BFS Application

Requirements

- Compiler with std c++11 support (e.g GNU C/C++ v4.8.1+)
- UNIX-like OS with CUDA 6+ support

Installation, Setup and Execution of the BFS code

Downloading

Create an account in https://bitbucket.org Request access for bfs_multinode repository

Available repositories are: (Replace the token Your_Bitbucket_User)

- https://Your_Bitbucket_User@bitbucket.org/julianromera/bfs_multinode.git (Updated daily)
- https://Your_Bitbucket_User@bitbucket.org/jyoung3131/bfs_multinode.git (Updated weekly)

```
$ REPLACE the token Your_Bitbucket_User
$ git clone https://Your_Bitbucket_User@bitbucket.org/julianromera/bfs_multinode.git
$ git checkout -b architectural_tuning
$ git pull origin architectural_tuning
```

Setting up the environment

1. Add CUDA libraries and binaries to your path:

Add this text to your ~/.bashrc

```
UPDATE NEXT LINE
export CUDA_PATH=/usr/local/cuda-7.0
export PATH=$CUDA_PATH/bin:$PATH
export LD_LIBRARY_PATH=$CUDA_PATH/lib64:$CUDA_PATH/lib:$CUDA_PATH/extras/CUPT
```

1. In case of using EXTOLL add these lines to your ~/.bashrc

```
Adjust next line to your actual path
export EXTOLL_HOME=/extoll2
if [ -d /extoll2 ]; then
        source $EXTOLL_HOME/extoll_env.bash
        # ensure this path is correct
        export MPI_PATH=$EXTOLL_HOME/mpi/openmpi-1.6.4
        export PATH=$MPI_PATH/bin:$PATH

        export LD_LIBRARY_PATH=$EXTOLL_HOME/lib:$MPI_PATH/lib:$LD_LIBRARY_PATH
        ulimit -l 4194304
        ulimit -s unlimited
fi
```

Building

The code to compile is in the folder `cpu_2d/`. It is built using `Make`.

The currently available Makefiles are:

- Makefile.gcc
- Makefile.gcc.keeneland

The Makefiles in the list below are currently being developed:

- Makefile.Intel
- Makefile.cpu

The file `Makefile` is a symbolic link to the Makefile.___ being used. It may be created/edited using:

```
$ cd bfs_multinode/cpu_2d
$ rm -f Makefile
$ ln -s Makefile.gcc Makefile
```

```
$ cd bfs_multinode/cpu_2d
$ make
$ cd ../eval
```

Makefile Configuration

The Makefiles in the `cpu_2d/` folder are configurable. Edit the Makefile file to access the documentation for the variables. The documentation is placed in the header of the file.

The default variables/ values for Makefile.gcc.Keeneland are:

```
nvidia_architecture                    :="auto"
nvidia_ptxas_otimize                   :="no"
manual_profiler_cuda                   :="no"
manual_profiler_other_compilers        :="yes"
openmp_on_cuda                         :="no"
openmp_on_other_compilers              :="no"
custom_openmpi                         :="no"
custom_openmpi_basedir                 :=/home/CHANGE_ME/openmpi
scorep_profiler_enable                 :="no"
scorep_profiler_automatic_instrumentation :="yes"
scorep_custom                          :="yes"
scorep_custom_basedir                  :=/home/CHANGE_ME/scorep
use_avx_instructions                   :="yes"
enable_simd_compression                :="yes"
enable_simd_compression_benchmark      :="no"
debug                                  :="no"
quiet_output                           :="no"
use_cuda                               :="yes"
code_optimization_level                :="O4"
code_optimization_flags                :=-funroll-loops -flto
```

Test Scenarios

Test scenarios are in the folder `eval/`

SBatch relevant Test-cases in the previous list are:

- o4p2n-roptim.rsh [SCALE_FACTOR]
- o4p2n-coptim.rsh [SCALE_FACTOR]
- o4p2n-noptim.rsh [SCALE_FACTOR]
- o9p8n.rsh [SCALE_FACTOR]

- o16p8n.rsh [SCALE_FACTOR]

The name in these files explains what they do. For example:

The part "4p2n" indicates 4 Processes will be run in 2 Nodes. These scripts are run through Slurm. "16p8n" would indicate 16 processes distributed in 8 Nodes.

Runing tests

Ensure that the Slurm queue is empty

```
$ squeue
JOBID PARTITION NAME USER ST TIME NODES NODELIST(REASON)
```

Use Sbatch to execute the .rsh file. Add a numeric argument at the end (Sf)

```
$ sbatch o4p2n_roptim.rsh 21
Submitted batch job 423
$ sbatch o4p2n_coptim.rsh 21vv
Submitted batch job 424
$ sbatch o4p2n_noptim.rsh 21
Submitted batch job 425
```

Automation Scripts

List of scripts

- r-ify.sh
- r-compare.sh
- check-all.sh
- scorep_install.sh

Installing scripts

Scripts are located under the `scripts/` folder

To install

- r-ify.sh
- r-compare.sh
- check-all.sh

Run:

```
$ cd bfs_multinode
$ # Make a symlink from eval to the script file
$ cd eval
$ ln -s ../scripts/r-ify.sh r-ify.sh
$ ln -s ../scripts/r-compare.sh r-compare.sh
$ ln -s ../scripts/check-all.sh check-all.sh
$ chmod u+x *.sh
```

To install (Useful for Profiling and Tracing)

- scorep_install.sh

Run:

```
cp scripts/scorep_install.sh ~
cd ~
chmod u+x scorep_install.sh
./scorep_install.sh
...
```

Script r-ify.sh

Description

It uses the execution traces to generate R-code. This R-code (once run) shows the time meassurements of the Phases of the BFS code. This script uses result files with same Scale Factor. The results are represented as a Barplot.

Execution

```
$ ./r-ify.sh 423 424 425

-> File slurm-423.out. Validation passed (Tasks: 4, GPUS/Task: 1, Scale Factor: 21).
-> File slurm-424.out. Validation passed (Tasks: 4, GPUS/Task: 1, Scale Factor: 21).
-> File slurm-425.out. Validation passed (Tasks: 4, GPUS/Task: 1, Scale Factor: 21).
Enter new labels for the X-Axe? (y/n) [n] y
Enter a total 3 label(s) between quoutes. Separate them with spaces: "4p2n-Roptim" "4p2n-Cop
-> Created file "file-423-424-425.r".
-> R-Code successfully generated.
```

Open the file `file-423-424-425.r` with your R editor and run the code.

Script r-compare.sh

Description

As the previous script, this also uses the execution traces to generate R-code. This differs from the previous one in that it can compare several files from several Scale Factors. Results are visualized as a Lineplot.

Execution

```
$ ./r-ify.sh
(to be completed)
```

Script check-all.sh

Description

This script automatizes the execution of tests for different Scale Factors.

Execution

```
$ check-all.sh 15 30
```

This will run the tests with format `o*.rsh` in the `eval/` folfer for Scale Factors 15 to 30.

Process is shown in ncurses-like format.

Script scorep_install.sh

Description

This script Download, Uncompress, Builds and Install locally Score-P, Scalasca and CUBE, and a compatible version of OpenMPI. This script is configurable. The configuration may be changed editing the script.

Root priviledges are not required to run the script.

Execution

```
$ cd $HOME
$ ./scorep_install.sh
```

Other

Profiling and Tracing

This BFS application allows the code to be instrumented in Zones using Score-P with very low overhead. This requires Score-P and Scalasca to be installed in the system. The results may be analyzed either visually (using CUBE) or through console using `scorep-score`

These tools may be installed locally (No privileged user is needed) using the scorep_install.sh aforementioned.

The names of the instrumentable Zones are listed below. Others may be added if needed.

```
BFSRUN_region_vertexBroadcast
BFSRUN_region_nodesTest
BFSRUN_region_localExpansion
BFSRUN_region_testSomethingHasBeenDone
BFSRUN_region_columnCommunication
BFSRUN_region_rowCommunication
```

The first step is to update the system variables. This may be done either on .bashrc or in a separate script.

Update the paths in the variables below

```
$ cat >> ~/.bashrc << EOF
export G500_ENABLE_RUNTIME_SCALASCA=yes

export SCOREP_CUDA_BUFFER=48M
export SCOREP_CUDA_ENABLE=no
export SCOREP_ENABLE_PROFILING=true
export SCOREP_ENABLE_TRACING=false
export SCOREP_PROFILING_FORMAT=CUBE4
export SCOREP_TOTAL_MEMORY=12M
export SCOREP_VERBOSE=no
export SCOREP_PROFILING_MAX_CALLPATH_DEPTH=330

export LD_LIBRARY_PATH=$HOME/cube/lib:$LD_LIBRARY_PATH
export PATH=$HOME/cube/bin:$PATH

export LD_LIBRARY_PATH=$HOME/scorep/lib:$LD_LIBRARY_PATH
export PATH=$HOME/scorep/bin:$PATH

export LD_LIBRARY_PATH=$HOME/scalasca/lib:$LD_LIBRARY_PATH
export PATH=$HOME/scalasca/bin:$PATH

export MPI_PATH=/home/jromera/openmpi
export PATH=$MPI_PATH/bin:$CUDA_PATH/bin:$PATH

export LD_LIBRARY_PATH=$MPI_PATH/lib:$CUDA_PATHo/lib64:$CUDA_PATHo/lib64/stubs:$
export LD_LIBRARY_PATH=$HOME/scorep/lib:$LD_LIBRARY_PATH
export PATH=$HOME/scorep/bin:$PATH
EOF
```

The variable `G500_ENABLE_RUNTIME_SCALASCA` set to yes will enable the required runtime instrumentor of Scalasca.

Results will be stored on a folder with format `scorep-*` in the `/eval` folder.

To instrument graphically with CUBE run:

```
$ cd eval/scorep-____FOLDER_NAME____
$ cube profile.cubex
```

To instrument through the console run:

```
$ cd eval/scorep-____FOLDER_NAME____
$ scorep-score -r profile.cubex
```

Current Limitations

Out-Of-Memory errors and CUDA memory size limitations:

For some high Score-Factors (e.g 22 as of the day of writing this guide), the resulting Slurm trace will be:

```
Using SCALE-FACTOR 22
Tue Jul 14 15:42:51 CEST 2015
 ========================   JOB MAP   ========================

 Data for node: creek01 Num procs: 2
        Process OMPI jobid: [1404,1] Process rank: 0
        Process OMPI jobid: [1404,1] Process rank: 2

 Data for node: creek02 Num procs: 2
        Process OMPI jobid: [1404,1] Process rank: 1
        Process OMPI jobid: [1404,1] Process rank: 3

 =============================================================
row slices: 2, column slices: 2
graph_generation:               7.749108 s
Input list of edges genereted.
6.710886e+07 edge(s) generated in 8.499692s (7.895447 Medges/s on 4 processor(s))
Adjacency Matrix setup.
2.956432e+06 edge(s) removed, because they are duplicates or self loops.
1.283049e+08 unique edge(s) processed in 18.308235s (7.008041 Medges/s on 4 processor(s
[../b40c/graph/bfs/csr_problem_2d.cuh, 697] CsrProblem cudaMalloc frontier_queues.d_values
[cuda/cuda_bfs.cu, 486] Reset error. (CUDA error 2: out of memory)
MPI_ABORT was invoked on rank 0 in communicator MPI_COMM_WORLD
with errorcode 1.
```

Validation errors for the BFS runs

Looking at your resulting Sbatch trace you may see:

"Validation: failed" / or an output trace similar to the one below:

The issue seems to be related with compiler version errors. The version of the compiler used for the CUDA code may not be met. See the Requirements section above.

```
$ cat slurm-JOBID.out
...
BFS Iteration 24: Finished in 0.111173s
max. local exp.:     0.004383s(3.942371%)
max. queue handling: 0.056909s(51.189917%)
est. rest:           0.049881s(44.867712%)
max. row com.:       0.038369s(34.513278%)
max. col com.:       0.069516s(62.529783%)
max. pred. list. red:0.007413s(6.667911%)
(3:0) Edge [26849(2273):244(244)] with wrong levels: 3 -1
Validation of iteration 24 finished in 0.022564s
Result: Invalid 524279 Edge(s) processed, 4.715888 MTeps
(0:3) Edge [244(244):26849(2273)] with wrong levels: -1 3
(0:3) Edge [244(244):29960(5384)] with wrong levels: -1 4
(3:0) Edge [29960(5384):244(244)] with wrong levels: 4 -1
BFS Iteration 25: Finished in 0.096601s
max. local exp.:     0.000905s(0.937128%)
max. queue handling: 0.048780s(50.496083%)
est. rest:           0.046916s(48.566789%)
max. row com.:       0.032854s(34.009833%)
max. col com.:       0.054057s(55.958921%)
max. pred. list. red:0.009008s(9.324883%)
Validation of iteration 25 finished in 0.020493s
Result: Valid 524280 Edge(s) processed, 5.427272 MTeps
invalid_level
invalid_level
...
```

Troubleshooting

- Problem: In the .out file of Slurm/ Sbatch execution I get the text:

  ```
  S=C=A=N: Abort: No SCOREP instrumentation found in target ../cpu_2d/g500
  ```

- Solution:

  The instrumentation is activated for the runtime execution (i.e: the binary is being run prefixed with scalasca).

  Disable it with:

  ```
  $ export G500_ENABLE_RUNTIME_SCALASCA=no
  ```

About this document

- Version 1.0
- Last revision: 12th August 2015

License

This code contains a subset of Duane Merrill's BC40 repository of GPU-related functions, including his BFS implementation used in the paper, Scalable Graph Traversals.

External Resources

https://www.rstudio.com/products/RStudio/