

TPC H As Linked Data

Bringing Linked Data to Analytics

Table of Contents

- [Introduction](#)
- [Mapping](#)
- [Queries](#)
- [Dereferenceable URIs](#)
- [Running the Samples](#)
- [Running the Samples Against External Data](#)

Introduction

This article takes the industry standard TPC-H decision support benchmark and shows how its data can be exposed as linked data from any relational database.

Virtuoso can either host the data itself or attach the relevant tables from any other database. Here we will look at a miniature sample of the TPC-H benchmark database that comes with the demo database.

Exposing this data consists of a mapping declaration which translates the contents of the tables into a virtual RDF graph. The triples involved are not physically stored but are created on demand as requested by SPARQL queries or dereferencing the URIs representing the logical entities in the database.

Mapping

See the attached SQL file:

[tpch.sql](#)

Queries

See the attached SQL file:

[tpch.sql](#)

As of this writing, not all the queries are supported. Thus you will see some of them commented-out. The relevant optimizations and SPARQL extensions for running these are being developed at the time of this writing.

Dereferenceable URIs

See the attached SQL file:

[tpch.sql](#)

This file defines a virtual directory and URI-rewrite rules that capture references to the TPC-H entities. The URL rewrite rules in this file catch references to the the TPC-H entities and redirect these into a SPARQL describe queries on the URI in question. The redirect makes an HTTP 303 redirection reply that goes to the SPARQL endpoint with a `describe` query for the object in question.

Running the Samples

There is a live sample of this demo running at

<http://demo.openlinksw.com/DAV/home/demo/tpch/>

Virtuoso Open Source Edition will come with an implementation of the TPC-H data and queries from version 5.0.6 onwards.

Start the demo server and connect to <http://localhost:8890/DAV/home/demo/tpch> with a web browser.

The host part of the URIs given to the TPC-H data is given by the DefaultHost entry in the URIQA section of the virtuoso.ini file.

Running the Samples Against External Data

With the commercial release of Virtuoso, the tables do not have to reside in Virtuoso itself. They can be attached from any supported database using the Virtual Database layer (VDB). These can then be mapped into a graph of their own by re-running the mapping statements with the graph and tables renamed.

[CategoryWebSite](#) [CategoryVirtuoso](#) [CategoryOpenSource](#)