What is Data Conversion

January 25, 2008   Editorial Team+   Data Warehouse Basics   No comments

Data Conversion, as the name implies, deals with changes required to move or convert data from one physical environment format to that of another, like moving data from one electronic medium or database   product onto another format.

Every day, data is being shared from one computer to another. This is a very common activity especially in data warehouses where database severs gather, extract, transform and load data from different sources at every moment. Since these data gathered and shared from different computers which may have different hardware and software platforms, there should be a mechanism in dealing with data so that each computer server   receiving them can understand what information the data contains.

Data conversion is technical process of changing the bits contained in the data from one format to another format for purpose of interoperability between computers. The most simple example of data conversion is a text file converted from one character encoding to another. Some of the more complex conversions involve conversion of office file formats and conversion of audio, video and image file format which needs to consider different software applications   to play or display them.

Data conversion can be difficult and painstaking process. While it may be easy for a computer to discard information, it is difficult to add information. And adding information is not just simply padding bits but sometimes is would involve human judgment. Upsampling, the process of converting data to make it more feature rich, is not about adding data. It is about making room for addition, a process which also needs human judgement.

To illustrate, a true color image can be easy to convert to grayscale but not the other way around. A Unix text file can be converted to Microsoft file by simply adding a CR byte but adding color information to a grayscale image cannot be programmatically dne because only human judgment can know which colors are appropriate for each section of the image; this is not rule based that can be easily done by a computer   .

Despite that fact that data conversion can be done directly from format to another desire format, may application use pivotal encoding in converting data files. For instance, converting Cyrillic text files from KO18-R to Windows-1251 is possible with direct conversion using a look up table between encodings. But a more people use conversion from by first converting the KOI8-R file to Unicode before converting to Windows-1251 because of manageability benefits. Conversion with character encoding is a lot easier this way because having a lookup table and all permutations of character encodings involves hundreds of records.

Data conversion sometimes results to loss of information. For instance, converting a Microsoft Word file to plain text files results in a lot of data loss   because the text file removes the Word formatting feature. To prevent this from happening, the target format must support the same data constructs and features of the source data.

Inexactitude can also be a result of data conversion. This means that the result of the conversion can be conceptually different from the source files. An example would be the extant in word processors, WYSIWG paradigm and desktop publishing applications compared to the structural descriptive paradigm found in XML and SGML.

It is important to know the workings of both source and target format when converting data. If the format specifications are unknown, reverse engineering can be applied to carry out any conversion as this can attain close approximation of the original specification although there is no assurance that there can be no error or inconsistency. In any case, there applications that can detect errors so appropriate actions can be done.