

NAREGIによる分散天文データ 解析機構の試験構築

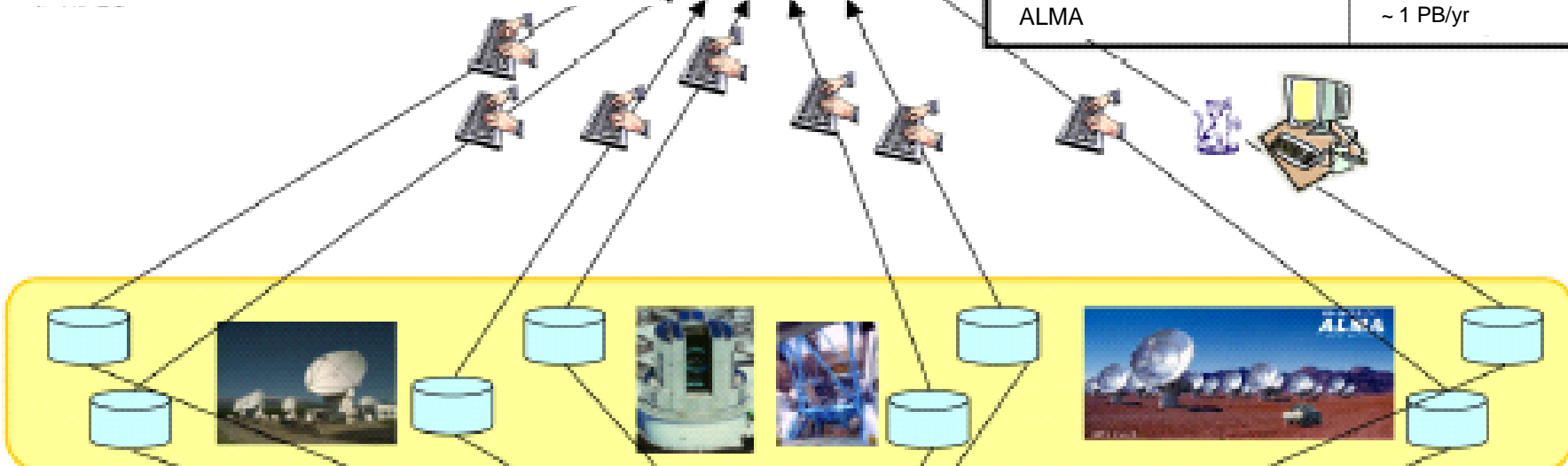
大石雅寿・白崎裕治
国立天文台，総合研究大学院大学

三浦 昭
JAXA/ISAS



**Too hard to collect and analyze whole data.
Need Cat's help**

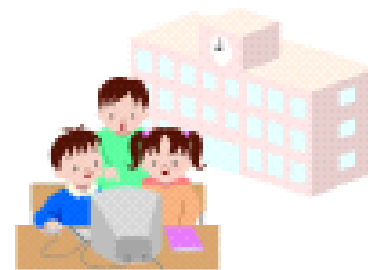
	Data Rate
Nobeyam Radio Obs.	~ 1 TB/yr
SUBARU telescope	~ 20 TB/yr
ALMA	~ 1 PB/yr



Accelerate astronomical research, and sufficient time for research itself !!

Virtual Observatory

Even for educational use



Accessible from anywhere at any time

JVO Portal

<http://jvo.nao.ac.jp/portal/>

Several new features have been added

More than 4000 resource can be accessible

The screenshot shows the JVO Portal website. At the top, there is a navigation bar with the JVO logo (Japanese Virtual Observatory) and links for Top, Search, VOServices, Subaru, Analysis, Bookmark, and JVOSpace. The user is logged in as Masatoshi Ohishi. The main content area is divided into several sections:

- News:** A list of recent updates, including Suprime-Cam mosaic image updates and new data services like JVO Sky.
- Service Contents:** A central menu with sub-sections:
 - Data Search:** Includes Quick Search, Single VO Service, Multiple VO Services (circled in red), JVO Sky (circled in red), Xmatch Search, and JVOQL Search.
 - Subaru:** Includes Suprime-Cam and HDS.
 - Surveys:** Includes Subaru Deep Survey and IRSF Survey.
 - JVO Space:** Includes Home.
- Service Search:** Includes Keyword Search, Category Search (Auto), Category Search (Manual), and Advanced Search.
- Astronomical Tools:** Includes Source Extractor and HyperZ.
- Bookmark:** Includes Bookmark of VOService and Bookmark of JVOSpace.
- Admin:** Includes Admin.

At the bottom, there is a contact information section with the JVO logo and the email address help_desk@jvo.nao.ac.jp.

Toward All Sky Data Query



JVO JAPANESE VIRTUAL OBSERVATORY p00 ver.100118 News | FAQ(J) | Help(J) | Bugs(J) Masatoshi Ohishi
=> Location: Top Page > JVO Sky

Subaru SuprimeCAM

JVOSky: footprints
by means of
Google sky API

→ Easy to find
needed images

Next step: select
multiple images to
query, increase
contents other than
SuprimeCAM

JVO Sky

all object name image Suprime-Cam/Subaru

Coodinate or ObjectName : M31 (ra, dec)=(10217, 41714)=(0h40m52.15s, 41d42m49.9s)

OK

▼ spcam

- 00h37m53+41d56m24 (Minor_1 W-C-IC)
- 00h39m31+41d46m17 (Minor_1 W-C-IC)
- 00h38m13+41d57m41 (Minor_1 W-J-V)
- 00h40m42+41d40m16 (Minor_0 W-C-IC)
- 00h38m49+41d50m42 (Minor_0 W-C-IC)
- 00h39m21+41d47m08 (Minor_0 W-J-V)
- 00h40m04+41d37m54 (Minor_0 W-J-V)
- 00h40m43+41d34m05 (M31_C W-C-IC)
- 00h44m26+41d36m19 (M31_1 W-S-Z+)
- 00h44m45+41d43m11 (M31_0 W-S-Z+)
- 00h42m45+41d16m53 (M31-North W-C-)
- 00h41m02+41d29m52 (M31-North W-C-)
- 00h42m45+41d29m01 (M31-North W-C-)
- 00h42m07+41d01m42 (M31 W-C-RC)
- 00h39m49+41d43m59 (Minor_1 W-J-V)
- 01h14m54+18d58m17 (STRING_COC0 W-
- 01h16m13+19d18m04 (STRING_COC0 W-

科学衛星データの世界への発信

天文学

太陽物理学

太陽地球系物理学

月惑星科学

よくある質問

天文学

- MAXI

- あかり

- すざく

- はるか

- あすか

- IRTS

- ぎんが

- てんま

- JUDO

太陽物理学

- ひので

- ようこう

太陽地球系物理学

- わいめい

- ジオテイル

- あけぼの

- テミス (ミラー)

- CEF

今月のDARTS

一般向け解説

▶ [English](#)

DARTS (Data ARchives and Transmission System)は、天文学、太陽物理学、および、太陽地球系物理学といった多分野にわたる宇宙科学のデータアーカイブです。主に日本の科学衛星によって取得したデータを公開しています。DARTSは、宇宙航空研究開発機構・宇宙科学研究本部、科学衛星運用・データ利用センター (C-SODA) が運用しています。

▶ お知らせ

<http://darts.jaxa.jp/index.html.ja>

(30 April 2010) [MAXIに関するページ](#)がDARTS/天文に加われました。

(30 March 2010) 「あかり」赤外線天体カタログ (1st version)が公開されました。[AKARI CAS](#)を使ってカタログを検索してください。

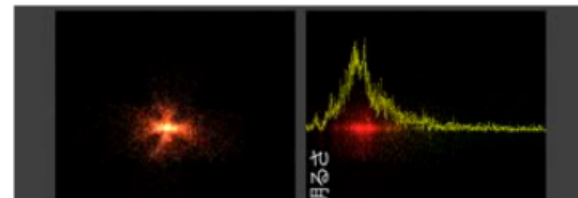
(30 March 2010) JUDOに[Swift 全天画像](#)が追加されました。

[過去のお知らせ](#)

▶ 今月のDARTS

X-ray movies

X線天文衛星は、2次元座標の他に、エネルギーと時刻をあわせて、4次元の観測データを得る事ができます。例えばUDONで、[みずがめ座AE星を表示してみます](#)。「すざく」の観測データをもとに、論文などでおなじみの「スペクトル」

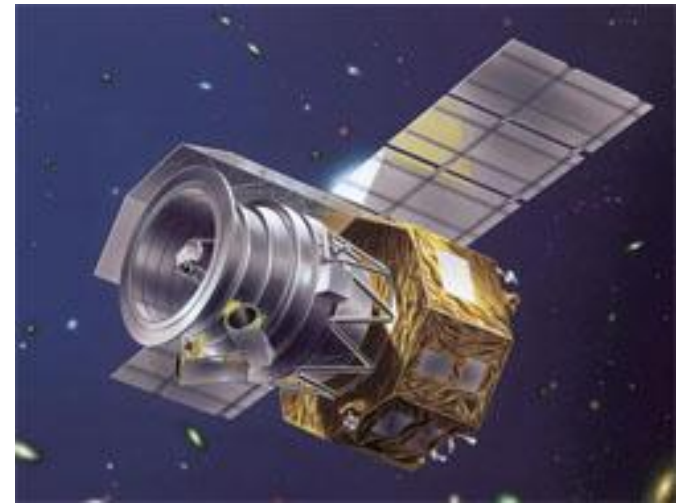




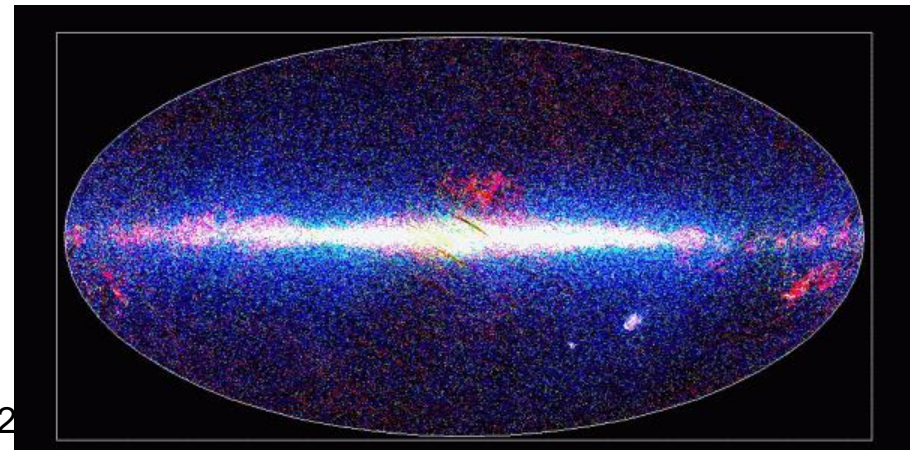
AKARI catalogues public
release from DARTS
(March 30)

- IRC bright source
9/18 μm
- FIS bright source
65/90/140/160 μm

Released from JVO on
May 10



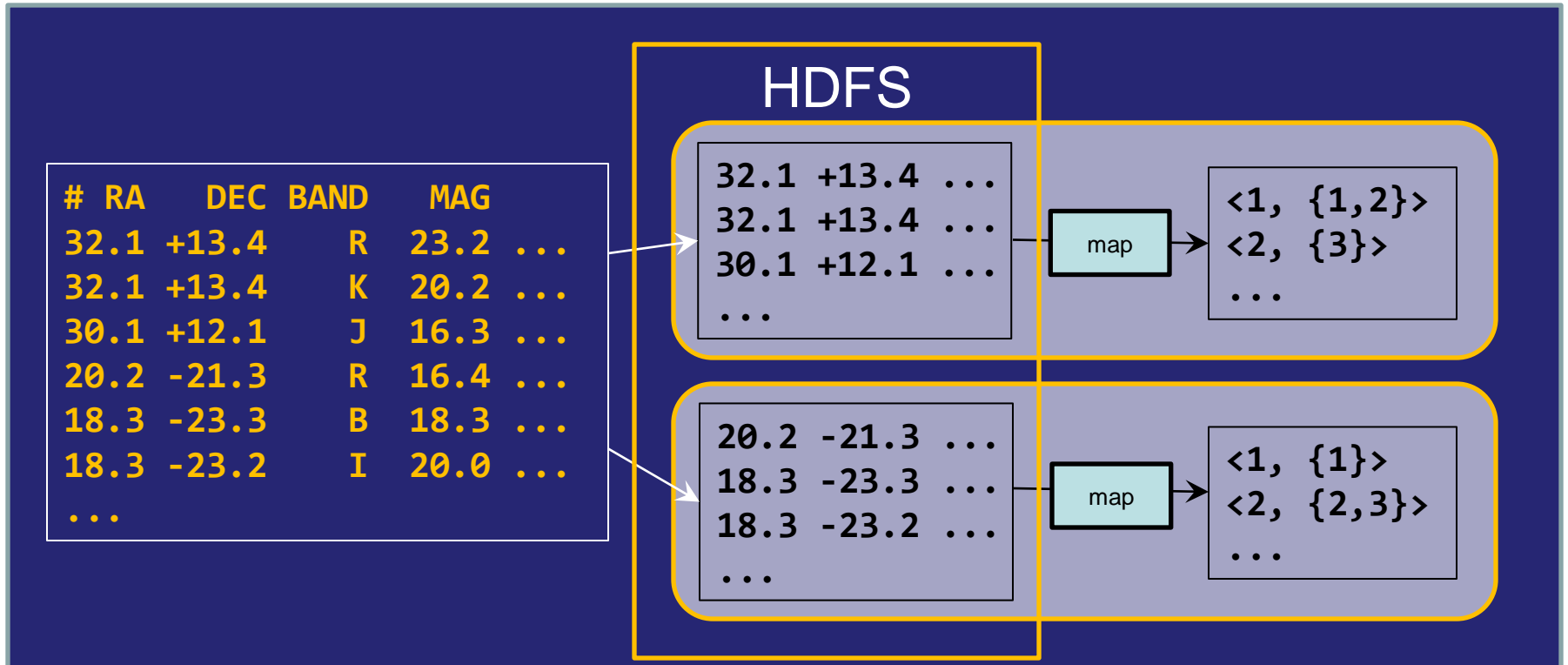
9 – 160 μm , 68.5 cm telescope



100億レコードのクロスマッチ



- 様々なカタログの天体をクロスID
- Hadoop を利用した分散処理で実行
- 180日かかる処理を3日で実現



CSI委託事業 平成21年度の目標



- 宇宙科学研究所に NAREGI 拠点を設置
 - Naokek VO (国立天文台、KEK、広島工業大)
- データグリッド (Gfarm) の試験利用
 - 国立天文台、宇宙科学研究所間で Gfarm を利用した分散ファイルシステム
 - ファイアーウォール越え試験
- 100億レコードからなる天体データベースのクロスマッチ
 - Hadoop 実装との比較

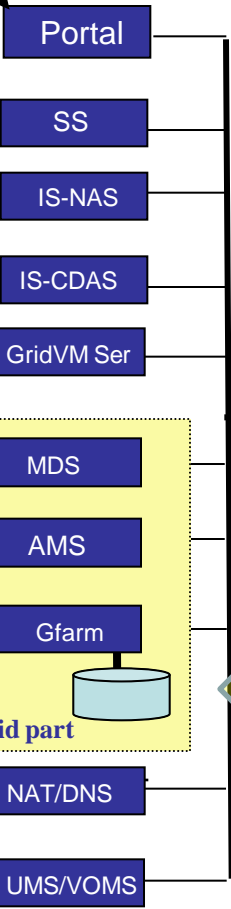
DataGrid Testbed between NAOJ and ISAS

March 2010

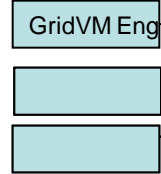


User Browsers

NAOJ Site



Private network



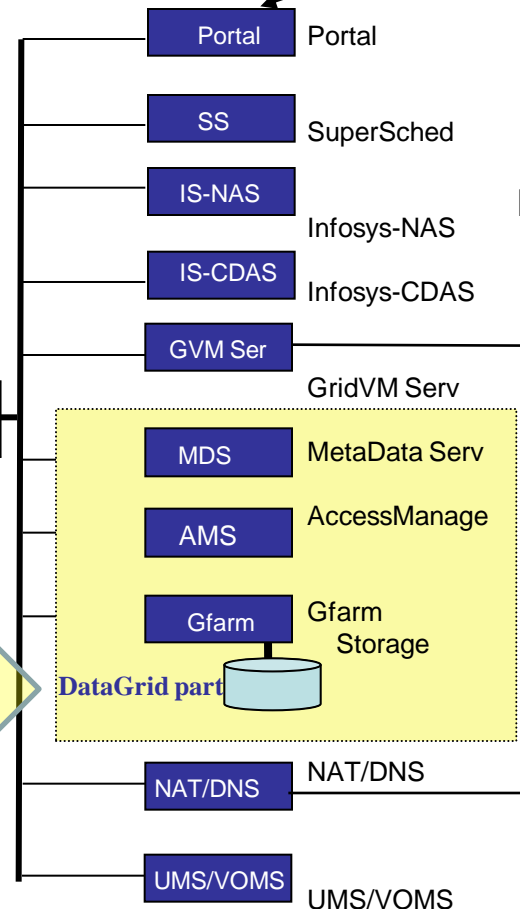
Compute nodes

DataGrid part

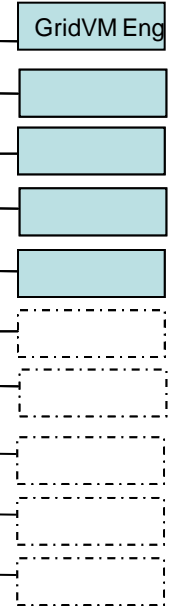


User Browsers

ISAS / C-SODA Site



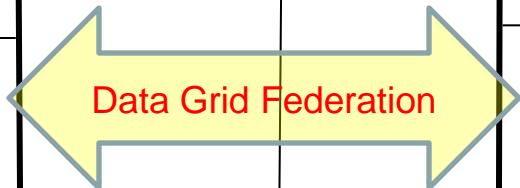
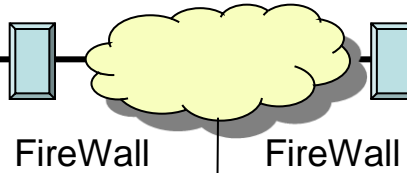
Private network



Compute nodes

DataGrid part

SINET3



VOMS/KEK

Test VO: naokek

データグリッド構築結果



- 利用するポートを特定し、ファイアーウォールの穴あけを行い、疎通は成功した。
- 解析ノードのディスクを gfarm で共有できない？
 - 解析データは解析するノードにあらかじめある方が、データ転送にかかる時間を短縮できる。
 - 解析ノードはプライベートアドレスであるため、異なるLAN 間ではディスク共有ができないようだ。

```
yshirasa@naregi-gridvm:~  
ファイル(E) 編集(E) 表示(V) 端末(T) タブ(B) ヘルプ(H)  
[yshirasa@naregi-gridvm duniv]$ cd  
[yshirasa@naregi-gridvm ~]$ gdfd  
  1K-blocks      Used      Avail Capacity  Host  
    18156292    17092232    126896     99%  naregi-gfarm.gc.isas.jaxa.jp  
  17576040704  186941216 17389099488     1%  naregi-gridvm.mtk.nao.ac.jp  
[yshirasa@naregi-gridvm ~]$
```

Workflow Tool を使って実行



The screenshot shows the Workflow Editor interface. At the top, there is a menu bar (File, Edit, View, Window) and a title bar indicating the user is 'yshirasa'. Below the menu is a grid of 10 yellow job icons, each labeled with a job ID from c58000 to c58009. Each icon has a small circle and an 'E' button. Below the grid is a 'Property' dialog box for job 'c58007'. The dialog has two tabs: 'By Menu' and 'By JSDL'. The 'By JSDL' tab is active, showing a table of job specifications. The 'Argument' field is highlighted with a red box and contains the value 'merge_58007.psv.gz'. Below the table is an 'Input/Output' section with a table for defining symbols and files. At the bottom of the dialog are 'OK' and 'Cancel' buttons.

Name	Detail	Value
JobName	-	Program
Executable	-	/home/ys Shirasa/vo_dev/build/linux-x86_64/bin/...
Argument	1	merge_58007.psv.gz
Output	-	out.dat
Error	-	err.dat
WorkingDirectory	-	/tmp/ys Shirasa
Environment	-	
WallTimeLimit	-	300
MemoryLimit	-	
CPUTimeLimit	-	
VirtualMemoryLimit	-	
CandidateHosts	-	
OperatingSystemName	HostName	naregi-gridvm.mtk.nao.ac.jp
IndividualCUPUCount	-	
TotalResourceCount	-	

Symbol	Type	File
1		
2		
3		
4		
5		

- 100億レコードを 10,049ファイルに分割
- 各ファイルを分散処理
- 今回は10ファイル分のみ。

結果



- 実行時に Gfarm 上のファイルが読めなくなるとい
う問題が発生。
 - 実行前には gfexport で読めていたが、実行後に読
めなくなった。
 - 読めないけれども、gfls ではファイルはリストされる。
 - 直接解析ノードにログインしてスクリプトを実行すると
成功する。
- 現在原因調査中

平成 21 年度の成果



- 宇宙科学研究所を NAREGI 拠点として NAOKEK VO に組み込むことに**成功**
- 国立天文台、宇宙科学研究所間でのファイアーウォール越えでのデータグリッド環境構築に**成功**
- データグリッドを利用した「100億レコードクロスマッチ」は成功にはいたっていない。
 - 実行前には読めていたファイルが、実行中に読めなくなる。実行後には読めない。

ヴァーチャル天文台(JVO) におけるNAREGI活用案

- H21
NAREGI 1.0によるJVO (NAOJ三鷹)-JAXA/ISASの連携
PSE / Data Grid
- H22
NAREGI 1.0のNAOJハワイ,
NAOJ野辺山, などへの展開,
資源の相互利用
- その後
NAREGIと欧米のGridミドルウェア連携を踏まえたJVOと他国VO間のグリッド環境連携

