

DWH

**Фулл-стек аналитика корпоративного
хранилища данных**

Бажина Екатерина
Luxoft

Что такое DWH?

Предметно ориентированный, интегрированный, неизменный, поддерживающий хронологию набор данных, предназначенный для поддержки принятия решений.

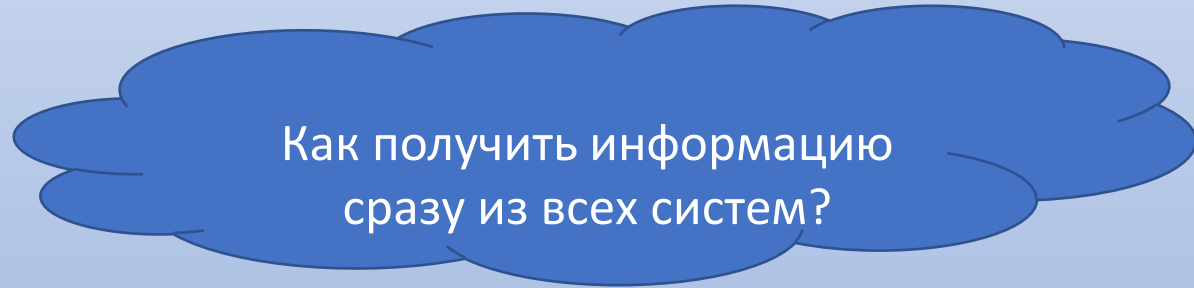
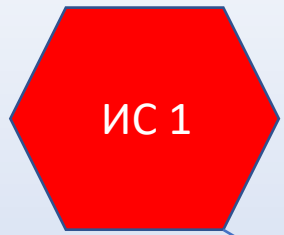
Функции

- а) сбор, организация и подготовка данных для анализа в виде постоянно наращиваемой базы данных
- б) собственно анализ как элемент принятия решений

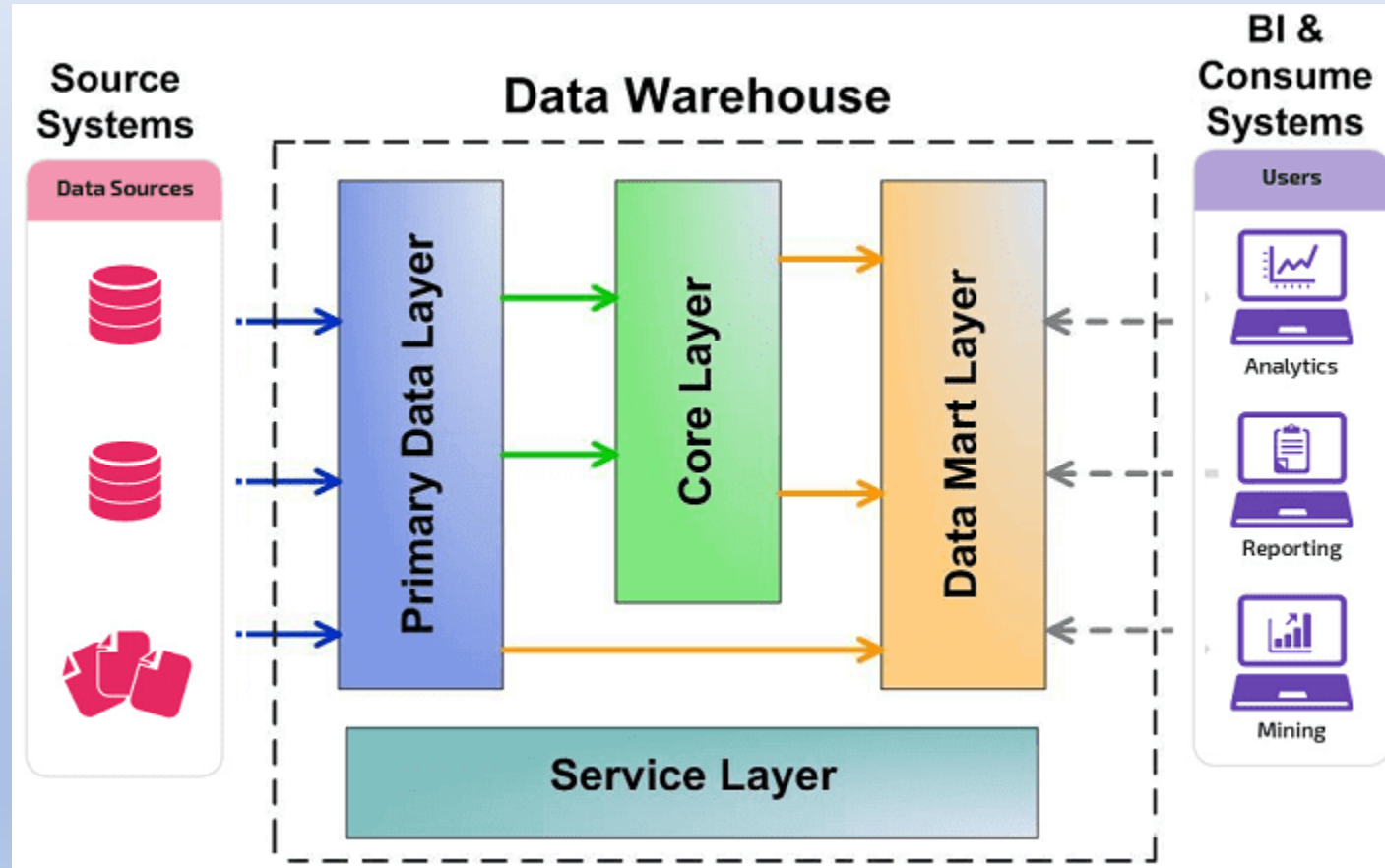
Б. Инмон

Зачем нам DWH?

- обеспечить пользователя (компанию и ее ключевых лиц) возможностью принимать верные решения в ключе управления бизнесом на основе целостной информационной картины

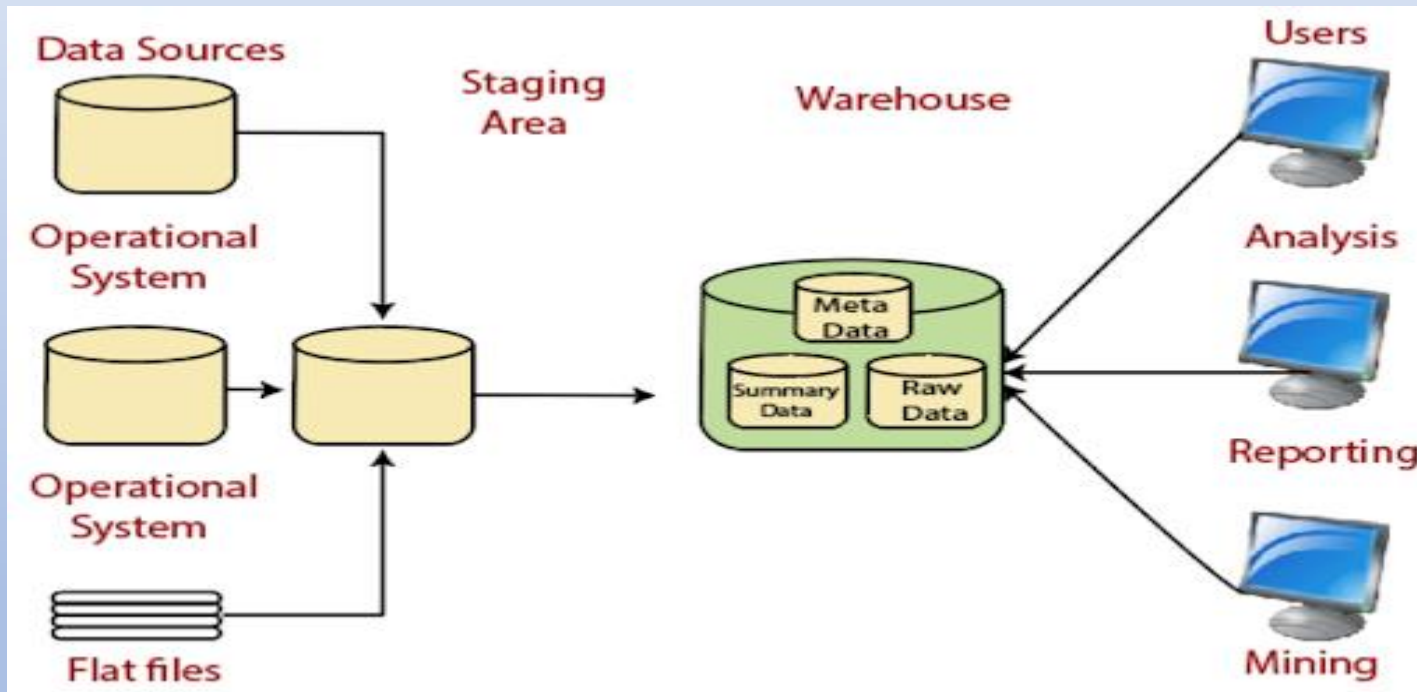


Структура DWH



LSA – слоеная архитектура DWH: как устроено хранилище данных

Источники данных

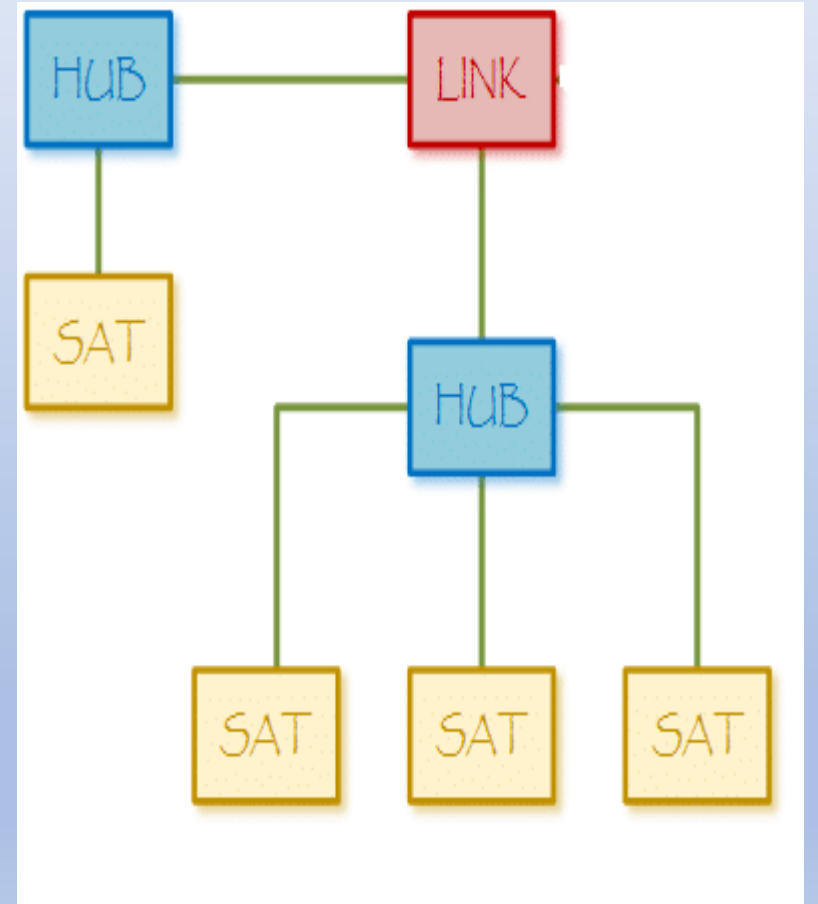


А все ли нам необходимо из систем –
источников?



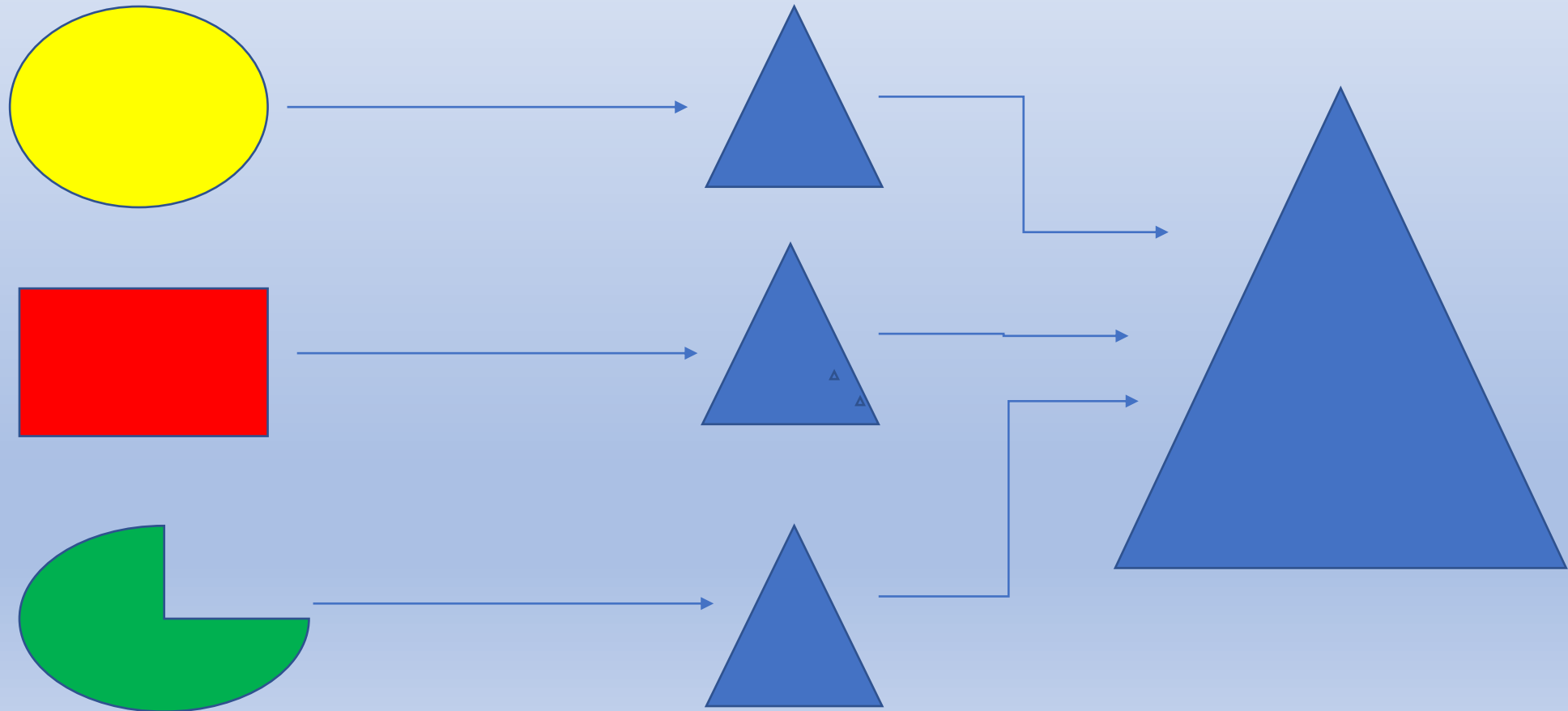
Анатомия DataVault как метода моделирования DWH

-
- HUB — позволяют обеспечить бизнес-ориентированность хранилища и обеспечивают возможность интеграции дополнительных источников данных.
- LINK — обеспечивают связь между сущностями.
- SATELLITE — хранят характеристики и обеспечивают историческое хранение данных.



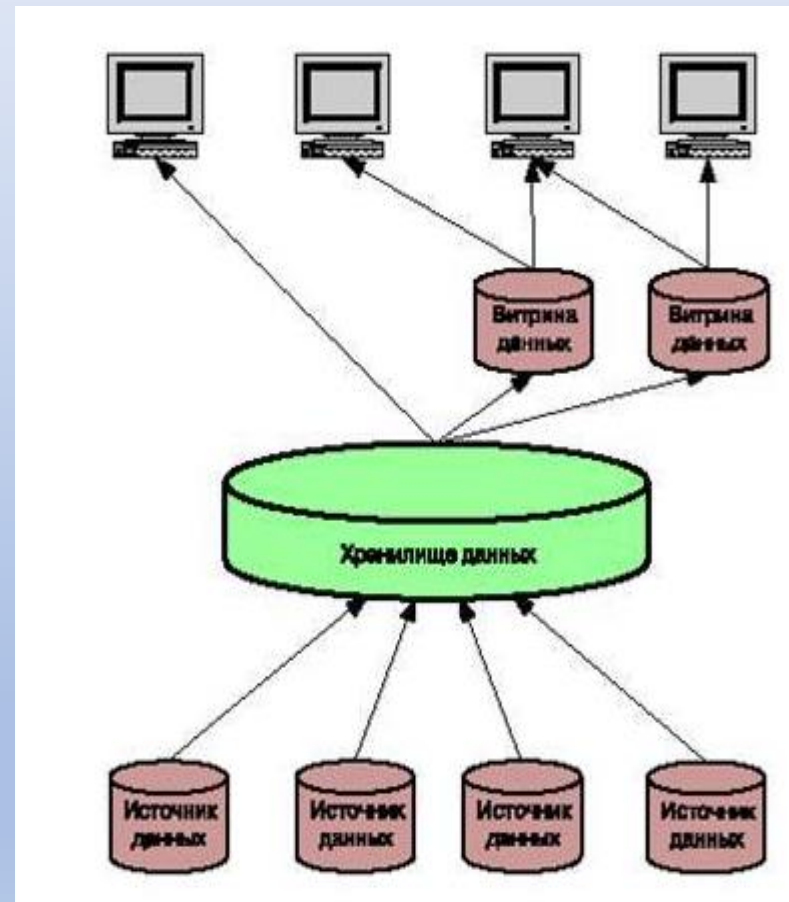
Моделирование DWH

- Что мы делаем?



Витрины данных

- Срез базы, хранилища данных, который призван представлять собой массив узкоспециализированной, тематической информации, ориентированный под запросы сотрудников определенного департамента, вектора работы организации

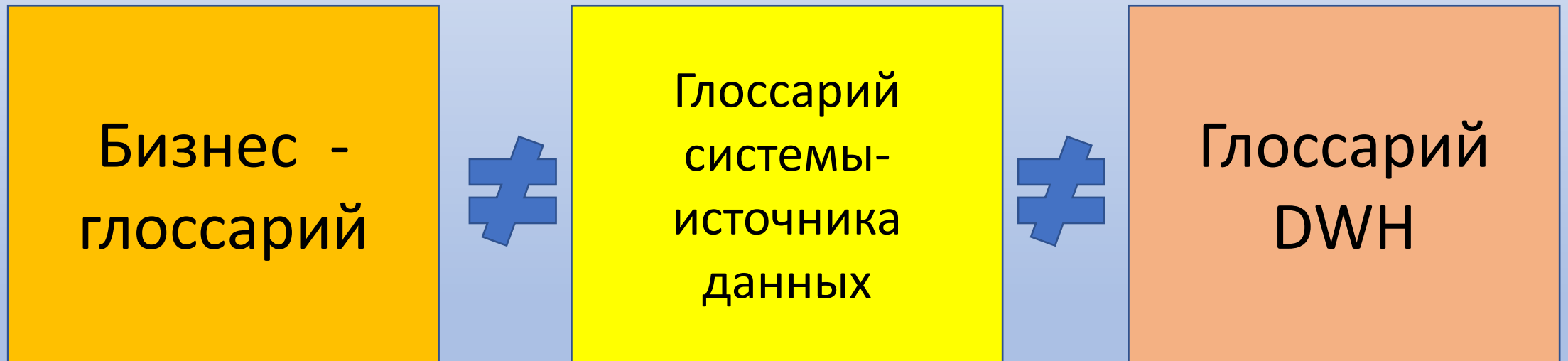


Аналитик на проекте по построению DWH

1. Понимает бизнес-область, умеет общаться с заказчиком, умеет трансформировать бизнесовые требования и формировать технические
2. Может предложить аналитическое решение, понять, как решать задачу, какие нужны данные, какие надо “придумать”, какие надо сделать допущения
3. Умеет работать с источниками данных в сыром виде, понимает, как устроено хранение данных
4. Знает структуру хранилища данных, понимает, как оно спроектировано и замоделировано, понимает при каких условиях возникает необходимость расширения структуры/атрибутивного состава хранилища
5. Знает, как нужно пересобрать данные из источников, как их методологически обрабатывать, и куда их необходимо поместить в хранилище
6. Понимает, как постать данные из хранилища для построения необходимой бизнесу витрины
7. Умеет визуализировать свои результаты и докладывать своим заказчикам

Анализ бизнес - требований

- Глоссарии



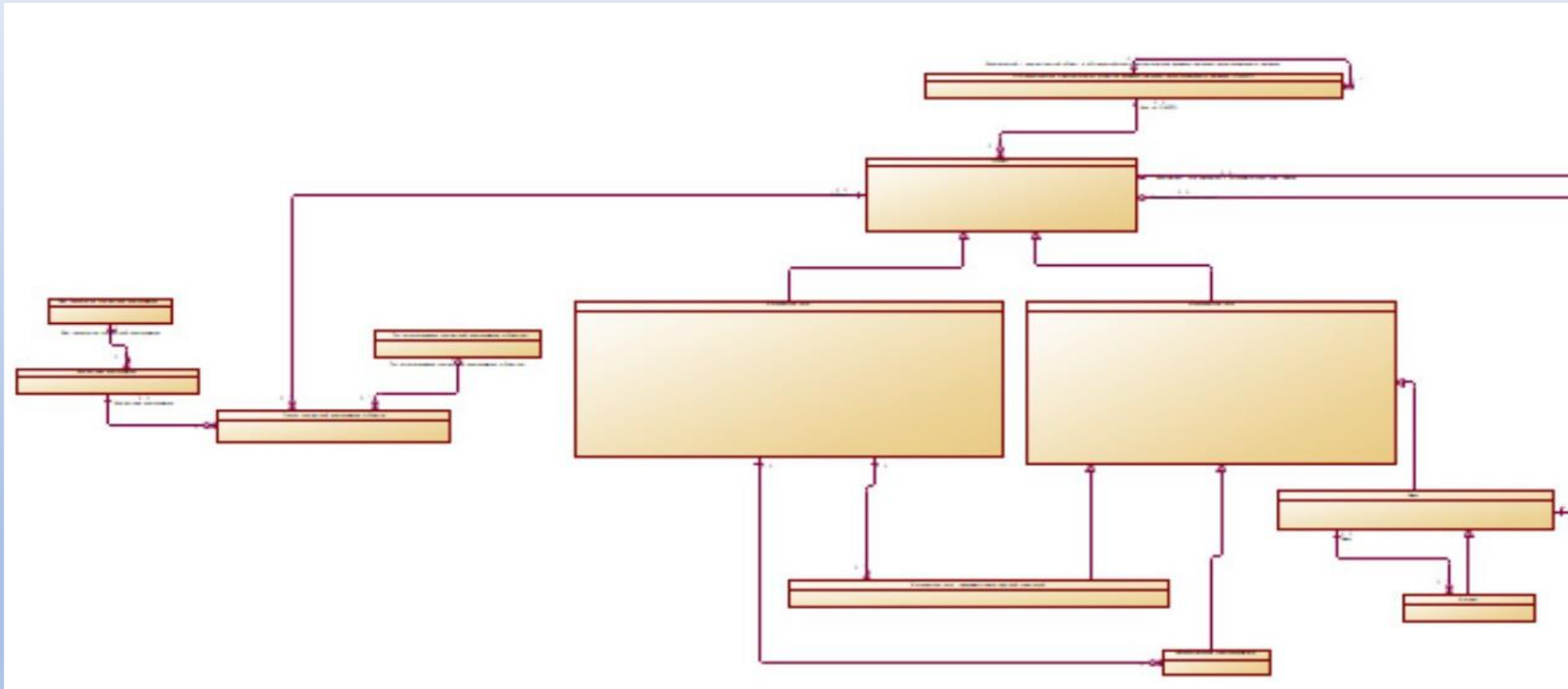
Мэппинг бизнес-требований на системы-источники

- Составление локального глоссария системы – источника
- Определение необходимого атрибутивного состава системы – источника
- Расширение состава реплики источника в контуре DWH

Мэппинг терминов системы-источника на логическую модель

Есть ли для нашего атрибута (сущности) местечко в модели?

Расширение логической модели DWH



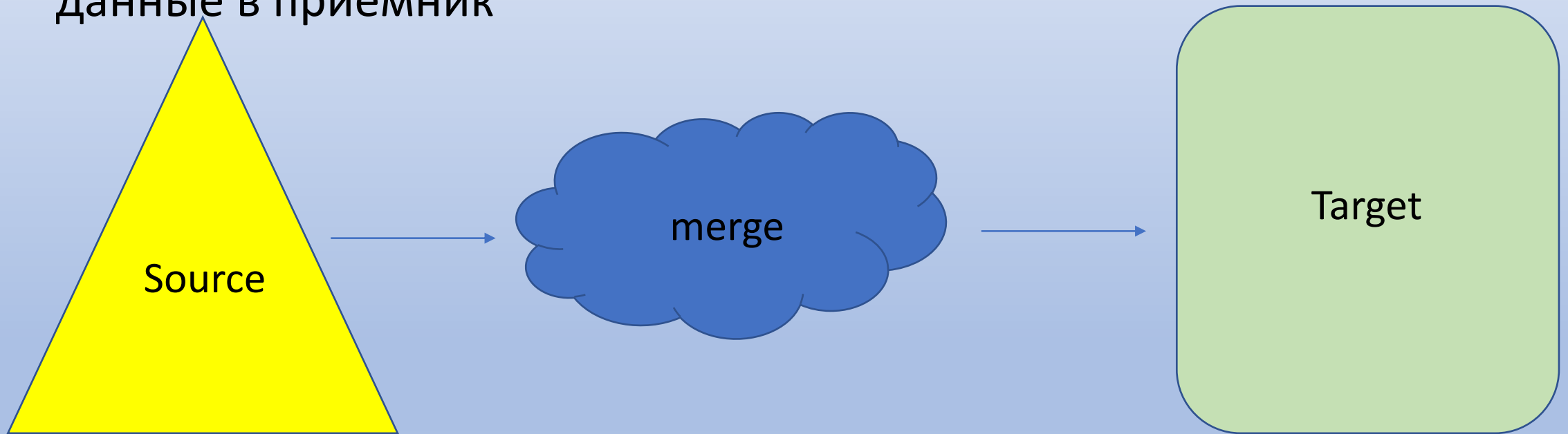
Трансформация логической модели в физическую

А что дальше?

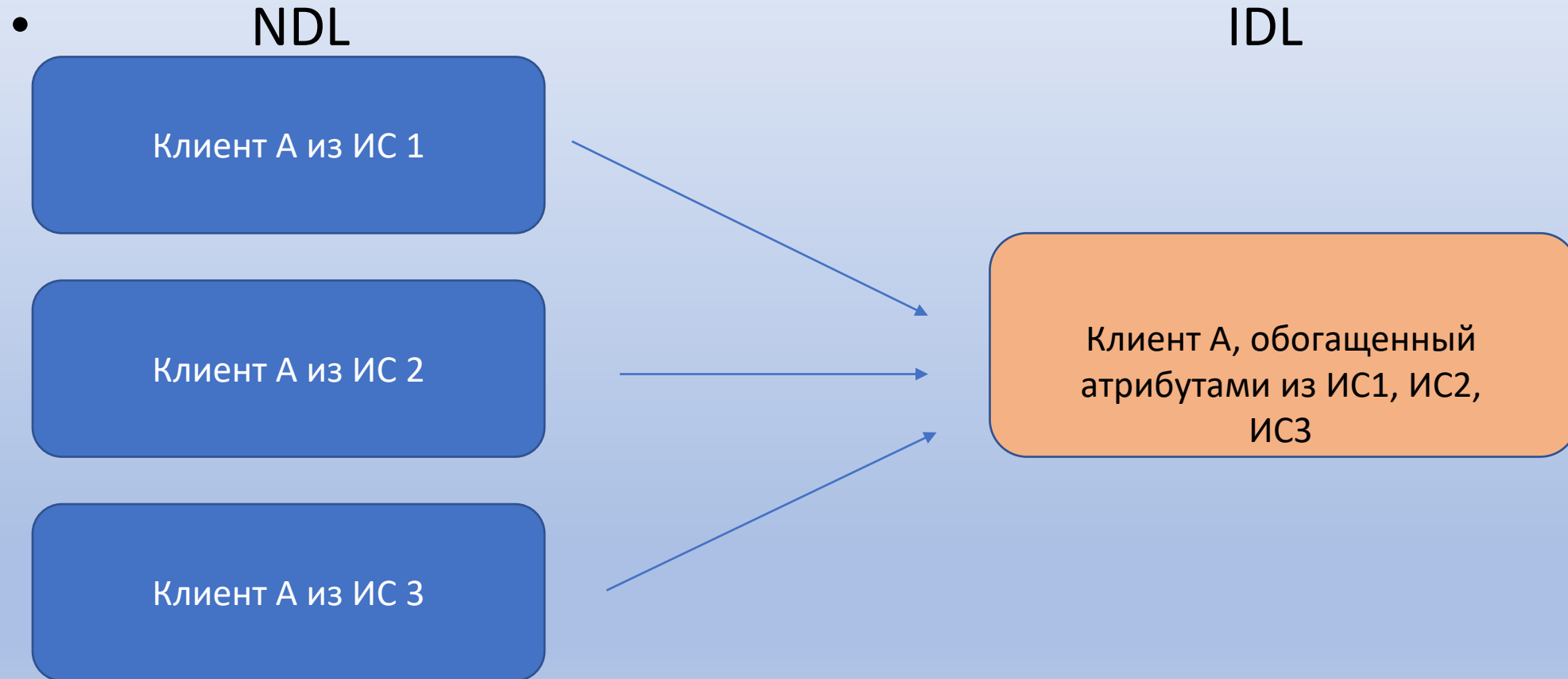
- 1) Пересборка данных на уровне ndI (недифференцированные данные)
- 2) Проливка данных на слой idI (дифференцированные данные). Унификация
- 3) Построение витрин из унифицированных данных

Концепция S2T

- Source to Target – отобрать из источника необходимые данные, пересобрать под концепцию приемника, сложить пересобранные данные в приемник



Унификация на слое idl



Таким образом, аналитик DWH

Совмещает технические и аналитические компетенции

Аналитический цикл длинный

Много интересных и разнообразных задач

Требуется микс знаний – как бизнесовых, как и системных

Большое поле для творческих решений

Вместо заключения

Что интересного почитать по теме DWH?

- Patrick Cuba: The Data Vault Guru. A Pragmatic Guide on Building a Data Vault. Selbstverlag, ohne Ort 2020, ISBN 979-86-9130808-6.
- Kent Graziano: Better Data Modeling. An Introduction to Agile Data Engineering Using Data Vault 2.0. Data Warrior, Houston 2015.
- Hans Hultgren: Modeling the Agile Data Warehouse with Data Vault. Brighton Hamilton, Denver u. a. 2012, ISBN 978-0-615-72308-2.
- Daniel Linstedt: Super Charge Your Data Warehouse. Invaluable Data Modeling Rules to Implement Your Data Vault. Linstedt, Saint Albans, Vermont 2011, ISBN 978-1-4637-7868-2.