

阿里云开发者社区
ALIBABA CLOUD DEVELOPER COMMUNITY

阿里云数据中台解决方案 系列白皮书

金融行业新一代 数据仓库解决方案

2020云栖大会 特别版

数智俱乐部



钉钉扫描二维码 加入

阿里云数据中台 出品

see more please visit: <https://homeofbook.com>



钉钉扫码加入
阿里云数据中台交流群



阿里云开发者“藏经阁”
海量免费电子书下载

| 目录

| | |
|---------------------|----|
| 1.行业挑战 | 4 |
| 2.解决方案 | 5 |
| 2.1 新一代数据仓库与传统数仓的区别 | 5 |
| 2.2 新一代数据仓库的逻辑架构 | 5 |
| 2.3 新一代数据仓库的数据架构 | 7 |
| 3.产品优势 | 8 |
| 4.演变路径 | 10 |
| 法律声明 | 11 |

1.行业挑战

数据仓库的建设及应用在金融机构中发挥着重要作用,其整合了企业内部绝大多数结构化数据,在此基础上按照主题域进行模型设计,通过数据清洗、加工、计算为业务分析提供数据,支撑监管报送、风险管理、业务决策等多种类型的数据分析应用。当前,金融机构数据仓库平台多采用国外厂商的产品,这些产品目前正在面临一些困难和挑战,主要为以下几点:

1) 数据的使用

几十年前建立的金融行业数据标准模型,无法适应快速变化的业务需求,传统的方法论不适应新的业务趋势,缺乏新的方法论支撑;传统架构不能有效服务业务系统,数据服务化能力(API)缺失,与业务系统不能形成良好互动。本质上是业务与数据脱节造成上述局面,需要基于业务视角提供“好数据”,能够很方便地找数据、放心用数据的能力。

2) 数据的管理

技术人员投入大量精力维护着几千个作业,上万张表,十多万个指标,但不知道它们是否能用,在用,谁用;大量的文本文件和 T+1 批处理作业,不能实时的捕获所有变化;大量割裂的工具集,如 ETL 工具,建模工具,数据库, BI 工具,元数据管理工具等各成体系,不能有机整合,开发体验差。

3) 昂贵的成本

传统数仓产品采用一体机方式售卖(集成了服务器、存储、操作系统和数据库)软硬件强绑定的封闭式架构使得建设成本、扩容成本、运维成本都非常高,性价比低;随着数据量的增长,只能纵向扩容无法水平扩展,导致扩展能力有上限,扩容需要中断业务并且费用昂贵。

4) 国产化的趋势

随着国产化政策的持续深入, OLAP 市场同样需要国产化平台替换原有国外产品。监管机构已明确要求在 2022-2023 年之前,完成国外厂商数据仓库平台的升级替换。

2. 解决方案

2.1 新一代数据仓库与传统数仓的区别

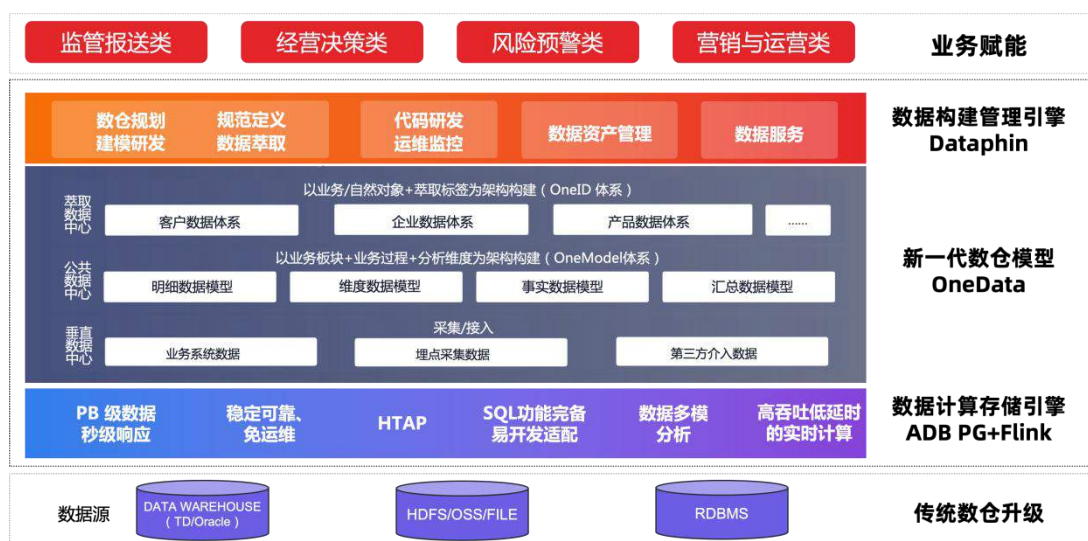
新一代数仓与传统数仓相比，在数据处理时效性、一体化数据研发效率、企业级数据架构体系以及支撑的业务模式等方面的能力都有显著的改变和提升。

- 1) 在数据时效性与处理场景方面，传统数仓以 T+1 的批量数据处理为主，而新一代数仓需要兼具离线、准实时、实时，不同场景的数据采集与处理能力，同时能够提供在线查询的交互能力。
- 2) 在数据研发效率方面，传统数仓在模型设计，脚本开发、统一调度、数据治理方面也采用了相应的工具，但工具之间彼此割裂，解决了单独的技术问题，却无法从整体的角度联动起来，进而提升整体研发效率。新一代数仓具备一站式的数据研发工具体系，实现从数据采集、逻辑模型设计、物理模型落地、脚本开发，任务调度到数据资产管理、数据服务的全链路数据研发体系，能够从根本上提升数据研发和数据管理的整体效率。

在数据架构方面，传统数仓更多聚焦在数仓内部的数据架构设计，围绕减少数据冗余的目标，进行数据模型设计与落地，而新一代数仓将从企业级数据架构视角，围绕业务便捷使用数据的目标进行数据架构的设计。

2.2 新一代数据仓库的逻辑架构

基于阿里云 MPP 分布式架构数据库引擎和新一代维度建模方法论体系以及智能数据构建管理引擎的数据仓库解决方案可以帮助金融机构从容应对上述挑战，新一代金融数据仓库解决方案架构如下：



新一代数据仓库从逻辑上主要划分为三部分，分别是批量计算与在线分析，实时计算以及数据构建与管理。

批量计算与在线分析引擎，采用 MPP 分析型数据库 ADB for PG，支持 PB 级数据计算与存储，兼容 PostgreSQL、GreenPlum、且高度兼容 Oracle 语法生态。具有存储计算分离，在线弹性平滑扩容的特点。支持任意维度的在线分析探索和高性能离线数据处理。

实时计算引擎，采用阿里云企业版 Flink，提供多种编程语言，对流式语义完全支持，并提供高吞吐、低延时、可扩展架构，轻松支持超大规模 QPS。企业版 Flink 与开源社区版内核完全兼容，同时补充了社区版没有的开发、测试、运维等方面相应的工具。

智能数据仓库构建与管理引擎采用 Dataphin 平台，可同时对接 ADB for PG 和 Flink 实现离线计算与实时计算在数据研发、数据管理层面的统一构建。从功能层面，Dataphin 提供了数仓规划，数据规范定义、数据建模、数据研发、标签萃取、调度运维、数据资产管理、数据服务等全链路的体系化工具。并且可对接关系型数据库，NoSQL 数据库数据文件、日志等多种数据源，实现批量采集、准实时采集和实时采集的能力。

2.3 新一代数据仓库的数据架构

企业中的数据经过采集后，按照离线、准实时、实时通道进入新一代数据仓库，在数据仓库内。逻辑上依然采用分层架构体系，自下而上依次为垂直数据中心、公共数据中心和萃取数据中心。

垂直数据中心在经过清洗加工后保留贴源数据。公共数据中心按照维度建模，构建事实数据模型，维度数据模型和汇总数据模型。同时可根据客户的现状及需求，在保留现有数仓三范式模型的基础之上，再按照维度建模，构建明细数据和汇总数据。可实现基于现有三范式模型支持的分析应用的平滑迁移。在公共数据中心之上，构建萃取数据中心，实现数据的标签化、资产化。基于整合汇总后的业务数据，加工成有业务价值的数据资产，即分析指标和数据标签。形成企业级、高质量、计算逻辑可配置的指标中心和标签中心。基于分析指标支撑监管报送、统计分析等类型的数据应用，基于标签中心支撑客户画像，精准营销与客户运营等类型的数据应用，并可进一步将数据标签、指标进行封装，提供主题式的数据查询、API调用以及相应的数据产品，体现了新一代数仓数据资产服务化的能力。

3. 产品优势

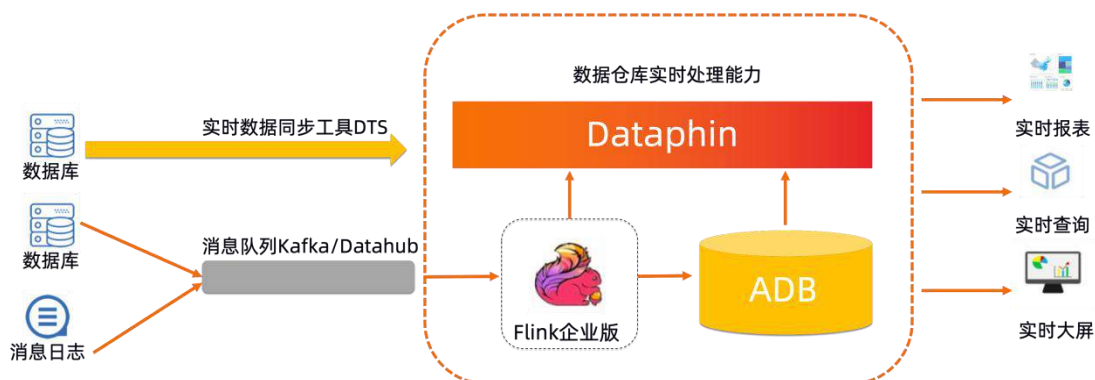


核心优势 1：同时支持批量计算与交互分析的 MPP 高性能分析型数据库

阿里云分析型数据库 AnalyticDB PostgreSQL 提供 PB 级 MPP 全并行架构数据仓库解决方案。兼容 Greenplum、广泛兼容 PostgreSQL/Oracle 语法生态，新一代向量引擎性能超越传统数据库引擎 3 倍以上，可以实现对海量数据的即席查询分析、ETL 处理及可视化探索。同时适配主流的 x86 服务器、国产化 ARM 服务器等国产化方案。

核心优势 2：基于 Flink 的实时数据处理能力

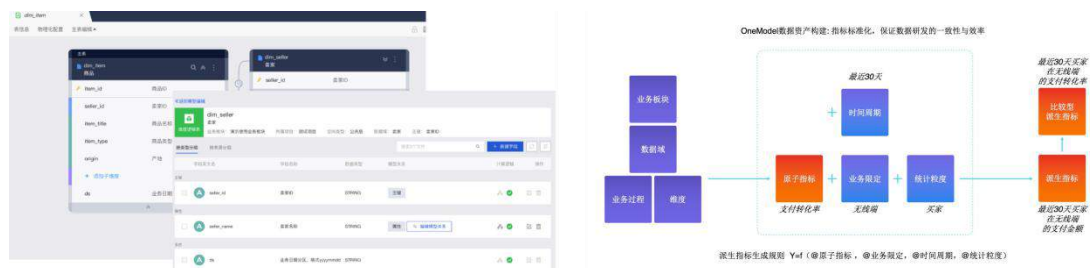
相比于传统数仓以离线处理任务为主的现状，阿里云新一代数据仓库解决方案增强了实时数据处理能力，可同时支持批量离线任务、实时数据计算、在线交互查询等多种数据处理模式，扩大了数据仓库在金融机构的应用场景。实时数据处理能力可从实时数据采集、实时数据计算、在线交互查询几个方面体现。



核心优势 3：可视化数据建模+规范化分析指标构建

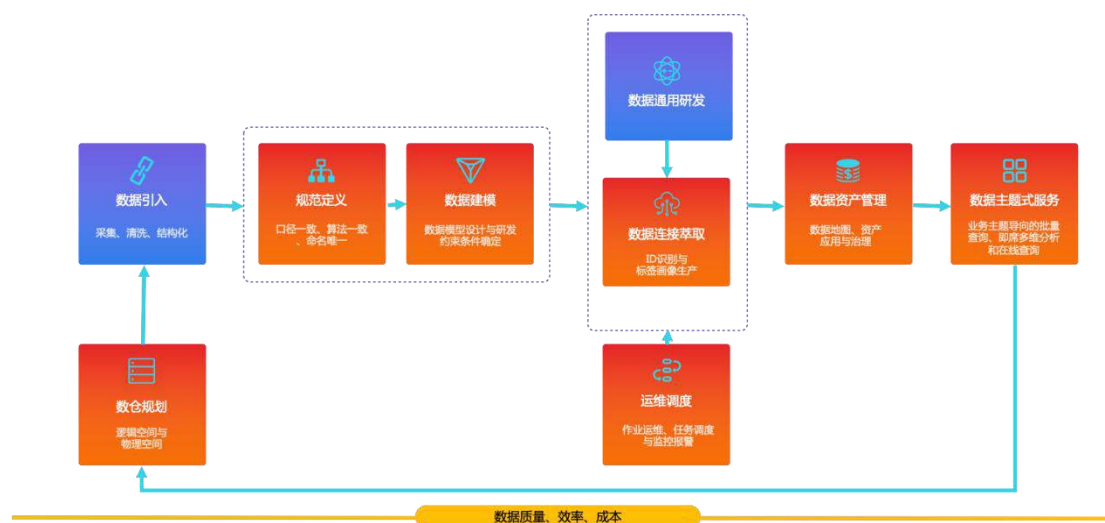
阿里云智能数据构建及管理引擎 Dataphin 提供了可视化逻辑建模工具，支持数据模型的快速构建，并支持从逻辑数据模型到物理数据模型的自动映射，用户只需根据业务进行逻辑模型可视化设计，后台会自动生成相应的物理表，实现了数据模型设计即开发的能力。

在分析指标方面，阿里云智能数据仓库解决方案提供了标准化、规范化的指标设计体系，包括原子指标、衍生指标、业务限定、时间维度等元素，并通过工具实现了从指标设计到计算任务的自动化关联，在提升指标计算效率的同时，保障了指标业务含义的准确性与一致性，提升了指标的数据质量。



核心优势 4：一站式全链路数据研发体系

支撑稳态+敏态数据应用的数据仓库，在其自身的数据加工、供应流程中，各个环节形成一个有机的整体，共同支撑敏捷的数据研发能力。在新一代数据仓库解决方案中，基于智能数据构建与管理平台 Dataphin，提供了数据引入、规范定义、数据建模、数据研发、运维调度、数据资产管理、数据主题式服务等全流程的能力支撑，构成完整的数据采集、研发、管理、服务全链路。



4. 演变路径

金融行业用户在数据领域已积累了丰富的建设成果，数据中台建设并非从零开始，而是基于企业现有数据架构体系，可先进行局部的升级与改造，再逐步迭代实现整体的能力组合升级。MPP+大数据平台是金融行业用户普遍采用的数据平台架构，阿里云数据中台解决方案可基于此架构，进行数据中台能力的建设，并为用户量身定制数据中台的发展演进路径。

从数据仓库项目实施层面，项目的实施内容可分为两个部分，一方面是现有数据仓库平台的迁移。阿里云能够提供从现有平台迁移至 ADB 的脚本转换工具，目前已支持的包括 Teradata、Xdata，GreenPlum，能够有效提升平台迁移的效率。在原有平台迁移完成的基础之上，进一步进行数仓能力的提升，包括实时数据采集与计算，可视化数据建模，规范化指标构建，数据标签、数据资产管理、数据服务等方面的能力。这些能力可根据用户的现状及需求的迫切程度，进行分阶段建设。

与此同时，借鉴数据仓库能力升级的建设，需要从企业全局的视角，对相关数据平台及整体数据架构进行统筹规划，借鉴阿里自身的实践，对数据仓库未来向数据中台演进，设计发展路径。新一代数据仓库可作为数据中台建设的第一步，逐步完成数据资产化、资产服务化、服务价值化的数据能力体系建设。与数据仓库建设相对应的是数据团队的发展与壮大，从需求响应型的技术服务组织，向服务驱动型的数据运营组织发展，培养更多的复合型人才，为金融行业数字化转型提供充足的动力。



法律声明

阿里云提醒您在阅读或使用本文档之前仔细阅读、充分理解本法律声明各条款的内容。如果您阅读或使用本文档的，您的阅读或使用行为将被视为对本声明全部内容的认可。

- 1) 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档，且仅能用于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息，您应当严格遵守保密义务；未经阿里云事先书面同意，您不得向任何第三方披露本文档内容或提供给任何第三方使用。
- 2) 未经阿里云事先书面许可，任何单位、公司或个人不得擅自摘抄、翻译、复制本文档内容的部分或全部，不得以任何方式或途径进行传播和宣传。
- 3) 由于产品版本升级、调整或其他原因，本文档内容有可能变更。阿里云保留在没有任何通知或者提示下对本文档的内容进行修改的权利，并在阿里云授权通道中不时发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠道下载、获取最新版的用户文档。
- 4) 本文档仅作为用户使用阿里云产品及服务的参考性指引，阿里云以大数据集成服务平台的“现状”、“有缺陷”和“当前功能”的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的操作指引，但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的，阿里云不承担任何法律责任。在任何情况下，阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害，包括用户使用或信赖本文档而遭受的利润损失，承担责任（即使阿里云已被告知该等损失的可能性）。
- 5) 阿里云网站上所有内容，包括但不限于著作、产品、图片、档案、资讯、资料、网站架构、网站画面的安排、网页设计，均由阿里云和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意，任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外，未经阿里云事先书面同意，任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称（包括但不限于单独为或以组合形式包含“阿里云”、“Aliyun”、“AliCloud”、“万网”等阿里云和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司）。
- 6) 如若发现本文档存在任何错误，请与阿里云取得直接联系。



钉钉扫码加入
阿里云数据中台交流群



阿里云开发者“藏经阁”
海量免费电子书下载