

Evolução Integrada

Plataforma de Dados e Estratégias de Testes



Olá, eu sou o Samuel!

Ele/he/él

Engenheiro de Dados Sr
+10 anos como tecnologista
ThoughtWorker há 5 anos
Gaúcho





Olá, eu sou a Vanessa!

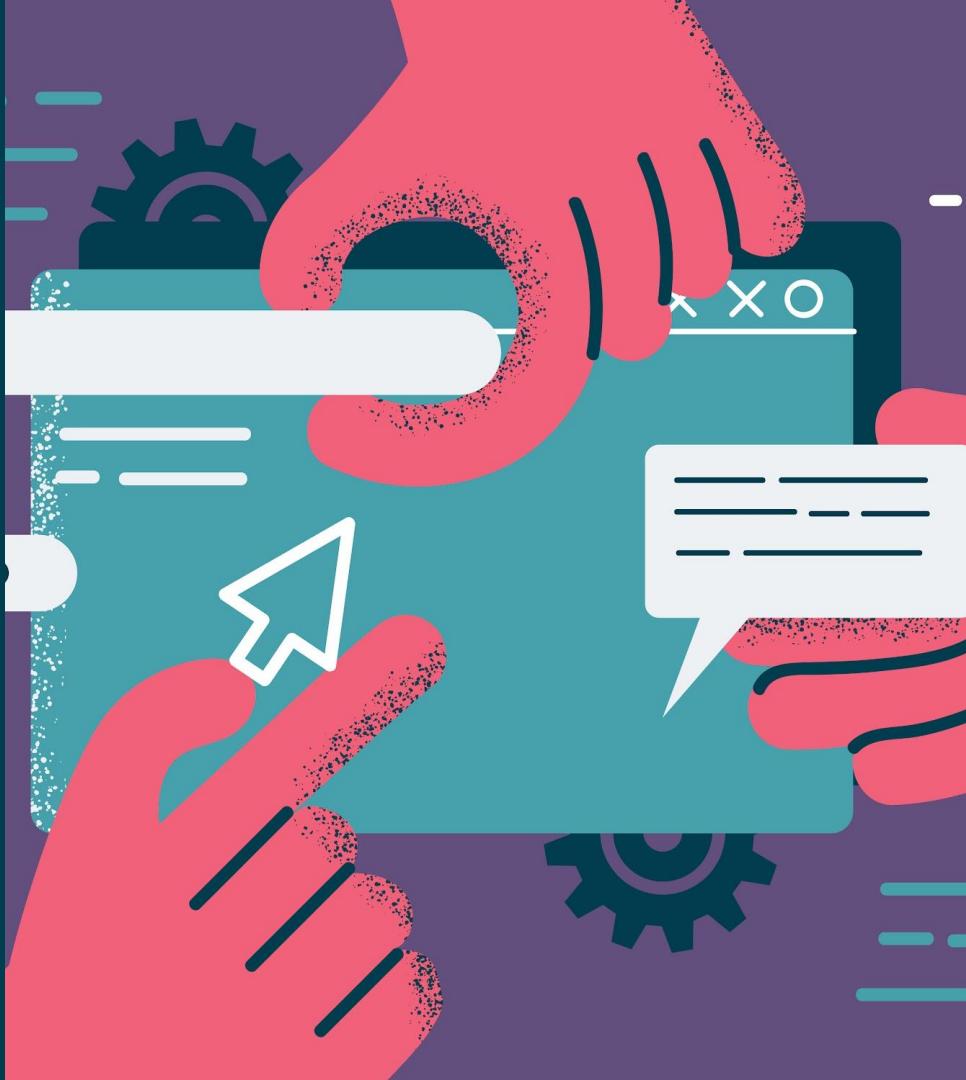
Ela/She/ella

Engenheira de Dados Sr
8 anos como tecnologista
ThoughtWorker há +3 anos
Paraense-paulistana



Entendendo o cenário

/thoughtworks



O caso de uso

Recomendação de conteúdo

"A união entre o acolhimento humano e a eficiência da tecnologia, pautado em inteligência artificial."



Aplicação de recomendação de conteúdo baseado no perfil da hóspede para indicações personalizada.

Desafios do projeto

Objetivo

Construir uma plataforma para habilitar aplicações de **Inteligência Artificial e Machine Learning** para transformar a experiência de hospedagem.

Projeto
Green Field

Jornada inicial
em dados

Orçamento
enxuto e prazo
curto

Dependência de
fornecedores
externos



Pilares de Engenharia de Excelência

Arquitetura Evolutiva

*Portões de qualidade
(Quality gates)*

Teste Automatizado

Integração Contínua (CI)

Entrega Contínua (CD)

Infraestrutura como Código

Cloud First

Programação em Par

Construir com segurança integrada





Plataforma & Produto de dados

Recomendação de Conteúdo

Arquitetura de solução

Fontes externas



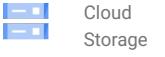
Aplicação interna

Google Cloud

Ingestão



Armazenamento



Pipeline de dados & IA



Apresentação



Aplicação



Orquestração



Cloud Composer

Tudo Restaurantes Bares e baladas

Bares e baladas

Santana

Drinks autorais



Tokio SP

Karaokê



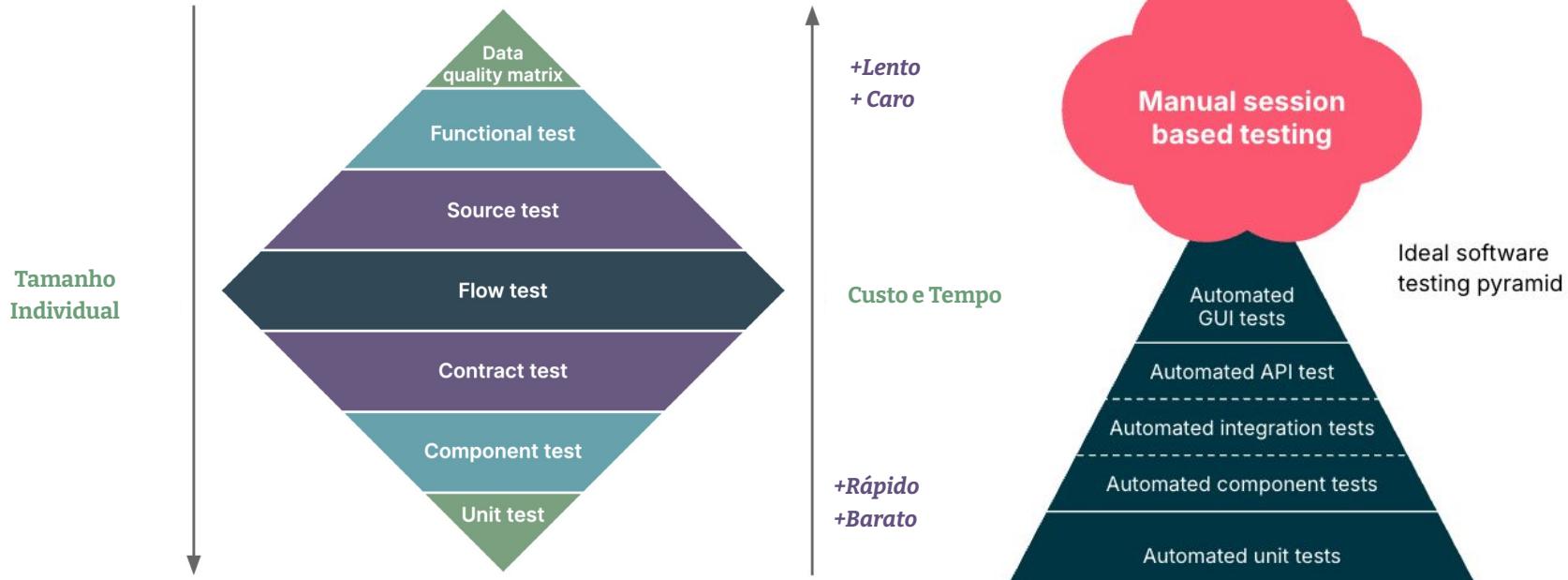
Caracol Bar

Drinks e petiscos sofisticados



Estratégia de testes

Diamante de Testes



O nosso Diamante de Testes



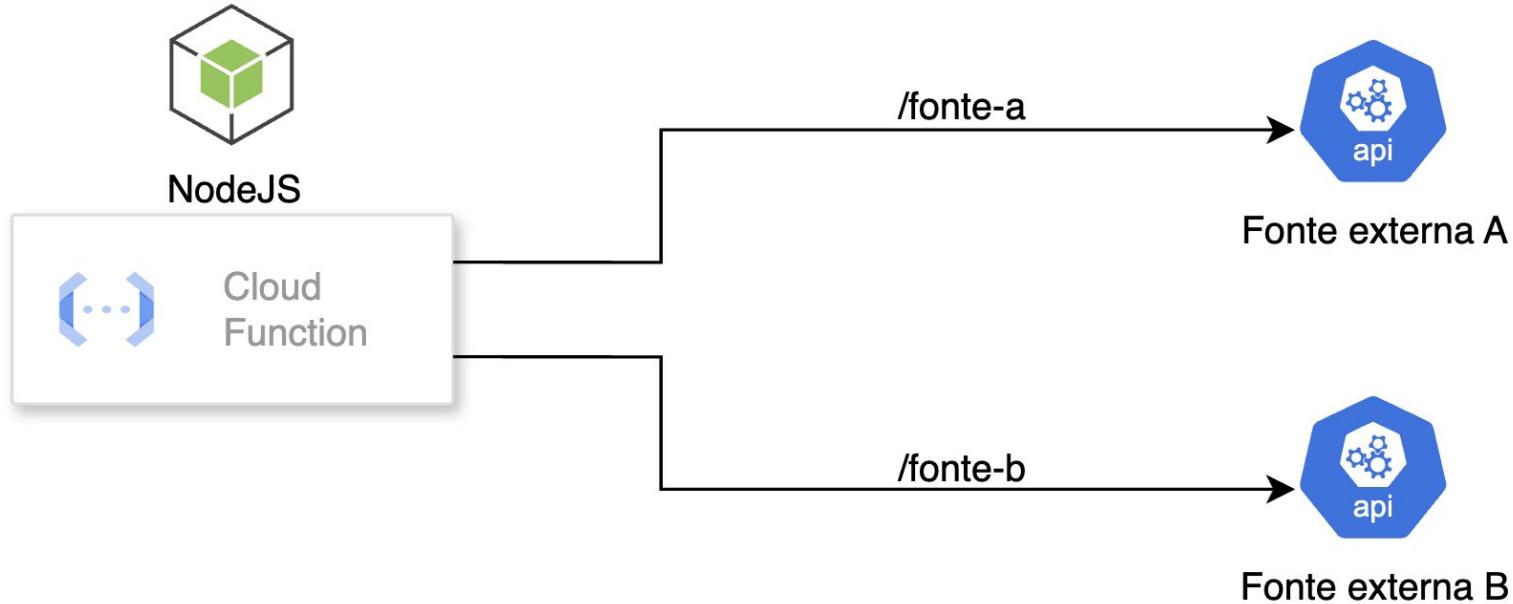
Mas e os dados?

/thoughtworks



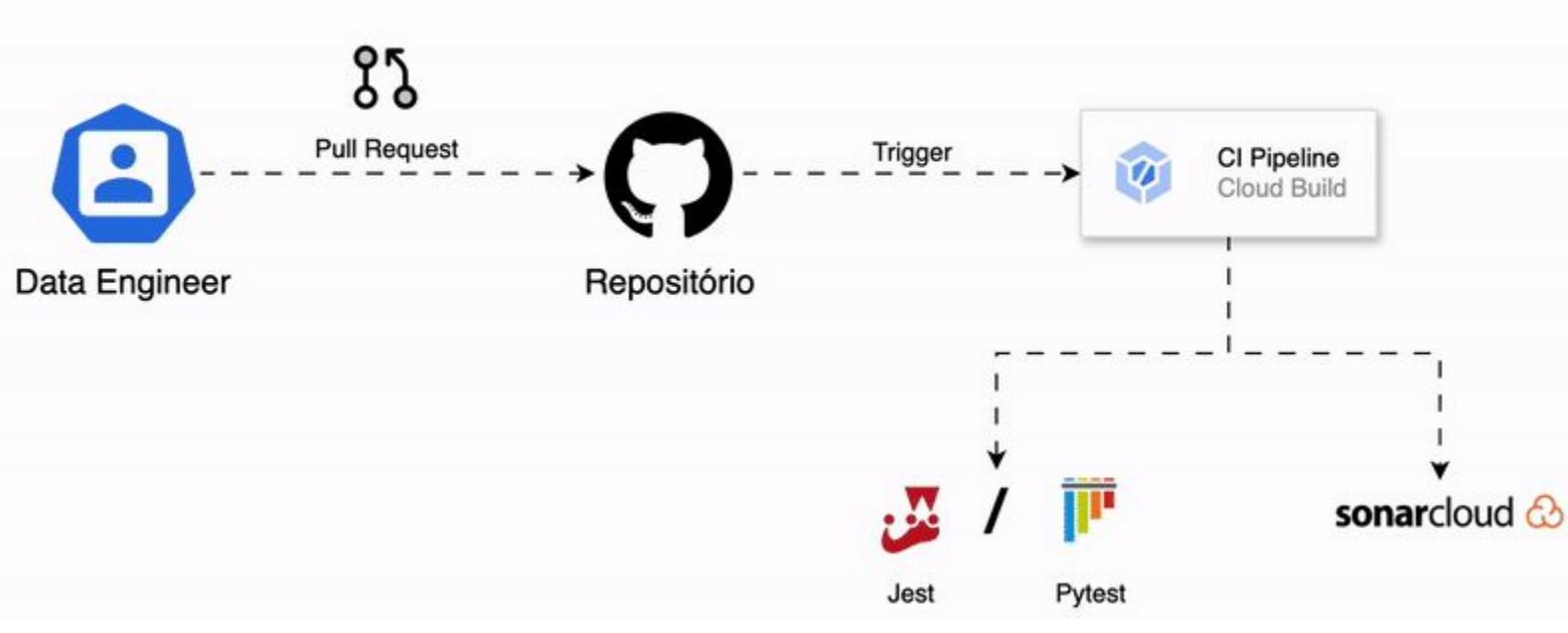
Ingestão de Dados

Contexto

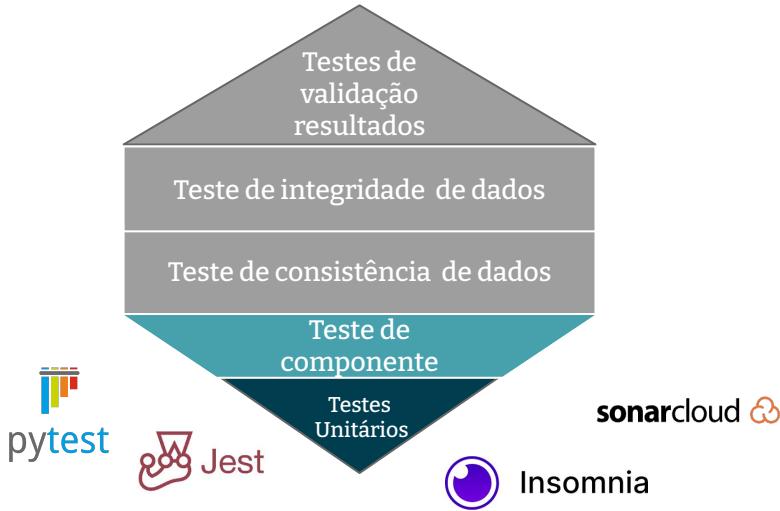


Ingestão de Dados

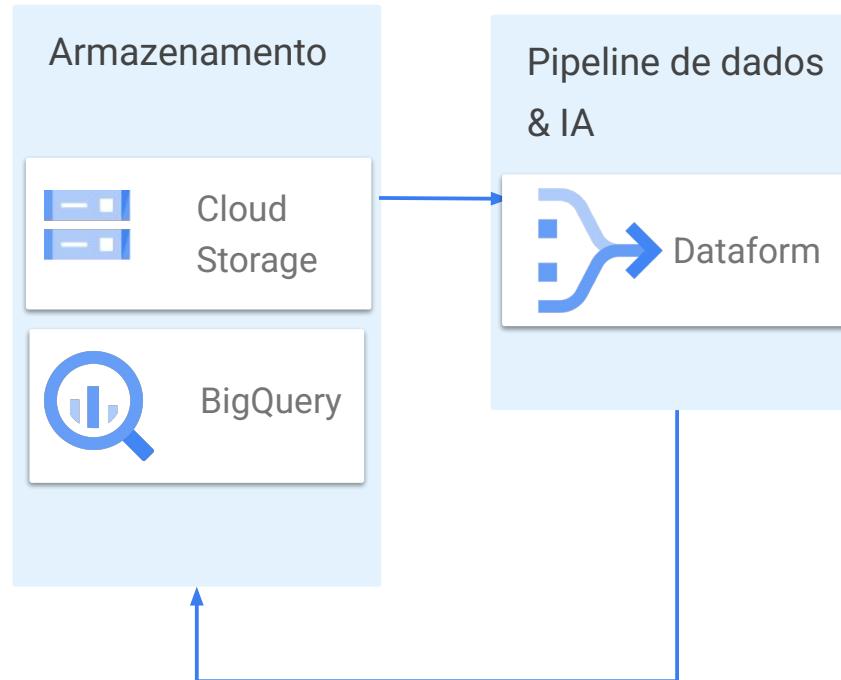
Pipeline de CI



Construindo o Diamante de testes

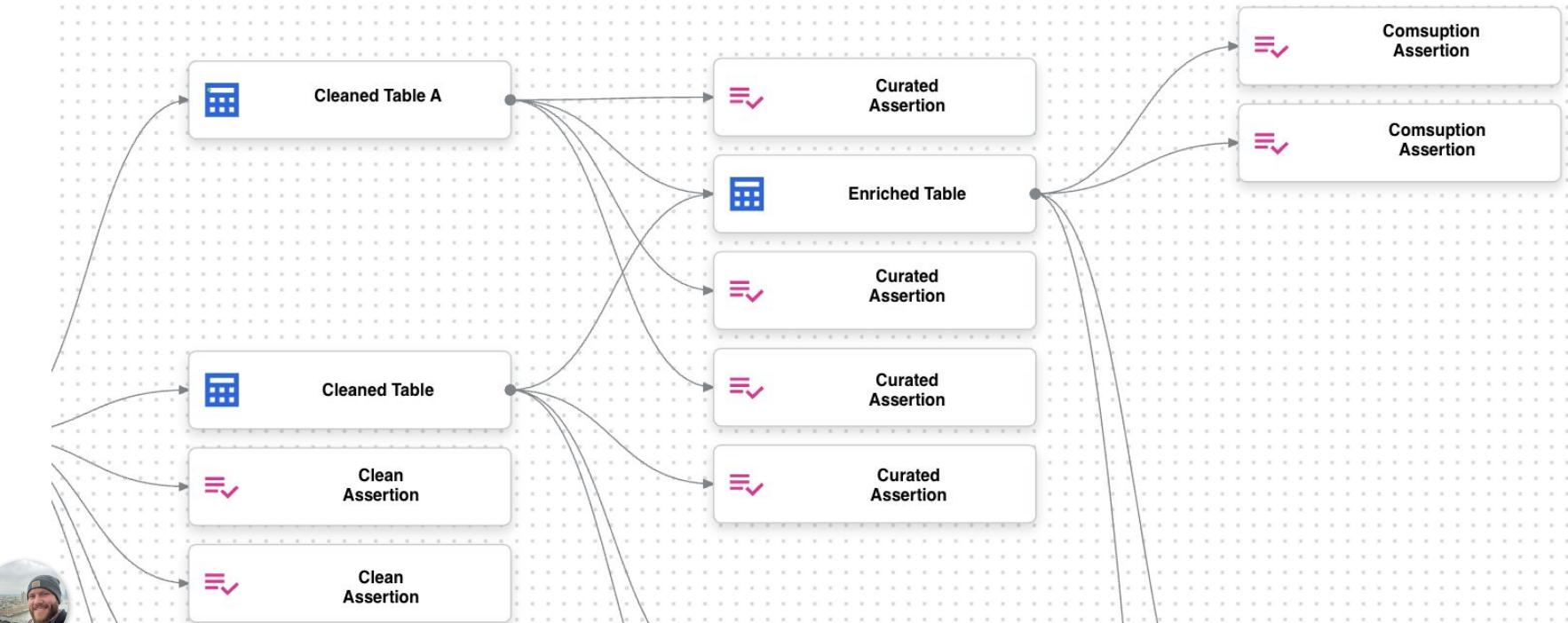


Limpeza e enriquecimento dos Dados



Limpeza e enriquecimento dos Dados

Movimentando dados com Dataform



Limpeza e enriquecimento dos Dados

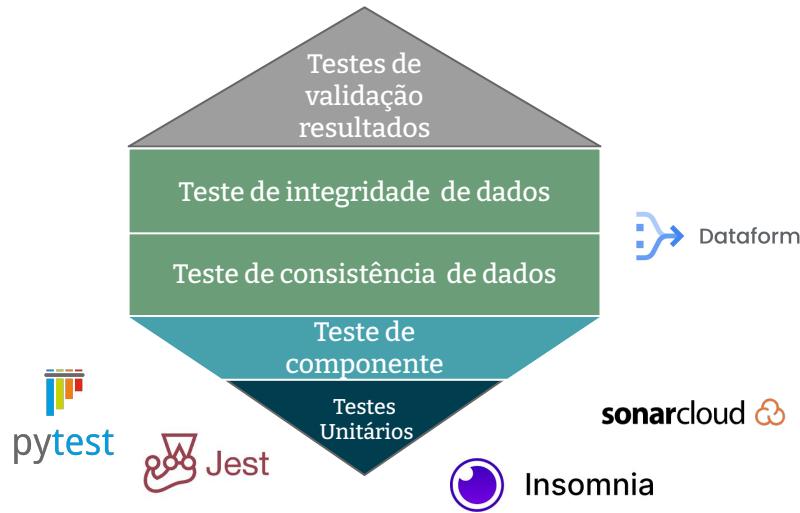
Assertions com Dataform

```
config {  
  type: "table",  
  assertions: {  
    uniqueKey: ["user_id"],  
    nonNull: ["user_id", "customer_id"],  
    rowConditions: [  
      'signup_date is null or signup_date > "2019-01-01"',  
      'email like "%@%.%"'  
    ]  
  }  
}  
SELECT ...
```

```
config { type: "assertion" }  
  
SELECT  
  *  
FROM  
  ${ref("sometable")}  
WHERE  
  a IS NULL  
  OR b IS NULL  
  OR c IS NULL
```



Evoluindo o Diamante de testes



Consumo e Disponibilização de Dados

Validação dos resultados



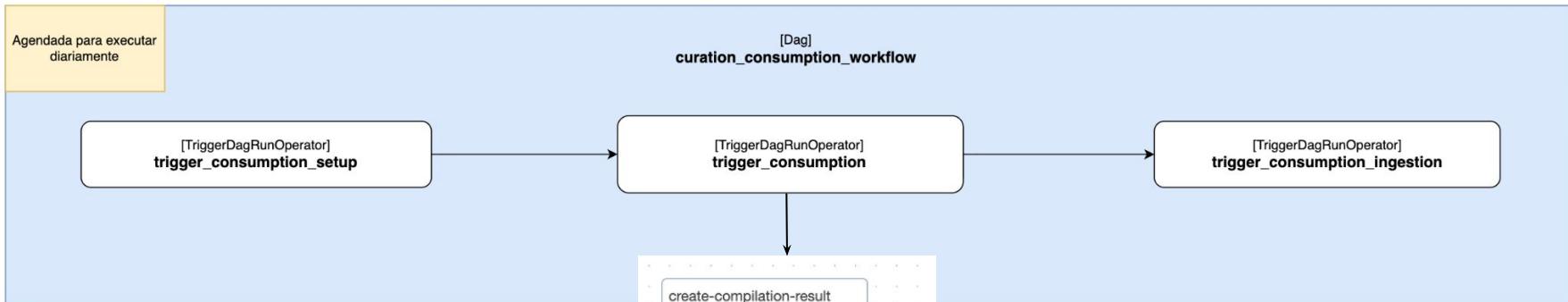
A map of a portion of São Paulo, Brazil, showing streets like Av. Rebouças, R. Cônego Eugênio Leite, and Rua Atlântica. A red marker indicates the location of 'Maní'. Other points of interest shown include 'Bacio di Latte - Rua Dos Pinheiros' and 'Padoca do Maní'. A small circular icon in the top right corner of the map area contains three circles: two white and one blue.

Ranking	Nome	Endereço	Tipo
1	Maní	R. Joaquim Antunes, 210 - São Paulo - SP	Restaurante
2	Parque Ibirapuera	Av. Pedro Álvares Cabral, s/n - São Paulo - SP	Ponto Turístico
3	GP de Fórmula 1	Av. Sen. Teotônio Vilela, 261 - São Paulo - SP	Evento
...

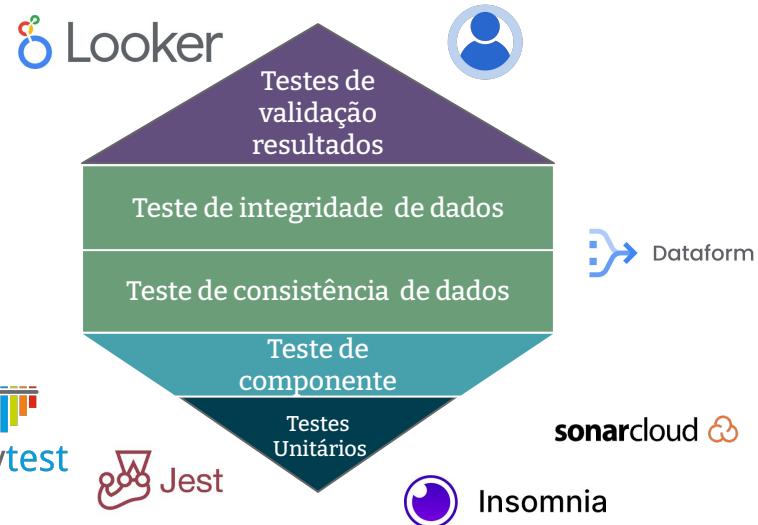


Consumo e Disponibilização de Dados

Orquestração



O caminho até aqui



Aprendizados



Entregas enxutas e palpáveis para a cliente



Entregas de plataforma associadas as funcionalidades prioritárias do produto



Paramento entre: Dados + Produto + Design + QA + Infra



Time multidisciplinar e **diverso**



Perguntas ?



Obrigada

Vanessa Anjos

Engenheira de Dados

<https://www.linkedin.com/in/vanessa-p-anjos/>

Samuel Locatelli

Engenheiro de Dados

<https://www.linkedin.com/in/sam-locatelli/>

Material apresentado

https://github.com/vpanjos/tdc_sp_25_trilha_eng_dados/

/thoughtworks

