TAP implementation in VizieR

Gilles Landais (CDS)

9 mai 2012

Sommaire

The VizieR database characteristics

To manage the VizieR volumetry

- **METAdata** : ∼10.000 catalogs, ∼20.000 tables and ∼300.000 columns
- **Big catalogs** : 2MASS(∼400G), GSC2.3(∼1T), ...

To manage the VizieR data

- different kinds of storage : database (Sybase or PostgreSQL), binary files (2 formats)
- data stored in an adapted database type

To manage the heterogeneous coordinates systems

- **Coordinate system, equinox, epoch depends of the catalog**
- VizieR compute positions with taking in account equinox, epoch and **propper motions**

The technology used

Storage system	PostgreSQL database (size ∼4Tb)
Positions indexation	H3C (healpix index using the NASA library)
Parser/ADQL translator	Java library (G.Mantelet)
TAP	Java library (G.Mantelet)
Convert coordinate system	AS4 (F.Ochsenbein)

The Java ADQL/TAP implementation

An helpfull library to :

- include VizieR METAdata into the ADQL tree
- Adaptation to Q3C/H3C functions
- Adaptation to the AS4 convert functions
- Verify the ADQL consistency with the VizieR data (add warnings..)
- Computation of columns in the adapted storage with taking in account the precision.
- Optimization depending of the Q3C/H3C library (reorder functions depending of the tables size for join usage)

The METAdata

- ∼20.000 tables, ∼300.000 columns
- XML describing the TAP_SCHEMA (provided by the TAP service)
  XML size which contains the tables+columns name only ∼25Mb
  XML size for the complete description ∼80Mb
- To decrease the output volume : ⇒ cut the XML output in :
  1. one XML containing ALL tables descriptions :
     ↪ the output contains a tag $<$ *accessURL* $>$ with the URL of the column's definitions
  2. for each table a XML containing the entire definition of columns

```
<tableset xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xmlns:vod="http://www.ivoa.net/xml/VODataService/v1.1" xsi:type="vod:TableSet">
<schema>
  <name>vizls</name>
  <description>Large surveys - big catalog</description>
  <table type="base_table">
      <name>vizls.c2mass</name>
      <description>2MASS All-Sky Catalog of Point Sources (Cutri+ 2003)</description>
      <acessURL>/vizier/tap/column?c2mass</acessURL>
  </table>
  .......
</schema>
</tableset>
```

```
<tableset xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xmlns:vod="http://www.ivoa.net/xml/VODataService/v1.1" xsi:type="vod:TableSet">
<schema>
    <name>vizls</name>
    <description>Large surveys - big catalog</description>
    <table type="base_table">
        <name>vizls.c2mass</name>
        <description>2MASS All-Sky Catalog of Point Sources (Cutri+ 2003)</description>
        <column std="true">
            <name>RAJ2000</name>
            <description>(ra) Right ascension (J2000)</description>
            <ucd>pos.eq.ra;meta.main</ucd>
            <dataType xsi:type="vod:TAPType">REAL</dataType>
        </column>
        .....
    </table>
</schema>
</tableset>
```

Index Tables with PostgreSQL

- **Pgsphere** : user-friendly
- **Q3C** : more efficient with large volumetry, index size smaller
- **H3C** : Healpix Tree C - standardization of index in used in CDS (a standard ?)

H3C characteristics

- similar to **Q3C** but using healpix instead of Qbox
- use the PostgreSQL functional index
- the same functions than Q3C : *h3c_radial_query, h3c_join, etc.*
- available for convex polygon only !
- as efficient than **Q3C** when merging 2MASS and hipparcos :

Q3C, H3c	15minutes
Pgsphere	48 minutes

Manage heterogeneous coordinate system

Standardization of coordinates system

⇒ add physically (if not exist) the ICRS columns

Understanding ADQL function in VizieR Tap

- *What happend if coordinates systems in ADQL and stored data are different ?*
  POINT('ICRS', rab1950, deb1950)
  VizieR management :
  for computing — ignore the user coordinate system (somewhere else than in *select* part)
  for output display — make a change of coordinate system (in *select* part)
- *What happend if two functions in different coordinate system are joined ?*
  CONTAINS(POINT('ICRS',....), CIRCLE('FK4', ...))
  ⇒ VizieR compute the change of coordinate system
  ↪ index is not used !

AS4 library usage

double precision[2] **as4_convert**(ra, dec, csys_in, csys_out)
double precision[2] **as4_convert**(ra, dec, csys_in, equinox_in, epoch_in, csys_out, equinox_out,epoch_out)
Example : *select as4_convert(ra+n*pmra, dec+n*pmde, 'ICRS', 'FK4') ....*

Progress in the developpement

Database mirroring	mirror software done + big catalog are partially stored (1.5Tb currenly, 4Tb expected...)
Homogenize tables with ICRS	program done. TODO : execution on catalogs
H3C index	done
Parsing ADQL and translation to SQL	Almost done. TODO : tests, adjust precision...
TAP implementation	Partially done. TODO : stored file using IRODS + TAP_SCHEMA output
WEB interface	on action. TODO : the asynchrounous call