



Running NVIDIA Parabricks on Azure

Table of contents

What is NVIDIA Parabricks?

Starting a Compute Instance

Installing Parabricks

Testing Parabricks

Closing Remarks

List of Figures

Figure 0. Image Starting A Compute Instance

Figure 1. Image Create Pb Resource

Figure 2. Image Set Instance Details

Figure 3. Image Select Image

Figure 4. Image Choose Size

Figure 5. Image Admin Acct

Figure 6. Image Validation Passed

Figure 7. Image Go To Resource

Figure 8. Image Connect

Figure 9. Image Ngc

Figure 10. Image Docker Pull

Figure 11. Image Wget Sample Data

Figure 12. Image Untar

Figure 13. Image Output

This guide shows how to run Parabricks on a compute instance on Azure.

What is NVIDIA Parabricks?

Parabricks is an accelerated compute framework that supports applications across the genomics industry, primarily supporting analytical workflows for DNA, RNA, and somatic mutation detection applications. With industry leading compute times, Parabricks rapidly converts a FASTQ file to a VCF using multiple, industry validated variant callers and also includes the ability to QC and annotate those variants. As Parabricks is based upon publicly available tools, results are easy to verify and combine with other publicly available data sets.

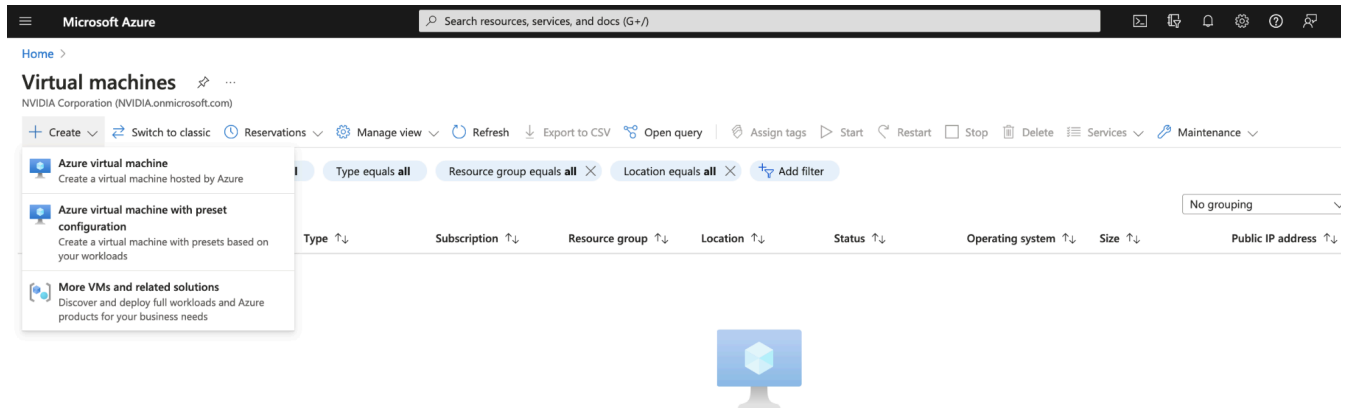
More information is available on the [Parabricks Product Page](#).

Detailed installation, usage, and tuning information is available in the [Parabricks user guide](#).

Starting a Compute Instance

In this section, we will show how to start a Compute Instance on Azure.

Begin by navigating to the [Azure Portal homepage](#) and from the menu on the top left, select Virtual machines. This will take us to the Virtual machines page.



At the top of the page, select Create, and Azure virtual machine. Here we can configure all the settings for our Virtual machine. You will need to choose your Subscription, then choose or create a new Resource group. For our example we will create a Resource group called "Parabricks".

Home > Virtual machines >

Create a virtual machine ...

Basics Disks Networking Management Monitoring Advanced Tags Review + create

Create a virtual machine that runs Linux or Windows. Select an image from Azure marketplace or use your own customized image. Complete the Basics tab then Review + create to provision a virtual machine with default parameters or review each tab for full customization. [Learn more](#)

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription * ⓘ

Resource group * ⓘ [Create new](#)

Instance details

Virtual machine name * ⓘ

Region * ⓘ

Availability options ⓘ

Security type ⓘ

A resource group is a container that holds related resources for an Azure solution.

Name *

We then add the instance details, we will name our VM "Parabricks", and select an appropriate region. We then choose the Security type as "Standard" as shown below.

Create a virtual machine ...

Instance details

Virtual machine name * ⓘ

Parabricks ✓

Region * ⓘ

(US) East US

Availability options ⓘ

No infrastructure redundancy required

Security type ⓘ

Standard

Image * ⓘ

Standard
The basic level of security for your virtual machines.

Trusted launch virtual machines
Protects against persistent and advanced attacks on Gen 2 virtual machines with configurable features like secure boot and virtual Trusted Platform Module (vTPM).

Confidential virtual machines
On top of Trusted launch, Confidential virtual machine offers higher confidentiality and integrity guaranteed with hardware-based trusted execution environment.

[See all sizes](#)

VM architecture ⓘ

Run with Azure Spot discount ⓘ

Size * ⓘ

We choose an image by clicking on See all images. You can search for nvidia, and choose "NVIDIA GPU-Optimized VMI with vGPU driver" that will take care of the GPU Driver, and docker installation.

Select an image ...

Other Items

- My Images
- Shared Images
- Community Images
- Direct Shared Images (PREVIEW)

Marketplace

All

- Recently created
- Private products

Categories

- Compute (22)
- Analytics (14)
- Developer Tools (14)
- AI + Machine Learning (8)
- Containers (3)
- Databases (1)
- Blockchain (0)
- DevOps (0)

Marketplace

Publisher name: All X Security
 Azure benefit eligible only ⓘ Pricing: All X

Showing 1 to 20 of 34 results for 'nvidia' with 1 selected filters. [Clear search + filters](#)

| | | |
|--|---|---|
| NVIDIA GPU-Optimized VMI NVIDIA Corporation Virtual Machine Maximize GPU performance for your Deep Learning, Machine Learning & HPC applications with this image Select v | NVIDIA GPU-Optimized VMI with vGPU driver NVIDIA Corporation Virtual Machine Maximize GPU performance for your Deep Learning, Machine Learning & HPC applications with this image Select v | NVIDIA AI Enterprise NVIDIA Corporation Virtual Machine The end to end platform for building accelerated production AI. Starts at \$2.00/user/3 years Select v |
| NVIDIA GPU-Optimized VMI - v23.03.0 - x64 Gen 1 NVIDIA GPU-Optimized VMI - v23.03.0 - x64 Gen 2 NVIDIA GPU-Optimized VMI - v22.06.0 - x64 Gen 1 NVIDIA GPU-Optimized VMI - v22.06.0 - x64 Gen 2 NVIDIA GPU-Optimized VMI - v22.03.0 - x64 Gen 2 NVIDIA GPU-Optimized VMI - v22.03.0 - x64 Gen 1 Virtual Ubuntu 18.04 Cloud Computing | | NVIDIA CloudXR Introductory Offer - NVIDIA Corporation Virtual Machine CloudXR is NVIDIA's streaming SDK |

We then choose the Size for the GPUs that we want to be using.

Instance details

| | |
|--------------------------------|---|
| Virtual machine name * ⓘ | <input type="text" value="Parabricks"/> |
| Region * ⓘ | <input type="text" value="(US) East US"/> |
| Availability options ⓘ | <input type="text" value="No infrastructure redundancy required"/> |
| Security type ⓘ | <input type="text" value="Standard"/> |
| Image * ⓘ | <input type="text" value="NVIDIA GPU-Optimized VMI - v22.03.0 - x64 Gen2"/> See all images Configure VM generation |
| VM architecture ⓘ | <input type="radio"/> Arm64 <input checked="" type="radio"/> x64 i Arm64 is not supported with the selected image. |
| Run with Azure Spot discount ⓘ | <input type="checkbox"/> |
| Size * ⓘ | <input type="text" value="Standard_NV72ads_A10_v5 - 72 vcpus, 880 GiB memory (Loading price...)"/> See all sizes |

For the Administrator account, You can either use an SSH public key, or as shown below, use a Username and Password.

Administrator account

| | |
|-----------------------|---|
| Authentication type ⓘ | <input type="radio"/> SSH public key <input checked="" type="radio"/> Password |
| Username * ⓘ | <input type="text" value="pbuser"/> |
| Password * ⓘ | <input type="password" value="....."/> |
| Confirm password * ⓘ | <input type="password" value="....."/> |

[Review + create](#)

[< Previous](#)

[Next : Disks >](#)

If you need Disks, Networking or other specifications related to your work, click on “Next:Disks >”. However, if you don’t need any of these click on “Review + create”. If everything is configured well you will have a page to review, and the message “Validation passed” at the top.

Home >

Create a virtual machine ...

Validation passed

Basics Disks Networking Management Monitoring Advanced Tags Review + create

Cost given below is an estimate and not the final price. Please use [Pricing calculator](#) for all your pricing needs.

Price

NVIDIA GPU-Optimized VMI
by NVIDIA Corporation
[Terms of use](#) | [Privacy policy](#)

Not covered by credits ⓘ

0.0000 USD/hr

1 X Standard NV72ads A10 v5
by Microsoft
[Terms of use](#) | [Privacy policy](#)

Subscription credits apply ⓘ

[Pricing for other VM sizes](#)

TERMS

By clicking "Create", I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed above; (b) authorize Microsoft to bill my current payment method for the fees associated with the offering(s), with the same billing frequency as my Azure subscription; and (c) agree that Microsoft may share my contact, usage and transactional information with the provider(s) of the offering(s) for support, billing and other transactional activities. Microsoft does not provide rights for third-party offerings. See the [Azure Marketplace Terms](#) for additional details.

Name

Preferred e-mail address

Preferred phone number

Basics

Subscription

Resource group (new) Parabricks

Virtual machine name Parabricks

Region East US

Availability options No infrastructure redundancy required

Security type Standard

Image NVIDIA GPU-Optimized VMI - v22.03.0 - Gen2

[Create](#) [< Previous](#) [Next >](#) [Download a template for automation](#)

Review the details and click Create at the bottom of the page. You will land on a page saying "... Deployment in progress". When it is complete, click the "Go to resource" button.

∨ **Deployment details**

∧ **Next steps**

[Setup auto-shutdown](#) Recommended

[Monitor VM health, performance and network dependencies](#) Recommended

[Run a script inside the virtual machine](#) Recommended

[Go to resource](#)

[Create another VM](#)

You should see a page like below. Click on Connect.

The screenshot shows the Microsoft Azure portal interface. At the top, there is a search bar and the text "Microsoft Azure". Below this, the breadcrumb navigation shows "Home > CreateVm-nvidia.nvidia-gpu-optimized-vmi-a10-nvid-20230829154145 | Overview >". The main heading is "Parabricks" with a sub-heading "Virtual machine". Below the heading, there is a search bar and a row of action buttons: Connect, Start, Restart, Stop, Capture, Delete, Refresh, Open in mobile, and Feedback. On the left side, there is a navigation menu with "Overview" selected, followed by "Activity log", "Access control (IAM)", "Tags", "Diagnose and solve problems", and "Settings". Under "Settings", there are "Networking" and "Connect" options. On the right side, under the "Essentials" section, the following details are listed: Resource group (move) : Parabricks, Status : Running, Location : East US, Subscription (move) : , Subscription ID : , Health state : -, and Tags (edit) : Add tags.

Once the instance is running click on Connect and use the IP information to ssh into your instance.

Verify that the driver is installed using *nvidia-smi*.

Installing Parabricks

We will install Parabricks into our instance that we just created. To do this, we will use the NVIDIA GPU Cloud (NGC) to download the Parabricks Docker image.

Visit the [Parabricks page on NGC](#) to get the Docker pull command for the latest version of Parabricks.

Catalog > Containers > Nvidia Clara Parabricks

Nvidia Clara Parabricks

Copy Image Path

Overview Tags Layers Security Scanning Related Collections

These instructions and commands are valid for Clara Parabricks v4.0.0-1 only. For earlier versions, please visit [Parabricks user guides](#) for each specific older version.

Note, you will need an installer for versions prior to v4.0.0-1. Instructions for this are also in the [Parabricks user guides](#).

What is Nvidia Clara Parabricks?

Nvidia Clara Parabricks is an accelerated compute framework that supports applications across the genomics industry, primarily supporting analytical workflows for DNA, RNA, and somatic mutation detection applications. With industry leading compute times, Parabricks rapidly converts a FASTQ file to a VCF using multiple, industry validated variant callers and also includes the ability to QC and annotate those variants. As Parabricks is based upon publicly available tools, results are easy to verify and combine with other publicly available datasets.

More information is available on the [Clara Parabricks Product Page](#).

Detailed installation, usage, and tuning information is available in the [Parabricks user guide](#).

The Clara Parabricks docker image can be obtained by running the following command:

```
$ docker pull nvcr.io/nvidia/clara/clara-parabricks:<TAG>
```

An example run of the fq2bam tool using the container will be as follow:

```
#This command assumes all the inputs are in <INPUT_DIR> and all the outputs go to <OUTPUT_DIR>.  
$ docker run --rm --gpus all -v <INPUT_DIR>:/workdir \  
-v <OUTPUT_DIR>:/outputdir \  
-v <TMP_DIR>:/raid/myrun -w /workdir \  
nvcr.io/nvidia/clara/clara-parabricks:<TAG> \  
pbrun fq2bam \  
--ref /workdir/$(REFERENCE_FILE) \  
--in-fq /workdir/$(INPUT_FASTQ_1) /workdir/$(INPUT_FASTQ_2) \  
--knownSites /workdir/$(KNOWN_SITES_FILE) \  
--out-bam /outputdir/$(OUTPUT_BAM) \  
--out-recal-file /outputdir/$(OUTPUT_RECAL_FILE)
```

Description
Nvidia Clara Parabricks is an accelerated compute framework that supports applications across the genomics industry, primarily supporting analytical workflows for DNA, RNA, and somatic mutation detection applications

Publisher
Nvidia

Latest Tag
4.0.0-1

Modified
November 8, 2022

Compressed Size
2.02 GB

Multinode Support
No

Multi-Arch Support

Back in our EC2 instance, let's run the docker pull command:

```
$ docker pull nvcr.io/nvidia/clara/clara-parabricks:4.3.1-1
```

```
[pbuser@Parabricks:~]$ docker pull nvcr.io/nvidia/clara/clara-parabricks:4.1.1-1
4.1.1-1: Pulling from nvidia/clara/clara-parabricks
df6635ed1257: Pull complete
6ceabd2ff7b0: Pull complete
cd189d71cce3: Pull complete
b0b6463464d3: Pull complete
0dbb8e45df45: Pull complete
518385bcb747: Pull complete
4cb5dcef316f: Pull complete
Digest: sha256:7a44179fcaeb6422c52f9e1f8b966110ab3fd50a57f06793c74659ad38eedb66
Status: Downloaded newer image for nvcr.io/nvidia/clara/clara-parabricks:4.1.1-1
nvcr.io/nvidia/clara/clara-parabricks:4.1.1-1
```

Parabricks is now installed. Let's run some sample data to test it.

Testing Parabricks

Parabricks provides a small sample dataset as a test for the installation and hardware which can be downloaded using:

```
$ wget -O parabricks_sample.tar.gz \
  "https://s3.amazonaws.com/parabricks.sample/parabricks_sample.tar.gz"
```

```
[pbuser@Parabricks:~]$ wget -O parabricks_sample.tar.gz \
> "https://s3.amazonaws.com/parabricks.sample/parabricks_sample.tar.gz"
--2023-08-01 12:57:31-- https://s3.amazonaws.com/parabricks.sample/parabricks_sample.tar.gz
Resolving s3.amazonaws.com (s3.amazonaws.com)... 52.217.232.16, 52.216.54.128, 52.217.227.184, ...
Connecting to s3.amazonaws.com (s3.amazonaws.com)|52.217.232.16|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 9924454379 (9.2G) [application/x-tar]
Saving to: 'parabricks_sample.tar.gz'
```

When the download completes we can untar the data using:

```
$ tar xvf parabricks_sample.tar.gz
```

And you should see something like below:

```
[pbuser@Parabricks:~]$ tar xvf parabricks_sample.tar.gz
parabricks_sample/
parabricks_sample/Data/
parabricks_sample/Data/sample_2.fq.gz
parabricks_sample/Data/sample_1.fq.gz
parabricks_sample/Ref/
parabricks_sample/Ref/Homo_sapiens_assembly38.fasta
parabricks_sample/Ref/Homo_sapiens_assembly38.fasta.pac
parabricks_sample/Ref/Homo_sapiens_assembly38.fasta.ann
parabricks_sample/Ref/Homo_sapiens_assembly38.known_indels.vcf.gz.tbi
parabricks_sample/Ref/Homo_sapiens_assembly38.fasta.amb
parabricks_sample/Ref/Homo_sapiens_assembly38.dict
parabricks_sample/Ref/Homo_sapiens_assembly38.fasta.fai
parabricks_sample/Ref/Homo_sapiens_assembly38.known_indels.vcf.gz
parabricks_sample/Ref/Homo_sapiens_assembly38.fasta.bwt
parabricks_sample/Ref/Homo_sapiens_assembly38.fasta.sa
```

Finally, we can run any of the Parabricks pipelines on it. Let's run the [germline pipeline](#) using the following command:

```
$ docker run \ --rm \ --gpus all \ --volume `pwd`:`pwd` \ --workdir
`pwd`/parabricks_sample \ nvc.io/nvidia/clara/clara-parabricks:4.3.1-1 \ pbrun
germline \ --ref Ref/Homo_sapiens_assembly38.fasta \ --in-fq Data/sample_1.fq.gz
Data/sample_2.fq.gz \ --knownSites
Ref/Homo_sapiens_assembly38.known_indels.vcf.gz.tbi \ --out-bam output.bam \ --
out-variants germline.vcf \ --out-recal-file recal.txt
```

We can tell that Parabricks started correctly when we see the Parabricks banner and the ProgressMeter begins to populate with values:

```

pbuser@Parabricks:~$ docker run \
> --gpus all \
> --rm \
> --volume $(pwd):/workdir \
> --volume $(pwd):/outputdir \
> nvr.io/nvidia/clara/clara-parabricks:4.1.1-1 \
> pbrun fq2bam \
> --ref /workdir/parabricks_sample/Ref/Homo_sapiens_assembly38.fasta \
> --in-fq /workdir/parabricks_sample/Data/sample_1.fq.gz /workdir/parabricks_sample/Data/sample_2.fq.gz \
> --out-bam /outputdir/fq2bam_output.bam \
> --low-memory

[Parabricks Options Msg]: Checking argument compatibility
[Parabricks Options Msg]: Automatically generating ID prefix
[Parabricks Options Msg]: Read group created for /workdir/parabricks_sample/Data/sample_1.fq.gz and
/workdir/parabricks_sample/Data/sample_2.fq.gz
[Parabricks Options Msg]: @RG\tID:HK3TJBCX2.1\tLB:lib1\tPL:bar\tSM:sample\tPU:HK3TJBCX2.1
[PB Info 2023-Aug-01 13:08:01] -----
[PB Info 2023-Aug-01 13:08:01] ||                               Parabricks accelerated Genomics Pipeline                               ||
[PB Info 2023-Aug-01 13:08:01] ||                               Version 4.1.1-1                               ||
[PB Info 2023-Aug-01 13:08:01] ||                               GPU-BWA mem, Sorting Phase-I                               ||
[PB Info 2023-Aug-01 13:08:01] -----
[M::bwa_idx_load_from_disk] read 0 ALT contigs
[PB Info 2023-Aug-01 13:08:03] GPU-BWA mem
[PB Info 2023-Aug-01 13:08:03] ProgressMeter      Reads          Base Pairs Aligned
[PB Info 2023-Aug-01 13:08:14] 5043564          580000000
[PB Info 2023-Aug-01 13:08:19] 10087128 1170000000

```

This should take ~10 minutes to finish running. When it's done, we should see the output files in the sample data directory.

Closing Remarks

We encourage you to expand on the demo in this guide by using your own data, trying other pipelines, and generally exploring what Parabricks has to offer. Check out the [documentation](#) for more information about the different pipelines available. You can also find our online developer community on the [Parabricks forum](#), where you can ask questions and search through answers while you are learning how to use Parabricks.

© Copyright 2024, Nvidia.. PDF Generated on 06/05/2024