

Подготавливаем ландшафт  
для внедрения любой BI-системы  
на примере Visiology

Облегчаем будущие миграции  
с помощью Loginom  
и компонентов BI2BUSINESS



**Евгений Стучалкин**, опыт разработки бизнес-аналитики - 9 лет

Эксперт по организации единого аналитического ландшафта в бизнесе. Разработчик алгоритма автоматической генерации моделей данных для BI-приложений.

# Где найти аналог Power BI, Qlik, Tableau?

**Можно выдохнуть. Их нет.**

Даже наиболее продвинутые российские BI-системы подразумевают жесткие компромиссы.

Поэтому нет смысла стремиться заменить все одной коробкой.

Нужно научиться выстраивать процессы создания бизнес аналитики.



# Навыки архитектора аналитики становятся критическими

Даже если вы работаете в компании СМБ-сегмента



Специфика бизнес-аналитики:

- Много источников;
- Требуется преобразование данных по бизнес-логике;
- Много таблиц;
- Непредсказуемые изменения;
- Отсутствие архитектуры превращает жизнь в боль;

# Ваш аналитический стек должен быть вездеходным

И простым в обслуживании

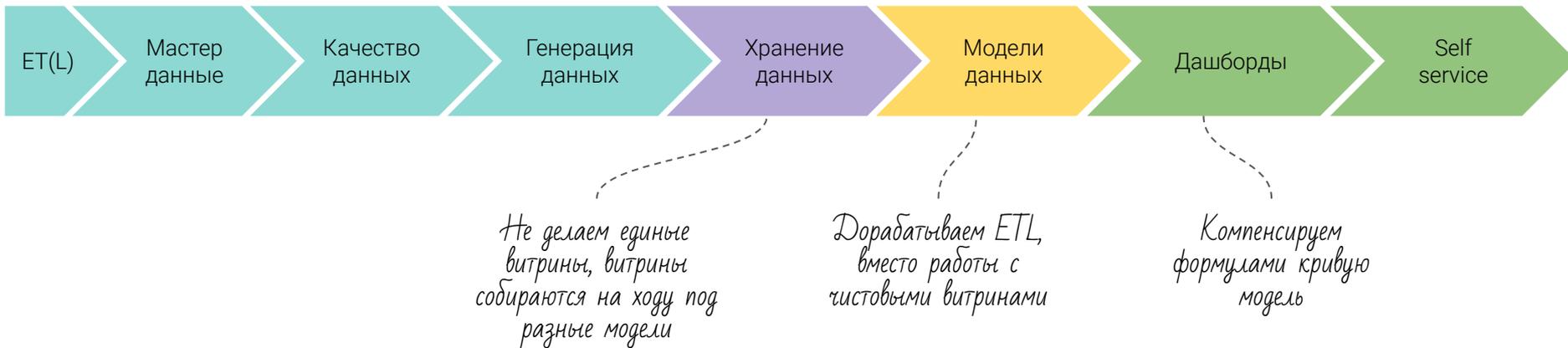


✗ Нет



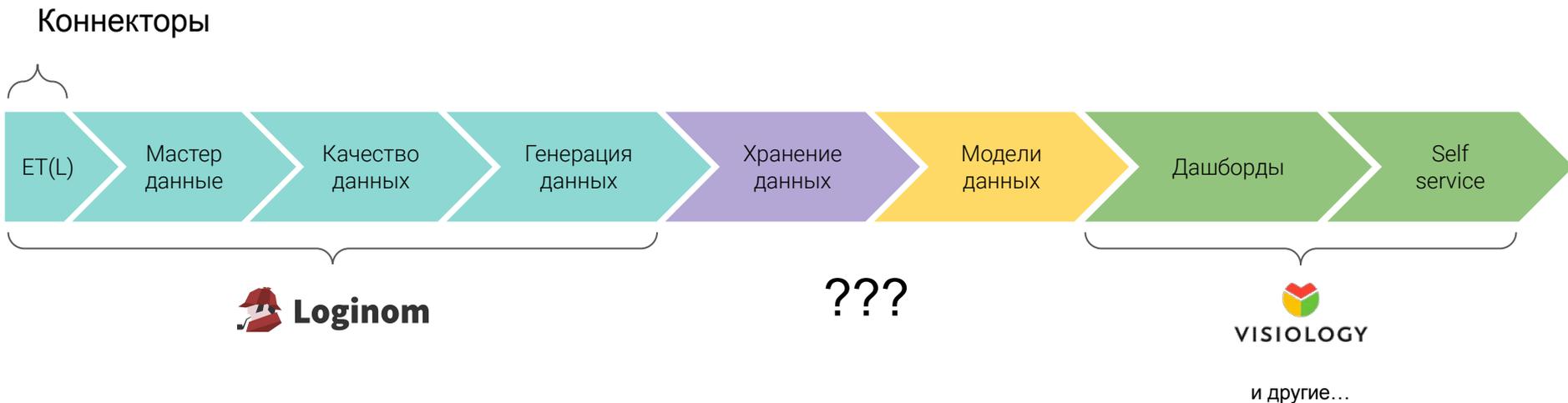
✓ Да

# Наличие единого инструмента не гарантирует правильной архитектуры



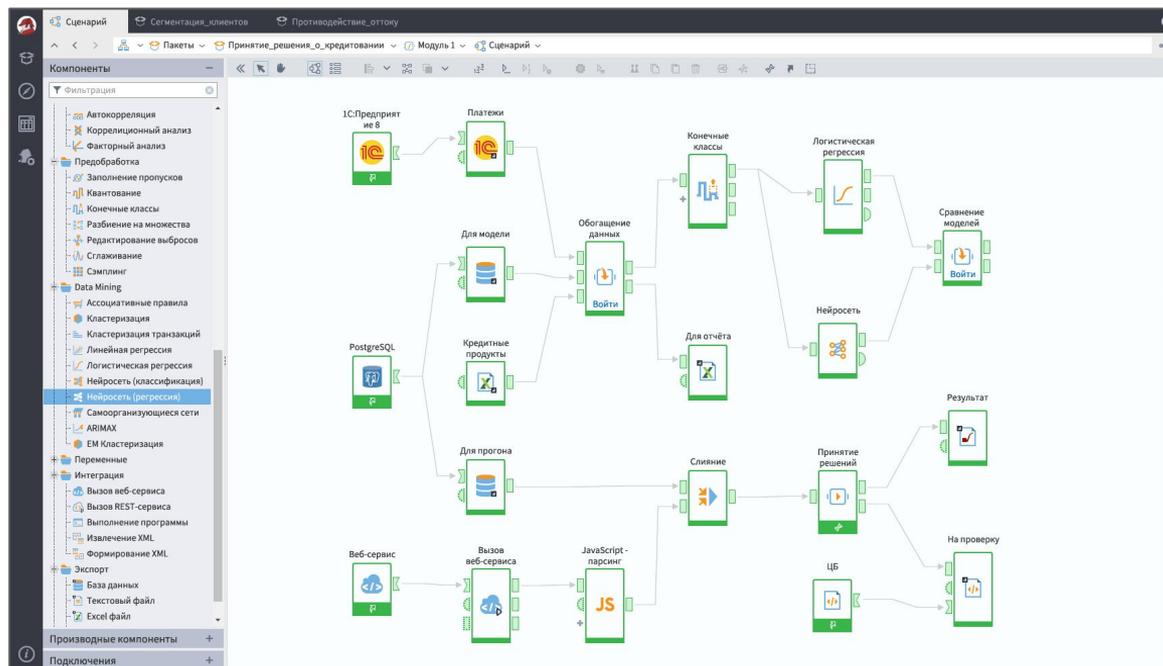
# Четкое ролевое разделение систем - залог успеха

Нужно забирать лучшее из нескольких компонентов



# LogiNot - система для визуального проектирования процессов обработки данных

А также Low-code Data Science и разработки переиспользуемых модулей



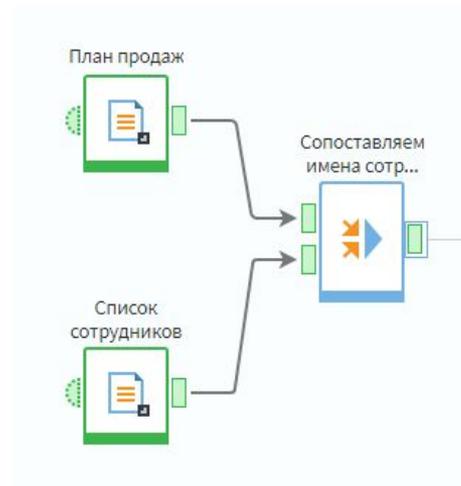
## 8 критериев правильной аналитической витрины

- Нормализация ключей;
- Наличие первичного ключа, уникального для каждой строки;
- Даты приведены к формату первого дня периода;
- Исчерпывающий нейминг полей (не Date, а Sale\_date. Не Sales, Sales\_sum\_VAT);
- Никаких кростабличных вычислений;
- Предрасчитанные поля ( $Cost * Amount$  as Sales\_sum\_VAT);
- Денормализация (если справочник используется как свойство только одной таблицы);
- Уникальный префикс витрины в названии полей;

# Все ключевые поля должны быть приведены к одинаковой форме

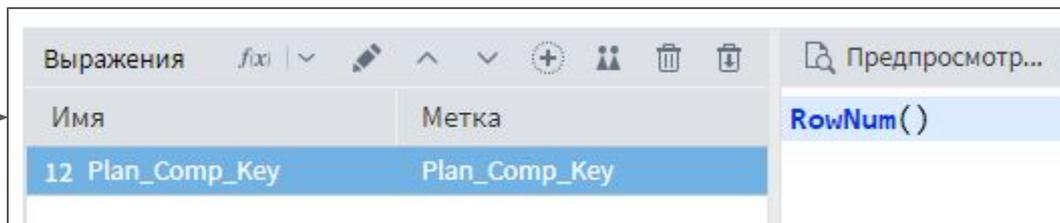


#	ab Сотрудник	31 Месяц	12 План продаж	12 User_ID
1	Лазуткин Святослав	01.01.2021, 00:00	900 000	1
2	Лазуткин Святослав	01.02.2021, 00:00	1 500 000	1
3	Лазуткин Святослав	01.03.2021, 00:00	2 100 000	1
4	Лазуткин Святослав	01.04.2021, 00:00	2 400 000	1
5	Лазуткин Святослав	01.05.2021, 00:00	2 430 000	1
6	Лазуткин Святослав	01.06.2021, 00:00	3 000 000	1
7	Лазуткин Святослав	01.07.2021, 00:00	2 700 000	1
8	Лазуткин Святослав	01.08.2021, 00:00	2 670 000	1
9	Лазуткин Святослав	01.09.2021, 00:00	3 600 000	1



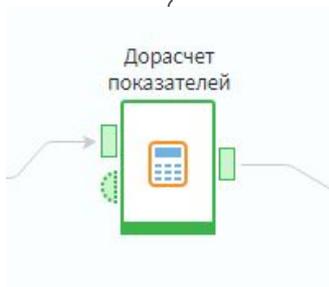
# Обязательный первичный ключ. Если отсутствует, то заменяется нумерацией строк, или составным ключом

Ключ на базе нумерации строк



Имя	Метка
12 Plan_Comp_Key	Plan_Comp_Key

Предпросмотр...  
RowNum()



Составной ключ на базе других полей

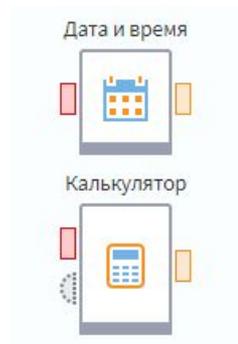


Имя	Метка
ab Sales_Comp_Key	Sales_Comp_Key

Предпросмотр... | AND OR  
Sale\_ID+"\_" + Position\_ID

## Даты приведены к первому дню периода

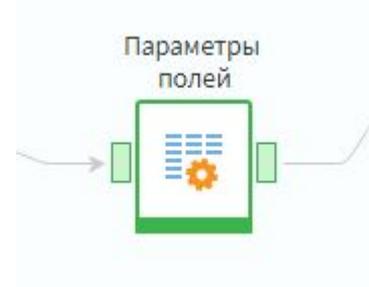
Пример	Как надо
Январь 2019	01.01.2019
Q2 2020	01.04.2020
2021	01.01.2021
01.01.2022 12:56:43	01.01.2022



31	AddDay (Дата, Количество)
31	AddMonth (Дата, Количество)
31	AddQuarter (Дата, Количество)
31	AddWeek (Дата, Количество)
31	AddYear (Дата, Количество)
ab	DateTimeToStr (Дата_время [, Формат_да...
ab	DateToStr (Дата [, Формат])
12	Day (Дата)
12	DayOfWeek (Дата)
12	DaysBetween (Дата1, Дата2)
31	EncodeDate (Год, Месяц, День)
31	EncodeDateTime (Год, Месяц, День, Часы, ...
31	EncodeTime (Часы, Минуты, Секунды)
12	Hour (ДатаВремя)
12	Minute (ДатаВремя)
12	Month (Дата)

# Исчерпывающий нейминг полей

Как не надо	Как надо
Date	Sales_date
Sum	Sales_sum_VAT



Метка	Имя	Вид данных	Назначение
ab Сотрудник	User	Дискретный	Не задано
12 Plan_Comp_Key	Plan_Comp_Key	Непрерывный	Не задано
31 Месяц	Plan_month	Дискретный	Не задано
12 План продаж	Sales_plan	Непрерывный	Не задано
12 User_ID	User_ID	Непрерывный	Не задано

# Никаких кросстабличных вычислений!

Большая ошибка

Amount	PriceKey
3	a
4	b
5	c

PriceKey	Price
a	100
b	200
c	300

Приемлемый формат

Amount	PriceKey	Price
3	a	100
4	b	200
5	c	300

sum(Amount\*Price)

# Предрасчитывайте все, что можно предрасчитать

Приемлемый формат

Amount	PriceKey	Price
3	a	100
4	b	200
5	c	300

sum(Amount\*Price)

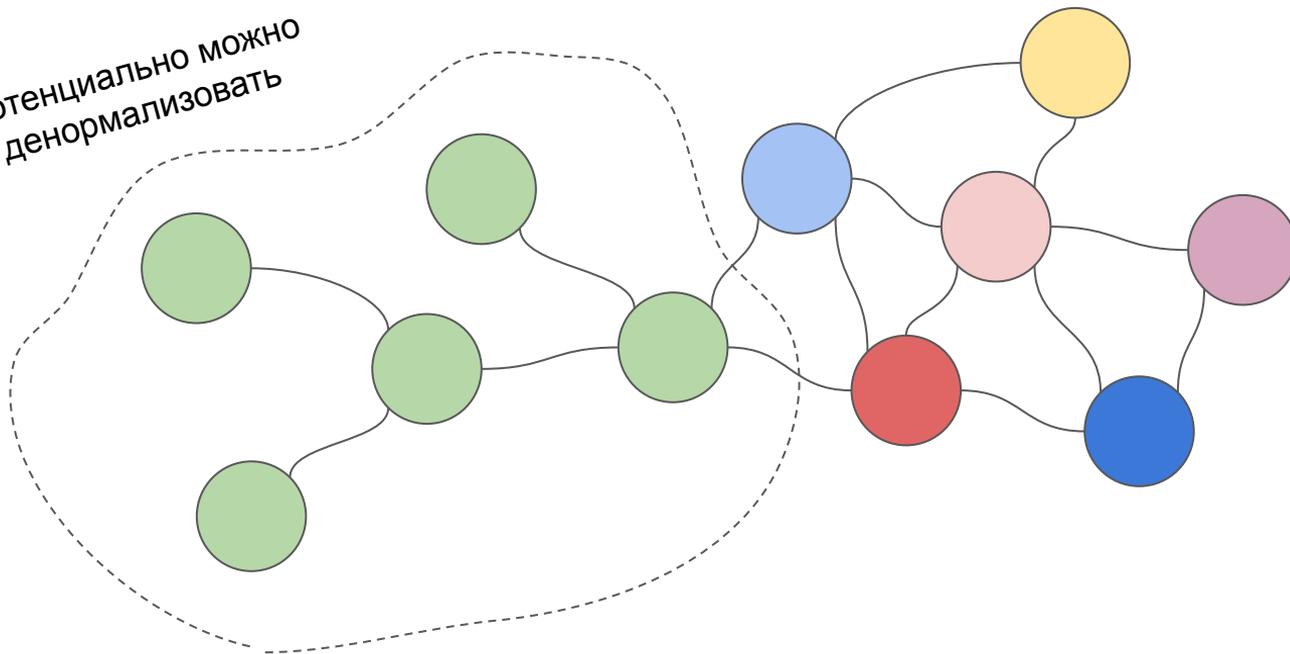
Рекомендуемый формат

Amount	PriceKey	Price	Sales_sum
3	a	100	300
4	b	200	800
5	c	300	1500

sum(Sales\_sum)

## По возможности, денормализуйте витрины

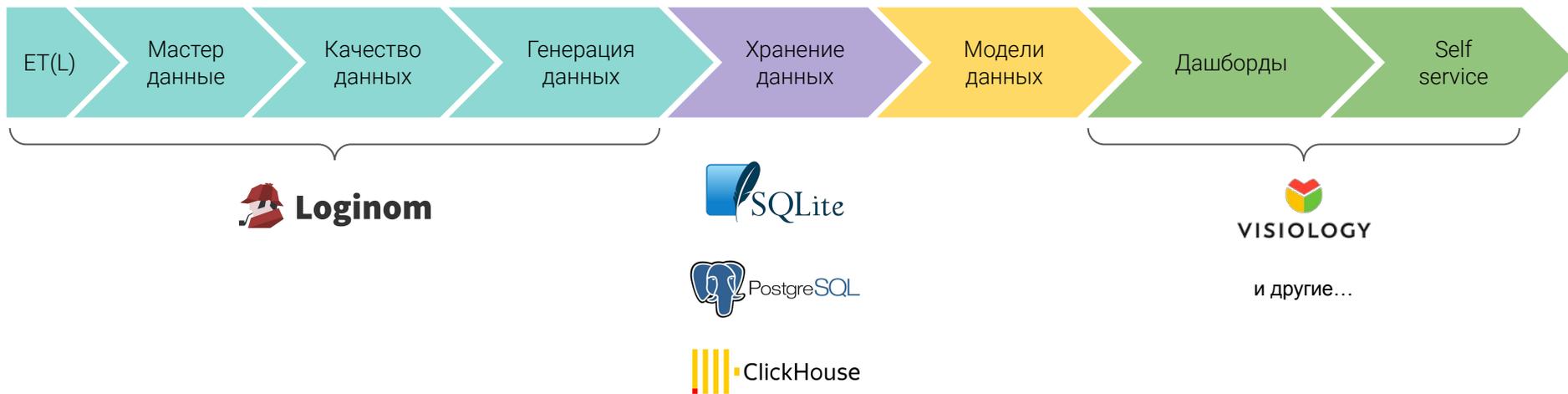
Потенциально можно  
денормализовать



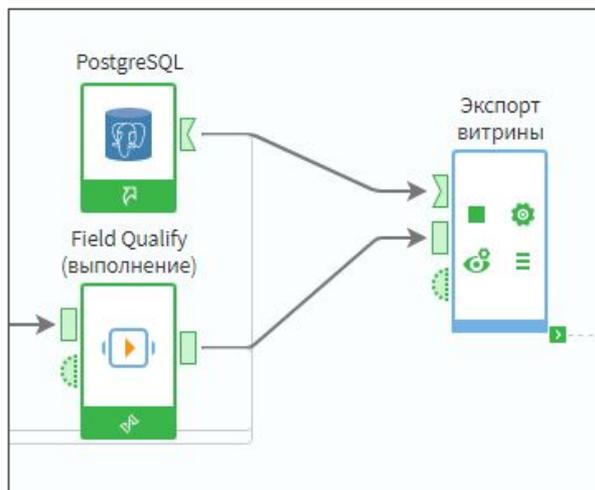
## Уникальный префикс в имени полей витрины

	<u>Data Output</u>	Explain	Messages	Notifications				
	SLS__Sale_ID bigint	SLS__Position_ID bigint	SLS__Sales_Sum_VAT double precision	SLS__Cost_Sum_VAT double precision	SLS__Sales_Comp_Key character varying (50)	SLS__Amount bigint		
1	1	1	2608.59762210402	1827.96705752499	1_1	9		
2	1	2	1276.50090350212	485.0066426024	1_2	4		
3	1	3	11168.6413564715	2669.85183368442	1_3	11		
4	1	4	541.80724930018	446.92592721666	1_4	2		
5	1	5	208879.441117173	43911.513119190604	1_5	9		
6	1	6	2038.8635656796	529.9673091853	1_6	10		
7	2	1	248.71738933492497	289.81927138566897	2_1	3		
8	3	1	94465.95759550101	17757.10263662	3_1	10		
9	3	2	5929.0742685087	1691.38834761456	3_2	6		
10	3	3	1303.48987311124	1080.39785156026	3_3	2		

# Где взять СУБД для хранения данных?

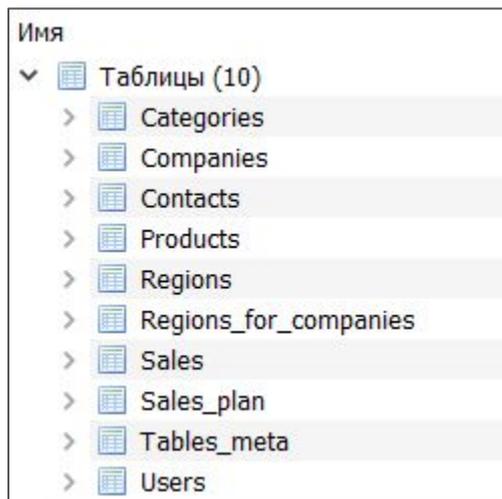


# Экспорт в БД осуществляется с помощью спец. узлов

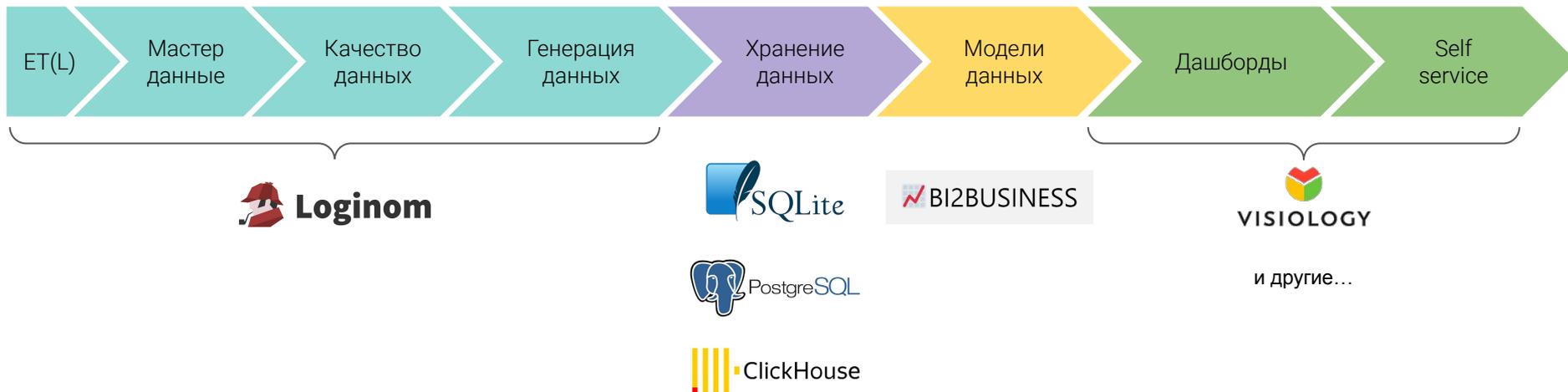


База данных	localhost:5432:BI
Имя таблицы	Sales_plan
Тип экспорта	
<input type="radio"/> Дополнить таблицу данными	
<input checked="" type="radio"/> Очистить таблицу и заполнить данными	
<input type="radio"/> Удалить записи по ключевым полям	
<input type="radio"/> Удалить записи по ключевым полям и дополнить таблицу данными	
<input type="radio"/> Обновить существующие записи в таблице	

# Мы получаем набор витрин, готовых к визуализации в BI

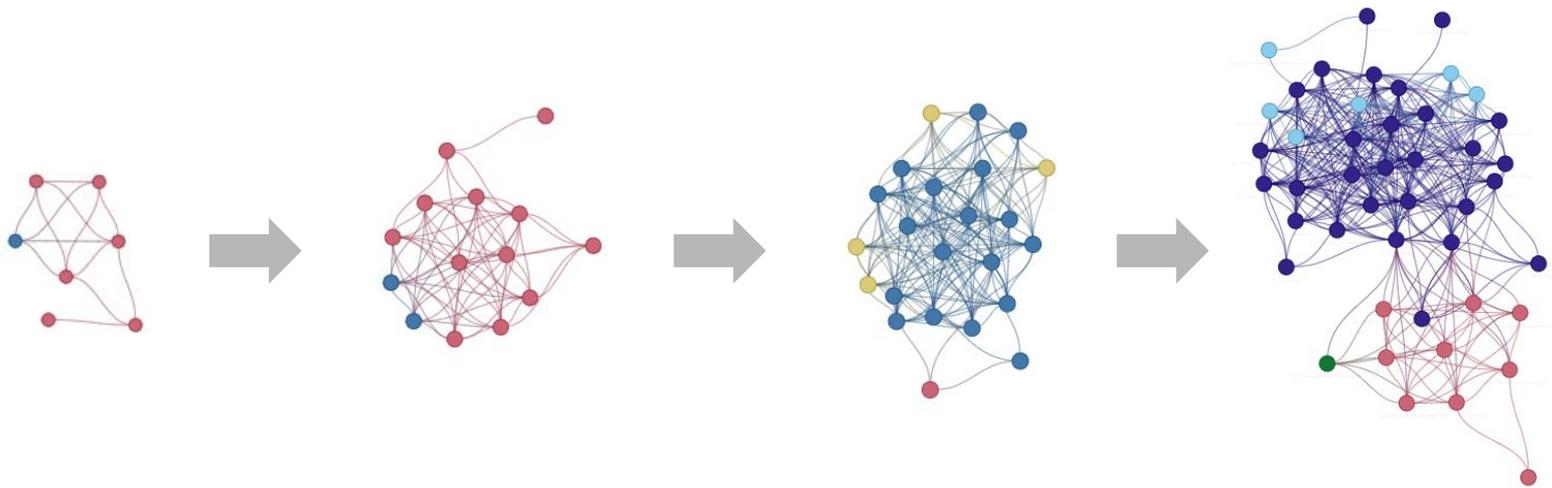


# Как построить модель данных?

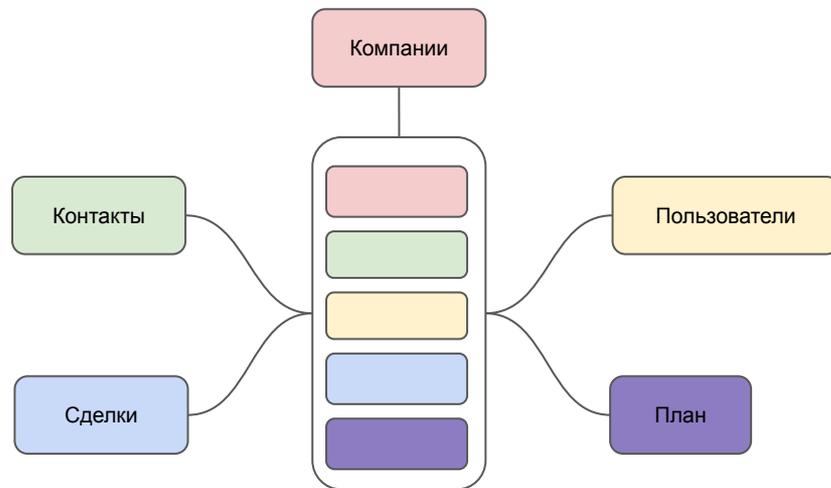
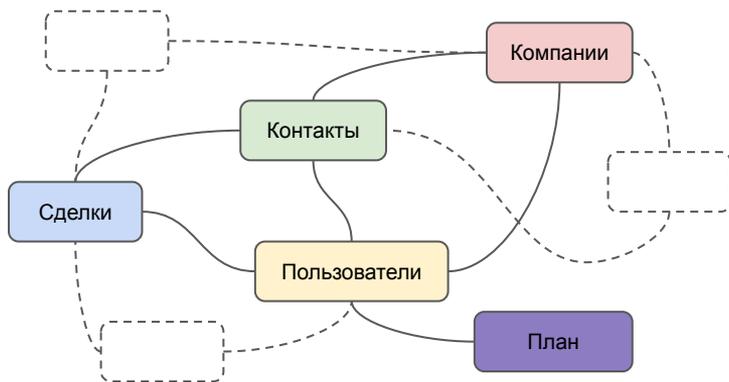


# Зачем нужна автоматическая генерация моделей?

Вам понадобятся нерушимые стандарты для управления усложняющимся ландшафтом



# Звезда - универсальная топология моделей данных для аналитики



# Есть несколько сценариев организации звезды

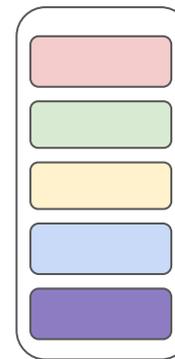
Ключи  
в центре, данные в лучах



Меры и ключи  
в центре, справочники в лучах



Все данные  
в одной таблице



# Особенности топологии Звезда

## Плюсы:

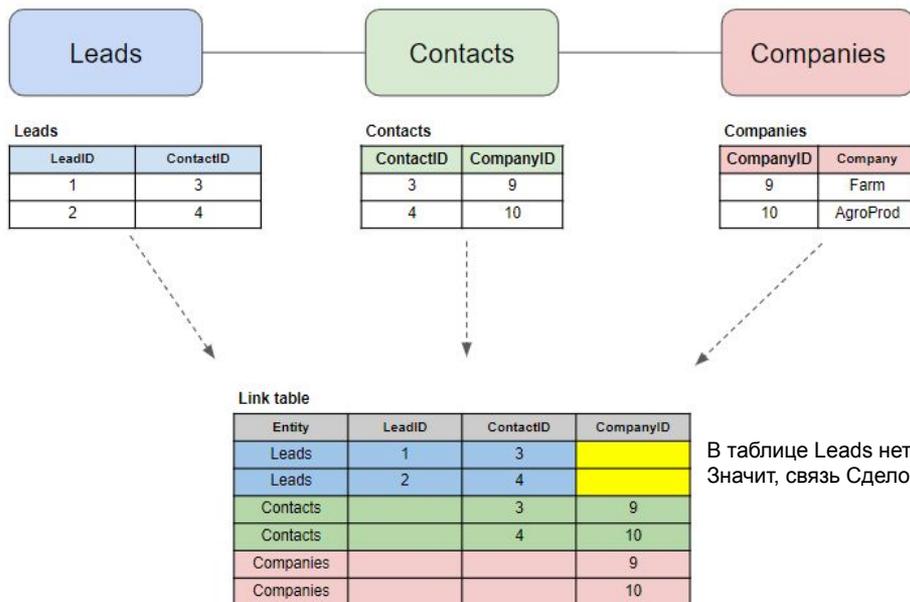
- Простота эксплуатации;
- Единообразие формирования;
- Гибкость;

## Минусы:

- Избыточность (*решается: генерация моделей, агрегированные витрины, фильтры витрин, DQ*);
- Сложность разработки;
- Сложность поддержки;

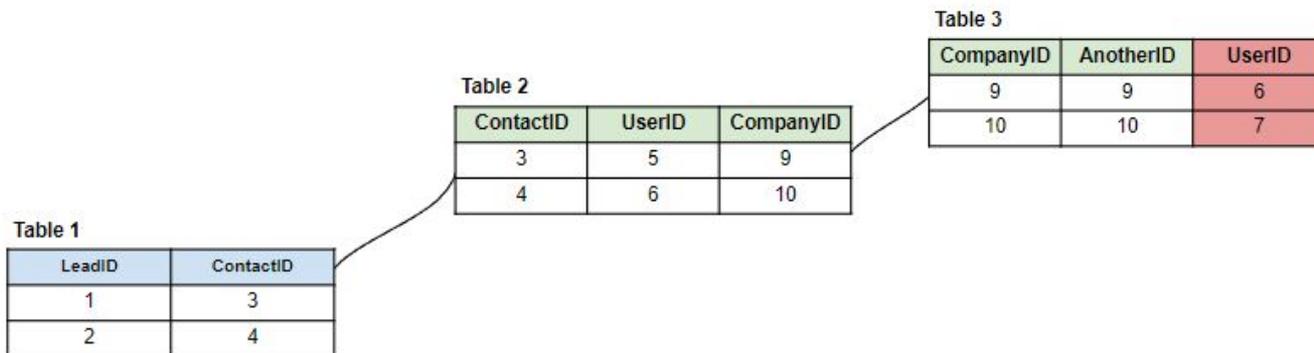
# В Звезде не достаточно использовать собственные ключевые поля таблицы

Каждый блок связи должен содержать полный набор ключевых полей

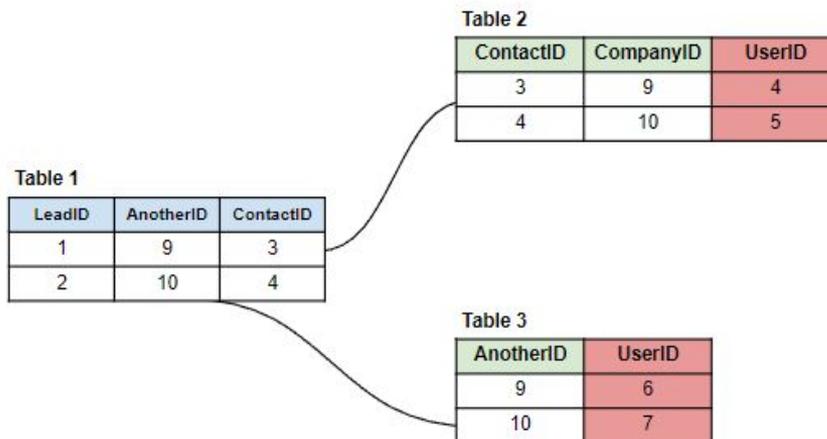


В таблице Leads нет поля CompanyID.  
Значит, связь Сделок и Компаний работать не будет

# Конфликт #1: ключевое поле добавлено ранее



## Конфликт #2: ключевое в двух присоединяемых таблицах



# Для генерации модели необходимы метаданные

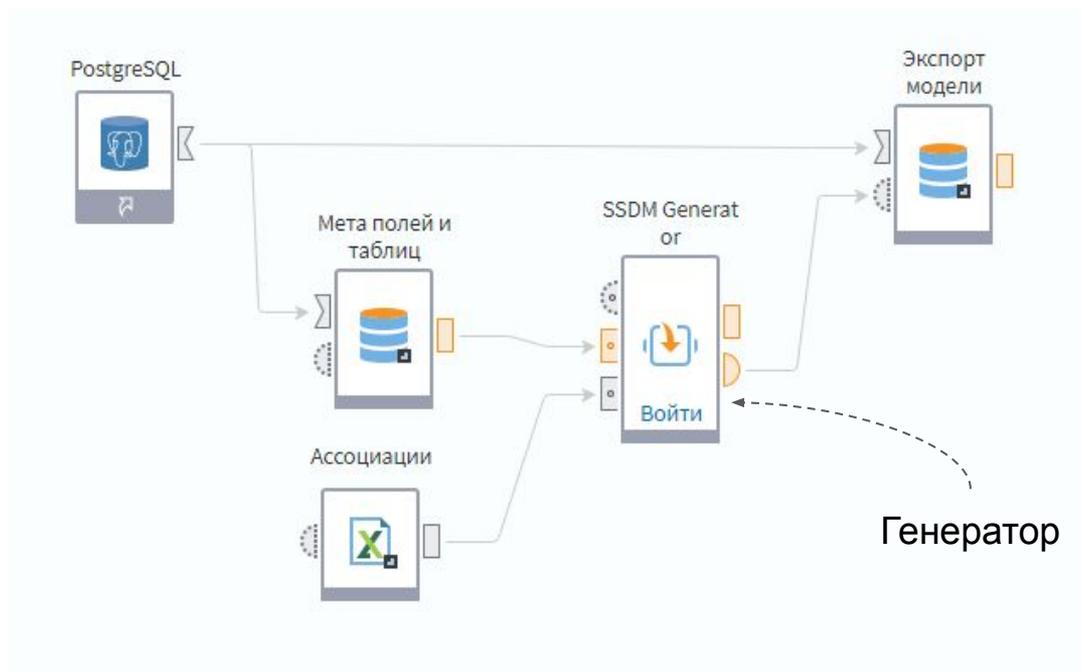
## Список всех полей

#	ab TableName	ab PrimaryKey	ab Name
1	Contacts	CON__Contact_ID	CON__Contact_ID
2	Contacts	CON__Contact_ID	CON__Contact_Name
3	Contacts	CON__Contact_ID	CON__Company_ID
4	Contacts	CON__Contact_ID	CON__Phone
5	Companies	COM__Company_ID	COM__Company_Name
6	Companies	COM__Company_ID	COM__Company_ID
7	Companies	COM__Company_ID	COM__Region_ID
8	Companies	COM__Company_ID	COM__UserID
9	Categories	CAT__Category_ID	CAT__Category_ID
10	Categories	CAT__Category_ID	CAT__Category_Name
11	Products	PRD__Product_ID	PRD__Product_ID
12	Products	PRD__Product_ID	PRD__Category_ID
13	Products	PRD__Product_ID	PRD__Price
14	Products	PRD__Product_ID	PRD__Cost
41	Products	PRD__Product_ID	PRD__Product_Name

## Список ассоциаций

#	ab Поле	ab Ассоциация
8	REG__Region_ID	REG__Region_ID
9	USR__Region_ID	REG__Region_ID
10	USR__User_ID	USR__User_ID
11	PLN__User_ID	USR__User_ID
12	SLS__User_ID	USR__User_ID
13	COM__UserID	USR__User_ID
14	SLS__Sales_Comp_Key	SLS__Sales_Comp_Key
15	REGC__Region_ID	REGC__Region_ID
16	COM__Region_ID	REGC__Region_ID
17	PRD__Product_ID	PRD__Product_ID
18	SLS__Product_ID	PRD__Product_ID
19	SLS__Amount	<null>
20	SLS__Sales_Sum_VAT	<null>
21	SLS__Cost_Sum_VAT	<null>
22	PLN__Sales_plan	<null>

# Мета-данные подаются в генератор, на выходе SQL-запрос



# BIG DEMO

# Модель Звезда обеспечивает простую и одинаковую логику выражений в разных инструментах



Добавить показатель

Расчетные показатели

: Выручка с НДС

Добавить расчетный показатель

Добавить измерение

Строки

Фильтр DATE\_TYPE (Тип)

Фильтр SLS\_Status\_ID (



```
Sum({<SLS_Status_ID={2}, [DATE_TYPE]= {"SLS_CloseDate"}>} [SLS_Sales_Sum_VAT])
```



```
=ЕСЛИ( И([@SLS_Status_ID]=2;[@DATE_TYPE]="SLS_CloseDate");[@SLS_Sales_Sum_VAT];0)
```

## Когда вам подойдет Звезда?

- Вы аналитик, на котором висит все;
- Вы внедряете аналитику проектно;
- Когда вам нужно внедрить управляемый Self-service;
- Когда вам нужна консолидация всей аналитики компании;

# Что делать дальше

- Добавляться в группу телеграм <https://t.me/bi2business>
- Записываться на консультацию:
  - Telegram - @stuchalkin;
  - email - es@bi2business.ru

Спасибо за внимание!  
Задайте ваши вопросы