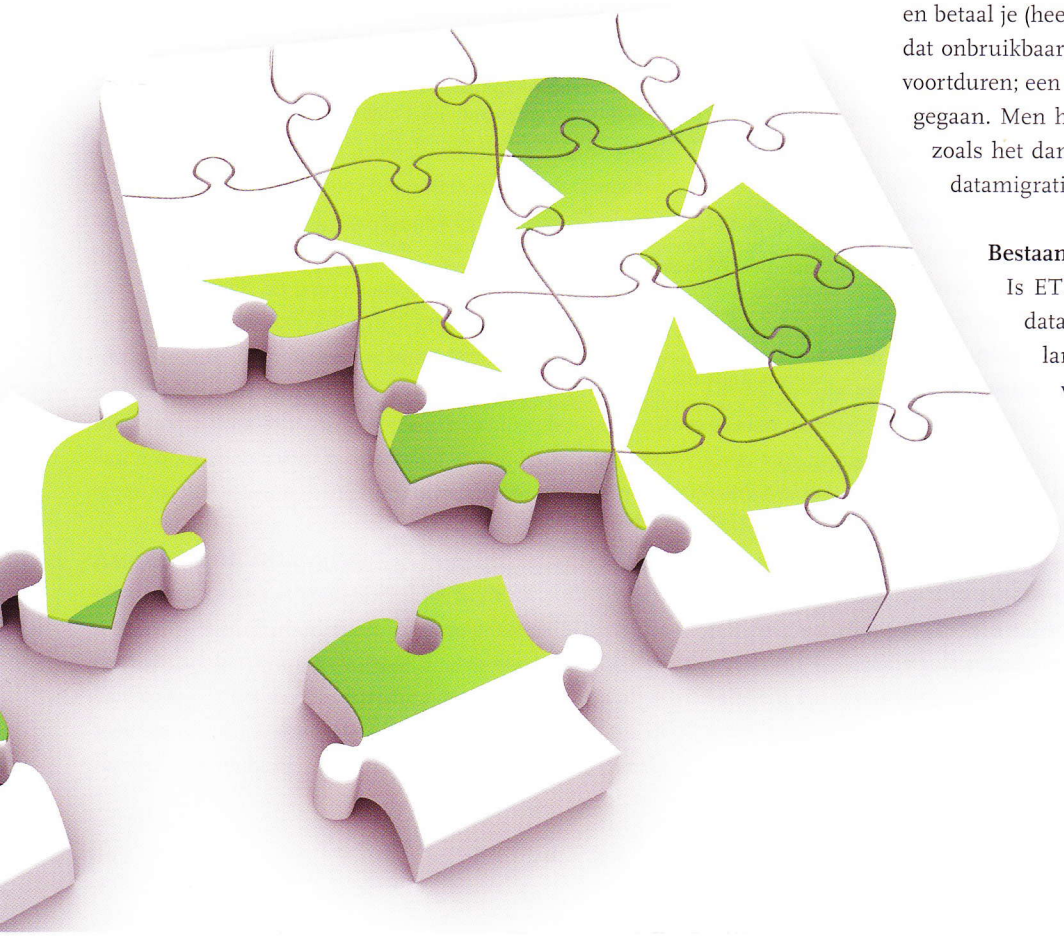
Datamigratie hoeft niet lastig te zijn

De invoer van een nieuw systeem verloopt lang niet altijd pijnloos. Dat ligt vaak niet zozeer aan het systeem zelf, maar meer aan het feit dat – wellicht door gebrek aan kennis – te makkelijk wordt gedacht over de implementatie. Symptomatisch is de manier waarop wordt omgegaan met een onderliggend, maar cruciaal proces: datamigratie. Het blijkt al snel niet zo eenvoudig als gedacht om de data uit het oude systeem te halen en geschikt te maken voor het nieuwe systeem. Datamigratie is vaak een langdurig, kostbaar en moeilijk beheersbaar proces met een aanzienlijke kans op mislukking. En zonder goede data is elk systeem onbruikbaar. Kan dat niet anders? Jawel. Een tool met functies voor ETL– Extraction, Transformation & Load - kan het proces aanzienlijk versimpelen.

door: Dave den Hartog en Jeroen Beltman



Organisaties stappen over naar een nieuw systeem als het oude systeem niet meer voldoet. Het is niet langer in staat de bedrijfsprocessen te ondersteunen, is moeilijk beheersbaar of er is simpelweg iets mooiers op de markt gekomen. Vaak gaat het om legacy systemen, waaraan in de loop der jaren allerlei subsystemen zijn gehangen. Er zijn dus diverse bronnen in verschillende formaten, die allemaal ingevoerd moeten worden in het nieuwe systeem.

Voor datamigratie bestaat veelal geen plan, geen tool, geen werkomgeving en geen beheer. De aandacht gaat vooral uit naar de selectie van het nieuwe systeem, terwijl de complexiteit van de datamigratie onvoldoende wordt onderkend. De migratie moet uiteindelijk met veel improvisatie en handwerk worden uitgevoerd.

En het is nogal wat, als je het hele datamigratieproces handmatig moet doorlopen. Bronbestanden moeten ingevoerd worden in Excel, de – soms wel miljoenen – records omgezet naar een formaat waarmee het nieuwe systeem overweg kan en alles moet geschift worden op relevantie en redundantie. Vervolgens moet het nieuwe systeem het resultaat nog accepteren, wat niet bepaald gegarandeerd is. Als het systeem het bestand weigert, kan dat liggen aan één enkel record; daar moet je dan dus naar op zoek. Event logging is meestal niet beschikbaar, dus bijhouden wat er is veranderd en waar,kan niet. Het is onmogelijk om alle overbodige informatie te schiften, waardoor de kans groot is dat het nieuwe systeem wordt gevuld met een vervuild bestand. Bovendien is het proces onbeheersbaar omdat er geen rapportagemogelijkheid is. De bedrijfsvoering heeft geen handvatten om het proces te monitoren en mee te beslissen.

Het gevolg is dat je aan de ene kant een nieuw systeem hebt en aan de andere kant een onwillig pak data en daartussen een diepe kloof van onbegrip. Ondertussen houdt het proces de hele ict-afdeling af van het dagelijkse werk en betaal je (heel stevige) licentiekosten voor een systeem dat onbruikbaar is. Het is een situatie die jarenlang kan voortduren; een situatie waaraan bedrijven ten onder zijn gegaan. Men heeft zich 'verslikt in de implementatie', zoals het dan heet. In feite is de complexiteit van de datamigratie onderschat.

Bestaande techniek, nieuwe toepassing

Is ETL iets nieuws, speciaal ontwikkeld voor datamigratie? Nee. De methode wordt al langer toegepast bij het creëren en vullen van datawarehouses, als tool om data uit verschillende bronnen te onttrekken, te keuren, schoon te maken en geschikt te maken voor gebruik in het datawarehouse. Hiermee wordt data

uit de bronsystemen omgevormd tot rapporteerbare data. Vaak wordt er een rapportagetool aan gekoppeld. Niet alleen om de data voor rapportagedoeleinden te kunnen gebruiken, maar ook om een vertaalslag te maken naar de bedrijfsvoering en het ETL-proces beheersbaar te houden. Het moge duidelijk zijn waarom ETL niet alleen bij datawarehousing, maar ook bij datamigratie bruikbaar is: beide toepassingen lijken nogal op elkaar. ETL kan daarom ook de datamigratie een stuk effectiever laten verlopen. De kans op fouten wordt kleiner en het proces wordt gemakkelijker en sneller. Bovendien wordt de migratie betaalbaarder: om ETL toe te passen is weliswaar in veel gevallen expertise van buiten nodig, maar de kosten die daarmee zijn gemoeid, vallen in het niet bij de financiële impact van een langdurig handmatig datamigratieproces. Veel bedrijven hebben ETL bovendien al op de plank liggen, omdat er al eens een datawarehouse is ingevoerd. Dat de tool nog een keer kan worden gebruikt is qua *return on investment* natuurlijk erg prettig. En de kans is groot dat er al ETL-kennis in huis is.

Hoe werkt het?

Het ETL-proces bestaat uit drie fases: Staging & Reject Handling, Transformation & Validation en Loading (zie *figuur 1*).

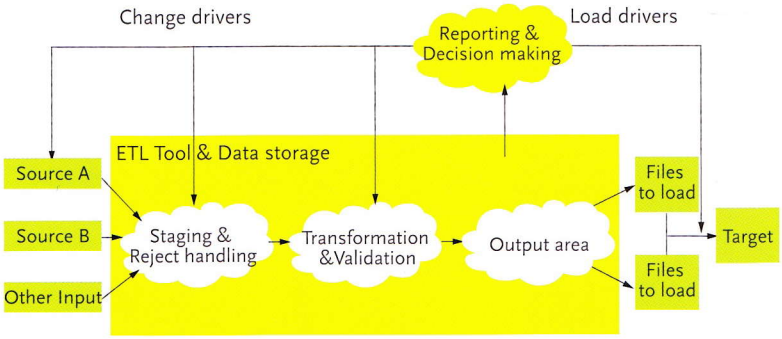
Staging & Reject Handling. In deze eerste fase wordt data aan de aanwezige bronnen onttrokken en opgeslagen in een database. Alles wordt, met andere woorden, bij elkaar gezet. Op basis van parameters (*business rules*) wordt vastgesteld welke data niet voldoet; de geweigerde data wordt vervolgens geïsoleerd. Aangezien de reject handling in deze fase al gebeurt, hoeft de onbruikbare data niet door het hele proces geleid te worden. Dat scheelt tijd, aangezien het kan gaan om miljoenen records. De bedrijfsleiding kan met de reporting-mogelijkheden ondertussen precies bijhouden wat is besloten en waarom, en direct invloed uitoefenen op de business rules. Doordat die vertaalslag naar de business wordt gemaakt, kan worden bewaakt dat de juiste beslissingen om de juiste redenen worden genomen.

Transformation & Validation. Zoals gezegd gaat bij handmatig transformeren van alles mis. De kans op fouten bij transformation & validation via een ETL-tool is veel kleiner. Elke business rule die wordt doorgevoerd, wordt gelogd en als er ergens iets fout gaat, wordt meteen gerapporteerd wat en waar. Bovendien kan op verschillende plaatsen in het proces ingegrepen worden. Wijzigingen in de bron worden automatisch doorgevoerd in de transformatie- of loadfase en vice versa. Je kunt teruggaan naar een vorige versie van de output als de nieuwe problemen oplevert en bij elke wijziging is meteen zichtbaar wat de gevolgen zijn voor het acceptatiepercentage in het nieuwe systeem. Juist in deze fase zorgt ETL ervoor dat de kans van slagen van datamigratie drastisch toeneemt.

Loading. In deze fase wordt de bewerkte data aangeboden aan het laadprogramma, dat er dan voor zorgt dat het geheel wordt ingevoerd in het nieuwe systeem. Soms treedt ook hier een fout op; de rapporteertool zorgt ervoor dat precies duidelijk wordt wat fout is gegaan en waar. Het probleem kan dan in een van de vorige fases verholpen worden.

BEZINT EER GE BEGINT...

De overstap naar een nieuw systeem is niet zelden een prestigekwestie. Het klinkt nu eenmaal goed om te kunnen zeggen dat je bedrijf draait op Hyperion, SAP of IBM. Bedrijven zijn echter dikwijls niet in staat de impact en implicaties van de implementatie te overzien – met in potentie dramatische gevolgen. Het heeft geen zin het nieuwe systeem te proberen aan te passen aan de oude data; de oude data moet geschikt gemaakt worden voor het nieuwe systeem. Een taak die serieus genomen moet worden. Bovendien: een omgeving als SAP omvat ook voorschriften voor de inrichting van bedrijfsprocessen. De fabrikant heeft daarover nagedacht en maakt gebruik van de input van duizenden gebruikers – en die hebben het niet allemaal mis. Het is daarom contraproductief om te proberen de bestaande bedrijfsprocessen op te leggen aan het nieuwe systeem. Het maakt de kans kleiner dat de organisatie echt baat heeft bij de investering. Feitelijk betekent de invoer van een nieuw systeem dus dat het hele business process management op de schop moet.
Dit alles betekent dat zowel de datamigratie als de implementatie van het nieuwe systeem op het juiste niveau belegd moet worden: aan de top. Het is geen klusje dat alleen aan ict'ers overgelaten kan worden, want de manier waarop met data en systeem wordt omgegaan, heeft verstrekkende gevolgen voor de hele organisatie.
In dit licht bezien is de toepassing van ETL wellicht niet meer dan symptoombestrijding. ETL kan helpen het datamigratieproces in goede banen te leiden, maar de implementatie van een nieuw systeem lukt alleen als de voorwaarden daarvoor en gevolgen daarvan op juiste waarde worden geschat.



Figuur 1: Het ETL-proces

Conclusie

Bij de aanschaf van een nieuw systeem is één gedachte cruciaal: alles begint met data. De data, dat is de informatie waarop het bedrijf wordt gerund. De geschiktheid van de data bepaalt of het nieuwe systeem de investering gaat terugbetalen. Alles hangt daarom af van de datamigratie: als die mislukt is het nieuwe systeem niets waard.
Dit artikel breekt een lans voor de toepassing van een ETL-tool om de migratie in goede banen te leiden, maar bevat ook een waarschuwing: staar je niet blind op de verworvenheden van het nieuwe systeem. Begin bij de basis en zorg ervoor dat je beseft wat de implementatie van een nieuw systeem behelst.

Over de auteurs: Dave den Hartog en Jeroen Beltman zijn werkzaam bij Ensior