Data Processing for SUBARU Telescope using GRID

Yuji Shirasaki
National Astronomical Observatory of Japan
Astronomy Data Center
on behalf of the JVO Project team

Contents

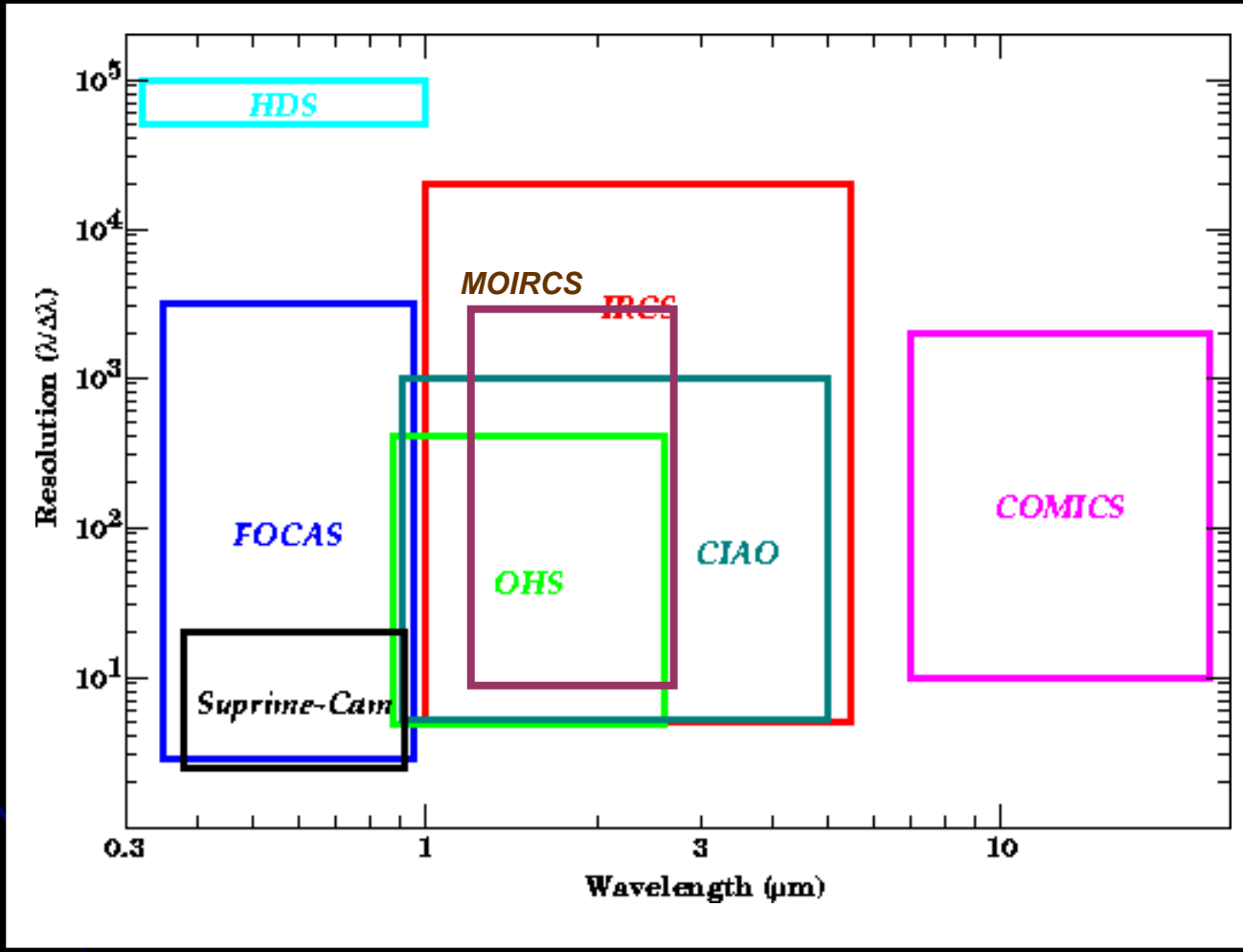
- Subaru Telescope and Instruments
- Virtual Observatory
- CCD Image Reduction
- Subaru GRID Data Analysis System

Subaru Telescope

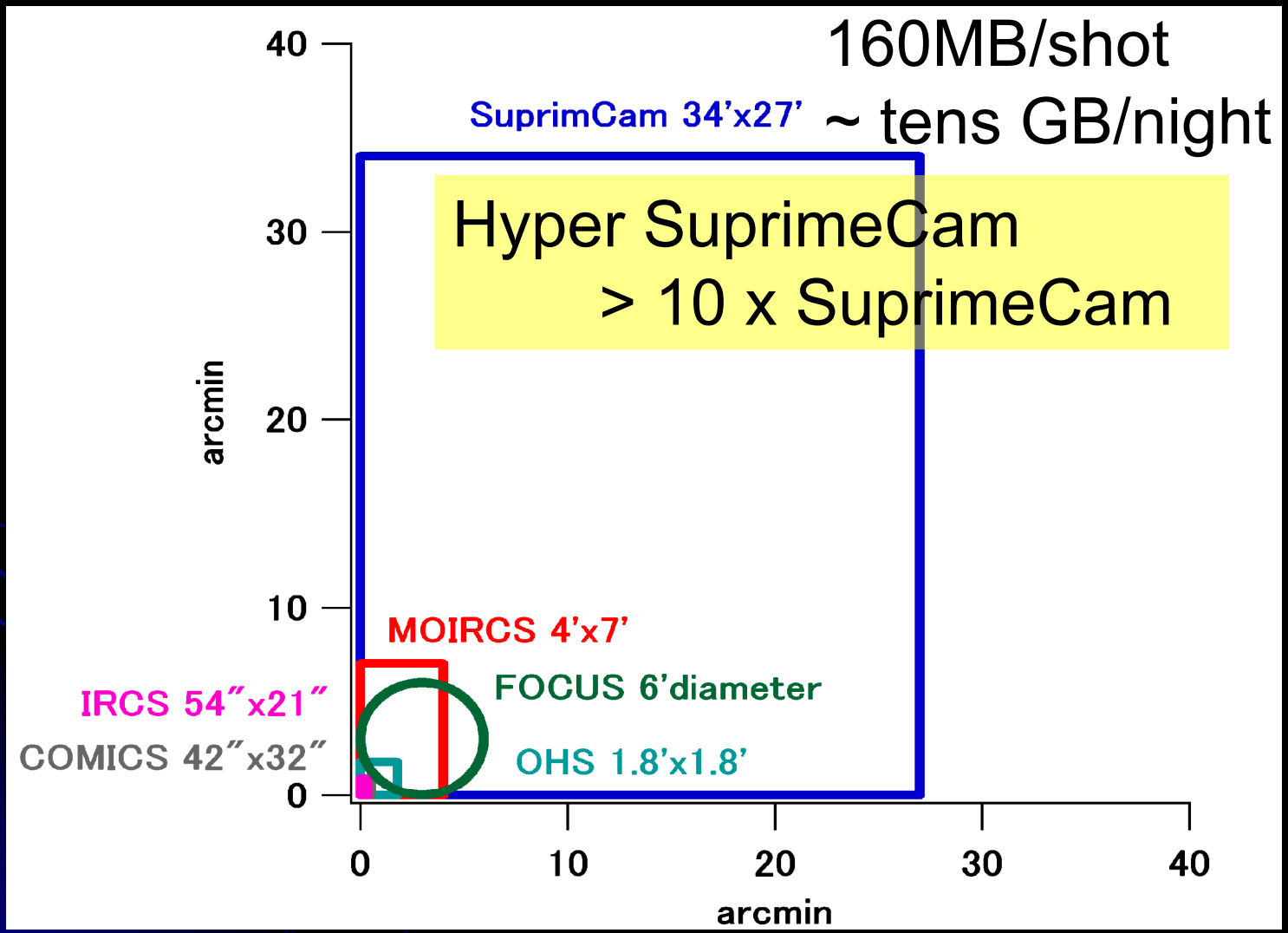
Subaru Telescope is an optical-infrared 8.2 m telescope at Hawaii, operated by NAOJ.

Instruments for Subaru

- CIAO
- COMICS
- FOCUS
- IRCS
- MOIRCS
- SuprimeCam
- HDS
- AO/CIAO



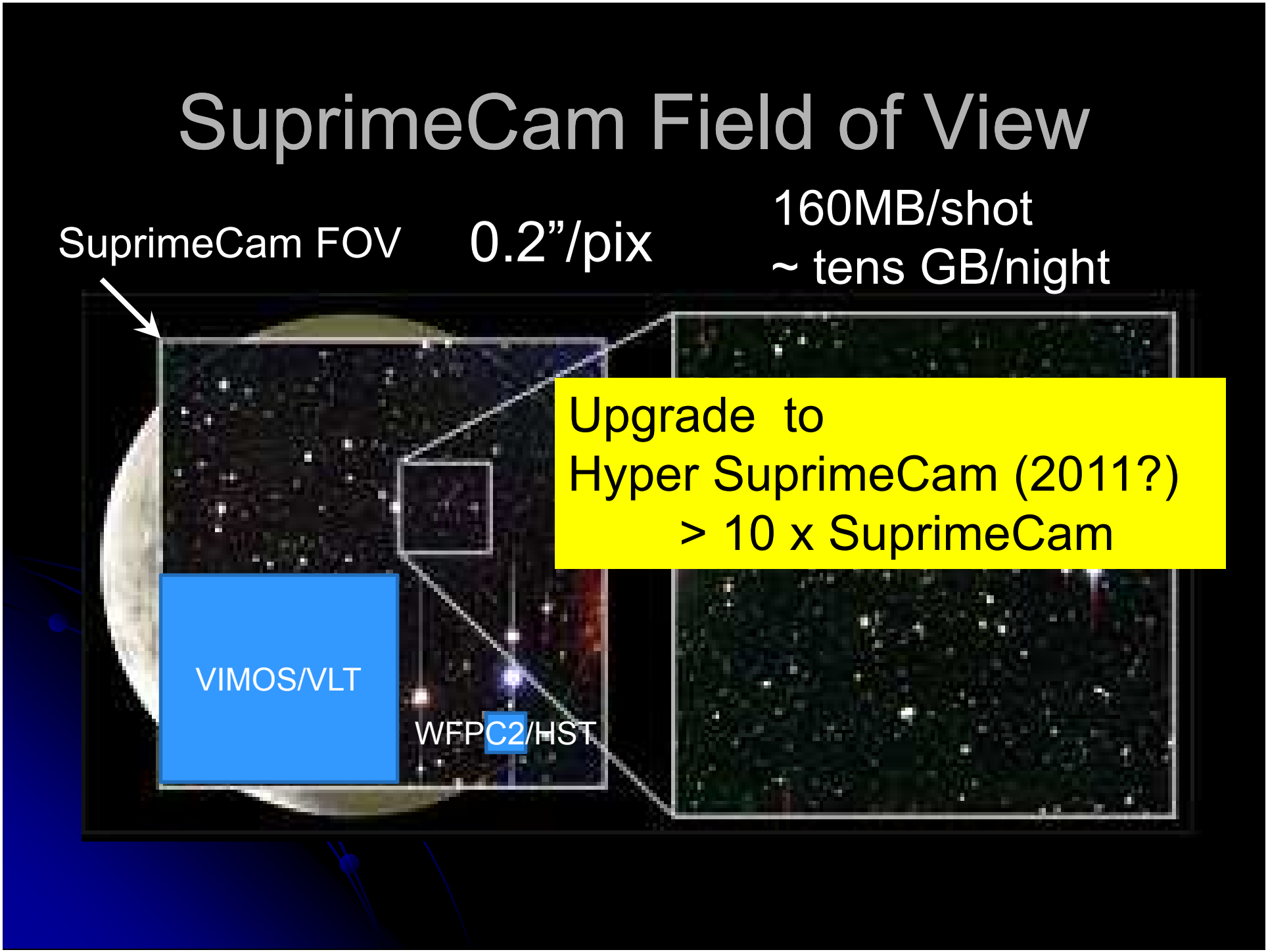
Instruments' Field of View



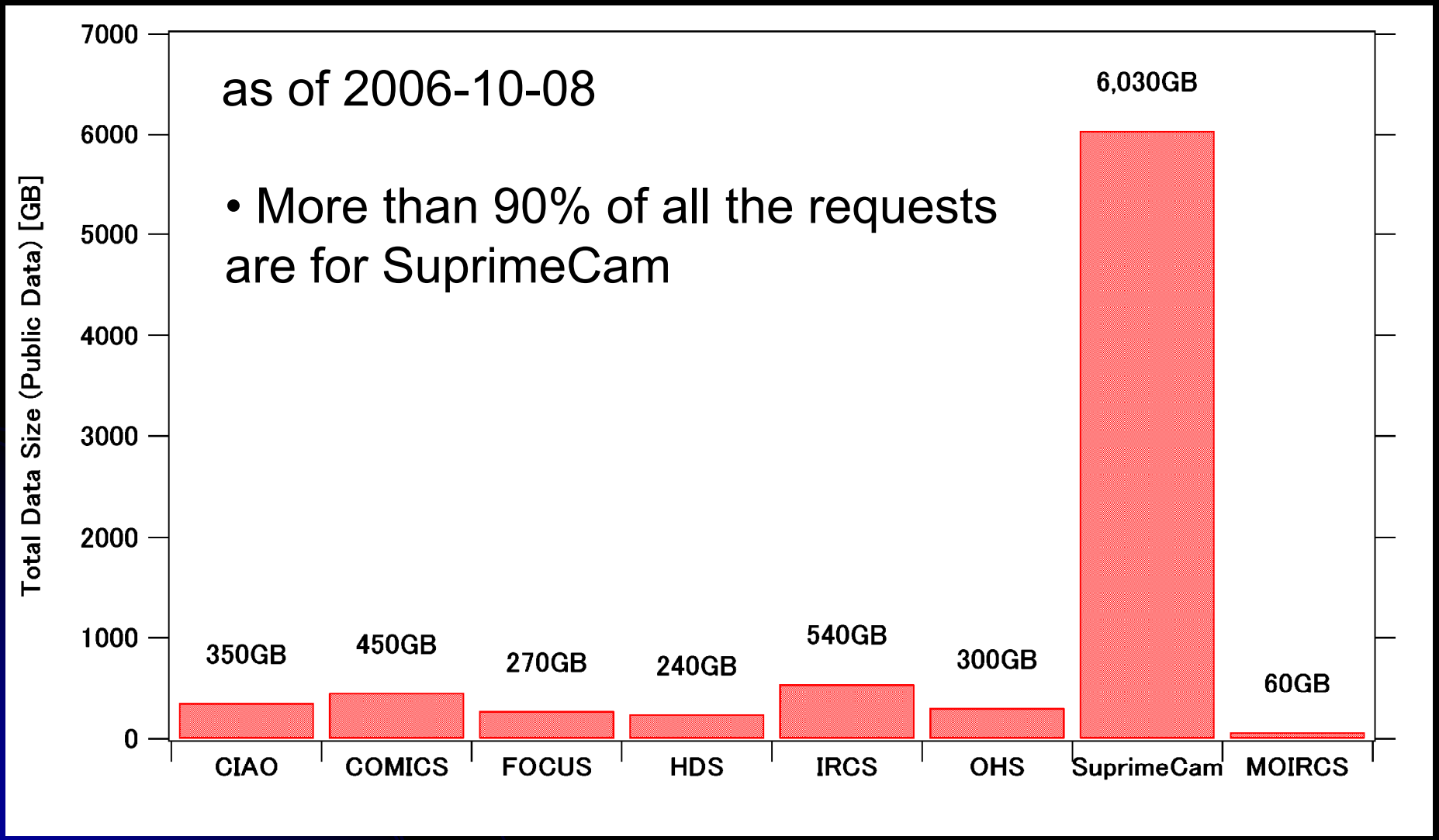
Suprime-Cam

Size: 1035 mm x 960 mmϕ
Weight: 295 kg
Power: 420 W
F: 1.86
FOV: 30' ϕ
CCD: MIT/LL CCID20
Format: 4096x2048 per CCD
# of CCD: 10 (2 x 5)
Filter: B,V,Rc,Ic,g',r',I',z'

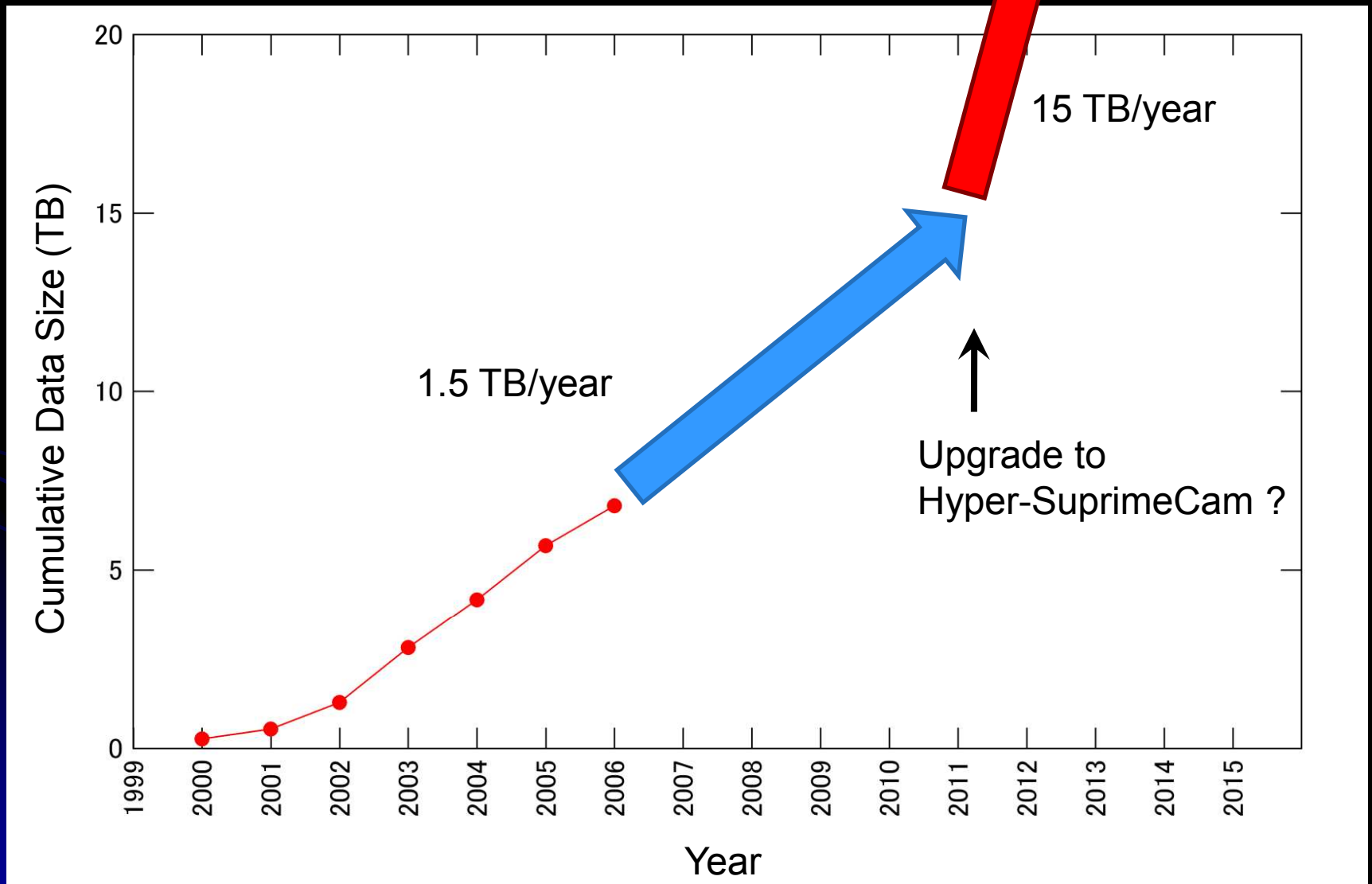
SuprimeCam Field of View



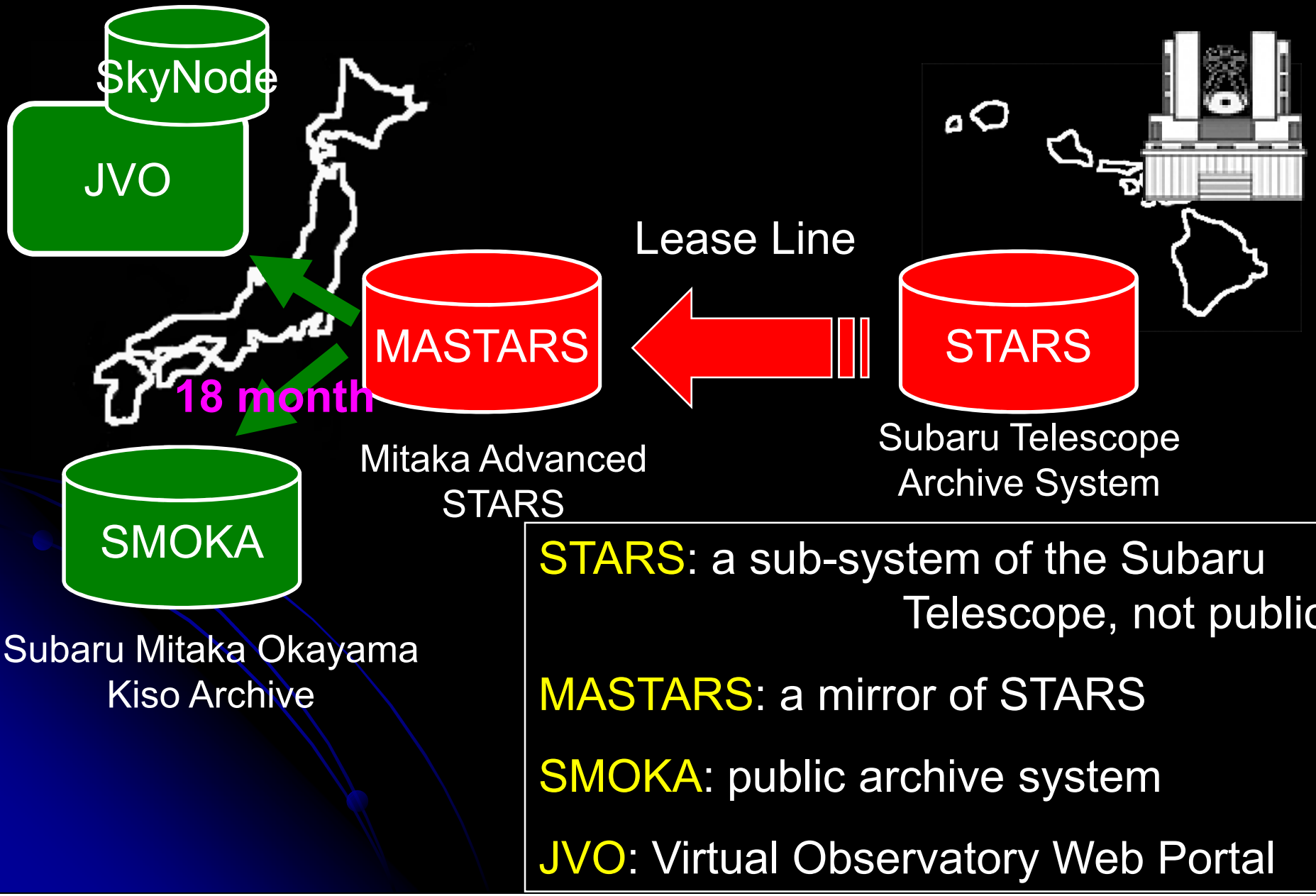
Total amount of public data



Cumulative Data Volume of SuprimeCam



Subaru Data Archive

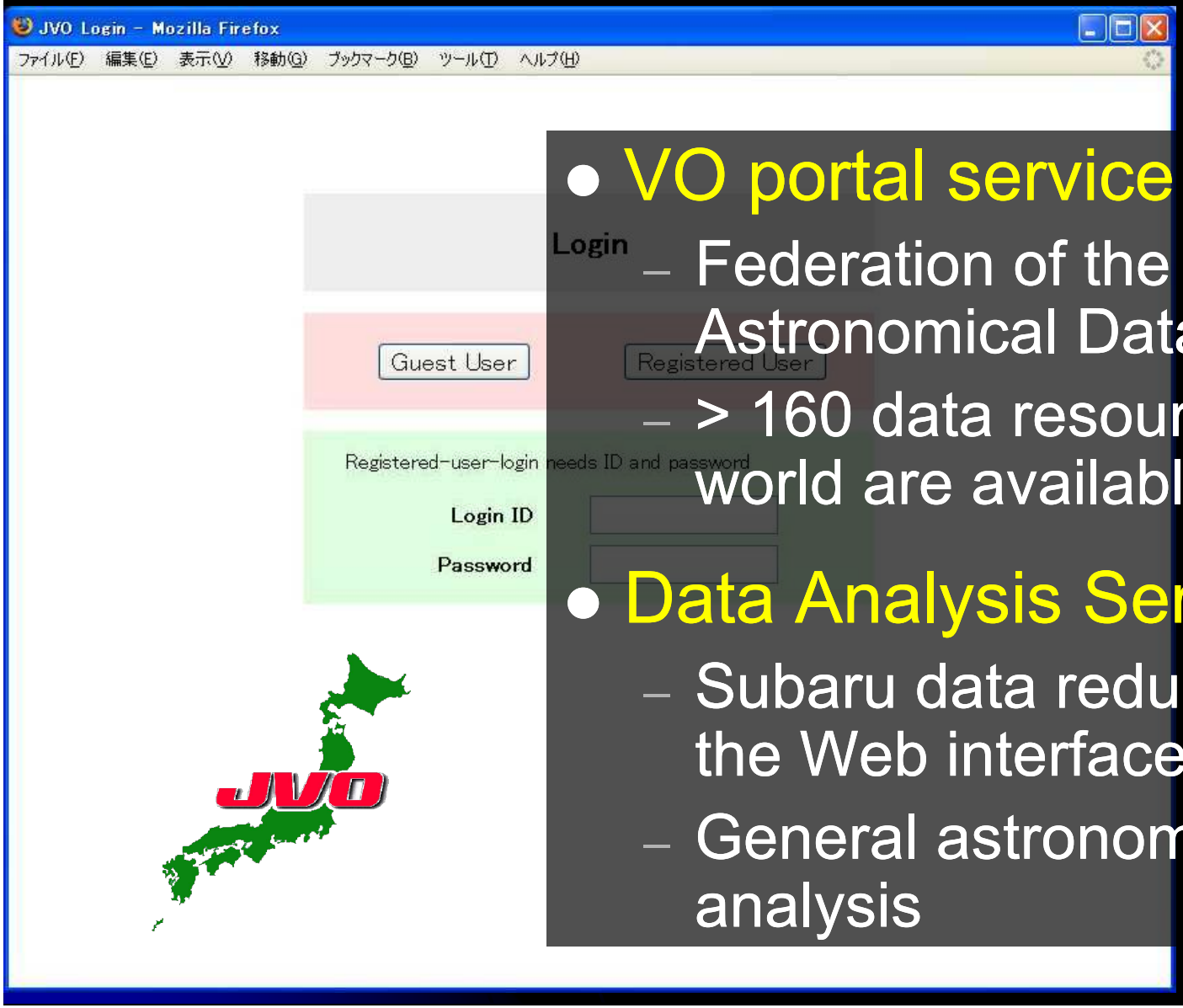


International Virtual Observatory Alliance (IVOA)

- 2002 Jun ~
- 16 VO projects from Armenia, Australia, Canada, China, Europe, France, Germany, Hungary, India, Italy, Japan, Korea, Russia, Spain, the United Kingdom, and the United States.
- Uniform access to the distributed databases
  → Standard of data access interface

Japanese Virtual Observatory (JVO)

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



- **VO portal service**
  - Federation of the Distributed Astronomical Databases
  - > 160 data resources from the world are available
- **Data Analysis Service**
  - Subaru data reduction through the Web interface
  - General astronomical data analysis

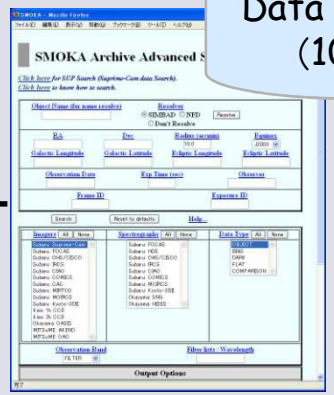
Current situation in astronomy

Manual (Instr.A)

Manual (Instr.B)

Manual (Instr.C)

Subaru Telescope Data Archive (10TB)

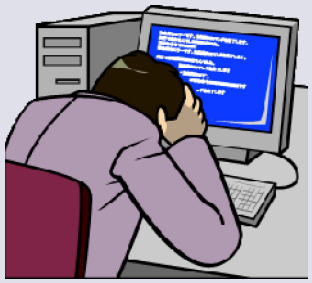


Download by hand

Software for instr. A

Software for instr. B

Software for instr. C



Subaru Data Analysis and Virtual Observatory

User disk space

Subaru Telescope Data Archive (10TB)

4. Retrieve reduced data

3. Save the reduced data

1. Request Data

JVO Portal Service

2. Reduce raw data

Necessity of the Resource Aggregation

- Limited Data transfer rate

  Bandwidth between the Data Center and each institute is still below 100 MB/s.

- Improvement of computing power is saturating.

  Just upgrading the CPU does not necessarily shorten the computing time.

- Computing power/cost is still rapidly increasing

  e.g. Core 2 Quad 2.4GHzx4 US$600 (Now)

  Pen4 1.8GHzx1 US$600 (2001)

- Progress on the software technology

  Computing Grid / Data Grid /…

Pros and Cons of Server Side Processing

- Pros:
  - User doesn't need to take care about installation / update of analysis software
  - User doesn't need to have a large storage
  - User doesn't need to have a high-spec CPU
- Cons:
  - User cannot modify the software

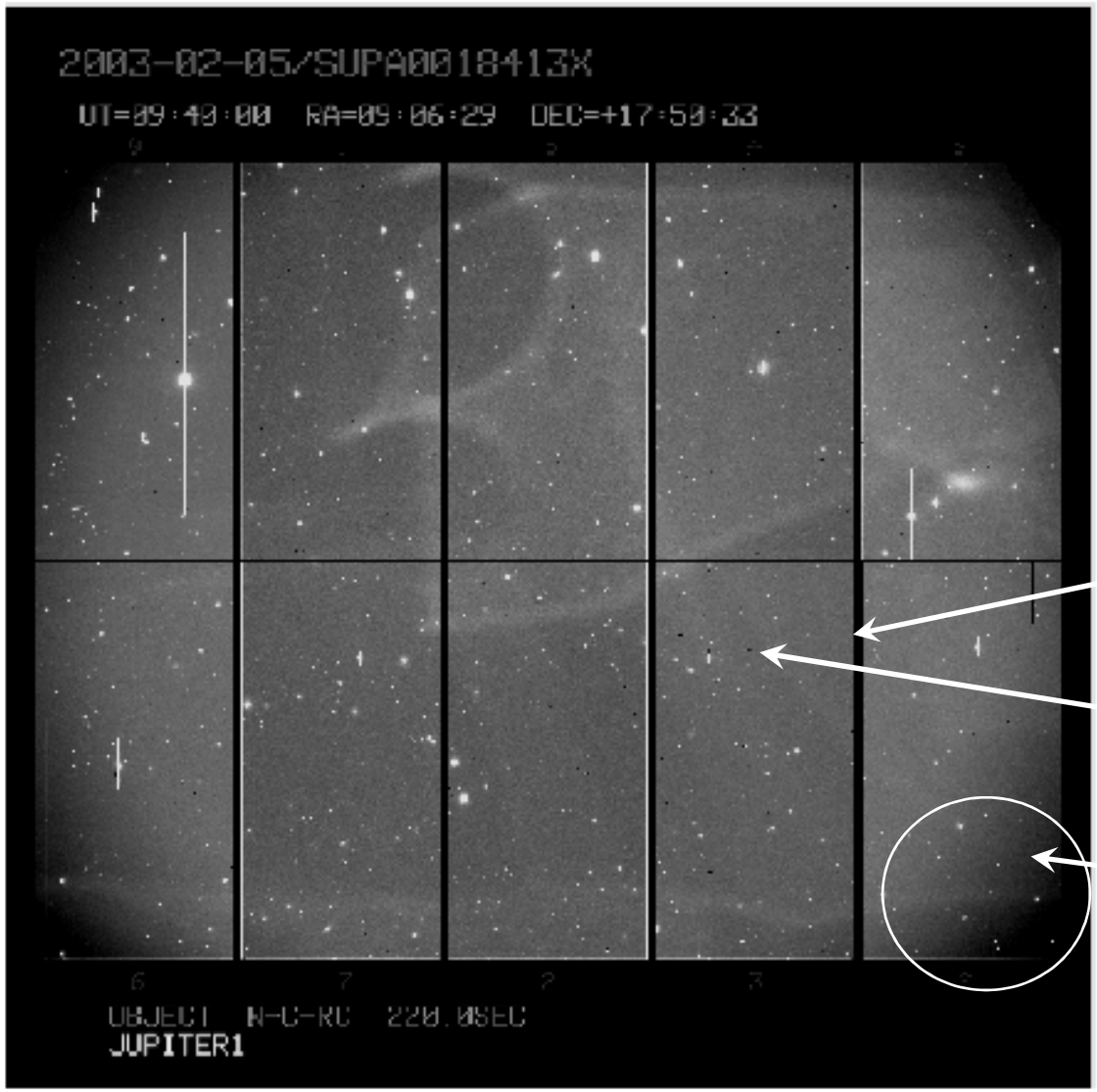
    The primary target of this system is for novice user
  - If unexpectedly large number of users submit jobs at the same time, the computing time slow down.

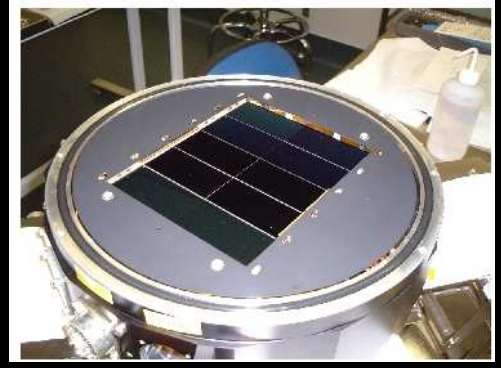
    No resolution, be patient. It may be still faster than preparing all the resource in the local machine ?

Data Analysis in Astronomy

1. Raw data reduction
   - Very complex : The reduction procedure depends on instrument. Require the knowledge on the instrument.
   - Time consuming : 10GB of data must be used to get one image (300MB). It takes half a day to reduce data taken in one night (SuprimeCam).
2. Parameter measures from reduced data
   - Relatively easy : Many kinds of open software is available
   - Problem: need to find, install, and know the software
3. Interpretation of the derived parameters
   - The way of the analysis depends on each science case.

SuprimeCam CCD



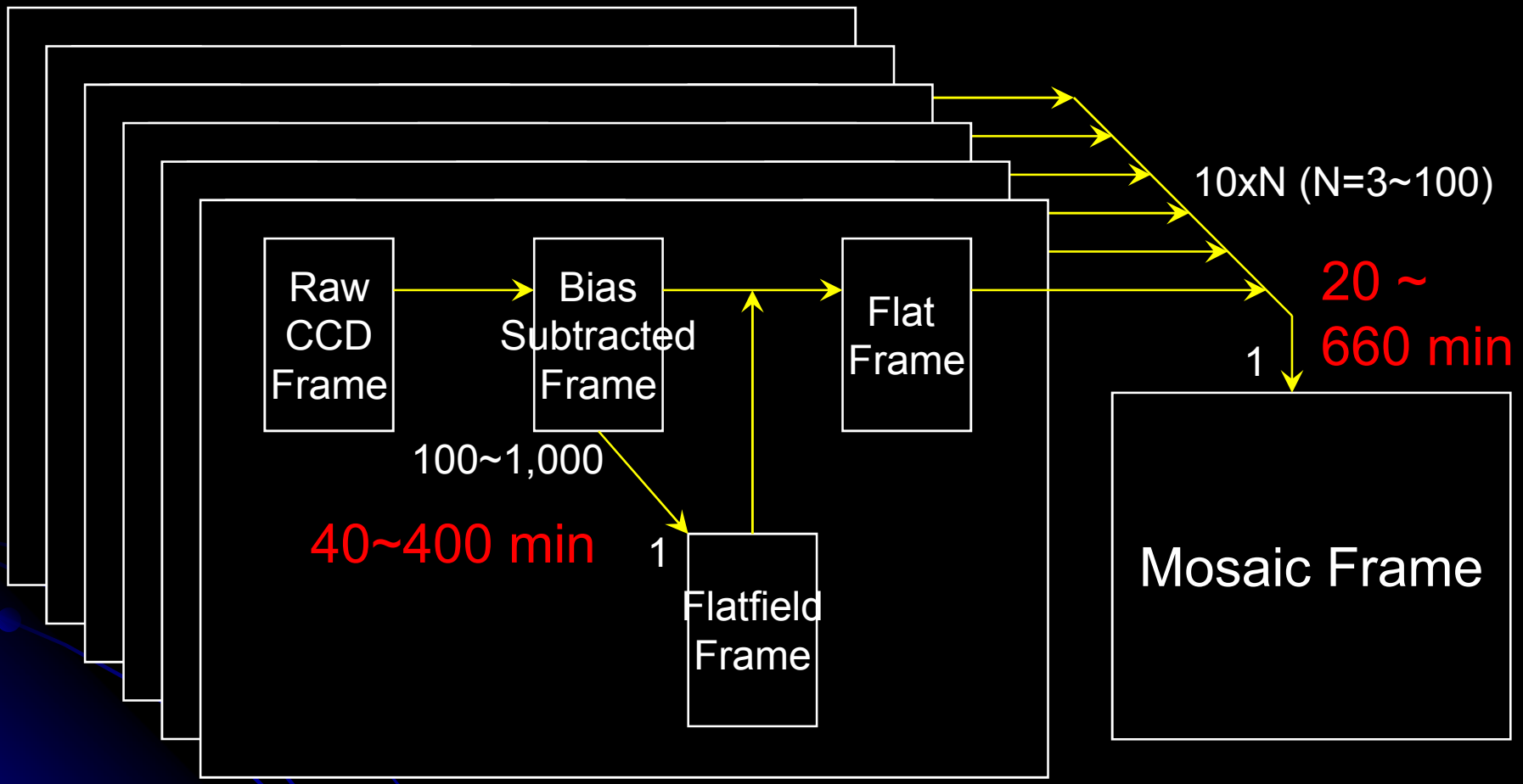


CCD gap

Bad pixel /
Cosmic Ray

Non-uniformity

CCD Image Calibration / Reduction

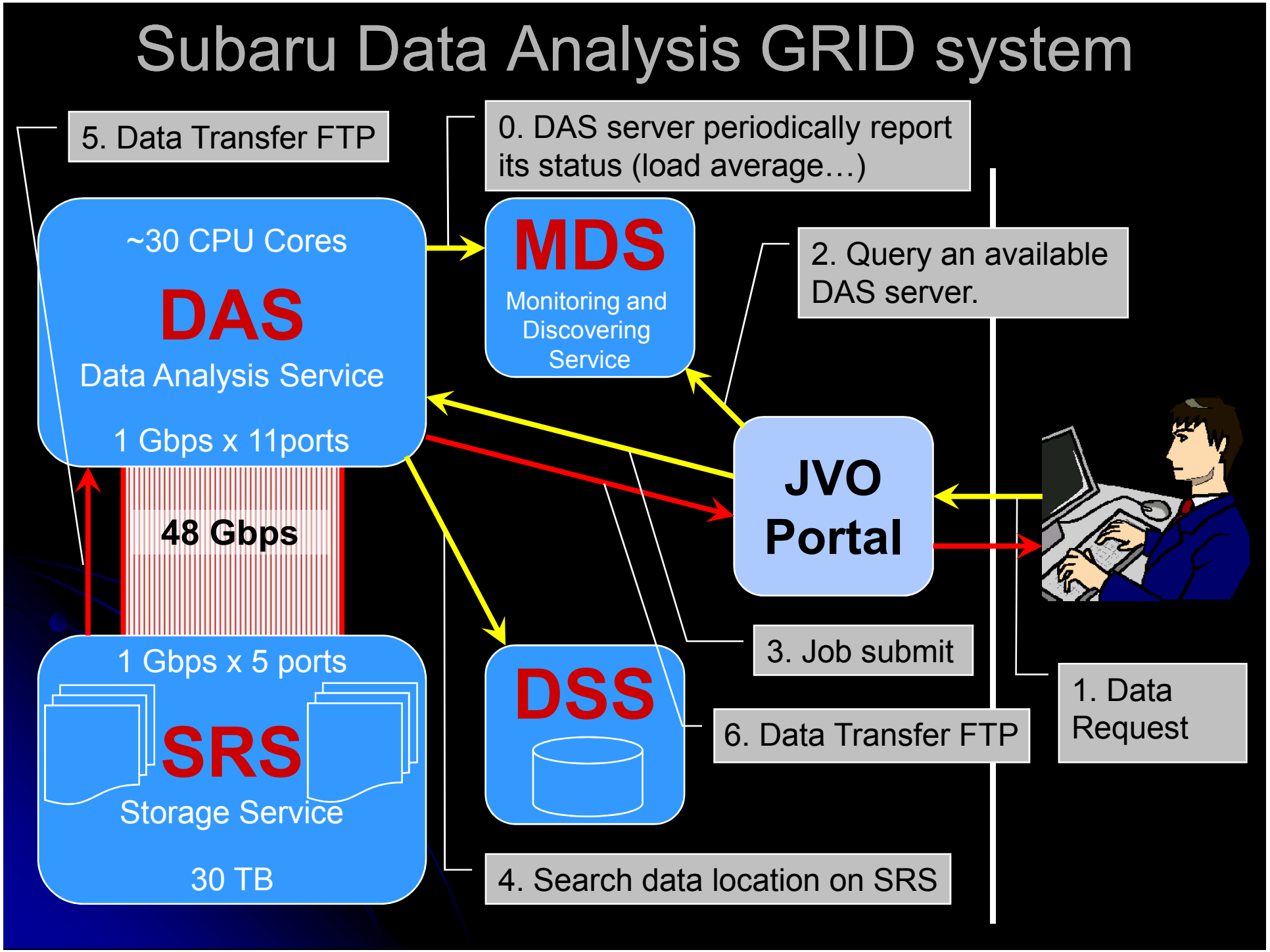


One day for each observation night

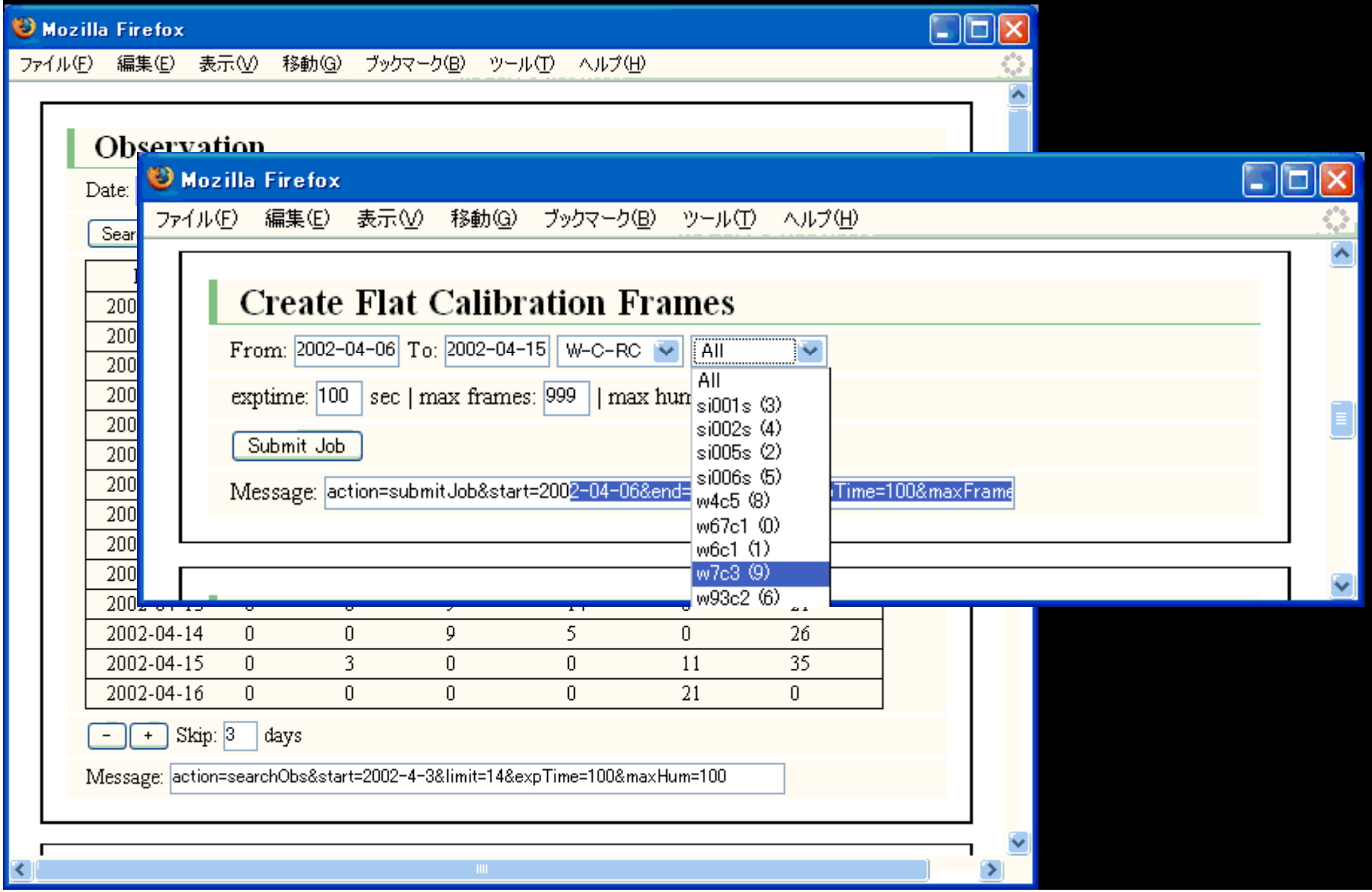
Components of the Grid system

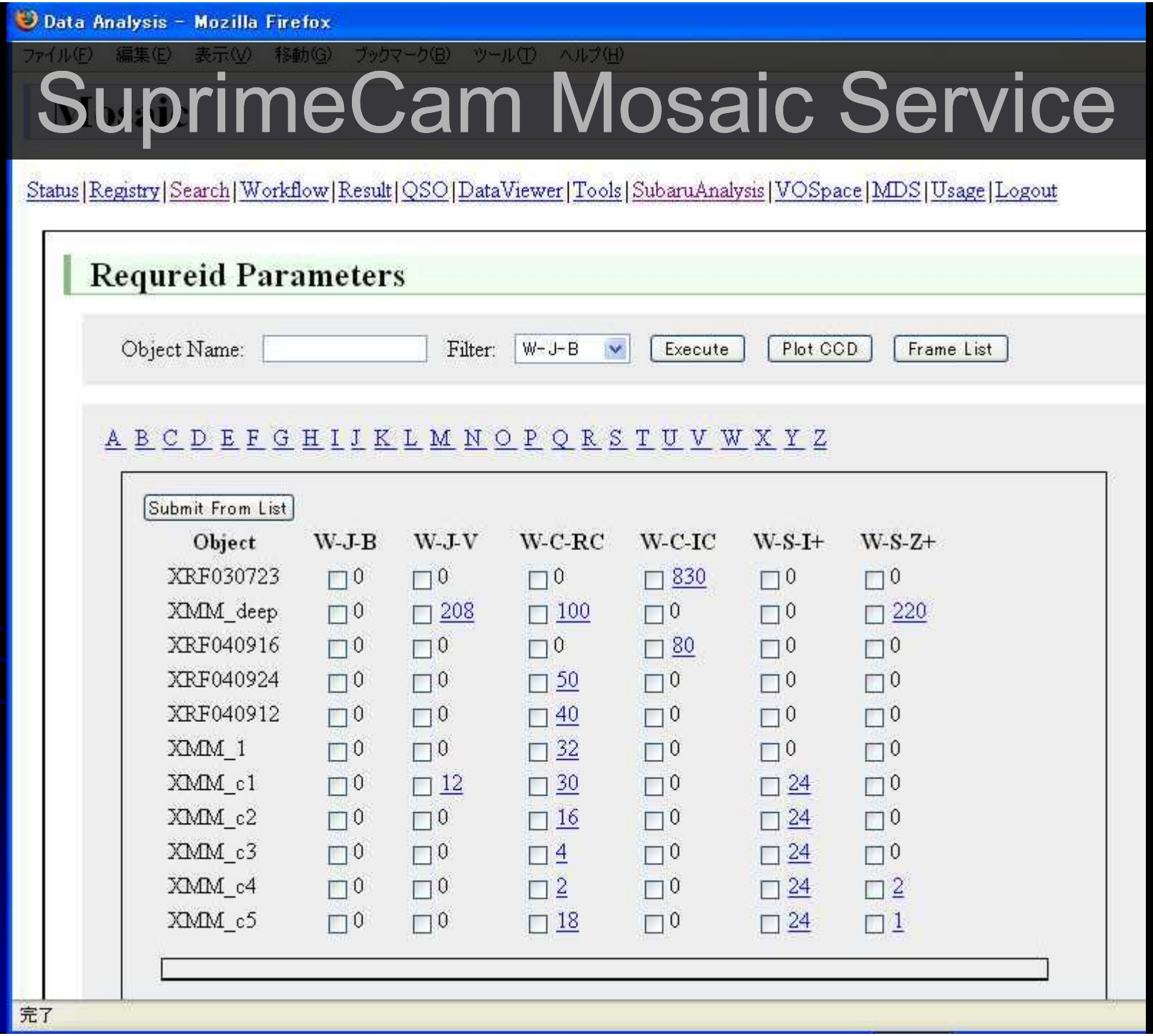
- Computing resource management
  - Monitoring and Discovery Service (MDS)
- Data transfer
  - FTP or HTTP get
- Remote Job execution
  - Web service (Tomcat + AXIS)
- All the components were developed by ourselves
  - Plan to introduce NAREGI GRID middleware

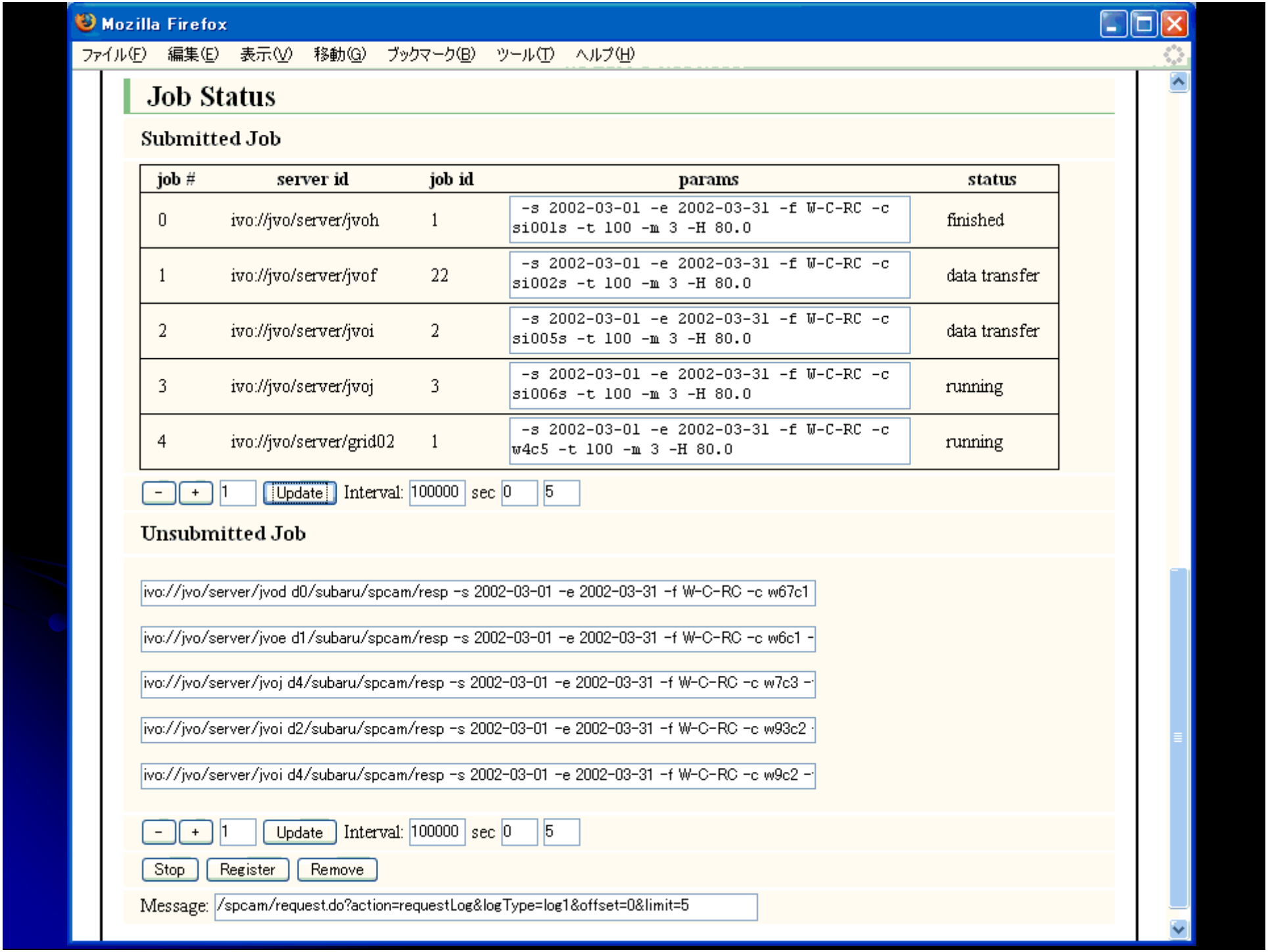
Subaru Data Analysis GRID system

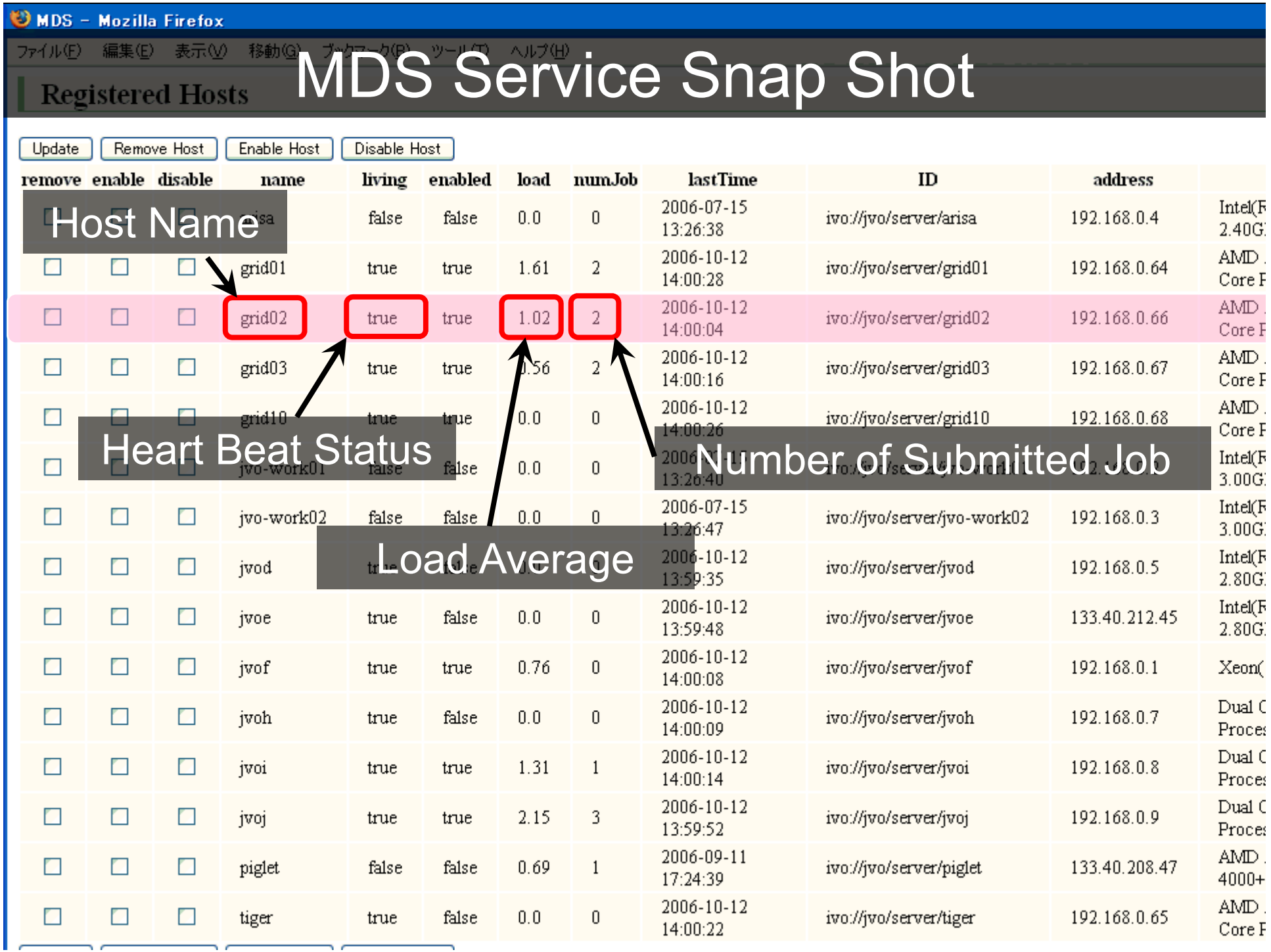


Flat-field frame calculation GUI



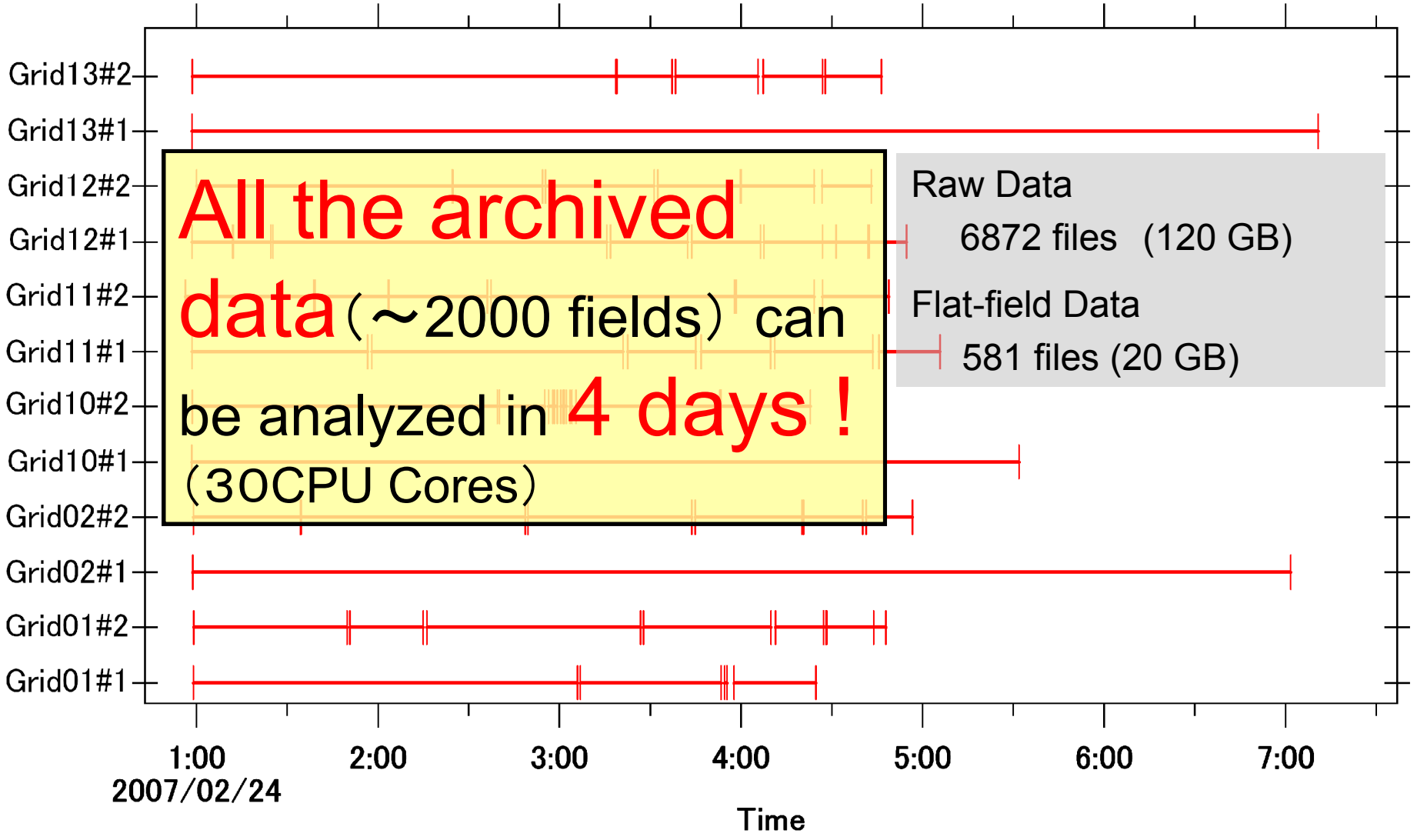

Data Analysis - Mozilla Firefox
ファイル(F) 編集(E) 表示(V) 移動(G) ブックマーク(B) ツール(T) ヘルプ(H)
SuprimeCam Mosaic Service
Status | Registry | Search | Workflow | Result | QSO | DataViewer | Tools | SubaruAnalysis | VOSpace | MDS | Usage | Logout
Requreid Parameters
Object Name:
Filter: W-J-B
Execute
Plot CCD
Frame List
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Submit From List
Object W-J-B W-J-V W-C-RC W-C-IC W-S-I+ W-S-Z+
XRF030723 0 0 0 830 0 0
XMM_deep 0 208 100 0 0 220
XRF040916 0 0 0 80 0 0
XRF040924 0 0 50 0 0 0
XRF040912 0 0 40 0 0 0
XMM_1 0 0 32 0 0 0
XMM_c1 0 12 30 0 24 0
XMM_c2 0 0 16 0 24 0
XMM_c3 0 0 4 0 24 0
XMM_c4 0 0 2 0 24 2
XMM_c5 0 0 18 0 24 1
完了


Mozilla Firefox
ファイル(F) 編集(E) 表示(V) 移動(G) ブックマーク(B) ツール(T) ヘルプ(H)
Job Status
Submitted Job
job # | server id | job id | params | status
0 | ivo://jvo/server/jvoh | 1 | -s 2002-03-01 -e 2002-03-31 -f W-C-RC -c si001s -t 100 -m 3 -H 80.0 | finished
1 | ivo://jvo/server/jvof | 22 | -s 2002-03-01 -e 2002-03-31 -f W-C-RC -c si002s -t 100 -m 3 -H 80.0 | data transfer
2 | ivo://jvo/server/jvoi | 2 | -s 2002-03-01 -e 2002-03-31 -f W-C-RC -c si005s -t 100 -m 3 -H 80.0 | data transfer
3 | ivo://jvo/server/jvoj | 3 | -s 2002-03-01 -e 2002-03-31 -f W-C-RC -c si006s -t 100 -m 3 -H 80.0 | running
4 | ivo://jvo/server/grid02 | 1 | -s 2002-03-01 -e 2002-03-31 -f W-C-RC -c w4c5 -t 100 -m 3 -H 80.0 | running
- + 1 Update Interval: 100000 sec 0 5
Unsubmitted Job
ivo://jvo/server/jvod d0/subaru/spcam/resp -s 2002-03-01 -e 2002-03-31 -f W-C-RC -c w67c1
ivo://jvo/server/jvoe d1/subaru/spcam/resp -s 2002-03-01 -e 2002-03-31 -f W-C-RC -c w6c1 -
ivo://jvo/server/jvoj d4/subaru/spcam/resp -s 2002-03-01 -e 2002-03-31 -f W-C-RC -c w7c3 -
ivo://jvo/server/jvoi d2/subaru/spcam/resp -s 2002-03-01 -e 2002-03-31 -f W-C-RC -c w93c2 -
ivo://jvo/server/jvoi d4/subaru/spcam/resp -s 2002-03-01 -e 2002-03-31 -f W-C-RC -c w9c2 -
- + 1 Update Interval: 100000 sec 0 5
Stop Register Remove
Message: /spcam/request.do?action=requestLog&logType=log1&offset=0&limit=5



remove	enable	disable	name	living	enabled	load	numJob	lastTime	ID	address	
			arisa	false	false	0.0	0	2006-07-15 13:26:38	ivo://jvo/server/arisa	192.168.0.4	Intel(R 2.40G
			grid01	true	true	1.61	2	2006-10-12 14:00:28	ivo://jvo/server/grid01	192.168.0.64	AMD Core F
			grid02	true	true	1.02	2	2006-10-12 14:00:04	ivo://jvo/server/grid02	192.168.0.66	AMD Core F
			grid03	true	true	0.56	2	2006-10-12 14:00:16	ivo://jvo/server/grid03	192.168.0.67	AMD Core F
			grid10	true	true	0.0	0	2006-10-12 14:00:26	ivo://jvo/server/grid10	192.168.0.68	AMD Core F
			jvo-work01	false	false	0.0	0	2006-07-15 13:26:40			Intel(R 3.00G
			jvo-work02	false	false	0.0	0	2006-07-15 13:26:47	ivo://jvo/server/jvo-work02	192.168.0.3	Intel(R 3.00G
			jvod	true	false	0.0	0	2006-10-12 13:59:35	ivo://jvo/server/jvod	192.168.0.5	Intel(R 2.80G
			jvoe	true	false	0.0	0	2006-10-12 13:59:48	ivo://jvo/server/jvoe	133.40.212.45	Intel(R 2.80G
			jvof	true	true	0.76	0	2006-10-12 14:00:08	ivo://jvo/server/jvof	192.168.0.1	Xeon(
			jvoh	true	false	0.0	0	2006-10-12 14:00:09	ivo://jvo/server/jvoh	192.168.0.7	Dual C Proces
			jvoi	true	true	1.31	1	2006-10-12 14:00:14	ivo://jvo/server/jvoi	192.168.0.8	Dual C Proces
			jvoj	true	true	2.15	3	2006-10-12 13:59:52	ivo://jvo/server/jvoj	192.168.0.9	Dual C Proces
			piglet	false	false	0.69	1	2006-09-11 17:24:39	ivo://jvo/server/piglet	133.40.208.47	AMD 4000+
			tiger	true	false	0.0	0	2006-10-12 14:00:22	ivo://jvo/server/tiger	192.168.0.65	AMD Core F

Experiment (58mosaic/12CPU Core)



Summary

- We developed GRID computing system for Subaru data analysis
  - One night observation
    One day (1CPU) → One hour
  - All the archived data (6 years)
    More than one year (1CPU) → One week
  - Accessible from Web browser.
- Operation System will be completed next year.
- Experimental use of NAREGI Grid middleware is underway