

Handling special multi-byte characters using Double-Byte Character Support

This article is linked to KB Item [23577](#).

This article concerns the need to handle multi-byte characters in COBOL programs particularly of special characters (ie non-standard English language characters). It arose as a result of configuring an Oracle database as UTF-8 and, thus, multi-byte characters. Problems were encountered when COBOL programs needed to contain the result of a read to the database resulted in the read field becoming truncated. To this end it was necessary to define the field lengths in the COBOL program in a more appropriate manner.

The solution to this problem is Double-Byte Character Set Support.

This kind of support exists because there are many (perhaps thousands) of different characters pertaining to the world's stock of different languages. The majority of computer systems use bytes which are 8 bits in length; in that they assign an 8 bit code to represent each different character. However, this method has its limitations and can represent up to a maximum of 256 characters.

Normally, a programmer does not need to bother him/herself with the internal code which is used to represent characters. The Single-Byte Character Set (SBCS) of the American Standard Code for Information Interchange (commonly known as ASCII) is quite sufficient for programmers operating in the English language. However, for programmers operating in non-standard English languages the limitation of 256 can become somewhat restrictive.

To this end Double-Byte Character Set (DBCS) is provided in which each character is represented by a code length of 16 bits ie. double the standard length of 8 bits and hence the name "Double-Byte" Character Set. Each character occupies a pair of adjacent bytes and the resulting array of possible characters runs into the thousands. The assignment of DBCS character codes to characters varies from country to country.

Micro Focus Extensions for Double-Byte Character Support is enabled by the NCHAR or JAPANESE Compiler directives. Double-Byte Character set support is sensitive to the DBCS Compiler directive.