



Curso de Engenharia de Sistemas e Informática - 5º Ano

Ficha T. Prática n.º 4

Objectivo:

Investigação sobre OLAP, suas abordagens e ferramentas disponíveis.

Proposta de Trabalho

Este caso debruça-se sobre o paper “*Providing OLAP (On-Line Analytical Processing) to User-Analysts: An IT Mandate*”, onde E.F. Codd & Associates lançam as bases do OLAP. Este paper pode ser consultado no ficheiro fornecido em anexo “codd.pdf” e também disponível em on-line na página da disciplina em http://web.estv.ipv.pt/jloureiro/ESI_AID2006_2007/fichas/codd.pdf, recomendando-se uma leitura genérica, ainda que aligeirada, só focando nos excertos abaixo apresentados, dele retirados.

Excerto 1 (pág. 4). Logo no início, mostra-se a evolução nos sistemas de bases de dados, salientando as virtudes do modelo relacional (não se tratasse do “pai” do modelo relacional), mas também as suas fronteiras relativas à função de disponibilização de vistas sobre os dados.

Tecendo algumas considerações sobre este último assunto, mostra a necessidade para o OLAP, dada a incapacidade quer das RDBMS de modo directo ou de muitas das ferramentas de Query / Reporting de “proporcionar a consolidação, análise e visualização dos dados de acordo com múltiplas dimensões, sob formas que façam sentido aos analistas de negócio. Este requisito é chamado de análise de dados multidimensional”.

Excerto 2 (pág. 5). Mais à frente, Codd acrescenta “tentar forçar uma tecnologia ou ferramenta a satisfazer uma necessidade particular para a qual outra ferramenta é mais efectiva e eficiente é como tentar meter um parafuso numa parede com um martelo, tendo uma chave de parafusos à mão: o parafuso poderá eventualmente entrar na parede, mas a que custos?”

Excerto 3 (pág. 6). Acrescentando ainda, a este propósito “As bases de dados relacionais têm sido, são hoje e continuarão a ser a tecnologia mais apropriada para as bases de dados das empresas. A necessidade que existe não é para outra tecnologia de base de dados, mas para ferramentas de OLAP robustas para análise de dados nas empresas que complementem os sistemas de gestão de base de dados das empresas já existentes e que sejam suficientemente exactos para antecipar e facilitar os tipos de análises de dados de negócio inerentes em OLAP.”

Excerto 4 (pág. 6). “As actividades dessas ferramentas analíticas sobre a base de dados constitui uma transacção. Contudo, a duração dessa transacção pode ser muito mais prolongadas do que a sua contrapartida tradicional em OLTP (dias ou semanas versus

segundos ou talvez minutos). Outra diferença entre estes tipos de transacções é que as aplicações OLTP necessitam de trabalhar com dados actuais (instantâneos) exactos, enquanto que o que os autores chamam “OLAP” trabalham quase exclusivamente com dados históricos exactos num dado ponto no tempo, especificamente o início da transacção OLAP.”.

Excerto 5 (pág. 6). Mais à frente refere “OLAP é feito de muito cenários, numerosos e especulativos do tipo “what-if” e/ou “why” executados dentro do contexto de alguma base e perspectiva histórica específica.”

Excerto 6 (pág. 6). E ainda “A consolidação de dados é o processo de sintetizar bocados de informação em blocos simples de conhecimento essencial. O nível mais alto num caminho de consolidação de dados é referido como essa dimensão de dados. Uma dada dimensão de dados representa uma perspectiva específica dos dados incluídos no seu caminho de consolidação associado. Há tipicamente um número de diferentes dimensões a partir das quais um dado conjunto de dados pode ser analisados. Esta perspectiva plural, ou visão conceptual multidimensional parece ser a forma como a maioria dos analistas de negócio vêm naturalmente a sua empresa. Cada uma dessas perspectivas é considerada como uma dimensão de dados complementar. A análise simultânea de dimensões múltiplas de dados é referida como análise de dados multidimensional”.

Excerto 7 (pág. 8). “Uma vez os dados tenham sido consolidados de acordo com um ou mais caminhos de consolidação, ‘perfurar’ em direcção a maior nível de detalhe é possível. *drilling down* refere-se ao movimento de maior para menores níveis de consolidação. Inversamente *rolling up* refere-se ao movimento de menor para maior nível de consolidação.

Os caminhos de consolidação determinam que níveis de detalhe são visíveis aos utilizadores finais quando fazem drill down. “

Excerto 8 (pág. 12). Finalmente, ao referir uma abordagem complementar às bases de dados relacionais através de ferramentas implementando análise multidimensional, introduz as famosas doze regras de avaliação de produtos OLAP, que são:

1. visão conceptual multidimensional;
2. transparência;
3. acessibilidade;
4. desempenho em consultas consistente;
5. arquitectura cliente / servidor;
6. dimensionalidade genérica;
7. manuseamento dinâmico de matrizes dispersas;
8. suporte multiutilizador;
9. operações dimensionais não restringidas;
10. manipulação de dados intuitiva;
11. emissão de relatórios flexível;
12. níveis ilimitados de dimensões e agregação.

1ª Parte

1. No primeiro excerto do texto, referem-se as virtudes das bases de dados relacionais, mas também as suas fronteiras. Concorda com as virtudes e fronteiras definidas? Fundamente a sua opinião.
2. Quando se fala, no 2º excerto, em tentar forçar uma tecnologia ou ferramenta a satisfazer

uma necessidade particular para a qual outra ferramenta é mais efectiva. A que se refere? Concorda com os autores do artigo? Justifique.

3. Comente a perspectiva defendida no 3º excerto.
4. Mostre a diferença entre uma transacção OLTP e OLAP, focada no 4º excerto, mostrando quais as principais implicações que isso poderá ter na própria tecnologia da ferramenta e na das tecnologias de suporte.
5. No excerto 5 fala-se de análises what-if e why e perspectivas históricas específicas. Comente o seu inter-relacionamento. A extrapolação para o futuro será aqui antevista? Justifique.
6. Mostre o conceito de consolidação de dados e caminhos de consolidação, focando na visão nativa dos dados de um analista de negócio.
7. Dê a sua definição fundamentada de dimensão de dados.
8. No excerto 7 fala-se de caminhos de consolidação e de operações de drill down e roll. Mostre como estão inter-relacionados.
9. Comente o último parágrafo do penúltimo excerto.

2ª Parte

Já no excerto 8 indicam-se as famosas 12 regras de avaliação de produtos OLAP, à imagem das 12 regras relativas a bases de dados relacionais, anteriormente formuladas pelo mesmo autor.

1. Comente, à luz dos conhecimento apreendidos quanto às diversas abordagens tecnológicas em OLAP: MOLAP, ROLAP; RAP e HOLAP, o seu suporte efectivo quanto a cada uma dessas regras.
2. Lançando mão do mundo www e sua informação, cite, localize e refira algumas características mais relevantes de alguns produtos disponíveis no mercado.
3. Nas pesquisas feitas para responder à questão anterior, deu decerto conta de alguma tendência recente de ligação de ferramentas analíticas com aplicações CRM e muitas no domínio do data mining. Cite exemplos e tente dar uma justificação para ao facto.