



OFICIOS DIGITALES

PROGRAMACIÓN PYTHON

Introducción a las tecnologías IT con Python

LIC. GUILLERMO ALFARO



Introducción a las tecnologías IT con Python

Introducción

A menudo cuando nos proponemos comenzar el aprendizaje de alguna nueva tecnología o lenguaje de programación que despierta nuestro interés, no nos detenemos a ver en detalle el ámbito de aplicación de lo que queremos aprender.

No es un tema menor ya que saber exactamente cuáles son las aplicaciones posibles de lo que estamos encarando, nos ayuda a elegir lo que realmente necesitamos y aprenderlo convencidos de su utilidad futura.

Es por eso que comenzaremos explorando el maravilloso mundo de la programación con Python, conociendo un poco de las múltiples aplicaciones que este lenguaje tiene y sus usos presentes y futuros, ya que es un lenguaje moderno, versátil y con una inmensa comunidad de programadores detrás.

Python

Python es un lenguaje de programación de alto nivel y propósito general, creado en 1991.

Actualmente es uno de los lenguajes más utilizados principalmente por su simpleza y su versatilidad, lo que hace de este, uno de los lenguajes más indicados para iniciarse en el mundo de la programación.



Los programas en este lenguaje suelen ser bastante compactos, por lo general suelen ser más cortos que el mismo programa escrito en otros lenguajes, como por ejemplo C.

Es un lenguaje interpretado, multiplataforma y orientado a objetos.

Python nos permite la codificación orientada a objetos, en forma imperativa y en menor medida, la programación funcional.

Otro detalle no menor es la gratuidad de su intérprete que tiene versiones para casi cualquier plataforma.

Características

En lo que su versatilidad se refiere, Python nos permite crear, desde aplicaciones de uso local o web hasta entrenar algoritmos de Machine Learning y hoy en día es el lenguaje más utilizado en todo lo que se refiere a inteligencia artificial, redes neuronales y ciencia de datos.

Sus principales características son:

- Sintaxis legible.
- Versatilidad.
- Es interpretado.
- Posee múltiples implementaciones.
- Multiparadigma.
- Posee tipado dinámico.

Que sea de sintaxis legible, significa que podemos leer e interpretar con relativa facilidad el código escrito por otros programadores, captando rápidamente cual es la función del mismo y sus principales elementos.

Cuando hablamos de versatilidad, nos referimos a que podemos desarrollar aplicaciones, ya sea para equipos móviles, computadoras hogareñas, podemos hacer SCRIPTING, levantar servidores y páginas web y aplicar toda su potencia en el análisis de datos o sencillez para representar, mediante código, ideas complejas de manera muy sencilla.



Estas aplicaciones del lenguaje, a su vez, están potenciadas por una enorme comunidad de programadores e investigadores que hay por detrás y que posibilita que tengamos a nuestro alcance un sinfín de documentos, ejemplos e información de todo tipo.

Para quienes no conocen la diferencia entre lenguaje compilado e interpretado, al decir que Python es interpretado, decimos que necesita de un intérprete para traducirle al dispositivo el código que genera el programador.

Esto nos permite entre otras cosas, poder ejecutar el código en bloques sin necesidad de hacer una ejecución completa, permitiendo el avance en nuestro desarrollo de una manera mucho más ágil detectando y corrigiendo errores y verificando si el resultado de la ejecución de dicho bloque es la esperada o no.

Cabe aclarar que, la ejecución de un programa compilado es más rápida que la de un programa interpretado.

Otro de sus atributos es que posee múltiples implementaciones en otros lenguajes, como por ejemplo Java, permitiendo escribir código para ser utilizado en dicho lenguaje o por el contrario, hacer uso de las diferentes librerías de Java.

Un paradigma de programación es un estilo, una forma de escribir código de manera sistemática obedeciendo determinadas reglas o imposiciones de dicho paradigma, lo que nos aporta el beneficio de generar código de manera más ordenada, lo que redunda en las ventajas a la hora de mantener o actualizar nuestros desarrollos.

Python admite trabajar con varios paradigmas, como por ejemplo la programación orientada a objetos, la programación estructurada o la programación funcional.

En este curso en particular, nos vamos a enfocar en lo que es la programación orientada a objetos para poder sacar un mejor provecho de toda la potencia de este lenguaje.

Finalmente, cuando decimos que Python tiene un tipado dinámico, significa que, cuando definimos una variable u otro almacén temporal de datos, no debemos especificar el tipo de datos que va a contener, pudiendo en una misma variable cargar un dato numérico, una cadena de texto o cualquiera de los tipos que maneja el lenguaje.

Resumiendo lo visto hasta, podemos armar un cuadro resumen de las ventajas y desventajas de Python, atributos a tener en cuenta al momento de decidir que herramientas vamos a utilizar para llevar a buen término nuestro desarrollo.



Ventajas	Desventajas
Facilidad de uso	Ejecución lenta
Popularidad	No se tiene control de la gestión de memoria
Open Source	Requiere testing en tiempo de ejecución
Fácil integración con sistemas empresariales	
Desarrollo asíncronico	

Versiones de Python

Python posee dos versiones diferentes del intérprete actualmente en uso

- Python 2.
- Python 3.

Si bien la versión 2 fue deprecada, es decir, ya no tiene más actualizaciones y su uso está totalmente desalentado, se continúa usando ya que hay muchos sistemas armados en base a esta versión aún funcionando y su desarrollo sobre la nueva implica en muchos casos, un desarrollo desde cero.

Big Data

Sin duda una de las primeras y más populares aplicaciones de Python, fue todo el ecosistema referido al mundo del Big Data.

Los informáticos estábamos acostumbrados a trabajar con grandes volúmenes de datos, verificar su veracidad y hacer desarrollos basados en dichos datos, de manera eficiente.

Con el paso de los años las empresas, en particular las de telecomunicaciones y grandes redes sociales, comenzaron a ver la gran cantidad de datos que los usuarios generamos día a día y decidieron invertir en analizar si era posible sacar algún rédito económico de toda esa enorme cantidad de datos, lo que comúnmente llamamos monetizar la información.

Fue así como se fue desarrollando una nueva tecnología, un nuevo paradigma de ver el mundo de los datos que hoy conocemos como Big Data.



Dicha definición empezó a circular por el año 1.999, en una publicación académica acerca de la exploración de datos en tiempo real.

Luego se definieron las bases para trabajar el concepto de Big Data, tomando como pilares fundamentales las tres V, volumen, velocidad y variedad.

Como casi todas las nuevas ideas, esta también fue impulsada por el avance constante de internet y el crecimiento exponencial de usuarios en todo el mundo.

En 2.007 surge la definición actual de Big Data, en 2.010 el entonces presidente ejecutivo de Google, Eric Schmidt, dice en una conferencia que la cantidad de datos que en ese momento se estaban creando cada dos días, era mayor que la creada desde el comienzo de la civilización humana hasta el 2.003.

Si bien es un dato difícil de comprobar, no estaba tan alejado de la realidad y la cantidad de datos que generamos, muchos de ellos de manera inconsciente, es de un volumen muy, pero muy grande a lo que estábamos acostumbrados a manejar.

Finalmente, la ONU formaliza una definición formal en el año 2.012, definiendo al Big Data como:

Volumen masivo de datos, tanto estructurados como no estructurados, los cuales son demasiado grandes y difíciles de procesar con las bases de datos y el software tradicional.

Resumiendo, definimos a Big Data como un gran volumen de datos, estructurados y no estructurados y la tecnología necesaria para procesarlos.

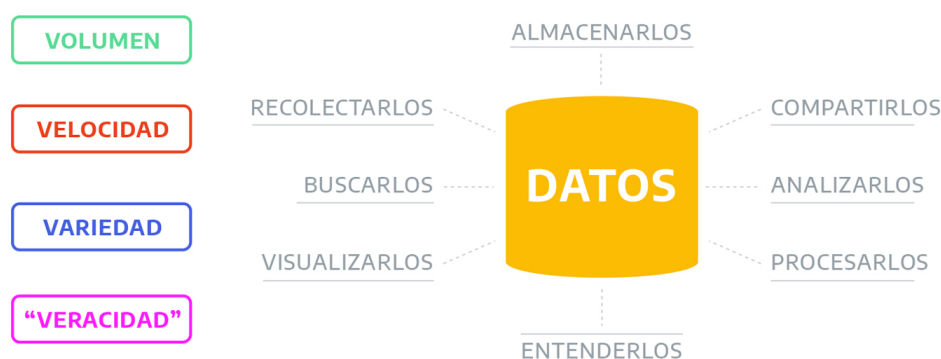
Cuando hablamos de tecnología, nos referimos tanto a la de extracción y almacenamiento como a la de análisis y proceso, que es justamente dónde Python se hace fuerte entre los profesionales del área.

Dentro de las empresas, Big Data es el sector de IT que hace referencia a grandes conjuntos de datos que por la velocidad a la que se generan, la capacidad para tratarlos y los múltiples formatos y fuentes de origen, es necesario procesarlos con mecanismos distintos a los tradicionales.

A las clásicas tres V fundamentales en las primeras ideas del Big Data, sumamos una cuarta referida a la veracidad de la información, completando así el concepto formal de todo el trabajo necesario para desempeñarse en este ámbito.



Un detalle no menor y que contribuyó al desarrollo de esta nueva tecnología, es el hecho de que el precio del almacenamiento de datos fue bajando año a año, llegando a ser muchísimo más barato almacenar información que procesarla, esto implicaba directamente que la captura de información en tiempo real era un hecho posible y relativamente fácil de integrar con las tecnologías disponibles.



Aplicaciones del Big Data en los negocios.

Ya establecida esta tecnología, fue madurando y tanto desarrolladores como usuarios, rápidamente fueron encontrando aplicaciones comerciales derivadas de las características de la nueva información disponible, porque justamente, lo que antes eran solo datos almacenados, pasaron a ser información útil y la demanda de nuevos usos creció casi al ritmo que el Big Data evolucionaba.

Básicamente, en donde tenemos mucha gente generando muchos datos, es ahí donde necesitamos aplicar esta tecnología, para poder relacionarnos en tiempo real con nuestros usuarios y clientes actuales o potenciales, gracias al análisis de los datos con Python, entre los principales lenguajes.

Sistemas de gestión de relacionamiento con clientes (CRM), redes sociales, compras, ventas, oferta de servicios y marketing en tiempo real, son algunos de los principales usos del Big Data.

Esto posibilitó un conocimiento más detallado de los clientes, pasamos de modelo estático a uno dinámico, se crearon las campañas publicitarias personalizadas, la retención de clientes insatisfechos, mejora de experiencia de usuarios, anticipar el comportamiento de las personas en cuanto a sus perfiles de comportamiento y



muchas aplicaciones más se hicieron posibles gracias a lenguajes como Python, que es, dentro del mundo de la ciencia de datos, el más utilizado y los profesionales especializados, son muy requeridos.

Casos de uso

Dentro de los casos de uso más conocidos, nos encontramos con el gigante Amazon, que hace una evaluación continua en tiempo real del comportamiento de sus potenciales clientes, trabajando con modelos analíticos (son los que trabajan sobre datos ya generados) y con modelos predictivos (aquellos que anticipan los gustos y necesidades del cliente).

Otro caso muy conocido es el de la serie de Netflix, HOUSE OF CARDS, una producción que se creó en base a los gustos de los usuarios, estableciendo un patrón de consumo en base a más de 40 millones de usuarios, detectando que era lo que más les atraía y en base a ello, crearon la serie.



Otro detalle es que se crearon diez avances diferentes de la serie y en base al perfil de quien quisiera verlo, le mostraban el que más se ajustaba a sus preferencias.

Las empresas de venta en línea y los bancos trabajan mucho con el tema del comportamiento del usuario para detectar posibles fraudes, algo que también se aplica al uso de las redes sociales.

Bases de datos

Todos hemos trabajado con bases de datos alguna vez, desde una simple agenda telefónica hasta un complicado sistema de facturación.

Cuando hablamos de Big Data, se mencionaron los tipos de datos, un tema no menor cuando hablamos de bases.

Todos conocemos las bases de datos tradicionales, en las que puedo guardar los tipos de datos más usados, ya sean numéricos, de texto, fechas y todo lo que se derive de cada uno.

Para cada dato, conocemos exactamente su tipo y por ello es relativamente sencillo modelar las bases de datos donde serán guardados.

A estos datos los llamamos estructurados.

Pero en los volúmenes de datos que comenzamos a trabajar en Big Data, nos encontramos con la necesidad de almacenar datos como, por ejemplo, fotos, audios, videos, geolocalizaciones y muchos otros. A este tipo de datos los llamamos no estructurados y fue necesario avanzar en el concepto tradicional de base de datos para aprender a trabajar con esta nueva información.

Para quienes no trabajaron nunca con bases de datos tradicionales, los datos se consultaban escribiendo las consultas en un lenguaje particular, el principal para la explotación de este tipo de datos, el SQL.

Es un lenguaje estructurado para la ejecución de consultas a bases de datos.

Pero para los nuevos tipos de datos, necesitábamos algo diferente, fue así como comenzamos a hablar de bases NoSQL (NOT ONLY SQL).

Son sistemas de gestión de bases de datos que difieren del modelo clásico de las bases relacionales, no usan SQL como lenguaje de consulta y los datos almacenados no requieren tener una estructura fija como una tabla.



Surgieron para complementar las bases de datos tradicionales y no para reemplazarlas.

Dentro de estos nuevos ecosistemas de datos, el más conocido y utilizado es HADOOP, un sistema de código abierto que se utiliza para almacenar, procesar en paralelo y analizar grandes volúmenes de datos.

DataWarehouse

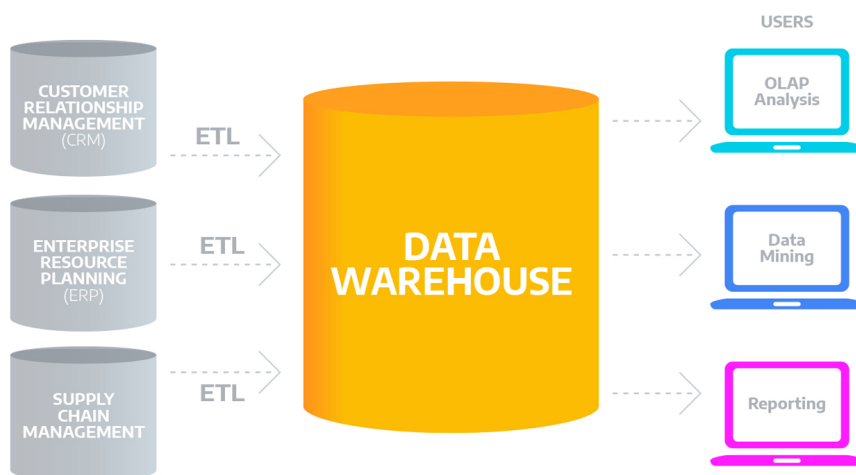
Un DataWarehouse es un repositorio de datos integrado, no volátil, variable en el tiempo, con clara orientación al negocio, organizado de forma tal que facilita el análisis de grandes volúmenes de datos para la toma de decisiones.

En las empresas habitualmente se trabaja con más de un sistema de información, por ejemplo, sistemas de RR. HH, sistemas de cobros y facturación, sistemas de venta, etc.

Si tuviéramos que tomar decisiones en base a los datos de todos ellos, sería un poco complicado, es por ello que existe el DataWarehouse, que se encarga de realizar la extracción de datos de todos estos sistemas, transformarlos a un modelo sencillo orientado al modelo de negocio de la empresa y cargarlos en el para que estén disponibles.

Es una gran base de datos que contiene la información más relevante de todos los sistemas de la compañía, en un formato que sea de fácil lectura y uso.

En base a estos datos se arman reportes, tableros de datos y diferentes estudios o análisis de los principales indicadores de la compañía.





El DataWareHouse es una herramienta fundamental para la toma de decisiones, en cualquier área funcional, basándose en información integrada y global de la empresa.

Facilita mucho el uso y aplicación de técnicas estadísticas de análisis y modelización, para encontrar las relaciones ocultas entre los datos almacenados, y obtener de esa forma, un valor extra para el negocio.

Además, proporciona la capacidad de aprender de los datos del pasado y predecir situaciones futuras, en diversos escenarios.

Cuando decimos que un DataWareHouse (DW) debía ser integrado, nos referíamos a que debe proporcionar información proveniente de sistemas heterogéneos (diferentes tipos de origen de datos) además de contar con procesos de integración de dichos datos y limpieza de la información.

Por no volátil, nos referimos a que los datos deben persistir en el tiempo.

Es, como ya mencionamos, un repositorio de datos históricos, con un tiempo de conservación de los mismos mucho mayor al de los sistemas tradicionales, en general, la fecha es un dato fundamental para poder analizar en el tiempo, trabajar en el estudio de las evoluciones.

La orientación al negocio del DW implica que los datos que allí se almacenan, están organizados y presentados de la manera en que se manejan en el negocio para facilitar su explotación.

Los datos deben tener el nivel de detalle y estructura necesarios para los que toman decisiones.

Inteligencia artificial

Cuando hablamos de inteligencia artificial, nos basamos en dos ramas de trabajo principales, la interpretación del lenguaje natural y la imitación del razonamiento humano para la resolución de problemas, añadiendo la capacidad de memoria y procesamiento que nos ofrecen los equipos de hardware hoy en día.

Python basa su versatilidad en dos premisas muy útiles, la filosofía DRY (Don't Repeat Yourself) y RAD (Rapid Application Development).



Ambas características han hecho de Python el lenguaje por excelencia para el desarrollo de aplicaciones de inteligencia artificial (AI).

Es un lenguaje que permite desarrollar ideas muy complejas de una manera muy sencilla por la sintaxis y semántica que tiene, muy fácil de entender y con una facilidad para el aprendizaje que ha posibilitado que mucha gente lo utilice.

Al lenguaje en si se le pueden agregar de manera muy simple muchas funcionalidades, lo que conocemos habitualmente como módulos.

MACHINE LEARNING (ML)

El MACHINE LEARNING o aprendizaje automático, es un derivado de la inteligencia artificial, cuya principal finalidad es desarrollar técnicas para que las máquinas aprendan una determinada tarea o proceso.

Dentro de lo que hace un desarrollador en el proceso de creación de dichas técnicas, está el explorar los datos para identificar el problema a resolver, y para ello Python nos proporciona herramientas tales como Pandas y Numpy.

El estudio de dichas librerías es parte del curso avanzado de ciencia de datos.

Otro trabajo típico de esta rama es el de graficar los datos explorados, con librerías como por ejemplo MATPLOTLIB.

Una vez llegado a este punto, llega el momento de empezar a aplicar técnicas de ML para lo cual contamos con excelentes librerías como SCIPY y SKLEARN.

Redes neuronales.

A menudo se habla de redes neuronales, como una forma de emular algunas características propias del ser humano, asociándolo solo a la capacidad de memorizar y de asociar ciertos hechos.

Cuando nos encontramos con uno de esos problemas que no pueden ser resueltos con un algoritmo, notamos que todos tienen algo en común, la experiencia o el entrenamiento requeridos para su resolución.

Los seres humanos somos capaces de resolver estos problemas, apelando a la experiencia y a la asociación de la situación a resolver con otras vividas previamente.



Es así como surge esta idea de acudir a la construcción de sistemas que sean capaces de reproducir esta conducta humana para la resolución de problemas.

Resumiendo, las redes neuronales no son más que un modelo artificial y simplificado del cerebro humano, que es el modelo más acertado del que disponemos para emular y desarrollar un sistema que sea capaz de adquirir conocimiento a través de la experiencia.

Es un sistema para el tratamiento de la información, cuya entidad básica de procesamiento está inspirada en la célula fundamental del sistema nervioso humano, la neurona.

Casi todos los procesos del cuerpo humano se relacionan en menor o mayor medida con la actividad de dichas neuronas.

Como sabemos, el pensamiento tiene lugar en el cerebro humano, que consta de billones de neuronas que están interconectadas, por lo tanto, las redes neuronales presentan las siguientes características:

- Son unidades de procesamiento que intercambian datos e información.
- Sirven para reconocer patrones y secuencias de tiempo.
- Son capaces de aprender y mejorar su funcionamiento.

