

# Análise e Visualização de Dados

## Projectos BI - Introdução

Cláudio Sapateiro

Mestrado em Engenharia do Software

ESTSetúbal – IPS

[claudio.sapateiro@estsetubal.ips.pt](mailto:claudio.sapateiro@estsetubal.ips.pt)

# Conteúdos

- Projectos *Business Intelligence* (BI)
- Sistemas OLTP vs OLAP
- *Data Warehouse* e Sistemas de Apoio à Decisão
- Projecto de *Data Warehouses*
- O Processo ETL: overview
- Modelação Dimensional: Introdução
  - A metáfora do cubo de dados

# Projectos BI

## Sistemas Operacionais vs SAD

- Sistemas Operacionais
  - OLTP: OnLine Transaction Processing
  - Importante fonte de entrada/recolha de dados
  - Dimensionados para transacções curtas, repetitivas e bem estabelecidas
  - Processamentos sobre dados actuais/correntes
- Sistemas de Apoio à Decisão (SAD)
  - OLAP: OnLine Analytical Processing
  - Agregadores de dados de várias fontes
  - Dimensionados para promover a navegação e análises sobre os dados
  - Suportam processamentos exploratórios, mais longos (e exigentes em termos de recursos)
  - Processamentos sobre imagem estática de dados históricos
- *Data Warhouses são desenhados especificamente para OLAP*

# Projectos BI

## Dados Operacionais vs Dados de Apoio à Decisão

<b>Dados Operacionais</b>	<b>Dados de Apoio à Decisão</b>
Orientados a transações	Orientados a análises
Processamento repetitivo	Processamento exploratório
Detalhados	Agregados
Acessos a poucos dados	Acessos a conjuntos de dados
Válidos no momento	Abrangem intervalos de tempo
Fontes de dados internas	Fontes de dados internas e externas
Atualizados em contínuo	Atualizações planeadas
Estrutura estática	Estrutura dinâmica
Relatórios pré-definidos	Relatórios personalizados
Alta disponibilidade	Disponibilidade relaxada
Não redundantes	Redundantes

# Projectos BI

## Operações OLTP vs Operações OLAP (exemplos)

- Operações OLTP
  - Transferências bancárias
  - Colocação de itens em carrinhos de compras
  - Pagamentos automáticos (e.g. portagens)
- Operações OLAP
  - Total de vendas de cada unidade de negócio por mês
  - Identificação de artigos mais vendidos por categoria
  - Avaliação de associações (e.g. sazonalidade de vendas)

# Projectos BI

## Modelos de Dados OLTP vs OLAP

- Modelos Dados de Suporte a OLTP
  - Mais complexos e normalizados
  - Muitas tabelas e relações
  - Interrogações analíticas requerem maior esforço
    - Mais difíceis de formular
    - Muitas junções de tabelas aumentam o tempo de resposta
- Modelos de Dados de Suporte a OLAP
  - Menos tabelas (que reúnem dados de múltiplas fontes)
  - Interrogações analíticas ao alcance dos decisores
  - Espelham mais directamente processos de negócio

# Projectos BI

## Data Warehouse e Apoio à Decisão

- Mais dados
  - Extracção periódica de dados dos (múltiplos) sistemas operacionais
- Melhores dados
  - Processos específicos para a transformação de dados
    - **Limpeza:** valores em falta, corrompidos ou duplicados
    - **Conformação:** uniformização de nomes e unidades de medida
- Melhores Ferramentas
  - Vista multidimensionais e inteligíveis sobre os dados
  - Suporte a análises exploratórias com vários níveis de detalhe
  - Aplicação e validação de modelos de prospecção de dados

## Exercicio

3 + 3

- Grupos de 2
- Escrever 3 características de dados OLTP
- Escrever 3 características de dados OLAP

**10:00**



# Exercicio

3 + 3

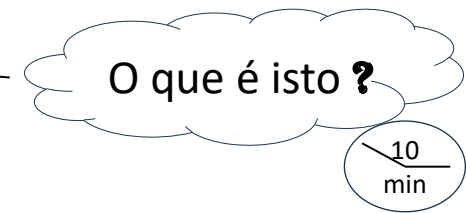
- Escrever 3 características de dados OLTP
- Escrever 3 características de dados OLAP
- Cada grupo apresenta 1 de cada

# PROJETOS BI/ANALYTICS

# Projectos BI

## Principais Objectivos do Projecto de DW

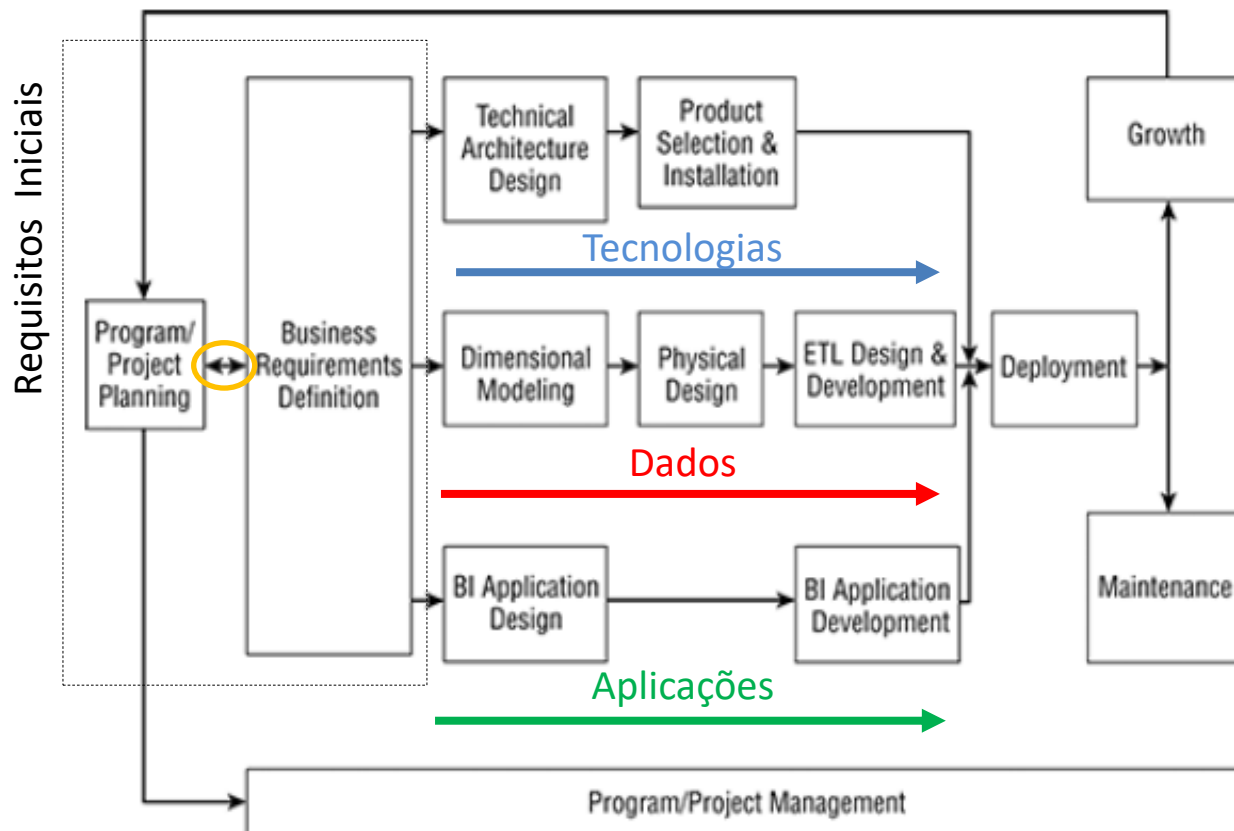
- Simplificar a consulta aos dados representativos dos processos de negócio



- Armazenar dados de forma coerente
  - Requer tratamento/processamento prévio
    - Identificação/caracterização
    - Consolidação de múltiplas fontes (*Limpeza* e *Conformação*)
- Acomodar alterações
  - Manutenção de histórico consistente

# Projectos BI

## Projecto de um Data Warehouse [Kimball]



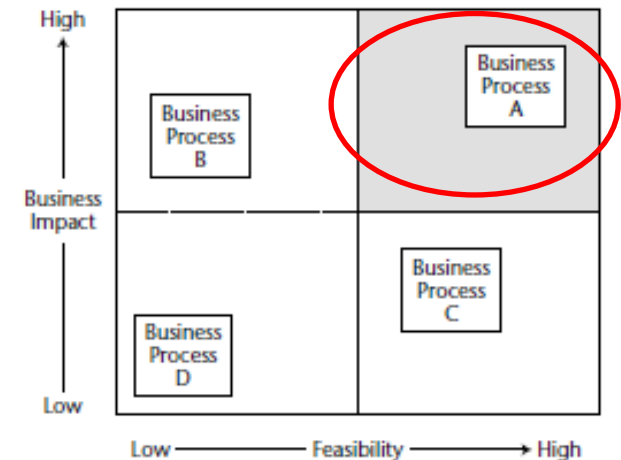
# Projectos BI

## Projecto de um Data Warehouse: *Requisitos Iniciais*

- Identificação das necessidades do negócio
  - Necessidades de informação para um suporte informado à tomada de decisão por parte dos utilizadores finais
  - Recurso a entrevistas e reuniões para estabelecer
    - objectivos (e.g. tipo de interrogações analíticas desejadas)
    - Prioridades: **matriz de exequibilidade/valor** →
    - Definição da **bus matrix**

↓

BUSINESS PROCESSES	COMMON DIMENSIONS							
	Date	Product	Store	Promotion	Warehouse	Vendor	Contract	Shipper
Retail Sales	X	X	X	X				
Retail Inventory	X	X	X					
Retail Deliveries	X	X	X					
Warehouse Inventory	X	X			X	X		
Warehouse Deliveries	X	X			X	X		
Purchase Orders	X	X			X	X	X	X



# Projectos BI

## Projecto de um Data Warehouse: *Requisitos Iniciais*

- Estabelecer as politicas relativamente a
  - necessidades temporais na actualização dos dados do DW
    - *batch* (periódica) *vs streaming oriented* (mais exigente e menos comum)
    - automática *vs* iniciada por alguém
    - fluxo do processamento ETL das dimensões (horizontal/independente *vs* vertical/sincronizado)
  - armazenamento de processamentos intermédios dos dados
  - segurança dos dados (e.g. backups, acessos)
- Identificar/adequar necessidades em termos de competências da equipa de projecto

# Projectos BI

## Projecto de um Data Warehouse

- Vertente de Tecnologias
  - Desenho da arquitectura
    - Baseado na infra-estrutura existente e apostas estratégicas em tecnologia (e.g. licenciamentos)
  - Selecção e instalação do produto
    - Avaliação comparativa de produtos existentes no mercado
- Vertente de Dados
  - Modelação dimensional
    - Processos prioritários, factos, medidas de negócio, dimensões , relatórios
  - Desenvolvimento do sistema ETL
    - Extracção, Transformação (*Limpeza e Conformação*) e carregamento de dados para o Data Warehouse

# Projectos BI

## Projecto de um Data Warehouse

- Vertente de Aplicações
  - Desenho de aplicações analíticas
    - Estudo prático de soluções para as necessidades dos decisores
    - Desenvolvimento de ***dashboards*** com vistas sobre o negócio e relatórios dinâmicos
- Implantação e Exploração (Deployment)
  - Integração na infra-estrutura, documentação e formação dos decisores
- Manutenção (Maintenance)
  - Monitorização e afinação de desempenho, cópias de segurança,...
- Crescimento (Growth)
  - Inclusão de mais dados e fontes, adaptação a novas necessidades



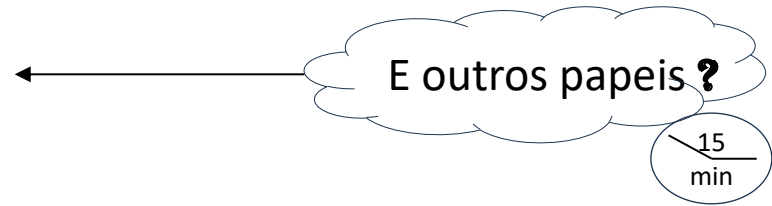
# Projectos BI

## **Projecto de um Data Warehouse:** *Considerações da Qualidade*

- Tratamento de Excepções
  - Políticas e mecanismos de suporte ao tratamento de excepções geradas pelo sistema ETL
    - e.g. uma BD contendo a identificação da excepção, do processo que a originou contextualizada pelo respectivo *timestamp*  
+ um sistema que permita a monitorização e acompanhamento (classificação de severidade e estado das acções de resolução)
- Recuperação (*Recovery & Restart*)
  - Acautelar efeitos de uma execução subsequente a uma execução prévia falhada
    - Atender a que à dados que podem eventualmente já ter sido processados

# Projectos BI

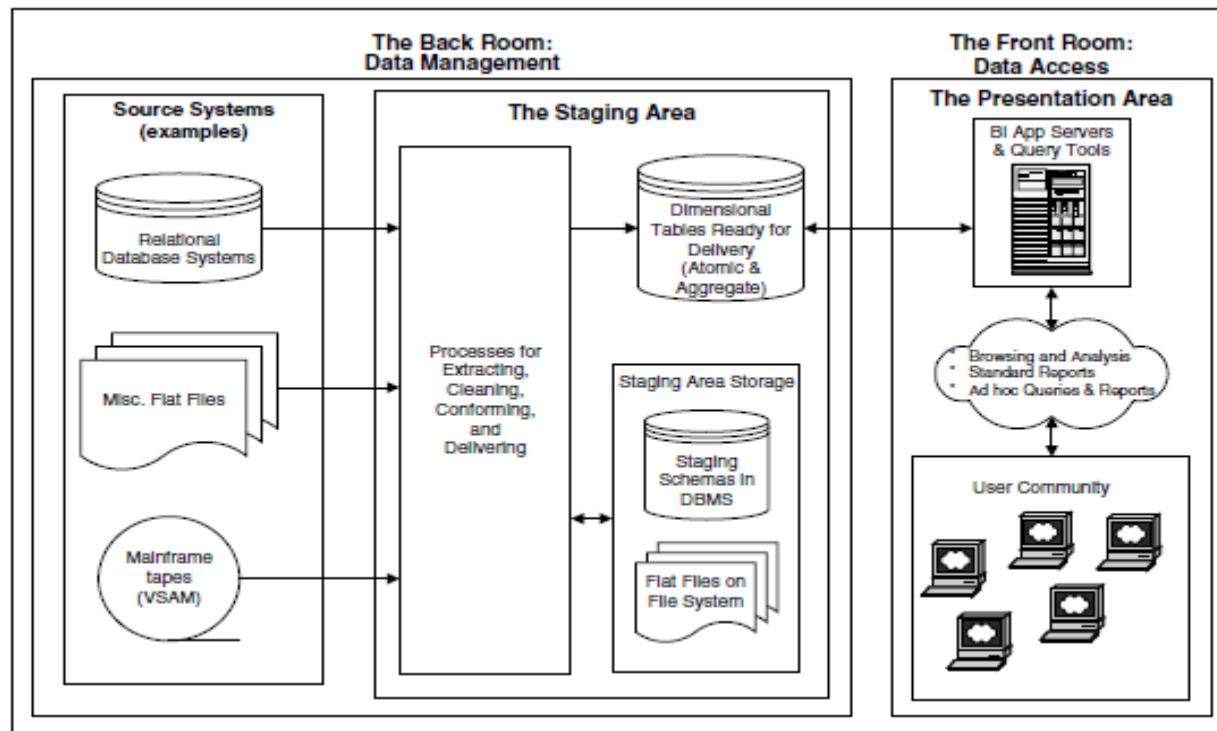
## Intervenientes no Projecto de DW



- Lado do Negócio
  - Sponsor: defensor da necessidade do DW
  - Líder do Projecto: perito do negócio que acompanha o projecto
  - Utilizadores: informam sobre os requisitos finais de utilidade do produto
- Lado da equipa de desenvolvimento
  - Gestor do Projecto: gere/coordena recursos, tarefas e *deliverables*
  - Analista de Sistemas: traduz necessidades dos utilizadores em requisitos
  - Engenheiro de Sistemas: avalia alternativas e define arquitectura
  - Consultor de Dados: define o modelo de dados
  - Programador da *staging area*: codifica processos do sistema ETL
  - Programador de aplicações analíticas: codifica ferramentas para os decisores
  - Administrador de BD: assegura e optimiza o funcionamento do DW
  - Técnico de manutenção: acompanha a implantação e exploração do DW

# Projectos BI

## Principais Componentes do DW



# Exercicio sumário

**10:00**

- Grupos de 2
- Sintetizar os 3 fluxos principais da framework do kimball
- Sintetizar os dois instrumentos de apoio à fase de levantamento e análise de requisitos

# Exercicio sumário

**10:00**

- Sintetizar os 3 fluxos principais da framework do kimball  
(Data, Technology, Applications)
- Sintetizar os dois instrumentos de apoio à fase de levantamento e análise de requisitos  
(Bus matrix e matriz de exequibilidade)

# Análise e Visualização de Dados

## Projectos BI - Introdução

Cláudio Sapateiro

Mestrado em Engenharia do Software

ESTSetúbal – IPS

### *Referências:*

- The Microsoft Data Warehouse Toolkit 2nd edition. Joy Mundy, Warren Thornthwaite, Ralph Kimball. Wiley. ISBN: 978-0-470-64038-8.

[claudio.sapateiro@estsetubal.ips.pt](mailto:claudio.sapateiro@estsetubal.ips.pt)