Binary Data Formats - Advanced Detective Work

Binary Data Format

When reading binary data, origins generate a record with a single byte array field at the root of the record.

When the data exceeds the user-defined maximum data size, the origin cannot process the data. Because the record is not created, the origin cannot pass the record to the pipeline to be written as an error record. Instead, the origin generates a stage error.

When writing binary data, destinations and processors write binary data to a single field in the record.

How is a binary file used?

Binary files are not human readable and require a special program or hardware processor that knows how to read the data inside the file. Only then can the instructions encoded in the binary content be understood and properly processed.

The following screenshot shows part of the content from a file on a Mac computer. The content can be understood only by specific system architectures such as Mac. The file in this figure is a Unix executable binary file. As is typical with binary files, the file's content is a series of sequential bytes, lined up one after the other. Although we can view the bytes that make up the binary data, the file can be properly interpreted only by the target platform.

What is a binary file?

A binary file is a file whose content is in a binary format consisting of a series of sequential bytes, each of which is eight bits in length. The content must be interpreted by a program or a hardware processor that understands in advance exactly how that content is formatted and how to read the data. Binary files include a wide range of file types, including executables, libraries, graphics, databases, archives and many others.