Single Node Amazon Web Services Deployment

Creating a New Elastic Compute Cloud Instance for Partek Flow Software

Note: This guide assumes all items necessary for the Amazon elastic Comput Clout (EC2) instance does not exist, such as Amazon Virtual Private Cloud (VPC), subnets, and security groups, thus their creation is covered as well.

Log in to the Amazon Web Services (AWS) management console at https://console.aws.amazon.com

Click on **EC2**

Switch to the region intended to deploy Partek Flow software. This tutorial uses *US East (N. Virginia)* as an example.

On the left menu, click on **Instances**, then click the **Launch Instance** button. The **Choose an Amazon Machine Image (AMI)** page will appear.

Click the **Select** button next to **Ubuntu Server 16.04 LTS (HVM), SSD Volume Type - ami-f4cc1de2**. *NOTE:* Please use the latest Ubuntu AMI. It is likely that the AMI listed here will be out of date.

Choose an Instance Type, the selection depends on your budget and the size of the Partek Flow deployment. We recommend **m4.large** for testing or cluster front-end operation, **m4.xlarge** for standard deployments, and **m4.2xlarge** for alignment-heavy workloads with a large user-base. See the section AWS instance type resources and costs for assistance with choosing the right instance. In most cases, the instance type and associated resources can be changed after deployment, so one is not locked into the choices made for this step.

NOTE: New instance types will become available. Please use the latest mX instance type provided as it will likely perform better and be more cost effective than older instance types.

On the **Configure Instance Details** page, make the following selections:

Set the **number of instances** to *1*. An **autoscaling group** is not necessary for single-node deployments

Purchasing Option: Leave **Request Spot Instances** *unchecked*. This is relevant for cost-minimization of Partek Flow cluster deployments.

Network: If you do not have a virtual private cloud (VPC) already created for Partek Flow, click **Create New VPC**. This will open a new browser tab for VPC management.

Use the following settings for the VPC:

Name Tag: *Flow-VPC*

IPv4 CIDR block: *10.0.0.0/16*

Select **No IPv6 CIDR Block**

Tenancy: *Default*

Click **Yes, Create**. You may be asked to select a **DHCP Option set**. If so, then make sure the dynamic host configuration protocol (DHCP) option set has the following properties:

Options: *domain-name = ec2.internal;domain-name-servers = AmazonProvidedDNS;*

DNS Resolution: leave the defaults set to *yes*

DNS Hostname: change this to *yes* as internal DNS resolution may be necessary depending on the Partek Flow deployment

Once created, the new *Flow-VPC* will appear in the list of available VPCs. The VPC needs additional configuration for external access. To continue, right click on *Flow-VPC* and select **Edit DNS Resolution**, select *Yes*, and then Save. Next, right click the *Flow-VPC* and select **Edit DNS Hostnames**, select *Yes*, then Save.

Make sure the DHCP option set is set to the one created above. If it is not, right-click on the row containing *Flow-VPC* and select **Edit DHCP Option Sets.**

Close the **VPC Management** tab and go back to the **EC2 Management Console**.

Click the refresh arrow next to **Create New VPC** and select *Flow-VPC*.

Click **Create New Subnet** and a new browser tab will open with a list of existing subnets. Click **Create Subnet** and set the following options:

> **Name Tag**: *Flow-Subnet*

> **VPC**: *Flow-VPC*

> **VPC CIDRs**: This should be automatically populated with the information from Flow-VPC

> **Availability Zone**: It is OK to let Amazon choose for you if you do not have a preference

> **IPv4 CIDR block**: *10.0.1.0/24*

Stay on the **VPC Dashboard Tab** and on the left navigation menu, click **Internet Gateways**, then click **Create Internet Gateway** and use the following options:

> **Name Tag**: *Flow-IGW*

> Click **Yes, Create**

The new gateway will be displayed as **Detached**. Right click on the *Flow-IGW* gateway and select **Attach to VPC**, then select *Flow-VPC* and click **Yes, Attach**.

Click on **Route Tables** on the left navigation menu.

If it exists, select the route table already associated with *Flow-VPC*. If not, make a new route table and associate it with *Flow-VPC*. Click on the new route table, then click the **Routes** tab toward the bottom of the page. The route *Destination = 10.0.0.0/16 Target = local* should already be present. Click **Edit**, then **Click Add another route** and set the following parameters:

> **Destination**: *0.0.0.0/0*

> **Target** set to *Flow-IGW* (the internet gateway that was just created)

Click **Save**

Close the **VPC Dashboard** browser tab and go back to the **EC2 Management Console** tab. *Note that you should still be on **Step 3: Configure Instance Details***.

Click the refresh arrow next to **Create New Subnet** and select *Flow-Subnet*.

Auto-assign Public IP: *Use subnet setting (Disable)*

Placement Group: *No placement group*

IAM role: *None*.

> Note: For multi-node Partek Flow deployments or instances where you would like Partek to manage AWS resources on your behalf, please see *Partek AWS support* and set up an IAM role for your Partek Flow EC2 instance. In most cases a specialized IAM role is unnecessary and we only need instance ssh keys.

Shutdown Behaviour: *Stop*

Enable Termination Protection: select *Protect against accidental termination*

Monitoring: leave **Enable CloudWatch Detailed Monitoring** *disabled*

EBS-optimized Instance: Make sure **Launch as EBS-optimized Instance** is enabled. Given the recommended choice of an m4 instance type, EBS optimization should be enabled at no extra cost.

Tenancy: *Shared - Run a shared hardware instance*

Network Interfaces: leave as-is

Advanced Details: leave as-is

Click **Next: Add Storage**. You should be on **Step 4: Add Storage**

For the existing root volume, set the following options:

> **Size**: *8 GB*

> **Volume Type**: *Magnetic*

> Select **Delete on Termination**

> Note: All Partek Flow data is stored on a non-root EBS volume. Since only the OS is on the root volume and not frequently re-booted, a fast root volume is probably not necessary or worth the cost. For more information about EBS volumes and their performance, see the section EBS volumes.

Click **Add New Volume** and set the following options:

Volume Type: *EBS*

Device: */dev/sdb* (take the default)

Do not define a snapshot

Size (GiB): *500*

Note: This is the minimum for ST1 volumes, see: EBS volumes

Volume Type: *Throughput optimized HDD (ST1)*

Do not delete on terminate or encrypt

Click **Next: Add Tags**

You do not need to define any tags for this new EC2 instance, but you can if you would like.

Click **Next: Configure Security Group**

For **Assign a Security Group** select **Create a New Security Group**

Security Group Name: *Flow-SG*

Description: *Security group for Partek Flow server*

Add the following rules:

SSH set **Source** to *My IP* (or the address range of your company or institution)

Click **Add Rule:**

Set **Type** to *Custom TCP Rule*

Set **Port Range** to *8080*

Set **Source** to *anywhere (0.0.0.0/0, ::/0)*

*Note: It is recommended to restrict **Source** to just those that need access to Partek Flow.*

Click **Review and Launch**

The AWS console will suggest this server not be booted from a magnetic volume. Since there is not a lot of IO on the root partition and reboots are will be rare, choosing **Continue with Magnetic** will reduce costs. Choosing an SSD volume will not provide substantial benefit but it OK if one wishes to use an SSD volume. See the EBS Volumes section for more information.

Click **Launch**

Create a new keypair:

Name the keypair *Flow-Key*

Download this keypair, the run **chmod 600 Flow-Key.pem** (the downloaded key) so it can be used.

Backup this key as one may lose access to the Partek Flow instance without it.

The new instance will now boot. Use the left navigation bar and click on **Instances**. Click the pencil icon and assign the instance the name Partek *Flow Server*

Enabling External Access to the Partek Flow Elastic Compute Cloud Instance

The server should be assigned a fixed IP address. To do this, click on **Elastic IPs** on the left navigation menu from the **EC2 Management Console**.

Click **Allocate New Address**

Assign **Scope** to *VPC*

Click **Allocate**

On the table containing the newly allocated elastic IP, right click and select **Associate Address**

For **Instance**, select the instance name *Flow Test Server*

For **Private IP**, select the one private IP available for the Partek Flow EC2 instance, then click **Associate**

Note: For the remaining steps, we refer to the elastic ip as elastic.ip

SSH to the new Flow-Server instance:

$ chmod 600 Flow-Key.pem

$ ssh -i Flow-Testing.pem ubuntu@elastic.ip

Attaching the Amazon Elastic Block Store Volume for Partek Flow Data Storage

Attach, format, and move the *ubuntu* home directory onto the large ST1 elastic block store (EBS) volume. All Partek Flow data will live in this volume. Consult the AWS EC2 documentation for further information about attaching EBS volumes to your instance.

$ sudo su

$ mkfs -t ext4 /dev/xvdb

> *Note: Under **Volumes** in the EC2 management console, inspect **Attachment Information**. It will likely list the large ST1 EBS volume as attached to /dev/sdb. Replace "s" with "xv" to find the device name to use for this mkfs command.*

Make a note of the newly created UUID for this volume

Copy the *ubuntu* home directory onto the EBS volume using a temporary mount point:

$ mount -t ext4 /dev/xvdb /mnt/

$ rsync -avr /home/ /mnt/

$ umount /mnt/

Make the EBS volume mount at system boot:

Add the following to /etc/fstab: *UUID=the-UUID-from-the-mkfs-command-above /home   ext4   defaults,nofail   0   2*

$ mount -a

Disconnect the ssh session, then log in again to make sure all is well

Installing Partek Flow on a New Elastic Compute Cloud Instance

Note: For additional information about Partek Flow installations, see our generic Installation Guide

Before beginning, send the media access control (MAC) address of the EC2 instance to MAC address of the EC2 instance to licensing@partek.com. The output of *ifconfig* will suffice. Given this information, Partek employees will create a license for your AWS server. MAC addresses will remain the same after stopping and starting the Partek Flow EC2 instance. If the MAC address does change, let our licensing department know and we can add your license to our floating license server or suggest other workarounds.

Install required packages for Partek Flow:

$ sudo apt-get update

$ sudo apt-get install software-properties-common

$ sudo add-apt-repository -y ppa:openjdk-r/ppa

$ sudo apt-get install openjdk-8-jdk python python-pip python-dev zlib1g-dev python-matplotlib r-base python-htseq libxml2-dev perl make gcc g++ zlib1g libbz2-1.0 libstdc++6 libgcc1 libncurses5 libsqlite3-0 libfreetype6 libpng12-0 zip unzip libgomp1 libxrender1 libxtst6 libxi6 debconf

$ sudo pip install --upgrade pip && pip install --upgrade --upgrade-strategy eager --force-reinstall virtualenv numpy pysam cnvkit

Install Partek Flow:

Note: Make sure you are running as the ubuntu user.

$ cd (we will install Partek Flow to ubuntu's home directory)

$ wget --content-disposition packages.partek.com/linux/flow-release

$ unzip PartekFlow*.zip

$ ./partek_flow/start_flow.sh

Partek Flow has finished loading when you see *INFO: Server startup in xxxxxxx ms* in the *partek_flow/logs/catalina.out* log file. This takes ~30 seconds.

Alternative: Install Flow with Docker. Our base packages are located here: **https://hub.docker.com/r/partekinc/flow/tags**

Open Partek Flow with a web browser: http://elastic.ip:8080/

Enter license key

Set up the Partek Flow admin account

Leave the library file directory at its default location and check that the free space listed for this directory is consistent with what was allocated for the ST1 EBS volume.

Done! Partek Flow is ready to use.

Partek Amazon Web Services Support

After the EC2 instance is provisioned, we are happy to assist with setting up Partek Flow or address other issues you encounter with the usage of Partek Flow. The quickest way to receive help is to allow us remote access to your server by sending us *Flow-Key.pem* and amending the SSH rule for *Flow-SG* to include access from IP 97.84.41.194 (Partek HQ). We recommend sending us the *Flow-Key.pem* via secure means. The easiest way to do this is with the following command:

$ curl -F "file=@FlowKey.pem" https://installfeedback.partek.com/fupload

We also provide live assistance via GoTo meeting or TeamViewer if you are uncomfortable with us accessing your EC2 instance directly. Before contacting us, please run **$ ./partek_flow/flowstatus.sh** to send us logs and other information that will assist with your support request.

General Recommendations

With newer EC2 instance types, it is possible to change the instance type of an already deployed Partek Flow EC2 server. We recommend doing several rounds of benchmarks with production-sized workloads and evaluate if the resources allocated to your Partek Flow server are sufficient. You may find that reducing resources allocated to the Partek Flow server may come with significant cost savings, but can cause UI responsiveness and job run-times to reach unacceptable levels. Once you have found an instance type that works, you may wish to use *reserved instance* pricing which is significantly cheaper than *on-demand instance* pricing. Reserved instances come with 1 or 3-year usage terms. Please see the EC2 Reserved Instance Marketplace to sell or purchase existing reserved instances at reduced rates.

The network performance of the EC2 instance type becomes an important factor if the primary usage of Partek Flow is for alignment. For this use case, one will have to move copious amounts of data back (input fastq files) and forth (output bam files) between the Partek Flow server and the end users, thus it is important to have as what AWS refers to as *high network performance* which for most cases is around 1 Gb/s. If the focus is primarily on downstream analysis and visualization (e.g. the primary input files are *ADAT*) then network performance is less of a concern.

We recommend HVM virtualization as we have not seen any performance impact from using them and non-HVM instance types can come with significant deployment barriers.

Make sure your instance is *EBS optimized* by default and you are not charged a surcharge for EBS optimization.

T-class servers, although cheap, may slow responsiveness for the Partek Flow server and generally do not provide sufficient resources.

We do not recommend placing any data on *instance store* volumes since all data is lost on those volumes after an instance stops. This is too risky as there are cases where user tasks can take up unexpected amounts of memory forcing a server stop/reboot.

Amazon Web Services Instance Type Resources and Costs

The values below were updated April 2017. The latest pricing and EC2 resource offerings can be found at http://www.ec2instances.info

Instance Type	Memory	Cores	EBS throughput	Network Performance	Monthly cost
m4.large	8.0 GB	2 vCPUs	56.25 MB/s M	Medium	$78.840
r4.large	15.25 GB	2 vCPUs	50 MB/s H(10G int)	High (+10G interface)	$97.09
m4.xlarge	**16.0 GB**	**4 vCPUs**	**93.75 MB/s H**	High	**$156.950**
r4.xlarge	30.5 GB	4 vCPUs	100 MB/s H	High	$194.180
m4.2xlarge	**32.0 GB**	**8 vCPUs**	**125 MB/s H**	High	**$314.630**
r4.2xlarge	61.0 GB	8 vCPUs	200 MB/s H(10G int)	High (+10G interface)	$388.360

Single server recommendation: **m4.xlarge** or **m4.2xlarge**

Network performance values for US-EAST-1 correspond to: Low ~ 50Mb/s, Medium ~ 300Mb/s, High ~ 1Gb/s.

Elastic Block Store Volumes

Choice of a volume type and size:

This is dependent on the type of workload. For must users, the Partek Flow server tasks are alignment-heavy so we recommend a throughput optimized HDD (ST1) EBS volume since most aligner operations are sequential in nature. For workloads that focus primarily on downstream analysis, a general purpose SSD volume will suffice but the costs are greater. For those who focus on alignment or host several users, the storage requirements can be high. ST1 EBS volumes have the following characteristics:

Max throughput 500 MiB/s

$0.045 per GB-month of provisioned storage ($22.5 per month for a 500 GB of storage).

Note that EBS volumes can be grown or performance characteristics changed. To minimize costs, start with a smaller EBS volume allocation of **0.5 - 2 TB** as most mature Partek Flow installations generate roughly this amount of data. When necessary, the EBS volume and the underlying file system can be grown on-line (making ext4 a good choice). Shrinking is also possible but may require the Partek Flow server to be offline.

Additional Assistance

If you need additional assistance, please visit our support page to submit a help ticket or find phone numbers for regional support.

Your Rating: ☆☆☆☆☆   Results: ★★★★★ 43 rates