

SZSD

数 字 山 东 技 术 规 范

SZSD06 0014—2026

代替 SZSD06 0027—2020

一体化大数据平台设计规范

Design specification of integrated government big data platform

2026 - 08 - 19 发布

2026 - 10 - 01 实施

烟台市大数据局 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 2

5 设计要求 2

6 设计原则 2

 6.1 可靠性原则 2

 6.2 开放性原则 3

 6.3 安全性原则 3

 6.4 高效性原则 3

 6.5 友好性原则 3

 6.6 集约性原则 3

 6.7 可伸缩性原则 3

 6.8 可扩展性原则 3

7 一体化大数据平台设计 3

 7.1 平台工作模式 3

 7.2 平台工作流向 3

 7.3 平台功能设计 5

 7.4 平台部署设计 6

 7.5 数据资源规划架构设计 7

参考文献 9

前 言

本文件按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》给出的规则起草。

本文件代替SZSD06 0027—2020《政务数据归集平台设计规范》，与SZSD06 0027—2020相比，除结构调整和编辑性改动外，主要技术变化如下：

- a) 由“政务数据归集平台”变更为“一体化大数据平台”（见标题）；
- b) 由“归集”为核心变更为以“一体化大数据平台”为核心；
- c) 旧定义强调“归集+管理”，新定义覆盖“汇、治、用、安”全链条（见3术语和定义）。

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。

本文件由烟台市大数据中心提出，由烟台市大数据局归口。

本文件起草单位：烟台市大数据中心、浪潮软件集团有限公司。

本文件主要起草人：孙伟、陈大炜、王蕾、耿晓燕、苏家志、姜斐、孙浩荃、王涵霄、王珏、顾雪平、周子钦、邓鹏健、张倩楠、王佳慧。

本文件及其所代替文件的历次版本发布情况为：

2020年首次发布为SZSD06 0027—2020，本次为第一次修订。

一体化大数据平台设计规范

1 范围

本文件规定了一体化大数据平台的设计要求、设计原则以及一体化大数据平台设计等内容。
本文件适用于非涉密的一体化大数据平台的建设与管理。

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。

SZSD 0027.1—2024 一体化大数据平台 第1部分：总体架构

3 术语和定义

下列术语和定义适用于本文件。

3.1

一体化大数据平台 Integrated Government Big Data Platform

是指市级政务数据资源汇聚整合、规范治理、共享交换、安全管控的一体化信息技术系统，是全省一体化政务大数据体系的市级节点，对接省级与国家平台，统筹本市跨部门、跨层级政务数据，具有目录管理、数据湖存储、治理加工、共享服务、安全防护等核心能力。

3.2

中心数据湖 central data lake

是指汇聚各类政务数据、以原始格式集中存储结构化、半结构化、非结构化全类型数据的存储与管理平台，支撑政务数据资源的统一存储、处理、治理与共享应用。

3.3

元数据 metadata

是指描述信息资源特征的数据。

3.4

数据仓库 data ware house

是指出于分析性报告和决策支持目的而创建的，提供所有类型数据支持的战略集合，是面向主题、集成性、稳定性和时变性的大型数据存储库。

3.5

单点登录 sign in

是指在多个应用系统中，用户只需要登录一次就可以访问所有相互信任的应用系统。

3.6

数据脱敏 data desensitization

是指对政务数据的某些敏感信息通过脱敏规则进行数据的变形与隐藏，实现敏感隐私数据的可靠保护。

3.7

数据清洗 data cleaning

是指发现并纠正汇聚的政务数据文件中可识别的错误的最后一道程序，包括检查数据一致性，处理无效值和缺失值等。

3.8

防火墙 firewall

是指对经过的数据流进行解析，并实现对一体化大数据平台访问控制和安全防护功能的网络安全产品。

3.9

网闸 gatekeeper

是指连接两个独立主机系统的信息安全设备，从逻辑上隔离、阻断对政务外网具有潜在攻击可能的一切网络连接，使外部攻击者无法直接入侵、攻击或破坏内网，保障了内部主机的安全。

3.10

前置机 front-end processor

是指部署在政务部门核心业务系统之前，用于实现部门之间数据交换的数据库服务器。

4 缩略语

以下缩略语适用于本文件：

LDAP：轻型目录访问协议（Lightweight Directory Access Protocol）。

SOAP：简单对象访问协议（Simple Object Access Protocol）。

UDDI：用于描述、发现、集成Web Service的技术（Universal Description Discovery and Integration）。

Web Service：是一个平台独立的，低耦合的，自包含的、基于可编程的web的应用程序。

XML：可扩展标记语言（Extensible Markup Language）。

5 设计要求

一体化大数据平台应满足以下设计要求：

- d) 满足集中政务数据资源存储的软、硬件配置要求；
- e) 满足目录管理系统、资源管理系统、数据交换系统、共享开放网站等数据资源平台建设的数据支撑要求；
- f) 满足各政务部门对数据汇聚、存储、交换、管理过程中的数据安全需求；
- g) 满足政务数据汇聚、交换过程中的网络安全要求。

6 设计原则

6.1 可靠性原则

一体化大数据平台在设计时应满足政务数据汇聚共享工作所涉及的功能和性能的要求，并同时保证各个子系统之间的可靠性和可用性，尤其是保证关键业务的连续不间断运作和对非正常情况的可靠处理。

6.2 开放性原则

应全面支持LDAP、SOAP、UDDI、Web Service、XML等当前受到普遍支持的国际开放标准，保证平台能够从其他平台的应用系统、数据库等采集数据并进行应用级的互操作性和互连性。

6.3 安全性原则

既要注重一体化大数据平台本身防御外部攻击的能力，又要提高数据传输的安全性，遵循保密规范，保证一体化大数据平台安全保密性、查错纠错力、抗病毒等能力，实行信息安全等级保护制度，严格落实国家网络安全等级保护制度，一体化大数据平台整体安全保护等级不低于第三级。

6.4 高效性原则

一体化大数据平台的运行效率应主要包括：处理能力、处理速度和响应时间等。

一体化大数据平台应在满足其他各项功能需要的前提下，提高平台运行效率。

6.5 友好性原则

应具有人性化的人机交互界面，要求用户界面标准，统一集成，使用简单，提高使用效率。

6.6 集约性原则

应按照集约化建设要求，依托统一的政务云服务大数据基础设施建设。

6.7 可伸缩性原则

应符合多层浏览器/服务器体系结构，应能满足功能新增的需求，保证一体化大数据平台之后的平滑升级。

6.8 可扩展性原则

应充分考虑现有技术，保证一体化大数据平台具有较好的开放性和结构的扩展性，在一体化大数据平台设计中应尽量采用模块化结构，提高各模块的独立性，尽可能减少模块间的数据耦合，使各子系统间的数据依赖程度减至最低限度，同时要适当兼顾一体化大数据平台功能扩展的需要。

7 一体化大数据平台设计

7.1 平台工作模式

数据管理部门按照物理集中与逻辑集中相结合的模式，通过库表、接口和文件上传的汇聚方式实现政务数据资源的统一汇总、存储。

7.2 平台工作流向

数据从各数据源汇聚至中心数据湖，根据汇聚方式不同分为以下两种类型：

- a) 以库表或文件形式汇聚的政务数据，根据数据来源的多样性和数据传输过程中的安全风险分析，数据汇聚流向见图 1：

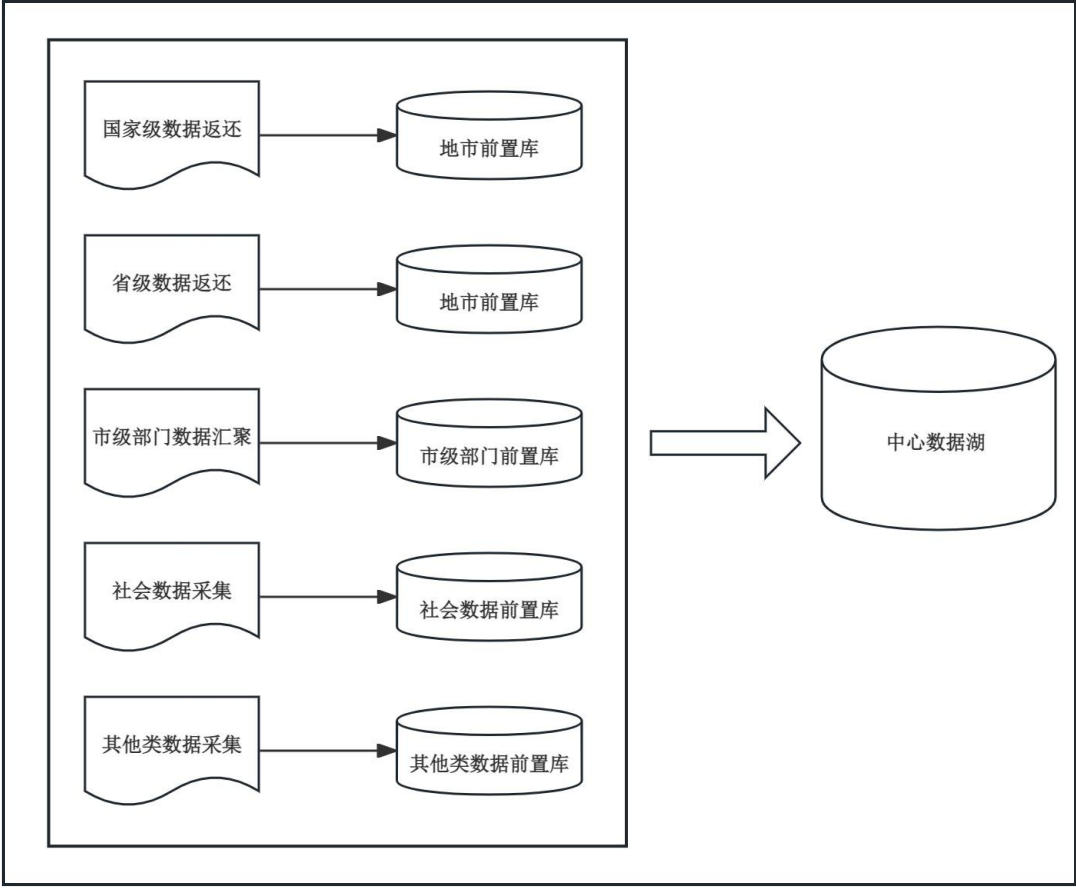


图1 库表文件方式汇聚的数据流向图

- b) 根据数据提供方数据保密性或技术限制性的约束，一体化大数据平台应提供接口形式汇聚的政务数据，接口方式汇聚的数据流向见图 2：

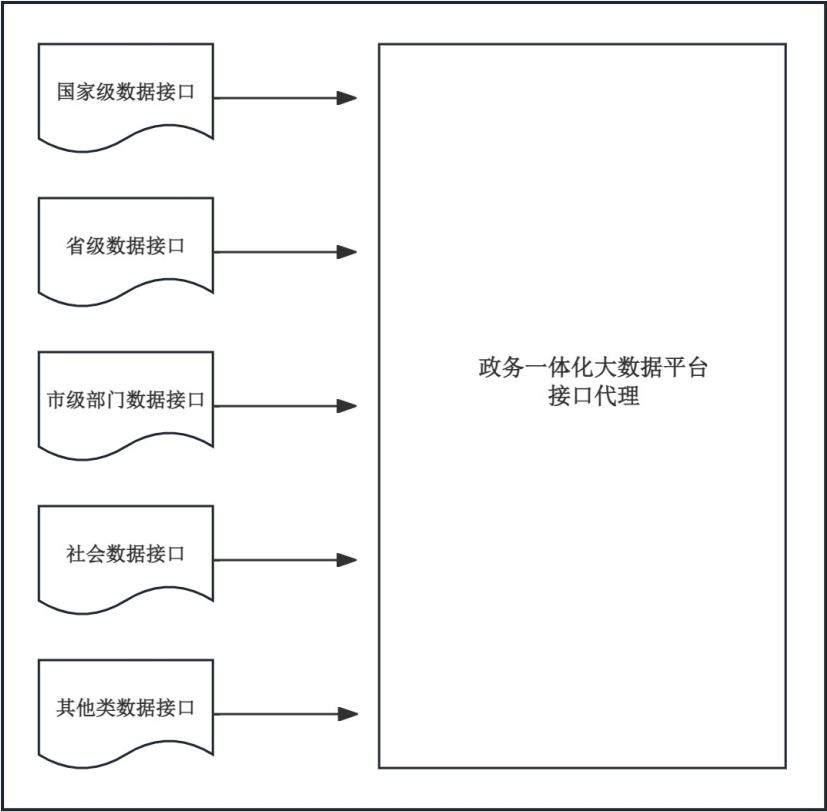


图2 接口方式汇聚的数据流向图

7.3 平台功能设计

一体化大数据平台的功能设计应满足保障政务数据汇聚工作的各个方面，平台功能架构设计见图3：

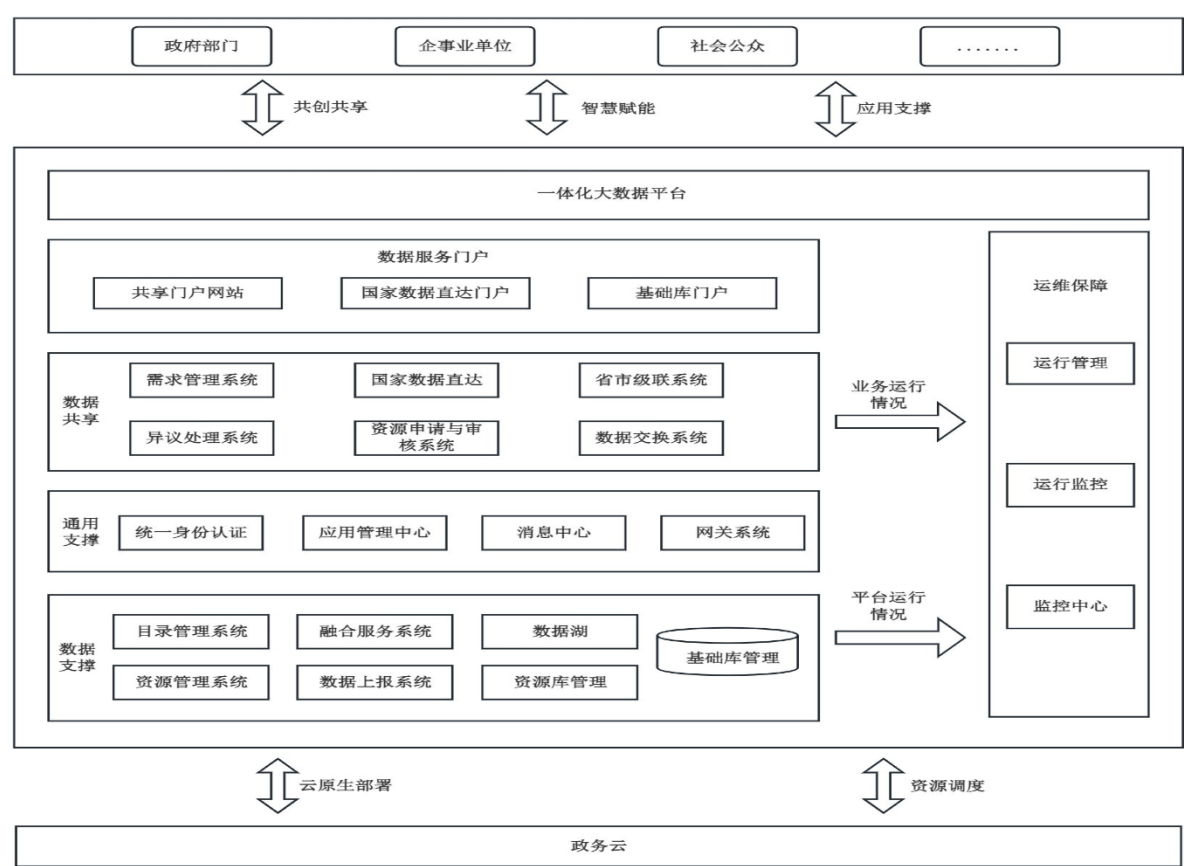


图3 平台功能架构图

依据一体化大数据平台功能架构设计模型，一体化大数据平台所涉及的应用系统建设应包括但不限于：

- a) 目录管理系统：可以支持政务信息资源目录资源的梳理、采集、描述、编目、分类目录管理等功能；
- b) 供需对接系统：可以支持各部门业务需求的申请、审核以及答复等功能；
- c) 级联对接系统：可以实现国家、省级、市级三级数据级联对接；
- d) 资源管理系统：可以支持对政务数据资源全生命周期的管理功能，包括库表、文件和服务接口方式汇聚的三种类型的数据。
- e) 资产管理系统：可以实现对政府内部信息资产进行管理和维护的功能；
- f) 数据交换系统：可以提供库表数据抽取、文件交换、全流程监控等功能；
- g) 融合服务系统：可以支持通过服务网关，承载服务注册、代理的功能；
- h) 用户管理系统：可以实现对用户、应用、组织机构等信息进行配置和管理的功能；
- i) 数据湖系统：可以实现对数据进行分析演练及配置管理功能，并根据业务需求生成可复用的业务数据模型；
- j) 运行管理系统：可以实现一体化大数据平台各模块运行日志采集、监控、平台巡检及发现问题等功能；
- k) 安全管理系统：可以实现数据全生命周期的安全管理功能，并对应用系统级数据库提供安全保障。

7.4 平台部署设计

一体化大数据平台建设应综合考量网络环境、基础设施资源、技术实现等多个方面要素，平台部署架构见图4：

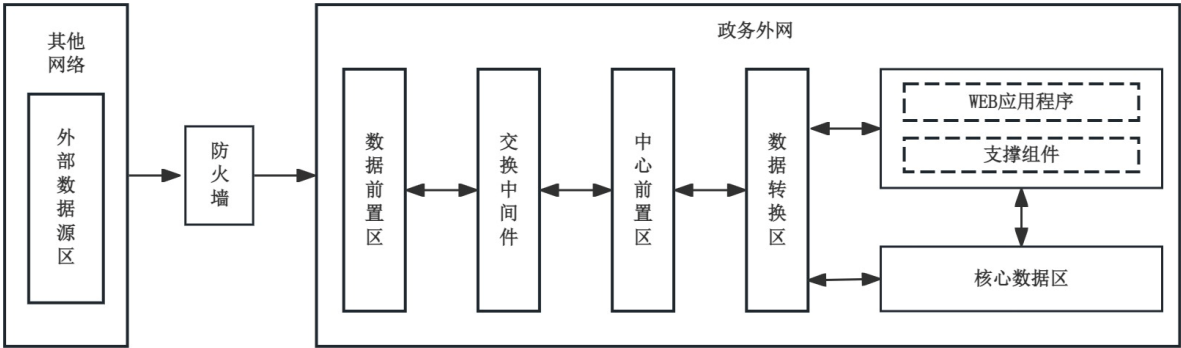


图4 平台部署架构图

一体化大数据平台根据政务数据汇聚的业务要求、技术要求、存储要求、网络环境要求选择基础设施配置清单，应包括但不限于：

- a) 应描述一体化大数据平台设计的逻辑架构，包括模块与组件之间的层次关系以及依赖关系；
- b) 应确认政务服务部署架构，包括云平台资源所在的物理位置和逻辑位置、区域内和跨区域网络通信方式、高可用方案、容灾容错的逻辑视图等；
- c) 应根据资源配置清单，维护配置管理中资源配置和组件关系信息，形成部署架构设计说明书。

7.5 数据资源规划架构设计

7.5.1 架构设计

一体化大数据平台的数据资源规划可以分为外部数据源和中心数据湖两部分，数据资源规划架构图见图5：

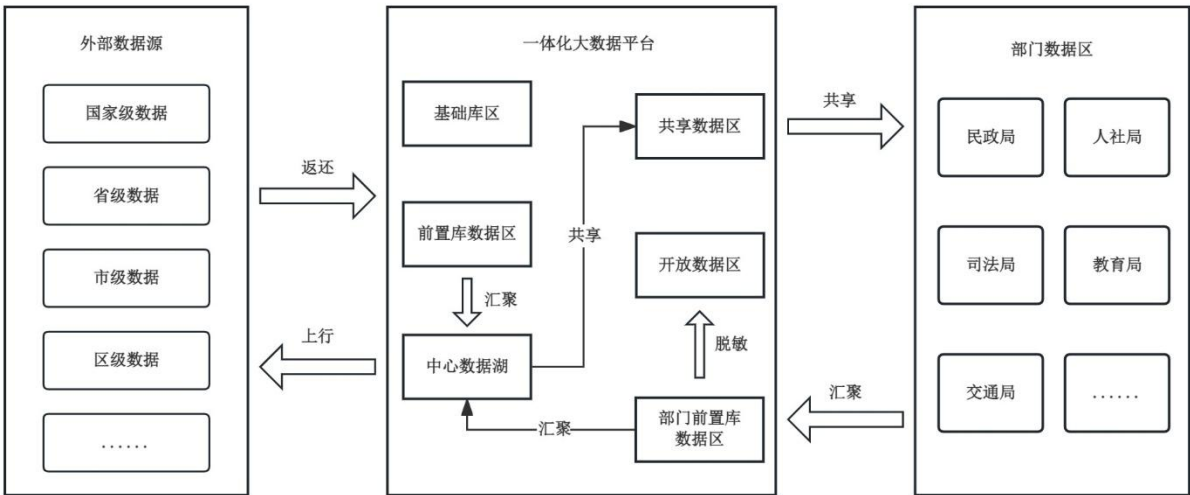


图5 数据资源规划架构图

架构体设计参考DB 37/T 3520和DB 37/T 3522。

7.5.2 外部数据源

一体化大数据平台建设应获取一切可获取的有效政务数据资源，其中包括：国家级、省级、市级、区县级四级垂直管理的政务数据资源以及其他渠道可以获得的政务数据资源，并将可获取到的数据资源汇聚至中心数据湖。

7.5.3 中心数据湖

中心数据湖中存在结构化数据区、非结构化数据区和内存数据库区，中心数据湖中的各个数据库可根据数据的类别、使用场景，选择不同的存储区。中心数据湖应具备更新扩展的能力，其中包含但不限于以下数据分区：

- a) 部门数据区：根据数据所属部门进行划分，存储部门源数据；
- b) 共享数据区：根据数据共享属性进行划分，存储可共享的数据；
- c) 开放数据区：根据数据开放属性进行划分，存储可开放的数据；
- d) 基础库区：由政府集中建设、统一管理，包括但不限于人口、法人单位、空间地理、宏观经济、电子证照、公共信用等信息资源库；
- e) 主题库区：由政府单独或牵头建设的信息资源库，包括但不限于住房保障、医疗保障、生态环境、民生保障、社会保障等信息资源库；
- f) 配置库区：存储配置库保存的数据目录、元数据及其他配置数据；
- g) 分析演练区：根据业务需求对数据进行分析演练，形成可直接使用并可复用的业务数据模型。

参 考 文 献

- [1] 《政务数据共享条例》（国务院令第809号）
 - [2] 《山东省数字政府建设一体化综合改革方案》（鲁政办发〔2025〕2 号）
 - [3] 《山东省电子政务和政务数据管理办法》（省政府令第 329 号）
 - [4] DB37/T 3522—2019 政务信息资源 数据交换规范
 - [5] DB37/T 3520—2019 政务数据资源体系 总体架构
-