello a un precio competitivo.

# La solución

## Una plataforma abierta, segura y responsable adecuada para deep learning

*Enseguida vimos que OVHcloud era la mejor elección como partner tecnológico para el alojamiento y la explotación de nuestro marketplace. El espíritu de OVHcloud tenía la misma esencia que el marketplace. Nuestras exigencias en cuanto a flexibilidad y potencia nos condujeron directamente a la solución Public Cloud.*

**Jean Senellart, presidente-director general de SYSTRAN**

## Una solución técnica que aúna potencia, flexibilidad y predictibilidad

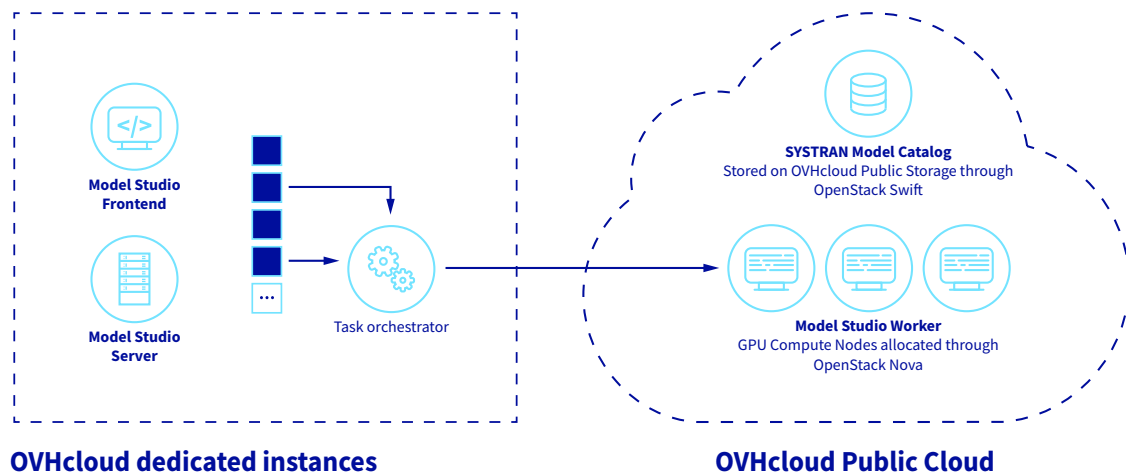
Para llevar el proyecto a buen término, SYSTRAN eligió la solución **Public Cloud**, que permite mantener los costes bajo control y acceder a una gran variedad de servidores y servicios complementarios. Asimismo, proporciona la flexibilidad necesaria para entrenar modelos neuronales bajo demanda y gestionar volúmenes de traducción variables en el tiempo.

SYSTRAN Model Studio —una solución única desarrollada por SYSTRAN para que los expertos lingüistas y los especialistas de otros ámbitos de conocimiento puedan entrenar sus propios modelos de traducción— necesita acceder bajo demanda a los procesadores gráficos (GPU) más potentes del mercado. La disponibilidad instantánea de las instancias de cálculo no representa un problema, ya que el entrenamiento de modelos neuronales se basa en ciclos que pueden durar desde unas horas hasta una semana.

Model Studio es un orquestador de tareas capaz de generar una secuencia de iteraciones que corresponden a un entrenamiento determinado. Utiliza la API Nova de OpenStack para iniciar instancias de cálculo dinámicamente.

En este esquema, la fiabilidad de las instancias es esencial, puesto que un error en una iteración provocaría el fallo del entrenamiento correspondiente y, como resultado, la pérdida de días de cálculo.

Model Studio también requiere una enorme capacidad de almacenamiento: cada iteración de un entrenamiento implica que una red de neuronas es archivada y probada. Cabe destacar que cada modelo representa miles de millones de parámetros, es decir, varios gigabytes almacenados en **Object Storage** a través del servicio Swift de OpenStack organizado en contenedores.



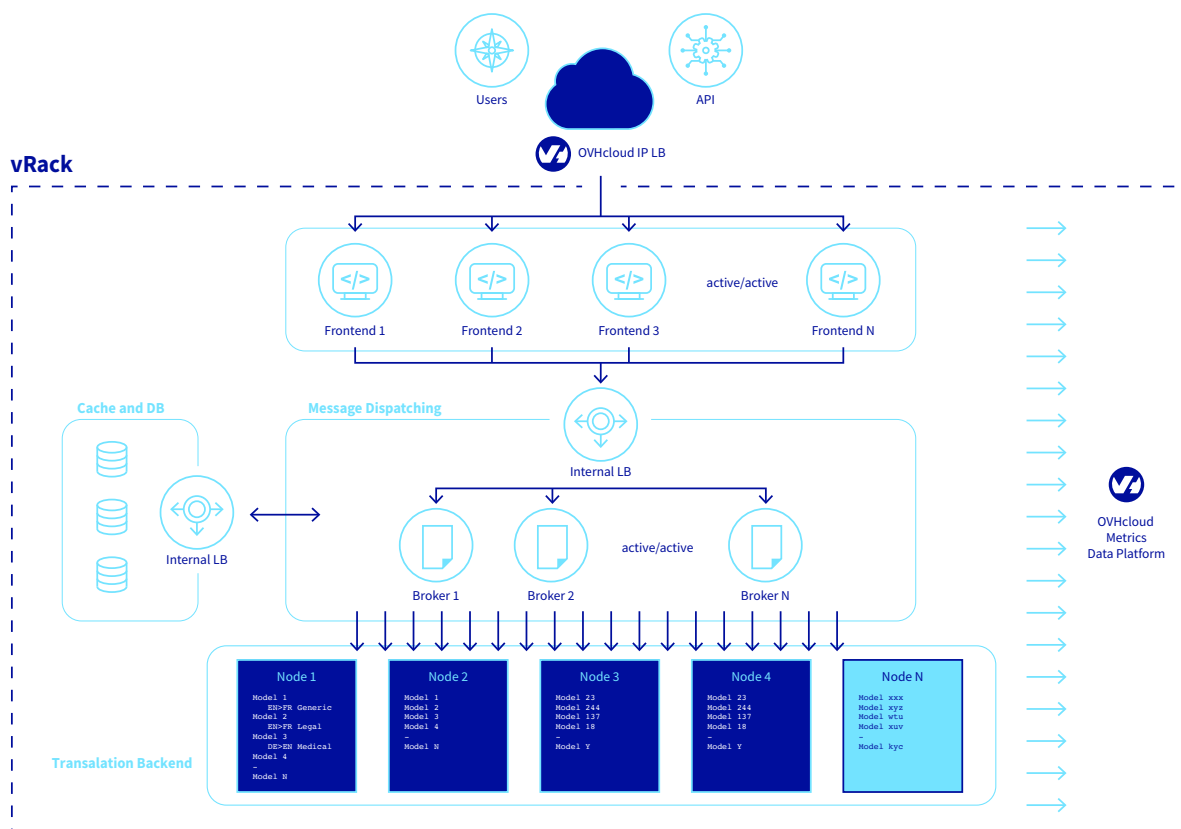
SYSTRAN ha tardado un año en desarrollar esta infraestructura. Durante este tiempo, su equipo ha podido entrenar cientos de modelos utilizando un pool basado en **servidores NVIDIA DGX-1**, pools complementarios de Public Cloud basados en instancias **GPU NVIDIA Tesla V100**. La plataforma ya está a disposición de los «entrenadores» del marketplace para que estos puedan crear sus propios modelos con total autonomía.

En lo relativo a la inferencia, el problema es el contrario: el servicio debe estar disponible durante las 24 horas del día y debe adaptarse al volumen de las peticiones en cualquier momento, teniendo en cuenta que además debe utilizar instancias optimizadas para el cálculo. Asimismo, cada petición debe tratarse en cuestión de milisegundos, y requiere una combinación de instancias estáticas y dinámicas.

El punto de entrada de la infraestructura de la plataforma SYSTRAN Translate es un **balanceador de carga**. Su papel es crucial, ya que reparte la carga entre los distintos servicios alojados en **datacenters** y protege la aplicación contra los **ataques DDoS**. Este dispositivo también asegura la escalabilidad de la infraestructura en caso de pico de carga. Por último, permite garantizar la alta disponibilidad del servicio y optimizar los tiempos de respuesta.



En julio de 2019, la infraestructura estaba formada por 74 instancias de Public Cloud de tipo GPU. Dicha infraestructura está protegida mediante el **vRack**, una interconexión privada exclusiva de OVHcloud. Para ir más allá, el equipo ha añadido un componente dinámico al servicio, basado en **Kubernetes**, que permite ofrecer disponibilidad instantánea y dimensionamiento flexible de la infraestructura. Esta última está monitorizada por la plataforma administrada Metrics Data Platform, que permite la supervisión en tiempo real de cada uno de los componentes y el seguimiento de los tiempos de respuesta y los volúmenes de traducción para todos los pares de lenguas y modelos.



## Una plataforma basada en estándares abiertos

El desarrollo de la infraestructura del marketplace se ha visto facilitado en gran medida por los servicios de OVHcloud. Al disponer de una API open source, los equipos de desarrollo pueden utilizarlos desde el primer momento.

*La elección y la inversión en soluciones open source garantizan a los usuarios finales la mejor tecnología disponible, y evita que los desarrolladores y contribuidores del marketplace se vean atrapados en tecnologías propietarias.*

**Yannick Douzant, director de Productos y Tecnologías de SYSTRAN**

Además de facilitar el uso, este enfoque open source es una parte importante de la filosofía en torno al desarrollo de software que comparten las dos empresas, tanto SYSTRAN, que desarrolla y mantiene todo el código de traducción neuronal del proyecto OpenNMT, como OVHcloud, que ha optado por utilizar numerosos estándares abiertos para su solución Public Cloud.

## Un enfoque responsable

*El compromiso de OVHcloud en materia de ecorresponsabilidad, tanto en el diseño de los servidores y su exclusivo sistema de watercooling como en su política de energía sostenible y de reciclado de los componentes al final de su vida útil, ha sido un criterio decisivo en elección de la infraestructura de nuestro marketplace.*

**Jean Senellart, presidente-director general de SYSTRAN**

En cuanto a los datos, estos están protegidos y cuentan con la garantía de no abandonar el territorio europeo para asegurar el cumplimiento del **Reglamento general de protección de datos** (RGPD).





# El resultado

Gracias a la tecnología utilizada y al asesoramiento de los expertos de OVHcloud, el equipo técnico de SYSTRAN solo necesitó dos semanas para desplegar y publicar el servicio SYSTRAN Translate.

Solo cinco meses después de su lanzamiento, más de un millón de usuarios de 190 países ya han traducido miles de millones de palabras, convirtiéndose en una herramienta muy popular en Europa, y especialmente en Francia, Reino Unido, Bélgica y Alemania.

El servicio de traducción automática ofrece más de 40 idiomas y 400 modelos de traducción. El objetivo es alcanzar los 5000 modelos en el próximo año gracias a la expansión de la comunidad de expertos.

Esto no ha hecho más que empezar, ya que SYSTRAN Translate es solo una parte de una nueva solución destinada a los profesionales: SYSTRAN Marketplace, cuyo objetivo es ofrecerles el mayor catálogo de modelos especializados junto con la gama más amplia de soluciones de traducción desplegadas on-premises o en la nube, en modo privado o público, para satisfacer todo tipo de necesidades, sin limitación de volumen, con la misma calidad.

OVHcloud es un proveedor mundial de cloud hiperescalable que ofrece a las empresas valor y prestaciones de referencia en el sector. El grupo, fundado en 1999, gestiona y mantiene 30 datacenters ubicados en cuatro continentes, despliega su propia red mundial de fibra óptica y controla la totalidad de la cadena de alojamiento. Basándose en sus propias infraestructuras, OVHcloud proporciona soluciones y herramientas simples y potentes que permiten poner la tecnología al servicio de las empresas y revolucionan la forma de trabajar de más de un millón de clientes en todo el mundo. El respeto por las personas, la libertad y la igualdad de oportunidades de acceso a las nuevas tecnologías siempre han sido principios firmemente anclados en la empresa. *Innovation for Freedom.*