

|  |
| --- |
| Datawarehouse for Sales Analysis  RELOUIS COMPANY |
| Relouis-Cosmetics |

| Related Artifacts | |
| --- | --- |
| Ref. | Name |
|  |  |
|  |  |

|  |  |
| --- | --- |
| Abbreviations and Acronyms | |
|  |  |
|  |  |

Contents

[1 Описание предметной области 4](#_Toc498585782)

[1.1 бизес основа 4](#_Toc498585783)

[1.2 недостатки существующих решений 4](#_Toc498585784)

[1.3 преимущества внедрения хранилища данных 4](#_Toc498585785)

[2 Бизнес измерения 5](#_Toc498585786)

[2.1 описание бизнес-процесса 5](#_Toc498585787)

[2.2 Определение зерна 5](#_Toc498585788)

[2.3 описание измерений 5](#_Toc498585789)

[2.4 Описание фактовой таблицы 8](#_Toc498585790)

[3 логическая схема 9](#_Toc498585791)

[3.1 схема слоя bl\_3nf: 9](#_Toc498585792)

[3.2 Схема слоя bl\_dm 10](#_Toc498585793)

[4 Объекты партиционирования 10](#_Toc498585794)

[5 Бизнес процессы 10](#_Toc498585795)

[6 Модель данных 10](#_Toc498585796)

[6.1 описание типов измерений 10](#_Toc498585797)

[6.2 слои хранилища данных 11](#_Toc498585798)

# Описание предметной области

## бизес основа

Компания Relouis – это белорусская компания, которая на данный момент производит более 600 наименований и включает все группы декоративной косметики. На данный момент насчитывается 5 коллекций и 5 основных групп товара.

## недостатки существующих решений

На данный момент данные поступают из различных источников и в «грязном» виде. Будет производится их чистка и улучшение performance, т.к. прежнее хранилище уже плохо справляется с данной нагрузкой. Когда разрабатывалось первое хранилище, никто не думал, что компания за столько небольшой срок наберет такую широкую массу покупателей и так хорошо наладит бизнес, поэтому и возникла проблема в создании нового хранилища данных.

## преимущества внедрения хранилища данных

* улучшиться доступ к данным,
* расширится круг разрезов, по которым менеджеры смогут проводить аналитику,
* ускорить процесс обработки данных

# Бизнес измерения

## описание бизнес-процесса

Учет продажи товаров для будущего анализа компании с использованием различных метрик.

## Определение зерна

Зерном будет являться продажа какого(их)–то товаров определенном покупателю, определнным сотрудником в конкретном магазине.

## описание измерений

Схема будет содержать следующие измерения:

| DIM\_Customers | | |
| --- | --- | --- |
| Название колонки | Тип колонки | Описание (Пример) |
| Customer\_id | NUMBER(10) | ПК,Уникальный идентификатор для каждого покупателя (20222) |
| First\_name | VARCHAR2(40 BYTE) | Имя покупателя (Алина) |
| Last\_name | VARCHAR2(40 BYTE) | Фамилия покупателя (Макарец) |
| Phone | VARCHAR2(40 BYTE) | Телефон покупателя (+375294444444) |
| Email | VARCHAR2(40 BYTE) | Почтовый адрес для рассылки с информацией о новинках/акциях (alinamakarets@tut.by) |
| Age | NUMBER(4) | Возраст покупателя (20) |
| Discount | NUMBER(2) | Персональная скидка для покупателя (5-имеется в виду 5%) |
| Address\_id | NUMBER(10) | ВК для таблицы с указанием развернутого адреса покупателя (2032) |
| Insert\_dt | DATE | Дата вставки строки |
| Update\_dt | DATE | Дата обновления строки |

| DIM\_Locations | | |
| --- | --- | --- |
| Название колонки | Тип колонки | Описание (Пример) |
| Address\_id | NUMBER(10) | ПК, развернутый адрес магазина или покупателя (2032) |
| Address | VARCHAR2(40 BYTE) | Указание улицы и номера дома (Дзержинского 95) |
| Postal\_code | VARCHAR2(40 BYTE) | Почтовый индекс (225002) |
| City\_id | NUMBER(4) | Код города ( 17) |
| City\_name | VARCHAR2(40 BYTE) | Название города (Минск*, только в случае с Минском будет дублироваться название региона и города, т.к. в нашем случае это одно и тоже)* |
| Region\_id | NUMBER(4) | Код региона (7) |
| Region\_name | VARCHAR2(40 BYTE) | Название региона (Минск) |
| Country\_code | VARCHAR2(10 BYTE) | Уникальный код страны (BY) |
| Country\_name | VARCHAR2(40 BYTE) | Название страны (Беларусь) |
| Insert\_dt | DATE | Дата вставки строки |
| Update\_dt | DATE | Дата обновления строки |

| DIM\_Products | | |
| --- | --- | --- |
| Название колонки | Тип колонки | Описание (Пример) |
| Product\_id | NUMBER(10) | ПК, для обозначения уникальности продукта (203) |
| Product\_name | VARCHAR2(40 BYTE) | Название продукта (Тушь с эффектом накладных ресниц) |
| Product\_code | VARCHAR2(10 BYTE) | Уникальный код продукта (456-12) |
| Color\_name | VARCHAR2(40 BYTE) | Название цвета (Черный) |
| Color\_id | NUMBER(4) | ВК, Код цвета (85) |
| Category\_name | VARCHAR2(40 BYTE) | Название категории (Глаза) |
| Category\_id | NUMBER(4) | ВК, Код категории ( 4) |
| Collection\_name | VARCHAR2(40 BYTE) | Название коллекции (Professional, *также возможно без указания названия коллекции, в этом случае пишется N/A*) |
| Collection \_id | NUMBER(4) | ВК, Код коллекции (1, *-999 для N/A)* |
| Price | NUMBER(10) | Цена (12) |
| Insert\_dt | DATE | Дата вставки строки |
| Start\_dt | DATE | Дата старта продаж коллекции |
| End\_dt | DATE | Дата окончания продаж коллекции |
| Is\_current | VARCHAR2(4 BYTE) | Показывает актуальность продаж по данной коллекции |

| DIM\_Collections | | |
| --- | --- | --- |
| Название колонки | Тип колонки | Описание (Пример) |
| Collection \_id | NUMBER(4) | ПК, Код коллекции (1, *-999 для N/A)* |
| Collection\_name | VARCHAR2(40 BYTE) | Название коллекции (Professional, *также возможно без указания названия коллекции, в этом случае пишется N/A*) |
| Collection\_date | DATE | Дата начала продажи коллекции (28-03-2016) |
| Collection\_desс | VARCHAR2(80 BYTE) | Описание коллекции ( Предназначена для пользования визажистами) |
| Insert\_dt | DATE | Дата вставки строки |
| Start\_dt | DATE | Дата старта продаж коллекции |
| End\_dt | DATE | Дата окончания продаж коллекции |
| Is\_current | VARCHAR2(4 BYTE) | Показывает актуальность продаж по данной коллекции |

| DIM\_Categories | | |
| --- | --- | --- |
| Название колонки | Тип колонки | Описание (Пример) |
| Category\_id | NUMBER(4) | ПК, Код категории ( 4) |
| Category\_name | VARCHAR2(40 BYTE) | Название категории (Глаза) |
| Category\_desс | VARCHAR2(80 BYTE) | Описание категории (тушь, тени для век, карандаши для глаз, подводки, карандаши для бровей и пр. ) |
| Insert\_dt | DATE | Дата вставки строки |
| Update\_dt | DATE | Дата обновления строки |

| DIM\_Colors | | |
| --- | --- | --- |
| Название колонки | Тип колонки | Описание (Пример) |
| Color\_id | NUMBER(4) | ПК, Код цвета (85) |
| Color\_name | VARCHAR2(40 BYTE) | Название цвета (Черный) |
| Color\_desс | VARCHAR2(80 BYTE) | Описание цвет (глубокий черный ) |
| Insert\_dt | DATE | Дата вставки строки |
| Update\_dt | DATE | Дата обновления строки |

| DIM\_Stores | | |
| --- | --- | --- |
| Название колонки | Тип колонки | Описание (Пример) |
| Store\_id | NUMBER(4) | ПК, Код магазина (101) |
| Store\_name | VARCHAR2(40 BYTE) | Название цвета (Фирменный Магазин Relouis) |
| Phone | VARCHAR2(40 BYTE) | Телефон магазина (+375172569849) |
| Address\_id | NUMBER(10) | ВК, развернутый адрес магазина (2092) |
| Manager\_id | NUMBER(10) | ВК, определяет ответственного за магазин (2223) |
| Insert\_dt | DATE | Дата вставки строки |
| Update\_dt | DATE | Дата обновления строки |

| DIM\_Employees | | |
| --- | --- | --- |
| Название колонки | Тип колонки | Описание (Пример) |
| Employee\_id | NUMBER(10) | ПК,Уникальный идентификатор для работника (2237) |
| First\_name | VARCHAR2(40 BYTE) | Имя работника (Александра) |
| Last\_name | VARCHAR2(40 BYTE) | Фамилия работника (Стародубец) |
| Phone | VARCHAR2(40 BYTE) | Телефон работника (+375295268896) |
| Email | VARCHAR2(40 BYTE) | Почтовый адрес работника (a.astarodybets@tut.by) |
| Age | NUMBER(4) | Возраст работника (27) |
| Position\_Name | VARCHAR2(40 BYTE) | Позиция работника (Продавец-Консультант) |
| Manager\_id | NUMBER(10) | ВК для таблицы **Employees** для указания менеджера работника (2036) |
| Start\_dt | DATE | Дата принятия на работу |
| End\_dt | DATE | Дата окончания работы на нанимателя |
| Is\_current | VARCHAR2(4 BYTE) | Показывает актуальность работы на данный момент ( работает такой человек или нет) |

| DIM\_Time\_Day | | |
| --- | --- | --- |
| Название колонки | Тип колонки | Описание (Пример) |
| Date\_id | DATE | ПК, Дата (04-01-2017) |
| Day\_of\_week | NUMBER(2) | День недели, число (3) |
| Day\_name\_of\_week | VARCHAR2(20 BYTE) | Среда |
| Day\_of\_month | NUMBER(8) | День в месяце (4) |
| Day\_of\_year | NUMBER(8) | День в году (4) |
| Week\_of\_month | NUMBER(8) | Неделя в месяце (2) |
| Week\_of\_year | NUMBER(8) | Неделя в году (2) |
| Month\_number | NUMBER(8) | Число месяца (1) |
| Month\_name | VARCHAR2(20 BYTE) | Название месяца (Январь) |
| Quarter | NUMBER(2) | Число квартала (1) |
| First\_day\_of\_month | DATE | Первый день месяца (1) |
| Last\_day\_of\_month | DATE | Последний день месяца (31) |
| Year | NUMBER(8) | Год (2017) |
| Year-Quater | VARCHAR2(20 BYTE) | Год-Квартал (2017 1 Квартал) |

## Описание фактовой таблицы

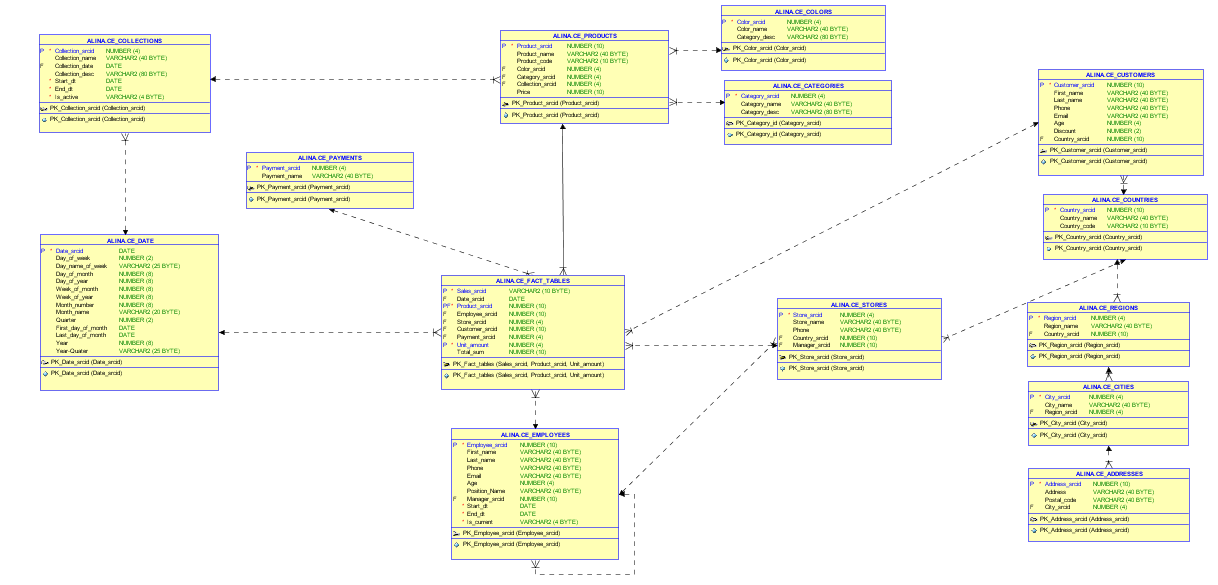
Фактовая таблица будет выглядеть следующим образом:

| FCT\_Sales | | |
| --- | --- | --- |
| Название колонки | Тип колонки | Описание (Пример) |
| Sales\_id | VARCHAR2(10) | Номер покупки (22897) |
| Event\_dt | DATE | ВК к таблице с датой (04-01-2017) |
| Product\_id | NUMBER(10) | ВК к таблице с продуктами (203) |
| Employee\_id | NUMBER(10) | ВК к таблице с сотрудниками (2237) |
| Store\_id | NUMBER(4) | ВК к таблица с магазинами (101) |
| Customer\_id | NUMBER(10) | ВК к таблице с покупателями (20222) |
| Payment\_method | VARCHAR2(25) | Тип Платежа. Доступно Наличными/ По карте |
| Unit\_amount | NUMBER(4) | Количество продуктов (2) |
| Total\_sum | NUMBER(10) | Общая сумма (20) |
| Insert\_dt | DATE | Дата вставки строки |
| Update\_dt | DATE | Дата обновления строки |

# 

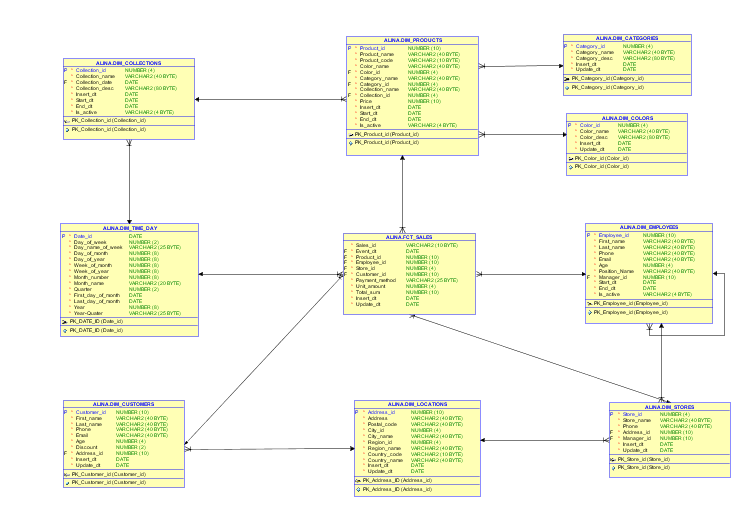
# логическая схема

## схема слоя bl\_3nf:



Picture 1 bl\_3nf

## Схема слоя bl\_dm



Picture 2 bl\_dm

В данном проекте целосообразно использовать “Snowflake” (в виде снежинки) схему, т.к. измерения DIM\_CUSTOMERS и DIM\_STORES имеют связь с измерением DIM\_LOCATIONS, в данном случае не целесообразно распределять измерение с локацией в данные два измерения. Помимо этого измерение DIM\_COLLECTIONS ссылается на измерение с датами.

# Объекты партиционирования

Для таблицы, содержащая факты, будет создана партиция, разделенная методом Range по кварталам. В своя очередь, каждая партиция по кварталу будет содержать партиции по категориям. Это будет сделано с целью будущего анализа по категориям за кварталы, для того, чтобы выявить есть ли какие-то сезонные тренды на ту или иную категорию товаров.

# Бизнес процессы

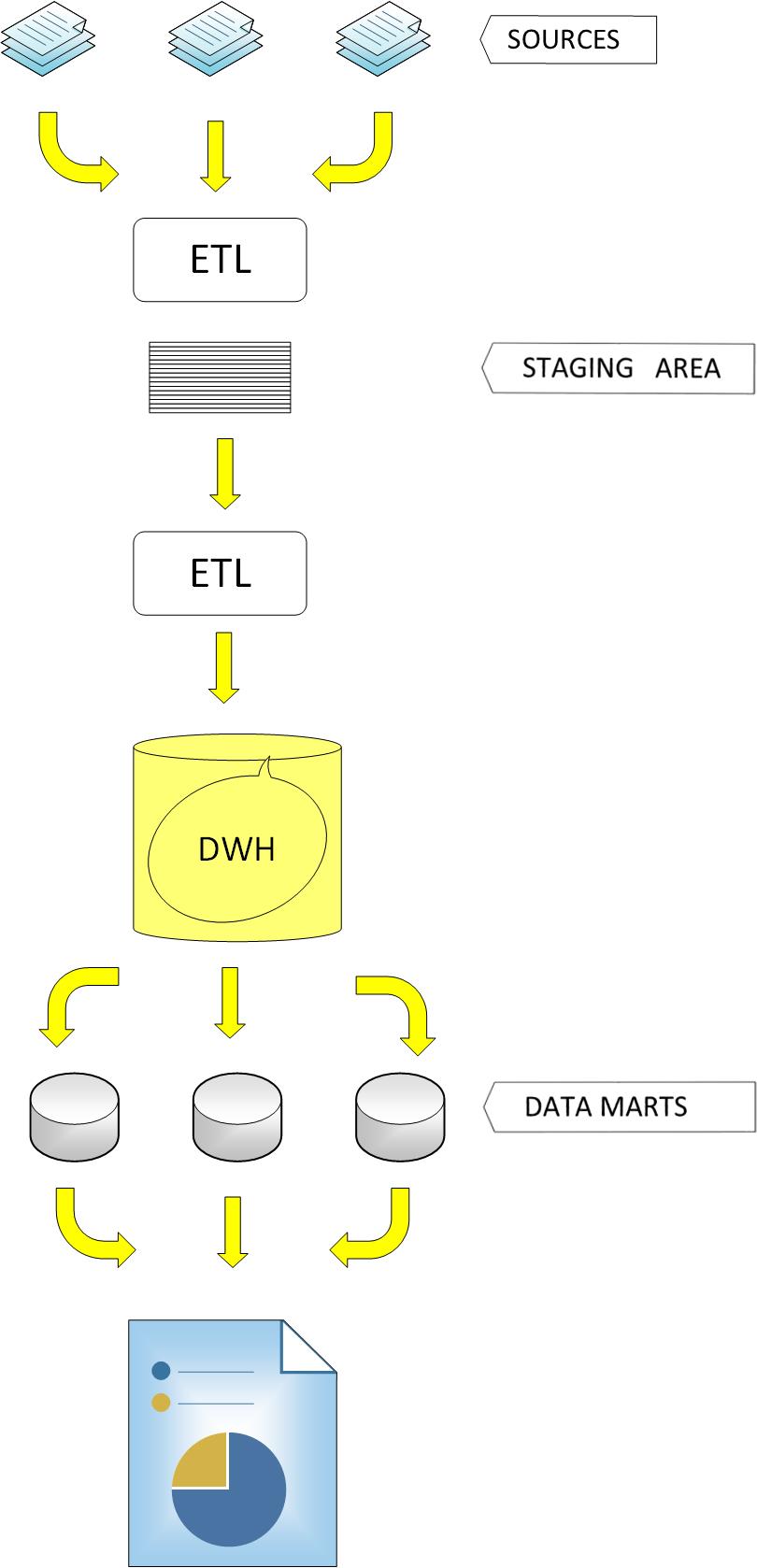
(добавить главу про примеры отчетов)

# Модель данных

## описание типов измерений

В данном хранилище данных использовано 9 измерений, 3 из которых имееют тип SCD2 type с атрибутами “Start\_dt”, “End\_dt”, “Is\_current”. В данном хранилище данных использовано измерение каленадрного типа DIM\_TIME\_DAY, позволяющее нам делать анализ продаж в разрезе лет, месяцев и даже дней.

## слои хранилища данных



Picture 3 Data\_flow

Изначально наши данные представляют собой набор фалов, содержащие сырые данные. Для загрузки их в наше хранилище данных необходимо произвести процесс трансофрмации и обработки данных и поместить их на Staging Area. После этого шага, если нам еще необходимо производить процесс трансформации и обработки, то мы выполняем этот процесс, если нет, то загружаем данные в наше хранилище. На базе нашего хранилища разрабыватся различные витрины данных, которые позволят делать качественные отчеты.

| REVISION HISTORY | | | | | |
| --- | --- | --- | --- | --- | --- |
| Ver. | Description of Change | Author | Date | Approved | |
| Name | Effective Date |
| 3.1 | Изменена тема шаблон | Alina Makarets | 16-NOV-2017 |  | 16-NOV-2017 |
|  |  |  |  |  |  |
|  |  |  |  |  |  |