

Klaas Brant over IBM's DB2 Universal Database

# No, no... my name is Antonio!

Toen ik onlangs les gaf in Portugal kwam na afloop een man naar mij toe. In zijn beste Engels zei hij: "Thank you very much, this was very useful!" Mijn standaard antwoord is "You are welcome". Tot mijn verbazing zei hij toen "No, no... my name is Antonio". Het is duidelijk, waarschijnlijk heeft Antonio gedurende de week wel meer niet begrepen. Thuis vroeg ik mij af of we dergelijke 'communicatieproblemen' ook met onze databases kunnen verwachten.

Als we problemen in de communicatie aan de orde stellen, dan moeten we het probleem opsplitsen in twee problemen. Ten eerste kunnen we elkaar wellicht echt niet begrijpen. Vele mensen in de wereld spreken slechts één taal en als twee partijen ieder een andere taal spreken dan hebben we een probleem.

Ook in IT bestaan dergelijke problemen. Zo zijn er diverse manieren om onze digitale informatie op te slaan. De bekendste zijn ASCII en EBCDIC; twee 'talen' die elkaar absoluut niet verstaan. Zelfs als we 'dezelfde' taal spreken dan kan het fout gaan. Dialecten kunnen de communicatie soms stevig verstoren. Binnen het dialect is er geen probleem, maar erbuiten kan miscommunicatie ontstaan. In de IT heten de dialecten code pages en zolang beide partijen maar hetzelfde dialect (lees code page) spreken dan gaat het goed.

Als iedereen in de toekomst Unicode spreekt dan zijn de problemen weg.

Onze beste Antonio sprak ook een soort Unicode, namelijk Engels en toch ging het mis. Waarom? Omdat als we het eerste probleem van de taal opgelost hebben, we automatisch met het tweede probleem geconfronteerd worden. Het tweede probleem is het inhoudelijk begrijpen van de boodschap. Mocht u nu denken dat dit niets te maken heeft met Unicode dan heeft u het mis. Heeft u ooit gehoord van *endianess*? De manier waarop twee byte-code wordt opgeslagen is op diverse platformen verschillend. Zo spreekt men van *little endian* op Intel-machines en *big endian* op Unix, Macintosh en mainframes. De volgorde van het opslaan van bytes (high en low order) is precies omgekeerd! Als Unicode in twee bytes wordt opgeslagen (UTF-16) dan zal een file van het ene platform totaal onbruikbaar zijn op het andere platform.



Om het probleem nog een beetje erger te maken, wil het omkeren van de bytes niet direct zeggen dat de file geen geldige Unicode meer is. Het omkeren van de bytes heeft meer tot effect dat er letterlijk Chinees ontstaat op het andere platform (een normale N wordt een Chinese 1 als men de bytes omkeert).

Dus als u informatie via het file-systeem van uw Windows DB2 naar uw mainframe of Unix DB2 transporteert kan dit probleem optreden. Om deze interpretatieproblemen te voorkomen wordt aangeraden om het eerste character van een Unicode file een zogenaamd Byte Order Mark (BOM) te laten zijn, om aan te geven of de rest van de file big endian (0xFEFF) of little endian (0xFFFE) is. De BOM characters zijn uniek, maar de ontvangende software moet wel in staat

zijn ze te interpreteren en of te negeren. Bij een directe koppeling van op diverse platforms draaiende DB2-versies, zal de communicatie-software dit soort problemen automatisch voor u oplossen.

Dus als we Unicode goed aanpakken dan zijn er geen problemen meer? Helaas, als laatste houden we nog het moeilijkste probleem over. We spreken dezelfde taal en toch begrijpen we elkaar niet, zoals in het geval van Antonio.

Wist u dat als men in Brabant zegt "Ik ben om zeven aangeden", dat dit betekent dat men om zeven uur vertrokken is en niet dat men om zeven uur tegen een andere auto is gereden? Vaak wordt er naar XML gewezen als het om inhoudelijke interpretatie gaat. Maar ook XML zal geen einde maken aan dit soort Babylonische spraakverwarringen. XML gaat er hooguit voor zorgen dat het juiste veld wordt aangewezen, als u bijvoorbeeld een leverdatum zoekt. Maar als de maker van de file de dag waarop het artikel de fabriek verlaat bedoelt en u denkt dat dit de dag is waarop u de artikelen ontvangt, dan kan er nog heel veel mis gaan.

Hiermee komen we alweer aan het eind van de laatste column van 2003. Ik wens u bijzonder fijne feestdagen en heel veel goeds voor 2004.

**Klaas Brant** (kbrant@kbce.nl) is DB2-specialist en directeur van KBCE b.v. Meer informatie over DB2 is te vinden op [www.kbce.nl](http://www.kbce.nl) en [www.db2-times.com](http://www.db2-times.com)