Event-Based Triggering and Management of Scientific Workflow Ensembles

Suraj Pandey[1], Karan Vahi[2], Rafael Ferreira da Silva[2], Ewa Deelman[2], Ming Jiang[3], Albert Chu[3], Cyrus Harrison[3], Henri Casanova[1]

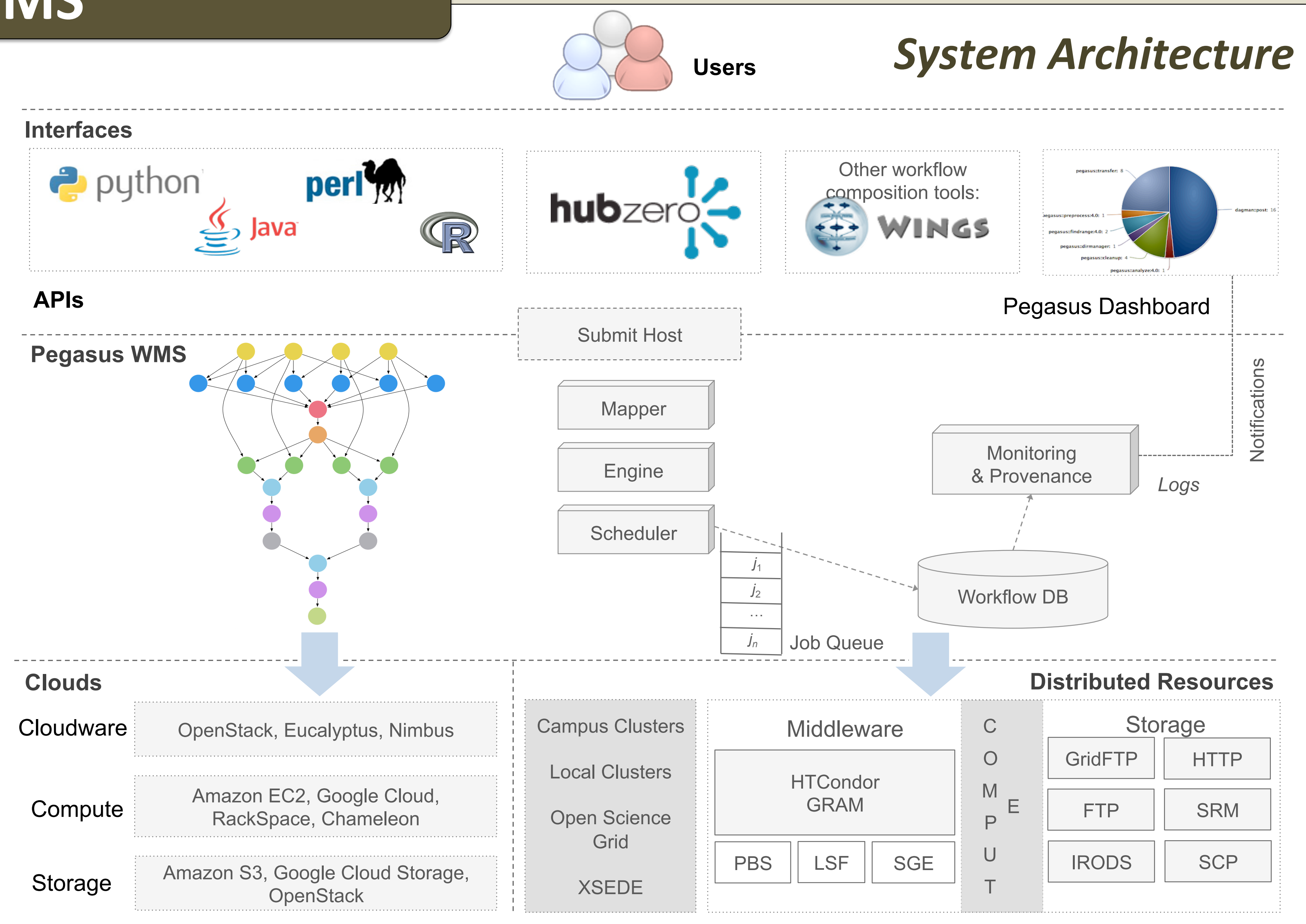
[1]*University of Hawaii,*

[2]*USC Information Sciences Institute,*

[3]*Lawrence Livermore National Laboratory*

Pegasus WMS

- Pegasus is a system for mapping and executing abstract application workflows over a range of execution environments.
- The same abstract workflow can, at different times, be mapped different execution environments such as XSEDE, OSG, commercial and academic clouds, campus grids, and clusters.
- Pegasus can easily scale both the size of the workflow, and the resources that the workflow is distributed over. Pegasus runs workflows ranging from just a few computational tasks up to 1 million.
- Workflows often consume tens of thousands of hours of computation and involve transfer of many terabytes of data.
- **Workflows have a DAG model**
  - **A node in the DAG is started only when all the parent nodes have successfully finished.**



System Architecture

Ensemble Manager

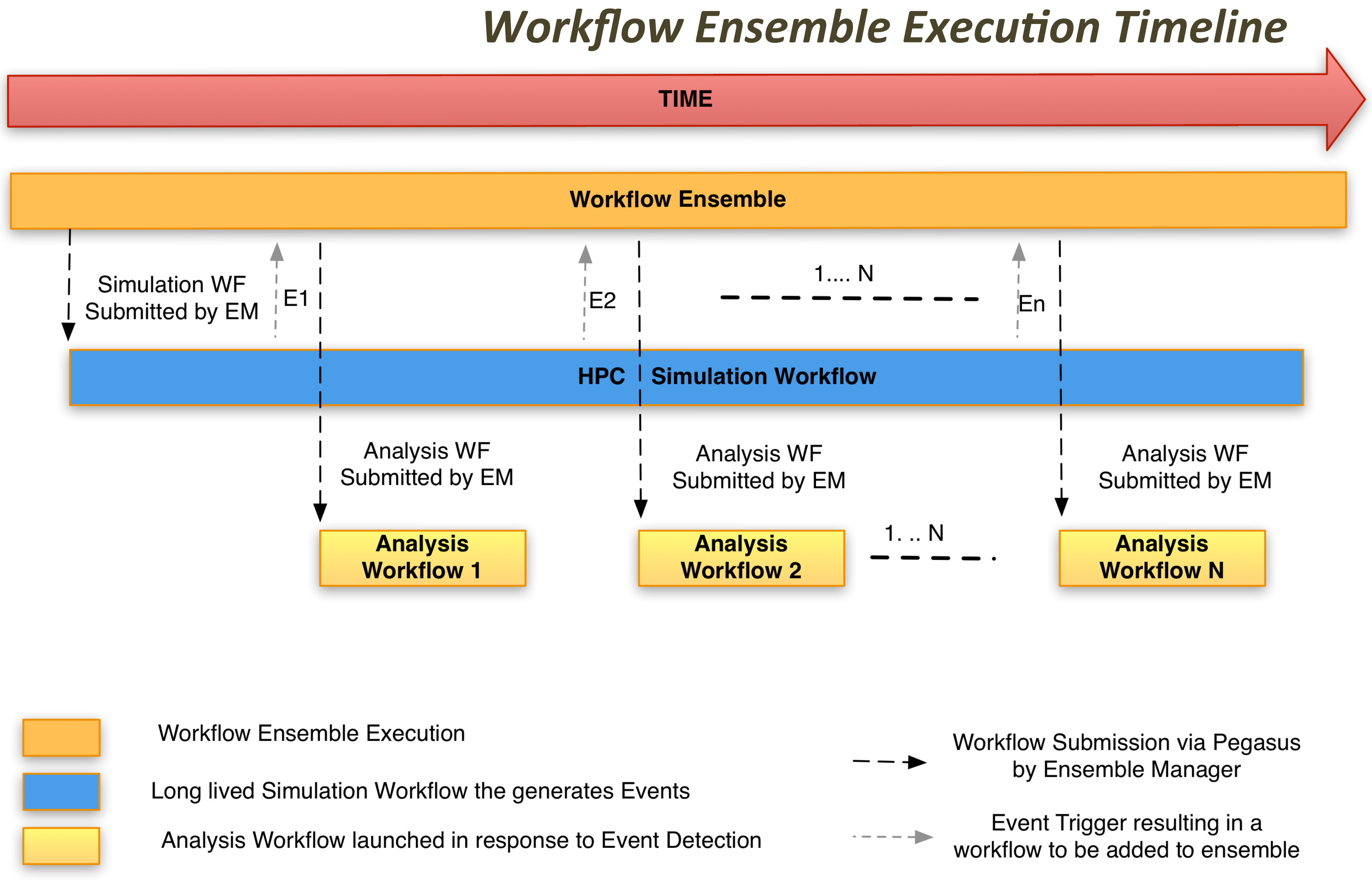
Problems mapping certain computations to DAG workflows

- With push towards extreme scale computing it is possible run traditional HPC simulation codes simultaneously on tens of thousands of cores.
- Generated data often needs to be periodically analyzed using Big Data Analytics frameworks.
- Integrating Big Data analytics with HPC simulations is a major challenge for the current generation of scientific workflow management systems.
- Need an ability to automatically spawn and manage the analysis workflows, as the long running simulation workflow executes.

Solution

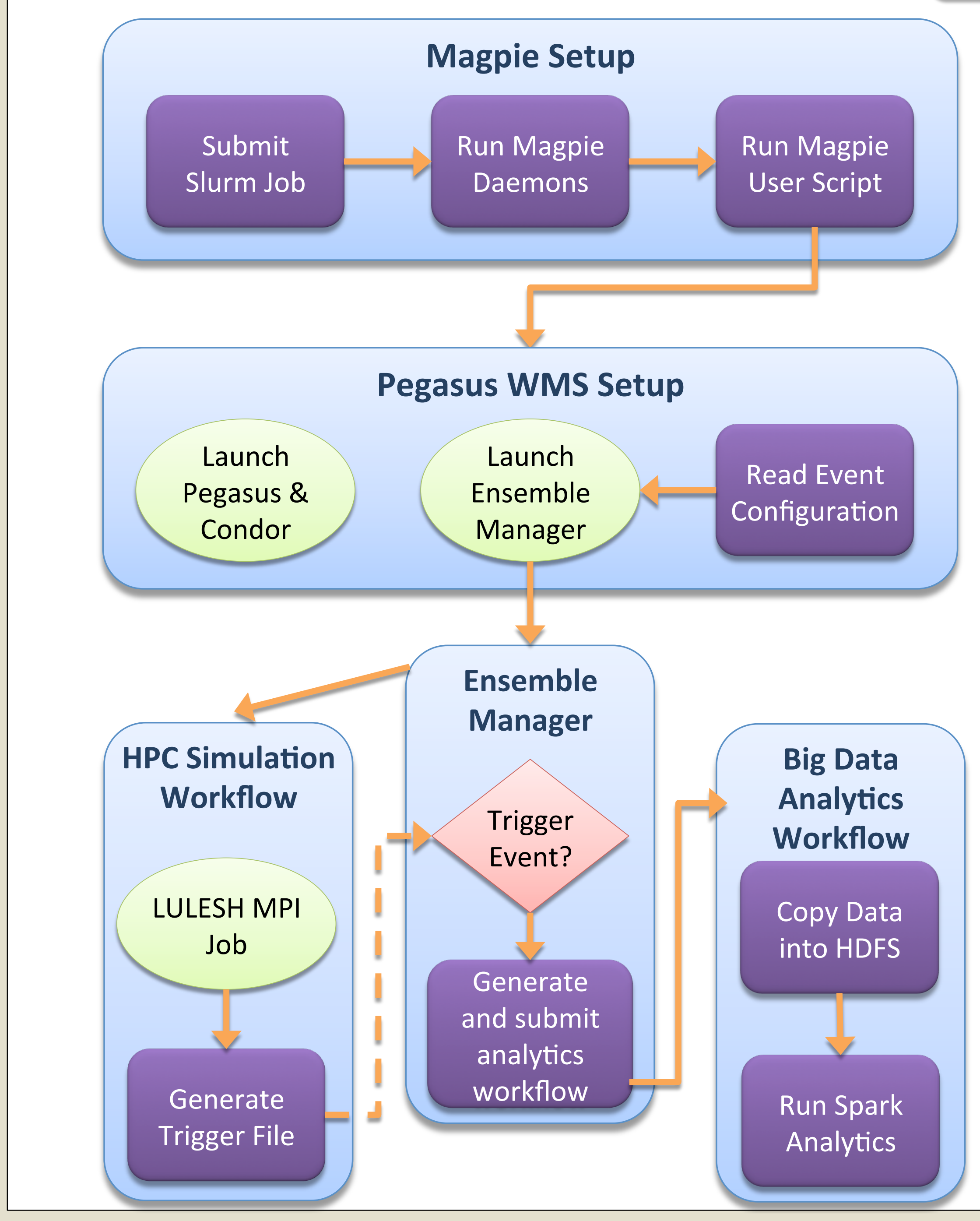
- Pegasus has an Ensemble Manager Service that allows user to submit a collection of workflows called ensembles.
- We extended the Ensemble Manager to support ***event triggers*** that can trigger addition of new workflows to an existing ensemble.
- We support the following types of triggers
  a) File based event triggers – a file gets modified
  b) Directory based event triggers – files appear in a directory
- Initially, ensemble has a single workflow consisting of the long running HPC simulation workflow.
- The HPC simulation workflow periodically generates output data that in a directory that is tracked by the ensemble manager.
- A new analysis workflow is launched automatically as the output data is detected.

Workflow Ensemble Execution Timeline



TIME
Workflow Ensemble
Simulation WF Submitted by EM
E1
E2
1.... N
En
HPC Simulation Workflow
Analysis WF Submitted by EM
Analysis WF Submitted by EM
Analysis WF Submitted by EM
Analysis Workflow 1
Analysis Workflow 2
1. .. N
Analysis Workflow N
Workflow Ensemble Execution
Long lived Simulation Workflow the generates Events
Analysis Workflow launched in response to Event Detection
Workflow Submission via Pegasus by Ensemble Manager
Event Trigger resulting in a workflow to be added to ensemble

Experimental Setup at LLNL Catalyst Cluster



- Reliably and repeatedly test that implemented solution works.
- Tested the implementation on LLNL Catalyst Cluster (150 teraFLOP/s system with 324 nodes, each with 128 GB of DRAM and 800 GB of non volatile memory) .

Experimental Setup

1. On catalyst, a Magpie SLURM job is submitted that does
   - Determines which nodes will be “master” nodes, “slave”nodes, or other types of nodes.
   - Sets up, configures, and starts appropriate Big Data daemons to run on the allocated nodes. In our setup , we used the Magpie SPARK template to setup a dynamic Spark cluster
   - Reasonably optimizes configuration for the given cluster hardware that it is being run on. Magpie then executes a user specified script to give control back to the user.
2. The user script sets up Pegasus WMS and starts the Ensemble Manager.
3. Ensemble Manager submits the HPC Simulation Workflow consisting of LULESH application
   - Every 10 simulation cycles LULESH writes out outputs to a directory on the shared filesystem.
   - This directory is tracked by the ensemble manager as part of the event trigger specified.
   - Ensemble Manager invokes a script that generates the Big Data Analytics Workflow on the newly generated datasets.

https://github.com/pegasus-isi/pegasus-llnl

Project is supported by the Department of Energy under contract DE-AC52-07NA27344 and by LLNL under project 16-ERD-036.

Pegasus is currently funded by the National Science Foundation(NSF) under the OAC SI2-SSI grant 1664162 .