



Data Science Academy

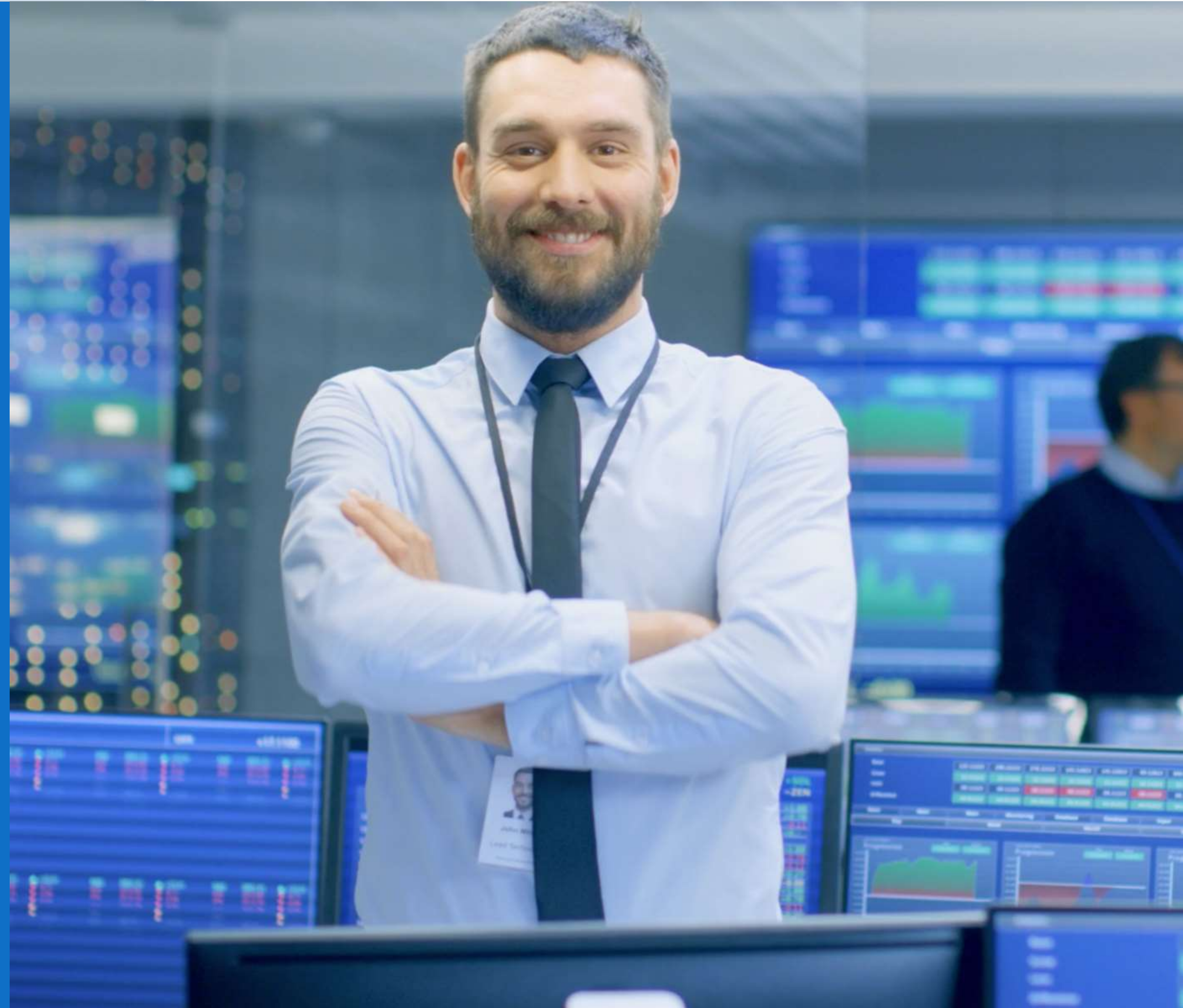


Design e Implementação de Data Warehouses



Data Science Academy

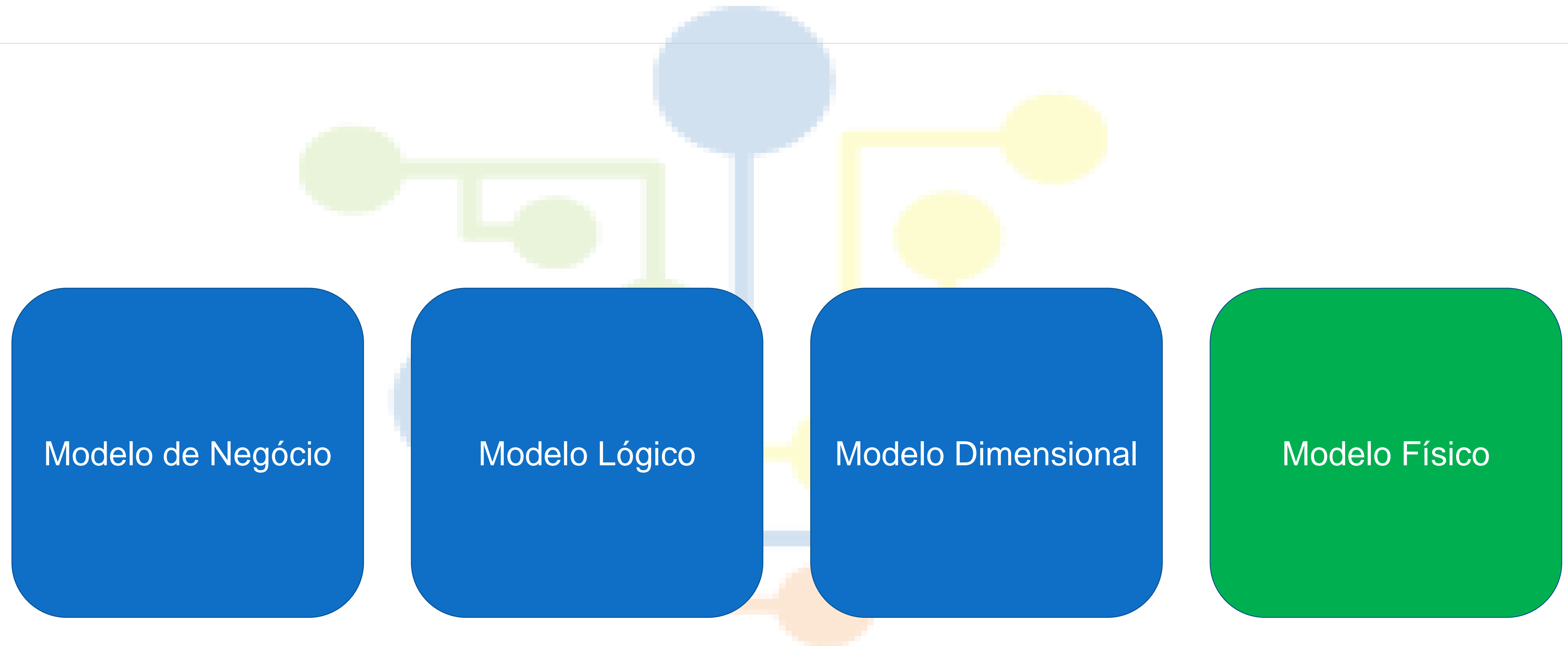
Banco de Dados Oracle Parte 2



The Oracle logo is displayed in red. In the background, there is a large, faint diagram consisting of a central blue vertical line with several horizontal and diagonal branches. These branches end in circles of different colors: light blue, light green, light yellow, and light orange. The word "ORACLE" is written in red, with a registered trademark symbol (®) to its upper right.

ORACLE®

- Parte 1
 - Ferramentas que serão usadas
 - Sistema operacional
 - Banco de Dados
 - Instalação
- Parte 2
 - Configuração e Fundamentos
 - Database Sizing
 - Particionamento e Indexação
 - Paralelismo
 - Segurança

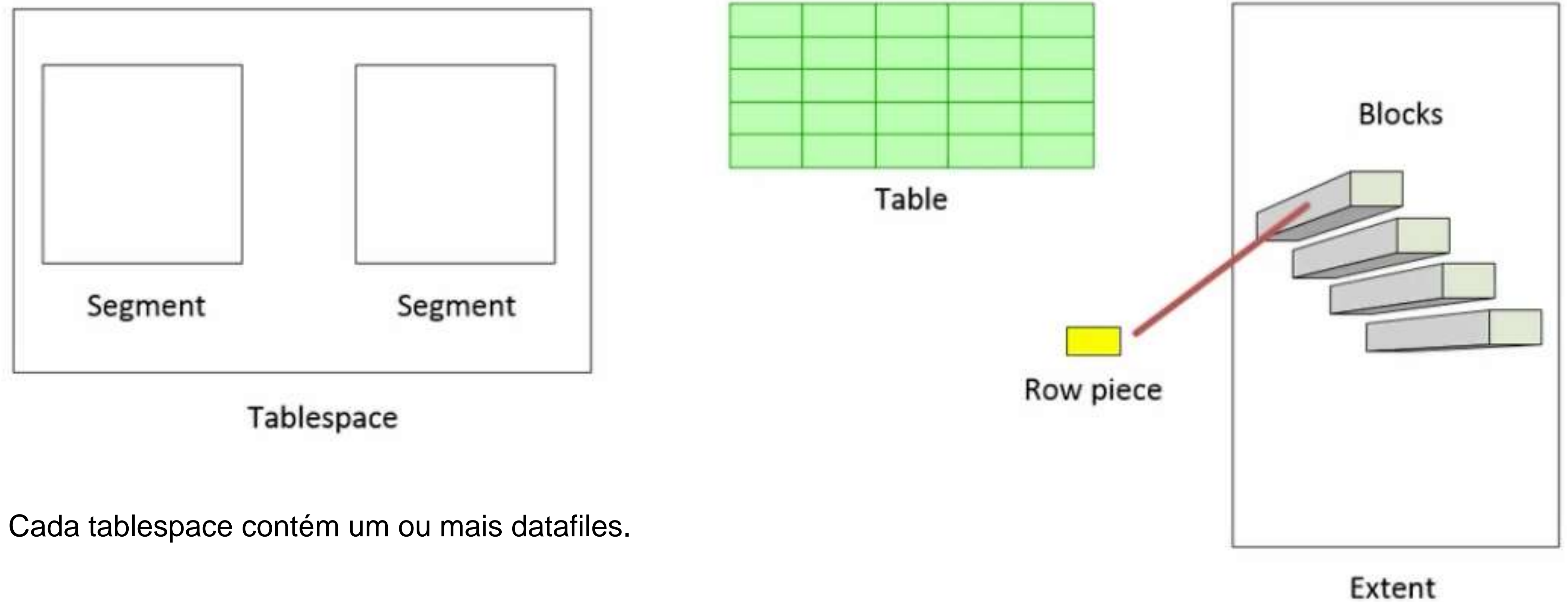




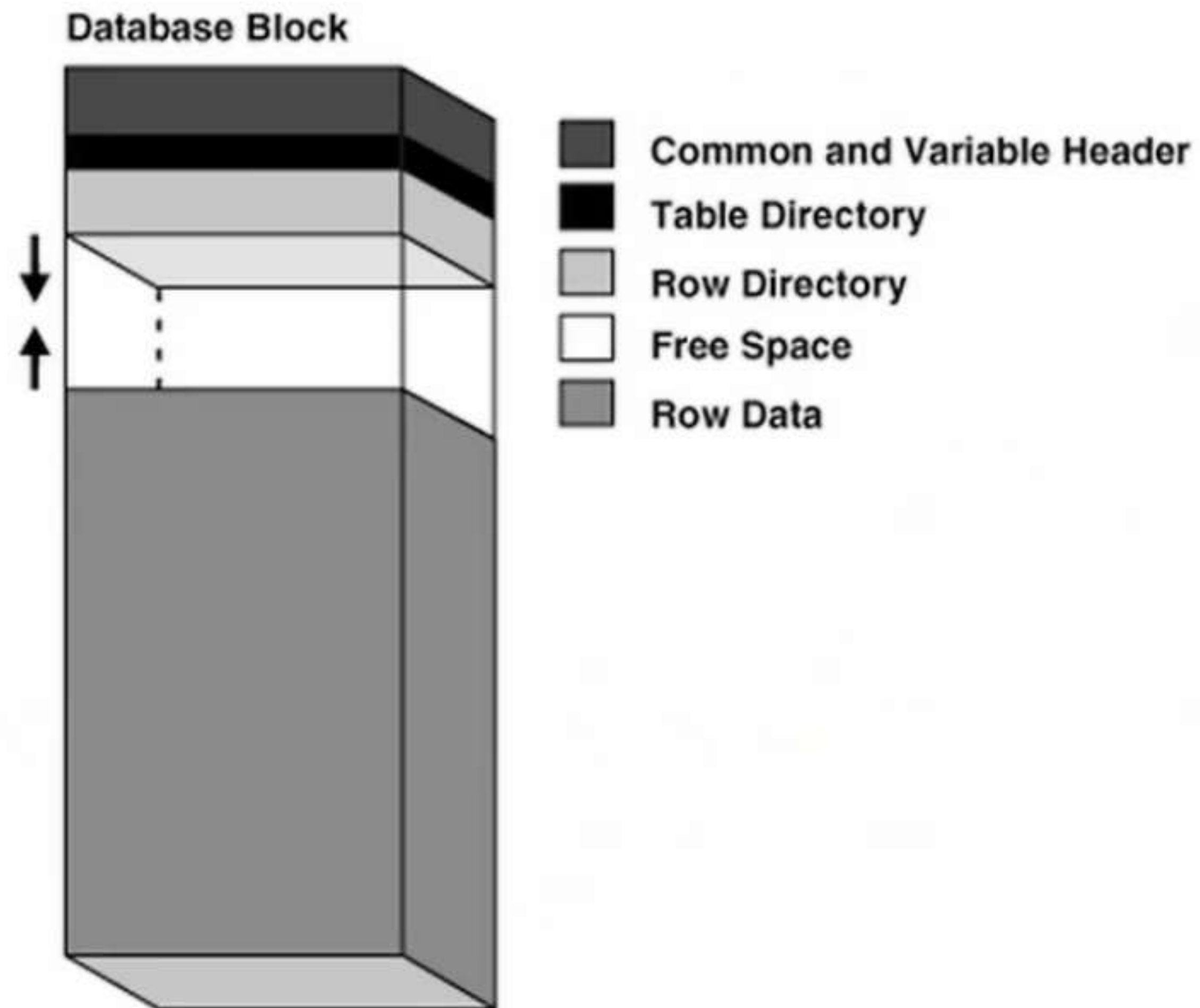
Data Science Academy

Como os Dados São Armazenados





Cada tablespace contém um ou mais datafiles.



Bloco é a estrutura mínima de armazenamento no Oracle.



Data Science Academy

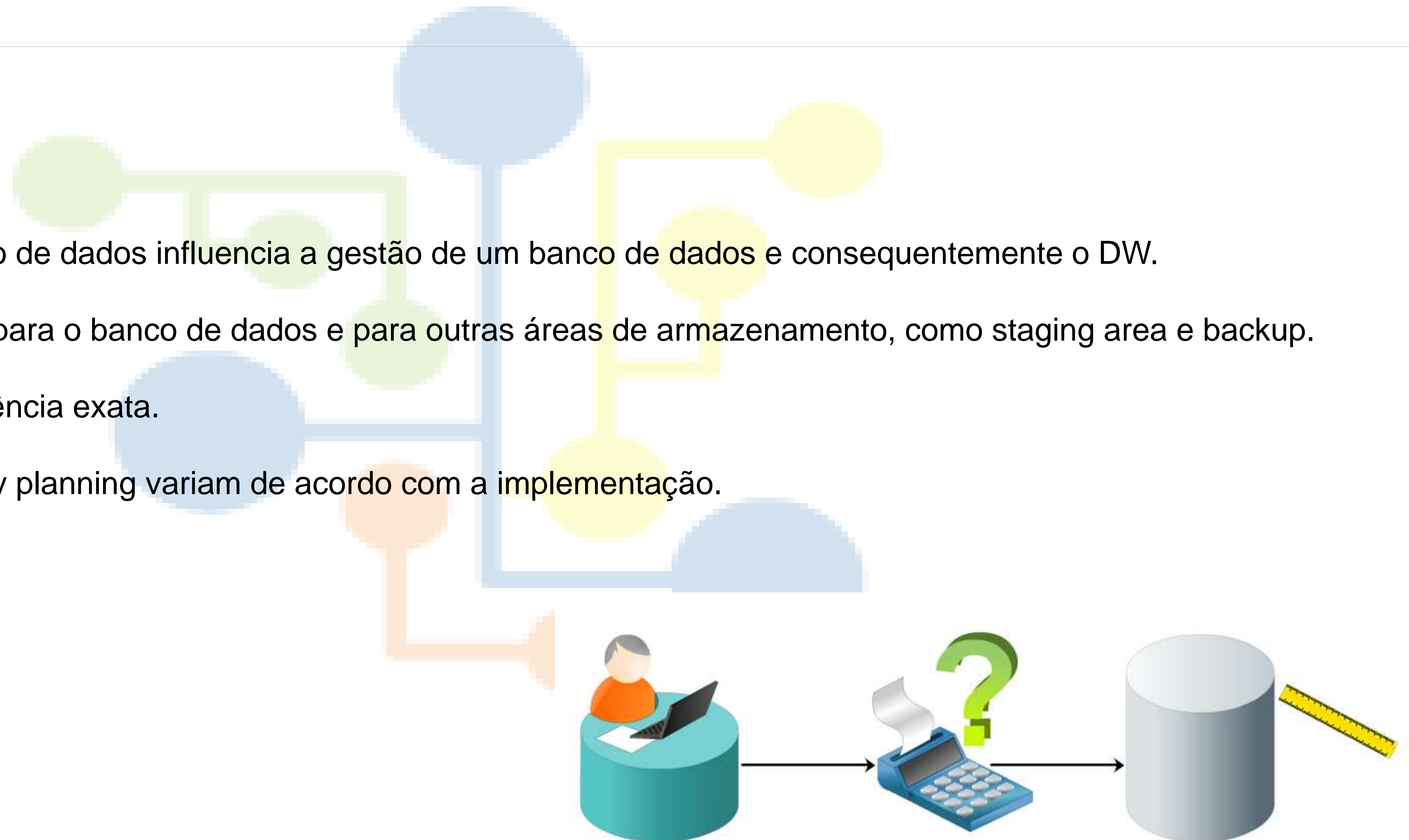
Capacity Planning e Database Sizing





Sizing

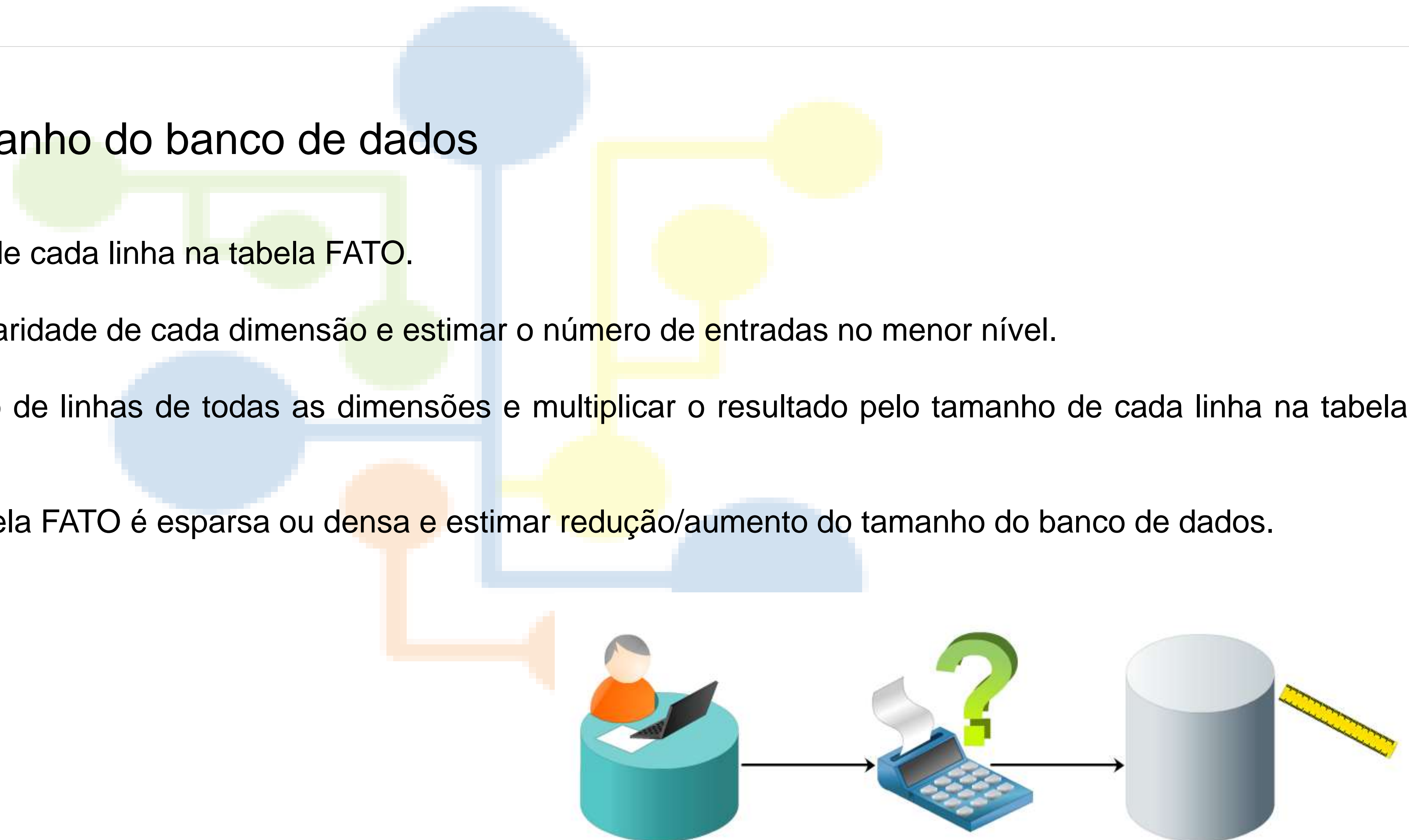
- O tamanho do banco de dados influencia a gestão de um banco de dados e consequentemente o DW.
- Sizing é importante para o banco de dados e para outras áreas de armazenamento, como staging area e backup.
- Sizing não é uma ciência exata.
- Técnicas de capacity planning variam de acordo com a implementação.





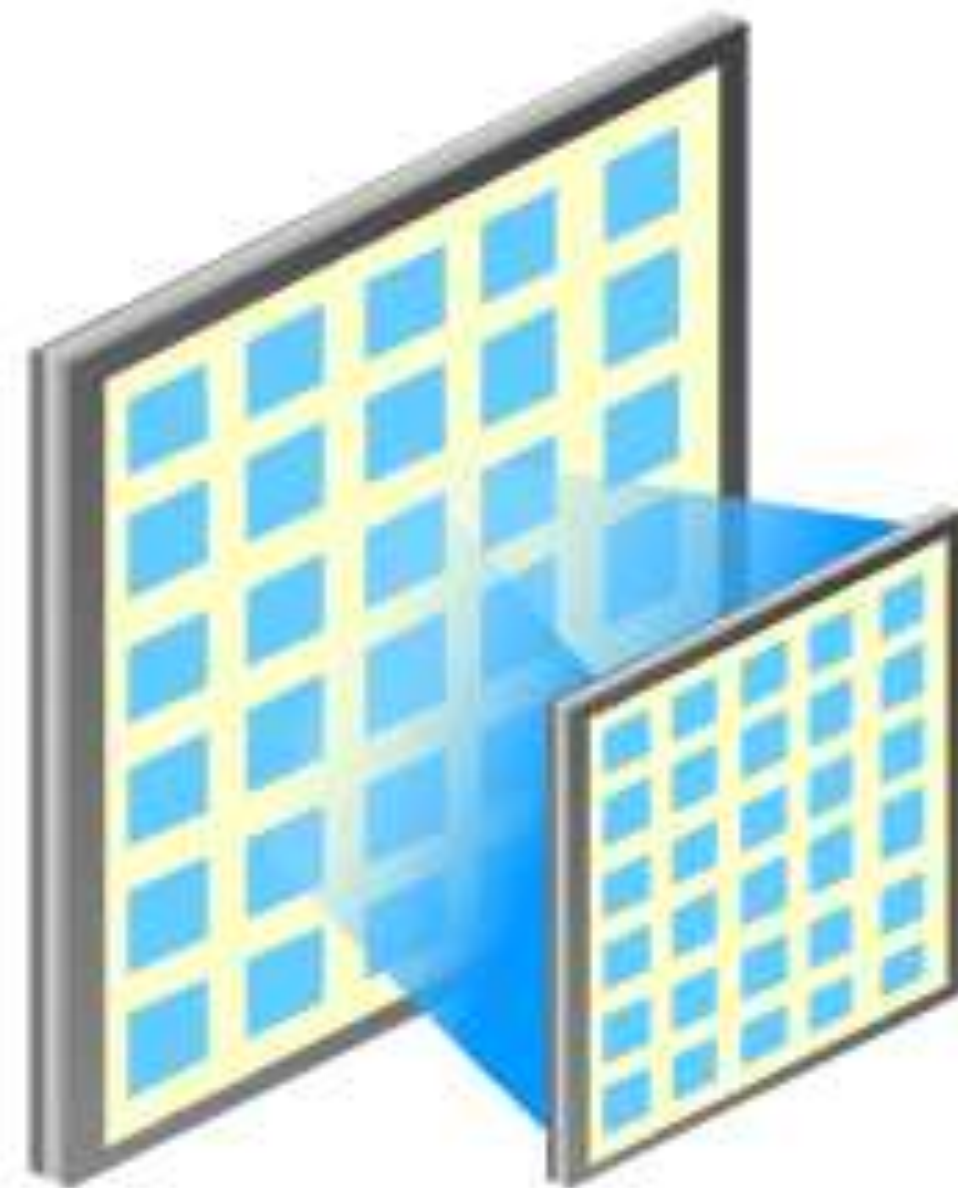
Estimando o tamanho do banco de dados

- Estimar o tamanho de cada linha na tabela FATO.
- Determinar a granularidade de cada dimensão e estimar o número de entradas no menor nível.
- Multiplicar o número de linhas de todas as dimensões e multiplicar o resultado pelo tamanho de cada linha na tabela FATO.
- Determinar se a tabela FATO é esparsa ou densa e estimar redução/aumento do tamanho do banco de dados.





Descrição	Estimativa
1. Estimar o tamanho de uma linha na tabela FATO.	52 bytes
2. Determinar a granularidade de cada dimensão e estimar o número de entradas no menor nível.	DIM_CANAIS - 5 canais DIM_CLIENTES - localizados em 23 países DIM_PRODUTOS - 72 itens DIM_HISTORICO - 48 meses
3. Multiplicar o número de linhas de todas as dimensões e multiplicar o resultado pelo tamanho de cada linha na tabela FATO.	$(5 \times 23 \times 72 \times 48) \times 52 = 20.666.880$ bytes
4. Determinar se a tabela FATO é esparsa ou densa e estimar redução/aumento do tamanho do banco de dados.	$20.666.880 \times 0.1 = 2.066.688$ $20.666.880 - 2.066.688 = 18.600.192$ bytes
Tamanho Estimado do Banco de Dados	18.6 MB (aproximadamente)



Carregar dados de exemplo, garantindo que os dados representem a população, considerando:

- Dados de diferentes períodos
- Operações do dia a dia
- Dados sazonais
- Índices

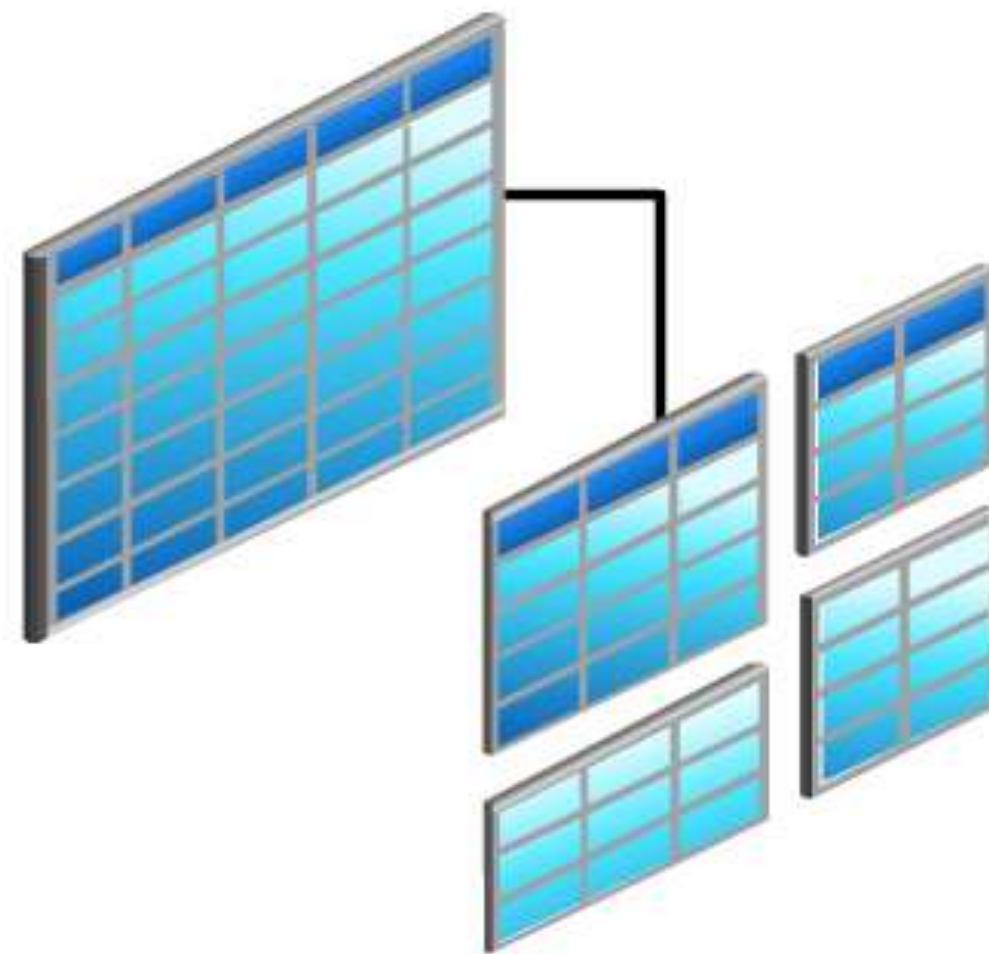




Particionamento

DWs pode crescer consideravelmente tornando-se VLDBs (Very Large Database).

- Grandes tabelas FATO e muitos dados históricos.
- Grandes tabelas são mais vulneráveis a falhas em disco.
- Índices de grandes tabelas levam muito tempo para serem construídos ou reconstruídos.
- Deletes parciais podem levar horas ou mesmo dias.
- Scan de grandes tabelas e índices leva muito tempo.
- Scan de porções menores da tabela aumenta a performance.

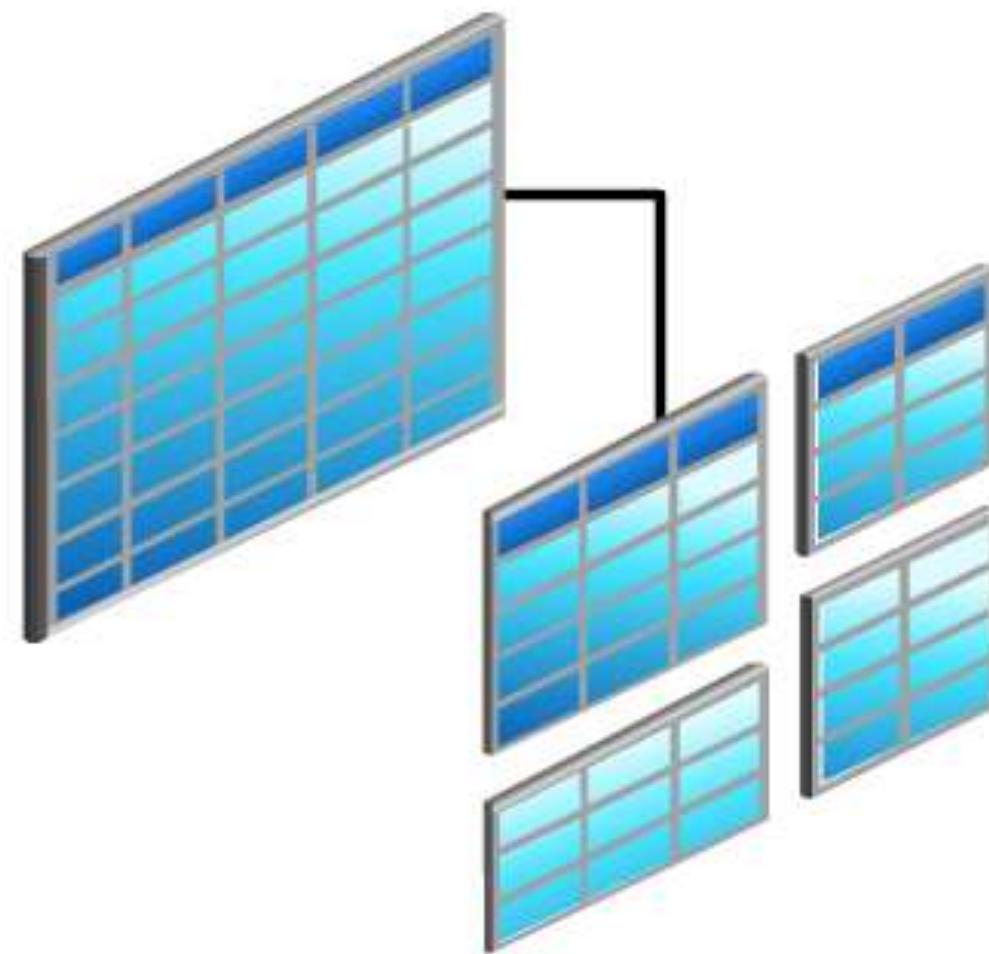




Particionamento

DWs pode crescer consideravelmente tornando-se VLDBs (Very Large Database).

- Dividir os dados em porções menores, que podem ser tratadas de forma independente.
- Operações de Join podem ser otimizadas por partição.
- Partições podem ser distribuídas por diferentes dispositivos físicos.
- O otimizador do banco de dados elimina partições que não precisam ser escaneadas.



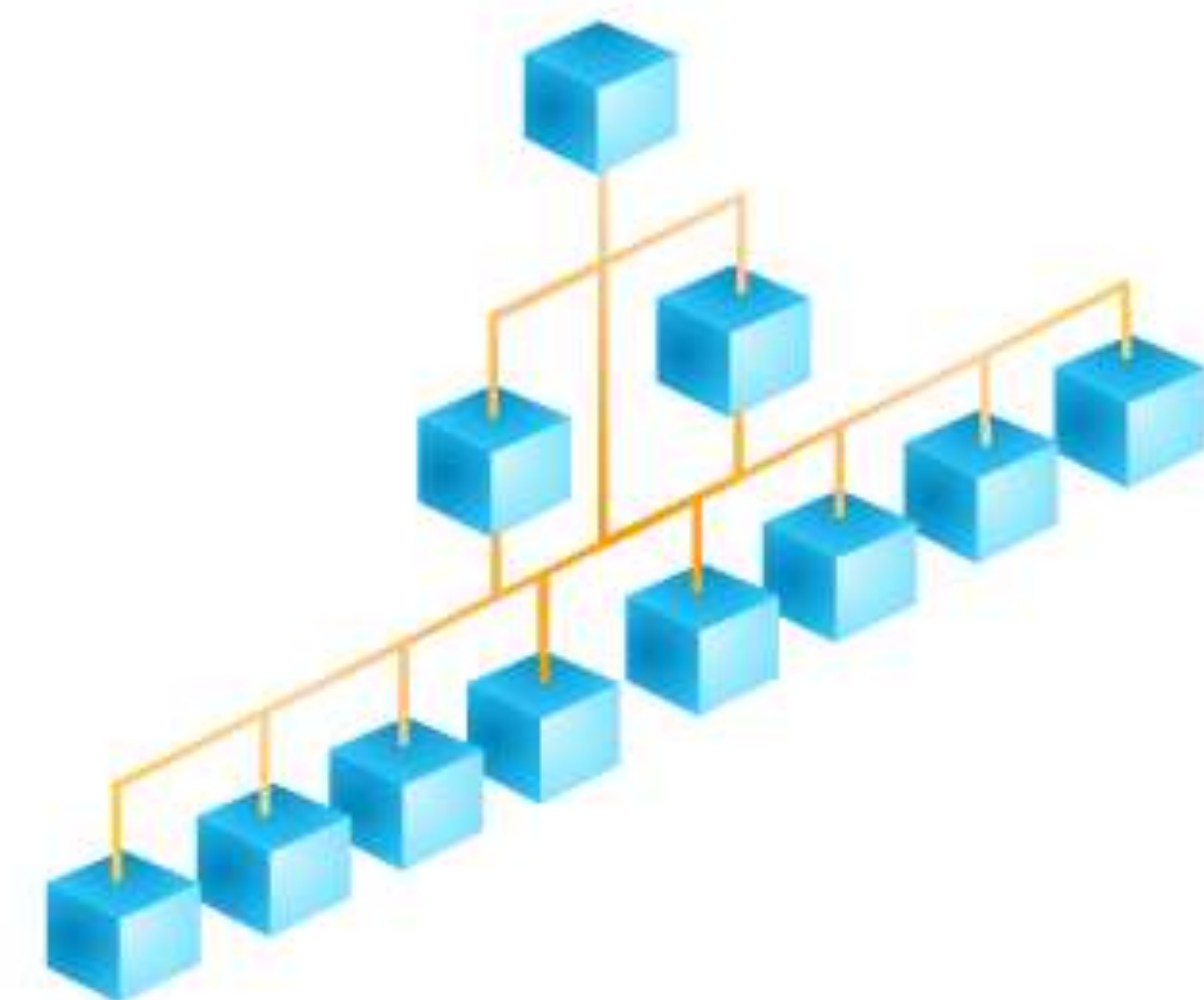


Estratégia de Particionamento	Descrição
Range	Dados são mapeados para partições com base no range de valores das “partition keys” para cada partição.
Hash	Dados são mapeados para partições com base em um algoritmo de hash nas “partition keys” para cada partição.
List	Dados são mapeados para partições usando uma lista de valores discretos para a coluna de Particionamento.
Interval	Dados são mapeados para partições com base em ranges que são automaticamente criados pelo banco de dados seguindo um intervalo específico.
System	A aplicação pode mapear os dados para partições.
Composite	Combinação de diferentes estratégias de particonamento.
Index	Utiliza o índice da tabela como estratégia de Particionamento.



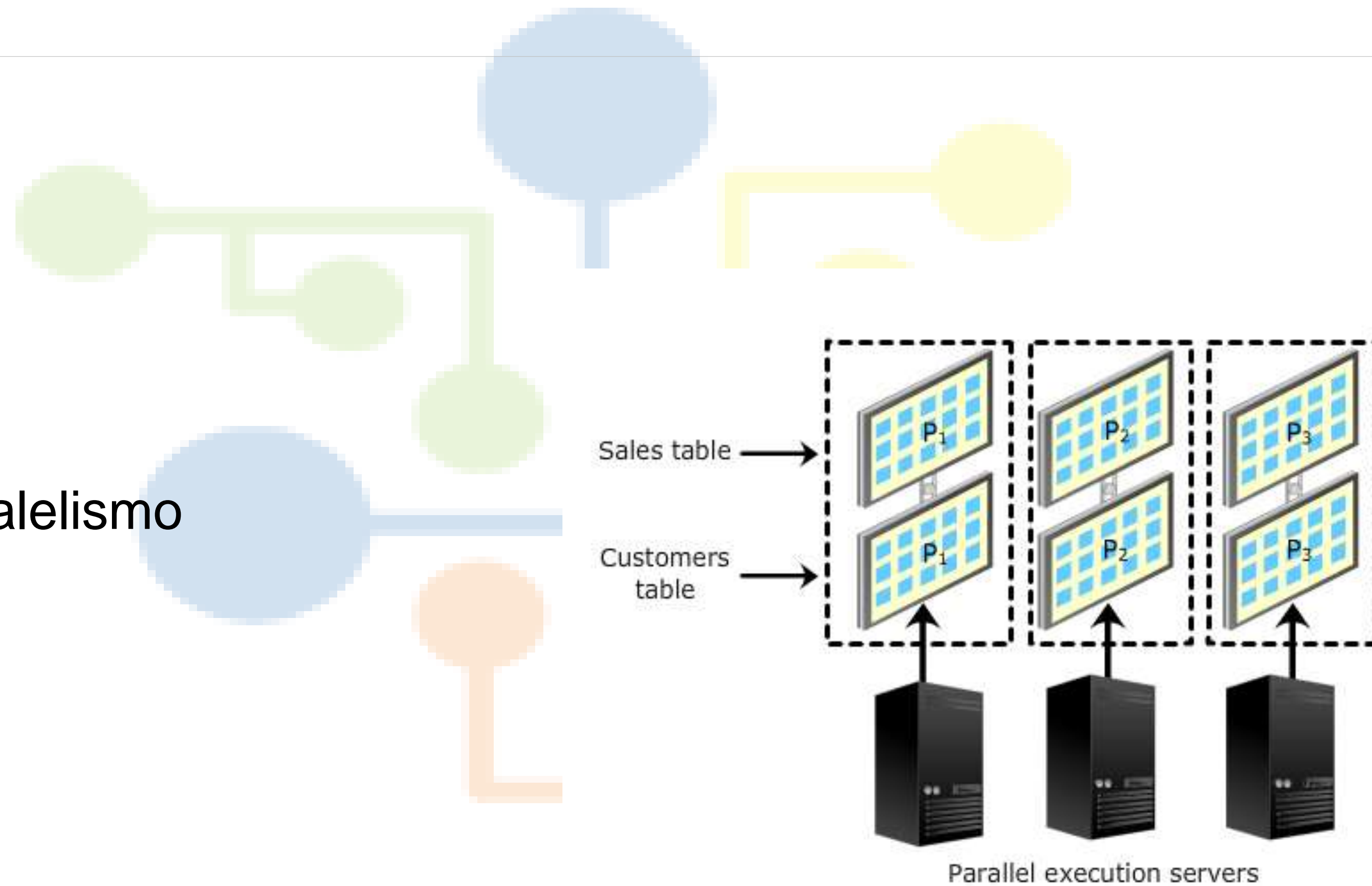
Indexação

- Melhora performance e escalabilidade
- Substitui full-table scan



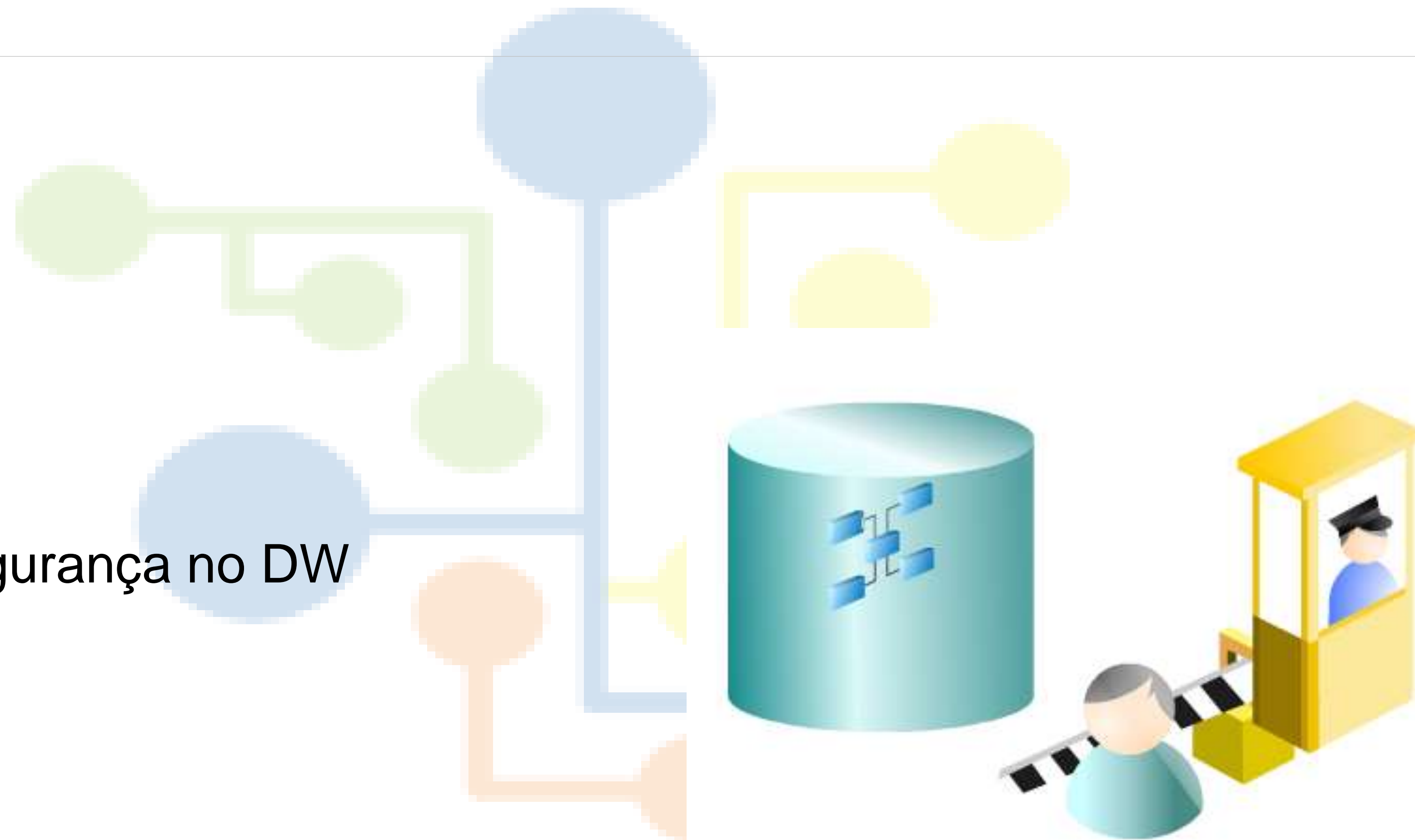


Paralelismo





Segurança no DW





Muito Obrigado!

É um prazer ter você aqui.
Tenha uma excelente jornada de aprendizagem.



Oportunidade



Disponibilidade



Conhecimento