

Исследование производительности подсистемы ввода/вывода суперкомпьютера BlueGene/P

С. Коробков¹

¹Факультет вычислительной математики и кибернетики
МГУ им. М.В. Ломоносова

Всероссийская научная конференция
«Научный сервис в сети Интернет: все грани параллелизма»
Абрау-Дюрсо, Морьяк, 2013

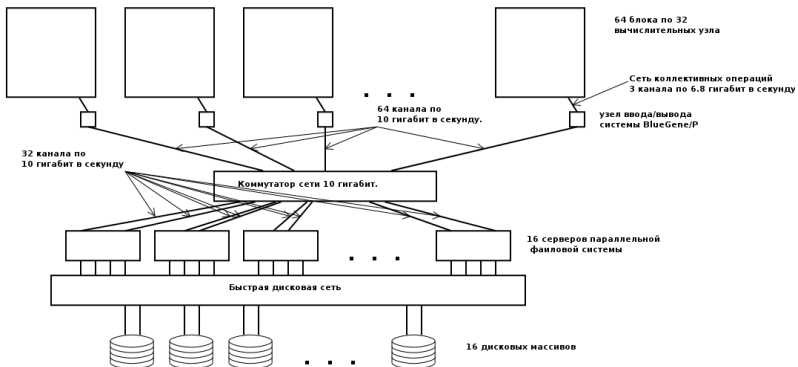
Содержание

- 1 Введение
- 2 Подсистема ввода/вывода BlueGene/P
- 3 Тест на узле ввода/вывода
- 4 Тесты с вычислительных узлов
 - POSIX IO
 - MPI IO
 - Socket IO
- 5 Заключение

Введение

Часто возникает проблема что заявленная скорость ввода/вывода не соответствует реальной скорости. Это приводит к проблемам при вводе/выводе большого объема данных. Определим узкие места вычислительной системы.

Схема BlueGene/P в МГУ



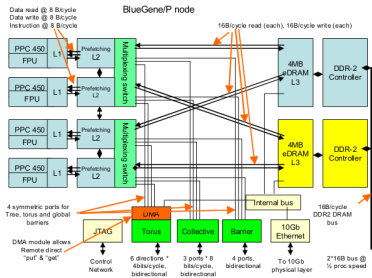
Узкие места подсистемы ввода/вывода BlueGene/P

- Узел ввода/вывода BlueGene/P
- Связь файловых серверов с коммутатором
- Скорость записи на жёсткий диск

Схема вычислительного и узла ввода/вывода

Основные процессы узла ввода/вывода:

- `ciod` – управляющий процесс ввода/вывода
- `mmfsd` – клиент файловой системы
- `iorproхu` – процесс выполнения операций ввода/вывода проброшенных с вычислительных узлов



Тест на узле ввода/вывода

- Записывалось 10 Гб случайных данных порциями по 100мб
- Записывалось время записи данных
- Замерялась загрузка узла ввода/вывода

Тест с записью на диск: 695 Мб/сек

Тест без записи на диск: 880 Мб/сек

Методы ввода/вывода данных

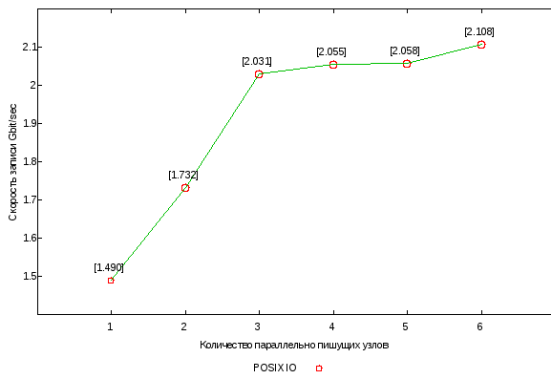
Для ввода/вывода данных мы можем использовать:

- POSIX IO функции
- MPI IO функции
- Функции для работы с сокетами.

Описание теста

- Запись данных выполняется с помощью обычных операций write с вычислительных узлов
- Каждый узел пишет 10 гигабайт случайных данных порциями по 100 мб
- Каждый узел пишет данные в свой файл

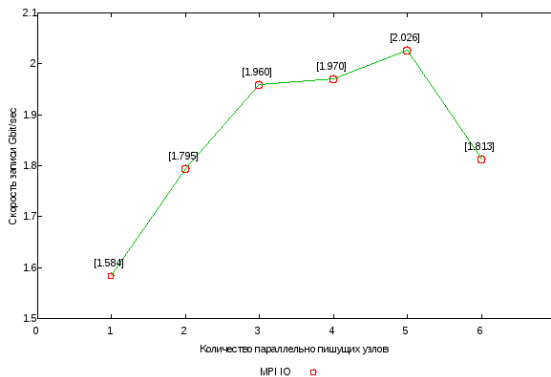
Скорость теста



Описание теста

- Запись данных выполняется с помощью функцией `MPI_File_write_at` с вычислительных узлов
- Каждый узел пишет 10 гигабайт случайных данных порциями по 100 мб
- Каждый узел пишет данные в свою область файла, открытого при помощи MPI IO

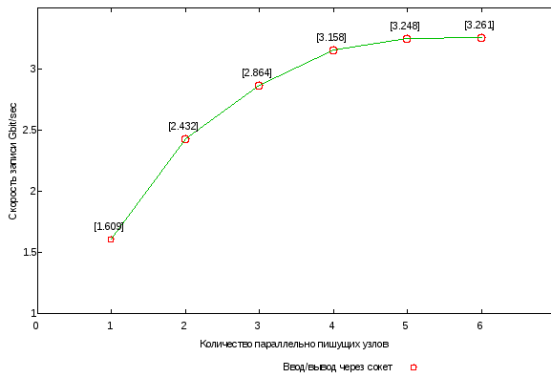
Скорость теста



Описание теста

- Вывод данных осуществляется путём записи в сокет
- Каждый узел выводит 10 гигабайт случайных данных порциями по 100 мб
- Данные считываются из сокета и пишутся на диски на серверах файловой системы

Скорость теста



Заключение

- Клиент файловой системы создаёт значительную нагрузку на узел ввода/вывода
- Эту нагрузку можно переложить на узлы файловой системы используя вывод через сокеты
- Используя для вывода 4 - 5 вычислительных узлов можно достичь максимальной производительности ввода/вывода 3.2 Gbit/s

Спасибо за внимание

Вопросы?