

CLC Genomics Cloud Engine security and reliability

Abstract: This white paper describes how the CLC Genomics Cloud Engine (GCE) delivers a high level of security and reliability. Users can trust the system to keep their data secure, ensuring that it can be used in a way that is compliant with national or international regulations.

Introduction to CLC Genomics Cloud Engine

The CLC Genomics Cloud Engine is a system that can be deployed on a user-owned Amazon Web Services (AWS) account, enabling NGS analysis of data stored in S3. It is ideal for use in cases where the same analysis is performed in a highly repetitive fashion, for example, on multiple samples.

Stability

When the CLC Genomics Cloud Engine is deployed, it will set up network configurations, provision server instances and install software.

Multi-location redundancy

GCE has redundant frontend and database servers to ensure the availability of the service. If a single server unexpectedly fails, another server will seamlessly handle all requests. A server that shuts down will automatically be replaced by a new one.

The redundant servers are placed in multiple AWS Availability zones (data centers). Placing the servers in physically dispersed locations ensures that a local disaster will not interfere with the system's availability.

The design of GCE removes the risk of cascading failures, where the failure of one server causes other servers to fail and ultimately result in complete system failure. GCE servers depend on highly reliable AWS messaging services (AWS SQS) for communication. Each server can continue to function regardless of the failure of other servers in the system.

Analysis isolation

An analysis performed by GCE is run independently and will not be affected by any other analysis being processed at the same time. This ensures safe, reliable and highly scalable operation.

World class cloud storage

GCE works in tandem with AWS S3, which is the most feature-rich object storage platform available in the cloud today. It is designed to deliver 99.999999999% durability and 99.99% availability through massive redundancy and security features (see <https://aws.amazon.com/s3/>).

Certified quality assurance

The CLC Genomics Cloud Engine has been thoroughly tested and documented by QIAGEN® Bioinformatics. All development is carried out by QIAGEN Aarhus A/S under development standards, which have been ISO 9001 certified.

Secure service access

When the CLC Genomics Cloud Engine is deployed, it integrates with the customer's own authentication service. The integration is based on the OAuth2 standard and provides a single sign-on experience.

Depending on the required level of security, the customer can connect to GCE in three different ways of increasing security:

1. Internet gateway
2. VPN connection
3. Direct connection

The servers running the GCE service can be set up to automatically install security updates on a weekly basis, to ensure that known security holes cannot be exploited.



Protected network for data processing

The analysis of samples is performed on backend servers placed in an isolated subnet. Communication is limited to the communication needed for job distribution and reporting of progress and performed exclusively over Amazon® SQS messaging. This keeps the software that accesses sensitive data isolated and safe.

Secure data access

When using the CLC Genomics Cloud Engine, input and output data is stored in AWS S3. S3 can be configured in several different ways to keep your data safe, including support for multi-factor login.

GCE supports working with data in S3 that is server-side encrypted using the 256-bit Advanced Encryption Standard (AES-256).

S3 allows fine grained access control of files, to ensure that you can specifically control who has access to what. GCE supports this by enforcing that a user cannot perform any analysis on data that the person does not have access to read.

Compliance

Distribution of responsibilities

The CLC Genomics Cloud Engine is a package comprising of both server software and infrastructure configurations. The package is deployed on the customer's own AWS account. Consequently, the GCE user will be a customer of AWS and can enter a Business Associate Agreement with Amazon and utilize Amazon's extensive experience with a long list of regulations and standards.

AWS compliance

GCE works in tandem with AWS S3, where GCE is responsible for processing data and S3 is responsible for storage of data. Due to this segregation, storage-related compliance is ensured by AWS, while compliance related to the processing is ensured by GCE.

To learn more about AWS policies related to privacy and data security, please read the following pages:

- **AWS customer agreement**
- **EU data protection**

Some regulations are related to the geographical location of the servers running the system. With AWS, you can freely choose where to store your data and run your analysis. Amazon has data centers in several countries around the world and are constantly expanding. A list of current and coming regions can be found at <https://aws.amazon.com/about-aws/global-infrastructure/>.

For up-to-date licensing information and product-specific disclaimers, see the respective user manual. QIAGEN user manuals are available at www.qiagenbioinformatics.com or can be requested from QIAGEN Technical Services or your local distributor.