

Escola Superior de Gestão, Hotelaria e Turismo

Ano letivo 2022/23

Introdução ao *Business Intelligence*

Material de apoio à unidade curricular de
Tecnologias e Sistemas de Informação
Mestrado em Direção e Gestão Hoteleira

Marisol Correia

As TIC na Gestão da Informação e do Conhecimento

- A informação é um recurso que diferencia e valoriza as organizações.
- A gestão da informação e do conhecimento tem permitido (Santos & Ramos, 2017):
 - A implementação de estratégias para aumentar o sucesso das organizações;
 - A melhoria de processos e a inovação de produtos e serviços;
 - A medição do desempenho organizacional.
- As tecnologias de informação e comunicação têm um papel fundamental, ao proporcionarem ferramentas que permitem recolher, armazenar e analisar grandes volumes de informação com rapidez e eficácia.
- Estas ferramentas permitem identificar informação relevante e criar conhecimento, diminuindo a incerteza no processo decisório e proporcionando vantagens competitivas face aos concorrentes.

Business Intelligence (BI)

- Business Intelligence:
 - Sistema que combina a recolha, o armazenamento e a gestão do conhecimento com recurso a ferramentas de análise para apresentar informação interna complexa e competitiva aos gestores da organização/empresa.
- Business Intelligence:
 - Recolher e organizar
 - Analisar e visualizar
 - Partilhar e monitorizar
 - Suporte à tomada de decisão
 - Decisão com base em evidências
 - Medir o desempenho do passado
 - Planear o futuro



Referências principais: (Santos & Ramos, 2017) e (Gartner, 2013)

Business Intelligence

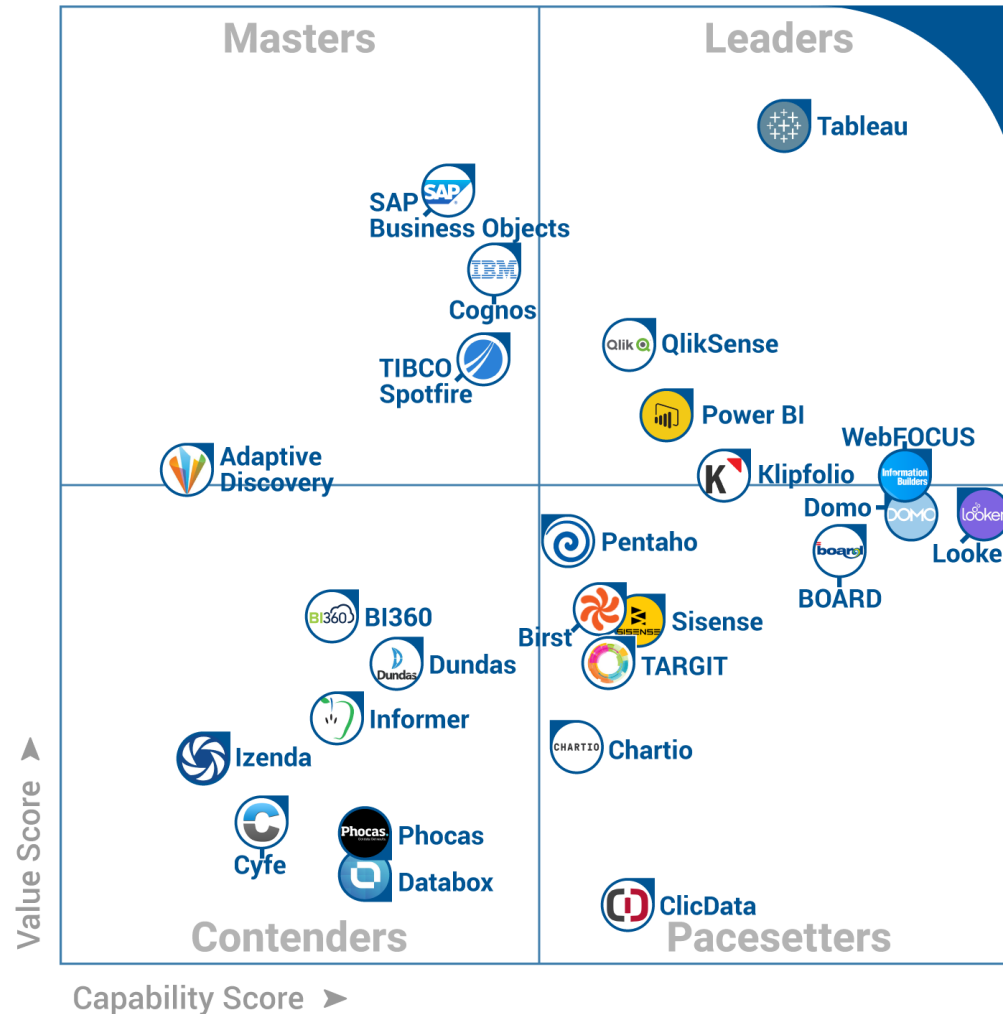
- Os sistemas de BI tem uma utilização bastante abrangente na análise de informação e estão geralmente associados às seguintes tarefas (Santos & Ramos, 2017):
 - Elaborar previsões baseadas em dados históricos, nos desempenhos passados e atuais da organização;
 - Criar cenários que evidenciam o impacto da alteração de diversas variáveis;
 - Permitir o acesso *ad-hoc* aos dados para responder a questões pré-definidas;
 - Analisar detalhadamente a organização, obtendo um conhecimento profundo da mesma.

Business Intelligence

- **Observar.** O que está acontecendo?
- **Compreender.** O que pode ocorrer?
- **Colaborar.** O que a equipa deve fazer?
- **Decidir.** Qual o caminho a seguir?
- **Prever.** O que aconteceria?

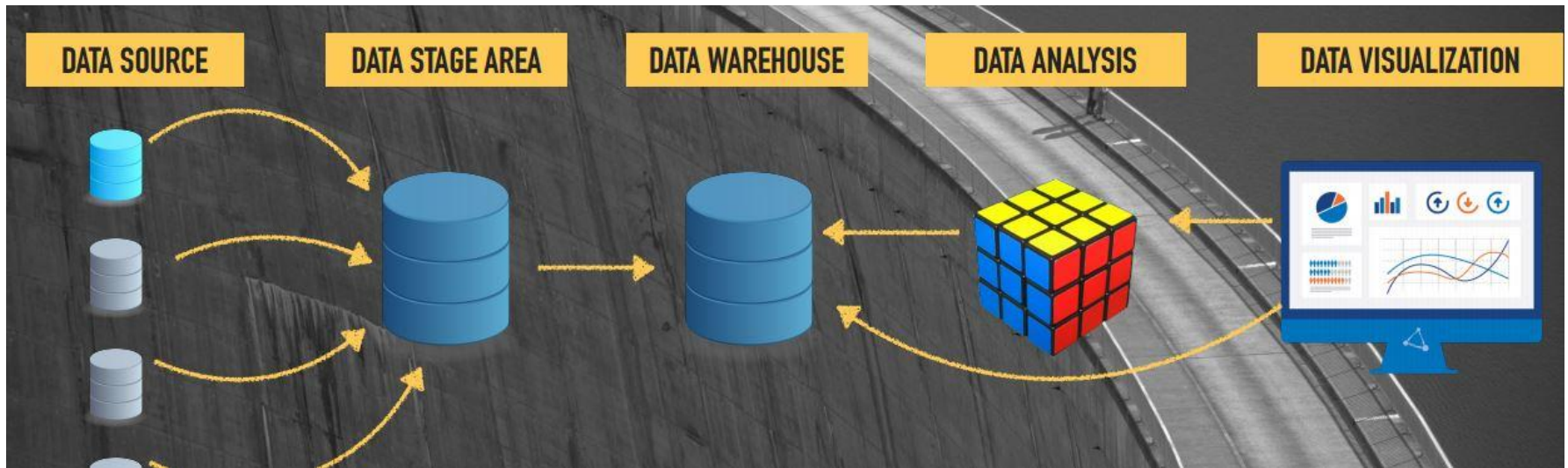


Plataformas de BI



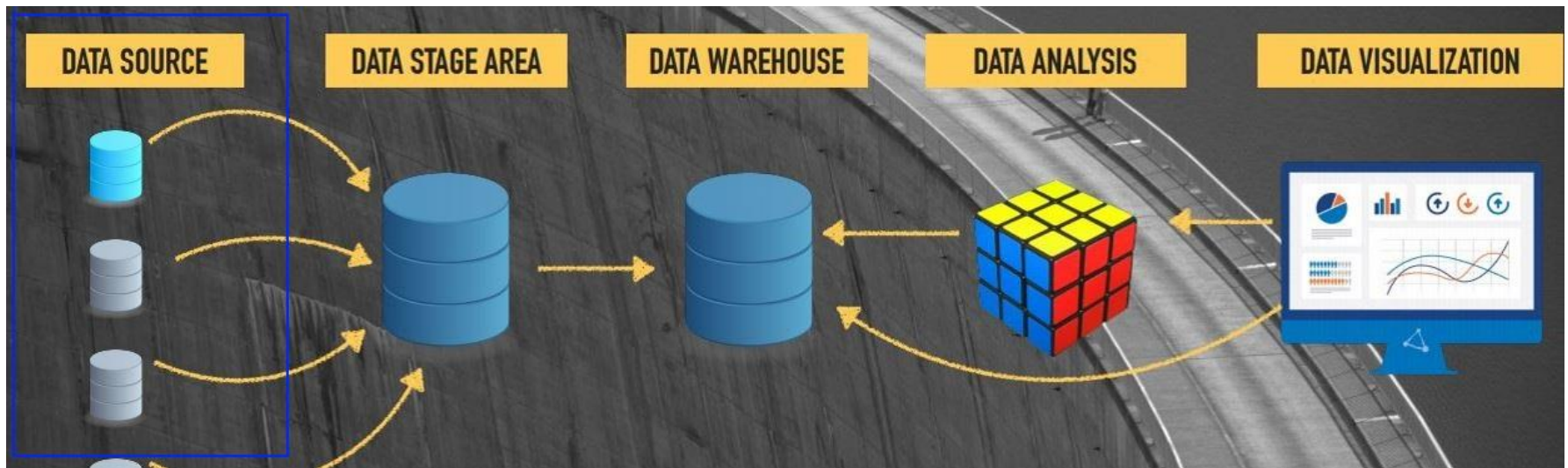
Fonte: (Gartner, 2017)

Arquitetura de BI



Fonte: (Piton, 2018)

Data Source



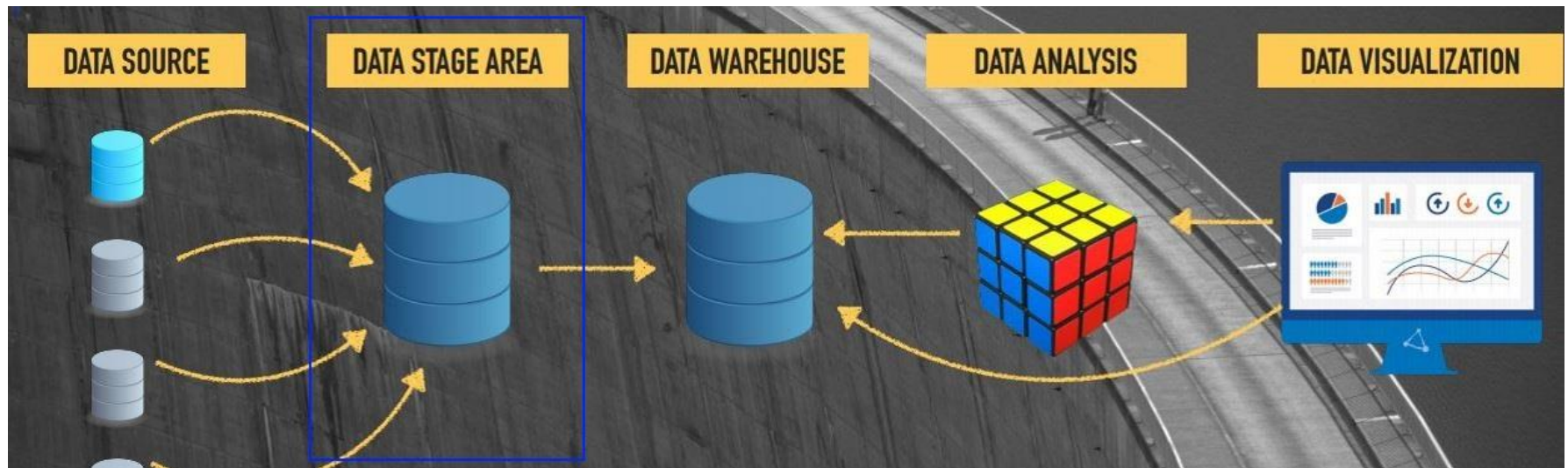
Fonte: (Piton, 2018)

Data Source

- CRM
- Marketing Automation System
- ERP
- Etc.



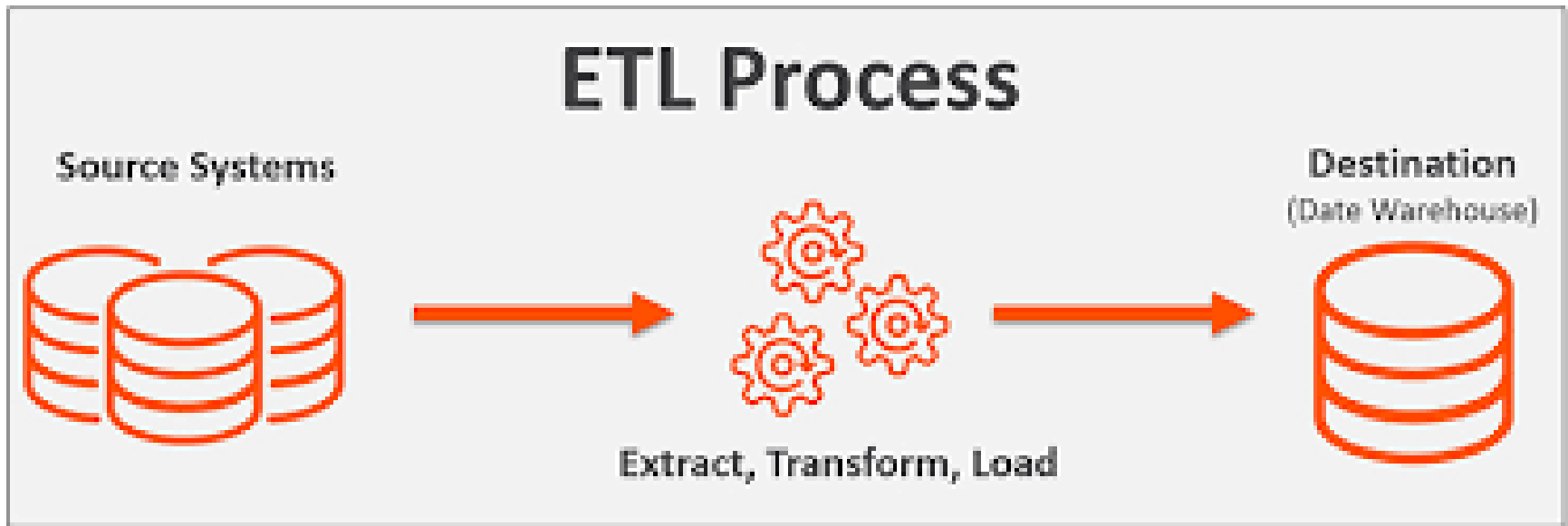
Data Stage Area



Fonte: (Piton, 2018)

Extract, Transform, Load (ETL)

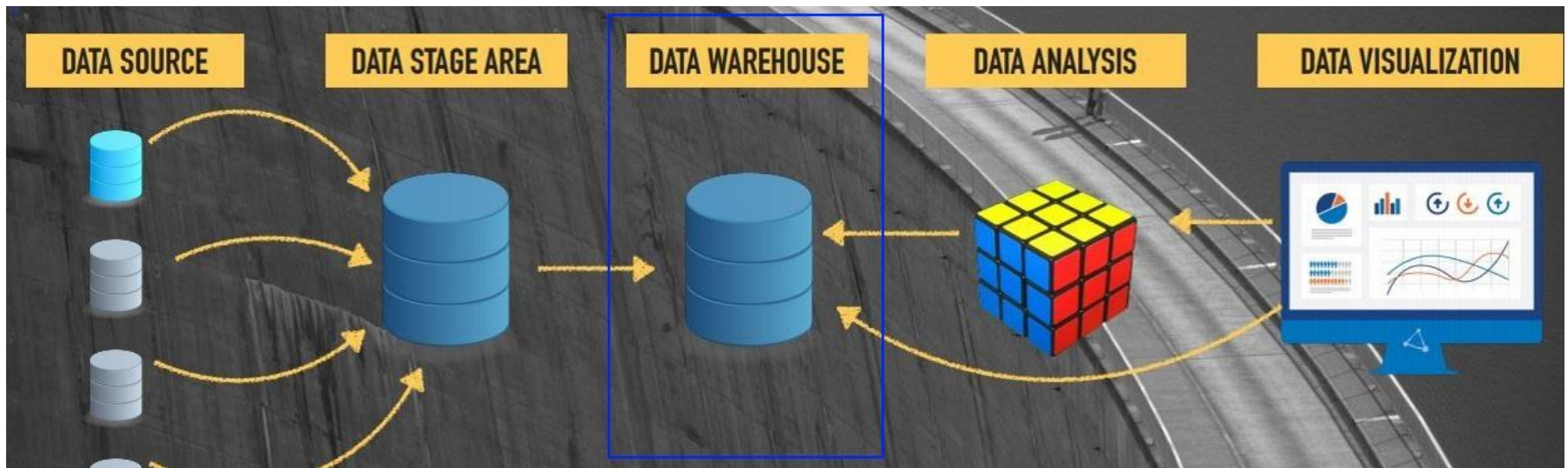
- Processo que efetua a extração, transformação e carregamento de uma ou mais fontes de dados para o *Data Warehouse (DW)*.



Fases do Processo ETL

- **Extração dos dados** - Recolha dos dados de uma ou mais fontes internas ou externas à organização.
- **Transformação dos dados:**
 - **Limpeza** – Tratamento de valores nulos ou omissos, uniformização de valores;
 - **Transformação** – Seleção das informações a inserir, junção de campos, agregações de valores e alteração de formatos.
- **Carregamento dos dados** - Carregar os dados analisados e transformados para o Data Warehouse.
- **Refreshamento** - Atualização do Data Warehouse através do carregamento de novos dados.

Data Warehouse



Fonte: (Piton, 2018)

Data Warehouse (DW)

- É um repositório de dados, para suporte ao processo de tomada de decisão, que se encontra logicamente e fisicamente separado das bases de dados operacionais da organização (Inmon, 2005).
- É uma cópia de dados de transações especificamente estruturadas para consulta e análise. Esta é uma visão funcional de um DW (Kimball, 2013).



Data Warehouse

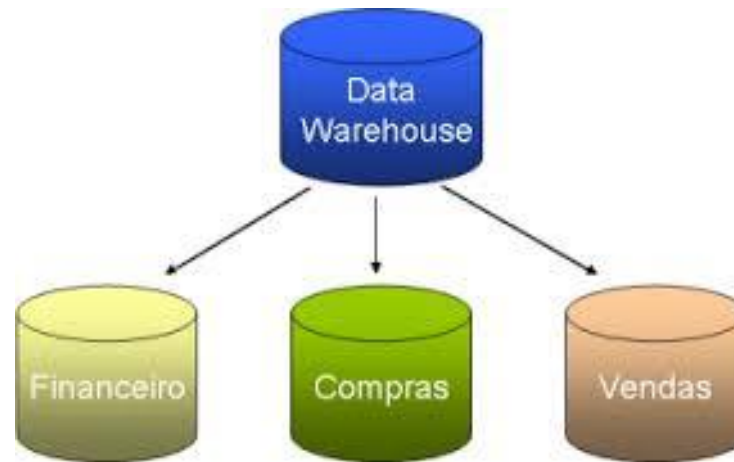
- **Orientados a assuntos:**
 - Orientado aos principais assuntos ou negócios da empresa como, por exemplo, clientes, vendas, produtos, apólices, tratamentos, seguros, viagens, etc.
- **Integrados:**
 - Reúne informações dispersas por diversas bases de dados e distintas plataformas e trabalha de forma a padronizar os termos e as estruturas técnicas que são utilizados nas bases de dados operacionais.
- **Variável ao longo do tempo:**
 - A estrutura dos dados do *Data Warehouse* contém sempre algum elemento de tempo, enquanto que nas bases de dados operacionais isso não acontece obrigatoriamente; nestes sistemas, o horizonte de tempo é normalmente de 2 a 3 meses, enquanto no *Data Warehouse* este horizonte é de 5 a 10 anos.
- **Não volátil:**
 - No ambiente operacional, os dados sofrem as alterações necessárias como: incluir, alterar ou excluir dados; porém, no *Data Warehouse* é só possível duas operações sobre os dados: a sua carga e as consultas.

Vantagens da utilização de um DW

- Reúne informações dispersas por diversas bases de dados e distintas plataformas numa estrutura única.
- A inserção das informações no Data Warehouse implica uma análise prévia e a resolução de inconsistências existentes nos dados.
- Diminuição do tempo que os gestores utilizam para a obtenção das informações necessárias aos seus processos decisórios, uma vez que a informação é estruturada tendo em vista facilitar a pesquisa da informação.
- A informação pode ser armazenada no Data Warehouse por longos períodos de tempo.

Data Mart

- Data Mart é um repositório de dados mais pequeno do que um DW, o qual armazena os dados em uma determinada área funcional ou departamento.



Bases de Dados Transacionais vs Data Warehouse

Bases de Dados Operacionais/Transacionais

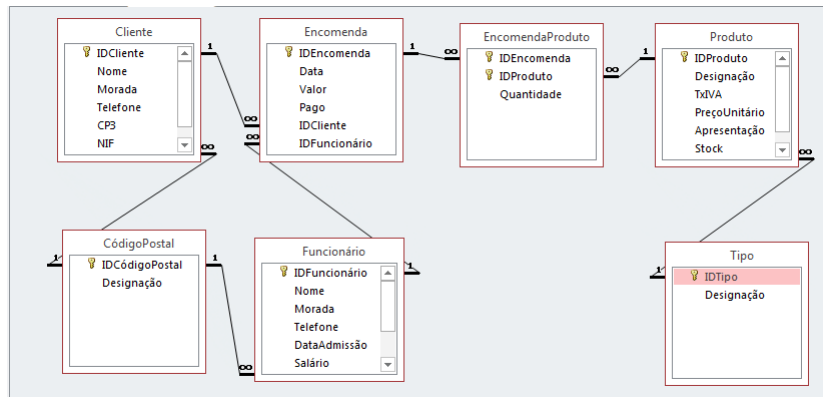
- Desenvolvido numa ótica de otimização do espaço de armazenamento dos dados e numa estrutura que tenta **minimizar os custos associados à redundância de dados**.
A preocupação com a maneira como os dados são **agrupados** e disponibilizados aos utilizadores é mínima ou inexistente.
- **Uma base de dados transacional suporta *on-line transaction processing* (OLTP).**

Data Warehouse

- A preocupação fundamental é otimizar a maneira como os **utilizadores percebem os dados**.
- **A repetição dos mesmos dados em múltiplas tabelas não é assumida como um inconveniente**, mas, ao contrário, como uma maneira de dar aos utilizadores a possibilidade de explorar as informações do *data warehouse* de maneira autónoma.
- **Um DW suporta *on-line analytical processing* (OLAP).**

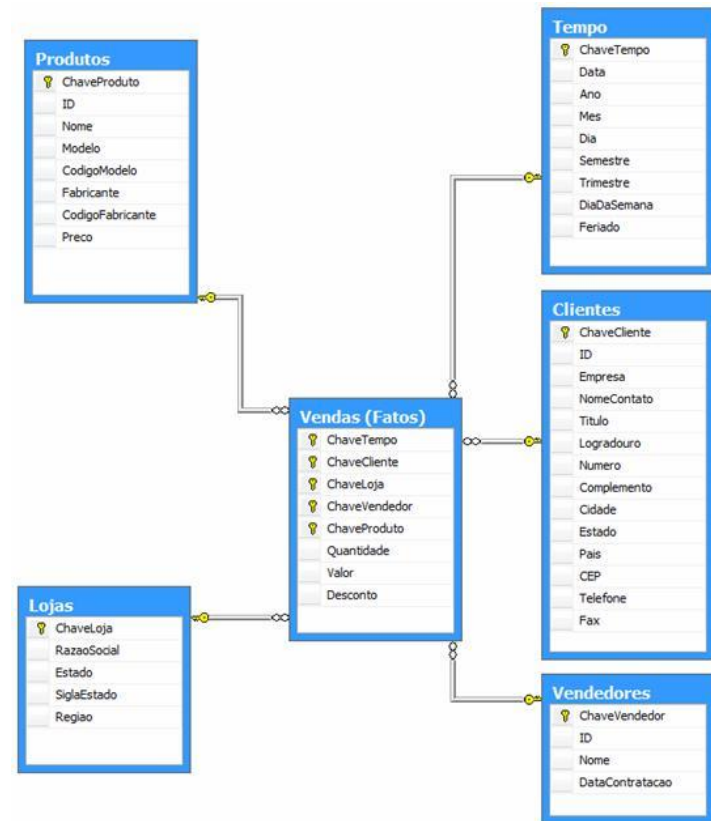
Base de Dados Transaccional vs Data Warehouse

Exemplo de um esquema de base de dados transaccional



Fonte Própria

Exemplo de um esquema de um DW



Fonte: <http://msdn.microsoft.com/pt-br/library/cc518031.aspx>

Modelação Dimensional/Multidimensional

- **Metodologia que procura alcançar os objetivos principais:**
 - Fácil interação com o utilizador final da aplicação;
 - Alto rendimento do processamento de consultas.
- **Conceitos importantes na modelação dimensional:**
 - Dimensões
 - Medidas
 - Factos
 - Agregações
 - Hierarquia
- **Tabelas de factos versus tabelas de dimensões:**
 - Tabela de factos – Armazena os indicadores do negócio, tantos básicos como elementos calculados.
 - Tabela de dimensões – Armazena os dados descritivos.

Dimensões vs Medidas

- **Dimensões:** estabelecem a organização dos dados, determinando possíveis consultas/cruzamentos. Por exemplo: Tempo, Produto, etc.. Os elementos das dimensões são considerados **membros**, que podem estar organizados em diferentes níveis hierárquicos. A dimensão Tempo, por exemplo, pode possuir várias hierarquias: calendário gregoriano (com os níveis ano, mês e dia), calendário fiscal (com os níveis ano, semana e dia), etc..
- **Medidas:** são os valores a serem analisados. Por exemplo: Quantidade de Unidades Vendidas, Valor das Vendas, etc..
- Exemplo (retirado do Contoso):

Rótulos de C		Dimensão: Tempo					Total Soma de SalesQuantity	Total Soma de SalesAmount	
		2007		2008		2009			
Rótulos de Linha	Soma de SalesQuantity	Soma de SalesAmount	Soma de SalesQuantity	Soma de SalesAmount	Soma de SalesQuantity	Soma de SalesAmount			
A. Datum	1035036	273.855.351,06 €	710853	184.006.055,89 €	654746	161.942.346,61 €	2400635	619.803.753,56 €	
Adventure Works	817579	332.911.489,47 €	762987	350.093.193,55 €	840867	406.729.565,72 €	2421433	1.089.734.248,74 €	
Contoso	4853792	587.486.834,25 €	5285335	450.421.544,27 €	7551225	460.012.389,84 €	17690352	1.497.920.768,36 €	
Fabrikam	1110570	744.027.625,54 €	942131	612.001.935,53 €	929336	518.426.293,28 €	2982037	1.874.455.854,35 €	
Litware	148132	75.886.011,99 €	235028	129.518.238,29 €	234083	122.490.560,16 €	617243	327.894.810,44 €	
Northwind Traders	63336	2.282.551,51 €	129446	7.152.268,20 €	186976	10.728.184,27 €	379758	20.163.003,97 €	
Proseware	1173515	418.836.302,04 €	915874	282.088.129,33 €	957346	253.615.789,16 €	3046735	954.540.220,53 €	
Southridge Video	951703	179.329.984,26 €	863209	149.693.316,75 €	990788	142.630.144,02 €	2805700	471.653.445,03 €	
The Phone Company	1064954	300.646.915,31 €	708989	199.240.307,79 €	653215	173.638.184,86 €	2427158	673.525.407,96 €	
Wide World Importers	633311	229.130.226,71 €	716547	278.198.227,45 €	779501	304.204.397,75 €	2129359	811.532.851,91 €	
Total Geral	11851928	3.144.393.292,13 €	11270399	2.642.413.217,03 €	13778083	2.554.417.855,67 €	36900410	8.341.224.364,83 €	

Dimensão: Marca

Medida: Quantidade de Unidades Vendidas

Medida: Valor das Vendas

Factos

- **Factos:** são os dados a serem agrupados, contendo os valores de cada medida para cada combinação das dimensões existentes. O tamanho da tabela que contém os factos merece atenção especial do analista.
- **Tabelas de factos:**
 - Contêm os dados, numéricos ou qualitativos, que correspondem às medidas de negócio;
 - Contêm um conjunto de chaves estrangeiras que a ligam às tabelas de dimensão;
 - Chave primária composta pelas chaves estrangeiras para as tabelas de dimensão;
 - São constituídas por poucos atributos, comparativamente às tabelas de dimensão, e por muitas linhas;
 - Todas as “medições” numa tabela de factos têm de ter o mesmo grão;
 - Ocupam aproximadamente 90% do DW.
- **Tabelas de dimensão:**
 - Contêm a descrição dos factos medidos, ou seja, as características dos factos a que os utilizadores acedem na exploração do DW;
 - São constituídas por um grande número de colunas, comparativamente às tabelas de factos, e por, relativamente poucas linhas;
 - Possibilitam a análise da tabela de factos sob diferentes perspetivas, permitindo responder a diversas questões, tais como: **quem, quando, onde, porquê, como**, entre outras.
 - Os atributos das dimensões são utilizados habitualmente como etiquetas ou cabeçalhos nas *queries*;
 - Chaves primárias, em geral, artificiais (*surrogate keys*);
 - Contém muitas vezes uma hierarquia de dimensões (sem normalização). Por exemplo: Dimensão Tempo: dia < mês < ano;
 - Ocupam normalmente entre 10% a 15% do total dos dados armazenados.

Agregação vs Hierarquia

- **Agregação:**

- Mecanismo de cálculo que a partir de factos com um alto grau de atomicidade produz um resultado sumariado;
- Esta operação incrementa velocidade de exploração dos dados;
- É executada utilizando atributos de, pelo menos, uma dimensão e com suporte nas métricas da tabela de factos.
- Factos fortemente agregados são analisados mais rapidamente do que os factos muito detalhados.
- Uma tabela de factos com dados agregados ocupa menos espaço na BD do que outra com dados mais pormenorizados.

- **Hierarquia:**

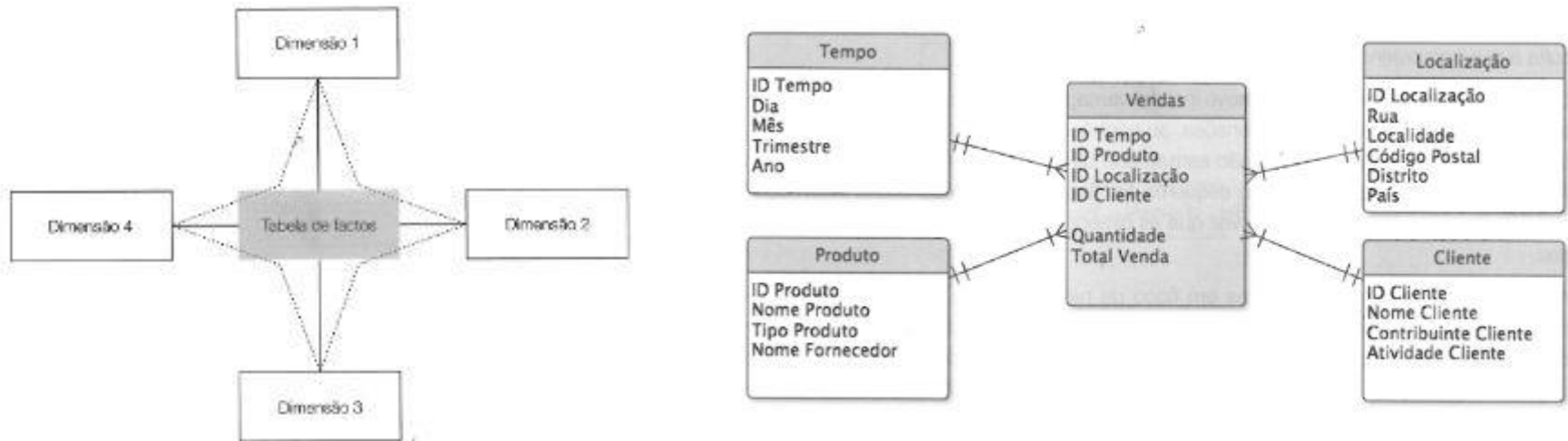
- Conjunto de associações entre os dados, em que alguns atributos pertencem a uma ou mais hierarquias;
- Cada atributo é pai ou filho de um outro atributo e esta relação pai-filho proporciona diferentes graus de sumarização dos dados;
- Exemplo: Rua, Cidade, Região, País, etc..

Esquemas do modelo dimensional

- **O modelo conceptual de um DW pode existir nas seguintes formas:**
 - Esquema em Estrela;
 - Esquema em Floco de Neve;
 - Esquema em Constelação de Factos.

Esquema em Estrela

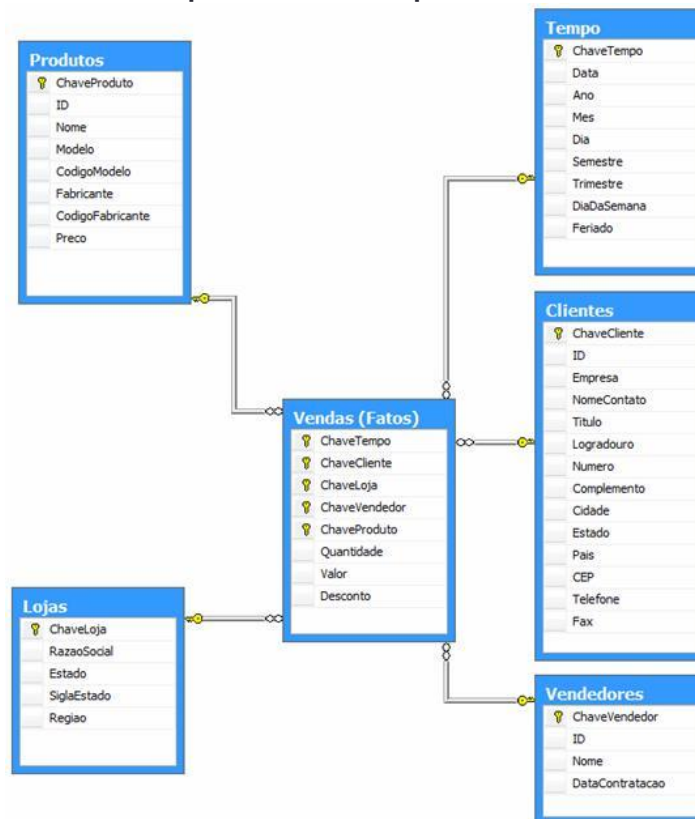
- Existem vários esquemas utilizados em ROLAP. Exemplos:
 - Esquema em Estrela:



Fonte: (Santos & Ramos, 2017)

Esquema em Estrela (cont.)

- É a forma mais elementar do modelo dimensional.
- Consiste numa tabela de factos (centro da estrela) que está interligada, num formato estelar, a um conjunto de tabelas de dimensão que contêm a descrição dos factos armazenados na tabela de factos.
- Cada dimensão é representada por uma única tabela.
- Exemplo:



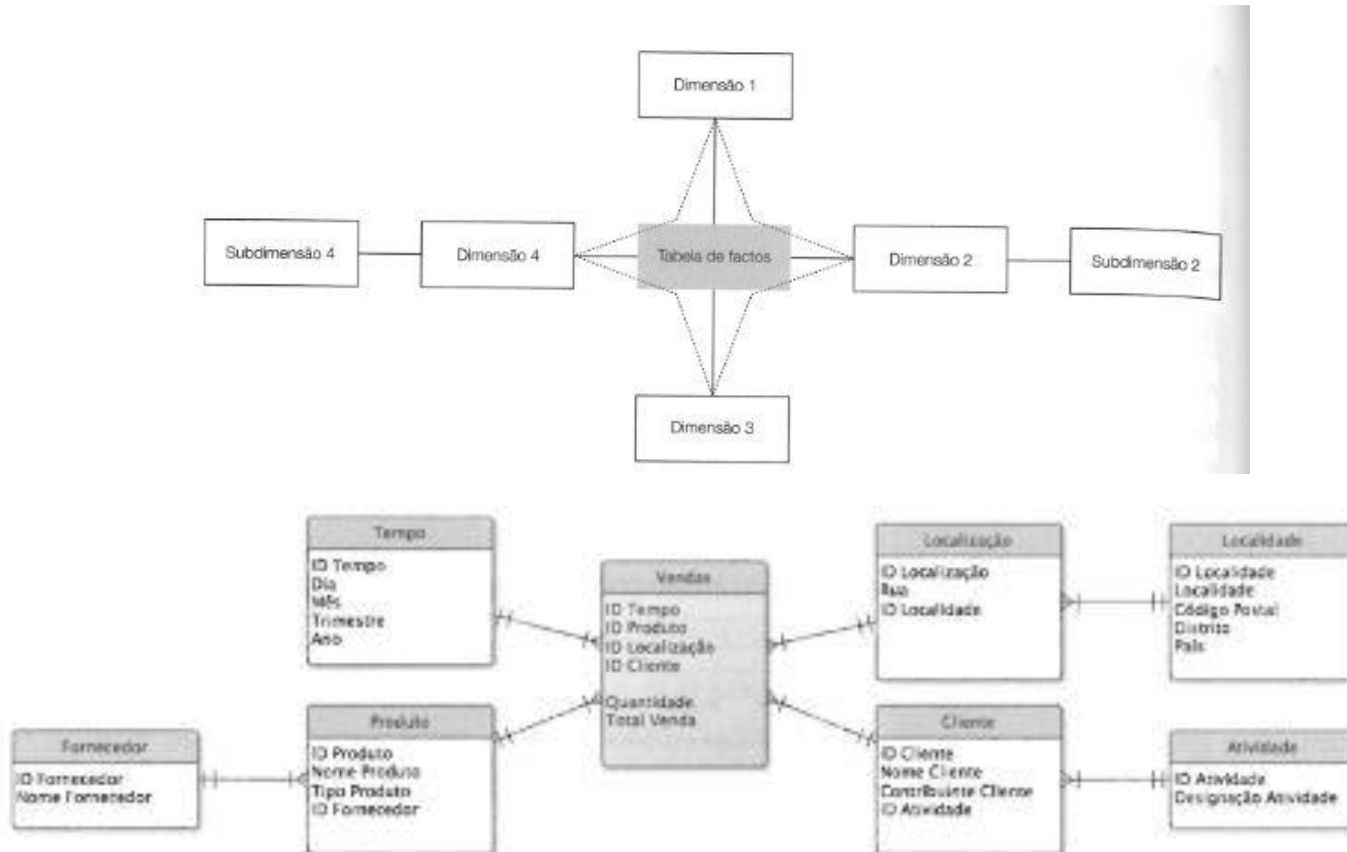
Aspectos a considerar:

- Maior eficiência dada pelo reduzido número de junções nas pesquisas e pelas chaves simples, e a facilidade de definir hierarquias.
- Tamanho considerável e a não normalização das tabelas de dimensões.

Fonte: <http://msdn.microsoft.com/pt-br/library/cc518031.aspx>

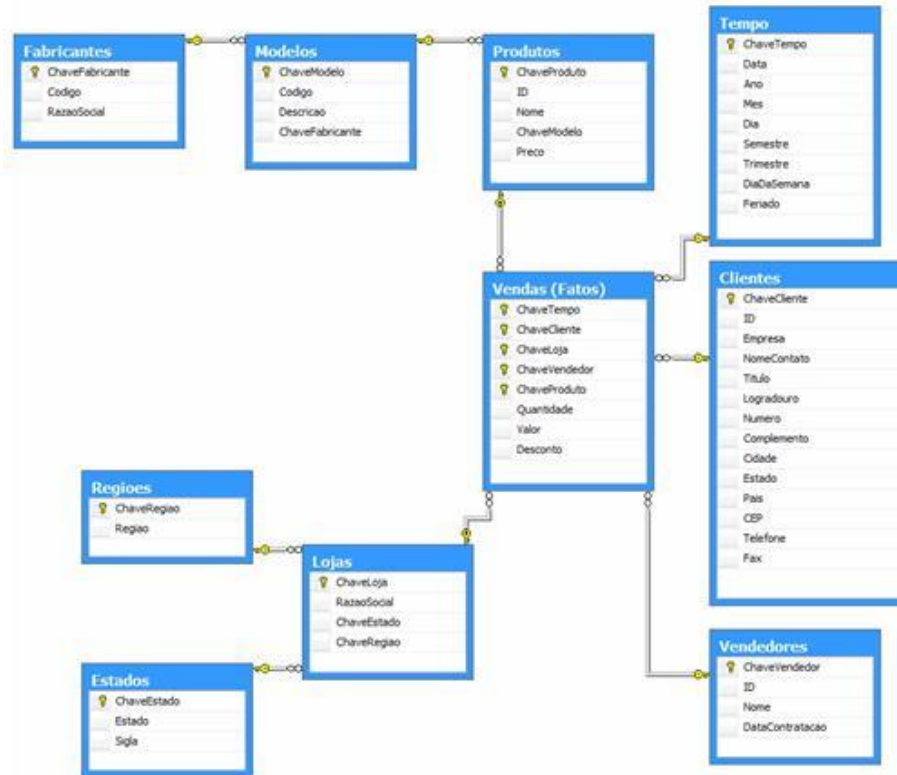
Esquema em Floco de Neve

- Esquema em Floco de Neve:



Esquema em Floco de Neve (cont.)

- Refinamento do esquema em estrela.
- Algumas dimensões estão normalizadas, estando os dados de cada dimensão separados em várias tabelas.
- Exemplo:

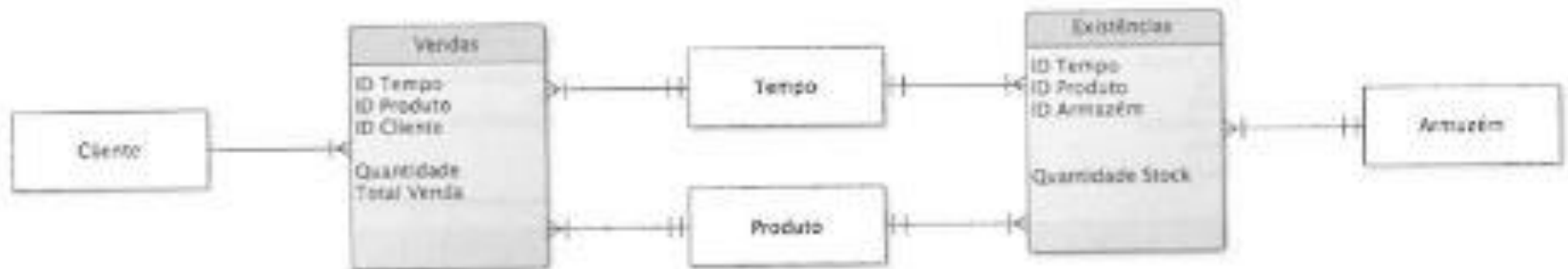
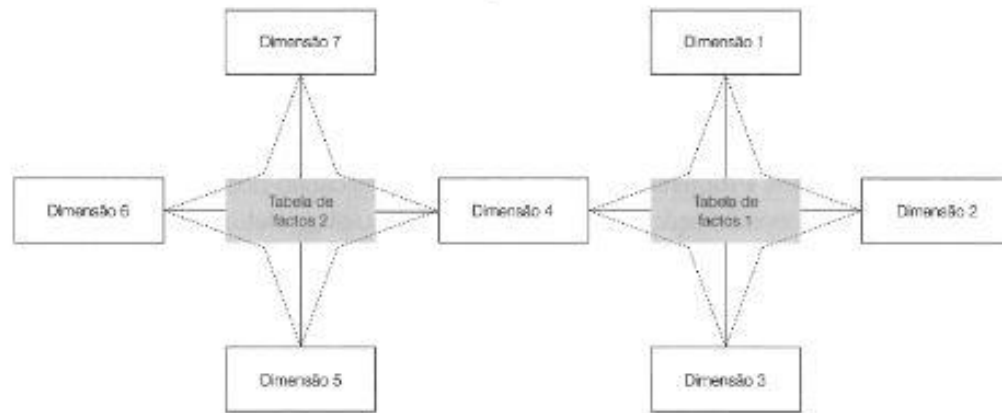


Aspectos a considerar:

- Redução nas redundâncias, o que resulta numa maior agilidade na manutenção.
- Modelo resulta num maior número de tabelas em junções, podendo existir uma diminuição do desempenho.

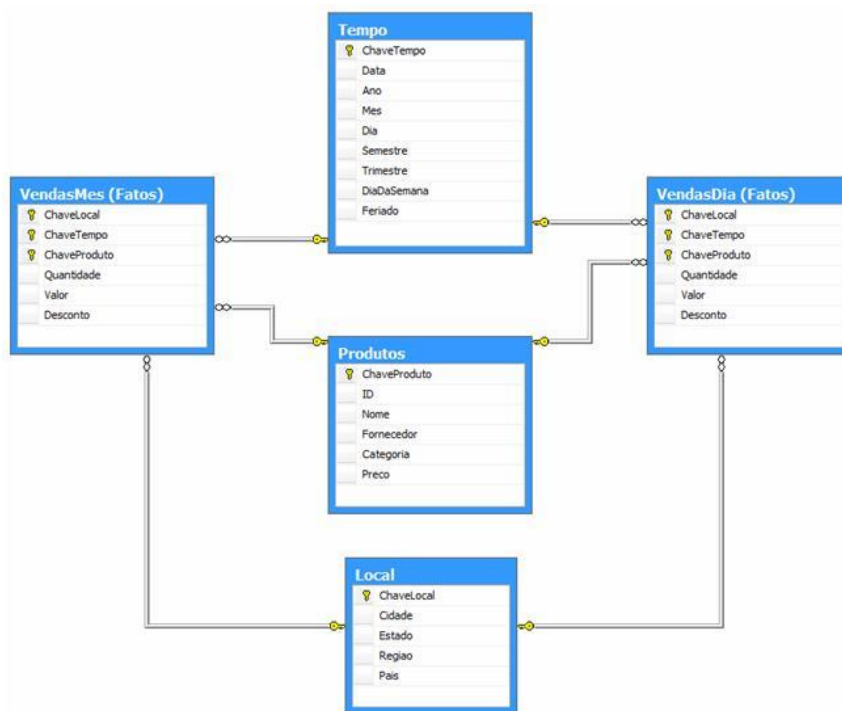
Esquema em Constelação de Factos

- Esquema em Constelação de Factos:



Esquema em Constelação de Factos (cont.)

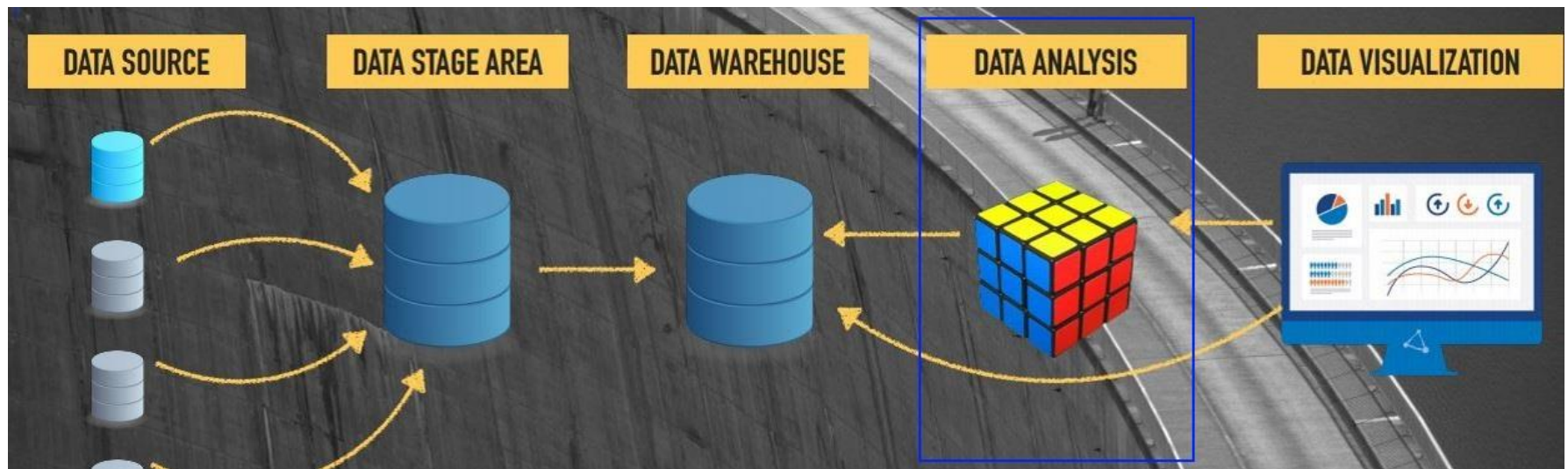
- Integra múltiplas tabelas de factos que partilham dimensões comuns.
- Exemplo:



Aspectos a considerar:

- Menos exigente quanto à sua manutenção, devido ao compartilhamento das tabelas de dimensão.

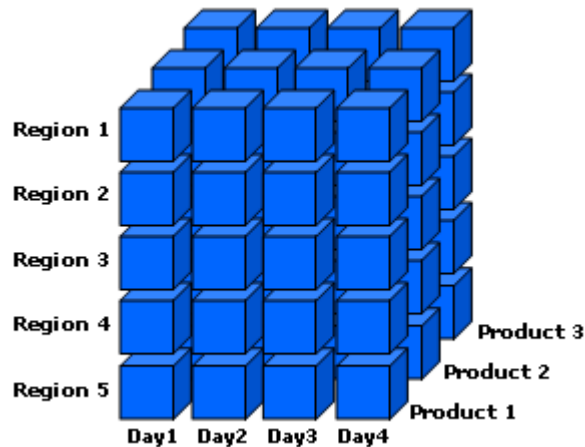
Data Analysis



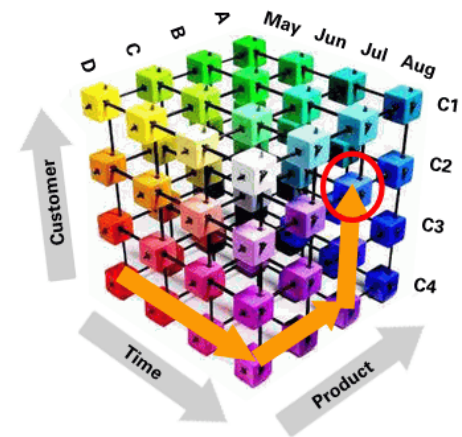
Fonte: (Piton, 2018)

OLAP

- Permite manipular e analisar um grande volume de dados em várias perspetivas.
- Representação lógica do modelo multidimensional em que as fases do cubo representam as dimensões e as células do cubo contêm valores das medidas (ou factos) para cada combinação de valores das dimensões.



Fonte: <http://msdn.microsoft.com/en-us/library/office/aa140038%28v=office.10%29.aspx>



Fonte: <http://www.oracle.com/technetwork/articles/sql/11g-dw-olap-100058.html>

Operações OLAP

- Técnicas de processamento simultâneo de grupos de dados:
 - *Roll-up (drill-up)*;
 - *Drill-down (roll-down)*;
 - *Slice* (permite executar uma seleção sobre uma dimensão do cubo);
 - *Dice*;
 - *Pivot (Rotate)* (permite reorientar o cubo ao representar a informação de 3D em vários planos de 2D).
 - *Drill Across* (permite navegar através de uma ou mais tabelas de factos).

Operações OLAP

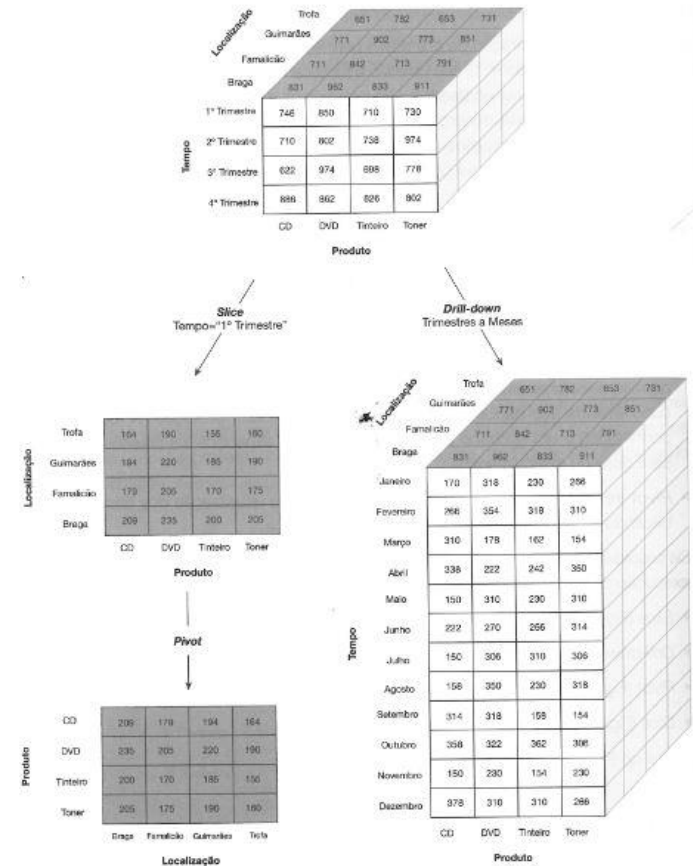
- Exemplo: Cubo em que um dos factos (total_venda) é analisado sob três dimensões: Tempo, Localização e Produto.
- Operações típicas OLAP incluem:
 - Drill-down:** Permite navegar de dados generalizados para dados mais detalhados. O objetivo é fornecer uma visão mais detalhada dos dados analisados. Ex: Lucro é inicialmente analisado por ano, depois por trimestre e finalmente por mês. Os níveis de drill-down dependem da granularidade dos dados na Data Warehouse.
 - Roll-up:** Operação oposta ao *drill-down*. Permite agregar os dados visualizados no cubo utilizando uma dada hierarquia. Ex: Analisar os dados inicialmente por mês, depois por trimestre e finalmente por ano.

Tempo	Localização				Produto			
	Trofa	Guimarães	Famalicão	Braga				
1º Trimestre	651	771	711	831	746	850	710	730
2º Trimestre	782	902	842	962	710	802	738	974
3º Trimestre	653	773	713	833	622	974	698	778
4º Trimestre	731	851	791	911	886	862	826	802

Fonte: (Santos & Ramos, 2017)

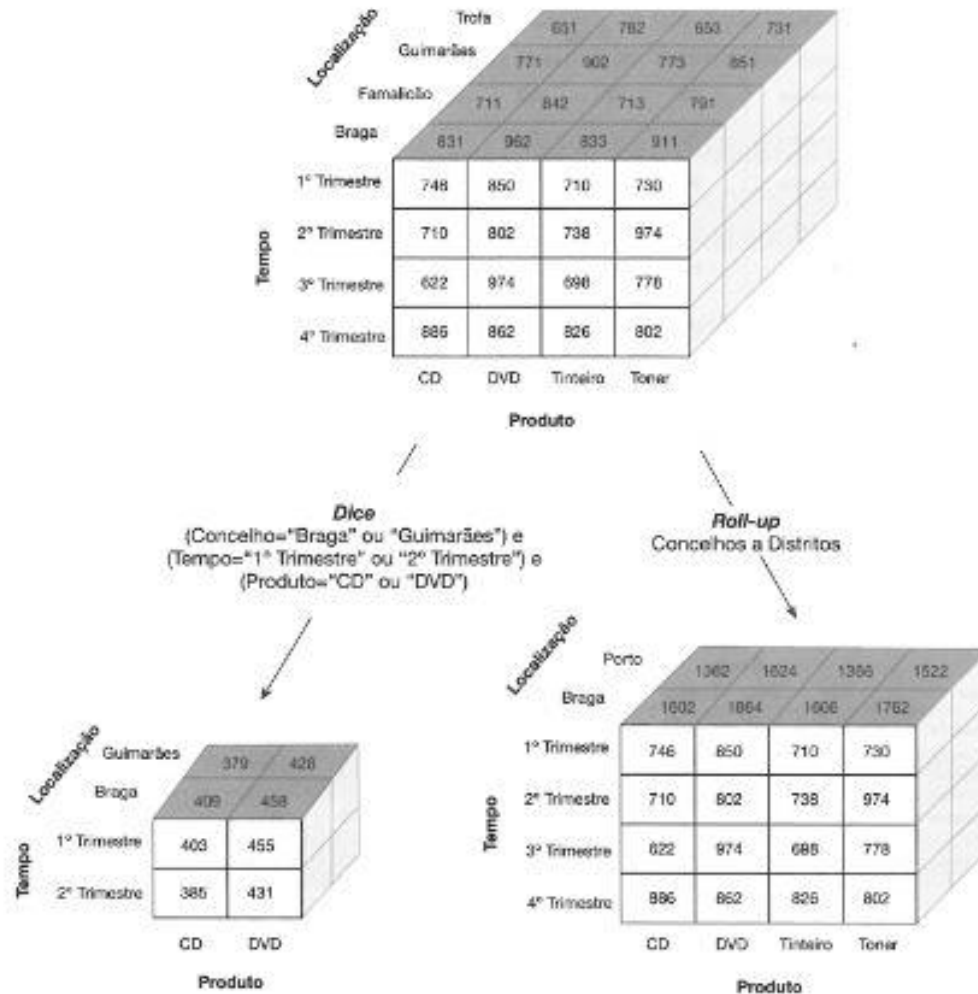
Operações OLAP (cont.)

- Slice and dice:** Permite restringir a informação a visualizar, utilizando o corte e a redução de um conjunto de dados. O corte (*slice*) permite selecionar um subconjunto de dados do cubo, em que uma das dimensões do cubo é limitada aos dados que obedecem determinada condição. A redução (*dice*) permite definir um sub-cubo para o qual podem ser especificados critérios de seleção para uma ou mais condições.
- Pivot (rotate):** A rotação permite rodar os eixos de visualização dos dados, disponibilizando uma representação alternativa dos mesmos.



Fonte: (Santos & Ramos, 2017)

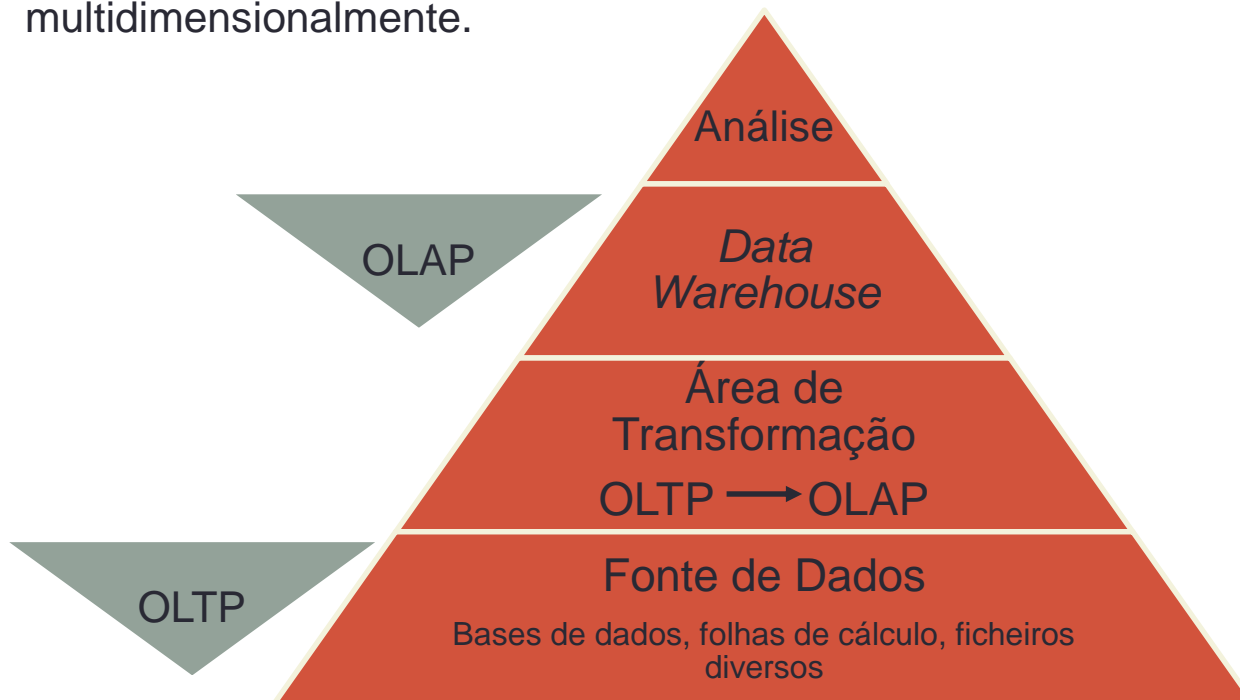
Operações OLAP (cont.)



Fonte: (Santos & Ramos, 2017)

OLAP vs OLTP

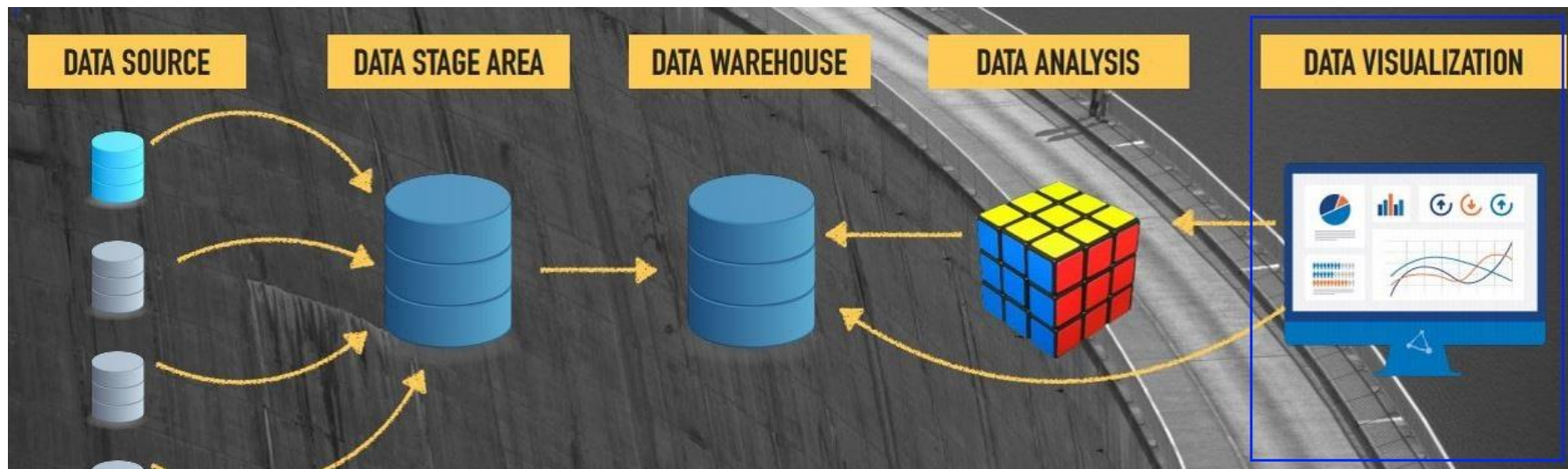
- Os sistemas OLAP contrapõem-se aos OLTP por serem **ferramentas de análise de dados**, onde o objeto das aplicações OLAP não é a forma como a informação é guardada num repositório de dados, mas o modo como essa mesma **informação pode ser eficientemente analisada**.
- Os processos OLAP abrangem grandes conjuntos de dados e, conseqüentemente, são tarefas demoradas que, raramente, são executados em tempo real.
- Para facilitar a visualização, os dados são tipicamente modelados multidimensionalmente.



Data Mining

- É o processo de descoberta de padrões em dados. Este processo pode ser automático ou semiautomático e os padrões descobertos devem ter significado, de forma a proporcionarem algum tipo de vantagem, normalmente de âmbito económico (Witten et al., 2011).
- É um conjunto de métodos e técnicas para explorar e analisar conjuntos de dados (normalmente de grande dimensão), de uma forma automática ou semiautomática, de forma a encontrar nestes dados, regras, associações ou tendências escondidas ou desconhecidas (Tuffery, 2011).

Data Visualization



Fonte: (Piton, 2018)

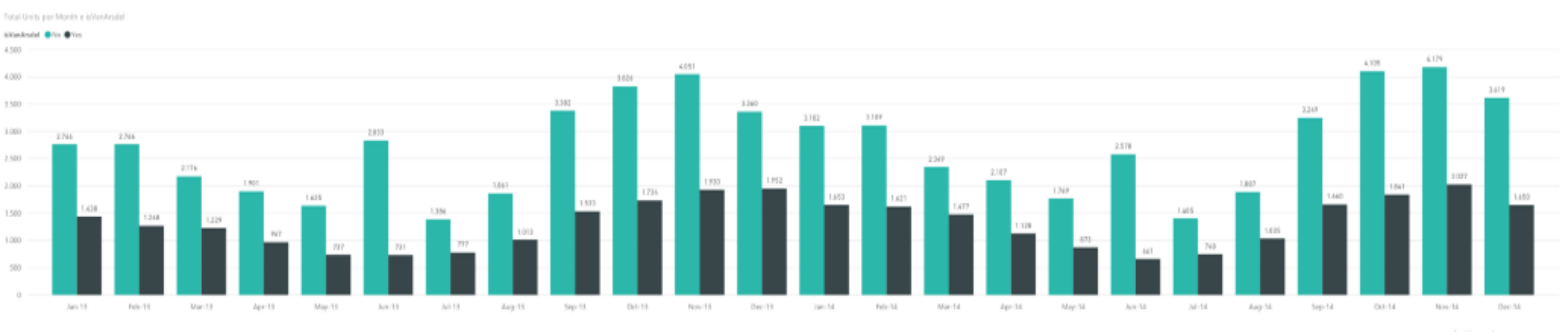
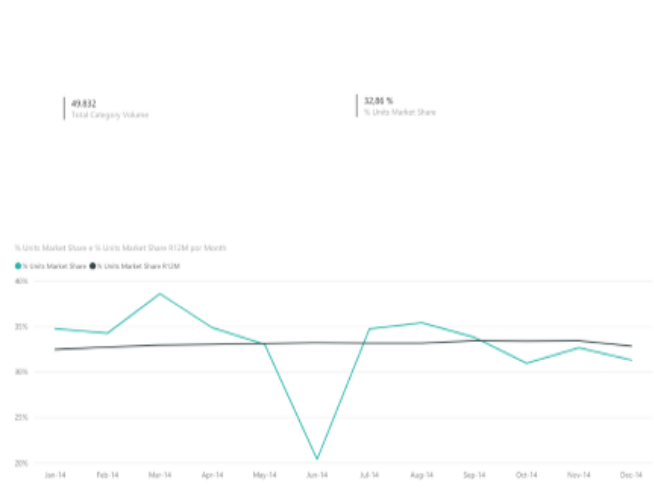
Painel de Controlo (Dashboards)



Fonte: Microsoft Power BI

Relatórios (Reports)

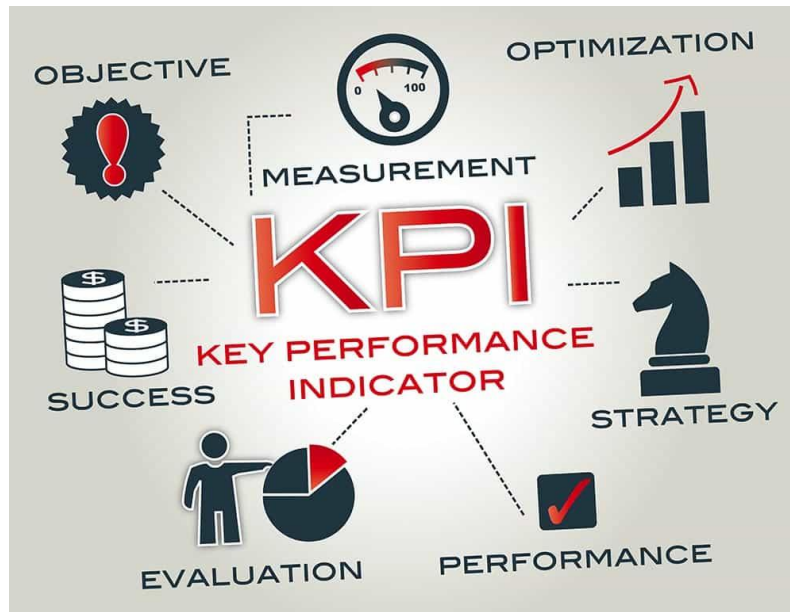
VanArsdel - Market Share



Fonte: Microsoft Power BI

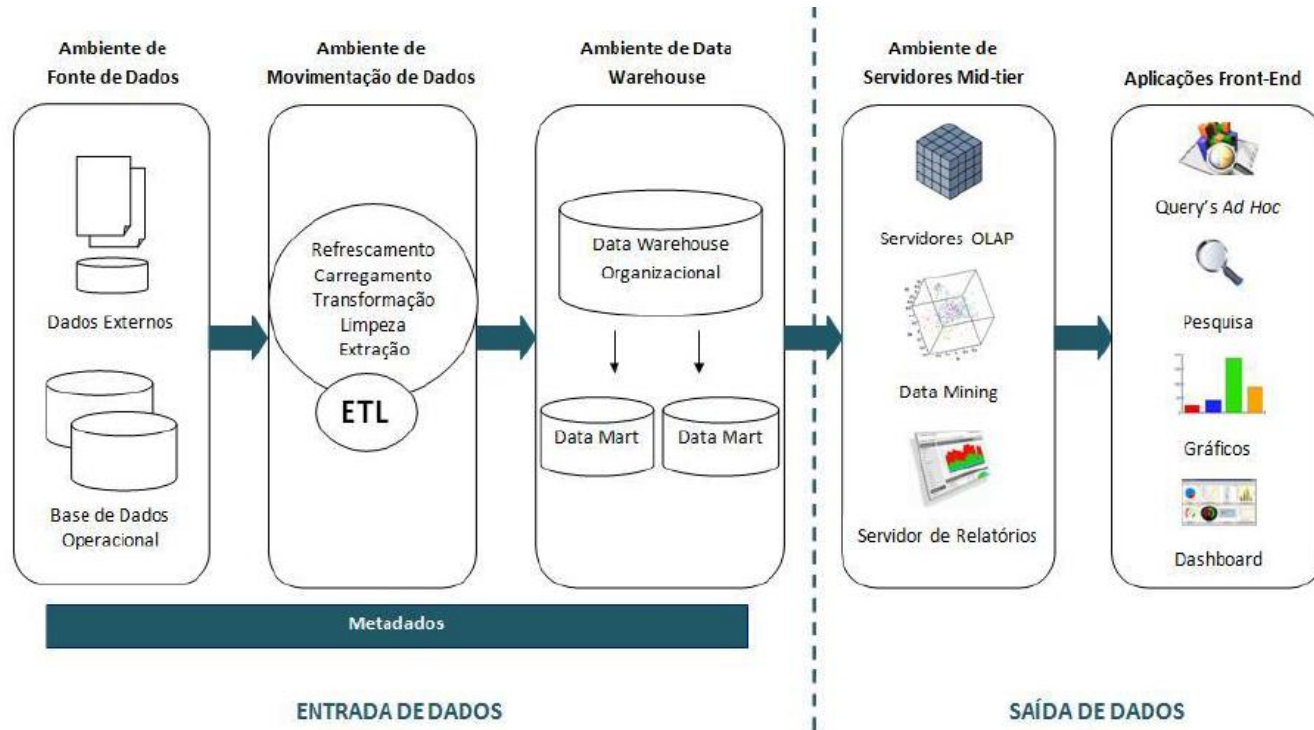
Key Performance Indicators - KPI

- E-commerce: quantidades em stock, visitantes no website, vendas realizadas e as diferenças nos resultados dentro de um determinado período.
- Desempenho de um departamento de marketing: acompanhar o desenvolvimento das vendas de um produto, pesquisas de satisfação do cliente, fontes de origem que tragam clientes para uma loja, performance do website, comportamento do consumidor no website ou blog, etc.



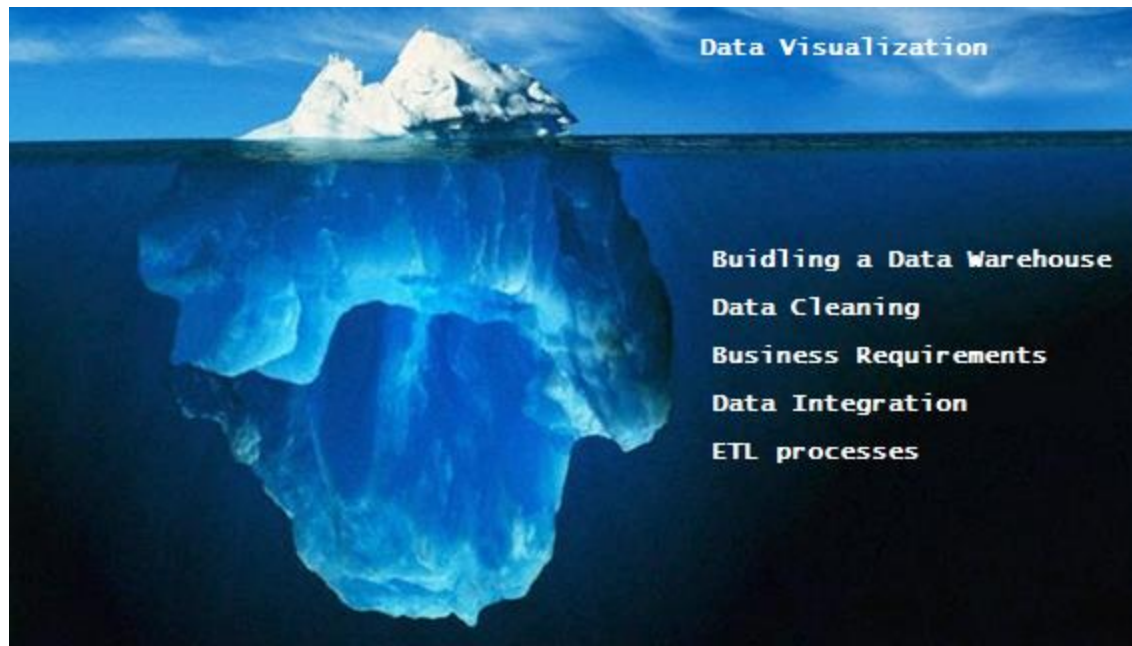
Product Sales KPIs						
Sales Date	Subcategory	Product	Quantity	Sales	Linear KPI	Stoplight KPI
1/5/2015	Accessories	Carrying Case	68	\$16,996.60		
		Mini Battery Charger	44	\$1,056.00		
	Digital	Slim Digital	112	\$18,752.40		
			44	\$8,357.80		
Total			156	\$26,410.40		
1/6/2015	Accessories	Telephoto Conversion Lens	18	\$1,380.00		
		Tripod	18	\$1,350.00		
		USB Cable	26	\$780.00		
Total			62	\$3,510.00		
1/7/2015	Digital	Compact Digital	84	\$10,836.00		
			84	\$10,836.00		
	Digital SLR	SLR Camera	88	\$26,576.00		
Total			172	\$37,412.00		
1/8/2015	Accessories	Budget Movie-Maker	9	\$3,798.00		
			9	\$3,798.00		
	Digital	Consumer Digital	17	\$2,550.00		
Total			17	\$2,550.00		
1/9/2015	Camcorders	Business Videographer	13	\$10,400.00		
			13	\$10,400.00		
	Digital SLR	SLR Camera 35mm	34	\$18,530.00		
Total			34	\$18,530.00		
1/10/2015	Camcorders	Social Videographer	60	\$3,000.00		
Total			60	\$3,000.00		
1/11/2015	Accessories	Lens Adapter	17	\$1,147.50		
			17	\$1,147.50		
	Digital	Advanced Digital	39	\$7,234.50		
Total			56	\$8,382.00		
Total			579	\$113,992.40		

Arquitetura de BI



Fonte: (Amarante, 2014)

Business Intelligence



Limitações das Tecnologias Tradicionais de Data Warehouse

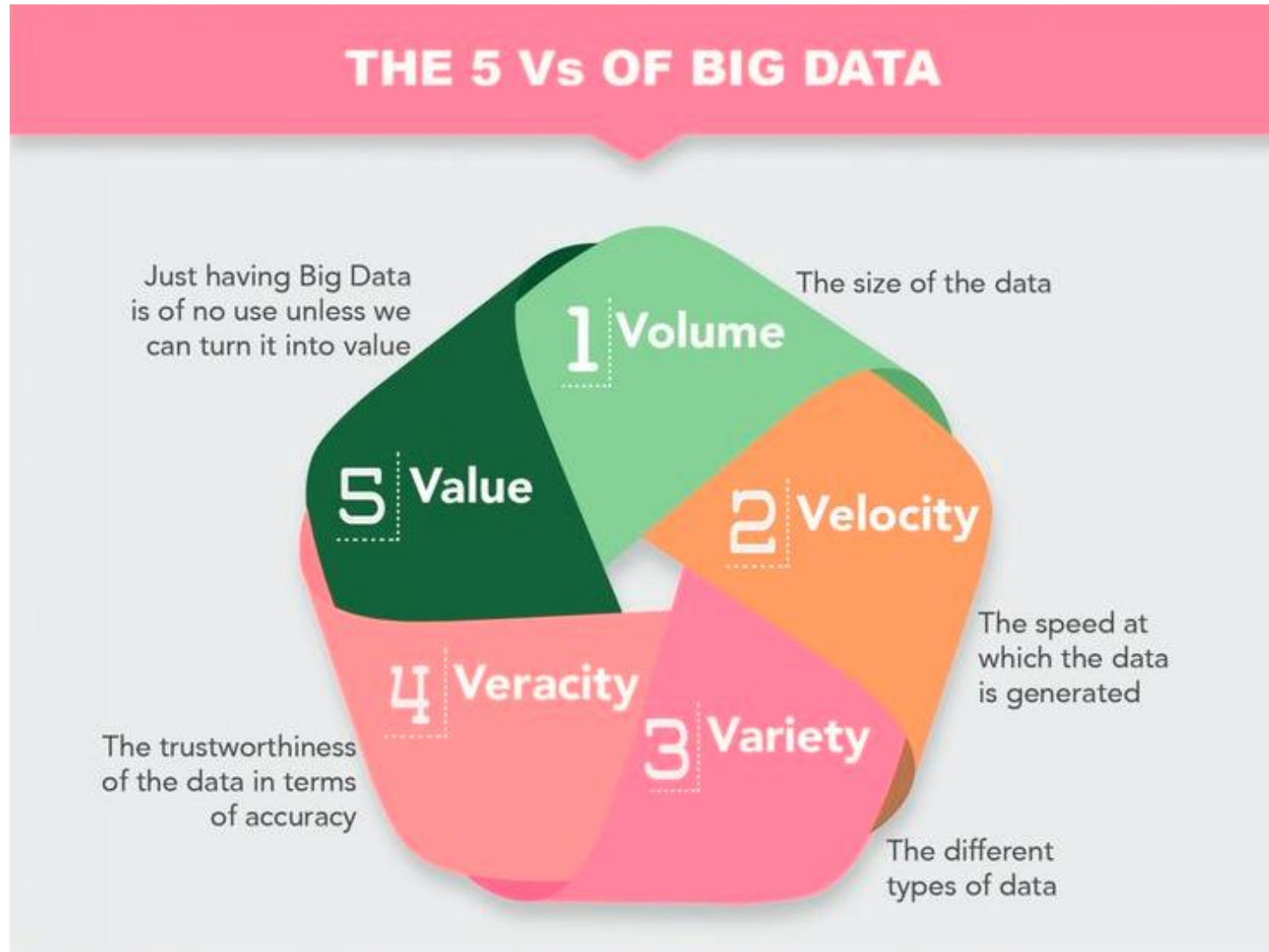
- Não são flexíveis e a sua expansão é dispendiosa (escalabilidade vertical);
- Tem limitações no processamento de grandes quantidades de dados em tempo real;
- Requerem a existência de uma estrutura (problema com dados semiestruturados e não estruturados);
- Requerem um planeamento inicial aprofundado (formatos dos dados, organização e relacionamentos).

Big Data

- O Big Data é a análise massiva de dados. Uma quantidade enorme de dados que os aplicativos de software de processamento de dados, que eram usados tradicionalmente, não são capazes de capturar e tratar em um tempo razoável.



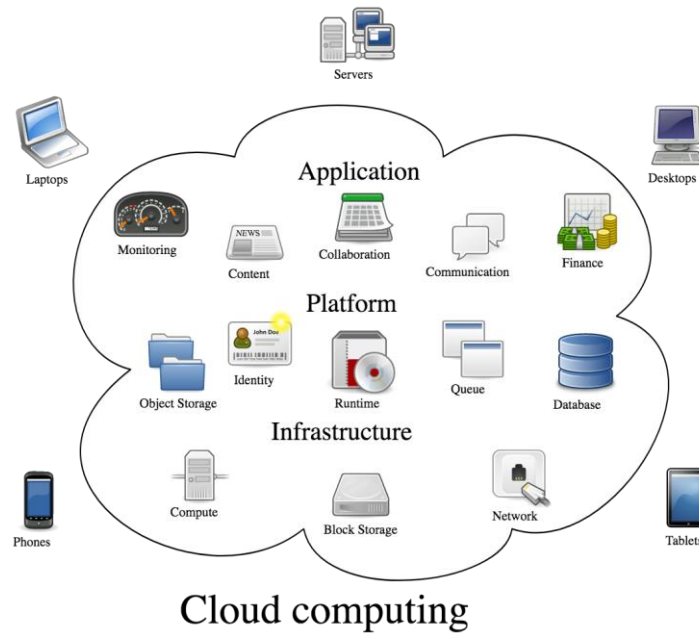
As V do Big Data



Big Data



Serviços Cloud



Microsoft
Azure



Google
Cloud Platform

Vídeos

- [Real-Time Hotel Business Intelligence](#)
- [What is Power BI?](#)
- [Travel Click](#)

Referências

- Amarante, R.P. (2014). *Business Intelligence no contexto de uma PME*. Relatório de Estágio do Mestrado em Gestão, Universidade de Coimbra.
- Caldeira, C. P. (2012). *Data Warehousing: Conceitos e Modelos*, 2.ª edição, Lisboa, Edições Sílabo.
- Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. (1996). *Advances in knowledge discovery and data mining*. *aaai.org*.
- Gartner. (2013). *Gartner IT Glossary*. Retrieved from <http://www.gartner.com/it-glossary/>
- Gartner. (2017). *Gartner Methodology*. Retrieved from <https://www.softwareadvice.com/bi/#top-products>
- Inmon, W. H. (2005). *Building the Data Warehouse*, Third Edition, Wiley Computer Publishing.
- Kimball, R., Ross, M. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*, Third Edition, Wiley.
- McGuire K. A. (2017). *The Analytic Hospitality Executive – Implementing Data Analytics in Hotels and Casinos*. John Wiley & Sons.
- Piton, R. (2018). *Business Intelligence*. Retrieved from <https://www.youtube.com/watch?v=MEzMDtz2Y1U>
- Santos, M., Ramos, I. (2017). *Business Intelligence - Da Informação ao Conhecimento*, 3.ª edição atualizada. FCA Editora de Informática.
- Tuffery, S. (2011). *Data Mining and Statistics for Decision-Making*. Wiley.
- Turban, E., Sharda, R., Delen, D., King, D. (2010). *Business Intelligence: A Managerial Approach*, 2.º Edition, Prentice Hall.
- Witten, I. H., Frank, E., Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann.