

Introducing BlueGranite's Azure Data Factory Connector for Illumina BaseSpace

Colby Ford

In the genomics space, Illumina sequencing instruments reign supreme when it comes to next-generation sequencing (NGS). NGS produces incredible amounts of data, and this mound of data continues to grow as new advancements in sequencing technologies allow for larger output lengths, increased number of sequencing reads per run, and faster processing times. Today, many users use Illumina's online service called [BaseSpace](#) to house their sequencing data and to connect to third-party apps for downstream analyses.



Here at BlueGranite, you've heard us talk about enhancing scalability when it comes to genomics. That is, being able to instantly scale secondary and tertiary analyses and machine learning in the cloud using Azure platform services like Azure Databricks or Azure Machine Learning.

But how do you get your data over into Azure from your BaseSpace Sequence Hub? The answer: our new Azure Data Factory connector.

Azure Data Factory is a fully-managed data orchestration service that provides powerful options for moving data around, transforming it, and analyzing your data to use in other services. Currently, Azure Data Factory includes over 90 data connectors to ingest your data from services like Microsoft SQL Server, Google Big Query, AWS S3, Teradata, FTP, and more.



We often use Data Factory to copy data from source systems into Azure Data Lake. Plus, we automate these pipelines such that little to no manual intervention is needed for continuous data pulls into your data lake. We've taken this capability a step further in the genomics direction by creating a connector for Illumina BaseSpace.

How our BaseSpace Connector Makes it Easier

Like other connectors in Azure Data Factory, we can automate the ingestion and processing of data from various sources. With BaseSpace, we authenticate at the project-level, so security around sensitive data stays in place. We then copy data from BaseSpace to Azure Data Lake, organizing the data by Project, Run, Dataset, and then data type.



This automated approach for retrieving your project samples, analysis outputs, and other datasets unlocks the ability to take advantage of the Azure cloud for secondary and tertiary analyses, machine learning, and more. For example, using Data Factory, we can kick off secondary analysis pipelines in the Azure Databricks Runtime for Genomics, which allows for massively-distributed jobs for alignment, filtering/quality control, variant calling, annotation, and more.

illumina

SEQUENCE HUB

Colby Ford

HOME

RUNS


PROJECTS

ANALYSES

BIOSAMPLES

APPS

DEMO DATA



BlueGranite Azure Data Factory Connector

BlueGranite, Inc.

Bookmarked

Help

Version

1.0.0

LAUNCH APPLICATION

This application works with BlueGranite's custom Azure Data Factory pipelines to copy data and metadata from Illumina BaseSpace and your Azure tenant. This app enables the connection between these services for data movement into Azure Storage/Data Lake services.

Note: You must have the accompanying Azure Data Factory pipelines installed by BlueGranite to use this app.

READ LESS

Additional Information

Release Notes

Version 1.0.0

Initial Release -This release uses the **datasets** notion to retrieve project- and run-based data from the BaseSpace environment.

Developer

Visit Website

Terms

This app has no agreements to view.

Categories

Resequencing

Small RNA

Targeted Sequencing

De Novo Assembly

RNA-Seq

[Learn more about the Azure Databricks Runtime for Genomics.](#)

Want to Learn More?

Interested in seeing how our Azure Data Factory connector for Illumina BaseSpace can help scale your genomics practice? BlueGranite's expertise in data, analytics, and bioinformatics can help your streamline your data processing in the cloud. [Contact us](#) today with questions or to schedule a demo.

[Learn more about Genomics at BlueGranite.](#)

Illumina[®] and BaseSpace[®] are registered trademarks of Illumina, Inc. BlueGranite nor this data connector are affiliated with or endorsed by Illumina.[1] Z. D. Stephens et al., *PLOS Biology*, 2019.