

A decorative graphic on the left side of the slide consisting of a grid of squares in various shades of blue and purple, arranged in a stepped pattern.

Fundamentos de *Business Intelligence* (3º Encontro)

Prof. MSc. Fernando Siqueira

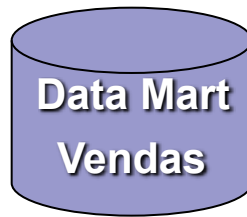


Data Mart

Data Mart

Principais Características

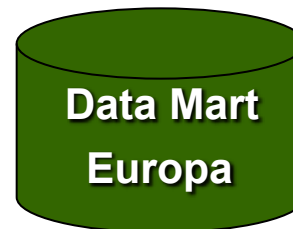
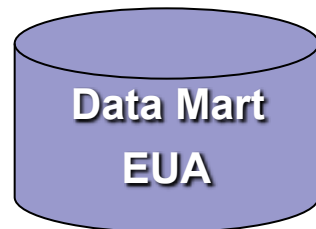
- Um DW de pequena capacidade, projetado para atender uma área de negócio específica
- Dados mantidos no DW são separados por assunto, em subconjuntos, de acordo com:
 - Estrutura interna da empresa
 - O processo de tomada de decisão



Data Mart

Principais Características

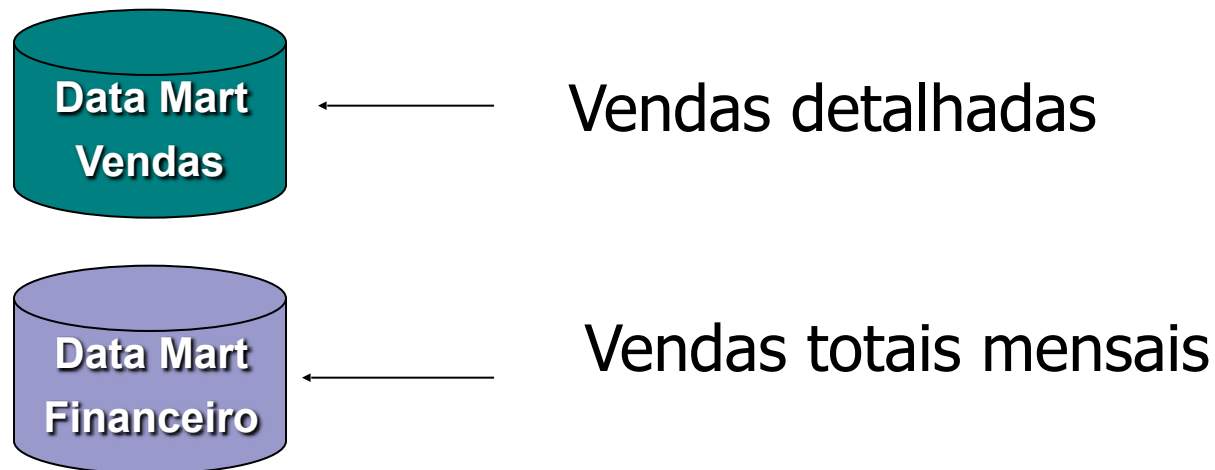
- Pode desempenhar o papel de um DW departamental, regional, de acordo com:
 - Estrutura interna da empresa
 - O processo de tomada de decisão



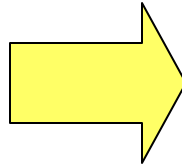
Data Mart

Principais Características

- Dados podem ser repetidos em dois ou mais *Data Marts*
- Os mesmos dados podem estar representados com granularidade diferente



Processo de Desenvolvimento do BI



Modelagem Dimensional

Construção dos DM's / DW

Extração e Transformação
(fontes internas e externas)

Desenvolvimento das Aplicações



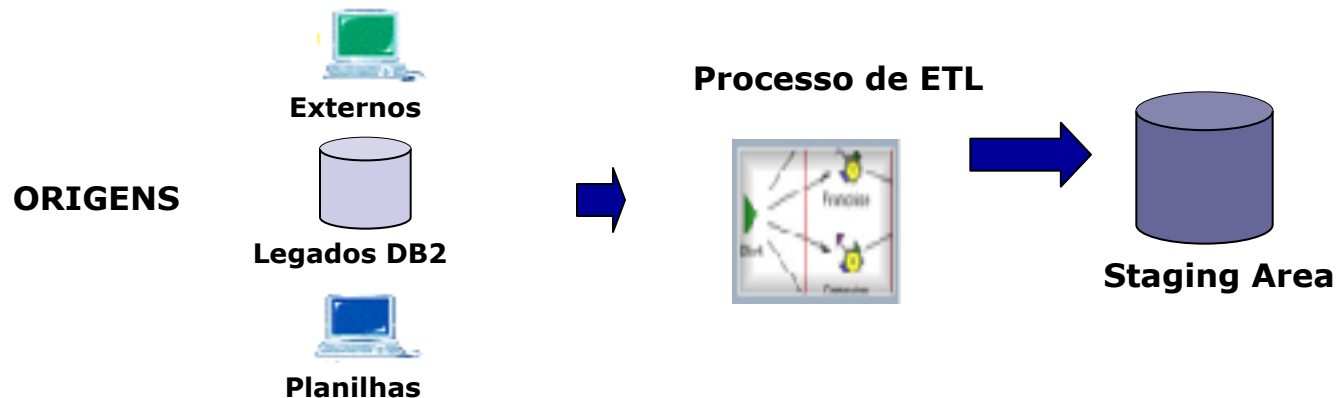
ODS, SA, DM/DW



Construção SA, ODS, DM/DW

- Staging Area
- Operational Data Store
- Data Mart/Data Warehouse

SA – Staging Area



- Conhecida também com “camada integração e transformação” - Ilmon.
- Características:
 - Padronização
 - Tratamento dos dados
 - Transformações

ODS – *Operational Data Store*



- Visão integrada do mundo Operacional.
- Características:
 - Orientada a assunto
 - Coletivamente Integrada
 - Volátil
 - Valores correntes
 - Apenas dados comparativos detalhados



Construção SA, ODS, DM/DW

- Projeto de Banco de Dados BI
 - O objetivo desta atividade é projetar os modelos de dados do projeto.
 - No contexto do BI são utilizados os seguintes tipos de Modelos de Dados:
 - Modelo de Dados do ODS (Operational Data Store) que contém as tabelas utilizadas no processo de recuperação de informações e na carga dos data marts.
 - Modelo de Dados Dimensional que contém as tabelas de dimensão e fato dos data marts.
 - Modelo de Dados da Stage Area que contém as tabelas da stage área usadas no processo de ETL.
 - Modelo de Dados do DW que contém as tabelas do Data Warehouse.
- Etapas:
 - Projeto de Banco de Dados
 - Definir Volumetria



Construção SA, ODS, DM/DW

- Etapas

- Projeto de Banco de Dados

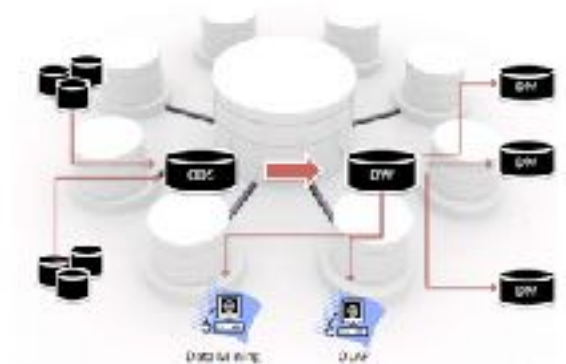
- A finalidade dos bancos de dados dimensionais é fornecer subsídio para realização de análises. Para tanto, sua arquitetura e até mesmo a terminologia empregada são distintas das utilizadas para bancos de dados transacionais.
 - Quando o modelo de dados começa a ser definido, elementos básicos de representação precisam ter sido estabelecidos, de modo a criar-se um padrão de modelagem. Em um modelo dimensional teremos dimensões e fatos representados em tabelas.

- Definir Volumetria

- A análise de volumetria auxilia nas decisões de: design, arquitetura, modelagem, armazenamento, proteção de dados, planos de backup, latência, tráfego de rede, etc.
 - A estimativa de volumetria deve ser aplicada aos modelos de Stage, Dimensional e ODS, deve contemplar o tamanho inicial da base de dados e a previsão de crescimento nos anos seguintes.

Projeto de *Data Warehouse*

Dicas





Projeto de Data Warehouse

Dicas de Revisão

- Granularidade
- Dimensão Tempo
- *Slow Changing Dimension*
- Conformidade dos Dados
- Chaves Substitutas (SKs)
- Modelo Dimensional
- Qualidade de Dados



Granularidade

- O grão é uma das mais importantes definições na modelagem de dados do DW.
- Trata-se do menor nível da informação e é definido de acordo com as necessidades levantadas no início do projeto.
- Quanto maior for a granularidade, menor será o detalhe (ou maior será a sumarização).
- Quanto menor for a granularidade, maior será o detalhamento (ou menor será a sumarização)



Dimensão Tempo

- A Dimensão temporal (ou dimensão data para alguns) é a mais importante perspectiva para a análise dos dados.
- Sem ela é difícil fazer a averiguação descritiva dos fatos.
- É imprescindível para a análise das ocorrências durante o tempo.
- É importante que seja avaliado a conformidade, analisando os níveis de detalhes exigidos para essa dimensão, mantendo a coerência de definição entre os diversos Data Marts.



Slow Changing Dimension

- O SCD tem papel fundamental para a visualização de informações históricas e armazenamento de versões dos dados de acordo com variações no tempo.
- Retrata as Dimensões que sofrem atualizações em seus campos e os classifica pelo tipo de mudança existente em cada uma delas.
- Exemplo:
 - Em modelo dimensional você tem a possibilidade de avaliar as vendas de cada filial e de cada vendedor.
 - Se um vendedor é transferido para uma outra filial, deve-se ter o tratamento dessa mudança para não ter uma avaliação incorreta do montante de uma venda para cada filial.



Conformidade dos Dados

- A conformidade propicia o compartilhamento semântico de uma mesma Dimensão entre vários *Data Marts*, mantendo a consistência das informações geradas pelas diferentes análises.
- Centraliza as alterações, e permite uma otimização no uso dos recursos computacionais.
- A conformidade de Dimensões é imprescindível para um projeto evolutivo de DW exitoso.



Chaves Substitutas (*Surrogate Keys*)

- Chaves substitutas ou artificiais (surrogate keys), são chaves desprovidas de inteligência de negócio.
- Utilizadas como artifício para o versionamento dos dados e junção entre tabelas Fatos e de Dimensões.



Modelo Dimensional

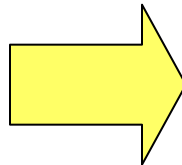
- Revise também o modelo dimensional resultante na busca de erros que só são evidentes após a implementação do DW.
- Verifique a existência de relacionamentos desnecessários e duplicidade de informações sem propósito e que oneram a performance das análises e consultas.



Qualidade de Dados

- A qualidade dos dados é um dos aspectos mais importante no processo de construção do DW.
- Devemos analisar os dados armazenados a procura de inconsistências e deformidades.
- Caso seja detectado alguma anormalidade nos dados, o processo de ETL deverá ser verificado e revisado para garantir que as informações armazenadas possuam valor para a utilização na tomada de decisão empresarial.

Processo de Desenvolvimento do BI



Modelagem Dimensional

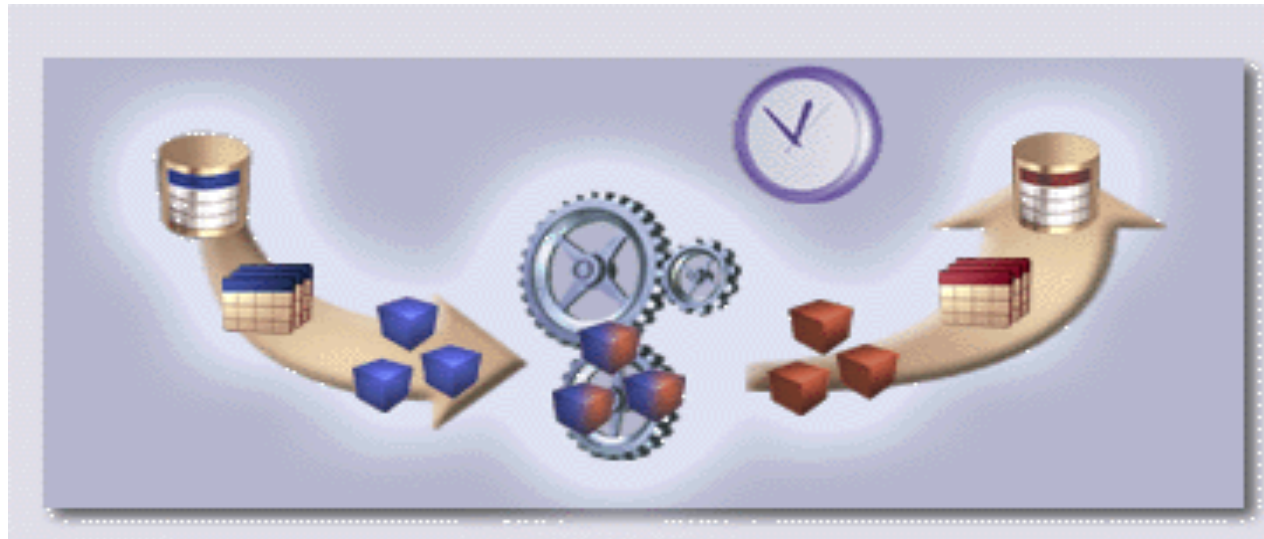
Construção dos DM's / DW

Extração e Transformação
(fontes internas e externas)

Desenvolvimento das Aplicações

ETL – *Extract, Transform and Load*

- Conjunto de técnicas para **mover os dados** de **diversas fontes**, de forma a promover a sua **consolidação** nas **aplicações a que se destinam**



ETL – *Extract, Transform and Load*

Processo de ETL



- É o processo de captura das fontes de dados a serem utilizadas em um ambiente BI, sua transformação, padronização e posterior carga no DW (ou DM ou ODS).
- Pode ser implementado via ferramenta específica ou através de linguagem de programação de banco de dados.
- Através de ferramentas são muitas as suas vantagens: metadados, facilidade de manutenção, gráfica, processo em uma só ferramenta, ...

ETL – *Extract, Transform and Load*

- Transformação

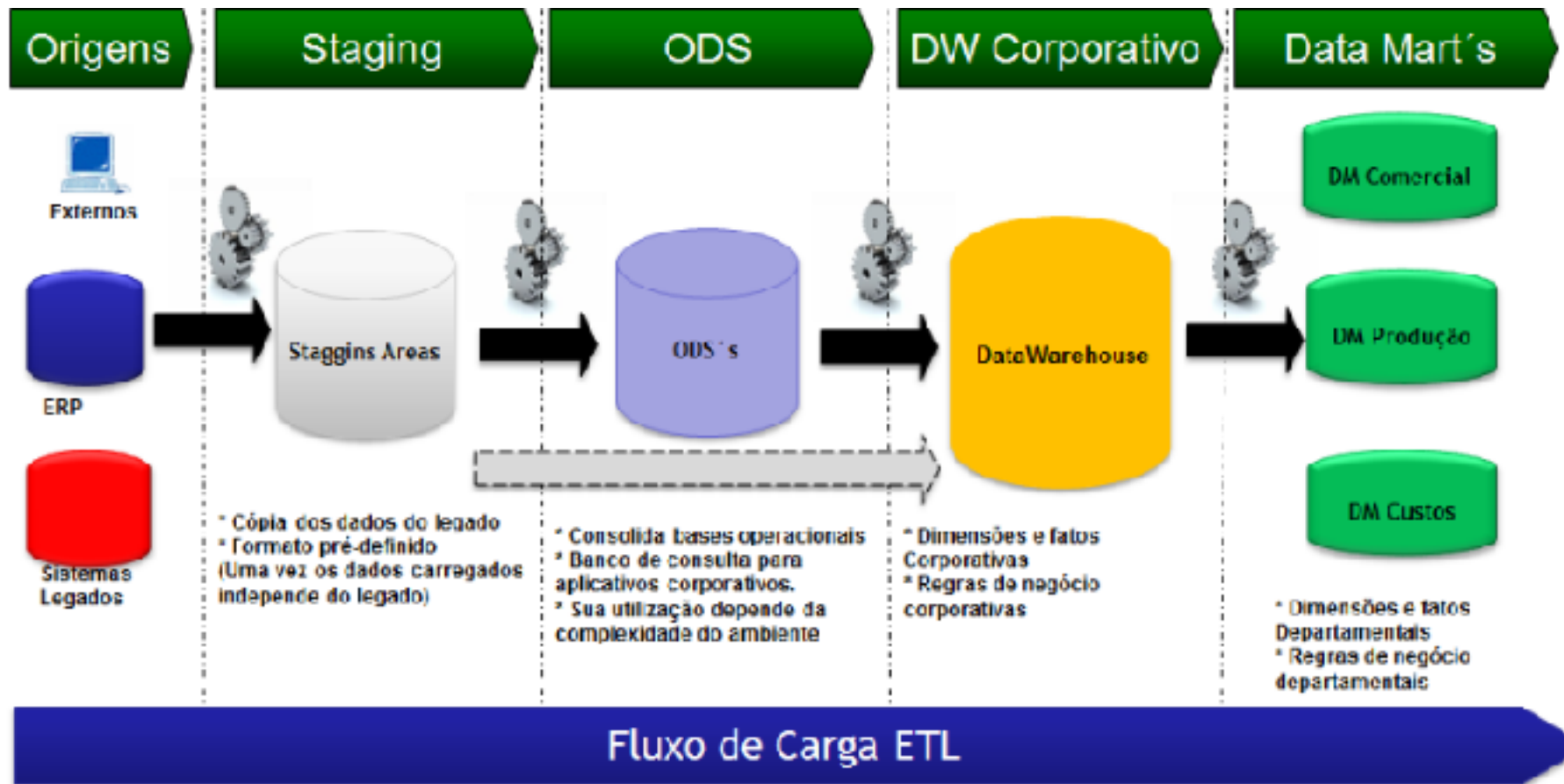
- Após os dados serem extraídos, esses podem sofrer algumas transformações como:
 - Padronização das unidades, domínios, tipos de dados
 - “limpeza” dos dados
 - Combinação das diversas fontes de dados
 - Criação de *surrogate keys* para cada registro dimensional
- Onde devem ser executadas as transformações:
 - Onde existe capacidade de processamento disponível
 - Nas plataformas que a organização está habituada a operar



ETL – *Extract, Transform and Load*

- Carga/População
 - Após o processo de transformação, os dados já estão aptos a serem carregados no DW
 - A carga consiste em povoar as tabelas do modelo dimensional (tabelas dimensões e fatos) criadas no repositório DW

ETL – *Extract, Transform and Load*



ETL – *Extract, Transform and Load*

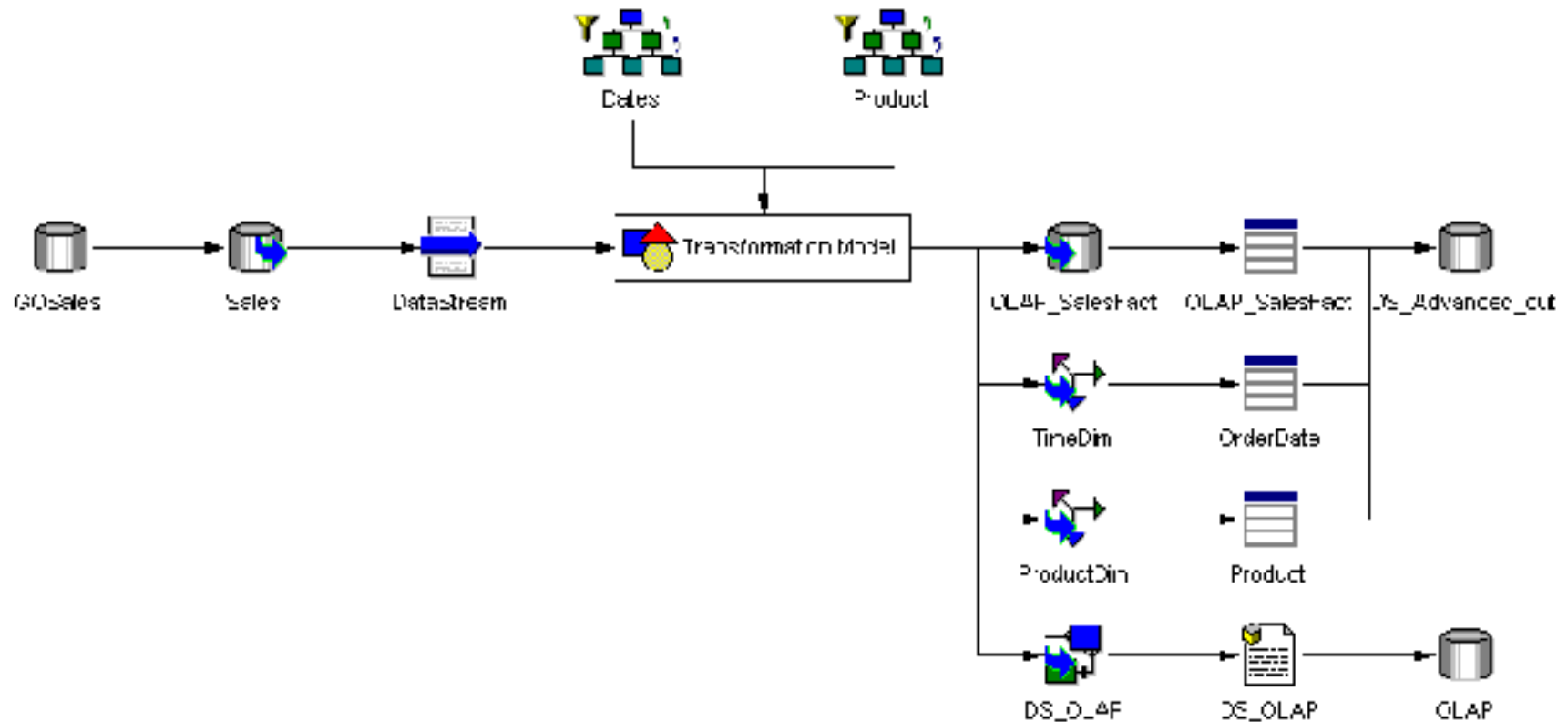
Table Name: *CUSTOMER*

Section Notes									
Mapping	Source Table	Source Column	Source Type	Transformation Logic	Target Table	Target Column	Target Type	Example	Comments
1.15	PURCHASER	BILL_CITY	Varchar(45)	Direct Map for target address.ADDRESS_TYPE = 'BILL'	ADDRESS_DIM	CITY	Varchar(25)	Farmersville	Join the CUSTOMER_DIM table to the ADDRESS_DIM table on CUSTOMER_DIM.BILLING_ADDRESS_ID = ADDRESS_DIM.ADDRESS_ID and pull ADDRESS_DIM.CITY
1.16	PURCHASER	BILL_STATE	Varchar(25)	Lookup state abbreviation and load for target address.ADDRESS_TYPE = 'BILL'	REGION_DIM	REGION_NAME	Varchar(5)	DE	Join the CUSTOMER_DIM table to the ADDRESS_DIM table on CUSTOMER_DIM.BILLING_ADDRESS_ID = ADDRESS_DIM.ADDRESS_ID and the ADDRESS_DIM table to the REGION_DIM table on ADDRESS_DIM.REGION_ID = REGION_DIM.REGION_ID and pull REGION_DIM.POSTAL_ABBRE
1.17	PURCHASER	BILL_POSTALCODE	Varchar(12)	Direct Map for target address.ADDRESS_TYPE = 'BILL'	ADDRESS_DIM	POSTAL_CODE	Varchar(12)	27745	Join the CUSTOMER_DIM table to the ADDRESS_DIM table on CUSTOMER_DIM.BILLING_ADDRESS_ID = ADDRESS_DIM.ADDRESS_ID and pull ADDRESS_DIM.POSTALCODE

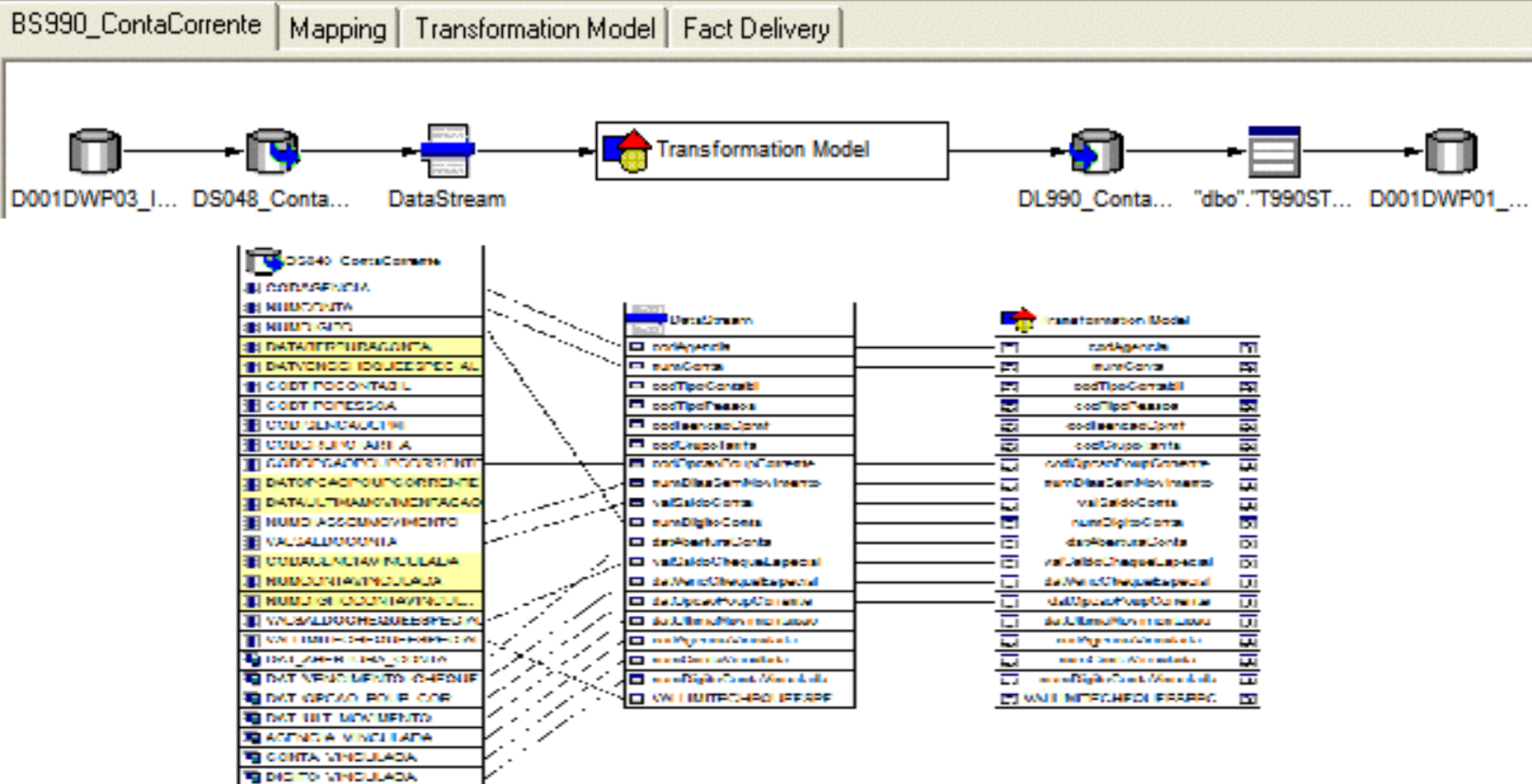
ETL – *Extract, Transform and Load*

[illegible]

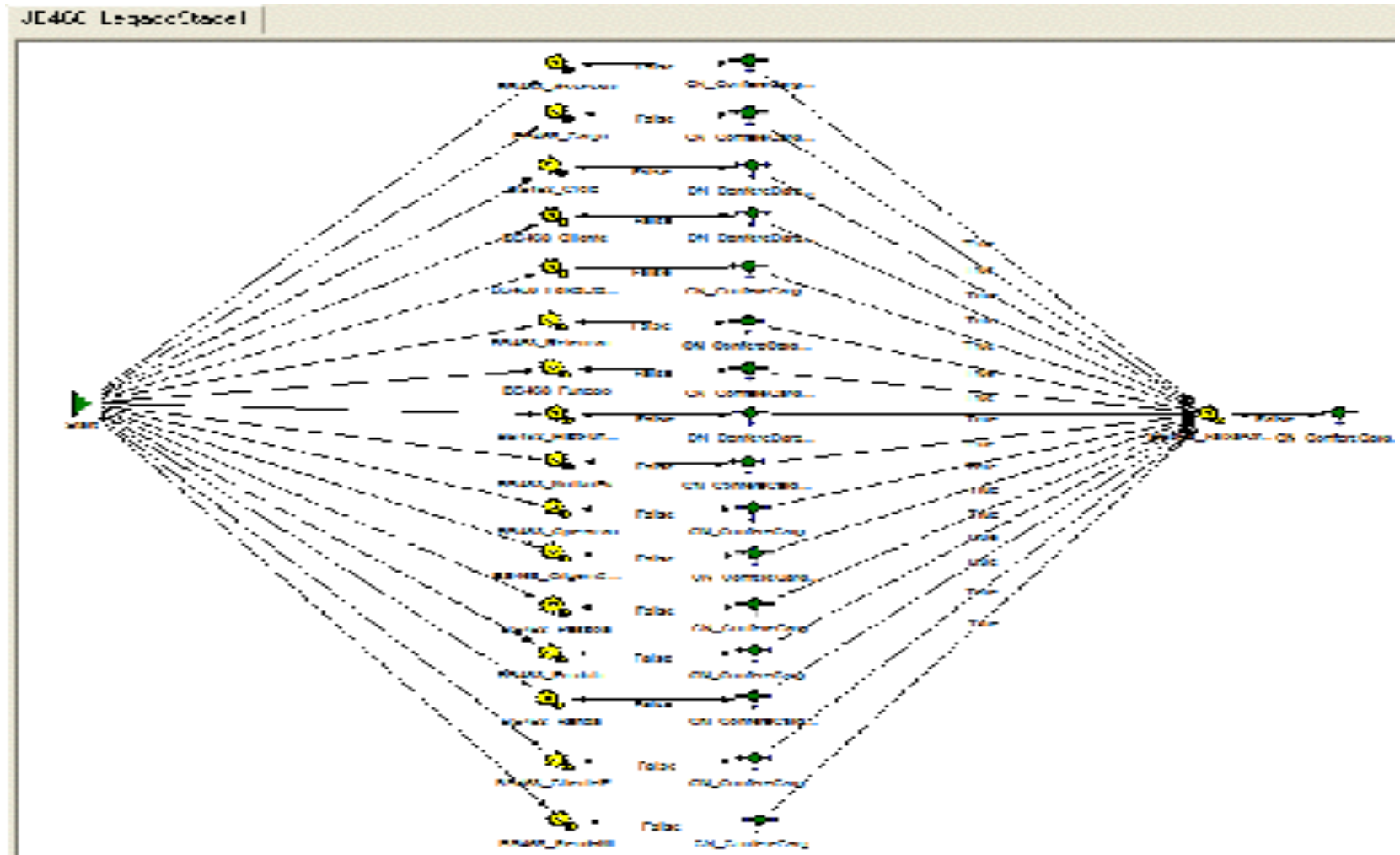
ETL – *Extract, Transform and Load*



ETL – *Extract, Transform and Load*



ETL – *Extract, Transform and Load*

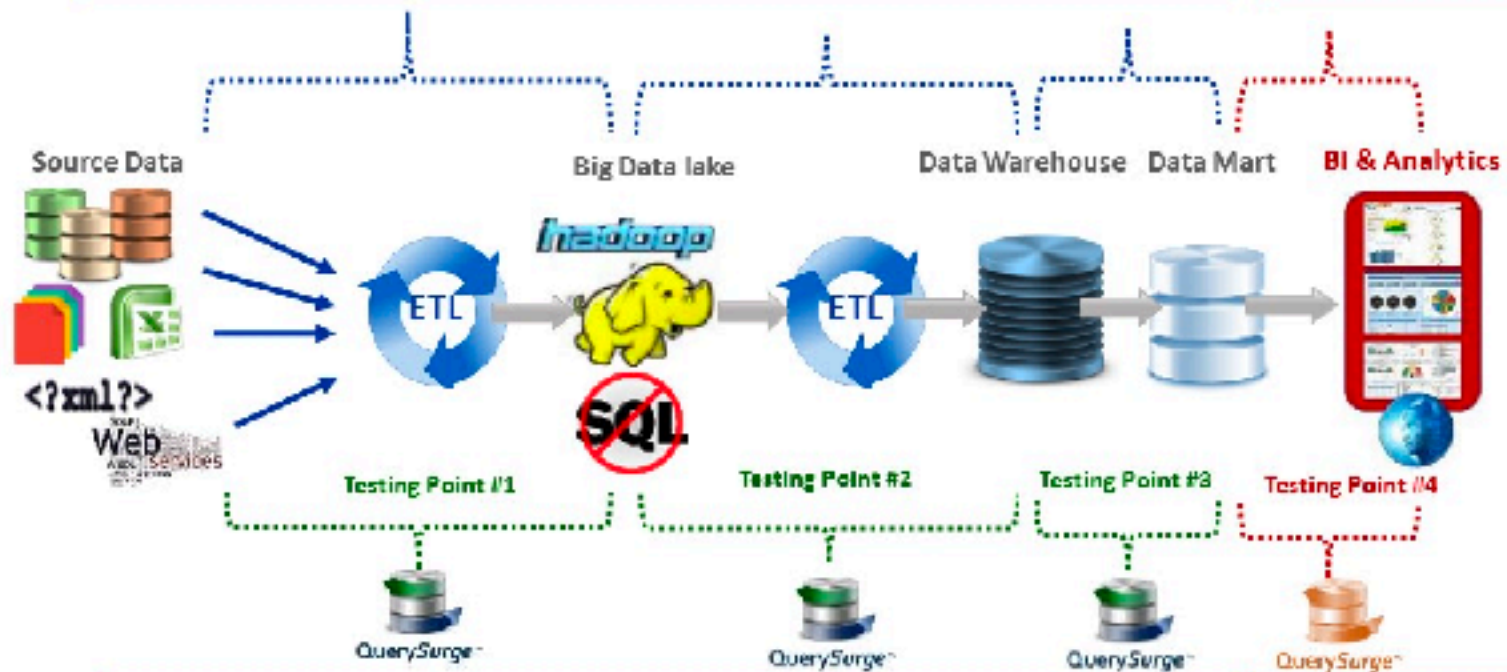


ETL – *Extract, Transform and Load*

Data Process: Developer & Tester 

ETL Developer: *Codes data movement based on Mapping Requirements*

BI User extracts data for reports



Data Tester: *Tests data movement based on Mapping Requirements*

Tester tests BI Reports



ETL – *Extract, Transform and Load*

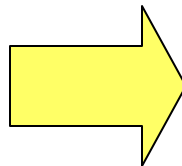
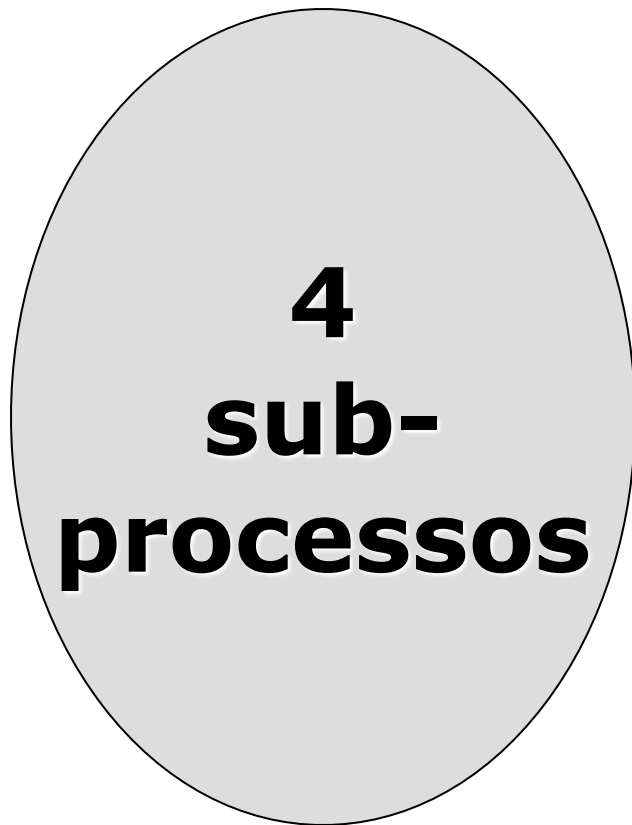
- Ferramentas de Testes ETL
 - RightData
 - Informatica Data Validation
 - QuerySurge
 - ICEDQ
 - Datagaps ETL Validator

ETL – *Extract, Transform and Load*

Figure 1. Magic Quadrant for Data Quality Tools



Processo de Desenvolvimento do BI



Modelagem Dimensional

Construção dos DM's / DW

Extração e Transformação
(fontes internas e externas)

Desenvolvimento das Aplicações

Dashboard

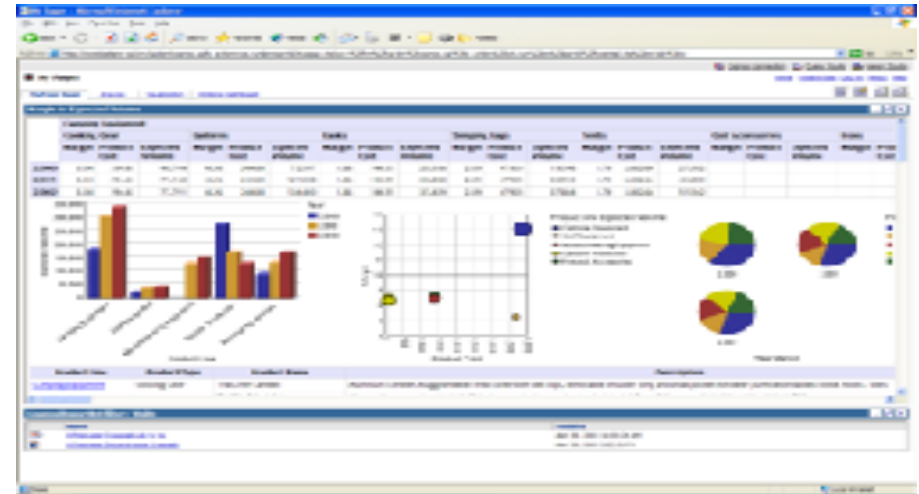
Conceito



- *Dashboard* é uma representação visual das informações **mais importantes** para atingir **um ou mais objetivos de negócio**, que tem sido consolidado em **uma única tela** para acompanhamento e monitoramento do negócio de **forma ágil**.
- Um *Dashboard* não é obrigatoriamente necessário ser apresentando em um computador. Pode ser um quadro, um **tablet** ou **smartphone**. O importante é o *Dashboard* não tenha barras de rolagem.

Dashboard

Conceito



- O **Dashboard** apresenta complexas informações de negócios de forma rápida e intuitiva. Com a visualização simultânea de diversos quadros possibilitando analisar as informações que se correlacionam e, assim, tomar decisões melhores.
- Tem impacto visual instantâneo.
- Exploração Interativa. Análise com mais detalhes (Zoom) uma área de interesse, podendo visualizar diferentes novas perspectivas.

Conceito

[illegible]

- *Dashboard* **não** é um relatório.
- Um relatório é a apresentação de dados transformados em informações formatadas e organizadas de acordo com requisitos de negócio específicos.
- Os relatórios podem ser imagens estáticas simples ou apresentações altamente interativas dos dados. Pode ordenar, filtrar e agregar, desagregar ou explorar dados em alguns tipos de relatório.

Dashboard

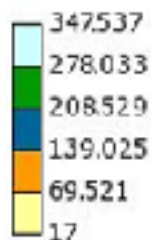
Exemplo

Posição: 2010/03/01

Valores Contratados por UF



Valor das Contrataç...



Selecione uma Fonte:

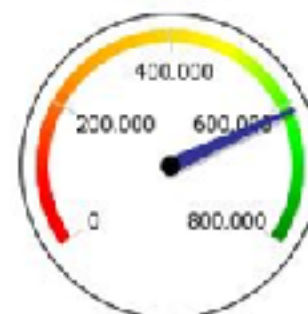
Fontes | R\$ Milhões



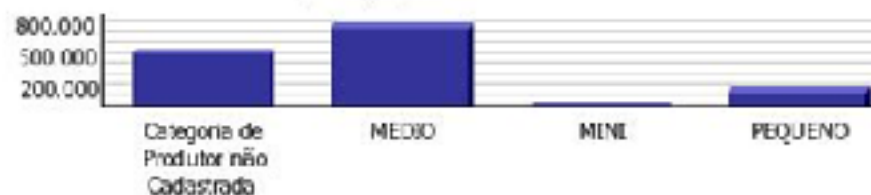
PRONAF | R\$ Milhões



MPE | R\$ Milhões



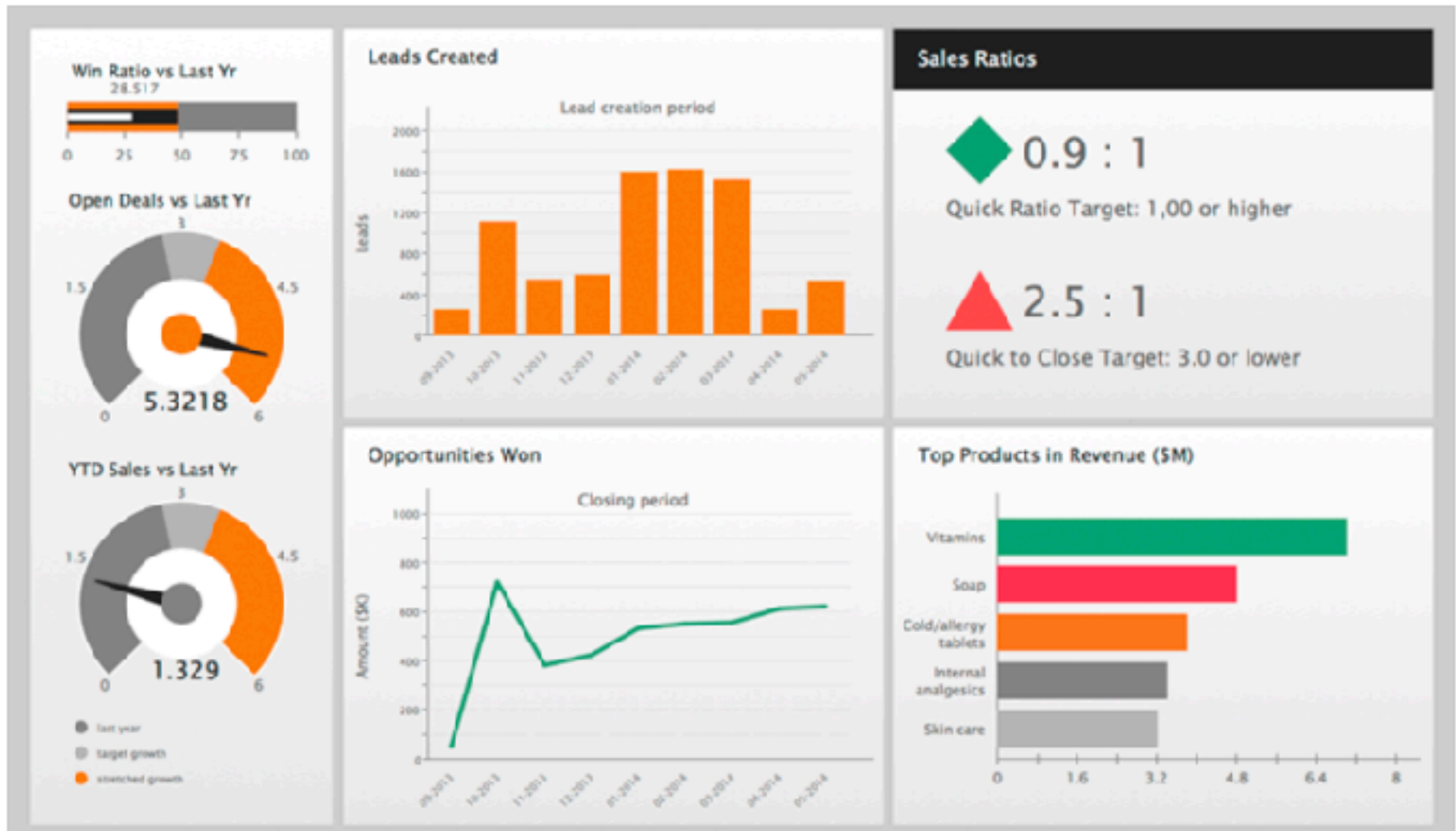
Categoria | Quantidade Contrada



Fonte de Recursos	Valor Contratado	Meta FNE
FNE	R\$ 935.292	R\$ 8.468.000

Dashboard

Exemplo





Dashboard

- *Design Dashboard*
 - O desafio fundamental do design do *dashboard* é exibir todas as informações necessárias em uma única tela, de forma clara e sem distrações, de forma que possa ser rapidamente assimilada
- As Camadas de Apresentação
 - Monitoramento
 - Análise
 - Gerenciamento

Dashboard

Steps for Selecting the Right Data Visualizations

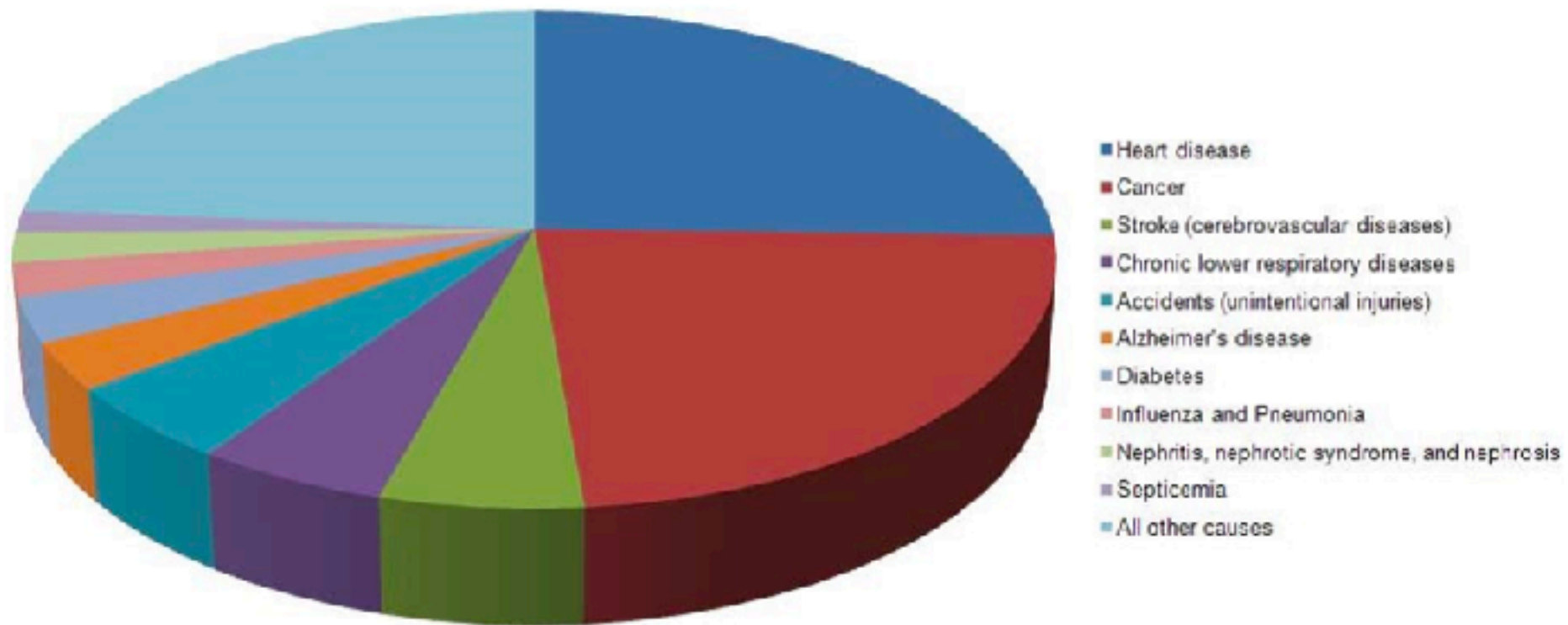


Source: Gartner 2019
ID: 383886

Dashboard

Example of Pie Chart: Leading Causes of Death in the U.S., 2007

Total Deaths in America by Cause in 2007



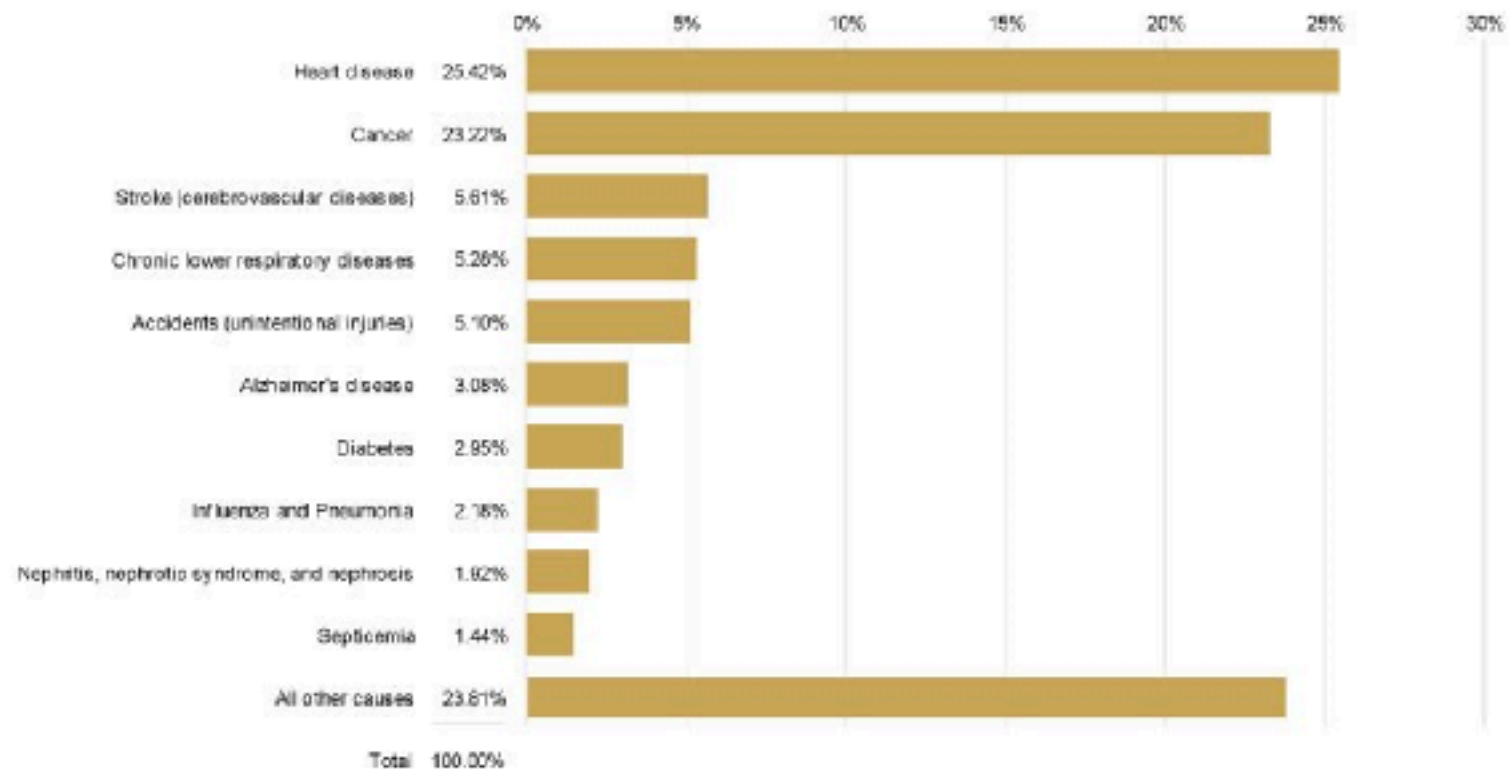
Source: Adapted from Stephen Few, "Data Visualization for Human Perception" in The Encyclopedia of Human Computer Interaction, 2nd ed. (CC BY-NC-ND 4.0 International License)

ID: 303006

Dashboard

Example of Bar Chart: Leading Causes of Death in the U.S., 2007

Total Deaths in America by Cause in 2007

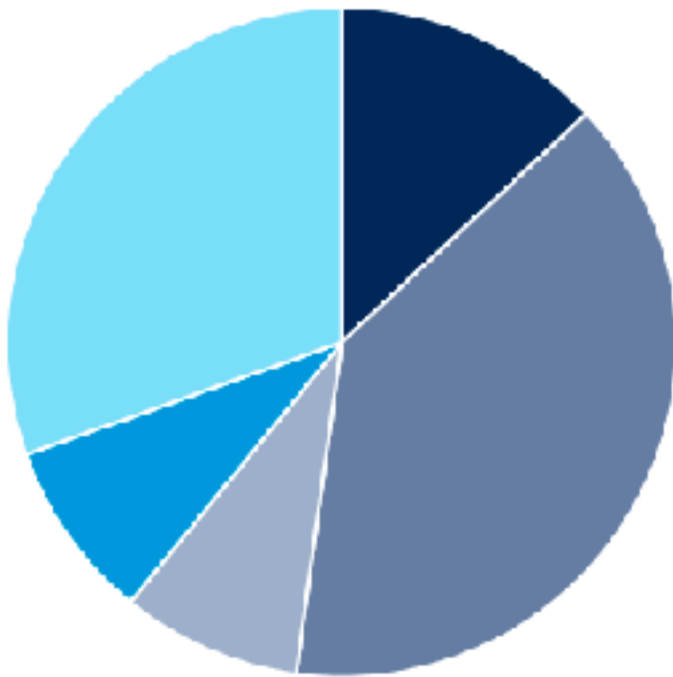


Total 100%

Source: Adapted from Stephen Few, "Data Visualization for Human Perception" in *The Encyclopedia of Human-Computer Interaction*, 2nd ed. (CC BY No Derivatives License).
ID: 383086

Dashboard

Comparing a Pie Chart and a Donut Chart With the Same Data



Source: Gardner
ID: 383886

Dashboard

Suitability of Visualization Types to Data Types

Visualization Type	Data Type		
	Categorical	Continuous	Combination of Types
Area Chart		✓	
Bar Chart	✓		
Box and Whisker Plot	✓		
Bubble Chart			✓
Bullet Graph			✓
Column Chart	✓		
Donut Chart	✓		
Dual Axis Chart			✓
Frequency Polygon		✓	
Funnel Chart	✓		
Heat Map	✓	✓	
Histogram		✓	
Line Chart		✓	
Marmaleke Chart	✓		
Meters and Gauges	✓	✓	
Multiseries Chart		✓	
Pictograph	✓		
Pie Chart	✓		
Scatter Plot		✓	
Stacked Bar Chart	✓		
Stacked Column	✓		
Timeline		✓	
TreeMap	✓		
Venn Diagram	✓		
Waterfall Chart	✓		
Word Cloud	✓		

Dashboard

Executive-Level Sales Dashboard Example

HIGH & LOW PERFORMING ACCOUNTS

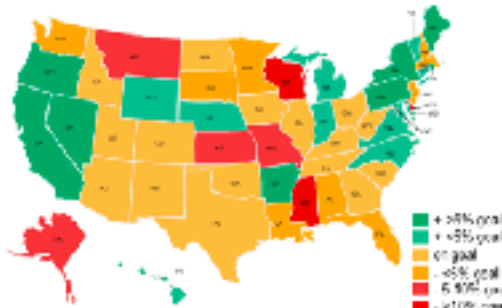
Top 5 Accounts (over target)

Acme Corp (#07234)	+362%
Account Lead: George S.	Region: Midwest
XYZ, Inc. (#48289)	+118%
Account Lead: Evelyn T.	Region: Pacific
Happy, Ltd. (#48111)	+112%
Account Lead: Evelyn T.	Region: Pacific
LaserCo (#19087)	+78%
Account Lead: Carl L.	Region: South
Food Group (#09133)	+34%
Account Lead: Patrick M.	Region: Northeast

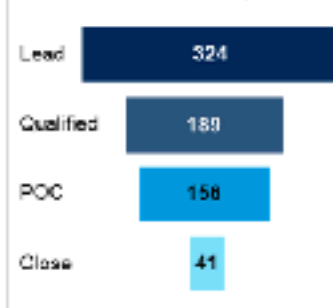
Bottom 5 Accounts (under target)

Fay Ent. (#22219)	-412%
Account Lead: Marc J.	Region: South
Omega7 (#73289)	-99%
Account Lead: Derek M.	Region: Midwest
City Service (#63477)	-89%
Account Lead: Shirley Y.	Region: South
ZS, Inc. (#00195)	-61%
Account Lead: Lori G.	Region: Northeast
VibPro (#76183)	-11%
Account Lead: Lori M.	Region: Midwest

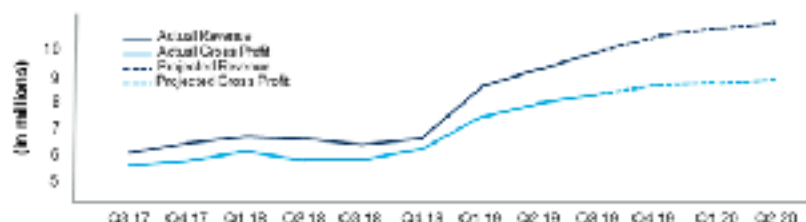
SALES PERFORMANCE BY REGION - FY 2019



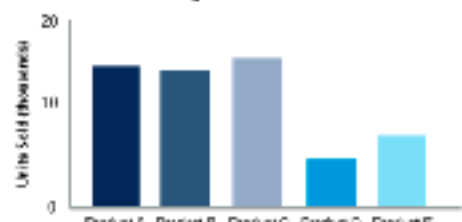
SALES FUNNEL - Q3 2019



REVENUE & GROSS PROFIT



UNITS SOLD By PRODUCT - Q3 2019



SOCIAL MEDIA SNAPSHOT

Overall Engagement

- 2.2%

compared to last week

New Followers

+ 97

across all platforms

Overall Sentiment

92% 8%

positive & neutral negative

Recent High Impact Post

Twitter 2 days ago

Ready for summer? Check out our latest products in our online showroom! <http://url.ly/15XK7/>

NEWS

Severe Weather Causing Major Travel Delays in Northeast

NATIONAL DAILY NEWS NETWORK

published 8 hours ago

Severe weather delays of travel, causing major disruptions in travel and causing major delays in travel. The weather is causing major delays in travel and causing major delays in travel. The weather is causing major delays in travel and causing major delays in travel.

Multination Trade Deal Nearing Completion; Good News for Manufacturers

INDUSTRY TRADE NEWS

published 2 hours ago

Multination trade deal nearing completion, good news for manufacturers. The deal is expected to be completed by the end of the year. The deal is expected to be completed by the end of the year. The deal is expected to be completed by the end of the year.

Two Indicted in Corruption Charges at City Hall

LOCAL DAILY NEWS NETWORK

published yesterday

Two indicted in corruption charges at city hall. The charges are related to the city hall. The charges are related to the city hall. The charges are related to the city hall.

ALERTS

Stock Price ▲ 12.16 (0.75%)

! Shipment Entering Port !

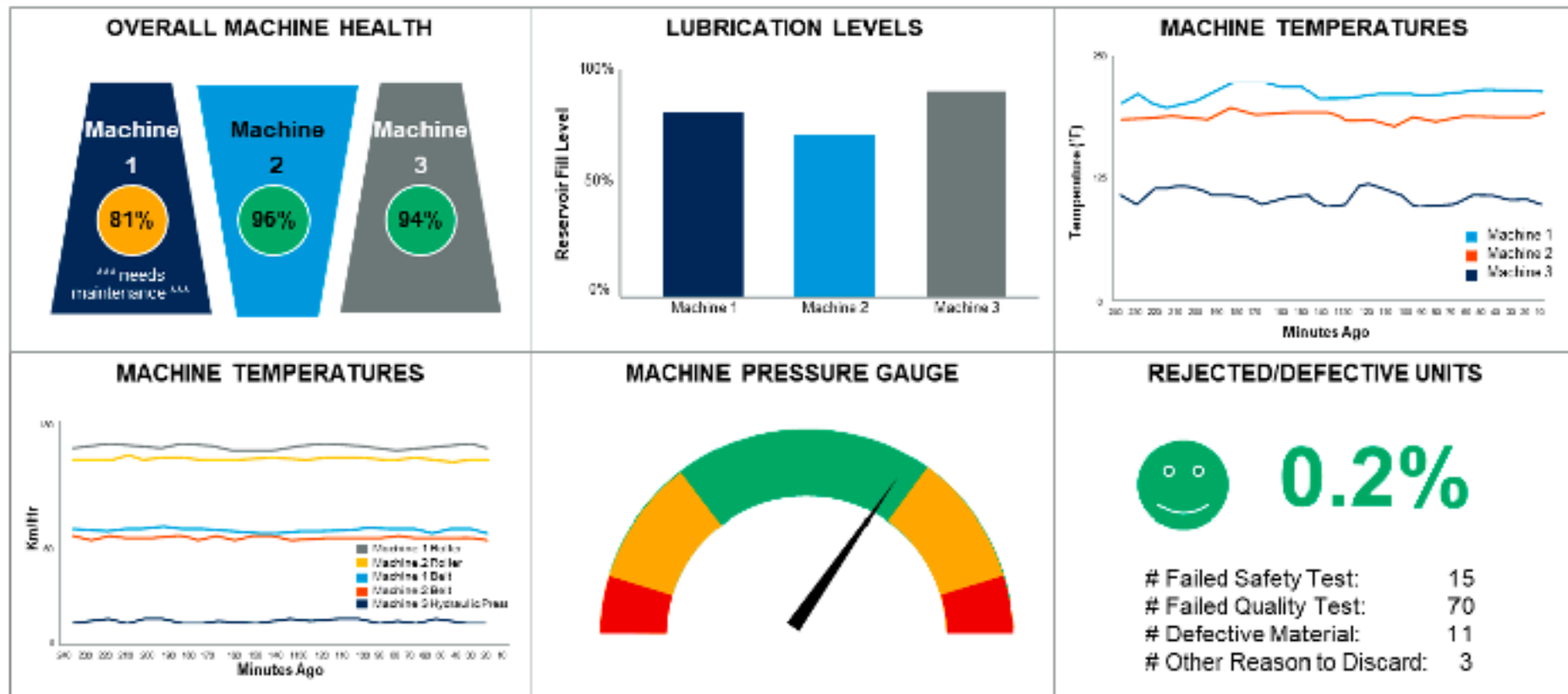
Estimated clearance date: 3 days
Estimated arrival at destination: 6 days

Upcoming Sales Pitches

South: Jane A. - Thurs - Miami, FL
Midwest: Alex B. - Fri - Chicago, IL
Pacific: Julia G. - Fri - Portland, OR

Dashboard

Drilled-Down Machine-Specific IoT Dashboard Example





Melhores Práticas de Design para *Dashboard*

- KPI de referência com padrões da indústria
- Envolver as métricas com metadados contextuais
- Validar o Design por um Especialista em Usabilidade
- Priorizar e classificar alertas e exceções
- Enriquecer o painel de controle com comentários de empresas e usuários
- Apresentar informações em três níveis diferentes
- Fornecer análise orientada



Dashboard

- O que procurar em um *dashboard*
 - Uso de componentes visuais para destacar dados e exceções que exigem ação
 - Transparente para o usuário, o que significa que requer treinamento mínimo e são extremamente fáceis de usar
 - Combinar dados de uma variedade de sistemas em uma visão única, resumida e unificada do negócio
 - Habilitar *drill-down* ou *drill-through* para fontes de dados ou relatórios subjacentes
 - Apresentar uma visão dinâmica do mundo real com dados atuais
 - Exigir pouca codificação para implementar, implantar e manter



Ferramentas de Visualização de Dados

Ferramentas de BI

MicroStrategy

 Microsoft

 **birst**

 **SAP**

 **sas**

 **alteryx**

 **tableau**
SOFTWARE

PROGNOZ

 **Information
Builders**

 **Logi**
ANALYTICS

 **TIBCO**

 **DATAWATCH**

 **pentaho**
A HCL Group Company

 **Qlik**

ORACLE

Ferramentas de BI



Visualização de Dados

"O uso de representações visuais para explorar, dar sentido e comunicar dados".

- Visualização de dados vs. Visualização de informações
 - Informação = agregação, resumo e contextualização de dados
- Relacionado a gráficos de informação, visualização científica e gráficos estatísticos
- Muitas vezes inclui gráficos, ilustrações, ...

Visualização Analítica

- Um termo recentemente cunhado
 - Visualização de informações + análise preditiva
- Visualização de informações
 - Descritivo, orientado para trás
 - "O que aconteceu" o que está acontecendo “
- Análise preditiva
 - Preditivo, focado no futuro
 - "O que vai acontecer" "por que isso vai acontecer“
- Há um forte movimento em direção à análise visual

visual analytics

O Crescimento da *Visualização de Dados e Analítica*

- Quadrante Mágico para Plataformas de *Business Intelligence e Analytics*
- Existe um movimento para as ferramentas de visualização



OLAP



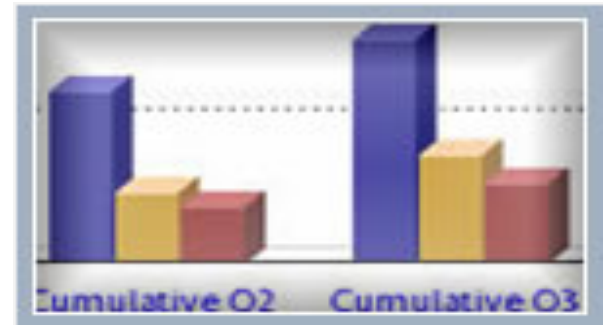


OLAP

- Conjunto de ferramentas voltadas para acesso e análise *ad-hoc* de dados
- Objetivo de uma ferramenta OLAP
 - “Transformar dados em informações capazes de dar suporte a decisões gerenciais de forma amigável e flexível ao usuário e em tempo hábil”

OLAP

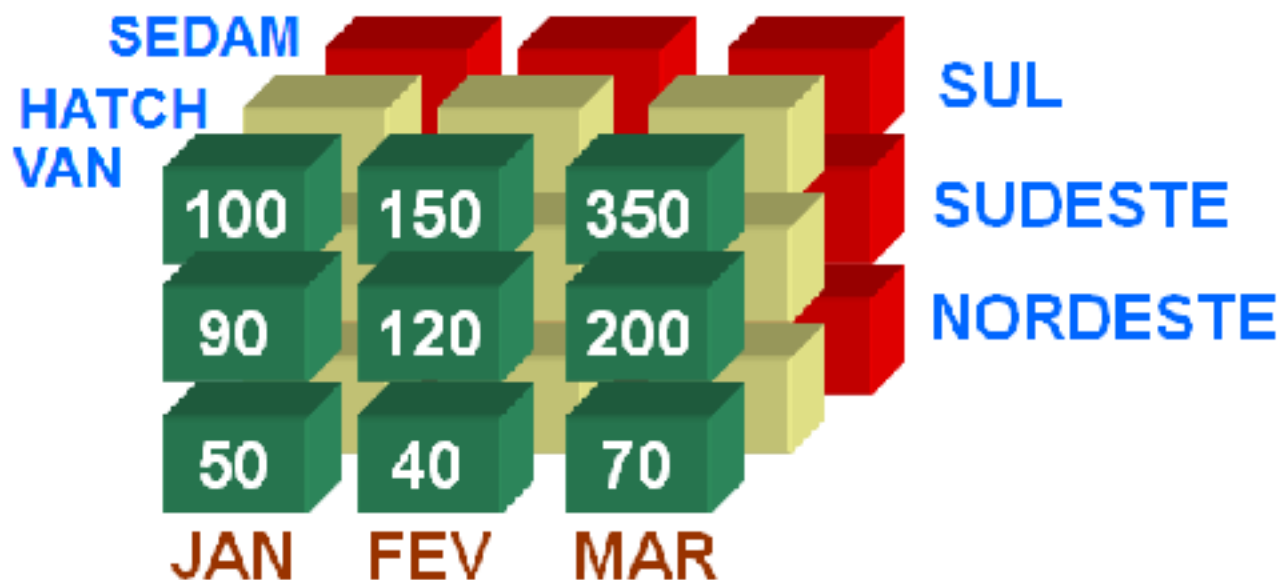
Características



- OLAP – *On-Line Analytic Processing*: fornece para organização um método de acessar, visualizar, e analisar dados corporativos com alta flexibilidade e performance.
- Disponibiliza os dados em forma de cubos para a análise dos dados em diversos ângulos de visão, sobre os aspectos de negócio. Trata as informações como Dimensões e Indicadores.

OLAP

- Cubo OLAP – Cubo de dados
 - Estrutura que facilita aos usuários visualizar os dados segundo suas dimensões



- Operações de navegação OLAP
 - Slice-dice / drill-down / roll-up / drill-up



OLAP

Operações

- Principais operações
 - Pivot
 - Drill-down
 - Driw-up (Roll-up)
 - Slice-and-Dice
- Operações podem ser combinadas para se obter a visualização desejada



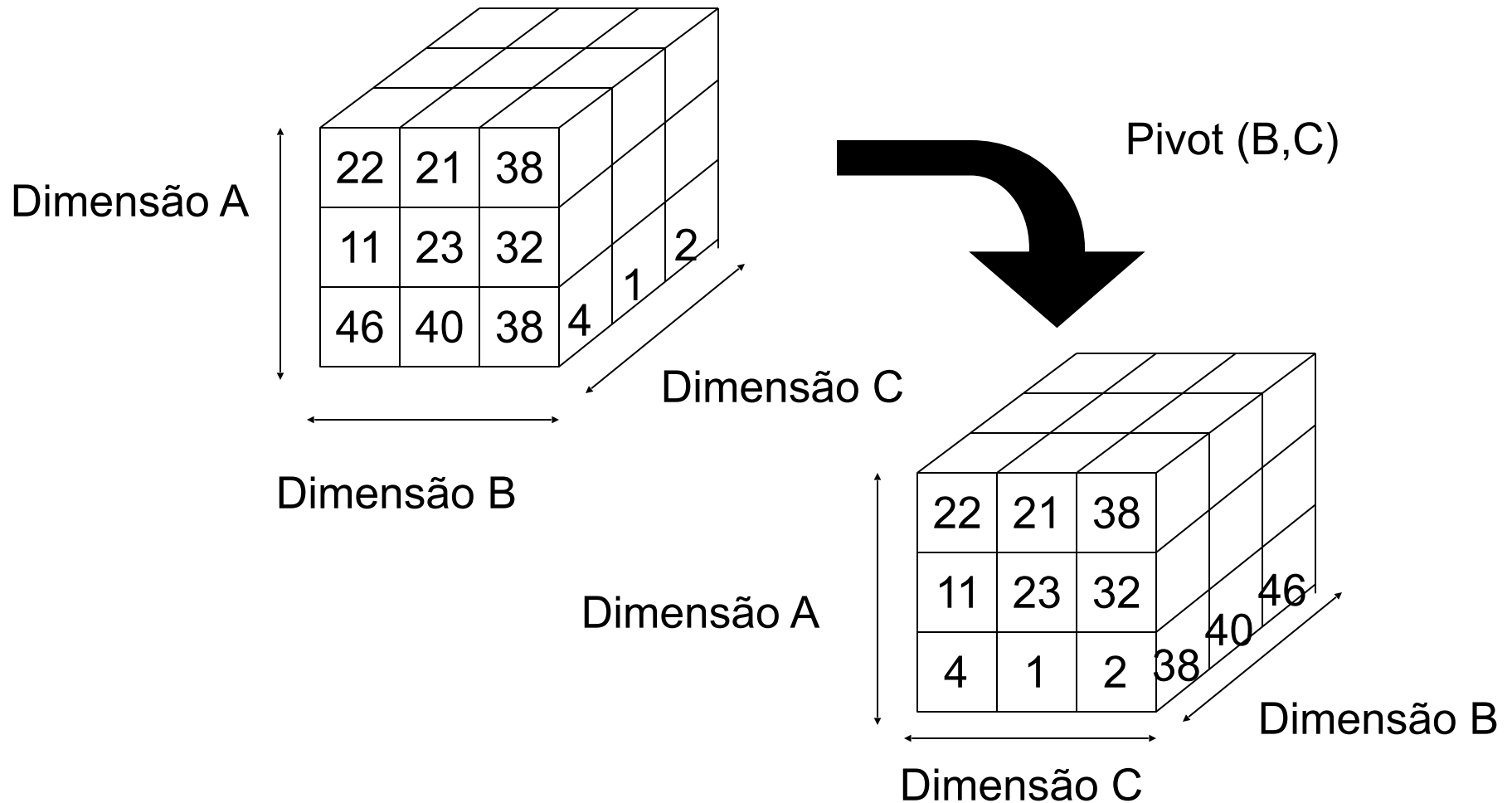
OLAP

Operações - PIVOT

- Rotação do hipercubo
- A forma mais comum de visualização
- Corresponde a inversão dos eixos das dimensões para, por exemplo, uma posterior rolagem.

OLAP

Operações - PIVOT

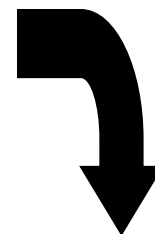


OLAP

Operações – PIVOT - Exemplo

Quantidade Vendida (milhares)	2019	
	Liquidificador	Micro-Ondas
SUL	77	152

Pivot (local, produto)



Quantidade Vendida (milhares)	2019
	SUL
Liquidificador	77
Micro-ondas	152



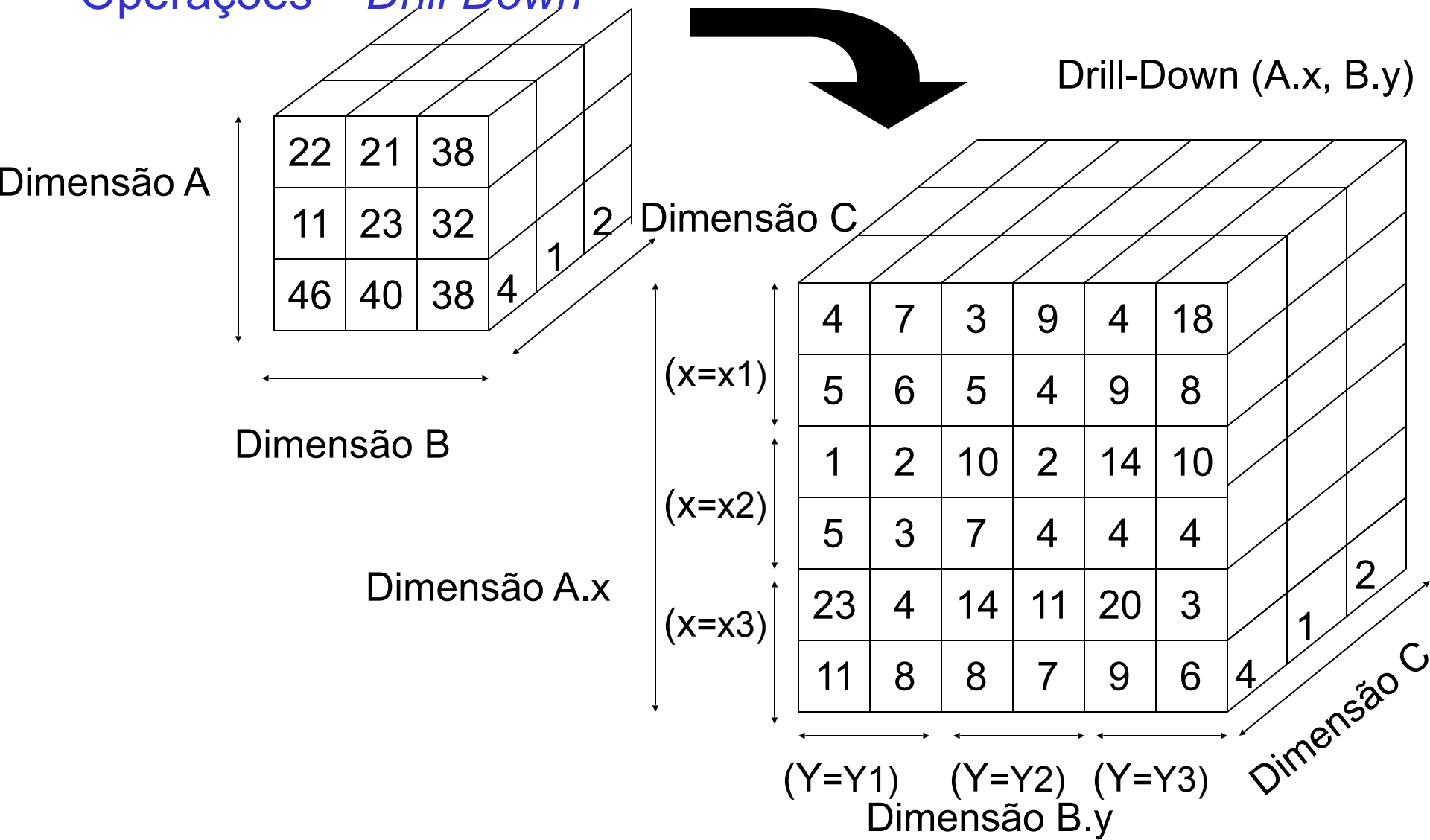
OLAP

Operações – *Drill Down*

- Determinam o detalhamento de um consulta
- As consultas são mais restritas se existirem mais detalhes nos critérios de seleção
 - Ex. País x Cidade
- Os resultados são mais detalhados

OLAP

Operações – *Drill Down*

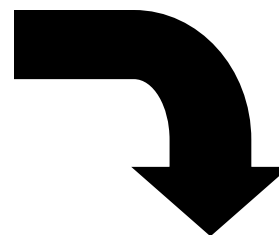


OLAP

Operações – Drill Down - Exemplo

Quantidade Vendida (milhares)	2019	
	Liquidificador	Micro-Ondas
SUL	77	152

Drill-Down
(*local.regiao.estado*,
produto.tipo.marca)



Quantidade Vendida (milhares)		2019			
		Liquidificador		Micro-Ondas	
		Walita	Faet	Sharp	Consul
SUL	RS	14	16	40	25
	SC	9	10	12	20
	PR	15	13	30	25



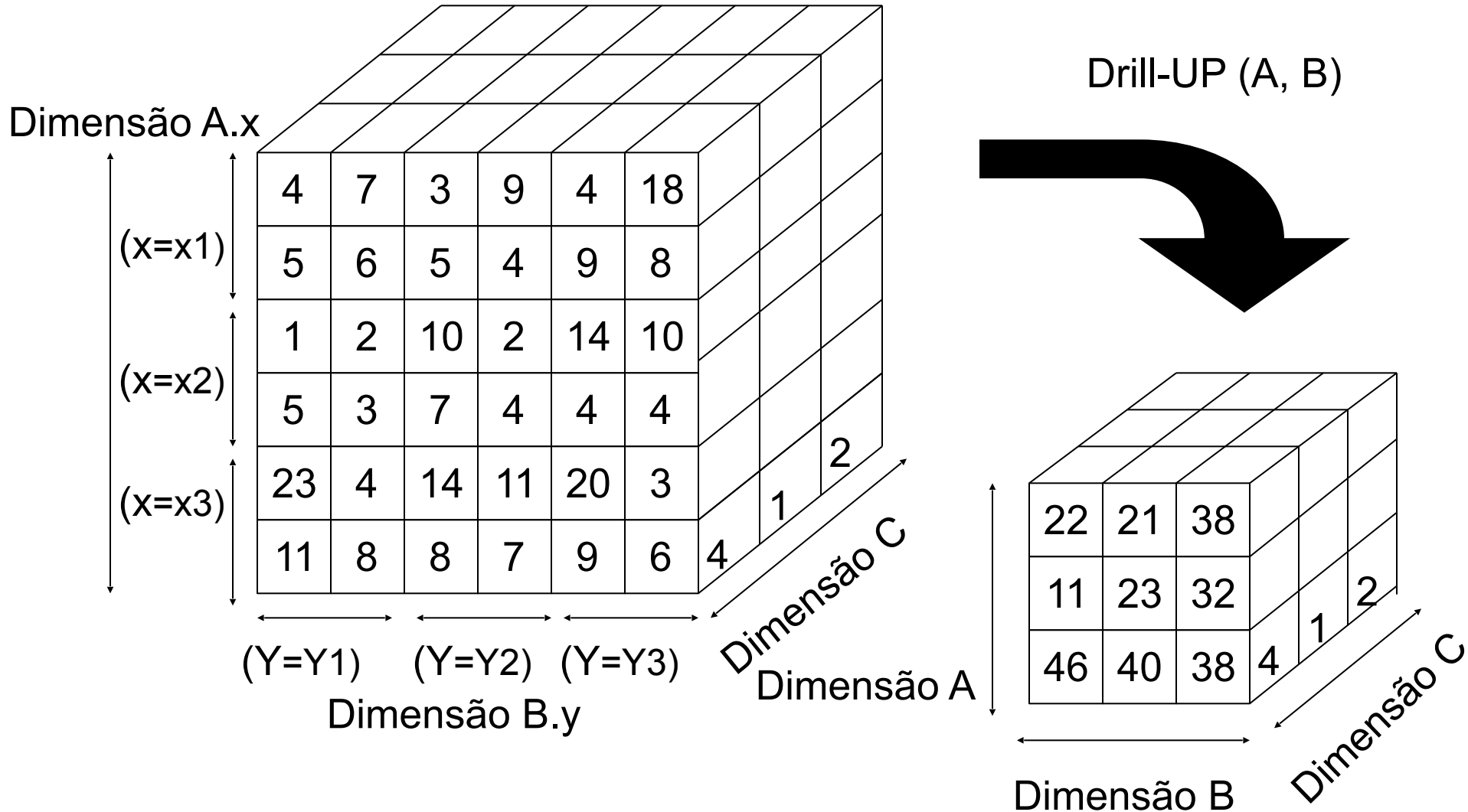
OLAP

Operações – *Drill Up*

- Determinam o detalhamento de um consulta
- As consultas são mais abrangentes se existirem menos detalhes
 - Ex. Cidade x pais
- O resultado é mais sumarizado

OLAP

Operações – *Drill Up*

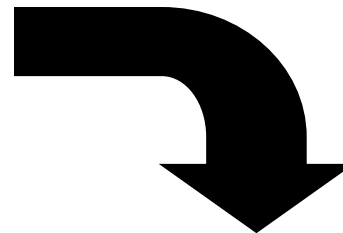


OLAP

Operações – Drill Up- Exemplo

Quantidade Vendida (milhares)		2019			
		Liquidificador		Micro-Ondas	
		Walita	Faet	Sharp	Consul
SUL	RS	14	16	40	25
	SC	9	10	12	20
	PR	15	13	30	25

Drill-up (produto.tipo)



Quantidade Vendida (milhares)		2019	
		Liquidificador	Micro-ondas
SUL	RS	30	65
	SC	19	32
	PR	28	55



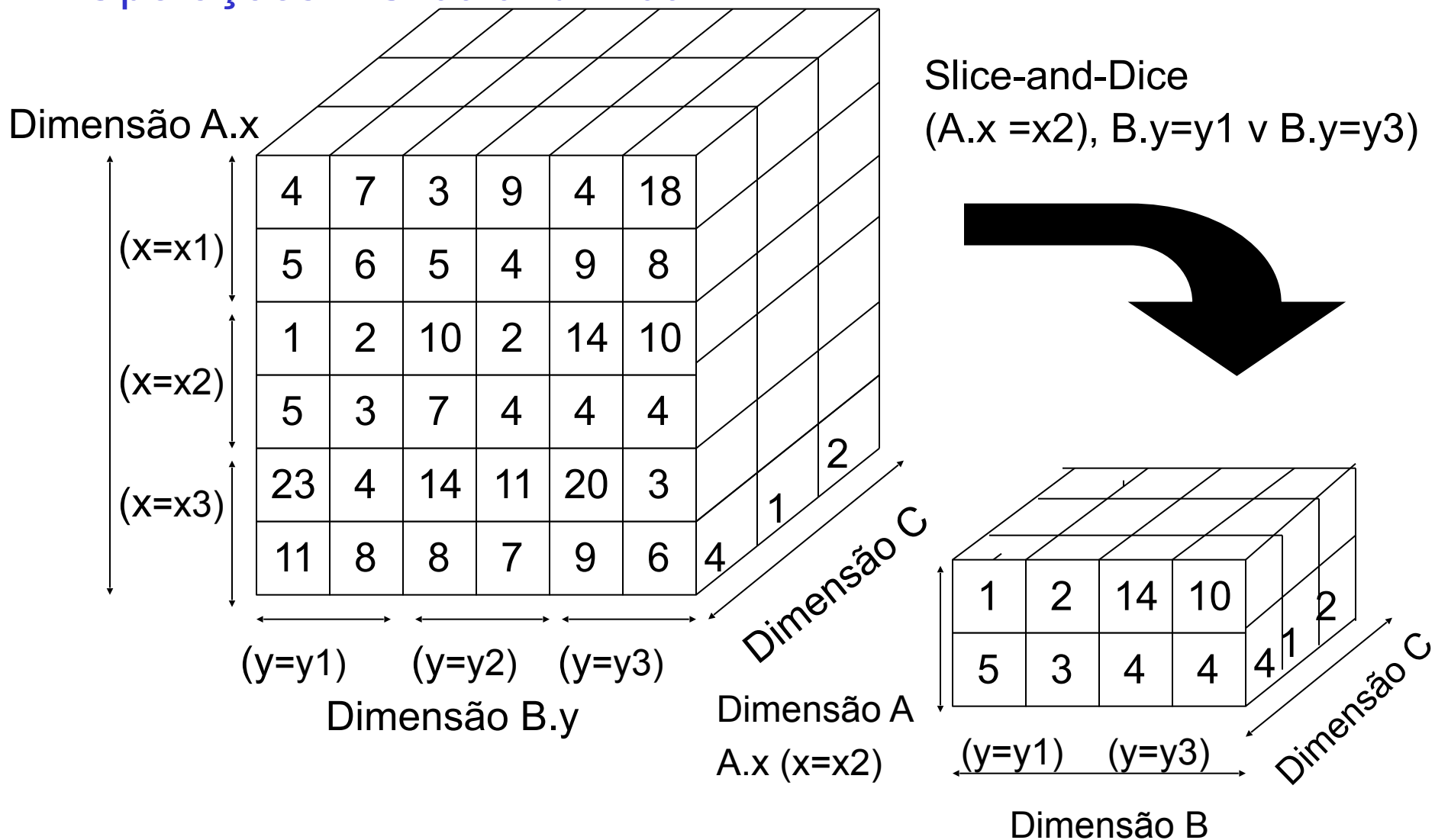
OLAP

Operações – *Slice and Dice*

- As consultas são realizadas de acordo com as restrições impostas
- As restrições correspondem a atributos de dimensões com determinado valor de consulta Quanto mais atributos de dimensão mais restrito é o conjunto de resposta.
- Uma escolha de atributos de dimensões representa uma fatia(Slice) na visualização das dimensões
- A troca dos atributos de dimensões é uma operação de rolagem (dice).

OLAP

Operações – *Slice and Dice*



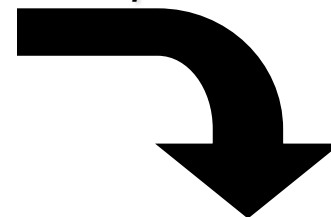
OLAP

Operações – *Slice and Dice*

Quantidade Vendida (milhares)		2019			
		Liquidificador		Micro-Ondas	
		Walita	Faet	Sharp	Consul
SUL	RS	14	16	40	25
	SC	9	10	12	20
	PR	15	13	30	25

Slice-and-Dice

*(local.região.estado = 'SC',
produto.tipo.marca = 'Walita',
produto.tipo.marca = 'Sharp')*



Quantidade Vendida (milhares)		2019	
		Liquidificador	Micro-ondas
		Walita	Sharp
SUL	SC	9	12

Especialização em Ciência de Dados com Big Data, BI e *Data Analytics*



Prof. Fernando Siqueira
fernando.siqueira@uni7.edu.br