

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





数据中台在光大银行的二次蜕变

演讲人：王磊



个人介绍



王 磊

- 光大银行资深架构师、数据中台团队负责人
- 信通院大数据产品评测专家评委
- 《分布式数据库30讲》专栏作家
- 金融数士公众号作者

目前主要负责光大银行数据中台、AI平台等系统建设和数据技术产品研发工作



数 / 造 / 未 / 来





数据中台的过度炒作

- 从数据分析师到大数据分析师
- 技术词汇的演变逻辑
- 包罗万象的“大词”是对专业性的伤害
- 探究数据中台的差异性

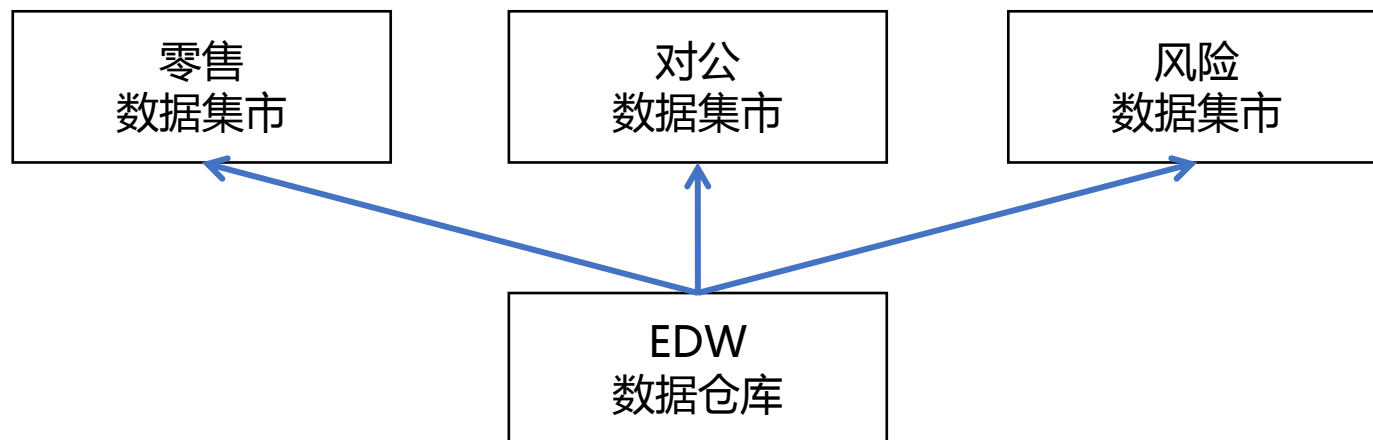


数据仓库的初心

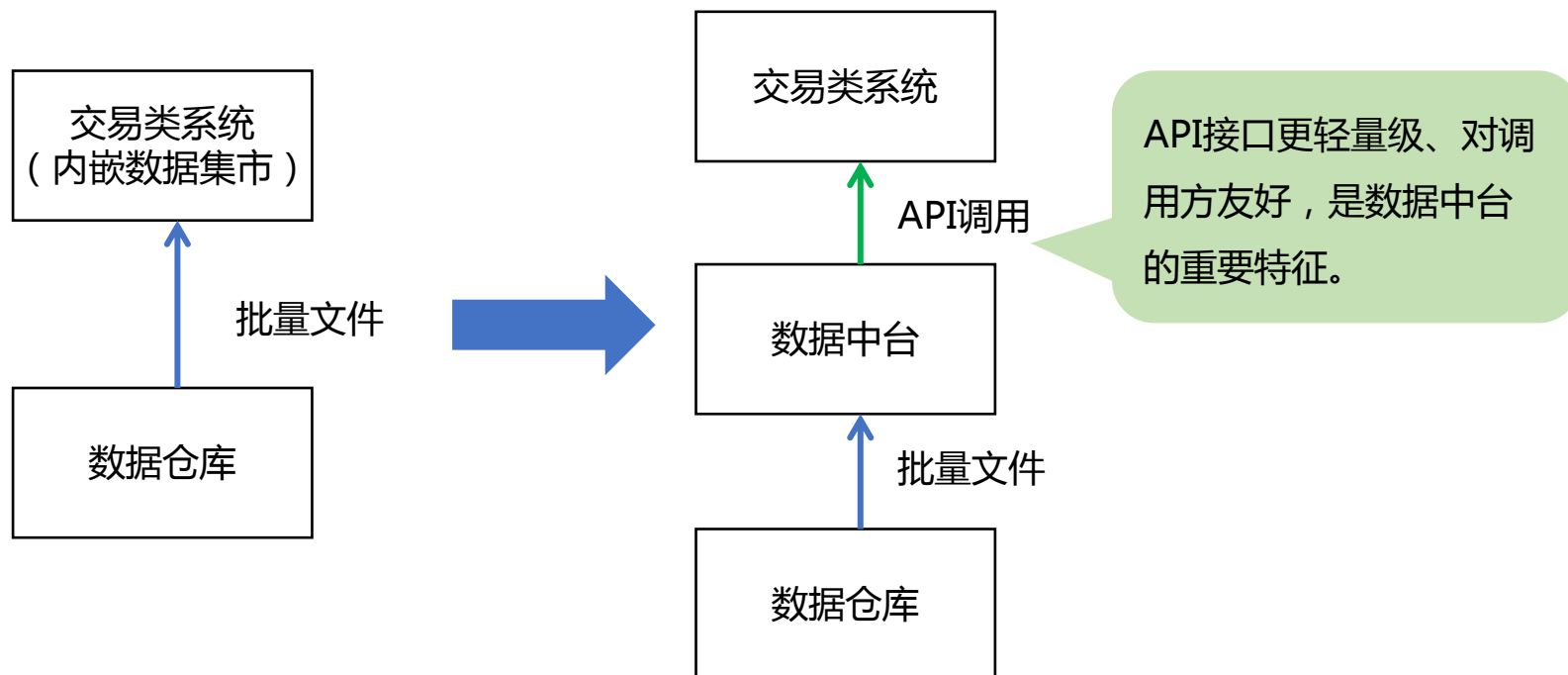
企业级能力复用是业界对数据中台的最大共识，但不是首次尝试。

数据仓库作为企业级数据平台，面向主题，提供复用能力；数据集市作为业务场景提供解决方面，面向应用。

在过去的近20年中，两者构成的数据仓库体系长期稳定运行，其中数据仓库的复用能力依赖于先验模型。



数据中台的突破





数据中台与数据仓库实现复用能力的逻辑不同

数据仓库：数据复用

方式：基于主题模型向上衍生

“复用”始终是数据平台追求的目标，传统数据仓库从源系统出发以企业级主题模型为核心，但距离应用场景较远，构建复用数据资产的难度更大。

主题模型的生命力？

数据中台：服务复用+数据复用

方式：双管齐下

1. API服务是从应用侧入手，在真实业务场景下，持续提炼原子服务，通过原子服务的归集，建立面向应用的数据域。
2. 共性数据平台延续数据仓库的思路，面向业务条线进行数据的逐步聚合。根据双方数据资产积累情况，分主题逐步与API数据域对接。



为什么要拆中台？



在讨论中台时，常采用“变速齿轮”理论。这意味着，一个成功的中台系统其变更频率一定是低于前台的。

怎么做到这一点呢？通常是因为中台预设好了使用模式，前台系统遵从，获得了复用的效果。如果前台系统要打破预设模式，就无法借助中台的能力。

所以，中台的预设模式会约束业务创新。

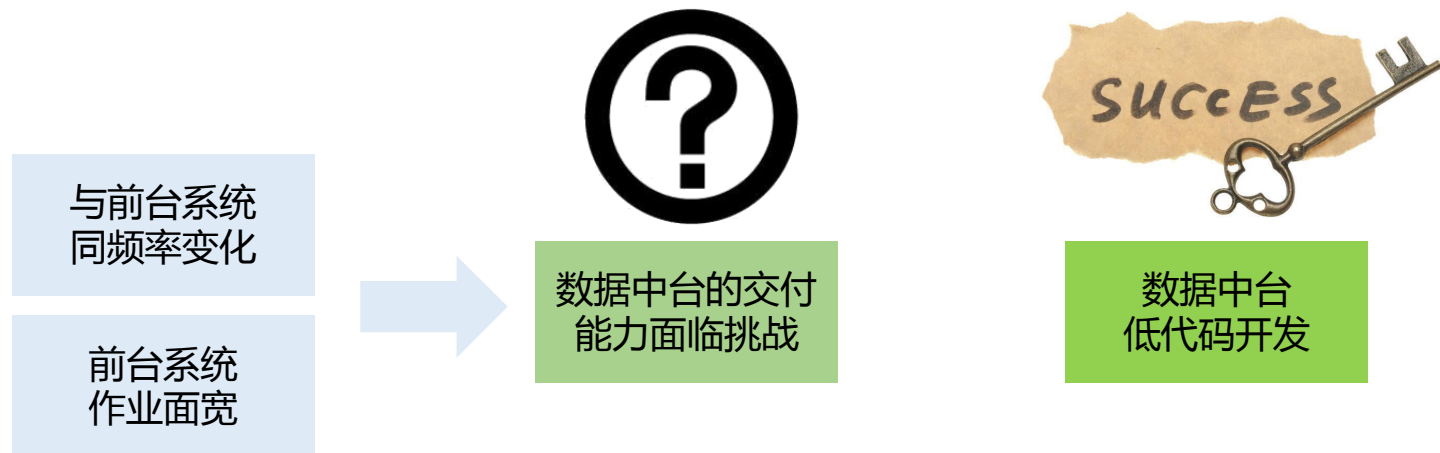




敏捷的数据中台

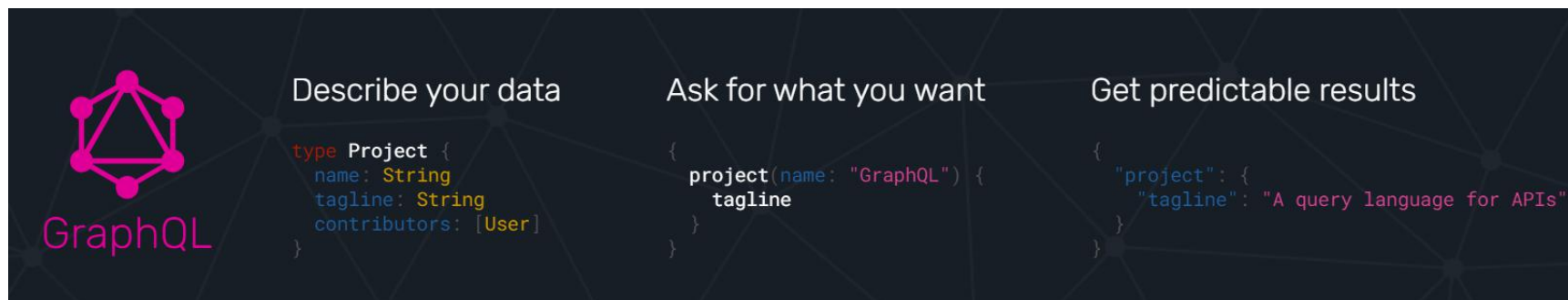
要做到灵活支持前台系统，中台就不能停留在固定模式下，应该与前台同频率变化。而前台系统覆盖企业的全部业务线，作业面非常宽，数据中台的交付能力面临挑战。

所以，数据中台必须是敏捷的



敏捷的数据中台——技术实现

低代码开发的基础是微服务和服务编排，服务编排的复杂性并不完全匹配数据中台的特点。
这使我们有机会选择更简单的技术“数据融合” GraphQL



GraphQL

```
{
  hero {
    name
    friends {
      name
      homeWorld {
        name
        climate
      }
      species {
        name
        lifespan
        origin {
          name
        }
      }
    }
  }
}
```

```
type Query {
  hero: Character
}

type Character {
  name: String
  friends: [Character]
  homeWorld: Planet
  species: Species
}

type Planet {
  name: String
  climate: String
}

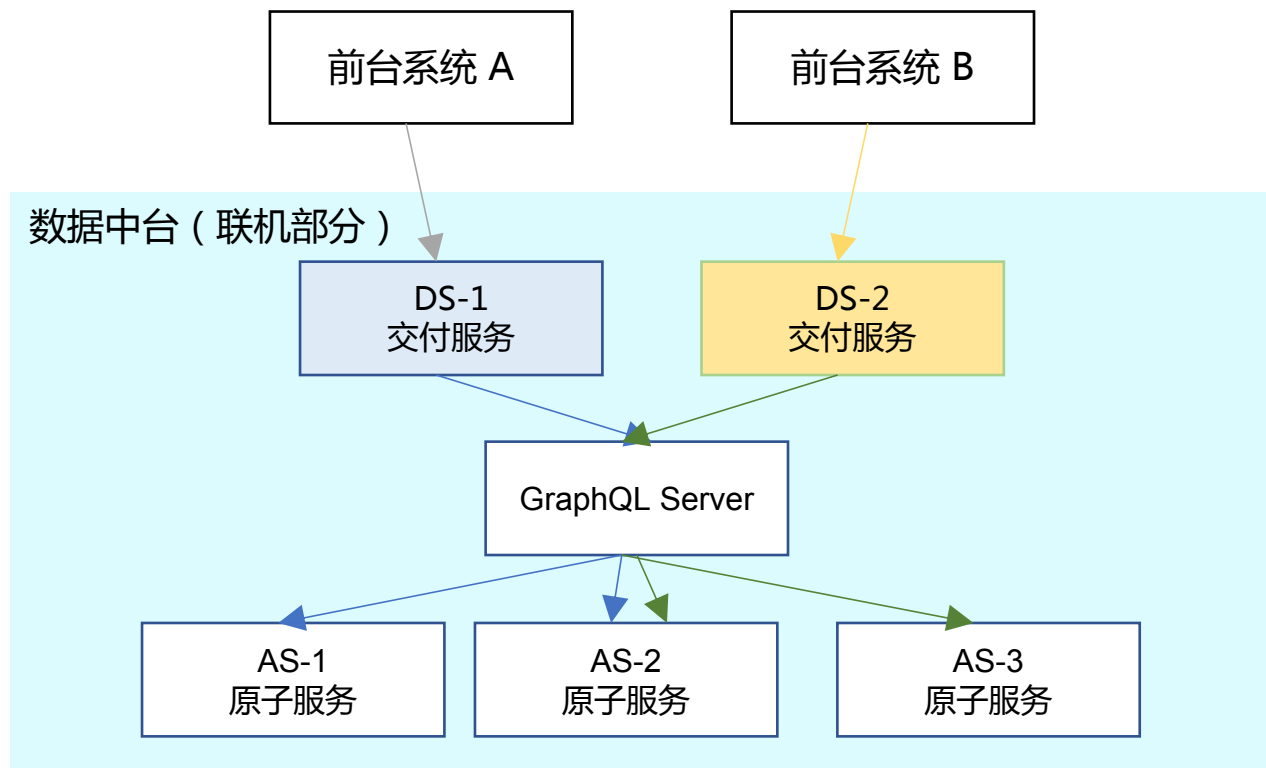
type Species {
  name: String
  lifespan: Int
  origin: Planet
}
```

1. 客户端定义接口
2. 兼容Json报文
3. 整合后台服务
4. 多语言实现
5. 支持异步非阻塞调用
6. 批量调用解决N+1问题





数据中台架构示例 (GraphQL)



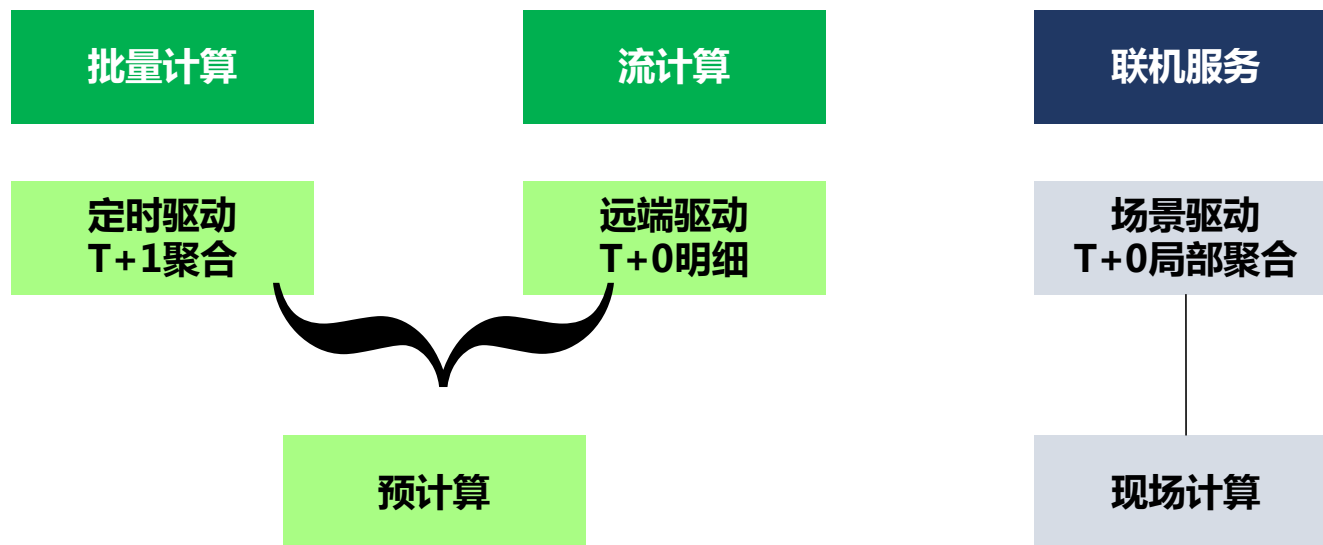
实时应用数据的诉求

作为大数据时代的重要标志，数据使用者不仅是企业中高层管理者，覆盖了更多的一线业务用户，这也是数据中台的主要驱动力之一。从一线业务用户实际操作出发，必然会提出更高的数据时效性要求。

所以，实时数据是数据中台必须响应的诉求。



实时的数据中台



实时的数据中台——技术实现

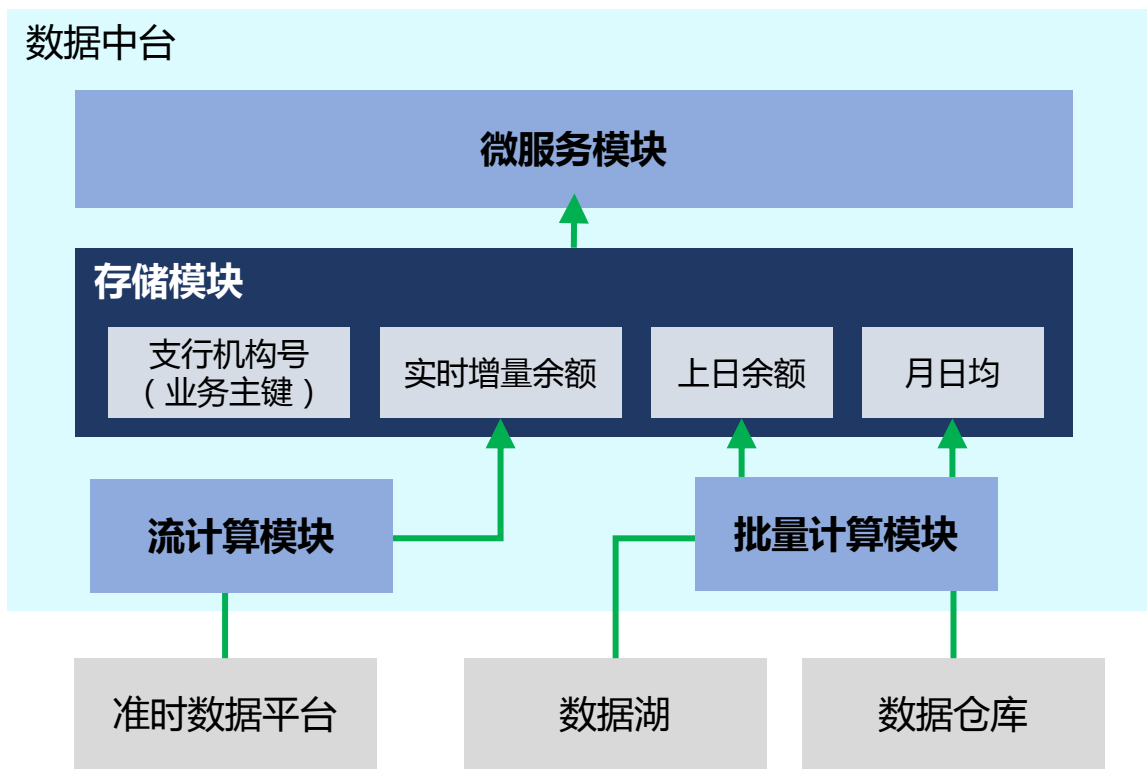
实时分布式计算框架的加持下，今天计算机系统的现场计算能力远强于20年前，这使我们有
可能将一部分ETL工作重新放回到实时计算框架中执行。

对ETL的裁剪也让数据能够按照更加通用样式存储，避免与业务场景的紧耦合，提升了中台的
复用能力。

ClickHouse和分布式数据库（HTAP）的发展都为这种技术的成熟性提供了背书。



现场计算理念的落地



业务场景：

面向零售客户经理，提供总、分、支等机构维度的零售客户余额查询。

设计特点：

数据时效达到秒级延迟，保证24小时不间断的数据供给，对上游批量延迟也有一定的容错能力。



We are hiring

数据中台技术经理& 基础软件研发

Java技术栈为主，有微服务经验尤佳

熟悉Hadoop大数据生态技术

联系邮箱：zh-wanglei@cebbank.com



金融数士



数，造，未，来





THANKS