

# Tópicos em Engenharia de Software: Relatório Técnico

Pedro Henrique Gonçalves Barcelos

<https://github.com/pedrobarcelos/pratica-4>

## Introdução

Neste relatório técnico, abordaremos o processo de implementação de testes para funções que processam grandes volumes de dados, com foco na função ‘avgAgeCountry’. Exploraremos os cenários de teste, os objetivos por trás desses testes e a importância de testar funções em ambientes de Big Data.

## 1 Desenvolvimento de Testes Unitários

### 1.1 Função avgAgeCountry

A função ‘avgAgeCountry’ é projetada para calcular a média de idades com base no país de origem dos indivíduos em um grande conjunto de dados. Para garantir a confiabilidade dessa função, implementamos vários cenários de teste:

- **\*\*Cenário de Teste: Dados Válidos - Caso de Sucesso:\*\*** Testamos a função com um conjunto de dados válido, onde todos os campos necessários (nome, idade e país) estão preenchidos corretamente.
- **\*\*Cenário de Teste: Dados Válidos com Transformação de Idade:\*\*** Extensão do cenário anterior, introduzindo uma função de transformação para as idades antes de calcular a média.
- **\*\*Cenário de Teste: Dados com Campos Ausentes:\*\*** Testamos a função com dados que têm campos ausentes, como registros que não possuem o campo "idade".
- **\*\*Cenário de Teste: Dados com Países Ausentes:\*\*** Testamos a função com dados onde alguns registros não têm o campo "país".
- **\*\*Cenário de Teste: Conjunto de Dados Vazio:\*\*** Verificamos o comportamento da função quando o conjunto de dados está vazio.

### 1.2 Outras Funções de Processamento

Além da função avgAgeCountry, criamos funções adicionais, como ‘read-json-file’ para ler arquivos JSON e ‘save-to-json’ para salvar dados em arquivos JSON. Essas funções desempenham um papel fundamental em fluxos de trabalho de processamento de dados e contribuem para a integridade dos processos de manipulação de dados em larga escala.

### 1.3 Função de Transformação

A função de transformação é aplicada à idade antes de calcular a média. Por exemplo, convertamos a idade de anos para meses. Isso é relevante em cenários onde as unidades de medida dos dados podem variar. Para testar a função de transformação, criamos um cenário de teste específico, aplicando uma função de transformação às idades e verificando se a média é calculada corretamente após a transformação.

## 2 Objetivos dos Testes

Os objetivos por trás dos testes desenvolvidos são:

- Prevenir problemas, como valores ausentes, campos nulos ou campos ausentes, garantindo a funcionalidade correta da função ‘avgAgeCountry’.
- Validar os resultados produzidos pela função em diferentes cenários, garantindo que a média de idades seja calculada corretamente.
- Demonstração da relevância do teste unitário, permitindo a criação de diferentes asserções para verificar o comportamento da função.
- Identificação de situações problemáticas e melhoria contínua da funcionalidade.

## 3 Reflexão sobre Testes em Big Data

Escrever testes para funções em ambientes de Big Data apresenta desafios significativos. A importância de testar funções que processam grandes volumes de dados reside na garantia da confiabilidade dos resultados. Com a crescente disponibilidade de dados, a integridade dos resultados é crucial. A otimização de desempenho, especialmente em ambientes PySpark, é essencial.

Para testar o desempenho em um cenário avaliado, poderíamos implementar testes de desempenho que medem o tempo necessário para a função ‘avgAgeCountry’ processar conjuntos de dados de diferentes tamanhos. Isso ajudaria a identificar gargalos de desempenho e aprimorar a eficiência da função em ambientes de Big Data.

## 4 Conclusão

Os testes desempenham um papel crítico na garantia da qualidade do software que processa grandes volumes de dados. Neste relatório, enfocamos os testes unitários da função ‘avgAgeCountry’ e destacamos a importância de verificar a validade dos resultados, identificar problemas em dados do mundo real e manter a funcionalidade ao longo do tempo.

Além disso, enfatizamos a relevância das funções adicionais, como leitura e gravação de dados em formatos como JSON, no ecossistema de processamento de dados. Garantir a confiabilidade e a eficácia dessas funções é igualmente importante para a integridade dos processos de manipulação de dados em larga escala. Portanto, é crucial que as funções e os testes sejam desenvolvidos e mantidos com o devido cuidado para garantir a qualidade dos resultados.