ICS xxxxxx

P xxx

|  |
| --- |
|  |

团体标准

T/xxx xxx—2021

|  |
| --- |
|  |

商业智能与大数据分析软件

business intelligence and big data analysis software

|  |
| --- |
|  |
|  |

2021 - ×× - ××发布

2020 - ×× - ××实施

广东省网络空间安全协会

信息技术创新联盟

发布

目 次

[前 言 III](#_Toc2910)

[1 范围 1](#_Toc19142)

[2 规范性引用文件 1](#_Toc12582)

[3 术语和定义 1](#_Toc22061)

[4 软件的概述 2](#_Toc971)

[5 基础软件需求 3](#_Toc3955)

[5.1 操作系统 3](#_Toc24887)

[5.2 数据库 3](#_Toc7170)

[5.3 中间件 3](#_Toc18509)

[6 软件功能需求 3](#_Toc9879)

[6.1 统一平台管理 3](#_Toc13993)

[6.2 数据管理 5](#_Toc18512)

[6.3 数据分析 9](#_Toc17843)

[6.4 增强分析 14](#_Toc3762)

[6.5 分享与协作 18](#_Toc10159)

[7 非功能需求 19](#_Toc6778)

[7.1 软件安全管理 19](#_Toc31687)

[7.2 开放兼容性 21](#_Toc4608)

[7.3 性能 21](#_Toc11109)

[7.4 拓展性 22](#_Toc27573)

[7.5 集成性 22](#_Toc25403)

[8 验收要求 22](#_Toc31100)

[8.1 测试标准 22](#_Toc3500)

[8.2 验收标准 23](#_Toc9206)

[8.3 其它要求 24](#_Toc8569)

[9 售后服务 24](#_Toc31989)

前 言

本文件按照GB/T 1.1-2020《标准化工作导则 第1部分：标准的结构和编写》给出的规则起草。

本文件由广州思迈特软件有限公司提出。

本文件起草单位：广州思迈特软件有限公司

本文件主要起草人：广州思迈特软件有限公司

本文件首次发布。

商业智能与大数据分析软件

# 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”）的软件定义、基础软件需求、软件功能需求、非功能需求。

本文件适用于商业智能与大数据分析软件。

# 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，凡是注日期的引用文件，仅注日期的版本适用于本文件。凡是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。

GB/T 26232 基于J2EE的应用服务器技术规范

GB/T 17544 信息技术软件包质量要求和测试

GB/T 16260 软件质量模型与度量介绍

# 术语和定义

下列术语和定义适用于本文件。

## 

## 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，且支持导出不同的文件格式。

## 

## 联机分析处理 OLAP

OLAP技术基于MDX查询语言，主要通过多维的方式来对数据进行分析、查询和生成报表。OLAP系统的应用主要是对用户当前的及历史数据进行分析，辅助决策。

## 

## 仪表板 Dashboard

借助图表、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，具备高度交互化。

## 

## 移动BI Mobile Business Intelligence

指通过移动设备提供报告、仪表板、可视化数据发现和基本分析功能。

## 

## 数据准备 Data preparation

数据准备一个为了使数据可用于分析为目的，而进行迭代和敏捷的过程，用于探索、组合、清理原始数据并将其转换为自助数据集成。

## 

## 数据处理 ETL

ETL是英文Extract-Transform-Load的缩写，用来描述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。

## 

## 元数据管理 Metadata management

用于描述数据属性的信息，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。

## 

## 数据探索 Data exploration

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

## 

## 数据目录 Data catalog

在安全受控的前提下，开放组织内经过清洗的数据给用户，业务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。

## 

## 信创产业Information technology application innovation industry

信创产业，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。

# 软件的概述

商业智能与大数据分析软件指在一个安全可控的平台上将企业现有的数据转化为知识，帮助企业做出明智的业务经营决策，实现从“数据”到“决策行动”。主要包括以下4类分析：

## 描述性分析

解决“发生了什么”的问题。通常由IT主导，基于预处理好的数据模型进行交互式的KPI和报表的创建，从而让管理者浏览查阅，帮助其做出分析决策。

## 诊断性分析

解决“为什么发生”的问题。由业务分析师主导，通过可视化的探索平台，进行自由的用户交互操作，创建用于监控和分析的仪表板，从而解决诊断性、探索性的问题。

## 预测性分析

解决“将来会怎么样”的问题。由业务分析师和消费者主导，基于开放式问题和结构化数据进行增强分析。包括增加了机器学习ML用于解决预测性的问题，增加了自然语言分析NLA用于人机交互等。

## 指示性分析

解决“我应该怎么做”的问题。通常面向消费者，进行数据的增强消费，解决指示性的问题。主要体现在：以机器学习为主导在上下文中为用户连续自动生成描述性，诊断性，预测性，说明性见解；以故事或新闻馈送的形式或可固定到仪表板的动态仪表板的形式提供洞察力。

# 基础软件需求

基础软件是指部署于运行商业智能与数据分析软件系统的基本依赖软件或应用，包含操作系统、数据库（存放系统元数据）、中间件等模块。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。

## 操作系统

系统应支持部署和在多种不同的操作系统上使用，包括但不限于基于Linux、Unix等系统平台。

## 数据库

系统的元数据应支持存放在国产化的数据库上。

## 中间件

中间件是介于应用系统和[系统软件](https://baike.baidu.com/item/%E7%B3%BB%E7%BB%9F%E8%BD%AF%E4%BB%B6/215962)之间的一类软件，它使用系统软件所提供的基础服务，衔接应用系统的各个部分或不同的应用，能够达到资源共享、功能共享的目的。需支持国产化中间件。

# 软件功能需求

## 统一平台管理

统一管理平台是系统平稳运行的基础，主要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。

### 数据源管理

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。

系统应具备界面化配置式的数据源管理能力，使用户能够连接和摄取本地和云中各种类型的存储数据库中包含的结构化和非结构化数据的功能使用户能够连接和提取本地和云中各种类型的存储数据库中包含的结构化和非结构化数据的功能。支持数据源包括但不限于：

* 1. 应支持国产数据库，且具备自主可控性。
  2. 应支持文本数据源，能够直接读取CSV、TXT等数据、数据分析包加载到系统中进行分析，无须预先定义模板。
  3. 应支持开源的NoSQL数据库。

### 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：

* 1. 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监控等分析显现能力。
  2. 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。

### 元数据管理

元数据是描述数据属性的数据，其主要是描述数据属性的信息，如数据存储位置、数据来源、文件记录等。系统应具备对元数据的管理能力，包括但不限于以下功能：

* 1. 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。
  2. 应提供对资源进行影响性分析和血缘分析的功能。血缘关系和影响分析两个关键特性为用户重建了整个元数据家族的构建过程，刻画了家族成员彼此连接的脉络和途径。比如当报表数据出现错误或者异常时，我们可通过血缘关系图向上分析锁定问题产生的源头（如表和字段）；当对某些字段数据进行修改时，可通过影响关系图向下分析，得到哪些数据实体中的数据会受到影响。

### 权限管理

权限管理指根据系统设置的安全规则或者[安全策略](https://baike.baidu.com/item/%E5%AE%89%E5%85%A8%E7%AD%96%E7%95%A5/2890163" \t "https://baike.baidu.com/item/_blank)，用户可以访问而且只能访问自己被授权的资源。系统应具备数据行权限、资源权限、操作权限的管理能力，包括但不限于以下功能：

* 1. 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。
  2. 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。
  3. 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。

## 数据管理

数据管理是一个为了使数据可用于分析为目的，而进行迭代和敏捷的数据准备过程，用于探索、组合、清理原始数据，并将其转换为BI分析成果来源的各种分析模型。包括但不但于数据目录、数据处理、数据缓存、虚拟语义模型、增强数据集等内容。

### 数据目录

在安全受控的前提下，开放组织内经过清洗的数据给用户，业务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。数据目录应包括但不限于以下功能：

* 1. 应提供数据快速定位的功能，用户可以在数据目录中通过全文检索的方式快速查找有平台内部有哪些数据可以使用。
  2. 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，系统对数据进行随机脱敏和处理，以保障数据的安全。
  3. 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，审核用户审核通过并授权后即可获取需要的数据。

### 数据处理

数据处理（ETL）指的是在数据分析之前对数据进行清洗、转换、加载的过程。源数据通常来自不同的业务系统，它们有不同的数据格式，也包含一些冗余的信息。ETL负责将分散的、异构数据源中的数据如关系数据、平面数据文件等抽取到临时中间层后，进行清洗、转换、集成，最后加载到数据仓库或数据集市中，成为联机分析处理、数据挖掘提供决策支持的数据。ETL数据处理应包括但不限于以下内容：

#### 数据源支持

应支持关系数据源、Kafka流、文本数据文件等多种数据源。

#### 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。

#### 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。

#### 丰富的数据处理组件

应具备丰富的数据处理组件，包括但不限于：数据拆分、字段过滤与映射、列选择、过滤空值、合并列、合并行、JOIN、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。

#### 自定义脚本

应支持自定义编写SQL、Python等数据处理脚本，能够轻松应对一些比较复杂的业务场景。

#### 支持大数据量表关联

应支持对两个亿级数据进行表关联处理，能够应对超大数据量处理。

#### 支持作业流及调度监控

应支持在Web端可视化的配置作业流，对ETL转换流进行可视化组装配置，支持作业流的运行进行调度监控。

### 数据缓存

应提供分布式的数据缓存引擎，用于性能加速，在进行数据分析的时候，可基于数据集进行数据增量或者全量抽取到分布式存储的数据缓存层，解决性能瓶颈。并且需要能够提供多种缓存库可灵活替换抽取方案。

### 虚拟语义模型

虚拟语义模型是指允许用户创建涵盖整个仓库的语义模型，可以按照业务逻辑对原始数据表进行重新建模，将数据库物理表、字段，封装重组成面向业务的逻辑视图，方便业务分析用户进行自助式的数据分析。通过构建主题视图，用户即可灵活、自助式地获取业务上所需要的信息，从而快速辅助决策，包括但不限于以下功能：

* 1. 应支持将数据库中的数据定义成明确有业务含义的名称，业务人员面对的不再是表、字段，而是业务术语。
  2. 应支持将数据库中的表字段进行拖拽重组，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。
  3. 应支持跨多种不同的数据库建立语义模型。
  4. 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。
  5. 应支持多表关联设置，支持左连接、右连接、内连接、全连接等方式，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。

### 增强数据集

系统需要具有数据建模能力，作为数据源和可视化展示的中间环节，承接数据源的输入，并作为可视化展示输出到仪表盘或报表，这通常通过增强数据集功能提供。

在增强数据集中，您可以对数据集关联、转换ETL高级查询，预览数据、抽取设置、自动创建维度表，标记计算度量等操作。

#### 多源组合能力

多源组合能力是指产品在BI建模的过程中，数据集可能需要连接多个数据源以创建报表，而创建不同数据源之间的逻辑连接，需要创建一种关系，即表与表之间的关系，了解表与表之间的关系，让您能够更快速/便捷地读取数据中的关键信息。包含但不限于以下功能：

* 1. 应支持跨源跨库：能够连接不同种类、不同类型的数据库；能够连接SQL数据集、存储过程数据集、脚本查询以及即席查询等二次分析结果
  2. 应支持数据关系管理：支持通过关联表功能，将多张数据表以雪花模型或星型模型方式关联。
  3. 应支持图形化设置：左联、内联、外联、全联。
  4. 应支持二次数据处理分析：能够新建或编辑度量、维度，支持高级ETL转换，将多来源的异构数据进行处理后得到具备完整性、一致性的数据模型，应支持的数据预处理方法包括采样、拆分、过滤与映射、列选择、空值处理、合并列、合并行、元数据编辑、JOIN、行选择、去除重复值、排序、增加序列号、聚合、分列、派生列等；能够直接在界面配置数据脱敏规则实现数据脱敏，并对不同用户组、用户、角色生效；能够直接将文本数据或数据分析包加载到系统中进行分析，无须预先定义模板。
  5. 应支持直连和抽取模式，其中抽取模式可基于数据集进行数据增量或者全量抽取，数据可抽取到国产数据缓存库。

#### 增强OLAP能力

增强OLAP主要是通过数据立方体（cube）进行数据分析，提供直观易懂的查询结果，满足多维环境下特定的查询和报表需求。其功能应包括但不限于：

* 1. 支持多维数据建模，需要能建立与设置如下功能:

1. 维度（Dimension）：是人们观察数据的特定角度，是考虑问题时一类属性的集合（时间维、地理维等）
2. 度量（Measure）：也称为测量值，是数值型的数据，通常是可聚合的。
3. 层次（level）：人们观察数据的某个特定角度（即某个维）还可以存在细节程度不同的各个描述方面（例如时间维：年、季度、月份、日期；地理维：国家、地区、省份、城市）。
4. 层次结构（Hierarchy）：维度中层次的上下级组织关系。（例如年-季-月是一个层次结构，年-月-日是另一个层次结构）同一个维度可以有多个层次结构。
5. 成员（Member）：维度上的某一个特定取值。
6. 成员属性（Properties）：维度成员的描述信息，name、caption等是成员的内置成员属性，可以定义其它附属的成员属性。例如商店维度的商店层次，有商店面积、店员人数等，这些信息属于这个商店这个分析维度，又不适合独立成一个维度。
7. 多维数据集（Cube）：一组分析维度和一组面向相同主题的度量，构成了一个多维数据集。
   1. 应支持的国产或者开源多维数据库。
   2. 应支持多维表达式(MDX):
8. 计算成员：计算成员是为一个动态添加的临时成员
9. 命名集：命名集是一个预先定义的集合，供后面的mdx子句使用。当一个集合重复出现多次时，可以通过命名集来简化语句，避免重复。
10. MDX函数：如数值函数、逻辑函数、字符串函数、成员函数、集合函数等

#### 增强计算能力

增强计算能力是指通过机器学习和AI，帮助用户准备海量数据进行分析，并提供见解。并助能够根据用户行为等不同维度的信息动态的提供预测，其功能包括但不限于：

* 1. 分布式计算。
  2. 直观的流式建模、拖拽式操作。
  3. Python扩展。
  4. 预测/聚类等成熟学习算法。
  5. 模型集中管理与发布功能。

#### 高性能

高性能主要是保障系统运行的生命周期和扩展能力，系统应具备高速缓存机制来提高系统性能，应对在处理大数据或超大数据时导致报表长时间刷不出来或者系统崩溃等现象。高性能应具备但不限于以下能力：

* 1. 应支持列式数据库
  2. 应支持MPP：

1. 应支持集群部署和大规模并行处理。
2. 应能够处理PB级别的、高质量结构化数据，同时为应用提供丰富的SQL和事务支持能力。
3. 应支持两个亿级表关联查询，秒级响应。
4. 对于海量大数据跨库查询，内置的跨库引擎应能够通过线性扩充、并行处理。

## 数据分析

数据分析指用适当的统计、分析方法对收集来的大量数据进行分析，将它们加以汇总和理解并消化，以求最大化地开发数据的功能，发挥数据的作用。数据分析是为了提取有用信息和形成结论而对数据加以详细研究和概括总结的过程。

### 报表

报表就是用表格、图表等格式来动态显示数据，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：

* 1. 应支持直接在Web上开发设计报表，以便更好的支持国产系统和软件。
  2. 应支持各种计算复杂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。
  3. 需兼容日常办公软件的表格函数。

### 数据探索

数据探索是指进行探索式的数据分析，用户尝试各种方法来探索发现数据中可能存在的关系。数据探索通常通过简单的操作，用户即可进行明细数据查询或是统计分析，还能自定义查询条件。

#### 即席查询

即席查询是用户根据自己的需求，灵活的选择查询条件，系统能够根据用户的选择生成相应的统计报表。即席查询与普通应用查询最大的不同是普通的应用查询是定制开发的，而即席查询是由用户自定义查询条件的。其功能需包括但不限于：

* 1. 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。
  2. 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。
  3. 应支持大数据量的批量导出。
  4. 应支持加密导出，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系统进行数据分析。

#### 透视分析

透视分析是指通过交互式的方式进行多维度统计分析。它可以进行某些计算，如求和与计数等，所进行的计算与数据跟维度的排列有关。之所以称为透视分析，是因为可以动态地改变分析维度，以便按照不同方式分析数据，每一次改变维度布置时，透视分析会立即按照新的布置重新计算数据。另外，如果原始数据发生更改，则可以更新透视分析。

* 1. 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。
  2. 应支持多种分析方法，包括但不局限于切片、钻取、旋转、行列互换、焦点项等。
  3. 应支持超多维度、甚至维度无法固定的分析场景
  4. 应支持通过自定义计算字段，扩展可分析的维度和度量
  5. 应支持超大数据量的查询性能，亿级数据秒级响应。
  6. 应支持自定义时间维度，并支持多种时间计算选项，包括但不局限于：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。

#### OLAP分析

OLAP分析使得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，以利于迅速做出正确的判断，辅助商业决策。其功能需包括但不限于涵盖多维分析、统计分析、自定义计算。

* 1. 多维分析

支持用户能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企业的数据，以多维的形式从多方面和多角度来观察企业的状态与变化。

1. 钻取:能够改变当前多维分析中维的层次，变换分析粒度的一类分析操作。包含但不限于向上钻取、向下钻取、层钻操作。向上钻取，应当支持以渐增概括方式汇总数据(例如，从周到季度，再到年度);向下钻取，应当支持使用户在多层数据中展现渐增的细节层次，获得更多的细节性数据;层钻，应当支持在维度树结构中根据查找条件搜索出当前维度中任意层次中的成员。
2. 聚合:系统应当支持用户以更为简便的表格操作方式实现维度成员的聚合。包括但不限于常用的汇总方法，如合计、平均值、最大值和最小值。
3. 切块:系统应当支持使人们能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。同时，应当支持用户为了实现更复杂的需求，能够自定义一类参数。
4. 行列互换:系统应当支持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。
5. 分级排序:系统应当支持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。
6. 系统设计好的多维分析报表发布到门户后，支持业务人员仍然能够在浏览界面（非编辑报表界面）继续对维度和度量等进行调整，继续做进一步探索。
   1. 统计分析

通过简便操作就能实现与业务有关的统计分析能力，包括但不限于现状分析和发展分析。

1. 现状分析:系统应当支持用户能够从不同角度对当前关键数据进行分析，从而可以了解机构业务或经营情况的现状，现状分析方法应该包括但不限于占比分析、排名分析等。
2. 发展分析 :系统应当支持用户能够对同一个指标不同时期的样本进行比较，从而总结出分析指标发展规律；发展分析方法应该包括但不限于环比分析、前期值分析、同期值分析、同比分析等。
   1. 自定义计算

支持用户无需编程就可以定义新的专门计算，实现复杂的业务计算需求。

1. 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。
2. 计算优先级:存在两个或多个自定义成员的计算规则时，能够设置自定义成员的计算优先级，以保证计算结果的正确性。

### 仪表板

仪表板代表了一种报告风格，它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。仪表板通过组合信息和交互能力来帮助改进决策。它还使用诸如刻度盘、仪表和“红绿灯”之类的信息图形来显示指标，这些信息图形指示显示着既定目标的进展。在某些情况下，仪表板正在演变为自己的“应用程序”。

#### 交互式可视化

交互式可视化是指将数据转换成图形或图像在屏幕上显示出来，用户可直接在图形或图像上进行交互式操作。需要支持的图形或图像包括但不限于：

* 1. 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、小又多图等等特色图形，支持3D航线图、3D散点图、3D柱图等3D图形。
  2. 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，内置支持百度地图，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、散点标记地图、热力地图和航线图等；
  3. 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。
  4. 应具备较好的扩展能力，能够扩展第三方图形。

#### 自助仪表盘

仪表盘是为了实现些某些特定目标而对重要信息进行的视觉传达方式，对一屏上的内容进行组织呈列使人一瞥便能掌握其所要传达的信息。自助仪表盘是人人都可以操作的仪表盘，用户只用托拉拽的操作全程自助式地完成仪表盘的操作。自助仪表盘应包括但不限于以下功能：

* 1. 应支持可视化拖拽操作，简单易用；还支持所见即所得，一屏完成仪表板的创建。
  2. 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。
  3. 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。
  4. 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；
  5. 应支持多种终端的适合，创建的成果支持发布电脑、手机、平板等。
  6. 应支持各种交互效果，包括但不局限于联动、钻取、跳转等。

#### 大屏可视化

大屏可视化主要是将数据和信息通过可视化的形式一屏全面展现。通过可视化大屏，企业可以观察数据的变化、了解业务的发展，及时发现问题、根据问题解决实际业务问题，帮助企业更有效地管理决策、监控运营等。大屏可视化应包括但不限于以下功能：

* 1. 应支持各种大屏种类，包括但不局限于管理驾驶舱、监控大屏、可视化展示报告等。
  2. 应支持自由布局及自适应，组件、控件任意摆放，高宽自适应。
  3. 应支持不同应用场景的业务互动，包括但不局限于多种筛选联动、不限层次的钻取、跳转资源、监控预警等。
  4. 应支持多终端展示，包括但不局限于PC、手机、平板、液晶显示屏、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。
  5. 应支持大屏和人工智能融合，比如通过智能的机器学习算法加持，可以自动分析波动原因。用户可以通过PC端、移动端语音命令或手动输入提问，即可进行统计分析。

### 分析故事

分析故事是将交互式数据可视化与叙事技术相结合，叙述从数据中发现的规律，以引人注目的、易于理解的形式打包和交付分析内容，以呈现给决策者。分析故事应包括但不限于以下功能：

* 1. 应包含多个故事点，每个故事点都能够利用数据讲述交互式故事
  2. 应支持在故事点中添加分析报表或仪表盘，便于使用一个屏幕讲述整个可视化故事
  3. 同个分析故事中应支持复用已有故事点，简单修改即可从不同维度传递另一个新的分析故事。
  4. 应支持添加文本，表达观点和最终结论
  5. 应支持设置分析故事显示大小和布局，以适应不同分辨率的显示设备
  6. 应支持与来源保持连接，基础数据、分析报表、仪表盘发生变化时，故事内容随之变化

### 移动BI

移动BI是通过移动设备（平板电脑和智能手机）提供报告、仪表板、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：

* 1. 应支持H5标准，能够与微信企业号、钉钉集成；
  2. 应支持报表分享，可以匿名将报表分享到微信等第三方系统。
  3. 应支持在离线状态下可以进行交互分析，包括但是不局限于钻取、跳转、改变查询条件等；支持离线包推送，用户接收后可以增量更新离线数据包。
  4. 应该提供数据安全保障的访问机制，包括但不局限于密码错误安全响应机制、数据权限控制、访问资源权限控制等。

## 增强分析

### 预测分析

预测分析是对商业数据中的大量业务数据进行抽取、转换、分析和其他模型化处理，从中提取辅助商业决策的关键性信息的过程。预测分析模块旨在为企业所做的决策提供预测性智能服务，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，且采用大规模分布式的计算引擎实现高效的模型训练。

#### 可视化建模

可视化建模是指通过图形化的实现，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流程的能力，包括但不限于：

* 1. 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，每个节点可以自定义参数；
  2. 应提供直观的流式建模方式，通过拖拉拽的方式快速构建机器学习模型；
  3. 应支持启动/停止任务，断点重跑；
  4. 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。

#### 特征处理

特征处理指的是根据业务需求，对数据表的字段进行加工、扩展和衍生，形成建模所需要的特征变量，最终生成一张统一的宽表数据进行建模处理，并能够对业务模型的特征变量进行重要性分析，帮助业务人员理解。

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：

* 1. 对于各种类别型特征，应支持可对其编码；
  2. 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；
  3. 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。
  4. 应提供多种消除特征之间的量纲影响的数据归一化方法；
  5. 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；
  6. 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；

#### 预测算法

机器学习是一类算法的总称，这些算法从大量历史数据中挖掘出其中隐含的规律，并用于预测或者分类。预测分析模块应具备多种机器学习算子，不需要写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：

* 1. 向量机、梯度提升决策树等二分类算法：
  2. 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；
  3. 线性回归、梯度提升回归树等回归算法；
  4. K均值、高斯混合模型等聚类算法；
  5. FP-Growth等关联规则算法
  6. 分词、停用词处理、TF-IDF、Word2Vec词向量、词频编码、词频统计、主题模型等文本分析算法，具有对文字类型的数据进行处理的能力。

#### 模型评估

模型评估功能包含一系列模型评估标准，通过多维度、多度量体系的评估指标体系来评判模型的好坏。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：

* 1. 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；
  2. 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；
  3. Calinski Harabasz指标、和方差、轮廓系数等聚类评估指标。

#### 模型部署

模型部署指的是将已经训练好的优化预测模型部署成为Web服务，提供API供实际业务使用。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，提供预测服务的能力，包括但不限于：

* 1. 应提供模型保存、模型导入、模型导出的功能。
  2. 调用服务后应支持依据训练好的预测模型返回预测结果；
  3. 应支持服务发布上线/下线；
  4. 应支持灰度测试功能，即多版本模型并行运行；
  5. 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。

#### 自动建模

自动建模旨在为不懂算法、编程的业务人员轻松构建符合业务应用的分析模型，将数据挖掘的门槛降到最低，节省企业成本。

在预测分析模块中，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：

* 1. 应支持输入数据、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；
  2. 应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。
  3. 应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。

#### 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科学。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：

* 1. 应支持动态扩展，可随时加入新的计算节点
  2. 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
  3. 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

### 自然语言分析

BI（Business Intelligence）系统中的自然语言分析是指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，而且其核心能力以API（Application Programming Interface）形式供第三方系统调用。

#### 自然语言指令解析

自然语言指令解析是指通过内置语义分析模型，对输入的自然语言进行语义分析，进行用户意图识别，并根据数据系统中实际情况进行反馈，其功能应包括但不限于：

* 1. 应提供资源搜索功能，能够对数据指标、报表、功能进行查找。
  2. 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，根据数据系统的实际情况生成查询语句，获取数据；
  3. 应提供预测分析功能，能够使用系统中内置预测模型，基于已有数据对数据未来发展趋势进行预测，并将预测结果返回。

#### 指令执行

指令执行是指通过自然语言完成标准化程度较高的操作，包括但不限于：

* 1. 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。
  2. 应支持报表查询访问，通过自然语言检索系统中的报表与分析。
  3. 应支持参数切换，通过自然语言切换报表、分析中的参数。

#### 多轮对话/会话式分析

多轮对话是用于构建系统与用户对话共识，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 rephrase），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：

* 1. 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。
  2. 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。

#### 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：

* 1. 登入接口：调用开始，对调用方身份进行识别，授权使用功能。
  2. 登出接口：功能调用结束，关闭授权。
  3. 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。
  4. 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。
  5. 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。

## 分享与协作

分享是将一件物品或是信息的使用权和知情权与他人共用拥有；协作是指在目标实施过程中，部门与部门之间、个人与个人之间的协调与配合。

企业间数据和应用的分享和协作，可以帮助企业将数据资源和分析应用的价值发挥到最大。

通过分享和协作可以真正提升企业的决策效率，解决企业存在无沉淀、无共享、无文化的问题，帮助企业构建完善的数据化运营的生态系统。

### 应用门户

应用门户是把数据分析资源、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。

* 1. 应支持显示在不同的终端，包括但不局限于PC端、移动端。
  2. 应满足不同部门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等；门户还应该能在其他方面支持个性化，比如布局、配色等。
  3. 应支持丰富的门户组件，包括但不局限于收藏夹、卡片列件、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。
  4. 应支持提供完备的门户组件接口，供第三方信息接入。

### 消息中心

在日常的办公中会产生各种消息通知，包括预警、推荐、公告、待办、@、私信等等，消息中心用于实现平台内消息的接收、存储、消息状态报告、推送等功能。信息资讯互联互通，所有分析相关、平台相关的资讯都在消息中心访问。

* 1. 根据消息来源和应用范围，应支持不同种类的消息，包括但不局限于系统消息、数据预警、审核信息、企业公告信息、个人私信、评论行为提醒等等。
  2. 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。
  3. 消息查看应该具有便捷性。满足的功能包括但不局限：入口明显，点击信息即可打开查看详细内容等。
  4. 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。
  5. 消息中心应该支持统一管理，包括但不局限于编辑重发，重发，删除等操作。

### 互动交流

互动交流是指用户对平台上的各种资源都可以进行点赞、评论、收藏等互动，并且也能把好的分析成果分享给其他人。通过互动交流企业能够动态传播更多优秀的分析成果，吸引更广泛用户参与到分析平台里面，形成企业独特的数据分析社交文化。

* 1. 互动的方式应包括但不局限于点赞、评论、收藏、分享。
  2. 应提供分享功能，系统内部用户间可互相分享优秀的分析资源，也可以通过OA、钉钉等向外部用户分享。
  3. 应支持一次分享多个资源，被分享者能收到消息提醒。
  4. 分享应保证分析资源的安全，可以控制的权限包括但是不限于资源范围、用户范围、时间范围、权限粒度等。

# 非功能需求

## 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。

### 应用安全

#### 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：

* 1. 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。
  2. 应支持自定义用户密码复杂度校验规则，用户名校验错误提示；比如：密码至少包含大写字母、小写字母、数字以及特殊字符中的三个。

#### 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，要求包括但不限于：

* 1. 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。
  2. 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，并且是否继承的方式可配置。
  3. 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。

#### 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](https://baike.baidu.com/item/%E5%AE%89%E5%85%A8%E7%AD%96%E7%95%A5/2890163)，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：

* 1. 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。
  2. 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。
  3. 应能够根据记录数据进行二次分析，并生成审计报表。

#### 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，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：

* 1. 应具备自研分发器，无需依赖应用服务器分发，系统可自行依据应用节点CPU、GC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。
  2. 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，无须用户重新登录，保障业务操作不中断、数据/模板不丢失。
  3. 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，从而提示存在的问题进行修复。

#### 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。

#### 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。

#### 安全补丁更新

应支持安全补丁热修复，即可在不重启设备的情况下，对设备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。

### 数据安全及备份恢复

#### 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。

#### 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（如测试环境、开发环境和生产环境）之间的快速资源迁移。

### 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，并提供严格的安全意识培训和安全审查。

## 开放兼容性

系统应具备良好的开放兼容性，必须能与主流的国产化产品完全兼容。

## 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、长时间不间断运行的情况下系统仍然健壮性良好，并且无论是前端访问还是数据后台都具备强大的计算能力，系统正常稳定运行为用户带来流畅且持续的使用体验。

* 1. 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；
  2. 在单节点8核CPU、32GB内存配置下，应支持不少于200并发用户量访问，并且登录平均耗时应小于1秒。
  3. 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。
  4. 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。
  5. 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。

## 拓展性

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。

### 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。

* 1. 应支持通过一定的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。
  2. 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。
  3. 应提供故障管理和控制的机制，故障切换组件使用冗余系统和数据来保证关键任务数据流不间断。
  4. 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。

### 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。

## 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：

* 1. 应提供Java 、Web Service、JS等多种技术二次开发接口，并有完整的中文说明文档。
  2. 应支持功能模块层面集成，在第三方系统内整合模块功能。
  3. 应支持用户自行通过JS脚本扩展或插件包扩展方式进行定制化开发，不直接修改产品源码，保证系统的稳定性与升级的方便性。

# 验收要求

## 测试标准

系统验收前，由用户方主导测试验收，组织对应用软件进行测试验收，并提供合格证明，内容包括但不限于：

### 功能测试

应对软件需求规格说明书中的所有功能项进行测试。

所有功能模块完成测试后，结束标志应为：系统功能满足设计需求。

### 业务流程测试

应对典型的业务流程进行测试。

典型业务流程完成测试后，结束标志应为：系统符合业务流程设计需求。

### 容错测试

软件应对错误操作具备提示功能，错误提示应包含但不限于以下场景：

* 1. 应对用户常见的错误操作具有提示。
  2. 应对软件错误具有提示。
  3. 应对重要数据的删除具有提示。

所有模块通过容错测试后，结束标志应为：软件容错测试满足要求。

### 安全性测试

应对软件安全进行测试，安全性测试内容包括但不限于加密、数据备份、恢复、病毒检测、权限管理、日志监控等测试。

软件安全测试通过后，结束标志应为：软件安全性符合要求。

### 性能测试

应对软件需求规格说明书中明确的软件性能进行测试。

软件性能测试通过后，结束标志应为：软件性能满足需求。

## 验收标准

### 软件错误的等级要求

应将软件错误的等级划为四等。

* 1. 等级1：不能执行正常功能或重要功能；
  2. 等级2：严重地影响系统要求或基本功能的实现；
  3. 等级3：严重地影响系统要求或基本功能的实现,但存在合理的解决办法；
  4. 等级4：使操作者不方便或遇到麻烦,但不影响执行正常功能或重要功能。

### 验收指标要求

* 1. 测试用例不通过数的比例< 1.5 %；
  2. 不存在错误等级为1 和2的错误；
  3. 错误等级为3 的错误数量≤ 5；
  4. 所有提交的错误都已得到更正。

### 验收文档要求

应提供系统验收文档，包括但不限于：

* 1. 项目验收申请报告；
  2. 软件需求说明书；
  3. 数据库设计要求说明书；
  4. 详细设计说明书；
  5. 软件测试报告；
  6. 操作手册；
  7. 合同要求或专家组要求的其他材料。

## 其它要求

系统验收应参考GB/T 17544 信息技术软件包质量要求和测试、GB/T 16260 软件质量模型与度量介绍的规定。

# 售后服务

自软件产品交付之日起，软件应为客户方提供一年免费的技术支持服务，确保客户方日常使用本软件时可能出现的问题能及时地得到解决。

免费服务期满后，根据客户方的需求，软件宜继续提供有偿的技术服务。有偿维护合同签署后，按照合同提供与免费维护期内同等的服务内容、组织、响应及服务质量。