

ICS 35.080

L 70

团 体 标 准

T/GDCSA 004—2021

商业智能与大数据分析软件通用技术规范

General technical specifications for business intelligence and big data
analysis software

2021 - 08 - 23 发布

2021 - 08 - 24 实施

广东省网络空间安全协会 发布

目 次

前言..... II

引言..... III

1 范围..... 1

2 规范性引用文件..... 1

3 术语和定义..... 1

4 支撑环境需求..... 4

 4.1 操作系统..... 4

 4.2 数据库..... 4

 4.3 中间件..... 4

5 软件功能需求..... 4

 5.1 统一平台管理..... 4

 5.2 数据管理..... 5

 5.3 数据分析..... 6

 5.4 增强分析..... 9

 5.5 分享与协作..... 11

 5.6 软件安全管理..... 12

6 非功能需求..... 14

 6.1 开放兼容性..... 14

 6.2 性能..... 14

 6.3 扩展性..... 14

 6.4 集成性..... 14

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》的规定起草。

本文件由广州思迈特软件有限公司提出。

本文件由广东省网络空间安全协会归口。

本文件起草单位：广州思迈特软件有限公司、网安联认证服务有限公司、广东新兴国家网络安全和信息化发展研究院、广州华南信息安全测评中心、广东关键信息基础设施保护中心、网安联信息技术有限公司。

本文件主要起草人：郑根养、邓庆芳、邢静、孙铭远、赖永红、刘健聪、杜健航、何毅、杨瑾瑜、刘芳、黄珊珊、钟翠英。

本文件为首次发布。

引 言

随着中国的发展，很多技术因为一些原因越来越受制于人，尤其是上游核心技术。为了解决这个问题，我国明确了“数字中国”建设战略，抢占数字经济产业链制高点。于是，国家提出“2+8”安全可控体系，中国 IT 产业从基础硬件-基础软件-行业应用软件有望迎来国产替代潮。这些都是为了实现信创发展的目标：自主可控。

信息技术应用创新发展是目前的一项国家战略，也是当今形势下国家经济发展的新动能。发展信创是为了解决本质安全的问题。本质安全也就是说，现在先把它变成我们自己可掌控、可研究、可发展、可生产的。信创产业发展已经成为经济数字化转型、提升产业链发展的关键，从技术体系引进、强化产业基础、加强保障能力等方面着手，促进信创产业在本地落地生根，带动传统 IT 信息产业转型，构建区域级产业聚集群。

而随着万物“数字化”的浪潮奔腾而来，商业智能与大数据分析软件在组织的数字化布局中也扮演着越来越重要的角色。通过商业智能与大数据分析软件可将“数据”转变成“决策行动”。在信创背景下，国产化的商业智能与大数据分析软件迎来了新一轮的爆发式增长，众多数据分析厂商携手共建我国自主可控的产业生态，商业智能与大数据分析产业标准也将迎来新的定义。

商业智能与大数据分析软件通用技术规范

1 范围

本文件规定了商业智能与大数据分析软件（以下简称“软件”）的支撑环境需求、软件基础软件需求、软件功能需求、非功能需求等方面内容。

本文件适用于信创领域的商业智能与大数据分析软件通用技术规范应用，为企业选购商业智能与大数据分析软件提供软件通用技术规范参考。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的，凡是标注日期的引用文件，仅注明日期的版本适用于本文件。凡是不注明日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 16260 信息技术 软件产品评价 质量特性及其使用指南

GB/T 17544 信息技术软件包质量要求和测试

3 术语和定义

下列术语和定义适用于本文件。

3.1

报表 report

用表格格式来动态显示业务数据，具备固定格式、不规则行列头、多区域报表、支持跨区域公式计算的特点，且以电子化形式展现时支持导出不同的文件格式。

3.2

联机分析处理 OLAP online analytical processing

使分析人员、管理人员或者执行人员能够从多种角度对从原始数据中转化出来的、能够真正为用户所理解的，并真实反映企业多维特性的信息进行快速、一致、交互的存取，从而获得对数据的更深入了解的一类软件技术。

3.3

仪表盘 dashboard

借助表格、图形和地图等可视化元素形成业务综合报告，供用户查看和了解数据中的趋势、异常和状态，从而改进决策效率。电子化的仪表盘具备高度交互化的特性。

3.4

移动 BI mobile business Intelligence

通过移动设备提供报告、仪表盘、可视化数据发现和分析功能。

3.5

数据准备 data preparation

从不同来源数据源中提取数据，进行准确性检查、数据转换和合并整理数据，供软件分析、应用的综合过程。

3.6

数据 ETL 处理 extract-transform-load

将数据从来源端经过抽取（Extract）、转换（Transform）、加载（Load）至目的端的过程。

3.7

元数据管理 metadata management

描述数据属性的数据，为数据存储位置、数据来源、资源查找、文件记录提供支撑，进而达成协助数据检索、提高数据使用效率的目的。

3.8

数据探索 data exploration

具备可视化操作的清单明细数据查询，以及多维度、深层次、自由式的数据分析过程。

3.9

数据目录 data catalog

在安全受控的前提下，开放经过清洗的数据给用户，用户可以在数据目录中快速找到自己需要的数据并提出申请，以获得数据的使用权限。

3.10

信创产业 information technology application innovation industry

即信息技术应用创新产业。信创涉及到的行业包括 IT 基础设施：CPU 芯片、服务器、存储、交换机、路由器、各种云和相关服务内容；基础软件：数据库、操作系统、中间件；应用软件：OA、ERP、办公软件、政务应用、流版签软件；信息安全：边界安全产品、终端安全产品等。

3.11

统一管理平台 unified management platform

系统平稳运行的基础，为系统提供底层服务，其功能需涵盖数据源管理、性能监控管理、元数据管理、权限管理等模块。

3.12

数据源管理 data source management

数据源是管理信息系统、办公自动化系统、决策支持系统等各类信息系统的核心部分，是进行数据分析和决策管理的重要技术手段和基础来源。

3.13

权限管理 authority management

根据系统设置的安全规则或者安全策略，用户只能访问自己被授权的资源和进行权限范围内的操作。

3.14

虚拟语义模型 virtual semantic model

按照业务逻辑对原始数据库里面的物理表进行关联，将数据库物理表、字段、数据格式，封装重组成为面向业务的逻辑视图，方便用户实现数据分析。

3.15

分布式计算 distributed computing

由多台计算机分别计算，再上传运算结果后，将结果统一合并得出数据结论的科学。分布式系统的例子来自有所不同的面向服务的架构、对等网络应用。

3.16

自然语言分析 natural language analysis

自然语言处理在数据分析领域的具体应用体现，是人工智能和语言学领域的分支学科。除了自然语言处理的认知、理解、生成等部分之外，自然语言分析最主要的作用是在数据分析场景，结合人工智能和语音技术，实现使用者完成数据分析零门槛要求的主要技术手段。

3.17

应用商店 app store

应用程序商店（通常简称应用商店）是应用程序的数字发行平台。本文所指应用商店是指在软件中将数据分析成果封装成应用，并通过软件实现分发的功能模块。

3.18

平滑加权轮询算法 smooth weight round robin balancing

一种网络调度算法，在一段时间内，不同的节点服务器被选择的次数分布和权重一致，调度算法比较均匀地选择服务器。

3.19

MDX 函数 MDX function

MDX 语句（MultiDimensional Expressions）是一种语言，支持多维对象与数据的定义和操作。它可以表达在线分析出来数据卡上的选择、计算和一些元数据定义等操作，并赋予用户表现查询结果的能力。MDX 函数就是使用 MDX 语法中的函数。函数可用于任何有效的 MDX 语句，并且用于查询、自定义汇总定义和其他计算。

3.20

事实表 the fact table

即为事实数据表的简称。主要特点是含有大量的数据，并且这些数据是可以汇总，并被记录的。

3.21

星型模型 star schema

星型模式是多维的数据关系，它由事实表（Fact Table）和维度表（Dimension Table）组成。每个维度表中都会有一个维度作为主键，所有这些维度的主键结合成事实表的主键。事实表的非主键属性称为事实，它们一般都是数值或其他可以进行计算的数据。

3.22

雪花模型 snowflake model

当有一个或多个维度表没有直接连接到事实表上，而是通过其他维度表连接到事实表上时，其图解就像多个雪花连接在一起，故称雪花模型。雪花模型是对星型模型的扩展。

4 支撑环境需求

商业智能与大数据分析软件部署所依赖的操作系统、数据库、中间件服务器等支撑环境，需要与国内信创产业需求高度贴合，充分体现核心信息技术产品和关键服务完整的可控性与自主性。

4.1 操作系统

软件应符合信创国产化的操作系统需求，支持部署和应用在多种不同的操作系统上。

4.2 数据库

软件的元数据应支持存放在国产化的数据库上。

4.3 中间件

软件部署依赖的中间件需支持国产化中间件。

5 软件功能需求

5.1 统一平台管理

5.1.1 数据源管理

系统应具备通过图形化界面配置数据源链接的管理功能，使用户能够直接读取公共或私有数据环境中的结构化和非结构化数据。支持数据源类型包括但不限于：

- a) 应支持连接国产的结构化数据库；
- b) 应支持连接文本数据源，能够直接读取 CSV、TXT、数据分析包等数据，加载到系统中进行分析，无须预先定义模板；
- c) 应支持开源的 NoSQL 数据库。

5.1.2 性能监控与管理

系统应提供性能监控的分析能力，能够跟踪和监控系统运行情况，分析及排查系统问题，性能监控与管理内容包括但不限于：

- a) 应提供统一的系统监控界面，方便系统管理员进行系统运行监控与问题排查，包括线程、缓存、堆打印、实验监控、服务监控、会话、网络、内存、SQL 监控等能力；
- b) 应提供耗时分析功能，报表可进行耗时分析，能查看服务器总耗时时间、SQL 执行时间、数据读取时间、服务器处理时间、数据集行数，并可对相关时间进行详情查看。

5.1.3 元数据管理

系统应具备对元数据的管理能力，包括但不限于以下功能：

- a) 应提供元数据搜索功能，能够搜索系统内数据源、数据集、参数等所有系统资源，并定位资源所在路径；
- b) 应提供对资源进行影响性分析和血缘分析的功能，支持筛选查找要分析的数据类型，如：参数、字段、图形主题等。

5.1.4 权限管理

系统应具备数据权限、资源权限、操作权限的管理能力，包括但不限于以下功能：

- a) 需提供细粒度权限控制的功能，能够将目录或报表授权给用户、用户组、角色等，实现查看、编辑、删除、再授权等不同粒度的权限控制；
- b) 需提供批量授权的功能，报表资源权限授权用户后，直接对用户进行该报表资源的引用资源的一键批量授权，从而减轻授权工作量；
- c) 需提供权限继承的功能，能够将下级目录或报表可自动继承上级目录的授权设置，无需逐个目录或报表进行授权，权限继承的方式可配置。

5.2 数据管理

应提供数据管理能力，满足用户进行数据准备的各种需求场景。

5.2.1 数据目录

应提供数据目录功能，在数据安全受控前提下，开放系统经过清洗后的表、数据模型等数据资源通过数据目录开放给用户，用户可以在数据目录中快速找到自己需要的数据并提出申请，以获得数据的使用权限。具体需求应包括以下功能：

- a) 应提供数据快速定位的功能，分析用户可以在数据目录中通过全文检索的方式快速查找软件内部有哪些数据可以使用；
- b) 应提供数据快查的功能，需提供数据结构、数据脱敏预览、数据溯源分析等基础的能力，用户可以直接查看权限范围内的数据，对于没有权限的数据，软件可实现数据脱敏处理，以保障数据的安全；
- c) 应提供在线数据申请授权的功能，对于未授权的数据，用户可以自主申请授权，审核通过并授权后即可使用需要的数据。

5.2.2 数据 ETL 处理

应提供数据 ETL 处理能力来满足数据分析之前的清洗、转换、加载等不同数据准备需求，保障数据的质量。应包括但不限于以下功能：

- a) 基于 Web 的可视化 ETL 作业流程配置：

应具备可视化流式操作界面，能够在界面上直接配置 ETL 转换作业流，业务人员能够在 Web 端通过拖拉拽的方式完成配置等操作，支持对作业流的运行状况实现调度监控；

b) 数据采样：

应支持数据进行随机采样、加权采样、分层采样，下采样，并且可选择按个数、比例进行采样；

c) 丰富的数据处理组件：

应具备丰富的数据处理组件，包括数据拆分、字段过滤与映射、列选择、过滤空值、合并列、合并行、连接、元数据编辑、行选择、去除重复值、排序、增加序列号、聚合、分列、派生列、类型转换、行列转换等操作；

d) 自定义脚本：

应支持自定义编写数据处理脚本，能够轻松应对复杂业务场景；

e) 支持大数据量表关联：

在超大数据量分析场景下，可支持对两个亿级数据表进行表关联计算。

5.2.3 虚拟语义模型

软件需要提供虚拟语义模型，应提供虚拟语义层，构建业务主题视图，让用户灵活、自助式地获取所需要的业务数据，从而快速辅助决策，包括但不限于以下功能：

- a) 应支持将数据库中的表字段实现拖放重组，形成面向业务分析的逻辑主题，并支持在主题中对表字段进行时间层次、转换规则、参与分组等设置；
- b) 应支持跨多种不同的数据库建立语义模型；
- c) 应支持对指标设置默认的浏览格式（如百分比、万元）和聚合方式（合计、计数等），在分析展现浏览数据时该字段应可自动以该格式展现，无需逐个设置；
- d) 应支持多表关联设置，支持左连接、右连接、内连接、外连接、全连接方式，并支持按列名、按别名、按外键自动创建表关系，支持一键清除表关系。

5.2.4 数据模型

系统应提供数据建模能力，应对不同来源的数据关联和聚合分析的场景。

a) 多源组合能力：

——应提供多种查询融合的手段，如：数据源表、可视化拖放的数据查询、文件数据查询、SQL 查询、存储过程查询、脚本查询等，并可实现不同查询手段之间相互关联；

——应支持跨源跨库能力，能够连接不同种类、不同类型的数据库进行关联分析；

——应支持模型关联能力，支持多张数据表以常见的雪花模型或星型模型方式关联，也能满足多事实表的数据模型关联；

——支持各种查询手段与数据 ETL 处理打通，通过二次数据处理，来满足更加多样的数据处理场景。

b) 支持多维计算能力：支持新建计算度量、计算成员、新建命名集和手工编写 MDX 函数满足复杂的计算；

c) 支持数据抽取功能：应支持直连和抽取模式，其中抽取模式可基于数据模型进行数据增量或者全量抽取，数据可抽取到软件内置的数据缓存库。

5.3 数据分析

应提供多种数据分析工具如数据查询、报表报告、可视化仪表盘等，来满足不同用户的数据分析需求。

5.3.1 报表

应支持编制与展现业务报表的功能，通过表格、图形等方式来动态显示数据，它对数据按一定的表格形式进行汇总反映和综合反映，其功能需包括但不限于：

- a) 应支持直接在 Web 上开发设计报表，以便更好地支持国产系统和软件；
- b) 应支持各种计算格式复杂的中国式报表，如不规则报表、交叉报表、多区域报表、分片报表等；
- c) 应兼容日常办公软件的表格函数。

5.3.2 数据探索

应支持数据探索的功能，满足用户灵活地查询数据和自动数据汇总统计的需求。

5.3.2.1 即席查询

应支持即席查询功能，让用户根据自己的需求，灵活地选择查询条件，系统能够根据用户的选择生成相应的统计报表。其功能需包括但不限于：

- a) 应支持条件筛选，可以选择任意输出字段也可以将任意字段作为筛选条件；
- b) 应支持灵活更改查询运算符，灵活改变查询条件间组合关系和顺序，还能根据字段类型智能显示参数控件类型；
- c) 应支持大数据量的批量导出；
- d) 可支持加密导出和导出后的数据分析包的导入，能够将报表导出为加密的数据分析包，数据分析包可以导入回软件进行数据分析。

5.3.2.2 透视分析

应支持透视分析的功能，通过交互式的方式进行多维度统计分析，并且进行计算，如求和与计数等。应提供如下功能：

- a) 应支持组合、汇总计算，不需要预先构建多维模型，任何字段都可直接作为输出字段或筛选条件；
- b) 应支持多种分析方法，如切片、钻取、旋转、行列互换、焦点项等；
- c) 应支持多个维度分析，如 50 维度、甚至维度无法固定的分析场景；
- d) 应支持通过自定义计算字段，扩展可分析的维度和度量；
- e) 应支持超大数据量的查询性能，亿级数据秒级响应；
- f) 应支持自定义时间维度，并支持多种时间计算选项，如：同期值、同期增长率、前期值、环比、环比增长率、期初期末值、日均、累计值、增长值等。

5.3.2.3 多维分析

应支持多维分析功能，使不同角色的用户可基于自身所关注的维度，进行灵活快捷、多维度的交互数据分析，从而使用户对大量复杂数据的分析变得轻松而高效，以便于迅速做出正确的判断，辅助商业决策。其功能需包括如下功能：

- a) 多维分析：

支持用户按指定维度层次进行汇总并呈现成表格或图形，并在表格或者图形上进行钻取、聚合、切块、行列互换、分级排序等分析操作，以供用户灵活地操纵数据，以多维的形式从多层次和多角度来观察数据状态与变化。

——钻取：向上钻取、向下钻取、层钻操作；

——聚合：常用的汇总方法，如合计、平均值、最大值和最小值；

- 切块:用户能够把维度放到多维分析切块区域, 实现在多个维上集中观察数据;
 - 行列互换:用户能够将行上所有的维与列上所有的维按照原有顺序进行整体调换, 从而得到不同视角的数据;
 - 分级排序:用户根据某数据列对表格区中的数据重新进行排列的操作;
 - 多维报表的二次修改: 设计好的多维分析报表仍然能够继续对维度和度量等进行调整, 作进一步探索。
- b) 自定义计算:
- 基于已有成员通过算术运算符、数字和 MDX 函数进行重新组合生成新的指标, 或者为重复使用而创建的维度成员集合, 将其作为分析的一部分。

5.3.3 仪表盘

应具备仪表盘能力, 来满足数据报告的需求。应具备聚合并显示各种指标, 使所有用户都能一目了然地检查数据的能力。

5.3.3.1 交互式仪表盘

应具备交互式可视化仪表盘能力, 将数据转换成图形或图像在 PC、移动设备或大屏幕上显示出来, 用户可直接在图形或图像上进行交互式操作。支持的功能包括但不限于:

- a) 应支持各种各样的图形, 并能可视化设置图形选项。图形类型包括但是不局限于柱图、线图、饼图、瀑布图、旭日图、关系图、雷达图、油量图、热力图、树图、桑基图、漏斗图、极坐标柱图、极坐标线图等图形, 还能支持迷你图、帕累托图、子弹图等特色图形, 支持 3D 航线图、3D 散点图、3D 柱图等 3D 图形;
- b) 应支持地理图形展现。地图包括但是不局限于中国、及各省行政区域地图, 内置支持国产地图, 支持自定义地图; 地图还应该包括多种地图样式供选择, 包括但不局限于区域地图、散点地图、热力地图和航线图等;
- c) 图形之间应支持各种交互操作, 包括但不局限于联动、跳转、钻取等;
- d) 应具备较好的扩展能力, 能够扩展第三方图形;
- e) 应支持多终端展示, 包括但不局限于 PC、手机、平板、大屏设备等, 还应支持单屏和多屏展示, 可以根据不同的分辨率进行自适应;
- f) 可支持人工智能交互技术扩充, 用户可以通过语音命令或手动输入提问, 即可在 PC 端、移动端进行数据分析。

5.3.3.2 自助配置的仪表盘

应提供用户拖放式的操作完成仪表盘的新建和配置的功能。自助仪表盘应包括但不限于以下功能:

- a) 应支持可视化拖拽操作, 简单易用; 支持所见即所得的方式完成仪表盘的创建;
- b) 应支持各种可视化组件, 包括但不局限于柱图、瀑布图、面积图、雷达图、地图、Tab 页组件和 URL 链接组件等;
- c) 应支持多种主题样式, 包括但不局限仪表盘主题、图形主题、组件主题等; 配色样式包括深色风格、浅色风格等, 主题和样式都支持自定义;
- d) 应支持多种布局方式, 包括但不局限于高宽自适应、宽度自适应、自由布局。布局还支持拖拽实现资源自由组合、缩放大小;
- e) 应支持多种终端的适配, 创建的成果支持发布到电脑、手机、平板, 可以根据不同的分辨率实现自适应;
- f) 应支持多种交互效果, 包括联动、钻取、跳转等。

5.3.4 移动 BI

应提供移动 BI 能力，通过移动设备（平板电脑和智能手机）提供报告、仪表盘、可视化数据发现和基本分析功能。通过在较小的屏幕上呈现 BI 内容，可以通过基于触摸的界面、查询远程或本地数据实现导航。移动 BI 应包括但不限于以下功能：

- a) 应支持 HTML5 标准，能够与第三方应用程序如微信、钉钉的集成；
- b) 应支持报表分享，可以匿名将报表分享到第三方国产化社交工具；
- c) 应支持在离线状态下可以进行交互分析，包括钻取、跳转、改变查询条件等；
- d) 支持离线数据包推送以实现离线分析的能力，用户接收后可以增量更新离线数据包；
- e) 应提供数据安全保障的访问机制，包括数据权限控制、访问终端控制、访问资源权限控制等。

5.4 增强分析

5.4.1 预测分析

应具备预测分析能力，其功能需涵盖可视化建模、特征处理、预测算法、模型评估、模型部署、自动建模等，且可采用大规模分布式的计算引擎实现高效的模型训练。

a) 可视化建模：

可视化建模是指通过图形化的操作方式，将复杂的建模过程转变为对用户更为友好且可理解的界面，降低业务人员的使用门槛。在预测分析模块中，提供可视化建模方式，具备无需编程技术即可自定义建模的能力，包括但不限于：

- 应支持将数据建模过程各个步骤的核心能力封装为独立的节点组件，每个节点可以自定义参数；
- 应提供直观的流式建模方式，通过拖放的方式快速构建机器学习模型；
- 应支持启动/停止任务，支持断点重跑（如果任务因故停止，可在停止处重启任务）机制；
- 应支持查看运行结果（如：数据、分析报告、保存模型等）、运行记录及运行日志。

b) 特征处理：

预测分析模块应具备特征处理能力，能够最大限度地从原始数据中提取特征，以供算法和模型使用。特征处理能力包括但不限于：

- 对于各种类型特征，应可对其编码；
- 应支持使用特征离散化算法将数据自动进行分桶和离散化，从而得到离散化非线性的特征指标；
- 分析样本表中每个特征和目标值之间的重要性关系，通过图形或者列表的方式展示特征对预测目标值的重要性系数，以此分析特征的异常现象，从而对模型特征进行调整优化；
- 应提供多种消除特征之间的量纲影响的数据归一化方法；
- 应支持采用某种映射方法，将高维空间中的数据点映射到低维度的空间以达到数据降维；
- 应支持对一份数据表做各种特征组合分析，自动快速探索并生成优化后的组合特征并评估生成的特征的重要性。

c) 预测算法：

预测分析模块应具备多种机器学习算法，无需输入任何与问题有关的特定代码，将数据输入泛型算法当中，在数据的基础上建立对应的逻辑，即可输出期望的结果。预测算法包括但不限于：

- 向量机、梯度提升决策树等二分类算法；
- 逻辑回归、朴素贝叶斯、决策树、多层感知机、随机森林等多分类算法；
- 线性回归、梯度提升回归树等回归算法；
- K 均值、高斯混合模型等聚类算法。

d) 模型评估：

预测分析模块应具备对于使用不同算法构建的模型，提供可视化的对比分析，包括算法训练性能对比、超参数对比、特征变量对比、评估指标对比等。对于每个模型可以一键生成模型评估报告，评估报告中需要包含特征变量、评估指标等内容，且能够根据算法类型提供不同的评估指标，模型评估指标包括但不限于：

- 应支持混淆矩阵、准确率、精确率、召回率、F1 值、假阳率、假阴率、特异度、ROC 曲线、AUC 值、KS 曲线等分类评估指标；
- 应支持平均绝对误差、均方误差、均方根误差、R 方、校正 R 方等回归评估指标；
- 可支持 Calinski Harabasz 指标、和方差、轮廓系数等聚类评估指标。

e) 模型部署：

预测分析模块应具备通过提供 REST 或者其他形式的接口的方式，将持久化后的模型对外发布，并提供预测服务能力。模型部署包括但不限于：

- 应支持模型保存、模型导入、模型导出的功能；
- 应支持服务发布上线/下线；
- 应支持灰度测试功能，即多版本模型并行运行；
- 应支持批量调用，可以通过计划任务的方式配置模型调用的频次和有效时间段。

f) 自动建模：

在预测分析模块中，建议可提供自动建模能力，使得机器学习模型无需人工干预即可被应用，包括但不限于：

- 可支持输入数据、配置相关选项后，即可完成数据预处理、特征工程、模型构建、超参数优化、模型评估等一系列数据挖掘的环节；
- 可支持自动将最优模型进行服务部署和模型自学习，轻松实现从建模到生产以及模型运维整个生命周期的全自动化和智能化学习，在帮助用户提升建模的效率同时保证模型的质量；
- 可支持自动化建模完成后可生成建模流程图，用户可对流程图进行修改，达到进一步提升模型质量的效果。

g) 分布式计算：

分布式计算是一种把需要进行大量计算的工程数据分割成小块，由多台计算机分别计算，在上传运算结果后，将结果统一合并得出数据结论的技术。预测分析模块需采用分布式计算，实现平台高效的模型训练。功能点包括但不限于：

- 应支持动态扩展，可随时加入新的计算节点；
- 应支持节点资源分配，可以指定节点上的资源使用如 CPU、内存；
- 应支持节点资源监控，具有监控界面能实时查看节点使用情况。

5.4.2 自然语言分析

系统中的自然语言分析应通过自然语言与数据分析软件功能的有机结合，将自然语言转换为系统指令，对系统中的数据进行查询、分析、展现等操作，支持多种终端适用。

a) 自然语言指令解析：

自然语言指令解析其功能应包括但不限于：

- 应提供资源搜索功能，能够对数据指标、报表、功能实现定位；
- 应提供数据查询功能，能够提取自然语言中关键信息，生成查询定义，根据数据库的实际情况生成查询语句，获取数据。

b) 指令执行：

具备指令执行能力，通过自然语言完成标准化程度较高的操作，包括但不限于：

- 应支持自助报表生成，通过自然语言指导系统中对查询结果的组合，完成报表制作；
- 应支持报表查询访问，通过自然语言检索系统中的数据并完成查询与分析；
- 应支持参数切换，通过自然语言切换报表与数据分析过程中的参数；

——应支持多轮对话/会话式分析。

应提供多轮对话能力，当系统识别用户的评估存在疑惑时，则提供正向反馈（如确认等行为）、负向反馈（如拒绝理解/要求重复），要求反馈、修复请求等响应机制。为提供更贴近用户真实意图的服务，多轮对话应包括但不限于以下功能：

- 应提供会话式分析功能，终端以会话界面进行系统数据展示，根据输入的自然语言指令，输出系统反馈的结果；
- 应提供数据探索分析功能，根据用户自然语言指令，不断对数据进行加工处理，处理方式根据自然语言解析结果选择，进行数据探索分析过程。

c) API 接口：

自然语言 API 接口主要对指令解析、指令执行、多轮对话进行封装，并提供一些预先定义的函数，目的是提供应用程序与开发人员基于某软件或硬件得以访问一组例程的能力，而又无需访问源码，或理解内部工作机制的细节，可以在第三方系统中使用上述功能，API 接口包括但不限于以下：

- 登入接口：调用开始，对调用方身份进行识别，授权使用功能；
- 登出接口：功能调用结束，关闭授权；
- 语音转文本接口：当输入系统指令为语音时，需要将语音文件转化为文本供系统识别分析；
- 自然语言查询接口：对输入的自然语言文本进行语义解析，进行系统指令归类，输出触发执行系统指令的命令；
- 关闭查询接口：根据语义分析结果，生成触发相应的系统指令，返回结果后，关闭查询。

5.5 分享与协作

5.5.1 应用门户

应用门户应能将数据分析资源、入口和相关应用等内容，集中地组织起来，以便作为统一的浏览入口。应用门户还可以针对每个用户进行个性化设置，这样有利于对信息的治理、传播和信任。应用门户包括但不限于以下：

- a) 应支持显示在不同的终端，包括但不局限于 PC 端、移动端；
- b) 应满足不同用户的个性化要求。门户内容包括但不局限于用户关注的分析资源、常用功能、关注的分析指标、外部系统的快捷入口等，门户还应该支持布局、配色等个性化需要；
- c) 应支持丰富的门户组件，包括收藏夹、卡片列表、应用商店、便捷入口等，操作上通过可视化拖拽、可视化设置实现；
- d) 应支持提供完备的门户组件接口，供第三方信息接入。

5.5.2 消息中心

系统应提供消息通知功能，包括预警、推荐、公告、待办、指向、私信等，消息中心可支持系统内消息的接收、存储、消息状态报告、推送等功能。

- a) 应根据来源和应用范围支持不同类型的消息，包括系统消息、数据预警、审核信息、企业公告信息、个人私信、评论行为提醒等；
- b) 消息的推送应该具有实时性，支持显示信息的发布人和发布时间等信息；
- c) 消息查看应该具有便捷性，入口明显，点击信息即可打开查看详细内容等；
- d) 消息中心应支持状态标记，如标记已读未读，未读消息支持高亮显示；
- e) 消息中心应该支持统一管理，包括编辑、重发、删除等操作。

5.5.3 应用商店

数据应用商店应支持如下功能：

- a) 支持将分析结果发布成应用；
- b) 内置发布流程，应用发布成功通过消息中心告知平台所有用户；
- c) 商店对应用上架、应用下架、应用重新上架等管理操作进行审核；
- d) 支持在“应用商店”搜索关键字查找自己想要的应用，或者按照安装状态、应用发布者分类查询；
- e) 可根据发布时间、访问量、下载量、评论数对应用进行排序；
- f) 支持通过权限申请将应用安装到“已安装应用”中进行使用。

5.5.4 互动交流

应支持在平台内进行用户的互动交流：

- a) 支持对分析、应用、问答等进行点赞、评论；
- b) 系统支持用户自定义自己的头像和个人信息；
- c) 支持对优秀的分析成果进行分享：

——对外可通过链接、二维码等形式进行分享，并可设置访问密码及访问截止日期，来提供安全的访问服务；

——对内可分享给指定组、用户、角色，控制浏览和再分享权限，限制访问截止日期。

5.6 软件安全管理

针对软件系统安全提供完整的解决方案，包括管理应用的用户访问和身份验证、数据的安全及备份、安全管理措施等方面内容。

5.6.1 应用安全

5.6.1.1 身份鉴别

身份鉴别是对登录用户进行身份标识和鉴别，保障系统的安全。身份鉴别应包括但不限于以下功能：

- a) 应支持在系统前端界面自定义用户名校验规则和用户名校验错误提示；
- b) 应支持自定义用户口令复杂度校验规则，用户名校验错误应有提示；
- c) 可支持设置用户口令的复杂度，防止暴力破解；
- d) 可支持国产化的加密标准。

5.6.1.2 访问控制

应提供访问控制系统功能操作、系统资源，控制用户组/用户/角色对系统功能和用户数据的访问；并且应由授权主体来配置访问策略，能严格限制用户的权限。访问控制的具体要求包括：

- a) 应支持将目录或报表授权给用户、用户组、角色等，实现查看、编辑、删除、再授权等不同粒度的操作权限控制；
- b) 下级目录或报表应支持自动继承上级目录的授权设置，无需逐个目录或报表进行授权，并且应支持直接勾选/取消等简易操作方式配置权限继承；
- c) 移动设备应支持访问限制，对注册服务平台的移动设备管理，如：审核、启用、禁用等。

5.6.1.3 安全审计

通过安全审计收集、分析、评估系统的用户行为信息、系统资源异常使用等信息，才能实现掌握安全的状况，从而制定行之有效的安全策略，确保整个安全体系的完备性、合理性和适用性，才能将系统调整到“最安全”和“最低风险”的状态。要求包括但不限于：

- a) 应提供全面记录用户的各种操作信息的记录表，包含访问时间、用户、操作类型、详细操作描述等信息；
- b) 应提供内存告警日志，包含告警的资源名称、类型、告警时间、告警原因等；
- c) 应能够根据记录数据进行二次分析，并生成审计报表。

5.6.1.4 软件容错性

系统在运行中应具备较好的从错误中恢复的能力，在故障存在的情况下系统不失效，仍然能够正常工作。本项要求包括但不限于：

- a) 可具备国产化应用代理分发器，系统可自行依据应用节点 CPU、GC (Garbage Collection) 时间独立进行分发，可在图形化界面进行节点配置，负载均衡采用平滑加权轮询算法保证系统稳定性；
- b) 应支持应用服务器共享会话信息，无论是否宕机、切换服务器，用户无须重新登录，保障业务操作不中断、数据/模板不丢失；
- c) 应支持一键进行系统全面检查，包括系统的完整性和正确性，提示存在的问题并进行修复。

5.6.1.5 资源控制

应能够对系统的会话进行控制与分析，能够一键进行批量或单个会话进行断开。提供针对不同时间范围内的会话情况、内存情况进行分析。

5.6.1.6 安全监控

系统应具备可视化、统一的系统监控界面，方便系统管理员进行系统运行监控与问题排查，包括进行线程、缓存、堆打印、实验监控、服务监控、会话、网络、内存、SQL 监控等分析。

5.6.1.7 安全补丁更新

应支持安全补丁热修复，即可在不重启服务器的情况下，对系统当前软件版本的缺陷进行修复。安全补丁更新方式包括在线更新和手动更新。

5.6.2 数据安全及备份恢复

5.6.2.1 数据脱敏

在界面配置实现数据脱敏功能，应支持掩码、字符替换、正则替换等脱敏算法，应支持指定不同用户组、用户、角色生效，并且可控制在浏览或者导出时候生效。

5.6.2.2 备份恢复

采用冗余技术设计系统拓扑结构，避免关键节点存在单点故障。应提供系统级别的一键备份与恢复机制，可自动备份；支持资源级别的备份与恢复，并且在导出资源时可一键添加依赖资源、资源权限、数据权限，实现资源在不同环境（如开发环境、测试环境和生产环境）之间的快速迁移。

5.6.3 安全管理要求

应制定系统的安全策略、安全管理制度和操作流程规范等一体化的信息安全管理制度体系。从事软件相关关键岗位的人员，需签署岗位安全协议，并进行严格的安全意识培训和安全审查。

6 非功能需求

6.1 开放兼容性

系统应具备良好的开放兼容性，能与信创生态体系里面的国产化主流产品兼容，实现自主可控。

6.2 性能

系统应采用标准且先进成熟的技术手段，应具有较高性能及较强的生命力，满足未来不断发展的业务需求。即使在高用户负载、长时间不间断的情况下系统仍然能够平稳运行。

有关性能指标的建议如下：

- a) 系统应具备良好的稳定性，应支持连续 7×24 小时不间断工作；
- b) 建议在单节点 8 核 CPU、32GB 内存配置下，可支持不少于 200 并发用户量访问，并且登录平均耗时应小于 1 秒；
- c) 建议在单节点 8 核 CPU、32GB 内存配置下，100 并发用户查询亿级清单表，选择 10 个维度 10 个指标在线清单数据查看，平均响应时间应小于 5 秒；
- d) 建议在单节点 8 核 CPU、32GB 内存配置下，100 并发用户操作使用亿级宽表数据进行 3 维度、4 指标交叉表汇总计算，平均响应时间应小于 5 秒；
- e) 建议在单节点 8 核 CPU、32GB 内存配置下，两个千万级表（10 维度、5 指标），关联查询结果，平均响应时间应小于 5 分钟。

6.3 扩展性

系统应具备良好的扩展能力，能应对业务不断发展带来的业务需求和系统架构的挑战。

6.3.1 系统横向扩展能力

系统应具备集群横向扩展能力，可通过集群部署扩展应用节点。集群可以利用多个计算机进行并行计算从而提升系统的负载能力。

- a) 应支持通过分发策略将用户请求分派到不同的应用节点上，以此实现负载均衡；
- b) 应支持集群部署，如果某个节点发生故障，则能够自动将当前会话切换到其它正常的节点上继续提供服务；
- c) 应提供故障管理和控制的机制，故障切换组件时使用冗余系统和数据来保证关键任务数据流不间断；
- d) 负载均衡分发器应支持图形可视化操作方式扩展集群节点，不需要手工进行额外的文件转移。

6.3.2 功能扩展

应提供宏代码、功能扩展包等插件方式灵活实现功能扩展，满足客户个性化需求。

6.4 集成性

系统按照一定开放标准对外提供 API、SDK，将分析的内容可以更好的嵌入到业务平台，包括用户集成、报表集成、二次开发接口等常用的技术手段。本项要求包括但不限于：

- a) 应提供二次开发接口，并有完整的中文说明文档；
- b) 应支持功能模块集成能力，在第三方系统内整合模块功能；
- c) 应支持用户自行通过 JS 脚本扩展或插件包扩展方式实现定制开发，不直接修改产品源代码，保证系统的稳定性与升级的便利性。