# **Tarefa #1 Estudo sobre índices**

Nessa tarefa vocês irão trabalhar com o benchmark TPC-H. O TPC é uma organização sem fins lucrativos, tendo como objetivo principal estabelecer critérios para se obter informações a respeito do desempenho de processamento de transações e de banco de dados por meio de benchmarks. Baseia-se em testes padronizados tais como o TPC-C, o TPC-W e o TPC-H para obter tais resultados e só assim divulgar os dados reais desse desempenho. Estes testes do TPC seguem normas extremamente rigorosas que testam diversos pontos do sistema - principalmente nos quesitos confiabilidade e durabilidade.

O benchmark TPC-H que utilizaremos nessa tarefa simulam um ambiente de Data Warehouse, sincronizado com bancos de dados de produção online. Ele se utiliza de consultas aleatórias de alto grau de complexidade, projetadas para responder a algumas questões de negócios do mundo real, como: perguntas sobre pré e promoções, oferta e demanda, lucro e receita e participação de mercado.

O schema em mysql se encontra anexo a esta tarefa, porém, quem quiser fazer "do zero" para aprender, basta seguir as instruções abaixo.

Como configurar o TPC-H

1) Download TPC-H:

<http://www.tpc.org>

2) Descompactar o tpc(versão).zip

3) Entrar na pasta do tpc $ cd /home/"nome\_da\_maquina"/tpc

4) Criar o makefile. $ cp makefile.suite Makefile

5) Editar o Makefile de acordo como abaixo:

$ vi Makefile

CC = gcc

DATABASE = MYSQL

MACHINE = ICL

WORKLOAD = TPCH

Sair do vi com o comando:

$ :wq!

6) Editar o arquivo tpcd.h

$ vi tpcd.h

ifdef MYSQL

define GEN\_QUERY\_PLAN "SET CURRENT EXPLAIN SNAPSHOT ON;"

define START\_TRAN "START TRANSACTION;"

define END\_TRAN "COMMIT;"

define SET\_OUTPUT ""

define SET\_ROWCOUNT "--

SET ROWS\_FETCH %d\n"

define SET\_DBASE "CONNECT TO %s ;\n"

endif

Salvar e sair:

$ :wq!

7) Instalar o GCC no yast

8) Realizar a compilação do sistema.

$ Make

9) Criar pasta data:

$ mkdir /var/data

10) Gerar base de 1gb.

$ cp /home/nome\_da\_maquina/tpc/dists.dss /var/data

$ cd /var/data

$ ~nome\_da\_maquina/tpc/dbgen -v -s 1

Apos concluir a criação da base de dados entre novamente na pasta /var e crie a pasta "Consultas".

11) Gerar consultas:

$ cd ..

$ mkdir consultas

$ ~nome\_da\_maquina/tpc/qgen > /var/data/consultas/all.sql

Cada grupo deve criar o esquema do TPC-H em um dos seguintes SGBDs:

a. mySQL -<https://www.mysql.com/>

b. Oracle -<http://www.oracle.com/index.html>

c. SQL Server - [www.microsoft.com/SQL\_Server](http://www.microsoft.com/SQL_Server)

d. PostgreSQL - [www.postgresql.org/](http://www.postgresql.org/)

O TPC-H nativo não possui índices para acelerar as consultas. A sua tarefa será propor uma série de índices para melhorar o desempenho das consultas do TPC-H. Cada grupo deve instanciar dois bancos no SGBD escolhido: um deles com o uso de índices e o outro sem o uso de indices (como vimos na última aula).

Cada um dos bancos deve ser populado usando a ferramenta DBGen (vide links a seguir ou usar o schema já anexo a tarefa) e as consultas do TPC-H devem ser executadas em cada um dos bancos (as consultas se encontram nos zips carregados nesse post).

Cada grupo deve entregar os dois esquemas (com e sem índice) e um relatório comparativo do desempenho das consultas em cada um dos bancos. Nesse relatório devem ser apresentadas as quantidades de registros do BD, os parâmetros utilizados na ferramenta DBGen para geração dos dados, o tempo de execução da consulta e uma explicação sobre a diferença de desempenho das mesmas.

Att.,

Daniel

Links úteis:

<http://kejser.org/tpc-h-data-and-query-generation/>

<http://planet.mysql.com/entry/?id=27807>

<http://www.tpc.org/tpch/>