

Arquitetura de Data Warehouse e Data Marts

Aula 9: Aplicações BI - O que é OLAP? Ferramentas e Aplicação

Apresentação

Nesta aula serão apresentados o conceito OLAP - On-line Analytical Processing, as características das ferramentas de OLAP, as operações de análise de dados: Slice, Drill-down, Roll-up, Pivot, Drill-across, Dice, Drill-through; a introdução ao ambiente de construção de consultas analíticas do Microsoft Power BI.

Objetivos

- Descrever o conceito On-line Analytical Processing (OLAP);
- Reconhecer as operações de análise de dados e características de ferramentas OLAP;
- Examinar funcionalidades da ferramenta OLAP Power BI.

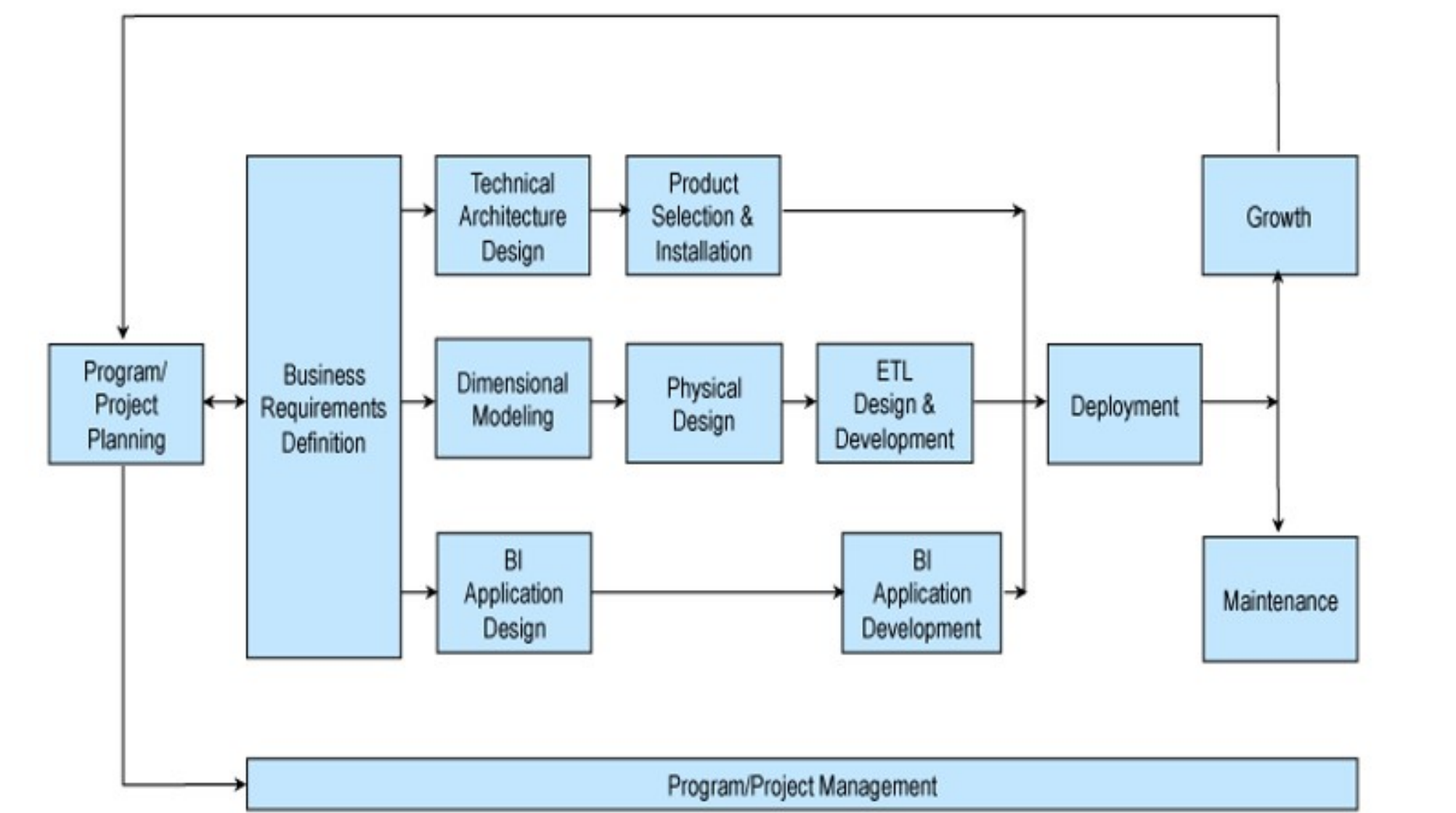
Aplicações de BI

A última trilha de atividades definidas por Ralph Kimball no ciclo de vida de um projeto de Data Warehouse é composta pela atividade de definição, **Especificação de Aplicação de BI** (BI Application Design), e construção da camada de visualização dos dados, **Desenvolvimento das Aplicações de BI** (BI Application Development).

Comentário

Na fase de levantamento de requisitos, as consultas analíticas principais foram mapeadas de acordo com as necessidades dos usuários. A fase de Especificação de Aplicação de BI deve utilizar esse material e aperfeiçoar a documentação acrescentando informações ou modificando-a de acordo com o amadurecimento do projeto.

As consultas predefinidas devem oferecer as principais análises realizadas ou desejadas pela organização, no entanto novas consultas poderão ser desenvolvidas posteriormente pelos usuários. Quando o uso das consultas definidas pelos usuários torna-se frequente por outros usuários, elas podem ser disponibilizadas para que os demais possam utilizá-las e assim, ela passa a ser uma consulta padrão como as criadas pelo analista do projeto.




 Ciclo de Vida de um Projeto de Data Warehouse por Kimball. Fonte: (Kimball 2013).

Saiba mais

O desenvolvimento das análises depende da base de dados criada com uma amostra de dados carregada e da ferramenta para construção das análises, definida na fase de seleção de ferramentas para o projeto.

Devido ao dinamismo das consultas analíticas do DW/DM, as ferramentas devem apresentar os recursos que atendam às necessidades desse cenário, o que as diferenciam de ferramentas voltadas para ambientes operacionais (OLTP - Online Transaction Processing). As ferramentas de construção de consultas em sistemas operacionais OLTP trabalham com alta velocidade na manipulação dos dados, no entanto são ineficientes em consultas nos sistemas analíticos.

OLAP

 Clique no botão acima.

Já as ferramentas OLAP (On-line Analytical Processing) são preparadas para construir e suportar as análises dimensionais presentes no ambiente analítico, otimizando a leitura de dados e consultas gerenciais realizadas em grandes volumes de dados históricos.

Como visto anteriormente, os dados organizados em um modelo de dados dimensional podem ser representados como um cubo de dados de processamento analítico on-line (OLAP), em que as dimensões estão localizadas nos eixos do cubo e as medidas, contidas na tabela Fato, estão localizadas na interseção das dimensões.

O processamento OLAP auxilia o processamento de análises multidimensionais e as operações de cálculos complexos. Essas operações têm como objetivo auxiliar o usuário na consulta específica de dados no cubo e extrair apenas as informações necessárias para a tomada de decisão.

O termo OLAP foi usado pela primeira vez por Edgar Frank Codd em 1993, que também descreveu 12 características que uma ferramenta deve apresentar para que seja considerada uma ferramenta OLAP:

1. Visão Conceitual Multidimensional – Deve oferecer visão multidimensional e permitir o cruzamento de dados.
2. Transparência – Deve ser adequada ao trabalho habitual dos usuários sem afetar as atividades realizadas e oferecer consultas conforme a solicitação.
3. Acessibilidade – Deve ser capaz de acessar as bases de dados origem, mapear os dados e acessar apenas dados necessários às análises.
4. Performance consistente de informações – Deve disponibilizar consultas sem complexidade ao usuário e questões referentes a quantidade de dimensões ou o aumento da base de dados não deve afetar a performance dos relatórios acessados pelo usuário.
5. Arquitetura cliente/servidor – O servidor deve ser capaz de atender às análises solicitadas pelos usuários, independentemente da estação que está sendo utilizada e do sistema operacional utilizado.
6. Dimensionalidade genérica – Deve responder às análises com as diversas dimensões contidas no modelo, podendo aplicar as mesmas funções a quaisquer dimensões disponíveis.
7. Tratamento dinâmico de matrizes esparsas – Matrizes esparsas são matrizes que possuem elementos nulos. Então, a maioria das posições são preenchidas por zeros. Assim, para economizar espaço em memória, apenas os termos diferentes de zero devem ser armazenados e manuseados visando melhorar a eficiência computacional.
8. Suporte a multiusuários – Deve oferecer acesso simultâneo com integridade dos dados e segurança.

- 9. Operações irrestritas em dimensões cruzadas – Deve permitir o cálculo e a manipulação de dados a partir de qualquer número de dimensões de dados e não deve restringir ou inibir qualquer relação entre eles.
- 10. Manipulação de dados intuitiva – A manipulação dos dados deve ser realizada pelo mecanismo da ferramenta sem a interferência do usuário, que deve navegar pelos dados de forma simples.
- 11. Consultas flexíveis – Os dados devem estar visivelmente organizados de forma que se aproximem das visualizações da organização permitindo ao usuário realizar qualquer tipo de consulta.
- 12. Dimensões e níveis de agregação ilimitados – É recomendado que a ferramenta OLAP suporte no mínimo 15 dimensões em um mesmo modelo e que estas permitam que vários usuários realizem diversos níveis de agregações aos dados.

Operações de análise de dados

As ferramentas OLAP possuem alta capacidade de realizar cálculos e manipulação entre linhas, trabalham com hierarquias, entre outras funções. Dispõem de navegabilidade permitindo realizar operações de maior detalhamento dos dados (Drill down) e menor detalhamento dos dados (Drill Up/Roll up).

Além das operações de maior e menor detalhamento dos dados, há outras como:

- Drill Across – Permite comparar medidas distintas em fatos diferentes relacionadas por, pelo menos, uma dimensão em comum. É como navegar de uma estrutura de dados para outra unindo-as por dimensões compartilhadas. Isso possibilita comparar a quantidade de produtos vendida e a quantidade de produtos contida no estoque pelas mesmas visões, como data e produto.

Vendas de Produto por Mês								
Mês	LIMPEZA		GRÃO			NÃO INFORMADO	HIGIENE	
	DETERGENTE	SABAO EM PO	ARROZ	FEIJAO	MILHO DE PIPOCA	SUCO DE UVA	PASTA DE DENTE	SABONETE
May-2020	3	0	2	10	2	1	0	0
Jun-2020	0	1	2	0	0	0	2	47
Total	3	1	4	10	2	1	2	47

Estoque de Produto por Mês								
Mês	LIMPEZA		GRÃO			NÃO INFORMADO	HIGIENE	
	DETERGENTE	SABAO EM PO	ARROZ	FEIJAO	MILHO DE PIPOCA	SUCO DE UVA	PASTA DE DENTE	SABONETE
May-2020	30	0	60	60	60	50	0	0
Jun-2020	0	20	40	0	0	0	30	80
Total	30	20	100	60	60	50	30	80

- Slice - Permite fatiar o cubo de dados selecionando-os de uma dimensão contida no conjunto de dados.

Vendas de Produto por Mês								
	LIMPEZA		GRÃO			NÃO INFORMADO	HIGIENE	
Mês	DETERGENTE	SABAO EM PO	ARROZ	FEIJAO	MILHO DE PIPOCA	SUCO DE UVA	PASTA DE DENTE	SABONETE
May-2020	3	0	2	10	2	1	0	0
Jun-2020	0	1	2	0	0	0	2	47
Total	3	1	4	10	2	1	2	47

Vendas de Produto por Mês

GRÃO				
Mês	ARROZ	FEIJAO	MILHO DE PIPOCA	Total
May-2020	2	10	2	14
Jun-2020	2	0	0	2
Total	4	10	2	16

- Dice – Permite cortar o cubo de dados de acordo com os atributos de duas ou mais dimensões contidas no conjunto de dados.

Vendas de Produto por Mês								
Mês	LIMPEZA		GRÃO			NÃO INFORMADO	HIGIENE	
	DETERGENTE	SABAO EM PO	ARROZ	FEIJAO	MILHO DE PIPOCA	SUCO DE UVA	PASTA DE DENTE	SABONETE
May-2020	3	0	2	10	2	1	0	0
Jun-2020	0	1	2	0	0	0	2	47
Total	3	1	4	10	2	1	2	47

Vendas de Produto por Mês

GRÃO				
Mês	ARROZ	FEIJAO	MILHO DE PIPOCA	Total
Jun-2020	2	0	0	2
Total	2	0	0	2

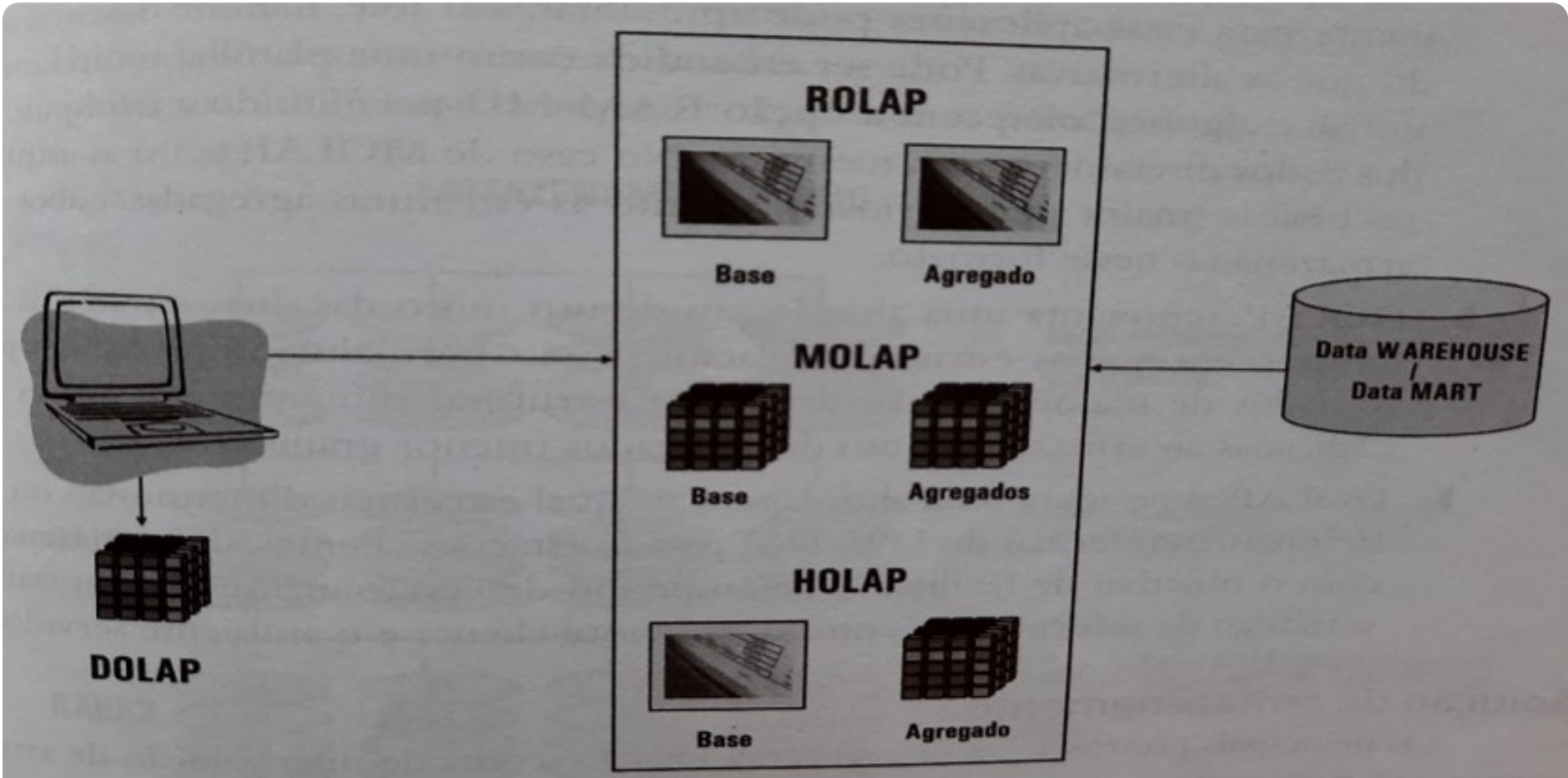
- Pivoting – Permite visualizar os dados com a mudança do eixo das dimensões.
- Drill through – Permite visualizar um maior nível de detalhe do dado que está sendo armazenado na tabela Fato, onde, por exemplo no cenário Supermercado, o menor nível contido é em relação ao produto vendido e quando foi vendido. O Drill through vai além dessa análise, utilizando o número do pedido armazenado na tabela Fato vendas e acessando o pedido no sistema transacional em que ele está armazenado.

Ele é semelhante ao Drill down, contudo acessando um ambiente externo ao esquema dimensional do DW/DM.

Atenção! Aqui existe uma videoaula, acesso pelo conteúdo online

As Ferramentas OLAP quanto ao tipo de armazenamento dos dados

No mercado há diversas ferramentas OLAP que possuem o objetivo em comum de gerar consultas analíticas baseadas em grandes volumes de dados. Elas possuem arquitetura baseada na estrutura cliente/servidor voltada ao suporte de multiusuários. No entanto, as ferramentas OLAP podem ser classificadas em quatro grupos: ROLAP, MOLAP, HOLAP e DOLAP.. Conheça-as a seguir.



[ROLAP \(Relational Online Analytical Processing\)](#)



As ferramentas do tipo ROLAP trabalham com os dados em estruturas relacionais, recuperam e processam-nos de acordo com a demanda do usuário. Possuem as vantagens dos SGBDs Relacionais, no entanto há a preocupação relacionada ao desempenho das consultas, pois o acesso às tabelas normalizadas compromete a recuperação dos dados. As tabelas Fatos e as tabelas Fatos agregadas são armazenadas nessa arquitetura.

Por meio de consultas SQL, a ferramenta realiza cálculos complexos em grandes volumes de dados. No entanto, quanto maior o conjunto de dados mais lenta é a resposta das consultas.

[MOLAP \(Multidimensional Online Analytical Processing\)](#)



As ferramentas do tipo MOLAP trabalham com características de armazenamento especial para o tratamento de dados dimensionais, em cubos OLAP, e contam com os dados resumidos e pré-calculados respondendo às consultas com maior rapidez. As tabelas Fatos e as tabelas Fatos agregadas são armazenadas no formato de cubos dimensionais.

Permitem a manipulação de dados em memória, mas apesar do benefício de maior rapidez nas respostas das consultas, essa arquitetura é limitada quanto a armazenamento.

[HOLAP \(Hybrid Online Analytical Processing\)](#)



As ferramentas do tipo HOLAP unem benefícios dos tipos anteriores, armazenando os dados — parte como o ROLAP e parte como o MOLAP —, onde a estrutura relacional geralmente é utilizada para as tabelas Fatos (nível maior de granularidade) e a estrutura dimensional é utilizada para o armazenamento das tabelas Fato agregadas (menor nível de granularidade).

[DOLAP \(Desktop Online Analytical Processing\)](#)



Nessa arquitetura, a estrutura dos dados é transferida para a estação desktop que solicitou a análise com o objetivo de minimizar o tráfego de informações na rede da organização.

Ferramentas OLAP

Anualmente, o Gartner Group analisa as empresas e ferramentas disponíveis no mercado de tecnologia e publica um relatório com o conhecido Quadrante Mágico, que classifica as ferramentas de BI em relação a diversos aspectos, como inovação tecnológica e abrangência de necessidades dos usuários e o cumprimento do que se propõe a solução.

Algumas das ferramentas contidas na análise Gartner são: IBM, Microsoft, MicroStrategy, Oracle, Qlik, Salesforce, SAP, SAS, Tableau

A Microsoft apresenta-se como a atual líder dentre as ferramentas OLAP e é

representada pela ferramenta Microsoft Power BI, uma ferramenta que possibilita acessar dados de fontes locais (planilhas Excel, SGBDs, entre outras) ou da nuvem Microsoft Azure.

Dica

Para a execução dos exemplos e exercícios, vamos utilizar a ferramenta Power BI Desktop.

Você pode baixar a ferramenta gratuitamente no site da Microsoft.

Atenção! Aqui existe uma videoaula, acesso pelo conteúdo online

Microsoft Power BI

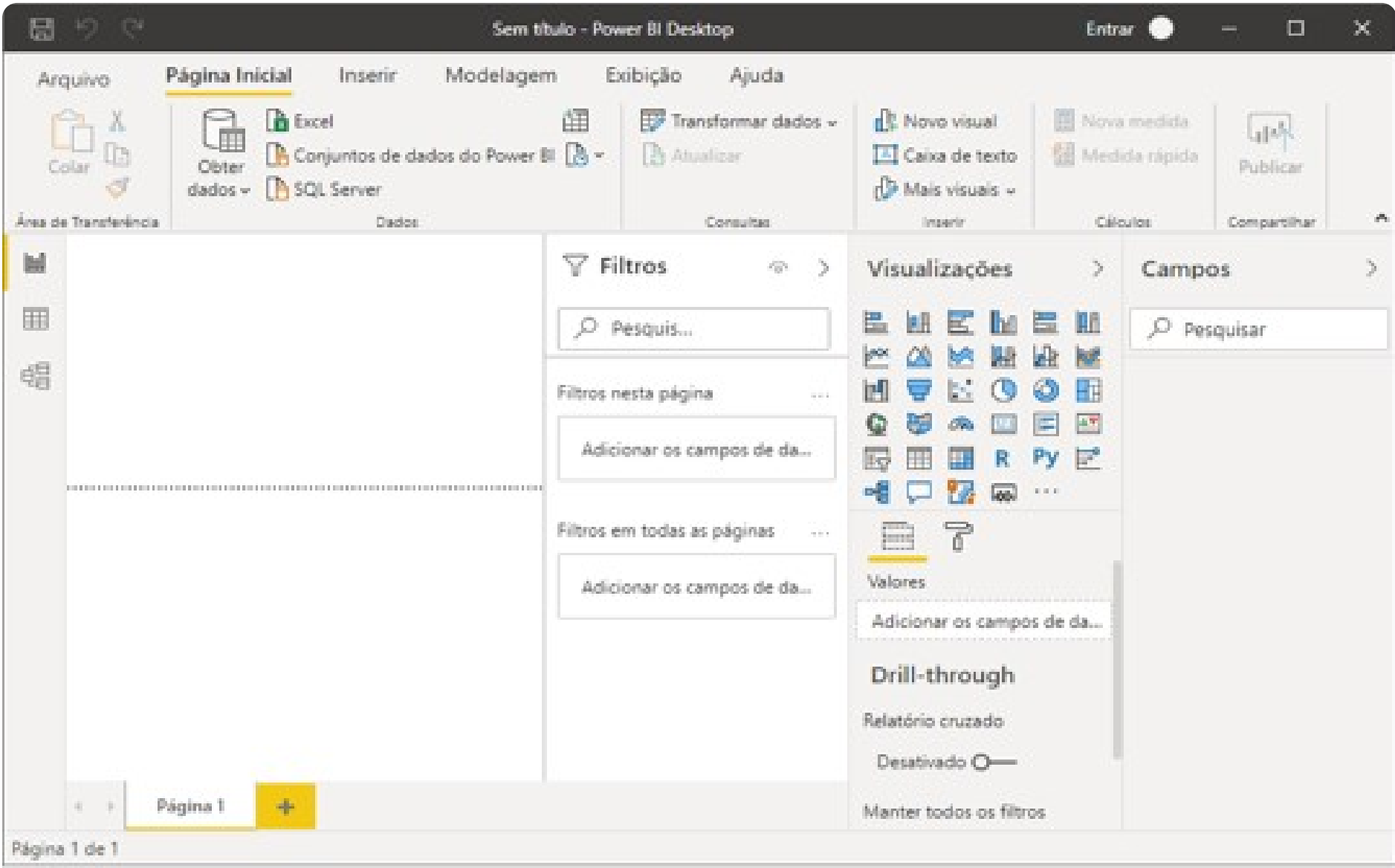
O Power BI é um conjunto de ferramentas voltadas a análises estratégicas, permitindo criar visualizações interativas e possui uma interface simples que facilita a criação de aplicações analíticas, inclusive pelos usuários finais.

Saiba mais

O Power BI possui versão gratuita que permite criar relatórios e dashboards, utilizar os recursos de modelagem e tratamento dos dados utilizando o editor Power Query e publicar os relatórios. Nessa versão, não há controle de usuários, então todo usuário que tiver o link poderá acessar as análises. Isso, em uma organização, geralmente não é praticado, pois os dados gerenciais não são abertos ao público ou a todos os usuários da organização.

Para essa necessidade existe a versão Pro (paga), que possui controle de acesso de usuários. No entanto, somente usuários que possuem um perfil Pro poderão acessar os relatórios.

A interface do Power BI é bastante intuitiva e facilita a navegação entre as funcionalidades disponíveis.



Na lateral da janela à esquerda estão posicionados três botões:

Relatório

Apresenta a área de trabalho onde são desenvolvidos os relatórios.

Dados

Onde é possível visualizar os dados carregados no ambiente.

Modelo

Permite relacionar os dados carregados.

Nesse painel é possível ver as tabelas carregadas e relacioná-las obedecendo às chaves criadas no modelo de dados dimensional.

- O painel de Filtros exibe as tabelas e os campos contidos no relatório, permitindo aplicar filtros nas consultas contidas no relatório e visualizar os filtros que foram aplicados.
- O painel Visualizações lista os elementos que apresentam os dados no relatório, como cartões, gráficos e tabelas. Além das Visualizações disponíveis, há outras que podem ser adicionadas ao painel conforme a necessidade das consultas que você está desenvolvendo.
- O painel Campos exibe as tabelas contidas no modelo de dados do Power BI, os campos e as medidas e atributos criados a partir dos campos existentes.

Obter Dados

A primeira ação a ser feita é conectar o Power BI à base de dados que iremos consultar para construir as análises. Para isso, basta clicar no botão Obter Dados. Nesse passo, abre-se um menu com as opções de fontes de dados. Caso a fonte de dados que você irá utilizar não esteja disponível, clique botão Mais e a janela ilustrada pela imagem será aberta.

Para os nossos exemplos utilizaremos a base de dados contida no PostgreSQL.

Obter Dados

Pesquisar

Tudo

Arquivo

Banco de Dados

Power Platform

Azure

Serviços Online

Outro

Tudo

Excel

Texto/CSV

XML

JSON

Pasta

PDF

Pasta do SharePoint

Banco de dados SQL Server

Banco de dados do Access

Banco de dados SQL Server Analysis Services

Oracle Database

Banco de dados IBM Db2

Banco de dados do IBM Informix (Beta)

IBM Netezza

Banco de dados MySQL

Banco de dados PostgreSQL

Conectores Certificados

Conectar

Cancelar

Após escolher a fonte de dados, clique no botão Conectar. Serão solicitados o servidor e o nome da base de dados. Preencha as informações e clique no botão **Ok**.

Banco de dados PostgreSQL

X

Servidor

localhost

Banco de Dados

DWSuper

Modo de Conectividade de Dados ⓘ

☒ Importar

☐ DirectQuery

> Opções avançadas

OK

Cancelar

Após a conexão com a base de dados, as tabelas Dimensões e tabelas Fatos serão exibidas para que sejam selecionadas e carregadas no ambiente do Power BI.

Atenção

O botão **Carregar** é responsável por carregar a estrutura de tabelas e dados diretamente na ferramenta, e o botão **Transformar** permite aplicar regras, limpezas, entre outros, antes do carregamento. Essa função é muito útil se você está lendo dados brutos e necessita transformá-los.

A imagem ilustra as tabelas contidas no modelo de dados dimensional do cenário Supermercado. Nessa etapa, devem ser selecionadas as tabelas que serão usadas na construção das análises.

No nosso exemplo, devem ser selecionadas as tabelas iniciadas com o prefixo dim_ (dimensão), ft_ (fato) e agr_ (agregadas).

Navegador

Opções de Exibição ▾

localhost: DWSuper [13]

✓

public.agr_lucro_produto

✓

public.agr_vendas_produto

✓

public.dim_categoria

✓

public.dim_cliente

✓

public.dim_data

✓

public.dim_produto

✓

public.ft_estoque

✓

public.ft_vendas

public.tmp_categoria

public.tmp_cliente

public.tmp_ft_estoque

public.tmp_ft_vendas

public.tmp_produto

public.ft_vendas

sk_produto_vendas	sk_cliente_vendas	sk_data_vendas
10	1	20200620
10	6	20200619
10	7	20200620
10	8	20200620
11	5	20200520
14	1	20200625
14	4	20200518
15	5	20200520
16	5	20200520
17	1	20200625
18	5	20200520
19	1	20200625

<

>

Selecionar Tabelas Relacionadas


Carregar



Transformar Dados



Cancelar



Após clicar no botão Carregar as tabelas são exibidas no Painel Campos.



Campos







 public agr_lucro_produto



 public agr_vendas_produto



 public dim_categoria

 public dim_cliente

 public dim_data

 public dim_produto

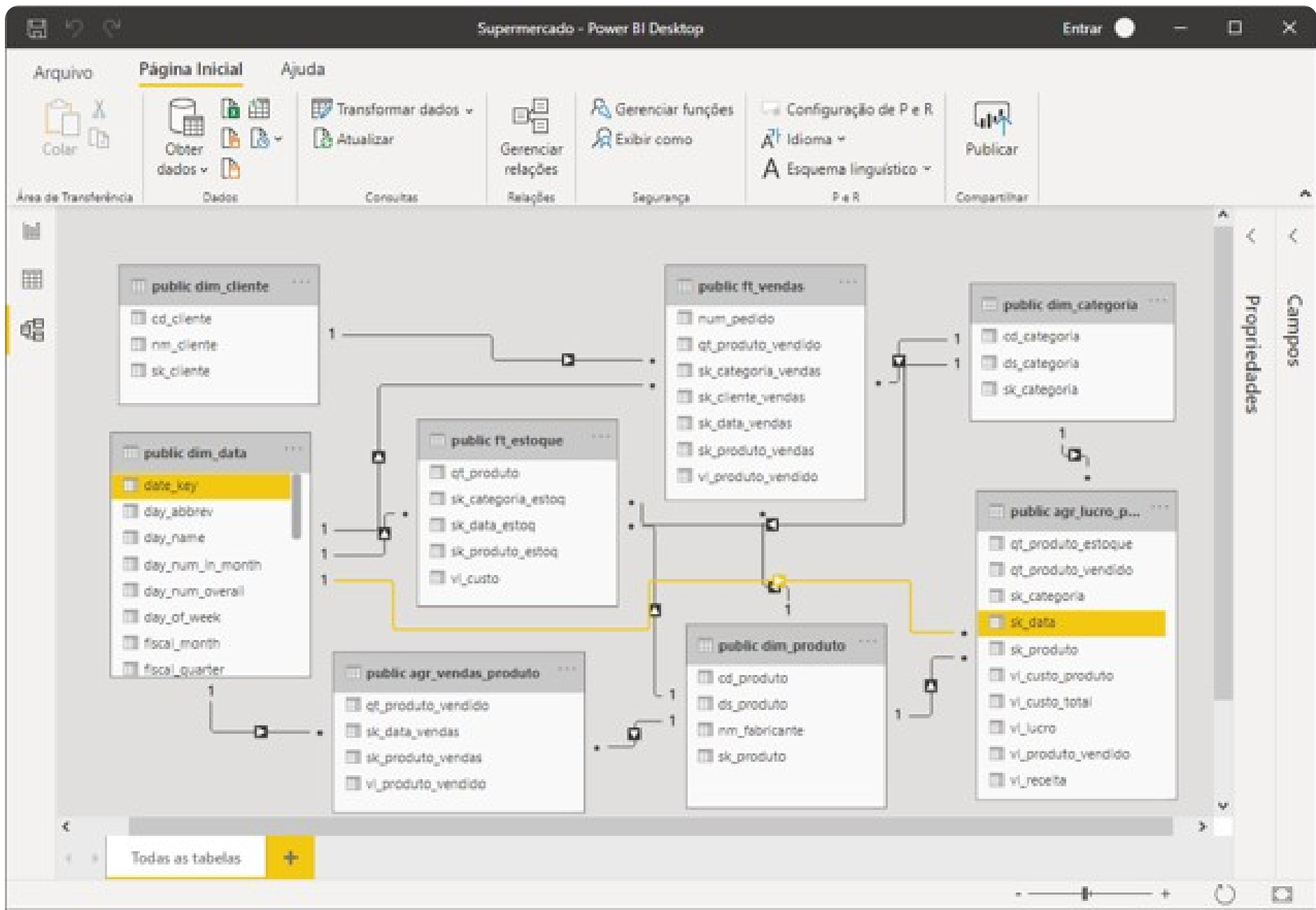
 public ft_estoque

 public ft_vendas

Após a carga das tabelas do modelo de dados dimensional, clique no botão **Modelo** (terceiro botão na lateral esquerda da janela) para verificar e relacionar corretamente as chaves SKs das dimensões com as tabelas Fato.

Nesse painel são exibidas as tabelas, bem como os relacionamentos entre elas. Ao carregar as tabelas, o Power BI verifica os nomes das colunas que parecem ser correspondentes e cria as relações automaticamente, atribuindo a cardinalidade, a direção e as propriedades da relação.

Verifique se há alguma relação inadequada. Para isso, clique com o botão direito sobre a relação (linha amarela em destaque) e escolha a opção **Excluir** para eliminar a relação ou **Propriedades** para ajustar a relação.



A linha amarela em destaque representa a ligação da dimensão Data com a tabela Fato agregada Agr_Lucro_Produto. A relação define que um elemento contido na dimensão Data pode estar relacionado a várias linhas contidas na tabela Fato Agr_Lucro_Produto.

A janela ilustrada a seguir apresenta as tabelas contidas na relação, os campos relacionados e a cardinalidade do relacionamento.

✕

Editar relacionamento

Selecione tabelas e colunas relacionadas umas às outras.

public ft_vendas ▾

sk_produto_vendas	sk_cliente_vendas	sk_data_vendas	sk_categoria_vendas	num_pedido	qt_produto_ve
10	1	20200620	1	98563	
10	6	20200619	1	95687	
10	7	20200620	1	95687	

public dim_data ▾

date_key	full_date	day_of_week	day_num_in_month	day_num_overall	day_name
20190102	quarta-feira, 2 de janeiro de 2019	3	2	2	Wednesday
20190109	quarta-feira, 9 de janeiro de 2019	3	9	9	Wednesday
20190116	quarta-feira, 16 de janeiro de 2019	3	16	16	Wednesday

Cardinalidade

Muitos para um (*:1) ▾

Direção do filtro cruzado

Único ▾

☒ Ativar este relacionamento

☐ Pressuponha integridade referencial

☐ Aplicar filtro de segurança em ambos os sentidos

OK

Cancelar

Uma vez que o modelo de dados está carregado no Power BI e relacionado corretamente, as consultas podem ser construídas.

Vale destacar que as ferramentas OLAP possuem formas de trabalhar, características e funcionalidades diferentes. A ferramenta deve ser escolhida com base nas necessidades e possibilidades da organização.

Atenção! Aqui existe uma videoaula, acesso pelo conteúdo online

Como exercício, realize os passos de preparação dos dados na ferramenta Power BI apresentados até aqui.

Nesta aula foram examinados o conceito OLAP, as características das ferramentas de OLAP, as operações de análise de dados: Slice, Drill-down, Roll-up, Pivot, Drill-across, Dice, Drill-through e a ferramenta OLAP Microsoft Power BI.

Agora, com base nos conceitos aplicados, vamos fixar o entendimento!

Atividade

Os dados organizados em um modelo de dados dimensional podem ser representados como um cubo de dados de processamento analítico on-line (OLAP). Sobre essa tecnologia é possível afirmar:

- a) Realiza análise de dados multidimensionais e processa operações de cálculos complexos.
 - b) Realiza consultas em dados operacionais.
 - c) Realiza análise de dados multidimensionais e não suporta cálculos complexos.
 - d) Não realiza análise de dados multidimensionais e processa operações de cálculos complexos.
 - e) Não realiza consultas em hierarquias de dados.
-

As ferramentas OLAP dispõem de navegabilidade nos dados permitindo realizar operações:

- a) Slicing, Pivoting e ETL.
 - b) MOLAP, Pivoting e Drill through.
 - c) MOLAP, Pivoting e Drill Across.
 - d) Slicing, ROLAP e Drill through.
 - e) Slicing, Pivoting e Drill through.
-

De acordo com a arquitetura das ferramentas OLAP, elas podem ser classificadas como:

- a) EOLAP, ROLAP, MOLAP e HOLAP.
- b) ROLAP, MOLAP, HOLAP e XOLAP.
- c) EOLAP, MOLAP, HOLAP e NOLAP.
- d) ROLAP, MOLAP, HOLAP e DOLAP.
- e) ROLAP, FOLAP, HOLAP e DOLAP.

Notas

Título modal ¹

Lorem Ipsum é simplesmente uma simulação de texto da indústria tipográfica e de impressos. Lorem Ipsum é simplesmente uma simulação de texto da indústria tipográfica e de impressos. Lorem Ipsum é simplesmente uma simulação de texto da indústria tipográfica e de impressos.

Título modal ¹

Lorem Ipsum é simplesmente uma simulação de texto da indústria tipográfica e de impressos. Lorem Ipsum é simplesmente uma simulação de texto da indústria tipográfica e de impressos. Lorem Ipsum é simplesmente uma simulação de texto da indústria tipográfica e de impressos.

Referências

BARBIERI, C. BI2 – **Business Intelligence**: Modelagem e Qualidade. Rio de Janeiro: Campus-Elsevier, 2011.

KIMBALL, M. R. R. **The Data Warehouse Toolkit - The Definitive Guide to Dimensional Modeling**. 3. ed. Indianapolis: John Wiley Sons, 2013.

Próxima aula

- A construção das análises na ferramenta Power BI;
- A linguagem de expressão de fórmulas DAX;
- Publicação dos relatórios Power BI.

Explore mais

- Conheça outras visualizações disponíveis no Power BI acessando o site da Microsoft.
- Caso você não tenha completado a carga dos dados no banco de dados dimensional nas aulas anteriores, poderá fazer o download aqui de um arquivo de backup para restaurar o database no PostgreSQL e realizar o exercício desta aula no Power BI.