

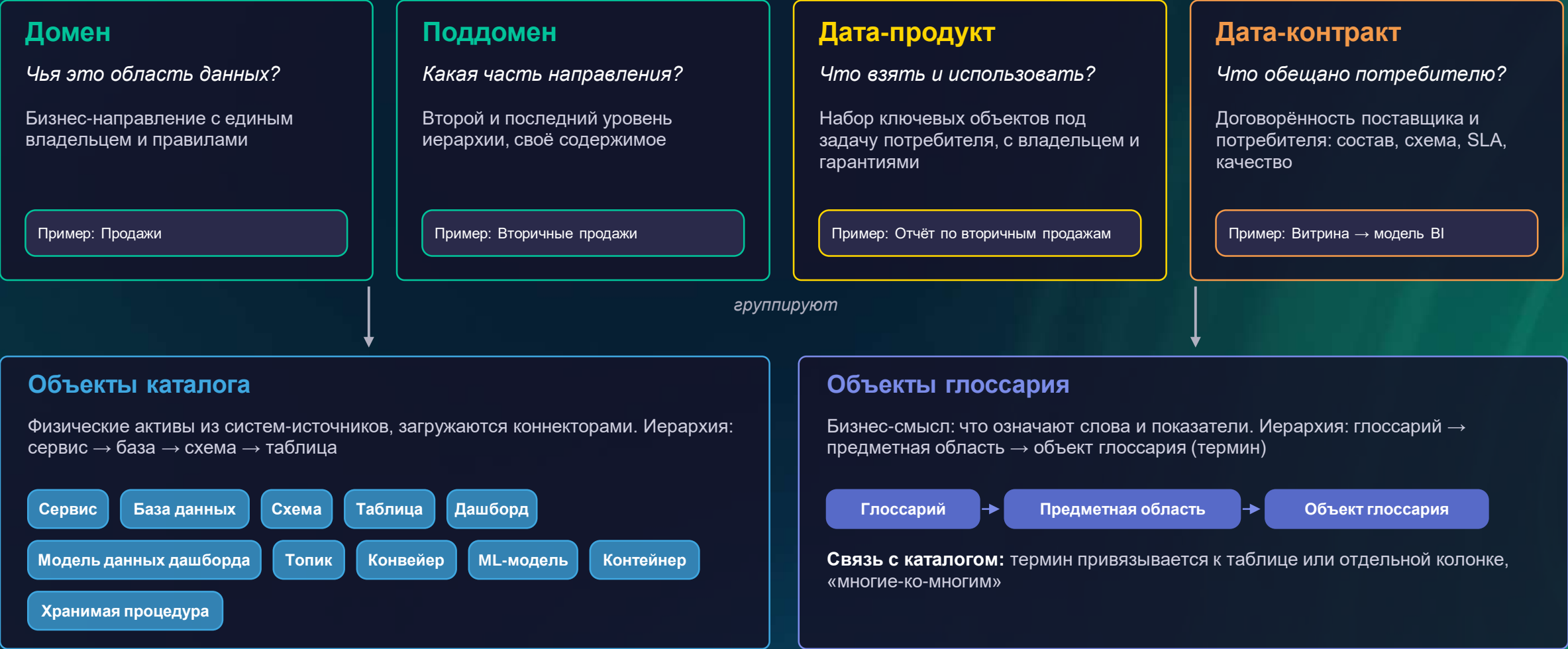
Домены, дата-продукты и глоссарий

Как в Arenadata Catalog группируются объекты каталога и глоссария — на сквозном примере вторичных продаж



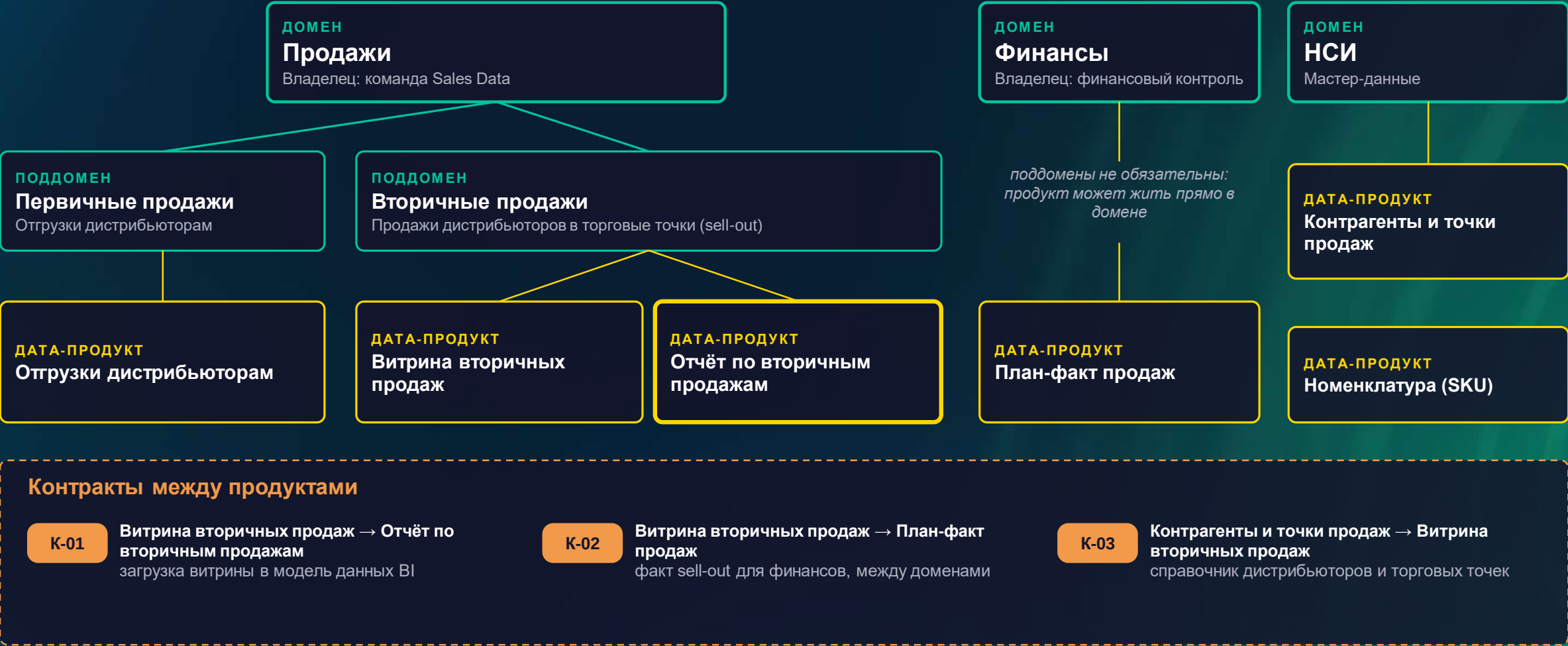
Из чего всё состоит

Внизу — сами объекты; вверху — логические «полки» и «коробки», в которые их раскладывают



Пример: карта домена «Продажи»

Домены и поддомены — «полки», дата-продукты — «коробки» на полках, контракты — договорённости между коробками



Дата-продукт «Отчёт по вторичным продажам»

Продукт собирает ключевые объекты каталога и глоссария под одну задачу потребителя

Домен: Продажи

>

Поддомен: Вторичные продажи

>

Дата-продукт: Отчёт по вторичным продажам

Объекты каталога

Дашборд

Вторичные продажи: динамика и покрытие

Дашборд

Вторичные продажи по дистрибьюторам

Модель данных дашборда

SM «Вторичные продажи»

Конвейер

refresh_sm_secondary_sales

ДАТА-ПРОДУКТ

Отчёт по вторичным продажам

Владелец: BI-команда коммерции

Потребители: КАМ, региональные менеджеры, коммерческий директор

Обновление: ежедневно до 08:00 мск

Уровень SLA: Silver

Этап: Активный

Источник: ДП «Витрина вторичных продаж» по контракту К-01

Объекты глоссария

Вторичные продажи (sell-out)

Выручка sell-out

Покрытие торговых точек

Дистрибьютор

Торговая точка

Чего в продукте нет: таблиц витрины. Они — состав продукта «Витрина вторичных продаж»; отчёт получает их данные по контракту К-01. Термины можно включить в состав (вкладка «Объекты») или связать через тип связи (вкладка «Связанные объекты»); ориентир — 4–6 ключевых ассетов и 4–7 терминов.

Один объект — один «дом», но несколько продуктов

Домен отвечает «чьё это», продукт — «где это используется». Поэтому это разные измерения

Объект	Тип	Домен объекта	Витрина вторичных продаж	Отчёт по вторичным продажам	Отгрузки дистрибьюторам	План-факт продаж
dm_secondary_sales.sales_fact	Таблица	Вторичные продажи ↓	●			●
dm_secondary_sales.distributor_stock	Таблица	Вторичные продажи ↓	●			
dm_mdm.dim_counterparty	Таблица	НСИ ↓	●		●	
dm_mdm.dim_sku	Таблица	НСИ ↓	●		●	●
load_secondary_sales	Конвейер	Вторичные продажи	●			
SM «Вторичные продажи»	Модель данных BI	Вторичные продажи		●		
Вторичные продажи: динамика	Дашборд	Вторичные продажи		●		
Выручка sell-out	Термин	Продажи + Вторичные ↓	●	●		●
Дистрибьютор	Термин	НСИ ↓	●	●	●	

↓ — домен унаследован (от схемы, глоссария или предметной области), без значка — назначен вручную

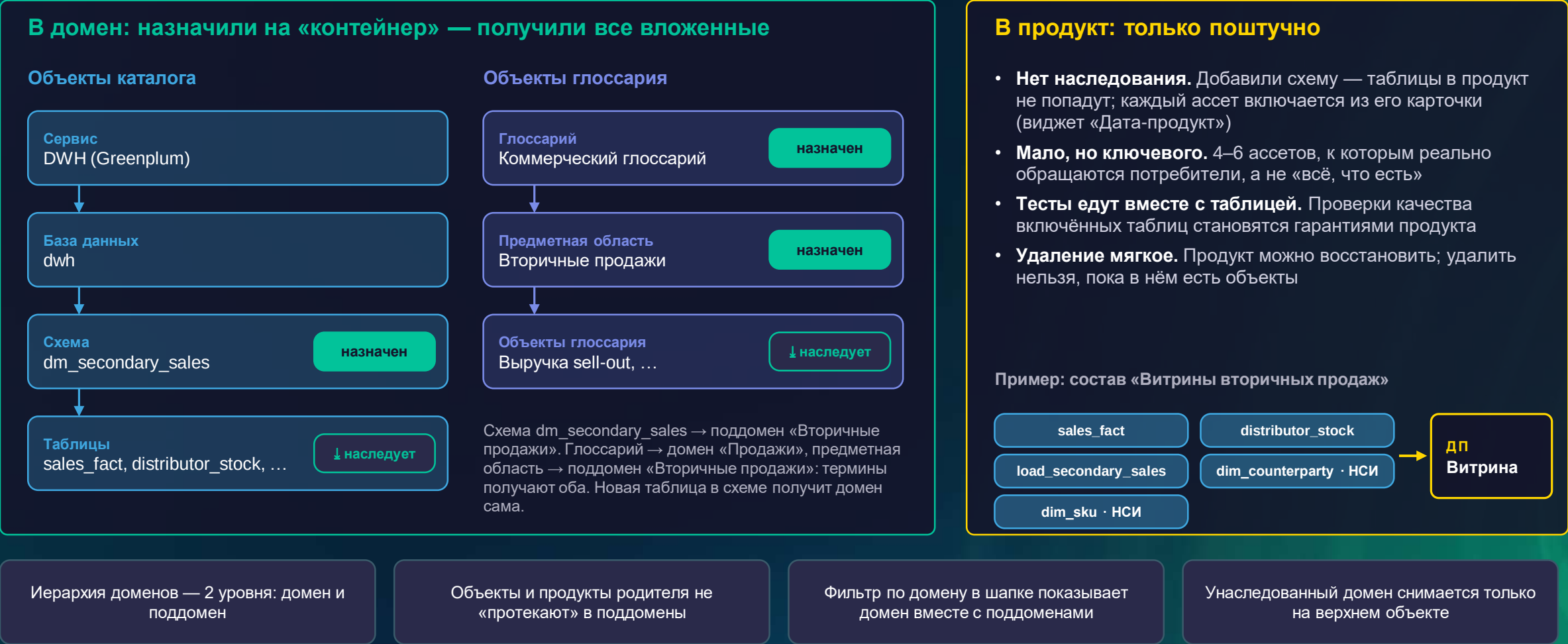
Домен: до 5 «домов». Вручную объекту назначают до пяти доменов; унаследованных — сколько угодно. Прямое назначение приоритетнее наследуемого.

Продукт: сколько угодно. dim_sku входит в три продукта. Состав продукта задаётся явно, объект не «переезжает» из своего домена.

Чужие объекты — можно. Витрина продаж берёт справочники из домена НСИ. В продукт можно включить объекты своего, чужого домена и без домена.

Как объекты попадают в домен и в продукт

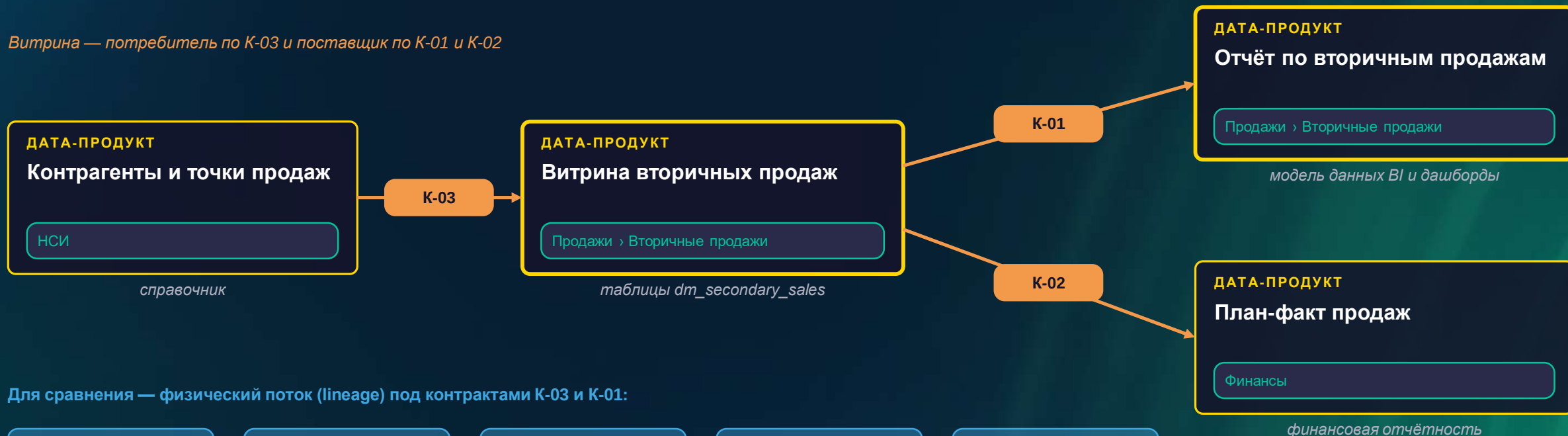
В домен — сверху вниз по наследованию; в продукт — только вручную, поштучно



Дата-контракты: как продукты связаны между собой

Контракт — обязательство продукта-поставщика перед продуктом-потребителем

Витрина — потребитель по K-03 и поставщик по K-01 и K-02



Для сравнения — физический поток (lineage) под контрактами K-03 и K-01:



Что фиксирует контракт

Состав и схему передаваемых данных, свежесть и SLA, правила качества, порядок изменений и уведомлений, версию и статус

Контракт ≠ lineage

Lineage показывает, как данные текут физически (таблица → модель BI). Контракт — что поставщик обещает и что будет при изменениях

Внутри и между доменами

K-01 связывает продукты одного поддомена, K-02 — доменов «Продажи» и «Финансы». Владелец с обеих сторон видит друг друга

Контракт К-01: витрина → модель BI

Что записано в контракте и как это оформляется в Arenadata Catalog 1.0

К-01 · Витрина → Отчёт по вторичным продажам

Поставщик	ДП «Витрина вторичных продаж» · владелец Sales Data
Потребитель	ДП «Отчёт по вторичным продажам» · BI-команда коммерции
Предмет	sales_fact, distributor_stock → модель данных SM «Вторичные продажи»
Схема	period_date, distributor_id, outlet_id, sku_id, qty, amount_rub
Свежесть	данные за D-1 до 06:00 мск; обновление модели BI в 06:30
Качество	ключ строки уникален, amount_rub ≥ 0, сверка с первичными продажами ±2%
Изменения	уведомление потребителя за 10 рабочих дней; несовместимые — только новой версией
Статус	Активный · версия 1.2

Значения условные — пример структуры

Как оформить сегодня

Расширенные атрибуты продукта

Атрибут «Дата-контракт» со ссылкой на документ К-01 v1.2, а также SLA, частота обновления, потребители

Тип связи между продуктами

Направленная связь «Поставляет данные / Получает данные» на вкладке «Связанные объекты» — видна с обеих сторон

Lineage и тесты качества

Lineage подтверждает поток sales_fact → модель BI; тесты таблиц витрины — проверяемая часть обещаний

Подписка потребителя

Владелец отчёта подписан на витрину и получает уведомления об изменениях её карточки

В планах развития ADC: контракт как отдельная сущность — с версионированием схемы, политикой совместимости и проверкой данных на соответствие. Оформленное сейчас ляжет в неё без переделок.

Взгляд с другой стороны: карточка таблицы

Каждый объект сам «знает», в каком он домене, в каких продуктах и какие термины его описывают

ТАБЛИЦА

dwh.dm_secondary_sales.sales_fact

Домен

Продажи > Вторичные продажи

↓ унаследован от схемы dm_secondary_sales

Дата-продукты

Витрина вторичных продаж

Продажи

План-факт продаж

Финансы

Термины

Вторичные продажи (sell-out)

на таблицу

Выручка sell-out

на колонку amount_rub

Дистрибьютор

на колонку distributor_id

Lineage

load_secondary_sales

→

sales_fact

→

SM «Втор. продажи»

→

Дашборд

Три пути к одним и тем же данным

От домена

Карточка домена: вкладки «Поддомены», «Дата-продукты», «Объекты», «Участники», «Связанные объекты»

От продукта

Карточка продукта: «Основная информация» с атрибутами, «Объекты» — состав, «Связанные объекты» — термины и контракты

От объекта

Виджеты «Домен» и «Дата-продукт» в карточке, термины на таблице и колонках, lineage

Фильтр по домену

Переключатель «Все домены» в шапке: «Обзор» и «Глоссарий» показывают только выбранный домен и его поддомены

Раньше: глоссарии и предметные области вместо доменов

Что давала прежняя схема и чего ей не хватало



Чего не хватало

- **Смысл вместо принадлежности.** Глоссарий отвечает «что значит слово», а не «чьи это данные и кто за них отвечает»
- **Нет состава области.** Список таблиц «Вторичных продаж» собирается поиском по тегам и неполон: нет наследования от схем
- **Нет владельца набора активов.** Владелец есть у термина и у таблицы, но не у области данных целиком
- **Нет упаковки для потребителя.** Негде зафиксировать SLA, потребителей и связи поставщик → потребитель

Глоссарий не заменяется доменом — они дополняют друг друга

Глоссарий и ПО

слой смысла: термины и определения

Домены и поддомены

слой ответственности: чьё и кто отвечает

Дата-продукты и контракты

слой потребления: что брать и с какими гарантиями

Как перенести прежнюю схему на домены

Глоссарий не трогаем — «привязываем» его к доменам и достраиваем недостающие слои

Было	Становится	Что сделать в каталоге
● Глоссарий «Продажи»	● Домен «Продажи»; глоссарий остаётся	Создать домен и назначить его на глоссарий — предметные области и термины унаследуют домен
● ПО «Вторичные продажи»	● Поддомен «Вторичные продажи»	Создать поддомен в карточке домена и назначить его на предметную область — термины получат и домен, и поддомен
● Общий корпоративный глоссарий	● Остаётся общим	Домен назначать не на весь глоссарий, а на отдельные предметные области
● Таблицы с тегами терминов ПО	● Объекты поддомена	Назначить поддомен на схему-«контейнер» (dm_secondary_sales); точно доразметить остальное
● Отчёт и его источники	● ДП «Отчёт...» и ДП «Витрина...»	Собрать 4–6 ключевых ассетов в каждый продукт, связать 4–7 терминов, заполнить атрибуты
● Договорённости «на словах»	● Контракты К-01, К-02	Атрибут «Дата-контракт» со ссылкой на документ + тип связи между продуктами
● Стюард предметной области	● Владелец поддомена	Назначить команду-владельца и бизнес-стюарда; добавить участников

- 1

Инвентаризация глоссариев, тегов, владельцев
- 2

Таблица «было → стало», согласовать со стюардами
- 3

Домены на глоссарии, ПО и схемы
- 4

Продукты в пилотном поддомене
- 5

Через 2 недели — отзывы, следующий домен

Шпаргалка: куда положить объект

Четыре вопроса и сущность, которая на него отвечает

Что это значит?	Глоссарий → предметная область → термин	Термин «Выручка sell-out» с определением и формулой; привязан к колонке amount_rub
Чьё это и кто отвечает?	Домен → поддомен	Продажи › Вторичные продажи; владелец — команда Sales Data. Назначается на схему, наследуется таблицами
Что брать для задачи?	Дата-продукт	«Отчёт по вторичным продажам»: 2 дашборда, модель данных BI, 5 терминов, SLA и потребители
Что обещано потребителю?	Дата-контракт	K-01: витрина → модель BI, D-1 к 06:00, сверка ±2%, изменения — за 10 рабочих дней

Ориентиры

6–10 корневых доменов для старта	2 уровня: домен и поддомен	2–5 продуктов в домене	4–6 ключевых ассетов в продукте	4–7 терминов на продукт	до 5 прямых доменов у объекта
-------------------------------------	-------------------------------	---------------------------	------------------------------------	----------------------------	----------------------------------