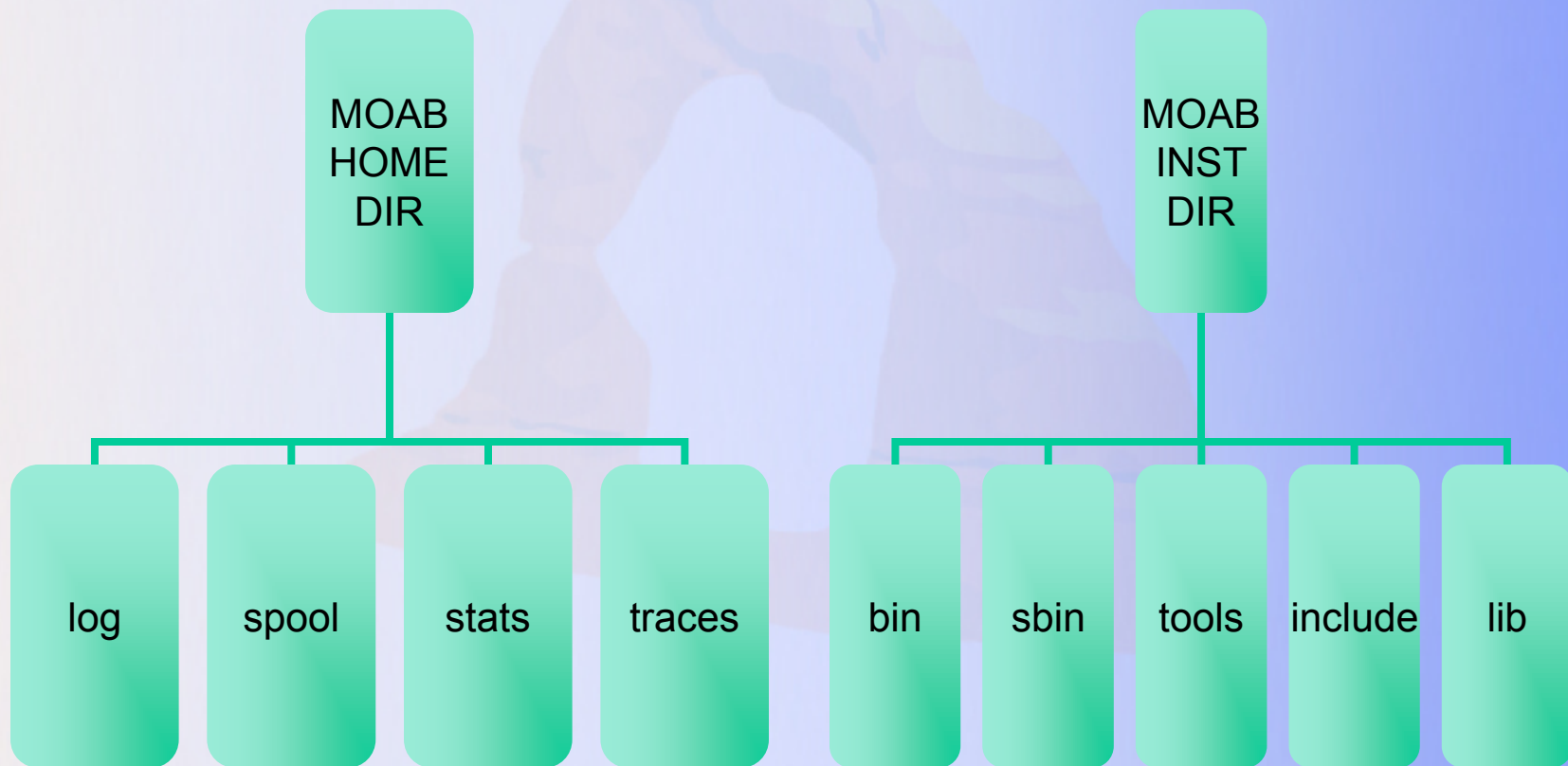


# Troubleshooting and Diagnostics

- File Locations
- Scheduling Modes
- Diagnostic Commands
- Logging and Event Files
- Checkpoint File
- Other useful tricks

<http://www.clusterresources.com/products/mwm/moabdocs/14.0troubleshootingandsysmaintenance.shtml>

# Directory Layout



# Moab Home Directory Contents

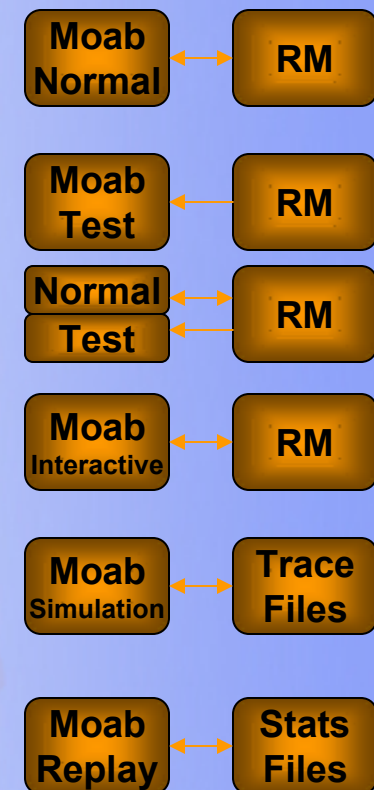
- Default is “/opt/moab”
- overridden by MOABHOMEDIR environment variable
- Files
  - **moab.cfg**
    - config file containing information required by both the Moab server and user interface clients
  - **moab-private.cfg**
    - config file containing private information required by the Moab server only
  - **.moab.ck**
    - Moab checkpoint file preserved state across restarts
  - **moab.dat**
    - Used only by Moab Cluster Manager (GUI)
- Subdirectories
  - **log** Contains Moab server and tool log files
    - **moab.log, moab.log.1, ...**
  - **Stats** Contains Moab statistics files
    - Moab stats files are of the format 'events.Dow\_Moy\_DD\_CCYY'
    - Moab fairshare data files are of the format 'FS.EpochTime')
  - **Spool** Contains temporary Moab files (i.e. job scripts)
  - **Traces** Contains trace files used in simulations

# Moab Install Directory Contents

- Default is “/usr/local”
- Configured via `--instdir` (`--prefix` in moab 5.2.0)
- Subdirectories
  - **bin**
    - Contains Moab client commands – `showq`, `checkjob`, `mdiag`, `mschedctl`, etc.
    - All the executables are identical copies of `mclient`
    - Needed on server and login nodes
  - **sbin**
    - Contains server daemons – `moab`, etc.
    - Needed on server node only
  - **tools**
    - Contains helper scripts for interfacing with external services
    - For example: `ipmi`, `ganglia`, `data-staging`, `license management`, `native resource manager interfaces` (`sgc`, `bproc`, `cray xt3`), `emulation clients` (`lsf`, `slurm`, `pbs`)
  - **lib**
    - Contains static Moab libraries and Perl modules for tools
  - **Include**
    - Contains Moab header files

# Scheduling Modes

- Normal (Default)
  - Used for production scheduling
  - Starts, cancels, modifies jobs and performs all scheduling services
- Test (Monitor)
  - Allows evaluation of new Moab releases, configurations, and policies in a risk-free manner
  - Loads live job and node information from resource managers
  - Does NOT start, cancel, or modify jobs
- Interactive
  - Allows “man-in-the-middle” scheduling – no action taken without admin approval
- Simulation
  - Allows a *test drive* of the scheduler
  - For evaluating how various policies can improve the current performance on a production system.
- Replay
  - Allows replay of historical events
  - For training and/or troubleshooting purposes



# Introducing New Policies

- Verifying Correct Specification of New Policies
  - If manually editing the moab.cfg file, check validity with the mdiag – C command
  - If done via Moab Cluster Manager, automatically verifies proper policy specification
- Verifying Correct Behavior of New Policies
  - Use INTERACTIVE mode and verify each decision
  - Use TEST mode and examine decisions with mdiag -S -v
- Determining Long Term Impact of New Policies
  - Use SIMULATION Mode and compare results to straight run

# Diagnostics

- Use the mdiag command to diagnose system health
- Moab's diagnostic commands present detailed information about the system and objects (jobs, nodes, reservations, etc.)
  - Object and system state and configuration
    - Attributes, policies, statistics
  - Scheduling and resource access problems
  - Failures and messages
  - Performs numerous internal health and consistency checks
  - Past and current performance
- Output can be provided in XML for parsing (--flags=xml)

# mdiag options

- Use `-v` when troubleshooting
- Most common diagnostics
  - Scheduler (`-S`)
  - Resource manager (`-R`)
  - Configuration (`-C`)
  - Jobs (`-j`)
  - Blocked jobs (`-b`)
  - Nodes (`-n`)
  - Priority (`-p`)
- Other diagnostics
  - Users (`-u`), Accounts (`-a`), Classes (`-c`), Groups (`-g`)
  - Reservations (`-r`), Standing Reservations (`-s`), Partitions (`-t`)
  - Fairshare (`-f`), QOS (`-q`), Triggers (`-T`)

```
> mdiag -R

RM[base] Type: PBS State: Active ResourceType: COMPUTE
Version: '1.2.0p6-snap.1124480497'
Nodes Reported: 4
Flags: executionServer,noTaskOrdering,typeIsExplicit
Partition: base
Event Management: EPORT=15004
NOTE: SSS protocol enabled
Submit Policy: NODECENTRIC
DefaultClass: batch
Variables: X=cat,Y=dog
RM Performance: AvgTime=0.00s MaxTime=1.03s (1330 samples)


RM[base] Failures:
Mon May 3 09:15:16 clusterquery 'cannot get node info (rm is unavailable)'
Mon May 3 10:25:46 workloadquery 'cannot get job info (request timed out)'

RM[Boeing] Type: NATIVE State: Active ResourceType: LICENSE
Cluster Query URL: file://$HOME/lic.dat
Licenses Reported: 3 types (3 of 6 available)
Partition: SHARED
License Stats: Avg License Avail: 0.00 (438 iterations)
Iteration Summary: Idle: 0.00 Active: 100.00 Busy: 0.00
RM Performance: AvgTime=0.00s MaxTime=0.00s (877 samples)

RM[GM] Type: NATIVE State: Active ResourceType: COMPUTE
Cluster Query URL: file:///tmp/gm.dat
Nodes Reported: 2
Partition: CM
RM Performance: AvgTime=0.00s MaxTime=0.00s (877 samples)

NOTE: use 'rmctl -f -r' to clear stats/failures
```

# Job Troubleshooting



Why won't  
my job  
run!

To determine why a particular job will not start, there are several commands which can be helpful:

- **checkjob -v**
  - Checkjob will evaluate the ability of a job to start immediately. Tests include resource access, node state, job constraints (ie, startdate, taskspnode, QOS, etc), job dependencies.
- **checknode**
  - Display detailed status of node
- **mdiag -b**
  - Display various reasons job is considered 'blocked' or 'non-queued'.
- **mdiag -j**
  - Display high level summary of job attributes and perform sanity check on job attributes/state.
- **showbf -v**
  - Determine general resource availability subject to specified constraints.

## Collecting a support snapshot

- `tools/support.diag.pl`
  - Creates a tarball to send to the technical support team that captures the state of the scheduler when a problem occurs.
  - Collects logs, and output from several diagnostic commands for use in troubleshooting the problem.

# Logging Facilities

- Moab Log
    - Report detailed scheduler actions, configuration, events, failures, etc
  - Syslog
    - Log INFO, WARN or DEBUG level info to system logging facility
  - Event Log
    - Report scheduler, job, node, and reservation events and failures
    - RECORDEVENTLIST  
JOBCANCEL,JOBEND,JOBSTART,SCHEDPAUSE,SCHEDSTART,SCHEDSTOP,TRIGEND,TRIGFAILURE,TRIGSTART
- 
- <http://www.clusterresources.com/products/mwm/moabdocs/a.fparameters.shtml#eventrecordlist>
  - <http://www.clusterresources.com/products/mwm/moabdocs/14.2logging.shtml>
  - <http://www.clusterresources.com/products/mwm/moabdocs/a.fparameters.shtml#usesyslog>

# Logging Parameters

- LOGDIR - Specifies directory for log files
- LOGFILE - Specifies name of log file
- LOGFILEROLLDEPTH – Specifies the maximum number of logs to maintain
- LOGFILEMAXSIZE - Specifies size of log file that causes it to roll
- LOGLEVEL - Specifies verbosity of logging (0 – 9)

## Moab Log Files

- Information about internal status is logged at all LOGLEVELs. Critical internal status is indicated at low LOGLEVELs while less critical and more verbose status information is logged at higher LOGLEVELs.

```
##moab.log
```

```
INFO:   job orion.4228 rejected (max user jobs)
```

```
INFO:   job fr4n01.923.0 rejected (maxjobperuser policy failure)
```

- NOTE: each log level increases the verbosity by an order of magnitude. High log levels (6+) can impact the performance of your system.

## Scheduler Warnings

- Warnings are logged when the scheduler detects an unexpected value or receives an unexpected result from a system call or subroutine.
- Alerts are logged when the scheduler detects events of an unexpected nature which may indicate problems in other systems or in objects.
- Errors are logged when the scheduler detects problems of a nature of which impact the scheduler's ability to properly schedule the cluster.

```
> grep -E "WARNING|ALERT|ERROR" moab.log
```

```
WARNING: cannot open fairshare data file '/opt/moab/stats/FS.87000'
```

```
ALERT: job orion.72 cannot run. deferring job for 360 Seconds
```

```
ERROR: cannot connect to Loadleveler API
```

## Enabling Syslog

- In addition to the log file, the Moab Scheduler can report events it determines to be critical to the UNIX syslog facility via the **daemon** facility using priorities ranging from INFO to ERROR.
- USESYSLOG = TRUE:local3
- The verbosity of this logging is not affected by the LOGLEVEL parameter. In addition to errors and critical events, user commands that affect the state of the jobs, nodes, or the scheduler may also be logged to syslog.
- Moab syslog messages are reported using the **INFO**, **NOTICE**, and **ERR** syslog priorities.

## Event Logs

- Major events are reported to both the Moab log file as well as the Moab event log. By default, the event log is maintained in the statistics directory and rolls on a daily basis, using the naming convention:
  - events.DoW\_MoY\_DD\_CCYY (e.g. **events.Fri\_Aug\_19\_2005**)

```
# stats/events.Wed_Mar_21_2007
10:38:27 1174495106 job 176 JOBSTART 0 1 scottmo scottmo 3600 Idle [batch:1] 1174495106
1174495106 1174495106 1174495106 - - ->= 0M >= 0M - 1174495106 0 0 :-
[RESTARTABLE][GLOBALQUEUE] - /var/moab/dev/spool/moab.job.Jjxszk - 0 0.00 ALL 1 0M
0M 0M 0 2140000000 keko internal -- [DEFAULT] -- 0.00 --- 0 --

10:38:28 1174495108 job 176 JOBEND 0 1 scottmo scottmo 3600 Completed [batch:1]
1174495106 1174495106 1174495106 1174495107 - - ->= 0M >= 0M - 1174495106 1 0 :-
[RESTARTABLE][GLOBALQUEUE] - /var/moab/dev/spool/moab.job.Jjxszk - 0 0.00 atlas 1 0M
0M 0M 0 2140000000 keko internal -- [DEFAULT] -- 1.00 --- 0 --
```

# showhist

- Information can be extracted from the stats files by using a beta tool called showhist
- In `–s tools/showhist.moab.pl bin/showhist`
- `showhist [-a account_name] [-c class_name] [-g group_name] [-j | -o object_type] [-n days] [-q qos_name] [-u user_name] [--show attribute_name[,attribute_name...]] [[-i] <object id>] [--help] [--man]`

```
# showhist -j 176
TimeStamp                JobId Event      User      Class      Account Procs  Status
-----
Wed Mar 21 10:38:26 2007 176  JOBSTART  scottmo  [batch:1] -      1      Idle
Wed Mar 21 10:38:28 2007 176  JOBEND    scottmo  [batch:1] -      1      Completed

# showhist -j -u scottmo -n 30
# showhist -o rsv
# showhist -o sched
```

# Checkpoint File

- Moab checkpoints its internal state so that it can remember information it has learned across restarts. The checkpoint file records statistics and attributes for jobs, nodes, reservations, users, groups, classes, and almost every other scheduling object.
- Stored in .moab.ck and .moab.ck1
  - Remove or edit these if you need to clear out the state
- Configuration parameters in moab.cfg
  - CHECKPOINTEXPIRATIONTIME - Indicates how long unmodified data should be kept after the associated object has disappeared. ie, job priority for a job no longer detected.
    - FORMAT - [[[DD:]HH:]MM:]SS
    - EXAMPLE - CHECKPOINTEXPIRATIONTIME 1:00:00:00
  - CHECKPOINTFILE - Indicates path name of checkpoint file
    - FORMAT - <STRING>
    - EXAMPLE - CHECKPOINTFILE /var/adm/moab/moab.ck
  - CHECKPOINTINTERVAL - Indicates interval between subsequent checkpoints.
    - FORMAT - [[[DD:]HH:]MM:]SS
    - EXAMPLE - CHECKPOINTINTERVAL 00:15:00

## Other Useful tricks

- If moab dies immediately at startup without any output, try starting moab in the foreground
  - Export MOABDEBUG=True or run moab -d
- If moab dies after some time, run moab under a debugger
  - MOABDEBUG=True gdb moab
- If you need to get a log file from Moab at a higher debug level than you are currently running
  - schedctl -L 9
  - creating temporary log '/var/moab/dev/log/moab.log.20070322202721' at log level 9
- If you want to stop moab from scheduling, but you still want to be able to run diagnostic and query commands, pause moab
  - schedctl -p
  - Step through iterations with schedctl -S 1
  - Resume normal scheduling with schedctl -r
- If you want to figure out your version, server host, install or home dir
  - moab --about

Questions?

