



PitchBook Direct Data

Data Feed Documentation

Contents

1	Feed Setup.....	3
1.1	Setting up the Universe.....	3
1.1.1	PitchBook Platform Search Universe.....	3
1.1.2	CSV File Universe.....	4
1.2	Setting Up a Scope.....	4
1.3	Selecting Data Points	4
1.3.1	Example.....	4
1.4	Data Feed Currency.....	4
1.5	Pricing.....	4
1.6	Initial Development Time	5
1.7	Client-side Implementation and Ingestion	5
1.7.1	SQL Scripts - CREATE TABLE Generator.....	5
1.7.2	Customer-Relationship Management (CRM) integration.....	5
2	Data Feed File Structure	5
2.1	File Type and delimiter.....	5
2.2	File Size	6
2.3	PitchBook ID (PBiD).....	6
2.4	Last Updated.....	6
2.5	Delta Files.....	6
2.5.1	Delta File Rows.....	7
2.5.2	Delta File Sample.....	7
3	Delivery.....	8
3.1	Delivery Time	8
3.2	Delivery Frequency.....	8
4	Entity Resolution Services	8
4.1	Entity Resolution and Scoring (entity matching).....	8
4.2	Providing PitchBook a Scope File to Match	8
4.2.1	Updating Scope File on PitchBook SFTP.....	9
4.2.2	Updating Scope File in Client Environment.....	10
4.3	InternalIdentifier column	10
4.4	Entity Resolution Turnaround Time.....	10
4.5	Example.....	10

5	Accessing, your data feed, Hosting, Security.....	11
5.1	Accessing your data feed.....	11
5.2	Hosting	11
5.2.1	Connecting over SFTP to PitchBook's server	11
5.2.2	FTP site directory.....	12
5.2.3	PitchBook IP addresses and whitelisting.....	12
5.2.4	Pushing to external client server.....	12
5.3	Whitelisting IP addresses.....	13
5.4	SSH File Transfer Protocol (SFTP) - default.....	14
5.5	SSH Public/Private Key Value pairs (optional).....	14
5.6	Connecting over HTTPS to PitchBook's server.....	14
5.6.1	HTTPS: Token endpoint.....	15
5.6.2	HTTPS: Basic Authentication.....	16
6	Data Feed Changes.....	17
6.1	Changing the universe.....	17
6.2	Changing the scope of data	17
6.3	Changing data points	17

1 Feed Setup

Data feeds are a group of relational text-based flat files designed to integrate in-depth PitchBook data into client applications. Data feeds are constructed around a custom universe, scope, and collection of data points. PitchBook delivers CSV files by default.

1.1 Setting up the Universe

A universe defines the entities and financial transactions that get included into the data feed. A universe can be constructed by specifying criteria in PitchBook's advanced search, providing a CSV file of entities to PitchBook, or a combination of both methods. Throughout the year the universe will often expand as PitchBook continues to capture new market activity.

1.1.1 PitchBook Platform Search Universe

Clients with access to the PitchBook platform can build an advanced search using a variety of criteria. This advanced search can then be shared with the Data Feed Team and used as the universe of entities that flow into the data feed. If a client does not have

access to the platform they can work with their PitchBook representative to build this search.

1.1.2 CSV File Universe

A CSV file containing an exported list of entities can be sent to the Data Feed Team for inclusion in the data feed.

1.2 Setting Up a Scope

The scope outlines the underlying data on the entities contained in the universe and how various datasets relate to each other. Client's will review the data dictionary and relay their scope to PitchBook.

1.3 Selecting Data Points

The data dictionary includes data points available to data feed clients. Clients review the data dictionary then communicate data points of interest to PitchBook.

1.3.1 Example

A client requests to capture all private equity firms in the U.S. that have provided financing in the last two years (universe). For each private equity firm in this universe the client requires all active portfolio companies and all financing history for the portfolio companies (scope). The client would also specify the underlying data points they'd like to receive for the private equity firms, portfolio companies, and financing transactions.

1.4 Data Feed Currency

PitchBook's data feed service supports multiple currencies including US Dollars (USD), British Pounds (GBP), Euros (EUR), Canadian Dollars (CAD), Chinese Yuan (CNY), Japanese Yen (JPY). Please reach out to your representative to get the full list of supported data points.

1.5 Pricing

There are three inputs that affect data feed pricing: the size of the universe/scope, data points selected, and the size of the audience who will be viewing the data.

1.6 Initial Development Time

Data feeds take approximately 1-4 weeks to build depending on complexity.

1.7 Client-side Implementation and Ingestion

Clients are provided a spec that outlines the universe, scope, data points, and relationships between each file in the feed. Upon request clients can be provided with a more traditional ERD. Data feed implementation and ingestion requires a developer or development team. As a data provider PitchBook does not market professional services to implement data feeds.

1.7.1 SQL Scripts - CREATE TABLE Generator

For clients using certain relational databases PitchBook can automatically generate SQL statements to create the necessary data feed tables allowing clients to load PitchBook data more seamlessly. PitchBook currently supports the following relational database management systems:

- SQL Server
- PostgreSQL
- MySQL
- MariaDB

Clients using a system other than one listed can reach out to their PitchBook rep to see if support can be provided on an ad-hoc basis.

1.7.2 Customer-Relationship Management (CRM) integration

In a CRM enrichment use case PitchBook data will be refreshed on an agreed upon cadence. Clients will often opt to programmatically send a CSV export of their CRM data periodically to their PitchBook SFTP site. PitchBook will detect the uploaded files and add all new entities into the data feed. Upon request PitchBook will include the client's internal identifiers to map back to the correct CRM account.

2 Data Feed File Structure

2.1 File Type and delimiter

The default file type is a text-based CSV file with a pipe (,) delimiter. Upon request files can be delivered in the DAT file type with a pipe delimiter (|). The text-based files are delivered in a

time-stamped zip folder, CLIENTNAME_YYYYMMDD. Each new delivery a single zip folder is generated.

2.2 File Size

Data feed size will depend on the size of the universe, scope and data points. Full files can be approximately 1MB to over 1GB for larger feeds. Delta files are generally less than 1MB.

2.3 PitchBook ID (PBiD)

All entities in PitchBook are classified as a company, investor, limited partner, fund, person, or combination and assigned a unique identifier known as a PitchBook ID or PBiD. In addition, every financial transaction is assigned a unique ID. These identifiers are used as the primary and foreign keys to relate data feed files allowing for the creation of relational databases.

2.4 Last Updated

Each CSV file in the data feed will have a “LastUpdated” column containing a date. The “LastUpdated” date represents when any field in a record last updated from the feed’s perspective. Therefore, the “LastUpdated” date field will always be a current or a past feed delivery date and not when the data is updated in PitchBook’s database.

Example: A data feed is refreshed on a monthly cadence with an initial delivery of 01/08/2022. On 01/25/2022 there is a change to the “Revenue” field for a record. In the 02/08/2022 delivery, this record will now have “LastUpdated” value of “2/8/2022”. All other records that have not been changed since the last delivery will have a “LastUpdated” field value of “1/8/2022” and all the same values.

2.5 Delta Files

Delta files contain records that reflect updates to data values, as well as new and deleted records, relative to the most recent delivery. Clients receive a delta file for each full file in their data feed that contains one of the updates above.

By default, clients are only delivered full files, but can elect to receive delta files upon request. Delta files are often requested for larger data feeds where ingesting large files may slow load time. Once activated, the default full files will continue to be delivered in conjunction with delta files. Three columns are added to all delta files: **IsNewRow**, **IsDeletedRow**, **UpdatedColumns**

2.5.1 Delta File Rows

Field Name	Description
IsNewRow	Column Values: "Yes" or "No" "No" = The record was previously in the last delivery. "Yes" = The record is new i.e. the record was not in the previously delivery
IsDeletedRow	Column Values: "Yes" or "No" "No" = The record was NOT deleted since the last delivery. "Yes" = The record dropped off the feed since the last delivery. Example: A data feed is setup to be delivered weekly. The first delivery is 3/18. On 3/21 "Company-A" drops off the feed. The 3/25 delivery will have IsDeletedRow populated tagged as "Yes" for "Company-A".
UpdatedColumns	A comma separated list of all columns updated since the last delivery. **Example:** A data feed is setup for weekly delivery. "Company-A" on 3/18 has Chicago listed as its HQ City. On 3/21 the HQ city updates to Seattle. In the 3/25 feed the UpdatedColumns will populate with "HQ City".

2.5.2 Delta File Sample

CompanyId	Website	TotalRaised	HQLocation	LastUpdated	IsNew	IsDeletedRow	UpdatedColumn
10367-92	pitchbook.com	14.95	Seattle	9/1/2023	Yes	No	
12362-68	morningstar.com	231.3	Chicago	8/7/2023	No	No	Website
10099-45	orrick.com		San Diego	8/5/2023	No	No	

PitchBook: *This is a new record. This record was not in the last delivery*

Orrick: *This record had one column with an updated value since the last delivery: Website*

3 Delivery

3.1 Delivery Time

Upon request clients can schedule a specific delivery time. If a delivery time is not requested PitchBook will select one. Data feeds are delivered on Central Standard Time.

3.2 Delivery Frequency

Delivery frequency is determined by client needs. Available cadences include daily, weekly, monthly, bi-monthly, semi-annual, and annual. Feeds cannot be delivered more than once per day.

4 Entity Resolution Services

4.1 Entity Resolution and Scoring (entity matching)

PitchBook's entity resolution service attempts to match a list of entities provided by the client to their respective record in PitchBook. This service takes various fields (described below) and uses PitchBook's scoring model to match to the correct entity. Each entity receives a score based on the number of fields matched; if the score meets a certain threshold, it will be matched to the PitchBook ID that has the highest score assigned.

PitchBook can match on the following entity types:

- company
- investor
- service provider
- limited partner

4.2 Providing PitchBook a Scope File to Match

Clients should provide PitchBook with a "scope file" - a comma delimited CSV UTF-8 file containing the following fields, ordered as listed in the table below:

Scope file format

- **FileName:** ClientName_Scope.csv {adjust ClientName}
- **File Format:** .csv
- **Delimiter:** CSV
- **Encoding:** UTF-8
- **Entity Types:** Companies, investors, limited partners, service providers, and people (people can't be matched, but they can be included in the scope file if the PbID is available for each person included).

Column #	Column Name	Description
1	PbID	Unique PitchBook identifier for entity (if known)
2	EntityName	Name of the entity you'd like to match
4	Website	Website or domain of the entity
5	HQAddress	HQ Address line 1
6	HQCity	HQ City
7	HQState	HQ State/Province
8	HQZip	HQ Zip/Postal code
9	HQCountry	HQ Country
10	InternalIdentifier	Any internal identifier you want to keep associated with this entity

In the scope file, if any of the above columns are not available, **please still include the column but leave the values blank but still include the column as a header.**

4.2.1 Updating Scope File on PitchBook SFTP

1. Login to ftp.pitchbook.com using your provided credentials.
2. Navigate to the file directory where PitchBook hosts the scope file: /active/scope.
3. Download the current scope file {e.g. ABCBank_Scope.csv}.
4. Open the file and append any new records to the end of the scope file.
5. Upload the scope file by overwriting the current scope file.

Upon request, PitchBook can enable new feed to be automatically generated when a new scope file is uploaded.

Curl command to upload file via HTTPS

Clients who access their data feed over HTTPS can use the below Curl command to upload their scope file.

```
curl -T ClientName_Scope.csv -u username:password
https://df.pitchbook.com/upload/active/scope/
```

4.2.2 Updating Scope File in Client Environment

1. Create a path on your server where the scope file will be uploaded.
2. Upload the scope file to the location you created in #1.
3. Provide your PitchBook representative with the following:
 - a. The path you created in #1
 - b. File name of the scope file
 - c. Read credentials to access the scope file.
4. To add new records into the scope file: append new records to the end of the scope file and overwrite the current file.

4.3 InternalIdentifier column

In the scope file clients may opt to include a column for their internal identifier (e.g Salesforce ID). Clients do this as a way of mapping back PitchBook data delivered in a data feed to the already existing records in their database. PitchBook can add a column in the data feed for the client's internal identifier; any matched entities that get added into the data feed will also include their respective internal identifier in this column.

4.4 Entity Resolution Turnaround Time

Successfully matched entities that meet the scoring threshold will be inserted into the data feed on the next or following delivery depending on the data feed's scheduled cutoff.

4.5 Example

A PitchBook client tracks 15,000 entities in its ever-growing internal platform. For each entity tracked in this platform the client is recording its legal name, HQ city, and HQ state. Also, every entity is assigned a unique 18-digit numeric identifier. Without having this internal identifier, the client may have difficulty mapping the data back to its internal platform. The client uploads an initial scope file of the 15,000 companies in CSV UTF-8 to the PitchBook FTP. PitchBook automatically detects the scope file and begins matching.

The first record has a name of "Company XYZ Inc", a city of "Redmond" and a state of "Washington". PitchBook tracks the entity using its formal name of "Company XYZ" but also has its legal variation as "Company XYZ Inc". The PitchBook scoring model would log an exact legal name match and a partial formal match (PitchBook.com may omit "Inc" from the formal name). PitchBook has the HQ city listed as "Bellevue" whereas the client has inadvertently tagged it to "Redmond", also a city in Washington. This field is not a match, however, both the client and PitchBook have the same state, "Washington". The scoring model sums up the total score for this record and in this case, it meets the scoring threshold. All

the matched entities get added to the next scheduled delivery that meets the cutoff time requirements. The 18-digit internal identifier is populated as well so the client can more seamlessly ingest the underlying PitchBook data.

5 Accessing, your data feed, Hosting, Security

5.1 Accessing your data feed

Data feeds can be easily accessed and downloaded with a trusted FTP client like FileZilla or WinSCP. Clients will automate the ingestion of their data feed with automated schedulers powered by custom batch scripts written to suit their use case. Please use the credentials described in section 5.2.1 along with the provide password to access your data feed. More information on the delivery is described above in sections 3.1-3.3. Clients should not use a web browser (e.g. Google Chrome, IE, etc) to access their data feed.

5.2 Hosting

Data feeds are hosted on PitchBook servers and delivered over SSH File Transfer Protocol (SFTP) by default. Clients may also opt for HTTPS delivery.

Data feeds can be delivered to a client's servers upon request e.g., Amazon S3, Azure, GCS, Snowflake.

5.2.1 Connecting over SFTP to PitchBook's server

To establish a connection to your PitchBook server, download, and ingest your feed please us the below credentials:

Production SFTP site

Hostname: ftp.pitchbook.com

Username: Custom, provided by your PitchBook rep

Password: Custom, provided by your PitchBook rep

Port: 22 (port 22 is required)

Sandbox SFTP site

To establish a test connection to the PitchBook FTP, download, and ingest a sample data feed please use the data feed sandbox credentials:

Hostname: ftp.pitchbook.com
Username: datafeedsandbox
Password: 7CWTetjs2dxFOIAHPY6b
Port: 22 (port 22 is required)

5.2.2 FTP site directory

- /active: The five most recent data feed deliveries.
 - /active/scope (if applicable – see section 4.2): This is the directory where the scope file should be pushed for entity resolution.
- /history: all deliveries beyond the most recent five deliveries.
- /matched_results (if applicable –see section 4.2): This is the directory where the matching results from the scope file are stored.

Filename	Filesize	Filetype
..		
matched_results		File folder
history		File folder
active		File folder

5.2.3 PitchBook IP addresses and whitelisting

Connecting to PitchBook's server: 35.247.86.240 It is recommended that clients use the hostname, ftp.pitchbook.com, to connect in lieu of the IP address as the IP address is subject to change.

Receiving Data from PitchBook: 35.247.15.51 If PitchBook is pushing the data feed to the client's server, this IP must be whitelisted if whitelisting is required on the client side.

5.2.4 Pushing to external client server

Upon request PitchBook pushes feeds to external locations. In the event you'd like PitchBook to push data feed files to your preferred location please provide PitchBook with all the necessary delivery credentials and whitelist the PitchBook IP if required.

Please find the list of delivery locations below:

External Host Type	Credentials PitchBook needs from Client
AWS S3	<ul style="list-style-type: none"> • Bucket Name • Secret • Key • Region • Path
Azure Blob	<ul style="list-style-type: none"> • Storage URL • Container Path • SAS Token
Google Cloud	<ul style="list-style-type: none"> • Bucket Name • Key • Path
SFTP	<ul style="list-style-type: none"> • Host • Port • Username • Password
Snowflake	<ul style="list-style-type: none"> • Cloud Platform • Region • Account Name • Locator ID

5.3 Whitelisting IP addresses

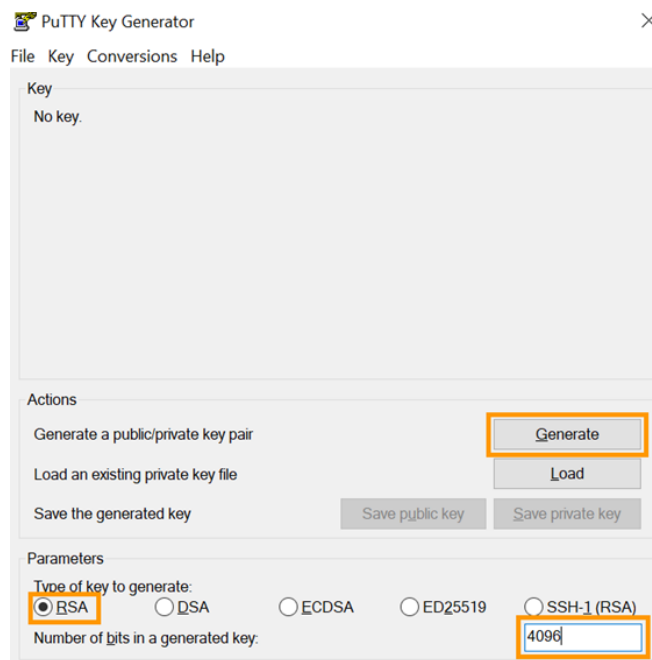
PitchBook does not require IP whitelisting to access its server. Upon request PitchBook can whitelist static public facing IP addresses or certain ranges of IP addresses depending on size. If a client is unable to provide an IP and would like enhanced security an SSH public/private key or HTTPS can be used. If you are receiving data from PitchBook please ensure you whitelist **35.247.15.51**

5.4 SSH File Transfer Protocol (SFTP) - default

By default, data feeds are delivered through SSH file transfer protocol (SFTP).

5.5 SSH Public/Private Key Value pairs (optional)

For clients that do not provide IP addresses (optional) or prefer to provide a public key, SSH public/private key values pairs may be used. Popular key generators like PuTTYgen can be used to generate SSH keys. To generate a key download PuTTYgen and select RSA as the key type, enter 4096 as the number of bits and click 'Generate'. Save both your private and public key in a safe place and provide PitchBook with the public key.



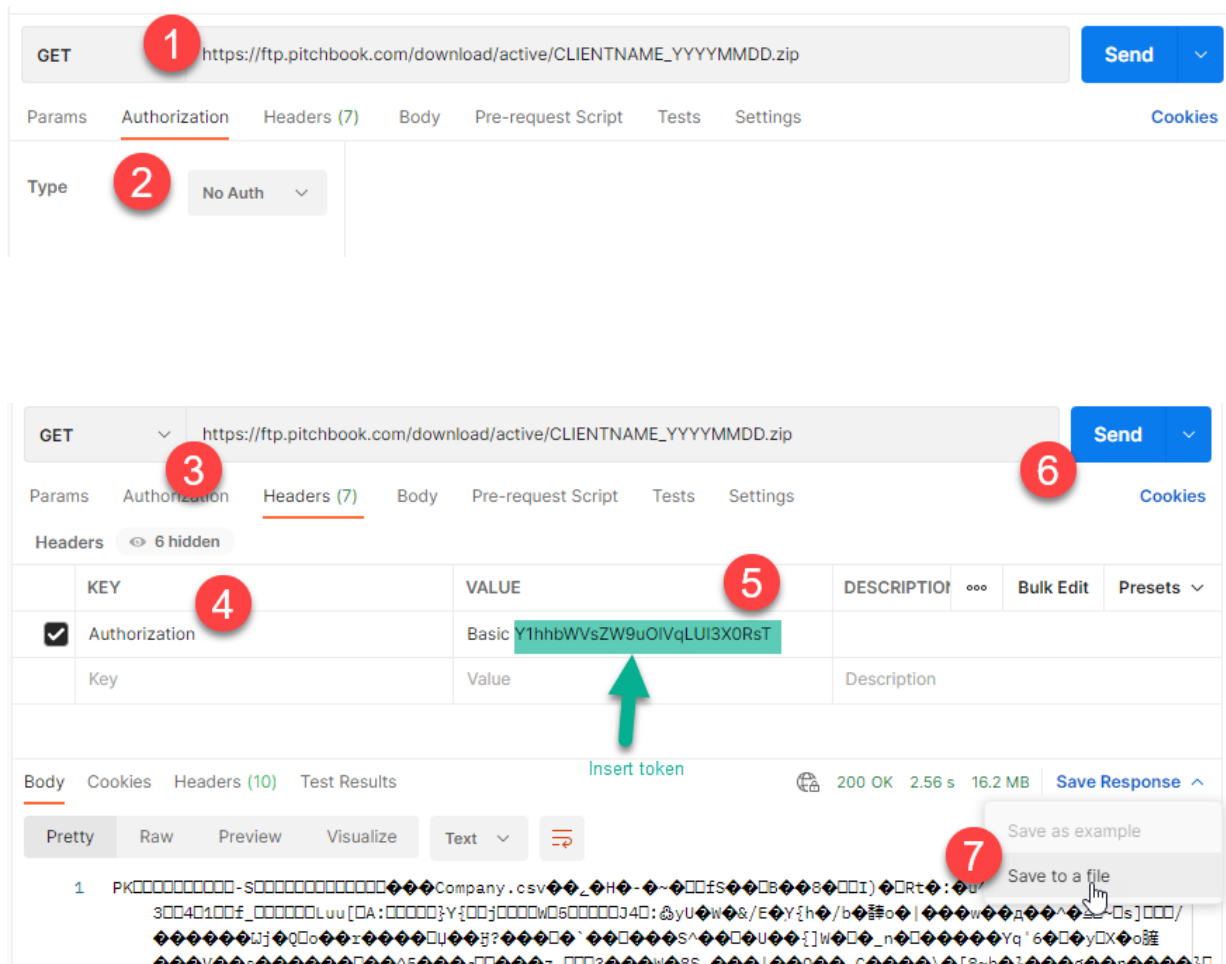
5.6 Connecting over HTTPS to PitchBook's server

PitchBook can support both token endpoint and basic authentication. For HTTPS PitchBook will provide either a token (token endpoint) or username and password (basic authentication). The full URL to access the data feed for both token endpoint and basic authentication is: <https://ftp.pitchbook.com/download/active/filename>. IP whitelisting is not required for HTTPS. For larger files clients may run into issues using free API clients like Postman and WinSCP. Files can be downloaded using curl commands or a web browser.

5.6.1 HTTPS: Token endpoint

Clients opting to use HTTPS token endpoint will be provided with an API token from PitchBook. If you're using Postman API client set the authorization to "no auth" and enter the API token provided as a header.

Download using Postman API client (token authentication)



Download data feed using Curl (token authentication)

Curl Structure: `curl -H "Authorization: Basic paste token here"`

`https://ftp.pitchbook.com/download/active/CLIENTNAME_YYYYMMDD.zip -o CLIENTNAME_YYYYMMDD.zip`

```
Microsoft Windows [Version 10.0.19042.1052]
(c) Microsoft Corporation. All rights reserved.

C:\Users\david.gardner.CORP>
C:\Users\david.gardner.CORP>curl -H "Authorization: Basic Yeejbefjb488sgfsfjsg3s" https://ftp.pitchbook.com/download/active/ABCBANK_20210913.zip -o ABCBANK_20210913.zip
```

5.6.2 HTTPS: Basic Authentication

Clients using HTTPS basic authentication will be provided a username and password in lieu of a token. Clients using this method should select “Basic Auth” as the authorization and enter their username and password provided by PitchBook. Examples below.

Download data feed using Postman API client (basic authentication)

The screenshot shows the Postman API client interface. At the top, the HTTP method is set to 'GET' (1) and the URL is 'https://ftp.pitchbook.com/download/active/CLIENTNAME_YYYYMMDD.zip' (4). Below the URL bar, the 'Authorization' tab is selected (2), and the 'Type' is set to 'Basic Auth'. A warning message states: 'Heads up! These parameters hold sensitive data. To keep this data secure while working in a collaborative environment, we recommend using variables. [Learn more about variables](#)'. The 'Username' field is filled with 'clientusername' (3), and the 'Password' field is masked with dots. There is a 'Show Password' checkbox below the password field. The 'Send' button is visible on the right.

Download data feed using Curl (basic authentication)

Curl structure: `curl -u username:password`

`https://ftp.pitchbook.com/download/active/CLIENTNAME_YYYYMMDD.zip -o CLIENTNAME_YYYYMMDD.zip`

Above, replace *username:password* with your unique credentials.


```
Microsoft Windows [Version 10.0.19042.1052]
(c) Microsoft Corporation. All rights reserved.

C:\Users\david.gardner.CORP>curl -u abcbank:ffsmw04_rst https://ftp.pitchbook.com/download/active/ABCBANK_20210913.zip -o ABCBANK_20210913.zip
```

Download in web browser (basic authentication)

Copy and paste full URL (including file name) into any modern browser and enter the username and password.

Sign in

<https://ftp.pitchbook.com>

Username

Password

6 Data Feed Changes

The data feed scope, universe, and data points can be changed throughout the year. Depending on the change there could be an impact on the data feed pricing.

6.1 Changing the universe

Generally, clients using an advanced search are requested to keep a relatively fixed universe in terms of criteria. Throughout the year clients can request to change their universe criteria should requirements change. This will require additional development time by PitchBook, and the client will be provided with a release date. Clients using a scope file can add new entities throughout the year as long as they are in compliance with their contract.

6.2 Changing the scope of data

Generally, clients are requested to keep a relatively fixed scope. Throughout the year clients can request to change their scope should requirements change. This will require additional development time by PitchBook and the client will be provided with a release date.

6.3 Changing data points

Throughout the year clients can request to add/remove data points should requirements change. This will require additional development time by PitchBook, and the client will be provided with a release date.