Pegasus Users Group

MEETING

USC

Pegasus is funded by the National Science Foundation under grant #1664162

Welcome!

Welcome to the Workshop

Ewa Deelman

deelman@isi.edu

February 23, 2021

Inaugural PUG Meeting

- **Goals:**
  - Hear from our **users**, what are your **needs**, how can we **improve**?
  - **Share information** about different usages of Pegasus across sciences
  - Understand **new applications, new types of workflows: ML4Science**
  - Tell you about **Pegasus 5.0** and what it can do

  - **Mark 20 years of Pegasus!**

How Did Pegasus Start?

Extend the concept of view materialization in DBs to distributed environments

NSF ITR: GriPhyN Project: Ian Foster (PI), Paul Avery, Carl Kesselman, Miron Livny, (co-Pis)

The Virtual Data Grid (VDG) Model

- Data suppliers publish data to the Grid
- Users request raw or derived data from Grid, without needing to know
  - Where data is located
  - Whether data is stored or computed

How do you translate the Computer Science idea to the needs of science?

Circa. 2001

Virtual Data Scenario

- (LIGO) "Conduct a pulsar search on the data collected from Oct 16 2000 to Jan 1 2001"
- For each requested data value, need to
  - Understand the request
  - Determine if it is instantiated; if so, where; if not, how to compute it
  - Plan data movements and computations required to obtain all results
  - Execute this plan

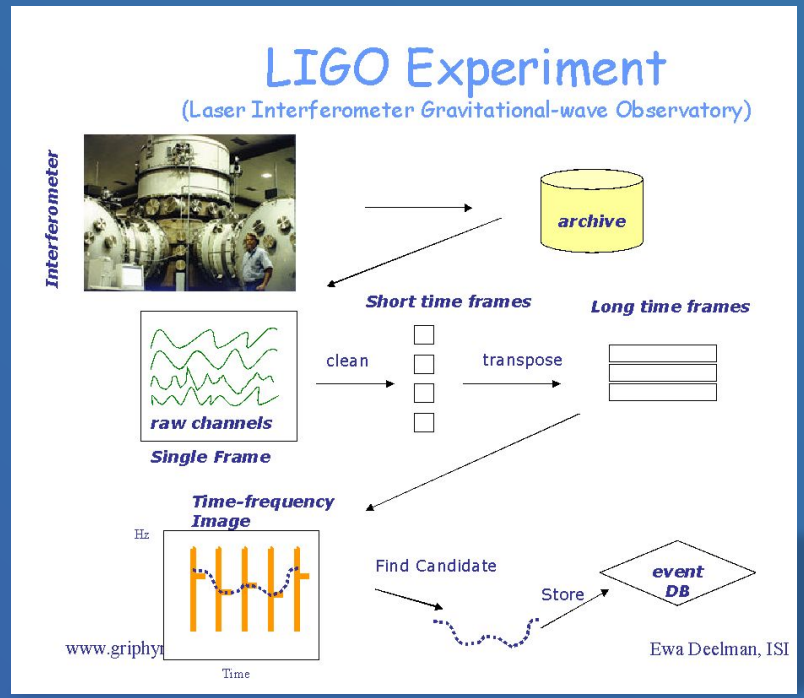
www.griphyn.org

Ewa Deelman, ISI

Challenge: How Translate a Science Request to an Actionable Plan?

Explore AI planning techniques

Work with Yolanda Gil and Jim Blythe



Lost in translation: high-level abstraction for this science domain
Found: new research direction: management of workflows in distributed environments

Challenges of Workflow Management

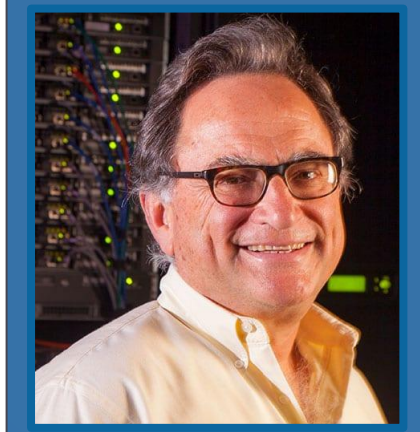
- **Working with applications (astronomy, earthquake science), gravitational-wave physics) found common challenges:**
  - Need to describe complex workflows in a simple way
  - Need to access distributed, heterogeneous data and resources
  - Need to deal with resources/software that change over time
- **Our focus:**
  - Separation between workflow description and workflow execution
  - Workflow planning and scheduling (scalability, performance)
  - Task execution (monitoring, fault tolerance, debugging)

Started a collaboration with Miron Livny and his HTCondor team

HTCondor
High Throughput Computing

Duncan Brown
Syracuse University

Scott Callaghan
USC

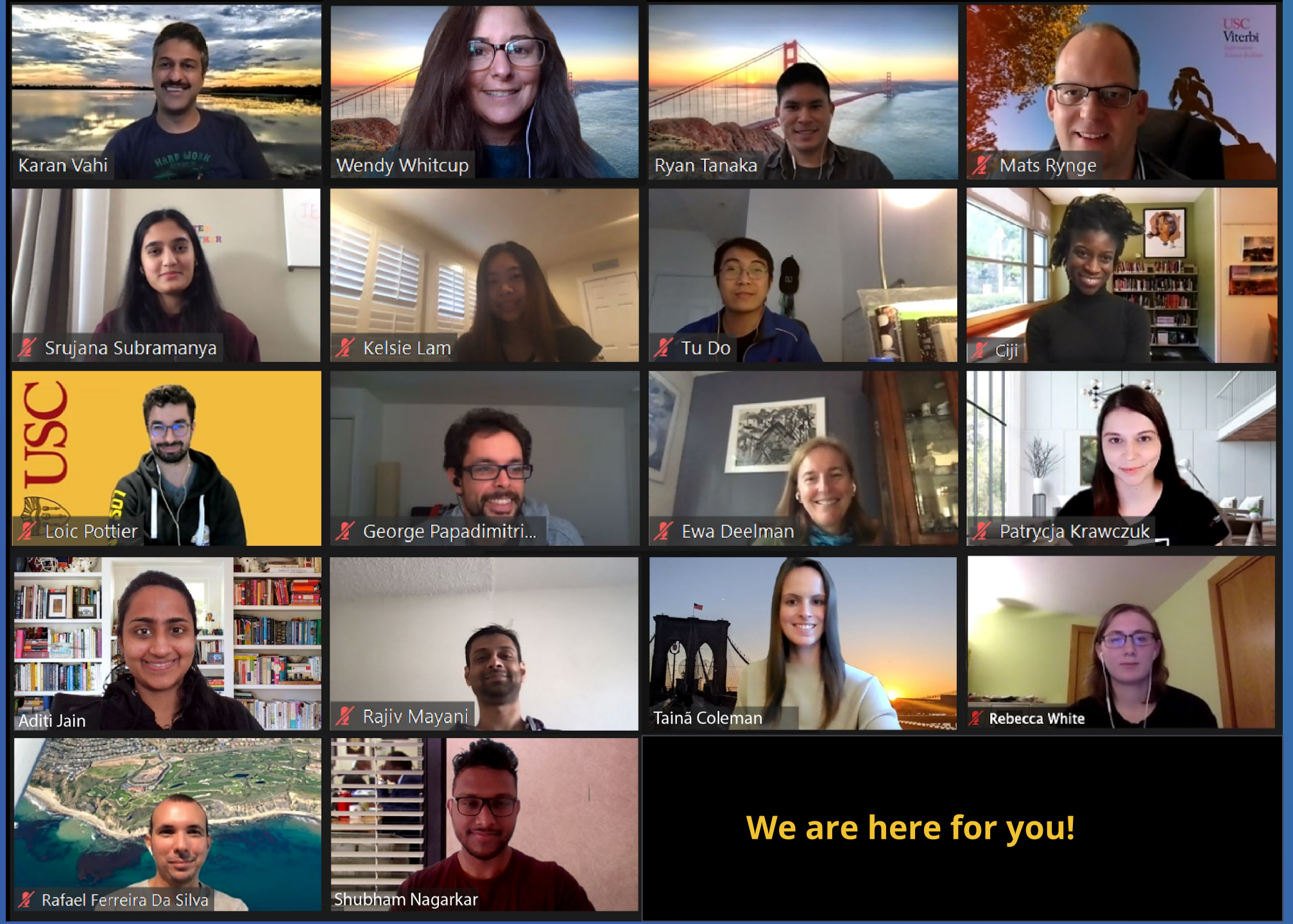


Miron Livny,UW Madison

Bruce Berriman, Caltech

Phil Maechling
USC

PEGASUS TEAM



PUG 2021

NSF

Thank You!