Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Data Provenance Hybridization Supporting Extreme-Scale Scientific Workflow Applications

TODD ELSETHAGEN

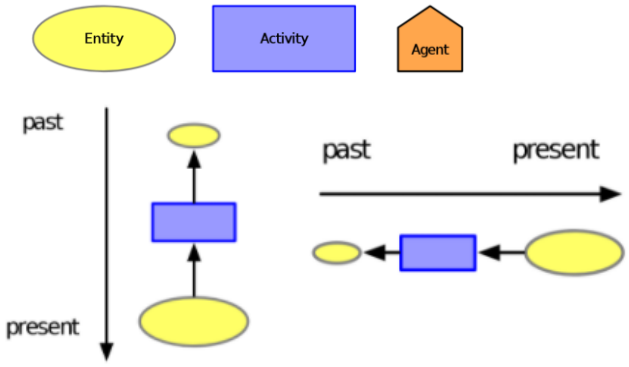
Pacific Northwest National Laboratory

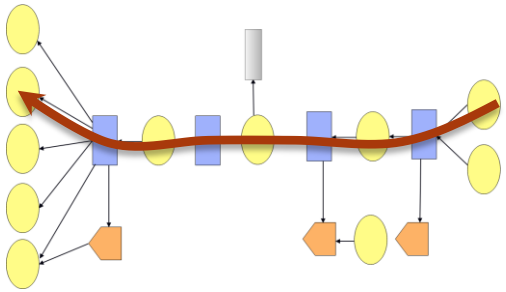
2016 New York Scientific Data Summit

U.S. DEPARTMENT OF ENERGY

Provenance Definitions

- The origin or source of something
- A detailed historical explanation of influential factors, process flows, data flows used to accomplish work.
- The chronology of the ownership, custody or location of a historical object
- Provenance is information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability or trustworthiness [W3C PROV].





Provenance Graph
Tracing Data Origin

Provenance at scale

- **Minimize impact,** control granularity (coarse to fine) and retention of provenance
- **Retrieval,** how to retrieve, explore, and analyze large amounts of collected provenance
- **Scalability,** provenance collection from concurrent large-scale scientific workflows will require a scalable solution
- **Dynamic interference,** provide real-time monitoring and analysis to support runtime workflow steering
- **Context**, integrate system level data to extend provenance descriptions
- **Provenance by Design,** provenance disclosure designed for workflow domain objectives:
  - *Reproducibility, Results Explanation, Performance Optimization, Anomaly Detection, Monitoring, Others…*

Discussion Outline

- Provenance Environment (ProvEn) Overview
- Producer API (PAPI) Library
- Provenance Cluster
- Hybrid Store
- Use Cases
- Next Steps

Discussion Outline

- Provenance Environment (ProvEn) Overview
- PAPI Library
- Provenance Cluster
- Hybrid Store
- Use Cases
- Next Steps

Provenance Environment (ProvEn) Overview

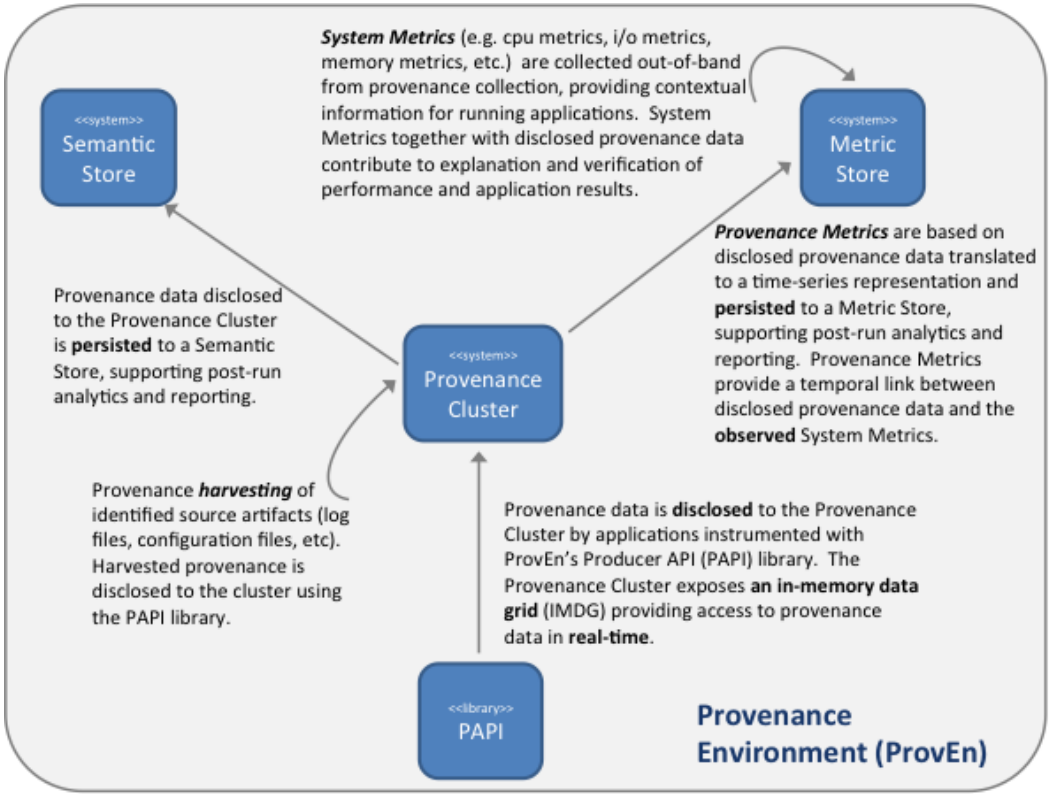
Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- **ProvEn** is a provenance management platform consisting of loosely coupled components supporting the disclosure, storage, and access to provenance information.

- ***Producer API (PAPI)***
  - ProvEn's provenance disclosure library. Scientific workflow applications instrumented with PAPI can produced and disclose their provenance data.

- ***Provenance Cluster***
  - ProvEn's scalable approach for collecting concurrent provenance data streams from PAPI sources.

- ***Hybrid Store***
  - ProvEn combines system level metrics (**Metric Store**) with the traditional disclosed provenance (**Semantic Store**) to create an extended provenance view.



Provenance Model

- **W3C PROV** data model published in 2013 defines a core data model for provenance for building representations of the entities, people and processes involved in producing a piece of data or thing in the world.

- **Workflow Performance Provenance** (WFPP) data model is an extension to PROV that will enable the empirical study of workflow performance characteristics and variability including complex source attribution.

- **Provenance Environment** (ProvEn) data model provides concepts specific to the ProvEn provenance management software platform.

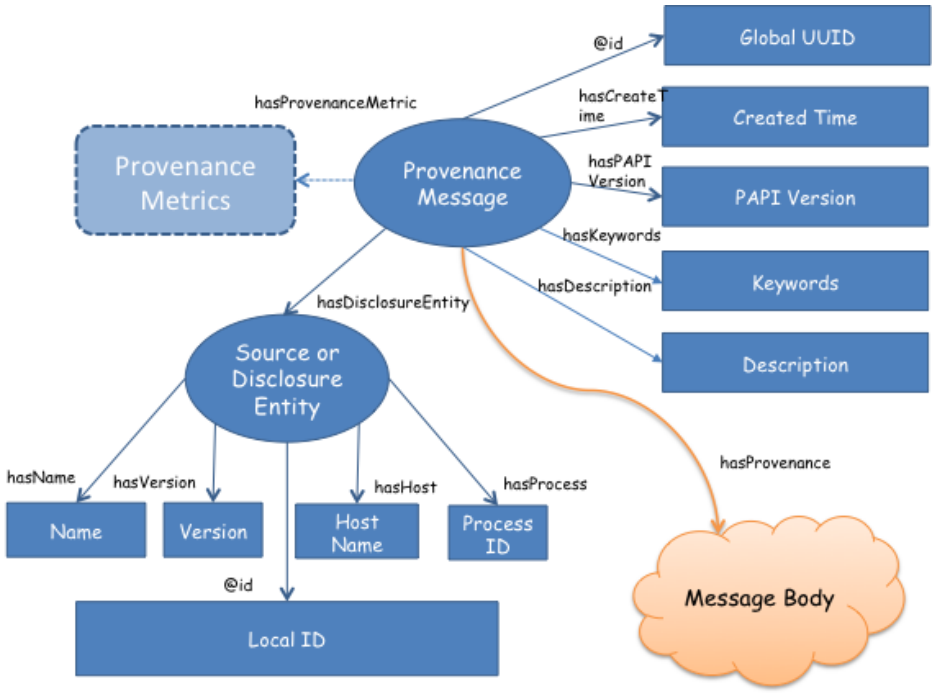
- **Domain Specific** integration

Discussion Outline

- Provenance Environment (ProvEn) Overview
- PAPI Library
- Provenance Cluster
- Hybrid Store
- Use Cases
- Next Steps

PAPI Library
What's Disclosed?

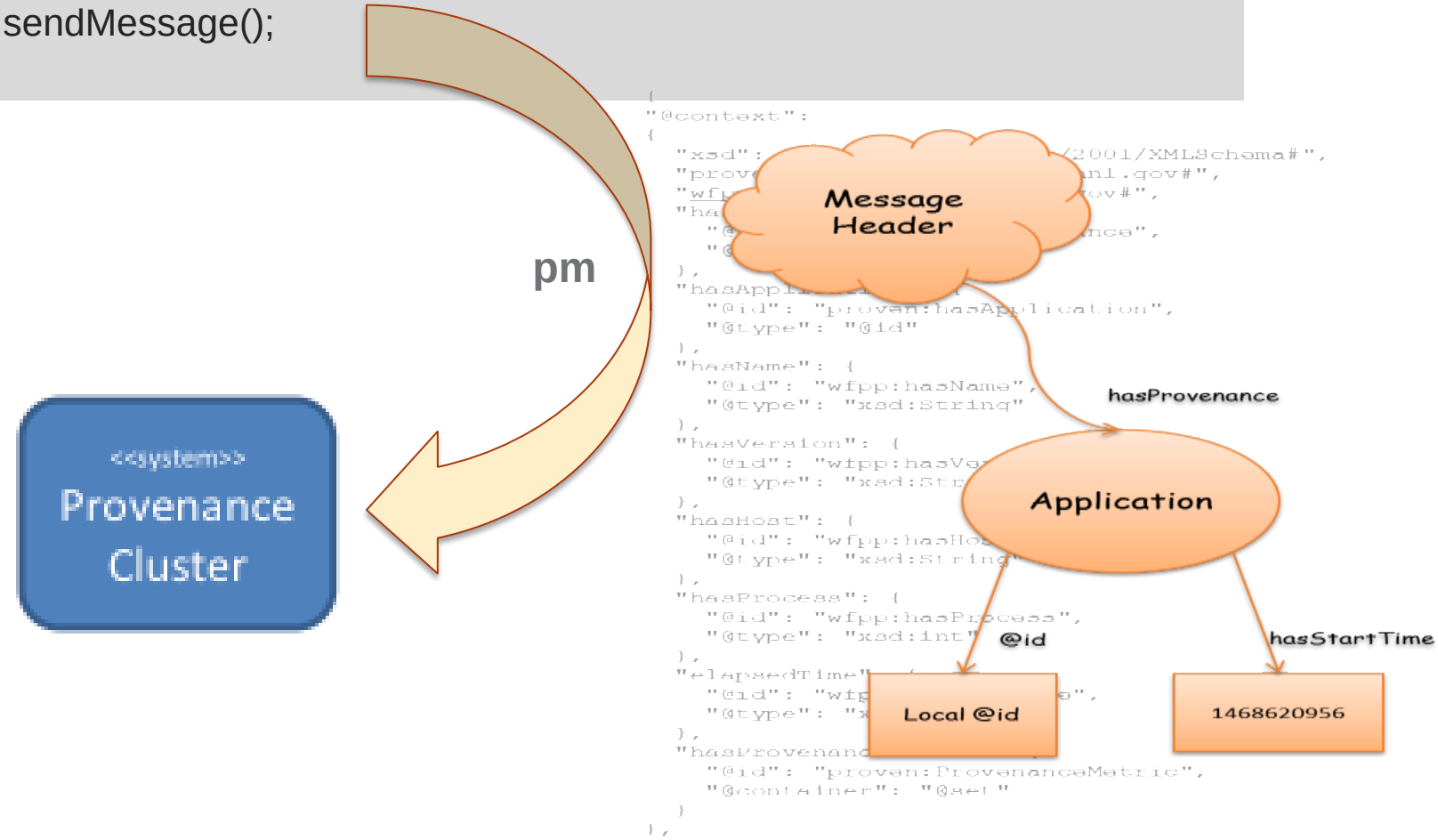
- **Provenance Message**
  - PAPI's "unit of" provenance
  - Each message is a fragment of the complete provenance graph
  - Every message created uses the same structure (*Header* + *Body*)
  - Provenance by design – messages tailored per PAPI distribution. Ad-hoc also supported
  - PAPI can be used outside workflow to harvest provenance from workflow application artifacts (e.g. log files)
  - Messages are serialized as JSON-LD for a direct interchange to Semantic Store – RDF Database
  - Provenance Metrics are identified
  - Offline messaging capability



PAPI Library
Message Example

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

```
…
ProvenanceMessage pm = createMessage(START_APPLICATION);
pm.sendMessage();
…
```



PAPI Library Feature List

- Lightweight minimal call interface.
- Java is reference implementation.
- No external dependencies.
- Compile-time checking for pre-defined messages.
- Provenance production details hidden from caller.
- Library can be used to disclose harvested provenance from application artifacts or direct disclosure from application.
- Warnings not errors; logging for post-run review.
- Disclosure errors retried in background, if failures occur.
- Identical disclosures can be made without side-effects to hybrid store.
- Can be used stand-alone, plans for other output formats.

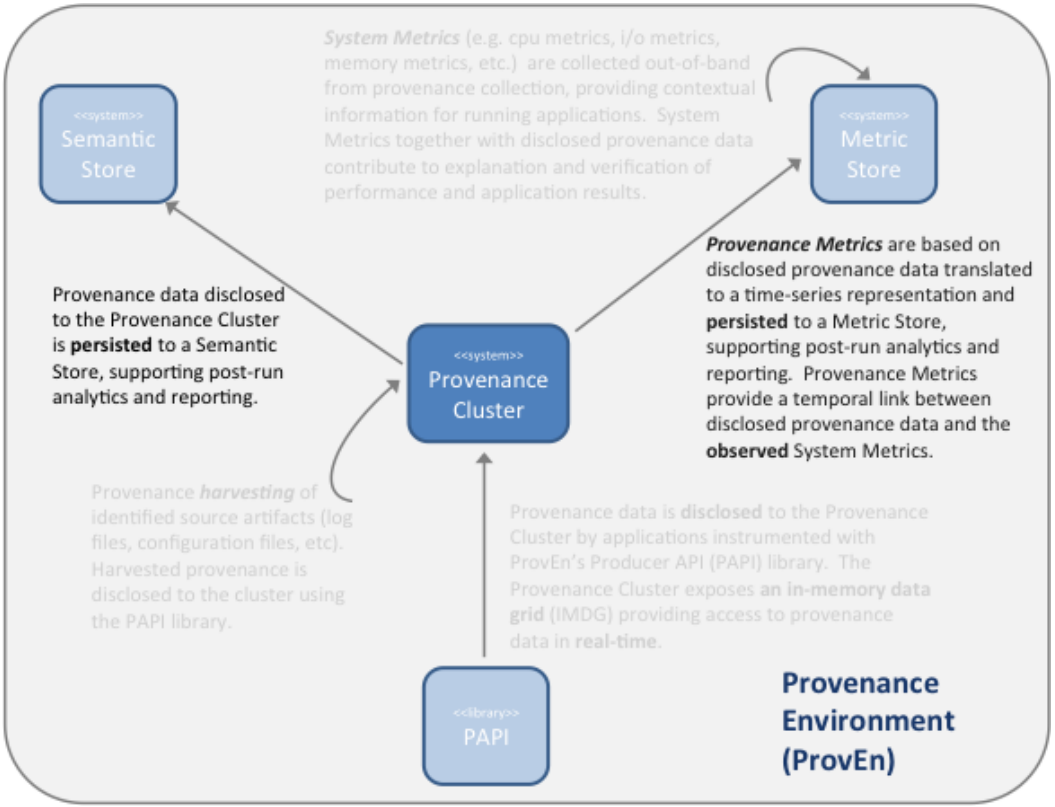
Discussion Outline

- Provenance Environment (ProvEn) Overview
- PAPI Library
- Provenance Cluster
- Hybrid Store
- Use Cases
- Next Steps

Provenance Cluster Overview

- **Provenance Cluster** is ProvEn's scalable approach for collecting concurrent provenance data streams.

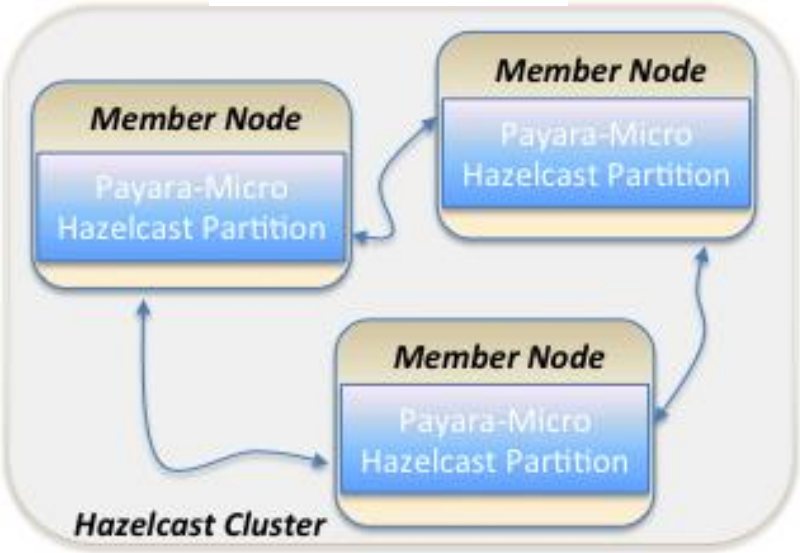
  - Cluster is composed of homogenous **member nodes** arranged in a master-less peer-to-peer network, and scales out/in by adding/removing member nodes
  - Member nodes provide services for storage, access, and monitoring provenance data
  - Member nodes retain provenance data in-memory for real-time access, as well as distributing to the Semantic and Metric Stores for post-run analytics



Provenance Cluster
What's a Member Node?

- **Payara-Micro**
  - Open source Java EE application container
  - Embedded server
  - Hazelcast Integration
  - Payara-Micro API
  - Microservices development approach
- **Hazelcast**
  - Open source in-memory data grid (IMDG) library
  - Auto-clustering provided by Hazelcast
  - Distributed data structures to provide provenance in real-time
  - Distributed compute to perform data distribution
- **Member Node = Payara-Micro + Hazelcast**
  - Elasticity
  - High Availability
  - Provenance processing via microservices

payara
Payara Micro



hazelcast

Provenance Cluster
Cluster Member Microservices

- ***Disclosure Services***: Capture disclosed provenance from PAPI instrumented applications. Incoming provenance is distributed to the IMDG as well as staged for persistence to the semantic and metric stores.

- ***Exchange Services***: Exchanges the staged provenance with the semantic and metric stores. Uses *Hybrid Services* to perform the actual exchange.

- ***Hybrid Services***: Provides storage and retrieval access to the constituents (semantic, metric, and IMDG) of the hybrid store.

- ***Member Services***: Provides status and statistics for a cluster member to help support scale-out decisions.

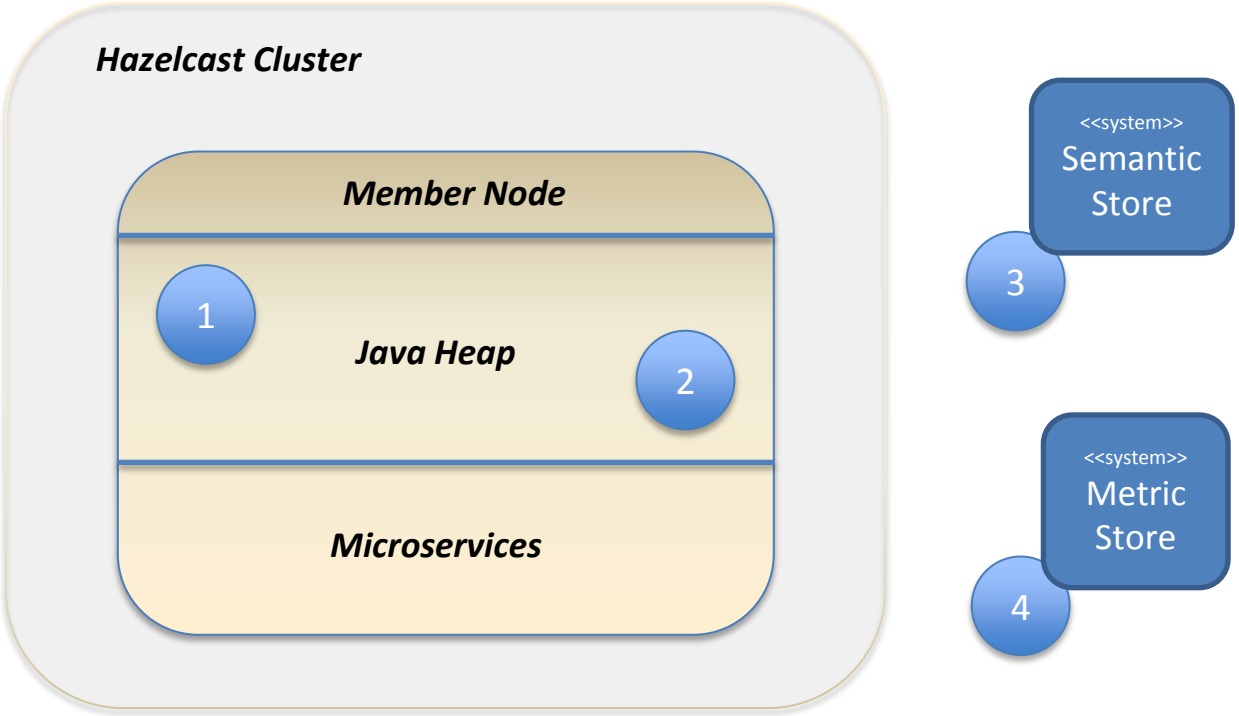
Provenance Cluster
Provenance Message Distribution

① Provenance message is stored in-memory in an **IMap** distributed Key/Val data structure where it can be queried.

② Provenance message is stored in-memory in an **IQueue** staged for persistence to registered stores (Producer – Consumer buffer)

③ Provenance message is saved to Semantic Store in their domain context

④ If provenance message contains Provenance Metrics, they are serialized to Metric Store

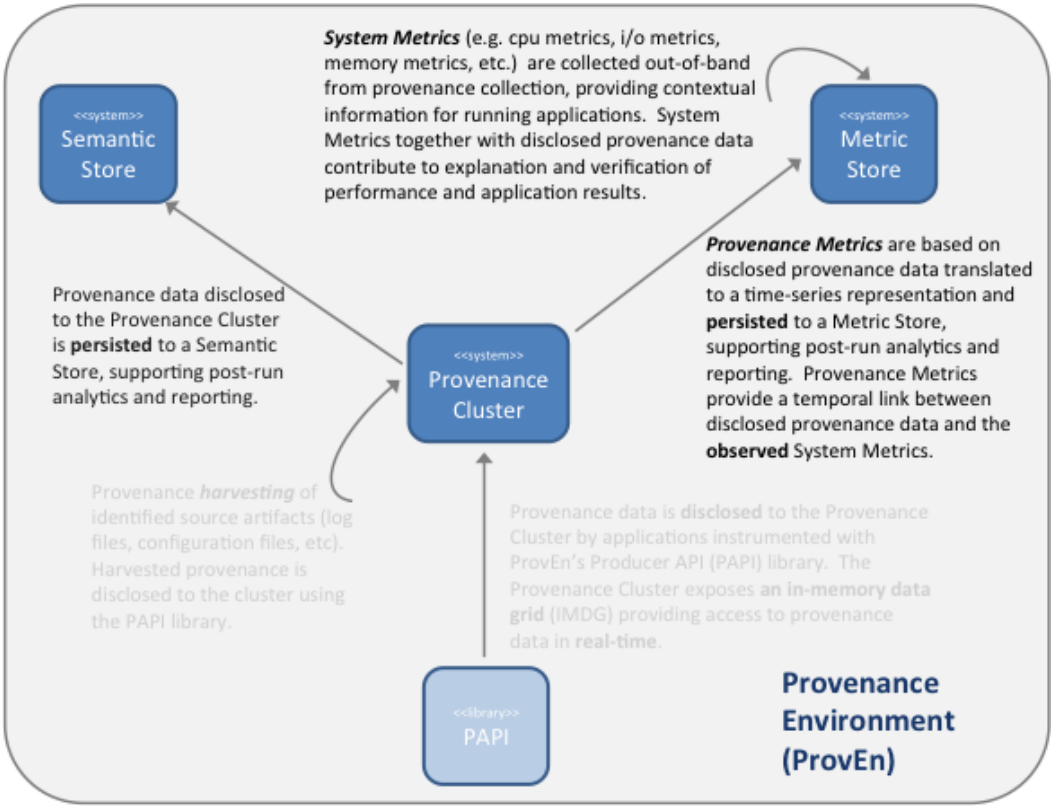


Discussion Outline

- Provenance Environment (ProvEn) Overview
- PAPI Library
- Provenance Cluster
- Hybrid Store
- Use Cases
- Next Steps

Hybrid Store

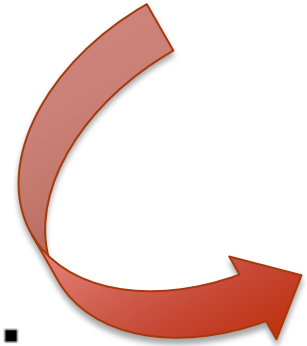
- **Hybrid store** is a fusion of semantic provenance with system level metrics, our *hybridization* approach; extending the provenance capture.
  - Semantic (*RDF Database*), Metric (*Time-Series Database*), and Provenance Cluster (IMDG) components contribute their data to the Hybrid Store.
  - Semantic and Metric stores are *registered* with the Provenance Cluster to support distribution and persistence of incoming provenance data.
  - Alignment of semantic and metrics data to form the Hybrid Store, is accomplished with what we are calling **Provenance Metrics**



Hybrid Store
What are Provenance Metrics?

- **Provenance Metrics** are discrete pieces of semantic provenance (a single triple) identified in a Provenance Message, and serialized into a time-series format for storage in a registered Metric Store.
- Occurs at time of disclosure, at a minimum alignment of data is by time

```
acme:simulation_1 wfpp:hasStartTime “1471355953002"^^xsd:dateTime
acme:simulation_1 wfpp:hasStopTime  “1471355959001"^^xsd:dataTime
```



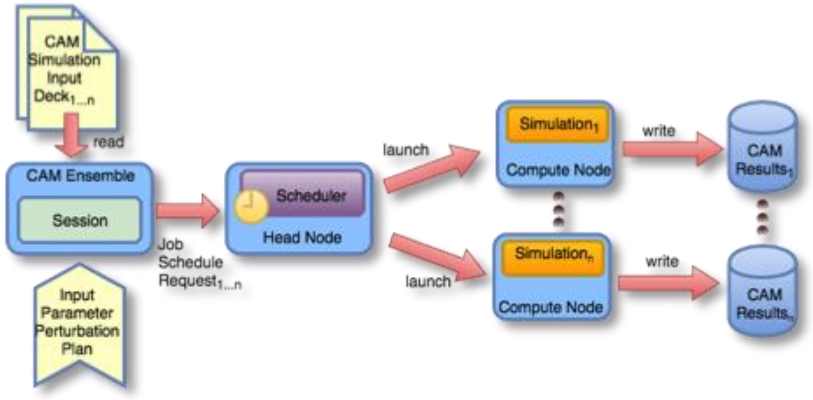
timestamp	node	sensor	value	state	message_id	app_id
1471355953002				**START**	**1**	**1**
1471355953004	pi06	CPU1	9.062			
1471355953004	pi06	MEM1	2.464			
1471355953004	pi06	CPU2	8.057			
1471355953004	pi06	MEM2	2.597			
…						
1471355959001				**STOP**	**100**	**1**
…						

Discussion Outline

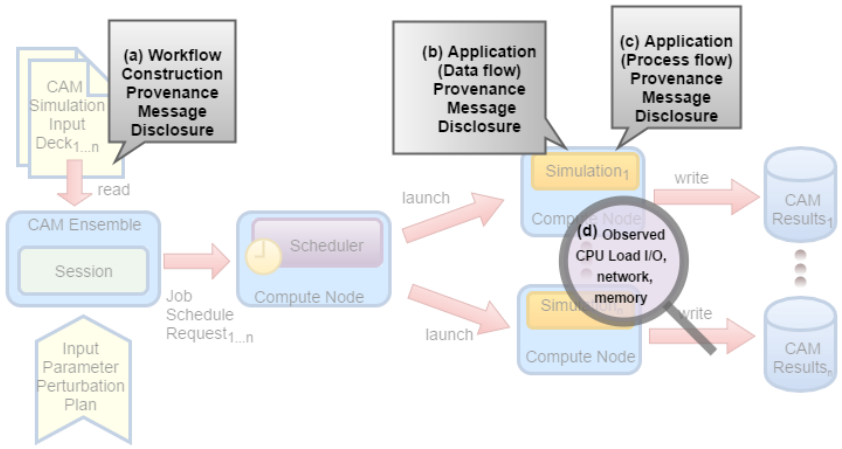
- Provenance Environment (ProvEn) Overview
- PAPI Library
- Provenance Cluster
- Hybrid Store
- Use Cases
- Next Steps

Accelerated Climate Modeling for Energy (ACME) – Ensemble Reproducibility Provenance Message Design

- Conceptual layout of ensemble workflow for a Community Atmosphere Model (CAM) sensitivity analysis study.



- Callout boxes represent provenance messages PAPI supports to produce provenance. Although not discussed in this use case, the magnifying glass (d) shows metrics that ProvEn can observe at the compute node during a simulation to extend the provenance descriptions.



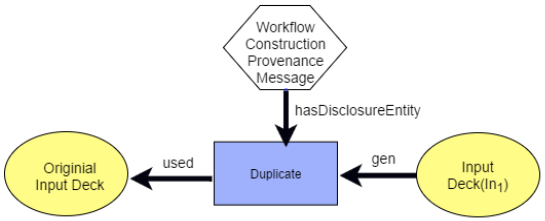
ACME – Example Experimental Setup

- **Simulation conditions**: One year coverage, first five days simulated of each month to test during different seasonal conditions.

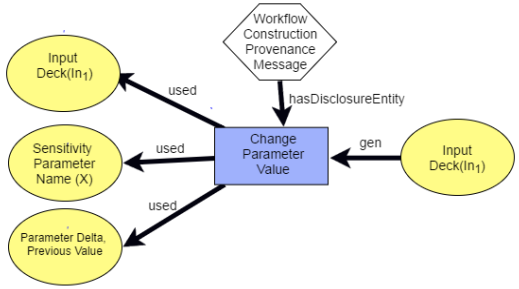
- **Experiment**: 12 parameters of 391 (land, sea, ice, atmosphere) parameters simultaneously perturbed.

- **Basic Reproducibility Questions**:
  - Which input deck for each ensemble simulation $\boldsymbol{S_1, S_2, S_3 \ldots S_n}$ corresponds to a given result?
  - Which parameter(s) was(were) perturbed in each simulation input deck $\boldsymbol{In_1}$-$\boldsymbol{I}_n$ for ensemble ***En***?

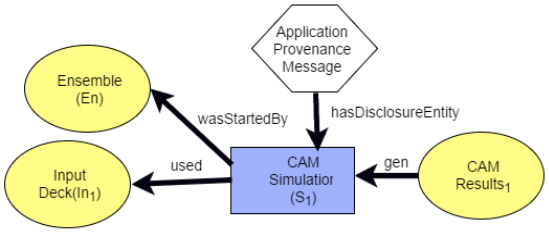
ACME - Message Disclosure and Collection



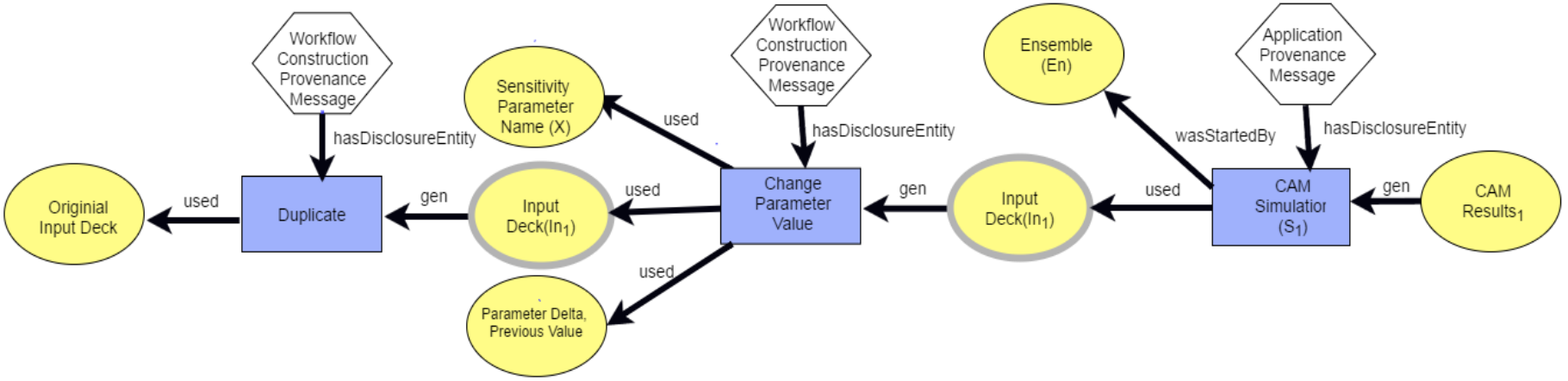
Input Deck Message



Perturbation Message



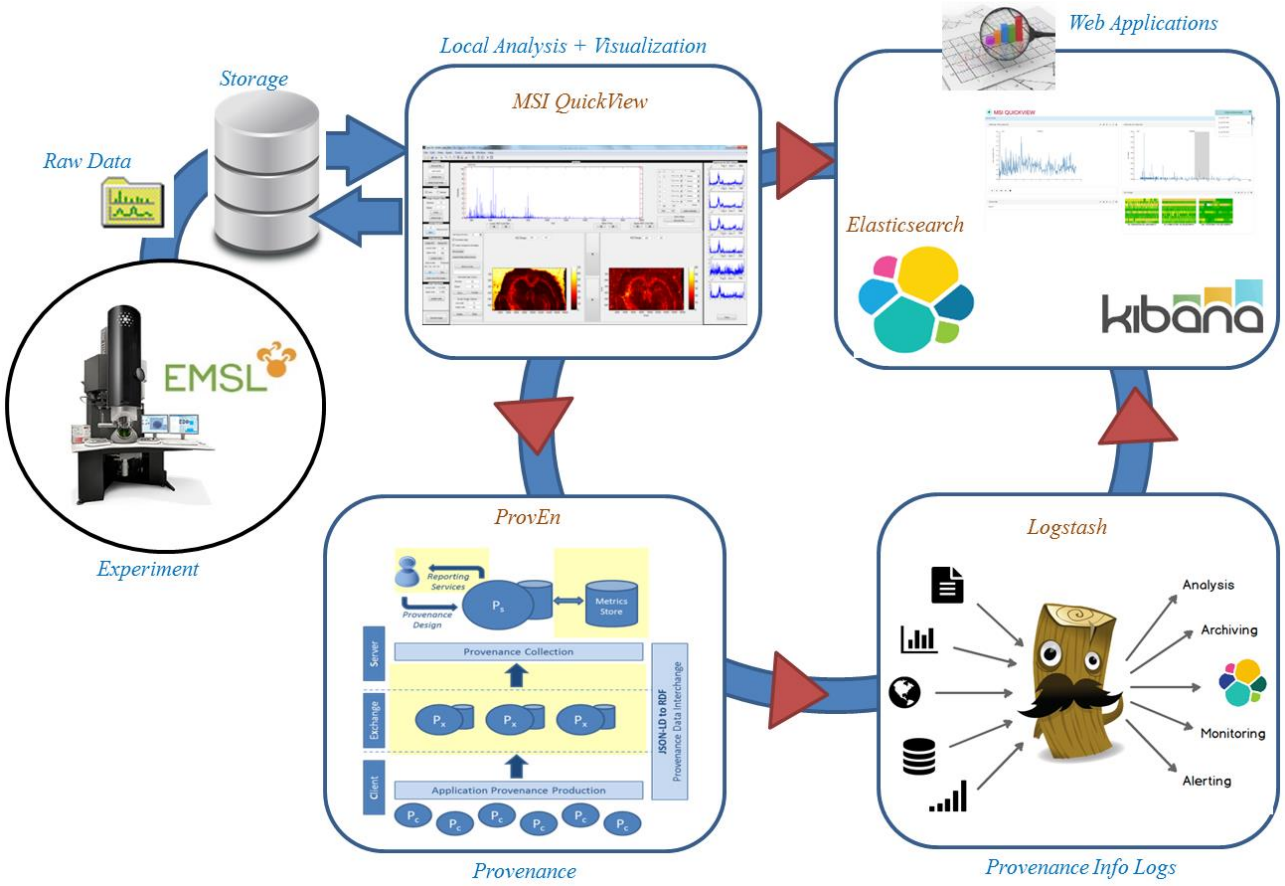
Results Message



When collected by ProvEn provenance message fragments are integrated into a connected provenance graph to answer the questions posed earlier. The gray outline on entity ovals indicates where messages are connected to form the complete provenance graph.

Enabling Reproducible Mass Spectrometry Data flows

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965



Thomas M, J Laskin, B Raju, EG Stephan, TO Elsethagen, NYS Van, and SN Nguyen. 2016. "Enabling Re-executable Workflows with Near-real-time Visualization, Provenance Capture and Advanced Querying for Mass Spectrometry Data." In *NYSDS 2016 - Data-Driven Discovery*. (accepted)

Seapearl

- **Seapearl** - a 52 node cluster, instrumented by the Penguin PowerInsight sensor system to provide out-of-band power and thermal monitoring.

- **Use Case - Anomaly Detection & Predictive Analytics**
  - ~100 million sensor measurements collected from Seapearl and saved to ProvEn's Metric Store, providing data for Predictive Analytic and Anomaly Detection experiments.
  - Detecting faulty sensors and using temperature readings to predict physical location of compute nodes in data center.

- **Use Case – Reproducibility**
  - CPU stress software instrumented with PAPI, messages included Provenance Metrics indicating START and STOP times for load and idle events.
  - Thermal and power metrics were captured out-of-band from stress software invocations.
  - Provenance metrics saved to Metric Store allowed us to properly sub-select time-series data for analysis and comparison to original experiment that was performed manually.

Discussion Outline

- Provenance Environment (ProvEn) Overview
- PAPI Library
- Provenance Cluster
- Hybrid Store
- Use Cases
- Next Steps

Next Steps

- Hybrid Store federated query capabilities
- Investigate improvements to message design
- Provenance Cluster monitoring and predictive analytics through microservice development
- Dynamic interference using our hybrid architecture

Acknowledgements

- Eric Stephan, Bibi Raju, Malachi Schram, Matt MacDuff, Darren Kerbyson - Pacific Northwest National Laboratory
- Kerstin Kleese van Dam - Brookhaven National Laboratory
- Ilkay Altintas, Alok Singh - San Diego Supercomputer Center & University of California, San Diego

- Project Acknowledgements
  - Integrated End-to-end Performance Prediction and Diagnosis for Extreme Scientific Workflows (IPPD) Project. IPPD is funded by the U. S. Department of Energy Awards FWP-66406 and DEC0012630
  - Accelerated Climate Modeling for Energy (ACME) project funded by the Office of Biological and Environmental Research (BER) in the U.S. Department of Energy (DOE) Office of Science.
  - Analysis In Motion (AIM) Initiative at Pacific Northwest National Laboratory (PNNL), which is conducted under PNNL's Laboratory Directed Research and Development Program