GPU Server Guide @ Robotics Lab

Account Creation:

To get an account for accessing GPU server, please send your following information at <u>ahkhan@kfupm.edu.sa</u>

Required Login Name: Full Name: Mobile Number: Email:

Note: You will be provided access to GPU server for only the current Term. If you need access for long duration then send the expected work duration with the approval from your advisor.

Login to Server:

For login to server, you need a ssh client such as "Putty". It is freely available on internet, you can download putty from http://www.putty.org/

Steps:

1. Open Putty

8	PuTTY Configuration	×
Category: Session Logging Terminal Keyboard Bell Features Window Appearance Behaviour Translation Selection Colours Connection Proxy Telnet Rlogin SSH Auth Tunnels Bugs	Basic options for your PuTTY session Specify your connection by host name or IP address Host Name (or IP address) Port I 23 Protocol: Baw I elnet Rlogin Saved Sessions Load Default Settings Load KAUST KAUST via Proxy NVIDIA K40 Delete Close window on exit: Only on clean exit	
About	<u>O</u> pen <u>C</u> ance	ł

2. Enter IP address of GPU Server: 172.16.0.70 and Select "SSH" in Protocol. Click Open

PuTTY Configuration				
Category: Session Logging Terminal Keyboard Bell Features Window	Basic options for your PuTTY session Specify your connection by host name or IP address Host Name (or IP address) Port 172.16.0.70 Protocol: O Raw Ielnet Rogin SSH			
 Appearance Behaviour Translation Selection Colours Connection Proxy Telnet Rlogin SSH 	Load, save or delete a stored session Sav <u>e</u> d Sessions Default Settings KAUST KAUST via Proxy NVIDIA K40 Delete			
Auth Tunnels Bugs	Close <u>w</u> indow on exit: Always Never Only on clean exit <u>O</u> pen <u>C</u> anc	el		

3. Click Yes on Putty Security Alert. This is only for the first login.

PuTTY Security Alert			
The server's host key is not cached in the registry. You have no guarantee that the server is the computer you think it is. The server's key fingerprint is: ssh-rsa 2048 4c:ae:7c:12:69:fc:bc:b0:f5:0b:37:f5:3a:f3:e4:7c If you trust this host, hit Yes to add the key to PuTTY's cache and carry on connecting. If you want to carry on connecting just once, without adding the key to the cache, hit No. If you do not trust this host, hit Cancel to abandon the connection.			
Yes <u>N</u> o Cancel			

4. Enter User name and Password.



5. You are not logged in to the system.

```
æ
                                                                      _ 🗆 🗙
                                  ayaz@rlk20: ~
login as: ayaz
ayaz@172.16.0.70's password:
Welcome to Ubuntu 14.04.1 LTS (GNU/Linux 3.13.0-30-generic x86 64)
 * Documentation: https://help.ubuntu.com/
  System information as of Mon Sep 15 21:44:23 AST 2014
  System load: 0.01
                                  Processes:
                                                       349
  Usage of /: 2.1% of 853.81GB Users logged in:
  Memory usage: 1%
                                  IP address for eth0: 172.16.0.70
  Swap usage:
               0%
 Graph this data and manage this system at:
   https://landscape.canonical.com/
12 packages can be updated.
7 updates are security updates.
Last login: Sun Sep 14 16:08:13 2014 from 10.13.7.175
ayaz@r1k20:~$
```

Prepared By Ayaz ul Hassan Khan

Transfer Files:

For login to server, you need a scp client such as "WinSCP". It is freely available on internet, you can download WinSCP from http://winscp.net/

1. Open WinSCP

5.	WinSCP Login	- 🗆 ×
New Site KAUST via Proxy mic@KAUST NVIDIA K40	Session File protocol: SFTP V Host name: User name: Save V	Port number: 22 - assword: Advanced
<u>T</u> ools ▼ <u>M</u> anage ▼	E Login	Close Help

2. Enter IP Address of GPU Server: 172.16.0.70 as Host name, Enter your user name and password. Click Login

5	WinSCP Login	- 🗆 ×
Image: New Site Image: KAUST via Proxy Image: mic@KAUST Image: NVIDIA K40	Session File protocol: SFTP V Host name: 172.16.0.70 User name: ayaz Save V	Port number: 22 💌 Password: Advanced 🔽
Tools	ge 🔻 🔁 Login	Close Help

3. Click Yes on Warning for host key. This is only for the first login.

Warning ? ×				
Continue connecting to an unknown server and add its host key to a cache?				
The server's host key was not found in the cache. You have no guarantee that the server is the computer you think it is.				
The server's rsa2 key fingerprint is: ssh-rsa 2048 4c:ae:7c:12:69:fc:bc:b0:f5:0b:37:f5:3a:f3:e4:7c				
If you trust this host, press Yes. To connect without adding host key to the cache, press No. To abandon the connection press Cancel.				
Yes No Cancel Copy Key Help				

4. You can now transfer files from server to local machine or local machine to server by drag and drop the files/folders from left to right or right to left respectively.

ē ⊻			Documents - ayaz@	172.16.0.70 - WinSCP					- 0 ×
Local Mark Files Commands Session Options Remote Help									
冊 🔡 📮 Synchronize 🔲 🐺 👔 🕼 Queue - Transfer Settings Default									
📮 ayaz@172.16.0.70 💣 (New Session								
My documents 🔹	🚰 🔽 (+ + + -) 🗈 🕻	t 🔐 🗶 🐁		🌗 ayaz 🔹 🗧	7	🔷 - 🗈 🖬 🏠 🎜	🖀 Find Files	20	
👔 🔐 Upload 🎲 📝 Edit 💲	K 🛃 🕞 Properties 📑 🕞	+ - V		Download 🙀 📝	Edit 🗙 🚮 🛛	Properties 📑 🕞 🛙	+ - 4		
C:\Users\Ayaz-ul-Hassan\Do	cuments			/home/ayaz					
Name Êxt	Size Type	Changed	^	Name Êxt	Size	Changed	Rights	Owner	
€.	Parent directory	6/30/2014 1:55:37 AM		<u>.</u>		9/15/2014 3:20:58 PM	rwxr-xr-x	root	
퉬 Arena Visual Designer	File folder	4/20/2014 5:56:31 PM		🐌 .cache		9/9/2014 4:42:02 PM	rwx	ayaz	
퉬 Bandicam	File folder	4/18/2013 4:12:37 PM		🌗 bin		9/9/2014 5:04:36 PM	rwxrwxr-x	ayaz	
퉬 Bluetooth Exchange F	File folder	11/28/2013 2:10:07 PM		퉬 cuda-samples		9/9/2014 5:04:23 PM	rwxrwxr-x	ayaz	
퉬 Custom Office Templ	File folder	10/29/2013 5:47:02 PM		.bash_history	1,494 B	9/14/2014 4:08:20 PM	rw	ayaz	
퉬 Fax	File folder	7/17/2013 5:06:34 PM		.bash_logout	220 B	9/9/2014 3:35:29 PM	rw-rr	ayaz	
퉬 GitHub	File folder	3/26/2014 7:23:47 AM		.bashrc	3,637 B	9/9/2014 3:35:29 PM	rw-rr	ayaz	
퉬 IISExpress	File folder	4/24/2013 9:38:11 AM		.profile	675 B	9/9/2014 3:35:29 PM	rw-rr	ayaz	
Integration Services S	File folder	6/12/2013 9:56:02 AM		.Xauthority	204 B	9/14/2014 4:08:13 PM	rw	ayaz	
Integration Services S	File folder	6/12/2013 9:55:50 AM							
퉬 MobaXterm	File folder	11/24/2013 9:00:26 AM							
퉬 Mobile Genie	File folder	1/5/2014 10:59:13 AM							
🛗 My Shapes	File folder	11/17/2013 11:46:50 AM							
퉬 My Web Sites	File folder	4/24/2013 9:38:11 AM							
NetBeansProjects	File folder	4/29/2013 5:36:37 PM							
퉬 Outlook Files	File folder	9/21/2013 8:20:42 AM							
퉬 SAP	File folder	2/23/2014 5:08:01 PM							
Scanned Documents	File folder	1/19/2014 10:52:00 AM							
퉬 SQL Server Managem	File folder	8/27/2014 2:56:41 PM							
퉬 Toad Data Modeler	File folder	5/6/2014 4:34:46 PM							
퉬 ViberDownloads	File folder	6/30/2014 1:55:39 AM							
퉬 Virtual Machines	File folder	6/20/2013 10:54:03 AM							
퉬 Visual Studio 2005	File folder	6/15/2013 8:24:54 AM							
퉬 Visual Studio 2008	File folder	7/10/2013 6:04:59 PM							
0 B of 863 KiB in 0 of 38	P2 2 11	*****	•	0 B of 6,230 B in 0 of 8					
									SFTP-3 🗐 0:01:53

CUDA Example:

```
Kernel File: kernel.cu
__global__ void matrix_scale(float *C, float const* __restrict__ A, int scale, int N)
{
  int tid = threadIdx.x;
  int bid = blockIdx.x;
  int ij = bid * BLOCKSIZE + tid;
  int i = (ij / N) * MERGE_LEVEL;
  int j = (ij % N) * SKEW_LEVEL;
  for (int m = 0; m < MERGE LEVEL; m++)
    for (int n = 0; n < SKEW_LEVEL; n++)</pre>
      C[((i + m)) * N + ((j + n))] = scale * A[((i + m)) * N + ((j + n))];
}
Main File: main.cu
#include<stdlib.h>
#include<stdio.h>
#include<string.h>
#include<math.h>
#include<time.h>
#include<cuda.h>
void checkCudaError(const char *msg) {
  cudaError_t err = cudaGetLastError();
  if (cudaSuccess != err) {
    printf("%s(%i) : CUDA error : %s : (%d) %s\n", __FILE__, __LINE__, msg, (int) err,
cudaGetErrorString(err));
    exit(-1);
  }
}
#include "params.h"
#include "kernel.cu"
int main(int argc, char *argv[]) {
  int N = 1024;
  int GPU = 0;
  if (argc > 1)N = atoi(argv[1]);
  if (argc > 2)GPU = atoi(argv[2]);
```

```
cudaSetDevice(GPU);
float *A, *C;
int memsize = N * N * sizeof (float);
cudaMallocManaged(&A, memsize);
cudaMallocManaged(&C, memsize);
```

```
A[0] = 1;
```

dim3 threads(BLOCKSIZE, 1); dim3 grid(N * N / BLOCKSIZE / MERGE_LEVEL / SKEW_LEVEL, 1);

float time; cudaEvent_t start, stop; // variables to record time of kernel start and stop

// pre-requisite to collect timings at kernel start and stop events
cudaEventCreate(&start);
cudaEventCreate(&stop);

//record the time at kernel start
cudaEventRecord(start, 0);

matrix_scale << <grid, threads >> >(C, A, 3.0, N); cudaDeviceSynchronize();

//record the time at kernel stop
cudaEventRecord(stop, 0);
cudaEventSynchronize(stop);

//calculate the time using start and stop timings
cudaEventElapsedTime(&time, start, stop);

printf("kernel execution time = %f sec\n", time * 1e-3);

```
printf("A[0] = %f, C[0] = %f\n", A[0], C[0]);
printf("End of Program\n");
cudaFree(A);
cudaFree(C);
cudaThreadExit();
```

}

Compiling and Running the Example:

Steps:

1. Goto the source directory containing kernel file, main file, other headers, and Makefile:

cd test_program/

2. To compile the program, use the Make utility (Makefile is provided in the example package):

make

3. To run the program, execute following command:

./main

Note: Example code with Makefile can be downloaded from the following link: <u>https://dl.dropboxusercontent.com/u/13524969/test_program.tgz</u>

Before running your CUDA program, make sure that no one else is using GPUs at the same time so there should not be any conflict among different cuda kernels. You can check this by running following command: nvidia-smi

🚰 ayaz@rlk20: ~ 🧧					
ayaz@rlk20:~\$ nvidia-smi Tue Sep 16 15:08:50 2014	^				
NVIDIA-SMI 340.29 Driver Version: 340.29	+				
GPU Name Persistence-M Bus-Id Disp.A Fan Temp Perf Pwr:Usage/Cap Memory-Usage	Volatile Uncorr. ECC GPU-Util Compute M.				
0 Tesla K20Xm On 0000:02:00.0 Off N/A 26C P8 18W / 235W 15MiB / 6143MiB					
1 Tesla K20Xm On 0000:03:00.0 Off N/A 32C P8 18W / 235W 15MiB / 6143MiB	0ff 0% Default				
2 Tesla K20Xm On 0000:84:00.0 Off N/A 24C P8 18W / 235W 15MiB / 6143MiB	Off 0% Default				
· · · · ·	+				
Compute processes: GPU Memory GPU PID Process name Usage					
No running compute processes found					
ayaz@rlk20:~\$					

To run the kernel on a particular GPU device, you need to use following API function:

cudaSetDevice(GPU);

Here, GPU is the ID of GPU to be used. It can be 0, 1, or 2.

For any help regarding GPU Server and CUDA: Contact Person: Ayaz ul Hassan Khan Email: <u>ahkhan@kfupm.edu.sa</u> Robotics Lab Available Hours: UT 3:00 PM – 5:00 PM