

APLICAÇÃO DE PRINCÍPIOS FAIR E ORQUESTRADORES DE FLUXOS DE TRABALHO NO CONTEXTO DE ANÁLISES DE DADO



Maria do Ceu Pontes Vieira¹; Marianna Pereira Silva Ramalho¹; Yasmin Yngrid Mendes de Brito¹; Gabriela Pessoa Lima de Souza Medeiros¹; Marcel da Câmara Ribeiro-Dantas^{2,3}.

Universidade Potiguar

¹Curso de Medicina. ²Programa de Pós-Graduação em Biotecnologia. ³Programa de Pós-Graduação em Administração. Campus Salgado Filho.

marcel.camara@unp.br

Introdução

A análise de dados na saúde enfrenta desafios crescentes devido ao volume, diversidade e complexidade dos dados gerados, como genomas, exames clínicos e imagens médicas. Orquestradores de fluxo de trabalho, como Nextflow, Snakemake e Apache Airflow, permitem o gerenciamento eficiente de pipelines complexos, facilitando a integração de ferramentas heterogêneas e promovendo a reprodutibilidade. Este estudo analisa a importância desses sistemas na análise de dados de saúde, com ênfase no impacto na medicina personalizada.

Objetivos

Realizar uma revisão narrativa da literatura buscando entender o panorama de orquestradores de fluxos de trabalho.

Metodologia

A revisão foi realizada através da base de dados PubMed, utilizando os descritores: “workflow management system” ou “workflow management systems” e scalability ou reproducibility ou portability ou FAIR, sendo “e” e “ou” os operadores lógicos. Foram incluídos artigos publicados de 2014 a 2024, em inglês, e com o artigo completo disponível de forma gratuita, totalizando 53 publicações. Após aplicados os critérios de exclusão, restaram 44 trabalhos. Além dos 44 trabalhos encontrados na nossa revisão, durante a leitura destes e investigação do tema, outros trabalhos foram encontrados e trazidos para discussão.

Agradecimentos

Ao Instituto Ânima e a Universidade Potiguar.

Resultados

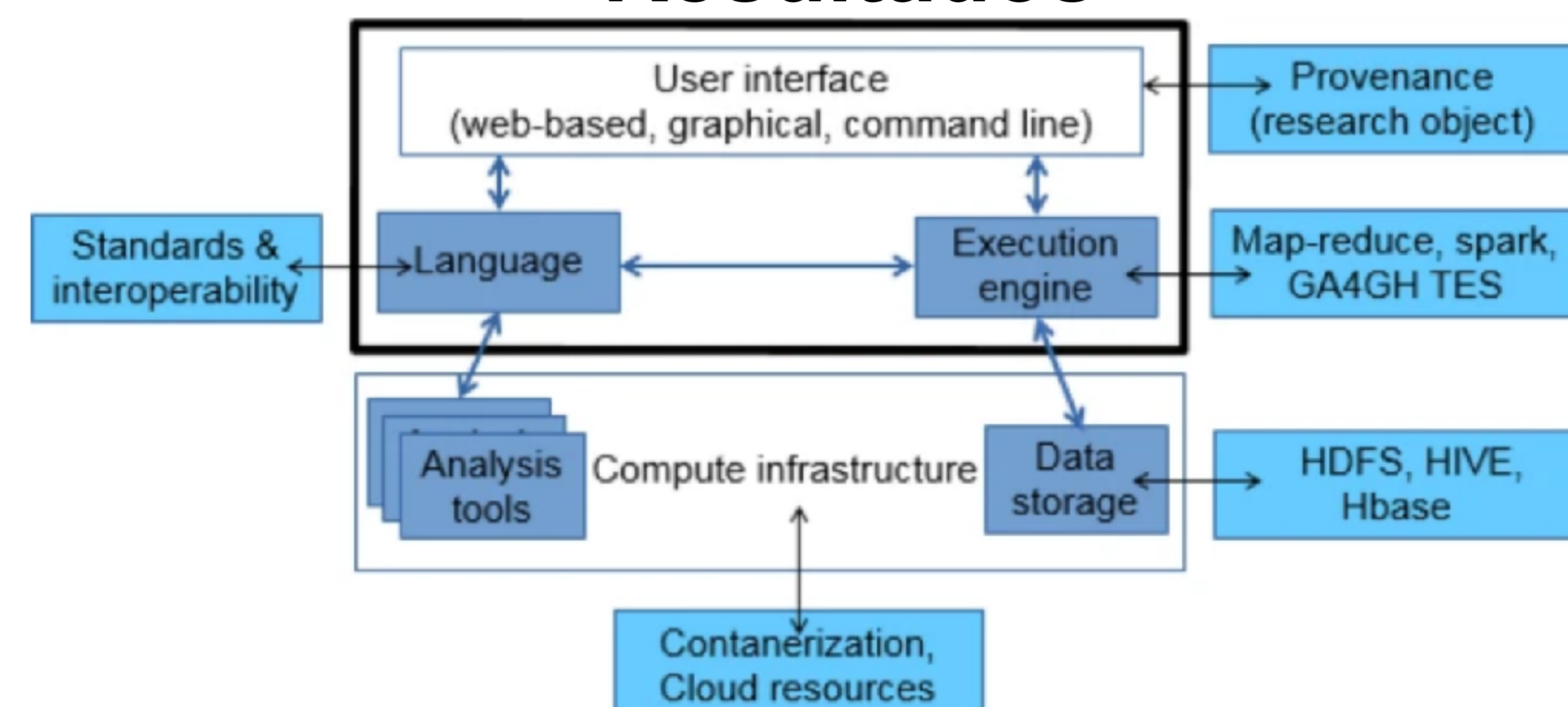


Figura1: Um WfMS é um middleware entre o analista e o ambiente computacional. Abrange as especificações de linguagem para interconectar os executáveis de análise e o mecanismo de execução para despachar tarefas e gerenciar dependências na infraestrutura de computação.

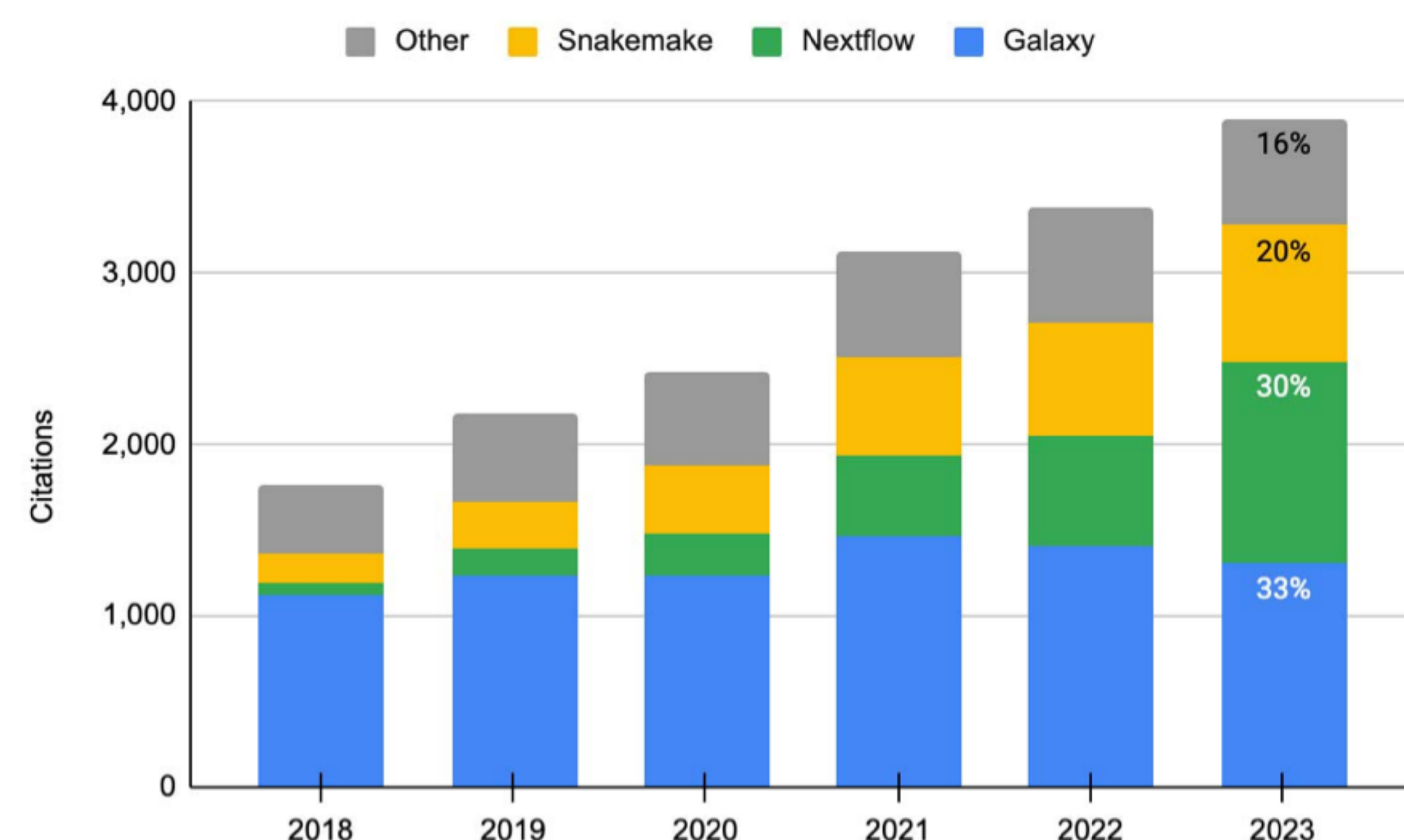


Figura2: Contagens de citações do Google Acadêmico para sistemas de gerenciamento de fluxo de trabalho de bioinformática. Soma de citações das principais publicações de Galaxy, Nextflow e Snakemake entre 2018 e 2023

Conclusões

Os orquestradores de fluxo de trabalho, como o Nextflow, são ferramentas indispensáveis na análise de dados de saúde, promovendo eficiência, reprodutibilidade e escalabilidade. Esta revisão reforça seu papel transformador na medicina personalizada, contribuindo para a análise de grandes volumes de dados clínicos de forma robusta, eficiente e reprodutível.

Bibliografia

- Ahmed, A.E., Allen, J.M., Bhat, T. et al. Design considerations for workflow management systems use in production genomics research and the clinic. Sci Rep 11, 21680 (2021). <https://doi.org/10.1038/s41598-021-99288-8>
- Langer, Bjorn E., et al. "Empowering bioinformatics communities with Nextflow and nf-core." bioRxiv (2024): 2024-05.