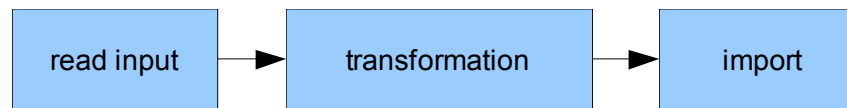


ETL Process for Individual Case Data

- what is individual case data?
 - opposite: aggregated (summed) data
 - one record == one line in the data
 - sums are calculated on the fly by „counting lines“
 - examples:
 - students
 - exams
 - personnel

ETL

- Extraction
- Transformation
- Loading



Extraction

- reading records from text files
- fixed field widths
- every field can have different width
- indexes have to be given:

example file contents:

05082008

05111967

...

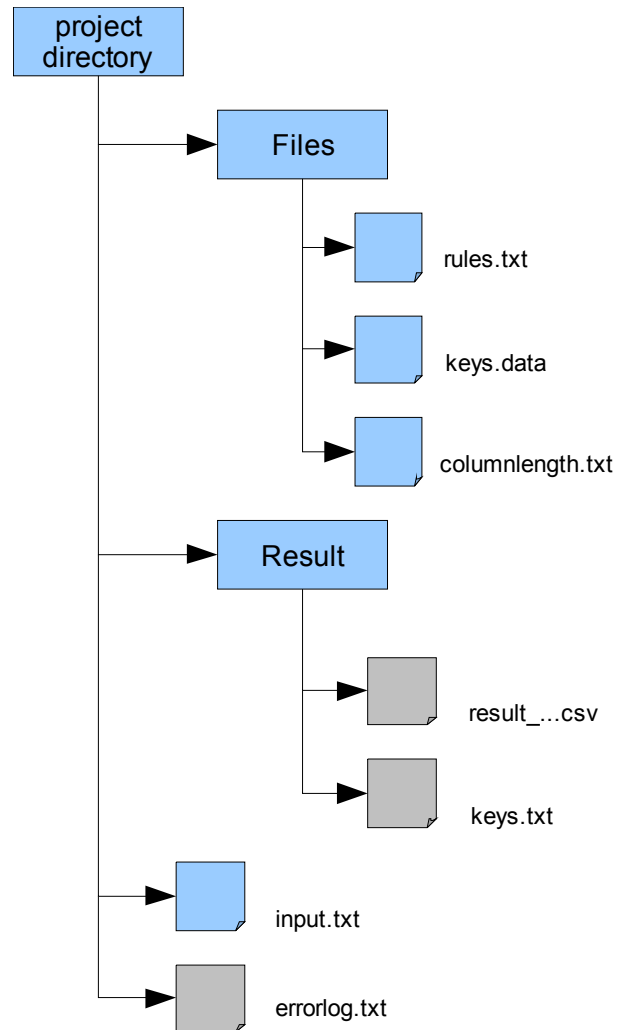
Transformation

- different data sources => different keys
- harmonisation of keys to make data comparable
- ICE key system
- transformation of keys
- creating new aggregates

Loading

- read transformed data
- load data into database table
- update metadata in database

Creating An Import Project



Creating An Import Project

Analysing Input Data

- example data:

20031500101

- 1st Field: 1-4 (2003) -> Year
- 2nd Field: 5 (1) -> Passed Exam
- 3rd Field: 6-8 (500) -> Course
- 4th Field: 9-10 (1) -> #Females
- 5th Field: 10-11 (0) -> #Males
- 6th Field: 12-13 (1) -> #sum

Writing the Rule File

- Create new stock in database
- Define identifier (for input file)
- Select characteristics for import
- Does system contain all keys?
- Define field indexes in rules.txt
- Define metadata line (JDBC connection...)
- Select rules

Testing and Refining

1. Start Transformation
2. Analyse errorlog.txt, schluessel.txt, result...
3. Repeat 1. & 2. until no more errors

Import into the Database

- Check virtual stock: New point in time exists?
- Generate example tables
- Compare with previous points in time