



Deployment of the Sentieon software to AWS

Release 202308.02

Sentieon, Inc

Feb 01, 2024

Contents

1	Introduction	1
1.1	Overview of the software	1
1.2	Platform requirements	2
2	Multi-Availability Zone Deployment on AWS	2
2.1	Prerequisites	3
2.2	Deployment of the Sentieon® license server	3
2.3	Deployment of one or more instances for genomic data processing	5
2.4	Upgrading the deployment and restarting the license server	5
3	Troubleshooting	6

1 Introduction

1.1 Overview of the software

The Sentieon® software provides fast and efficient data processing for genomic data. Supported applications include secondary analysis of sequence data from whole-genome, whole-exome, targeted, and RNA sequence data for alignment, preprocessing, germline and somatic variant calling.

The Sentieon® Genomics software enables rapid and accurate analysis of next-generation sequence data. The Sentieon® DNaseq pipelines enable germline variant calling on hundreds of thousands of samples simultaneously. The Sentieon® TNseq pipelines enable accurate calling of somatic variants in paired tumor-normal samples or in an unpaired tumor samples. The Sentieon® Genomics software produces more accurate results than other tools on third-party benchmarks with results obtained in one tenth the time of comparable pipelines.

1.2 Platform requirements

Sentieon® Genomics software is designed to run on Linux and other POSIX-compatible platforms.

For Linux systems, we recommended using the following Linux distributions or higher: RedHat/CentOS 6.5, Debian 7.7, OpenSUSE-13.2, or Ubuntu-14.04.

Sentieon® Genomics software requires at least 16GB of memory, although we recommend using 64 GB of memory. The alignment step is typically the most memory consuming stage in most pipelines.

Sentieon® Genomics software processes file data at speeds around 20-30MB/s per core using a state of the art server. In order to take advantage of the maximum speed that the software can provide, we recommend that you use high-speed SSD hard drives in your system, preferably two identical drives in a high speed RAID 0 striped volume configuration. When using a RAID 0 configuration, it is advised to use the drives to only store intermediate files, and move out any final results or important information out of those hard drives; this way, your server should have additional storage space to store the results.

The required hard disk space is pipeline and data specific and can vary significantly between use-cases. A general rule of thumb is that the software can use approximately 3x the input file size for intermediate file storage during data processing.

2 Multi-Availability Zone Deployment on AWS

The Sentieon® software is controlled by a license and requires an active Sentieon® license server to run. Deployment on AWS requires setup and configuration of the Sentieon® license server along with setup and configuration of one or more instances for genomic data processing. A minimum deployment will provision the following resources inside your default VPC and should take approximately ten minutes:

- Security groups that can be used to run the Sentieon® license server and the compute nodes.
- A CloudWatch log group for the license server logs.
- An IAM role and instance profile for the license server.
- A (t3.nano) EC2 instance running the Sentieon® license server.
- A private hosted zone in AWS Route53.

The Sentieon® software can be deployed onto additional instances for additional data processing capacity.

The provisioned IAM role will grant the license server instance read access to the Sentieon software package and to your Sentieon license file in AWS s3. It will also grant write access to the CloudWatch log group to record the license server logs in CloudWatch.

The root EBS disk for the license server instance will be encrypted. By default, the encryption will use an AWS-managed encryption key. Optionally, a KMS key can be specified to use a customer-managed encryption key.

With a minimum deployment, customers will be charged by AWS EC2 for the instance running the license server and for outbound communication from the Sentieon® license server to the Sentieon® master license server. Customers will also be charged for the Route53 hosted zone and for DNS queries resolved by Route53. Estimated AWS costs for the minimal license server deployment with a c6a.2xlarge instance for compute at 60% utilization is **\$134.03 per month in On-Demand costs**¹.

Adding additional compute instances to the deployment will increase the AWS EC2. Data processing may incur additional costs for data transfer, etc. The Sentieon® software is a proprietary software and this deployment requires a license subscription for the Sentieon® software.

This deployment is supported in the following AWS Regions:

- US East (Ohio)
- US East (N. Virginia)

¹<https://calculator.aws/#/estimate?id=099986a1d2f73bf694c8f74c08dc022299844ebd>

-
- US West (N. California)
 - US West (Oregon)
 - Africa (Cape Town)
 - Asia Pacific (Hong Kong)
 - Asia Pacific (Hyderabad)
 - Asia Pacific (Jakarta)
 - Asia Pacific (Melbourne)
 - Asia Pacific (Mumbai)
 - Asia Pacific (Osaka)
 - Asia Pacific (Seoul)
 - Asia Pacific (Singapore)
 - Asia Pacific (Sydney)
 - Asia Pacific (Tokyo)
 - Canada (Central)
 - Europe (Frankfurt)
 - Europe (Ireland)
 - Europe (London)
 - Europe (Milan)
 - Europe (Paris)
 - Europe (Spain)
 - Europe (Stockholm)
 - Europe (Zurich)
 - Middle East (Bahrain)
 - Middle East (UAE)
 - South America (São Paulo)

This deployment assumes basic familiarity with the VPC and EC2 services on AWS and familiarity with the Linux command-line interface.

2.1 Prerequisites

[Terraform](#)² is an open-source infrastructure as code (IaC) tool for the provisioning and management of cloud infrastructure. This guide uses Terraform to automate the provisioning of the required AWS infrastructure.

Terraform configuration files used in this deployment can be found on GitHub at, <https://github.com/Sentieon/terraform>.

The following are prerequisites for the deployment:

- The [Terraform CLI](#)³.
- The [AWS CLI](#)⁴.
- An AWS account and IAM credentials with permission to provision the required resources.

2.2 Deployment of the Sentieon® license server

In order to use the software on AWS, you will need to follow the following installation instructions to deploy a Sentieon® License server to your VPC.

1. Choose an AWS Region to work in. Choose a fully qualified domain name (FQDN) to associate with the license server. This might be something like `licsrvr.sentieon.example.com`.
2. Send your Sentieon® support representative the FQDN. Sentieon will send back a license file that can be used to start a Sentieon license server at the FQDN. Move the license file into an s3 bucket in the AWS region.

<https://www.terraform.io/>

<https://developer.hashicorp.com/terraform/downloads>

<https://docs.aws.amazon.com/cli/latest/userguide/getting-started-install.html>

Multi-Availability Zone Deployment on AWS

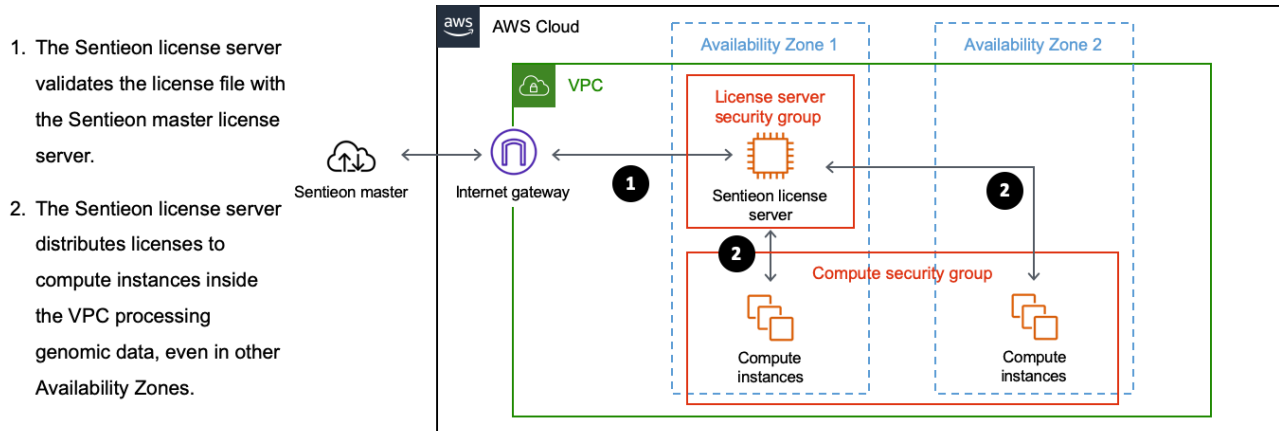


Fig. 2.1: Overview of the Sentieon deployment architecture on AWS

3. Use the following commands to download the Terraform configuration files to your local machine, configure your AWS credentials and initialize the directory with Terraform.

```
# Download the configuration files
git clone https://github.com/sentieon/terraform
cd terraform/aws_license-server

# Configure your AWS credentials
export AWS_ACCESS_KEY_ID=<AWS_ACCESS_KEY_ID>
export AWS_SECRET_ACCESS_KEY=<AWS_SECRET_ACCESS_KEY>
```

4. Use Terraform to provision the infrastructure and start the Sentieon license server. Use the `-var` arguments to pass the selected AWS region and FQDN along with the URI of your Sentieon® license file in s3 to Terraform. After the initial `terraform apply` command, Terraform will print the proposed infrastructure plan and ask to confirm the proposed changes. Replying yes will cause Terraform to provision the infrastructure.

```
# Provision the license server infrastructure
terraform apply \
  -var 'aws_region=<AWS_REGION>' \
  -var 'licsrvr_fqdn=<FQDN>' \
  -var 'license_s3_uri=s3://<S3_URI>'
```

The command should provision the infrastructure and start the Sentieon license server in under five minutes.

After the license server has been deployed, the license server can be monitored using the log files written to the CloudWatch log group, `/sentieon/licsrvr/LicsrvrLog`.

The infrastructure deployment can be destroyed by running `terraform apply` with the `-destroy` argument:

```
# Destroy the provisioned infrastructure
terraform apply \
  -destroy \
  -var 'aws_region=<AWS_REGION>' \
  -var 'licsrvr_fqdn=<FQDN>' \
  -var 'license_s3_uri=s3://<S3_URI>'
```

2.3 Deployment of one or more instances for genomic data processing

After the license server is deployed in your VPC, you can deploy the software to one or more additional instances for genomic data processing. These instances can be deployed to any Availability Zone in the VPC. The following steps provide guidance for deploying the Sentieon® software onto newly launched EC2 instances:

1. Start an EC2 instance that meets the platform requirements for the Sentieon® software inside your VPC. Assign the instance to the `sentieon_compute` security group created by Terraform.
2. Download the Sentieon® tools to the newly launched instance.
3. Set an environment variable in your system to indicate the location of the license server.

```
export SENTIEON_LICENSE=<FQDN>:8990
```

The new instance is now ready to process data and data processing does not require root privileges. Please see the Sentieon® software manual at, <https://support.sentieon.com/manual/> for more information on the functionality included in the Sentieon® software.

2.4 Upgrading the deployment and restarting the license server

Most new releases of the Sentieon® software will not affect the function of the license server. In the event of a software update affecting the license server, the following steps can be used to update the license server.

1. In the `main.tf` file, update the `sentieon_version` variable to the desired version.
2. Re-run the `terraform apply` command to terminate the running license server instance and start a new instance using the updated Sentieon software version.

```
# Update the license server
terraform apply \
  -var 'aws_region=<AWS_REGION>' \
  -var 'licsrvr_fqdn=<FQDN>' \
  -var 'license_s3_uri=s3://<S3_URI>'
```

In the event that the license server becomes unreachable is otherwise non-functional, you can recover the license server by destroying current instance and starting new instance. The instance can be destroyed with `terraform destroy`:

```
# Destroy the license server instance
terraform destroy \
  -target aws_instance.sentieon_licsrvr \
  -var 'aws_region=<AWS_REGION>' \
  -var 'licsrvr_fqdn=<FQDN>' \
  -var 'license_s3_uri=s3://<S3_URI>'
```

The license server can then be restarted with `terraform apply`:

```
# Restart the license server
terraform apply \
  -var 'aws_region=<AWS_REGION>' \
```

(continues on next page)

(continued from previous page)

```
-var 'licsrvr_fqdn=<FQDN>' \  
-var 'license_s3_uri=s3://<S3_URI>'
```

3 Troubleshooting

If you encounter issues with the software, please refer to the [Sentieon Software Manual](#)⁵ and [Quick Start Guide](#)⁶. You may also contact the Sentieon® support team at support@sentieon.com.

©Sentieon Inc.
160 E Tasman Dr STE 208, San Jose, CA 95134-1619
www.sentieon.com

<https://support.sentieon.com/manual>
https://support.sentieon.com/quick_start