

Estudo de caso 1(Parte 1) - Definindo estratégias de vendas

Você é Analista de Dados na empresa Ponto Maximo, uma rede de varejo que vende produtos eletrônicos e eletrodomésticos com lojas espalhadas por diversas cidades do Brasil. A empresa começou sua operação no Brasil em 2012 e atua nos quatro estados da região sudeste mais os estados do Paraná e Bahia.

A empresa está montando a estratégia de vendas para o próximo ano e precisa saber qual dos fabricantes dos produtos vendidos, apresenta melhor desempenho nas vendas. O objetivo é descartar os fabricantes cujos produtos possuem poucas vendas e tentar negociar melhores condições com os principais fabricantes.

Em paralelo a isso, a empresa gostaria de ter diferentes visões das vendas realizadas nos últimos 4 anos (período de 2012 a 2015). Deve ser possível segmentar os relatórios de vendas por diferentes informações e por diferentes ângulos. Estas informações irão suportar as estratégias da empresa para o próximo ano.

Sua fonte de dados é um arquivo Excel com dados coletados do sistema de vendas, CRM (Gerenciamento de Relacionamento com o Cliente) e ERP (Planejamento de Recursos Corporativos) da empresa. O conjunto de dados foi entregue pelo departamento de TI com as seguintes colunas

Coluna	Descrição
ID-Produto	Identificador único de cada produto
Produto	Nome do produto
Categoria	Categoria do produto
Segmento	Segmento do produto
Fabricante	Fabricante do produto
Loja	Loja onde foi efetuada a venda
Cidade	Cidade da loja onde foi efetuada a venda
Estado	Estado da loja onde foi efetuada a venda
Vendedor	Nome do vendedor
ID-Vendedor	ID-Vendedor
DataVenda	Data da venda
ValorVenda	Valor da venda

Haverá diversas reuniões para definição da estratégia de vendas e os relatórios poderão ser extraídos sob demanda, de acordo com a necessidade dos gestores. Por conta disso, você deve criar um modelo de dados que permita a extração de relatórios a qualquer momento e que permita extrair dados por diferentes visões e ângulos.

2.1 Estudo de caso 1(Parte 2) - Compreendendo o problema

Empresas existem para fornecer produtos ou serviços. Até mesmo instituições sem fins lucrativos, necessitam de alguém por trás para custear seus gastos, pois nada é realmente de graça.

Na maioria dos casos a empresa já tem os dados necessários para as análises, sendo necessário capturar esses dados de diversas fontes e de forma automatizada.

2.2 Estudo de caso 1 (Parte 3) - Extração de dados

Entre na pasta compartilhada e baixe os arquivos referentes ao Estudo de Caso.

Abra a planilha de vendas.xls e observe sua fonte de dados:

	A	B	C	D	E	F	G	H	I	J	K	L
1	ID-Produto	Produto	Categoria	Segmento	Fabricante	Loja	Cidade	Estado	Vendedor	ID-Vendedor	Data Venda	Valor Venda
2	SKU-0000001	LG K10 TV Power	Celulares	Corporativo	LG	SP8821	São Paulo	São Paulo	Ana Teixeira	1009	04/10/2012	R\$ 679,00
3	SKU-0000002	Geladeira Duplex	Eletrodomésticos	Doméstico	Brastemp	SP8821	São Paulo	São Paulo	Josias Silva	1006	01/01/2012	R\$ 832,00
4	SKU-0000003	Lavadora 11 Kg	Eletrodomésticos	Doméstico	Brastemp	SP8821	São Paulo	São Paulo	Josias Silva	1006	02/02/2012	R\$ 790,00
5	SKU-0000004	Lavadora 11 Kg	Eletrodomésticos	Doméstico	Brastemp	SP8821	São Paulo	São Paulo	Mateus Gonçalves	1003	03/03/2012	R\$ 765,32
6	SKU-0000005	Lavadora 11 Kg	Eletrodomésticos	Doméstico	Electrolux	SP8821	São Paulo	São Paulo	Artur Moreira	1004	04/04/2012	R\$ 459,89
7	SKU-0000006	Lavadora 11 Kg	Eletrodomésticos	Doméstico	Brastemp	SP8821	São Paulo	São Paulo	Rodrigo Fagundes	1005	04/05/2012	R\$ 590,98
8	SKU-0000007	Geladeira Duplex	Eletrodomésticos	Doméstico	Brastemp	SP8821	São Paulo	São Paulo	Josias Silva	1006	04/06/2012	R\$ 1.000,91
9	SKU-0000008	Geladeira Duplex	Eletrodomésticos	Doméstico	Electrolux	A9990	Belo Horizonte	Minas Gerais	Fernando Zambrini	1007	04/07/2012	R\$ 1.229,00
10	SKU-0000009	Geladeira Duplex	Eletrodomésticos	Doméstico	Brastemp	SP8821	São Paulo	São Paulo	Mateus Gonçalves	1003	04/08/2012	R\$ 1.300,00
11	SKU-0000010	Geladeira Duplex	Eletrodomésticos	Doméstico	Electrolux	SP8821	São Paulo	São Paulo	André Pereira	1002	04/09/2012	R\$ 1.290,00
12	SKU-0000011	Geladeira Duplex	Eletrodomésticos	Doméstico	Brastemp	SP8821	São Paulo	São Paulo	Ana Teixeira	1009	04/09/2012	R\$ 1.287,00
13	SKU-0000012	Geladeira Duplex	Eletrodomésticos	Doméstico	Brastemp	A9990	Belo Horizonte	Minas Gerais	Josias Silva	1006	04/11/2012	R\$ 1.651,00
14	SKU-0000013	Geladeira Duplex	Eletrodomésticos	Doméstico	Brastemp	SP8821	São Paulo	São Paulo	Josias Silva	1006	04/11/2012	R\$ 1.100,00
15	SKU-0000014	Geladeira Duplex	Eletrodomésticos	Doméstico	Brastemp	SP8821	São Paulo	São Paulo	Mateus Gonçalves	1003	04/12/2012	R\$ 1.190,00
16	SKU-0000015	Geladeira Duplex	Eletrodomésticos	Doméstico	Brastemp	SP8821	São Paulo	São Paulo	Artur Moreira	1004	04/12/2012	R\$ 1.190,98
17	SKU-0000016	Lavadora 11 Kg	Eletrodomésticos	Doméstico	Electrolux	A9990	Belo Horizonte	Minas Gerais	Rodrigo Fagundes	1005	02/01/2013	R\$ 877,00
18	SKU-0000017	Lavadora 11 Kg	Eletrodomésticos	Doméstico	Brastemp	SP8821	São Paulo	São Paulo	Josias Silva	1006	02/02/2013	R\$ 982,00

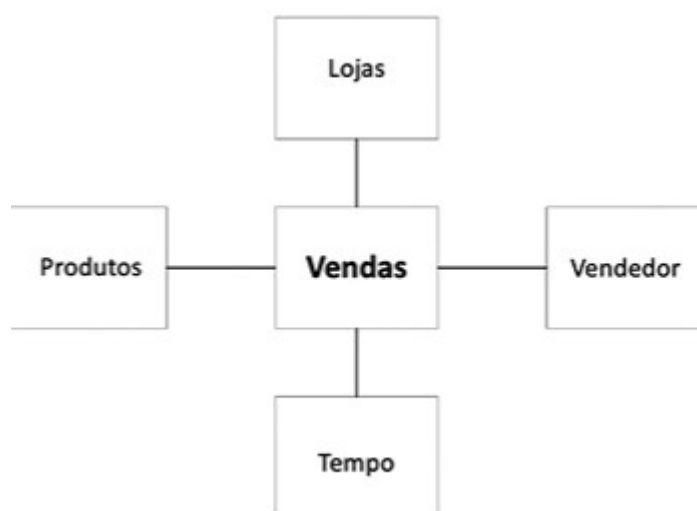
Confira se os títulos das colunas estão de acordo com o apresentado no problema e se existe alguma incompatibilidade com o tipo dos dados. Salve os dados em um repositório padrão de fácil acesso.

Basicamente temos um identificador do produto, o produto, a categoria, o segmento, o fabricante, a loja, a cidade, o estado, o nome do vendedor, o identificador do vendedor, a data da venda e o valor da venda. O identificador torna as linhas únicas, independente dos valores serem iguais nas outras colunas. Para isso a a coluna de identificador deve ser uma chave primária com auto incremento de valores, para não se ter valores repetidos.

2.3 Estudo de caso 1 (Parte 4) - Modelos de organização dos dados

- Posso enviar os dados para o Power BI da forma que estão?
 - Não! É necessário primeiramente tratar os dados e retirar quaisquer inconsistências.
- Quais são os requerimentos de organização dos dados?
 - Para isso é necessário compreender o problema, de preferência com auxílio da área de negócios da empresa. Mas lembre-se não é obrigação da área de negócios entender de organização de dados.
- Como mantenho os dados sobre um formato de atualização sob demanda?
 - O modelo deve prever as atualizações e se manter organizados.
- Como visualizar os dados por diferentes visões em ângulos?
 - Deve-se quebrar os dados, ou seja, organiza-lo em esquemas, de acordo com a necessidade, ou seja, primeiramente cria-se o modelo e depois carrega-se os dados, por se tratar de BI. Em data Science é possível carregar os dados primeiro e modelar depois.

Logo o modelo como é apresentado na planilha do Excel não é o ideal, o que precisamos é um modelo que facilite a visualização dos dados por diferentes ângulos e sob demanda.



Então para cada uma das tabelas seguintes: Lojas, Produtos, Vendedor e Tempo, temos **dimensões** específicas com um identificador para cada uma. Quando elas

se unem, temos um **fato**, ou nesse caso é uma venda. Podemos alterar as tabelas, ou até mesmo adicionar mais um fato.

Com isso podemos subir e descer facilmente na hierarquia dos dados, ou seja, na sua granulometria. **Sendo mais granular dados com maior detalhamento (categorizados por dia, por exemplo).**

- Como definimos o melhor modelo de organização dos dados?

-Devemos identificar claramente o processo de negócios e prever o processo de negócios

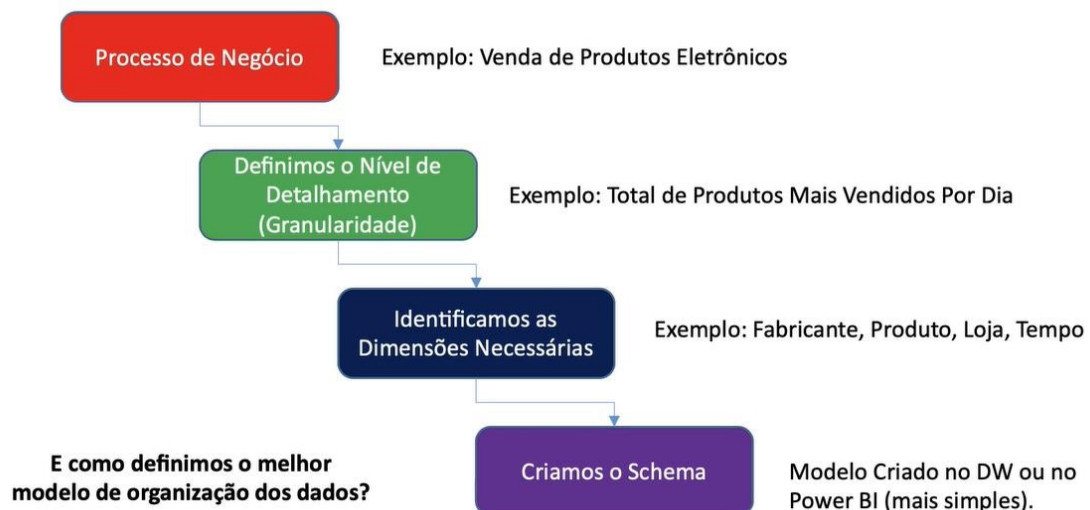
- Definir corretamente o nível de granularidade, que implica na definição do modelo físico (O nível de detalhamento deve atender as necessidades da empresa).

Modelo Lógico: Definição geral da organização dos dados;

Modelo físico: Implementação no banco de dados ou no nosso caso no Power BI.

- Devemos identificar as dimensões necessárias.

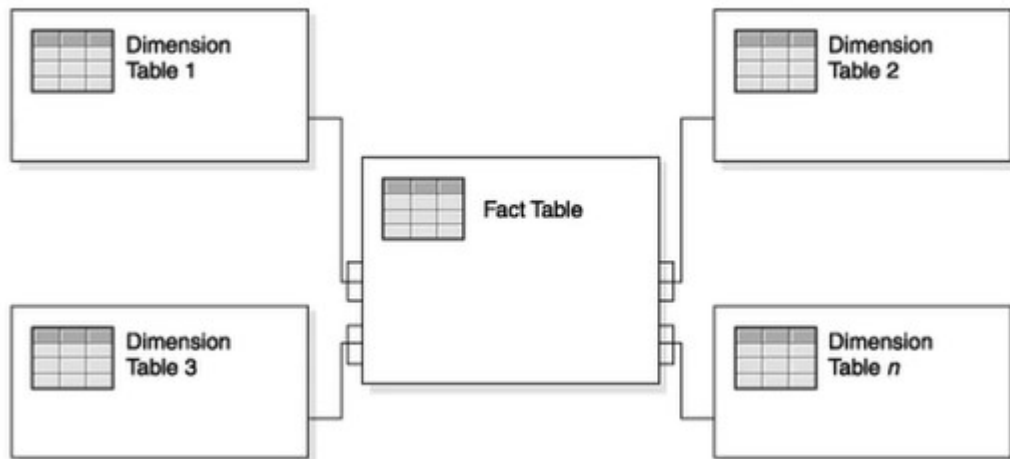
- Por fim determinamos o **Schema (Modelo de organização)**.



- Principais modelos de organização:

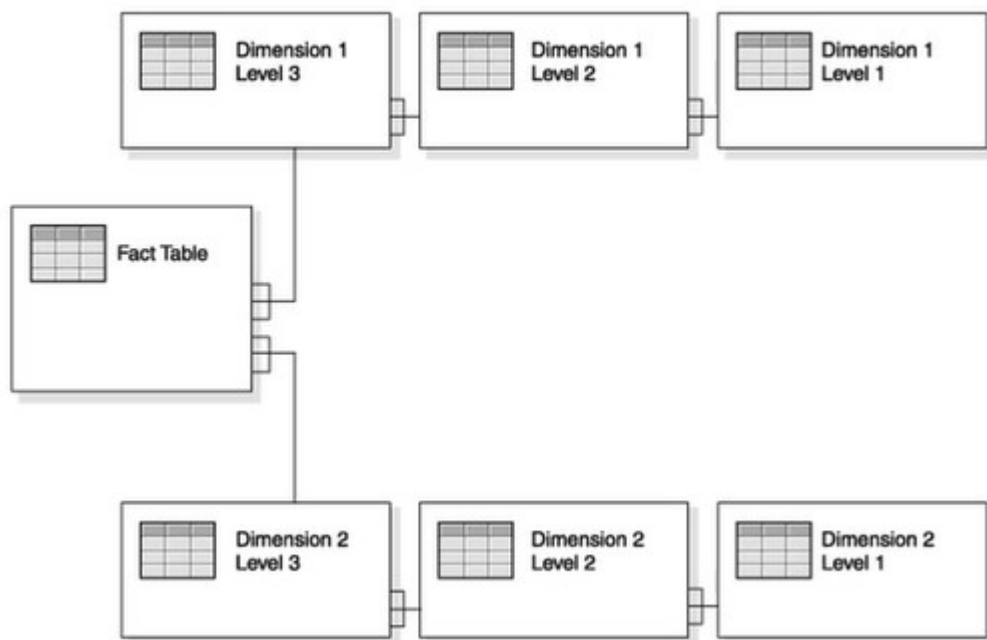
- O modelo **Star Schema** tem tabelas dimensão ligadas a uma única tabela fato diretamente.

Star Schema



O modelo **Snowflake Schema** tem um tabelas dimensão ligadas a tabela fato, e tabelas dimensão ligadas a outras tabelas dimensão. É um modelo mais robusto, porém mais complexo de ser utilizado.

Snowflake Schema



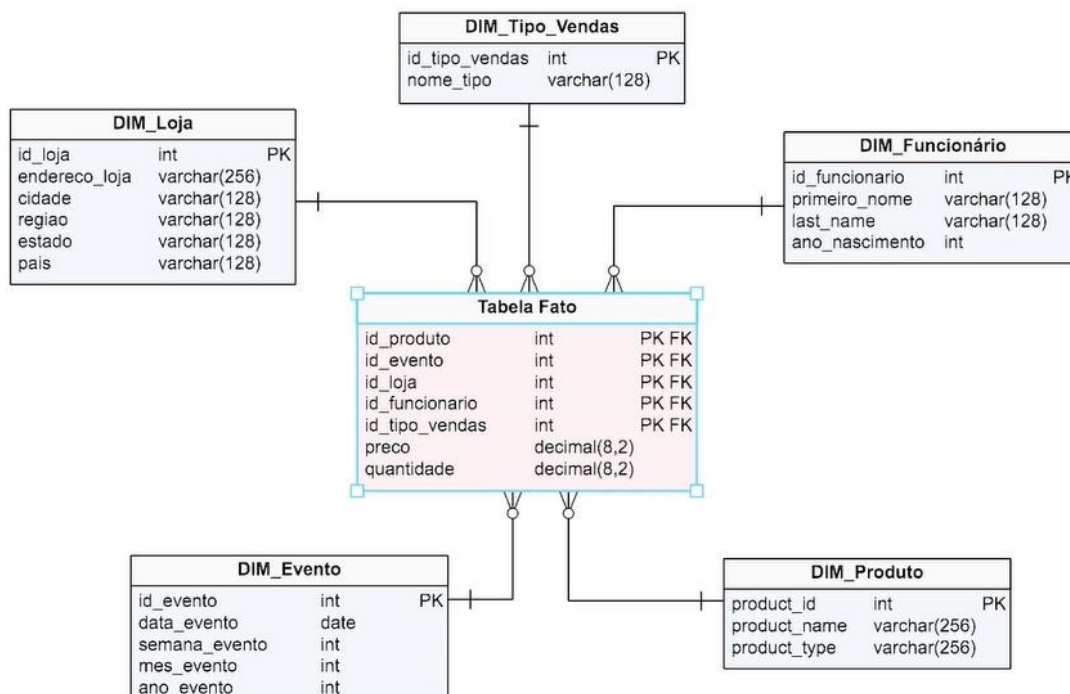
Agora um resumo do que foi apresentado até então:

- Business Intelligence é um conceito que envolve técnicas, ferramentas e procedimentos para transformar dados brutos em informação e conhecimento para os tomadores de decisão.
- O Power BI é uma ferramenta de Self-Service BI, cujo objetivo é permitir que pessoas sem conhecimento técnico avançado possam realizar análises em dados.

- Para trabalhar de forma eficiente e profissional com o Power BI, pode ser necessário criar um modelo de organização dos dados. Esse modelo permite realizar análises por diferentes ângulos, sob demanda e orientada à solução dos problemas de negócio.
- Existem 2 modelos principais de organização dos dados: Star Schema e Snowflake. Em ambos, o objetivo é criar uma estrutura de organização, que facilite a compreensão dos dados.
- Normalmente o modelo de dados é criado em um Data Warehouse (DW), um banco de dados que serve de fonte para processos de BI. As empresas extraem dados de sistemas transacionais, como CRM ou ERP, consolidam os dados e carregam no DW seguindo o modelo adotado. As ferramentas de BI podem conectar no DW e extrair os dados para análise.
- Para o Estudo de Caso deste capítulo, um modelo Star Schema não seria obrigatório, e nosso objetivo aqui é demonstrar como isso pode ser feito no Power BI.

2.5 Estudo de caso 1 (Parte 5) - Criando Tabelas Dimensão e tabela Fato

Normalmente o modelo star schema é implementado em um Data Warehouse, mas para este estudo de caso iremos implementar diretamente no Power BI.



As tabelas dimensão tem o pré-fixo DIM. Neste caso temos uma dimensão evento, uma dimensão produto, uma dimensão funcionário, uma dimensão loja e uma dimensão vendas. Para saber se uma tabela é uma dimensão, pergunte se seu problema necessita dele. Por exemplo, uma venda precisa de um produto? Se sim, o produto é uma

dimensão da venda. A tabela **fato** é o que estamos medindo no final das contas, unindose as dimensões.

Agora se atende aos símbolos dentro das tabelas:

- PK: Primary Key (Chave Primária)
- FK: Foreign Key (Chave Estrangeira)

Uma chave primária é um registro único, como por exemplo um cpf. A chave estrangeira é uma chave primária representada em outra tabelas, que no caso é a tabela fato, que reuni todas essas informações.