

# **An Introduction to Cluster Computing at PIK**

A beginner's guide

Ciaron Linstead, IT Services  
10th November 2011

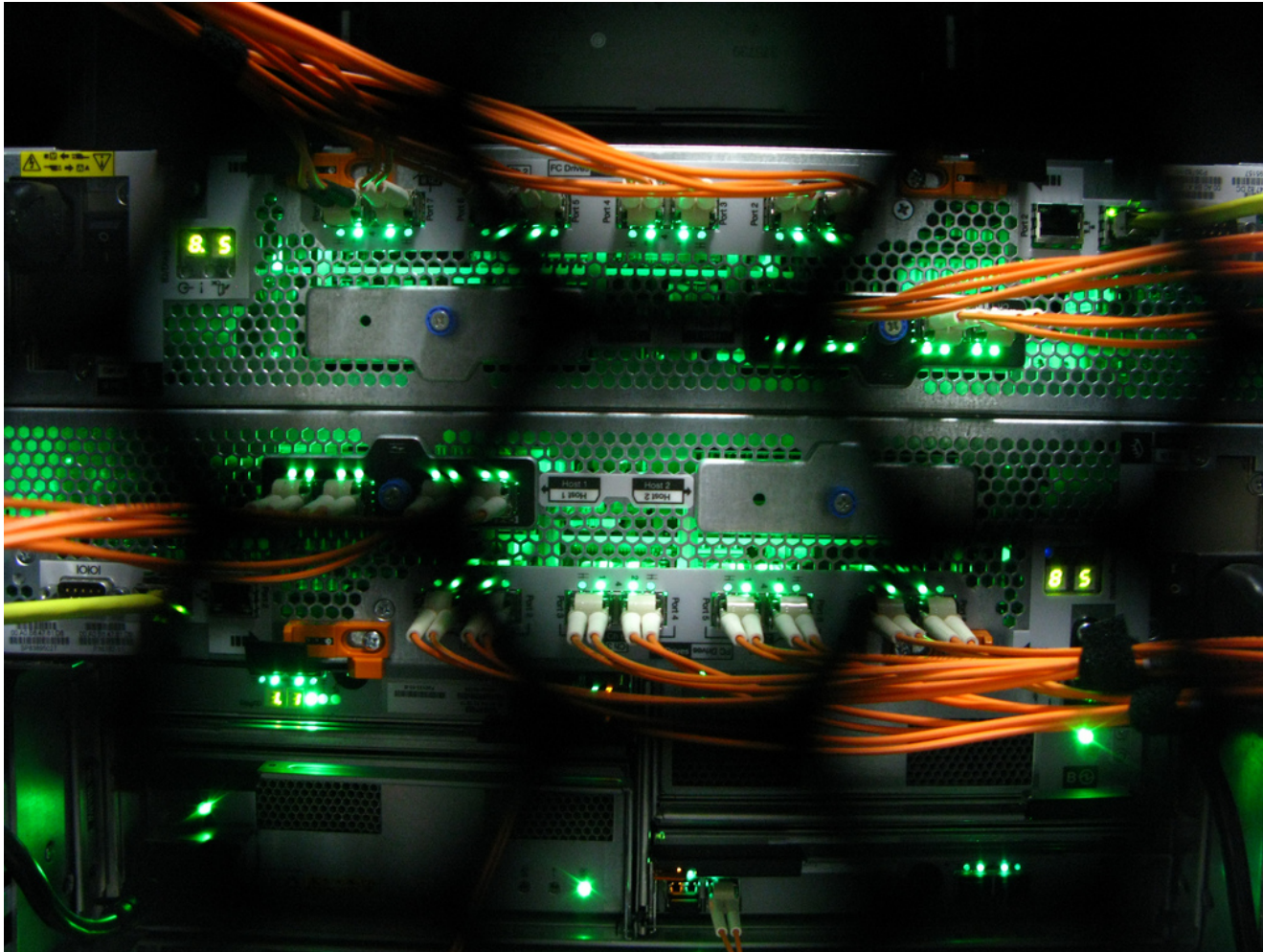




# Introduction

1. Description of the cluster
2. Preparing and running a job
3. Using the cluster effectively
4. Demo(s)

# 1. The Cluster

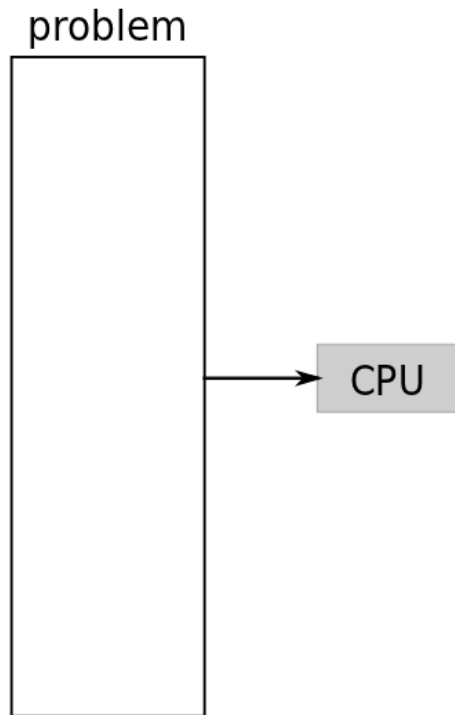


# What is a Compute Cluster?

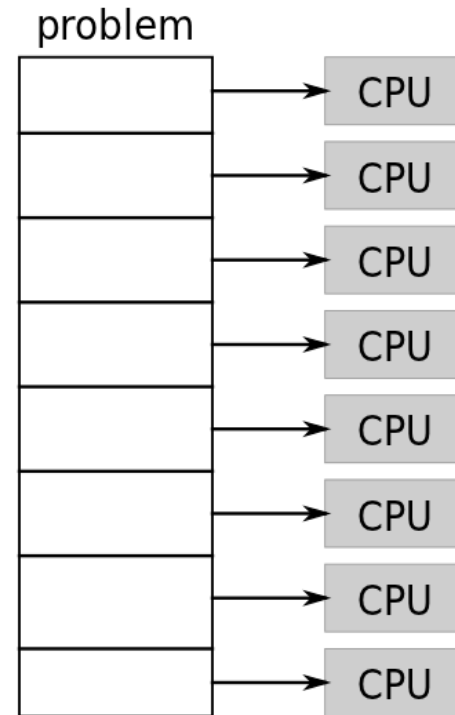
a group of computers ('nodes') connected by a dedicated high-speed network

- high performance
- high availability

# How is it used?

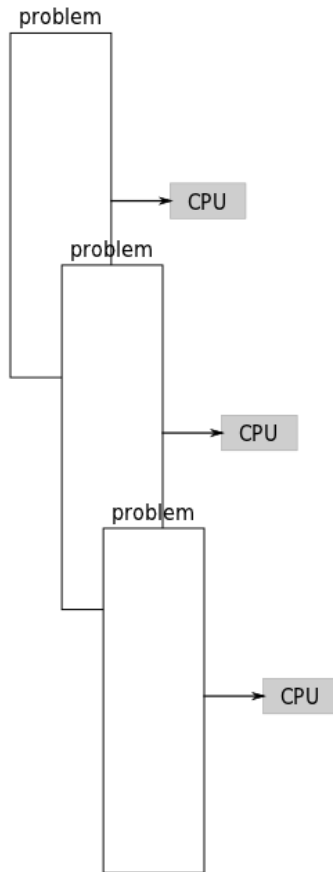


serial computing

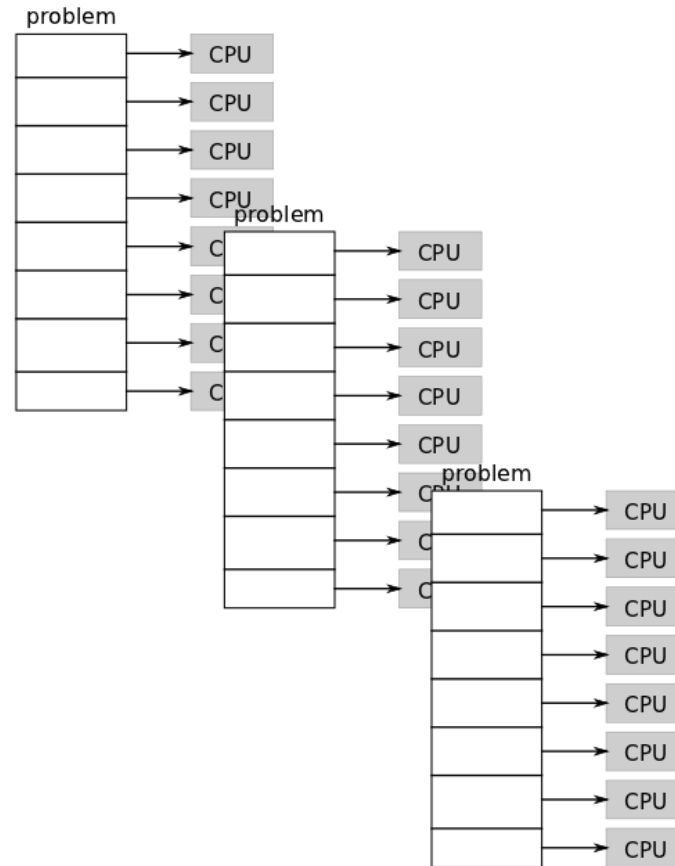


parallel computing

# How is it used?



serial computing



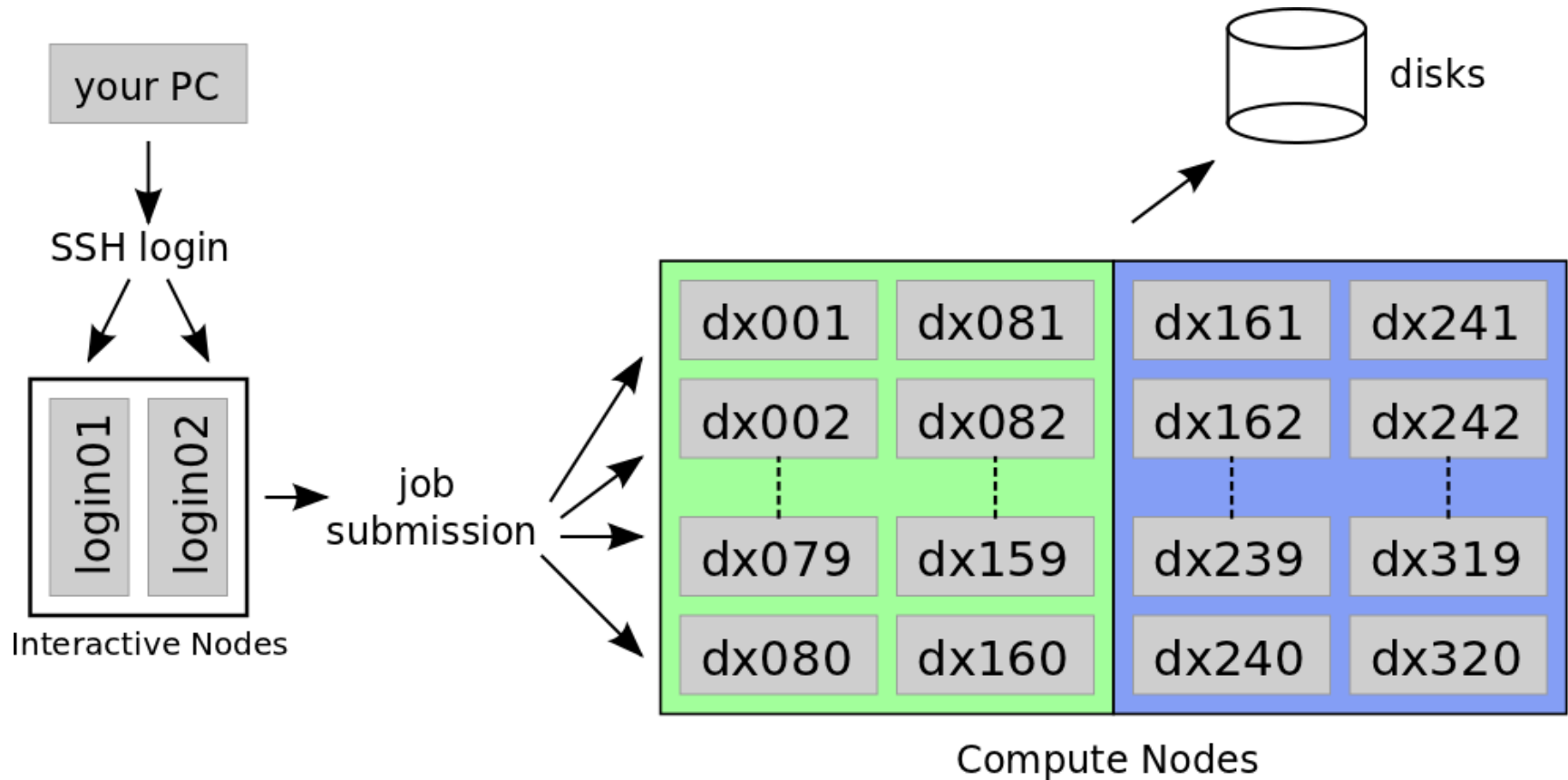
parallel computing

# How is it used?

- Log in
- (Prepare code)
- Prepare job for a *workload scheduler*
  - at PIK: LoadLeveler
- Submit

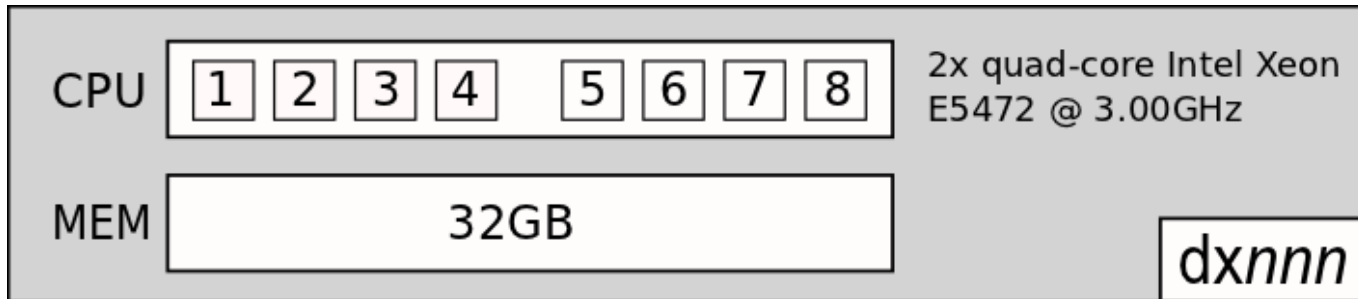
# The PIK Compute Cluster

- hardware



# The PIK Compute Cluster

- hardware





# The PIK Compute Cluster

## - filesystems

- `/home/ (your username)`
  - not for application data!
- `/ipler/01`
  - fast, parallel filesystem for project groups, backed up
- `/scratch/01`
  - fast, parallel filesystem. No backups, 30+ day old files deleted
- archiving and backups: IBM Tivoli Storage Manager (`dsmc`, `dsmj`)



# The PIK Compute Cluster

- software

## Development Tools

- Intel Cluster Studio (C, C++, Fortran, MPI, MKL)
- Matlab, Python, R
- source code management: Subversion
- standard Linux tools: GCC, sed, awk, make

# The PIK Compute Cluster

- software



Code profiling and tuning

- Intel VTune Amplifier XE2011
- Intel Trace Analyzer and Collector
- GNU profiler



# The PIK Compute Cluster

- software

## Debugging

- Totalview debugger
- Intel debugger
- GNU debugger

Most software: `/iplex/01/sys/applications`



# The PIK Compute Cluster

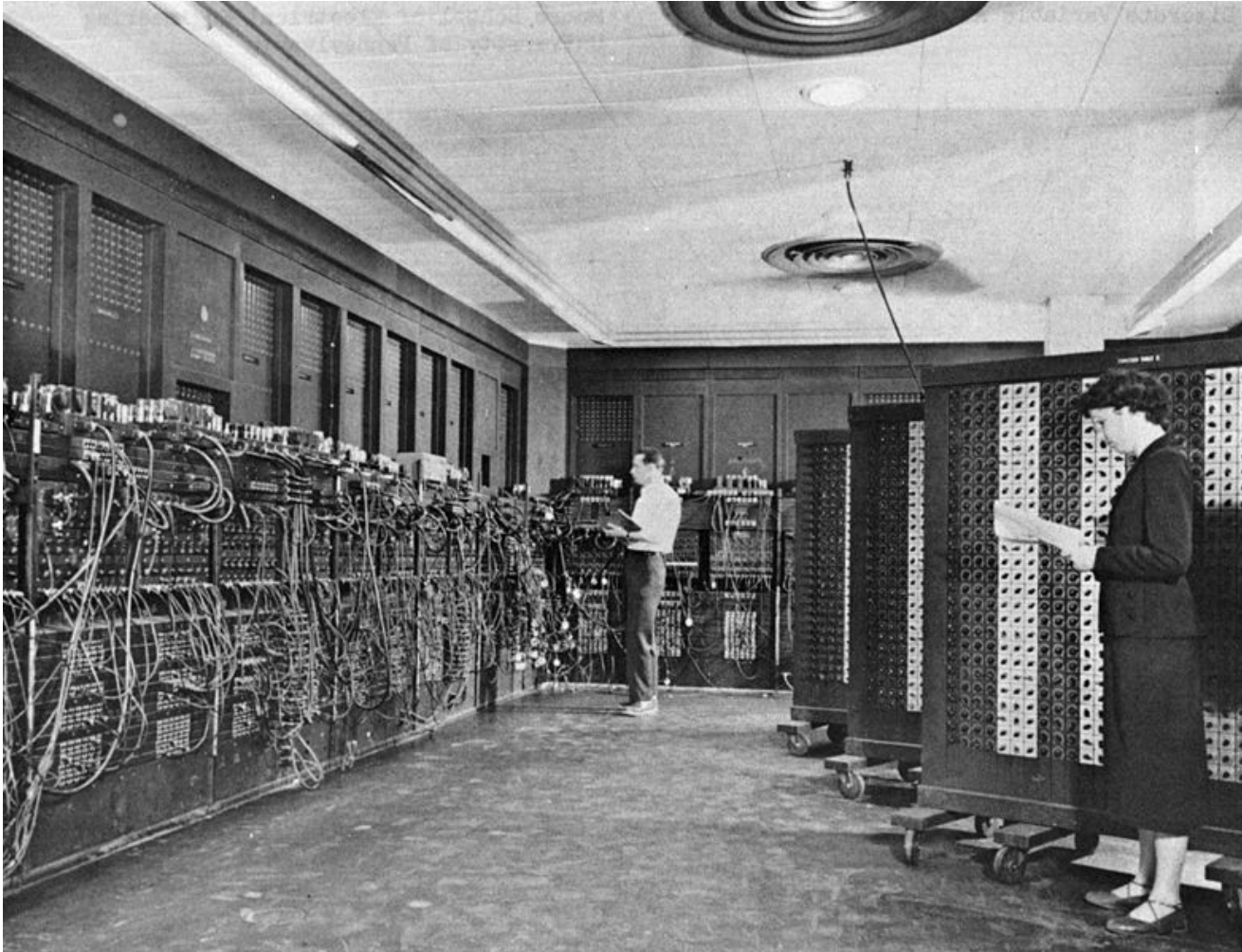
- so far this year:
  - 200000 jobs run
  - 9 million hours of CPU time



9M hours = 1026 years

985AD: Erik the Red  
colonises Greenland

## 2. Preparing & running a job



# Running a program

## - log in

- Login via Secure Shell (SSH) public key authentication
  - see HPC documentation
- From Linux or a Mac
  - `ssh cluster` on the command line / Terminal
  - `ssh cluster.pik-potsdam.de`
- From Windows
  - PuTTY (<http://putty.org>)

More: <http://www.pik-potsdam.de/members/gibietz/how-to/advanced-ssh-tunneling/overview-ssh-client-software-for-your-computer>

# Running a program

## - log in

```
From: Wordsmith <wsmith@wordsmith.org>
To: linstead@pik-potsdam.de
Subject: A Word A Day--apopemptic
Date: Mon, 7 Nov 2011 00:23:24 -0500 (11/07/2011 06:23:24 AM)
Wordsmith.org

LAST DOWNTIME

Cluster was down due to scheduled maintenance
August 14th 6:00 p.m. to August 15th 6:00 a.m.

UPCOMING MAINTENANCE to sponsor
celebrityplace.com: thinker's daily quote In My Book@
Carefully selected non-fiction book excerpt free to your email each day for sale in fine b
www.inmybook.com

Nov 7, 2011
This week's theme
Odds and ends

linstead@login01:~$
```

# Running a program

## - preparing a job

### The LoadLeveler Job Command File (JCF)

```
#!/bin/ksh
```

```
# @ class = short
```

```
# @ group = its
```

```
# @ notify_user = linstead@pik-potsdam.de
```

```
# @ output = example.out
```

```
# @ error = example.err
```

```
# @ queue
```

```
/home/linstead/examples/serial/hello_world
```

# Running a program

## - submitting a job

- `llsubmit example.jcf`

```
llsubmit: The job "cwsa.iplex.pik-potsdam.de.102812" has  
been submitted.
```

- `llq [ -u myusername ]`

```
cwsa.102812.0  linstead  11/7  12:10 I  50  short
```

# Running a program

## - submitting a job



**From:** [loadl@pik-potsdam.de](mailto:loadl@pik-potsdam.de)  
**To:** [linstead@pik-potsdam.de](mailto:linstead@pik-potsdam.de)  
**Subject:** cwsa.iplex.pik-potsdam.de.102812  
**Date:** Mon, 07 Nov 2011 12:11:28 +0100

From: LoadLeveler

LoadLeveler Job Step: cwsa.iplex.pik-potsdam.de.102812.0  
Executable: /home/linstead/examples/JCF/simple.jcf  
Executable arguments:  
State for machine: dx055.iplex.pik-potsdam.de  
LoadL\_starter: The program, simple.jcf, exited normally and returned an exit code of 0.

This job step was dispatched to run 1 time(s).  
This job step was rejected by Starter 0 time(s).  
Submitted at: Mon Nov 7 12:10:30 2011  
Started at: Mon Nov 7 12:11:28 2011  
Exited at: Mon Nov 7 12:11:28 2011

Real Time: 0 00:00:58  
Job Step User Time: 0 00:00:00  
Job Step System Time: 0 00:00:00  
Total Job Step Time: 0 00:00:00

Starter User Time: 0 00:00:00  
Starter System Time: 0 00:00:00  
Total Starter Time: 0 00:00:00

# Running a program

- but where's my output?

Unless I say otherwise, same place as my JCF

```
linstead@login01:~/examples/helloworld$ cat example.out
```

```
Hello world!
```

# Running a program

- more examples



```
/iplex/01/sys/load1/examples
```

# 3. Using the cluster effectively



# Using the cluster effectively

- filesystems



`$HOME` not for application data!

# Using the cluster effectively

## - useful JCF parameters

```
# @ class = ...
```

- short

- 1 day, 256 processes

- medium

- 7 days, 128 processes

- long

- 30 days, 4 processes

- largemem

- 7 days, 1 core, 14GB memory

- dev, io

# Using the cluster effectively

- useful JCF parameters

```
# @ wall_clock_limit = ...
```

- HH:MM:SS
- lets you jump the queue!

# Using the cluster effectively

## - useful JCF parameters

```
# @ job_name = ...
```

- easily identify your job in log files
- Recommended: use model name
  - e.g. LPJmL, Lagom, PISM-PIK etc

# Using the cluster effectively

- useful JCF parameters

```
# @ group = ...
```

- `run l1groups`
- pick the right one for your job

# Using the cluster effectively

## - useful JCF parameters

- OK

```
# @ output = example.out
```

```
# @ error = example.err
```



# Using the cluster effectively

- useful JCF parameters

- Better...

```
# @ output = /scratch/01/linstead/example.out
```



# Using the cluster effectively

- useful JCF parameters

- even better...

```
# @ output = /scratch/01/linstead/example_$(jobid).out
```



# Using the cluster effectively

- useful JCF parameters

- Better still...

```
# @ output = /scratch/01/$(user)/example_$(jobid).out
```



# Using the cluster effectively

## - useful JCF parameters

- Best!

```
# @ job_name = example
```

```
# @ initialdir = /scratch/01/$(user)
```

```
# @ output = $(job_name)_$(jobid)_$(stepid).out
```



# Using the cluster effectively

## - useful JCF parameters

### ● other variables

**\$(host)**: The hostname of the machine from which you submitted the job.

**\$(domain)**: The domain of the host from which you submitted the job.

**\$(jobid)**: A sequential number assigned to the job by the submitting machine. Equivalent to the `$(cluster)` variable.

**\$(stepid)**: The sequential number assigned to a job step when more than one queue statement appears in the Job Command file. The `$(stepid)` and `$(process)` variables are equivalent.

**\$(executable)**: Contains the name of the executable if you set the executable keyword.

**\$(base\_executable)**: Contains the name of the executable without the directory path if you set the executable keyword.

**\$(class)**: Contains the name of the job class that your job has been submitted to if you set the class keyword.

**\$(comment)**: Contains the comment text if you set the comment keyword.

**\$(job\_name)**: Contains the job name text if you set the `job_name` keyword.

**\$(step\_name)**: Contains the step name text if you set the `step_name` keyword.

# Using the cluster effectively

## - useful LoadLeveler commands

- `llcancel`
  - `llcancel cws02a.76469.0 -h cws02a`
  - `llcancel -u linstead -h cwsa -h cws02a`
- `llclass`
  - show available classes
- `llstatus`
  - show the load on the cluster



# Using the cluster effectively

## - common errors

### Job stays Idle

```
# @ output = /scartch/01/${user}/${job_name}_${cluster}.  
out
```

# Using the cluster effectively

## - common errors



## Wrong or missing group

```
llsubmit: 2512-080 Class "short" is not valid for group "itss".
```

```
llsubmit: 2512-051 This job has not been submitted to LoadLeveler.
```

```
llsubmit: 2512-078 Group "esm" is not valid for user "linstead".
```

```
llsubmit: 2512-051 This job has not been submitted to LoadLeveler.
```

# Using the cluster effectively

## - how to get help

- documentation

- <http://www.pik-potsdam.de/services/it/hpc>
- <http://publib.boulder.ibm.com/epubs/pdf/a2278818.pdf>
- /iplex/01/sys/loadl/examples

- email

- [cluster-admin@pik-potsdam.de](mailto:cluster-admin@pik-potsdam.de)

- submitting problem reports

- include job ID, path to JCF

# To conclude...

Try the examples.

If you get stuck:

- `cluster-admin@pik-potsdam.de`

Spring 2012:

- Advanced LoadLeveler - get your job to the top of the queue.

# 4. Demos

