



# 8 Building the DW

## Solutions

- Exercise I
  - What is a staging area?
  - Why are users not allowed to interact with the staging area?
  - When should we use flat files in the staging area?
  - What is ETL, and when should it be applied?



# 8 Building the DW

## Solutions

- Exercise 2

- a) Compose a regular expression which validates only strings which contain letters only, no spaces or other characters than letters, start with only one capital letter, and continue with at least one, up to 20 small letters. See examples in Annex 2.

- `[A-Z][a-z]{1,20}`

String	Evaluation
A	Bad
Aa	Good
AA	Bad
aa	Bad
A1	Very bad
Thomas	Good
Thomas Mann	Bad



# 8 Building the DW

# Solutions

- Exercise 2
  - b) Install Eobjects Data Cleaner. Use the regular expression from 2.a, and create a validation task, add as validation rule a “regex validation”, choose as data selections the CUSTOMER table, and as data subset the CONTACTLASTNAME and CONTACTFIRSTNAME attributes

Regex validation

Type: Regex validation

10 rows did not validate

CONTACTLASTNAME	CONTACTFIRSTNAME
de Castro	Isabel
Rancé	Martine
Fresnière	Jean
Da Cunha	Daniel
O'Hara	Anna
McKenna	Patricia
MacKinlay	Wales
Müller	Rita
McRoy	Sarah
Altagar, G M	Raanan

Regex validation

Type: Regex validation

4 rows did not validate

CONTACTFIRSTNAME	CONTACTLASTNAME
Frédérique	Citeaux
Wing C	Tam
Martín	Sommer
José Pedro	Roel



# 8 Building the DW

# Solutions

- Exercise 2
  - c) Give three examples (of different patterns) strings which pass the validation of the following regular expression, and one that doesn't:

- $(\d{1,2})?((\d{1,4})|(\d){3,5}[-/]?)((\d){1,5})$

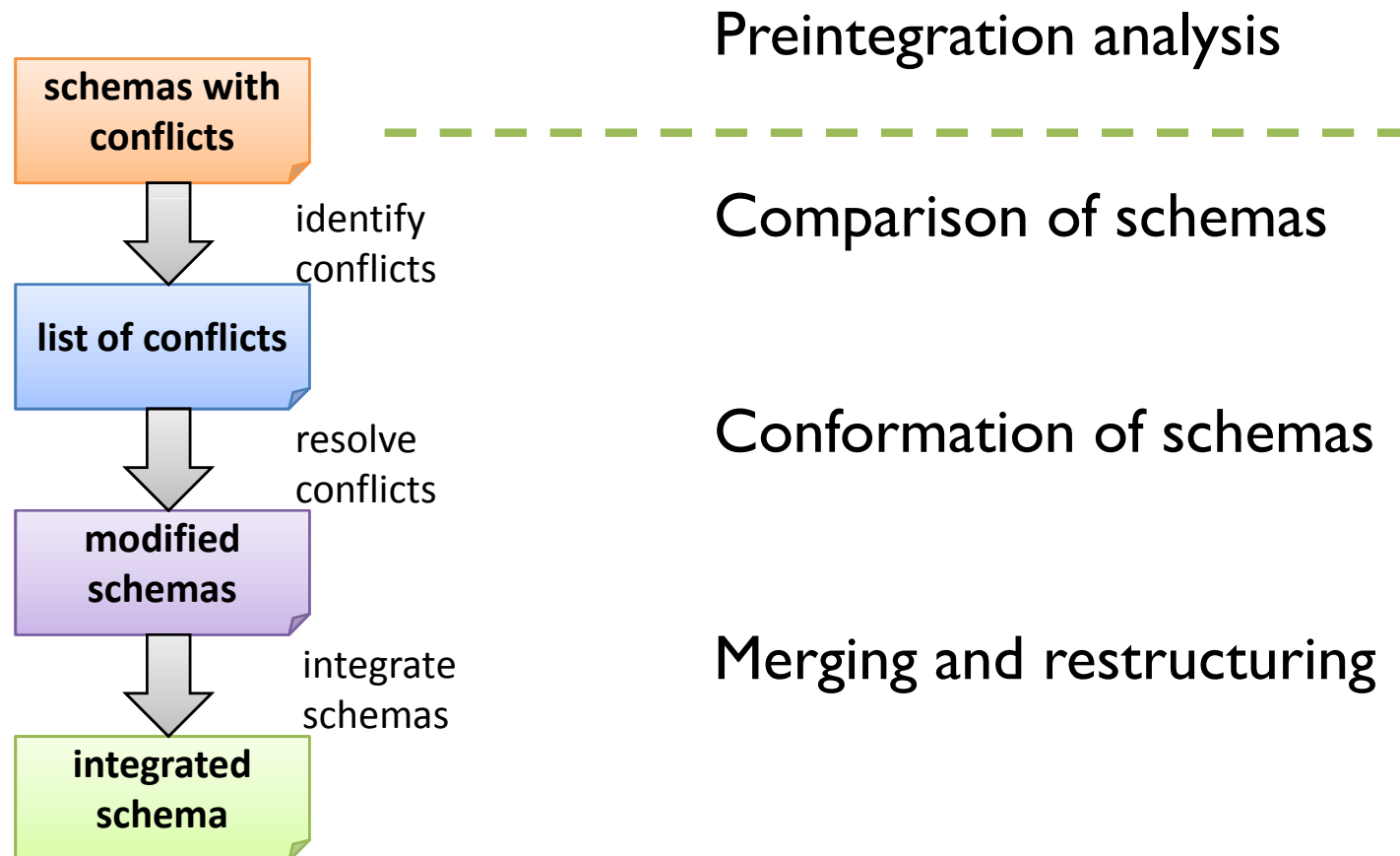


# 8 Building the DW

## Solutions

- Exercise 3

- Briefly describe the basic steps in schema integration?





# 8 Building the DW

## Solutions

- Exercise 3
  - Explain how schema mapping is performed in praxis.
    - Transform data to application specific XML → Transform to XML specific to other application / general schema → Transform back to binary
    - Note: the integration work still has to be done **manually**



## 8 Building the DW

## Solutions

- Exercise 3
  - When should we use bulk loading and why is it a good idea?
    - Loading fact table data
    - It bypasses integrity checks, concurrency, index maintenance, etc. so it is much faster than regular inserts