

PORTABLE CONTAINERS ORCHESTRATION AT SCALE WITH NEXTFLOW

Paolo Di Tommaso, Seqera Labs
ISC-HPC 2019 - Frankfurt



ENABLING TECHNOLOGY

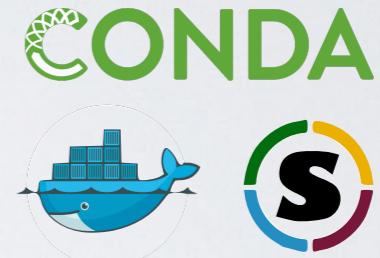
nextflow



code



orchestration



dependencies



deployment



sharing & reproducibility

WHAT DO YOU MEAN?

```
# satellite sequences reported by RepeatMasker.  
zcat rmsk.txt.gz \  
| grep Satellite \  
| cut -f6,7,8 \  
| sed s,^chr,, \  
| perl -pe 's/^[\^s_]+([^\s_]+)_random/$1.1/' \  
| tr "gl" "GL" \  
| sort -k1,1N -k2,2n \  
| bgzip > hs37d5.satellite.bed.gz
```

THE NEXTFLOW WAY

```
process filtering {
    input: file 'rmsk.txt.gz' from sequences_ch
    output: file 'hs37d5.satellite.bed.gz' into results_ch

    ...
    # satellite sequences reported by RepeatMasker.
    zcat rmsk.txt.gz \
        | grep Satellite \
        | cut -f6,7,8 \
        | sed s,^chr,, \
        | perl -pe 's/^[\s]+([^\s]+)_random/$1.1/' \
        | tr "gl" "GL" \
        | sort -k1,1N -k2,2n \
        | bgzip > hs37d5.satellite.bed.gz
    ...
}

Channel.fromPath('data/rmsk.txt.gz') | filtering | publishTo { '/path' }
```

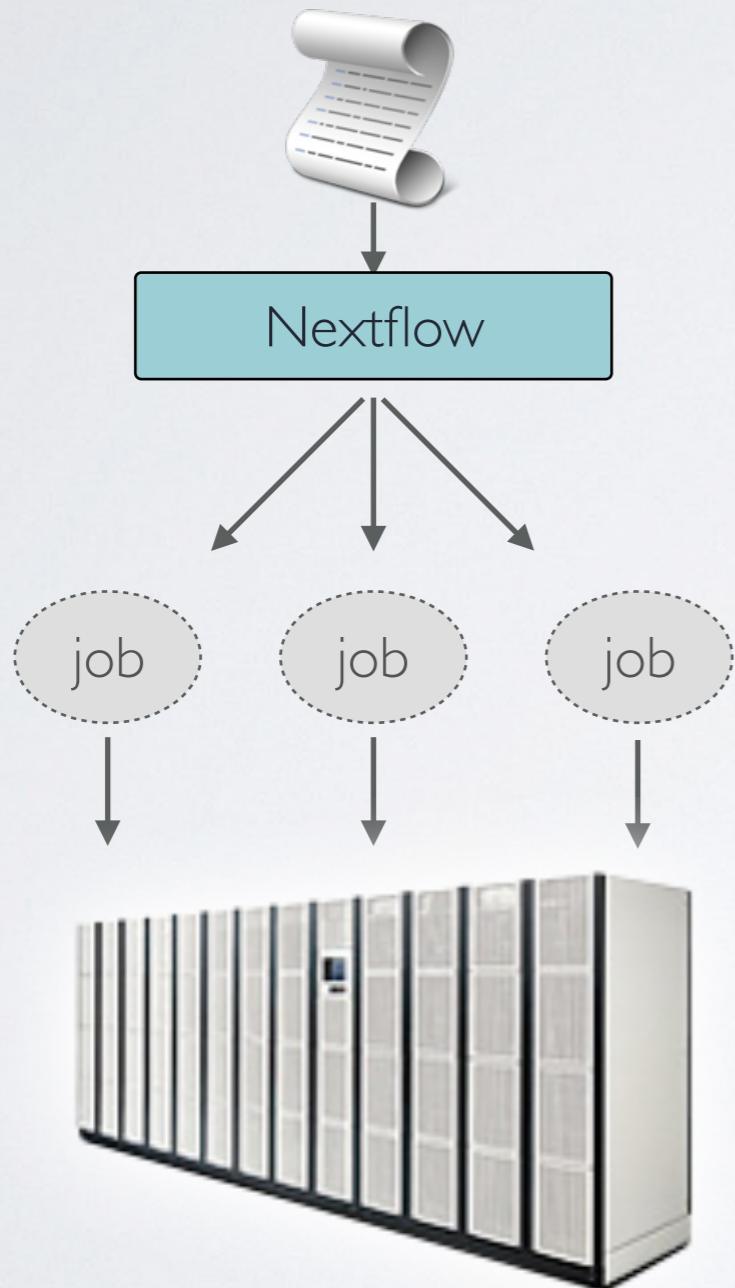
THE NEXTFLOW WAY

```
process filtering {
    input: file 'rmsk.txt.gz' from sequences_ch
    output: file 'hs37d5.satellite.bed.gz' into results_ch

    ...
    # satellite sequences reported by RepeatMasker.
    zcat rmsk.txt.gz \
        | grep Satellite \
        | cut -f6,7,8 \
        | sed s,^chr,, \
        | perl -pe 's/^[\s]+([^\s]+)_random/$1.1/' \
        | tr "gl" "GL" \
        | sort -k1,1N -k2,2n \
        | bgzip > hs37d5.satellite.bed.gz
    ...
}

Channel.fromPath('data/*.txt.fq') | filtering | publishTo { '/path' }
```

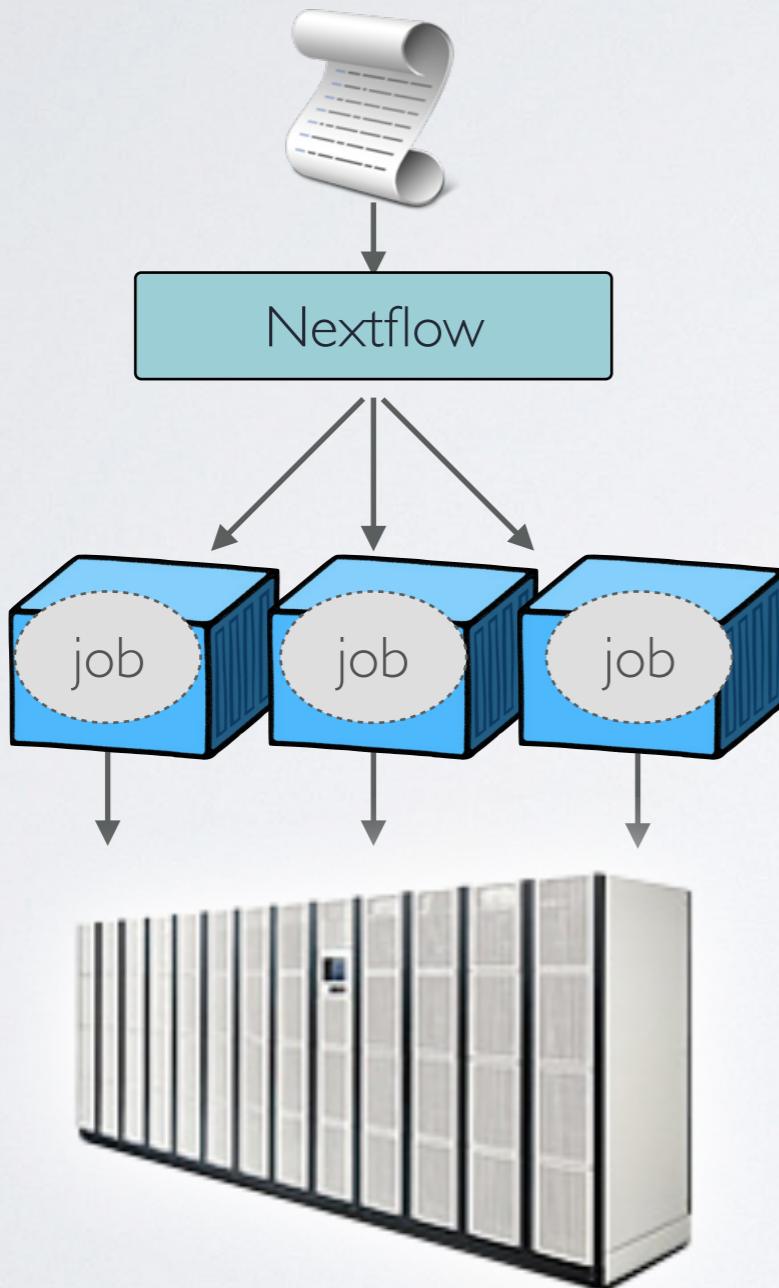
CONTAINERISATION



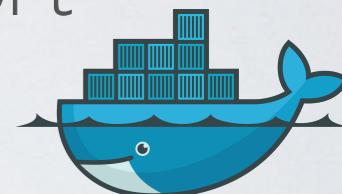
- Nextflow envisioned the use of software containers to fix computational reproducibility
- Mar 2014 (ver 0.7), support for Docker
- Dec 2016 (ver 0.23), support for Singularity



CONTAINERISATION



- Nextflow envisioned the use of software containers to fix computational reproducibility
- Mar 2014 (ver 0.7), support for Docker
- Dec 2016 (ver 0.23), support for Singularity

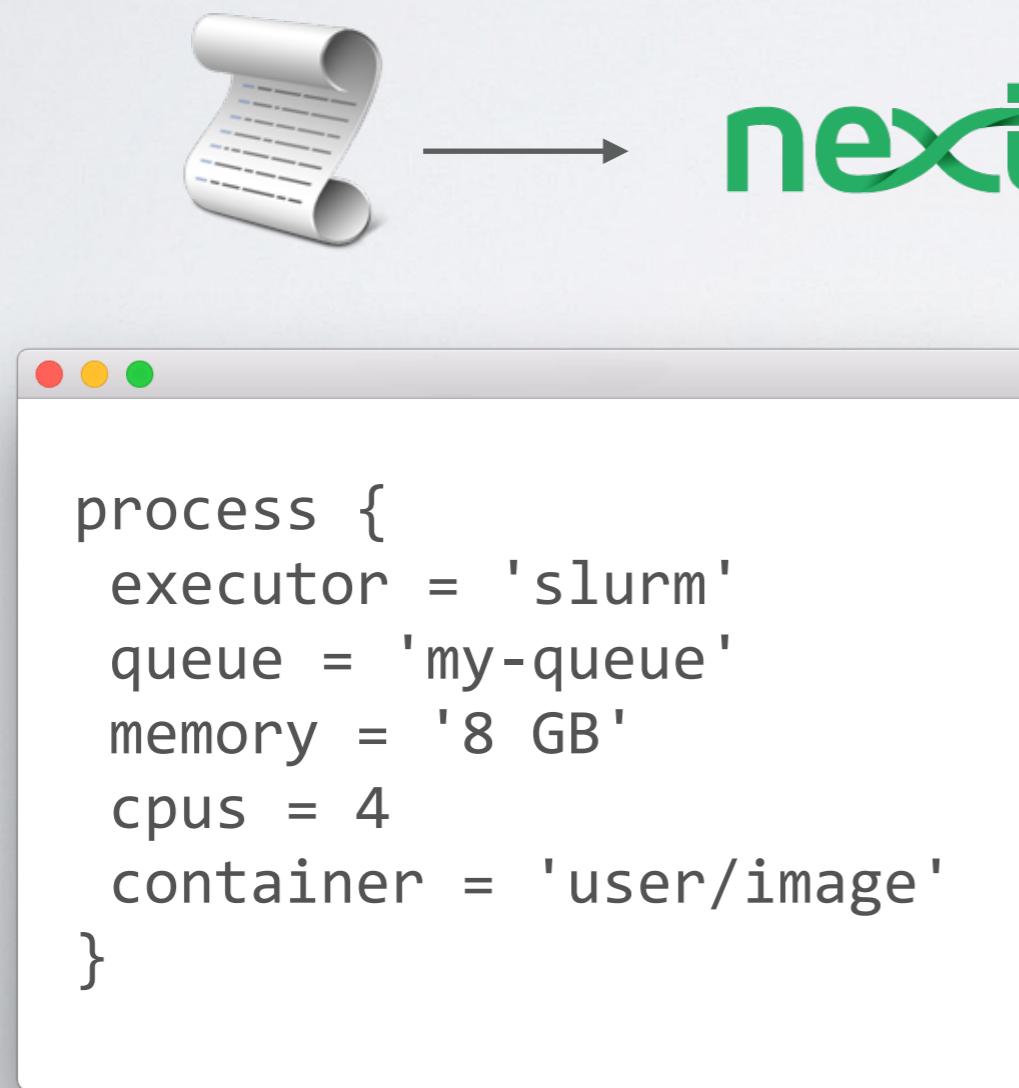


PORTABILITY

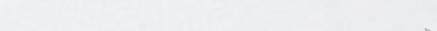


```
nextflow run your-script.nf -with-docker your/image
```

PORTABILITY



nextflow



slurm
workload manager

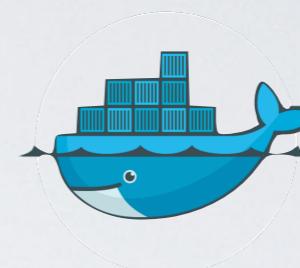


PORTABILITY



nextflow

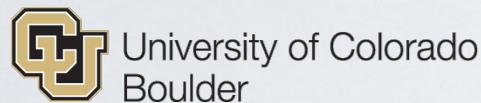
```
process {  
    executor = 'awsbatch'  
    queue = 'my-queue'  
    memory = '8 GB'  
    cpus = 4  
    container = 'user/image'  
}
```



slurm
workload manager



WHO IS USING NEXTFLOW?



Weill Cornell Medical College



Agency for
Science, Technology
and Research
SINGAPORE

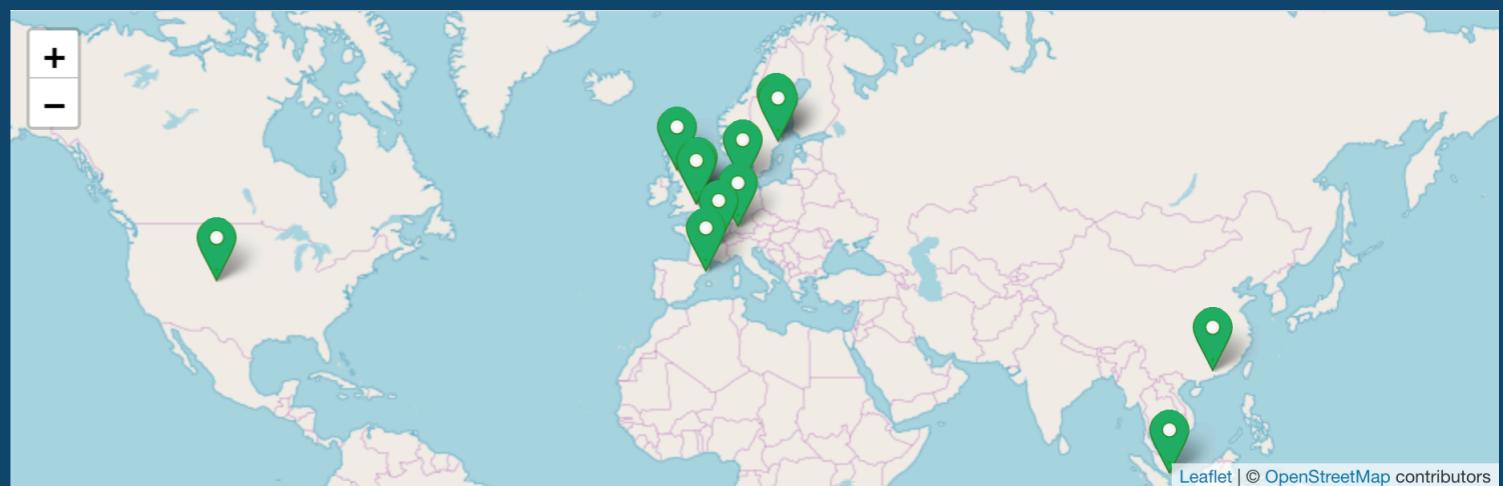
SYNTHETIC GENOMICS



nf-core



20 pipelines 38 members 12+ institutions



CANCER
RESEARCH
UK

BEATSON
INSTITUTE



NATIONAL
GENOMICS
INFRASTRUCTURE



Genome Institute
of Singapore



SLU



中山大學
腫瘤防治中心
SUN YAT-SEN UNIVERSITY CANCER CENTER

International Agency for Research on Cancer



University
of Colorado
Boulder



THANK YOU

nextflow

<http://nextflow.io>



seqeralabs

<http://seqera.io>