Scientific workflow management a way to enable e-science on both Grids and Clouds

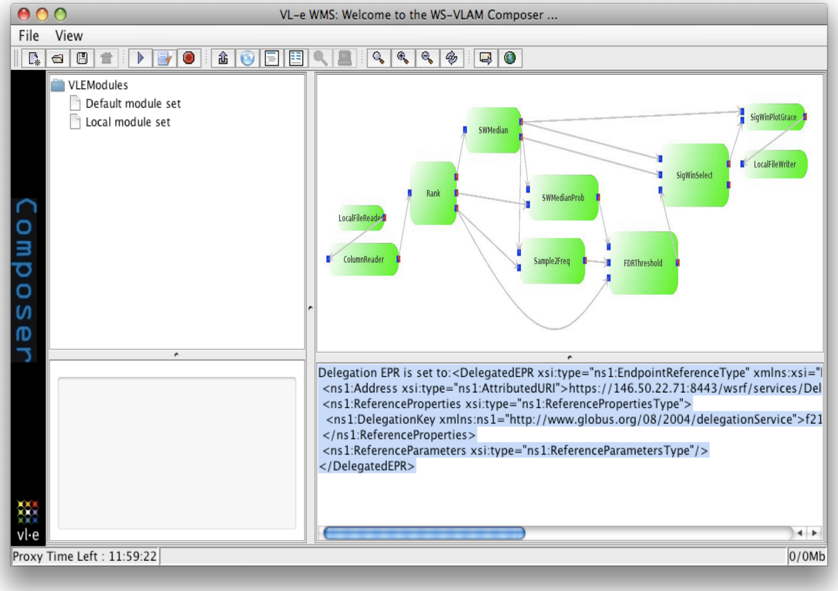
Adam Belloum
Institute of Informatics
University of Amsterdam
a.s.z.belloum@uva.nl

Outline

- Introduction
- Lifecycle of an e-science workflow
- Workflow management Systems
- Scientific workflows Applications
- Provenance
- Examples of Scientific workflow managements

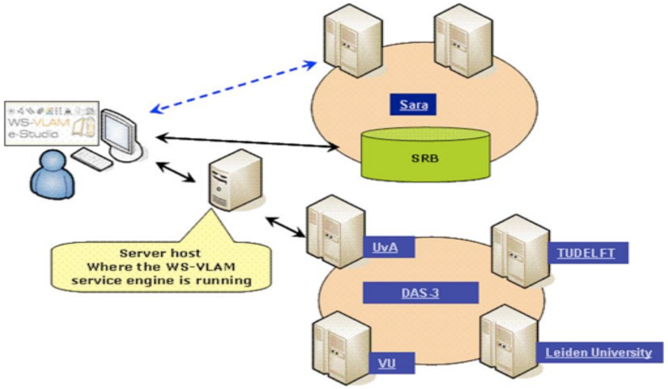
Workflow management system

- **Workflow management system** is a computer program that manages the execution of a workflow on a set of computing resources.



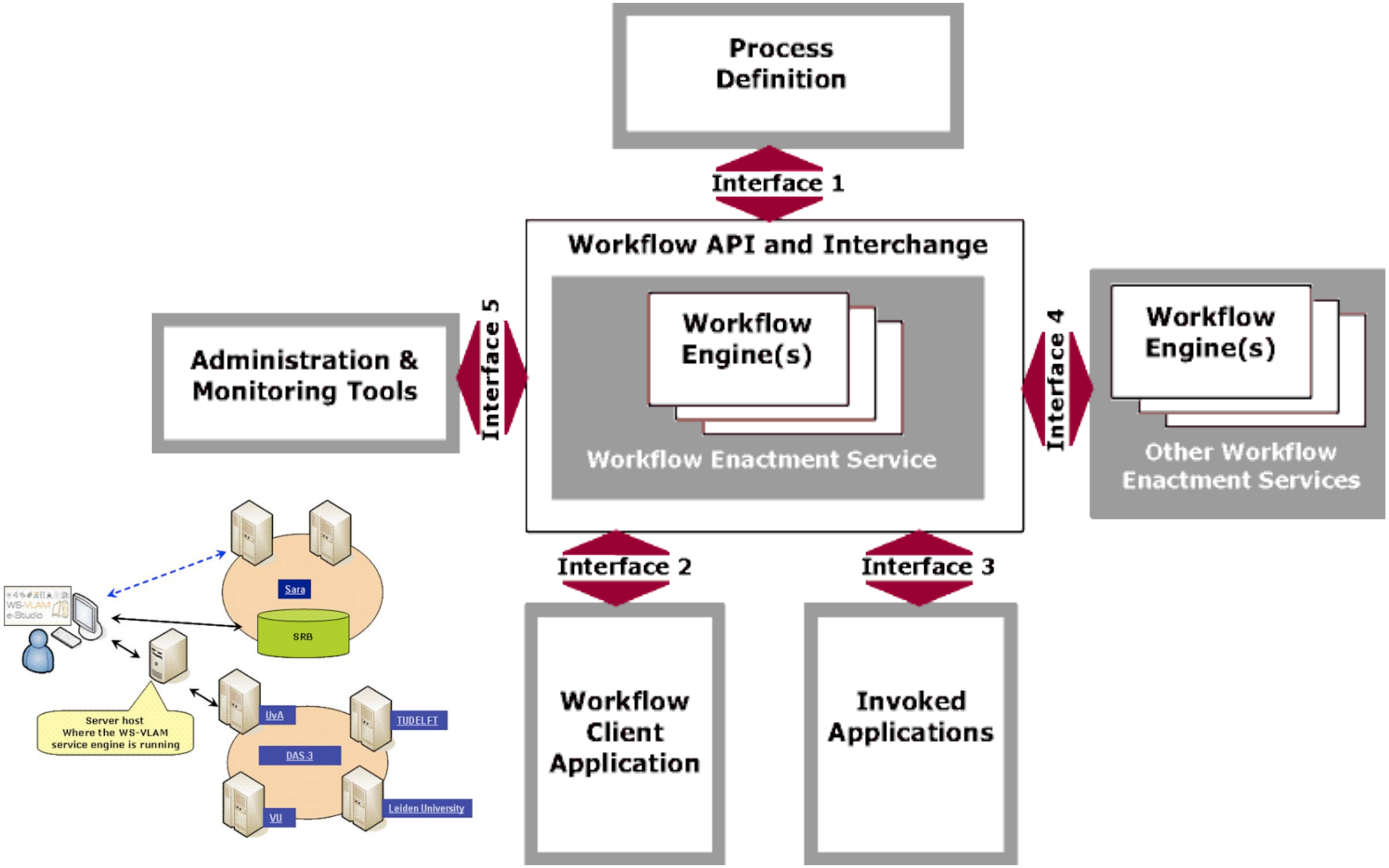
```
Delegation EPR is set to:<DelegatedEPR xsi:type="ns1:EndpointReferenceType" xmlns:xsi="
 <ns1:Address xsi:type="ns1:AttributedURI">https://146.50.22.71:8443/wsrf/services/Del
 <ns1:ReferenceProperties xsi:type="ns1:ReferencePropertiesType">
  <ns1:DelegationKey xmlns:ns1="http://www.globus.org/08/2004/delegationService">f21
 </ns1:ReferenceProperties>
 <ns1:ReferenceParameters xsi:type="ns1:ReferenceParametersType"/>
</DelegatedEPR>
```

The user interface of the WS-VLAM a workflow management system developed in the VL-e project to execute application workflow on geographically distributed computing resources



Deployed as service on Dutch super Computer (DAS3), and Dutch NGI (BigGrid) Clusters

Reference Model From WFMC


Process Definition
Interface 1
Workflow API and Interchange
Workflow Engine(s)
Workflow Enactment Service
Interface 5
Administration & Monitoring Tools
Interface 4
Workflow Engine(s)
Other Workflow Enactment Services
Interface 2
Interface 3
Workflow Client Application
Invoked Applications
WS-VLAM e-Studio
Sara
SRB
Server host Where the WS-VLAM service engine is running
UvA
TUDELFT
DAS-3
VU
Leiden University

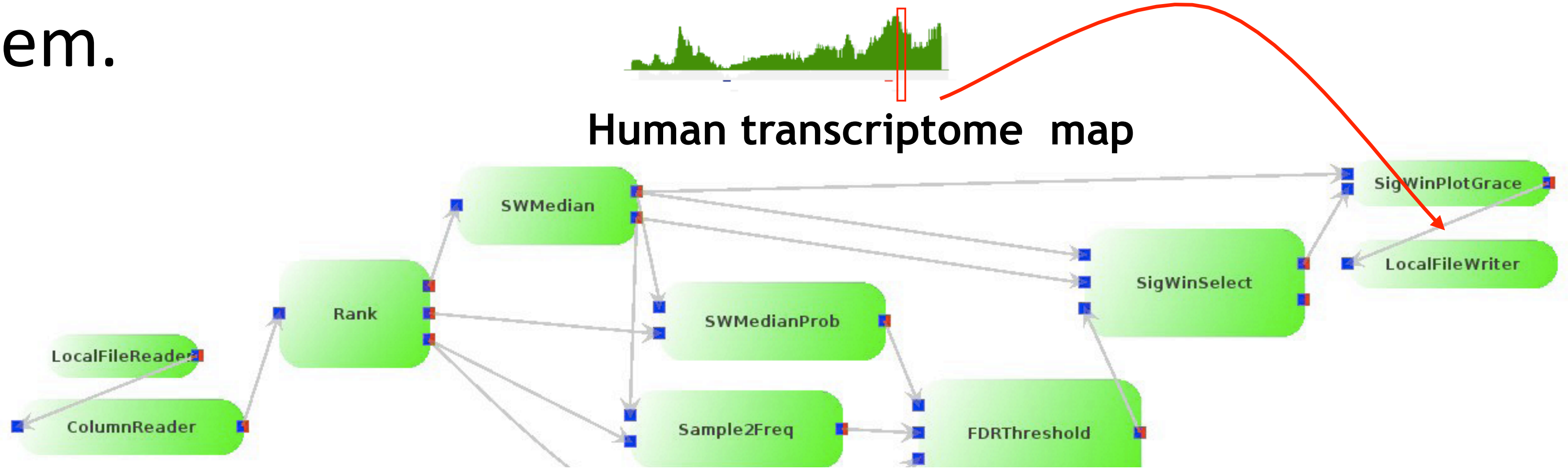
*The automation of a **business process**, in whole or parts, where documents, information or tasks are passed from one participant to another to be processed, according to a set of **procedural rules**. (WFMC definition of a Workflow)*

Challenges of running workflows on e-infrastructure (grids and clouds)

- ***co-allocate*** *resources needed for workflow enactment across multiple domains*?
- *achieve* ***QoS*** *for data centric application workflows that have special requirements on network connections*?
- *achieve* ***Robustness*** *and fault tolerance for workflow running across distributed resources*?
- *increase* ***re-usability*** *of Workflow, workflow components, and refine workflow execution*?

Workflow

A workflow is a model to represent a **reliably repeatable sequence** of operations/tasks by showing explicitly the interdependencies among them.

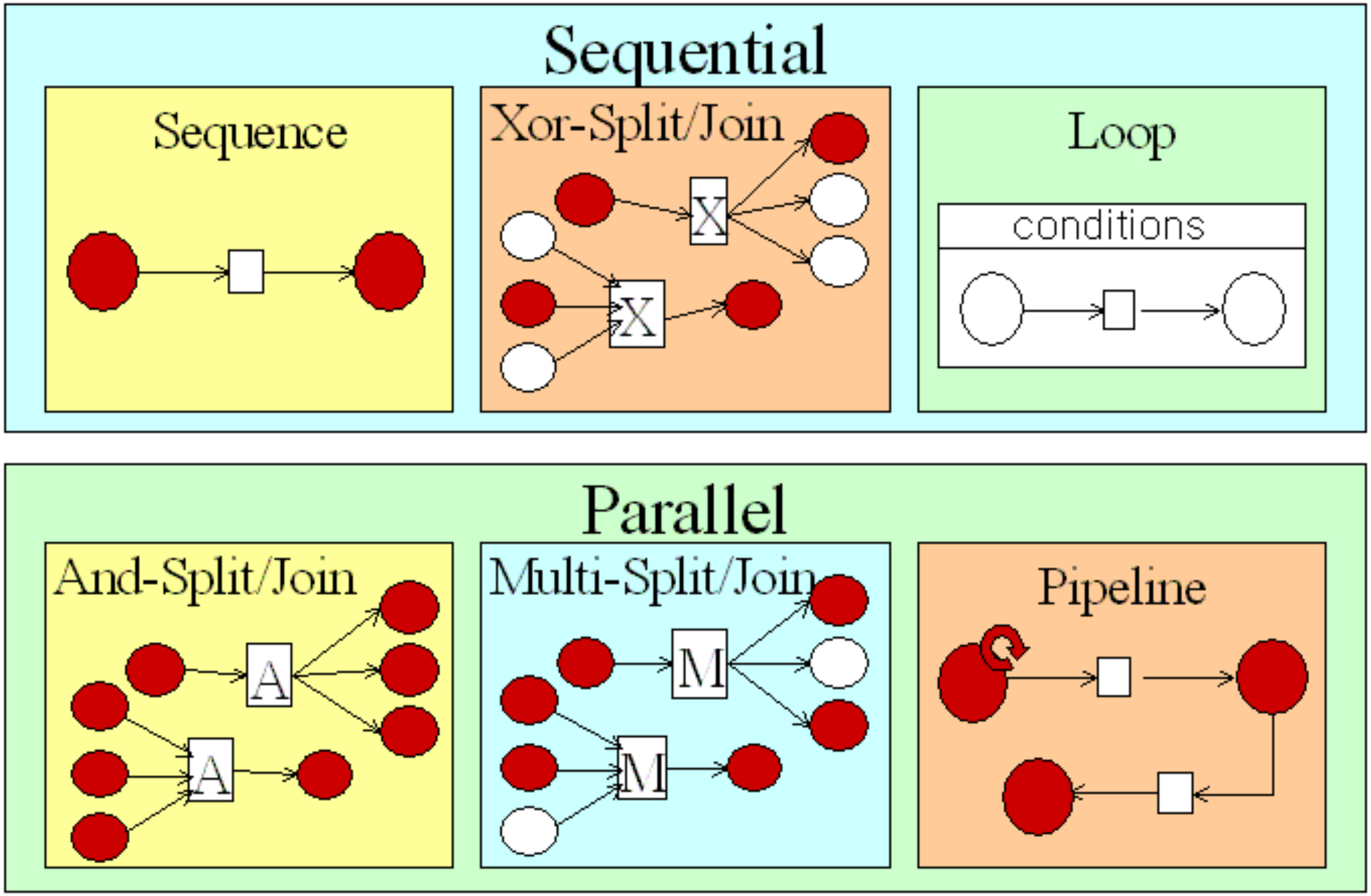


Human transcriptome map

http://www.youtube.com/watch?v=R6bTFrzaR_w&feature=player_embedded

SigWin-Detector workflow has been developed in the VL-e project to detect ridges in for instance a Gene Expression sequence or Human transcriptome map, BMC Research Notes 2008, 1:63 doi: 10.1186/1756-0500-1-63.

Workflow Pattern



Business vs Scientific Workflows (Similarities)

- **Capturing knowledge/best practices**
  - Capture **business process** based on the company policy
  - Capture **best practices of scientist**, expert from a specific domain
- **Series of structured activities and computations**
  - Both involves **repeated execution** of certain procedures, and both describes tasks within this procedures.
- **Incorporate human decision in the process**
  - There are exceptional cases that can not be automated both in business and scientific workflow

http://www.csc.ncsu.edu/faculty/mpsingh/papers/databases/workflows/sciworkflows.html

Business vs Scientific Workflows (Differences)

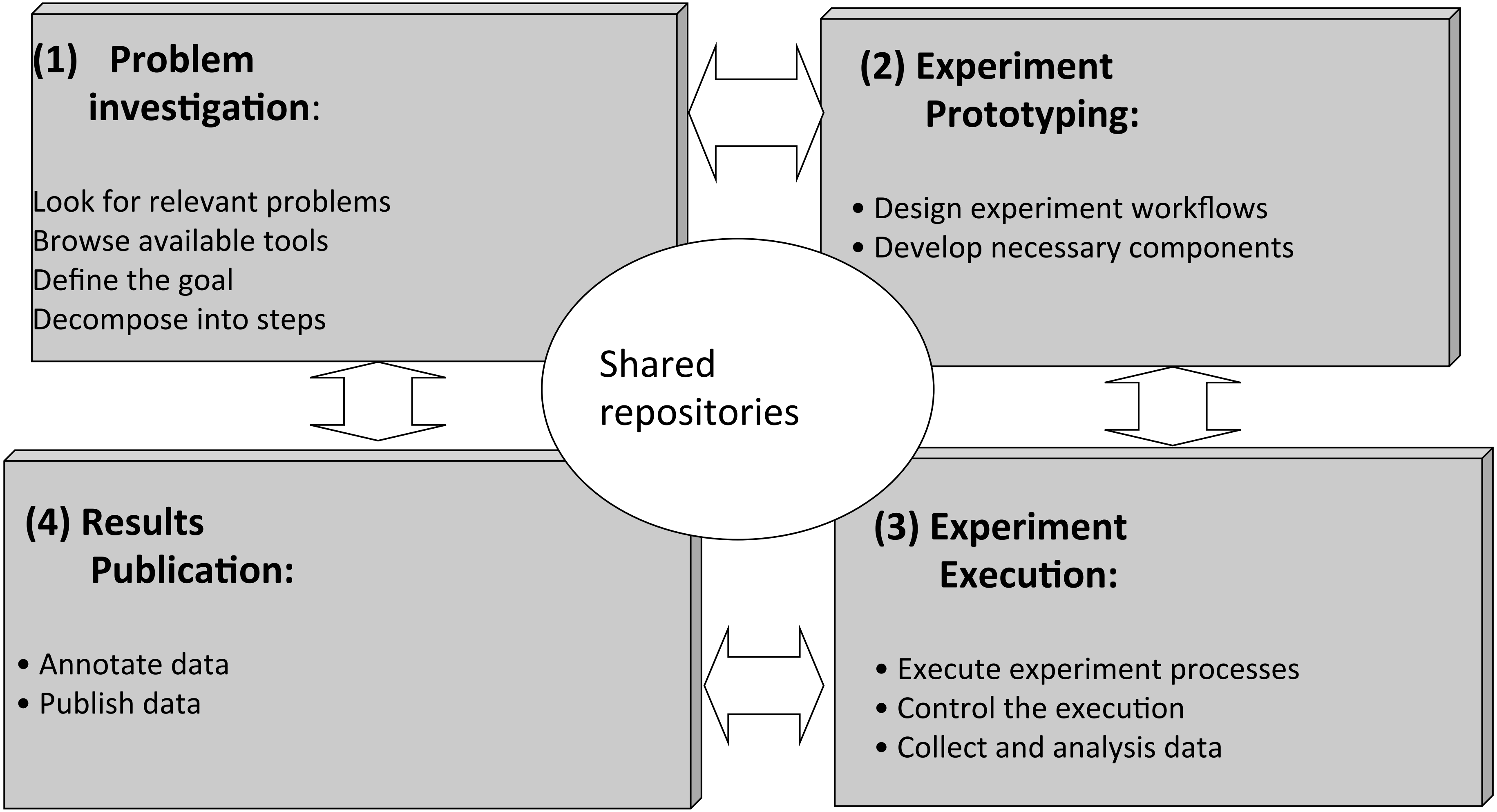
- **Business Process**
  - **Information**, task, procedural rules of a certain company
  - Driven by business profit goals
- **Static Procedures**
  - Reflecting certain policy within a company
  - Rigid, any changes require approval from management
- **Closed Environment**
  - Managed own resources
  - Within company, actual organization
- **Documents, task descriptions**
  - Flight reservation, credit approval, supply chain, billing, resource planning

- **Scientific process**
  - **Data** analysis, experiment, data manipulation recipes
  - Driven by problem solving goal
- **Dynamic**
  - Exploratory and speculative
  - Flexible, scientist manage their own business (they are their own user/ manager).
- **Open Environment**
  - Non Centralized grid environment
  - Across boundary, Virtual Organizations
- **Large Data**:
  - High energy physics data, bioinformatics micro array/ genomic data etc.

Scientific Workflow Specific Needs

- What makes scientific workflow different?
  - Need for **large** data flows support
  - Need to do **parameterized** execution of large number of jobs
  - Need to monitor and control workflow execution including **ad-hoc** changes
  - Need to execute in **dynamic** environment where resources are not known a priori and may need to adapt to changes
  - Hierarchical execution with sub-workflows created and destroyed when necessary

Complex Scientific experiments model



(1) Problem investigation:

Look for relevant problems
Browse available tools
Define the goal
Decompose into steps

(2) Experiment Prototyping:

- Design experiment workflows
- Develop necessary components

Shared repositories

(4) Results Publication:

- Annotate data
- Publish data

(3) Experiment Execution:

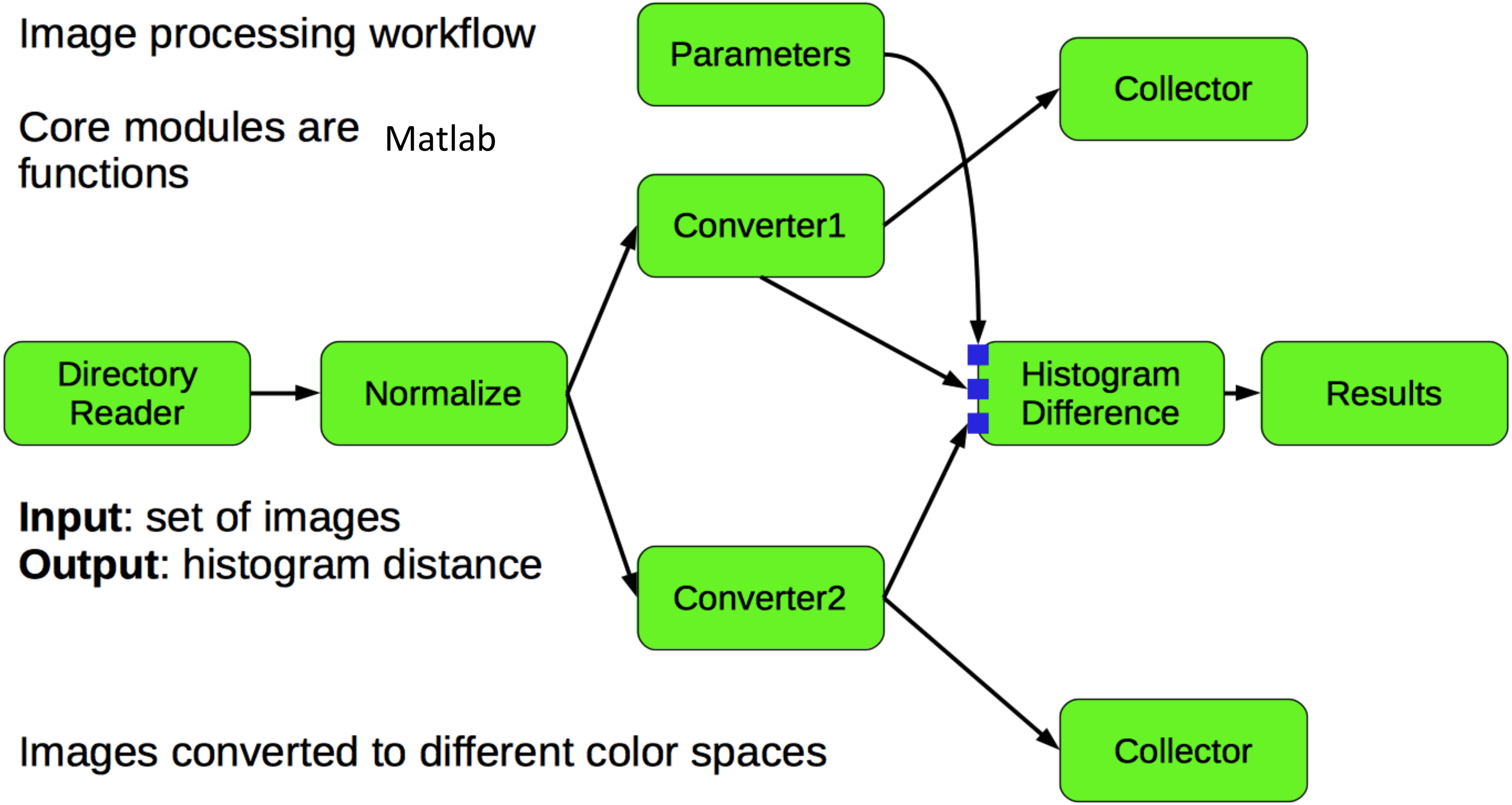
- Execute experiment processes
- Control the execution
- Collect and analysis data

Collaborative e-Science experiments: from scientific workflow to knowledge sharing *A.S.Z. Belloum, Vladimir Korkhov, Spiros Koulouzis, Marcia A Inda, and Marian Bubak JULY/AUGUST, IEEE Internet Computing, 2011*

Example of Scientific workflow

Image processing workflow

Core modules are Matlab functions



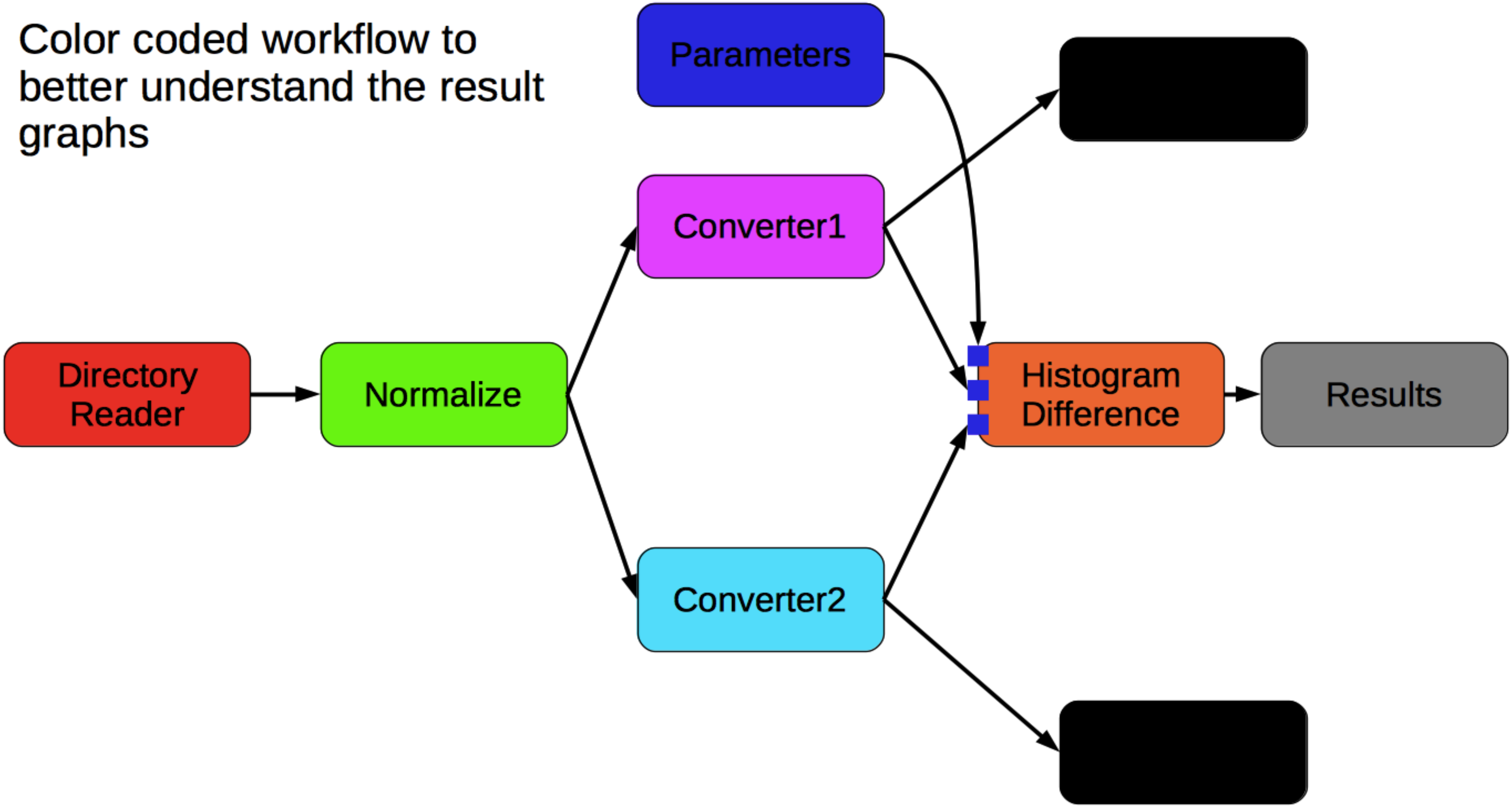
Input: set of images
Output: histogram distance

Images converted to different color spaces

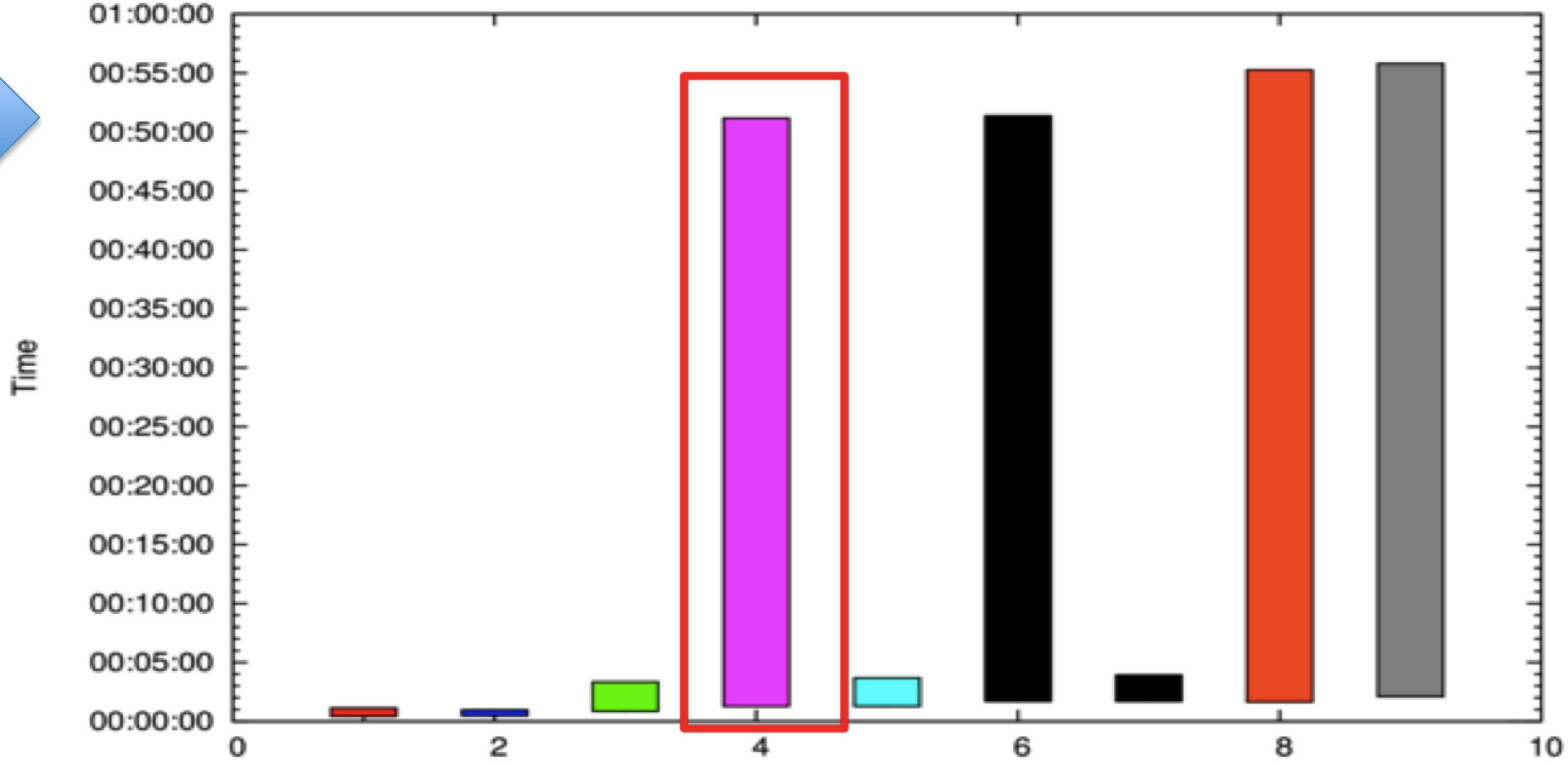
Histogram difference is calculated between color spaces

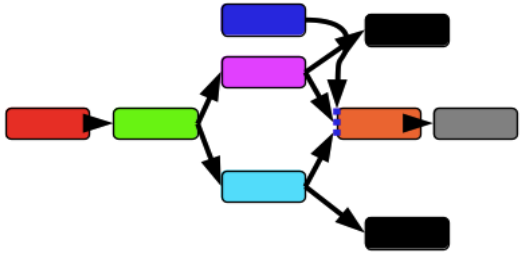
Example of Scientific workflow

Color coded workflow to better understand the result graphs



Workflow Without Scaling





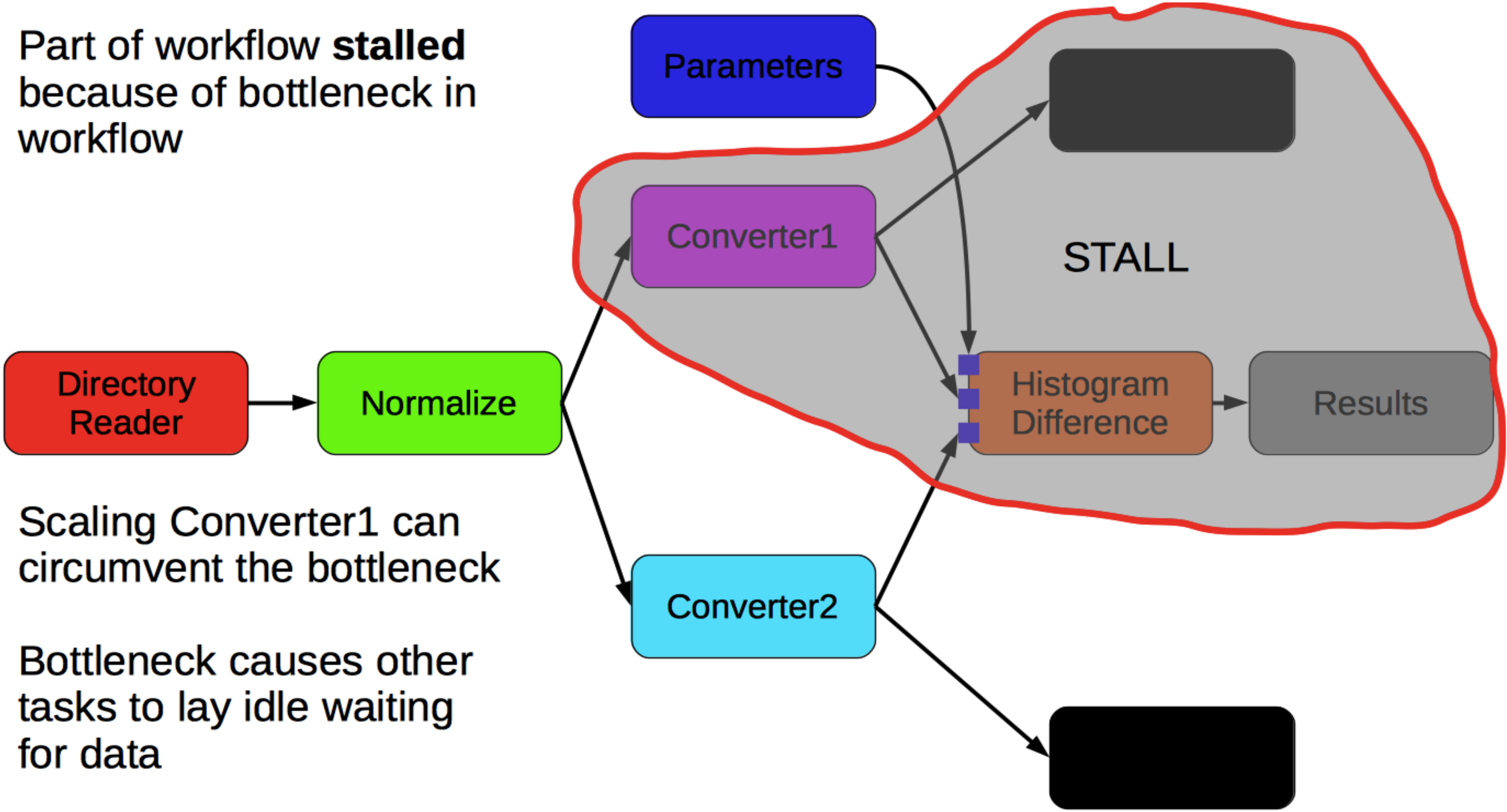
Slow task causing a **bottleneck** in the workflow

Example of Scientific workflow (1)

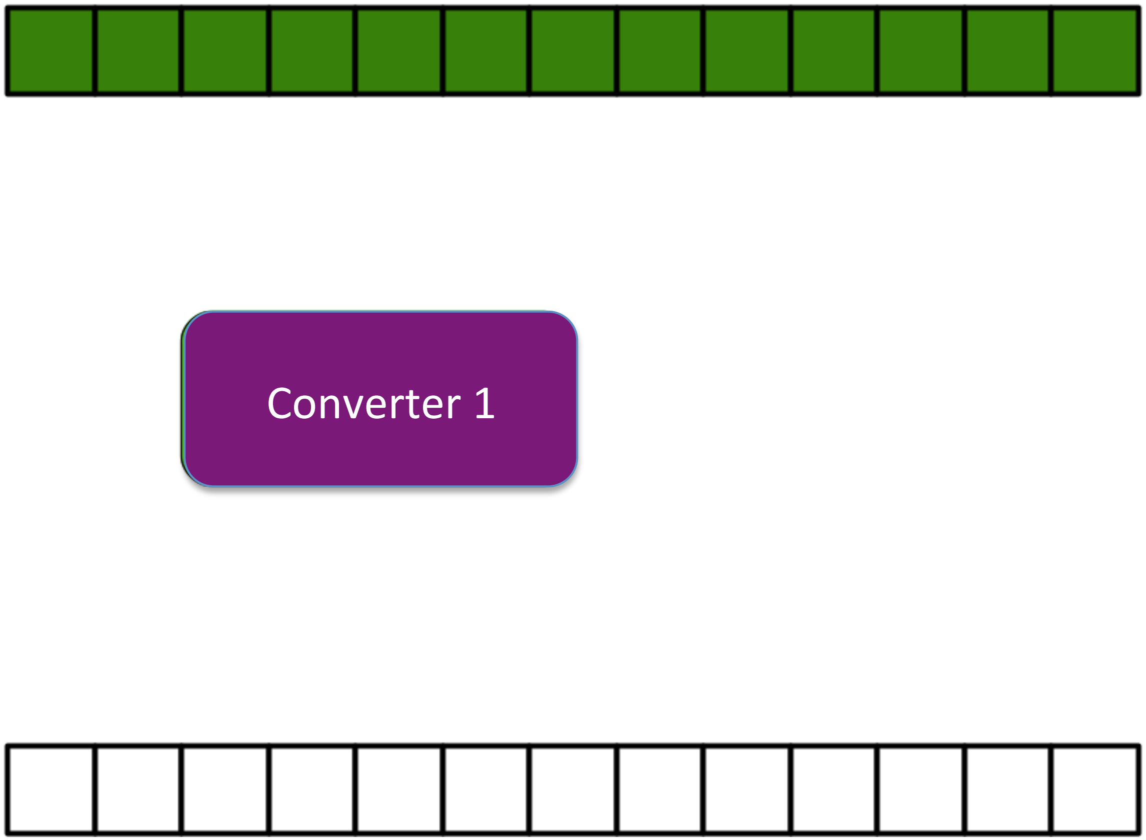
Part of workflow **stalled** because of bottleneck in workflow

Scaling Converter1 can circumvent the bottleneck

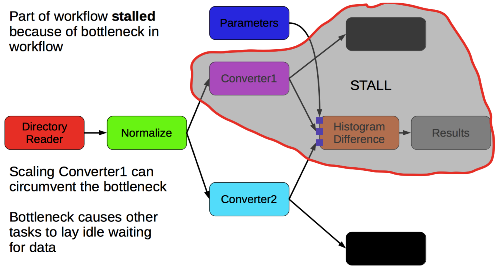
Bottleneck causes other tasks to lay idle waiting for data



Zooming into the Task Converter 1



Data organized in atomic parcels(messages)

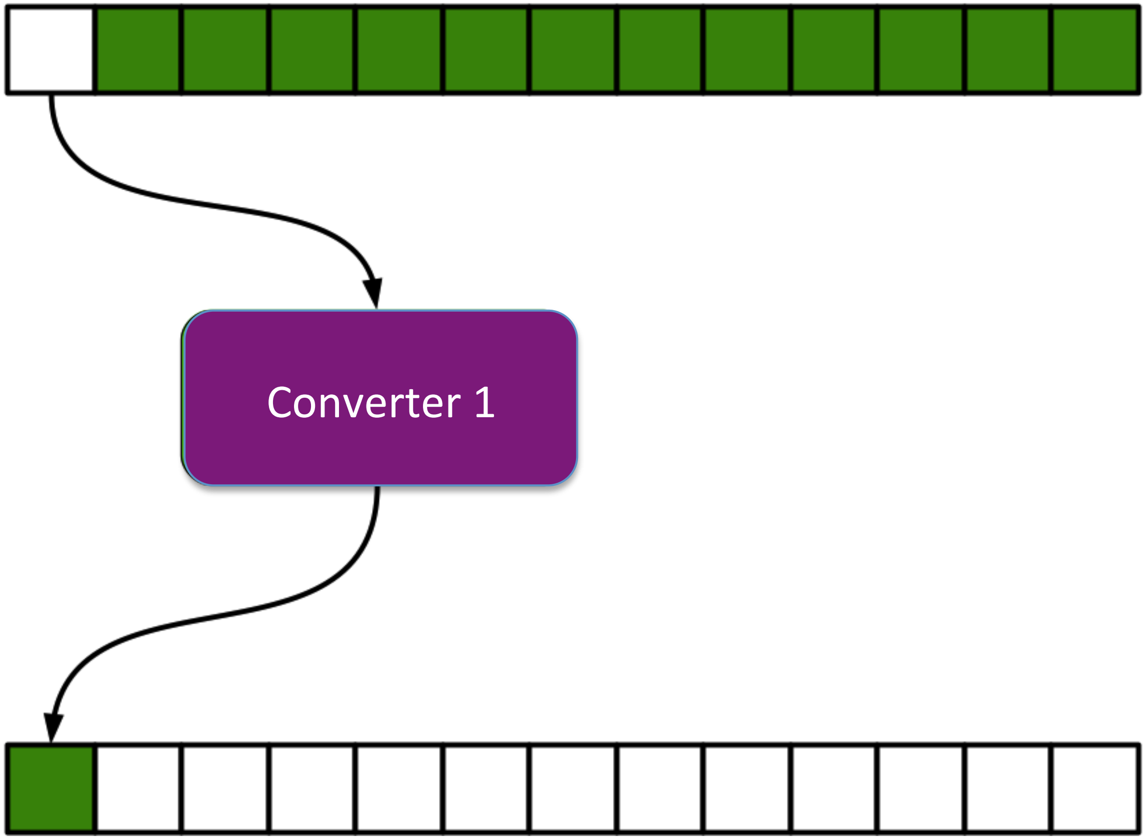


Part of workflow **stalled** because of bottleneck in workflow

Scaling Converter1 can circumvent the bottleneck

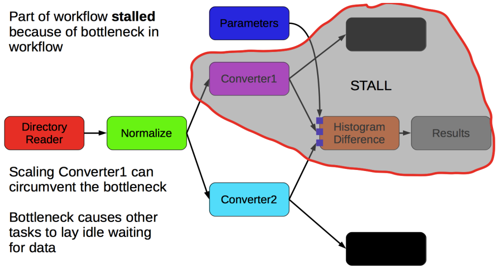
Bottleneck causes other tasks to lay idle waiting for data

Zooming into the Task Converter 1



Data organized in atomic parcels(messages)

Task processes data sequentially

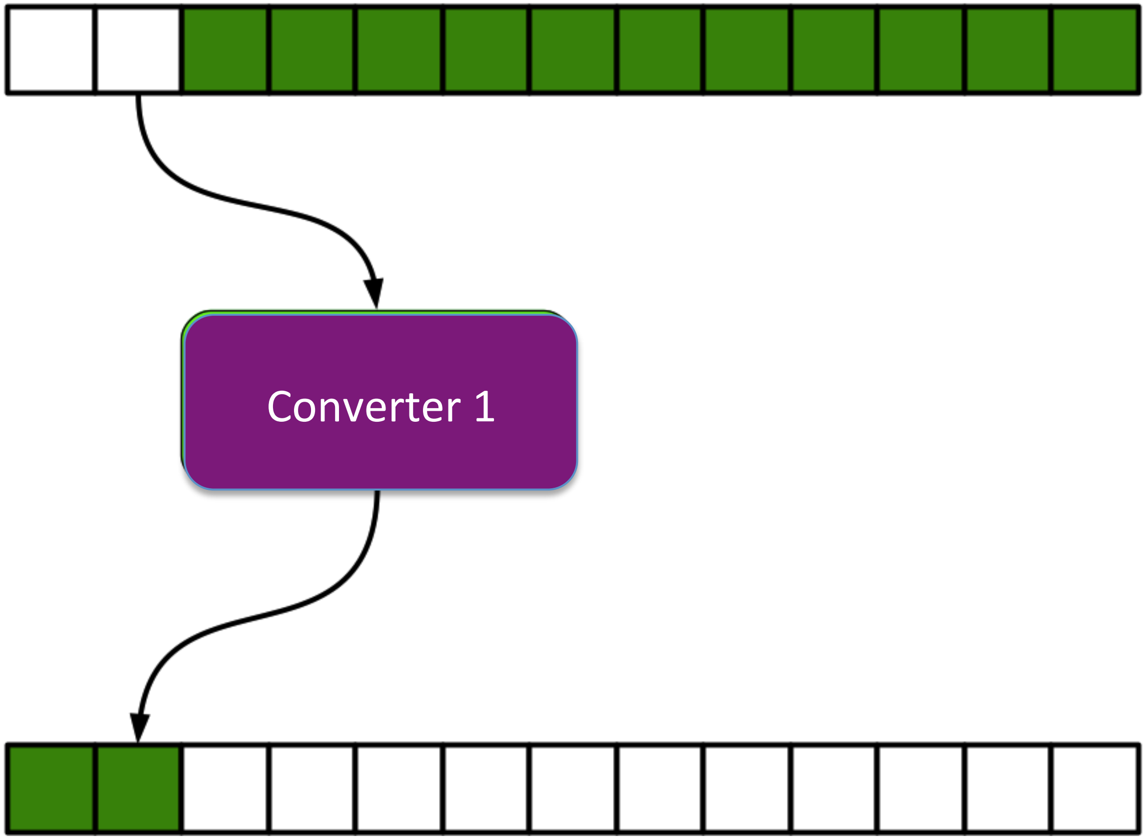


Part of workflow **stalled** because of bottleneck in workflow

Scaling Converter1 can circumvent the bottleneck

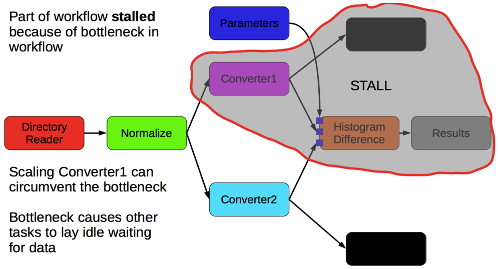
Bottleneck causes other tasks to lay idle waiting for data

Scaling Concepts



Data organized in atomic parcels(messages)

Task processes data sequentially

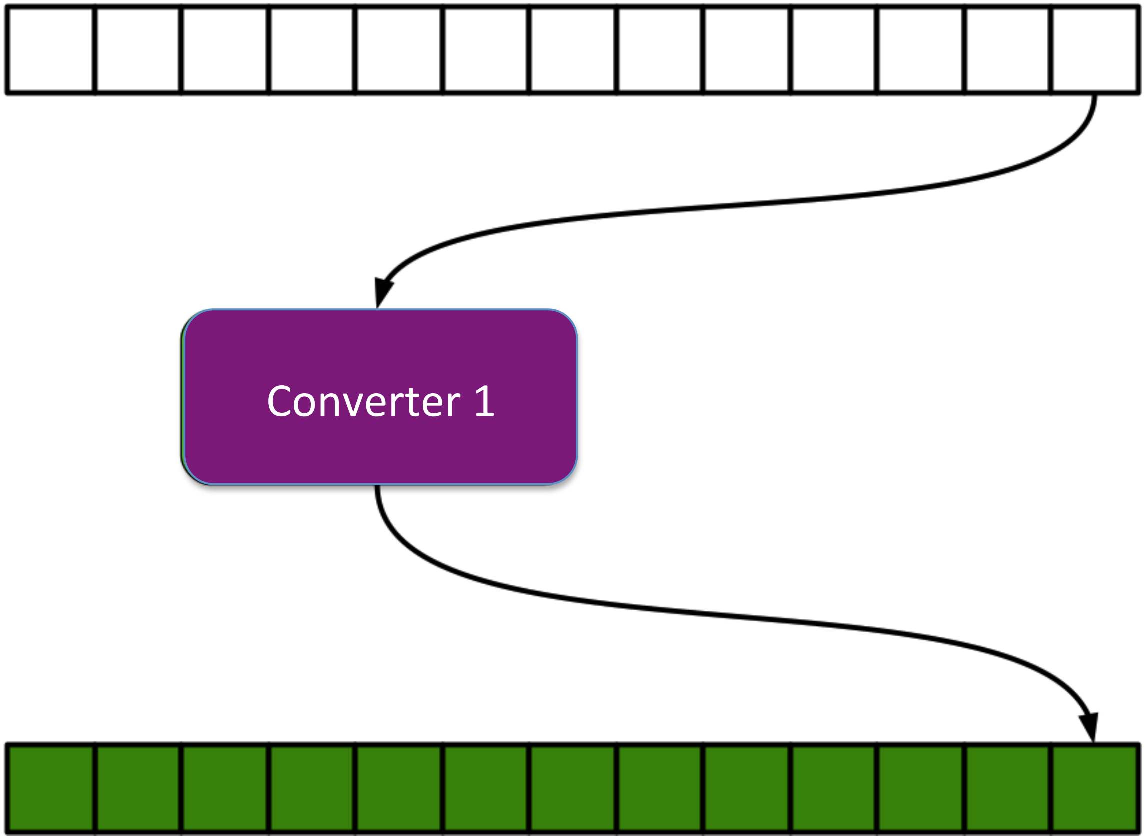


Part of workflow **stalled** because of bottleneck in workflow

Scaling Converter1 can circumvent the bottleneck

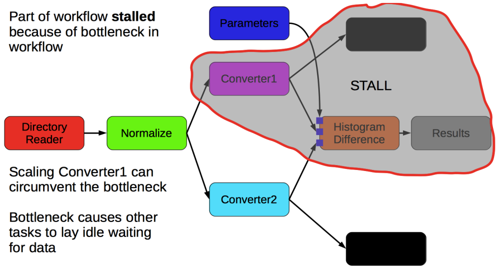
Bottleneck causes other tasks to lay idle waiting for data

Zooming into the Task Converter 1



Data organized in atomic parcels(messages)

Task processes data sequentially

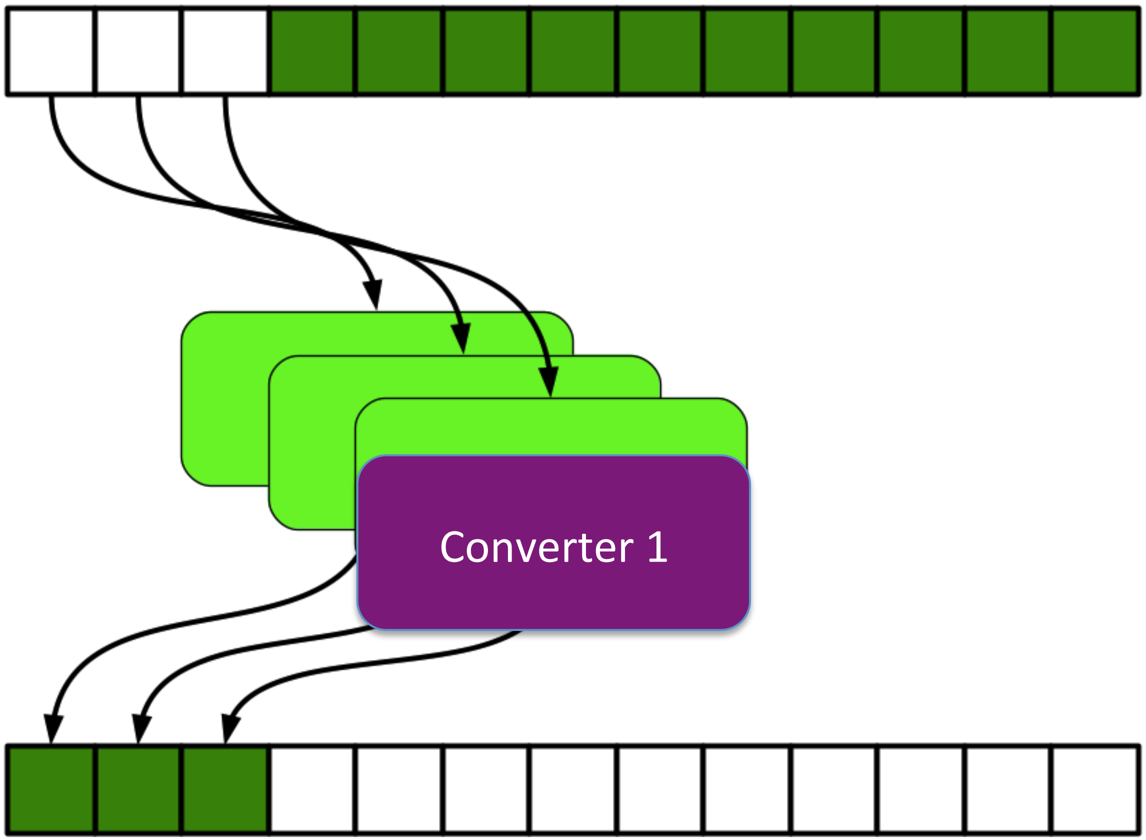


Part of workflow **stalled** because of bottleneck in workflow

Scaling Converter1 can circumvent the bottleneck

Bottleneck causes other tasks to lay idle waiting for data

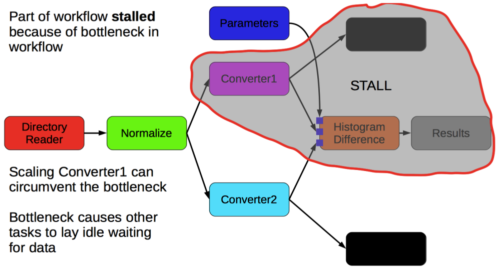
Zooming into the Task Converter 1



Data organized in atomic parcels(messages)

Tasks processes data **concurrently**

Adding more tasks increases **message consumption** rate

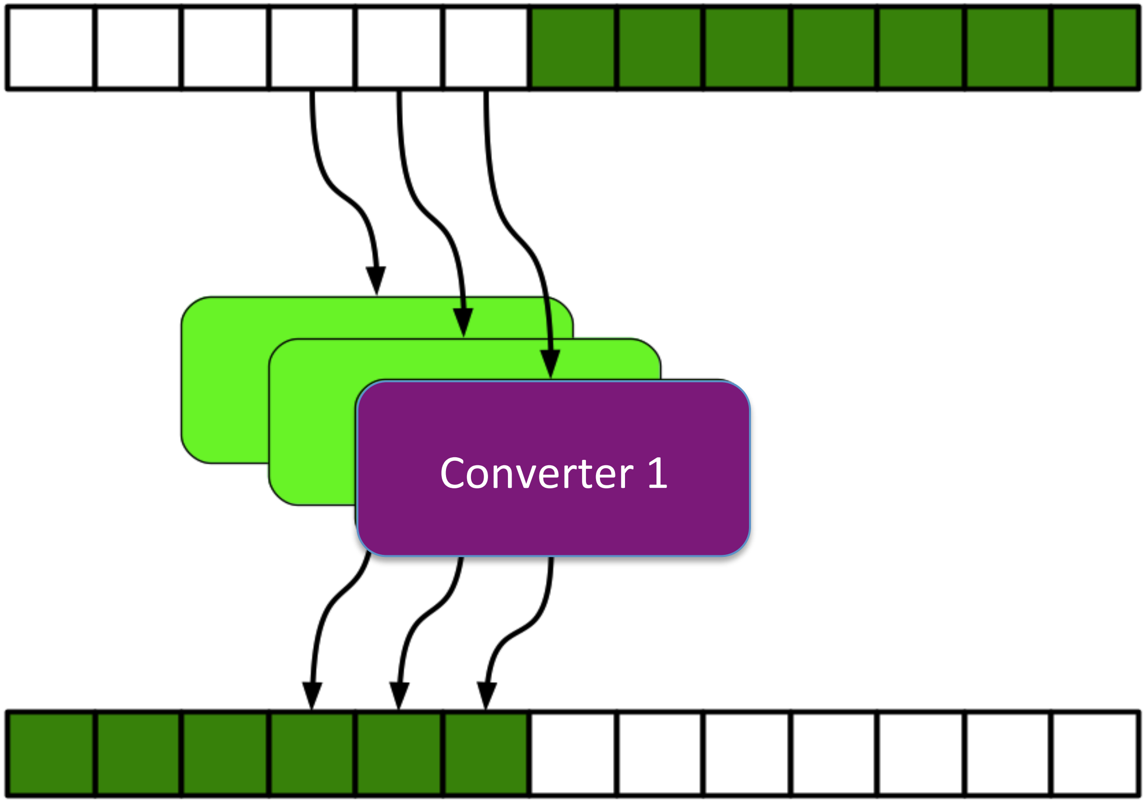


Part of workflow **stalled** because of bottleneck in workflow

Scaling Converter1 can circumvent the bottleneck

Bottleneck causes other tasks to lay idle waiting for data

Zooming into the Task Converter 1



Data organized in atomic parcels(messages)

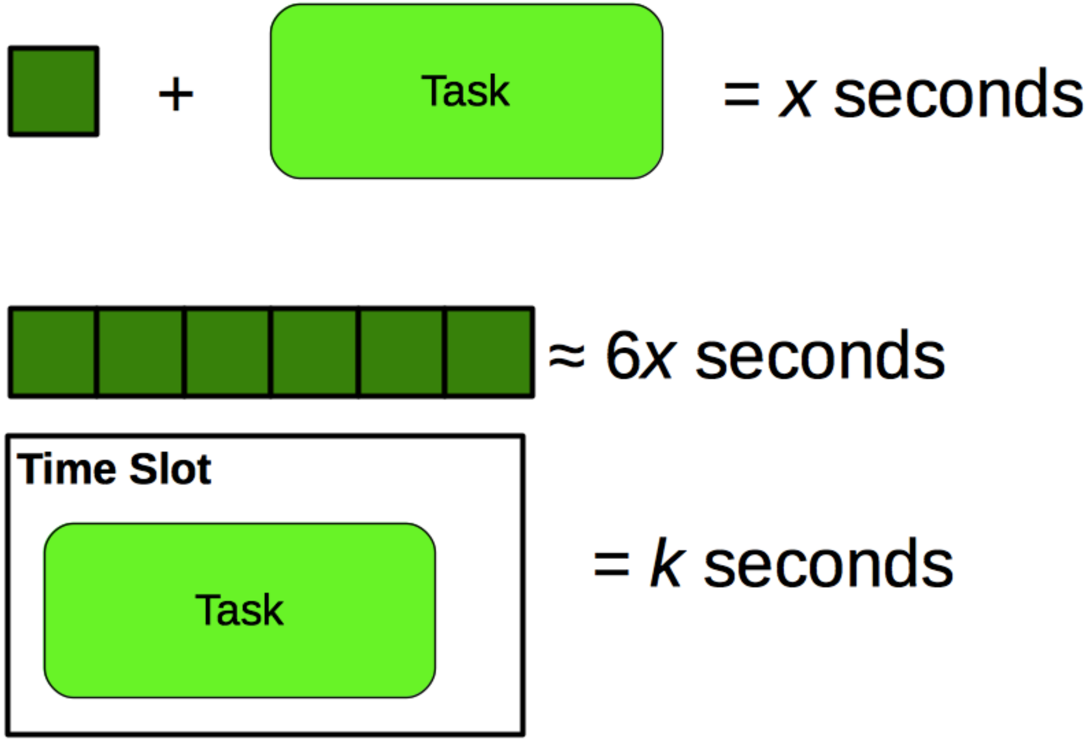
Task processes data sequentially

Adding more tasks increases **message consumption** rate

Challenge: How many tasks to create?

Too **many** and tasks get stuck on queues. Too **few** and optimal performance not achieved

Load Prediction



🟩 + Task = *x* seconds

🟩🟩🟩🟩🟩🟩 ≈ 6*x* seconds

Time Slot

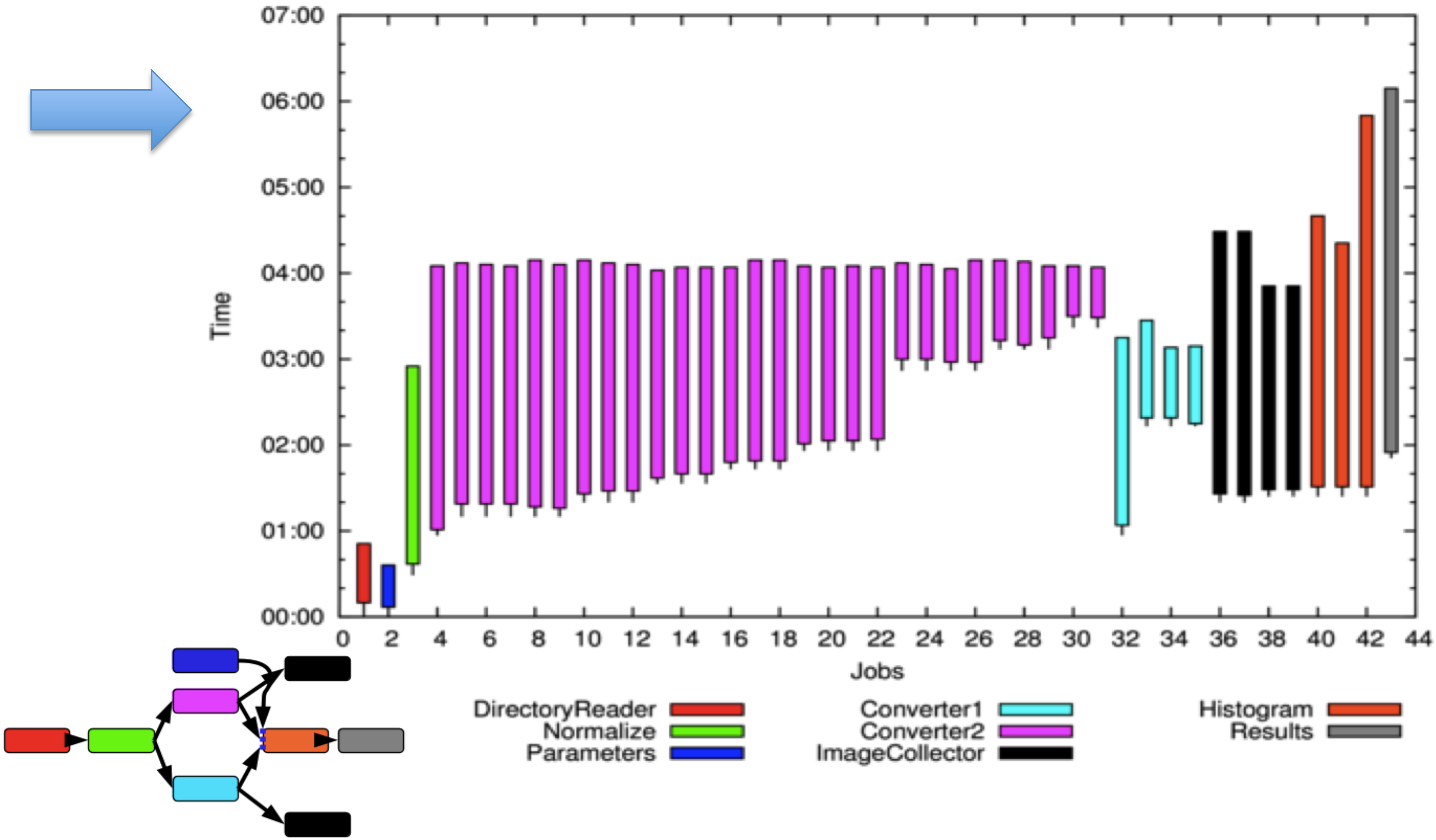
Task

= *k* seconds

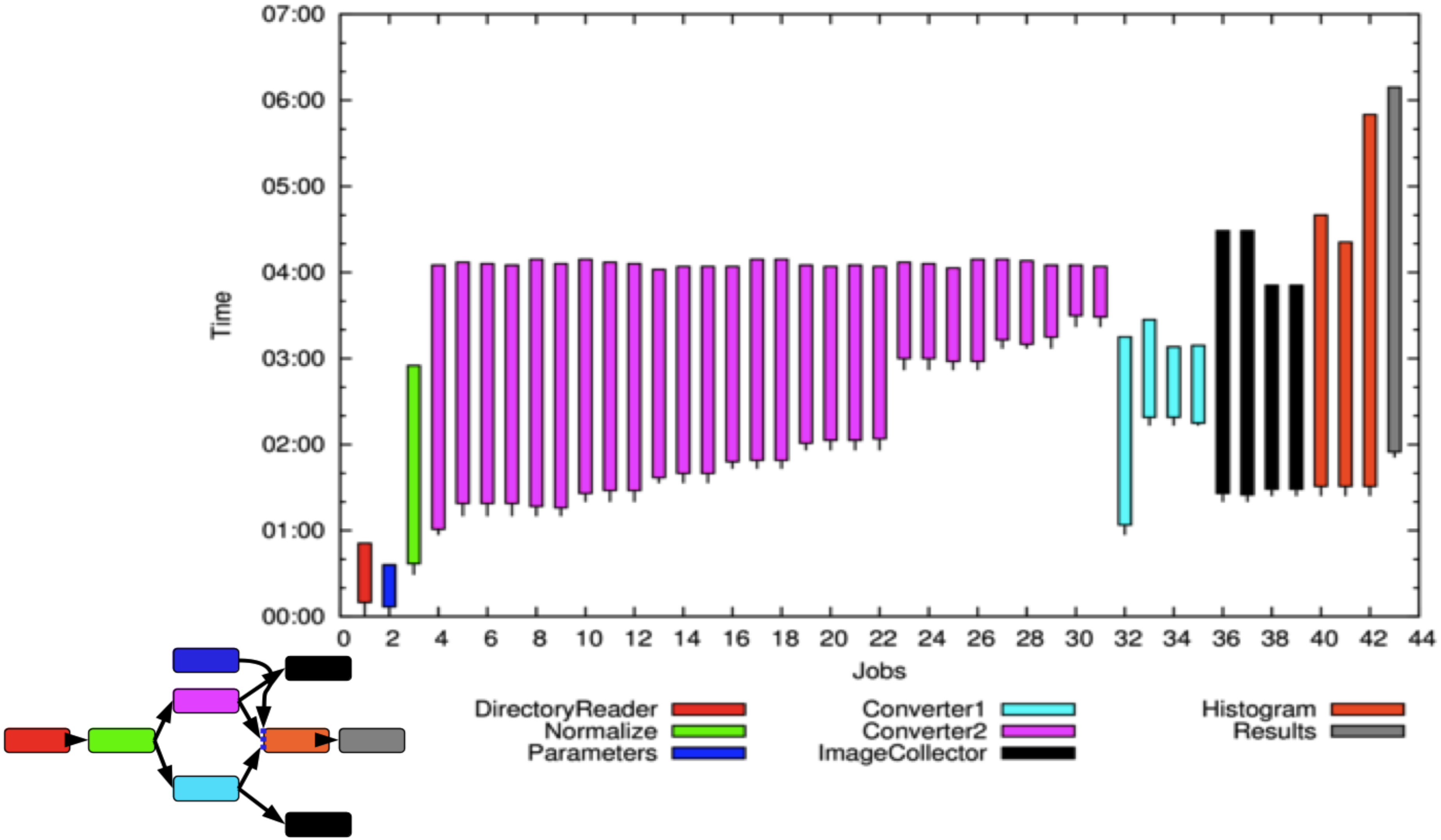
Simplified Load = 6*x*/*k* time slots

Assumption: Size of data directly proportional to computation time. May not always be the case

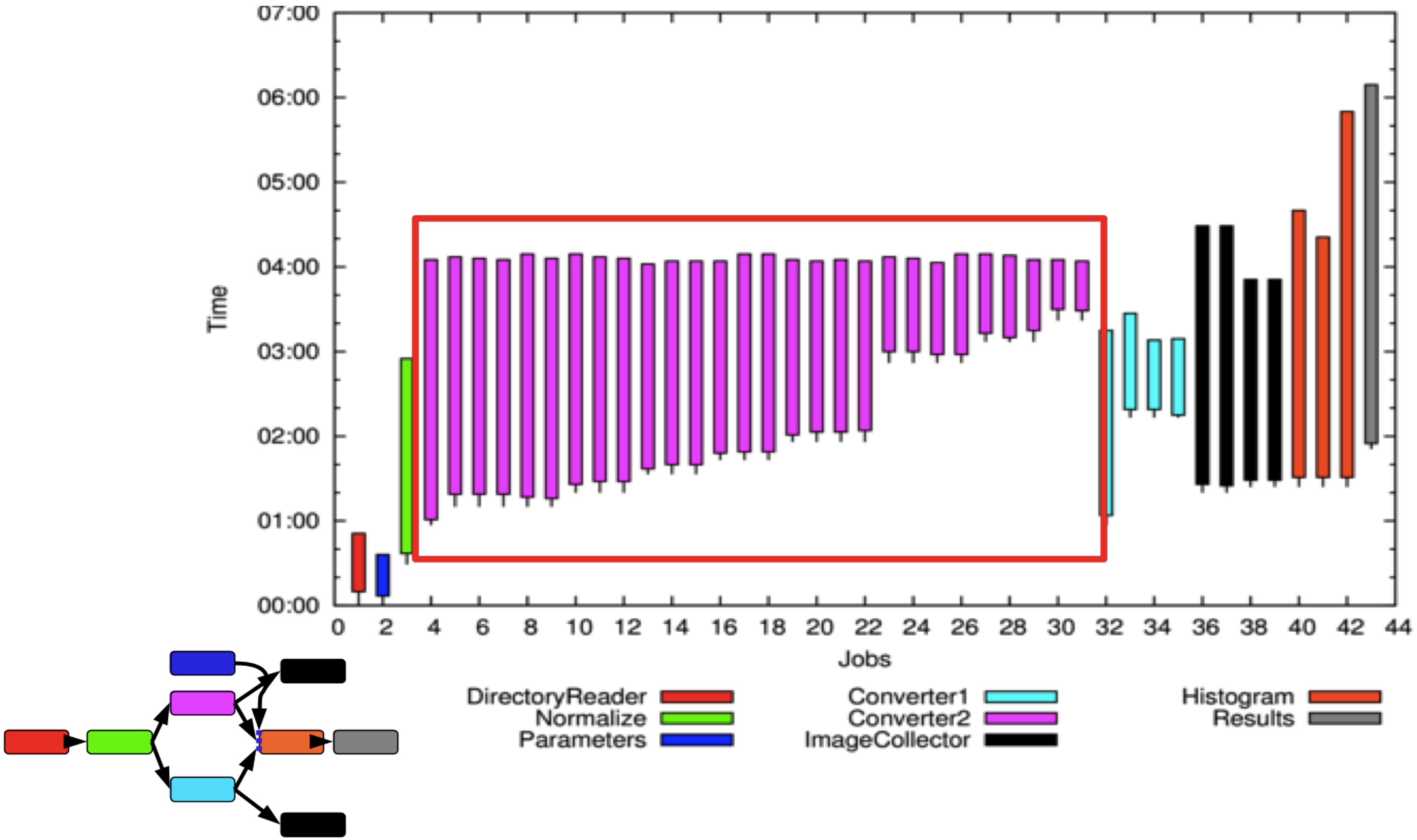
Workflow execution with Scaling



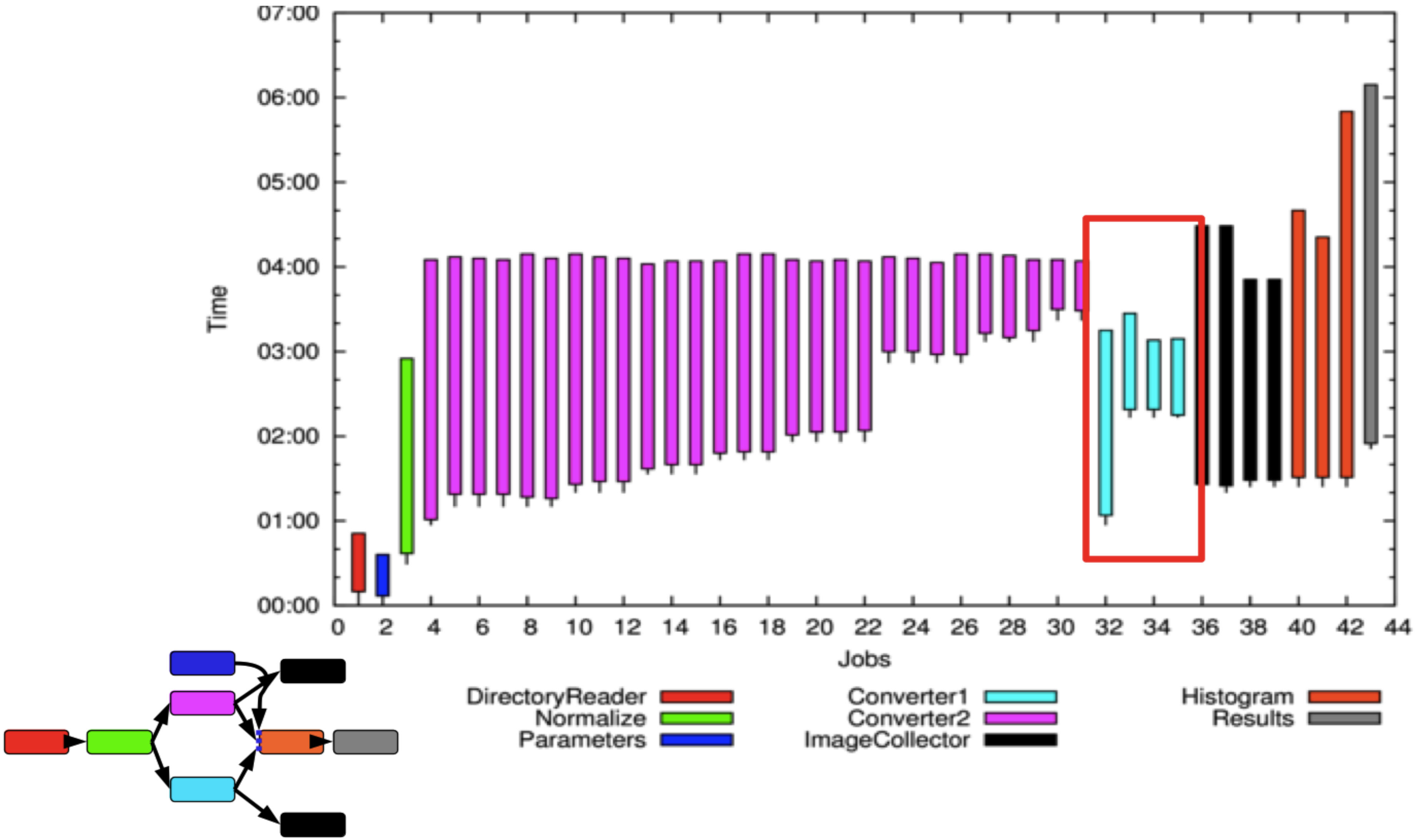
Workflow execution with Scaling



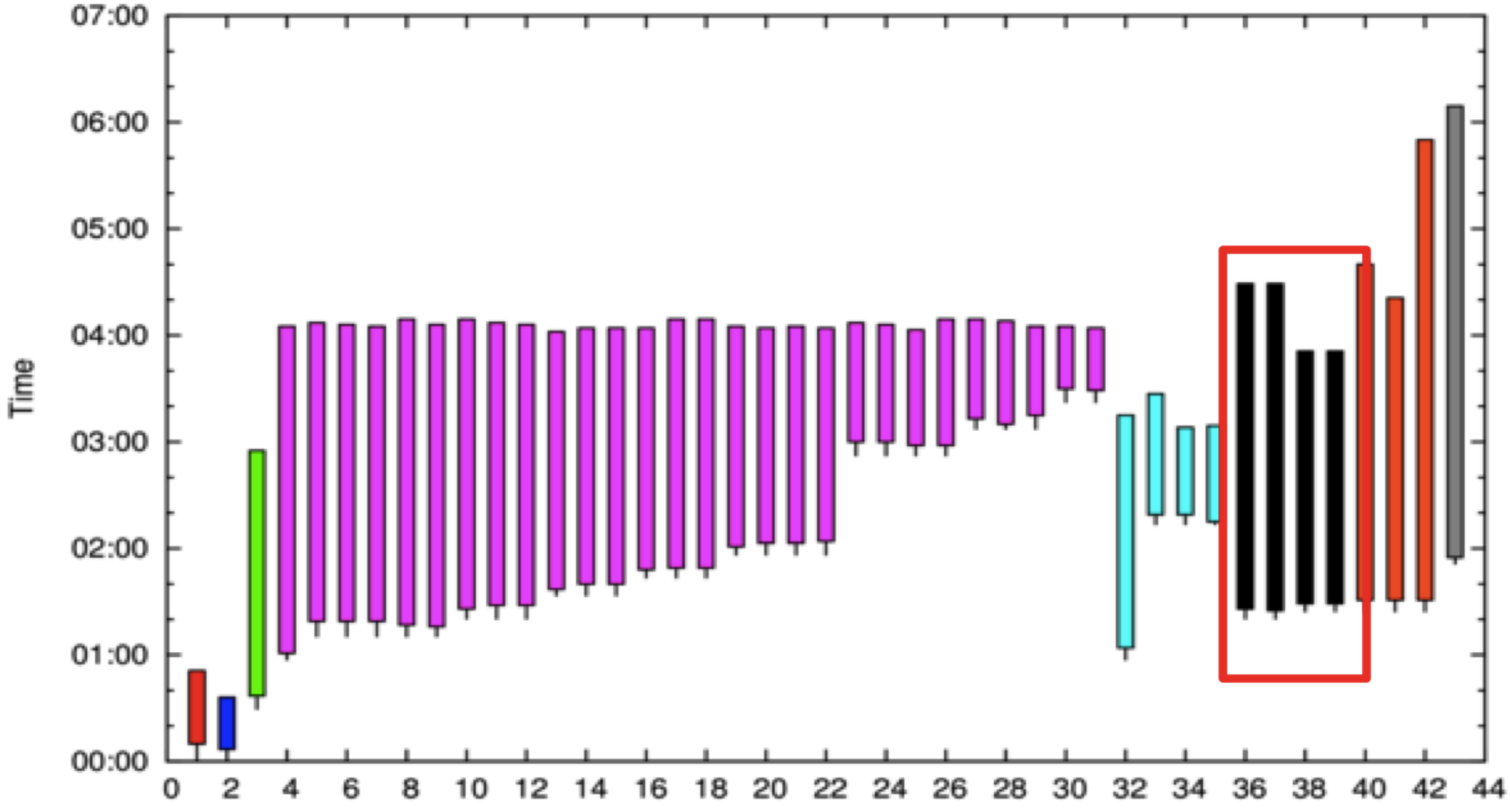
Workflow execution with Scaling Task -1

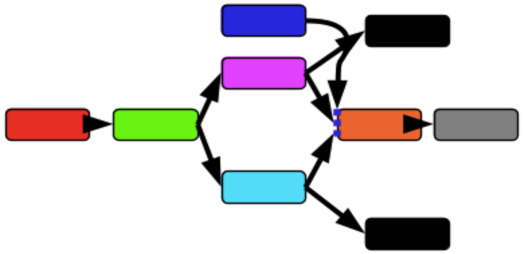


Workflow execution with Scaling Task -2



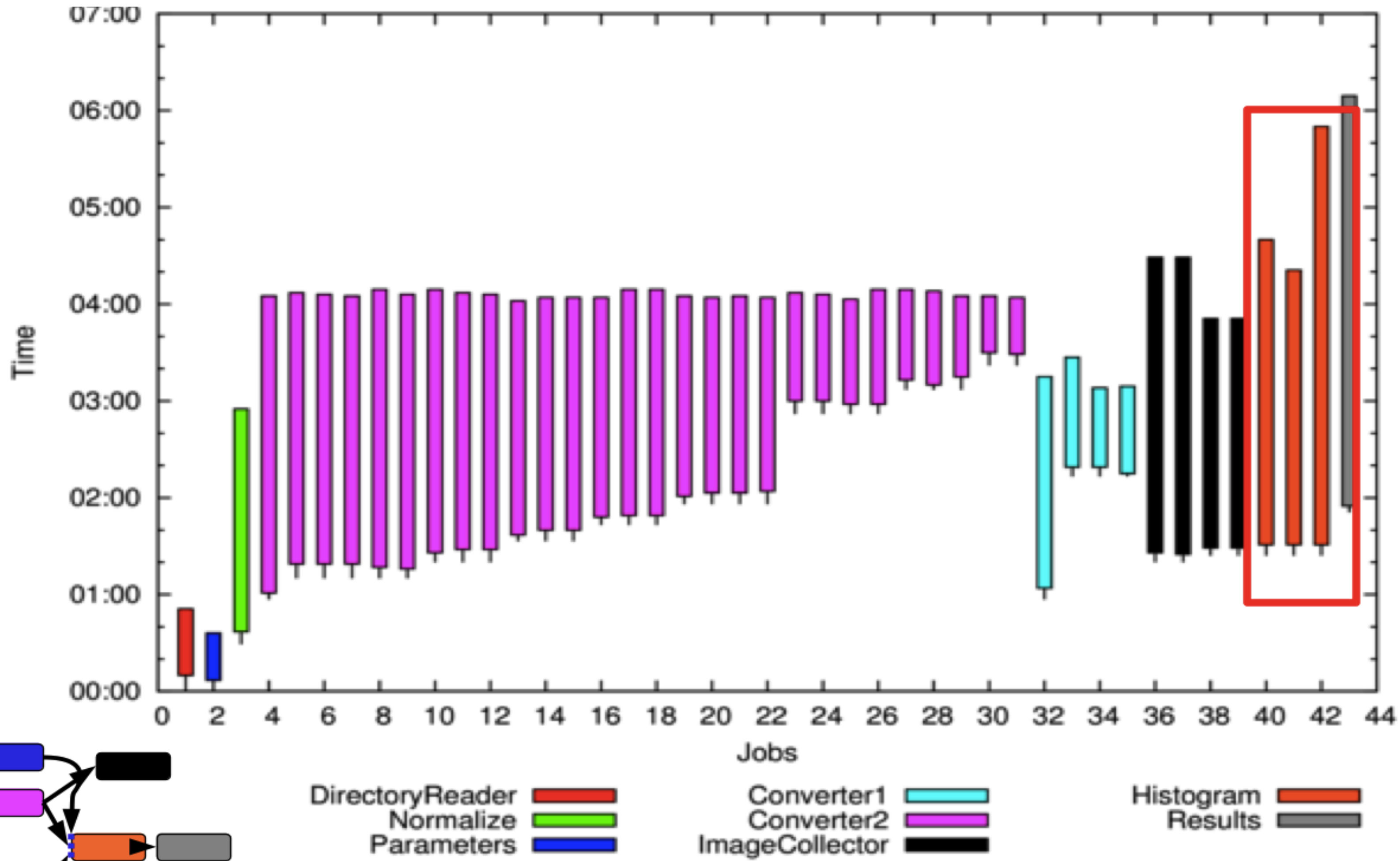
Other Scaled Task -1

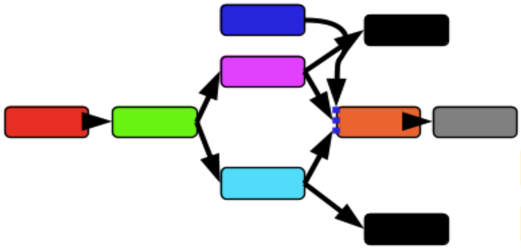




ImageCollector was set to a **fixed** amount (4)

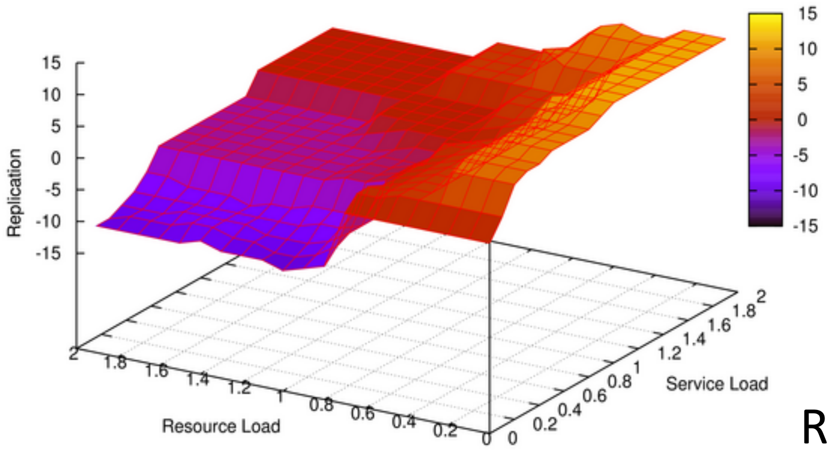
Auto Scaling Task -2



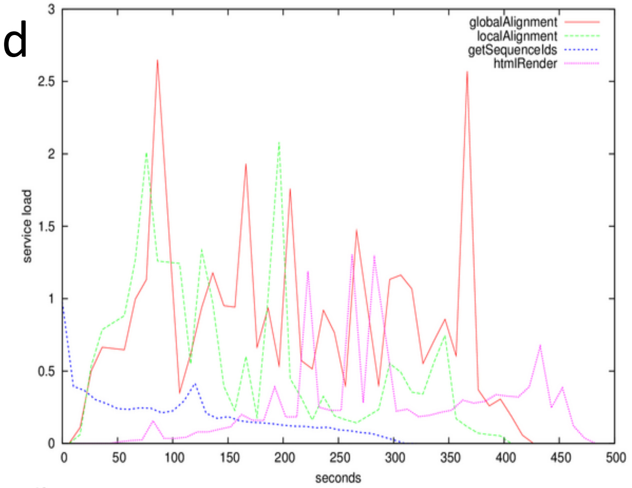


HistogramDifference was set to **one-to-one** scaling.
Each parameter generates a new task

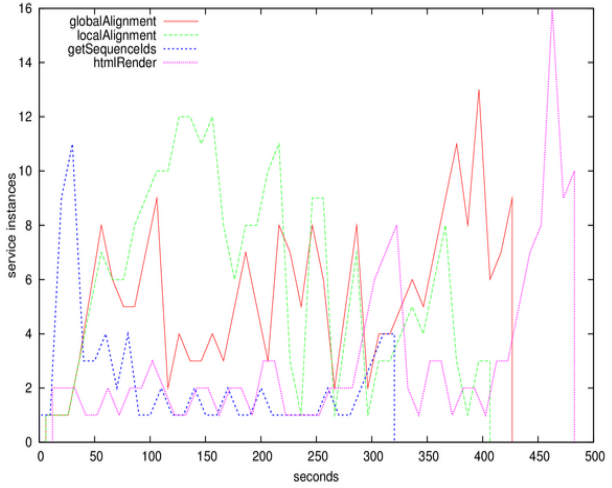
Example of Scientific workflow (2)

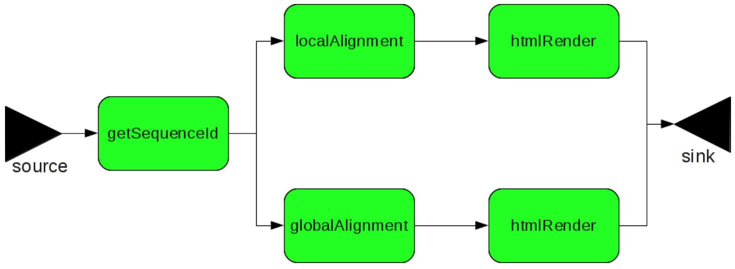


Service Load



Running Service instances

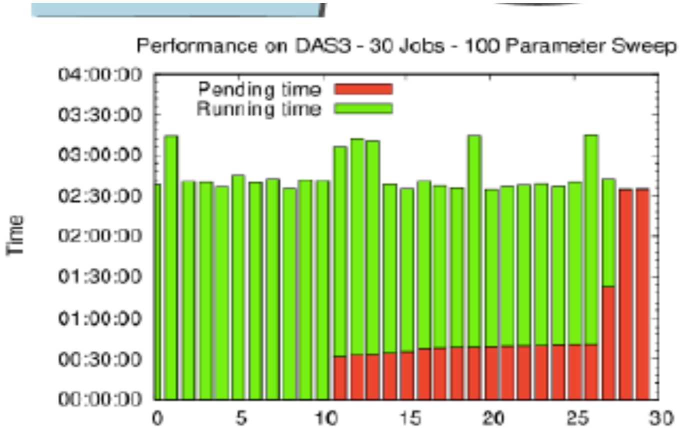


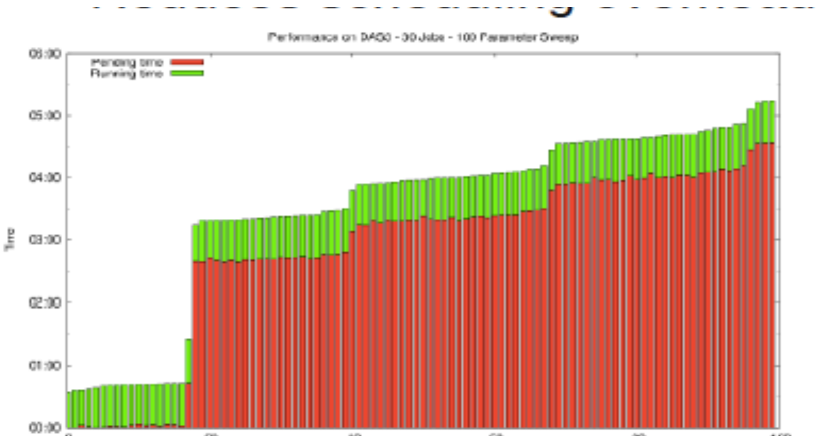


Reginald Cushing, Spiros Koulouzis, Adam S. Z. Belloum, Marian Bubak, **Dynamic Handling for Cooperating Scientific Web Services**, 7th IEEE International Conference on e-Science, December 2011, Stockholm, Sweden

Workflow as a Service (WFaaS)

- Once a workflow is initiated on the resources it stays alive and process data/jobs continuously
- Reduce the scheduling overhead





Reginald Cushing, Adam S. Z. Belloum, V. Korkhov, D. Vasyunin, M.T. Bubak, C. Leguy ECMLS'12, June 18, 2012, Delft, *Workflow as a Service: An Approach to Workflow Farming*, The Netherlands

Workflow Issues

- **Workflow description**

  How to capture **knowledge** of expert while still **hiding** complexity of underlying system.

  - **Workflow Models:** allow to **model** the tasks and **dependencies** between them (DAG, Petri Net)
  - **Workflow languages:** provide the required support to express the workflow model.

- **Workflow Enactment:** The functions provided by enactment are scheduling, fault management and data movement.
  - In the context of Grid environment workflow enactment service can be built on the top of low level Grid middleware

Workflow Enactment

- Workflow Refinement
  - Modification from the workflow description
  - Reduction of workflow if some data already exist
  - Additional data movement preparation if needed

- Mapping to actual resource
  - Resource discovery, allocation and management
  - Bind to real computing resource

- Workflow Fault Tolerance & Monitoring of Execution
  - Two level failure recovery techniques
    - Task Level
    - Workflow Level

Model of computation

- Model of computation: stream-based process network.
  - Engine **co-allocates** all workflows.
  - Components waste time idling.
  - Co-allocation difficult.
- Communication: time **coupled**
  - Assumes components are running
  - Simultaneously
  - Synchronized p2p
  - Fixed TCP/IP

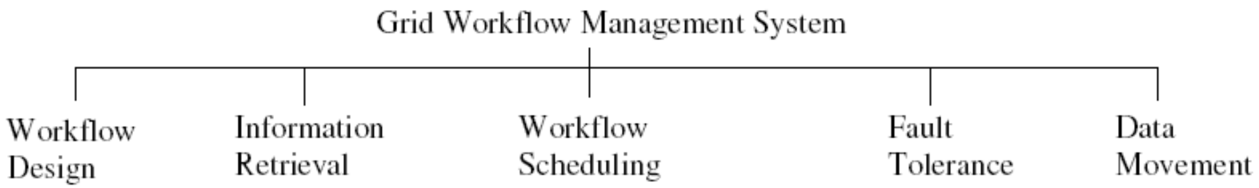
V. Korkhov et al. VLAM-G: Interactive data driven workflow engine for Grid-enabled resources, Scientific Programming 15 (2007) 173–188 173 IOS Press

Model of computation

- Model of computation: dataflow network
  - components **scheduled** depending on data
  - components **only activated** when data is available
  - **no need** for **co-allocation**

- Communication: time decouples
  - messaging communication system.
  - components not synchronized
  - communication not strictly TCP/IP

Workflow Taxonomy


Grid Workflow Management System

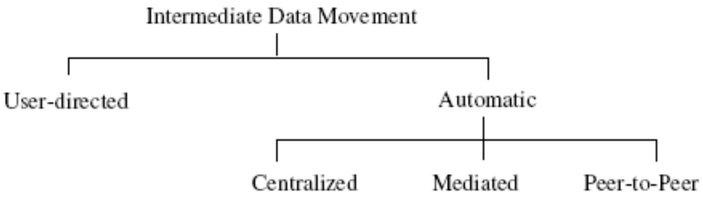
Workflow Design

Information Retrieval

Workflow Scheduling

Fault Tolerance

Data Movement


Intermediate Data Movement

User-directed

Automatic

Centralized

Mediated

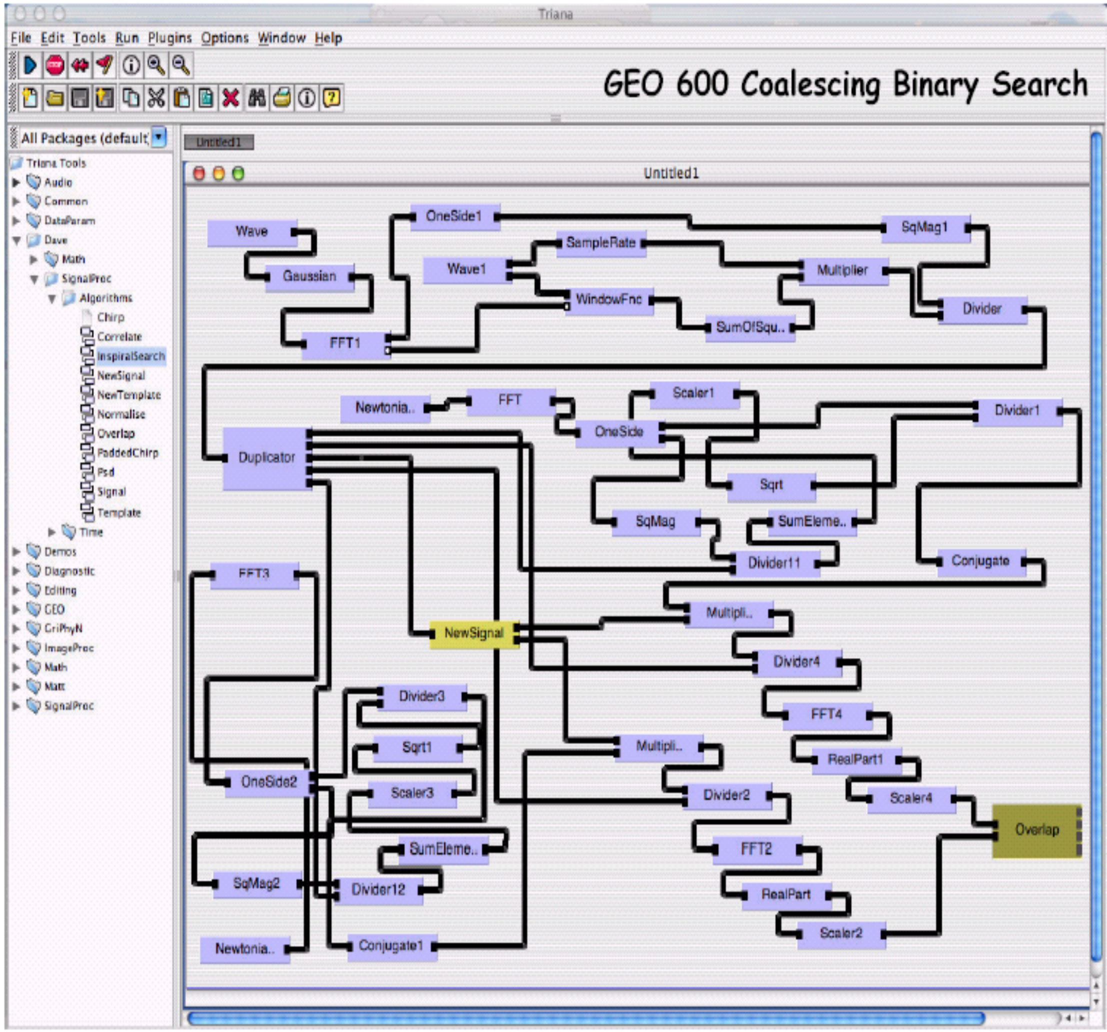
Peer-to-Peer

- For Grid workflow applications,
  - the input files of tasks need to be staged to a remote site before processing the task.
  - Similarly, output files may be required by their children tasks which are processed on other resources.
- The intermediate data has to be staged out to the corresponding Grid sites.

A Taxonomy of Workflow Management Systems for Grid Computing
Jia Yu and Rajkumar Buyya, http://www.cloudbus.org/reports/GridWorkflowTaxonomy.pdf

Component Based Workflow Description: Triana

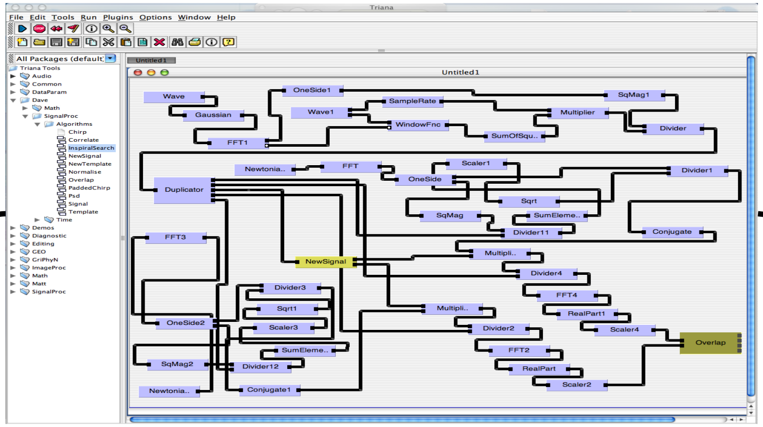
- Workflow design
  - workflow structure: Non-DAG
  - workflow model: concrete
  - workflow composition: user-directed: Graph-based Modeling: User-defined component
- Information retrieval
  - GAT
- Scheduling
  - Architecture: decentralized
  - Decision Making: local
  - Planning scheme: just-in-time
  - Scheduling strategy: GAT
- Performance estimation: N/A
- Fault tolerance :GAT
- Data movement: P2P



Triana, the GAT and the GAP

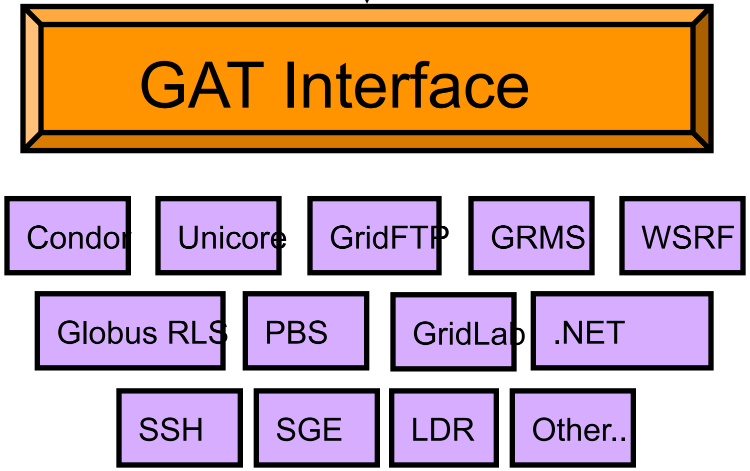
Grid Computing:

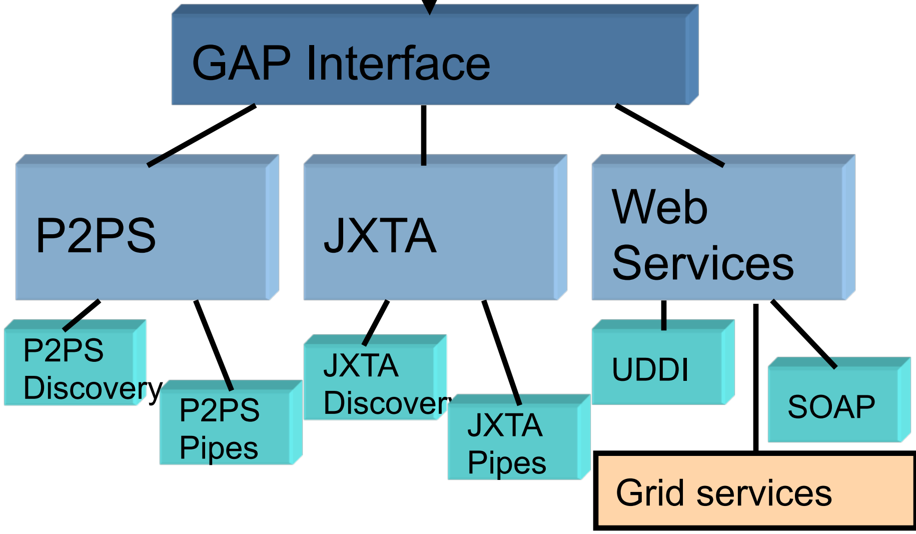
Job Submission, data services
On top of a number of Grid Middleware



Service Based Computing:

Deployment, discovery and communication with distributed services e.g. P2P and (GSI) Web services

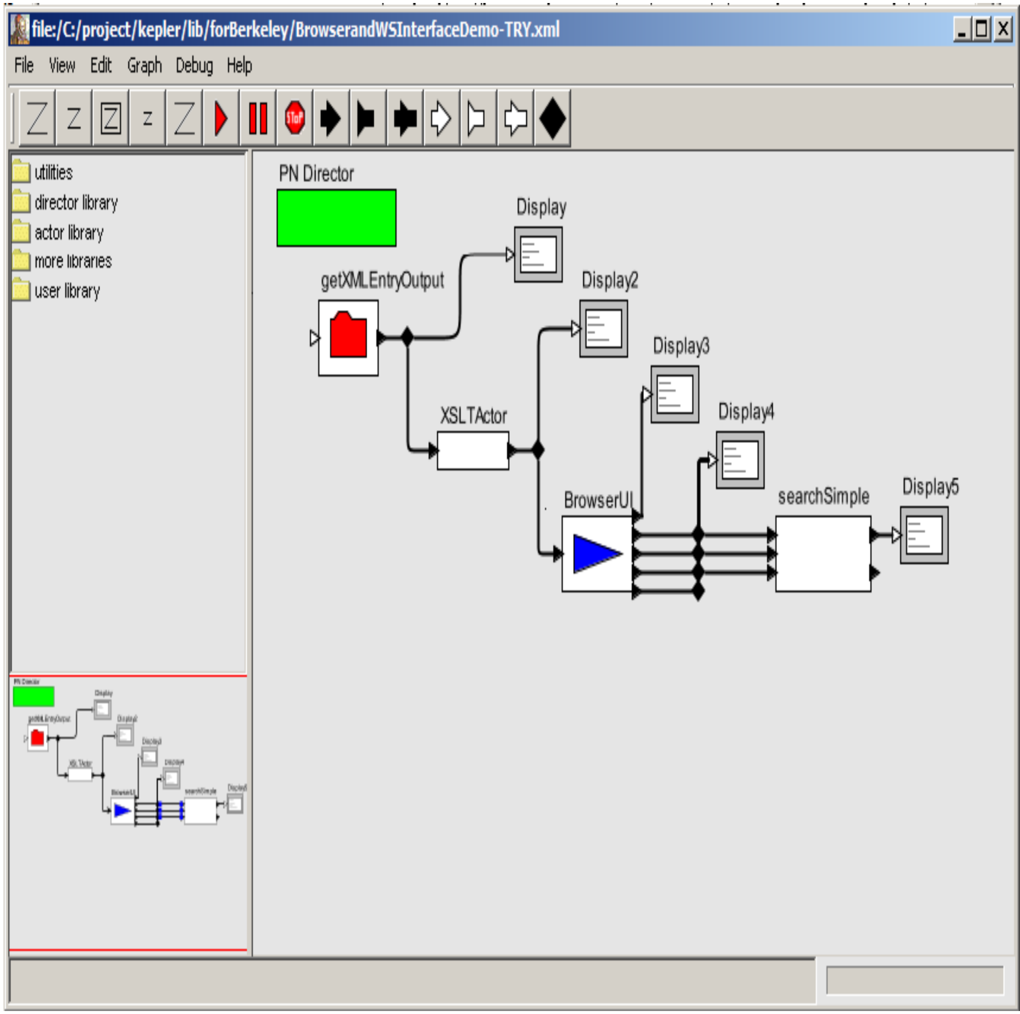




Component Based Workflow Description: Kepler

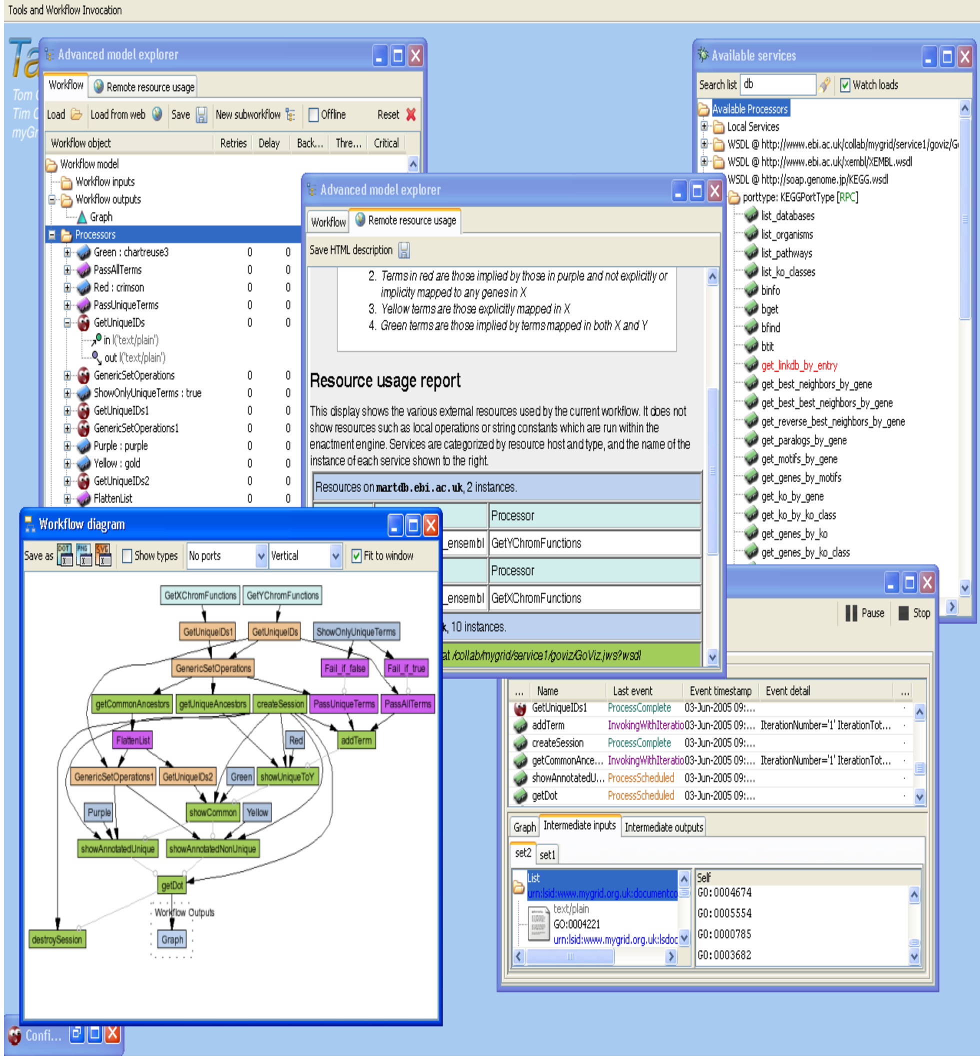
- Workflow design
  - workflow structure: Non-DAG
  - workflow model: concrete
  - workflow composition: user-directed: Graph-based Modeling: User-defined component

- Scheduling
  - architecture: centralized
  - Decision Making: local
  - Planning scheme: static
  - Scheduling strategy: Performance-driven
- Performance estimation: N/A
- Fault tolerance: N/A
- Data movement: P2P



Program/Application: workflow Based: Taverna

- Workflow design
  - workflow structure: DAG
  - workflow model: concrete
  - workflow composition: user-directed: Graph-based Modeling: User-defined component
- Information retrieval
  - Dynamic: execution related
- Scheduling
  - Architecture: centralized
  - Decision Making: local
  - Planning scheme: just-in-time
  - Scheduling strategy: N/A
  - Performance estimation: N/A
- Fault tolerance: Task-level (retry, Alternate)
- Data movement: Centralized



2. *Terms in red are those implied by those in purple and not explicitly or implicity mapped to any genes in X*
3. *Yellow terms are those explicitly mapped in X*
4. *Green terms are those implied by terms mapped in both X and Y*

Resource usage report

This display shows the various external resources used by the current workflow. It does not show resources such as local operations or string constants which are run within the enactment engine. Services are categorized by resource host and type, and the name of the instance of each service shown to the right.

Resources on **martdb.ebi.ac.uk**, 2 instances.	
	Processor
_ensembl	GetYChromFunctions
	Processor
_ensembl	GetXChromFunctions
k, 10 instances.	
at /collab/mygrid/service1/goviz/GoViz.jws?wsdl	

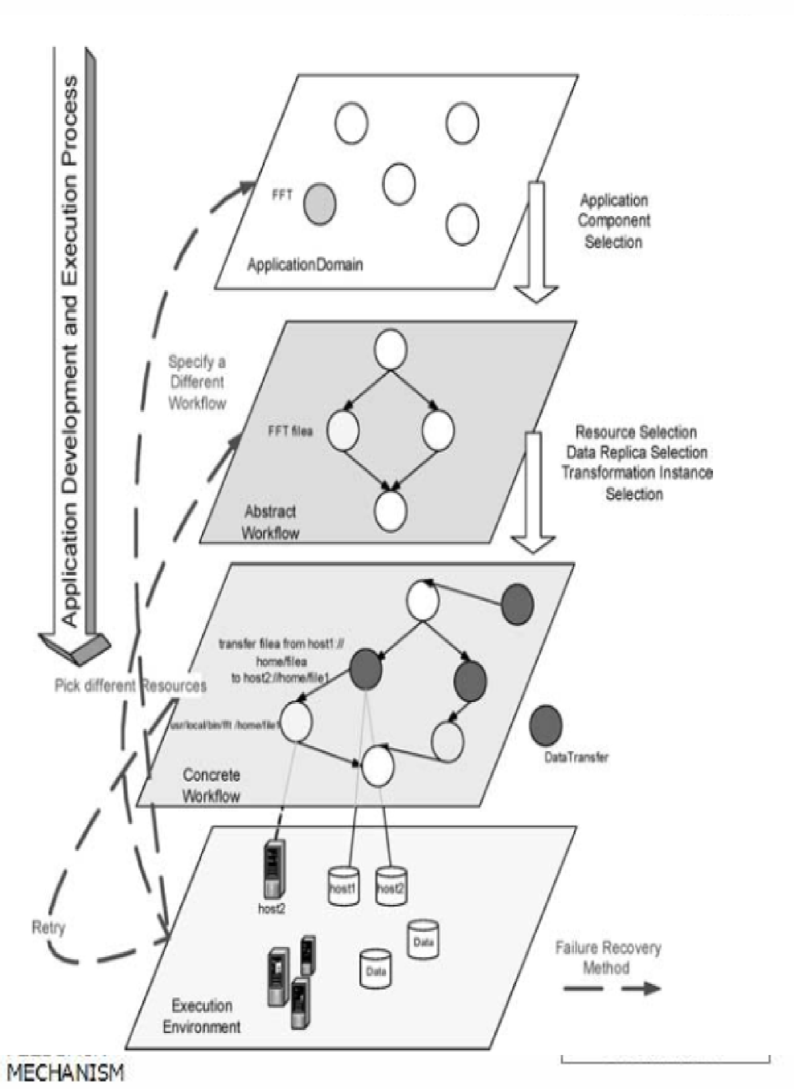
...	Name	Last event	Event timestamp	Event detail	...
	GetUniqueIDs1	ProcessComplete	03-Jun-2005 09:...		
	addTerm	InvokingWithIteratio	03-Jun-2005 09:...	IterationNumber='1' IterationTot...	
	createSession	ProcessComplete	03-Jun-2005 09:...		
	getCommonAnce...	InvokingWithIteratio	03-Jun-2005 09:...	IterationNumber='1' IterationTot...	
	showAnnotatedU...	ProcessScheduled	03-Jun-2005 09:...		
	getDot	ProcessScheduled	03-Jun-2005 09:...		

List
urn:lsid:www.mygrid.org.uk:documentco
text/plain
GO:0004221
urn:lsid:www.mygrid.org.uk:lsdoc

Self
GO:0004674
GO:0005554
GO:0000785
GO:0003682

Pegasus (GriPhyN)

- Pegasus Planner and Condor G Manager
- Pegasus convert abstract workflow into concrete Workflow
  - Prepare for data movements
  - Reduce workflow if data exist
  - Assign Resources to Processes
- Condor G Manager


Application Development and Execution Process
FFT
ApplicationDomain
Application Component Selection
Specify a Different Workflow
FFT filea
Abstract Workflow
Resource Selection Data Replica Selection Transformation Instance Selection
transfer filea from host1:// home/filea to host2://home/file1
/usr/local/bin/fft /home/file1
Pick different Resources
Concrete Workflow
DataTransfer
host2
host1
host2
Data
Data
Retry
Execution Environment
Failure Recovery Method
MECHANISM

The process of developing data intensive applications for Grid environments.

Source: http://www.cs.wisc.edu/condor/stork/papers/workflow-ccpe04.pdf

Workflow Refinement

- Example of a simple abstract workflow in which
  - the logical component *Extract* is applied to an input file with a logical filename F.a.
  - The resulting files F.b1 and F.b2, are used as inputs to the components identified by logical filenames *Resample* and *Decimate*.
  - Finally, the results are *Concatenated*

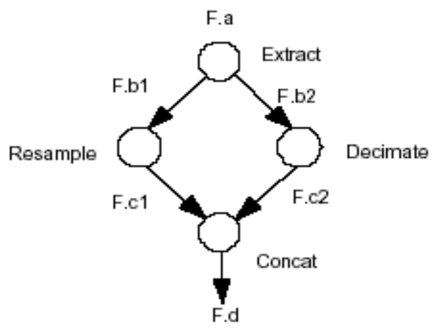


Figure 1.2. An example abstract workflow.

- If we assume that F.c2 is already available
  1. Reduces the workflow to 3 components, namely *Extract*, *Resample*, and *Concat*.
  2. Adds the transfer nodes for transferring F.c2 and F.a from their current locations.
  3. Adds transfer nodes between jobs that will run on different locations.
  4. Adds output transfer nodes to stage data out and registration nodes if the user requested that the resulting data be published and made available at a particular location.

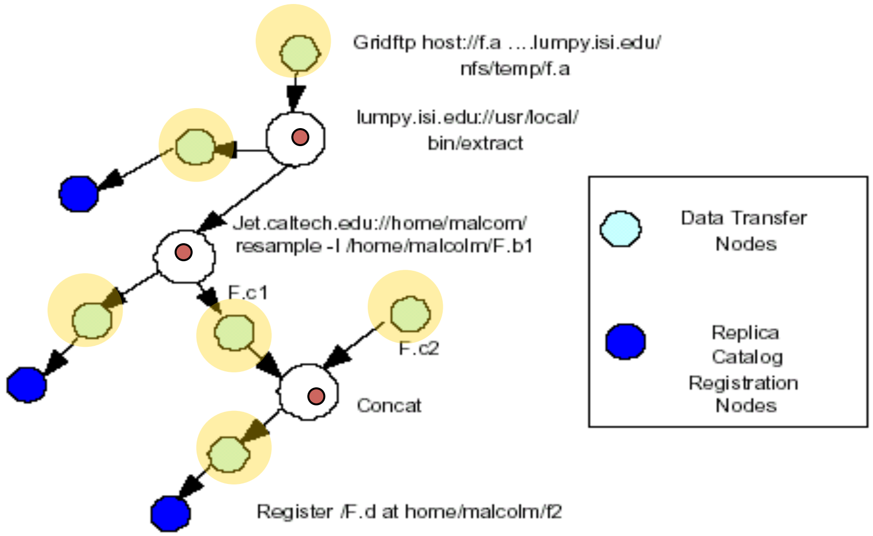
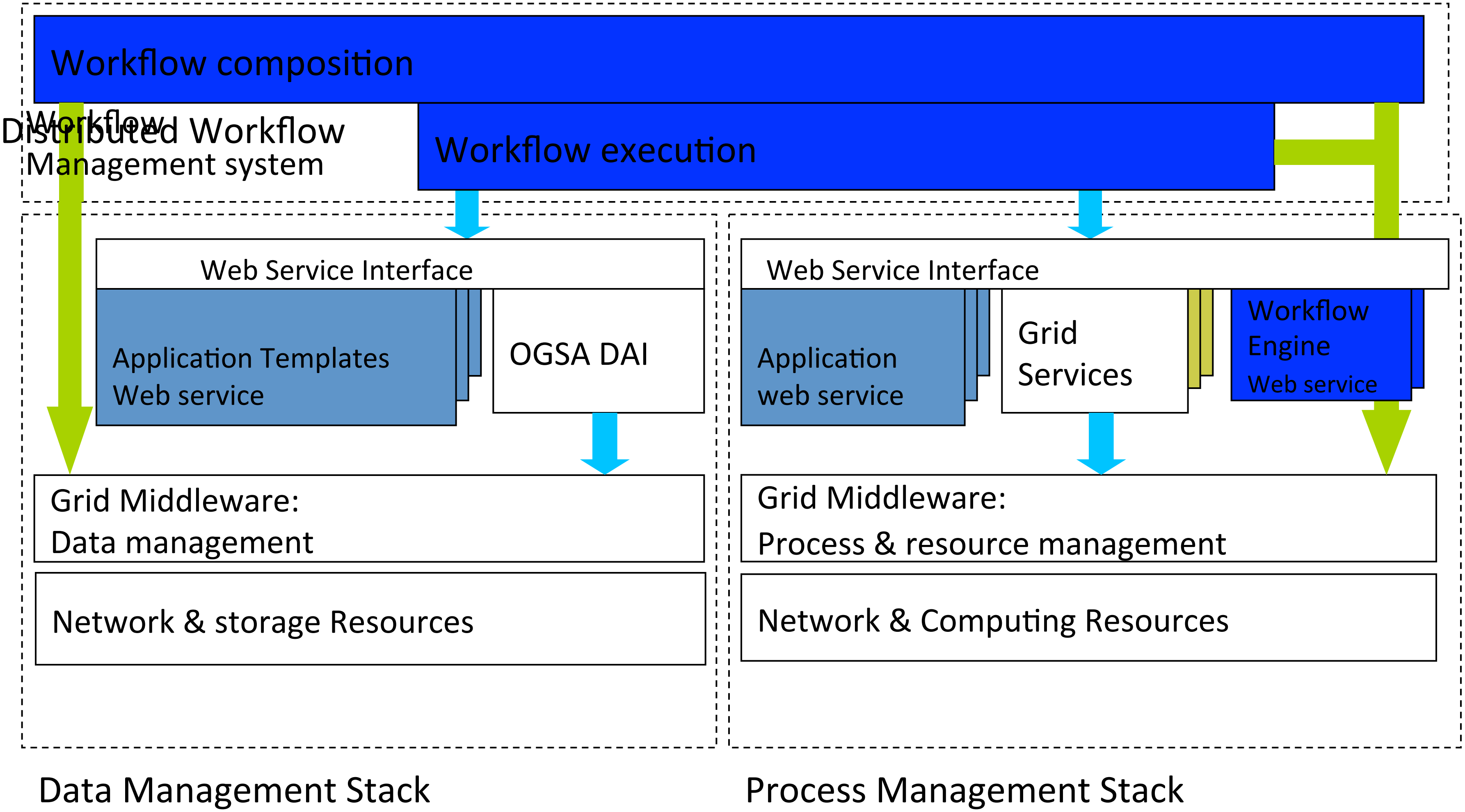

Gridftp host://f.a ....lumpy.isi.edu/
nfs/temp/f.a
lumpy.isi.edu://usr/local/
bin/extract
Jet.caltech.edu://home/malcom/
resample -I /home/malcolm/F.b1
F.c1
F.c2
Concat
Register /F.d at home/malcolm/f2
Data Transfer
Nodes
Replica
Catalog
Registration
Nodes

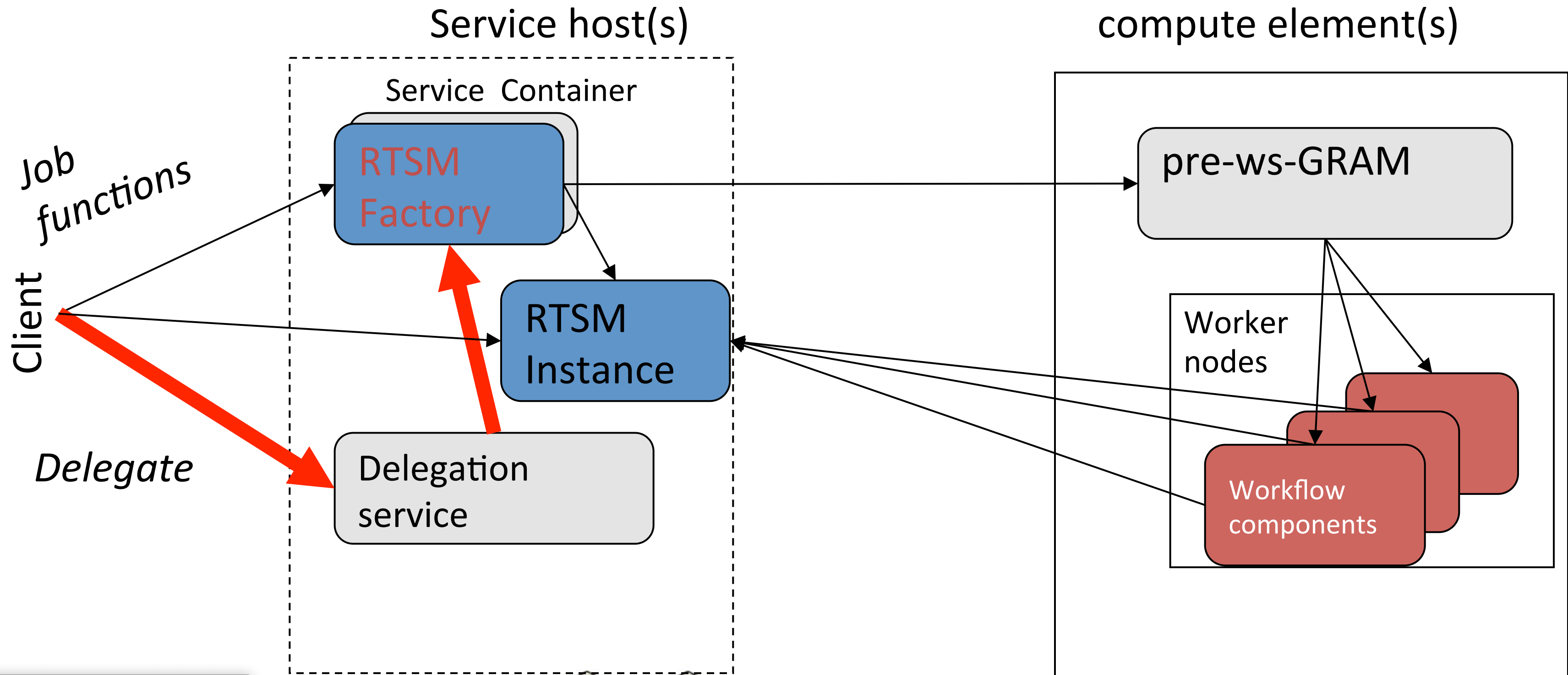
Figure 1.3. An example reduced, concrete workflow.

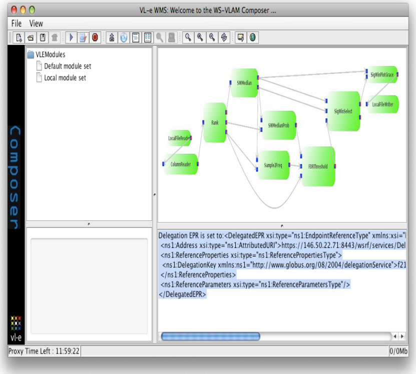
Distributed enabled workflow engines



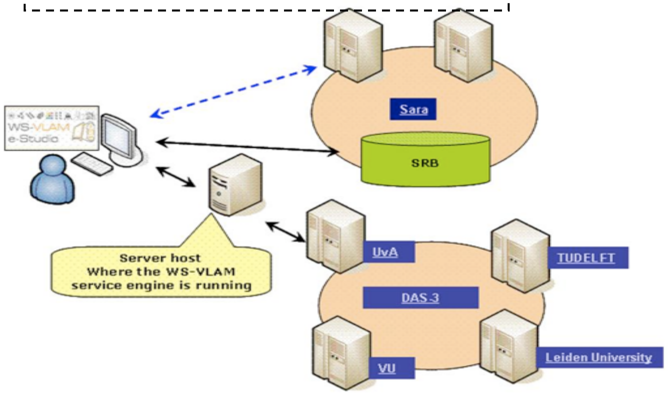
Bob Hertberger keynote talk at *2nd IEEE Conf on eScience & grid computing , Amsterdam 2006*

WS-VLAM

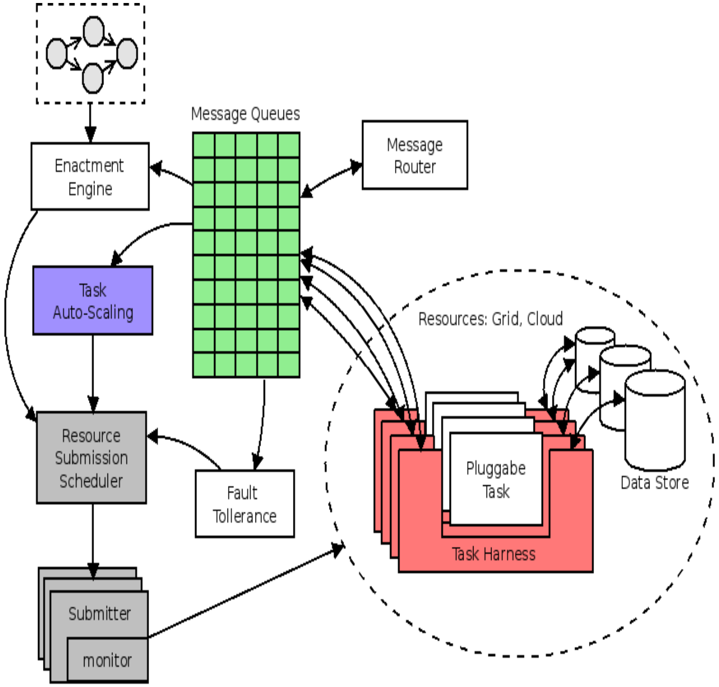


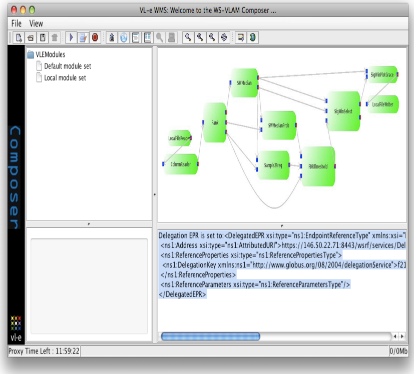


```
Delegation EPR is set to:<DelegatedEPR xsi:type="ns1:EndpointReferenceType" xmlns:xsi="
 <ns1:Address xsi:type="ns1:AttributedURI">https://146.50.22.71:8443/wsrf/services/Del
 <ns1:ReferenceProperties xsi:type="ns1:ReferencePropertiesType">
  <ns1:DelegationKey xmlns:ns1="http://www.globus.org/08/2004/delegationService">f21
 </ns1:ReferenceProperties>
 <ns1:ReferenceParameters xsi:type="ns1:ReferenceParametersType"/>
</DelegatedEPR>
```



WS-VLAM

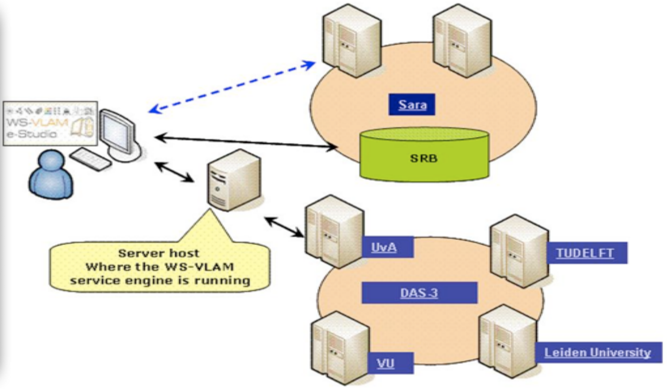



VL-e WMS: Welcome to the WS-VLAM Composer ...

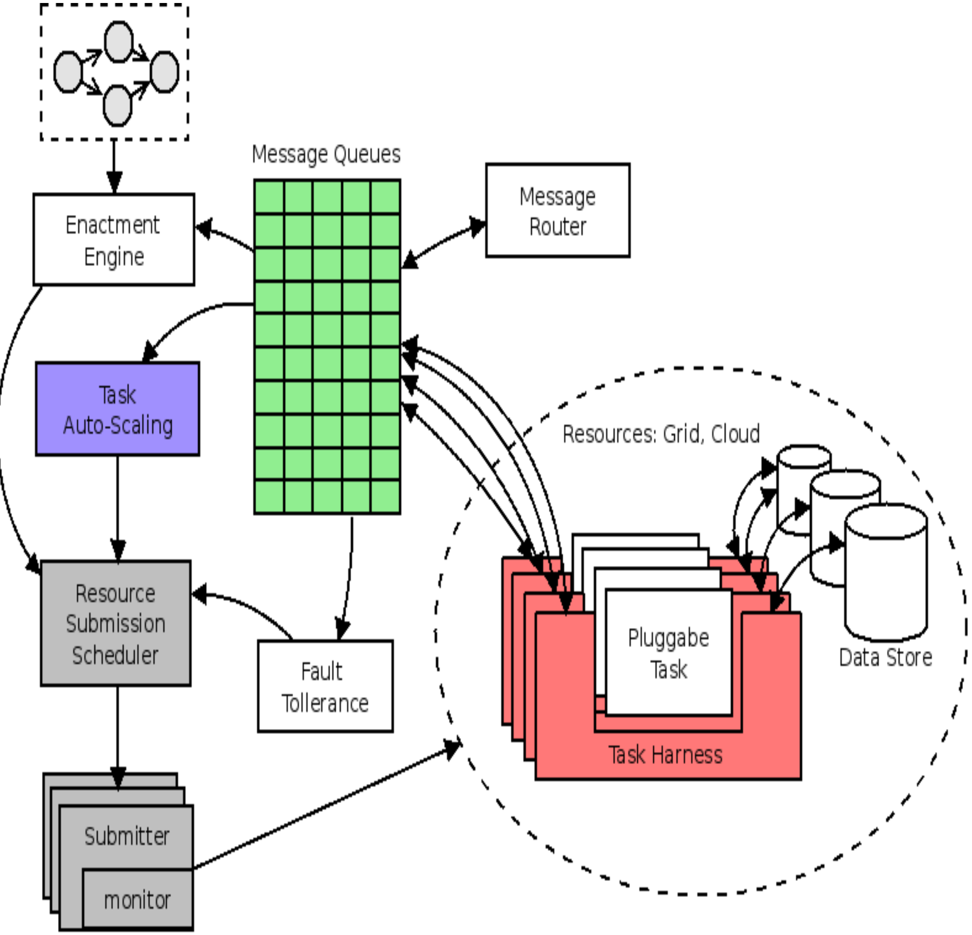
```
Delegation EPR is set to:<DelegatedEPR xsi:type="ns1:EndpointReferenceType" xmlns:xsi="
 <ns1:Address xsi:type="ns1:AttributedURI">https://146.50.22.71:8443/wsrf/services/Del
 <ns1:ReferenceProperties xsi:type="ns1:ReferencePropertiesType">
  <ns1:DelegationKey xmlns:ns1="http://www.globus.org/08/2004/delegationService">f21
 </ns1:ReferenceProperties>
 <ns1:ReferenceParameters xsi:type="ns1:ReferenceParametersType"/>
</DelegatedEPR>
```

Proxy Time Left : 11:59:22

0/0Mb



DataFluo engine



- Automatic scaling of workflow components based
  - Resource load
  - Application load
  - provenance data
- Scaling across various infrastructures
  - desktop
  - Grids
  - Clouds

Reginald Cushing, Spiros Koulouzis, Adam S. Z. Belloum, Marian Bubak, **Prediction-based Auto-scaling of Scientific Workflows**, Proceedings of the 9th International Workshop on Middleware for Grids, Clouds and e-Science, ACM/IFIP/USENIX December 12th, 2011, Lisbon, Portugal

Usage of Web Services in e-science

- WS offer interoperability and flexibility in a large scale distributed environment.
- WS can be **combined** in a **workflow** so that more complex operations may be achieved,
- but any workflow implementation is potentially **faced** with a *data transport problem*

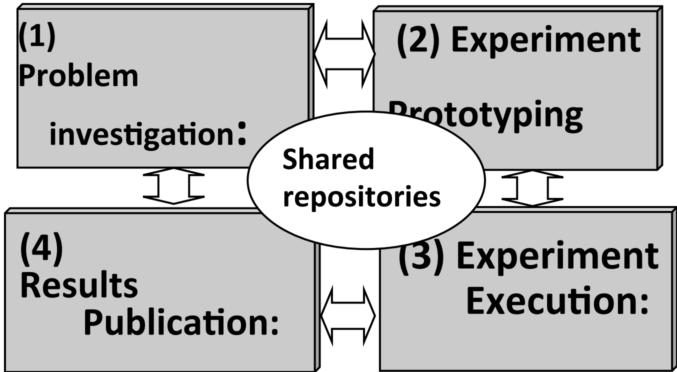
Resource management

- Within a single workflow services are **competing** for resources.
- Scaling one service without any regard to the whole workflow may starve parts of the workflow and hamper progress
- It would be ideal to have a mechanism to **greedily** consume resources if **no one** is using them but **donate back** resources once they are requested.

Fuzzy controller tries to do just that.

Outline

- Introduction
- Lifecycle of an e-science workflow
- Different approach to workflow scheduling
  - Workflow Process Modeling & Management In Grid/Cloud
  - Workflow and Web services (intrusive/non-intrusive)
- provenance


(1) Problem investigation:
(2) Experiment Prototyping
Shared repositories
(3) Experiment Execution:
(4) Results Publication:

Provenance/ reproducibility

- “A complete provenance record for a data object allows the possibility to reproduce the result and reproducibility is a critical component of the scientific method”
- Provenance: The recording of metadata and provenance information during the various stages of the workflow lifecycle

Workflows and e-Science: An overview of workflow system features
and capabilities Ewa Deelmana, Dennis Gannonb, Matthew Shields c, Ian Taylor, Future Generation Computer Systems 25 (2009)

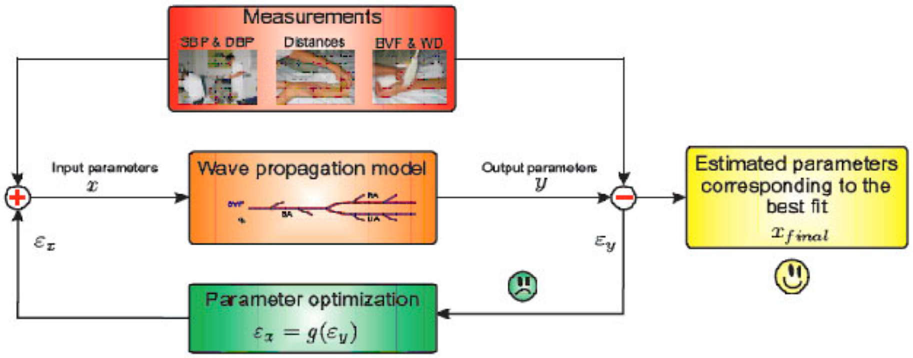
History-tracing XML (FH Aachen)

- provides data/process provenance following an approach that
  - maps the workflow graph to a layered structure of an XML document.
  - This allows an intuitive and easy processable representation of the workflow execution path,
  - which can be, eventually, electronically signed.

```
<patternMatch>
  <events>
    <PortResolved> provenance
data</PortResolved>
    <ConDone>provenance data
             </ConDone>
    ...
  </events>
  <fileReader2>
    <events> ... </events>
    <sign-fileReader2> ...
          </signfileReader2>
  </fileReader2>
  <sffToFasta>
    Reference
  </sffToFasta>
  <sign-patternMatch> ...
         </sign-patternMatch>
</patternMatch>
```

M. Gerards, Adam S. Z. Belloum, F. Berritz, V. Snder, S. Skorupa, **A History-tracing XML-base Proveannce Framework for workflows**, WORKS 2010, New Orleans, USA, November 2010

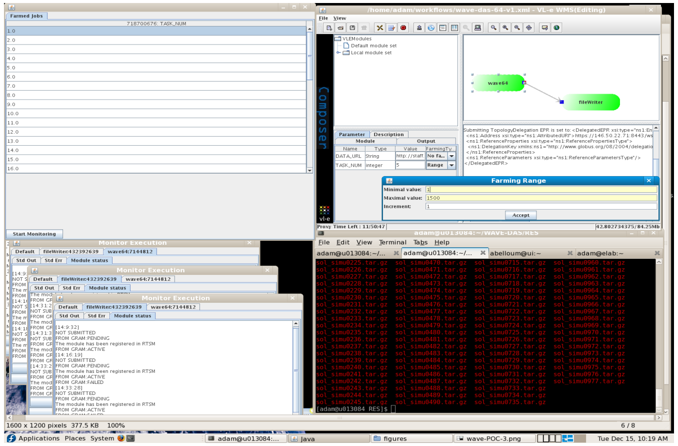
wave propagation model applications



[Biomedical engineering Cardiovascular biomechanics group TUE])

wave propagation model of blood flow in large vessels using an approximate velocity profile function:

a biomedical study for which **3000 runs** were required to perform a global sensitivity analysis of a blood pressure wave propagation in arteries



718700676: TASK_NUM
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
11.0
12.0
13.0
14.0
15.0
16.0

VLEModules
- Default module set
- Local module set

Module		Output	
Name	Type	Value	FarmingTy...
DATA_URL	String	http://staff.	No fa...
TASK_NUM	integer	5	Range

```
Submitting TopologyDelegation EPR is set to: <DelegatedEPR xsi:type="ns1:En
 <ns1:Address xsi:type="ns1:AttributedURI">https://146.50.22.71:8443/ws
 <ns1:ReferenceProperties xsi:type="ns1:ReferencePropertiesType">
  <ns1:DelegationKey xmlns:ns1="http://www.globus.org/08/2004/delegatio
 </ns1:ReferenceProperties>
 <ns1:ReferenceParameters xsi:type="ns1:ReferenceParametersType"/>
</DelegatedEPR>
```

Farming Range

Minimal value: 1

Maximal value: 1500

Increment: 1

Proxy Time Left : 11:50:47 42.802734375/84.25Mb

Monitor Execution

Default | fileWriter:432392639 | wave64:7144812

Std Out | Std Err | Module status

```
[14:9:32]
NOT SUBMITTED
FROM GRAM:PENDING
The module has been registered in RTSM
FROM GRAM:ACTIVE
[14:16:19]
NOT SUBMITTED
FROM GRAM:PENDING
The module has been registered in RTSM
FROM GRAM:ACTIVE
FROM GRAM:FAILED
[14:33:28]
NOT SUBMITTED
FROM GRAM:PENDING
The module has been registered in RTSM
FROM GRAM:ACTIVE
```

adam@u013084:~/WAVE-DAS/RES

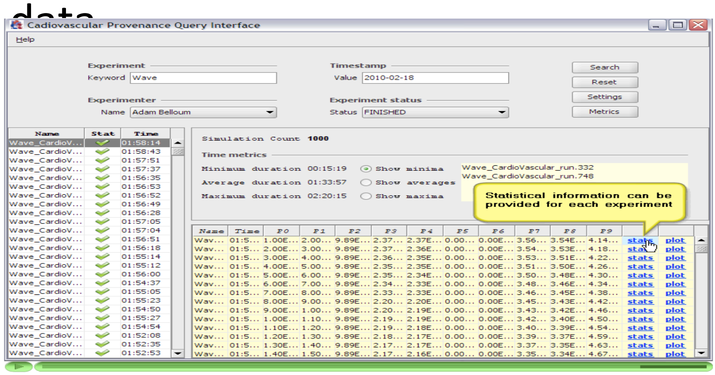
adam@u013084:~/... | adam@u013084:~/... | abelloum@ui:~ | adam@elab:~

```
sol_simu0225.tar.gz  sol_simu0470.tar.gz  sol_simu0715.tar.gz  sol_simu0960.tar.gz
sol_simu0226.tar.gz  sol_simu0471.tar.gz  sol_simu0716.tar.gz  sol_simu0961.tar.gz
sol_simu0227.tar.gz  sol_simu0472.tar.gz  sol_simu0717.tar.gz  sol_simu0962.tar.gz
sol_simu0228.tar.gz  sol_simu0473.tar.gz  sol_simu0718.tar.gz  sol_simu0963.tar.gz
sol_simu0229.tar.gz  sol_simu0474.tar.gz  sol_simu0719.tar.gz  sol_simu0964.tar.gz
sol_simu0230.tar.gz  sol_simu0475.tar.gz  sol_simu0720.tar.gz  sol_simu0965.tar.gz
sol_simu0231.tar.gz  sol_simu0476.tar.gz  sol_simu0721.tar.gz  sol_simu0966.tar.gz
sol_simu0232.tar.gz  sol_simu0477.tar.gz  sol_simu0722.tar.gz  sol_simu0967.tar.gz
sol_simu0233.tar.gz  sol_simu0478.tar.gz  sol_simu0723.tar.gz  sol_simu0968.tar.gz
sol_simu0234.tar.gz  sol_simu0479.tar.gz  sol_simu0724.tar.gz  sol_simu0969.tar.gz
sol_simu0235.tar.gz  sol_simu0480.tar.gz  sol_simu0725.tar.gz  sol_simu0970.tar.gz
sol_simu0236.tar.gz  sol_simu0481.tar.gz  sol_simu0726.tar.gz  sol_simu0971.tar.gz
sol_simu0237.tar.gz  sol_simu0482.tar.gz  sol_simu0727.tar.gz  sol_simu0972.tar.gz
sol_simu0238.tar.gz  sol_simu0483.tar.gz  sol_simu0728.tar.gz  sol_simu0973.tar.gz
sol_simu0239.tar.gz  sol_simu0484.tar.gz  sol_simu0729.tar.gz  sol_simu0974.tar.gz
sol_simu0240.tar.gz  sol_simu0485.tar.gz  sol_simu0730.tar.gz  sol_simu0975.tar.gz
sol_simu0241.tar.gz  sol_simu0486.tar.gz  sol_simu0731.tar.gz  sol_simu0976.tar.gz
sol_simu0242.tar.gz  sol_simu0487.tar.gz  sol_simu0732.tar.gz  sol_simu0977.tar.gz
sol_simu0243.tar.gz  sol_simu0488.tar.gz  sol_simu0733.tar.gz
sol_simu0244.tar.gz  sol_simu0489.tar.gz  sol_simu0734.tar.gz
sol_simu0245.tar.gz  sol_simu0490.tar.gz  sol_simu0735.tar.gz
[adam@u013084 RES]$ 
```

1600 x 1200 pixels 377.5 KB 100% 6 / 8

adam@u013084:... Tue Dec 15, 10:19 AM

User Interface to compose workflow (top right), monitor the execution of the farmed workflows (top left), and monitor each run separately (bottom left) data



Experiment

Keyword: Wave

Timestamp

Value: 2010-02-18

Experimenter

Name: Adam Belloum

Experiment status

Status: FINISHED

Name	Stat	Time
Wave_CardioV...	✓	01:58:14
Wave_CardioV...	✓	01:58:43
Wave_CardioV...	✓	01:57:51
Wave_CardioV...	✓	01:57:37
Wave_CardioV...	✓	01:56:35
Wave_CardioV...	✓	01:56:53
Wave_CardioV...	✓	01:56:52
Wave_CardioV...	✓	01:56:49
Wave_CardioV...	✓	01:56:28
Wave_CardioV...	✓	01:57:05
Wave_CardioV...	✓	01:57:04
Wave_CardioV...	✓	01:56:51
Wave_CardioV...	✓	01:56:18
Wave_CardioV...	✓	01:55:14
Wave_CardioV...	✓	01:55:12
Wave_CardioV...	✓	01:56:00
Wave_CardioV...	✓	01:54:37
Wave_CardioV...	✓	01:55:05
Wave_CardioV...	✓	01:55:23
Wave_CardioV...	✓	01:54:50
Wave_CardioV...	✓	01:55:27
Wave_CardioV...	✓	01:54:54
Wave_CardioV...	✓	01:52:08
Wave_CardioV...	✓	01:52:35
Wave_CardioV...	✓	01:52:53

Simulation Count **1000**

Time metrics

Minimum duration 00:15:19 ● Show minima

Average duration 01:33:57 ○ Show averages

Maximum duration 02:20:15 ○ Show maxima

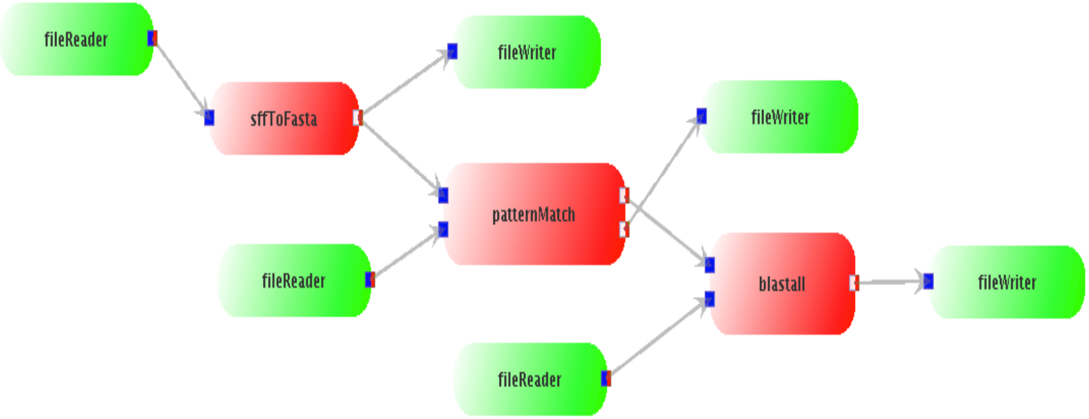
Wave_CardioVascular_run.332
Wave_CardioVascular_run.748

Statistical information can be provided for each experiment

Name	Time	P0	P1	P2	P3	P4	P5	P6	P7	P8	P9		
Wav...	01:5...	1.00E...	2.00...	9.89E...	2.37...	2.37E...	0.00...	0.00E...	3.56...	3.54E...	4.14...	stats	plot
Wav...	01:5...	2.00E...	3.00...	9.89E...	2.37...	2.36E...	0.00...	0.00E...	3.54...	3.53E...	4.18...	stats	plot
Wav...	01:5...	3.00E...	4.00...	9.89E...	2.36...	2.35E...	0.00...	0.00E...	3.53...	3.51E...	4.22...	stats	plot
Wav...	01:5...	4.00E...	5.00...	9.89E...	2.35...	2.35E...	0.00...	0.00E...	3.51...	3.50E...	4.26...	stats	plot
Wav...	01:5...	5.00E...	6.00...	9.89E...	2.35...	2.34E...	0.00...	0.00E...	3.50...	3.48E...	4.30...	stats	plot
Wav...	01:5...	6.00E...	7.00...	9.89E...	2.34...	2.33E...	0.00...	0.00E...	3.48...	3.46E...	4.34...	stats	plot
Wav...	01:5...	7.00E...	8.00...	9.89E...	2.33...	2.33E...	0.00...	0.00E...	3.46...	3.45E...	4.38...	stats	plot
Wav...	01:5...	8.00E...	9.00...	9.89E...	2.20...	2.20E...	0.00...	0.00E...	3.45...	3.43E...	4.42...	stats	plot
Wav...	01:5...	9.00E...	1.00...	9.89E...	2.20...	2.19E...	0.00...	0.00E...	3.43...	3.42E...	4.46...	stats	plot
Wav...	01:5...	1.00E...	1.10...	9.89E...	2.19...	2.19E...	0.00...	0.00E...	3.42...	3.40E...	4.50...	stats	plot
Wav...	01:5...	1.10E...	1.20...	9.89E...	2.19...	2.18E...	0.00...	0.00E...	3.40...	3.39E...	4.54...	stats	plot
Wav...	01:5...	1.20E...	1.30...	9.89E...	2.18...	2.17E...	0.00...	0.00E...	3.39...	3.37E...	4.59...	stats	plot
Wav...	01:5...	1.30E...	1.40...	9.89E...	2.17...	2.17E...	0.00...	0.00E...	3.37...	3.35E...	4.63...	stats	plot
Wav...	01:5...	1.40E...	1.50...	9.89E...	2.17...	2.16E...	0.00...	0.00E...	3.35...	3.34E...	4.67...	stats	plot

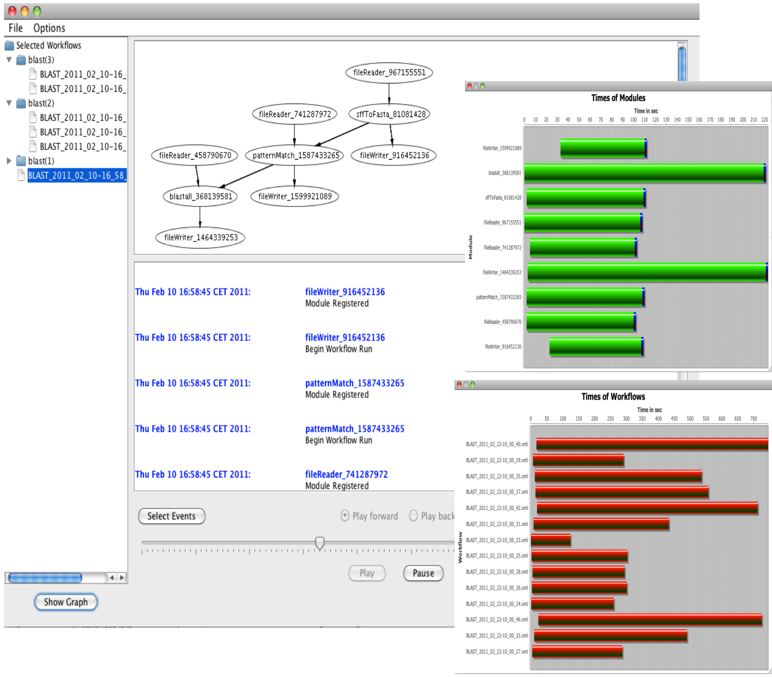
Query interface for the provenance data collected from 3000 simulations of the "*wave propagation model of blood flow in large vessels using an approximate velocity profile function*"

Blast Application



[Department of Clinical Epidemiology, Biostatistics and Bioinformatics (KEBB), AMC]

The aim of the application is the **alignment of DNA sequence** data with a given reference database. A workflow approach is currently followed to run this application on distributed computing resources.



Thu Feb 10 16:58:45 CET 2011: fileWriter_916452136
Module Registered

Thu Feb 10 16:58:45 CET 2011: fileWriter_916452136
Begin Workflow Run

Thu Feb 10 16:58:45 CET 2011: patternMatch_1587433265
Module Registered

Thu Feb 10 16:58:45 CET 2011: patternMatch_1587433265
Begin Workflow Run

Thu Feb 10 16:58:45 CET 2011: fileReader_741287972
Module Registered

For Each workflow run

The provenance data is collected an stored following the XML-tracing system

User interface allows to reproduce events that occurred at runtime (replay mode)

User Interface can be customized (User can select the events to track)

User Interface show resource usage

on-going work UvA-AMC-fh-aachen

More References

1. A.S.Z. Belloum, V. Korkhov, S Koulouzis, M. A Inda, and M. Bubak *Collaborative e-Science experiments: from scientific workflow to knowledge sharing* JULY/AUGUST, IEEE Internet Computing, 2011
2. Ilkay Altintas, Manish Kumar Anand, Daniel Crawl, Shawn Bowers, Adam Belloum, Paolo Missier, Bertram Ludascher, Carole A. Goble, Peter M.A. Sloot, Understanding Collaborative Studies Through Interoperable Workflow Provenance, IPAW2010, Troy, NY, USA
3. A. Belloum, Z. Zhao, and M. Bubak Workflow systems and applications , Future Generation Comp. Syst. 25 (5): 525-527 (2009)
4. Z. Zhao, A.S.Z. Belloum, et al., Distributed execution of aggregated multi domain workflows using an agent framework The 1st IEEE International Workshop on Scientific Workflows, Salt Lake City, U.SA, 2007
5. Zhiming Zhao, Adam Belloum, Cees De Laat, Pieter Adriaans, Bob Hertzberger Using Jade agent framework to prototype an e-Science workflow bus Authors Cluster Computing and the Grid, 2007. CCGRID 2007

Summary

- Workflow research especially in the grid environments are rapidly growing research subject
- VOs in Grid can benefits from the experience of workflows in the business community
- Scientific Workflow in Grid Environment have their own characteristics that need to be dealt with new approach
- Scientific Workflow research is highly related with various other research topics: resource management, fault tolerance, application performance, ontology.