



Risicoanalyse bij DWH-testen

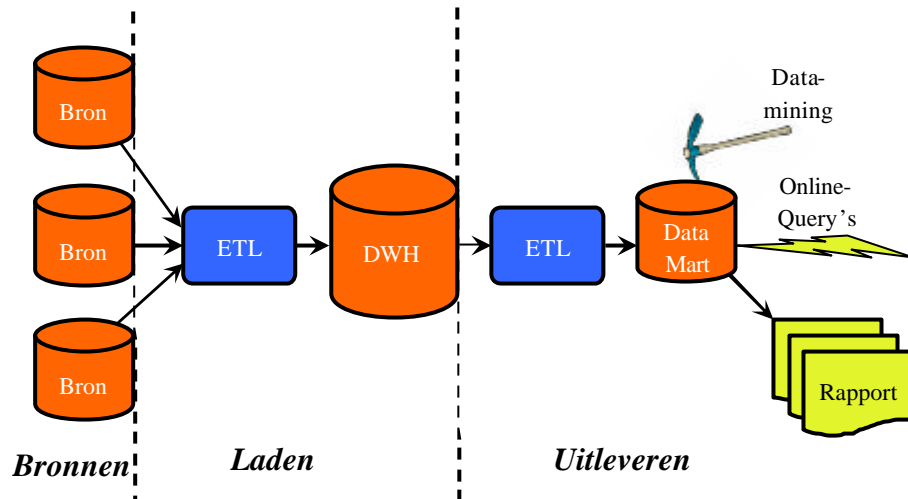
Cees Dulfer, Rabobank Nederland

Agenda

- Inleiding DHW
- Knelpunten
- Opzet PRA
- Testaanpak
- Vragen



Data Ware House



Datakwaliteit

Kwaliteit van het DWH wordt bepaald door de kwaliteit van:

- de brongegevens
- het ETL-proces voor het laden van het DWH
- het ETL-proces voor het vullen van de Datamart (s)
- de standaard rapportages
- de ter beschikking gestelde queries
- De data mining faciliteiten.



Product Risico Analyse

- Brongegevens:
Verantwoordelijkheid leverancier
- Uitleveren:
Afnemer(s) zijn bekend dus kun je de gebruikelijke Product Risico Analyse uitvoeren
- Laden:
Er is niet bekend:
 - Wie (toekomstige) afnemers zijn
 - Wat het (toekomstig) gebruik is *Hoe doe je dan een Product Risico Analyse?*



Overwegingen

- Toekomstig gebruik is onbekend dus uitgaan van de hoogste kwaliteitseis?
- Eisen aan de verwerking (performance, foutafhandeling, controles) zijn altijd gelijk
- Laadproces is in principe altijd hetzelfde: het bronbestand overzetten naar 1 of meer doelbestanden
- Conclusie:
Source to Target Mapping is enige variabele



Aanpak Product Risico Analyse

- Generieke risico's:
 - Voor alle bronnen gelijk
 - Meestal gerelateerd aan beheereisen
 - ✍ Eénmalig opzetten met stakeholder beheer
- Bron afhankelijke risico's:
Risico wordt bepaald door
 - Complexiteit van de verwerking
 - Belang van de attributen
 - ✍ Standaard aanpak opzetten



Standaard aanpak Risicoanalyse

Door de modelleur (Informatie Analist) op grond van zijn kennis, kunde en ervaring het risico ingeschat voor de verwerking van de bronbestanden

De volgende aspecten spelen hierbij een rol:

- De complexiteit van de verwerkingslogica (Complexiteit, Complexity).
Dit gegeven is vergelijkbaar met het begrip "**Kans**".
- Het belang dat een data attribuut uit het bronbestand heeft (Belang, Importance).
Dit gegeven is vergelijkbaar met het begrip "**Schade**".

Vaststellen Complexiteit

Per scenario wordt de Complexiteit bepaald:

Complexiteit	Criterium
1 Zeer eenvoudig	Overnemen van een record met standaardwaardes (defaults)
2 Eenvoudig	Vertaling naar codering, ander formaat of dmv formule óf foutafhandeling
3 Niet zo eenvoudig	Verwerking keuze (paden) óf levenscyclus functie
4 Complex	2-3 verschillende verwerkingen zoals onder complexiteit 2 of 3 beschreven
5 Bijzonder complex	Meer dan 3 verschillende verwerkingen zoals onder complexiteit 2 of 3 beschreven

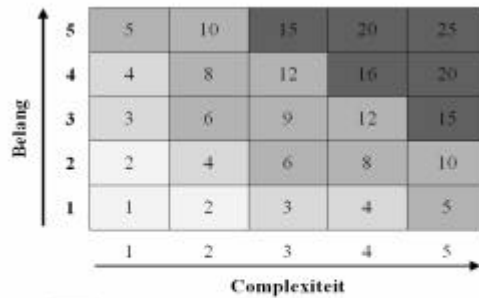
N.B.: Een bronbestand wordt verwerkt volgens meerdere scenario's, afhankelijk van het aantal doeltabellen.

Vaststellen Belang

Per attribuut wordt het belang bepaald:

Belang	Criterium
0	Totaal onbelangrijk, het gegeven wordt niet opgenomen
1	Wordt opgenomen, maar wie het ooit ergens voor gaat gebruiken is nog onbekend
2	Is voor één partij een informatief gegeven om bedrijfsprocessen te volgen of te sturen
3	Voor meerdere partijen is dit een informatief gegeven om bedrijfsprocessen te volgen of te sturen
4	Is voor één partij een kerngegeven om bedrijfsprocessen te volgen of te sturen
5	Voor meerdere partijen is dit een kerngegeven om bedrijfsprocessen te volgen of te sturen

Vaststellen Risico/Prioriteit



- M Must (15, 16, 20 en 25)
- S Should (5, 8, 9, 10 en 12)
- C Could (3, 4 en 6)
- W Would/Won't (1 en 2)

Testaanpak

De volgende typen testaanpak worden onderkend, waarbij ieder hoger level alle voorgaande levels omvat:

Type	Testaanpak
A	1 goed testrecord [1] en outputvoorspelling
B	1 goed testgeval
C	1 goed testgeval en visuele controle testresultaat
D	1 goed testgeval en 1 error testgeval en deltarun en visuele controle testresultaat
E	1 goed testgeval en 1 error testgeval, steekproef controle/visuele controle testresultaat
F	1 goed testgeval en 1 error testgeval en deltarun met outputvoorspelling
G	1 goed testgeval en 1 error testgeval, filters, outputvoorspelling
H	1 goed testgeval en 2e run, visuele controle testresultaat
I	1 goed testgeval en 2e run en 1 error testgeval, filters, outputvoorspelling

[1]) 1 goed testrecord die alle (zo mogelijk) scenario's raakt

Strategiematrix

In onderstaande tabel is aangegeven welke testaanpak wordt gevolgd afhankelijk van de typering van de tabellen en de prioriteit:

<i>Typing tabellen</i>	<i>W</i>	<i>C</i>	<i>S</i>	<i>M</i>
Master-tabellen	A	C	E	F
Relatie-tabellen	A	B	C	D
Basis -tabellen	A	C	E	F
Profile -tabellen	A	C	E	G
Summary -tabellen	A	B	C	G
Relatiedetail-tabellen	A	B	C	D
Nondelta -tabellen	A	B	C	G
Sequence -tabellen	A	B	H	I

Controle query's (1)

Queries welke gebruikt worden voor de analyse van de verwerkte data:

- Totalen van afgewezen records totalen rejects
- Totalen verwerkte records uit bronbestanden
- Totalen bedrag velden
- Totalen per maand
- Totalen van gevulde nullable velden.
- Volledigheid van relaties.

Controle query's (2)

- Datumcontroles
 - Er wordt met name gecontroleerd of periodes
 - Op elkaar aansluiten
 - Elkaar niet overlappen
 - Niet eindigen voordat ze begonnen zijn

Vragen ?