Realizing LIGO Virtual Data

Caltech: Kent Blackburn, Phil Ehrens, Albert Lazzarini, Roy Williams

ISI: Ewa Deelman, Carl Kesselman, Gaurang Mehta, Leila Meshkat, Laura Pearlman

UWM: Bruce Allen, Scott Koranda

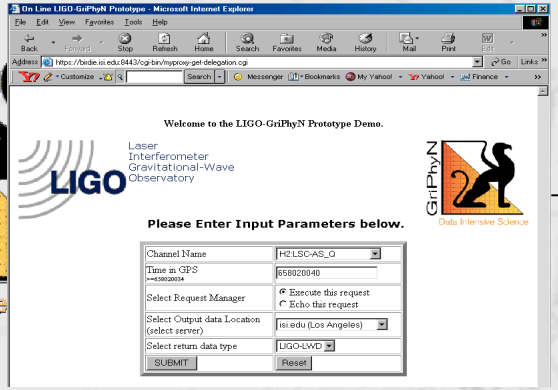
Outline

- GriPhyn/LIGO Prototype Overview
  - User interface
  - Request interpreter
  - Security model
  - Request planner
  - Request execution
- Plans for Year 2
  - Pulsar search
  - Virtual data
  - Request planning and execution
- Issues and Challenges

GriPhyN/LIGO prototype functionality

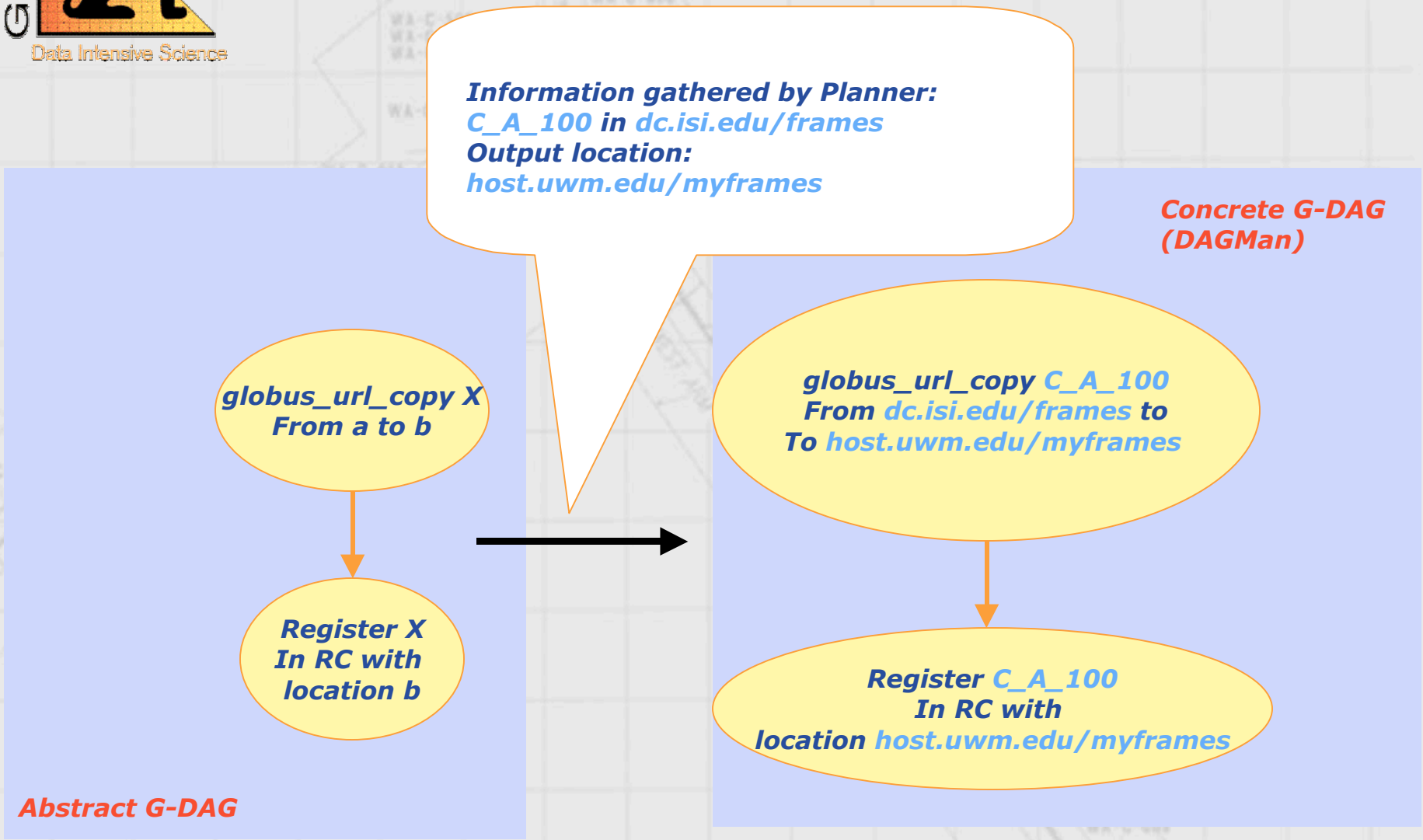
- Interpret an XML-specified request
- Acquire user's proxy credentials
- Consult replica catalog to find available data
- Construct a plan to produce data not available
- Execute the plan
- Return requested data in Frame or XML format

Compute resources (LDAS) at Caltech and UWM, storage resources at ISI, UWM and Caltech

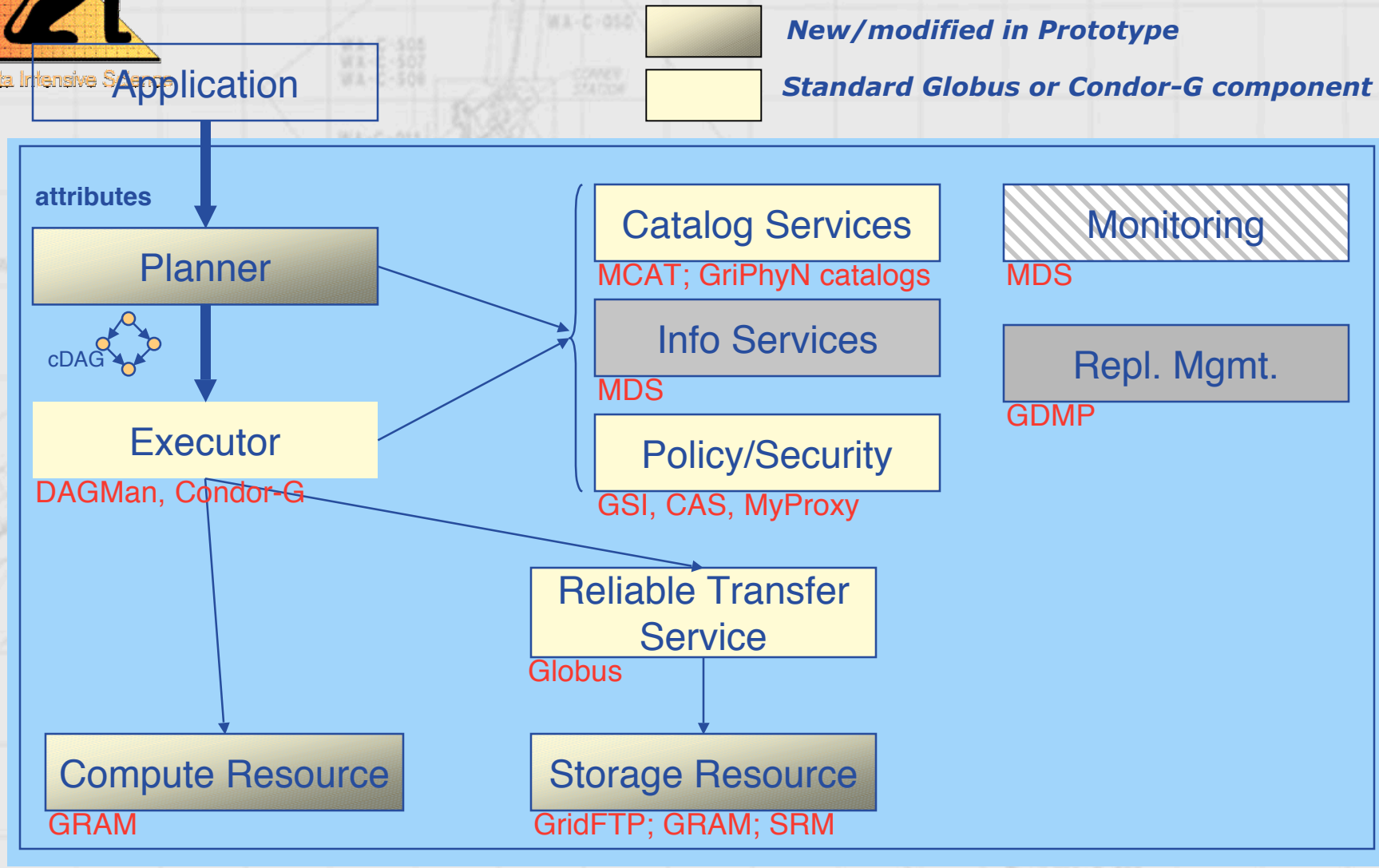

On Line LIGO-GriPhyN Prototype - Microsoft Internet Explorer
Welcome to the LIGO-GriPhyN Prototype Demo.
LIGO
Laser Interferometer Gravitational-Wave Observatory
GriPhyN
Data Intensive Science
Please Enter Input Parameters below.
Channel Name
H2:LSC-AS_Q
Time in GPS
658020040
Select Request Manager
Execute this request
Echo this request
Select Output data Location (select server)
isi.edu (Los Angeles)
Select return data type
LIGO-LWD
SUBMIT
Reset

GriPhyN
Data Intensive
xml
HTTP frontend
MyProxy server
Cgi interface
Transformation Catalog
Planner
Monitoring MDS
Replica Catalog
G-DAG (DAGMan)
Executor CondorG/ DAGMan
Logs
GridCVS
GridFTP
GRAM/LDAS
GridFTP
GRAM
LDAS
Storage Resource
Compute Resource

Template instantiation



Preliminary GriPhyN Data Grid Architecture



Accomplishments

- Simple demonstration of Virtual Data Concepts
  - Transparency with respect to location
  - Transparency with respect to materialization
- Provided a Globus interface to LDAS
  - basis for a secure access to LIGO resources
- Designed the Transformation Catalog
  - can be used in many systems
- Basic infrastructure for the development of Virtual Data concepts
  - foundation for Year 2

The Physics of LIGO's Pulsar Search

- EM pulsars not likely to generate strong GW signature
  - need to develop techniques for locating presently unidentified sources of continuous gravitational waves (" GW pulsars")
  - Support for the so-called "blind" or "all-sky" search for new sources.
- GW signals are frequency modulated by Doppler shift produced by Earth's rotation and barycentric motion around the Sun.
  - Need to account for this as a function of different lines of sight to putative sources, with different parameters.
  - Exploiting full SNR potential of LIGO data becomes a peta-flops class problem.
- Grid provides a fabric on which to process such searches as background tasks
  - target of next phase of development for GriPhyN-LIGO applications.

LIGO's Pulsar Search

(Laser Interferometer Gravitational-wave Observatory)

GriPhyN
Data Intensive Science

archive

raw channels

Single Frame

Extract channel

Short time frames

transpose

Long time frames

SFT

Extract frequency range

Construct image

Time-frequency Image

Hz

Time

Find Candidate

Store

event DB

LIGO-G020004-00-E
GriPhyN/LIGO prototype 1/2002

Ewa Deelman, ISI

Year 2 The Year of the Pulsar Search Mock Data Challenge

- Broaden the GRAM/LDAS interface
  - greater variability and functionality: SFTs, concatenation, decimation and resampling.
- Design a Data Discovery mechanism for discovery of data replicas on a Grid.
  - ability to interact with the LDAS Diskcache resources
- Implementation of the Data Discovery mechanism to support the pulsar search

Virtual Data Concepts

- Implement the Transformation Catalog.
- Explore the design of the Derived Data Catalog, which specifies how Virtual Data products are materialized.
- Unify the catalog schemes used by CMS and LIGO – base it on a common VDT 2.0 release.
- Apply replication concepts by developing a real-time international mirror, and a fault-tolerance replica at UW-Mil.
- Use of Catalogs to materialize Virtual data required in the pulsar search (including Transformation Catalog).

Planning and Fault Tolerance

- Specify the planning requirements
- Evaluate the available solutions
- Prototype a more sophisticated planner

- Specify LIGO's fault tolerance requirements, extrapolate to GriPhyN in general
- Assess existing fault and failure issues within LIGO
- Assess the applicability of existing techniques

Year 2 Challenges

- Explore bulk data operations
  - Finding new available data
  - Registering data into catalogs
- Deepen the understanding of Virtual Data naming
  - How do you ask for what you want?
- Planning and Fault Tolerance
  - Need to specify model
  - Explore existing planning solutions
  - Examine fault tolerance issues at the system level
- Scalable pulsar search to scientifically interesting levels of sensitivity at SC'2002