

Database - Feature #6457

finish COPY-LOB support

05/27/2022 12:16 PM - Eric Faulhaber

Status:	WIP	Start date:	
Priority:	Normal	Due date:	
Assignee:	Stanislav Lomany	% Done:	0%
Category:		Estimated time:	0.00 hour
Target version:		vendor_id:	GCD
billable:	No		
Description			
Related issues:			
Related to Base Language - Bug #6623: validate the memptr bytes in "copy-lob ...			New

History

#1 - 05/27/2022 12:17 PM - Eric Faulhaber

The remaining work:

- Encoding behavior needs better testing (at least [#6623](#) must be fixed).
- BOM handling for file source/target not implemented yet
- Some remaining issues (see [#4768-8](#))

#2 - 07/21/2022 07:20 AM - Greg Shah

- Related to Bug #6623: validate the memptr bytes in "copy-lob from memptr to clob/longchar" statement added

#3 - 07/28/2022 06:55 PM - Greg Shah

- Assignee set to Stanislav Lomany

#4 - 09/13/2022 07:04 AM - Greg Shah

The BOM handling has a testcase already written. See testcases/copy_lob/tests/bom.p.

#5 - 09/13/2022 07:27 PM - Stanislav Lomany

The BOM handling has a testcase already written. See testcases/copy_lob/tests/bom.p.

OK, the first issue I met in bom.p surprisingly was that the following sequence leads to invalid output file content in 4GL because it runs some conversion due to convert option. FWD doesn't do it and the output file is fine UTF-8.

```
inputLongchar = ?.
fix-codepage(inputLongchar) = 'UTF-8'.
inputLongchar = "A" + CHR(14844588, "UTF-8", "UTF-8") + "A".
copy-lob from inputLongchar to file 'copy_lob/output/out.tmp' convert target codepage 'UTF-8' no-error.
```

#6 - 09/20/2022 07:24 PM - Stanislav Lomany

- Status changed from New to WIP

Guys, I don't quite understand how 4GL produces output. Consider the following UTF-16 output example:

```
inputLongchar = ?.  
fix-codepage(inputLongchar) = 'utf-8'.  
inputLongchar = "A" + CHR(14844588, "UTF-8", "UTF-8") + "A".  
copy-lob from inputLongchar to file 'copy_lob/output/out16.tmp' convert target codepage 'UTF-16' no-error.
```

4GL produces this for A€A:

```
ff fe 41 00 e2 00 82 00 ac 00 41 00  
BOM A Euro???? A
```

While normal UTF-16 output is

```
ff fe 41 00 ac 20 41 00  
BOM A Euro A
```

Do you have any idea what and how 4GL produces?

#7 - 09/21/2022 01:45 AM - Constantin Asofiei

Stanislav, looking at the 14844588 value, in hex is e282ac. See this: <https://community.progress.com/s/article/P181822> Please experiment with -cpstream and see what happens. More, in the docs https://documentation.progress.com/output/ua/OpenEdge_latest/pdsoe/PLUGINS_ROOT/com.openedge.pdt.langref.help/rfi1424920632352.html it states:

If either the source or the target object is a file, the target's code page defaults to -cpstream.

On a side note, test abends in OE on Windows if the target path does not exist on disk.

#8 - 09/21/2022 12:11 PM - Ovidiu Maxiniuc

Stanislav Lomany wrote:

Do you have any idea what and how 4GL produces?

Try launching the 4GL using `-cpinternal utf8` command-line parameter. This may shed some light on how it works. In my opinion, it assigns the € representation in UTF-8 to longchar variable using some UTF to default CP conversion which breaks the content, ignoring the previous `fix-codepage` statement.

#9 - 10/06/2022 03:51 PM - Stanislav Lomany

Guys, I've been trying to experiment with `-cpstream` and `-cpinternal` parameters and found that when it comes to setting them to UTF-16/UTF-32 I can set `-cpstream` to to UTF-16 and that's it. For other cases it produces "Case table for code page UTF-16 and case name BASIC was not found in convmap.cp (1038)" error message or silently abends.

Do you have any idea how to fix it?

#10 - 10/06/2022 04:49 PM - Ovidiu Maxiniuc

I know, I had the same problems :(.

Try setting the same value for all these cp parameters, unless having them different is your goal.

#11 - 10/07/2022 03:25 AM - Stanislav Lomany

Try setting the same value for all these cp parameters

Errors are the same for this case too.

#12 - 10/12/2022 09:16 AM - Stanislav Lomany

I cannot change `-cpstream/-cpinternal` parameters on a customer's VM either, so I cannot experiment with it.

Moreover, `testcases/copy_lob/tests/bom.p` fails on this VM because output produced for UTF-8 is different (using default `-cpstream/-cpinternal`).

Greg, what should we do about these parameters if we cannot be sure how they work?

#13 - 10/12/2022 09:29 AM - Greg Shah

I cannot change -cpstream/-cpinternal parameters on a customer's VM either, so I cannot experiment with it.

Is the testing issue caused by the virtual machine not having the needed codepages installed? Otherwise I don't understand why OE would not honor these command line specifications.

#14 - 10/12/2022 09:58 AM - Stanislav Lomany

Is the testing issue caused by the virtual machine not having the needed codepages installed?

Some cases claim that "case table" for this particular case is missing in convmap.cp (quick googling didn't tell me how to add something to it). Some cases abend. And some lead to artefacts in screen output.

#15 - 10/12/2022 10:12 AM - Greg Shah

Please provide the list of codepages/scenarios which need testing (and which you cannot test).

Marian: I think we need your team to do this testing.

#16 - 10/13/2022 08:46 AM - Marian Edu

Greg Shah wrote:

Please provide the list of codepages/scenarios which need testing (and which you cannot test).

Marian: I think we need your team to do this testing.

Not sure what need to be tested here, the codepage/collation support might be different from a version to another and some customisation can be done but we've made some procedures that list:

- the supported codepages/collation combinations: i18n/4gl/cp_collation.xml
- the accepted conversion between codepages for codepage-convert: i18n/4gl/cp_conversion.xml
- the accepted conversion between codepages for chr: i18n/4gl/chr_conversion.xml
- the accepted conversion between codepages for input stream: i18n/4gl/is_conversion.xml

#17 - 10/16/2022 03:05 PM - Stanislav Lomany

i18n/4gl/cp_collation.xml i18n/4gl/cp_conversion.xml i18n/4gl/chr_conversion.xml i18n/4gl/is_conversion.xml

Marian, these are theoretically allowed conversions. The problem is that I cannot set -cpstream/-cpinternal parameters to test how they work. I either get an error that "case table" for this particular case is missing in convmap.cp or it just abends.
Do you think you'll be able to set -cpstream/-cpinternal with a working result? If so, I'll make a testcase to be tested with all -cpstream/-cpinternal combinations.

#18 - 10/17/2022 08:42 AM - Marian Edu

Stanislav Lomany wrote:

i18n/4gl/cp_collation.xml i18n/4gl/cp_conversion.xml i18n/4gl/chr_conversion.xml i18n/4gl/is_conversion.xml

Marian, these are theoretically allowed conversions. The problem is that I cannot set -cpstream/-cpinternal parameters to test how they work. I either get an error that "case table" for this particular case is missing in convmap.cp or it just abends.

What are the values you're trying to use for the codepages/collation tables - for cpstream and cpinternal?

Do you think you'll be able to set -cpstream/-cpinternal with a working result? If so, I'll make a testcase to be tested with all -cpstream/-cpinternal combinations.

Yes, the idea is that if the combination is valid then there shouldn't be any error although is quite rare to have different codepages exactly because of the conversion overhead - the client uses the codepage appropriate for the user, the server often use the same codepage as client or at times some UTF if the database needs to support various codepages.

#19 - 10/17/2022 10:33 AM - Stanislav Lomany

What are the values you're trying to use for the codepages/collation tables - for cpstream and cpinternal?

I want to test this ones:

UTF-8

UTF-16

UTF-16LE

UTF-16BE

UTF-32

UTF-32LE

UTF-32BE