

Processing mainframe / EBCDIC data in Apache NiFi

Date published: 2019-12-16

Date modified: 2024-12-10



Legal Notice

© Cloudera Inc. 2024. All rights reserved.

The documentation is and contains Cloudera proprietary information protected by copyright and other intellectual property rights. No license under copyright or any other intellectual property right is granted herein.

Unless otherwise noted, scripts and sample code are licensed under the Apache License, Version 2.0.

Copyright information for Cloudera software may be found within the documentation accompanying each component in a particular release.

Cloudera software includes software from various open source or other third party projects, and may be released under the Apache Software License 2.0 (“ASLv2”), the Affero General Public License version 3 (AGPLv3), or other license terms. Other software included may be released under the terms of alternative open source licenses. Please review the license and notice files accompanying the software for additional licensing information.

Please visit the Cloudera software product page for more information on Cloudera software. For more information on Cloudera support services, please visit either the Support or Sales page. Feel free to contact us directly to discuss your specific needs.

Cloudera reserves the right to change any products at any time, and without notice. Cloudera assumes no responsibility nor liability arising from the use of products, except as expressly agreed to in writing by Cloudera.

Cloudera, Cloudera Altus, HUE, Impala, Cloudera Impala, and other Cloudera marks are registered or unregistered trademarks in the United States and other countries. All other trademarks are the property of their respective owners.

Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES NOT WARRANT THAT CLOUDERA PRODUCTS NOR SOFTWARE WILL OPERATE UNINTERRUPTED NOR THAT IT WILL BE FREE FROM DEFECTS NOR ERRORS, THAT IT WILL PROTECT YOUR DATA FROM LOSS, CORRUPTION NOR UNAVAILABILITY, NOR THAT IT WILL MEET ALL OF CUSTOMER’S BUSINESS REQUIREMENTS. WITHOUT LIMITING THE FOREGOING, AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, CLOUDERA EXPRESSLY DISCLAIMS ANY AND ALL IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, QUALITY, NON-INFRINGEMENT, TITLE, AND FITNESS FOR A PARTICULAR PURPOSE AND ANY REPRESENTATION, WARRANTY, OR COVENANT BASED ON COURSE OF DEALING OR USAGE IN TRADE.

Contents

Use case overview.....	4
Before you begin.....	4
Configuring EBCDICRecordReader.....	4
Building your dataflow.....	6

Use case overview

Learn how to use NiFi to process Mainframe data encoded in the Extended Binary Coded Decimal Interchange Code (EBCDIC) scheme while building dataflows in CDP Public Cloud.

Cloudera Flow Management offers a Cloudera exclusive NiFi component designed to process Mainframe / EBCDIC encoded data. This Record Reader component can be used to convert Mainframe data into other structured formats such as JSON or Avro.

**Note:**

The EBCDICRecordReader component is currently in Technical Preview and will transition to General Availability based on customer feedback.

This controller service component operates under the assumption that the data has already been offloaded from the Mainframe and NiFi would retrieve the binary data files and convert them using a processor like ConvertRecord.

The Mainframe / EBCDIC data processing workflow involves:

1. Retrieving binary files exported from the Mainframe.
2. Converting the EBCDIC encoded data into a readable format, such as JSON.
3. Storing or further processing the transformed data.

This use case walks you through creating a dataflow that receives files using GetFile, converts the data using the ConvertRecord processor with the help of the EBCDICRecordReader, and writes the converted data with PutFile. It gets you started with creating a basic conversion dataflow. However, if you require a different data source or target, you have various processor options in NiFi.

Before you begin

Before setting up a NiFi dataflow to process Mainframe data encoded in the Extended Binary Coded Decimal Interchange Code (EBCDIC) scheme between moving the data between your source and target, you must meet the following prerequisites.

- Ensure that Cobol copybooks describing the schemas of the tables are available in NiFi.
- Ensure that binary files exported from the Mainframe are available in NiFi.

Configuring EBCDICRecordReader

Learn how to create and configure the EBCDICRecordReader controller service for your conversion dataflow.

About this task

Controller services provide shared services to be used by the processors in your dataflow. Configure the EBCDICRecordReader before building the NiFi dataflow to ensure that it is available during processor configuration.

The EBCDICRecordReader reads Extended Binary Coded Decimal Interchange Code (EBCDIC) encoded input data and outputs a record representation. It relies on a COBOL copybook schema that needs to be provided using the Copybook Path property. One instance of this controller service can handle one type of data described by the corresponding copybook.

**Note:**

The EBCDICRecordReader component is currently in Technical Preview and will transition to General Availability based on customer feedback.

Configure the EBCDICRecordReader controller service with the path to the COBOL copybook, allowing NiFi to interpret the Mainframe data correctly.

Configure Controller Service | EBCDICRecordReader 1.0.0.2.16.0-323

SETTINGS

PROPERTIES

COMMENTS

Required field

Property

Value

Copybook Path

?

#{Copybook Path}/mycopybook.cbl

→

Copybook Character Set

?

UTF-8

↕

Code Page

?

Common

↕

Record Description Word Header Expected

?

true

↕

Record Description Word Byte Order

?

Big-endian

↕

Record Description Word Part Of Length

?

true

↕

CANCEL

APPLY

Example

Example COBOL copybook file structure:

5

```

1      01 MONTH-END-EXT-BORR-ACCT-REC.
2      02 EBORR-DEM-DATA.
3          04 EBORR-ACCT-NUM.
4              06 EBORR-SSN          PIC X(9).
5              06 EBORR-SSN-SUFFIX   PIC X.
6              04 FILLER              PIC X(288).          00670829
7              04 EBORR-NUM-OF-ABCD   PIC 99.
8      02 MONTH-END-EXT-ABCD-DATA OCCURS 1 TO 20 TIMES
9                                  DEPENDING ON EBORR-NUM-OF-ABCD.
10     04 EABCD-RECORD.
11         06 EABCD-BINT-FLG          PIC X(1).
12         06 EABCD-BINT-PEND-AMT     PIC S9(7)V9(2) COMP-3.
13         06 EABCD-BINT-OUTS-AMT     PIC S9(7)V9(2) COMP-3.
14         06 EABCD-BINT-PT-PD-AMT    PIC S9(7)V9(2) COMP-3.
15         06 EABCD-CS-BKD-PRM-AMT    PIC S9(5)V9(2) COMP-3.
16         06 EABCD-NC-BKD-PRM-AMT    PIC S9(5)V9(2) COMP-3.
17         06 EABCD-BKD-PRM-RT-AMT    PIC S9(2)V9(3) COMP-3.
18         06 EABCD-FTD-PRM-RT-AMT    PIC S9(2)V9(3) COMP-3.
19         06 EABCD-APP-YR             PIC 9(4).
20         06 EABCD-EBR-FLG           PIC X(1).
21         06 EABCD-ORG-1ST-PMT-DT.
22             08 EABCD-ORG-1ST-PMT-YY PIC X(4).
23             08 EABCD-ORG-1ST-PMT-MM PIC X(2).
24             08 EABCD-ORG-1ST-PMT-DD PIC X(2).
25 L30753 06 EABCD-MPN-SEQ-NUM        PIC S9(9) COMP.
26 L30753 06 EABCD-MPN-TYP-CD         PIC X(1).
27 L30753 06 EABCD-MPN-STA-CD         PIC X(1).
28 L30753 06 EABCD-ORIG-LDR-ID        PIC X(6).
29 L30753 06 EABCD-ORIG-LDR-BR-ID     PIC X(4).
30 L30753 06 EABCD-PR-OWN-ID          PIC X(6).
31 L30753 06 EABCD-PR-OWN-BR-ID      PIC X(4).

```

Building your dataflow

Learn how to create a NiFi dataflow that processes Mainframe / EBCDIC encoded data.

About this task

The basic steps of setting up the elements of your dataflow in NiFi involve opening NiFi in a CDP Public Cloud Flow Management cluster, adding processors to your NiFi canvas, and connecting the processors.

Procedure

1. Open NiFi
 - a) To access the NiFi service in your Flow Management cluster, navigate to Management Console service Data Hub Clusters .
 - b) Click the tile representing the Flow Management Data Hub cluster that you want to work with.
 - c) Click the NiFi icon in the Services section of the cluster overview page to access the NiFi UI.

You will be logged into NiFi automatically with your CDP credentials.

2. Add and configure the `GetFile` processor for data input.

This processor creates `FlowFiles` from files in a directory.

- a) Drag and drop the processor icon into the canvas.

This displays a dialog that allows you to choose the processor you want to add.

- b) Use the Add Processor filter box to search for the `GetFile` processor and click Add.

You can use other processors (for example SFTP processors) to retrieve offloaded binary files from a given location.

3. Add and configure the `ConvertRecord` processor for data conversion.

This processor converts records from one data format to another using Record Reader and Record Write Controller Services. You can use it to transform your EBCDIC encoded data into JSON or other structured formats. Configure it with the `EBCDICRecordReader` previously described and a default JSON writer.

4. Add and configure the `PutFile` processor for data output.

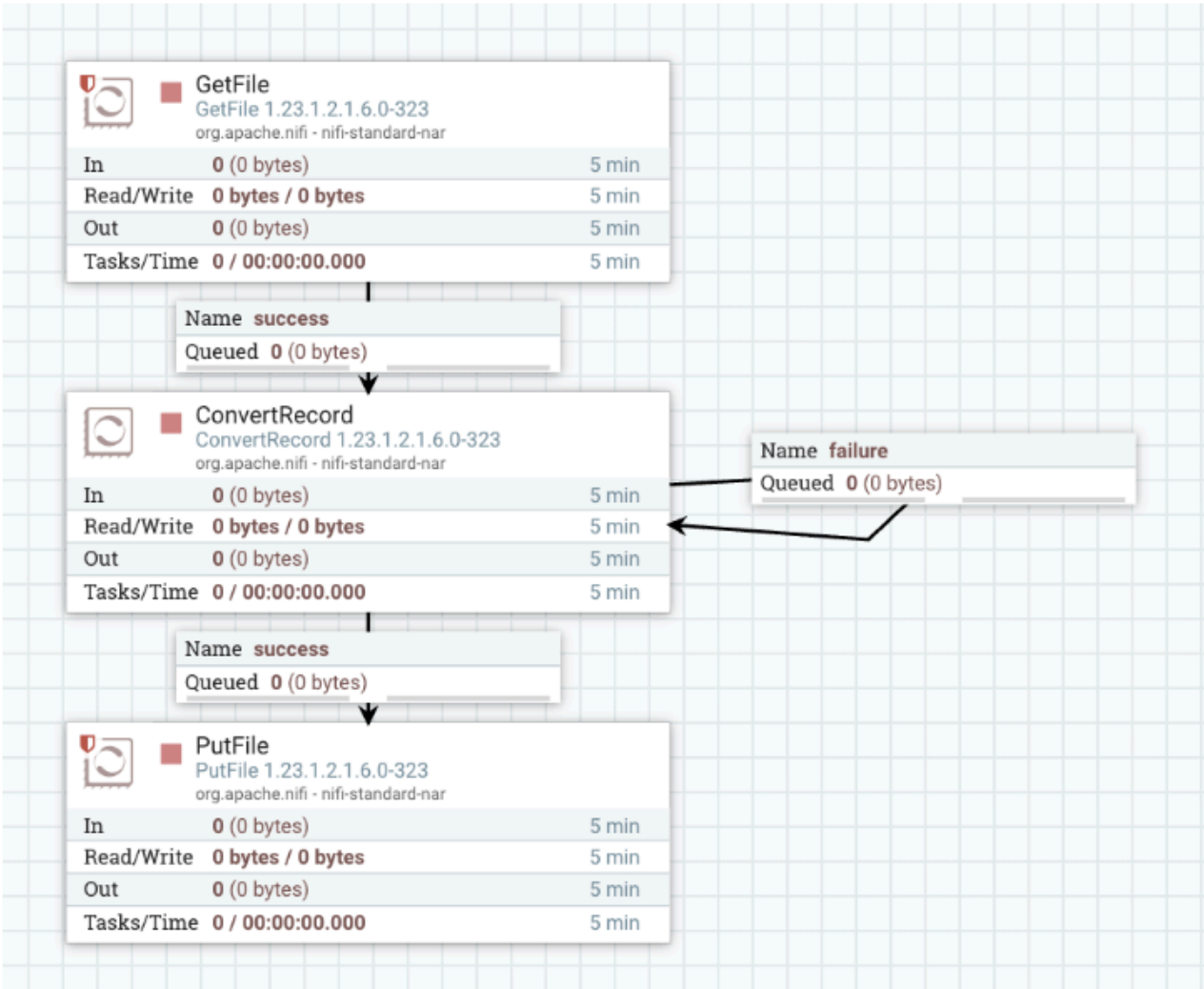
You can replace `PutFile` with any other sink processor appropriate for your use case (for example ingesting data into Iceberg, CDW, object store).

5. Connect the processors to create the flow.


- a) Drag the connection icon from the one processor, and drop it on the next processor.
- b) Configure the connection.
- c) Click Add to close the dialog box and add the connection to your flow.

Results

Your dataflow may resemble the following:



When this flow is running and receives a file from the Mainframe, it converts the data to JSON format, with a result that would look like the following:



View as: original ▼

```

1 [ {
2   "EBORR_DEM_DATA" : {
3     "EBORR_ACCT_NUM" : {
4       "EBORR_SSN" : "001030007",
5       "EBORR_SSN_SUFFIX" : "1"
6     },
7     "FILLER_0" : null,
8     "EBORR_NUM_OF_ABCDS" : 8
9   },
10  "MONTH_END_EXT_ABCD_DATA" : [ {
11    "EABCD_RECORD" : {
12      "EABCD_BINT_FLG" : " ",
13      "EABCD_BINT_PEND_AMT" : 0.00,
14      "EABCD_BINT_OUTS_AMT" : 0.00,
15      "EABCD_BINT_PT_PD_AMT" : 0.00,
16      "EABCD_CS_BKD_PRM_AMT" : 0.00,
17      "EABCD_NC_BKD_PRM_AMT" : 0.00,
18      "EABCD_BKD_PRM_RT_AMT" : 0.000,
19      "EABCD_FTD_PRM_RT_AMT" : 0.000,
20      "EABCD_APP_YR" : 0,
21      "EABCD_EBR_FLG" : " ",
22      "EABCD_ORG_1ST_PMT_DT" : {
23        "EABCD_ORG_1ST_PMT_YY" : "2007",

```