### Construction of Multiple-Catalog Database for JVO

Masahiro Tanaka, Yuji Shirasaki, Masatoshi Ohishi, Yoshihiko Mizumoto

National Astronomical Observatory of Japan, Tokyo, Japan

Yasuhide Ishihara, Jumpei Tsutsumi, Yoshihiro Machida

Fujitsu Ltd., Chiba, Japan

Hiroyuki Nakamoto, Yuusuke Kobayashi, Michito Sakamoto

Systems Engineering Consultants Co. Ltd., Tokyo, Japan

Abstract. We present a development of an efficient query system for the Multiple Catalog Database. In a case where a VO user collects all the available information about an astronomical object, it is necessary to send queries to all the VO services. However, this would be inefficient since most services tend not to have data on the target object. We thus developed a new query system for users to efficiently find all the available information from multiple catalogs. This system collects basic data (coordinates and fluxes) of astronomical objects from the VO data services and major astronomical catalogs, and store them into a local database built with PostgreSQL. The design of this database is one of key issues; an efficient query mechanism for more than billions of objects is required. We employed the Table-Partitioning technique and developed a method to build queries for the partitioned tables. We compared our method with the partitioning function of PostgreSQL and found that our method is more efficient by a factor of 7 to 150. We discuss the architecture of this system.

#### 1. Introduction

One of common tasks to astronomers is to search for available observational data on the target astronomical objects that they are studying. In order to collect all the available information about the objects from VO, it is necessary to send queries to all the VO services. This method would be inefficient since only a part of data services will provide data on the target objects. We thus developed a new Multiple Catalog Database in order to efficiently find all the available information from multiple catalogs. This database consists of a single table in which we store multiple catalogs collected through VO services and also obtained manually. In order to develop this database, we investigated the following issues: (1) Fast query mechanism for large number of objects, and (2) Design of unified table to store various kinds of astronomical catalogs. This database is available to public through JVO system (Ohishi et al. 2006, Shirasaki et al. 2006, 2007).

1

2 Tanaka et al.

# 2. Design of Table Partitioning

Query performance for a large database is one of key issues to build the Multiple Catalog DB. For example, 2MASS All Sky Catalog has 4.7 billion objects, SDSS DR6 Catalog has 287 million objects. If we simply store more than billions of objects into Relational Database Management System (RDBMS) as a single table, we do not obtain enough query performance. Therefore we employed the Table-Partitioning technique for large databases. In order to make position search faster, we used HTM (Hierarchical Triangular Mesh; Kunszt et al. 2000) as a partitioning key. Objects are grouped by upper HTM level 6, corresponding to  $8 \times 4^6 = 32768$  regions, and stored data into distinct tables, named as psc\_32768, psc\_32768, ..., psc\_65535. If this system receives the following simple region query:

```
select ra, dec, j_m
```

from psc where Region('Circle 0 0 1');

then the region condition is converted to the condition of table selection and the ranges of HTM lower ID as:

```
select ra, dec, j_m
```

```
from ( select * from psc_63488 where htm_id between 0 and 65535
union select * from psc_63488 where htm_id between 217088 and 218111
union select * from psc_47104 where htm_id between 0 and 65535
...
) psc;
```

We implemented a query converter in Java using HTM Java library developed at JHU and used PostgreSQL ver 8.2 for RDBMS.

# 3. Performance of Table Partitioning

To evaluate the performance of our method, we compared it with the partitioning function which PostgreSQL was equipped with as of the version 8.1. For this test, we used 2MASS All Sky Catalog. We performed this test on machine with Pentium-4 2.8GHz and 2GB memory. The result (Table 1) shows that our method is 7-150 times faster, suggesting that our partitioning method yields enough performance for large databases. We note the following points:

- 1. The elapsed time in our method does not include time to convert region conditions. It is less than 0.5 s.
- 2. To test the PostgreSQL partitioning function, we reduced the number of partitioned tables to 2048, 16 times smaller than our method. This is because PostgreSQL has a restriction on the number of tables to transact simultaneously since this transaction consumes a lot of shared memory. In contrast, our method allows us to transact larger number of partitioned tables.
- 3. Elapsed time of the PostgreSQL partitioning function depends on the number of HTM condition, rather than the search radius.

# 4. Design of the Multiple Catalog DB

Astronomical catalogs are provided in their own table formats; one record for one object and multiple columns for multiple band fluxes. Multiple catalogs in

Search radius	Result objects	Elaps	ed time (s	# of HTM conditions					
arcmin	#	Postgre SQL	Our method	ratio	$\begin{array}{c} \text{Postgre} \\ \text{SQL}^a \end{array}$	$\begin{array}{c} \text{Our} \\ \text{method}^b \end{array}$			
1	2	6.46	0.04	154	32	32			
10	165	3.81	0.03	127	16	16			
60	6697	6.47	0.11	60	32	32			
100	26720	2.02	0.31	7	4	16			
180	57246	9.04	0.71	13	48	72			

Table 1. Measured elapsed time to search partitioned tables

<sup>a</sup>the number of HTM between conditions in where clause

<sup>b</sup>the number of subqueries concatenated with union

such *object-based* formats are difficult to combine into a single table. We thus employed a simple *flux-based* format shown in Table 2 for the Multiple Catalog DB. In our format, each record has *only one* flux column. Multi-column flux data of original catalogs are decomposed and stored into distinct records. This method has been employed for the Catalog of Infrared Observation (Gezari et al. 1999). Multi-wavelength information of objects such as the SED (Spectral Energy Distribution) is obtained by grouping flux records in terms of position. Although the Multiple Catalog DB includes only a part of information from original catalogs, further information can be obtained from the original catalog whose access point is recorded in the link\_ref column.

category	$\operatorname{column}$	description
Object	id	Object ID
	name	Object name
Position	ra	Right Ascension
	dec	Declination
	pos_err	Position Error
	htm	HTM index
Wavelength	band_name	Band name
	band_unit	Unit of band
Flux	flux	Flux value in catalog
	flux_err	Flux error
	flux_unit	Unit of flux
	flux_srch	Flux in Jy
Reference	link_ref	Link URL to reference
	org_id	ID in original catalog
	cat_id	Catalog ID

Table 2. Table design for the Multiple Catalog DB

#### Tanaka et al.

🖯 JVD Data Search – Mozilla																
	20 7999-310 9-14D ALTU	Indiana Awards	0													
9.0-000000	p://yw.nax.ac.jp/dev/perial/top-page.do	• 🕨 🔂 • Occele	C JVD VOTable Ver	Maailla Fie											DE.	2
JVC	Top   Search   VO Services   Subaru   Analy		771112 編集2 書			2-1KD A	178	_	_					-		
IAPANESE VIETLAL OBSER	valoev ee 20070918	Tanaka Masahira (tanaka adein)		같 E Htp://www								- 4	G. Occe			101 E
About Acknonicssment	Service Contents			Tool	Search   VO 3	Services   Sub	ara   Anahais   W	forkfow I JV	O Soace	flagout						^
			INFAMILE VIETAN CEN							Masahira (tanaka adeini)						
Nerwa Versian 0.2 is open	Data Search	Service Search	=> Location Top Page	> Astro Tools > 5	Sextractor											
since 2007-07-01	GuickSeach	Reyword Search	Save Filter Metado	te Graphic Sec	arch Appears	ance Sky Me	p & SED Color-	-Calor Plat		_						
	<ul> <li>Search on a single VO Service <sup></sup></li> </ul>	Category Search	Tatal 11106 records	new 121451	01007	200				Pop-up	Men	u te	or v	aric	ous tu	nctions
	<ul> <li>Parallel search on multiple VO Services III</li> </ul>	Advanced Search	Alize Name	D0 (C1			01	61		ier ier		icro.	[g]]	01	gu A	
	Xmatch Search		group	ID NAME	RA	DEC	POSERR	LINK REF		BAND NAME BAND CE	NTER BAND UNT	FLUX	FLUX, ERP	FLUX	FLUX, SRCH	
	JVOQL Search	Astrnomical Tools	10 Check cownis	90 sort sort 187	10.386402	41 354405	set Sec-05	tee.	tee all	1 126	tee [	10.545	0.153	line line	0.0003352921	
		Source Extractor	1 1	188	10.386402	41.354465	5e-05	Link	two mees	H 1.6	1/76	16.217	0.165	rot	0.000342286	
	Subaru	<ul> <li>HyperZ</li> </ul>	2 🗆	189	10.386402	41.354465	5e-05	Link.	frio mattr	Kg 2.15	une	15.414		T-66	0.003459754	1
	Suprime-Cam		3 1	190	10.350114	41.050224	1.944444-05		frio mass	J 1.26 H 1.6	un	15.509	0.076	rag	0.000354147	
		Workflow	5 🗆	192	10.390114	41.050994	1.94444e-05		tino mass	Ks 2.15	Ling	14,452	0.072	1.00	0.00110491	
	JVO Space	Workflow Editor (Script)	0	192	10.390661	41.331158	6.833332e-06		two maos	J 1.26	119	16.515	0.143	not	0.00343382	
		Vibridiov Editor     Vibridiov Monitor	7 🗆	194	10.350061	41.331168	5.83332e-16		teo mass	H 1.6	ung	15.856	0.155	F-16	0.003477258	
	+ Home	Vieridiaw Monitor		196	10.39146	41.055534	7.5+-05	Link	140 8 800	J 1.26	Lin	16.742	0.167	n og	0.000327622	1
_			erosp check downlo	ad ID NAME	RA	DEC	POS_ERR	LINK, REP	GATID	BAND NAME BAND GE	NTER BAND UND	FLUX	FLUX, ERF	FLUX	FLUX,SRCH	
	Entry Page:		10	197	10.39146	41.055534	7.5+-15	Link	1 MO IE also	н 1.6	ung	15.614	0.150	nat	0.00053199	
L	inny raye.			198	10.39146	41.355534	7.5e-85	Link.	two mass	Ke 2.15	read	15.648	0.253	1.06	0.00035722	
	Categorized Se	rvicoc	12	199	10.394434	41,388909	5.55556e-05		two mass	J 1.26	119	16.692	0.155	not	0.000342758	1
L C	alegonzeu se	IVICES	¢					/	-						>	
			光7			_						-				ille.
-	10 JVD Quick Search - Nozilla Firefox		E F			Nc	w da	acia	ın T	able Vi	owor					
and a state	77-(1)(2) 編集(2) 表示(2) 展開(2) 7557-558	B 2-40 1678		<u></u>		INC	w ut	coig	, i i i	able vi	ewei					
	🕸 • 🕸 • 🗑 😳 🏠 🕒 http://prenas.ac.jp/de	w/portal/search2.ds/action-quick.	• Coople	14												
<u> </u>	Tap   Search   V	O Services   Subaru   Analysis   Workflow   JNO Space	Lopour	1	-			_								
	JAFANISE VIRTUAL DESERVATORY VEF.20070915		ka Masahiro Ranaka admir	2			wor - Mozilla   STOO ELECT		76 er 4/7							
	=> Location: Top Page > Search > Quick Sear	ndh			40 . 1		合 E Htp://w					× 1a	G+ Good		101	
	Quick Search				-	VOTable Vi		600							1.	
								a   Search   )		Ouic	:k Im	ane	s vir	ewe	r	
	Name or Coordinate	Samples: 34.5 -5.0 M H31			-	V	TOS		vo services						.',	
	Coordinate Frame		J0000 (793) 36		=> Loca	tion: Top Pag	> Astro Tools.	Sextractor	¢ .	<u> </u>	ED pl	otte	or e	atc		
	Search Region Radius		10 anein 18		Save	Filter Matad	ata Graphic S	Search App		NY Hap 8	-D pi	on	- , 、	cic.		
	Wavelength Band				Greep 3	D 0 -	+ Update	1					1			
	Wavelength Band	CAL CRedo CR Cotical C		2	group id	1 = 0				Position Map				00	x.4	
									41.25	342				ND.OENTER		
	Brightness Bange		C2 C0 C1 C16						41.25						C	
	Brightness Harge	brighter then: Max AB M fainter then:	Mos,AB M						41.25	585			1.0		um um	
			104/10 00			1000			41.28 13 41.28	384 -		2 0.11	2.		U.84	
	Object type specific criteria								8 4 28 0 4 28	542		Dens	13		UR	
	ALL AGN GRB		、						名 41.28 41.28	541		ă î	2		0.00	
	Submit Olew Check SQL	1							41.28	222		-	1.3		um	
									41.25	197 -		0.061	1.1		UR4	
		Simple Interfa	o to						41.28	10,46090 18,66885 1	1 662 DO . 10 .663**		1.		UR	
		Simple Interna								RA (dep			AL DAD	ND_GENTER	DAN	
	N/	Iultiple Catalog	Quary		00h42	nGB.600640 + 8.0 arcmin	41 dt 5m46.88290	0					1.1		uni	
	_¶‴V	iumple Catalog	Query	~		from archive.	na.org						2		-	
	光7				光7								-	296		

Figure 1. New User Interface of JVO portal

## 5. New User Interface of JVO Portal

The Multiple Catalog DB is build in the JVO Portal system and available to public users. We also developed new user interface (Figure 1) including a simple query interface to search this DB, categorized services with improved accessibility, and a table browser equipped with many new features.

Acknowledgments. This work was supported by the JSPS Core-to-Core Program and Grant-in-aid for Information Science (15017289, 18049074 and 19024070) carried out by the Ministry of Education, Culture, Sports, Science and Technology of Japan.

## References

Gezari, D. Y., Pitts, P. S., & Schmitz, M. 1999, Catalog of Infrared Observation, ed. 5
Kunszt, P. Z. et al. 2000, in ASP Conf. Ser. 216, ADASS IX, ed. N. Manset, C. Veillet,
& D. Crabtree (San Francisco: ASP), 141

- Ohishi, M. et al. 2006, in ASP Conf. Ser. 351, ADASS XV, ed. C. Gabriel, C. Arviset, D. Ponz, & E. Solano (San Francisco: ASP), 375
- Shirasaki, Y. et al. 2006, in ASP Conf. Ser. 351, ADASS XV, ed. C. Gabriel, C. Arviset, D. Ponz, & E. Solano (San Francisco: ASP), 456
- Shirasaki, Y. et al. 2007, in ASP Conf. Ser. 376, ADASS XVI, ed. R. A. Shaw, F. Hill, & D. J. Bell (San Francisco: ASP), [O10.2]