As If You Were There
Remote Queries in the Virtual Observatory
Markus Demleitner (msdemlei@ari.uni-heidelberg.de)

Agenda

• Locating data
• Sending queries
• Orchestrating servers
The Problem

Data

Data

Data

Data

User

Data

Data

Data

Data

?
The VO’s solution: TAP

The Table Access Protocol transmits (ADQL) queries and ancillary data to servers, and lets you manage jobs and results, and provides for some service introspection.

TAP’s typical pattern:

1. Define a job (set the query, add uploads, change limits)
2. Start it
3. Check progress (or not, if you’re not curious)
4. When you see it’s done, retrieve or use result
5. Clean up (if you’re nice)
Async? Show me.

In case there’s no net, a screenshot:

tapsh> select * from ppmx.data where cmag between 10 and 11 and deltaFloat>80

tapsh> start

tapsh> job

Job nicked zowize (remote id o8JfWP)

phase EXECUTING, destruction due 2010-09-05T09:52:02, time limit 3600

---------------------------------------------------------------

select * from ppmx.data where cmag between 10 and 11 and deltaFloat>80

tapsh> job

Job nicked zowize (remote id o8JfWP)

phase EXECUTING, destruction due 2010-09-05T09:52:02, time limit 3600

---------------------------------------------------------------

select * from ppmx.data where cmag between 10 and 11 and deltaFloat>80

tapsh> job

Job nicked zowize (remote id o8JfWP)

phase COMPLETED, destruction due 2010-09-05T09:52:02, time limit 3600

---------------------------------------------------------------

select * from ppmx.data where cmag between 10 and 11 and deltaFloat>80

tapsh>
What’s out there?

To find out what is already on TAP, you could ask the registry, use VO-Desktop, or query each server’s TAP_SCHEMA yourself. Or you could just use ⟨GloTS⟩.
Running Queries

Though many services will let your web browser operate a TAP services more or less crudely, you definitely want specialized TAP client software:

- ⟨VODesktop⟩
- ⟨TAP shell⟩
- ⟨tapquery⟩ – a python library that’s part of GAVO’s python votable package

- more soon, or so we hope...

Here, I’m running the TAP shell.
A quick example

Let’s look for really red objects close to quasars in the redshift range between 0.1 and 0.12.

**Parameters**

- Column words: redshift
- Table words: quasar

**Result**

Matched: 4

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
<th>UCD</th>
<th>Unit</th>
<th>Index?</th>
<th>Table description</th>
<th>Table name</th>
<th>id</th>
<th>Src</th>
</tr>
</thead>
<tbody>
<tr>
<td>z</td>
<td>Redshift</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>The catalog of all confirmed quasars in SDSS Data Release 3.</td>
<td>unknown_DR3QuasarCatalog</td>
<td></td>
<td>wfab.roc.ac.uk/sdssdr7-dsa</td>
</tr>
</tbody>
</table>
Query WFAU

That’s almost enough to let us query the service. In tapsh, using tab-completion, say:

```plaintext
server ivo://wfau.roe.ac.uk/sdssdr7-dsa
select * from DR5QuasarCatalog as t where t.z between 0.1 and 0.12
run
send to topcat
```
Now find a table with the UCD mag.phot;em.IR.K. Let’s use PPMXL. We need to upload the last result; therefore, we give it a fixed nick.

nick quas
server ivo://org.gavo.dc/__/system__/tap/run
select * from TAP_UPLOAD.uploaded as u join ppmxl.main as p on (1=contains(point('ICRS', p.raj2000, p.dej2000), circle('ICRS', ra, dec, 0.5))) where kmag-jmag>0.5
upload result quas as uploaded
run
send to topcat
What next?

- Clean up: `echo "purge all" | tapsh`
- ⟨Download tapsh⟩ (and, of course, ⟨topcat⟩); see http://www.g-vo.org
- Complain to whoever made your favourite data set if it’s not yet on TAP; tell them to talk to us if they don’t know how to get it there.
- Report bugs, problems, and ideas.
- Be almost there!