
Subject: Re: Changes in hashing
Posted by [mirek](#) on Fri, 05 Jun 2020 10:40:47 GMT
[View Forum Message](#) <> [Reply to Message](#)

Oblivion wrote on Fri, 05 June 2020 12:28Hello Mirek,

I'm somewhat confused about this.

Just to be clear: Does this mean that the client code can continue to use dword GetHashValue() variant on 64-bit machines?

I'm worried because, e.g, In Terminal ctrl I use a dword hash of incoming string data as unique ID for images and hyperlinks, in each cell.

Changing it to 64-bit would be very expensive (memory consumption will be significantly higher).

Truncating 64-bit to 32-bit probably won't do too much harm here, but still, it means information loss and I'd prefer to avoid that.

Or do I have to separately maintain the 32-bit hash funtions on 64-bit confiurations?

Best regards,
Oblivion

Well, first of all, using GetHashValue as unique ID is probably not a good idea, but I guess you mean something slightly different there.

Second, simply truncating to 32-bit bit would indeed, for memhash produced numbers, produce inferior hashes, just like taking lower bits in previous 32-bit incarnation, as fast hashing algorithms tend to accumulate entropy in highest bits. But exactly for this reason we have (for quite a long time) FoldHash function, which IMO seems ideal for you scenario. It takes hash_t and produces dword, while bringing entropy back to lowest bits (BTW, look at implementation, I think it is one of more clever ideas from me... :)

Mirek
