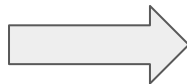
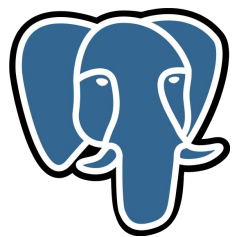


# Новая аналитическая платформа

# Текущее положение дел



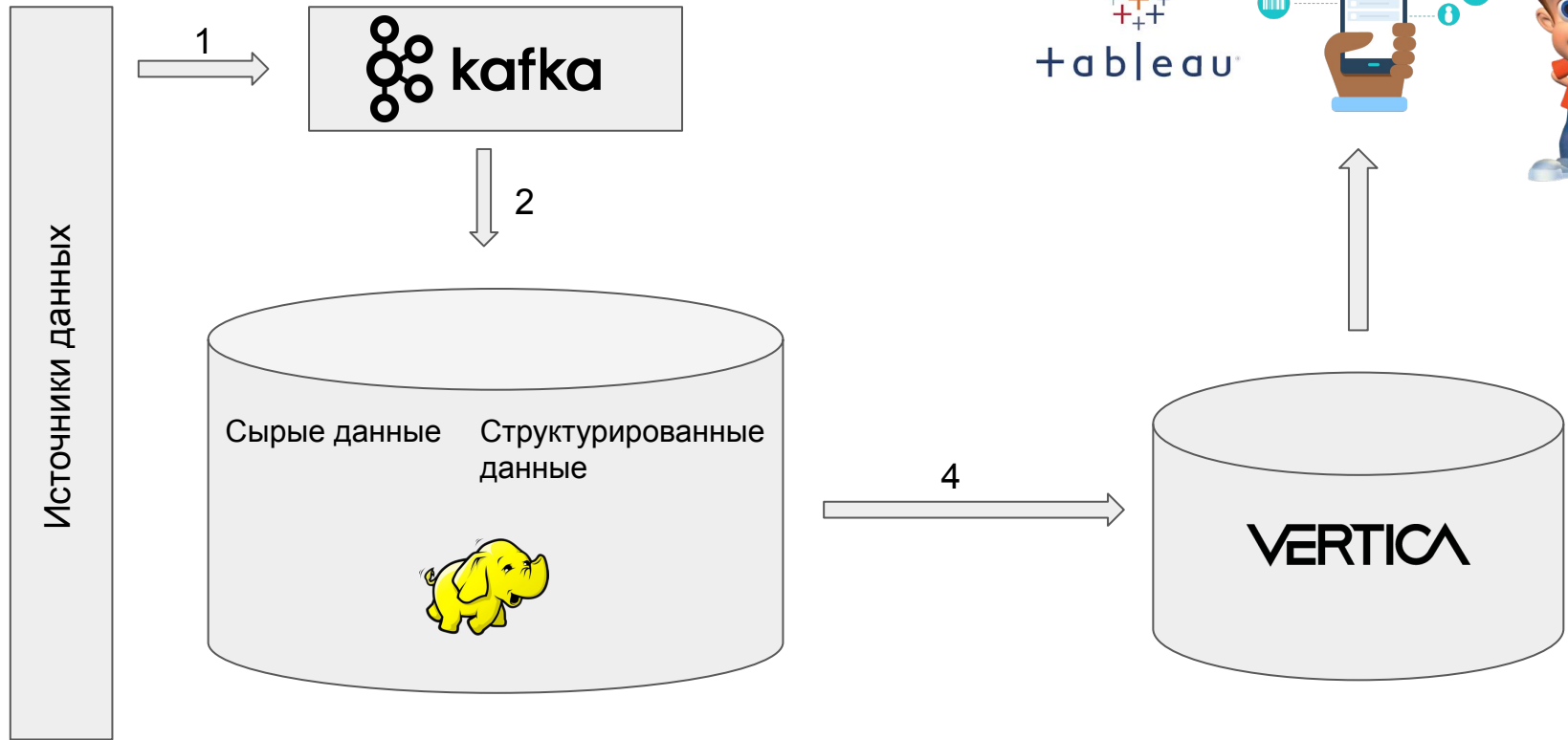
# Текущие проблемы

- 1) Сложная поддержка
- 2) Переизбыток данных на уровне MSBI
- 3) Частые сбои
- 4) Сложная интеграция новых источников данных (не SAP)
- 5) Нет реал тайма
- 6) Невозможность частичной интеграции (не SAP)
- 7) Долгие запросы
- 8) Нет доступа к данным через API(JDBC)
- 9) Отказоустойчивость
- 10) Один BI инструмент
- 11) Дата сайнтисты не удовлетворены

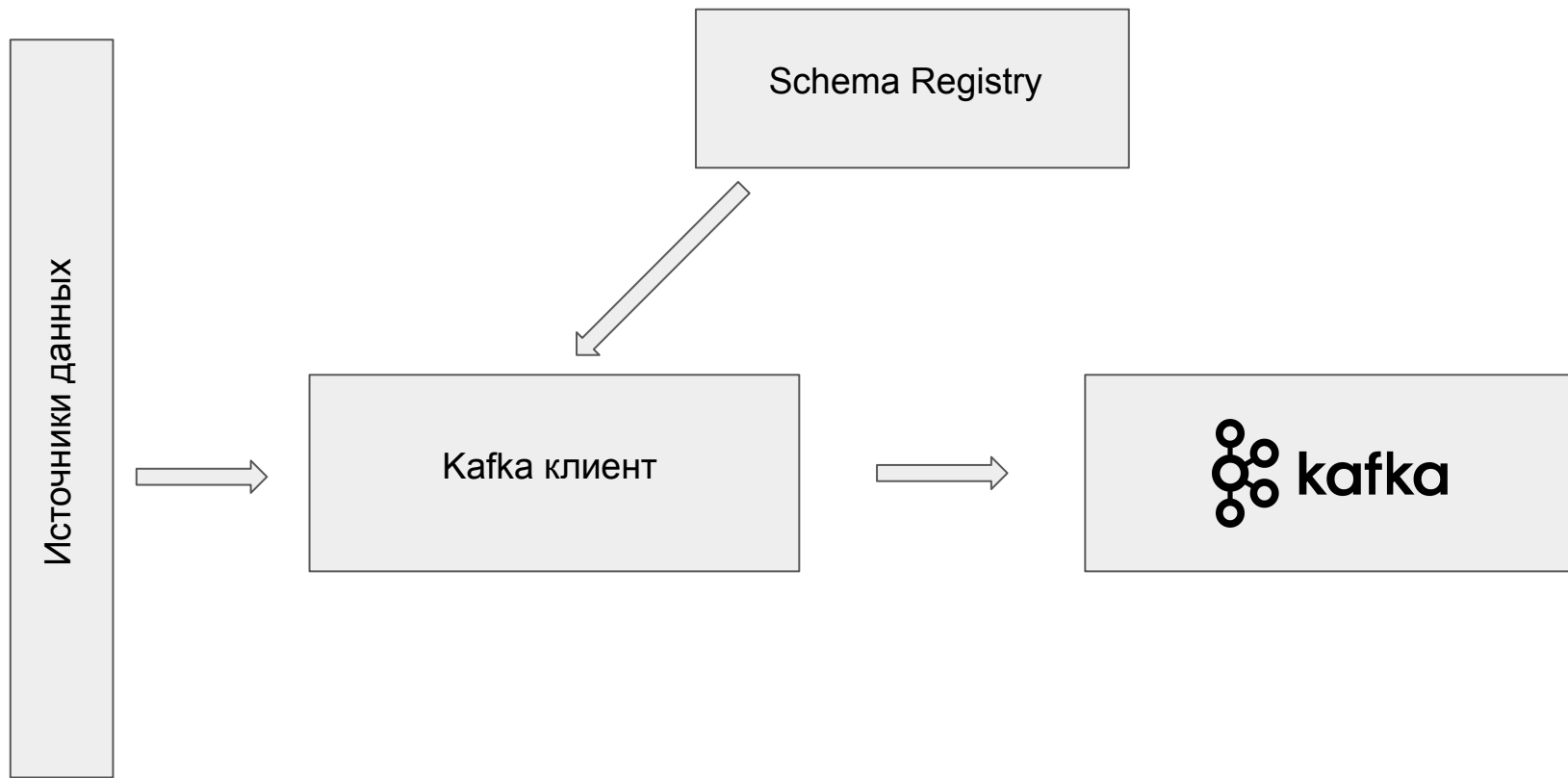
# Требования

- 1) Простое и быстрое подключение источников данных (сервисы, sap, google analytics)
- 2) Реал тайм аналитика
- 3) Скорость
- 4) Простая масштабируемость
- 5) Высокая отказоустойчивость
- 6) Быстрое восстановление после сбоев
- 7) Доступ по API
- 8) Удовлетворение датасайнтистов

# Вот так будет норм

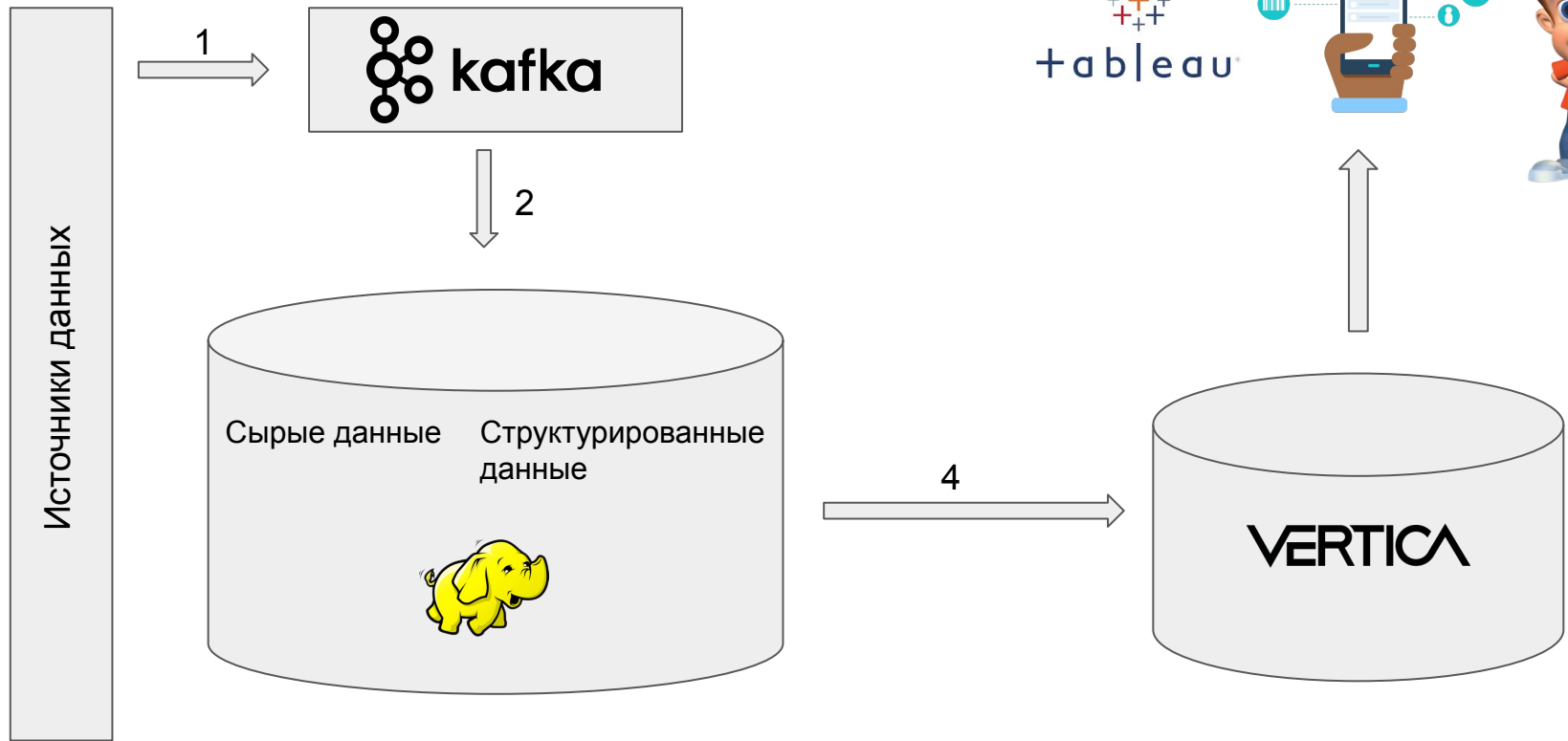


# Стрелочка 1



```
{
  "id":6666666,
  "state":"checkout",
  "orderChageAt":1566542606999,
  "userId":112039,
  "prices":{
    "total":1948.0
  },
  "positions":[
    {
      "id":6623453,
      "productId":12121196,
      "productId":"427324",
      "measureUnit":"ST",
      "quantity":1,
      "positionId":1,
      "price":{
        "base":1799.0,
        "red":null,
        "sale":1799.0
      }
    }
  ]
}
```

# Вот так будет норм



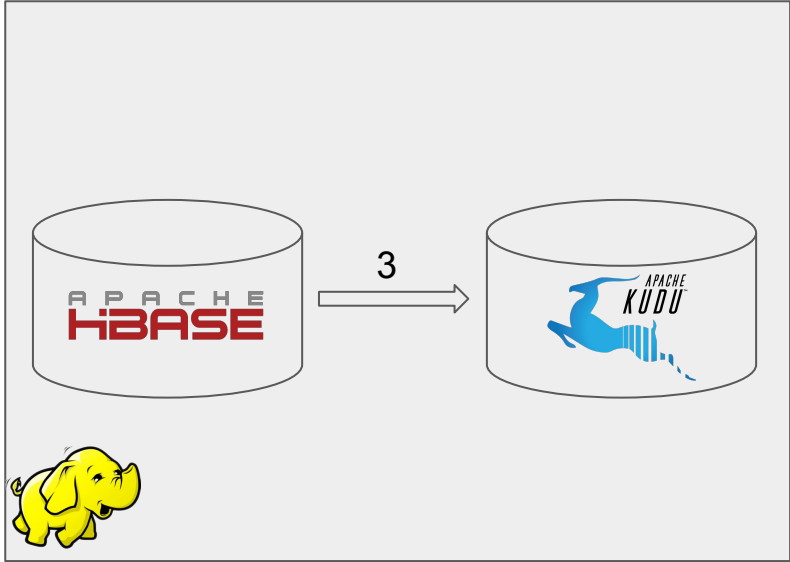
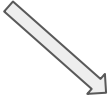


# Стрелочка 2

Schema Registry



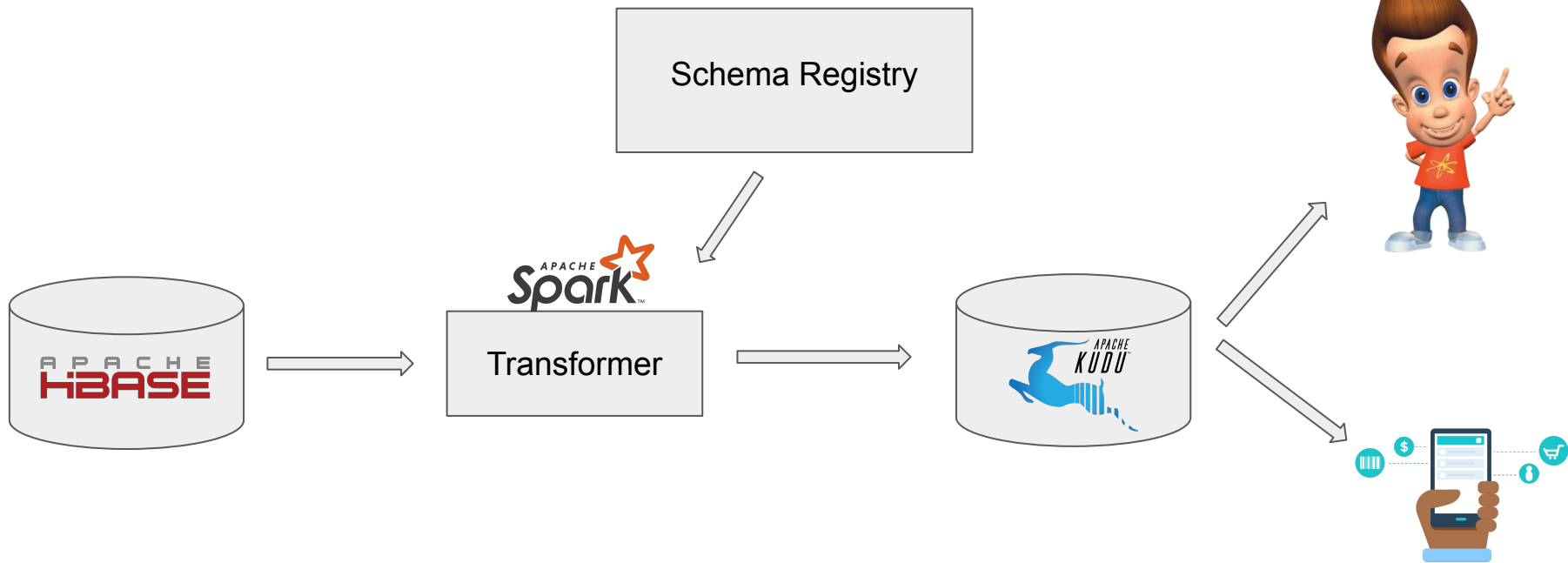
Analytic Receiver



3



# Стрелочка 3



ID	CHANGE_AT	DATA
hash(6666666)	1566542606999	{"id":6666666,"state":"check out" ...}



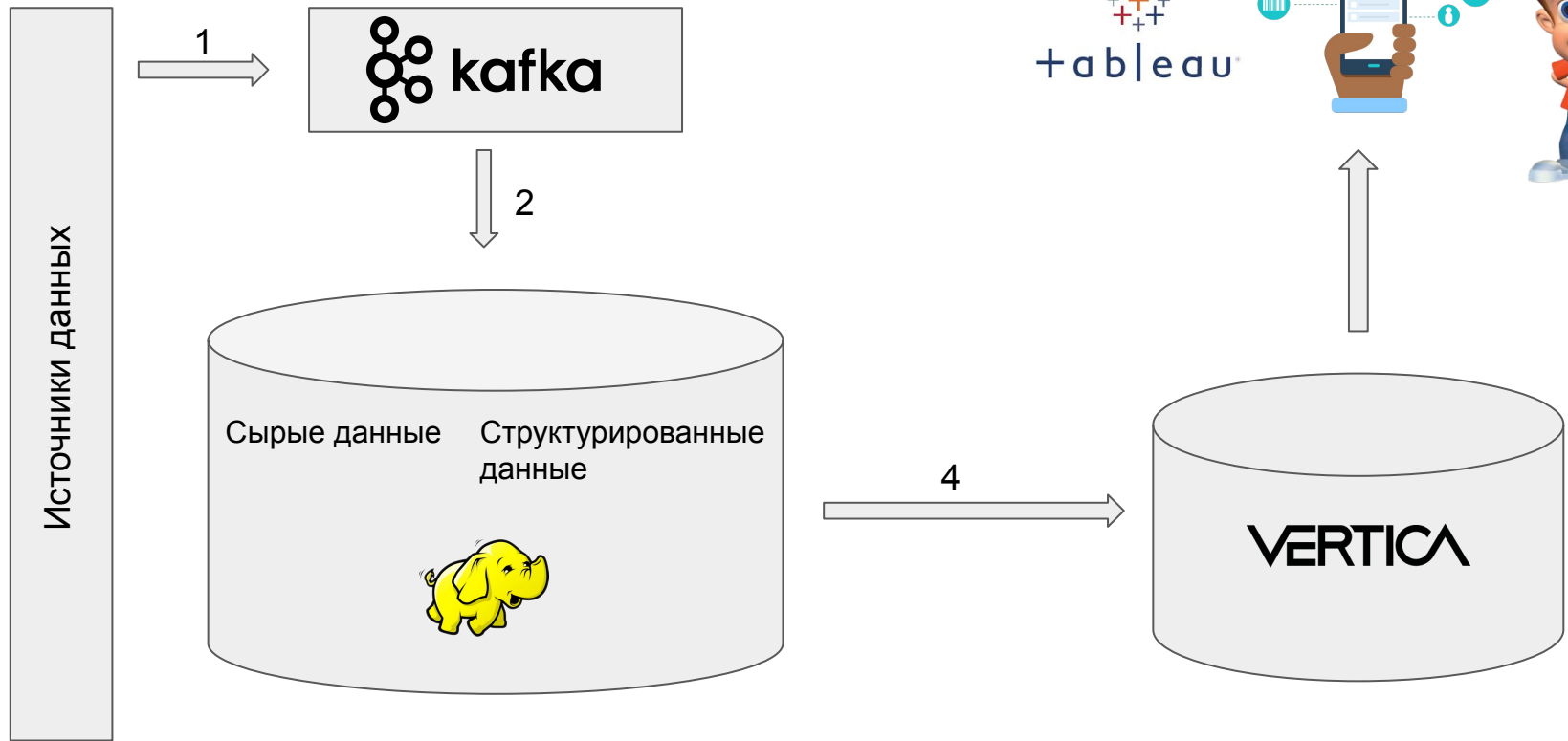
### ORDER

ID	CHANGE_AT	STATE	PRICE_TOTAL	USER_ID
6666666	1566542606999	checkout	1948	112039

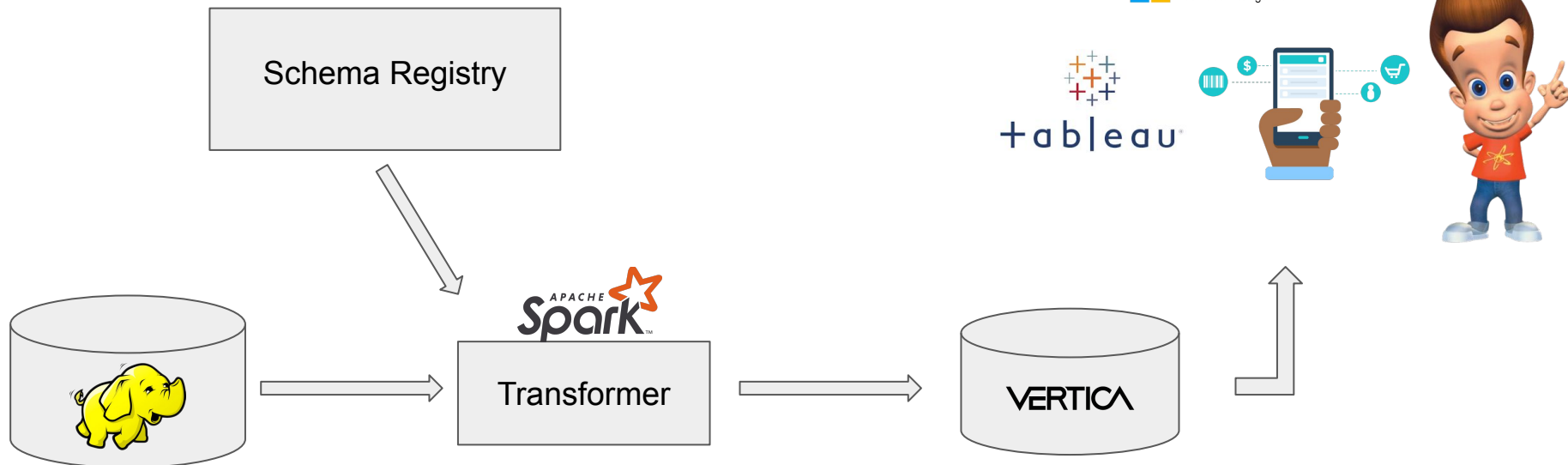
### ORDER\_POSITIONS

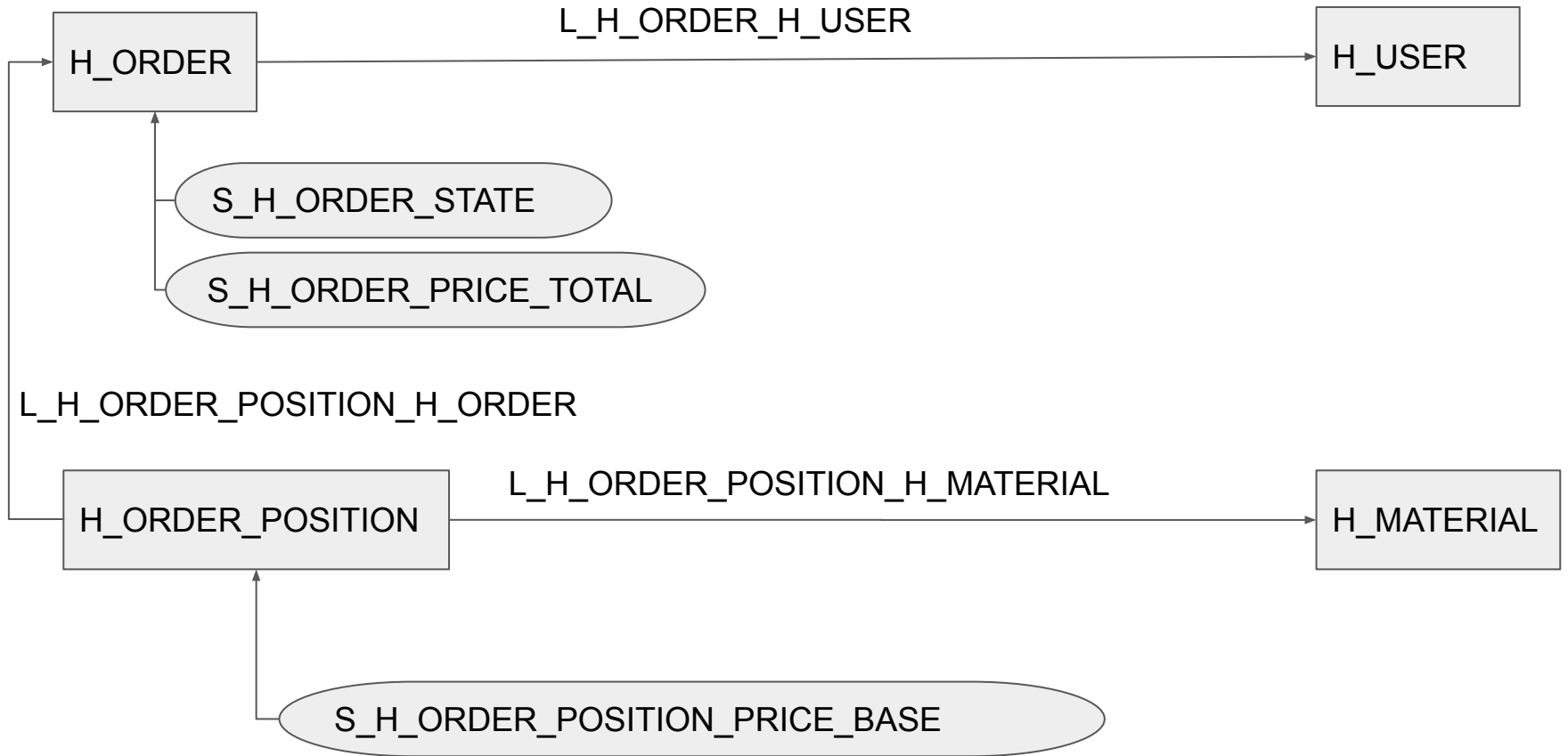
ID	ORDER_ID	PRODUCT_ITEM_ID	CHANGE_AT	PRICE_BASE
6623453	6666666	427324	1566542606999	1799

# Вот так будет норм



# Стрелочка 4





H\_ORDER

ID	ORDER_ID
1	6666666

S\_H\_ORDER\_STATE

H_ORDER_ID	STATE
1	checkout

H\_ORDER\_POSITION

ID	POSITION_ID
1	112039

S\_H\_ORDER\_POS\_PRICE\_BASE

H_ORDER_POS_ID	PRICE_BASE
1	1799

H\_MATERIAL

ID	PRODUCT_ITEM_ID
1	427324

S\_H\_MATERIAL\_NAME

H_MATERIAL_ID	NAME
1	Цемент

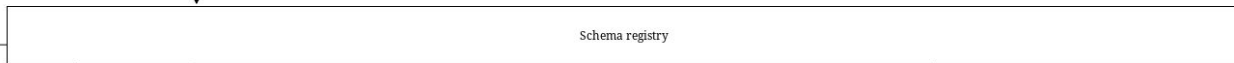
L\_H\_ORDER\_POS\_H\_ORDER

H_ORDER_POS_ID	H_ORDER_ID
1	1

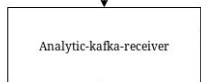
# Первичные настройки

Необходимы для интеграции / обработки.  
Включают в себя:

- 1) Наименование источника (producer)
- 2) Наименование объекта интеграции (integration\_object)
- 3) Заведение версии, ключей и схемы объекта

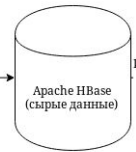


Первичные настройки (получение аналогично kafka-client)

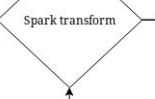


Настройки трансформации (при запуске приложения)

Сырые данные



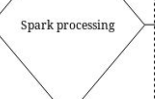
Измененные сырые данные



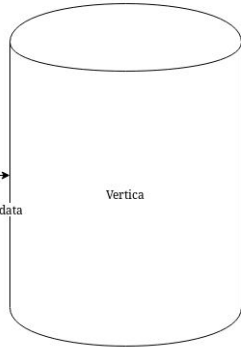
Преобразованные данные



Полуструктур. данные



Anchor model data



# Настройки трансформации

Включают в себя описание полей и структур:

- 1) Поле куда (имя поля, тип)
- 2) Информация, откуда взять (либо явное поле в сыром объекте, либо функция преобразования)
- 3) Если структура сложная, то можно указать разделение на таблицы

Дополнительно, с каждым описанием трансформации идут первичные настройки



# Что еще нужно сделать

- 1) Мониторинг всего и вся
- 2) Админка schema registry
- 3) SDK
- 4) Получатель данных из DWH
- 5) Переезд на локальное железо
- 6) Механизм сложных трансформаций и расчетов
- 7) Перевод всех источников на новую платформу
- 8) Частичное обновление