

Predição do impacto do fluxo de dados no desempenho de workflows científicos

Duílio Henrique Haroldo Elias

RELATÓRIO PARCIAL DE INICIAÇÃO CIENTÍFICA

Programa: Programa Institucional de Bolsas de
Iniciação Científica em Desenvolvimento Tecnológico
e Inovação – PIBITI/CNPq/USP

Orientadora: Kelly Rosa Braghetto

Departamento de Ciências da Computação
Instituto de Matemática e Estatística
Universidade de São Paulo

10 de fevereiro de 2014

Resumo

Nas últimas décadas os workflows científicos receberam grande atenção da comunidade acadêmica, devido à sua grande capacidade de manipulação, análise e simulação de experimentos que trabalham com grandes quantidades de dados. No entanto, as ferramentas existentes para tal fim são ainda escassas e muitos cientistas afirmam que isso seja o grande gargalo da produção científica. Neste trabalho, busca-se contribuir com este processo, a partir da identificação e caracterização de linguagens de modelagem de workflows científicos, criando modelos estocásticos capazes de prever o impacto causado pelo fluxo de dados no desempenho e na execução de workflows em ambientes computacionais distribuídos. Com isso, pretende-se desenvolver um algoritmo que gera modelos estocásticos a partir do fluxo de dados de workflows científicos, extraíndo índices capazes de prever o impacto do uso de diferentes abordagens de controle de fluxo de dados no desempenho dos workflows.

Palavras-chaves: Workflows científicos. Redes de Petri. Análise de Desempenho. Gerenciadores de Workflows.

Sumário

Sumário	3
1 Introdução	4
2 Objetivo	4
3 Conceitos Preliminares	5
3.1 Workflows científicos	5
3.2 Sistemas gerenciadores de workflows científicos	7
3.2.1 Taverna	7
3.2.2 Pegasus	8
3.3 Redes de Petri	9
3.3.1 PIPE	10
3.4 Simuladores de workflows científicos	11
3.4.1 WorkflowSim	12
4 Metodologia	12
5 Resultados Parciais	12
6 Conclusões Parciais	13
6.1 Tarefas Realizadas	13
6.2 Tarefas em Andamento	13
6.3 Tarefas a serem realizadas	13
Referências	15
7 Anexos	16
7.1 Anexo 1	16
7.2 Anexo 2	16
7.3 Anexo 3	18

1 Introdução

Um workflow científico pode ser compreendido como um conjunto de tarefas computacionais que constituem um experimento científico[Júnior, 2012]. Esse conjunto de tarefas pode ser dividido em modelagem, execução e análise. Um cientista pode definir seu workflow com o auxílio de uma ferramenta de software, com uma interface gráfica simples e intuitiva. Depois de criado um modelo, o gerenciador auxilia o pesquisador a executar o workflow de forma eficiente, automaticamente, com pouca ou nenhuma intervenção manual.

Os workflows científicos são conhecidos pelo seu poder de manipulação e execução de projetos que trabalham com grandes quantidades de dados. Neste trabalho, estamos interessados em melhorar o desempenho da execução de workflows em ambientes computacionais distribuídos, o que muitas vezes não é uma tarefa fácil. Para isso, o sistema gerenciador precisa considerar a relação de dependência que existe entre as tarefas do workflow, ou seja, precisa definir quais são os dados de entrada de uma determinada tarefa, a fonte desses dados e destino dos dados processados. Há também um custo envolvido nas transferências de dados, em termos de tempo, que precisa ser considerado e que depende do ambiente computacional.

Todas essas observações se relacionam ao modelo de workflow de duas formas diferentes: primeiro, pela relação de dependência existente entre as tarefas, ou seja, como os dados irão percorrer o workflow e, segundo, quais as características computacionais do ambiente em que o workflow será executado.

O conjunto de conexões que interligam as atividades pode ser caracterizado como fluxo de dados, esse fluxo de dados define as dependências entre as atividades e os dados manipulados[Teixeira, 2013].

2 Objetivo

Neste trabalho, pretende-se identificar e caracterizar os principais construtores de fluxo de dados encontrados em linguagens de modelagem de workflows científicos, bem como criar modelos estocásticos em Redes de Petri capazes de prever o impacto causado pelo fluxo de dados no desempenho da execução de workflows. Para validar os índices que espera-se obter do modelo estocástico criado, serão modelados em Redes de Petri um conjunto de workflows científicos bem conhecidos pela comunidade científica, como o Montage¹. As previsões obtidas a partir do modelo de

¹ <https://confluence.pegasus.isi.edu/display/pegasus/Montage+Characterization>

Redes de Petri são comparadas com as obtidas a partir da simulação da execução dos mesmos workflows na ferramenta de simulação CloudSim-DVFS. Com as experiências obtidas nas fases anteriores pretende-se criar um algoritmo capaz de gerar modelos estocásticos a partir de modelos de fluxos de dados de workflows científicos.

3 Conceitos Preliminares

3.1 Workflows científicos

O termo workflow tem sua origem ligada a processos de automação de escritórios que surgiram na década de 70. Com o objetivo de gerar uma tecnologia capaz de fornecer uma base de apoio à distribuição de documentos nos escritórios, o que resultou na diminuição do uso de papéis[Ogasawara, 2011]. Embora a sua origem esteja mais relacionada com a automação de tarefas de logística, essa teoria foi expandida para outras áreas do conhecimento, como Economia, Administração, até mesmo Física, Biologia e Oceanográfica. Segundo *Workflow Management Coalition*[WfM, 2014], um workflow pode ser definido como: "A automação de um processo de negócio, completo ou apenas parte dele, através do qual documentos, informações ou tarefas são transmitidos de um participante a outro por ações, de acordo com regras procedimentais".

Já o termo workflow científico está bem mais ligado com as ciências do que o termo workflow, por ter o seu foco na manipulação de dados e na execução de tarefas que estão diretamente relacionadas com o processamento de dados. Essa grande utilização de dados caracteriza os workflows científicos como uma estrutura baseada no fluxo de dados, ou seja, as conexões entre as atividades representam o fluxo de dados.[Teixeira, 2013]

Os sistemas de workflows científicos geralmente são compostos por um grande número de tarefas que podem depender dos dados produzidos por outras tarefas. Essas tarefas podem ser manuais ou automáticas e são responsáveis pela gerência de todo o ciclo de vida do experimento científico: composição, mapeamento, execução e análise. Cada etapa do ciclo do experimento é muito importante e precisa ser analisada com cuidado[Júnior, 2012]:

- Composição: especificações das tarefas que compõem o experimento, dos tipos de dados de entrada e saída e das dependências de dados existentes entre as atividades e o ambiente computacional em que o workflow será executado.

- Mapeamento: associação do modelo de workflow com os recursos computacionais disponíveis.
- Execução: realização das tarefas definidas anteriormente no ambiente computacional.
- Análise: estudo dos dados gerados na execução.

Todo esse processo pode ser automatizado de forma que o cientista possa reexecutá-lo até que os dados obtidos na fase de análise sejam os dados esperados.

A estrutura de um workflow é composta pelas atividades a serem executadas e as conexões entre elas. Compreende-se por atividade cada ação realizada que recebe um dado de entrada e produz um dado ou uma informação de saída. O conjunto de conexões entre as atividades pode ser caracterizado como fluxo de controle ou fluxo de dados. Neste trabalho estamos interessados em analisar apenas o fluxo de dados, pois é a estrutura que está mais interligada com os workflows científicos. O fluxo de dados representa a dependência entre as atividades e os dados manipulados, e a forma como os dados são utilizados no workflow, com atividades produzindo dados que serão consumidos por outras atividades[Teixeira, 2013].

As principais construções de fluxo de dados são[Teixeira, 2013]:

- *Pipeline*: são várias atividades combinadas sequencialmente, em que cada atividade recebe como entrada os dados produzidos como saída da atividade anterior;
- *Distribuição de dados(Data Partitioning)*: é feita por atividades que produzem dados de saída que são recebidos como entrada por múltiplas atividades, ou que apenas particionam grandes conjuntos de dados em subconjuntos menores para serem processadas por outras atividades;
- *Agregação de dados(Data Agregation)*: é feita por atividades que processam múltiplos conjuntos de dados de entrada, gerando uma combinação dos dados como saída;
- *Redistribuição de dados*: é feita por atividades que funciona como um ponto de sincronização com relação ao processamento de dados, recebendo como entrada várias porções de dados e produzindo também múltiplas saídas como resultado.

Veja a figura ?? alguns exemplos gráficos de fluxo de dados.

3.2 Sistemas gerenciadores de workflows científicos

Para apoiar os workflows científicos, diversos Sistemas de Gerencia de Workflows Científicos(SGWfC) foram desenvolvidos. Onde cada um deles apresentam uma particularidade para representação dos workflows e visa atender determinados tipos de aplicações de áreas específicas. Estes SGWfC são responsáveis pela coordenação das chamadas a programas, seja em ambiente local ou para ambientes distribuídos[Ogasawara, 2011]. Os Sistemas mais utilizados , segundo o numero de workflows compartilhados no *myexperiment*², estão descritos, de maneira bem básica, logo abaixo:

3.2.1 Taverna

O Taverna foi desenvolvido como parte do projeto *mygrid*³, é um ambiente para modelagem, execução e análise de workflows científicos, tem como foco principal apapar cientistas da área de Ciências Biológicas que normalmente não possuem conhecimento profundo de computação. O Taverna tem código aberto o que permite unir projeto de terceiros[Teixeira, 2013].

Um workflow no Taverna é especificado como um grafo direcionado onde os nós, chamados de processadores, são as atividades e representam componentes de software. Um nó processador consome dados que chegam nas suas entradas e produz dados de saída. Cada arco no grafo conecta um entrada com uma saída de dados e expressa uma dependência entre as atividades o que denota o fluxo de dados entre as atividades[Teixeira, 2013].

A construção e execução dos workflows no Taverna é feita com o auxilio de uma interface gráfica, que é representado posteriormente na linguagem Scufi(*Simplified Conceptual Workflow Language*) que descreve o workflow em um formato baseada em XML e depois é executado pelo *Motor de Execução*[Teixeira, 2013]. Observe a seguir uma imagem dessa interface gráfica:

A linguagem Schfl possui uma serie de componentes que permite a representar no seu formato todo o workflow. A construção mais básica do Taverna são os *Pipelines*, porém é possível criar construções de fluxos de dados como Agregação, Distribuição e Redistribuição de dados [Teixeira, 2013].

Como pode ser visto na figura acima o Taverna possui uma integração com o *myexperiment* para compartilhamento dos workflows mode-

² url="http://www.myexperiment.org/", repositório de workflows científicos

³ http://www.mygrid.org.uk/

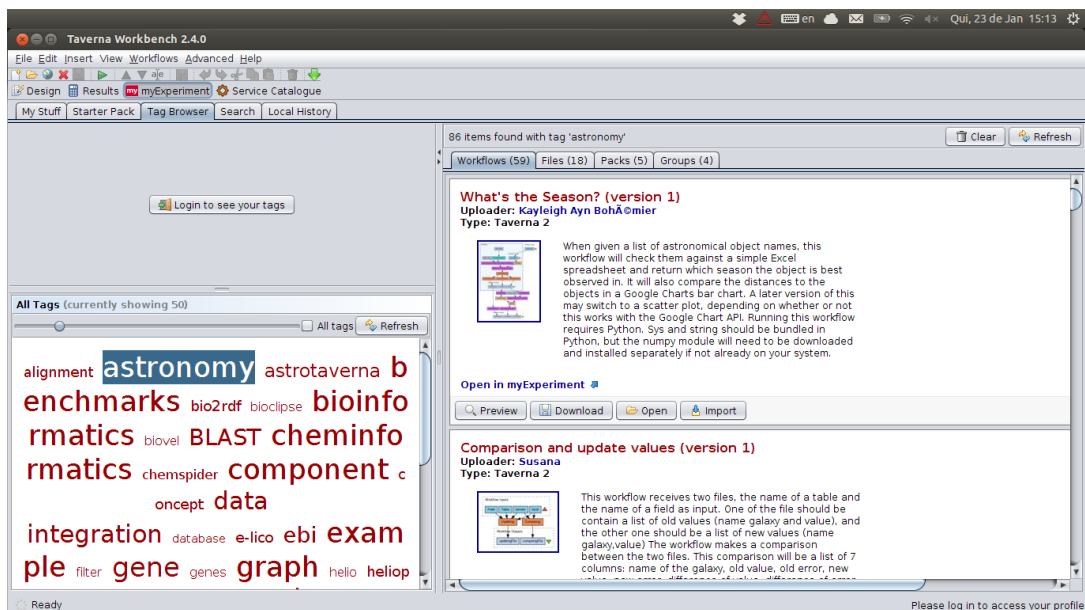


Figura 1 – Interface gráfica do Taverna

lados nele. O que talvez tenha o tornado um dos SGWC mais conhecidos, além de todo seu poder de construção de workflows científicos voltado para área Biológicas.

3.2.2 Pegasus

O SGWC Pegasus é um *framework* que permite o mapeamento de instâncias de workflows em um conjunto de recursos distribuídos, como uma grade. A descrição de um workflow no Pegasus é feita a partir de um arquivo de entrada na forma de um DAG em formato XML, na linguagem DAX (*Directed Acyclic Graph in XML*), que representa todo o fluxo de processamento e as dependências entre as tarefas, além dos recursos computacionais em que o workflow será executado [Teixeira, 2013].

O DAX descreve os workflows termos de atividades, onde cada atividade consiste de um código de identificação da atividade, os nomes dos arquivos de entrada e saída de dados e uma relação de dependência entre as atividades. A partir dessa definição de dependências é possível construir as principais estruturas de fluxos de dados, *pipelines*, distribuição, agregação e redistribuição de dados. O DAX também permite a definição de vários arquivos de saída que podem ser de tipos diferentes, dependendo da necessidade da atividade [Teixeira, 2013].

Código abaixo mostra como é representada uma atividade na lin-

guagem DAX⁴:

Código- 1 – Exemplo de uma atividade DAX

```
1
2 <job id="ID000002" namespace="vahi" name="findrange"
   version="1.0"
3 level="2" dv-namespace="vahi" dv-name="left" dv-version="
   1.0">
4   <argument>
5     -a left -T 6 -i
6     <filename file="f.b1"/> -o <filename file="f.c1"/>
7     -p 0.5
8   </argument>
9   <uses file="f.b1" link="input" dontRegister="false"
      dontTransfer="false" temporaryHint="true"/>
10  <uses file="f.c1" link="output" dontRegister="true"
      dontTransfer="false" temporaryHint="true"/>
11 </job>
```

Numa próxima fase o DAX é mapeado pelo Pegasus e as atividades definidas são executadas.

3.3 Redes de Petri

As Redes Petri(RdP) é um dos formalismos mais utilizados para modelagem analítica e análise de sistemas concorrentes mais utilizados, isso deve-se à fácil compreensão representação gráfica e devido ao seu potencial matemático para análise de modelos. Algumas características importantes de workflows científicos pôde ser facilmente representadas utilizando RdP, tais como a sincronização de tarefas, concorrência e controle de dependências[Braghetto, 2011]. Uma RdP pode ser formalmente definida pela tupla $RdP = L, T, A, P, m_0$, em que:

- $L = \{l_1, l_2, \dots, l_L\}$ é um conjunto de lugares;
- $T = \{t_1, t_2, \dots, t_T\}$ é um conjunto de transições;
- $A \subseteq (LxT) \cup (TxL)$ é um conjunto de arcos;
- $P : A \rightarrow \mathbb{N}$ é a função peso dos arcos;
- $m_0 = \{m_{01}, m_{02}, \dots, m_{0L}\}$ é a marcação inicial da rede;

Cada um dessas definições está associados a diferentes e importantes conceitos de RdP, que são:

⁴ <https://confluence.pegasus.isi.edu/display/pegasus/WorkflowGenerator>, Este exemplo foi retirado do DAX que representa o Montage

- *lugar* (representado graficamente por um círculo)- modela uma condição que deve ser satisfeita para que o disparo da transição seja realizado;
- *transição* (representado graficamente por um retângulo) - pode ser compreendido por uma ação ou um evento;
- *arco orientado* liga um lugar a uma transição ou o contrário, ligando condições e eventos;
- *fichas, marcas ou tokens* (representado por um ponto preto)- Representam um recurso ou um estado de um sistema;
- *peso* - os arcos possuem um peso associado a eles; os pesos indicam quantas fichas uma transição consome de um lugar de entrada ou quantas fichas uma transição acrescenta em um lugar de saída. Por *default* quando um arco não possui um peso o peso é 1;

Da necessidade de modelagem de sistemas grandes e complexos, como os encontrados na natureza e na sociedade, foram definidas sobre a base de Redes de Petri clássica outras redes de alto nível, como as Redes de Petri Temporizadas(RdPTs) e Redes de Petri Estocásticas(RdPEs). As RdPTs como o nome diz permite modelar situações que ocorrem entre o início e fim de cada tarefa em termos de tempos. O permitiu diferentes possibilidades para modelagem, como lugares temporizados, fichas temporizadas e todas as outras estruturas descritas acima puderam ser temporizadas e permitiram modelar uma série de sistemas reais. Já as RdPEs é uma subclasse das RdPTs, onde o tempo de atraso de uma transição é uma variável aleatória exponencialmente distribuída[Braghetto, 2011].

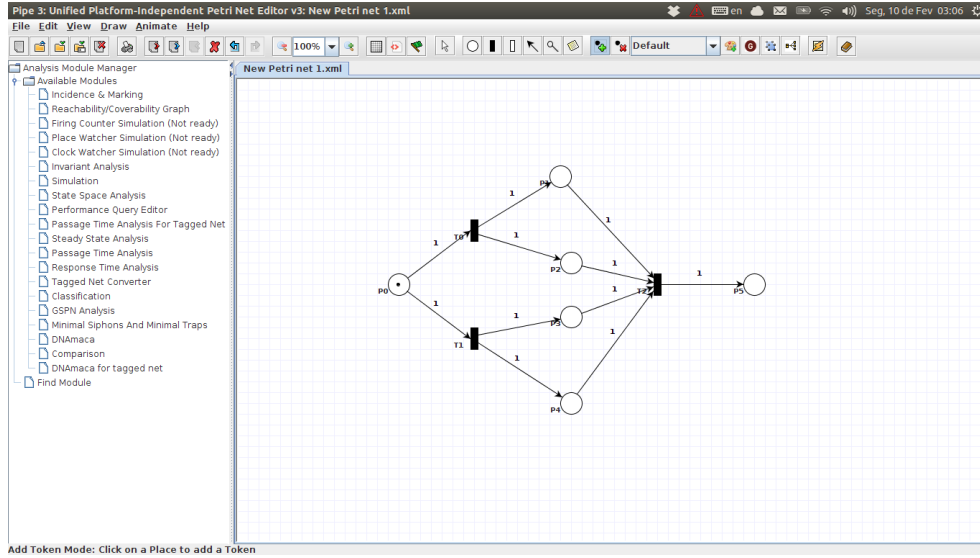
Ao se usar as redes de Petri para se modelar a execução de um workflow, cada vértice pode armazenar um ou mais tokens, os vértices de transição de redes de Petri podem consumir e produzir tokens de múltiplas localizações. Os vértices de transições agem em tokens de entrada por um processo denominado disparo. Quando uma transição é disparada, ela consome os tokens de suas posições de entrada, realiza alguma tarefa de processamento, e realoca um número específico de tokens nas suas posições de saída. Isso é feito atomicamente. Como os disparos são não-determinísticos, as redes de Petri são muito utilizadas para modelar comportamento concorrente em sistemas distribuídos.

3.3.1 PIPE

O PIPE[[of Computing at Imperial College London, 2002](#)] é uma ferramenta *open source* criada para análise de RdPEs, possui uma interface gráfica simples e intuitiva. Com um conjunto de módulos para

avaliar o comportamento estatístico das RdPEs, a representação de Lugares, Transições, Arcos e Fichas, segue o modelo gráfico descrito na seção anterior, como pode ser visto na figura abaixo:

Figura 2 – Interface gráfica-PIPE



Após a construção gráfica do modelo é possível observar a evolução dinâmica do sistema modelado e fazer a análise dos dados gerados, assim como exportar em formato XML ou o PNG da Rede Petri representada.

3.4 Simuladores de workflows científicos

Depois de criados os modelos em RdPEs para workflows científicos há algumas maneiras de validar o modelo, a primeira possibilidade é execução do workflows. Essa alternativa possui como problema o custo monetário envolvido na instanciação das máquinas, o que a torna inviável no primeiro momento. Uma outra alternativa mais viável é a utilização de um ambiente de simulação computacional. Essa alternativa abre a possibilidade de criar-se diferentes ambientes computacionais e facilidade de reprodução da simulação, o que torna viável a busca por gargalos no workflows utilizados. Além da possibilidade de ter um avaliar modelos analíticos em um ambiente controlado que é um dos objetivos desse trabalho.

Os simuladores possivelmente serão utilizados nesse trabalho estão brevemente descritos abaixo:

3.4.1 WorkflowSim

O WorkflowSim[Chen and Deelman, 2012] é uma ferramenta de simulação de workflows científicos baseada no CloudSim[Google, 2013], que também é uma ferramenta de simulação. No entanto, o CloudSim não possui certas características presentes no WorkflowSim, por exemplo, a simulação de ambientes heterogêneos e possibilidade de incluir falhas no sistema. As atividades no WorkflowSim são descritas na linguagem DAX(XML gerado para execução de atividades no Pegasus). O WorkflowSim, possibilita também, modelar todo um ambiente computacional para execução dos workflows(Quantidade de *Virtual Machine(VM)* e *DataCenters*). Observe no Anexo 3 como é descrito um ambiente computacional no WorkflowSim.

4 Metodologia

Nessa primeira fase do projeto foi realizado um estudo sobre os conceitos básicos envolvendo workflows, workflows científicos, Sistemas Gerenciadores de Workflows Científicos(SGWC), Redes de Petri, ferramentas de representação de redes de Petri, modelagem analítica de workflows científicos(como o Montage) e Simuladores de workflows científicos. Depois do estudo da bibliografia básica envolvendo workflow científicas e as motivações acadêmicas para a pesquisa de workflows científicas, foi feito o estudo de alguns SGWC, quanto a implementação de workflows e a representação do fluxo de dados em cada um deles.

Para aparar o formalismo matemático envolvendo workflows científicos e fazer uma modelagem analítica de alguns workflows bem conhecidos foram estudadas Redes de Petri, Redes de Petri Temporizadas e Redes de Petri Estocásticas, assim como uma ferramenta para representação gráfica e análise das RdPEs (PIPE).

Para finalizar a primeira parte do trabalho foi feito um estudo do WorkFlowSim, quanto possibilidade de ferramenta de simulação para o projeto.

5 Resultados Parciais

Na primeira etapa desse projeto foram obtidos conhecimentos básico de workflows científicos, SGWC, RdPEs, modelagem analítica de workflow científicos. Para uma melhor interpretação foi feita a modelagem analítica do fluxo de dados do Montage, partir do Anexo 1, [Pegasus Project, 2012] com o auxílio PIPE. Como pode vista na figura

do Anexo 2, que permitirá extrair índices de desempenho e verificar a validade do modelo a partir da simulação do Montage utilizando o WorkflowSim.

6 Conclusões Parciais

Todas as tarefas previstas no cronograma projeto foram realizadas com sucesso e estão listadas abaixo, assim como as atividades em andamento e as atividades futuras:

6.1 Tarefas Realizadas

- Estudo dos conceitos básicos relacionados a workflows científicos e seus respectivos modelos de representação de fluxo de dados;
- Estudo dos conceitos básicos relacionados à modelagem estocástica de workflows científicos;
- Identificação e caracterização do conjunto de construtores de fluxo de dados: Kepler, Taverna e Pegasus.
- Modelagem do Montage utilizando o PIPE;
- Estudo básico dos possíveis simuladores de workflows científicos para validar o modelo analítico criado.

6.2 Tarefas em Andamento

- **2 meses** - Estudo da viabilidade do uso de diferentes tipos de formalismos estocásticos (em particular Redes de Petri e as Álgebra de processos estocásticos) para a modelagem de fluxos de dados;
- **15 dias** - Estudo do funcionamento dos simuladores de workflows: WorkflowSim e Cloud-DVFS, quanto possibilidade de representação um ambiente computacional que possa ser modelado analiticamente;

6.3 Tarefas a serem realizadas

- **15 dias** - Estudo e definição de índices extraídos a partir dos modelos representados na etapa anterior, que possam refletir um impacto no desempenho do fluxo de dados;

- **15 dias**- Montar uma simulação de workflow científico em um dos simuladores estudado na etapa anteriores;
- **1,5 meses**-Desenvolvimento de um algoritmo que faça a conversão automática de fluxo de dados de um fluxo de dados de workflow em um modelo estocástico correspondente(selecionado dentre os tipos de modelos estudados).

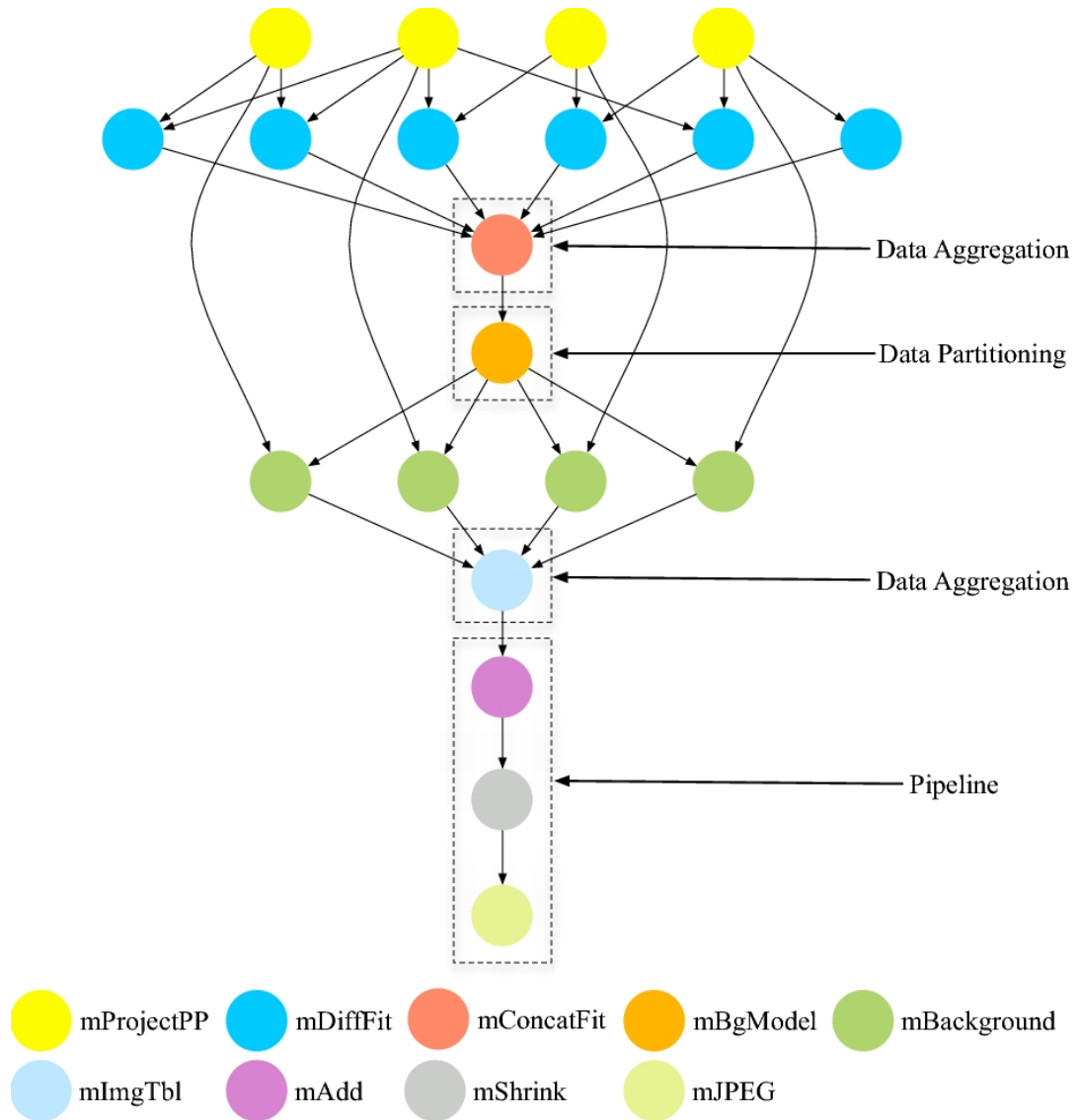
Referências

- [WfM, 2014] (2014). Workflow management coalition. <http://www.wfmc.org/>. Último acesso: 2014-02-06.
- [Braghetto, 2011] Braghetto, K. (2011). Técnicas de modelagem para a análise de desempenho de processos de negócio. Doutorado, Universidade de São Paulo, USP-São Paulo.
- [Chen and Deelman, 2012] Chen, W. and Deelman, E. (2012). Workflowsim: A toolkit for simulating scientific workflows in distributed environments. In *E-Science (e-Science), 2012 IEEE 8th International Conference on*, pages 1–8.
- [Google, 2013] Google (2013). Calheiros: Cloudsim: a toolkit for modeling and simulation of ... - google scholar. http://scholar.google.com.br/scholar?cites=272054103506592999&as_sdt=2005&sciodt=0,5&hl=pt-BR. [Online; acessado em 01 de fevereiro de 2014].
- [Júnior, 2012] Júnior, L. (2012). Gerência de proveniência em workflow científicos para e distribuídos. Doutorado, Universidade Federal do Rio de Janeiro, UFRJ-Rio de Janeiro.
- [of Computing at Imperial College London, 2002] of Computing at Imperial College London, D. (2002). Ferramenta de análise e representação de redes de petri. <http://pipe2.sourceforge.net/>. [Online; acessado em 01 de fevereiro de 2014].
- [Ogasawara, 2011] Ogasawara, E. S. (2011). Uma abordagem algébrica para workflows científicos com dados em larga escala. Doutorado, Universidade Federal do Rio de Janeiro, UFRJ-Rio de Janeiro.
- [Pegasus Project, 2012] Pegasus Project (2012). WorkflowGenerator. <https://confluence.pegasus.isi.edu/display/pegasus/WorkflowGenerator>. [Online; acessado em 1 de setembro de 2013].
- [Teixeira, 2013] Teixeira, E. (2013). Modelagem e escalonamento de workflows científicos para execução em plataforma de computação em nuvem. Doutorado-qualificação, Universidade de São Paulo, USP-São Paulo.

7 Anexos

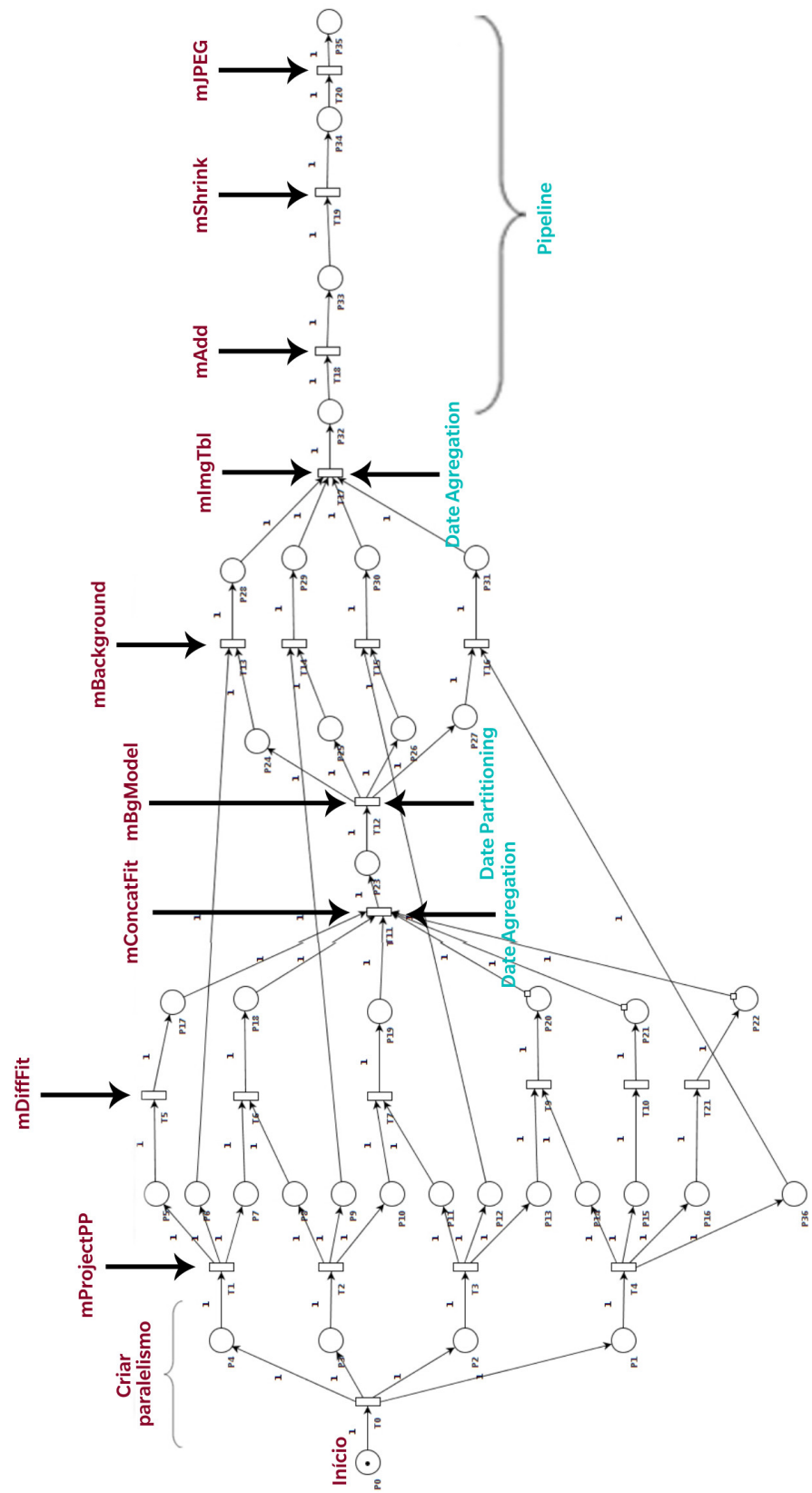
7.1 Anexo 1

Figura 3 – Ilustração do workflow do Montage



7.2 Anexo 2

Figura 4 – Representação Gráfica do Montagem usando Redes de Petri



7.3 Anexo 3

Código- 2 – Descrição de uma ambiente computacional no WorkflowSim

```
1
2 #####Overhead Related Parameters
3 #####The Workflow Engine Delay for level 1
4 #d1      = 8.1
5 #####The Workflow Engine Delay for level 2
6 #d2      = 18.44
7 #####The Workflow Engine Delay for other levels that are
      not specified
8 #d0      = 6.74
9 #####The Postscript Delay for level 1
10 #p1     = 78.05
11 #####The Postscript Delay for level 2
12 #p2     = 13.12
13 #####The Postscript Delay for other levels that are not
      specified
14 #p0     = 6.81
15 #####The Queue Delay for level 1
16 #q1     = 77.75
17 #####The Queue Delay for level 2
18 #q2     = 343.47
19 #####The Queue Delay for other levels that are not
      specified
20 #q0     = 161.96
21 #####The Clustering Delay for level 1
22 #c1     = 428.28
23 #####The Clustering Delay for level 2
24 #c2     = 67.44
25 #####The interval for Workflow Engine
26 #interval = 5
27
28 #####Failure related parameters
29 #####The Task Failure Rate for level 1
30 #a1     = 0.02
31 #####The Task Failure Rate for level 2
32 #a2     = 0.02
33 #####Fault tolerant clustering method
34 #ftc.method = FTCLUSTERING_DC
35 #####Fault tolerant clustering monitor mode
36 #ftc.monitor = MONITOR_ALL
37 #####Fault tolerant clustering failure generation mode
38 #ftc.failure = FAILURE_JOB
39
40 #####Clustering Related parameters
41 #####The number of clustered jobs per level
42 #clusters.num = 20
43 #####The size of a clustered job (number of tasks per job
      )
44 #clusters.size = 20
45 #####The clustering method
```

```

46 #clusters.method = horizontal
47 #clusters.method = balanced
48 #####The type of file system to use
49 file.system = LOCAL
50 #file.system    = SHARED
51
52 #####Scheduling Related Parameters
53 #####If you have specified planner.method, it will be
    disabled
54 scheduler.method= MINMIN_SCH
55
56 #####Planning Related Parameters
57 #planner.method= RANDOM
58 #deadline = 100000
59
60 #####Please replace it with your real physical path
61 dax.path = /home/duilio/workspace/WorkflowSim-1.0-master/
    config/dax/Montage_1000.xml
62 #####Reducer mode used to remove duplicate dependencies
63 #reduce.method = montage

```