Параллельные вычислительные технологии 2015 (ПаВТ-2015)
Екатеринбург, 31 марта – 2 апреля 2015 г.

# Применение сжатия информации при использовании многоядерных ускорителей для обработки баз данных

Беседин К.Ю, Костенецкий П.С.
*Южно-Уральский государственный Университет*

Работа выполнена при финансовой поддержке Минобрнауки РФ в рамках ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2014–2020 годы» (Соглашение № 14.574.21.0035).

# Актуальность

- Многоядерные сопроцессоры Intel Xeon Phi могут быть эффективно использованы в СУБД:
  - Сортировка;
  - поиск;
  - соединение;
  - агрегатные функции.
- шина PCI-E является узким местом при разработке ПО для Intel Xeon Phi;
- для сокращения времени передачи данных можно использовать сжатие:
  - Давно применяется в СУБД для снижения объема данных, хранимых на диске и участвующих в операциях ввода-вывода;
  - более эффективно для СУБД с поколоночным хранением данных;
  - возможна обработка данных в сжатом виде;
- исследований, посвященных сжатию данных на Intel Xeon Phi, пока не проводилось.

# Исследуемые методы сжатия

- RLE (Run Lenght Encoding)
- LZSS (Lempel-Ziv-Storer-Szymanski)
- Null Suppression
- RLE + NS

# Реализация

- Язык программирования — C++;
- технология параллельного программирования — OpenMP;

# Вычислительные эксперименты

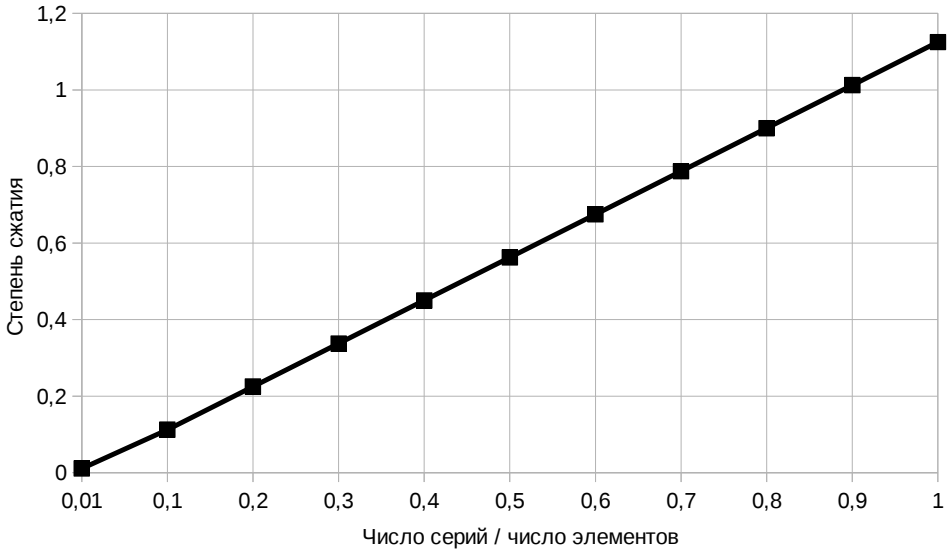
- Вычислительные эксперименты выполнялись на суперкомпьютере «Торнадо-ЮУрГУ»;
- проводились для 1500 МБ тестовых данных;
- тип сжимаемых данных — `uint64_t` (8 байт);
- для каждого метода сжатия варьировалась характеристика сжимаемых данных, влияющая на степень сжатия;
- под временем обработки понимается сумма времени передачи данных, их распаковки (если требуется) и время выполнения над ними агрегатной функции — вычисления суммы;
- во всех экспериментах предполагается, что данные изначально находятся в сжатом виде.

# Использованное оборудование

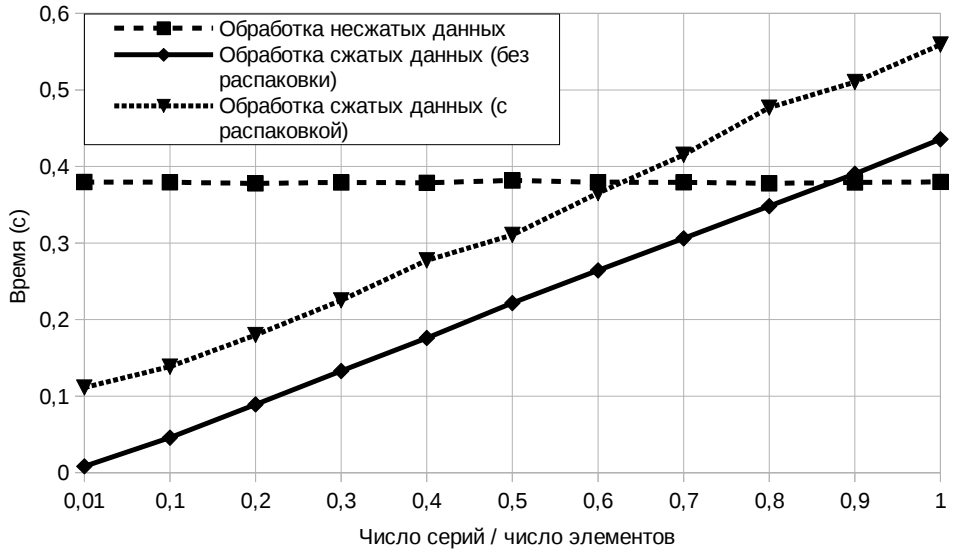
Суперкомпьютер «Торнадо ЮУрГУ»

| | | |
|---|---|---|
| Узел | Процессор | Intel Xeon 5680 3.33 GHz |
| | Объем ОЗУ | 24 / 48 GB |
| | Число процессоров | 2 |
| | Сопроцессор | Intel Xeon Phi 7110X |
| Число узлов | | 480 |

# Run Length Encoding



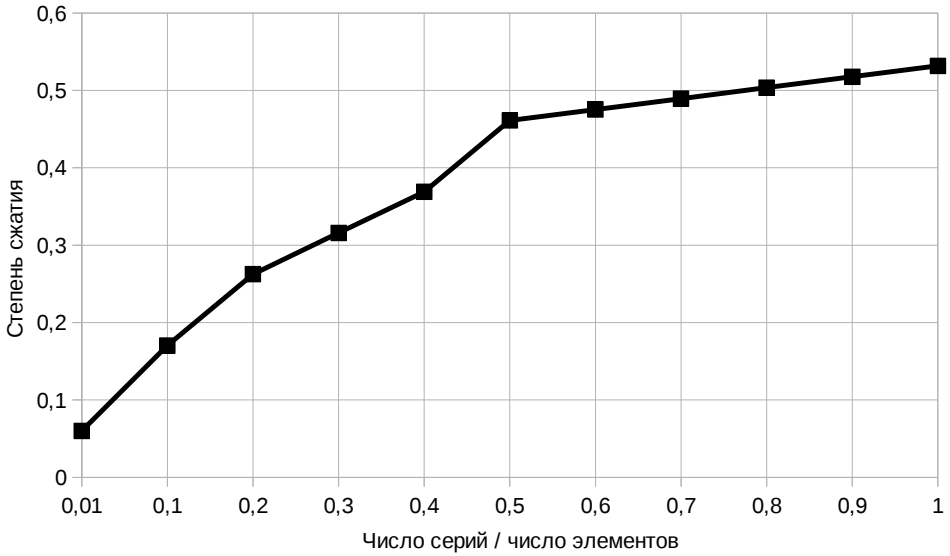

Степень сжатия



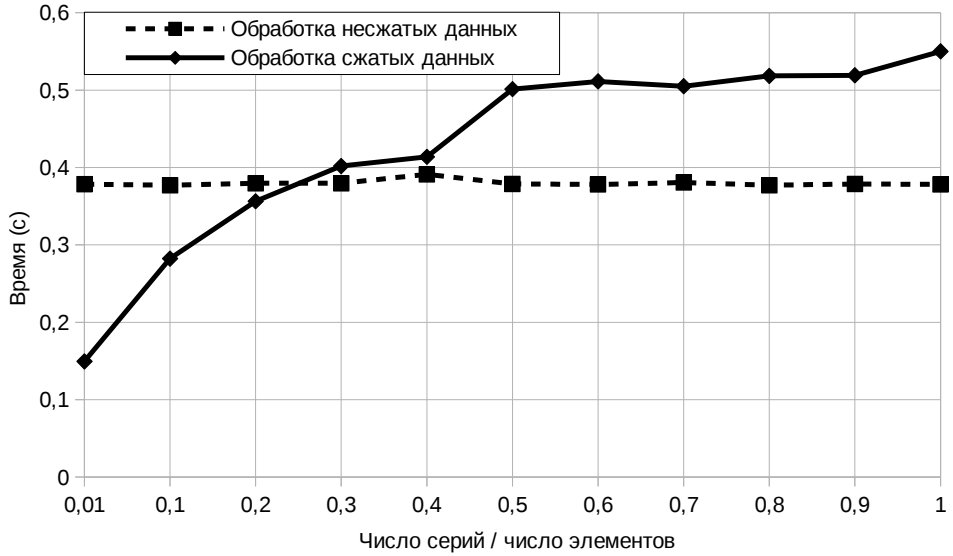

Время обработки данных

- Варьируемая характеристика — отношение числа серий на общее число элементов сжимаемых данных;
- метод RLE может использоваться для повышения эффективности обработки баз данных на Intel Xeon Phi если число серий в сжимаемых данных достаточно мало;
- обработка данных в сжатом виде позволяет дополнительно увеличить эффективность использования метода RLE.

# LZSS



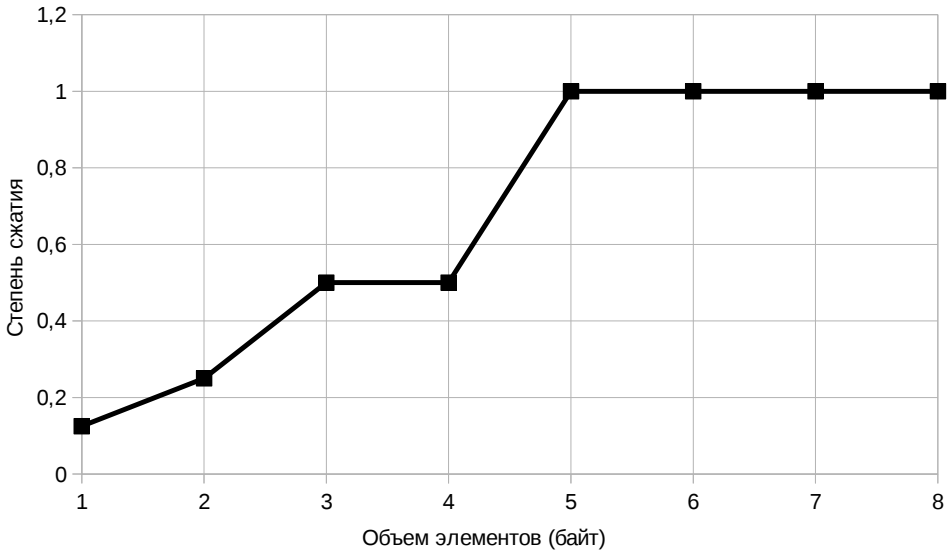

Степень сжатия



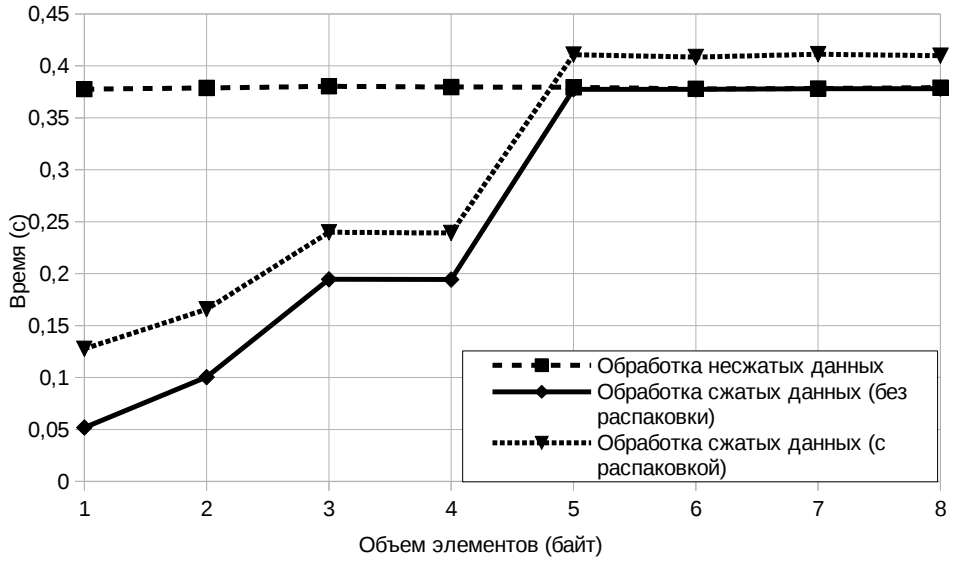

Время обработки данных

- Варьируемая характеристика — отношение числа серий на общее число элементов сжимаемых данных;
- метод LZSS может быть эффективно использован при передаче данных из основной памяти в память сопроцессора Intel Xeon Phi если число серий в сжимаемых данных достаточно мало.

# Null Suppression



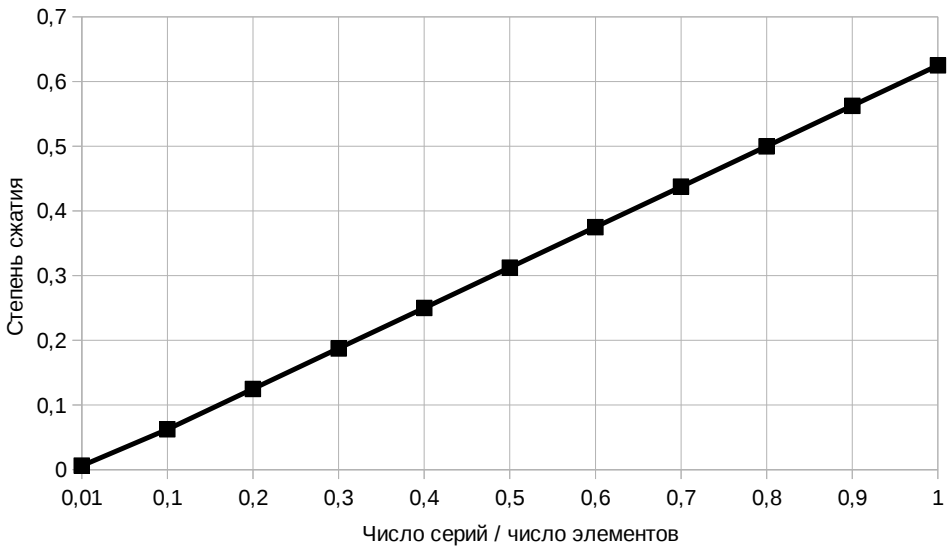

Степень сжатия



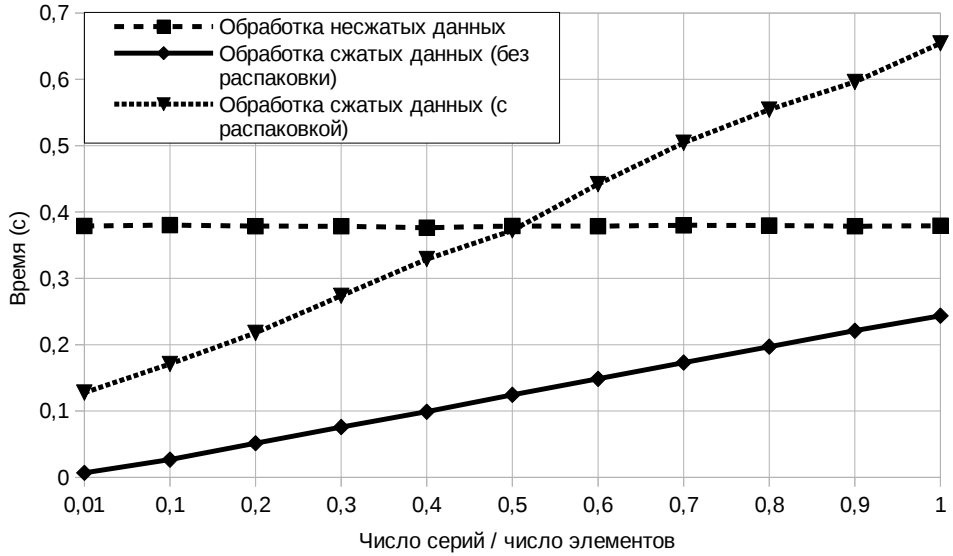

Время обработки данных

- Варьируемая характеристика — размер максимального элемента данных;
- метод Null Suppression может использоваться для повышения эффективности обработки баз данных на Intel Xeon Phi если значения сжимаемых данных лежат в ограниченном диапазоне;
- обработка данных в сжатом виде значительно повышает эффективность метода Null Suppression.

# RLE + Null Suppression




Степень сжатия




Время обработки данных

- Варьируемая характеристика — отношение числа серий на общее число элементов сжимаемых данных;
- комбинация методов RLE и Null Suppression может использоваться для повышения эффективности обработки баз данных на Intel Xeon Phi;
- комбинация методов RLE и Null Suppression может оказаться эффективнее данных методов по отдельности;
- обработка данных в сжатом виде значительно повышает эффективность метода RLE + NS.

# Выводы

1. Выбрано несколько методов сжатия, используемых в СУБД;
2. выполнена параллельная реализация выбранных методов сжатия для сопроцессоров Intel Xeon Phi и центральных процессоров;
3. исследована эффективность применения выбранных методов сжатия при передаче данных из основной памяти в память сопроцессора Intel Xeon Phi.