



João Pedro Pinheiro
Data Engineer



Migração de Arquitetura **BigData**

15 / MARÇO / 2021

Apresentação



João Pedro Pinheiro
Data Engineer



Formação Acadêmica



DEPARTAMENTO
DE INFORMÁTICA
PUC-RIO



Experiência Profissional



SIEVE

estante
virtual

OLX Brasil

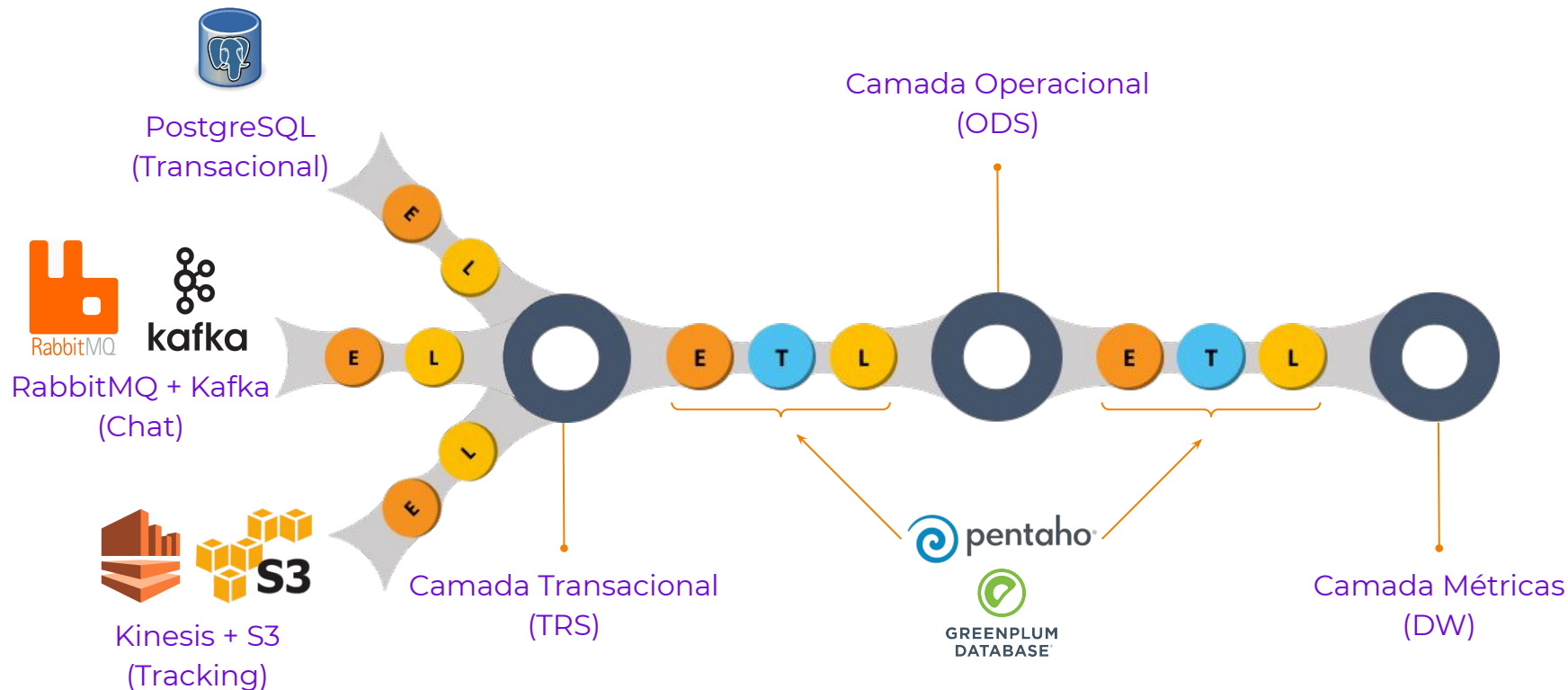


Empreendedorismo

Beni.

GLIO.

Contexto Pré-Migração



Motivação

Principais pontos que motivaram à realização dessa migração de arquitetura.



Planejamento



Aprendizados

Durante e, principalmente, após a entrega, levantamos os principais pontos de melhoria para futuros projetos



Planejamento

Não foi pensado um modo de **entrega por partes**, como costuma-se fazer em eng. de soft. Foi assumido que era **inviável** o sistema antigo co-existir com o sistema novo. E essa premissa se manteve até às vésperas da virada de chave, quando pensou-se em uma **solução**.

Priorização

Quebrou-se o desenvolvimento em **camadas** (ods e dw) e **tabelas**. Em um segundo momento, definimos **assuntos** (conjunto de tabelas) a serem entregues. Somente próximo ao fim do projeto, priorizamos entregas e validações das tabelas **mais utilizadas**.

Estimativas

Devido a complexidade das **regras de negócio** envolvidas e **falta de conhecimento** do processo como um todo da equipe, **subestimamos** a entrega final. Isso gerou algumas várias estimativas de entrega (virada de chave) **sem sucesso**.

Aprendizados

Durante e, principalmente, após a entrega, levantamos os principais pontos de melhoria para futuros projetos



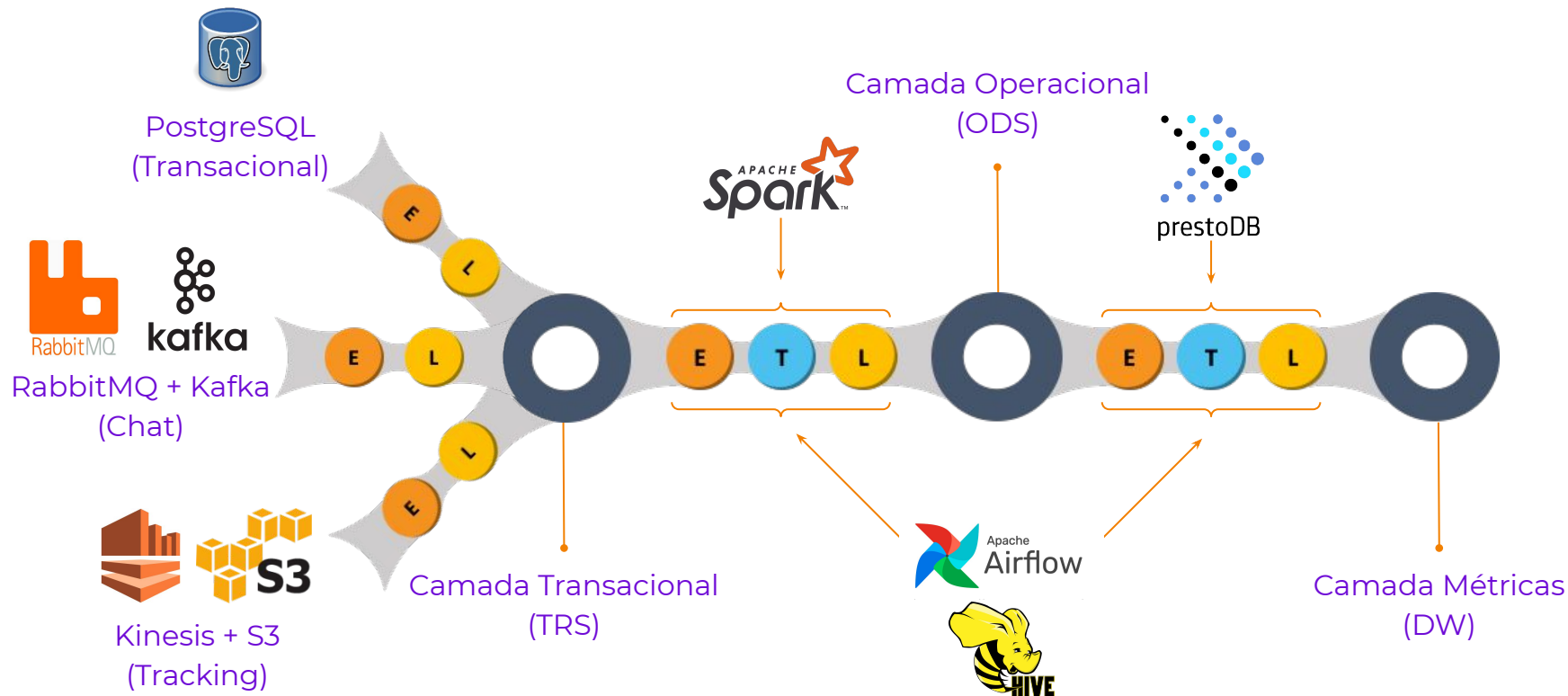
Mudanças na Equipe

Pessoas que detinham o **conhecimento** do **processo** como um todo, deixaram a empresa. Isso afetou, principalmente, no processo de **homologação** das tabelas.

Otimizações Precipitadas

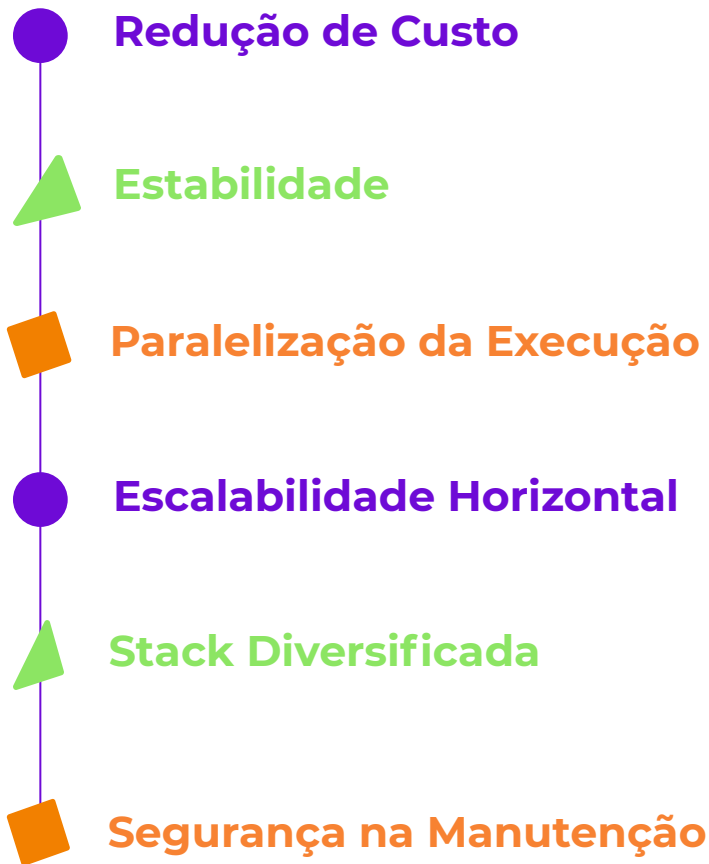
Estratégias de carga dos dados foram modificadas (com relação ao original), visando **melhor performance** e entrega dentro do **SLA**. Quando, na verdade, algumas delas tiveram que ser **revertidas**, pois não passaram na etapa de **homologação**.

Contexto Pós-Migração



Resultados

Principais mudanças conquistadas com a nova arquitetura.



Visão do Usuário

Principais mudanças conquistadas com a nova arquitetura.

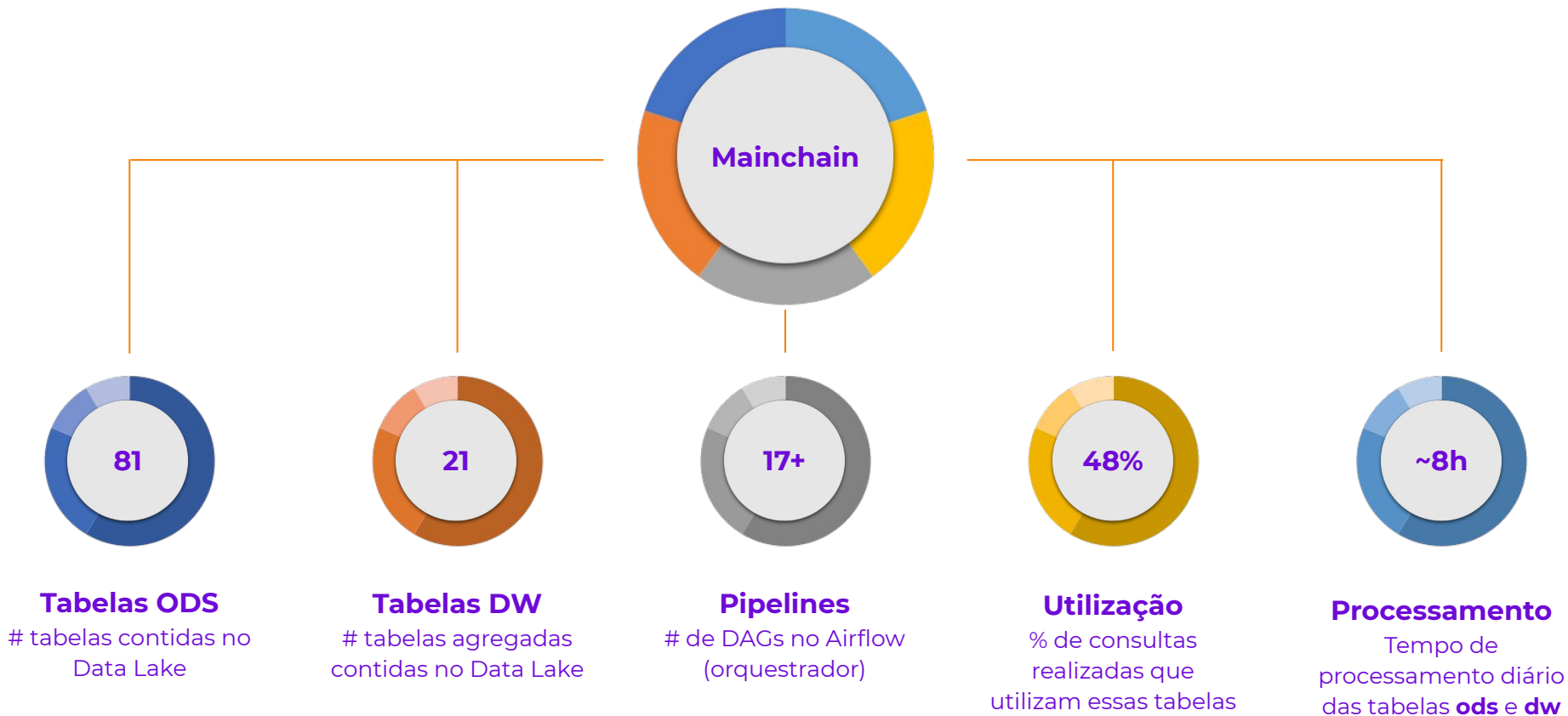


Delivery Control

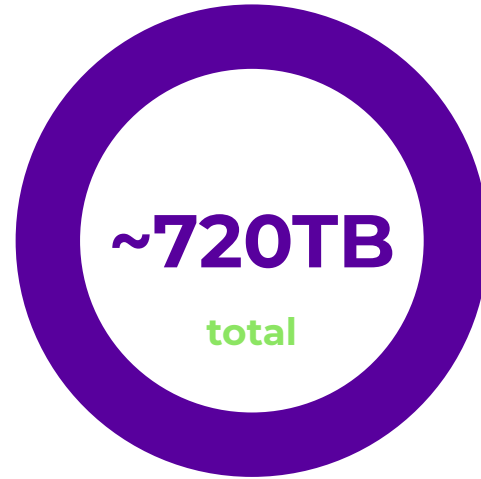


Entregas particionadas

Volumetria

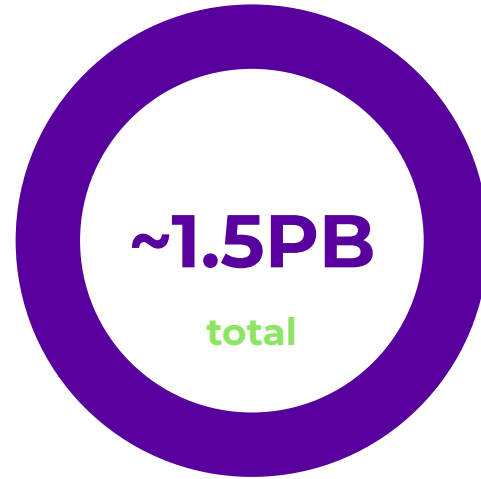


Volumetria




Lurker (tracking)

Volumetria



DataLake

Exemples



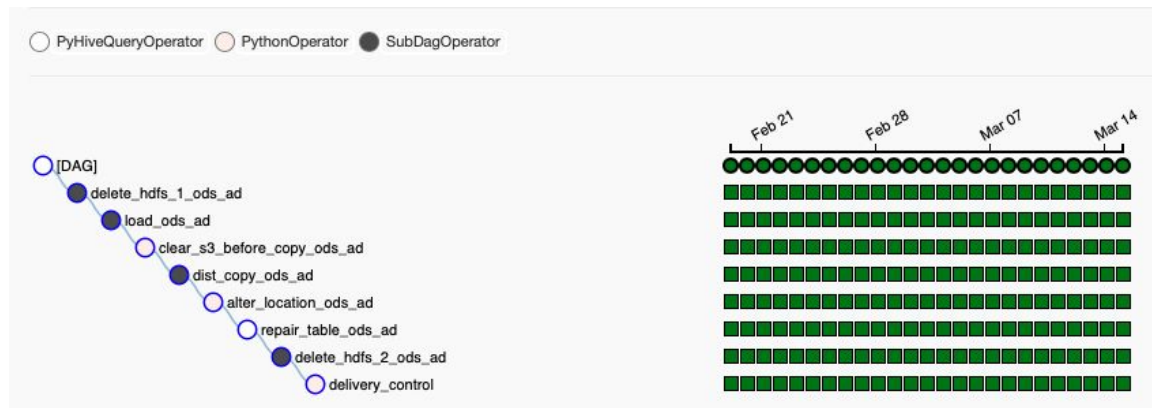
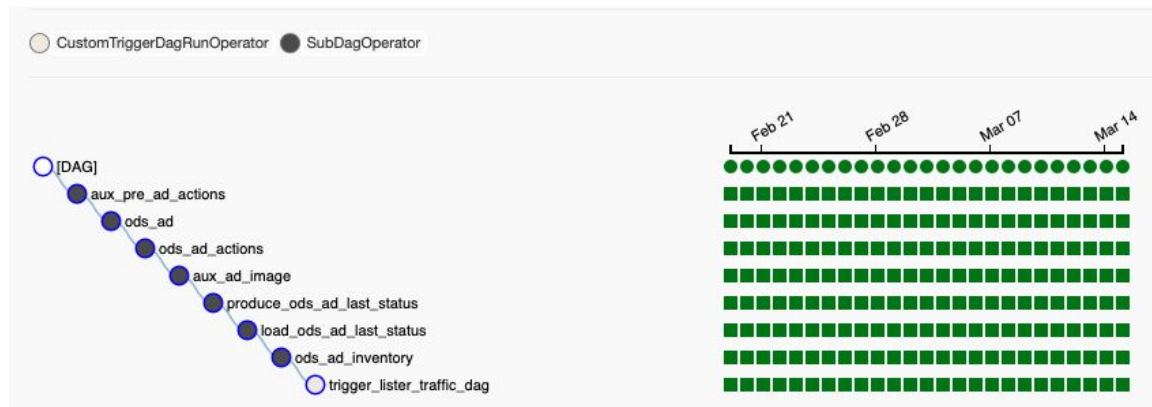
DAGs
Data Profiling
Browse
Admin
Docs
About

DAGs

Search:

		DAG	Schedule	Owner	Recent Tasks	Last Run
	<div>On</div>	dw-pipeline_0-start	None	corebl_bigdata	<div>8</div> <div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> </div>	2021-03-15 08:59
	<div>On</div>	dw-pipeline_1.1-ad-summary	None	corebl_bigdata	<div>9</div> <div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> </div>	2021-03-15 09:18
	<div>On</div>	dw-pipeline_1.2-ad-contacts	None	corebl_bigdata	<div>10</div> <div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> </div>	2021-03-15 10:22
	<div>On</div>	dw-pipeline_2.1-lister-daily	None	corebl_bigdata	<div>11</div> <div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> </div>	2021-03-15 09:20
	<div>On</div>	dw-pipeline_2.2-lister-weekly	None	corebl_bigdata	<div>11</div> <div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> </div>	2021-03-15 09:30
	<div>On</div>	dw-pipeline_2.3-lister-monthly	None	corebl_bigdata	<div>11</div> <div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> </div>	2021-03-15 10:10
	<div>Off</div>	dw-pipeline_3-monetization-total	30 13 ***	corebl_bigdata	<div>19</div> <div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> </div>	2020-11-27 00:06
	<div>On</div>	dw-pipeline_4.1-firebase-traffic-kpi	None	corebl_bigdata	<div>7</div> <div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> </div>	2021-03-15 13:00
	<div>On</div>	dw-pipeline_4.2-daily-metrics	None	corebl_bigdata	<div>26</div> <div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> </div>	2021-03-15 18:50
	<div>On</div>	dw-pipeline_4.3-weekly-metrics	None	corebl_bigdata	<div>26</div> <div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> </div>	2021-03-15 18:57
	<div>On</div>	dw-pipeline_4.4-monthly-metrics	None	corebl_bigdata	<div>29</div> <div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> </div>	2021-03-15 19:35
	<div>On</div>	dw-pipeline_5-automatic-tests	None	corebl_bigdata	<div>2</div> <div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> </div>	2021-01-19 15:17
	<div>On</div>	ext-pipeline_metrics_traffic	None	corebl_bigdata	<div>7</div> <div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> </div>	2021-03-15 13:08
	<div>On</div>	ods-pipeline_0-start	0 5 ***	corebl_bigdata	<div>7</div> <div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> </div>	2021-03-14 05:00
	<div>On</div>	ods-pipeline_1-pre-ads	None	corebl_bigdata	<div>9</div> <div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> </div>	2021-03-15 05:21
	<div>On</div>	ods-pipeline_2-dimensions	None	corebl_bigdata	<div>12</div> <div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> </div>	2021-03-15 05:30
	<div>On</div>	ods-pipeline_3-accounts	None	corebl_bigdata	<div>4</div> <div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> </div>	2021-03-15 05:39
	<div>On</div>	ods-pipeline_4-ads	None	corebl_bigdata	<div>9</div> <div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> </div>	2021-03-15 05:58
	<div>On</div>	ods-pipeline_5-lister-traffic	None	corebl_bigdata	<div>9</div> <div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> </div>	2021-03-15 07:45
	<div>On</div>	ods-pipeline_6-contacts	None	corebl_bigdata	<div>7</div> <div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> </div>	2021-03-15 08:59

Exemplos



Exemplos

Friday, March 12th ▾



BigDataBot APP 2:16 AM



ods pipeline started



BigDataBot APP 6:10 AM



dw pipeline started



BigDataBot APP 10:08 AM



ods pipeline finished



BigDataBot APP 10:54 AM



Failed task waiting to retry.

Task: repair_table_monetization_total

Dag: support_update-keys-payments-pipeline

Execution Time: 2021-03-12T13:00:00+00:00



BigDataBot APP 11:31 AM



Task failed and died.

Task: tableau_update_dw.monetization_total


Dag: dw-pipeline_4.4-monthly-metrics

Execution Time: 2021-03-12T13:43:20+00:00



Monetization refreshed

Exemplos

 **Delivery Control**

▼

Monitoramento

Delivery Control

Tabelas por Time Responsável

Estratégia

Reference Dates

2021-03-14

Search

×

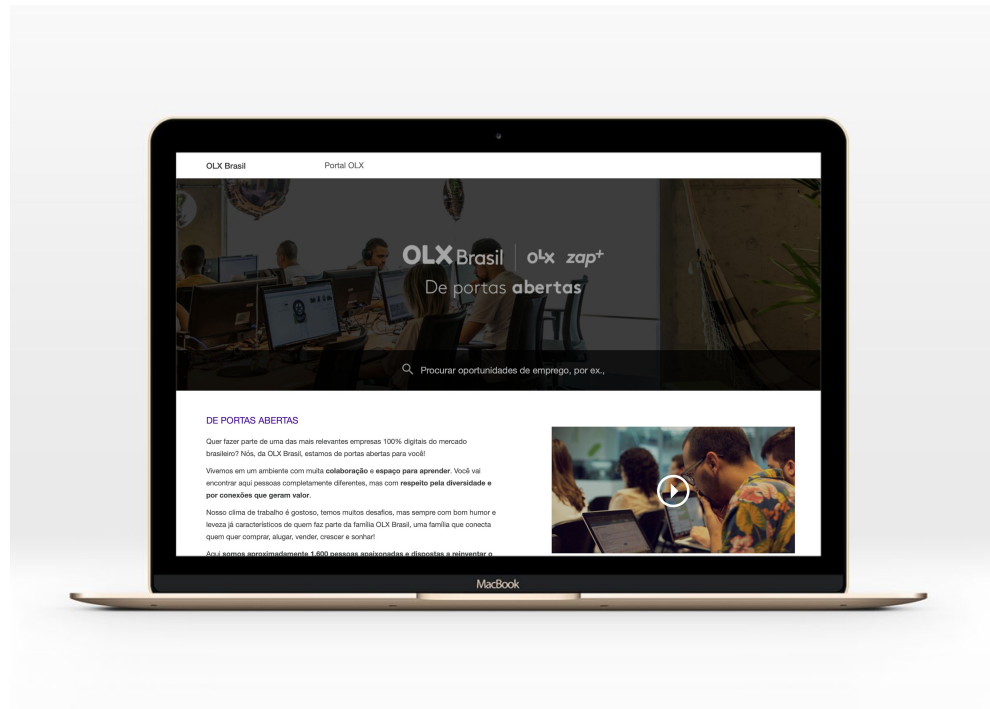
Tabela	Data de Referência	Entregue em	Time Resp.
stg.ad	2021-03-14	15/03/2021 02:30:02	Data Services
ods.dm_refusal_reason_text	2021-03-14	15/03/2021 02:32:49	Data Services
ods.dm_reason_removed_detail_path	2021-03-14	15/03/2021 02:32:48	Data Services
ods.dm_admin	2021-03-14	15/03/2021 02:32:53	Data Services
ods.dm_deciding_module	2021-03-14	15/03/2021 02:32:54	Data Services
ods.dm_queue	2021-03-14	15/03/2021 02:36:34	Data Services
ods.dm_location	2021-03-14	15/03/2021 02:39:29	Data Services
ods.dm_platform_path	2021-03-14	15/03/2021 02:39:29	Data Services
ods.dm_category	2021-03-14	15/03/2021 02:39:34	Data Services
ods.dm_refusing_module	2021-03-14	15/03/2021 02:39:34	Data Services

10 rows

1-10 of 78

Vagas Abertas

[smartrecruiters.com/
OLXBrasil](https://smartrecruiters.com/OLXBrasil)



Obrigado!

joao.pinheiro@olxbr.com
[linkedin.com/in/jpedrojp](https://www.linkedin.com/in/jpedrojp)