Intel® Parallel Studio XE 2015 Cluster Edition: Новые возможности библиотеки Intel® MPI 5.0 и Intel® Trace Collector and Analyzer 9.0

Дмитрий Сивков

11 марта 2015

Пакет разработки Intel® Parallel Studio XE 2015 - Помогает создать быстрый код быстрее

Решения для разработок в области HPC

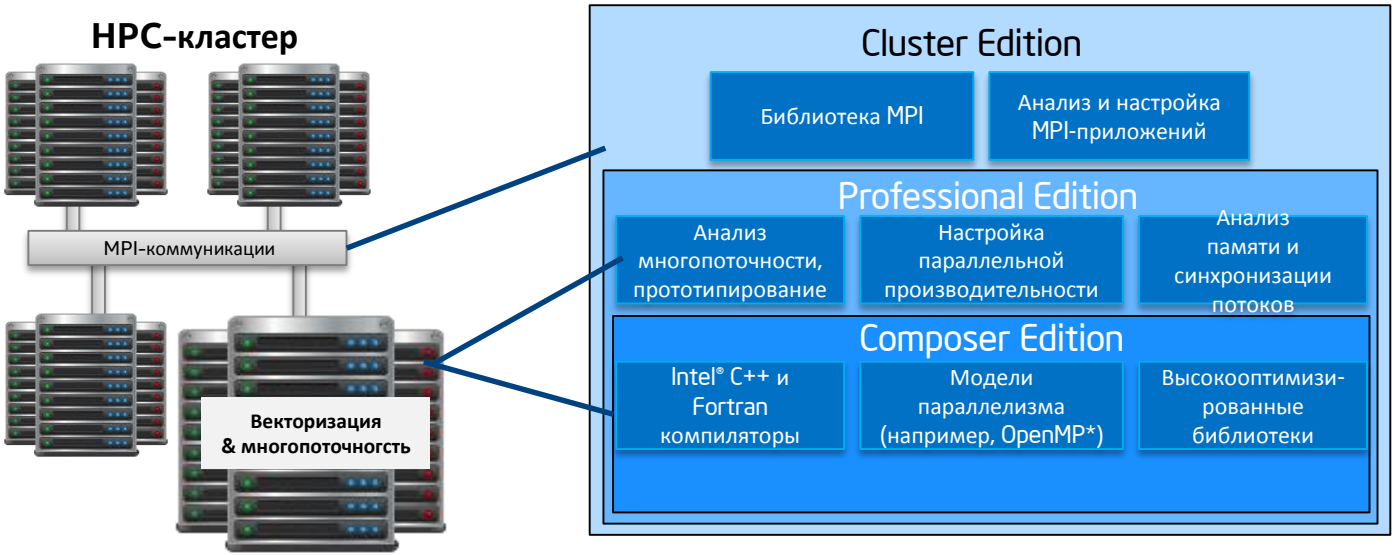
- Более 20 лет на рынке
- Взаимодействие с индустрией при разработке стандартов
- Нацелены на производительность и масштабируемость с аппаратным обеспечением Intel

Отвечает вызовам

- Повышение производительности
- Увеличение масштабируемости
- Увеличение продуктивности

Intel® Parallel Studio XE 2015 Cluster Edition

Помогает создать быстрый код быстрее для HPC



Библиотека Intel® MPI

Возможности

Что	• **Высокопроизводительная реализация MPI от Intel**
Почему	• **Производительность –** настроена на новейшие архитектуры Intel • **Масштабируемость –** готова для классических и многоядерных платформ • **Эффективность –** гибкий выбор фабрик, совместимость
Как	• **Основана на стандартах –** базируется на открытой реализации MPICH **Устойчивая масштабируемость –** Настроена на низкую латентность, большую пропускную способность, большое число процессов • **Поддержка многочисленных фабрик–** Поддерживает популярные высокопроизводительные сетевые фабрики

Библиотека Intel® MPI - обзор

Оптимизация производительности MPI-приложений

- Настройки на конкретное приложение
- Автоматическая настройка

Низкая латентность и поддержка различных вендоров

- Одни из лучших показателей латентности
- Поддержка новейших возможностей OFED и DAPL 2.x

Быстрые коммуникации MPI

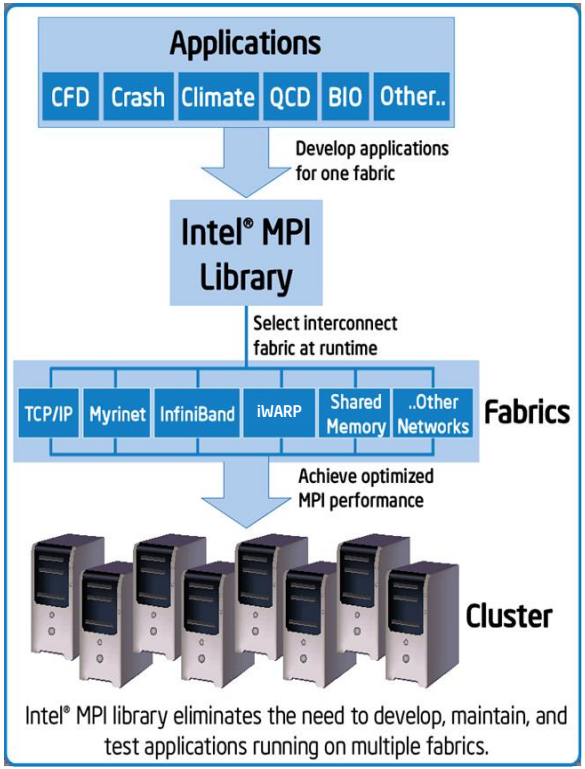
- Оптимизированные коллективные операции

Устойчивая масштабируемость до 262K процессов

- Встроенная поддержка InfiniBand* дает низкую латентность, высокую пропускную способность и снижает требования к используемой памяти

Более надежные MPI-приложения

- Бесшовная интеграция с Intel® Trace Analyzer and Collector



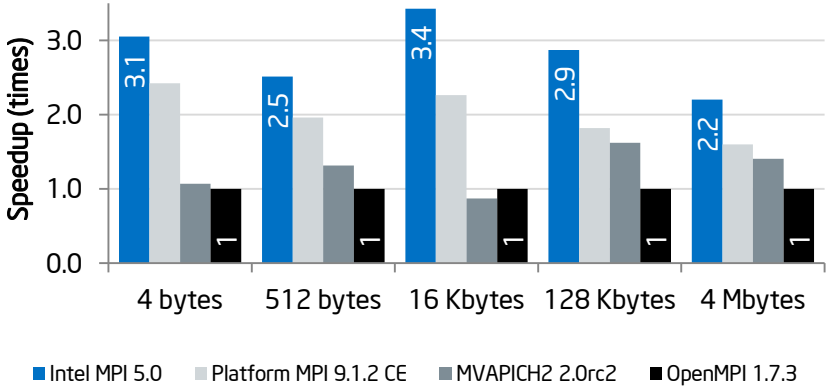
Меньше латентность – выше производительность

Intel® MPI Library

Впечатляющая производительность Intel® MPI Library 5.0

192 процессов, 8 узлов (InfiniBand + shared memory), Linux* 64

Relative (Geomean) MPI Latency Benchmarks (Higher is Better)



Configuration: Hardware: CPU: Dual Intel* Xeon E5-2697v2@2.70Ghz; 64 GB RAM. Interconnect: Mellanox Technologies* MT27500 Family [ConnectX*-3] FDR.. Software: RedHat* RHEL 6.2; OFED 3.5-2; Intel* MPI Library 5.0 Intel* MPI Benchmarks 3.2.4 (default parameters; built with Intel* C++ Compiler XE 13.1.1 for Linux*);

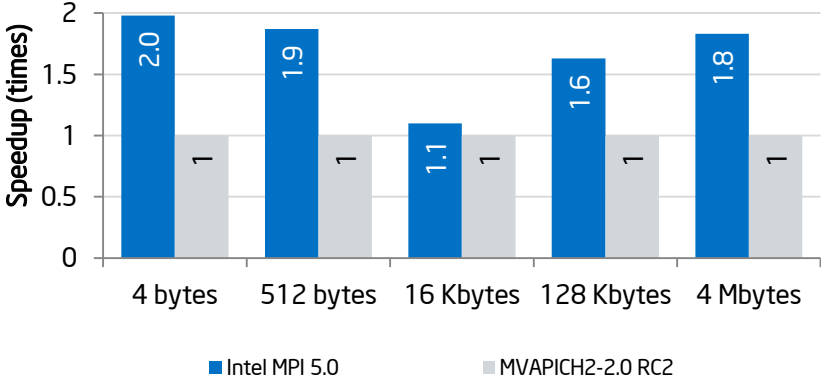
Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. * Other brands and names are the property of their respective owners. Benchmark Source: Intel Corporation

Optimization Notice: Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice. Notice revision #20110804 .

Впечатляющая производительность Intel® MPI Library 5.0

64 процесса, 8 узлов (InfiniBand + shared memory), Linux* 64

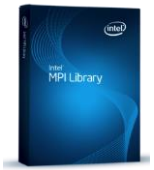
Relative (Geomean) MPI Latency Benchmarks (Higher is Better)



Configuration: Hardware: Intel* Xeon* CPU E5-2680 @ 2.70GHz, RAM 64GB; Interconnect: InfiniBand, ConnectX adapters; FDR. MIC: C0-KNC 1238095 kHz; 61 cores. RAM: 15872 MB per card. Software: RHEL 6.2, OFED 1.5.4.1, MPSS Version: 3.2, Intel* C/C++ Compiler XE 13.1.1, Intel* MPI Benchmarks 3.2.4.;

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. * Other brands and names are the property of their respective owners. Benchmark Source: Intel Corporation

Optimization Notice: Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice. Notice revision #20110804 .



Intel® MPI Library 5.0

Что нового

Поддержка MPI-3

- Неблокирующие коллективы
- Быстрые RMA
- Большие сообщения

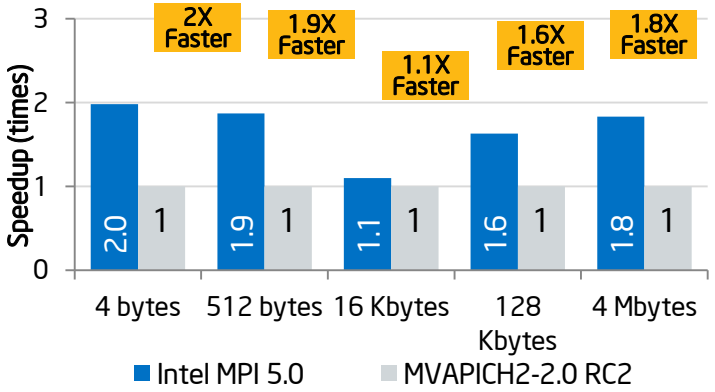
Бинарная совместимость с MPICH ABI

- Совместимость с MPICH* v3.1, IBM* MPI v1.4, Cray* MPT v7.0

Производительность и масштабирование

- Оптимизация потребления памяти
- Масштабирование до 262K ранков*
- До 35% сокращение времени коллективов
- Современный менеджер процессов Hydra для Windows* по умолчанию

Производительность Intel® MPI Library 5.0
64 процесса, 8 узлов(InfiniBand + shared memory), Linux* 64
Relative (Geomean) MPI Latency Benchmarks (Higher is Better)



Configuration: Hardware: Intel® Xeon® CPU E5-2680 @ 2.70GHz, RAM 64GB; Interconnect: InfiniBand, ConnectX adapters; FDR. MIC: C0-KNC 1238095 kHz; 61 cores. RAM: 15872 MB per card. Software: RHEL 6.2, OFED 1.5.4.1, MPSS Version: 3.2, Intel® C/C++ Compiler XE 13.1.1, Intel® MPI Benchmarks 3.2.4;

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. * Other brands and names are the property of their respective owners. Benchmark Source: Intel Corporation

Optimization Notice: Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice. Notice revision #20110804 .

Бинарная совместимость с MPICH

Для MPI реализаций Intel® MPI Library 5.0 и MPICH3.1 объявлена бинарная совместимость:

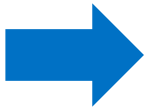
- Возможность заменить используемую библиотеку без перекомпиляции с сохранением работоспособности
- Также для IBM* MPI v1.4, Cray* MPT v7.0
-

```
$ cat ./version_info.c
#include "mpi.h"
#include <stdio.h>

int main (int argc, char *argv[]) {
    int rank, namelen;
    char mpi_name[MPI_MAX_LIBRARY_VERSION_STRING];

    MPI_Init (&argc, &argv);
    MPI_Comm_rank (MPI_COMM_WORLD, &rank);

    MPI_Get_library_version (mpi_name, &namelen);

    if (rank == 0)
        printf ("Hello world: MPI implementation:\n
%s\n", mpi_name);

    MPI_Finalize ();
    return (0);
}
```

Бинарная совместимость с MPICH

MPICH 3.1

```
$ mpiexec -n 2 ./version_info
Hello world: MPI implementation:
MPICH Version: 3.1.2
MPICH Release date: Mon Jul 21 16:00:21 CDT 2014
MPICH Device: ch3:nemesis
MPICH configure: --prefix=/home/user/mpich/3.1.2/
MPICH CC: gcc    -O2
MPICH CXX: g++   -O2
MPICH F77: gfortran   -O2
MPICH FC: gfortran   -O2
```



Intel® MPI Library 5.0

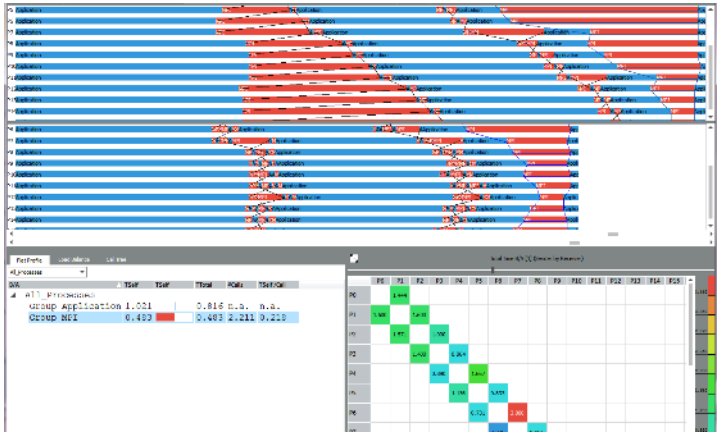
```
$ export
LD_LIBRARY_PATH=/opt/intel/impi/5.0.1.035/intel64/l
ib/:$LD_LIBRARY_PATH
$ mpiexec -n 2 ./version_info
Hello world: MPI implementation:
Intel(R) MPI Library 5.0 Update 1 for Linux* OS
```

Intel® MPI Library 5.0

```
$ source
/opt/intel/impi/5.0.1.035/intel64/bin/mpivars.sh
$ mpiexec.hydra -V
Intel(R) MPI Library for Linux* OS, Version 5.0
Update 1 Build 20140709
Copyright (C) 2003-2014, Intel Corporation. All
rights reserved.
$ mpiexec.hydra -n 2 ./version_info
Hello world: MPI implementation:
Intel(R) MPI Library 5.0 Update 1 for Linux* OS
```

Инструменты настройки производительности

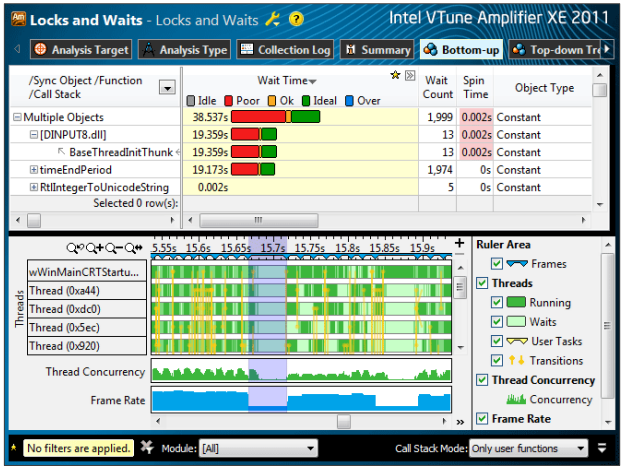
Intel® Trace Analyzer and Collector



Настройка межузловых взаимодействий MPI

- Визуализация вызовов MPI
- Оценка балансировки нагрузки MPI
- Поиск коммуникационных «узких мест»

Intel® VTune™ Amplifier XE



Настройка многопоточности внутри узла

- Визуализация потоков
- Оценка балансировки нагрузки
- Поиск «узких мест» синхронизации

Intel® Trace Analyzer & Collector

Что	• **Высокопроизводительный профилировщик производительность MPI приложений от Intel**
Почему	• **Производительность – Может работать с большим числом узлов** • **Масштабируемость – Готова для классических и многоядерных платформ** • **Эффективность – Настройка и анализ приложений**
Как	• Визуализация – Понять поведение параллельного приложения • Оценка – Параметры выполнения, балансировка нагрузки • Анализ – Автоматизированный поиск распространенных проблем MPI • Поиск – «Узких мест» коммуникаций

Полный инструментарий разработки приложений для систем с общей, распределенной памятью, гибридных

Компиляторы

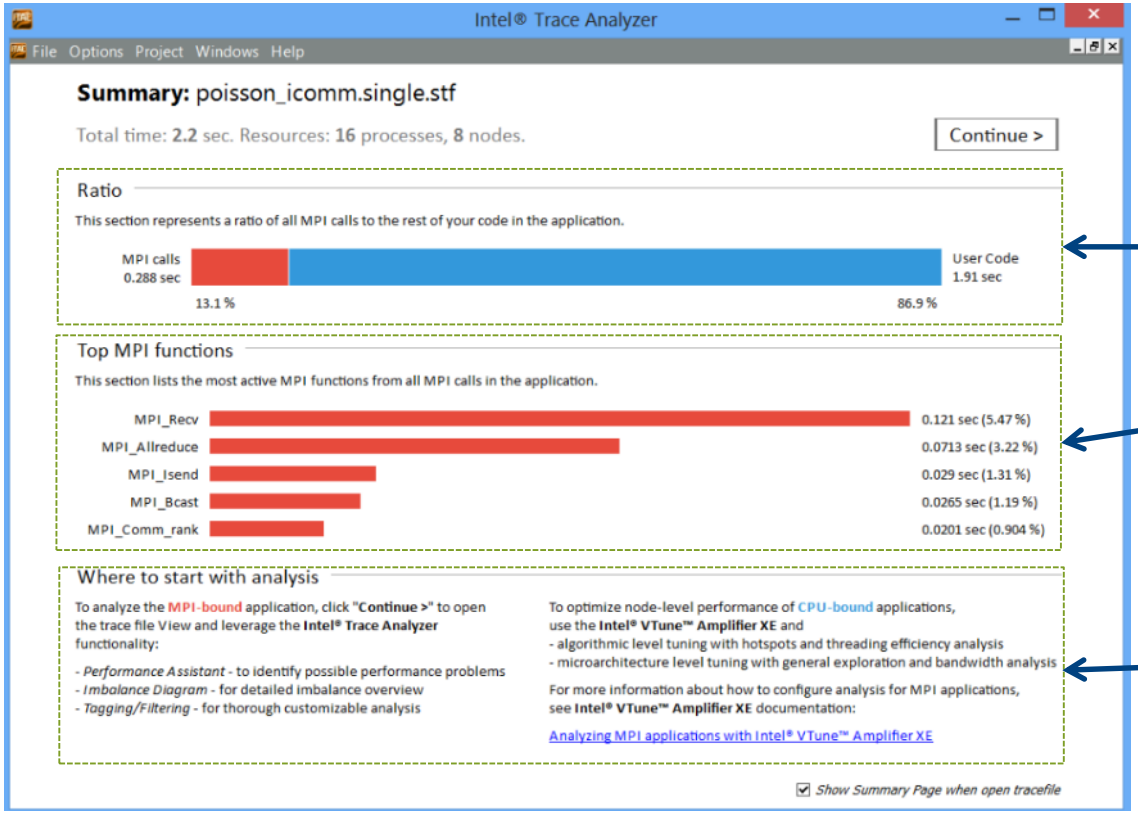
Инструменты анализа

Библиотеки

Высокая производительность MPI-приложений
с
Intel® Trace Analyzer and Collector

компонент
Intel® Parallel Studio XE Cluster Edition

Страница общего обзора MPI-коммуникаций



Соотношение
Вычисления
vs Коммуникации

Поиск наиболее
используемых MPI-функций

Совет как начать дальнейший
анализ

Что нового

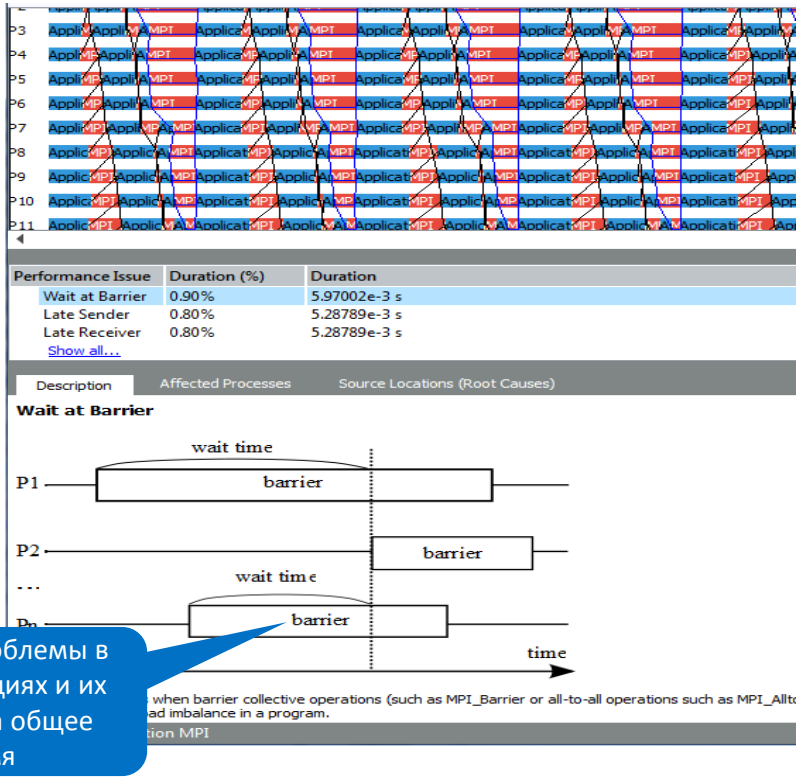
Intel® Trace Analyzer and Collector

Общий обзор MPI-коммуникаций

Поддержка MPI 3.0

Ассистент проблем производительности

- Обнаружение общих проблем с производительностью MPI
- Советы



Находит проблемы в коммуникациях и их влияние на общее время

MPI Performance Snapshot[1]

Легковесный масштабируемый инструмент для анализа производительности MPI-приложений.
Позволяет анализировать

- Соотношение времени вычислений и обменов
- Дисбаланс вычислительной нагрузки
- Время, затраченное в MPI-функциях
- Распределение размеров передаваемых сообщений

Очень прост в использовании.

```
$ source <IMPI_installdir>/intel64/bin/mpivars.sh
$ source <ITAC_installdir>/bin/mpi_perf_snapshot_vars.sh
$ mpirun -mps –n 2 ./myApp
…
======== GENERAL STATISTICS ========
WallClock: 21.077 sec (All processes)
MIN: 10.538 sec (rank 1)
MAX: 10.539 sec (rank 0)
```

[1] Доступен как превью в Intel® Parallel Studio XE 2015 Cluster Edition Update 1 для Linux OS

MPI Performance Snapshot

Минимальное влияние на приложение

```
$ source
<ITAC_installdir>/bin/mpi_perf_snapshot_var
s.sh
```

```
$ export
I_MPI_JOB_RESPECT_PROCESS_PLACEMENT=disable
I_MPI_PIN_DOMAIN=socket OMP_NUM_THREADS=14
MKL_NUM_THREADS=14
```

```
$ time ( mpirun -n 4 -ppn 2
./xhpl_hybrid_intel64_dynamic )
…
real    4m51.676s
user    98m52.230s
sys     0m36.415s
```

Минимальное влияние

```
$ time ( mpirun -mps -n 4 -ppn 2
./xhpl_hybrid_intel64_dynamic )
…
real    4m53.572s
user    99m14.129s
sys     0m35.119s
```

MPI Performance Snapshot

Использование

```
$ mpirun -mps -n 4 -ppn 2 ./xhpl_hybrid_intel64_dynamic
…
==================== GENERAL STATISTICS ====================
WallClock:          1169.466 sec (All ranks)
      MIN:           292.326 sec (rank 2)
      MAX:           292.402 sec (rank 0)

   MPI:    5.59%    NON_MPI:   94.41%

================= MEMORY USAGE STATISTICS =================
All ranks:  33116.691 MB
      MIN:   8231.441 MB (rank 0)
      MAX:   8330.074 MB (rank 1)
```

MPI Performance Snapshot

Использование – начальный анализ

```
$ mpirun -mps -n 4 -ppn 2 ./xhpl_hybrid_intel64_dynamic
…
================= MPI IMBALANCE STATISTICS =================
MPI Imbalance:            44.967 sec            3.845% (All ranks)
          MIN:            10.716 sec            3.665% (rank 1)
          MAX:            11.766 sec            4.024% (rank 0)

==================== OpenMP STATISTICS ====================
OpenMP Regions:           904.757 sec           77.365%               30 region(s) (All ranks)
           MIN:           224.699 sec           76.846%                8 region(s) (rank 0)
           MAX:           227.247 sec           77.734%                7 region(s) (rank 3)

OpenMP Imbalance:          14.309 sec            1.224% (All ranks)
             MIN:           2.614 sec            0.894% (rank 1)
             MAX:           4.290 sec            1.467% (rank 2)
```

MPI Performance Snapshot

Основные диаграммы

Последующий анализ включает:

- Использованная память и значения аппаратных счетчиков
- Информация об использованных функциях MPI
- Время в MPI для каждого ранка
- Время коллективных операций
- Распределение размеров сообщений
- Общий объем переданных данных между ранками

MPI Performance Snapshot

Основные диаграммы - пример

```
$ mpi_perf_snapshot -f ./stats.txt ./app_stat.txt
| Function summary for all ranks
|-------------------------------------------------------------------------------------------------------------
|           Function        Time(sec)          Time(%)       Volume(MB)        Volume(%)             Calls
|-------------------------------------------------------------------------------------------------------------
                Send          25.1177            44.018          39443.6               50            130069
               Probe          13.8178           24.2154                0                0          12182371
                Recv           10.337           18.1153          39443.6               50            130069
                Wait          7.12479            12.486                0                0            128997
                Init          0.66177           1.15973                0                0                 4
              Gather       0.00261283        0.00457891       0.000366211      4.64221e-07                 4
| [skipped 2 lines]
|=============================================================================================================
| TOTAL                      57.0623               100          78887.2              100          12571530
```

Онлайн-ресурсы

Intel® MPI Library

Страница Intel® MPI Library и бесплатная 30-дневная оценочная версия!

- www.intel.com/go/mpi

Страница Intel® Trace Analyzer and Collector

- www.intel.com/go/traceanalyzer

Форумы Intel® Clusters and HPC Technology

- http://software.intel.com/en-us/forums/intel-clusters-and-hpc-technology

Сообщество разработчиков для Intel® Xeon Phi™

- http://software.intel.com/en-us/mic-developer

Обмен данными с использованием MPI. Работа с библиотекой MPI на примере Intel® MPI Library

- http://habrahabr.ru/company/intel/blog/251357/

Legal Disclaimer & Optimization Notice

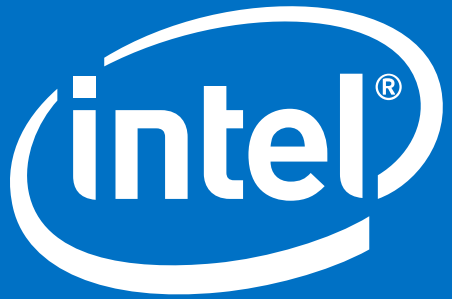
INFORMATION IN THIS DOCUMENT IS PROVIDED "AS IS". NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO THIS INFORMATION INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.

Optimization Notice

Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice.

Notice revision #20110804



Software

dmitry.sivkov@intel.com