

23 - Criando um pipeline no portal II

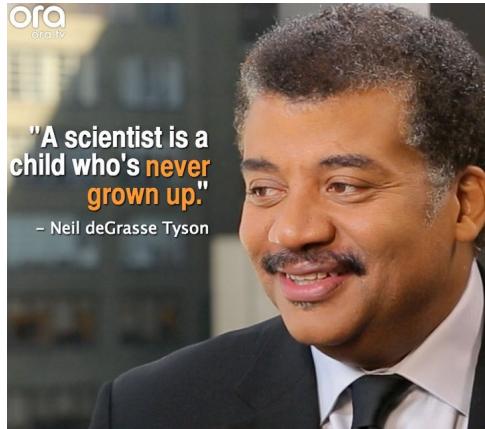
Tutorial

Julia Gschwend

Momento mensagem motivacional

Motivos para desenvolver seu science workflow no DES Science Portal (1.0):

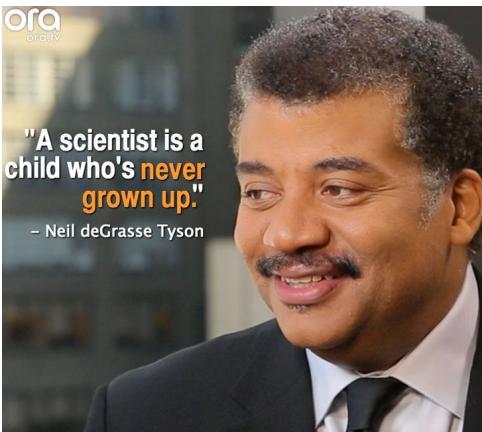
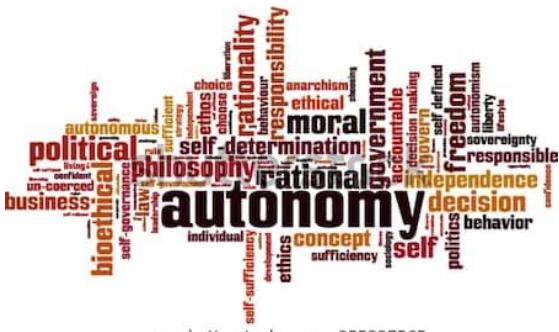
- Mais 2 anos de DES (no mínimo)
- Componentes poderão ser aproveitados no portal LSST
- Apoio da equipe LIneA (IT team + cientistas)
- Agilidade no desenvolvimento (E2E VS. Science)



Momento mensagem motivacional

Motivos para desenvolver seu science workflow no DES Science Portal (1.0):

- Mais 2 anos de DES (no mínimo)
- Componentes poderão ser aproveitados no portal LSST
- Apoio da equipe LIneA (IT team + cientistas)
- Agilidade no desenvolvimento (E2E VS. Science)



1. Desenhar workflow - criar doc pipeline description
2. Criar versão inicial do código no Jupyter (opcional)
3. Criar repositório para o novo pipeline contendo os arquivos:
 - pipeline.xml
 - VERSION (V00_00_00)
 - ChangeLog
4. Criar repositório para o componente 1 contendo os arquivos:
 - __init__.py
 - config/component_config.xml
 - input/component_input.xml
 - VERSION (V00_00_00)
 - ChangeLog
5. Criar repositório para o componente 2 contendo os arquivos:
 - __init__.py
 - output/component_output.xml
 - etc/ingest_table.rd
 - VERSION (V00_00_00)
 - ChangeLog
6. Atualizar repositório
7. Rodar no portal do desenvolvedor
8. Debugar/fine tuning
9. Versionar pipeline e componentes
10. Atualizar testing
11. Testar em testing
12. Fast deploy

1. Desenhar workflow - criar doc pipeline description
2. Criar versão inicial do código no Jupyter (opcional)
3. Criar repositório para o novo pipeline contendo os arquivos:
 - pipeline.xml
 - VERSION (V00_00_00)
 - ChangeLog
4. Criar repositório para o componente 1 contendo os arquivos:
 - __init__.py
 - config/component_config.xml
 - input/component_input.xml
 - VERSION (V00_00_00)
 - ChangeLog
5. Criar repositório para o componente 2 contendo os arquivos:
 - __init__.py
 - output/component_output.xml
 - etc/ingest_table.rd
 - VERSION (V00_00_00)
 - ChangeLog
6. Atualizar repositório
7. Rodar no portal do desenvolvedor
8. Debugar/fine tuning
9. Versionar pipeline e componentes
10. Atualizar testing
11. Testar em testing
12. Fast deploy

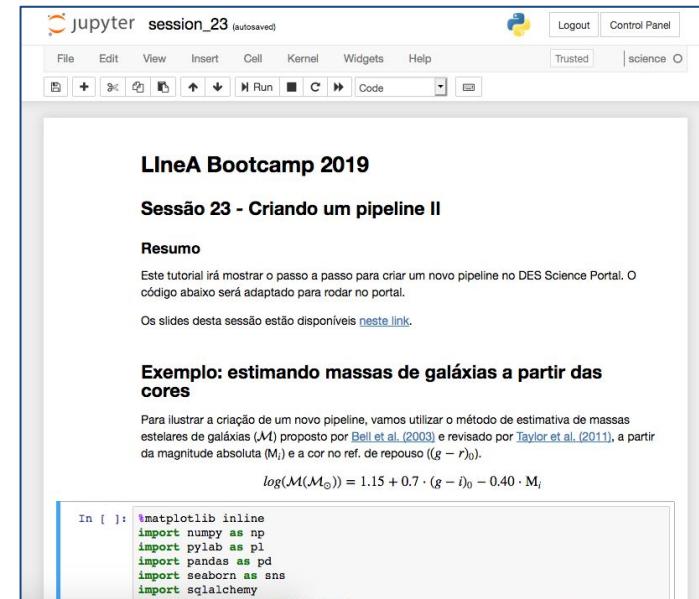
1. Desenhar workflow - criar doc pipeline description
2. Criar versão inicial do código no Jupyter (opcional)
3. Criar repositório para o novo pipeline contendo os arquivos:
 - pipeline.xml
 - VERSION (V00_00_00)
 - ChangeLog
4. Criar repositório para o componente 1 contendo os arquivos:
 - __init__.py
 - config/component_config.xml
 - input/component_input.xml
 - VERSION (V00_00_00)
 - ChangeLog
5. Criar repositório para o componente 2 contendo os arquivos:
 - __init__.py
 - output/component_output.xml
 - etc/ingest_table.rd
 - VERSION (V00_00_00)
 - ChangeLog
6. Atualizar repositório
7. Rodar no portal do desenvolvedor
8. Debugar/fine tuning
9. Versionar pipeline e componentes
10. Atualizar testing
11. Testar em testing
12. Fast deploy

2 - Código no Jupyter

Exemplo: estimando massas de galáxias a partir das cores (Bell et al. 2003 / Taylor et al. 2011).

$$\log(M(M_\odot)) = 1.15 + 0.7 \cdot (g - i)_0 - 0.40 \cdot M_i$$

jupyter session_23



LineA Bootcamp 2019

Sessão 23 - Criando um pipeline II

Resumo

Este tutorial irá mostrar o passo a passo para criar um novo pipeline no DES Science Portal. O código abaixo será adaptado para rodar no portal.

Os slides desta sessão estão disponíveis [neste link](#).

Exemplo: estimando massas de galáxias a partir das cores

Para ilustrar a criação de um novo pipeline, vamos utilizar o método de estimativa de massas estelares de galáxias (M) proposto por [Bell et al. \(2003\)](#) e revisado por [Taylor et al. \(2011\)](#), a partir da magnitude absoluta (M_i) e a cor no ref. de repouso ($(g - r)_0$).

$$\log(M(M_\odot)) = 1.15 + 0.7 \cdot (g - i)_0 - 0.40 \cdot M_i$$

```
In [ ]: %matplotlib inline
import numpy as np
import pylab as pl
import pandas as pd
import seaborn as sns
import sqlalchemy
```

Passo-a-passo

1. Desenhar workflow - criar doc pipeline description
2. Criar versão inicial do código no Jupyter (opcional)
3. **Criar repositório para o novo pipeline contendo os arquivos:**
 - pipeline.xml
 - VERSION (V00_00_00)
 - ChangeLog
4. Criar repositório para o componente 1 contendo os arquivos:
 - __init__.py
 - config/component_config.xml
 - input/component_input.xml
 - VERSION (V00_00_00)
 - ChangeLog
5. Criar repositório para o componente 2 contendo os arquivos:
 - __init__.py
 - output/component_output.xml
 - etc/ingest_table.rd
 - VERSION (V00_00_00)
 - ChangeLog
6. Atualizar repositório
7. Rodar no portal do desenvolvedor
8. Debugar/fine tuning
9. Versionar pipeline e componentes
10. Atualizar testing
11. Testar em testing
12. Fast deploy

3. Criar repositório para o novo pipeline

```
$ mkdir new_pipeline.tmp  
$ cd new_pipeline.tmp  
$ git init  
  
$ vim pipeline.xml  
$ vim VERSION  
$ vim ChangeLog  
  
$ git add .  
$ git commit -m "Initial commit for a new submodule"  
$ git tag `cat VERSION`
```

1. Desenhar workflow - criar doc pipeline description
2. Criar versão inicial do código no Jupyter (opcional)
3. Criar repositório para o novo pipeline contendo os arquivos:
 - pipeline.xml
 - VERSION (V00_00_00)
 - ChangeLog
4. Criar repositório para o componente 1 contendo os arquivos:
 - __init__.py
 - config/component_config.xml
 - input/component_input.xml
 - VERSION (V00_00_00)
 - ChangeLog
5. Criar repositório para o componente 2 contendo os arquivos:
 - __init__.py
 - output/component_output.xml
 - etc/ingest_table.rd
 - VERSION (V00_00_00)
 - ChangeLog
6. Atualizar repositório
7. Rodar no portal do desenvolvedor
8. Debugar/fine tuning
9. Versionar pipeline e componentes
10. Atualizar testing
11. Testar em testing
12. Fast deploy

4,5. Criar repositório para o novo component

```
$ mkdir new_component.tmp  
$ cd new_component.tmp  
$ git init  
  
$ vim __init__.py  
$ vim VERSION  
$ vim ChangeLog  
  
$ git add .  
$ git commit -m "Initial commit for a new submodule"  
$ git tag `cat VERSION`
```

helpdesk@linea.gov.br

Email subject: product creation: PRODUCTNAME VXX_YY_ZZ
(e.g.: pipeline-cluster_halo_matching V00_00_00, component-subsets_separation V00_00_00)

Email content:

Please could you deploy this new product?

Product Type: (pipeline/component)

Product Location: full path to newmodule.tmp

Does it involve a new data class? (y/n)

Does it involve a new version of package in Tawala/EUPS? (y/n)

Documentação: [NT 010-2018 - Release Procedure for Developers](#)

1. Desenhar workflow - criar doc pipeline description
2. Criar versão inicial do código no Jupyter (opcional)
3. Criar repositório para o novo pipeline contendo os arquivos:
 - pipeline.xml
 - VERSION (V00_00_00)
 - ChangeLog
4. Criar repositório para o componente 1 contendo os arquivos:
 - __init__.py
 - config/component_config.xml
 - input/component_input.xml
 - VERSION (V00_00_00)
 - ChangeLog
5. Criar repositório para o componente 2 contendo os arquivos:
 - __init__.py
 - output/component_output.xml
 - etc/ingest_table.rd
 - VERSION (V00_00_00)
 - ChangeLog
6. Atualizar repositório
7. Rodar no portal do desenvolvedor
8. Debugar/fine tuning
9. Versionar pipeline e componentes
10. Atualizar testing
11. Testar em testing
12. Fast deploy

6. Atualizar repositório (depois da resposta do ticket)

```
$ cd $PORTAL_ROOT/des  
$ git pull  
$ git submodule init  
$ git submodule update
```

Passo-a-passo

1. Desenhar workflow - criar doc pipeline description
2. Criar versão inicial do código no Jupyter (opcional)
3. Criar repositório para o novo pipeline contendo os arquivos:
 - pipeline.xml
 - VERSION (V00_00_00)
 - ChangeLog
4. Criar repositório para o componente 1 contendo os arquivos:
 - __init__.py
 - config/component_config.xml
 - input/component_input.xml
 - VERSION (V00_00_00)
 - ChangeLog
5. Criar repositório para o componente 2 contendo os arquivos:
 - __init__.py
 - output/component_output.xml
 - etc/ingest_table.rd
 - VERSION (V00_00_00)
 - ChangeLog
6. Atualizar repositório
7. Rodar no portal do desenvolvedor
8. Debugar/fine tuning
9. Versionar pipeline e componentes
10. Atualizar testing
11. Testar em testing
12. Fast deploy

7. Rodar no portal do desenvolvedor

BOOT
CAMP



INCT
e-Universo

>>

[Dashboard](#) [My Workspace](#) [Pipelines](#) [Tools](#) [Science Products](#) [Data Server](#) [Documentation](#) [Help](#)

DES Science Portal: Workflows

[Tweets by DES Science Portal](#)

The Science Portal has two instances:

- **Workflows:** hosts workflows for Quality Assessment (QA), for the creation of Value-Added Catalogs (VACs) and for Science Analysis.
- **Data Server:** provide access to the Catalog Server and published results

The system is designed to be self-evident, use the help icon "(?)" available on each page.

The Science Portal is a facility developed by [Linea](#). If you have any question please contact us through the helpdesk@linea.gov.br

Developer Portal Instance

Powered by [Linea](#)

1. Desenhar workflow - criar doc pipeline description
2. Criar versão inicial do código no Jupyter (opcional)
3. Criar repositório para o novo pipeline contendo os arquivos:
 - pipeline.xml
 - VERSION (V00_00_00)
 - ChangeLog
4. Criar repositório para o componente 1 contendo os arquivos:
 - __init__.py
 - config/component_config.xml
 - input/component_input.xml
 - VERSION (V00_00_00)
 - ChangeLog
5. Criar repositório para o componente 2 contendo os arquivos:
 - __init__.py
 - output/component_output.xml
 - etc/ingest_table.rd
 - VERSION (V00_00_00)
 - ChangeLog
6. Atualizar repositório
7. Rodar no portal do desenvolvedor
8. **Debugar/fine tuning**
9. Versionar pipeline e componentes
10. Atualizar testing
11. Testar em testing
12. Fast deploy

8. Debug /fine tuning



```
$ cd /mnt/scratch/users/<username>/master_des/0000<process number>
$ cd <task name>_1_sandbox

$ # change code
```

8. Debug /fine tuning

BOOT
CAMP



INCT do
e-Universo

```
$ cd /mnt/scratch/users/<username>/mast  
$ cd <task name>_1_sandbox  
  
$ # change code
```

```
$ cd $PORTAL_ROOT/<path to submodule>  
$ git checkout master  
$ git pull  
  
$ vim __init__.py # make changes  
$ git status  
$ git add __init__.py  
$ git commit -m "<commit message>"
```

Repetir quantas
vezes for
necessário

8. Debug /fine tuning



```
$ cd /mnt/scratch/users/<username>/master_des/0000<process number>
$ cd <task name>_1_sandbox

$ # change code

$ # run in sandbox (debug mode)
--- local: ---
$ python -c "import components.<component name> as alias; alias.run()"
--- using the cluster: ---
$ condor_run 'python -c "import components.<component name> as alias; alias.run()" ' &

--- if changes in XML: ---
$ ingestWorkflow -W <pipeline name> --status enabled
```

1. Desenhar workflow - criar doc pipeline description
2. Criar versão inicial do código no Jupyter (opcional)
3. Criar repositório para o novo pipeline contendo os arquivos:
 - pipeline.xml
 - VERSION (V00_00_00)
 - ChangeLog
4. Criar repositório para o componente 1 contendo os arquivos:
 - __init__.py
 - config/component_config.xml
 - input/component_input.xml
 - VERSION (V00_00_00)
 - ChangeLog
5. Criar repositório para o componente 2 contendo os arquivos:
 - __init__.py
 - output/component_output.xml
 - etc/ingest_table.rd
 - VERSION (V00_00_00)
 - ChangeLog
6. Atualizar repositório
7. Rodar no portal do desenvolvedor
8. Debugar/fine tuning
9. **Versionar pipeline e componentes**
10. Atualizar testing
11. Testar em testing
12. Fast deploy

9. Versionar component/pipeline

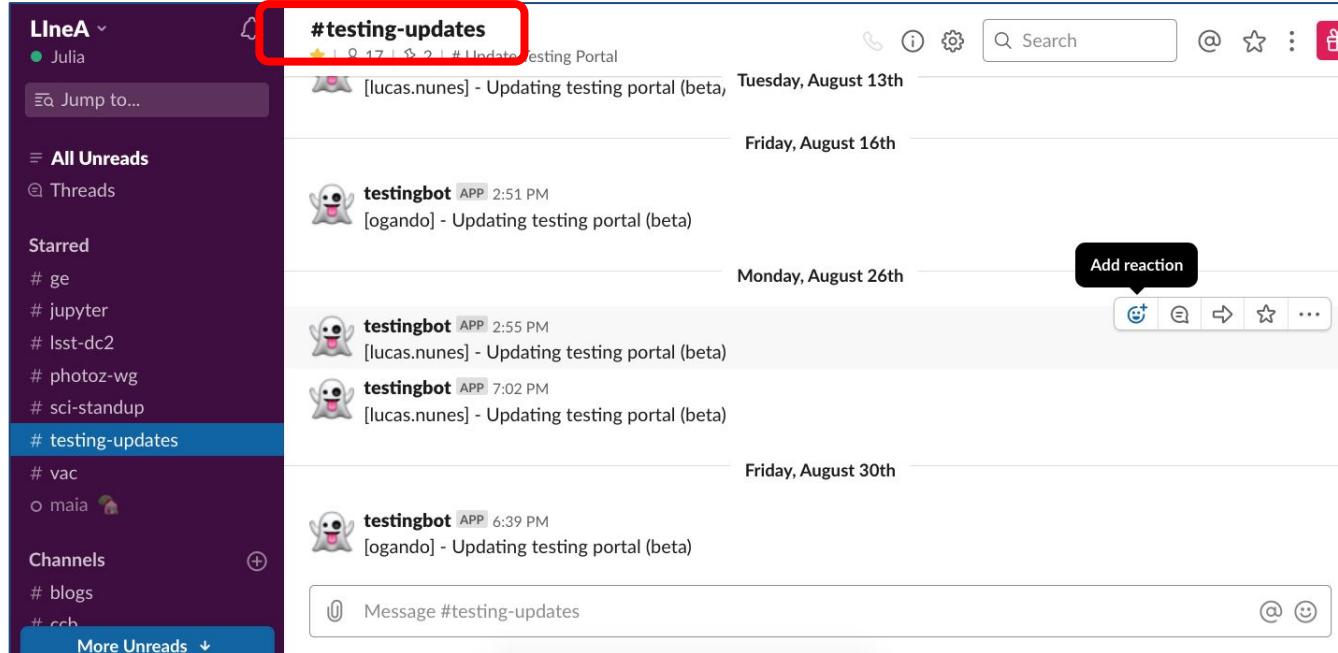


```
$ vim VERSION # change version number
$ vim ChangeLog # add high level description of changes
$ git add VERSION ChangeLog
$ git commit -m "Updating submodule with version `cat VERSION`."
$ git tag `cat VERSION`
$ git push origin
$ git push --tag
```

1. Desenhar workflow - criar doc pipeline description
2. Criar versão inicial do código no Jupyter (opcional)
3. Criar repositório para o novo pipeline contendo os arquivos:
 - pipeline.xml
 - VERSION (V00_00_00)
 - ChangeLog
4. Criar repositório para o componente 1 contendo os arquivos:
 - __init__.py
 - config/component_config.xml
 - input/component_input.xml
 - VERSION (V00_00_00)
 - ChangeLog
5. Criar repositório para o componente 2 contendo os arquivos:
 - __init__.py
 - output/component_output.xml
 - etc/ingest_table.rd
 - VERSION (V00_00_00)
 - ChangeLog
6. Atualizar repositório
7. Rodar no portal do desenvolvedor
8. Debugar/fine tuning
9. Versionar pipeline e componentes
10. **Atualizar testing**
11. Testar em testing
12. Fast deploy

10. Atualizar testing - Slack #testing-updates

```
$ sudo -H -u testing -i /home/testing/update_portal.sh
```



Passo-a-passo

1. Desenhar workflow - criar doc pipeline description
2. Criar versão inicial do código no Jupyter (opcional)
3. Criar repositório para o novo pipeline contendo os arquivos:
 - pipeline.xml
 - VERSION (V00_00_00)
 - ChangeLog
4. Criar repositório para o componente 1 contendo os arquivos:
 - __init__.py
 - config/component_config.xml
 - input/component_input.xml
 - VERSION (V00_00_00)
 - ChangeLog
5. Criar repositório para o componente 2 contendo os arquivos:
 - __init__.py
 - output/component_output.xml
 - etc/ingest_table.rd
 - VERSION (V00_00_00)
 - ChangeLog
6. Atualizar repositório
7. Rodar no portal do desenvolvedor
8. Debugar/fine tuning
9. Versionar pipeline e componentes
10. Atualizar testing
11. **Testar em testing**
12. Fast deploy

7. Rodar no portal Testing

BOOT
CAMP



INCT
e-Universo

>>

[Dashboard](#) [My Workspace](#) [Pipelines](#) [Tools](#) [Science Products](#) [Data Server](#) [Documentation](#) [Help](#)

DES Science Portal: Workflows

[Tweets by DES Science Portal](#)

The Science Portal has two instances:

- **Workflows:** hosts workflows for Quality Assessment (QA), for the creation of Value-Added Catalogs (VACs) and for Science Analysis.
- **Data Server:** provide access to the Catalog Server and published results

The system is designed to be self-evident, use the help icon "(?)" available on each page.

The Science Portal is a facility developed by [Linea](#). If you have any question please contact us through the helpdesk@linea.gov.br

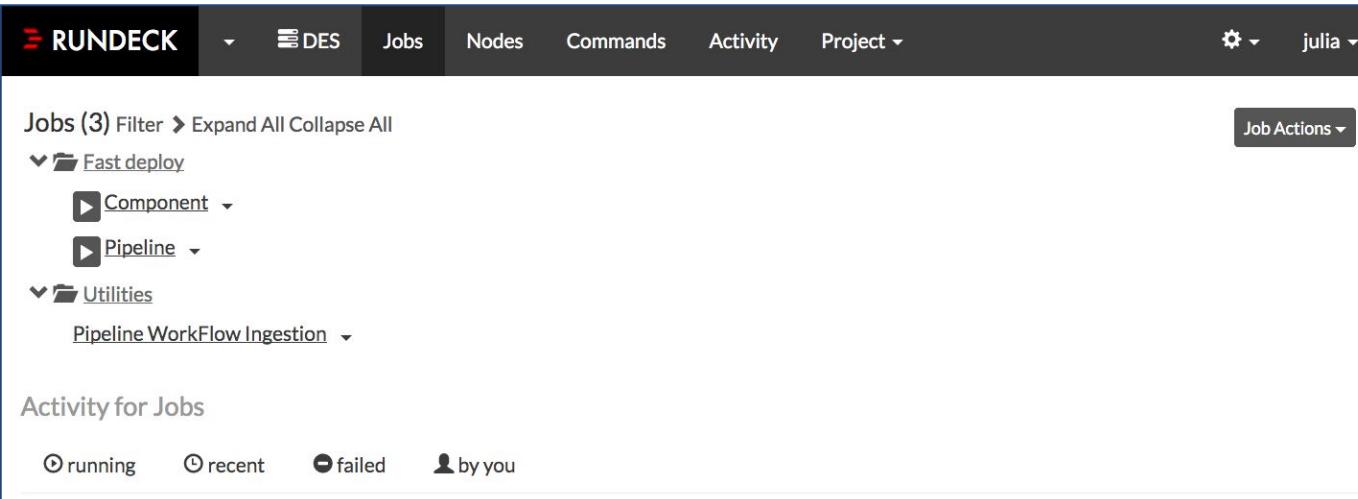
Developer Portal Instance

Powered by [Linea](#)

1. Desenhar workflow - criar doc pipeline description
2. Criar versão inicial do código no Jupyter (opcional)
3. Criar repositório para o novo pipeline contendo os arquivos:
 - pipeline.xml
 - VERSION (V00_00_00)
 - ChangeLog
4. Criar repositório para o componente 1 contendo os arquivos:
 - __init__.py
 - config/component_config.xml
 - input/component_input.xml
 - VERSION (V00_00_00)
 - ChangeLog
5. Criar repositório para o componente 2 contendo os arquivos:
 - __init__.py
 - output/component_output.xml
 - etc/ingest_table.rd
 - VERSION (V00_00_00)
 - ChangeLog
6. Atualizar repositório
7. Rodar no portal do desenvolvedor
8. Debugar/fine tuning
9. Versionar pipeline e componentes
10. Atualizar testing
11. Testar em testing
12. Fast deploy

12. Deploy em production

- E2E → update pipeline history (google drive) + email to change control board (ccb@linea.gov.br)
- Science Workflow → update pipeline history (google drive) + fast deploy (<https://fox.linea.gov.br/project/DES/jobs>)



Dúvidas?

Extra

1. Desenhar workflow
2. Criar versão inicial do código no Jupyter (opcional)
3. Criar repositório para o novo pipeline contendo os arquivos:
 - pipeline.xml
 - VERSION (V00_00_00)
 - ChangeLog
4. Criar repositório para o componente 1 contendo os arquivos:
 - __init__.py
 - config/component_config.xml
 - input/component_input.xml
 - VERSION (V00_00_00)
 - ChangeLog
5. Criar repositório para o componente 2 contendo os arquivos:
 - __init__.py
 - output/component_output.xml
 - etc/ingest_table.rd
 - VERSION (V00_00_00)
 - ChangeLog
6. Atualizar repositório
7. Rodar no portal do desenvolvedor
8. Debugar/fine tuning
9. Versionar pipeline e componentes
10. Atualizar testing
11. Testar em testing
12. Fast deploy

Sessão 23 - Criando um pipeline no portal II

BOOT
CAMP



INCT
e-Universo

Calendário do Bootcamp

	2	3	4	5	6
9:00 - 9:30	1- Introdução ao LineA L. Nicolaci - 45'	9- Introdução ao LineA II L. Nicolaci - 45'	17- Usando o Santos Dumont V. Lima - 30'	Colaborações Científicas	37- Documentação C. Silveira - 30'
9:30 - 10:00	2- LineA: Como se registrar G. Soares - 30'	10- Trabalhando no LineA A. Nunes - 45'	18- Usando o Santos Dumont: Casos de Uso H. Camacho, A. Pieres - 30'	25- DES - B. Santiago - 25'	38- Transferência de Dados J. Souza - 20'
10:00 - 10:30	Coffee break 15'		Foto e Coffee break 30'	26- DESI - F. Oliveira - 25'	39- Máquinas Virtuais, Containers e Kubernetes J. Souza - 20'
10:30 - 11:00	3- LineA: Infraestrutura C. Adean - 45'	Coffee break 15'	19- Uso Avançado da Infraestrutura C. Adean - 45'	27- SDSS - J. Schimoia - 25'	40- Rede Nacional de Ocupação G. Rossi - 30'
11:00 - 11:30	4- Landing Page do LineA C. Silveira - 30'	12- Ferramentas Astronômicas I C. Singulani - 30'	20- Portal Científico: Introdução J. Gschwend - 45'	28- TON - J. Camargo - 25'	Coffee break 20' (inicio 10:40)
11:30 - 12:00	Almoço 75'	Almoço 75'		Colaborações Científicas	Coffee break 20' (inicio 10:40)
12:00 - 12:30	Almoço 60'	Almoço 60'		29- LSST - L. Nicolaci - 30'	41- Ciberinfraestrutura para Apoio à e-Ciência A. Moura - 30'
12:30 - 13:00				30- Atividades Científicas A. Pieres - 30'	42- Atividades de EPO E. Nogueira - 30'
13:00 - 13:30	5- Spotlight talks R. Boufleur - 90'	13- Ferramentas Astronômicas II G. Costa - 60'	21- Criando um Pipeline no Portal I C. Singulani, L. Nunes - 60'	31- Analisando DES Y6 R. Gruendl - 60'	
13:30 - 14:00		14- Jupyter Notebooks: Criando J. Gschwend - 30'	22- Portal Científico: Workflows Científicos A. Pieres, M. Aguena - 45'	32- DES - Primary Products A. Pieres - 45'	43- Planejando o Futuro A. Nunes, A. Pieres - 150'
14:00 - 14:30	6- Serviços F. Massena - 30'	15- Jupyter Notebooks: Usando e Desenvolvendo A. Pieres, M. Aguena - 60'		33- DES Portal - Data Preparation J. Gschwend - 45'	
14:30 - 15:00	7- Jupyter Notebooks: Introdução G. Soares - 30'		23- Criando um Pipeline no Portal II J. Gschwend - 60'	Coffee break 30'	44- Perspectivas do LineA L. Nicolaci - 30'
15:00 - 15:30	Coffee break 30'	Coffee break 30'		34- Validando Imagens DES Y6 A. Pieres - 30'	Encerramento 30'
15:30 - 16:00		Coffee break 30'	Coffee break 30'	35- Validando Catálogos DES Y6 M. Aguena - 30'	
16:00 - 16:30	8- Jupyter Notebooks: Exemplos A. Pieres - 60'	16- Jupyter Notebooks: Casos de Uso F. Oliveira, R. Boufleur - 60'	24- Requisitos para o Portal do LSST M. Sanchez - 45'	36- Explorando Dados DES Y6 M. Aguena - 30'	Legenda Apresentação
16:30 - 17:00					Tutorial Hacksession/Hands on
17:00 - 17:30	Q&A 30'	Q&A 30'	Q&A 30'		

ATUALIZAR TODOS OS PIPELINES E COMPONENTS

/home/aguena/bin_share/gitthem

criar uma component nova

/home/aguena/bin_share/make_component

criar um pipeline novo

/home/aguena/bin_share/make_pipeline

Dar um tag e push no pipeline

/home/aguena/bin_share/tagueia

ADICIONAR COMMITS NO CHANGELOG

/home/aguena/bin_share/upchange