

DTCC

数 / 造 / 未 / 来

第十二届中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA 2021

2021 年 10 月 18 日 - 20 日 | 北京国际会议中心





数 / 造 / 未 / 来
第十二届中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2021

分布式架构下的数据共享与数据交换

海保人寿保险股份有限公司

架构师 胡赵勇

DTCC
2021



北京国际会议中心

🕒 2021/10/18-10/20



ChinaUnix^{.net}

ITPUB

目录

CONTENTS

01 背景介绍

02 数据的分类与处理

03 数据共享

04 数据交换





以应用为中心

应用复杂性的上升和微服务架构的流行，数据正变得越来越以应用为中心。带来可伸缩性的好处，也带来了融合计算的障碍。海保应用架构管理平台



信息孤岛问题突出

数据散落在不同的数据库、消息队列、文件系统中，计算平台如果直接访问这些数据，会遇到可访问性和数据传输延迟等问题



多变的数据诉求

实时和非实时的数据请求，落地与不落地的数据需要。必须将数据从孤立的数据源中采集出来，汇集到可被计算平台高效访问的目的地即数据中心

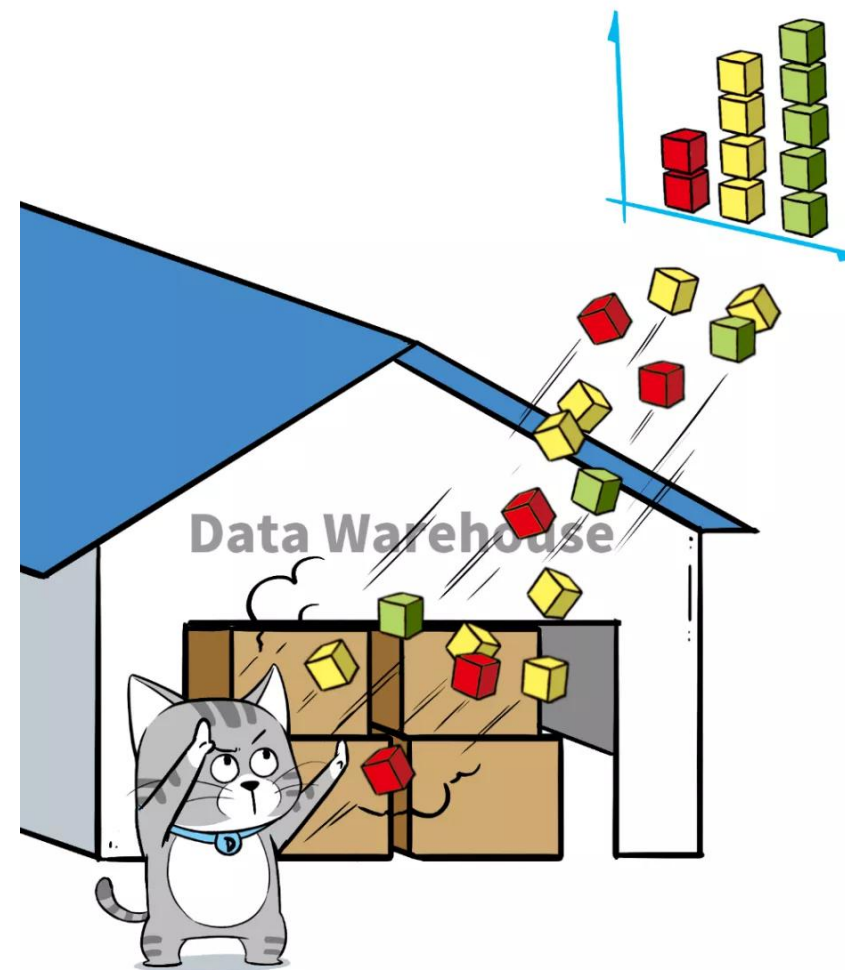
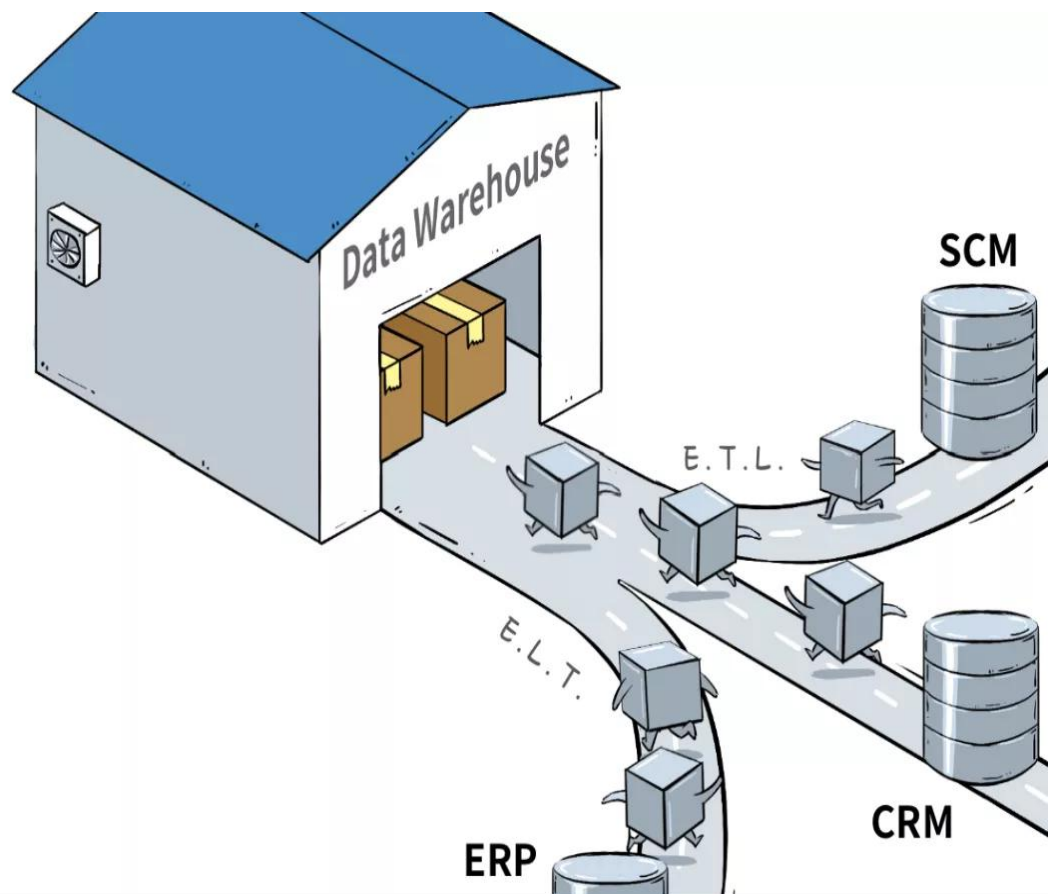


第三方数据

充分利用数据营销获客，会与拥有数据势能的第三方合作，寻求数据共享。这时，需要实时接入新的 API，且每个 API 会不断迭代，导致维护成本巨大。如何对接、统一管理几十个 API 接口成为问题



数据预处理



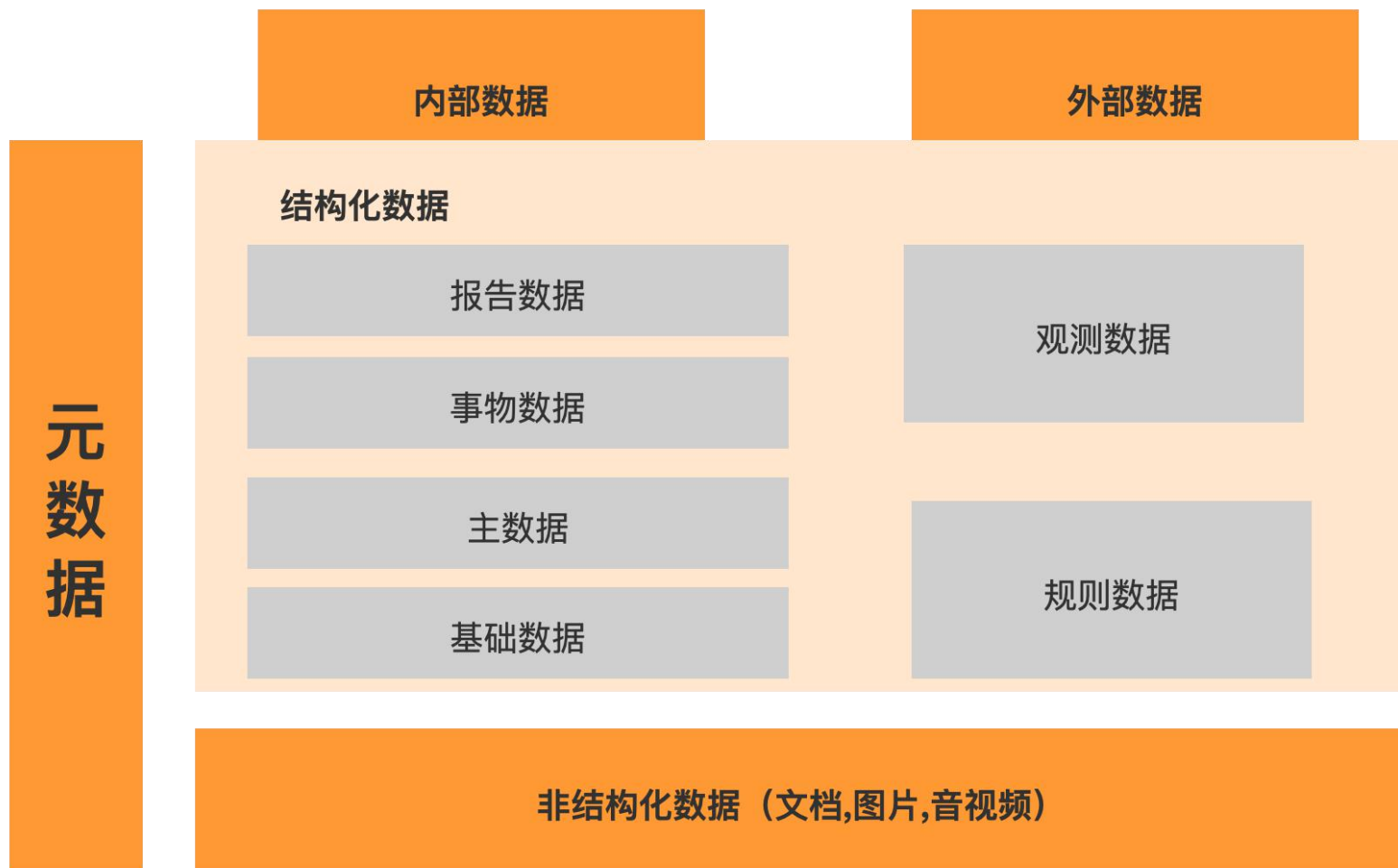


Part 02 数据的分类与处理

分类管理，有的放矢



数据分类



数据分类的目的

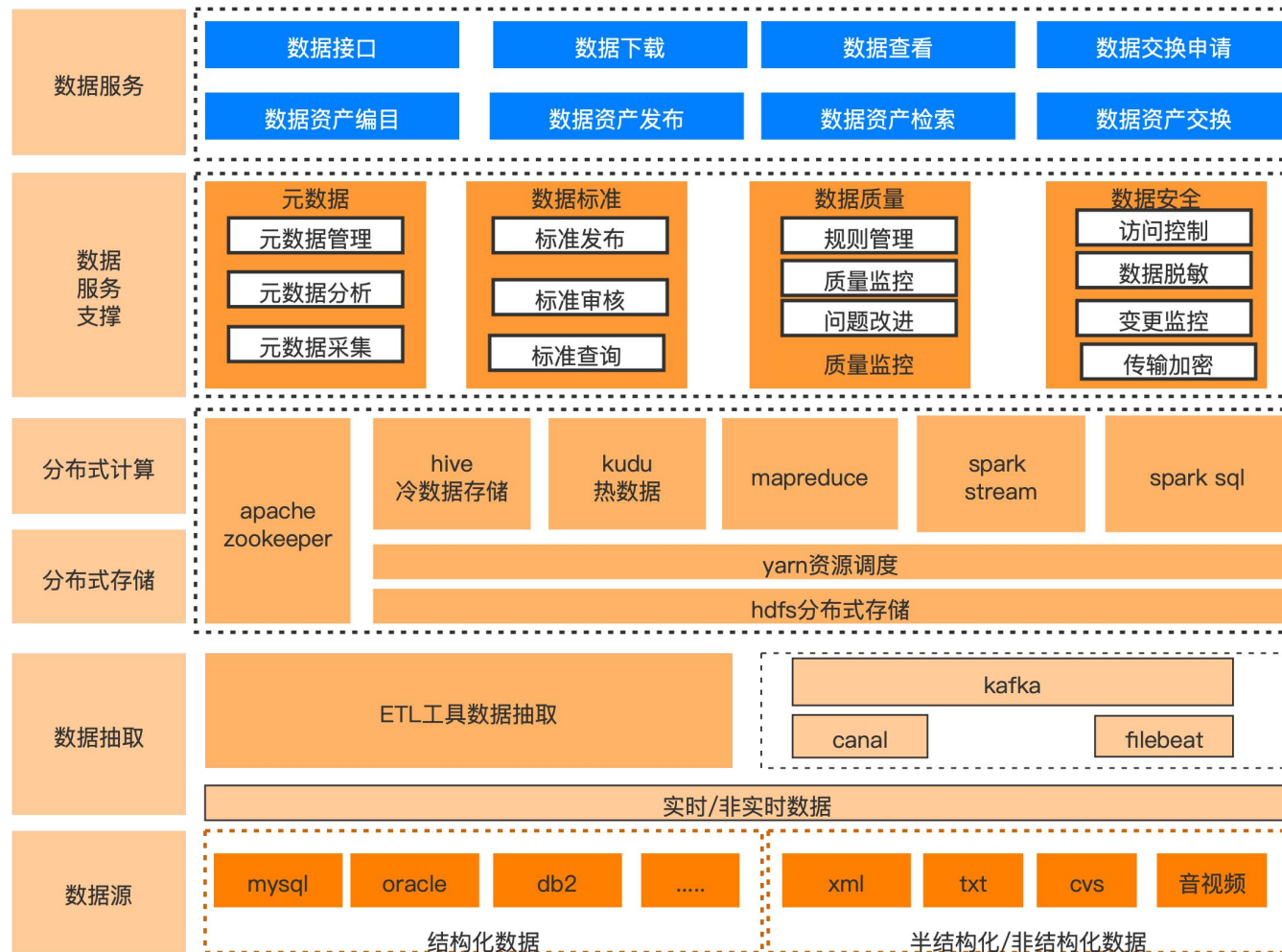
- 纷繁的系统产生杂乱的数据
- 分类管理，有的放矢
- 数据质量是共享和交换的前提
- 共享和交换是数据价值的体现



数据价值化



系统架构图



- 架构是危险的权衡艺术
- 应用伸缩性与应用复杂性的权衡
- 数据分散和数据资产价值化的博弈
- 应用摊薄，数据做厚





Part 03 数据共享

实时接口请求





数据共享的关注点

01 冷热数据存储
冷热数据的存储策略
数据的冷却

02 严格的数据质量
数据分类与预处理
元数据管理
数据质量检查

03 多变的业务诉求
单系统的数据请求
多系统的数据整合

04 数据安全性与监管
数据访问安全
数据存储安全
数据脱敏
有迹可查，有据可依



数据共享四大组件

数据定义

一方面描述企业数据和服务的分布情况，另一方面，作为语义层抽象，用于驱动整个企业数据的加工处理过程。

服务开发

海保批处理平台、实时接口服务、文件传输服务三种数据服务形式，并且提供在线、离线两种服务开发方式。过程中可以根据具体业务场景需求，制定服务发布和安全策略。

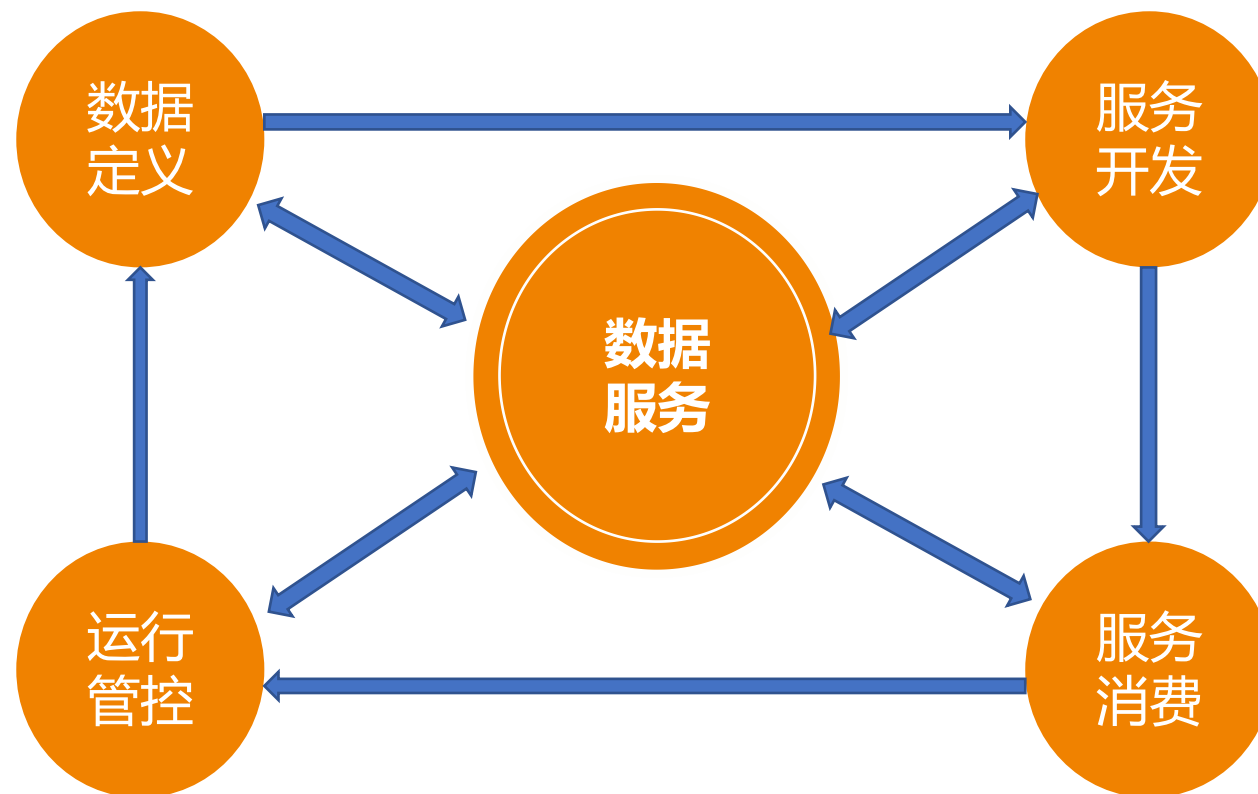
服务消费

平台为消费者提供数据预览、申请、审批等功能，具备单个、批量申请的能力，具备向消费方推送或由消费方拉取的服务方式

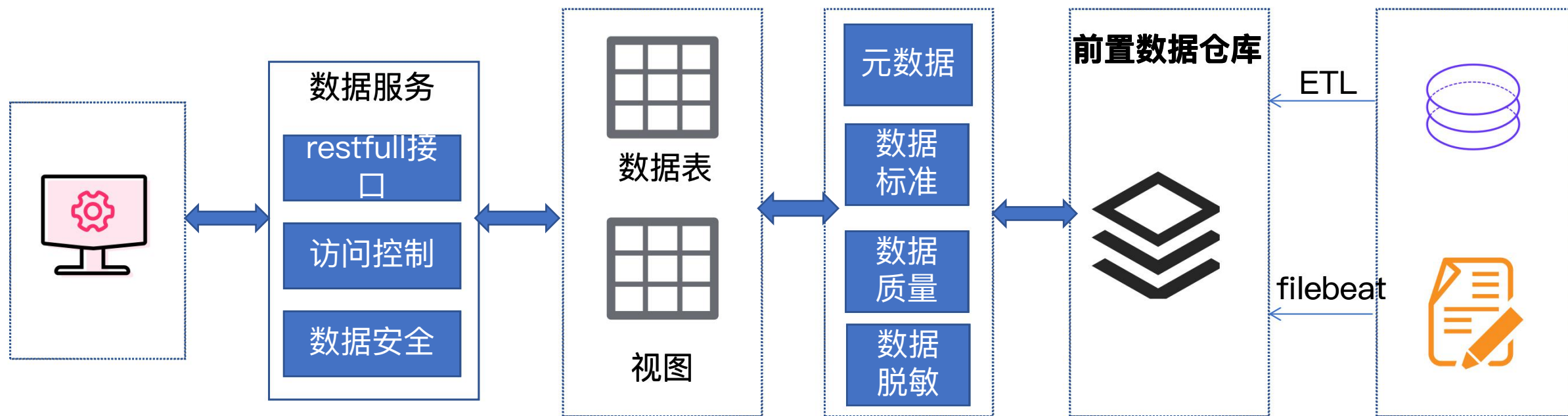
运行管控

实现对数据服务全生命周期管理、监控等；通过权限控制验证是否有权限访问数据；对提取的关键数据进行脱敏处理，保障数据安全性；实现对数据流向、批量任务、接口、文件传输





数据共享的流程



数据共享

* 应用名称:

* 责任人:

* 责任人邮箱:

* 是否加密:

* 共享库:

* 视图权限:

表或视图

表名	类型	备注	数据条数	创建时间
test	BASE TABLE		96	2011-02-25 17:57:59
test	BASE TABLE		113	2011-07-15 09:05:54
test	BASE TABLE		0	2011-12-10 17:41:57
test	VIEW			

元数据

序号	所属表	列名	数据类型	是否允许为空	备注
1	test	id	int(11)	NO	test
2	test		varchar(100)	NC	

序号	应用名称	来源IP	责任人	请求库	请求参数	执行sql	返回数据	创建时间	耗时/ms
1				test		select id from test where id=6		2011-12-10 17:41:57	4 ms





Part 04 数据交换

数据落地与后续逻辑处理



数据交换三大特性

灵活架构

灵活的体系架构，既可支持集中部署，也可支持分布部署，数据支持多样的传输方式。交换源端与目标端灵活配置，源与目标端多种方式可任意组合

01

高性能

02

可扩展集成

数据交换平台可以无缝集成更多的功能平台包括：海保数据平台，海保数据质量中心等，灵活匹配项目实际需求场景。

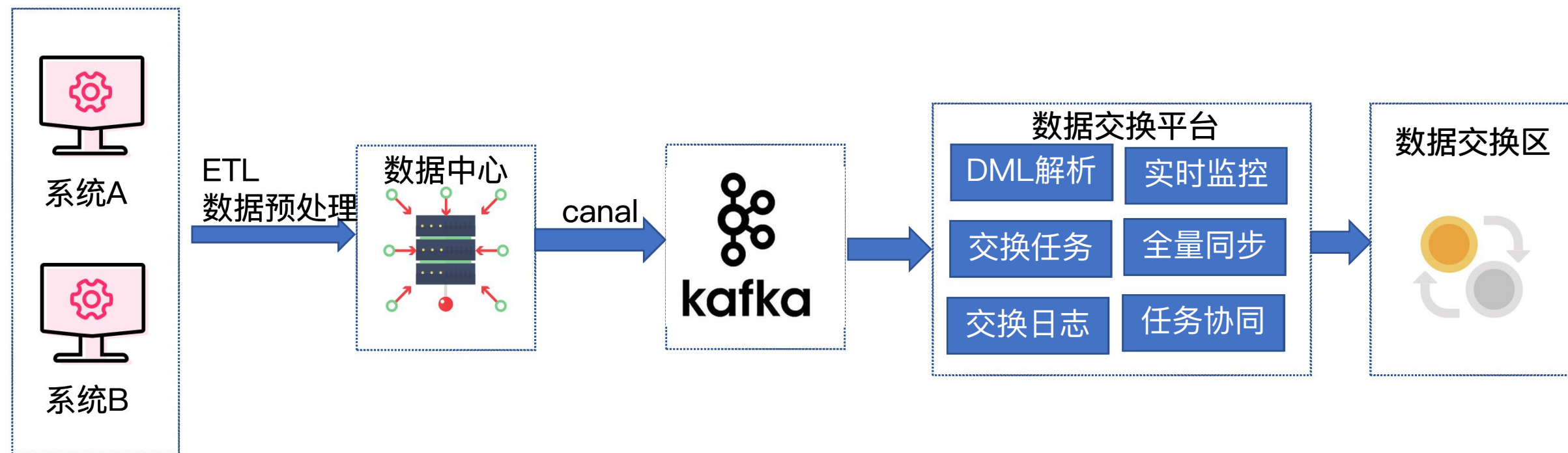
03

监控体系完备

全面的运行监控和系统监控，完整的日志追踪，保障系统稳定

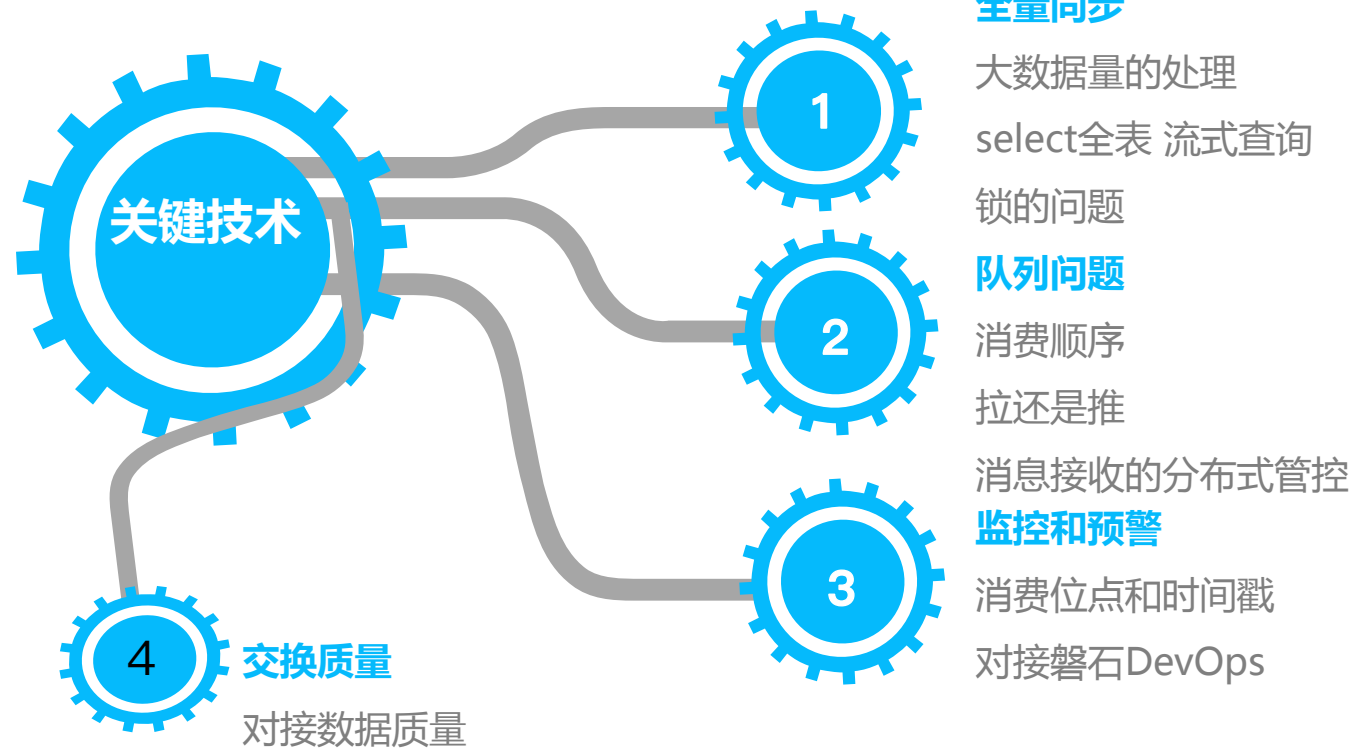


数据交换的流程





关键技术点



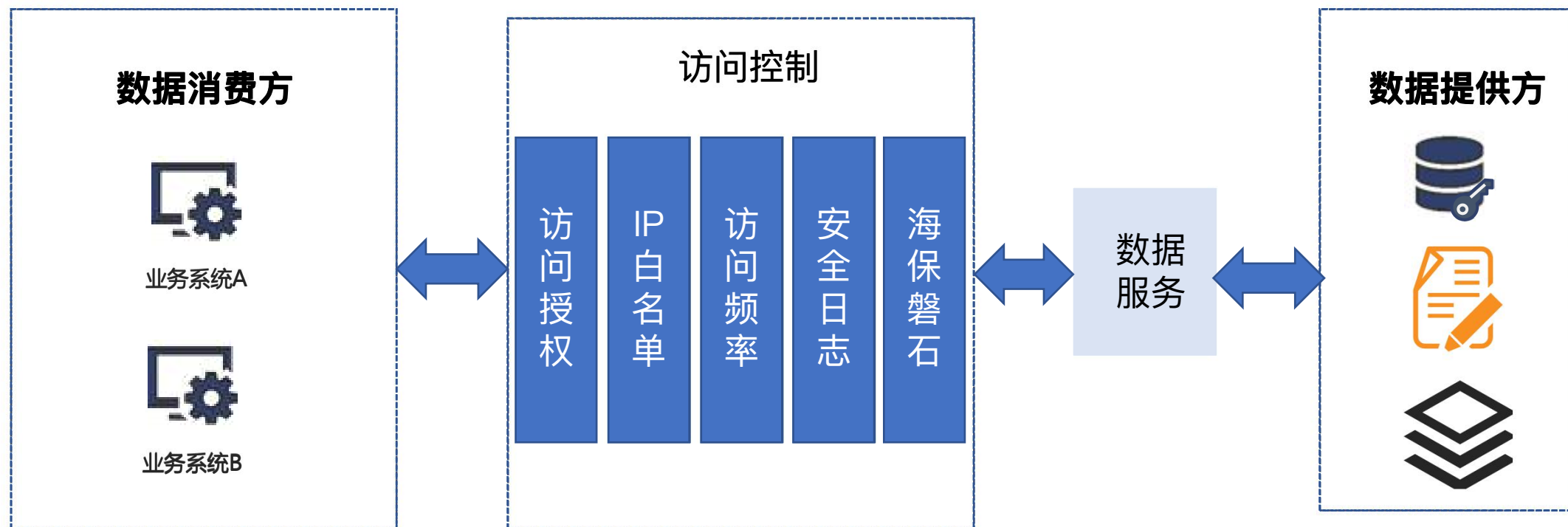
序号	任务名	状态	队列名	最新同步时间	操作
1	数据同步任务	已启动	Kafka-DB2	2023-10-08 10:08:56	<div>开启 <input checked="" type="checkbox"/> 关闭</div> <div>✓ 全量同步 ✓ 历史 ✓ 日志 ✓ 编辑 ✓ 查看</div> <div>删除</div>
2	数据同步任务	待启动	Kafka-DB2	2023-10-08 10:08:56	<div>开启 <input checked="" type="checkbox"/> 关闭</div> <div>✓ 全量同步 ✓ 历史 ✓ 日志 ✓ 编辑 ✓ 查看</div> <div>删除</div>

序号	任务名称	源数据表	目标表	任务类型	交换前	交换后	目标sql	日志文件	binlog位点	交换状态	信息	时间	操作
1	数据同步任务	mriskap	c_data	INSERT			insert in...	binlog_057662	7871931	成功	交换成功	1630	查看
2	数据同步任务	lr_app	c_cole	DELETE			delete fr...	binlog_057662	7871931	成功	交换成功	1630	查看



- 技术手段与管理方法结合
- MQ是数据缓冲的利器
- ZK是分布式协调的不错选择







THANKS