

Configuring the Complex Flat File Stage in InfoSphere Information Server

Support Technical Exchange

Marcia Stewart, IBM
May 21, 2015



Complex Flat File Stage

This stage provides a Graphical User Interface to guide the developer through an easier method to define schemas and create osh using a standard input/output operator, like a sequential stage, to process complex data files.



Complex Data File Considerations

- EBCDIC vs ASCII
- Fixed vs Variable length
 - Occurs depending on
 - Multiple record definitions
 - Record Descriptor Word
- Big Endian vs Little Endian
- IBM Codepage – NLS



Understanding EBCDIC

- Mainframe and AS/400 data is in EBCDIC

- Cobol data types used:

- Packed (COMP-3).
- Zoned (signed DISPLAY).
- Character (DISPLAY or PIC X).
- Integer (COMP. Integer)



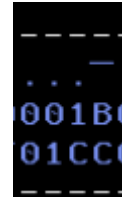
- Packed decimal data puts decimal digit in each nibble. There are two nibbles per byte. The sign is os dependent, but the value of C, A, E and F (positive) or D and B (negative) is placed in the last nibble.
- Zoned decimal data puts the decimal digit in the low order nibble and high order nibble is usually F. The last byte may contain the sign in the high order nibble (usually C or D) and last decimal in the low order nibble. Also known as display numeric. It helps to look at the data in hex, where they display as F1 through F9.

Understanding EBCDIC

■ Cobol data types used:

- Binary (same as COMP, but unsigned)
 - Decimal (COMP-2).
 - Float (COMP-1).
- comp-1 is single precision floating point and comp-2 is double precision floating point in IEE754 standard format.

They are seldom used.



Complex Data File Considerations

- **Verify fixed length file and record length**
 - Import Cobol file definition
 - View table definition via Layout tab, Cobol, last field end column
 - Take size of file in bytes and divide by the number above
 - Occurs – How do you want to process (to flatten or not)? Parallel occurs?
- **Determine record type: fixed versus variable length**
 - Occurs depending on
 - 01 VAR-RECORD.
 - 05 COUNTER-1 PIC S99.
 - 05 RECORD-1 PIC X(3)
 - OCCURS 1 TO 5 TIMES
 - DEPENDING ON COUNTER-1.
 - Multiple record types
 - Record indicator
 - Redefines
 - Multiple 01s



Occurs exceptions

- DataStage does not flatten array columns that have redefined fields.
- Columns with ODOs are not allowed to be flattened.
- Schema File property is not supported.



COBOL CFD and sample data

```
01 RECORD-1S.  
  02 FIELD1 PIC X(1).  
  02 FIELD2 PIC 9(13) COMP-3.  
  02 FIELD3 PIC 9(9) COMP.  
  02 FIELD4 PIC S9(9) COMP-3.  
  02 FIELD5  
    REDEFINES FIELD4  
    PIC 9(7)V9(2) COMP-3.  
  02 FIELD6 PIC 9(5)V9(2) COMP-3.  
  02 FIELD7 PIC 9(3) COMP-3.  
  02 FIELD8 PIC 9(5) COMP-3.  
  02 FIELD9 PIC X(1).  
  02 FIELD10 PIC X(1).  
  02 FIELD-GROUP.  
    03 FIELD11 OCCURS 6 TIMES  
      04 FIELD11A PIC X(1).  
      04 FIELD11B PIC X(1).  
  02 FIELD12 PIC 9(1).  
  02 FIELD13 PIC 9(13) COMP-3.  
  02 FIELD14 PIC X(1).
```

```
BROWSE      DSMMS1.TSTDemo  
Command ==>  
***** Top of Data *****  
  
-----  
M.b.....-.....le.000000000000123.....0  
D0832120001B0000000000982FFFFFFFFFFFFFFFF0000000F  
4021201F01CC0000C000F0F35F00000000000123000003F0  
-----  
M.Ê.....-.....βα..cË-000000000000234.....0  
D0712320002B00001005901875FFFFFFFFFFFFFFFF0000000F  
4021201F02CC0000C019F0F34F00000000000234000007F0  
-----
```



CFF Stage – Record Layout

Column	Picture clause	Starting column	Ending column	Storage length
01 RECORD_1 (49)				
02 FIELD1	PIC X(1).	1	1	1
02 FIELD2	PIC 9(13) COMP-3.	2	8	7
02 FIELD3	PIC 9(9) COMP.	9	12	4
02 FIELD4	PIC S9(9) COMP-3.	13	17	5
02 FIELD5	PIC 9(7)V9(2) COMP...	13	17	5
02 FIELD6	PIC 9(5)V9(2) COMP...	18	21	4
02 FIELD7	PIC 9(3) COMP-3.	22	23	2
02 FIELD8	PIC 9(5) COMP-3.	24	26	3
02 FIELD9	PIC X(1).	27	27	1
02 FIELD10	PIC X(1).	28	28	1
02 FIELD_GROUP		29	40	12
03 FIELD11		29	30	2
04 FIELD11A	PIC X(1).	29	29	1
04 FIELD11B	PIC X(1).	30	30	1
03 FIELD11_2		31	32	2
04 FIELD11A_2	PIC X(1).	31	31	1
04 FIELD11B_2	PIC X(1).	32	32	1
03 FIELD11_3		33	34	2
04 FIELD11A_3	PIC X(1).	33	33	1
04 FIELD11B_3	PIC X(1).	34	34	1
03 FIELD11_4		35	36	2
04 FIELD11A_4	PIC X(1).	35	35	1
04 FIELD11B_4	PIC X(1).	36	36	1
03 FIELD11_5		37	38	2
04 FIELD11A_5	PIC X(1).	37	37	1
04 FIELD11B_5	PIC X(1).	38	38	1
03 FIELD11_6		39	40	2
04 FIELD11A_6	PIC X(1).	39	39	1
04 FIELD11B_6	PIC X(1).	40	40	1
02 FIELD12	PIC 9(1).	41	41	1
02 FIELD13	PIC 9(13) COMP-3.	42	48	7
02 FIELD14	PIC X(1).	49	49	1



CFF Stage – File Options

The screenshot shows a dialog box titled "Complex_Flat_File_0 - PxCFF stage". The "File options" tab is selected. The "Stage name" field contains "Complex_Flat_File_0". The "File type" dropdown is set to "File(s)". The "File name(s)" field is empty. The "Record type" dropdown is set to "Fixed", and the "Record prefix" field is empty. The "Missing file mode" dropdown is set to "Depends", and the "Reject mode" dropdown is set to "Continue". The "Filter" field is empty. The "Multiple node reading" section has "Read from multiple nodes" unchecked and "Number of readers per node" set to 1. The "Report progress" and "Keep file partitions" checkboxes are also unchecked. The "Read first n rows" field is empty. At the bottom, there is a status bar showing "Fast Path: 1 of 5" and buttons for "OK", "Cancel", and "Help".

Complex_Flat_File_0 - PxCFF stage

Stage | Output

Stage name:
Complex_Flat_File_0

General | File options | Record options | Records | Records ID | Layout | NLS Map | Advanced

File type:
File(s)

File name(s):

Record type:
Fixed

Record prefix:

Missing file mode:
Depends

Reject mode:
Continue

Filter:

Multiple node reading:
 Read from multiple nodes
Number of readers per node:
1

Report progress
 Keep file partitions
Read first n rows:

Fast Path: 1 of 5

OK Cancel Help

CFF Stage – File Options

File requirements:

- EBCDIC files must be transferred in binary
- Variable length files must contain the Record Descriptor word
 - Ftp command: quote site rdw

Record Type:

Fixed - F,

Fixed Block - FB,

Variable - V,

Variable block - VB,

Variable spanned - VS,

Variable block spanned - VBS.

VR - This is provided for compatibility with Sequential stage



CFF Stage – Record Options

The screenshot shows a dialog box titled "Complex_Flat_File_0 - PxCFF stage" with two tabs: "Stage" and "Output". The "Stage" tab is active, and the "Record options" sub-tab is selected. The "Stage name" field contains "Complex_Flat_File_0". The "Record options" sub-tab includes the following settings:

- Float representation: IEEE (dropdown), Print fields
- General:
 - Byte order: Big-endian (dropdown)
 - Character set: EBCDIC (dropdown)
 - Data format: Binary (dropdown)
 - Record delimiter: (empty dropdown)
- Decimal:
 - Rounding: Nearest value (dropdown)
 - Separator: Project default (dropdown)
 - Allow all zeros
- Default values:
 - Character: (empty text box)
 - Decimal: 0 (text box)
 - Integer: 0 (text box)

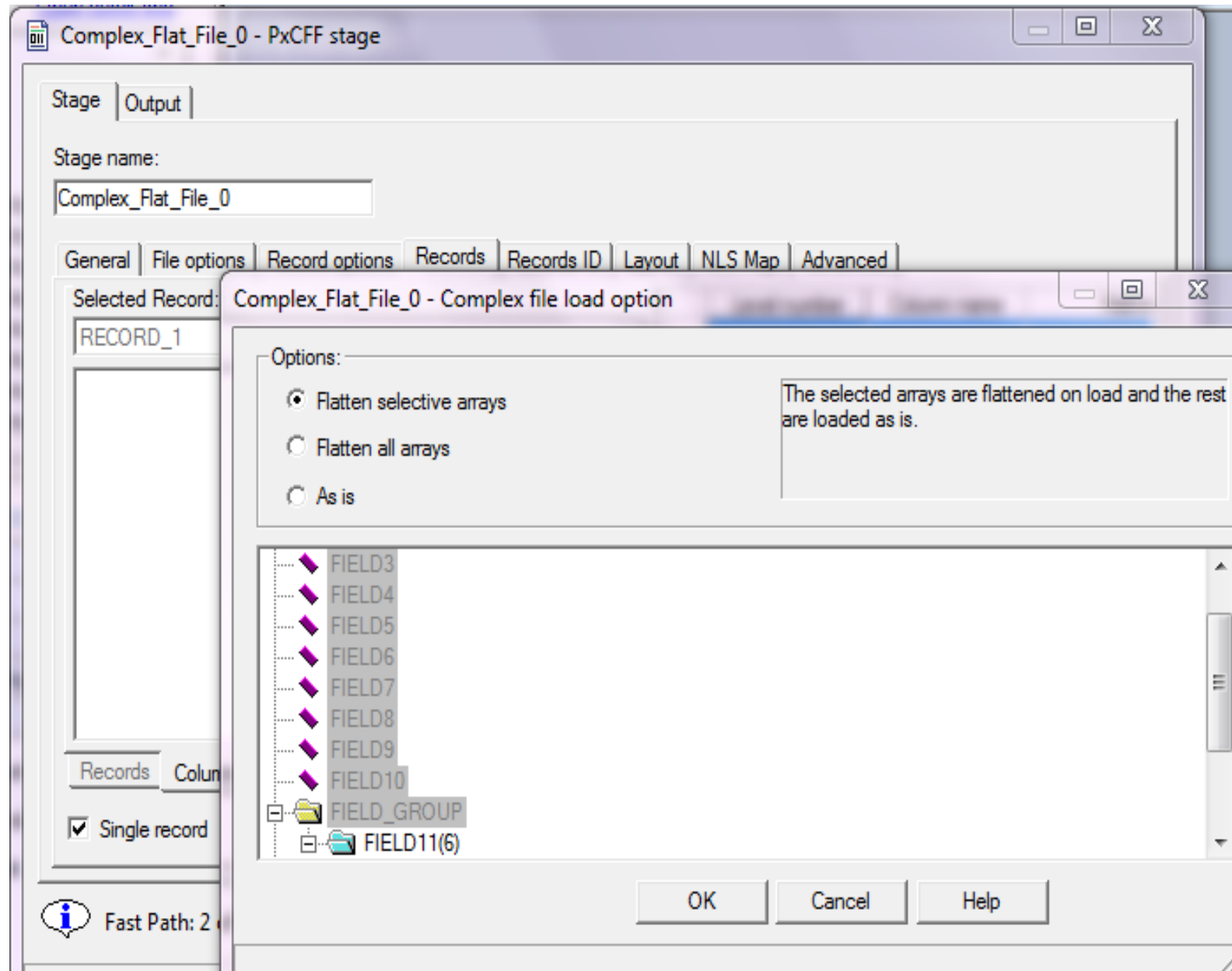
At the bottom, there is a "Fast Path: 1 of 5" indicator with left and right arrow buttons, and "OK", "Cancel", and "Help" buttons.



CFF Stage – Record Options

- Byte order – Endianess (usually default Native) – set Big Endian if mainframe file
- Data Format – binary or text
- Character set – EBCDIC or ASCII
- Record delimiter
- Allow all zeros
- Default values for Null handling
- Print fields – this option is helpful for debugging, though best to use a subset of data, especially if the file has many fields





Stage | Output

Stage name:
Complex_Flat_File_0

General | File options | Record options | Records | Records ID | Layout | NLS Map | Advanced

Selected Record:
RECORD_1

	Level number	Column name	Native type	Extended	Length	Scale	NLS map	Description
1	02	FIELD1	CHARACTER		1			
2	02	FIELD2	DECIMAL		13			
3	02	FIELD3	BINARY	Unsigned	9			
4	02	FIELD4	DECIMAL		9			
5	02	FIELD5	DECIMAL		9	2		
6	02	FIELD6	DECIMAL		7	2		

FIELD1
FIELD2
FIELD3
FIELD4
FIELD5
FIELD6
FIELD7
FIELD8
FIELD9
FIELD10
FIELD_GROUP
FIELD11
FIELD11A
FIELD11B
FIELD11_2
FIELD11A 2

Records | Columns

Properties:

- General
 - Column name = FIELD4
 - Native type = DECIMAL
 - Length = 9
 - Scale = 0
- Extended Attributes
 - Level number = 02
 - Sign indicator = Signed
 - Usage = COMP-3
- Derived Attributes

Value:

Available properties to add:

Single record

Save As... Clear All Load

Fast Path: 1 of 4

OK Cancel Help



CFF Stage – Column selection

Stage Output

Output name: DLink2 Columns... View Data...

General Selection Constraint Columns Advanced

Available columns:

- RECORD_1
 - FIELD1
 - FIELD2
 - FIELD3
 - FIELD4
 - FIELD5
 - FIELD6
 - FIELD7
 - FIELD8
 - FIELD9
 - FIELD10
- FIELD_GROUP
 - FIELD11
 - FIELD11A
 - FIELD11B
 - FIELD11_2
 - FIELD11A_2
 - FIELD11B_2
 - FIELD11_3

Selected columns:

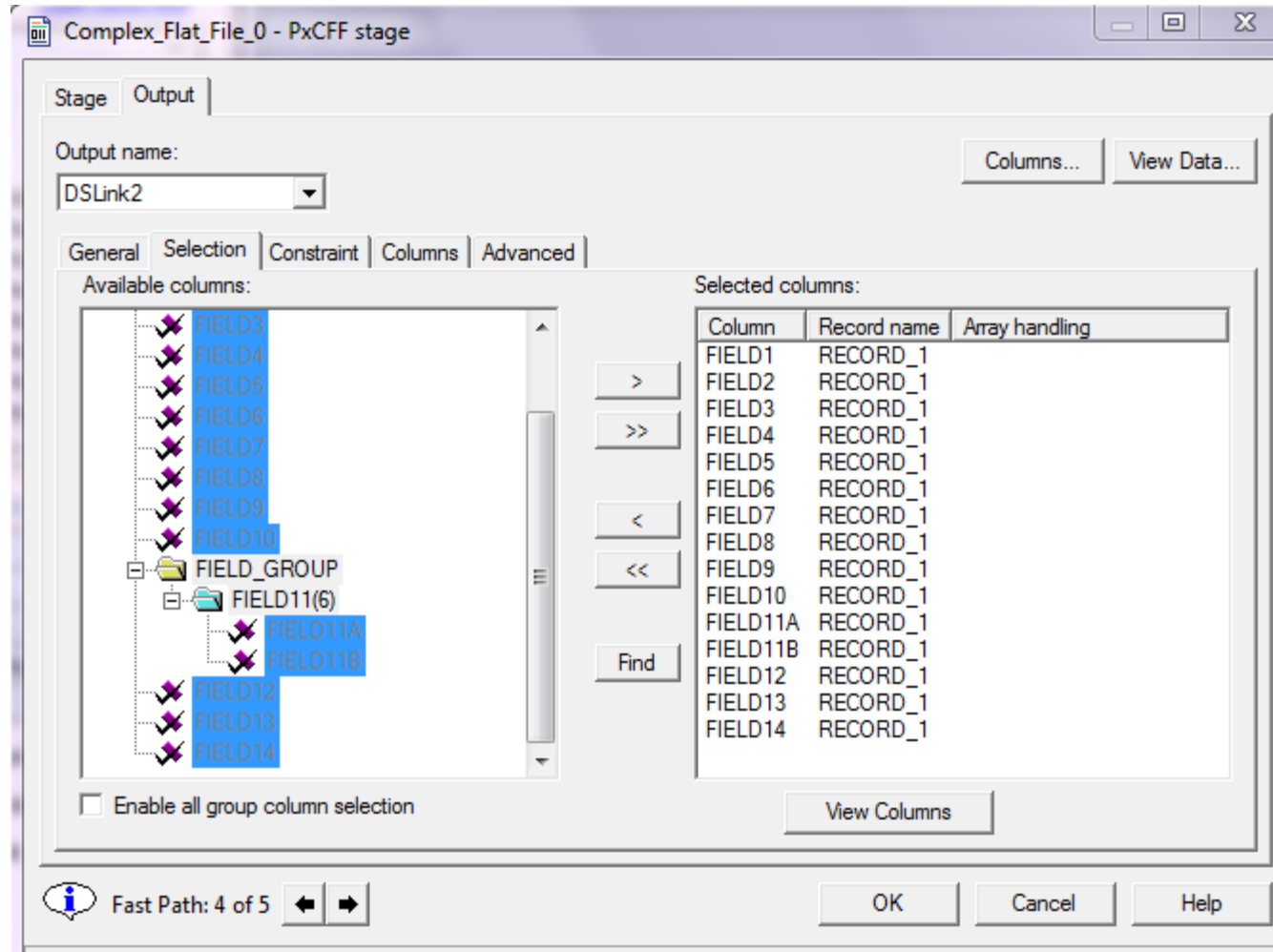
Column	Record name	Array handling
FIELD1	RECORD_1	
FIELD2	RECORD_1	
FIELD3	RECORD_1	
FIELD4	RECORD_1	
FIELD5	RECORD_1	
FIELD6	RECORD_1	
FIELD7	RECORD_1	
FIELD8	RECORD_1	
FIELD9	RECORD_1	
FIELD10	RECORD_1	
FIELD11A	RECORD_1	
FIELD11B	RECORD_1	
FIELD11A_2	RECORD_1	
FIELD11B_2	RECORD_1	
FIELD11A_3	RECORD_1	
FIELD11B_3	RECORD_1	
FIELD11A_4	RECORD_1	
FIELD11B_4	RECORD_1	
FIELD11A_5	RECORD_1	
FIELD11B_5	RECORD_1	
FIELD11A_6	RECORD_1	

Enable all group column selection View Columns

Fast Path: 3 of 4 OK Cancel Help



Column selection normalized occurs



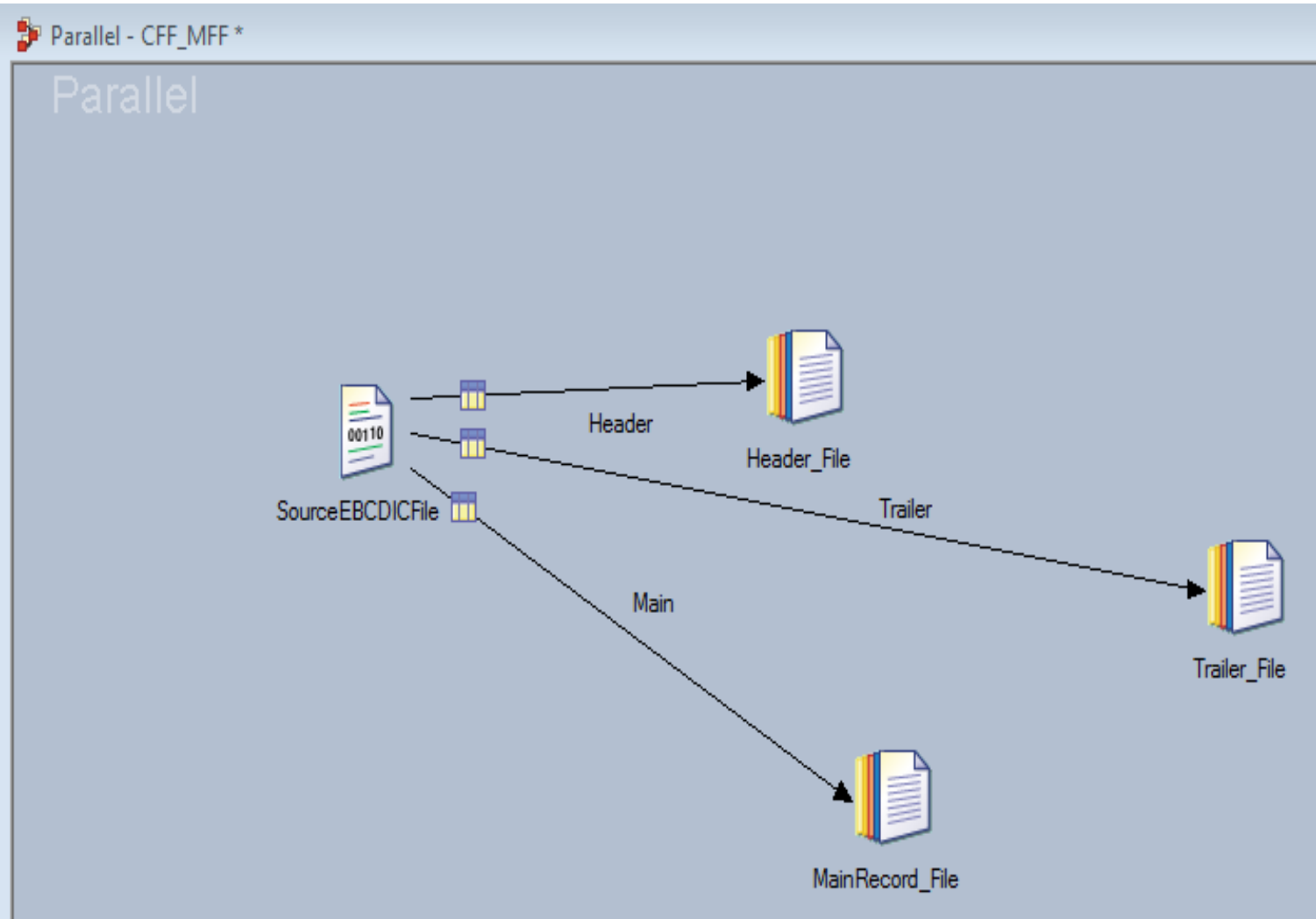
Multiple Record type

```
Menu Utilities Compilers Help
-----
BROWSE      DSMMS1.MFTDEMO
Command ==>
***** Top of Data *****

-----
H.b....._.....le.00000000000000123.....0
C0832120001B000000000000982FFFFFFFFFFFFFFFF0000000F
8021201F01CC0000F000F0F35F0000000000000123000003F0
-----
M.Ê....._.....βπ..cÈ~00000000000000234.....0
D0712320002B00001005901875FFFFFFFFFFFFFFFF0000000F
4021201F02CC0000F019F0F34F00000000000000234000007F0
-----
```



Multiple record type output links



Load Multiple Record Definitions

SourceEBCDICFile - PxCFE stage

Stage | Output

Stage name:
SourceEBCDICFile

General | File options | Record options | Records | Records ID | Layout | NLS Map | Advanced

	Level number	Column name	Native type	Extended	Length	Scale	NLS
1	02	FIELD1	CHARACTER		1		
2	02	FIELD2	DECIMAL		13		
3	02	FIELD3	BINARY	Unsigned	9		
4	02	FIELD4_H	DECIMAL		9		
5	02	FILLER_2	CHARACTER		32		

RECORD_HDR
RECORD_MAIN
RECORD_TR

Records | Columns

Properties:

- General
 - Column name = FIELD4_H
 - Native type = DECIMAL
 - Length = 9
 - Scale = 0
- Extended Attributes
 - Level number = 02
 - Sign indicator = Signed
 - Usage = COMP-3
- Derived Attributes

Value:

Available properties to add:

Single record

Save As... Clear All

Fast Path: 1 of 8

OK



Record Indicator Constraint

Stage | Output

Stage name:

General | File options | Record options | Records | Records ID | Layout | NLS Map | Advanced

Records	Column	Op	Value
RECORD_HDR RECORD_MAIN RECORD_TR	FIELD1	=	'H'



Column Selection

The screenshot shows the configuration window for the 'SourceEBCDICFile - PxCFF stage'. The 'Output' tab is active, and the 'Output name' is set to 'Main'. The 'Columns' sub-tab is selected, showing a tree view of available columns and a list of selected columns.

Available columns:

- RECORD_HDR
 - FIELD1
 - FIELD2
 - FIELD3
 - FIELD4_H
 - FILLER_2
- RECORD_MAIN
 - FIELD1_2
 - FIELD2_2
 - FIELD3_2
 - FIELD4
 - FIELD5
 - FIELD6
 - FIELD7
 - FIELD8
 - FIELD9
 - FIELD10
- FIELD_GROUP
 - FIELD11

Selected columns:

Column	Record name	Array handling
FIELD1_2	RECORD_MAIN	
FIELD2_2	RECORD_MAIN	
FIELD3_2	RECORD_MAIN	
FIELD4	RECORD_MAIN	
FIELD5	RECORD_MAIN	
FIELD6	RECORD_MAIN	
FIELD7	RECORD_MAIN	
FIELD8	RECORD_MAIN	
FIELD9	RECORD_MAIN	
FIELD10	RECORD_MAIN	

Buttons: >, >>, <, <<, Find, View Columns

Enable all group column selection



CFF Stage - NLS

- Default EBCDIC code page IBM037
- EBCDIC stateful encoding contains two code pages
 - Single byte characters
 - Double byte characters
- Multibyte character data utilizes control characters to indicate when switching to double byte. Shift in and shift out surround the multibyte characters (0x0E and 0x0F in hex)
- Character fields must be set with unicode when may contain multibyte data
- Use of column level code page setting
- HELPFUL LINK <http://www-01.ibm.com/software/globalization/cdra/>



Stage | Output

Stage name:
Complex_Flat_File_0

General | File options | Record options | Records | Records ID | Layout | NLS Map | Advanced

Map name:
Job default (ISO_8859-1:1987)

Allow per-column mapping

- Job default (ISO_8859-1:1987)
- Adobe-Standard-Encoding
- ANSI_X3.4-1968
- ASCL_ASCII
- ASCL_ASCII-PC1
- ASCL_BIG5
- ASCL_C0-CONTROLS
- ASCL_C1-CONTROLS
- ASCL_EBCDIC
- ASCL_EBCDIC-037
- ASCL_EBCDIC-1026
- ASCL_EBCDIC-500V1
- ASCL_EBCDIC-875
- ASCL_EBCDIC-CTRLS
- ASCL_EBCDIC-IBM1364
- ASCL_EBCDIC-IBM1371
- ASCL_EBCDIC-IBM933
- ASCL_EBCDIC-IBM937
- ASCL_EBCDIC-JP-CTRLS
- ASCL_EBCDIC-JP-KANA
- ASCL_EBCDIC-JP-KANA-E
- ASCL_EBCDIC-JP-KANA-HW
- ASCL_GB2312

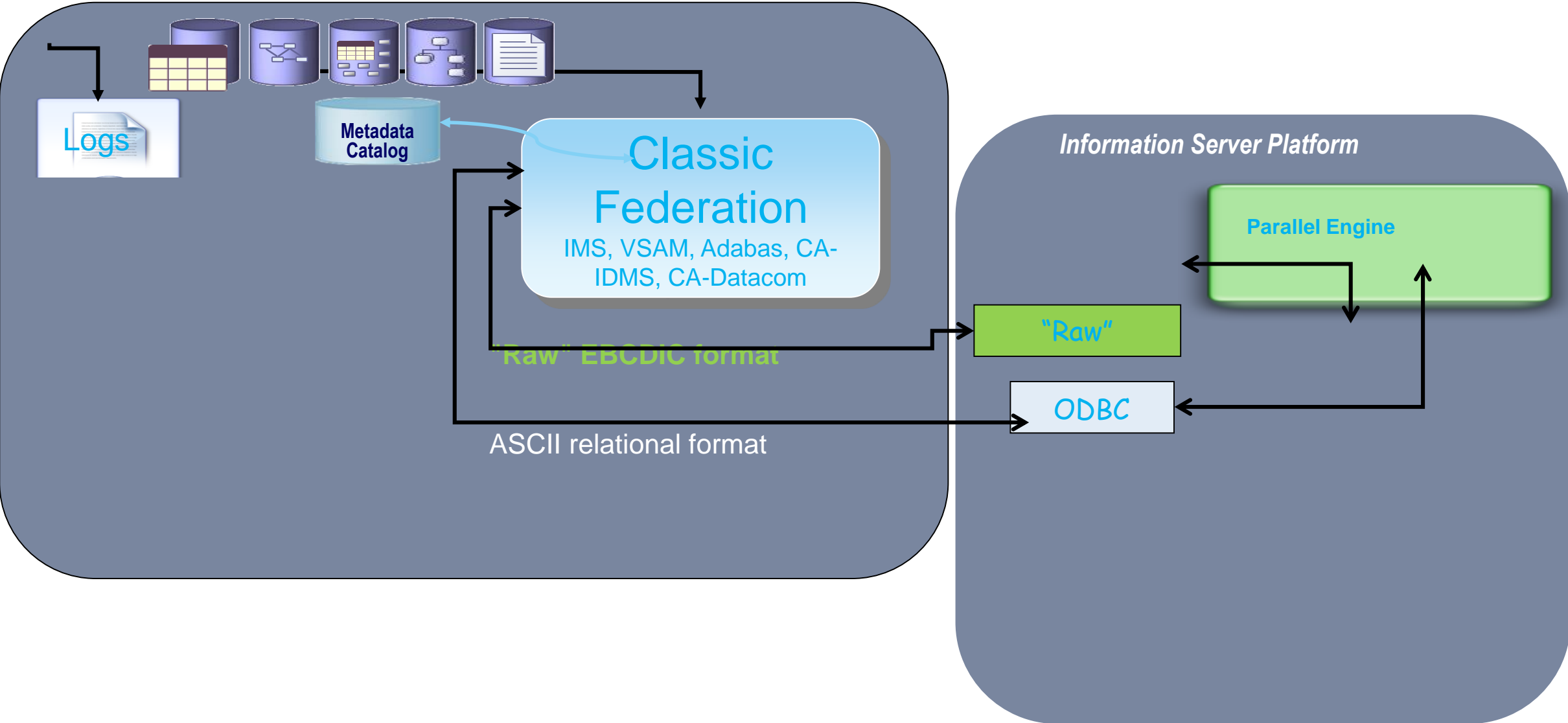


CFF Stage – Gotchyas

- Column definitions must contain level number
- quote site rdw – be sure to ftp variable length files with this command
- Single byte binaries not supported in CFF stage. Must use sequential stage.
 - To confirm data type, view data in hex to see if the values look like 0xF1, .. versus 0x01... which is difference if value is display numeric vs single byte binary.
 - Workaround, in Sequential stage define column as tiny int or binary with length of one.
- VSAM file can only be read if fixed off sets
 - Workaround, have the VSAM file dumped to sequential file
 - Use Classic Federation
- Reject record option only supported with fixed length records
 - Single record type
 - No occurs depending on
 - Code page must only contain single byte characters



InfoSphere z/OS Data **ACCESS** Integration with DataStage



Attend our next Support Technical Exchange

- Topic: Setting up basic DataStage Users
- Date: Tuesday June 16, 2015
- Time: 11:00am ET
- Speaker: Julia Hirlinger, IBM Technical Support, DataStage
- Register: <https://events.na.collabserv.com/register.php?id=d57874bed5&l=en-US>

