

Advanced Data Modeling

10DTS

Advanced Data Modeling - 10DTS

O que veremos nesta disciplina:

- Contexto "sócio-tecno-econômico" dos dados
- Arquiteturas de Solução de DW, DL e LH
- Bancos de Dados
 - Tipos de Dados
 - Tipos de Bancos de Dados
- Modelagem de dados estruturados
 - Modelagem relacional
 - Modelagem dimensional
- Modelagem de dados não estruturados
 - Modelagem chave e valor
 - Modelagem de documentos
 - Modelagem de grafos
 - Modelagem colunar

Avaliação:

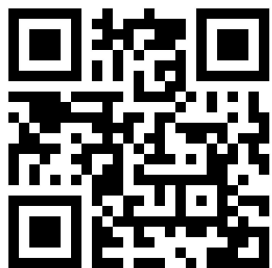
- Trabalhos práticos em sala de aula
- Entrega no Portal FIAP até a aula seguinte.



Advanced Data Modeling

Lourenço Taborda

Arquiteto de Solução



<https://linktr.ee/devtb>
d



Keylla Saes

FIAP MBA+

profkeylla.saes@fiap.com.br br.linkedin.com/in/keyllasaes

Formação Acadêmica

- Colégio Técnico: Processamento de dados – Colégio Brasília
- Graduação: Ciência da Computação – Universidade São Judas Tadeu
- Pós-graduação: Administração de Banco de Dados – FIAP
- Pós-graduação: MBIS – MBA em Gestão de Projetos – PUC/SP
- Mestrado acadêmico: Sistemas de Informação – EACH/USP
- Doutorado: Aluna especial – USP

Certificações

- Exin - Privacy and Data Protection
- Exin - Iso 27001 Foundation – Information Security Certification
- DP-900 Microsoft Azure Data Fundamentals
- AI-900 Microsoft Azure AI Fundamentals

Experiência Profissional

- Nasajon Sistemas
- Hexasolution Consultoria
- Accenture do Brasil

Experiência Acadêmica

- Coordenadora na FIAP do MBA Data Science & IA
- Professora na FIAP nos cursos de BI & Analytics e Engenharia de Dados

Vivência na área de TI

- Consultora em TI atuando com dados desde 2001
- Projetos em indústria de Varejo, Telecom, Financeira, Seguros e Aviação nas áreas de database, e-commerce, CRM, BI, Integração, Migração, Data Quality, MDM, Data Management, Governança de dados, Analytics, Big Data, Inteligência Artificial e RPA



HOBBIES

- Séries
- Ler
- Dançar
- Andar de Bike

Modelagem Dimensional

Business Intelligence

Business Intelligence é usar da coleta de dados, organização, análise, ação e monitoramento para tomar **melhores decisões** e saber se os investimentos feitos estão trazendo bons resultados.

Como a explicação acima deixa claro, **BI não é uma ferramenta**, como muitos pensam. Apesar de depender de softwares robustos para entregar todo o valor que se espera, o Business Intelligence vai muito além disso.

Os líderes e gestores devem ter em mente que BI é um conjunto de processos que tem por objetivo entregar a **informação certa, para a pessoa certa, na hora certa**.

KPI - Key Performance Indicator

Os KPI's, ou indicadores de desempenho, facilitam a **transmissão da visão e missão de uma determinada empresa** para funcionários que não ocupam cargos elevados.

Um indicador chave de desempenho funciona como um **veículo de comunicação**, garantindo que os trabalhadores entendam como os seus trabalhos são importantes para o sucesso ou falta de sucesso da organização. Um KPI pode ser quantidade **quantitativo** ou **qualitativo**.



Exemplos de indicadores:

- indicadores de atraso;
- indicadores de processo;
- indicadores de resultados;
- indicadores financeiros.

Exemplos de KPI's de “Mercado”

- Quantidade de vendas
- Quantidade de vendas por região
- Quantidade de vendas por região e por categoria
- Quantidade de clientes
- Quantidade de clientes por região, por segmento, por renda
- Valor ticket médio
- Valor do maior ticket do mês
- Venda média por vendedor
- Crescimento mensal por região

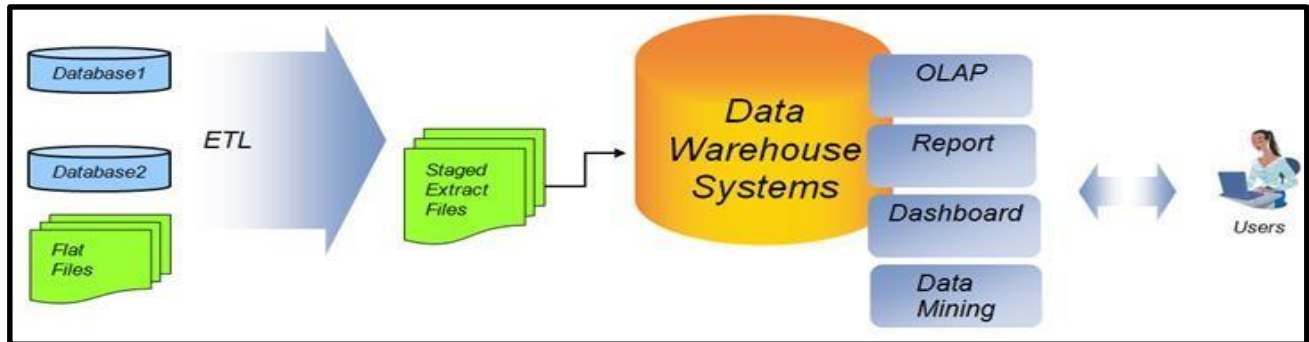


Arquitetura tradicional

**Carga e
Transformação**

**Armazenamento e
Organização**

Análise

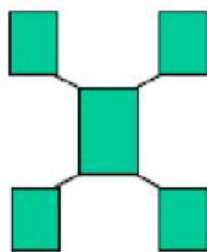


Dados
Transacionais

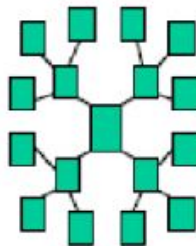
Dados
Analíticos

A **modelagem multidimensional**, ou dimensional como às vezes é chamada, é a técnica de modelagem de banco de dados para o auxílio às consultas do Data Warehouse nas mais diferentes perspectivas.

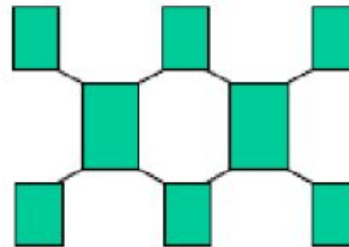
Tipos de Modelagem Dimensional



Modelo Estrela



Modelo Floco de Neve



Modelo Multi-Estrela

Ilustração 5: Tipos de modelo dimensional

Fonte: BALLARD, *et al.*, 2006

- O **modelo star schema (estrela)** consiste em uma única tabela de fatos, e diversas tabelas de dimensão não normalizadas.
- O **modelo snowflake (floco de neve)** utiliza de técnicas de normalização em suas tabelas de dimensão. Este modelo é uma extensão do modelo estrela com dimensões normalizadas. Com a normalização das tabelas de dimensão o modelo se assemelha a um floco de neve, daí seu nome.
- Já o **modelo multistar (multi-estrela)** possui diversas tabelas de fatos unidas através de dimensões.

Modelo Dimensional Star Schema (Estrela)

O conceito de Star Schema, ou modelo estrela, foi idealizado por Ralph Kimball.

Star Schema



Modelo composto por tabelas **conceitualmente** denominadas de fatos e dimensões.

Dimensão

É uma coleção de dados descritivos distintos que irão classificar, definir e esclarecer as informações relacionadas ao **fato**.

Por exemplo, quando eu faço uma venda, quero saber por onde a venda foi feita, que produto foi vendido, ou para quem.

- Dimensão cliente
- Dimensão produto
- Dimensão cidade
- Dimensão bairro
- Dimensão sexo
- Dimensão status venda



Tipo do dado

- Dados Mestres
- Dados de Referência

Fato

Armazena o que ocorreu, é o fato propriamente dito, por isso ela tem esse nome, porque é o fato ocorrido. A tabela fato está sempre ligada a duas ou mais dimensões.

Citando um exemplo de varejo, um fato seria: Fato Vendas, Fato RH, Fato Faturamento, etc.

- Fato Venda
- Fato Cliente
- Fato Voo
- Fato Chamado
- Fato Sinistro
- Fato Aula



Tipo do dado

- Dados transacionais

Modelo Conceitual Dimensional

Star Schema (Estrela)



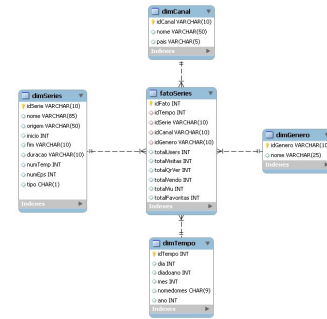
Modelo estrela representando o assunto vendas
Fonte: MACHADO, Felipe Nery Rodrigues. Tecnologia e Projeto de Data Warehouse

Processo de Modelagem Dimensional

Negócio



TI



Selecione o processo de negócio a ser modelado

Identifique a granularidade dos fatos

Identifique e modele as dimensões

Identifique e modele os fatos

Porque?

Quanto?

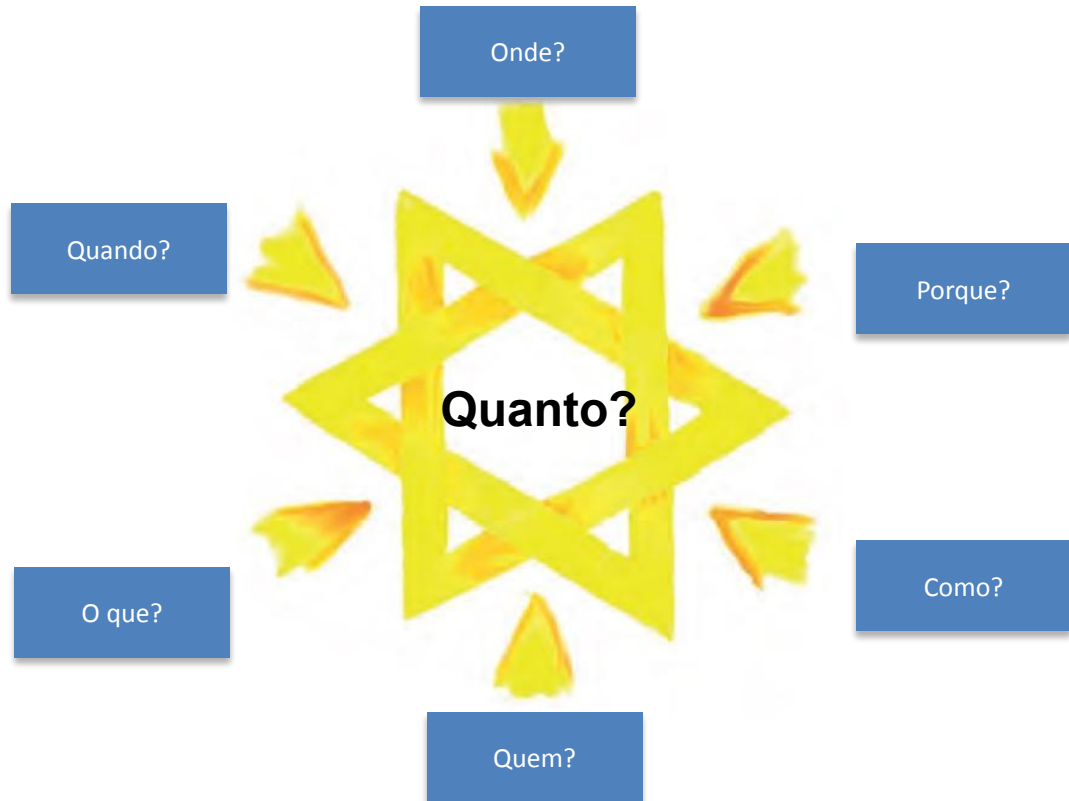
Quando/Onde/Quem

O Que ?

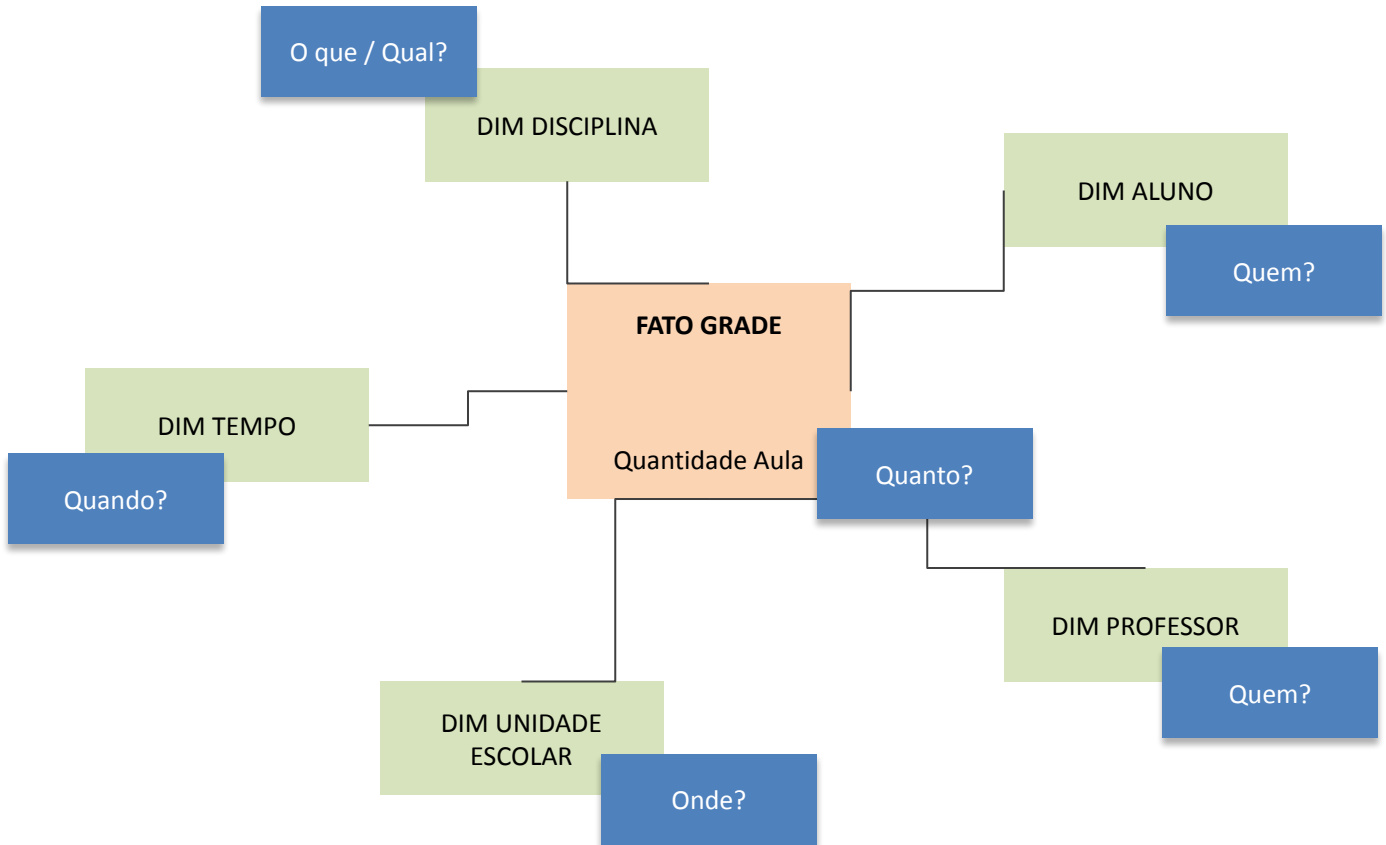
Responda as perguntas

Processo de Modelagem Dimensional

O Modelo Dimensional deve ser planejado para responder algumas perguntas:

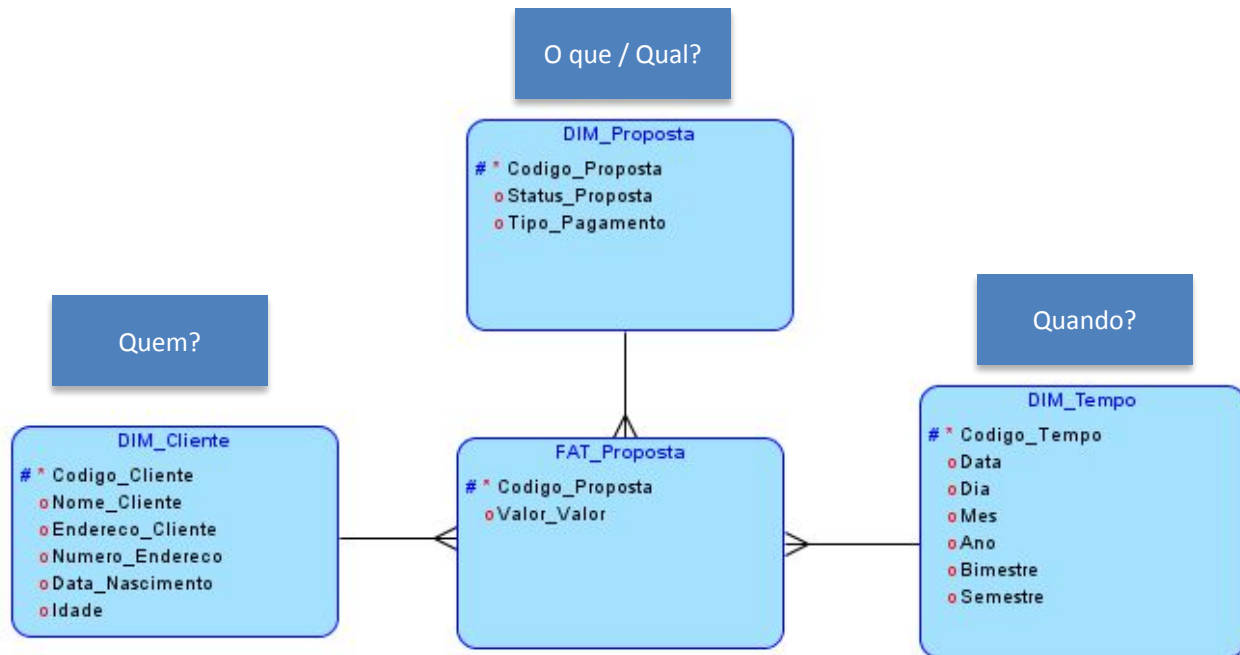


Modelo Conceitual: Fato Grade



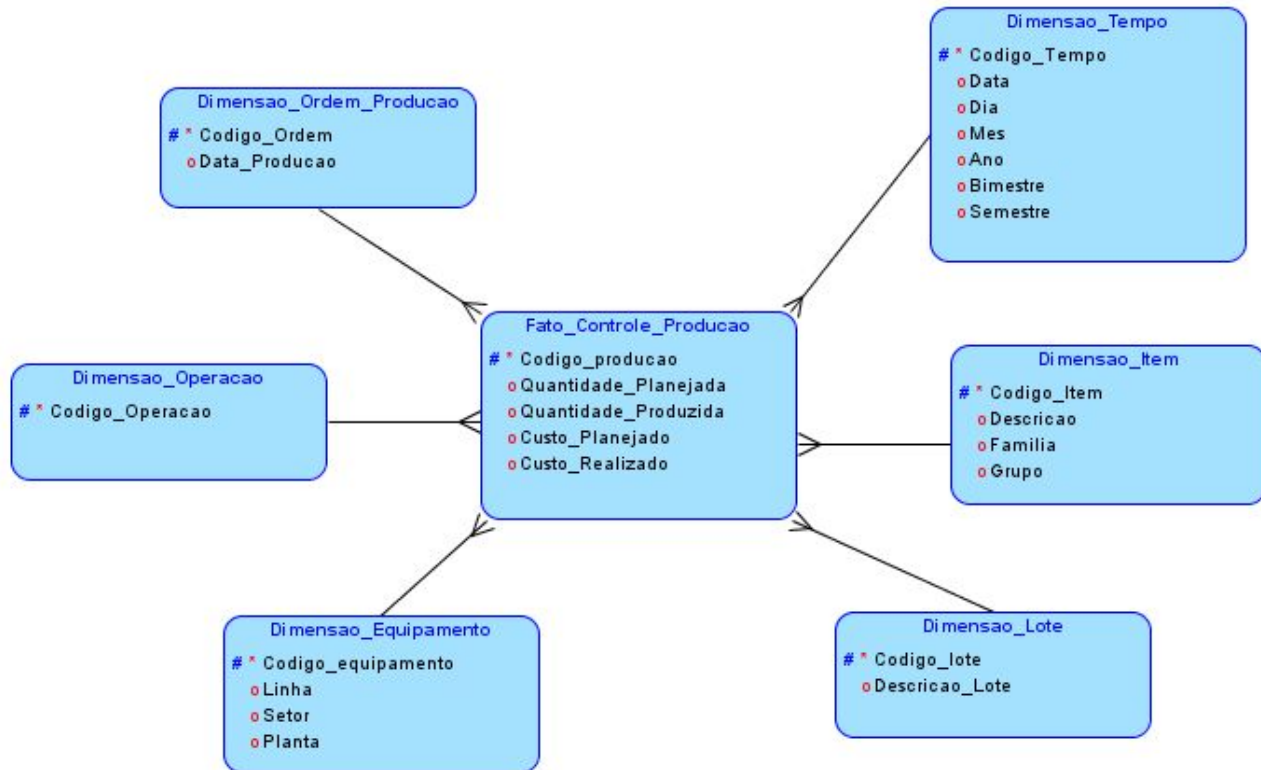
Modelo Dimensional – Star Schema FIAP MBA⁺

No Oracle SQL DataModeler – Modelo Lógico

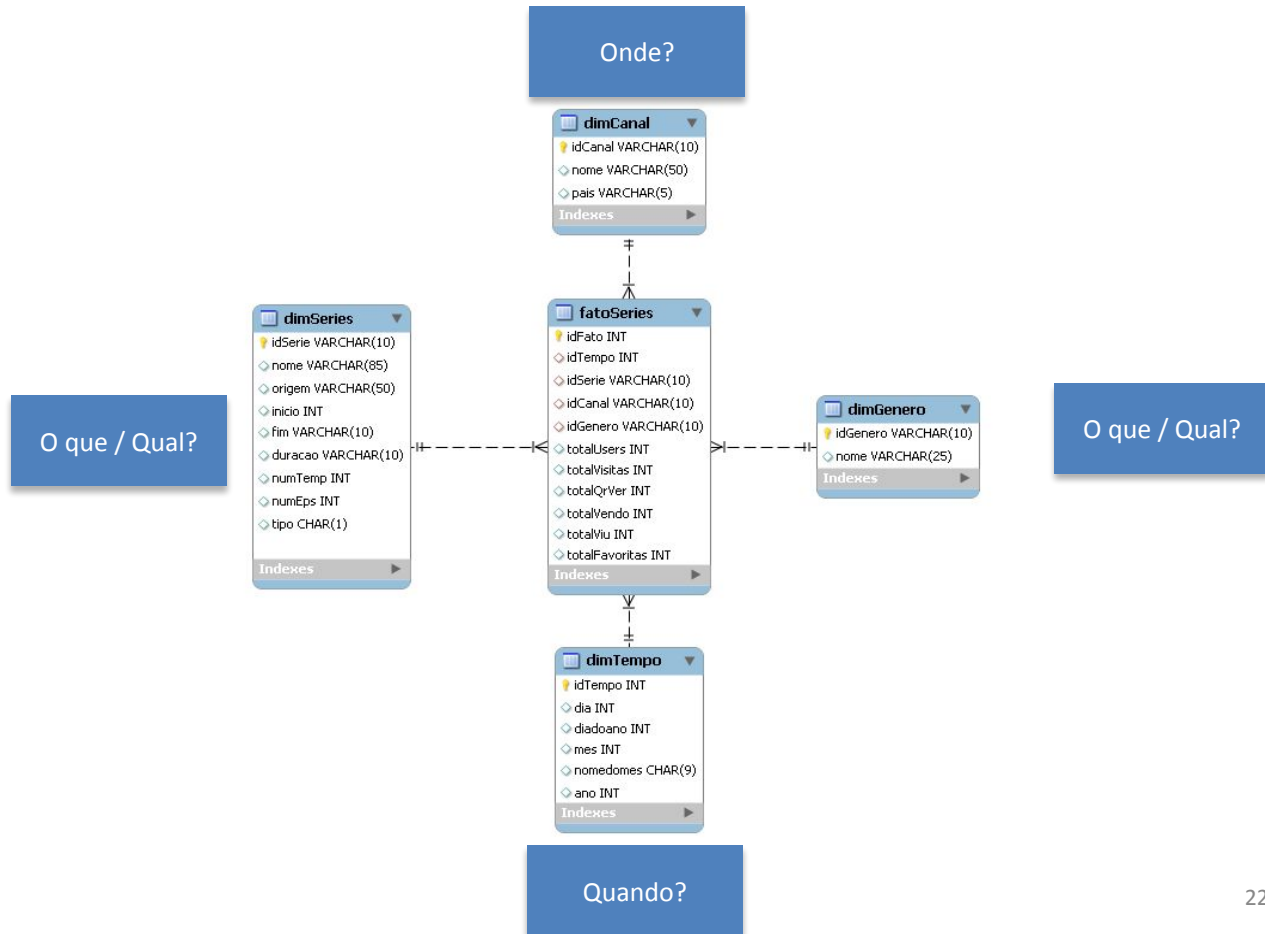


Modelo Dimensional – Star Schema

No Oracle SQL DataModeler – Modelo Lógico

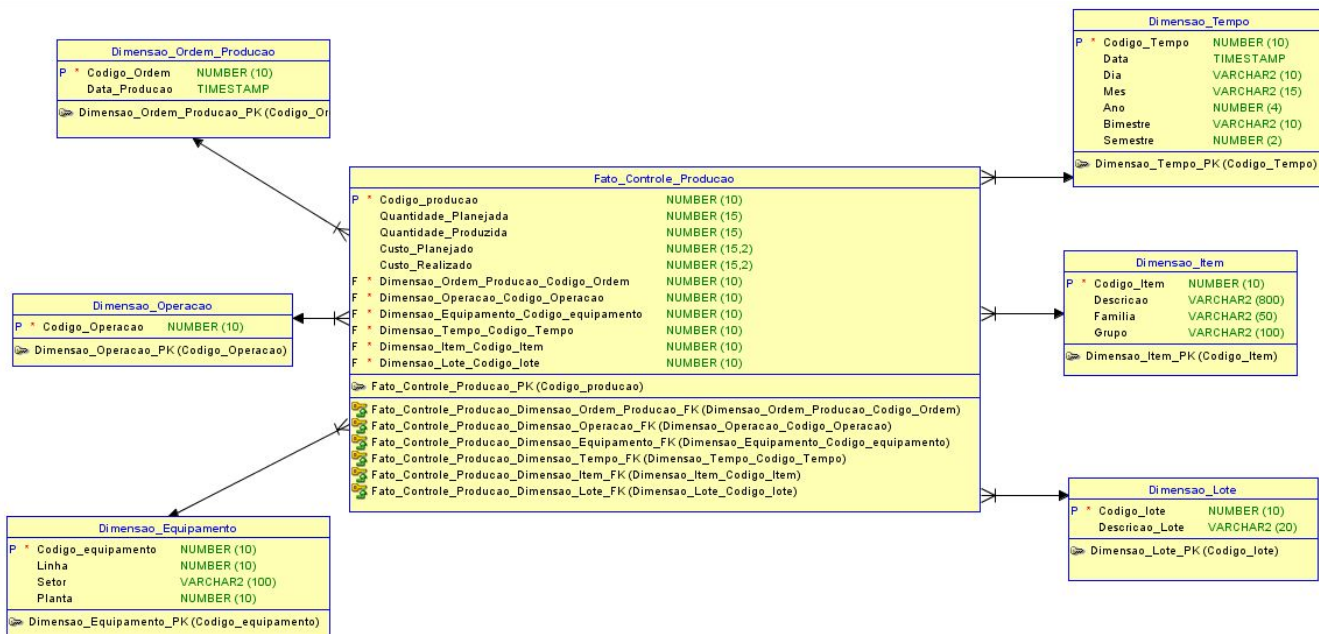


Respondendo perguntas



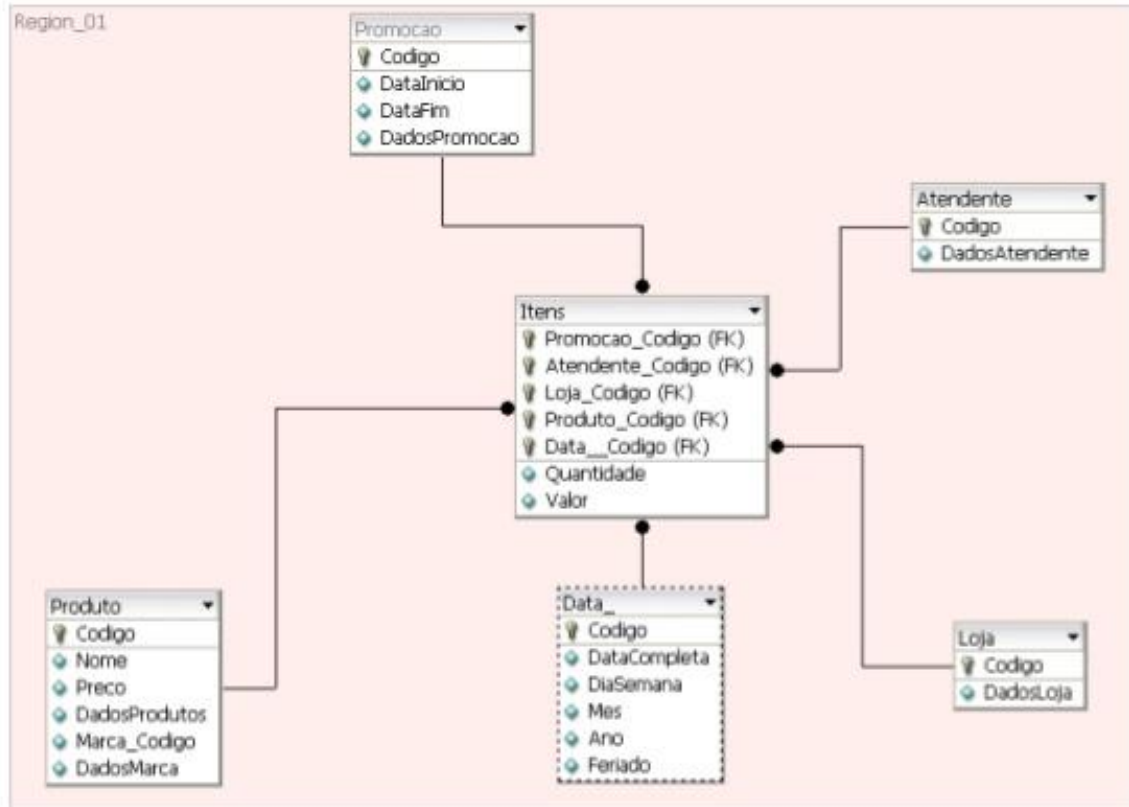
Modelo Dimensional – Star Schema

No Oracle SQL DataModeler – Modelo Físico



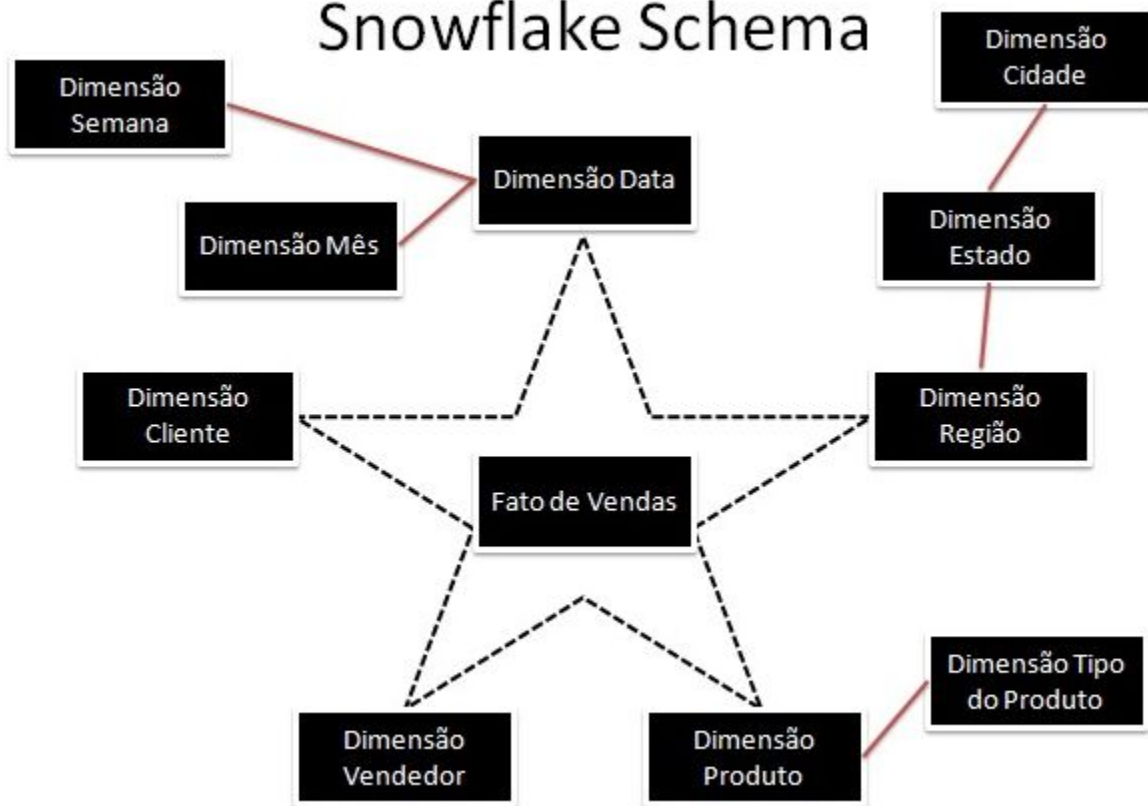
Modelo Dimensional – Star Schema

Em outras ferramentas – Modelo Físico

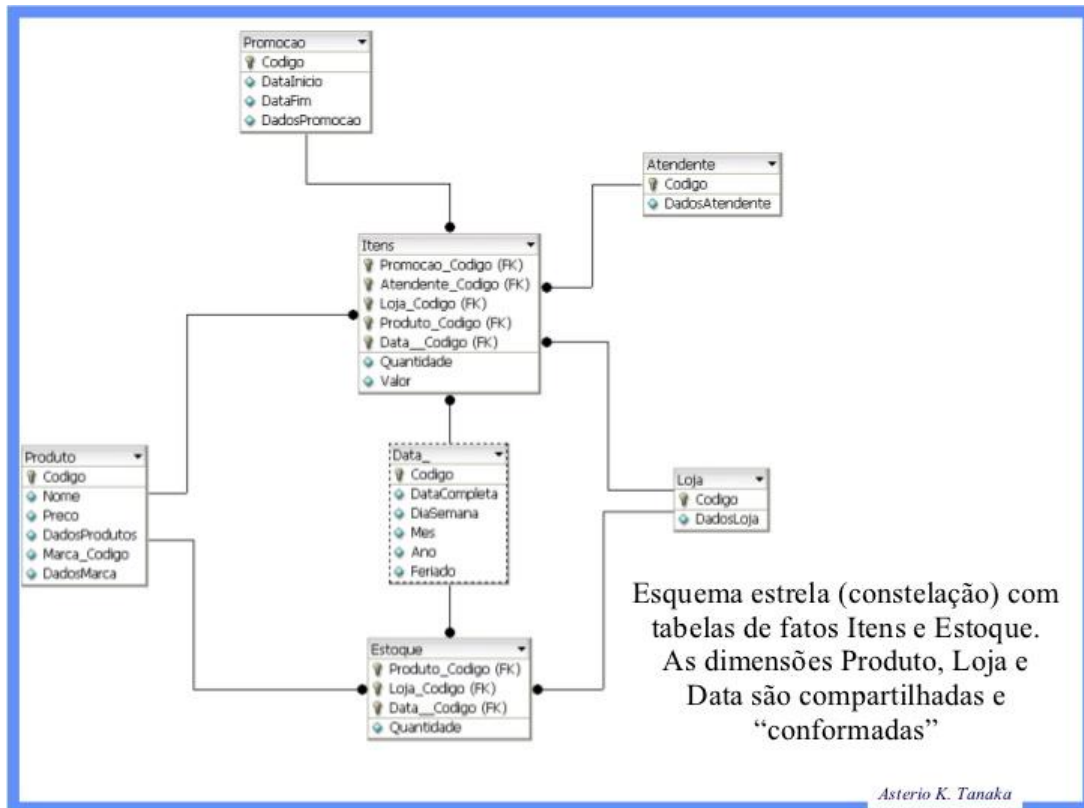


Modelo Dimensional

Snowflake Schema



Modelo Dimensional – Constelação



Exercitando:

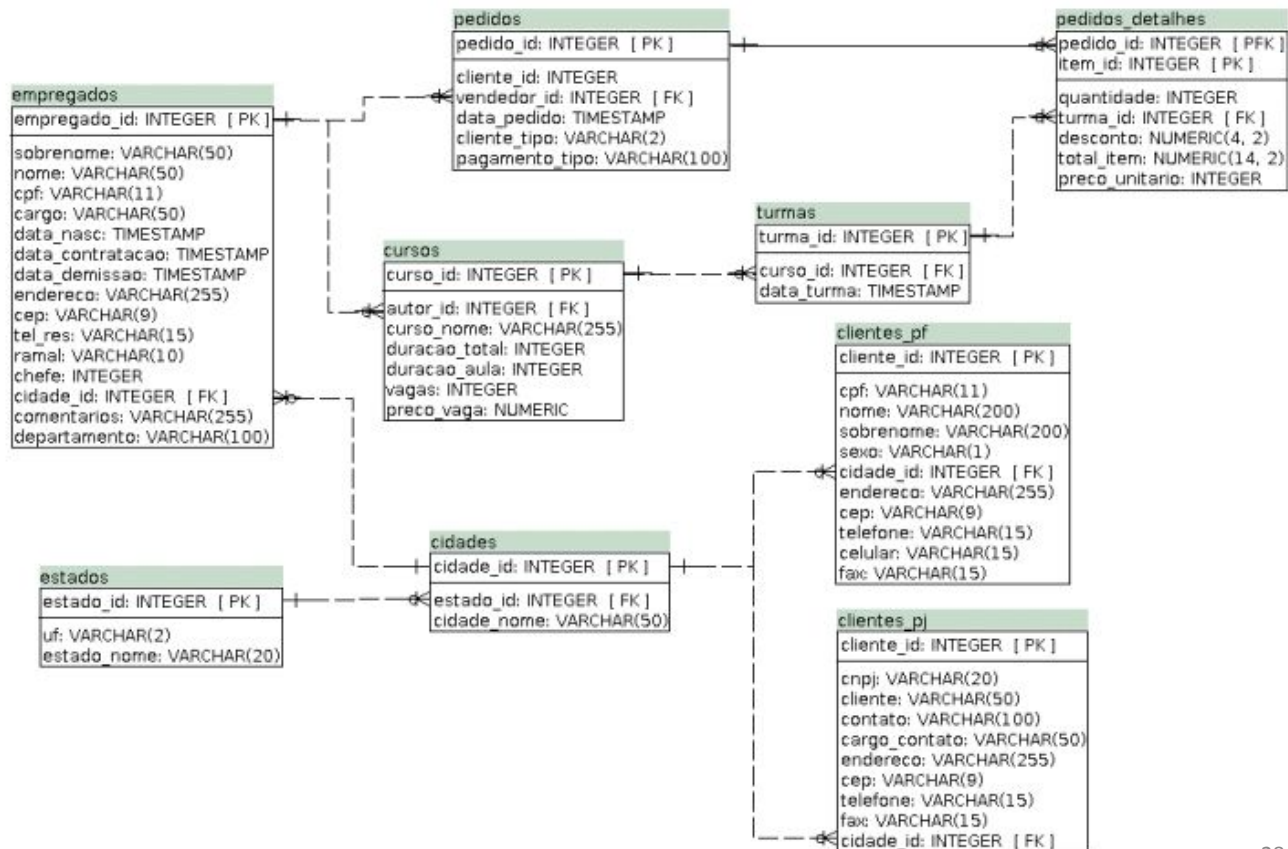
Modelo Relacional Beltrano School

A Beltrano School é uma escola particular, que como uma empresa como tantas outras, cresceu usando uma aplicação *customizada* para atender suas necessidades. Estes casos, normalmente, contam com algum sistema que registra as vendas em tabelas mais ou menos bagunçadas, um aspecto embutido no banco.

Crie um **modelo dimensional lógico** que responda aos seguintes KPI's, baseado no modelo da aplicação customizada.

- Valor vendido por mês
- Valor vendido por vendedor
- Valor vendido por curso
- Valor vendido por cidade

Modelo Relacional Beltrano School



Modelagem Relacional x Modelagem Dimensional

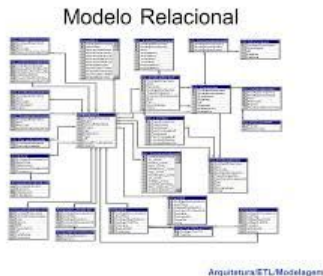
Sistemas e Modelos de Dados

Sistemas Transacionais

Formulário de cadastro de cliente para a loja NOME FANTASIA. O formulário contém campos para Nome Fantasia, CPF, Data de Nascimento, Data de Emissão, e uma tabela para registrar informações adicionais. No topo, há um cabeçalho com o nome da loja e o endereço. No rodapé, há uma seção para o nome e o endereço do cliente.



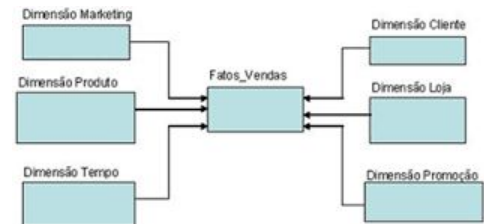
Modelo de Dados Relacional



Sistemas Analíticos



Modelo de Dados Dimensional



Modelo Relacional: Sistemas Transacionais

Regras de negócio (Conceitos):

1 – O cadastro dos usuários com endereço completo inclusive com a informação do curso do usuário e data de inicio de termino previsto.

2 - O cadastro das obras da biblioteca com a devida classificação: científicos, periódicos científicos, periódicos informativos, periódicos diversos, entretenimento, etc.

3- Língua que se encontra a obra

4 - Mídias que se encontra a obra

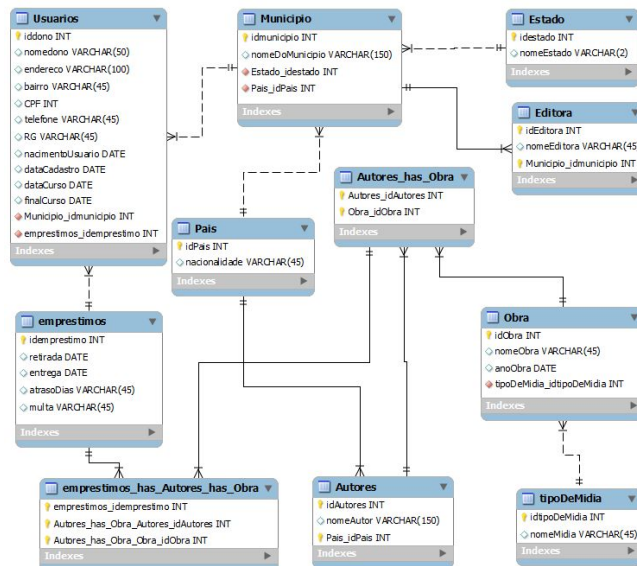
5 - Os autores da obra com o controle da nacionalidade do autor

6 - As editoras dos exemplares com o ano de edição

7- Os usuários estão classificados entre alunos, professores e funcionários. Sendo que para os professores não serão considerados limites de volumes para empréstimos, nem possíveis atrasos na devolução desses.

8 - histórico dos empréstimos.

Modelo de Dados para um Sistema Transacional de controle de Biblioteca



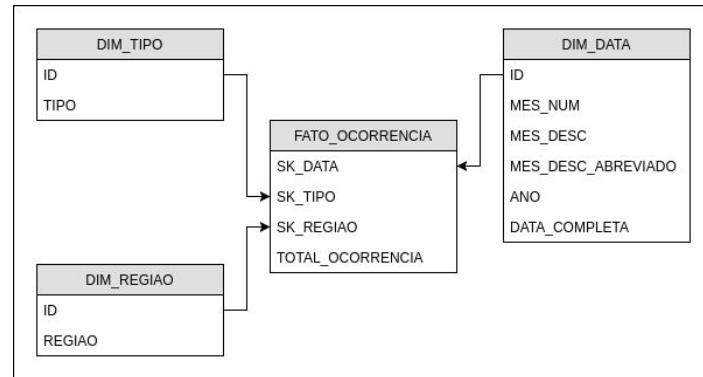
<http://www.webeder.com.br/2011/06/modelo-er.html>

Modelo Dimensional: Sistemas Analíticos

Regras de negócio (Conceitos):

- 1 – Total de Ocorrência por tipo
- 2 – Total de ocorrência por região
- 3 – Total de ocorrência por data
- 4 – Total de ocorrência por tipo e data
- 5 – Total de ocorrência por região e data
- 6 – Média de ocorrência por mês
- 7 – Média de ocorrência por dia
- 8- Média de ocorrência por região e por ano

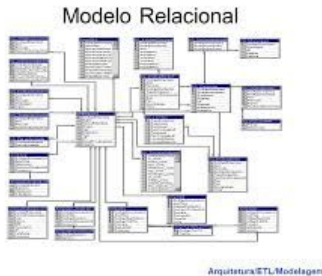
Modelo de Dados para um Sistema Dimensional de análises de ocorrência



<http://www.danielinternet.com.br/dashboard-violencia-contra-mulheres-parte-2-populando-dimensoes-do-dw-com-pentaho/>

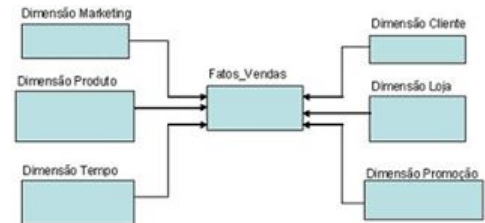
Como transformar transações em indicadores de negócio?

Modelo de Dados Relacional



ETL

Modelo de Dados Dimensional

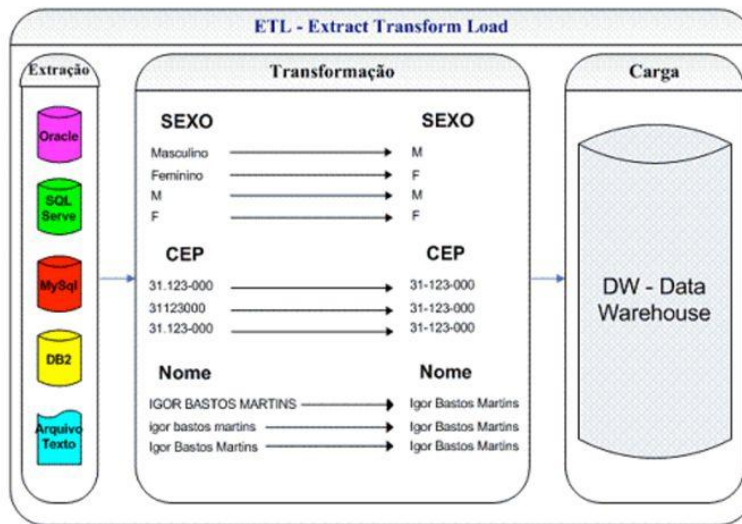


1. **E**xtraindo os dados do modelo origem
2. **T**ransformando os dados para o modelo destino
3. **C**arregando os Dados

Regras de transformação de dados

Realizado por TI, porém
direcionado pelo negócio.

ETL- EXTRAÇÃO DOS DADOS

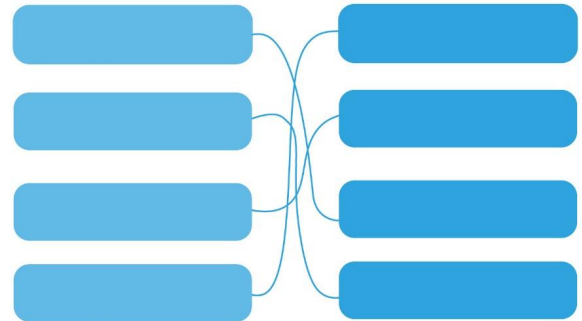


Data Mapping: Mapeamento de dados



- Atividade realizada durante o desenho do processo de ETL, responsável por aplicar as regras necessárias e traduzir dados em informações.

- Selecionar os banco de dados
- Selecionar as entidades necessárias
- Selecionar os atributos necessários



É importante que o profissional conheça o modelo de destino e as informações necessárias para ir em busca dos dados de origem.

Exemplo de um template de projeto

[illegible]

Como

?

ETL - Mapeamento de Dados

Origem					
Nome Origem	Nome Físico do Campo	Tipo	Tam	Null	PK



Destino										
Nome Destino	Campo Destino	Nome Semântico	Tipo	Tam	Null	PK / FK	SK	SCD	SUBST	Transformação

Regras técnicas ou funcionais

Conversão de Base de Dados Relacional para Dimensional

Relacional

Dimensional

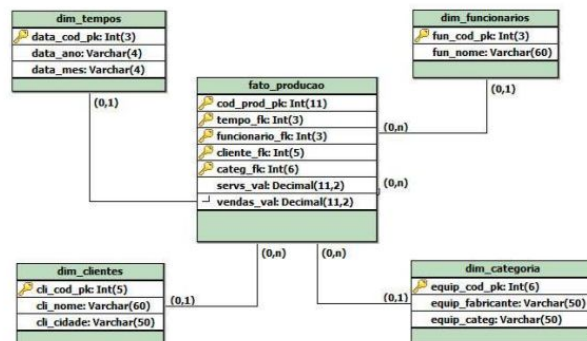
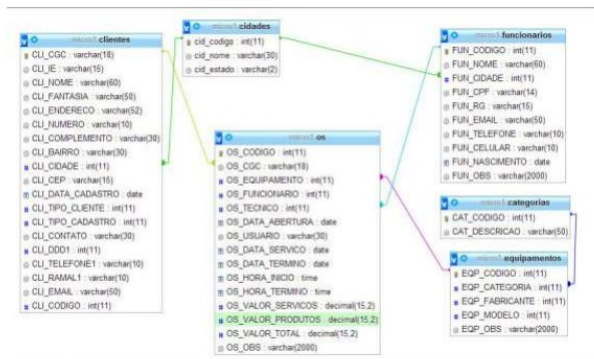


Tabela 2. Quantidade de linhas nas tabelas do Banco Relacional.

Tabela	Linhas
Categorias	33
Cidades	60
Clientes	235
Funcionários	40
Equipamentos	4.611
Os	201.298

Fonte : Carlos Henrique Cardoso e Robert David Nebo

Questão	Faturamento por	Relacional (segundos)	Dimensional (segundos)	%
1	Funcionário	22.68	0.09	0.4%
2	Categoria	22.4	0.03	0.1%
3	Clientes	19.07	0.06	0.3%

Fonte: Carlos Henrique Cardoso e Robert David Nebo

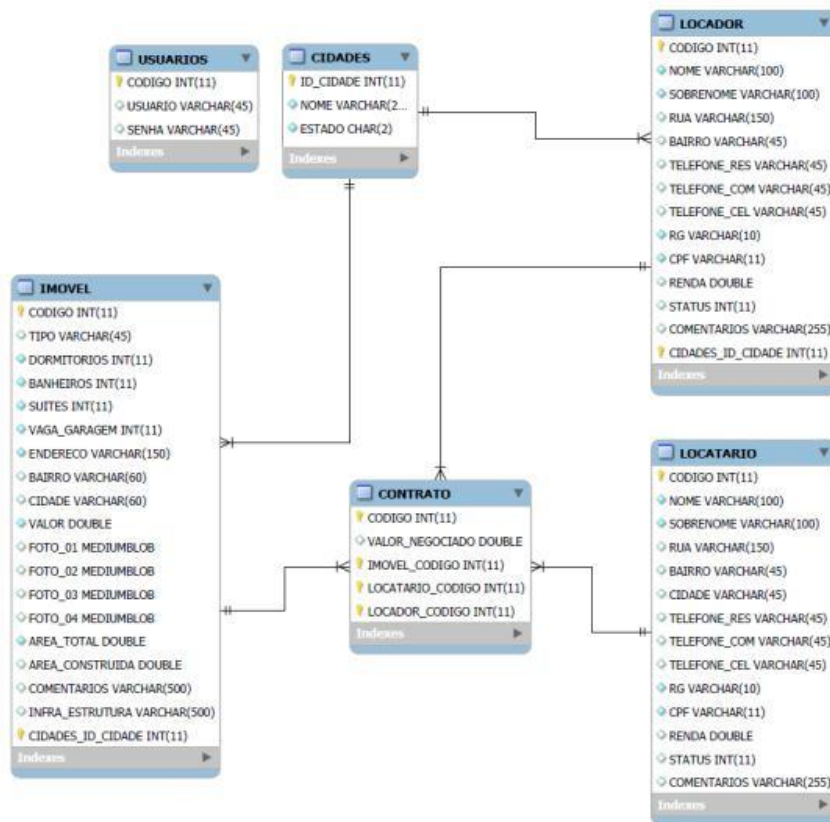
Exercitando:

Modelo Relacional Locação de Imóveis

Desenvolver no Data Modeler um **modelo dimensional lógico (Star Schema)** que responda às seguintes perguntas de negócio, baseado no modelo de dados do sistema transacional de Locação de imóveis.

- Quantidade de imóveis locados da cidade de São Paulo
- Valor da locação maior que \$1000
- Quantidade de imóveis locados no último mês
- Quantidade de imóveis locados para locatários maiores de 40 anos

DER – Sistema de Locação

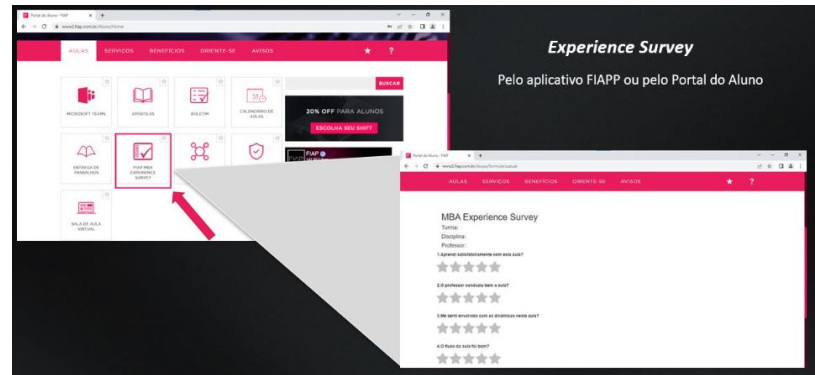
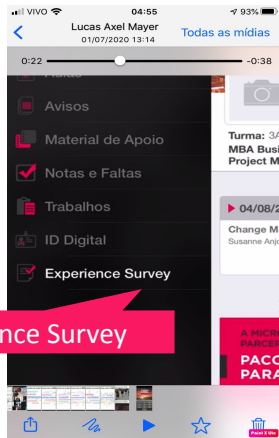


O que você achou da aula de hoje?

FIAP MBA+

Pelo aplicativo da FIAPP

Portal do Aluno



Pelo link: <https://fiap.me/Pesquisa-MBA>

FIAP