

PHENOMANAGER : UMA ABORDAGEM PARA GERÊNCIA DE HIPÓTESES DE FENÔMENOS CIENTÍFICOS

Leonardo Sousa Lima Ramos
e-mail: leosramos@gmail.com

Banca examinadora:

- Daniel de Oliveira – Orientador
- Fabio Porto – Co-orientador
- Kary Ocaña
- Antônio Tadeu
- Aline Marins

SUMÁRIO



1. Introdução
2. Fundamentação Teórica
3. Abordagem proposta: **PhenoManager**
4. Avaliação Experimental
5. Conclusão
6. Agradecimentos

INTRODUÇÃO

MOTIVAÇÃO



- ✗ O processo de experimentação é uma das formas usadas para apoiar as teorias baseadas em um método científico
- ✗ Aumento da utilização de simulações computacionais para apoiar experimentos em diversos domínios da ciência
- ✗ Projetos podem estar associados a vários experimentos e experimentos podem estar associados a diversos workflows
- ✗ Validação de hipótese pode requerer análise da execução de Workflows
- ✗ Necessidade de conhecimento computacional
- ✗ Equipes podem estar dispersas geograficamente

INTRODUÇÃO

OBJETIVO

- ✕ Necessidade de uma ferramenta que auxilie no apoio do método científico, ajudando na documentação, compartilhamento dos dados obtidos e na facilitação da reprodução dos experimentos realizados
- ✕ Desenvolvimento do PhenoManager
- ✕ Simplificação de configuração de execução de simulações computacionais e facilitação na reprodução dos experimentos
- ✕ Interfaces amigáveis
- ✕ Integração com SciManager para gerência de projetos científicos

FUNCAMENTAÇÃO TEÓRICA

PROJETOS CIENTÍFICOS

- ✕ É a unidade de trabalho de mais alto nível e tem como objetivo principal funcionar como um roteiro de trabalho ou instrumento de planejamento, além de ser o elemento direcionador da pesquisa
- ✕ Conta com grupos de pessoas para as quais as tarefas e responsabilidades relativas ao projeto serão atribuídas
- ✕ Pode ser composto por um ou mais experimentos científicos

FUNCAMENTAÇÃO TEÓRICA

EXPERIMENTOS CIENTÍFICOS



- ✘ Definição: “um teste executado sob condições controladas, que é realizado para demonstrar uma verdade conhecida, examinar a validade de uma hipótese, ou determinar a eficácia de algo previamente não explorado”. *Oxford Dictionary of English*
- ✘ Geralmente representam um encadeamento e execução de diferentes programas e podem ser modelados em um Workflow Científico

FUNCAMENTAÇÃO TEÓRICA

WORKFLOWS CIENTÍFICOS

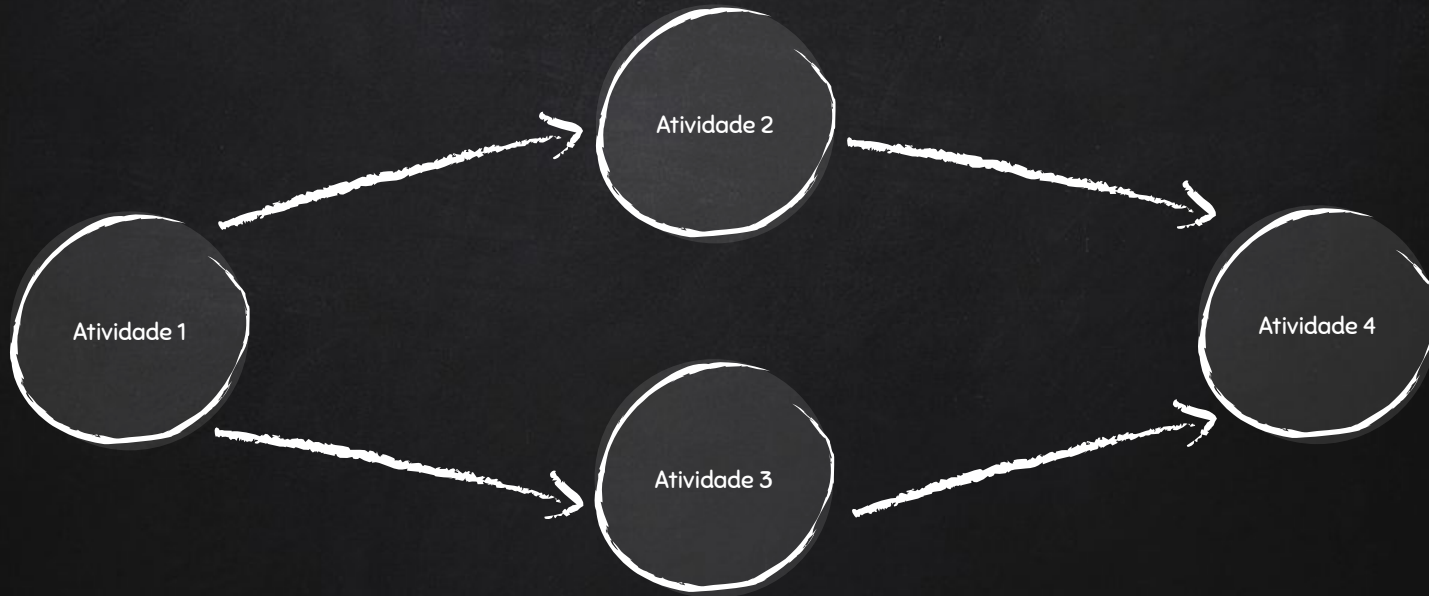
- x É uma abstração que modela o encadeamento de atividades (invocações de programas) e dados sendo executados por Sistemas de Gerência de Workflow Científicos (SGWfC)

FUNCAMENTAÇÃO TEÓRICA

WORKFLOWS CIENTÍFICOS



- ✗ Podemos representá-lo como um grafo direcionado e acíclico



FUNCAMENTAÇÃO TEÓRICA

PROVENIÊNCIA DE DADOS

- ✗ Metadados associados as diversas etapas do experimento científico
- ✗ Os dados de proveniência descrevem o histórico da execução de um experimento e são fundamentais para a reprodução e/ou a análise dos resultados
 - Mapear o fluxo de dados
 - Determinar a utilização de recursos por cada etapa do workflow, script ou aplicação;
 - Detectar erros na geração de dados e/ou no processo;
 - Estimar a qualidade e/ou confiabilidade dos dados baseando-se na origem dos dados;
 - A replicação ou derivação de dados;
 - A realização consultas baseadas nos metadados de origem para a descoberta de dados.

FUNCAMENTAÇÃO TEÓRICA

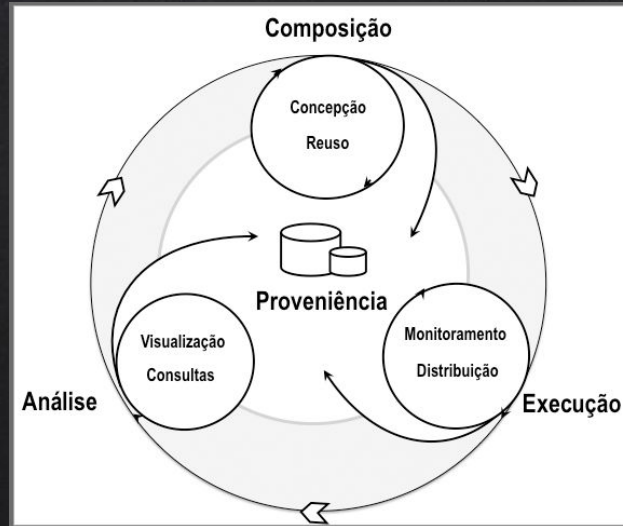
PROVENIÊNCIA DE DADOS

- ✕ Prospectiva
 - Plano de execução
- ✕ Retrospectiva
 - Execução

FUNCAMENTAÇÃO TEÓRICA

CICLO DE VIDA DO EXPERIMENTO CIENTÍFICO

- ✗ Consiste basicamente em múltiplas interações, a serem realizadas por cientistas, durante o curso de um experimento

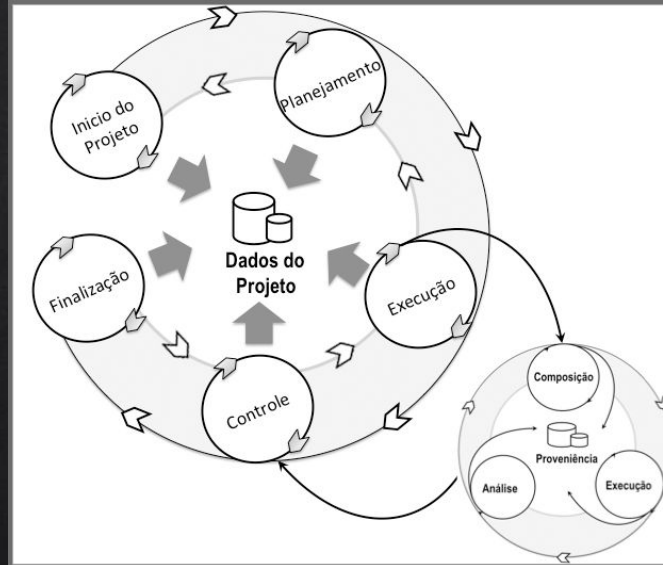


Ciclo simplificado

FUNCAMENTAÇÃO TEÓRICA

CICLO DE VIDA DO PROJETO CIENTÍFICO

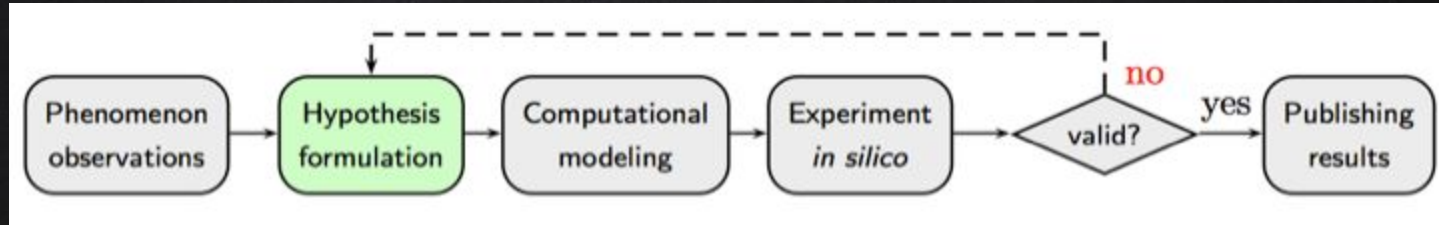
- ✗ Consiste em um conjunto de fases que são executadas pelos membros do projeto
- ✗ Baseado em técnicas/metodologias de gerência de projetos



FUNCAMENTAÇÃO TEÓRICA

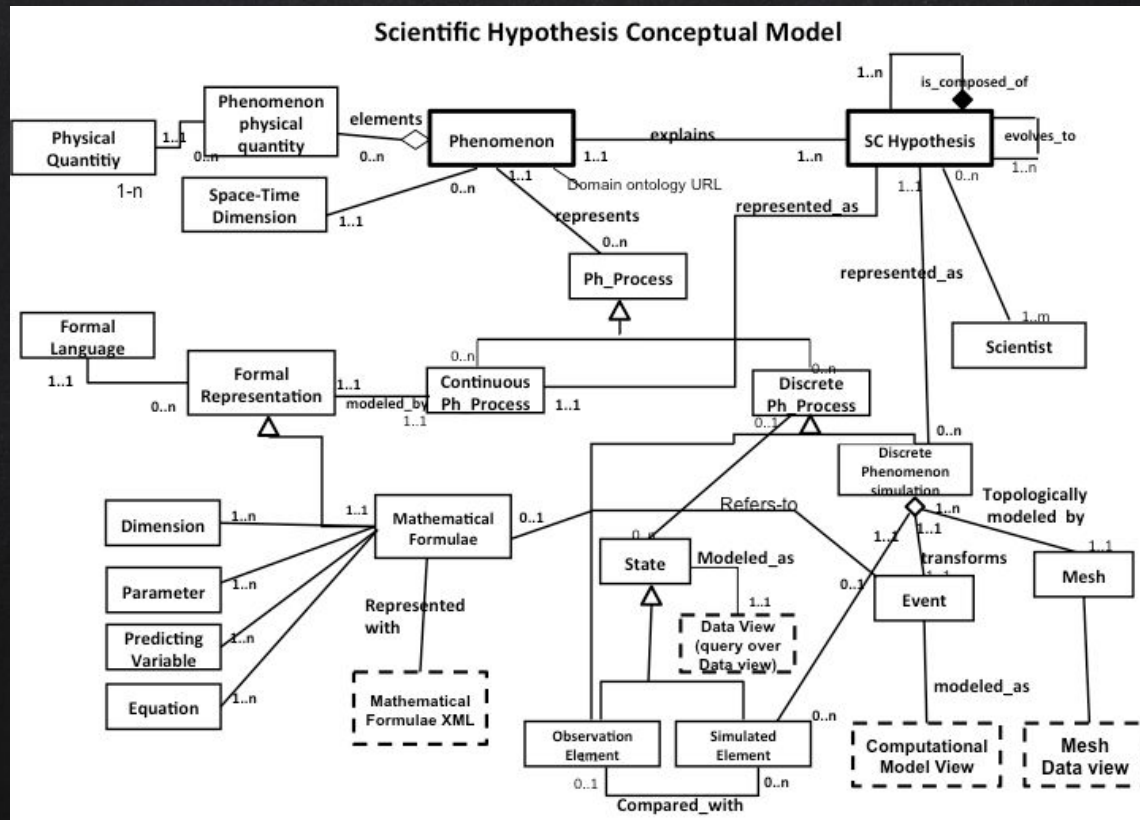
O MODELO DE DADOS DE HIPÓTESIS CIENTÍFICAS

✕ Ciclo de vida da exploração científica *In-Silico*



FUNCAMENTAÇÃO TEÓRICA

O MODELO DE DADOS DE HIPÓTESIS CIENTÍFICAS



FUNCAMENTAÇÃO TEÓRICA

AMBIENTES DE PROCESSAMENTO DE ALTO DESEMPENHO (PADs)

- ✗ Clusters (Homogêneo)
- ✗ Grade (Heterogêneo)
- ✗ Ambiente de computação voluntária
- ✗ Nuvem de computadores (máquinas virtuais)

FUNCAMENTAÇÃO TEÓRICA

RESEARCH OBJECT

- ✕ Facilitar reuso
- ✕ Facilitar reprodução
- ✕ Documentar dados de proveniência
- ✕ Formato publicável



O PHENOMANAGER ARQUITETURA



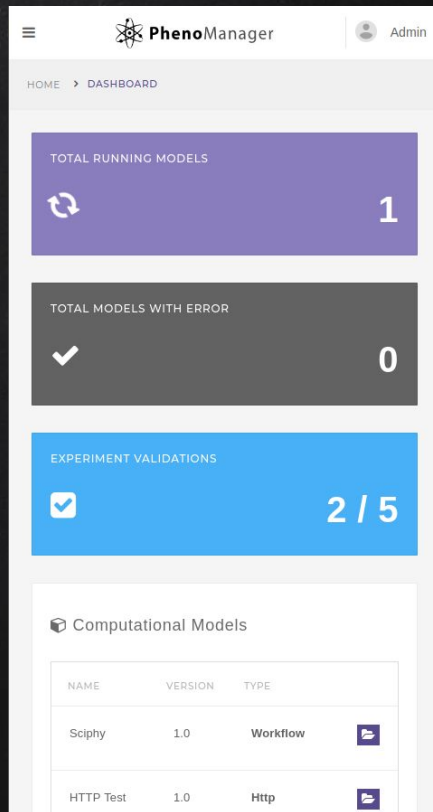
- ✗ Microserviços em Java (1.8)
- ✗ Utilização do Arcabouço Spring Boot para os Artefatos
 - Convection Over Configuration
- ✗ Artefato de estáticos que consume dados de APIs



O PHENOMANAGER ARQUITETURA



- ✗ Frontend desenvolvido de maneira responsiva
 - Se adapta a qualquer tamanho de tela (viewport)
 - Tamanho dos componentes baseado em porcentagens
 - Arcabouço Twitter Bootstrap

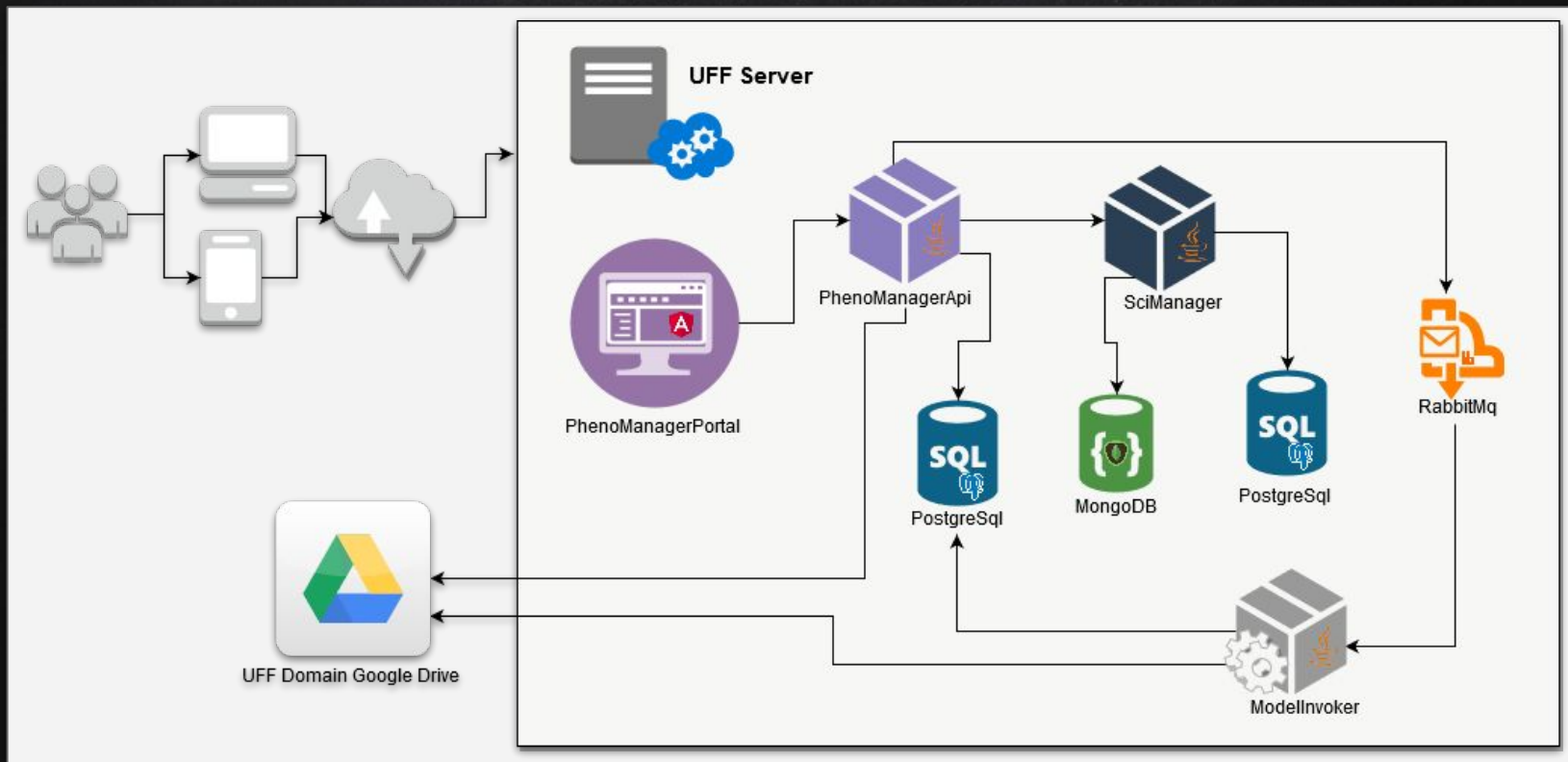




O PHENOMANAGER



ARQUITETURA - BIG PICTURE





O PHENOMANAGER FUNCIONALIDADES





PhenoManager

Login

Enter your data

EMAIL

scimanager.admin@uff.br

PASSWORD

Enter

TOTAL RUNNING MODELS



1

TOTAL MODELS WITH ERROR



0

EXPERIMENT VALIDATIONS



2 / 5

Computational Models

NAME	VERSION	TYPE	
Sciphy	1.0	Workflow	
HTTP Test	1.0	Http	
Model Test	1.0	Workflow	

Hypotheses

NAME	PARENT HYPOTHESIS	PHENOMENON	STATE	
Sciphy improvement	--	Biology Research	Formulated	
Branch hypothesis	Hypothesis Test	Phenomenon Test	Formulated	

PROJECTS

NAME

Biology Project







Astronomy Project



Computer Cience Project



	NAME	EMAIL	INSTITUTION	ROLE	CREATION DATE	
	Kary Ocanã	kary@gmail.com	LNCC	Regular user	20/05/2019	≡ ✎ ✖
	Daniel Oliveira	danielcmo@uff.ic.br	UFF	Admin	20/05/2019	≡ ✎ ✖
	Leonardo Ramos	leoslramos@gmail.com	UFF	Admin	20/05/2019	≡ ✎ ✖
	Admin	scimanager.admin@uff.br	UFF	Admin	25/10/2017	≡ ✎ ✖

[«](#) [<](#) [1](#) [>](#) [»](#)

Update team

NAME

Admin Team

*

TEAM USERS *

✕ Admin

✕ Daniel Oliveira

Cancel

Save

USERS



Create permission

PERMISSION

*

Write



USER



TEAM

USERS *

Select

Cancel

Save

+ New

«

<

1

>

»

PROJECTS

[+ New](#)

NAME	CREATION DATE	
Biology Project	14/05/2019	= 📄 ✎ 🗑
Astronomy Project	14/05/2019	= 📄 ✎ 🗑
Computer Cience Project	23/04/2019	= 📄 ✎ 🗑
		« < 1 > »

PROJECTS > BIOLOGY PROJECT

* PHENOMENONS

Permissions

+ New

NAME

RESEARCH DOMAIN

CREATION DATE

Biology Research

Biology

14/05/2019



« < 1 > »

DETAILS



Name

Biology Project

Creation Date

14/05/2019

Description

Project responsible for biology related aspects

PROJECTS > BIOLOGY PROJECT > BIOLOGY RESEARCH

HYPOTHESES PERMISSIONS

+ New

NAME	PARENT HYPOTHESIS	STATE	CREATION DATE
Sciphy improvement	--	Formulated	14/05/2019



« ‹ 1 › »

DETAILS



Name

Biology Research

Creation Date

14/05/2019

State

Biology

Description

Sciphy

EXPERIMENTS

CHILDREN HYPOTHESES

PERMISSIONS

+ New

NAME

CREATION DATE

Sciphy GPU

14/05/2019



« < 1 > »

DETAILS



Name

Sciphy improvement

Creation Date

14/05/2019

State

Formulated

Description

Test Sciphy with different SciCummulus iterations

Create child hypothesis

PARENT HYPOTHESIS STATE

Improved

NAME

Sciphy GPU Core



DESCRIPTION

Improved performance|



Cancel

Save

PROJECTS > BIOLOGY PROJECT > BIOLOGY RESEARCH > SCIPHY IMPROVEMENT > SCIPHY GPU

COMPUTATIONAL MODELS PHASES CONCEPTUAL PARAMS VALIDATION ITEMS PERMISSIONS

+ New

NAME	VERSION	TYPE	CREATION DATE
Sciphy	1.0	Workflow	14/05/2019



« < 1 > »

DETAILS



Name

Sciphy GPU

Creation Date

14/05/2019

Description

Sciphy with GPU SciCummulus

PROJECTS > BIOLOGY PROJECT > BIOLOGY RESEARCH > SCIPHY IMPROVEMENT > SCIPHY GPU

COMPUTATIONAL MODELS PHASES CONCEPTUAL PARAMS VALIDATION ITEMS PERMISSIONS

+ New

VALIDATED

EXPECTED VALUE DESCRIPTION

EVIDENCE



keep execution cost bellow X



execution time bellow N seconds



« < 1 > »

DETAILS



Name

Sciphy GPU

Creation Date

14/05/2019

Description

Sciphy with GPU SciCummulus

Validate Item

Do you wish to validate this item?

UPLOAD VALIDATION EVIDENCE 

Cancel

Validate

+ New



Update computational model

NAME

Sciphy

*

CURRENT VERSION

1.0

*

DESCRIPTION

Sciphy GPU

//



PUBLIC



NOT PUBLIC

Cancel

Save

RESEARCH > SCIPHY IMPROVEMENT > SCIPHY



Last execution Status: i

METADATA EXTRACTORS

INSTANC

FINISH DATE

USER
AGENT

25/10/2017 -
18:03:02

Admin

Sciphy
SciCummulus

--

LNCC Santos Dumont
Cluster

--

Admin



Type: Cluster

DETAILS

Name

Sciphy

Creation Da

14/05/2019

Type

Workflow

Is public

Yes

Description

Sciphy GPU



Update execution environment

TYPE

Cluster



TAG

LNCC Santos Dumont



HOST ADDRESS

xxx.xxx.x.xx

USERNAME

user.name

PASSWORD

CLUSTER PROJECT NAME

LNCC Santos Dumont

☐ VPN ☒ CISCO VPN ☐ NONE

VPN CONFIGURATION

IPSec gateway xxx.xxx.x.xx

IPSec ID vpnID

IPSec secret *****

Xauth username user

Xauth password *****



Cancel

Save

PASSWORD

CLUSTER NAME

SECRET KEY

ACCESS KEY

IMAGE

☐ VPN ☐ CISCO VPN ☒ NONE

VIRTUAL MACHINES

Type

Financial cost

Disk space

RAM

Gflops

Platform

Number of cores



Cancel

Save

Update executor

TAG

Sciphy SciCummulus



UPLOAD EXECUTOR (ZIP) 

EXECUTION COMMAND

java -jar SCSetup.jar



ABORT COMMAND

killall -9 java



JOB NAME (FOR CLUSTERS)

sciphy-phenomanager

Cancel

Save

Update executor

TAG

HTTP Rest



HTTP PROTOCOL TYPE

REST



URL BASE

http://localhost:9501/v1/projects



HTTP VERB

POST



HTTP BODY

```
{  
  "name": "test"  
}
```



HTTP HEADERS

"Authorization=Bearer
ADSHADJKHjkhjdsk98789KDSAHDJkhjk2132132"



Cancel

Save

PROJECTS > BIOLOGY PROJECT > BIOLOGY RESEARCH > SCIPHY IMPROVEMENT > SCIPHY GPU > SCIPHY

ACTIVE EXECUTOR



Tag: Sciphy SciCummulus

Last execution Status: idle

ACTIVE ENVIRONMENT



Tag: LNCC Santos Dumont

Type: Cluster

EXECUTIONS RESULTS

EXECUTOR

METADATA EXTRACTORS

INSTANCE PARAMS

ENVIRONMENT

PERMISSIONS

[+ New](#)

KEY	VALUE	DESCRIPTION	CREATION DATE	
fasta		Biology Dataset	18/05/2019	
mafft		Biology Dataset	18/05/2019	

[<<](#) [<](#) [1](#) [>](#) [>>](#)

DETAILS



Name

Sciphy

Creation Date

14/05/2019

Type

Workflow

Is public

No

Description

Sciphy GPU

Update metadata extractor

TAG

scicummulus extractor

*

UPLOAD EXTRACTOR (EXECUTABLE) 

EXECUTION COMMAND

.extract_from_scicummulus_database.sh

*

//

Cancel

Save

PROJECTS > BIOLOGY PROJECT > BIOLOGY RESEARCH > SCIPHY IMPROVEMENT > SCIPHY GPU > SCIPHY

ACTIVE EXECUTOR

Tag: Sciphy SciCummulus

Last execution Status: idle

ACTIVE ENVIRONMENT

Tag: LNCC Santos Dumont

Type: Cluster

EXECUTIONS RESULTS EXECUTOR METADATA EXTRACTORS INSTANCE PARAMS ENVIRONMENT PERMISSIONS

STATUS	START DATE	FINISH DATE	USER AGENT	EXECUTOR	EXTRACTORS	ENVIRONMENT	
Finished	25/10/2017 - 11:03:02	25/10/2017 - 18:03:02	Admin	Sciphy SciCummulus	scicummulus extractor	LNCC Santos Dumont Cluster	
Running	21/10/2019 - 11:03:02	--	Admin	Sciphy SciCummulus	--	LNCC Santos Dumont Cluster	
Aborted	25/10/2017 - 11:03:02	25/10/2017 - 11:03:02	Admin	Sciphy SciCummulus	--	LNCC Santos Dumont Cluster	

DETAILS

Name

Sciphy

Creation Date

14/05/2019

Type

Workflow

Is public

No

Description

Sciphy GPU

Execution output

Running 

Start Date: 21/10/2019 - 11:03:02

Finish Date: *Still Running*

```
10/05/2019 - 23:25:50: Starting execution of modelExecutor [Sciphy]
10/05/2019 - 23:26:11: Setting up Cluster environment [LNCC Santos
Dumont]...
10/05/2019 - 23:27:42: Finish setting up Cluster environment [LNCC
Santos Dumont]
. . .
```

Close

 DETAILS

Name

Sciphy

Creation Date

14/05/2019

Type

Workflow

Is public

No

Description

Sciphy GPU

Type: Cluster



Sciphy
SciCummulus

LNCC Santos Dumont
Cluster



```

{
  "@context":{
    "schema":"http://schema.org/",
    .
    .
    .
  },
  "@graph":[{
    "@type":[
      "ro:ResearchObject",
      "ore:Aggregation"
    ],
    "@id":"4B471432FCD146018593817458D6E21D"
  },{
    "schema:name":"Sciphy"
  },{
    "dc:creator":"QWE123987POEIQPEWQ12687EWQEWQEF"
  },{
    "dc:abstract":"Sciphy GPU"
  },{
    "dc:contributor":["QWE123987POEIQPEWQ12687EWQEWQEF"]
  },{
    "dc:title":"Sciphy"
  },{
    "Ore:aggregates":[{
      "@type":"ro:Resource",
      "@id":"http://localhost:9500/PhenoManagerApi/v1/computational_models/4B471432FCD146018593817458D6E21D/instance_params/3EE92B89774345BD9A8CA4DF77FB148A/value_file"
    },{
      "@type":"ro:Resource",
      "@id":"http://localhost:9500/PhenoManagerApi/v1/computational_models/4B471432FCD146018593817458D6E21D/instance_params/E4A3F7035B8D45D29ABA0754E89FE8F5/value_file"
    },{
      "@type":"ro:Resource",
      "@id":"http://localhost:9500/PhenoManagerApi/v1/computational_models/4B471432FCD146018593817458D6E21D/model_executors/4C6212B68E2C47929F509B88A037583B/executor"
    },{
      "@type":"ro:Resource",
      "@id":"http://localhost:9500/PhenoManagerApi/v1/computational_models/4B471432FCD146018593817458D6E21D/model_result_metadata/ND5JDKHAJKD789HDSJGAJD"
    }
  ]
  },{
    "prov:wasAttributedTo":["QWE123987POEIQPEWQ12687EWQEWQEF"]
  },{
    "schema:contributor":["QWE123987POEIQPEWQ12687EWQEWQEF"]
  },{
    "foaf:name":"Admin",
    "@@type":"foaf:Person",
    "@id":"QWE123987POEIQPEWQ12687EWQEWQEF",
    "schema:name":"Admin"
  }
]
}

```

Sync project with SciManager

If you confirm, all the data related to this project will be replicated in SciManager application.

Do you wish to sync project?

Cancel

Confirm

14/05/2019

Biology Project

Projetos científicos > Biology Project

Dashboard

Documentação

Experimentos

Fases

Quadro de tarefas

Lista de tarefas

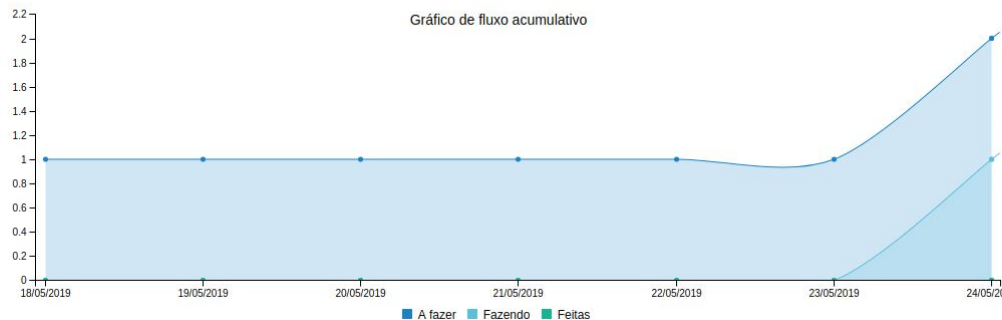
Métricas

18/05/2019 - 24/05/2019 ▾

Filtro avançado ▾

Acumulativo

Diário



Experimentos Científicos

Criar experimento +

Nome do experimento científico

Grupos responsáveis

Sciphy with GPU SciCummulus

Teste2, Teste



Detalhes

Biology Project ✎

Grupos responsáveis:

Teste2, Teste

Experimentos Científicos: 1

Workflows Científicos: 2

Tarefas a fazer: 1

Tarefas fazendo: 1

Tarefas concluídas: 0

Tarefas atrasadas: 0

Tarefas

Criar tarefa +

Executar workflow

Estimativa: 45



execução

Planejar concepção dos testes



Teste

teste

Biology Project

Projetos científicos > Biology Project

Dashboard Documentação Experimentos Fases **Quadro de tarefas** Lista de tarefas

Criar tarefa +

📅 Data inicial

📅 Data final

Buscar por...

▼ Buscar

Todas: ☒ Minhas: ☐

A fazer

(1 de 1 tarefas)

Planejar concepção dos testes



Teste

teste

Em progresso

(1 de 1 tarefas)

Executar workflow



Estimativa: 45



execução

Concluído

(0 de 0 tarefas)

AVALIAÇÃO EXPERIMENTAL



- ✗ Estudo de caso - Workflow Sciphy
 - Gerar árvores filogenéticas com máxima verossimilhança
- ✗ Execução via SciCumulus
- ✗ Ambiente Cluster Santos Dumont do LNCC
- ✗ Dois casos:
 - Fila vazia
 - Fila cheia
 - 1000 mensagens
 - 10 threads para o consumidor

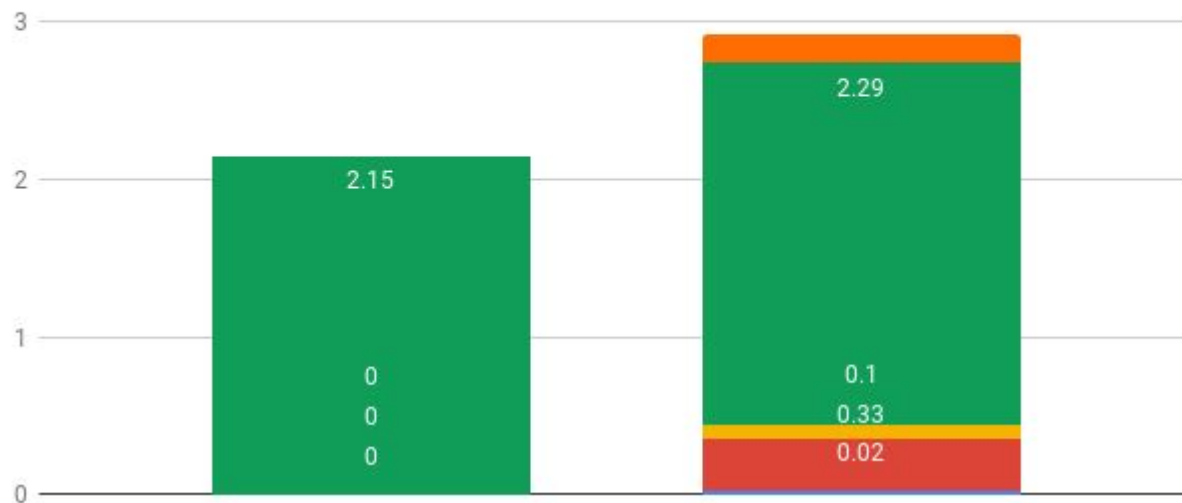
AVALIAÇÃO EXPERIMENTAL

FILA VAZIA



Tempos de execução

Upload de Metadado Execução Upload do executor Download do executor
RabbitMq



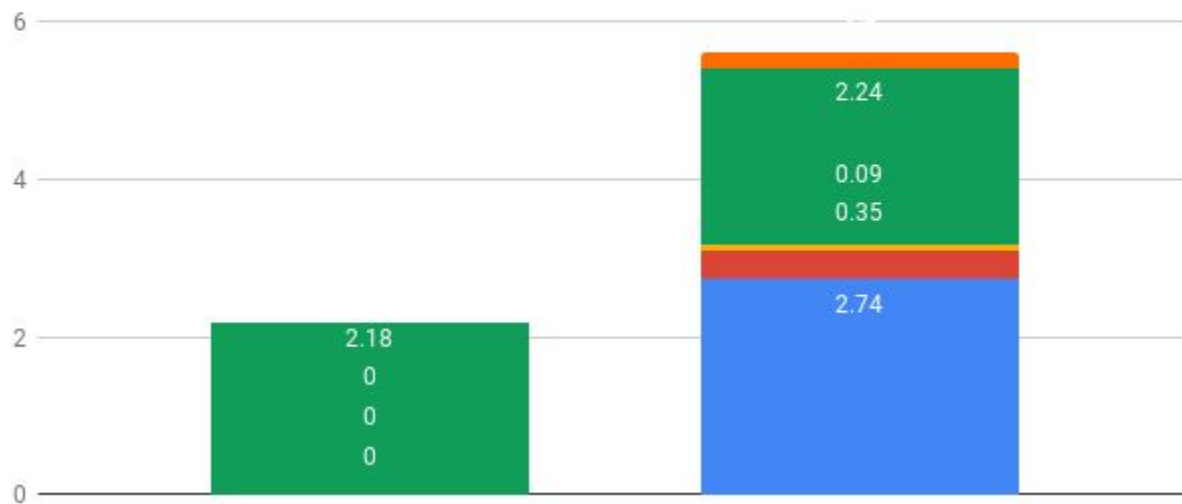
AVALIAÇÃO EXPERIMENTAL

FILA VAZIA



Tempos de execução

Upload de Metadado Execução Upload do executor Download do executor
RabbitMq



AVALIAÇÃO EXPERIMENTAL



- ✗ De forma a avaliar a abordagem proposta foi utilizado o modelo de avaliação denominado TAM (Technology Acceptance Model)
- ✗ Avaliar a receptividade/comportamento de um usuário no que se refere a utilidade e a facilidade da tecnologia/ferramenta que está sendo proposta
- ✗ Utilidade refere-se ao quanto o usuário acredita que a abordagem proposta o auxiliará em suas tarefas e a facilidade se refere ao quão fácil/simple será utilizar tal abordagem

AVALIAÇÃO EXPERIMENTAL



- ✗ Criação de um questionário indagando:
 - qualidade do treinamento (T)
 - a facilidade de uso do PhenoManager (F)
 - Utilidade do PhenoManager (U)
- ✗ Candidatos que participaram da avaliação: alunos da graduação e pós graduação com um mínimo conhecimento prévio sobre o domínio abordado

AVALIAÇÃO EXPERIMENTAL

PERGUNTAS – TREINAMENTO



ID	Questão
T1	O treinamento oferecido pelos avaliadores foi completo?
T2	O treinamento me ofereceu subsidios para utilizar o PhenoManager?
T3	O treinamento foi adequado em relação ao tempo e ao detalhamento?

AVALIAÇÃO EXPERIMENTAL

PERGUNTAS – FACILIDADE



ID	Questão
F1	Você gostou de utilizar o PhenoManager?
F2	O acesso ao PhenoManager é simples?
F3	Utilizar o PhenoManager é uma boa idéia?
F4	As funcionalidades no PhenoManager são simples de serem compreendidas?
F5	É fácil encontrar as informações que desejo no PhenoManager?
F6	O PhenoManager possui uma interface atraente e amigável?
F7	Mesmo antes de clicar em uma funcionalidade, você já consegue prever o que vai acontecer?

AVALIAÇÃO EXPERIMENTAL

PERGUNTAS – UTILIDADE



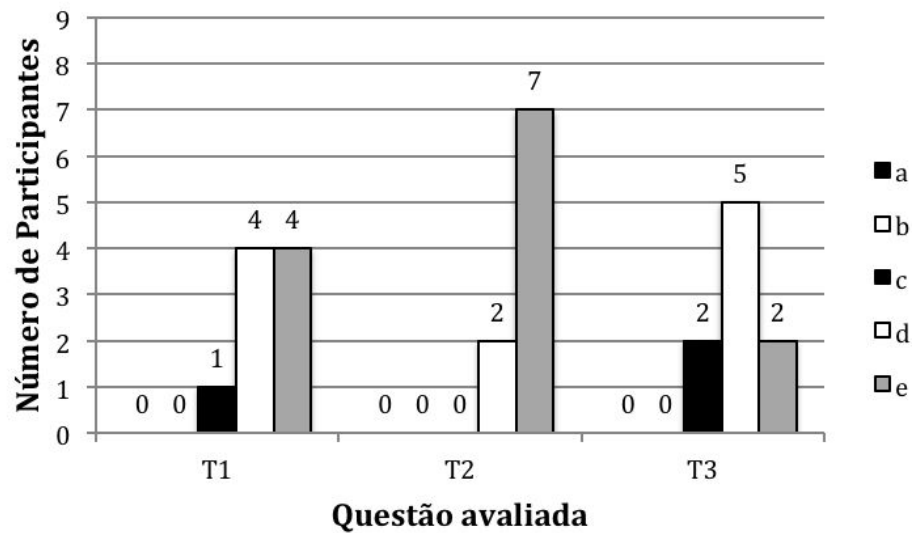
ID	Questão
U1	O uso do PhenoManager agrega valor aos experimentos que eu executo?
U2	Utilizar o PhenoManager é útil para ensinar iniciantes a trabalhar em projeto?
U3	Utilizar o PhenoManager pode melhorar o desempenho do meu projeto científico?
U4	O uso do PhenoManager pode facilitar meu trabalho?
U5	Seria factível integrar o uso do PhenoManager na minha rotina de trabalho?
U6	Eu recomendaria o PhenoManager no desenvolvimento de outros projetos?
U7	Os conceitos de experimentação e projeto científico foram abordados por completo no PhenoManager?

AVALIAÇÃO EXPERIMENTAL

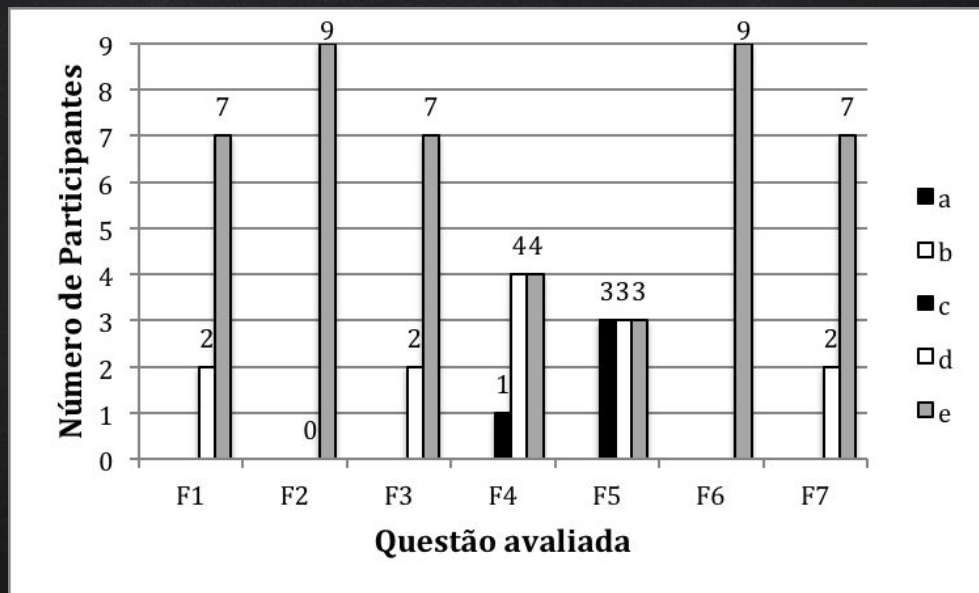


- ✕ Respostas possíveis:
 - a) Discordo totalmente
 - b) Discordo Parcialmente
 - c) Neutro
 - d) Concordo Parcialmente
 - e) Concordo totalmente

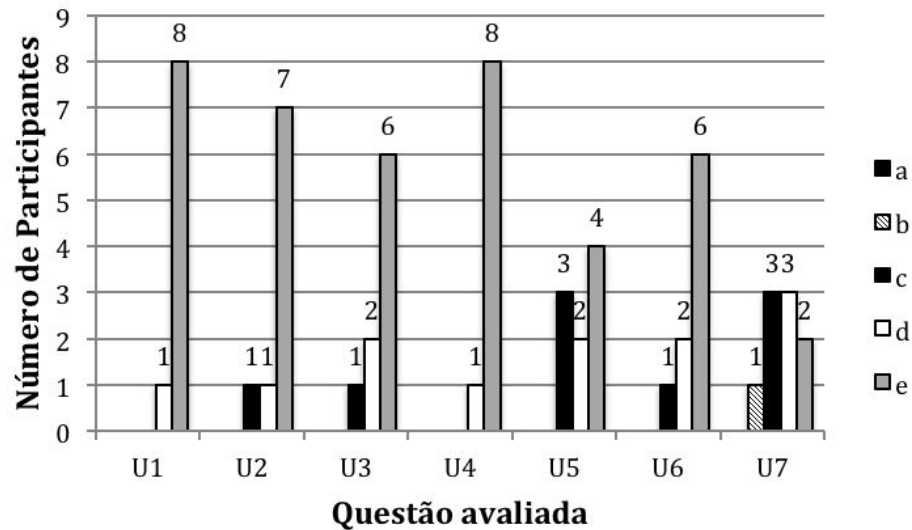
AVALIAÇÃO EXPERIMENTAL



AVALIAÇÃO EXPERIMENTAL



AVALIAÇÃO EXPERIMENTAL



AVALIAÇÃO EXPERIMENTAL



- ✗ A avaliação experimental mostrou que mais de 88% dos avaliadores acredita que o uso do PhenoManager agrega valor ao seu trabalho e, possivelmente melhorará o desempenho do projeto científico

CONCLUSÃO



- ✗ Praticidade e facilidade x Overhead
- ✗ Os resultados foram positivos e 88% dos avaliadores concordaram que o PhenoManager possivelmente traria melhorias no desempenho do Projeto
- ✗ Trabalhos futuros incluem modelagem de Workflows Científicos na ferramenta, monitoramento de saúde dos ambientes computacionais

OBRIGADO

