Aprovisionamiento de un recurso de servicios de Azure Al

Los servicios de Azure AI incluyen una amplia gama de funcionalidades de inteligencia artificial que puede usar en sus aplicaciones. Para usar cualquiera de los servicios de Azure AI, debe crear los recursos adecuados en una suscripción de Azure para definir un punto de conexión en el que se pueda consumir el servicio, proporcionar claves de acceso para el acceso autenticado y administrar la facturación del uso del servicio por parte de la aplicación.

Recurso de varios servicios

Puede aprovisionar un recurso de Servicios de AI que admita varios servicios de IA diferentes. Por ejemplo, podría crear un único recurso que le permita usar los servicios Idioma de Azure AI, Visión de Azure AI, Voz de Azure AI, entre otros.

Este enfoque le permite administrar un único conjunto de credenciales de acceso para consumir varios servicios en un único punto de conexión y con un único punto de facturación para el uso de todos los servicios.

Recurso de un único servicio

Cada servicio de Servicios de AI se puede aprovisionar individualmente, por ejemplo mediante la creación de distintos recursos de **Idioma de AI** y **Visión de AI** en su suscripción de Azure.

Este enfoque le permite usar puntos de conexión independientes para cada servicio (por ejemplo, para aprovisionarlos en regiones geográficas diferentes) y administrar las credenciales de acceso para cada servicio de forma independiente. También permite administrar la facturación por separado para cada servicio.

Por lo general, los recursos de un solo servicio ofrecen un nivel gratuito (con restricciones de uso), lo que los convierte en una buena opción para probar un servicio antes de usarlo en una aplicación de producción.

Recursos de entrenamiento y predicción

Aunque la mayoría de los servicios de IA se pueden usar a través de un único recurso de Azure, algunos ofrecen (o requieren) recursos independientes para el *entrenamiento* y la *predicción* de modelos. Esto le permite administrar la facturación para el entrenamiento de modelos personalizados de forma separada del consumo de modelos por parte de las aplicaciones, y en la mayoría de los casos le permite utilizar un recurso específico del servicio dedicado para entrenar un modelo, pero un recurso genérico de **servicios de IA** para que el modelo esté disponible para las aplicaciones con fines de inferencia.

Identificación de puntos de conexión y claves

Al aprovisionar un recurso de servicio de Servicios de Azure Al en su suscripción de Azure, está definiendo un punto de conexión a través del cual una aplicación puede consumir el servicio.

Para consumir el servicio a través del punto de conexión, las aplicaciones requieren la siguiente información:

- El URI del punto de conexión. Se trata de la dirección HTTP en la que se puede acceder a la interfaz REST del servicio. La mayoría de los kits de desarrollo de software (SDK) de los servicios de IA usan el URI del punto de conexión para iniciar una conexión al punto de conexión.
- Una clave de suscripción. El acceso al punto de conexión está restringido en función de una clave de suscripción. Las aplicaciones cliente deben proporcionar una clave válida para consumir el servicio. Al aprovisionar un recurso de servicios de AI, se crean dos claves, y las aplicaciones pueden usar cualquiera de ellas. También puede volver a generar las claves según sea necesario para controlar el acceso al recurso.
- La ubicación del recurso. Al aprovisionar un recurso en Azure, normalmente se asigna a una ubicación, que determina el centro de datos de Azure en el que se define el recurso. Aunque la mayoría de los SDK usan el URI del punto de conexión para conectarse al servicio, algunos requieren la ubicación.

Uso de una API REST

Los servicios de Azure Al proporcionan interfaces de programación de aplicaciones (API) REST que las aplicaciones cliente pueden usar para consumir servicios. En la mayoría de los casos, se puede llamar a las funciones de servicio mediante el envío de datos en formato JSON a través de una solicitud HTTP, que puede ser una solicitud POST, PUT o GET en función de la función específica a la que se llame. Los resultados de la función se devuelven al cliente como una respuesta HTTP, a menudo con contenido JSON que encapsula los datos de salida de la función.

El uso de interfaces REST con un punto de conexión HTTP significa que cualquier lenguaje de programación o herramienta capaz de enviar y recibir JSON a través de HTTP se puede usar para consumir servicios de IA. Puede usar lenguajes de programación comunes como Microsoft C#, Python y JavaScript, así como utilidades como Postman y cURL, que pueden ser útiles para las pruebas.

Uso de un SDK

Puede desarrollar una aplicación que use servicios de Azure AI mediante interfaces REST, pero es más fácil crear soluciones más complejas mediante el uso de bibliotecas nativas para el lenguaje de programación en el que está desarrollando la aplicación.

Los kits de desarrollo de software (SDK) para lenguajes de programación comunes abstraen las interfaces REST para la mayoría de los servicios de IA. La disponibilidad del SDK varía según los distintos servicios de IA, pero para la mayoría de los servicios hay un SDK para lenguajes como:

- Microsoft C# (.NET Core)
- Python
- JavaScript (Node.js)
- Go
- Java

Cada SDK incluye paquetes que puede instalar para usar bibliotecas específicas del servicio en el código, y documentación en línea para ayudarle a determinar las clases, métodos y parámetros adecuados que se usan para trabajar con el servicio.