Optimizing Resource Allocation for Workflow Execution in Multi-site Clouds

Daniel de Oliveira
danielcmo@ic.uff.br

Motivation

Comparative Genomics (CG)

- A bioinformatics domain used for exploring complete genomes.
- Phylogenetics, phylogenomics and evolutionary analysis play an important role in CG.
  - composed by the execution of several programs in a coherent flow of activities:

(i) identify homologues sequences

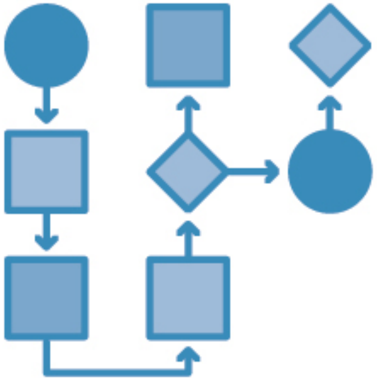
(ii) construct phylogenetic trees

(iii) construct phylogenomic trees

(iv) infer hypotheses about evolutionary relationships.

Motivation

- Each analysis is composed by a set of applications.
  - the outcome of one application is used as input of the next one in the flow.



- These analysis are modeled as **scientific workflow.**
  - a set of activities connected by a **dataflow**.
- **Comparative genomics workflows** are computing intensive by nature.
  - activities are executed repeatedly times by changing the input proteins to interpret the quality of the result by each analysis executed.

⇩

High Performance Computing (HPC) environments

Motivation

- Most of scientific workflows execute on traditional HPC environments.

- Many users do not have access to large clusters or grids.
  - They have adopted **clouds** to execute their experiments in parallel.

Main Advantage of Clouds

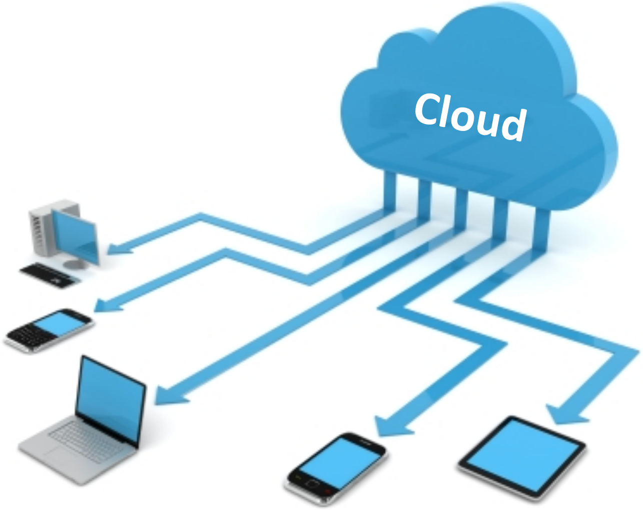
- huge variety of resources for the general public offered via Internet:

virtual machines (VMs)

storage

as-a-service model

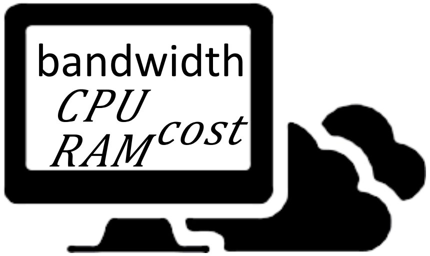
on demand model



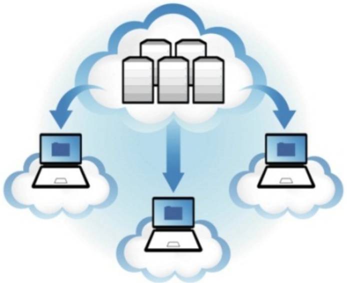
Motivation

Infrastructure-as-a-Service

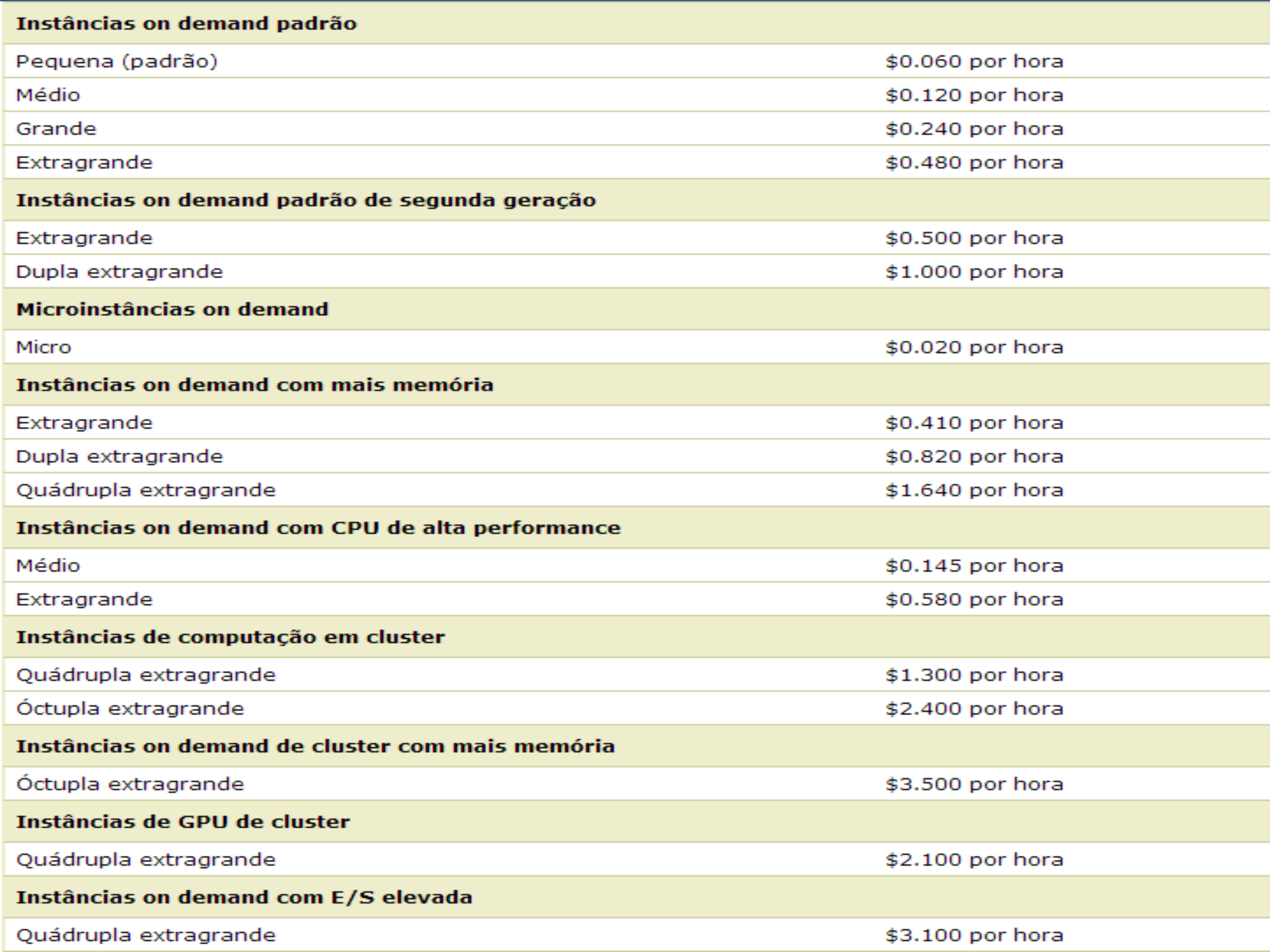
- Large number of offered VM types
  - each VM type has its own characteristics such as bandwidth, CPU, memory and financial cost.



- These VMs can be used to create a **virtual cluster** in the cloud to execute the workflow in parallel.
  - new VMs can be instantiated according to an estimated workload.



- **Problem: what is the amount of VMs to instantiate of each type and for how long?**



Instâncias on demand padrão	
Pequena (padrão)	$0.060 por hora
Médio	$0.120 por hora
Grande	$0.240 por hora
Extragrande	$0.480 por hora
Instâncias on demand padrão de segunda geração	
Extragrande	$0.500 por hora
Dupla extragrande	$1.000 por hora
Microinstâncias on demand	
Micro	$0.020 por hora
Instâncias on demand com mais memória	
Extragrande	$0.410 por hora
Dupla extragrande	$0.820 por hora
Quádrupla extragrande	$1.640 por hora
Instâncias on demand com CPU de alta performance	
Médio	$0.145 por hora
Extragrande	$0.580 por hora
Instâncias de computação em cluster	
Quádrupla extragrande	$1.300 por hora
Óctupla extragrande	$2.400 por hora
Instâncias on demand de cluster com mais memória	
Óctupla extragrande	$3.500 por hora
Instâncias de GPU de cluster	
Quádrupla extragrande	$2.100 por hora
Instâncias on demand com E/S elevada	
Quádrupla extragrande	$3.100 por hora

Motivation

- Some solutions for dimensioning the virtual cluster for scientific applications were proposed

SciDim	GraspCC
• Based on genetic algorithms	• Near-optimal estimations
• Non-optimal estimations	• HPC applications
• Embedded existing SWfMS	• Not designed for scientific workflows

- Non-optimal estimations may produce over and under dimensioning.

Goal

- To analyze the feasibility of the GraspCC approach for dimensioning the virtual cluster for this class of bioinformatics experiments.

Applying GraspCC in Workflows

- Underestimation or Overestimation

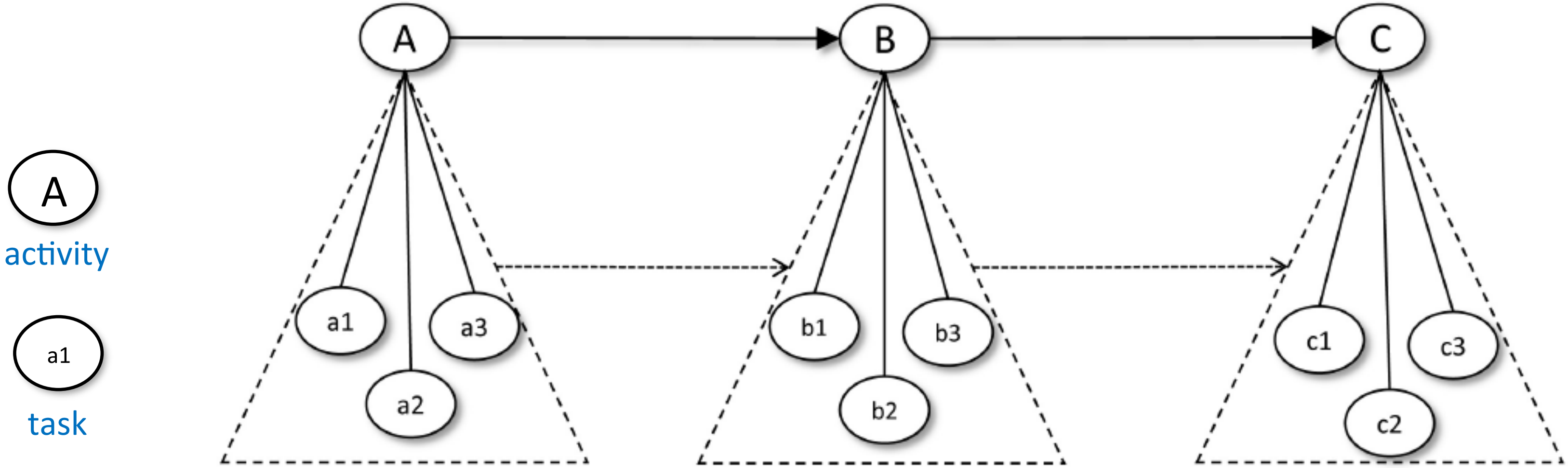
- It is fundamental to find the **best tradeoff** between performance and financial cost.

- **GraspCC**
  - provides **optimal or near-optimal estimations** of the amount VMs to be instantiated for general application executions in a **reasonable time.**
  - but… no data dependency among programs is considered.

Applying GraspCC in Workflows

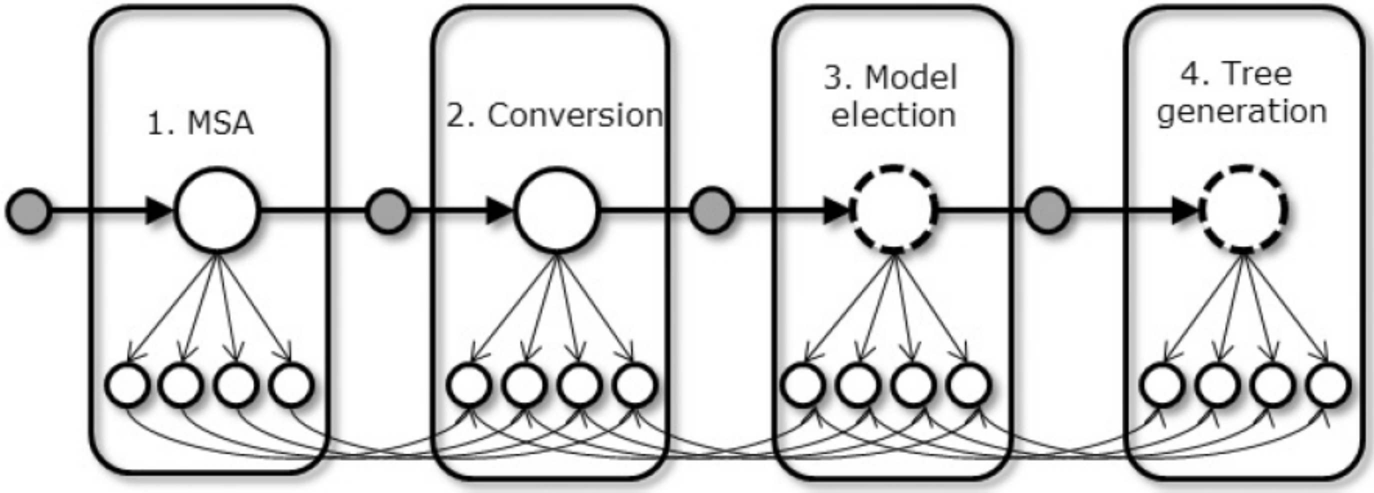
- The use of GraspCC was **adapted** for scientific workflow engine.

- GraspCC was invoked immediately before each activity execution
  - First Activity First (FAF) dataflow strategy.
    - each activity is divided into several tasks that can be executed in parallel in different VMs.



Applying GraspCC in Workflows

- SciPhy workflow execution with the invocations of GraspCC

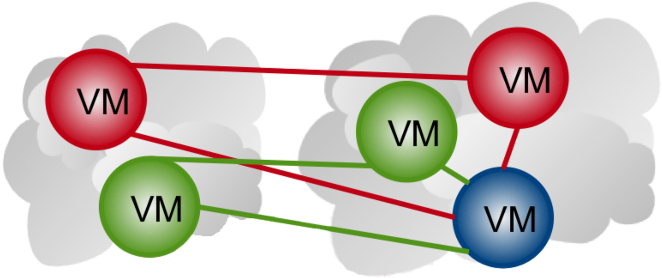


◯ activity ○ task ● GraspCC is invoked

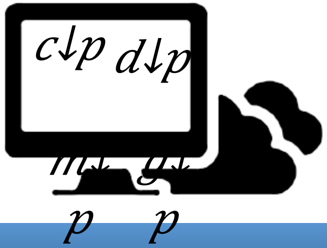
Applying GraspCC in Workflows

Modeling the problem...

- Let P be the set of VMs types offered by a cloud provider during a set of time periods.



- Each VM type $p \in P$ has:
  - cost $c{\downarrow}p$
  - computing resources:
    - disk storage $d{\downarrow}p$
    - memory capacity $m{\downarrow}p$
    - processing power of $g{\downarrow}p$ Gflop per period of time

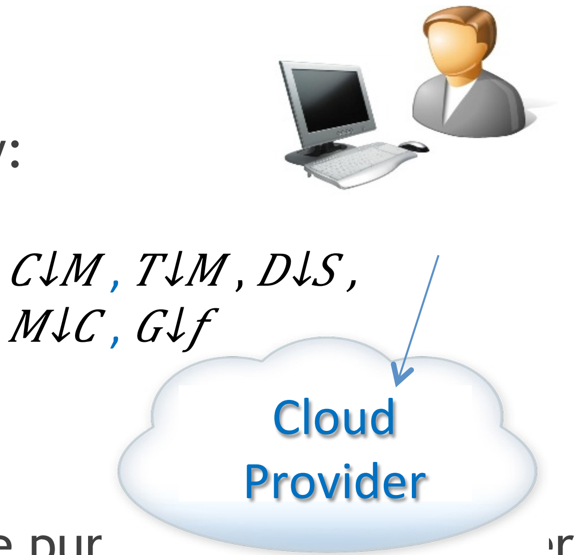


Applying GraspCC in Workflows

Modeling the problem...

- The set of user requirements is defined by:
  - maximum cost
  - maximum time
  - disk storage
  - memory capacity
  - processing demand of Gflop

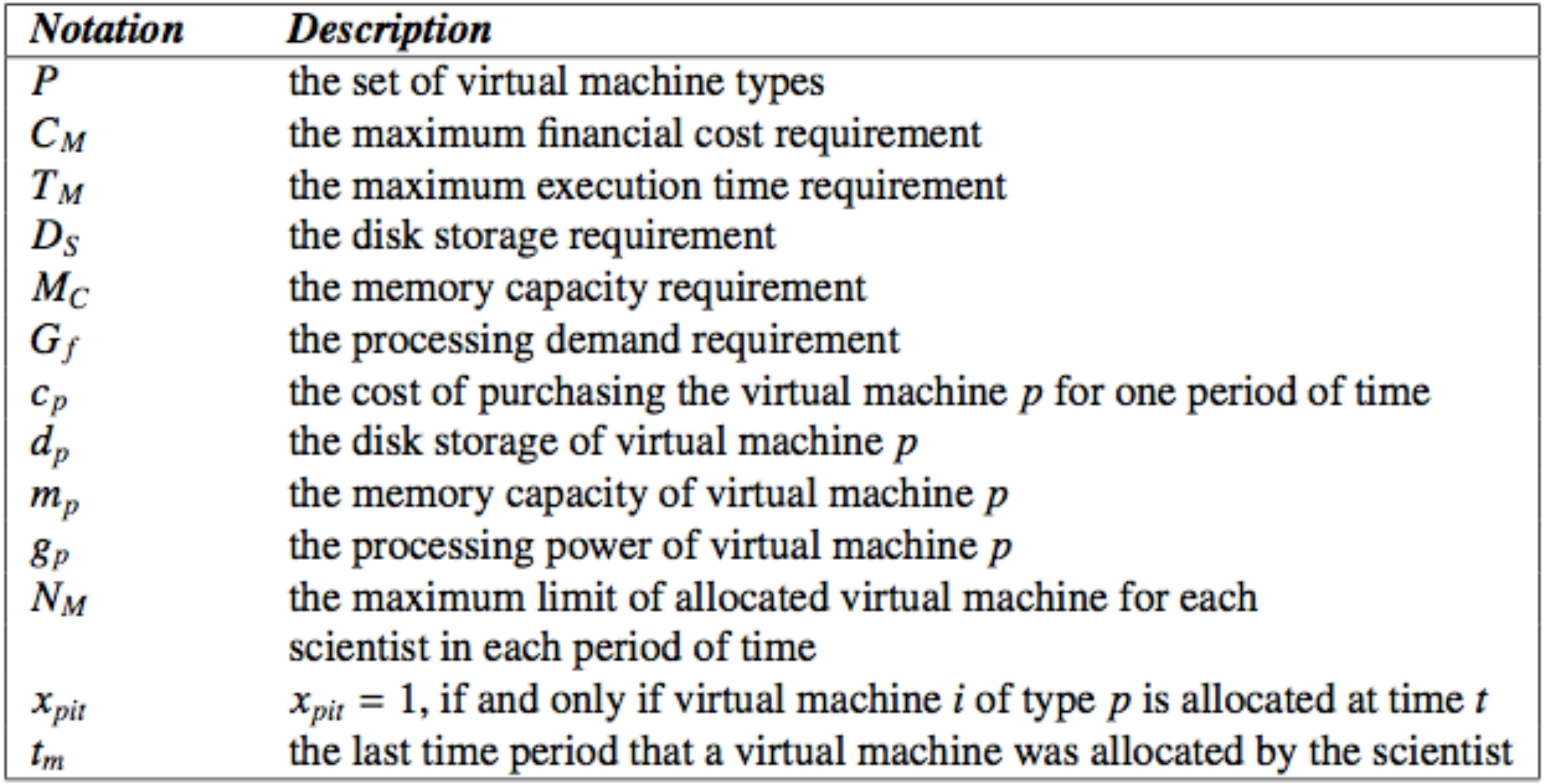
- Let , the maximum limit of VM that can be pur r in each period of time.



C_M, T_M, D_S, M_C, G_f

- Let be the last period that a VM was selected.

Modeling



Notation	*Description*
P	the set of virtual machine types
C_M	the maximum financial cost requirement
T_M	the maximum execution time requirement
D_S	the disk storage requirement
M_C	the memory capacity requirement
G_f	the processing demand requirement
c_p	the cost of purchasing the virtual machine p for one period of time
d_p	the disk storage of virtual machine p
m_p	the memory capacity of virtual machine p
g_p	the processing power of virtual machine p
N_M	the maximum limit of allocated virtual machine for each scientist in each period of time
x_{pit}	$x_{pit} = 1$, if and only if virtual machine i of type p is allocated at time t
t_m	the last time period that a virtual machine was allocated by the scientist

$$\min(\alpha_1 \sum_{p\in P} \sum_{i=1}^{N_M} \sum_{t\in T} c_p x_{pit} + \alpha_2 t_m) \tag{1}$$

$$\sum_{p\in P} \sum_{i=1}^{N_M} \sum_{t\in T} c_p x_{pit} \leq C_M \tag{2}$$

$$\sum_{p\in P} \sum_{i=1}^{N_M} d_p\ x_{pit} \geq D_S\ x_{p'i't}, \qquad \forall t \in T, \forall p' \in P, \ \forall i' \in \{1, \ldots, N_M\} \tag{3}$$

$$\sum_{p\in P} \sum_{i=1}^{N_M} m_p\ x_{pit} \geq M_C\ x_{p'i't}, \qquad \forall t \in T, \forall p' \in P, \ \forall i' \in \{1, \ldots, N_M\} \tag{4}$$

$$\sum_{p\in P} \sum_{i=1}^{N_M} \sum_{t\in T}\ g_p x_{pit} \geq G_f \tag{5}$$

$$\sum_{p\in P} \sum_{i=1}^{N_M} x_{pit} \leq N_M, \qquad \forall t \in T \tag{6}$$

$$t_m \geq t\ x_{pit}, \qquad \forall t \in T, \forall p \in P, \ \forall i \in \{1, \ldots, N_M\} \tag{7}$$

$$x_{pit+1} \leq x_{pit}, \qquad \forall t \in T, \forall p \in P, \ \forall i \in \{1, \ldots, N_M\} \tag{8}$$

$$x_{pi+1t} \leq x_{pit}, \qquad \forall t \in T, \forall p \in P, \ \forall i \in \{1, \ldots, N_M - 1\} \tag{9}$$

$$x_{pit} \in \{0, 1\}, \qquad \forall t \in T, \forall p \in P, \ \forall i \in \{1, \ldots, N_M\} \tag{10}$$

$$t_m \in \mathbb{Z} \tag{11}$$

Applying GraspCC in Workflows

Modeling the problem...

- Define a *cost function* which will measure the quality of the solution:

$$F(s) = (\alpha_1 \sum_{(p,i,t)\in s} c_p + \alpha_2 t_m(s))$$

$$+ \lambda_1(\max\{0, t_m(s) - T_M\}) + \lambda_2(\max\{0, \sum_{(p,i,t)\in s} c_p - C_M\})$$

Applying GraspCC in Workflows

GraspCC

- The heuristic *GraspCC* is composed of two phases:
  - a construction phase *coCC*
  - a local search phase *lsCC*

- GraspCC consists to perform the *coCC* following by the *lsCC* until the maximum number of iterations without improvement in the best solution found is satisfied.

Modeling Comparative Genomic Workflows

SciHmm: a workflow for homologue sequence identification

- First workflow that has to be executed in order to identify candidate drug targets for a specific disease.
- Performs a cross validation procedure to evaluate the specificity and sensibility of each Multiple Sequence Alignment (MSA) method.

OCAÑA, K. ; OLIVEIRA, D. ; DIAS, J. ; OGASAWARA, E. ; MATTOSO, M. L. Q. . Optimizing Phylogenetic Analysis Using SciHmm Cloud-based Scientific Workflow. In: The seventh IEEE e Science conference, 2011, Estocolmo. Proceedings of the seventh IEEE e Science conference. New York: IEEE Computer Society, 2011.

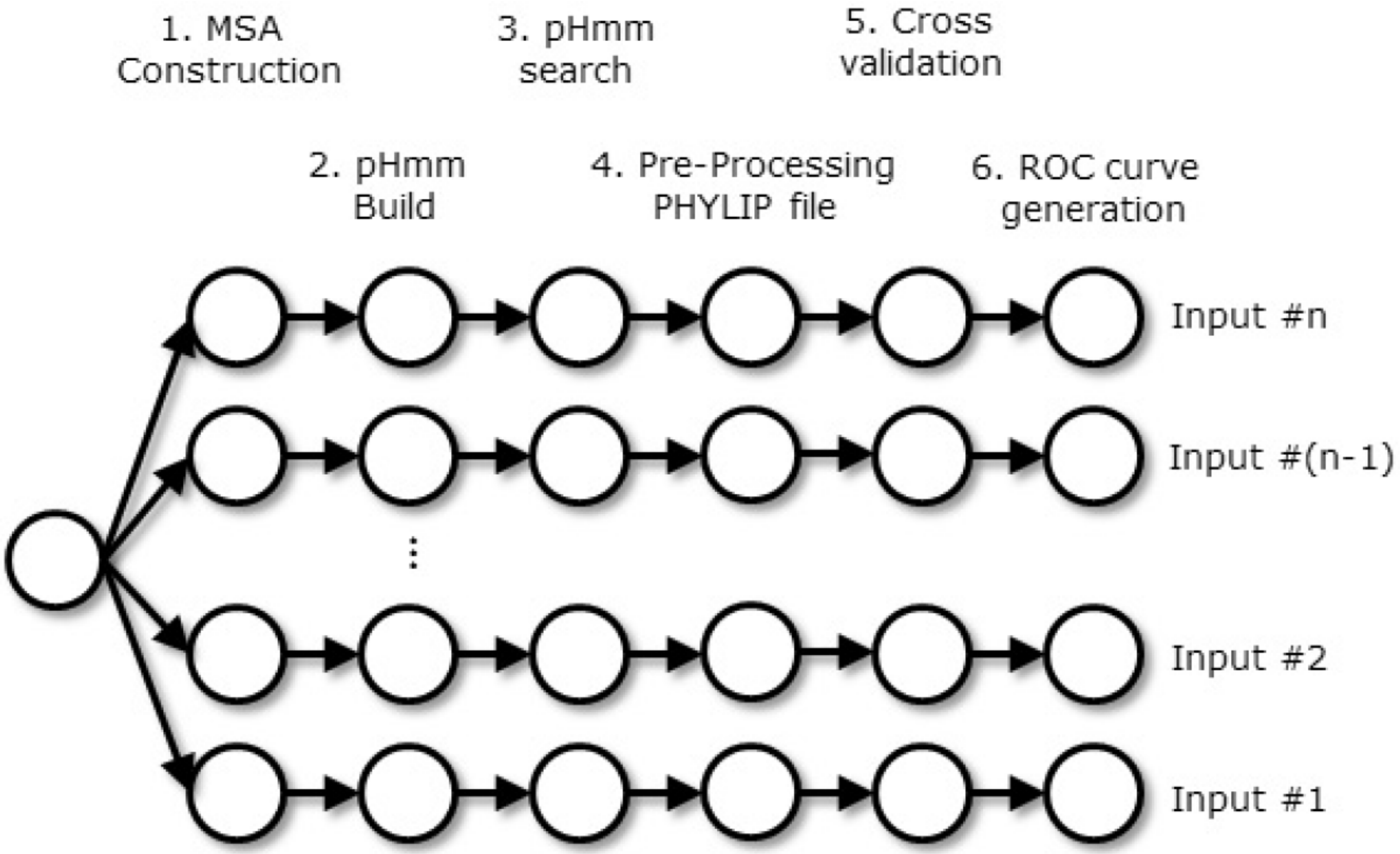
Modeling Comparative Genomic Workflows

SciHmm: a workflow for homologue sequence identification

- Composed by five activities:

1. MSA Construction
2. Contruction of Profile Hidden Markov Model (pHMM)
3. pHMM search against a target database
4. Cross-validation analysis
5. Generation of Receiver-Operating Characteristic (ROC) curves.

Modeling Comparative Genomic Workflows



Modeling Comparative Genomic Workflows

SciPhy: a workflow for phylogenetic analysis

- A phylogenetic analysis workflow that produces phylogenetic trees.
- First step towards identifying candidate drug targets enzymes in genomes.

OCAÑA, K. ; OLIVEIRA, D. ; OGASAWARA, E. ; DAVILA, A. M. R. ; LIMA, A. A. B. ; MATTOSO, M. L. Q. . SciPhy: A Cloud-based Scientific Workflow for Phylogenetic Analysis of Drug Targets in Protozoan Species. In: Brazilian Simposium of Bioinformatics 2011, 2011, Brasilia, DF. Proceedings of the Brazilian Simposium of Bioinformatics 2011, 2011.

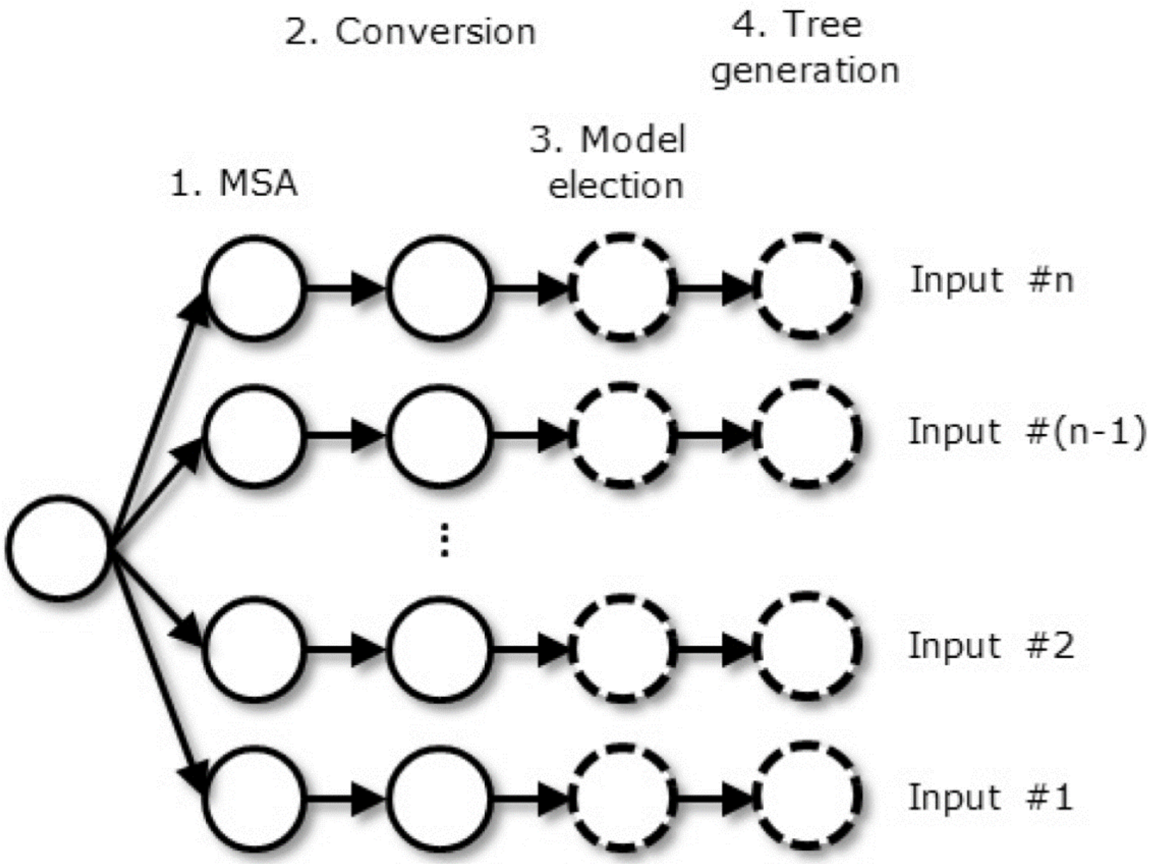
Modeling Comparative Genomic Workflows

SciPhy: a workflow for phylogenetic analysis

- Composed by four activities:

1. Sequence alignment
2. Sequence conversion
3. Search for the best evolutionary model
4. Construction of the phylogenetic trees.

Modeling Comparative Genomic Workflows



Modeling Comparative Genomic Workflows

SciPhylomics: a workflow for phylogenomic analysis

- A scientific workflow that aims inferring evolutionary relationships between homologous genes of different species.

- Composed by nine activities:

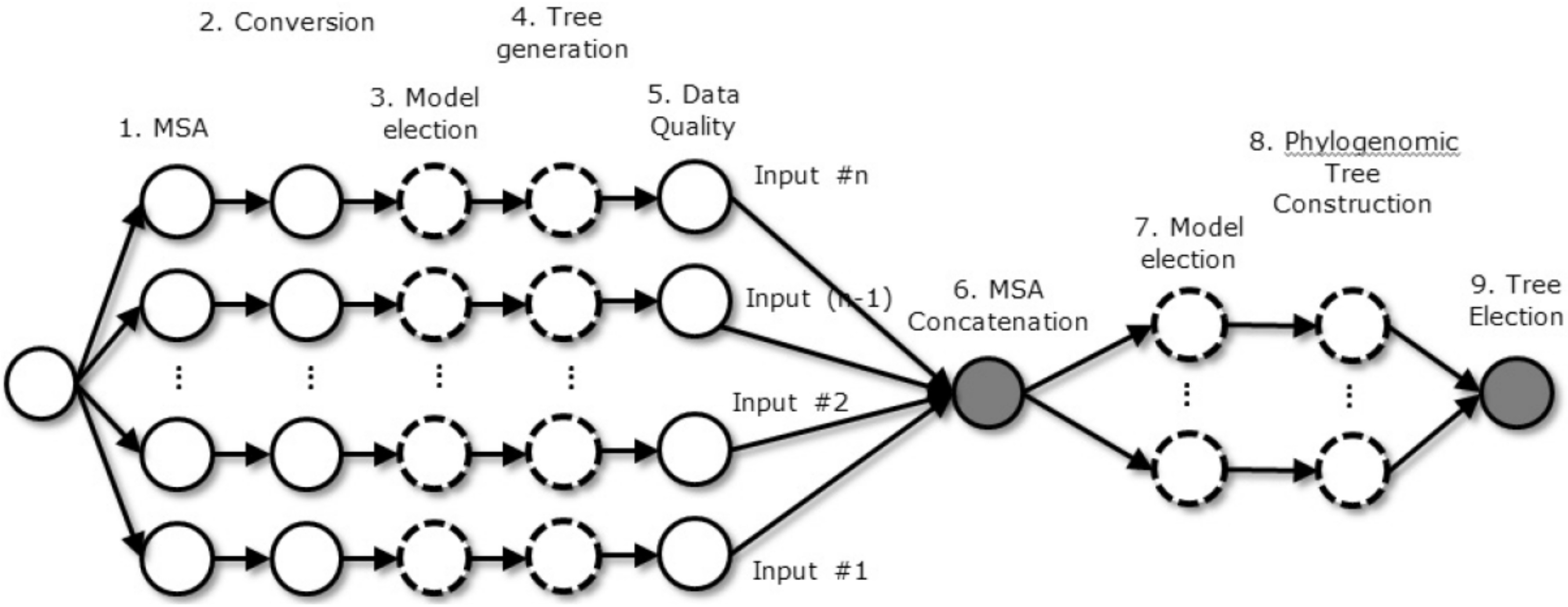
1-4. Sub-workflow SciPhy

5. Filter results that do not comply with a given quality criteria.

6-9. Phylogenomic analysis

OLIVEIRA, D. ; OCAÑA, K. ; OGASAWARA, E. ; DIAS, J. ; GONCALVES, J. ; BAIAO, F. ; MATTOSO, M. L. Q. . Performance evaluation of parallel strategies in public clouds: A study with phylogenomic workflows. Future Generation Computer Systems, v. 29, p. 1816-1825, 2013.

Modeling Comparative Genomic Workflows



Modeling Comparative Genomic Workflows

SciEvol: a workflow for molecular evolutionary analysis

- A scientific workflow that aims detecting positive Darwinian selection on genomic data.

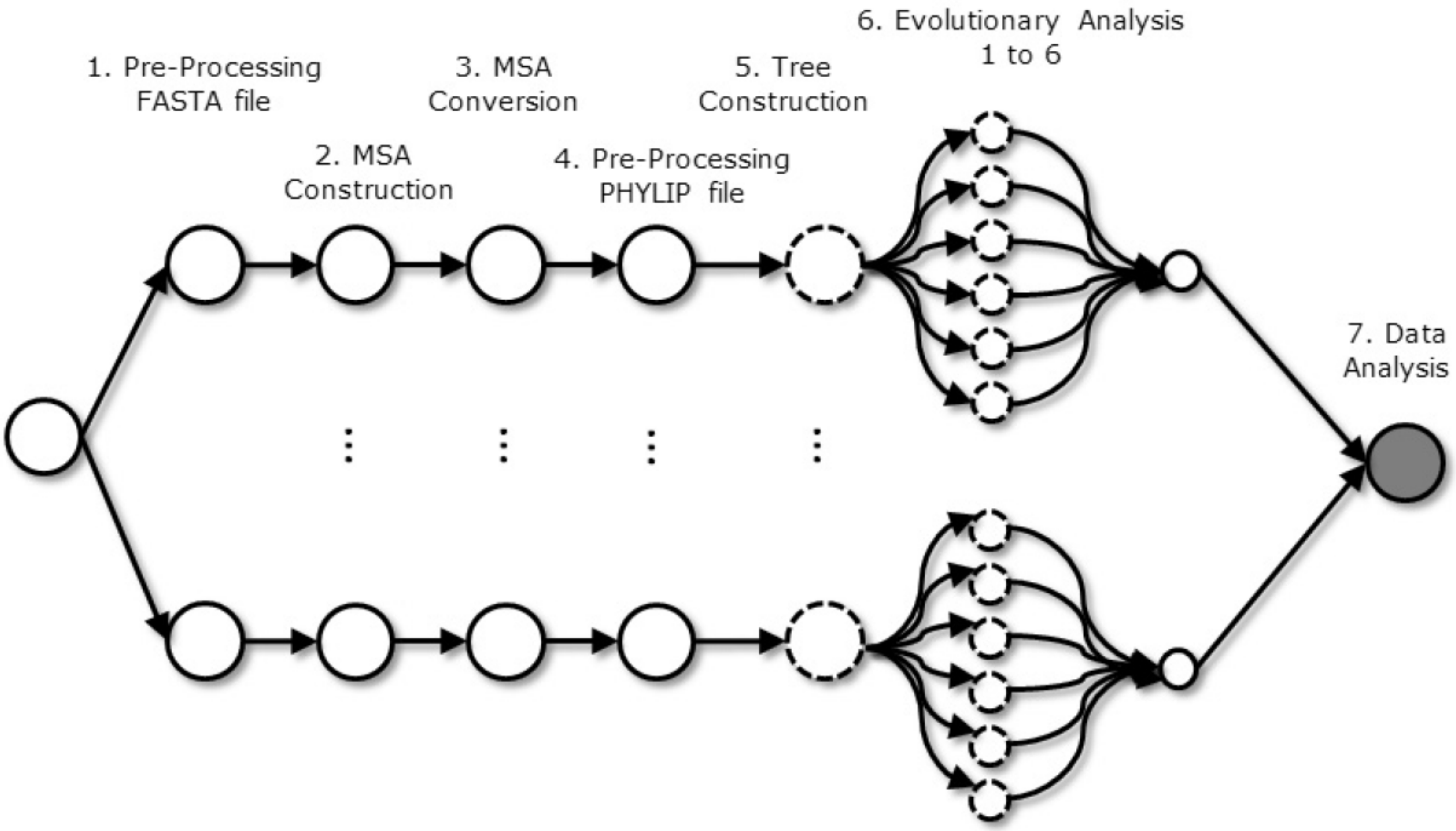
- Composed by eleven activities:

1. Stop codons removal
2. MSA construction
3. MSA format conversion to the PHYLIP format
4. Phylogenetic tree construction

5-10. Evolutionary analysis execution

11. Evolutionary data analysis.

Modeling Comparative Genomic Workflows


1. Pre-Processing FASTA file
2. MSA Construction
3. MSA Conversion
4. Pre-Processing PHYLIP file
5. Tree Construction
6. Evolutionary Analysis 1 to 6
7. Data Analysis

Experimental Results

- Evaluation of GraspCC to estimate the amount of VM for instantiating four real comparative genomics workflows executed in parallel in a cloud environment.

- Comparison the real performance and financial costs with the estimated by GraspCC:
  - Is it suitable to dimension the virtual cluster for this class of bioinformatics workflows?

Experimental Results

Cloud Environment Setup

- 5 types of VM: m1.small, m3.medium, m3.large, m3.xlarge and m3.2xlarge
- Linux Cent OS 5 (64-bit) and Secure Shell (SSH)
- Amazon image ami-6e1a8907 contains all programs installed and it is stored in the cloud.
- SciCumulus creates a virtual cluster to execute the experiment based on this image.
- GraspCC implemented in ANSI C
  - Executed in an isolated computer with processor equivalent to Intel Core i5 2.5GHz and 6Gb of RAM under
  - Linux (Ubuntu 12.04) operating system.

Experimental Results

Experiment Setup

- **Focus:** cysteine proteases (CPs) as a candidate drug target for protozoan diseases.
  - malaria are candidates for novel approaches utilizing CPs as targets.
- **Input:** a dataset of 100 fasta files of target protein sequences of falcipain CPs from Plasmodium species to identify CPs.
  - using scientific workflows that were executed using cloud with SciCumulus engine.
- **Workflows:** SciHmm, SciPhy, SciPhylomics and SciEvol.
- BLAST 2.2.18, MAFFT v7.012b, Muscle v3.7, MODELLER 9-11, NEST (Jackal 1.5), CONGEN Version 2.2.1, SEGMOD v1.0, and PROCHECK v.3.5.4, RAxML 7.2.8-ALPHA, Mega 5.1, ReadSeq 2.1.26, Molsoft ICM browser 3.7-2c and PAML 4.7.
  - using default parameters.

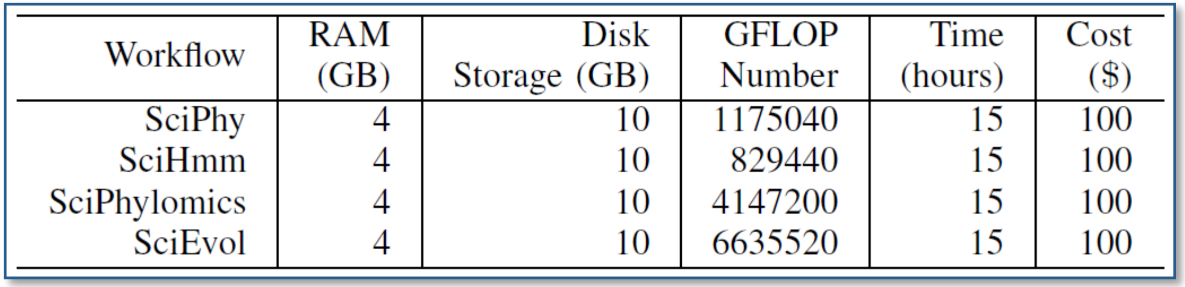
Experimental Results

Performance Results

- **First step to use GraspCC:** calculate the necessary GFlops for each workflow
  - historical execution of all previous executions.

- Requirements for each workflow:
  - RAM, disk space, maximum execution time allowed and maximum financial cost allowed.

- The maximum number allowed is 20.



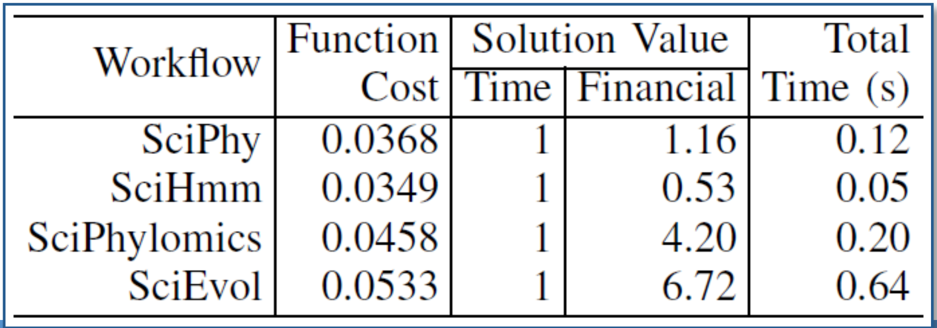
Workflow	RAM (GB)	Disk Storage (GB)	GFLOP Number	Time (hours)	Cost ($)
SciPhy	4	10	1175040	15	100
SciHmm	4	10	829440	15	100
SciPhylomics	4	10	4147200	15	100
SciEvol	4	10	6635520	15	100

Experimental Results

Performance Results

- GraspCC produced the following estimations (with $\alpha_1 = 0.5$ and $\alpha_2 = 0.5$):

1. SciPhy – 1 m1.small and 4 m3.xlarge for 1 hour
2. SciHmm - 12 m1.small virtual machines for 1 hour
3. SciPhylomics - 1 m3.xlarge and 7 m3.2xlarge virtual machines for 1 hour
4. SciEvol – 12 m3.2xlarge virtual machines for 1 hour.

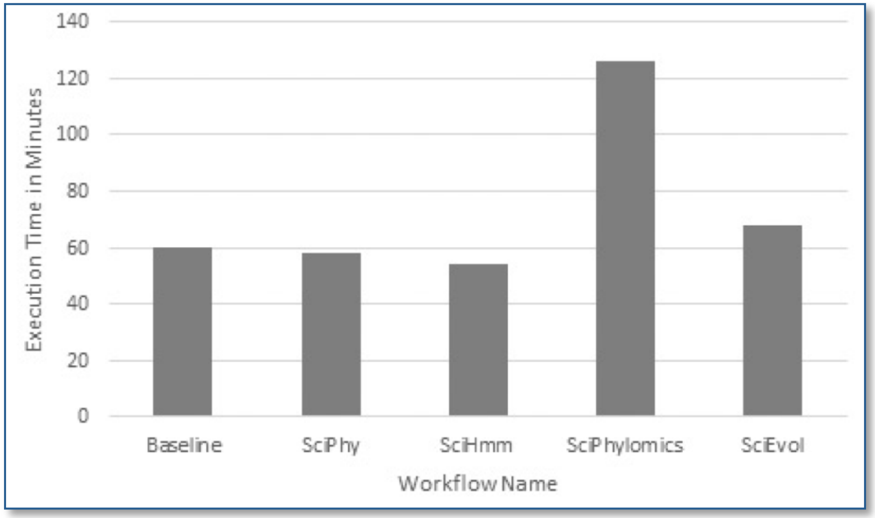


Workflow	Function Cost	Solution Value		Total
		Time	Financial	Time (s)
SciPhy	0.0368	1	1.16	0.12
SciHmm	0.0349	1	0.53	0.05
SciPhylomics	0.0458	1	4.20	0.20
SciEvol	0.0533	1	6.72	0.64

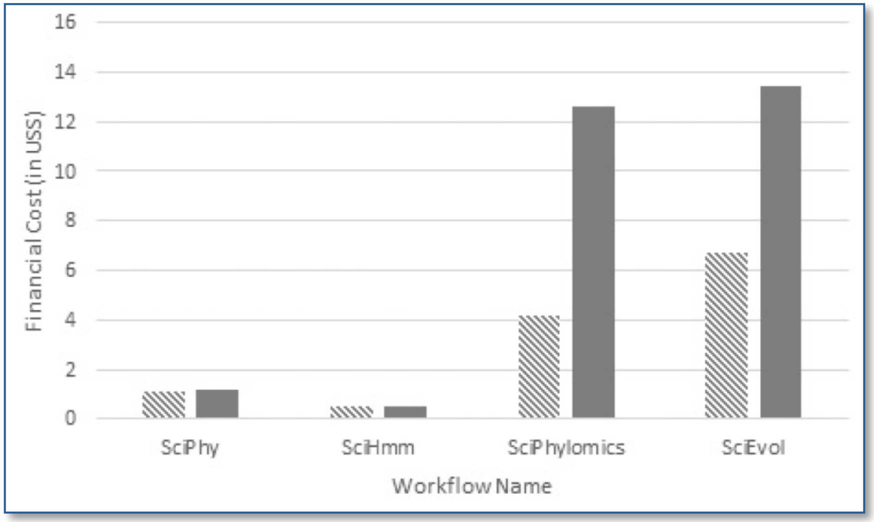
Experimental Results

Performance Results

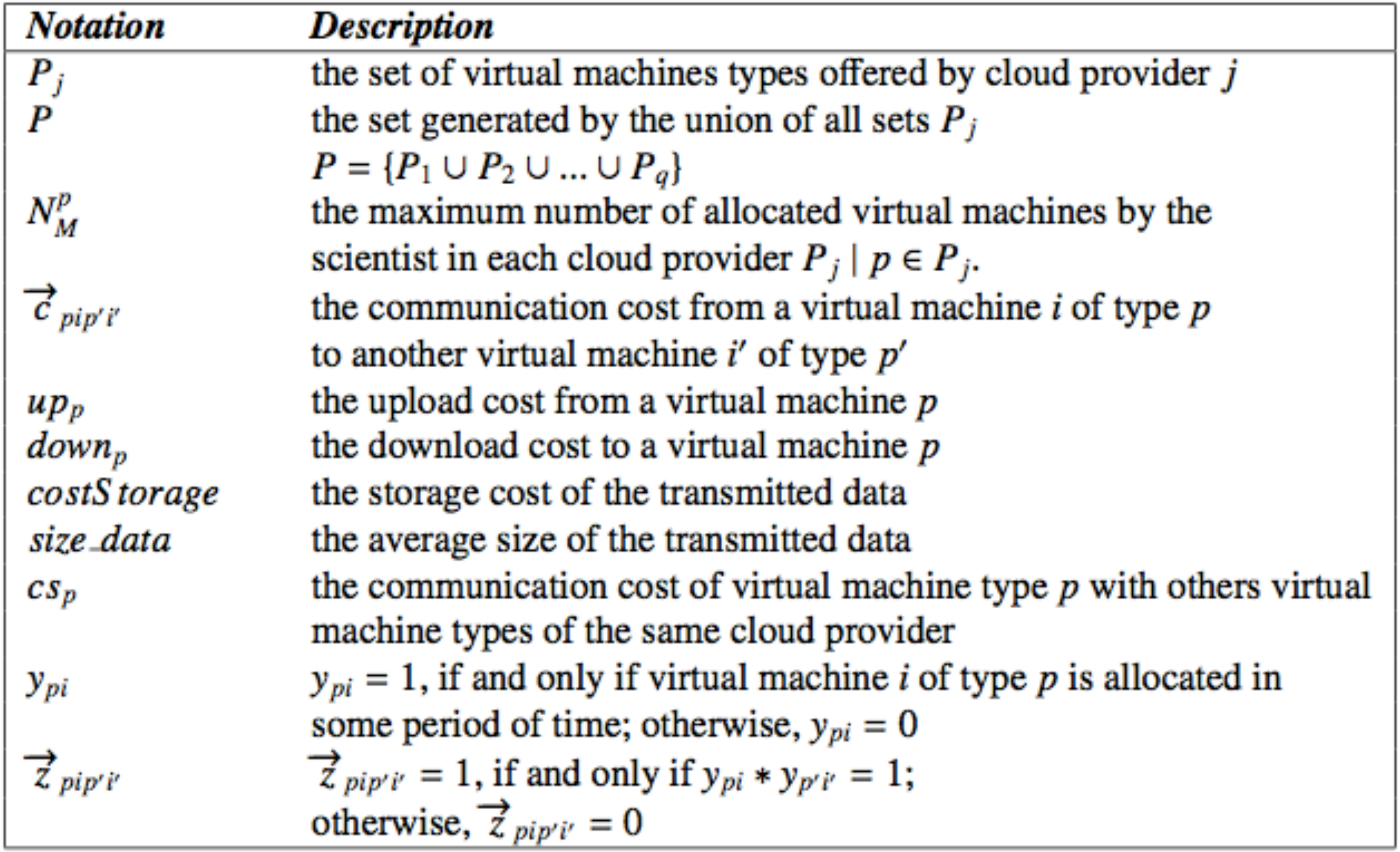
- After the GraspCC estimation, the workflows were executed using the suggested configuration.



Total execution time per workflow



Total financial cost per workflow



Notation	***Description***
P_j	the set of virtual machines types offered by cloud provider j
P	the set generated by the union of all sets P_j $P = \{P_1 \cup P_2 \cup \ldots \cup P_q\}$
N_M^p	the maximum number of allocated virtual machines by the scientist in each cloud provider $P_j \mid p \in P_j$.
$\overrightarrow{c}_{pip'i'}$	the communication cost from a virtual machine i of type p to another virtual machine i' of type p'
up_p	the upload cost from a virtual machine p
$down_p$	the download cost to a virtual machine p
$costStorage$	the storage cost of the transmitted data
$size\_data$	the average size of the transmitted data
cs_p	the communication cost of virtual machine type p with others virtual machine types of the same cloud provider
y_{pi}	$y_{pi} = 1$, if and only if virtual machine i of type p is allocated in some period of time; otherwise, $y_{pi} = 0$
$\overrightarrow{z}_{pip'i'}$	$\overrightarrow{z}_{pip'i'} = 1$, if and only if $y_{pi} * y_{p'i'} = 1$; otherwise, $\overrightarrow{z}_{pip'i'} = 0$

(CC-IP-fed)

$$\begin{aligned}&\min(\alpha_1(\sum_{p\in P}\sum_{i=1}^{N_M^p}\sum_{t\in T}\overrightarrow{c}_p x_{pit}+\\ &\sum_{p\in P}\sum_{p'\in P}\sum_{i=1}^{N_M^p}\sum_{i'=1}^{N_M^{p'}}\overrightarrow{c}_{pip'i'}\overrightarrow{z}_{pip'i'})+\alpha_2 t_m)\end{aligned} \tag{17}$$

$$\sum_{p\in P_j}\sum_{i=1}^{N_M^p} x_{pit} \leq N_M^p, \qquad \forall j = 1...q, \forall t \in T \tag{18}$$

$$y_{pi} \geq \overrightarrow{z}_{pip'i'}, \qquad \begin{aligned}&\forall p, p' \in P,\\ &\forall i \in \{1, \ldots, N_M^p\}, i' \in \{1, \ldots, N_M^{p'}\}\end{aligned} \tag{19}$$

$$y_{p'i'} \geq \overrightarrow{z}_{pip'i'}, \qquad \begin{aligned}&\forall p, p' \in P,\\ &\forall i \in \{1, \ldots, N_M^p\}, i' \in \{1, \ldots, N_M^{p'}\}\end{aligned} \tag{20}$$

$$y_{pi} + y_{p'i'} - 1 \leq \overrightarrow{z}_{pip'i'}, \qquad \begin{aligned}&\forall p, p' \in P,\\ &\forall i \in \{1, \ldots, N_M^p\}, i' \in \{1, \ldots, N_M^{p'}\}\end{aligned} \tag{21}$$

$$\sum_{t\in T} x_{pit} \leq y_{pi}|T|, \qquad \begin{aligned}&\forall p \in P,\\ &\forall i \in \{1, \ldots, N_M^p\}\end{aligned} \tag{22}$$

$$y_{pi} \in \{0, 1\}, \qquad \begin{aligned}&\forall p \in P,\\ &\forall i \in \{1, \ldots, N_M^p\}\end{aligned} \tag{23}$$

$$\overrightarrow{z}_{pip'i'} \in \{0, 1\}, \qquad \begin{aligned}&\forall p, p' \in P,\\ &\forall i \in \{1, \ldots, N_M^p\}, i' \in \{1, \ldots, N_M^{p'}\}\end{aligned} \tag{24}$$

“Conclusions”

- It was possible measure the accuracy of the GraspCC estimations in large-scale and parallel scientific workflow scenarios.
- The results indicate that the real executions of comparative genomics workflows with virtual cluster dimensioned by GraspCC have 100% of precision in 2 of the 4 executed workflows (SciHmm and SciPhy).
- The experiments with multi-site clouds are ongoing work…