



BPD Data Lake Data Consumption API

#### Version: 1.8

#### Update Date: 02/23/2023

#### Created on: 07/14/2022

You are reminded that the information contained in this document is confidential, proprietary to the Blue Cross Blue Shield Association, and intended only for use by Blue Cross and Blue Shield Plans. Unauthorized use, disclosure or copying of this information or any part thereof is prohibited.

Revision History

|  |  |  |  |
| --- | --- | --- | --- |
| Date | Version | Description | Author |
| 07/21/2022 | 1.1 | Initial Draft with basic section contents | Ankur Chourasiya |
| 08/22/2022 | 1.2 | Updates details:  Sec 2: API Workflow  Sec 3: API URL  Sec 4: Curl Command  Sec 5.1: API Key | Ankur Chourasiya |
| 08/24/2022 | 1.3 | Update Details:  Sec 3: Api details updated, added new input to API , data\_type  Sec 4: Sample input command and output | Ankur Chourasiya |
| 10/4/2022 | 1.4 | Update details:  Update the API URL in all sections  Updated the API Key required to access the API | Ankur Chourasiya |
| 11/30/2022 | 1.5 | Update details:  Updated Sec 3 and Sec 4 for the text file output format from unzipped to zipped format | Ankur Chourasiya |
| 02/01/2023 | 1.6 | Update details:  Updating the document to add the details for Delta file download using the API  Updating the Full file download details to not add data\_date key while doing API call | Ankur Chourasiya |
| 02/17/2023 | 1.7 | Update details:  Updating the curl command to call the API | Ankur Chourasiya |
| 02/23/2023 | 1.8 | Update details:  Added Cross account access pattern details for API Gateway  Call API from Glue Job | Ankur Chourasiya |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Table of Contents**

Contents

[1 Introduction 3](#_Toc128064869)

[2 API Architecture 3](#_Toc128064870)

[3 API Details 3](#_Toc128064871)

[4 How to consume data 4](#_Toc128064872)

[Step 1 4](#_Toc128064873)

[Step 2 6](#_Toc128064874)

[5 Security Baseline 7](#_Toc128064875)

[5.1 Authentication 7](#_Toc128064876)

[5.2 Availability 7](#_Toc128064877)

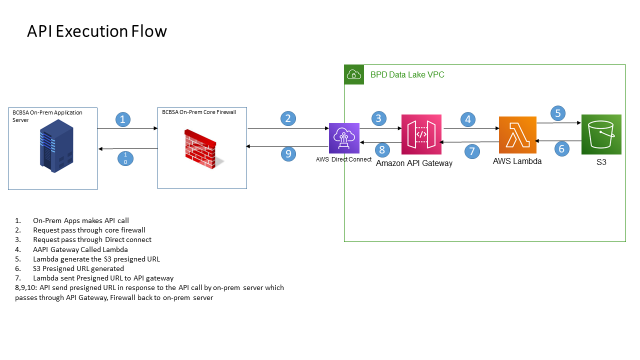
[6 Cross AWS account access Setup 7](#_Toc128064878)

[7 Call API from Glue Job 8](#_Toc128064879)

# Introduction

This document will provide information about using the API for consuming the data from BCBSA BPD Data Lake.

# API Architecture



Please see this space for updated architecture with OAuth2 Okta integration in future

# API Details

**URL:** https://vpce-0812496c903bbaf60-8dbu62mr.execute-api.us-east-1.vpce.amazonaws.com/ProdStage/presigned\_url

**Inputs Required:**

**X-API-Key** :[Mandatory] <<API Key>> (see sec 3.1 for API Key for calling this API)

Sample API Key : KQcIak4HJO5bitgf9X3Hf4fhlMnK6Pd73VIM97xl

**x-apigw-api-id** : rveg8q00bk

**application\_name**: :[Mandatory] <<Caller Application Name>>

Supported Applications: NCCT, PRP, PEX, NDW, PDA, NPD, NRS

**file\_type:** :[Mandatory] <<Type of file required from Data Lake>>

Supported Types: Txt, Parquet

**data\_mode**: :[Mandatory] <<Full or delta >>

**data\_date** : :[Optional] Data date in MMDDYYYY format for which the Delta file to be downloaded. Error will be thrown out if Delta file is not available for given date. If data\_date not provided then presigned url of latest delta file will be provided.

For full extract data\_date is not expected, if provided data\_date will be ignored and latest data file presigned url will be shared.

**Sample API calling example below using curl command:**

**Full File presigned URL API Call:**

curl -X POST -H "Accept: application/json" -H "Content-Type: application/json" --header "X-API-Key : <<API Key>>" –header “x-apigw-api-id : <<API ID>>”--header "application\_name : <<Caller Application Name>>" --header "file\_type : <<File Type (txt/parq)>>" --header "data\_mode: << (full/delta)>>" –header “data\_date : <<Data date for delta file presigned url, for full file this option is not needed>> ‘<<API URL>>'

**API Response:**

API response will be a JSON, with following keys:

{

    "application\_name": "<<Caller application name>>",

    "data\_type": "<<data type PII or NON PII >>",

    "data\_mode": "<<Data Mode Full or Delta>>",

    "file\_type": "<<File Type Txt or Parquet>>",if file type txt is chosen, zipped file will be downloaded

    "data\_date": "<<Date of the data>>",

    "url": "<<presigned url of S3 file>>”

# How to consume data

Below are the series of steps to be followed to consume the data using API call.

## Step 1

Make the PAI call with required inputs e.g. API Key, application name and File type

**Full File download example below using curl command:**

Application Name: NRS

File Type : Txt

API Key: see sec 3.1 for API Key

Data mode: full

**curl -X POST -H "Accept: application/json" -H "Content-Type: application/json" --header "X-API-Key : yXNpZc3Rkdlwd1BjJ9ir7DosgTGenP211E2Pg9z7" --header "application\_name : BHI" --header "file\_type : txt" --header "data\_mode : full" --header "x-apigw-api-id : rveg8q00bk" 'https://vpce-06cd7f47f725aa996-bknh28z6.execute-api.us-east-1.vpce.amazonaws.com/QaPeStage/presigned\_url'**

**API Response:**

Sample output of above curl command:

{"application\_name": "NRS", "data\_type": "pii", "data\_mode": "full", "file\_type": "txt", "data\_date": "11182022", "url": "https://bpd-datalake-qape-zone.s3.amazonaws.com/delivery-zone/pii/txt/11182022/part-00000-827e290d-f548-4525-90ab-f687b0e7bca5-c000.csv.gz?X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=ASIARNE2UKNAV2X6KO3W%2F20221130%2Fus-east-1%2Fs3%2Faws4\_request&X-Amz-Date=20221130T130956Z&X-Amz-Expires=3600&X-Amz-SignedHeaders=host&X-Amz-Security-Token=IQoJb3JpZ2luX2VjEF4aCXVzLWVhc3QtMSJGMEQCIGsAznq0PUAL91itHKm2PaUX2eobFDUUZ2rSMZpqbyCpAiBUobR0sy%2B609%2BbS8uDpHIXb%2FFKS76Tra9AAWXKOt2EVCqZAwh2EAIaDDA5Njk2MTA1NzYwMSIManbfDY%2Bpl9UKetgJKvYC02tVl8n2XcmfBGKjQXvGhkItU7zwNv4PzSsNlxRZQfEiC70pxdHoXIw%2Fz%2Fk2%2FzGF%2BQ%2FH6F8S%2FuEzjdmEEKmqj%2Fr3iSsKO1sDGixnRIApLKB8Jnx9zu%2FmJzJNaq49OvlfV2uOK%2BBz31FZheHlMJmeT0PrFuvRGrcAQRSiSwlADRLZiKxqjzSMqZddCsUrahdCsBUt2oQ4SbbbWZw0dtFpXNkjfF1ak8bB6ESQopQg8psS%2BAStnxADOEweRIa%2B%2FLM3UoJLUPe4TveJPIPyt3xfbfKDW4Yp2mNE6Nen1eqve2f763dYr5ozaUNOcPu4WmIC7O0mkZRgRpXjKWffJCMVdPH9px2I6xNRrCV5MrrCBkw1qHHQINifIPYZVzIKXi4jzPvfE6nT5oQWCjNgSllTMG81yfZW2ebtUYVJItkN3Gib8LOjFpiNx6B4aj0eIC3D9KFxC1VgHVjHbvgw81eSF81OhtiNw8x7%2FADK6nEhuwMjkpQ0DZcwoaydnAY6ngFavLv4eXEjzRjq%2Fu097aqr7paAdZ69dujcFAMfnfYr7f0fW9Y7ArdyjjmqTospc5xgqod%2BWaxJOY22SM%2FCui%2B8Aqf9yYseP8%2B%2BSWcQtkXArIqMKAsyq%2FHME2IGBrINQUFTSF60%2BCrsPKq8yhHaKF3oAQfjOGlejplX0t9ebqEMu1ropMguu4ADQkzbt0riw0eKt5DmLLCl8PQVPmFeXQ%3D%3D&X-Amz-Signature=03609025176defbd5a741d45b83283f91c26e2a65606e5d39134301f78005da8"}

**Note: if file type is chosen to txt then zipped file will be downloaded. Application team will have to unzip the file as per data consumption requirements.**

**Delta File download example below using curl command:**

Application Name: BHI

File Type : Txt

API Key: see sec 3.1 for API Key

Data mode: Delta

Data\_date: Date for which the delta file is needed

**curl -X POST -H "Accept: application/json" -H "Content-Type: application/json" --header "X-API-Key : yXNpZc3Rkdlwd1BjJ9ir7DosgTGenP211E2Pg9z7" --header "application\_name : BHI" --header "file\_type : txt" --header "data\_mode : delta" --header "data\_date : 01092023" --header "x-apigw-api-id : rveg8q00bk" 'https://vpce-06cd7f47f725aa996-bknh28z6.execute-api.us-east-1.vpce.amazonaws.com/QaPeStage/presigned\_url'**

**API Response:**

Sample output of above curl command:

{"application\_name": "BHI", "data\_type": "pii", "data\_mode": "delta", "file\_type": "txt", "data\_date": "12022022", "url": "https://bpd-datalake-qape-zone.s3.amazonaws.com/delivery-zone/delta\_extract/pii/txt/12022022/bpd-datalake-pii-txt-extract-12022022.csv.gz?X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=ASIARNE2UKNAUYFELYR2%2F20230102%2Fus-east-1%2Fs3%2Faws4\_request&X-Amz-Date=20230102T112530Z&X-Amz-Expires=3600&X-Amz-SignedHeaders=host&X-Amz-Security-Token=IQoJb3JpZ2luX2VjEHQaCXVzLWVhc3QtMSJIMEYCIQDcXvc34TfBCgZCgo51%2Fwab6PG2uksL1P5iVP2IBE78rQIhAL%2FKynT4fNwXM497YnLCan7IkF40rYB8WwpCXilJLmkOKqIDCL3%2F%2F%2F%2F%2F%2F%2F%2F%2F%2FwEQAhoMMDk2OTYxMDU3NjAxIgzG%2F3vpg6CzzlkHXQ0q9gIps0KPznb%2FO5v70geGf6WCqBF%2BGZggUJ7ipSRGK6%2FgxdSJltJB8gHEHN%2FmKiItg%2BNQGxj4I5ybXCk%2BTcsrH%2Fu36uImFkMX4yOGbLVGkUY6V5Vfve1FOCOArKZGymbNIphc63wDvaPqcIxYM8BmlAgCs0aqfKKnXSDO2j%2BKm1oUV6zAyLGsW2yhmoYS6qNZ6pMpWd1ifVejKjrgNrZBtcBCFK%2BxLUgaqgoTCa5BHuABRYuQBIt5u5xQx1bmsNnnWORcft2ymdyfBjc%2BRy%2BC4zVkCECGPQpXO2ccxQfx%2BlafVa3zNJkh89iEoSecpIFD2vmpSTcdJn9qmsIqHB5J3IL2zS7xySgIDnwQBgjnzgLfXataaUDzlYovPZD3Aqi1DUbspu%2B26EgaZX6EbAO97d8%2BGPk%2FMPFsL48HMcymR6t8p1X52PjQ8nFBf6lIG2jQaLpRPEvdPhLntMvAhF8INKsJkC6u8nMy5QmcllhHxgSHbgu5Y1kOdjCo%2FsqdBjqcAWgR06O%2BfAdj1YD85DxWAACpjlt0oHvUHUQjq2YCmLKBZPvyNecPvYvUfNJWCzsCbM5zv7Wp8NWmHkCM4RjV1l53U9JENPl%2FWhGQWF%2BCujfbA1DvfQUO1KJ%2FRAKedzLdi%2BaMs4CJ3ATEZgyXZjTzWHxjkgXj10x4HVte16g7dri1a9p5WyZPVCe7hCQdx8aXZ9%2FFlYtPOwL1fxburQ%3D%3D&X-Amz-Signature=fb48d53ecc5ae5738efd8c6ea4451d7c5d536d4013bf5386457505f47f3cedad"}

**Note: if file type is chosen to txt then zipped file will be downloaded. Application team will have to unzip the file as per data consumption requirements.**

## Step 2

Use the URL in API response to download the data file from BPD Data Lake.

Example below using the wget command with output filename option:

**wget ‘**https://bpd-datalake-qape-zone.s3.amazonaws.com/delivery-zone/pii/txt/11182022/part-00000-827e290d-f548-4525-90ab-f687b0e7bca5-c000.csv.gz?X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=ASIARNE2UKNAV2X6KO3W%2F20221130%2Fus-east-1%2Fs3%2Faws4\_request&X-Amz-Date=20221130T130956Z&X-Amz-Expires=3600&X-Amz-SignedHeaders=host&X-Amz-Security-Token=IQoJb3JpZ2luX2VjEF4aCXVzLWVhc3QtMSJGMEQCIGsAznq0PUAL91itHKm2PaUX2eobFDUUZ2rSMZpqbyCpAiBUobR0sy%2B609%2BbS8uDpHIXb%2FFKS76Tra9AAWXKOt2EVCqZAwh2EAIaDDA5Njk2MTA1NzYwMSIManbfDY%2Bpl9UKetgJKvYC02tVl8n2XcmfBGKjQXvGhkItU7zwNv4PzSsNlxRZQfEiC70pxdHoXIw%2Fz%2Fk2%2FzGF%2BQ%2FH6F8S%2FuEzjdmEEKmqj%2Fr3iSsKO1sDGixnRIApLKB8Jnx9zu%2FmJzJNaq49OvlfV2uOK%2BBz31FZheHlMJmeT0PrFuvRGrcAQRSiSwlADRLZiKxqjzSMqZddCsUrahdCsBUt2oQ4SbbbWZw0dtFpXNkjfF1ak8bB6ESQopQg8psS%2BAStnxADOEweRIa%2B%2FLM3UoJLUPe4TveJPIPyt3xfbfKDW4Yp2mNE6Nen1eqve2f763dYr5ozaUNOcPu4WmIC7O0mkZRgRpXjKWffJCMVdPH9px2I6xNRrCV5MrrCBkw1qHHQINifIPYZVzIKXi4jzPvfE6nT5oQWCjNgSllTMG81yfZW2ebtUYVJItkN3Gib8LOjFpiNx6B4aj0eIC3D9KFxC1VgHVjHbvgw81eSF81OhtiNw8x7%2FADK6nEhuwMjkpQ0DZcwoaydnAY6ngFavLv4eXEjzRjq%2Fu097aqr7paAdZ69dujcFAMfnfYr7f0fW9Y7ArdyjjmqTospc5xgqod%2BWaxJOY22SM%2FCui%2B8Aqf9yYseP8%2B%2BSWcQtkXArIqMKAsyq%2FHME2IGBrINQUFTSF60%2BCrsPKq8yhHaKF3oAQfjOGlejplX0t9ebqEMu1ropMguu4ADQkzbt0riw0eKt5DmLLCl8PQVPmFeXQ%3D%3D&X-Amz-Signature=03609025176defbd5a741d45b83283f91c26e2a65606e5d39134301f78005da8**' -O ndw\_file.txt.gz**

**Sample output of above command:**

Resolving bpd-datalake-dev-output-zone.s3.amazonaws.com (bpd-datalake-dev-output-zone.s3.amazonaws.com)... 54.231.163.25

Connecting to bpd-datalake-dev-output-zone.s3.amazonaws.com (bpd-datalake-dev-output-zone.s3.amazonaws.com)|54.231.163.25|:443... connected.

HTTP request sent, awaiting response... 200 OK

Length: 2663 (2.6K) [text/plain]

Saving to: ‘**ndw\_file.txt.gz**’

100%[========================================================================================================================================================================>] 2,663 --.-K/s in 0s

2022-07-14 14:30:08 (118 MB/s) - ‘nrs\_file.txt’ saved [2663/2663]

Below is the screen capture of execution flow in cloud environment:



**Note: This is sample to showcase the functionality, respective internal application team can use the preferred way to consume the data.**

**Current Demo support application name = NRS and Txt file can be downloaded with full data mode using API**

# Security Baseline

## 5.1 Authentication

Currently the API requires and API Key to provide to authenticate the caller request.

**API Key to call this API is below:**

KQcIak4HJO5bitgf9X3Hf4fhlMnK6Pd73VIM97xl

## 5.2 Availability

BPD Data Lake data consumption API is the private API bind to VPC end point of BCBSA AWS VPC network. It will be accessible only within the BCBSA network.

# Cross AWS account access Setup

To setup the cross AWS account connectivity between API Gateway and Target account below steps needs to be followed.

Vocabulary:

Source Account: Account in which Presigned URL API is deployed

Target Account: Any AWS account which require access to Presigned URL API

Detailed Steps:

1. Create a API Gateway VPC endpoint in Target Account
2. Whitelist the IP of the EC2 instance or the CIDR range of subnet in which Glue job or lambda are running from which the API to be called in VPC endpoint security group created in step 1.
3. Share the VPC endpoint ID created in step 1 with Source Account Team.
4. Source account team to add the Target account VPC endpoint ID in API Gateway resource policy to allow access to invoke API for traffic coming from Target account VPC endpoint.
5. Once steps 1 to 4 are completed test the API connectivity using the curl command provided above in section 4.

# Call API from Glue Job

To call the Presigned URL API from Glue Job either from source account or target account below steps can be followed:

Vocabulary:

Source Account: Account in which Presigned URL API is deployed

Target Account: Any AWS account which require access to Presigned URL API

**Glue job running within the source account:**

1. Create a Network type connection in Glue.
2. Attach the connection in Glue Job to run it within VPC
3. Whitelist the CIDR range of subnet of Network connection in API Gateway VPC endpoint security group to allow HTTPS access
4. Below is the sample python command to call API from Glue job:

import os

import requests

import sys

import subprocess

url = "https://vpce-06cd7f47f725aa996-bknh28z6.execute-api.us-east-1.vpce.amazonaws.com/QaPeStage/presigned\_url"

headers = {

"X-API-Key" : "T9fhUTedLW4bdBCiIjqLx92zB60FbZG91eKzv195",

"Accept" : "application/json",

"Content-Type": "application/json",

"application\_name" : "NRS",

"file\_type" : "txt" ,

"data\_mode" : "full",

"x-apigw-api-id" : "rveg8q00bk"

}

response = requests.post(url, headers=headers,)

data = response.json()

print(data)

**Glue job running in target account:**

1. Complete the step written in Section 6 from Cross AWS account connectivity
2. Once connectivity is setup, follow the steps 1 to 4 from above section.