

Data Warehouse

1º Semestre



Mestrado em Desenvolvimento de Software e Sistemas Interativos

Instituto Politécnico de C.Branco

11ª Edição – 2021/22

Slides #6 – Ferramentas

Eurico Lopes

V.8-23.10.20

1

Índice

■ Ferramentas

- Tools – Seleção: Condução do Processo
- Tools - Modelagem de Dados
 - Modelagem de Dados: Características
 - Soluções Modelagem de Dados
- Tools – ETL
 - ETL: Características
 - Soluções ETL
- Tools – Condução da Análise de Dados
- Tools - Ferramentas vs Propósitos
- Tools – Posicionamento
- Tools - Exploração de Dados
 - OLAP – MOLAP
 - OLAP – ROLAP
 - OLAP – HOLAP
 - OLAP – DOLAP
 - Tools - Arquitetura Two Tier
 - Tools - Arquitetura Three Tier
 - Tools - Arquitetura Three+ Tier

■ Ferramentas (cont.)

- Tools - Reporting /Ad-hoc: Características
- Tools - Reformulação de Queries
- Tools - Redireccionamento de Queries e Multipass SQL
- Tools - Aggregate Awareness / ReWrite SQL
- Tools – Query Governing
- Tools – MOLAP
- Tools - Exploração de Dados – Exemplos
- Tools - Exercício Prático

2

DW Tools - Ferramentas

Sumário

- Neste capítulo pretende-se apresentar os diferentes tipos de ferramentas associados ao desenvolvimento de um DW.
- Em alguns casos complementar com alguma informação técnica sobre as soluções.
- Pretende-se também identificar recomendações para o processo de seleção dessas mesmas ferramentas.

Tools – Seleção: Condução do Processo

- Continuar focado nos objetivos de negócio
 - “a tecnologia serve o negócio e não o contrário”
 - ponto de vista importante sobretudo na seleção de ferramentas de exploração - as ferramentas para os utilizadores do negócio
- Garantir a sustentação pós-entrega do DW e o ciclo de manutenção evolutiva
 - aqui os aspetos técnicos têm relevância - as ferramentas para os informáticos que constroem e suportam o DW
 - se o DW funciona bem e pode evoluir o negócio fica também a ganhar!

Tools – Seleção: Condução do Processo

- Características técnicas, funcionais e de usabilidade:
 - Matrizes de características/funcionalidades
 - Relevância e criticidade (“must have” vs “nice to have” ou classificação numérica)
 - Classificação das ferramentas de mercado perante a matriz

Tools – Seleção: Condução do Processo

- Análise do mercado
 - Recolha de prospetos comerciais dos produtos
 - Recolha de “white papers”/“case studies” patrocinados pelos “vendedores”
 - Web
 - Relatórios especialistas (Ex. Ovum)
- Pré-seleção grosseira
 - O processo de seleção pode transformar-se num processo caro!

Tools – Seleção: Condução do Processo

- Seleção “fina” através de demonstração de produtos pelos seus distribuidores:
 - Indispensável uma boa preparação de quem assiste às sessões de demonstração
 - Recomendar ao “vendedor”: “não queremos uma apresentação comercial”
- Possibilidade de prototipagem:
 - Indexar a seleção definitiva a um miniprojecto de prototipagem em parceria com o “vendedor”

Tools – Seleção: Condução do Processo

- Características do “**vendedor**”:
 - Estratégia para a evolução do produto
 - Presença Nacional
 - Formação
 - Suporte
- Característica dos “**mercado**”
 - “Market share”
 - Viabilidade da estratégia para a evolução do produto
 - Disponibilidade de “know how” no mercado

Tools - Modelagem de Dados

■ Objetivo

- Ferramenta que suporte a modelagem de dados e respetiva implementação, através da geração de scripts para criação do esquema de bases de dados.

Tools - Modelagem de Dados

- Facilitar o processo de desenho e modelagem multidimensional
 - Modelo conceptual
- Garantir documentação sobre os modelos de dados
 - Suportar manutenção evolutiva e análises de impacto
 - Suportar as fases de construção das “aplicações” de exploração de dados
- Desenhar e gerar o modelo físico a implementar
 - Scripts DDL

Tools - Modelagem de Dados: Características

■ Características técnicas e funcionais a analisar:

Características		Obrigatório (S/N)	Peso (1-10)
1000	Técnicas		
1005	Suporta a metodologia ER		
1010	Suportada por um repositório partilhado		
1015	Repositório implementado em multi-RDBMS de mercado		
1020	Configuração de notação		
1030	Modelos e sub-modelos		
1040	Features de star schemas		
1050	Features de particionamento		
1060	Acesso via web browser		
1070	Geração de modelos em formatos web		
1080	Geração de modelos físicos a partir de modelos lógicos		

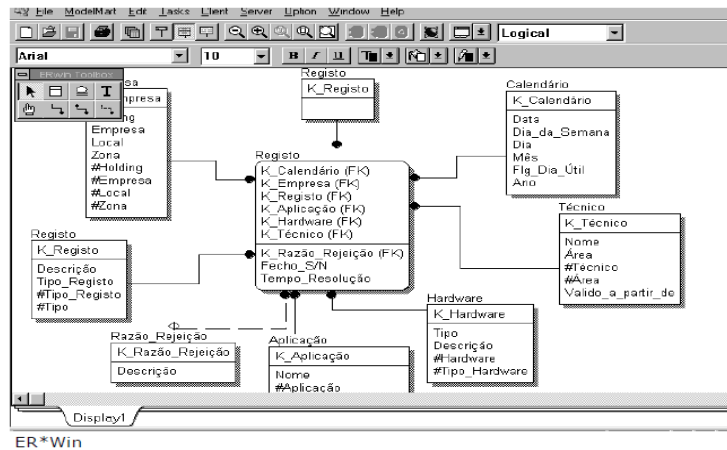
Tools - Modelagem de Dados: Características

■ Características técnicas e funcionais a analisar:

Características		Obrigatório (S/N)	Peso (1-10)
1000	Técnicas		
1090	Reverse engineering de modelos físicos de diferentes RDBMS		
1100	Geração de scripts para criação de modelos físicos em diferentes RDBMS		
1120	Geração de alterações a modelos físicos a partir de alterações ods respectivos modelos lógicos		
1130	Geração de scripts para alteração de modelos físicos já implementados		
1140	Capacidade de impressão e formatação de reports		
1150	Copy/Paste para ferramentas MSOffice		
1160	Inclusão de comentários nos diagramas		
1170	Gestão de versões de modelos		

Tools – Soluções Modelagem de Dados

- Exemplos: Power Designer, ERWin, ...



Tools – Soluções Modelagem de Dados

- Uma comparação:
 - Competitive Analysis of Market Leaders in Data Modeling: PowerDesigner, ERwin and ER/Studio
 - http://www.sybase.com.br/files/White_Papers/DM-Industry-Leaders-Analysis-110608-wp.pdf
 - Disponível em Textos Auxiliares

Tools - ETL

■ Objetivo

- ❑ Ferramenta para suportar a extração, transformação e carregamento do DW a partir dos sistemas operacionais.
- ❑ Pode suportar a execução e monitorização do processo ETL.
- ❑ Alternativa a uma solução de "hand coding" (PL*SQL, shell scripting, etc)

Tools - ETL

- Solução CASE alternativa ao "hand coding"
- Acelerar o processo de programação das transformações
- Potenciar auto-documentação dos processos de transformação
- Assegurar a execução dos planos de Extração-Transformação-Carregamento
- Control e monitorização da execução dos planos

Tools – ETL: Características

■ Características técnicas e funcionais a analisar:

Características	Obrigatório (S/N)	Peso (1-10)
1000 Técnicas		
1010 Integração de dados em formatos: Oracle, Sybase, DB2, AS400 (EBCDIC), Flat File		
1020 Expressões de transformação lógica sobre os dados embebidas no carregamento		
1030 Limpeza de Dados (Regras de Parse, Detecção e Tratamento de Registos Duplicados, ...)		
1040 Agregações incrementais embebidas no carregamento		
1050 Partilha e reuso de Módulos de Transformação		

Tools – ETL: Características

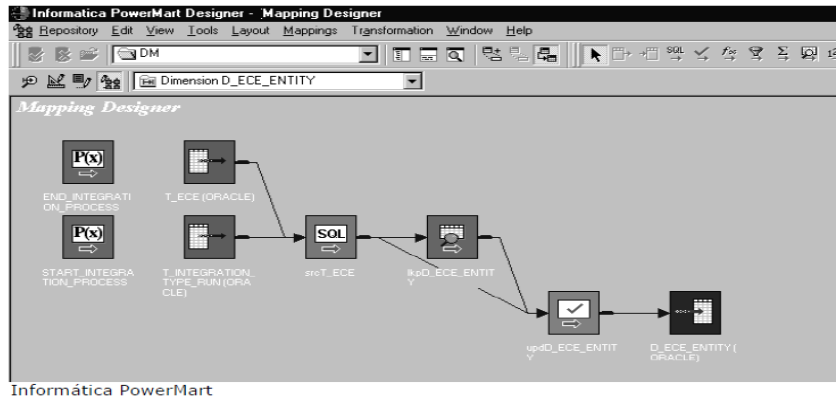
■ Características técnicas e funcionais a analisar:

Características	Obrigatório (S/N)	Peso (1-10)
1000 Técnicas		
1060 Escalonamento de Carregamentos		
1070 Criação e edição do Data Flow		
1080 Aplicação de Filtros no carregamento		
1090 Definição da estratégia de Update a registos existentes		
1100 Acesso a procedimentos externos		
1110 Instâncias de Transformações		
1120 Gestão de Utilizadores e Permissões		
1130 Leitura e Geração de Standards de Metadados		
1140 Monitorização de Sessão		
1150 Utiliza API's do RDBMS para carregamento massivo de dados (em vez de INSERT as SELECT...)		
1160 Paralelização e sincronismo de carregamentos		

Tools – Soluções ETL

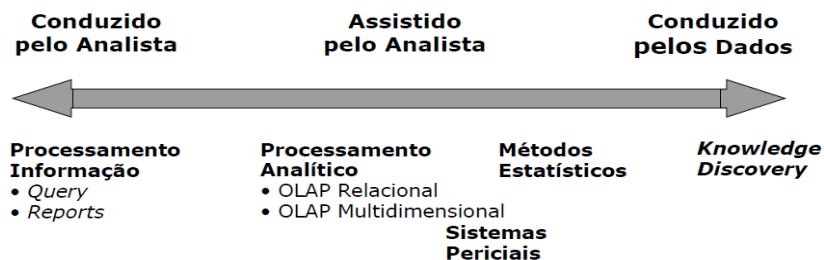
Exemplos:

- Prims Warehouse Executive, Sagent, Informática PowerMart Oracle Data Mart Builder, Oracle Warehouse Builder, ...

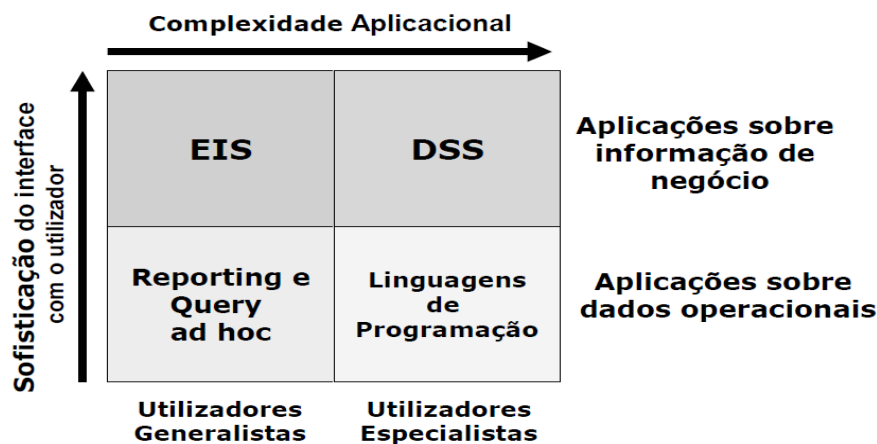


Tools – Condução da Análise de Dados

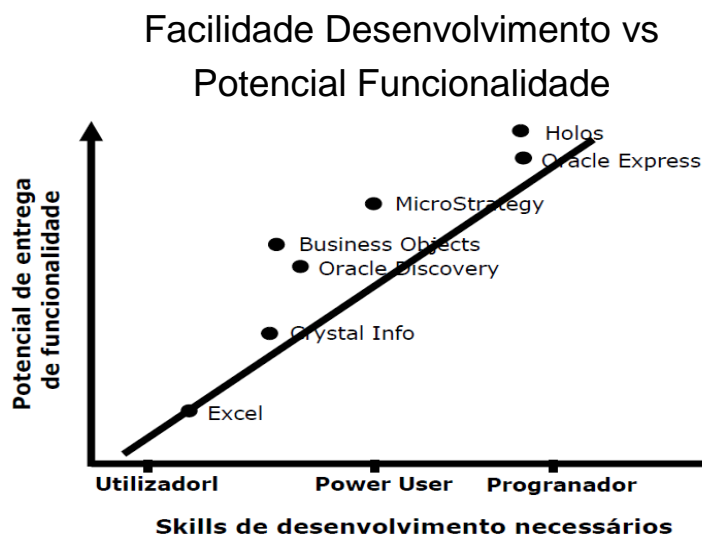
O Papel do Analista de Negócio vs o Processo de Análise dos Dados



Tools - Ferramentas vs Propósitos



Tools - Posicionamento



Tools - Exploração de Dados

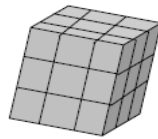
- Constatação de que não existe A ferramenta perfeita e completa
- O Excel transformou-se no instrumento de manipulação de dados com vista à produção de resultados ou conclusões
- O volume de dados introduz uma complexidade na seleção de ferramentas – performance

Tools - Exploração de Dados

- Diferentes ferramentas não estão, normalmente, integradas, mesmo que do mesmo “vendedor”
- Custos de licenciamento elevados e por vezes incompatíveis com a democratização do acesso à informação
- Características Fat e Thin do Servidor e Cliente

Tools: OLAP - MOLAP

- Multidimensional On-Line Analytic Processing



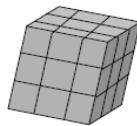
Vista do utilizador

Dados Atômicos e Agregados

- Base de dados Multidimensional (estruturas de dados persistentes)

Tools: OLAP - ROLAP

- Relational On-line Analytical Processing



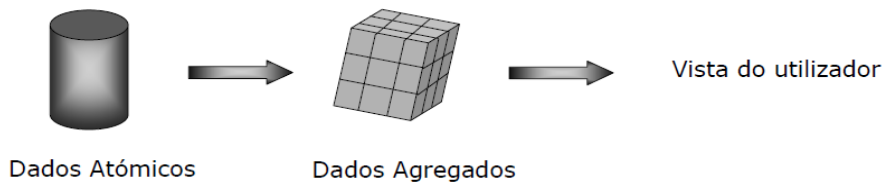
Vista do utilizador

Dados Atômicos e Agregados

- RDBMS + Servidor ROLAP (estruturas de dados voláteis)

Tools: OLAP - HOLAP

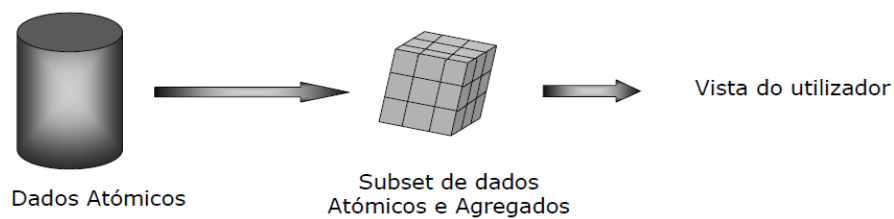
■ Hybrid On-line Analytical Processing



■ Estruturas de dados persistentes e voláteis

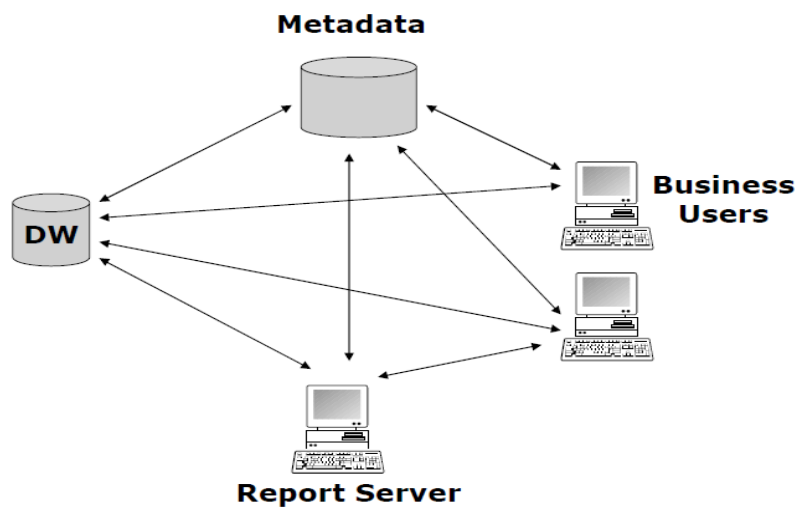
Tools: OLAP - DOLAP

■ Desktop On-line Analytical Processing

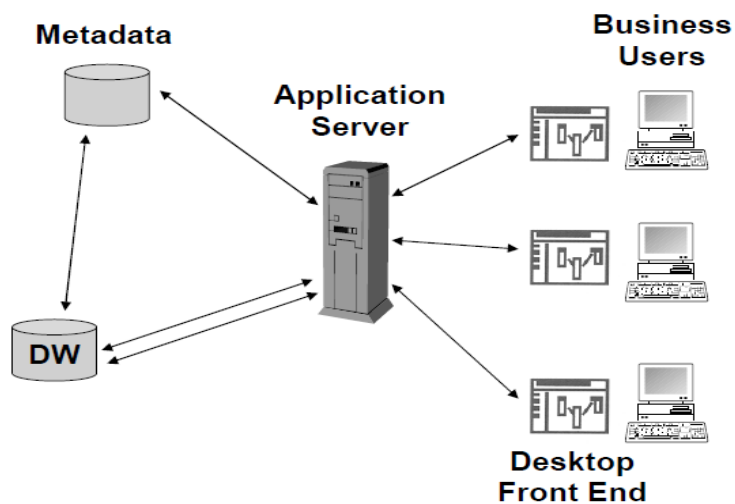


■ Desktop

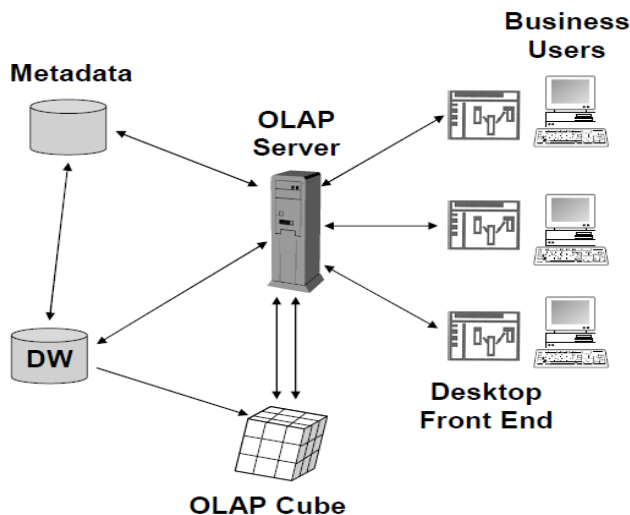
Tools - Arquitetura Two Tier



Tools - Arquitetura Three Tier



Tools - Arquitetura Three+ Tier



Tools - Reporting /Ad-hoc: Características

■ Características técnicas e funcionais a analisar

Características		Obrigatório (S/N)	Peso (1-10)
1000	Técnicas		
	Suportada por ambiente de gestão de execução, agendamento e partilha de relatórios		
1010			
1020	Multipass SQL		
1030	Aggregation Navigator		
	Restrição de execução online baseada em tempo ed execução ou volume de dados a retornar		
1040			
1050	Permite interrupção de execução		
1060	Relatórios baseados em mais do que um query (resultado na mesma página física)		

Tools - Reporting /Ad-hoc: Características

■ Características técnicas e funcionais a analisar

Características	Obrigatório (S/N)	Peso (1-10)
1000 Técnicas		
1070 Tabela e gráfico num mesma página física (formatações complexas)		
1080 Exportação para Excel do dados retornados por cada SQL		
1090 Exportação para Excel do relatório		
1100 Operável a partir de um web browser		
1110 Capacidade de auditar utilização		
1120 Acesso a BD de diferentes fornecedores		
1130 Acesso a dados em ficheiros Excel		
1140 Possibilidade de combinar no mesmo reports dados de diferentes BD e/ou Excel		
1150 Suporta clientes PC e MAC		
1160 Slice-and-dice		

Tools - Reporting /Ad-hoc: Características

■ Características técnicas e funcionais a analisar

Características	Obrigatório (S/N)	Peso (1-10)
1000 Técnicas		
1170 Drill-down		
1180 Modelos de previsão (forecasting)		
1190 complexas		
1200 Mecanismos de segurança		
1210 Capacidades gráficas		
1220 Campos calculados		
1230 Pivoting		
1240 Ordenação		
1250 Alertas (cores)		

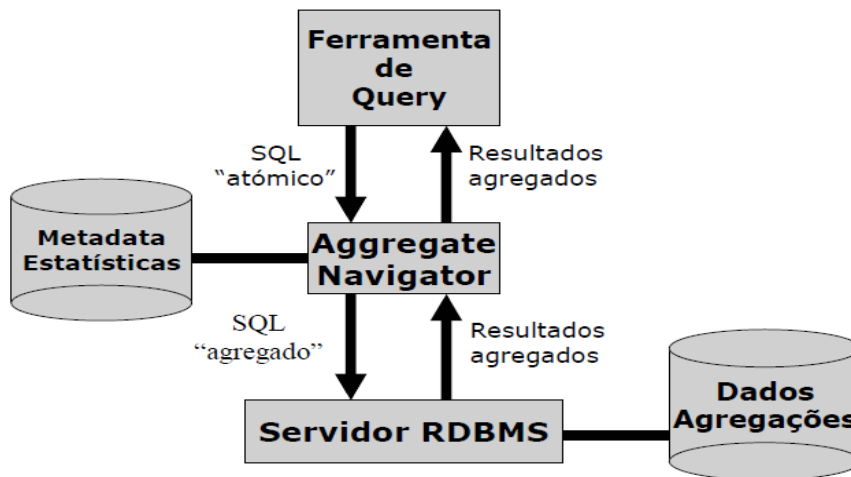
Tools - Reporting /Ad-hoc: Características

- Características/funcionalidades de destaque:
 - Reformulação de queries
 - Redireccionamento de queries
 - “Multipass” SQL
 - “Aggregate awareness”
 - “Query governing”
 - “Date awareness”

Tools - Reformulação de Queries

- Limitações da linguagem de acesso aos dados
SELECT [<função>] atributo, ...
FROM ...
WHERE <condição_lógica>
- GAP entre a pergunta de negócio relevante e a pergunta SQL correspondente
 - “Quanto é que o negócio cresceu no último ano?”
 - “Que contas cresceram mais de 100 %?”
- Capacidade de traduzir uma pergunta de negócio num ou vários SQL complexos

Tools - Aggregate Awareness / ReWrite SQL



Tools – Query Governing

- O query ad hoc é imprevisível !
 - Funcionalidade básica:
 - Limitar número de linhas que um query pode retornar
 - Limitar tempo de execução máximo para um query
 - Em ambos os casos a funcionalidade reage passivamente e não impede que se desperdice o tempo limite
 - Funcionalidade avançada:
 - A validação do plano de execução
 - Usar estatísticas de queries semelhantes

Tools - MOLAP

- A monitorização de alto nível sobre o comportamento do negócio deve ser suportada por instrumentos eficientes
- A transformação dos dados operacionais em informação de negócio introduz forte sumarização (agregações de agregações)
- A representação multidimensional introduz vantagens de performance no acesso aos dados relativamente ao relacional

Tools - MOLAP: Características

■ Interface

- Conjunto completo de funcionalidades disponíveis no interface com o utilizador (todas ou quase todas as disponíveis nas ferramentas de reporting e query ad hoc)
- Capacidade de gerir alertas e gerar gráficos sofisticados
- Sistema de navegação fácil e intuitivo.
Capacidade do utilizador criar as suas próprias “bookmarks”
- Suportar modelos de previsão

Tools - MOLAP: Características

■ Base de Dados Multidimensional

- ❑ 4GL potente para programação do carregamento e atualização dos cubos de dados
- ❑ Meta cubos ou cubos virtuais
- ❑ Algoritmos adequados à dimensão e dos cubos - cubos pequenos podem residir em memória principal

Tools - Exploração de Dados - Exemplos

- Reporting:
 - ❑ CRYSTAL REPORTS AND XCELSIUS SOFTWARE
 - <http://www.sap.com/solutions/sapbusinessobjects/sme/reporting-dashboarding/index.epx>
 - ❑ Ad-Hoc Query (MicroStrategy, Brio, Business Objects, Oracle Discovery, ...)
 - <http://www.microstrategy.com/>
 - <http://businessintelligence.ittoolbox.com/groups/vendor-selection/bi-select/brio-vs-other-bi-solutions-brio-response-128225>
 - <http://www.sap.com/solutions/sapbusinessobjects/index.epx>
 - <http://www.oracle.com/technology/products/discoverer/index.html>
 - ❑ OLAP (MOLAP: Seagate Holos, Oracle Express, ...)

Tools - Exercício Prático

■ Objetivo:

- Apresentação de um protótipo em Business Objects, sobre uma base de dados Access (ou MS/SQL Server, 2008 de um sistema de exploração (reports pré-definidos + ambiente de query ah doc) sobre uma parte do caso em estudo.
 - *Microsoft SQL Server 2008 Business Intelligence*. [Online] Available: <http://www.microsoft.com/sqlserver/2008/en/us/business-intelligence.aspx>
 - *Integration Services*. [Online] Available: <http://msdn.microsoft.com/pt-br/library/ms130214.aspx>
 - *Introdução ao Excel Services e Excel Web Access*. [Online] Available: <http://office.microsoft.com/pt-pt/sharepoint-server-help/introducao-ao-excel-services-e-excel-web-access-HA010105476.aspx>

Tools - Exercício Prático

■ Ferramentas Comuns:

■ Azure SQL Data Warehouse

- <https://azure.microsoft.com/pt-pt/services/sql-data-warehouse/>

■ POWER BI

■ TARGIT

■ SAS Visual Analytics

■ Datometry **DATOMETRY**

- <https://datometry.com/>

