



Introdução a Banco de Dados e Linguagem SQL

Data Warehouse

Agenda

- ❑ Modelagem Dimensional
- ❑ Consultas OLAP



Modelagem Dimensional

- ❑ A modelagem dimensional é uma técnica de projeto lógico para estruturar os dados de maneira que sejam intuitivos para os usuários do negócio e para que forneçam um alto desempenho de acesso aos dados.
- ❑ A fim de fornecer subsídios aos relatórios de negócio e às necessidades analíticas, a modelagem dimensional requer uma abordagem diferente daquelas usadas no projeto de bancos de dados transacionais.
- ❑ O modelo dimensional é composto de dois tipos de tabelas: **fato** e **dimensão**.

Modelagem Dimensional

- ❑ A tabela fato é a tabela central do modelo dimensional, armazena medições numéricas do negócio, na qual cada uma das medições é obtida pela interseção das dimensões, possui chaves de várias dimensões e deve conter o grão mais detalhado capturado a partir do processo de negócio.
- ❑ A tabela dimensão é uma entidade independente em um modelo dimensional que serve como mecanismo para aplicar o recurso de separação e combinação das métricas da tabela fato em um modelo dimensional.

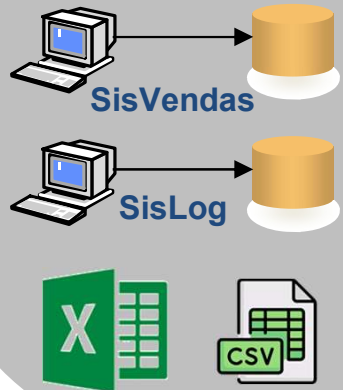
Modelagem Dimensional

- ❑ O processo de modelagem dimensional pode ser resumido em quatro passos:
 - ❑ **Escolher o Processo de Negócio:** prioriza os processos mais importantes da organização.
 - ❑ **Declarar o Grão:** deve-se escolher o menor grão (maior nível de detalhamento) possível. Cada tabela fato deve ter um único grão.
 - ❑ **Identificar as Dimensões:** a partir da escolha do grão, é possível escolher as dimensões compatíveis.
 - ❑ **Identificar os Fatos:** escolhe-se as métricas aplicáveis ao processo selecionado.

Escolher o Processo de Negócio

- ❑ Vamos utilizar o processo de negócio relacionado a vendas de uma empresa de varejo fictícia denominada “Best Sale” com filiais em vários estados brasileiros.
- ❑ Serão considerados dados internos e externos a essa organização para a criação e carga do modelo dimensional.

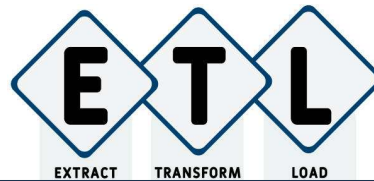
Fontes Internas



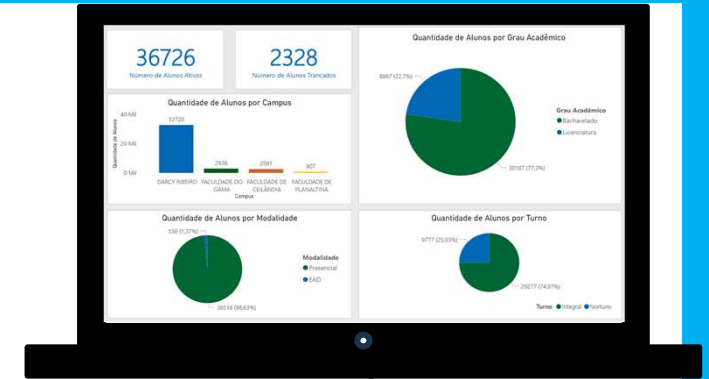
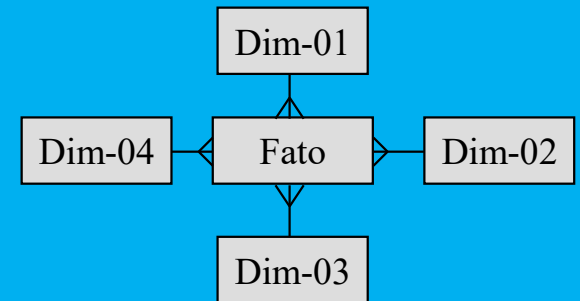
Fontes Externas



Área de organização de dados



Data Warehouse

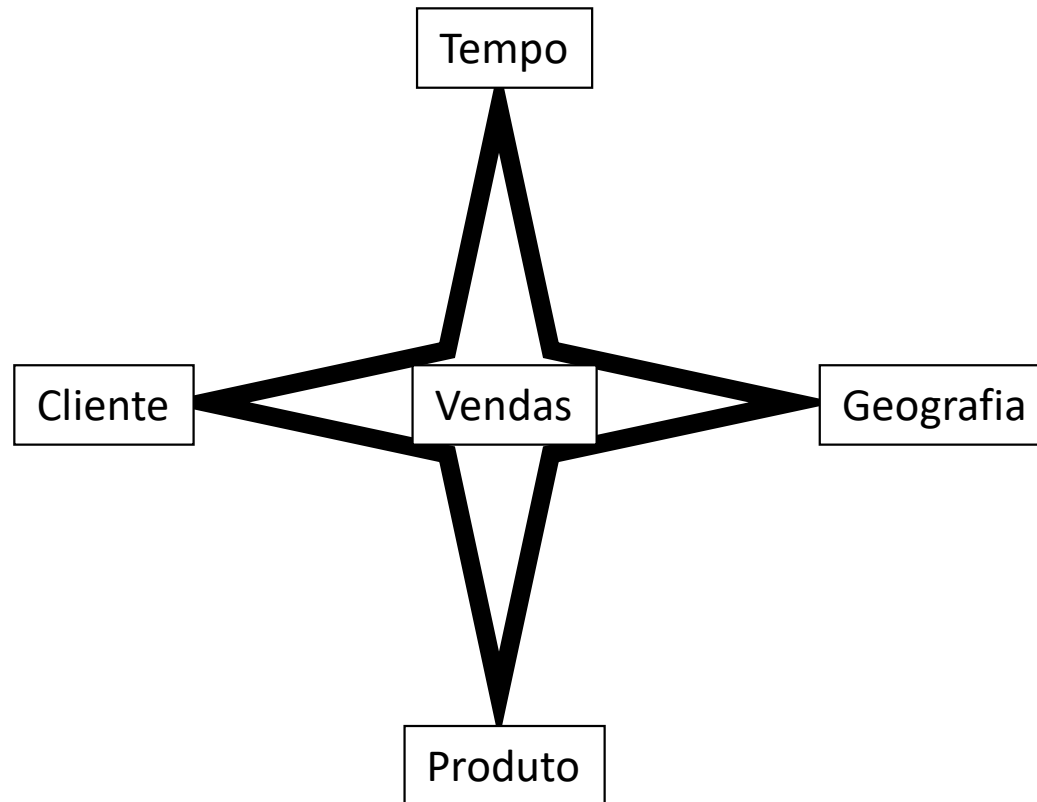


Declarar o Grão

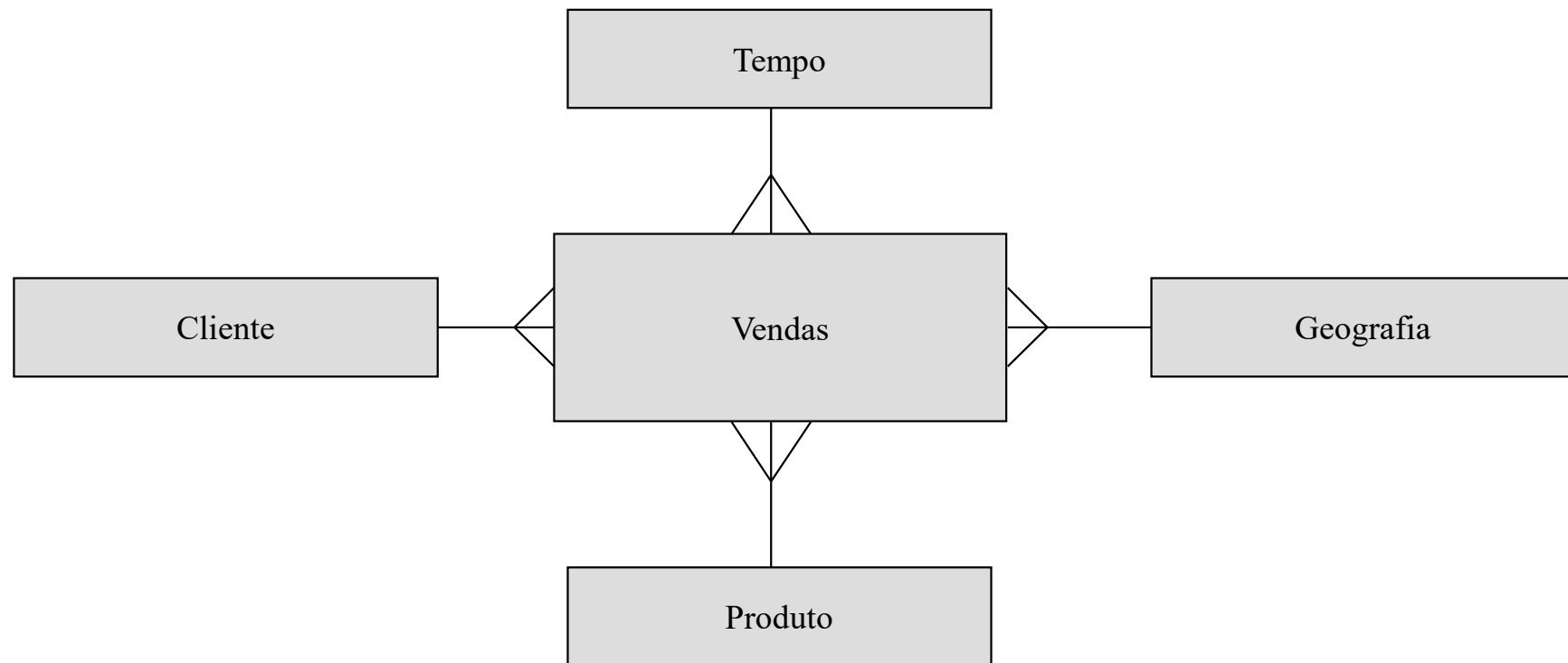
- Uma linha de uma nota fiscal emitida pela empresa:

Empresa Best Sale					
Nº da Venda:		1	Data da Venda:		05/12/2022
CPF Cliente:		111	Nome Cliente:		Ana Maria
Matrícula Vendedor:		852	Nome Vendedor:		José Silva
Nr Ordem	Código	Produto	Preço	Qtd	Total
1	2222	Impressora	450,00	3	1.350,00
2	3333	Fogão	650,32	2	1.300,64
3	4444	Televisão	810,65	1	810,65
Total					R\$ 3.461,29

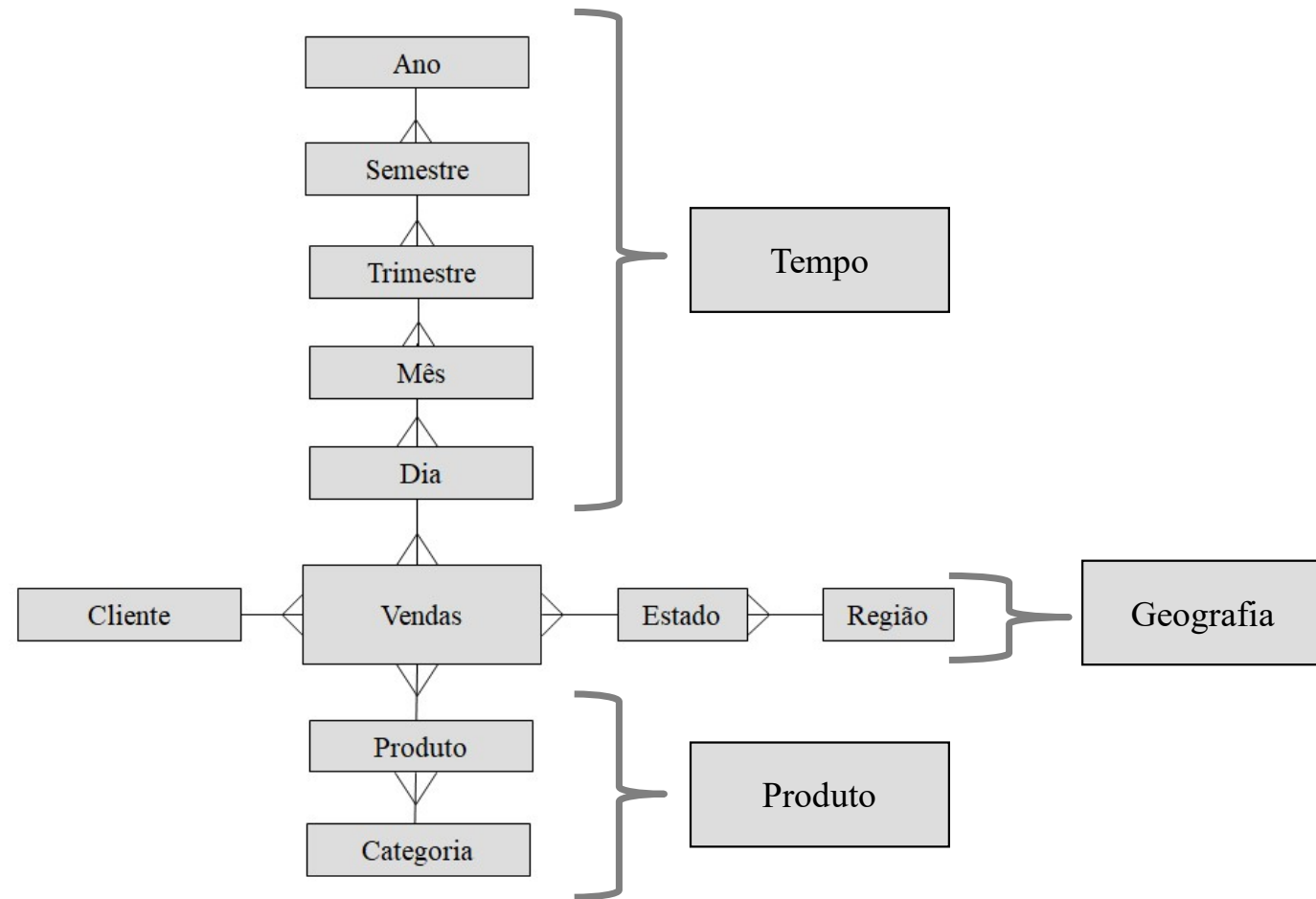
Identificar as Dimensões e Fatos



Identificar as Dimensões e Fatos



Identificar as Dimensões e Fatos



pk_cliente	codigo	sexo	faixa_etaria	nivel_instrucao	renda_mensal
1	1001	M	1	1	1
2	1002	M	2	2	2
3	1003	M	3	3	3
4	1004	M	4	4	4
5	1005	M	5	1	1

Dim_Cliente

pk_tempo	data	ano	mes	dia	trimestre	nome_dia	nome_mes	semestre
1	2022-01-01	2022	1	1	1	Sábado	Janeiro	1
2	2022-01-02	2022	1	2	1	Domingo	Janeiro	1
3	2022-01-03	2022	1	3	1	Segunda	Janeiro	1
4	2022-01-04	2022	1	4	1	Terça	Janeiro	1
5	2022-01-05	2022	1	5	1	Quarta	Janeiro	1

Dim_Tempo

pk_dim_tem	pk_dim_geo	pk_dim_prod	pk_dim_cli	preco_unitario	quantidade	Valor
3	1	2	4	3500.00	3	10500.00
201	2	3	5	4000.00	2	8000.00
173	3	2	5	3500.00	3	10500.00
353	3	1	4	5000.00	2	10000.00
40	4	5	4	5000.00	3	15000.00

Fato_Vendas

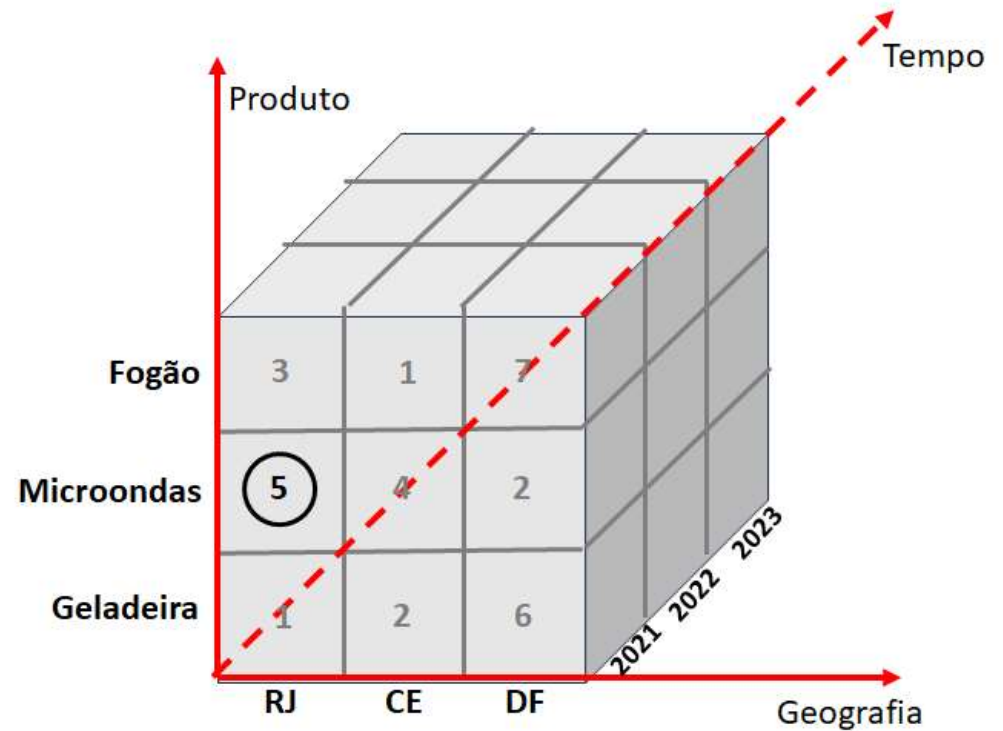
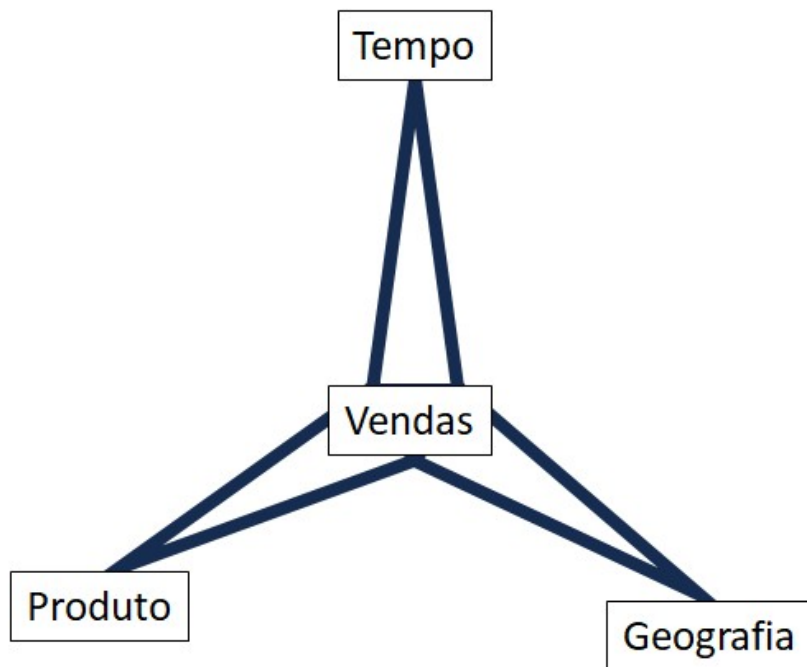
pk_produto	nome	marca	categoria	preco
1	PC	Apple	Informática	5000.00
2	PC	Dell	Informática	3500.00
3	PC	HP	Informática	4000.00
4	Notebook	Apple	Informática	8000.00
5	Notebook	Dell	Informática	5000.00

Dim_Produto

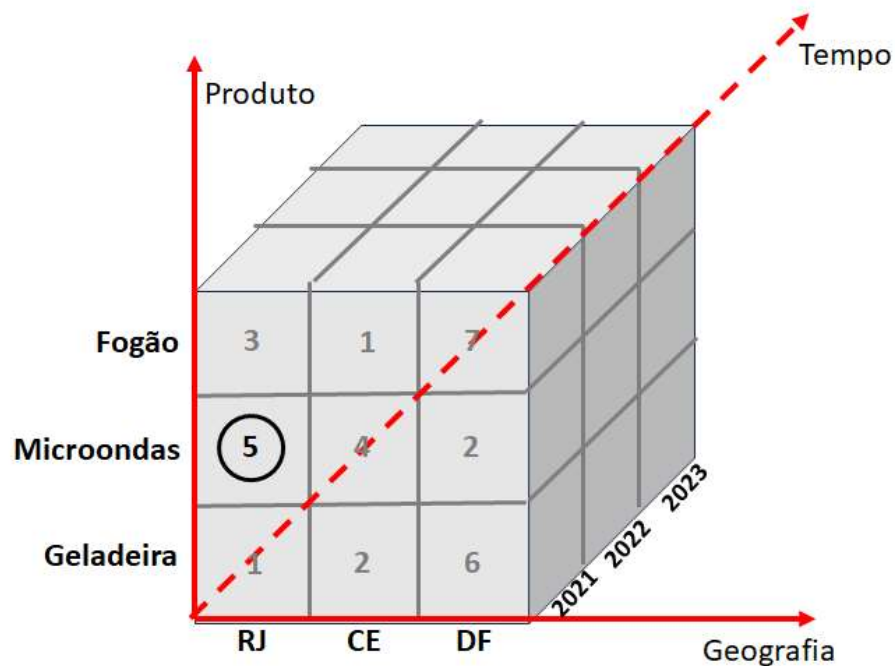
pk_geografia	sicla	uf	nome_uf	sicla_regiao	nome_regiao
1	RO	Rondônia	N	Norte	Norte
2	AC	Acre	N	Norte	Norte
3	AM	Amazonas	N	Norte	Norte
4	RR	Roraima	N	Norte	Norte
5	PA	Pará	N	Norte	Norte

Dim_Geografia

Modelo Estrela x Cubo



Relatório gerado a partir do cubo

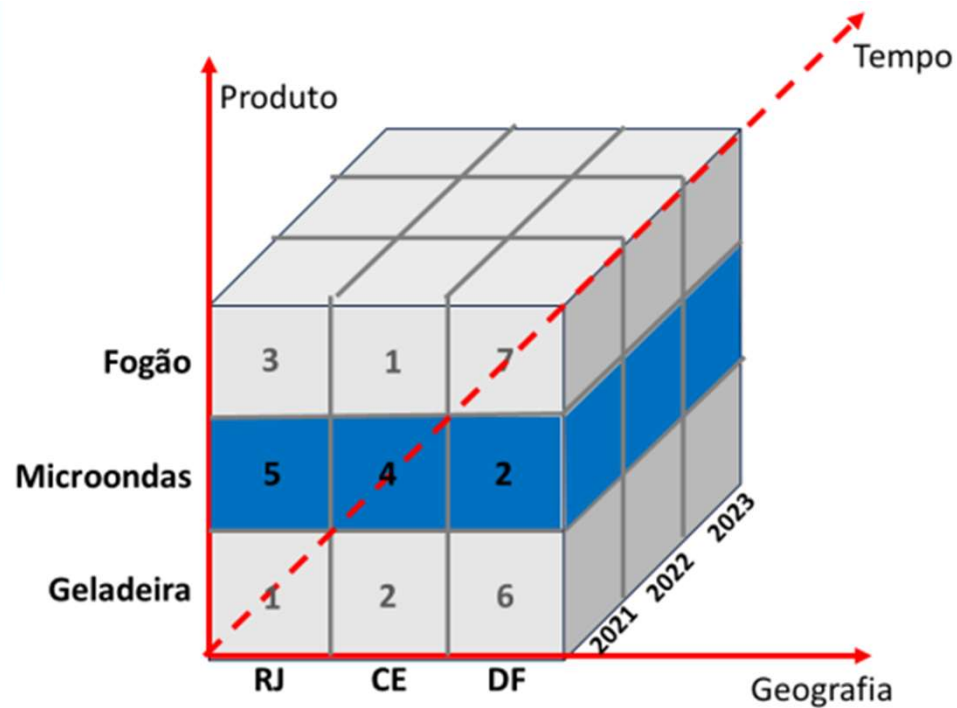


	2021			2022			2023		
	RJ	CE	DF	RJ	CE	DF	RJ	CE	DF
Fogão	3	1	7	2	4	2	3	7	9
Microondas	5	4	2	4	9	7	5	6	1
Geladeira	1	2	6	2	1	4	5	7	6

Consultas OLAP

- ❑ As consultas OLAP mais comuns são:
 - ❑ **Slice-and-Dice (fatiar e cortar):** realizar a operação de projeção nas dimensões.
 - ❑ **Pivot (pivoteamento):** rotação do cubo ou hipercubo.
 - ❑ **Roll-up ou Drill-up (agregação):** os dados são resumidos com generalização crescente (dia, mês, ano).
 - ❑ **Drill-down (desmembramento):** níveis crescentes de detalhes são revelados (ano, mês, dia).

Slice

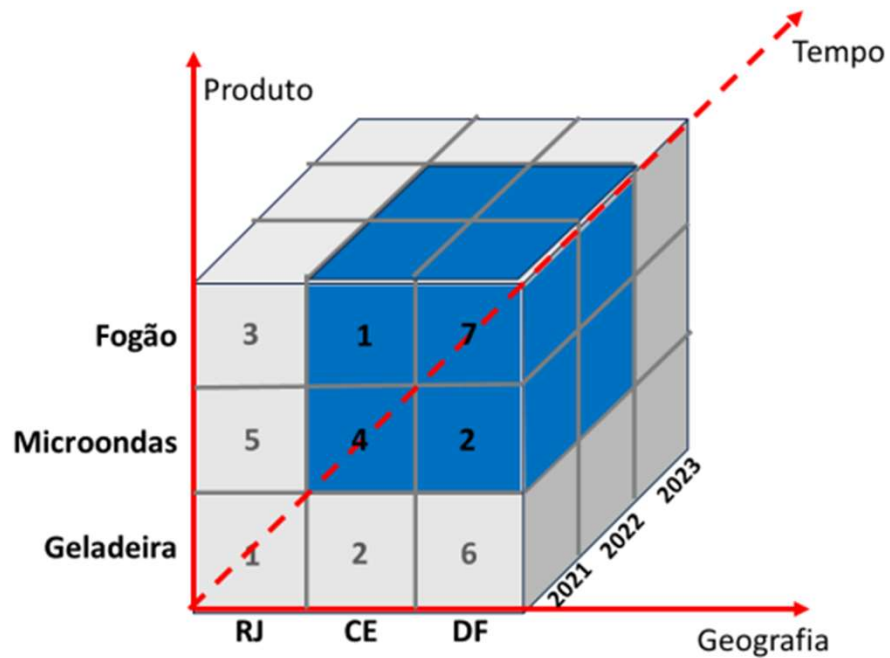


	2021			2022			2023		
	RJ	CE	DF	RJ	CE	DF	RJ	CE	DF
Fogão	3	1	7	2	4	2	3	7	9
Microondas	5	4	2	4	9	7	5	6	1
Geladeira	1	2	6	2	1	4	5	7	6



	2021			2022			2023		
	RJ	CE	DF	RJ	CE	DF	RJ	CE	DF
Microondas	5	4	2	4	9	7	5	6	1

Dice

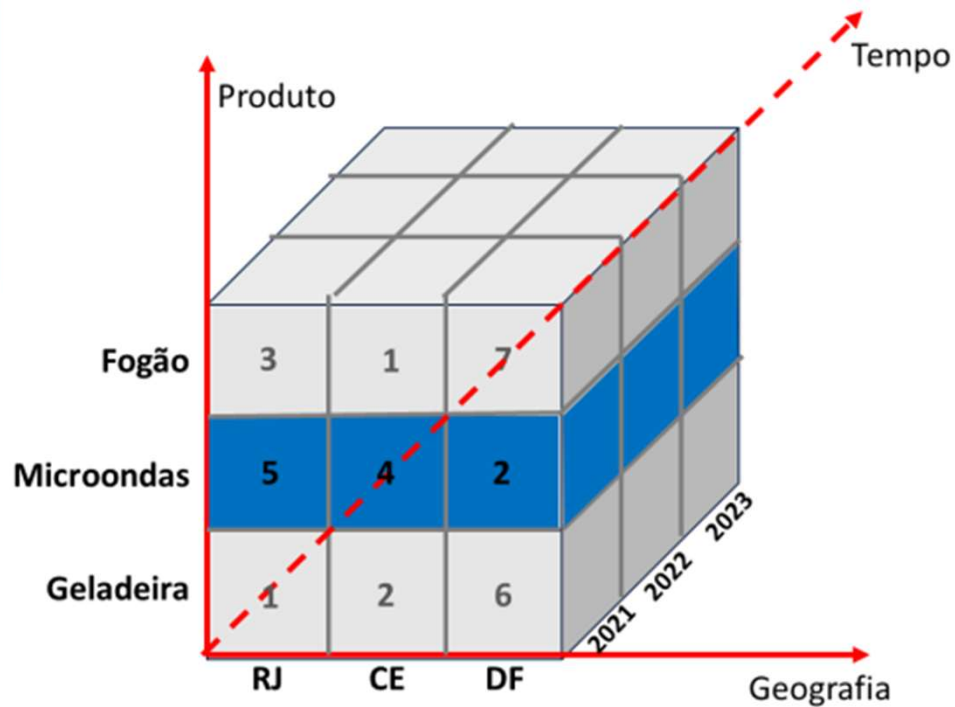


	2021			2022			2023		
	RJ	CE	DF	RJ	CE	DF	RJ	CE	DF
Fogão	3	1	7	2	4	2	3	7	9
Microondas	5	4	2	4	9	7	5	6	1
Geladeira	1	2	6	2	1	4	5	7	6



	2021		2022	
	CE	DF	CE	DF
Fogão	1	7	4	2
Microondas	4	2	9	7

Slice-and-Dice



	2021			2022			2023		
	RJ	CE	DF	RJ	CE	DF	RJ	CE	DF
Fogão	3	1	7	2	4	2	3	7	9
Microondas	5	4	2	4	9	7	5	6	1
Geladeira	1	2	6	2	1	4	5	7	6



	2021			2022			2023		
	RJ	CE	DF	RJ	CE	DF	RJ	CE	DF
Microondas	5	4	2	4	9	7	5	6	1

Pivot

	2021			2022			2023		
	RJ	CE	DF	RJ	CE	DF	RJ	CE	DF
Fogão	3	1	7	2	4	2	3	7	9
Microondas	5	4	2	4	9	7	5	6	1
Geladeira	1	2	6	2	1	4	5	7	6



	2021			2022			2023		
	Fogão	Microondas	Geladeira	Fogão	Microondas	Geladeira	Fogão	Microondas	Geladeira
RJ	3	5	1	2	4	2	3	5	5
CE	1	4	2	4	9	1	7	6	7
DF	7	2	6	2	7	4	9	1	6

Roll-up ou Drill-up

		2021			2022			2023		
		Fogão	Microondas	Geladeira	Fogão	Microondas	Geladeira	Fogão	Microondas	Geladeira
Sul	PR	3	5	1	2	4	2	3	5	5
	SC	1	4	2	4	9	1	7	6	7
	RS	7	2	6	2	7	4	9	1	6



		2021			2022			2023		
		Fogão	Microondas	Geladeira	Fogão	Microondas	Geladeira	Fogão	Microondas	Geladeira
Sul		11	11	9	8	20	7	19	12	18

Drill-down

	2021			2022			2023		
	Fogão	Microondas	Geladeira	Fogão	Microondas	Geladeira	Fogão	Microondas	Geladeira
Sul	11	11	9	8	20	7	19	12	18

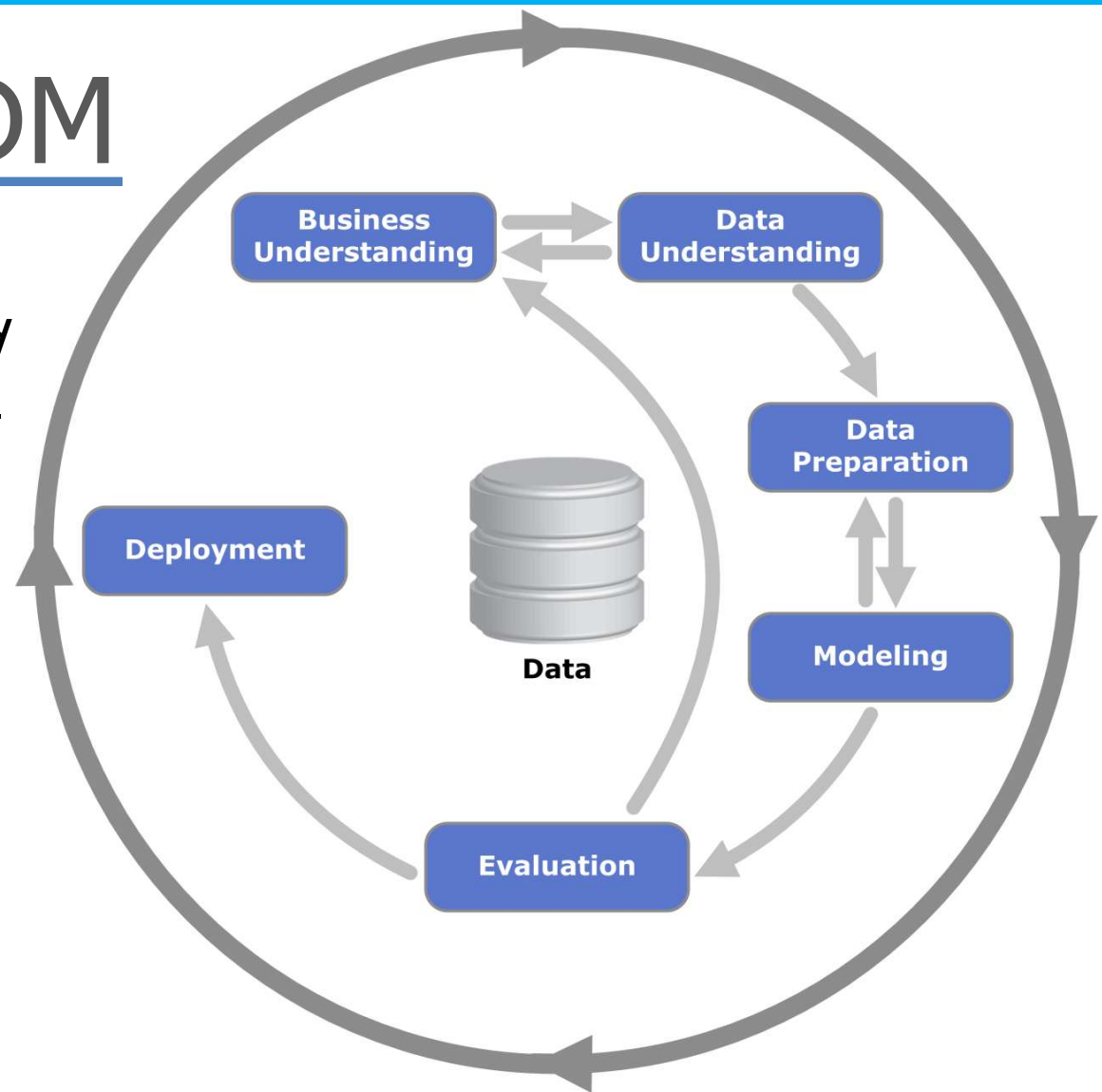


		2021			2022			2023		
		Fogão	Microondas	Geladeira	Fogão	Microondas	Geladeira	Fogão	Microondas	Geladeira
Sul	PR	3	5	1	2	4	2	3	5	5
	SC	1	4	2	4	9	1	7	6	7
	RS	7	2	6	2	7	4	9	1	6

Modelo CRISP-DM

Modelo CRISP-DM

❑ O modelo de referência *Cross Industry Standard Process for Data Mining* – CRISP-DM, está organizado nas seguintes fases:



Modelo CRISP-DM

- ❑ **O Entendimento do Negócio (*Business Understanding*):** foca o entendimento dos objetivos e requisitos do projeto, da perspectiva do domínio, a relevância do conhecimento prévio e os objetivos do usuário final;
- ❑ **O Entendimento dos Dados (*Data Understanding*):** realiza a coleta inicial de dados, descreve e explora os dados e verifica a qualidade dos dados;
- ❑ **O Pré-Processamento dos Dados (*Data Preparation*):** consiste na seleção de atributos, limpeza, construção, integração e formatação dos dados de entrada;
- ❑ **A Modelagem (*Modeling*):** seleciona modelos e parâmetros, com o uso direcionado para a obtenção de *insights*.
- ❑ **A Avaliação (*Evaluation*):** avalia, do ponto de vista de análise dos dados, a qualidade dos modelos obtidos, além de verificar se os objetivos do negócio foram atingidos conforme os critérios de sucesso adotados.
- ❑ **A Implantação (*Deployment*):** incorpora o modelo selecionado ao processo de tomada de decisão da organização.

Ciclo da Aprendizagem



Fim