

CDS xmatch service updates

Thomas Boch
François-Xavier Pineau
Sébastien Derrière



*IVOA Interop Meeting@Naples
May 17 2011 - Apps 2 - T. Boch et al.*



Service presentation

- Purpose
 - **Positional cross-identification** of sources in 2 tables among:
 - VizieR tables (including very large surveys)
 - SIMBAD
 - User-uploaded tables
 - Focus on **large-scale** cross-match
 - UI : web application on top of UWS service

Demo

CDS X-Match Service

Tables managementX-match

thomas Preferences Admin Logout

Choose tables to cross-match

NOMAD

X

USNOB1

My storeVizieRSIMBAD

My storeVizieRSIMBAD

Hide options

Cross-match criteria

By position

Radius: 2 arcsec

By position including error

Sigma: 3.43935 (completeness: 99.73 %)

Max. distance: 5 arcsec

Cross-match area

All sky

Cone

Center: LMC

Radius: 1 deg

Healpix cell

Nside: 1

Index:

Begin the X-Match

Visualize and manage your cross-match jobs

List of X-match jobs

Table 1	Table 2	Options	Begin	Status	Actions
DENIS	2MASS	fixed radius +	13/05/2011 at 14:12	completed	Get result

For the selected job(s):

Delete

Job executed in 1h2min9s
10min59s to correlate
51min9s to generate file
Result: 272,005,494 rows

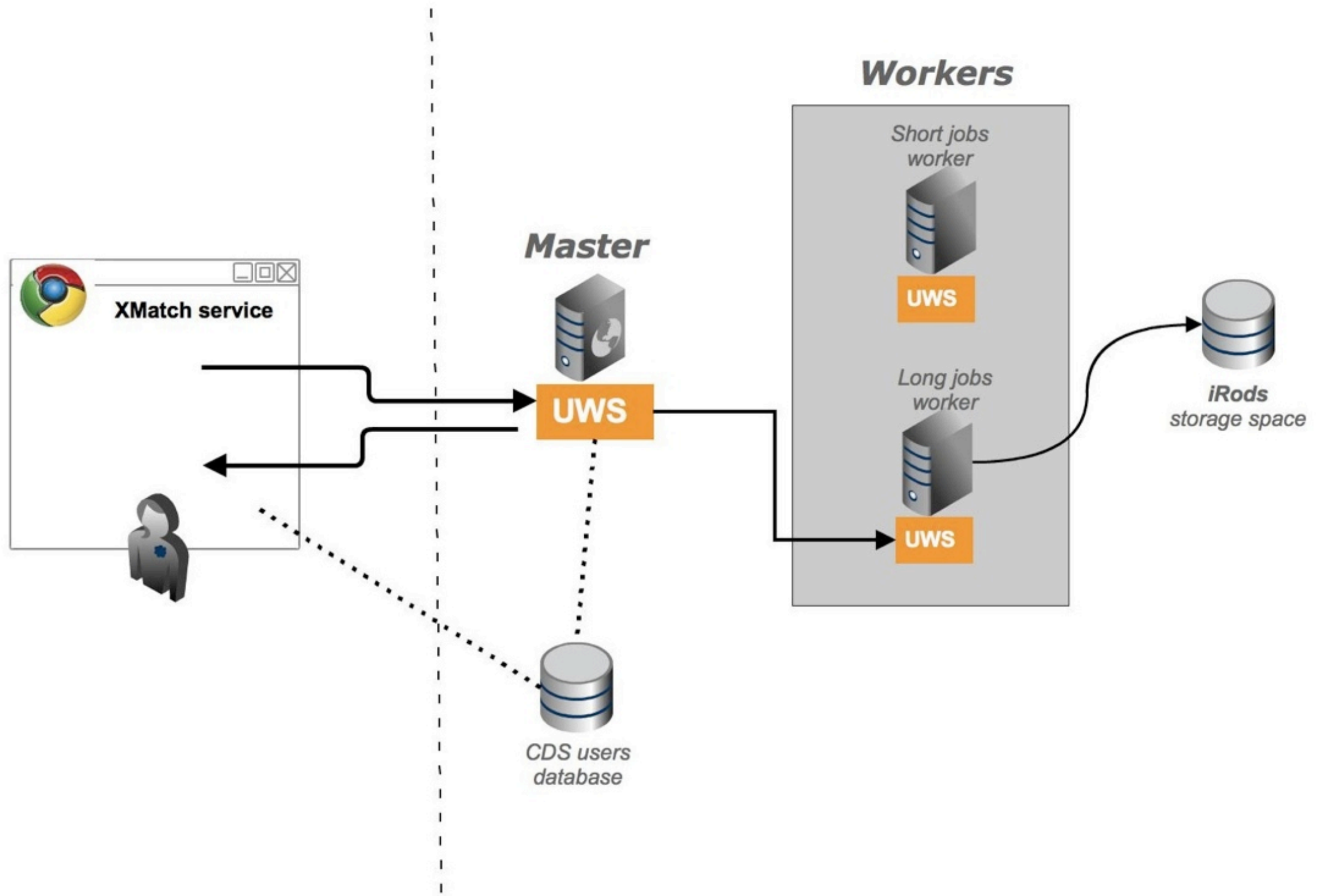

CENTRE DE DONNÉES
ASTRONOMIQUES DE STRASBOURG


EURO
ICE
International
Cooperation Empowerment

IVOA Interop@Naples
T. Boch, CDS


IVOA

Architecture



Updates

- Since Nara Interop:
 - cross-match jobs **dispatched on 2 machines** (one for short jobs, one for long jobs)
 - inclusion of **VizieR tables > 10M rows** (USNOB1, 2MASS, SDSS7, NOMAD, PPMXL, GLIMPSE, ...)
 - cross-match on all-sky **or in a cone** (position or object name + radius) **or for a given HEALPix cell**
 - choice of position+errors metadata (for user-uploaded tables)
 - faster HDDs (15,000 rpm)
 - faster output generation

Service performances

Cross-match at 5 arcsec :

Table 1	Table 2	Computation time	Result generation	Result size	Total time
SDSS <i>357M rows</i>	2MASS <i>470M rows</i>	7 min	12 min	13 GB	19 min
DENIS <i>355M</i>	2MASS <i>470M</i>	11 min	51 min	58 GB	1 hour 2 min
GLIMPSE <i>104M</i>	NOMAD <i>1.1 billion</i>	6 min	17 min	19 GB	23 min
SIMBAD <i>5M</i>	USNOBI <i>1 billion</i>	3 min	1 min	1 GB	4 min
SIMBAD <i>5M</i>	PPMX <i>18M</i>	20 seconds	20 seconds	440 MB	40 sec

I/O limitations

- When computing the associations :
 - **Reading input catalogues** is the main limiting factor
- When generating the result file :
 - Performance mainly limited by the **network bandwidth between workers and iRods** ($\approx 12\text{Mbyte/s}$)
- When the user downloads the result :
 - Downloading SDSS-2MASS result at 1Mbyte/s takes 3.7 hours

Hardware



- Worker 1:
 - 2x Quad Core 2.27 GHz → 16 threads
 - RAM : 24GB @1333MHz
- Worker 2:
 - 2x Six Core 2.27 GHz → 24 threads
 - RAM : 32GB @1333MHz
- 6TB RAID5 array with 15,000 rpm disks
 - Read : 570 MB/s
 - Write : 130 MB/s