EMOD MPI Test Report

Zhaowei Du

(April 2024)

1 Table of Contents

1   Overview

The following is a typical command to run simulation with EMOD on SLURM:

singularity exec <user_sif_file> Assets/Eradication --config config.json --dll-path ./Assets --input-path ./Assets\;.

NYU user would like to efficiently use system cores to improve performance, SLURM has several related parameters, such as the following, user is not sure how to choose from:

#SBATCH --nodes=3
#SBATCH --ntasks=5
#SBATCH --cpus-per-task=2
#SBATCH --ntasks-per-core=4

Instead of using above SLURM parameters, user did a test by inserting '**mpirun -n #**' to the above command as follows, and wonders if it is a good solution:

singularity exec <user_sif_file> mpirun -n 4 Assets/Eradication --config config.json --dll-path ./Assets --input-path ./Assets\;.

This document is a report based on my test on all kinds of scenarios. All the tests have been done on NU SLURM system.

We run the **same** experiment/simulations in different scenarios.

Note: our running test's demographic file has **node_count = 3**.

The following is the normal case result without '**mpirun -n #**' being inserted:

📁 [..]		4/17/2024 1:18:47 PM
InsetChart.json	7,862 KB	4/17/2024 1:35:46 PM
PropertyReport.json	1,027 KB	4/17/2024 1:35:46 PM
ReportEventRecorder.csv	9,079 KB	4/17/2024 1:35:43 PM
SpatialReport_Adult_Vectors.bin	257 KB	4/17/2024 1:35:46 PM
SpatialReport_Daily_Bites_Per_Human.bin	257 KB	4/17/2024 1:35:46 PM
SpatialReport_Population.bin	257 KB	4/17/2024 1:35:46 PM
SpatialReport_Prevalence.bin	257 KB	4/17/2024 1:35:46 PM
state-21900.dtk	58,360 KB	4/17/2024 1:35:45 PM
transitions.json	1 KB	4/17/2024 1:18:47 PM

We will use this to compare the results generated from each scenarios test.

2   Scenarios and results

On NU system, tested the following scenarios.

2.1   Test Cases

(1)   singularity exec <user_sif_file> mpirun -n # Assets/Eradication --config
config.json --dll-path ./Assets --input-path ./Assets\;.

Case: mpirun -n 1
Case: mpirun -n 2
Case: mpirun -n 3
Case: mpirun -n 4

Note: tested mpiexec as well.

(2)   mpirun -n # singularity exec <user_sif_file> Assets/Eradication --config
config.json --dll-path ./Assets --input-path ./Assets\;.

Case: mpirun -n 1
Case: mpirun -n 2
Case: mpirun -n 3
Case: mpirun -n 4

Note: tested mpiexec as well.

(3)   COMPS approach
#!/usr/bin/bash
#SBATCH --ntasks=<Number of instances to run>
Srun –mpi=pmi2 singularity exec <user_sif_file> Assets/Eradication --config config.json --
dll-path ./Assets --input-path ./Assets\;.

Case: #SBATCH --ntasks=1
Case: #SBATCH --ntasks=2
Case: #SBATCH --ntasks=3
Case: #SBATCH --ntasks=4

Remark: the tests show that all above approaches work and behavior the same way.
Refer to the next Section for details.

2.2 Test Results

Expected results:

|			
⬆️ ..			4/17/2024 1:18:47 PM
InsetChart.json		7,862 KB	4/17/2024 1:35:46 PM
PropertyReport.json		1,027 KB	4/17/2024 1:35:46 PM
ReportEventRecorder.csv		9,079 KB	4/17/2024 1:35:43 PM
SpatialReport_Adult_Vectors.bin		257 KB	4/17/2024 1:35:46 PM
SpatialReport_Daily_Bites_Per_Human.bin		257 KB	4/17/2024 1:35:46 PM
SpatialReport_Population.bin		257 KB	4/17/2024 1:35:46 PM
SpatialReport_Prevalence.bin		257 KB	4/17/2024 1:35:46 PM
state-21900.dtk		58,360 KB	4/17/2024 1:35:45 PM
transitions.json		1 KB	4/17/2024 1:18:47 PM

2.2.1 Cases: mpirun -n 1 & #SBATCH --ntasks=1

Results
- Run succeeded.
- Output as expected.

Example output

⬆️ ..			4/17/2024 12:56:15 PM
InsetChart.json		7,862 KB	4/17/2024 1:12:07 PM
PropertyReport.json		1,027 KB	4/17/2024 1:12:07 PM
ReportEventRecorder.csv		9,079 KB	4/17/2024 1:12:03 PM
SpatialReport_Adult_Vectors.bin		257 KB	4/17/2024 1:12:07 PM
SpatialReport_Daily_Bites_Per_Human.bin		257 KB	4/17/2024 1:12:07 PM
SpatialReport_Population.bin		257 KB	4/17/2024 1:12:07 PM
SpatialReport_Prevalence.bin		257 KB	4/17/2024 1:12:07 PM
state-21900.dtk		58,360 KB	4/17/2024 1:12:06 PM
transitions.json		1 KB	4/17/2024 12:56:15 PM

stdout.txt

...
00:15:48 [0] [I] [Simulation] Update(): Time: 21900.0 Rank: 0 StatPop: 3310 Infected: 397
00:15:48 [0] [I] [SerializedPopulation] Writing state to 'output/state-21900.dtk'
00:15:51 [0] [I] [Simulation] Finalizing 'InsetChart.json' reporter.
00:15:52 [0] [I] [Simulation] Finalized 'InsetChart.json' reporter.
00:15:52 [0] [I] [Simulation] Finalizing 'PropertyReport.json' reporter.
00:15:52 [0] [I] [Simulation] Finalized 'PropertyReport.json' reporter.

00:15:52 [0] [I] [Simulation] Finalizing 'SpatialReport' reporter.
00:15:52 [0] [I] [Simulation] Finalized 'SpatialReport' reporter.
00:15:52 [0] [I] [Simulation] Finalizing 'ReportEventRecorder.csv' reporter.
00:15:52 [0] [I] [Simulation] Finalized 'ReportEventRecorder.csv' reporter.
00:15:52 [0] [I] [Controller] Exiting execute_internal
00:15:52 [0] [I] [Eradication] Controller executed successfully.

2.2.2 Cases: mpirun -n 2 & #SBATCH --ntasks=2

Results:
- Run succeeded.
- Less time.
- Differrent output.

Example output

Name	Size	Changed
↑		4/17/2024 12:19:40 PM
InsetChart.json	7,866 KB	4/17/2024 12:29:36 PM
PropertyReport.json	1,027 KB	4/17/2024 12:29:36 PM
ReportEventRecorder.csv	9,209 KB	4/17/2024 12:29:35 PM
SpatialReport_Adult_Vectors.bin	257 KB	4/17/2024 12:29:36 PM
SpatialReport_Daily_Bites_Per_Human.bin	257 KB	4/17/2024 12:29:36 PM
SpatialReport_Population.bin	257 KB	4/17/2024 12:29:36 PM
SpatialReport_Prevalence.bin	257 KB	4/17/2024 12:29:36 PM
state-21900-000.dtk	41,513 KB	4/17/2024 12:29:36 PM
state-21900-001.dtk	41,478 KB	4/17/2024 12:29:36 PM
transitions.json	1 KB	4/17/2024 12:19:41 PM

stdout.txt
...
00:09:54 [1] [I] [Simulation] Update(): Time: 21900.0 Rank: 1 StatPop: 1676 Infected: 328
00:09:54 [0] [I] [SerializedPopulation] Writing state to 'output/state-21900-000.dtk'
00:09:54 [1] [I] [SerializedPopulation] Writing state to 'output/state-21900-001.dtk'
00:09:55 [0] [I] [Simulation] Finalizing 'InsetChart.json' reporter.
00:09:55 [0] [I] [Simulation] Finalized 'InsetChart.json' reporter.
00:09:55 [0] [I] [Simulation] Finalizing 'PropertyReport.json' reporter.
00:09:55 [0] [I] [Simulation] Finalized 'PropertyReport.json' reporter.
00:09:55 [0] [I] [Simulation] Finalizing 'SpatialReport' reporter.
00:09:55 [1] [I] [Controller] Exiting execute_internal
00:09:55 [1] [I] [Eradication] Controller executed successfully.
00:09:55 [0] [I] [Simulation] Finalized 'SpatialReport' reporter.
00:09:55 [0] [I] [Simulation] Finalizing 'ReportEventRecorder.csv' reporter.
00:09:55 [0] [I] [Simulation] Finalized 'ReportEventRecorder.csv' reporter.

00:09:55 [0] [I] [Controller] Exiting execute_internal
00:09:55 [0] [I] [Eradication] Controller executed successfully.

2.2.3 Cases: mpirun -n 3 & #SBATCH --ntasks=3

Results
- Run failed.
- Different output.

Example output

↰			4/17/2024 12:38:20 PM
ReportEventRecorder.csv		9,081 KB	4/17/2024 12:47:59 PM
SpatialReport_Adult_Vectors.bin		257 KB	4/17/2024 12:47:57 PM
SpatialReport_Daily_Bites_Per_Human.bin		257 KB	4/17/2024 12:47:57 PM
SpatialReport_Population.bin		257 KB	4/17/2024 12:47:57 PM
SpatialReport_Prevalence.bin		257 KB	4/17/2024 12:47:57 PM
state-21900-000.dtk		41,308 KB	4/17/2024 12:48:01 PM
state-21900-001.dtk		40,659 KB	4/17/2024 12:48:01 PM
state-21900-002.dtk		23,857 KB	4/17/2024 12:48:00 PM
transitions.json		1 KB	4/17/2024 12:38:20 PM

stdout.txt

…
00:09:41 [0] [E] [Eradication]

IllegalOperationException:
Exception in baseReportLib/ChannelDataMap.cpp at 248 in Reduce.
Channel=Variant Fraction-PfEMP1 Major from rank=2 had size=16331 while this rank=0 had size=21900

…
00:09:41 [1] [E] [Eradication]

IllegalOperationException:
Exception in baseReportLib/ChannelDataMap.cpp at 248 in Reduce.
Channel=Variant Fraction-PfEMP1 Major from rank=2 had size=16331 while this rank=1 had size=21900

…
00:09:41 [2] [E] [Eradication]

IllegalOperationException:

Exception in baseReportLib/ChannelDataMap.cpp at 248 in Reduce.

Channel=Variant Fraction-PfEMP1 Major from rank=0 had size=21900 while this rank=2 had size=16331

2.2.4 Cases: mpirun -n 4 & #SBATCH --ntasks=4

Results:
- Run failed.
- No output.

stdout.txt

...

00:00:01 [1] [I] [Simulation] Rank map contents not displayed until NodeRankMap::ToString() (re)implemented.

00:00:01 [2] [I] [Simulation] Merged rank 2 map now has 3 nodes.

00:00:01 [2] [I] [Simulation] Rank map contents not displayed until NodeRankMap::ToString() (re)implemented.

00:00:01 [2] [I] [Simulation] Initialized 'InsetChart.json' reporter

00:00:01 [3] [I] [Simulation] Merged rank 3 map now has 3 nodes.

00:00:01 [3] [I] [Simulation] Rank map contents not displayed until NodeRankMap::ToString() (re)implemented.

00:00:01 [3] [W] [Simulation] Rank 3 wasn't assigned any nodes! (# of procs is too big for simulation?)

00:00:01 [3] [I] [Eradication] Controller execution failed, exiting.

00:00:01 [0] [I] [Simulation] Merged rank 0 map now has 3 nodes.

00:00:01 [0] [I] [Simulation] Rank map contents not displayed until NodeRankMap::ToString() (re)implemented.

00:00:01 [0] [I] [Simulation] Initialized 'InsetChart.json' reporter

00:00:01 [0] [I] [Simulation] Initialized 'PropertyReport.json' reporter

00:00:01 [1] [I] [Simulation] Initialized 'InsetChart.json' reporter

00:00:01 [1] [I] [Simulation] Initialized 'PropertyReport.json' reporter

00:00:01 [2] [I] [Simulation] Initialized 'PropertyReport.json' reporter

00:00:01 [2] [I] [Simulation] 00:00:01 [0] [I] [Simulation] Initialized 'SpatialReport' reporter

00:00:01 [1] [I] [Simulation] Initialized 'SpatialReport' reporter

00:00:01 [1] [I] [Simulation] Initialized 'ReportEventRecorder.csv' reporter

00:00:01 [1] [I] [SimulationEventContext] Time for campaign event. Calling Dispatch...

00:00:01 [1] [I] [SimulationEventContext] 1 node(s) visited.

00:00:01 [1] [I] [SimulationEventContext] Time for campaign event. Calling Dispatch...

00:00:01 [1] [I] [SimulationEventContext] 1 node(s) visited.

00:00:01 [1] [I] [SimulationEventContext] Time for campaign event. Calling Dispatch...

00:00:01 [1] [I] [SimulationEventContext] 1 node(s) visited.

00:00:01 [1] [I] [StandardEventCoordinator] UpdateNodes() distributed 'NodeLevelHealthTriggeredIV' intervention to node 2

00:00:01 [1] [I] [StandardEventCoordinator] UpdateNodes() distributed 'NodeLevelHealthTriggeredIV' intervention to node 2
00:00:01 [1] [I] [StandardEventCoordinator] UpdateNodes() distributed 'NodeLevelHealthTriggeredIV' intervention to node 2
Initialized 'SpatialReport' reporter
00:00:01 [2] [I] [Simulation] Initialized 'ReportEventRecorder.csv' reporter
00:00:01 [2] [I] [SimulationEventContext] Time for campaign event. Calling Dispatch...
00:00:01 [2] [W] [SimulationEventContext] No nodes were visited. No nodes were added to the event coordinator.
00:00:01 [2] [I] [SimulationEventContext] Time for campaign event. Calling Dispatch...
00:00:01 [2] [W] [SimulationEventContext] No nodes were visited. No nodes were added to the event coordinator.
00:00:01 [2] [I] [SimulationEventContext] Time for campaign event. Calling Dispatch...
00:00:01 [2] [W] [SimulationEventContext] No nodes were visited. No nodes were added to the event coordinator.

2.3 Special test

This is a normal test without invoking MPI manually, but with more SLURM directives setup:

#SBATCH --nodes=2
#SBATCH --ntasks=3
#SBATCH --cpus-per-task=2
#SBATCH --ntasks-per-core=3
#SBATCH --mem-per-cpu=8192

[zdf1921@quser34]$ sacct -X -P --format='jobid,state,elapsed, reqmem, reqcpus, ncpus' -j 8662087
JobID|State|**Elapsed**|ReqMem|ReqCPUS|NCPUS
8662087|COMPLETED|**00:13:47**|48G|6|6

output

⬆ ..		4/19/2024 1:00:39 PM
InsetChart.json	7,862 KB	4/19/2024 1:14:21 PM
PropertyReport.json	1,027 KB	4/19/2024 1:14:21 PM
ReportEventRecorder.csv	9,079 KB	4/19/2024 1:14:18 PM
SpatialReport_Adult_Vectors.bin	257 KB	4/19/2024 1:14:21 PM
SpatialReport_Daily_Bites_Per_Human.bin	257 KB	4/19/2024 1:14:21 PM
SpatialReport_Population.bin	257 KB	4/19/2024 1:14:21 PM
SpatialReport_Prevalence.bin	257 KB	4/19/2024 1:14:21 PM
state-21900.dtk	58,360 KB	4/19/2024 1:14:20 PM
transitions.json	1 KB	4/19/2024 1:00:39 PM

Note: there is no big performance difference when running single simulation with just default settings.

3   Summary

Based on the testing results, we get a feeling that

- Seems like Eradication works differently with MPI cores #.
  Q1: what do we want to achieve with MPI cores in EMOD?
  Q2: should the results depend on (or be affected by) the MPI cores?
  Q3: how to use the results if the results got affected by MPI cores #?

- Not sure User's suggestion is acceptable.

- We can't improve single simulation performance dramatically. Perhaps we may gain some overall performance for all simulations execution by choosing more nodes, mem, or cpus,…

What do you think?