業界最安値の GPUaaS 「GPUSOROBAN」を実際に試してみた

株式会社 PALTEK

業界最安値の GPUaaS 「GPUSOROBAN」を実際に試してみた

発行日：2022 年 6 月 27 日

【概要】

本資料は最新の GPU が使える上に、とても安いと噂の「GPUSOROBAN」がどんな感じのものなのか、CUDA 初心者が 3 日間の無償評価で会員登録から CUDA のサンプルを動かすまでのレポートです。

「GPUSOROBAN」は株式会社ハイレゾが運営する AI 開発や高解像度映像のレンダリング、交通インフラや医療などのシミュレーションなどに活用される GPU リソースを、低コスト・定額料金で利用できる業界最安値の GPU クラウドサービスです。

本資料はインスタンスを使い始める入り口までですが、今回は CUDA が使える状態まで進めておき、次回の資料では何か動かしてみたいと思います。

Windows でのインスタンスの作成から接続までの手順は「GPUSOROBAN」公式 Web サイトの「SOROBAN で仮想インスタンスを作成するには～WINDOWS10 の場合～」にも掲載されています。

【目次】

1. 利用前の確認

弊社内の「GPUSOROBAN」担当と連絡を取り、無償評価をしたいことを伝えました。インスタンスにインストールされている CUDA のバージョンが分からなかったので、それも聞いてみたのですが、そこで Docker についての説明もありました。
なお、無償評価については、以下問合せ画面から問合せていただくとスムーズかと思います。
URL：https://www.paltek.co.jp/solution/gpucloud/index.html

◆サマリ

・評価開始希望日を伝えると、その日から 3 日間が無償評価期間になる
・会員登録はいつでもできる
・インスタンスの作成は評価開始日から可能になる
・CUDA のバージョンは 11.1
・Docker を使うには仮想インスタンスではなく、ベアメタルインスタンスでの提供になる

現在の私の開発環境は、CUDA が 11.4、環境は Docker で作っています。Docker を使いたかったのでベアメタルインスタンスを相談中です。ベアメタルインスタンスは仮想インスタンスのように即座に使用可能とはいかないので仮想インスタンスで進めます。

2. 会員登録

① サイトへアクセス

「GPUSOROBAN」の Web サイトにアクセスして「新規申し込み」をクリックします。

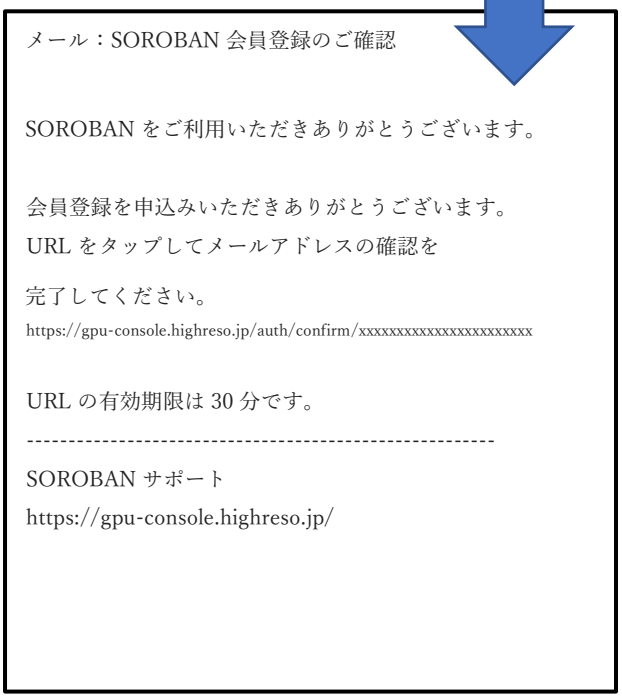
URL：https://soroban.highreso.jp/

① 会員登録

会員登録画面が表示されますので、名前、メールアドレス、パスワードを入力し、利用規約に同意をチェックして「登録」ボタンを押します。

次に、登録完了の表示に切り替わり、登録したメールアドレスに登録認証メールが送信されます。

最後に届いたメールに記載されている URL をクリックすると登録完了です。



② ログイン

下記の URL よりログインしてみましょう。

ログイン URL

https://gpu-console.highreso.jp/auth/login

ログインするとコントロールパネルが表示されます。

3. インスタンスの作成

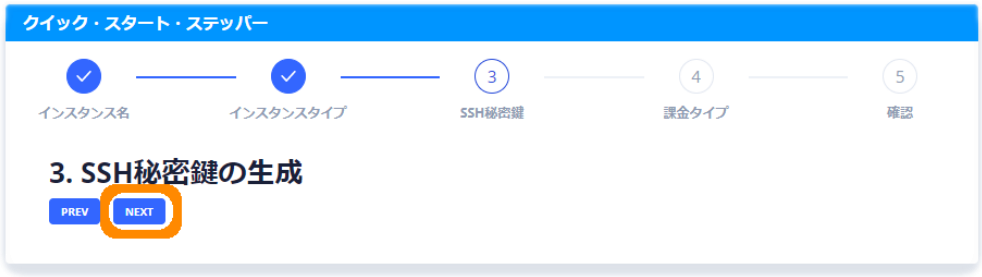
ここからはインスタンスを作成していきます。

今回は、ウィザード形式で進められる「クイック・スタート」で行います。

① 左のメニューバーより「インスタンス作成」、「クイック・スタート」の順にクリックし、作成するインスタンス名を入力して「NEXT」ボタンをクリックします。

② 後々機械学習をやってみたいので、インスタンスに nv4-1dl を選択します。

③ インスタンスにアクセスするための“SSH 秘密鍵の生成”を行います。
ここは「NEXT」をクリックして進めるだけで、OK です。



④ 今回は無償試用なので「従量」のまま「NEXT」をクリックします。

⑤ 確認画面が表示されますので、内容に問題のないことを確認して「CREATE」をクリックします。従量に金額が表示されていますが、試用中は課金されないのでご安心ください。

以上の設定でインスタンスを作成できました。

4. インスタンスの起動

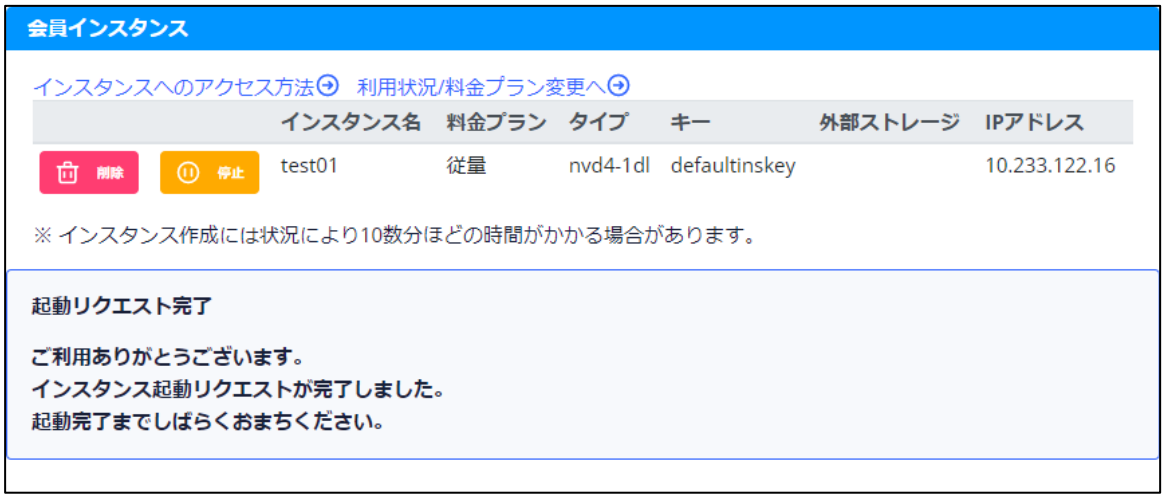
作成したインスタンスは「会員インスタンス」に表示されます。

① 「会員インスタンス」より「起動」をクリックしてインスタンスを起動します。

② 課金に関する確認のダイアログが表示されますのでお読みいただき、「内容を確認しました」をチェックして、「起動する」をクリックします。

③ 「IP アドレス」欄が「停止中」から「Pending」に変わりました。

インスタンスを作成後、ちょっと離席してから起動しました。その間にインスタンスの作成は終わっていたようで、15 秒程度で起動して「IP アドレス」欄に IP アドレスが表示されました。



5. インスタンスへの接続と環境の確認

「会員インスタンス」の下の方に OS 別の「接続方法」が記載されています。今回は Linux から接続してみます。

① 「アクセス鍵ダウンロード」、「インスタンス鍵ダウンロード」を順にクリックして、接続に使用する Linux PC に秘密鍵をダウンロードします。

② 今回は、ユーザのホームディレクトリ下の.ssh 内にダウンロードして、パーミッションを変更しました。

```
kirin@xeon-E3:~$ ls -la ~/.ssh/
合計 16
drwxrwxr-x 2 kirin kirin 4096  6 月 16 10:36 .
drwxr-xr-x 7 kirin kirin 4096  6 月 16 10:28 ..
-rwxr--r-- 1 kirin kirin 3243  6 月 16 10:36 ackey.txt
-rwxr--r-- 1 kirin kirin 3243  6 月 16 10:36 mykey.txt
kirin@xeon-E3:~$ chmod 600 ~/.ssh/ackey.txt ~/.ssh/mykey.txt
```

インスタンスへの接続経路は以下のようになっています。

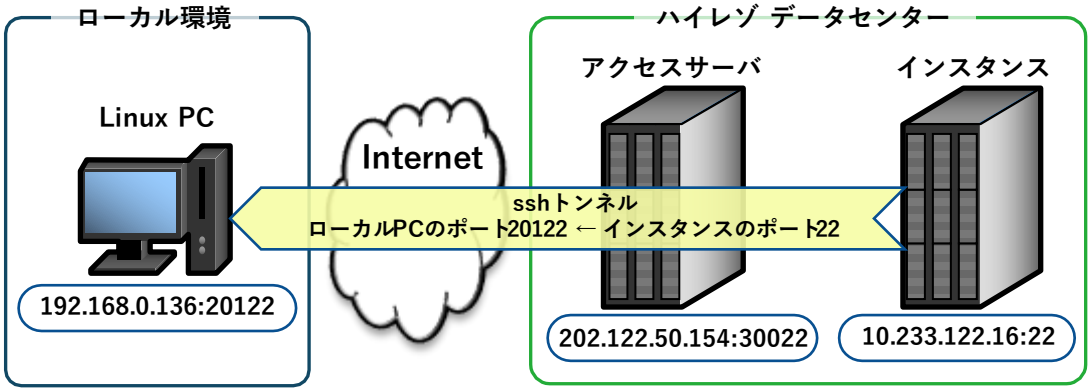


図 1 インスタンスへの接続経路

インスタンスに接続するにはアクセスサーバを経由する必要があり、インスタンスに直接 ssh でログインすることができません。

「インスタンスへのアクセス方法」を参考に ssh のポートフォワーディングを使ってアクセスサーバと ssh で接続し、インスタンスのポート 22(ssh 用ポート)をローカルの Linux PC のポート 20122 にフォワードします。

※上記リンク先は会員登録をしていないと開けません

③ インスタンスの IP アドレスは、インスタンスを再起動するたびに変化しますので、環境変数に割り当てました。では、ssh でトンネルを掘ります。ssh の鍵はアクセスサーバ用秘密鍵を使用します。

```
kirin@xeon-E3:~$ export VM_ADR=10.233.122.16  # インスタンスの IP アドレス
kirin@xeon-E3:~$ export AC_ADR=202.122.50.154 # アクセスサーバの IP アドレス
```

```
kirin@xeon-E3:~$ export CONNECT_PORT=20122     # インスタンスの ssh ポートを割り当てるローカル PC のポート
kirin@xeon-E3:~$ ssh -L $CONNECT_PORT:$VM_ADR:22 -l user $AC_ADR -p 30022 -i ~/.ssh/ackey.txt
Highreso GPU Advance.
This is access server.
```

これでローカルの Linux PC からインスタンスに ssh で接続できるようになりました。ローカル PC のポート 20122 にアクセスするとインスタンスのポート 22 にフォワードされます。このターミナルは、このままにしておきます。

④ インスタンスに ssh でログインします。ssh の鍵はインスタンス用秘密鍵を使用します。

インスタンスに作られているアカウントは「user」です。

```
kirin@xeon-E3:~$ ssh user@localhost -p 20122 -i ~/.ssh/mykey.txt
The authenticity of host '[localhost]:20122 ([127.0.0.1]:20122)' can't be established.
ECDSA key fingerprint is SHA256:FDO/Kab5WuIJF44iXFiVfdeaNz18XMOT2nH5m8FuA+k.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added '[localhost]:20122' (ECDSA) to the list of known hosts.
Welcome to Ubuntu 18.04.5 LTS (GNU/Linux 4.15.0-167-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage

The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.
```

ログインに成功しました。

※下記のエラーが発生した場合はインスタンスの IP アドレスの指定に間違いがないかご確認ください。

```
kirin@xeon-E3:~$ ssh user@localhost -p 20122 -i ~/.ssh/mykey.txt
ssh_exchange_identification: read: Connection reset by peer

# ssh トンネルを掘ったターミナルには下記のエラーが表示されている
# channel 3: open failed: connect failed: Host is unreachable
```

CUDA 関連の情報を見てみます。窓口から聞いた通り 11.1 ですね。

```
(base) user@test01:~$ nvidia-smi
Thu Jun 16 02:44:21 2022
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 455.23.05    Driver Version: 455.23.05    CUDA Version: 11.1     |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|                               |                      |               MIG M. |
|===============================+======================+======================|
|   0  A100-PCIE-40GB      Off  | 00000000:C1:00.0 Off |                    0 |
| N/A   44C    P0    38W / 250W |      0MiB / 40536MiB |      0%      Default |
|                               |                      |             Disabled |
+-------------------------------+----------------------+----------------------+

+-----------------------------------------------------------------------------+
| Processes:                                                                  |
|  GPU   GI   CI        PID   Type   Process name                  GPU Memory |
|        ID   ID                                                   Usage      |
|=============================================================================|
|  No running processes found                                                 |
+-----------------------------------------------------------------------------+
(base) user@test01:~$ cat /proc/driver/nvidia/version
NVRM version: NVIDIA UNIX x86_64 Kernel Module  455.23.05  Fri Sep 18 19:37:12 UTC 2020
GCC version:  gcc version 7.5.0 (Ubuntu 7.5.0-3ubuntu1~18.04)
(base) user@test01:~$ cat /usr/local/cuda/version.txt
CUDA Version 11.0.228
(base) user@test01:~$ nvcc -V
nvcc: NVIDIA (R) Cuda compiler driver
Copyright (c) 2005-2020 NVIDIA Corporation
```

```
Built on Wed_Jul_22_19:09:09_PDT_2020
Cuda compilation tools, release 11.0, V11.0.221
Build cuda_11.0_bu.TC445_37.28845127_0
```

他にも気になる gcc、Python、Anaconda、Docker の状態を見てみます。

```
(base) user@test01:~$ gcc --version
gcc (Ubuntu 7.5.0-3ubuntu1~18.04) 7.5.0
Copyright (C) 2017 Free Software Foundation, Inc.
This is free software; see the source for copying conditions.  There is NO
warranty; not even for MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.
(base) user@test01:~$ python -V
Python 3.7.6
(base) user@test01:~$ python3 -V
Python 3.7.6
(base) user@test01:~$ pip list | egrep "(tensor|cv)"
(base) user@test01:~$
(base) user@test01:~$ conda -V
conda 4.9.2
(base) user@test01:~$ docker
-bash: docker: コマンドが見つかりません
```

事前に聞いていた通り Docker はないですね。Docker はインストールしても使えないそうです。

6. CUDA のサンプルでベンチマークを実行

CUDA のサンプルをビルドします。

サンプルのソースは/usr/local/cuda-11.0 にあるのでこちらを使用します。

```
(base) user@test01:~$ cd /usr/local/cuda-11.0/bin
(base) user@test01:/usr/local/cuda-11.0/bin$ ./cuda-install-samples-11.0.sh ~/cuda_sample
Copying samples to /home/user/cuda_sample/NVIDIA_CUDA-11.0_Samples now...
Finished copying samples.
(base) user@test01:/usr/local/cuda-11.0/bin$ cd ~/cuda_sample/NVIDIA_CUDA-11.0_Samples/
(base) user@test01:~/cuda_sample/NVIDIA_CUDA-11.0_Samples$ time make -j16
make[1]: ディレクトリ '/home/user/cuda_sample/NVIDIA_CUDA-11.0_Samples/0_Simple/cppOverload' に入ります
：省略
Finished building CUDA samples
```

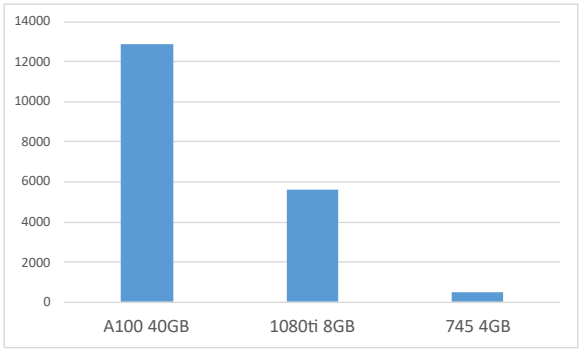
無事、ビルドに成功しました。

サンプルのベンチマークを実行して動作確認をします。今回は nbody を使用し、パラメータの numbodies は 4194304 にします。

```
(base) user@test01:~$ cd ~/cuda_sample/NVIDIA_CUDA-11.0_Samples
(base) user@test01:~/cuda_sample/NVIDIA_CUDA-11.0_Samples$ ./bin/x86_64/linux/release/nbody ¥
  -benchmark -numbodies=4194304
Run "nbody -benchmark [-numbodies=<numBodies>]" to measure performance.
        -fullscreen       (run n-body simulation in fullscreen mode)
        -fp64             (use double precision floating point values for simulation)
        -hostmem          (stores simulation data in host memory)
        -benchmark        (run benchmark to measure performance)
        -numbodies=<N>    (number of bodies (>= 1) to run in simulation)
        -device=<d>       (where d=0,1,2.... for the CUDA device to use)
        -numdevices=<i>   (where i=(number of CUDA devices > 0) to use for simulation)
        -compare          (compares simulation results running once on the default GPU and once on the
CPU)
        -cpu              (run n-body simulation on the CPU)
        -tipsy=<file.bin> (load a tipsy model file for simulation)

NOTE: The CUDA Samples are not meant for performance measurements. Results may vary when GPU Boost is
enabled.

> Windowed mode
> Simulation data stored in video memory
> Single precision floating point simulation
> 1 Devices used for simulation
GPU Device 0: "Ampere" with compute capability 8.0

> Compute 8.0 CUDA device: [A100-PCIE-40GB]
number of bodies = 4194304
4194304 bodies, total time for 10 iterations: 541977.688 ms
= 324.592 billion interactions per second
= 6491.849 single-precision GFLOP/s at 20 flops per interaction
```

10 分近くかかりました。

A100、GTX 1080Ti、GTX 745 で同じベンチマークを走らせてみました。A100 の得意なところを引き出せるベンチマークではないのかなという印象です。ベンチマークはよく考えなければいけませんね。



7. まとめ

会員登録の開始から 5 分位でインスタンスを起動できました。

起動したインスタンスへの接続には踏み台（アクセスサーバ）を経由する必要があり、ローカルの PC、踏み台、インスタンスそれぞれのポートを整理するために絵を描いたり、インスタンスの IP アドレスを打ち間違えてエラーが発生したりして 10 分位かかってしまいました。そのようなことをしなければインスタンスへの接続は数分で終了すると思います。

CUDA のドライバや toolkit はインストールされていたので、CUDA のサンプルはビルドするだけで動かすことができました。ベンチマークの実行中の待ちが 10 分位かかりましたが、それを除けば 5 分位で終了します。

登録開始からサンプルを動かすまで、特に急ぐことなく進めて 30 分あれば十分でした。Docker は不要であれば、今回作成したインスタンスを使って始められそうです。

では、次回お会いしましょう。

改版履歴

Version	日付	改版内容
1.0	2022/6/27	・新規作成