

**СОФИЙСКИ УНИВЕРСИТЕТ “СВ.КЛИМЕНТ ОХРИДСКИ”
ФАКУЛТЕТ ПО МАТЕМАТИКА И ИНФОРМАТИКА**

МАГИСТЪРСКА ПРОГРАМА «ЕЛЕКТРОНЕН БИЗНЕС»

ДИПЛОМНА РАБОТА

ТЕМА:

**Интегриране на данни за нуждите на Business Intelligence в
търговска верига Бриколаж**

ДИПЛОМАНТ:

Диян Славчев Йорданов

Факултетен номер: М 21710

НАУЧЕН РЪКОВОДИТЕЛ:

Доц. Петко Русков

**СОФИЯ
Юли, 2007**

Съдържание

Увод	3
I. Теоретична постановка:	5
Характеристики на интеграцията на данни	8
Техники за интеграция на данни	9
Консолидацията на данни	9
Федерацията на данни	12
Разпространението на данни	13
Улавяне на промени в данните	14
Качество на данните	16
Технологии за интеграция на данни	16
Extract, Transform and Load	17
Enterprise Information Integration	18
Enterprise Application Integration.....	22
Enterprise Data Replication.....	23
Master Data Management и Customer Data Integration	24
II. Постановка на интеграционния казус в Бриколаж	26
Описание на средата	27
Необходими BI приложения	32
Главна контекстна област	34
Интегрирана информация.....	34
Интерактивни справки	36
III. Модел на интеграция на данните в Бриколаж.....	41
Основополагаща концепция.....	41
Модел на данните	43
Дименсия на магазините	45
Дименсия на доставчиците	47
Дименсия на промоциите.....	48
Дименсия на продуктите	49
Дименсия на касовите бележки	50
Факт таблица на продажбите.....	51
Заклучение	54
Използвана литература.....	56

Увод

В съвременната среда на свободно движение на капитали, на разширяване на пазарите и все по-строги изисквания към бизнеса актуалният въпрос пред всяка организация е как да се приспособи бързо и адекватно към предизвикателствата на средата. Независимо от сферата на дейност, всички организации в различна степен се намират под силния натиск на настъпващите промени, които изискват непрекъснато адаптиране на дейността и усъвършенстване според условията на заобикалящия ги свят.

Българският бизнес все по-често се изправя пред предизвикателства, които до скоро са били чужди за бизнес средата в България. Разглеждайки в частност пазара на търговията на дребно със стоки от типа „Направи си сам“, за последните 5 години се наблюдава засиленото присъствие на търговски вериги, които си оспорват лидерската позиция на тази разрастваща се пазарна ниша. Бриколаж е първата верига от този тип, стъпила на българския пазар и като такава първа се изправя пред предизвикателството да овладее процесите на разрастване във всички функционални отдели. Освен чисто организационните мерки, които се вземат в подобни случаи, ключова е и ролята на информационните системи, които поддържат работните процеси в бизнес организацията.

Но освен да подпомогнат бурното разрастване на операциите в търговските процеси, информационните технологии се разпознават и като източник на конкурентно предимство от ръководството на Бриколаж. Business Intelligence като модерен подход за задълбочен поглед на

случващото се в организацията и за проактивно насочване на процесите в желаната от бизнес лидерите в организацията посока е посочен за ключов фактор за успеха на веригата на българския пазар.

Първата стъпка в един Business Intelligence проект е да се подготвят данните, върху които ще се изградят бъдещите аналитични приложения. Тази дипломна работа описва процеса, през който премина проекта за внедряване на Business Intelligence в търговска верига Бриколаж и в частност големия интеграционен казус, който е решен за създаването на стабилна основа за цялостното аналитично решение.

В първата част от дипломната работа се описват основните методи, техники и технологии, наложени в областта на интеграцията на данни. Поставят се основните въпроси, на които трябва да се отговори преди да се предприеме всеки проект за интегриране на данни в една разнородна информационна среда.

Втората част представя обекта на интеграционния проект – конкретна търговска верига в лицето на Бриколаж. Разглежда се информационната инфраструктура и нейното развитие във времето. Представят се и изискванията на бизнес потребителите – техните ясни заявки за анализи в различни търговски области.

Третата част описва реализирания интеграционен процес, етапите и трансформациите, през които преминават данните от оперативните системи преди да достигнат до своето място в единния корпоративен data warehouse на Бриколаж.

I. Теоретична постановка:

Дисциплината **интеграция на данните** се състои от практики, техники за дизайн на архитектурата и инструменти за постигане на консистентен достъп и предоставяне на данни в пълната ширина от приложни области и видове структури от данни в една организация за да се посрещнат изискванията при използване на данни от всички приложения и бизнес процеси. ^[1]

Интеграцията на ниво цялостна организационна структура се осъществява на следните три нива – *данни, приложения и бизнес процеси* (фиг. 1).



Фиг. 1

Интеграцията на данни предоставя обединен изглед на **бизнес данни**, които иначе са разпръснати по цялата организация. Този обединен изглед може да се постигне с използването на различни методи и технологии. Той може да бъде **физически изглед** на данни, извлечени от множество самостоятелни източници на данни и консолидирани в интегрирано хранилище на данни (*data store*), каквито са *data warehouse* и *operational data store*. Той може да бъде

и **виртуално обединен (federated) изглед** на иначе несравними данни, динамично сглобяван в момента на достъп. Трета възможност е да се представи изглед към данни, интегрирани посредством разпространяване на данни от една база данни в друга – като обединяване на клиентски данни от CRM база данни в ERP база данни например.

Интеграцията на приложения предоставя обединен изглед на **бизнес приложения**, които присъстват в или извън рамките на една организация. Този обединен изглед се постига с управление и координация на потока от събития (транзакции, съобщения или данни) между приложенията.

Интеграцията на бизнес процеси предоставя единен изглед към **бизнес процесите** в организацията. Инструментите за дизайн на бизнес процеси дават възможност на специалистите да анализират, моделират и симулират бизнес процеси и изгаждащите ги дейности. BPM (business process management) инструменти имплементират и управляват тези процеси, опирайки се на технологиите в областта на интеграцията на приложения. Основният принос от интеграцията на бизнес процеси е в разделянето на дизайна на бизнес процесите от тяхното физическо управление и промени в софтуерните приложения.

Трите нива на интеграция не функционират изолирано едно от друго. В една цялостно интегрирана бизнес среда протича взаимодействие между отделните нива на интеграция. В среда, базирана на *data warehouse* инструментите за интеграция на данни работят с приложенията за интеграция на бизнес софтуер за улавяне на събития, протичащи по време на бизнес операциите,

трансформиране и зареждане на транзакциите в *operational data store (ODS)* или *data warehouse*.

Едно от предизвикателствата при изграждане на среда, способна да поддържа BI системите в организацията е разнообразието от сорс системи. Има три главни типа източници на данни – оперативни, лични и външни.^[2]

Оперативни OLTP и batch системи предоставят вътрешни оперативни данни за наличните оперативни области в организацията. Такива може да бъдат:

- Финанси
- Логистика
- Продажби
- Заявки
- Персонал
- Billing
- Нови разработки и развитие

Личните данни обикновено идват от персоналните компютри на бизнес анализатори, статистици, мениджъри и други създаващи знание служители. Примери за този тип данни са:

- Електронни таблици с продуктова информация
- Бази данни с потенциални клиенти

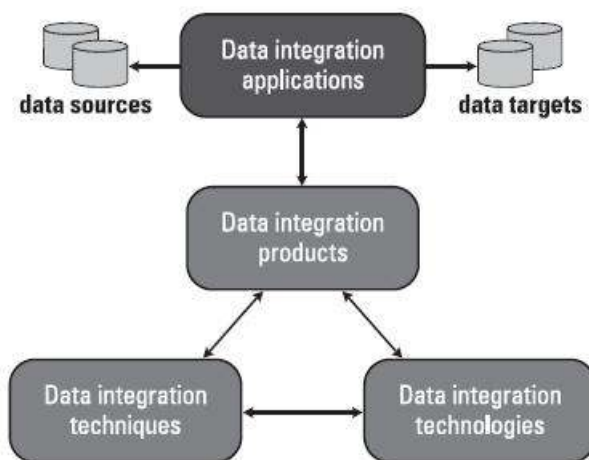
Организациите често купуват външни данни от компании, които са специализирани в събирането на

специфични за индустрията данни. Обикновено информацията от външни източници се обединява в следните категории:

- Маркетинг и продажби
- Кредитни данни
- Данни за конкуренцията
- Данни за индустрията
- Икономически данни
- Демографски данни
- Метеорологична информация

Характеристики на интеграцията на данни

Интеграцията на данни се базира на рамка, изградена от **приложения, техники, технологии и инструменти**, които представят обединен и последователен изглед на бизнес данните в цялата организация (фиг. 2). [3]



Фиг. 2 - Компоненти на интеграционното решение

- **Приложенията** са специално разработени или готови решения, които използват възможностите на един или повече инструмента за интегриране на данни.
- **Инструментите** за готови комерсиални решения, които поддържат една или повече технологии за интегриране на данни.
- **Технологиите** имплементират една или повече техники за интеграция.
- **Техниките** са технологично независими концепции за осъществяване на интеграцията.

Техники за интеграция на данни

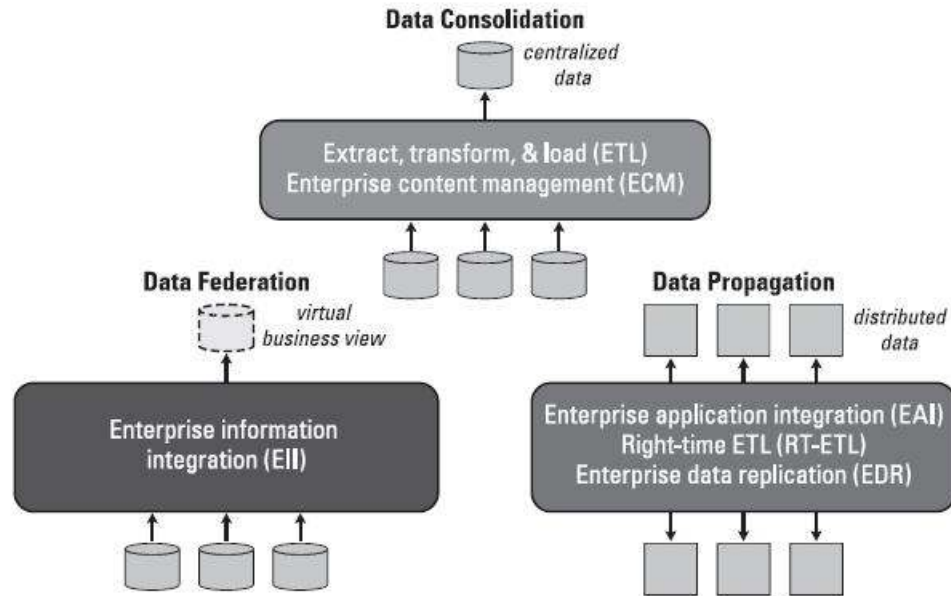
Има три основни техники за интеграция на данни: консолидация, федерация и пропагиране (фиг. 3).

Консолидацията на данни (Data Consolidation) улавя данни от множество системи и ги интегрира в един непроменящ се data store. Той от своя страна може да се използва за репортинг и анализ като при data warehouse или да служи като източник на данни при нови приложения като operational data store.

Когато се използва техниката консолидация на данни обикновено съществува известно забавяне между времето, в което се осъвременяват данните в сорс системите и времето, в което се осъвременяват данните в единния data store. В зависимост от нуждите на бизнеса това забавяне може да варира от няколко секунди до много дни. Терминът *near-real-time* се използва често за да опише забавяне от

няколко секунди, минути или часове. Истински *real time* данни не могат да се извлекат с тази техника.

Единните data stores, съдържащи данни с забавяне над един ден се изграждат най-често с *batch* приложения, които **изтеглят** данни от сорс системите на зададени интервали.



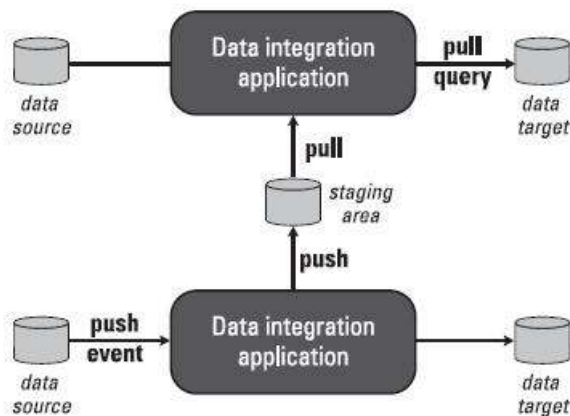
Фиг. 3 - Основни техники за интеграция на данни

Този подход използва заявки за данни, които създават периодични „снимки“ на сорс данните. Въпреки, че тези заявки ще извлекат текущата версия на данните те няма да отразяват промените, настъпили от оследната снимка – сорс данните може да са се променили няколко пъти между двете снимки. Изтеглящият подход е базиран на заявки (on demand).

Единни data stores с малък период на забавяне се обновяват от on-line приложения за интеграция, които динамично улавят и **изпращат** данни при всяка промяна в сорс системата. По този начин се улавят и промените в

данните, обект на консолидиране. Изпращащият подход е задвижван от събития (event driven).

Изтеглящия и изпращащия подход може да се използват едновременно - on-line приложения за интеграция събира данни за промените в междинна база данни, която от своя страна се достъпва на заложи интервали от batch изтеглящо приложение (фиг. 4).



Фиг. 4 – Изпращане и изтегляне на данни

Бизнес приложения, които използват консолидиран data store могат да изпълняват заявки, справки и анализи върху данните в него. Те обаче обикновено не могат да връщат обратно информация към него поради проблеми със синхронизацията към сорс системите. Изключение от това правило са специфични бизнес процеси, които позволяват „засичане“ на данни в даден момент – края на деня, седмицата месеца. Такива процеси са планирането, бюджетирането и прогнозирането, дефиниране на ценови модели и др.

Предимството на техниката консолидация на данни е в това, че тя позволява големи обеми от данни да бъдат

трансформирани по време на техния преход от сорс системата към единния data store. Недостатъкът и е в значителните изчислителни ресурси, необходими за поддържане на процесите по консолидация, както и големия обем дисково пространство, необходим за поддържане на самия data store.

Федерацията на данни позволява общ виртуален изглед към един или повече файлове с данни. Когато бизнес приложение използва заявка към този виртуален източник, инструментът за федерация на данни извлича данните от необходимия източник, интегрира ги във вида, необходим за отговор на заявката и връща резултата към бизнес приложението. По дефиниция федерацията на данни винаги **изтегля** данни от сорс системите на база конкретна заявка. Всички необходими трансформации на данните се извършват в момента на изтегляне то им от сорс файловете. Enterprise information integration (EII) е пример за технология, която използва подходът за федерация на данните.

Един от ключовите елементи на системите, използващи федерация са метаданните, използвани за достъп до сорс данните. В някои случаи метаданните се състоят единствено от виртуална дефиниция на връзките между сорс данните и интегрираните данни (mapping). В по развити решения метаданните може да съдържат и детайлна информация за обема на сорс данните, методът за тяхното достъпване и др. Тази разширена информация може да подпомогне и подобри достъпът на федериращата система до сорс системите.

Някои решения за федерация на данни може да предоставят допълнителни бизнес метаданни, които описват семантичната връзка между данните в сорс системите. Пример за подобен подход са данните за клиенти. Метаданните може да съдържат общ идентификатор на клиента, който да е мапнат към различните идентификатори в различните сорс системи.

Основното предимство на подходът за федерация на данни е, че предоставя достъп до текущи данни и премахва необходимостта от консолидиране на няколко сорс източника в единен data store. От друга страна федерирането не е подходящо за използване в случаи, когато се извличат големи обеми данни или за приложения, където са налице значителни проблеми с качеството на данните в сорс системите. Трябва да се има в предвид и допълнителното натоварване на транзакционните системи, водещо до намаляване на тяхната производителност.

Техниката за федериране на данни може да се прилага в случаи, когато разходите за прилагане на техники за консолидиране на данни са по-големи от ползите, които бизнесът ще извлече от прилагането им. Оперативният репортинг е пример, попадащ в тази категория. Федерацията на данни може да е решение и в случаи, когато политики за сигурност, лицензиране или други ограничения не позволяват копиране на сорс данните. Често федерацията се използва като краткосрочно решение при обединение на две компании.

Разпространението на данни (data propagation) е техника, в която едно приложение копира данните от една точка в друга. Такива приложения работят онлайн и

изпращат данните към таргет системите, което означава, че те се задвижват от събития (event driven). Промени в сорс системите модар да се разпространяват синхронно или асинхронно към таргет системите. Синхронното разпространение изисква и сорс и таргет системите да се променят едновременно с една и съща физическа транзакция. Независимо от типа разпространение, този подход гарантира „доставката“ на сорс данните в таргет системата. Тази гаранция е в основата на техниката за разпространение на данни. Повечето технологии за разпространение позволяват двупосочна размяна на данни между сорс и таргет системата. Enterprise application integration (EAI) и enterprise data replication (EDR) са пример за технологии, използващи този подход.

Голямо предимство на разпространението на данни е във възможността техниката да се използва в real-time и near-real-time приложения. Друго предимство е гаранцията за пренос на данните и двупосочния трансфер. Затова и често разпространението се използва за workload balancing, backup, recovery и disaster recovery.

Техниките, използвани от приложенията за интеграция на данни зависят едновременно от бизнес изискванията и технологичните особености в компанията. Често се прилагат смесени подходи, използващи възможностите на различни техники за постигане на поставените от бизнеса цели.

Улавяне на промени в данните

Консолидирането на данни и разпространението на данни създават и поддържат копие на сорс данните.

Предизвикателството пред двете техники е как да се справят с промените в данните, настъпващи в сорс системите. Едно от решенията е да се изгражда наново таргет data store системата на определен интервал от време и така да се поддържа актуалността на данните. Това разбира се е възможно само за малки data stores. За това е необходима функционалност, позволяваща да се улавят промените в сорс данните.

Ако сорс данните имат time stamp, която да показва кога данните са променени за последно, то тя може да се използва за маркиране на данни, които са се променили от момента на последно осъвременяване на данните. Освен в случаите, когато се създава нов запис и/или версия при всяка промяна в данните, приложението улавящо промените ще засече последната от няколкото възможни промени, настъпили от последното осъвременяване.

Ако сорс данните нямат time stamp, тогава сорс системите имат две възможности - или да бъдат променени така, че да генерират time stamp или да се започнат да генерират отделен файл, в който да записват последователно всички промени в данните.

Стандартен подход в релационните бази данни е да се добави database update trigger, който да създава копие на променените данни. Друг източник на информация за настъпили промени в релационната база е recovery log таблицата. Enterprise data replication (EDR) решенията често поддържат точно този подход, базиран на тригери и recovery logs. Тригерите обикновено имат по-негативно влияние върху производителността на системата, защото записът на данните в сорс системата и тригера се

осъщесвява в един и същ момент. Използването на `recovery log` има значително по-слабо влияние, защото е независимо от текущото натоварване на сорс системата.

Качество на данните

Качеството на данните само по себе си е самостоятелна дисциплина, но когато се разглежда въпросът за интеграция на данни не може да се пропусне ефектът, който има качеството на данните върху техниките за интеграция.

Два аспекта на качеството на данните трябва да се имат в предвид при интеграция на данните. Първият е анализа на сорс `data stores` за да се установи тяхното съдържание и качество. `Data profiling` инструментите са от голяма полза в този случай.

Вторият аспект за анализ е подходът за пречистване на данните с лошо качество. Това често се постига чрез вместване на процес за трансформация на данните в цялостния интеграционен `workflow`. Процесът на трансформация включва преструктуриране, пречистване, съгласуване и агрегиране. Най-често проблеми с качеството на данни се появява в етапите на пречистване и съгласуване.

Технологии за интеграция на данни

Съществува широк набор от технологии, които предоставят възможност за прилагане на техниките за интеграция, представени по-горе. Основните са:

- extract, transform and load (ETL)
- enterprise information integration (EII)
- enterprise application integration (EAI)
- enterprise data replication (EDR)
- master data management (MDM)
- customer data integration (CDI)

Extract, Transform and Load

Както подсказва името, ETL технологиите извличат данни от сорс системите, трансформират ги така, че да отговорят на изискванията на бизнеса и ги зареждат в таргет системите. Ролята на сорс и таргет най-често играят бази данни и файлове. ETL технологията поддържа техниката за консолидация при интеграция на данни.

Данните може да се извличат от базиран на график изтеглящ подход или от базиран на събития изпращащ подход. И двата подхода може да прилагат техники за улавяне на промени в данните. Изтеглящият подход обикновено се изпълнява от batch процеси, докато изпращащият подход се изпълнява от разпространяващи промените на данните онлайн процеси.

Трансформацията на данните може да включва реструктуриране и съгласуване, пречистване на съдържанието и агрегиране. Зареждането може да предизвиква пълно зареждане на таргет системата или може да се изпълни чрез осъвременяване на съществуващите записи. Интерфейсите използвани за зареждане се базират на станадарти като ODBC, JDBC, JMS или присъщи на конкретните бази данни приложни интерфейси.

В зората на ETL технологията, тя се е използвала batch задачи изпълнявани на фиксирани интервали за да улавя данни от flat files и релационни бази данни и да ги консолидира в data warehouse, управлявана от система за управление на бази данни(DBMS). Сега към тези инструменти са добавени още функционалности:

- Допълнителни сорс системи - legacy данни, XML файлове, Web logs, EAI sources, Web services, неструктурирани данни
- Допълнителни таргет системи - EAI системи, Web services
- Подобрени възможности за трансформация - user defined изходни формати, data profiling и data quality management, поддръжка за standard programming languages, Web services
- Управление на метаданни, job scheduling и error recovery
- По-добра производителност - parallel processing, load balancing, кеширане
- Поддръжка на техники за федерация на данни

Enterprise Information Integration

EII предоставя виртуалне бизнес изглед към разпокъсани данни. Този изглед може да се използва за заявки и достъп до оперативни транзакционни бизнес данни, до data warehouse и/или неструктурирана информация при поискване. EII поддържа федеративния подход към интегрирането на данни.

Целта на EII е да позволи на приложенията да „виждат“ разпръснати данни все едно са разположени в една общ база данни. По този начин се прикрива сложността на извличането на данни от множество локации, където данните може да се различават по семантика, формат и интерфейс за достъп.

В базовия си вид, EII достъпът до разпокъсани данни включва разбиване на заявката към виртуалния изглед на отделни компоненти и изпращането им към локациите, където данните се намират. След това EII инструментът комбинира извлечената информация и изпраща окончателния резултат към приложението, изпратило заявката.

Продуктите, използващи EII са се развили от две коренно различни технологии – системи за управление на релационни бази данни и XML. Тенденцията е да се предложи поддръжка и на двата интервейса в едно решение – SQL (ODBC и JDBC) и XML (XQuery и XPath).

Системите, които са произлезли от DBMS често се възползват от изследванията в областта на distributed database management systems (DDBMS). Основната цел на този подход е да се постигне прозрачен достъп до разпределени данни, както и пълен read и write достъп. Основен недостатък при DDBMS се отнася до влиянието върху производителността на критични за дейността системи. Той се проявява особено силно когато е включена поддръжката на права за писане към разпределени данни. За да преодолеят проблема повечето EII продукти позволяват read-only достъп до хетерогенни данни.

Важно е да се подчертае, че EII подходът за федерация на данни не може да замени традиционния ETL подход към консолидация, използван при data warehousing. Изцяло федериран data warehouse не е за предпочитане поради потенциално слаба производителност и проблеми с консистентността на данните. За сметка на това EII може да допълни класическата data warehouse среда предлагайки решение за специфични бизнес нужди.

EII е силна технология, способна да разреши специфични проблеми при интеграция на данните, но е важно да се разбира цената, която се плаща при използване на федерирани данни. Проблемът се задълбочава в случаи, когато федерирани заявки достигат оперативна транзакционна система.

Друг поренциален проблем стоящ пред всяко EII решение е как да трансформира данни, идващи от множество различни сорс системи. Същият проблем стои и при дизайна на ETL процес, който да захранва data warehouse. В никакъв случай EII не намалява необходимостта от задълбочено моделиране и анализ. На практика дизайна може дори да е по-сложен поради близката до real-time същност на трансформацията на данни в EII среда.

И все пак EII може да бъде подходяща алтернатива на ETL в следните случаи:

- Достъп при поискване до променливи данни – улавянето и консолидирането на бързо променящи се данни от транзакционна система може да бъде скъпо удоволствие. EII може да се използва за директен достъп до текуща информация, а

федерираната заявка може да се обръща едновременно и към оперативни и към исторически данни от data warehouse.

- Директен достъп и write права към сорс данни
- Когато е трудно да се консолидират данните – при наистина много хетерогенна среда създаването на единно консолидирано хранилище може да е невъзможно.
- Федерирането на данни е по-евтино от тяхното консолидиране
- Когато е забранено копирането на сорс данни – в случай на достъп до чувствителни бизнес данни, данни от друга организация или други условия, свързани със сигурността.

Аргументите в полза на използване на ETL са при работа в среда, където:

- Трансформацията на данните е сложна –
- Нуждите на бизнес потребителите са предвидими в близко бъдеще
- Потребителите се нуждаят от исторически данни и тенденции
- Изискванията към производителността на системата и нейната достъпност са значителни
- Достъпът до данни е read-only

Enterprise Application Integration

Интеграция на приложения системи, при която се позволява те да комуникират и разменят бизнес трансакции, съобщения и данни помежду си на базата на стандартни интерфейси се описва в литературата с термина **enterprise application integration** (EAI). По този начин се позволява приложенията да достъпват данни прозрачно и без знание за тяхното местоположение и формат. EAI обикновено се използва в случаи на обработка на real-time оперативни трансакционни данни. Техниката, поддържана от EAI е тази за разпространение на данни (data propagation).

Посоката, в която се движи дисциплината интеграция на приложения е в посока използване на enterprise service bus (ESB), която да поддържа връзката между legacy системите и съвременните приложения, а също и връзка към уеб услуги за осъществяване на service-oriented architecture (SOA).

От гледна точка на интеграцията на данни EAI може да се използва да транспортира данни между приложенията и да рутира събития в реално време към други приложения за интеграция като ETL процеси например. Достъпът до сорс и таргет системите се осъществява чрез Web services, Microsoft .NET интерфейси, JMS, legacy application интерфейси и адаптери и др.

Дизайна на EAI е насочен към разпространение на малки обеми данни от едно приложение към друго. Разпространението може да бъде синхронно или асинхронно. В случай на асинхронно разпространение, всяка бизнес трансакция може да се разбие на множество физически трансакции. Това се налага, защото възможностите за

трансформация и поддръжка на метаданни в EAI системите са фокусирани към обработка на опростени транзакции, за разлика от ETL системите например.

Въпреки разностранните твърдения, EAI и ETL не са конкуриращи се технологии. В много ситуации те могат да се допълват, EAI да служи като сорс система за ETL, а ETL да служи като услуга към EAI.

Една от основните цели на интеграцията на приложения е да се предостави прозрачен достъп към широкия кръг от системи съществуващи в организацията. Затова връзката EAI-ETL може да предостави на ETL продукта достъп до тези данни. Тази взаимовръзка може да се осъществи посредством Web services или подредени съобщения. По този начин се премахва необходимостта от разработване на адаптери за достъп до данните на всяка една сорс система.

Enterprise Data Replication (EDR)

Изброявайки технологии за интергация на данни трябва да се отбележи и репликацията на данни като значим инструмент в ръцете на практиците. Въпреки че не е толкова видима като ETL, EII и EAI, тя се използва не по-малко в интеграционни проекти. Причината може би се крие във факта, че EDR често присъства в цялостните пакети на доставчиците на решения. Репликацията се използва не само за интеграция на данни, но и за backup и recovery, data mirroring и workload balancing сценарии.

Инструментите, използващи репликация често използват тригери и/или recovery logs за да улавят промени в сорс данните и в следствие да ги разпространяват до една или повече бази данни. В повечето случаи разпространението става асинхронно, но понякога се поддържа и двупосочно синхронно разпространение на данни между множество бази данни. Трансформация в процеса на разпространение е по силите на EDR.

Най-значимата разлика между репликацията на данни и EAI е, че първата прави трансфер на данни между бази данни, докато втората предава съобщения между приложения. Обикновено EDR транспортира значително по-голям обем данни от EAI.

Репликацията на данни понякога се съвместно с ETL инструмент. Например използвайки EDR може да се улавят и транспортират големи обеми данни и промени в данните до staging зона. От там те се извличат на определени интервали от batch ETL процес, който ги трансформира и консолидира в data warehouse.

Master Data Management и Customer Data Integration

Управлението на мастер данни (MDM) предоставя консистентен изглед към референтните данни в една организация, която иначе би могла да е разпръсната в множество отделни системи^[4]. Такива са данни относно клиенти, продукти, служители, някои финансови данни и др. Специализирани в обработката на клиентски мастер данни

продукти често се реферират като Customer Data Integration (CDI) инструменти.

Това, което управлението на мастер данни добавя към технологиите за интеграция на данни е бизнес семантика относно референсираните данни. MDM и CDI data stores може да се представят като core системи за data warehouse приложения.

Жизненият цикъл при управлението на мастер данни следва следните етапи:

- Оценка на данните (Data assessment)
- Хармонизиране (Data harmonization)
- Зареждане на MDM системите
- Създаване на оперативен процес, който се грижи за интегритета на данните
- Установяване на data governance процедури за текуща оценка.

II. Постановка на интеграционния казус в Бриколаж

Дружеството Доверие Брико АД е създадено през 1999 г. с цел да въведе в България търговската марка Mr. Bricolage. Веригата хипермаркети, предлагащи стоки за строителството, ремонта и обзавеждането на дома и градината Mr. Bricolage навлиза в България през 2000 г., когато е построен първият магазин в София. В момента у нас работят 8 магазина от веригата в София, Пловдив, Варна, Благоевград, Бургас, Стара Загора и Плевен, а до края на 2007 г. се планира да бъдат открити още три. Всички те са изградени по проверения в над 25-годишната практика модел на френската верига Mr. Bricolage, обединяваща 450 обекта по целия свят.

През последните години търговските компании и в глобален, и в локален мащаб демонстрират подчертан интерес към обогатяването на своите информационни системи с аналитичен инструментариум, тъй като това е свързано с възможността да развиват техния бизнес в перспективни направления. Тази тенденция е напълно валидна за магазини като Mr. Bricolage, фокусирани върху задачата да отговорят на потребностите на един широк кръг от клиенти.

С разрастване на магазинната мрежа на Mr. Bricolage, ръководството на дружеството предприе инициатива за внедряване на аналитична система, която да консолидира данните от всички магазини и да предостави единен източник на информация, който да подпомогне ръководните кадри на всички нива в процеса на вземане на решения.

Описание на средата

Развитието на дружеството през годините от създаването му през 1999 г. е оставило своя отпечатък върху организационната структура, бизнес процесите и съответните приложни системи. По своята динамика и обхват то представлява едно продължаващо и в момента предизвикателство за топ мениджмънта и ръководителите на функционални отдели на всички управленски нива.

Описанието на условията, в които възниква нуждата от насочване на усилия към въвеждане на Business Intelligence решение в информационната среда на дружеството е важно за ясното разбиране на причините за избор на конкретен подход при интегриране на данните. Резултат от текущата организационна среда и визията за бизнес процесите в дружеството в дългосрочен план е определянето и на аналитичните инструменти, задоволяващи нуждите на бизнес потребителите от информация, подпомагаща вземането на решения.

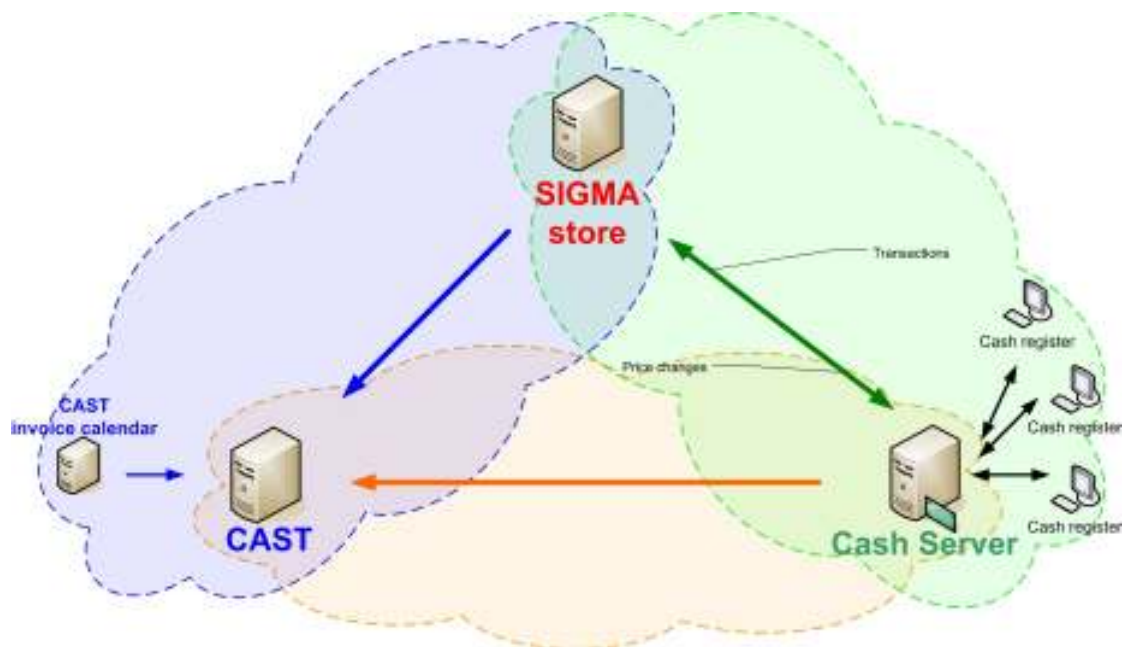
Търговската верига Бриколаж отваря първия си магазин през 2000 г. Информационната среда се състои от следните информационни системи:

- Система за управление на магазин (SIGMA) – собствена разработка на компанията майка Mr. Bricolage SA, специализирана за управление на магазин.
- Касова система – касови апарати с фискална памет
- Счетоводна система

Управлението на магазин с площ от 6000 м² и 40 хиляди активни артикула изисква строга организация и бизнес правила, които са тясно приплетени с ресурсите, предоставяни от информационните системи. Опитът, натрупан с първия магазин на веригата се копира и във втория магазин. Доброто познаване на системите и изградените по поръчка техники и средства за комуникация между тях позволяват гладкото протичане на бизнес процесите и намаляване на общата цена на притежание (total cost of ownership – TCO) за веригата като цяло.

От своя страна всяка информационна система представлява отделен продукт, който се развива според плановете на компаниите-производители. Не всички производители обаче имат ресурса и желанието да развиват продуктите си в посока интегриране на данни – както между две инсталации на собствените си продукти, така и между своя продукт и продуктите на трети доставчици. Решението за изграждането на интеграция е в ръцете на бизнеса, а бизнеса се нуждае от сериозни доказателства за необходимостта за всяка инвестиция.

Работата в информационна среда, разпределена между няколко информационни системи налага създаването на стабилни връзки между тях по начин, осигуряващ оперативността на магазина и служителите в него. Утвърденият в годините модел може да се опише с диаграмата на фиг. 5.



Фиг. 5 - начален модел

Оформят се три информационни потока, които обхващат съобщенията, протичащи между приложенията в Бриколаж.

Синият поток на данни включва информация за доставки на стоки. Данните протичат между системата за управление на магазин SIGMA и счетоводната система. Комуникацията е еднопосочна, като след изпращане на данните SIGMA не търси потвърждение за случващото се в счетоводната система. Осъществява се веднъж дневно по време на нощната обработка.

Оранжевият поток включва данни за издадени фактури на клиенти. Тази информация протича между касовата система и счетоводната система. Комуникацията и в този случай е еднопосочна от касите към счетоводството, като се разчита на счетоводител да провери дали трансфера е протекъл коректно. Осъществява се веднъж дневно.

Зеленият поток включва най-важната за функционирането на магазина информация. Той свързва

системата за управление на магазин и касовата система. Трансфера на данни е двупосочен. От една страна SIGMA изпраща актуалният списък с продукти и цени към касовата система. Този трансфер се изпълнява автоматично по време на нощната обработка или при поискване от оторизиран служител. От друга страна касовата система изпраща данни за осъществените продажби. Тези данни се трансферират автоматично на всеки 15 минути или при поискване от оторизиран служител.

Целта е да се извлече максимума от възможностите на всяка система, като се използват силните страни на софтуерните решения в областите, в които те са се специализирали.

Успешният модел на внедряване на ИТ системи в магазините на Бриколаж се мултиплицира три пъти – в Пловдив, Варна и София. Добрите резултати от локалните решения във всеки магазин се сблъскват с ново предизвикателство – опериране на по-високо ниво, а именно на ниво дружество. Пред мениджмънта на търговска верига Бриколаж се очертават ясно сериозните недостатъци от работа в разпокъсани информационни системи:

- Липса на знание какво се случва в дружеството като цяло – в среда с четири работещи магазина, мениджърите покупки, търговския, финансовия и изпълнителния директор нямат пряк достъп до търговска информация на ниво дружество. Всяка нужда от подобна информация се получава на база данни от ИТ системата на отделните магазини, консолидирани в електронна таблица.

- Невъзможен достъп до данни – ИТ системите включват пакет от фиксирани справки за крайни потребители. Всяка заявка за специфично сечение на данни или нестандартен времеви период се отхвърля или поставя ИТ отдела в несвойствена роля да извлича данни директно от базите данни на приложението.
- Повтаряемост на бизнес операции – много бизнес процеси включват повтаряема във всяка ИТ система операция, която осигурява консистентност на данните във всяка SIGMA.

Необходими BI приложения

Професионалистите в областта на Business Intelligence (BI) споделят едно общо разбиране – независимо от използваните инструменти, технологии, дизайн и архитектура на техните системи, крайният резултат от тяхната работа е **аналитично приложение**.

Аналитичното приложение се състои от множество групи логически интегрирани, интерактивни справки, включително дашбордове и скоркарти, които позволяват на широк кръг поребители да достъпват, анализират и да предприемат действия спрямо интегрирана информация в контекста на бизнес процеси и задачи, които те управляват в дадена област – продажби, услуги, операции и др.^[5]

Четири ключови елемента се открояват в дефиницията на аналитичното приложение:

- Логически интегрирани – аналитичните приложения трябва да имат вградена бизнес логика, която да помага на потребителите да се ориентират в поредността от задачи, свързани с конкретна цел. В транзакционните системи приложението преминава през поредица от екрани, в които се въвежда информацията, необходима за приключване на транзакцията. В аналитичните приложения е необходимо потребителите да се водят през поредица от интерактивни справки или изгледи към дименционални данни, в идеалният случай водейки към момент за предприемане на действие

като вземане на решение, план или заявка за още данни.

- Интерактивни справки, които позволяват на широк кръг потребители да достъпват, анализират и да предприемат действия – докато справките в транзакционните системи представляват екрани с фиксирани колони, аналитичните справки са интерактивни. Това означава, че потребителите могат да дрилват от по-високо ниво на детайлност към по-ниско за повече информация. А при използването на аналитичен дашборд потребителя може да с един клик върху светещия с червена светлина виртуален светофар да получи детайлна справка за прилежащите данни.
- Интегрирана информация – архитектурно погледнато, аналитичното приложение е парче от корпоративния data warehouse, ориентирано към конкретна бизнес област. Един data warehouse може да поддържа десетки аналитични приложения. Понякога организациите използват аналитичните си приложения в отделни data marts, но този разпределен подход може да доведе до провлеми с консистентността на корпоративно ниво.
- В контекста на бизнес област – аналитичните приложения се дефинират от информационните изисквания на бизнеса – в областта на продажбите, производството, услугите, дистрибуцията и т.н. Тези области често са взаимосвързани в един общ логически модел, защото много от бизнес областите споделят общи

референтни данни – за клиентите, продуктите, географски данни и др.

Мениджърският екип на търговска верига Бриколаж преминава през няколко етапа на анализ относно характера на аналитичните приложения, необходими на дружеството. Следвайки добри практики в областта на Business Intelligence се определя референтна рамка за бъдещия проект.

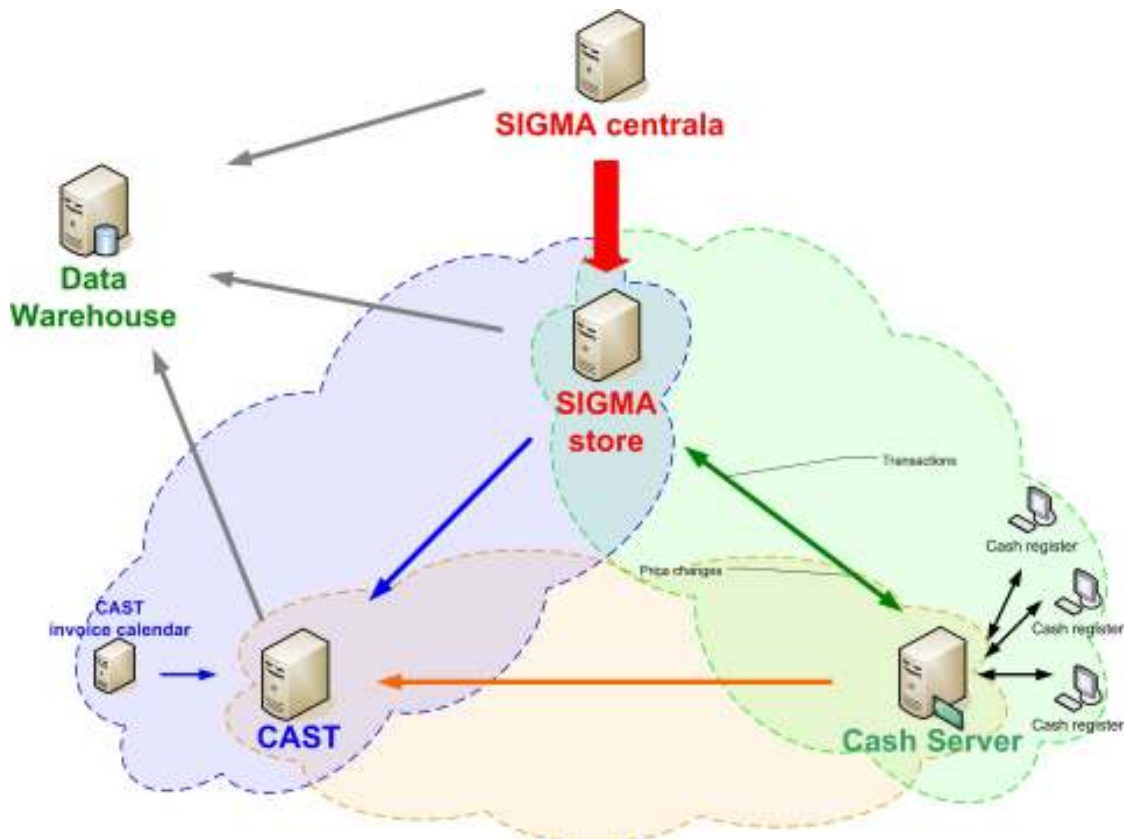
Главна контекстна област на аналитичното приложение да бъдат продуктите, техните продажби и нива на наличност. Основният бизнес на Бриколаж е да продават, за това знанието за това какво се продава, колко, кога, къде, на кой и при какви условия е най-ценния продукт, който едно аналитично приложение може да достави на организацията. Поддържаща, но ключова от гледна точка задоволяване на потребителското търсене, а и на финансовите резултати е ролята на информация, относно нивата на наличност на продуктите. Да имаш точното количество стока в точния момент е ключов фактор за успеха на всяка търговска верига.

Интегрирана информация е в основата на успешното реализиране на всяка инициатива в областта на Business Intelligence. Описаните по-горе подходи за интеграция на данни бяха разгледани детайлно при избора на решение за интеграция на данните в Бриколаж. Основните характеристики на използвания в Бриколаж подход са:

- Техниката за интеграция на данни да бъде тази на **консолидация на данни**.

- Технологията за интеграция на данни да бъде **Extract, Transform and Load (ETL)** .
- Инструментът за интеграция да бъде **WhareScape RED** - продукт, предоставящ интегрирана среда за разборка, изграждане и управление на data warehouse.

Като се използват на дефинираните вече контекстна област и подход за интегриране на данни може да се представи графично бъдещия модел на информационната среда в дружеството (фиг.6) .



Фиг. 6 - бъдещ модел

Интерактивни справки, които да подпомогнат мениджъри, анализатори и редови служители в магазините във всекидневната им работа и постигането на ефективен и ефикасен работен процес трябва да предоставят функционалност в следните области:

Седмична и месечна справка за оборота

	Metrics	Продажби в лв. с ДДС	Себестойност продажби в лв.	Марж в лв.	Марж в %	Стокова наличност в лв.
Магазин Район						

Седмичните и месечни справки показват обобщена информация за продажби и наличност по дименсии. В практиката на Бриколаж най-често се анализират седмичните и месечни резултати по маразин и продуктов район. Интерактивно може да се дрилва от продуктов район към продуктови семейства и под-семейства, както и да се групират магазини според техните особености. Ключовите измерители на бизнеса, наричани метрики, са продажбите в лв. с ДДС, Себестойността на продажбите, Брутната печалба в стойност и процент (във възприетата от Бриколаж терминология – марж) и нивото на стокова наличност към края на анализирания период.

Дневна справка за оборота

Дата	Магазин	Metrics	Дял на продажбите към дневните общи продажби в %	Продажби в лв. с ДДС	Брутна печалба	Търговски марж в %	Брой касови бележки без сторно	Средна кошница за магазин

Дневните справки дават по-детайлен поглед върху динамиката на търговските процеси във веригата магазини. Проследяването на динамиката изисква интерактивност по отношение определянето на деня от седмицата и съпоставка на резултатите за един и същ ден от седмицата в няколко последователни седмици. Метриците, правещи този анализ пълноценен за служители от търговските отдели, а и от ръководството на магазините са отново продажбите и маржа, както и броя касови бележки за деня (косвен показател за броя клиенти и натовареността на касите) и средната кошница (средния размер на една клиентска покупка).

Анализ продажби по продуктови район, семейство

Под-семейство	Магазин	Текущо налично количество	Продадени количества предходните три месеца	Продадени количества за една година назад
	Metrics			

Ключов за успешния бизнес на всяко дружество е процесът на анализ на продажбите по продуктови групи. На база този анализ се извършва планиране на продажбите за

следващия период. Именно тук е изключително полезно лесното движение по степента на детайлност на справката – нагоре към продуктово семейство и район и надолу към конкретните продукти. Метриките в този анализ са продажбите за предходни три месеца, за последната година и текущото налично количество от конкретната стока.

Анализ продажби по продукт

Metrics	Продажби в лв. с ДДС	Продадени количества	Дял от общите продажби в %	Брутна печалба	Търговски марж в %
Продукти					

При детайлен анализ на продуктовата селекция в магазините на Бриколаж се налага отговарящите за дадена продуктова група да анализират продуктовата номенклатура продукт по продукт. Избирайки конкретен период се проследява представянето на всеки продукт и дела му в общите продажби на съответната група. За анализатора е важно да може да групира продукти по различни характеристики, както и да дрилва към дейлно ниво, разбиващо показателите по магазин. Ключовите измерители в този анализ са продажбите в стойност и брой, както и брутната печалба, донесли на дружеството за избрания период.

Анализ продажби по доставчик

			Metrics	Продадени количества	Продажби в лв. с ДДС	Покупна цена	Продажна цена	Реален коэффициент
Продукти	Текущ статус	Основен доставчик						

Централно място при определяне на продуктовата селекция има доставчикът на продуктите и договорените с него условия за доставка на стоки. Част от договаряните условия се отнасят за покупната цена за всеки продукт в зависимост от търгуваните обеми за зададен период. Анализът на продажбите по доставчик трябва да позволи.

Анализ продажби по промоция

		Metrics	Продадени количества	Продажби в лв. с ДДС	Търговски марж в %	Налично количество
Промоция	АНПФ					

В силно конкурентна среда всяко търговско дружество търси различни способы за привличане на повече клиенти в своите обекти. Промоциите са силен и неотменен инструмент в тази борба за клиенти. Но дали промоциите успяват да постигнат своите цели е въпрос, чиито отговор се открива само с достатъчно детайлна информация, консолидирана от всички ключови източници на данни и гъвкав интерактивен инструмент за анализ.

Всички тези области на анализ на продажбите са взаимосвързани. Всички те се описват едно ключово събитие в търговският процес на Бриколаж – продажбата. За това

освен анализът на отделните области, изброени по-горе е необходимо да се осъществи и анализ на взаимовръзките между тези области. Примери за подобен анализ, търсен от Бриколаж чрез реализацията на интеграционен и Business Intelligence проект са:

- Анализ на промоционалните продажби по доставчик
- Анализ на продажбите за период в зависимост от промоциите по месеци
- Анализ на покупките по клиент и промоция или клиент и доставчик (марка продукти)
- Анализ на географските различия в потреблението на продуктив район и др.

III. Модел на интеграция на данните в Бриколаж

Основополагаща концепция

Един от най-важните активи във всяка организация е информацията. Този актив е почти винаги държан в две форми – текущ запис в оперативните системи и исторически запис в data warehouse.

Служителите, работещи с оперативните системи са тези, които записват поръчки, регистрират потребители, приемат стока. Почти винаги те се фокусират върху един запис към даден момент.

От другата страна са служителите, клиенти на data warehouse, които наблюдават какво се случва в организацията. Те броят новите поръчки и сравняват броят им с броят от миналата седмица, питат защо броят на регистрирани потребители пада, сравняват реално получената стока с направената поръчка.

Ясно е, че изискванията, клиентите, структурата и ритъма на работа на един data warehouse дълбоко се различават от тези на оперативната система. Ако трябва да изведем изискванията към един data warehouse те могат да се опишат в следните няколко точки:

- Да направи информацията на организацията лесно достъпна –
- Да представи информацията на организацията в консистентен вид –
- Да бъде адаптируема и гъвкава към промяна –
- Да осигурява сигурността на информацията

- Да служи като основа за подобрения процес за вземане на решения
- Да бъде възприета от бизнес потребителите като своя територия –

При изграждането на data warehouse водеща роля има концепцията за дименсионално моделиране. **Дименсионалното моделиране** е методология за логическо моделиране на данни за нуждите на тяхното ефективно и лесно използване, което се базира на набор от основни измеруеми събития.^[7] В среда на релационна система за управление на бази данни се създава една факт таблица, съдържаща по един запис за всяко дискретно събитие. Тази факт таблица се „обгражда“ от множество дименсионни таблици, описващи ясно какво още се знае за контекста на измереното събитие. Поради характерната структура на дименсионалния модел, често се използва терминът **стар схема** (star scheme) (фиг. 7).

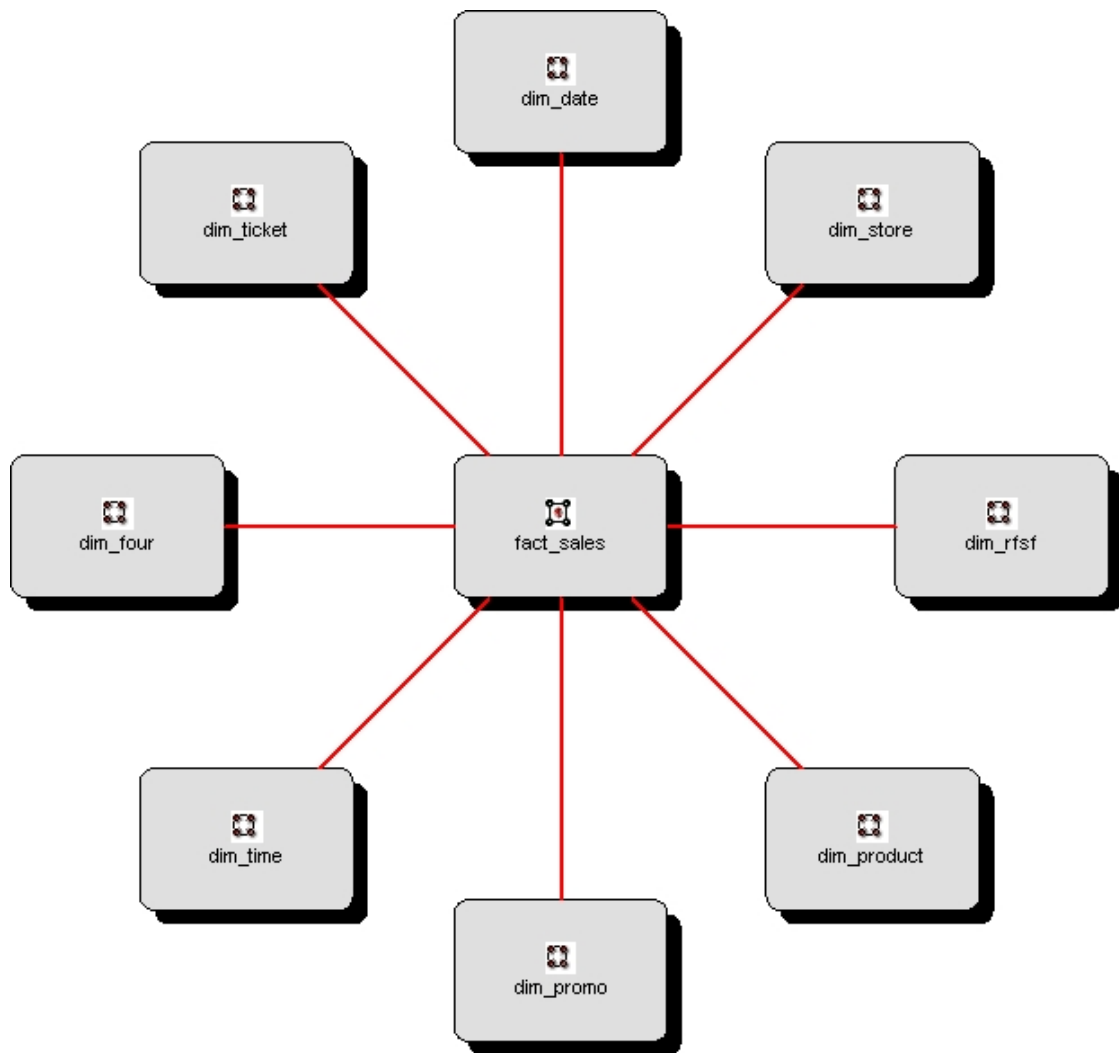


Фиг. 7

Модел на данните

Описаните в част втора изисквания на бизнес потребителите в търговска верига Бриколаж очертаха насоката за работа при дефиниране на дименционалния модел за бъдещия data warehouse (фиг. 8). В него намират място следните дименционални таблици

- Dim_date - стандартна времева дименсия на календарните дни, разширена с допълнителен фискален календар, определяне на национални празници и др.
- Dim_store - дименсия за магазините във веригата
- Dim_rfsf - дименсия за продуктовете групи
- Dim_product - дименсия за продуктовата номенклатура
- Dim_promo - дименсия за промоциите
- Dim_four - дименсия за доставчиците на стоки
- Dim_ticket - дименсия за обобщена информация за всяка продажба (касова бележка)
- Dim_time - стандартна дименсия за часовете и минутите от деня



Фиг. 8

Изграждането на факт таблицата и всяка от дименсиите е основния консолидационен казус, стоящ пред дружеството. Консолидирането на данните до степен, която да позволи пълноценното им използване и осигуряване изпълнението на изискванията към един добър data warehouse се нуждае от задълбочено познаване на оперативните системи от една страна и на ясно формулирани бизне нужди от друга.

Дименсия на магазините



Дименсията на магазините служи за определяне на мястото, където се случват измерваните събития. Събирането и поддръжката на данните за магазините не изисква сериозни консолидационни усилия, тъй като в SIGMA системата, играеща роля на централна се поддържа единен списък с магазините. Поради малкия размер на тази номенклатура и липсата на промяна във вече създаден запис за магазин това не създава трудности нито за служителите, нито за data warehouse. Създадена е процедура, която следи за изменение и допълнение на номенклатурата в централната SIGMA (фиг. X). Тя се състои в ежедневно експортиране на пълната номенклатура от SIGMA във файла **store.d**, който се зарежда в **load_store** със следната процедура (публикувана със съкращения):

```
@echo off
SET LOAD_FILE="\\centra\oracle\bricodw\data\store.d"
echo load of %LOAD_TABLE% > %FILEAUD%
SET /A RESULT_CODE=1
echo loading %FILE_NAME% >> %FILEAUD%
echo load data > %FILECTL%
echo infile %FILE_NAME% >> %FILECTL%
echo badfile '%FILEBAD%' >> %FILECTL%
echo into table %LOAD_TABLE% >> %FILECTL%
echo fields terminated by "  " >> %FILECTL%
echo optionally enclosed by '^"' >> %FILECTL%
echo trailing nullcols >> %FILECTL%
echo ( >> %FILECTL%
echo store_code,>> %FILECTL%
echo store_name,>> %FILECTL%
echo store_short_name>> %FILECTL%
echo ) >> %FILECTL%
exit
```

След зареждане на данните в load таблицата следва въведет на димензионната таблица:

```
CURSOR c_Load IS
```

```

SELECT
    load_store.store_code store_code
  , load_store.store_name store_name
  , load_store.store_short_name store_short_name
FROM load_store
MINUS
SELECT DISTINCT
    store_code
  , store_name
  , store_short_name
FROM dim_store;
BEGIN

v_step := 1;
v_DimRec.dss_update_time := sysdate;
v_DimRec.dss_create_time := sysdate;
v_DimRec.store_code := SUBSTR('N/A',1,6);
v_DimRec.store_name := SUBSTR('N/A',1,64);
v_DimRec.store_short_name := SUBSTR('N/A',1,64);
v_step := 30;
FOR v_LoadRec IN c_Load LOOP

    v_DimRec.store_code := v_LoadRec.store_code;
    v_DimRec.store_name := v_LoadRec.store_name;
    v_DimRec.store_short_name := v_LoadRec.store_short_name;

    =====
    -- Update the dim_store
    =====

v_step := 80;
UPDATE dim_store
SET
    store_name = v_DimRec.store_name,
    store_short_name = v_DimRec.store_short_name,
    dss_update_time = sysdate
WHERE
    store_code = v_DimRec.store_code;

    =====
    -- If the record is not found then insert it
    =====

v_step := 100;
IF SQL%NOTFOUND THEN
    INSERT INTO dim_store
    (
        dim_store_key,
        store_code,
        store_name,
        store_short_name,
        dss_update_time,
        dss_create_time
    )
VALUES
    (
        dim_store_seq.NEXTVAL,
        v_DimRec.store_code,
        v_DimRec.store_name,

```

```

        v_DimRec.store_short_name,
        SYSDATE,
        SYSDATE
    );

```

Дименсия на доставчиците



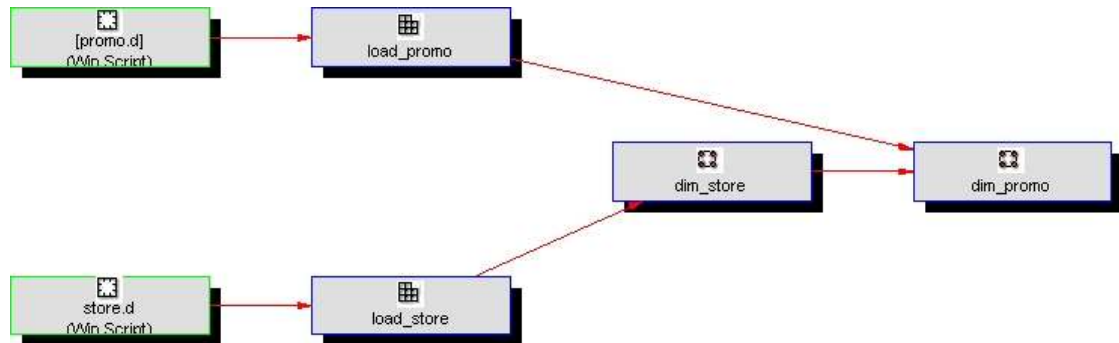
Търговската политика в търговска верига Бриколаж позволява сключването на договор с нов доставчик или промяната на условията със съществуващ доставчик да се извършва само от специализиран отдел в централния офис на дружеството. На тези служители е поверена и отговорността да поддържат съответните данни по доставчик в актуален вид. От гледна точка на интеграцията, данните се намират и актуализират на едно място – в централната SIGMA, за това процедурата по извличане и зареждане не се различава от тази за дименсията на магазините. В дименсията на доставчиците обаче влиза много по-разнообразна и често актуализирана информация. Това налага прилагането на базови трансформационни похвати, с които са се изчистват зарежданите в димензионната таблица данни:

```

CURSOR c_Load IS
  SELECT
    load_four.cod_four cod_four
  , load_four.four_name four_name
  , .....
  , decode(load_four.four_dat_creat, '?', cast(Null as
date), to_date(load_four.four_dat_creat, 'dd/mm/rr')) four_dat_creat
  , .....

```

Дименсия на промоциите



За разлика от дименсиите на магазините и доставчиците, дименсията на промоциите трябва да събере своите данни от всички функциониращи магазини. Търговската практика на Бриколаж допуска създаването на локални промоции, съответстващи на местните особености на всяко населено място. Поради тази причина информацията за провежданата маркетингова политика по места се съдържа в базите данни на оперативните системи на всеки магазин.

За да се извлекат тези данни се използва подхода за ежедневен експорт на данни от всяка SIGMA в нейната локална файлова система. След това използвайки изградения VPN консолидиращото приложение изтегля експортираните файлове:

```
SET STORES=platfo sofia sofia2 plovdi varna blago
SET LOAD_FILE=f:\export\promo.d

REM Concatenate all Sigma promo data to one single file
rsh centra for d in %STORES%; do f=data/$d/promo.d; if [ -f $f ] ;
then cat $f; fi; done >%LOAD_FILE%
```

Данните се зареждат в `load_promo`. От там зареждането в `dim_promo` преминава през определяне на уникален ключ за всяка промоция, но и през добавянето на един външен ключ – този за магазина, в който се случва промоцията:

```
CURSOR c_Load IS
```

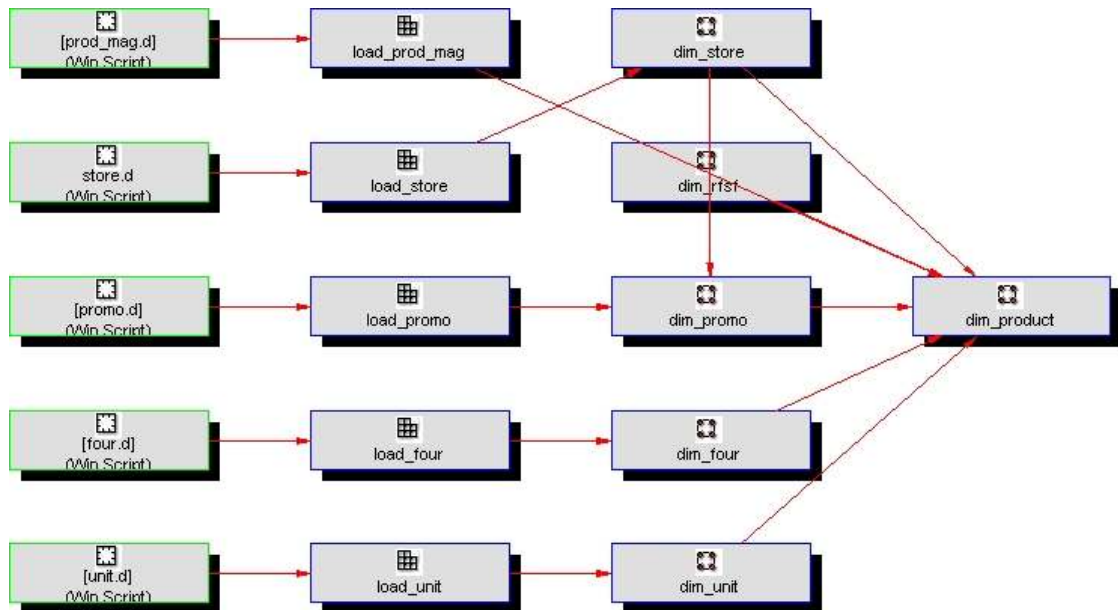


```

SELECT
  load_promo.id_pro id_pro
, load_promo.cod_pro cod_pro
, load_promo.year_pro year_pro
, load_promo.promo_name promo_name
, load_promo.start_date_delivery start_date_delivery
, load_promo.end_date_delivery end_date_delivery
, load_promo.start_date_sale start_date_sale
, load_promo.end_date_sale end_date_sale
, load_promo.cod_mag cod_mag
, dim_store.dim_store_key dim_store_key
FROM
  load_promo,
  dim_store
WHERE load_promo.cod_mag = dim_store.store_code

```

Дименсия на продуктите



Дименсията на продуктите е най-ценната дименсия в цялата схема на Бриколаж. Тя осигурява контекста за анализ на търговските показатели в търсения Business Intelligence модел в дружеството.

Нейното поддържане е и най-сложния момент в изграждането на дименционалните таблици в data warehouse.

За интегритета на данните е необходимо предварително да са актуализирани данните за магазините, промоциите, доставчиците.

При извличане на данните от сорс системите се използва подхода, описан в дименсията на промоциите:

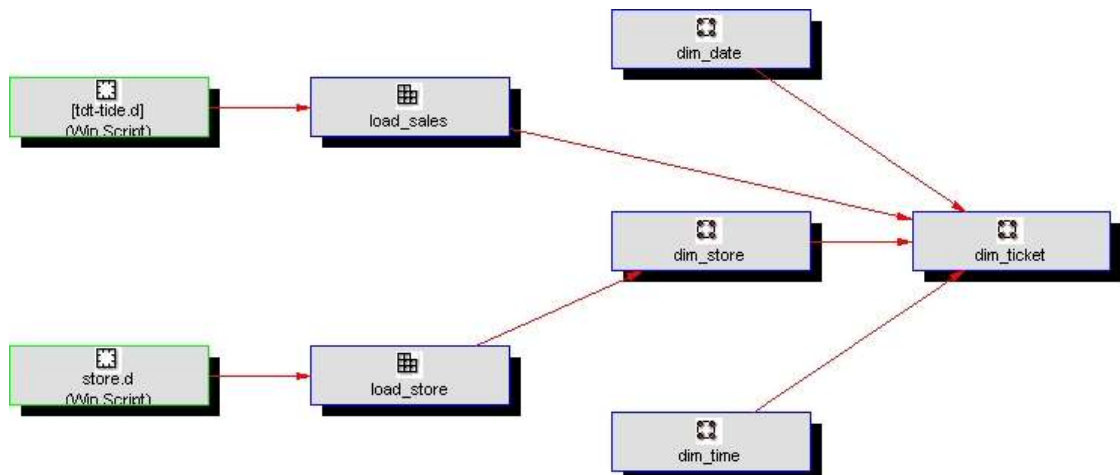
```
SET STORES=platfo sofia sofia2 plovdi varna blago
SET LOAD_FILE=f:\export\prod_mag.d
```

```
REM Concatenate all Sigma prod_mag data to one single file
rsh centra for d in %STORES%; do f=data/$d/prod_mag.d; if [ -f $f ] ;
then cat $f; fi; done >%LOAD_FILE%
```

Зареждането на данните и тук изисква тяхното изчистване и добавяне на външни ключове. В случая се добавят ключове към дименсиите на магазините, промоциите, доставчиците, мерните единици и продуктите групи.

В дименсията на продуктите освен описателни характеристики на продуктите могат да се намерят и някои факти. Такъв факт, използван често в анализите на бизнес потребителите е стоквата наличност на продуктите. Тя се актуализира всеки ден от ETL процеса.

Дименсия на касовите бележки



Дименсията на касовите бележки предоставя еднозначна референция за осъществяването на продажба. Всяка касова бележка може да съдържа много редове, но всички те ще съдържат референцията на една касова бележка.

Всяка SIGMA система експортира данните за дневните продажби във файла tdt-tide.d . В него във всеки ред се съдържат както детайлите от продажбата, така и обобщаващата информация (хедър). При зареждане на новите касови бележки за деня се взема информацията от полетата, описващи хедъра:

```
CURSOR c_Load IS
  SELECT
    load_sales.cac_code cac_code
  , load_sales.tdc_acc1 tdc_acc1
  , load_sales.total_amt total_amt
  , load_sales.invoice_no invoice_no
  , load_sales.bulstat bulstat
  , load_sales.tax_no tax_no
  , dim_store.dim_store_key dim_store_key
  , dim_date.dim_date_key dim_date_key
  , dim_time.dim_time_key dim_time_key
  FROM
    load_sales,
    dim_store,
    dim_date,
    dim_time
  WHERE load_sales.cod_mag = dim_store.store_code
  AND
  to_number(substr(load_sales.time,1,2)||substr(load_sales.time,4,2)) =
  dim_time.time_hhmm
  AND load_sales.dat_arc = dim_date.calendar_date
```

Факт таблица на продажбите



Изгаждането на факт таблицата е най-важният процес в ETL процедурата. Преди да се пристъпи към него е необходимо всички референтни данни да са обработени и

заредени коректно, така че ключовете, свързващи всеки ред от факт таблицата с димензионните таблици да съществуват.

Заредените вече данни в load_sales влизат в staging зоната, където се осъществяват всички необходими трансформации и връзки:

```
CURSOR c_Load IS
SELECT
  load_sales.cac_code cac_code,
  load_sales.tdc_accl tdc_accl,
  load_sales.tdc_lign tdc_lign,
  load_sales.tdd_remi tdd_remi,
  load_sales.tdd_qute tdd_qute,
  load_sales.tdd_grat tdd_grat,
  load_sales.ard_pvte ard_pvte,
  load_sales.tdd_pven tdd_pven,
  load_sales.tvc_code tvc_code,
  load_sales.cod_anpf cod_anpf,
  load_sales.tdd_pvht tdd_pvht,
  load_sales.pv_ttc_ac pv_ttc_ac,
  load_sales.pa_pond pa_pond,
  load_sales.dat_arc dat_arc,
  load_sales.time time,
  load_sales.cod_mag cod_mag,
  load_sales.cod_four cod_four,
  load_sales.tdd_comm cod_rfsf,
  load_sales.id_pro id_pro,
  load_sales.id_gencod
FROM
  load_sales;
```

Ключовете към дименционалните таблици се извличат с процедури като тази:

```
=====
-- Key lookup for the dimension dim_store
=====
v_step := 10;
p_status := get_dim_store_key(
  v_StageRec.cod_mag,
  'N', 'N',
  p_sequence, p_job_name, p_task_name, p_job_id, p_task_id,
  v_StageRec.dim_store_key);

-- See if we added a new key or failed to find one
IF p_status = 2 THEN
  v_dim_store_add := v_dim_store_add + 1;
ELSIF p_status != 1 THEN
  v_dim_store_nul := v_dim_store_nul + 1;
END IF;
```

С това приключва процесът по консолидиране на данните за продажби от всички магазини в дружеството. Освен консолидирането на данните за продажбите, с посочените процедури се постигат и други консолидационни задачи – да се предостави единен източник на данни за продуктите и промоциите в дружеството.

Но работата по Business Intelligence проекта в търговска верига Бриколаж не спира до тук. Изграденият data warehouse е само основата за разработването на цялостни аналитични приложения, които да подпомогнат мениджъри, анализатори и редови служители в магазините във всекидневната им работа и постигането на ефективен и ефикасен работен процес.

Заклучение

Успешното управление на търговско дружество, което оперира в отворена пазарна среда, изисква вземането на стратегически решения по много и различни бизнес казуси. Интеграцията на данните, които съществуват в разпръснатите информационни системи, е стъпка в посока улесняване на процеса по вземане на решения.

Интеграцията на данни в една организация е сам по себе си стратегически въпрос, който трябва да намери своя отговор на база задълбочено проучване на наличната информационна инфраструктура и визията на бизнеса за бъдещите предизвикателства, с които тя ще трябва да се справи.

Подходът на консолидация на данните, избран от Бриколаж за водещ в интеграционния проект, доказва своята жизнеспособност и ефективност в последващите етапи на Business Intelligence инициативата на търговската верига. Чрез създадените batch задачи ежедневно се издърпват големи обеми транзакционни данни от оперативните системи на вече седем магазина. Последващата обработка в staging зоната не натоварва оперативните системи, а ресурсите, заделени за извършване на трансформациите са планирани да поемат и следващите четири магазина на Бриколаж, без да се налага промяна в дизайна на процесите.

Използването на утвърдени в логическото моделиране практики позволява прозрачното разширяване на модела на данните в общия data warehouse с нови области на анализ като анализ на веригата за доставки и анализ на стоките запаси.

Разглеждайки проекта за интеграция на данни в контекста на Business Intelligence задачите, решавани от дружеството не може да не бъдат забелязани известни пропуски в този първоначален етап на реализация на решението. Такъв е липсата на специфична factless fact table, която да отчита участието на продуктите в промоция, с което да се улавят продукти, включени в промоция, но непродавани поради някаква причина.

Това обаче не намалява удовлетвореността на крайните потребители на консолидираните данни. Информация изисквана от потребителите и доставяна от ИТ отдела след няколко дни вече се достъпва без посредничеството на ИТ служители и в рамките на минути. Разрези, немислими за фиксираната структура на данните в оперативните системи, показват интуитивно налучквани преди това зависимости. Появява се една изцяло нова култура на служителите по отношение вземането на решения, базирани на информация.

Посоката за развитие на информационната стратегия в търговската верига Бриколаж е вече ясно дефинирана – информацията е стратегически ресурс и като такъв ще бъде обект на старателен и итеративен процес на анализ и оптимизиране.

Използвана литература

1. Gartner (2006), Magic Quadrant for Data Integration Tools
2. Larissa T. Moss, Shaku Atre (2003) "Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications"
3. Colin White (2005), Data Integration: Using ETL, EAI, and EII Tools to Create an Integrated Enterprise
4. Philip Russom (2006), Master Data Management: Consensus-Driven Data Definitions for Cross-Application Consistency
5. Wayne W. Eckerson (2005), Development Techniques for Creating Analytic Applications
6. <http://tdwi.org/> - The Data Warehousing Institute
7. Ralph Kimball (2002), The data warehouse toolkit