**Hoja de Trabajo 3**

## Competencias a desarrollar:

Diseña programas paralelos en computadoras heterogéneas usando GPUs, CUDA y OpenACC.

## Instrucciones

De forma individual, realice las siguientes acciones para establecer conexión a la instancia remota de CUDA en AWS. Complete la hoja antes de la siguiente clase del día miércoles (martes 17 a media noche). Suba su código y deje evidencia de todo su trabajo. .

## Conexión a CUDA en AWS (15 pts)

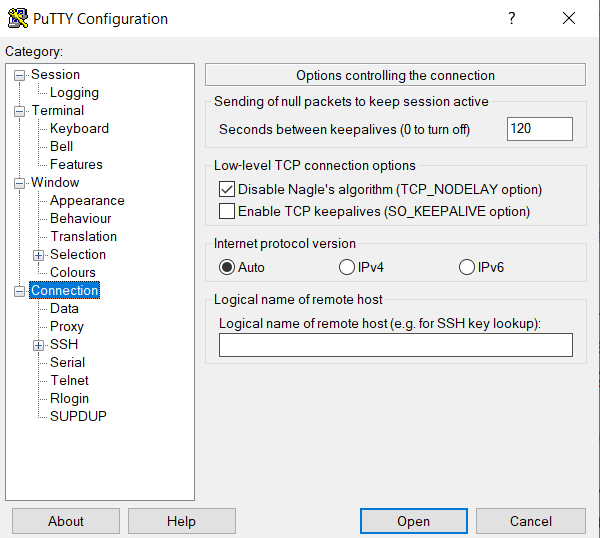
Establezca la conexión SSH a la instancia de CUDA en AWS. Para esto necesita consultar el nombre de DNS público de la instancia, que será puesto en el espacio de la Hoja 3 en Canvas. **Elija la opción adecuada para su sistema**.

**Conexión mediante PuTTY (Windows):**

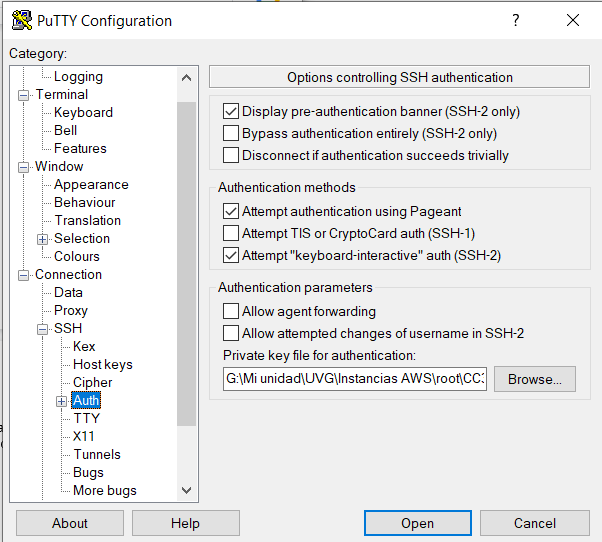
1. Descargue e instale PuTTY.
2. Cree una nueva conexión SSH (default) y coloque el nombre de DNS público para ese día.


       PuTTY configuration - Session
      

1. Configure que su sesión envíe una señal para ‘keepalive’ que evitará que el servidor desconecte la sesión si no hay actividad. En la ventana derecha de **Category**, elija **Connection** y en el campo ‘Connection between keepalives (0 to turn off)’ escriba 120. Este es el número de segundos de intervalo para la señal.



1. Asocie la llave privada a su sesión. En la ventana derecha de **Category**, elija **Connection**, luego **SSH** y **Auth**. En el campo ‘Private key file for authentication:’, presione Browse para buscar la ubicación del archivo con la llave privada.



1. Cada vez que se conecte mediante PuTTY y cambie el nombre de DNS, PuTTY mostrará un mensaje de alerta preguntando si confía en el host al que se está conectando. Acepte el mensaje para guardar la firma digital de la conexión.
2. Guarde su conexión. Cada vez que se conecte, elija la conexión y presione Load para cargar la configuración guardada. Una vez hecho esto, puede modificar el valor de ‘Host name or IP address’ con el DNS para ese día.

**Conexión mediante terminal (Linux / MacOS):**

1. Descargue la llave privada de Canvas (llave.pem) El nombre puede variar, pero el archivo debe terminar en .pem
2. Guarde la llave en un directorio fácil de acceder desde la terminal.
3. Modifique los permisos de la llave:

$ sudo chmod 600 llave.pem

1. Descargue y modifique el script de conexión (script.sh). El nombre puede variar, pero el archivo debe terminar en .sh. Debe actualizar la ruta donde se encuentra la llave y el nombre de usuario.
2. Modifique el script con la ruta correcta a donde se encuentra la llave privada.
3. Modifique los permisos del script:

$ sudo chmod +x script.sh

1. Para conectarse a la instancia, copie el DNS público para ese día. Luego escriba el nombre del script y pegue la dirección pública de DNS: (*\*\* Este nombre deberá cambiarlo cada vez que se conecte, pues el nombre de la instancia cambia toda vez se reinicia el servidor \*\**)

$ ./script.sh <dns-copiado-desde-canvas-para-ese-dia>

1. Al presionar Enter, si la ruta de la llave y sus permisos están bien configurados, recibirá un mensaje indicándo si desea agregar la firma digital. Confirme agregar (Y) para conectarse. Incluya una captura de pantalla de su conexión al servidor.

***(15 PTS) CAPTURA DE PANTALLA DE SU CONEXIÓN INICIAL A LA INSTANCIA***

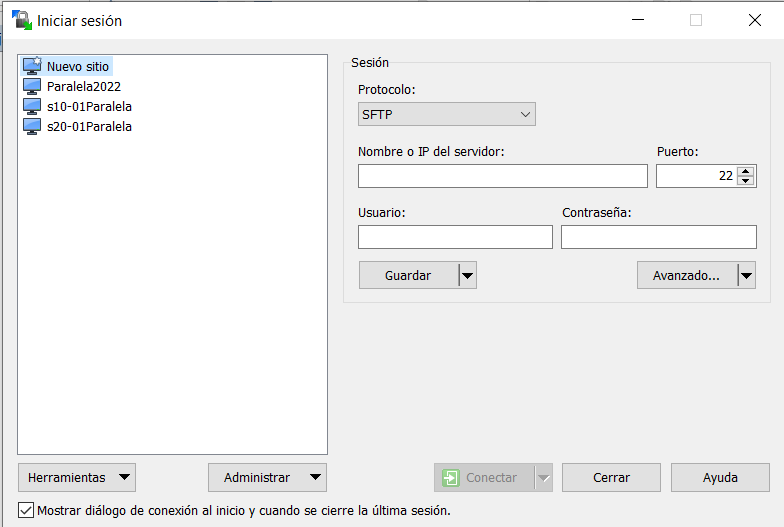
## 

## Transferencia de archivos (15 pts)

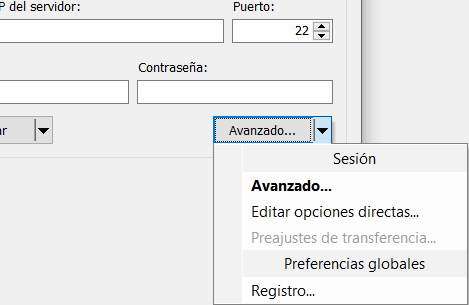
Para transferir archivos al servidor, deberá configurar un cliente SFTP o SCP. Esta acción también necesita una llave privada y el DNS para el día. **Elija la opción adecuada para su sistema**.

**Conexión mediante WinSCP (Windows):**

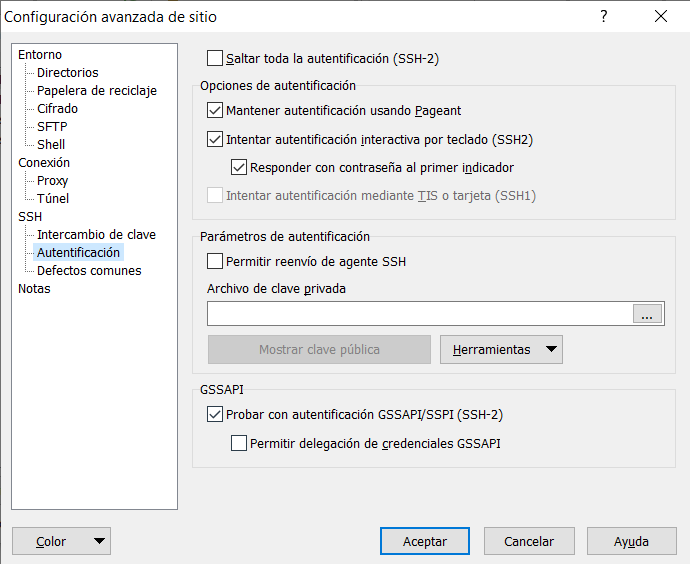
1. Descargue e instale WinSCP (<https://winscp.net/eng/download.php>). Cuando abra la aplicación, desplegará la ventana para crear una nueva conexión. Elija en la ventana derecha la opción **Nuevo sitio**.



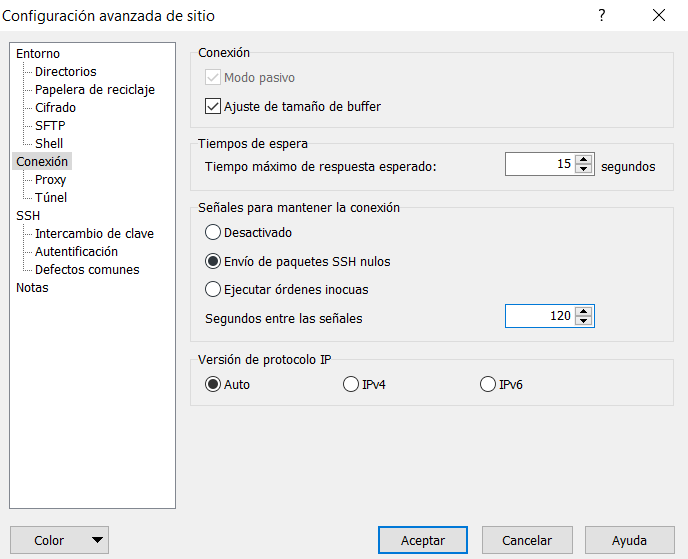
1. Escriba el nombre de DNS para ese día y su usuario asignado.
2. Presione el botón Avanzado.



1. En la siguiente ventana, elija en el panel derecho **SSH** y luego **Autenticación**. En el campo ‘Archivo de clave privada’ presione el botón de tres puntos (...) para seleccionar la ubicación de su llave privada.ppk.



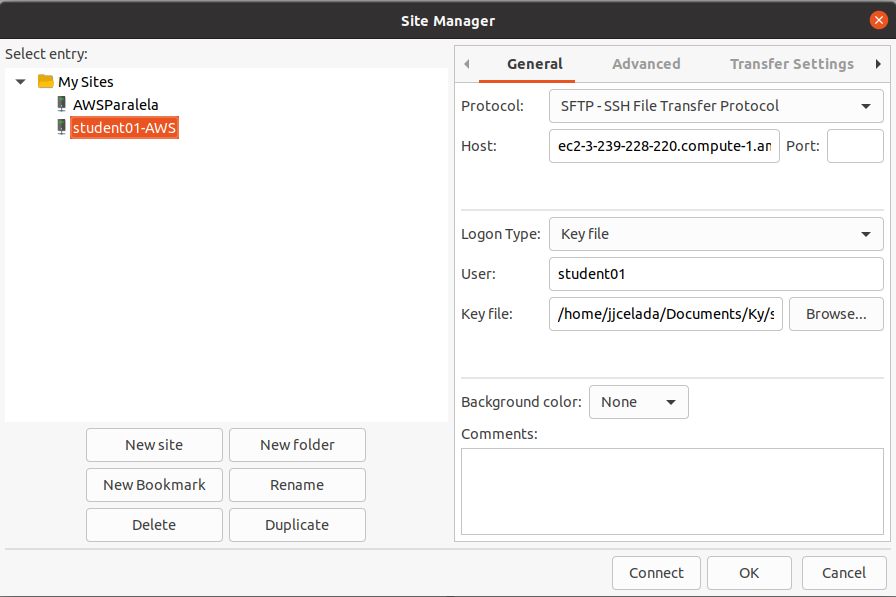
1. Acepte la configuración.
2. Para configurar un intervalo de señales ‘keepalive’ elija del panel derecho en la ventana de ‘Configuración avanzada de sitio’ la opción **Conexión**. Seleccione ‘Envío de paquetes SSH nulos’ y modifique el valor a 120.



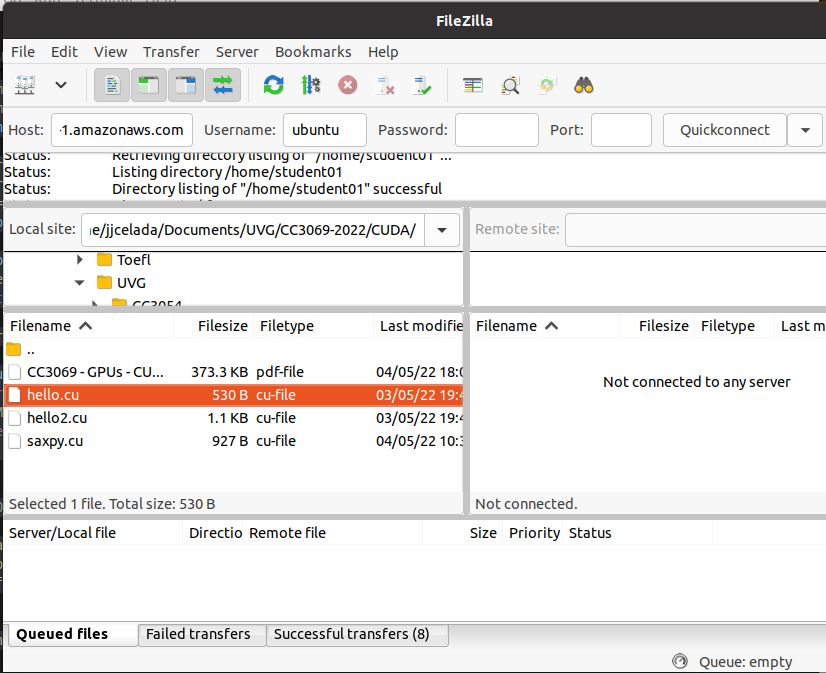
1. Guarde la configuración. La siguiente vez que necesite conectarse, recuerde Editar el nombre o IP del servidor.

**Conexión mediante Filezilla (opcional). Puede usar cualquier otro cliente o línea de comando (Linux / MacOS)**:

* 1. Descargue e instale Filezilla (<https://filezilla-project.org/>) Una vez instalada la aplicación, vaya a File > Site Manager.



* 1. Agregue un nuevo sitio y configure el Protocolo: SFTP
  2. Pegue nuevamente el DNS de la instancia para ese día (*\*\* Este nombre deberá cambiarlo cada vez que se conecte, pues el nombre de la instancia cambia toda vez se reinicia el servidor \*\**).
  3. En Logon Type: seleccione Key File.
  4. Escriba el nombre del usuario que se le brindó en Canvas
  5. Agregue la ruta al Key file mediante Browse para buscar la ruta donde está el archivo llave.ppk Esta llave es la misma que la usada para la conexión mediante SSH, pero en un formato que Filezilla entiende. Ambas puede descargarlas del espacio de Canvas.
  6. Descargue de canvas el archivo hello.cu y súbalo al servidor. Elija el archivo en la ventana derecha y dele click derecho. Seleccione la opción Upload.



***(15 PTS) CAPTURA DE PANTALLA DE SU PRIMER ARCHIVO SUBIDO EXITOSAMENTE AL SERVIDOR***

## 

## hello.cu (35 pts)

El programa hello.cu ilustra la forma básica del modelo de ejecución para CUDA. Realice las siguientes acciones para comprender el efecto de la configuración del kernel y su relación con el Compute Capability de una tarjeta.

* 1. Compile el programa (ignore la advertencia sobre código deprecado):

$ nvcc hello.cu -o hello

* 1. Ejecute el programa. Observe cuántas veces se imprime el mensaje y su conexión con la configuración de la llamada al kernel − hello<<<g,b>>>():

$ ./hello

* 1. Modifique el programa para correr 2 bloques de 1024 hilos. Busque en el despliegue de consola el mensaje del último hilo de la serie (1023).

***(5 PTS) CAPTURA DE PANTALLA DE LA EJECUCIÓN CON 2 BLOQUES DE 1024 HILOS***

* 1. Busque en el sitio de Nvidia el Compute Capability de la tarjeta T4 para Datacenters. Escriba acá el valor de CC y busque la tabla resumen con las características del CC:

***(5 PTS) Compute Capability:***

* 1. Modifique el programa para correr 1 bloque de 2048 hilos. Coloque una captura de pantalla de la salida y busque en la tabla de CC el siguiente dato:

***(5 PTS) Maximum number of threads per block:***

***(10 PTS) CAPTURA DE PANTALLA DE LA EJECUCIÓN CON 1 BLOQUE DE 2048 HILOS***

***(10 PTS) EXPLIQUE EN POCAS PALABRAS EL RESULTADO***

## 

## hello2.cu (35 pts)

El programa hello2.cu ilustra la forma para calcular un identificador global al momento de usar hilos que pertenecen a bloques diferentes. Realice las siguientes acciones para comprender el efecto de la configuración del kernel y su relación con la forma de calcular el ID único de los hilos.

* 1. Descargue, compile y ejecute hello2.cu. Observe la relación de la configuración de la llamada al kernel con la geometría de los hilos y el resultado. Escriba la respuesta a los dos enunciados:
     1. ***Máximo ID de los hilos:***
     2. ***Ejecución de los hilos en orden:***
  2. Observe que la fórmula genérica para cálculo del ID global está en los comentarios. Realice la siguiente modificación al programa y use la fórmula genérica para crear la nueva fórmula:

dim3 g (4,2);

dim3 b (32,16);

hello <<<g, b>>>();

***(10 PTS) FÓRMULA PARA CALCULAR EL ID GLOBAL Y SALIDA DE PANTALLA***

***(5 PTS) CAPTURA DE PANTALLA DE LA NUEVA CONFIGURACIÓN (con el mensaje impreso por el hilo con el máximo ID global)***

* 1. Revise nuevamente la información del Compute Capability respecto a las dimensiones máximas de hilos-bloque en x, y, & z para una grilla. Cree una configuración para lanzar exitosamente el kernel para procesar 100,000 datos. (Sugerencia: busque una configuración que lance como mínimo 100,000 hilos. Modifique el kernel para que imprima el mensaje únicamente si es el ID global máximo)

***(10 PTS) CONFIGURACIÓN USADA***

***(5 PTS) FÓRMULA PARA CALCULAR EL ID GLOBAL Y SALIDA DE PANTALLA***

***(5 PTS) CAPTURA DE PANTALLA DE LA NUEVA CONFIGURACIÓN (con el mensaje impreso por el hilo con el máximo ID global)***

## Verificación

* 1. Incluya capturas de pantalla en la forma como se indica y para la actividad correspondiente. Siempre verifique que la captura muestre la identificación de su host y/o terminal. La terminal debe tener un tamaño de letra legible.
  2. Incluya el código fuente completo en su entrega de Canvas. Cuando un mismo archivo haya tenido varias modificaciones, deje comentado el código original y el de las modificaciones anteriores, dejando sin comentar únicamente el código de la última modificación.