BaseSpace® Sequence Hub Data Security

Data is safe in the Illumina cloud analysis and storage platform.

Introduction

Next-generation sequencing (NGS) has fundamentally altered biomedical research, generating large amounts of analytical data. This large-scale data requires a scalable, robust, and secure storage and analysis solution. The BaseSpace Sequence Hub is a genomics cloud analysis platform built by Illumina using Amazon Web Services (AWS). AWS is a leader in cloud-based infrastructure, hosting customer-facing services, and critical operations for both private industry and government departments including Treasury, DOE, and State.

Data security is a key concern in making the decision to move to cloud-based genomic storage and analysis. Illumina provides a combination of Amazon's comprehensive and well-tested approach to platform security and Illumina's own security testing and procedures. The result provides a cloud genomics solution that meets or exceeds the security provided by many institutional IT infrastructures.

The BaseSpace Hub Data Model

A sequencing run contains log files, instrument health data, run metrics, and base call information (*.bcl files), which are demultiplexed in the BaseSpace Hub to create the samples used in secondary analysis.[1]

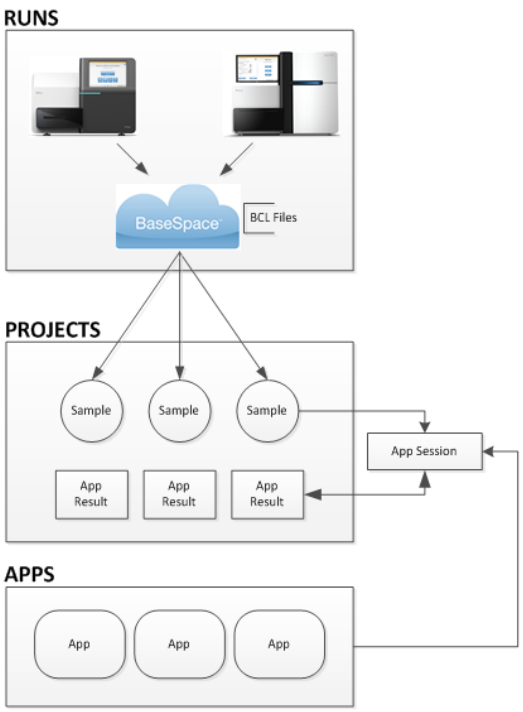

Figure 1: BaseSpace Hub Data Model—BaseSpace Hub data are stored according to a data model that includes runs, samples, app results, and app sessions.

Samples are analyzed by launching BaseSpace Apps. BaseSpace Apps are processing software and routines that interact with BaseSpace data through the API. User-level authentication and inflight data encryption are enforced for every app that requests access to BaseSpace data. Files that are output from apps are stored in an object called AppResults. For example, when a resequencing app executes alignment and variant calling, an AppResult is created for each sample. AppResults can be used as inputs to apps as well. App sessions are created to record every time an app is launched. Finally, projects are simple containers that store samples and AppResults (Figure 1).[1]

Security of Data in Flight

Data transfer is the major part of communication between the genomic sequencing instruments and the data analysis and storage servers. Illumina has implemented several security measures to make sure your data are protected in flight.

Secure Connection to Instrument

The user always makes the decision to send data to the BaseSpace Sequence Hub during run set-up. If BaseSpace Hub is chosen, the run is authenticated against and tracked to a user-specific BaseSpace / MyIllumina account. The user can simultaneously store data locally and in BaseSpace Hub if desired (Figure 2).

Figure 2: Run Setup Options (MiSeq® System Software)—(top) Users initiate each data session as part of the sequencing run process, and it is authenticated against a BaseSpace Hub - MyIllumina account. (bottom) The sequencer software does not maintain any hosting/IIS services and does not carry a publicly addressable IP.

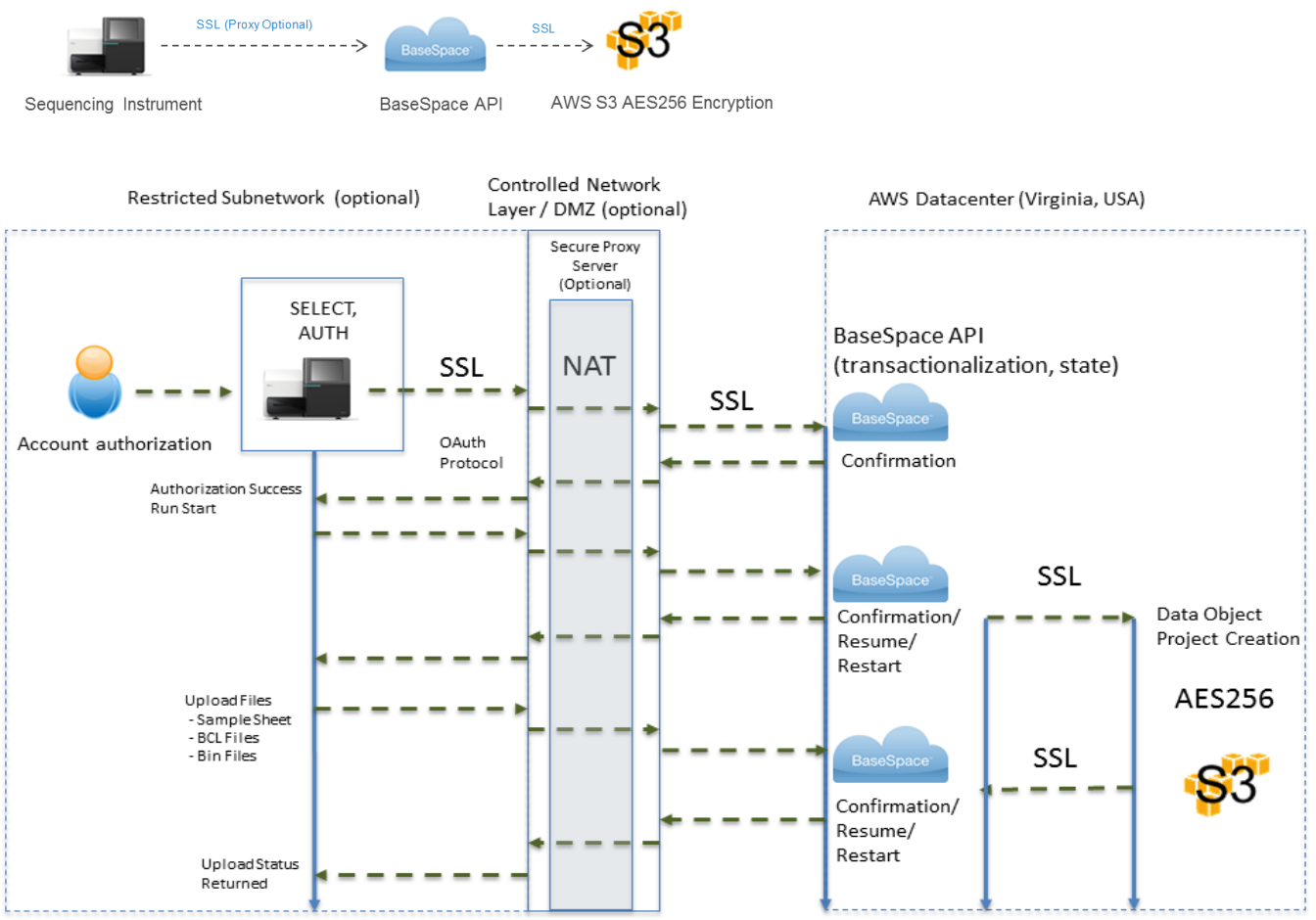

Figure 4: BaseSpace Sequence Hub Data Stream Event Map—(top) Overview of communication between sequencer and the BaseSpace Hub. (bottom) Streaming data to the BaseSpace Hub begins with user account authorization and selection of data streaming options. The data stream is fully encrypted to the BaseSpace Hub. The BaseSpace Hub does not initiate any activities to the sequencer.

Both the MiSeq and HiSeq® instruments include onboard software to encrypt, transmit, and verify small base calling files generated over the course of a sequencing run. This brokering software interacts with the BaseSpace API and allows network interrupts, latencies, and incomplete transmissions to be caught and requeued automatically.

Encryption and Authentication

Communication between instruments and the BaseSpace Hub is done through an http based API (Figure 3). This communication has the following characteristics:

- Transfer of data to the BaseSpace Hub is encrypted using the AES256 standard, which is a high-level security measure, safe enough for top secret US government information.[2]
- All traffic is over Secured SSL Transport. SSL (Secure Sockets Layer) is an internet standard that has been developed to encrypt sensitive communications as they pass over the internet. SSL is regularly updated to allow for new technologies and new threats. Just as millions of people and institutions rely on SSL every day to protect their financial transactions, we use SSL to protect BaseSpace Hub data uploads and downloads.
- All service methods require API key signatures; service is refused to all others
- Requests are monitored for abuse and IPs and/or API keys could be blacklisted.

Protection of Data at Rest in the Cloud

BaseSpace Sequence Hub can be described as Software-as-a-Service components (eg the UI and the Apps) developed on the BaseSpace Platform-as-a-Service (eg the databases and workflows). This is then built on the Amazon Web Services Infrastructure-as-a-Service (Figure 4). Data storage within this infrastructure stack needs to be secure; Amazon has taken many measures to make sure that data in the cloud is protected.

AWS Security Standards

Amazon's own security processes and standards are publicly available for review.[3,4] AWS standards and accreditation include:

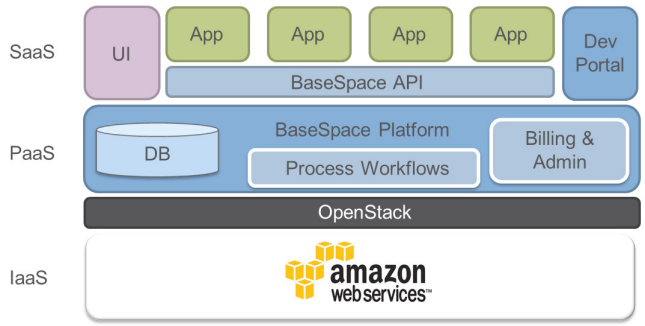

Figure 3: BaseSpace Sequence Hub Architecture—The BaseSpace Hub architecture stack consists of Amazon Web Services (AWS) Infrastructure-as-a-Service (IaaS), the BaseSpace Platform-as-a-Service (PaaS), and Software-as-a-Service (SaaS) components.

- SOC 1/SSAE 16/ISAE 3402.
  The Service Organization Controls 1 audit verifies that AWS' controls to protect customer data are properly designed and that the individual controls are operating effectively.
- FISMA moderate.
  This is an accreditation granted by the US Federal Government to strengthen federal information system security. For reference, the NIH's own data centers are rated FISMA moderate.
- PCI DSS Level 1.
  The Payment Card Industry Data Security Standard is set up to increase electronic payment security. AWS is rated at the highest level.
- ISO 27001.
  This is a widely recognized international security standard that specifies security management best practices and comprehensive security controls.
- FIPS 140-2.
  The Federal Information Processing Standard (FIPS) Publication 140-2 is a US government computer security standard that specifies the requirements for cryptography modules.

Data Storage

All run and sample data are stored in S3 and made private. No files are allowed to be accessed publicly and must be either created through the API or website and can only be downloaded through the website.

BaseSpace Sequence Hub, through AWS, synchronously stores your data across multiple facilities, performs regular data integrity checks, and is automatically self-healing. You are protected against any accidental loss of data. However, be conscious that BaseSpace Hub is not an unlimited back-up system. If you decide to delete your data, there is no mechanism to retrieve it. Also, if you transfer ownership, the new owner can delete the data, without you being able to retrieve it.

Also, AWS data centers are protected by security staff and controlled access procedures. Staff with system access undergoes background checks, and all hardware is located behind firewalls that are configured by default to block all traffic. Operating security patches are automatically applied to AWS servers, including BaseSpace Hub servers.

Comparison with Institutional Security

The BaseSpace Sequence Hub is a secure environment for your data, and meets the safety standards maintained at your institute. It may even exceed the security provided by some institutional IT infrastructures. Ask your IT personnel if they maintain the following practices:

- Data Encryption.
  We encrypt all uploaded data using the AES256 standard, ensuring that even if all other security precautions were circumvented, the stolen data could not be read. This is rarely done in the institutional IT setting.
- Active Firewall Monitoring.
  It is highly likely that any computer used for storing genomic data is already connected to the internet, or at the least on an intranet that is in turn connected to the internet. Secure isolation from the internet is typically provided by a firewall device configured to protect the internal network from outside attack. BaseSpace AWS computers are also protected by firewalls; however, AWS actively monitors its firewalls to check for vulnerabilities, a service beyond the resources of most institutions.
- Third-Party Audits.
  To verify end-to-end security, a third-party computer security firm has been retained to assess our architecture for security risks. This group also runs penetration tests on BaseSpace Hub to identify potential vulnerabilities.
- Physical Access Protection.
  AWS data centers are physically protected by security staff and controlled access procedures; staff with system access undergoes background checks.
- Security Patches.
  Operating security patches are automatically applied to AWS servers, including BaseSpace Hub servers.

References

1. BaseSpace API data model: https://developer.basespace.illumina.com/docs/content/documentation/rest-api/data-model-overview
2. Advanced Encryption Standard: http://en.wikipedia.org/wiki/Advanced_Encryption_Standard
3. Amazon Web Service security and compliance Information: http://aws.amazon.com/security/
4. Amazon Web Services: Overview of Security Processes (white paper, 2011): http://d36cz9buwru1tt.cloudfront.net/pdf/AWS_Security_Whitepaper.pdf

Illumina • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

For Research Use Only. Not for use in diagnostic procedures.

 Pub. No. 970-2013-005 Current as of 16 March 2016

illumina®