EXPLORING WORKFLOW FRAGMENTS IN MULTISITE CLOUD

Vítor Silva

Graduated from UFRJ in 2013

Master Student at COPPE/UFRJ,

started in March 2013,

expected to finish in Sept 2014

Advisor: Marta Mattoso

Co-advisor: Daniel de Oliveira

Internship at LIRMM (UMR CNRS)

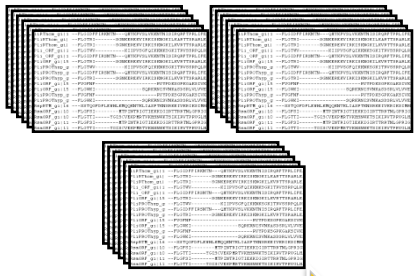
Member of SBC

COPPE UFRJ
Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia

Scientific Workflow Scenario

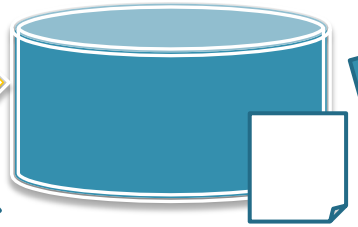
The analysis uses a chain of programs that may require HPC, *e.g.* clouds

1. Data is generated and collected

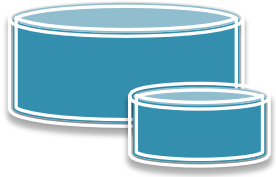


2. It is locally Evaluated

Provenance Data

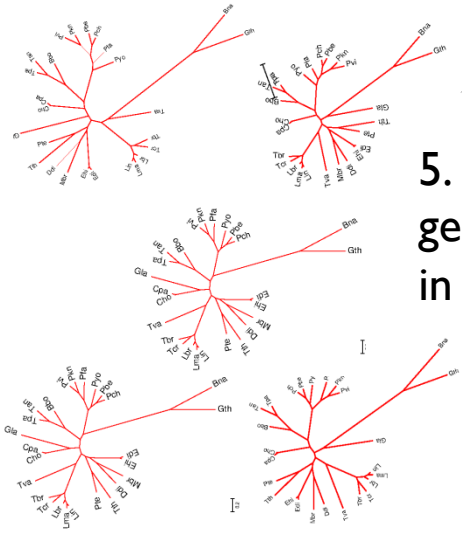


3. Large volume of data produced ...



4. ...which need to be processed by a computing-intensive environment

5. Final results generated in a feasible time



Phylogenetic trees

Aspects to Execute Scientific Workflows in HPC environment

- A parallel Scientific Workflow Management System (SWfMS) or a workflow engine is required to execute workflows in HPC environments
- Workflow or environment features
  - Data locality
  - Scientist locality
  - Storage capacity
  - Software permissions
- These restrictions motivate **workflow partitioning**

Workflow Fragments

- Workflow Definition (Ogasawara et al. PVLDB'11)
  - "Workflows can be defined as models of processes, which consist in series of activities and their dependencies" [1]
- Workflow Fragment Definition (Ogasawara et al. PVLDB'11)
  - Subset of the activities of a workflow
- Opportunities to explore workflow fragments
  - Use different parallel execution strategies for each fragment [1]
  - Generate optimized execution plan, based on workflow fragments
  - Use workflow fragments as the unit of distribution in multiple cloud sites

[1] E. Ogasawara, J. Dias, D. Oliveira, F. Porto, P. Valduriez, and M. Mattoso, "Optimization for Parallel Execution of Data-Intensive Scientific Workflows", JVLDB'11, 2011

Exploring Workflow Fragments in Multisite Cloud

- Aims to…
  - Evaluate the potential of workflow distributed execution in multiple cloud sites
- Considering
  - Activity and data dependencies between fragments
  - Site
    - A set of computational resources placed in the same region of an HPC environment
- Motivation
  - Existing SWfMS do not present an automatic mechanism…
    - To partition a scientific workflow
    - To manage the execution of fragments in multiple cloud sites

Exploring Workflow Fragments in Multisite Cloud

- Development during internship

  (joint work with Ji, Pacitti, Valduriez, Oliveira and Mattoso)

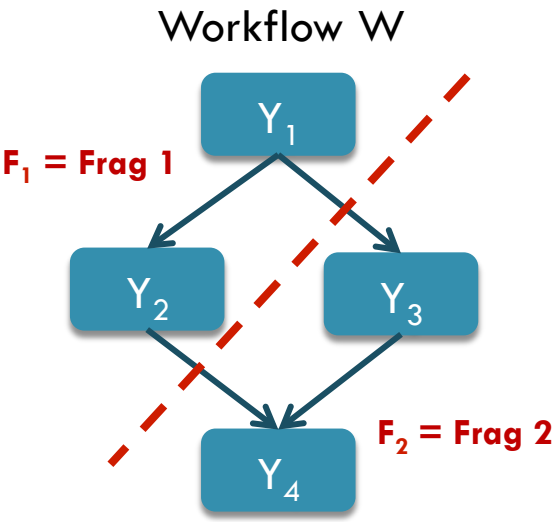
  - Partition workflow into several **fragments** [1]
  - Distribute each fragment to a specific site (manually)
  - Adapting Chiron to execute under in multiple cloud sites
  - Modeling a real scientific workflow
  - Workflow execution for big data analysis (Buzz workflow)
  - Execution in Amazon EC2 with 2 cloud sites
    - Using StarCluster (cloud resource provisioning) and Chiron (parallel workflow engine)
  - Paper under revision for submission

[1] E. Ogasawara, J. Dias, D. Oliveira, F. Porto, P. Valduriez, and M. Mattoso, "Optimization for Parallel Execution of Data-Intensive Scientific Workflows", JVLDB'11, 2011

Background (1)

- Workflow fragment (proposed by Ogasawara et al.)

A workflow W includes a set of activities $Y = \{Y_1, \ldots, Y_n\}$. Given $Y_i \mid (1 \leq i \leq n)$, let $R = \{R_1, \ldots, R_m\}$ be the input relation set for activity Y_i, then $Input(Y_i) \supseteq R$. Also, let T be the output relation set produced by activity Y_i, then $Output(Y_i) \supseteq T$. We denote the dependency between two activities as $Dep(Y_j, Y_i) \leftrightarrow \exists R_{k'} \in Input(Y_j) \mid R_{k'} \in Output(Y_i)$. Additionally, a fragment of a workflow, *fragment* for short, is a subset F of the activities of a workflow W, such that either F is an unitary set or $\forall Y_j \in F, \exists Y_i \in F \mid (Dep(Y_i, Y_j)) \vee (Dep(Y_j, Y_i))$.



$F_1 = \{Y_1, Y_2\}$; $F_2 = \{Y_3, Y_4\}$; $W = F_1 \cup F_2$

- Activation

Given a workflow W, a set $X = \{x_1, \ldots, x_k\}$ of activations is created for its execution. Each activation x_i belongs to a particular activity Y_j, which is represented as $Act(x_i) = Y_j$.

E. Ogasawara, J. Dias, D. Oliveira, F. Porto, P. Valduriez, and M. Mattoso, "An Algebraic Approach for Data-Centric Scientific Workflows," *Proc. of VLDB Endowment*, vol. 4, no. 12, pp. 1328–1339, 2011.

Background (2)

- Dataflow strategy
  - First Activity First (FAF) x First Tuple First (FTF)
- Dispatching strategy
  - Static x Dynamic

> given a workflow W, an associated workflow activations set $X=\{x_1, \ldots, x_k\}$ is evaluated according to a schedule. The schedule of activations depends on the dataflow strategy assigned to the corresponding workflow fragment. Thus, given a fragment F_i and a dataflow strategy DS_i, a mapping function $DSF(F_i, DS_i)$ assigns a dataflow strategy to a fragment of the workflow. In this context, given a set of activations $X'=\{x_1, \ldots, x_m\}$ associated to a fragment F_i, a dataflow strategy (DS_i) imposes a partial activation order among activations of X'

E. Ogasawara, J. Dias, D. Oliveira, F. Porto, P. Valduriez, and M. Mattoso, "An Algebraic Approach for Data-Centric Scientific Workflows," *Proc. of VLDB Endowment*, vol. 4, no. 12, pp. 1328–1339, 2011.

Expected Results from Exploring Workflow Fragments in Multisite Cloud

- Definition of some heuristics based on…
  - Data locality
    - Location of input data and workflow configuration files
  - Scientist locality
    - Monitoring at runtime
    - Modifications in workflow specification
- After (*manually*) workflow partitioning, each fragment is allocated in a different cloud site and performance was measured by
  - Total elapsed time with data transfer cost

Experiments

- We evaluated our approach using **Buzz Workflow**, a workflow for big data analysis
  - DBLP Computer Science Bibliography (http://dblp.uni-trier.de/db)

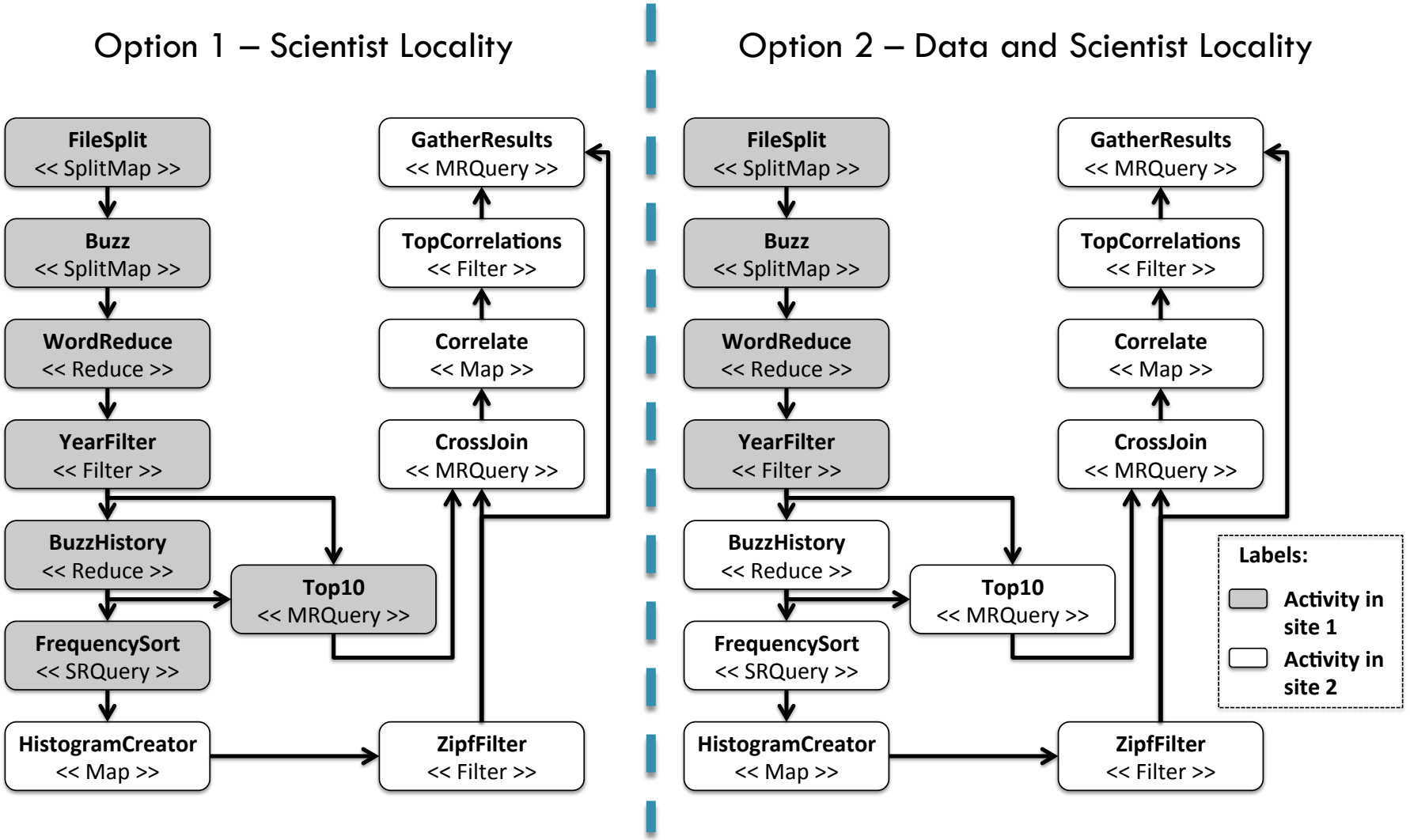
- Technologies
  - Chiron
    - Modifications in this parallel workflow engine to support execution in multisite cloud
  - StarCluster
    - An open source cluster computing toolkit to build and to configure clusters of virtual machines in cloud environments
  - Amazon EC2 → Cloud provider
    - Two different sites (or regions): US East (North Virginia) and US West (N. California)
    - Each site presents 2 m1.xlarge virtual machines with an amount of 16 cores and 1.7 GB of RAM memory

StarCluster - http://star.mit.edu/cluster/docs/latest/index.html

Chiron - Algebraic Operators

- Program invocation
  - Map (1:1)
  - SplitMap (1: n)
  - Reduce (n : 1)
  - Filter (1: 0-1)
- Relational algebra expressions
  - SRQuery → Single Relation Query
  - MRQuery → Multiple Relation Query

Buzz Workflow: a Study Case for Workflow Partitioning



Buzz Workflow: a Study Case for Workflow Partitioning

- Evaluation metrics for workflow fragments execution
  - Total elapsed time
  - Data transfer cost
  - Query performance

Time elapsed time and data transfer cost

Approach	Time in minutes				
	Site 1	*Site 2*	*Elapsed time without data transfer*	*Data transfer*	*Elapsed time with data transfer*
Sequential	289.3	0.0	289.3	0.0	289.3
First option	145.0	111.5	256.5	104.5	361.0
Second option	117.0	157.6	274.6	9.7	284.3

Query performance

Query to select some histogram files according to specific buzzwords

Sequential → 5,812 ms

Second option → 439 ms

THANK YOU!

EXPLORING WORKFLOW FRAGMENTS IN MULTISITE CLOUD

Vítor Silva