

ИСПОЛЬЗОВАНИЕ ОТКРЫТЫХ СУБД, ПРИЛОЖЕНИЙ И КОМПОНЕНТОВ ДЛЯ РАЗРАБОТКИ АНАЛИТИЧЕСКИХ СИСТЕМ.

Докладчик: Микитюк С.В. директор Silentium Company

Russian Open Source Summit'2014

Введение

Наша компания занимается разработкой программного обеспечения для бизнеса и госучреждений. Хотелось бы поделиться опытом в этой области и постараться расставить акценты, показать преимущества использования свободного программного обеспечения при выполнении IT-разработок.

Любой проект в области информационных технологий начинается с красивой идеи. И, как правило, такая идея изначально кажется достаточно простой и легко реализуемой на практике. Однако первое впечатление может быть обманчиво, до тех пор, пока не начинается расчет бюджета проекта. Когда финансовые выкладки готовы, итоговые суммы, зачастую получаются в разы больше, чем казалось на первый взгляд. Далее есть два пути: либо проект откладывается на неопределенное время, либо находятся способы снижения издержек. Наиболее рациональные из них это:

- Сокращение цикла разработки
- Активное использование ресурсов заказчика
- Использование свободного ПО

Однако снижение стоимости не является самой целью. Гораздо важнее дать заказчикам уверенность в том, что финансирование проекта является надежным вложением средств в организацию (бизнес, госструктуру). Именно поэтому руководству и акционерам нужно быть уверенными в том, что в конце проекта будет получен результат не только в виде работающей системы. Лучший вариант, когда в руках заказчика остается открытая для дальнейшего развития платформа с прозрачной архитектурой и возможностью привлечения как своих специалистов, так и программистов сторонних компаний. Следует также отметить, что вопрос открытости неразрывно связан с безопасностью бизнеса, надежностью работы контролируемых процессов.

Вот почему вопрос открытого кода, масштабируемых платформ выходит на первое место. Процесс развития даже небольших систем становится предметом труда различных специалистов (аналитиков, системных архитекторов, интеграторов, системных администраторов, дизайнеров). Присутствие в составе компании или организации нескольких работающих вместе информационных систем требует не просто внедрения их на отдельных участках, но и их совместное функционирование. Наличие единых корпоративных стандартов обеспечивает надежность и прозрачность. Все вышеперечисленное может предоставить открытая платформа, на которую могут опираться дополнительные подсистемы. Требуется ядро, которое будет прозрачно, логично, доступно для дальнейшего развития. Современному предприятию требуется именно такая платформа, на которую можно опереться для подключения новых модулей, приложений и сервисов.

В области привлечения разработчиков также давно наблюдаются тенденция к частой смене места работы, что может негативно сказаться на функционировании программных продуктов. Даже если команда программистов меняется процесс развития информационной структуры должен продолжаться и быть независимым от того где будут трудиться ее нынешние разработчики. Даже если имеются все исходные тексты программ, это не значит, что новый инженер, придя в проект, сможет быстро что-либо в нем изменять, не видя картины в целом. Вот почему помимо открытого кода требуется стройная логичная архитектура разрабатываемых продуктов. Всему должно быть свое логическое место. Требования к проектам должны быть следующие:

- Открытая архитектура
- Использование платформы для построения приложений
- Возможность замены отдельных узлов информационной системы
- Наличие всех исходных кодов программ

Для того чтобы достичь перечисленных выше целей требуется план развития, который начинается с выбора платформы, СУБД, возможностей интеграции, использования готовых программ, экспорта данных в офисные пакеты, подключения средств анализа данных.

1. Как использование свободного ПО позволяет привлекать потенциальных клиентов

Для того чтобы снизить расходы на лицензии мы предлагаем нашим клиентам следующие шаги:

- Установка серверных операционных систем семейства Linux
- Использование OpenOffice.org в качестве офисного пакета
- Использование бесплатных версий СУБД
- Использование для web-доступа Apache, PHP

Такие шаги не являются обязательными и имеют как положительные, так и отрицательные стороны и зависят от готовности персонала перейти на перечисленные программные продукты. Выбор СУБД так же может быть неоднозначным и зависит от внутренних стандартов компании/организации. Кроме того, многие лицензии могут быть закуплены уже задолго до начала проекта.



Рис. 1 Направления использования свободного ПО.

Наша задача как профессионального разработчика систем – предложить клиентам различные варианты использования программ в разной комплектации. Следовательно, при построении систем требуется универсальность подключения баз данных, офисных приложений и выбора операционных систем. Это же касается и web-разработок.

2. Преимущества разработки баз данных с использованием Firebird и mySQL.

Выбор базы данных в основном обусловлен сложностью проекта, необходимостью тиражирования баз данных, подготовленностью системных администраторов, разработчиков. Наиболее эффективные решения следующие:

- mySQL
- Firebird
- Oracle DB



Рис. 2 Использование популярных СУБД.

Соответственно для функционирования web-решений максимально подходит mySQL, тем более что в его последних версиях появилась возможность использования триггеров и хранимых процедур. Выполненные не нем проекты удобно разворачивать на хостинг-провайдере. Практически на всех хостингах установлена эта популярная СУБД.

Firebird лучше всего подходит для Windows-приложений и удобен для быстрого развертывания системы с помощью простого инсталлятора. Он неприхотлив в эксплуатации. У Firebird имеется большой набор средств для написания серверного кода. Он достаточно демократичен в проектировании таблиц, не нужно тратить время на тонкие мелочи, которые зачастую не пропускает Oracle. Также есть возможность подключения библиотек на стороне сервера в виде dll-расширений.

Oracle имеет массу достоинств, но он сложен в администрировании и требует отличной подготовки разработчиков. Он имеет высокую стоимость владения. Oracle особенно эффективен для больших объемов данных и сложных алгоритмов. К ним относятся, например, прогнозирование, моделирование, оптимизация запасов. Эта СУБД позволяет полностью реализовать расчеты на стороне сервера и максимально отойти от настольных систем.

3. Интеграция с офисными приложениями на примере OpenOffice

Очень часто заказчики информационных систем требуют возможности универсальной выгрузки данных для дальнейшего наращивания инструментов обработки информации. Наиболее удобно использовать для этого MS Excel, или OpenOffice Calc, так как это с одной стороны структурированный поток данных с которым могут работать другие системы, с другой стороны эти данные можно легко просматривать, переносить и при необходимости исправлять.

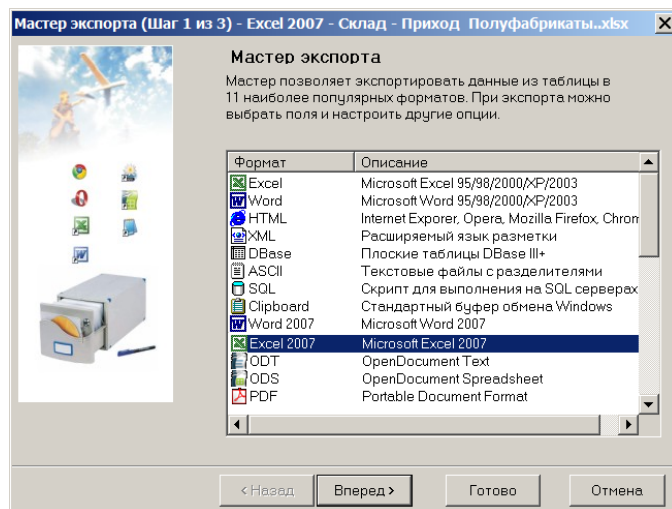


Рис. 3 Возможность экспорта данных.

Также необходимы и другие не менее важные форматы: HTML, XML, PDF и так далее. Мы решили эти вопросы одним большим блоком экспорта, который может легко подключаться к любому журналу, справочнику, каталогу, реализованному с помощью нашей платформы - продукта Silentium ETNI Platform.

Помимо универсального экспорта была реализована возможность построения более сложного вывода информации в виде ведущедетализированных конструкций, группировки данных с помощью встроенных инструкций. Ниже приведен простой пример выгрузки плоской таблицы.

[ROLLTEXT.Имя_источника]

Наименование: **[GOODSNAME]** Количество: **[QUANTITY]** Ед.изм.: **[UNITNAME]**

[/ROLLTEXT.Имя_источника]

Рис. 4 Макроподстановки экспорта.

Таким образом, получается возможность не только выгрузки данных, но и формирования сложных первичных документов, отчетов с возможностью настройки (доработки) шаблонов непосредственно опытным пользователем или системным администратором. Заложенная в проект гибкость дает свободу по улучшению вывода на печать, наращивания возможностей без увеличения стоимости проекта.

Для облегчения интеграции с OpenOffice нашими разработчиками был создан специальный компонент TSiOpenOffice. Для интеграции в приложениях на Object Pascal вынесены SheetName, SheetIndex свойства и реализованы основные методы:

- Connect, Disconnect,
- CreateFile, OpenFile,
- SaveFile, SaveAsFile
- WriteCellText, WriteCellFormula, WriteCellTextByName
- GoToCell, InsertRows, InsertCols

Общий подход создания универсальных компонентов оправдал себя. Он позволяет резко уменьшить цикл написания программ и дает возможность разнести большой объем работ по программированию на разных специалистов.

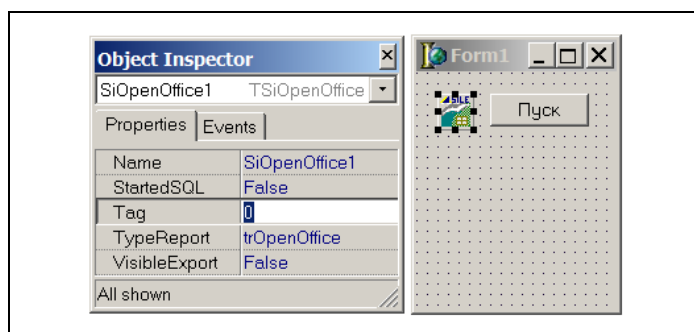


Рис. 5 Использование готовых компонентов.

Кроме того, со временем необходимо не только улучшать написанные компоненты, но и полностью замещать их новыми в соответствии с требованием времени. При этом компоненты тестируются и пишутся отдельно, а затем достаточно легко встраиваются внутрь работающих проектов.

Такой подход требует не просто разнесения кода программ на отдельные компоненты, но и четко выстроенной архитектуры приложений. При этом крайне желательно чтобы такая архитектура не сильно отличалась от проекта к проекту. Это позволяет быстро наращивать функциональность, проводя работы параллельно и независимо разными командами.

Предлагаемая нами архитектура приложений подразумевает для разных проектов различные базы данных по структуре и по типу (производителю). Несмотря на это клиентские места являются универсальными и настраиваемым в зависимости от потребностей заказчика. То есть компания Silentium предлагает универсальную платформу для информационных систем. Специфику каждого проекта для каждого клиента необходимо выносить за рамки программного кода. В таком случае для появления новых справочников, журналов, отчетов, аналитических выборок нет необходимости заново компилировать существующие модули.

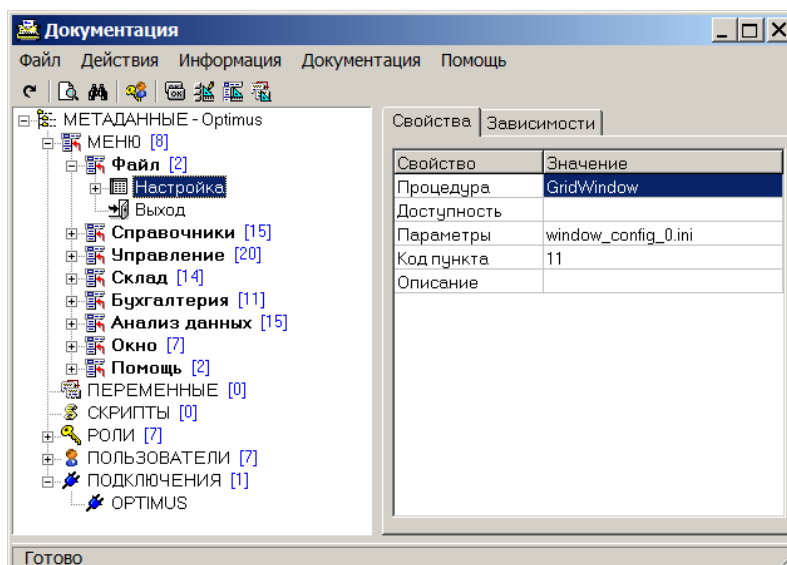


Рис. 6 Навигация по настройкам платформы.

Это сокращает трудоемкость программирования во много раз. Там где раньше требовалось потратить 2-3 дня уходит от двух до трех часов, что дает время для тестирования, внедрения новых решений. Что касается специфических настроек проекта, то они также должны быть максимально открытыми. Для этого:

- Хранение настроек организовано в виде текстовых (ini) файлов
- Создана специальная утилита для вывода документации по системе
- Организован сквозной поиск нужных настроек, ссылок
- Сделана диагностика противоречий, отступлений от стиля с выводом предупреждений

4. Построение многомерных кубов OLAP средствами PHP.

Для аналитической обработки был разработан механизм многомерного анализа данных средствами PHP с расчетом агрегатов на стороне сервера. Библиотека классов SiWebCube также имеет разделение на реализацию интерфейса и настройки.

```
[SQL]
SELECT
  BYDATE, BYMONTH, MAKERNAME, GOODSNAME, GROUPGOODSNAME,
  CUSTOMERNAME, USERNAME, REGIONNAME, ZONENAME,
  ROUND(PARTY_PRICE, 0) AS PARTY_PRICE,
  ROUND(PROFIT, 0) AS PROFIT
FROM RAIN_SELLS

[FIELDS]
BYMONTH      | Месяц      | | | DIM |
GROUPGOODSNAME | Группа     | | | DIM |
GOODSNAME     | Товар      | | | DIM |
MAKERNAME     | Производитель | | | DIM |
CUSTOMERNAME  | Покупатель | | | DIM |
USERNAME      | Менеджер   | | | DIM |
REGIONNAME    | Район      | | | DIM |
ZONENAME      | Зона       | | | DIM |
BYDATE        | Дата       | | | DIM |
PARTY_PRICE   | Сумма      | | | SUM |
PROFIT        | Прибыль    | | | SUM |
[END ALL]
```

Рис. 7 Настройка доступа к данным.

На картинке представленной выше показан файл настройки куба в упрощенном варианте. Одна секция файла отвечает за SQL-запрос который готовит данные, вторая секция помогает подписать выходные поля и сообщает, как с каждым из них работать. Для того чтобы снизить объем вычислений, получаемые данные сначала преобразуются в дерево, затем делаются расчеты и только после этого результаты выводятся в виде сложной таблицы, которая затем может быть транспонирована, перестроена, подвергнута проверке на предмет исходных данных.

Web-Cube 2.1

Доступно	Выбрано	Вычисления
Месяц Группа Товар Производитель Покупатель > < << >>	Месяц Группа < >	Сумма Сумма Прибыль < >

Месяц	ВАФЛИ	ДРАЖЕ	СУХАРИ	ШОКОЛАД	Итого
01 январь	60744845	5315418	41005309	73270483	180336055
02 февраль	66211209	12048188	41128206	90529037	209916640
03 март	66692017	14026495	53969612	126102092	260790216
04 апрель	62110678	15327815	45602883	63925717	186967093
05 май	70377649	5517549	35659708	69357453	180912359
06 июнь	50331651	6798436	30382463	29251800	116764350
07 июль	60160494	12246464	29741998	37108837	139257793
08 август	57867635	8714299	33398355	43496508	143476797
09 сентябрь	57185787	8491915	31894668	60690880	158263250
10 октябрь	49320389	4092566	31015310	49358638	133786903
11 ноябрь	53162495	223430	31287882	34334335	119008142
12 декабрь	58681453	280012	41365901	100680677	201008043
Итого	712846302	93082587	446452295	778106457	2030487641

Рис. 8 Развертывание многомерного куба.

Таким образом, реализация одного компонента позволяет настраивать множество ini-файлов и разработка аналитики сводиться к программированию SQL-запросов, проверке получаемых результатов, внедрению у

заказчика. Рекомендуем разработчикам при такой настройке сверять данные с регулярными отчетами и первичными документами на контрольных примерах.

Открытый код РНР позволяет и далее развивать механизмы библиотеки классов. Блоки программ оформлены в виде самостоятельных процедур и классов, которые могут наследоваться. Однако ознакомление с ними требует дополнительного времени, поэтому по ним готовится специальная документация с прилагаемыми примерами.

5. Бесплатное приложение ETNI Fly, как мост между настольными приложениями и централизованными базами данных.

Для построения небольших приложений компанией Silentium в 2012 году был выпущен конструктор пользовательского интерфейса ETNI Fly. Его задача построение пилотных модулей будущих приложений, а также обеспечение настольной обработки данных. Результатом настройки являются специальные etni-файлы, имеющие открытый формат и содержащие в себе не только метаданные, вспомогательную информацию, настройки подключения, но и непосредственно данные пользователя, разворачивающиеся в виртуальной СУБД.



Рис. 9 Конструктор интерфейса ETNI Fly.

Основные функции ETNI Fly

- Предоставление интерфейса ввода информации
- Навигация, поиск, фильтрация, экспорт
- Построение отчетов
- Построение графиков
- Многомерный анализ данных (OLAP)
- Сбор и передача данных на web-серверы
- Доставка данных с web-серверов на локальные компьютеры

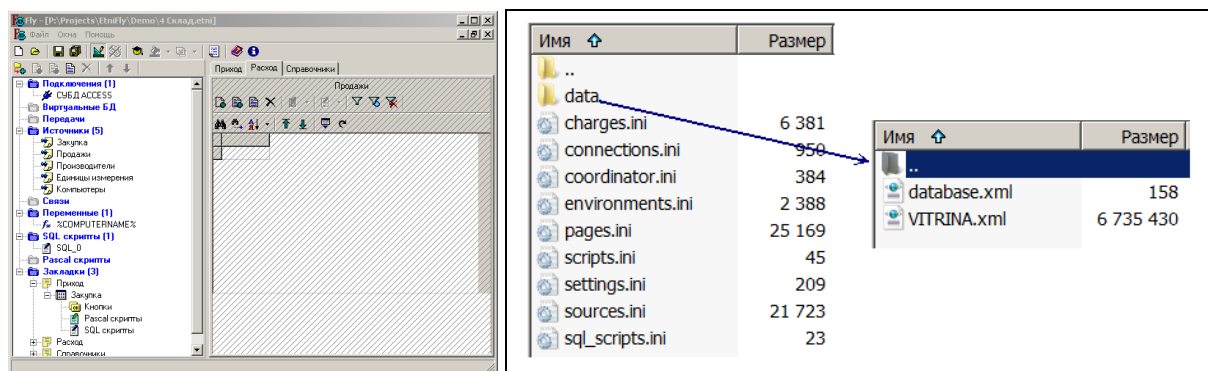


Рис. 10 Возможности настройки конструктора и структура внутренних файлов.

Все вышеперечисленные технологии и решения выполнены исходя из требования, что информационная система больше не может развиваться силами только одного человека, группы лиц или даже компании. Современная система должна контролироваться заказчиком, развиваться вместе с его бизнесом, даже, несмотря на ее возможную сложность.

Микитюк Сергей Владимирович

Компания Silentium

Москва

11.04.2014 г.