

Why does COM express GUIDs in a mix of big-endian and little-endian? Why can't it just pick a side and stick with it?

 devblogs.microsoft.com/oldnewthing/20220928-00

September 28, 2022



Raymond Chen

Wikipedia claims that the COM textual representation of GUIDs is mixed-endian.

Is it, really?

No, it is little-endian all the way. But if you don't understand how GUIDs are formed, it might look like some parts are big-endian.

The parts of a GUID as defined in the specification are

Field	Type
time_low	32-bit integer
time_mid	16-bit integer
time_hi_and_version	16-bit integer
clock_seq_hi_and_reserved	8-bit integer
clock_seq_low	8-bit integer
node	6-byte MAC address

The **GUID** structure breaks it down as

```
struct GUID
{
    uint32_t Data1;
    uint16_t Data2;
    uint16_t Data3;
    uint8_t Data4[8];
}
```

Let's line up the two pieces against each other.

Field	Type	Structure
time_low	32-bit integer	Data1
time_mid	16-bit integer	Data2
time_hi_and_version	16-bit integer	Data3
clock_seq_hi_and_reserved	8-bit integer	Data4[0]
clock_seq_low	8-bit integer	Data4[1]
node	6-byte MAC address	Data4[2..7]

If you print out each piece of the GUID structure, with hyphens between each part, then you get

33221100-5544-7766-88-99-AA-BB-CC-DD-EE-FF

Notice that everything is still little-endian. We didn't have to do any byte flipping when printing:

```
printf("%04x-%04x-%04x-%02x-%02x-%02x-%02x-%02x-%02x-%02x",
    Data1, Data2, Data3,
    Data4[0], Data4[1], Data4[2], Data4[3],
    Data4[4], Data4[5], Data4[6], Data4[7]);
```

My guess is that the folks who designed the string format thought there were too many dashes, so they removed the byte dashes, except for the one that separates the clock bytes from the MAC address.

```
printf("%04x-%04x-%04x-%02x%02x-%02x%02x%02x%02x%02x%02x",
    Data1, Data2, Data3,
    Data4[0], Data4[1], Data4[2], Data4[3],
    Data4[4], Data4[5], Data4[6], Data4[7]);
```

33221100-5544-7766-8899-AABBCCDDEEFF

The result is that the last two pieces of the stringified GUID look big-endian, but they're not. They're just little-endian with some dashes missing.

Raymond Chen

Follow



