SC12 Presentation Schedule
RENCI/North Carolina, Booth #3640

Salt Palace Convention Center, Salt Lake City, Utah

Monday, November 12

7 p.m. – 9 p.m.: Enterprise iRODS and the E-iRODS Consortium

Presenters: Leesa Brieger and Charles Schmitt, RENCI

Description: This presentation will contain an outline of the advantages furnished by iRODS technology that explain the wide popularity of iRODS in data-intensive domains. We will outline the need for enterprise-grade support for this technology and describe the development of E-iRODS. The description of the E-iRODS Consortium will make the point that this represents a first step toward enhancing sustainability for the technology by broadening its funding base, while also bringing users together to help coordinate the E-iRODS road map. (Rolling slide show during the Gala—will go for the full two hours unless someone else wants their work to roll for one of the hours)

Tuesday, November 13

10:30 a.m.: Workflow-Driven Elastic IaaS for Data Intensive Workflow Applications

Presenters: Paul Ruth and Anirban Mandal, RENCI
Contributors: Ilia Baldine, Yufeng Xin, Chris Heermann, Jeff Tilson and Steve Cox, RENCI; Jeff Chase, Duke University

Description: This demonstration will show the ability of ORCA to provision infrastructure on-demand to execute a scientific workflow. More specifically, ORCA allocates a slice of resources including computational resources from multiple ExoGENI sites including RENCI (Chapel Hill, NC), BBN (Boston, MA), University of Houston (Houston, TX), Florida International University (Maimi, FL). In addition, ORCA provides bandwidth-provisioned network connections between sites (using Starlight, NLR, ESnet, and BEN).   The slice will be a private sandbox in which a complex scientific workflow that is generated by requests for resources in response to the workflow's current resource requirements. For more information, see www.exogeni.net.

11:30 a.m.: Supporting Persistent Queries for PerfSONAR Measurement Data to Monitor Application and Network Performance

Presenters: Anirban Mandal and Paul Ruth
Contributors: Ilia Baldine, Yufeng Xin, and Chris Heerman, RENCI; Jeff Chase, Duke University

Description: We will demonstrate a Persistent Query Agent (PQA) that supports persistent queries for perfSONAR measurement data. Applications/clients would be able to register queries concerning performance metrics of interest. They would get asynchronous notification through an XMPP pubsub mechanism when their queries are satisfied. We have used a complex event processing (CEP) engine called Esper for managing and executing the queries expressed in a standard SQL-like query language - the Esper Event Processing language (EPL). EPL enables expressing complex matching conditions that include temporal windows, joining of different event streams, as well as filtering, aggregation,

and sorting. The PQA also includes a perfSONAR WS-client that feeds events into the Esper engine. The perfSONAR client would generate the event streams by querying one or more perfSONAR measurement archives (MA) services.

1:30 p.m.: Enterprise iRODS and the E-iRODS Consortium
Presenters: Leesa Brieger and Charles Schmitt, RENCI

Description: This presentation will contain an outline of the advantages furnished by iRODS technology that explain the wide popularity of iRODS in data-intensive domains. We will outline the need for enterprise-grade support for this technology and describe the development of E-iRODS. The description of the E-iRODS Consortium will make the point that this represents a first step toward enhancing sustainability for the technology by broadening its funding base, while also bringing users together to help coordinate the E-iRODS road map.

2:30 p.m: iRODS Data Grid

Presenter: Reagan Moore, RENCI and UNC-Chapel Hill

Description: The ability to register and share data sets has been augmented with the ability to register, share, and re-execute workflows. Within the iRODS data grid, researchers can collaborate on research projects, register the input files and parameters for a workflow, register the workflow, and store the workflow result sets. A workflow can be re-executed with changed parameters, and the results will be automatically save, enabling comparison across analyses. Examples of workflows will be given, along with a demonstration of the iRODS data grid.

3:30 p.m.: Approaches in large-scale informatics for clinical genomics

Presenter: Charles Schmitt, RENCI

Description: A serious challenge in reaching the promise of next generation genomic sequencing is dealing with the scale of informatics infrastructure that is required to manage data and produce meaning knowledge from the data. These challenges include everything from high performance computing needed to process hundreds to thousands of samples, to management of Terabytes to Petabytes of data being used by various researchers and clinicians, to annotation of genomic data that ties raw sequences data to clinically-relevant knowledge. Such informatics most operate in an environment that facilitate research into the relationship between genetics and medical phenotypes while still delivering quality results to clinicians. We review the informatics technologies that have been developed at UNC to facilitate analysis and management of thousands of patient whole and exome genomic data for clinical care.

4:30 p.m. Data Science at RENCI: Enabling Data Usage in the Scientific Enterprise.

Presenter: Howard Lander, RENCI

Description: The Renaissance Computing Institutes at UNC Chapel Hill has a strong research focus in the area of data science as applied to the scientific enterprise, particularly in data management. This focus has led us to a number of interesting projects. This presentation includes an overview and short descriptions for a selection of these projects.

Wednesday, November 14

10:30 a.m. Enterprise iRODS and the E-iRODS Consortium
Presenters: Leesa Brieger and Charles Schmitt, RENCI

Description: This presentation will contain an outline of the advantages furnished by iRODS technology that explain the wide popularity of iRODS in data-intensive domains. We will outline the need for enterprise-grade support for this technology and describe the development of E-iRODS. The description of the E-iRODS Consortium will make the point that this represents a first step toward enhancing sustainability for the technology by broadening its funding base, while also bringing users together to help coordinate the E-iRODS road map.

11:30 a.m. Scientific Data Management at RENCI

Presenter: *Nassib Nassar, RENCI*

Description: Complex science and the large volumes of disparate data required to analyze risks from coastal hazards makes it difficult to communicate these risks to government and business decision makers. Factoring in the potential impacts of climate change further complicates this process. The RENCI team will present a tool that helps communicate risk to non-technical audiences: an immersive visualization environment that integrates data from high-resolution images, sensed and measured sources, model output and more. The environment scales from the desktop to dome theatre venues.

1:30 p.m.: Workflow-Driven Elastic IaaS for Data Intensive Workflows Applications

Presenters: Paul Ruth and Anirban Mandal, RENCI
Contributors: Ilia Baldine, Yufeng Xin, Chris Heermann, Jeff Tilson and Steve Cox, RENCI; Jeff Chase, Duke University

Description: This demonstration will show the ability of ORCA to provision infrastructure on-demand to execute a scientific workflow. More specifically, ORCA allocates a slice of resources including computational resources from multiple ExoGENI sites including RENCI (Chapel Hill, NC), BBN (Boston, MA), University of Houston (Houston, TX), Florida International University (Maimi, FL). In addition, ORCA provides bandwidth-provisioned network connections between sites (using Starlight, NLR, ESnet, and BEN). The slice will be a private sandbox in which a complex scientific workflow that is generated by requests for resources in response to the workflow's current resource requirements. For more information, see www.exogeni.net.

2:30 p.m.: Supporting Persistent Queries for PerfSONAR Measurement Data to Monitor Application and Network Performance

Presenters: Anirban Mandal and Paul Ruth
Contributors: Ilia Baldine, Yufeng Xin, and Chris Heerman, RENCI; Jeff Chase, Duke University

Description: We will demonstrate a Persistent Query Agent (PQA) that supports persistent queries for perfSONAR measurement data. Applications/clients would be able to register queries concerning performance metrics of interest. They would get asynchronous notification through an XMPP pubsub mechanism when their queries are satisfied. We have used a complex event processing (CEP) engine called Esper for managing and executing the queries expressed in a standard SQL-like query language

- the Esper Event Processing language (EPL). EPL enables expressing complex matching conditions that include temporal windows, joining of different event streams, as well as filtering, aggregation, and sorting. The PQA also includes a perfSONAR WS-client that feeds events into the Esper engine. The perfSONAR client would generate the event streams by querying one or more perfSONAR measurement archives (MA) services.

3:30 p.m. Approaches in large-scale informatics for clinical genomics

Presenter: Charles Schmitt, RENCI

Description: A serious challenge in reaching the promise of next generation genomic sequencing is dealing with the scale of informatics infrastructure that is required to manage data and produce meaning knowledge from the data. These challenges include everything from high performance computing needed to process hundreds to thousands of samples, to management of Terabytes to Petabytes of data being used by various researchers and clinicians, to annotation of genomic data that ties raw sequences data to clinically-relevant knowledge. Such informatics most operate in an environment that facilitate research into the relationship between genetics and medical phenotypes while still delivering quality results to clinicians. We review the informatics technologies that have been developed at UNC to facilitate analysis and management of thousands of patient whole and exome genomic data for clinical care.

4:30 p.m. Data Science at RENCI: Enabling Data Usage in the Scientific Enterprise.

Presenter: Howard Lander, RENCI

Description: The Renaissance Computing Institutes at UNC Chapel Hill has a strong research focus in the area of data science as applied to the scientific enterprise, particularly in data management. This focus has led us to a number of interesting projects. This presentation includes an overview and short descriptions for a selection of these projects.

Thursday, November 15

10:30 a.m.: Workflow-Driven Elastic IaaS for Data Intensive Workflows Applications

Presenters: Paul Ruth and Anirban Mandal, RENCI
Contributors: Ilia Baldine, Yufeng Xin, Chris Heermann, Jeff Tilson and Steve Cox, RENCI; Jeff Chase, Duke University

Description: This demonstration will show the ability of ORCA to provision infrastructure on-demand to execute a scientific workflow. More specifically, ORCA allocates a slice of resources including computational resources from multiple ExoGENI sites including RENCI (Chapel Hill, NC), BBN (Boston, MA), University of Houston (Houston, TX), Florida International University (Maimi, FL). In addition, ORCA provides bandwidth-provisioned network connections between sites (using Starlight, NLR, ESnet, and BEN).   The slice will be a private sandbox in which a complex scientific workflow that is generated by requests for resources in response to the workflow's current resource requirements. For more information, see www.exogeni.net.

11:30 a.m.: Supporting Persistent Queries for PerfSONAR Measurement Data to Monitor Application and Network Performance

>*Presenters:* *Anirban Mandal and Paul Ruth*
>*Contributors: Ilia Baldine, Yufeng Xin, and Chris Heerman, RENCI; Jeff Chase, Duke University*

Description: We will demonstrate a Persistent Query Agent (PQA) that supports persistent queries for perfSONAR measurement data. Applications/clients would be able to register queries concerning performance metrics of interest. They would get asynchronous notification through an XMPP pubsub mechanism when their queries are satisfied. We have used a complex event processing (CEP) engine called Esper for managing and executing the queries expressed in a standard SQL-like query language - the Esper Event Processing language (EPL). EPL enables expressing complex matching conditions that include temporal windows, joining of different event streams, as well as filtering, aggregation, and sorting. The PQA also includes a perfSONAR WS-client that feeds events into the Esper engine. The perfSONAR client would generate the event streams by querying one or more perfSONAR measurement archives (MA) services.

1 p.m.: Enterprise iRODS and the E-iRODS Consortium

>*Presenters:* *Leesa Brieger and Charles Schmitt, RENCI*

Description: This presentation will contain an outline of the advantages furnished by iRODS technology that explain the wide popularity of iRODS in data-intensive domains. We will outline the need for enterprise-grade support for this technology and describe the development of E-iRODS. The description of the E-iRODS Consortium will make the point that this represents a first step toward enhancing sustainability for the technology by broadening its funding base, while also bringing users together to help coordinate the E-iRODS road map. (Rolling slide show during the Gala)

2 p.m.: Scientific Data Management at RENCI

>*Presenter:* *Nassib Nassar, RENCI*

Description: Complex science and the large volumes of disparate data required to analyze risks from coastal hazards makes it difficult to communicate these risks to government and business decision makers. Factoring in the potential impacts of climate change further complicates this process. The RENCI team will present a tool that helps communicate risk to non-technical audiences: an immersive visualization environment that integrates data from high-resolution images, sensed and measured sources, model output and more. The environment scales from the desktop to dome theatre venues.