



#### ://terapix.iap.fr

# A wide-field images processing suite

TERAPIX (Traitement Élémentaire, Réduction et Analyse des PIXels de megacam) is an astronomical data reduction centre dedicated to the processing of extremely large data flows from digital sky surveys. Located at the IAP (Institut d'Astrophysique de Paris), its primary tasks are : - to develop image processing and pipeline software for MegaCam (the new giant CCD camera of the CFHT telescope in Hawaii); - to develop and provide tools for handling of large CCD images;

- to operate the final reduction pipeline to produce calibrated images and catalogues of MegaCam images over the next 5 years;
- to provide technical assistance and TERAPIX computing facilities to MegaCam users.



### Snoop

downloads Perl daemon that automatically data via a http or ftp protocol.



Based on wget (free tool for non-interactive files download from the web), Snooppix can scan a web page and download data sending simultaneous wget on each web page's link.



## DataTransfer

Set of perl and php scripts that dispatches data and managing transfers on a cluster.



In order to transfer and sort out quickly images coming from CADC on our cluster, we have developped a software based on php and perl scripts which allows to dispatch data across a cluster.





**Processing tools to produce calibrated** data from large astronomical images



To reduce MegaCam data (the world biggest CCD camera : one square degree), we have designed a pipeline software named Spica (Software Pipeline for Images and

Catalogs Analysis). It produces calibrated images (astrometry, photometry and coaddition/resampling, catalog extraction). It can be installed as a standalone application or as a client/server one.

Once the data are stored on the cluster, Spica checks their integrity and loads them in the SQL database (see RDBix). MegaCam images are uncompressed if needed and sorted out by filter and runid.

The quality of each image is controlled by Qualityfits assessment (see companion poster) and this information is used to select which data could be used by Spica. There are two ways to use it :

DataTransfer is managed from a web page. So, you can transfer images by simple clicks. DataTransfer is not only a tool for data transfer, it can also display : - FITS header

- files distribution across nodes. - transfer history
- free disk space on each nodes

# **RDBix**

RDBIX is the memory of the whole pipeline. It is a MySQL relational database designed for the storage of catalogues and metadata produced by the processing software. RDBIX organizes these metadata to maintain history of the images processing.



Each step of the pipeline process communicates with RDBIX to get data about images (coordinates, storage path, ...) and to inform RDBIX, via a piece of software called dbclient, about images modifications. An XML file, called instruction file, serves as a support for this communication with RDBIX. Other data, such as acceptance files (quality data) are also integrated into RDBIX.

RDBIX is based on a completely dynamic design, which allows the database to grow without heavily impacting the performances. Finally, RDBIX has been designed to run in a distributed environment : each node of the cluster runs a client version of dbClient (and its own pipeline package), while the master node runs a server version of



On each node, Perl scripts manage input/output from each pipeline application, database storage and send command line scripts to the spica which processing run daemon according to priority rules. Generated files are sent to the output storage disk. Input data are processed with rules written in configuration files.

A web interface allows authorised users to reduce their own data. They can choose :

1) which data to process according to their preferences,

which applications (including 2) configurations) to run.

All steps are recorded in the database for further analysis.

A	A											
Back F	orward R	🍊 🚷 eload Sto	p \land http://cli	x.iap.fr:808	10/spica/admin/	admin-spi	ca/qualityfits/in	dex.php			🚽 📮	M
(ome > Too	TEF ls > Data redu	RA   etion > Spica	ЭIX	Quality	ProcessID : yfits Incomin	g data			a			
mages proc	essed : 58		RunId		Filter		Node					
earch imag	es : 58	Name	Number %	Name	Number 9	Nam	e Number %					
//dAilliulli II	iages dishiay	03ALO	2 13 22 %	i	2 3	% node	1 13 22 5	%				
		03AF1	9 45 77 %	r	23 39	% node	2 45 77 1	%				
				g	23 39	% ~						
a 1 D'	1 75 7.1	T*11	N I F	u .	10 17							
Search Dis	play Runia	Futer	Node Exp	_ume (mi	n/max) Airma	iss (min/n	hax) Miseeing	(min/max) IV	bkg (min/max			
G0 15		≥   ≥	I  13	/ (860	1.034	/ [1.508	p.7317 /	1.3833  18	3.48 /  23.49			
Image	RunID	Filter	ExpTime (s)	Date	e I	RA	Dec	Equinox	Airmass	Mseeing (')	Mbkg (mag)	Node
705487 o	03AL02	i	620	2003-06-	09 14:17:	55.61	51:43:01.0	2000	1.418	0.941	18.51	node1
7054890	03AL02	i	620	2003-06-	09 14:17:	52.93	51:43:31.0	2000	1.508	0.8624	18.48	node1
7072750	03AL02	r	500	2003-06-	23 14:17:	56.52	54:28:16.2	2000	1.219	1.0495	21.12	node1
7072770	03AL02	g	500	2003-06-	-23 14:17:	52.27	54:30:01.2	2000	1.231	1.146	21.41	node1
7072790	03AL02	g	500	2003-06-	23 14:17:	55.72	54:31:00.8	2000	1.248	1.0453	21.44	node1
7074330	03AL02	r	500	2003-06-	24 14:17:	56.47	53:32:16.0	2000	1.204	0.9882	20.98	node1
7074350	03AL02	g	500	2003-06-	24 14:17:	52.32	53:34:00.8	2000	1.214	1.2637	21.33	node1
7074370	03AL02	g	500	2003-06-	24 14:17:	55.68	53:35:01.1	2000	1.23	1.2551	21.36	node1

ak - Stop & Stop & ttp://clix.iap.fr:8080/spica/login.p TERAFIC Welcome to SPICA <u>Group change</u>
 <u>Ganglia</u>
 <u>3dm</u> Define a new process Config management ocess configuration History

• View waiting/running runid 60 selected : launch delete archive Archive

• <u>View archive</u>

dbClient and hosts the MySQL RDBIX database which is used by every node.

7074450	02 A E10	r	560	2003-06-24	15:08:29:69	5:00:10.4	2000	1.111	0.8100	20.93	node
7074460 (	02 A E10	1	560	2002-06-24	15:00:29:29	5:00:10.4 4:50:40.4	2000	1.133	0.0102	20.94	node
7074460 (	02 4 510		500	2000 0/ 24	15.00.20.70	4.50.40.4	2000	1 1 50	0.0102	20.74	

7074410 03AF19 r 560 2003-06-24 15:08:28.99 4:59:55.3 2000 1.064 0.7635



Results

#### Typical processing of MegaCam images.

On the left is a raw U band 860 seconds exposure. On the right, a stacked U+G+R image composed of 45 Megacam images obtained with Swarp software.

Processing time is around 40 hours on a dual XP1800+ computer (will be down to 30 hours with Opteron processor).

Color image generated by the stiff too which is able also to convert image from FITS format to PNG.



Part of a U band raw image



Same area after image processing

Administration interface written in PHP can check the pipeline status. It can also display spica produced data.



The users have their personal web spaces allowing to create a customised environment. Forms help users to build their setup processes.

		Configurat	10n	- Step 2		
		SW	arp			
	(*) indicates parameters which o	an be omitted from this config file				
Output	Weight type Map weight 📩	Help				
Astrometry	Celestial type	Native 💌		Help		
	Projection type	Distorted tangentiel	<u> </u>	Help		
	Center type	All		Help		
	Coordinates of the image center	00:00:00.0, +00:00:00.0		Help		
	Pixel scale type	Median 🔄	_	Help		
	Pixel scale	0.0		Help		
	Image size	0	_	Help		
Resampling	Method Oversampling in each dimension Interpolate bad input pixels ? Flux-scaling factor	Lanczos3 y         Help           n 0         Help           Yes y         Help           Fixed y         Help				
Background subtraction	Subtraction sky background ? Background mesh size in pixels Background map filter range in	Yes     Help       126     Help       meshes     3				
Co-addition	Combine resampled images ? Combine type	Veighted T Help				
Miscellaneous	Write information about each is List of FITS keywords to propa Number of CPU	nput file in the output image header? gate from the input to the output head	Iers Fil	TER,EXPTIME,PHOT_C,PH	Hel IOT_K,AIRMAS Hel Hel	lp lp
		Submit	Rese	et		