- сейчас у нас приоритетное направление в проекте - это так называемые **VEL**

- это значит **Ventes En Ligne** = онлайновые продажи

- речь идет о продажах абонементов на газеты и покупках электронных версий статей или других предметов

- сначала мы их называли PVL = Projet des Ventes en Lignes, т.е. "проект онлайновых продаж", поэтому объекты в базе данных, которые относятся к этой части, имеют префикс PVL\_

**- в рамках этих VEL мы получаем от клиента ежедневно по четыре файла**

- файлы следующие:

- **1- каталог абонементов и продуктов**. Название файла: **CatalogueDesOffres**. Таблица, куда он загружается: **import.PVL\_CatalogueOffres**

- **2 - клиенты, т.е. покупатели, пользователи этой услуги**. Файл называется **Accounts**, таблица: **import.PVL\_Utilisateur**

- (utilisateur = user)

**- 3 - покупки, заказы**. Файл: **DailyOrders**, таблица**: import.PVL\_Achats**

- (achat по-фр. - покупка)

**- 4 - абонементы, подписка**. Файл: Subscriptions, таблица: **import.PVL\_Abonnements**

- имя конкретного файла включает в себя, как обычно у нас, также timestamp, т.е. дату выгрузки данных клиентом, и суффикс .csv например, файл может называться Subscriptions-25032015.csv

- структуры таблиц, форматные файлы, процедуры валидации (rejeter) и публикации (Publier) уже существуют. в этом можно убедиться, если посмотреть схемы import и rejet : там есть объекты с префиксом PVL

- беда тут вот в чем:

- каждое издание имеет свои особенности в тех данных, которые они присылают

- поэтому структура файлов одинакова, а вот с данными приходится обращаться немного по-разному, в зависимости от того, присылает ли их газета EQUPE, или Le Parisien, или вот теперь новый клиент той же группы - газета France Football - это "дочка" Equipe, посвященная целиком футболу

- вот она http://www.francefootball.fr/

- ну так вот. Андрей, ты следишь? всё ли пока понятно?

- отлично.

- ну так вот. сначала, еще в конце прошлого года, мы сделали VEL Equipe

- Equipe первая перешла на VEL, а теперь подтягиваются остальные

- VEL Equipe мы сдали в производство в январе. когда потом в марте мы получили наконец данные Le Parisien, оказалось, что код, написанный для Equipe, на этих данных не работает, и его пришлось переделывать. собственно, этим я и занимался в последние две недели, после поездки к вам

- таким образом, сейчас база отличается в части PVL от той, которая есть у тебя

- все процедуры валидации и публикации существуют отныне в двух вариантах: EQ и LP

- в то время как таблицы - в одном варианте, прежнем, без суффиксов LP или EQ, поскольку структура файлов идентична

- процедуры я пришлю. в общем да, но и в EQ есть некоторые изменения, поэтому я пришлю все

- теперь наконец прислали данные FF - Франс Футбол

- под эти данные нужно будет разработать третью серию процедур, с суффиксом \_FF

- приказ командования -

- поручить это тебе

- делать там нужно будет не очень много, но процедуры довольно сложные, особенно в части абонементов

- а поскольку на следующем проекте, для газет Юго-Запада (Бордо) этого будет всего много - тоже покупки, абонементы, клиенты, каталоги - то сейчас это удобный случай вникнуть в эту кухню

- а я буду в это время заниматься репортингом - это различные статистики, новая серия тоже, и там сроки начинают подпирать

- хорошо?

- тогда для начала я сейчас пришлю новые версии процедур, чтобы ты их мог скомпилировать у себя

- а потом еще пообъясняю, в чем там дело

- послал по почте

- хорошо. попробуй, пожалуйста, их скомпилировать на своем сервере

- ну как? получается, или не хватает каких-то объектов?

- каких?

- хорошо

- там наверняка не хватает - пока можно всю команду IF (EXISTS(... EXECUTE - это для пересчета статистики по отброшенным записям, такая финтифлюшка у нас стоит в конце каждой процедуры публикации

- там где execute (@S) тоже убрать в комментарий: при компиляции это незаметно, а потом при исполнении процедуры будет ошибка

- но рамку, т.е. цикл while оставить, пусть он впустую крутится

- всё получилось? других проблем при компиляции не было?

- очень хорошо. тогда: зачем это всё нужно и как это всё работает?

- в файлах PVL содержится информация о товарах (газетах, статьях), об отдельных покупках, о подписках на издания, и, наконец, самое главное - о людях, которые осуществляют подписку и покупку.

- на основе этих данных, наши процедуры публикации заполняют следующие таблицы базы:

- ref.CatalogueAbonnements - каталог абонементов

- ref.CatalogueProduits - каталог товаров для разовой продажи

- dbo.AchatsALActe - разовые покупки (achat à l'acte = разовая покупка)

- dbo.Abonnements - абонементы, подписки на издания. абонемент отличается от разовой покупки тем, что имеет протяженность во времени. хотя в некотором смысле и то и другое - покупка

- brut.ConsentementsEmail - уже известные нам оптины

- внимание: PVL не заполняет brut.Contacts

- потому что все клиенты онлайновых продаж должны быть нам известны из других источников

- вот здесь-то и прячется рыба под камнем

- сейчас объясню

- собственно, процедуры EQ отличаются от процедур LP тем, как мы разыскиваем, какой человек ("контакт") совершил покупку или подписку на издание

- да, именно так: детективная история :)

- поэтому сейчас нужно сказать два слова о том, какие у нас есть вообще источники данных

- сделай, пожалуйста запрос select \* from ref.Sources

- источник номер 1: Neolane L'Equipe

- Neolane - это такая система, инструмент для маркетинговых кампаний - то есть учет клиентов, рассылка и т.п.

- произносится "неолЕн". у газеты Equipe есть своя система Neolane, откуда они нам выгружают данные об электронных подписчиках -

- и до того, как они перешли на VEL, источник Neolane нам поставлял также данные о покупках и подписках

- Андрей, у меня не подключен микрофон

- сейчас подключу

- что-нибудь случилось?

- а, хорошо. конечно, письменно рассказывать дольше, зато всё остается

- понятно :)

- продолжаю повествование

- итак

- данные о клиентах Eqyipe продолжает хранить в неолене

- но в части разовых продаж и электронных абонементов они перешли на VEL

- поэтому получается вот что: таблицу контактов мы заполняем данными из неолена, а когда получаем данные о продажах из VEL, мы должны поставить пользователя VEL в соответвие клиенту из Neolane

- то есть, мы не создаем нового человека, а разыскиваем старого

- поэтому если ты посмотришь в процедуре import.PublierPVL\_Utilisateur\_EQ как мы находим ProfilID

- (по-фр. trouver = найти)

- то там будет такая связка таблиц:

- наша таблица #T\_Trouver\_ProfilID с данными из PVL\_Utilisateur сначала связвается с неоленовской таблицей import.NEO\_CusCompteEFR по ClientUserId=sIdCompte

- это таблица соответствий, где мы находим iRecipientId, который является первичным ключом контакта в неолене

- и поэтому OriginalID в brut.Contacts для SourceID=1

- так что вторым шагом мы запросто находим ProfileID по

- inner join brut.Contacts b on a.iRecipientId=b.OriginalID and b.SourceID=1

- но этот прием работает только для Equipe

- для Le Parisien он не годится, потому что там другие источники данных о клиентах: во-первых источник 2 = SSO Le Parisien

- во-вторых, источник 4 = Prospects Le Parisien

- в третьих, источник 3 = SDVP для марок LP (Le Parisen) и AF (Aujourd'hui en France)

- и стратегия поиска там другая

- думаю, что на сегодня хватит

- уже, наверное, ты устал, да и я тоже

- а мне еще нужно программку постараться написать до конца дня

- так что помедитируй пока над сказанным и как максимум сравни, как отыскивается ProfilID в процедурах PublierPVL\_Utilisateur для EQ, с одной стороны, и для LP, с другой.

- три источника: 2, 4, 3 - в этом порядке приоритетности

- источник 1 мы не имеем права брать для Le Parisien даже если не нашли в тех трех

- с источником 10 = PVL мы не создаем контактов

- хотя создаем покупки и абонементы

- если контакт не найден, покупка вообще не создается: без контакта она не имеет смысла

- те строчки, для которых невозможно найти контакт, мы стараемся отбрасывать уже на стадии rejet

- для пользователей нашей системы имеют смысл только те покупки, которые мы можем привязать к человеку

- в абсолюте, все покупки должны находить своего покупателя

- но поскольку нет в мире совершенства, то иногда бывает, что не находим - какое-то несоответствие значит есть в данных клиента

- мы обсуждали это с нашими пользователями из Equipe и Le Parisien, по поводу даных источника Milibris, где нет персональных данных, и мы им предлагали это как вариант -

- но они не согласились

- потому что Noname тогда будет тяжелым камнем лежать именно на репортинге, статистиках

- и всё перекашивать в свою сторону

- белое пятно, которое будет сильно нарушать пропорции

- ну хорошо, Андрей. я желаю тебе доброго вечера, а сам займусь своей программкой. завтра на свежую голову станет яснее.

- Доброе утро, Андрей! Как дела? Сегодня после обеда продолжим VEL ?

- Отлично. В понедельник у нас во Франции выходной - католическая Пасха. В Украине выходной следующий понедельник, так?

- хорошо

- Минут через 10, хорошо?

- Андрей, ты готов продолжить сейчас?

- Какие у тебя есть вопросы по вчерашним объяснениям?

- в dbo.Contacts его нет, потому что в dbo.Contacts находятся уже "чистые" контакты, полученные в результате дедупликации. Дедупликация позволяет сказать: вот этот контакт, информацию о котором мы получили с одной стороны, из Neolane, а с другой стороны, из SDVP - это один и тот же человек. Поэтому в dbo.Contacts у нас синтетическая запись, которая может быть составлена из нескольких строчек brut.Contacts

- да, разумеется. в процедурах публикации мы работаем с контактами только на уровне brut

- понятно ли тебе соотношение brut.Contacts и dbo.Contacts или еще не вполне?

- нет, нельзя, хотя бы потому что "чистого" может еще не быть, в то время как брутто-контакт уже существует

- потому что по фазам процесса - мы рисовали эту схемку, что когда происходит в дневном цикле - сначала публикация, то есть создание записей brut всех таблиц, а затем дедупликация и Build

- но у нас во Франции есть поговорка: "зачем делать просто, когда можно сделать сложно?"

- поэтому некоторые таблицы dbo мы заполняем уже на стадии публикации, в том числе dbo.AchatsALActe и dbo.Abonnements

- потом процедура etl.BuildContacts подставит MasterID, если нужно

- вот так:

- update X set MasterID = bC.MasterID

from dbo.Abonnements X inner join brut.Contacts bC

on X.ProfilID = bC.ProfilID and (X.MasterID != bC.MasterID or X.MasterID is null)

- строчки абонементов или покупок остаются прежними до и после дедупликации, поэтому нам не нужна таблица brut.Abonnements или brut.AchatsALActe

- понятна ли эта логика, Андрей? это момент очень важный и не очень простой, поэтому я на нем останавливаюсь подробнее

- хорошо. со временем это прояснится как следует.

- моя мысль такая: вообще-то твоим заданием будет сейчас адаптировать процедуры PVL Equipe для PVL France Football

- в основном, как я думаю, это затронет поиск контактов

- но при этом мне хотелось бы также объяснить и понятие абонемента - это самая сложная часть, - и дать представление всё более и более ясное о том, как вообще всё устроено

- но начнем с покупок, это проще

- покупка = контакт + товар + момент покупки

- в принципе да, хотя здесь кво=1

- кажется

- потому что может иметь смысл купить две одинаковые бумажные газеты, но нет особого смысла покупать два одинаковых файла PDF

- но в общем случае ты совершенно прав: количество экземпляров

- в источнике SDVP где продаются бумажные издания,есть понятие количества экземпляров, поэтому в таблице dbo.AchatsALActe оно есть

- при этом даже электронные товары могут иметь количество: например, билеты на футбольный матч

- так что да

- но всё-таки уникальность покупки определяют эти три элемента: покупатель+товар+датавремя покупки

- сведения о покупателях VEL содержатся в таблице import.PVL\_Utilisateur куда мы загружам файл Accounts или AccountReport, как он называется в ФФ

- сведения о товарах - в каталоге товаров: таблица import.PVL\_CatalogueOffres, строки из которой мы публикуем в таблицу ref.CatalogueProduits

- сведения о конкретных покупках - в таблице import.PVL\_Achats которую мы публикуем в dbo.AchatsALActe

- это всё просто и ясно?

- можешь задавать вопросы по мере их возникновения, пожалуйста

- или просто, когда чувствуешь, что перестал понимать

- или не успеваешь следить

- ...и момент покупки

- ведущая таблица у нас здесь import.PVL\_Achats

- открой, пожалуйста, код процедуры - или скорее - а я себе пока чаю налью :)

- как обычно, я создаю здесь в начале временную табличку, в которую буду собирать данные из разных исходных и справочных таблиц

- временная таблица покупок #T\_Achats

- пройдемся по полям этой таблицы

- ProfilID - это брутто-профиль покупателя: его предстоит найти, сначала он будет NULL

- ClientUserID - это идентификактор покупателя в системе клиента

- ему-то и надо будет найти соответствие в нашей системе, т.е. сответствующий ему ProfilID

- OriginalID - это "зачем делать просто..." - здесь это код товара (продукта) в системе клиента

- в то время как в предыдущей задаче это был код контакта в системе клиента

- надо не запутаться

- ProduitID - это код товара в нашей системе, т.е. в таблице ref.CatalogueProduits - его тоже предстоит найти, а сначала он NULL

- SourceID - это код источника, для PVL это 10 (см. таблицу ref.SourceID)

- да, ref.Sources, разумеется!

- Marque - это марка издателя, то есть Equipe или France Football

- мы ее найдем в каталоге, а пока тоже не знаем

- справочная таблица для марок:

- select \* from ref.Misc a where a.TypeRef=N'MARQUE'

- NomProduit - название товара

- nom по-фр. = name по-англ.

- AchatDate - датавремя покупки

- это у нас всегда тип datetime, потому что минуты-секунды важны

- ExProdNb - это количество экземпляров, неуклюжее сокращение от Exemplaires Produit Nombre

- Reduction - это процент скидки, здесь не используется

- MontantAchat - это сумма, стоимость покупки

- montant (произносится "монтан") = "денежная сумма", очень употребительное в нашем деле слово

- то же самое, что англ. amount

- OrderID - уникальный номер заказа, т.е. единичной покупки, строчки в файле DailyOrder

- ProductDescription - описание товара, более подробное чем название

- MethodePaiement - метод оплаты (как правило, Credit Card)

- CodePromo - код акции, от слова promotion (произносится "промосьон") - коммерческая акция, очень популярное во французском народе слово

- Provenance - происхождение покупки, как правило Web

- CommercialId - имя продавца (если продажа оффлановая, таких почти нет)

- не имя, а ID, конечно

- SalonId - Id салона при оффлановой покупке (тоже почти нет)

- ModePmtHorsLigne - метод оффлановой покупки (как правило, чек)

- hors ligne (читается "ор линь") = off line

- iRecipientId - вот, наконец, самое интересное

- iRecipientId - это идентификатор покупателя в системе Neolane. Его нам предстоит найти сначала, а затем по нему - ProfilID в нашей системе

- для ФФ его тоже нужно будет тебе найти, но другим способом, чем для Экип

- ну и наконец ImportID - это уже знакомый тебе идентификатор строчки в таблице импорта - в данном случае это будет ImportID таблицы import.PVL\_Achats

- уф

- да

- это относится к методу поиска контакта для VEL Le Parisien

- в этой процедуре не применяется

- и для ФФ тоже не будет применяться

- там будет только источник номер 1 = Neolane

- Группа Амори, которой принадлежит весь этот базар, разделена на две логические части

- первая - эта спортивная пресса, то есть Экип и ее дочки

- вторая - общественно-политическая, это Ле Паризьен и его производная Aujourd'hui en France ("Сегодня во Франции")

- люди из этих двух частей сидят по разным домам

- и не могут видеть данных друг друга

- в нашей системе

- поэтому мы не можем использовать профили из Ле Паризьен для покупок Экип, даже если они существуют - один и тот же человек ведь может быть подписан на то и на то

- и это часто бывает

- но внутри системы мы их разделяем по полю Appartenance (произносится "апартенанс") = принадлежность

- в источнике Neolane у нас только контакты Экип

- да, и есть еще головная группа маркетинга, которая имеет доступ к обеим частям системы

- Appartenance - это, как мы уже говорили, битовая маска, числовые значения у нее 1 = Экип, 2 = Паризьен, 3 = оба, т.е. "Группа Амори"

- в двоичном коде: 01, 10 и 11

- контакт в dbo.Contacts

- dbo

- но не brut

- может иметь любую из трех "принадлежностей"

- но Абонемент или покупка не могут быть "Группа" - они всегда либо Экип, либо Паризьен

- не очень понятно?

- вот именно

- потому что в brut мы всегда знаем точно, из какого источника мы получили контакт

- а в dbo - это уже синтез из нескольких записей brut

- поэтому в dbo он может иметь Appartenance=3

- источники делятся так: 1 = Neolane - только Экип

- 2 = SSO и 4 = Prospects - только Паризьен

- 3 = SDVP - могут быть и Экип, и Паризьен, и мы их различаем по префиксу в OriginalID : EQ = Equipe, LP и AF = Le Parisien

- 5 и 6 - это Le Parisien, больше не используются, это была разовая загрузка

- запутанно?

- со временем прояснится

- во всяком случае, в этой процедуре у нас всё относительно просто

- здесь только Экип

- и Франс Футбол - это тоже Экип по принадлежности

- источник контактов здесь только один - это Neolane

- пойдем дальше по процедуре?

- в понедельник, поскольку меня не будет, собирался ли ты целый день работать по своему другому проекту или как всегда половину дня по нашему?

- не слишком ли будет сложно пока разбираться одному тебе с процедурами публикации PVL ?

- как хочешь: можешь в понедельник их сам поразбирать - сначала покупки, затем абонементы - но абонементы нужно объяснять, я думаю, там сильно всё накручено

- это да, но надо чтобы я вначале сказал, что такое абонемент, в системном смысле

- тогда я скажу это сейчас, не вдаваясь особо в детали, просто как ориентир

- а затем продолжим по покупкам сколько успеем

- для нас абонемент - это совокупность подписок, перерыв между которыми не больше 180 дней

- если поле Recurrent=1 - то есть "продление абонемента по умолчанию"

- а если Recurrent=0, т.е. подписка на твердый срок, то они не объяединяются, а 1 подписка = 1 абонемент

- подписка - технически это строчка таблицы PVL\_Abonnements файла Subscriptions

- в процедуре import.PublierPVL\_Abonnements\_EQ самая сложная часть -это как раз объединение отдельных подписок в абонемент

- но ее менять для ФФ скорее всего не придется

- но тот же принцип - в абонементах всех источников

- и в новом проекте клиенты из Бордо сказали, что они так же хотят

- поэтому и там придетмся это программировать

- ага, вот

- получил обработанные данные ФФ

- сейчас пришлю тебе их и к ним форматные файлы для загрузки

- могу еще прислать батчи для загрузки - валидации - публикации всех серий

- прислать?

[03.04.2015 18:00:35] Андрей Брагар: да. чтобы произвести весь цикл