TclGlobus

ITR2003 Meeting

Argonne National Laboratory

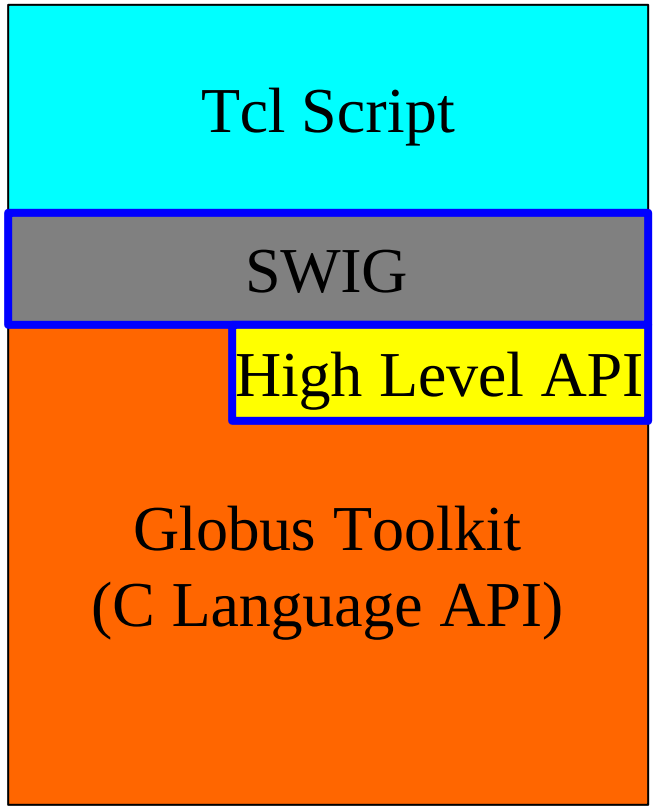
May 10th, 2004

Kent Blackburn

Definition of Task

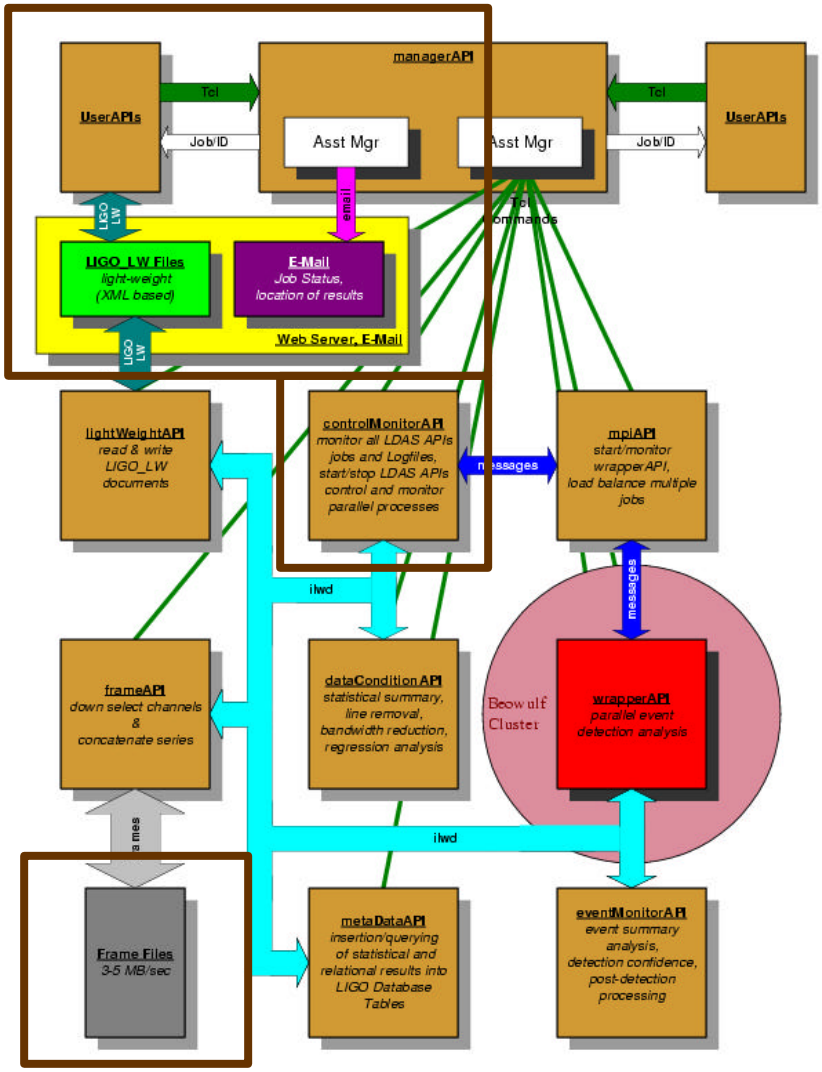
- **Develop a Globus API applicable to the Tck/Tk scripting language.**
  - Analogous to PyGlobus
- **Use this new API to extend LIGO's Data Analysis System (LDAS) to the Grid:**
  - Authentication / Authorization using GSI certificates.
  - Publishing/moving LIGO data products around on the Grid.
  - Grid level monitoring of LDAS systems sharing the Grid.
- **Provide *packaged* deliverable to the larger community of Tcl/Tk developers for Grid Applications.**

Wrapping Globus in Tcl



- **Use SWIG – Simple Wrapper Interface Generator to expose the C language interface in Globus to Tcl.**
  - **Plan to wrap most of the Globus Toolkits primitives for use in Tcl.**
  - **Add high level interfaces for commonly grouped Globus calls needed by LDAS (*and possibly other Tcl Applications*) for greater efficiency.**
- **Architecturally analogous to the PyGlobus design/implementation.**

Grid-enabling LDAS



- **LDAS datapipeline flow control handled with Tcl.**
- **Users currently make a request to LDAS by connecting with a Tcl Socket on the managerAPI:**
  - **username, encrypted password**
  - ***migrate to GSI certificates***
- **Data Products "pulled" from LDAS using web services:**
  - **E-mail notification to user when data available**
  - ***migrate to data publishing on grid***
  - ***use of gridFTP to move data***
  - ***publish system statistics on grid, allowing monitor capabilities on grid.***

Principle Participants

- **Caltech taking the lead on TclGlobus under this ITR:**
  - Kent Blackburn: Cognizant Scientist, *~0.1 FTE*.
  - Ed Maros: Software Developer, *~0.2 FTE*.
  - Hari Pulapaka: Software Developer, *~0.1 FTE*.
  - Looking for an additional 1.5 FTE.
- **On February 24th, 2004 Ed Maros and Kent Blackburn visited the PyGlobus development Team at LBNL:**
  - Keith Jackson, David Konerding principle developers of PyGlobus.
  - PyGlobus is being redesigned internally – good timing for us!
  - Agreed to collaborate on a common architecture – "SWIGlobus".
  - SWIGlobus would establish foundation for both Tcl & Python scripting languages as well as future Globus script languages (Ruby, Perl, others).
  - Leverage off of existing infrastructure at LBNL (code repository, etc.).
  - One area for concern is joint licensing of the source code.
    - PyGlobus will strickly adhere to BSD software license.

Technical Details

- **Chose Globus 2.4 for its stable C interface.**
  - If Globus 3 provides a stable C interface we will consider using it.
  - Performance of Java API has raised concerns.
- **Developing with most current Tcl/Tk version 8.4.6.**
  - LDAS currently based on Tcl/Tk 8.3.x, but plan to migrate 8.4.x in time for TclGlobus integration.
  - Otherwise, we may become “motivated” to support TclGlobus under Tcl/Tk 8.3.x as well.
- **Using most recent version of SWIG (version 1.3.21).**
  - Same as current PyGlobus development.
- **Adopted the automake/autoconfig for target platform configuration management.**
  - PyGlobus project prefers not to use these tools, but we are working with them to allow their preferences (Python’s distutils) to coexist with ours.

Technical Infrastructure

Caltech

- **Set up a dual Intel Linux (RH9) development system.**
  - Allows testing of threaded code.
  - Has local *CVS* repository for TclGlobus - *obsolete*.
- **Have necessary accounts at LBL's code repository.**
- **Plan to set up website for TclGlobus**
- **Plan to use *doxygen* to for documentation generation.**
- **Interacting with LBL emails and telephone on weekly bases.**

LBNL

- **Using existing Linux (RH9) server for PyGlobus and SWIGlobus development.**
- **Set up code repository (using *subversion*).**
  - Directory structure supports requirements of both projects.
- **Setup problem tracking system (using *bugzilla*).**
- **Plan to setup WIKI for SWIGlobus.**
- **Using *epydoc* for PyGlobus documentation generation.**

Technical Challenges

- **Understanding Globus Toolkit:**
  - Software team (PyGlobus as well as TclGlobus) unhappy with level of documentation provided.
  - This is particularly an issue with developing SWIG wrappers where function parameters (input vs. output) poorly documented.
- **Python and Tcl differ in management of threads.**
  - We have adopted Tcl Thread extension library (version 2.5.2) to overcome significant issues with native Tcl.
  - Allows Tcl implementation to more closely follow Python's.
- **Long term maintenance issues:**
  - Tcl/Tk changed significantly in 8.4.x versus previous 8.3.x.
  - This has prevented migration of current LIGO software.
  - Globus 2.x support slated to end-of-life in late 2005.
  - Support for C language API in Globus 3.x uncertain at this time and will certainly come late in the TclGlobus projects planning if at all.

Benefactors

(***within LIGO Scientific Collaboration***)

- **The LIGO Data Analysis System will be able to integrate well with the Grid opening up greater resources for LIGO's data analysis.**
- **Current need for LIGO Scientific Collaboration (LSC) members to use several methods of authentication and authorization will be simplified with LDAS' migration to GSI standard X.509 digital certificates.**
- **Many of the client side tools used in conjunction with LDAS are based on Tcl/Tk language and will be able to utilize the TclGlobus package for connectivity with LDAS and the larger Grid (*e.g., job submission, data movement, monitors*).**
  - The SWIGlobus collaboration will also provide an API for Python users within the LSC (*and possible other scripting languages in the future*).

Benefactors

(*beyond the LIGO Scientific Collaboration*)

- **Combining PyGlobus with TclGlobus through the SWIGlobus collaboration allows each project to reach out to a larger community.**
- **SWIGlobus is providing a template for other scripting languages (Perl, Ruby, etc) to wrap the Globus Toolkit.**
- **Interest in Tcl interface to Globus has been expressed at "All-Hands-Meetings" of GriPhyN/iVDgL Projects.**
- **Astronomy community commonly uses Tcl, e.g., SDSS.**

Work Schedule

(year one: August 2003-August 2004)

- **August 2003 – present:**
  - **Hire 2 FTEs at Caltech for TclGlobus Development**
    - ☑ **Software developer to address SWIG interface development (*transitioning existing staff*)**
    - **Postdoc/staff to address high level interfaces needed by LDAS and overs ee project**
- **December 2003- January 2004:**
  - **Setup TclGlobus Infrastructure**
    - ☑ **Source code version control system**
    - ☑ **Problem tracking system**
    - **Web server at Caltech for TclGlobus Project**
- **January 2004 – June 2004:**
  - **Setup SWIGlobus *super-project* in collaboration with PyGlobus Group at LBNL**
    - ☑ **Source code version control system (*using subversion at LBNL*)**
    - ☑ **Problem tracking system (*using bugzilla at LBNL*)**
    - **WIKI information sharing system (*will be setup at LBNL, so far email has sufficed*)**
- **January 2004 – August 2004:**
  - **Understand Globus Toolkit function calls and the PyGlobus project's API model**
    - ☑ **Establish build environment and code repository directory structure for SWIGlobus**
    - **Wrap up a few test cases (starting with sockets and moving on to gridFTP client/servers) for both Tcl and Python and explore documentation tools (*just starting*)**

Work Schedule

(year two: August 2004-August 2005)

- **August 2004 – December 2004:**
  - **Implement first set of primitive API into TclGlobus**
    - **Focus on GSI socket communications and authentication**
    - **Implement standalone Tcl/Tk scripts that exercise this subset**
    - **Implement standardized documentation for these APIs with examples**
- **December 2004 – March 2005:**
  - **Based on experiences with standalone Tcl scripts, test primitives in LDAS**
    - **Add GSI certificate authentication to managerAPI to development LDAS**
    - **Add gridFTP data product client/server movers to development LDAS**
- **March 2005 – August 2005:**
  - **Develop high level API for use in LDAS based on experiences from early integration**
    - **Implement standalone Tcl/Tk test scripts based on high level API**
    - **Document new high level functionality of TclGlobus**
    - **Integrate high level API into development LDAS for GSI authentication and data product movement**
    - **Prepare first alpha release of TclGlobus, distribute under SWIGlobus**
    - **Release a version of LDAS that is based on this alpha release of TclGlobus**

Work Schedule

(year three: August 2005-August 2006)

- **August 2005 – December 2005:**
  - **Begin prototyping of Monitor and Discovery Services (MDS)**
    - **Develop high level APIs that publish simulated data representative of LDAS in standard grid accessible formats**
    - **Test these high level MDS APIs in test Tcl scripts**
    - **Document high level MDS APIs**
    - **Integrate high level MDS APIs in LDAS development.**
- **December 2005 – August 2006:**
  - **Finalize design of TclGlobus based on prior experiences under LDAS**
    - **Prepare beta release of TclGlobus, distribute under SWIGlobus**
    - **Release a version of LDAS based on beta version of TclGlobus which includes MDS**
    - **Create website for TclGlobus providing general information and access to the beta version**

Work Schedule

(year four: August 2006-August 2007)

- **August 2006- August 2007:**
  - **Put polishing touches on TclGlobus deliverables**
    - **Close out an many open issues as possible**
    - **Complete the documentation**
    - **Prepare version 1.0 official release of TclGlobus**
  - **Integrate version 1.0 release of TclGlobus into LDAS**
    - **Use development LDAS to integrate TclGlobus 1.0 release**
    - **Release version of LDAS based on TclGlobus 1.0**
  - **Support community of client-tool developers to move over to TclGlobus**
    - **Develop client tools for interacting with LDAS**