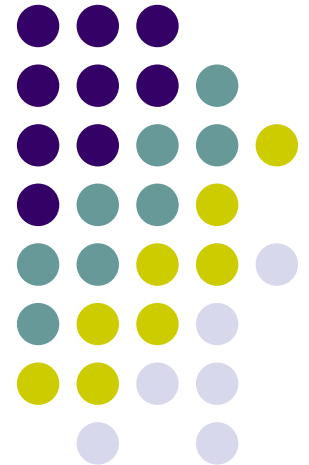


# WFM-RG

# Workflow Management

---

Ewa Deelman

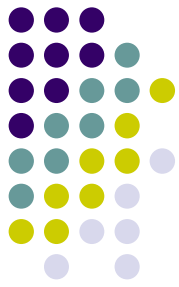


# Ongoing Workflow Management Work

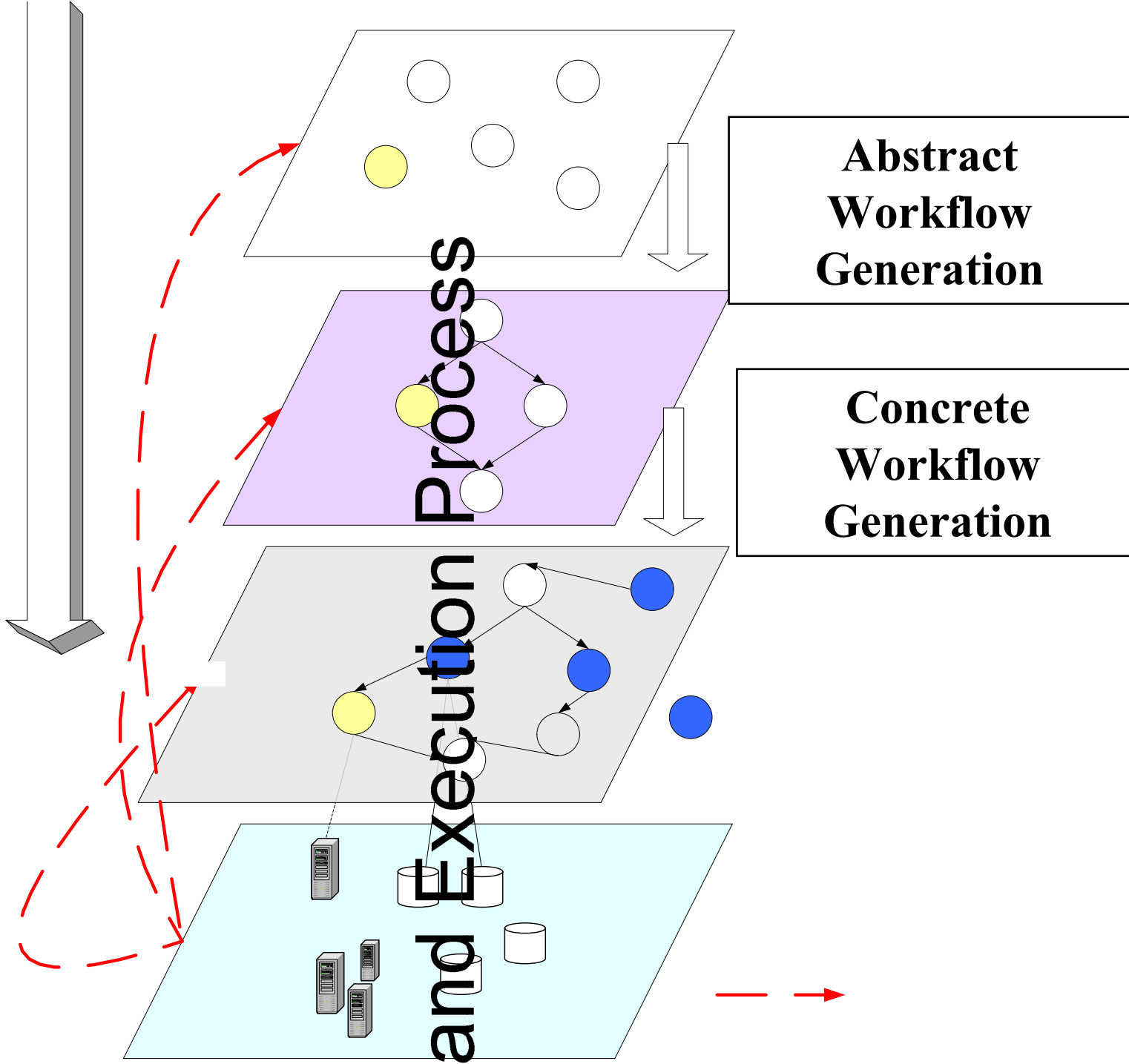


- Part of the NSF-funded GriPhyN project
- Supports the concept of Virtual Data, where data is materialized on demand
- Data can exist on some data resource and be directly accessible
- Data can exist only in a form of a recipe
- The GriPhyN Virtual Data System can seamlessly deliver the data to the user or application regardless of the form in which the data exists
- GriPhyN targets applications in high-energy physics, gravitational-wave physics and astronomy

# Workflow management in GriPhyN

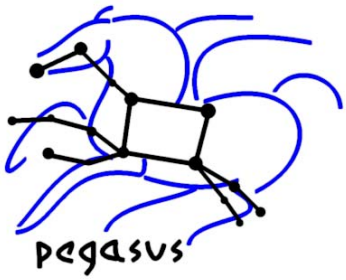


- Workflow Generation: how do you describe the workflow (at various levels of abstraction)? (Chimera)
- Workflow Mapping/Refinement: how do you map an abstract workflow representation to an executable form? (Pegasus)
- Workflow Execution: how to you reliably execute the workflow? (Condor's DAGMan)

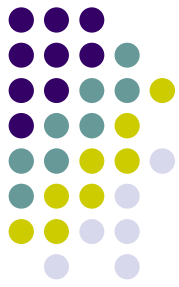


FFT

Applica

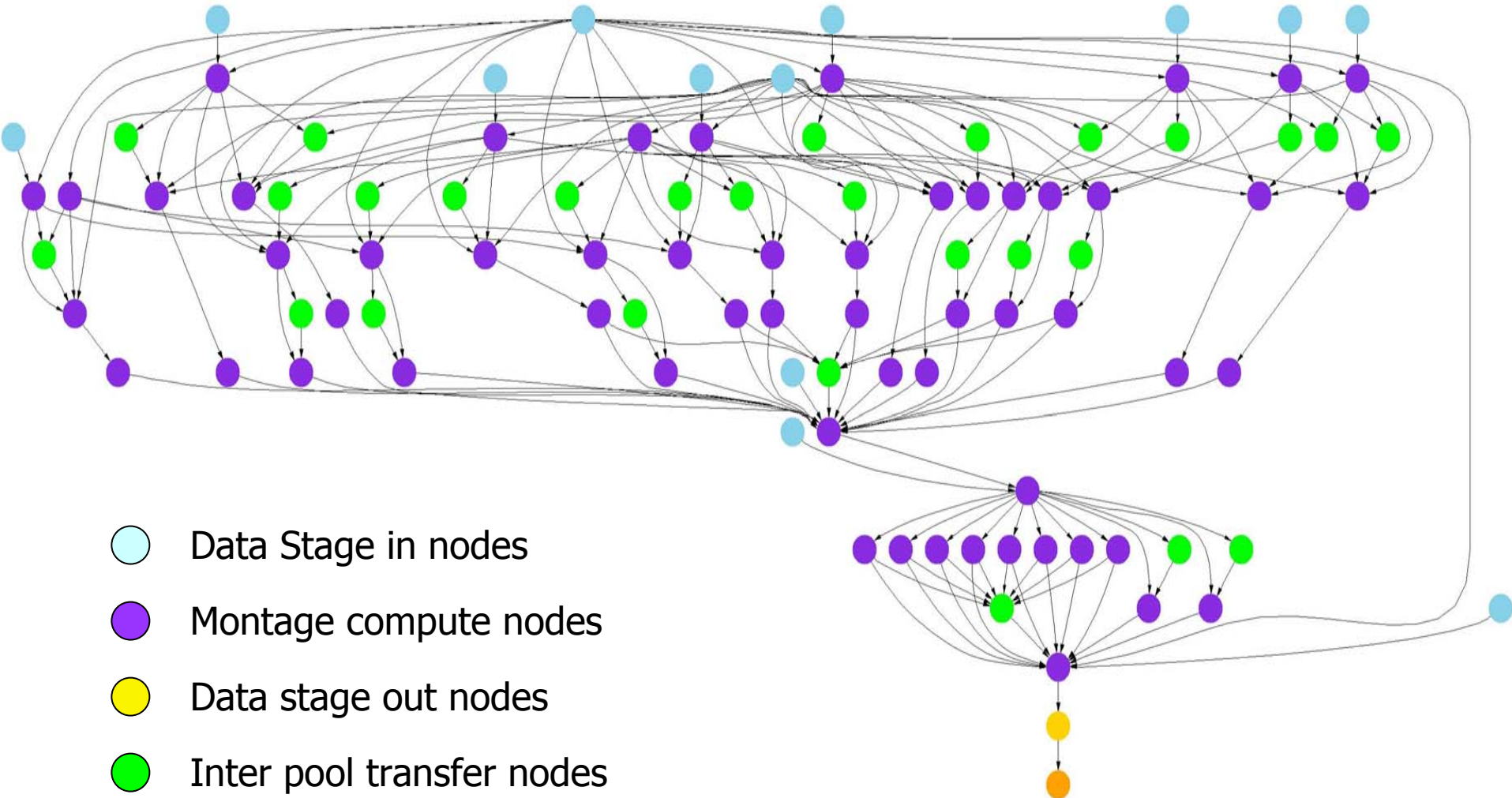
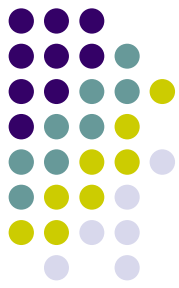


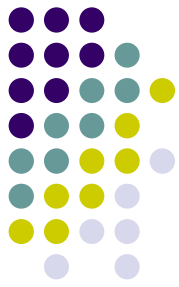
# Pegasus, Planning for Execution in Grids



- Performs the mapping from an abstract workflow to a concrete workflow, which can be executed on the Grid
- Isolates the user from many Grid details
- Automatically locates physical locations for both components (transformations) and data (RLS, TC)
- Finds appropriate execution resources (MDS)
- Whenever several alternatives are possible (e.g., alternative physical files, alternative resources) it makes a random choice, or uses simple heuristics
- Stages data in and out of computations, publishes newly derived data products
- Reuses existing data products where applicable

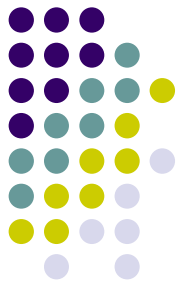
# Montage-an astronomy workflow





# Workflows

- Composition of applications or services
- Difference between business workflows and Grid workflows
  - Data being transferred between workflow activities      **Dataflow?**
  - The nature of the execution environment: very dynamic, prone to faults



# Workflow Evolution

- Workflow description
  - Metadata
  - Partial, abstract description
  - Full, abstract description
  - A concrete, executable workflow
- Workflow refinement
  - Take a description and produce an executable workflow
- Workflow execution



# Workflow Refinement

- The workflow can undergo an arbitrarily complex set of refinements
- A refiner can modify part of the workflow or the entire workflow
- A refiner uses a set of Grid information services and catalogs to perform the refinement (metadata catalog, virtual data catalog, replica location services, monitoring and discovery services, etc. )

# Workflow Refinement and execution

