ETL

jingmi@gmail.com
2010.08.20

Principle

- Simple

- Trust nothing

- One-pass

- Less Disk Access(Input/Output)

- Log

Simple Data Structure

- List

- Hash

Validate(1)

- Validate input(formation)

Validate(2)

- Group Value(find new/illegal/unknown options)

One-Pass Parser

• One-Pass Compiler?

Index

- Key --> Offset

- Key --> line no

Split/Map

- Large DataSet

- Muti-files

- When

Validate(3)

- ASCII

ID generating

- Generate ID from specified number

- Generate ID for foreign key