

Clase 5: Git Remoto y Trabajo en Equipo

Objetivos de la Clase

Al finalizar esta clase, serás capaz de:

- Comprender el concepto de **repositorios remotos** y su papel fundamental en la colaboración en equipo.
- Configurar y utilizar comandos para **sincronizar tu repositorio local** con un repositorio remoto (`git remote`, `git push`, `git pull`, `git fetch`).
- Entender la función de plataformas como **GitHub, GitLab o Bitbucket** en el ciclo de vida del desarrollo de proyectos de datos.
- Dominar la **clonación de repositorios remotos** y el **manejo de ramas en un entorno remoto**.
- Familiarizarte con el **flujo de trabajo colaborativo** basado en Pull Requests (PRs) y revisión de código.

1. Repositorios Remotos y su Configuración

Hasta ahora, hemos trabajado con repositorios Git de forma local en nuestra máquina. Si bien esto es útil para el control de versiones personal, la verdadera potencia de Git se revela cuando colaboramos con otros. Para ello, necesitamos **repositorios remotos**.

1.1. ¿Qué son los Repositorios Remotos?

Un **repositorio remoto** es una versión de tu repositorio que está alojada en un servidor externo, accesible a través de la red (internet o una red privada). Sirve como un **punto central de sincronización y colaboración** para todos los miembros de un equipo.

- **Punto de Verdad:** Es el lugar donde se encuentran las versiones más actualizadas y compartidas del código y los recursos del proyecto (scripts, notebooks, datasets).
- **Colaboración:** Permite que múltiples científicos de datos puedan empujar sus cambios y traer los cambios de otros, facilitando la integración del trabajo.
- **Respaldo:** Actúa como una copia de seguridad externa de tu proyecto, protegiéndote contra la pérdida de datos local.

1.2. Plataformas de Alojamiento de Repositorios Remotos

Existen varias plataformas populares que proporcionan servicios de alojamiento de repositorios Git remotos. Las más utilizadas son:

- **GitHub:** La plataforma más grande y popular para el desarrollo de software colaborativo, especialmente proyectos de código abierto.
- **GitLab:** Ofrece una solución completa de DevOps que incluye control de versiones, CI/CD, seguridad y monitoreo, disponible tanto en la nube como para auto-hospedaje.
- **Bitbucket:** Popular para proyectos privados y equipos que utilizan Jira y Trello de Atlassian, ya que ofrece una fuerte integración.

Aunque usaremos GitHub para nuestros ejemplos, los conceptos y comandos de Git son universales y aplicables a todas estas plataformas.

1.3. Conectar un Repositorio Local con uno Remoto (**git remote**)

Cuando inicias un proyecto con **git init** de forma local y luego decides publicarlo en una plataforma como GitHub, necesitas "enlazar" tu repositorio local con el remoto.

Paso a Paso:

1. **Crea un repositorio vacío en la plataforma remota** (ej. GitHub). No lo inicialices con README, .gitignore o licencia en este paso, ya que tu repositorio local ya tiene contenido.
2. **Obtén la URL del repositorio remoto.** Esta puede ser HTTPS o SSH. Se recomienda SSH por su seguridad y comodidad (sin credenciales repetitivas, que veremos en la Clase 6).
 - Ejemplo de URL HTTPS:
`https://github.com/tu-usuario/mi-proyecto-datos.git`
 - Ejemplo de URL SSH:
`git@github.com:tu-usuario/mi-proyecto-datos.git`
3. **Añade el repositorio remoto a tu configuración local.** Utiliza el comando **git remote add**. Por convención, el primer remoto se llama **origin**.

Unset

```
# En tu repositorio local (donde ya hiciste 'git init' y commits)
git remote add origin
https://github.com/tu-usuario/mi-proyecto-datos.git
# O con SSH
# git remote add origin
git@github.com:tu-usuario/mi-proyecto-datos.git
```

- **origin** es un alias o nombre corto que Git le da a la URL del repositorio remoto. Puedes tener múltiples remotos (ej. **upstream** para el repositorio original si estás en un **fork**).

4. Verifica los remotos configurados:

Unset

```
git remote -v
# Salida esperada:
# origin https://github.com/tu-usuario/mi-proyecto-datos.git
(fetch)
# origin https://github.com/tu-usuario/mi-proyecto-datos.git
(push)
```

2. Clonación de Repositorios Remotos (**git clone**)

Cuando un proyecto ya existe en un repositorio remoto y quieres empezar a trabajar en él en tu máquina local, usas **git clone**. Este comando no solo descarga los archivos, sino que también inicializa un repositorio Git local, configura el remoto llamado **origin** y descarga todo el historial.

2.1. Clonar un Repositorio Existente

Sintaxis:

Unset

```
git clone <URL_del_repositorio_remoto>
```

- Git creará una nueva carpeta con el mismo nombre que el repositorio remoto (ej. **analisis-predictivo/**) y colocará el contenido dentro.
- Se configurará automáticamente un remoto llamado **origin** que apunta a la URL que proporcionaste.

Ejemplo para Científicos de Datos: Descargando un proyecto de análisis de mercado:

Unset

```
git clone https://github.com/empresa-datos/analisis-mercado.git  
# 0 con SSH  
# git clone git@github.com:empresa-datos/analisis-mercado.git
```

2.2. Clonar en un Directorio Específico

Si quieres que la carpeta local tenga un nombre diferente al del repositorio remoto, puedes especificarlo:

Unset

```
git clone <URL_del_repositorio_remoto>  
<nombre_de_la_carpeta_local>
```

Ejemplo:

Unset

```
git clone https://github.com/empresa-datos/analisis-mercado.git  
mi-analisis-local
```

2.3. Clonación Superficial (**--depth**)

Para repositorios muy grandes con un historial extenso, clonar todo puede llevar mucho tiempo y espacio. Si solo necesitas el historial más reciente, puedes usar la opción **--depth**.

Unset

```
git clone --depth=1 <URL_del_repositorio_remoto>
```

- **--depth=1** solo descarga el último commit. Es útil para CI/CD o si solo quieres ver el código actual.
- Puedes traer el historial completo más tarde con **git pull --unshallow**.

3. Sincronización con el Repositorio Remoto

Una vez que tu repositorio local está enlazado con uno remoto, necesitas comandos para intercambiar cambios.

3.1. Traer Cambios del Remoto (`git fetch` y `git pull`)

`git fetch`: Obtener cambios sin fusionar

`git fetch` descarga los objetos (commits, ramas) del repositorio remoto a tu repositorio local, pero **no los integra automáticamente en tu rama de trabajo actual**. Solo actualiza tus referencias remotas (ej. `origin/main`).

- **Uso:** Para ver qué cambios han ocurrido en el remoto sin modificar tu trabajo local.

Unset

```
git fetch origin
# Verás mensajes como: from github.com:usuario/repo * [new
branch] main -> origin/main
```

- Después de `git fetch`, puedes usar `git log origin/main` para ver los commits que están en el remoto pero no en tu rama `main` local.

`git pull`: Obtener y fusionar/rebasar cambios

`git pull` es un comando conveniente que es, en esencia, una combinación de `git fetch` y `git merge` (o `git rebase`). Trae los cambios del remoto y los integra inmediatamente en tu rama local actual.

- **Sintaxis:**

Unset

```
git pull <alias_remoto> <rama_remota>
```

- **Ejemplo:**

Unset

```
git pull origin main # Trae los cambios de 'main' en 'origin' y los fusiona en tu 'main' local.
```

- **Estrategias de pull:** Git puede ser configurado para usar `merge` o `rebase` cuando haces `git pull` y hay historias divergentes.
 - **`git config pull.rebase false` (default antiguo):** Cuando `git pull` detecta que tu rama local y la remota han divergido, intentará hacer un `merge commit`.
 - **`git config pull.rebase true` (recomendado en algunos flujos):** Cuando `git pull` detecta divergencia, intentará hacer un `rebase` de tus commits locales sobre los remotos. Esto mantiene un historial más lineal (veremos `rebase` en la Clase 9).
 - **`git config pull.ff only`:** Solo permite `fast-forward` merges. Fallará si hay divergencia.
 - **Configurar globalmente:** `git config --global pull.rebase true`

Resolución de Conflictos en `git pull`: Si al hacer `git pull` surgen conflictos (porque tus cambios locales colisionan con los cambios remotos), el proceso se pausará, y deberás resolverlos manualmente (como vimos en la Clase 4).

3.2. Enviar Cambios al Remoto (`git push`)

`git push` es el comando para enviar tus commits desde tu repositorio local al repositorio remoto.

- **Sintaxis:**

Unset

```
git push <alias_remoto> <rama_local>
```

- **Ejemplo:**

Unset

```
git push origin main # Envía los commits de tu rama 'main' local  
a la rama 'main' en 'origin'.
```

- **Configurar upstream (rama de seguimiento):** La primera vez que haces **push** de una nueva rama, se recomienda establecer la rama de seguimiento para no tener que especificar **origin <rama>** cada vez.

Unset

```
git push -u origin feat-nueva-visualizacion  
# Después, solo necesitarás: git push
```

Errores Comunes en **git push**:

- **[rejected] ... (non-fast-forward)**: Este es el error más común. Ocurre cuando intentas hacer **push**, pero la rama remota ya tiene commits que tu rama local no tiene (es decir, alguien más hizo **push** antes que tú).
 - **Solución: ¡No uses **git push --force** a menos que sepas exactamente lo que haces!** Primero, debes integrar los cambios remotos en tu rama local.

Unset

```
git pull origin main # Trae y fusiona los cambios remotos  
# Resuelve cualquier conflicto si los hay  
git push origin main # Intenta el push de nuevo
```

- Piensa que tu historial local ha divergido del remoto. Git te pide que resuelvas esa divergencia localmente antes de enviar los cambios al servidor.

4. Manejo de Ramas en Repositorios Remotos

El manejo de ramas en un entorno remoto es crucial para la colaboración en ciencia de datos.

4.1. Crear y Enviar una Rama Remota

1. **Crea la rama localmente** (si aún no lo has hecho):

Unset

```
git switch -c feat-analisis-clusters # Crea y cambia a la rama
```

2. **Haz commits** en tu nueva rama local.
3. **Envía la rama a tu repositorio remoto.** La primera vez que envías una rama, se recomienda usar `-u` (o `--set-upstream`) para vincular tu rama local con la nueva rama remota.

Unset

```
git push -u origin feat-analisis-clusters
# Esto creará 'feat-analisis-clusters' en 'origin' y la
configurará como la rama de seguimiento para tu rama local.
```

4.2. Integrar Cambios en Otra Rama (Flujos Principales)

Una vez que has terminado de trabajar en tu rama (ej., `feat-analisis-clusters`), hay dos formas principales de integrar tus cambios en la rama principal (`main`):

- **Opción A: Merge Local y Push (Menos común en equipos grandes)**

1. Asegúrate de que tu rama `main` local esté actualizada:

Unset

```
git switch main
git pull origin main
```

2. Fusiona tu rama de trabajo en `main` localmente:

Unset

```
git merge feat-analisis-clusters
```



```
# Resuelve conflictos si los hay, luego commit el merge si no es fast-forward.
```

3. Envía los cambios de la fusión a la rama remota `main`:

Unset

```
git push origin main
```

4. Elimina la rama local `feat-analisis-clusters` (y la remota si ya no la necesitas):

Unset

```
git branch -d feat-analisis-clusters  
git push origin --delete feat-analisis-clusters # Eliminar del remoto
```

- **Opción B: Flujo de Pull Request (PR) / Merge Request (MR) (Recomendado en equipos)** Este es el flujo de trabajo colaborativo estándar en la mayoría de los proyectos modernos. Tus cambios se revisan antes de ser integrados.

5. Flujo de Trabajo Colaborativo: Pull Request (PR) y Revisión de Código

Las **Pull Requests (PRs)** (o **Merge Requests** en GitLab) son el corazón de la colaboración moderna con Git en plataformas como GitHub, GitLab o Bitbucket. Una PR es una propuesta para integrar los cambios de una rama a otra, que permite la discusión, revisión y aprobación antes de la fusión.

5.1. ¿Qué es una Pull Request (PR)?

Una PR es esencialmente una **petición para "extraer" (pull) tus cambios de tu rama** (en tu repositorio remoto) a la rama de destino (ej., `main`) de otro repositorio o del mismo.

[Image of Pull Request Flow](#)

Pasos de un Flujo de PR para Científicos de Datos:

1. **Crea una Rama de Característica/Experimento:** Desde la rama `main` (o `develop`), crea una nueva rama local para tu trabajo (ej. `feat-nueva-visualizacion`, `exp-modelo-lineal`).

Unset

```
git switch main
git pull origin main # Asegúrate de que tu main esté actualizada
git switch -c feat-nueva-visualizacion
```

2. **Desarrolla y Haz Commits:** Realiza tus cambios (escribe scripts, ajusta notebooks, genera visualizaciones) y haz commits locales de forma regular en tu rama.

Unset

```
# Modifica tus archivos Python, Jupyter, etc.
git add .
git commit -m "feat: Implementar gráfica de dispersión para correlaciones"
```

3. **Envía tu Rama al Remoto:** Una vez que tu trabajo local esté listo para revisión, envía tu rama a tu repositorio remoto.

Unset

```
git push -u origin feat-nueva-visualizacion
```

4. *La plataforma remota te indicará una URL para abrir la PR.*
5. **Abre la Pull Request (en la Interfaz Web):**
 - Ve a la plataforma (GitHub/GitLab). Verás una notificación para crear una PR desde tu rama recién enviada.
 - Selecciona la **rama base** (donde quieres fusionar, ej. `main`) y la **rama a comparar** (tu rama, ej. `feat-nueva-visualizacion`).
 - Asigna un **título claro** y una **descripción detallada** a tu PR, explicando el problema que resuelve, la nueva funcionalidad o el experimento realizado. Incluye resultados, gráficos (si aplica), o cualquier contexto necesario.

- Puedes abrir la PR como **borrador (Draft PR)** si aún no está lista para revisión final.

6. Revisión de Código y Colaboración:

- Los miembros del equipo revisan tu PR. Pueden dejar comentarios, hacer preguntas, sugerir cambios o incluso añadir commits directamente (si tienen permisos).
- Para científicos de datos, esto puede incluir revisar la lógica del script, la correcta aplicación de algoritmos, la reproducibilidad, la calidad del código Python, o la interpretación de los resultados.
- Si hay peticiones de cambio, realizas los ajustes en tu rama local, haces nuevos commits y `git push` de nuevo. La PR se actualizará automáticamente.

7. Resolución de Conflictos en PRs:

- Si durante la revisión, la rama base (`main`) avanza y tus cambios en la PR entran en conflicto con ella, la plataforma te lo indicará ("This branch has conflicts that must be resolved").
- **Solución:** Debes actualizar tu rama de PR con los cambios de la rama base, resolver los conflictos localmente, y luego empujar los cambios a tu PR.

Unset

```
git switch feat-nueva-visualizacion # Tu rama de PR
git pull origin main                # Trae los cambios de main a
tu PR y resuelve conflictos
git push origin feat-nueva-visualizacion
```

- Esto es un `merge` de `main` en tu rama de PR. También se puede hacer `rebase`.

8. Aprobación y Fusión:

- Una vez que la revisión es satisfactoria y todos los checks automatizados (CI/CD) pasan, alguien con permisos de fusión aprueba y combina la PR.
- Las plataformas ofrecen opciones de fusión:
 - **Merge commit (predeterminado):** Crea un nuevo commit de fusión en la rama base.
 - **Squash and merge:** Combina todos los commits de la PR en un solo commit en la rama base. Ideal para mantener un historial lineal.
 - **Rebase and merge:** Rebasa los commits de la PR sobre la rama base y luego los aplica linealmente.

9. **Limpieza:** Una vez fusionada la PR, la plataforma a menudo te da la opción de eliminar la rama de origen (ej. `feat-nueva-visualizacion`) del repositorio remoto. Tú también puedes eliminarla localmente:

Unset

```
git branch -d feat-nueva-visualizacion
```

6. Actividad Práctica (Laboratorio 5)

Vamos a simular un flujo de trabajo colaborativo utilizando GitHub.

Paso 1: Crear un Repositorio Remoto en GitHub

1. Ve a github.com e inicia sesión (o crea una cuenta).
2. Haz clic en el botón **+** (esquina superior derecha) y selecciona **New repository**.
3. Dale un nombre al repositorio (ej. **mi-analisis-colaborativo-ds**).
4. Marca **Public** o **Private** según tu preferencia.
5. **¡IMPORTANTE! No marques la opción "Add a README file" ni ninguna otra inicialización.** Queremos un repositorio remoto vacío para que podamos empujar nuestro repositorio local a él.
6. Haz clic en **Create repository**.
7. En la página que aparece, copia la **URL HTTPS** (o SSH, si ya configuraste tu clave SSH). Se verá algo como
`https://github.com/tu-usuario/mi-analisis-colaborativo-ds.git`.

Paso 2: Enlazar tu Repositorio Local con el Remoto

1. Abre tu terminal y navega al directorio de tu proyecto local (ej. **clase4-lab-ds** o un nuevo directorio con algunos commits).

Unset

```
# Si no tienes un repo local con commits, créalo y haz un commit inicial:
# mkdir mi-analisis-colaborativo-ds
# cd mi-analisis-colaborativo-ds
# git init
# touch script_base.py
# echo "print('Análisis base v1')" > script_base.py
# git add script_base.py
# git commit -m "feat: Initial base script for analysis"
```

2. Añade el repositorio remoto que creaste en GitHub:

Unset

```
git remote add origin <URL_del_repositorio_remoto_que_copiaste>
# Ejemplo: git remote add origin
https://github.com/tu-usuario/mi-analisis-colaborativo-ds.git
```

3. Verifica que el remoto se ha añadido correctamente:

Unset

```
git remote -v
```

4. Envía tus commits locales a la rama `main` del repositorio remoto por primera vez. Utiliza `-u` para establecer la rama de seguimiento.

Unset

```
git push -u origin main
# Te pedirá tus credenciales de GitHub si usas HTTPS.
```

5. Si tu rama principal local se llama `master` en lugar de `main`, usa `git push -u origin master`.

Paso 3: Simular un Cambio Remoto y Hacer `git pull`

1. **En GitHub.com:** Ve a tu repositorio (`mi-analisis-colaborativo-ds`).
2. Haz clic en `Add file > Create new file`.
3. Nombra el archivo `README.md`.
4. Añade un texto simple: `# Mi Análisis Colaborativo de Datos`.
5. Haz clic en `Commit new file` (en la parte inferior).
6. **En tu Terminal Local:** Intenta enviar un cambio local *sin haber hecho `pull` primero*.
 - Modifica `script_base.py` (o cualquier archivo existente) localmente:

Unset

```
# script_base.py
print('Análisis base v1')
print('Función de preprocesamiento añadida.')
```

- Haz un commit local:

Unset

```
git add script_base.py
git commit -m "feat: Add new preprocessing function"
```

- Intenta hacer push:

Unset

```
git push origin main
# Observarás el error: [rejected] ... (non-fast-forward)
```

7. Ahora, trae los cambios del remoto y resuelve la divergencia:

Unset

```
git pull origin main
# Git hará un merge de los cambios remotos. No debería haber
conflicto si solo se creó README.md.
```

8. Envía tus cambios locales al remoto de nuevo:

Unset

```
git push origin main
# Ahora debería funcionar.
```

Paso 4: Flujo de Trabajo con Ramas Remotas y Pull Request

1. Crea una nueva rama local para una funcionalidad de análisis:

Unset

```
git switch -c feat-nueva-metrica-evaluacion
```

2. Crea un nuevo archivo `metrics.py` en esta rama:

Unset

```
# metrics.py
def calculate_f1_score(precision, recall):
    # Fórmula simplificada
    return 2 * ((precision * recall) / (precision + recall))

if __name__ == '__main__':
    print(f"F1 Score: {calculate_f1_score(0.8, 0.7)}")
```

3. Haz un commit en esta rama:

Unset

```
git add metrics.py
git commit -m "feat: Implement F1 score calculation in metrics.py"
```

4. Envía esta nueva rama a tu repositorio remoto en GitHub:

Unset

```
git push -u origin feat-nueva-metrica-evaluacion
# Observa la URL que te sugiere Git para crear la Pull Request.
```

5. En **GitHub.com**:

- Ve a tu repositorio. Verás un banner amarillo/verde que te invita a **Compare & pull request**. Haz clic en él.
 - Asegúrate de que la rama base sea **main** y la rama de comparación sea **feat-nueva-metrica-evaluacion**.
 - Escribe un **título descriptivo** (ej. "feat: Añadir cálculo de F1-Score") y una **descripción** (explica para qué sirve esta métrica, cómo se usa, etc.).
 - Haz clic en **Create pull request**.
6. **Simular Revisión y Fusión de la PR:**
- Imagina que un colega revisa tu código.
 - Haz clic en la pestaña **Files changed** para ver tus modificaciones.
 - Si todo está bien, haz clic en **Merge pull request** y luego **Confirm merge**.
 - GitHub te dará la opción de **Delete branch** (la rama remota). Haz clic en ella.
7. **En tu Terminal Local:**
- Vuelve a la rama **main** local:

```
Unset  
git switch main
```

- Trae los cambios del remoto (la fusión de la PR):

```
Unset  
git pull origin main  
# Verás que se ha fusionado la rama.
```

- Elimina tu rama local, ya que la PR fue fusionada y la rama remota eliminada:

```
Unset  
git branch -d feat-nueva-metrica-evaluacion
```

7. Posibles Errores y Soluciones

1. **Error: fatal: remote origin already exists. al hacer git remote add origin ...**
 - **Causa:** Ya tienes un remoto llamado **origin** configurado en tu repositorio local. Esto sucede si, por ejemplo, clonaste el repositorio en lugar de iniciarlo, o si lo añadiste previamente.

- **Solución:** Puedes verificar los remotos existentes con `git remote -v`. Si quieres cambiar la URL de `origin`, usa `git remote set-url origin <nueva_url>`. Si quieres eliminarlo para añadirlo de nuevo, usa `git remote rm origin`.
- 2. **Error: `src refspec <rama> does not match any` al hacer `git push origin <rama>`**
 - **Causa:** La rama local que intentas empujar no existe, o su nombre está mal escrito.
 - **Solución:** Verifica los nombres de tus ramas locales con `git branch`. Asegúrate de que estás en la rama correcta o especifica el nombre correcto.
- 3. **Error: `Updates were rejected because the tip of your current branch is behind its remote counterpart`. (el famoso `non-fast-forward` push rechazado)**
 - **Causa:** Alguien más (o tú mismo desde otro lugar) ha hecho `push` de nuevos commits a la rama remota que tú no tienes en tu rama local. Tu historial local y el remoto han divergido.
 - **Solución:** No fuerces el `push` (`--force`) a menos que tengas muy claro lo que haces. Primero, trae los cambios remotos y fúndelos con los tuyos:

Unset

```
git pull origin <nombre_rama>
# Resuelve cualquier conflicto que pueda surgir
git push origin <nombre_rama>
```

- 4. **Conflictos al hacer `git pull` o fusionar una PR:**
 - **Causa:** Tus cambios locales colisionan con los cambios traídos del remoto (misma línea modificada, archivo eliminado/modificado, etc.).
 - **Solución:** Git te indicará los archivos en conflicto. Debes editar esos archivos, resolver las marcas de conflicto (`<<<<<<, =====, >>>>>>`), luego `git add <archivo_resuelto>`, y finalmente `git commit` para completar la fusión.
- 5. **`Permission denied (publickey)` o `Authentication failed` al interactuar con el remoto (clonar, push):**
 - **Causa:** Si usas SSH, tu clave pública no está configurada en GitHub/GitLab, o la clave privada no está en tu sistema/agente SSH. Si usas HTTPS, tus credenciales son incorrectas o no están almacenadas.
 - **Solución:**
 - **Para SSH:** Revisa los pasos de configuración de clave SSH (Clase 6). Asegúrate de que tu clave pública esté en tu perfil de GitHub y que tu clave privada esté en tu máquina y añadida al agente SSH (`ssh-add`).

- **Para HTTPS:** Git puede pedirte las credenciales cada vez, o puedes configurar un "credential helper" para almacenarlas de forma segura.

Conclusión de la Clase 5

¡Excelente trabajo, científicos de datos! Has dado un salto gigantesco en tu dominio de Git. Hoy hemos explorado el mundo de los **repositorios remotos**, aprendiendo a clonar, configurar y sincronizar tus proyectos con plataformas como GitHub. Crucialmente, has practicado el **flujo de trabajo colaborativo** a través de la creación y gestión de **Pull Requests**, una habilidad esencial para trabajar en equipo en cualquier proyecto de ciencia de datos.

En la próxima clase, profundizaremos en las **buenas prácticas** en el uso de Git, que te ayudarán a mantener tu historial de commits limpio, tus ramas organizadas y tu colaboración aún más eficiente.