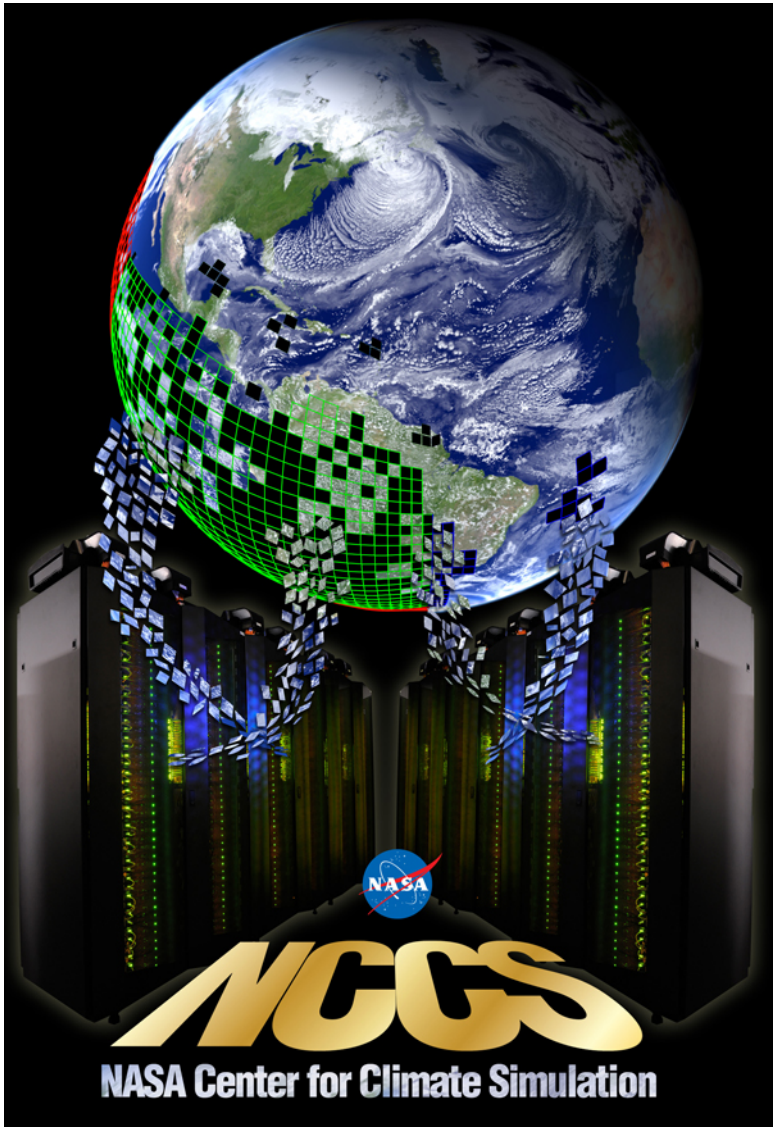


Debugging and Profiling GMAO Models with Allinea's DDT/MAP

Georgios Britzolakis
Georgios.Britzolakis-1@nasa.gov
April 30, 2015

Today's Objectives

Objectives:

- Outline Allinea's MAP/DDTs features
- Some preliminary settings on how to set up SSH connections to Discover that Allinea needs to work.
  - Configure the SSH settings (Review the NCCS primer)
  - Setup the Remote Installation directory for Allinea
- Demonstrate how to use DDT to debug GMAO models using the remote client
- Some guidance on how the *.qtf Allinea job files are generated
  - Describe the qtf file structure
  - Distinguish how different the *.qtf file is from the original job file.
- Demonstrate how to use MAP to profile GMAO modes using the remote client.
  - Present some of the features that MAP provides
- Give a real time example for DDT and MAP
- Questions

Allinea Features

- DDT:
  - Allinea DDT is a straight forward, scalable, graphical debugger capable of debugging a wide variety of scenarios found in today's development environments. With Allinea DDT it is possible to debug:
    - Single process with multithreaded software
    - OpenMP
    - Heterogeneous software such as that written to use GPUs
    - Hybrid codes mixing paradigms such as MPI + OpenMP or MPI + CUDA
    - Multi-process software of any form, including client server applications
  - Allinea DDT supports many compiled languages that are found in mainstream and high-performance computing including:
    - C, C++, and all derivatives of Fortran, including Fortran 90
    - Parallel languages/models including MPI, UPC, and Fortran 2008 Co-arrays
    - GPU languages such as HMPP, OpenMP accelerators, CUDA and CUDA Fortran

Allinea Features

- MAP:
  - Allinea MAP is a parallel profiler that aims to be powerful but easy to use, scalable to hundreds of thousands of processes while maintaining low overhead. Allinea MAP features:
    - A sampling profiler with adaptive sampling rates to keep the data volumes collected under control. Samples are aggregated at all levels to preserve key features of a run without drowning in data.
    - A folding code and stack viewer allows you to drill down to time spend on individual lines and draw back to see the big picture across nodes of routines
    - Both interactive and batch nodes for gathering profile data
    - A unified job control interactive with Allinea DDT – configure it once and both tools just work, with a familiar, usable interface
    - Metrics that allow memory usage, floating point calculations and MPI usage to be seen through a program run and across processes
    - All the above with just 5% application slowdown even with thousands of MPI processes

Preliminary Settings to Configure SSH for Allinea

- This assumes you have installed Allinea on your local machine
- Set up your .ssh/config (for linux/MAC systems) so it contains the following:

  host discover.nccs.nasa.gov dirac.nccs.nasa.gov dali.nccs.nasa.gov discover

  User ***<<Your User ID>>***

  LogLevel Quiet

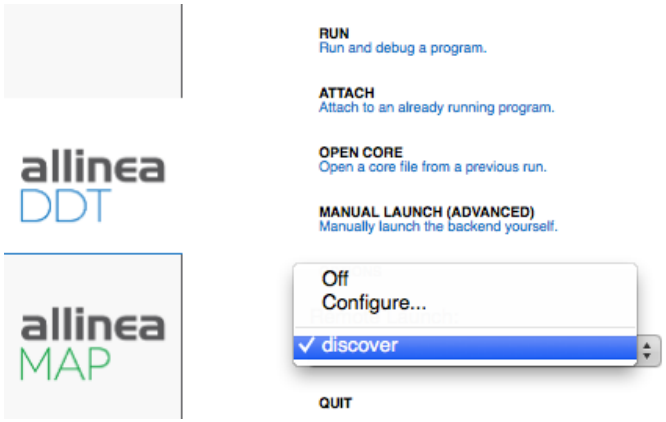
  ProxyCommand ssh -l ***<<Your User ID>>*** **login.nccs.nasa.gov direct %h**

  Protocol 2

  ConnectTimeout 30 ***(optional)***
- Add "Remote Launch" connection from the Allinea menus:
  - Go under "Remote Launch"->"Configure"-> "Add"
    - Connection Name: **discover**
    - Host Name: ***<<Your User ID>>*** **@discover**
    - Remote Installation Directory: **/usr/local/allinea/5.0.1/tools** **(for Allinea v5.0.1)**
    - Press OK to add the new connection
    - Back on the Main menu, select the **discover** connection so the connection is established

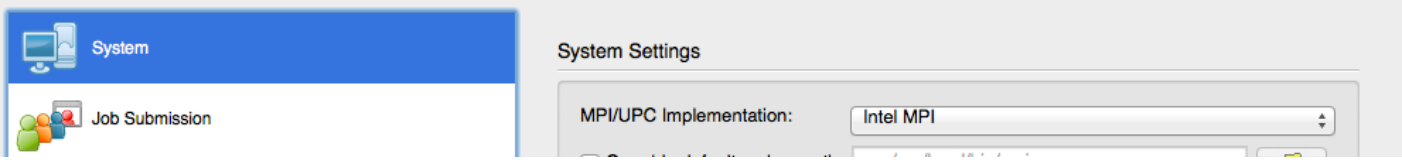
Set Up DDT with GMAO's Models

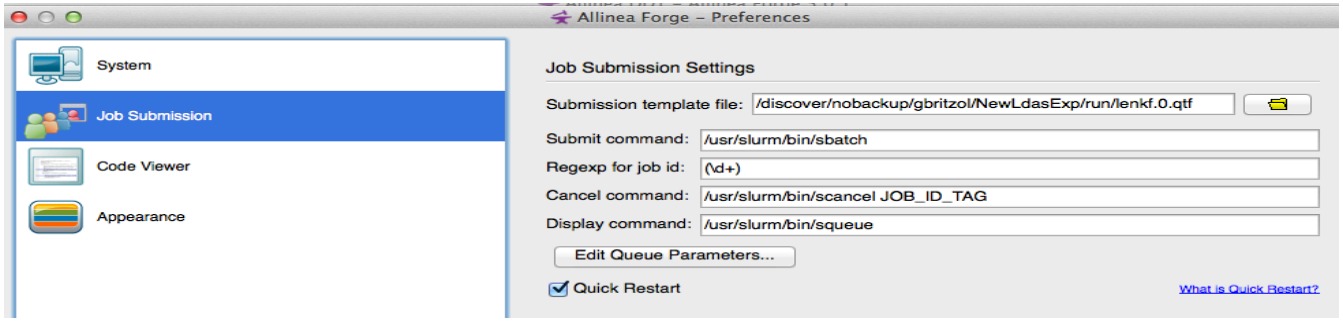
- Build your code using option: **BOPT=Og**
- Install your experiment test case by running ldsetup.py (or ldsetup) script
- In addition to the job script a *.qtf file was generated under the run directory.
- Start Allinea and select "discover" under the remote Launch section (as it was configured before).



Set Up DDT with GMAO's Models (Cont'd)

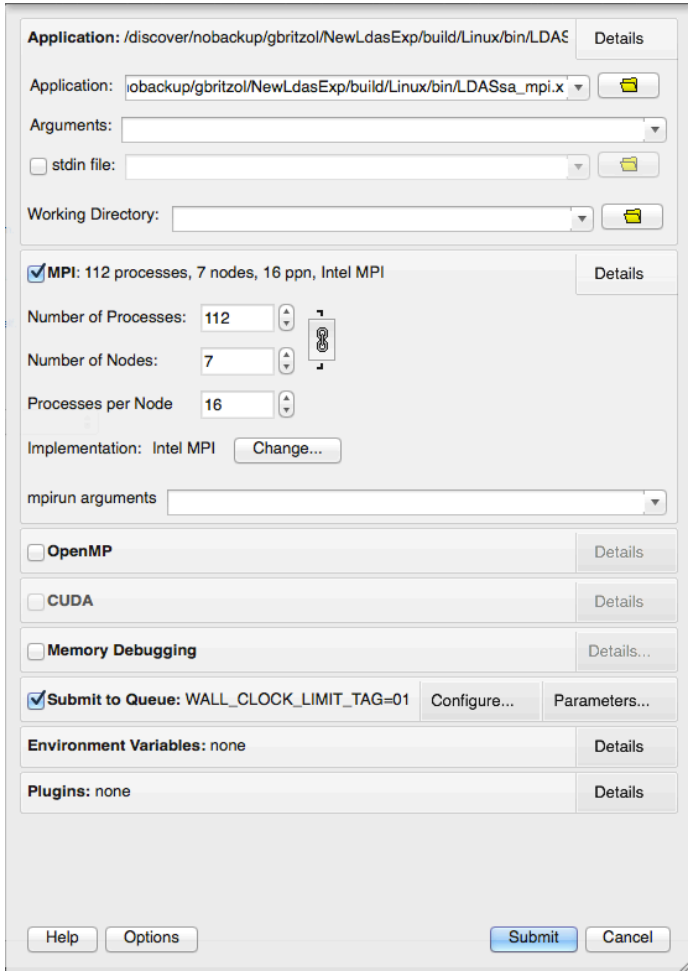
- On the main window, select "Options":
  - "System":
    - MPI/UPC Implementation: **"Intel MPI"**
  - "Job Submission":
    - Submission File: The location of your qtf file (under the run directory)
    - Edit Queue Parameters: Set the time you would like the resources to be requested for





Set Up DDT with GMAO's Models (Cont'd)

- On the main screen go to "Run"
- Identify the application that is going to be debugged
- Identify the number of Processes, Number of Nodes, Processes per Node. (Specify the same number of cores that the original job file includes.)
- Check Submit to Queue



Compare *.qtf Allinea Job File with Regular Job File

lenkf.0.j <-> lenkf.0.qtf

File Edit View Global Region Line Options Display Windows Help

lenkf.0.j

```
#!/bin/csh -fx

#SBATCH --account=k3001
#SBATCH --time=1:00:00
#SBATCH --ntasks=112

#SBATCH --qos=debug
#SBATCH --export=NONE
#SBATCH --constraint=sand
#SBATCH --output=../output/SMAP_EASEv2_M09_GLOBAL/rc_out/Y2014/M05/NewLdasExp.ldas_log.20140
#SBATCH --error=../output/SMAP_EASEv2_M09_GLOBAL/rc_out/Y2014/M05/NewLdasExp.ldas_out.201405

limit stacksize unlimited
source ../build/Linux/bin/g5_modules

setenv MKL_CBWR SSE4_2 # ensure zero-diff across archs
setenv MV2_ON_DEMAND_THRESHOLD 8192 # MVAPICH2

mpirun -np 112 \
 /discover/nobackup/gbritzol/NewLdasExp/build/Linux/bin/LDASsa_assim_mpi.x \
 -work_path ../output \
 -run_path . \
 -exp_id NewLdasExp \
 -exp_domain SMAP_EASEv2_M09_GLOBAL \
 -start_year 2014 -start_month  5 -start_day  1 -start_hour  0 -start_min  0 -start_sec  0 \
 -end_year   2014 -end_month    5 -end_day    1 -end_hour   15 -end_min    0 -end_sec    0 \
 -N_ens 24 \
 -spin .false. \
 -force_dtstep 3600 \
 -first_ens_id 0 \
 -resolution SMAP_EASEv2_M09 \
 -restart .true. \
 -restart_path ../input/restart/SMAP_Nature_v03 \
 -restart_domain SMAP_EASEv2_M09_GLOBAL \
 -restart_id SMAP_Nature_v03 \
 -tile_coord_file SMAP_EASEv2_M09_3856x1624.til \
 -met_tag e5110_fp__precCPCUG5FPv1 \
 -met_path ../input/met_forcing/GEOS5_land_forcing \
 -ens_prop_inputs_path . \
 -ens_prop_inputs_file m3_15_assim_test-ensprop_v01.nml \
 -driver_inputs_path . \
 -driver_inputs_file m3_15_assim_test-driver_v01-RRpaths.nml \
 -ens_und_inputs_path . \
```

lenkf.0.qtf

```
#!/bin/csh -fx
#
# WARNING: If you install a new version of the Allinea tools to the same
#          directory as this installation, then this file will be overwritten.
#          If you customize this script at all, please rename it.
#
# Name: NCCS INTEL SLURM
#
# submit: /usr/slurm/bin/sbatch
# display: /usr/slurm/bin/squeue
# job regexp: (\d+)
# cancel: /usr/slurm/bin/scancel JOB_ID_TAG
# show num_nodes: no
#
#SBATCH --account=k3001
#SBATCH --nodes=NUM_NODES_TAG
#SBATCH --ntasks-per-node=PROCS_PER_NODE_TAG
#SBATCH --time=WALL_CLOCK_LIMIT_TAG
#SBATCH --job-name="ddt"
#SBATCH --qos=debug
#SBATCH --export=NONE
#SBATCH --constraint=sand
#SBATCH --output=/discover/nobackup/gbritzol/NewLdasExp/allinea-%j.out
#SBATCH --error=/discover/nobackup/gbritzol/NewLdasExp/allinea-%j.err

cd /discover/nobackup/gbritzol/NewLdasExp/run

/usr/local/other/policeme/policeme.exe -d /discover/nobackup/gbritzol/NewLdasExp

limit stacksize unlimited
source ../build/Linux/bin/g5_modules
module load tool/allinea-tools-5.0.1
setenv MKL_CBWR SSE4_2 # ensure zero-diff across archs
setenv MV2_ON_DEMAND_THRESHOLD 8192 # MVAPICH2

DDT_MPIRUN_TAG DDT_DEBUGGER_ARGUMENTS_TAG MPI_ARGUMENTS_TAG PROGRAM_TAG  -work_path ../outpu

 -run_path . \
 -exp_id NewLdasExp \
 -exp_domain SMAP_EASEv2_M09_GLOBAL \
 -start_year 2014 -start_month  5 -start_day  1 -start_hour  0 -start_min  0 -start_sec  0 \
 -end_year   2014 -end_month    5 -end_day    1 -end_hour   15 -end_min    0 -end_sec    0 \
 -N_ens 24 \
 -spin .false. \
 -force_dtstep 3600 \
 -first_ens_id 0 \
 -resolution SMAP_EASEv2_M09 \
 -restart .true. \
 -restart_path ../input/restart/SMAP_Nature_v03 \
 -restart_domain SMAP_EASEv2_M09_GLOBAL \
 -restart_id SMAP_Nature_v03 \
 -tile_coord_file SMAP_EASEv2_M09_3856x1624.til \
 -met_tag e5110_fp__precCPCUG5FPv1 \
 -met_path ../input/met_forcing/GEOS5_land_forcing \
 -ens_prop_inputs_path . \
 -ens_prop_inputs_file m3_15_assim_test-ensprop_v01.nml \
 -driver_inputs_path . \
 -driver_inputs_file m3_15_assim_test-driver_v01-RRpaths.nml \
 -ens_und_inputs_path . \
```

Translate Allinea's Tags to Real Parameters

Allinea Tag	Real value
NUM_NODES_TAG	Number of Nodes
PROCS_PER_NODE_TAG	Number of Processes per Node
WALL_CLOCK_LIMIT_TAG	Processing time
DDT_MPI_RUN	mpirun/mpiexec/mpiexec.hydra
DDT_DEBUGGER_ARGUMENTS_TAG	Debugger's arguments
MPI_ARGUMENTS_TAG	The arguments that the mpirun needs
PROGRAM_TAG	The path for the executable

Header of the qtf:

```
# submit: /usr/slurm/bin/sbatch
# display: /usr/slurm/bin/squeue
# job regexp: (\d+)
# cancel: /usr/slurm/bin/scancel JOB_ID_TAG
# show num_nodes: no
#
```

Set Up MAP with GMAO's Models

- The Allinea setup to use MAP is similar to the DDT procedure except when the code is built it needs to be build with options:**BOPT=Og APROF=y**
- Install your experiment test case by running ldsetup.py (or ldsetup) script if this is not done during the debug process
- The qtf is the same as the one in the debug step
- Remaining settings are the same as in the debug procedure

MAP Profile Using LDAS

/gpfsm/dhome/gbritzol/LDASsa_mpi_112p_1t_2015-04-13_10-37.map - Allinea MAP - Allinea Forge 5.0.1

Profiled: LDASsa_mpi.x on 112 processes, 112 cores (1 per process) Started: Mon Apr 13 10:37:14 2015 Runtime: **976s** Hide Metrics...

Main thread activity

CPU floating-point (%) 0.0 - 100.0 (13.7 avg)

Memory usage (MB) 44.56 - 4,883 (368.5 avg)

10:37:14-10:53:28 (974.315s): Main thread compute **43** %, MPI **57** %, File I/O **0** %, Sleeping **0** % | CPU floating-point **13.7** %; Memory usage **368.5** MB; Metrics Select All

clsm_ensdrv_main.F90 | ens_driver_routines.F90 | clsm_ensdrv_pert_routines.F90 | process_cat.F90

```
1
2  program CLSM_ens_driver
3
4    !*********************************************************************
5    !
6    !  Ensemble driver for the NASA Catchment land surface model
7    !
8    !  Rolf Reichle, 23 Aug 2005
9    !
10   !*********************************************************************
11   !
12   ! runtime options are read in three steps:
13   !
14   ! 1.) read options from default namelist files called
15   !     clsm*default*inputs.nml in run path (run path specified at command line)
```

Input/Output | Project Files | Main Thread Stacks | Functions

Main Thread Stacks

Total core time	MPI	Function(s) on line	Source	Position
		LDASsa_mpi.x [program]		
		clsm_ens_driver	program CLSM_ens_driver	clsm_ensdrv_main.F90:2
		ENS_DRIVER_ROUTINES::propagate...	call propagate_ensemble(&	clsm_ensdrv_main.F90:1869
		PROCESS_CAT::propagate_cat	call propagate_cat(&	ens_driver_routines.F90:322
12.1%		CATCHMENT_MODEL::catchment	call catchment(N_cat, tile_coord%com_lon, tile_coord%com_lat, &	process_cat.F90:559
4.6%		PROCESS_CAT::sfc_turb	call sfc_turb(N_cat, met_force, cat_progn, cat_diagS, zol, dze, veg_param%lai, &	process_cat.F90:300
1.2%		CATCHMENT_MODEL::catch_calc...	call catch_calc_soil_moist(&	process_cat.F90:237
7.3%		136 others		
2.3%		CLSM_ENSDRV_PERT_ROUTINES:...	call apply_progn_pert(N_progn_pert, N_catd, &	ens_driver_routines.F90:343
1.5%		CLSM_ENSDRV_PERT_ROUTINES:...	call apply_force_pert(N_force_pert, N_catd, &	ens_driver_routines.F90:247
0.8%		14 others		
13.1%	8.5%	CLSM_ENSDRV_PERT_ROUTINES::i...	call initialize_perturbations(&	clsm_ensdrv_main.F90:1497
11.7%	11.6%	CLSM_ENSDRV_PERT_ROUTINES::i...	call io_pert_rstrt('w', work_path, exp_id, ens_id(n_e), &	clsm_ensdrv_main.F90:2247

Showing data from 112,000 samples taken over 112 processes (1000 per process) Allinea Forge 5.0.1 Main Thread View

Example – High Memory Usage Detection

/gpfsm/dhome/gbritzol/LDASsa_mpi_112p_1t_2015-04-13_10-37.map - Allinea MAP - Allinea Forge 5.0.1

Profiled: LDASsa_mpi.x on 112 processes, 112 cores (1 per process) Started: Mon Apr 13 10:37:14 2015 Runtime: **976s**

Hide Metrics...

Main thread activity

CPU floating-point (%)
0.0 - 100.0 (13.7 avg)
0.0 - 40 (0.1 avg)

Memory usage (MB)
44.56 - 4,883 (368.5 avg)
167.8 - 4,883 (250.2 avg)

10:38:16-10:39:51 (95.579s, 9.9% of total): Main thread compute **33** %, MPI **67** %, File I/O **0** %, Sleeping 0 % | CPU floating-point **0.1** %; Memory usage **250.2** MB;

Metrics Select All

clsm_ensdrv_main.F90 clsm_ensdrv_pert_routines.F90

```
1490        !
1491        ! initialize random seed and perturbation fields
1492
1493        if (N_force_pert>0 .or. N_progn_pert>0) then
1494
1495           ! initialize random seed and perturbations
1496
31.9%  1497        call initialize_perturbations(                          &
1498             N_catl, N_ens, ens_id, start_time,                        &
1499             restart_path, restart_domain, restart_id, work_path, exp_id,  &
1500             tile_coord_l, pert_grid_f, pert_grid_l,                   &
1501             N_force_pert, N_progn_pert,                               &
1502             force_pert_param, progn_pert_param,                       &
1503             Pert_rseed, Force_pert_ntrmdt_l, Progn_pert_ntrmdt_l,     &
1504             Force_pert_tile_new, Force_pert_tile_old,                 &  (...)
```

Input/Output | Project Files | Main Thread Stacks | Functions

Main Thread Stacks

Total core time	MPI	Function(s) on line	Source	Position
		LDASsa_mpi.x [program]		
		clsm_ens_driver	program CLSM_ens_driver	clsm_ensdrv_main.F90:2
42.2%	42.2%	mpi_scatterv_	call MPI_SCATTERV(&	clsm_ensdrv_main.F90:1179
31.9%	1.0%	CLSM_ENSDRV_PERT_ROUTINES::i...	call initialize_perturbations(&	clsm_ensdrv_main.F90:1497
10.0%	10.0%	CLSM_ENSDRV_VEGALB_ROUTINE...	call get_ALB(N_catg, N_catf, N_catl, f2g, N_catl_vec, low_ind, &	clsm_ensdrv_main.F90:1371
7.6%	6.6%	CLSM_ENSDRV_VEGALB_ROUTINE...	call get_VEG('GRN', N_catg, N_catf, N_catl, f2g, N_catl_vec, low_ind, &	clsm_ensdrv_main.F90:1361
4.0%	4.0%	CLSM_ENSDRV_VEGALB_ROUTINE...	call get_VEG('LAI', N_catg, N_catf, N_catl, f2g, N_catl_vec, low_ind, &	clsm_ensdrv_main.F90:1366
2.0%	2.0%	CLSM_ENSDRV_VEGALB_ROUTINE...	call get_veg_and_alb_times(N_catg, N_catf, res_ftag, &	clsm_ensdrv_main.F90:1343
1.5%	1.5%	mpi_scatterv_	call MPI_SCATTERV(&	clsm_ensdrv_main.F90:1399
0.9%		2 others		

Showing data from 11,088 samples taken over 112 processes (99 per process)

Allinea Forge 5.0.1 Main Thread View

Example – High Floating Point Activity

/gpfsm/dhome/gbritzol/LDASsa_mpi_112p_1t_2015-04-13_10-37.map – Allinea MAP – Allinea Forge 5.0.1

Profiled: LDASsa_mpi.x on 112 processes, 112 cores (1 per process) Started: Mon Apr 13 10:37:14 2015 Runtime: **976s**

Hide Metrics...

Main thread activity

CPU floating-point (%)
0.0 - 100.0 (13.7 avg)
0.0 - 100.0 (39.3 avg)

Memory usage (MB)
44.56 - 4,883 (368.5 avg)
296.6 - 1,351 (396.6 avg)

10:41:54-10:42:23 (28.284s, 3.0% of total): Main thread compute **88** %, MPI **12** %, File I/O **0** %, Sleeping **0** % | CPU floating-point **39.3** %; Memory usage **396.6** MB;

Metrics Select All

clsm_ensdrv_main.F90 | clsm_ensdrv_pert_routines.F90

```
1
2  program CLSM_ens_driver
3
4    !*********************************************************************
5    !
6    !  Ensemble driver for the NASA Catchment land surface model
7    !
8    !  Rolf Reichle, 23 Aug 2005
9    !
10   ! *********************************************************************
11   !
12   ! runtime options are read in three steps:
13   !
14   ! 1.) read options from default namelist files called
15   !      clsm*default*inputs.nml in run path (run path specified at command line)
```

Input/Output | Project Files | Main Thread Stacks | Functions

Main Thread Stacks

Total core time	MPI	Function(s) on line	Source	Position
		LDASsa_mpi.x [program]		
		clsm_ens_driver	program CLSM_ens_driver	clsm_ensdrv_main.F90:2
78.5%		ENS_DRIVER_ROUTINES::propagate...	call propagate_ensemble(&	clsm_ensdrv_main.F90:1869
6.3%	6.3%	mpi_scatterv_	call MPI_SCATTERV(&	clsm_ensdrv_main.F90:1722
4.4%	4.4%	mpi_scatterv_	call MPI_SCATTERV(&	clsm_ensdrv_main.F90:1727
3.7%		[partial traces]		
1.7%		CLSM_ENSDRV_DRV_ROUTINES::ze...	call zenith(&	clsm_ensdrv_main.F90:1742
1.5%		CATCH_TYPES::add_cat_progn	cat_progn_ensavg(i) = cat_progn_ensavg(i) + cat_progn(i,n_e)	clsm_ensdrv_main.F90:1840
1.2%		CLSM_ENSDRV_DRV_ROUTINES::b...	call balance_calcs('ini', N_cat1, real(model_dtstep), &	clsm_ensdrv_main.F90:1848
2.7%	0.9%	8 others		

Showing data from 3,360 samples taken over 112 processes (30 per process)

Allinea Forge 5.0.1 Main Thread View

Real-time Demo with DDT/MAP for LDAS Model

- Demonstrate how to use use the debugger (i.e., submit a job, step through the code, view variables)
- Some hints on how to use the profiler:
  - Submit a job
  - Detect hot spots in the code
  - View the appropriate code section for each hot spot