



中 华 人 民 共 和 国 金 融 行 业 标 准

JR/T 0237—2021

金融大数据平台总体技术要求

General technical requirements of financial big data platform

2021-12-29 发布

2021-12-29 实施

中国人民银行 发 布

目 次

前言 II

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 缩略语 3

5 金融大数据信息价值链 3

6 金融大数据平台总体目标 4

7 金融大数据平台功能技术要求 4

 7.1 概述 4

 7.2 数据接入功能要求 5

 7.3 数据存储功能要求 6

 7.4 数据处理功能要求 7

 7.5 数据分析功能要求 8

 7.6 数据服务功能要求 9

 7.7 平台运维功能要求 10

8 金融大数据平台非功能技术要求 12

 8.1 安全性要求 12

 8.2 可靠性要求 12

 8.3 性能要求 12

9 金融大数据平台接口技术要求 13

 9.1 数据接入接口技术要求 13

 9.2 数据服务接口技术要求 13

 9.3 管理接口技术要求 13

附录（资料性）金融大数据平台应用场景 14

参考文献 16

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》给出的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中国人民银行提出。

本文件由全国金融标准化技术委员会（SAC/TC 180）归口。

本文件起草单位：招商银行股份有限公司、中国人民银行清算总中心、华为技术有限公司、中国邮政储蓄银行股份有限公司、北京国家金融标准化研究院有限责任公司。

本文件主要起草人：俞吴杰、杨鹏、陈翔、黄波、王鹏程、张天舟、张芮玲、王庆刚、孔昭龙、方方、那勇、王嘉琪、张兰兰、王伟利、白阳、符海芳、赵华、蒋健源、沈澍、李琪。

金融大数据平台总体技术要求

1 范围

本文件规定了面向金融业的大数据平台总体技术要求的框架结构、功能技术要求、非功能技术要求以及接口技术要求。

本文件适用于金融大数据平台的设计、开发和应用。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

JR/T 0236—2021 金融大数据 术语

ISO/IEC 20546:2019 信息技术 大数据 概述和术语

3 术语和定义

JR/T 0236—2021《金融大数据 术语》界定的以及下列术语和定义适用于本文件。

3.1

大数据 big data

海量的数据集，其数据在本质上具有体量大、种类多、变化快、变数多的特征，需要一种易扩展的技术来有效存储、处理、管理和分析。

注：1. 大数据通常以多种不同方式使用，例如，作为某种用于处理大数据海量数据集的易扩展技术的名称。

2. GB/T 35295—2017《信息技术 大数据 术语》中2.1.1对大数据的定义为“具有体量巨大、来源多样、生成极快、且多变等特征并且难以用传统数据体系结构有效处理的包含大量数据集的数据”。

3. 在某些非工程性研讨的场合，“大数据”一词的外延可能被扩大到所有的数据。

[来源：JR/T 0236—2021，3.1]

3.2

大数据参考体系结构 big data reference architecture

一种用作工具以便于对数据内在要求、设计结构和运行进行开放性探讨的高层概念模型。

注：比较普遍认同的大数据参考体系结构一般包含：系统协调者、数据提供者、大数据应用提供者、大数据框架提供者和数据消费者等5个逻辑功能构件。

[来源：JR/T 0236—2021，4.1]

3.3

系统协调者 system orchestrator

大数据参考体系结构中的一种逻辑功能构件，可定义所需的数据应用活动并将其整合到可运行的垂直系统中。

注：1. 系统协调者可以是人、软件或这二者。

2. 系统协调者一般包括：业务领导者、咨询师、数据科学家、信息体系结构设计师、软件体系结构设计师、安全体系结构设计师、个人信息保护体系结构设计师和网络体系结构设计师。

[来源：JR/T 0236—2021，4.2]

3.4

数据提供者 data provider

大数据参考体系结构中的一种逻辑功能构件，可将新的数据或信息引入大数据系统。

注：数据提供者一般包括：企业、公共机构、科学家、调研人员、从事数据搜索的工程师、网络应用软件、网络运营商和末端用户。

[来源：JR/T 0236—2021，4.3]

3.5

大数据应用提供者 big data application provider

大数据参考体系结构中的一种逻辑功能构件，可执行数据生命周期操作，以满足系统协调者定义的需求以及安全和隐私保护的需求。

注：大数据应用提供者一般包括：大数据应用领域专家、大数据平台领域专家和大数据咨询师。

[来源：JR/T 0236—2021，4.4]

3.6

数据消费者 data consumer

大数据参考体系结构中的一种逻辑功能构件，是使用大数据应用提供者所提供的应用的末端用户或其他系统。

注：数据消费者一般包括：末端用户、调研人员、应用和系统。

[来源：JR/T 0236—2021，4.6]

3.7

分布式文件系统 distributed file system

多个结构化数据集分布在一个或多个服务器集群的各计算节点的文件系统。

注：此类系统中，数据可能分布在文件或数据集层，更为普遍的是在数据块层级分布，同时支持集群中多个节点与大型文件或数据集的不同部分交互。

[来源：JR/T 0236—2021，6.12]

3.8

非结构化数据 unstructured data

特征为除了记录或文件级别外没有任何结构的数据。

注：总体而言，非结构化数据不是由数据元素组成。

示例：自由文本是非结构化数据的一个例子。

[来源：JR/T 0236—2021，3.30]

3.9

工作流 workflow

指具有明确起点和终点以描述任务的活动序列。

注：工作流是一类活动模型，且可能具有分支、决策点和事件。

[来源：GB/T 32854—2020，3.1]

4 缩略语

下列缩略语适用于本文件。

API: 应用程序接口 (Application Program Interface)
AI: 人工智能 (Artificial Intelligence)
CEP: 复杂事件处理 (Complex Event Processing)
CPU: 中央处理器 (Central Processing Unit)
DDL: 数据库模式定义语言 (Data Definition Language)
FTP: 文件传输协议 (File Transfer Protocol)
FTPS: 基于SSL的文件传输协议 (File Transfer Protocol Over SSL)
HTML: 超文本标记语言 (Hyper Text Markup Language)
I/O: 输入/输出 (Input/Output)
JDBC: Java数据库连接 (Java Database Connectivity)
JAR: 一种软件包文件格式 (Java Archive)
JSON: JavaScript对象简谱 (JavaScript Object Notation)
LPA: 标签传播算法 (Label Propagation Algorithm)
MPI: 信息传递接口 (Message Passing Interface)
ODBC: 开放数据库互连 (Open Database Connectivity)
REST: 表述性状态传递 (Representational State Transfer)
RPC: 远程过程调用 (Remote Procedure Call)
SCP: 安全复制 (Secure Copy)
SDK: 软件开发工具包 (Software Development Kit)
SFTP: 安全文件传送协议 (Secret File Transfer Protocol)
SQL: 结构化查询语言 (Structured Query Language)
SSL: 安全套接层协议 (Secure Sockets Layer)
UI: 用户界面 (User Interface)
Web: 全球广域网 (World Wide Web)
XML: 可扩展标记语言 (Extensible Markup Language)

5 金融大数据信息价值链

从信息价值链的角度来看,金融大数据平台面向金融大数据应用提供者、系统协调者及大数据框架提供者,核心活动在于从数据提供者处收集、处理、分析金融活动过程中的相关业务数据、用户信息、行为数据等,并结合金融业务特征支撑金融机构的各项活动和服务。金融大数据信息价值链见图1。

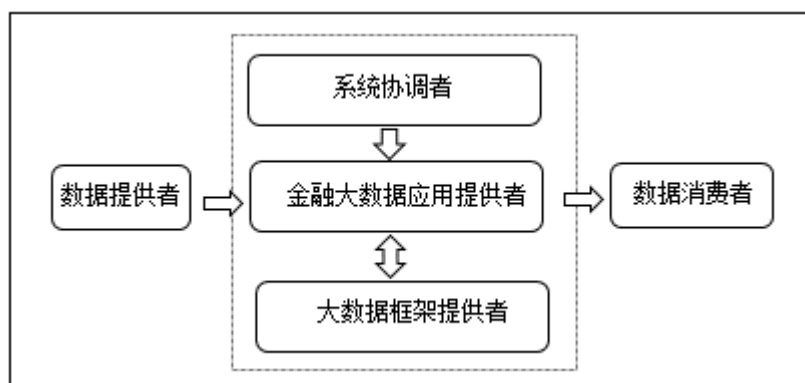


图1 金融大数据信息价值链

数据提供者金融大数据平台提供数据或信息，包括数据源提供者、数据流通平台提供者和数据API提供者。

系统协调者在金融领域规范和协调各类所需的数据应用活动，包括系统的配置管理、作业调度、资源调度、运行监控等活动。

金融大数据应用提供者满足金融领域大数据服务需求，包括收集、预处理、分析、可视化和访问等活动。

大数据框架提供者金融大数据应用提供者在创建具体应用时提供资源和服务。

数据消费者是金融大数据应用的最终使用者或其他系统。

6 金融大数据平台总体目标

金融大数据平台的总体目标是帮助金融机构更高效、更快速地完成金融大数据应用的开发、部署和管理，从以交易为中心转向以数据为中心，以应对更多维、更大量、更实时的数据和互联网业务的挑战。金融机构以防范系统性风险为底线，对数据安全和业务连续性保障的技术要求通常高于其他行业。

7 金融大数据平台功能技术要求

7.1 概述

金融大数据平台是企业级、分布式、开放、统一的大数据平台，应包括数据接入、数据存储、数据处理、数据分析及数据服务相关组件。从功能的必要性及重要性角度看，金融大数据平台具体功能技术要求可以划分为基本要求和增强要求。增强要求是从技术的发展趋势和金融用户的前瞻性需求入手提出的。

数据接入的基本要求应包括作业调度、数据传输、消息队列、批量采集，增强要求宜包括实时采集；数据存储的基本要求应包括分布式关系数据库、分布式存储系统（文件、对象、块）等，增强要求宜包括分布式列数据库、分布式图数据库；数据处理的基本要求应包括批处理、流处理、图计算，增强要求宜包括批流融合计算、内存计算；数据分析的基本要求应包括模型库、数据处理引擎等，增强要求宜包括数据可视化；数据服务的基本要求应包括风险管理、用户画像、营销管理、客户服务和经营分析等。金融大数据平台技术架构见图2。

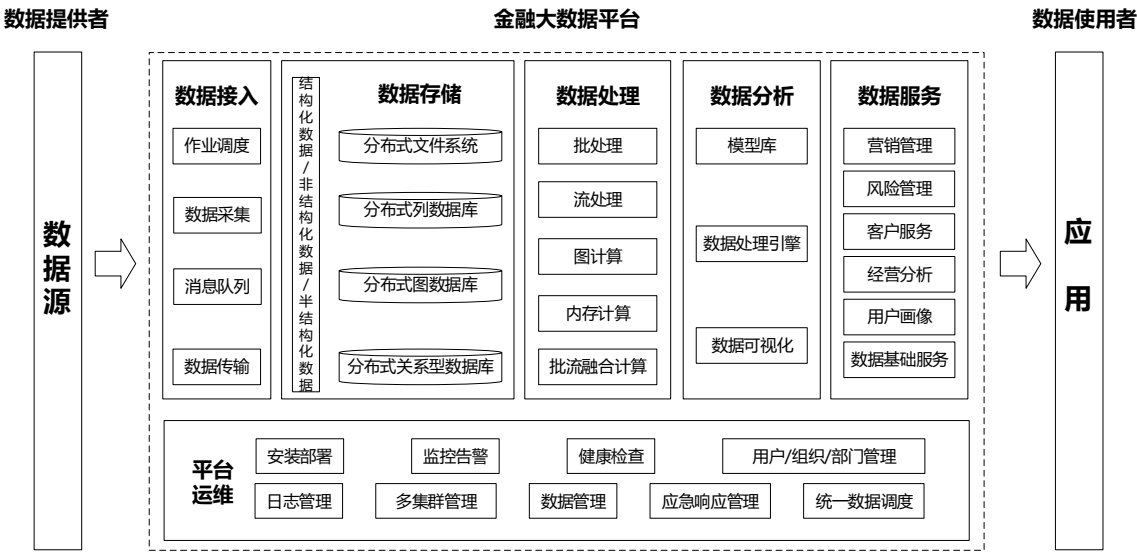


图2 金融大数据平台技术架构

金融大数据平台由数据接入、数据存储、数据处理、数据分析、数据服务和平台运维等部分组成，主要内容如下：

- a) 数据接入提供结构化数据、非结构化数据或半结构化数据的采集、数据的导入导出以及数据流的定义等功能，支持多系统集成，支持多协议接入，如支持RESTful、RPC、异步消息接入。结构化数据包括数据库文件以及结构化文本文件等。非结构化数据包括图片、音频、视频、文档及应用数据等。半结构化数据包括XML、表格、JSON、HTML、电子邮件等。
注：REST指的是一组架构约束条件和原则，满足这些约束条件和原则的应用程序或设计称为RESTful。
- b) 数据存储提供大数据的分布式存储管理，涵盖多种存储方式和文件系统。
- c) 数据处理支持结构化、半结构化及非结构化数据，提供丰富的API接口和SDK开发包，支持分布式计算技术、流处理技术、内存计算技术，实现稳定、强大的数据处理能力。
- d) 数据分析采用数据挖掘、机器学习等技术，支持多个数据处理引擎，具备数据可视化、多种算法、建模分析等深度挖掘分析能力。
- e) 数据服务提供海量金融交易数据、用户信息数据、金融业务数据的访问、控制等基础服务，提供营销管理、风险管理、客户服务、经营分析及用户画像等高级数据服务，金融大数据平台支撑的应用场景见附录。
- f) 平台运维具备高可靠、安全、容错、易用的集群管理能力，支持大规模集群的安装部署等功能，提供分钟级别的统一风险管理，具备各类日志的指标实时采集能力、通用的SQL查询能力、可视化监控能力。

7.2 数据接入功能要求

7.2.1 作业调度

作业调度的功能要求如下：

- a) 应支持 workflow 任务的调度。
- b) 应支持 workflow 的编排。
- c) 应支持 workflow 的批量启停、导入、导出。
- d) 应支持通过接口方式进行 workflow 的增、删、改、查管理。

- e) 宜支持工作流的可视化。

7.2.2 数据采集

数据采集的功能要求如下：

- a) 应支持数据库与文件系统间数据、文件的导入和导出。
- b) 应支持数据的批量采集。
- c) 宜支持数据的实时采集。
- d) 应支持任务的并行执行，在任务失败后可以重新调度。
- e) 应支持通过可视化配置等形式进行的数据采集。
- f) 宜支持异构数据库之间的建表语句转换DDL建表功能。

7.2.3 消息队列

消息队列的功能要求如下：

- a) 应支持高吞吐、分布式消息的处理。
- b) 应支持消息的缓存。
- c) 应支持离线应用和在线应用对消息的使用和处理。
- d) 宜支持多种客户端。

7.2.4 数据传输

数据传输的功能要求如下：

- a) 应支持通用的数据传输协议。
- b) 应支持断点续传机制。
- c) 应支持同步、异步传输机制。
- d) 应支持传输安全加密机制。
- e) 宜支持高速大容量数据传输机制，如FTP、FTPS、SCP、SFTP等。

7.3 数据存储功能要求

7.3.1 分布式文件系统

分布式文件系统的功能要求如下：

- a) 应支持文件的上传、下载、读写、复制、移动、删除、访问控制等。
- b) 应支持文件的搜索、批量操作、回收站、快照等。
- c) 应支持文件容错机制和系统高可用机制，包括数据块的备份、系统快速恢复等。
- d) 应支持文件数据的校验和同步，保证数据的完整性与一致性。
- e) 应支持RESTful接口对分布式文件系统进行基本操作。
- f) 应支持金融大数据平台对国内外主流操作系统的兼容性，如Linux、Unix、AIX、MacOS、Kylin、EulerOS等。

注：Linux、Unix、AIX、MacOS、Kylin、EulerOS为操作系统名称。Linux是一种免费使用和自由传播的操作系统；Unix是20世纪70年代初出现的一种操作系统，除了作为网络操作系统之外，还可以作为单机操作系统使用；AIX是国际商业机器公司开发的操作系统，名称来自先进交互运行系统（Advanced Interactive executive，AIX）；MacOS是由苹果公司开发的操作系统；Kylin是一个中国自主知识产权操作系统；EulerOS是华为自主研发的服务器操作系统。

- g) 宜支持将小文件自动打包为大文件进行集中存储。
- h) 宜支持对不同型号CPU的兼容性。
- i) 宜支持弹性扩展功能、动态添加操作以及删除节点操作。
- j) 宜支持对数据资源的统一检索、编目、增加和删除操作。

7.3.2 分布式列数据库

分布式列数据库的功能要求如下：

- a) 应支持数据以键值形式进行存储。
- b) 应支持基于表级、行级和列级的用户权限管理。
- c) 应支持对数据库数据进行列加密。
- d) 应支持多级索引。
- e) 宜支持将多个具有类似功能或存在关联的业务表合并存储。

7.3.3 分布式图数据库

分布式图数据库的功能要求如下：

- a) 应支持由节点及边组成（即节点间关系）的数据模型。
- b) 应支持图查询、图遍历及图分析。
- c) 应支持主流开发接口，如RESTful、Gremlin等的接口。

注：Gremlin是一种图遍历语言。

- d) 宜支持单节点、多节点多层关系扩线查询。
- e) 宜支持顶点、属性的继承操作。
- f) 宜支持长任务异步会话机制。

7.3.4 分布式关系型数据库

分布式关系型数据库的功能要求如下：

- a) 应支持结构化数据的分布式存储机制。
- b) 应支持SQL实现数据的查询操作，包括并发事务控制、存储过程等。
- c) 应支持多表关联。
- d) 应保障数据分布式存储的一致性。
- e) 应支持关系型数据库间的数据进行跨库关联。
- f) 宜支持行列混合存储，表按行或列格式存储。
- g) 宜支持与ODBC接口的兼容性。

7.4 数据处理功能要求

7.4.1 批处理

批处理的功能要求如下：

- a) 应支持多种数据类型的离线分析，包括结构化数据、半结构化数据、非结构化数据。
- b) 应支持离线计算任务进度与状态的实时上报。
- c) 宜支持执行多节点离线任务联动。
- d) 宜支持多种语言分析任务的开发接口。
- e) 宜支持分散-聚集的处理方式。

7.4.2 流处理

流处理的功能要求如下：

- a) 应支持数据的实时获取、处理、输出和持久化。
- b) 应支持用户级别的访问控制。
- c) 应支持对消息处理任务进行全生命周期管理，包括创建、浏览、中止、激活、去激活等。
- d) 应支持滚动窗口和滑动窗口两种窗口方式。
- e) 应提供SQL或类SQL的数据操作接口。
- f) 应支持在出现故障情况下，使用容错机制处理事件。

- g) 应具备高容错能力，如节点、进程等出现异常时，能够重新部署该处理单元。
- h) 宜支持与JDBC接口的兼容性。
- i) 宜支持基于多种分布式流处理引擎的查询语言，如flink、spark等。
注：flink是一种开源流处理框架，spark是一种专为大规模数据处理而设计的快速通用的计算引擎。
- j) 宜提供画布、SQL和JAR等多种流计算开发模式。

7.4.3 图计算

图计算的功能要求如下：

- a) 应支持同步计算模型或异步计算模型编写迭代算法。
- b) 应支持基于属性图模型的图数据表达，包含结点或边上的标签和属性类型定义。
- c) 应支持内置常用图指标计算功能，以描述图的拓扑结构特征。
- d) 应支持实现水平扩展的分布式图计算和图查询。
- e) 宜支持明细数据的多种导入方式，包括全量导入、增量导入以及自定义导入。
- f) 宜支持索引，提供在线图分析和图查询功能。
- g) 宜支持PageRank、社区发现、最短路、LPA等多种图计算能力。

注：PageRank是谷歌排名运算法则的一部分，是谷歌用来标识网页的等级或重要性的一种方法。

7.4.4 内存计算

内存计算的功能要求如下：

- a) 应支持负载均衡和水平扩展。
- b) 应支持对多种数据类型的离线分析，包括结构化数据、半结构化数据、非结构化数据。
- c) 应支持高度抽象算子，快速构建分布式的数据处理应用。
- d) 宜支持标准SQL语法。
- e) 宜支持读取非关系型数据库数据。

7.4.5 批流融合计算

批流融合计算的功能要求如下：

- a) 宜支持统一查询SQL语言。
- b) 宜支持多种场景下的流式SQL，如位置信息分析等。
- c) 宜支持常用时间窗口，包括滚动窗口、滑动窗口等。
- d) 宜支持基于SQL的批流数据的模式识别。
- e) 宜支持事件驱动的流处理，降低处理延迟。
- f) 宜支持处理乱序事件流、窗口计算、CEP等。
- g) 宜支持对复杂任务的调度，如支持深度学习的训练、MPI任务等。

7.5 数据分析功能要求

7.5.1 模型库

模型库的功能要求如下：

- a) 应支持具备机器学习、深度学习、自然语言处理等能力的数据挖掘及分析算法模型。
- b) 应支持动态扩展。
- c) 应支持模型有效性校验及评估。
- d) 应支持用户灵活的自定义模型。
- e) 应提供算法I/O注册、算法注册的接口，算法厂商的算法要与存储解耦。
- f) 算法镜像文件存储宜对接分布式文件系统，保证镜像数据的可靠性。
- g) 应为Web类常驻应用提供访问代理和负载均衡服务，允许用户自定义域名。

- h) 应提供标准的RESTful接口，方便业务系统集成。
- i) 宜支持对多算法厂商多算法的统一注册、管理。
- j) 宜支持各算法厂商独立维护、升级算法。
- k) 宜提供友好直观的UI设计，方便用户操作管理。
- l) 宜对外提供在线或离线算法服务。

7.5.2 数据处理引擎

数据处理引擎的功能要求如下：

- a) 应支持图计算引擎。
- b) 应支持流处理引擎。
- c) 宜支持语义推理引擎。
- d) 宜支持规则引擎。
- e) 宜支持预测算法、机器学习、推荐算法、相似度算法、分类聚类、文本挖掘。

7.5.3 数据可视化

数据可视化的功能要求如下：

- a) 数据应可视化，支持多维度、多种形式的数据展现。
- b) 数据模型应可视化，支持可视化交互、调整模型。
- c) 数据处理过程宜可视化，包括数据获取、数据分析、数据展现等。
- d) 宜支持信息流可视化，包括交易过程等。
- e) 宜支持统计分析可视化，包括用户视图、交易视图、资金视图等。

7.6 数据服务功能要求

7.6.1 营销管理

营销管理的功能要求如下：

- a) 应提供客户画像和客户标签服务。
- b) 应提供拓展客户、营销、销售、运营、服务、考核等客户营销全流程服务。
- c) 应提供多渠道营销服务。
- d) 应提供实时营销服务。
- e) 宜提供AI智能营销服务。

7.6.2 风险管理

风险管理的功能要求如下：

- a) 应提供客户征信信息服务。
- b) 应提供客户风险识别、风险计量和评估服务。
- c) 应提供客户风险监测、风险控制、风险预警和处理服务。
- d) 应支持用户风险识别分析、用户行为预测分析。
- e) 应支持用户风险评估和监测分析、实时发现交易风险、实时处理风险、实时处理反欺诈等。
- f) 应支持对经过识别和评估的风险采取分散、对冲、转移、规避和补偿等措施进行分析，支持对抵押、担保、金融衍生品等风险缓释工具的分析。

7.6.3 客户服务

客户服务的功能要求如下：

- a) 应提供客户管理信息服务。
- b) 应提供客户账单管理、积分管理、财富体验、财务规划等增值服务。
- c) 应提供远程开户、智能客服等多渠道客户服务。

- d) 宜提供刷脸支付、无感支付等多种智能支付服务。
- e) 宜提供智能投顾、智能保单识别、智能金融咨询等多种智能金融服务。

7.6.4 经营分析

经营分析的功能要求如下：

- a) 应提供资产负债管理服务。
- b) 应提供财务管理、管理会计服务。
- c) 应提供多维度绩效考核分析服务。
- d) 应提供运营管理服务。
- e) 应提供合规审计服务。

7.6.5 用户画像

用户画像的功能要求如下：

- a) 应支持用户自然属性数据分析。
- b) 应支持用户系统日志、交易数据分析。
- c) 应支持用户标签及多维度细分，包括用户价值、活跃度、忠诚度、影响力、消费能力、用户属性、消费偏好等。
- d) 应支持多用户的关联分析。
- e) 宜支持用户社交网络等第三方数据分析。

7.6.6 数据基础服务

数据基础服务的功能要求如下：

- a) 应支持交易数据访问控制，包括实时交易数据、历史交易数据等。
- b) 应支持用户数据访问控制，包括用户基本信息、金融业务偏好等。
- c) 应支持金融业务数据访问控制，包括负债业务、资产业务、中间业务等。
- d) 应支持上层应用注册机制、鉴权、管理。
- e) 应支持服务功能查询机制。
- f) 应支持开放的API调用接口。
- g) 应支持批量数据查询获取。
- h) 应支持数据订阅功能。
- i) 应支持元数据驱动的数据地图管理及服务，包括元数据自动获取、数据血缘分析等。

7.7 平台运维功能要求

7.7.1 安装部署

安装部署的功能要求如下：

- a) 应支持用户根据业务需要对金融大数据平台容量进行调整，包括对主机、服务实例等的调整。
- b) 应支持用户定期升级系统，支持关键业务滚动升级。
- c) 应支持补丁上传、安装、卸载。
- d) 宜支持对金融大数据平台所有组件的自动化、向导式安装部署。
- e) 宜提供开箱即用的组件开发包。
- f) 宜支持集群、组件和管理系统统一进行安装部署。

7.7.2 监控告警

监控告警的功能要求如下：

- a) 应支持对主机、服务及服务实例等进行实时监控。
- b) 应支持系统自动感知故障并生成告警信息。

- c) 应支持以告警的方式展示运行过程中的系统故障，便于用户及时定位及处理系统故障。
- d) 应支持多种北向网管协议，支持与第三方告警和监控系统对接。
- e) 宜提供告警恢复的方法，帮助用户快速解决系统故障。
- f) 宜支持可视化的监控告警。

7.7.3 健康检查

健康检查的功能要求如下：

- a) 应支持对系统运行环境进行自动健康检查。
- b) 应支持健康检查报告的导出及问题分析。
- c) 宜提供健康检查的界面。
- d) 宜支持手工健康检查。

7.7.4 用户/组织/部门管理

用户/组织/部门管理的功能要求如下：

- a) 应支持动态添加和删除用户/组织/部门。
- b) 应支持管理模型，实现将用户动态配置到相应的用户/组织/部门中。
- c) 应支持对用户/组织/部门进行统一管理及认证。
- d) 应支持对用户配置不同角色及与角色对应的权限管理。

7.7.5 日志管理

日志管理的功能要求如下：

- a) 应支持分布式日志收集。
- b) 应支持对日志的采集、聚合和传输。
- c) 应满足日志可读性、操作痕迹可追溯性和细化日志埋点的要求。
- d) 宜支持对日志的可视化操作，包括日志收集、日志下载、日志查询等。
- e) 宜支持日志审计和日志导出。

7.7.6 多集群管理

多集群管理的功能要求如下：

- a) 应支持对多个金融大数据集群（以下简称多集群）进行集中统一管理和监控。
- b) 应针对不同集群进行资源队列的分层管理，支持动态刷新资源池。
- c) 应对多集群的存储资源进行全局配额管理。
- d) 宜支持多集群及自建集群运维管理，如启动、停止、修改等。
- e) 宜支持展示纳管集群的运维信息，如主机信息、服务信息、组件信息。

7.7.7 数据管理

数据管理的功能要求如下：

- a) 应支持统一元数据管理，实现对元数据的自动捕获及更新。
- b) 应支持数据质量管理，实现对数据质量规则的可视化管理及监控、自动化调度及监控。
- c) 应支持对数据血缘的自动捕获及更新。

7.7.8 应急响应管理

应急响应管理的功能要求如下：

- a) 应制定应急响应运营管理制度，明确人员岗位职责，实现 24 小时不间断监控和响应服务。
- b) 应具备应急管理接口或具体功能，可纳管或处理突发情况。
- c) 运维管理模块应具有冗余部署功能，防止运维管理模块单点故障。

7.7.9 统一数据调度

统一数据调度的功能要求如下：

- a) 宜支持对各种作业任务的调度。
- b) 宜支持周期性调度，可配置年、季、月、周、天，甚至是小时级别的调度周期，可配置作业启动时间。
- c) 宜支持依赖调度。
- d) 宜支持参数调度。
- e) 宜支持作业任务监控和作业任务重跑等。

8 金融大数据平台非功能技术要求

8.1 安全性要求

金融大数据平台的安全性要求如下：

- a) 应支持数据传输链路冗余机制，保障数据传输可靠性。
- b) 应支持数据采集、传输、存储、处理等全流程安全加密机制，防止信息泄露。
- c) 应支持数据资产分级、权限访问控制及数据处理策略，权限访问控制需要根据资产敏感等级决定权限控制的粒度，如角色组级别、表级别、字段级别等。
- d) 应支持数据清洗、转换及使用过程中的安全性及可追溯性。
- e) 应支持个人隐私信息保护，应支持个人信息授权同意、更正、删除。
- f) 应支持基于用户角色的安全控制机制。
- g) 应支持应用、设备与外部服务组件安全管理。
- h) 应支持数据的脱敏、变形。
- i) 宜支持金融大数据平台与业界主流云计算技术栈的集成与兼容。
- j) 宜支持多租户数据安全。
- k) 宜支持行为审计与分析审计。

8.2 可靠性要求

金融大数据平台的可靠性要求如下：

- a) 应支持系统切换过程与数据库平滑切换，保障数据不丢失。
- b) 应支持在系统切换过程中产生告警或事件上报。
- c) 应支持对系统集群的备份恢复。
- d) 宜提供对管理节点的网络平面和业务平面分离部署功能。
- e) 宜支持指定数据中心存储备份数据或副本数据。
- f) 宜具备集群互备能力。
- g) 宜支持系统冗余部署。

8.3 性能要求

金融大数据平台的性能要求如下：

- a) 负载应满足如下要求：
 - 支持对多台云服务器进行流量分发的负载均衡服务，扩展应用系统对外服务能力，提升应用系统的可用性，满足业务系统的需求。
 - 易于扩展，随数据增加而扩展。
 - 在容量扩展时，金融大数据平台负载性能同步得到提升。
 - 支持构建跨地域统一存储资源池。

- 支持在多个数据中心建立统一存储资源池。
- 选择标准化的部件，利于灵活替换和容量扩展。
- b) 资源利用率应提供租户级百分比和绝对值的混配策略，满足各种灵活的租户资源调度诉求。
- c) 并发应满足如下要求：
 - 支撑潜在的来自全部用户高并发查询、分析业务需求。
 - 高并发情况下的响应速度需在可接受范围内，保证系统不宕机。
 - 多集群的运维任务可并发执行、互不影响，应支持在任务管理中心显示集群名称，支持按集群检索任务。
- d) 处理效率应满足如下要求：
 - 对查询响应时间要求较高的业务，查询速度快，其响应时间在 1 秒内。
 - 查询条件简单，80%查询是主键查询，其他查询是简单条件组合查询。

9 金融大数据平台接口技术要求

9.1 数据接入接口技术要求

数据接入接口的技术要求如下：

- a) 应支持对数据源的注册、鉴权及管理。
- b) 应支持高性能批量导入。
- c) 应支持数据库、数据仓库、文件、Hadoop、NoSQL等多种类型的数据源。
注：Hadoop是一种分布式系统基础架构。NoSQL指不遵循关系模型的数据库。
- d) 宜支持大量在线实时数据采集。

9.2 数据服务接口技术要求

数据服务接口的技术要求如下：

- a) 应支持对应用的注册、鉴权及管理。
- b) 应支持应用数据查询、获取及更新。
- c) 应支持批量数据查询及在线数据交互式查询获取。
- d) 应支持数据分析、数据请求及数据订阅功能。
- e) 应支持RESTful、API、RPC等接口技术。

9.3 管理接口技术要求

管理接口的技术要求如下：

- a) 应支持用户管理。
- b) 应支持灰度升级维护。
- c) 应支持安全维护。
- d) 宜支持对软硬件资源的配置管理。

附录 (资料性)

金融大数据平台应用场景

1 风险管理

1.1 场景描述

风险管理以反欺诈场景为例。金融欺诈风险主要来自针对借记卡和信用卡的身份盗用、持卡人欺诈、第三人欺诈、商户欺诈，以及针对电子渠道的钓鱼网站、中间人欺诈、木马欺诈等。银行业金融机构应有效地甄别、防范和控制风险。

1.2 解决方案

金融大数据平台实时反欺诈功能通过客户的自然属性、行为属性、客户信用度、资产负债状况、交易环境等信息进行实时交易行为监测，通过交易记录、频度、位置等信息，实时分析判断交易风险，并针对不同类别和级别的疑似欺诈交易做相应处理，将传统的风险事后跟踪转变为事中控制。实时反欺诈大数据分析可以通过统一管理银行业金融机构内部多源异构数据结合依法查询外部征信数据，建立完善的风险防范体系。

2 市场营销

2.1 场景描述

市场营销以精准营销场景为例。在金融产品标准化、同质化的背景下，传统的营销模式效果不佳。银行业金融机构需要选择合适的时间、合适的产品推广给真正需要的客户，动态掌握市场需求的变化，提升产品营销效率。

2.2 解决方案

金融大数据平台通过对客户数据进行挖掘、追踪、分析，将不同客户群体进行聚类，分析客户的消费习惯、风险收益偏好等特征信息，对细分的客户进行个性化智慧营销。例如，根据客户的购买习惯和风险偏好进行动态的理财产品组合营销。根据客户行为轨迹分析，主动推送关联产品服务，真正做到个性化的主动营销服务，提升客户整体满意度。

3 客户服务

3.1 场景描述

数字化时代的银行业金融机构在互联网的冲击下，需要满足不断涌现的各种新需求，更好地服务客户，提升电话银行、网络银行、手机银行及社交网络等多渠道的个性化、专业化服务水平。

3.2 解决方案

通过金融大数据平台改善银行业金融机构与客户之间的交互模式，为客户提供增值服务，不断增强银行业金融机构核心竞争力，增强客户粘性。银行业金融机构可将客户行为转化为信息流，分析客户的个性特征和偏好，更深层地理解客户的交互习惯，为用户画像，智能化分析和预测客户需求，从而进行产品创新和服务优化。

4 经营分析

4.1 场景描述

金融服务竞争力某种程度上就体现在整体服务效率上，如快速放款、产品组合优化等涉及多部门、多环节审批处理流程，需要系统化地分析并优化金融服务流程，提升整体经营效率。

4.2 解决方案

通过大数据分析的方法改善经营决策,为管理层提供可靠的数据支撑,使经营决策更加高效、敏捷、准确,降低人工处理成本,提高流程效率。金融大数据平台可以监控不同业务、产品的服务质量和效益,从而进行产品组合设计、推广、合作渠道的调优。同时,银行业金融机构也可以选择合适的渠道推广合适的金融产品或金融服务,从而进一步提升经营效率和质量。

参 考 文 献

- [1] GB/T 22239—2019 信息安全技术 网络安全等级保护基本要求
 - [2] GB/T 32854—2020 自动化系统与集成 制造系统先进控制与优化软件集成
 - [3] GB/T 35273—2020 信息安全技术 个人信息安全规范
 - [4] GB/T 35274—2017 信息安全技术 大数据服务安全能力要求
 - [5] GB/T 35589—2017 信息技术 大数据 技术参考模型
 - [6] GB/T 36073—2018 数据管理能力成熟度评估模型
 - [7] GB/T 37988—2019 信息安全技术 数据安全能力成熟度模型
 - [8] JR/T 0171—2020 个人金融信息保护技术规范
 - [9] ISO/IEC 20546:2019 Information technology — Big data — Overview and vocabulary
 - [10] ISO/IEC TR 20547 Information technology — Big data reference architecture
-