Workflow management with GNU Guix

Roel Janssen

February 5, 2017
Data analysis
Processes

Initial dataset

Undeniable proof

grep -v "LowQuality" \
DataSet.txt > \
FilteredDataSet.txt

Rscript sig_test.R  \
FilteredDataSet.txt
The workflow

Workflow management with GNU Guix
Roel Janssen
Parallel steps

Initial dataset → A → Undeniable proof
G → F → B → C → N
H → E → D → K → M
I → J → L

Workflow management with GNU Guix
Roel Janssen
Executing on a computer cluster

Workflow management with GNU Guix
Roel Janssen
Process record type

(define quality-control
  (process
    (name "quality-control")
    (version "1.0")
    (package-inputs `(('"fastqc" ,fastqc-0.11.4)))
    (data-inputs `("/path/to/sample_R1_001.fastq.gz"))
    (output-path "/path/to/rnaseq-out")
    (run-time (complexity (space (* 1024 1024))
                        (time 60)))
    (procedure #~(begin (map (lambda (sample)
                             (system* (string-append
                                        #$@(assoc-ref package-inputs "fastqc")
                                        "/bin/fastqc") sample "-o" #$output-path))
                        #$data-inputs)))
    (synopsis "Report read quality of FastQ files")
    (description "...")))

Workflow management with GNU Guix Roel Janssen
Workflow record type

(define rnaseq-workflow
 (workflow
   (name "rnaseq-workflow")
   (version "1.0")
   (input "/path/to/input/files")
   (output "/path/to/output")
   (processes
    (quality-control
     align-reads
     add-read-groups
     sort-and-index))
   (restrictions
    ((add-read-groups align-reads)
     (sort-and-index add-read-groups)))
   (synopsis "RNA sequencing workflow")
   (description "...")))
How it works

#<process> #<derivation>

- programs
- job script
Get started

- Tutorial: https://gwl.roelj.com
- Source code: https://git.roelj.com/guix/guix-wl
- Questions: roel@gnu.org
Acknowledgements

- Pjotr Prins
- Joep de Ligt
- Ricardo Wurmus
- Ludovic Courtès
- GNU Guix contributors
- GNU Guile contributors