



Risk Analyses for DWH-testing

Cees Dulfer, Rabobank Nederland

Agenda



- Rabobank
- Product Risk Analyses
- Data Ware Houses
- Bottlenecks
- Product Risk Analyses DWH
- Test Approach
- Questions

Rabobank Group essentials



- Cooperative banking model
 - Originated in 1896
 - 183 affiliated banks
 - 2.765 Service points and 341 Cash points
 - Affiliates for leasing, asset management, insurance etc.
 - Worldwide operations (60.000+ employees)
 - 9.000.000 customers, 1.700.000 members
 - Net income EUR 5.679 million
- Largest Internet Bank in Europe
 - www.rabobank.be
- Only private bank with an AAA rating

Rabobank Group Strategy



- To remain the largest, best and most innovative all-finance service provider in the Netherlands
- To become the global leader in financing in the food & agri market
- To be the most sustainable, socially involved bank

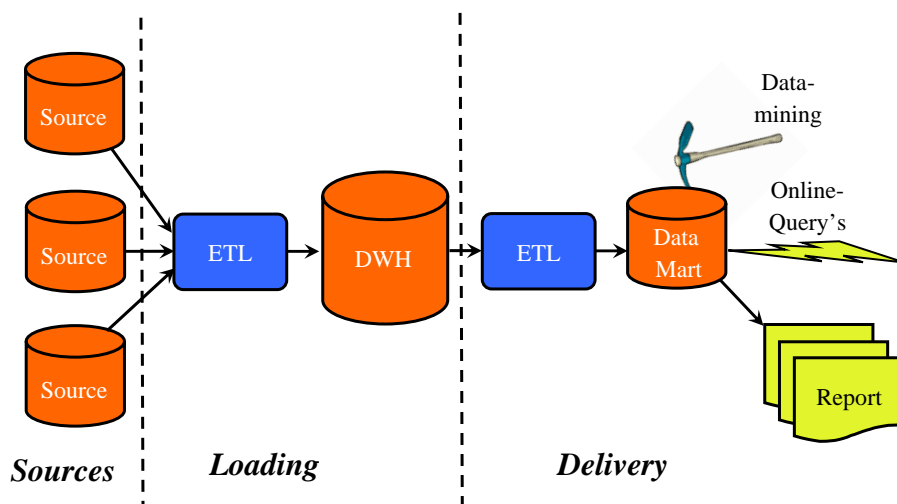
Test Center Strategy

- To be the best test centre in the European financial community in 2010

Product Risk Analyses

- Involve the stakeholders
- Elaboration of Business risks into user and system risks
- Supply different views of the system (requirements, acceptance criteria, risks)
- Improve the completeness of the requirements by matching these views
- Prioritise Risk (= Impact * Chance)
- Establish a base for the test strategy

Data Ware House





Data quality

The quality of a DWH is determined by the quality of:

- source data
- ETL-process to load the DWH
- ETL-process to load the Datamart(s)
- standard reports
- standard query's
- data mining facilities.



Product Risk Analyses DWH

- Source data:
responsibility of the supplier
- Delivery:
Customer(s) are known so a regular PRA can be performed
- DWH Loading:
It is not (yet) known:
 - Whether the data will ever be used
 - What will be the (future) use
 - Who will be the (future) customers

– *How to do a PRA without stakeholders?*



Considerations

- Future use is unknown, should we assume the highest quality requirements?
- Operational requirements (performance, error handling, controls) are always the same
- Loading process is always identical: a source file is transferred to 1 or more target tables
- Source to Target Mapping is the only variable



Approach Product Risk Analyses

- **Generic risks:**
 - For all sources identical
 - Usually related to operational requirements
 - ➔ Once only determination with operations staff
- **Source dependent risks:**
 - Define a risk driven standard approach
 - Risk is defined by:
 - Complexity of the processing
 - Importance of the data items



Standard approach Risk Analyses

Based on his knowledge, skill and experience the modeller (Information Analyst) estimates the risks related to the processing of the source files

Aspects that play a role in this estimation are:

- The Complexity of the processing
Complexity can be compared with “**Chance, Probability**”.
- The Importance of a data items in the source file.
Importance can be compared with “**Impact, damage**”.



Complexity

Per scenario the Complexity is determined:

Complexity	Criterion
1 Very simple	1:1 transfer with default values
2 Simple	Rule based translation to a code or a different format
3 Moderate	Processing of decision path or life cycle function
4 Complex	2-3 different processing modes with complexity 2 or 3
5 Very complex	More than 3 different processing modes with complexity 2 or 3

NB.: A source file is processed according to 1 or more scenarios depending on the number of target tables



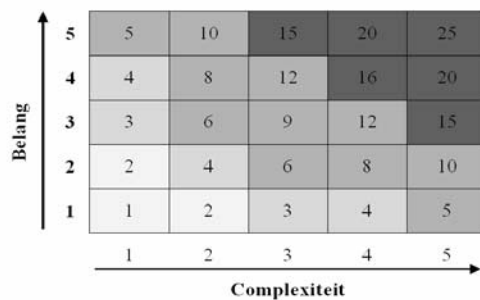
Importance

For every data item the importance is determined:

Importance	Criterion
0	The item is not include in the target file
1	The item is included but whether, when and by whom is not yet known
2	The item is for one customer informative to follow or direct the business processes
3	The item is for more than one customer informative to follow or direct the business processes
4	The item is for one customer crucial to follow or direct the business processes
5	The item is for more than one customer crucial to follow or direct the business processes



Determine Priority



- 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)



Test types

The following kind of test types are recognized. Each level comprises all previous levels:

Type	Test type
A	1 correct test record with output prediction
B	1 correct test case
C	1 correct test case plus 1 error test case with visual inspection of the result
D	1 correct test case plus 1 error test case plus a delta run with visual inspection of the result
E	1 correct test case plus 1 error test case with random check and visual inspection of the result
F	1 correct test case plus 1 error test case plus a delta run with output prediction
G	1 correct test case plus 1 error test case plus filters with output prediction
H	1 correct test case plus second run with output prediction
I	1 correct test case plus second run plus error test case with output prediction



Strategy matrix

Depending on the type of the target table and the priority a test type is chosen:

Type of tables	W	C	S	M
Master-tables	A	C	E	F
Relation-tables	A	B	C	D
Base-tables	A	C	E	F
Profile-tables	A	C	E	G
Summary-tables	A	B	C	G
Relation-detail-tables	A	B	C	D
Non-delta-tables	A	B	C	G
Sequence-tables	A	B	H	I



Control query's (1)

Queries that are used to analyse the processed data:

- Number of input records
- Number of rejected records
- Number of target records per target table
- Totals of currency fields
- Totals per month
- Referential integrity.



Control query's (2)

- Date controls
Specific checks whether time frames:
 - Link up
 - Don't overlap
 - Don't end before they are started



Questions?