

阿里巴巴大数据实践 - 数据服务

原创

阿里云技术 于 2020-08-28 14:35:58 发布 404 收藏

文章标签: 阿里云

版权声明: 本文为博主原创文章, 遵循 [CC 4.0 BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) 版权协议, 转载请附上原文出处链接和本声明。

本文链接: https://blog.csdn.net/weixin_43970890/article/details/108279210

版权

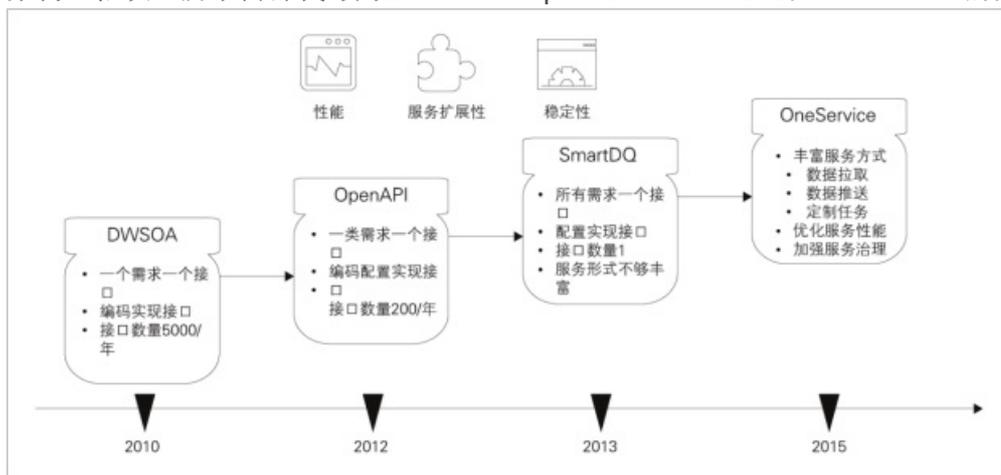
数据部门产出的海量数据, 如何能方便高效地开放出去, 是我们一直想要解决的难题。在没有数据服务的年代, 数据开放的方式简单、粗暴, 一般是直接将数据导出给对方。这种方式不仅低效, 还带来了安全隐患等诸多问题。

为此, 我们在数据服务这个方向上不断探索和实践。最早的数据服务雏形诞生于2010年, 至今已有7个年头。在这期间, 随着我们对业务的理解不断加深, 同时也得益于新技术的持续涌现, 对数据服务架构也进行了多次升级改造。服务架构的每次升级, 均在性能、稳定性、扩展性等方面有所提升, 从而能更好地服务于用户。

来源: 数智化转型俱乐部

1. 服务架构的演进

阿里数据服务架构演进过程如图6.1所示。基于性能、扩展性和稳定性等方面的要求, 我们不断升级数据服务的架构, 依次经历了内部代号为DWSOA、OpenAPI、SmartDQ和OneService的四个阶段。



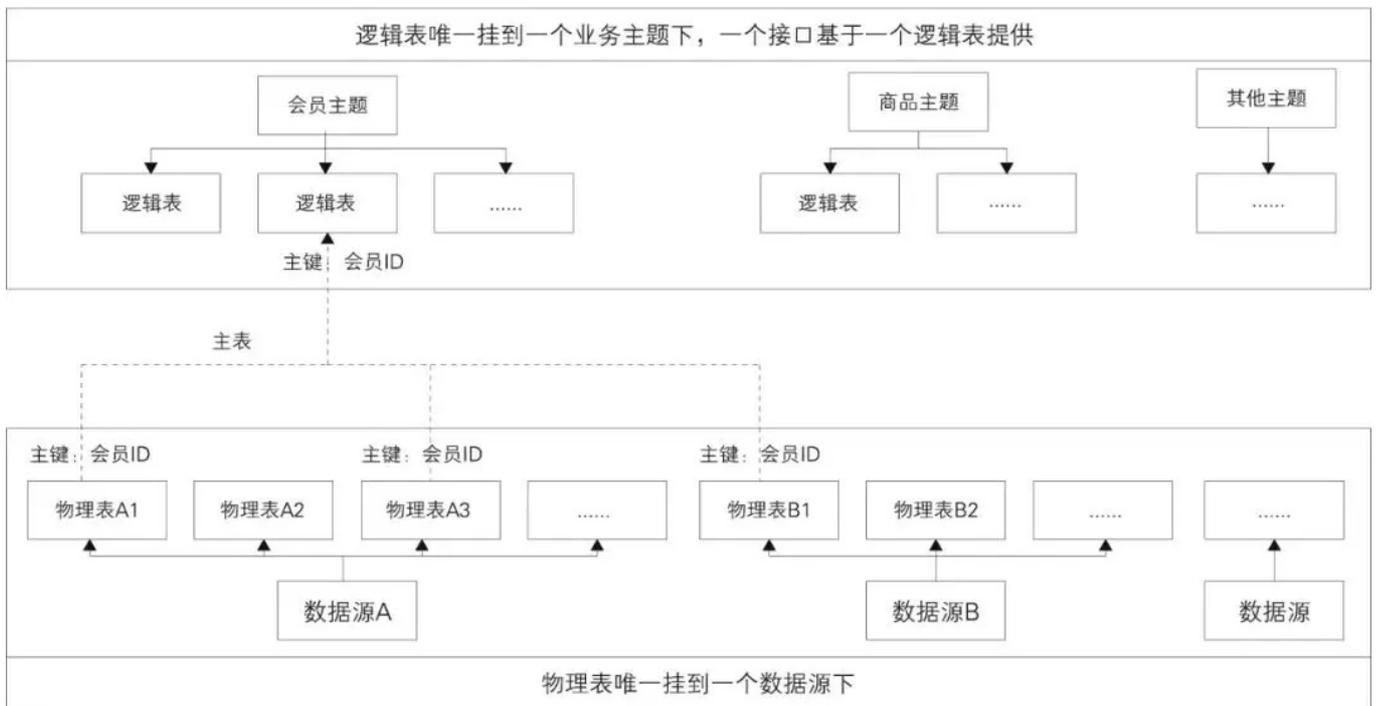
阿里数据服务架构演进过程

其中, 第四个阶段是统一的数据服务层(即OneService)。大家心里可能会有疑问: SQL并不能解决复杂的业务逻辑啊。确实, SmartDQ其实只满足了简单的查询服务需求。我们遇到的场景还有这么几类: 个性化的垂直业务场景、实时数据推送服务、定时任务服务。所以OneService主要是提供多种服务类型来满足用户需求, 分别是OneService-SmartDQ、OneService-Lego、OneService-iPush、OneService-uTiming。

在OneService阶段, 开始真正走向平台化。我们提供数据服务的核心引擎、开发配置平台以及门户网站。数据生产者将数据入库之后, 服务提供者可以根据标准规范快速创建服务、发布服务、监控服务、下线服务, 服务调用者可以在门户网站中快速检索服务, 申请权限和调用服务。

2. 技术架构

- SmartDQ

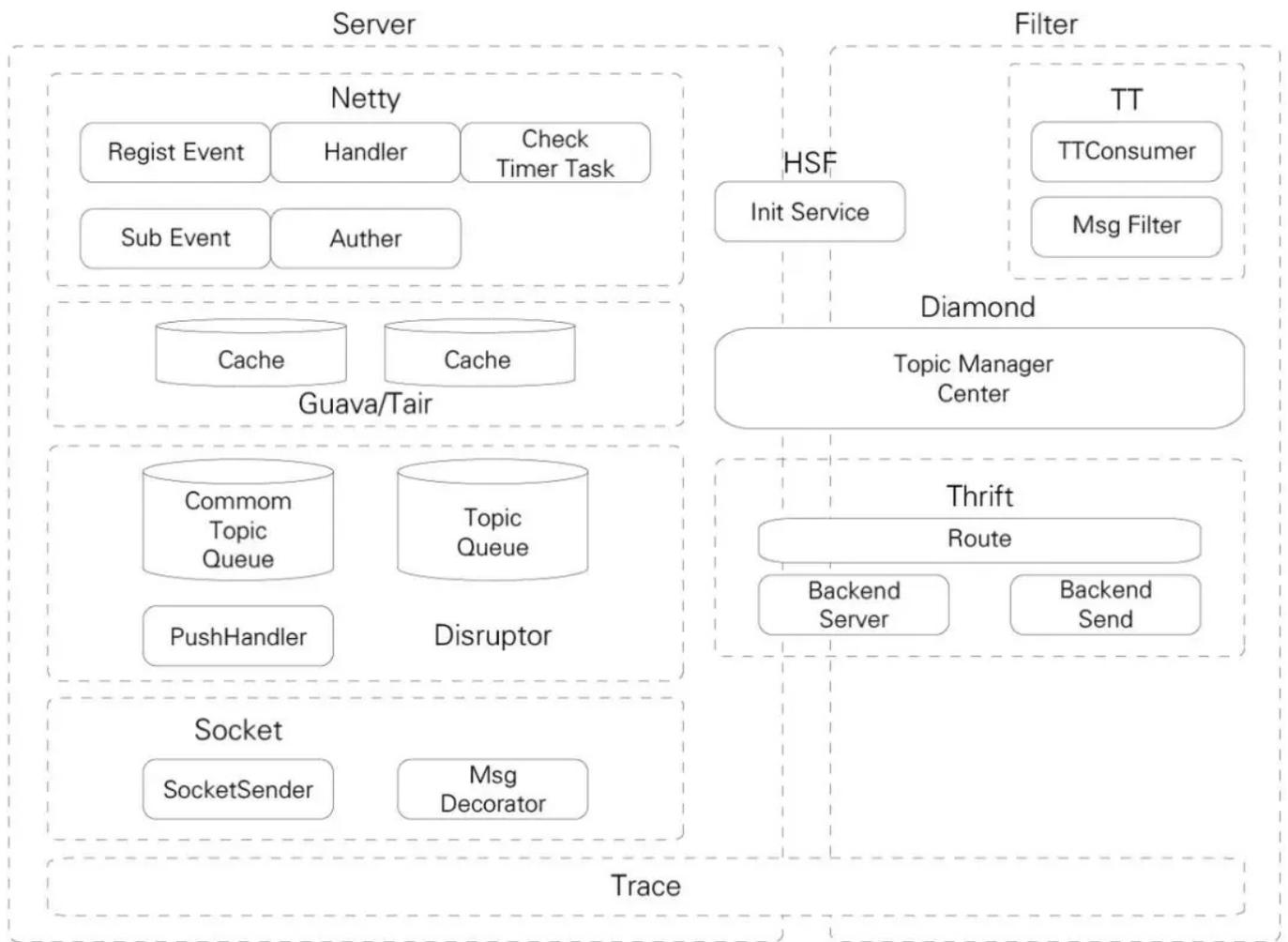


SmartDQ的元数据模型架构示意图

SmartDQ的元数据模型，简单来说，就是逻辑表到物理表的映射。自底向上分别是：

- (1) 数据源：SmartDQ支持跨数据源查询，底层支持接入多种数据源，比如MySQL、HBase、OpenSearch等。
- (2) 物理表：物理表是具体某个数据源中的一张表。每张物理表都需要指明主键由哪些列组成，主键确定后即可得知该表的统计粒度。
- (3) 逻辑表：逻辑表可以理解为数据库中的视图，是一张虚拟表，也可以看作是由若干主键相同的物理表构成的大宽表。SmartDQ对用户展现的只是逻辑表，从而屏蔽了底层物理表的存储细节。
- (4) 主题：逻辑表一般会挂载在某个主题下，以便进行管理与查找。

- iPush



iPush应用架构示意图

iPush应用产品是一个面向TT、MetaQ等不同消息源，通过定制过滤规则，向Web、无线等终端推送消息的中间件平台。iPush核心服务器端基于高性能异步事件驱动模型的网络通信框架Netty 4实现，结合使用Guava缓存实现本地注册信息的存储，Filter与Server之间的通信采用Thrift异步调用高效服务实现，消息基于Disruptor高性能的异步处理框架（可以认为是最快的消息框架）的消息队列，在服务器运行中Zookeeper实时监控服务器状态，以及通过Diamond作为统一的控制触发中心。

- Lego

Lego被设计成一个面向中度和高度定制化数据查询需求、支持插件机制的服务容器。它本身只提供日志、服务注册、Diamond配置监听、鉴权、数据源管理等一系列基础设施，具体的数据服务则由服务插件提供。基于Lego的插件框架可以快速实现个性化需求并发布上线。

Lego采用轻量级的Node.JS技术栈实现，适合处理高并发、低延迟的IO密集型场景，目前主要支撑用户识别发码、用户识别、用户画像、人群透视和人群圈选等在线服务。底层根据需求特点分别选用Tair、HBase、ADS存储数据。

- uTiming

uTiming是基于在云端的任务调度应用，提供批量数据处理服务。uTiming-scheduler负责调度执行SQL或特定配置的离线任务，但并不直接对用户暴露任务调度接口。用户使用数据超市工具或Lego API建立任务。注：本书中出现的部分专有名词、专业术语、产品名称、软件项目名称、工具名称等，是淘宝（中国）软件有限公司内部项目的惯用词语，如与第三方名称雷同，实属巧合。

节选自《大数据之路：阿里巴巴大数据实践》已受版权保护，未经授权不得转载

数据中台是企业数智化的新基建，阿里巴巴认为数据中台是集方法论、工具、组织于一体的，“快”、“准”、“全”、“统”、“通”的智能大数据体系。目前正通过阿里云数据中台解决方案对外输出，包括[零售](#)、[金融](#)、[互联网](#)、[政务](#)等领域。

[原文链接](#)

本文为阿里云原创内容，未经允许不得转载。