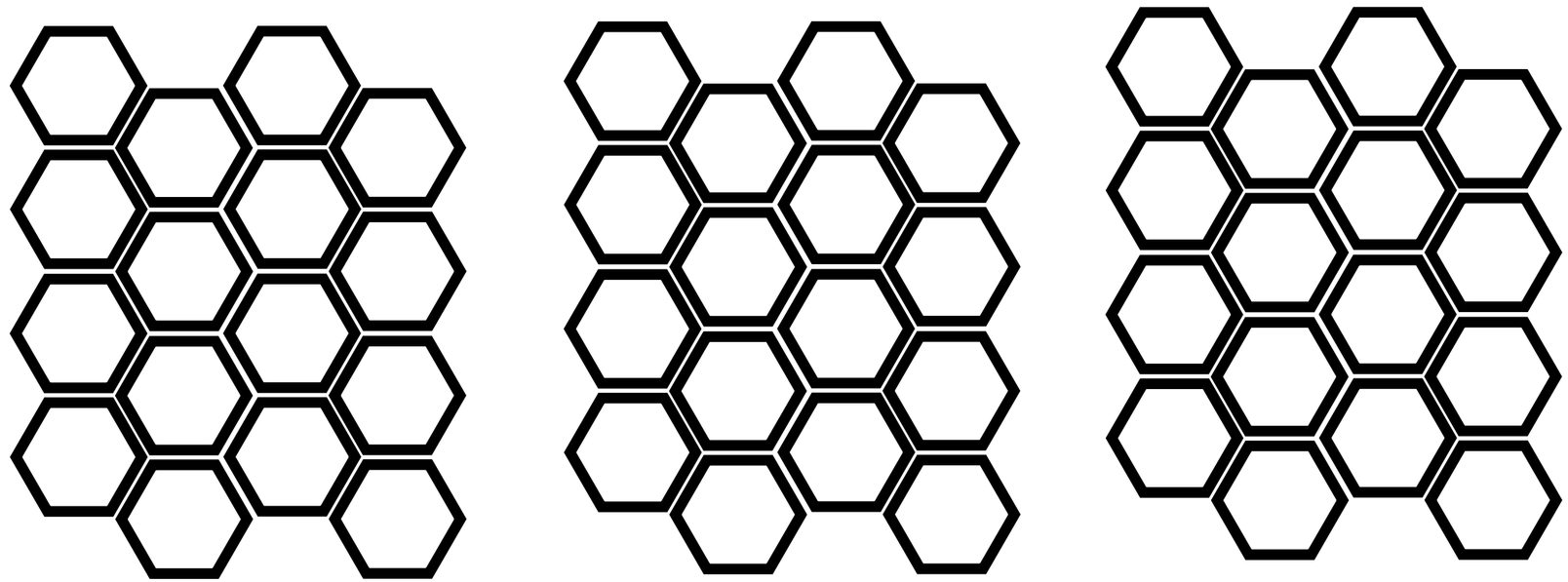
HyperQueue

Jakub Beránek, Ada Böhm

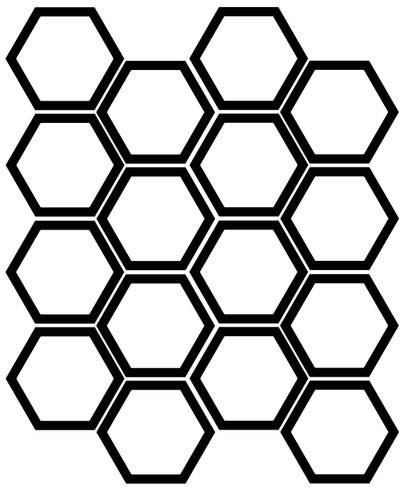
IT4Innovations

jakub.beranek@vsb.cz github.com/kobzol



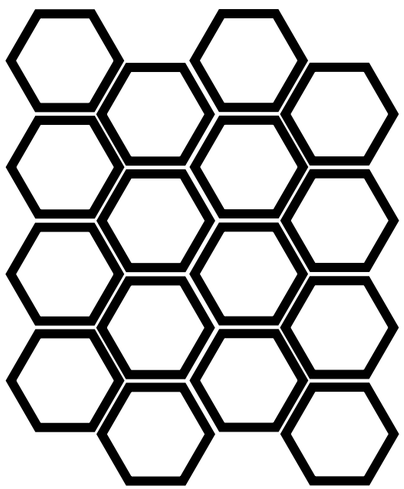
Login nodes

System scheduler



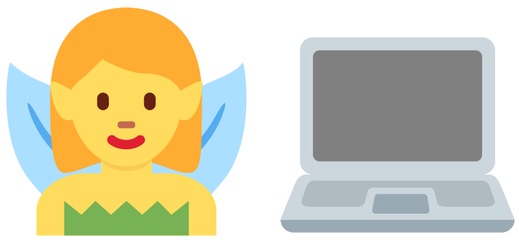
Compute nodes

Login nodes



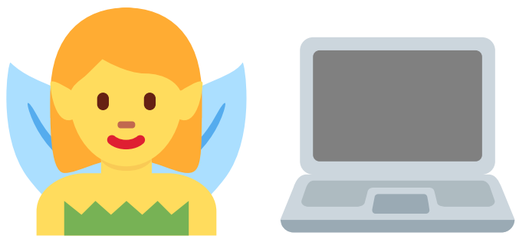
System scheduler

Compute nodes

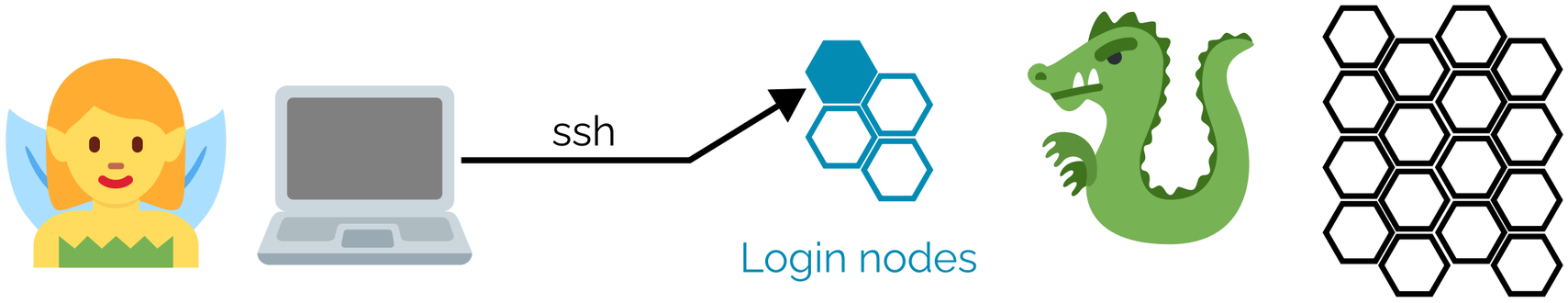



Login nodes
System scheduler
Compute nodes

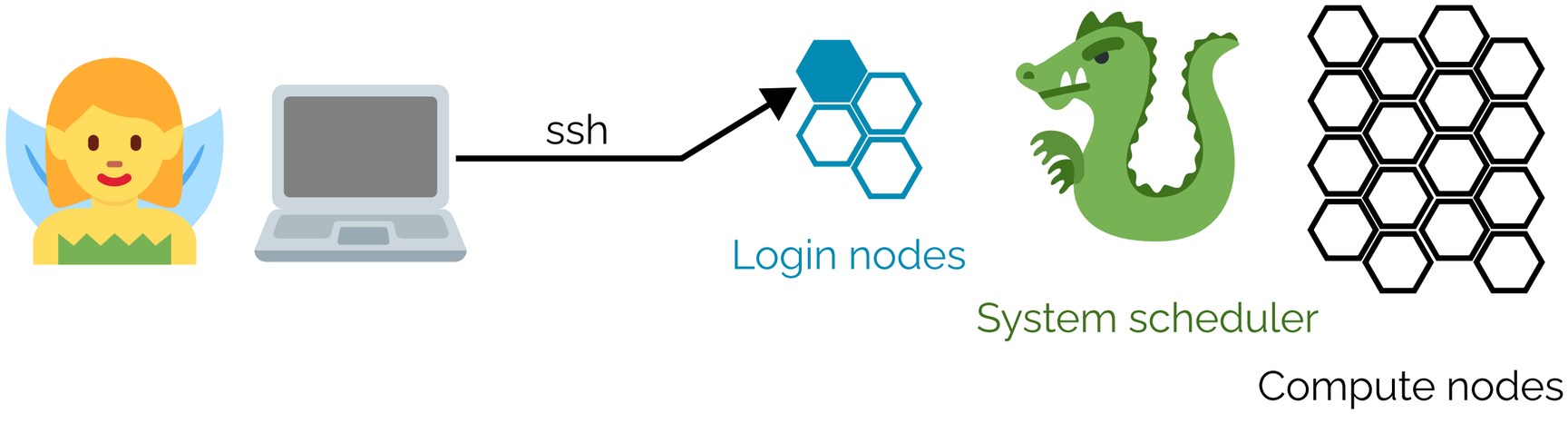
```
$ _
```



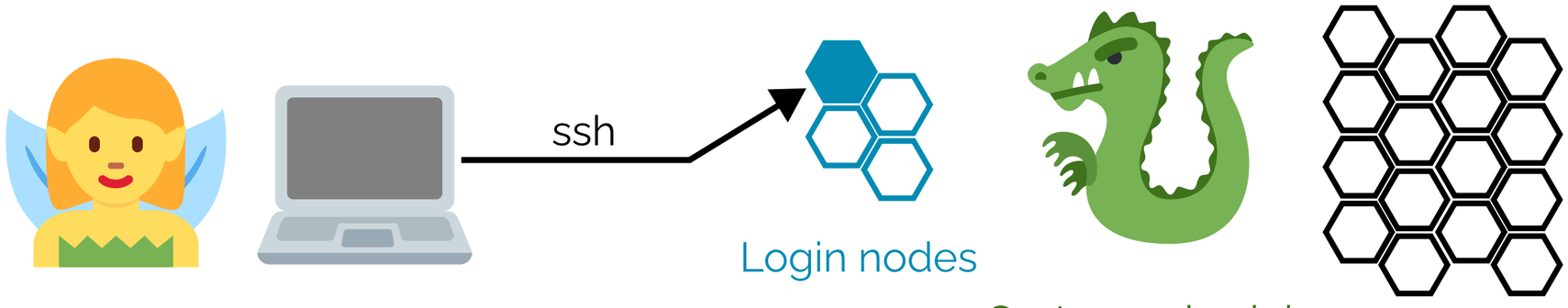

```
$ ssh karolina.it4i.cz
```



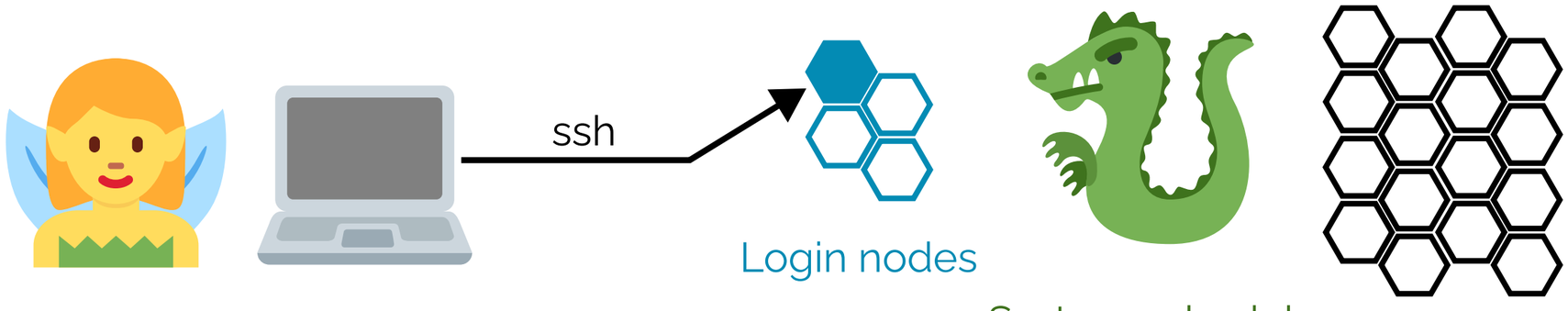

```
$ ssh karolina.it4i.cz
```



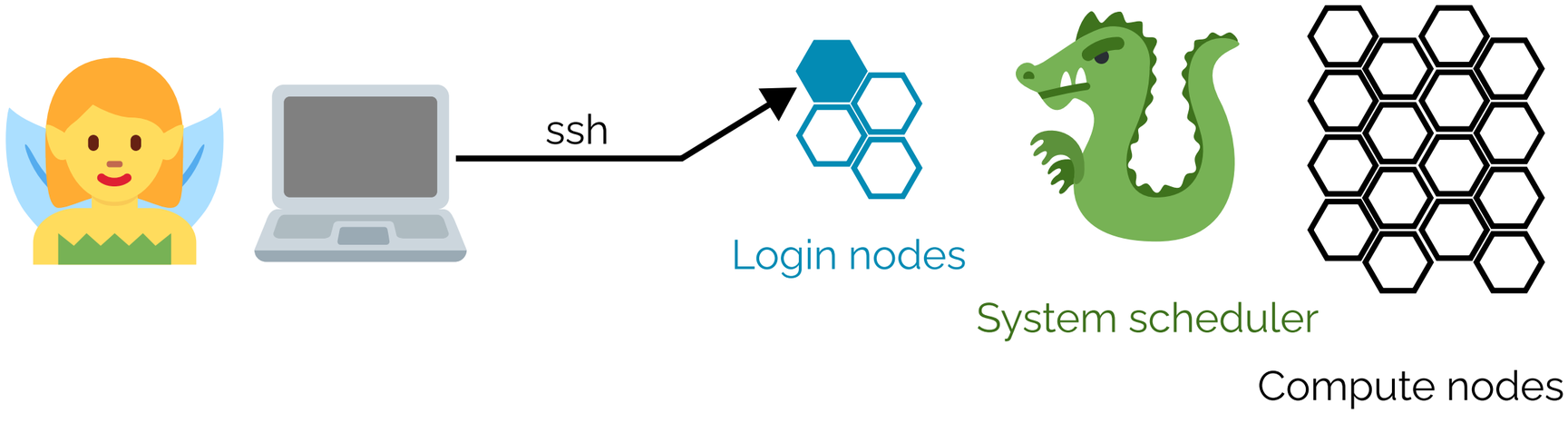

```
$ ssh karolina.it4i.cz


 _   __                  _ _
| | / /                 | (_)
| |/ / __ _ _ __ ___ ___| |_ _ __   __ _
|    < / _` | '__/ _ \ | | | '_ \ / _` |
| . \ (_| | | | (_) | | | | | | | (_| |
|_|\_\__,_|_|  \___/|_|_|_| |_|\__,_|

 ...running on Red Hat Enterprise Linux 7.x

login1.karolina$ _
```



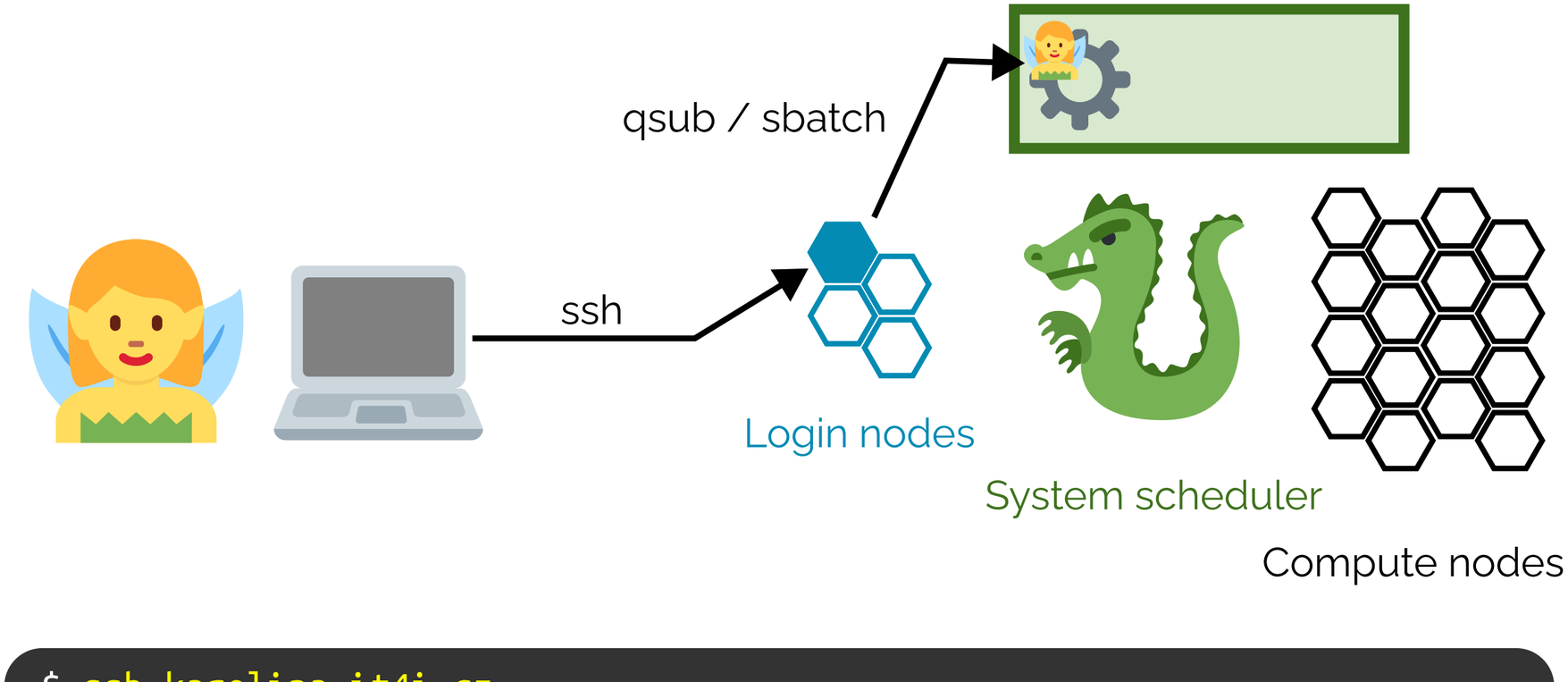

```
$ ssh karolina.it4i.cz


 _    __                  _ _
| |/ /                   | (_)
| ' / __ _ _ __ ___ ____ | |_ _ __   __ _
|  < / _` | '__/ _ \| | | | '_ \ / _` |
| . \ (_| | | | (_) | | | | | | | (_| |
|_|\_\__,_|_|  \___/|_|_|_| |_|\__,_|

 ...running on Red Hat Enterprise Linux 7.x

login1.karolina$ ./my-computation
```



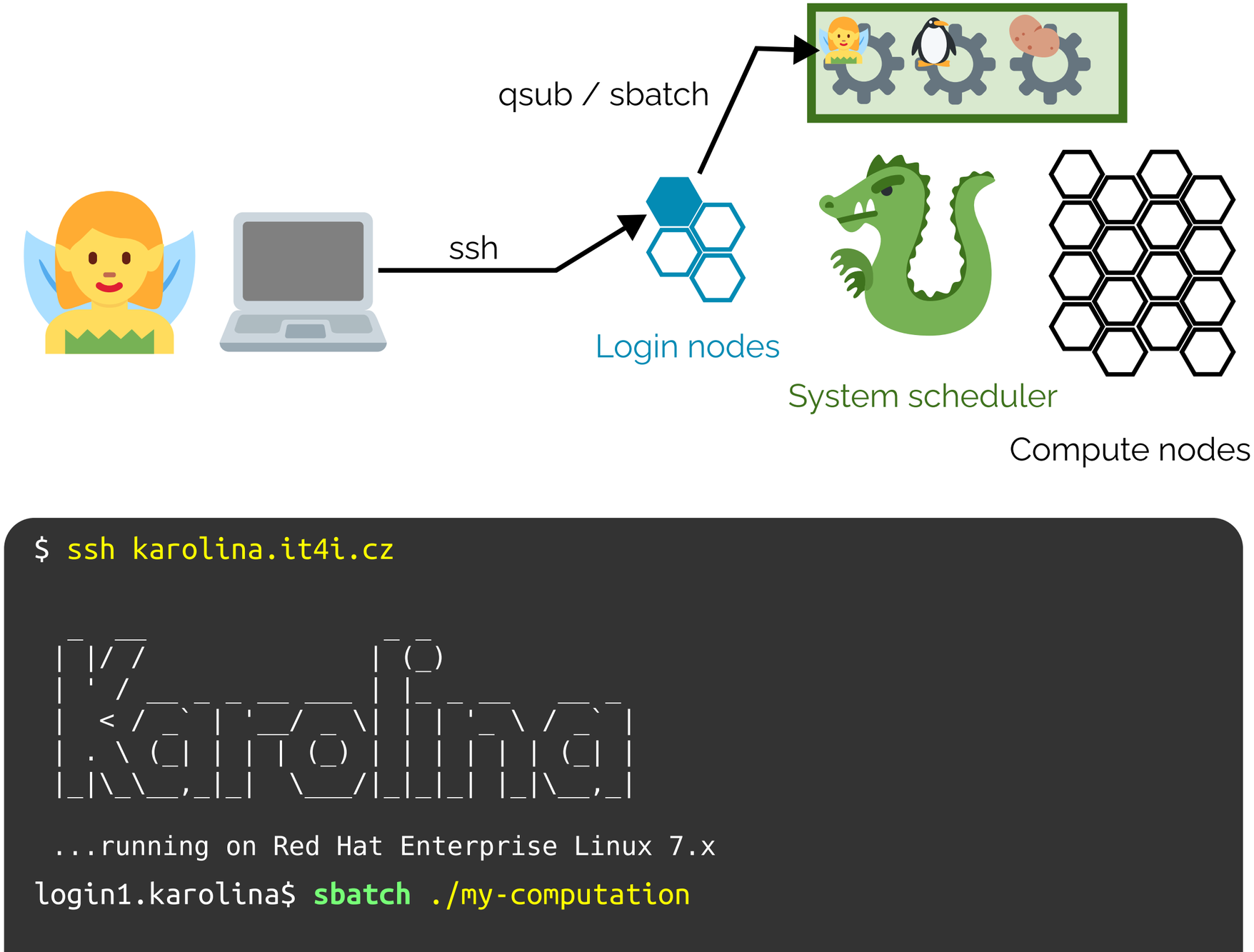

```
$ ssh karolina.it4i.cz


 _    __                  _ _
| |/ /                   | (_)
| ' / __ _ _ __ ___ | |_ _ __   __ _
|  < / _` | '__/ _ \| | | '_ \ / _` |
| . \ (_| | | | (_) | | | | | | (_| |
|_|\_\__,_|_|  \___/|_|_|_| |_|\__,_|

 ...running on Red Hat Enterprise Linux 7.x

login1.karolina$ ./my-computation
```



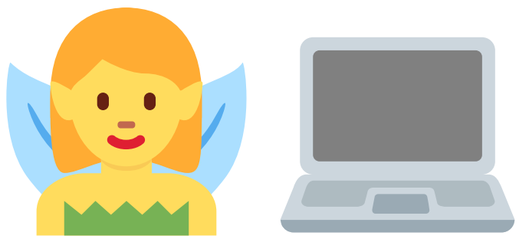

```
$ ssh karolina.it4i.cz


 _    __                  _ _
| |/ /                   | (_)
| ' / __ _ _ __ ___ ____ | |_ _ __   __ _
|  < / _` | '__/ _ \| | | | '_ \ / _` |
| . \ (_| | | | (_) | | | | | | | (_| |
|_|\_\__,_|_|  \___/|_|_|_| |_|\__,_|

 ...running on Red Hat Enterprise Linux 7.x
login1.karolina$ sbatch ./my-computation
```

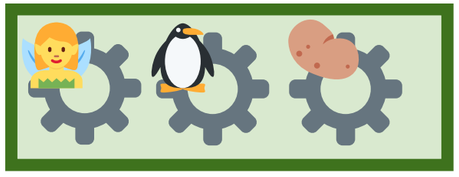



```
$ ssh karolina.it4i.cz


 _   __                  _ _
| |/ /                 | (_)
| ' / __ _ _ __ ___ | |_ _ __   __ _
|  < / _` | '__/ _ \| | | '_ \ / _` |
| . \ (_| | | | (_) | | | | | | (_| |
|_|\_\__,_|_|  \___/|_|_|_| |_|\__,_|

 ...running on Red Hat Enterprise Linux 7.x
login1.karolina$ sbatch ./my-computation
```

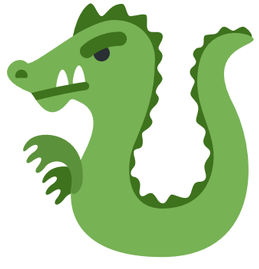



```
$ ssh karolina.it4i.cz


 _    __                  _ _
| |/ /                   | (_)
| ' / __ _ _ __ ___ ____ | |_ _ __   __ _
|  < / _` | '__/ _ \| | | | '_ \ / _` |
| . \ (_| | | | (_) | | | | | | | (_| |
|_|\_\__,_|_|  \___/|_|_|_| |_|\__,_|

 ...running on Red Hat Enterprise Linux 7.x
login1.karolina$ sbatch ./my-computation
```

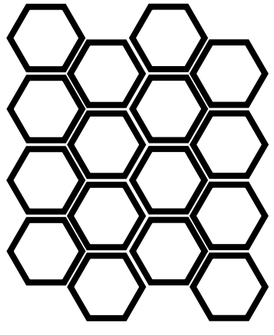




Login nodes

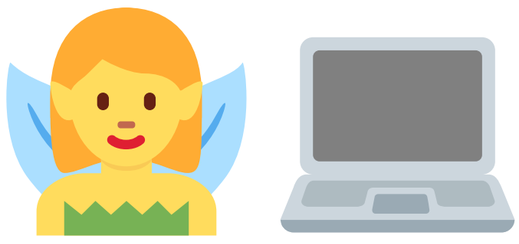


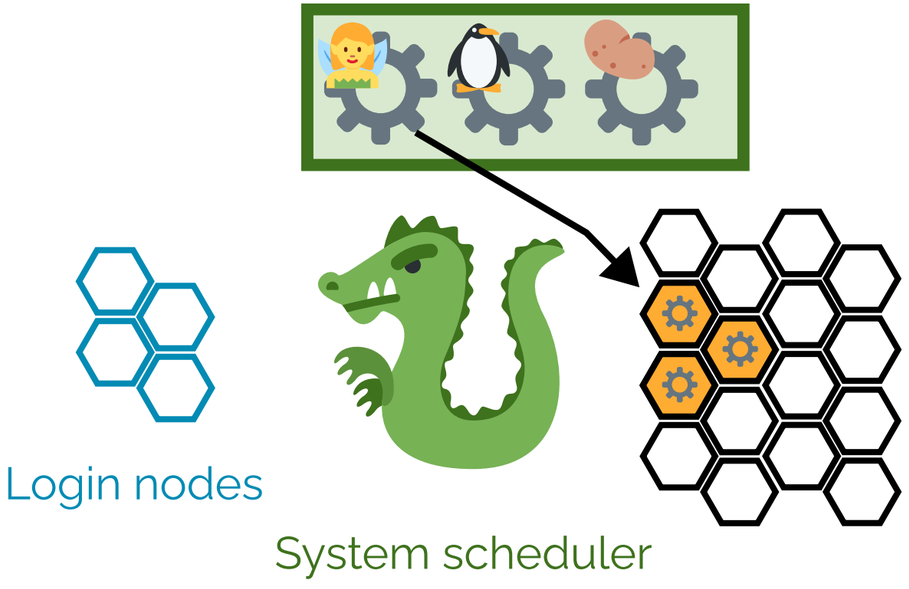
System scheduler



Compute nodes

```
$ ssh karolina.it4i.cz


 _    __                  _ _
| |/ /                   | (_)
| ' / __ _ _ __ ___ ___  | |_ _ __   __ _
|  < / _` | '__/ _ \| | | | '_ \ / _` |
| . \ (_| | | | (_) | | | | | | | (_| |
|_|\_\__,_|_|  \___/|_|_|_| |_|\__,_|

 ...running on Red Hat Enterprise Linux 7.x
login1.karolina$ sbatch ./my-computation
```

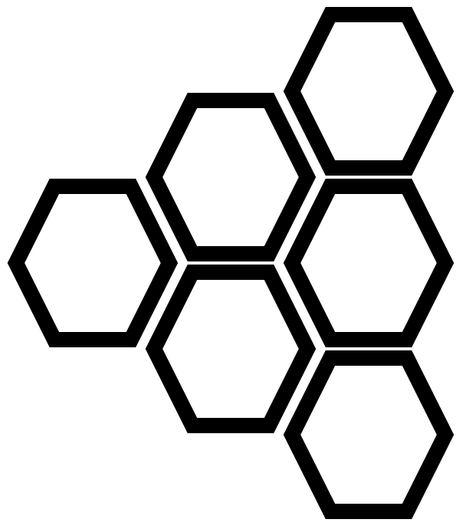




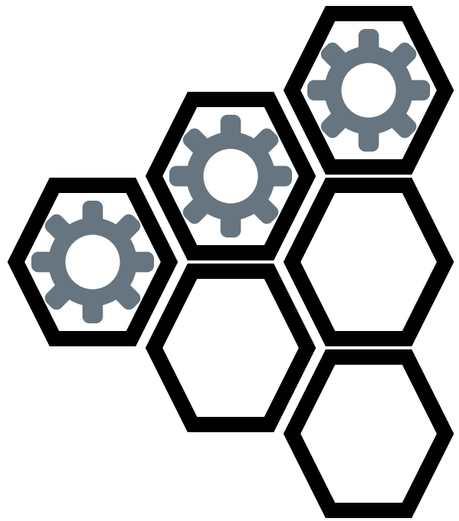

```
$ ssh karolina.it4i.cz


 _    __                  _ _
| |/ /                   | (_)
| ' / __ _ _ __ ___ ___  | |_ _ __   __ _
|  < / _` | '__/ _ \| | | | '_ \ / _` |
| . \ (_| | | | (_) | | | | | | | (_| |
|_|\_\__,_|_|  \___/|_|_|_| |_|\__,_|

 ...running on Red Hat Enterprise Linux 7.x
login1.karolina$ sbatch ./my-computation
```

What's the problem?

A few large tasks

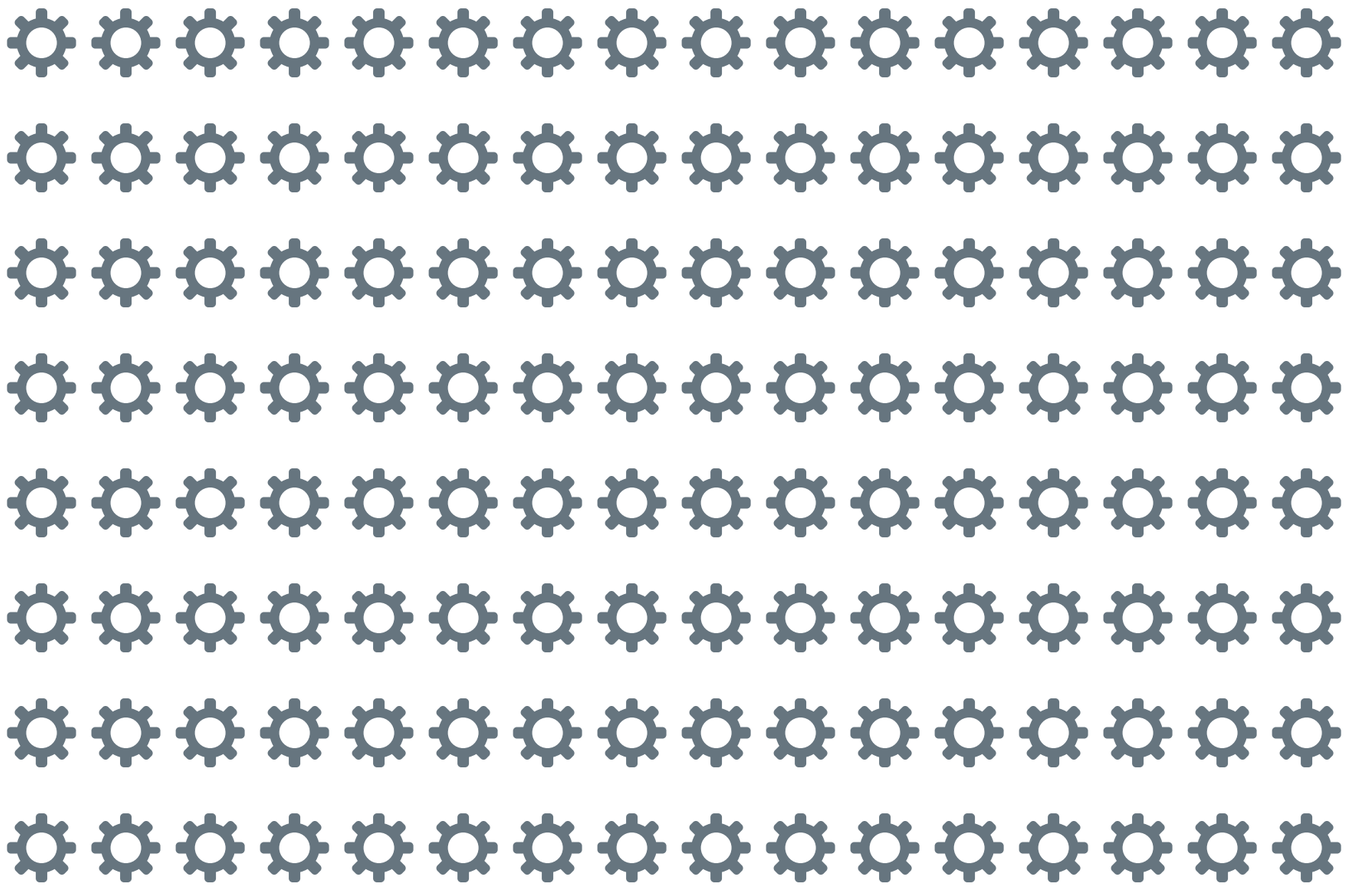


A few large tasks



Many tasks

Many tasks

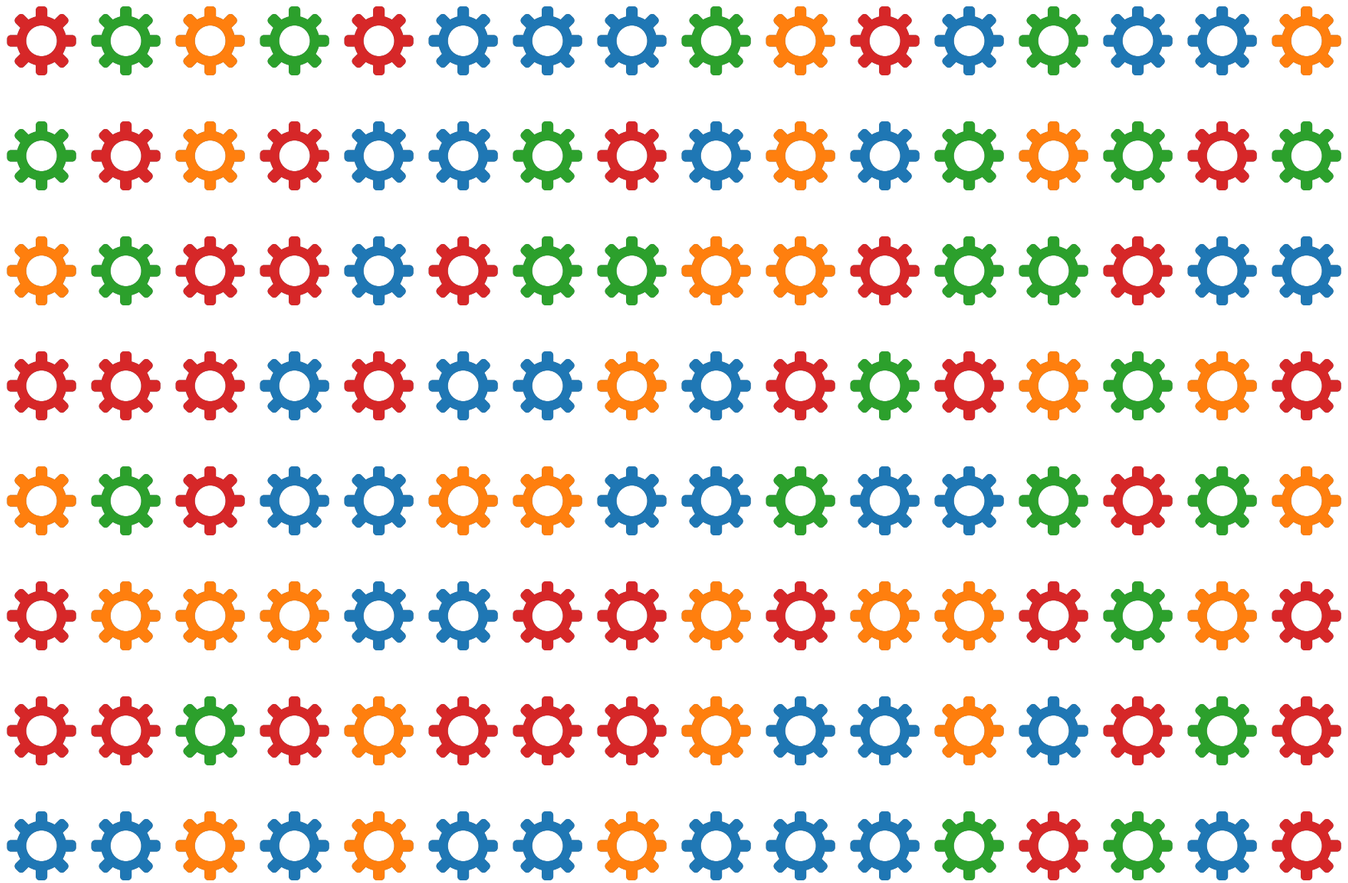


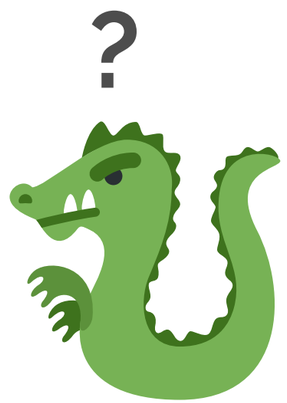
```
sbatch script.sh
sbatch script.sh
sbatch script.sh
sbatch script.sh
sbatch script.sh
sbatch script.sh
sbatch script.sh
sbatch script.sh
sbatch script.sh
sbatch script.sh
sbatch script.sh
sbatch script.sh
...
```

Many tasks

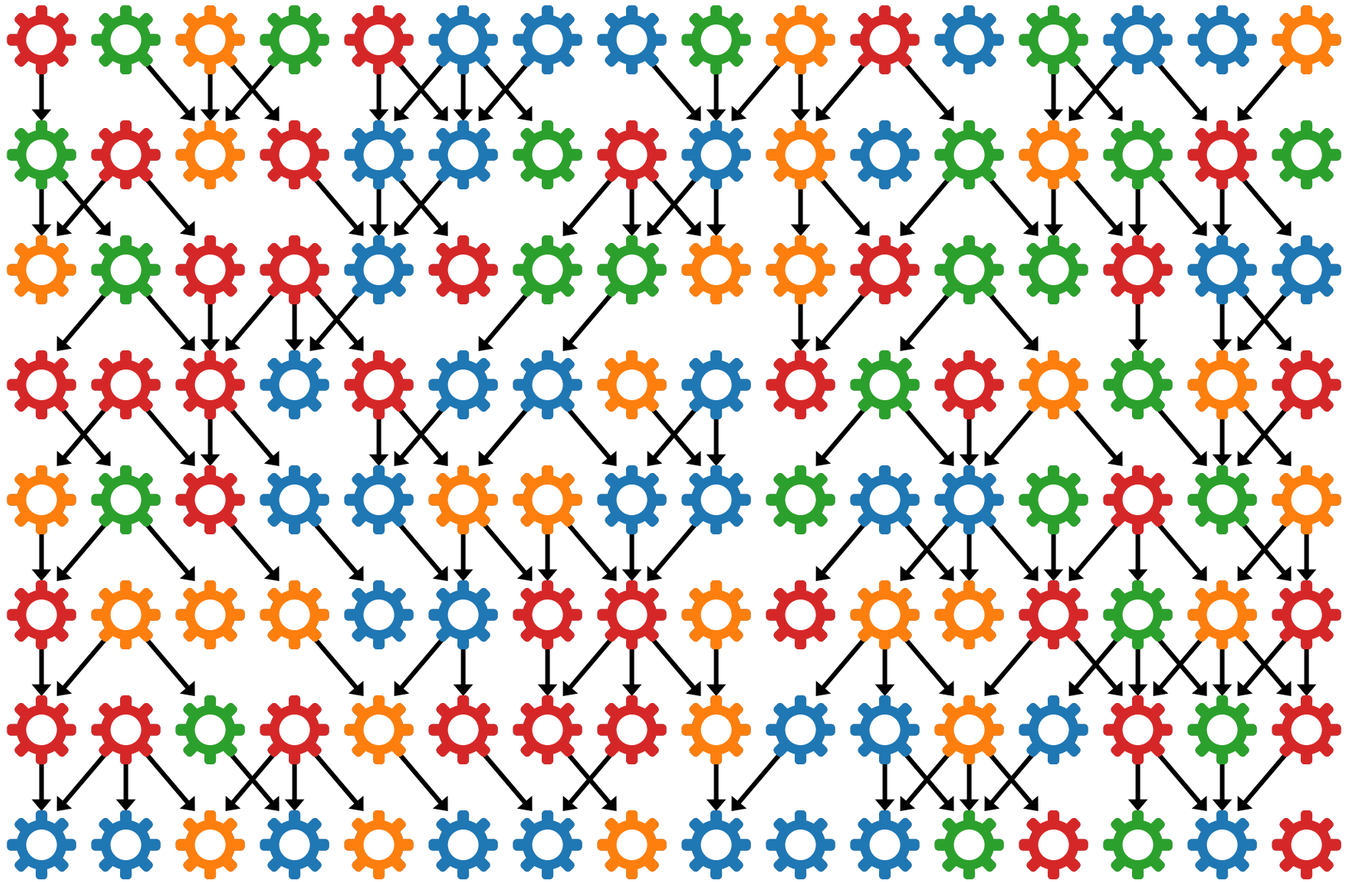
sbatch script.sh
sbatch script.sh
sbatch script.sh
sbatch script.sh
sbatch scri
sbatch pt.sh
sbatc sh
sbatc scrip
sbatc script.sh
sbatch ript.sh
sbatch scr
sbatch script.sh
sbatch script.sh
sbatch script.sh
sbatch script.sh
...

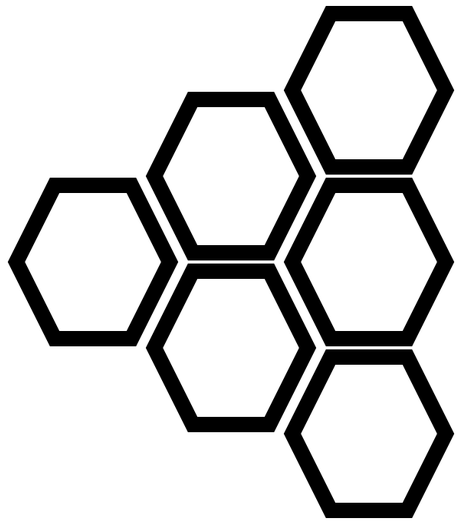
Heterogeneous tasks

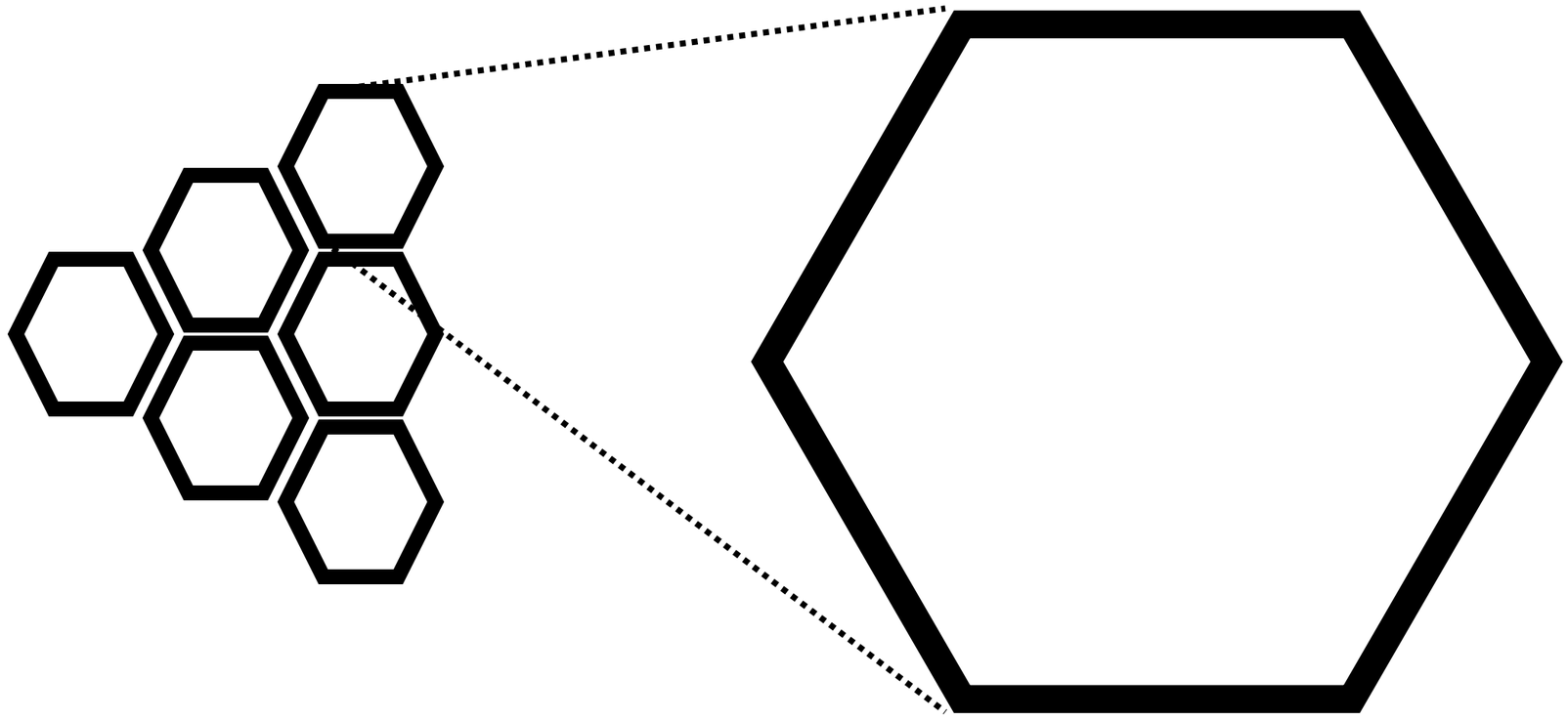


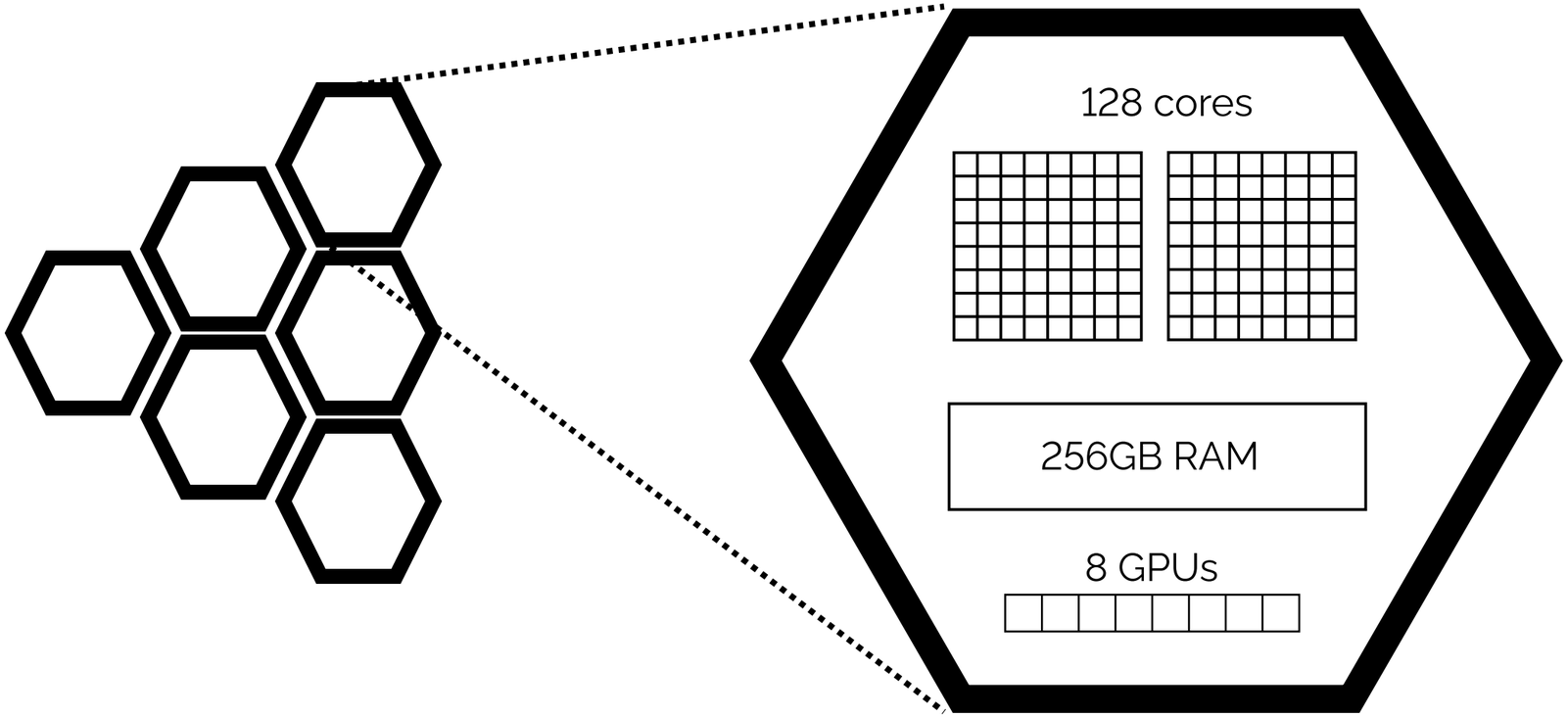


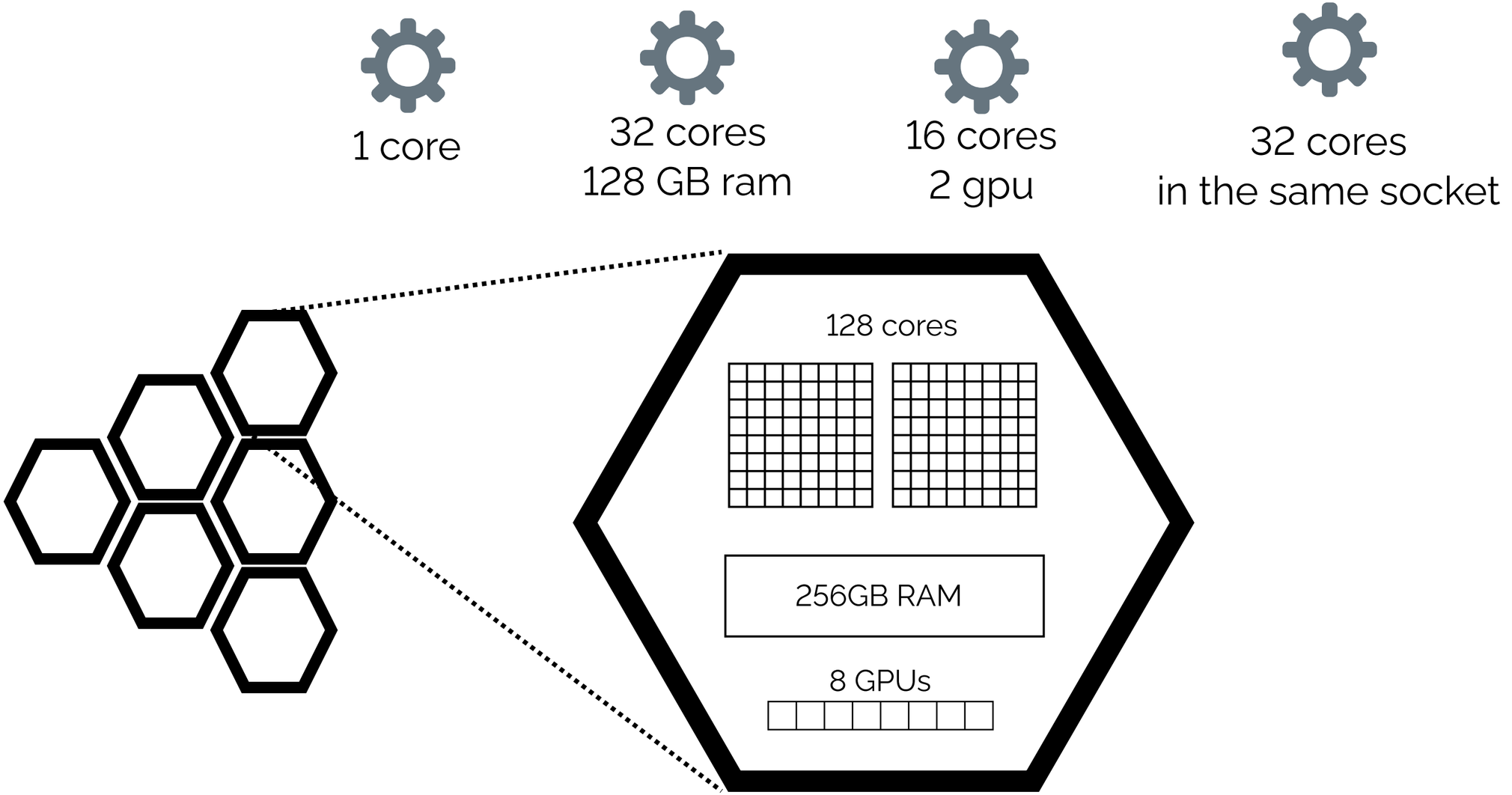
Task dependencies








128 cores
256GB RAM
8 GPUs

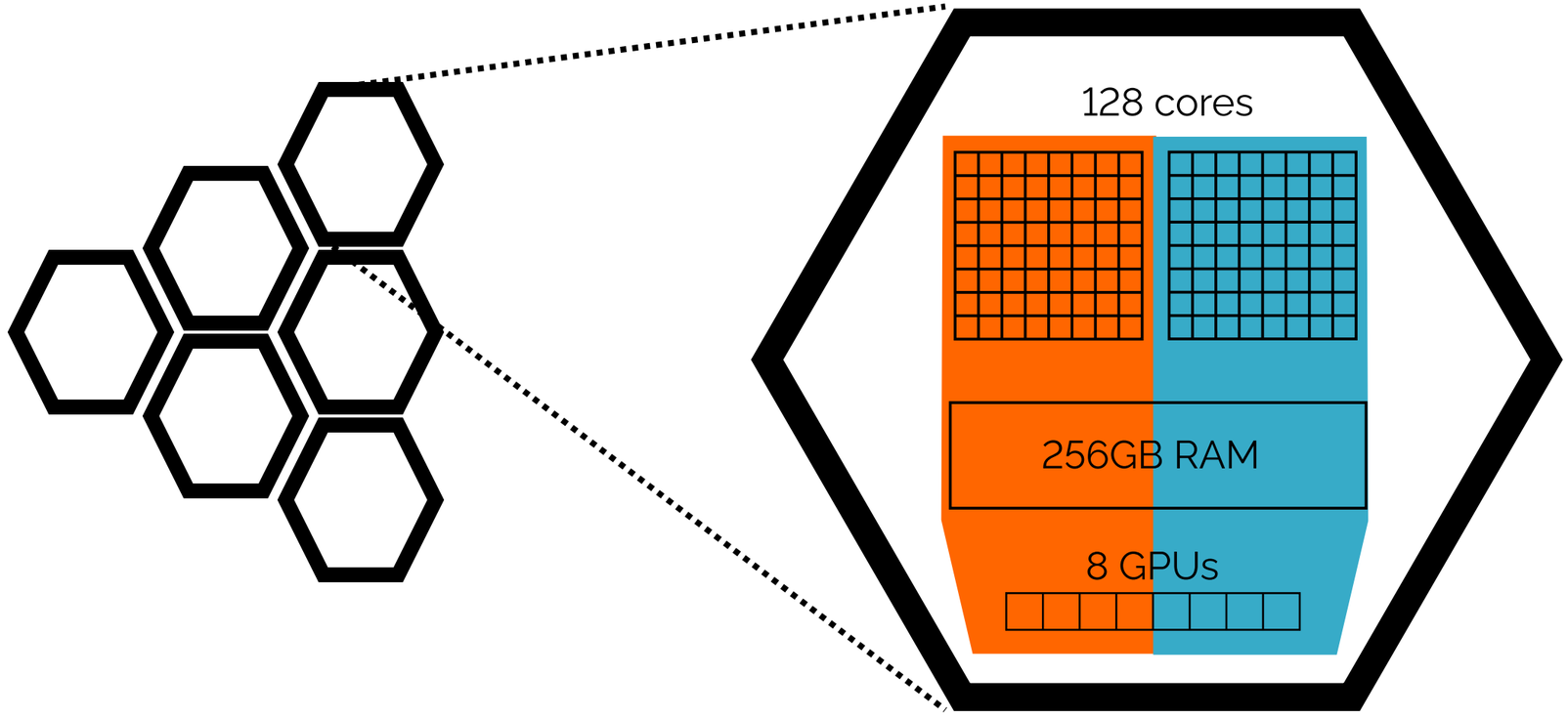

1 core
32 cores
128 GB ram
16 cores
2 gpu
32 cores
in the same socket
128 cores
256GB RAM
8 GPUs

1 core

32 cores
128 GB ram

16 cores
2 gpu

32 cores
in the same socket

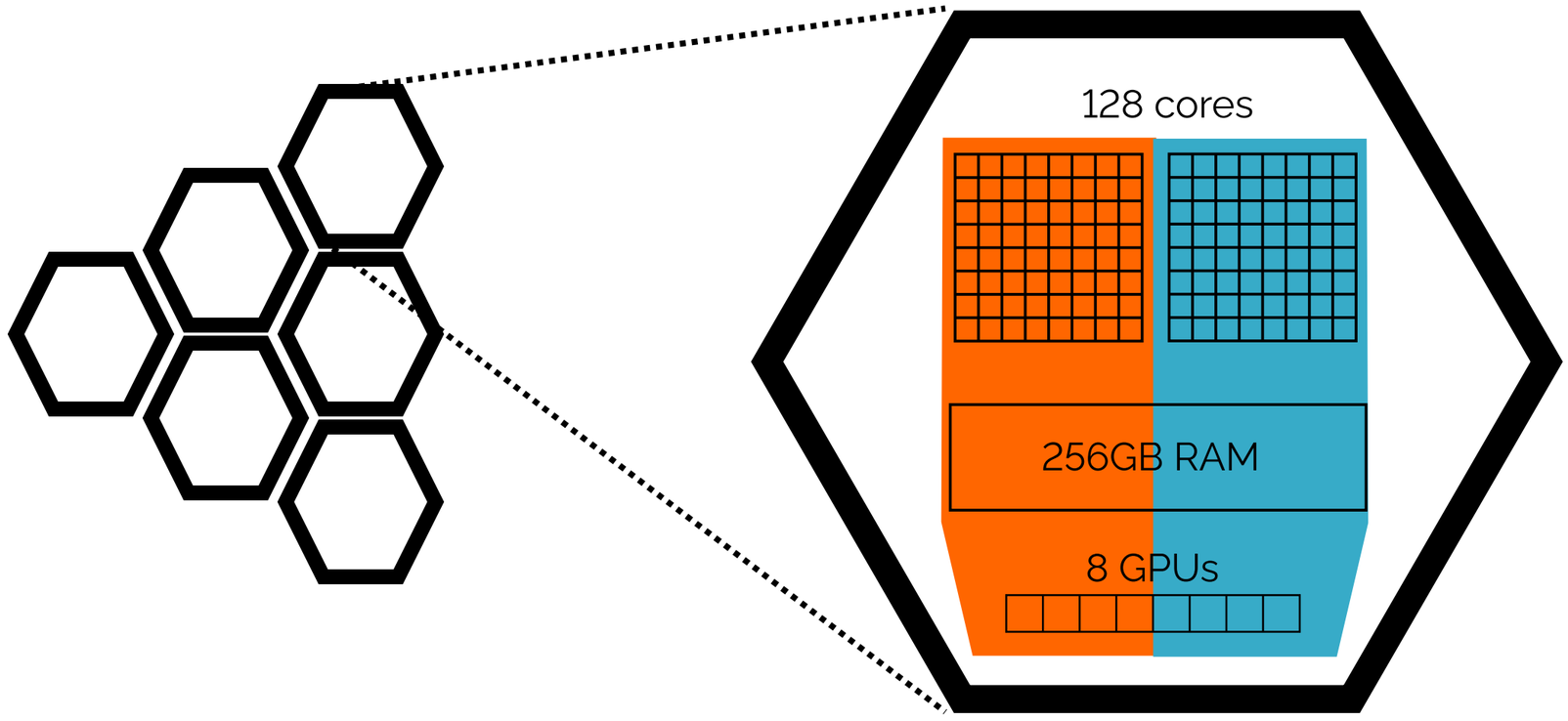


1 core

32 cores
128 GB ram

16 cores
2 gpu

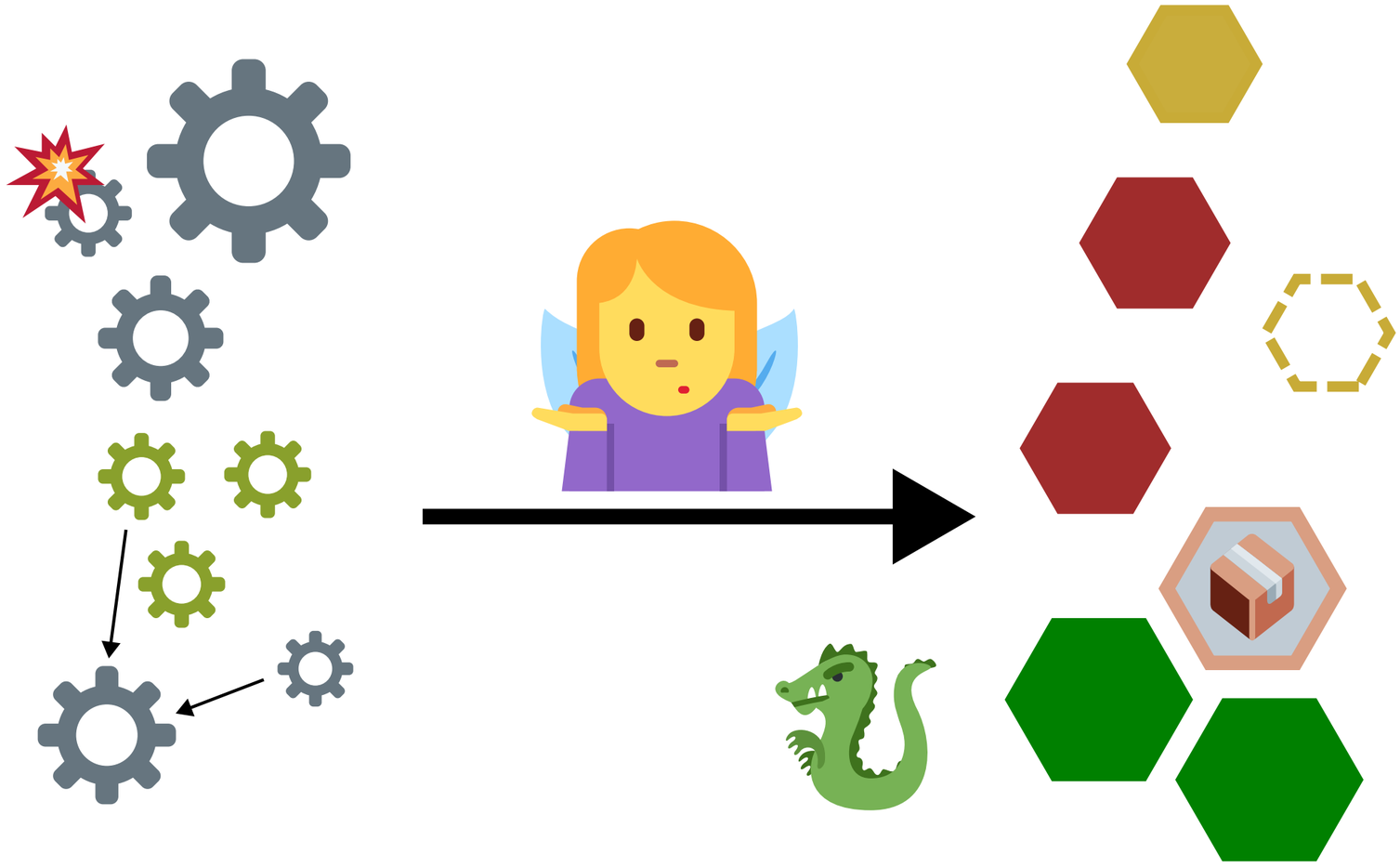
32 cores
in the same socket



128 cores

64 cores
8 GPUs

32 cores
24TB ram



Slurm/PBS approach:

Slurm/PBS approach:

What to compute (tasks) +

Slurm/PBS approach:

What to compute (tasks) +
Where to compute it (nodes)

Slurm/PBS approach:

What to compute (tasks) +
Where to compute it (nodes)
=> always defined together

HyperQueue

https://github.com/it4innovations/hyperqueue

HyperQueue

https://github.com/it4innovations/hyperqueue

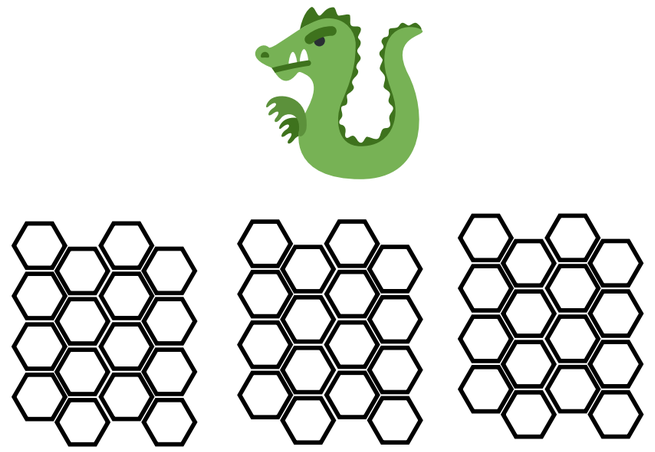
Distributed task runtime

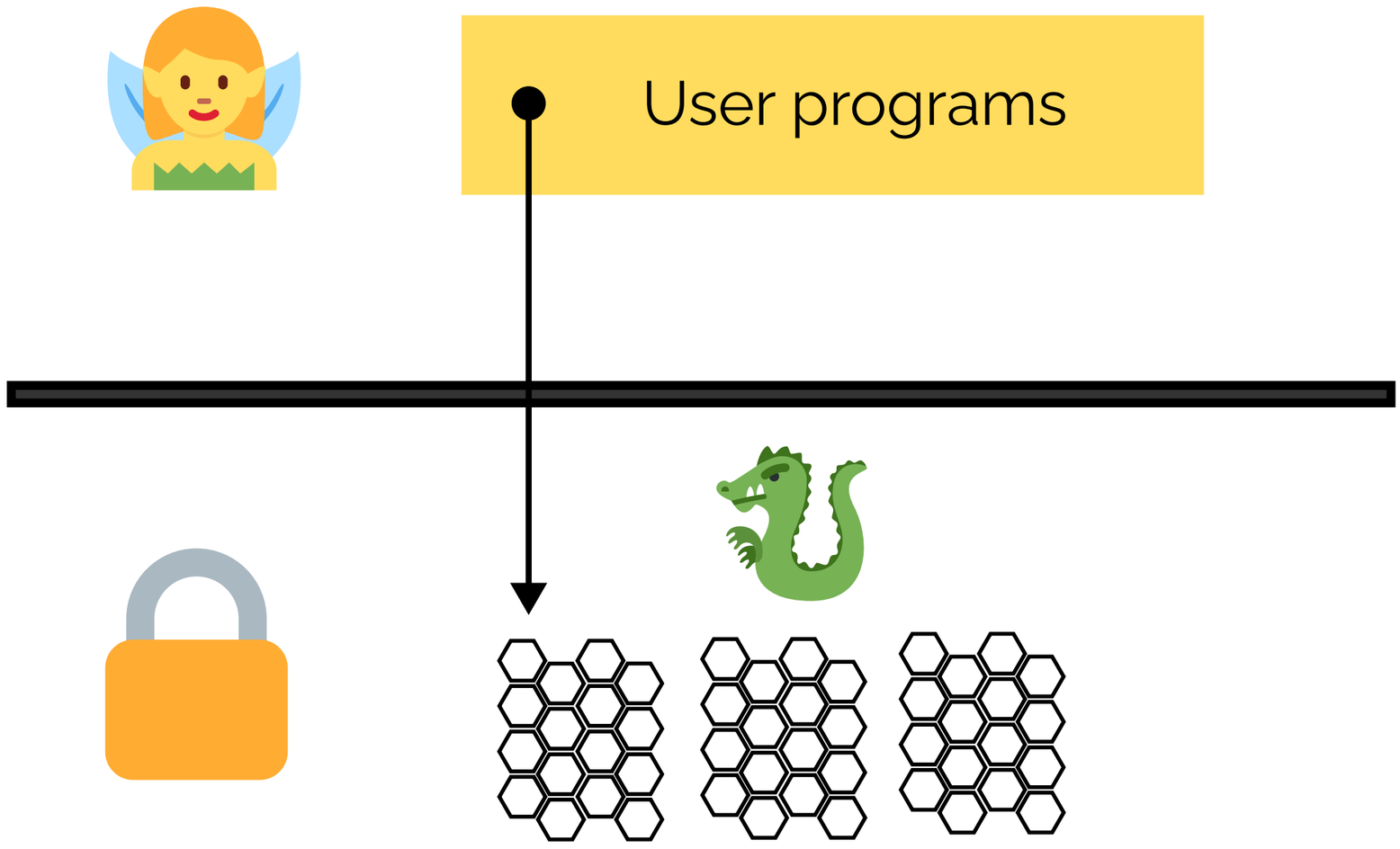
HyperQueue

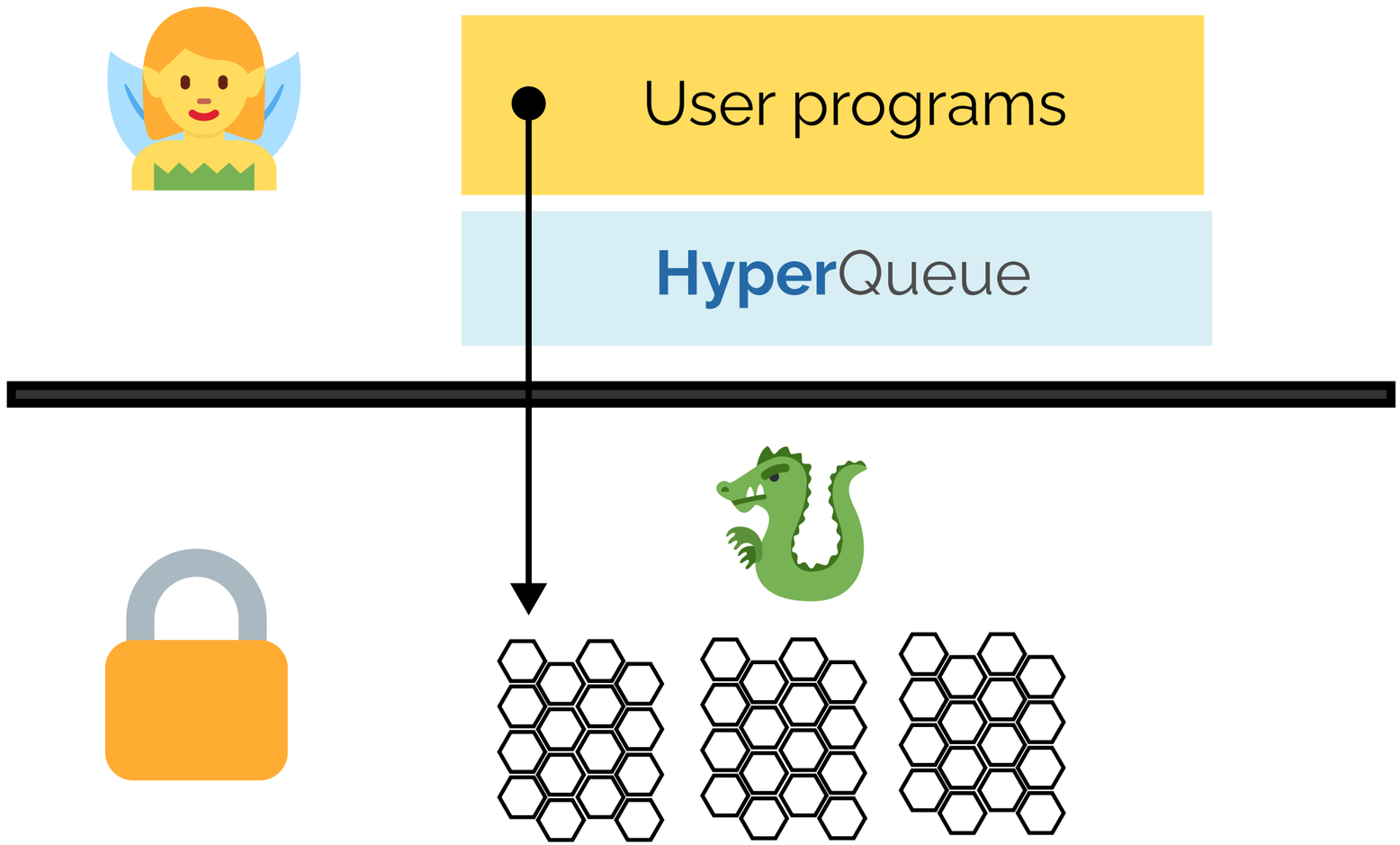
https://github.com/it4innovations/hyperqueue

Distributed task runtime

Efficient and ergonomic task (graph) execution on HPC clusters




User programs


User programs
HyperQueue

HyperQueue

Installation

HyperQueue

Installation

- Single binary

HyperQueue

Installation

- Single binary
- No dependencies

HyperQueue

Installation

- Single binary
- No dependencies
- No admin privileges needed

HyperQueue

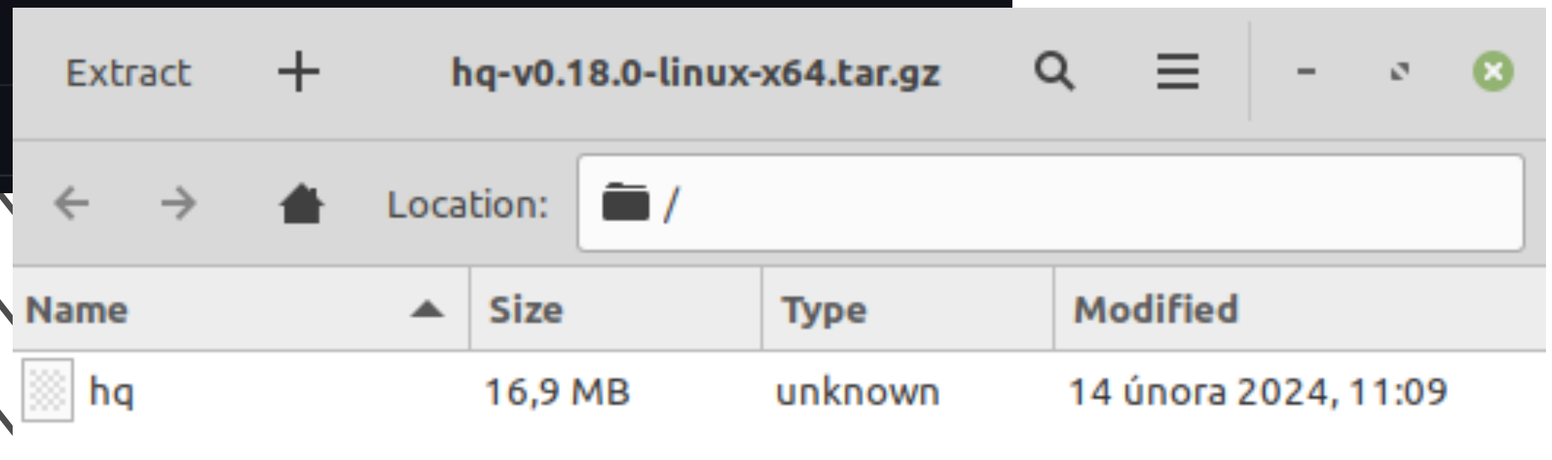
Installation

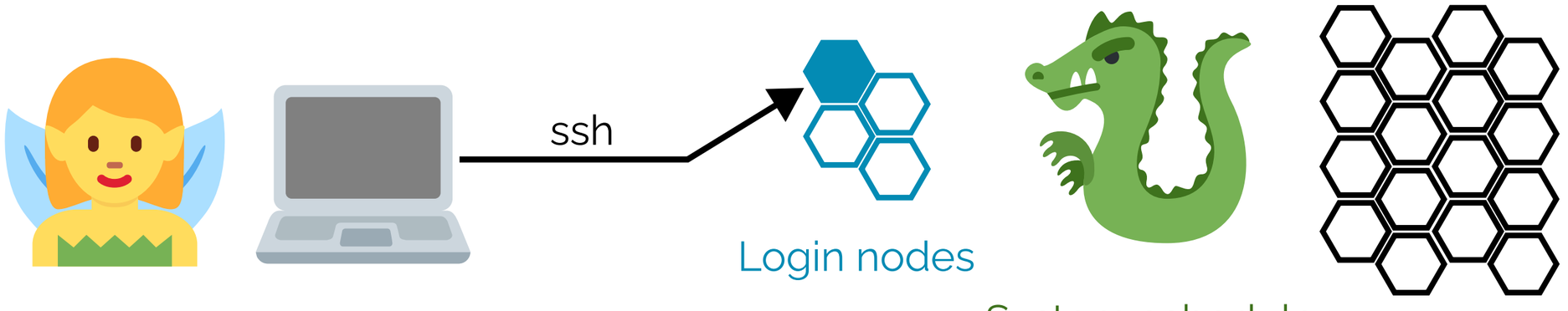
- Single binary
- No dependencies
- No admin privileges needed
- No configuration

- No dependencies
- No admin privileges needed
- No configuration

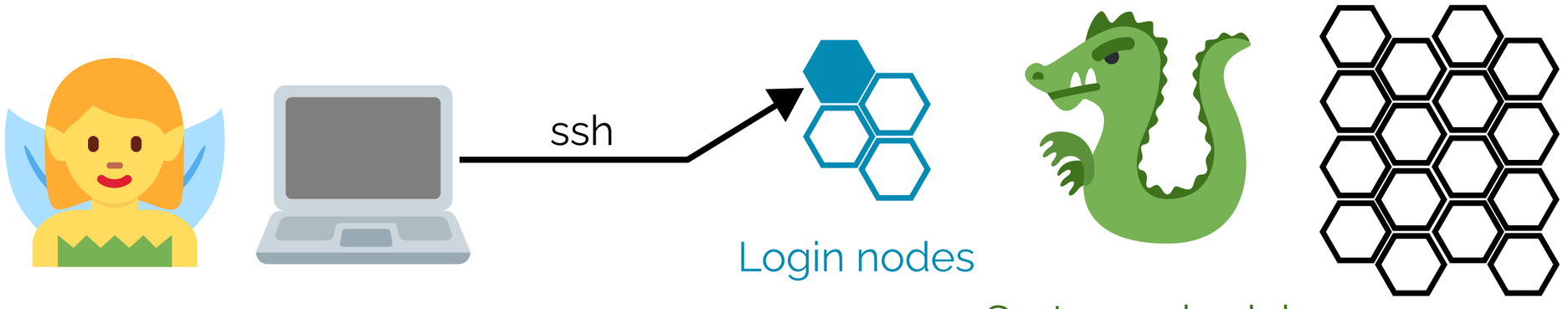
Assets 7

hq-v0.18.0-linux-arm64-linux.tar.gz	8.62 MB
hq-v0.18.0-linux-powerpc64.tar.gz	8.75 MB
hq-v0.18.0-linux-raspberry-pi.tar.gz	8.66 MB
hq-v0.18.0-linux-x64.tar.gz	11.1 MB
hyperqueue-0.18.0-cp36-abi3-manylinux_2_17_x86_64.manylinux2014_...	5.63 MB
Source code (zip)	
Source code (tar.gz)	

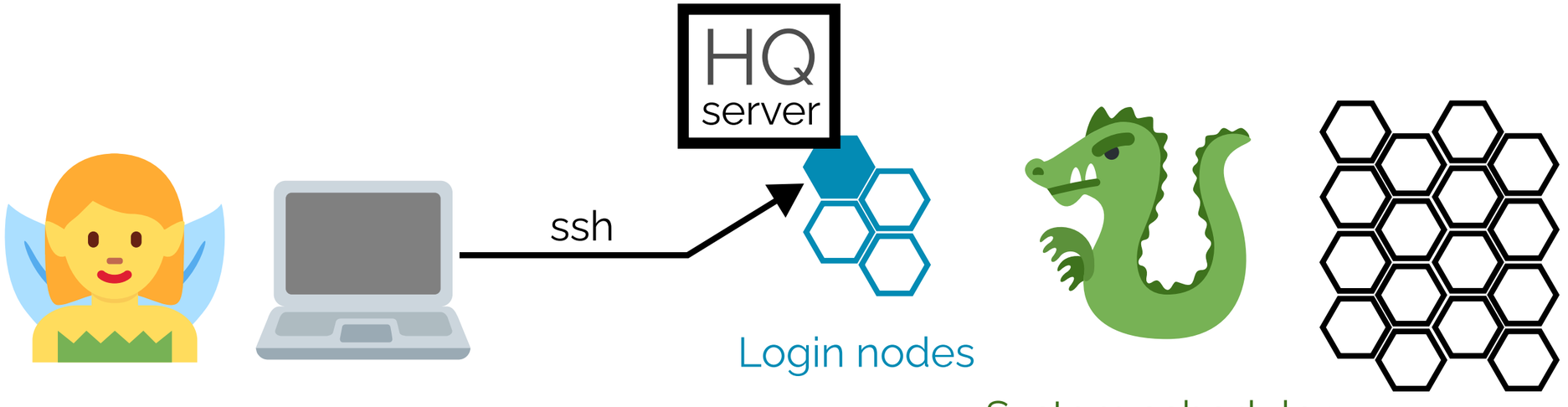




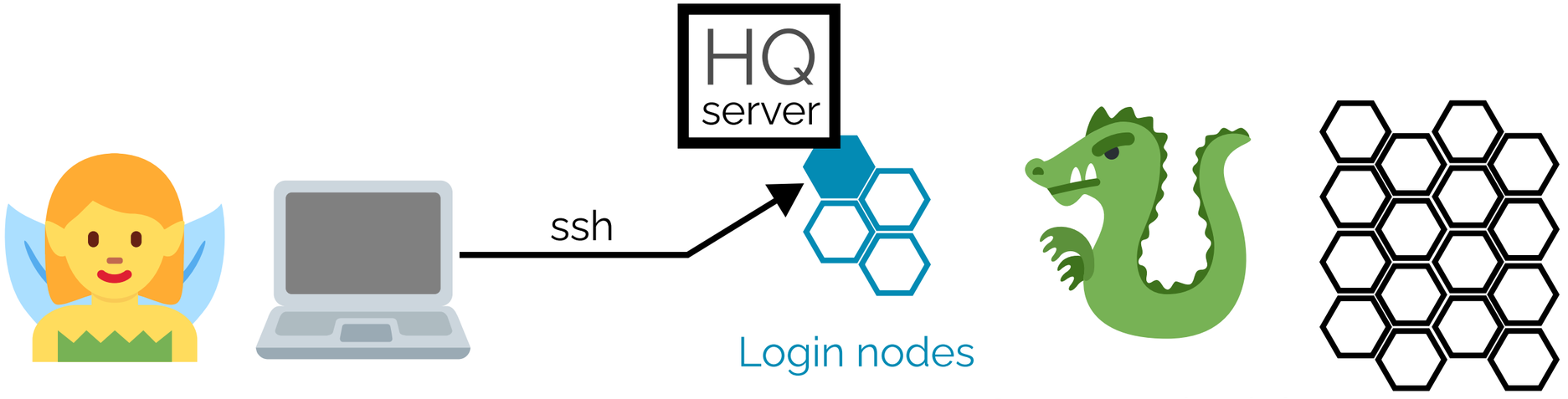

```
login1.karolina$ _
```



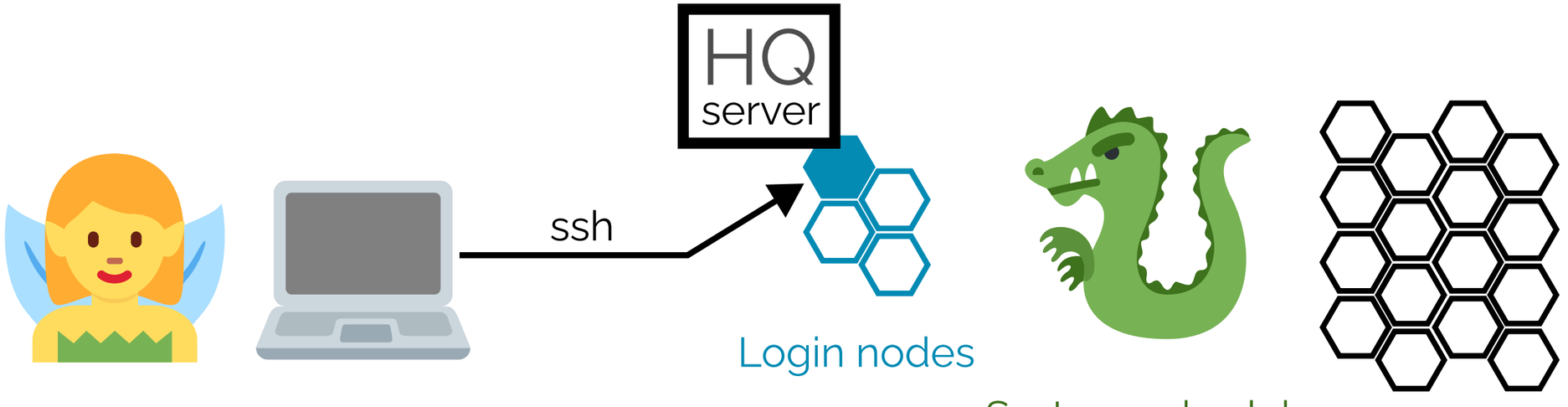

```
login1.karolina$ hq server start
```



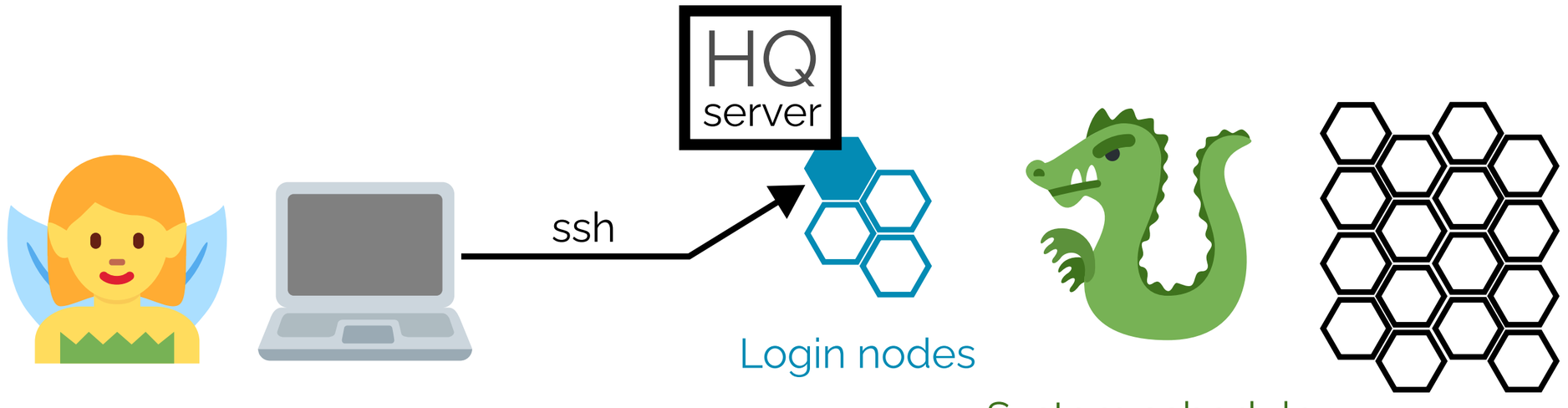

```
login1.karolina$ hq server start
```



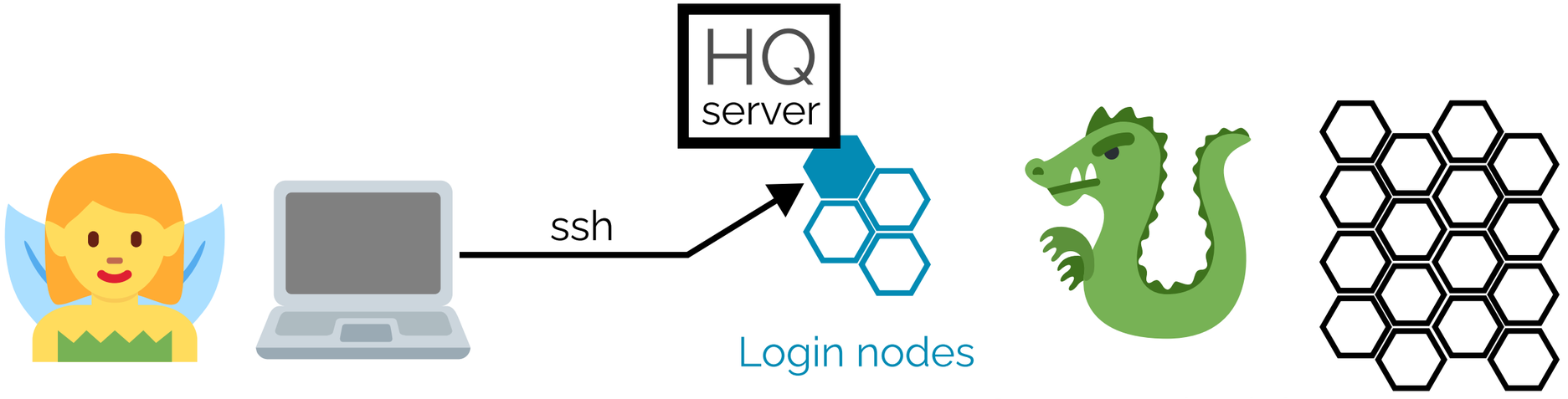

```
login1.karolina$ hq server start
INFO No online server found, starting a new server
INFO Saving access file as '/home/boh126/.hq-server/046/access.json'
+------------------+------------------------+
| Server directory | /home/boh126/.hq-server |
| Server UID       | xaTcNZ                  |
| Host             | login1.karolina         |
| Pid              | 51752                   |
| HQ port          | 44367                   |
| Workers port     | 44783                   |
| Start date       | 2024-04-30 12:05:00 UTC |
| Version          | 0.18.0                  |
+------------------+------------------------+
```



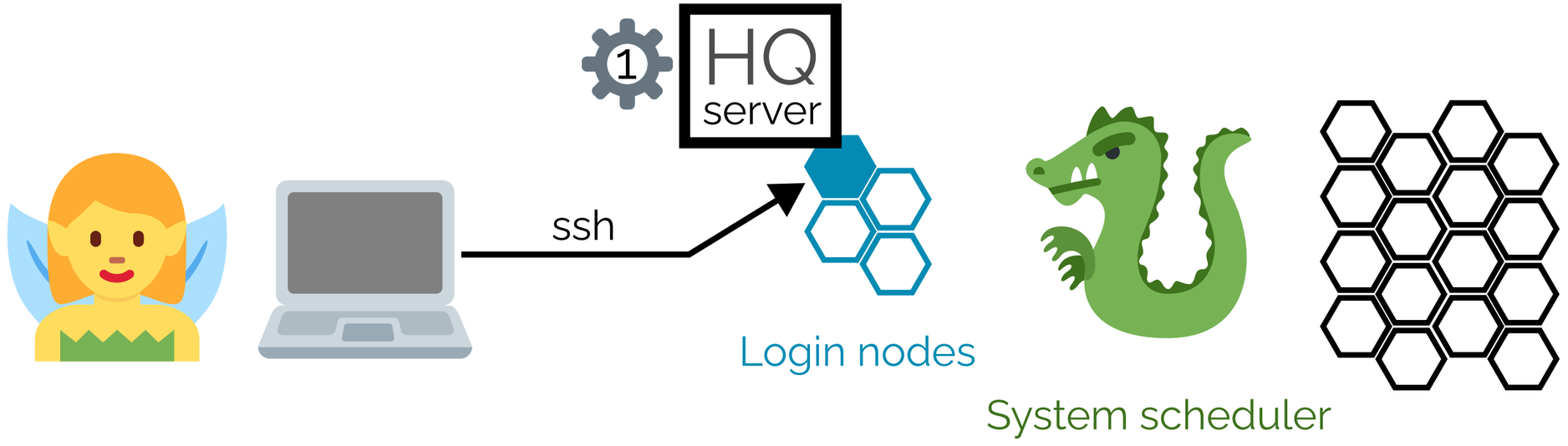

```
login1.karolina$ _
```



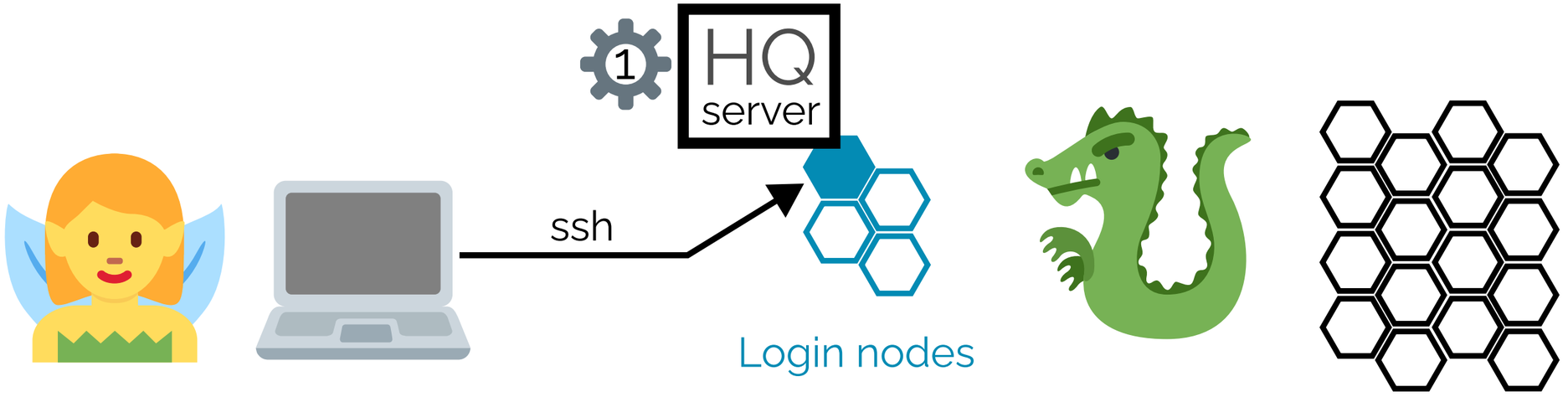

```
login1.karolina$ hq submit ./my-computation
```




```
login1.karolina$ hq submit ./my-computation
Job submitted successfully, job ID: 1
```

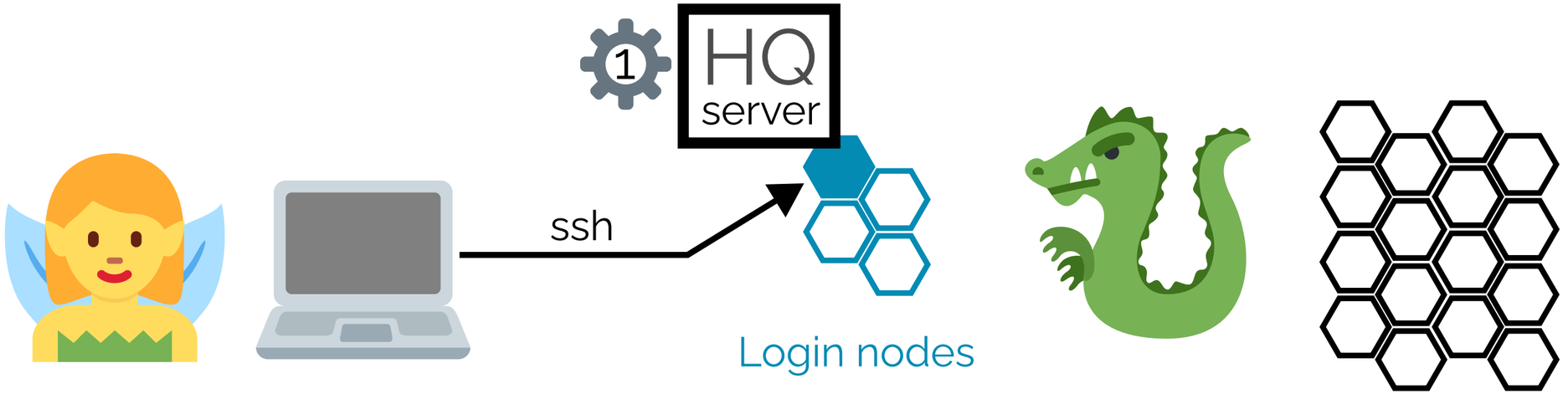



```
login1.karolina$ hq submit ./my-computation
Job submitted successfully, job ID: 1
```



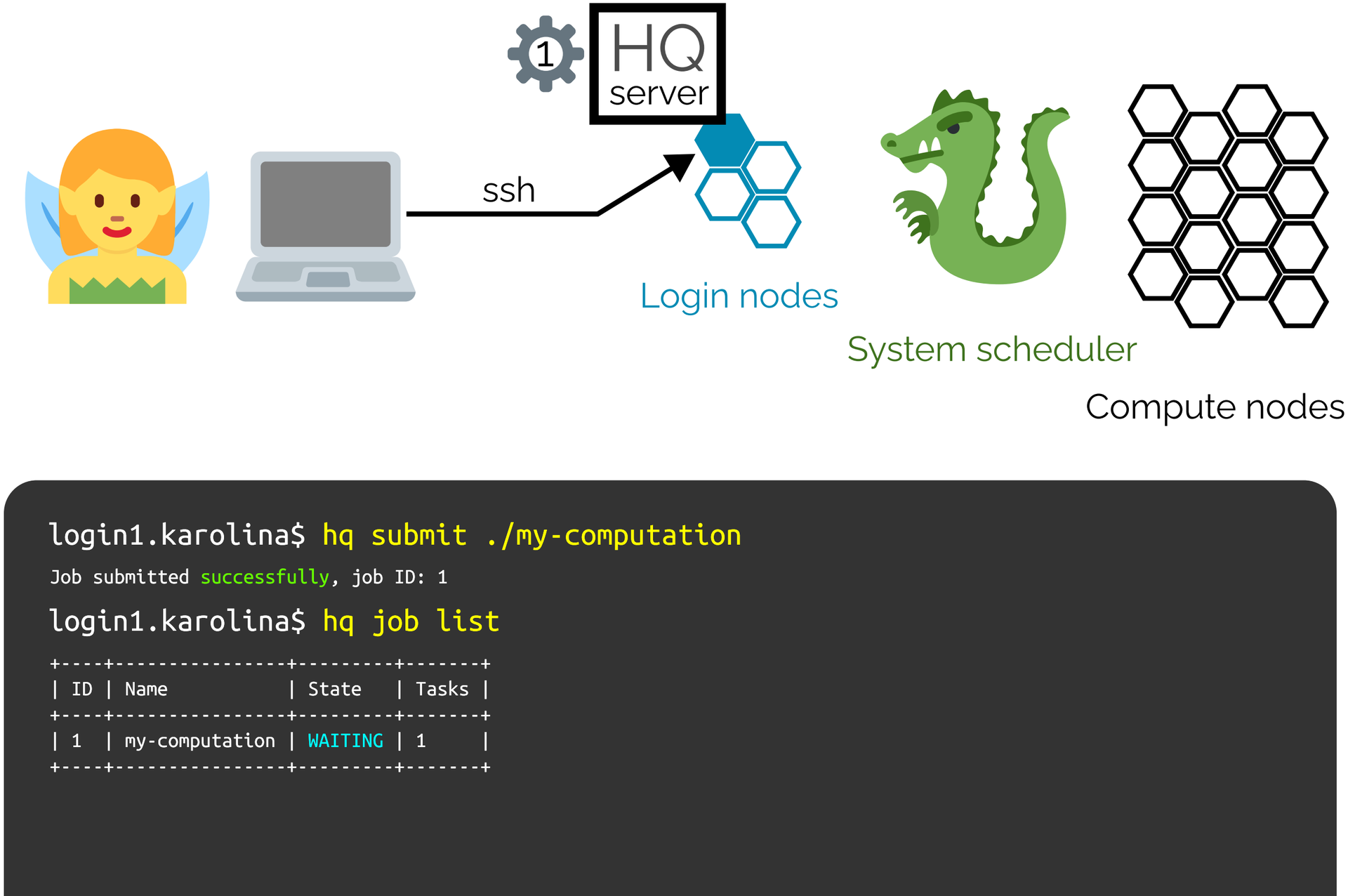

```
login1.karolina$ hq submit ./my-computation
Job submitted successfully, job ID: 1
login1.karolina$ hq job list
+----+----------------+---------+-------+
| ID | Name           | State   | Tasks |
+----+----------------+---------+-------+
| 1  | my-computation | WAITING | 1     |
+----+----------------+---------+-------+
```

```
machine$ _
```

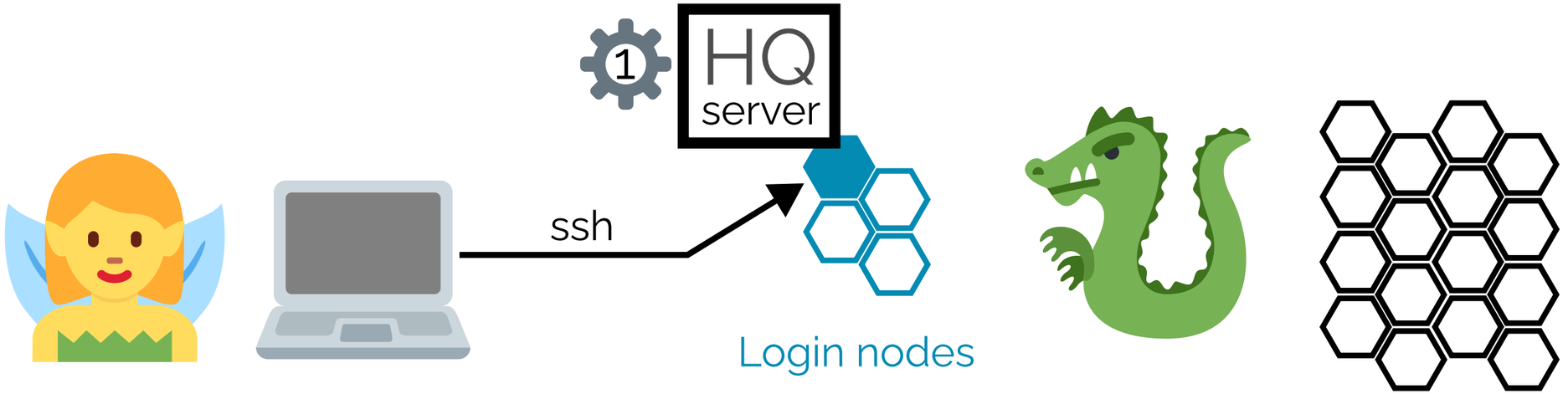



```
login1.karolina$ hq submit ./my-computation
Job submitted successfully, job ID: 1
login1.karolina$ hq job list
+----+----------------+---------+-------+
| ID | Name           | State   | Tasks |
+----+----------------+---------+-------+
| 1  | my-computation | WAITING | 1     |
+----+----------------+---------+-------+
```

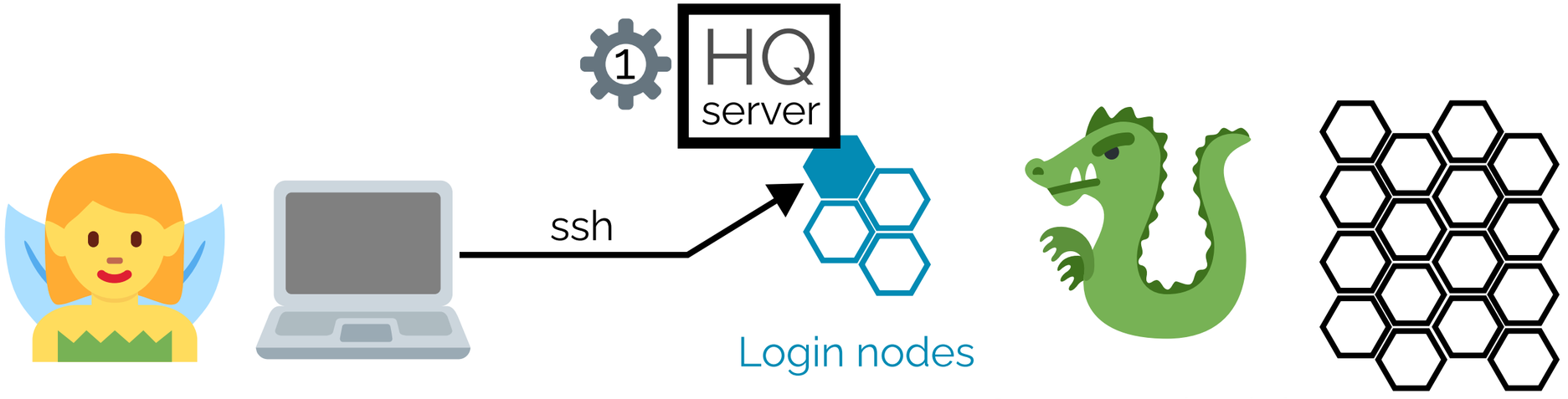
```
machine$ hq worker start
```



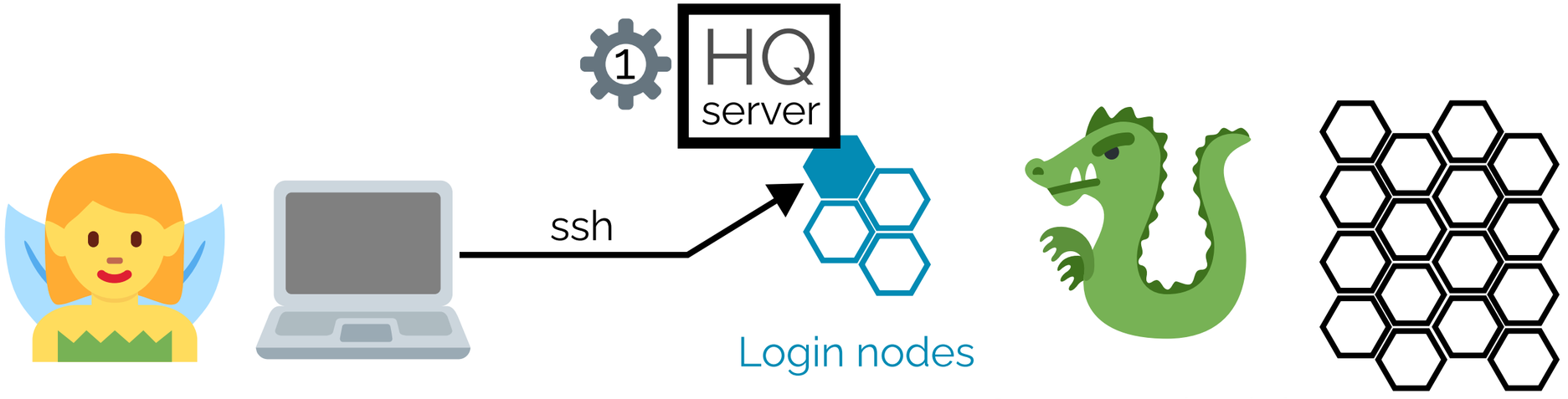

```
login1.karolina$ hq submit ./my-computation
Job submitted successfully, job ID: 1
login1.karolina$ hq job list
+----+----------------+---------+-------+
| ID | Name           | State   | Tasks |
+----+----------------+---------+-------+
| 1  | my-computation | WAITING | 1     |
+----+----------------+---------+-------+
```



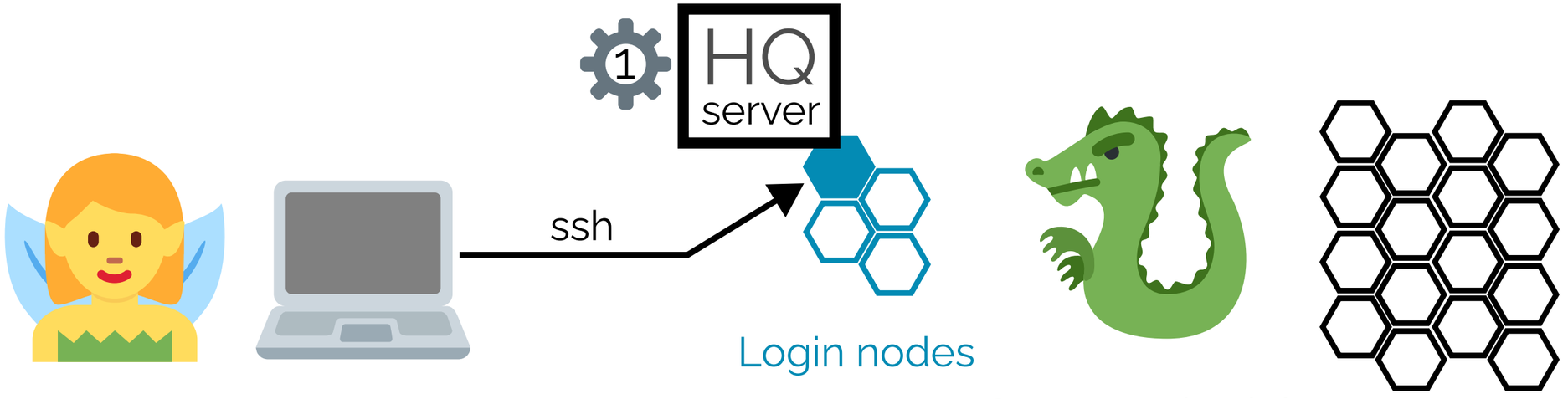

```
login1.karolina$ hq submit ./my-computation
Job submitted successfully, job ID: 1
login1.karolina$ hq job list
+----+---------------+---------+-------+
| ID | Name          | State   | Tasks |
+----+---------------+---------+-------+
| 1  | my-computation | WAITING | 1     |
+----+---------------+---------+-------+
login1.karolina$ hq alloc add <SYSTEM_SCHEDULER>
```



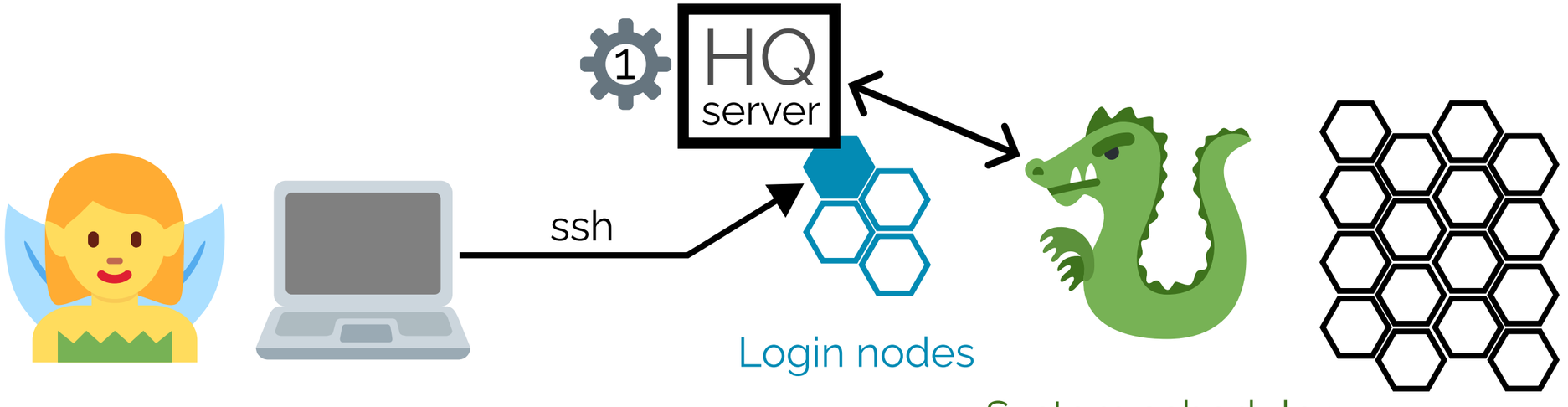

```
login1.karolina$ hq submit ./my-computation
Job submitted successfully, job ID: 1
login1.karolina$ hq job list
+----+---------------+---------+-------+
| ID | Name          | State   | Tasks |
+----+---------------+---------+-------+
| 1  | my-computation | WAITING | 1     |
+----+---------------+---------+-------+
login1.karolina$ hq alloc add slurm
```



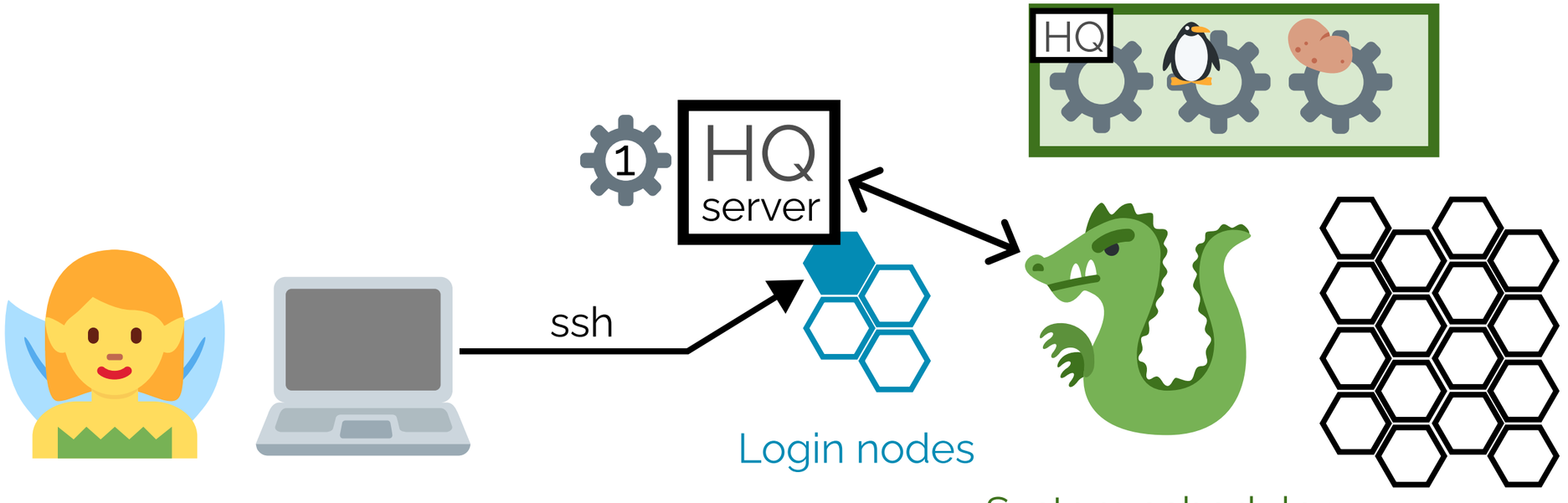

```
login1.karolina$ hq submit ./my-computation
Job submitted successfully, job ID: 1
login1.karolina$ hq job list
+----+----------------+---------+-------+
| ID | Name           | State   | Tasks |
+----+----------------+---------+-------+
| 1  | my-computation | WAITING | 1     |
+----+----------------+---------+-------+
login1.karolina$ hq alloc add slurm --timelimit=1h
```



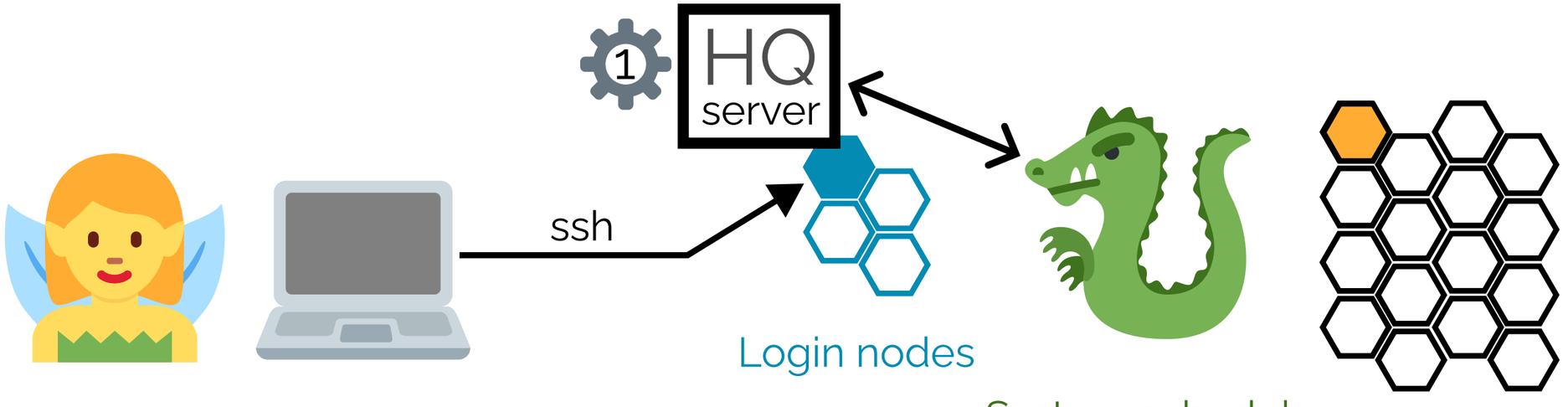

```
login1.karolina$ hq submit ./my-computation
Job submitted successfully, job ID: 1
login1.karolina$ hq job list
+----+---------------+---------+-------+
| ID | Name          | State   | Tasks |
+----+---------------+---------+-------+
| 1  | my-computation | WAITING | 1     |
+----+---------------+---------+-------+
login1.karolina$ hq alloc add slurm --timelimit=1h -- -pstandard
```



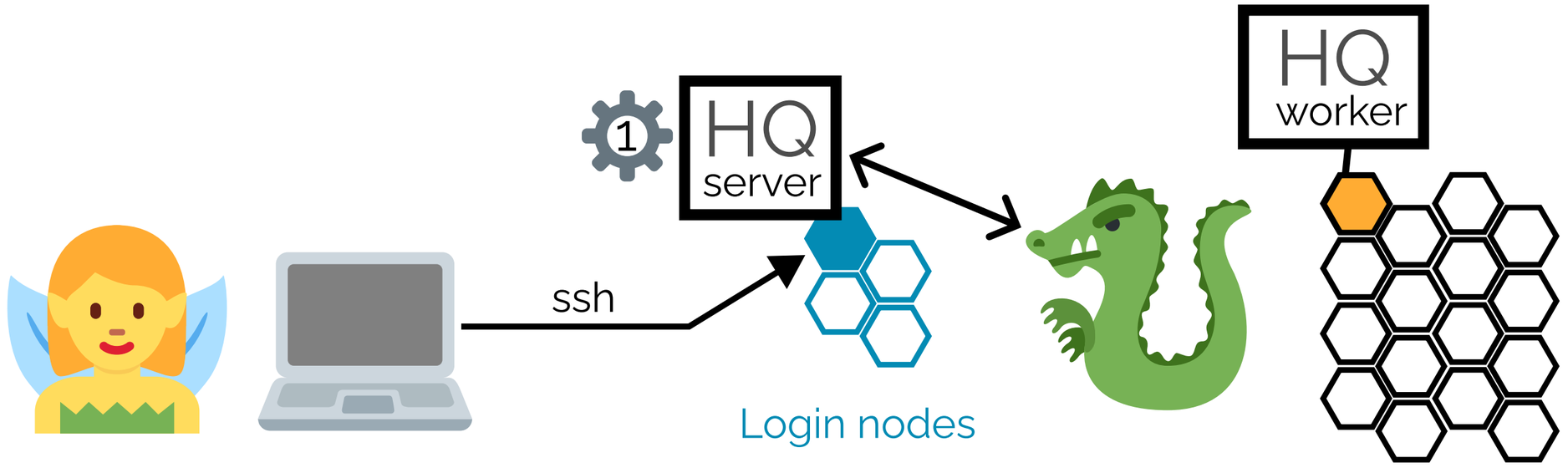

```
login1.karolina$ hq submit ./my-computation
Job submitted successfully, job ID: 1
login1.karolina$ hq job list
+----+----------------+---------+-------+
| ID | Name           | State   | Tasks |
+----+----------------+---------+-------+
| 1  | my-computation | WAITING | 1     |
+----+----------------+---------+-------+
login1.karolina$ hq alloc add slurm --timelimit=1h -- -pstandard
```



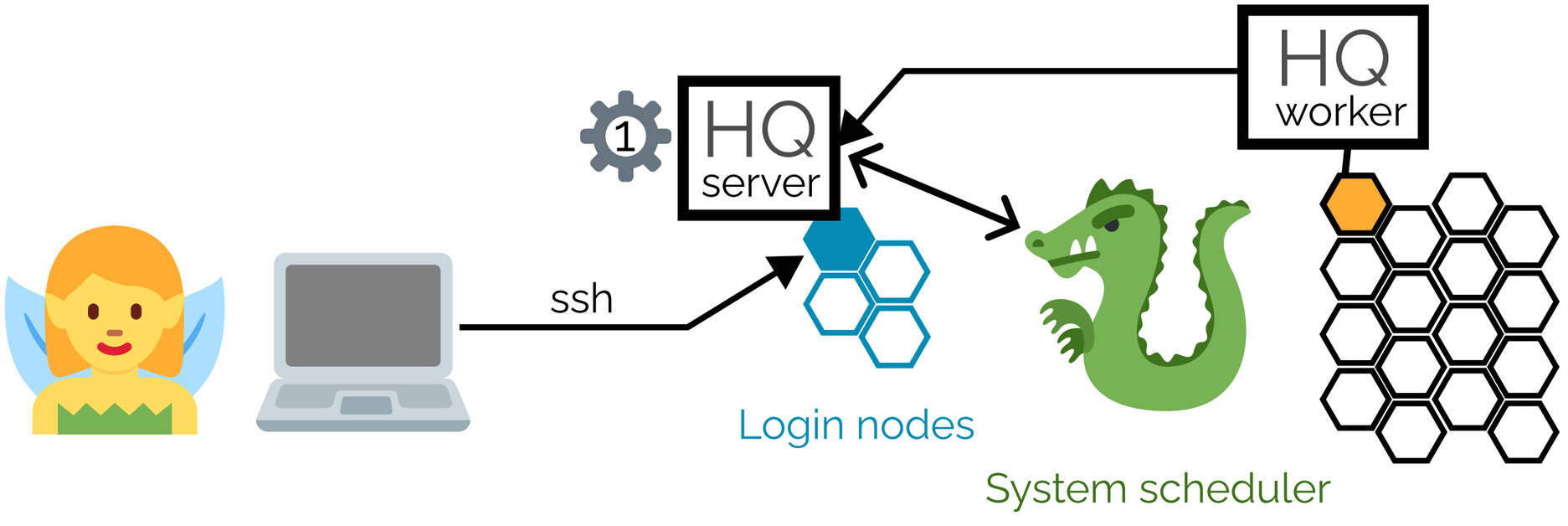

```
login1.karolina$ hq submit ./my-computation
Job submitted successfully, job ID: 1
login1.karolina$ hq job list
+----+----------------+---------+-------+
| ID | Name           | State   | Tasks |
+----+----------------+---------+-------+
| 1  | my-computation | WAITING | 1     |
+----+----------------+---------+-------+
login1.karolina$ hq alloc add slurm --timelimit=1h -- -pstandard
```



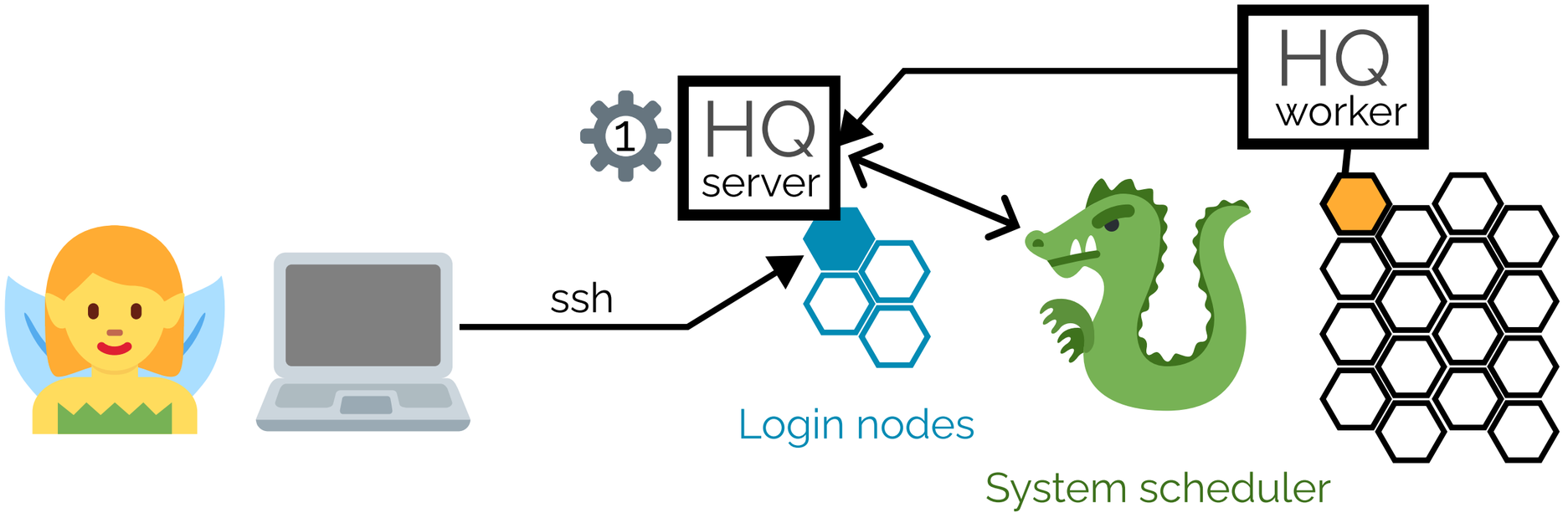

```
login1.karolina$ hq submit ./my-computation
Job submitted successfully, job ID: 1
login1.karolina$ hq job list
+----+----------------+---------+-------+
| ID | Name           | State   | Tasks |
+----+----------------+---------+-------+
| 1  | my-computation | WAITING | 1     |
+----+----------------+---------+-------+
login1.karolina$ hq alloc add slurm --timelimit=1h -- -pstandard
```



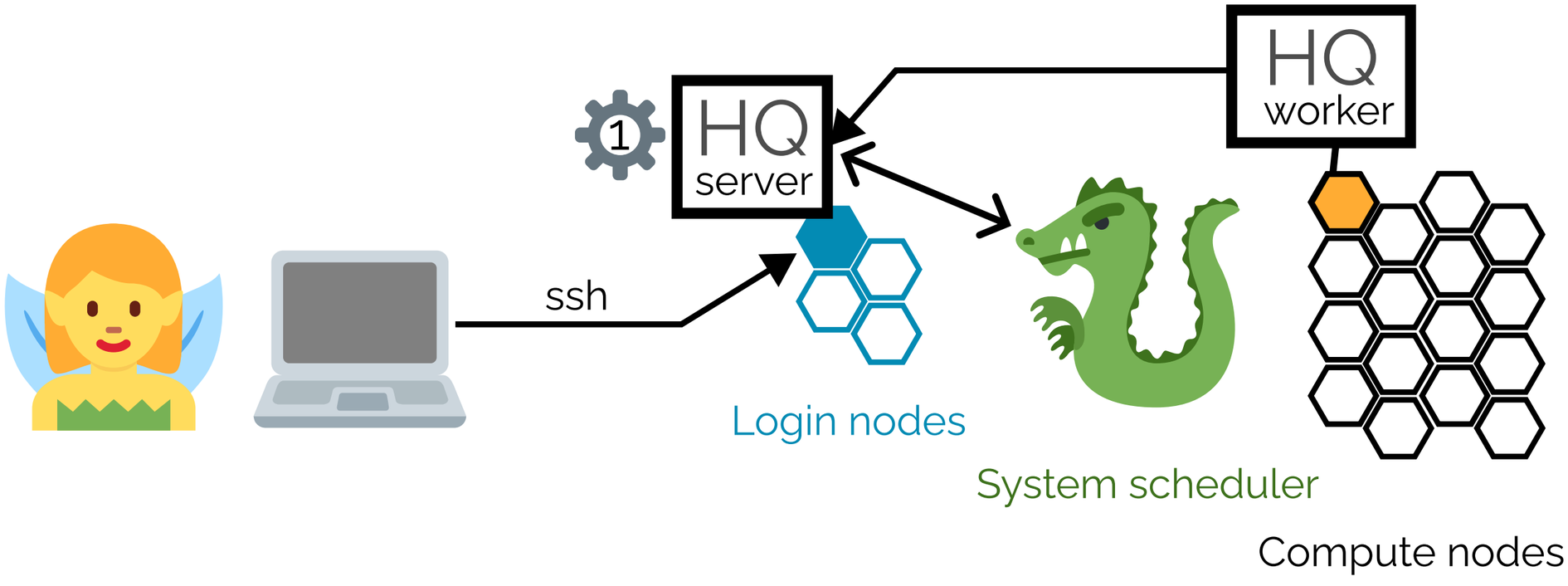

```
login1.karolina$ hq submit ./my-computation
Job submitted successfully, job ID: 1
login1.karolina$ hq job list
+----+----------------+---------+-------+
| ID | Name           | State   | Tasks |
+----+----------------+---------+-------+
| 1  | my-computation | WAITING | 1     |
+----+----------------+---------+-------+
login1.karolina$ hq alloc add slurm --timelimit=1h -- -pstandard
```



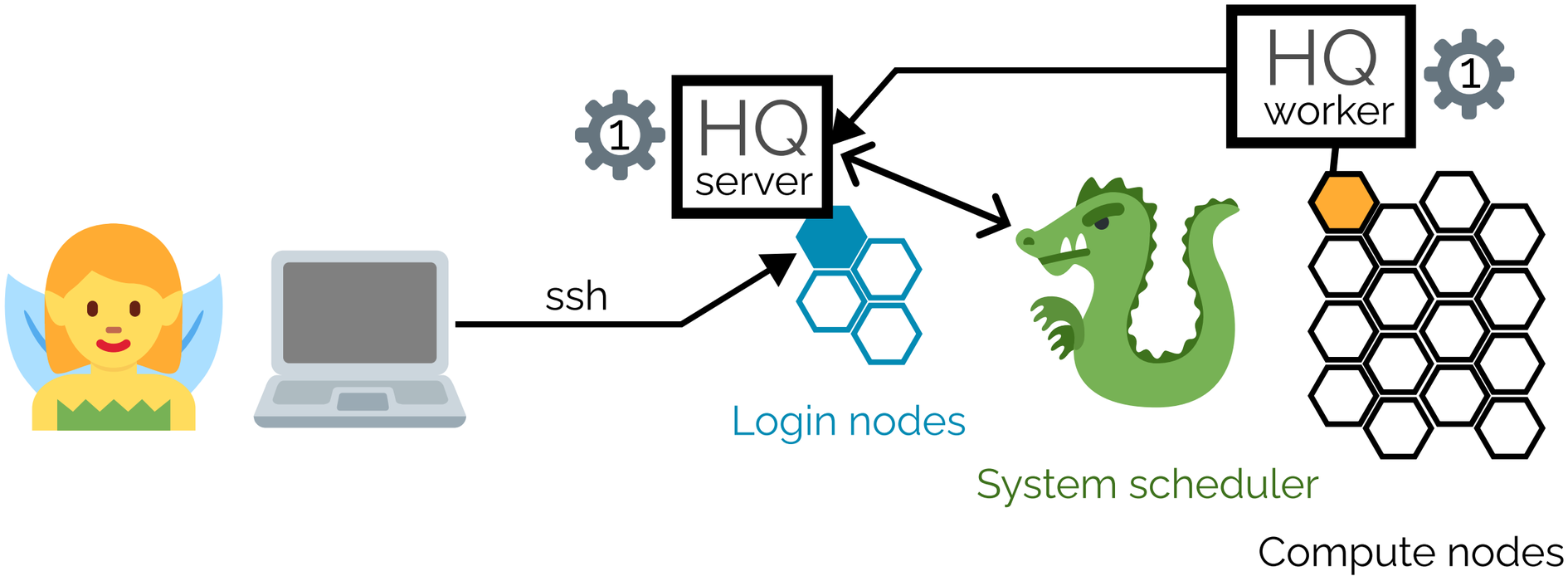

```
login1.karolina$ hq submit ./my-computation
Job submitted successfully, job ID: 1
login1.karolina$ hq job list
+----+----------------+---------+-------+
| ID | Name           | State   | Tasks |
+----+----------------+---------+-------+
| 1  | my-computation | WAITING | 1     |
+----+----------------+---------+-------+
login1.karolina$ hq alloc add slurm --timelimit=1h -- -pstandard
```



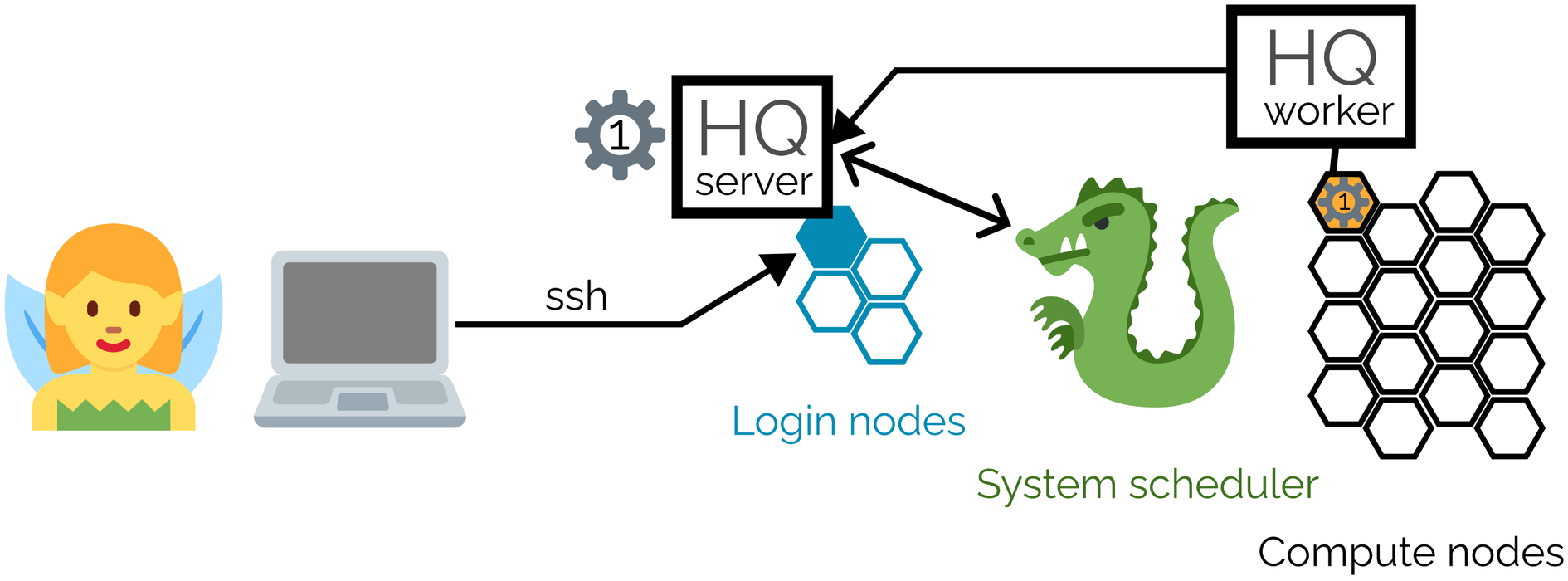

```
login1.karolina$ hq worker list
```



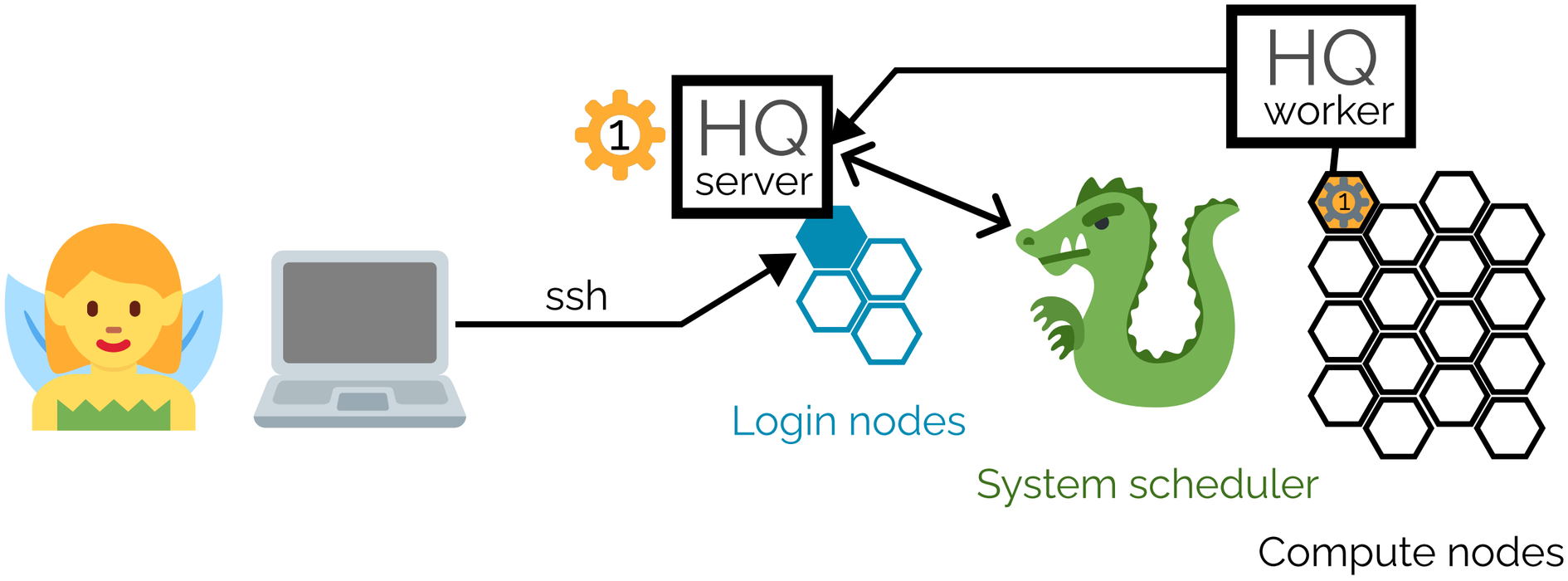

```
login1.karolina$ hq worker list
+----+---------+-----------------+-----------+---------+----------------+
| Id | State   | Hostname        | Resources | Manager | Manager Job Id |
+----+---------+-----------------+-----------+---------+----------------+
| 1  | RUNNING | cn.lumi         | 2x64 cpus | SLURM   | 550463         |
+----+---------+-----------------+-----------+---------+----------------+
```



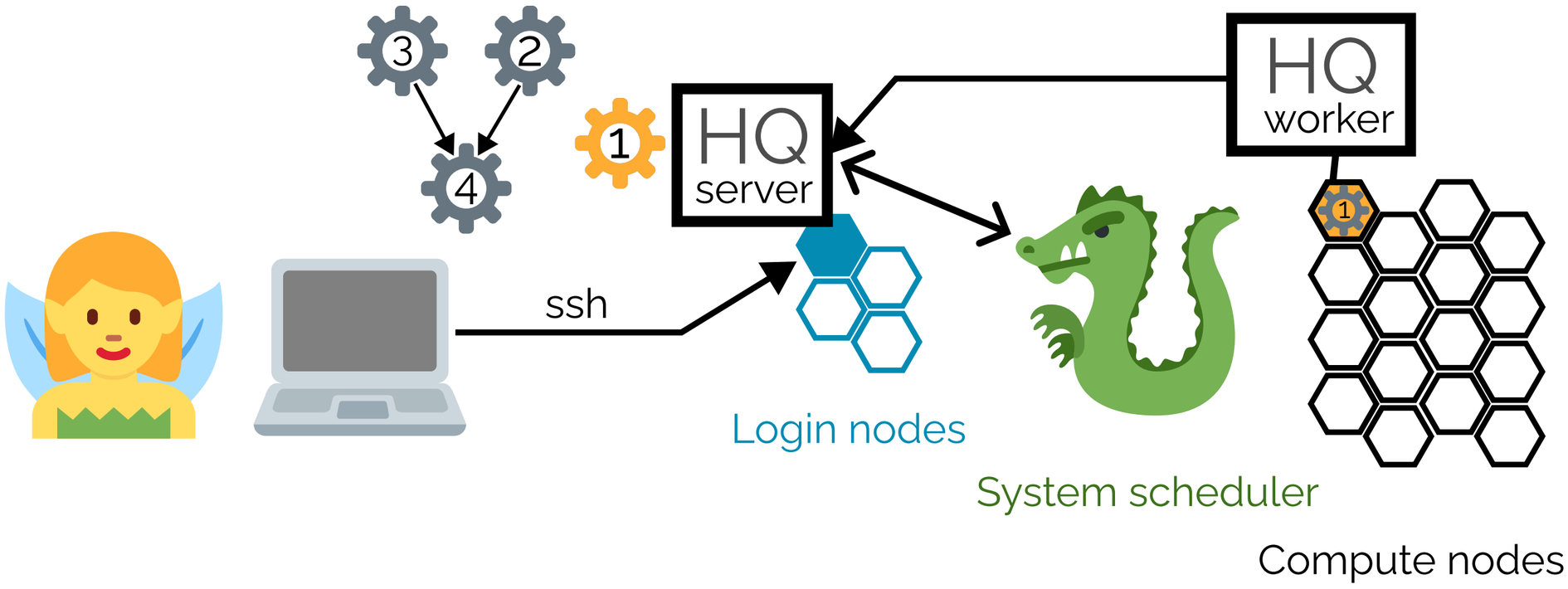

```
login1.karolina$ hq worker list
+----+---------+-----------------+-----------+---------+----------------+
| Id | State   | Hostname        | Resources | Manager | Manager Job Id |
+----+---------+-----------------+-----------+---------+----------------+
| 1  | RUNNING | cn.lumi         | 2x64 cpus | SLURM   | 550463         |
+----+---------+-----------------+-----------+---------+----------------+
```



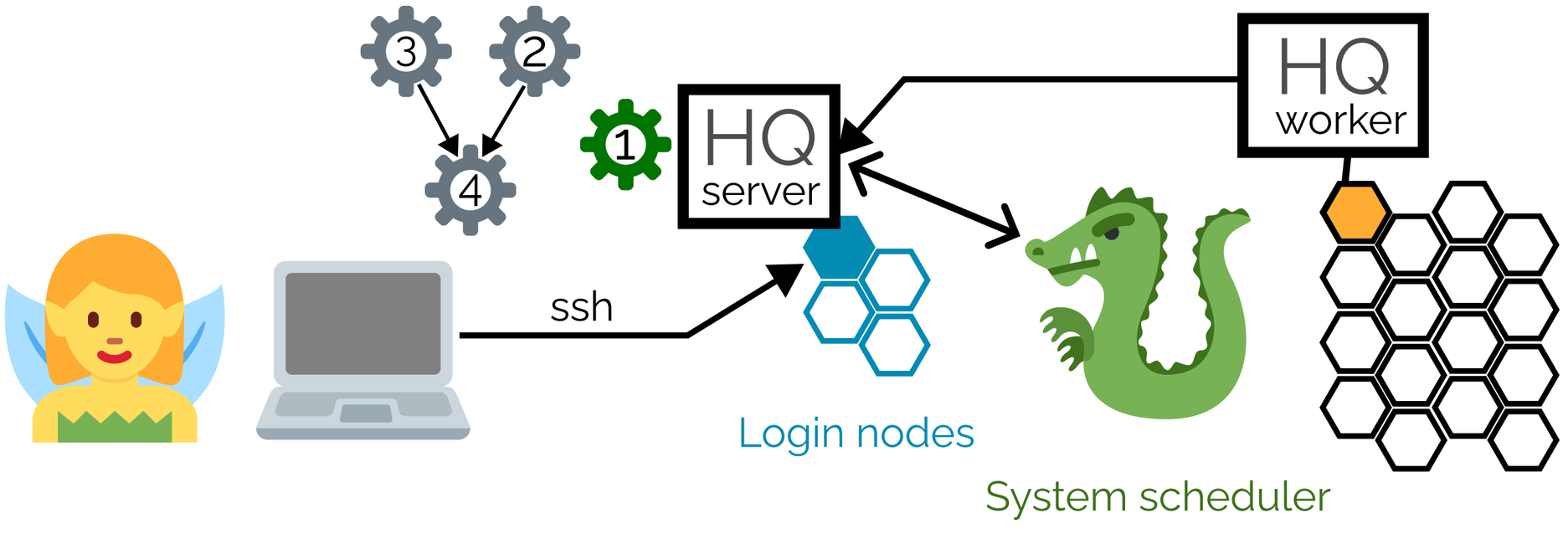

```
login1.karolina$ hq worker list
+----+---------+-----------------+-----------+---------+----------------+
| Id | State   | Hostname        | Resources | Manager | Manager Job Id |
+----+---------+-----------------+-----------+---------+----------------+
| 1  | RUNNING | cn.lumi         | 2x64 cpus | SLURM   | 550463         |
+----+---------+-----------------+-----------+---------+----------------+
```



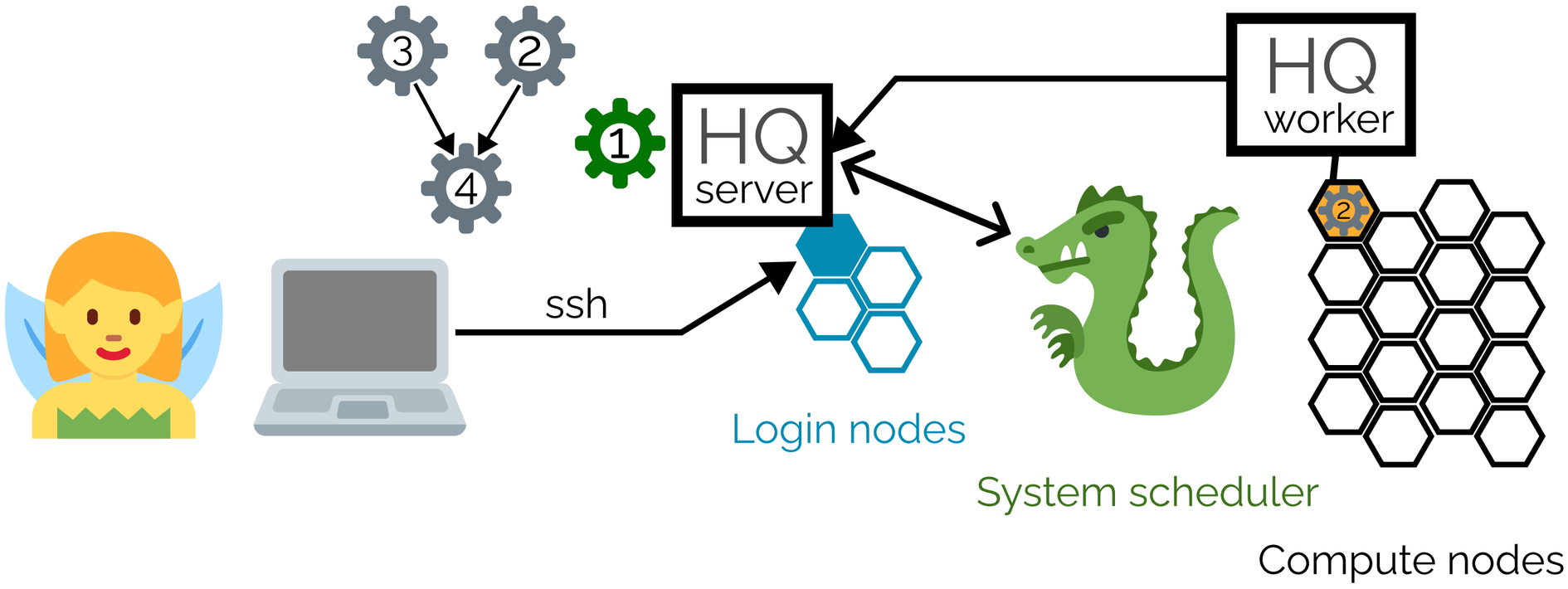

```
$ hq job list
+----+----------------+---------+-------+
| ID | Name           | State   | Tasks |
+----+----------------+---------+-------+
| 1  | my-computation | RUNNING | 1     |
+----+----------------+---------+-------+
```



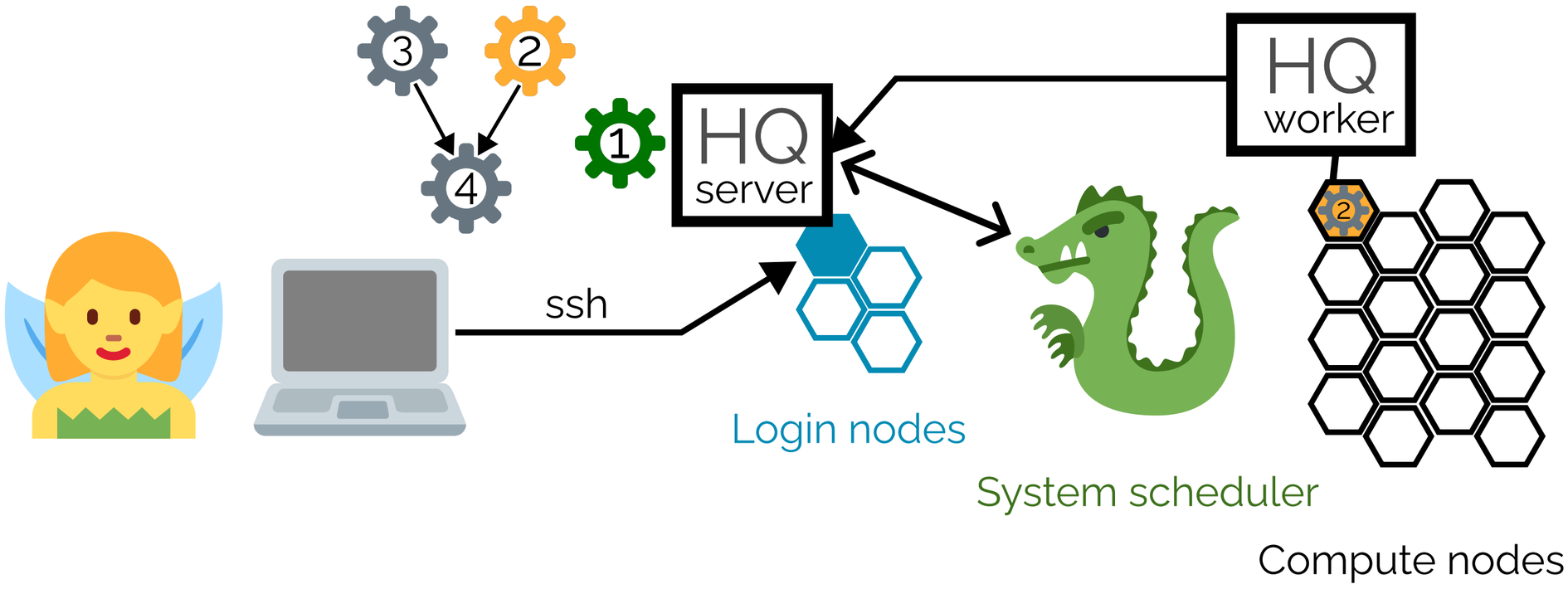

```
$ hq job list
+----+----------------+---------+-------+
| ID | Name           | State   | Tasks |
+----+----------------+---------+-------+
| 1  | my-computation | RUNNING | 1     |
+----+----------------+---------+-------+
| 2  | my-workflow    | WAITING | 3     |
+----+----------------+---------+-------+
```



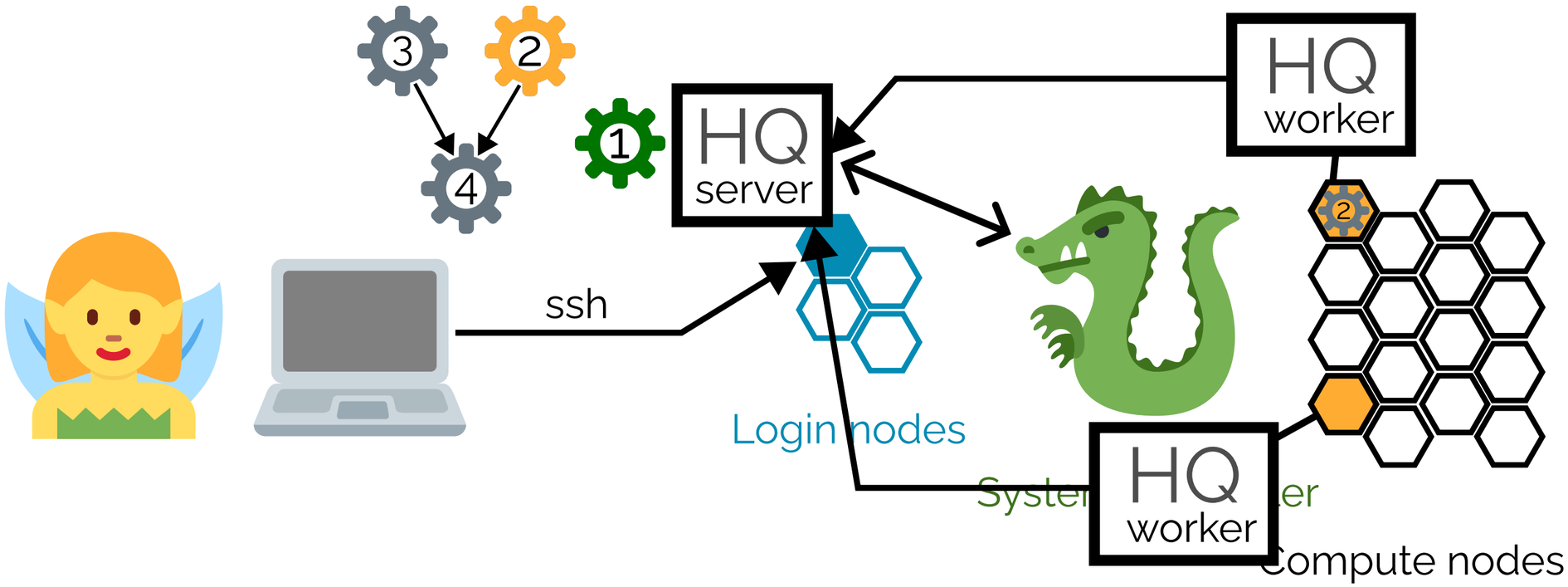

```
$ hq job list
+----+----------------+----------+-------+
| ID | Name           | State    | Tasks |
+----+----------------+----------+-------+
| 1  | my-computation | FINISHED | 1     |
+----+----------------+----------+-------+
| 2  | my-workflow    | WAITING  | 3     |
+----+----------------+----------+-------+
```



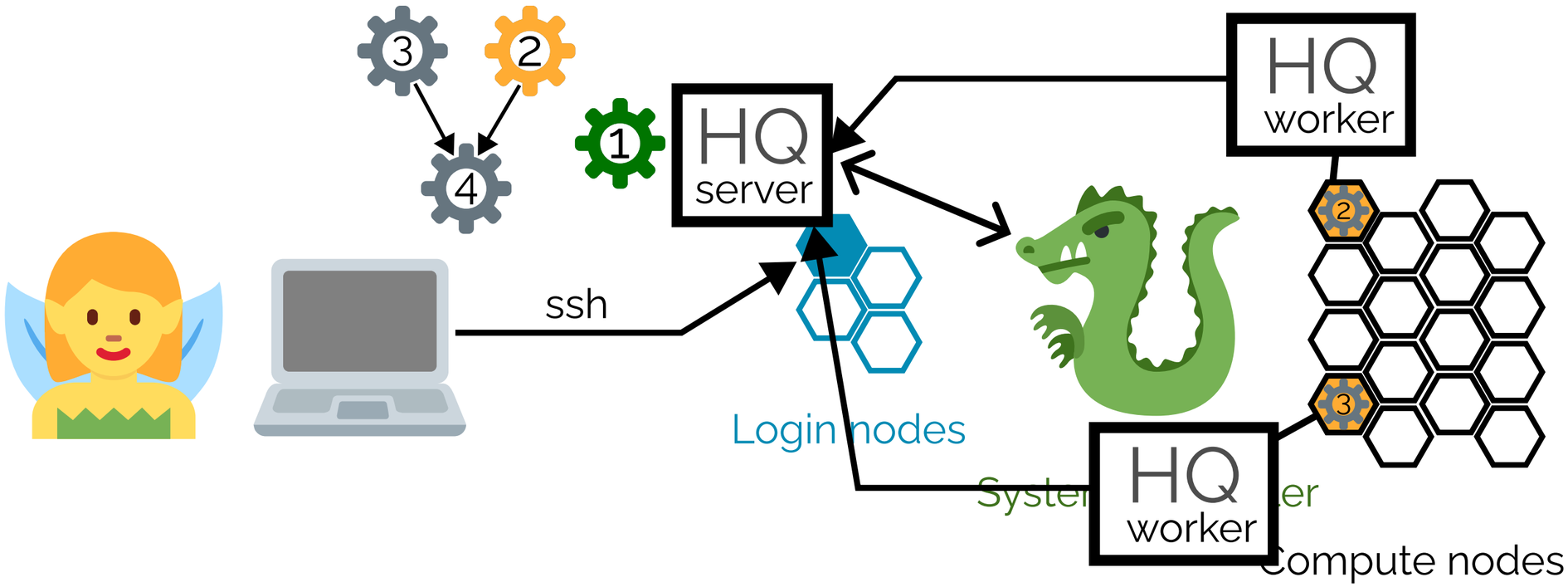

```
$ hq job list
+----+----------------+----------+-------+
| ID | Name           | State    | Tasks |
+----+----------------+----------+-------+
| 1  | my-computation | FINISHED | 1     |
+----+----------------+----------+-------+
| 2  | my-workflow    | WAITING  | 3     |
+----+----------------+----------+-------+
```



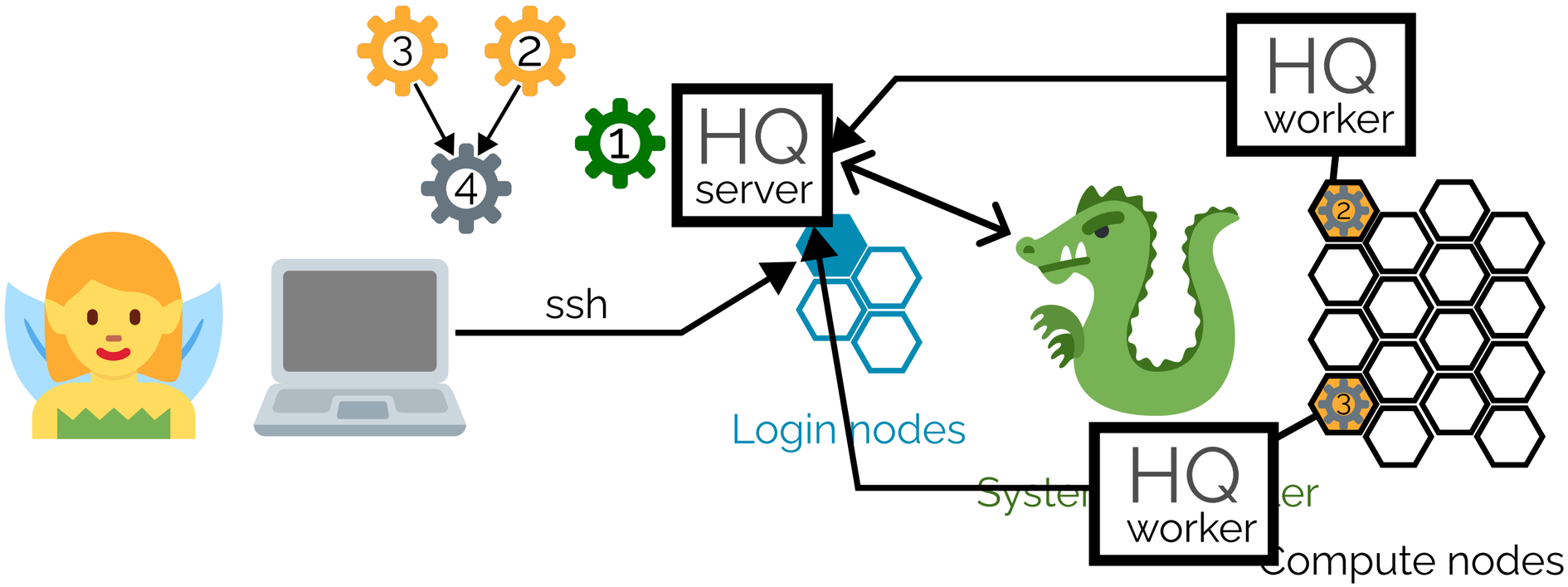

```
$ hq job list
+----+----------------+----------+-------+
| ID | Name           | State    | Tasks |
+----+----------------+----------+-------+
| 1  | my-computation | FINISHED | 1     |
+----+----------------+----------+-------+
| 2  | my-workflow    | RUNNING  | 3     |
+----+----------------+----------+-------+
```



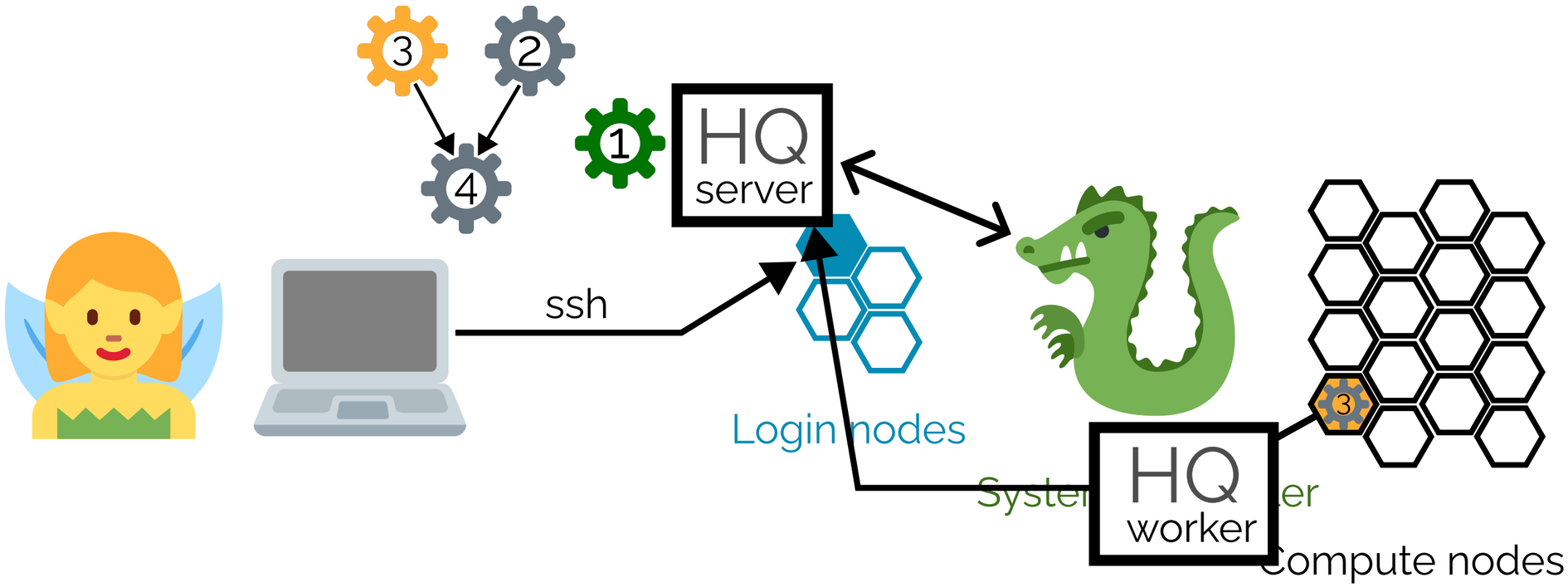

```
$ hq job list
+----+----------------+----------+-------+
| ID | Name           | State    | Tasks |
+----+----------------+----------+-------+
| 1  | my-computation | FINISHED | 1     |
+----+----------------+----------+-------+
| 2  | my-workflow    | RUNNING  | 3     |
+----+----------------+----------+-------+
```




```
$ hq job list
+----+----------------+----------+-------+
| ID | Name           | State    | Tasks |
+----+----------------+----------+-------+
| 1  | my-computation | FINISHED | 1     |
+----+----------------+----------+-------+
| 2  | my-workflow    | RUNNING  | 3     |
+----+----------------+----------+-------+
```




```
$ hq job list
+----+----------------+----------+-------+
| ID | Name           | State    | Tasks |
+----+----------------+----------+-------+
| 1  | my-computation | FINISHED | 1     |
+----+----------------+----------+-------+
| 2  | my-workflow    | RUNNING  | 3     |
+----+----------------+----------+-------+
```




```
$ hq job list
+----+----------------+----------+-------+
| ID | Name           | State    | Tasks |
+----+----------------+----------+-------+
| 1  | my-computation | FINISHED | 1     |
+----+----------------+----------+-------+
| 2  | my-workflow    | RUNNING  | 3     |
+----+----------------+----------+-------+
```

HyperQueue approach:

HyperQueue approach:

What to compute (tasks) +

HyperQueue approach:

What to compute (tasks) +
Where to compute it (nodes)

HyperQueue approach:

What to compute (tasks) +
Where to compute it (nodes)
=> disentangled

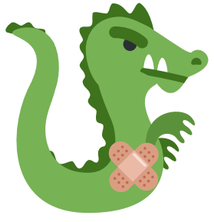
HyperQueue approach:

What to compute (tasks) +
Where to compute it (nodes)
=> disentangled

Load balancing across all available resources

Why we cannot "fix" Slurm/PBS?

Why we cannot "fix" Slurm/PBS?



Fairness

Why we cannot "fix" Slurm/PBS?

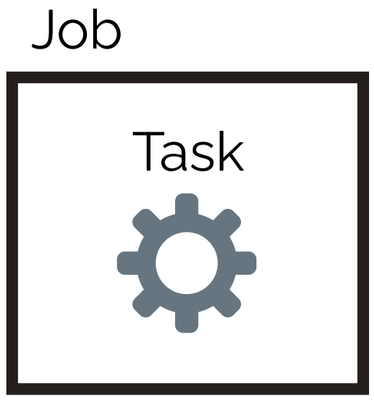
Fairness

Security

Task arrays

```
$ hq submit ./my-computation
```

Task arrays

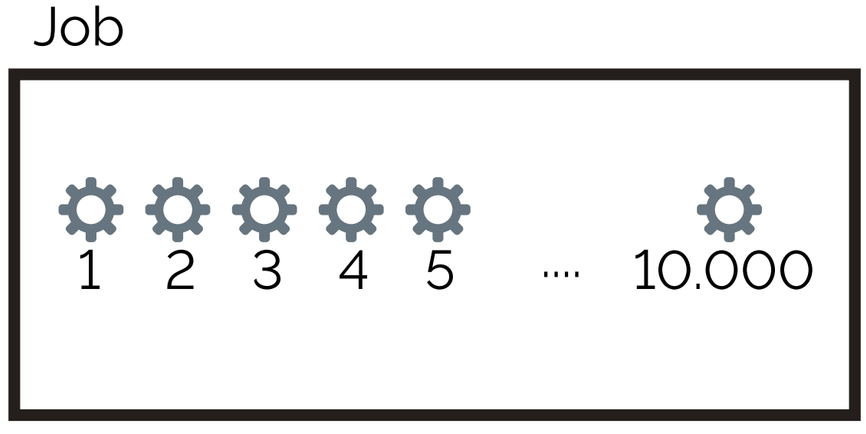



```
$ hq submit ./my-computation
```

Task arrays

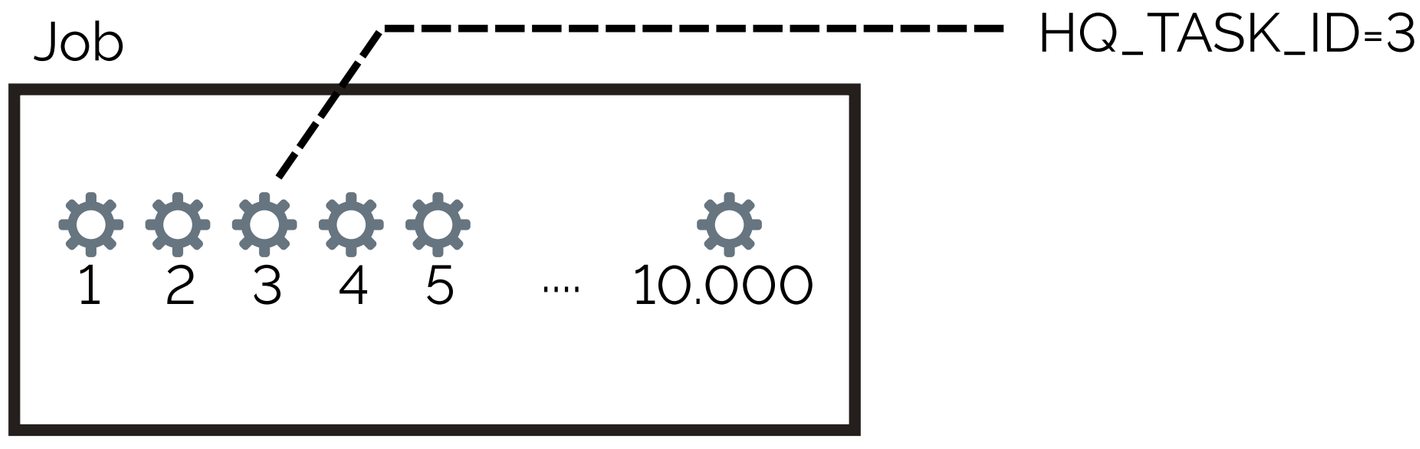
```
$ hq submit --array=1-10_000 ./my-computation
Job submitted successfully, job ID: 1
```

Task arrays




```
$ hq submit --array=1-10_000 ./my-computation
Job submitted successfully, job ID: 1
```

Task arrays




```
$ hq submit --array=1-10_000 ./my-computation
Job submitted successfully, job ID: 1
```

Task arrays

```
$ hq submit --array=1-10_000 ./my-computation
Job submitted successfully, job ID: 1
$ hq progress 1
[###.............................] 0/1 jobs, 232/10000 tasks (4 RUNNING, 232 FINISHED)
```

Task arrays

myfile.txt

```
lineA
lineB
lineC
lineD
lineE
```

```
$ hq submit --array=1-10_000 ./my-computation
Job submitted successfully, job ID: 1
$ hq progress 1
[###.............................] 0/1 jobs, 232/10000 tasks (4 RUNNING, 232 FINISHED)
```

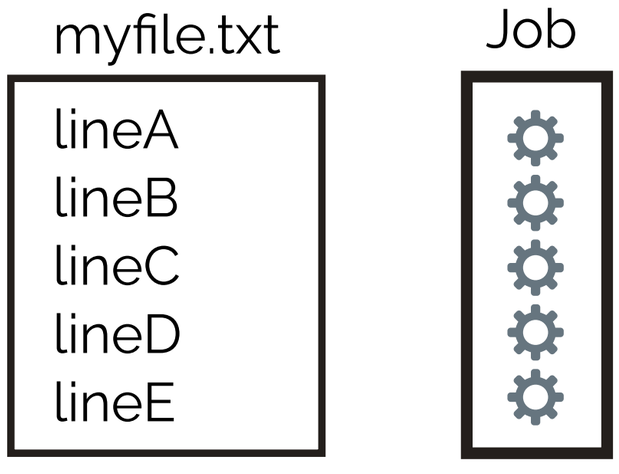
Task arrays

myfile.txt

```
lineA
lineB
lineC
lineD
lineE
```

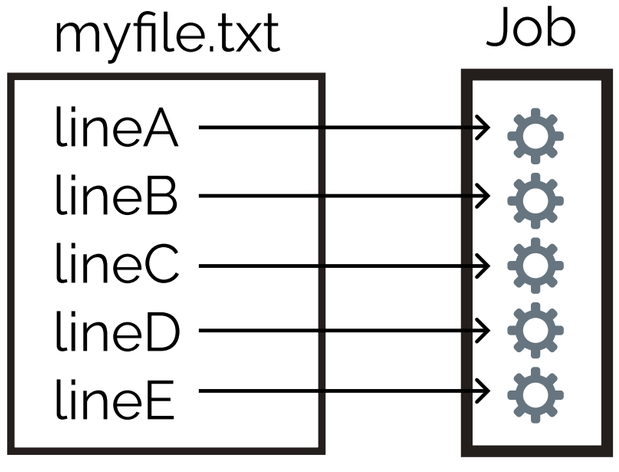
```
$ hq submit --array=1-10_000 ./my-computation
Job submitted successfully, job ID: 1
$ hq progress 1
[###.............................] 0/1 jobs, 232/10000 tasks (4 RUNNING, 232 FINISHED)
$ hq submit --each-line=myfile.txt ./my-computation
```

Task arrays



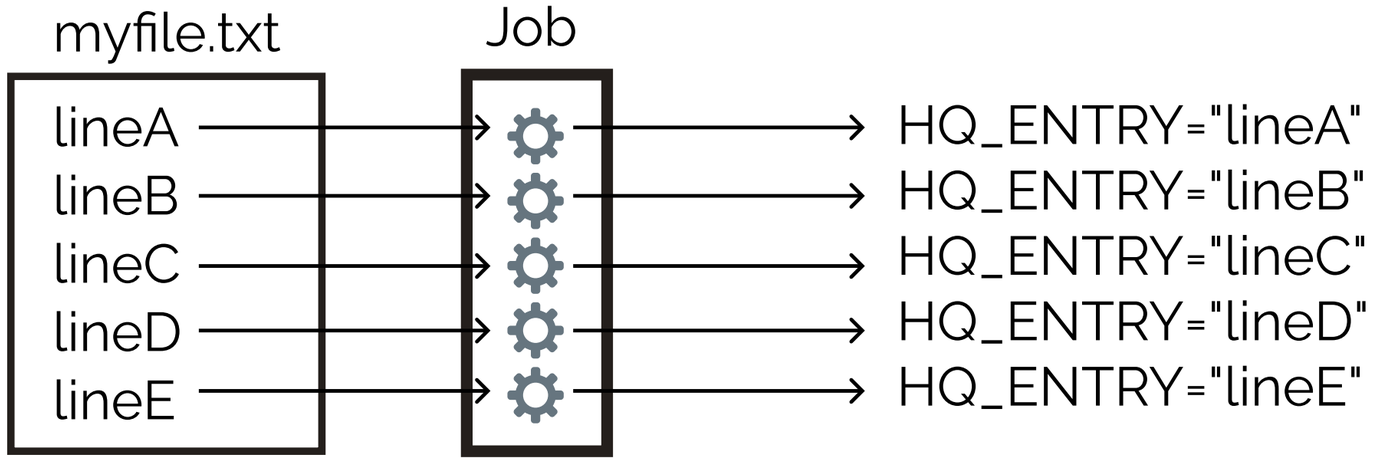

```
$ hq submit --array=1-10_000 ./my-computation
Job submitted successfully, job ID: 1
$ hq progress 1
[###.............................] 0/1 jobs, 232/10000 tasks (4 RUNNING, 232 FINISHED)
$ hq submit --each-line=myfile.txt ./my-computation
```

Task arrays



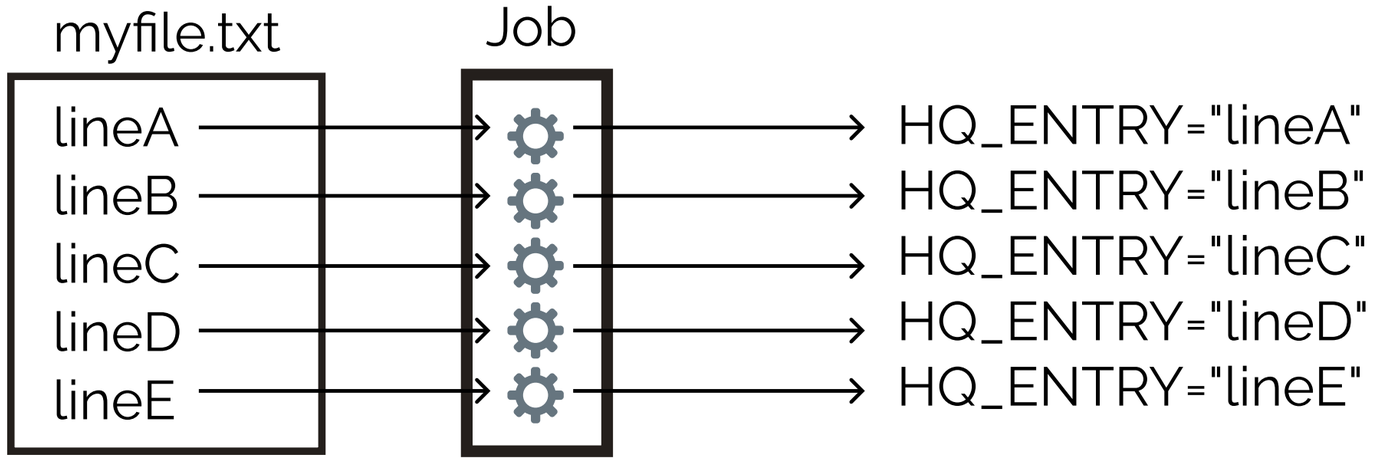

```
$ hq submit --array=1-10_000 ./my-computation
Job submitted successfully, job ID: 1
$ hq progress 1
[###.............................] 0/1 jobs, 232/10000 tasks (4 RUNNING, 232 FINISHED)
$ hq submit --each-line=myfile.txt ./my-computation
```

Task arrays



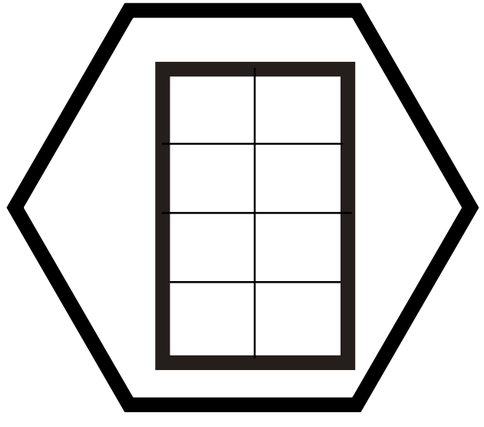

```
$ hq submit --array=1-10_000 ./my-computation
Job submitted successfully, job ID: 1
$ hq progress 1
[###.............................] 0/1 jobs, 232/10000 tasks (4 RUNNING, 232 FINISHED)
$ hq submit --each-line=myfile.txt ./my-computation
```

Task arrays

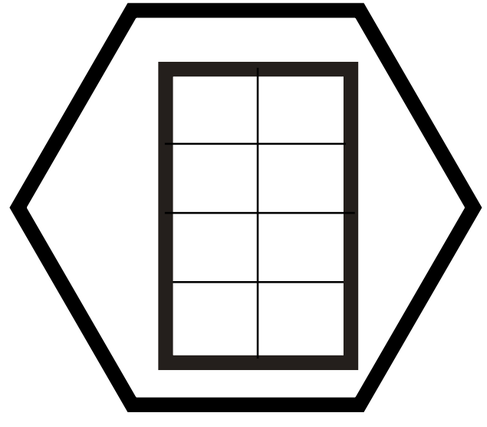



```
$ hq submit --array=1-10_000 ./my-computation
Job submitted successfully, job ID: 1
$ hq progress 1
[###.............................] 0/1 jobs, 232/10000 tasks (4 RUNNING, 232 FINISHED)
$ hq submit --each-line=myfile.txt ./my-computation
$ hq submit --from-json=items.json ./my-computation
```

CPU resource requirements

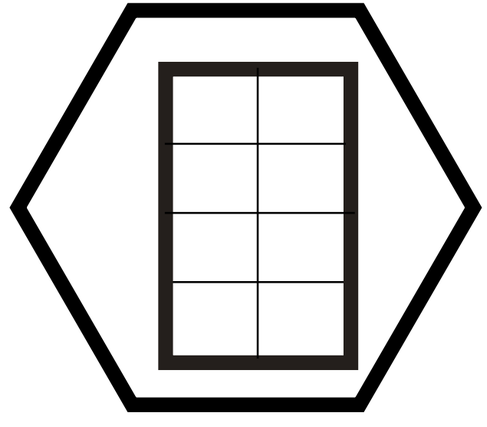


CPU resource requirements



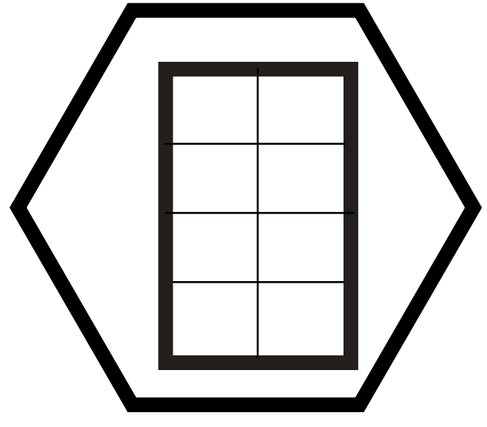
```
$ hq submit ./my-computation
```

CPU resource requirements



```
$ hq submit --cpus=<NUMBER_OF_CPUS> ./my-computation
```

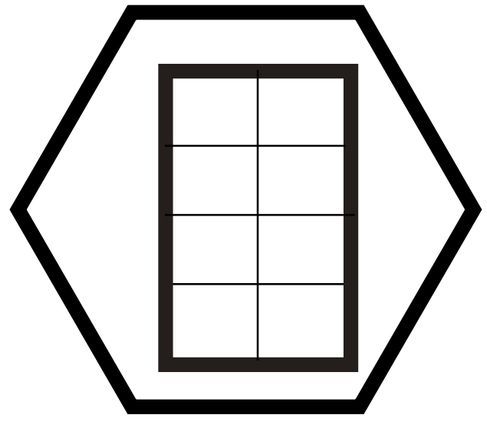
CPU resource requirements



```
$ hq submit --cpus=4 ./my-computation
```

CPU resource requirements

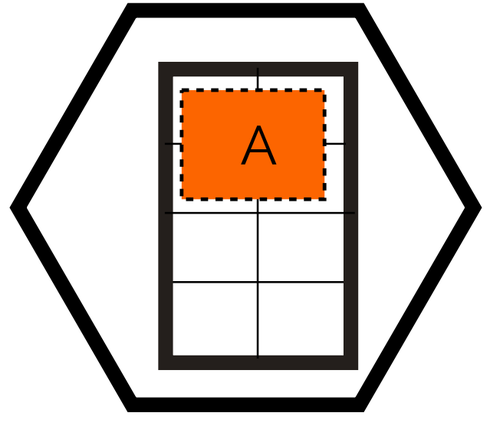
cpus=4



```
$ hq submit --cpus=4 ./my-computation
```

CPU resource requirements

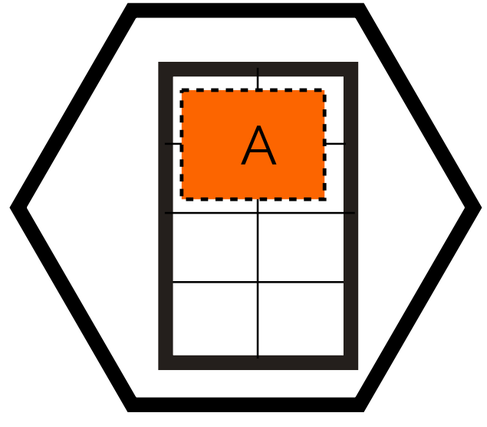
cpus=4




```
$ hq submit --cpus=4 ./my-computation
```

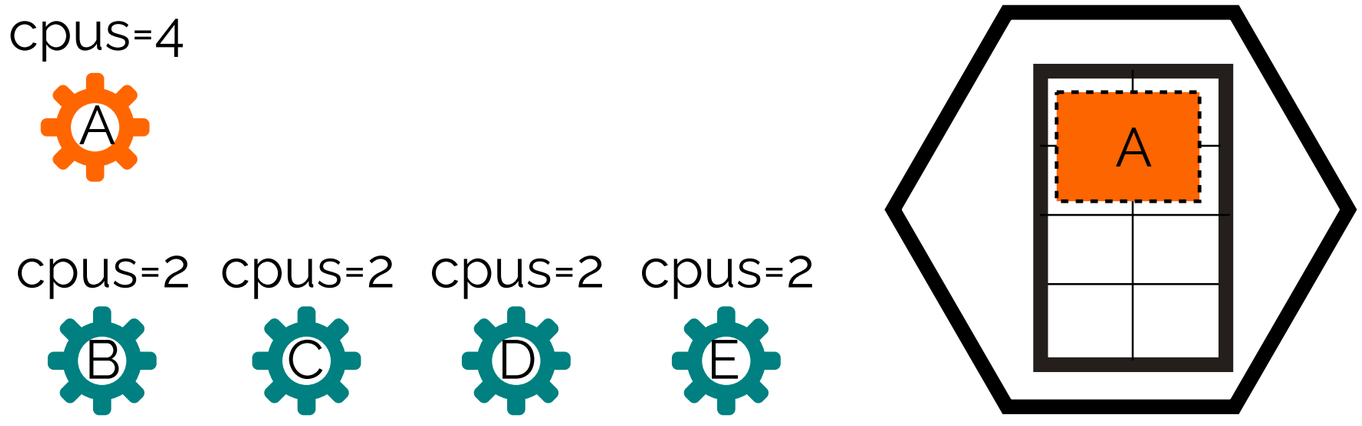
CPU resource requirements

cpus=4



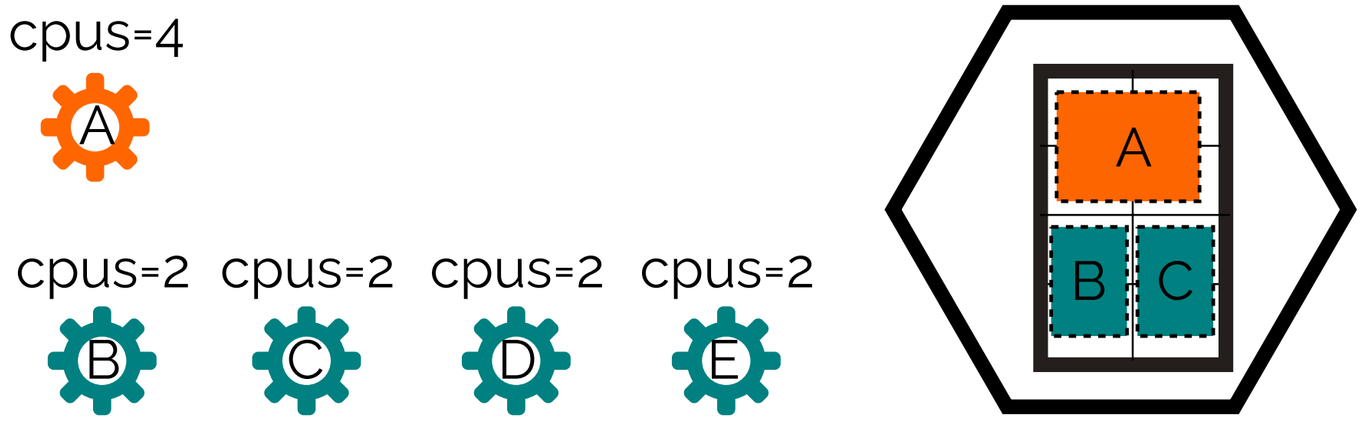

```
$ hq submit --cpus=4 ./my-computation
$ hq submit --array=1-4 --cpus=2 ./my-computation
```

CPU resource requirements



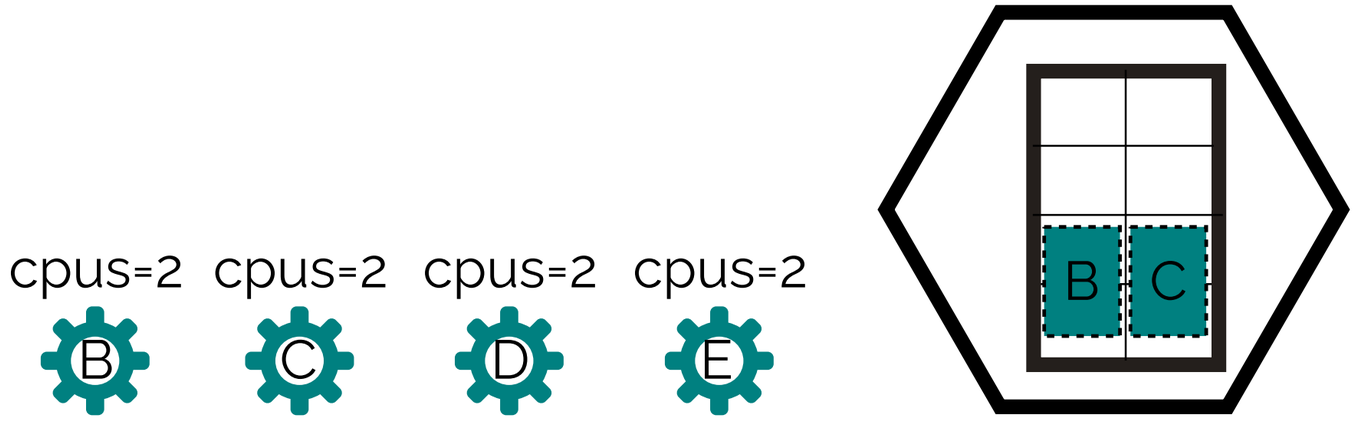

```
$ hq submit --cpus=4 ./my-computation
$ hq submit --array=1-4 --cpus=2 ./my-computation
```

CPU resource requirements



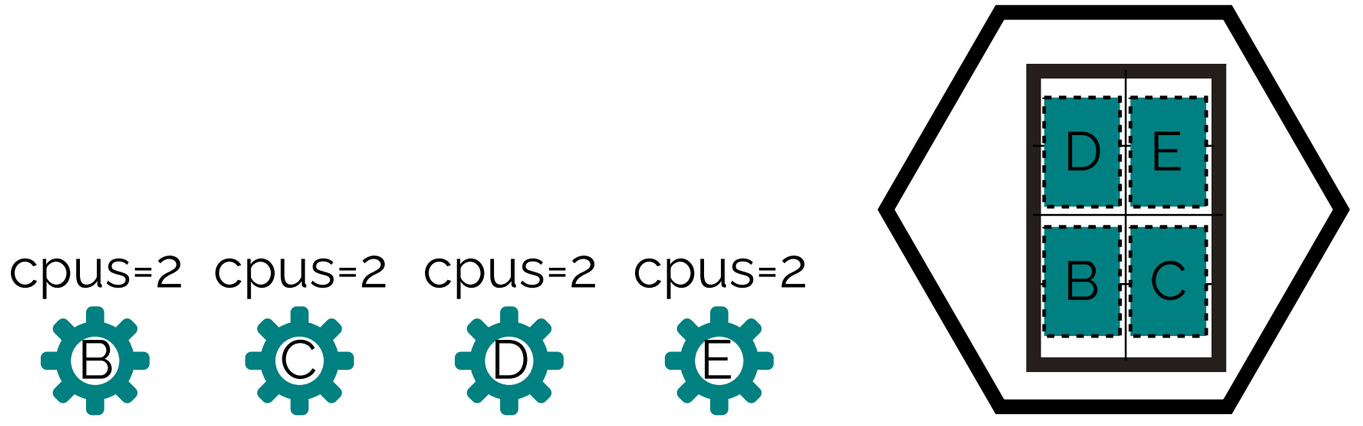

```
$ hq submit --cpus=4 ./my-computation
$ hq submit --array=1-4 --cpus=2 ./my-computation
```

CPU resource requirements




```
$ hq submit --cpus=4 ./my-computation
$ hq submit --array=1-4 --cpus=2 ./my-computation
```

CPU resource requirements




```
$ hq submit --cpus=4 ./my-computation
$ hq submit --array=1-4 --cpus=2 ./my-computation
```

General resource requirements

General resource requirements

```
$ hq submit --resource NAME=AMOUNT ./my-computation
```

General resource requirements

```
$ hq submit --resource foo=2 ./my-computation
```

General resource requirements

```
$ hq worker start --resource foo=[5,6,7,8]
```

```
$ hq submit --resource foo=2 ./my-computation
```

General resource requirements

```
$ hq worker start --resource foo=[5,6,7,8]
```

HQ_RESOURCE_foo="5,8"

```
$ hq submit --resource foo=2 ./my-computation
```

General resource requirements

```
$ hq submit --resource foo=2 ./my-computation
$ hq submit --resource gpus/nvidia=4 ./my-computation
```

General resource requirements

HQ_RESOURCE_gpus/nvidia="0,1,3,4"
CUDA_VISIBLE_DEVICES="0,1,3,4"

```
$ hq submit --resource foo=2 ./my-computation
$ hq submit --resource gpus/nvidia=4 ./my-computation
```

Automatic detection of worker resources

Automatic detection of worker resources

- CPUs (including NUMA sockets)

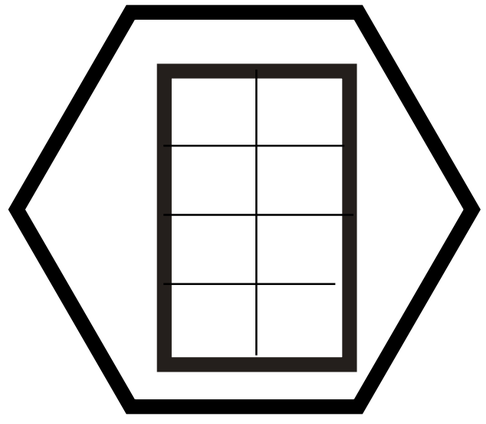
Automatic detection of worker resources

- CPUs (including NUMA sockets)
- Memory

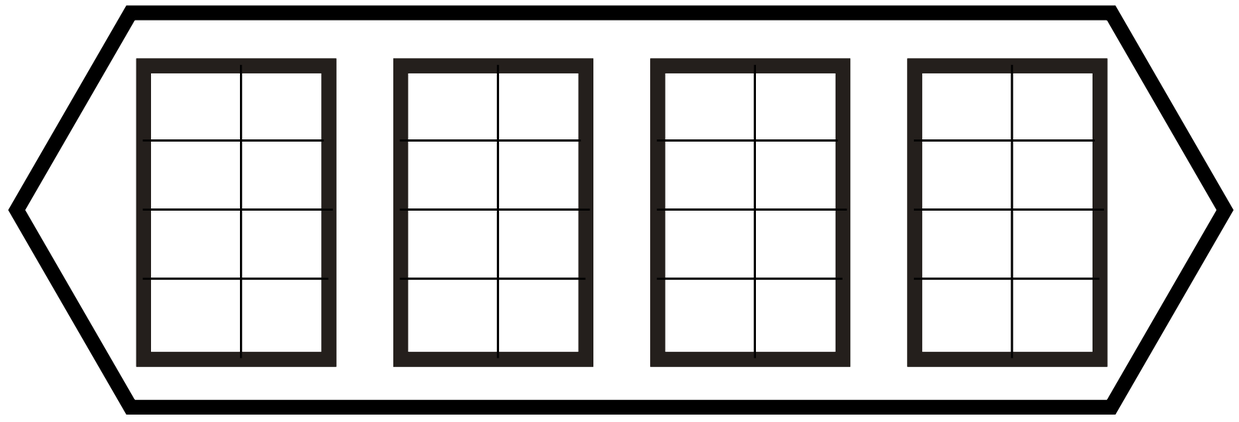
Automatic detection of worker resources

- CPUs (including NUMA sockets)
- Memory
- Nvidia/AMD GPUs

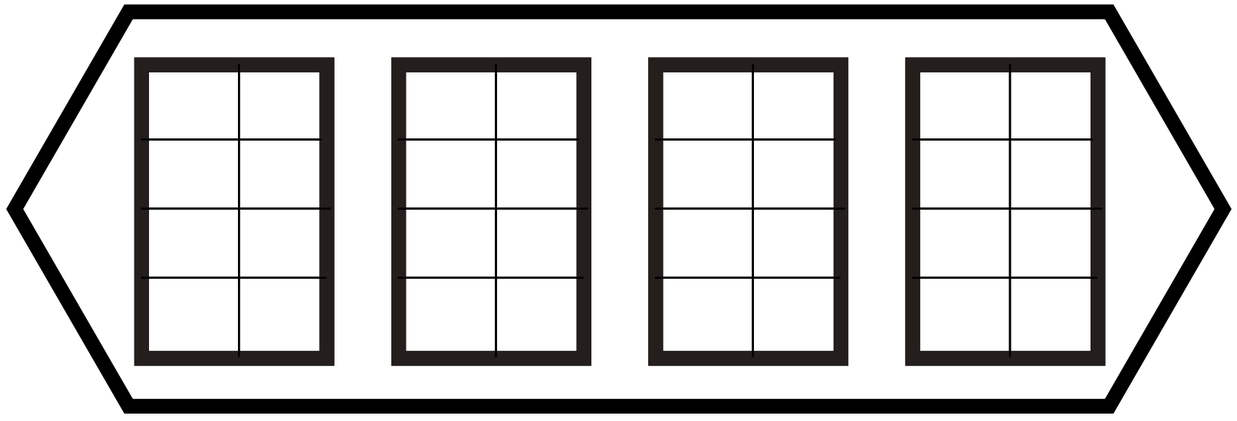
NUMA handling



NUMA handling

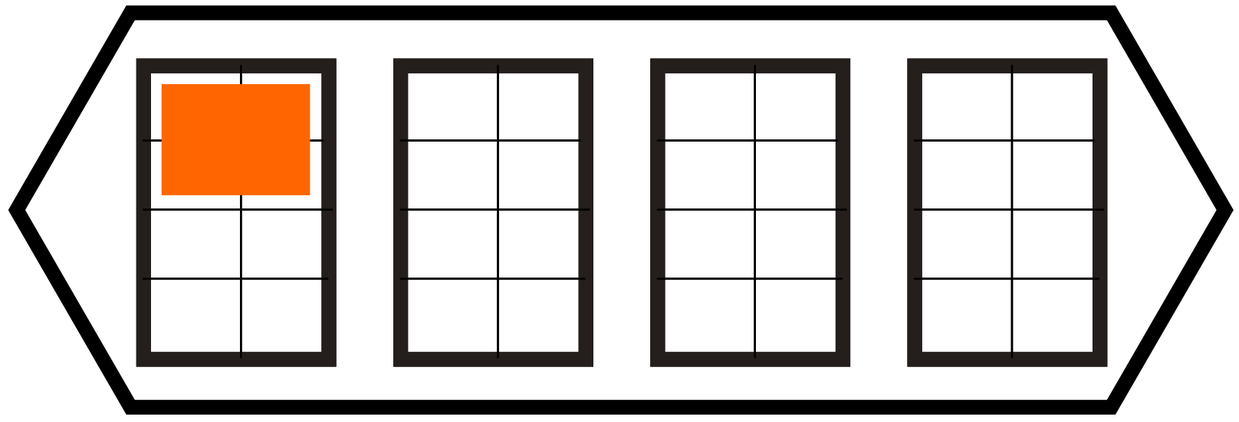


NUMA handling



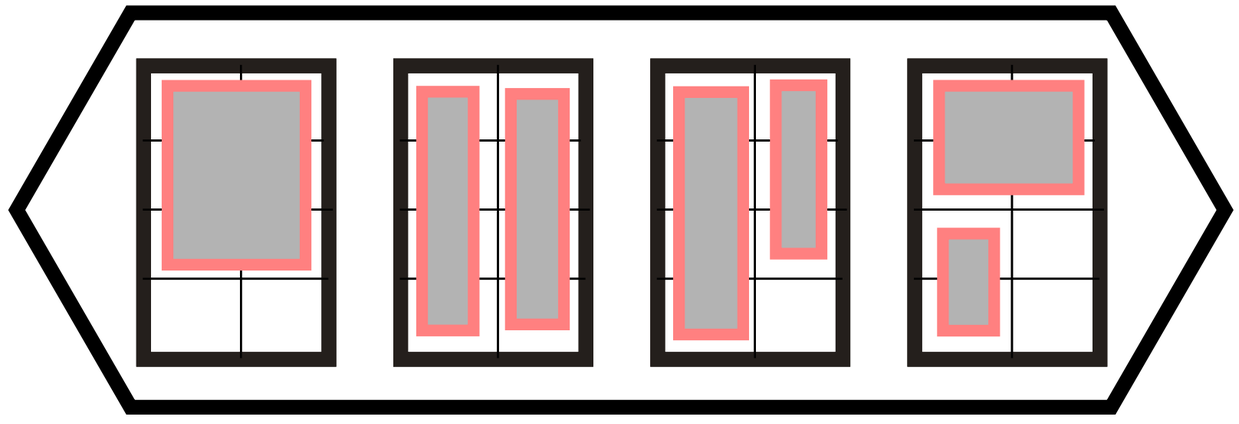
```
$ hq submit --cpus=4 ./my-computation
```

NUMA handling



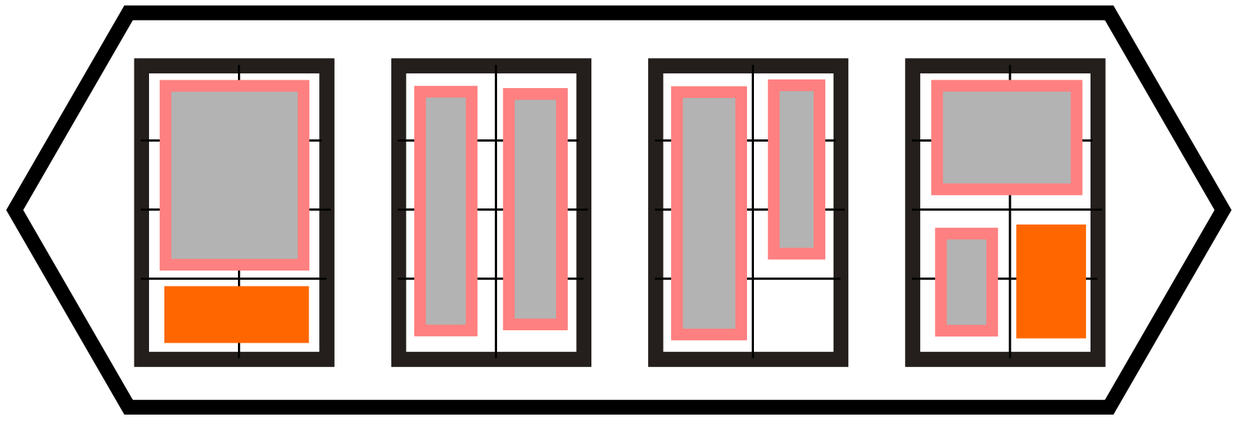
```
$ hq submit --cpus=4 ./my-computation
```

NUMA handling



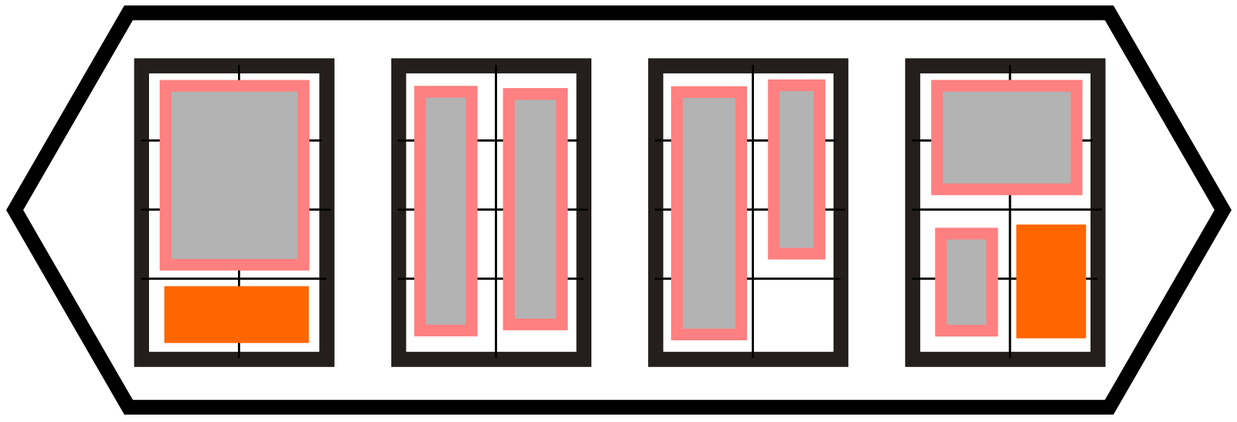
```
$ hq submit --cpus=4 ./my-computation
```

NUMA handling



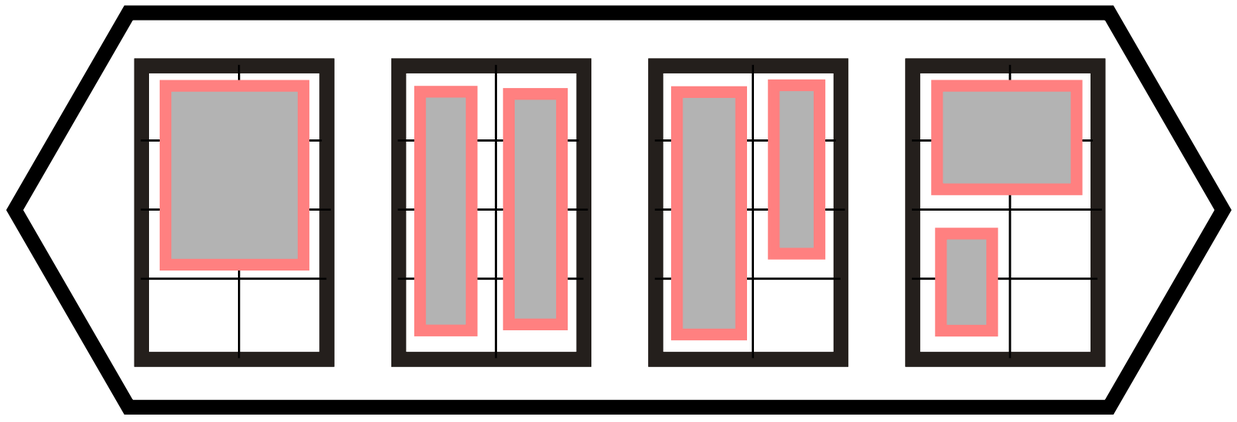
```
$ hq submit --cpus=4 ./my-computation
```

NUMA handling



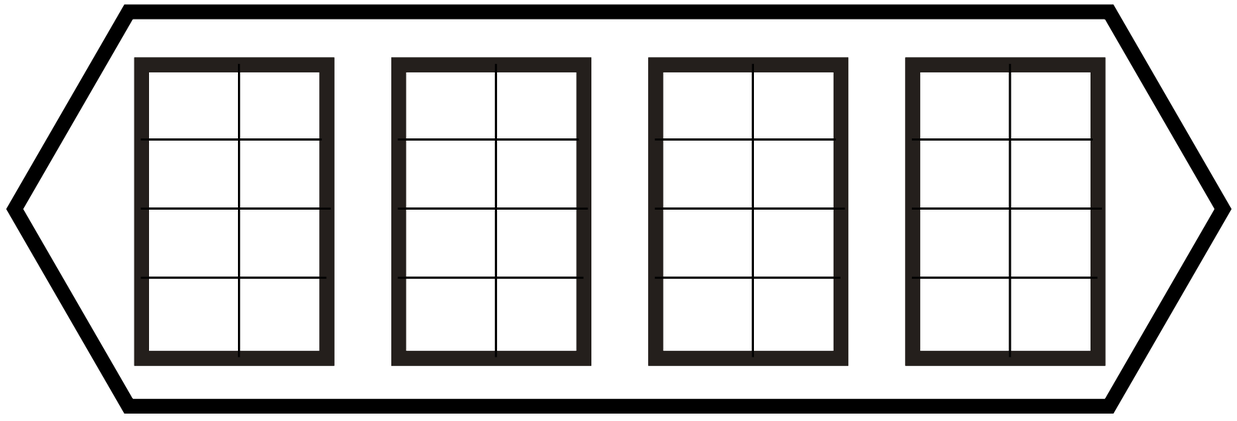
```
$ hq submit --cpus=4 ./my-computation
$ hq submit --cpus="4 compact" ./my-computation
```

NUMA handling



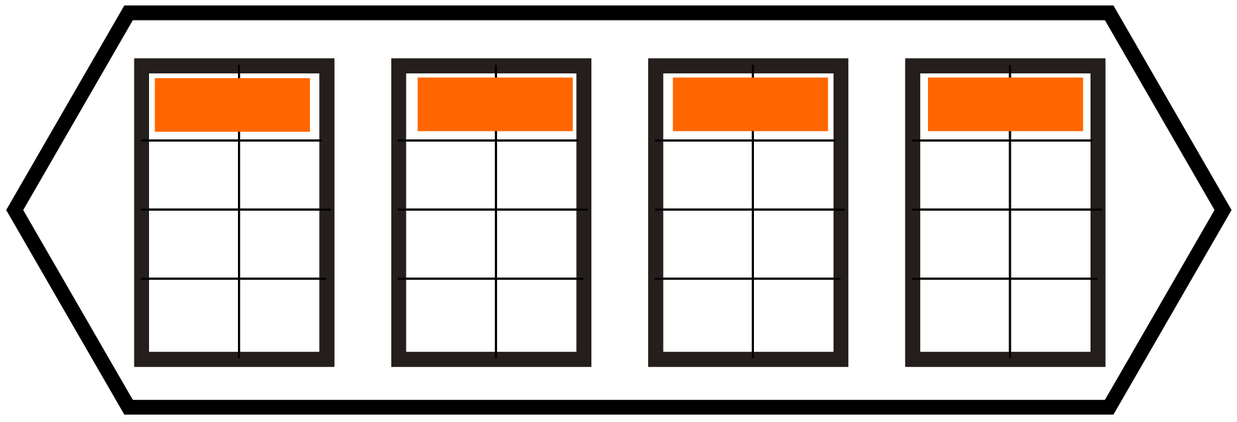
```
$ hq submit --cpus=4 ./my-computation
$ hq submit --cpus="4 compact" ./my-computation
$ hq submit --cpus="4 compact!" ./my-computation
```

NUMA handling



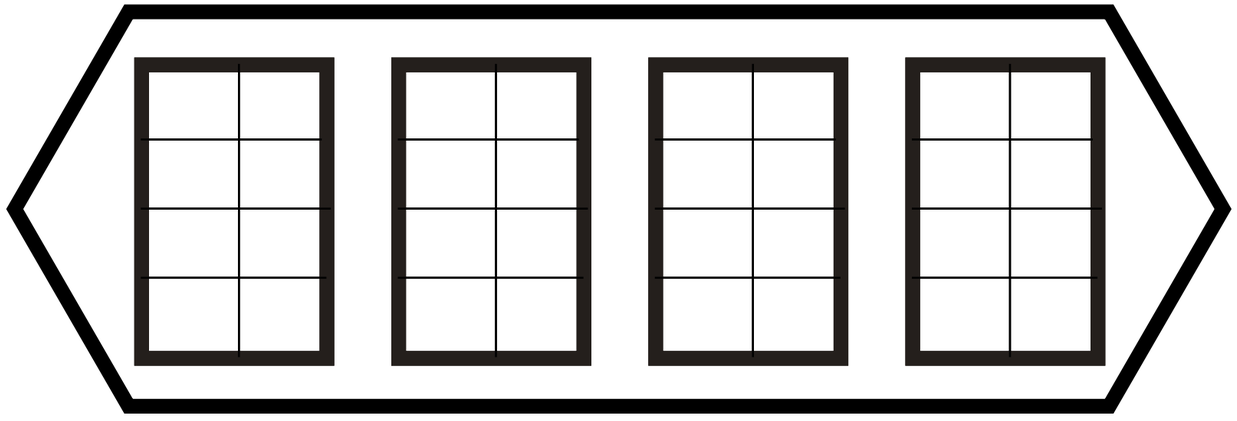
```
$ hq submit --cpus=4 ./my-computation
$ hq submit --cpus="4 compact" ./my-computation
$ hq submit --cpus="4 compact!" ./my-computation
$ hq submit --cpus="8 scatter" ./my-computation
```

NUMA handling



```
$ hq submit --cpus=4 ./my-computation
$ hq submit --cpus="4 compact" ./my-computation
$ hq submit --cpus="4 compact!" ./my-computation
$ hq submit --cpus="8 scatter" ./my-computation
```

NUMA handling



```
$ hq submit --cpus=4 ./my-computation
$ hq submit --cpus="4 compact" ./my-computation
$ hq submit --cpus="4 compact!" ./my-computation
$ hq submit --cpus="8 scatter" ./my-computation
$ hq submit --pin --cpus=... ./my-computation
```

Complex resource requirements

```
$ hq submit --time-limit=1h ./my-computation
```

Complex resource requirements

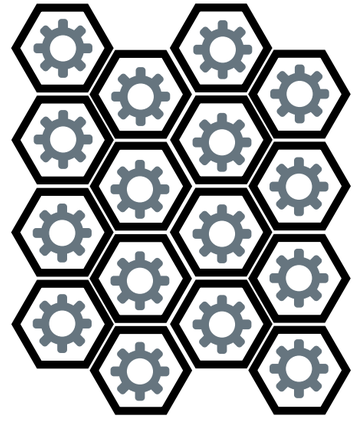
```
$ hq submit --time-limit=1h ./my-computation
$ hq submit --time-request=10m ./my-computation
```

Complex resource requirements

```
$ hq submit --time-limit=1h ./my-computation
$ hq submit --time-request=10m ./my-computation
$ hq submit --resource=gpus/nvidia=0.5 ./my-computation
```

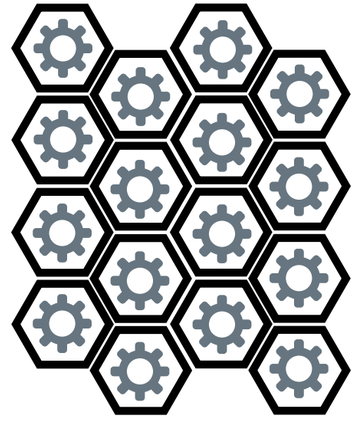
Multinode tasks

mpirun -np 16 my_computation



Multinode tasks

mpirun -np 16 my_computation



```
$ hq submit --nodes=16 ...
```

I/O streaming

```
$ hq submit --array 1-4 ./my-computation
```

I/O streaming

```
$ hq submit --array 1-4 ./my-computation
$ ls job-1
```

I/O streaming

```
$ hq submit --array 1-4 ./my-computation
$ ls job-1
1.stderr   2.stderr   3.stderr    4.stderr
1.stdout   2.stdout   3.stdout    4.stdout
```

I/O streaming

```
$ hq submit --array 1-4 ./my-computation
$ ls job-1
1.stderr   2.stderr   3.stderr    4.stderr
1.stdout   2.stdout   3.stdout    4.stdout
$ hq submit --array 1-1000_000 ./my-computation
```

I/O streaming

```
$ hq submit --array 1-4 ./my-computation
$ ls job-1
1.stderr   2.stderr   3.stderr    4.stderr
1.stdout   2.stdout   3.stdout    4.stdout
$ hq submit --array 1-1000_000 ./my-computation
$ ls job-2
```

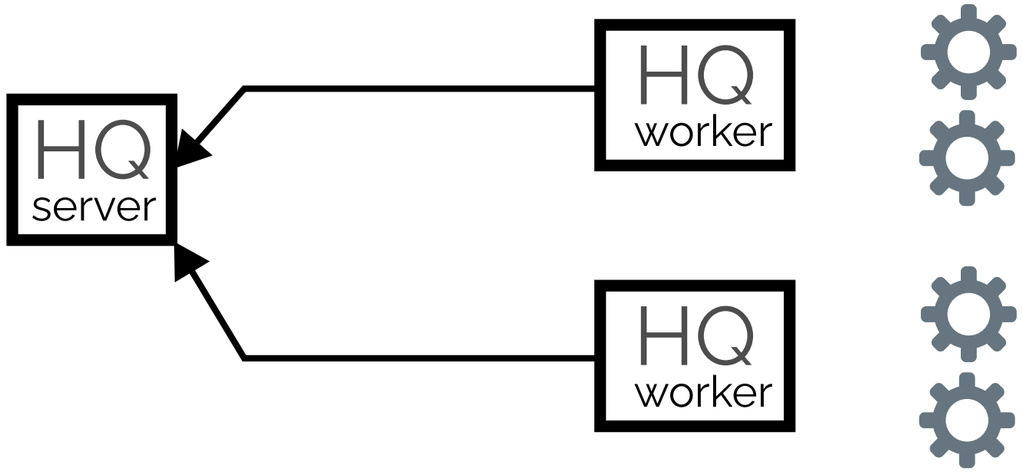
I/O streaming

```
$ hq submit --array 1-4 ./my-computation
$ ls job-1
1.stderr   2.stderr   3.stderr    4.stderr
1.stdout   2.stdout   3.stdout    4.stdout
$ hq submit --array 1-1000_000 ./my-computation
$ ls job-2
```

I/O streaming

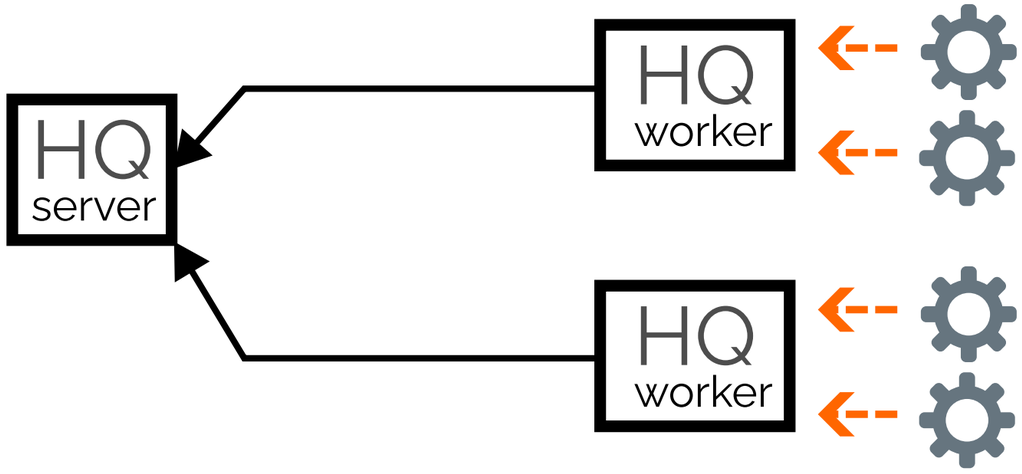
```
$ hq submit --array 1-4 --log=my.log ./my-computation
```

I/O streaming



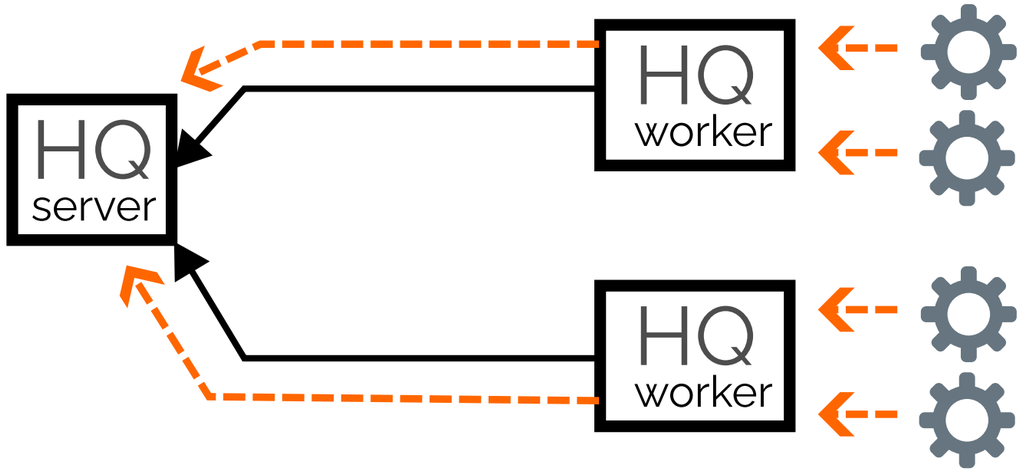

```
$ hq submit --array 1-4 --log=my.log ./my-computation
```

I/O streaming



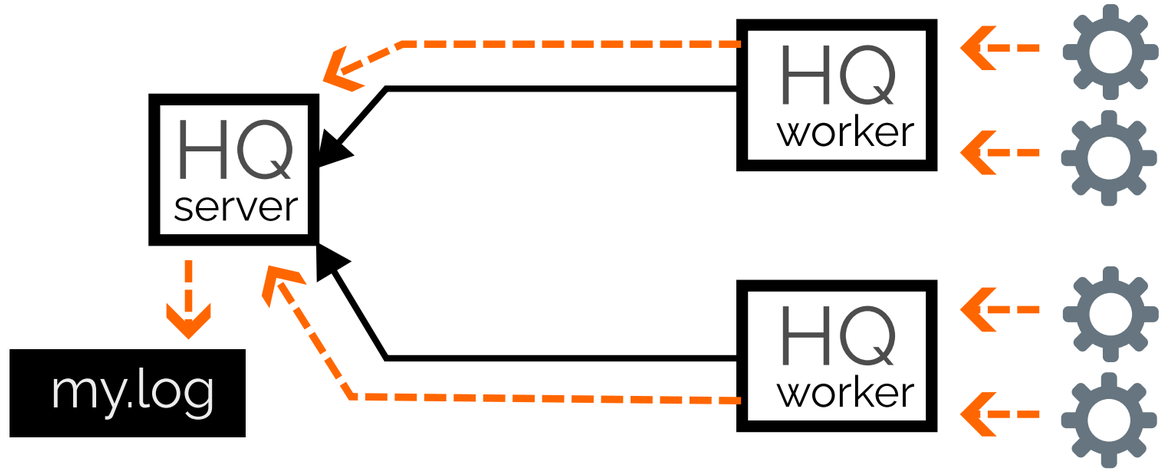

```
$ hq submit --array 1-4 --log=my.log ./my-computation
```

I/O streaming



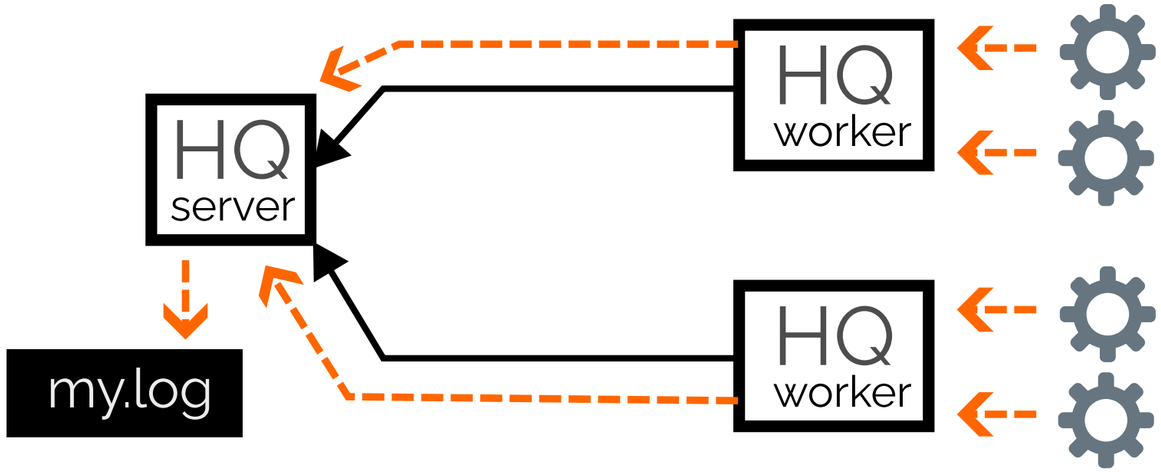

```
$ hq submit --array 1-4 --log=my.log ./my-computation
```

I/O streaming



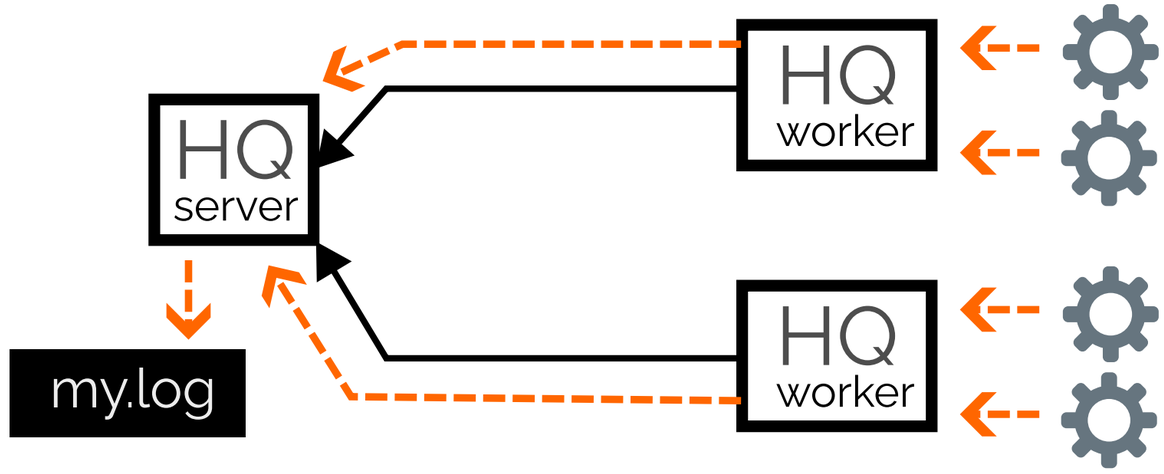

```
$ hq submit --array 1-4 --log=my.log ./my-computation
```

I/O streaming



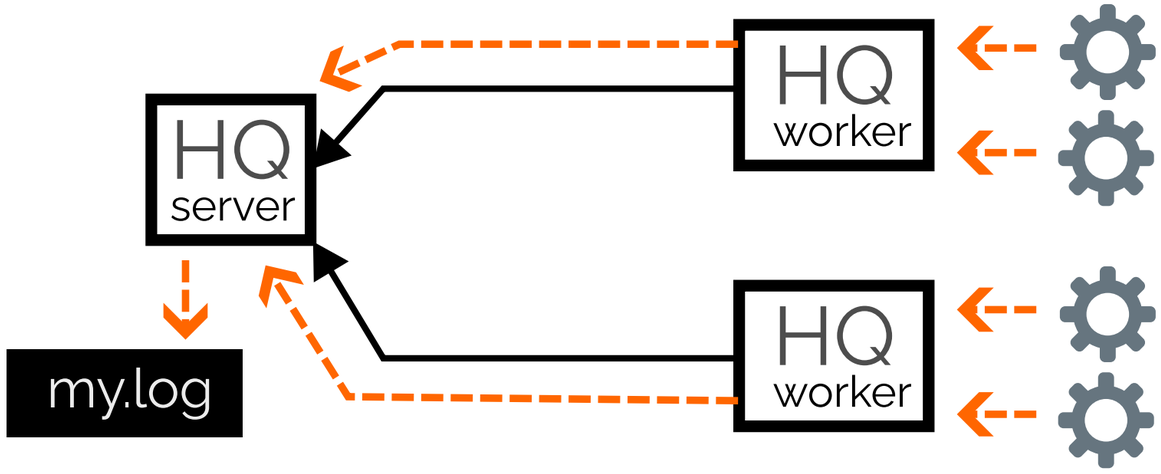

```
$ hq log my.log show
```

I/O streaming




```
$ hq log my.log show
2:0> Computation started ...
1:0> Computation started ...
3:0> Computation started ...
4:0> Computation started ...
3:0> Result is 3.2
3: > stream closed
2:0> Result is 5.2
2: > stream closed
1:0> Result is 1.2
1: > stream closed
4:0> Result is 4.0
4: > stream closed
```

I/O streaming




```
$ hq log my.log cat stdout --task=1
Computation started ...
Result is 1.2
```

Python API

```python
import hyperqueue as hq

client = hq.Client()

job = hq.Job()
job.program(args=["./my-computation"])

job_id = client.submit(job)
client.wait_for_jobs([job_id])
```

Python API

```
import hyperqueue as hq

client = hq.Client()

job = hq.Job()
job.program(args=["./my-computation"])

job_id = client.submit(job)
client.wait_for_jobs([job_id])
```

Python API

```python
import hyperqueue as hq

client = hq.Client()

job = hq.Job()
job.program(args=["./my-computation"])

job_id = client.submit(job)
client.wait_for_jobs([job_id])
```

Python API

```python
import hyperqueue as hq

client = hq.Client()

job = hq.Job()
job.program(args=["./my-computation"])

job_id = client.submit(job)
client.wait_for_jobs([job_id])
```

Python API

```python
import hyperqueue as hq

client = hq.Client()

job = hq.Job()
job.program(args=["./my-computation"])

job_id = client.submit(job)
client.wait_for_jobs([job_id])
```

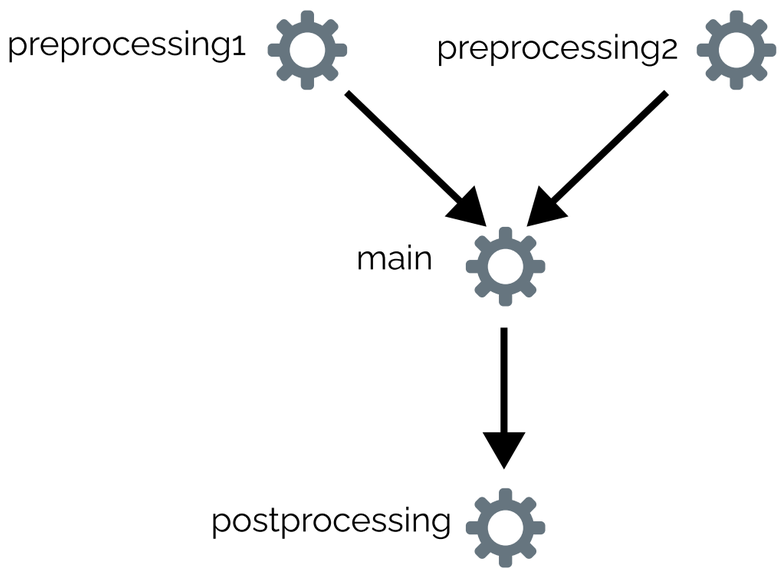
Python function tasks

```python
import hyperqueue as hq

def my_computation():
    import tensorflow as tf
    ...

client = hq.Client()

job = hq.Job()
resources = ResourceRequest(resources={"cpus": 32, "gpus/amd": 1})
job.function(my_computation, resources=resources)

job_id = client.submit(job)
client.wait_for_jobs([job_id])
```

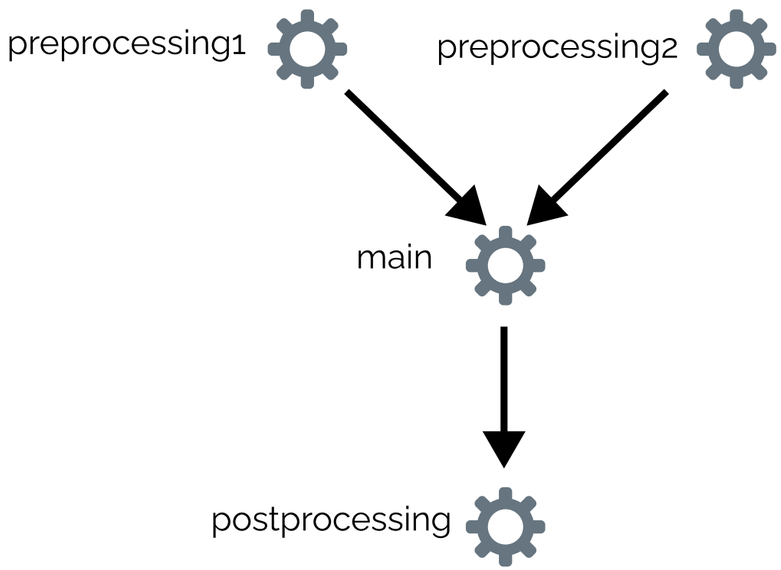
Python function tasks

```python
import hyperqueue as hq

def my_computation():
    import tensorflow as tf
    ...

client = hq.Client()

job = hq.Job()
resources = ResourceRequest(resources={"cpus": 32, "gpus/amd": 1})
job.function(my_computation, resources=resources)

job_id = client.submit(job)
client.wait_for_jobs([job_id])
```

Task dependencies

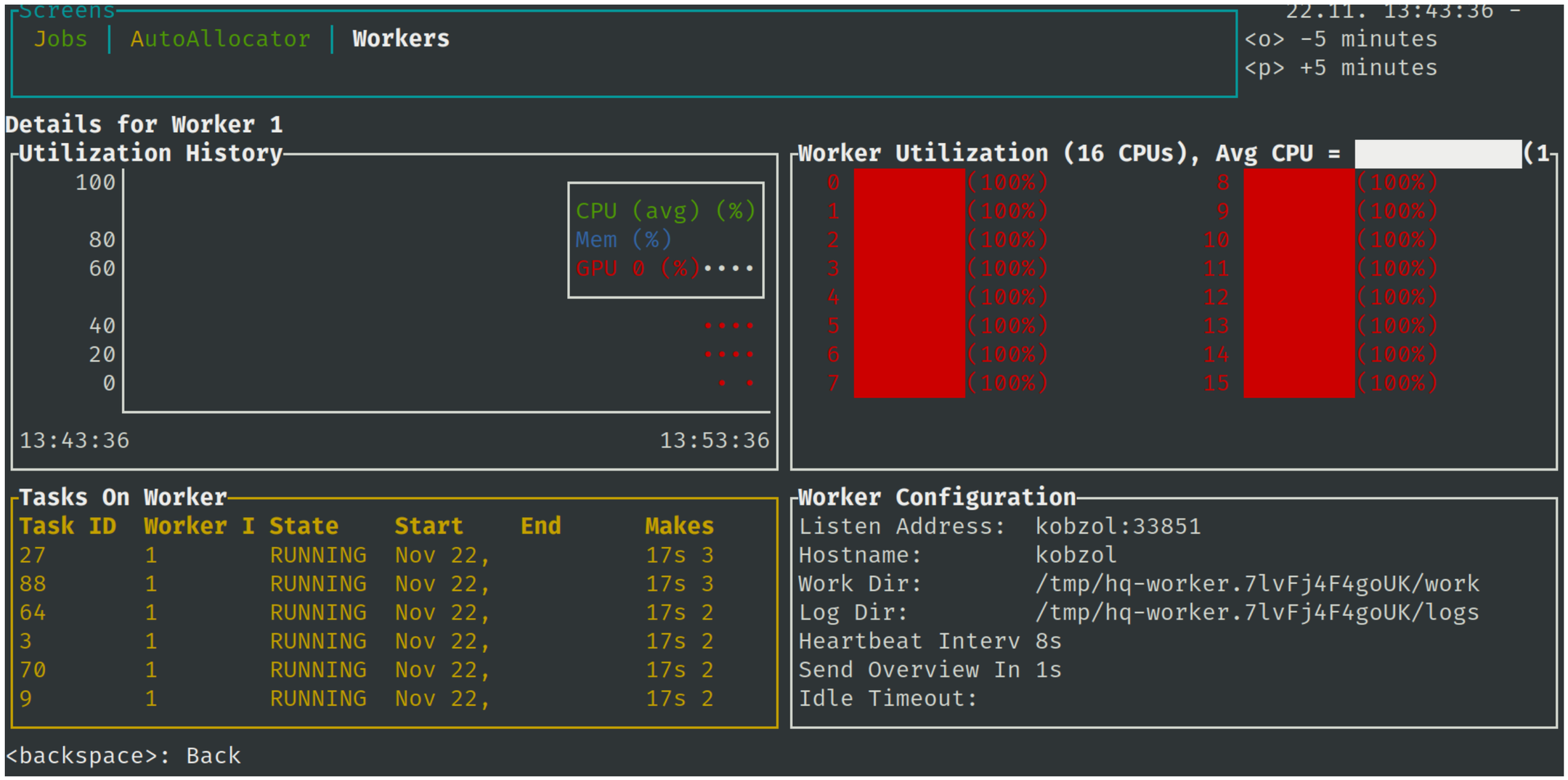


Task dependencies




```python
import hyperqueue as hq

client = hq.Client()

job = hq.Job()
p1 = job.program("preprocessing1")
p2 = job.program("preprocessing2")
main = job.function(my_function, deps=[p1, p2])
job.program("postprocessing", deps=[main])

job_id = client.submit(job)
client.wait_for_jobs([job_id])
```

Dashboard

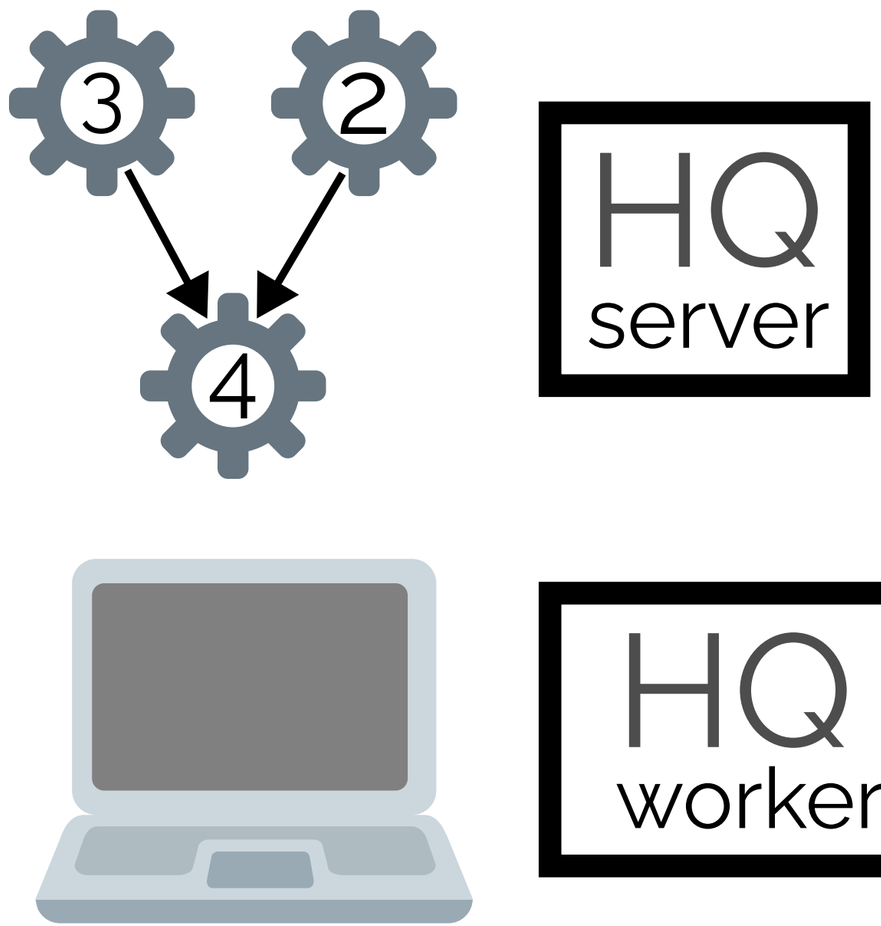

Screens
Jobs | AutoAllocator | Workers
22.11. 13:43:36 -
<o> -5 minutes
<p> +5 minutes
Details for Worker 1
Utilization History
100
80
60
40
20
0
CPU (avg) (%)
Mem (%)
GPU 0 (%)
13:43:36
13:53:36
Worker Utilization (16 CPUs), Avg CPU =
0 (100%)
1 (100%)
2 (100%)
3 (100%)
4 (100%)
5 (100%)
6 (100%)
7 (100%)
8 (100%)
9 (100%)
10 (100%)
11 (100%)
12 (100%)
13 (100%)
14 (100%)
15 (100%)
Tasks On Worker
Task ID Worker I State Start End Makes
27 1 RUNNING Nov 22, 17s 3
88 1 RUNNING Nov 22, 17s 3
64 1 RUNNING Nov 22, 17s 2
3 1 RUNNING Nov 22, 17s 2
70 1 RUNNING Nov 22, 17s 2
9 1 RUNNING Nov 22, 17s 2
Worker Configuration
Listen Address: kobzol:33851
Hostname: kobzol
Work Dir: /tmp/hq-worker.7lvFj4F4goUK/work
Log Dir: /tmp/hq-worker.7lvFj4F4goUK/logs
Heartbeat Interv 8s
Send Overview In 1s
Idle Timeout:
<backspace>: Back

Local prototyping

Local prototyping

Local prototyping

Local prototyping



LIGATE

EVEREST

DESIGN ENVIRONMENT
FOR EXTREME-SCALE BIG DATA ANALYTICS
ON HETEROGENEOUS PLATFORMS

ACROSS

VSB TECHNICAL UNIVERSITY OF OSTRAVA | IT4INNOVATIONS NATIONAL SUPERCOMPUTING CENTER

LUMI

AiiDA

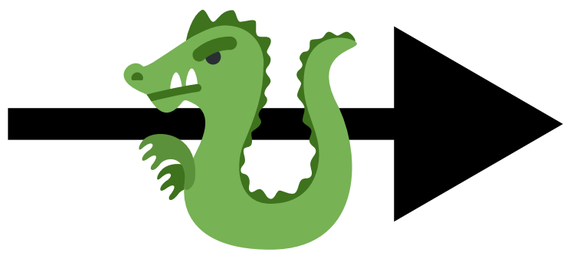
nextflow

CERN

HyperQueue

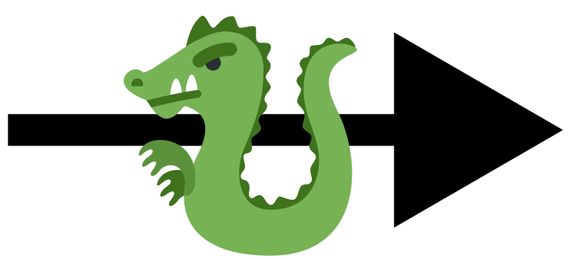
HyperQueue

Transparent over
PBS / SLURM

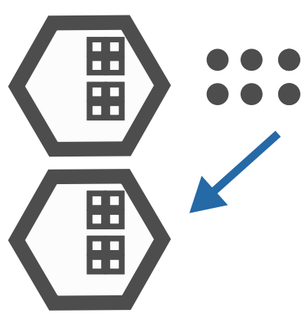


HyperQueue

Transparent over
PBS / SLURM

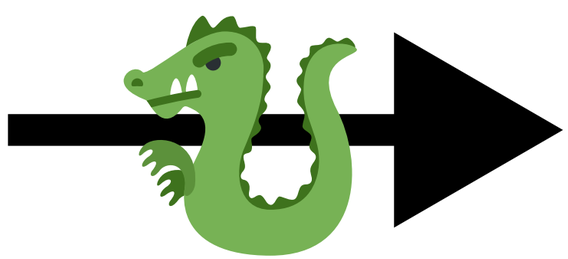


Dynamic
load balancing

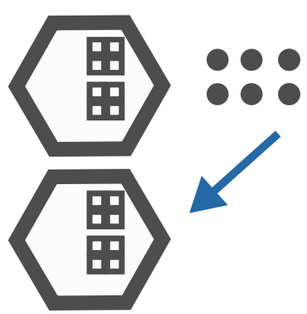


HyperQueue

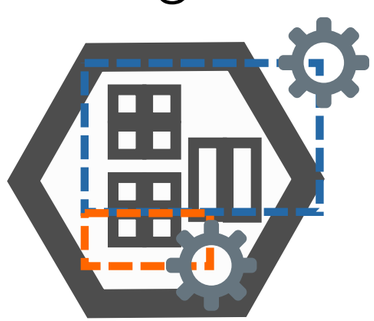
Transparent over PBS / SLURM



Dynamic load balancing

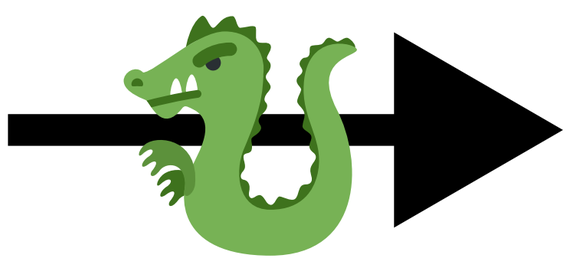


Resource management

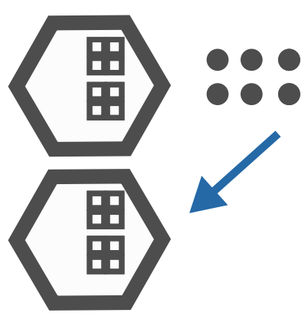


HyperQueue

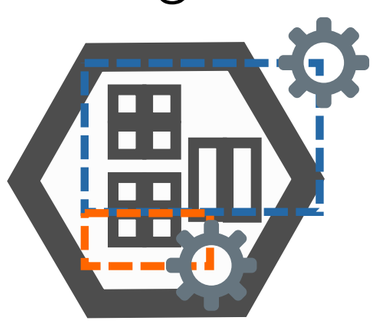
Transparent over
PBS / SLURM



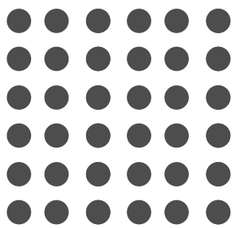
Dynamic
load balancing



Resource
management



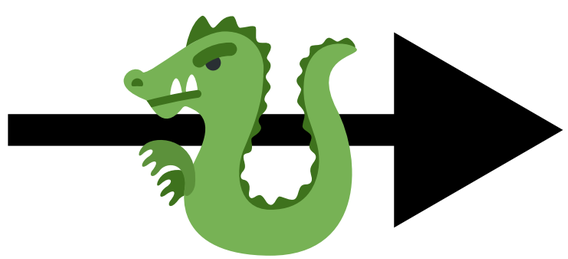
Scalable



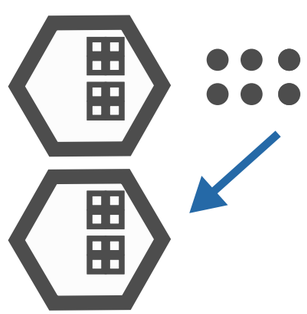
~100 µs
overhead per task

HyperQueue

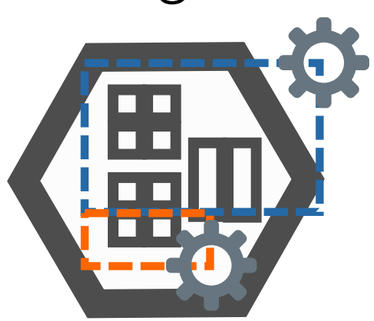
Transparent over
PBS / SLURM



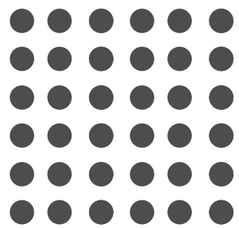
Dynamic
load balancing



Resource
management



Scalable

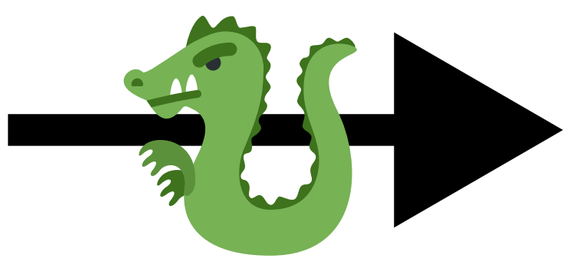


~100 µs
overhead per task

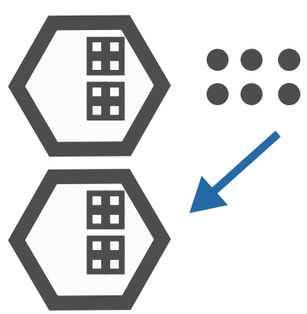
Easy to deploy

HyperQueue

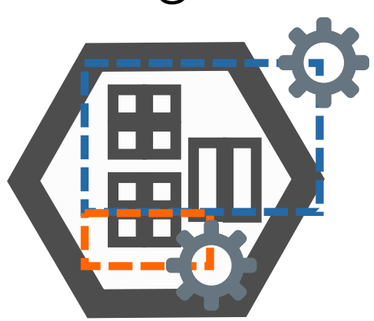
Transparent over
PBS / SLURM



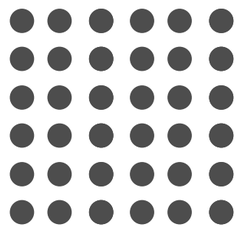
Dynamic
load balancing



Resource
management



Scalable



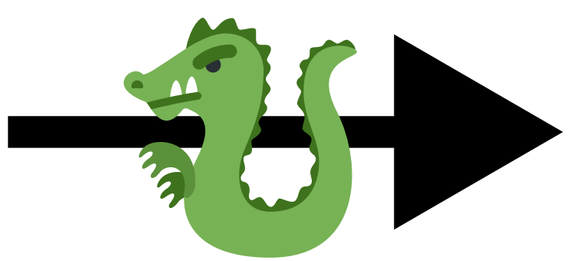
~100 µs
overhead per task

Easy to deploy

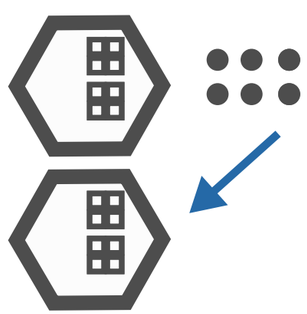
Secure

HyperQueue

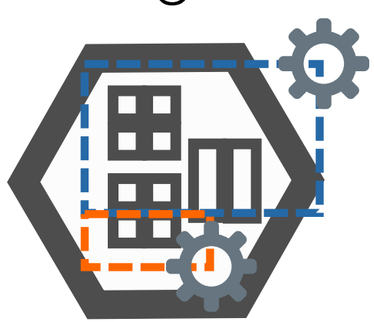
Transparent over
PBS / SLURM



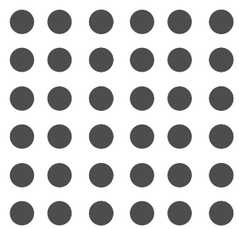
Dynamic
load balancing



Resource
management



Scalable



~100 µs
overhead per task

Easy to deploy

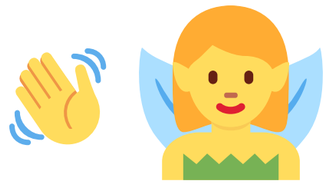
Open Source

Secure

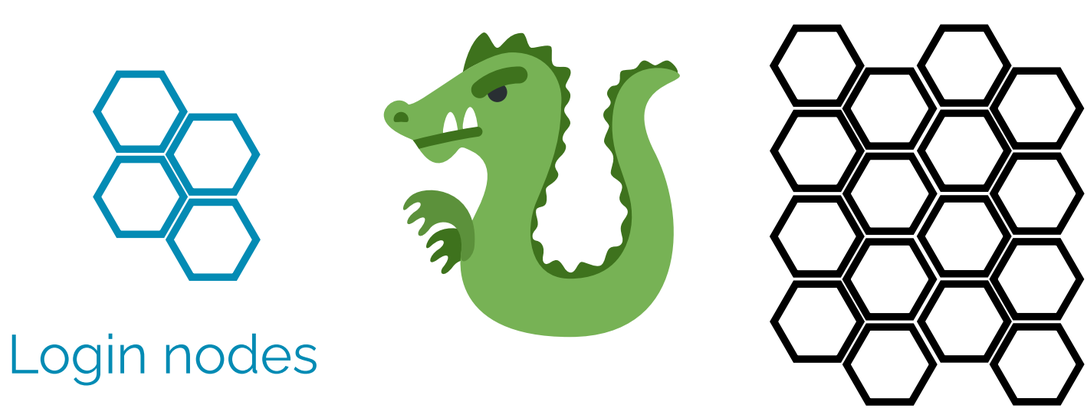
HyperQueue

https://github.com/it4innovations/hyperqueue

Thank you for your attention



Icons in slides:
https://twemoji.twitter.com



Login nodes

System scheduler

Compute nodes

Login nodes

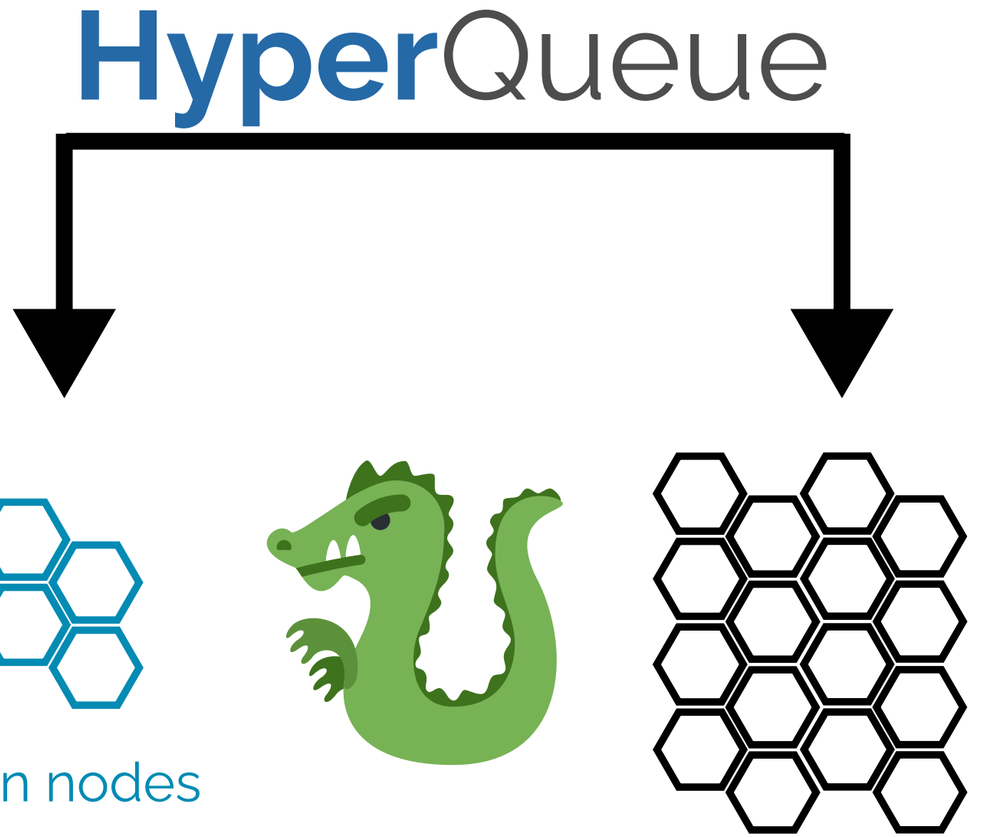
System scheduler

Compute nodes

Login nodes

System scheduler

Compute nodes


HyperQueue
Login nodes
System scheduler
Compute nodes