

Проблемы исследований и разработок в области управления данными



Сергей Кузнецов

Предыдущие отчеты

- Лагуна Бич, 1988 г.
 - Будущие направления исследований в области баз данных: десять лет спустя, http://www.citforum.ru/database/articles/future_01.shtml
- Пало-Альто, 1990 г.
 - <http://infolab.stanford.edu/~hector/lagi.ps>
- Пало-Альто, 1995 г.
 - Базы данных: достижения и перспективы на пороге 21-го столетия, <http://old.osp.ru/dbms/1996/03/103.htm>
- Кембридж, шт. Массачусетс, 1996 г.
 - Стратегические направления в системах баз данных, http://old.osp.ru/dbms/1997/04/4.htm#part_1
- Асилomar, 1998 г.
 - http://www.citforum.ru/database/digest/asil_01.shtml
- Лоуэлл, шт. Массачусетс, 2003 г.
 - Крупные проблемы и текущие задачи исследований в области баз данных, <http://www.citforum.ru/database/articles/problems/>

Клермонтский отчет, май 2008 г.

- The Claremont Report on Database Research
- Rakesh Agrawal, Anastasia Ailamaki, Philip A. Bernstein, Eric A. Brewer, Michael J. Carey, Surajit Chaudhuri, AnHai Doan, Daniela Florescu, Michael J. Franklin, Hector Garcia Molina, Johannes Gehrke, Le Gruenwald, Laura M. Haas, Alon Y. Halevy, Joseph M. Hellerstein, Yannis E. Ioannidis, Hank F. Korth, Donald Kossmann, Samuel Madden, Roger Magoulas, Beng Chin Ooi, Tim O'Reilly, Raghu Ramakrishnan, Sunita Sarawagi, Michael Stonebraker, Alexander S. Szalay, Gerhard Weikum
- <http://db.cs.berkeley.edu/claremont/claremontreport08.pdf>
- http://www.citforum.ru/database/articles/claremont_report/

Клермонтский отчет, май 2008 г.

- В 2008 г. общий смысл встречи был необычен и достаточно ясен:
 - исследования баз данных и индустрия управления данными находятся в точке поворота, что создает благоприятные возможности для технических и интеллектуальных достижений, предпринимательства и наращивания воздействия на науку и общество
- При наличии большого числа возможностей исследовательскому сообществу важно обратиться к решению проблем, оказывающих максимальное воздействие внутри сообщества баз данных, внутри компьютерного сообщества в целом и во внешних областях

Исследовательские возможности

Пересмотр архитектуры серверов баз данных

- В последнее десятилетие появилось много популярных задач, связанных с обработкой больших объемов данных, для которых реляционные СУБД обеспечивают плохое соотношение «цена/производительность», и при решении которых от использования РСУБД пришлось отказаться:
 - индексирование текста, обслуживание Web-страниц и доставка мультимедийного контента
- Новые рабочие нагрузки появляются в научных приложениях и приложениях в стиле Web 2.0

Исследовательские возможности

Пересмотр архитектуры серверов баз данных

- К числу наиболее важных исследовательских тем в этой области относятся:
 - разработка систем для кластеров многоядерных процессоров, в которых имеется ограниченный и неоднородный доступ к памяти вне кристалла;
 - использование удаленной основной и флэш-памяти в качестве среды персистентного хранения данных в дополнение к памяти на магнитных дисках;
 - **Гоц Грейф. Правило пяти минут двадцать лет спустя, и как флэш-память изменяет правила.** http://citforum.ru/database/articles/five_minute_rule/
 - разработка унифицированного подхода к постоянно выполняемой адаптации и самонастройке оптимизации запросов и физических структур хранения данных;
 - сжатие и шифрование данных на уровне хранения, интегрированное со структурой хранения и оптимизацией запросов;
 - разработка систем, опирающихся на не реляционные модели данных, вместо того, чтобы «впихивать» эти данные в таблицы;
 - нахождение компромиссов между согласованностью и доступностью для достижения лучшей производительности и масштабируемости уровня тысяч машин;
 - разработка СУБД, учитывающих потребление энергии, которые ограничивают энергопотребление без ущерба для масштабируемости

Исследовательские возможности

Декларативное программирование для новых платформ

- Хотя разработка новых парадигм программирования не является проблемой баз данных как таковых, декларативное программирование и оптимизация на основе оценок обеспечивают обнадеживающий угол атаки
- Имеются существенные основания, чтобы полагать, что подходы, ориентированные на обработку данных, в краткосрочной перспективе могут сильно воздействовать на программирование в целом

Исследовательские возможности

Декларативное программирование для новых платформ

- Одним из примеров потенциала этих подходов является наблюдаемая в последнее время популярность Map-Reduce
 - <http://labs.google.com/papers/mapreduce-osdi04.pdf>
- Другим примером может служить то, что в последнее время разрабатываются новые декларативные языки, часто основывающиеся на Datalog для разных проблемно-ориентированных систем в различных областях:
 - сетевых и распределенных систем,
 - компьютерных игр,
 - машинного обучения и робототехники,
 - компиляторов, протоколов безопасности и
 - извлечения информации

Исследовательские возможности

Декларативное программирование для новых платформ

- Третий пример происходит из области программирования корпоративных приложений
- В недавно появившихся языковых расширениях, таких как Ruby on Rails
 - <http://www.rubyonrails.ru/>
- и LINQ
 - <http://msdn.microsoft.com/en-us/netframework/aa904594.aspx>, поддерживается логика запросов в паттернах конструирования программ
- Одним из существующих языков является XQuery
 - <http://www.citforum.ru/internet/articles/xqlzxml.shtml>, который может способствовать этому виду декларативного программирования

Исследовательские возможности

Взаимосвязь структурированных и неструктурированных данных

- Значительной долговременной целью сообщества баз данных является переход от управления традиционными базами данных к намного более сложной задаче управления обширными коллекциями структурированных, полуструктурированных и неструктурированных данных, распределенных по многим репозиториям предприятий и Web
- Иногда это называют проблемой управления пространствами данных
- *Про пространства данных см. перевод статьи «От баз данных к пространствам данных: новая абстракция управления информацией», http://www.citforum.ru/database/articles/from_db_to_ds/ и мой обзор «Предвестники новых манифестов управления данными», <http://www.citforum.ru/database/articles/premanifest/>*

Исследовательские возможности

Взаимосвязь структурированных и неструктурированных данных

- Разработке методов эффективной обработки запросов к результирующему множеству разнородных данных и обеспечения понимания результатов этих запросов
- Конкретной проблемой является обеспечение ответов на запросы по ключевым словам над большими коллекциями разнородных источников данных
- Требуется анализировать запросы для извлечения их предполагаемой семантики и направлять данный запрос к уместным источникам данных в коллекции
- Не предполагаются наличие семантического отображения источников данных и известность прикладной области запроса и источников данных

Исследовательские возможности

Облачные службы данных

- Экономические факторы приводят к появлению инфраструктур, обеспечивающих программные и вычислительные средства в виде служб, обычно называемых *облачными (cloud)* сервисами, или облачным компьютерингом
- Имеется набор разнообразных облачных сервисов, включая
 - прикладные сервисы (salesforce.com),
 - сервисы хранения Amazon S3),
 - вычислительные сервисы (Google App Engine, Amazon EC2, и
 - сервисы данных (Amazon SimpleDB, Microsoft SQL Server Data Services, Google's Datastore)
- Многие будущие приложения, ориентированные на обработку данных, будут опираться на облачные сервисы данных

Исследовательские возможности

Облачные службы данных

- В облачных средах особенно важным качеством является управляемость
- Потребность в управляемости делает более срочной разработку технологий самоуправления баз данных, которые исследовались в последнее десятилетие
- Отдельной проблемой является абсолютный масштаб облачного компьютеринга
- Сегодняшние SQL-ориентированные системы баз данных просто не могут масштабироваться на тысячи узлов при размещении в облачном контексте
- При совместном использовании физических ресурсов в облачной инфраструктуре требуется обеспечение безопасности и конфиденциальности данных, которые не могут гарантироваться за счет наличия физического разграничения машин или сетей
- Следовательно, облачные сервисы обеспечивают плодородную почву для усилий по объединению и ускорению исследований, выполняемых сообществом баз данных в этих областях

Исследовательские возможности

Мобильные приложения и виртуальные миры

- Имеется новый класс приложений, примерами которых являются мобильные службы и виртуальные миры, характеризуемый потребностью
 - в управлении крупными объемами разнообразных данных, создаваемых пользователями,
 - их интеллектуальной синтезе и
 - обеспечении служб реального времени
- Сообщество управления данными начинает понимать проблемы, стоящие перед этими приложениями, но здесь требуется намного больше исследований
- Соответственно, обсуждение этих тем на встрече носило более умозрительный характер, чем тем, которым посвящались предыдущие подразделы, но участники встречи сочли их заслуживающими внимания в своем отчете

Исследовательские возможности

Другие темы

- Кроме тем, обсуждаемых ниже, во время встречи многократно упоминались следующие проблемы и области исследований:
 - управление неопределенной информацией,
 - конфиденциальность и безопасность данных,
 - электронная наука и другие приложения, ориентированные на ученых,
 - антропоцентрические взаимодействия с данными,
 - социальные сети и Web 2.0,
 - персонализация и контекстуализация задач, относящихся к выполнению запросов и поиска,
 - потоковые и сетевые данные,
 - самонастраиваемые и адаптивные системы, а также
 - проблемы, порождаемые новыми технологиями аппаратуры и ограничениями энергопотребления