

Visualization on the UK National Grid Service using GAPtk, a generic toolkit

Srikanth Nagella and Lakshmi Sastry.

CCLRC e-Science Centre.

Abstract

Grid Enabling proven and essential data analysis applications and augmenting these with real-time visualization components on the grid is a key that facilitates and accelerates the use of grid technology in everyday science. End users should be able to access data and compute intensive applications on the grid from their commodity desktop systems and via user interface tools that are familiar and appropriate for their applications. In addition, they should have access to advanced visualization utilities that can be linked with their applications to provide near-real time analysis capability for a complete exploitation of the grid for computational efficiency. GAPtk is a multi-layered toolkit which contains a core set of advanced visualisation and interactive services that execute on distributed clusters, which can take input data in a range of formats and outputs geometry that can be imported into a variety of desktop toolkits. GAPtk layers are incorporated into applications using API. The services can be invoked as independent units using WSDL.

1. Introduction

1.1 Motivation

CCLRC is home to some of the world's largest scientific facilities such as the ISIS Neutron Spallation facility and the Synchrotron Light Resource and large science departments that generate huge volumes of data from experiments and simulations of physical processes. High performance compute and data resources on the grid are combined to provide scalable, near real time visualization capability for applications to analyse these data on the UK National Grid Service (NGS). The visualization techniques of importance here are those which are in themselves computationally intensive or for which the pre-processing is or need to handle very large datasets. This motivated in developing a toolkit that solve the visualisation operations under conditions stated above on clusters such as NGS and which provides a services based interface and local rendering on the client desktop.

1.2 Background

The Grid middleware for performing high end computations and data storage is maturing rapidly. NGS is providing a stable environment for UK and international access to a large number of users. Hence, it provides an ideal platform for the implementation of ur GAPtk visualization services. The visualization pipeline process involves the raw data undergoing a series of filtering and other forms

of transformations, mapped onto abstract visualization objects which are then represented in terms of graphics primitives such as points, lines and triangles for rendering. These primitives can be displayed on a computer screen or printed on a sheet of paper. One of the drawbacks of traditional approach is that it is not often scalable and hence slower when the data sizes increase or the filtering and mapping involve complex algorithms. One of the methods to perform scalable near real time data analysis and visualization is to introduce a parallel paradigm into traditional serial system as well as distribute the nodes of the pipeline on a grid for faster/better computational outcome. The schematic of Grid enabled data flow pipe line (courtesy Wood and Brodlie) is shown below.

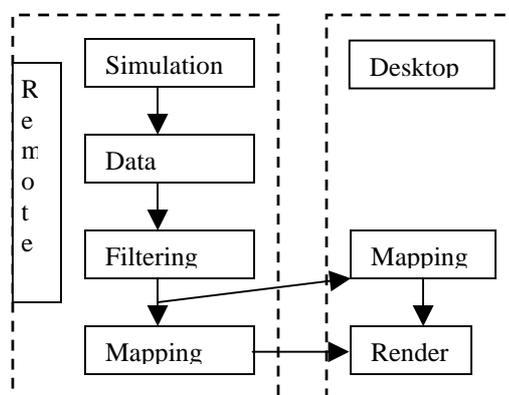
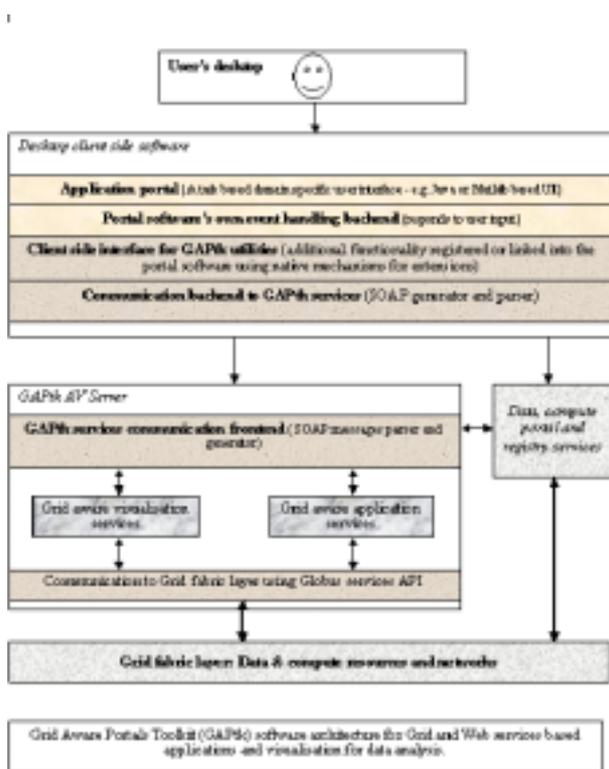


Fig 1. Schematic of grid-enabled dataflow pipeline model for visualization (Wood and Brodlie)

Our GAPtk model is an extension of the above. Our method involves parallel implementation of the visualization algorithms alongside the adoption of a services oriented interface and communication protocol with GridFTP based third party data transfer to reduce load on the desktop client. The interface on the client desktop and the visualization and analysis use cases and scenarios were drawn from extensive requirements capture from a range of application specific projects such as GODIVA[1], Integrative Biology[2] and eVe[3] projects and a generic specification derived from these. The core visualization services are now available on the UK NGS and are being made use of in a variety of applications and projects. We are developing a script interface to import output from the GAPtk services to the RAVE interface too [4].

2. Architecture

The overview GAPtk architecture is given in Figure 2 below:



GATk services are invoked using a variety of communication protocols such as HTTP, SOAP as well as using GT2 API directly from within client interfaces. Desktop interfaces to services are built using thin and thick clients. The services front-end are provided by an application visualisation server which is able to

invoke remote Grid-based services without the client needing to use Grid software itself. The client communication backend is a generic downloadable client-side library. This library also provides modules to interface the results from the server onto native graphics toolkits and problem solving environments such as VTK, IRIS Explorer and Matlab.

2.1 User scenarios

A typical client may browse, query and request data from any third party or external databases and pass the URL of the requested data to the application server. GAPtk server gets client's grid credentials from MyProxy Server, generates the necessary input information for the application and the RSL script to invoke the application on NGS. If the third party data services do not support GridFTP, the server fetches the data using appropriate data transfer protocols.

The application on NGS computes any and all requested derived quantities, transforms numerical data into geometric representations using GAPtk core visualization services and then passes back this reduced quantity of data to the client for local visualization and interaction. The user is able to specify dynamically the resolution of data he/she desires and conversely, the system is able to gather information on the hardware capabilities of the user's application environment and send data in default resolution and format.

The system also provides an interface to Storage Resource Broker (SRB) This enables the visualization services on the NGS to access simulation data directly from SRB without going through the client to generate geometry.

3. Services

GAPtk is based on services oriented architecture which provides the flexibility for any client tool to access the services through a uniform, standard API. This level of transparency provides compatibility and portability.

GAPtk provides two types of services:

- Data manipulation services
- Visualisation services

In data manipulation services, there is several derived quantity, data extraction and filtering computation services are available. Most of these services are generic and others are specific to research projects that were used to gather the requirements.

Visualisation services are classified into basic and advanced visualisation services based on the complexity and usability. There are several readers and writers provided for reading in data to and writing out data from the services. Most popular file formats such as netCDF, HDF4/5, VTK and OpenInventor are catered for. The data is read from the input files are internally mapped onto GAPtk generic data structures and all the visualisation operations are performed on this data structure. This allows us to expand or modify the services independent of readers and clientside tools.

3.1 Basic Visualization Services

Following gives a sample list of visualization services that are implemented on the Grid and available using WSDL:

- *Contour generation service* generates a 2D contour for the given data input in parallel.
- *Isosurface generation service* cater to both structured and unstructured mesh based data. GAPtk isosurface generation services have marching cubes and tetrahedron algorithms implemented to exploit multi processor architecture of the NGS.
- *Cutting plane generation service* generates a series of cut planes across the geometry in any user specified direction where the parameters of the orientations of the plane are gathered from the clientside as the user interacts with the plane. This service also allows the images to be stitched together to generate a movie at the end of an interactive session.
- *Image processing service*, this service takes images as inputs and performs different image processing algorithms such as edge detection, etc.
- *3D Reconstruction service*, this service will take a series of images from MRI scans or any experiment and generate a 3D surface.
- *Time Series Animation service* generates a view fixed, time based animation of time series data. The animation can be stored in MPEG-1/2 or AVI format.
- *Script based Animation service* generates animation based on the script which contain camera position. The GAPtk client-side library captures the mouse coordinates and

the camera parameters and records these in script while the user interacts with the visualised data.

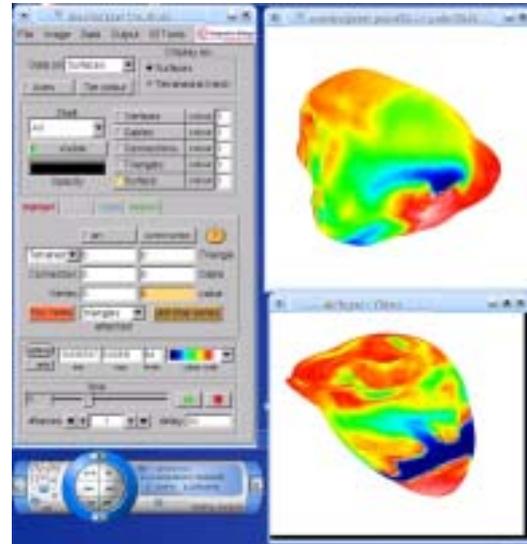


Fig.3 GAPtk grid enabled animation service visualized using Meshalyzer

Figure 3 above show electrical potential distribution across the surface of heart. The heart at the top is interacted with directly and also using the GUI to its left. The heart at the bottom is the associated movie with the MPEG player controls to the left.

- *Publication quality Postscript service* generates a postscript image from the data.

4. Client Side Library

GAPtk Client side library is usable from a variety of desktop problem solving environments and visualization toolkits used by scientists. The core is developed in C++ and an interface layer is provided in a variety of high level languages such as Python and Java. Specific mapping to Matlab, IRIS Explorer are also available.

Figures 4 and 5 below are some sample client applications integrated with GAPtk services to generate real time visualization for GODIVA project.

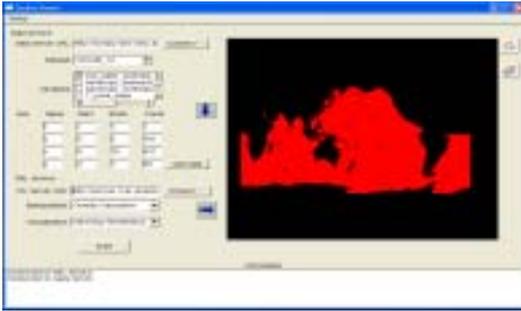


Fig. 4 A GODIVA client developed using fltk and GAPtk client libraries.

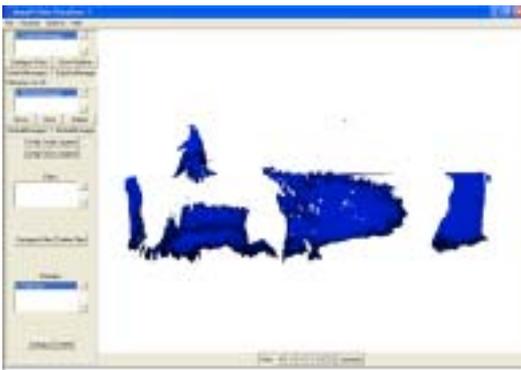


Fig. 5 Imported isosurface output into MayaVi

5. Conclusion

We have now implemented a few data analysis grids, using grid enabled visualization services from GAPtk. The application grids support seamless access to the resource-critical applications on the grid from desktop commodity processors, providing near real-time interactivity. The visualization data is produced in a generic data format and mapped onto a variety of other proprietary format for specific applications. The examples are applications under the NERC funded GODIVA project and the EPSRC funded Integrative Biology project. These application grids are taken up by their end-users as part of their everyday scientific activity.

6. References

- <http://www.e-science.cclrc.ac.uk/web/projects/godiva>
- <http://www.integrativebiology.ox.ac.uk>
- <http://www.e-science.cclrc.ac.uk/web/projects/eve/>
- <http://www.wesc.ac.uk/projects/rave>
- <http://mayavi.sourceforge.net/>