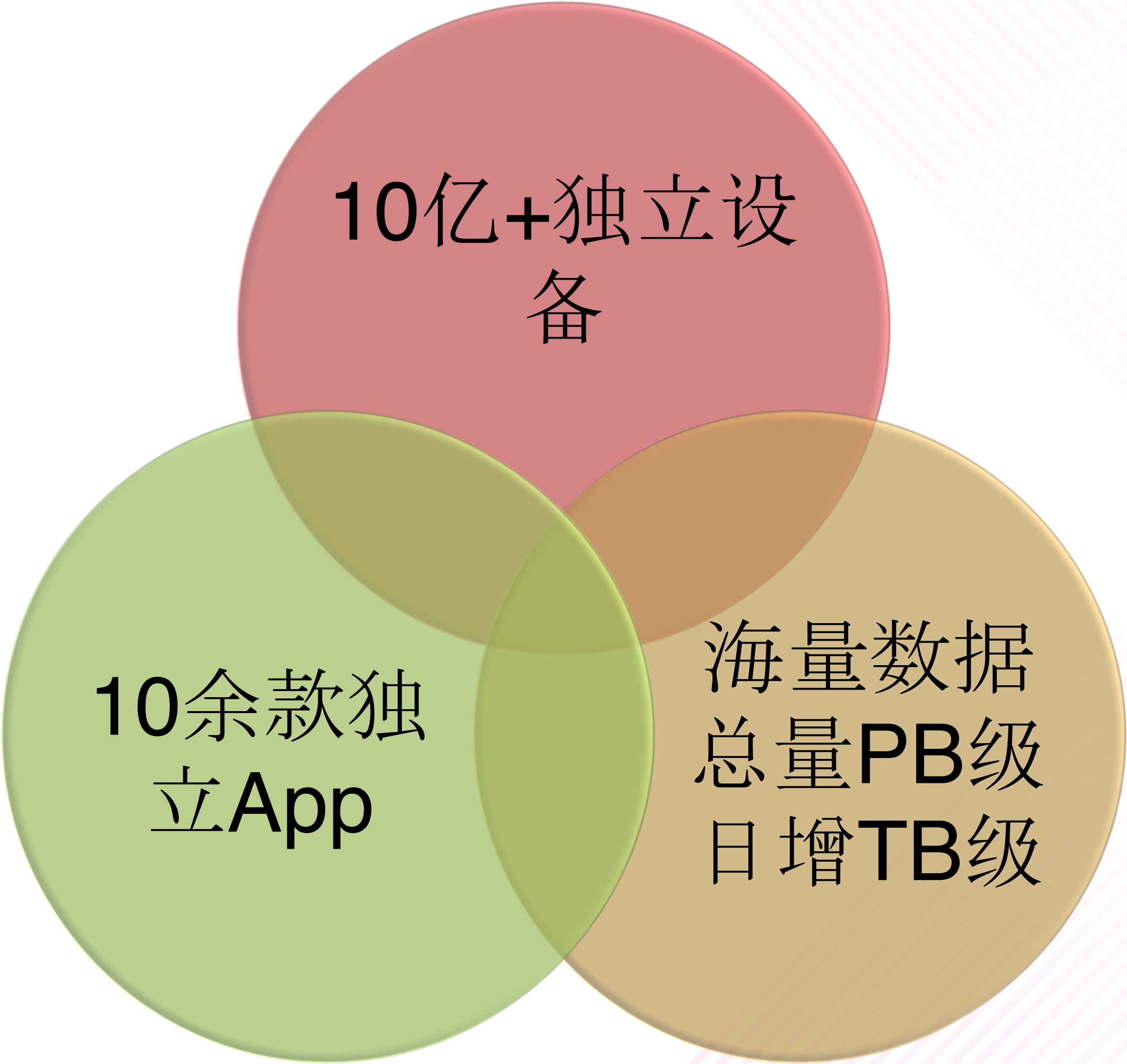


美图大数据统计解决方案

陈满意 技术经理 2017.11.04

- 美图数据统计发展
- 美图大数据开放平台
- 未来规划

美图公司与大数据统计



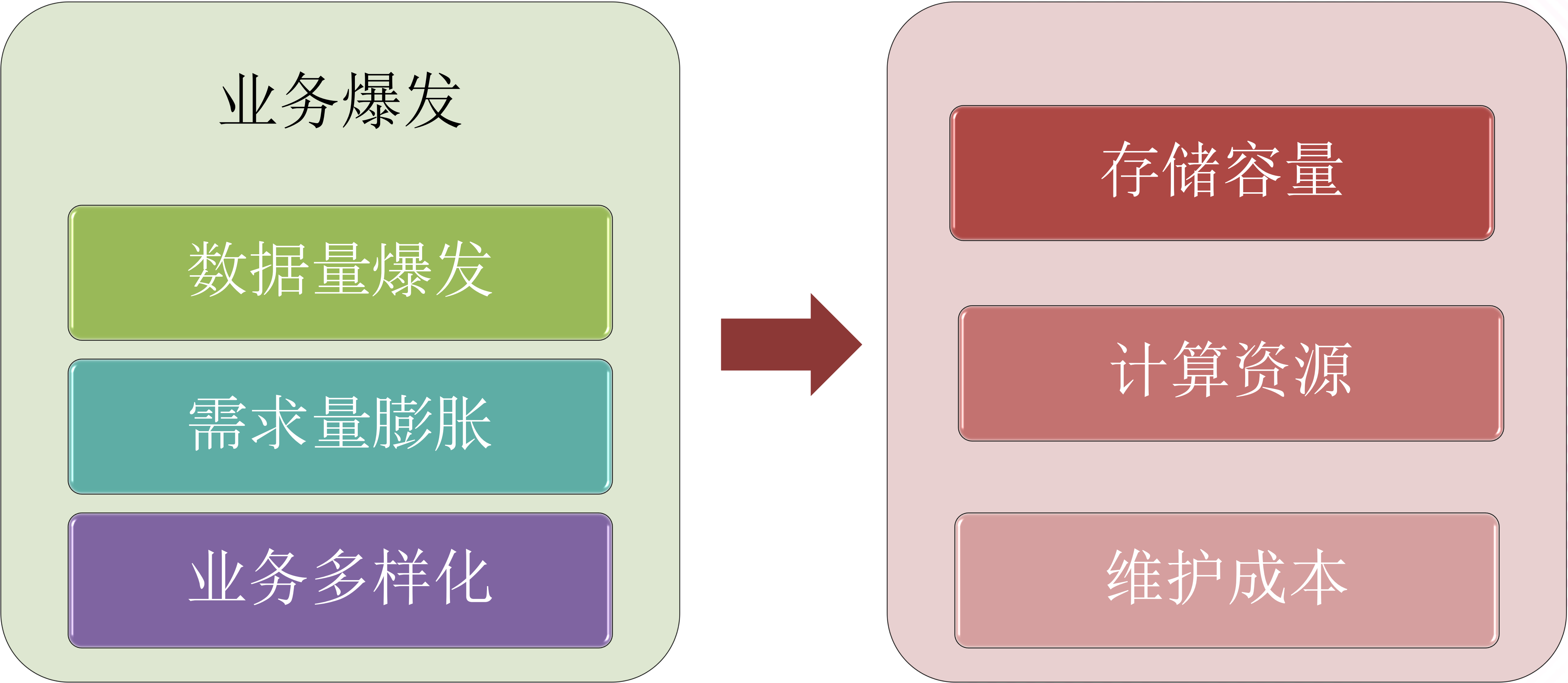
美图数据统计发展历程

从无到有

- 简单实现、快速上线



业务与数据的爆发



数据开放平台

数据集群

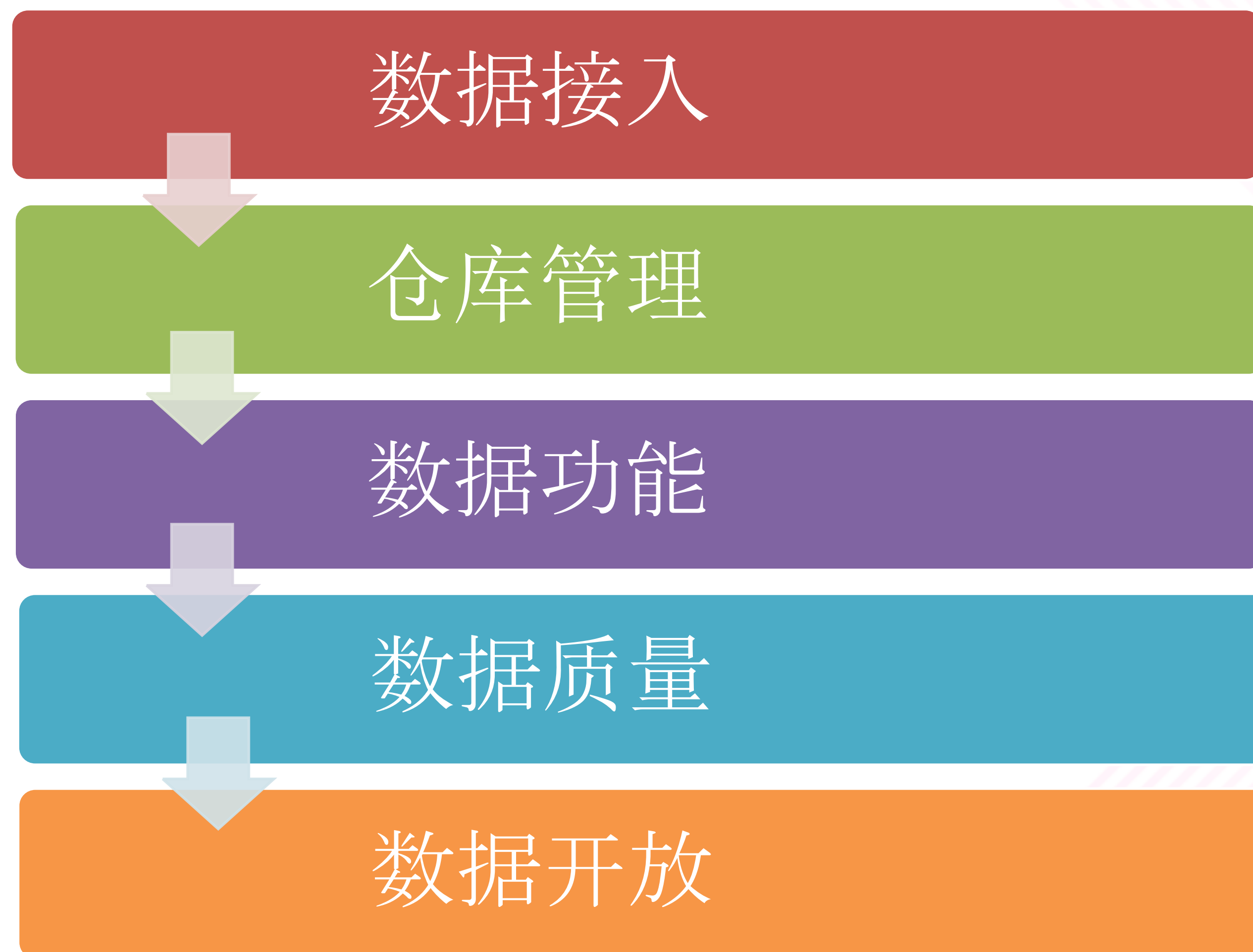
存储

计算

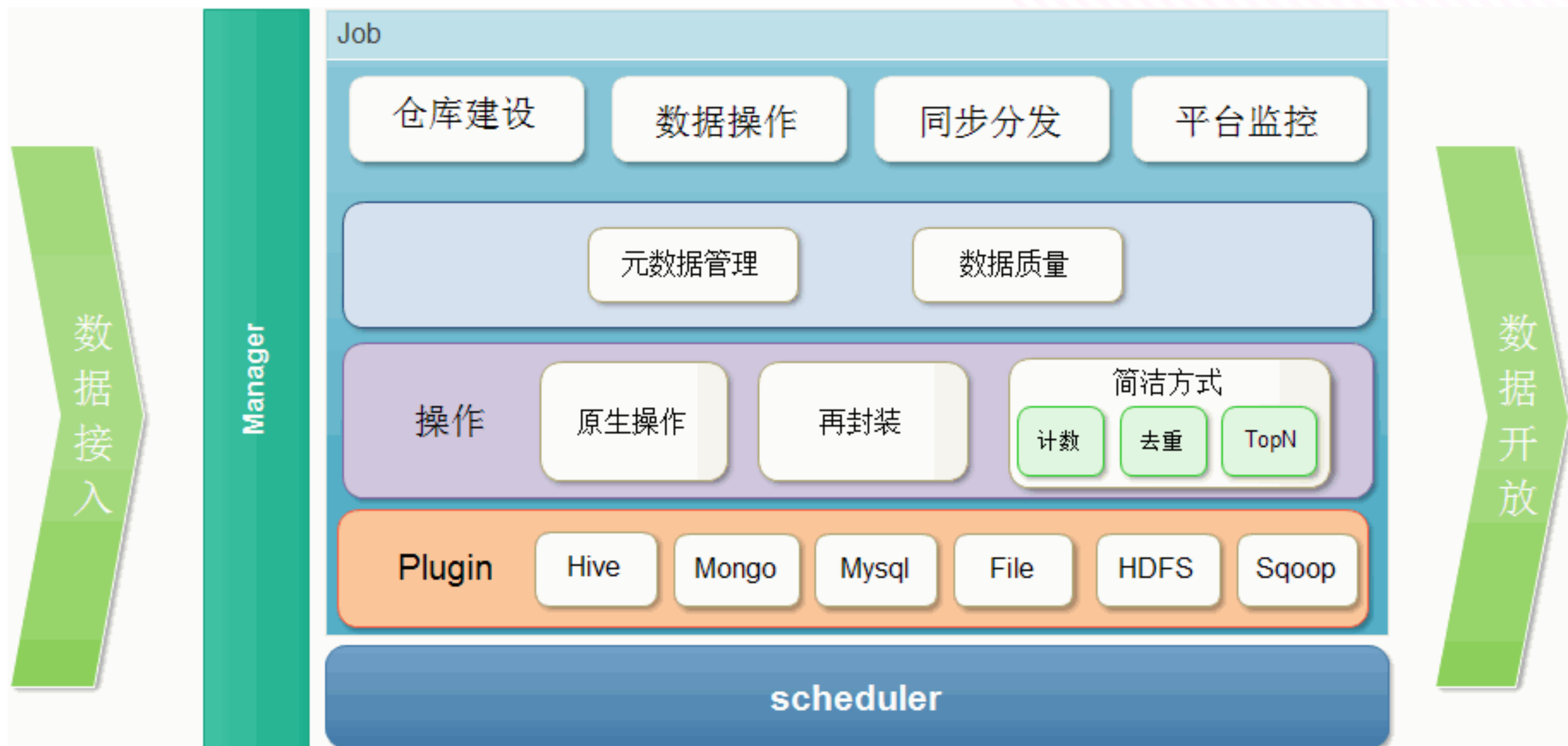
大数据开放平台

数据开放平台

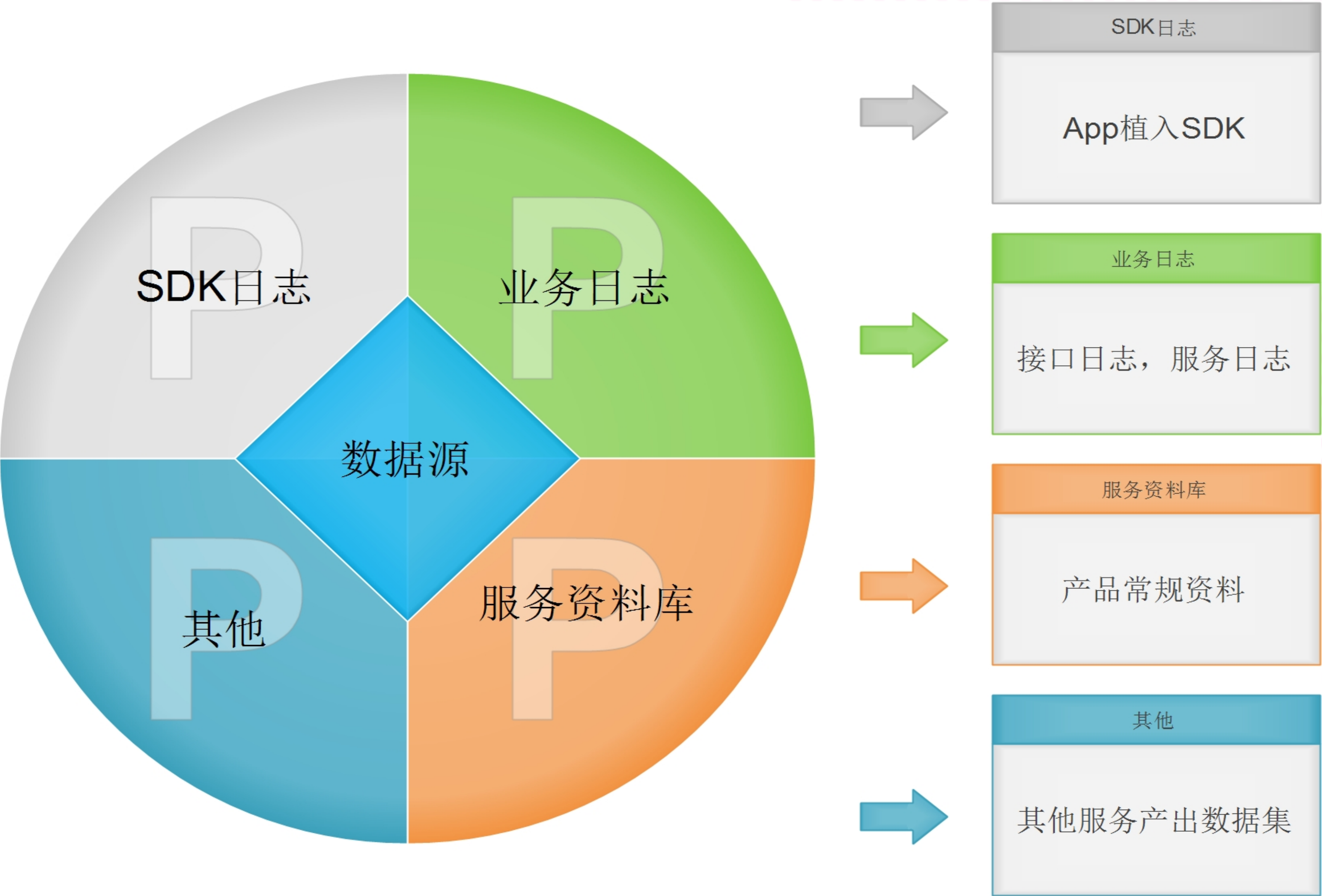
- 美图大数据体系中提供数据服务的开放平台



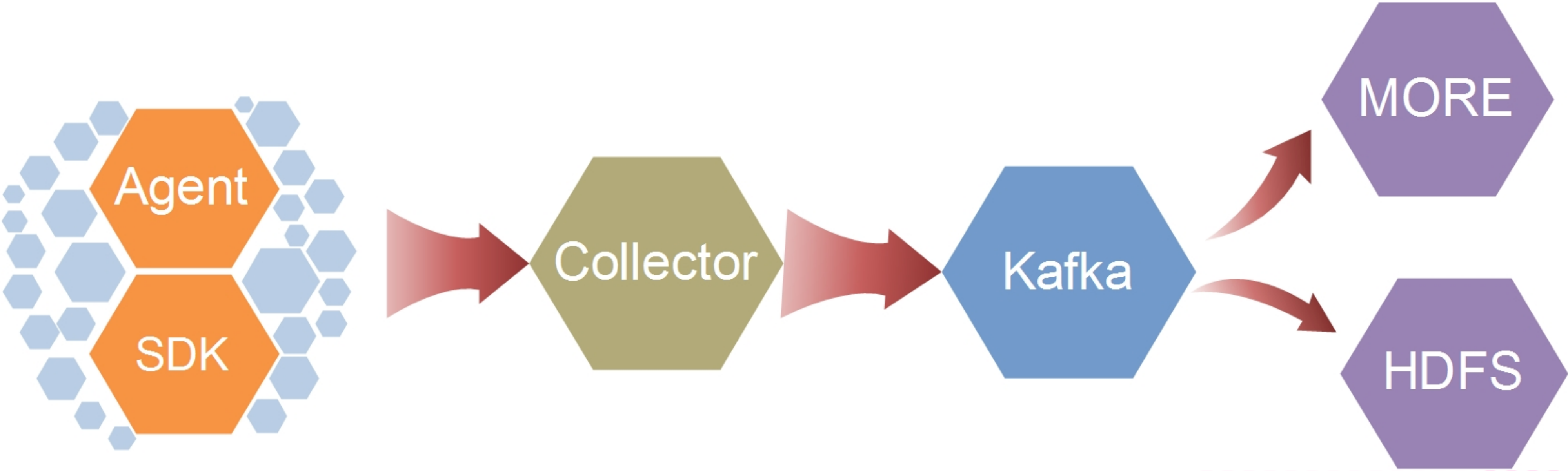
平台架构



数据接入-来源



数据接入-流程



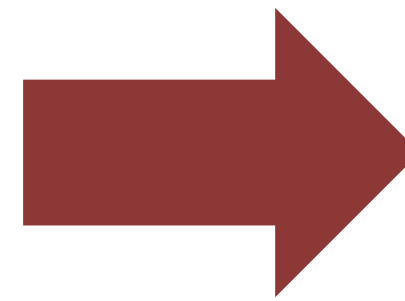
失控的HDFS与Hive

数据无归类

命名随意

缺乏数据描述

无数据权限

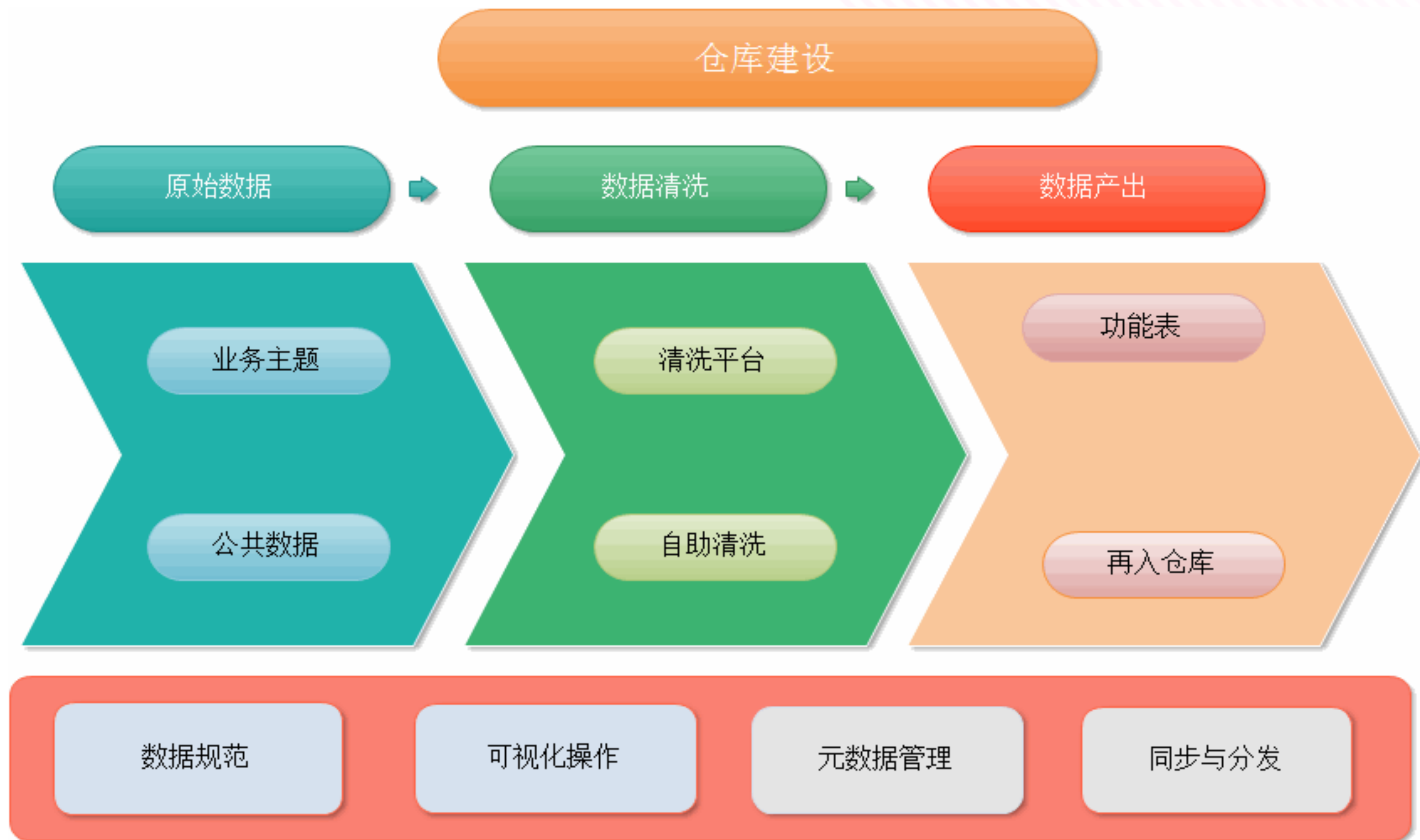


仓库紊乱

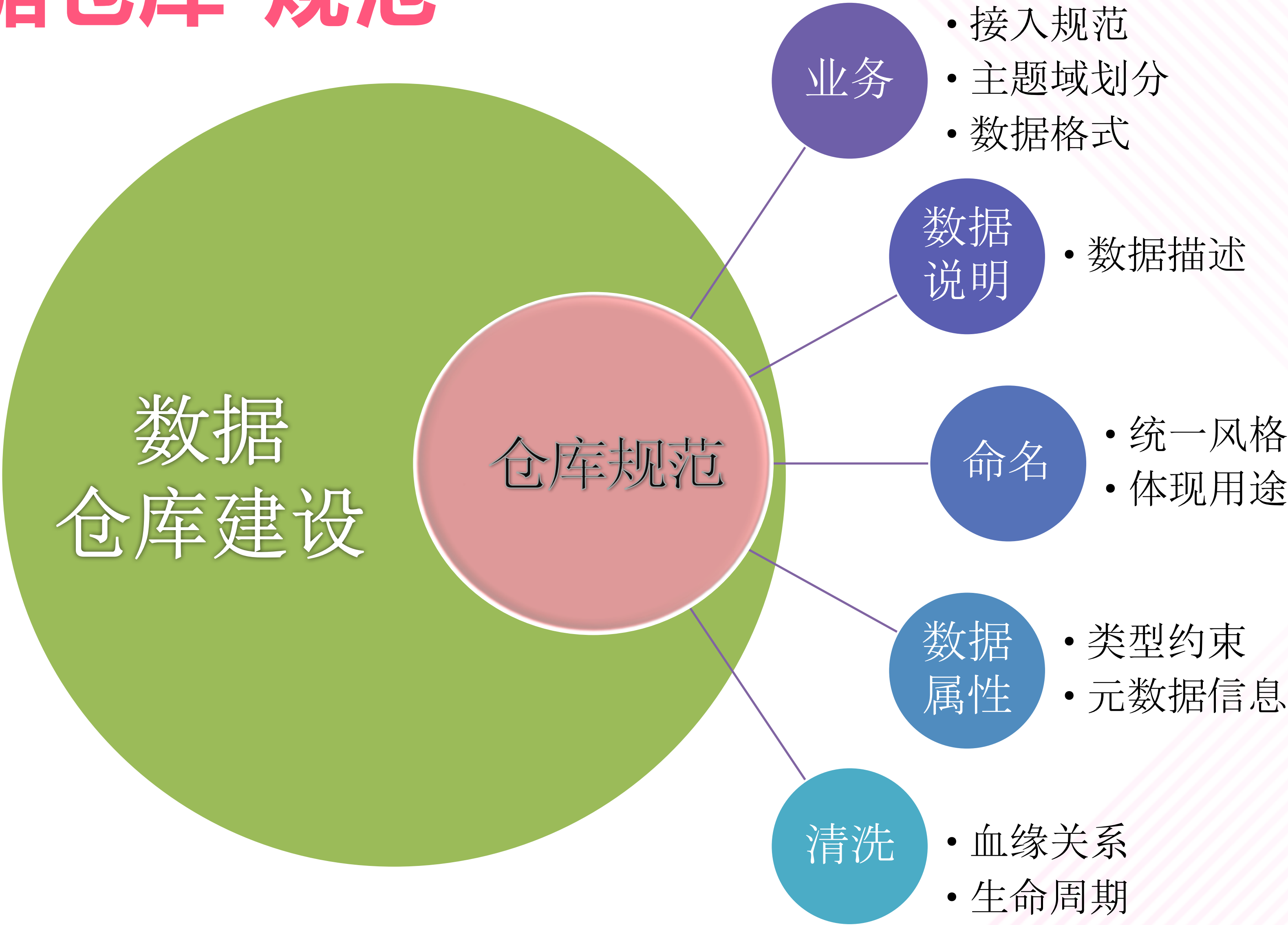
数据不透明

缺乏安全性

数据仓库-建设



数据仓库-规范



数据统计

- 丰富的统计插件
- 快速自定义插件
- 支持UDF

业务自治

- 非研发人员快速接入
- 可视化的任务配置
- 计数\去重\TopN等
- 自动化生成接口

数据分析

- 自助查询
- 语法校验

- 不受控的语法操作

```
graph LR; A(( )) --- B[低级错误语法]; A --- C[危险操作语法]; A --- D[高资源损耗的语法];
```

低级错误语法

危险操作语法

高资源损耗的语法

数据查询-语法验证

- 基于HQL的语法验证



自定义语法
校验

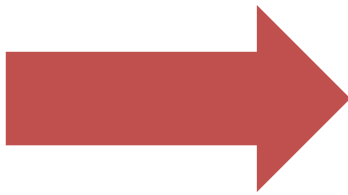
快速校验

危险语法限制

高资源损耗验证

不合理语法验证

数据查询-多维度查询



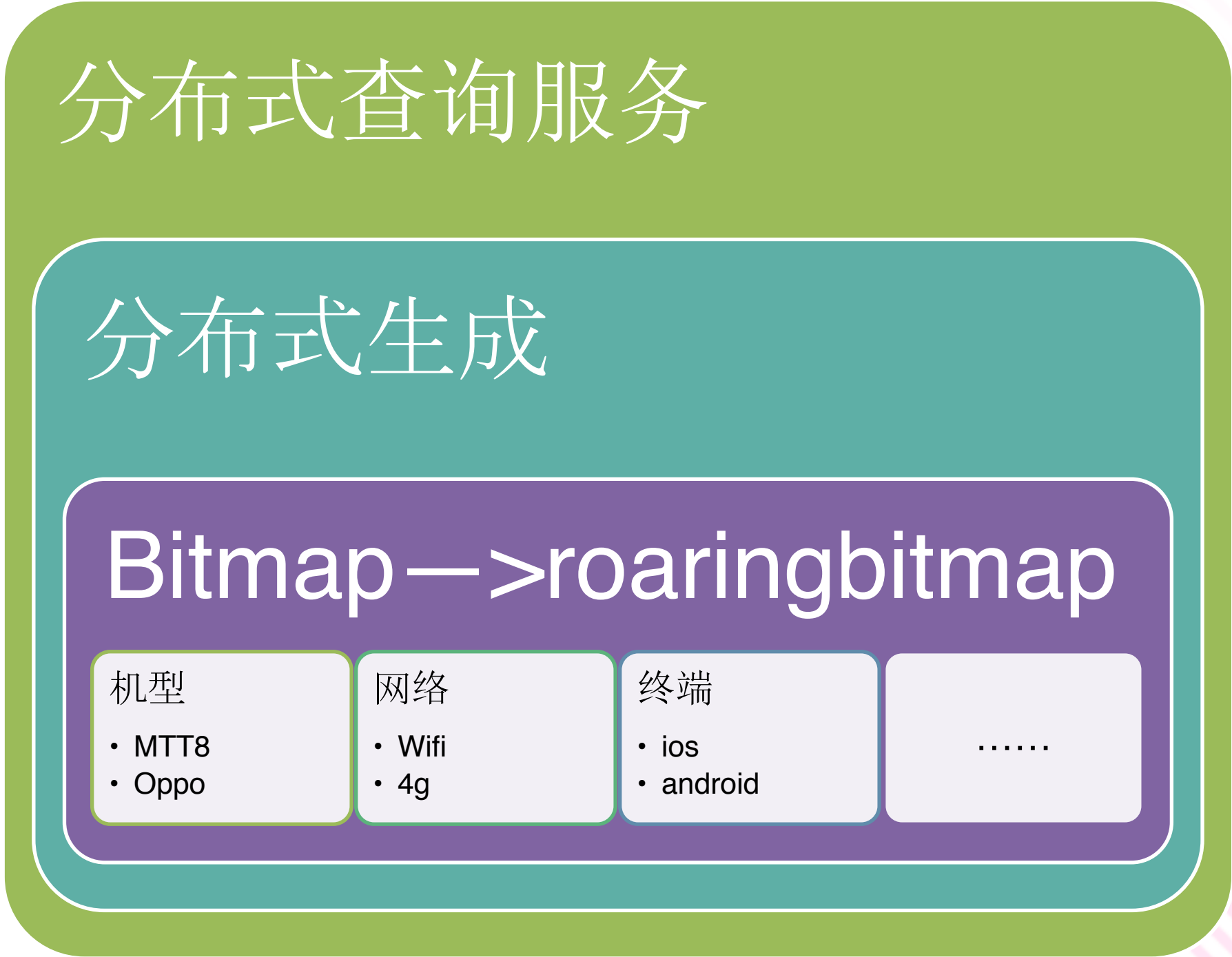
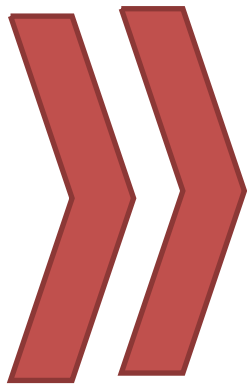
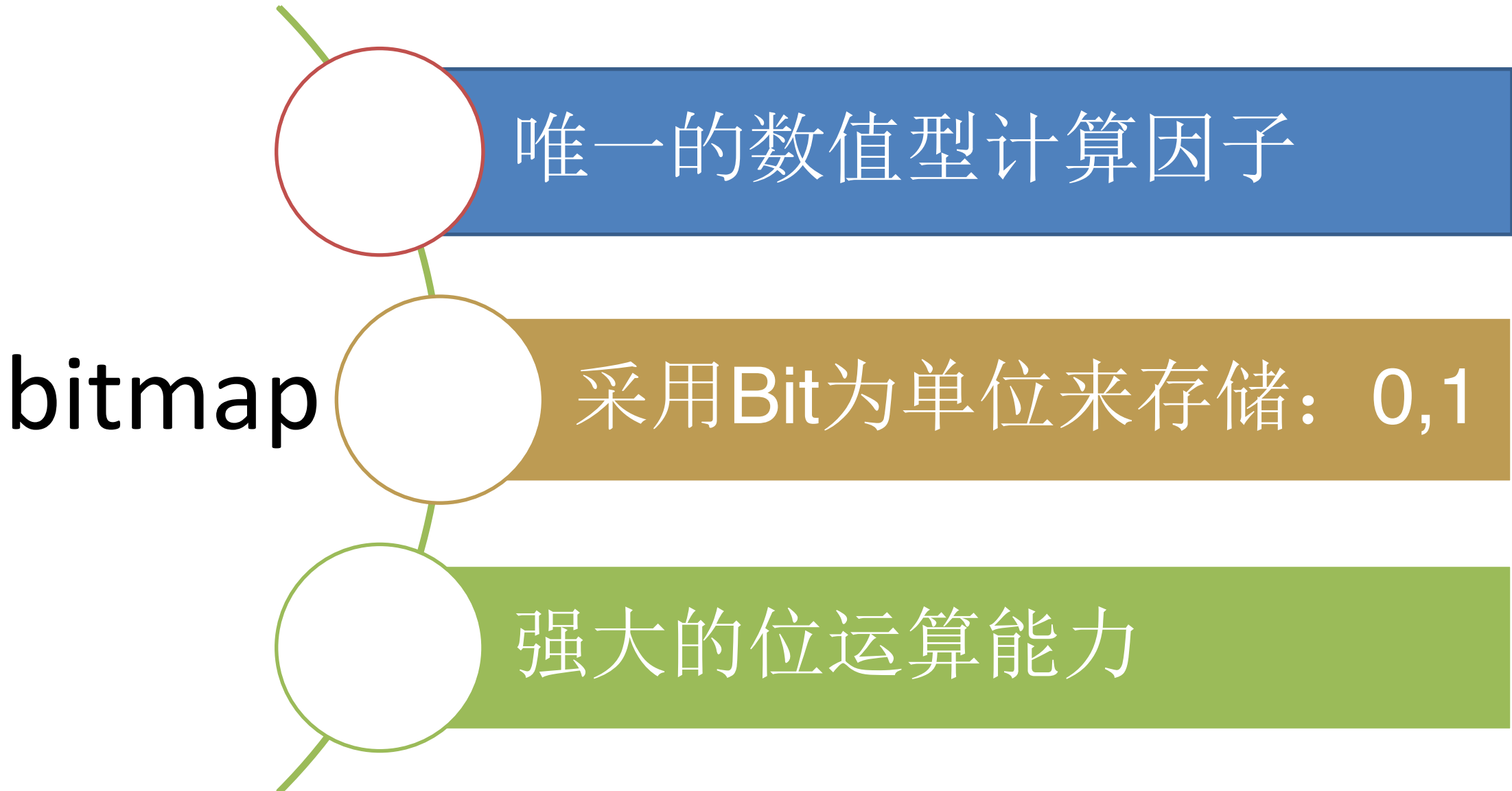
多维度运算

多维交叉

时间跨度长

即时响应

数据查询-多维度查询



作业调度器

- 各种查询作业是运行在作业调度器上



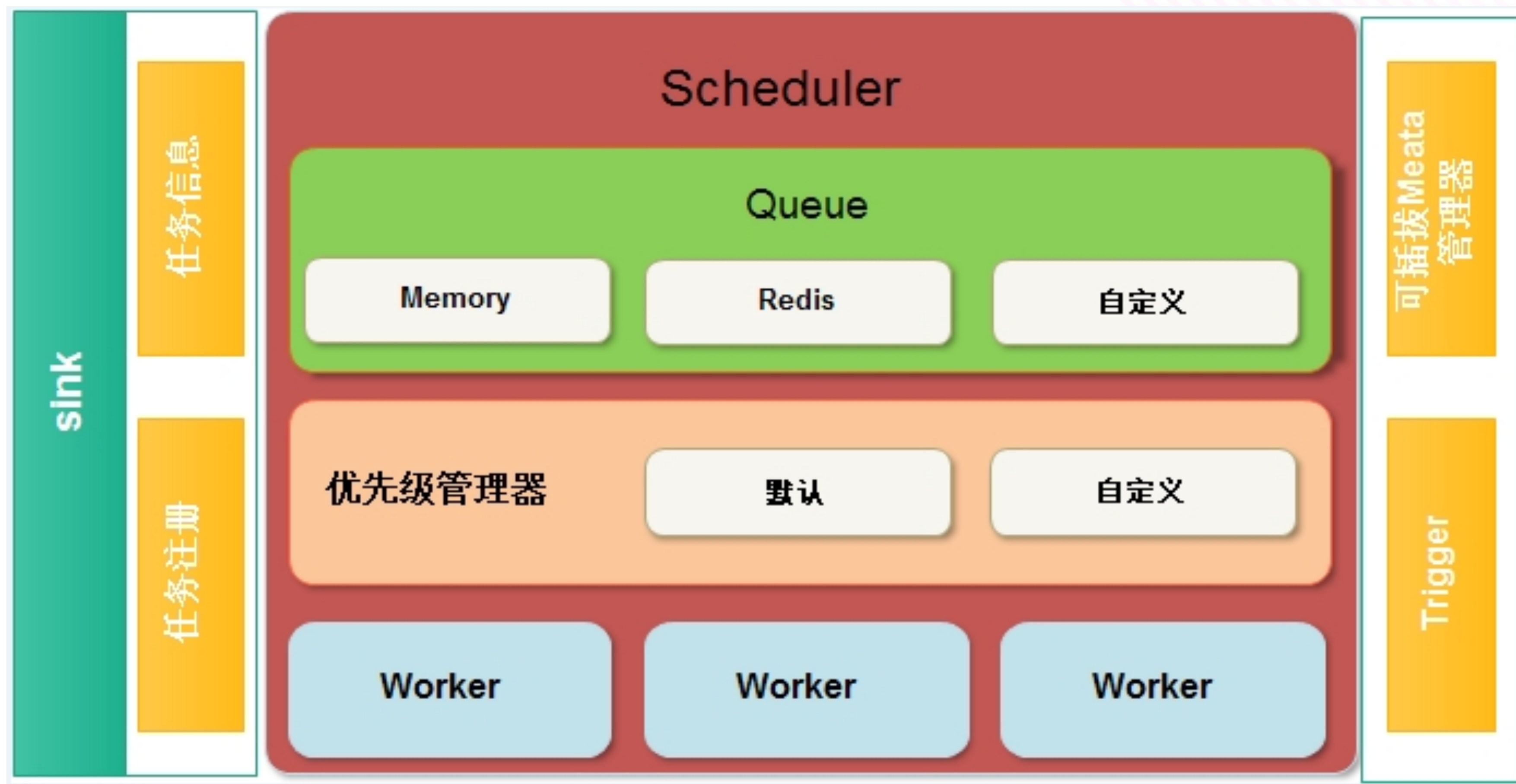
支持即时/定时的作业调度

支持工作流模式

支持作业优先级等参数设置

支持调度器横向扩展

作业调度器



统计数据存储方案

- mongodb

模式自由

- 简化任务创建

快速

- 近期指标为热数据

可靠性

- 副本集

可扩展

- 分片机制

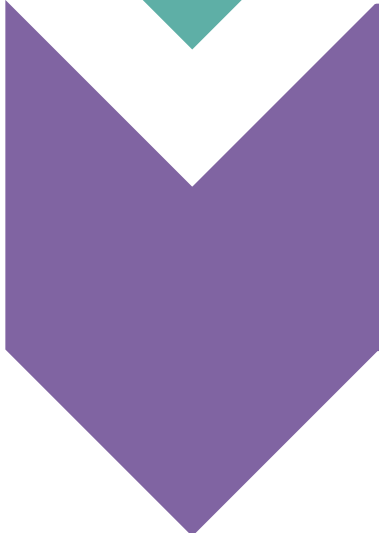
数据质量



- 统一的唯一设备ID

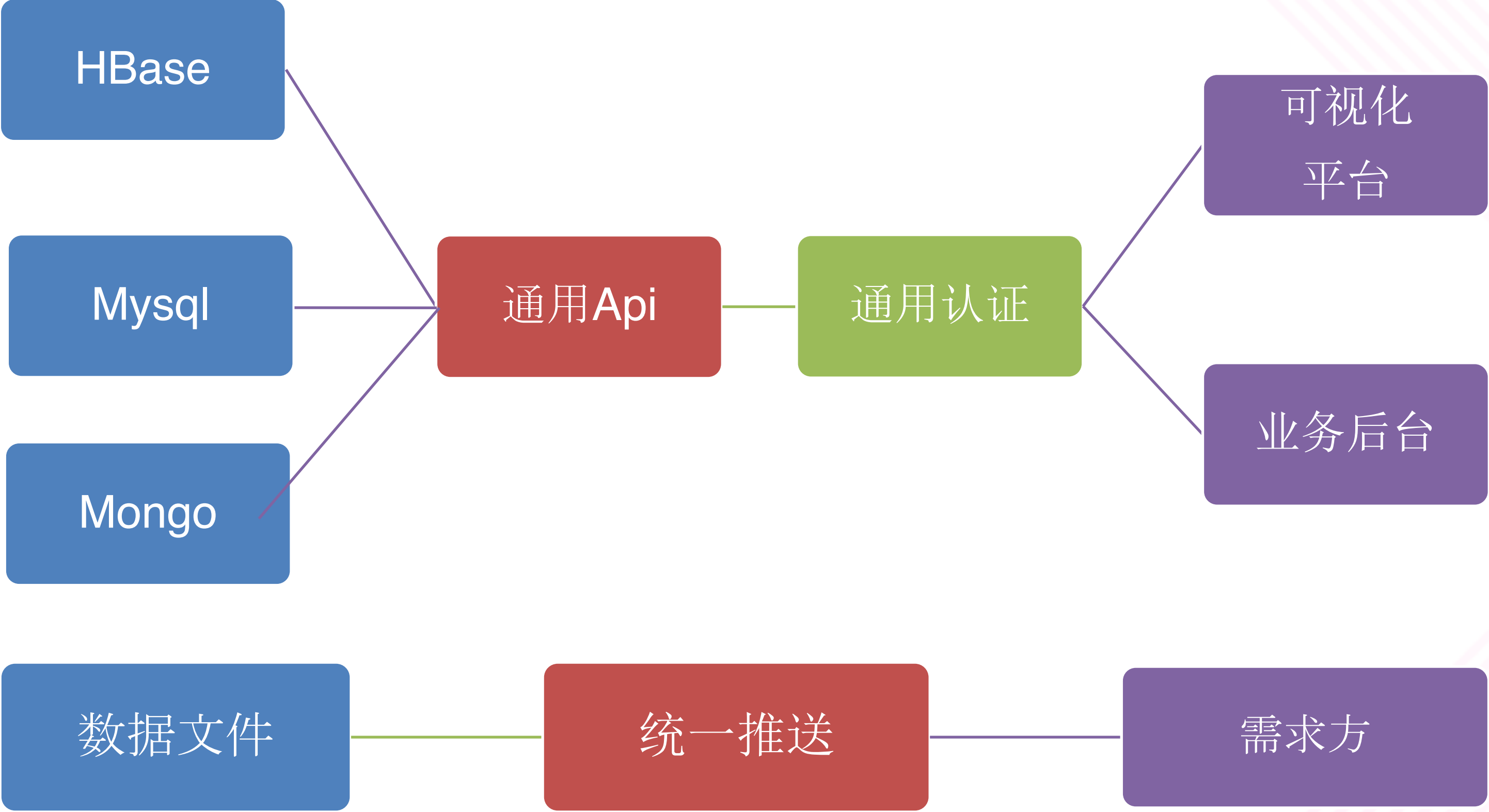


- 统一的地理位置服务



- 数据监控服务

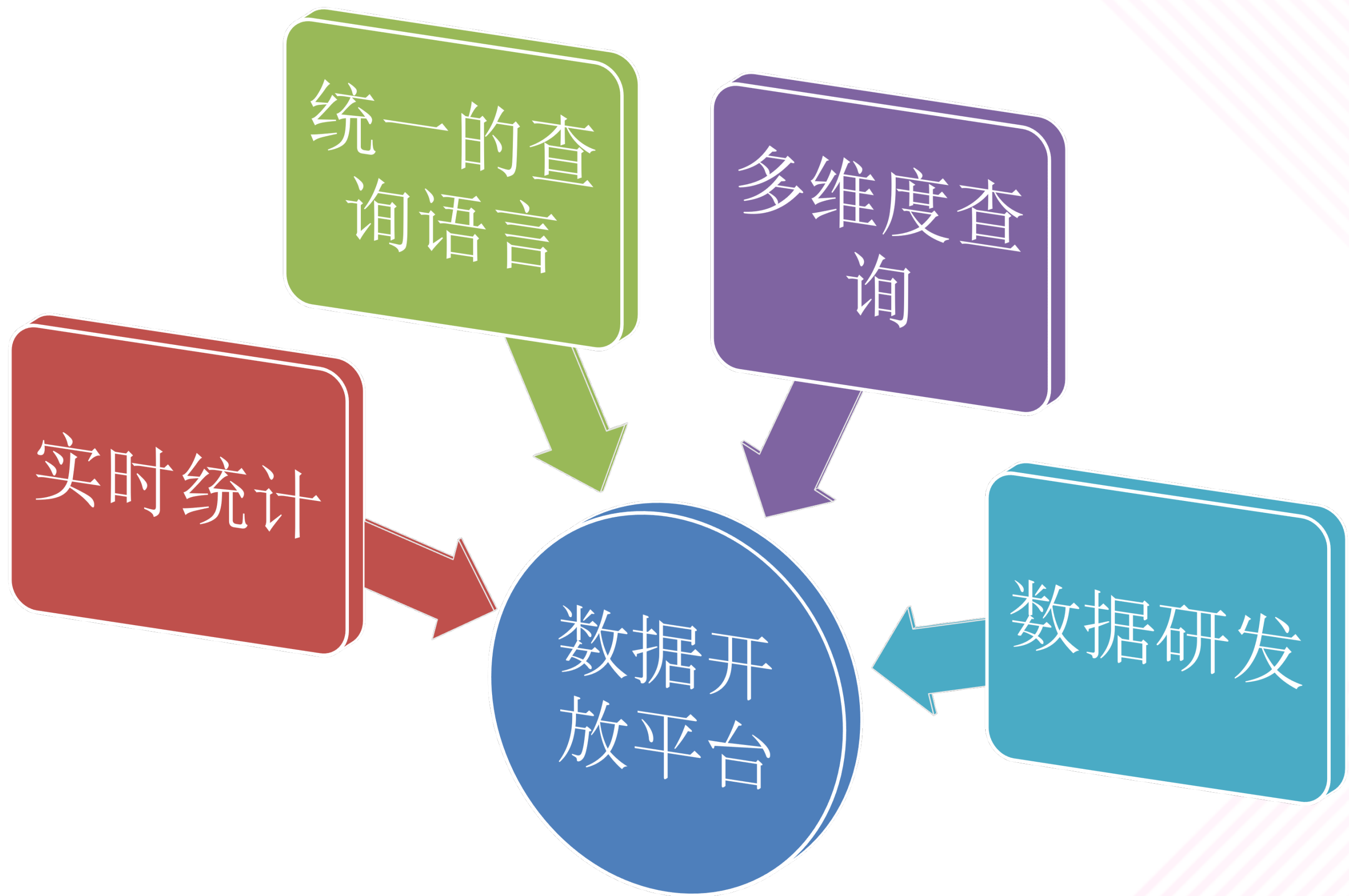
数据开放



未来规划

构建强大的大数据开放平台

美图大数据开放平台



Thank You !