

ИСПОЛЬЗОВАНИЕ ТЕХНОЛОГИЙ ETL И OLAP В АНАЛИЗЕ ФИНАНСОВЫХ ОРГАНИЗАЦИЙ

А.О. Мелащенко

Институт кибернетики им. В.М. Глушкова НАН Украины,
03028, Киев, проспект Науки 43, кв. 70.
Тел.: 524 5330, javatask@ukr.net

Рассматриваются вопросы использования решений BI в финансовых организациях Украины. Приводятся практические рекомендации по построению OLAP решений и их применению для прогнозирования бухгалтерского баланса.

In the report questions of use of decisions BI in the financial organisations of Ukraine are considered. Practical recommendations about construction OLAP of decisions and their application for forecasting of accounting balance are resulted.

Введение

Потребность в постоянном совершенствовании бизнес-процессов, повышения прибыльности, снижение себестоимости, расширение рынков сбыта предприятий невозможно без анализа. Как правило, бизнес анализ необходим на всех стадиях жизненного цикла продукта и во всех подразделениях предприятия. Что в свою очередь приводит к потребности обработки громадных объемов информации, ее структурирование, написание интерфейсов доступа к ней, обеспечения безопасности, целостности и других проблем связанных с анализом. Все это требует высокопрофессиональных сотрудников отдела IT, что в свою очередь является дорогим «удовольствием».

С другой стороны, в частности, деятельность финансовых организаций, в частности банков, достаточно формализована для создания типовых решений. Основными составляющими базиса анализа качества функционирования банка являются:

1. Клиентская база, которая в свою очередь делится на:
 - а) связанных лиц;
 - б) несвязанных лиц.
2. Баланс банка в разрезе балансовых счетов.
3. Кредитный портфель.
4. Депозитный портфель.

Имея удобные и гибкие средства, позволяющие отобразить нужные разрезы этой информации, управленцам и аналитикам, можно принимать оптимальные решения для улучшения всех показателей функционирования. Как технология обработки информации OLAP, допускают многомерные представления данных плюс оптимизацию скорости доступа к данным в сравнении с OLTP.

Но OLAP является законченной формой хранения данных, которые извлекаются из OLTP БД. Для преобразования данных в нужные структуры используются, так называемые ETL инструменты.

Также для доступа к данным OLAP нужны «красивые» средства, нежели API или MDX. Такими средствами являются или программы обработки электронных таблиц, которые обычно включают клиента OLAP БД или широко используются системы построения отчетов и клиенты pivot-таблиц.

Три технологии ETL, OLAP и генераторы отчетов совместно работают на BI.

В этом докладе, приведена концепция построения типовых OLAP решения и пример предоставления базового отчета для анализа баланса банка.

Ключи: OLAP, BI, SSIS, REPORTS, ETL

ГЛОССАРИЙ

БД – База данных
OLAP (Online analysis processing) – Аналитическая обработка в реальном времени
OLTP (Online transaction processing) – Обработка транзакций в реальном времени
ETL (Extract Transform Load) – Извлечь Преобразовать Загрузить
MDX (Multidimensional Expressions) – Многомерный язык запросов

BI (Business Intelligent) – Бизнес анализ
VFSFS – Visual studio for sql server
ПО – Программное обеспечение
SQL (Structural query language) – Структурированный язык запросов
DDL (Data definition language) – Язык описания структуры данных
НБУ – Национальный банк Украины
ОДБ – Операционный день банка

1. Постановка задачі

В любом банке более 80% отчетов и нормативов строится на основе бухгалтерского баланса, соответственно при проверке нормативов и отчетности аналитику и управленцу критично необходимо «видеть» баланс в удобной структурированной форме.

Использование готовых западных OLAP-решений сомнительна и даже невозможна. Это связано с громадным количеством отчетности, требуемых контролирующими органами в Украине, как государственными так и частными. В связи с этим, отечественные разработчики идут по пути наименьшего сопротивления и создают продукты, которые учитывают множество избыточной информации, необходимой для построения отчетности. Поэтому множество тонких моментов, важных для функционирования банка, теряются в громадных объемах информации.

Выходом из такой ситуации является ведение двух БД, одна содержит всю необходимую информацию для построения отчетности, как правило, это OLTP БД, вторая содержит «чистые данные», позволяющие адекватно оценивать качество функционирования банка, как правило, это OLAP БД. Так что приходится заново преодолевать путь построения OLAP систем.

Краткое изложение основных этапов построения OLAP решений продемонстрировано на задаче построения пользовательского разреза баланса. Эта задача поставлена в отделе аналитики одного крупного банка. Также продемонстрированы результаты прогнозирования состояния банка.

Детали задачи. Количество балансовых счетов бухгалтерского анализа состоит из порядка 1000 счетов, которые характеризуются так называемыми активностью и пассивностью. Эти счета сгруппированы, но каждый банк имеет практику выстраивать свою иерархию для отображения «видения» управляющего персонала. Каждый балансовый счет ниже по иерархии содержит лицевые счета, которые составляют сумму балансового счета.

Задача стоит так: построить разрез структурированного баланса, по дням с отображением остатков по категориям.

Эта задача демонстрирует основной подход и предположения используемые при построении системы BI на предприятии.

2. Этапы построения OLAP решения

Этап построения законченного аналитического приложения можно разбить на три базовых этапа.

1. Создание схемы OLAP БД.
2. Загрузка и преобразование (опционально) данных из OLTP БД в OLAP БД.
3. Построение необходимого куба (кубов), метрик и измерений.
4. Построение отображения данных пользователю.

Этап 1. Создание схемы OLAP БД. На данном этапе строится реляционная схема OLAP БД. Как правило, она состоит из:

- таблиц фактов – содержит в себе факты, например, остаток на счете, объем продаж, число привлеченных клиентов;
- таблиц измерений – содержит в себе возможные разрезы таблицы фактов, например, дата, менеджеры, типы продукции;
- отношений между таблицами.

Более детальную информацию о возможных схема OLAP БД можно найти в [1, 2]

Этап 2. Загрузка и преобразование данных из OLTP БД в OLAP. Основное различие OLAP и OLTP – это степень нормализации данных в БД. Как правило, OLTP БД имеют высокую степень нормализации, а OLAP глубоко денормализованы для повышения быстродействия операций выборки. Ввиду этого различия, чаще всего необходимы преобразования данных из OLTP в OLAP.

В реальных системах вводится, как минимум, ежедневное обновление аналитических БД. Для автоматизации актуализации данных в OLAP БД используется комплекс программ под общим названием ETL, он в буквальном смысле означает:

- извлечь, т.е. получить данные из указанных таблиц OLTP БД;
- преобразовать, т.е. преобразовать схемы OLTP БД в схему OLAP БД;
- загрузить, т.е. непосредственно загрузить данные в OLAP БД.

Этап 3. Построение куба, метрик и измерений. Над реляционной схемой OLAP БД строятся куб(ы).

Куб – концепт, состоящий из множества метрик (measures), таблиц фактов и измерений (dimensions) (таблиц измерений).

Ключевыми составляющими являются метрики, состоящие из данных содержащихся в таблице фактов; измерения – данные содержащиеся в таблице измерений и связанные с мерами, использующие связи один ко многим. Отметим, что измерения могут выстраиваться в иерархии.

Этот этап необходим для того, чтобы аналитик мог выбрать нужные ему разрезы и проводить анализ без помощи технического специалиста. На основе описанной схемы проходит проверка целостности, оптимизация, выборки.

Этап 4. Построение отображения данных пользователю. Этот этап отвечает за создание удобного и функционального интерфейса для пользователя. Интерфейс должен быть в состоянии:

- подключиться к соответствующей службе OLAP;
- иметь понятный и удобный интерфейс;
- обеспечить пользователю выбирать измерения по рядкам и столбцам, а также метрики, которые ему нужны для анализа;
- возможность экспортировать данные в нужные пользователю форматы;
- опционально, фильтровать данные.

3. Практическая реализация OLAP-решения

Для конкретной реализации выбран BI-инструментарий MS SQL Server 2005, как один из удобных и гибких для построения решений такого типа. Наличие утилит и инструментов разработчика позволяет в полной степени покрыть процесс построения аналитического решения, концепция которого раскрыта в разделе 2.

Краткий обзор утилит и инструментов:

1. Для создания реляционной схемы OLAP БД можно использовать любой инструмент, который генерирует на выходе SQL DDL. Специально для целей MS SQL Server 2005 использует расширение Visio 2003 (2007).

2. Типы проектов в приложении к Visual Studio 2005, которое идет вместе с поставкой MS SQL Server 2005 (этапы 2, 3, 4):

а) проект службы интеграции (Integration services project). Этот проект отвечает за выгрузку, преобразование и загрузку данных в OLAPo БД (и не только), т.е. это ETL;

б) проект аналитической службы (Analysis services project). В этом проекте создается куб с измерениями и мерами. Концепция построения состоит в выборе источника данных, определения отношений между таблицами, создание дополнительных вычисляемых полей, создания базового куба непосредственно, создание/редактирование/удаления измерений/мер и множественного другого функционала, который не является стандартным, и который мы не будем трогать в данном приложении;

в) проект службы отчетов (Report server project (Report server project wizard)). В этом проекте создаются базовые виды отчетов для отображения конечному пользователю, включая pivot-таблицы, отчеты, графики, диаграммы и т.п.

На этап 1 создаются схемы OLAP БД.

Схема БД показана на рис. 1.

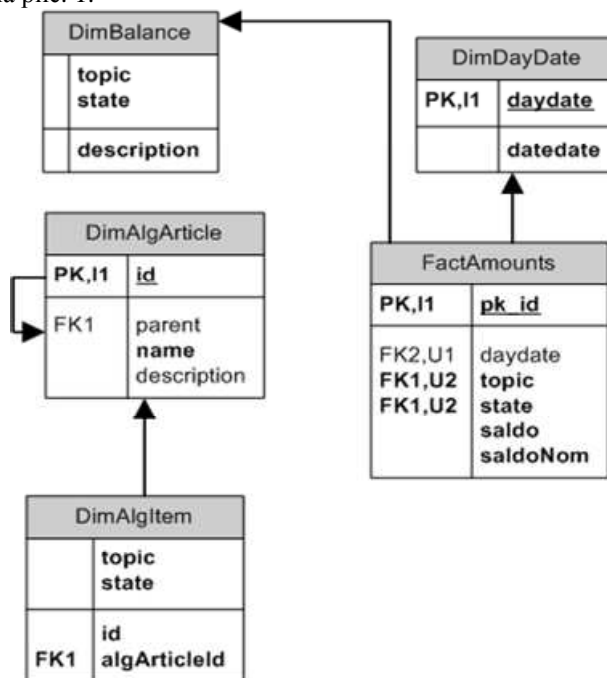


Рис. 1. Схема БД

Где DimBalance – таблица, содержащая стандартную разбивку НБУ баланса; DimDayDate – таблица, содержащая даты. Выделена отдельная таблица для возможностей поделить даты на месяцы, годы, дни недели и т.п.; DimAlgArticle – иерархия групп счетов, может быть уникальной для каждого банка, департамента и т.п.; DimAlgItem – балансовые счета групп; FactAmounts – таблица, содержащая факты, как остатки на счетах.

Этап 2 для загрузки и преобразования данных из OLTP БД в OLAP. Этот этап опущен из соображений высокой сложности и необходимости привязки к конкретным ОДБ. 2.3.

Этап 3 для построение необходимого куба(кубов), метрик и измерений .

Общая схема нашего куба, показана на рис. 2.

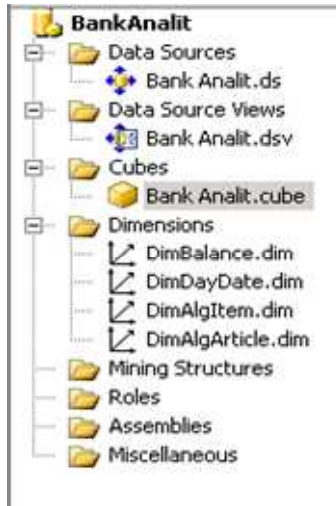


Рис. 2. Схема OLAP БД

Где, источник данных является БД (также возможны множественные источники данных). Здесь указана БД, именуемая Bank Analit.ds; отображение источника данных(Data Sources Views), является схема БД указанная на этапе, описанном в подразделе 3.1. Здесь именуемая, как Bank Analit.dsv; куб, именуемый Bank Analit.cube. Куб содержащий группу мер, рис 3:

- а) сальдо – остаток на балансовом счету в гривне;
- б) сальдо в номинале – остаток на балансовом счете в валюте счета;
- в)

измерения, соответствуют назначению таблиц, см. подраздел 3.1: DimBalance.dim; DimDayDate.dim; DimAlgItem.dim; DimAlgArticle.dim.

Этап 4. Построение отображения данных пользователя. Для отображения пользователю конечных результатов будут использованы два интерфейса MS Excel 2003, рис. 4; MS SQL Server 2005 Reports Services, браузерный рис. 5.

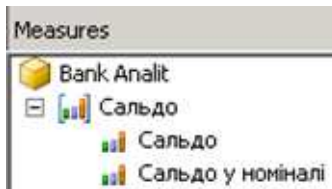


Рис. 3. Меры OLAP куба

Отметим, что MS Excel 2003 имеет встроенные средства для работы с pivot-таблицами, которые предоставляют конечному пользователю удобный интерфейс для работы с данными и позволяют строить любые разрезы без помощи технического специалиста.

Для построения браузерных отчетов, приложения VSFS содержат удобные и понятные визарды для быстрого построения и установки отчетов на сервер отчетов.

4	Опис	2006-06-30 00:00:00	2007-01-03 00:00:00	Общий итог
5	Інші необоротні матеріальні активи	-10940125	-2318674	-13258799
6	Інша дебіторська заборгованість за операціями з клієнтами банку	-647515,8125	-265519,25	-913035,063
7	Інша дебіторська заборгованість за розрахунками з працівниками банку та іншими о	-1897,050049		-1897,05005
8	Інша сумнівна дебіторська заборгованість за операціями з клієнтами банку	-1223,199951		-1223,19995
9	Інші акції та цінні папери з нефіксованим прибутком у портфелі банку на продаж	-1645734,25	-1646019,5	-3291753,75
10	Інші нараховані доходи	-242854,25	-1219216,25	-1462070,5
11	Аванси працівникам банку на витрати з відрядження	-11650		-11650
12	Аванси працівникам банку на господарські витрати	-2450		-2450
13	Акції та інші цінні папери з нефіксованим прибутком що випущені небанківськими ф	-1846690	-1838580	-3685270
14	Банківські метали в банку	-5307507	-3107456,25	-8414963
15	Банківські метали в дорозі	-677935,5		-677935,5
16	Банківські метали у відділенні банку	-1228755,5	-300764,4688	-1529520
17	Банкноти та монети в банкоматах	-31512636	-21960680	-53473316
18	Банкноти та монети в дорозі	-2228338,25		-2228338,25
19	Банкноти та монети в касі банку	-59345948	-62097104	-121443056
20	Банкноти та монети в касі відділень банку	-27450188	-5942059	-33392248
21	Банкноти та монети в обмінних пунктах	-1751643,125	-841159	-2592802
22	Банкноти та монети інкасовані до перерахування	0		0
23	Боргові цінні папери випущені банками в портфелі банку до погашення	-48195,48047	-49565,19141	-97760,6719
24	Боргові цінні папери випущені небанківськими фінансовими установами у портфелі б	0	-64970000	-64970000
25	Боргові цінні папери нефінансових підприємств у портфелі банку на продаж	-258648352	-299652512	-558300864
26	Боргові цінні папери нефінансових підприємств у портфелі банку до погашення	-349280,3438	-25023,67969	-374304,031
27	Боргові цінні папери нефінансових підприємств у торговому портфелі банку	-27000000	-28735500	-55735500
28	Вимоги що придбані за операціями факторингу із суб'єктами господарювання	-213155,4531	-307933,9063	-521089,375
29	Витрати майбутніх періодів	-699273,4375	-512257,6875	-1211531,13
30	Власні акції (частки паї) що викуплені в акціонерів (учасників)	-2000		-2000
31	Дебіторська заборгованість за операціями з готівкою	-27973,5		-27973,5
32	Дебіторська заборгованість за операціями з готівкою між філіями банку	-223697,4063		-223697,406
33	Дебіторська заборгованість за послугами	-1306096,25	-466172,8438	-1772269,13
34	Дебетові суми до з'ясування	-34269,60156	-196625,7344	-230895,344

Рис. 4. Пример отображения данных пользователю в MS Excel 2003

	2006-06-30 00:00:00	2007-01-03 00:00:00
Інші необоротні матеріальні активи	-10940125.00	-2318674.00
Інша дебіторська заборгованість за операціями з клієнтами банку	-647515.81	-265519.25
Інша дебіторська заборгованість за розрахунками з працівниками банку та іншими о	-1897.05	
Інша сумнівна дебіторська заборгованість за операціями з клієнтами банку	-1223.20	
Інші акції та цінні папери з нефіксованим прибутком у портфелі банку на продаж	-1645734.25	-1646019.50
Інші нараховані доходи	-242854.25	-1219216.25

Рис. 5. Пример отображения данных пользователя в браузере

4. Прогнозирование состояния банка

Результаты построения, вышеизложенные использованы в [4] для построения прогнозов данных, основанных на макроекономических инвариантах выбранных показателей банка. Отметим, что построения прогноза и поиск инвариантов был бы чрезвычайно сложен или невозможен, без построения аналитических разрезов баланса банка.

Основная идея построения прогноза такая:

1. Выбор показателя функционирования банка, состоящего из определенного аналитиком набора балансовых счетов.
2. Построение динамики этого показателя.
3. Поиск среди макроекономических показателей наиболее «подходящего», т.е. инварианта к выбранному показателю, используя корреляцию и автокорреляцию.
4. Построение или выбор прогноза макроекономического показателя.
5. Построение прогноза показателя банка.

В работе [4] продемонстрированы, практические результаты этого подхода. Некоторые из них представлены далее.

Для проверки практичности методологии проведен эксперимент на примере крупного Украинского банка. Как генеральные совокупности выбраны «средства на текущих счетах в долларах США» и «средства на текущих и депозитных счетах в гривне», а инвариантами служили такие макропоказатели НБУ, как «курс доллара США» и «Параметр M0» соответственно. Генеральная выборка состояла из усредненных данных поквартально с 2001 года по 2006. «Средства на текущих и депозитных счетах в гривне» и «Параметр M0» скорректированы на инфляцию за период анализа.

Корреляция составила (рис. 6 и 7):

1. «Средства на текущих счетах в долларах США» и «курс доллара США» – $-0,774$.
2. «Средства на текущих и депозитных счетах в гривне» и «Параметр M0» – $0,947$.

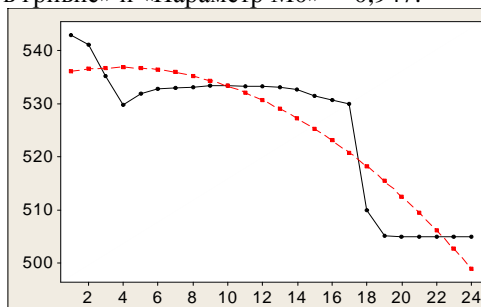
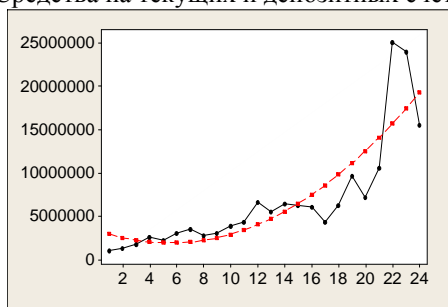


Рис. 6. Анализ тренда для «Средства на текущих счетах в долларах США» и «курс доллара США»

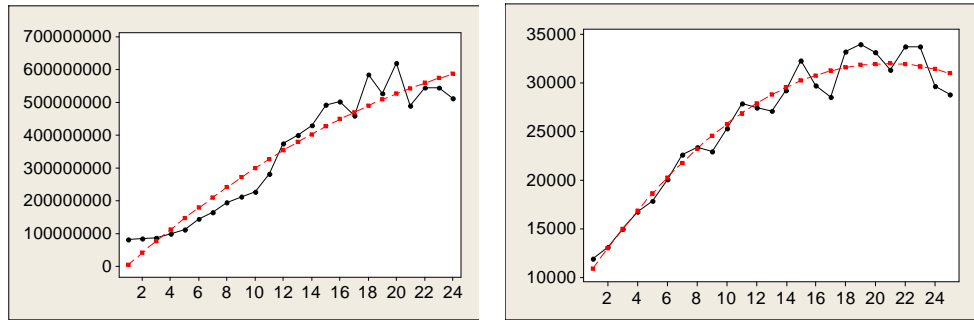


Рис. 7. Анализ тренда для «Средства на текущих и депозитных счетах в гривне» и «Параметр M0»

Исходя из анализа тренда видно, что при отрицательной корреляции имеем зеркально отображенный тренд на графике «курс доллара США». Также тренды генеральных совокупностей и инвариантов повторяют тенденции один другого, что позволяет сделать вывод о правильности выбора инвариантов.

Результаты автокорреляционного анализа, показаны на рис. 8 и 9, показывает соответствие генеральных выборок и инвариантов.

Наличие инвариантов позволяет прогнозировать состояние макропоказателей (или использовать уже готовые прогнозы). Поскольку макропоказатели агрегируют большие объемы данных, таким образом сглаживая аномалии, прогнозы построенные на их базе являются более точными. В таблице приведены результаты прогнозирования для 4-х вышеуказанных выборок.

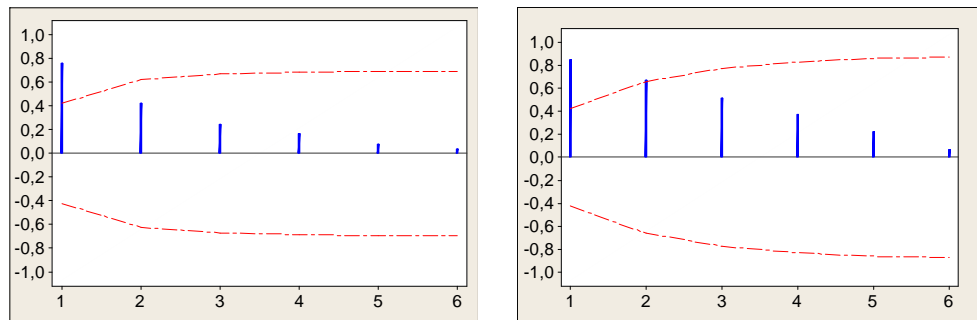


Рис. 8. Автокорреляционный анализ для «Средства на текущих счетах в долларах США» и «курс доллара США»

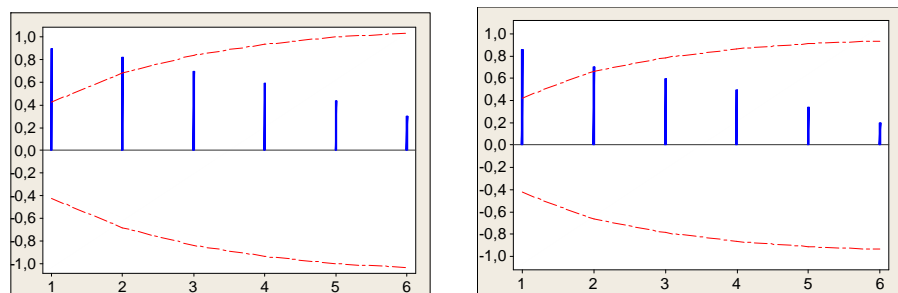


Рис. 9. Автокорреляционный анализ для «Средства на текущих и депозитных счетах в гривне» и «Параметр M0»

Таблица 2. Прогнозные значения

Год	Квартал	Курс доллара США	Средства на текущих счетах в долларах США	Параметр M0	Средства на текущих и депозитных счетах в гривне
2007	I	505	10593418	28813,82289	398274185,2
Прогнозное		504,05	13875012	29592	443301392
MAPE		0,4501	31,54	6	10,85
MAD		2,3556	2 375 730	1514	32 158 800

Как видно из результатов прогнозирования, MAPE как минимум в 1,8 раза меньше для инвариантов, чем для генеральных совокупностей.

Эксперимент поставлен на обычном персональном компьютере, что обусловлено отсутствием ресурсоемкости задач. При анализе больших объемов данных можно воспользоваться распараллеливанием задач.

Выводы

Построение OLAP решений на базе MS SQL Server 2005 достаточно хорошо структурировано, практично и востребовано. На практике такие решения создаются в тесном сотрудничестве со специалистами конкретной предметной области. Отметим что, такая деталь, как именование метрик и измерений играет важную роль в совершенствовании продукта и снижает усилия для перехода на новый продукт.

В настоящее время решения на базе Web-технологий еще не совсем привычны, несмотря на их множественные преимущества как для разработчиков, так и для конечных пользователей. В этом случае рекомендуется использовать двойной интерфейс, к примеру, один – в MS Excel 2003 (2007), второй – Web.

Изначально корректно построенная схема OLAP БД экономит время используя VSFSS, так как в этот продукт включено множество визардов и интеллектуальных систем, позволяющих построить 80% OLAP-куба, основываясь на схеме БД.

В докладе освещены актуальные проблемы финансового сектора Украины. Приведено решения одной из таких проблем, на примере задача прогнозирования данных бухгалтерского баланса. Показана практичность используемых методов и инструментов.

1. http://www.appdev.com/prodfamily.asp?catalog_name=AppDevCatalog&category_name=SQASProduct&MSCSProfile=E643FF728F825C96AE54FD4E74A9047A9651E2D9FB6FE2649138B7244C446CBF4CCBE3E9A63026ECA5BC42F66CDA19BFBD88BDB7F89F403B08324B2A9269A89A5B048B0B634BC2522F873F74224AB1A057E4B0E37BC4C1D4814FAA5B160F815013A80BD0992FD38D0940319C4D4D1E24F936EA07B6611D9BCF8D09956F4820E0CA6D0A34C85323B5
2. *Thomsen Erik*. OLAP Solutions Building Multidimensional Information Systems Second Edition.– Toronto, Canada: 2002.
3. <http://www.osp.ru/win2000/2005/08/380209/>
4. *Мелашенко А.О.* Метод инвариантов для прогнозирования // Компьютерная математика. – 2008. – Вып 1. – С. 15–23.