

Deployment

Data Warehouse и Data Lake

Skillbox

образовательная платформа

Данные бывают

- 1 Структурированные
- 2 Слабоструктурированные
- 3 Неструктурированные

Структурированные данные — это табличные
данные.

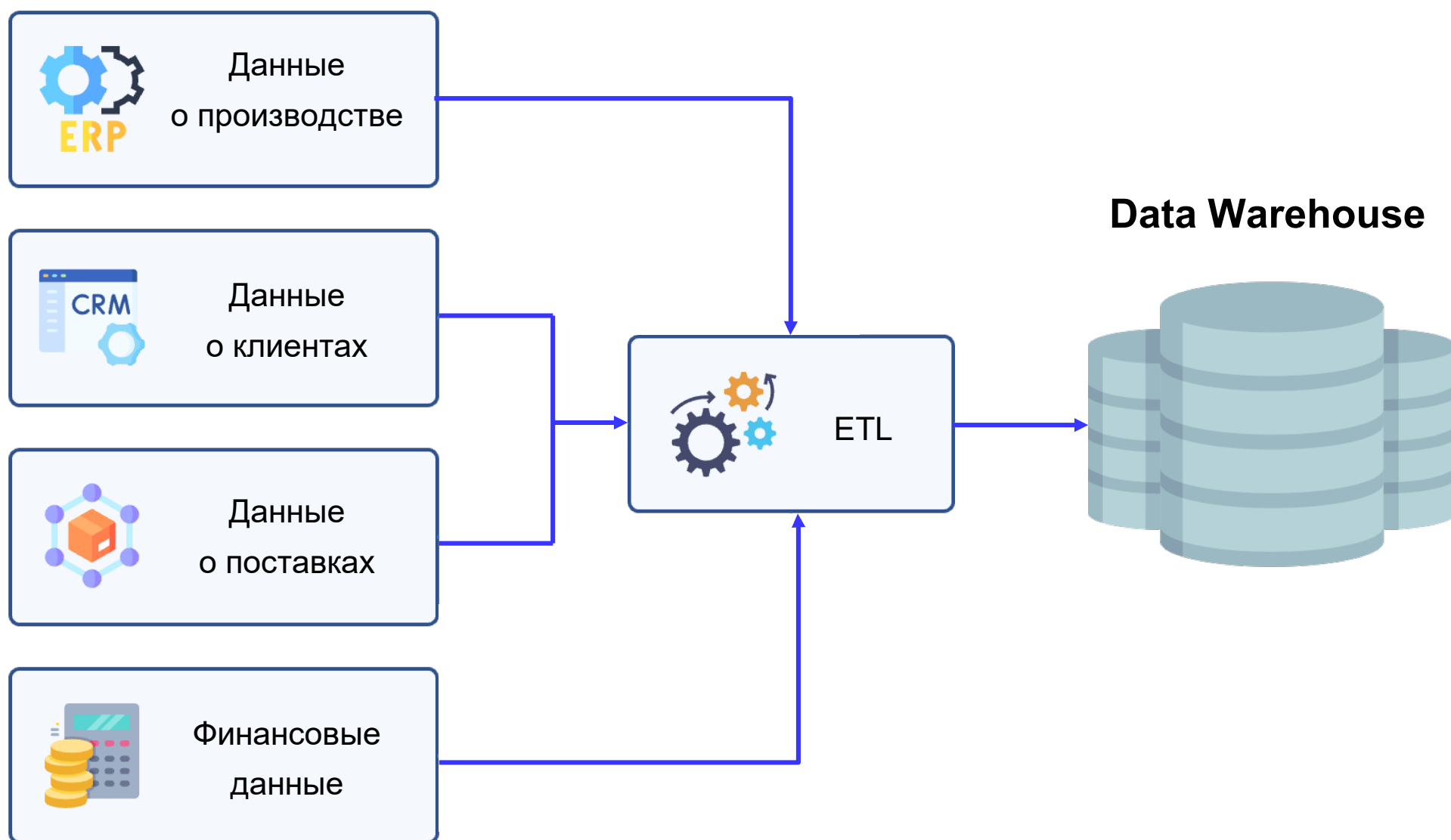
Слабоструктурированные данные имеют гибкую,
адаптивную структуру.

Неструктурированные данные — это тексты, видео, изображения и прочие данные без какой-либо структуры.

Большие данные o_O



Хранилище данных



Хранилище данных консолидирует всю важную
информацию о бизнесе.

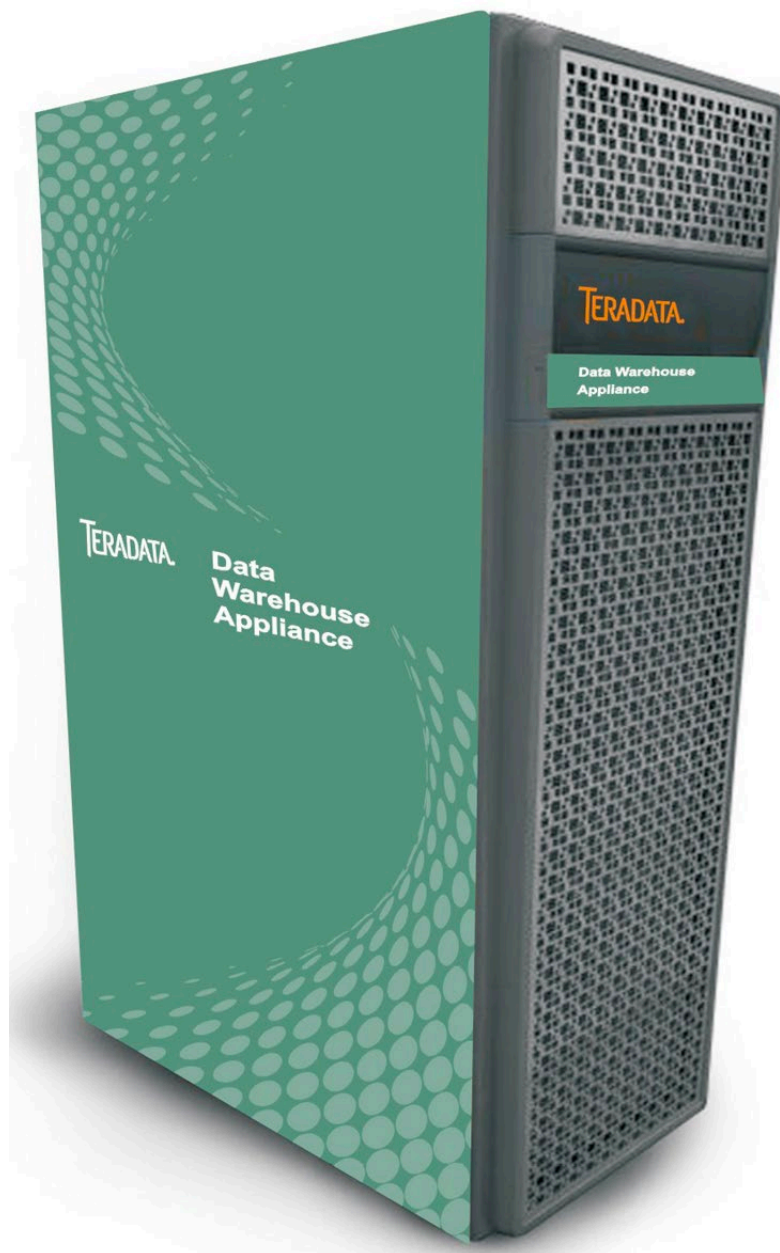
DVC/1012 1983 год



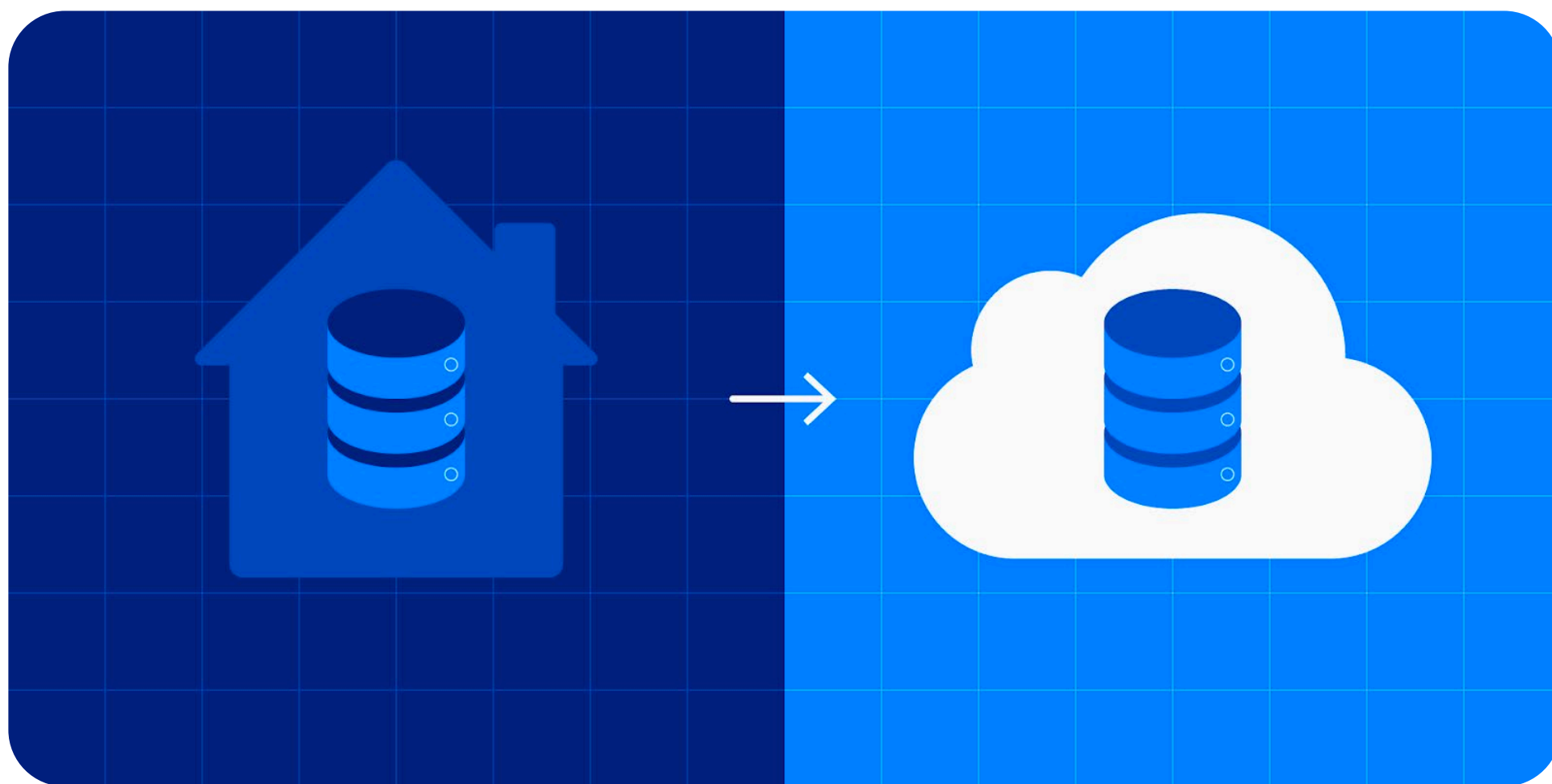
iPhone 2021 год



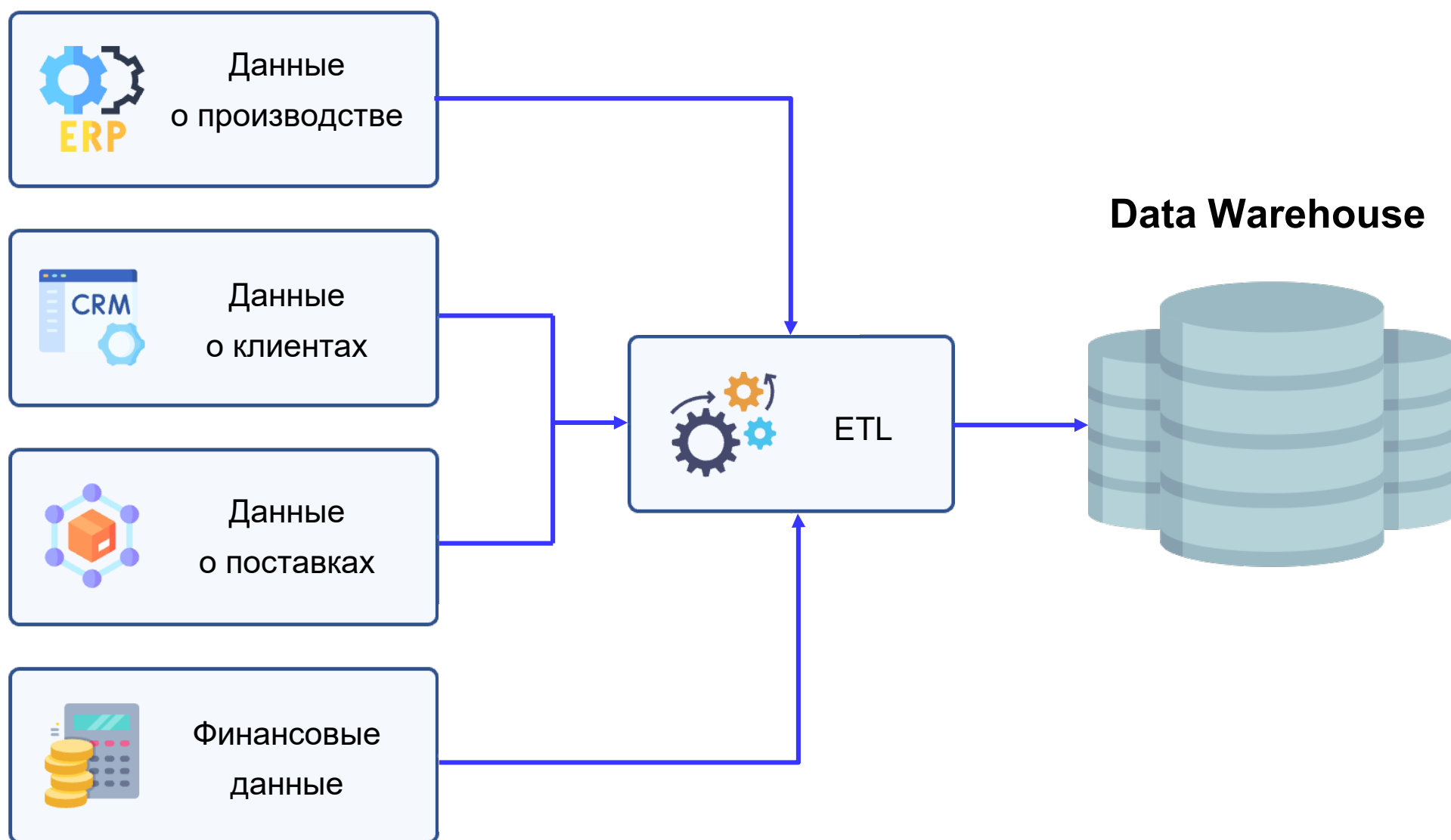
Teradata Data Warehouse



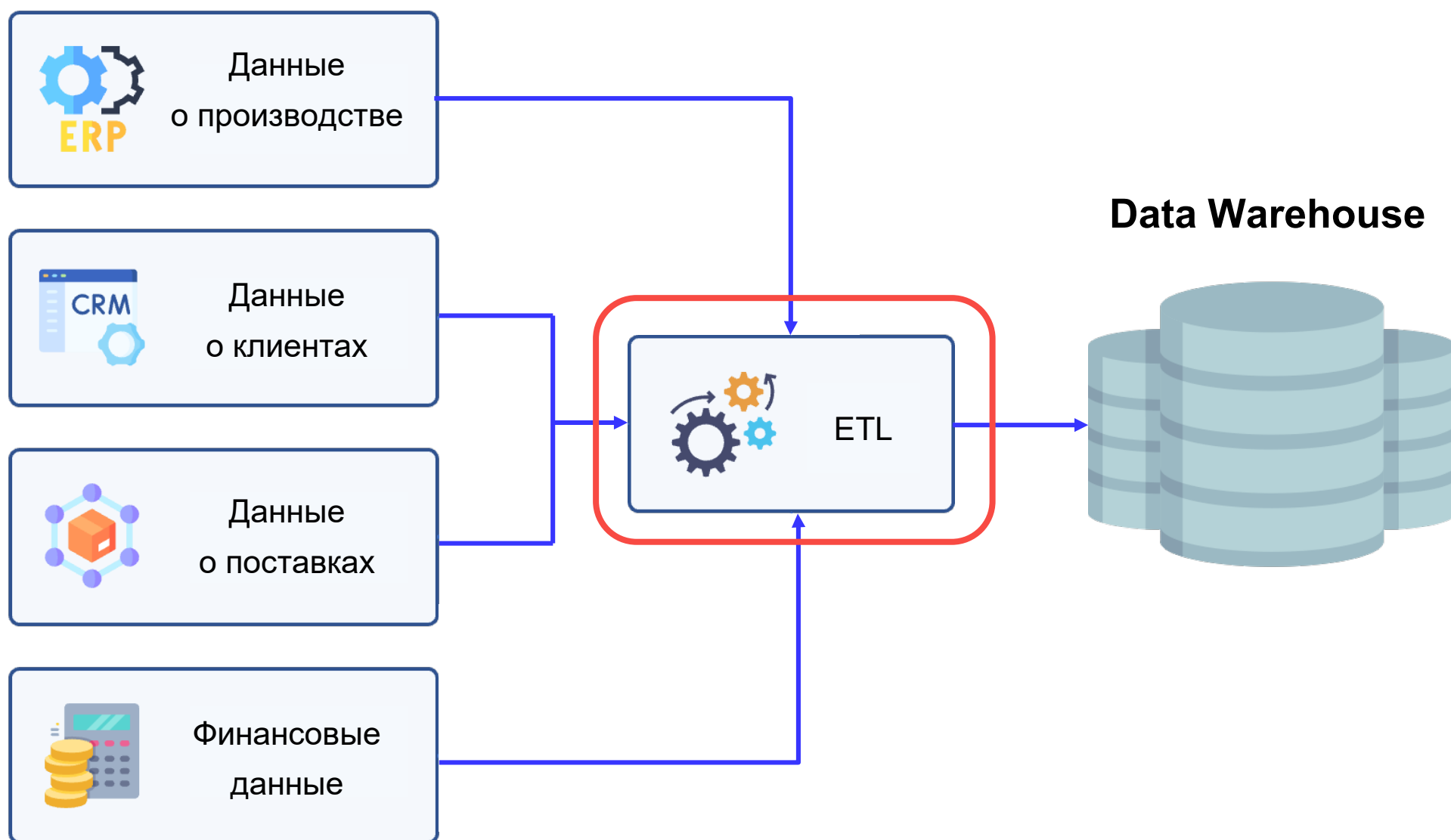
Облачные Warehouse



Хранилище данных



Хранилище данных



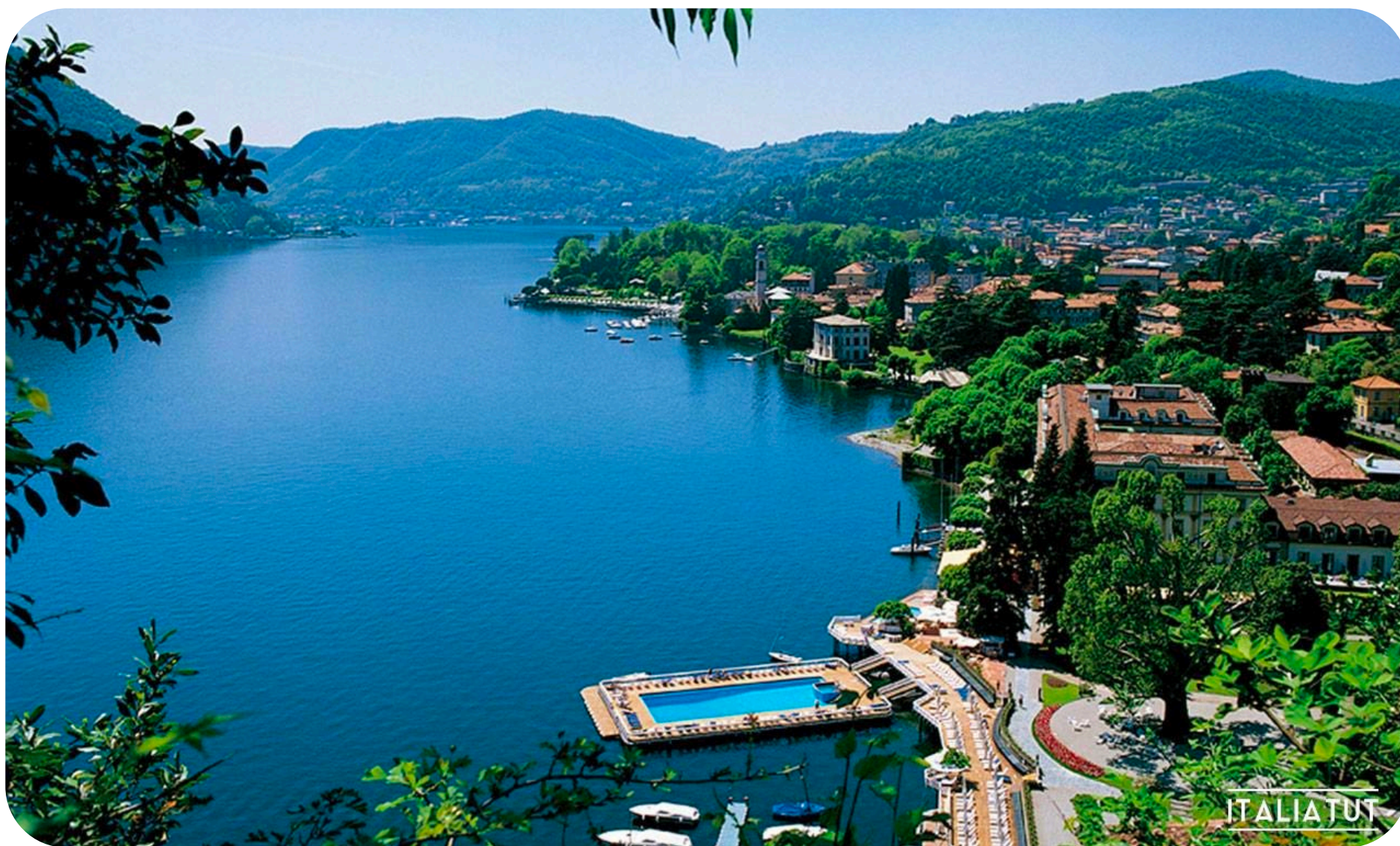
Некоторые данные

- 1 Плохо структурируются

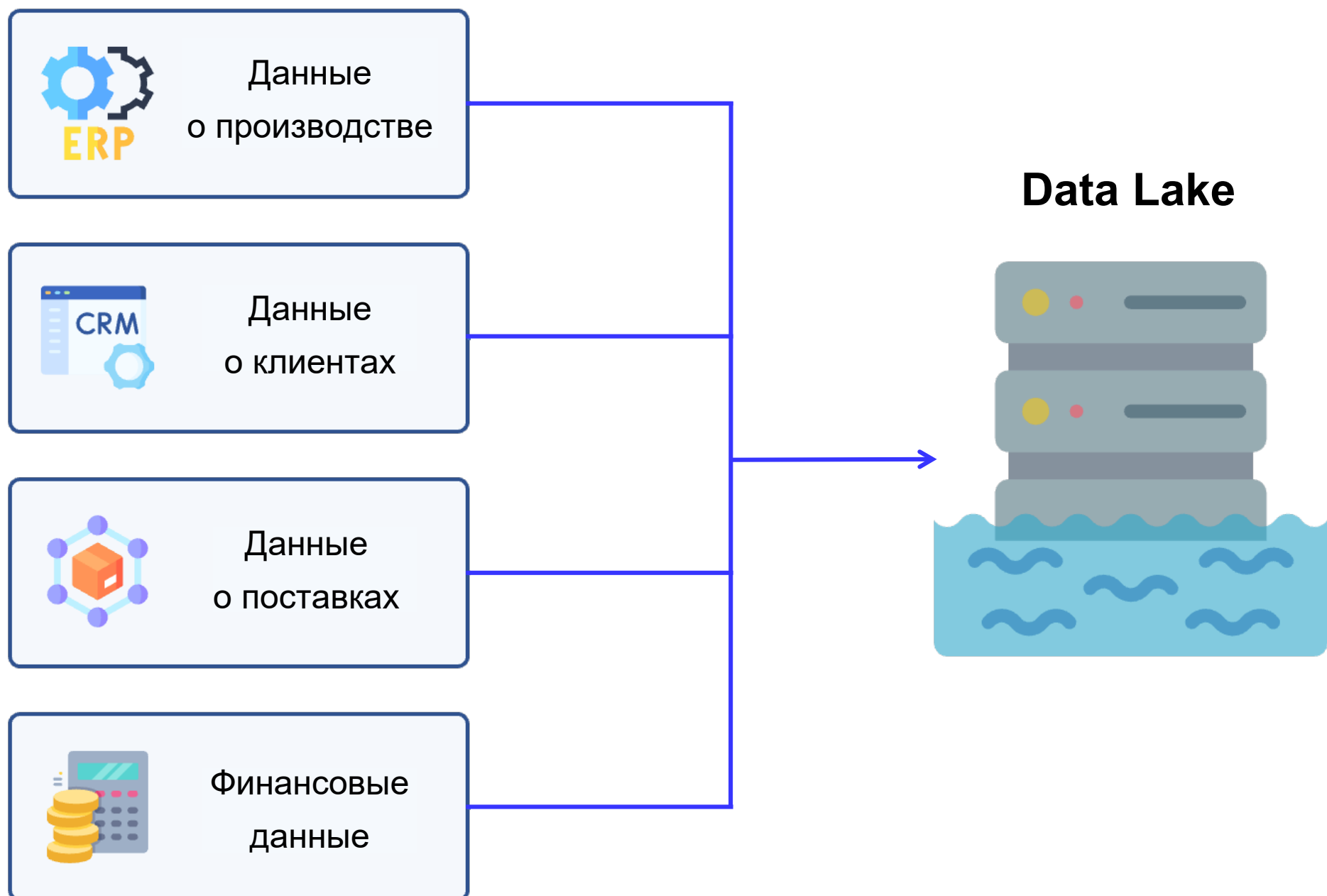
Некоторые данные

- 1 Плохо структурируются
- 2 Их очень много

Озеро (Lake)

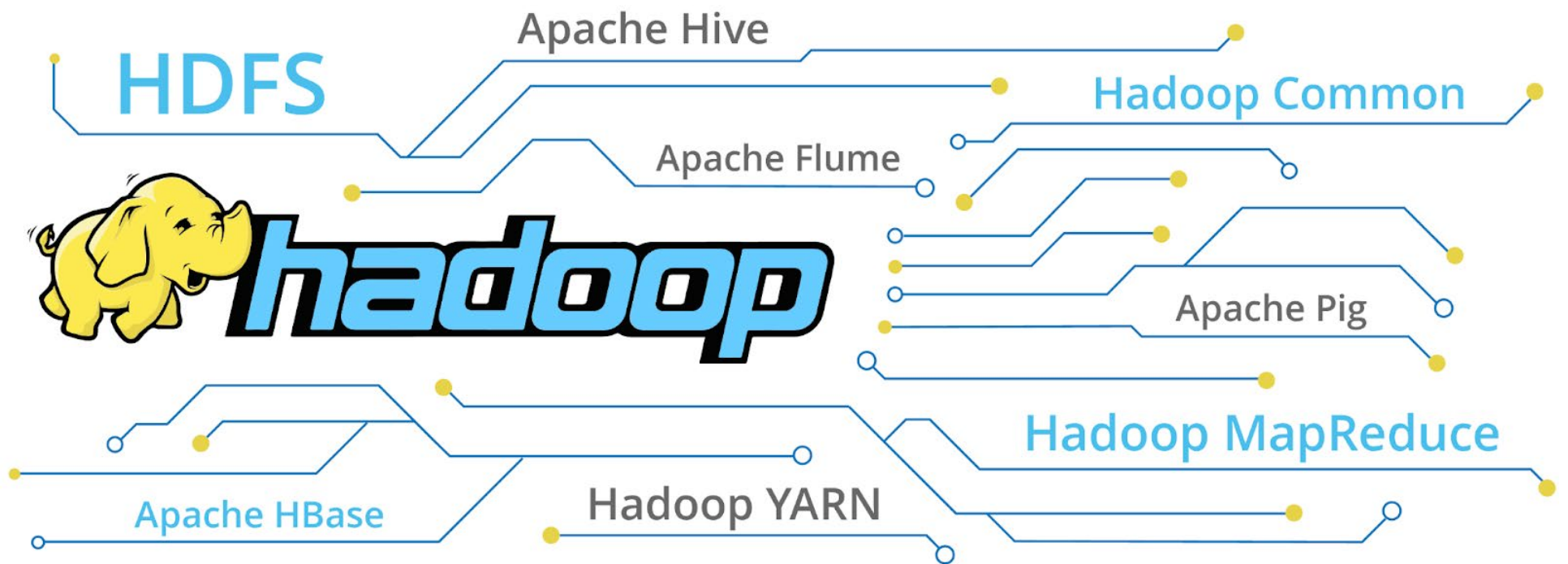


Озеро данных

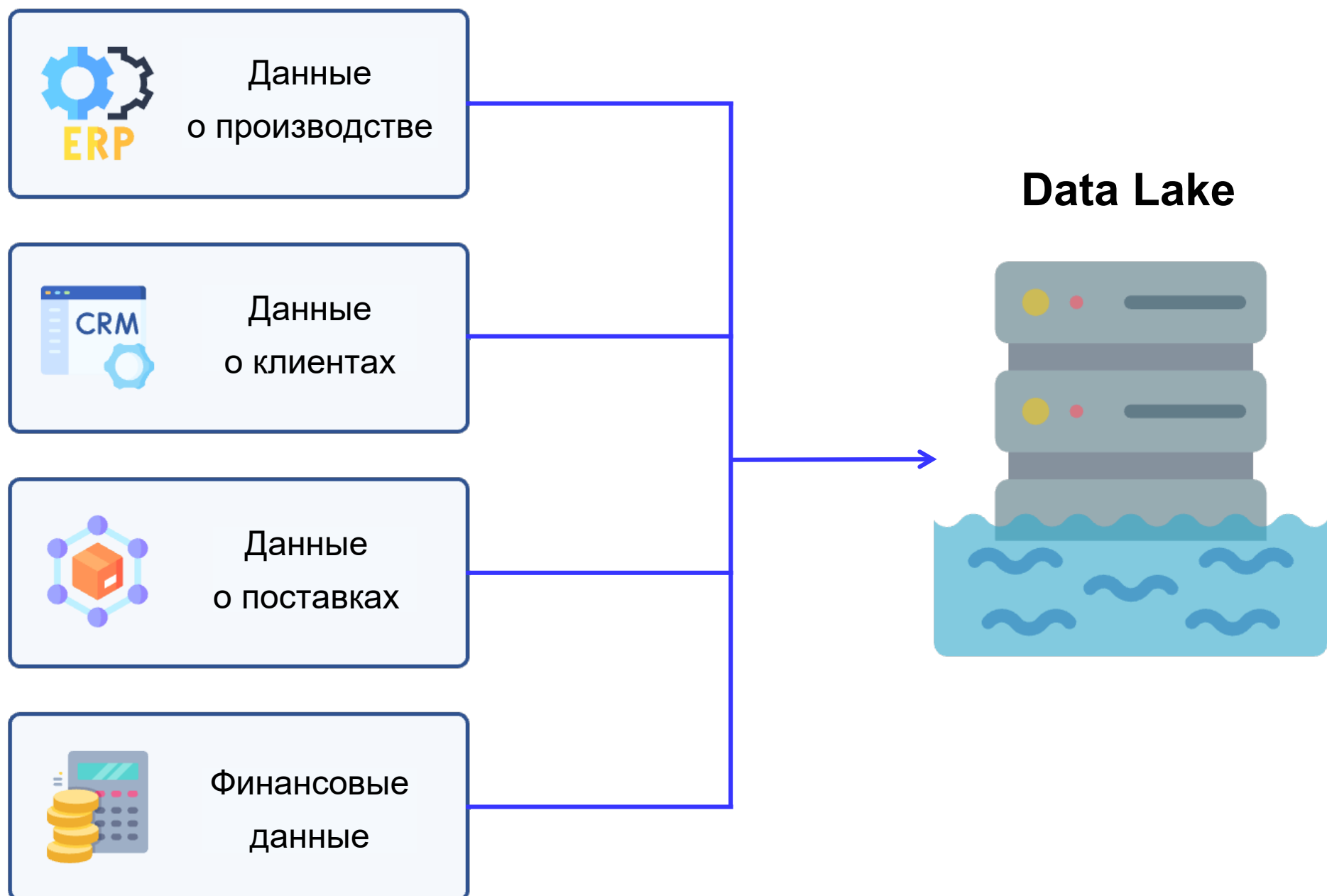


Кластер — группа компьютеров, представляющая для пользователя единый аппаратный ресурс.

Hadoop

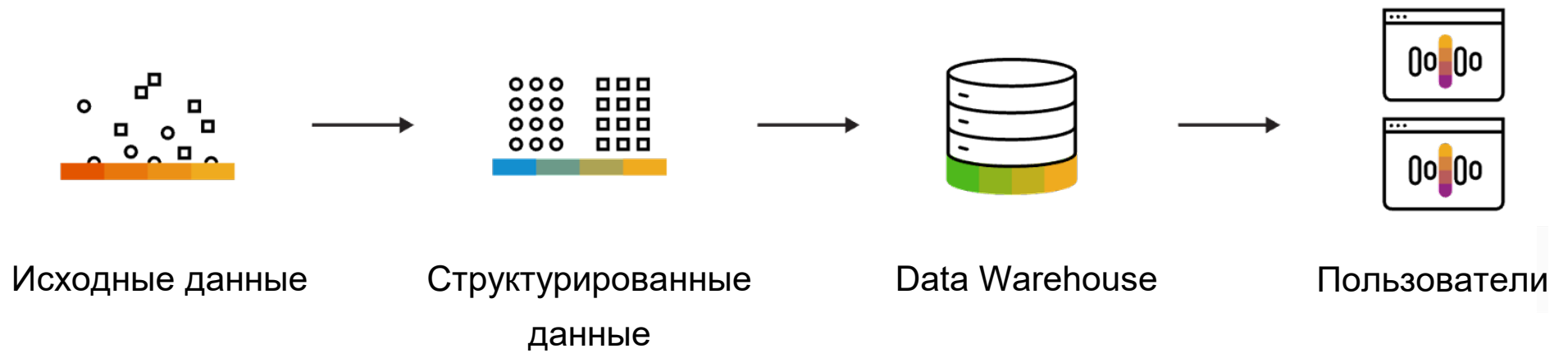


Озеро данных

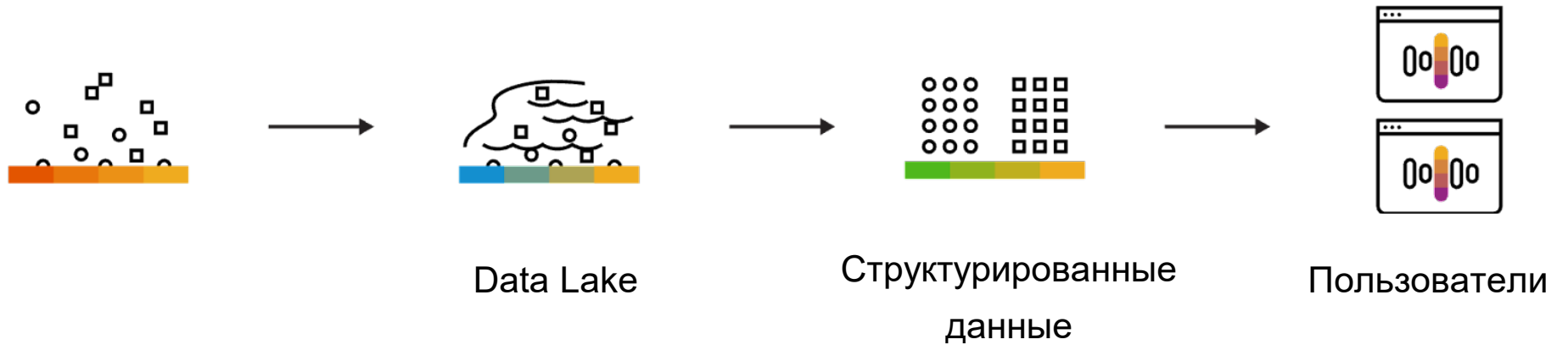


Lake vs Warehouse

Data warehouse



Data lake



Data Warehouse — это больше про людей
и классическую аналитику.

Data Lake — это больше про машинное обучение
и искусственный интеллект.

Озеро не болото!

