

# How to efficiently use the cluster computing power

Dominik Stroh

25.5.2021

## Introduction

In contrast to small computing tasks like downloading packages and inspection of datasets via head/tail, computing intensive operations like STAR, htseq-count and majiq should not be performed on a single workstation but on the whole cluster. For these tasks, the cluster uses the Slurm workload manager. In principle, you create a file where you tell the cluster what to do (in Bash) and submit this to the cluster as a job. It then computes the commands without your supervision and tells you when it is finished.

## Jobs

Jobs are a way to manage many computing requests and optimally utilise the computing power.

## Creating a job

Create a file named job.sh

```
cd [DESIRED DIRECTORY]
touch job.sh
```

The file should contain the following header before your code:

```
#!/bin/bash
#SBATCH --partition=all
#SBATCH --cpus-per-task=4
#SBATCH --mem-per-cpu=16gb
#SBATCH --job-name=[NAME OF YOUR JOB]
#SBATCH --array=1
#SBATCH --output=[NAME OF YOUR JOB].out
#SBATCH --error=[NAME OF YOUR JOB].err
#SBATCH --mail-user=[YOUR EMAIL ADDRESS]
#SBATCH --mail-type=ALL

#Your code should start after this header:
[YOUR CODE]
```

## Submitting a job

In order to submit a job to the cluster use the following command:

```
cd [DIRECTORY OF THE JOB FILE]
sbatch job.sh
```

## Progression of a job

When you want to see whether your job has been accepted, which machine computes it or how the workload of the cluster is use the following command:

```
squeue
```

## Cancelling a job

When you realise, something went wrong with your code (which you should double-check beforehand!), you can cancel a running job with the command:

```
scancel [YOUR JOBNUMBER]
```