TAP and ADQL

Hendrik Heinl *(heinl@ari.uni-heidelberg.de)*

GAVO / ARI

Agenda

- Why bother?

- A first query

- ADQL

T(able) A(ccess)
P(rotocol)

A(stronomical) D(ata)
Q(uery) L(anguage)

Data Intensive Science

Data-intensive science means:

1. Using many data collections

2. Using large data collections

Point (1) requires standard formats and access protocols to the data, point (2) means moving the data to your box and operating on it with FORTRAN and grep becomes infeasible.

The Virtual Observatory (VO) in general is about solving problem (1), TAP/ADQL in particular about (2).

TAP

TAP services

- provide database interfaces

- "understand" ADQL

- accept uploaded tables for computing (code to data)

- return table data (tuples of tuples)

ADQL

- Designed to fit to the needs of astronomers

- First recommended as ADQL 1.0 in 2005, based on SQL

- Since 2008 recommended as ADQL 2.0

- ADQL 2.1 is expected to be recommended in 2017

- Implemented on 104 services (SIMBAD, ESA-Archive) and

- several Clients, most popular TOPCAT and STILTS

(You may want to have a look at the most recent developments in PyVO.)

Why SQL?

The SELECT statement is written in ADQL, a dialect of SQL ("sequel"). Such queries make up quite a bit of the science within the VO.

SQL has been chosen as a base because

- Solid theory behind it (relational algebra)

- Lots of high-quality engines available

- Not Turing-complete, i.e., automated reasoning on "programs" is not very hard

Relational Algebra

At the basis of relational data bases is the relational algebra, an algebra on sets of tuples ("relations") defining six operators:

- unary *select*

- unary *project*

- unary *rename*

- binary *cartesian product*

- binary *union*

- binary *set difference*

Good News: You don't *need* to know any of this.

Why not plain SQL?

Lack of

- geometries and regions

- some useful mathematical functions

- Xmatching on spherical positions

SELECT for real

ADQL defines just one statement, the SELECT statement, which lets you write down expressions of relational algebra. Roughly, it looks like this:

SELECT [TOP *setLimit*] *selectList* FROM *fromClause* [WHERE *conditions*] [GROUP BY *columns*] [ORDER BY *columns*]

TOP

setLimit: just an integer giving how many rows you want returned.

▷ 1   SELECT TOP 5 * FROM tgas_source

A First Query

You don't have to follow the example, there will be plenty in the hands on sessions.

What the heck was that ?

If the former was beyond your ADQL skills, you may be a good candidate to attend the hands on tutorial introductiion to ADQL after lunch in room D52. Who already is familiar with ADQL may still benefit from the scientific use cases presented in room B3/B5

Topcat installation session at 2:30pm in room D52.

Use it for accessing Gaia data

... but also on other TAP services out there. Get involved by reporting bugs and demanding new functions.

For news and updates fee free to register at:

http://mail.ivoa.net/mailman/listinfo/dal

Gaia Data Center at ARI:

http://gaia.ari.uni-heidelberg.de/tap

Thank you for your attention.