illumina®

BaseSpace™ データセキュリティ

イルミナのクラウド解析と保管サービスでデータを安全に

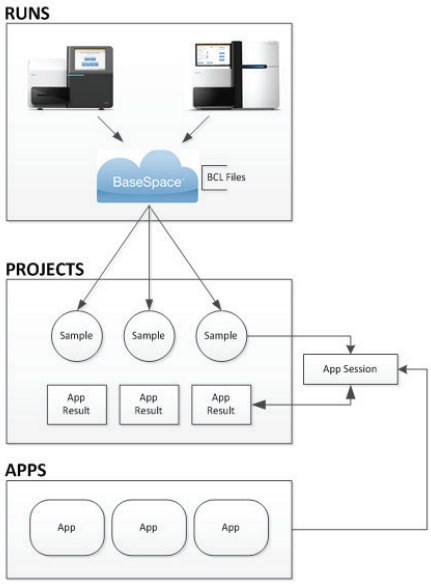
膨大なデータを産出する次世代シーケンサーは、生物医学研究を根本的に変えました。この大容量のデータが必要とするのは拡張性のある、強固で安全な解析ソリューションです。BaseSpace は、イルミナがアマゾンウェブサービス（AWS）上に構築した、ゲノミクスクラウド解析環境および保管サービスです。AWS は、民間企業のみならず、米国財務省、エネルギー省、国務省などの省庁に対しても顧客対応サービスや重要なオペレーションの機能を担っており、イルミナは、クラウドベースのインフラストラクチャーをリードするパートナーとして AWS と提携することを選びました。

クラウドベースのゲノム解析やデータ保管への移行を決定する際、主要な問題となるのがデータセキュリティです。イルミナは、包括的で数多くのテストを重ねたアマゾンのプラットフォームセキュリティへのアプローチと、イルミナ独自のセキュリティテストおよびプロシージャーとを組み合わせることで、多くの施設で提供される IT インフラストラクチャーのセキュリティレベルに適合し、さらにはそれを凌駕する、クラウドのゲノム研究ソリューションを提供します。

BaseSpace のデータモデル

シーケンサー上のフローセル処理は、ランと総称され、さまざまなファイルが作成されます。ランには、ログファイル、装置のヘルスデータ、ラン指標、ベースコール情報（*bcl files）が含まれ、これらは BaseSpace 内でデマルチプレックスされ、二次解析で使用するサンプル毎のデータを作ります [1]。

図 1：BaseSpace データモデル



BaseSpace データは、RUN、Sample、PROJECT、App Result、および App Session で構成されるデータモデルに準じて保管されます。

BaseSpace App を立ち上げてサンプルを解析すると、Base Space App は、API を通過して、BaseSpace データと通信するソフトウェアおよびルーチンを実行します。App が Base Space データへのアクセスを要求するたびに、ユーザーレベルを認証し、転送中のデータを暗号化します。App から出力されるファイルは AppResult と呼ばれるオブジェクトに格納されます。例えば、リシーケンス時の App がアライメントやさまざまなコールを実行しているとき、すべてのサンプルについて AppResult が作成されます。AppResult は App へのインプットとして利用することも可能です。App が起動するたびに、App セッションが作成され、記録を行います。最終的に、Project は、サンプルと AppResult を保管する単一のコンテナとなります（図 1）。[1]

転送中のデータのセキュリティ

ゲノムシーケンス装置とデータ解析および保管サーバー間のコミュニケーションにおいて大部分を占めるのが、データの転送です。イルミナは、いくつかのセキュリティ対策を実行し、インフ転送中のデータを確実に守ります。

装置への安全な接続

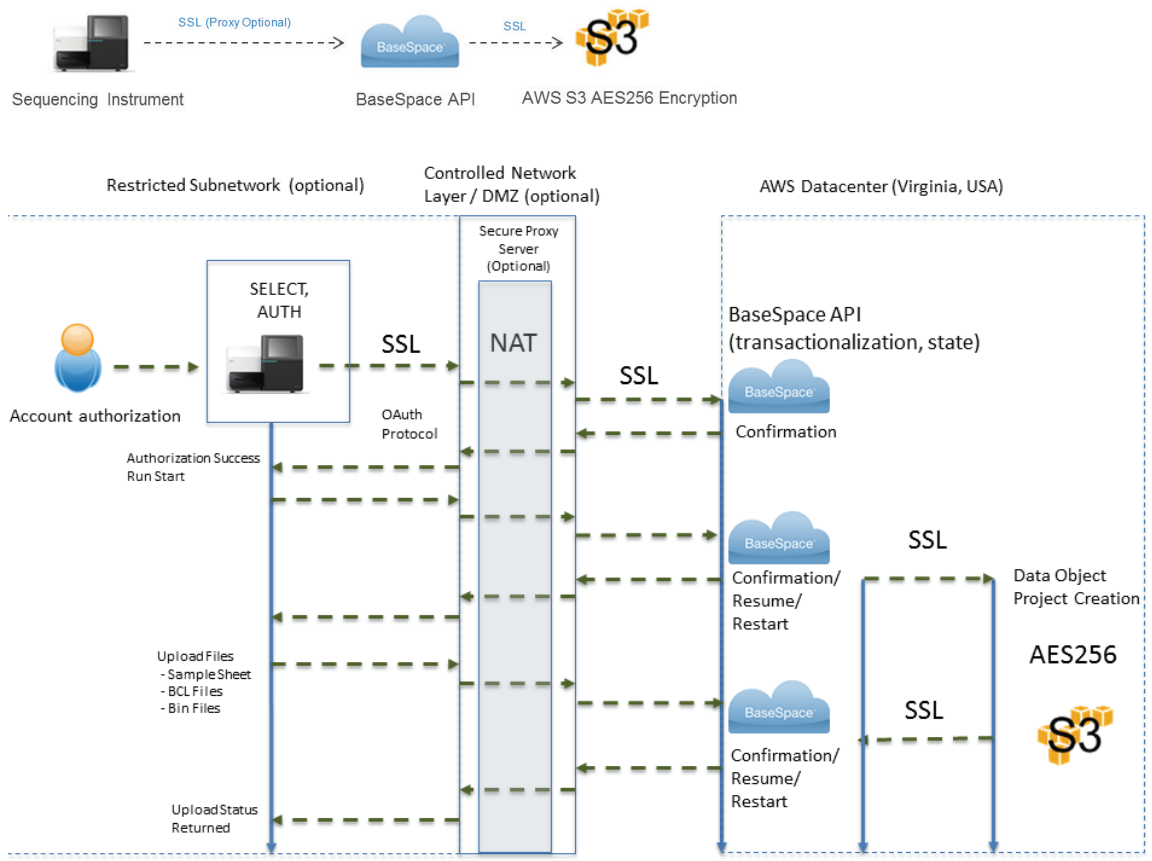
BaseSpace へのデータの送付の有無は、ランのセットアップ時に常にユーザーが決定します。BaseSpace を選択すると、ユーザーの BaseSpace/MyIllumina アカウントに対してランを認証し、トラッキングを行います。ユーザーは、データを同時にローカルにも保管することができ、必要に応じて BaseSpace に保管することもできます（図 2）。

図 2：ランのセットアップオプション（MiSeq システムソフトウェア）

図 2 上：それぞれのデータストリームセッションは、シーケンスランプロセスの一部としてユーザーが起動し、BaseSpace/MyIllumina アカウントにより認証を行います。

下：シーケンサーのソフトウェアは、いかなるホスティングサービスおよび IIS サービスの提供、および、パブリックに参照可能な IP アドレスを保持しません。

図 3：BaseSpace データの流れ – 概観図



上：シーケンサーと BaseSpace 間のコミュニケーションの概要

下：ユーザーのアカウントが認証され、データストリーミングのオプションが選択されると、BaseSpace へのデータ転送が始まります。BaseSpace へ転送されるデータは、完全に暗号化されます。BaseSpace およびいずれのクラウドサービスも、シーケンスを実行している装置に影響することはありません。

MiSeq および HiSeq システムには、シーケンスのラン実行中に作成される小さいサイズの確認用ベースコールファイルを暗号化し、送信し、検証するソフトウェアを搭載しています。この仲介役となるソフトウェアは、BaseSpace の API と通信することで、ネットワークの中断、待機、不完全な送信を自動的に検出して再び待機させることを行います。

暗号化と認証

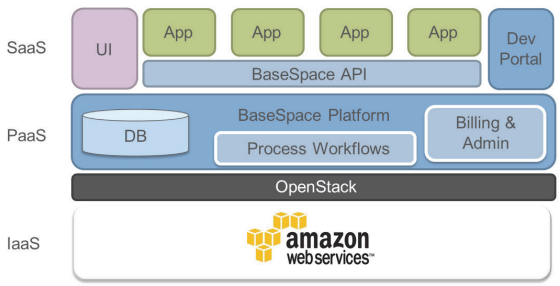
装置および BaseSpace 間の通信方式は、http ベースの API により行われます（図 3）。このコミュニケーションの特性として：

- 非常に高度なセキュリティ対策として、米政府の極秘情報の守秘にも十分安全な AES（新暗号規格–Advanced Encryption Standard）256 規格に基づいてデータを暗号化し、BaseSpace へ転送します。
- データの送受信はすべて、安全性の保障された SSL 通信で行われます。SSL（Secure Sockets Layer）は、慎重な扱いが要求される情報を、インターネットで通信を行う際に暗号化するために開発されたインターネット規格です。SSL は、新しい技術や脅威を考慮し、定期的にアップデートされています。数百万におよぶ個人や組織が日々の金融取引に関する情報の保護に SSL を利用しています。そして私たちは、アップロードおよびダウンロード時の BaseSpace データの保護に、SSL を利用しています。
- API キーシグネチャを要求するサービス方式を採用しています。本人以外へのサービスは拒否されます。
- 要求はモニターされており、悪用されると、IP および/または API キーはブラックリストに載る恐れがあります。

クラウドに長期間保管するデータのバックアップ

BaseSpace は、BaseSpace のサービスとしてのプラットフォーム（Platform-as-a-Service：データベースやワークフローなど）上に開発された、UI や App と同様に、サービスとしてのソフトウェア（Software-as-a-Service）の構成要素として特徴づけられ、アマゾンウェブサービスのサービスとしてのインフラストラクチャー（the Amazon Web Services Infrastructure-as-a-Service）上に構築されています（図 4）。このインフラストラクチャースタック内のデータは安全に保管される必要があるため、アマゾンでは、クラウド内のデータが確実に保護されるよう数多くの対策を実施しています。

図 4：BaseSpace のアーキテクチャスタック



BaseSpace アーキテクチャスタックは、アマゾンウェブサービス（AWS）のサービスとしてのインフラストラクチャー（IaaS）、BaseSpace のサービスとしてのプラットフォーム（PaaS）、およびサービスとしてのソフトウェア（SaaS）で構成されています。

AWS セキュリティ規格

アマゾン独自のセキュリティプロセスおよび規格は公開されており、閲覧が可能です。AWS 規格および認定に含まれるものとして:

- SOC 1/SSAE 16/ISAE 3402

  Service Organization Controls（SOC）1 監査は、顧客のデータを保護する AWS 管理が適切に設計されているか、個別の管理が効果的に稼働しているかを検証します。

- FISMA 中位レベル

  連邦政府の情報システムのセキュリティを強化することを目的とする、米連邦政府が付与する認定です。参考までに、NIH が所有するデータセンターは、FISMA の中位レベルと評価されています。

- PCI DSS レベル 1

  クレジットカード業界のセキュリティ基準は、電子決済のセキュリティ強化のために設定されました。AWS は最高レベルとして評価されています。

- ISO 27001.

  広く認知されたセキュリティの国際標準規格であり、セキュリティマネジメントのベストプラクティスおよび包括的なセキュリティ管理を規定しています。

- FIPS 140-2.

  連邦情報処理規格（FIPS）の公告 140-2 は、米国政府のコンピュータセキュリティ規格であり、暗号作成モジュールの要件を規定しています。

データの保管

ランおよびサンプルデータはすべて非公開のデータとして S3 に格納されます。この時点で、ファイルに公にアクセスすることは許可されなくなり、API またはウェブサイトを通して作成することが義務付けられ、ウェブサイトでのみダウンロードが可能となります。

AWS を通して、BaseSpace は、複数の施設で同期を取りながらお客様のデータを格納し、定期的にデータの完全性を確認し、自動で自己回復を実行します。お客様のデータは、不慮の損失から守られているのです。しかし、BaseSpace は無制限のバックアップシステムではありません。データの削除を実行してしまうと、データを回復させる方法はありません。また、所有権を譲渡した場合、お客様によるデータ回復は行えないまま、新たな所有者によるデータの削除が可能になります。

さらに、AWS のデータセンターはセキュリティスタッフおよびアクセスプロシージャーによる管理によって守られています。システムへアクセスするスタッフのバックグラウンドの確認を行い、ハードウェアはすべて、デフォルトであらゆる通信をブロックするように設定されたファイアウォールにより守られています。BaseSpace サーバーを含む AWS サーバーに対するセキュリティパッチは自動的に適用します。

施設のセキュリティとの比較

私たちは、BaseSpace がお客様のデータを守る非常に安全な環境であり、お客様のサーバー環境が維持する安全規格に適合すると自負しています。一部の施設で提供される IT インフラストラクチャーのセキュリティレベルを凌駕する場合もあります。

以下のプラクティスが、お客様の IT 環境で実施されているか、IT 担当者までお尋ねください：

- データの暗号化

  あらゆるセキュリティ予防策がうまく回避されデータが盗まれた場合でも、盗まれたデータが解読されないよう、他の施設ではあまり見られない例として、アップロードされるデータをすべて AES256 規格に準じて、確実に暗号化しています。

- 能動的なファイアウォールのモニタリング

  ゲノムデータの保管に使用しているコンピュータが既にインターネットへ接続されている可能性や、少なくともイントラネットを順にインターネットへ接続する可能性は非常に高いといえます。内部のネットワークを外部の攻撃から保護するよう配置されたファイアウォール装置が、通常、安全にインターネットからの切り離しを行います。BaseSpace AWS コンピュータもファイアウォールにより守られていますが、AWS は積極的に自身のファイアウォールをモニターし、脆弱性を確認します。これは、ほとんどの施設のリソースでは補えないサービスの一つです。

- 第三者による監査

  徹底したセキュリティを実証するため、コンピュータセキュリティーに関する第三者機関を設け、セキュリティリスクに対するアーキテクチャの評価を行っています。この機関は、BaseSpace への侵入テストも実施し、予測可能な脆弱性を特定します。

- データセンターへのアクセスの制限

  セキュリティスタッフが AWS データセンターに常駐し、アクセスプロシージャーを管理しています。システムにアクセスするスタッフは、バックグラウンドのチェックを受けます。

- セキュリティパッチ

  BaseSpace サーバーを含む AWS サーバーでは、セキュリティパッチが自動的に適用します。

参考資料

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

イルミナ株式会社

〒108-0014 東京都港区芝 5-36-7 三田ベルジュビル 22 階
Tel (03) 4578-2800 Fax (03) 4578-2810
www.illuminakk.co.jp

www.facebook.com/illuminakk

代理店

本製品の使用目的は研究に限定されます。 販売条件：www.illuminakk.co.jp/tc

© 2015 Illumina, Inc. All rights reserved.

Illumina, IlluminaDx, BaseSpace, BeadArray, BeadXpress, cBot, CSPro, DASL, Design Studio, GAIIx, Genetic Energy, Genome Analyzer, GenomeStudio, GoldenGate, HiScan, HiSeq, Infinium, iSelect, MiSeq, Nextera, NextSeq, NuPCR, SeqMonitor, Solexa, TruSeq, TruSight, VeraCode, the pumpkin orange color, the Genetic Energy streaming bases design は、Illumina, Inc. の商標または登録商標です。
その他の会社名や商品名は、各社の商標または登録商標です。予告なしに仕様および価格を変更する場合があります。

Pub. No. 970-2013-J005 00JUN2015

illumina®