Blue Prism Cloud Data Vault access details

Overview of Azure Data Lake and the Parquet Format

Data Lakes are large repositories of structured, semi-structured, unstructured, and binary data stored in its natural/raw format. Azures' Data Lake is built on Azure Blob storage (Binary Large Object) and exists as a v2 storage account.

Blue Prism Cloud utilizes this storage to maintain security, provide a preconfigured workspace, and store the archived log data with less financial and production performance cost.

Session metadata and logs are stored in a compressed Parquet file format. The Parquet file format stores data in columnar format making it ideal for querying its contained data. These files utilize a date-based hierarchical namespace storage structure which enables partition elimination for efficient querying.

Access to the Data Vault is authorized via a shared access signature (SAS) token. There are many ways to access the Data Lake with your token. The information below contains some basic examples.

Copying or syncing data from Data Lake

To download a copy of the archived parquet files, or perform a one way sync to keep any downloaded data up to date, AzCopy can be used. This data should not be copied or replicated onto the Virtual Machines within your Blue Prism Cloud platform.

If you would like the Data Vault to target your own Azure Data Lake to avoid having to copy or sync the data, please submit a service request using the SS&C | Blue Prism support portal.

AzCopy setup

AzCopy does not require installation, but where it is stored impacts how it can be run. To find the latest download, see the Download AzCopy links and download the OS specific software.

Using AzCopy you can copy or sync directly to your Azure Data Lake from the Data Vault Data Lake. For more information, see the following Microsoft articles:

- AzCopy Sync
- AzCopy Copy

PowerBI setup

PowerBI enables you to visualize and report on the data that is in the Data Lake. Within PowerBI there are several ways you can connect to the Data Lake depending on your license level. For more information, see the following Microsoft articles:

- Power Query Parquet Connector
- Power BI Direct Lake

Azure Storage Explorer setup

For simple viewing of the hierarchy within the container, Blue Prism Cloud recommends using Azure Storage Explorer.

You can find the download for your OS on Microsoft's website here: Azure Storage Explorer.

You can connect to the Data Vault Data Lake storage account with the SAS token you are provided with. Information on how to create that connection can be found in Microsoft's article: Get started with Storage Explorer.