



基于 ES + MongoDB 的实时数据融合平台架构分享

Arthur, Tapdata Solution Architecture

目 录

1. 客户背景及痛点
2. 我们的方案
3. 技术实现
4. 客户反馈、展望未来

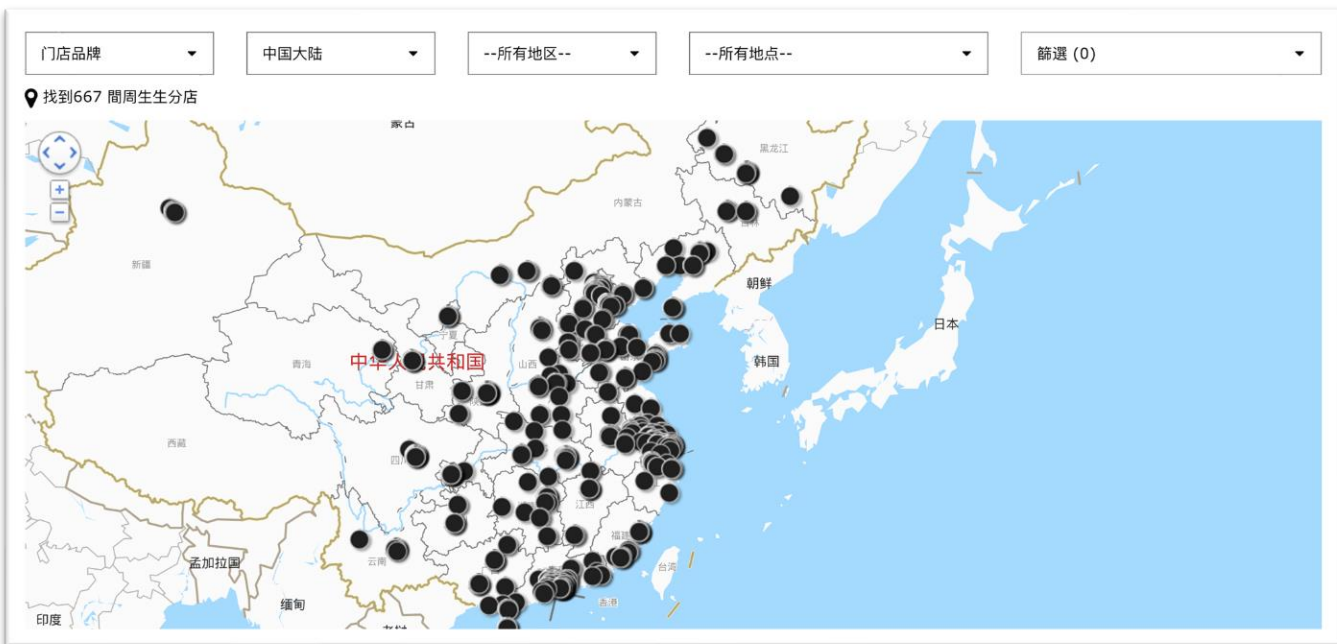
背景介绍

背景介绍

零售行业巨头

- ✓ 中国内地、香港、澳门、台湾扬名为显赫的黄金珠宝企业
- ✓ 大陆门店数量 ~ 700,
- ✓ 2019年营业额超 150 亿港元

- ✓ 旗舰系列 x 8
- ✓ 国际品牌 x 7
- ✓ 系列品牌总数 x 27



全部品牌系列	国际品牌	全部品牌系列	全部品牌系列
旗舰系列			
Infini Love Diamond 「全爱钻」	Hodel	Charme	薄荷系列
Charme	Kashikey	生生有礼	Infini Love Diamond 「全爱钻」
Promessa	Lady Heart	生生有蕙	Noir
g* 系列	Marco Bicego	文化祝福	One Piece
文化祝福	One Piece	Daily Luxe	PetChat
V&A 博物馆	V&A 博物馆	Diamond in Motion 「炫动钻饰」	Promessa
生生有蕙	故宫宫廷文化	g* 系列	V&A 博物馆
Let's Play 「好玩」		Happy Floret 「幸福花语」	故宫宫廷文化
		Hodel	
			Lace 「蕾丝」
			Lady Heart
			Let's Play 「好玩」
			Love Decode 「爱情密语」
			Loving Hearts 「心影」
			Marco Bicego

背景介绍 互联网+

- ✓ 拥抱互联网
- ✓ 与第三方电商、社交、媒体平台打通



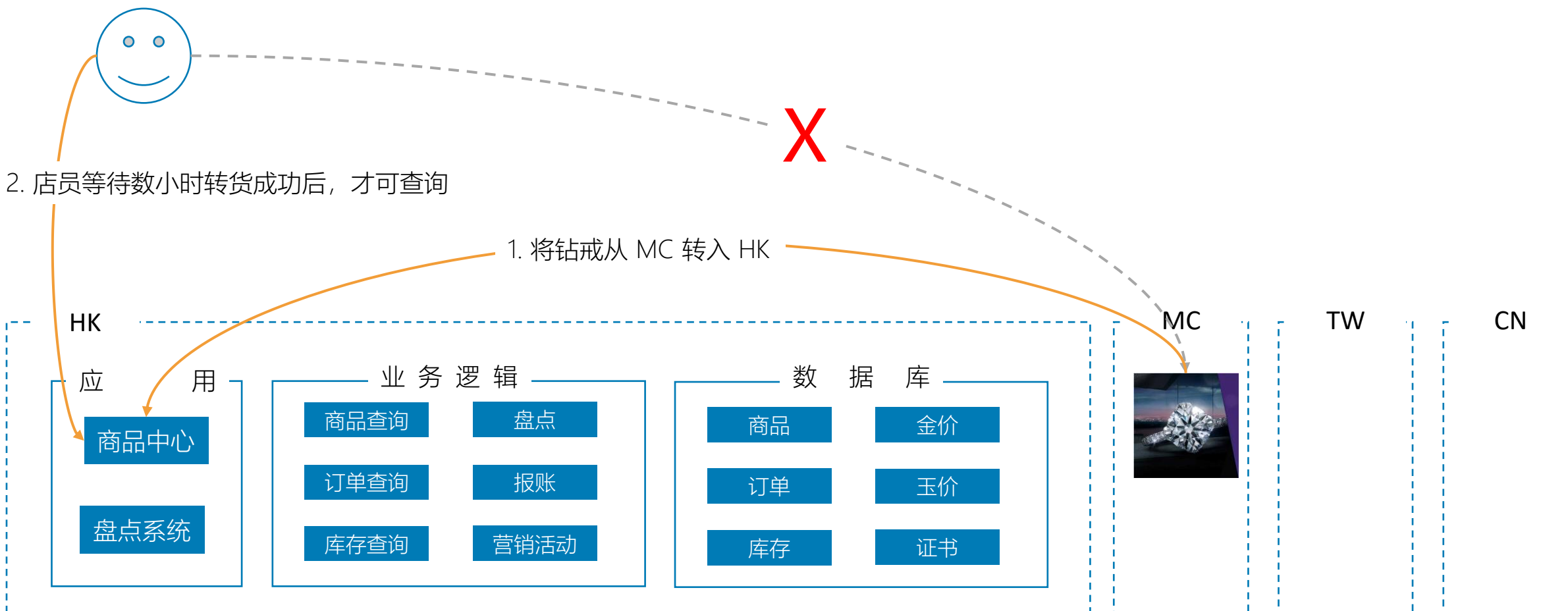
- ✓ 2019 年数百家门店线上线下活动过万
- ✓ 业务需求旺盛，产品快速迭代适应市场



背景介绍

割裂的商品中心

如果香港门店店员想要在 HK 卖一件货，但这件货在 MC

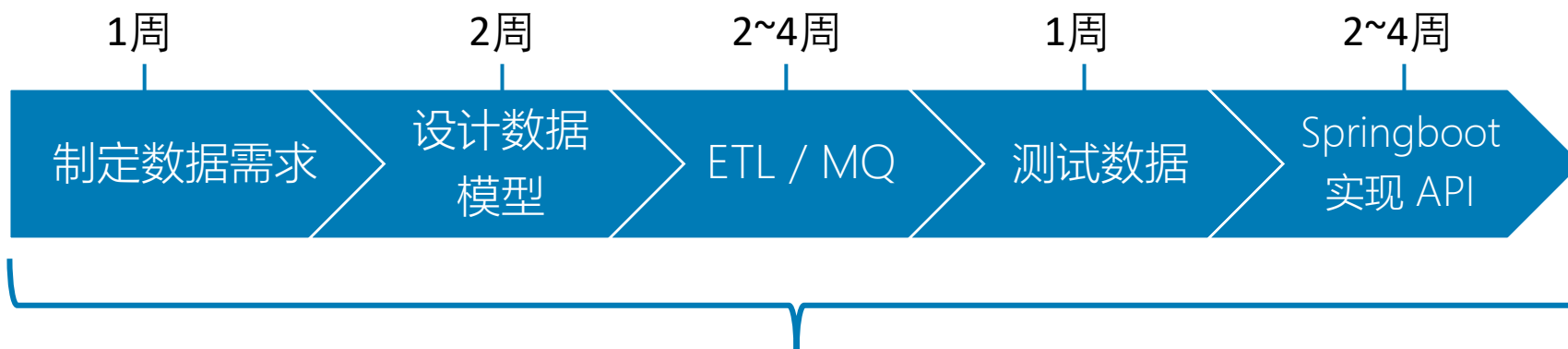


背景介绍

研发赶不上市场变化



业务以为的研发周期 **1~2 周**



IT 提供的研发周期 **4~12 周**



背景介绍

当前面临的挑战

烟囱设计

- 数据孤岛
 - 中港台澳 4 地 9 套 Oracle, 维护成本高
 - 库存、订单多地冗余, 状态无法统一
 - 无法获取完整订单信息, 商品属性, 库存记录

- 系统多、时间久远
 - 历史系统 20+ 年
 - 核心系统维护者只有几人, 业务无法扩展, 只能维护
 - 新老数据交替存在数据库中, 数据难以理解

难以维护

研发无法快速响应业务

- 业务需求旺盛
 - 开发人员需要应对不同业务组提交的数据请求, 并快速响应
 - 堆积的开发需求最早可追溯到 3 年前

- 跨部门协作难
 - 要一份数据需要经过多部门沟通, 时间周期长
 - 对数据不理解, 需要与负责同事交流

取数难

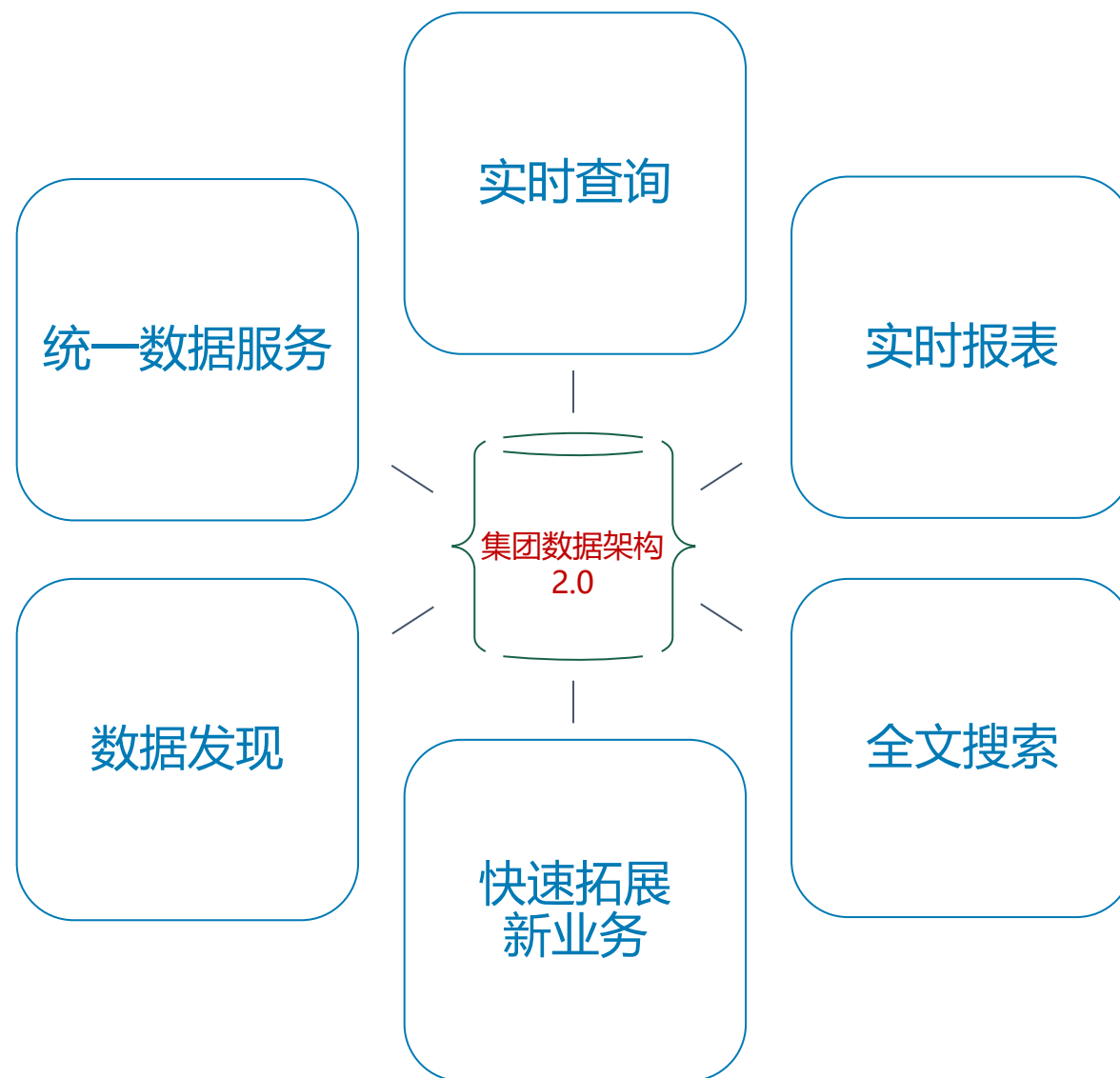


背景介绍

集团对新一代数据架构的诉求

业务诉求

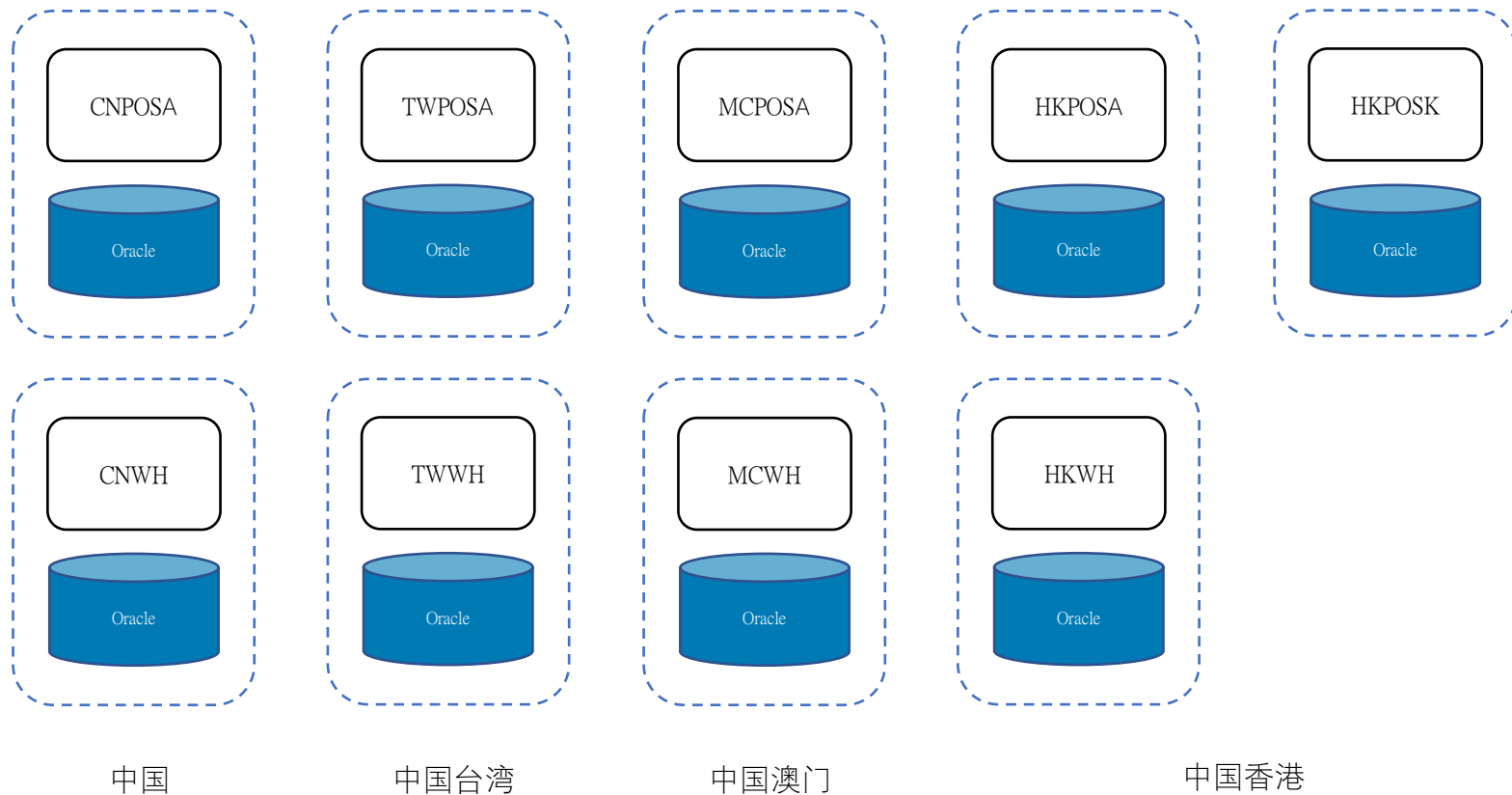
- ❖ 数据模型线性增长，关系型数据结构越来越复杂，修改字段时，像在电源插座上找唯一能用的插位
- ❖ 20年前的 IT 架构结构复杂，大量存储过程，学习周期慢，改动困难
- ❖ 研发速度跟不上需求变化
- ❖ 需要一个高效的数据开发服务平台支撑企业内部敏捷开发和 DevOps
- ❖ 重复建设用户数据、订单数据，统一数据困难



我们的方案

我们的方案

产生 IT 瓶颈的原因



10 年前为了更好的服务客户
为每个地区架设单独的系统

今天这些重复的系统意味着

- 分割的商品数据
- 冗余的库存信息
- 延时的订单消息

我们的方案

三大数据痛点

10 万商品种类信息
分布在 20 张表

- 3 级关联
- 传统 IT 设计模式
- 跨实例查询

100 万商品型号
需要通过映射表查询

- 字典表难以理解
- 业务映射表关系复杂
- 跨实例查询

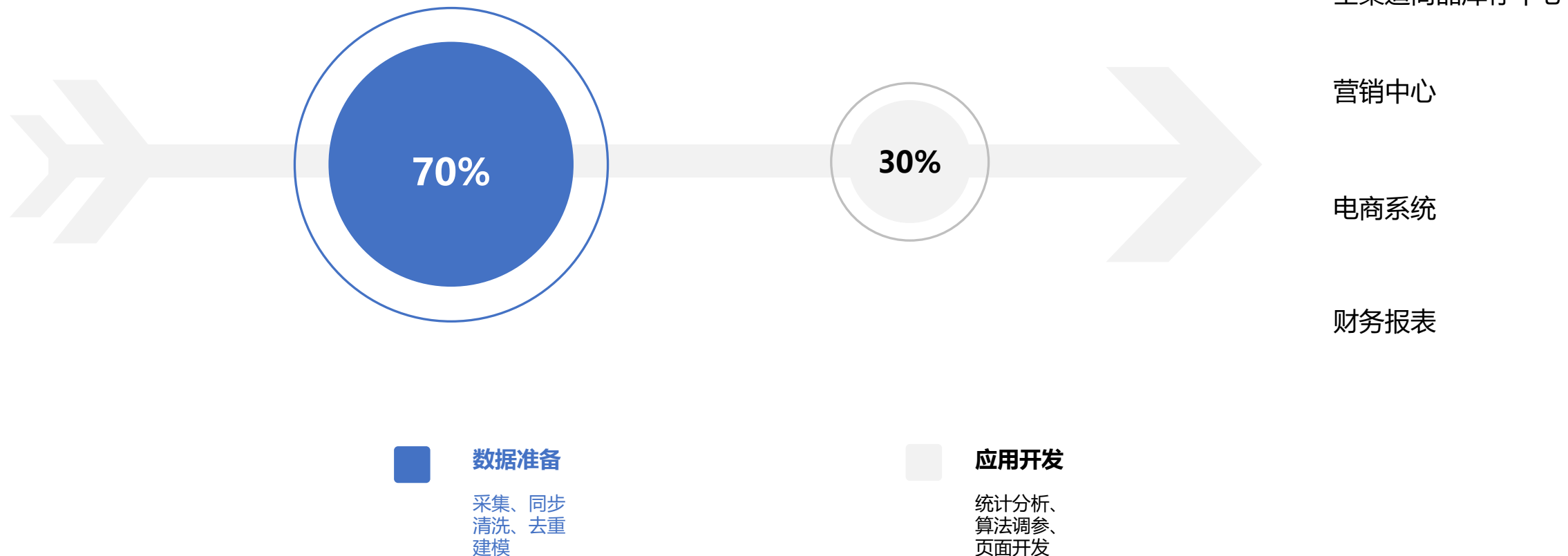
1.5 亿库存订单
存放在 9 个 Oracle 中

- 异地订单状态更新
- 数据库分布多地，数据延时
- 订单逻辑复杂，不易扩展

我们的方案

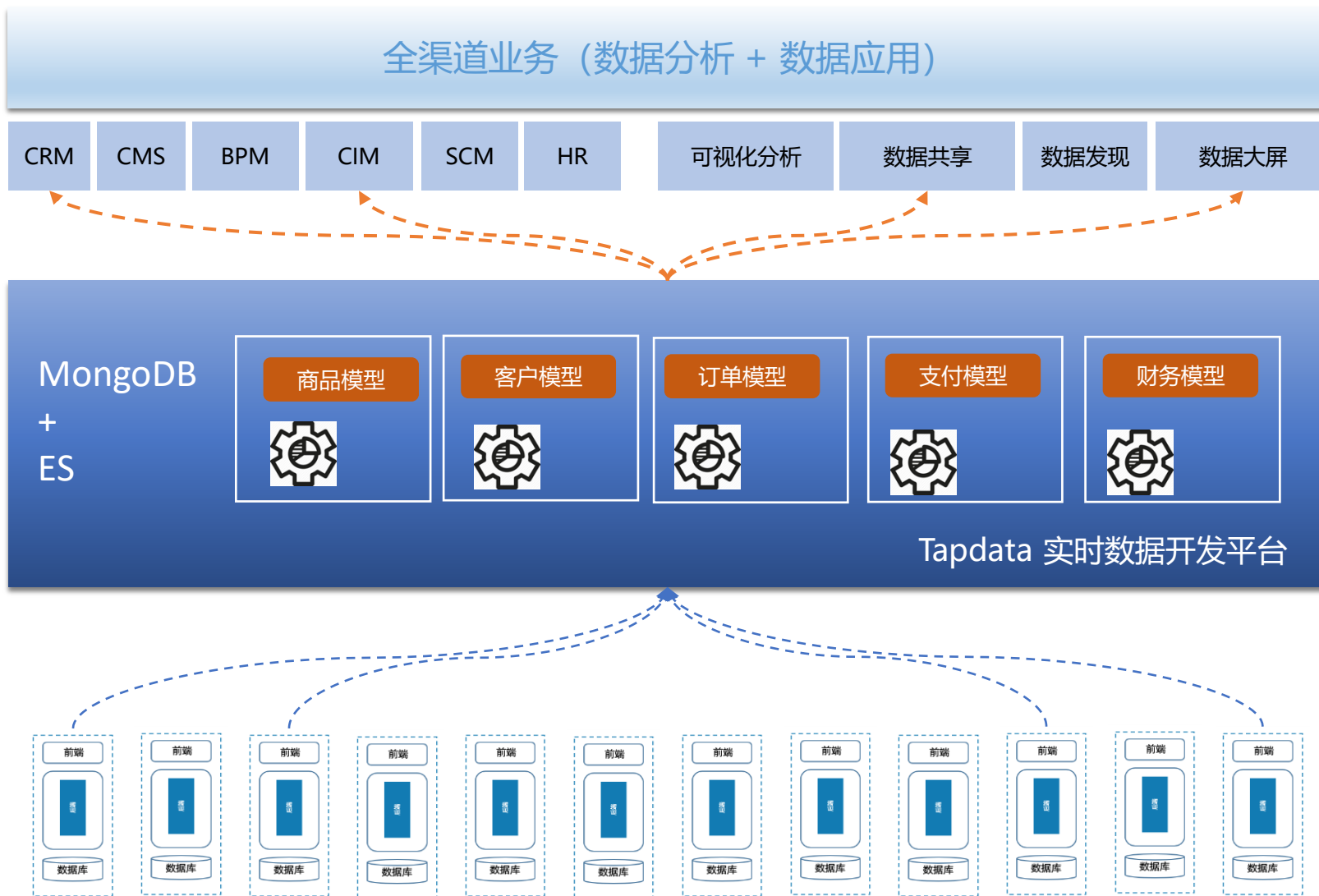
数据能力成为数字化的瓶颈，制约集团 IT 生产力

数据准备消耗了 70% 的工作量，仅剩下 30% 时间用于核心业务开发



我们的方案

通过数据中台解决企业数据孤岛、研发周期长的痛点



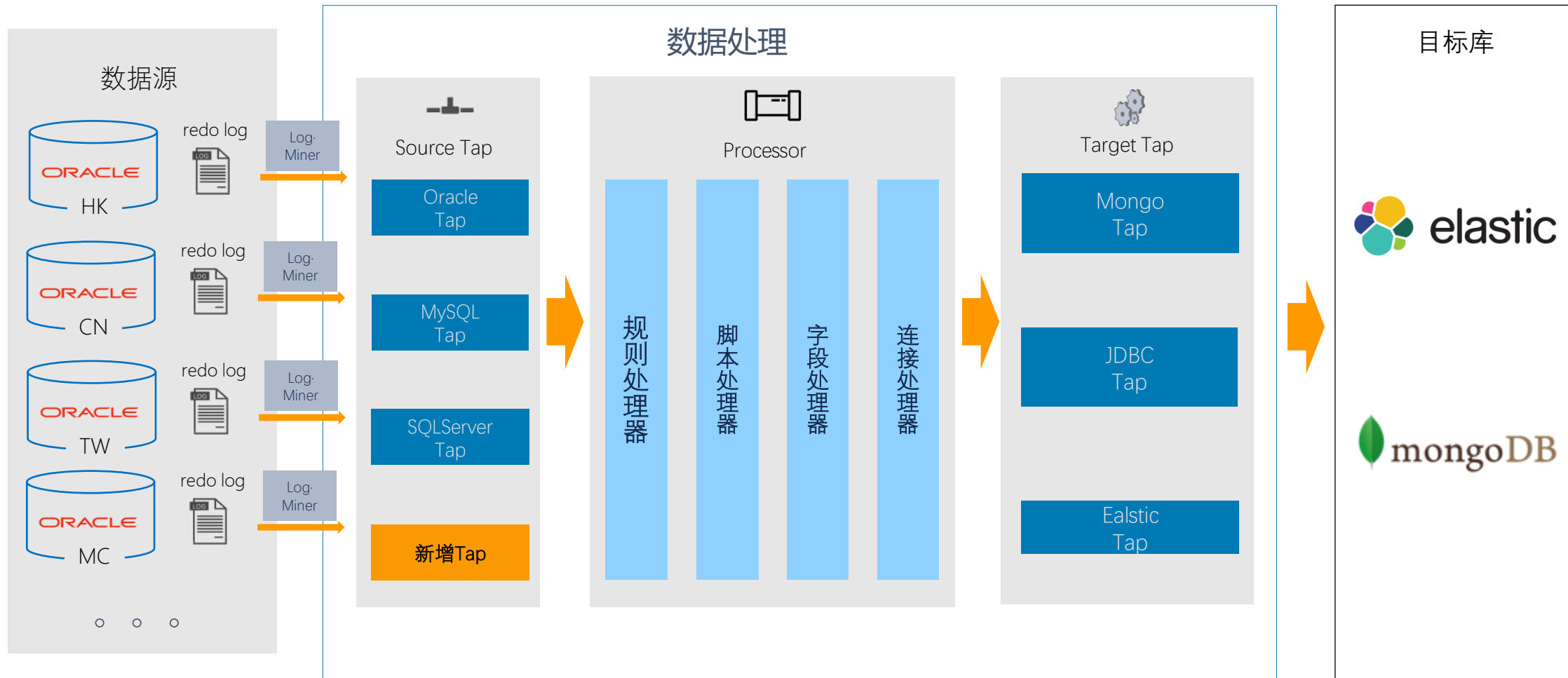
3. 基于中台快速开发数据应用

2. 构建三层业务数据模型

1. 数据融合, 打破孤岛

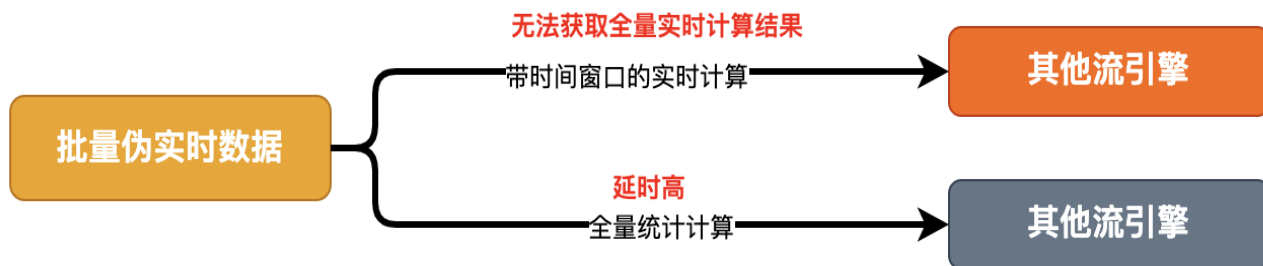
我们的方案

实时数据同步

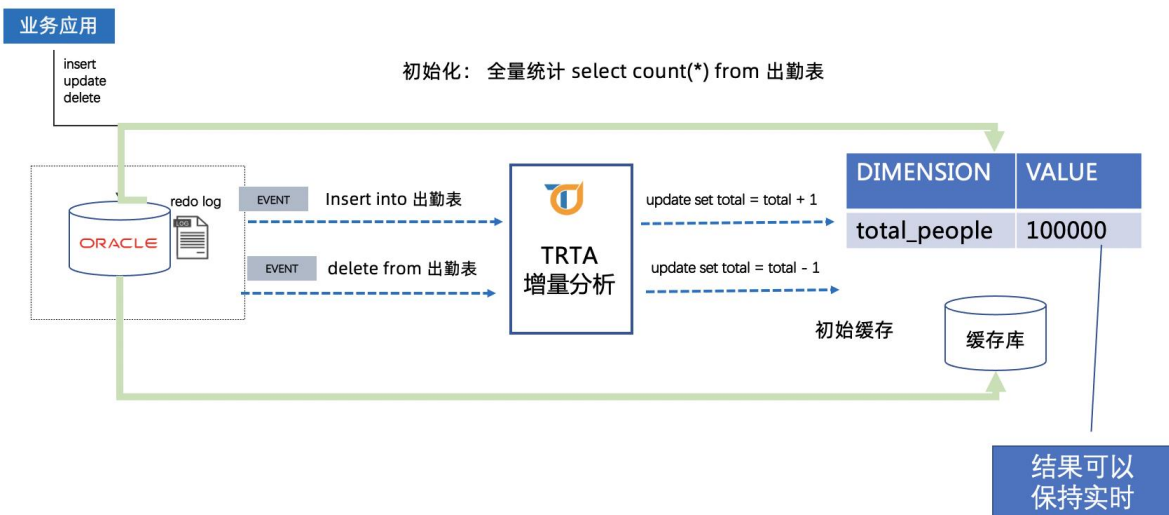


我们的方案

实时增量聚合 vs 批量大数据计算



VS



常见同步工具能力

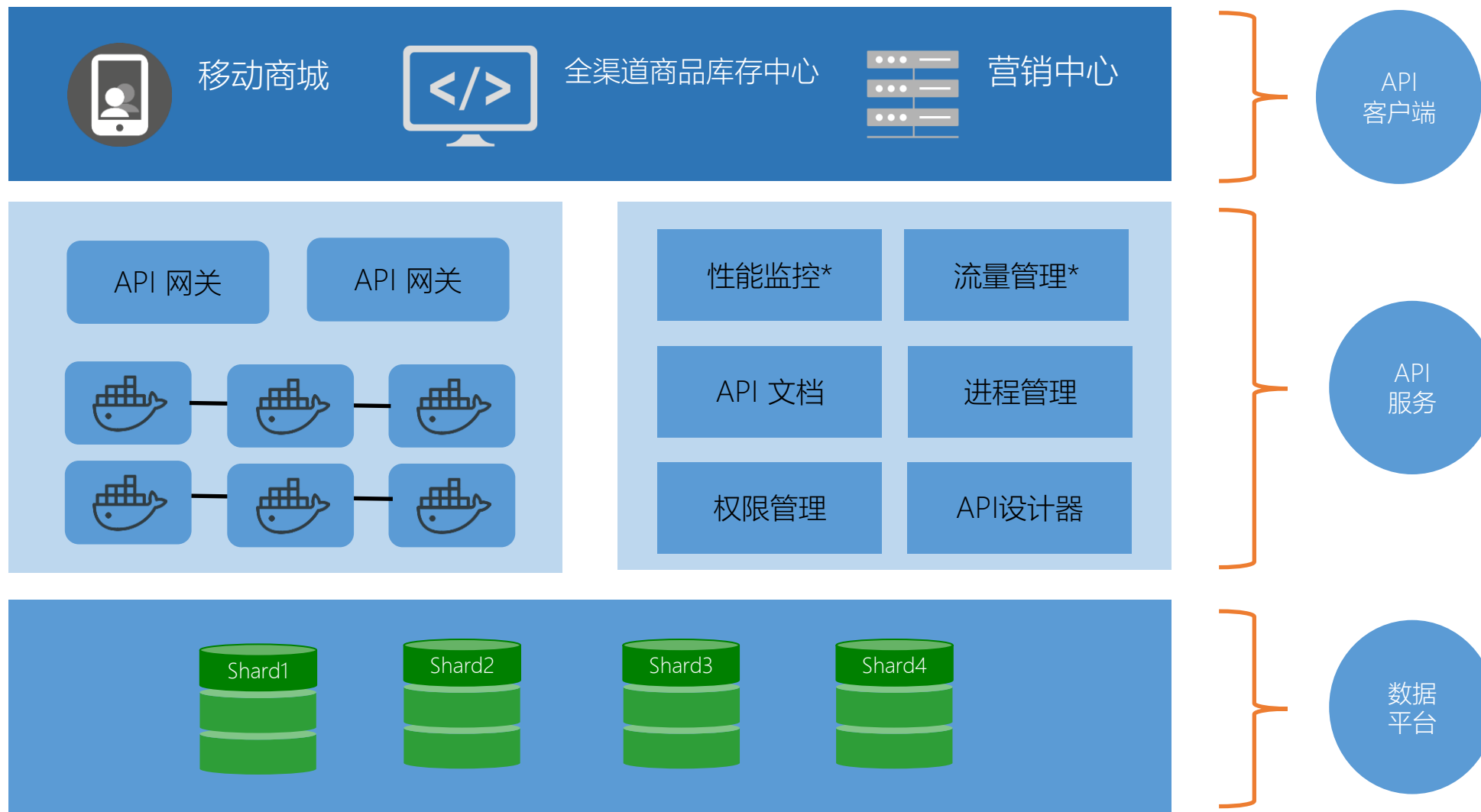
- ✓ T+1, 凌晨跑批数据, 只能看到前一天的数据
- ✓ 往往用于分析前一天的应用结果

TAPDATA 毫秒级实时同步

- ✓ 基于隔离Snapshot 计算首次结果
- ✓ 在缓存库保留相关全量计算原始值
- ✓ 增量数据到达时, 根据修改模式和最近记录修改状态, 计算delta, 更新结果集
- ✓ 用户可实时查看数据大屏

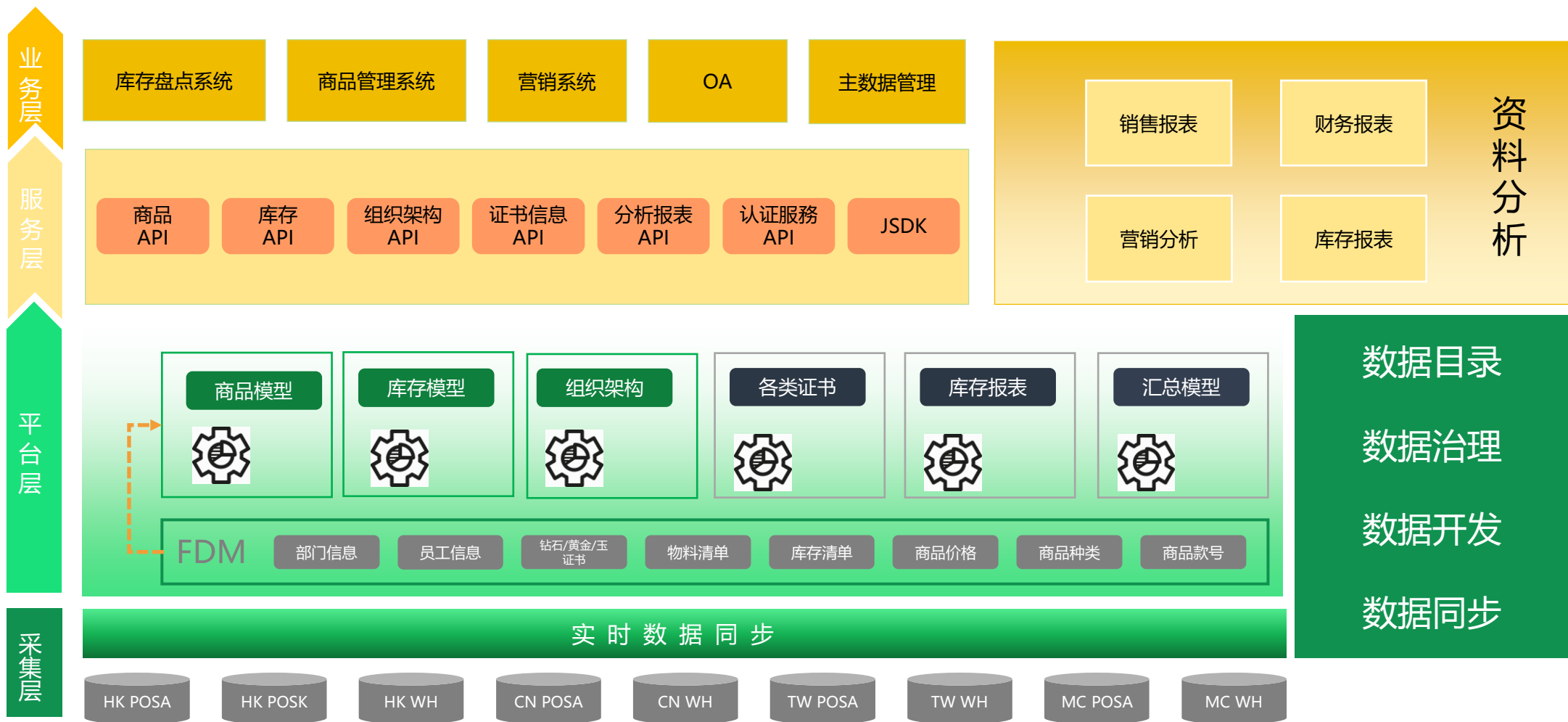
我们的方案

统一数据服务

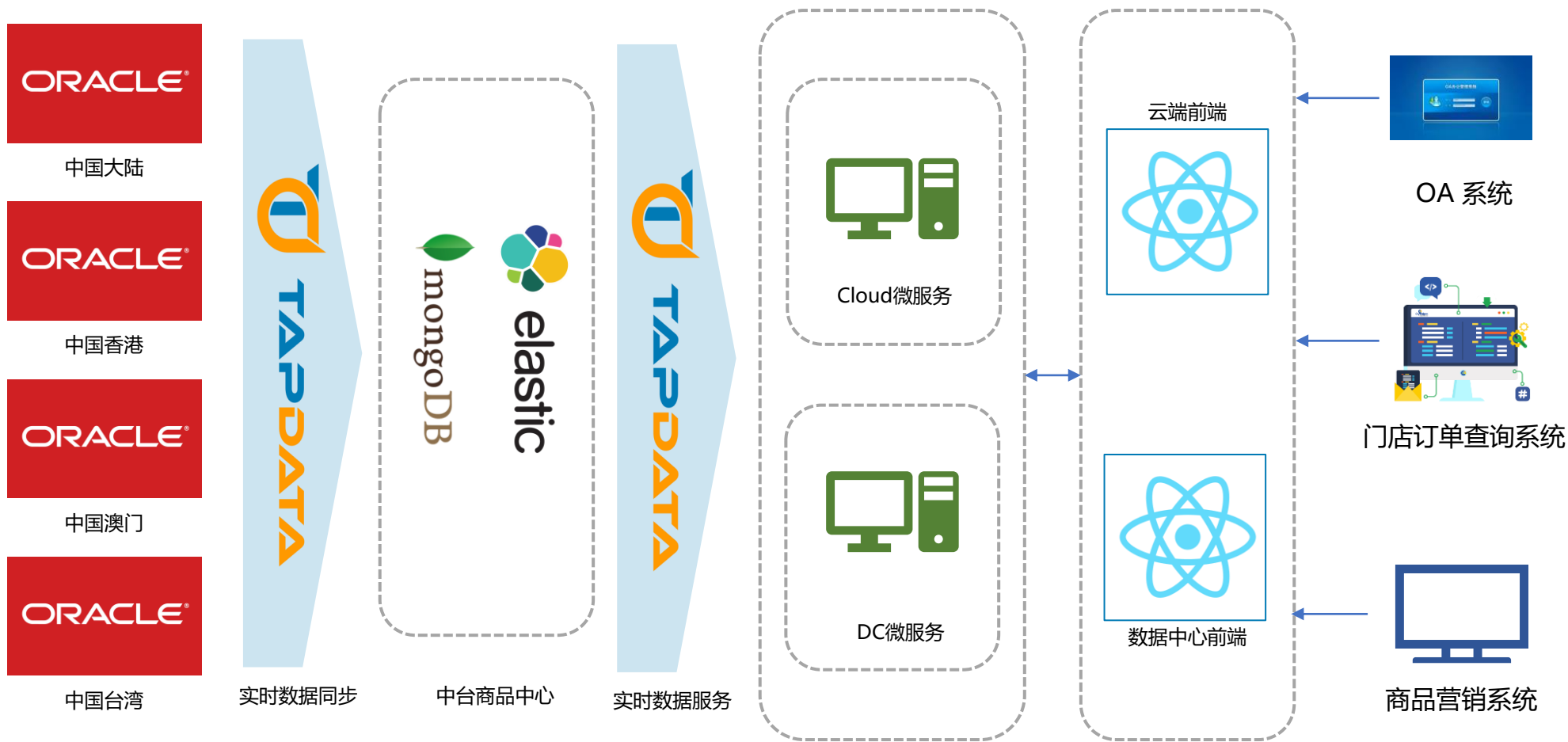


我们的方案

构建实时数仓和业务数据双底座

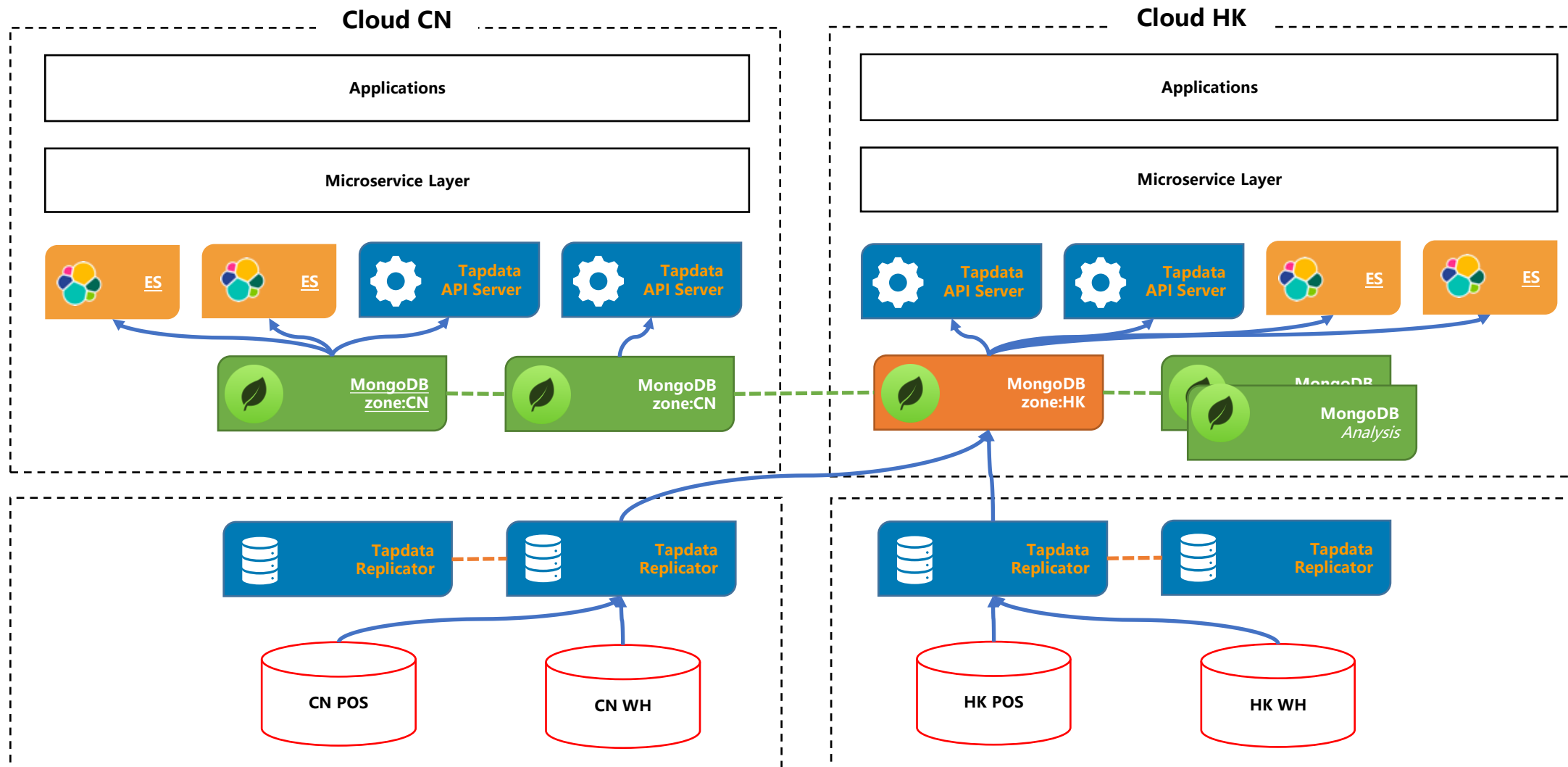


我们的方案 技术平台搭建



我们的方案

DaaS 物理架构图



我们的方案 DaaS 技术栈



技术实现

DaaS 项目规划 – 关键步骤

步骤	描述	里程碑
计划	项目需求确认 组建团队 确认技术栈	团队成员确认 逻辑、物理架构图 项目范围界定 数据需求确认 数据源确认
准备工作	创建数据库, 准备数据开发工具	tapdata 可成功连接所有数据源
FDM 层同步	初始化全量同步 Oracle 数据到 DaaS 制定数据规则	DaaS 存储全量 Oracle 数据
MDM 建模	根据需求创建 MDM 模型 为数据打标签	成功创建数据模型
ADM 建模	根据需求创建 ADM 模型 创建业务 API 测试 API	一个应用基于 DaaS 的 API 成功运行, 并保持数据实时更新
测试、上线	监控, 告警, 备份, 多数据中心部署	

DaaS 项目规划 – 人员安排

角色	职责
数据架构师	设计全链路架构，技术评估，确定用户案例 最好是对业务比较熟悉的
数据工程师	创建、维护数据模型 创建、维护、测试、验收数据同步任务
运维工程师	搭建 DaaS 所需硬件，包括数据库、中间件等 保证系统稳定运行，监控，数据备份还原，告警 系统高可用，按需扩容
数据管理者	指定数据标准，规则 设计数据目录，维护元数据
数据分析师	提供商业智能分析报告 创建、维护数仓模型 大数据分析
项目经理	协调团队，明确需求，为 DaaS 赋能 保证项目准时、高质量交付

DaaS 项目规划 – 人天预估

Service Scope	Description	Deliverable
Assessment	Current Data Architecture, Data Systems, Data models, Data Processing	Comprehensive Report DaaS Proposal
Data Architecture Design	Based on the customer infrastructure and future IT planning, design DaaS architecture that fits the cloud/hardware/system requirement	Architectural design and physical design
DaaS Application Design	Work with application team, provide system architecture, data / API design and best practices	Application design documentation pertaining to DaaS integration
DaaS Implementations	Including system provisioning, security hardening, oracle data migration, data processing, data modeling etc	The achievement of Mature Level of DaaS
Data Governance	Create a metadata system for customer's data set, define standard data models for common data types, establish data governance rules and build enforcement on DaaS platform	Data catalog and metadata structure Governance rules

Man days: 100

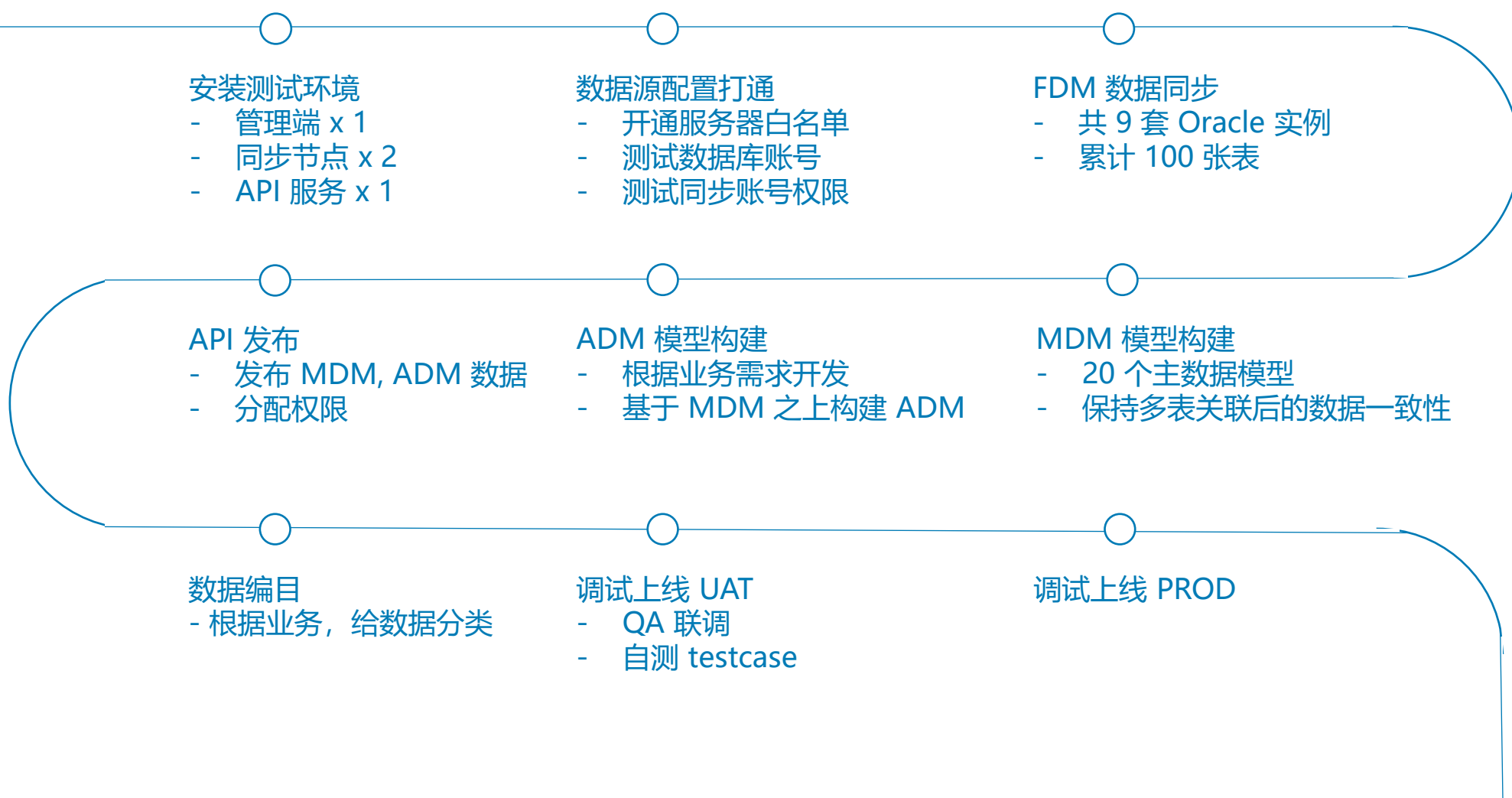
Validity: 1 year

Delivery Method: Onsite or Remote

Recommended Frequency: 2 days per week

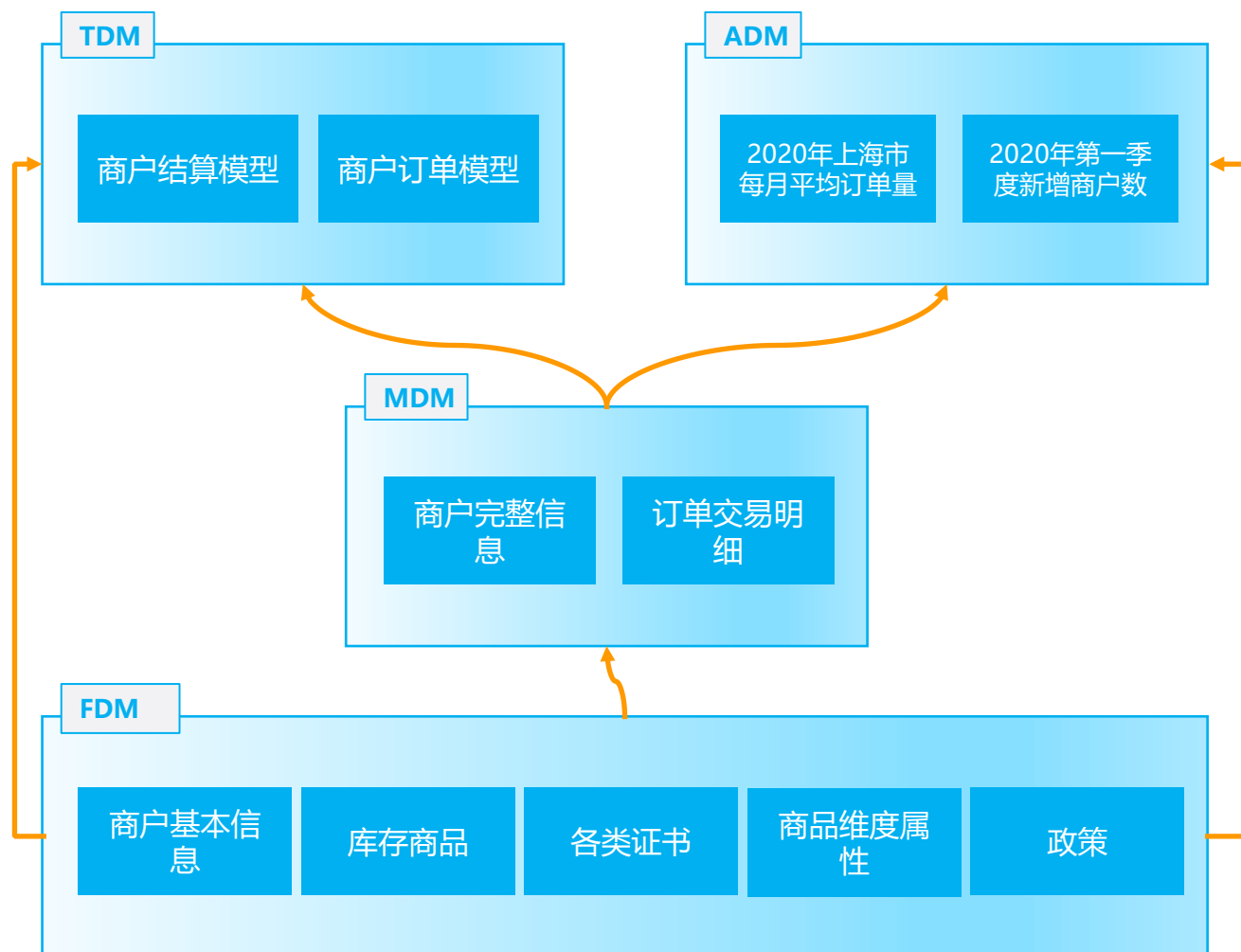
服务器	用途
CPU: 8 RAM: 16GB	广州: Tapdata 管理端 x 2, Tapdata API 服务 x 2 香港: Tapdata 管理端 x 2, Tapdata API 服务 x 2
CPU: 16 RAM: 64GB	广州: Tapdata 同步节点 x 2 香港: Tapdata 同步节点 x 2

DaaS 项目实际实施进程



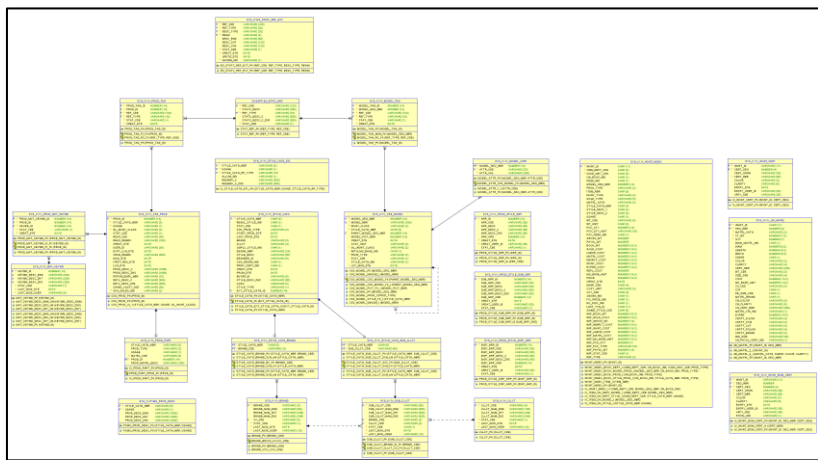
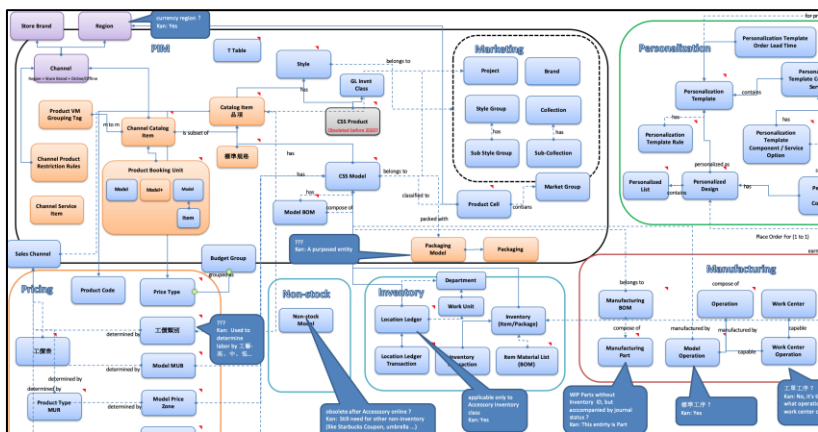
技术实现 三层模型

- FDM, Foundational Data Model:
 - FDM层存储来自于业务源系统的采集数据;
- MDM, Master Data Model
 - 根据业务主题对所有相关业务源数据进行整合。
 - 按照自然模型进行数据建模
- TDM, Transactional Data Model
 - 面向业务流程类系统进行模型设计
- ADM, Analytical Data Model
 - 面向企业分析型应用进行建模设计



技术实现

使用自然模型代替传统三范式



PRODUCT

```

{
  product_id:xx
  catalogItem: xx,
  productStatus: xx,
  inventoryClass: string,
  model: {
    modelNbr:xx,
    modelSeqNbr:xx,
    productType:xx
  }
  style:xx,
  usage:xx,
  productName:xxx,
  productDesc:xxx
  channelClusterCode:xx
}

```

INVENTORY

```

{
  lastModifiedDate: xx,
  masterCatalogItem: xx,
  modelSequenceNumber: xx,
  productType: xx,
  sku: xx,
  inventoryId: xx,
  itemNumber: xx,
  originalPrice: xx,
  physicalWeight: ,
  regionCode: ,
  departmentCode: ,
  pendingIndicator: boolean,
  onBookIndicator: boolean,
  inventoryStatus: xx
}

```

自从有了 DaaS 之后

有了 DaaS 之后 全渠道商品订单库存中心

- ✓ 中港台澳门店店员使用同一套后台，查询全国商品
- ✓ 前台查询一次，获取商品所有信息
- ✓ 查询次数从 5 次缩减为 1 次
- ✓ 查询响应时间从 10 秒缩减为 3 秒内



有了 DaaS 之后 企业数据目录

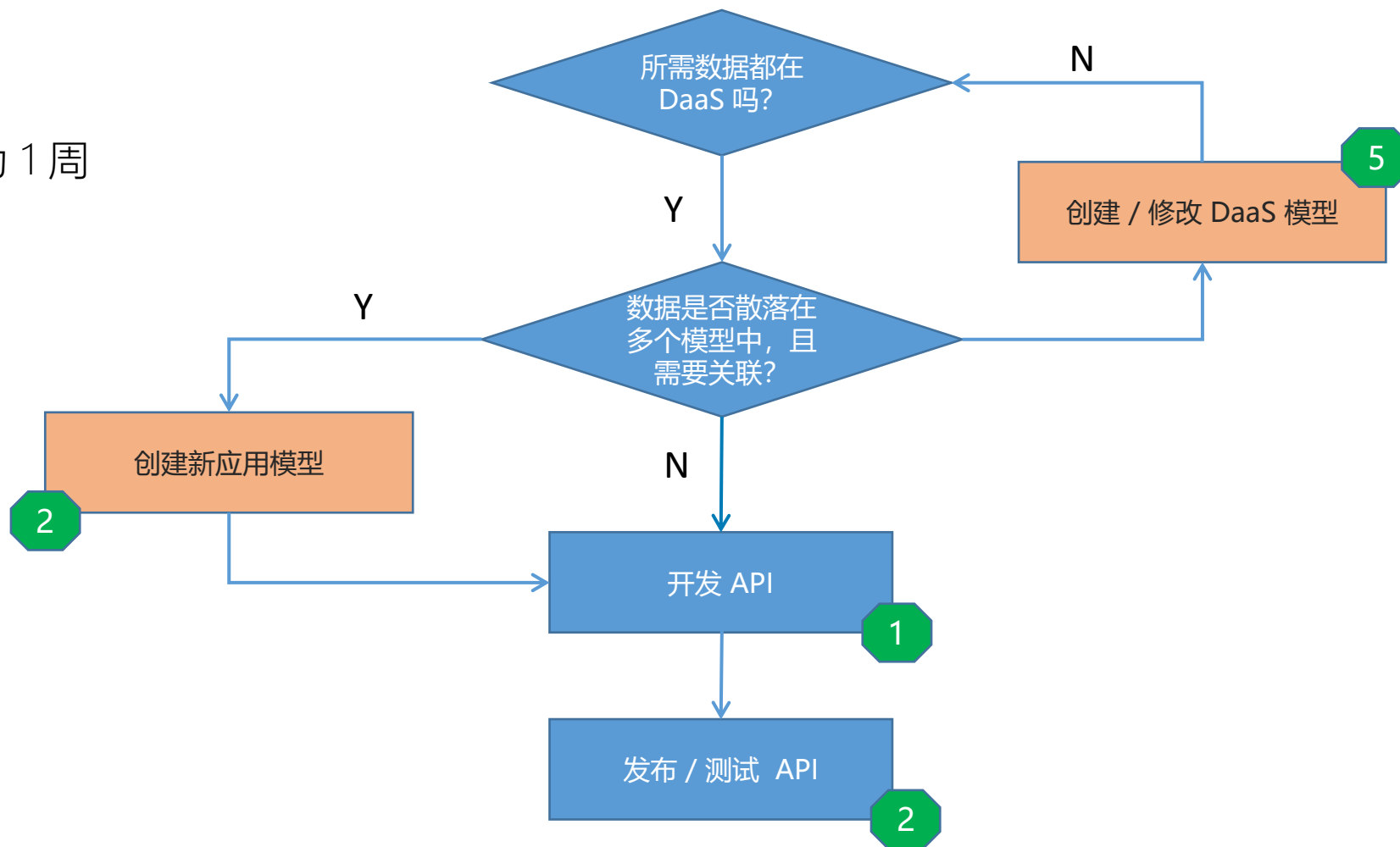
- ✓ 研发在平台上可以快速查到想要的数据库
- ✓ 减少 DBA 日常 70% 查询工作
- ✓ 提升研发 90% 开发效率，无需跨部门沟通数据

The screenshot displays the TAPDATA data catalog interface. On the left, there is a 'Tags' sidebar with a tree view of categories including DaaS, FDM, MDM, ADM, TDM, SYSTEM, All Connections, and Data API. The main area shows a search for 'IM_COMPANY' in the 'Collection' field, which is highlighted with a red box. Below the search bar, there are buttons for 'BATCH CLASSIFICATIONS' and 'CREATE MODEL'. A table lists search results with columns for Name, Superior node, Meta Type, Last update user, Last updated, and Operate. The table contains five rows of data, all with 'Collection' as the Meta Type and 'admin@admin.com' as the Last update user.

<input type="checkbox"/>	Name	Superior node	View	Meta Type	Last update user	Last updated ↓	Operate
<input type="checkbox"/>	POSS_IM_COMPANY	POSS_FDM_DEV	View	Collection	admin@admin.com	2021-02-02 14:45:21	
<input type="checkbox"/>	FDM_ALPHA_IM_COMPANY	T_Agent_Local_DaaS	View	Collection	admin@admin.com	2021-01-28 23:15:25	
<input type="checkbox"/>	FDM_ALPHA_IM_COMPANY	S_Agent_Local_DaaS_1	View	Collection	admin@admin.com	2021-01-19 14:49:49	
<input type="checkbox"/>	FDM_ALPHA_IM_COMPANY	S_Agent_Local_DaaS_3	View	Collection	admin@admin.com	2021-01-08 17:47:12	
<input type="checkbox"/>	FDM_ALPHA_IM_COMPANY	S_Agent_Local_DaaS_2	View	Collection	admin@admin.com	2021-01-08 17:45:21	

有了 DaaS 之后 API 研发流程调整

- ✓ 研发周期从 2~3 个月缩减为 1 周
- ✓ 研发当天可验证 N 次结果



2 工时：人天

有了 DaaS 之后 API 研发流程调整

里程碑	事项	参与者	交付物	工时 (人天)
提交需求	向 DaaS 负责人提出需求	应用负责人	需求清单	
分析需求 & 确认方案	明确需求 沟通方案 预估人天	PM Data Engineer Consultant	方案 预估人天 实施计划	2
开发 & 测试 API	Configure API in DaaS Test and Adjustment	Data Engineer	API API 文档	1 ~ N N: 需要新的数据模型
UAT 测试	根据测试用例验收 API	QA	测试报告	1
上线生产				1

有了 DaaS 之后 API 研发流程调整

- ✓ 研发在平台上自主调用、测试 API
- ✓ 研发在平台上直接查看 API 文档说明

The screenshot displays an API documentation interface. At the top, there is a list of endpoints:

- GET** `/api/v1/MasterData/IM_COMPANY/batch` Download batch import excel template
- GET** `/api/v1/MasterData/IM_COMPANY/count` count data
- POST** `/api/v1/MasterData/IM_COMPANY/find` get record list by page and limit
- GET** `/api/v1/MasterData/IM_COMPANY/{id}/delete` delete record by id

The selected endpoint is the DELETE endpoint. Below the endpoint list, there is a "Parameters" section with a "Try it out" button. The parameters table is as follows:

Name	Description
id ★ required	
string	
(path)	

Below the parameters section, there is a "Responses" section. The response table is as follows:

Code	Description	Links
204	<code>MasterDataIM_COMPANY DELETE success</code>	No links

我们的方案

DaaS 实现的价值

能力沉淀与覆用

高效共享价值信息
轻松复用历史数据
快速发现数据价值

服务化程度提高

灵活的交互式服务
缩短开发周期，提升市场竞争力

更新的技术栈

高可用 & 负载均衡
多租户 / 云计算
给用户带来数据价值

来自 IT 价值

向 企业及价值



更好的支持市场和营销
增加营收



降低运营成本
提升运营利润



改善 用户体验

下一步

- 免费云版体验产品
cloud.tapdata.net
- 联络 team@tapdata.io 进行POC

