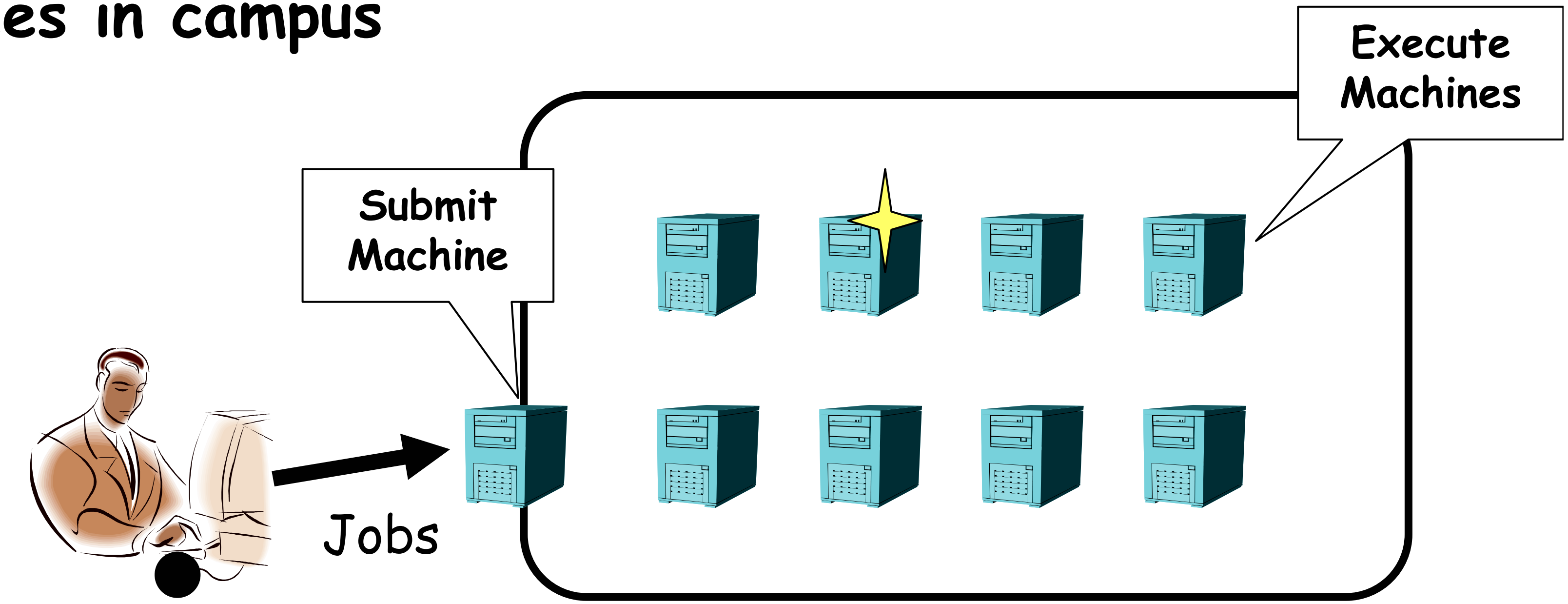
Design and Implementation of Condor-UNICORE Bridge

Hidemoto Nakada [1,4], Jaime Frey [2], Motohiro Yamada [3], Yasuyoshi Itou [3], Yasumasa Nakano [3], Satoshi Matsuoka [1,5]

1.National Institute of Advanced Industrial Science and Technology, Japan
2.University of Wisconsin
3.Fujitsu Limited
4.Tokyo Institute of Technology
5.National Institute of Informatics

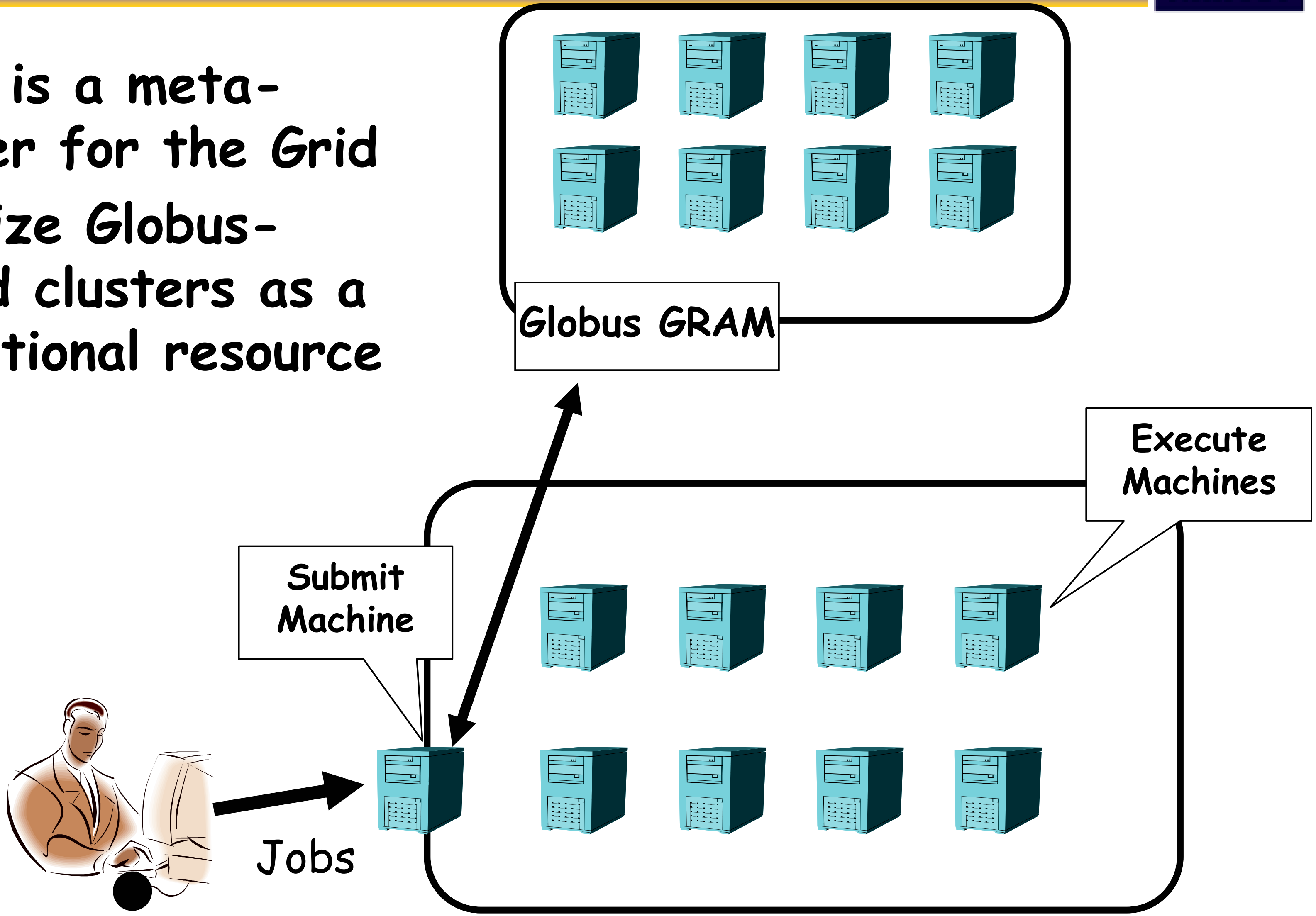
Condor Overview (1)

- Job scheduling system developed by Univ. Wisconsin, from '80
- The initial goal was to utilize computational resources in campus



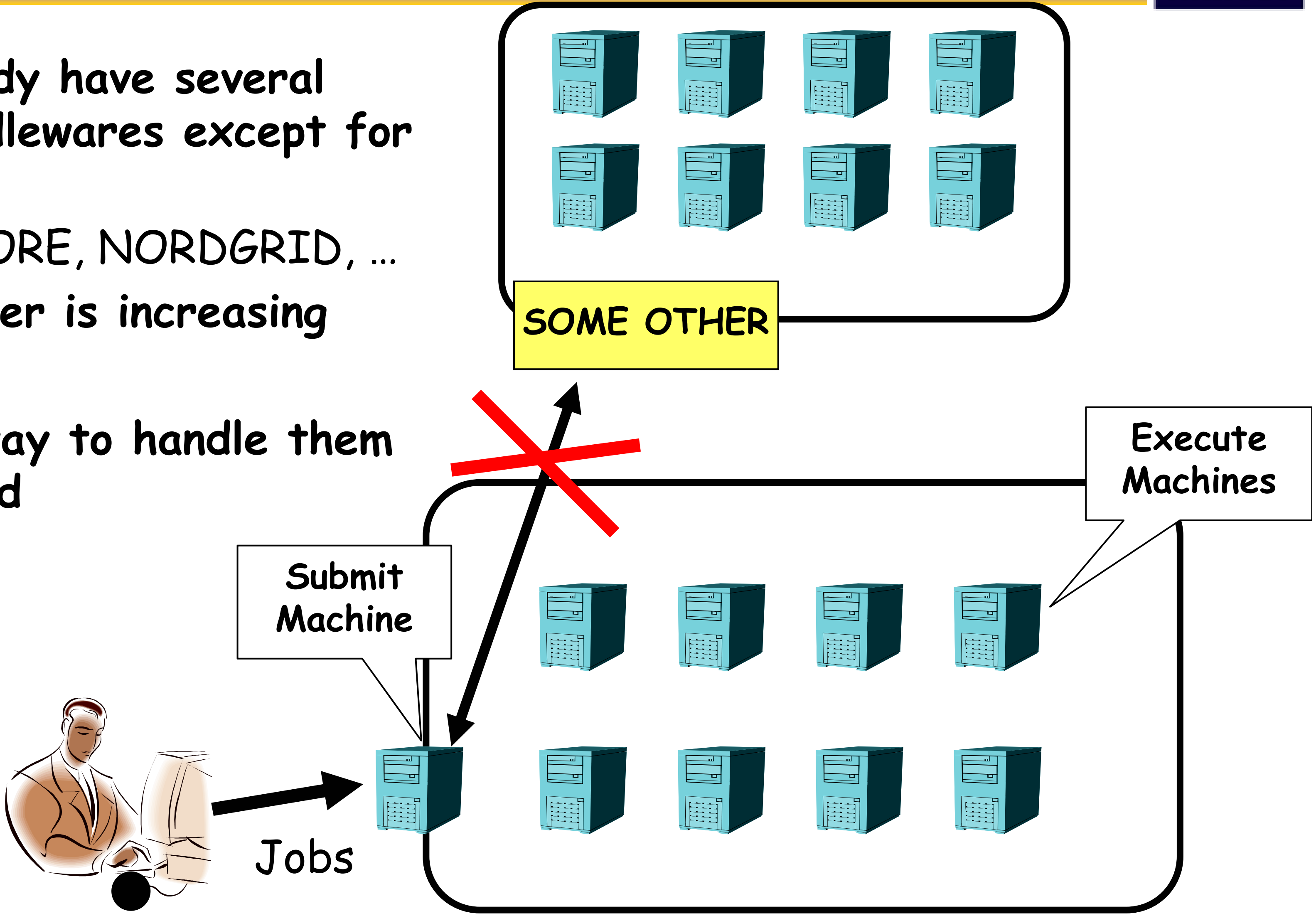
Condor Overview (2)

- Now, it is a meta-scheduler for the Grid
- Can utilize Globus-managed clusters as a computational resource pool



The problem

- We already have several Grid Middlewares except for Globus
  - UNICORE, NORDGRID, …
- The number is increasing
- Generic way to handle them is required



The Goal

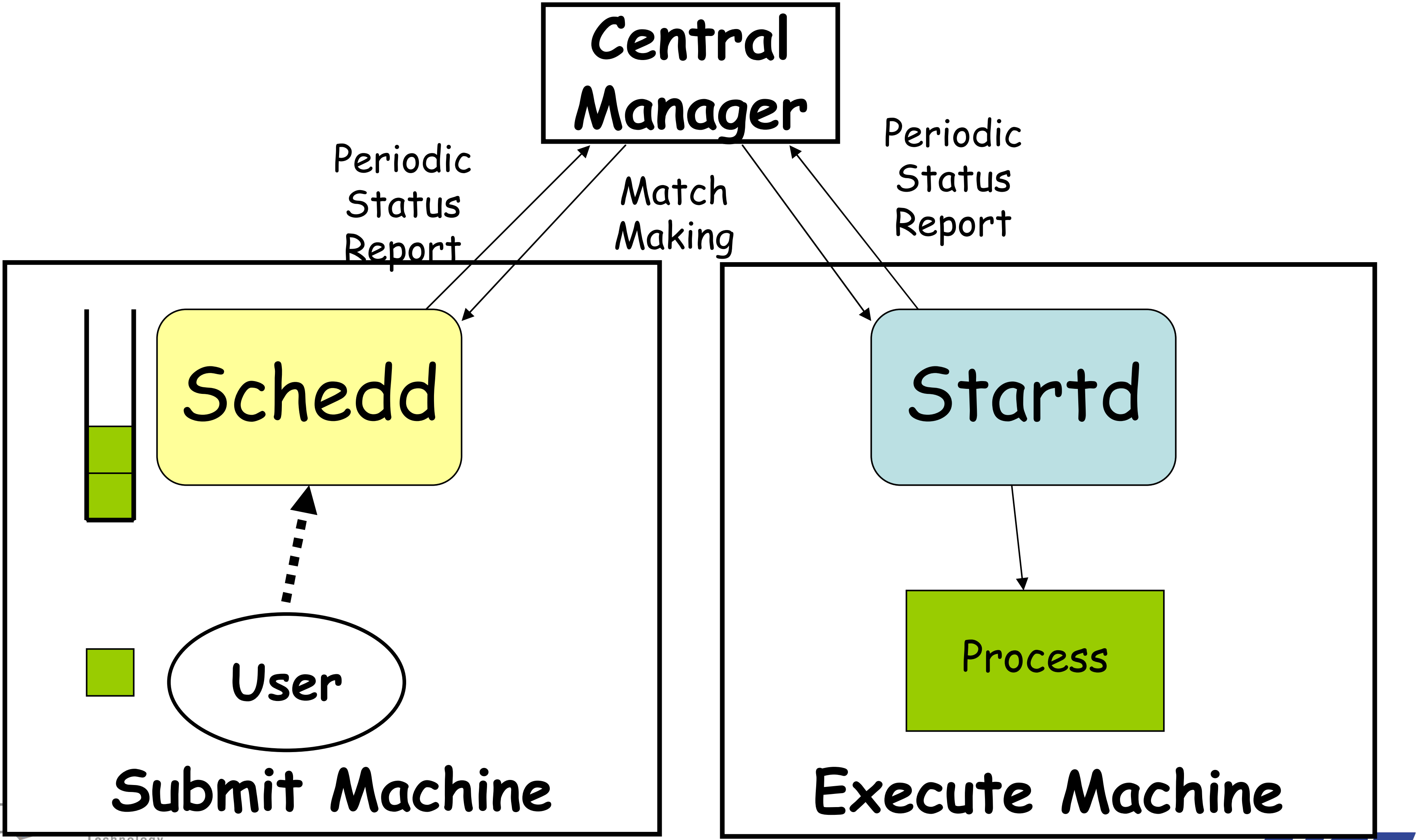
- **To give a generic framework that enables**
  - Easy implementation of the bridge to Grid Middlewares
  - Without modifying the Condor main modules

- **Validate the framework with implementing a bridge to UNICORE**

Outline of the talk

- **Condor architecture**
  - Existing bridging mechanism
- **Proposed bridging mechanism**
- **Implementation of a bridge to UNICORE**
- **Summary**

Condor Architecture



Central Manager

Periodic Status Report

Match Making

Periodic Status Report

Schedd

User

Submit Machine

Startd

Process

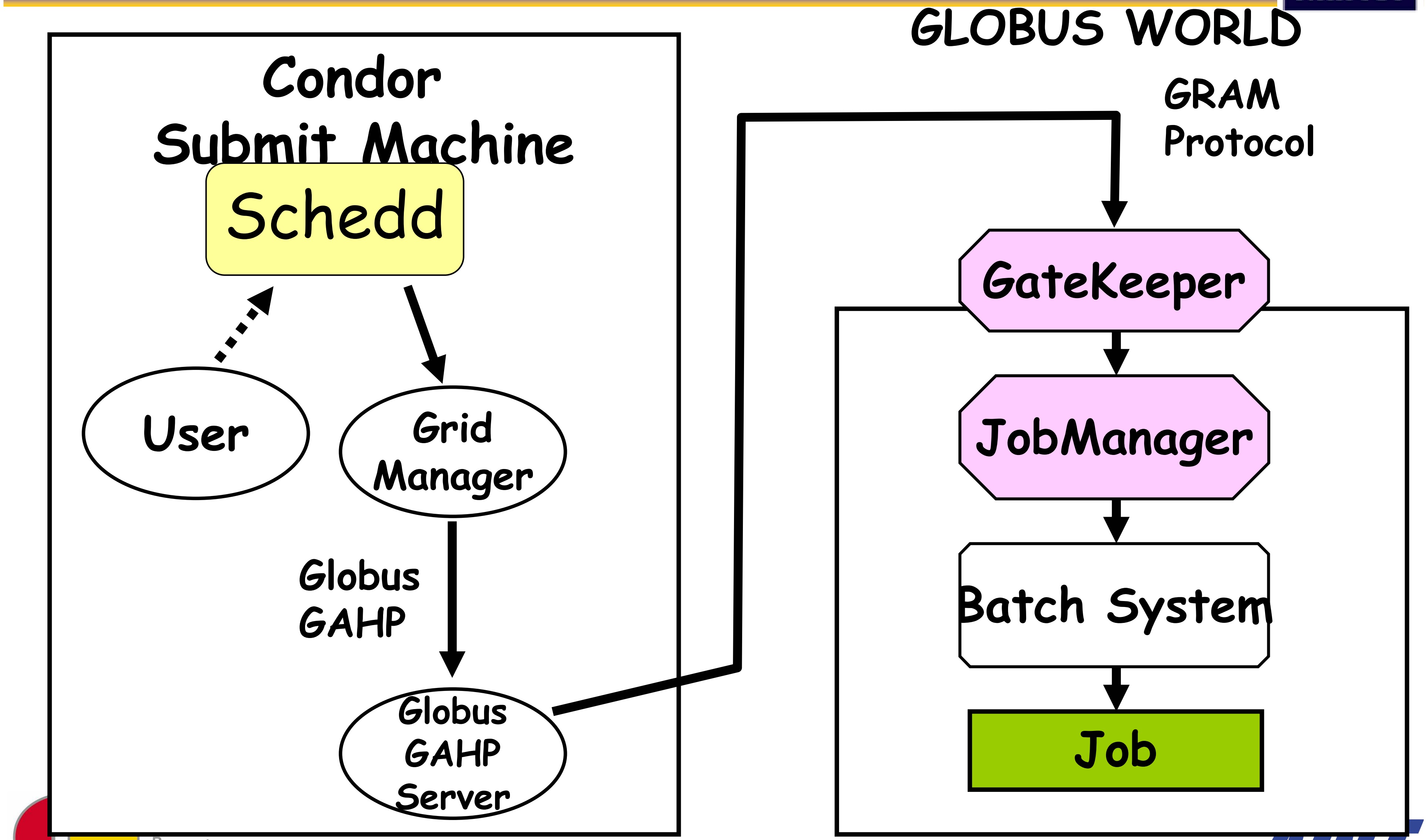
Execute Machine

Condor ClassAd: source of flexibility

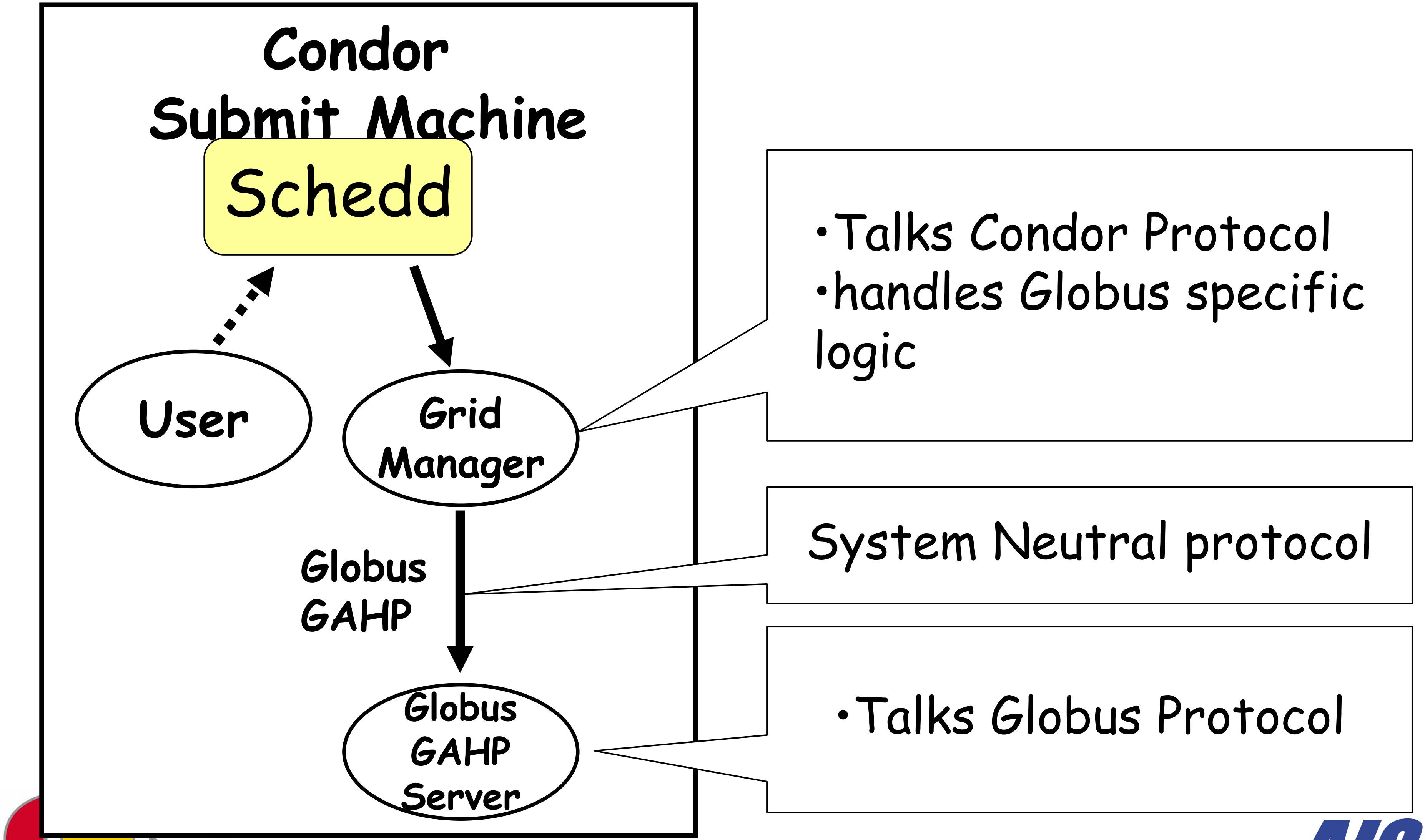
- **Data format used in whole condor system**
- **Attribute Key and Value pair**
  - Expressions are supported
- **All the information and requirements on jobs and machines are represented in classAd form**
- **Central manager 'matchmakes' them, assigning job to machine**

```
MyType = "Job"
TargetType = "Machine"
ClusterId = 292
User = "nakada@a02.aist.go.jp"
In = "/dev/null"
Out = "A.out"
Err = "/dev/null"
```

Existing bridging mechanism: Condor-G



Existing bridging mechanism: Condor-G



Condor
Submit Machine

Schedd

User

Grid
Manager

Globus
GAHP

Globus
GAHP
Server

- Talks Condor Protocol
- handles Globus specific logic

System Neutral protocol

- Talks Globus Protocol

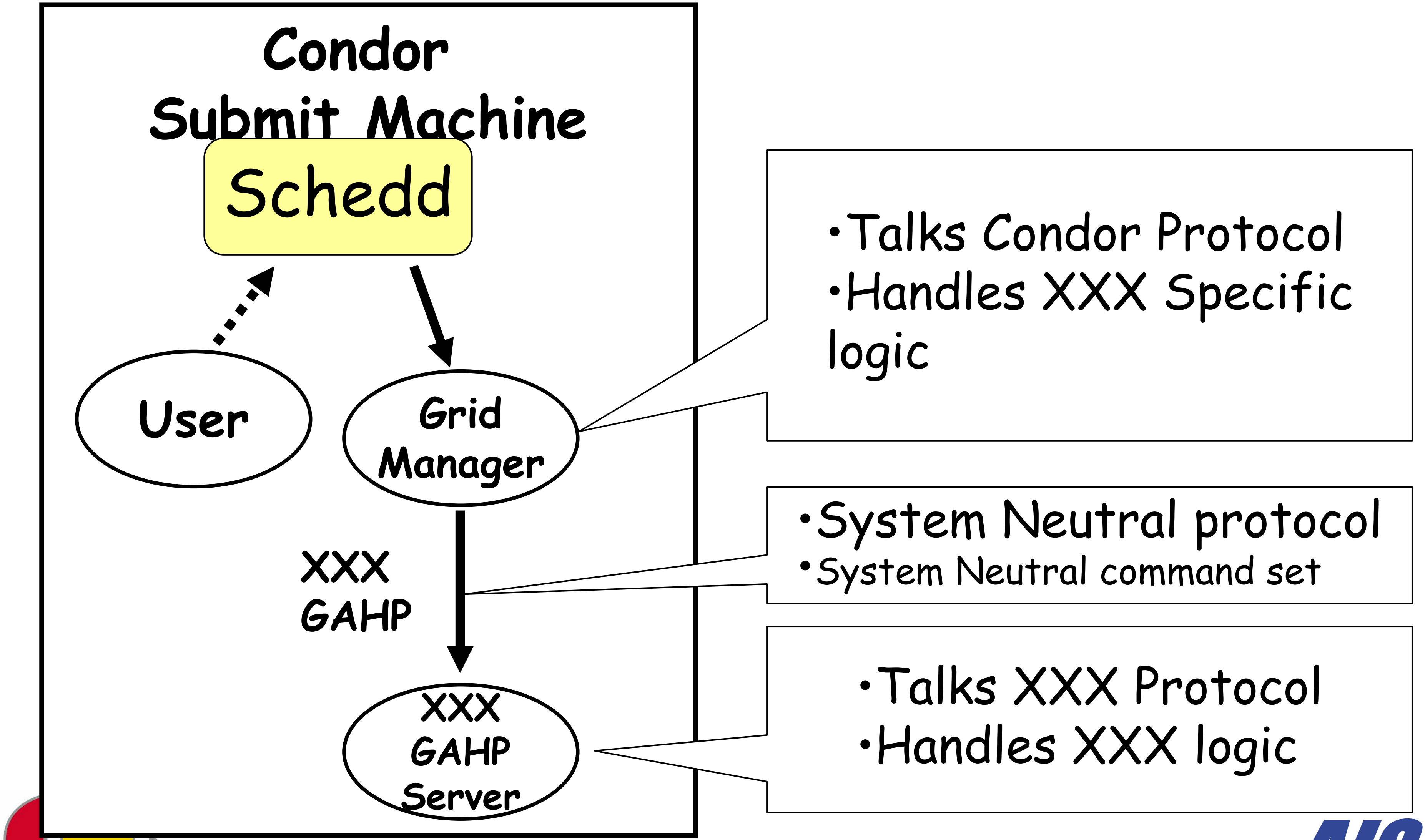
Condor-G concept

- **Clear separation between Condor and Globus specific codes**
  - It is really hard to Link them together.
  - Connect them with System neutral protocol
- **Great?, but in reality,**
  - The Globus specific logic is handled within the GridManager, not in the GAHP Server
  - GAHP protocol itself was system neutral, but the command set used on the protocol was very Globus Specific

GAHP command set for Globus

- INITIALIZE_FROM_FILE
- INITIALIZE_FROM_MYPROXY
- COMMANDS
- VERSION
- ASYNC_MODE_ON
- ASYNC_MODE_OFF
- QUIT
- RESULTS
- GRAM_CALLBACK_ALLOW
- GRAM_ERROR_STRING
- GRAM_JOB_REQUEST
- GRAM_JOB_CANCEL
- GRAM_JOB_STATUS
- GRAM_JOB_SIGNAL
- GRAM_PING
- GRAM_JOB_CALLBACK_REGISTER
- GASS_SERVER_INIT
- REFRESH_PROXY_FROM_FILE
- MYPROXY_REFRESH
- MYPROXY_RETRIEVE
- PROXY_INFO
- MYPROXY_DESTROY
- MYPROXY_DELEGATE

Proposed architecture



- Talks Condor Protocol
- Handles XXX Specific logic

- System Neutral protocol
- System Neutral command set

- Talks XXX Protocol
- Handles XXX logic

Proposed architecture

- **To support a new Grid Middleware system**
  - Only the GAHP server is required to be re-implemented
    - Java classes are provided to talk GAHP protocol
  - GridManager can be remain untouched
    - It is not easy to modify it for third party

What we did

- **Redesign the GAHP command set**
  - Simple and Generic as much as possible
  - Note: The GAHP protocol is not changed

- **Implement some logic in the Gridmanager to support Generic commands**
  - Can be reused for other systems

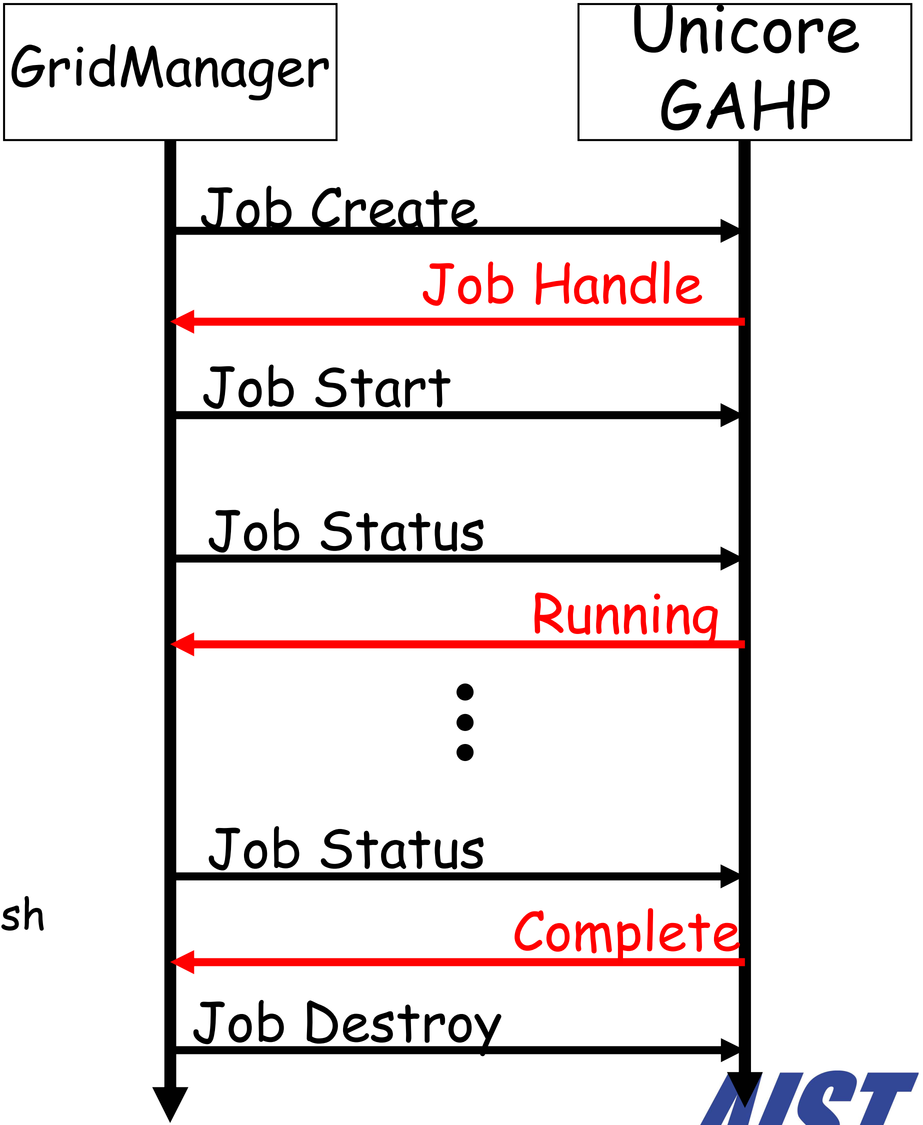
Design principle of the Generic GAHP command set

- **Simple and Generic**
  - High level commands
    - Just 5 commands - c.f. 23 commands for Globus

- **Use ClassAd as a command argument and a return value**
  - To ensure the genericity of the command set
  - System specific things are encapsulated in the ClassAd
  - You can extend the functionality by just extending the ClassAd attribute, without touching the Command set itself

Generic GAHP command set

- **Job Create**
  - Create a Job
  - Input: ClassAd
  - Output : Job Handle
- **Job Start**
  - Invoke the Job
  - Input : Job Handle
- **Job Status**
  - Query the Job status
  - Input : Job Handle
  - Output : Status ClassAd
- **Job Destroy**
  - Destroy information stored in the GAHP server
  - Input : Job handle
- **Job Recover**
  - Reconnect already running job. Used when GridManager recovers from crash
  - Input: Job handle, ClassAd



Condor-U

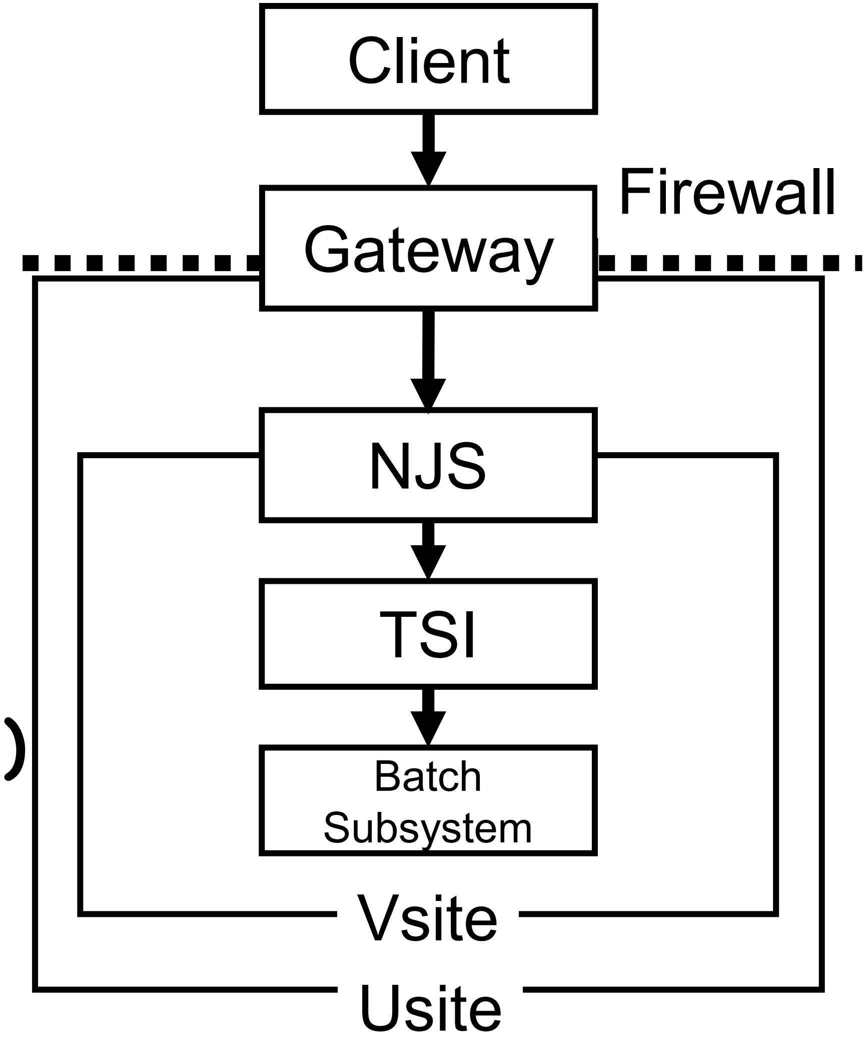
- **The bridge to invoke Condor job via UNICORE**
- **We implemented it to validate the architecture**
- **Implementation steps**
  - Design ClassAd for UNICORE
  - Implement the GAHP Server which understand UNICORE specific ClassAd and talks with UNICORE

The UNICORE

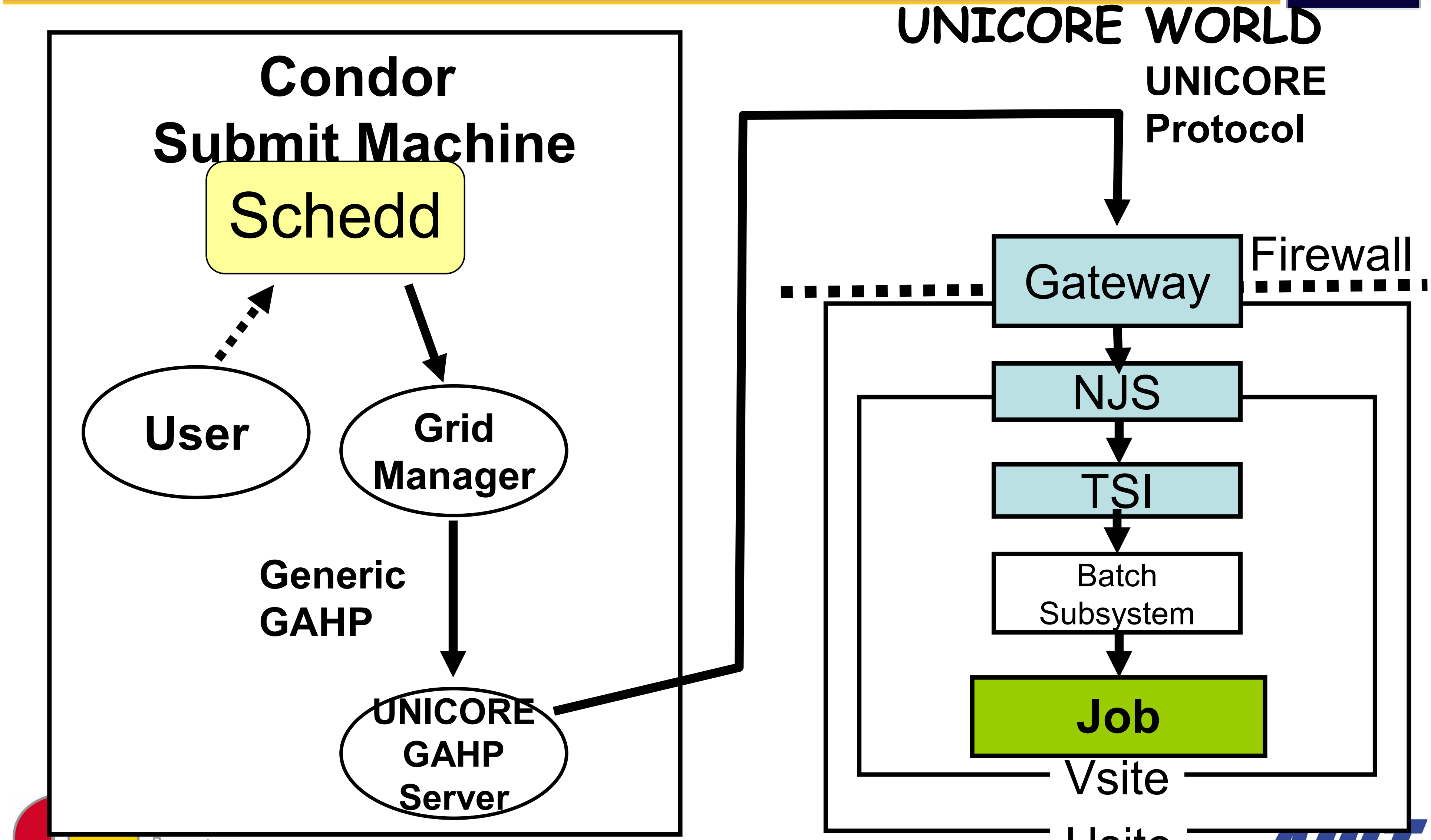
- **A Grid middleware developed mainly by Fujitsu Lab. Europe**
  - Owned by UNICORE Forum
  - Designed to utilize several supercomputers installed distributed supercomputer centers
  - SSH based security model
    - No PROXY CERT.
  - Firewall Aware (c.f. Globus)
    - Connection is one-way:
      - Can be used from private addressed network
      - Submit and Disconnect!
  - Totally in Java (except for small perl scripts)

The UNICORE architecture

- **Gateway**
  - Application level Router
  - Runs on a Firewall
  - Relay all communications
  - SSL based security
- **NJS (Network Job Supervisor)**
  - Workflow engine
  - Interpret AJO and execute
- **TSI (Target System Interface)**
  - Wrap batch sub system
  - Implemented in Perl



Condor-U Overview


UNICORE WORLD

UNICORE Protocol

Condor Submit Machine

Schedd

User

Grid Manager

Generic GAHP

UNICORE GAHP Server

Gateway

Firewall

NJS

TSI

Batch Subsystem

Job

Vsite

Usite

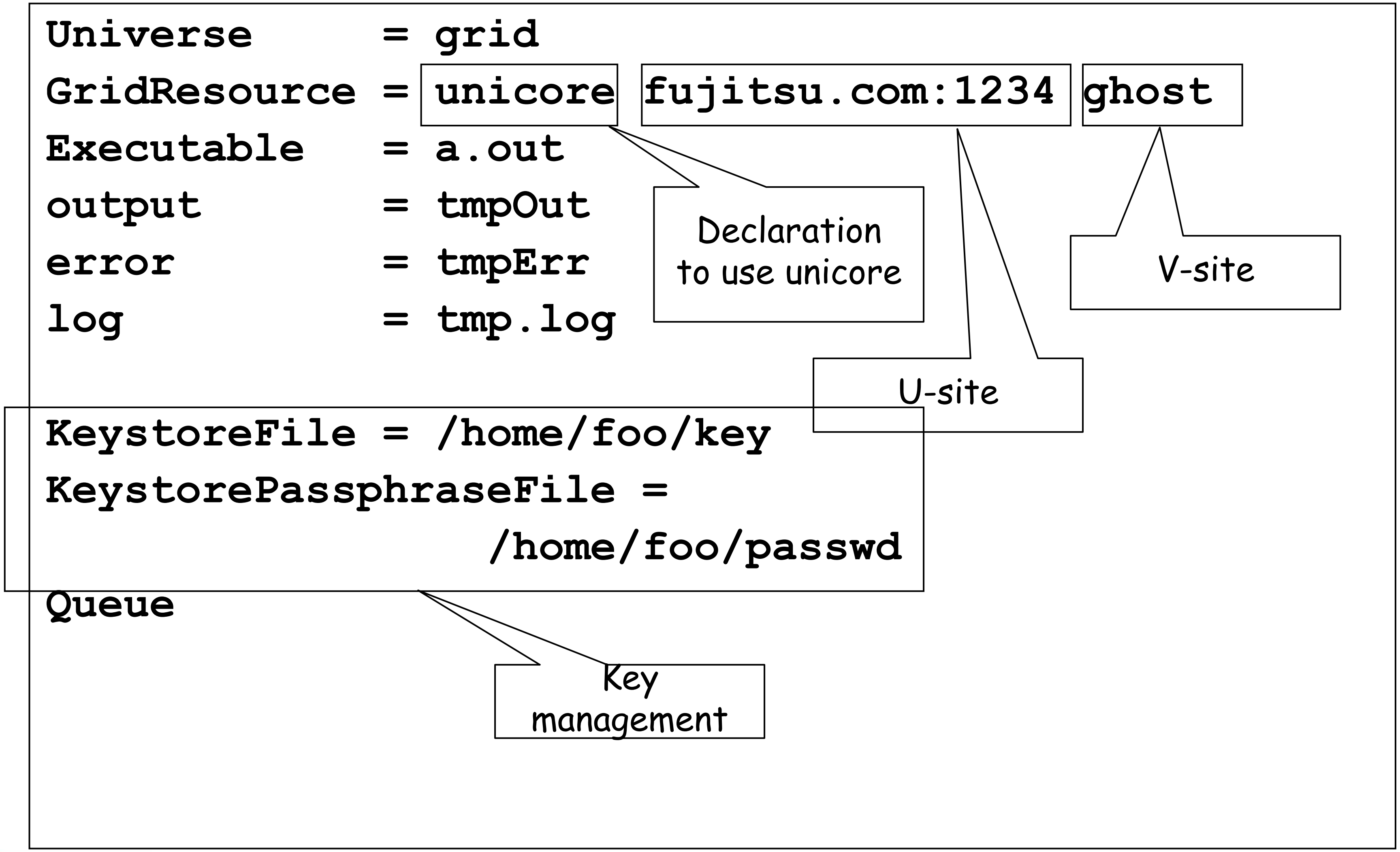
ClassAd Attributes (1) generic

Cmd	pathname of the command to execute	/home/foo/a.exe
Args	arguments for the execuatble	arg1, -a
Env	environment variables	LANG=en_US
	current directory	home/nakada/condor
In	standard input	input.dat
Out	standard output	ouput.dat
Err	error	Error.dat
TransferInput	Stage in files	a.exe, input.dat
TransferOutput	Stage out files	out.dat
JobStatus	Condor job status	jobClassAdAttributes
ErrMessage	Error messages	Script reported no errors
RemoteWallClockTime	Execution wall clock time	123.0
ByteSent	Bytes sent	1023004
ByteRecvd	Bytes received	1023004
ExitBySignal	The job process quitted by signal?	TRUE
ExitCode	Exit code of the process	1
ExitSignal	Exit signal of the process	9

ClassAd attributes (2) Unicore Dependent

GridResource	-gahp server name -FQDN and port number for the Unicore Usite gateway -Unicore Vsie Name	Unicore fujitsu.com:1234 NaReGI
KeystoreFile	Keystore file name	/home/foo/key
KeystorePassphraseFile	Pass phrase file name	/home/foo/passwd
UnicoreJobId	Job ID used as the handle	fujitsu.com:1234/NaReGI/1374036929
UnicoreJobStatus	UNICORE job status	
UnicoreLog	UNICORE log filename	/var/log/unicore.log

Submit file sample for Condor-U



```
Universe        = grid
GridResource = unicore fujitsu.com:1234 ghost
Executable    = a.out
output          = tmpOut
error           = tmpErr
log             = tmp.log

KeystoreFile = /home/foo/key
KeystorePassphraseFile =
                         /home/foo/passwd
Queue
```

Declaration to use unicore

U-site

V-site

Key management

Implication of the Condor-U

- **Condor users can use resources managed by the UNICORE**
  - From out side of the sites
  - Users can use Condor as a job-scheduler for UNICORE managed resources

- **Condor GlideIn might be used on it**
  - Note: Communication between nodes have to be assured – it is not common for UNICORE setup

Current Status

- **Merged into the main trunk of the Condor CVS**
  - Will be shipped as a part of the next development release

Summary and Future Work

- **We defined a generic and simple command set for Condor external interface**
  - Unicore GAHP server implementation showed feasibility of the command set

- **Future work**
  - Implements GAHP servers for other Grid systems with the command set
    - UNIGRID ?
    - NAREGI Middleware ?

Thank you

Acknowledgement:

A part of this research was supported by a grant from the Ministry of Education, Sports, Culture, Science, and Technology (MEXT) of Japan through the NAREGI (National Research Grid Initiative) Project.