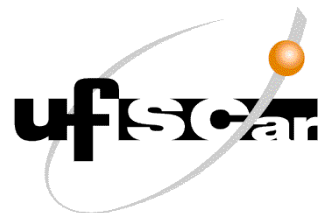

Data Warehouse Toolkit: Telecomunicações e Utilitários (Cap. 10)

Vinícius Ramos Toledo Ferraz

Prof. Dr. Ricardo Rodrigues Ciferri



Agenda

- Dicas para revisão de projetos
- O que está errado neste estudo de caso?
- Correções!
- Dimensão Posição Geográfica

Dicas para revisão de projetos

■ Granularidade

- *Qual o grão da tabela de fatos?*

“Declarar uma definição clara e concisa do grão da tabela de fatos é fundamental para um esforço de modelagem produtivo”

“Devem ser identificados os fatos que são consistentes com a declaração do grão”

Dicas para revisão de projetos

■ Granularidade

- *Qual o grão da tabela de fatos?*

Fato da transação de vendas a varejo no POS

Chave da data (FK)
Chave do produto (FK)
Chave da loja (FK)
Chave da promoção (FK)
Número da transação POS (DD)
Quantidade de vendas
Total de vendas em dólar
Valor de custo em dólar
Valor de lucro bruto em dólar

↑
**Grão = 1 linha por
linha de transação POS**

Fato da cesta de mercado no POS

Chave da data (FK)
Chave do produto A (FK)
Chave do produto B (FK)
Chave da loja (FK)
Chave da promoção (FK)
Contagem da cesta
Quantidade de vendas do produto A
Quantidade de vendas do produto B
Total de vendas em dólar do produto A
Total de vendas em dólar do produto B

↑
**Grão = 1 linha para cada par
de produtos vendidos em um
dia por loja e promoção**

Dicas para revisão de projetos

■ Granularidade

- *Qual o grão da tabela de fatos?*

“Fatos agregados como totais anuais são perigosos, pois não são totalmente aditivos”

“Ir para o nível mais baixo de granularidade possível não implica utilizar a maior quantidade de dados detalhados disponíveis na empresa”

Dicas para revisão de projetos

■ Granularidade

- *Qual o grão da tabela das tabelas de dimensão?*

“Cada dimensão deve utilizar um único valor com cada linha de medições da tabela de fatos”

“Devemos recolher as hierarquias de atributos das dimensões, sempre que possível”

Dicas para revisão de projetos

■ Granularidade

- *Qual o grão da tabela das tabelas de dimensão?*

*“A economia de disco com o **snowflaking** raramente compensa suas desvantagens em complexidade e desempenho da consulta”*

*“Usar **outriggers** em um cluster de cardinalidade baixa ou atributos reaproveitados com frequência deve ser a exceção, e não a regra”*

Dicas para revisão de projetos

- **Dimensão Data**

A que se refere a Dimensão Data?

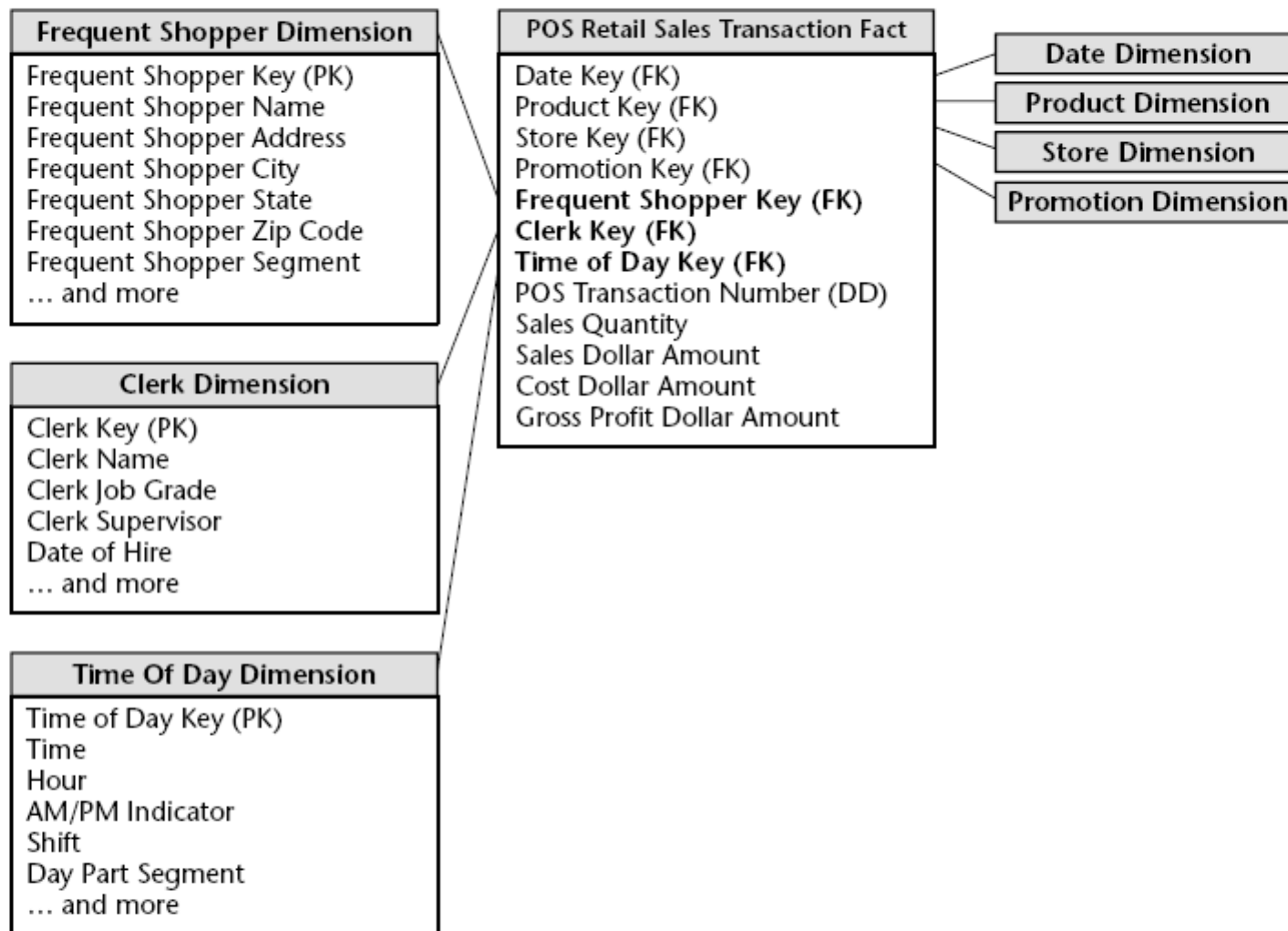
Dicas para revisão de projetos

■ Dimensões de degeneração

- ❑ *O grão da tabela de fatos representa uma transação ou item de linha da transação?*
- ❑ *Se tivéssemos uma dimensão separada para as informações de cabeçalho de uma transação, seríamos redundantes?*

“Os números de controle operacional normalmente dão origem a dimensões vazias e são representados como dimensões de degeneração”

Dicas para revisão de projetos



Dicas para revisão de projetos

■ Chaves substitutas

- *Há identificadores operacionais sendo utilizados como chave?*

“Em vez de contar com chaves ou identificadores operacionais, recomendamos o uso de chaves substitutas no seu projeto dimensional”

“Nenhuma chave do DW deve ser inteligente”

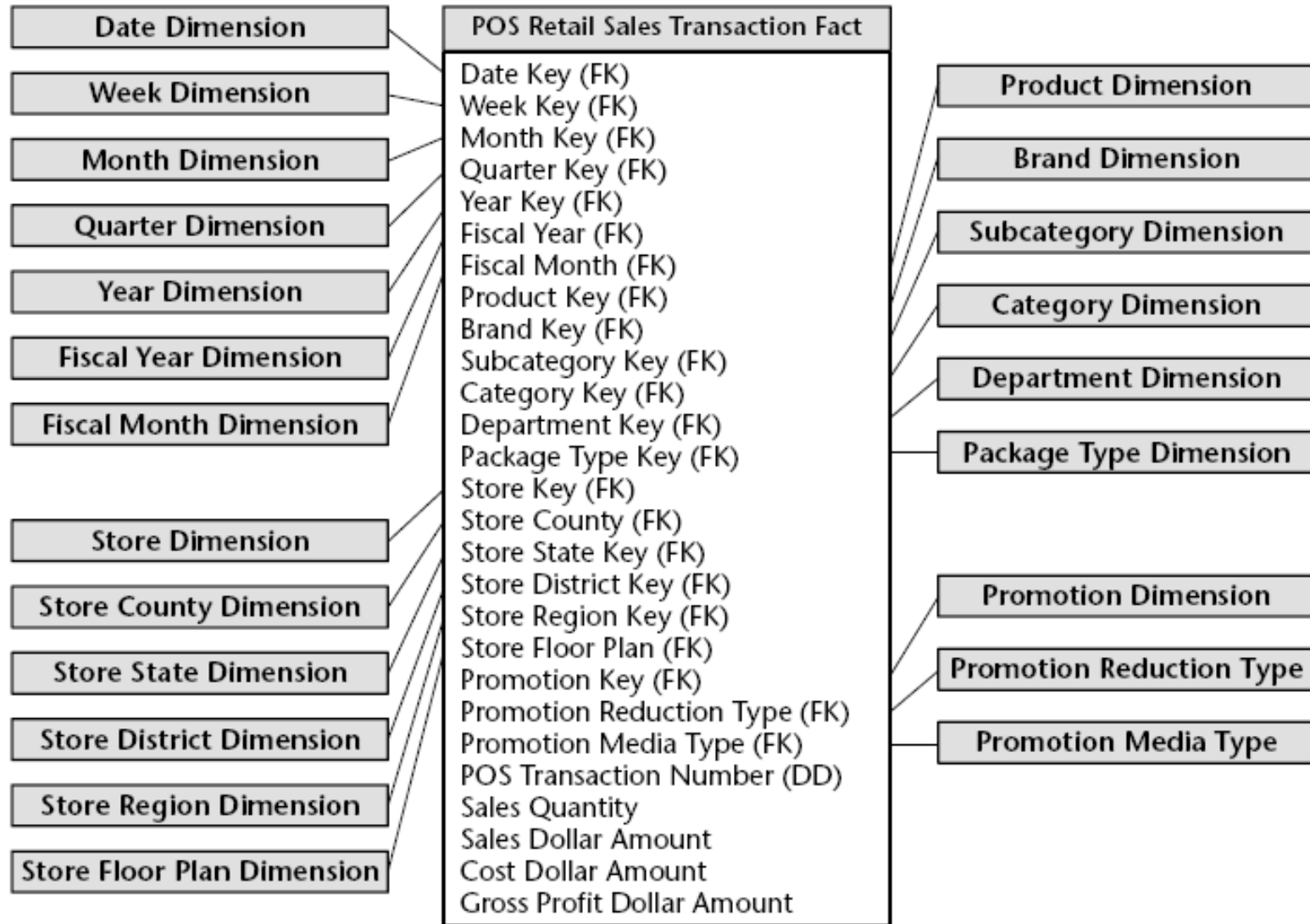
Dicas para revisão de projetos

- **Dimensões em excesso (ou em falta)**

- *O projeto contém entre 5 e 15 dimensões?*

“Um número muito grande de dimensões normalmente é um sinal de que várias delas não são completamente independentes e devem ser combinadas.”

Dicas para revisão de projetos



Agenda

- Dicas para revisão de projetos
- O que está errado neste estudo de caso?
- Correções!
- Dimensão Posição Geográfica

O que está errado neste projeto?

- **Data Warehouse de uma grande empresa de telecomunicações wireless**

	Data	Cliente	Produto	Plano de Tarifa	Canal de vendas	Linha de serviço	Switch	Fornecedor	Conta do LR	Empresa	Funcionário	Tipo da chamada	Status da chamada
Bilhetagem do cliente	X	X	X	X	X	X							
Tráfego de detalhes da chamada	X	X	X	X		X	X						
Compras	X		X					X	X	X			
Estoque do distribuidor	X		X		X			X		X			
Vendas por canal	X	X	X	X	X	X		X					
Chamadas de serviço	X	X	X							X	X	X	X
Itens de reparo	X	X	X							X	X	X	X

O que está errado neste projeto?

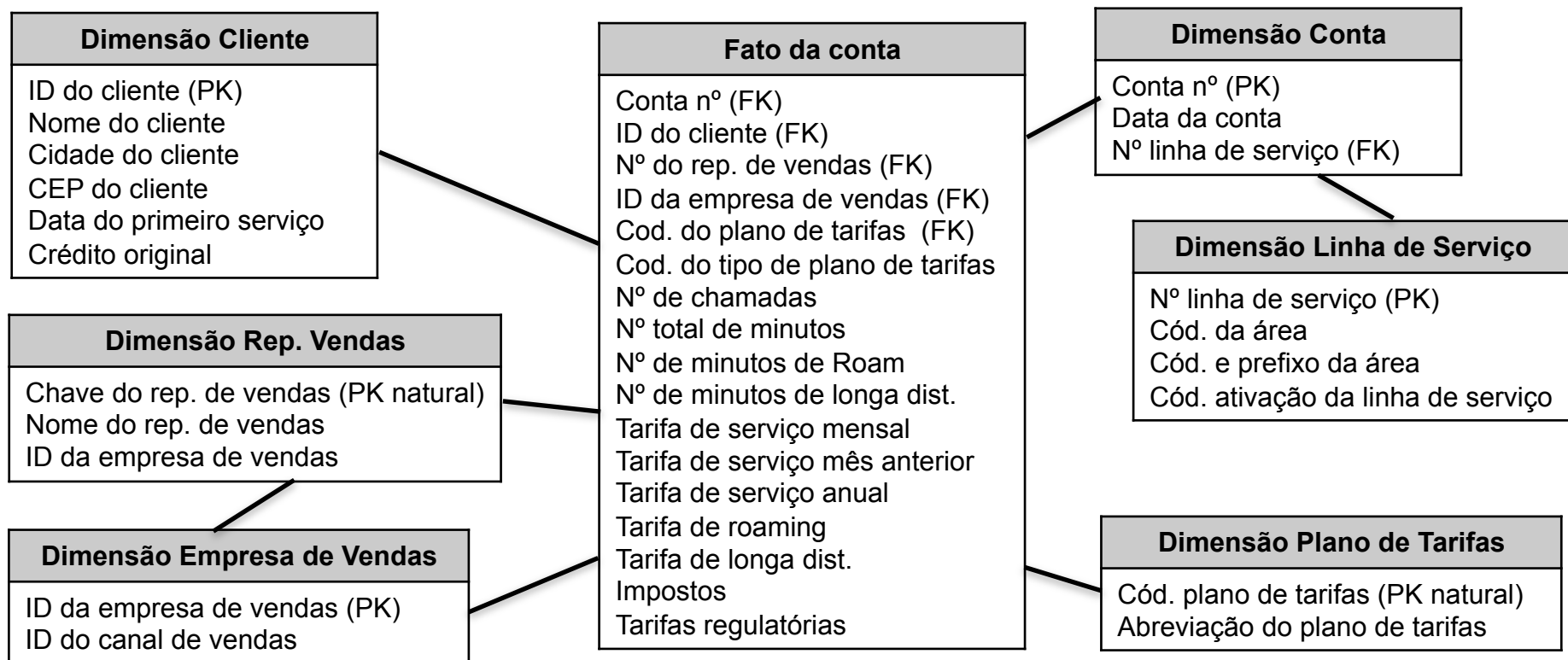
- Foco no processo de bilhetagem do cliente, pois deseja-se poder analisar...
 - Uso mensal
 - Receita gerada pelo cliente (bilhetagem)
 - Organização de vendas
 - Plano de tarifas
- ...para medir o desempenho do canal de vendas e seu representante, além de analisar a eficácia dos planos de tarifas.

O que está errado neste projeto?

- Dinâmica dos serviços da empresa:
 - Todo mês, é gerada uma conta para cada linha de serviço (número telefônico)
 - Cada linha de serviço é associada a um único cliente
 - Um cliente pode ser várias linhas de serviço, que aparecem como itens de linha separados na mesma conta
 - Cada linha de serviço possui seu próprio conjunto de métricas de bilhetagem, como o número de minutos usados e a tarifa de serviço mensal
 - Há um único plano de tarifa associado a cada linha de serviço em uma dada conta. Mas tal plano pode mudar
 - Um rep. de vendas (e sua empresa/canal de vendas respectivos) é associado a cada linha de serviço para avaliar a receita gerada por cada rep. de vendas.

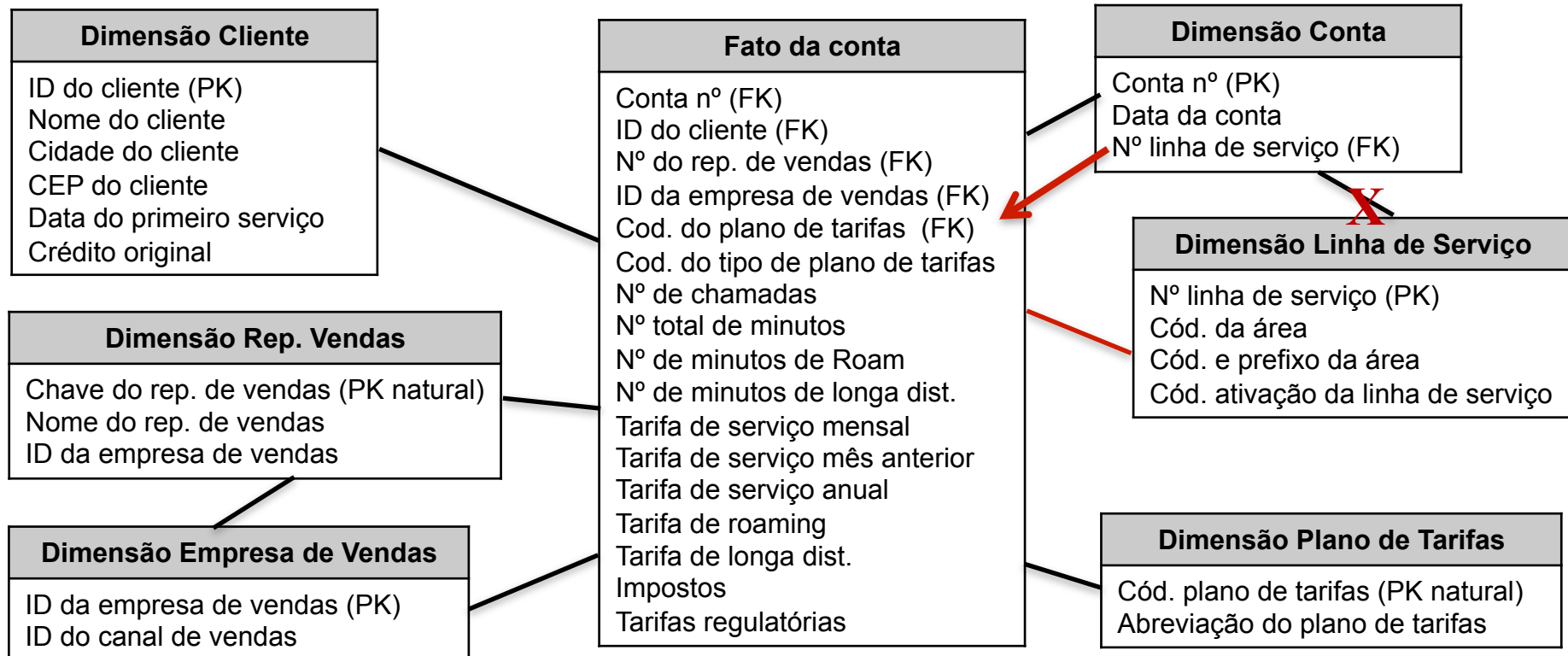
O que está errado neste projeto?

- A equipe de DW expõe orgulhosamente sua “obra-prima” da modelagem dimensional e aguarda o SEU pronunciamento...



Grão = 1 linha por conta a cada mês

Revisando o projeto...



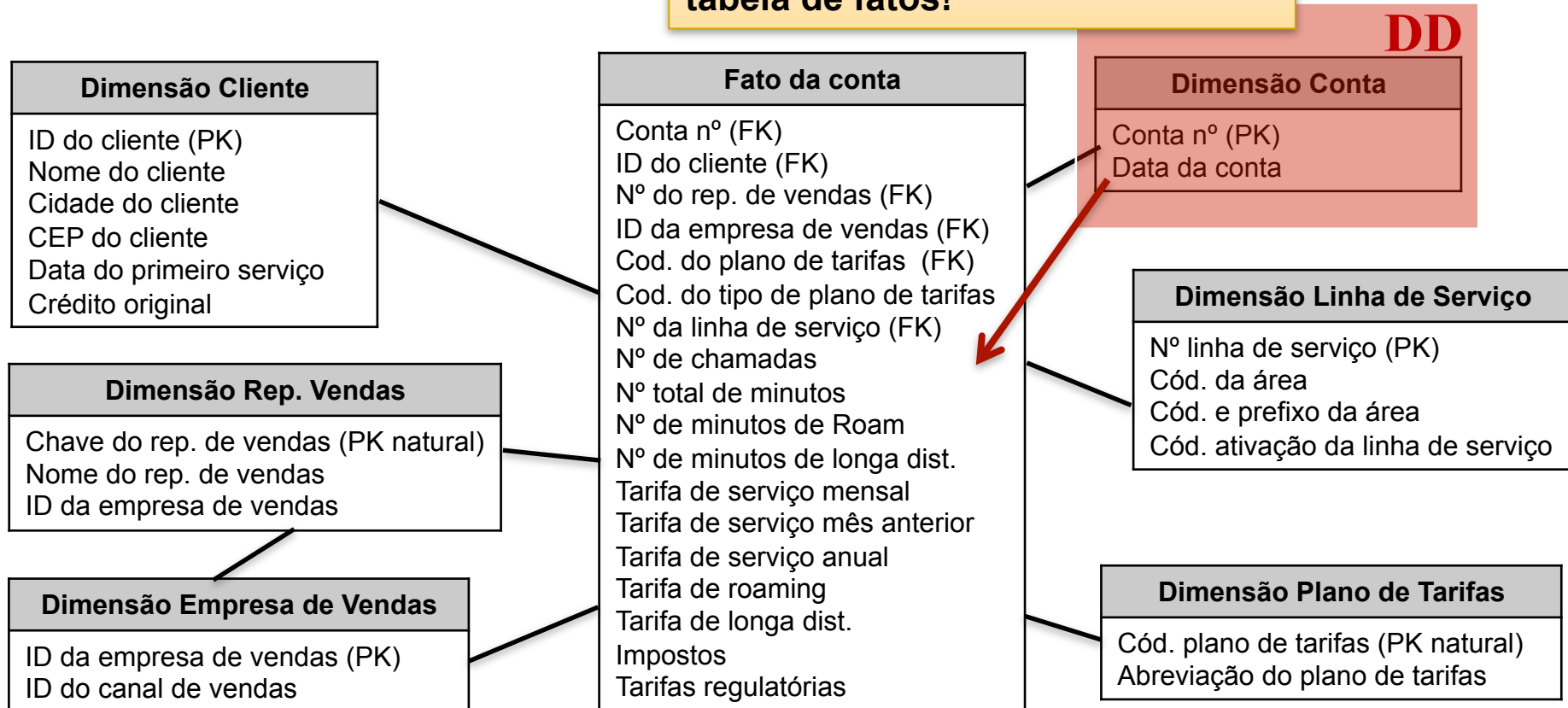
Grão = 1 linha por conta a cada mês

O grão mais baixo seria: 1 linha por linha de serviço em uma conta

Revisando o projeto...

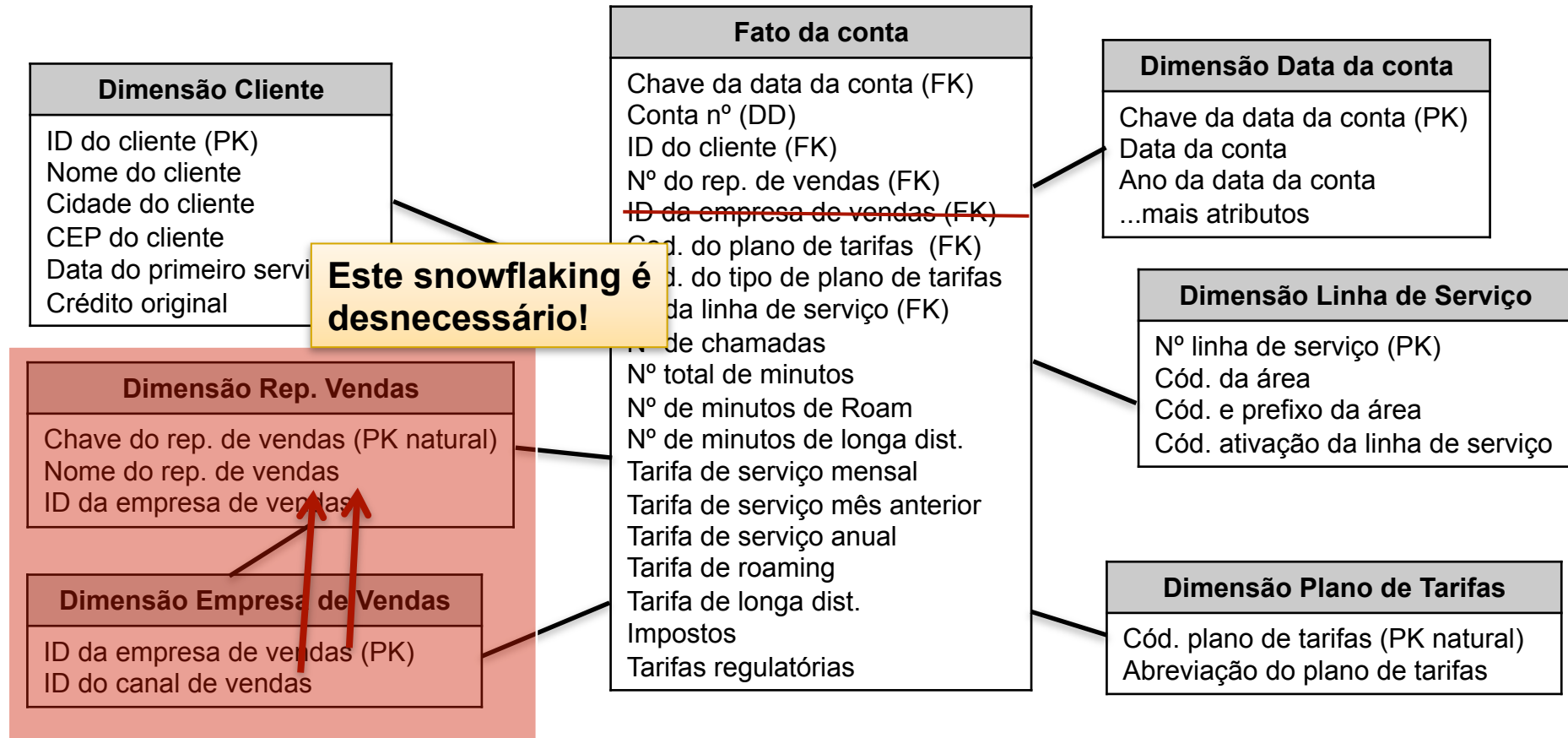
Vamos criar uma dimensão data!

Esta dimensão provavelmente terá o mesmo nº de tuplas que a tabela de fatos!



Grão = 1 linha por linha de serviço em uma conta

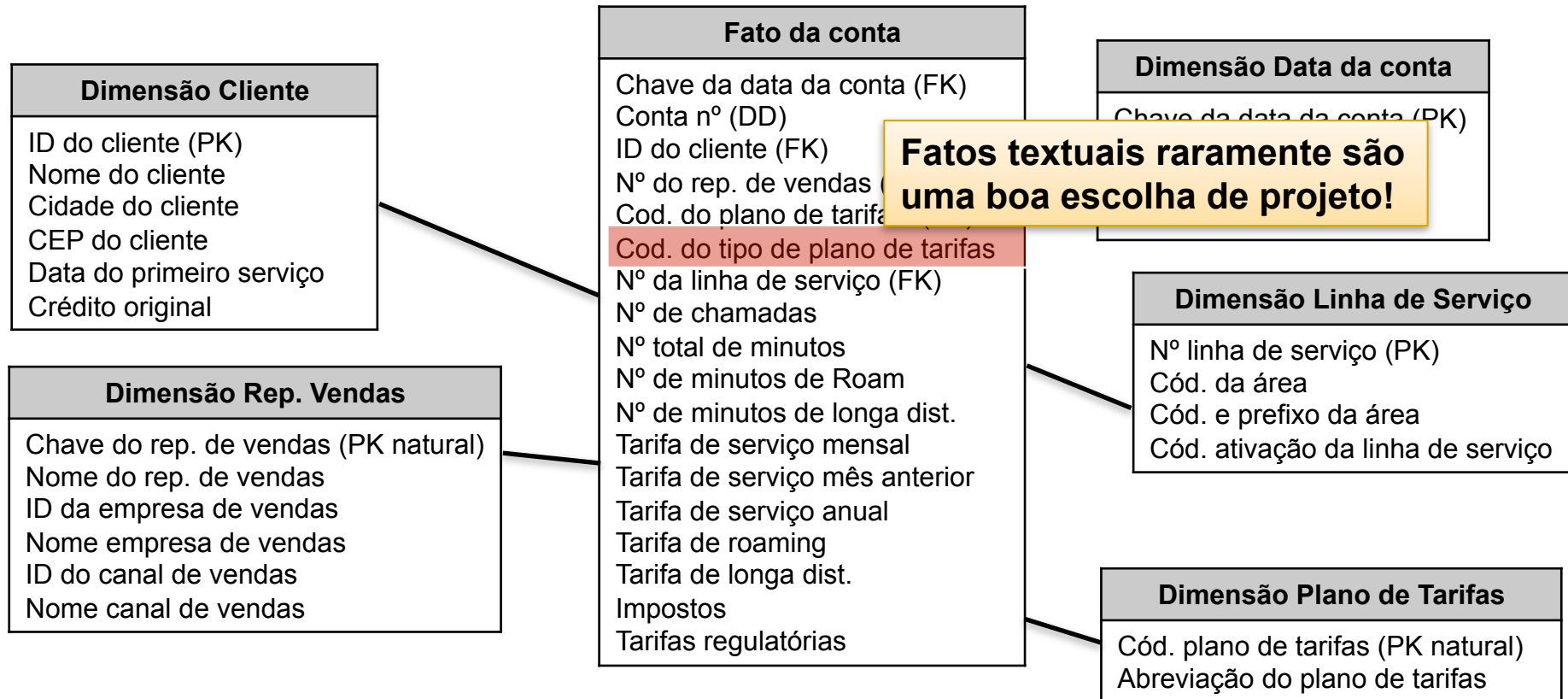
Revisando o projeto...



Grão = 1 linha por linha de serviço em uma conta

Revisando o projeto...

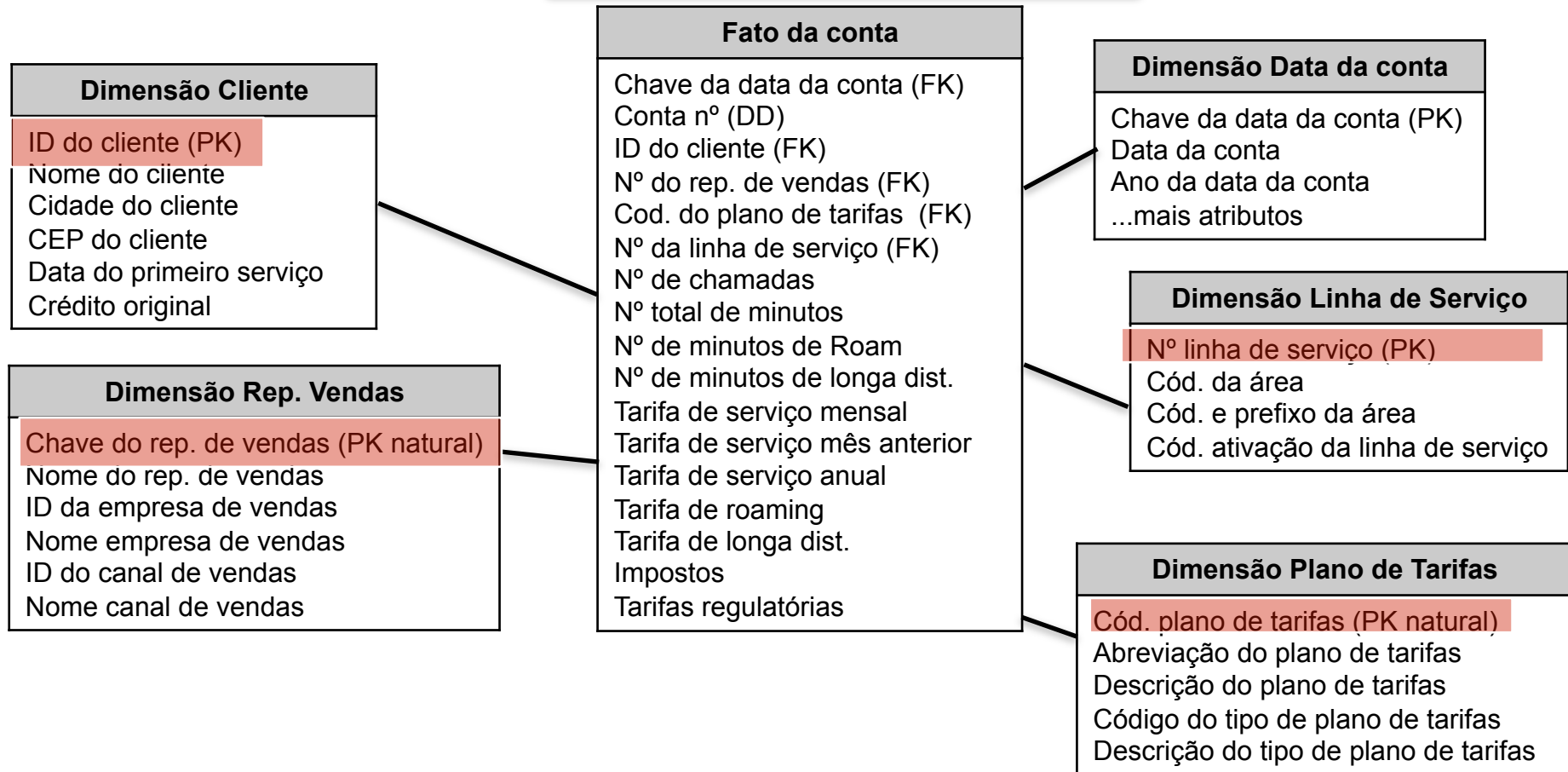
Vamos agupar este fato textual e seus atributos descritivos na Dimensão Plano de Tarifas!



Grão = 1 linha por linha de serviço em uma conta

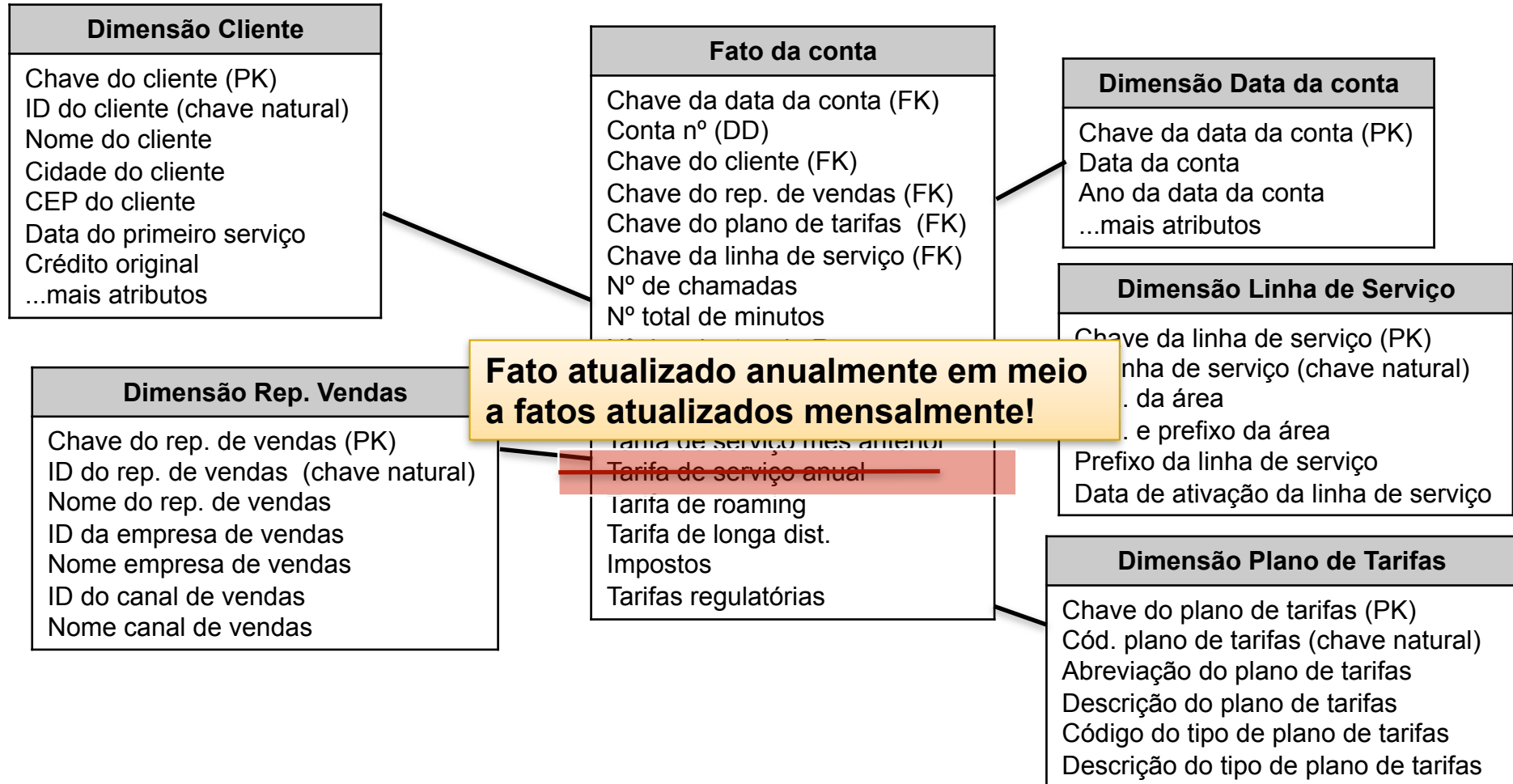
Revisando o projeto...

Incentiva-se a implementação de chaves substitutas!



Grão = 1 linha por linha de serviço em uma conta

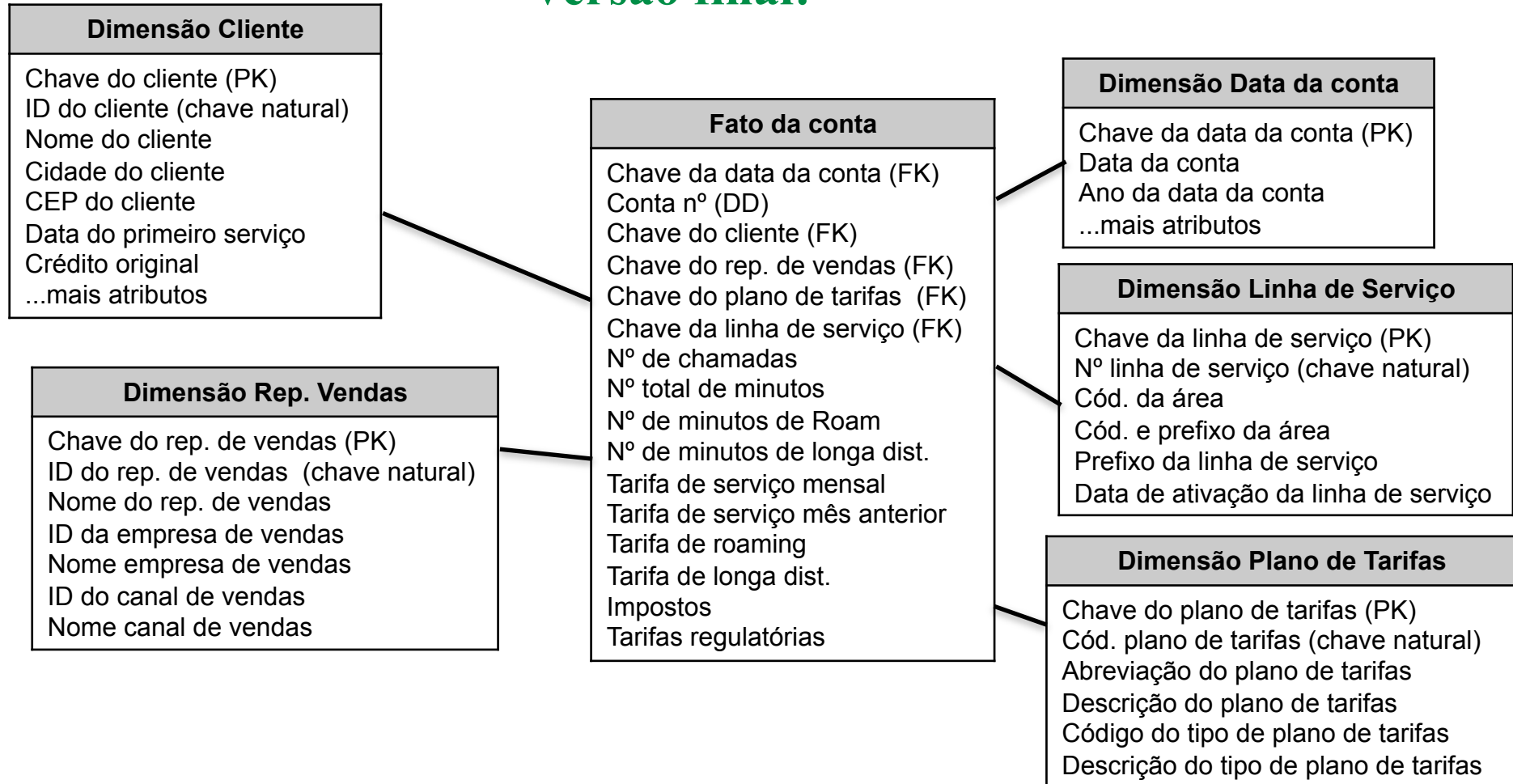
Revisando o projeto...



Grão = 1 linha por linha de serviço em uma conta

Revisando o projeto...

Versão final!



Grão = 1 linha por linha de serviço em uma conta

Dimensão Posição Geográfica

- Utilizada quando a localização geográfica faz parte do conjunto de atributos
 - Pode ser uma rua, cidade, estado CEP ou mesmo um conjunto de latitudes e longitudes
 - Cria-se uma tabela geográfica mestra, onde os dados são padronizados e depois reutilizados
 - Cada linha desta tabela é um uma localização geográfica.

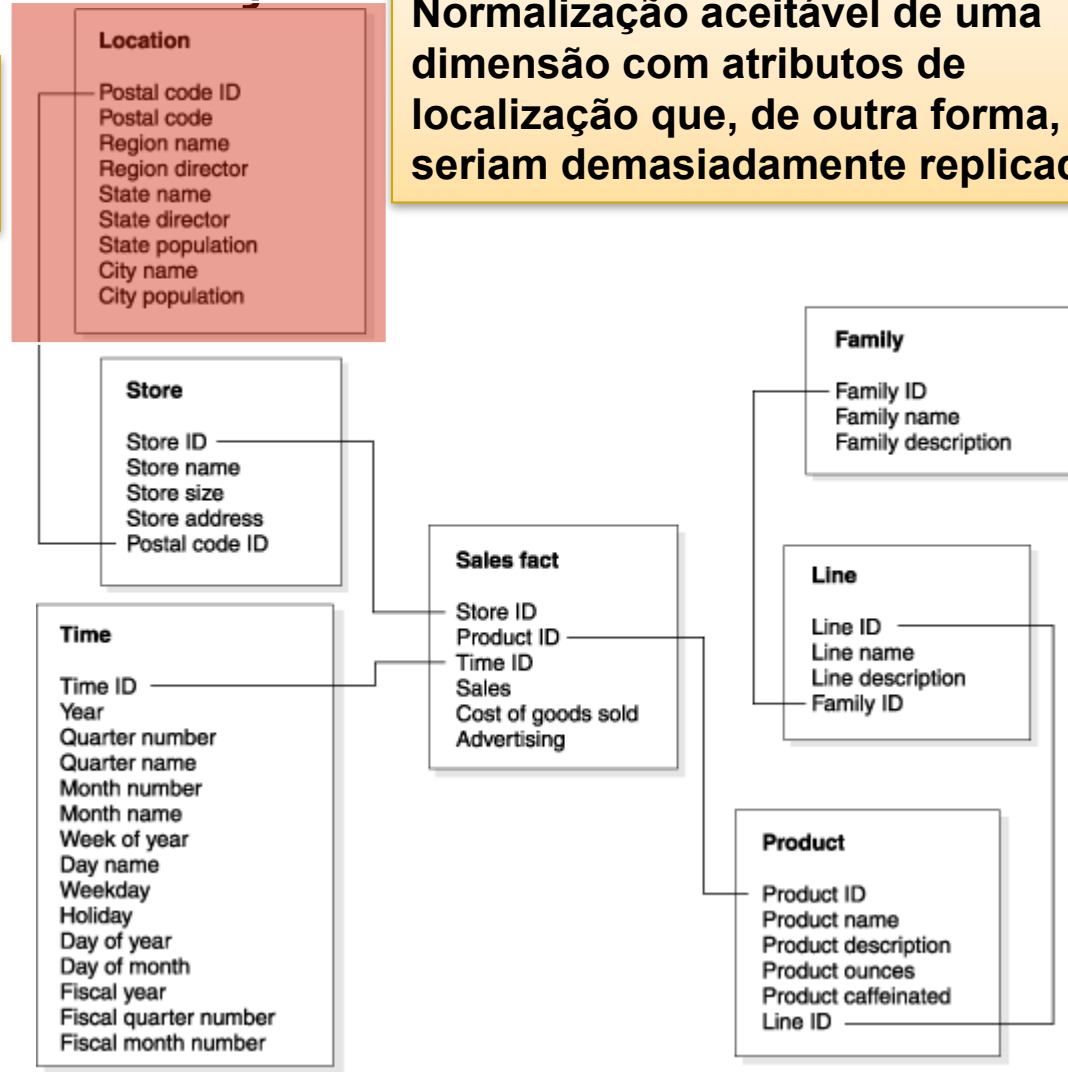
Dimensão Posição Geográfica

■ Outrigger de localização

Notem que a hierarquia de localização não foi normalizada!

Outrigger de localização: Normalização aceitável de uma dimensão com atributos de localização que, de outra forma, seriam demasiadamente replicados.

Notem também que este exemplo não aplica algumas das dicas apresentadas neste capítulo ;-)



Dimensão Posição Geográfica

- Integração com GIS
 - Possibilita visualização espacial das consultas, permitindo evidenciar informações antes “escondidas” em planilhas e relatórios tradicionais
 - Permite que o GIS tire proveito dos recursos espaciais dos SGBD's

Perguntas

Vinícius R. T. Ferraz

vinicius_ferraz@dc.ufscar.br

