First Brazil-France Workshop
On High Performance Computing and Scientific Data Management Driven by Highly Demanding Applications
COPPE UFRJ

EXPLORING PROVENANCE IN HIGH PERFORMANCE SCIENTIFIC COMPUTING

Marta Mattoso
COPPE/Federal University of Rio de Janeiro

Group 2 at COPPE/UFRJ

- Marta Mattoso (P.I. Distributed Database Group)
- Post-Docs:
  - ✓ Kary Ocana, D.Sc 2010 - Bioinformatics
  - ✓ Eduardo Ogasawara, DSc 2011 - Workflow engine
  - ✓ Daniel Oliveira, DSc 2012 - Cloud Workflows
- DSc Students:
  - ✓ Jonas Dias
  - ✓ Flavio Costa

Group 2 at COPPE/UFRJ

- **On going Collaborations** (now also in HOSCAR)
- INRIA:
  - ✓ Patrick Valduriez
    - CAPES, CNPq, FAPERJ and INRIA Equipes Associées
- LNCC:
  - ✓ Fabio Porto
  - ✓ Luiz Gadelha Jr

National Laboratory Scientific Computing

Group 2 at COPPE/UFRJ

- **On going Collaborations** (now also in HOSCAR – GROUP 1)
- COPPE- Numerical Methods:
  - ✓ Alvaro Coutinho
    - Renato Elias
    - José Camata
  - ✓ Fernando Rochinha
    - Gabriel
  - ✓ Luiz Landau
    - Josias

NACAD High Performance Computing Center

Group 2 at COPPE/UFRJ

Other International Collaborations (US)

- U.Chicago/ Argonne:
  - ✓ Ian Foster
  - ✓ Mike Wilde
- New York - Polytechnic Institute of NYU:
  - ✓ Juliana Freire

HOSCAR - Objectives

- "…taking full benefits of the processing capabilities of future high performance massively parallel architectures in the framework of very large-scale datasets and numerical simulations…"
- Group 3 (parallel architectures): basis for everyone
-
- Intergroup collaboration Group 1 (numerical simulations) and Group 2 (large-scale datasets)

HOSCAR - Objectives

- "…taking full benefits of the processing capabilities of future high performance massively parallel architectures in the framework of very large-scale datasets and numerical simulations…"
- Intergroup collaboration Group 1 (numerical simulations) and Group 2 (large-scale datasets)
- Why, how to collaborate ?
  - ✓ Some of our (large datasets) previous experiences collaborating with (simulations) COPPE& Petrobras
  - ✓ Manage scientific resources (files, programs, images)

1st Workshop on High-Performance Computing Meets Databases

SC11 Seattle, WA

Exploring Provenance in High Performance Scientific Computing

Luiz Gadelha,
Marta Mattoso
COPPE, Federal Univ Rio de Janeiro, Brazil

Michael Wilde,
Ian Foster
CI, U. Chicago / ANL, USA

COPPE UFRJ — THE UNIVERSITY OF CHICAGO — Argonne NATIONAL LABORATORY

1st International Workshop on Scalable Workflow Enactment Engines and Technologies

SWEET'12

Colocated with SIGMOD'12

EVALUATING PARAMETER SWEEP WORKFLOWS IN HIGH PERFORMANCE COMPUTING

Fernando Chirigati*, Vítor Silva, Eduardo Ogasawara, Daniel de Oliveira, Jonas Dias, Fábio Porto, Patrick Valduriez and Marta Mattoso

COPPE / Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

LNCC, Rio de Janeiro, Brazil

INRIA & LIRMM, Montpellier, France

* Currently at Polytechnic Institute of NYU

CCSA 2012 @ CCGrid

2nd Cloud Computing and Scientific Applications (CCSA)

Cloud-based Phylogenomic Inference of Evolutionary Relationships: A Performance Study*

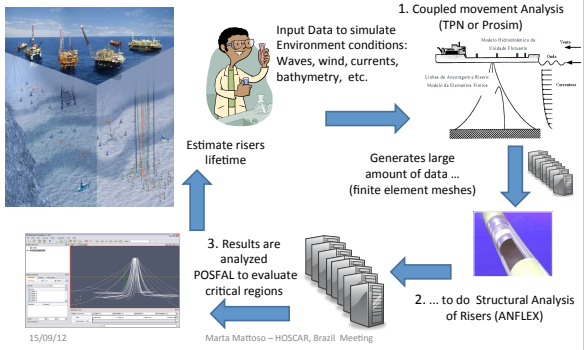
Daniel de Oliveira, Kary Ocaña, Eduardo Ogasawara, Jonas Dias João Gonçalves and Marta Mattoso

Federal University of Rio de Janeiro
Rio de Janeiro, Brazil

* Best paper award

COPPE UFRJ

Risers' Fatigue Analysis in Ultra-Deep Waters



HPC problems potentially amenable to database techniques

COPPE UFRJ

- as HPC computations become longer-running, there is a need to make them more interactive.
- Ad hoc science questions such as "Which regions generated by experiment X and parameter Y exhibit critic above some threshold?"
- Ad hoc monitoring questions such as "Which regions should I visualize?"

- Modeled, expressed, and optimized as queries
- Using Provenance / Scientific Workflow approach

Scientific Computing with Workflows

- Scientific experiments are often defined by a composition of many computational tasks linked through data flow
- These are also called *scientific workflows*
- Data sets manipulated can be large in size and quantity
- Analyzing the outcome of these large-scale computations can be difficult
- Provenance support is the key

Scientific Workflow Management Systems

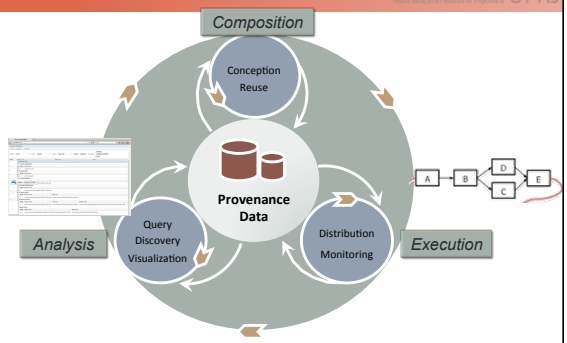
- Native parallel execution
  Ex.: Swift, Pegasus, Triana
- Desktop serial execution
  Ex.: Kepler, VisTrails, Taverna
- Coupled Parallel/Serial execution
  Ex.: VisTrails+Hadoop, Kepler+Hadoop

Scientific Workflow Management Systems

- Native parallel execution
  Ex.: Swift, Pegasus, Triana
- Desktop serial execution
  Ex.: Kepler, VisTrails, Taverna
- Coupled Parallel/Serial execution
  Ex.: VisTrails+Hadoop, Kepler+Hadoop

Provenance is only available after the execution

Our view of workflows

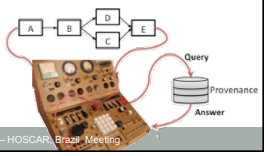


Data explorations with Wf

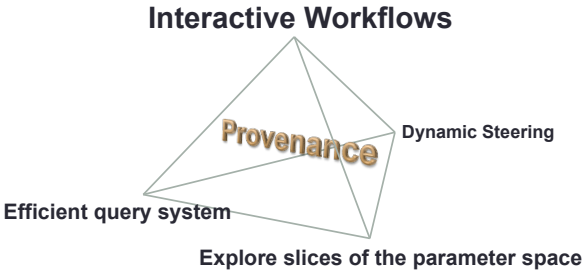
- The execution of variations upon the same workflow repeatedly
  - ✓ Parameter Sweeps
  - ✓ Fine-tuning (tolerance, solver options, …)
  - ✓ Iterative methods
    - Stop criteria defined by the scientist evaluation
- Interfere in the dataflow
  - ✓ Input parameters
  - ✓ Workflow configurations
- Dynamic nature of simulations/ large data

Objectives

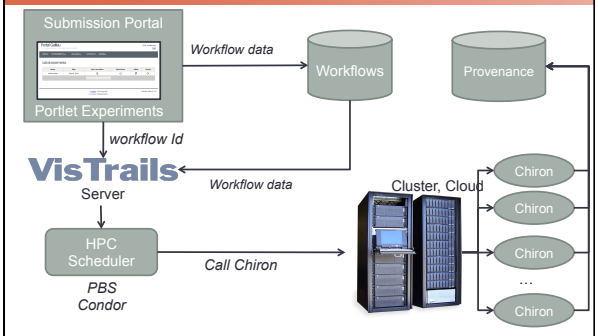
- Interactive Workflows
- Adaption and user-steering on workflow execution
  - ✓ Parameter space slicing
    - To compare results from different explored slices
  - ✓ Workflow adjustments
    - Human intervention
    - Dynamic steering
- Improvements
  - ✓ Quality of results
  - ✓ Execution efficiency
  - ✓ Scientific Data analysis



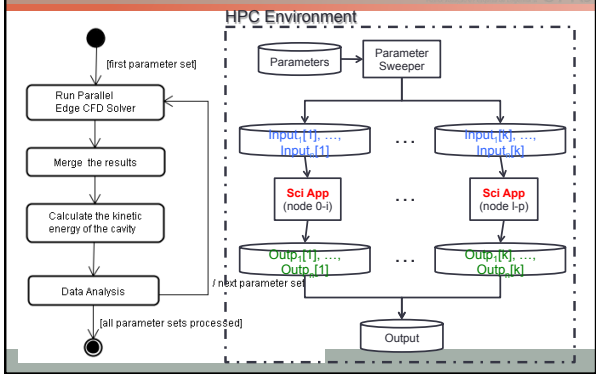
Using provenance



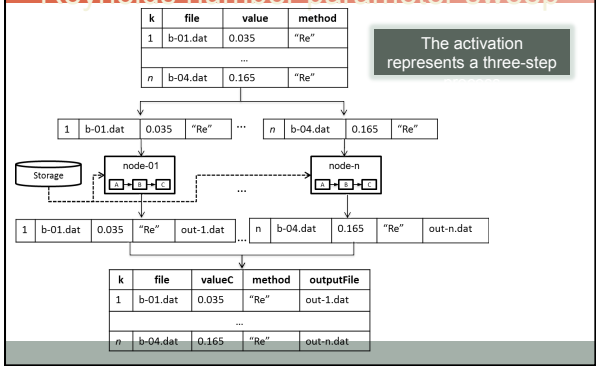
Chiron Workflow Execution



Chiron Support for Parameter Sweep



Activation in Chiron: Example of a Reynolds number parameter sweep

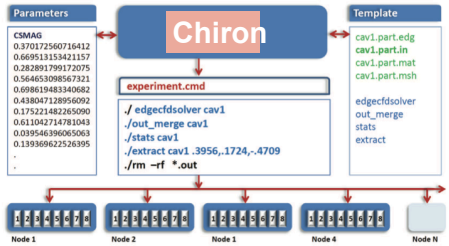


Provenance querying

```
SELECT t.workspace
FROM Task t, ParameterValue pv,Parameter p
WHERE p.ParameterID = pv.ParameterID
AND pv.TaskID = t.TaskID
AND p.ParameterName = 'Reynolds_Number'
AND  pv.NumericValue > 0.090
AND  pv.NumericValue < 0.110;
```

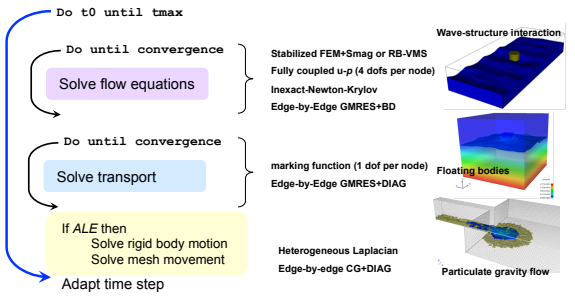
Provenance DB:
- files consumed and produced, standard output and standard error,
- execution time and the computer node that executed the program,
- used in real time to verify if any deterministic solver failed

Large Eddy Simulation Workflow Parallel Execution in Chiron

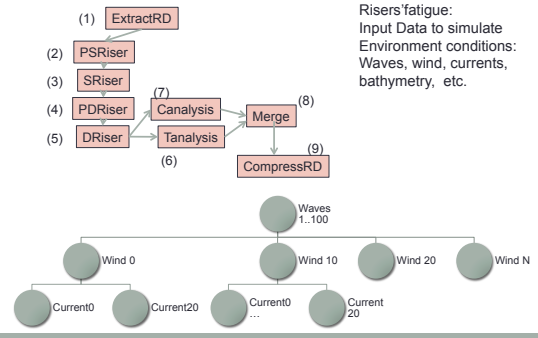


- Provenance gathers data to be used to evaluate statistical moments
- We can track the error between MC and SGC and determine an interpolation level to satisfy a minimum error
- Provenance capabilities in Chiron can automatically increase the interpolation level and resubmit the experiment to obtain statistical moments with a prefixed error

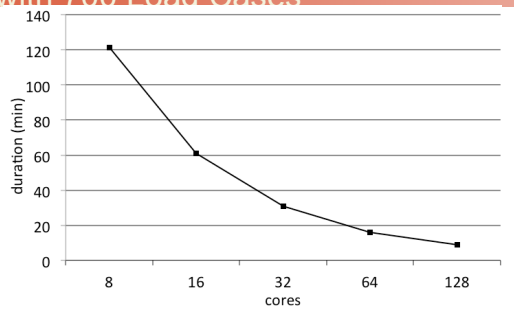
EdgeCFD Flow Solver



Exploring Slices Parameter

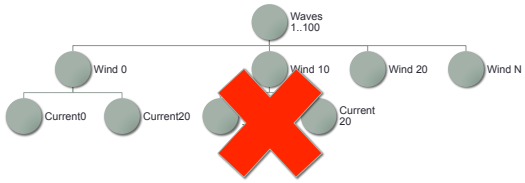


Evaluation of RFA Workflow with 700 Load Cases



1438 activations, 16765 files

Exploring Slices Parameter



Suppose, after some analysis, there is no need to evaluate slice Wind 10 → less files to generate

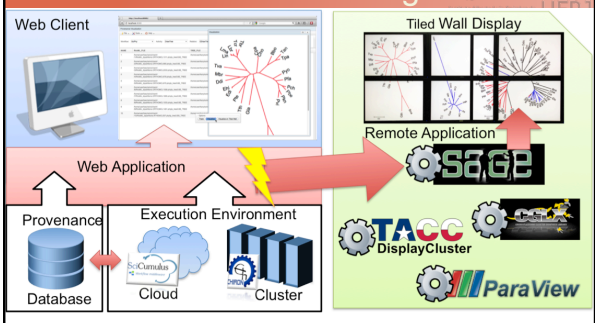
Large-Scale Provenance-enriched Visualization

1. Visualizes experiment data through the workflow
   - ✓ Displays provenance data for information browsing
   - ✓ Provenance data collected by **Chiron [5]** or **SciCumulus [6]** workflow engines
2. Allows for selections and filtering
   - ✓ **Let scientists select just the results that interest them**
3. Stage-out only the data the scientists selected
4. Tracks which input parameters produced a given result
5. Display the results locally or on a display environment
   - ✓ Integration with tiled wall displays technologies
   - ✓ Currently implemented to use SAGE [7]
   - ✓ Can be extended to other platforms such as CGLX or DisplayCluster

Visualization needs in large-scale

1. Go fast to the results you need to analyze
   - ✓ Easy navigation over workflow data and provenance data
2. Filter results and stage out only the data you need to analyze
   - ✓ Staging data out can be costly (remote clusters, clouds)
3. Use high-resolution display environments
   - ✓ Visualize multiple results to establish comparisons
     - On parameters exploration scenarios
   - ✓ Analyze highly detailed images and simulations
   - ✓ **Can take advantage of available tiled display technologies such as SAGE, CGLX, TACC DisplayCluster and Paraview**

Visualization needs in large-scale

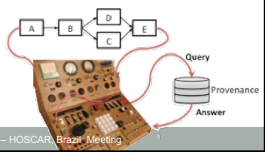


Data explorations with Wf

- The execution of variations upon the same workflow repeatedly
  - ✓ Parameter Sweeps
  - ✓ Fine-tuning (tolerance, solver options, …)
  - ✓ Iterative methods
    - Stop criteria defined by the scientist evaluation
- Interfere in the dataflow
  - ✓ Input parameters
  - ✓ Workflow configurations
- Dynamic nature of simulations/ large data

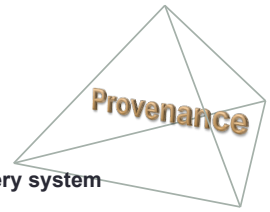
Objectives

- Interactive Workflows
- Adaption and user-steering on workflow execution
  - ✓ Parameter space slicing
    - To compare results from different explored slices
  - ✓ Workflow adjustments
    - Human intervention
    - Dynamic steering
- Improvements
  - ✓ Quality of results
  - ✓ Execution efficiency
  - ✓ Scientific Data analysis



Using provenance

Interactive Workflows



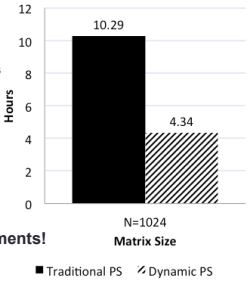
Dynamic Steering

Efficient query system

Explore slices of the parameter space

Dynamic Experiments

- Optimization problems
  - ✓ Genetic Algorithms
- Uncertainty Quantification
  - ✓ Adaptive sparse grid colocation methods
- Reduced order models
  - ✓ Numerical solutions for heat conduction
- Bioinformatics
  - ✓ Comparative genomics
  - ✓ Phylogeny
- **Good results on preliminary experiments!**
  - ✓ **(Dias et al., WORKS at SC 2011)**



Parameter Space Slicing

- Workflow executes choosing different slices of the parameter space at run time
  - ✓ To take initial decisions
  - ✓ To compare results
- We can make the slices addressable
- Parameter Space = Relation
  - ✓ Slices can be represented as horizontal fragments

$$R_i = \sigma_{F_i}(R), 1 \leq i \leq w$$

where F_i is the fragmentation predicate or a minterm

Example

R

IID	ModelFile	Flexibility	Curvature
1	I-01	3.2	0.8
2	I-02	4.5	0.9
3	I-03	5.6	1.2
4	I-04	5.2	1.1

R_1, R_2

T ← SlicedMap(ga, M, R)

T

IID	Mãe	Pai	ModelFile	Flexibility	Curvature
1	2	2	I-02M	3.1	0.5
2	1	1	I-01M	5.1	1.4
3	3	3	I-03	5.6	1.2
4	4	4	I-04	4.9	1.0
5	3	4	I-03x04	6.1	1.5

T_1, T_2

$M = \{ m_1, m_2 \mid$
m_1 : *Curvature* < 1.0,
m_2 : *Curvature* ≥ 1.0 $\}$

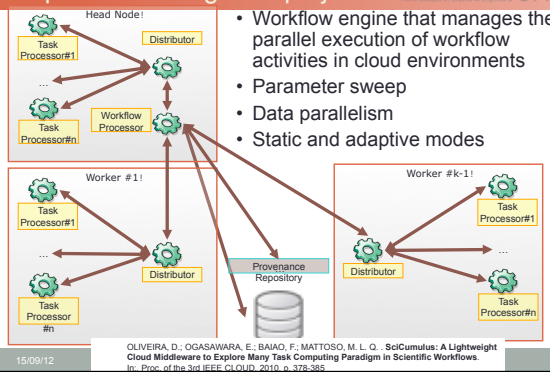
Evaluating our approach

- Real experiments as **Algebraic Workflows**
  - ✓ Bioinformatics experiments
    - Working with the biochemist Kary Ocaña
  - ✓ Numerical Simulations
    - Partnership with prof. Alvaro Coutinho's team
  - ✓ Uncertainty Quantification
    - Partnership with Prof. Fernando Rochinha's team
  - ✓ Oil and Gas experiments
    - Petrobras
  - ✓ Characterization of PS Workflows - Evaluation Framework
    - PS workflow patterns • workload configuration • performance metrics

Scientific Workflows in Clouds

- Many scientists are migrating their experiments to clouds, including bioinformatics ones
- They are not required to assemble expensive computational infrastructure to execute their experiments
- They do not need to configure many pieces of software
- However, clouds are changing environments and they may be susceptible to performance fluctuations during the execution course of the workflow

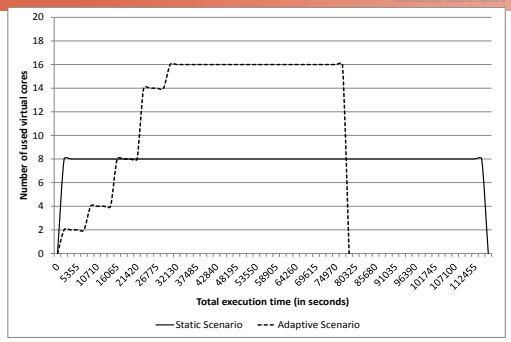
SciCumulus Workflow Engine
https://sourceforge.net/projects/scicumulus/



- Workflow engine that manages the parallel execution of workflow activities in cloud environments
- Parameter sweep
- Data parallelism
- Static and adaptive modes

OLIVEIRA, D.; OGASAWARA, E.; BAIAO, F.; MATTOSO, M. L. Q. . **SciCumulus: A Lightweight Cloud Middleware to Explore Many Task Computing Paradigm in Scientific Workflows**. In: Proc. of the 3rd IEEE CLOUD, 2010. p. 378-385

Analysis of the Adaptive Execution



Querying provenance data

- "Get *workflow* ,executions, activity names that executed without error

```
SELECT w.tag,
       a.tag,
       t.exitstatus,
       t.processor,
       t.workspace,
       t.status,
       t.endtime,
       t.starttime,
       extract ('epoch' from (t.endtime-t.starttime))|| ' ' as duration
from hworkflow w, hactivity a, htask t
where w.wkfid = a.wkfid
and a.actid = t.actid
and not exists (select * from htask a2
                where a2.actid = a.actid
                and a2.exitstatus = 0)
order by w.wkfid
```

NaffAdaptive	naff5	0	1	/root/exp/naff5/47/	FINISHED	2011-08-16 14:28:47.54+03	2011-08-16 14:28:42.157+03	5.201,
NaffAdaptive	naff5	0	1	/root/exp/naff5/62/	FINISHED	2011-08-16 14:29:27.54+03	2011-08-16 14:29:22.013+03	4.308,
NaffAdaptive	naff5	0	2	/root/exp/naff5/77/	FINISHED	2011-08-16 14:30:03.497+03	2011-08-16 14:30:00.82+03	4.677,
NaffAdaptive	naff5	0	1	/root/exp/naff5/115/	FINISHED	2011-08-16 14:31:19.211+03	2011-08-16 14:31:13.874+03	5.377,
NaffAdaptive	naff5	0	2	/root/exp/naff5/120/	FINISHED	2011-08-16 14:32:02.499+03	2011-08-16 14:31:55.972+03	6.467,
NaffAdaptive	naff5	0	1	/root/exp/naff5/136/	FINISHED	2011-08-16 14:32:48.431+03	2011-08-16 14:32:43.829+03	4.602,
NaffAdaptive	naff5	0	2	/root/exp/naff5/149/	FINISHED	2011-08-16 14:33:26.907+03	2011-08-16 14:33:21.177+03	5.73,
NaffAdaptive	naff5	0	2	/root/exp/naff5/164/	FINISHED	2011-08-16 14:34:10.702+03	2011-08-16 14:34:06.119+03	4.583,
NaffAdaptive	naff5	0	1	/root/exp/naff5/186/	FINISHED	2011-08-16 14:35:06.452+03	2011-08-16 14:35:01.38+03	5.072,
NaffAdaptive	naff5	0	2	/root/exp/naff5/200/	FINISHED	2011-08-16 14:35:42.759+03	2011-08-16 14:35:38.148+03	4.611,
NaffAdaptive	naff5	0	1	/root/exp/naff5/1/	FINISHED	2011-08-16 14:26:21.113+03	2011-08-16 14:26:09.956+03	11.157,
NaffAdaptive	naff5	0	1	/root/exp/naff5/3/	FINISHED	2011-08-16 14:26:30.246+03	2011-08-16 14:26:21.317+03	8.928,
NaffAdaptive	naff5	0	2	/root/exp/naff5/2/	FINISHED	2011-08-16 14:26:39.512+03	2011-08-16 14:26:09.957+03	29.555,
NaffAdaptive	naff5	0	1	/root/exp/naff5/4/	FINISHED	2011-08-16 14:26:40.491+03	2011-08-16 14:26:30.45+03	10.041,

Other real Bio Experiments... (one more in E-Science 2012)

Adaptive Execution and cost model

OLIVEIRA, D., Ocaña, K. A. C. S., Ogasawara, E., Dias, J., Goncalves, J., Mattoso, M., (2012), "**Cloud-based Phylogenomic Inference of Evolutionary Relationships: A Performance Study**". 2nd International Workshop on Cloud Computing and Scientific Applications (CCSA), Ottawa.

SciPhylomics
Phylogenomic analysis

Parallel execution with blocking activities

Ocaña, K. A. C. S., OLIVEIRA, D. de, Horta, F., Dias, J., Ogasawara, E., Mattoso, M., (2012), "Exploring Molecular Evolution Reconstruction Using a Parallel Cloud-based Scientific Workflow". Brazilian Symposium on Bioinformatics (BSB 2012), Campo Grande, MT.

SciEvol
Evolutionary analysis

Adaptive Execution

OCANA, K. ; OLIVEIRA, D. ; DIAS, J. ; OGASAWARA, E. ; MATTOSO, M. L. Q. . Optimizing Phylogenetic Analysis Using SciHmm Cloud-based Scientific Workflow. 7th IEEE e Science conference. IEEE Computer Society, 2011.

SciHmm
Comparative genomics

Conclusions

- Exploratory nature of experiments
  - ✓ Parameter explorations
  - ✓ Iterative methods
- Algebraic approach for workflows
  - ✓ Data uniformity
  - ✓ Workflow execution optimizations
- Interactive workflows
  - ✓ Fine-tuning adjustments during Wf execution
    - Parameter space slicing
    - Iterative support
    - Real time provenance
  - ✓ Experiment Optimization

Acknowledgements

- CNPq, CAPES, FAPERJ and Petrobras
- INRIA
- at COPPE / UFRJ:
  - ✓ NACAD's support team and CPU time

Matisse series of parallel paintings "execution"

Matisse-paires-et-series- George Pompidou, 2012

EXPLORING PROVENANCE IN HIGH PERFORMANCE SCIENTIFIC COMPUTING

THANKS!

marta@cos.ufrj.br