The top half of the slide features a dark blue background with a complex, abstract pattern of light blue particles and lines, resembling a network or data flow. The particles are concentrated in some areas and sparse in others, creating a sense of depth and movement.

金数据的数据库架构方案

—尉刚强



金数据

人人可用的数据平台，助力工作，成就商业。

金数据，人人可用的数据平台，帮助个人和企业客户快速设计表单，搭建业务数据系统，如调查、预约、登记、获客、抽奖、考试、订单等等。金数据收集和管理日常工作和业务运转中的数据，助力个人工作效率提升，成就企业商业价值。

自2012年上线以来，金数据已经为超过400万的个人和企业客户，提供高效、安全、可靠的数据服务，赢得了客户的认可，帮助客户获得成功。

提纲

- 金数据的业务特性
- 数据库选型
- 金数据数据架构方案

基本概念介绍

表单

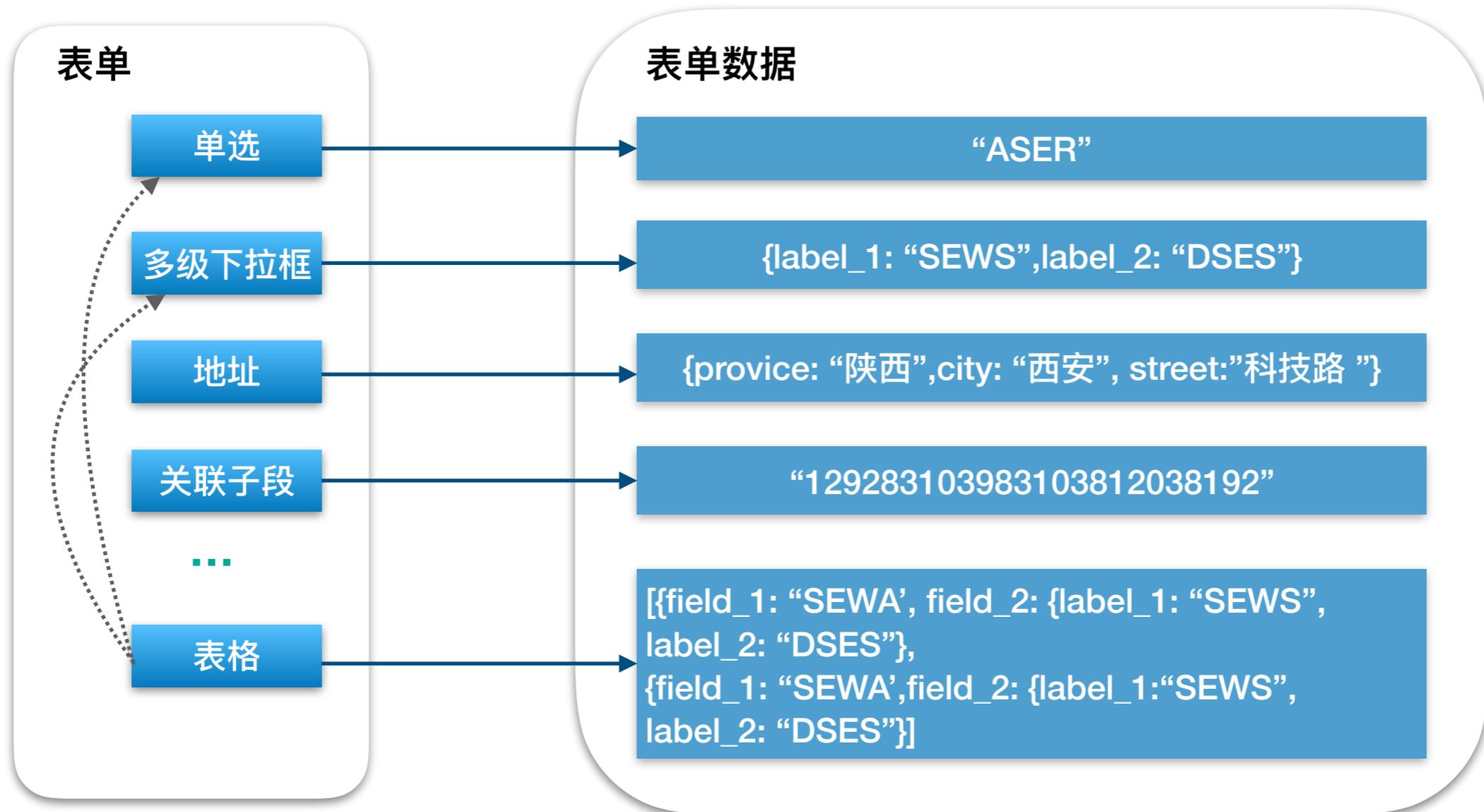
- 可以是一个在线的调查问卷，或者在线投票，每个表单中包含很多题目，可以是选择题、填空题、下拉菜单题目等，用户可以填写并提交数据。

表单数据（表单记录）

- 代表用户填写表单的记录，包括每个题目用户提交的内容，例如选项值，文本内容等。

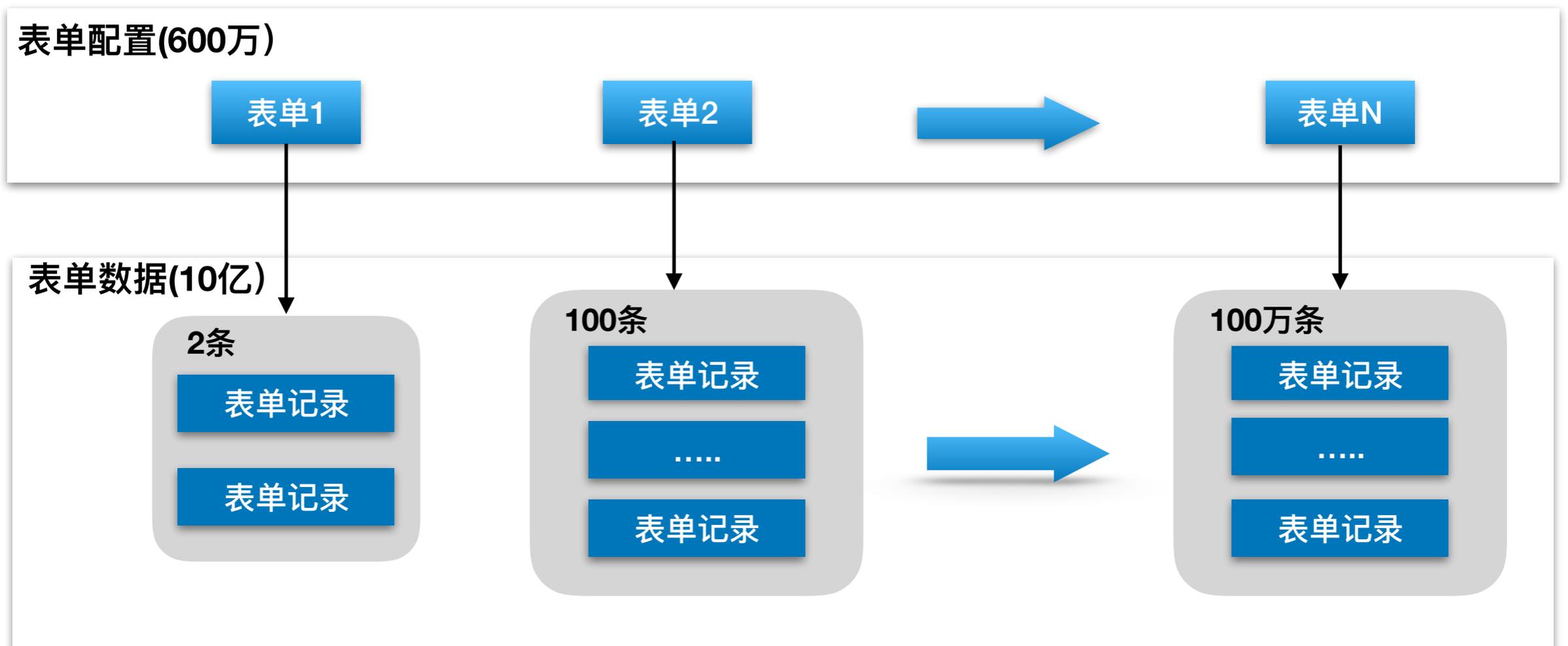
数据格式详解

- 目前表单包括字段类型约30种，后续可能还会增加。
- 其中有些字段类型是组合类字段，可以由多个其他类型组合。
- 每个表单包含字段个数不限制，同一个字段类型的个数也不限制。



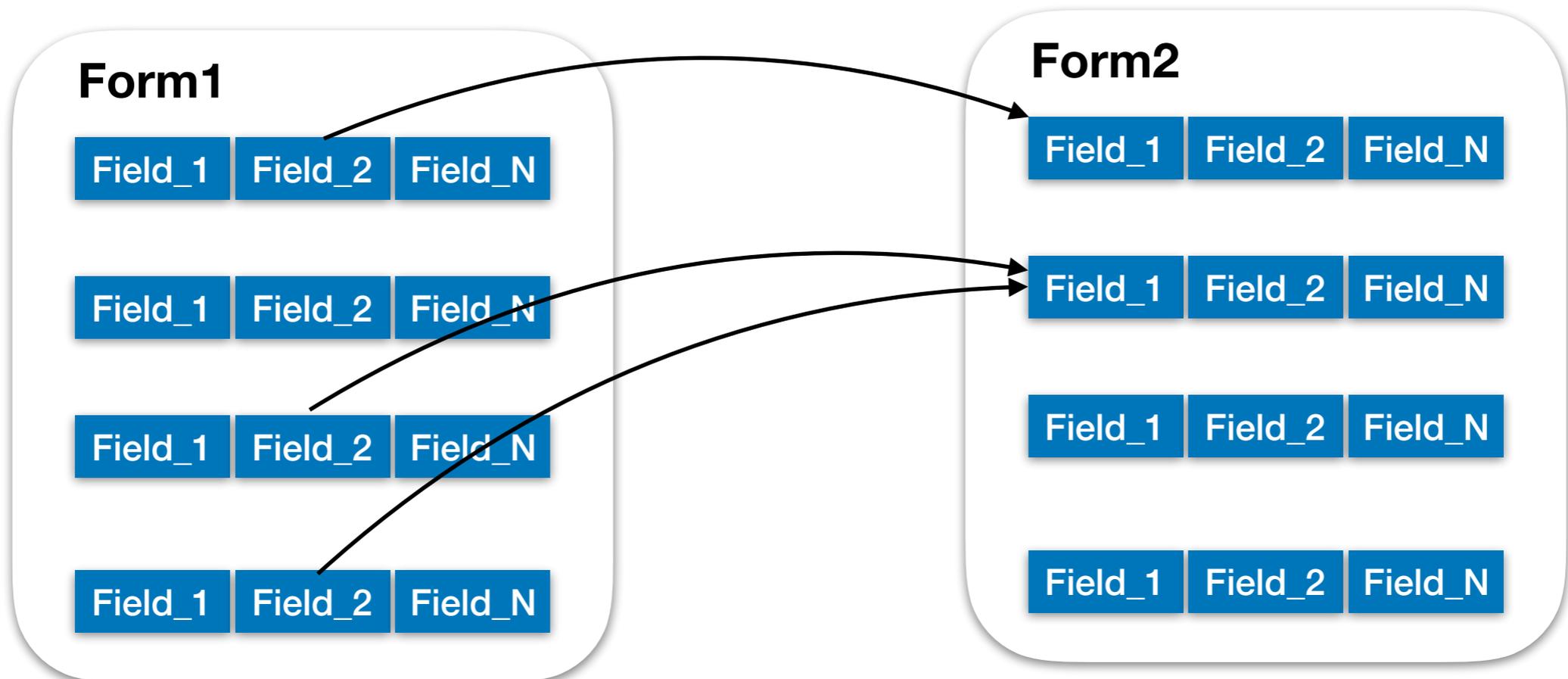
数据存储

- 数据保存在mongoDB中
- 表单配置的元数据保存在一个collection中，约600万
- 所有的表单数据保存在一个collection中，10亿规模



表单关联

- 一个表单中的某个字段可以关联另外一个表单中的一个字段
- 一个表单可以同时关联一个表单的多个字段，或者多个表单的多个字段
- 表单关联没有方向限制。



业务逻辑

- 100W记录以上的表单，包含多个条件和关联子段组合查询在1S内完成。
- 100W记录以上的表单，包含多个条件和关联子段聚合计算在1S内完成。

组合查询



1.子段1单选为A，子段5多选包含B和C，子段11表格中第一列选择F，同时文本输入包含“天空”的前100条记录和总行数。
2.子段2单选为C，子段3关联字段的另外一个表格的字段7的选项为C的前100条记录和总行数

汇聚计算



1. 字段1单选为C，字段7表格第二列选择为F的情况下，表格第三列各个选项的汇总数据。
2. 子端1单选为D，地段8地址在陕西西安，子段10关联的表单中第5字段数字收入的最大值最小值，平均值。

金数据的业务需求

对客户提供对外查询和报表，
并发度高，不完全可plan

插入的数据可以及时的查询出来

高并发

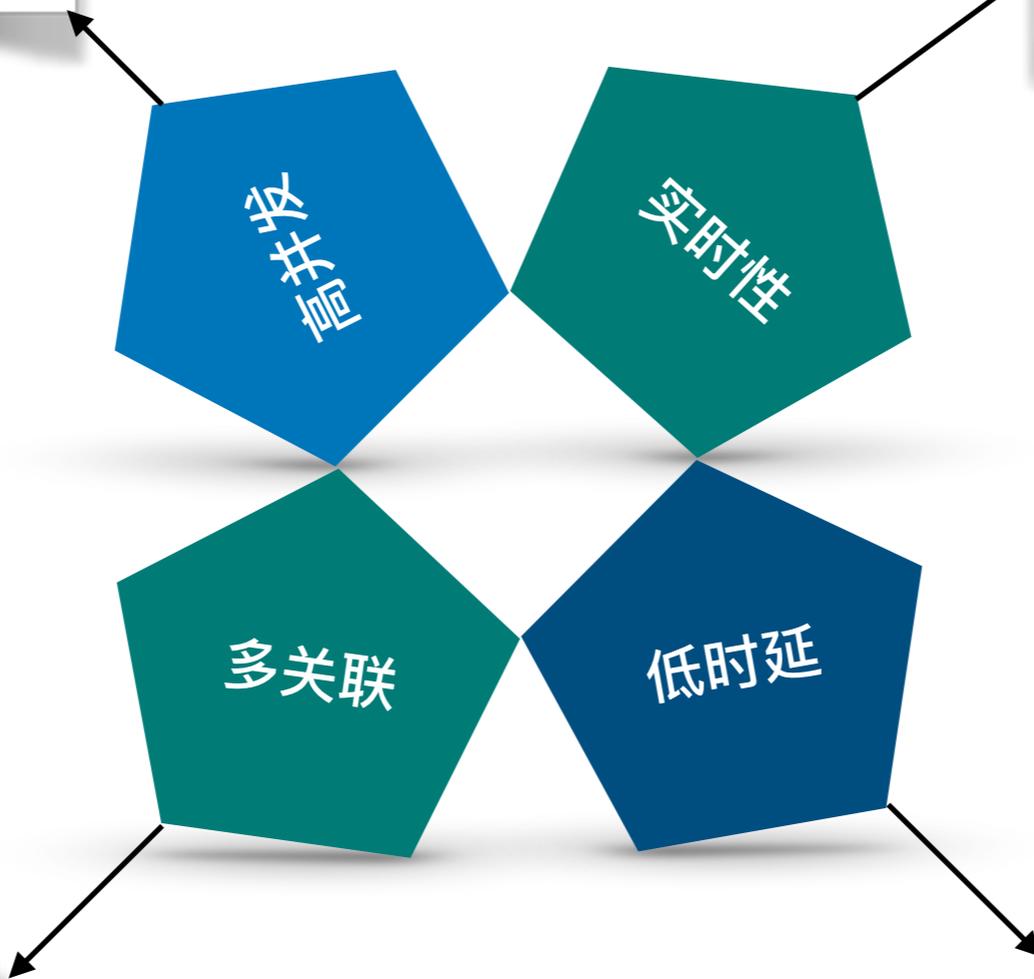
实时性

多关联

低时延

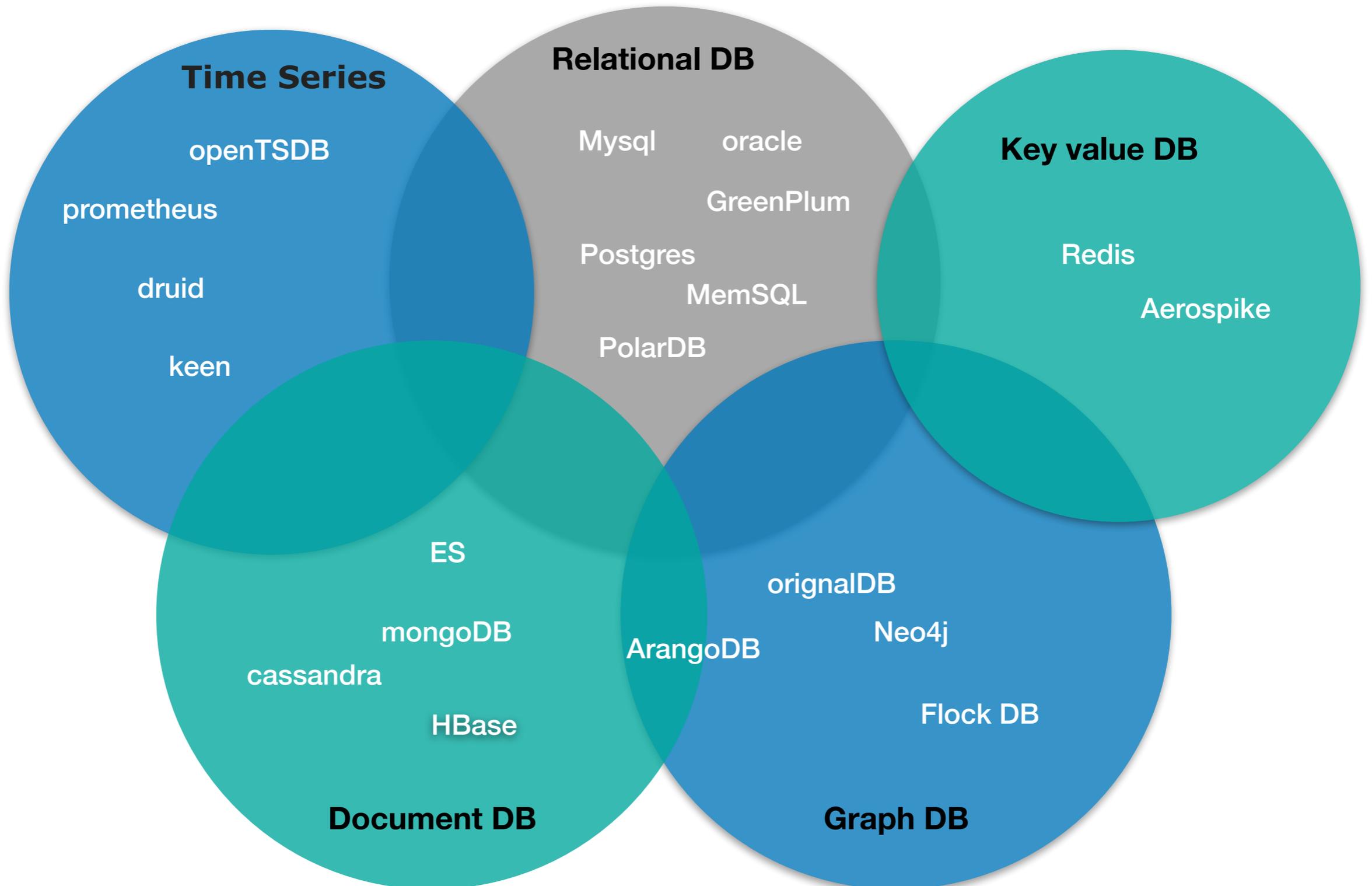
表单字段关联关系复杂

数据分析页面可以急速响应

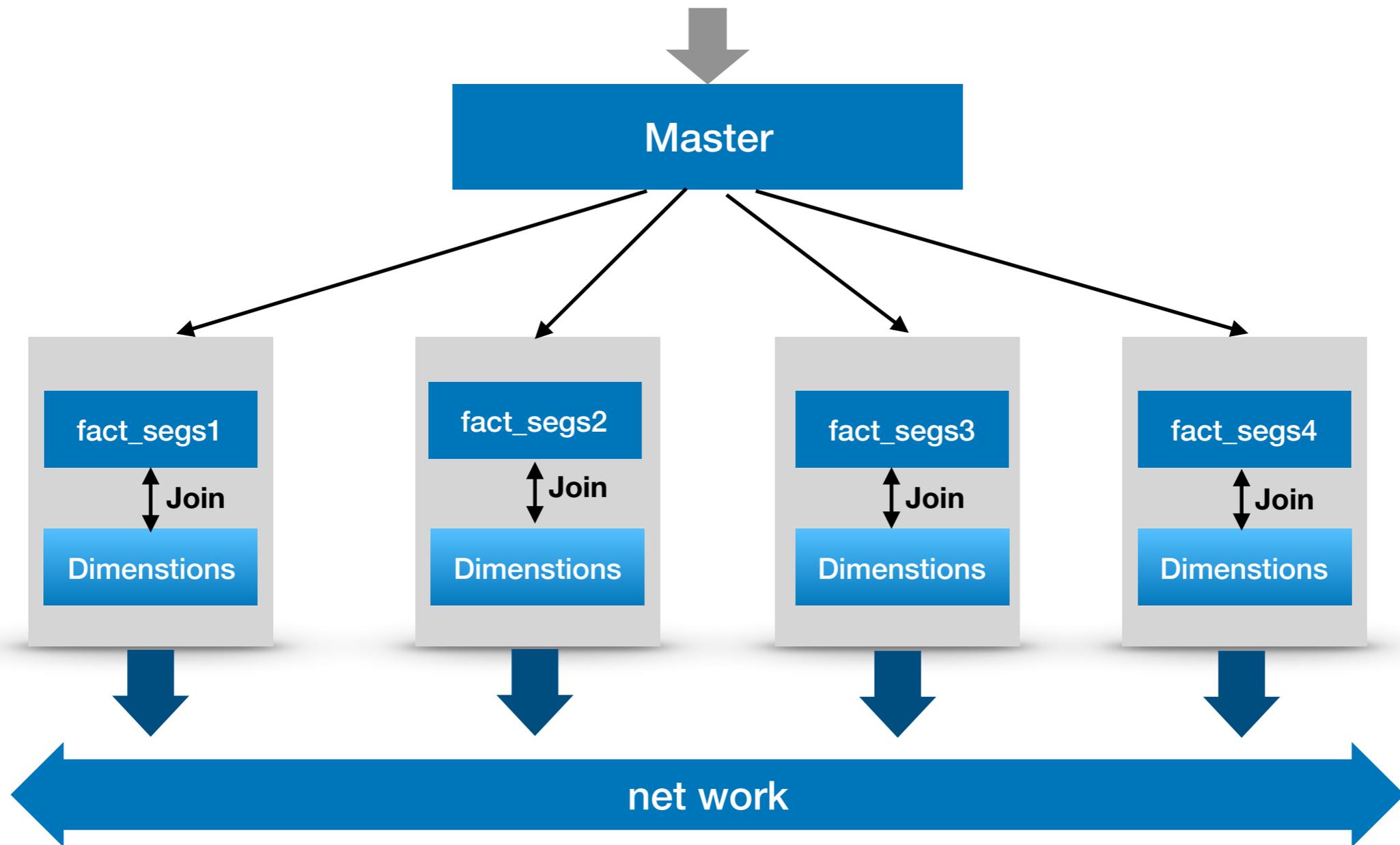


数据库选型

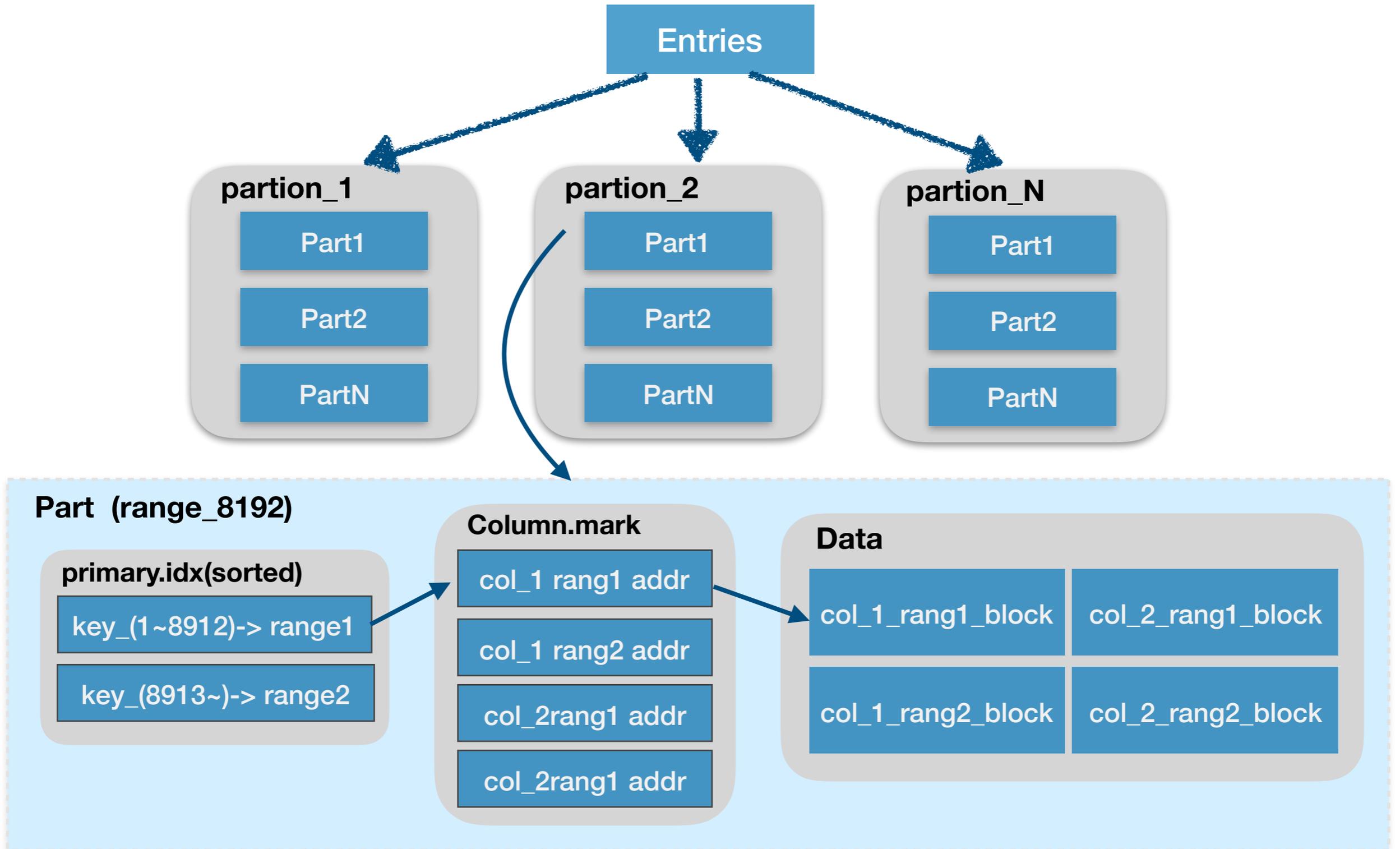
数据库选型



主流分析型数据架构

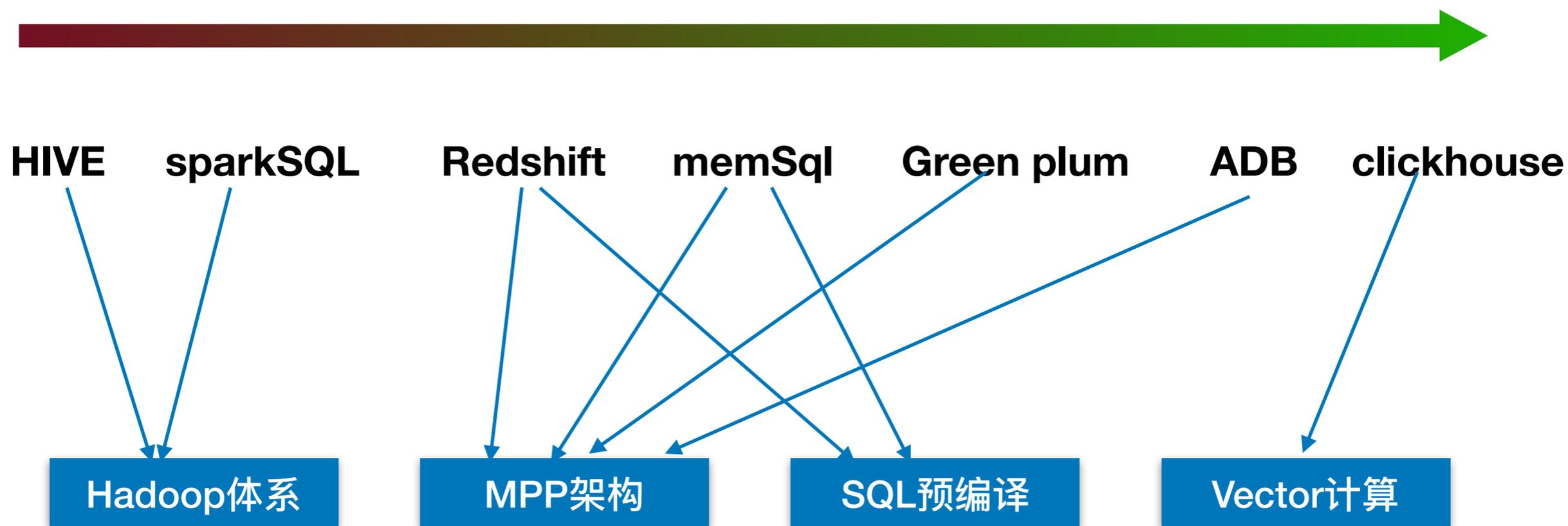


clickhouse存储架构



基于关系数据库的分析计算

启动速度



数据选型结果

Mongo

```
graph LR; Mongo[Mongo] --> Adb[Adb]; Mongo --> Clickhouse[Clickhouse];
```

Adb

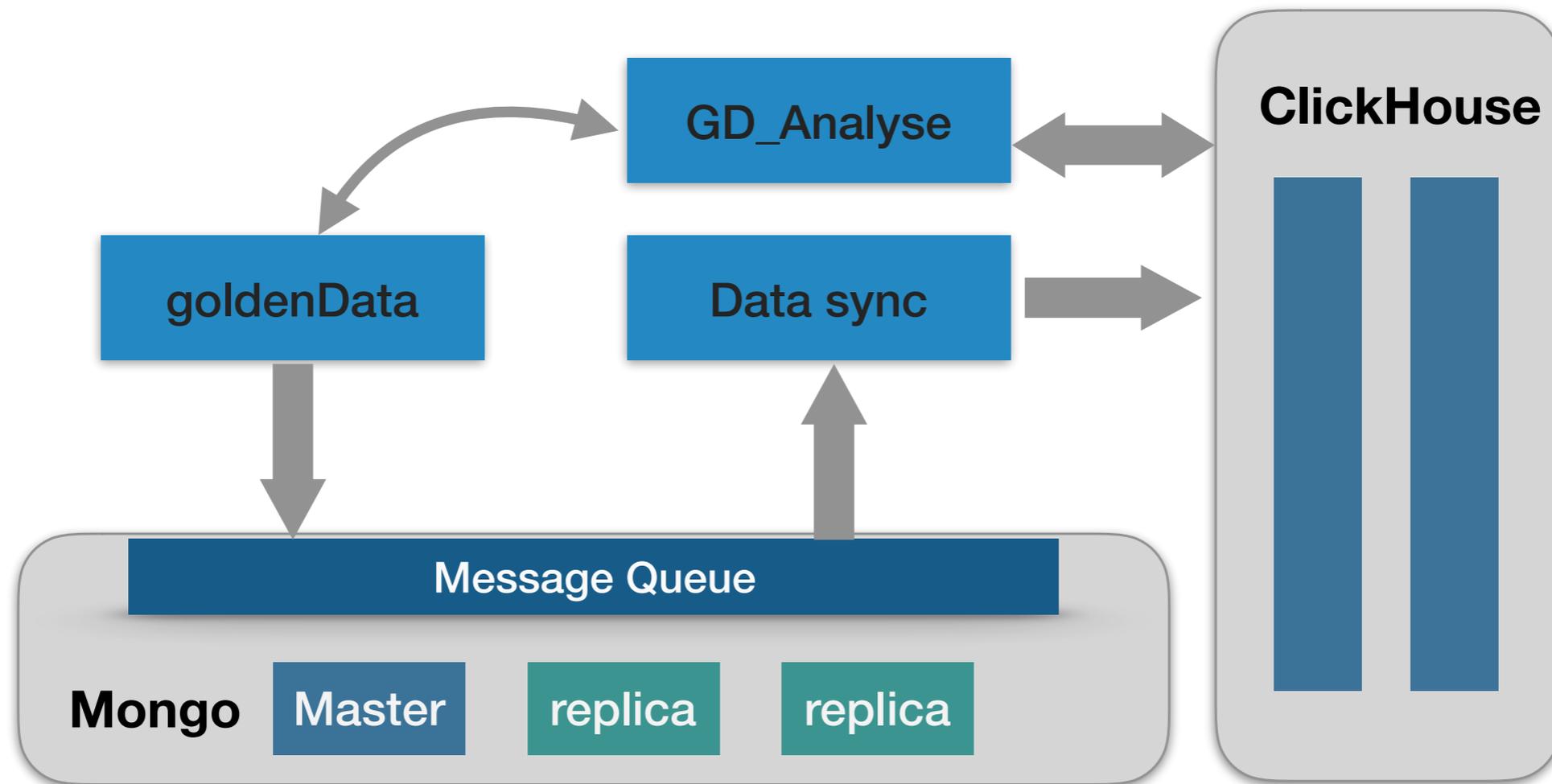
- 快速的分析查询能力（万亿数据秒内）
- 并发查询和写入能力强
- 私用化部署不方便
- 阿里云和AWS专线成本高
- Key的唯一性可以保证
- 事务能力弱
- 关联查询性能较弱

Clickhouse

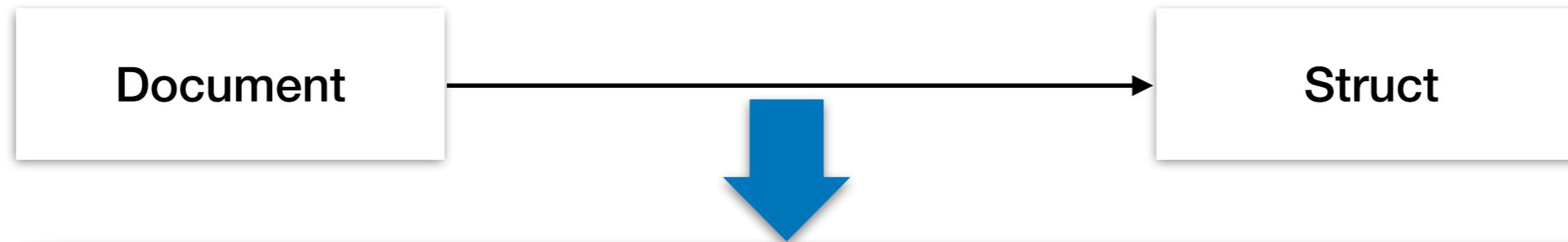
- 快速的分析查询能力（万亿数据秒内）
- 并发查询和写入能力较弱
- 私用化部署方便
- 没有专项经费
- 不保证key的唯一性
- 事务能力弱
- 关联查询能力较好

金数据数据架构方案

金数据业务优化架构



数据字段的映射



MongoDb

["SEWS", "DSES"]

SqlDB

"SEWS,DSES"

Clickhouse

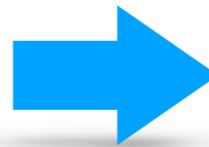
```
select choice, count(*) from entries_test  
ARRAY JOIN splitByChar(',', field_1) as  
choice group by choice
```

ADB

```
with T2 as (select split_part (a.field_1, ',', b.subindex) as  
choice, 1 as count from `entry_0002` as a join incre_table  
as b on b.subindex < (length(a.field_1) -  
length(replace(a.field_1, ','))+1)) select choice, COUNT(*)  
FROM T2 GROUP BY choice
```

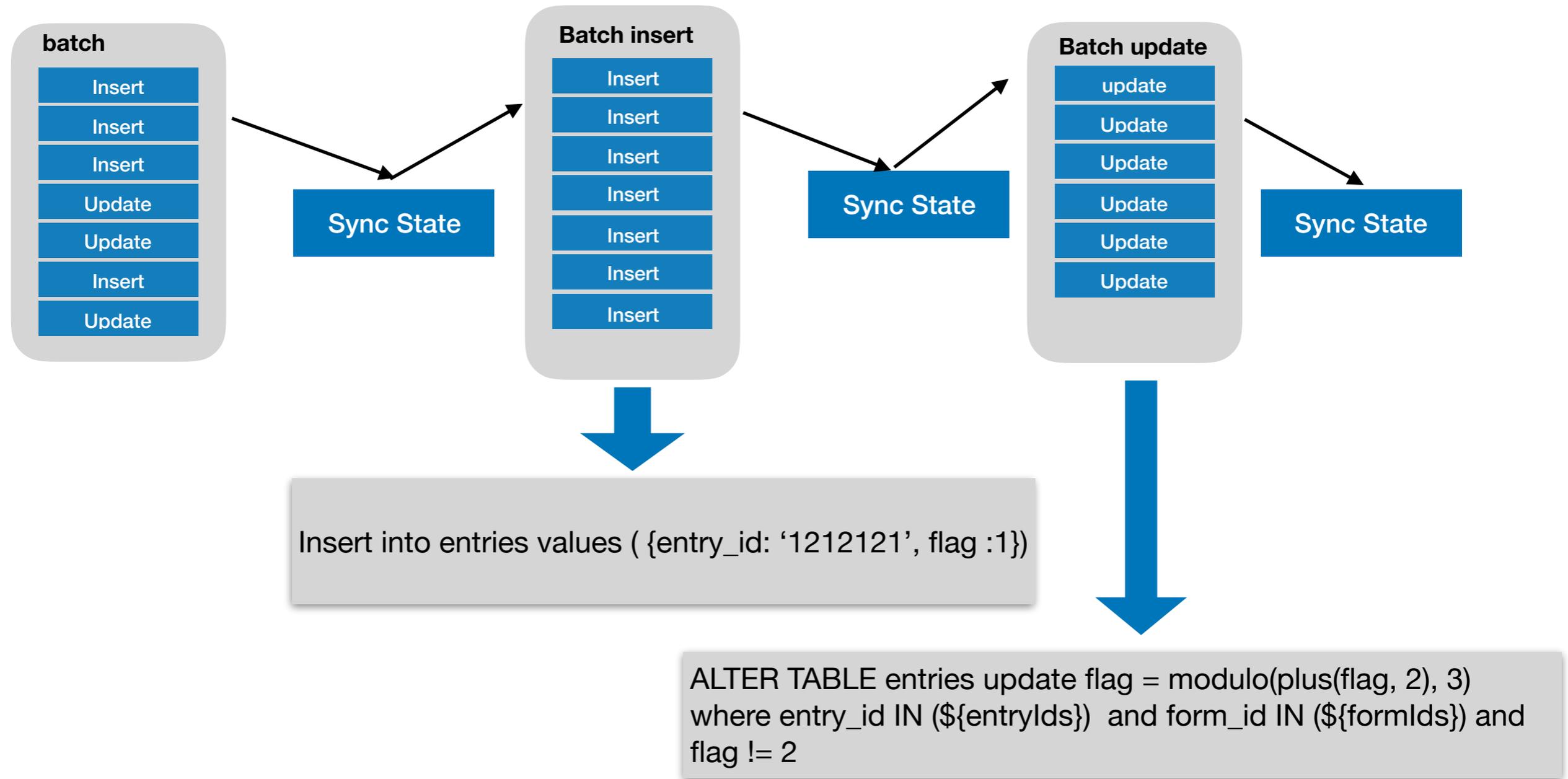
数据字段的映射

```
{
  "field_1": "2323",
  "field_2": {
    "asdw": {
      "field_1": "ssdf",
      "field_2": { "label1": "sdfd", "label2": "sdsd" }
    },
    "sedw": {
      "field_1": "swrw",
      "field_2": { "label1": "ssdd", "label2": "sser" }
    }
  },
  "field_4": [
    {
      "field_1": "sd23",
      "field_2": { "label1": "ser", "label2": "sdwesd" }
    },
    {
      "field_1": "sd23",
      "field_2": { "label1": "ser", "label2": "sers" }
    }
  ],
  "field_5": [
    { "field_code": "sdsfs", "amount": 1 },
    { "field_code": "sdsfs", "amount": 1 }
  ],
  "field_289": [
    { "field_code": "sdsfs", "amount": 1 },
    { "field_code": "sdsfs", "amount": 1 }
  ]
}
```



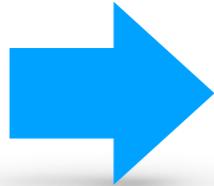
```
{
  "field_1": "2323",
  "field_5": "sdsfs:1, owuew:2,",
  "field_table": [
    "field_4;field_1;sd23",
    "field_4;field_2;label1:wer, label2:3wew",
    "field_4;field_1;sd23",
    "field_4;field_2;label1:wer, label2:3wew"
  ],
  "field_matrix": [
    "field_2;ssdf;field_1;sdrw",
    "field_2;ssdf;field_2;label1:wer, label2:3wew",
    "field_2;swes;field_1;sdrw",
    "field_2;sdwe;field_2;label1:wer, label2:3wew"
  ],
  "field_other": ["field5; sdsfs:1, owuew:2,"]
}
```

同步服务逻辑



数据分析服务

query



```
SELECT
  title,
  choice,
  count(*)
FROM
(
  SELECT
    substring(subject, 1, position(subject, ':') - 1) AS title,
    substring(subject, position(subject, ':') + 1) AS choice
  FROM
  (
    SELECT field_4
    FROM
    (
      SELECT substring(field_value, 9) AS ori_id
      FROM
      (
        SELECT field_extra
        FROM entries
        WHERE (form_id = '5cf5cd7577fa10d8caf1bf3a') AND (partition_id = 22) AND (flag != 2) AND (field_1 IN
        (
          SELECT toString(entry_id)
          FROM entries
          WHERE (form_id = '5cf5cb9a77fa10d8d1f1bf0e') AND (partition_id = 21) AND (flag != 2) AND match(field_2, '.*(S2K9
          )I(Zak7)I(SATp).*)' ) AND match(field_4, '.*(CZI3:FMC2)I(CZI3:6skD)I(CZI3:Yflu).*)' ) AND arrayExists(x -> match(x, 'field_10;.*(sUvk:eY
          mw)I(sUvk:cdmj)I(sUvk:V6n5).*'), field_extra) AND arrayExists(x -> match(x, 'field_8;.*(j6Bk)I(v18Y).*'), field_extra)
        ))
      )
      ARRAY JOIN field_extra AS field_value
      WHERE startsWith(field_value, 'field_5;')
    )
    INNER JOIN
    (
      SELECT
        toString(entry_id) AS ass_id,
        field_4
      FROM entries
      WHERE (form_id = '5cf5cb9a77fa10d8d1f1bf0e') AND (partition_id = 21) AND (flag != 2) AND notEmpty(field_4)
    ) ON ori_id = ass_id
  )
  ARRAY JOIN splitByChar(',', field_4) AS subject
)
GROUP BY
  title,
  choice
title choice count
| UCr0 | Yflu | 2 |
```