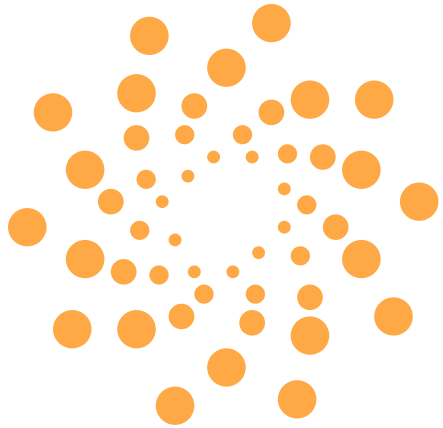
# *workflow-installer*: Simplifying Nextflow Workflow Deployment on HPC Clusters

Kisaru Liyanage

Wenjing Xue

Matthew Downton



NCI AUSTRALIA

# Nextflow Workflows on HPCs

- Nextflow is one of the best WMS for bioinformatics workflows[1]
- HPC clusters are powerful but can be challenging for workflow deployment[2]
- Researchers often face barriers setting up workflows on HPCs
  - Software or container (e.g., Singularity) dependencies
  - Restricted internet access
  - Fine-tuning configurations (e.g., resource allocation)


[1] V. Spišaková et al. 'Nextflow in Bioinformatics: Executors Performance Comparison Using Genomics Data', *Future Generation Computer Systems* 142, 328–339, 2023.
[2] M. Djaffardjy et al., 'Developing and reusing bioinformatics data analysis pipelines using scientific workflow systems', *Computational and Structural Biotechnology Journal*, vol. 21, 2075–2085, 2023.

# Simplifying Workflow Deployment on HPCs

- We introduce *workflow-installer,* a lightweight framework for centrally deploying versioned Nextflow pipelines as Environment Modules
- Users can load a workflow with `module load <workflow>/<version>`
- Features & Benefits:
  - Pre-fetched dependencies → offline execution
  - Fine-tuned configurations → improved performance
  - Demo data → quick installation testing
  - Simplified Seqera integration
  - Documentation → easy onboarding

ENVIRONMENT MODULES

# How *workflow-installer* Works: Before Installation

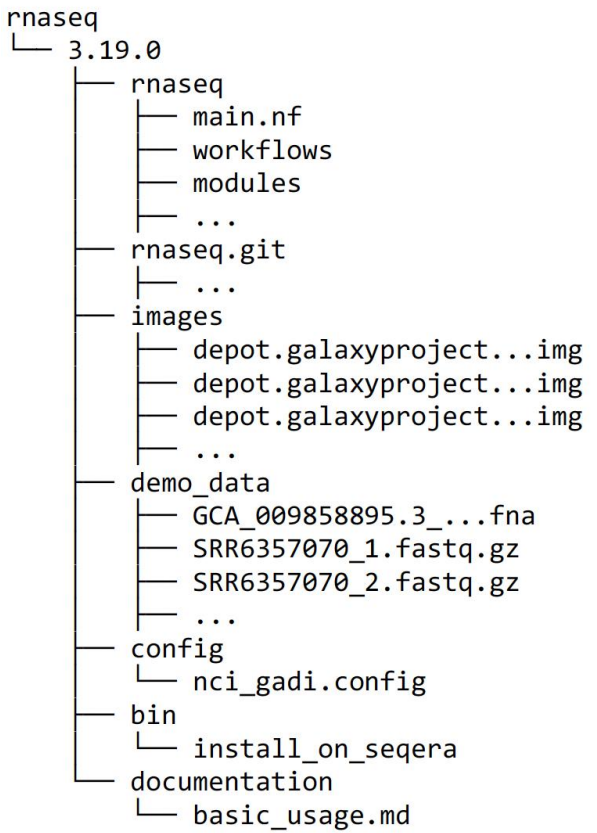
- Files used for an installation:
  - `install.sh` : The installer script
  - `container.list` : File containing Singularity container URIs
  - `nci_gadi.config` : Nextflow configuration fine-tuned for the cluster
  - `install_on_seqera` : Script for registering the workflow on Seqera
  - `basic_usage.md` : Documentation file

# How *workflow-installer* Works: The Installer Script

- An installer script is created for each workflow/version which automates all deployment tasks
- Key steps:
    1. Pulling versioned workflow code (e.g., from GitHub)
    2. Applying patches for offline execution
    3. Creating a bare Git repository (for Seqera)
    4. Retrieving Singularity container images
    5. Downloading demo data
    6. Copying script for Seqera integration, cluster-specific configurations and documentation
    7. Generating the final modulefile
- A separate helper script provides functions for common tasks such as caching, pulling images, and transforming container URIs.

# How *workflow-installer* Works: After Installation

- A directory named `<workflow>/<version>` is created



# How *workflow-installer* Works: After Installation

- A modulefile named `<workflow>/<version>` is created

```
#%Module1.0
setenv WF_HOME /path/to/workflows/rnaseq/3.19.0
setenv NXF_SINGULARITY_CACHEDIR /path/to/workflows/rnaseq/3.19.0/images
setenv NXF_OFFLINE true
prepend-path PATH /path/to/workflows/rnaseq/3.19.0/bin
```

# Usage Example

- An installed workflow can be loaded and launched as below:

```
module use /path/to/workflows/modulefiles
module load rnaseq/3.19.0
module load nextflow

nextflow -c ${WF_HOME}/config/nci_gadi.config run ${WF_HOME}/rnaseq ...
```

- It can be registered on the Seqera platform as below:

```
module use /path/to/workflows/modulefiles
module load rnaseq/3.19.0

install_on_seqera [-t API_ACCESS_TOKEN] [-c COMPUTE_ENV_ID] ...
```

# Conclusions

- *workflow-installer* simplifies workflow deployment on HPC clusters
- Enables fully offline execution with pre-fetched dependencies
- Delivers cluster-optimised performance
- Enhances reproducibility and accessibility of Nextflow workflows

➢ **Reduces user setup time to virtually zero, allowing researchers to focus on science rather than infrastructure!**

# Future Work

- Developing a CI/CD pipeline to automatically trigger workflow installations when a workflow is added or updated (in progress by Wenjing Xue)
- Make the framework open source
- Enable deployment on other HPC facilities for broader adoption
- Develop a catalogue for the workflows installed at NCI

# Acknowledgments


- The work is based on a framework originally developed by Dale Roberts (NCI/ANU)
- This work is supported by the Australian BioCommons as part of the Workflow Commons and GUARDIANS projects

# Q & A

# Thank You!