

Workflow Científico - Tema 5

Tópicos em Banco de Dados (INFO-7058)

Gabriel Silva Hermida (gash18@inf.ufpr.br)
Jefferson Luis Stefani Alves Pereira (jlsap17@inf.ufpr.br)
Josiney de Souza (josiney.souza@ifc.edu.br)
Lucas Sidnei dos Santos (lss16@inf.ufpr.br)

UFPR / DInf

10 de maio de 2023

Sumário

- 1 Introdução
 - Workflow
 - Workflow Científico
 - Provenance
- 2 Bharathi et al., (2008)
- 3 Cuevas-Vicenttin et al., (2012)
- 4 Silva et al. (2013)
- 5 Mattoso et al. (2015)
- 6 Miksa, Rauber, & Mina (2016)
- 7 Comentários/Impressões/Consolidações de conhecimento

Sumário

1 Introdução

- Workflow
- Workflow Científico
- Provenance

2 Bharathi et al., (2008)

3 Cuevas-Vicenttin et al., (2012)

4 Silva et al. (2013)

5 Mattoso et al. (2015)

6 Miksa, Rauber, & Mina (2016)

7 Comentários/Impressões/Consolidações de conhecimento



Workflow

Workflow é:

- Fluxo de trabalho
- Automação
- Controle
- Workflow Management System (WfMS)
- Sistema de Gerenciamento de Workflow

WfMS

Um sistema de gerenciamento de fluxo de trabalho fornece uma infraestrutura para a configuração, desempenho e monitoramento de uma sequência definida de tarefas, organizada como um aplicativo de fluxo de trabalho.

Fonte: https://en.wikipedia.org/wiki/Workflow_management_system

Exemplos

Exemplos de workflows:

- Apache Ant
- Make
- Monday.com
- Pipefy
- Adobe Workfront
- Trello

Fontes:

- <https://thedigitalprojectmanager.com/tools/workflow-management-software/>

- <https://hive.com/blog/workflow-management-software/>

-

<https://www.softwaretestinghelp.com/workflow-management-software/>

- <https://www.proofhub.com/articles/workflow-management-software>

Workflow Científico

Workflow científico é:

- Fluxo de trabalho **científico**
- Automação
- Controle
- **Scientific** Workflow Management System (SWfMS)
- Sistema de Gerenciamento de Workflow **Científico**
- Modelar e executar experimentos científicos de larga escala

SWfMS

Um sistema de fluxo de trabalho científico é uma forma especializada de um sistema de gerenciamento de fluxo de trabalho projetado especificamente para compor e executar uma série de etapas computacionais ou de manipulação de dados, ou fluxo de trabalho, em um aplicativo científico.

Fonte: https://en.wikipedia.org/wiki/Scientific_workflow_system

Exemplos

Exemplos de workflows **científicos**:

- Apache Airflow
- Prefect
- Kubeflow
- MLFlow
- Kepler
- VisTrails
- Taverna

Fonte 1: descrição da “Palestra: ”Data science pipelines from head to toes- Profa. Genoveva Vargas-Solar”, de 21/03/23, recebida por e-mail do SIGA

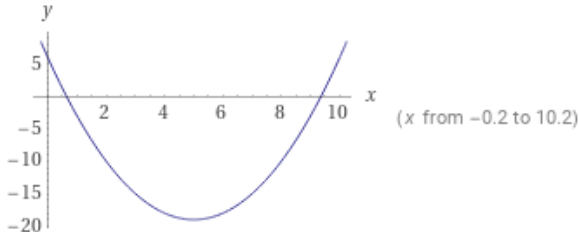
Fonte 2: Tomasz Miksa, Andreas Rauber, Eleni Mina - Identifying impact of software dependencies on replicability of biomedical workflows

Exemplo de Workflow

Função: $x^2 - 10x + 6$

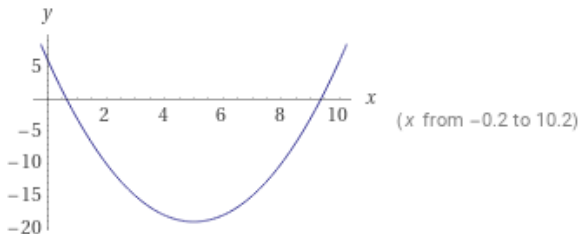
Raiz 1: $x = 5 - \text{sqrt}(19) =$
0.641101056459326447763018016140384340862996074767555063109

Raiz 2: $x = 5 + \text{sqrt}(19) =$
9.358898943540673552236981983859615659137003925232444936890



Fonte: <https://www.wolframalpha.com/input?i=x%5E2+-10x+%2B6>

Exemplo de Workflow



Fonte: <https://www.wolframalpha.com/input?i=x%5E2+-10x+%2B6>

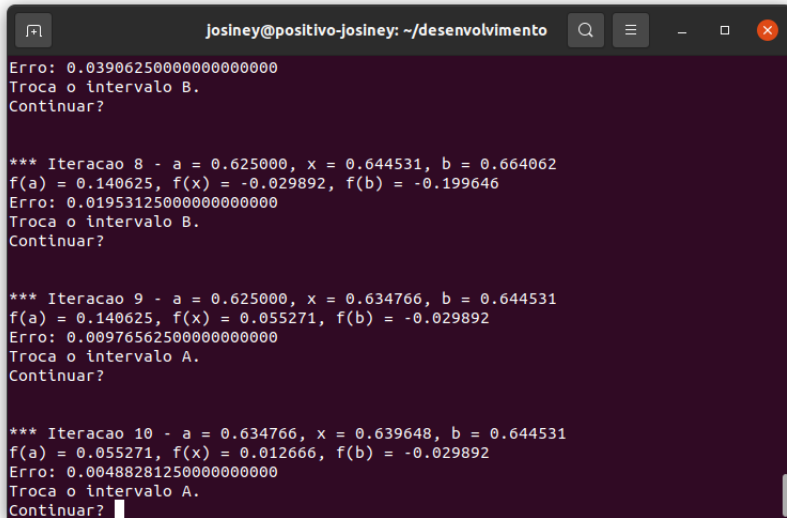
Disciplina: Métodos Numéricos (CI-181 -
<https://web.inf.ufpr.br/dinf/disciplinas-atuais/>)

Cálculo: Zero de função

Método: Bisseção

Intervalo: 0 a 5 (raiz 1)

Exemplo de Workflow



```
josiney@positivo-josiney: ~/desenvolvimento

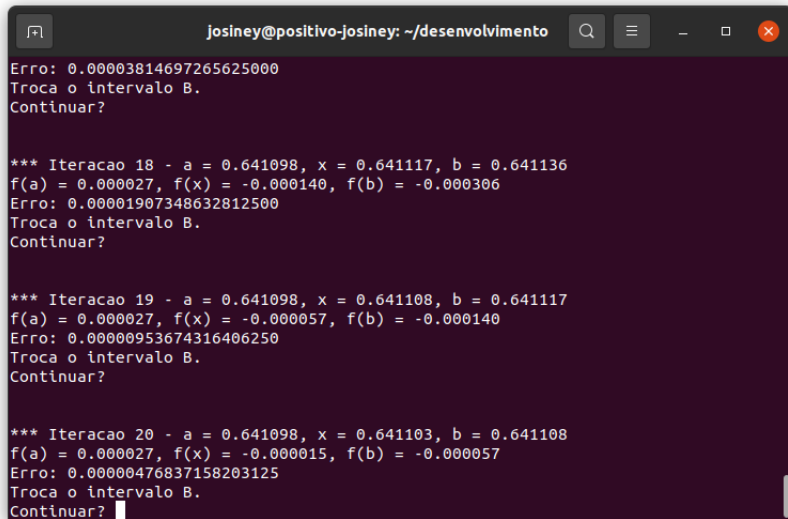
Erro: 0.039062500000000000
Troca o intervalo B.
Continuar?

*** Iteracao 8 - a = 0.625000, x = 0.644531, b = 0.664062
f(a) = 0.140625, f(x) = -0.029892, f(b) = -0.199646
Erro: 0.019531250000000000
Troca o intervalo B.
Continuar?

*** Iteracao 9 - a = 0.625000, x = 0.634766, b = 0.644531
f(a) = 0.140625, f(x) = 0.055271, f(b) = -0.029892
Erro: 0.009765625000000000
Troca o intervalo A.
Continuar?

*** Iteracao 10 - a = 0.634766, x = 0.639648, b = 0.644531
f(a) = 0.055271, f(x) = 0.012666, f(b) = -0.029892
Erro: 0.004882812500000000
Troca o intervalo A.
Continuar? 
```

Exemplo de Workflow



```
josiney@positivo-josiney: ~/desenvolvimento

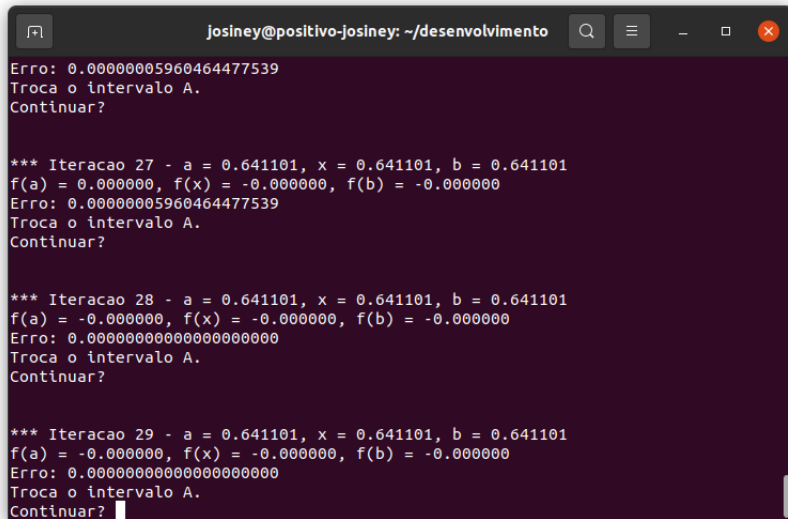
Erro: 0.00003814697265625000
Troca o intervalo B.
Continuar?

*** Iteracao 18 - a = 0.641098, x = 0.641117, b = 0.641136
f(a) = 0.000027, f(x) = -0.000140, f(b) = -0.000306
Erro: 0.00001907348632812500
Troca o intervalo B.
Continuar?

*** Iteracao 19 - a = 0.641098, x = 0.641108, b = 0.641117
f(a) = 0.000027, f(x) = -0.000057, f(b) = -0.000140
Erro: 0.00000953674316406250
Troca o intervalo B.
Continuar?

*** Iteracao 20 - a = 0.641098, x = 0.641103, b = 0.641108
f(a) = 0.000027, f(x) = -0.000015, f(b) = -0.000057
Erro: 0.00000476837158203125
Troca o intervalo B.
Continuar? 
```

Exemplo de Workflow



```
josiney@positivo-josiney: ~/desenvolvimento
Erro: 0.00000005960464477539
Troca o intervalo A.
Continuar?

*** Iteracao 27 - a = 0.641101, x = 0.641101, b = 0.641101
f(a) = 0.000000, f(x) = -0.000000, f(b) = -0.000000
Erro: 0.00000005960464477539
Troca o intervalo A.
Continuar?

*** Iteracao 28 - a = 0.641101, x = 0.641101, b = 0.641101
f(a) = -0.000000, f(x) = -0.000000, f(b) = -0.000000
Erro: 0.00000000000000000000
Troca o intervalo A.
Continuar?

*** Iteracao 29 - a = 0.641101, x = 0.641101, b = 0.641101
f(a) = -0.000000, f(x) = -0.000000, f(b) = -0.000000
Erro: 0.00000000000000000000
Troca o intervalo A.
Continuar? 
```

Provenance Data

Provenance data é:

- Procedência dos dados
- Proveniência dos dados
- Linhagem
- Histórico
- Pedigree
- Árvore genealógica
- Rastreio
- Registro
- Log

Provenance Data

Provenance Data

No contexto dos fluxos de trabalho científicos, proveniência geralmente significa a linhagem e o histórico de processamento de um produto de dados e o registro dos processos que levaram a ele.

A proveniência captura o design do fluxo de trabalho e o histórico de execução.

A proveniência ajuda a rastrear as entradas e saídas do fluxo de trabalho, processos e pontos de interseção de dados, para que os experimentos possam ser verificados, repetidos e, quando possível, reproduzidos de maneira precisa.

A proveniência também permite a comparação entre diferentes versões de fluxo de trabalho, reexecuções inteligentes e recuperação de falhas.

Fonte: <https://words.sdsc.edu/words-data-science/provenance>

Sumário

- 1 Introdução
 - Workflow
 - Workflow Científico
 - Provenance
- 2 Bharathi et al., (2008)
- 3 Cuevas-Vicenttin et al., (2012)
- 4 Silva et al. (2013)
- 5 Mattoso et al. (2015)
- 6 Miksa, Rauber, & Mina (2016)
- 7 Comentários/Impressões/Consolidações de conhecimento



Texto 1a

Título: Characterizing of Scientific Workflows

Autores: Shishir Bharathi, Ann Chervenak, Ewa Deelman, Gaurang Mehta, Mei-Hui Su, Karan Vahi

Origem: University of Southern California Information Sciences Institute

Publicado em: Third Workshop on Workflows in Support of Large-Scale Science

Ano: 2008

Editora: IEEE

Link: <https://ieeexplore.ieee.org/abstract/document/4723958>

Texto 1b

Título: Characterizing and Profiling Scientific Workflows

Autores: Gideon Juve, Ann Chervenak, Ewa Deelman, Shishir Bharathi, Gaurang Mehta, Karan Vahi

Origem: University of Southern California Information Sciences Institute

Publicado em: Future Generation Computer Systems

Ano: 2013

Editora: Elsevier

Link: <https://www.sciencedirect.com/science/article/pii/S0167739X12001732>

O problema

No desenvolvimento do texto, é possível ver que ...

① **introduz o problema**

- Pesquisadores necessitam de acesso a workflows para seus trabalhos
- Há diversas ferramentas para diferentes objetivos
- É difícil haver uma caracterização de cada um deles
 - Tipos/Perfil de tarefas
 - Tarefas que consomem mais tempo
- Benchmark

② mostra uma visão geral dos workflows e cita exemplos

③ caracteriza workflows

④ caracteriza os workflows de exemplos

⑤ apresenta o gerador de workflow

⑥ discute brevemente os trabalhos relacionados

⑦ conclui o trabalho

Exemplos de workflows

No desenvolvimento do texto, é possível ver que ...

- 1 introduz o problema
- 2 **mostra uma visão geral dos workflows e cita exemplos**

Montage: gerar mosaicos personalizados do céu

CyberShake: caracterizar riscos de terremotos em uma região

Broadband¹: simular e calcular terremotos para engenheiros sísmicos

Epigenomics: automatizar operações de sequenciamento genômico

LIGO Inspiral Analysis Workflow: analisar dados de estrelas de neutron binárias e buracos negros

SIPHT: buscar sRNAs para replicons bacterianos

- 3 caracteriza workflows
- 4 caracteriza os workflows de exemplos
- 5 apresenta o gerador de workflow
- 6 discute brevemente os trabalhos relacionados
- 7 conclui o trabalho

¹Apenas no trabalho de 2013

Estruturas básicas de workflows

No desenvolvimento do texto, é possível ver que ...

- ① introduz o problema
- ② mostra uma visão geral dos workflows e cita exemplos
- ③ **caracteriza workflows**
 - Processo
 - Pipeline
 - Distribuição de dados
 - Agregação de dados
 - Redistribuição de dados
- ④ caracteriza os workflows de exemplos
- ⑤ apresenta o gerador de workflow
- ⑥ discute brevemente os trabalhos relacionados
- ⑦ conclui o trabalho

Estruturas básicas de workflows

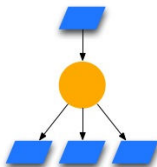
Process



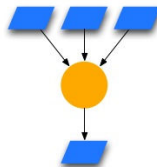
Pipeline



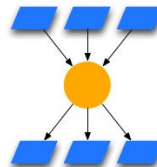
Data Distribution



Data Aggregation



Data Redistribution



Job



Data

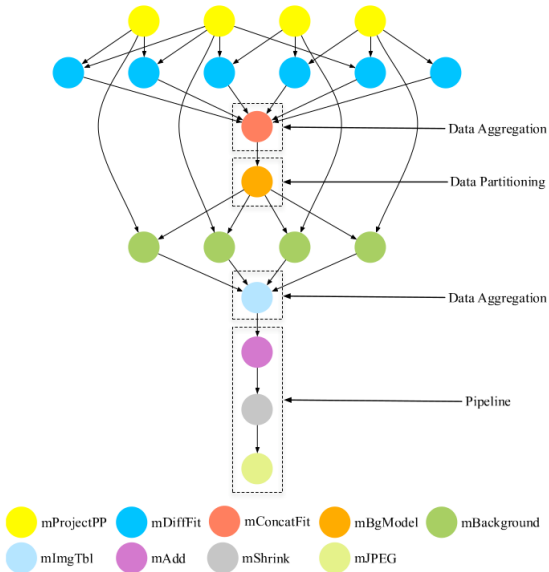
Fonte: Bharathi et al. (2008)

Características de cada workflow

No desenvolvimento do texto, é possível ver que ...

- ① introduz o problema
- ② mostra uma visão geral dos workflows e cita exemplos
- ③ caracteriza workflows
- ④ **caracteriza os workflows de exemplos**
 - Grafo de estruturas
 - Nome das tarefas
 - Número das tarefas
 - Tempo de execução
 - Entradas
 - Saídas
- ⑤ apresenta o gerador de workflow
- ⑥ discute brevemente os trabalhos relacionados
- ⑦ conclui o trabalho

Características de cada workflow



Características de cada workflow

Table 1. Example of a Montage Execution Profile

Executed on viz-cluster at ISI - 8 nodes with dual Intel Xeon 2.4 GHz processors, 2 GB Memory

Job	Count	Runtime		Inputs		Mean(MB)	Outputs	Variance
		Mean(s)	Variance	Mean(MB)	Variance			
mProject	45	13.59	6.00e-02	4.03	0	7.94		4.14e-04
mDiffFit	107	10.59	1.00e-02	15.88	1.37e-03	0.54		7.92e-02
mConcatFit	1	13.60	0	0.03	0	0.02		0
mBgModel	1	10.88	0	0.03	0	0.00		0
mBackground	45	10.74	3.00e-02	7.95	4.14e-04	7.94		4.14e-04
mImgtbl	1	10.69	0	357.27	0	0.01		0
mAdd	1	30.34	0	357.28	0	330.86		0
mShrink	1	12.26	0	165.43	0	6.62		0
mJPEG	1	10.96	0	6.62	0	0.32		0

Fonte: Bharathi et al. (2008)

Gerador de workflow

No desenvolvimento do texto, é possível ver que ...

- ① introduz o problema
- ② mostra uma visão geral dos workflows e cita exemplos
- ③ caracteriza workflows
- ④ caracteriza os workflows de exemplos
- ⑤ **apresenta o gerador de workflow**
 - Gerador sintético de workflow
 - Expertise em analisar workflows reais
 - Estima a quantidade de tarefas, de entradas e de saídas
 - Parâmetro de comparação
- ⑥ discute brevemente os trabalhos relacionados
- ⑦ conclui o trabalho

Sumário

- 1 Introdução
 - Workflow
 - Workflow Científico
 - Provenance
- 2 Bharathi et al., (2008)
- 3 Cuevas-Vicenttin et al., (2012)
- 4 Silva et al. (2013)
- 5 Mattoso et al. (2015)
- 6 Miksa, Rauber, & Mina (2016)
- 7 Comentários/Impressões/Consolidações de conhecimento

Texto 2

Título: Scientific Workflows and Provenance: Introduction and Research Opportunities

Autores: Víctor Cuevas-Vicentín, Saumen Dey, Sven Köhler, Sean Riddle, Bertram Ludäscher

Origem: Espanha, USA, ...

Publicado em: Datenbank Spektrum

Ano: 2012

Editora: Springer

Link: <https://link.springer.com/article/10.1007/s13222-012-0100-z>

O problema

No desenvolvimento do texto, é possível ver que ...

1 **introduz o problema**

- Comunidade científica de BD estudou workflows para negócios
- Workflows científicos foram deixados de lado pela comunidade de BD
- Introduz o assunto de workflows científicos e provenance
- Tarefas repetitivas são feitas com scripts
- Dificuldade em gerenciar manualmente grandes volumes de dados
- Escalabilidade e tolerância a falhas
- Reprodutibilidade, provenance, debug

2 **mostra aspectos de workflows científicos**

3 **explica provenance e a relação com workflows científicos**

4 **sumariza e conclui o texto**

Diferentes aspectos

No desenvolvimento do texto, é possível ver que ...

- 1 introduz o problema
- 2 **mostra aspectos de workflows científicos**

Modelos de computação: Um workflow executado com parâmetros pré-definidos em uma base de dados para gerar os dados resultantes

- Redes de processo, Dataflow síncrono, DAG

Execução de workflow: distribuição, paralelismo

Design de workflow: fácil e rápido para salvar ciclos humanos
OBS.: De 560 workflows científicos, 30% (168) são adaptações de soluções independentes

Gerenciamento e reuso: publicidade, versionamento, gerenciamento de visões, configuração, gerenciamento de contexto

- 3 explica provenance e a relação com workflows científicos
- 4 **sumariza e conclui o texto**

Provenance e Workflows

No desenvolvimento do texto, é possível ver que ...

- 1 introduz o problema
- 2 mostra aspectos de workflows científicos
- 3 **explica provenance e a relação com workflows científicos**

Modelos de Provenance: RWS (read, write, state-reset); Open Provenance Model (OPM); W3C PROV

Obtenção de Provenance: sobrecarga mínima

Armazenamento e consultas: acessibilidade (reachability) e caminho (RPQ)

- “que itens de dados estão envolvidos na produção de um certo resultado parcial?”
- “este ator usou saídas de algum desses outros dois atores?”

Interoperabilidade: integração de provenances de outros sistemas com formatos distintos

Aplicações/Usos: ciência de privacidade; sobrecarga de informação; debug; tolerância a falhas

- 4 **sumariza e conclui o texto**

Sumário

- 1 Introdução
 - Workflow
 - Workflow Científico
 - Provenance
- 2 Bharathi et al., (2008)
- 3 Cuevas-Vicenttin et al., (2012)
- 4 Silva et al. (2013)
- 5 Mattoso et al. (2015)
- 6 Miksa, Rauber, & Mina (2016)
- 7 Comentários/Impressões/Consolidações de conhecimento



Texto 5

Título: Capturing and Querying Workflow Runtime Provenance with PROV: a Practical Approach

Autores: Flavio Costa, Vítor Silva, Daniel de Oliveira, Kary Ocaña, Eduardo Ogasawara, Jonas Dias, Marta Mattoso

Origem: UFRJ, CEFET-RJ

Publicado em: EDBT '13: Proceedings of the Joint EDBT/ICDT 2013 Workshops

Ano: 2013

Editora: ACM

Link: <https://dl.acm.org/doi/abs/10.1145/2457317.2457365>

Problemas

No desenvolvimento do texto, é possível ver que ...

❶ introduz o problema e possível solução

- Experimentos são heterogêneos para executar em apenas um SWfMS
- “Big data” → “Big provenance”
- Provenance in runtime (tempo de execução)
- Apenas Chiron permite pesquisar; apenas SciCumulus tem suporte
- Uso do PROV para manter uma base de consulta sobre provenance

❷ explica o porquê de usar dados de provenance em tempo de execução

❸ detalha modelo de provenance e coleta distribuída

❹ apresenta um estudo de caso em bioinformática com 2 SWfMS

❺ discute trabalhos relacionados

❻ conclui o trabalho e apresenta trabalhos futuros

❼ agradece os fomentos

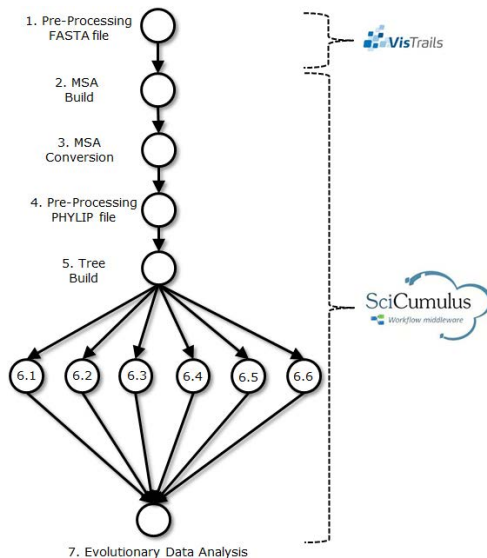
❽ referencia o trabalho

Porquê usar

No desenvolvimento do texto, é possível ver que ...

- ① introduz o problema e possível solução
- ② **explica o porquê de usar dados de provenance em tempo de execução**
 - Tratamento de falhas
 - Monitoramento de workflow
 - Escalonamento adaptativo
- ③ detalha modelo de provenance e coleta distribuída
- ④ apresenta um estudo de caso em bioinformática com 2 SWfMS
- ⑤ discute trabalhos relacionados
- ⑥ conclui o trabalho e apresenta trabalhos futuros
- ⑦ agradece os fomentos
- ⑧ referencia o trabalho

Modelo de Provenance em tempo de execução

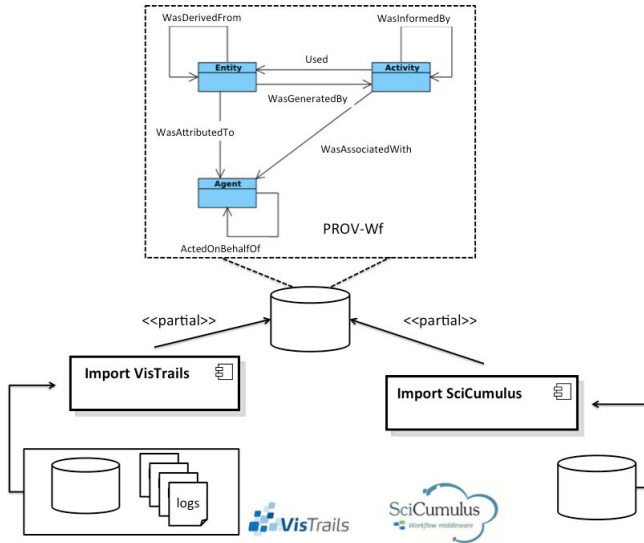


Conclusões

No desenvolvimento do texto, é possível ver que ...

- ① introduz o problema e possível solução
- ② explica o porquê de usar dados de provenance em tempo de execução
- ③ detalha modelo de provenance e coleta distribuída
- ④ apresenta um estudo de caso em bioinformática com 2 SWfMS
- ⑤ discute trabalhos relacionados
- ⑥ **conclui o trabalho e apresenta trabalhos futuros**
 - Modelo de dados em provenance baseado em PROV
 - Coleta em tempo de execução, em paralelo e em ambientes diferentes
 - Criado cartuchos para 2 Workflows (VisTrails e SciCumulus)
 - Testar em outros SWfMS e fazer uma API
- ⑦ agradece os fomentos
- ⑧ referencia o trabalho

Cartuchos



Sumário

- 1 Introdução
 - Workflow
 - Workflow Científico
 - Provenance
- 2 Bharathi et al., (2008)
- 3 Cuevas-Vicenttin et al., (2012)
- 4 Silva et al. (2013)
- 5 Mattoso et al. (2015)**
- 6 Miksa, Rauber, & Mina (2016)
- 7 Comentários/Impressões/Consolidações de conhecimento

Texto 3

Título: Dynamic steering of HPC scientific workflows: A survey

Autores: Marta Mattoso, Jonas Dias, Kary A.C.S. Ocaña, Eduardo Ogasawara, Flavio Costa, Felipe Horta, Vítor Silva, Daniel de Oliveira

Origem: UFRJ, CEFET-RJ, UFF

Publicado em: Future Generation Computer Systems

Ano: 2015

Editora: Elsevier

Link: <https://www.sciencedirect.com/science/article/pii/S0167739X14002519>

Problemas

No desenvolvimento do texto, é possível ver que ...

① **introduz o problema e possível solução**

- execução de forma “off-line” nos workflows
- não há o Human-in-the-Loop (HIL) nas execuções
- execuções podem demorar dias, semanas, meses, ...
- SWfMS são genéricos e possuem proveniência de dados
- em nuvens, os recursos são pagos
- dados ou resultados não guiados
- processos em larga escala e/ou distribuída

② foca em exemplos de workflows data-intensive em Bioinformática

③ traz uma taxonomia

④ mostra abordagens existentes

⑤ apresenta desafios em aberto

⑥ conclui o trabalho

Taxonomia e abordagens existentes

No desenvolvimento do texto, é possível ver que ...

- ① introduz o problema e possível solução
- ② foca em exemplos de workflows data-intensive em Bioinformática
- ③ **traz uma taxonomia**
 - Monitoramento
 - Análise
 - Adaptação
 - Notificação
 - Interface de interação
 - Modelo computacional
- ④ **mostra abordagens existentes**
- ⑤ apresenta desafios em aberto
- ⑥ conclui o trabalho

Desafios em aberto

No desenvolvimento do texto, é possível ver que ...

- ➊ introduz o problema e possível solução
- ➋ foca em exemplos de workflows data-intensive em Bioinformática
- ➌ traz uma taxonomia
- ➍ mostra abordagens existentes
- ➎ **apresenta desafios em aberto**
 - **Na análise em tempo de execução**
 - Preparação de dados: melhorar a preparação de dados
 - Análise *in situ*: fazer *in loco* (geralmente remoto)
 - Ferramentas de suporte à decisão: transformar, correlacionar e analisar
 - **Na adaptação dinâmica**
 - Motores dinâmicos: trocar programas, reexecutar, passos adicionais
 - Fatias de parâmetros: explorar espaços de entradas
 - Otimização de experimento: ajustar os parâmetros e tomar decisões
- ➏ conclui o trabalho

Sumário

- 1 Introdução
 - Workflow
 - Workflow Científico
 - Provenance
- 2 Bharathi et al., (2008)
- 3 Cuevas-Vicenttin et al., (2012)
- 4 Silva et al. (2013)
- 5 Mattoso et al. (2015)
- 6 Miksa, Rauber, & Mina (2016)
- 7 Comentários/Impressões/Consolidações de conhecimento



Texto 4

Título: Identifying impact of software dependencies on replicability of biomedical workflows

Autores: Tomasz Miksa, Andreas Rauber, Eleni Mina

Origem: Austria, The Netherlands

Publicado em: Journal of Biomedical Informatics

Ano: 2016

Editora: Elsevier

Link: <https://www.sciencedirect.com/science/article/pii/S1532046416301484>

Os problemas

No desenvolvimento do texto, é possível ver que ...

1 **introduz o problema e possível solução**

- Baixa reprodutibilidade e replicabilidade
- Preocupação com o contexto de software, infraestrutura e dependências
- Mesmos dados e mesma pilha de software
- Workflows escondem complexidades (infraestrutura e códigos)
- 30% de 1500 (450) workflows do Taverna no myExperiment executam

2 apresenta trabalhos relacionados

3 mostra uma visão geral do VFramework

4 descreve como o VFramework captura a execução do workflow

5 descreve os passos para verificar e validar re-execução

6 apresenta um estudo de caso

7 provê discussão dos resultados e recomendações

8 conclui o trabalho

Conclusões

No desenvolvimento do texto, é possível ver que ...

- ① introduz o problema e possível solução
- ② apresenta trabalhos relacionados
- ③ mostra uma visão geral do VFramework
- ④ descreve como o VFramework captura a execução do workflow
- ⑤ descreve os passos para verificar e validar re-execução
- ⑥ apresenta um estudo de caso
- ⑦ provê discussão dos resultados e recomendações
- ⑧ **conclui o trabalho**
 - É possível descrever o workflow e dependências com o VFramework
 - Workflows automáticos, não padronizados, podem não fazer registros
 - Faz orientações para melhorar a replicabilidade

Sumário

- 1 Introdução
 - Workflow
 - Workflow Científico
 - Provenance
- 2 Bharathi et al., (2008)
- 3 Cuevas-Vicenttin et al., (2012)
- 4 Silva et al. (2013)
- 5 Mattoso et al. (2015)
- 6 Miksa, Rauber, & Mina (2016)
- 7 Comentários/Impressões/Consolidações de conhecimento

Principais pontos dos artigos

De forma geral, os arquivos falam de:

- Workflows científicos (visão geral e exemplos)
- Provenance (“off-line” e em tempo de execução)
- Reprodutibilidade
- Human-in-the-Loop
- Paralelização
- Execução distribuída

Impressões

O que notamos sobre workflows:

- Assim como informática, workflow é um meio para se alcançar um fim
- Banco de dados está presente em provenance
- Para se ter dados abertos, necessário pensar em provenance e reprodutibilidade
- Duas grandes abordagens
 - Execução
 - Provenance
- Brasil tem representatividade nessa área de pesquisa
- Apresentação dos dados como em BI (Business Intelligence)?

Impressões

O que notamos sobre workflows:

- Vindo do trabalho do grupo 4 (GitHub para dados científicos)
 - Crise de reprodutibilidade em 2019
 - FAIR (Interoperabilidade)

Impressões

O que notamos sobre workflows:

- Vindo do trabalho do grupo 4 (GitHub para dados científicos)
 - Crise de reprodutibilidade em 2019
 - FAIR (Interoperabilidade)
- Vindo do trabalho do grupo 7 (Análise comparativa dos repositórios de dados)
 - Os dados dos repositórios podem vir dos workflows
 - Variedade de classificações
 - Preparação de dados

Impressões

O que notamos sobre workflows:

- A Survey of Data-Intensive Scientific Workflow Management - 2014 - Ji Liu, Esther Pacitti, Patrick Valduriez, Marta Mattoso
 - Mesmo dentro da computação, é uma área multidisciplinar
 - BD, paralelismo, escalonamento, redes, sistemas distribuídos, engenharia de software
 - Science gateway framework (WS-PGRADE/gUSE)

Science gateway framework

“um conjunto específico de ferramentas, aplicativos e coleções de dados da comunidade que são integrados por meio de um portal da web ou um aplicativo de desktop, fornecendo acesso a recursos e serviços de infraestruturas de computação distribuídas”

Fonte: <https://doi.org/10.1002/cpe.6872>

- Na plataforma de simulação SHIWA: **Interoperable** Workflow Intermediate Representation (IWIR) language

Próximo trabalho

Possibilidades para o próximo trabalho:

- Estudar a Base de Dados Científicos da UFPR - <https://bdc.c3s1.ufpr.br/>
- Estudar o RE3DATA - <https://www.re3data.org/>
- Verificar provenance de repositórios de dados
- Ler sobre Science gateway frameworks
- Tentar executar um workflow e reproduzir um experimento
- Atualizar o estado da arte com artigos novos
 - Toward a reference architecture based science gateway framework with embedded e-learning support - 2022 - Gabriele Pierantoni, Tamas Kiss, Alexander Bolotov, Dimitrios Kagialis, James DesLauriers, Amjad Ullah, Huankai Chen, David Chan You Fee, Hai-Van Dang, Jozsef Kovacs, Anna Belelaki, Themistocles Herekakis, Ioanna Tsagouri, Sandra Gesing

Workflow Científico - Tema 5

Tópicos em Banco de Dados (INFO-7058)

Gabriel Silva Hermida (gash18@inf.ufpr.br)
Jefferson Luis Stefani Alves Pereira (jlsap17@inf.ufpr.br)
Josiney de Souza (josiney.souza@ifc.edu.br)
Lucas Sidnei dos Santos (lss16@inf.ufpr.br)

UFPR / DInf

10 de maio de 2023

