



#### Monitoramento de Dados com Dashboards

**AULA 5.1. ETL ELT ELTL** 

PROF. LEONARDO MAURÍCIO CERQUEIRA

#### Nesta aula



- ☐ Os processos de cargas existentes:
  - > ETL: Extração Transformação e Carga.
  - > ELT: Extração Carga e Transformação.
  - > ELTL: Extração Carga Transformação e Carga.

#### Processos de carga - ETL



#### Conceito:

A disciplina de integração de dados compreende as práticas, técnicas de arquitetura e ferramentas para alcançar o acesso e a entrega consistentes de dados, em todo o espectro de áreas de assunto de dados e tipos de estrutura de dados na empresa, para atender aos requisitos de consumo de dados de todos os aplicativos e processos de negócios.

(...) Nos últimos anos, a maior parte da atividade esteve no mercado de ferramentas ETL.



# TRANSFORM

EXTRACT & LOAD



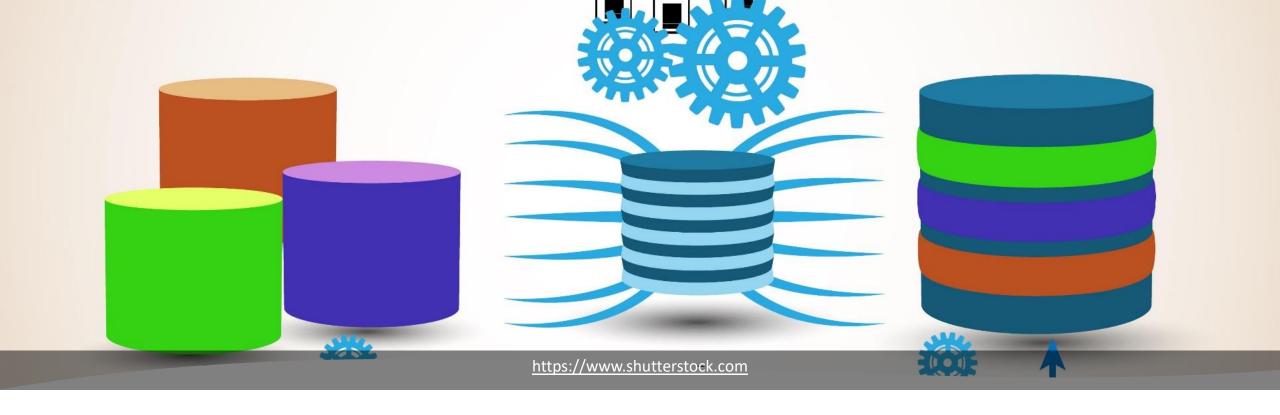
TRACT

TRANSFORM



LOAD





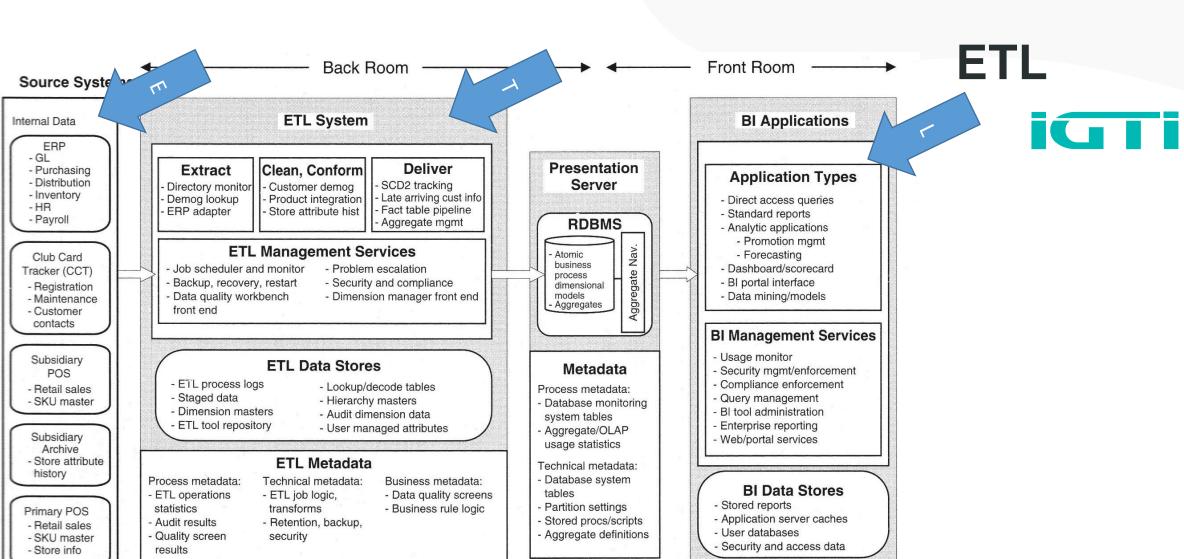
- **E** Extração
- **T** Transformação
- L Carga

- Extração: mapeamento das fontes de dados, conectores e estratégias de janelas de carga.
- **Transformação**: tratamentos diversos, limpeza de dados, filtros, regras de cálculo, unificação de fontes etc.
- Carga (Loader): armazenamento de dados, persistência das informações coletadas e tratadas.



"Construir um sistema de ETL é extraordinariamente difícil, porque está sujeito a realidades inevitáveis".

Ralph Kimball, 2005



External Data

Lookup Service

Demographics

#### **BI** Metadata

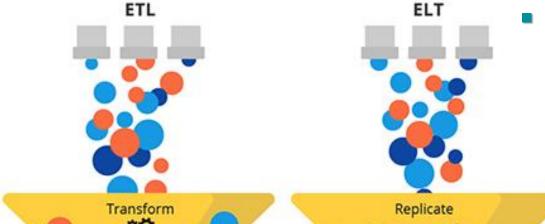
Process metadata:

- Usage statistics
- Technical meta data:
- BI tool semantic laver
- Standard query/report definitions
- Security groups/assignments Business metadata:
- Conformed attribute list Hear decumentation

**ETL->** O dado é armazenado após um tratamento de REGRA.







**ELT ->** A transformação (limpeza, análise, tratamentos em geral) vem depois do dado armazenado.

PRONTO para uso.



Uso APÓS Tratamento.

#### DATA LAKE

igti

É um termo recente, criado pelo CTO (Chief Technical Officer) do Pentaho, James Dixon.

A ideia é ter um único repositório dentro da empresa, para que todos os dados brutos estejam disponíveis a qualquer pessoa que precise fazer análise sobre eles.

https://jamesdixon.wordpress.com/2010/10/14/pentahohadoop-and-data-lakes/



#### **ETL - DATA LAKE**

iGTi

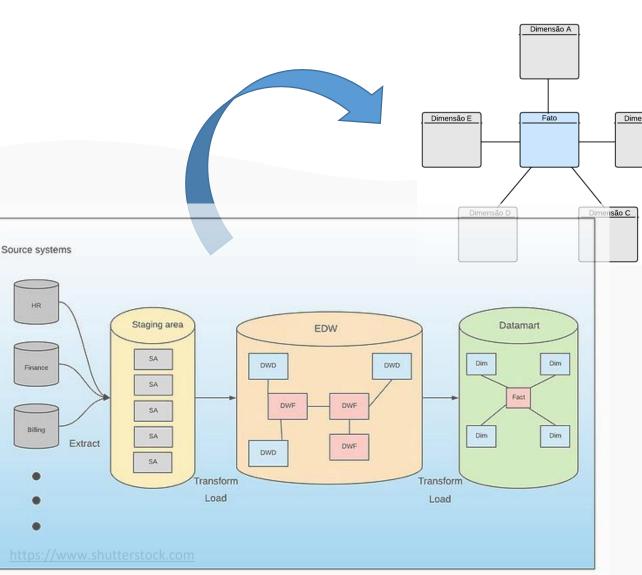
Um lago de dados é como um reservatório.

Primeiro você cria a estrutura (um cluster) e depois enche de água (dados). Depois que o lago estiver pronto, você começa a usar a água (dados) para várias finalidades, como geração de energia e consumo (análises preditivas etc.).



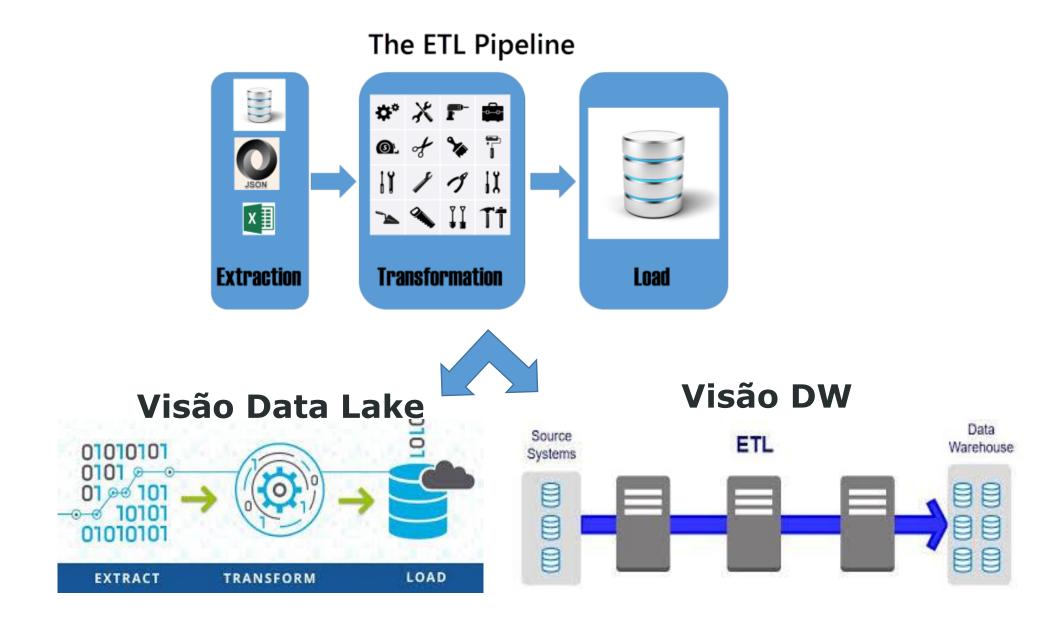
#### **ETL - Data Warehouse**





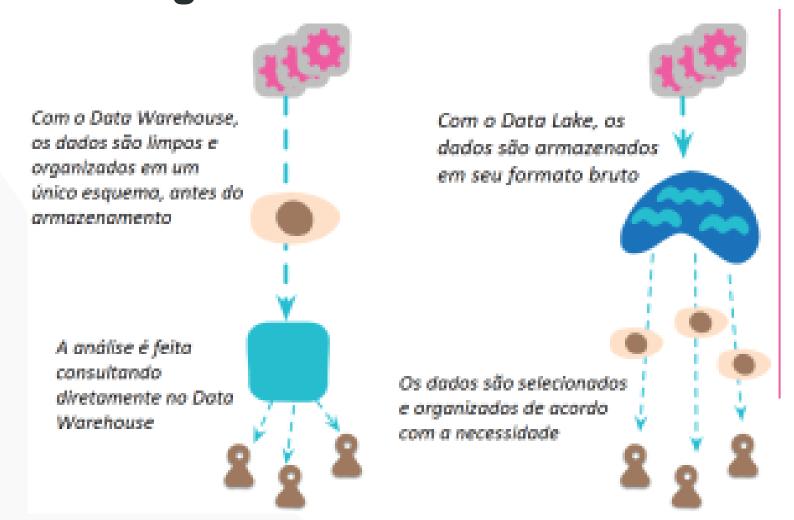
Base de dados utilizada para armazenar informações relativas às atividades de uma organização em bancos de dados, de forma consolidada. O desenho da base de dados favorece os relatórios, a análise de grandes volumes de dados e a obtenção de informações estratégicas, que podem facilitar a tomada de decisão.

#### **ETL - Conceitos**



#### Abordagem Data Warehouse x DataLake









#### **ETL – Data Warehouse**

## Traditional Approaches

Current state of a data warehouse

LOB

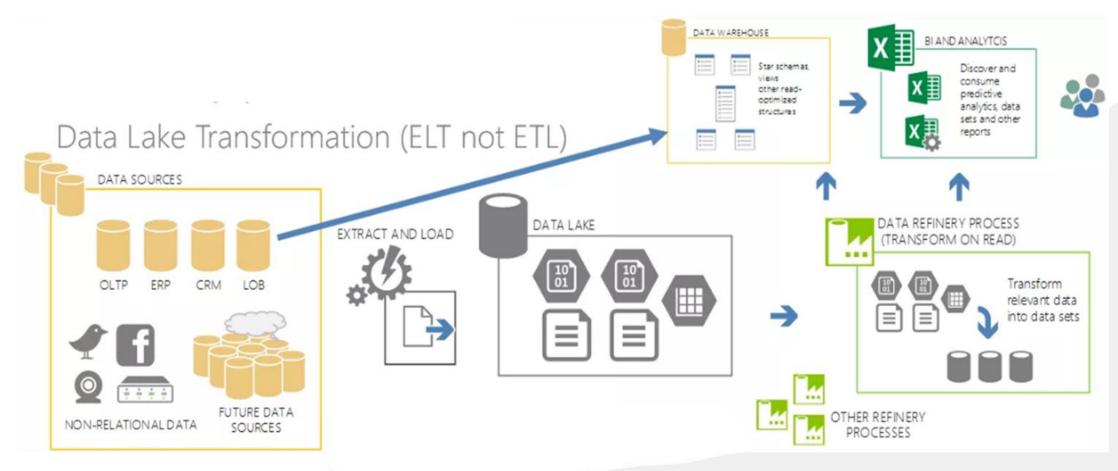
DATA SOURCES



MONITORING AND TELEMETRY

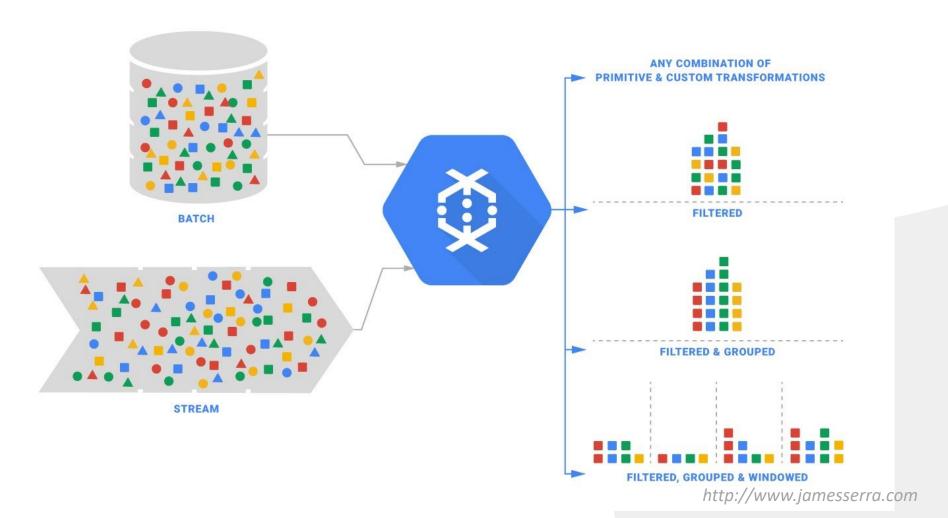


## ETL x ELT x ELTL

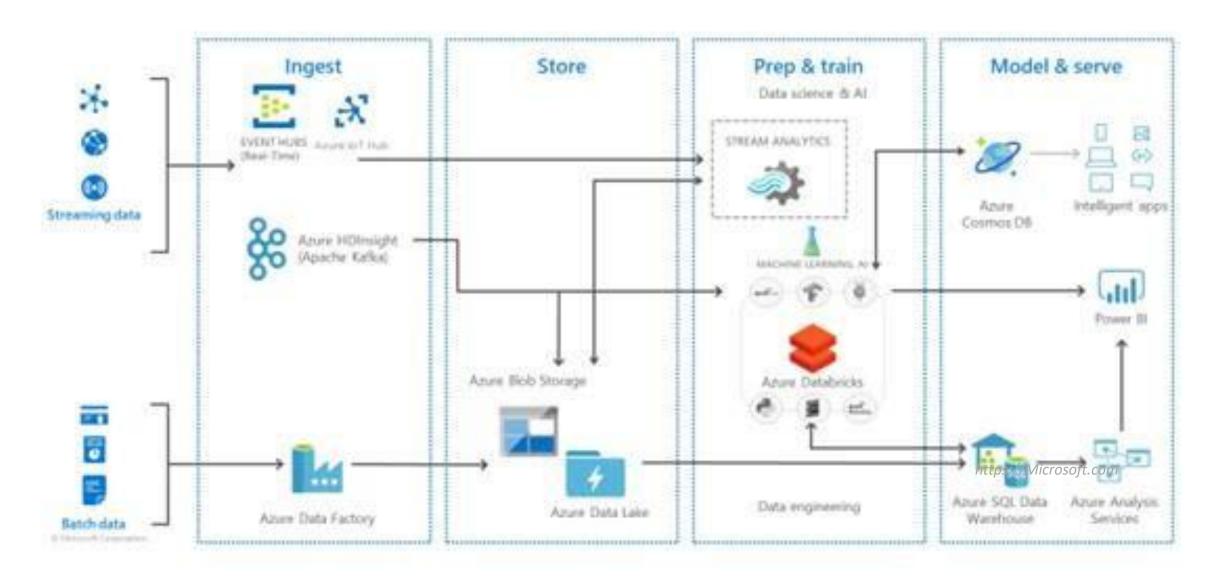




## Arquitetura



#### Arquitetura Moderna de BI



## "In God we trust, but all others bring data"

Eric Schmidt, ex-CEO Google



#### Laboratório

Laboratório - O processo ETL/ELT na carga do POWERBI:

5.2.1 – Carga Sem Transformações.

. Arquivo texto COVID19

5.2.2 – Carga com transformações.

.Fonte de dados de Vendas







Análise de dados tratados sobre os casos e mortes de COVID 19 no Brasil.

Ferramenta: POWERBI

Fonte: Flat File (COVID 19)





iGTi

Análise de dados de Vendas ABC.

Ferramenta: POWERBI.

Fonte: Flat File Conversão Dólar.

Fonte: Baco de Dados na SQL (AZURE).



#### Conclusão



- ✓ Os processos de cargas existentes:
  - ✓ ETL: Extração Transformação e Carga.
  - ✓ ELT: Extração Carga e Transformação.
  - ✓ ELTL: Extração Carga Transformação e Carga.

#### Próxima aula

igti

- ☐ Definição de Linhas de cálculo:
  - Média.
  - Mediana.
  - Máximas e Mínimos.
  - Projeções.

#### Monitoramento de Dados com Dashboards

**AULA 5.2. LINHAS DE CÁLCULO** 

PROF. LEONARDO MAURÍCIO CERQUEIRA

#### Nesta aula



- □ Definição de Linhas de cálculo:
  - Média, mediana, máximas e mínimos.
  - Projeções.
  - > Indicadores KPIs.

A média aritmética, Moda e Mediana, são considerados métodos estatísticos para avaliação de tendências diversas com focos diferenciados, para trazer atenção a tendências.

No caso deste estudo, é importante perceber que hoje muitos desses cálculos são naturalmente calculados pelas ferramentas de BI.

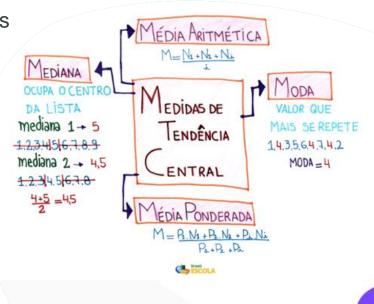
Por exemplo: média é o resultado da <u>divisão</u> do somatório dos números dados pela quantidade de números somados.

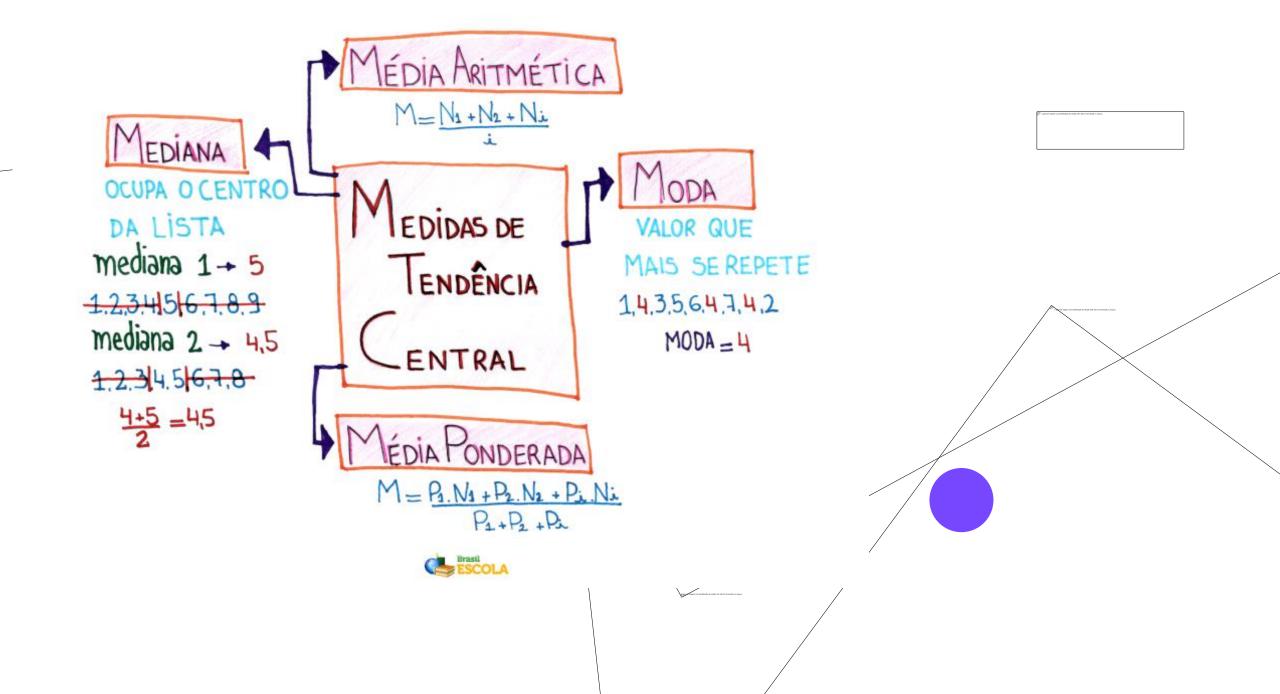
Vamos determinar a média dos números 3, 12, 23, 15, 2.

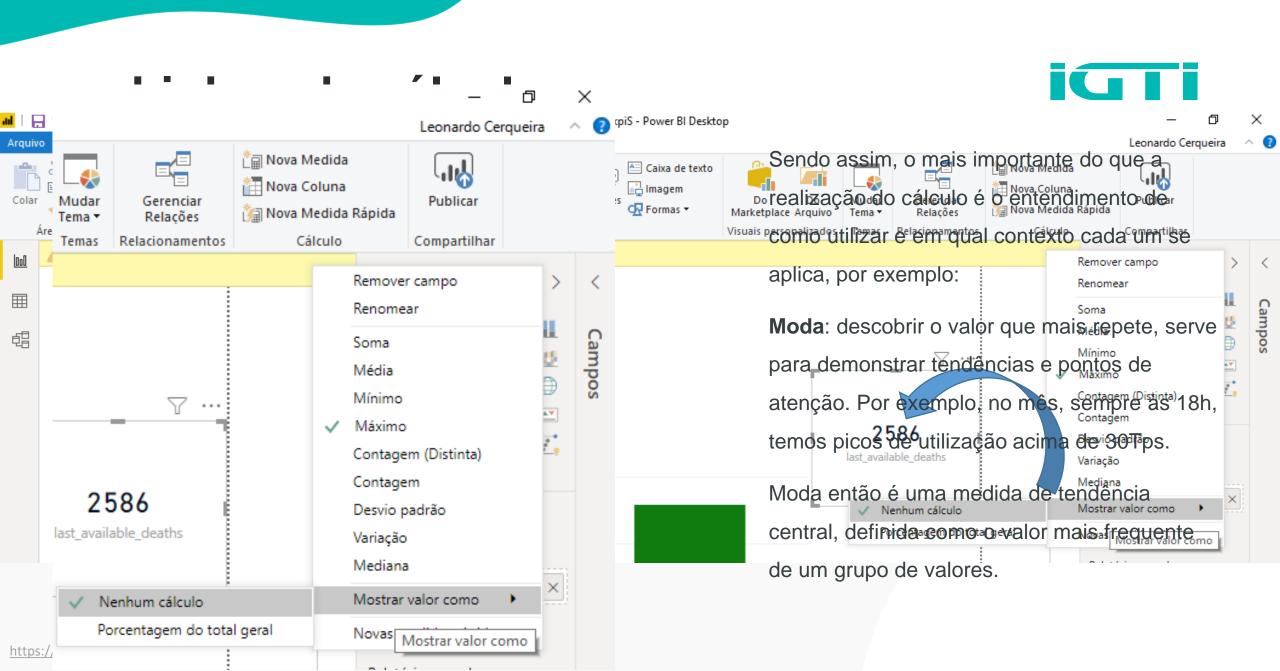
$$Ma = (3+12+23+15+2) / 5$$

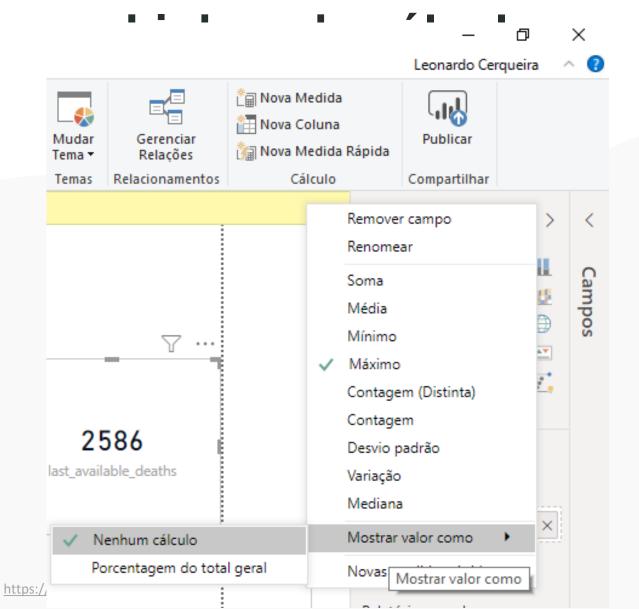
$$Ma = 55 / 5$$

Ma = 11 -> A **média** dos números é igual a 11.











Média: é considerada uma medida de tendência central e é muito utilizada no cotidiano. Contudo, com valores muitos distintos não é uma boa medida, pois um valor "esconde" o efeito do outro.

A Média Salarial de um setor é de R\$ 36.933,34.

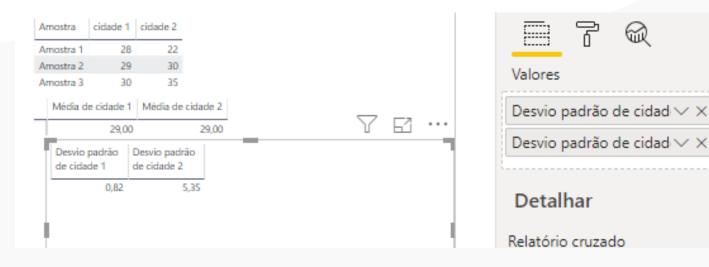
1 Gerente: 105.900,00.

1 Analistas: R\$ 4.500,00 (cada).

1 Estagiários: R\$400,00 (cada).

CUIDADO!!





Desvio padrão: se medirmos a temperatura máxima durante três dias em uma cidade e obtivermos os seguintes valores: 28°, 29° e 30°, e, em outra cidade, as temperaturas máximas nesses mesmos dias podem ter sido 21°, 30° e 35°, em ambos a média desses três dias foi 29°.

Mas, as duas cidades viveram situações iguais?

0,82 5,35

Quanto menor o desvio padrão, mais homogênea é a amostra.



AMOSTRA	SUCO DE LARANJA
Garrafa 1	446 ml
Garrafa 2	450 ml
Garrafa 3	554 ml
Garrafa 4	547 ml
Garrafa 5	486 ml
Garrafa 6	498 ml
Garrafa 7	440 ml
Garrafa 8	560 ml
Garrafa 9	451 ml
Garrafa 10	568 ml
Média	500 ml
Desvio Padrão	52,7 ml

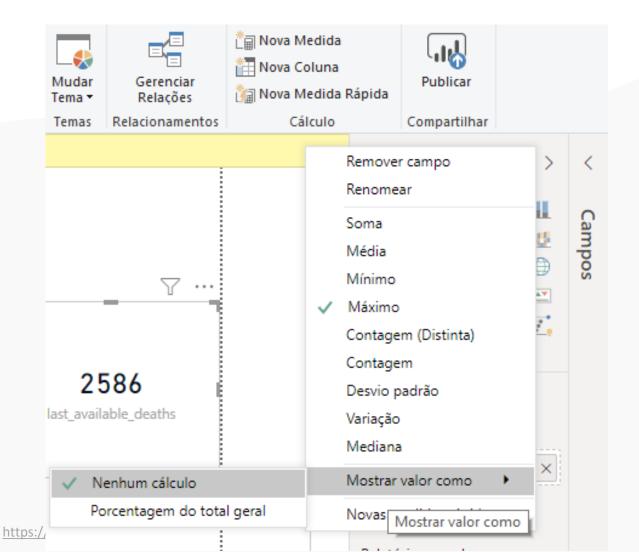
AMOSTRA	SUCO DE U
Garrafa 1	506 ml
Garrafa 2	507 ml
Garrafa 3	502 ml
Garrafa 4	498 ml
Garrafa 5	496 ml
Garrafa 6	495 ml
Garrafa 7	492 ml
Garrafa 8	503 ml
Garrafa 9	508 ml
Garrafa 10	493 ml
Média	500 ml
Desvio Padrão	6,0 ml

Desvio padrão: este tipo de medida deve ser usado para trazer uma noção do quão homogênea é a amostra, pois somente a média aritmética pode nos induzir a interpretações, erradas sobre a qualidade da amostra e a realidade do que está sendo exposto.

Veja mais um exemplo, como ao lado onde as médias são idênticas, mas a o DP (desvio padrão) demonstra que a qualidade do "suco de Laranja" é pior, tem mais variações que o "suco de uva".

https://operdata.com.br/blog/desvio-padrao-e-erro-padrao/





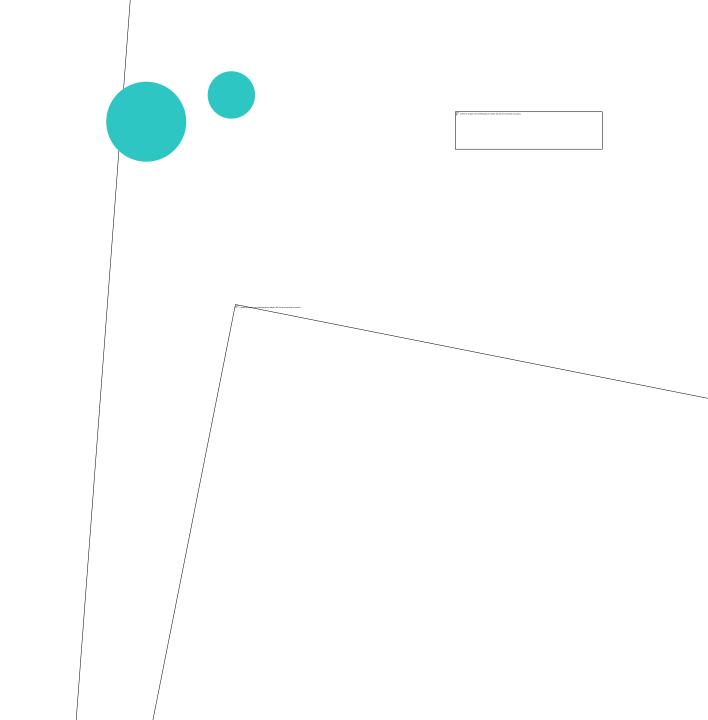
**Mediana**: a mediana é geralmente utilizada para retornar a tendência central para distribuições numéricas distorcidas.

**Mínimo e máximo**: menor e maior valor de uma amostragem, serve para destacar informações fora de um padrão.

## Laboratório linhas de cálculo

Ferramenta: POWER BI.

Arquivo Flat File de Vendas 2.



#### Conclusão



- ✓ Definição de Linhas de cálculo:
  - ✓ Média, mediana, máximas e mínimos.
  - ✓ Projeções.
  - ✓ Indicadores KPIs.

#### Próxima aula

- ☐ Construção e utilização de filtros em dashboards.
- ☐ Operações OLAP.
- ☐ Filtros de Objetos:
  - > Filtros de Página.
  - > Filtros de Relatório.
  - Filtros de Detalhamento (Drill Trought).



#### Monitoramento de Dados com Dashboards

AULA 5.3. CONSTRUÇÃO E UTILIZAÇÃO DE FILTROS EM DASHBOARDS

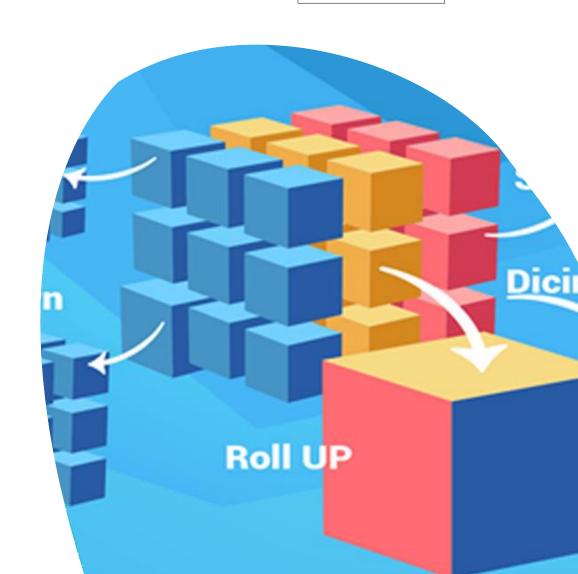
PROF. LEONARDO MAURÍCIO CERQUEIRA

#### Nesta aula



- Construção e utilização de filtros em dashboards.
- Operações OLAP.
- ☐ Filtros de Objetos:
  - Filtros de Página.
  - > Filtros de Relatório.
  - > Filtros de Detalhamento (Drill Trought).

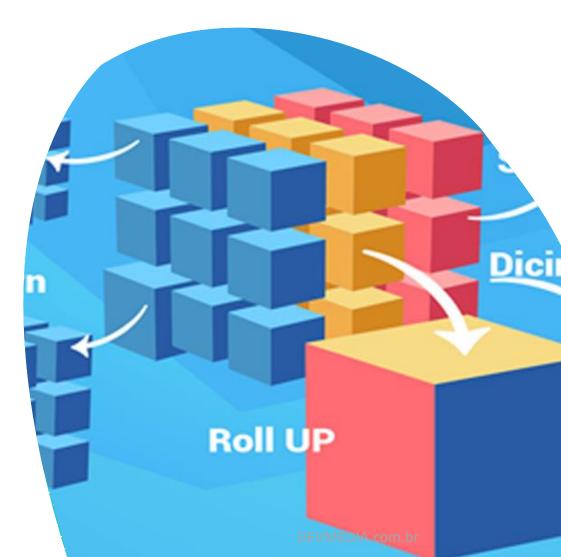
Segundo o glossário do Gartner, as plataformas de Business Intelligence (BI) permitem que as empresas criem aplicativos de BI, fornecendo recursos em três categorias: análise, como processamento analítico online (OLAP); entrega de informação, tais como relatórios e painéis; e integração de plataforma, como gerenciamento de metadados de BI e um ambiente de desenvolvimento.



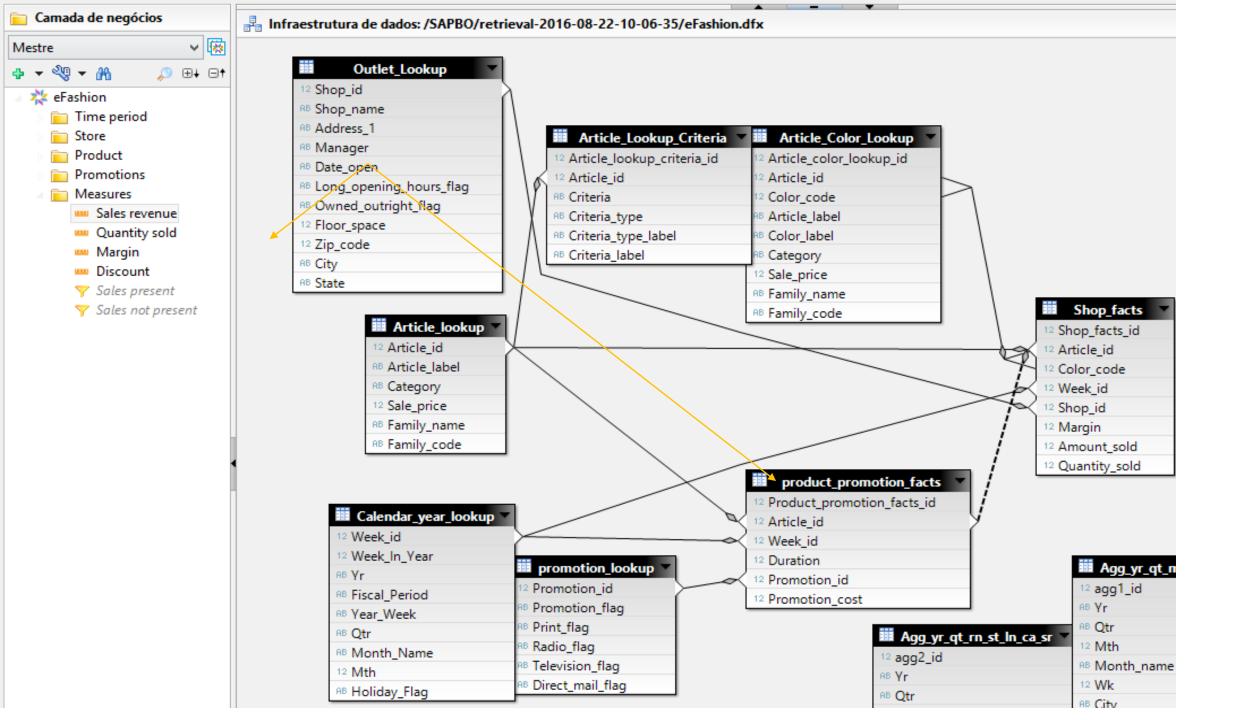
https://gartner.com.br/glossary

Segundo Nery (2006), as ferramentas OLAP permitem ao usuario analisar o motivo dos resultados obtidos. Refere-se a um conjunto de ferramentas voltadas para análise de informações, sob diversas perspectivas.

Com o objetivo final de transformar dados em informações capazes de dar suporte às decisões gerenciais, de forma amigável e flexível ao usuário e em tempo hábil. (..) É uma ferramenta muito importante no contexto gerencial, ajudando a analisar de forma mais eficiente a quantidade de dados crescente armazenada pelas organizações, transformando-os em informação útil.



- Ponto importante definido pela hierarquia das dimensões/tabelas/conjunto de dados.
- Permite que atributos de um conjuntos de dados relacionem-se com outros atributos do mesmo.
- Especifica níveis de agregação e, portanto, granularidade dos itens de dados.
- Pode ter múltiplas hierarquias além de outros atributos descritivos.

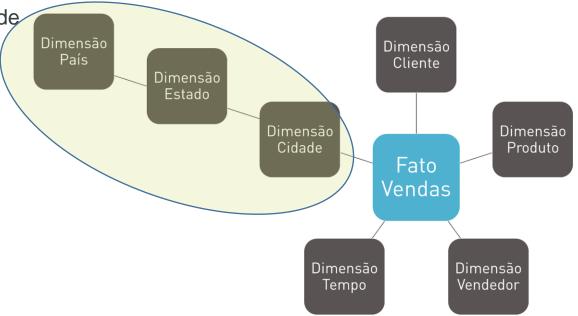


### Hierarquia

- Uma dimensão pode ter múltiplas hierarquias, além de outros atributos descritivos.
- Permite que atributos de uma dimensão relacionem-se com outros atributos da mesma dimensão.

 Especifica níveis de agregação e, portanto, granularidade dos itens de dados.

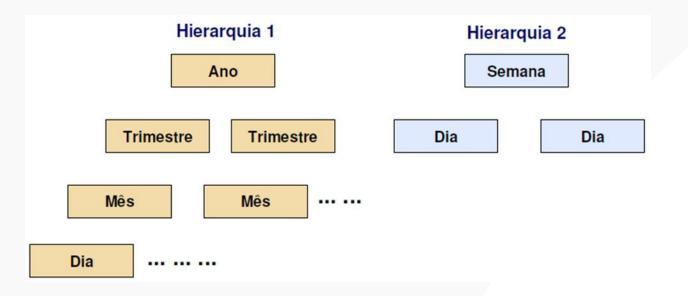
- Exemplo para uma empresa atacadista:
  - Geografia física:
    - o cep, cidade, estado, região, país
  - Geografia de vendas:
    - território, região, zona



#### Hierarquia

iGTi

- É uma classificação de dados dentro de uma dimensão.
- Exemplo para uma empresa atacadista:
  - o Geografia física: cep, cidade, estado, região, país.
  - o Geografia de vendas: território, região, zona.

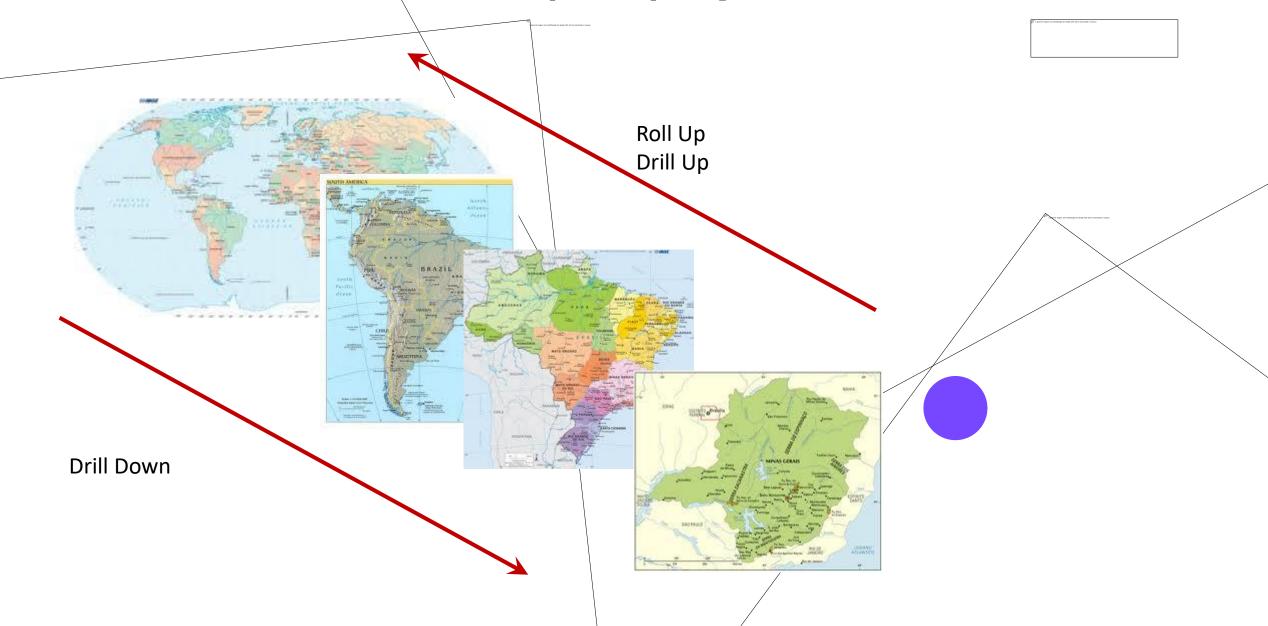


#### Drill down X Drill (Roll) Up



- Drill down: alto nível → baixo nível.
  - Ocorre quando o usuário aumenta o nível de detalhe da informação, diminuindo o grau de granularidade.
- Roll up: baixo nível → alto nível.
  - Ocorre quando o usuário diminui o nível de detalhe da informação, aumentando o grau de granularidade.
- Com o Drill Down e o Roll up, você pode "subir ou descer" dentro do detalhamento do dado, como, por exemplo, analisar uma informação tanto diariamente quanto anualmente, partindo da mesma base de dados.

## Drill down X Drill (Roll) Up



## Drill down X Drill (Roll) Up



MES	BANCO	REGIAO	QUANTIDADE EMITIDA
jan/13	Banco do Brasil	Sul	100
jan/13	Santander	Sul	120
jan/13	Banco do Brasil	Sudeste	250
jan/13	Santander	Sudeste	160
fev/13	Bradesco	Sul	150
fev/13	Banco do Brasil	Sul	130
fev/13	Santander	Sul	120
fev/13	Bradesco	Sudeste	200
fev/13	Banco do Brasil	Sudeste	230
fev/13	Santander	Sudeste	220

Roll up

	DANICO	DECLEO	FOTABO	OLIANTIDADE
MES	BANCO	REGIAO	ESTADO	QUANTIDADE EMITIDA
jan/13	Banco do Brasil	Sul	Rio Grande do Sul	35
jan/13	Banco do Brasil	Sul	Santa Catarina	32
jan/13	Banco do Brasil	Sul	Paraná	33
jan/13	Santander	Sul	Rio Grande do Sul	41
jan/13	Santander	Sul	Santa Catarina	39
jan/13	Santander	Sul	Paraná	40
jan/13	Banco do Brasil	Sudeste	Minas Gerais	60
jan/13	Banco do Brasil	Sudeste	São Paulo	80
jan/13	Banco do Brasil	Sudeste	Rio de Janeiro	70
jan/13	Banco do Brasil	Sudeste	Espírito Santo	40
jan/13	Santander	Sudeste	Minas Gerais	40
jan/13	Santander	Sudeste	São Paulo	55
jan/13	Santander	Sudeste	Rio de Janeiro	35
jan/13	Santander	Sudeste	Espírito Santo	30
fev/13	Bradesco	Sul	Rio Grande do Sul	60
fev/13	Bradesco	Sul	Santa Catarina	40
fev/13	Bradesco	Sul	Paraná	50
fev/13	Banco do Brasil	Sul	Rio Grande do Sul	50

**Drill down** 

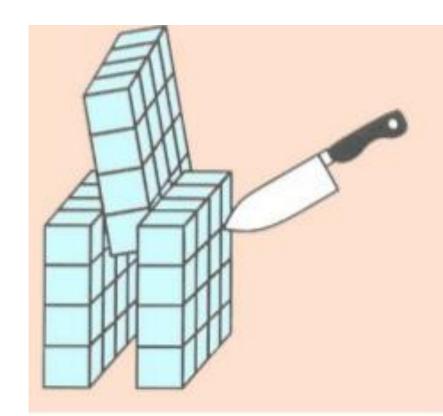
## Slice and Dice

Slice and Dice significa fatiar o Cubo:

SLICE: selecionar dados de uma única dimensão.

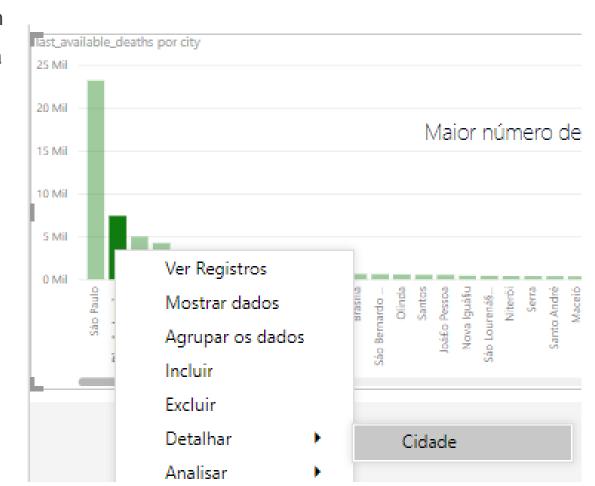
**DICE**: extrai um subcubo da seleção de duas ou mais dimensões.

Em resumo seria o mesmo que filtrar, a aplicação e filtros trazem informações e mais controle sobre os dados que se pretende analisar.



## **Drill Trought**

A capacidade de detalhar uma informação do CUBO em um subconjunto de dados mais detalhados, algumas vezes fora do context do cubo, por exemplo indo para outros sistemas e/ou fontes de dados.



#### **Drill Across**

A Drill Across significa saltos na hierarquia da dimensão, ou seja, imagine que temos uma dimensão tempo que é composta por ano, semestre, trimestre, mês e dia.

O usuário estará executando um Drill Across quando ele passar de ano direto para o mês.

Volume de Produção		Telefone Celular		Pagers	
(em milhar	(em milhares)		1002	2001	2002
Região Sul	RS	33	12	8	12
	SC	45	34	20	23



Dimensão localização geográfica

Membro RS (de ano para mês)

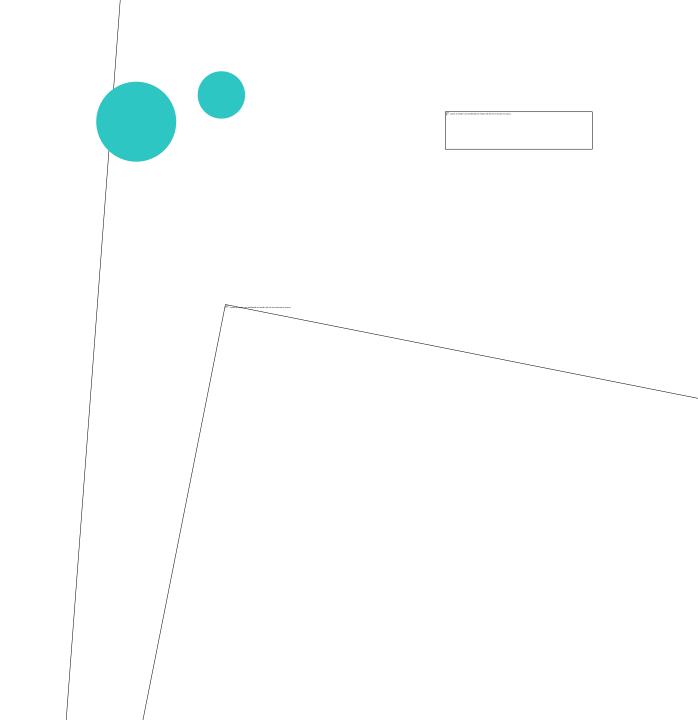
Volume de Produção		Telefone Celular		Pagers	
(em milhar	es)	Janeiro 01	Janeiro 02	Janeiro 01 Janeiro	
Região	RS	2	4	1	4
Sul	SC	5	3	3	2

Operação	Definição
drill-down	analisa os dados em níveis de agregação progressivamente mais detalhados, ou de menor granularidade
roll-up	analisa os dados em níveis de agregação progressivamente menos detalhados, ou de maior granularidade
slice and dice	restringe os dados sendo analisados a um subconjunto destes dados slice: corte para um valor fixo dice: seleção de faixas de valores
pivot	reorienta a visão multidimensional dos dados, oferecendo diferentes perspectivas dos mesmos dados
drill-across	compara medidas numéricas distintas que são relacionadas entre si através de pelo menos uma dimensão em comum

# Laboratório Operações OLAP

Ferramenta: POWER BI.

Arquivo Flat File de Vendas 2.



#### Conclusão



- ✓ Construção e utilização de filtros em dashboards.
- ✓ Operações OLAP.
- √ Filtros de Objetos:
  - ✓ Filtros de Página.
  - ✓ Filtros de Relatório.
  - ✓ Filtros de Detalhamento (Drill Trought).

#### Próxima aula



- ☐ Análises geográficas:
  - Utilização de latitude e longitude.
  - Utilização de nomes geográficos (cidades, estados, países...).
  - Gráficos de mapa tipo bolhas.
  - Gráficos Corolopléticos.
  - Mapas ARCGIS.



#### Monitoramento de Dados com Dashboards

**AULA 5.4. ANÁLISES GEOGRÁFICAS** 

PROF. LEONARDO MAURÍCIO CERQUEIRA

#### Nesta aula



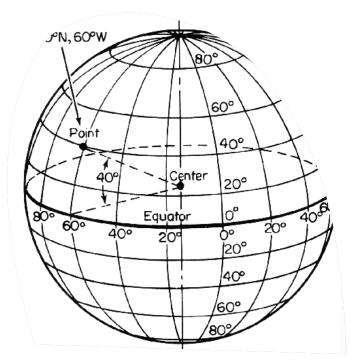
- Análises geográficas:
  - Utilização de latitude e longitude.
  - Utilização de nomes geográficos (cidades, estados, países...).
  - Gráficos de mapa tipo bolhas.
  - Gráficos Corolopléticos.
  - Mapas ARCGIS.

#### Coordenadas geográficas

Latitude e longitude são sistemas de linhas, usados para indicar a localização de qualquer lugar na Terra.

#### Temos três notações utilizadas:

- Graus, minutos e segundos (DMS, na sigla em inglês): 41°24'12.2"N 2°10'26.5"L.
- Graus e minutos decimais (DMM, na sigla em inglês): 41 24.2028, 2 10.4418.
- Graus decimais (DD, na sigla em inglês): 41.40338, 2.17403.



Esta Foto de Autor Desconhecido está licenciado em CC BY-SA

#### Coordenadas geográficas

#### No site 'Ajuda Google MAPS':

- 1. Ver as coordenadas de um lugar.
- No computador, abra o <u>Google Maps</u>. Se estiver usando o Google Maps em <u>modo Lite</u>, o desenho de um raio aparecerá na parte inferior, e você não poderá ver as coordenadas de um lugar.
- 3. Clique com o botão direito do mouse no lugar ou na área no mapa.
- 4. Selecione "O que há aqui?".
- 5. Na parte inferior, você verá um cartão com as coordenadas.



#### Power BI

Um mapa coroplético usa sombreamento, tonalidade ou padrões para exibir como um valor difere na proporção em uma localização geográfica ou de região. Exibe rapidamente essas diferenças relativas com sombreamento que varia de claro (menos frequente/inferior) para escuro (mais frequente/mais).





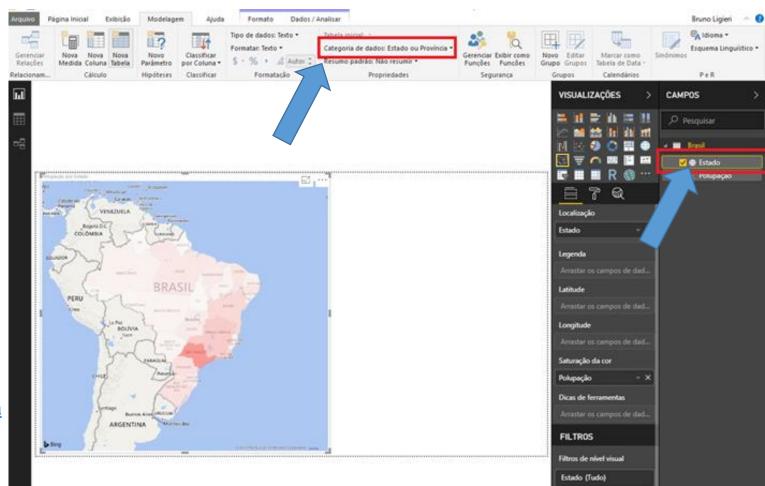
#### igti

#### Power BI

Power BI é integrado ao Bing para fornecer as coordenadas de mapa padrão (um processo chamado geocódigo). Quando você cria uma visualização de mapa no serviço do Power BI ou no Power BI Desktop, os dados nos buckets Local, Latitude e Longitude (que estão sendo usados para criar essa visualização), são enviados ao Bing.

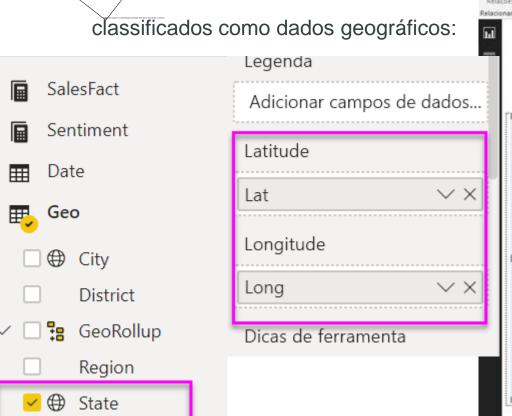
#### Permitir o acesso às URLs:

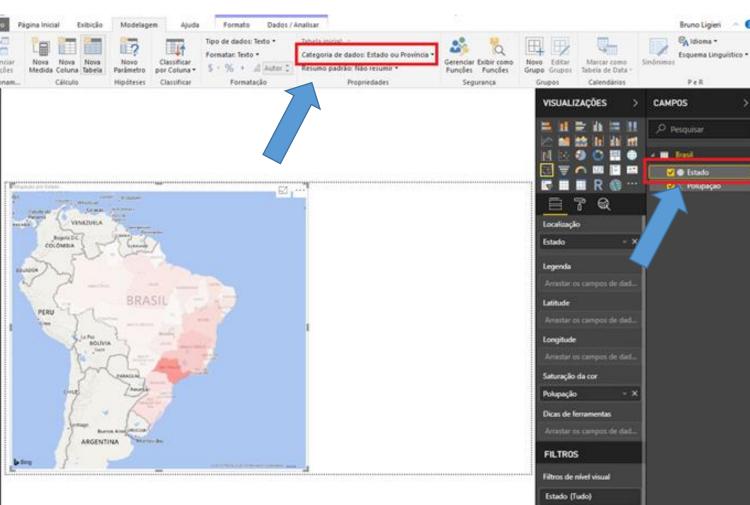
https://dev.virtualearth.net/REST/V1/Locations
https://platform.bing.com/geo/spatial/v1/public/Geodata
https://www.bing.com/api/maps/mapcontrol



#### Power BI

Os dados do POWER BI para correto funcionamento, devem ser classificados como dados geográficos







#### Power BI

Tipo de local	Palavras-chaves suportadas	Geometria
Endereço	cidade, cidades, end, endereço, rua, cidade, capital	pontos
Estado dos Estados Unidos	estado, estados	Polígonos
CEP	zip, zips, zipcode, zipcodes, cep, ceps, código postal, códigos postais	Polígonos
Município dos Estados Unidos	município, municípios	Polígonos
Cidade mundial	cidade, cidades	pontos
País mundial	país, países	Polígonos

# Laboratório Operações OLAP

Ferramenta: POWER BI.

Arquivo base para MAPAS.



#### Conclusão



- ✓ Análises geográficas:
  - ✓ Utilização de latitude e longitude.
  - ✓ Utilização de nomes geográficos (cidades, estados, países...).
  - ✓ Gráficos de mapa tipo bolhas.
  - ✓ Gráficos Corolopléticos.
  - ✓ Mapas ARCGIS.

#### Próxima aula

- ☐ Capítulo 6
- ☐ Princípios da Gestalt:
  - > Proximidade.
  - > Similaridade.
  - Unificação.
  - > Fechamento.

