TAPVizieR, a new way to acess the VizieR database

Gilles Landais - François Ochsenbein

A dedicated database

The VizieR application provides 10,000 catalogs, 22,000 tables and 320,000 columns stored in a transactionnal database, or in dedicated binary files for the large catalogs.

The TAP implementation of VizieR provides **ADQL** (a SQL extension containing astronomicals capabilities) access for users and remote applications.

We chose the transactionnal database **PostgreSQL** to gather all the VizieR data in a unique database. The size for data and indexes is currently about 3.5Tb.

Some typical figures : (*)

2MASS	450M records,	size = 247 GB
SDSS8	700M records,	size= 673 GB
PPMXL	910M records,	size = 275GB
USNOB1	1G records,	size = 372 GB

(*) The size number takes into account the storage in a PostgreSQL database of the couple data+index

Architechture TAPVizieR

TAPVizieR   **VizieR**

## Positional index with H3C in PostgreSQL

The H3C index is a btree index on the **HEALPix** number associated to the position.
H3C is largely inspired from Q3C (*Koposov*). The testellation of the sphere in HEALPix cells needs to work in a deep resolution to be as efficient as Q3C. This implementation was possible thank to the recent improvment of the HEALPix 64bits C++ library maintained by the Nasa.

Some comparisons between libraries :

	PgSphere	Q3C	H3C	H3C (cluster index)
conesearch in 2mass (M1, radius=2arcsec)	340ms	360ms	380ms	
conesearch in 2mass (M1, radius=2arcmin)	500ms	390ms	400ms	
conesearch in 2mass (M1, radius=2deg)	88s	4.3s	1.26s	
crossmatch Tycho-hipparcos	110s	4.5s	4.5s	3s
crossmatch 2mass-hipparcos	48min	15min	14min	6.5min
crossmatch 2mass-Tycho	4h30	49min	48min	11.5min

Comparisons of the positional indexation applied to Hipparcos (~100K records), Tycho (~2.5M records) and 2mass (~450M records) catalogues. Tests performed on the same Linux computer (99G RAM) using aPostgreSQL (version 9.1) database. The cache disk was deleted before each test.
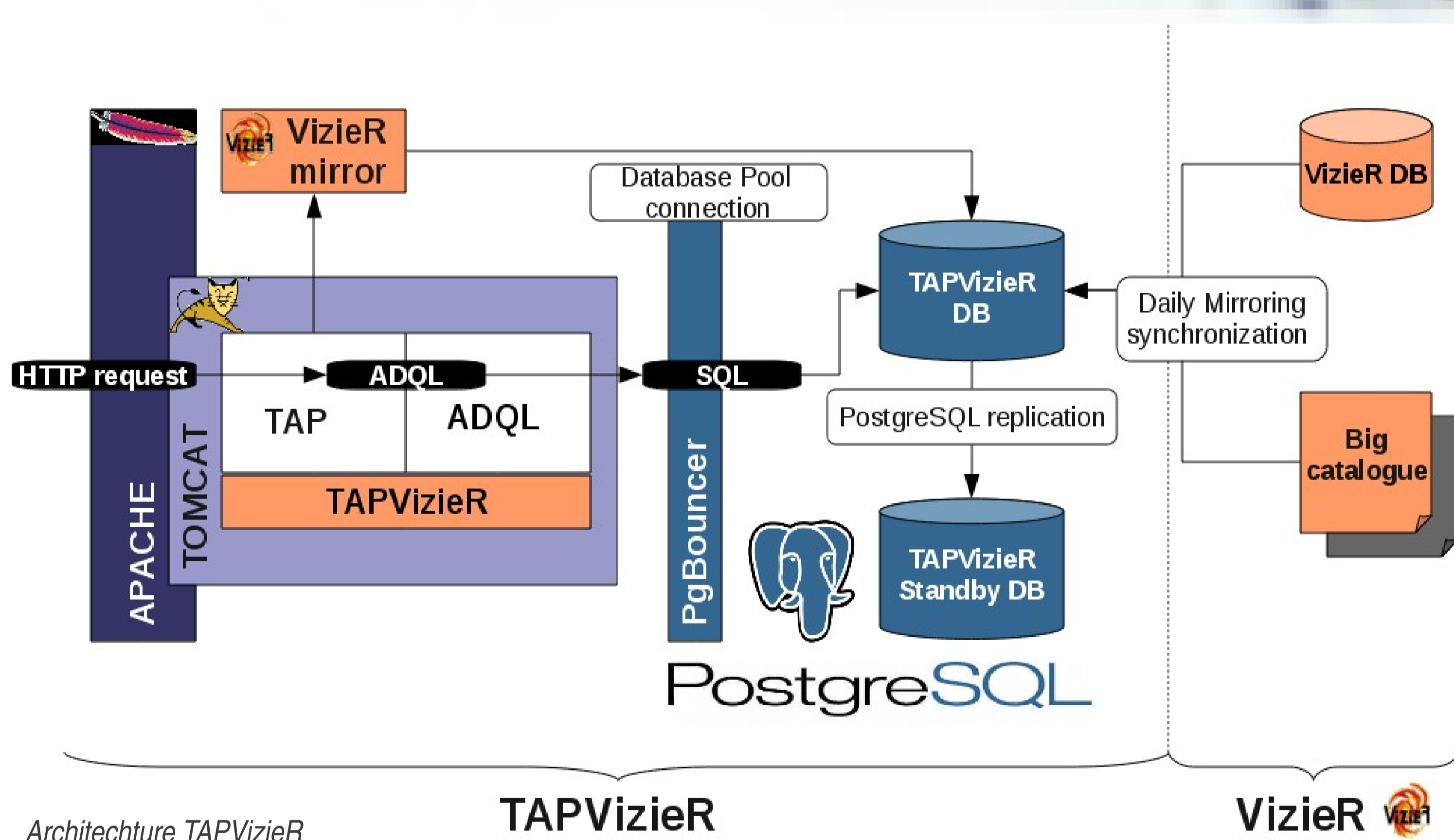
Homogenization of the coordinate systems

The coordinate System used in VizieR depends on the catalog. However, VizieR can compute positions of every catalog in an other coordinate system than the original, taking into account the proper motions when these are described in the VizieR METAdata.

Working in a unique coordinate system is obviously better in ADQL (especially in crossmatches). So, we are adding the position in **ICRS** in each table. The new columns (ra.icrs, de.icrs) are computed at the epoch J2000 if proper motions are known.

Some departures from standard in the first version.

➔ **the TAP XML output describing the database :** (/tables)

VizieR (currently) returns the table descriptions without the column descriptions (~ 3.5Mb) to limit the volumetry of the XML output (~86MB to be TAP compliant).
However, we added a service (i.e. an other URL) to give the entire description for a single table. This service is not yet a TAP standard.

➔ **Service to find the interesting tables :** (/search)

The rich METAdata available in VizieR are used in the way to retrieve tables from position, keywords, authors, etc.

➔ **Understanding the coordinate system in TAPVizieR :**

TAPVizieR makes the expected change of coordinate system in a join of geometrical areas. However, the coordinate system stored in the VizieR METAdata is used even if the ADQL query specifies another system.
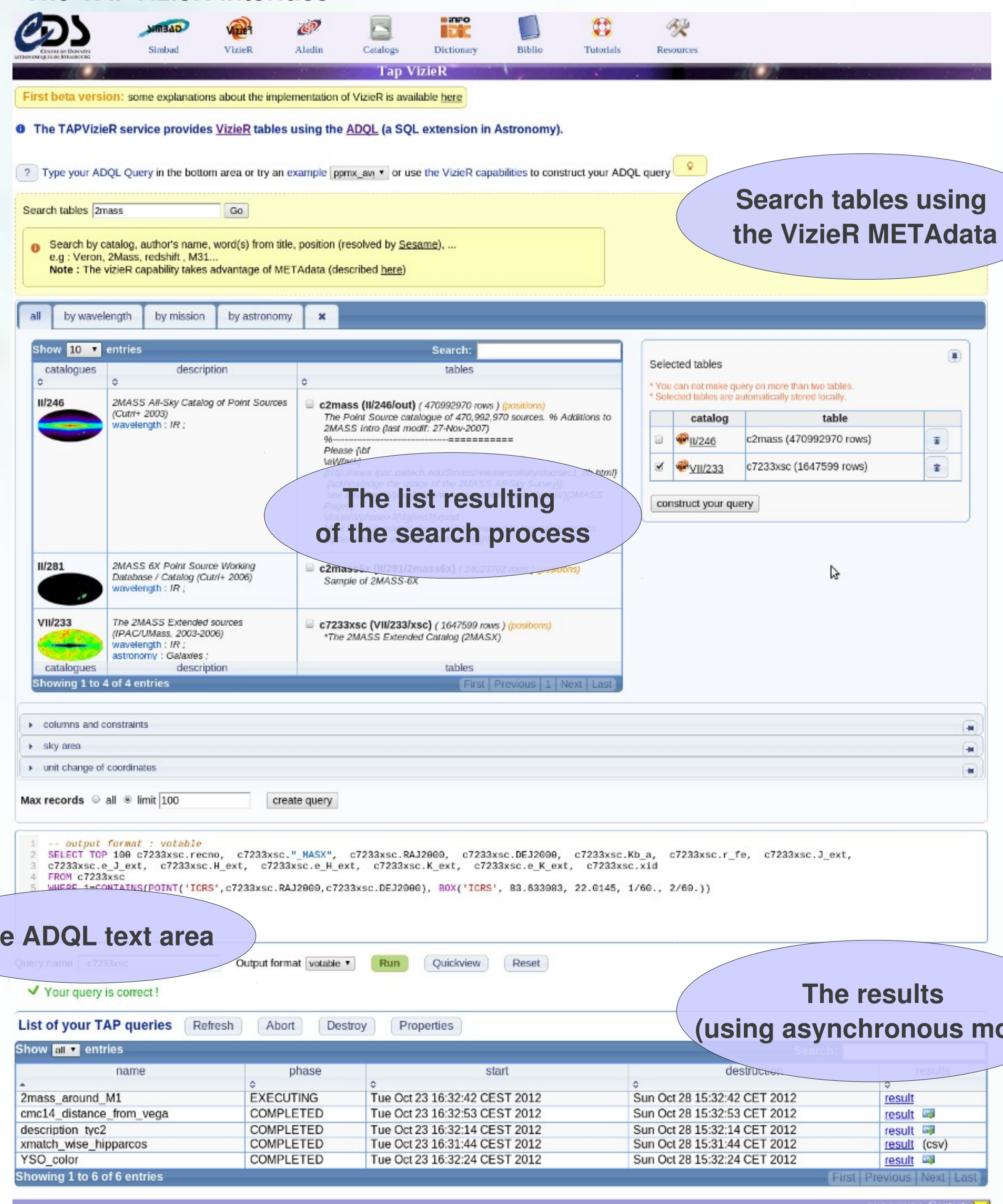
Ongoing developpement

➔ The upload capabilities, described in TAP would be implemented
➔ Using the HEALPix indexation to provide a MOC service
  (*P.Fernique: oral presentation*)

The TAPVizieR interface

Search tables using the VizieR METAdata

The list resulting of the search process

The ADQL text area

The results (using asynchronous mode)