

标准研究

2023 年第 7 期（总第 29 期）

证标委秘书处

2023 年 10 月 24 日

指数数据服务管理标准及设计介绍

【摘要】 数据服务通常被看作一个技术领域的问题，管理的重点在于如何实现业务需求。指数数据服务在大数据背景下，承担着指数供应商的价值传递，因此需要优化指数数据服务管理，提升基于安全、合规等要求的数据服务能力。

本文通过领域驱动设计的思想，对指数数据服务管理流程进行了分析，提出一套指数数据服务管理的设计，用来描述复杂业务场景下的工作流程，并将其作为关键技术引入的参考依据，同时对于重点功能领域进行了识别，便于职能部门之间进行沟通，减少技术实现与实际需求的差异所带来的风险。

关键词：指数数据服务管理 业务流程 功能领域

一、指数数据服务管理概述

数据服务指数据供应商通过技术手段提供数据采集、计算、存储、传输、交换等服务。在大数据时代的背景下，数据服务更被作为数据中台的最后一公里，所以数据服务通常被理解为一个技术范畴的问题，围绕着数据提供一组技术服务，包括但不限于标准化、分析及可视化等。

指数是金融行业的一个细分领域，但是随着人工智能在大数据时代的广泛应用，指数数据服务将承载不可估量的经济影响和社会效应，我国也意识到数据安全的重要性，近年颁布了《中华人民共和国数据安全法》及针对人工智能的管理规范《生成式人工智能服务管理办法》，因此当下指数数据服务在应用方面，需要满足安全、合规、可控以及高效四项基本要求，要满足这四方面要求，就需要增强对于指数数据服务管理的能力。

常规指数数据服务管理的重点是如何实现业务需求，基于这一思路的系统设计往往是单向直接的，即设计产生于需求，而系统又产生于设计。由于每个业务需求都有其独特性，随着业务需求的变更及系统复杂性的增加，单向且直接的设计方式将导致数据服务结构性问题的逐步突出，如系统架构、决策管理、可扩展性、数据安全以及性能等，在增加运维风险的同时，容易造就烟囱状系统。此外，指数数据服务承载了数据价值的传递，使得指数数据服务不能仅仅局限于

技术，更是一个管理问题。对指数数据服务管理的规范化，可以提高指数数据供应商的运营效率和竞争力。

二、指数数据服务管理设计

本节将围绕安全、可控、合规以及高效四个指数数据服务管理的基本要求，对指数数据服务业务流程进行分析，并为后续关键技术的引入提供参考依据。

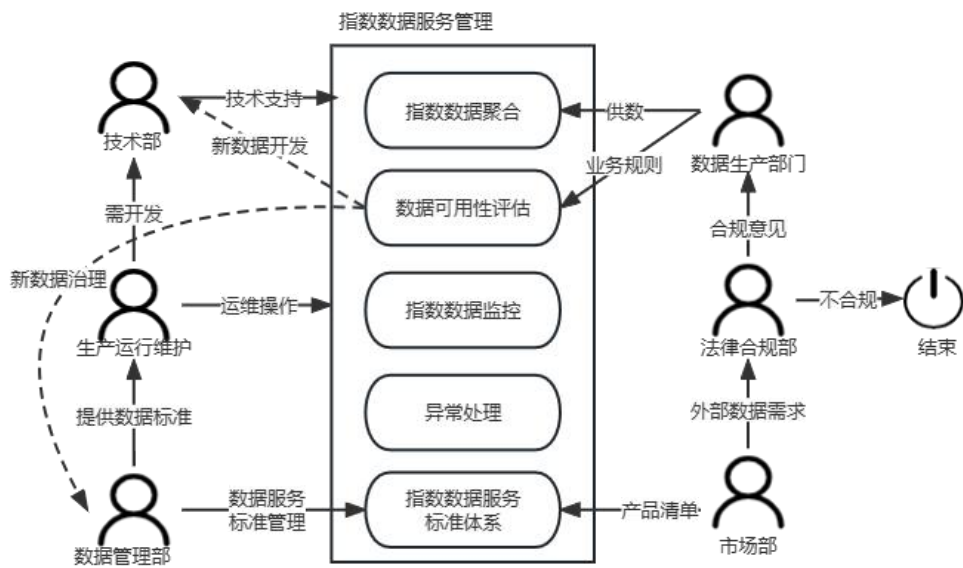


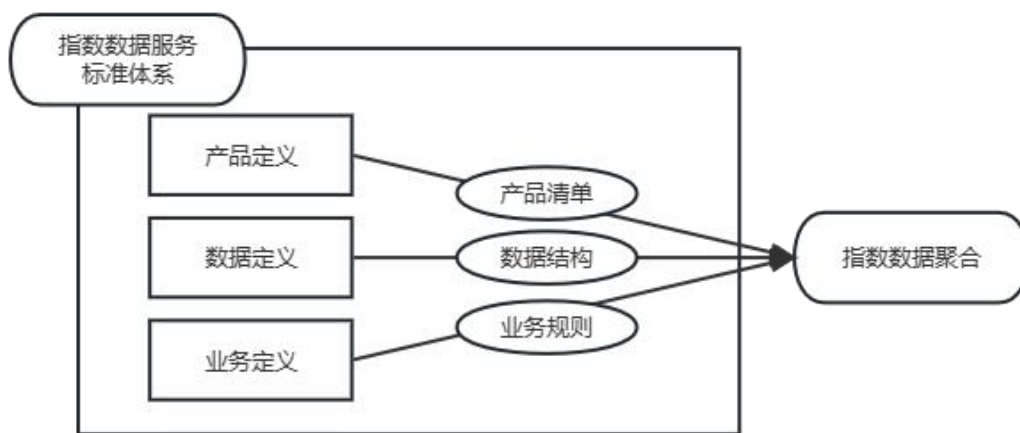
图 1 指数数据服务业务流程图

指数数据服务业务流程通常涉及多个职能部门，图 1 梳理了指数数据服务的主要业务流程。根据领域驱动设计（以下简称“DDD”）中领域保护的设计原则，指数数据服务管理可分为 5 个互相独立的功能领域：指数数据标准体系领域、异常处理领域、指数数据监控领域、数据可用性评估领域以及数据聚合领域。

下文中的子领域指某功能领域独有的领域；专业子领域指存在于领域上下文需要歧义消解的领域。

（一）指数数据标准体系

指数数据标准体系领域是指数数据供应商核心价值的体现，通常是基于数据治理的成果，形成的一系列数据产品并包含其商业政策说明。



上图方框为指数数据标准体系领域的子领域，具体如下：

1. 产品定义

产品由一组指数相关数据定义组成，产品定义有两部分功能：1.标准产品定义，标准产品定义包括但不限于：产品名、引用的一组指数相关数据定义（数据源、组合方式，以下简称“数据定义”）、生产频次、接口说明（API、接口文件等）、版本以及使用说明；2.非标准产品定义，产品名、数据模板、引用的一组数据定义（数据源、组合方式）、版本、权限以及使用说明。

2.数据定义

数据定义用于描述数据来源以及其聚合方式，具体如下：

(1) 元数据定义，元数据定义通常基于数据治理的结果，用于提供具体指数业务基础数据的数据来源、精度、描述等信息；

(2) 标准数据定义，标准数据基于上述元数据信息定义，包括标准数据名称、一组元数据及组合方式、数据来源、数据类型、数据精度等相关内容，为产品定义提供基础库。

(3) 复合数据定义，指一组标准数据定义的聚合方式，如在标准数据定义的基础上附加其他数据或是进行简单计算（求和、平均等算法），复杂或关键数据计算需要经过数据治理。此外，还提供基于标准数据定义的修订，如汇率精度使用扩位，常用于指数数据标准的修订或者客户定制类数据需求。

(4) 码表关系映射，需要建立内、外部两套码表以及映射关系的管理。内部代码主要用于数据聚合，外部码表作为指数数据服务标准数据体系一部分，用于管理对外的数据标准。

3.业务定义

用于定义和管理指数数据服务业务类型，如公共或基础数据服务、定制类数据业务，此外还需包括指数数据的生成

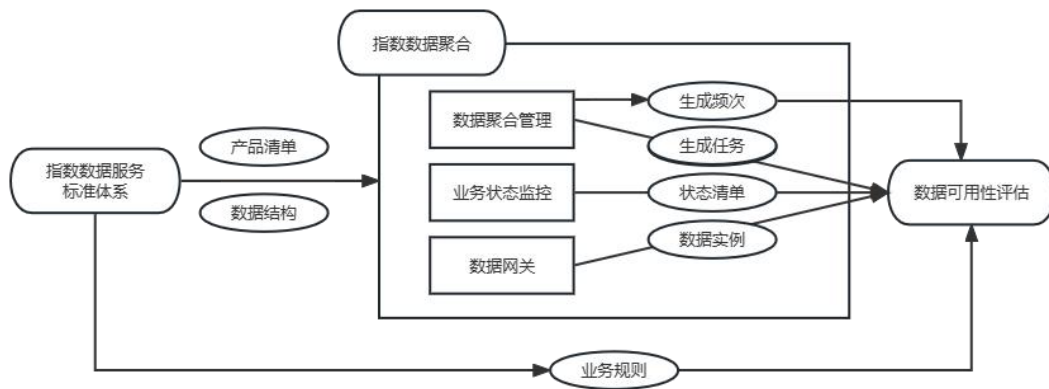
频次、业务校验规则（如搭载、过滤、转义等业务场景）等，其中定制类数据业务因实际情况会出现多层次的数据结构类型，通常建议在复合数据定义中进行定义和管理。

上述三个子领域存在上下文关系，需要提炼相同内涵专业子领域：产品清单、数据结构以及业务规则，用于领域上下文进行歧义消解，同时也能降低烟囱状系统出现的可能。

除此以外，指数数据标准体系领域需要提供接口文档动态生成的功能，用以保证指数数据标准体系与接口文档的一致性。

（二）指数数据聚合

指数数据聚合领域指基于指数数据标准体系的数据集成、聚合以及生命周期管理。



指数数据聚合需实现功能子领域如下：

1. 数据聚合管理

数据聚合管理包括但不限于数据集成、定义数据生成函数、可插拔式的数据生成组件、版本控制以及生成频次等。

(1) 数据生成函数参考了 FAAS 的建设思想，作为业务与后端服务之间的业务代理层，这一设计可以提升数据服务整体框架的敏捷程度，在不影响业务运行的前提下支持后端服务能力的横向扩展，该功能常用于提供实时指数数据服务。

(2) 可插拔式的数据生成组件及版本管理是用来支持数据服务的 DevOps，对公指数数据服务业务的上线并没有一个明确的成功标志，开发阶段无法模拟出所有下游客户实际的使用环境，此功能是为了快速地解决外部客户的诉求。

(3) 生成频次，是对指数数据服务数据标准体系在生产运行过程中的补充，是精益化指数数据服务业务管理的关键模块，常见的生成频次有：按工作日、交易日、周频、定调期、当前交易日下一日、当前交易日下一个交易日等，并且需要支持规则的可扩展性。

(4) 生成任务

基于任务驱动模型，生成任务通过可视化数据聚合流程配置以及流程引擎的功能实现。数据聚合流程配置，若流程不涉及人工交互，可采用流式处理框架，基于 DDD 的单一性设计原则，数据聚合流程使用模型驱动架构的框架进行设计，并对功能进行模块化开发，包括但不限于：

流程组件：如 JBPM 等支持模型驱动架构的开发框架。

装饰器组件：提升指数数据服务运行的安全等级，通过装饰器组件的开发 **Retry** 组件,在数据读取、数据聚合发生错误时，允许先自行执行错误补偿机制；日志组件，为与日志平台的对接提供标准化方法；跟踪组件，记录任务执行过程，对接数据聚合监控功能模块；独占组件，在分布式或多线程的环境下，提供按条件排他的组件功能。

监听组件：提供具有一定侵入性的监听功能，如业务状态监控。

业务功能模块：指不具备复用条件的业务编码。

2.业务状态监听

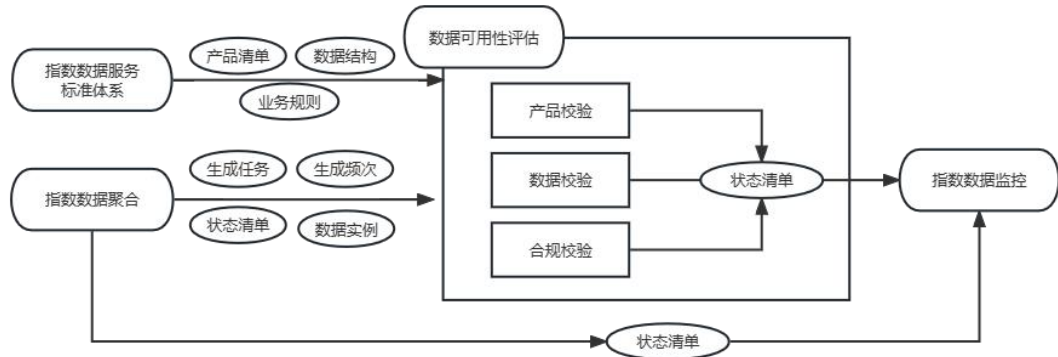
业务状态监听需定义一组业务状态，并支持主动以及被动两种模式。业务状态即包括外部驱动数据聚合任务时提供的业务状态，又包括内部驱动数据聚合任务时提供的处理状态。

同上一小节，指数数据聚合领域也需要提炼相同内涵专业子领域：生成频次、生成任务、状态清单以及数据实例，作为相关联领域的上下文。

（三）数据可用性评估

数据可用性评估领域的作用主要有两个，第一是作为数据产品对外分发前的最后一道防火墙，负责对数据产品中的数据、商业政策进行稽核；其次，提供数据聚合过程的反向追溯能力，应对外部用户对于数据产品的咨询。

该功能领域主要的作用是实现指数数据聚合的闭环校验，需实现的功能子领域如下：



1. 产品校验

该功能用于数据产品是否符合指数数据标准体系要求，包括数据字段使用、数值精度等，接合公司实际可附加权限稽核功能。

2. 数据校验

通常由数据生产者提供校验规则或校验结果；对于衍生数据（需经计算的数据），若该数据未经过数据治理，则必须提供额外的校验功能，该校验逻辑的实现不能复用数据聚合领域的代码，但是生产环境通常会出现一种情况，就是两套或多套异构校验逻辑之间的差异性不明显，则校验的方向转向所使用的计算函数，校验衍生数据计算是否使用了经过登记的计算函数。如上述校验不通过，则进入异常处理领域进行仲裁。

3. 合规校验

校验指数数据是否符合业务规则，是否按约定频次执行数据服务，此外还包括搭载数据、过滤或转义信息等场景。

对于自有数据，经过产品校验、数据校验之后，判断是否符合约定的分发频次，分发目标客户是否具有有效授权；对于搭载数据，需要判断是否数据搭载服务在有效期内；搭载数据服务是否需要进行脱敏或修订。

过滤需求通常来源于指数状态的改变、业务规则的修订以及临时的特殊事项，需要对外屏蔽部分数据。对于基于业务规则或临时性的过滤需求，合规校验用于判断是否对外披露了上述数据。

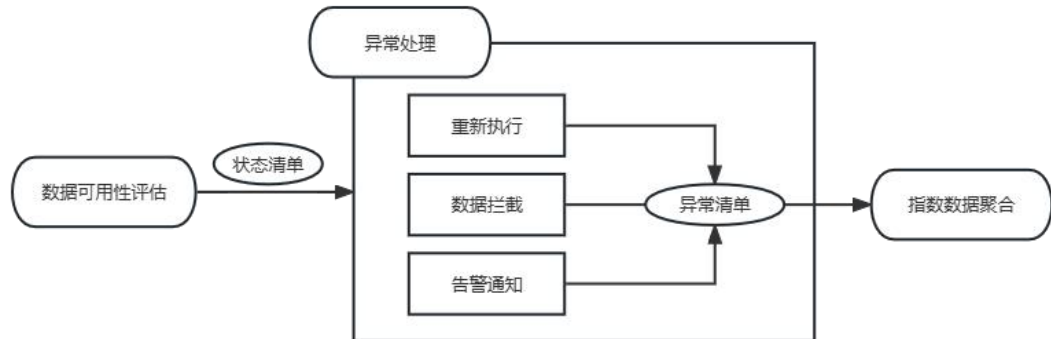
转义信息，包括内外部数据转义映射关系的管理，涉及标准指数数据以及客户定制类数据，此外转义信息还包括多语言的翻译功能。转义信息合规检查的核心在于是否经过业务部门的确认。

（四）指数数据监控

实现对于指数数据聚合的全过程监控功能，包括但不限于数据来源状态、时间戳、聚合流程状态、数据可用性评估状态、是否发生过拦截、是否重新生成、生成过程日志等等。根据指数数据异常的业务类型以及告警级别，给相关的人员进行告警通知。

（五）异常处理

异常处理领域是指数数据服务运维管理切入点，异常处理领域的子领域如下：



1.重新执行

重新执行是一种异常补偿机制，可以是业务驱动型，即数据生产部门提供的数据更新状态，从而触发数据聚合任务重新执行；可以是自驱动型，针对偶发性的技术问题提供补偿机制，尝试自动修复相关异常。

2.数据拦截

根据系统运维的情况，首先需要实现数据拦截功能，拦截功能有文件级、数据级两种类型。文件级拦截的使用场景有三种：1) 不进行数据聚合；2) 不进行指数数据服务文件生成；3) 不进行指数数据服务文件分发，4) 对于已经分发的指数数据服务文件进行删除，并告知已经下载相关文件的用户相关异常。数据级拦截场景与文件级拦截场景类似，差别在于数据聚合过程中屏蔽需拦截数据或者在已经聚合完成的数据产品中进行数据剔除，对于已经生成文件的数据进行重新生成，如该文件已经被用户下载，需告知客户相关异

常并提醒用户重新下载，以 API 形式读取数据客户的流程同上。

3.告警通知

告警通知需根据异常清单以及状态清单，评估受影响的指数数据产品范围，及时通知受影响的用户，告警通知通常需要对接监控平台。

异常事件为专业子领域，用来登记和记录所有指数数据产品生成过程中的异常及处理状态，该子领域对接对外通知功能。

三、指数数据服务关键领域

上一章节对于指数数据服务管理功能进行了设计，本节对于重点功能领域进行了识别，发现交互最多的专业子领域是状态清单以及异常事件（详见附件），状态清单以及异常事件处理通常涉及多个职能部门，基于状态清单的异常事件处理，需要结合公司的实际情况制定相应的应急操作流程。

附件：专业子领域状态清单及异常事件

附件

领域/子领域		专业子领域	数据形式	数据结构	校验规则	生成频率	生成任务	状态清单	数据实例	异常事件清单
指数数据	标准体系	产品定义	●							
		数据定义		●						
		业务定义			●					
指数数据聚合		数据聚合管理				●	●			
		业务状态监听						●		
		数据网关			●				●	
数据可用评估		产品校验			●			●		
		数据校验			●			●		
		合规校验			●			●		
数据		生成状态						●		●

	数据检索						●		●
	任务状态						●		●
异常处理	文件拦截						●		●
	数据拦截						●		●
	重收流程						●		●

(中证指数有限公司技术开发部高级工程师曹逸玮供稿)