

Running Metview in batch

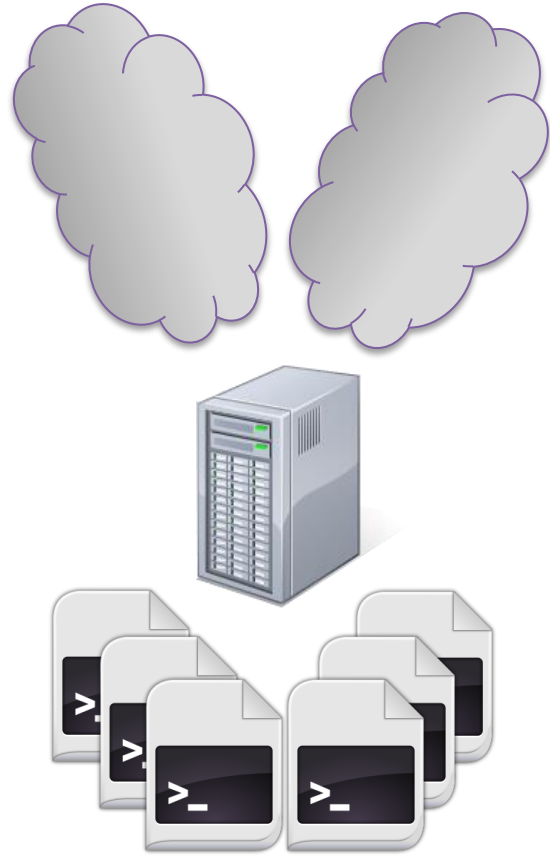
Submitting batch jobs

Xavi Abellan

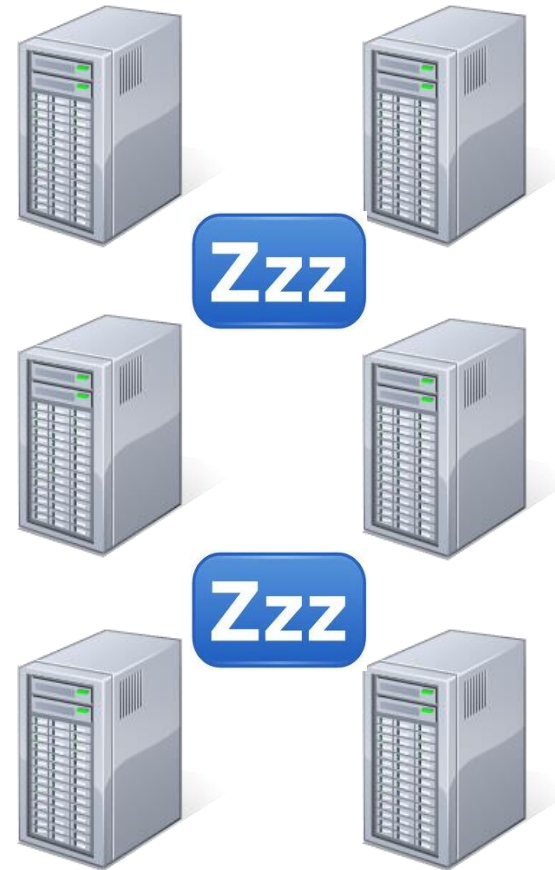
User Support

Xavier.Abellan@ecmwf.int

Interactive vs Batch

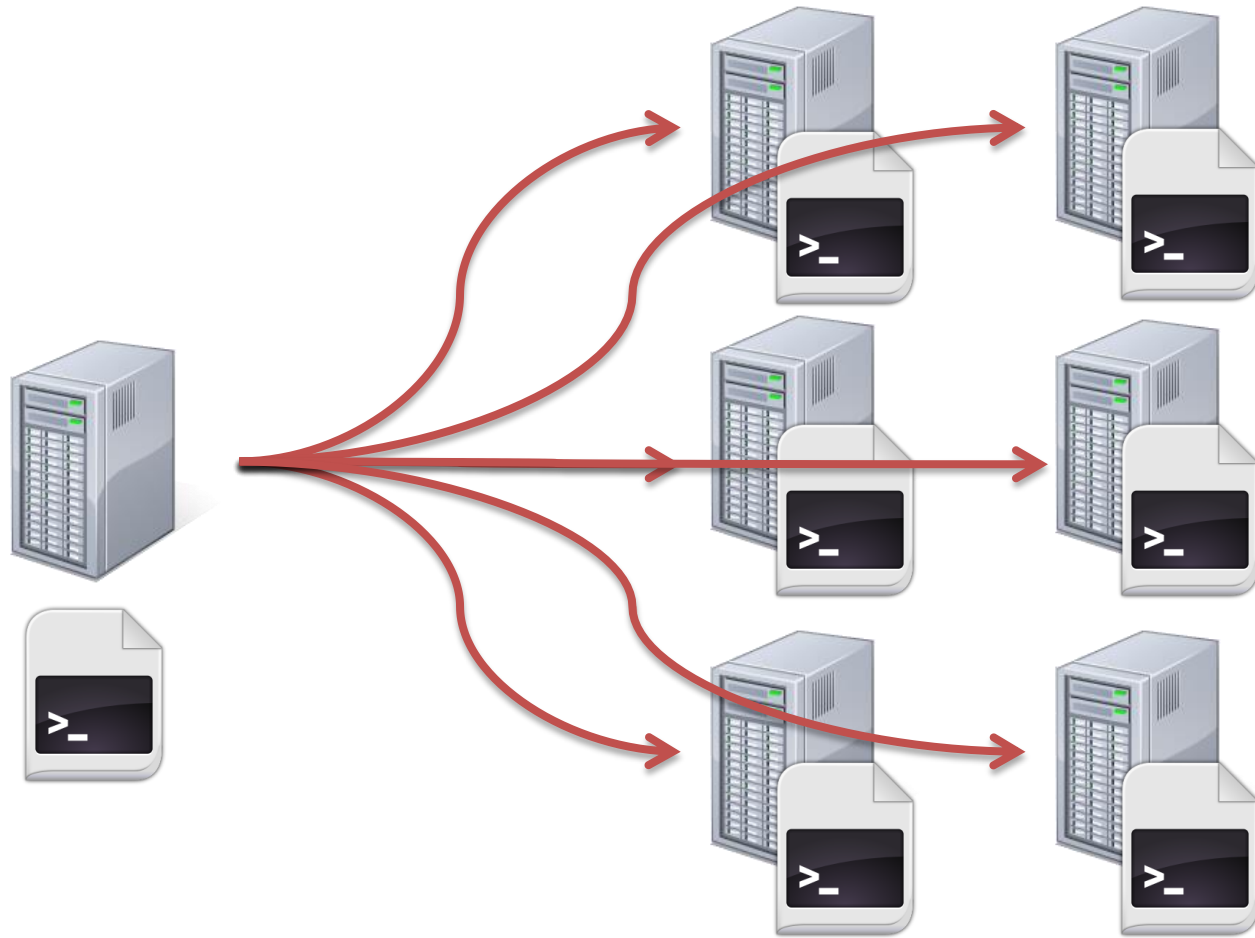


Login node



Computing (batch) nodes

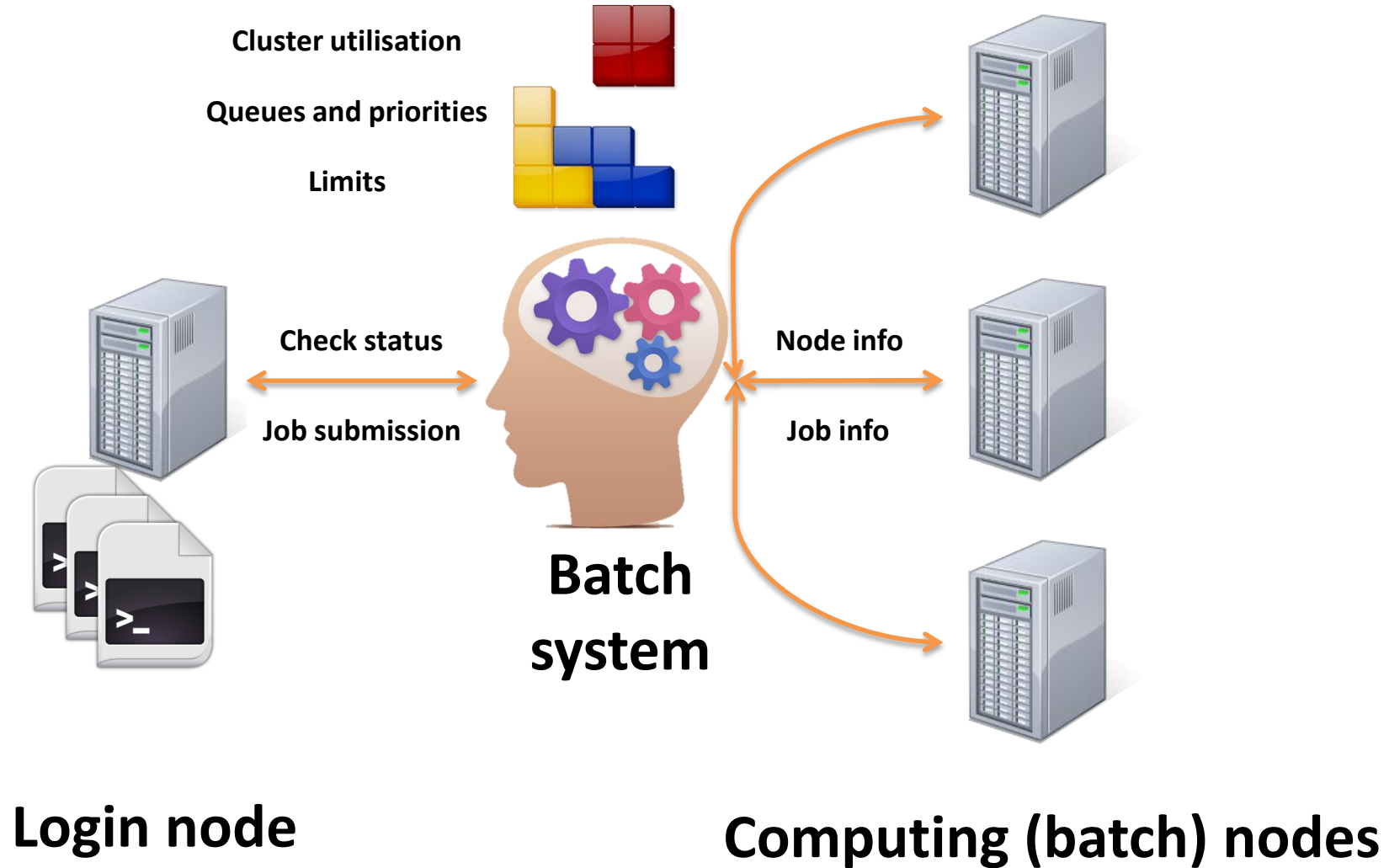
Interactive vs Batch



Login node

Computing (batch) nodes

How does it work?



Quality of service (queues)

- In Slurm, QoS (Quality of Service) = queue
- The queues have an associated priority and have certain limits

QoS	Description	Priority	Wall Time Limit	Total Jobs	User Jobs
normal	Suitable for most of the work. This is the default queue	Standard	1 day	256	32
express	Suitable for short jobs	High	3 hours	256	32
long	Suitable for long jobs	Low	7 days	32	4
timecrit1	Automatically set by EcAccess for Time Critical Option 1 jobs. Not accessible directly by the user	Highest	8 hours	128	16

Batch job script

- A job is a shell script
 - bash/ksh/csh
- Directives are shell comments:
 - starting with **#SBATCH**
 - Lowercase only
 - No spaces in between
 - No variable expansion
- All directives are optional
 - System defaults in place

```
#!/bin/bash
# The job name
#SBATCH --job-name=helloworld
# Set the error and output files
#SBATCH --output=hello-%J.out
#SBATCH --error=hello-%J.out
# Set the initial working directory
#SBATCH --workdir=/scratch/us/usxa
# Choose the queue
#SBATCH --qos=express
# Wall clock time limit
#SBATCH --time=00:05:00
# Send an email on failure
#SBATCH --mail-type=FAIL

# This is the job
echo "Hello World!"
sleep 30
```

Job directives

Directive	Description	Default
<code>--job-name=...</code>	A descriptive name for the job	Script name
<code>--output=...</code>	Path to the file where standard output is redirected. Special placeholders for job id (%j) and the execution node (%N)	slurm-%j.out
<code>--error=...</code>	Path to the file where standard error is redirected. Special placeholders for job id (%j) and the execution node (%N)	output value
<code>--workdir=...</code>	Working directory of the job. The output and error files can be defined relative to this directory.	submitting dir
<code>--qos=...</code>	Quality of service (queue) where the job is to be submitted	normal*
<code>--time=...</code>	Wall clock limit of the job (not cpu time limit!) Format: m, m:s, h:m:s, d-h, d-h:m or d-h:m:s	qos default
<code>--mail-type=...</code>	Notify user by email when certain event types occur. Valid type values are BEGIN, END, FAIL, REQUEUE, and ALL	disabled
<code>--mail-user=...</code>	Email address to send the email	submit user
<code>--hold</code>	Submit the job in held state. It won't run until released with <code>scontrol release <jobid></code>	not used

Batch Basic tasks

- Submit a job:

```
$> sbatch hello.sh  
Submitted batch job 1250968
```

- Check the queue / status of the job:

```
$> squeue -u $USER  
JOBID      NAME      USER      QOS      STATE      TIME  TIMELIMIT  NODELIST (REASON)  
1250968 helloworld usxa      express  RUNNING    0:08      5:00      ecgb07
```

- Cancel a job:

```
$> scancel 1250968  
$> cat hello-1250968.out  
Hello world!  
slurmd[ecgb07]: *** JOB 1250968 CANCELLED AT 2014-02-28T17:08:29 ***
```


Checking limits and general usage: sqos

- **sqos**: Utility to have an overview of the different QoSs, including usage and limits
 - This utility is ECMWF specific (not part of a standard Slurm installation)

```
$> sqos
      QoS  Prio    Max Wall  Total Jobs  User Jobs  Max CPUS  Max Mem
-----  -
  express  400    03:00:00   11 / 256   7 / 32     1 10000 MB
  normal   300    1-00:00:00  23 / 256   4 / 32     1 10000 MB
  long     200    7-00:00:00   7 / 32     4 / 4     1 10000 MB
  large    200    08:00:00    0 / 8      0 / 4     1 10000 MB
  timecrit1 500    08:00:00    0 / 96     0 / 16     1 10000 MB
```

Total: 43 Jobs, 41 RUNNING, 2 PENDING

```
  Account  Def QoS  Running Jobs  Submitted Jobs
-----  -
  *ectrain  normal   15 / 50      17 / 1000
```

User trx: 17 Jobs, 15 RUNNING, 2 PENDING

Practical 1: Basic job submission

- Practicals must be run on **ecgate**, so make sure you log in there first!

```
$> ssh ecgate  
$> cd $SCRATCH  
$> tar xvzf ~trx/mv_data/batch_job.tar.gz  
$> cd batch_job/
```

1. Have a look at the script “run_metview_plot.job”
2. Submit the job and check whether it is running
 - What QoS is it using? What is the time limit of the job?
3. Where did the output of the job go? Have a look at the output
4. Submit the job again and then once it starts cancel it
5. Check the output



MS Time-critical Applications: Option 1

- Enhanced ECaccess batch system
 - Scheduled run of jobs and retry mechanism (in ectrans) were already available in ECaccess.
- New concept of events, also known as notifications, added to ECaccess
 - Events are defined by one user; they can be made publicly available.
 - Event name: “fc12h240”
 - Event description: “at this stage, the 10 day forecast data from the high resolution 12UTC run is available”
 - Users can subscribe their own jobs to “public events”; these jobs will remain in standby mode until ...

MS Time-critical Applications: Option 1

- New concept of events, also known as notifications, added to ECaccess
 - ... until the event owner sends a notification to an event; ECaccess will then submit the jobs subscribing to that event.
 - Environmental variables can be passed to the jobs when the notification is given to the event, e.g. a date, time, ...
 - Last but not least, soon after the notification of an event, ECaccess will schedule a new version of the jobs subscribing to the event, ready to be submitted at the next notification.
- More than 1300 jobs for about 170 users in ~60 events.

MS Time-critical Applications: Option 1

ECMWF ecaccess service > Jobs > Submit

Submit a script.

Source script

Please write your script :

```
#  
# Batch request script:  
#
```

Or upload it from :

Subscribe to notification(s).

Subscription(s)

<input type="checkbox"/>	Id	Name	Comment
Notification(s) list			
<input type="checkbox"/>	326	ef00hplumes	At this stage, the EPS plume charts at 00UTC have been updated.
<input type="checkbox"/>	341	ef00h240	At this stage, the ensemble forecast model at 00UTC (step 240) is complete
<input type="checkbox"/>	323	ef12h504	At this stage, the ensemble forecast model at 12UTC - for step 504 (21 days) is complete
<input type="checkbox"/>	342	ef12h240	At this stage, the ensemble forecast model at 12UTC - step 240 - is complete
<input type="checkbox"/>	324	ef00h504	At this stage, the ensemble forecast model at 00UTC - for step 504 (21 days) is complete
<input type="checkbox"/>	343	bc00h072	At this stage, the boundary condition forecast at 00UTC - step 72 - is complete.
<input type="checkbox"/>	167	an00h000	At this stage, the analysis at 00UTC is complete.
<input type="checkbox"/>	168	an12h000	At this stage, the analysis at 12UTC is complete.

Script content editor

With the "script content editor" you can either enter your text line by line, or cut and paste the text from your computer.

Do not upload

As an alternative to the "script content editor", you can press the "Browse" button to navigate on your hard disk and select a script file from your computer.

Current PATH

The script you are going to submit will be executed in your ECSCRATCH domain. All non absolute PATH used in your script will be relative to this domain.

ECaccess target queue

Done

buaccess.ecmwf.int

MS Time-critical Applications: Option 1

ECMWF eaccess service > Jobs > Submit - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

<input type="checkbox"/>	285	mofcrefc	At this stage, the Monthly Forecast hindcast products have been updated.
<input type="checkbox"/>	286	wm00h120	At this stage, the the European Shelf model at 00UTC is complete (step 120).

Total: 34

automatically renew subscription
 one script to one notification

Settings of your job request.

Job settings

Do not run before (yyyy-MM-dd HH:mm): 2006-05-10 20:
Job expires after (in days): 7
Select the ECaccess target queue: ecgate (LoadLeveler) [v]
 script contains no directives
 redirect stderr to stdout

Man page for your job:

man page for ECMWF operators

Retry frequency (in seconds): 600
Retry count (<0:disabled): 144

Send mail to: usl
 at the beginning
 at the end (on success)
 at the end (on failure)
 when restarted (on retry)

Submit job

Done buaccess.ecmwf.int

start ECMWF eaccess ser... documents time_crit_opt_1 Ug - ... untitled - Paint EN 97% 21:40

automatically
renew subscription

one script to one
notification

man page for
ECMWF operators

retry count and
frequency

MS Time-critical Applications: Option 1

- Command line client:
 - ecaccess-job-submit
 - ecaccess-job-list
 - ecaccess-job-delete
 - ecaccess-job-get
 - ecaccess-event-list

Additional Info

- General Batch system and SLURM documentation:
 - <https://software.ecmwf.int/wiki/display/UDOC/Batch+Systems>
 - <https://software.ecmwf.int/wiki/display/UDOC/SLURM>
 - <https://software.ecmwf.int/wiki/display/UDOC/Slurm+job+script+examples>
- ECAccess Documentation
 - <https://software.ecmwf.int/wiki/display/ECAC/ECaccess+Home>
- SLURM website and documentation:
 - <http://www.schedmd.com/>
 - <http://www.schedmd.com/slurmdocs/documentation.html>
 - <http://www.schedmd.com/slurmdocs/tutorials.html>

Questions?