[SFTP] File not picked up in SFTP or SftpPathNotFound error

Last updated by | Supreeth Vasudevamurthy | Mar 14, 2022 at 3:06 PM PDT

Contents

Issue

SFTP files not picked up in Copy activity or you get an error with errorCode "SftpPathNotFound"

Resolution steps

Step 0) Check if the file is present in the SFTP path using a third party tool like WinSCP or Putty

Step 1) SFTP reports the file modified date to ADF, this modified date is used by SFTP server to pick those files, whenever Customer is using modifiedDateTimeStart and modifiedDateTimeEnd filter Customer should check the time reported by the SFTP Server on when it was last modified and this time should be between the range provided in the filter, if you see that the last modified timestamp of the file is between the modifiedDateTimeStart and modifiedDateTimeEnd time range but still not picked by ADF, please go to step 2

Step 2) Some files will have a setting where it can be downloaded only once, Check with SFTP Server admin if they have this setting and increase the limit to a higher number, if this number is already high (like 8 or 10) and you still see the same issue, please go to Step 3

Step 3) Check the maxConcurrentConnections, which is a server side setting which can be controlled by ADF client, there is no upper limit to this number, and ensure to have maxConcurrentConnections adjusted according to the SFTP server bandwidth, this setting will ensure that those many number of connections (at most) are opened and the file is read in parallel from those connections. if adjusting this number (vary from 1 up to 10) did not make much of a difference, please go to Step 4,

Step 4) ADF behavior is, it will get the file length first and divide the files into multiple parts and read them in parallel, Some SFTP servers may not support this behavior of multiple connections reading a single file at same time, so to workaround this you can add "disableChunking": true under storeSettings of pipeline JSON, this would ensure the file is read only with one connection from start to end, but the downside to this is parallel reading will be disabled. This setting is available in UI also

```
            },
            "userProperties": [],
            "typeProperties": {
                "source": {
                    "type": "JsonSource",
                    "storeSettings": {
                        "type": "SftpReadSettings",
                        "recursive": true,
                        "enablePartitionDiscovery": false,
                        "disableChunking": true
                    },
                    "formatSettings": {
                        "type": "JsonReadSettings"
                    }
                }
```

Also note that SFTP uses reused stream to read the data, what it means is When copying data form SFTP, by default ADF tries to get the file length first, then divide the file into multiple parts and read them in parallel. And when reading these parts, ADF reuses already opened read stream. For example: the desired file size 240MB, and there will be 15(240/16) parts in total, by default ADF opens 10 connections, when one of the first 10 connections read a part completely, this connection will seek to the proper offset to download the 11th part. But when reused connection seeking to another offset, some SFTP servers may return wrong data. Obviously, this is a server-side issue, the server should return correct data in any cases. Customer need to contact the SFTP server vendor for help.

If there are issues reaching to SFTP server, you can also follow below steps to see what is the output and if you are able to reach to SFTP server and corresponding file path, this tool is written based on the SFTP SDK used in ADF

1. Download and unzip the tool: https://guosen.blob.core.windows.net/share/tool/TestSftpSDK.zip ⤤
2. Edit host.json to specify host, port, username, password, filepath.
3. Run TestCurrentSdk.exe and TestNewSdk.exe respectively to see how the results. The test tool will try to connect to the SFTP server, get file attributes, download file and disconnect