OGSA-DAI Supported Asynchronous Data Access Service (VO-DAS)

Chao Liu, Haijun Tian, Dan Gao, Yang Yang, Yong Lu
China-VO
National Astronomical Observatories, CAS, China

Goals

- Handle high volume data
- Interlink distributed and heterogeneous archives
  - catalogs, images, spectrums
- **Make a software that works for astronomers**

OGSA-DAI Overview

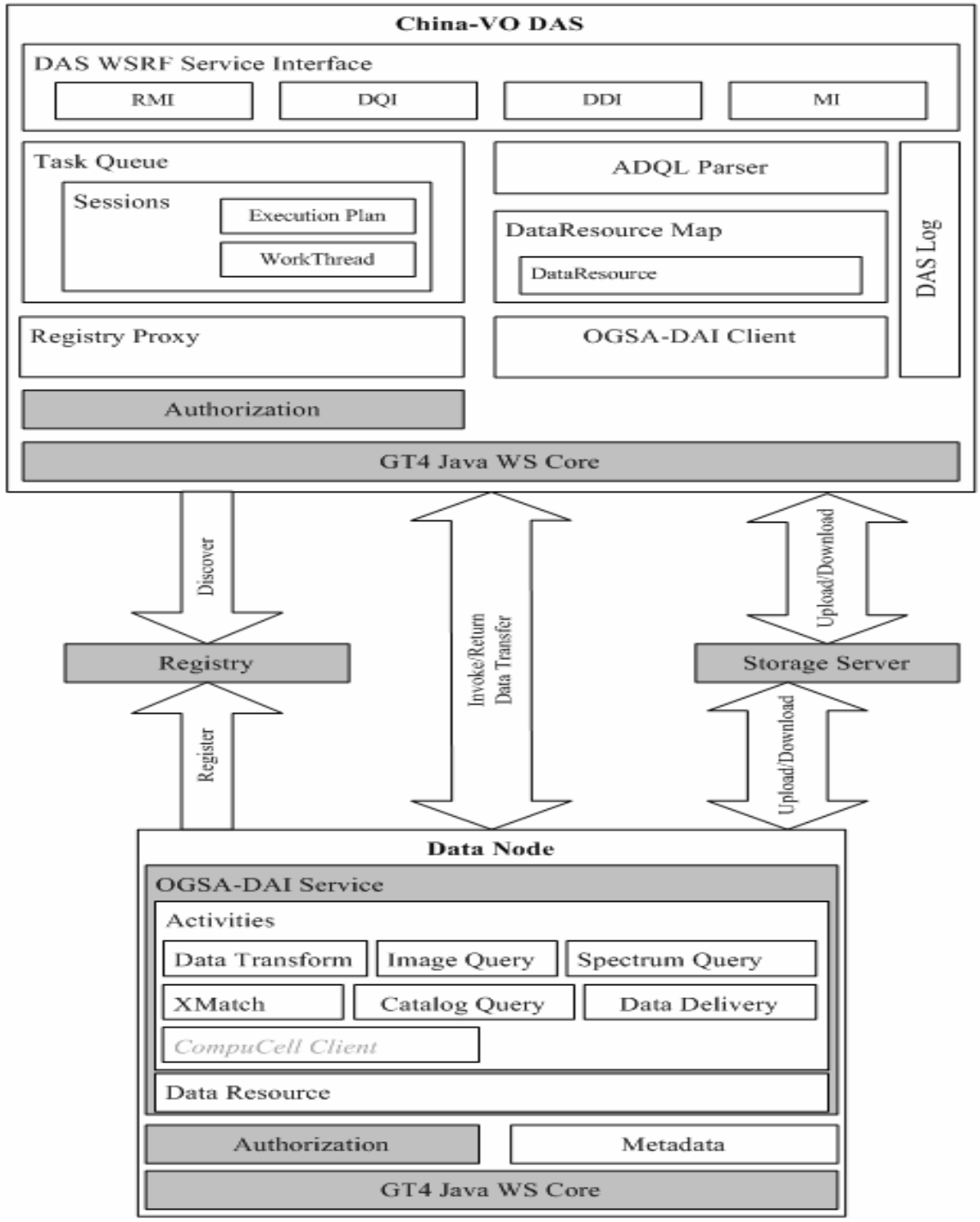
- OGSA-DAI is a middleware product which supports the exposure of data resources, such as relational or XML databases, on to grids.
- Work with Globus Toolkit 4

Our Evaluation

- Pros
  - Good wrapper for variant databases
  - Asynchronous query supported
  - Available for large dataset
  - WSRF interface
  - Flexible data transportation solution
  - Good extension capability
  - Multiple data format and extensible
- Cons
  - Still not stable
  - Cannot provide table metadata

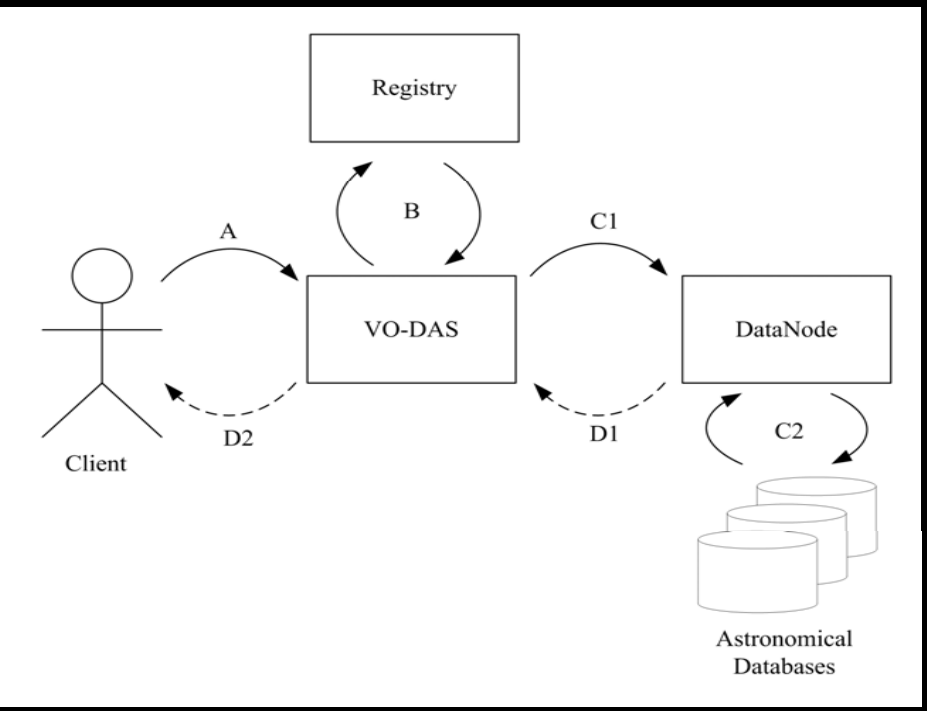
VO-DAS Architecture

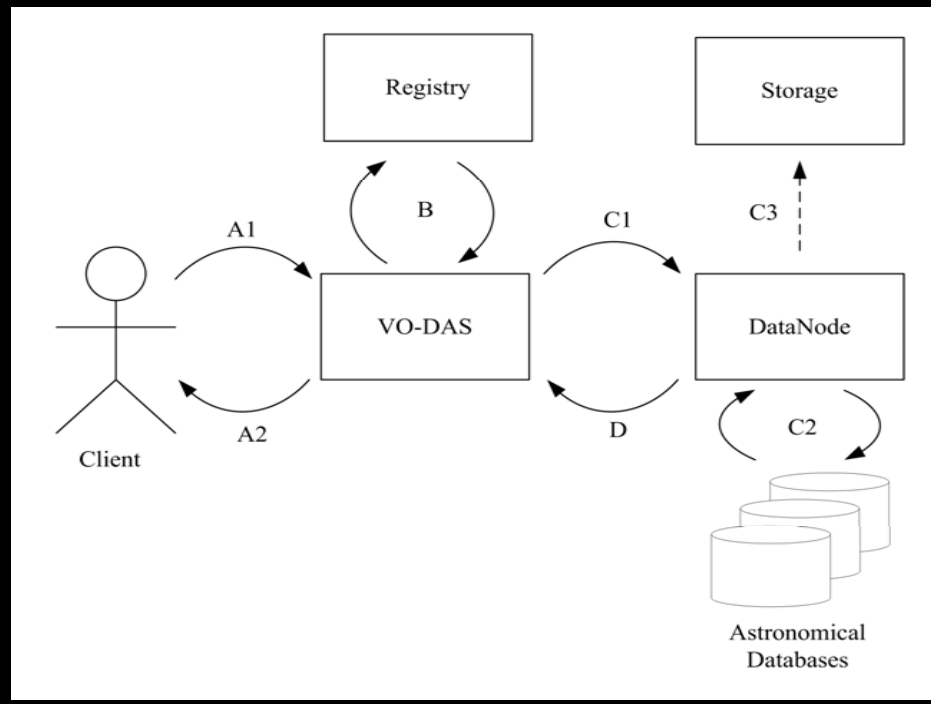
- VO-DAS Server
  - WSRF based web service
  - Integrates OGSA-DAI services
  - Job controlling and monitoring
- DataNode
  - OGSA-DAI wrapped data resources
- Interfaces
  - Resource Metadata Interface
  - Data Query Interface
  - Data Delivery Interface
  - Management Interface

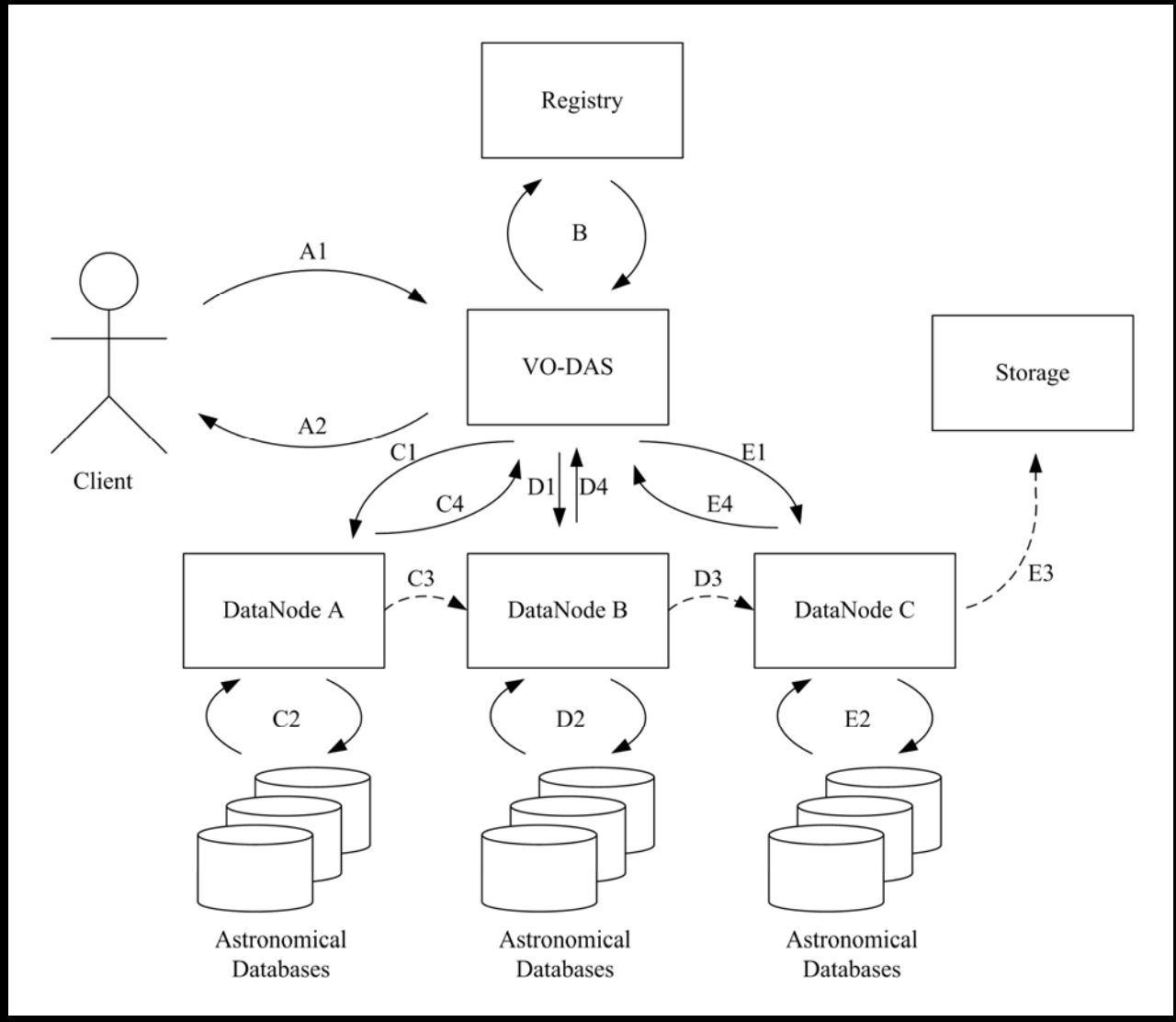

China-VO DAS
DAS WSRF Service Interface
RMI
DQI
DDI
MI
Task Queue
Sessions
Execution Plan
WorkThread
ADQL Parser
DataResource Map
DataResource
DAS Log
Registry Proxy
OGSA-DAI Client
Authorization
GT4 Java WS Core
Discover
Invoke/Return
Data Transfer
Upload/Download
Registry
Storage Server
Register
Upload/Download
Data Node
OGSA-DAI Service
Activities
Data Transform
Image Query
Spectrum Query
XMatch
Catalog Query
Data Delivery
CompuCell Client
Data Resource
Authorization
Metadata
GT4 Java WS Core

VO-DAS Action 1

- Query Workflow
  - Synchronous Query
  - Asynchronous Query
  - Asynchronous Query on multiple DataNodes


Registry
B
A
C1
VO-DAS
DataNode
Client
D2
D1
C2
Astronomical
Databases


Registry
Storage
B
A1
C1
C3
VO-DAS
DataNode
Client
A2
D
C2
Astronomical
Databases


Registry
B
A1
VO-DAS
Storage
A2
Client
C1
E1
C4
D1
D4
E4
DataNode A
C3
DataNode B
D3
DataNode C
E3
C2
D2
E2
Astronomical Databases
Astronomical Databases
Astronomical Databases

VO-DAS Action 2

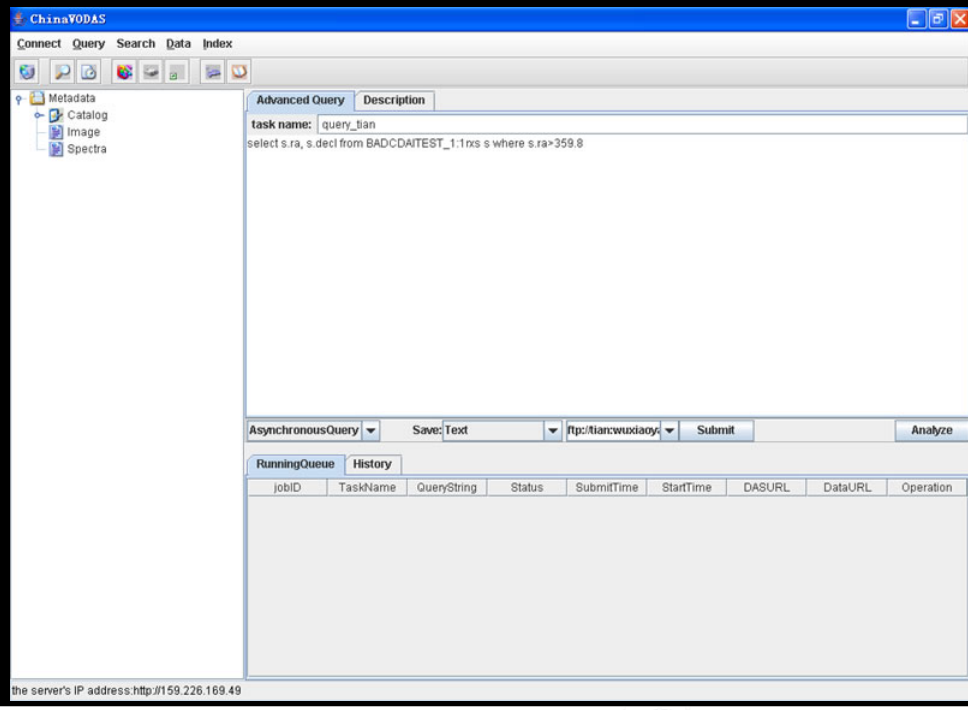
- Asynchronous Query Management
  - An End-Point Reference as session
  - Output destination setting while query
  - Result data format setting while query
- Example
  - *session=NewSession()*
  - *session.AsynQuery(<adql string>, <destination url>, <data format>)*
  - *session.GetStatus()*
  - *session.Destroy ()*

Performance

- High Volume Data Query Test
  - Long time to query and retrieve big data file
- Frequently Data Query Test
  - Multiple user simultaneously query
  - A user frequently access data from server
- Conclusion
  - A million rows can be handled successfully
    - Depend on memory size
  - It is not stable when simultaneously query (high loaded case)
    - Reason: OGSA-DAI's defect? Or database problem?
  - It is complicated to handle failure

Status

- An Alpha version of VO-DAS is ready
  - Simple ADQL access by a GUI java client
  - ConeSearch and XMatch between multiple OGSA-DAI services
  - VOTable, ASCII, CSV data format
  - Simple job controlling



Future Work

- Short Term
  - A *commercial* system
    - Used by astronomers in a small community
- Long Term
  - VO-DAS + VOSpace?
  - VO-DAS + Other VO applications?
  - VO-DAS + Workflow control?
- Question?