技术规范书

1.1 大数据共享交换与管理平台

1.1.1 系统概述

数据共享交换与管理平台是大数据建设的核心，借助大数据相关技术进行数据治理。实现数据共享交换与数据管理，达到数据标准化、规范化、准确化、共享化、安全化。然后将沉淀的真实数据汇聚为大数据资源池向外提供调用服务。

1.1.2 数据共享交换功能

1.1.2.1 建设原则

1.标准性。设计应该遵循相应的国际、国内、地方或行业标准。按照国标 GB/T 21062 政务信息资源交换体系、GB/T 21063 政务信息资源目录体系、GB/T 28168 消息传输等相关规范，以及当前国际主流的 WebService、SOA 参考架构等进行设计，并建立数据交换标准规范体系，保障平台可靠稳定长效运行。

2.安全性。项目建设应对数据、应用的安全性进行充分考虑，充分利用接入系统认证、用户角色管理、数据加解密、服务安全认证、操作事后审计、内外网不同网段之间等安全技术手段确保平台安全可靠运行。

3.稳定性。为保障平台的稳定可靠运行，平台建设的底层核心技

术是关键。核心技术必须是经过多年验证应用的成熟技术，必须采用成熟、稳定、可靠的商用中间件技术支撑软件。

4. 可扩展性。接入系统类型和部门逐步增多，平台必须支持大规模运行需要，并随着业务应用发展进行接入量、并发量的动态扩展，不影响已完成平台的运行。

5. 易用性。平台建设充分考虑面向用户的易用性，将底层技术核心进行二次封装，采用 B/S 架构实现数据交换的 Web 配置、共享数据目录和共享服务的 B/S 注册管理及调用。

1.1.2.2 技术路线

1.1.2.2.1 基础能力

1. 时效性。接入节点具有可扩展性。应用页面的响应时间不超过 5 秒。

2. 安全性。数据共享交换系统建立安全防护技术保障体系和安全管理保障体系，设计安全防护策略，通过节点身份认证机制、数据加密和访问控制等多种安全防护措施，确保政务服务数据共享交换系统、交换数据的安全符合等保三级安全要求。系统具备基本的安全防护能力，能够防护如 sql 注入、csrf 跨站请求攻击、xss 跨站脚本注入等常见的网络请求攻击行为。

3. 稳定性。数据共享交换系统满足为交换共享、业务协同提供基

础支撑，稳定、可靠的要求，具备 7*24 小时连续运行能力。

4.扩展性。随着各部门、各县（市、区）各项业务的扩展和信息化建设的不断推进，数据共享交换系统具备横向的功能扩展和纵向的性能扩展能力，能够支撑未来全市业务持续增长带来的新增的数据交换与共享需求。系统支持跨平台部署，同时支持 windows 和 linux 操作系统服务器，也可支持基于 docker 容器进行部署。系统支持 SQL Server、Mysql、MPP、Hbase 等多种数据库类型。系统具备分布式架构部署的能力，可以支持横向拓展。系统容量可以支撑现有信息资源的存量，并能支持未来 3 年的信息资源增长的能力。

5.维护性。随着市内各部门、各县（市、区）各项业务的扩展和信息化建设的不断推进，数据共享交换系统的功能会逐渐扩展，部署环境所需要的基础设施、系统软硬软件会逐渐增加，数据共享交换系统具备良好的维护性和较低的后期运维成本。

1.1.2.2.2 系统功能指标

1.1.2.2.2.1 数据存储功能

政府数据量很大，海量数据的处理与传统架构完全不同，原有小型机加商业数据库的方式已不能完全胜任，部分领域数据需要引入新架构。大数据存储于处理架构包括分布式存储、分布式资源管理及分布式文件架构。

1.1.2.2.2.2 数据汇聚功能

数据汇聚功能要求实现将各部门相关的信息资源统一采集交换到数据中心前置库中，满足多种采集方式，能够满足不同的网络环境、不同的数据类型等情况下进行数据的采集，并对采集的数据进行处理，进入中心库保存。

1.1.2.2.2.3 数据治理功能

政府数据在共享共用、开放运营及行业应用过程中，数据来源于各个业务系统，只有建立对数据质量的信任，才能放心地进行使用。所以数据治理和质量保障在政府数据开放共享平台建设中显得由为重要，数据开放共享平台数据治理功能主要包括信息资源目录管理、元数据管理和数据质量管里等内容。

1.1.2.2.3 数据处理能力

系统对数据的处理采用负载均衡技术，并与提供硬件设备厂商的虚拟管理软件相集成，使数据流量在系统中具有高可靠性和强大的并发处理能力，能够满足大量用户访问时所产生的压力，能够保证系统的高可靠和可用性。

1. 数据存储量：对数据容量理论上没有限制，总容量依赖于数据库存储容量，系统对数据的存储容量也没有限制，要求当达到数据库的存储容量或性能降低时，可以通过数据库（服务器）的扩充来解决。

2. 并发处理性能：该系统性能将随着客户端数量、数据库服务器

性能、管理信息种类和数量的变化而变化。以下是进行典型的文本类型数据（信息内容不超过 1M）管理操作的性能要求：

数据入库操作：≤3 秒；

数据修改入库时间：≤3 秒；

数据按照指定的样式进行显示页面合成时间：一般单条信息合成时间少于 2 秒，一般复杂检索的合成时间少于 5 秒；

数据检索时间：采用全文检索模式进行关键字检索的返回结果时间≤5 秒；

数据同步实时性（数据交换实时性）：数据同步服务器的响应依赖于网络传输速度，在局域网内（忽略网络速度的影响），一条数据（数据大小约为 10k）同步的时间在 3 秒内完成。

1.1.2.2.4 数据采集能力

作为政务数据共享交换平台的核心功能，数据采集在整个“互联网+政务服务”体系中尤为重要。政务数据共享交换平台提供强大的数据采集功能，具备多种数据对接能力，可应对多种数据采集业务场景。常用的对接方式包括：数据库方式、RESTful API 接口方式、Excel 导入方式、消息队列方式等。与各委办局调研过程中，可根据网络环境、业务规范、系统特点等实际情况，灵活选择数据对接方式。

1.1.2.2.5 数据交换中间件技术要求

信息交换平台由交换中间件、应用适配器系统组成，各部分设计要求如下：

1.1.2.2.5.1 交换中间件技术要求

交换中间件技术要求主要包括：数据传输要求、数据转换要求、可靠性要求、安全性要求、差错处理要求、Web 服务支持要求、跨平台要求、基本性能要求等几个方面。

1. 数据传输要求

在数据传输方面，交换中间件应能满足以下要求：

（1）传输协议

交换中间件必须支持 HTTP/HTTPS 传输协议。支持消息传输与文件传输。

（2）消息基础协议

交换中间件支持国家政务信息资源交换体系标准规定的消息格式。

（3）大文件支持

交换中间件支持大文件传输。

（4）消息交换模式

交换中间件支持消息主动发送、请求/应答、订阅/发布三种消息交换模式。

（5）消息路由

交换中间件支持消息路由的功能，包括基于消息内容的路由和基于消息头的路由。

2.数据转换要求

交换中间件具有数据转换的功能，支持数据格式转换、数据内容转换，提供图形化数据转换规则生成工具。

3.可靠性要求

交换中间件具有可靠传输的功能。通过断点续传、消息确认和消息重发机制，实现在网络发生故障、系统产生异常或发生崩溃的情况下实现数据交换“不丢、不错、不重”。

4.安全性要求

交换中间件支持安全的数据传输功能，支持消息加密、消息数字签名功能。

5.差错处理要求

交换中间件支持政务信息资源交换体系标准规定的错误处理功能，如身份验证失败、权限验证失败、消息超时、消息 ID 重复等错误处理功能。

6.Web 服务支持要求

交换中间件支持 Web 服务调用功能。

7.交换流程管理要求

交换中间件应支持动态交换流程配置，采用图形化的界面进行交换流程与交换内容配置管理。

8.跨平台要求

交换中间件具有跨平台特点，支持 UNIX、Linux、Microsoft Windows 等不同的操作系统，支持不同平台间平滑移植。

9.基本性能要求

交换中间件基本性能指标，如单位时间内传递消息、单位时间内字节流量、并发用户数量及在大负载情况下系统资源使用情况等满足交换平台的业务需求。

10.管理监控要求

（1）交换结点状态监控要求

中心交换管理系统能够监视中心交换服务器、部门交换前置机等各交换结点的运行状态。

（2）交换服务状态监控要求

中心交换管理系统能够监视中心交换服务器、部门交换前置机等各交换结点上部署的交换服务的运行状态，能够远程启动、停止交

换服务。

（3）交换过程监控要求

中心交换管理系统能够监视信息交换过程。

（4）交换日志管理要求

中心交换管理系统支持交换日志管理功能，如可配置的分级别日志记录功能、日志查询功能。

（5）远程部署要求

中心交换管理系统支持远程部署功能，实现交换服务及其配置文件的远程部署。

1.1.2.2.5.2 应用适配器系统技术要求

应用适配器技术要求主要包括：异构数据库支持、文件支持、图形化配置要求、适配器组件开发框架要求等几个方面。

1.异构数据库支持的要求

应用适配器支持不同类型数据库，主要包括 Oracle、MS SQLServer、Sybase、DB2、Mysql 等主流关系数据库。

2.文件支持的要求

应用适配器支持结构化文件与非结构化文件的读写，对 XML、EXCEL、TXT 等结构化文件提供文件内容解析功能，支持大数据文件的读取。

3. 图形化配置要求

应用适配器提供图形化的适配器配置工具。

4. 适配器组件开发框架要求

应用适配器提供开放的适配器组件开发框架，用户可以按照框架要求开发新的适配器组件，满足其个性化的需求。

1.1.2.2.6 平台安全保障

系统的安全性包括对物理安全性、服务器安全性、访问安全性、存储安全性、数据传输安全性上均做出了安全考虑，并针对系统本身的应用、保密安全，采取了权限分配、身份认证、访问控制、日志、备份恢复等相应安全措施，以保证项目的软件安全、应用安全、数据安全、操作安全、运行环境安全。

1.1.2.2.6.1 软件平台安全

计算机信息安全可通过安全体系结构来反映计算机信息系统安全需求和体系结构的共性，其构成要素是安全特性、系统单元及开发互联参考模型层次，在安全特性坐标中，描述了计算机信息系统的安全服务和安全机制，包括身份鉴别、访问控制、数据保密、数据完整、防止否认、审计管理、可用性和可靠性。排在安全特性坐标最前面的是：身份鉴别、访问控制和数据保密等。但安全是分层次的，且随各种环境的不同而有所变化。对于一个信息系统，怎样来考虑其安全技术策略问题，就要具体情况具体分析和对待。但无论如何，应当抓住

问题的实质，找出信息安全的最薄弱的环节，制定策略，加以防范。为保证系统的数据传输安全，通过身份识别（手机短信验证码）、权限认证等各种安全技术，保障合法用户才能访问软件，避免不同合法用户间会出现越权信息共享问题。

1. 数据安全保密性。满足在数据存储、传输过程中的安全保密性需求。政府在进政务服务批工作中涉及大量的敏感数据，在其处理过程中，特别是与各级单位数据交换过程中，要保证数据的安全保密性。

2. 数据完整性。满足在数据存储、传输过程中的完整性需求。在内部保证数据存储和传输过程中不被篡改和破坏；在与各级单位数据传输的过程中，保证数据不被篡改和破坏。系统具备完善的数据备份与恢复机制，支持数据的自动备份，还可以通过远程灾备中心进行数据备份，实现完整的数据备份。支持数据库导入、导出、修复等功能，保证系统在服务器掉电等意外情况发生时，确保数据的安全性以及系统地快速恢复。

3. 不可否认性。满足用户行为和系统行为不可抵赖性的需求。用户每天都利用数据中心处理大量的事务，事务处理过程的可管理、效率的可审计、行为的可审计等，需要行为的不可抵赖性来解决，本项目建设中要保证在所有数据处理过程中，办公人员行为和系统行为的不可抵赖，以便审计和监督。本系统针对管理人员的每一次操作都提供日志管理和审计功能，日志记录内容包括操作人、操作时间、操作等详细内容。日志的范围包括，对于与安全密切相关的用户登录事

件、访问控制事件以及身份认证、访问控制、重点操作等行为安全事件等应该进行安全审计操作，并记录系统安全日志。并提供对日志内容的检索和分析功能，保证在发生安全相关问题的时候能够做到追踪问责，通过对安全日志记录的查询和分析以及相关的审计操作找到安全问题的根源所在。

4. 对象和行为的可授权性。实现对数据资源的自主授权和访问控制的功能。针对本系统数据交换共享工作的特点，系统具有对对象灵活授权的功能，包括用户对用户的授权、系统对用户的授权、系统对系统的授权等，以及授权过程的审计监督。系统采用基于角色和功能模块的授权模式，用户只能访问其角色范围内容的数据，并通过对数据域的设置，只能访问特定模块的功能。

5. 统一信任与授权策略。对于涉及多个业务部门、若干业务系统的审批系统而言，安全性的实现不仅体现在各个部门、各个业务系统中，更重要的是在不同业务部门的不同系统实现互联后，如何保障数据、业务系统在互通后的信任、授权的一致性，因此在本系统中，必须建立统一的信任策略、授权策略，实现跨部门、跨系统的信任和授权服务的一致性，杜绝由于不同部门、不同业务系统不同的安全策略、不同的安全等级带来的安全漏洞和安全隐患。

6. 数据中心统一安全监管性。由于本项目涉及政府多个部门，因此需要实现数据交换、共享过程的可管理，实现对内部和对各级单位相关的业务处理的可审计性；系统中有大量的数据资源，为使这些资

源协同工作，需要实现对实体（用户或数据）进行统一的管理；系统对用户行为和系统行为进行记录和统计，对系统日志进行分析和统计，提供对用户和系统行为的审计监督。这种统一的安全监管必须以可靠的技术和严格的管理来保证。

1.1.2.2.6.2 运行环境安全

1. 安全管理，软件平台安全的基础是云平台安全，包括对虚拟安全设备镜像的管理。

2. 运维安全。软件平台运行所需的云平台日常运维管理，对系统运行状态、异常事件等各种安全事件进行监控、记录、维护。

3. 管理制度。软件平台运行所需的云平台管理制度的制定和落实、文档记录、以及应急预案。

1.1.2.2.6.3 数据备份

采用的数据备份机制为分表全库备份，所谓分表全库备份方式为数据库备份文件是以数据库表为基础单元进行备份。备份周期根据业务数据区别待定。为了节省空间，数据库的备份文件将会自动被压缩存放。数据备份文件将会被保存在多个物理磁盘中，维护人员也会每月将数据刻录成光盘进行外存。由于采用了分表全库备份方式，恢复数据可以恢复指定的数据库表，也可进行全库恢复。最大限度的保证数据的完整性与一致性。

备份类型包括全量备份及增量备份；全量备份不同于整体数据

库备份。完全数据文件备份是包含文件中所有已用数据块的备份。

增量备份是级别为 0 的备份，其中包含数据文件中除从未使用的块之外的所有块；或者是级别为 1 的备份，其中仅包含自上次备份以来更改过的那些块。级别为 0 的增量备份在物理上与完全备份完全一样。唯一区别在于级别为 0 的备份可用作级别为 1 的备份的基础，但完全备份不可用作级别为 1 的备份的基础。制定合理的备份策略，确保备份能够方便快速的从最新一次的备份进行数据恢复。