

СТЕНДОВЫЕ ДОКЛАДЫ

Комплекс "Библиотечная статистика" - новый взгляд на сбор и обработку данных

**Поникаровский Константин Викторович, программист,
Институт корпоративных библиотечно-
информационных систем Санкт-Петербургского
государственного политехнического университета,
Санкт-Петербург**

В докладе рассматриваются задачи, решаемые программным комплексом "Библиотечная статистика". Обоснована необходимость применения специального программного обеспечения для автоматизации сбора статистики работы библиотеки по широкому кругу показателей. Раскрыты основные особенности комплекса - независимость от используемых в библиотеке автоматизированных средств (АБИС), распределительные возможности комплекса, использование новых технологий (MOLAP) и новых решений - работа с отчётами непосредственно в MS Excel.

По мере автоматизации библиотечных процессов появляется возможность получать объективные данные о работе различных сотрудников, отделов и библиотеки в целом по широкому кругу показателей.

Традиционно программы библиотечной статистики настраиваются на набор заранее сформированных запросов, позволяющих получить типовые показатели работы библиотеки. Однако при желании сформировать иной запрос, получить новые показатели или же узнать зависимости показателей от новых параметров необходимо обращаться к программистам-разработчикам для доработки программы.

Современные технологии баз данных предлагают новое решение – аналитические системы, позволяющие варьировать запросы пользователей в очень широком диапазоне без изменения программного кода. Кроме того, аналитические базы данных организованы особым образом, позволяющим выполнять обработку значительных массивов данных за весьма непродолжительное время – секунды. Всё это стало возможно благодаря MOLAP технологии.

MOLAP (Multidimensional Online Analytical Processing, аналитическая обработка в реальном времени со многими измерениями) — технология обработки информации, включающая составление и динамическую публикацию отчётов и документов. Используется для быстрой обработки сложных запросов к базе данных. Причина использования MOLAP для обработки запросов — это скорость. MOLAP делает мгновенный снимок реляционной БД и структурирует её в пространственную модель для

запросов. Заявленное время обработки запросов в MOLAP составляет около 0.1 % от аналогичных запросов в реляционную БД.

Реляционные БД хранят сущности в отдельных таблицах, которые обычно хорошо нормализованы. Эта структура удобна для операционных БД, но сложные многотабличные запросы в ней выполняются относительно медленно. Более хорошей моделью для запросов, а не для изменения, является организация данных в виде гиперкубов.

Гиперкуб - это MOLAP-структура, созданная из рабочих данных. Гиперкуб создаётся из соединения таблиц с применением схемы звезды. В центре «звезды» находится таблица фактов или параметров, по которым делаются запросы. Множественные таблицы с измерениями присоединены к таблице фактов. Эти таблицы показывают, как могут анализироваться агрегированные реляционные данные. Количество возможных агрегирований определяется количеством способов, которыми первоначальные данные могут быть иерархически отображены.

Например, показатели книговыдачи могут быть сгруппированы по библиотекам, отделам библиотек и сотрудникам библиотек, таким образом, 50 сотрудников, 8 отделов и 2 библиотеки составят 3 уровня иерархии с 60 членами.

MOLAP-куб содержит в себе базовые данные и информацию об измерениях (агрегатах). Куб потенциально содержит всю информацию, которая может потребоваться для ответов на любые запросы. Из-за громадного количества агрегатов, зачастую полный расчёт происходит только для некоторых измерений, для остальных же производится «по требованию».

Итак, главные достоинства MOLAP технологии:

- совершенно новый подход к хранению данных в гиперкубах, позволяет делать так называемые срезы данных по любым заложенным в гиперкуб измерениям, то есть существует возможность осуществления любых «срезов» и «углублений» в определённой структуре БД;
- самое быстрое построение отчётов в сравнении как с реляционными базами данными так и с другими вариантами OLAP подхода;
- слабое влияние объёма аналитической базы данных на скорость построения отчётов.

Основной недостаток технологии обусловлен новыми подходами к хранению данных и заключается в сложной и зачастую уникальной для каждого продукта модели запросов. То есть стандартный механизм SQL-запросов к кубам данным не применим. Однако в большинстве случаев, там где используется MOLAP технология - нет необходимости в SQL, и того механизма запросов, который предлагается разработчиками продукта достаточно.

Программный комплекс "Библиотечная статистика" служит для автоматического сбора статистики работы библиотеки и быстрого, а главное, удобного представления результатов конечному пользователю. Скорость

обеспечивается новыми технологиями, а удобство - новыми решениями. Например, вся работа по созданию и редактированию отчётов ведётся непосредственно в программе MS Excel.



Microsoft Excel - По отделам и сотрудникам (с выбором даты).xls

Настройки

Дата: 30.04.2008

Удаление записей сотрудников, у которых показатели выдачи/возврата меньше: 1

Пересчёт

Пересчёт

Распределение книговыдачи по отделам и сотрудникам

		Выдача	Возврат	Продление
14	ОУЛ Итого	162,00	144,00	6,00
15	Акимжанова Т.А.	8,00	2,00	0,00
16	Валягина Е.П.	6,00	6,00	0,00
17	Васильева Т.И.	24,00	14,00	2,00
18	Чегнева Н.Ю.	4,00	5,00	0,00
19	Шейпацкая Е.Б.	17,00	30,00	0,00
20	ОНЛ Итого	295,00	65,00	29,00
21	Елагина Ю.К.	218,00	35,00	2,00
22	Загарьева А.И.	7,00	4,00	1,00
23	Михайлова О.П.	34,00	4,00	21,00
24	Суратова Н.А.	13,00	15,00	4,00

Программный комплекс меняет представление о построении статистических отчётов:

- Нет никаких предустановленных шаблонных отчётов, обычно предлагаемых разработчиками, можно придать любую функциональность и оформление отчёту. За функциональность отвечает MOLAP технология, а в оформлении поможет программа MS

Excel, возможностями которой может воспользоваться составляющий отчёт - это, например, и написание макросов и построение диаграмм;

- Отчёты любой сложности строятся за считанные секунды, при этом напрямую никакие данные с сервера "Руслан" не берутся;
- Сбор статистики полностью автоматизирован. Комплекс устанавливается и настраивается один раз, а статистика собирается постоянно без участия человека. Обновление базы статистики настраивается - например, Вы можете указать ночное время для пополнения базы статистики, при этом минимизируя нагрузку на сервер "Руслан". Новые данные извлекаются в течение нескольких минут и хранятся в не зависящем от АБИС " Руслан" хранилище по технологии гиперкубов;
- Комплекс является распределённой системой. Одновременно с базой статистики может работать любое количество пользователей. При этом каждый из них работает в Excel на своём компьютере, условие здесь одно - все компьютеры должны быть связаны между собой в локальную сеть;
- В одном отчёте отображается до шести различных параметров. При этом используется не одно, а несколько временных измерений - отчёты можно строить за любой временной интервал (от года до пяти минут).

В комплексе "Библиотечная статистика" реализовано построение гиперкуба по книговыдаче, имеющего следующие параметры, по которым может производиться "срез" данных:

В первую очередь, это основные показатели по книговыдаче. К их числу относятся:

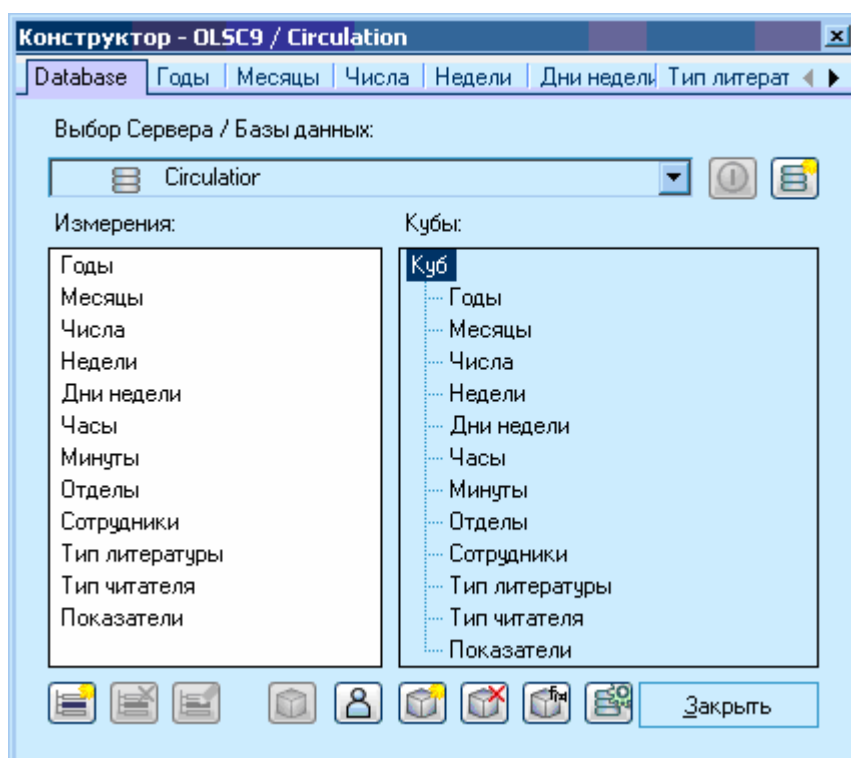
- выдача;
- возврат;
- продление литературы.

А так же в базу статистики введены такие измерения (параметры), как:

- отделы библиотеки;
- сотрудники библиотеки;
- тип литературы;
- контингент читателей.

К числу временных измерений относятся:

- год;
- месяц;
- число;
- неделя;
- день недели;
- час;
- пятиминутный интервал.



Именно статистика по книговыдаче была выбрана как первоочередная задача. Это связано с тем, что ручной сбор подобной статистики - наиболее трудоёмок. А существующие автоматизированные аналоги зачастую не предоставляют достаточное количество различных параметров (просмотр статистики по отделам, сотрудникам и т.д., а так же любым комбинациям этих показателей) и детализированностью временных показателей (например, распределение нагрузки по часам за сутки или по неделям за конкретный месяц и многое другое). Немаловажное значение в статистике обслуживания читателей имеет и учёт продления литературы.

Для каких целей могут использоваться отчёты комплекса "Библиотечная статистика" по обслуживанию читателей?

Прежде всего, для руководителей абонементов и читальных залов и не только для построения квартальных или годовых отчётов, но так же, например, и для определения нагрузки своих сотрудников. Для предоставления информации кафедрам и факультетам о предоставлении услуг их студентам, если например, с кафедрами и факультетами установлены деловые? отношения. А так же для наблюдения в целом за работой библиотеки тем, кому это будет интересно - получение фактических данных, которые раньше получить было невозможно - это и определение контингента реально обслуженных читателей за тот или иной период (в отличие от просто зарегистрированных в библиотеке) и определение интересов читателей по типам литературы.

Важно отметить, что комплекс "Библиотечная статистика" в данный момент совместим с АБИС "Руслан", откуда берутся данные для накопления в аналитической базе данных. Соответственно, может быть осуществлена

доработка комплекса для других АБИС или любых других источников, из которых можно извлечь необходимые для построения гиперкуба данные.

В качестве хранилища под аналитическую базу данных используется Palo Server (Jedox GmbH). Это продукт фирмы Jedox, который имеет бесплатную лицензию (GNU General Public License) и открытые исходные коды.

Комплекс "Библиотечная статистика" не стоит на месте и постоянно развивается. Работа над построением гиперкуба выбранной первоочередной статистики по книговыдаче завершена, и идёт работа по повышению функциональности комплекса - сбору статистики в других сферах библиотечной деятельности. Это в первую очередь статистика по книгохранению и по читателям. Для достижения функциональности по этим показателям, требуется разработка и построение новых гиперкубов данных.

Ознакомительную версию комплекса, более подробное описание, а также подробную документацию можно загрузить с сайта ООО «Открытые библиотечные системы» <http://obs.ruslan.ru/>.