



Tools

Tarea 01

Luis Florian

0710082

Maestría en Data Science
Ciencia de Datos en Python
Sección U
Universidad Galileo

Guatemala, febrero de 2021

Introducción

Cuando hablamos de desarrollo de software, programación o elaboración de proyectos que involucran código, resulta útil el manejo de herramientas que nos facilitan el trabajo evitando cometer errores comunes y tareas repetitivas.

El siguiente documento presenta una investigación de tres herramientas importantes al momento de programar. Git, Github y Anaconda resultan ser herramientas principales no solo para este curso, si no para el trabajo diario de un programador o en nuestro caso Data Scientist.

Anaconda

La mayoría de cursos, artículos y blogs de Data Science en Python o Python en general mantienen algo en común llamado Anaconda Distribution. Resulta ser una herramienta que brinda muchos beneficios debido a que abarca una serie de aplicaciones, librerías o módulos para el desarrollo de Ciencia de Datos no solo en Python si no también en R.

Una de estas herramientas es llamada **Conda** y nos permite gestionar entornos o ambientes con librerías y versiones específicas para el trabajo de diferentes proyectos. Es común encontrarnos en la necesidad de instalar versiones específicas de Python y librerías para el desarrollo um nuestro proyecto, por lo que resulta conveniente crear un solo entorno para el manejo de cada proyecto.

Instalación de Anaconda

A continuación el procedimiento de Instalación de Anaconda en Linux.

1. Actualizar el Local Package Manager de Linux

```
sudo apt-get update
```

2. Descargar la última versión de Anaconda del sitio oficial

```
https://www.anaconda.com
```

3. Ejecutar el script de instalación

```
bash Anaconda3-2020.02-Linux-x86_64.sh
```

4. Activar y probar la instalación

```
source ~/.bashrc
```

```
conda info
```

5. Crear y activar ambientes

```
conda create --name galileo_env python=3
```

```
conda activate galileo_env
```

6. Crear y activar ambientes

```
conda create --name galileo_env python=3
```

```
conda activate galileo_env
```

Git y Github

Git es una herramienta de control de versiones distribuida que admite flujos de trabajo distribuidos no lineales al proporcionar garantía de datos para desarrollar software de calidad permitiendo revertir errores mediante la disponibilidad de versiones anteriores del proyecto.

En un sistema de control de versiones distribuido, se tiene un repositorio/ servidor central y también repositorios locales. Y los desarrolladores al principio realizan sus cambios en sus repositorios locales y luego envían mediante la instrucción push esos cambios al repositorio central, también actualizan sus repositorios locales con todos los archivos nuevos que se introducen en el repositorio central mediante una operación llamada pull.

Entonces, GitHub será mi repositorio central y git es la herramienta que me permite crear mis repositorios locales. Entonces, git es solo una herramienta de administración de control de versiones y github es una plataforma de alojamiento de código para el sistema de control de versiones.

Github también es una red social para desarrolladores, me permite compartir código y visualizar el trabajo de demás personas.

Comandos básicos de Git

Los siguientes comandos son los más básicos utilizados para la creación y manejo de repositorios.

1. Crear un nuevo repositorio local en el directorio actual

```
git init
```

2. Copiar un repositorio en un servidor remoto

```
git clone nombredeusuario@host:/path/to/repository
```

3. Agregar archivos al área de preparación.

```
git add <Nombre del archivo>
```

4. Crear una instantánea (versión) de los cambios y guardarla en el directorio git

```
git commit -m "mensaje"
```

5. Enviar confirmaciones locales a la rama maestra del repositorio remoto.

```
git push origin <master>
```

Conclusiones

Anaconda es sin duda una de las herramientas principales a instalar para un Data Scientist, contiene una suite de aplicaciones y herramientas útiles y pensadas para Ciencia de Datos.

Git es la herramienta que me permite el manejo de control de versiones distribuidas y Github además de ser mi repositorio/server central en donde guardo mis repositorios y versiones, es una red social de developers en donde puedo encontrar el trabajo de personas individuales y grupos. Es sin duda el sitio principal para buscar algún proyecto si se desea realizar una colaboración.

Referencia

Anusha Sheth, July 29, 2020, Learning Resources of univ.ai. *Machine Learning Prerequisites: Git and Anaconda*.

<https://www.univ.ai/post/machine-learning-prerequisites-git-and-anaconda-resources>

Luigys Toro, *Anaconda Distribution: La Suite más completa para la Ciencia de datos con Python*

<https://blog.desdelinux.net/ciencia-de-datos-con-python/>

Na8, March 23, 2018. *Instalar ambiente de Desarrollo Python Anaconda para Aprendizaje Automático*.

<https://www.aprendemachinelearning.com/instalar-ambiente-de-desarrollo-python-anaconda-para-aprendizaje-automatico/>

Sitio oficial de Anaconda

<https://www.anaconda.com>

Sitio oficial de github

<https://github.com/>

Sitio oficial de git

<https://git-scm.com/>