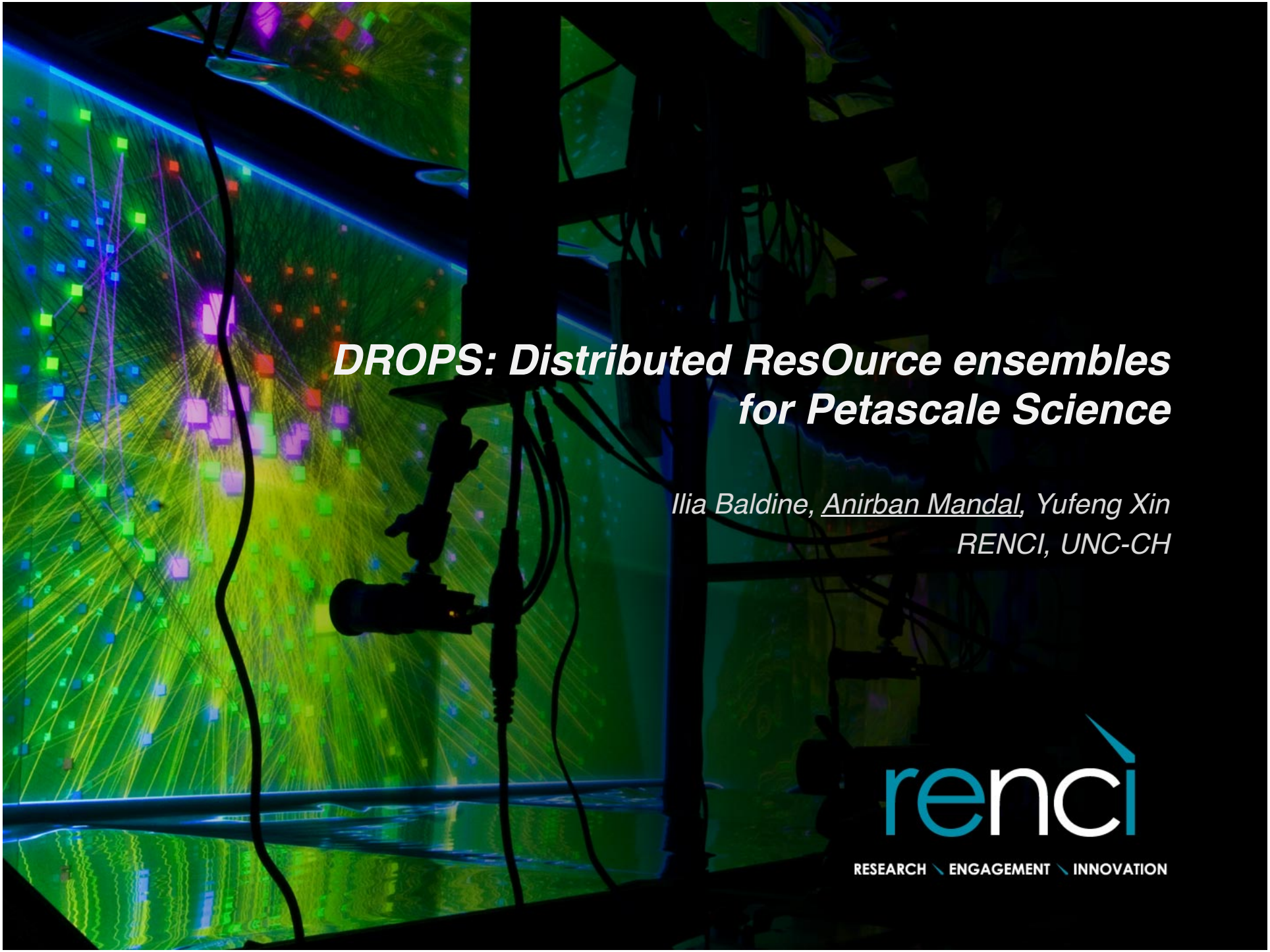


DROPS: Distributed ResOurce ensembles for Petascale Science

Ilia Baldine, Anirban Mandal, Yufeng Xin
RENCI, UNC-CH

renci
RESEARCH \ ENGAGEMENT \ INNOVATION

DROPS: Distributed Resource Ensembles for Petascale Science

Objectives

- Provide an API to provision distributed heterogeneous resources.
- Select representative science applications and access performance on networked clouds.
- Create persistent query mechanism for closed-loop performance feedback
- Support application performance monitoring and resource provisioning
- Extend semantic resource representations to measurements

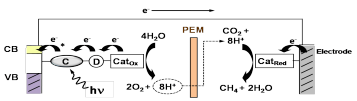
Impact

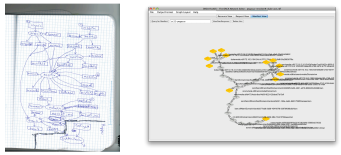
- Impact design of application driven resource provisioning systems.
- Semantic resource descriptions will standardize abstractions in resource requests.
- Closed-loop feedback help steer workflow execution and adjust resource allocation.
- Impact how a broader set of DOE applications can leverage high-performance multi-layer networked clouds.

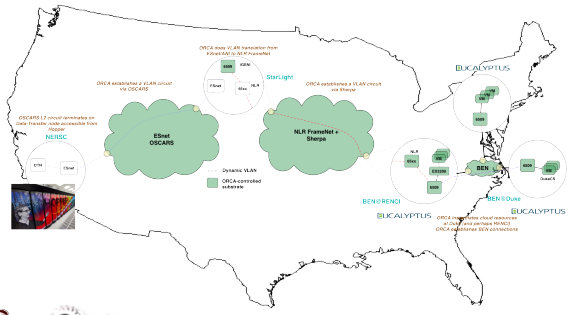
Year 1 and 2 Accomplishments

- Accessed performance of Hadoop based map/reduce applications and a solar fuels workflow application on networked clouds
- Created simple API for resource request using semantic resource descriptions
- Demonstrated it with above applications
- Developed a persistent query agent (PQA) for closed-loop performance feedback

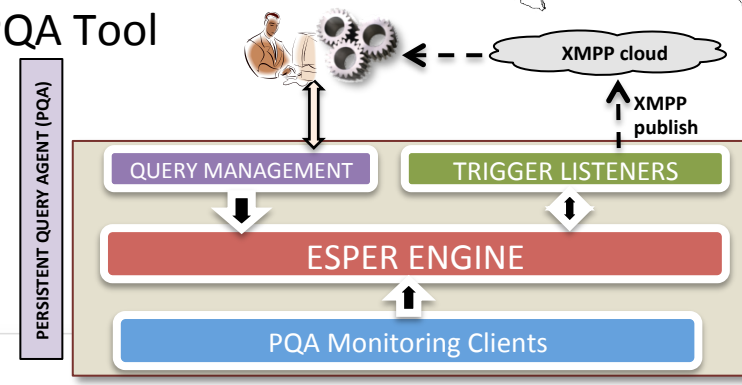
renci







PQA Tool



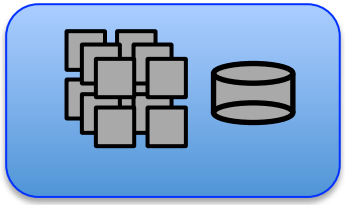
Overview

- DROPS Quad Chart
- Background – Networked clouds and ORCA
- DROPS overview and Year 1 work review
- Year 2 work – Persistent Query Agent
- Future directions

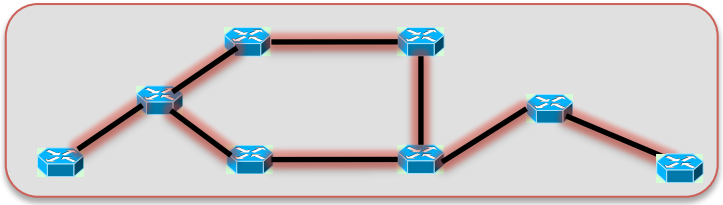
renci

NIaaS: Compute and Network Virtualization

Virtual Compute and Storage Infrastructure

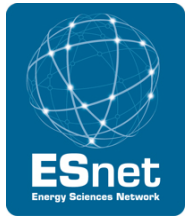


Virtual Network Infrastructure



Cloud APIs (Amazon EC2 ..)

Network Provisioning APIs (DOE ESNet OSCARS, Internet2 DRAGON, OGF NSI …)

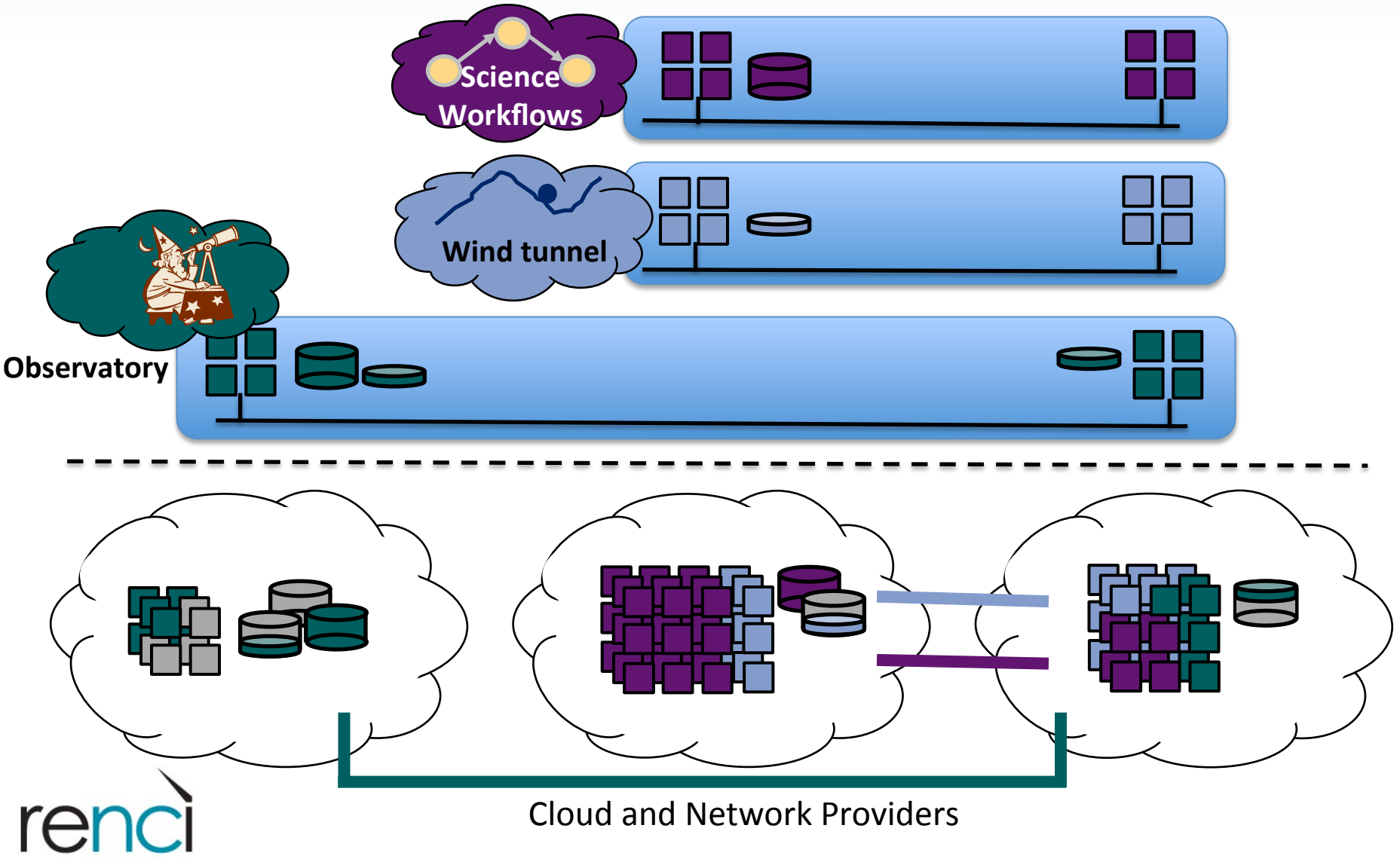


renci

Cloud Providers

Network Transit Providers

Networked Clouds



renci

ORCA

Open Resource Control Architecture

- ORCA is a “wrapper” for off-the-shelf cloud technologies and circuit networks etc., enabling federated orchestration:
  - + Resource brokering
  - + VM image distribution
  - + Topology embedding
  - + Stitching
  - + Authorization
- GENI, NSF SDCI
- http://networkedclouds.org
- http://geni-orca.renci.org

EUCALYPTUS

ESnet Energy Sciences Network

NLR

openstack

INTERNET2

XCAT

amazon web services™

BEN Breakable Experimental Network

renci

ORCA Capabilities

- Co-scheduling/co-provisioning of heterogeneous resources (primarily compute and network)
  - Automatic binding of resources to available sites
  - Automatic splitting of resources between sites
- Stitching of resources into connected topologies
  - Deducing and honoring resource dependencies
- Semantic resource descriptions used on user-facing APIs (NDL-OWL)
- Multi-layered network provisioning on BEN
  - Fiber, DWDM (Infinera) and L2 (Cisco, Juniper) provisioning
  - Using a combination of heuristics and ILP
- Support for
  - OpenStack, Eucalyptus, OpenFlow, OSCARS, Sherpa
  - Low-level drivers for Cisco and Juniper switches

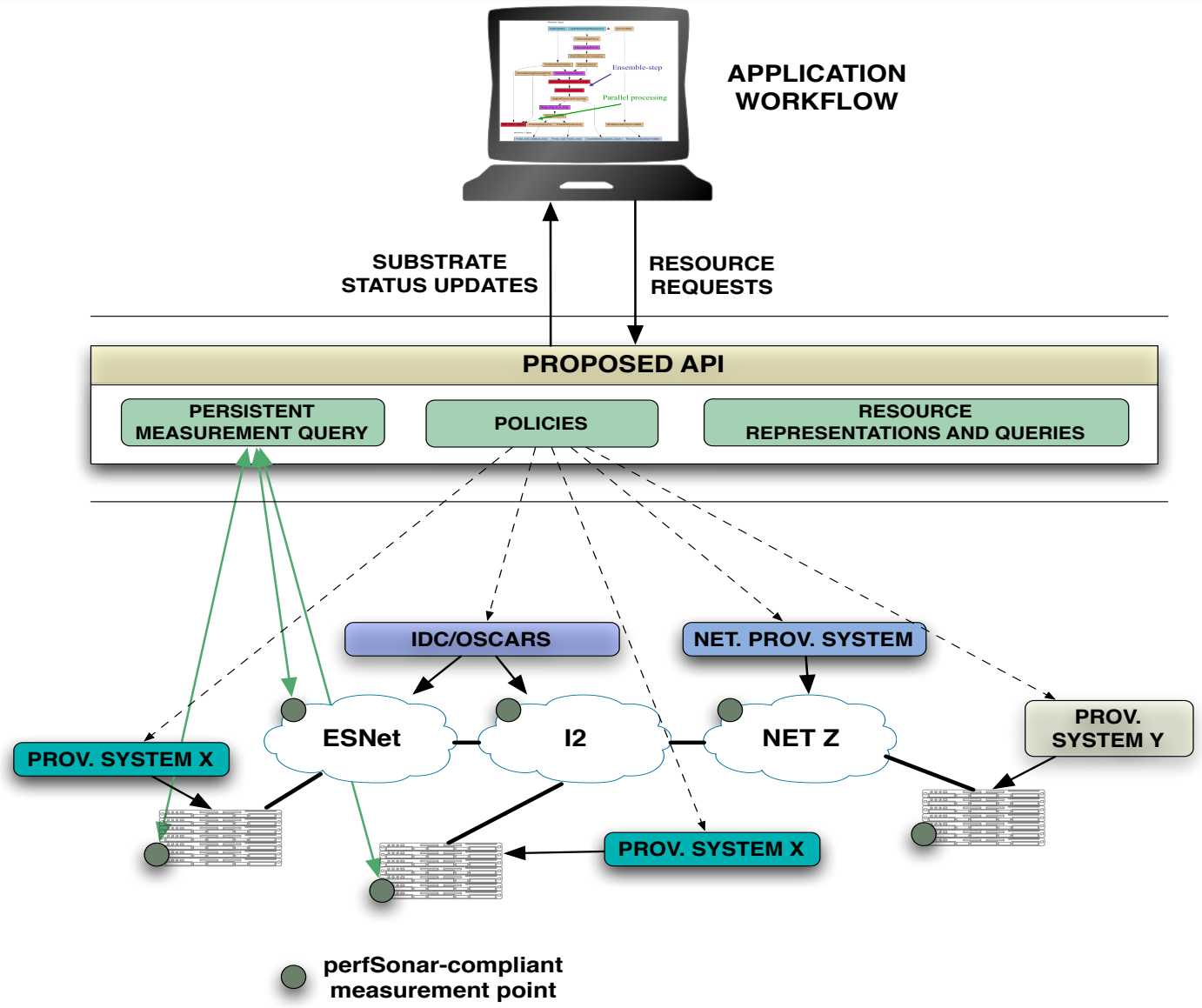
renci

Overview

- DROPS Quad Chart
- Background – Networked clouds and ORCA
- DROPS overview and Year 1 work review
- Year 2 work – Persistent Query Agent
- Future directions

renci

DROPS: Application-driven Resource Orchestration



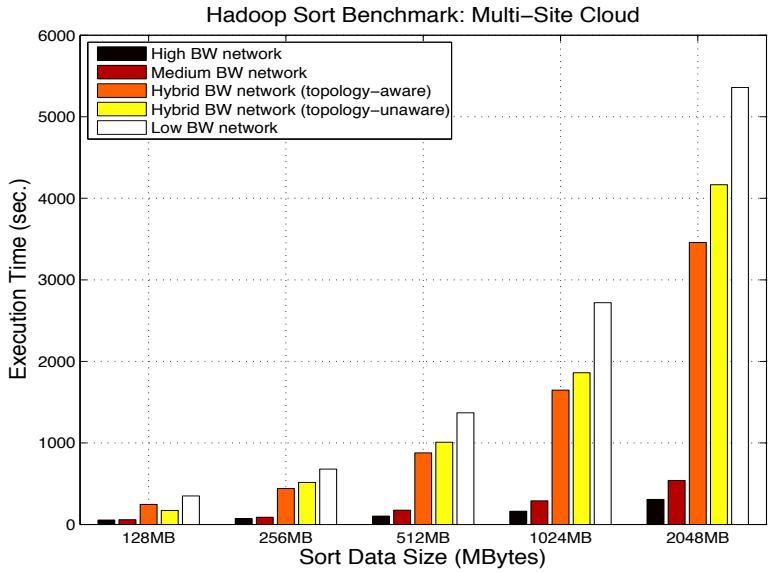
DROPS Goals

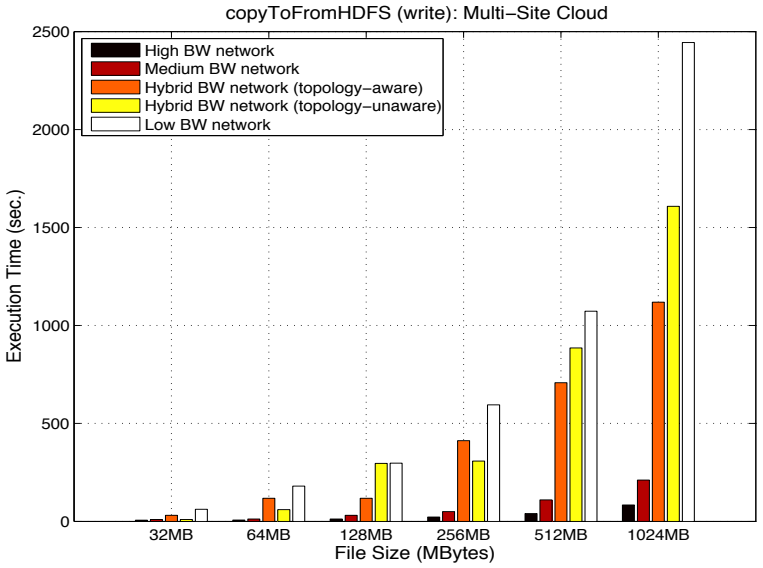
- Select representative scientific applications
  - Map/Reduce, EFRC Solar Fuels Pipeline
  - Assess their performance
- Create API for applications to create and modify ‘slices’
  - Simple API, complex semantic resource representations open to reasoning and inference
- Create persistent query mechanism for perfSonar and other measurement sources to support closed feedback loop for application performance monitoring and resource provisioning
- Extend Semantic Resource resource representations to performance measurements and metrics

renci

Year 1 work

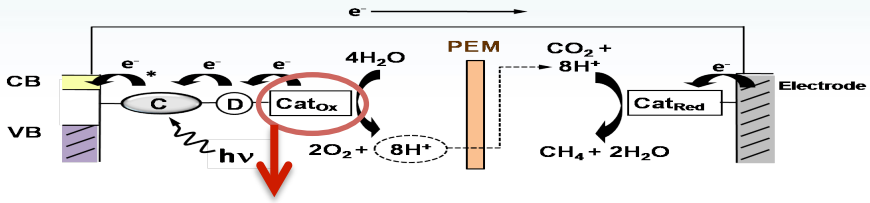
- Assess performance of Hadoop in slices with varying topologies and link latencies
- Convert EFRC workflow to Pegasus. Evaluate launching the workflow in a slice.





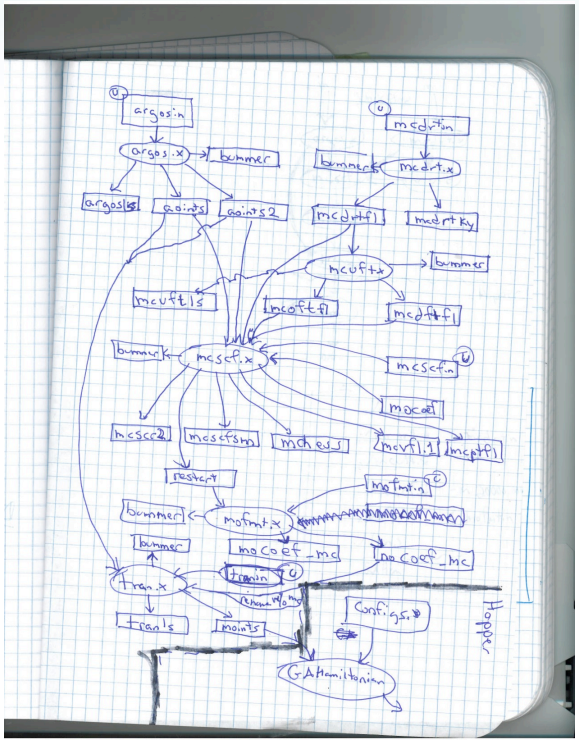
renci

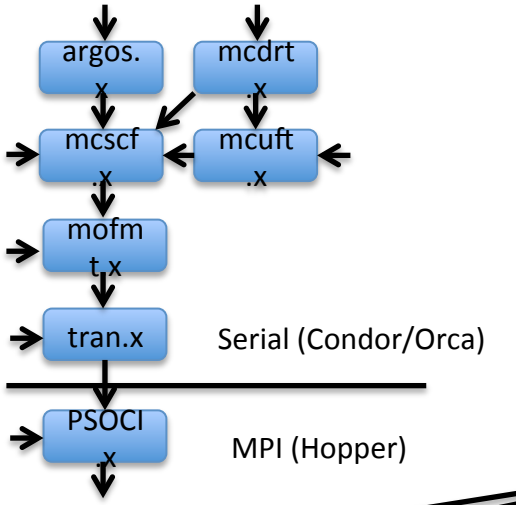
Year 1 - Solar Fuels Workflow – EFRC (DOE)

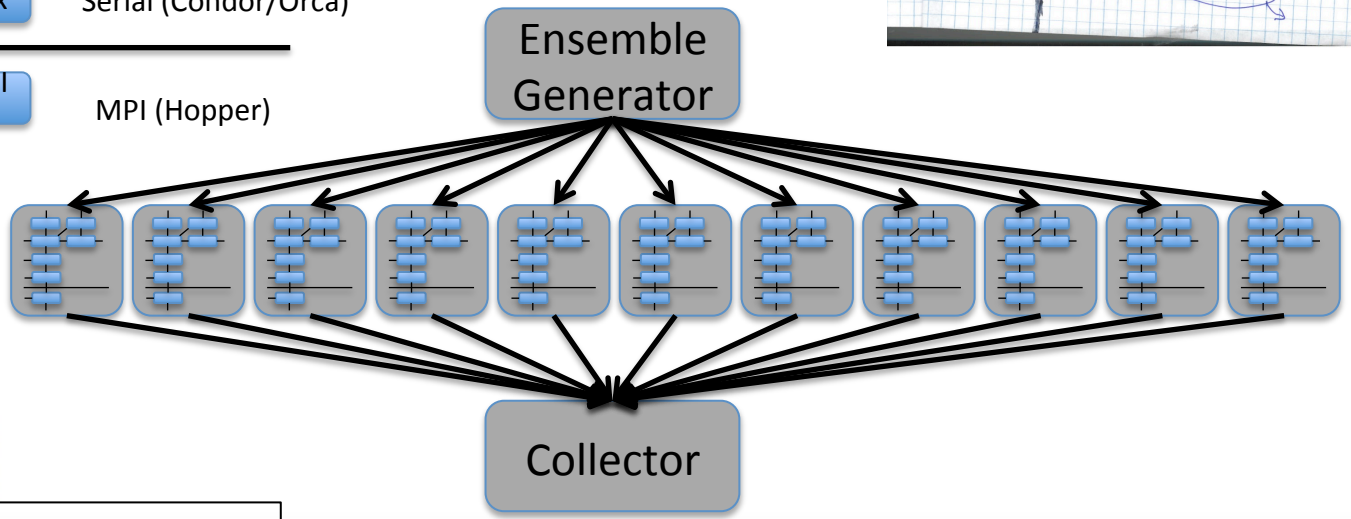


Oxidation catalysts

Image provided by UNC-CH EFRC: http://www.efrc.unc.edu/



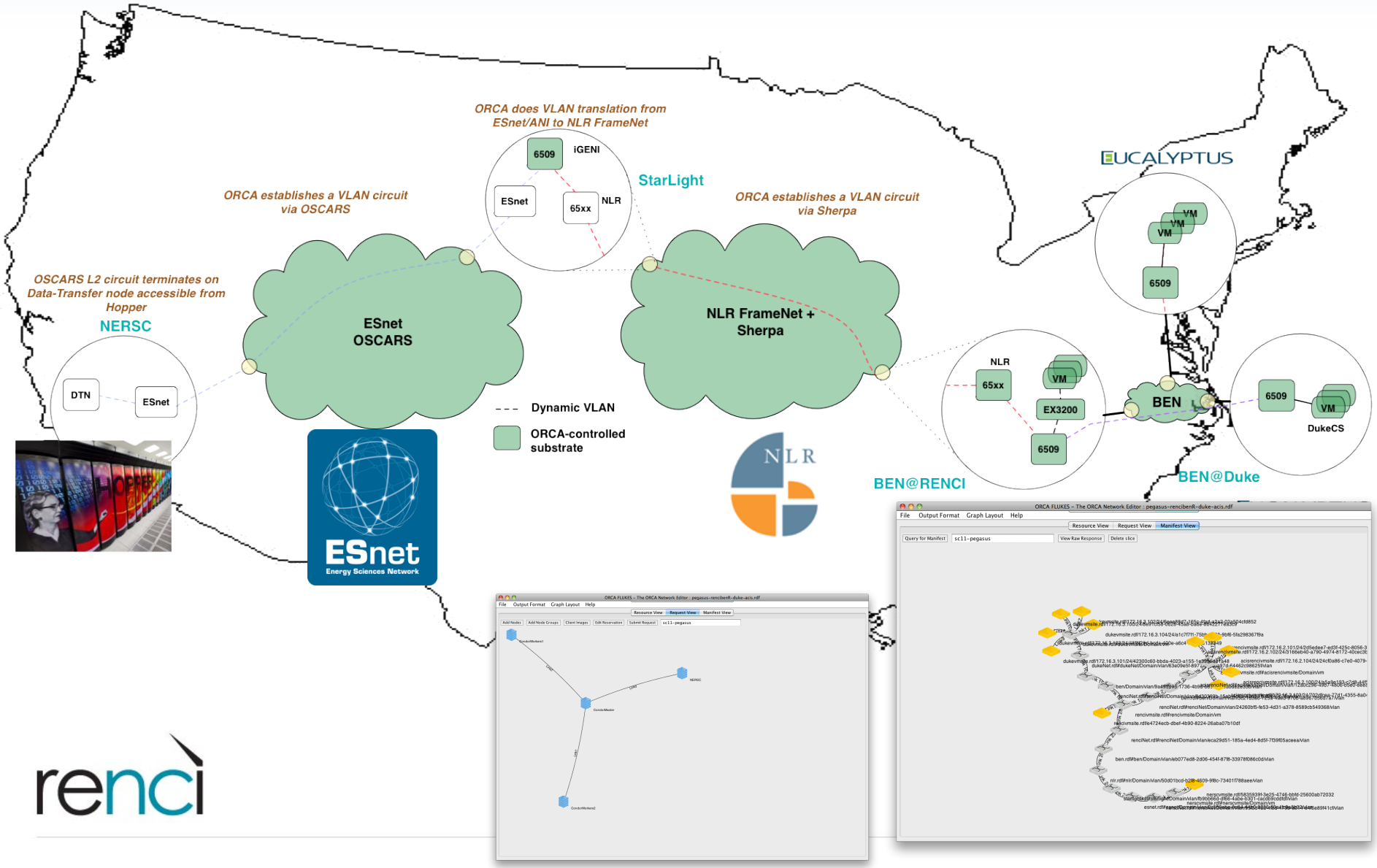




renci

Slide courtesy: Paul Ruth, Jeff Tilson

Resources used to run EFRC workflow



renci

Overview

- DROPS Quad Chart
- Background – Networked clouds and ORCA
- DROPS overview and Year 1 work review
- Year 2 work – Persistent Query Agent
- Future directions

renci

Goals

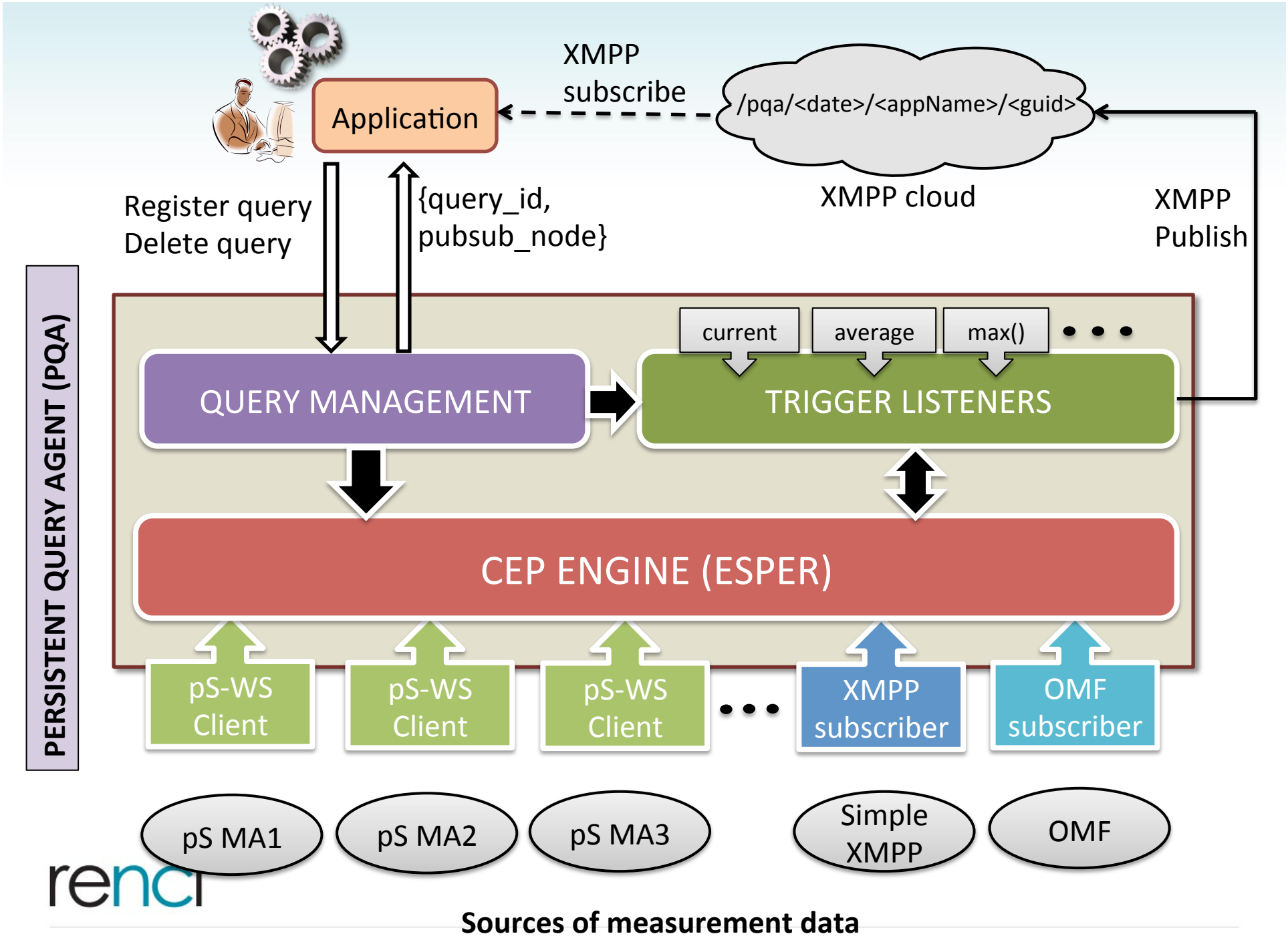
- Federated performance monitoring from diverse aggregates
  - Manage performance data from several measurement sources - perfSONAR MAs, xmpp based measurement streams, OMF, and others
- Scalability – manage volume and velocity of monitoring data
  - a large portion of monitoring data is not useful
- Real-time performance feedback to applications/tools
  - Enables applications to steer based on performance feedback
  - Enables resource provisioning adaptations
- Applications/clients should be able to express what metrics and threshold conditions are important
  - System should be able to asynchronously send notifications only when “interesting” things happen
- Gain high-level insights important to distributed applications
  - Entails taking a cross-aggregate view of application and network performance

renci

Persistent Query Agent (PQA) Tool

- PQA supports persistent queries for perfSONAR and other measurement data to monitor application and network performance
  - Interacts with diverse aggregates to obtain measurement data – federated monitoring
- Applications/clients can register queries with PQA concerning performance metrics and events of interest
- PQA runs persistent queries in real-time without storing measurement data
  - Addresses scalability
- Applications/clients get real-time, asynchronous notification using an XMPP pubsub mechanism when their queries get satisfied
  - Enables closed-loop performance feedback

renci



Sources of measurement data

Persistent Query Agent (PQA)

- Leverages 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
  - open-source, available from http://www.espertech.com
- ESPER behaves like a database turned “upside-down”
  - Queries “persist” in the ESPER system
  - Data and events are not stored, rather “watched” and analyzed as they pass by
- PQA Includes pluggable “trigger listeners”
  - These are activated when query conditions are met
  - Can peek into current event stream when activated and apply functions
  - Publish to designated nodes in the XMPP space using XMPP pubsub (XEP-0060 extensions for XMPP)

renci

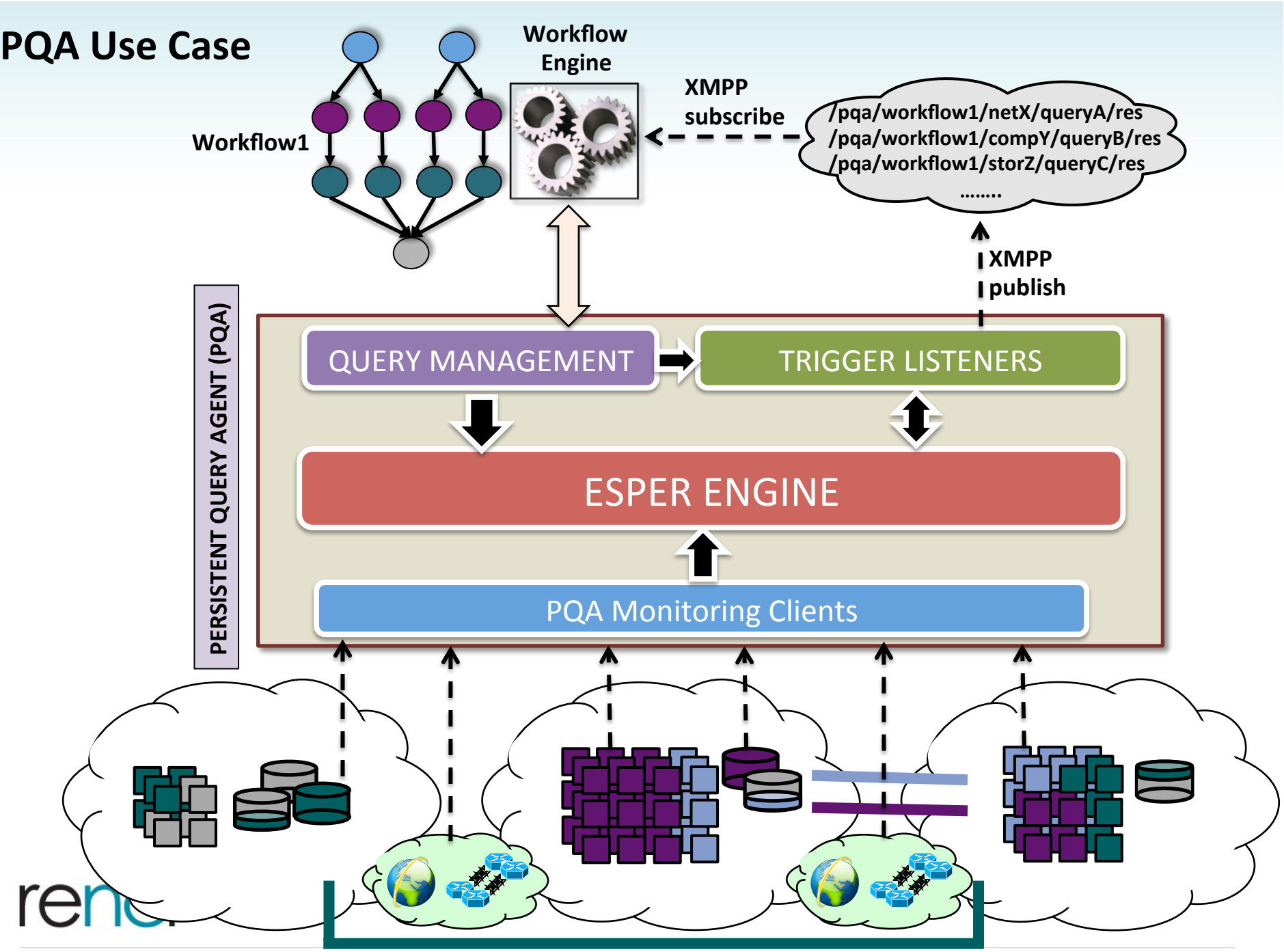
Persistent Query Agent (PQA)

- Query management
  - Applications/clients register EPL queries with PQA using an XML-RPC interface
  - A pubsub node handle is returned to the application/client
  - Application/client subscribes to this node handle to get asynchronous notifications
- PQA includes perfSONAR & XMPP-based monitoring clients
  - Feed events into the ESPER engine
  - perfSONAR client generates event streams by querying one or more perfSONAR measurement archives (MA) services
  - XMPP based monitoring client subscribes to pubsub nodes associated with monitored application metrics and generates ESPER event streams when new events are published

renci

*"**select** b.metricName as metricName1, b.metricValue as metricValue1, m.metricName as metricName2, m.metricValue as metricValue2*
from
BWUtilization.win:length(1) as b, MemoryUtilization.win:length(1) as m
***where** b.metricValue > 1.40012E9 and m.metricValue > 70"*

PQA Use Case


Workflow Engine
Workflow1
XMPP subscribe
/pqa/workflow1/netX/queryA/res
/pqa/workflow1/compY/queryB/res
/pqa/workflow1/storZ/queryC/res
……..
XMPP publish
PERSISTENT QUERY AGENT (PQA)
QUERY MANAGEMENT
TRIGGER LISTENERS
ESPER ENGINE
PQA Monitoring Clients

Related Work

- Multi-domain monitoring tools
  - perfSONAR provides hooks for delivering performance measurements in federated environments – responsibility of higher level tools to utilize them in an application specific manner
  - MonALISA, INTERMON, ENTHRONE, EuQoS, configurable multi-domain monitoring architecture (Belghith et. al ICOIN'11)
  - OMF – gathering measurements from experiments on network test-beds
- Automated ways of using (perfSONAR) measurement data
  - OnTimeDetect tool (Calyam et. al) - network anomaly detection and notification for perfSONAR deployments
  - Pythia (Kanuparthy et. al) - a data analysis tool that makes use of perfSONAR data to detect, localize and diagnose wide-area network performance problems
  - Kissel et. al (IEEE NOMS'12) - measurement and analysis framework to automate troubleshooting of end-to-end network bottlenecks using perfSONAR data

renci

Overview

- DROPS Quad Chart
- Background – Networked clouds and ORCA
- DROPS overview and Year 1 work review
- Year 2 work – Persistent Query Agent
- Future directions

renci

Future DROPS Directions

- Application integration with PQA
  - Close the loop – use performance feedback for future application planning
- Ontology extensions for measurements
  - Performance measurement resources
  - Application performance measurement metrics
- Slice elasticity
  - Allocate resources in reaction to workflow behavior

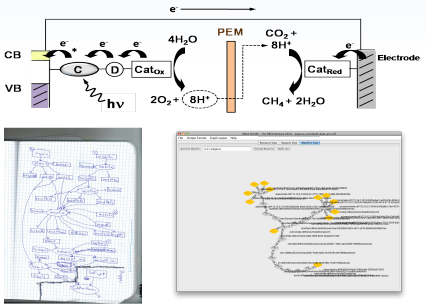
renci

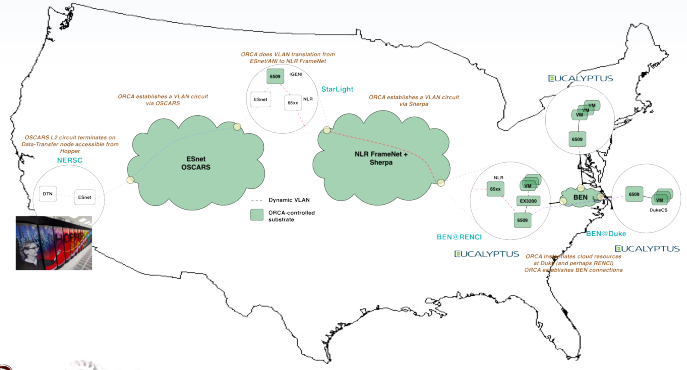
Recent publications

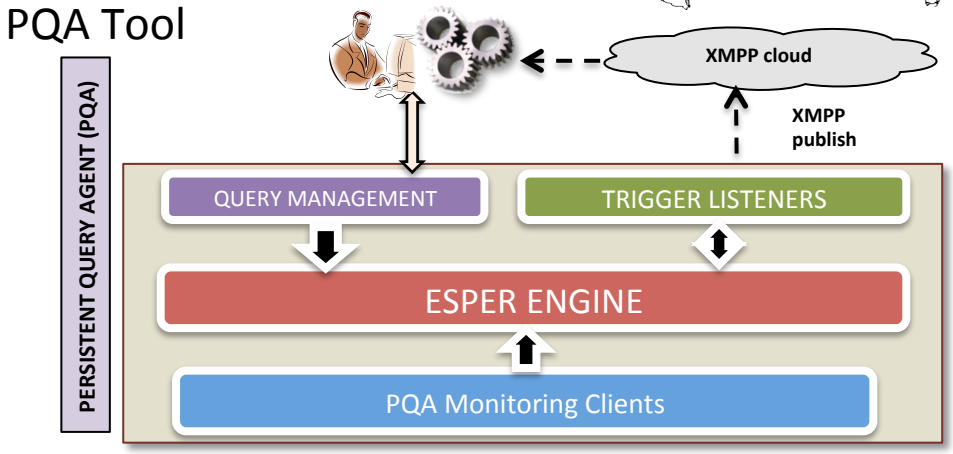
- ***Enabling Persistent Queries for Cross-aggregate Performance Monitoring.*** A. Mandal, I. Baldine, Y. Xin, P. Ruth, C. Heerman. In submission to IEEE Communications Magazine 2013.
- ***Dynamic Network Provisioning for Data Intensive Applications in the Cloud***. P. Ruth, A. Mandal, Y. Xin, I. Baldine, C. Heerman, J. Chase. In proceedings of IEEE 8th International Conference on E-Science, vol., no., pp. 1,2, 8-12 Oct. 2012.
- **Provisioning and Evaluating Multi-domain Networked Clouds for Hadoop-based Applications.** A.Mandal, Y.Xin, I.Baldine, P.Ruth, C.Heermann, J.Chase, V.Orlikowski and A. Yumerefendi. In Proceedings of IEEE CloudCom 2011, Dec 2011.
- ***Embedding Virtual Topologies in Networked Clouds.*** Y. Xin, I. Baldine, A. Mandal, C. Heermann, J. Chase, A. Yumerefendi. In Proceedings of CFI 2011, June 2011
- ***Autonomic Cloud Network Orchestration: A GENI Perspective.*** I.Baldine, J. Chase, Y.Xin, D. Irwin, V. Marupadi, A. Mandal, C. Heermann, A. Yumerefendi. IEEE International Workshop on Management of Emerging Networks and Services (IEEE MENS 2010).

renci

Thank you







DOE Award#: DE-FG02-10ER26016/DE-SC0005286

renci

U.S. DEPARTMENT OF ENERGY

renci
RESEARCH \ ENGAGEMENT \ INNOVATION