Технология совместного использования свободной СУБД PostgreSQL и сопроцессоров Intel Xeon Phi для решения задач интеллектуального анализа данных*

Т.В. Речкалов

ФГБОУ ВПО "Южно-Уральский государственный университет" (НИУ)

В настоящее время актуальной проблемой является разработка эффективных методов и алгоритмов интеллектуального анализа сверхбольших объемов данных. Одним из подходов к решению данной проблемы является интеграция интеллектуального анализа данных и параллельной СУБД [1]. Интеграция позволяет избежать накладных расходов на экспорт информации из базы данных в систему интеллектуального анализа данных. Актуальной темой исследований в области технологий баз данных является также эффективное использование в обработке запросов многоядерных ускорителей [2].

В докладе представлен подход к реализации операций агрегации данных по столбцам на основе использования хранимых процедур, частично исполняемых на сопроцессоре Intel Xeon Phi. Процедуры объединены в библиотеку с названием pgDM (PostgreSQL Data Mining). Библиотека pgDM предназначена для проведения интеллектуального анализа данных внутри реляционной СУБД PostgreSQL. Пользователем данной библиотеки является прикладной программист, решающий задачи проведения интеллектуального анализа данных, хранящихся в реляционной СУБД PostgreSQL.

Модульная структура библиотеки pgDM представлена на рис.1. Библиотека pgDM состоит из двух подсистем. Внешний интерфейс *pgdm-fe* представляет собой интерфейс прикладного программиста. Внутренний интерфейс *pgdm-be* представляет собой реализацию функций внешнего интерфейса посредством хранимых процедур PostgreSQL, частично исполняемых на сопроцессоре Intel Xeon Phi. Внутренний интерфейс, в свою очередь, состоит из двух модулей: *pgdm-be-wrapper* и *pgdm-be-Phi*.

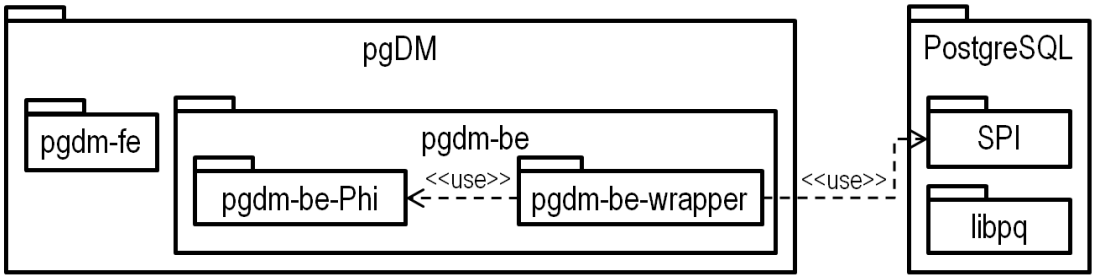

Рис. 1. Модульная структура библиотеки pgDM

Модуль *pgdm-be-wrapper* представляет собой обертку функций, реализованных на языках PL/pgSQL и C, которые выполняют интеллектуальный анализ данных. Модуль *pgdm-be-Phi* предоставляет функции, каждая из которых реализует на сопроцессоре Intel Xeon Phi одну вычислительную операцию над набором записей реляционной таблицы. В реализации модуля *pgdm-be-wrapper* используется стандартный модуль PostgreSQL SPI (Server Programming Interface), который обеспечивает работу с таблицами при помощи курсоров.

Литература

1. Ordonez C. Can we analyze big data inside a DBMS? // Proceedings of the Sixteenth International Workshop on Data Warehousing and OLAP. ACM, 2013. P. 85-92.
2. Agrawal R. et al. The Claremont report on database research //Communications of the ACM. 2009. T. 52. №. 6. P. 56-65.

* Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 12-07-00443-а.