

# Архитектура данных в стиле Agile или проектирование архитектуры данных с примерами компании Магнит

Гаврилков Григорий

Руководитель отдела системного анализа



## Гаврилков Григорий

Руководитель отдела  
системного анализа

- В IT 20 лет, специализируюсь в области управления данными (DWH, BI, аналитические системы и т.д.). Связан с системным анализом

- Последние 8 лет работал на руководящих позициях

- Уже 3 года представляю Магнит в роли руководителя системного анализа, до этого работал в ТОП-5 финансовых компаниях

## О компании



**№1**

ритейлер по  
количеству  
магазинов

**29 165**

магазинов

**>360 000**

сотрудников

**45**

распределительных  
центров

**20**

производств



**>260**

активных IT-проектов в  
работе

**4000**

человек работает в  
IT-блоке Magnit Tech

**>800**

информационных  
систем и сервисов

**>1 ПБ**

данных в платформе  
данных

# О чем сегодня поговорим

**Архитектура  
данных** как часть  
корпоративной  
архитектуры

**Моделирование  
и жизненный цикл  
данных**

**Метрики для  
оценки качества  
архитектуры**

# Зачем системным аналитикам Архитектура данных?

01

Данные — это актив компании (наравне с финансовым) и требуют трепетного отношения и управления

02

Чтобы воспользоваться данными, они должны быть описаны и подчиняться требованиям определенной архитектуры

03

Архитекторов в компаниях мало, а архитекторов данных еще меньше или нет вообще.

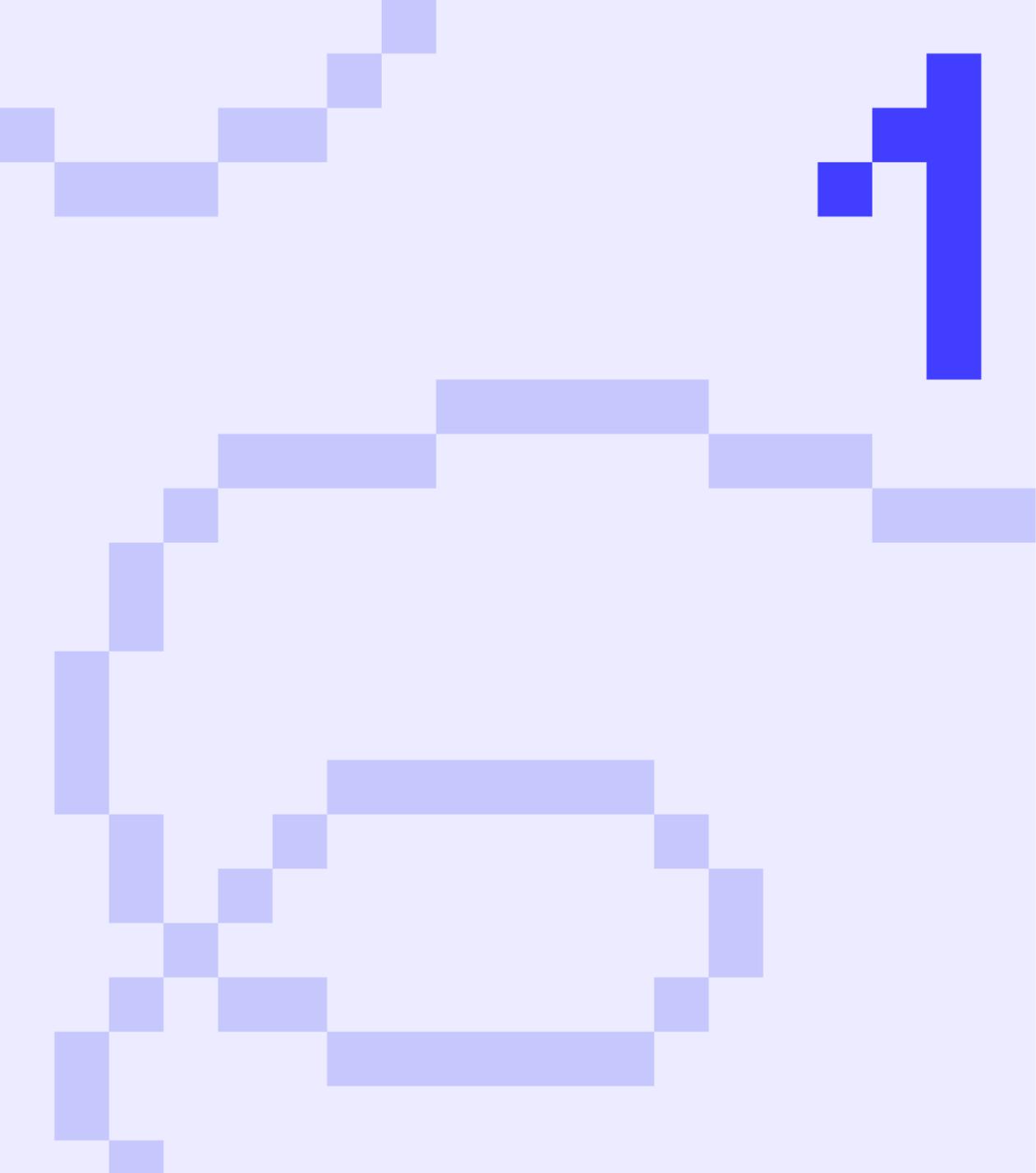
Поэтому, проработка ЛМД\* ложится на плечи системных аналитиков, так как они проводят анализ и пишут требования к развитию системы

**Архитектура  
данных** как часть  
корпоративной  
архитектуры

**Моделирование**  
и жизненный цикл  
данных

**Метрики для**  
оценки качества  
архитектуры





# Что такое данные?

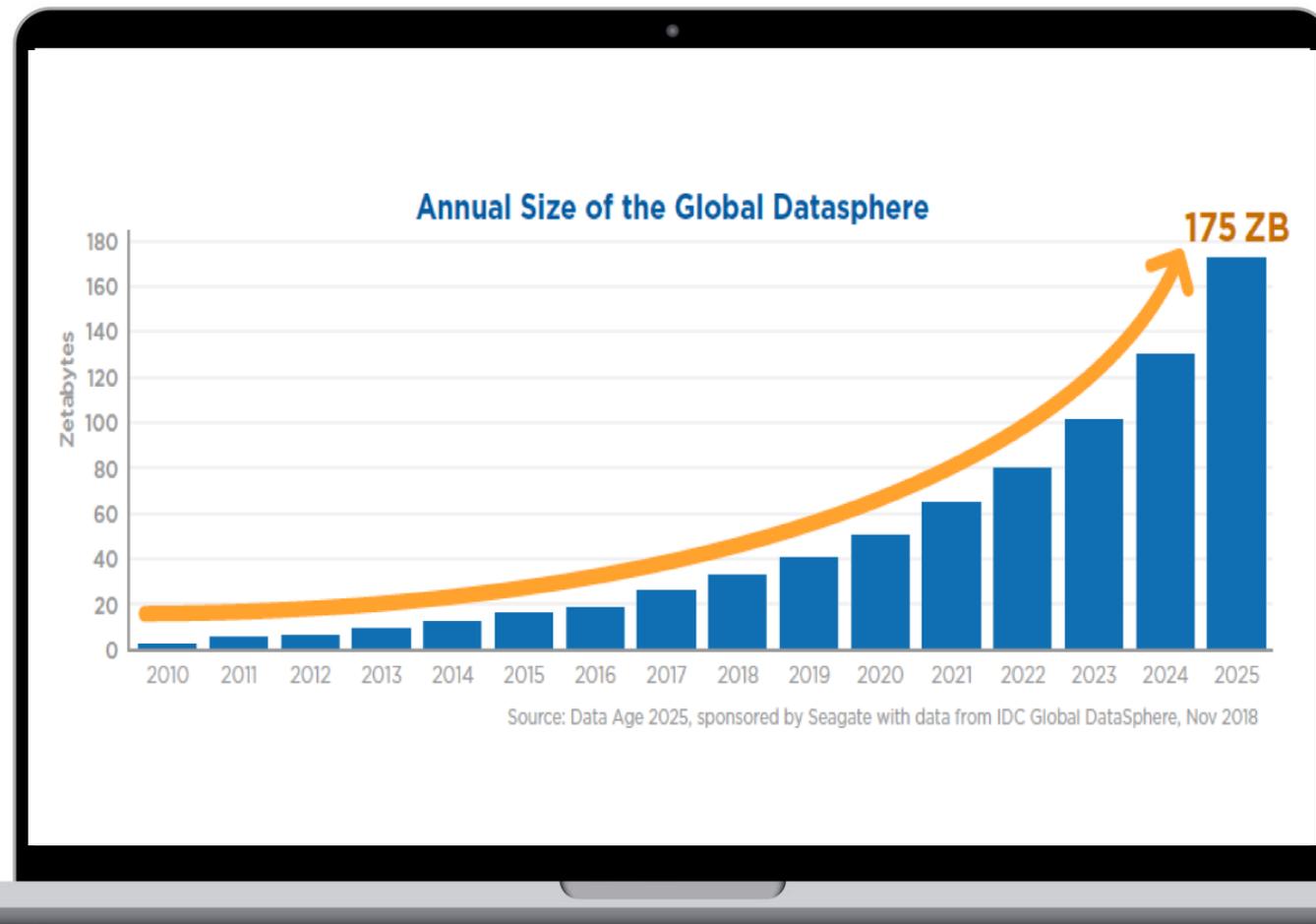
**Данные – это информация, сохраненная в цифровой форме**

*Данные — это одновременно интерпретация представляемых ими предметов и, сами по себе, предметы, подлежащие интерпретации (Sebastian-Coleman, 2013). Иными словами, в отрыве от контекста данные становятся бессмысленными.*

# Данных очень много

В платформе данных Магнит < 1 ПБ.

Для получения выгоды от такого объема, данными необходимо управлять.



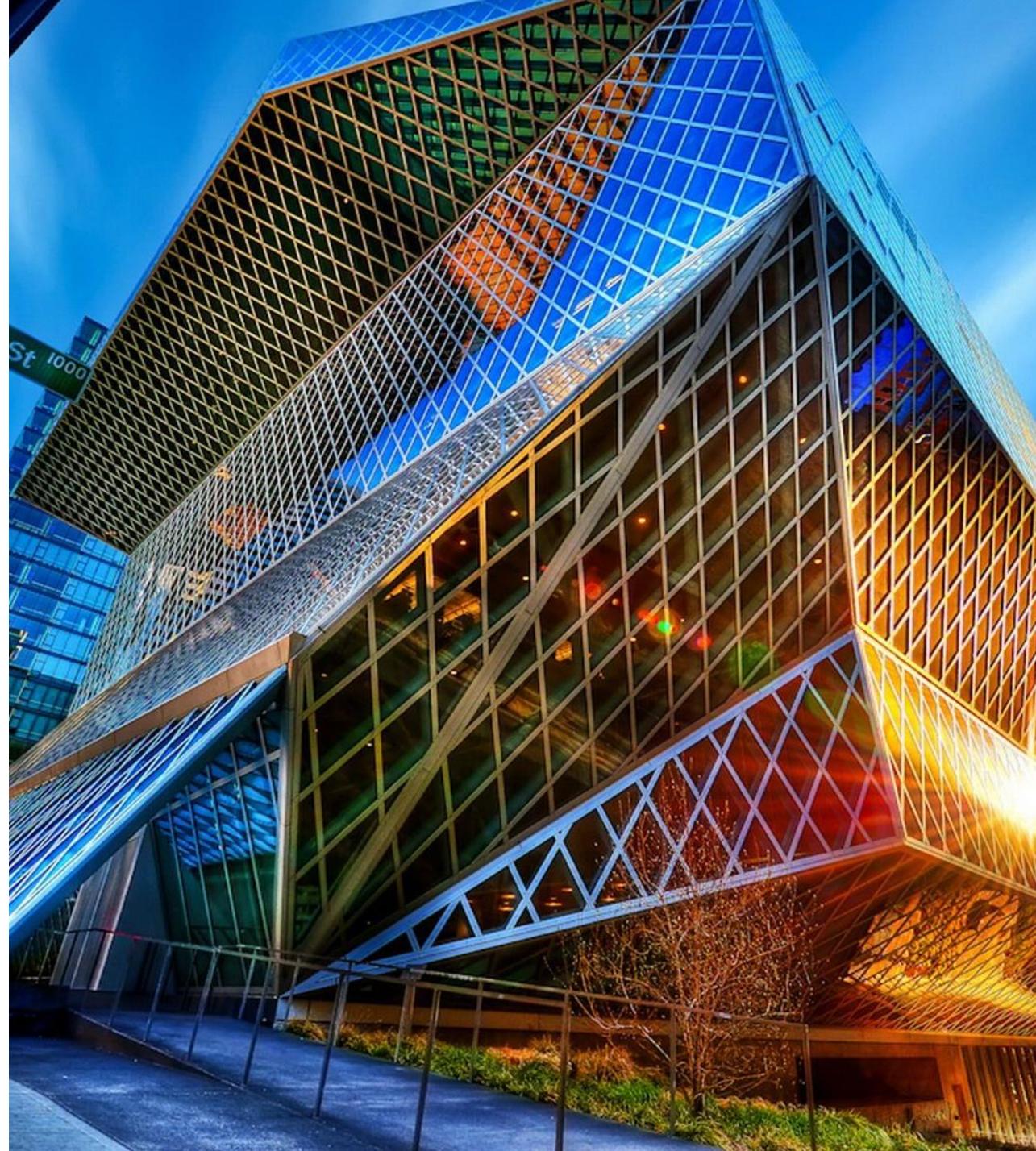
# Управление данными

**Управление данными** подразумевает эффективную, экономичную и безопасную организацию процессов сбора, хранения и использования данных.

Цель — обеспечить точность, последовательность, доступность и безопасность данных в любое время.

# Что такое архитектура?

Архитектурой принято называть как искусство и науку о проектировании и строительстве чего-либо, так и результаты этой деятельности — то есть сами строения



# А что если ...?!

- Компания не знает системы через которые распространяются данные?
- Компания не знает свои данные?

# Что такое архитектура данных?

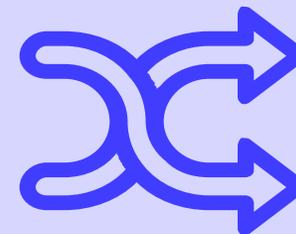
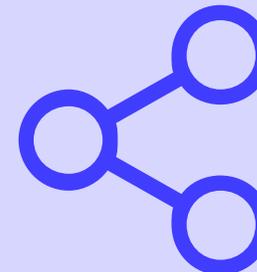
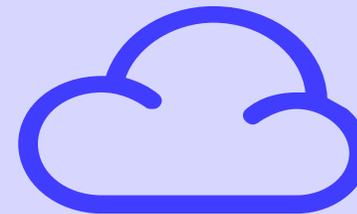
Архитектура данных создана для поддержания знания компании о данных и системах, через которые эти данные распространяются. Такие знания позволяют компании управлять данными как активом и повышать получаемую от них пользу



**Описание архитектуры данных** должны включать как **корпоративные модели данных (КМД)**, так и описания **потоков данных**

# Потоки данных

**Потоки данных.** Отражают перемещение данных с целью использования в бизнес-процессах, а также отдельными техническими компонентами

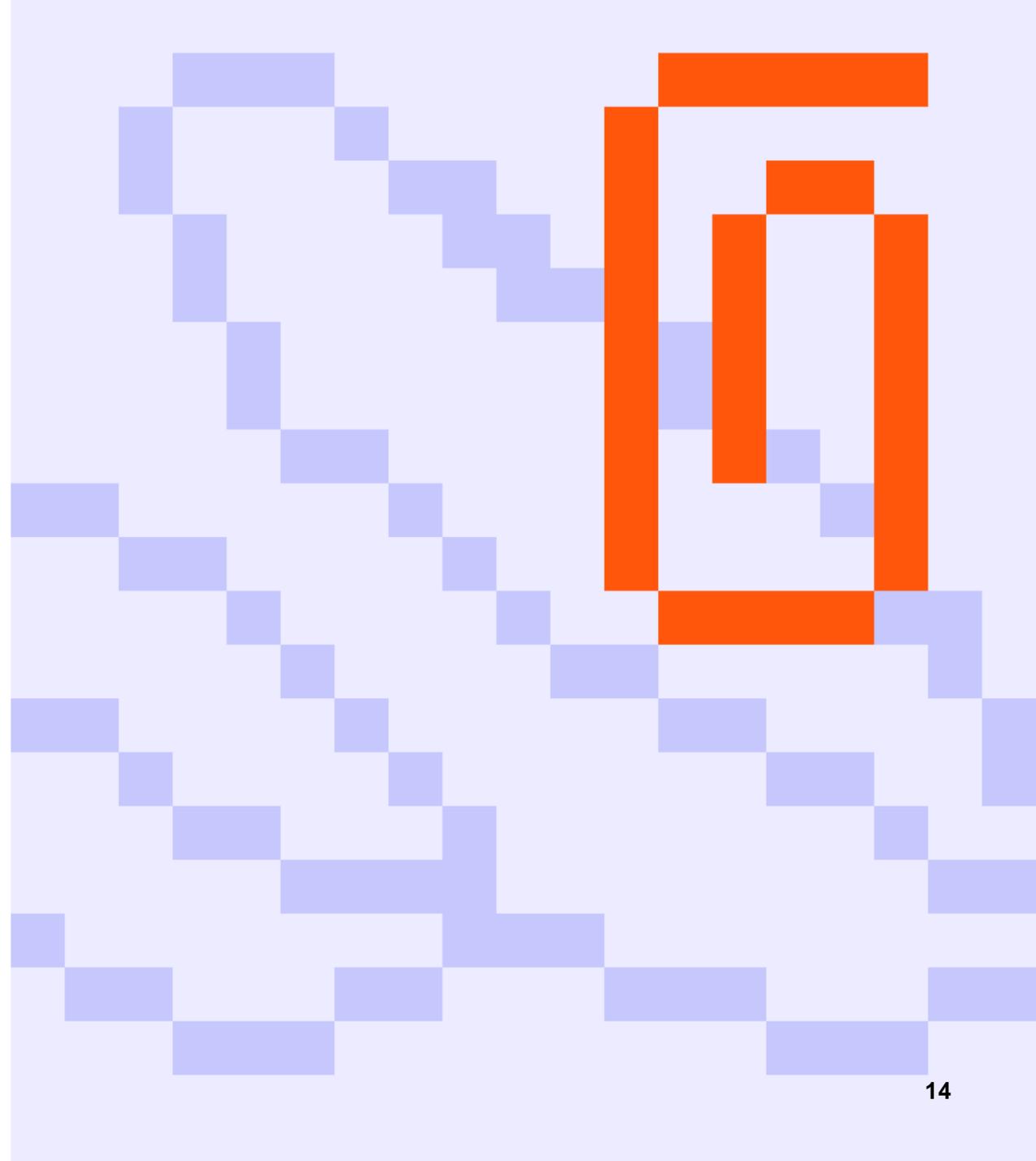


# Важно!

Создав архитектуру данных, ее необходимо развивать и поддерживать в актуальном состоянии. Т.е. все изменения должны проходить обязательно с отражением в архитектуре.

Чем больше компания зависит от данных, тем важнее вопрос по проработанной и актуальной архитектуре.

- **Вести архитектуру данных обязательно**



**Архитектура  
данных** как часть  
корпоративной  
архитектуры

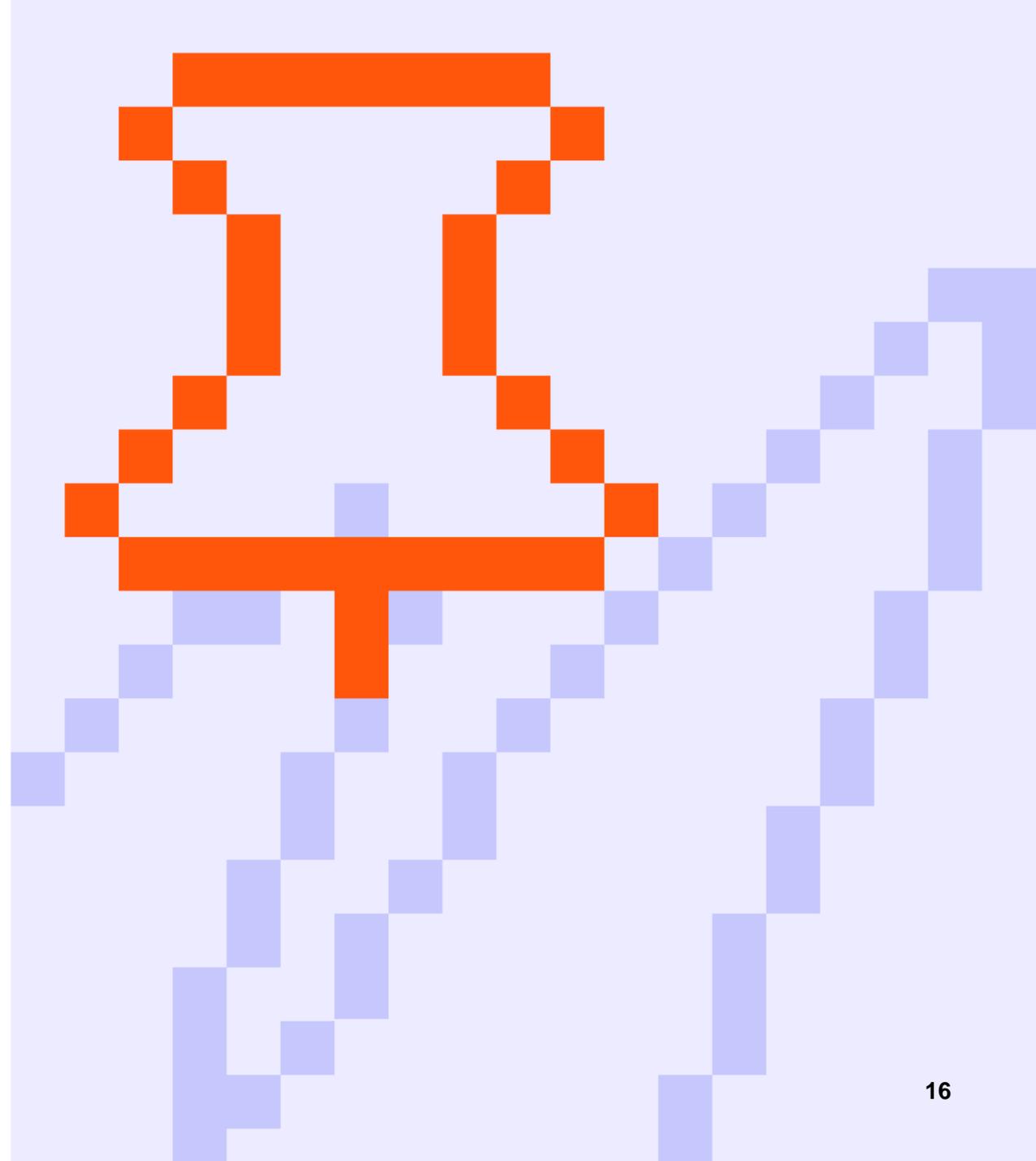
**Моделирование  
и жизненный цикл  
данных**

**Метрики для  
оценки качества  
архитектуры**



# А что если ...?!

- Нет понимания о концептах данных в компании?
- При реализации проектов отсутствует информация о доступных блоках данных?
- Отсутствует понимание о связи данных друг с другом?



# Корпоративная модель данных

**Корпоративная модель данных** представляет собой целостную модель данных, отражающую единый согласованный взгляд на данные в масштабах всей организации.

**1** Концептуальная модель данных

**2** Логическая модель данных

**3** Физическая модель данных

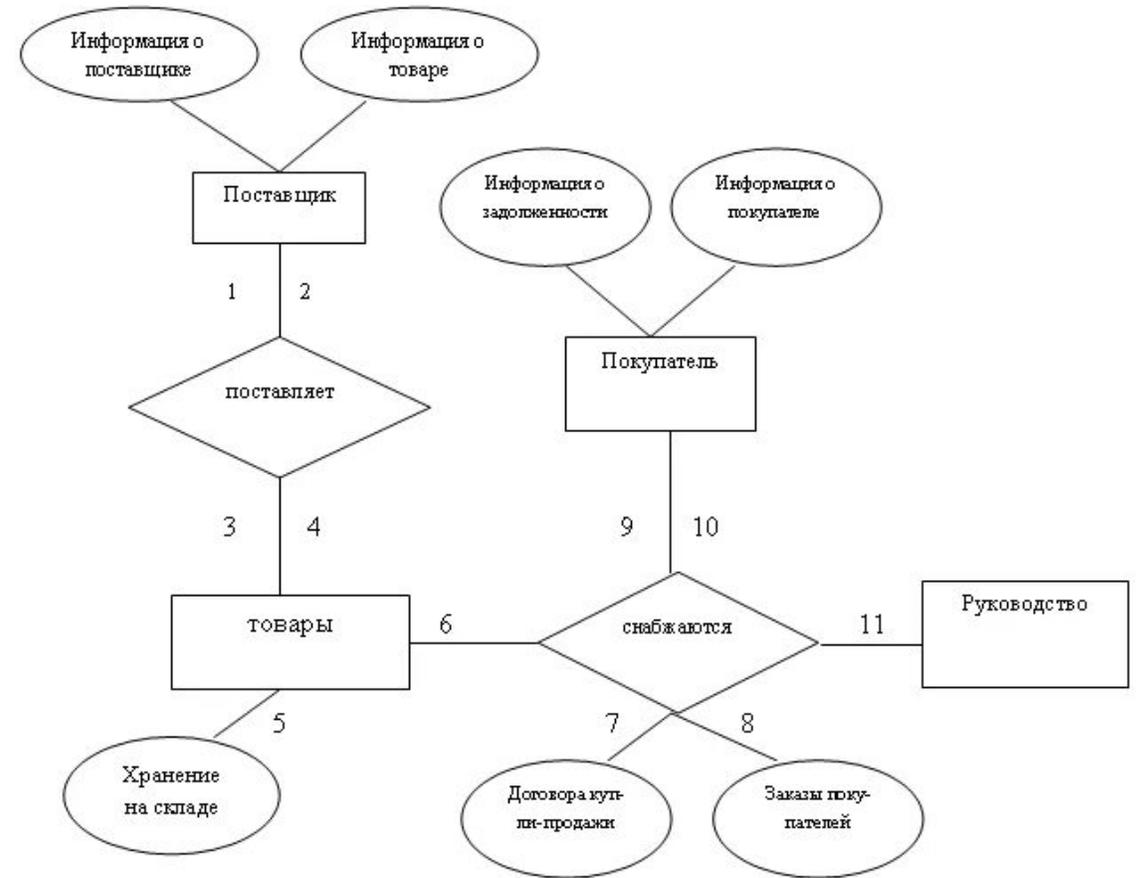
# Концептуальная модель данных

Концептуальная модель данных (КМД) - оперирует концептами разного уровня детализации.



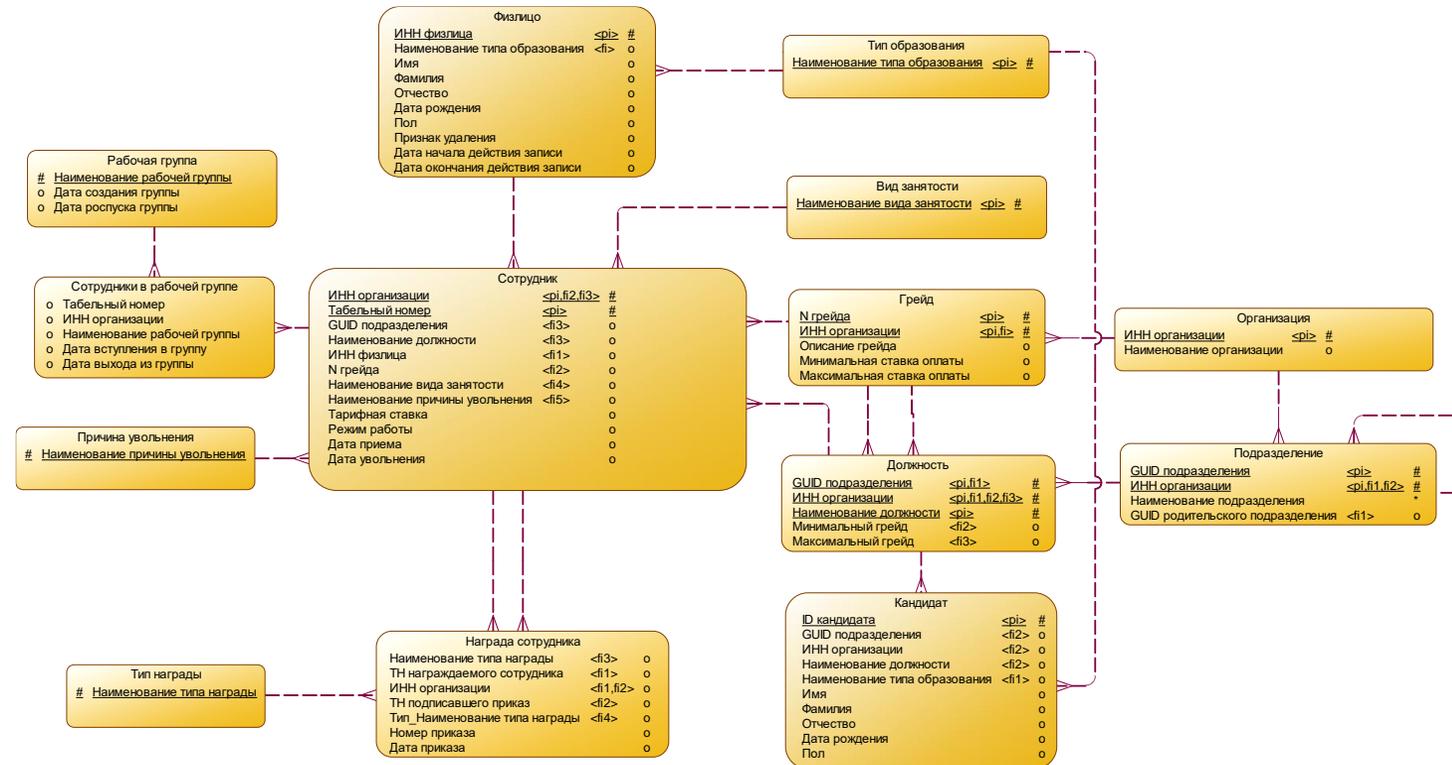
# Логическая модель данных

**Логическая модель данных (ЛМД)** - предназначенные для формализованного, логически полного и непротиворечивого документирования структуры данных, используемых в различных бизнес-процессах.



# Физическая модель данных

**Физическая модель данных (ФМД)** - определяет фактический дизайн базы данных. На этой модели базируется код, используемый для определения структуры хранения данных (создания таблиц, представлений, ограничений целостности).



# Роли в проработках моделей

## Концептуальная модель

Прорабатывается на уровне архитектора модели данных

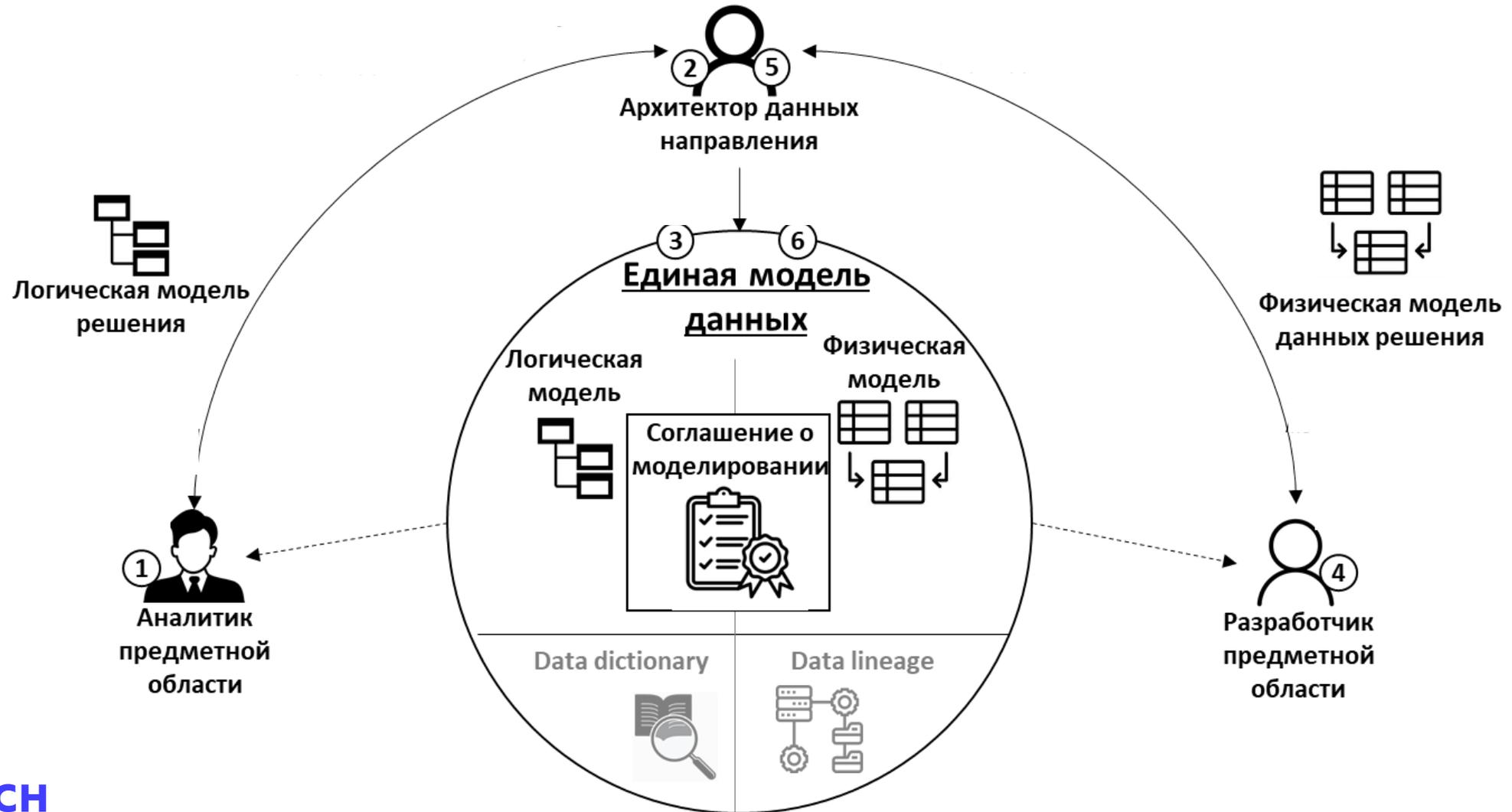
## Логическая модель

Разрабатывается СА и согласовывается на уровне архитектора модели

## Физическая модель

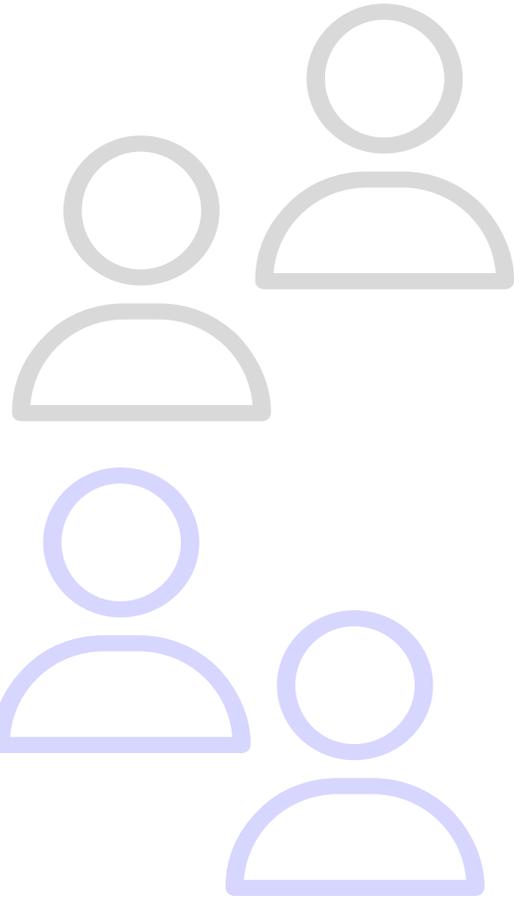
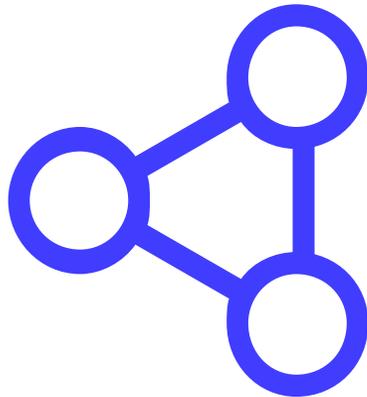
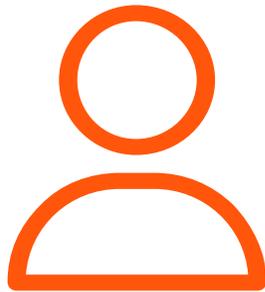
Создается на этапе разработки Dev и контролируется на уровне архитектора модели данных на соответствие с логической

# Разработка и ведение моделей данных



# Продуктовая команда

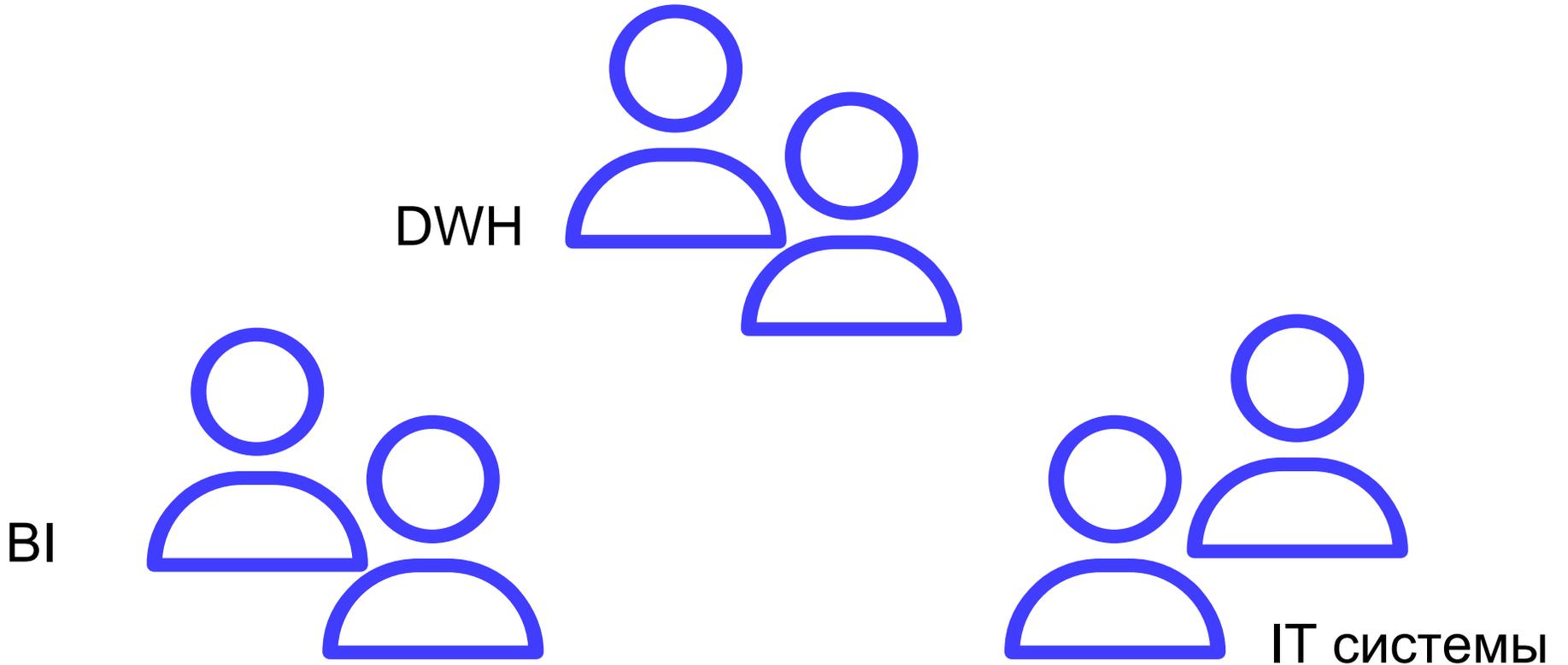
Владелец  
продукта



Представители  
бизнеса

Представители IT

# Представители IT



# Команда DWH

Системные аналитики



Разработчики



# Требования на реализацию DWH

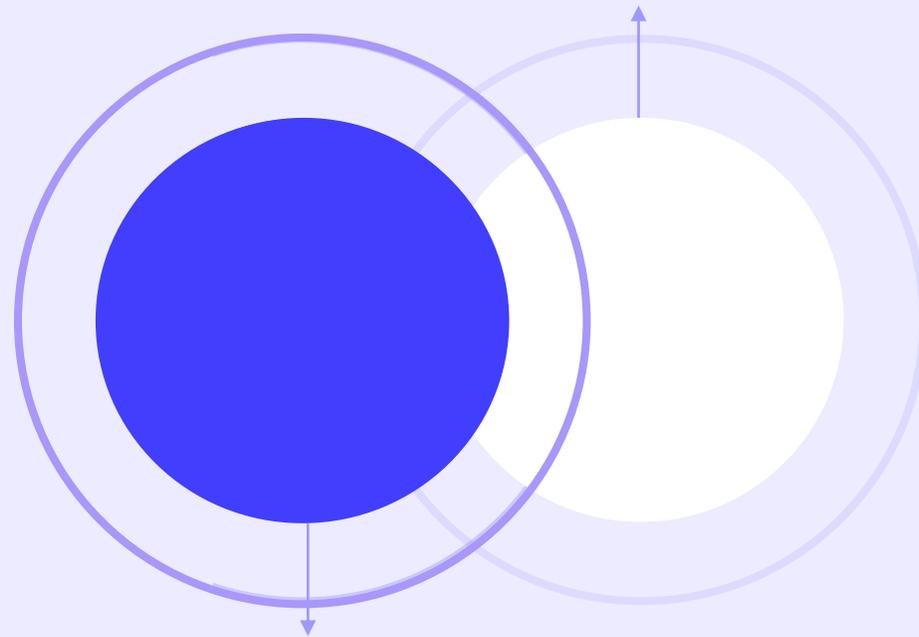
## Системные требования в части DWH:

- требования интеграции с учетными системами и системой BI
- требования хранения данных
- модель данных
- ...

В результате работы команды DWH реализуются витрины, которые смогут использовать в BI инструментах

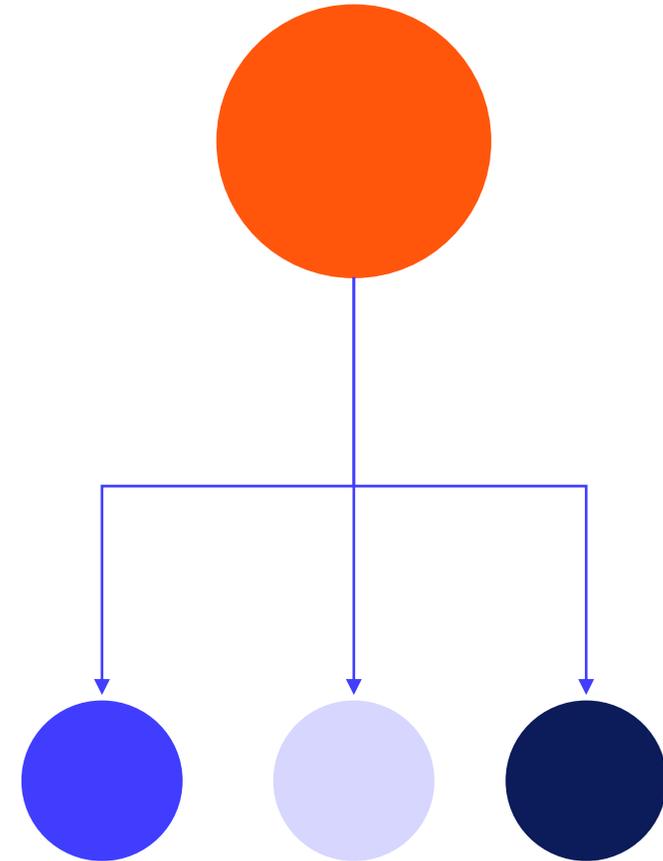
# Требования к данным

Для СА моделирование данных заключается в последовательном выявлении, анализе и формулировании основных требований к данным с последующим их представлением и распространением в точно определенной форме, называемой моделью данных



# Моделирование данных

Процесс моделирования требует выяснения и документирования того, как данные организации соотносятся друг с другом в рамках общей картины. В то же время моделирование заключается в разработке решений в отношении компоновки данных и их связи (Simsion, 2013).



## 2 способа моделирования

### Моделирование сверху вниз

От идеальной концептуальной модели данных до детализации предметных областей

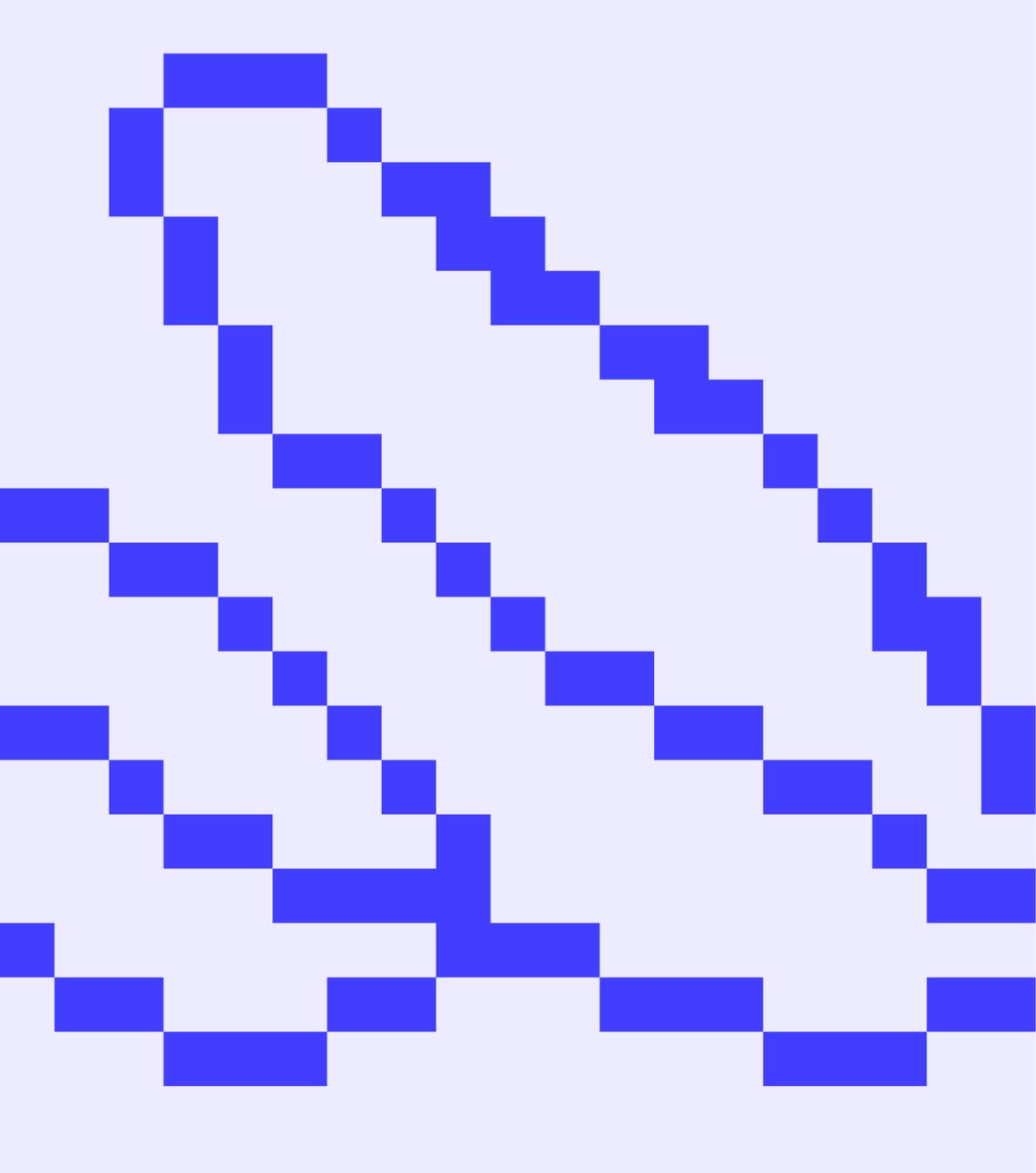
### Моделирования снизу вверх

От бизнес потребности в данных формируется предметная область и встраивается в общую концептуальную модель данных



В Магнит мы пошли по гибриднему пути, от потребности бизнеса прорабатывается предметная область и интегрируется с другими, и как пазл складывается общая модель данных.

Создана общая верхнеуровневая концептуальная модель предметных областей, гибко и оперативно дополняемая по мере необходимости детализациями на локальных уровнях в рамках развития DWH



## В чем преимущества подхода Магнит?

Такой подход позволяет корпоративной архитектуре данных Магнит развиваться пошагово, за счет небольших приращений.

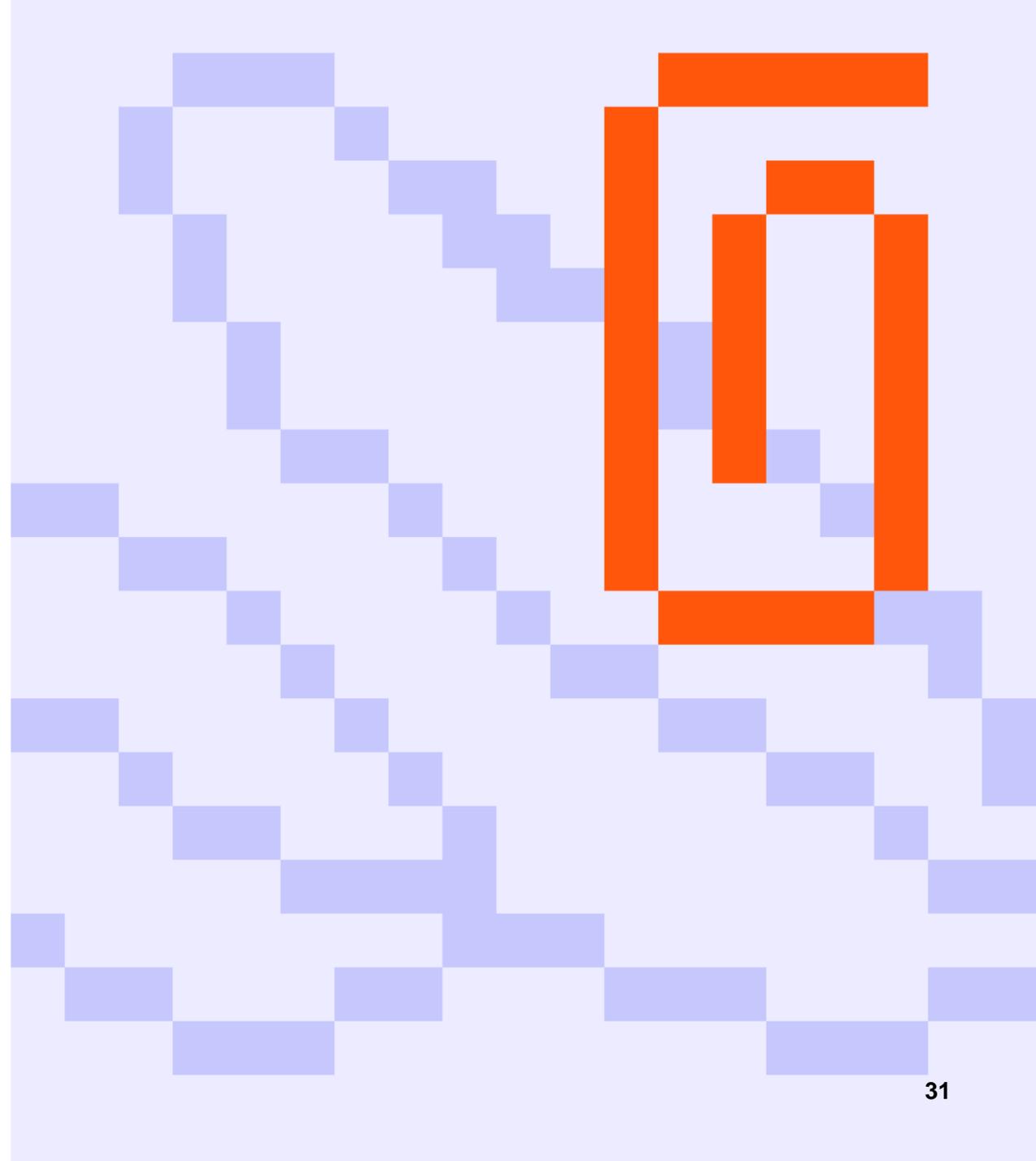
Это требует обязательного участия архитекторов моделей данных во всех инициативах в области разработки с самого их начала, поскольку в условиях инновационной культуры они развиваются очень быстро

# Важно!

У данных есть жизненный цикл и на каждом этапе важно их соответствие актуальной модели данных.

Процесс работы с изменениями содержит согласование с архитектором модели данных, этого можно добиться организационными или платформенными ограничениями

- **Все изменения должны проходить через актуализацию модели**



**Архитектура  
данных** как часть  
корпоративной  
архитектуры

**Моделирование  
и жизненный цикл  
данных**

**Метрики для  
оценки качества  
архитектуры**



# А что если ...?!

- Нет понимания о качестве архитектуры данных компании?
- Архитектура данных не внедрена или не соблюдается?

# Метрики для оценки качества архитектуры

**Метрики** для оценки качества архитектуры.

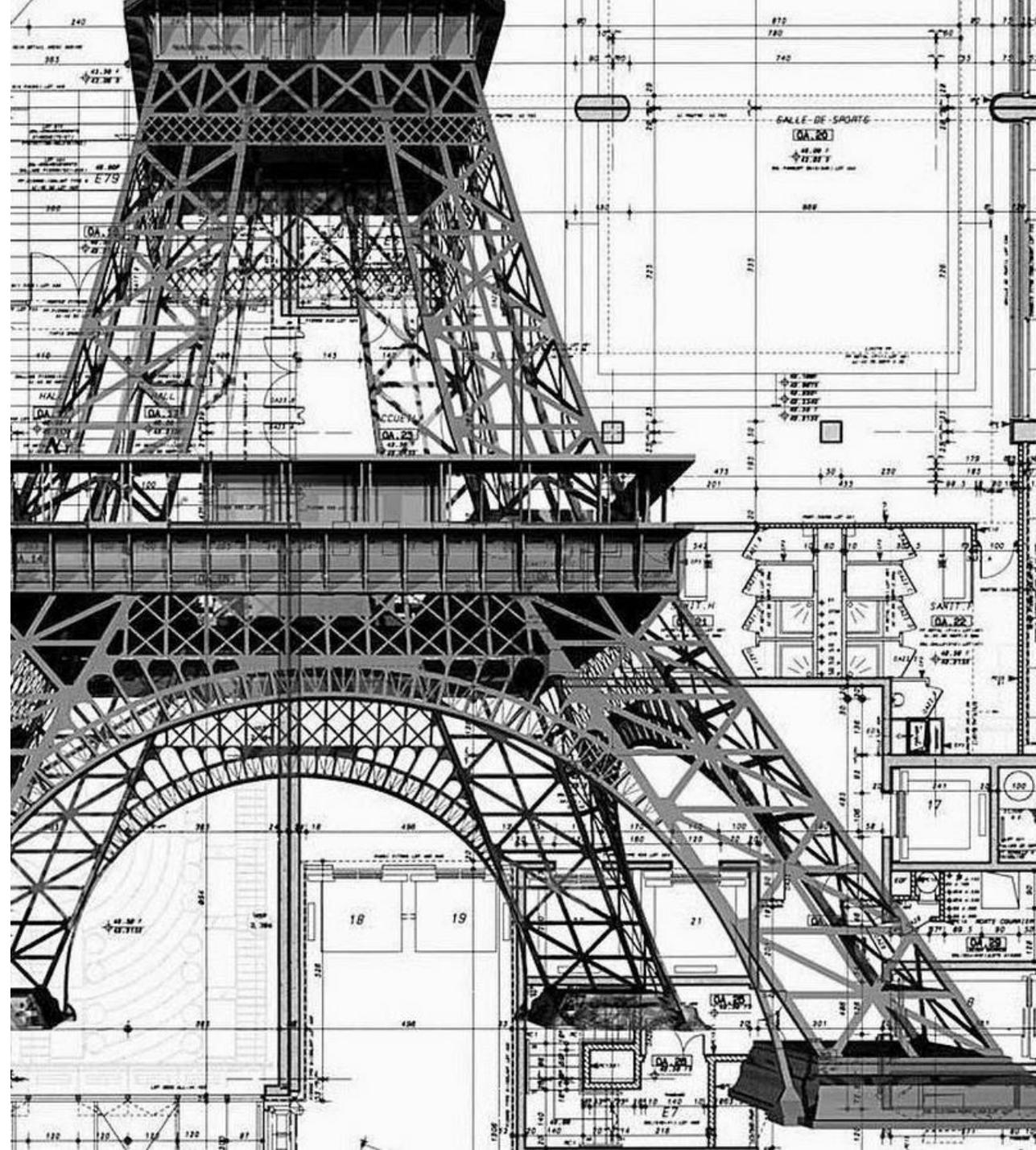
Уровень соблюдения архитектурных стандартов

Тренды внедрения

Оценка ценности для бизнеса

# Уровень соблюдения архитектурных стандартов

Уровень соблюдения архитектурных стандартов, таких как Правила моделирования, Нейминг конвеншен и т.д. Отражает, насколько строго продукт соблюдает требования принятых архитектур данных и придерживаются практики учета в своих процессах корпоративной архитектуры



# Соблюдение архитектурных стандартов

где  $Cas$  — число архитектурных стандартов;

$Sas$  — число изменений с соблюдением архитектурных стандартов;

$Sv$  — общее число изменений;

Метрики отражающие случаи отказа от соблюдения архитектурных требований полезны для понимания узких мест и препятствий для принятия архитектурной практики

$$Mas = \frac{\sum Sas}{Sv * Cas}$$

# Тренды внедрения



## Оценки количества используемых / повторно используемых / замененных / отмененных архитектурных артефактов.

Соотношение этих показателей судит об интенсивности внедрения и доле новых архитектурных разработок



## Оценки окупаемости продукта.

Оцениваются сроки и затраты на реализацию продуктов в целях выявления резервов для совершенствования за счет применения повторно используемых артефактов

# Оценки окупаемости продукта

Где  $P$  — резервы для совершенствования продукта;

$Z$  — затраты на реализацию продуктов;

$Z_a$  — затраты на повторно используемые артефакты;

$$P = \sum Z - Z_a$$

# Оценка ценности для бизнеса

## Повышение гибкости бизнеса.

Показатели, отражающие выгоды, полученные за счет усовершенствований жизненного цикла данных, или стоимость потерь, обусловленных задержками

## Качество операционной деятельности.

Показатели повышения эффективности:

- повышение точности;
- снижение затрат времени;
- снижение расходов на исправление ошибок, обусловленных некорректными данными и т. п.

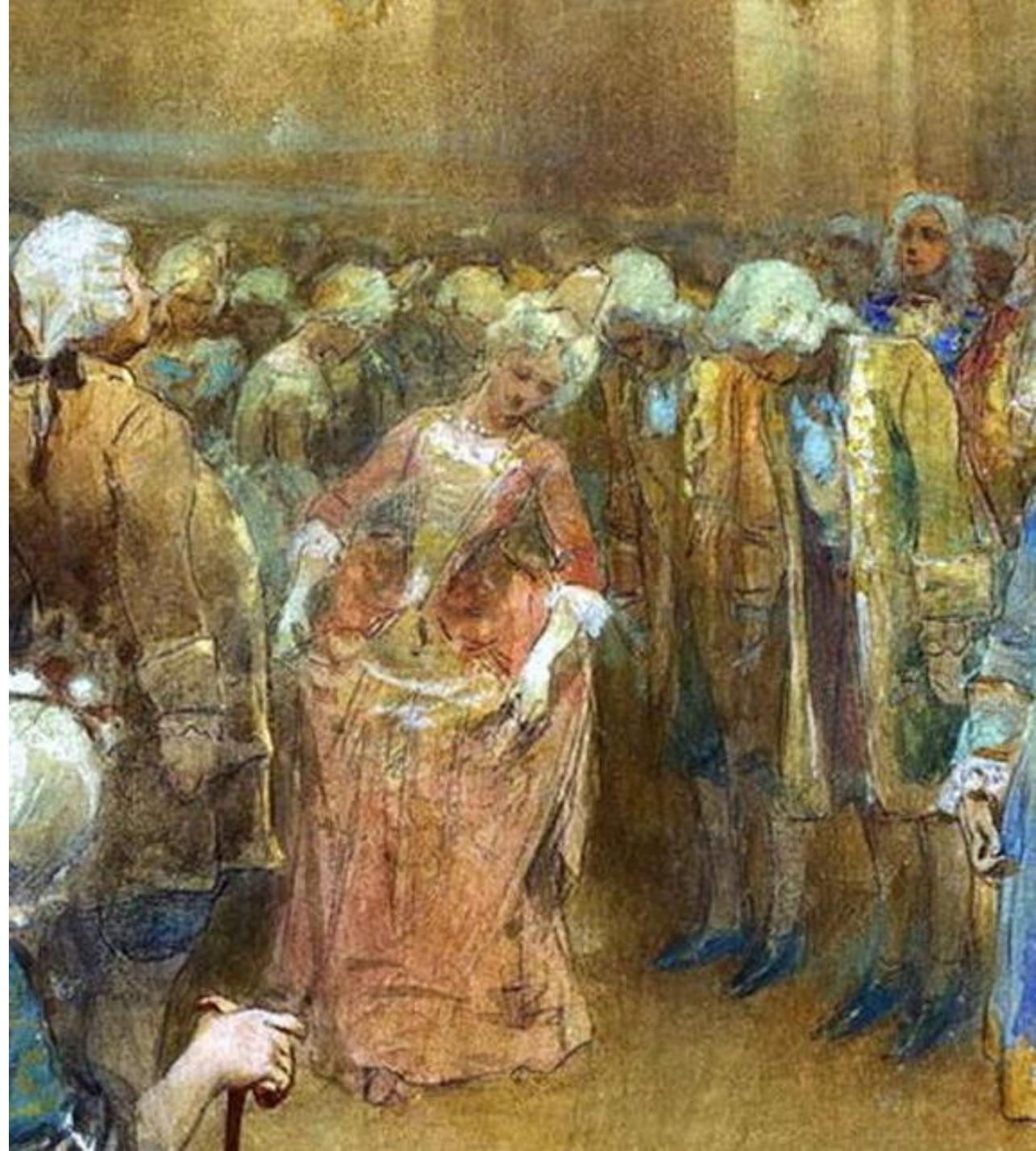
# Культура компании

Скорость с которой компания осваивает архитектурную практику зависит от того, насколько адаптивна ее культура.

Сфокусированные на результате, стратегически ориентированные компании находятся в лучшем положении с точки зрения восприятия архитектурных практик.

Таким компаниям свойственны:

- целеустремленность;
- знание проблем клиентов и партнеров;
- расстановка приоритетов



# Важно!

**Метрики** позволяют оценить качество архитектуры и ее стандартов.

**Тренды внедрения** позволяют отслеживать насколько корпоративная архитектура повлияла на улучшение способности компании реализовывать продукт

- **Необходимо измерять качество архитектуры**

# Что с собой забрать

**Архитектура  
данных** как часть  
корпоративной  
архитектуры

**Моделирование  
и жизненный цикл  
данных**

**Метрики для  
оценки качества  
архитектуры**



# Что с собой забрать

01

**Вести архитектуру  
данных обязательно**

02

**Все изменения  
должны проходить  
через актуализацию  
модели**

03

**Необходимо измерять  
качество архитектуры**

04

Управлять данными это  
– знать о них все,  
поддерживать качество  
и актуальность.

# Спасибо за внимание!



**Гаврилков Григорий**

Руководитель отдела системного  
анализа

Читай нас на:

HABR



[Habr.com/ru/  
company/magnit](https://habr.com/ru/company/magnit)

VC



[Vc.ru/magnit-tech](https://vc.ru/magnit-tech)

Magnit



[www.magnit.tech](https://www.magnit.tech)

Присоединяйтесь  
к нашей команде: