

# DRAGEN on the AWS Cloud

## Quick Start Reference Deployment

*October 2018*

*Illumina*

*Aaron Friedman and Vinod Shukla, Amazon Web Services*

### Contents

Quick Links .....	2
Overview.....	3
DRAGEN on AWS .....	3
Costs and Licenses.....	3
Architecture.....	4
Prerequisites .....	5
Technical Requirements.....	5
Specialized Knowledge .....	6
Deployment Options .....	6
Deployment Steps .....	6
Step 1. Prepare Your AWS Account.....	6
Step 2. Subscribe to the DRAGEN AMI.....	7
Step 3. Launch the Quick Start .....	8
Step 4. Test the Deployment .....	14
Best Practices Using DRAGEN on AWS.....	18
Security.....	18
FAQ.....	18
GitHub Repository .....	21

Additional Resources .....	21
Document Revisions .....	22

This Quick Start was created by Illumina in collaboration with Amazon Web Services (AWS).

[Quick Starts](#) are automated reference deployments that use AWS CloudFormation templates to deploy key technologies on AWS, following AWS best practices.

## Quick Links

The links in this section are for your convenience. Before you launch the Quick Start, please review the architecture, security, and other considerations discussed in this guide.

- If you have an AWS account, and you're already familiar with AWS services and DRAGEN, you can launch the Quick Start to build the architecture shown in [Figure 1](#) in a new or existing virtual private cloud (VPC). The deployment takes approximately 15 minutes. If you're new to AWS or to DRAGEN, please review the implementation details and follow the [step-by-step instructions](#) provided later in this guide.

**Launch**  
(for new VPC)

**Launch**  
(for existing VPC)

- If you want to take a look under the covers, you can view the AWS CloudFormation templates that automate the deployment.

**View template**  
(for new VPC)

**View template**  
(for existing VPC)

## Overview

This Quick Start guide provides step-by-step instructions for deploying Dynamic Read Analysis for GENomics (DRAGEN) on the AWS Cloud.

The Quick Start is for users who would like to conduct bioinformatic analysis on next-generation sequencing data using DRAGEN on AWS.

## DRAGEN on AWS

This Quick Start was created to get you up and running with DRAGEN Complete Suite (CS) on AWS. DRAGEN CS offerings enable ultra-rapid analysis of next-generation sequencing (NGS) data, and significantly reduce the time required to analyze genomic data while also improving accuracy.

DRAGEN CS includes [many bioinformatics pipelines](#) that harness the power of the DRAGEN platform and provide highly optimized algorithms for mapping, aligning, sorting, duplicate marking, and haplotype variant calling. These pipelines include DRAGEN Germline V2, DRAGEN Somatic V2 (Tumor and Tumor/Normal), DRAGEN Virtual Long Read Detection (VLRD), DRAGEN RNA Gene Fusion, DRAGEN Joint Genotyping, and GATK Best Practices. The DRAGEN Germline V2 and DRAGEN Somatic V2 pipelines have improved accuracy in calling single nucleotide polymorphisms (SNPs) and small insertions and deletions (INDELs) compared with industry standards. For performance results, see the [PrecisionFDA Hidden Treasures – Warm Up](#) challenge and the [DRAGEN evaluation](#) on the *Inside DNAnexus* blog.

## Costs and Licenses

You are responsible for the cost of the AWS services used while running this Quick Start reference deployment. There is no additional cost for using the Quick Start.

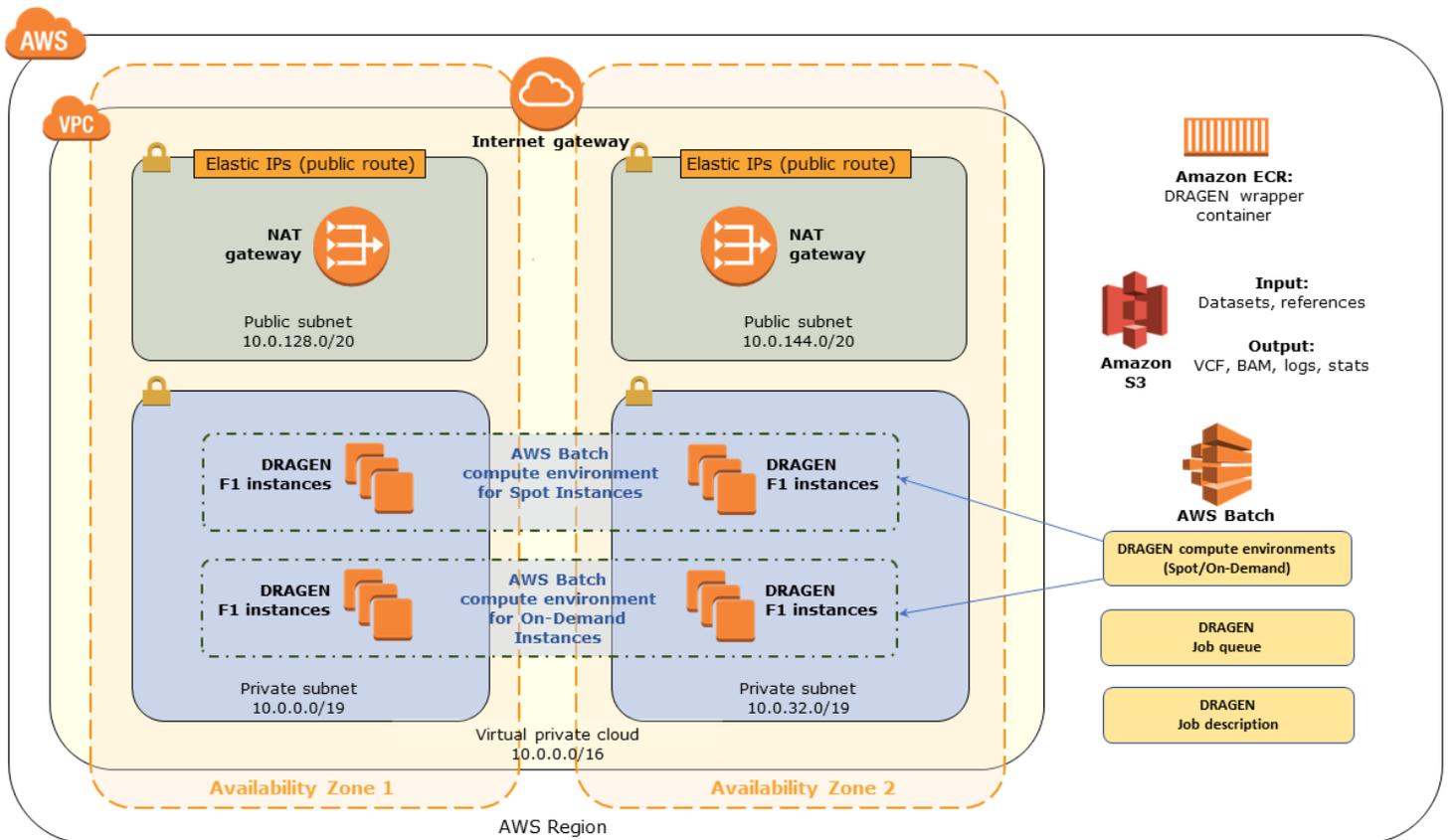
The AWS CloudFormation template for this Quick Start includes configuration parameters that you can customize. Some of these settings, such as instance type, will affect the cost of deployment. For cost estimates, see the pricing pages for each AWS service you will be using. Prices are subject to change.

**Tip** After you deploy the Quick Start, we recommend that you set up the [AWS Cost and Usage Report](#) to track costs associated with the Quick Start. This report delivers billing metrics to an S3 bucket in your account. It provides cost estimates based on usage throughout each month, and finalizes the data at the end of the month. For more information about the report, see the [AWS documentation](#).

This Quick Start requires a subscription to the Amazon Machine Image (AMI) for [DRAGEN Complete Suite](#), which is available with per-hour pricing from AWS Marketplace.

## Architecture

Deploying this Quick Start for a new virtual private cloud (VPC) with **default parameters** builds the following DRAGEN environment in the AWS Cloud.



**Figure 1: Quick Start architecture for DRAGEN on AWS**

The Quick Start sets up the following:

- A highly available architecture that spans two Availability Zones.\*
- A virtual private cloud (VPC) configured with public and private subnets according to AWS best practices. This provides the network architecture for DRAGEN deployment.\*
- An internet gateway to provide access to the internet.\*
- In the public subnets, managed NAT gateways to allow outbound internet access for resources in the private subnets.\*

- An AWS CodePipeline pipeline that builds a Docker image and uploads it into an Amazon Elastic Container Registry (Amazon ECR) repository that interfaces with AWS Batch and runs in the DRAGEN AMI.
- Two AWS Batch compute environments: one for Amazon Elastic Compute Cloud (Amazon EC2) Spot Instances and the other for On-Demand Instances.
- An AWS Batch job queue that prioritizes submission to the compute environment for Spot Instances to optimize for cost.
- An AWS Batch job definition to run DRAGEN.
- AWS Identity and Access Management (IAM) roles and policies for the AWS Batch jobs to run.

\* The template that deploys the Quick Start into an existing VPC skips the tasks marked by asterisks and prompts you for your existing VPC configuration.

## Prerequisites

### Technical Requirements

- **Limit increases for F1 instances.** DRAGEN runs on [Amazon EC2 F1 instances](#) because it requires a field-programmable gate array (FPGA). This Quick Start supports both f1.2xlarge and f1.16xlarge instance types. You should request limit increases for F1 instances, to support the maximum number of simultaneous DRAGEN jobs that you expect to run.
- **S3 bucket for genomic data.** You must have an Amazon Simple Storage Service (Amazon S3) bucket in the AWS Region where you plan to deploy the Quick Start. This S3 bucket should contain:
  - The genomic input datasets that you want to run in the Quick Start environment
  - The DRAGEN-specific reference hash table directories that are provided by Illumina or that you create from FASTA files
  - An output folder for DRAGEN job outputs such as the binary alignment map (BAM) and variant call format (VCF) files

You'll be prompted for the bucket name when you deploy the Quick Start. The appropriate roles and policies to read from, and write to, this S3 genomics bucket are created during the Quick Start deployment.

## Specialized Knowledge

Before you deploy this Quick Start, we recommend that you become familiar with the following AWS services. (If you are new to AWS, see [Getting Started with AWS](#).)

- [Amazon Elastic Block Store \(Amazon EBS\)](#)
- [Amazon EC2](#)
- [Amazon ECR](#)
- [IAM](#)
- [Amazon S3](#)
- [Amazon Virtual Private Cloud \(Amazon VPC\)](#)
- [AWS Batch](#)

## Deployment Options

This Quick Start provides two deployment options:

- **Deploy DRAGEN into a new VPC** (end-to-end deployment). This option builds a new AWS environment consisting of the VPC, subnets, NAT gateways, security groups, and other infrastructure components, and then deploys DRAGEN into this new VPC.
- **Deploy DRAGEN into an existing VPC**. This option provisions DRAGEN in your existing AWS infrastructure.

The Quick Start provides separate templates for these options. It also lets you configure CIDR blocks, instance types, and DRAGEN settings, as discussed later in this guide.

## Deployment Steps

### Step 1. Prepare Your AWS Account

1. If you don't already have an AWS account, create one at <https://aws.amazon.com> by following the on-screen instructions.
2. Use the region selector in the navigation bar to choose the AWS Region where you want to deploy DRAGEN on AWS.

**Important** This Quick Start includes F1 instances, which aren't available in all AWS Regions. For a list of supported regions, see the [Amazon EC2 Pricing](#) webpage.

3. Create a [key pair](#) in your preferred region.

4. If necessary, [request a service limit increase](#) for the Amazon EC2 F1 instance type. You might need to do this if you already have an existing deployment that uses this instance type, and you think you might exceed the [default limit](#) with this deployment.

## Step 2. Subscribe to the DRAGEN AMI

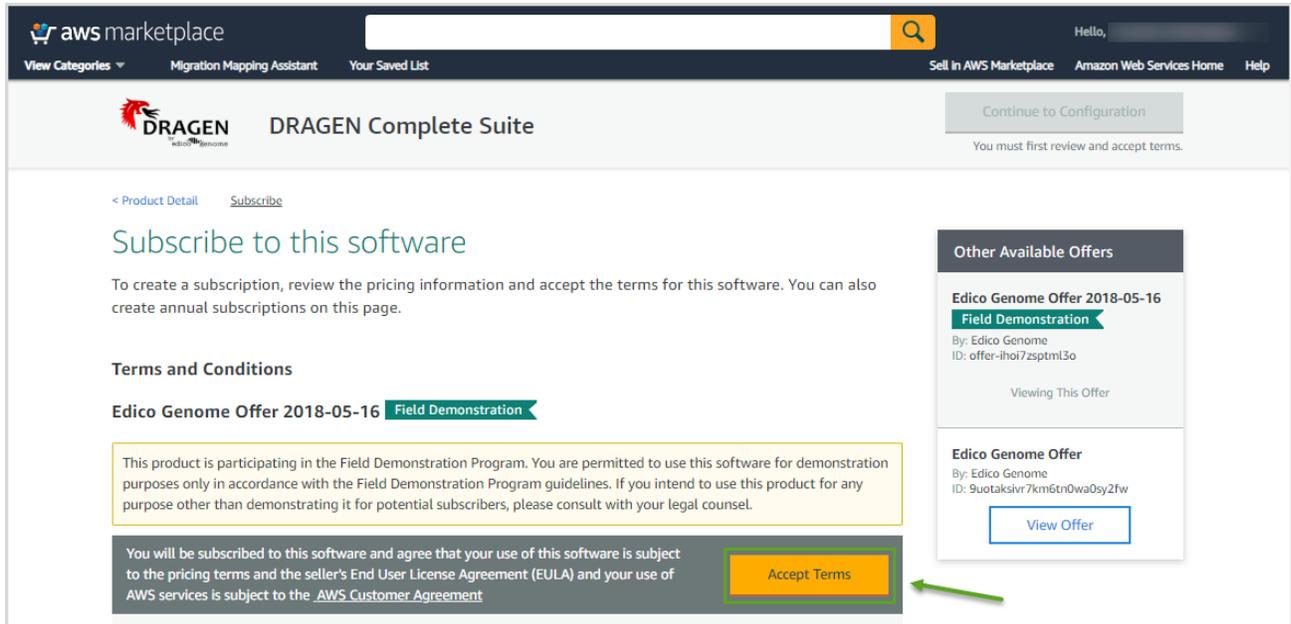
This Quick Start uses the AMI for DRAGEN Complete Suite in AWS Marketplace and requires that you accept the terms for the AMI from the AWS account where the Quick Start will be deployed.

1. Log in to the AWS Marketplace at <https://aws.amazon.com/marketplace>.
2. Open the page for [DRAGEN Complete Suite](#).

The screenshot shows the AWS Marketplace interface for the DRAGEN Complete Suite. At the top, there's a navigation bar with 'aws marketplace' and a search icon. Below that, there are tabs for 'View Categories', 'Migration Mapping Assistant', and 'Your Saved List'. The main content area features the product title 'DRAGEN Complete Suite' by Edico Genome, with the latest version 2.6.3. A description states it enables ultra-rapid analysis of NGS data. The operating system is Linux/Unix, and it has a 5-star rating with 1 review. A 'Free Trial' button is visible. A prominent orange 'Continue to Subscribe' button is highlighted with a green box and arrow. Below it is a 'Save to List' button. A pricing box indicates a typical total price of \$35.300/hr. The page also includes a 'Product Overview' section with a detailed description and a 'Highlights' section listing 'Ultra-Rapid', 'Highest Accuracy', and 'Lowest Cost'.

**Figure 2: DRAGEN Complete Suite in AWS Marketplace**

3. Choose **Continue to Subscribe** to view the license terms and pricing information.
4. Choose **Accept Terms**. For detailed instructions, see the [AWS Marketplace documentation](#).



**Figure 3: Accepting software terms in AWS Marketplace**

- When the subscription process is complete, exit out of AWS Marketplace without further action. **Do not** provision the software from AWS Marketplace—the Quick Start will deploy the AMI for you.

### Step 3. Launch the Quick Start

**Note** You are responsible for the cost of the AWS services used while running this Quick Start reference deployment. There is no additional cost for using this Quick Start. For full details, see the pricing pages for each AWS service you will be using in this Quick Start. Prices are subject to change.

- Choose one of the following options to launch the AWS CloudFormation template into your AWS account. For help choosing an option, see [deployment options](#) earlier in this guide.

<p><b>Option 1</b></p> <p><b>Deploy DRAGEN into a new VPC on AWS</b></p> <p style="background-color: #0070C0; color: white; padding: 5px 20px; border-radius: 5px; font-weight: bold; font-size: 1.2em;">Launch</p>	<p><b>Option 2</b></p> <p><b>Deploy DRAGEN into an existing VPC on AWS</b></p> <p style="background-color: #0070C0; color: white; padding: 5px 20px; border-radius: 5px; font-weight: bold; font-size: 1.2em;">Launch</p>
---	---

**Important** If you're deploying DRAGEN into an existing VPC, make sure that your VPC has two private subnets in different Availability Zones for the AWS Batch compute environments. These subnets require [NAT gateways or NAT instances](#) in their route tables, to allow the instances to download packages and software without exposing them to the internet. You will also need the domain name option configured in the DHCP options as explained in the [Amazon VPC documentation](#). You will be prompted for your VPC settings when you launch the Quick Start.

Each deployment takes about 15 minutes to complete.

2. Check the region that's displayed in the upper-right corner of the navigation bar, and change it if necessary. This is where the network infrastructure for DRAGEN will be built. The template is launched in the US East (Virginia) Region by default.

**Important** This Quick Start includes F1 instances, which aren't available in all AWS Regions. For a list of supported regions, see the [Amazon EC2 Pricing](#) webpage.

3. On the **Select Template** page, keep the default setting for the template URL, and then choose **Next**.
4. On the **Specify Details** page, change the stack name if needed. Review the parameters for the template. Provide values for the parameters that require input. For all other parameters, review the default settings and customize them as necessary. When you finish reviewing and customizing the parameters, choose **Next**.

In the following tables, parameters are listed by category and described separately for the two deployment options:

- [Parameters for deploying DRAGEN into a new VPC](#)
- [Parameters for deploying DRAGEN into an existing VPC](#)

- **Option 1: Parameters for deploying DRAGEN into a new VPC**

[View template](#)

*Network Configuration:*

Parameter label (name)	Default	Description
<b>Availability Zones</b> (AvailabilityZones)	<i>Requires input</i>	The list of Availability Zones to use for the subnets in the VPC. The Quick Start uses two Availability Zones from your list and preserves the logical order you specify.

Parameter label (name)	Default	Description
<b>VPC CIDR</b> (VPCCIDR)	10.0.0.0/16	The CIDR block for the VPC.
<b>Private Subnet 1 CIDR</b> (PrivateSubnet1CIDR)	10.0.0.0/19	The CIDR block for the private subnet located in Availability Zone 1.
<b>Private Subnet 2 CIDR</b> (PrivateSubnet2CIDR)	10.0.32.0/19	The CIDR block for the private subnet located in Availability Zone 2.
<b>Public Subnet 1 CIDR</b> (PublicSubnet1CIDR)	10.0.128.0/20	The CIDR block for the public (DMZ) subnet located in Availability Zone 1.
<b>Public Subnet 2 CIDR</b> (PublicSubnet2CIDR)	10.0.144.0/20	The CIDR block for the public (DMZ) subnet located in Availability Zone 2.

### DRAGEN Quick Start Configuration:

Parameter label (name)	Default	Description
<b>Key Pair Name</b> (KeyPairName)	<i>Requires input</i>	A public/private key pair, which allows you to connect securely to your instance after it launches. When you created an AWS account, this is the key pair you created in your preferred region.
<b>Instance Type</b> (InstanceType)	f1.2xlarge	The EC2 instance type for DRAGEN instances in the AWS Batch compute environment. This must be an instance type in the <a href="#">F1 instance family</a> , because DRAGEN requires a field-programmable gate array (FPGA).  You should make sure that the F1 instance limits in your AWS account support the maximum number of simultaneous DRAGEN jobs that you expect to run, as discussed in the <a href="#">Technical Requirements</a> section.
<b>Spot Bid Percentage</b> (BidPercentage)	50	The bid percentage set for your AWS Batch managed compute environment with Spot Instances. Specify a value between 1 and 100.  The bid percentage specifies the maximum percentage that a Spot Instance price can be when compared with the On-Demand price for that instance type before instances are launched. For example, if you set this parameter to 20, the Spot price must be below 20% of the current On-Demand price for that EC2 instance. You always pay the lowest (market) price and never more than your maximum percentage.
<b>Min vCPUs</b> (MinvCpus)	0	The minimum number of virtual CPUs for your AWS Batch compute environment. You can specify a value between 0 and 1000. We recommend keeping the default value of 0.

Parameter label (name)	Default	Description
<b>Max vCPUs</b> (MaxvCpus)	<i>Requires input</i>	The maximum number of virtual CPUs for your AWS Batch compute environment. You can specify a value between 0 and 1000.
<b>Desired vCPUs</b> (DesiredvCpus)	0	The desired number of virtual CPUs for your AWS Batch compute environment. You can specify a value between 0 and 1000. We recommend that you use the same number as the <b>Min vCPUs</b> parameter to optimize costs.
<b>Genomics Data Bucket</b> (GenomicsS3Bucket)	<i>Requires input</i>	The S3 bucket to be used for reading and writing genomics data. The bucket name can include numbers, lowercase letters, uppercase letters, and hyphens, but should not start or end with a hyphen.  This must be an existing S3 bucket that contains your genomic input datasets, DRAGEN-specific reference hash tables, and an outputs folder, as explained in the <a href="#">Technical Requirements</a> section.
<b>AWS Batch Retry Number</b> (RetryNumber)	1	The number of times the AWS Batch job will be retried if it fails. You can specify a value between 1 and 10. For more information, see the <a href="#">AWS Batch documentation</a> .

### AWS Quick Start Configuration:

Parameter label (name)	Default	Description
<b>Quick Start S3 Bucket Name</b> (QSS3BucketName)	aws-quickstart	The S3 bucket you have created for your copy of Quick Start assets, if you decide to customize or extend the Quick Start for your own use. The bucket name can include numbers, lowercase letters, uppercase letters, and hyphens, but should not start or end with a hyphen.
<b>Quick Start S3 Key Prefix</b> (QSS3KeyPrefix)	quickstart-illumina-dragen/	The <a href="#">S3 key name prefix</a> used to simulate a folder for your copy of Quick Start assets, if you decide to customize or extend the Quick Start for your own use. This prefix can include numbers, lowercase letters, uppercase letters, hyphens, and forward slashes.

- **Option 2: Parameters for deploying DRAGEN into an existing VPC**

[View template](#)

### Network Configuration:

Parameter label (name)	Default	Description
<b>VPC ID</b> (VPCID)	<i>Requires input</i>	The ID of your existing VPC (e.g., vpc-0343606e).

Parameter label (name)	Default	Description
<b>Private Subnet 1 ID</b> (PrivateSubnet1ID)	<i>Requires input</i>	The ID of the private subnet in Availability Zone 1 in your existing VPC (e.g., subnet-a0246dcd).
<b>Private Subnet 2 ID</b> (PrivateSubnet2ID)	<i>Requires input</i>	The ID of the private subnet in Availability Zone 2 in your existing VPC (e.g., subnet-b58c3d67).

### *DRAGEN Quick Start Configuration:*

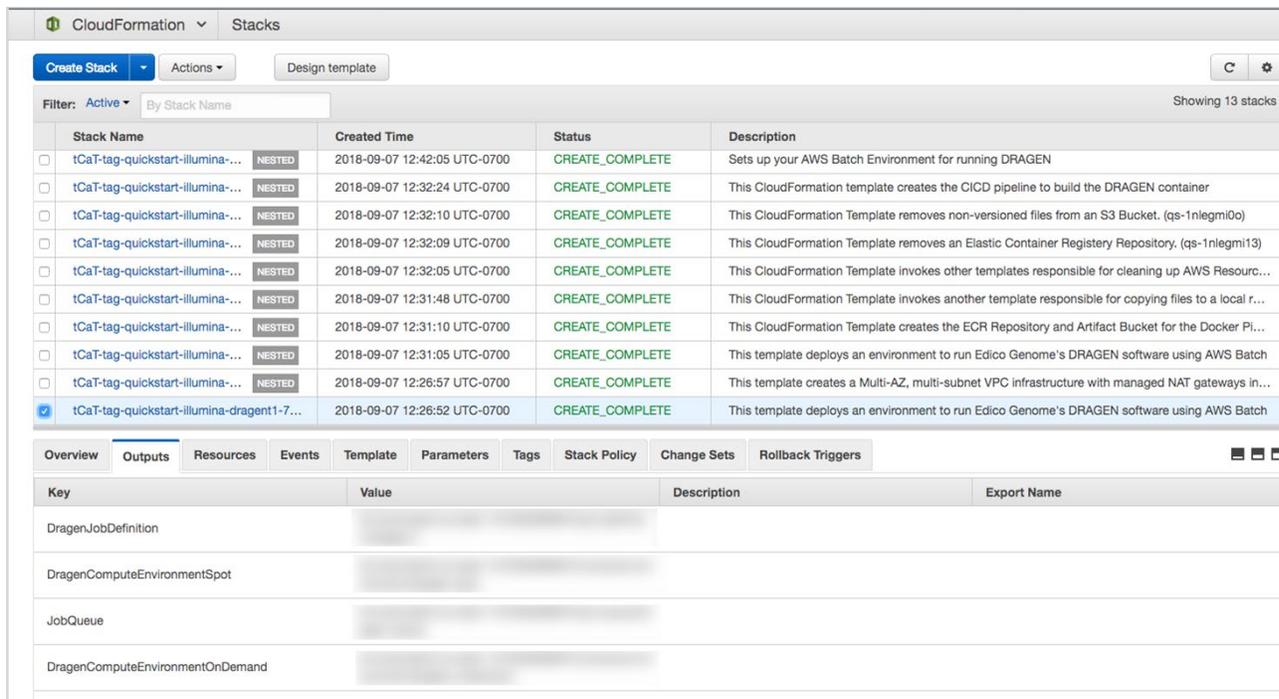
Parameter label (name)	Default	Description
<b>Key Pair Name</b> (KeyPairName)	<i>Requires input</i>	A public/private key pair, which allows you to connect securely to your instance after it launches. When you created an AWS account, this is the key pair you created in your preferred region.
<b>Instance Type</b> (InstanceType)	f1.2xlarge	The EC2 instance type for DRAGEN instances in the AWS Batch compute environment. This must be an instance type in the <a href="#">F1 instance family</a> , because DRAGEN requires a field-programmable gate array (FPGA). You should make sure that the F1 instance limits in your AWS account support the maximum number of simultaneous DRAGEN jobs that you expect to run, as discussed in the <a href="#">Technical Requirements</a> section.
<b>Spot Bid Percentage</b> (BidPercentage)	50	The bid percentage set for your AWS Batch managed compute environment with Spot Instances. Specify a value between 1 and 100. The bid percentage specifies the maximum percentage that a Spot Instance price can be when compared with the On-Demand price for that instance type before instances are launched. For example, if you set this parameter to 20, the Spot price must be below 20% of the current On-Demand price for that EC2 instance. You always pay the lowest (market) price and never more than your maximum percentage.
<b>Min vCPUs</b> (MinvCpus)	0	The minimum number of virtual CPUs for your AWS Batch compute environment. You can specify a value between 0 and 1000. We recommend keeping the default value of 0.
<b>Max vCPUs</b> (MaxvCpus)	<i>Requires input</i>	The maximum number of virtual CPUs for your AWS Batch compute environment. You can specify a value between 0 and 1000.
<b>Desired vCPUs</b> (DesiredvCpus)	0	The desired number of virtual CPUs for your AWS Batch compute environment. You can specify a value between 0 and 1000. We recommend that you use the same number as the <b>Min vCPUs</b> parameter to optimize costs.

Parameter label (name)	Default	Description
<b>Genomics Data Bucket</b> (GenomicsS3Bucket)	<i>Requires input</i>	The S3 bucket to be used for reading and writing genomics data. The bucket name can include numbers, lowercase letters, uppercase letters, and hyphens, but should not start or end with a hyphen.  This must be an existing S3 bucket that contains your genomic input datasets, DRAGEN-specific reference hash tables, and an outputs folder, as explained in the <a href="#">Technical Requirements</a> section.
<b>AWS Batch Retry Number</b> (RetryNumber)	1	The number of times the AWS Batch job will be retried if it fails. You can specify a value between 1 and 10. For more information, see the <a href="#">AWS Batch documentation</a> .

### *AWS Quick Start Configuration:*

Parameter label (name)	Default	Description
<b>Quick Start S3 Bucket Name</b> (QSS3BucketName)	aws-quickstart	The S3 bucket you have created for your copy of Quick Start assets, if you decide to customize or extend the Quick Start for your own use. The bucket name can include numbers, lowercase letters, uppercase letters, and hyphens, but should not start or end with a hyphen.
<b>Quick Start S3 Key Prefix</b> (QSS3KeyPrefix)	quickstart-illumina-dragen/	The <a href="#">S3 key name prefix</a> used to simulate a folder for your copy of Quick Start assets, if you decide to customize or extend the Quick Start for your own use. This prefix can include numbers, lowercase letters, uppercase letters, hyphens, and forward slashes.

- On the **Options** page, you can [specify tags](#) (key-value pairs) for resources in your stack and [set advanced options](#). When you're done, choose **Next**.
- On the **Review** page, review and confirm the template settings. Under **Capabilities**, select the check box to acknowledge that the template will create IAM resources.
- Choose **Create** to deploy the stack.
- Monitor the status of the stack. When the status is **CREATE\_COMPLETE**, the DRAGEN cluster is ready, as shown in Figure 4.
- Use the URLs displayed in the **Outputs** tab for the stack to view the resources that were created.



The screenshot shows the AWS CloudFormation console. At the top, there are tabs for 'Overview', 'Outputs', 'Resources', 'Events', 'Template', 'Parameters', 'Tags', 'Stack Policy', 'Change Sets', and 'Rollback Triggers'. The 'Outputs' tab is selected, displaying a table with columns for 'Key', 'Value', 'Description', and 'Export Name'. The table lists four outputs: 'DrogenJobDefinition', 'DrogenComputeEnvironmentSpot', 'JobQueue', and 'DrogenComputeEnvironmentOnDemand'. The 'Value' column contains redacted information.

Key	Value	Description	Export Name
DrogenJobDefinition	[REDACTED]		
DrogenComputeEnvironmentSpot	[REDACTED]		
JobQueue	[REDACTED]		
DrogenComputeEnvironmentOnDemand	[REDACTED]		

Figure 4: Stack outputs

## Step 4. Test the Deployment

When the deployment is complete, you can verify the environment by running a DRAGEN job. This job reads the input datasets and the reference hash table from the S3 genomics data bucket you specified when you deployed the Quick Start, and outputs the analysis results into the designated output location in the S3 bucket.

As a quick test, we recommend using an input dataset consisting of a smaller FASTQ pair, such as one of the following data samples:

- [NIST7035\\_TAAGGCGA\\_L001\\_R1\\_001\\_trimmed.fastq.gz](#)
- [NIST7035\\_TAAGGCGA\\_L001\\_R2\\_001\\_trimmed.fastq.gz](#)

Use AWS Batch to launch the job either through the AWS Command Line Interface (AWS CLI) or the console. Provide the DRAGEN job parameters as commands (in the command array within the `containerOverrides` field in the AWS CLI or the **Command** field in the console). For a full list of options, see the **ILLUMINA DRAGEN Getting Started Guide** or **ILLUMINA DRAGEN Bio-IT Platform User Guide** available on the [ILLUMINA DRAGEN Bio-IT Platform Support Resources](#) website.

Most of the options function exactly as they are described in the guide. However, for options that refer to local files or directories, provide either an HTTPS pre-signed URL (for files) or specify the full path to the S3 bucket that contains the genomics data (for directories).

In the following sections, we've provided instructions for running an end-to-end DRAGEN job using both methods (AWS CLI and console). In this example, the DRAGEN job handles mapping, aligning, sorting, deduplication, and variant calling. An input dataset with a paired FASTQ file is used to generate a variant call format (VCF) output file.

### *Option 1: Use the AWS CLI*

The simplest way to run a DRAGEN job by using the AWS CLI is to create an input JSON file that describes the job. Here's an example of a JSON input file named `e2e-job.json`:

```
{
  "jobName": "e2e-job",
  "jobQueue": "dragen-queue",
  "jobDefinition": "dragen",
  "containerOverrides": {
    "vcpus": 8,
    "memory": 120000,
    "command": [
      "--ref-dir",
      "s3://<bucket/path>",
      "-1",
      "https://<S3_URL>",
      "-2",
      "https://<S3_URL>",
      "--output-directory",
      "s3://<bucket/path>",
      "--enable-duplicate-marking",
      "true",
      "--output-file-prefix",
      "NA12878_S1_L004_R1_001",
      "--enable-map-align",
      "true",
      "--output-format",
      "BAM",
      "--enable-variant-caller",
      "true",
      "--vc-sample-name",
      "DRAGEN_RGSM"
    ]
  },
  "retryStrategy": {
    "attempts": 1
  }
}
```

```
}  
}
```

You can then launch the job from the command line by using the [submit-job](#) command and specifying the `e2e-job.json` file as input:

```
> aws batch submit-job --cli-input-json file:///e2e-job.json
```

### *Option 2: Use the AWS Batch Console*

To run the DRAGEN job from the console:

1. Open the AWS Batch console at <https://console.aws.amazon.com/batch/>.
2. From the navigation bar, choose the AWS Region you used for the Quick Start deployment.
3. In the navigation pane, choose **Jobs, Submit job**.
4. Fill out these fields, as shown in Figure 5:
  - **Job name:** Enter a unique name for the job.
  - **Job definition:** Choose the DRAGEN job definition that was created by the Quick Start and displayed in the **Outputs** tab of the AWS CloudFormation console in step 3(9).
  - **Job queue:** Choose `dragen-queue`, which was created by the Quick Start.
  - **Job type:** Choose **Single**.
  - **Command:** Specify the DRAGEN-specific parameters shown in the JSON command array in [option 1](#).
  - **vCPUs, Memory, Job attempts, Execution timeout:** Keep the defaults that are specified in the job definition.

For more information, see the [AWS Batch documentation](#).

5. Choose **Submit job**.

Jobs > Submit job  
**Submit an AWS Batch Job**

Select a job definition to run and apply any command, parameter, or environment variable overrides.

Job run-time

Job name

Job definition

Job queue

Environment

Job Type  Single  
 Array

Job depends on

Command

```
--ref-dir s3://<bucket/path> -1 https://<S3_URL> -2 https://<S3_URL> --output-directory  
s3://<bucket/path> --enable-duplicate-marking true --output-file-prefix NA12878 --  
enable-map-align true --output-format BAM --enable-variant-caller true --vc-sample-  
name DRAGEN_RGSM
```

Text is delimited by spaces and converted to a JSON array to be submitted with the job.  
JSON result:

```
[  
  "--ref-dir",  
  "s3://<bucket/path>",  
  "-1",  
  "https://<S3_URL>",  
  "-2",  
  "https://<S3_URL>",  
  "--output-directory",  
  "s3://<bucket/path>",  
  "--enable-duplicate-marking",  
  "true",  
  "--output-file-prefix",  
  "NA12878",  
  "--enable-map-align",  
  "true",  
  "--output-format",  
  "BAM",  
  "--enable-variant-caller",  
  "true",  
  "--vc-sample-name",  
  "DRAGEN_RGSM"  
]
```

For more details on how the command field is used in your job definition, view our [documentation](#).

vCPUs   
AWS batch will find a place to run your work that has at least the amount of vCPUs you specify.

Memory (MiB)   
The amount of memory you need for your job to run.

Job attempts

Execution timeout   
Time (in seconds) to allow each job attempt to run. If your job runs longer than the specified time, it will stop and be moved

Parameters

Key	Value
<a href="#">Add parameter</a>	

Environment variables

Key	Value
<a href="#">Add environment variable</a>	

Figure 5: Running a DRAGEN job from the AWS Batch console

6. Monitor the job status in the AWS Batch window to see if it succeeded or failed. For more information about job states and exit codes, see the [AWS Batch documentation](#).

## Best Practices Using DRAGEN on AWS

For simplicity, we recommend that you create your S3 bucket in the AWS Region that you are deploying the Quick Start into. In some use cases, as outlined in the [DRAGEN User Guide](#) on the Illumina website, you might need to attach EBS volumes to instances. The DRAGEN guides are available as links from the [DRAGEN Complete Suite webpage](#) in AWS Marketplace (see the *Usage Information* section on that page).

## Security

DRAGEN doesn't enforce any specific security requirements. However, for security, this Quick Start deploys DRAGEN into private subnets that aren't externally reachable from outside the VPC (they can access the internet only through NAT gateways). Please consult your IT and security teams for image hardening, encryption, and other security requirements.

## FAQ

**Q.** I encountered a `CREATE_FAILED` error when I launched the Quick Start.

**A.** If AWS CloudFormation fails to create the stack, we recommend that you relaunch the template with **Rollback on failure** set to **No**. (This setting is under **Advanced** in the AWS CloudFormation console, **Options** page.) With this setting, the stack's state will be retained and the instance will be left running, so you can troubleshoot the issue.

**Important** When you set **Rollback on failure** to **No**, you will continue to incur AWS charges for this stack. Please make sure to delete the stack when you finish troubleshooting.

For additional information, see [Troubleshooting AWS CloudFormation](#) on the AWS website.

**Q.** I encountered a size limitation error when I deployed the AWS CloudFormation templates.

**A.** We recommend that you launch the Quick Start templates from the links in this guide or from another S3 bucket. If you deploy the templates from a local copy on your computer or from a non-S3 location, you might encounter template size limitations when you create the stack. For more information about AWS CloudFormation limits, see the [AWS documentation](#).

**Q.** How can I find out whether my DRAGEN job has completed?

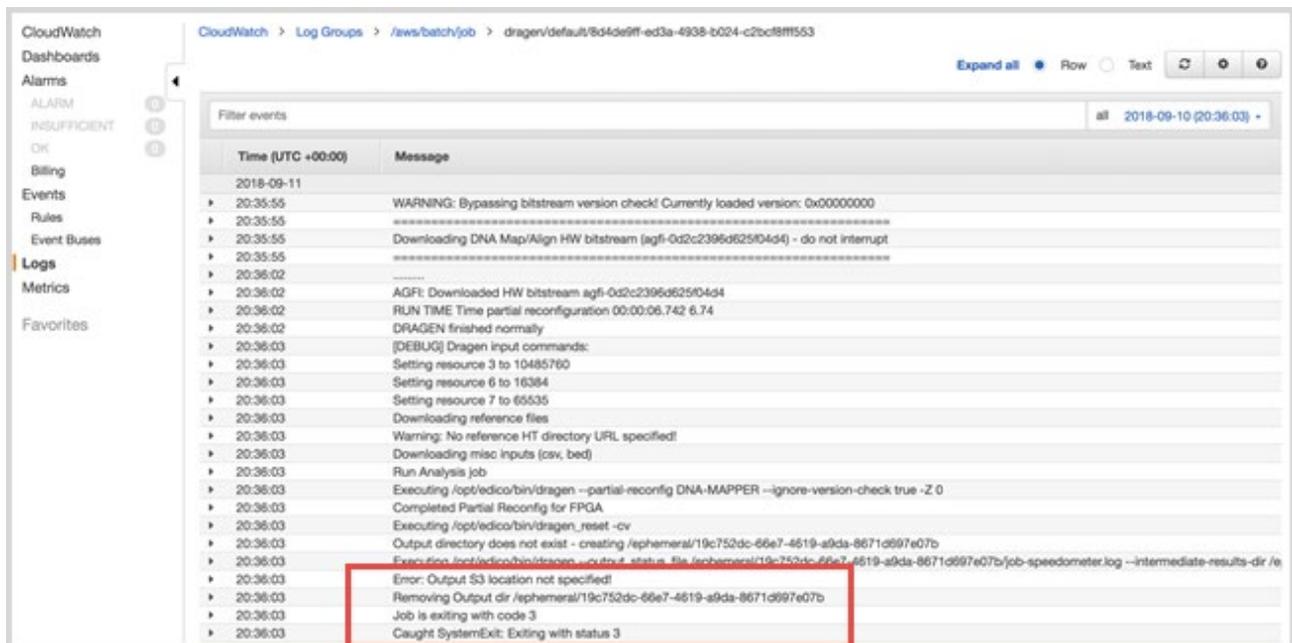
**A.** When you submit a DRAGEN job by using the AWS CLI or the AWS Batch console, as described previously in [step 4](#), you will get an AWS Batch job ID. You should monitor the job status to check for completion. When the job has completed successfully, its job state will be displayed as SUCCEEDED and the outputs will be uploaded into the output S3 bucket location that is specified in the job. For more information about job states, see the [AWS Batch documentation](#).

**Q.** What should I do if the DRAGEN job fails?

**A.** If your DRAGEN job fails, more information will be available in Amazon Cloudwatch Logs. You can access the log either from the CloudWatch console (see Figure 6) or through the **View logs** link in the AWS Batch job (see Figure 7).

In these examples, the user didn't specify an S3 bucket for the **Genomics Data Bucket** parameter when they deployed the Quick Start, which caused the error:

```
20:36:03. Error: Output S3 location not specified!
```



**Figure 6: Accessing logs from the CloudWatch console**

**Job status**

Status FAILED

Status reason Essential container in task exited

Container message

Created at 10:25:11 am 09/13/18

Started at 10:29:30 am 09/13/18

Queue

**Job attributes**

Job id d867d40a-2e90-4381-85c0-0ab660cc9f60

Job name test

Job definition

Depends on

**Attempts**

Attempts	Exit code	Start time	End time
1 of 1	<a href="#">3 - View logs</a>	10:29:30 am 09/13/18	10:29:52 am 09/13/18

Figure 7: Accessing logs from the AWS Batch console

When you run a DRAGEN job, you use the `--output-directory` parameter to specify an S3 output bucket, as described in [step 4](#). If the DRAGEN job fails, more information about the failure is provided in the DRAGEN process output log, which is uploaded to the S3 bucket you specified. Look for the object in the S3 bucket that has the key name “`dragen_log_timestamp.txt`” for the output log.

Generally, DRAGEN process logs will give adequate information about failures, but if you need further help, please contact Illumina at [techsupport@illumina.com](mailto:techsupport@illumina.com).

## GitHub Repository

You can visit our [GitHub repository](#) to download the templates and scripts for this Quick Start, to post your comments, and to share your customizations with others. The steps to clone the GitHub repository for customization is documented in the [Quick Start Contributor's Guide](#).

## Additional Resources

### AWS services

- Amazon EC2  
<https://aws.amazon.com/documentation/ec2/>
- Amazon VPC  
<https://aws.amazon.com/documentation/vpc/>
- AWS CloudFormation  
<https://aws.amazon.com/documentation/cloudformation/>

### DRAGEN documentation

- Illumina DRAGEN Bio-IT Platform Support Resources  
[https://sapac.support.illumina.com/sequencing/sequencing\\_software/dragen-bio-it-platform.html](https://sapac.support.illumina.com/sequencing/sequencing_software/dragen-bio-it-platform.html)
- Performance studies
  - [PrecisionFDA Hidden Treasures – Warm Up](#)
  - [How to Train Your DRAGEN – Evaluating and Improving Edico Genome's Rapid WGS Tools](#)

### Public genomic data

- Human genome reference  
<http://hgdownload.cse.ucsc.edu/goldenPath/hg19/bigZips/chromFa.tar.gz>
- Small data samples for testing
  - [ftp://ftp-trace.ncbi.nlm.nih.gov/giab/ftp/data/NA12878/Garvan\\_NA12878\\_HG001\\_HiSeq\\_Exome/NIST7035\\_TAAGCGA\\_L001\\_R1\\_001\\_trimmed.fastq.gz](ftp://ftp-trace.ncbi.nlm.nih.gov/giab/ftp/data/NA12878/Garvan_NA12878_HG001_HiSeq_Exome/NIST7035_TAAGCGA_L001_R1_001_trimmed.fastq.gz)

- [ftp://ftp-trace.ncbi.nlm.nih.gov/giab/ftp/data/NA12878/Garvan\\_NA12878\\_HG001\\_HiSeq\\_Exome/NIST7035\\_TAAAGCGA\\_L001\\_R2\\_001\\_trimmed.fastq.gz](ftp://ftp-trace.ncbi.nlm.nih.gov/giab/ftp/data/NA12878/Garvan_NA12878_HG001_HiSeq_Exome/NIST7035_TAAAGCGA_L001_R2_001_trimmed.fastq.gz)

## Quick Start reference deployments

- AWS Quick Start home page  
<https://aws.amazon.com/quickstart/>

## Document Revisions

Date	Change	In sections
May 2020	Updated documentation links	
October 2018	Initial publication	–

© 2020, Amazon Web Services, Inc. or its affiliates, and Illumina. All rights reserved.

### **Notices**

This document is provided for informational purposes only. It represents AWS's current product offerings and practices as of the date of issue of this document, which are subject to change without notice. Customers are responsible for making their own independent assessment of the information in this document and any use of AWS's products or services, each of which is provided "as is" without warranty of any kind, whether express or implied. This document does not create any warranties, representations, contractual commitments, conditions or assurances from AWS, its affiliates, suppliers or licensors. The responsibilities and liabilities of AWS to its customers are controlled by AWS agreements, and this document is not part of, nor does it modify, any agreement between AWS and its customers.

The software included with this paper is licensed under the Apache License, Version 2.0 (the "License"). You may not use this file except in compliance with the License. A copy of the License is located at <http://aws.amazon.com/apache2.0/> or in the "license" file accompanying this file. This code is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.