

**DISCIPLINA: PROJETO DE SISTEMAS APLICADO AS MELHORES PRÁTICAS EM QUALIDADE DE SOFTWARE E GOVERNANÇA DE TI**

**AULA:**  
**23 – DATA QUALITY - EXERCICIOS**

**PROFESSOR:**  
**RENATO JARDIM PARDUCCI**

PROFRENATO.PARDUCCI@FIAP.COM.BR

[Renato Parducci - YouTube](#)

## EXERCÍCIOS RESOLVIDOS

## 1. Exercício resolvido

Escreva um procedimento de trabalho para:

Considerando que existem duas bases de dados diferentes contendo cadastros de produtos com Código de identificação, Número de série do fabricante, Nome do produto, Unidade de medida de movimentação...

Defina uma regra de comparação de registros que evite a duplicação de dados.

Considere que as duas bases de dados são distribuições verticais da base de dados de produtos – cada uma das duas filiais da empresa administra uma parte dos produtos.

*\*Solução no final deste material*

## 2. Exercício resolvido

Qual a causa do problema de qualidade reportado a seguir.

Nosso processo de carga de dados de movimentos de vendas de filiais ocorre de forma batch, durante a noite e a captação dura em torno de 3 horas sendo que ela precisa iniciar as 22h e concluir até 01h da manhã seguinte para não comprometer outras rotinas de processamento de dados que precisam rodar antes dos sistemas de produção online iniciarem seu uso as 6h da manhã.

O disparo da rotina é automatizado pelo horário mas, devido a demora nos fechamentos das posições de vendas das filiais, parte dos valores de movimentação diária não é capturado e a empresa tem que lidar com dados consolidados parcialmente atualizados no dia seguinte.

*\*Solução no final deste material*

### 3. Exercício resolvido

Associe corretamente o que gerou o problema de qualidade de dados reportado.

Segundo o dono da informação, custo médio de estoque deve ser calculado uma vez por mês, com base nos preços de compras praticados para os materiais que ainda estão no estoque, mas o programador está atualizando a média a cada chegada de material com base no valor unitário da última compra. Inicialmente, quando o sistema entrou no ar, os estoques foram carregados com o custo médio que constava no sistema legado e que estava correto.

- ☐ Faltou harmonização de conceitos
- ☐ Faltou limpeza de dados na carga inicial de cadastros de estoques

*\*Solução no final deste material*

## 4. Exercício resolvido

Associe corretamente o(s) parâmetro(s) de qualidade envolvido(s).

As formulações de medicamentos de uma indústria farmacêutica possuem um sistema de administração regulatória com o cadastro de elementos químicos e concentrações que é reportado para a ANVISA, porém, nas suas fábricas, existem formulações cadastradas nas máquinas de controle do chão de fábrica que não coincidem sempre com as formulações licenciadas na agência regulatória. O produto final é retrato da formulação que consta no chão de fábrica.

- ☐ Acurácia
- ☐ Completude
- ☐ Consistência
- ☐ Atualidade
- ☐ Validade
- ☐ Unicidade

*\*Solução no final deste material*

## 5. Exercício resolvido

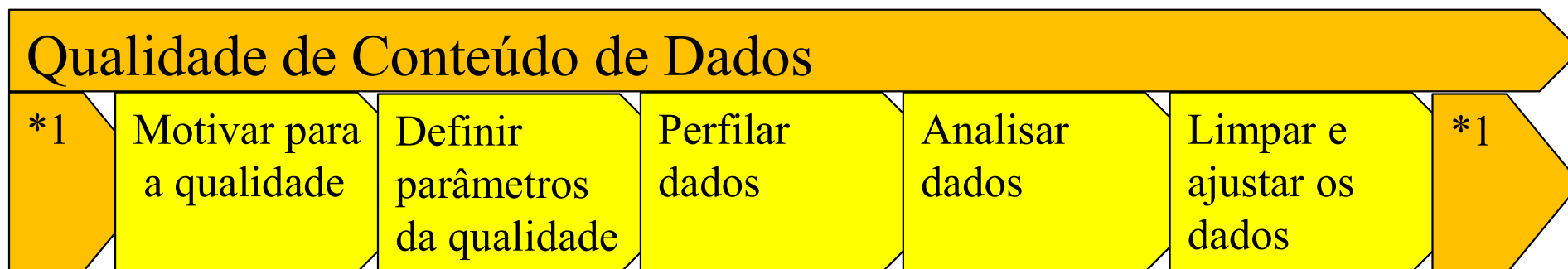
Para resolver problemas de definição de formulas de cálculo de comissões quando da fusão de duas empresas e suas áreas comerciais é necessário realizar...

- ☐ Harmonização de definições
- ☐ Cleansing na formação de um novo banco de dados com base nos legados

*\*Solução no final deste material*

## 6. Exercício resolvido

Faça uma associação correta o PDCA com os processos de qualidade de dados



*\*Solução no final deste material*



## EXERCÍCIOS PROPOSTOS

## 1. Exercício proposto

Escreva um procedimento de trabalho para:

Considerando que existem duas bases de dados diferentes contendo cadastros de clientes com RG, CPF,, nome, logradouro, número, CEP, sigla do estado, país, DDI, DDD e número do telefone de contato...

Defina uma regra de comparação de registros que evite a duplicação de dados.

Considere que as duas bases de dados são distribuições verticais da base de dados de clientes, sendo que um mesmo cliente pode constar nas duas réplicas mas não deveria constar duas vezes na mesma base de dados.

*\*Resolva e tire suas dúvidas com o professor*

## 2. Exercício proposto

Qual a causa do problema de qualidade reportado a seguir.

Nosso processo de carga de dados de movimentos de vendas de filiais ocorre de forma batch, durante a noite e a captação dura em torno de 3 horas sendo que ela precisa iniciar as 22h e concluir até 01h da manhã seguinte para não comprometer outras rotinas de processamento de dados que precisam rodar antes dos sistemas de produção online iniciarem seu uso as 6h da manhã.

O disparo da rotina é automatizado pelo horário mas, quando falta conexão de internet, o software não consegue capturar as posições de vendas das filiais sem conexão e a empresa tem que lidar com dados consolidados parcialmente atualizados no dia seguinte.

*\*Resolva e tire suas dúvidas com o professor*

### 3. Exercício proposto

Associe corretamente o que gerou o problema de qualidade de dados reportado.

Segundo o dono da informação, custo médio de estoque deve ser calculado uma vez por mês, com base nos preços de compras praticados para os materiais que ainda estão no estoque, mas o programador está atualizando a média a cada chegada de material com base no valor unitário da última compra. Inicialmente, quando o sistema entrou no ar, os estoques foram carregados com o custo médio que constava no sistema legado e que era calculado conforme a regra que o programador usou (a qual estava equivocada).

- ☐ Faltou harmonização de conceitos
- ☐ Faltou limpeza de dados na carga inicial de cadastros de estoques

*\*Resolva e tire suas dúvidas com o professor*

## 4. Exercício proposto

Associe corretamente o(s) parâmetro(s) de qualidade envolvido(s).

Estudantes não conseguiram postular para vagas em universidades federais em função do sistema único de avaliação com base no exame nacional do ensino médio não ter suportado a quantidade de transações e dados de simulação de vagas e notas por curso, instituição e candidato.

- |                                       |  |
|---------------------------------------|--|
| <input type="checkbox"/> Acurácia     | <input type="checkbox"/> Razoabilidade |
| <input type="checkbox"/> Completude   | <input type="checkbox"/> Privacidade   |
| <input type="checkbox"/> Consistência | <input type="checkbox"/> Atualidade    |
| <input type="checkbox"/> Atualidade   |  |
| <input type="checkbox"/> Validade     |  |
| <input type="checkbox"/> Unicidade    |  |

*\*Resolva e tire suas dúvidas com o professor*

## 5. Exercício proposto

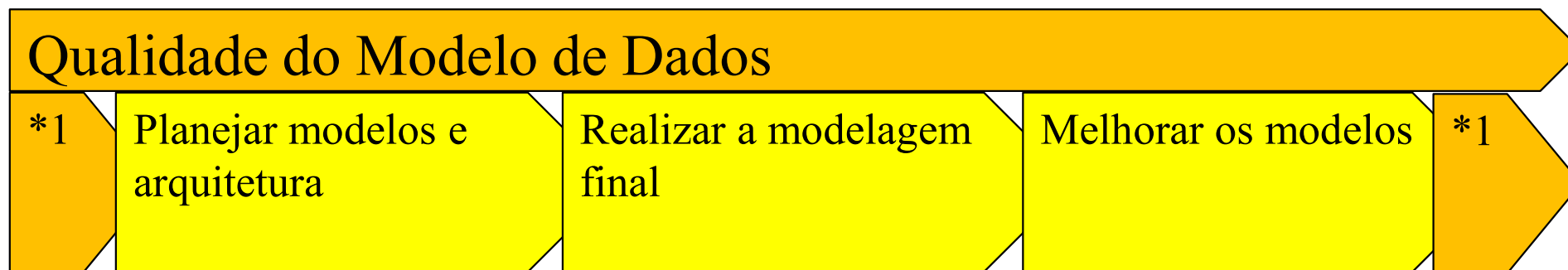
Quais, da lista a seguir, são resultados de Data Profiling?

- ☐ 93% dos clientes que compram esmaltes são mulheres
- ☐ A regional de vendas de Belo Horizonte é a mais lucrativa
- ☐ Existem 3.870 registros de contratos no sistema
- ☐ Ocorrem 59.301 acessos ao registro de cadastro da matriz por dia
- ☐ O campo de idade e sexo dos cadastros de clientes está em branco em 78% das linhas da tabela de cadastro
- ☐ 18% das pessoas que se interessam pelos nossos produtos são mulheres loiras altas e magras,
- ☐ 73% dos clientes que tornam a recomprar da empresa estão em Minas gerais, 17% em São Paulo e o restante em outros estados.

*\*Resolva e tire suas dúvidas com o professor*

## 6. Exercício proposto

Faça uma associação correta o PDCA com os processos de qualidade do modelo de dados



*\*Resolva e tire suas dúvidas com o professor*

## 7. Exercício proposto

Crie uma tabela usando a sua criatividade e instancie registros com problemas de qualidade (dados incompletos, dados duplicados, formatação despadronizada).

Utilize o DataCleaner para perfilar os dados, encontrando os problemas.

*\*Resolva e tire suas dúvidas com o professor*



## 8. Exercício proposto

Utilizando o FIAP-QualidProjSW-Aula-19-DataQuality – EXERCICIOS, aplique o DataCleaner para perfilar os dados, encontrando os problemas.

*\*Resolva e tire suas dúvidas com o professor*

## **SOLUÇÃO DOS EXERCÍCIOS RESOLVIDOS**

## 1. Exercício resolvido - SOLUÇÃO

No caso da avaliação de duplicação de registros de produtos, seria suficiente avaliar a inexistência de duplicidade do número de identificação do fabricante, mas por segurança, é aconselhado verificar a não duplicidade de nomes de produtos entre as duas bases de dados, uma vez que, como as duas bases são partições verticais da base de produtos, cada uma representa de forma parcial e única uma parte do conjunto de produtos cadastrados.

*\*Solução*

## 2. Exercício resolvido

Qual a causa do problema de qualidade reportado a seguir.

O problema está nas pessoas que descumprem as metas de prazos de fechamento.

*\*Solução*

### 3. Exercício resolvido

Associe corretamente o que gerou o problema de qualidade de dados reportado.

Segundo o dono da informação, custo médio de estoque deve ser calculado uma vez por mês, com base nos preços de compras praticados para os materiais que ainda estão no estoque, mas o programador está atualizando a média a cada chegada de material com base no valor unitário da última compra. Inicialmente, quando o sistema entrou no ar, os estoques foram carregados com o custo médio que constava no sistema legado e que estava correto.

- ☒ Faltou harmonização de conceitos
- ☐ Faltou limpeza de dados na carga inicial de cadastros de estoques

*\*Solução*

## 4. Exercício resolvido

Associe corretamente o(s) parâmetro(s) de qualidade envolvido(s).

As formulações de medicamentos de uma indústria farmacêutica possuem um sistema de administração regulatória com o cadastro de elementos químicos e concentrações que é reportado para a ANVISA, porém, nas suas fábricas, existem formulações cadastradas nas máquinas de controle do chão de fábrica que não coincidem sempre com as formulações licenciadas na agência regulatória. O produto final é retrato da formulação que consta no chão de fábrica.

- ☒ Acurácia
- ☐ Completude
- ☐ Consistência
- ☐ Atualidade
- ☒ Validade
- ☒ Unicidade

*\*Solução*

## 5. Exercício resolvido

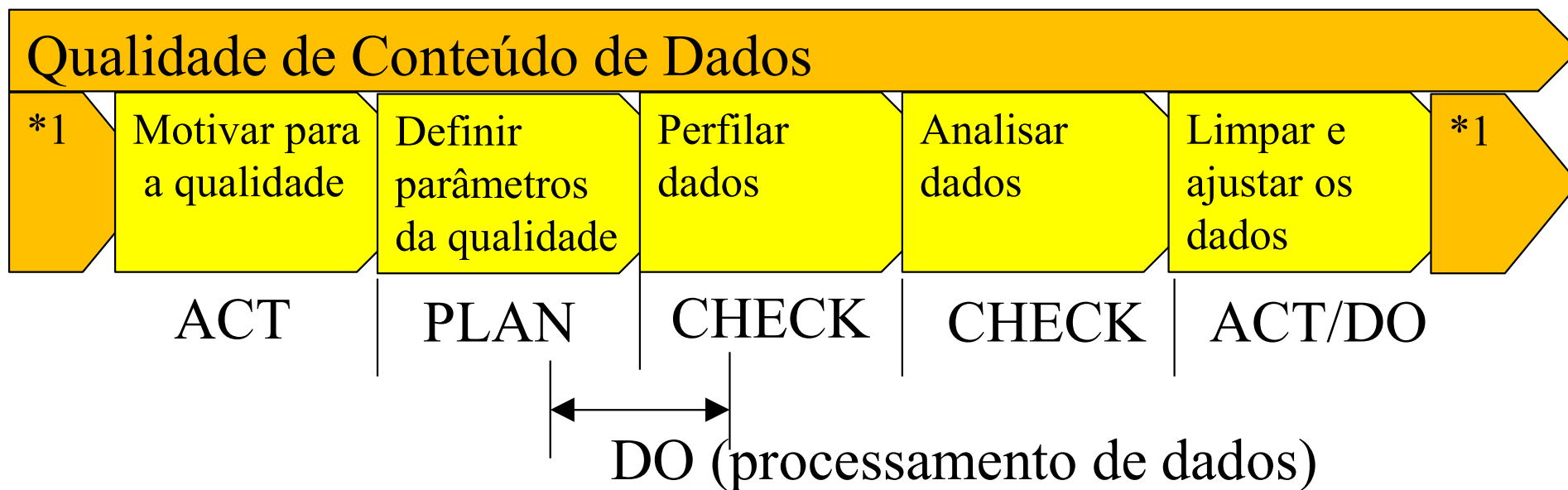
Para resolver problemas de definição de formulas de cálculo de comissões quando da fusão de duas empresas e suas áreas comerciais é necessário realizar...

- ☒ Harmonização de definições
- ☒ Cleansing na formação de um novo banco de dados com base nos legados

*\*Solução*

## 6. Exercício resolvido

Faça uma associação correta o PDCA com os processos de qualidade de dados



*\*Solução no final deste material*