TOPCAT and TAP/ADQL

Mark Taylor (Bristol)

ASTERICS Tech Forum #1

14 September 2015

$Id: tapcat.tex,v 1.7 2015/09/16 22:12:12 mbt Exp $

Outline

- TAP Refresher
- TOPCAT recent TAP enhancements
- TAP status

TAP/ADQL Overview

- TAP = Table Access Protocol
- ADQL = Astronomical Data Query Language
- TAP with ADQL lets you run SQL-like queries on remote databases
- Many standards involved for a TAP/ADQL client:
  - service discovery: RegTAP, VOResource
  - service/table metadata retrieval: VODataService, VOResource, TAPRegExt, DALI, VOTable, TAP
  - standard data models: ObsCore, EPNCore, RegTAP, TAP
  - query submission/execution: ADQL, DALI, VOTable, TAP
  - asynchronous job management: UWS
  - result retrieval: VOTable, TAP
- Non-ADQL TAP query languages are, in principle, available

Why?

We've had CasJobs for years, so why TAP?

- So the same client code can talk to different services
- Which makes life easier for data consumers:
  - Humans can learn one tool and use it to make complex queries on many different databases
  - Software tools can execute prepared queries on different remote datasets without custom code for each one

TAP Implementation Status

Clients

- Generic TAP clients
  - TAPHandle (CDS)
  - tapsh (ARI)
  - Seleste (SAO)
  - TOPCAT/STILTS (Bristol)
  - ... others?
- DM-specific clients
  - SPLAT (ObsTAP)
  - VESPA (EPN-TAP)
  - ... more?
- Service-specific clients
  - DaCHS, GACS, TAPVizieR, Simbad, DSA, ...

Services:

- 82 registered TAP services (Sept 2015), plus several unregistered
- Service toolkits: DaCHS (GAVO), TAP Lib (CDS/ARI), OpenCADC (CADC), Saada (CDS), DSA, IA2TAP, ...

TOPCAT TAP

TOPCAT TAP client history:

- Mar 2010: TAP 1.0 IVOA RECcomendation
- May 2011: TOPCAT v3.8, includes TAP client (thanks Markus/GAVO)
  - TAP working, but poor service discovery, not suitable for services with very many tables, query editing limited
- Aug 2015: TOPCAT v4.3, major TAP client overhaul
  - Better!

Availability

- GUI client present in TOPCAT application
- Command-line capability available in STILTS application
- Full GUI client or lower-level TAP libraries also available for standalone/embedded use

TOPCAT v4.3 TAP Client (1)

Improvements to TAP client in recent release:

- Service discovery
  - Can locate services by table, not just service, metadata
  - Uses TAP/GLoTS not RI1.0 - much faster and more reliable
- Metadata retrieval
  - (Optionally) multi-step metadata acquisition: now scalable to VizieR-size services
- Metadata display
  - More service metadata displayed (DMs, UDFs, VOResource description)
  - GUI now scales to VizieR-size services (JTree + instant filtering)
- Examples
  - Better example browse/display (but could still be improved?)
  - More canned examples listed, including DM-specific ones
  - Service-provided examples
- Query editing
  - Multi-tab query editing
  - Undo/Redo in query editing panel
  - Better ADQL syntax highlighting (bug fixes + UDF-awareness) — thanks Grégory

TOPCAT v4.3 TAP Client (2)

Improvements to TAP client in recent release ... continued

- Configurability options (TAP menu)
  - Metadata acquisition options (TAP_SCHEMA + variants, TableSet + variants)
  - Response VOTable serialization format (TABLEDATA, BINARY, BINARY2, default)
  - Upload VOTable serialization format (TABLEDATA, BINARY, BINARY2)
  - Service discovery mechanism (GLoTS, IVOA Registry)
  - HTTP-level compression (gzip, none)
- Diagnostics
  - Better error reporting in case of (incorrect) non-VOTable errors
  - Logging of `curl(1)` equivalent for TAP queries

Demo

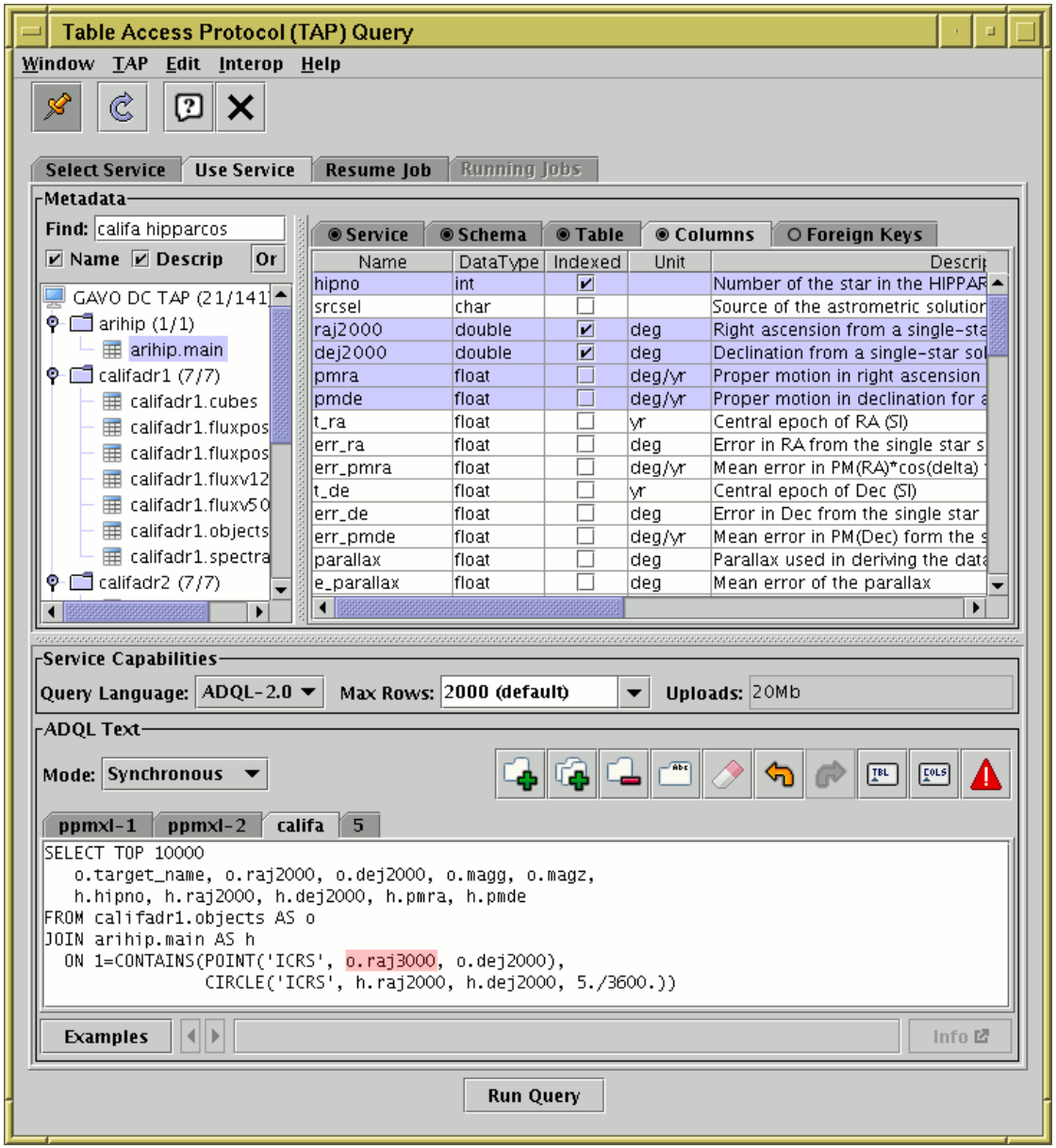
Table Access Protocol (TAP) Query
Window TAP Edit Interop Help
Select Service
Use Service
Resume Job
Running Jobs
Metadata
Find: califa hipparcos
Name Descrip Or
GAVO DC TAP (21/141)
arihip (1/1)
arihip.main
califadr1 (7/7)
califadr1.cubes
califadr1.fluxpos
califadr1.fluxpos
califadr1.fluxv12
califadr1.fluxv50
califadr1.objects
califadr1.spectra
califadr2 (7/7)
Service Schema Table Columns Foreign Keys
Name DataType Indexed Unit Descrip
hipno int Number of the star in the HIPPAR
srcsel char Source of the astrometric solution
raj2000 double deg Right ascension from a single-sta
dej2000 double deg Declination from a single-star sol
pmra float deg/yr Proper motion in right ascension
pmde float deg/yr Proper motion in declination for a
t_ra float yr Central epoch of RA (SI)
err_ra float deg Error in RA from the single star s
err_pmra float deg/yr Mean error in PM(RA)*cos(delta)
t_de float yr Central epoch of Dec (SI)
err_de float deg Error in Dec from the single star
err_pmde float deg/yr Mean error in PM(Dec) form the s
parallax float deg Parallax used in deriving the data
e_parallax float deg Mean error of the parallax
Service Capabilities
Query Language: ADQL-2.0
Max Rows: 2000 (default)
Uploads: 20Mb
ADQL Text
Mode: Synchronous
ppmxl-1 ppmxl-2 califa 5
SELECT TOP 10000
o.target_name, o.raj2000, o.dej2000, o.magg, o.magz,
h.hipno, h.raj2000, h.dej2000, h.pmra, h.pmde
FROM califadr1.objects AS o
JOIN arihip.main AS h
ON 1=CONTAINS(POINT('ICRS', o.raj3000, o.dej2000),
CIRCLE('ICRS', h.raj2000, h.dej2000, 5./3600.))
Examples
Info
Run Query

TAP Status

- Since TAP v1.0 REC (2010):
  - More/better clients
  - Better services (server library developments, validator availability, experience with protocol)
  - More services registered
  - Enhancements to standards (RegTAP, TAPRegExt, /examples endpoint, more on the way)
- Usage
  - I am now approaching *"Not embarrassed"* to show TOPCAT/TAP to astronomers
    - ▷ you don't need to be a VO expert to use it any more
    - ▷ you do need to have a slight understanding of SQL
    - ▷ post-SDSS, many astronomers have at least a rough idea, but generally need/like a bit of help *(examples!)*
  - Astronomers can/should be using TAP to do science now
    - ▷ for single-archive queries
    - ▷ for multi-archive data integration(?)

      — are they?

What's Missing? (1)

Service-provided Examples

- Clarify examples document format
  - Currently there are two competing variants proposed (DALI and TAPNote) for the RDFa format of the `/examples` endpoint. DAL WG needs to back one.
- Encourage services to provide examples
  - Examples is a recent addition to TAP protocol suite; most service providers don't know it exists
- Provide validator tool for service-provided examples
  - Action: me (`taplint` enhancement required)
- Better example presentation in clients?
  - TOPCAT presentation could be improved
  - Not currently displayed in other clients?

```
<html>
  <div typeof="example" id="basic-select" resource="#basic-select">
    <p>This is a simple query on the PhotoObj table:</p>
    <pre property="query">
      SELECT TOP 10 ra, dec, u, g, r, i, z FROM sdss.PhotoObj
    </pre>
  </div>
</html>
```

What's Missing? (2)

Consistent service capabilities

- Cross-match syntax

  "Standard" ADQL positional xmatch is of form

  ```
  JOIN ON 1=CONTAINS( POINT('ICRS', t1.ra, t1.dec),
                      CIRCLE('ICRS', t2.ra, t2.dec, radius_in_deg))
  ```

  - ▷ Geom functions are optional — not all services support this
  - ▷ Some services support it but with very poor performance (not indexed?)
    $\longrightarrow$ it's hard for clients:
    - ▷ to perform programmatic crossmatches
    - ▷ to help users write crossmatch queries
  - ▷ (It's also ugly and hard to remember)

- Upload capability
  - ▷ some services do, some don't

Together, these make multi-archive data integration hard to do;
is that a problem to be fixed, or just a fact of life?

What's Missing? (3)

Correctness and compliance

- Many services still broken/unreliable/slow

User education/publicity

- Tutorials, worked examples of TAP-based science *(hackathon?)*

Better service registration

- Public services slow to arrive in Registry $\longrightarrow$ GLoTS?

Unresolved technical issues (less urgent)

- Better RegTAP-based service discovery
  - ▷ Currently, not enough metadata is stored in the IVOA Registry to support table-based service discovery. TOPCAT bypasses this by using (non-standard) GLoTS service at Heidelberg. Registry WG are working on this.
- Scalable metadata acquisition
  - ▷ There are two competing proposals for multi-stage table metadata retrieval; CADC and VizieR. DAL WG needs to back one.

... plus plenty more items in TAP Implementation Notes 1.0

Summary

TAP

- Powerful ...
- ... but complicated
- Client software (TOPCAT and others) can make it usable by astronomers. I hope.
- Examples help; encourage service providers to supply them
- Some issues with consistent service capabilities remain
- Some loose ends to be tied up by IVOA