
TPC-DS Benchmark: como avaliar os sistemas de apoio a decisão da sua empresa?

Proponentes: Jorge Bernardino, Pedro Furtado

Durante a última década, os investigadores e a indústria na área de bases de dados usaram o benchmark TPC-H para avaliar o desempenho da tecnologia na área de apoio à decisão. Reconhecendo a mudança de paradigma na indústria e a necessidade de uma referência padrão para medir o desempenho dos sistemas DSS (Decision Support Systems) o Transaction Processing Performance Council desenvolveu recentemente um novo benchmark para apoio à decisão, o TPC-DS (TPC-Decision Support).

As diferenças entre os sistemas de apoio à decisão de hoje e a especificação benchmark TPC-H são múltiplas. O esquema TPC-H, que, embora suficientemente complexo para testar os primeiros sistemas, não é representativo das implementações de apoio de decisão mais complexas de hoje. Os esquemas usados atualmente são tipicamente compostas por um grande número de tabelas e colunas. Além disso, a escolha de implementação na indústria mudou de esquemas 3NF puros para variações do esquema em estrela, como flocos de neve (snowflakes).

O TPC-DS pode ser aplicado a qualquer indústria que pretenda transformar os seus dados operacionais e externos em business intelligence. Ele modela as tarefas de apoio à decisão de um típico fornecedor de vendas a retalho.

O benchmark abstrai a diversidade de operações encontradas usualmente em aplicações de análise de informação, mantendo as características essenciais de desempenho. Como é necessário executar um grande número de consultas e transformações de dados para gerir completamente qualquer ambiente de análise de negócios, o TPC-DS define 99 consultas distintas e 12 operações de manutenção de dados cobrindo DSS típicos como os tipos de consulta, como ad-hoc, relatórios iterativos (drill down / drill up) e consultas de extração e atualização periódica da base de dados.

Neste mini-projecto pretende-se implementar o benchmark TPC-DS de forma a que qualquer investigador ou utilizador interessado possa usá-lo definindo apenas os seus principais parâmetros e configurações. Deste modo, isto evita o enorme esforço de adaptação ao ambiente de trabalho, sistema operativo, base de dados a testar, dimensão da base de dados, transformação do SQL, etc.

Este trabalho também será de grande utilidade para os estudantes de mestrado, principalmente de Inteligência no Negócio, Sistemas de Gestão de Dados, Computação de Alto Desempenho e para os investigadores do CISUC, devendo ser disponibilizado de forma livre e gratuita na Web.

1. Estudo da benchmark TPC-DS (dias 0 a 15)
2. Transformação das 99 pesquisas SQL e das 12 operações de manutenção para os diferentes ambientes de teste.
3. Propostas de implementações (dias 24 a 50)
4. Elaboração e avaliação do protótipo (dias 51 a 90)
5. Escrita do relatório (dias 75 a 90)

Referências bibliográficas:

- Lin Shi, Bin Huang, Hao Xu, Xiao-jun Ye: Implementation of TPC-DS Testing Tool. DBTA 2010:1-4
- Meikel Pöss, Raghunath Othayoth Nambiar, David Walrath: Why You Should Run TPC-DS: A Workload Analysis. VLDB 2007:1138-1149
- Raghunath Othayoth Nambiar, Meikel Poes: The Making of TPC-DS. VLDB 2006:1049-1058
- TPC-DS Standard Specification Version 1.1.0, April 2012, <http://www.tpc.org/tpcds/tpcds.asp>