Doing Your Science While You’re in Orbit

Mark L. Green
Tech-X Corporation

Stephen D. Miller, Sudharshan S. Vazhkudai, and James R. Trater
Oak Ridge National Laboratory
mlgreen@txcorp.com

Abstract. Large-scale neutron facilities such as the Spallation Neutron Source (SNS) located at Oak Ridge National Laboratory need easy-to-use access to Department of Energy Leadership Computing Facilities and experiment repository data. The Orbiter thick- and thin-client and its supporting Service Oriented Architecture (SOA) based services (available at https://nucleus.sns.gov) consist of standards-based components that are reusable and extensible for accessing high performance computing, data and computational grid infrastructure, and cluster-based resources easily from a user configurable interface. The primary Orbiter system goals consist of (1) developing infrastructure for the creation and automation of virtual instrumentation experiment optimization, (2) developing user interfaces for thin- and thick-client access, (3) provide a prototype incorporating major instrument simulation packages, and (4) facilitate neutron science community access and collaboration. The secure Orbiter SOA authentication and authorization is achieved through the developed Virtual File System (VFS) services, which use Role-Based Access Control (RBAC) for data repository file access, thin- and thick-client functionality and application access, and computational job workflow management. The VFS Relational Database Management System (RDMS) consists of approximately 43 database tables describing 429 user accounts with 437 groups over 451,000 directories with 739,618 repository files. Over 39 million NeXus file metadata records are associated to the 12,794 unique NeXus file field/class names generated from the 42,894 repository NeXus files. Services that enable (a) summary dashboards of data repository status with Quality of Service (QoS) metrics, (b) data repository NeXus file field/class name full text search capabilities within a Google like interface, (c) fully functional RBAC browser for the read-only data repository and shared areas, (d) user/group defined and shared metadata for data repository files, (e) user, group, repository, and web 2.0 based global positioning with additional service capabilities are currently available. The SNS based Orbiter SOA integration progress with the Distributed Data Analysis for Neutron Scattering Experiments (DANSE) software development project is summarized with an emphasis on DANSE Central Services and the Virtual Neutron Facility (VNF). Additionally, the DANSE utilization of the Orbiter SOA authentication, authorization, and data transfer services best practice implementations are presented.

1. Background and Development
The Spallation Neutron Source (SNS) [1] and the High Energy Isotope Reactor (HFIR) [2] support worldwide users on neutron instruments for scientific research utilizing neutron scattering science. The Neutron Science TeraGrid Gateway (NSTG) [3] has initiated a work force consisting of HPC software experts and SNS instrument scientists to make TeraGrid resources available to facility researchers. A limited
software facility infrastructure has been designed, developed, and deployed on the prototype NSTG. It is being tested on select resources across TeraGrid and available to SNS/HFIR as well as other neutron scientists. The SNS Portal project is based on a Java Applet application and is fully functional serving the SNS/HFIR user communities [4].

While SNS is operating at full power, it will offer unprecedented performance for neutron-scattering research, with more than an order of magnitude higher flux than any existing facility. To realize the potential this offers for research in many fields, a world-class suite of instruments is being developed that makes optimal use of the SNS beams and that is suited to the needs of users across a broad range of disciplines. Instruments will be available to researchers with varying degrees of experience, from new graduate students and first-time neutron users to experienced users with an interest in instrument design. This broad range of users, novice to sophisticated, provides an ideal user community for integration and development of the Multitier Portal Architecture (MPA) in addition the user community requirements are also quite disparate based on their science domains. The Tech-X Corporation [5] has been working with the Scientific Computing Group at SNS in developing a MPA for this purpose.

It is extremely important that this system be as flexible as possible to support the wide range of use cases presented by the user community. The SNS software services are available from remote web clients (thin-clients) and desktop clients and user applications (thick-clients). Instrument scientists responsible for the operation of the instruments during experiments require the use of thin- and thick-clients for performing their duties and subsequent calibration and data analysis.

The basic portal architecture is modular with well-defined interfaces facilitating a distributed development environment with the ability to incorporate new technologies seamlessly. Today many portals are required to perform in an extremely dynamic environment where the quantity, sophistication, and requirements of portal users change rapidly. It is highly desirable to scale the portal capabilities with the requirements and sophistication of the user. In order to effectively and efficiently provide a scalable, robust, and flexible presentation tier we present the following multitiered approach providing thin-client servlets, applets, portlets, and web services in addition to thick-client Rich Client Platform (RCP) and Integrated Development Environment (IDE) support. The MPA system capabilities are shown in Figure 1. The Multitier Portal Architecture has the potential to scale to the expertise and requirements level of the user while maximizing the reuse of the developed infrastructure and services in successive tiers. With a carefully planned and defined service architecture each subsequent tier can significantly reuse and leverage existing services provided by the less sophisticated tiers. The simplest MPA Tier 1 can be accessed from a commodity browser whereas the feature rich MPA Tier 4 requires the local installation of an Integrated Development Environment (IDE) that has the capability of modifying/enhancing/customizing the entire MPA architecture.
Tech-X has obtained a Small business Innovative Research (SBIR) grant from the Department of Energy for the development of a prototype MPA at the SNS (award #: DE-FG02-08ER85000). As illustrated in the system architecture diagram, Figure 2, the Multitiered Portal Architecture (MPA) Orbiter Overview the MPA Orbiter RCP Thick-Client and the MPA Orbiter Thin-Client Portal share several system components within the Orbiter Service Oriented Architecture (SOA) implemented in this prototype development effort.

The cyber security measures implemented at the SNS are quite extensive and have required constant attention during the Orbiter development process. The XCAM-UCAMS [6] authentication system is an in-house authentication/authorization authority that serves as the core cyber security identity management service at ORNL (we have adopted this system as our identity management system for the Orbiter project). The SNS Portal Services provide valuable system components for integrating with the SNS-ORNL computational Resources and External Computational Resources (i.e. Application Manager) [7] and is a project funded by the SNS Scientific Computing Group. We have worked in close collaboration with this group during our entire Orbiter development project.

One of the most valuable core Orbiter components is the Orbiter Virtual File System (VFS) SOA services. The VFS Relational Database Management System (RDMS) consists of approximately 43 database tables describing 429 user accounts with 437 groups over 451,000 directories with 739,618 repository files. Over 39 million NeXus [8] file metadata records are associated with the 12,794 unique NeXus file field/class names generated from the 42,894 repository NeXus files. Services that enable (a) summary dashboards of data repository status with Quality of Service (QoS) metrics, (b) data
repository NeXus file field/class name full text search capabilities within a Google like interface, (c) fully functional RBAC browser for the read-only data repository and shared areas, (d) user/group defined and shared metadata for data repository files, (e) user, group, repository, and web 2.0 based global positioning with additional service capabilities are currently available. The baseline system in place utilizes the free open-source SQL full-text search engine Sphinx [9] for queries. The Orbiter VFS is a key component enabling the robust system development and feature rich user interface and system metrics, Figure 3. The Amazon Web Service (AWS) [10] model is used for securing RESTful web service requests within the Orbiter SOA.

2. Conclusion and Future Work
The current prototyping project conducted in collaboration with the SNS was a tremendous success. Several of the Orbiter SOA services are performing within the SNS production environment providing value to the scientific researchers. In particular the Orbiter find NeXus file service has received nearly 6,000 service requests over a 24-hour period and the Orbiter search NeXus file service provides full index searches, typically taking on the order of tenths of seconds, on over 39 million distinct field/class values. Using the Sphinx indexer that is approximately 5,000 times faster than a comparable SQL based search in this case enhances the Orbiter search capabilities. Our collaboration with the Scientific Computing Group at the SNS has been extremely productive and we continue to leverage the Orbiter infrastructure for other related projects.

The Tech-X Corporation SBIR Principal Investigator, Dr. Green, has completed the DOE SBIR Phase I feasibility and prototype that is currently served by the SNS. The subsequent SBIR Phase II proposal is currently pending and proposes several Orbiter system enhancements based on the prototype collaboration investigation. The key features that will be integrated into the Orbiter system will be a) enhanced support for virtual instrumentation, b) development of the thick-client user interface design, c) integrate ISAW [11] and DANSE [12] project applications as Software as a Service (SaaS), and d) build support for the Orbiter VFS SOA distributed base station deployment. This proposed work as well as additional enhanced capabilities would be performed in the DOE SBIR Phase II effort over a two-year project term if it were awarded.

References
[1] SNS: http://neutrons.ornl.gov/facilities/facilities_sns.shtml
[2] HFIR: http://neutrons.ornl.gov/hfir/hfir_facilities.shtml
[3] NSTG: http://www.teragrid.org/gateways/projects.php?id=56
[4] SNS Portal: http://neutrons.usr
[5] Tech-X Corporation: http://www.txcorp.com/
[6] ORNL XCAMS: https://xcams.ornl.gov/xcams/
[7] OIC Clusters: https://oic.ornl.gov/
[8] NeXus: http://www.nexusformat.org/Main_Page
[9] Sphinx: http://www.sphinxsearch.com/
[10] Amazon Web Services: http://aws.amazon.com/
[11] ISAW: http://www.pns.anl.gov/computing/isaw/index.shtml
[12] DANSE: http://danse.us/