

М. Серік¹, С.К. Жумагулова^{1*}, А.В. Копыльцов², Г.Б. Абилдаева³

¹Евразийский национальный университет имени Л.Н. Гумилева, Астана, Казахстан;

²Государственный университет аэрокосмического приборостроения, Санкт-Петербург, Россия;

³Карагандинский технический университет имени Абылкаса Сагинова, Караганда, Казахстан

(*Корреспондирующий автор. E-mail: saulesha_81@mail.ru)

ORCID 0000-0002-2801-432X¹

ORCID 0000-0003-0437-7545²

ORCID 0000-0001-6587-0507³

ORCID 0000-0003-3262-2552⁴

Использование технологий реализации параллельных вычислений в базе данных в современном образовании

В статье изложены технологии, применяемые при обучении параллельным вычислениям в базе данных в образовательном процессе вуза. Сегодня парадигму профессионального образования сменила компетентностная, цель которой состоит в ориентации образовательного процесса вуза на формирование и развитие комплекса ключевых компетентностей обучающихся. Как показали результаты анализа педагогических подходов, применяемых сегодня к определению характеристик компетентности будущих разработчиков программного обеспечения в области баз данных, а также изучение основ их формирования, можно наблюдать все более возросший интерес общества к качественной подготовке специалистов в области баз данных. При этом можно наблюдать недостаточность разработанности модели формирования специалиста, обладающего такой высокой компетентностью. Авторами изложены основные принципы работы моделей параллельных вычислений, а также вопросы использования алгоритмов распараллеливания заданий для оптимизации работы с базами данных. Обоснован предлагаемый методический подход в реализации параллельных вычислений в базе данных с помощью изложенных моделей. Авторы исследования предлагают подход к формированию навыков программирования параллельных вычислений в базе данных на основе внедрения практики решения задач из реального опыта софтверных компаний. Как показали результаты проведенного эксперимента, различия в средних экспертных оценках между контрольной и экспериментальной группами являются статистически значимыми. Полученные результаты исследования могут быть имплементированы в развитие специалистов по параллельным вычислениям в базах данных и способствовать развитию связей с компаниями, предоставляющими рабочие места выпускникам соответствующих направлений подготовки.

Ключевые слова: базы данных, параллельные вычисления, обучающиеся, учебный процесс, профессиональные компетенции, методология, система, IT-специалист.

Введение

Тенденции современного времени, социально-экономические факторы развития общества и особенности современных потребителей образовательных услуг требуют от системы профессионального образования значительных изменений в принципах проектирования и реализации имеющихся образовательных программ, а также новых форм и моделей обучения.

Следствием этого стал тот факт, что возникла потребность замены парадигмы профессионального образования на компетентностную, где образовательный процесс вуза направлен на формирование и развитие определенных ключевых компетентностей обучающихся. При этом целью профессионального образования служит не подготовка узких специалистов в конкретной отрасли, а развитие личности специалистов IT-профиля в сфере разработки программного обеспечения, включая базы данных (БД), а также повышение уровня его профессиональной компетентности.

Авторами научных трудов в области баз данных (С.А. Лупин, М.А. Посыпкин, А.В. Поначугин, С.А. Немнюгин и другими) и параллельных вычислений в базах данных были исследованы вопросы формирования профессиональной компетентности обучающихся IT-специальностей в ходе учебного процесса, результаты анализа которых позволяет утверждать о растущем интересе к данной проблеме. Одновременно с этим, в соответствии с данными анализа педагогической литературы, можно сделать вывод о том, что, несмотря на то, что сегодня имеются факты об освещенности в научных трудах проблемой формирования профессиональной компетентности IT-специалистов в параллельных

вычислениях в базах данных, отсутствуют научно обоснованные методики формирования данной компетентности у обучающихся.

В настоящее время имеется ряд педагогических подходов, определяющих характеристики компетентности специалистов по базам данных. Их анализ, а также изучение теоретико-практических основ формирования данной компетентности позволили обнаружить возросшую потребность общества в качественной подготовке специалистов в области баз данных. При этом можно наблюдать недостаточность степени разработки модели формирования специалиста с такой высокой компетентностью.

Еще при зарождении компьютерной эры создатели машин задумались над идеей применять параллельные вычисления в компьютерах. Причиной этому явился тот факт, что наращивание производительности только лишь за счет совершенствования аппаратного обеспечения машин требует немало материальных затрат, а также существует вероятность возникновения налагаемых физическими законами ограничений. Метод распараллеливания вычислений возник довольно продолжительное время назад. Для реализации организации совместной работы большого количества независимых процессоров необходимо выполнение глубоких теоретико-практических исследований, без которых сложная и дорогостоящая многопроцессорная установка не превосходит, а в большинстве случаев даже уступает по производительности обычному компьютеру [1].

В настоящее время параллельные системы БД применяются в системах самого различного назначения, начиная с сервера приложений и заканчивая сервером систем поддержки принятия решений. В данных разработках содержится обработка баз данных и реализация запросов с помощью параллельных систем. Своим успехом параллельные системы баз данных, главным образом в приложениях поддержки принятия решений (хранилища данных), обязаны оптимизации параллельных запросов. Учитывая SQL-запрос, целью оптимизации параллельных запросов служит поиск параллельного плана, который обеспечивает результат запроса за весьма короткий интервал времени. Различные оптимизирующие решения должны быть реализованы для параллельных БД.

На сегодняшний день актуальной является проблема недостатка методических разработок для обучения параллельным вычислениям в базах данных в вузах, таким образом, имеются проблемы со всесторонним освоением соответствующих технологий. Тот факт, что на практике чаще всего дается теоретическое обучение данного направления, влияет на скорее отрицательную динамику, чем положительную. Лишь в последнее время можно наблюдать появление специальных технологий, позволяющих ввести практику параллельных вычислений при обучении предмету баз данных.

При написании параллельных программ повышенного уровня сложности структуры алгоритмов должны быть подвергнуты намного более глубокому анализу, чем при обычном последовательном программировании, при этом нет возможности реализовать определенные подходы в случае отсутствия серьезных изменений мышления IT-специалистов. Для получения практически значимых результатов, наряду с теоретическим анализом, необходимо регулярно практиковать навыки разработки, поскольку имеется потребность в постоянной практике в написании параллельных программ [2].

Таким образом, принимая во внимание существующий минимальный объем академических исследований, посвященных формированию сложных взаимосвязанных технических навыков, к примеру, параллельных вычислений и управления базами данных, задачей данного исследования послужило предложение методики обучения будущих IT-специалистов в рамках дисциплин образовательных программ вуза.

Методы и материалы

Для написания исследовательской работы была изучена научная литература, а также труды отечественных и зарубежных ученых, опубликованные в научных журналах на тему исследования. В процессе исследования были применены такие методы как теоретические (анализ, сравнение, индукция, дедукция и пр.) и эмпирические (наблюдение, беседа, цифровое анкетирование).

Результаты и их обсуждение

Предназначением модели параллельных вычислений служит обеспечение высокоуровневого подхода к определению характеристик и сравнения времени реализации различных программ, при этом не берутся в расчет аппаратное обеспечение и детали реализации.

Первой моделью параллельных вычислений послужил компьютер с параллельным случайным доступом (PRAM — Parallel Random Access Machine), обеспечивший абстракцию машины с разделяемой памятью. Каждый процессор этой модели всегда имеет доступ к любой ячейке памяти. Базовая модель

PRAM осуществляет поддержку параллельных считываний и записи, при этом подмодели PRAM придерживаются правил, исключающих конфликтных ситуаций в случаях одновременного обращения ряда процессоров к общей памяти [3].

В модели BSP (Bulk Synchronous Parallel) объединены абстракции распределенной и разделенной памяти. В данной модели команды реализуются процессорами в синхронном порядке. К ключевым командам относят запись и считывание из памяти, а также логические и арифметические операции.

Более высокая производительность может быть достигнута в результате применения алгоритмов распараллеливания заданий. В СУБД принято различать следующие области применения подобных алгоритмов:

- параллельный ввод/вывод;
- параллельные средства и административные утилиты;
- параллельная обработка запросов к базе данных.

Распараллеливание ввода/вывода целесообразно использовать для реализации одновременного доступа к фрагментированным таблицам и индексам, местом расположения которых служат несколько физических дисков. Благодаря данному алгоритму имеется возможность во много раз увеличить скорость операций с внешними устройствами, имеющими относительно медленную производительность.

Утилиты администрирования (построение индексов, сортировка, загрузка и выгрузка данных, восстановление и архивирование баз данных и пр.) также должны быть реализованы полностью параллельно.

Наиболее сложным из перечисленных алгоритмов является реализация параллелизма в обработке запросов к базе данных. Теоретически возможность распараллеливания запросов в реляционной СУБД можно обосновать следующим свойством, которое носит название реляционного замыкания. Суть его состоит в том, что результатом каждого реляционного оператора: SELECT — выборка подмножества строк таблицы; PROJECT — выборка подмножества столбцов; JOIN — объединение двух таблиц, служит новое отношение. Поскольку запросы можно разбить на иерархию более простых операторов, целесообразна их параллельная реализация. Последовательность атомарных операций при выполнении запроса обработке запроса определяет оптимизатор из нескольких имеющихся вариантов. Для того, чтобы обеспечить параллелизм в сервере баз данных, производителю необходимо определить совокупность атомарных операций, способных решать свои части задачи в одно и то же время. При этом необходимо обеспечить возможность объединения результатов их работ так, как если эти задачи имели единого исполнителя [4].

При обработке сложных запросов, которым присуще низкое быстродействие, параллельная обработка может стать целесообразным решением. Также она может быть использована при крупномасштабных размерах обрабатываемых данных, к примеру, массивов обновляемых таблиц больших размеров. Таким образом, благодаря параллельной обработке небольших порций данных, параллелизм способен существенно повысить производительность систем пакетной обработки.

В этой цели крайне важно, чтобы на реализацию параллельной обработки запроса не влияли средства межзадачного взаимодействия конкретной вычислительной системы. Процесс создания множества тиражируемых копий одной атомарной операции, реализующих параллельно одну задачу, называют горизонтальным параллелизмом. Дополнительная степень параллелизма — вертикальный параллелизм — может быть достигнут в случае параллельного исполнения различных атомарных операций, обычно реализуемых по очереди. Благодаря применению горизонтального и вертикального параллелизма можно обеспечить рост общей производительности автоматизированной системы. При этом рост объемов обрабатываемых данных происходит за счет вертикального и/или горизонтального масштабирования аппаратной платформы.

Как правило, решение задач эффективной подготовки будущих IT-специалистов к реализации задач, с которыми им придется столкнуться на рабочем месте, рассматривается в академических исследованиях в более широком аспекте повышения эффективности подготовки, а также совершенствования отдельных компетенций. Обучение работе с базами данных либо параллельным вычислениям в БД рассматриваются как часть программы подготовки специалиста, при этом происходит оценка методов обучения специализированным навыкам в рамках специальных дисциплин.

Авторами академических исследований в рамках изучения технологий параллельных вычислений в базах данных был отмечен разрыв между потребностями данных постоянно развивающихся областей и навыками, характерными для существенной части уже работающих в определенной области IT-специалистов. Зачастую корень проблемы состоит в том, что имеется потребность в освоении новых

навыков, отличных от тех, которые имеются у специалиста, либо в разработке специфических технологий, предоставляющих возможность абстрагировать неисследованную для IT-специалиста информацию [5].

Применение для усиления обучения решения реальных задач, стоящих перед специалистами, работающими с базами данных, наблюдалось и ранее, но, в отличие от предлагаемой методики, решение реальных задач предлагалось в качестве уже решенных кейсов для изучения и запоминания. Предлагаемая авторами методика, подразумевающая предъявление к решению сложных задач в качестве вызова, направляющего формирование индивидуального трека подготовки обучающегося, может быть во много раз эффективнее при подготовке специалиста.

Сегодня мир предъявляет множество вызовов, которые обращены к параллельным вычислениям. Эти вызовы содержат геоинформационные системы (GIS), квантовые вычисления, распределенные системы научных расчетов с привлечением множества компьютеров, управление сложными процессами и пр. Значительная доля исследований изучает методы программирования, а не педагогические подходы к их освоению, поскольку чаще всего принято, что специалист осваивает необходимые ему современные технологии на рабочем месте.

При этом специализированные методики, связанные с освоением параллельных вычислений как одной из ключевых технологий современного программирования, рассматриваются весьма редко. Зачастую это можно объяснить тем, что реальные задачи параллельных вычислений практически всегда связаны с обработкой либо интенсивного потока данных (например, запросов к базе данных), либо с обработкой больших массивов данных, либо с оптимизацией вычислений, для которых в приоритете скорость получения результатов.

Применение в обучении будущих IT-специалистов инновационных подходов и платформ, таких как специфические высокоуровневые языки запросов, средства декомпозиции параллельных программ и реорганизации процессов их реализации и иных методов, сегодня не исследовано с точки зрения кратковременной и долговременной эффективности, что можно объяснить методологической сложностью реализации таких исследований, которую осознают и сами исследователи. Данный факт свидетельствует о том, что в данном исследовании значительную роль играет обращение к решению задач, максимально приближенных к реальным задачам, как форме проверки эффективности обучения и соответствия полученных знаний поставленным задачам обучения.

Формирование сложных навыков и знаний, связанных с компетенциями одновременно из нескольких различных технологий программирования, представляет собой весьма сложную задачу. Решение данной задачи остается одной из приоритетных педагогических проблем подготовки будущих IT-специалистов, но и сегодня практически не находит отражения в академической литературе. Представленное авторами исследование позволяет сделать акцент на выводы, сделанные рядом исследователей ранее: компетенции и навыки разработки ПО и иных IT-профессий, в том числе в области параллельных вычислений в базах данных, могут быть эффективнее изучены во время обучения в вузах, а не на рабочем месте, что требует смены парадигмы обучения [6].

Заключение

Разработка и управление базами данных тесно взаимосвязаны с технологиями параллельных вычислений, все более глубоко переплетающихся в реальной практике софтверных компаний. Но при этом академические исследования не уделяют в достаточной мере внимание специальным педагогическим подходам, направленным на подготовку кадров, способных сразу реализовать данные технологии на рабочем месте.

В настоящее время имеется относительно мало академических исследований, направленных на создание специализированных подходов к обучению специалистов в данной области в процессе обучения в вузе.

Представленное исследование предлагает подход к формированию навыков программирования параллельных вычислений в базах данных на основе внедрения практики решения задач из реального опыта софтверных компаний. Участниками эксперимента для апробации предлагаемой методики стали обучающиеся третьего курса образовательных программ IT-направления. Они в течение трех месяцев обучались, опираясь на построенную собственную индивидуальную обучающую траекторию для решения выданных им учебных задач. Для формирования задач был привлечен персонал софтверных компаний, специализирующийся на решении и реализации задач параллельного программирования в базах данных.

Для оценки качества реализации обучающимися выданных им заданий с позиции конечного пользователя программного продукта была сформирована группа экспертов. В целях сравнительной оценки эффективности подхода идентичные задачи были получены обучающимися третьего курса с аналогичным уровнем подготовки, не являющимися участниками эксперимента (контрольная группа). Определенная часть обучающихся экспериментальной группы показала способность к оперативному повышению уровня владения технологиями до приемлемой профессиональной компетенции; при этом участники контрольной группы не показали подобного результата.

Как показали результаты проведенного эксперимента, различия в средних экспертных оценках между контрольной и экспериментальной группами являются статистически значимыми. В реальности результаты проведенного исследования могут быть имплементированы в развитие специалистов по параллельным вычислениям в базах данных и способствовать развитию связей с компаниями, предоставляющими рабочие места выпускникам соответствующих направлений подготовки.

Список литературы

- 1 Богачев К.Ю. Основы параллельного программирования: учеб. пос. / К.Ю. Богачев. — М.: БИНОМ; Лаборатория знаний, 2003. — 127 с.
- 2 Лупин С.А. Технологии параллельного программирования: учеб. пос. / С.А. Лупин, М.А. Посыпкин. — М.: Форум; Инфра-М, 2008. — 208 с.
- 3 Миллер Р. Последовательные и параллельные алгоритмы: Общий подход / Р. Миллер, Л. Боксер. — М.: БИНОМ; Лаборатория знаний, 2006. — 406 с.
- 4 Поначугин А.В. Проблемы организации СУБД при параллельной архитектуре многопроцессорных вычислительных систем / А. В. Поначугин. // Молодой ученый. — 2016. — № 8 (112). — С. 159–163.
- 5 Немнюгин С.А. Параллельное программирование для многопроцессорных вычислительных систем / С.А. Немнюгин, О.Л. Стесик. — СПб.: БХВ-Петербург, 2002. — 400 с.
- 6 Серик М. Основные вопросы обучения реализации параллельных вычислений в базе данных в учебном процессе вузов / М. Серик, С.К. Жумагулова, Д.А. Казимова // Шамовские педагогические чтения научной школы управления образовательными системами: материалы XIII Междунар. науч.-практ. конф. (23 января–1 февраля 2021 года). — М.: МАНПО, 2021. — С. 787–789.

М. Серік, А.В. Копыльцов, С.К. Жумагулова, Г.Б. Абилдаева

Заманауи білім берудегі деректер базасында параллельді есептеулерді жүзеге асыру технологияларын қолдану

Мақалада жоғары оқу орнының білім беру процесінде деректер базасында параллельді есептеулерді оқытуда қолданылатын технологиялар көрсетілген. Бүгінгі күнде кәсіптік білім беру парадигмасын құзыреттілікке өзгертті, оның мақсаты жоғары оқу орнының білім беру процесін білім алушылардың негізгі құзыреттіліктерінің кешенін қалыптастыруға және дамытуға бағыттау болып табылады. Қазіргі уақытта болашақ бағдарламалық жасақтама жасаушылардың мәліметтер базасы саласындағы құзыреттілігінің сипаттамаларын анықтауға, сондай-ақ оларды қалыптастыру негіздерін зерттеуге қолданылатын педагогикалық тәсілдерді талдау нәтижелері көрсеткендей, бүгінде қоғамның мәліметтер базасы саласындағы IT-мамандарды сапалы даярлауға деген қызығушылығының арта түскенін байқауға болады. Бұл жағдайда осындай жоғары құзыреттілікке ие маманды қалыптастыру моделінің дамымағандығын байқауға болады. Жұмыста параллельді есептеу модельдерінің жұмысының негізгі принциптері, сондай-ақ мәліметтер базасымен жұмысты оңтайландыру үшін тапсырмаларды параллельдеу алгоритмдерін қолдану мәселелері көрсетілген. Ұсынылған әдістемелік тәсіл берілген модельдердің көмегімен мәліметтер базасында параллельді есептеулерді жүзеге асыруда негізделген. Ұсынылған зерттеу бағдарламалық жасақтама компанияларының нақты тәжірибесінен есептерді шешу тәжірибесін енгізу негізінде мәліметтер базасында параллельді есептеулерді бағдарламалау дағдыларын қалыптастыру тәсілін ұсынады. Жүргізілген эксперименттің нәтижелері көрсеткендей, бақылау және эксперименттік топтар арасындағы орташа сараптамалық бағалаулардағы айырмашылықтар статистикалық маңызды. Зерттеудің алынған нәтижелері мәліметтер базасында параллельді есептеу бойынша мамандарды әзірлеуде жүзеге асырылуы мүмкін және тиісті дайындық бағыттарының түлектеріне жұмыс орындарын ұсынатын компаниялармен байланысты дамытуға ықпал етуі мүмкін.

Кілт сөздер: деректер базасы, параллельді есептеу, білім алушылар, оқу процесі, кәсіби құзыреттіліктер, әдістеме, жүйе, IT-маман.

M. Serik, A.V. Kopyltsov, S.K. Zhumagulova, G.B. Abildaeva

The use of technologies for implementing parallel computing in a database in modern education

The article outlines the main issues of teaching the implementation of parallel computing in a database in the educational process of a university. Today, the paradigm of vocational education is being replaced by a competence-based one, where the purpose of the orientation of the educational process at the university should be the formation and development of a set of key competencies of students. The analysis of pedagogical approaches used today to determine the essential characteristics of the competence of IT specialists in the field of parallel computing in databases, as well as the study of the theoretical and practical foundations of its formation allowed us to discover the increased need of society for high-quality training of specialists in the field of databases and the insufficient development of a model for the formation of such a highly competent specialist. The paper outlines the basic principles of parallel computing models, as well as the use of task parallelization algorithms to optimize work with databases. The proposed methodological approach to the implementation of parallel computing in a database using the above models is substantiated. The presented research suggests an approach to the formation of parallel computing programming skills in a database based on the implementation of problem solving practices from the real experience of software companies. As the results of the experiment showed, the differences in the average expert estimates between the control and experimental groups are statistically significant. The obtained research results can be implemented in the development of specialists in parallel computing in databases and contribute to the development of relations with companies offering jobs to graduates of relevant training areas.

Keywords: databases, parallel computing, students, learning process, professional competencies, methodology, system, IT specialist.

References

- 1 Bogachev K.Yu. (2003). *Osnovy parallelnogo programmirovaniia* [Fundamentals of parallel programming]. Moscow: BINOM; Laboratoriia znaniia [in Russian].
- 2 Lupin, S.A., & Posypkin, M.A. (2008). *Tekhnologii parallelnogo programmirovaniia* [Technologies of parallel programming]. Moscow: Forum; Infra-M [in Russian].
- 3 Miller, R., & Bokser, L. (2006). *Posledovatelnye i parallelnye algoritmy: Obshchii podkhod* [Sequential and parallel algorithms: General approach]. Moscow: BINOM; Laboratoriia znaniia [in Russian].
- 4 Ponachugin, A.V. (2016). Problemy organizatsii SUBD pri parallelnoi arkhitekture mnogoprotsessornykh vychislitelnykh sistem [Problems of DBMS organization in parallel architecture of multiprocessor computing systems]. *Molodoi uchenyi — Young scientist*, 8 (112), 159–163 [in Russian].
- 5 Nemniugin, S.A. & Stesik O.L. (2002). *Parallelnoe programmirovanie dlia mnogoprotsessornykh vychislitelnykh sistem* [Parallel programming for multiprocessor computing systems]. Saint Petersburg: BKhV–Peterburg [in Russian].
- 6 Serik, M., Zhumagulova, S.K., & Kazimova, D.A. (2021). Osnovnye voprosy obucheniia realizatsii parallelnykh vychislenii v baze dannykh v uchebno protsesse vuzov [The main issues of teaching the implementation of parallel computing in a database in the educational process of universities]. *Shamovskie pedagogicheskie chteniia nauchnoi shkoly upravleniia obrazovatelnyimi sistemami: XIII Mezhdunarodnaia nauchno-prakticheskaia konferentsiia (23 yanvaria–1 fevralia 2021 goda) — Shamov Pedagogical Readings of the Scientific School of Management of Educational Systems: Proceedings of the XIII International Scientific and Practical Conference (January 23 - February 1, 2021)*, 787-789. Moscow: MANPO [in Russian].