Prepare for Deployment

Every Bright Cluster can be extended into the AWS or Microsoft Azure public clouds. Bright provides graphical cloud extension wizards that ask questions and use your answers to deploy and configure the cloud extension. It is important to think through your cloud deployment ahead of time because that will help you provide accurate answers to the questions. That is why I highly recommend that you read through this Getting Started Guide and think about how you would answer the questions.

Prerequisites

This Getting Started Guide assumes that you have installed your head node, activated your product key, and updated the head node and default software image.

To extend your cluster to AWS you will also need to meet the following three prerequisites:

1. You have a valid AWS account

   This can be a root or IAM account. You should be able to log into the AWS console at https://signin.aws.amazon.com/. The Cloud Extension wizard will authenticate you to AWS using your AWS Access Key ID and Secret Key.

2. You have registered your product key on the Bright Customer Portal


   Register your product key on the Customer Portal

   Please obtain your product key from the “Welcome to Easy8” email you received after requesting your free Easy8 product key. To ensure accuracy, I recommend that you select and copy the product key so you can paste it in the next step.
Easy8 Product Key for Robert Stober at Bright Computing

Dear Robert Stober,

Here is your Bright Cluster Manager Easy8 product key:
097293-210325-920411-628106-152316

This key is valid for 1 year from the day of activation. Please use the download link below to download a Bright Cluster Manager ISO.


TIP: You can also get your product key by executing the ‘cm-get-product-key’ command on the head node.

On the Customer Portal, select “Register Product Key”.

Paste your product key into the “Register Product Key” form, then press “Register”.

Welcome to the Customer Portal, Robert

Below are descriptions of the main Customer Portal sections.

1. Overview — The page you are currently viewing.
   - Account Overview — Displays all the basic information on your Customer Portal account and your login history.
   - Edit Account — Allows you to edit your account details.
   - Edit Billing — Allows you to edit your billing details.
   - Edit Password — You can use this page to change your password.

2. Product
   1. My Clusters — Go here if you want to list all your clusters.
   2. My Product Keys — Go here if you want to list all your product keys.

3. Unlock Product Key
   - Go here if you want to unlock your product key.

4. Register Product Key — Go here if you want to register product keys and discover all your clusters and product keys.

5. Cluster on Demand — This feature enables you to set up a free trial cluster on the website at Bright Computing. This page allows you to:
   - Start a cluster.
3. You have outgoing connectivity on port 1194 UDP

By default, an outbound UDP connection to port 1194 is used for the OpenVPN tunnel from the head node to the cloud director’s external IP. The OpenVPN tunnel securely connects your cluster to the AWS VPC that the cloud wizard will create for your cloud nodes.

**Deploy the Bright Cluster Extension to AWS**

Open your web browser and log into Bright View using the following URL:

https://<head node external IP>:8081/bright-view

Select Cloud -> AWS -> AWS Wizard
Check the Prerequisites

The wizard first shows the prerequisites for cluster extension. These are the same prerequisites we discussed in “Prepare for Installation” (above). Please press the “Next” button to continue if the prerequisites have been met.

Provide your AWS credentials

Enter your AWS access key ID and secret key, then press “Validate” to verify your credentials. Verification is optional, but strongly recommended.
Select Regions
Select the AWS regions you want to extend your cluster into. The wizard will create a VPC in each region you select.

Select Availability Zones
For each region you selected above, the wizard will create a public and private subnet. For best results, we recommend that you choose the same availability zone for both subnets.
Select software images

This cluster has two software images; the default-image and ai-image. If your cluster only has one software image, or if you want to use both software images on your cloud nodes, you should just accept the defaults by pressing “Next”.

If I were to accept these defaults both software images would be uploaded to the cloud director. Since I only intend to use the ai-image on my cloud compute nodes, I don’t need to upload the default-image to my cloud director and will remove it.

Note: You can use as many software images as you want on your cloud nodes. Each selected software image will automatically be uploaded to the cloud director and will consume disk space on the cloud director’s EBS volume. Therefore, the more software images you select the longer it will take to initially provision the cloud director, and the more AWS will charge you for EBS storage. While these costs are minimal compared to the instance costs, it makes sense to select only the software images you need. You can always add additional software images later.
Remove software images you do not need
To remove the default-image you need to select a different software image from the “Software image for cloud nodes” select list. I selected “ai-image”, and then I cleared the checkbox next to default-image in the left column. As a result, only the ai-image will be uploaded to the cloud director.

Choose cloud node instance types
This screen allows you to specify a permanent instance type for the cloud director and a default instance type for the cloud compute nodes. I have selected g4dn.xlarge, which provides a single NVIDIA T4 GPU.
TIP: You can change the default instance type for cloud nodes, the instance type for individual cloud nodes, or the instance type for groups of cloud nodes (node categories) after deployment. You just terminate the affected nodes, change the instance type, and power them back on. To change the cloud director’s instance type you would have to terminate it, change the instance type and power it back on. This will cause the cloud director to be rebuilt from scratch. You will need to rebuild the cloud director.

Review summary and deploy

Please review the summary. If you agree with it, please check the “Ready for deployment” checkbox, and then press “Deploy”. If you need to change anything you can use the “Back” button to navigate to the relevant step, or you can click on that step in the breadcrumbs at the top.
Cluster extension to AWS is now being deployed. In addition to the high-level installation log shown here, there’s a detailed log file in /var/log/cm-cluster-extension.log.

After a few minutes the deployment has completed successfully and the “Finish” button, which had been disabled during the deployment, is now enabled. Press the “Finish” button to close the wizard.

When you close the wizard, Bright View automatically shows the list of configured providers. Click on “Edit” to review the configuration that the wizard created.
Optional post-deployment configuration

There are two settings here that you might want to change: “Use paid marketplace AMIs”, and “Cloud job tagging”. By default, “Cloud job tagging” is set to “No”, and “Use paid marketplace AMIs” is set to “NEVER”.

Use paid marketplace AMIs

If you leave “Use paid marketplace AMIs” at the default setting (“NEVER”), Bright will never use the marketplace AMI, which means that you will only be able to start as many cloud nodes as you have available (unused) Bright licenses. For example, if you have 100 Bright subscription licenses, and you are using 90 of them on-prem, you will be able to start 10 cloud nodes (100 - 90). AWS will charge you for the use of those 10 cloud nodes, but neither AWS nor Bright will charge you for the Bright licenses since you have already purchased them.
If you set it to “ALWAYS”, you can start any number of cloud nodes (subject to AWS limits) without regard to the number of Bright licenses you own. AWS will charge you for the cloud nodes you use, and since you are using the marketplace AMI, their charge will include a small hourly cost for the Bright licenses. This is the best setting if all of your Bright licenses are in use, but if you have available Bright licenses the “AS_NEEDED” setting is better.

If you select “AS_NEEDED”, Bright will automatically use the regular (non-paid) AMI while you have available Bright licenses, and the marketplace AMI when you have no available Bright licenses. Your AWS charges will include a small hourly cost for the Bright licenses when the marketplace AMI is used. I recommend that you select this setting because it gives you the best of both worlds: you get to use your available Bright licenses if you have them, but you are not limited by licensing requirements when you have tight deadlines to meet and need to burst to the cloud for additional compute resources.

Note: Bright will automatically use the regular (non-paid) AMI while you have available Bright licenses, but once the cloud nodes have been started using the marketplace AMI they will continue to be used until they are powered, cycled, or rebooted.

Cloud job tagging
Turning on cloud job tagging allows Bright to automatically obtain the cost of running each job from AWS, and you will be able to see how much money each user spent in AWS during the time period in Bright Workload Accounting and Reporting. It is off by default, but I recommend that you turn it on.

Provision the Cloud Director
You are ready to power on the cloud director. The node installer automatically provisions the cloud director from the head node on first boot. During provisioning, the node installer installs the operating system, the /cm/shared directory, and the software images selected during deployment onto the cloud director’s EBS volume. For a default cluster, that comes to about 9GB, which, depending on your bandwidth, can take anywhere from 30 minutes to several hours. Most people just power on the cloud director and do something else while it provisions.

To power on the cloud director, select Devices -> Cloud Nodes, then press the cloud director’s “Use” button, and select Power -> On from the menu as shown in the following two images.
Confirm the power on operation.
Then just wait for the cloud director to be provisioned. Again, provisioning the cloud director is a one-time operation. After this, it will boot from its local disk hard drive.

If you want to see what’s going on behind the scenes while the cloud director is provisioning, you can have a look at the AWS console log from Bright View. From the cloud director’s “Use” button, select Misc -> Console Log.
You can also read the node-installer log file on the head node, which provides a detailed account of what the node installer is doing.

[root@headnode ~]# tail -f /var/log/node-installer

May 7 18:38:16 10.42.0.80 node-installer: Proc cmdline: /tmp/proc-cmdline
May 7 18:38:16 10.42.0.80 node-installer: Proc cmdline: BOOT_IMAGE=/boot/vmlinux-3.10.0-1062.12.1.el7.x86_64 ro vconsole.keymap=us crashkernel=auto vconsole.font=latarcyrheb-sun16 biosdevname=0 rootdelay=300 net.ifnames=0 earlyprintk=ttys0,115200 bcmcloud=aws bcmvirt=hvm console=ttys0 ip=10.42.0.80:10.42.0.1:10.42.0.1:255.255.255.0 BOOTIF=01-0a-4d-93-ee-10-f4
May 7 18:38:16 10.42.0.80 node-installer: Starting Node Installer (node-installer.143109_f559c80e5f) using instance ID: i-066d99efab483fb95 (running on Cloud Node)
May 7 18:38:16 10.42.0.80 node-installer: Node Installer was started in a non-interactive mode.
May 7 18:38:16 10.42.0.80 node-installer: Verifying the coherency of communication facilities.
May 7 18:38:16 10.42.0.80 node-installer: Calling verifyAPI on the CMDaemon at PXE master, https://172.30.255.254:8081
May 7 18:38:17 10.42.0.80 node-installer: Master cmdaemon url: https://172.30.255.254:8081
May 7 18:38:17 10.42.0.80 node-installer: No certificate and private key available.
May 7 18:38:18 10.42.0.80 node-installer: Requesting CMDaemon certificate directly from active head node.
May 7 18:38:19 10.42.0.80 node-installer: Certificate requested, request ID: 4, session ID: 42949672978
May 7 18:38:19 10.42.0.80 node-installer: Waiting for certificate to be issued.

When the cloud director has finished provisioning, a green UP arrow will be displayed next to it. Now you are ready to power on the cloud compute nodes. Note that the cloud compute nodes belong to the us-west-2-cloud-node category.

Provision the Cloud Compute Nodes

Once instantiated by AWS, the cloud nodes are provisioned from the cloud director. Although the first boot generally takes longer than subsequent boots (because the entire software image has to be provisioned onto each node) it still usually takes less than 5 minutes, but it could take longer depending on the availability of the selected instance type, and other factors.

To power on the cloud compute nodes, Select Grouping -> Categories, then Power -> On from the us-west-2-cloud-node categories “Use” button.
The cloud nodes are provisioning.

The cloud nodes are now UP and running.
That completes the deployment of the Bright Cluster extension to AWS. What you do next depends on how you intend to use the cloud. Here are some suggestions:

- Configure Slurm to use the cloud nodes
- Configure Slurm to use the GPUs on the cloud nodes
- Configure Bright Auto Scaler to dynamically burst into the cloud
- Deploy a Kubernetes cluster in the cloud
- Configure Bright Auto Scaler to scale a Kubernetes namespace

We will be adding more of these Getting Started Guides in the future, so please stay tuned. If you have any issues or questions, please refer to the Beacon User Community. If you would like to upgrade your Bright Cluster license to receive support or to extend your scope, please contact us.