

The SACC logo is rendered in a bold, white, sans-serif font with a blue glow effect. It is positioned in the upper right quadrant of the image, above the main conference title. The background features a blue wireframe architectural design with a perspective view of a city skyline and a large gear-like structure at the bottom left.

2021 中国系统架构师大会

SYSTEM ARCHITECT CONFERENCE CHINA 2021

数字转型 架构重塑

IT168.com

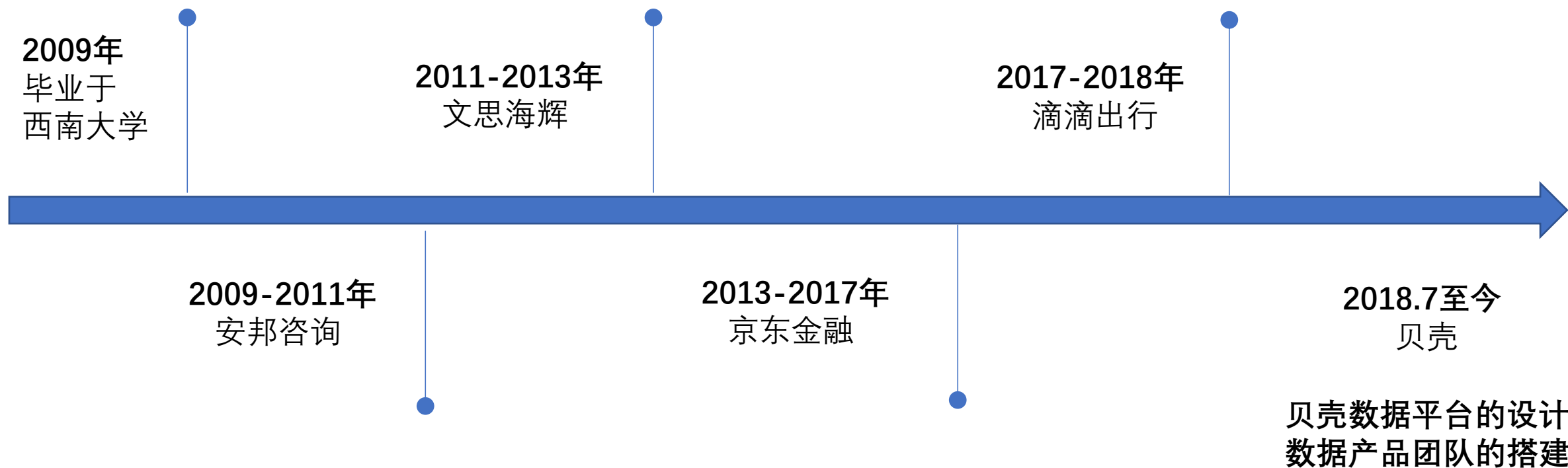
ChinaUnix

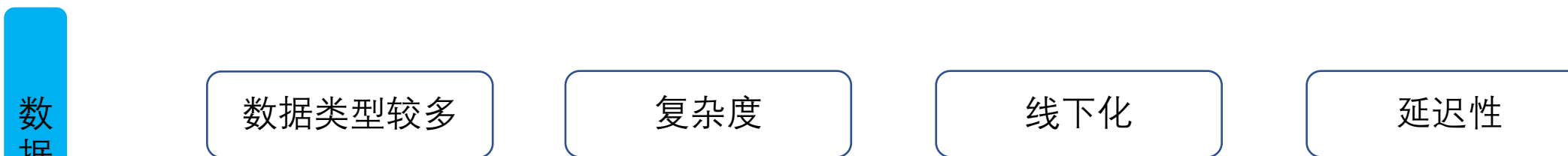
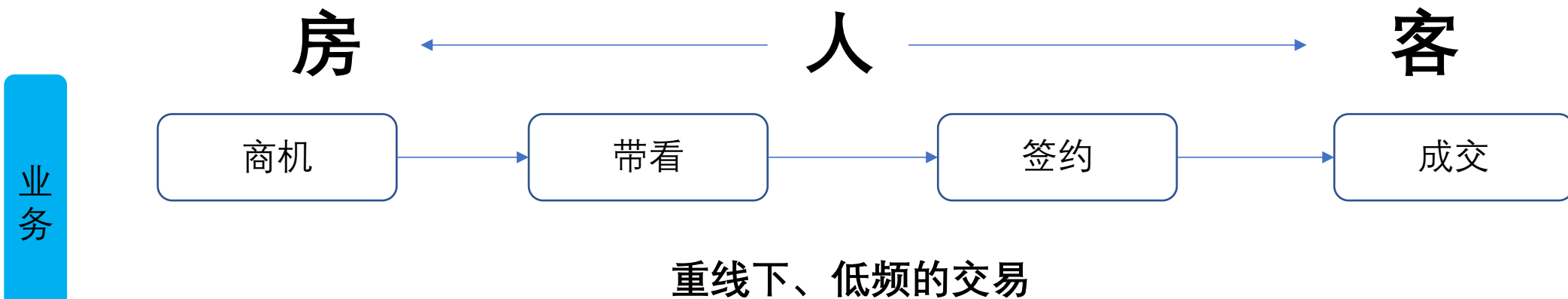
ITPUB

云上会议 网络直播 | 2021.5.20-2021.5.22

贝壳找房数据平台演进之路

自我介绍





贝壳都有哪些数据用户？



数据分析师

500+人

- 数据提供
- 探索分析
- 报告输出



产品/运营

4500+

- 策略孵化/制定
- 过程监控
- 向上汇报



管理人员

3000+

- 管理抓手
- 决策依据
- 目标制定
- 任务拆解

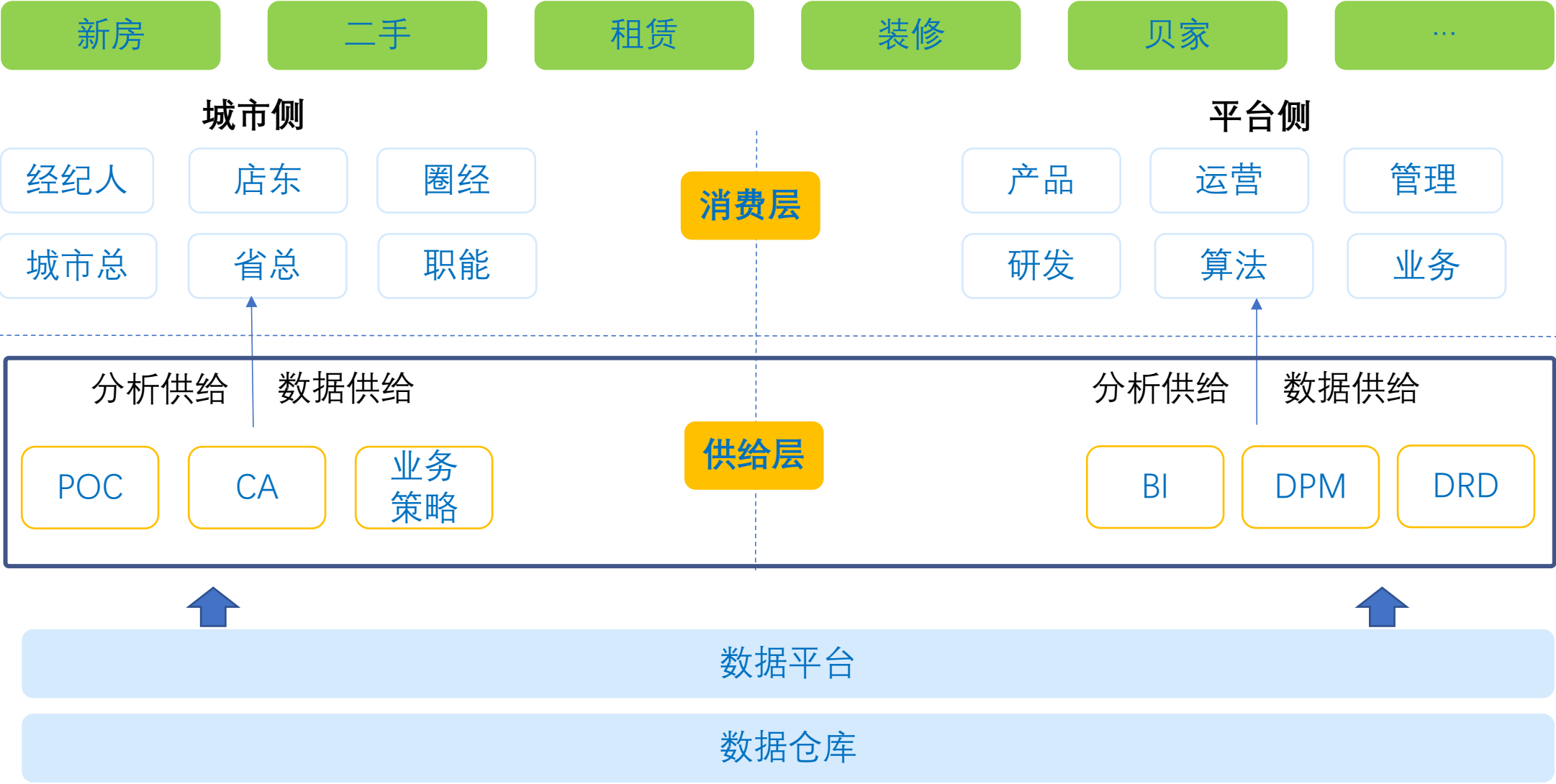


业务人员

20W+

- 业务管理
- 过程监控
- 向上汇报
- 述职

贝壳“数据业务”的组织形式



不同角色的数据应用场景与问题

用户

数据应用场景

管理视角

平台、城市
管理层

- ◆ 区首、省总、城市总述职
- ◆ OKR管理
- ◆ 绩效管理
- ◆

数据使用

管理抓手作业赋能

效率

- ◆ 以城市为例，由于不同城市的组织、人员、业务不同，城市自建数据库，城市人员加工效率和运维成本更高

作业视角

Ca/AE
店东/经纪人

- ◆ Ca作业数据赋能
- ◆ 城市职能数据赋能
- ◆ 市占率，健康度，门店分级
- ◆

易用/
安全

- ◆ 根据调研，数据传播90%以上基本是通过图片以群的形式传播，数据安全无法保障，数据易用性不足

产研视角

BI/PM/
RD/职能

- ◆ 通过数据分析优化迭代产品
- ◆ 周期性报告（日、周、月）
- ◆ 基于数据进行活动运营效果评估
- ◆

数据分析

系统对接

重复/
质量

- ◆ 业务线只能通过大数据获取数据，技术能力获取较少，不断孵化出多种数据产品，重复建设，导致数据出口多

公司视角

公司

- ◆ 作业过程中是否由数据进行指导
- ◆ 数据资产的价值，（楼盘字典）
- ◆ 数据使用成本
- ◆

数据评估

成本

- ◆ 只有上线，很少下线，重复、无用的数据量越来越多，计算、存储成本陡增，数据价值无法精确衡量

效率、易用、质量、成本为各类视角最突出的四大问题

要解决的核心问题



提升城市数据使用效率:

城市加工效率

指标一致性、复用率

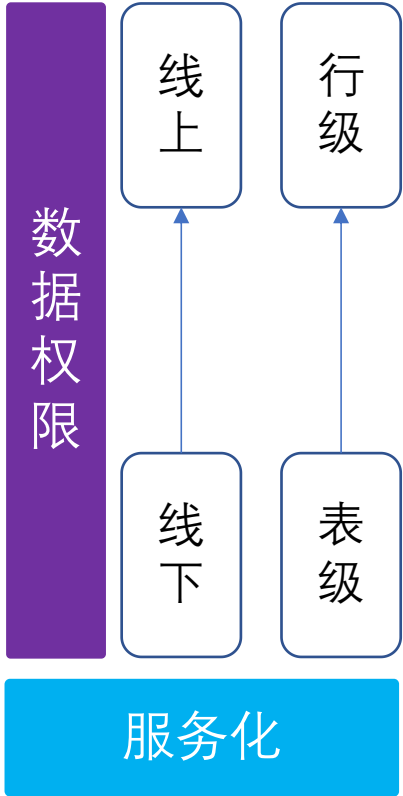
数据质量成本控制:

存储、计算资源使用率

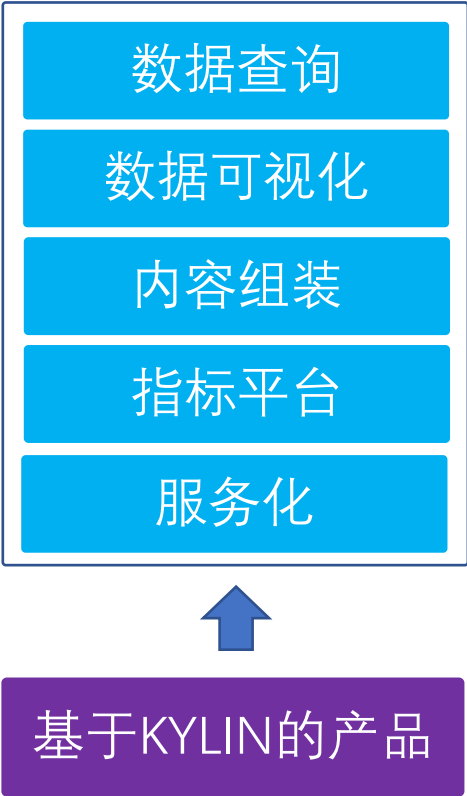
数据质量提升

解决思路

Setp1: 提高用户获取数据的效率



Setp2: 解耦+整合
数据产品体系



Setp3: 提升底层服务能力，数据资产化



Setp4: 运营体系搭建

产品解决方案

管理层\业务
职能\产研

同端
同源

经管场景应用

驾驶舱

阿记okr..

作业场景应用

Ca赋能

店东数据赋能..

分析场景应用

个性化分析

数据探索..

分析师

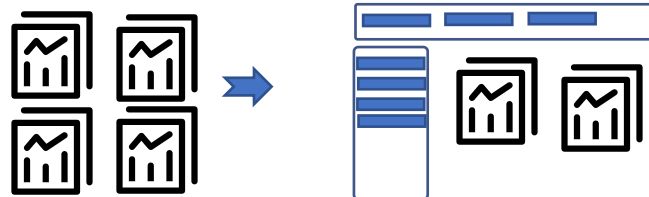
内容
组装

数据门户

报表

指标

基于指标、报表为数据源的数据门户配置



数据应用 (app)

指标

基于指标为数据源的数据应用配置平台

交互
框架

数据
配置

发布
应用

奥丁
生态

配置
化

内容
创作

数据接入

表

Hive\mysql
Excel\csv...

数据建模

模型

Sql建模
其他语言

数据可视化

报表



数据指标

指标

认证

数据查询

提数

探索

模块
化

研发

数据采集

业务库

流量

数据开发

离线

实时

数据地图

数据资产

元数据

数据服务

数据api

数据质量







规则引擎

服务层：监控、质量、通知、保障服务、工单...

原子，衍生+复合，构建指标体系

以往

指标数量8000+

- 重复建设： →  CASE:近7天北京链家带看量
-  →  CASE:近3天北京链家带看量
-  →  CASE:近5天北京链家带看量
- 口径传递：指标口径更新无法得到全链路传递
- 开发周期长：只要是新指标均有开发时长

基于KYLIN的指标构建

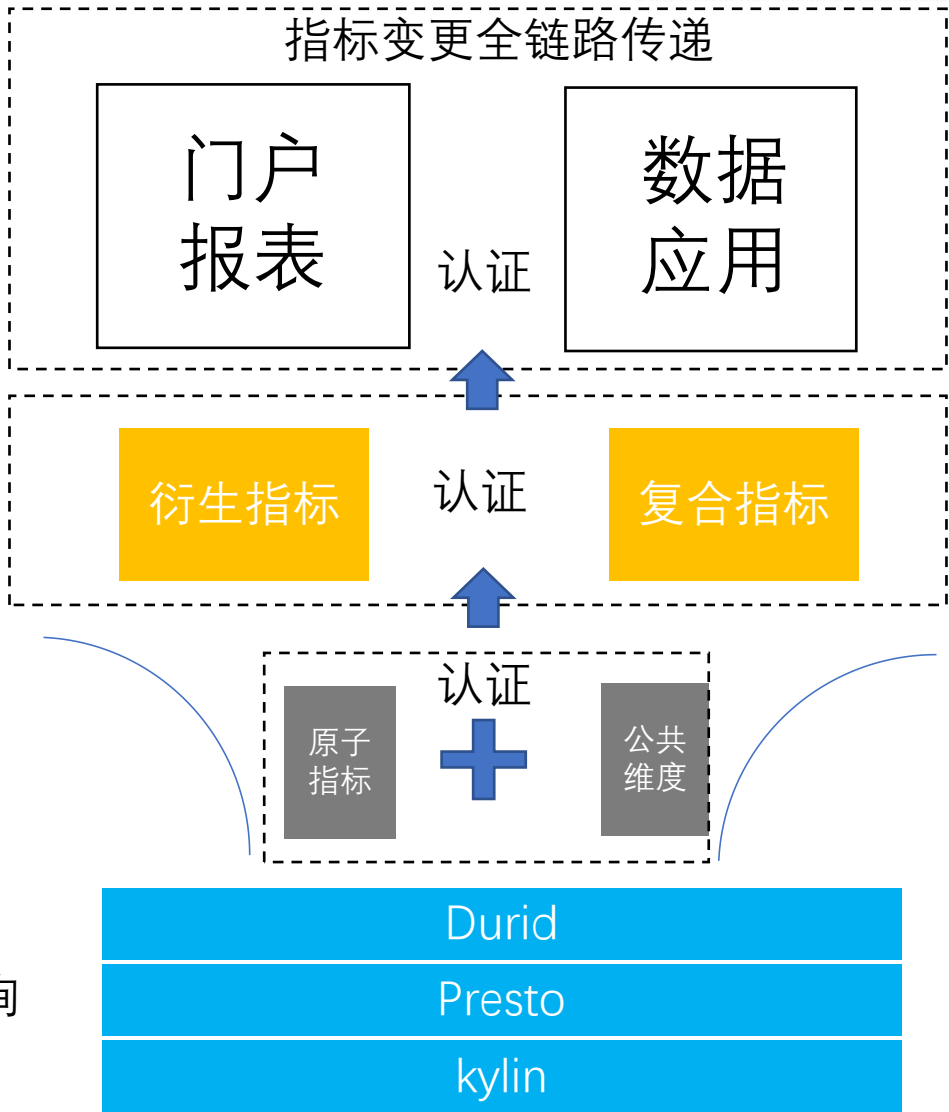
现在



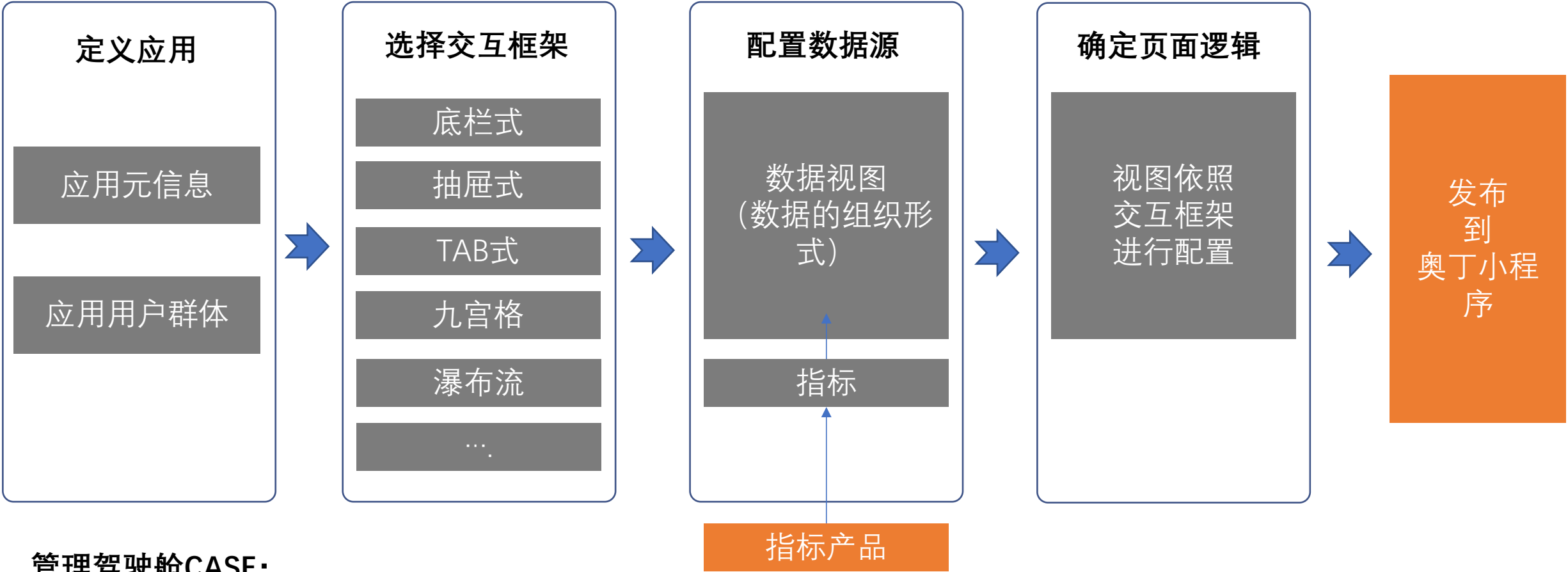
万级组合

百级开发

分层查询



指标+交互框架5分钟创建数据应用



管理驾驶舱CASE:



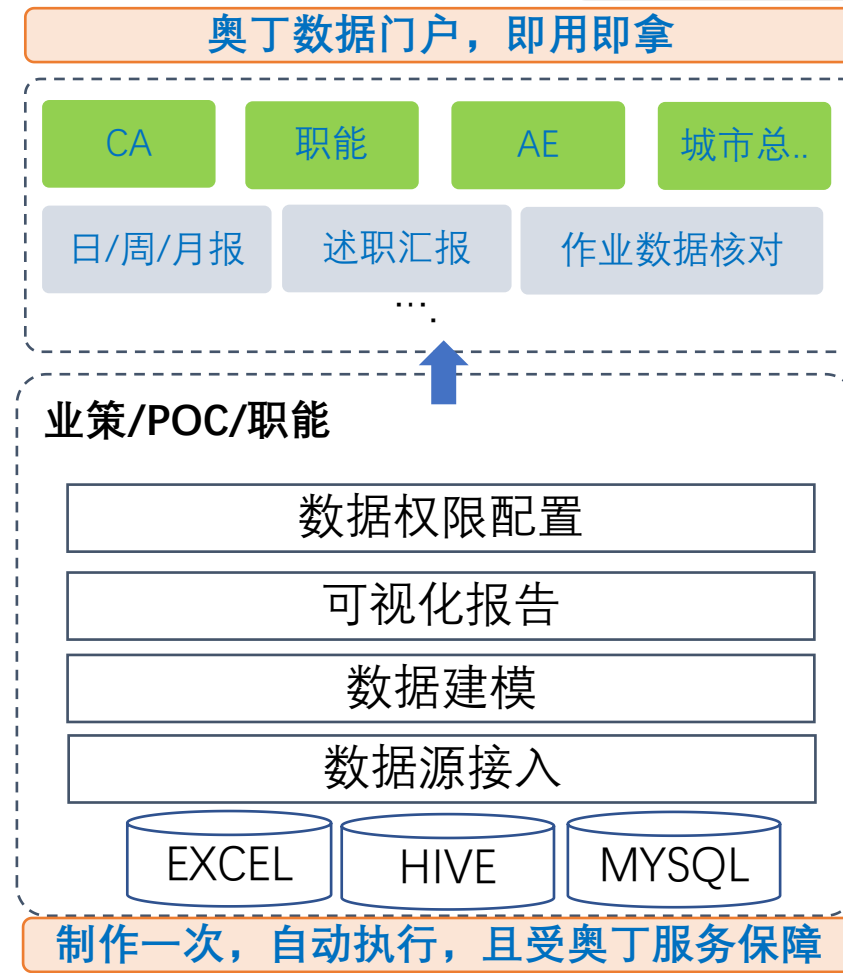
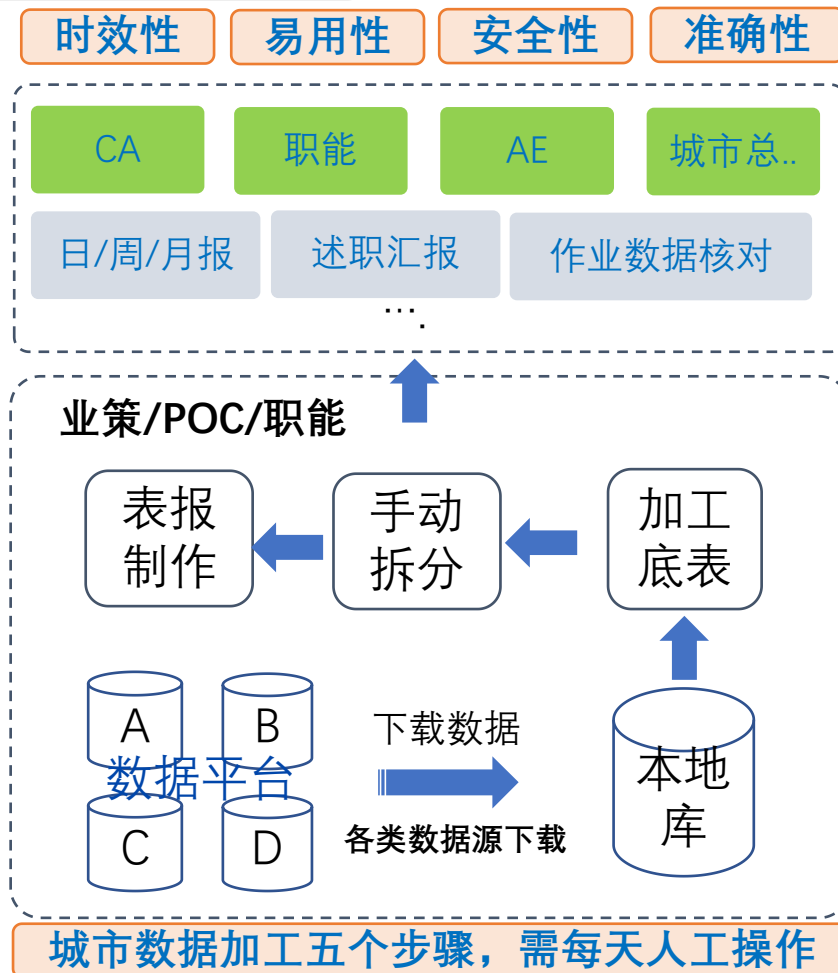
数据建设效率提升

SACC 2021

- 下午看数据
- 想看历史要翻群
- PDF\PPT漫天

过去 ———— 以城市举例 ———— 现在

- 上午8点
- 即用即拿
- 安全保障



由散到整，去“数据孤岛”，抓住效率痛点与供给侧产生共鸣，提高供给侧数据建设效率，释放人力，下游体验提升

平台产品对接，大数据能力复用

SACC 2021

- 定制化开发
- 数据开发周期长
- 数据权限独立开发

过去

现在

- 数据可视化能力复用
- 页面可嵌入
- 数据权限灵活可配置

例如19年7月的CHO效能看板，用时2周上线，提效显著

定制化开发需

发布

测试

权限开发

指标开发

页面开发

指标设计

页面设计

定制化开发流

需求/场景抽象

新房

二手

装修

...

数据场景化应用（例如图灵）

发布

嵌入

可视化报表、权限

指标开发

指标设计

配置化

ODIN能力复用

需求/场景抽象

新房

二手

装修

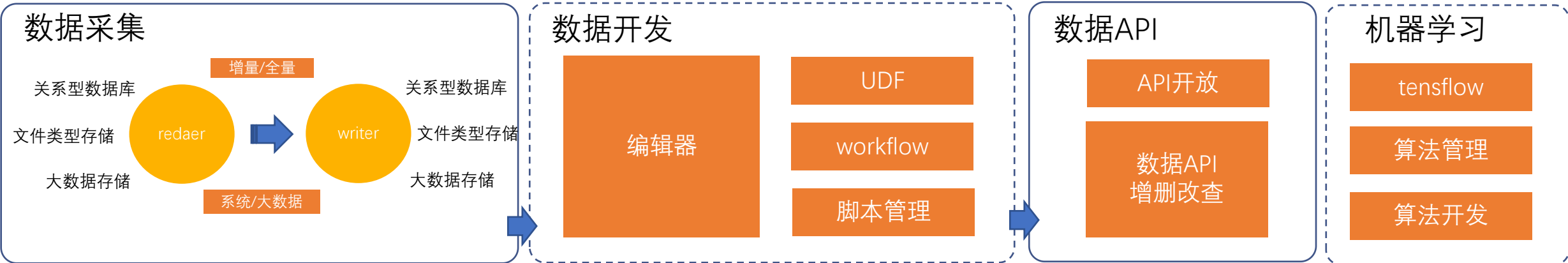
...

数据场景化应用（例如图灵）

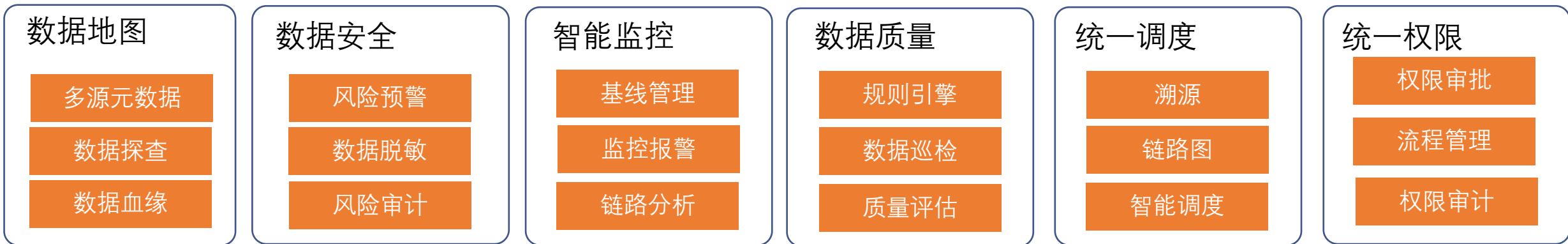
可视化、权限、指标、监控等大数据能力复用，提高了业务线系统“数据相关”开发的效率，并受大数据统一服务

构建一站式云上数据开发生态

一站式开发平台



6个底层基础服务能力

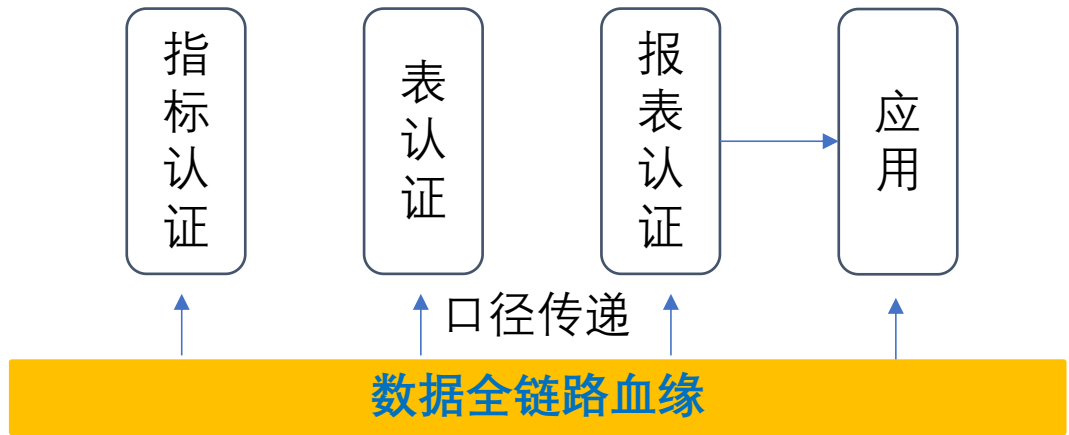


项目管理

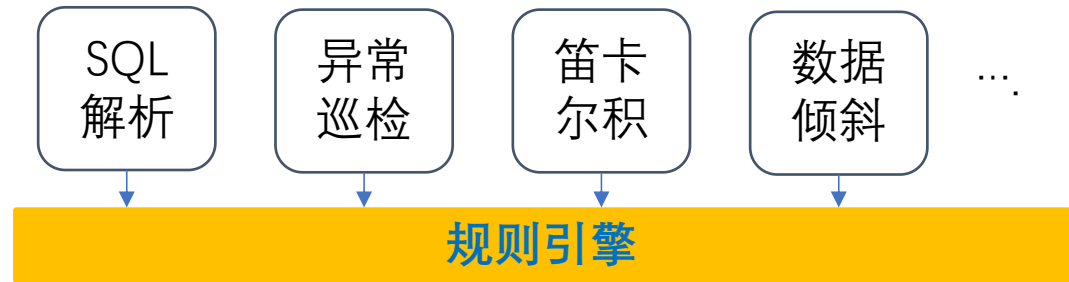
数据质量、数据成本可控

数据质量

通过认证链路解决数据准确、一致性问题

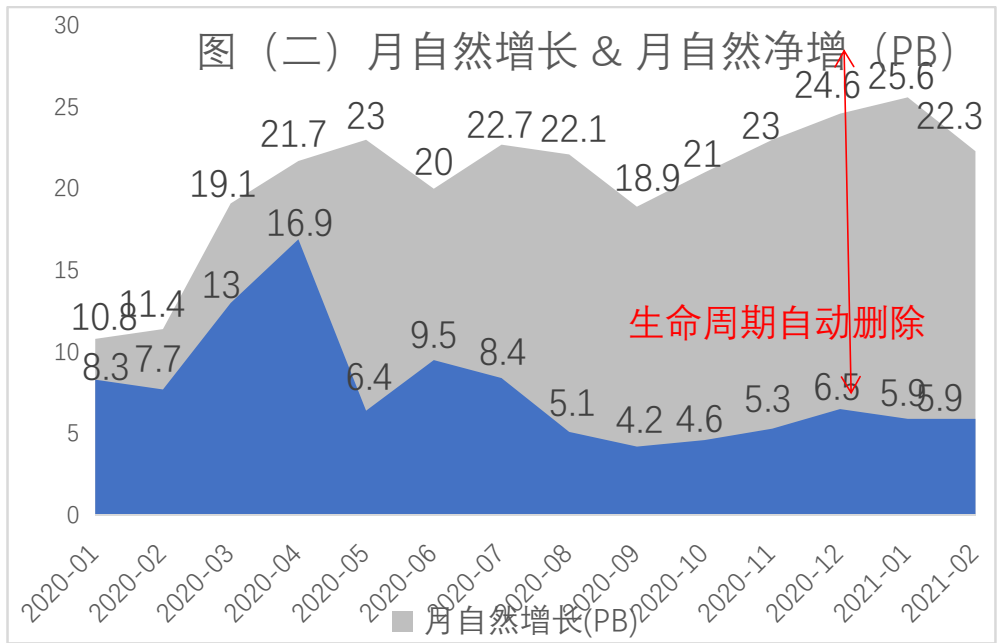
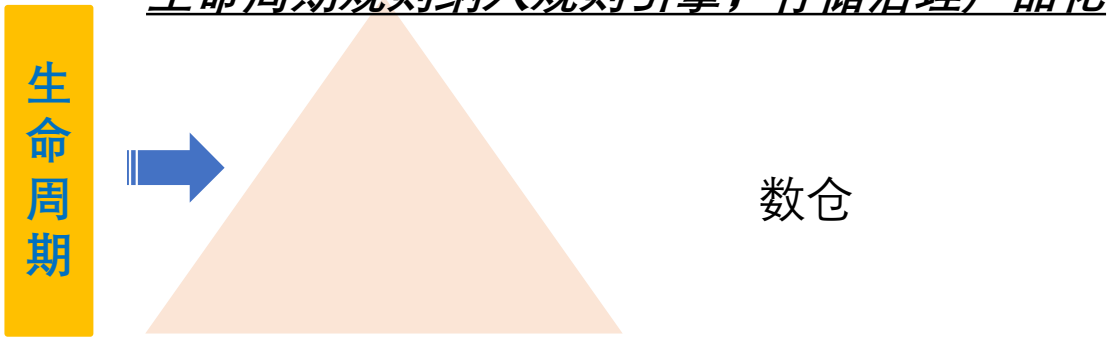


建立规则引擎，把控数据任务质量，巡检数据资产

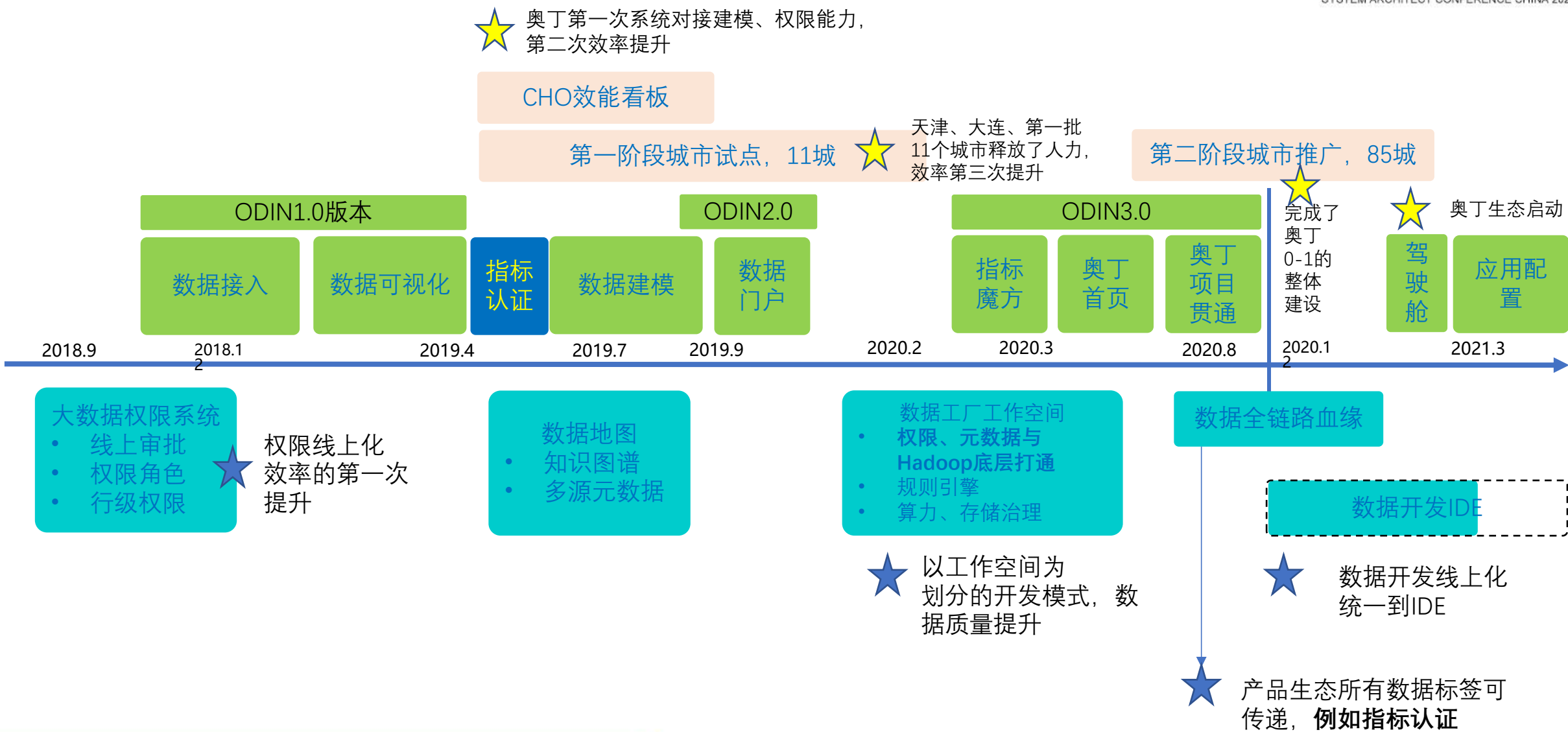


数据成本

生命周期规则纳入规则引擎，存储治理产品化



2018年至今产品建设过程-重要节点



从规模到标准，从产品到生态

重供给侧

由散到整，规模化

现在

未来

由量到质，标准化

完成0-1基础产品能力建设

ODIN

罗盘

ABtest

数据工厂

部落式开发

odin
数据工厂

需求输入

贝壳用户+场景

抽象提炼

更多的底层能力传递至上层

数据消费
产品体系

+

数据生产
产品体系

+

基础能力
产品体系

贝壳数据产品生态

功能模块解耦

配置封装能力

阅读者用户体验

如何快速的找到数据?
数据解读?
数据社交?

用数据创造无限可能

The background is a deep blue with a complex, abstract pattern of glowing blue wireframe cubes and rectangular prisms. These shapes are arranged in a way that creates a sense of depth and perspective, with some appearing to recede into the distance. A bright, horizontal lens flare or light streak cuts across the center of the image, passing behind the word 'THANKS'. In the upper left, there are some faint, stylized geometric shapes resembling flags or banners. The overall aesthetic is futuristic and digital.

THANKS