

Sistemas de Apoio à Decisão

Transformação num DW

Transformação

- Objetivos
 - Garantir qualidade de dados
- Critérios de qualidade de dados
 - Clareza, Consistência, Correção, Completude
- Pontos de garantia de qualidade
 - E->TL , ET->L

Transformação

- Critérios de qualidade de dados (exemplos)
 - Dados são CLAROS se não há dúvidas sobre o seu significado

Dados *não claros*

Locais

Id	Nome
1	New York
2	New York
3	Lisboa
4	Lisboa

Dados **CLAROS**

Locais

Id	Nome	Tipo
1	New York	Estado
2	New York	Cidade
3	Lisboa	Cidade
3	Lisboa	Distrito

Transformação

- Critérios de qualidade de dados (exemplos)
 - Dados são CONSISTENTES se são representados sempre da mesma forma

Fontes
(dados *não* CONSISTENTES)

Cientes

Id	NIF	País
192	210116324	USA
119	211075757	PT
186	211009929	PT
100	211009929	ES

Cientes

Id	NIF	País
192	210116324	EUA
119	211075757	POR
208	207446644	POR
15	208747475	SPA

DW
(dados CONSISTENTES)

Cientes

Id	NIF	País
192	210116324	USA
119	211075757	PT
186	211009929	PT
100	211009929	ES
208	207446644	PT
15	208747475	ES

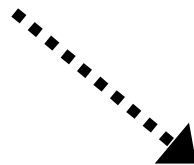
Transformação

- Critérios de qualidade de dados (exemplos)
 - Dados são CORRETOS se o seu conteúdo respeita uma versão oficial (atenção: correto \neq verdadeiro)

Cientes

Id	dataEntrada	Morada
2000	12-Jun-2014	Rua Filipe Andrade
2001	12-Jun-2014	Avenida Maria Faria Filhos
2003	15-Jun-2014	Estrada Conto do Vigário
2006	15-Jun-2014	Alto do Vieiro

Fontes
(dados *não* CORRETOS)



DW
(dados CORRETOS)

Cientes

Id	dataEntrada	Morada
2000	12-Jun-2014	Rua Filipe de Andrade
2001	12-Jun-2014	Avenida Maria Faria e Filhos
2003	15-Jun-2014	Estrada C anto do Vigário
2006	15-Jun-2014	Alto do Vieiro

Transformação

- Critérios de qualidade de dados (exemplos)
 - Dados são COMPLETOS se existem quando é esperado que existam

DW

(dados *não* completos? Por semana há habitualmente 7200 entregas)

Fontes

...



T_DATA_ENTREGAS

Id	data	codPostal
771616252	12-Jun-2014	3080-122
771616253	12-Jun-2014	3090-881
771616254	15-Jun-2014	2400-100
771616255	15-Jun-2014	2400-102

Transformação

- Critérios de qualidade de dados (exemplos)
 - Dados são COMPLETOS se existem quando é esperado que existam

DW
(dados não completos)

Fontes

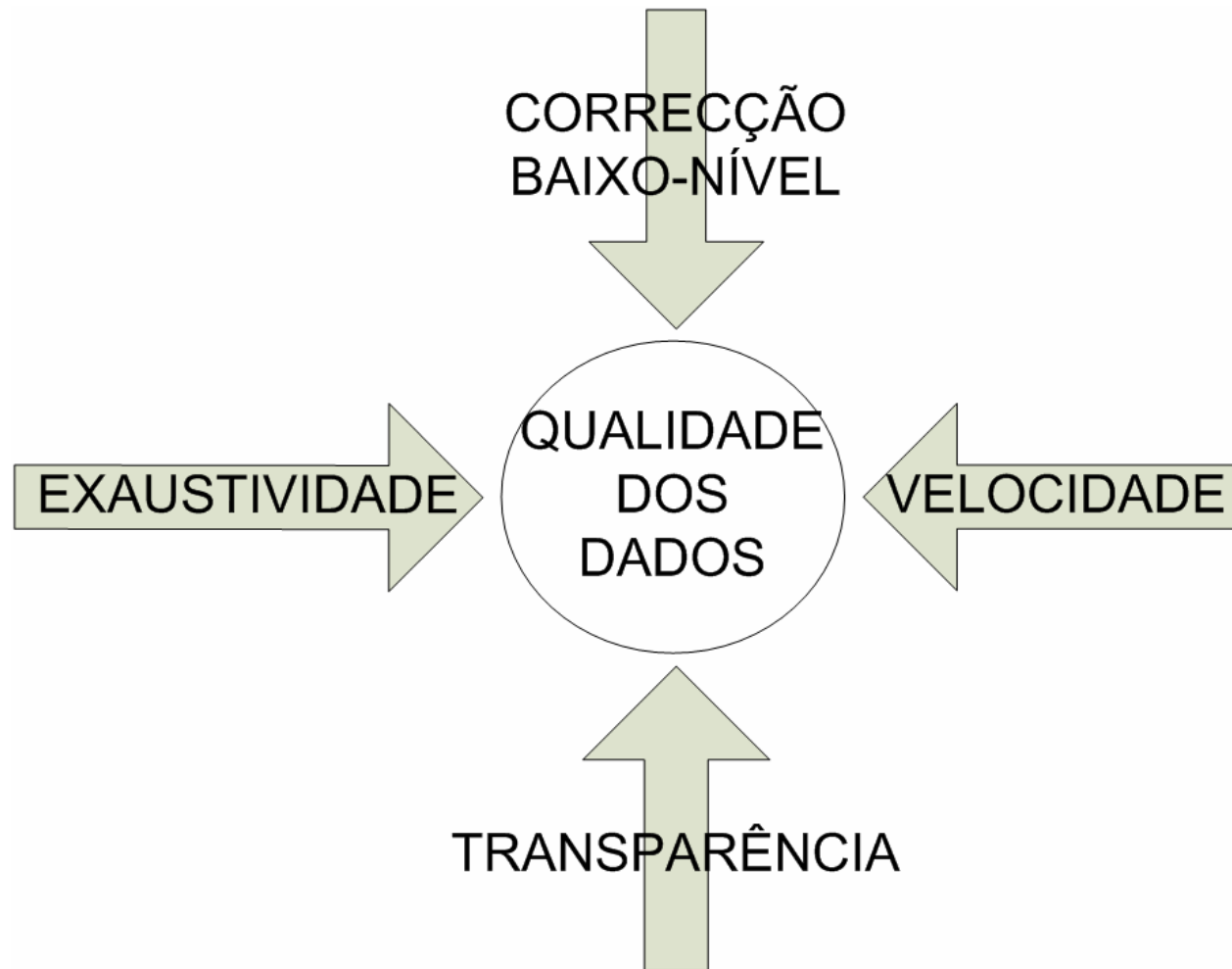
...



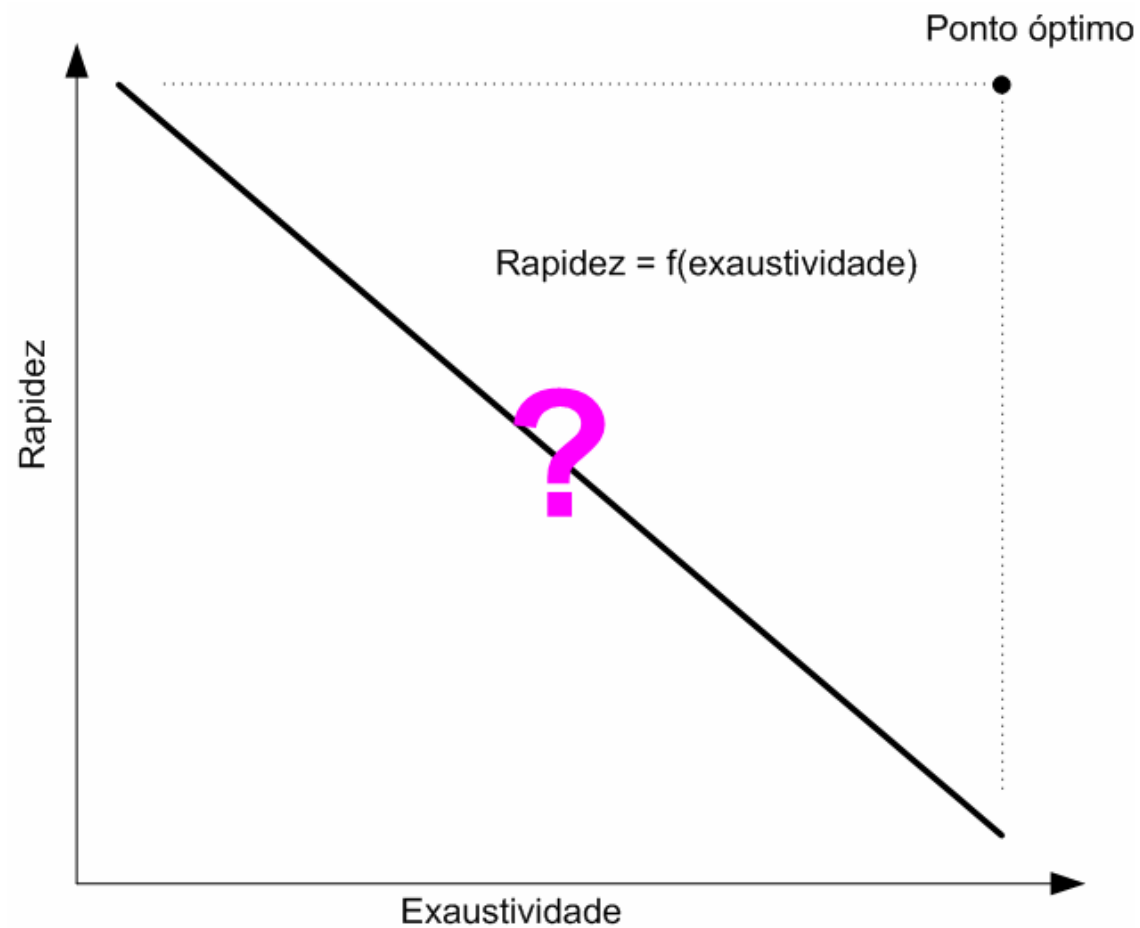
T_DATA_ENTREGAS

Id	data	codPostal
771616252	12-Jun-2014	3080-000
771616253	12-Jun-2014	3090-881
771616254	15-Jun-2014	2400
771616255	15-Jun-2014	
771616257	16-Jun-2014	3090-354

Conflitos & Prioridades



Conflitos & Prioridades



Gestão de falhas de qualidade

- Atitudes perante a falha
 - *Ignorar, abortar ou corrigir?*
 - Escolher atitude após uma **decisão informada**
 - Registrar exaustivamente todas as falhas detetadas

Gestão de falhas de qualidade

- Estrutura para registo de falhas
 - Dados não voláteis
 - Grande volume de dados
 - Alvo de análise estatística temporal e histórica

=

Sistema de Apoio à Decisão

Testes de qualidade dos dados

- Chamados “*screens*”
- Tipos de testes (exemplos)
 - Colunas estão preenchidas?
 - Os valores extraídos são os esperados?
 - O total de linhas extraído é o esperado?