1.1 数据资源体系建设

鹿城分平台是区数字化改革的数据底座，是推动政府治理现代化的基础平台，公共数据是支撑数字经济、数字社会发展的重要资源。鹿城区已建成功能实用的公共数据平台，此次建设的内容是开发数据接口实现与市级公共数据平台（以下简称市级平台）对接。进一步完善鹿城区范围内个性化数据的归集、治理和安全管理能力，支撑县域治理现代化建设。

1.1.1 数据门户

1.1.1.1 分平台门户

鹿城公共数据分平台门户展示的内容包含但不限于以下内容：

- 通知功能展示消息和待办通知；
- 最近访问展示最终用户最近访问的功能菜单；
- 公告展示系统发布的公告信息；

我的申请展示最终用户当前申请的资源的审批状态信息，列表字段包括序号、申请内容、申请时间、状态和操作。

1.1.1.2 单点登录

单点登录，即用户只登录一次，在其随后访问所有的授权的资源时，都不需要再次登录，真正实现“一次登录、全网漫游”。

各应用系统无需自己实现登录，直接使用用户中心单点登录模块平台提供的访问接口，取得用户登录信息，根据登录信息判断用户的能否进行相关操作。

系统需提供多种单点登录集成方式，分别针对新建的应用系统和原有的应用系统；对于已有应用系统，可提供账号映射和转换服务，使用统一账号登录后先转换为原有系统账号再进行访问，降低原有系统的改造成本。

1.1.1.3 用户多登录方式支持

实现与浙政钉统一用户、组织体系的对接，可通过钉钉账号或扫码实现

单点登录。建立浙政钉用户、组织体系与易和用户、组织的映射，实现与省市县三级系统用户、组织的对接。

用户中心支持以下登录方式

- 用户名/密码

目前最常见的用户认证方式，使用方便，但无法证明是用户本人，用户名、密码存在被破解和盗用的风险，为了提高安全性，一般还需要配合验证码、输入次数限制、提示问题等安全机制。适合大部分非核心类的服务。

除了传统的用户名，还需要支持手机号、身份证等别名方式登录。

- 手机短信

这种方式安全性比较高，除非手机被劫持否则很难被破解。使用方便快捷，适合绝大部分场景，但是需要走运营商发短信，存在一定的使用成本。

- 钉钉扫码

这种方式安全性很高，操作方便；适合企业用户或公务人员使用。

1.1.1.4 公共数据平台报表

对数据目录、数据归集、数据开放等相关工作进行汇总展示。内容包括：

- 数据目录指标
- 统计数据指标
- 部门接口调用次数
- 对接接口数
- 数据开放情况

1.1.2 数据编目

根据鹿城市相关文件要求，区县级平台目录系统直接复用市级公共数据平台目录系统，市目录系统开设区县目录专区，县级部门统一利用市目录系统开展信息系统普查和数据目录编制工作，构建一体化数据资源体系。

1.1.2.1 数据目录的梳理服务

对鹿城各级部门本地自建、在用的信息系统的进行梳理和调研，形成信息化的数据资源目录。依托全市统一的公共数据平台目录系统，持续补充完善

信息系统要素。

1.1.2.1.1 制定数据目录调研表

数据目录编制的过程中，首先应制定科学合理的调研计划，根据业务部门的工作安排，对调研计划和步骤进行不断的调整完善；结合业务实际，给出合理的需求调研样例，便于开展工作；应结合鹿城的实际业务系统和温州市数据目录体系，给出数据分类、目录代码、元数据及编码情况等的合理建议，推动数据目录编制工作有效推进。

1.1.2.1.2 信息系统调研及分析

对鹿城的数据资产状态、信息系统现状等进行全面、彻底的调研分析，主要包括如下工作内容：对鹿城政府机关单位部门的数据状况开展调研，通过温州市大数据局的数据目录和数据状况的分析比对，初步形成鹿城政府机关单位数据可能存在的状况，然后制定详细的数据调研表，协助相关鹿城机关单位一同对数据状况进行调研分析。

1.1.2.1.3 信息系统梳理工具

采用信息系统规划工具，辅助政府部门梳理信息系统。

一、部门信息系统汇总表是对部门信息系统梳理的支撑表格工具。

部门信息系统汇总表 1　　表 5.6-1

序号	系统名称	系统应用部门	系统主要功能	系统的数据库位置	系统的网络接入情况	开发商名称	开发时间	备注
1	请填写	请填写	请填写	请选择	请选择	请填写	请填写	
2								
3								
4								
5								
6								
7								

二、部门信息系统汇总表是对部门涉及的信息系统梳理的支撑表格工具。

部门信息系统汇总表 2　　表 5.6-2

序号	职能域	业务过程	信息资源编码	信息资源名称	信息资源描述	资源类型	管理方式	隶属系统	共享方式	公开范围	提供/更新周期	交付方式	资源主题	服务对象	管理分类	备注
1	请填写	请填写	34050010/000001	请填写	请填写	数据库	系统	请填写	请选择	请选择	请选择	请选择	请选择	请选择	请选择	
2																
3																
4																
5																
6																
7																
8																

三、信息系统明细表是对部门信息系统明细梳理的支撑表格工具。

部门信息系统明细表　　表 5.6-3

	信息资源编码	34050010/000001	信息资源名称	
序号	指标项编号	指标项名称	对应数据元标记	数据类型
1	34050010/000001001	姓名	XM	请选择
2	34050010/000001002	性别	XB	
3	34050010/000001003	出生年月	CSNY	
4	34050010/000001004	家庭人口数	JTRKS	
5	34050010/000001005	保障人口数	BZRKS	
6				
7				
8				
9				
10				

1.1.2.2 数据目录管理服务

以梳理的数据目录为基础，利用温州市数据目录系统，提供鹿城目录管理服务，挂载鹿城数据仓内相应的数据资源，方便查询和掌握数据仓内现有的数据资源信息和可提供使用的资源及方式。根据本地信息系统情况，按照《公共数据资源目录编制指南》，在市级公共数据平台目录系统上进行公共数据的目录新增、删除、修改、查询等操作，针对已编目的公共数据及数据项进行动态管理。

1.1.2.2.1 分类管理服务

分类管理分为部门分类管理、主题分类管理两部分，方便用户查询数据资源信息。

（1）部门分类：部门分类支持新增、编辑、上线、下线、迁移、删除等操作，支持 4 级树型结构部门分类。

（2）主题分类：主题分类支持新增、编辑、上线、下线、迁移、删除等操作，支持 4 级树型结构主题分类。

1.1.2.2.2 目录管理服务

帮助目录内容提供者对所有的所在市级资源目录进行管理维护，主要包括目录注册、目录变更、目录撤销、导入目录等功能。目录的新增可通过目录注册和目录导入两种方式进行。同时可在目录管理内，通过填写目录分类、目录代码、目录提供方、目录名称等基本信息对特定目录进行精准定位。

1.1.2.2.3 目录审核服务

利用市级数据目录平台对报送的目录进行最终的审核。审核所有内容，重点关注上报周期、目录级别、字段等信息，审核通过后发布该目录，审核未通过则将该目录退回至数据资源提供单位。

1.1.2.2.4 目录发布服务

利用市级目录平台对审核通过后的目录进行最终确认发布。可查看所有目录发布情况和详细内容。如对于每一条待发布的数据目录进行状态的显示，可以显示其目录编码、目录名称、编制部门、上报周期、专题数据库、编制日期、修改日期、目录状态、建表进度等信息。主要功能如下：

（1）对于每一条待发布的数据目录进行状态的显示，可以显示其目录编码、目录名称、编制部门、上报周期、专题数据库、编制日期、修改日期、目录状态、建表进度等信息。

（2）对于每一条通过审核的数据目录，可以执行查看详情、建表、发布操作。

（3）建表功能需对表单中的字段做出验证，无字段的表单不可建表，建表可以刷新。已完成建表的数据目录可以发布，完成建表过程。

1.1.3 数据归集

根据温州市公共数据平台建设要求，市级平台与县级分平台按照责任分工，需要分别建立完善的数据归集、数据交换系统，以持续推进全市公共数据归集和交换工作，确保数据的完整性、有效性和及时性。

市区两级平台承担全市数据归集工作，全市公共数据统一归集到市级平

台。鹿城分平台重点做好区域范围内个性化数据归集，建立完善归集任务管理、数据同步通道、配置管理等体系，完成数据归集流程自动创建与运行，实现数据归集入库。鹿城分平台向市级平台定时报送任务执行情况和数据归集情况，形成全市统一的数据归集视图，实现数据血缘分析。鹿城平台应提供人工采集功能。人工采集通过定时生成采集任务给对口人员，实现数据的人工数据上报工作。

本方案提供数据交换系统、人工数据采集系统、内外网数据跨网安全传输平台建设和数据归集、数据仓建设服务。

1.1.3.1 数据交换系统

在用户原有数据交换系统（一期）基础上进行升级，保障原有数据仓数据安全，对于非实时归集的场景，可以按照交换方式进行数据采集。数据采集分人工采集和自动采集两种方式。数据交换平台提供人工采集服务，通过定时生成采集任务给对口人员，实现数据的人工数据上报工作。自动采集由业务部门主动将数据推送至前置机或数据交换平台主动到业务部门数据库抽取数据，实现定时数据归集。

数据交换系统的核心功能是任务调度，通过前置机可以抽取本地相应数据，并经过转换、加载。但是，由于数据之间存在依赖关系，导致各任务之间存在一定的执行先后限制，分组、多依赖、定时任务运行的任务调度系统能够定时开始任务，合理调度各个任务，实现并发，并且具有简洁的图形化管理界面。

1.1.3.1.1 任务调度系统

通过前置机可以抽取本地相应数据，并经过转换、加载。但是，由于数据之间存在依赖关系，导致各任务之间存在一定的执行先后限制，分组、多依赖、定时任务运行的任务调度系统能够定时开始任务，合理调度各个任务，实现并发，并且具有简洁的图形化管理界面。

（一）变量管理

全局变量，可以定义后续任务配置中可以使用的变量，方便配置，可避免后续因路径调整而导致全部任务需要重新配置的情况。

（二）周期管理

任务周期管理，用于定义任务生成的周期。可以根据 python 表达式自定义周期，包括每年、每月、每日、每周、每时、每分、每秒，以及更加复杂的周期。

（三）引擎管理

引擎作为整个调用系统的核心单元，具有执行任务的功能。不仅支持 kettle，还支持任何命令行相关的指令。对引擎的管理变得相当重要，一方面可以方便配置引擎，另一方面需要实时掌握引擎的运行状态。系统支持多引擎配置，不同任务安排不同的引擎执行，界面上的最后心跳时间，可以确定引擎是否正常运行。

（四）权限分组管理

权限分组用来管理任务的分组，不同的权限分组，标示不同的权限。权限分组没有任何逻辑上的强关联，只是对任务进行分组，方便管理。

（五）同步分组管理

同步任务，表示同一分组的任务，在同步分组中，必须同一周期内的所有任务都完成，才会进行下一周期的任务。同步分组的任务看成一个整体，体现了任务的前置性。

（六）任务配置管理

任务配置管理用于管理各个任务配置。需要定义运行时间、周期、权限分组、同步分组、运行程序、参数格式、引擎等参数，并且可以定义任务依赖关系。任务配置对 kettle 进行了单独的处理，采用 jdk 调用的方式启动 java 任务。通过复制任务，可以快速配置一个新的任务。

（七）任务实例管理

任务实例管理可以按周期重建实例、重建选择实例。对每一个实例可以执行重建、删除、强制执行、强制终止、查看日志等功能。

（八）任务图形化管理

任务图形化管理，通过图的方式展示任务间的依赖关系。通过条件筛选，可以快速定位任务。右键可以对每个任务进行相关操作，包括立即启动、重建实例、删除、修改状态、终止任务、查看日志、修改任务配置。

1.1.3.1.2 人工数据采集系统

手工数据采集系统是一个 excel 上报系统。当数源部门，信息化程度较低，没有信息技术团队提供数据库、接口等能力时，数据采集平台能发挥关键作用。

系统定义了 3 种角色，任务配置员，任务管理员，任务操作员。任务配置员，主要是系统维护者，需要了解 IT 知识，配置任务，包括归集任务周期、归集任务分组、归集任务管理员、归集任务操作员等。任务管理员，主要是部门负责人或系统维护者；可以看到任务的汇总统计、超时情况、完成情况。任务操作员，主要对任务进行数据上传操作。也可支持在线编辑。

任务定义了周期，每隔一个周期，会重新生成任务实例，要求任务操作员重新上传最新数据。任务配置页面，主要配置任务名称、任务周期、主管部门、管理人员、操作人员。

（一）任务周期管理

任务周期管理，允许定义周期，默认有每日、每周、每月、每年。每隔一个周期，会重新生成任务实例，要求任务操作员重新上传最新数据。

（二）任务配置管理

任务配置页面，主要配置任务，包括任务名称、任务周期、主管部门、管理人员、操作人员。

一个任务配置对应一个 excel 格式，因此每个任务需要单独配置字段，主要填写字段名、字段说明、是否可为空以及数据校验表达式。后续模版会以这个配置自动生成，并且上传数据会严格按照此配置进行数据校验。

（三）任务管理

任务管理页面，主要用于查看任务的完成情况，并且可以查看具体的完成情况。 在任务管理页面，可以看到任务的完成进度，当前还有谁未处理，并且可以通过钉一下的功能提醒对方。

在我的任务页面，主要是上传、在线编辑、预览等功能。上传采用引导方式引导用户上传 excel，在线编辑，则直接在页面上渲染数据，修改并保存。所有操作完成后，都应通过预览确保数据正确性。

1.1.3.2 数据归集服务

1.1.3.2.1 数据归集路径

区级部门系统由于所属网络不同，在数据归集时有同网归集和跨网归集两种方式。部门的信息系统中有很多由国家、省、市统一建设，数据也都存放在上级部门，鹿城区在归集这部分数据时可以采用条线归集和集中归集两种方式。来自社会公众的第三方数据结构化数据，通过接口调取或者离线批量数据直接导入的方式对第三方数据进行汇聚。

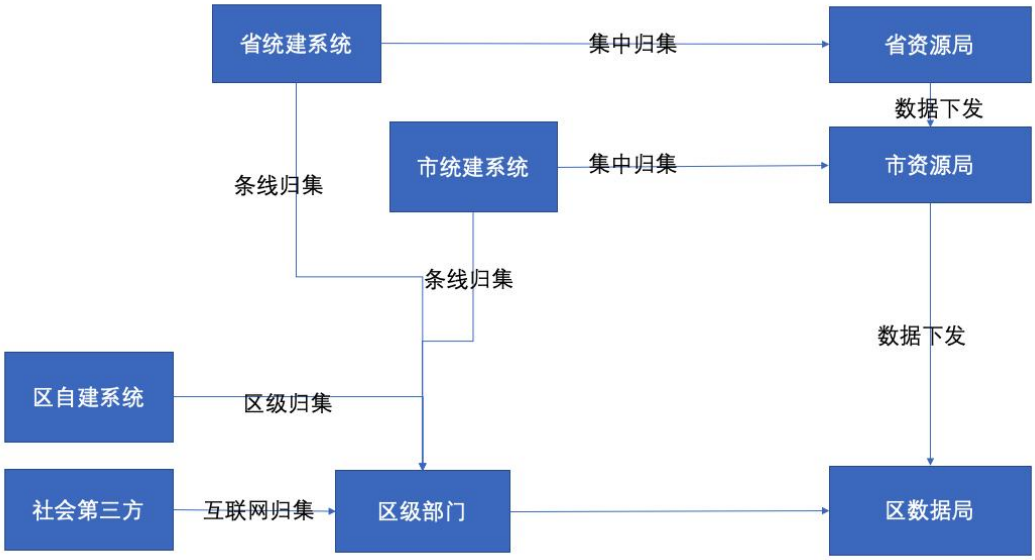


图 1.1–1 数据归集路径示意图

1.1.3.2.2 手工数据采集

鹿城部门在归集数据的过程中，存在大量 excel 手工处理的数据，在归集过程中缺乏部门技术人员支持，业务部门无法实现数据归集，造成工作力度不足，无法实现长效"精细"管理的局面。交换平台提供了 excel 归集的工具，部门按照固定模板填写数据，并通过文件传输系统将文件上传到文件服务器，交换平台实施人员通过配置自动解析流程将数据解析入库。

为部门提供一个 excel 数据在线数据填报和导入数据的直报平台，在数据上报的时候对数据的格式校验、数据的内容进行校验，保证数据入库的正确性和及时性，周期性将采集的数据入数据仓。

1.1.3.2.3 自动数据归集

数据归集根据数据编目清单，从数据中心的中心库中获取业务数据的过程。数据来源为业务系统和文件系统,归集方式为根据具体业务进行全量抽取或增量抽取,根据具体业务制定抽取的时间、频率、这些参数都是可配置的。归集方式包括全量归集和增量归集。

全量归集。该抽取方式一般在数据初始化的时候使用。将数据源中的数据原封不动的从数据库中抽取出来。

增量归集。我们采用基于时间戳的增量抽取，步骤如下：

1. 建立数据库连接

2. 定义一张数据字典表，定义需要进行处理的任务，其中主要包括业务数据库和目标数据库的表名、字段列表、以及条件等。

3. 对比源表和目标表的目前最大时间，抽取数据时间断为【目标表的最大时间】一【原表的最大时间】。

4. 根据设置的抽取频率循环抽取。

1.1.3.2.4 数据高铁服务

数据高铁（实时归集）采集实施包含对各业务系统的数据采集技术支持和系统对接联调。利用省市数据高铁（实时归集）服务能力，结合区县业务需求，直接读取业务系统库日志，从各业务系统向分平台推数据，转变为主动从业务系统拉数据，减少数据流转环节，大幅提升数据及时性、完整性。

1.1.3.3 数据仓建设

1.1.3.3.1 汇聚库建设

汇聚库按数据源分类，目标是与数据源数据相互对应，采取非历史表设计。从前置机加载数据，再进行数据清洗和转换，充分利用数据库特性，简化数据清洗的开发过程。

1.1.3.3.2 归集库建设

保证数据与部门提供的原始数据高度一致，并实现可历史追溯到任意时

间点。数据的历史追溯能力，不仅可以为后续基于时间维度的分析做准备，也可以有效不同时间点，回答查询结果不一致等问题。数据从汇聚层抽取，按照历史拉链表结构加载到业务镜像历史库。数据表按照业务逻辑关系，按第三范式设计。

1.1.4 数据治理

根据省市关于数据治理要求规范，鹿城区数据治理系统围绕提升数据准确性、完整性和及时性，让数据可用、好用、易用的核心目标，通过建立数据清洗规则、数据质量评估标准、数据问题反馈机制、数据使用标准等相关规范，对本地数据进行治理，形成数据治理闭环，提升数据质量。

数据治理系统应具备对归集数据进行清洗任务的管理能力，并可对清洗规则、数据标准进行相应的定义和配置，并建立一套相应的数据治理机制。一是建立问题反馈机制，形成问题发现/分析、问题处理、问题跟踪、问题评估，实现对问题数据的闭环管理。二是建立数据质量监控机制，根据清洗任务的运行情况及时作出告警，并通知数据处理负责人及时处理。三是参照《浙江省大数据发展管理局关于印发浙江省公共数据治理工作细则的通知》（浙数局发〔2019〕1 号），各县（市、区）应结合实际情况对数据清洗规则进行不断迭代优化，根据业务要求细化清洗颗粒度，降低问题数据误判率，提高数据质量问题的识别率，提升数据治理质量效果，并及时将经过治理的数据推送至市公共数据平台。四是数据治理系统及时向市级公共数据平台上报数据问题，市级公共数据平台应通过数据治理系统实现问题数据闭环处理。

1.1.4.1 清洗任务管理

通过统一数据库标准，表单配置规则，清洗任务管理这三个模块，系统管理人员可以以表和字段为维度进行数据清洗时的规则配置，并对清洗的任务进行统一的管理，包括任务周期、任务执行情况、任务异常等。清洗任务管理主要包括数据标准管理，表单规则配置和清洗任务管理。

数据标准管理展现了在库数据表的标准信息，展示数据源部门、数据库类型、数据库名称、IP 地址、端口号等配置信息，并提供链接测试功能，方便系统管理员根据不同情况的需求对数据库进行配置。

表单规则配置展示各数据表的清洗规则配置情况。

清洗任务管理是管理人员可以对平台清洗任务进行统一管理，在页面中展示数据表名、清洗状态、清洗任务节点、表清洗记录等。管理员可以对清洗任务进行启停、查看结果、查看规则等操作。

1.1.4.2 清洗规则配置

清洗规则配置页面将以列表的形式，展示各数据表的清洗规则配置情况。根据规则生成的类型有预制规则和自定义规则。预制规则是系统中预先设置的规则，自定义规则是用户根据实际需要和业务规范，配置好相应的核查标准规则。根据对数据的处理方式分为校验规则和处理规则，校验规则是对数据做校验，区分优质数据和问题数据；处理规则可以直接对问题数据进行纠正。

清洗规则包括但不限于：

1、对空值的校验和处理：校验数据空值，可以将空值替换或过滤。

2、身份证号码格式校验：对数据中的身份证号码按照国家标准校验是否正确。

3、验证数据正确：根据业务情况，校验数据范围，剔除不符合业务实际的数据。

4、数据格式转换：将不符合业务格式的数据转换成特定格式，比如日期字符串格式转换。

1.1.4.3 问题数据反馈

问题数据反馈系统是为了解决数据清洗过程中出现的问题数据反馈到数源部门。系统相关管理人员可实时跟踪数据质量情况，实现问题数据在各部门之间有序流转。做到问题数据的及时反馈，及时整改，提高问题数据解决效率。为鹿城区公共政务数据的质量的提升做支撑。在问题数据反馈系统中根据工单的状态分为我接受的工单、我提交的工单、需分派的工单和需审核的工单。

我接收的工单：问题数据工单发派到数源部门后，相应的部门用户登录问题反馈系统，可以在我接收的工单中查看需整改的问题数据，展示内容包括工单编号，工单类型，问题原因，联系人，联系方式等信息，可以对问题数据进行下载。数源部门对问题数据审核后，根据实际情况可以对问题数据进行整

改、退回、协商处理等。

我提交的工单：问题数据提交部门在我提交的工单页面，可以对新的数据问题工单发起提交，并可以对在权限范围内的问题工单进行提交、接收、待分派的情况进行查看。

需分派的工单：系统管理人员，可以查看待分配的问题工单的各种信息，如工单类型、问题来源、提交人，提交人联系方式等信息。在选择分派部门、处理时间、是否紧急等信息后分派工单。

需审核的工单：是各部门对问题工单协商和延期申请时产生的工单，反馈信息将在需审核的工单中进行统一展示。系统管理人员可以查看需审核的工单的延期原因和协商情况，并选择通过或驳回。

1.1.4.4 工单统计

工单统计是可以通过多种维度如数量、质量、异常、渠道等维度，对登录人权限范围内的数据问题工单进行可视化展示。各部门用户通过可视化展示可以方便的了解问题工单的各维度情况。

1、数量统计

从数量维度统计问题数据工单及其来源部门、处理情况等以可视化展示的方式对处理工单量、工单提交量、部门工单提交量、部门工单处理量等关键指标灵活的进行统计排名。

2、质量统计

可通过工单质量统计了解各部们的工单问题解决情况，督促问题工单部门及时对问题数据进行审核、整改。对部门问题数据工单解决率，工单处理时长等进行统计分析。针对问题工单解决率、异常问题占比率、处理时常等指标进行评分，直观展示各部门的数据整改情况。

3、异常统计

异常情况是问题工单在提交相关部门后，工单处理超时，处理未解决、延期未解决等情况。异常统计模块既是对这些异常问题的统计分析。

4、渠道统计

问题数据上报的渠道分为接口上报、手工填报、数据清洗系统三种渠道，渠道统计模块，既是对这些渠道维度的问题数据进行可视化分析展示。

1.1.4.5 元数据管理

元数据是关于数据的数据，元数据主要用于描述数据及其环境，它是在主题数据库建设过程中所产生的有关数据源定义，目标定义，转换规则等相关的关键数据。

在数据管控中，元数据管理对各数据的实体定义和流程管控管理两方面的元数据进行管理，并提供相应的对外服务。从数据源到后续的逐层加工以及稽核，元数据将各类的数据实体进行定义，约束；元数据管理贯穿于整个流程，提供相应的服务，并与各环节有效的互动。

元数据管理包括元数据基础数据管理和元数据应用，主要功能点包括元数据自动获取、元数据检索、数据模型管理、元数据管理、血缘关系、分工监控流程等。

1.1.4.6 数据质量检测分析

根据数据的完整性、准确性、唯一性、关联性、一致性和规范性对鹿城区各部门数据质量进行检测分析，建设查询问题数据、数据质量报告、数据质量评估模型、数据质量展现等模块。完成对鹿城区各部门数据质量的统一展示，对问题数据情况的展示和报告。

将数据质量评估过程中累积的统计数据进行汇总分析，形成包括数据整体质量报告、各部门数据质量报告，按照时间周期（年/月）生成相应的数据质量报告。通过报告的形式可以使大数据管理局和各部门领导了解鹿城区数据质量的情况，后续可以改进数据质量，提高处理问题数据的能力。

数据质量评估模型，按照浙江省大数据质量相关管理办法及要求，结合温州市实际情况，根据数据的完整性、准确性、唯一性、关联性、一致性、规范性，建立数据质量评估模型。方便对部门数据质量进行评定，了解各部门的数据质量情况，方便各部门对数据进行整改。

数据质量可视化展现，按照全局和部门两个方向，通过对数据质量的完整性、准确性、唯一性、关联性、一致性、规范性对鹿城区政务公共数据质量进行统计分析，并以可视化图表的形式对各部门数据质量进行展现。主要内容包括部门数据质量评分，数据质量排名，方便对各部门在数据质量方面进行考

核，督促各部门提升数据质量。

1.1.5 数据建模

本项目的数据建模是指对政府相关数据和业务进行抽象，通过整合各局委办数据，利用行业数据规范、统一业务口径、建立标准业务模型，构建用于数据查询、数据分析、数据挖掘的数据模式的过程。数据建模以数据治理为依托，是数据治理的上层逻辑和目标。

数据建模的项目实施包含下列四个阶段：

业务建模：汇总各局委办单位的业务，按照权力事项划分，对各个部门之间业务工作范围进行界定，理清各业务部门内部和部门之间的关系。重点分析部门数据共享和业务协同的现状和需求，明确流程并将其程序化。

领域建模：从全社会角度宏观分析并抽取政府关键业务概念，并将之抽象化。按照各业务主线聚合类似概念，细化分组概念，理清分组概念内的业务流程并抽象化。理清分组概念之间的关联，形成完整的政务领域概念模型。建立人口、法人、信用、经济、地理信息、事件等领域模型。

逻辑建模：将业务概念实体化，完善其具体的属性和概念之间的管理，输出逻辑模型。

物理建模：针对具体的数据库选型，将逻辑模型物理化，并根据数据库系统类型做出相应的技术调整。实现高效存储和快速访问。

数据建模包含下列四个数据加工阶段：

（1）过渡层（STAGE）

过渡层是用于统一采集来自于各委办局的数据，用于后续数据加工使用，不做长期保存。过渡层根据数据源类型的不同，采取不同的存储方式。结构化数据采取关系型数据库系统存储，非结构化数据采用网关加对象存储模型。如视频流数据采用视频网关加对象存储，文件和图像数据采用文件网关加对象存储。物联网数据采用物联网网关加关系型数据库方式存储。

数据抽取方式按照具体使用场景需求而定。结构化数据可采用周期性定时抽取或数据高铁准实时主动推送，视频数据采用主动推送方式进行数据捕捉，物联网数据可根据物联网平台主动推送或定时轮询方式进行数据采集，文件数据可利用人工数据采集平台定时由提醒相关人员手工上传。

过渡层数据的数据结构与数据源保持一致，并额外添加增量标识、采集时间戳、数据来源标识等元数据信息。

（2）沉淀层（DWD）

沉淀层是指将采集处理到过渡层的数据按照数据治理标准进行过滤、去重、校验、格式转换、编码转换等清洗和标准化操作之后，按照业务数据模型结构进行数据沉淀，数据结构和业务模型保持一致，但实现历史记录的保存，确保数据可以按照使用要求，重现任意时点的数据现场。

沉淀层将实施统一数据异常处理、标准化处理和数据质量监控，数据异常处理内容包括缺失值填充、异常值清洗、数据重复处理等，标准化处理内容包括字段取值列表标准化、字段取值格式标准化（例如日期、时间等格式）、字段数据类型标准化处理等。数据质量监控包括数据值分布、关键业务字段空置率等质控指标。

沉淀层将采用历史拉链表的方式保存数据，实现既能保留所有数据变动记录，又不保留重复数据，可以方便的实现历史任意时点的数据现场重现，并提供高效的查询能力。

数据沉淀层采用统一数据定义和命名规则。

（3）主题层（DWS）

在沉淀层之上根据领域建模的成果建立主题层。按照数据域对数据的划分，将数据归纳到一个主题下，构建星型模型和雪花型模型的事实表和维度表。形成编码统一、标准统一、概念统一的主题数据。

（4）应用集市层（ADM）

应用集市层是将主题层数据模型针对具体业务应用进行数据整合和提炼，按照不同应用业务场景的需求形成相应的数据宽表。应用集市层数据模型的建设将有助于提高业务指标或特征的开发效率。

应用集市层可以针对行业，如公安、交通运输、平安、城管、环保、卫健、

文旅等各政府部门的数据实现预定义部分常用数据集，为其他委办局在数据开发和数据分析应用提供基础。然后再根据具体需求整合其他数据集，以适应不用的应用场景需求。

应用集市层可以汇总数据，形成面向主题组织数据的数据立方体，通常是星状和雪花状数据，从数据粒度讲，它是轻度汇总级别的数据，数据存储方式可按照 ROLAP、MOLAP 或 HOLAP 存储，兼顾查询性能和存储效率，前端可采用 BI 工具实现数据立方体的多维分析和查询。

应用集市层可按需组合各类数据，形成数据宽表，通过数据挖掘工具，实现数据聚类、分类、回归分析、关联规则等机器学习算法，从数据中挖掘内在规律，为决策支持提供数据依据。制定出更加合理的指标为城管领域、财政收入支出领域、各产业生产消费统计领域、发改领域、市场监管领域、环保领域、信访领域、最多跑一次等领域服务。

1.1.6 数据共享

根据温州市公共数据平台建设要求，公共数据平台共享系统是全市数据接口共享的总枢纽，鹿城通过市统建的公共数据共享平台，接入鹿城数据和服务接口，实现省市县数据共享整体统一，鹿城数据管理独立。

1.1.6.1 数据共享分平台

1.1.6.1.1 数据共享分平台概述

数据共享平台是公共数据平台的重要组成部分。它负责将本地数据转换为服务能力，并整合上级或其他部门提供的数据接口，形成格式规范、权限统一、安全可控的对外服务能力，提供给其他应用使用，是公共数据平台的核心系统。下图为市区一体化数据共享平台架构图，其中微服务资源和鹿城管理模块需要在本地部署。

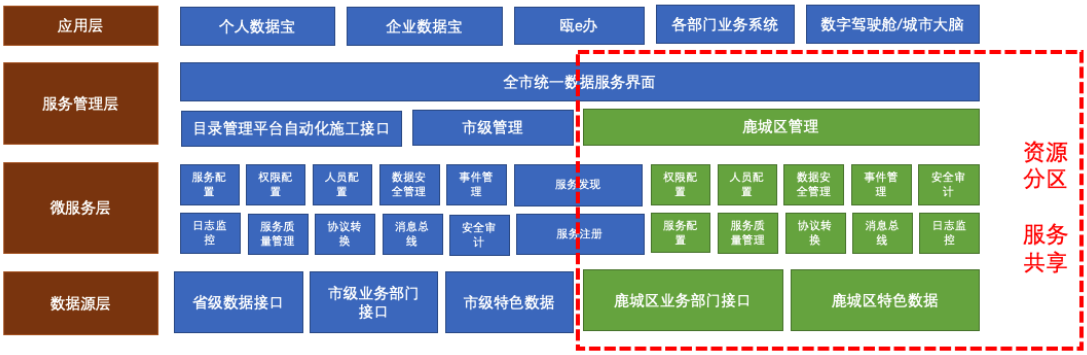


图 1.1–2 市区一体化数据共享平台架构图

鹿城数据共享分平台是对全市数据共享平台的拓展，通过建设鹿城管理功能，接入鹿城级数据和服务接口，实现全市数据共享整体统一，鹿城级数据共享独立管理的目标。

鹿城数据共享分平台的数据来源是归集的本地特色数据和鹿城自建的业务系统提供的数据接口。鹿城独立部署微服务资源，将数据和接口按照数据鲜活程度的不同，通过协议转换，封装成标准接口或消息服务，实现被动调用和主动推送服务。来自不同数源单位的数据将按标准进行规范化，注册到全市统一数据服务中心。同时和市级统一编目系统打通，形成以数据编目为驱动的数据管理体系，实现数据采集、数据清洗、数据共享申请、接口共享申请的全流程自动化。

鹿城对本地特色数据和省市下发的全量数据具有自主的权限控制，对本地数据形成的接口和市级已授权接口进行统一权限管理。鹿城数据管理工作从数据采集、加工、入库、整合、服务输出、权限分配，均保持完全的逻辑独立。

市区一体化数据共享平台由市、区分级维护、共同使用，在保证区级灵活性和自主性的前提下，可以有效降低区职能部门运维数据共享平台的压力，降低基础架构的投入。架构在平台上的优质应用案例，可以无缝推广到全市其他区域，实现市、区应用融合。

1.1.6.1.2 接入层数据源获取

数据接入层数据源可从多个渠道获取，包括本地业务部门系统数据库、本地物联感知网和温州市下发数据。各类数据的获取，需要鹿城相关数据管理部门与各相关政府单位协商，同意对数据的获取。

1.1.6.1.3 安全控制模块

安全控制主要围绕数据安全与服务器安全两方面。数据安全包括设置白名单限制部门的查询范围，设置黑名单限制访问敏感数据，保存完整的查询日志用户事后审计。服务器安全包括限定客户端 IP 访问，规定查询工作时间，设置每小时查询总次数最大值、每天查询总数据量最大值、查询间隔最小值等，登录时必须输入验证码并验证手机随机短信密码。

1.1.6.1.4 共享服务模块

共享服务以网站形式为提供系统管理、数据搜索、结果打印等功能。

1、系统管理

系统管理员可对系统进行管理，主要包括：登录认证、组织架构管理、角色管理、用户管理、菜单管理、部门库管理和接口管理 8 大模块。

（1）登录认证

已注册成功的管理员的用户名对应一个手机号码，在登录前需要输入用户名和验证码，并获取短信密码，短信密码由系统自动发送短信到绑定手机号，管理员输入信息全部正确后，成功登录。也可通过钉钉扫码登录。

（2）组织架构管理

部门组织架构通过树形图进行展示，可以增加或删除用户部门，添加部门时填写的优先级主要用于排序，默认从小到大进行排列，子部门优先级与大部门相同。

（3）部门角色管理

添加部门角色，可以给不同的部门角色调整不同的主页菜单的使用权限。勾选的权限表示该部门拥有此菜单权限，设置成功后对应的菜单会在主页显示，其余不会显示。

（4）用户管理

可以对用户进行增加，修改和删除操作。新增用户需要输入账号，姓名和手机号，并选择相应的部门和角色，便于查找和管理。同时部门树形图会同时更新，点击各个部门会显示部门下的用户。

（5）菜单管理

对主页的菜单进行管理，增加或者删除某项菜单。修改菜单管理的菜单栏，主页菜单会对应修改。

（6）部门库管理

将信息库放到归属的部门下，规范管理和操作。在接口管理中有一项归属部门库，方便接口与部门对应。先增加所属部门，再增加对应的信息库。

（7）接口管理

一个业务需要对应一个接口进行管理。添加接口时需要填写接口名称、接口类型、说明信息等内容，接口类型与对应业务的类型一致，主要为个人相关、企业相关、机构相关和其他 4 中，根据功能填选。在主页菜单栏的业务查询中按此分类显示。按照功能，接口主要分为两大类，即查询接口和编程接口，查询接口有界面显示，编程接口无界面显示。

另外可以对接口的权限进行管理。先选择角色，角色可在角色管理中添加。选择后先点查询确认已有的权限，在下方的接口清单里把需求接口的勾打上，添加权限， 添加的接口权限会自动保存

2、数据搜索与输出

用户可申请个人数据信息的查询，事前需提供纸质查询申请表，勾选查询数据项菜单，填写查询项总数并签字。持申请单，通过刷取个人身份证件，查询相关信息，并拍照上传查询申请表，通过身份证号码与查询工作挂钩的方式进行溯源。平台操作人员经用户同意后，选择数据源（包括部门、数据库等）进行查询，输出查询结果。

各认证与查询信息到位后，鹿城大数据共享服务平台可直接对查询结果打印输出。

1.1.6.1.5 安全日志审计

鹿城数据共享平台操作人员查询的各个环节将通过全流程日志记录其行为，并对查询参数和查询结果进行加密留档。安全日志可用于查询结果比对、事后审计与历史追溯。

1.1.6.1.6 安全浏览器定制

鹿城数据共享平台使用独立客户端定制浏览器进行系统管理和数据访问，

能够做到防插件与无缓存的功能。市民需要查询个人信息时需要持本人身份证，通过纷纷正读卡器进行认证，才能读取所需的信息。另外，限制运行平台的客户电脑，只有特定通过认证的电脑才能运行平台，操作人员在使用时通过电脑硬件采集到指纹，通过指纹进行授权部署。

1.1.6.2 统一支撑平台

1.1.6.2.1 统一运行监控

为运营团队人员提供管理工具，支持服务发布和运行监控，作业调度管控，数据资产管控和安全审计等。不仅需要对数据库、数据服务等资产的运行状态进行管控，还需要实时把握能力开放情况，通过动态监控和预警帮助运维管理人员随时掌握系统运行状态，提前预防及处理问题。

1.1.6.2.2 统一消息服务

基于消息队列的消息发送服务，提供短信、钉钉、系统内等方式。

定时推送消息，支持指定时间点、指定相隔多少时间，以满足不同业务场景的需要。

通过正则匹配，不同业务系统，只能发送规则正则表达式的内容，以达到对垃圾消息进行拦截的目的。

1.1.6.2.3 统一日志服务

基于 elk 实现日志的收集、查询、分析。

通过监控日志文件，实现不同业务系统的日志的统一采集，将需要采集的业务系统接入进来，并且在业务系统端主动发送相关的日志信息给存储组件。支持几乎任何类型的日志，包括系统日志、错误日志和自定义应用程序日志。

提供基于 Web 的图形界面，用于搜索、分析和可视化存储在系统指标中的日志数据。用户创建他们自己的数据的定制仪表板视图，还允许他们以特殊的方式查询和过滤数据。通过工具使用，可以实现采集日志的查看，并且支持日志的分析。

1.1.7 数据资源管理及分析工具

1.1.7.1 海量数据搜索分析系统

海量数据搜索分析系统能够实现对数据资源池上的数据进行全库全表“一站式”搜索和分析、支持并行计算、检索引擎等先进技术，提供适用于结构化数据、半结构化数据、非结构化数据的高效、便捷、智能的数据搜索服务，并支持信息的进一步钻取。

搜索分析系统对于搜索到的数据可以进行全景展示，既可以输出摘要信息，也可以查看其详细数据。摘要结果可按不同的信息关联度进行排行，并且支持摘要结果统计指标的自定义配置。详情展示可通过表格、地图、饼图、八爪图等多种可视化方式进行搜索对象详情信息的全维度展现，同时支持相关联主题数据的层层钻取，通过超链接的方式实现主题数据快速切换，引导用户进行相关联数据的快速探索。

1.1.7.1.1 大数据搜索分析首页

大数据搜索分析首页是入口页面，在页面完成首次搜索的输入功能。

（1）搜索主题设计

紧密贴合用户需求，按照业务域对所有搜索资源进行维度划分，所分维度显示于搜索框之上。每个维度又设计若干主题，用户在每个维度下，自定义配置主题，进而支持用户按主题进行数据检索。另外，搜索平台也会对各主题搜索频度进行统计排行，对搜索频度高的主题，通过“常用”模块呈现于搜索框之上。

智慧云搜搜索框，支持用户选中主题后，自动计算出该主题下的搜索信息条数，方便用于及时获取信息总量；另外，搜索首页所呈现的维度和主题，均可根据用户需要进行灵活、可视化的自定义配置。

（2） 搜索框功能设计

- 信息自动补全

在关键词输入过程中，系统后台会根据当前用户所输入的字段，结合系统关键字检索历史，自动补全搜索关键词。

- 搜索字段选择

支持不同主题下，对搜索字段的进一步过滤。用户在当前主题下，对关键词所搜索字段做进一步的选择，可以单选也可以多选，对于没有选择的字段范围不予搜索。这种过滤方式可以进一步缩小搜索范围，过滤无效信息，使搜索结果更有针对性。

- 搜索模式选择

支持精确搜索和模糊搜索两种搜索模式。精确搜索模式下，其输入关键词需为完整的字段，不能有空格，搜索结果也需精确匹配输入的关键词；模糊搜索模式下，搜索结果只需要包含部分关键词，且支持多关键词搜索。

- 多关键词搜索

当搜索模式为模糊搜索时，搜索框支持多个关键词的输入。多关键词搜索包括两种组合方式，与、或。当多关键词之间以“空格”分割、或者输入“且”时，则数据检索需要匹配所有的关键词；若关键词之间输入“或”时，则数据检索只需匹配一个关键词即可。

（3）全文搜索功能设计

目前，智慧云搜搜索框所具备的搜索能力如下所示：

- 单个的中文字/词、英文字/词、阿拉伯数字及相关短语；
- 空格分隔的两个及以上的中文字/词、英文字/词、阿拉伯数字及相关短语。其中非中文字词部分支持不完全匹配输入，企业注册号、身份证号码、报关单号等由英文字母和阿拉伯数据构成的字符串需完整输入，当输入的词全部存在时，即命中结果。此功能多用于已知多个碎片化信息时的搜索；
- 逗号分隔的两个及以上的中文字/词、英文字/词、阿拉伯数字及相关短语。其中非中文字词部分支持不完全匹配输入，企业注册号、身份证号码、报关单号等由英文字母和阿拉伯数据构成的字符串需完整输入，当输入的词存在一个，即命中结果。此功能多用于根据企业编码、报关单号、集装箱号等的批量搜索；
- 更多语法输入请参考该页面右上角的“搜索帮助”；
- 单击精确可切换为模糊查询，默认为精确查询；

➢ 对于报关单等分类，还可以在搜索前设置精确匹配条件。

1.1.7.1.2 搜索分析摘要页

摘要结果显示页展示搜索到的多条信息。每条信息涵盖标题部分和摘要部分，点击标题部分进入该信息的详情信息。同时支持所显示字段在后台系统中的自定义配置。

（一）分组统计

针对枚举类型的字段，支持按搜索结果条数进行分类统计。

（二）搜索结果排序

支持“按关联度排序”、“按时间顺序”、“按时间逆序”等多种搜索结果排列方式。所排序字段可在后台系统进行配置。

（三）取词搜索

支持用鼠标在搜索结果中选中若干个词，然后按所选词进行重新搜索。

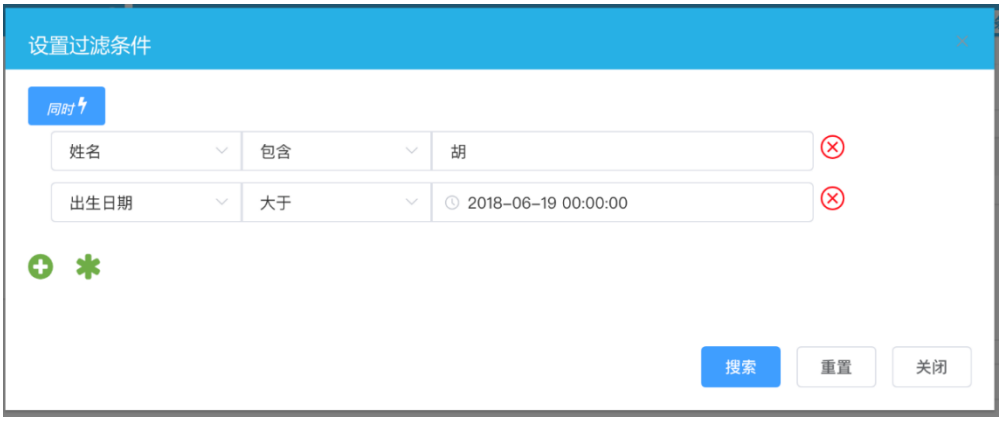
（四）展示分页

当搜索结果数较多且无法在一页中显示全部信息时，对所搜索到的摘要信息进行分页显示。

（五）条件过滤

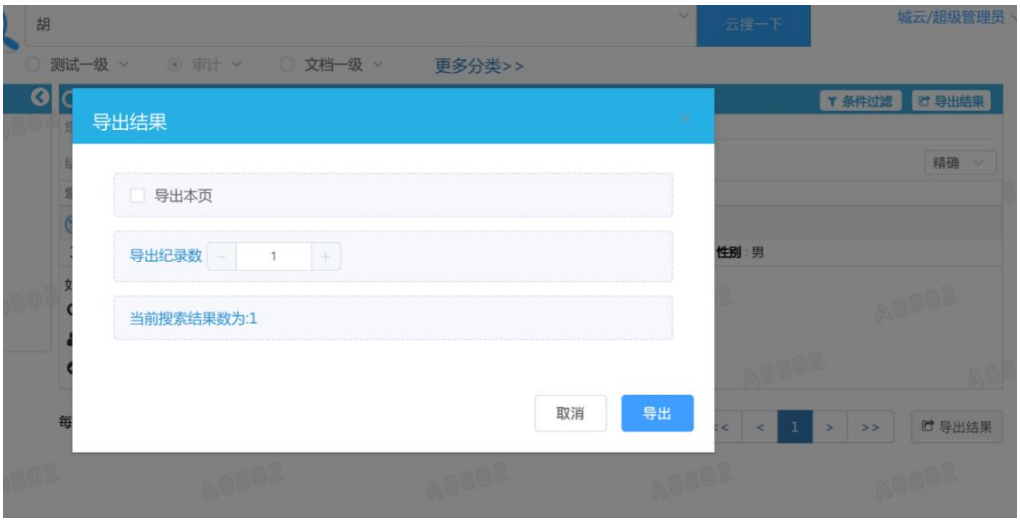
支持对搜索结果进行二次筛选，二次筛选支持如下操作：

- 按字段关系（与、或）进行多个字段过滤；
- 字段名称过滤；
- 按字段关系（等于、包含）进行多个字段过滤；
- 单过滤条件进行过滤；
- 多过滤条件进行过滤。



（六）导出

支持对所搜索的摘要结果保存为本地文件。所导出字段可在后台系统中配置。



（七）关联搜索

支持在关联数据表中，对搜索关键字段展开关联搜索，并自动统计出所关联信息条数。

1.1.7.1.3 搜索分析详情页

搜索分析详情页是搜索平台的核心内容，支持从多个维度、以多种可视化方式，展现被搜索对象的详情信息、统计结果、分析研判以及关联关系等。

详情页实现对所搜索对象的全方位信息展现。所显示模块由表单形式、表格形式和图形式构成。支持后台对显示方式、显示内容、显示格式的自定义配置，同时支持数据字段超链接配置。

（一）基本信息

搜索对象基本信息通常展现搜索对象的基本属性。按照智慧城市建设常用的维度划分，一般涵盖人、企、房、部件、事件等。对于以人为维度的对象，基本信息可以涵盖姓名、年龄、性别、民族、身份证号、学历、工作等信息；对于企业为维度的对象，基本信息涵盖企业名称、法人、注册地址、人数、联系方式等；对于以房为维度的对象，基本信息涵盖房屋地址、经纬度、面积、性质、所有人等信息；对于部件为维度的对象，基本信息涵盖部件编码、位置、作用、所属单位、责任人等信息。其他维度对象的基本信息可根据实际情况，进行自定义配置。

（二）详情信息

详情信息是基本信息的补充和完善，主要用以对企业的某类属性进行详细而完整的展现，如表单形式对企业工商注册信息的完整展现；或者表格形式对法人所担任企业信息的完整罗列等。

（三）关联关系分析

智系统会对搜索对象进行关联关系分析，并以八爪图的方式，直观显示多类相关信息，每条相关信息既可以信息统计的方式进行展现，也可进一步下钻为新的详情展示页面。

（四）信息钻取

智慧云搜详情页支持信息的进一步钻取，钻取使信息的获取面更广、更深。钻取既可下钻至对象某属性的详细表格，也可下钻至其他主题对象。

（五）可视化展现

通过地图、表单、表格、柱状图、饼图等多种可视化展现方式，对搜索结果的详情信息进行全面的可视化展现。

- 表单形式

单个记录信息显示为表格形式，主要显示由搜索结果列表页进入时显示的基本信息和相关扩展信息。

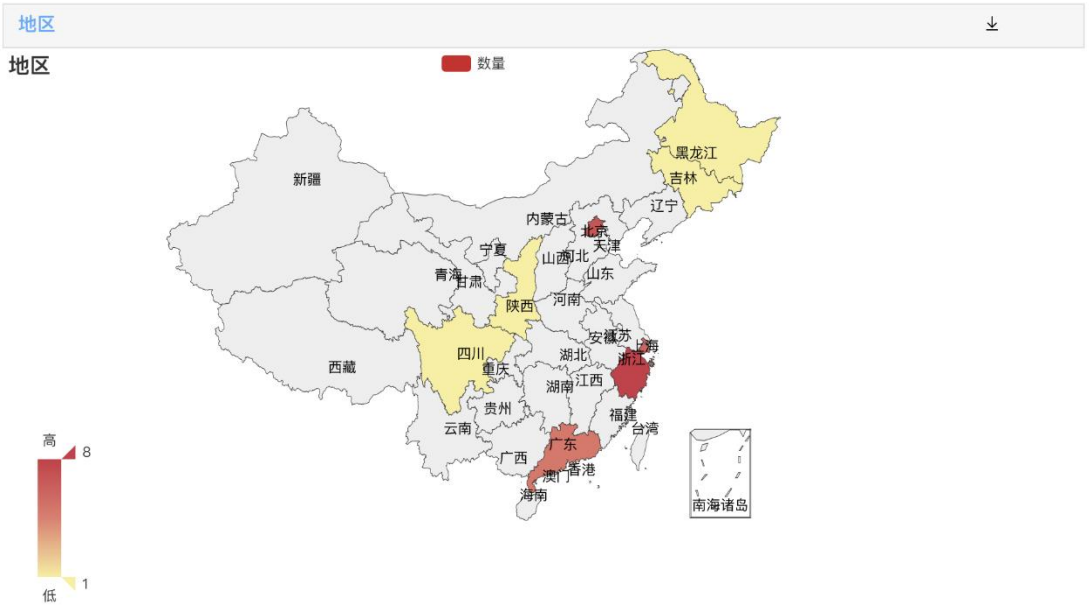
➢ 表格形式

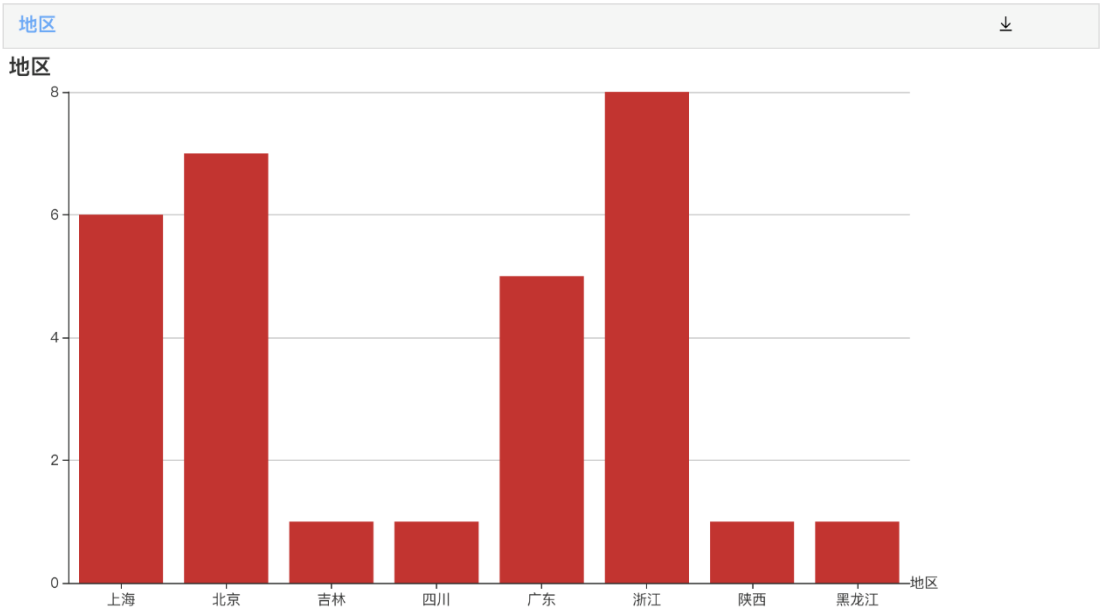
多行记录信息显示为带有分页的表格的形式。

序号	企业名称	法人姓名	组织机构代码	地区
1	浙江AB投资有限公司	迟为国	78883116-X	浙江
2	网易（杭州）网络有限公司	丁磊	60958075-X	浙江
3	杭州网易梦幻科技有限公司	丁磊	60958075-X	浙江
4	网之易信息技术（北京）有限公司	丁磊	60958075-X	北京
5	杭州朗和科技有限公司	丁磊	60958075-X	浙江
6	CY科技（中国）有限公司	蒋忆	59958075-X	浙江
7	广州博冠信息科技有限公司	丁磊	60958075-X	广东
8	广州网易信息科技有限公司	丁磊	60958075-X	广东
9	广州网易计算机系统有限公司	丁磊	60958075-X	广东
10	广州市凌怡电子科技有限公司	丁磊	60958075-X	广东

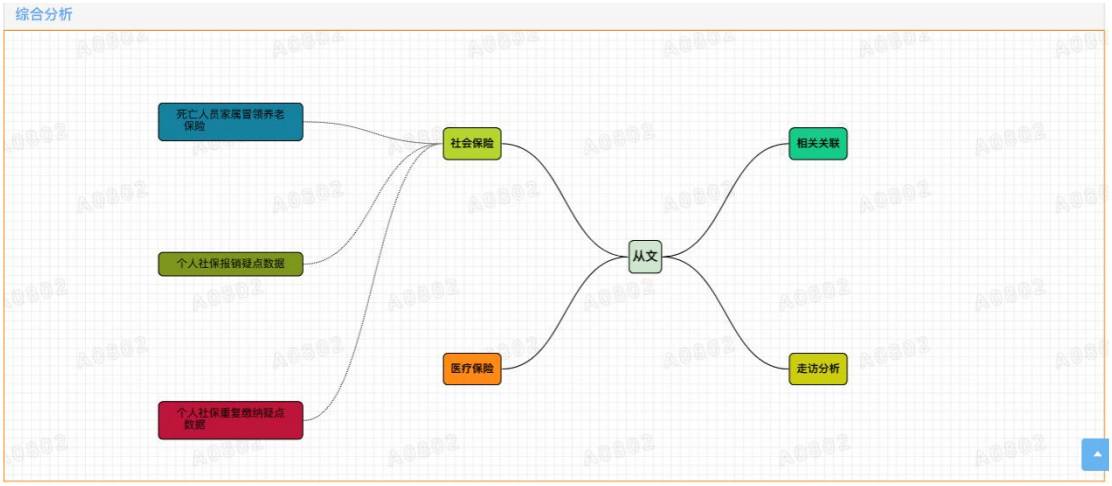
➢ 图形形式

图形形式主要包括地图、柱状图、饼图等显示方式。




地区
地区
8
6
4
2
0
上海
北京
吉林
四川
广东
浙江
陕西
黑龙江
地区

- 八爪图


综合分析
死亡人员家属冒领养老保险
个人社保报销疑点数据
个人社保重复缴纳疑点数据
社会保险
医疗保险
从文
相关关联
走访分析

- 图表联动

实现图形与表格中数据间的联动。如上图所示，点击柱状图中的“浙江”，表格中的数据自动按照地区“浙江”进行过滤。联动功能如何关联需要在后台管理系统中进行过滤。

1.1.7.1.4 系统后台配置

搜索系统后台管理的主要功能是把不同系统的数据汇总到搜索系统，提供统一的搜索和展示功能。 接入数据的一般步骤包括：数据注册 —> 主题配置 —> 详情配置。

搜索系统后台首页显示搜索集群的概况，包括文档数、文档占用空间以及搜索集群节点概况。

（一）数据注册

数据注册是搜索和展示的基础。数据注册的目的是把不同系统中的数据接入到搜索系统中，并按照搜索系统的要求完成对表结构的配置，配置搜索结果的展示方式为数据的搜索展示做准备。

搜索系统支持 mysql、oracle、impala、ads、gbase 等关系型数据库系统数据的接入。

- 数据库连接

数据注册实现新建数据库添加功能。

- 注册数据表

数据源配置完成后就可以获取数据源中数据表的信息。将数据源中的表结构信息读取到系统中并完成表结构的配置。

数据表注册详情如下图所示，并支持按表名查询出待注册的数据表。

普通的 mysql、oracle 数据库在数据表注册时不输入关键词可以列出待注册的数据表列表，在 odps 数据库中由于数据表较多需要给出表名称的前三个字母才能列车数据表信息。

- 表结构配置

按照搜索系统的要求，对已注册数据表重构表结构。

表结构配置说明：

- 字段名（中文）：字段的中文名称，在搜索结果列表页中显示该名称。
- 字段类型：
- 整数、长整数、小数、高精度小数：数值类型，系统自动识别无需修改
- 字符：证件号码、编码和较短的无需分词的字符串
- 短语：较长的一段描述，需要分词的字符串
- 日期(年月日时分秒)： 年月日 时分秒 完整的日期时间类型
- 日期(年月日)： 年月日
- 日期(年月)：年月
- 主键： 构成主键的多个字段
- 检索字段：在输入关键词进行搜索时是否在该字段中搜索，一般选择证件编码、描述短语等字段。
- 搜索标题：搜索结果中该字段显示在搜索标题部分。
- 搜索摘要：搜索结果中该字段显示在搜索摘要部分。
- 排序：在搜索结果列表中可以选择按照某个字段进行排序即是该配置项，一般为时间类型字段。
- 分组：搜索结果列表左侧的分组统计功能，选中该项把该字段加入其中，一般为枚举类型的字段比如状态等字段。
- 任意匹配：对于编码类型的字段比如身份证号码，如果我们需要输入部分编码需要匹配则需要选中该项，比较影响性能请谨慎选择。
- 是否导出字段：搜索结果列表页中可以导出搜索结果，这里设置是否导出该字段。

- 顺序：控制字段在搜索列表页中的前后显示顺序，数字越小显示靠前。

数据注入

待表结构配置完毕后，即可编辑任务信息，进行数据注入。

（二）主题分类配置

后台支持对一级主题、二级主题分类和名称的自定义配置。

- 新建主题分类

进入主题新增页面，新建搜索主题。

数据表类型表示数据来源为数据库中的数据表，可以绑定主表或关联表。文档类型表示该主题下存放的是用户上传的文档文件，无法绑定数据表。

- 主题分类关联数据表

支持二级主题和二级主题分类与某一数据表进行关联。

当用户进行关键字主题搜索时，关联表作为关联信息显示于主题搜索列表下方。

（三）展示块配置

搜索系统中，每一个详情展示页称作看板，看板中的每一个展示部分称作展示块。

- 新建展示块

实现展示块与后台所配置模型的关联，并支持 sql 语句对模块进行编辑和修改。

- 编辑修改展示块

支持对已建展示块信息的编辑和修改。

- 配置显示方式

支持对展示块配置显示方式。

- 表单显示方式

表单显示方式主要用作一条记录信息的展示。

- 表格显示方式

表格显示方式用来显示多条记录。

- 图形显示方式

图形显示以柱状图、线图、饼图、地图的方式形象的展示数据。

- 标题显示方式
- 展示块过滤条件

（四）看板配置

每一个详情页称为看板，看板由多个展示块拖拽布局而成。

- 新建看板
- 添加展示块
- 配置导航
- 过滤条件配置

看板的过滤条件表示该看板可以从外部接收的过滤条件。每一个过滤条件需要配置该过滤条件如何作用到展示块。

- 级联配置

级联用于配置两个图形和表格以及图形显示方式之间的联动。比如柱状图地区点击该柱状图，该地区的值作为过滤条件传递给其他展示块。

- 八爪图配置

八爪图实现与预先配置的看板之间的关联。

- 保存看板

实现对所配置看板的保存与发布。

- 详情页关联

详情页对应为一个看板，创建详情页的过程即为给一个数据表关联一个看板的过程。

1.1.7.2 可视化建模平台

数据可视化建模平台基于数据底盘库，支持数据建模分析、统计分析、挖掘分析、趋势分析等，实现基于业务分析的可视化建模。结合实际的业务场景，通过在画布中对相关组件进行拖拽和关联等。

本平台充分利用大数据云平台的分布式存储和计算能力，能对超过 1 亿条以上的建模数据进行及时处理。支持大数据云平台上数据库和其他通用数据库，包括 ODPS、ADS、RDS、HBASE、以及关系数据库（Oracle、MySQL、SQL Server 及其他）等。

1.1.7.2.1 建模工厂

模型管理设计

建模平台主要分为组件区、工作区、属性区三部分。

- 支持选择需要分析的数据源（数据源的具体内容包括选择表、选择字段）。

- 支持选择数据处理工具对数据进行加工分析，如数据分组、数据过滤等。
- 支持选择数据可视化工具对模型运行结果进行展示，如列表，图形展示。

1、新建模型

支持新建模型，并填写模型的基本信息，基本信息包括模型名称、模型分类、选择对应数据库（公共分析库查询数据快，但只有近 3 年的数据；大数据计算库数据全，但查询速度较慢）、是否预警模型（若选择为“是”后，则该模型会成为预警模型。发布应用中预警模型选择状态为“是”时才能选到对应的预警模型）、使用说明需手动填写或选择（使用说明可不填）；点击确定，页面自动进入组件。

2、修改模型

支持修改模型功能，点击修改按钮，可进入修改模型页面，可修改内容包括“模型名称”、“模型分类”、“使用说明”等等。

3、删除模型

支持删除模型功能，点击删除按钮，弹出删除模型确认提示框，点击确认，则完成删除模型操作。

4、复制模型

支持复制模型功能，选择需要复制的模型，点击从当前模型新建按钮，则弹出复制模型弹出框，复制后修改复制模型界面信息完成复制。

5、模型放大缩小

支持对模型放大缩小操作，可缩小模型对其行进全局展现。

6、运行模型

支持对模型的运行功能，对所建的模型进行调试。

7、运行节点

支持对模型的节点运行，节点运行可查看当前节点的字段内容。

8、节点删除

支持对某模型的节点进行删除操作，点击删除节点按钮，选中需要删除的模型节点，点删除当前节点按钮，即可删除当前节点。

9、获取节点 SQL

支持获取模型节点 SQL 语句，即可查看当前节点的 SQL 语句。选中一个节点，点击获取节点 SQL，页面会弹出当前节点的 SQL 语句，方便查询当

前节点的 SQL 运行。

10、成为正式模型

创建的模型点击成为正式模型后，成为正式应用，不在以草稿模型存在，在开发者中心中可发布为公共应用。

1.1.7.2.2 我的数据

数据管理

数据管理支持新增两种格式数据：文件与加载应用的结果数据。

文件格式目前支持上传 csv 类型，选择上传的文件，必填项名称，填写相应描述，读取出文件中的数值，红色星号为必填项，也可选择导入到不同的数据源中，供建模工厂中使用，选择所需填写的数值即可。

加载应用则需先对应用进行创建。填写名称，选择应用，并对应用进行配置，同时对调度进行配置，包括优先级、执行周期。

支持对文件和加载的应用进行数据管理。包括对数据的：编辑，启动，查看数据，数据导出，删除操作。

- 编辑：可以重新编写一定内容。
- 启动：文件格式没有启动操作按钮，应用可以启动和暂停。
- 查看数据： 查看创建的文件和应用里面的数据，进行展示。
- 数据导出：下载数据保存成文档格式。
- 删除：删除选中数据。

分类管理

分类管理点击新增按钮：名称为必填项，输入信息后保存。

新创建的场景分类操作列分为：修改，维护，同步，删除。

- 修改：修改选择的场景分类的内容。
- 维护：对选择的场景分类进行数据快关联。
- 同步： 同步数据。
- 删除：删除选中数据。

调度查看

调度查看主要对数据管理中创建的文件进行调查查看。状态栏分为三个模式：成功，运行中，失败。操作列：详情可以查看选择的日志（详情只可看

状态为失败的记录）。

1.1.7.2.3 我的应用

应用查询

支持对本地模型和在应用商城分享的模型进行查询，支持模糊查询与精确查询。

应用运行

支持对应用进行运行操作，点击运行后，跳转应用运行结果页面。

应用开发

支持对应用进行二次开发。选择一个要二次开发的应用，点击二次开发按钮，输入密码，跳转到建模平台，对该应用进行二次开发。

1.1.8 数据开放服务

根据市里数据平台要求，县级平台不再单独建设数据开放网站，由市级数据开放网站设立县级开放专区，鹿城依托市级数据开放网站实现数据开放需求，持续做好本地个性化数据的管理和监测工作，不断提升本地数据开放比例，跟踪数据开放利用效果，为数据开放应用提供高质量数据，推动优秀数据开放应用落地孵化。

鹿城将结合市级公共数据目录编制工作，基于市公共数据开放目录，增加本县特色开放数据，汇总形成市级数据开放目录，明确数据项名称、数据格式、更新周期、开放属性、开放单位等。与民生密切相关、社会迫切需求、经济增值潜力显著的公共数据，应当优先纳入开放重点。

1.1.9 数据安全

建立系统、科学、高效的数据安全管理制度，健全审批流程体系和评价体系，促进安全工作标准化、流程化、规范化开展，各部门切实增强数据安全主体责任意识，加强公共数据共享开放安全管理，建立全局、可控、智能、主动的数据安全体系，明确安全责任边界，实现涵盖数据采集、传输、存储、处理、使用、销毁等数据全生命周期安全管理，防止数据篡改、泄漏和数据滥用。

1.1.9.1 数据采集与安全审计

浙江省大数据发展管理局于今年 10 月发布了《浙江省大数据发展管理局关于征求浙江省公共数据安全日志审计规范（征求意见稿）意见的函》，该规范定义了公共数据的日志采集、存储和审计等过程。

在数据采集汇聚与安全审计层面，通过主动扫描、流量采集、日志采集等方式为数据安全管控平台提供运行所需的全部安全数据，通过统一的数据标准和接口标准归集到数据安全管控平台，实现数据的集中式存储分离，并为上层数据安全应用提供数据支撑。

同时，在进行各类数据采集等运维操作的过程中，需要避免越权操作、权限分配不合理等问题，因此需要专业化的运维审计能力进行支撑。

1.1.9.1.1 运维审计服务

由于信息化建设、业务不断拓展，在各信息系统中各种网络设备不断增加，对目标主机的管理必须经过各种认证和登录过程。在某个主机及账户被多个管理人员共同使用的情况下，引发了如账号管理混乱、授权关系不清晰、越权操作、数据泄漏等各类安全问题，并加大了运维审计的难度。

在上述场景下，运维审计服务以运维审计系统为抓手，通过多样化的运维手段，遵循《网络安全法》以及等级保护相关法律法规，对系统权限进行合理分配，通过全面的运维协议，让运维审计工作无死角推进。

运维审计服务的审计规则包括但不限于下列内容：

（1）未经授权的登录；

（2）用户账号和权限变更；

（3）操作系统的启动、停止信息；

（4）系统服务和配置修改；

（5）特殊权限使用和操作。

运维审计服务具体能力包括：

- 用户分权：支持多种用户角色：超级管理员、部门管理员、配置管理员、审计管理员、运维员、审计员、系统管理员、密码管理员，每种用户角色的权限都不同，为用设立不同的角色提供了选择，并且满足合规

对三权分立的要求。

- 集中授权：通过集中授权，梳理用户与主机直接的关系，并且提供一对一、一对多、多对一、多对多的灵活授权模式。
- 单点登录：托管主机的账户和密码，运维人员直接点击<登录>即可成功自动登录到目标主机中进行运维操作，无需输入主机的账户和密码。
- 统一审计：对所有的操作进行详细记录，并提供综合查询功能；审计日志可以在线播放也可以离线播放，所有的审计日志支持自动备份和自动归档。
- 自动运维：全面自动化运维，实现自动化的运维任务并将执行结果通知相关人员。
- 命令控制：集中命令控制策略，实现基于不同的主机、不同的用户设置不同的命令控制策略，策略提供命令阻断、命令黑名单、命令白名单、命令审核四种动作条件。
- 工单流程：操作人员向管理员申请需要访问的设备，申请时可以选择：设备 IP、设备账户、运维有效期、备注事由等，并且运维工单以邮件方式通知管理员。管理员对运维工单进行审核之后以邮件方式通知给运维人员；如果允许，则运维人员才可访问；否则就无法访问。
- 系统自审：不仅对操作行为进行审计，还能对系统自身变化信息进行审计，并且形成系统报表分析。

1.1.9.1.2 流量采集服务

在数据流转过程中，涉及到多层次的数据流量，各类流量中的敏感信息需要在监测时做到快速定位，因此需要专业化的流量采集服务对数据流转时的流量进行采集分析，同时具备网络入侵攻击报文检测能力，匹配到恶意流量时立刻触发告警，记录入侵攻击事件。

针对关键节点流量进行采集，通过流量镜像方式进行业务威胁分析，实时感知威胁情报分析和全流量监测威胁数据，实时发现违法行为和攻击行为。包括不限于入侵检测告警数据采集、僵木蠕毒检测告警数据采集、邮件攻击检测告警数据采集、DNS 流量检测告警数据采集、DDoS 异常流量检测数据采集等。采集结果上传至数据安全管控平台，最终通过大数据深度分析挖掘威胁趋

势，感知安全态势以及重点单位安全预警。

采集的流量对接数据安全管控平台，采用流量分析技术可对出口流量进行快速抓包协议还原，深度解析各种传播数据，发现各种已知威胁攻击，实现包括入侵检测、僵木蠕毒沙箱检测、异常流量分析在内的各类威胁监测，感知网络安全态势。

通过对上述流量的汇聚，从而形成数据安全管控的基础数据支撑资源，形成各类数据资源库供上层应用使用。

1.1.9.1.3 日志采集与审计服务

针对各类安全设备告警日志进行采集，主要包括入侵检测系统、堡垒机、入侵防御系统、Web 应用防护系统、漏洞扫描、防火墙、数据库审计等网络安全设备的告警日志，通过 Syslog 形式将日志信息输送至数据安全管控平台，实现安全日志数据的统一收集。

（一）日志审计

日志审计层面，当前仍存在海量日志无法管理、各类日志无法关联等问题。日志审计服务针对当前场景，对各种安全事件日志(攻击、入侵、异常)、各种行为事件日志(内控、违规)、各种弱点扫描日志(弱点、漏洞)、各种状态监控日志(可用性、性能、状态)、安全视角的事件描述：事件目标对象归类、事件行为归类、事件特征归类、事件结果归类、攻击分类、检测设备归类，进行统一存储，满足《网络安全法》针对日志留存 6 个月的明确要求。

日志采集范围：针对大数据平台上层业务系统及政务外网安全防护的安全设备告警日志进行采集，主要包括入侵检测系统、堡垒机、入侵防御系统、Web 应用防护系统、漏洞扫描、防火墙、数据库审计等网络安全设备的告警日志，通过 Syslog 形式将日志信息输送至综合日志审计系统并对接数据安全管控平台，实现安全设备数据的统一收集。

日志审计服务具体能力包括：

- 日志接收：收到相应的数据报文后，转换为相应的格式标准（SYSLOG 报文转换为字符串格式，SNMP TRAP 报文转换为 SNMP PDU 数据格式），并且附加上来源地址信息。
- 日志解析：进行规则库的加载，加载各种日志格式的解析、映射

定义，加载完成后，进行日志的解析处理。

- 聚合处理：对解析后的日志，根据标准化的通用事件格式，对各个标准化字段，进行信息的直接映射、非直接映射处理。

（二）数据库审计

数据库审计服务以数据库审计系统为基础，对数据库访问行为进行实时审计，对数据库的恶意攻击、数据库违规访问进行全面记录，为数据追踪溯源提供技术支撑。

数据库审计服务的审计规则包括但不限于下列内容：

（1）未经授权的数据库非法连接；

（2）未经授权直连数据库后的增删改等操作；

（3）用户的关键变更操作，如增删改用户及其权限；

（4）数据库服务启动和停止；

（5）数据库系统核心配置文件的修改。

数据库审计服务具体能力包括：

- 查询分析：对 SQL 报文、数据库命令执行时长、执行的结果集、客户端工具信息、客户端 IP 地址、服务端端口、数据库账号、客户端 IP 地址、执行状态、数据库类型、报文已经报文长度等内容进行审计并提供查询分析。
- 报表分析：审计结果支持塞班斯报表、综合分析报表、性能分析报表、等保参考分析报表、语句分析类报表、会话分析类报表、告警分析类报表和其他报表输出。
- 规则管理：利用各类规则模型，区分 SQL 注入攻击、漏洞攻击、账号安全、数据泄露和违规操作等攻击类型。
- 权限管理：支持四种不同的用户角色，分别是超级管理员、安全管理员、系统管理员和审计管理员，不同的用户角色对应不同的权限。

1.1.9.2 数据安全应用

数据应用层主要将数据分析结果和原始数据进行整合，为用户提供数据应用服务。数据应用层主要包括资产与风险管理模块、数据全生命周期防护模

块、数据安全监管监测模块以及数据安全运营模块。

1.1.9.2.1 数据安全防护服务

数据共享交换等平台实现了数据大集中的同时，也导致了数据的风险大集中。如何识别数据风险，进而采取有针对性的数据安全防护控制措施，来缓解、转移、规避数据安全风险，是数据安全防护体系建设必须考虑的一环。以数据为核心，构建覆盖数据全生命周期的安全防护体系，在数据采集、数据传输、数据存储、数据处理、数据交换和数据销毁环节采取相应的安全技术防护措施来保障数据安全。

（一）采集阶段

采集汇聚数据的数据源鉴别，避免因混入伪造、仿冒或者非法数据源，而导致数据分析结果失真情况；越权、违规采集数据的审计；采集数据的定位和溯源；控制采集数据的一致性、完整性，保障数据质量等。

（二）传输阶段

内部网络传输节点的非法登录和越权访问控制；传输加密，防护外部传输链路被监听、嗅探，导致的数据被篡改、窃取；合规业务或合法身份调用和访问敏感数据过程中的数据泄密防护等。

（三）存储阶段

逻辑存储、容器的自身安全管控；存储数据的合法调用和访问控制；数据违规、随意存储检查；按不同的数据级别定义不同的数据备份策略，避免误操作、宕机、停电等灾害导致数据丢失和损毁等。

（四）处理阶段

终端使用中通过对外设、即时通讯等管控，防止数据泄露；数据分析业务调用生产环境中的敏感数据进行脱敏处理；严格的访问控制，防止 BI 分析人员越权、违规操作数据。

（五）交换阶段

数据加密，防止链路被监听、嗅探，导致数据被篡改、窃取；敏感数据、涉密数据交换、发布的脱敏和涉密检查。

（六）销毁阶段

重要存储介质维修/报废前缺乏数据清除管控；对清理后的存储介质进行

敏感数据检查，避免擦除、销毁不彻底等。

1.1.9.2.2 数据安全管控平台

（一）单点登录

数据安全管控平台集成多项数据安全防护能力，通过单点登录服务，可让用户实现在多个数据安全管控服务中，只需要登录数据安全管控平台一次，就可以访问所有相互信任的其余数据安全管控服务，使得操作运维更加便捷，数据安全能力无壁垒。

（二）数据安全管控

数据安全管控平台涉及大量的政务信息以及市民个人隐私信息，一旦泄露，将对城市的安全防护体系造成不良影响，对个人信息所有者带来严重的安全隐患。因此需要采取有效的数据安全防护措施来应对。

目前窃取信息的手段越来越高明，使得通信信息和重要数据更容易被窃取,因此很有必要采取相应的手段确保信息的安全。法律和制度及管理手段上加强数据的安全在一定程度上可以保障数据的安全，但是，针对黑客和不法入侵者，如果利用密码技术对数据信息进行加密，可以达到对重要信息隐藏保护的目的，非常有必要应用在数据安全管控平台的建设中。

1、多因素认证

目前数据库在使用过程中，大部分是通过单独的账号密码登陆，账号和操作人是否对应，是否有过度的访问数据，是否有越权访问并导出重要数据、敏感数据等行为很难控制，因此我们采用多因素认证进行管理。

多因素认证管理提供用户身份鉴别处理机制，运维人员、开发人员、运维工具、开发工具、业务程序等访问数据资源库的过程中，必须经过失败鉴别处理和多维度身份认证处理。

基于多维身份认证与多因素访问控制的身份鉴别技术可以帮助数据管理员、驻场合作伙伴、运维、开发人员、业务程序进行身份统一管理，运维合规，隔离敏感数据，使运维操作更加规范、透明、可控。

通过种类繁多的应用程序访问控制，实现对应用程序进行敏感标签标记、实现高粒度的应用程序审计和操作访问白名单控制。利用 Ukey 安全客户端同时支持多种 SQL 工具访问控制（NAVICAT、TOAD、SQLPLUS、PLSQLDEV）

和其他应用程序访问身份鉴别。

通过多维度、多因素的身份鉴别处理，有效防止核心数据库的账号被暴力破解。防止核心数据库访问身份被假冒风险，如果出现暴力破解、彩虹表尝试数据库主机破解，能够通过多因素认证服务有效的禁止该操作行为，锁定该设备 IP 地址，断开会话连接，使其无法尝试连接。

多因素认证对于密码泄露和暴力破解过程中采用锁定终端的方式，而非锁定数据库账号密码的方式，采用锁定终端的方式优势在于不影响正常的业务会话活动。

采用锁定账号密码的方式，在出现密码攻击，将会直接将该账户会话锁定，这种锁定方式会将正常的业务会话和黑客会话一起锁定，从而导致正常业务同样不能继续使用。采用锁定终端的模式，一旦检测到密码攻击后，能够精确检测到是哪个终端进行的密码攻击，从而有针对性的锁定该终端不能正常进行登录操作，而不影响业务的正常访问。

2、密码托管

将数据库账号密码托管到数据安全管控平台，开发、运维人员在连接数据库的过程中无需直接输入数据库账号密码，只需要通过数据安全管控平台选择需要登录的数据库实例名，数据安全平台会自动填充需要登录的数据库账号和密码，因此无需将真实账号密码告知使用者，真正做到账号密码安全可控。

3、数据库权限管理

在数据分平台的建设过程中，平台建设承建方、业务厂商、第三方代维均会提供驻场开发、运维人员，通过数据采集、清洗工具、运维工具直连数据库进行数据操作，该过程中这些技术人员都能够接触政务业务系统的原始数据和数据分平台归集的原始数据，因此在对这些技术人员所需要的数据库权限进行托管，通过数据安全平台进行权限细分和最小化管理。

4、敏感数据访问控制

在对通过多因素认证允许合法的用户或者身份接入数据资源库后，需要对数据资源库的具体操作过程进行访问控制管理，主要从特权账号权限访问控制、行数返回控制、访问频次控制、高危操作控制等方面入手，全方位保障访问控制体系的安全性。

为了解决 SYSDBA 和 DBA 等用户拥有数据库中访问任何数据的最高权限带来的安全隐患，利用数据安全平台隔离 SYSDBA、DBA、Schema User、其他 any 权限等特权，使其权限最小化，只能访问授权范围内的敏感表格数据，有效的防止黑客入侵、账户提权等操作。

针对敏感数据访问操作，需要得到安全管理员的事先授权，并通过安全管理员授予的 USB-KEY 与安全证书的授权机制验证身份再访问敏感数据，最大化降低可能涉及的风险。

根据全流量的协议解析，包含数据请求、返回数据解析、跨语句、跨多包的绑定变量名及绑定变量值的解析，可实现返回行数阈值管控、修改和删除行数控制。

5、返回行限制

提供基于敏感表格访问的返回行控制技术，避免相关人员利用合法的语句导出大量用户敏感信息，最大限度控制敏感信息的安全性。同时能够对大量返回行或更新行等事件做出告警、能够对频繁的相同语句做出告警。

支持对用户修改、删除的行数进行管控，安全管理员可预先设置最大可接受的被修改行数的阈值，一旦操作语句中的行数高于阈值，则语句无法正常执行，有效保证数据的安全。

6、访问频次限制

针对一定时间内频繁访问数据的行为进行访问频次控制，有效保证数据安全，可以设置每秒或者每分钟时间内访问数据的频次，超出访问频次的访问行为将被阻断。

访问频次限制能够有效防范一些违规人员不定期批量获取数据并将获取的数据集合后泄露的情况。

7、危险操作防范

除了访问过程中的权限控制以外，还有一部分的危险性操作如 Drop Table，Truncate Table 等操作是数据库面临的巨大安全风险。充分考虑实际应用的灵活性，当运维人员必须进行某些危险性操作或者需要访问敏感数据时，可提交临时授权工单，由安全管理员进行逐级审批后方可进行操作。

8、危险操作恢复

合法的用户在合法的操作过程中，不可避免会产生误操作行为，支持对表格 Drop Table，Truncate Table 的动作进行数据进行恢复，用户一旦发生误操作行为，安全管理员可在管理端页面进行语句追踪，找回误删除的数据。

9、应用防假冒

数据分平台在使用数据资源库时除了使用应用工具以外还会有各种运维工具，比如 PLSQL、Navicat 等工具，而这些工具一般都是在网络上下载的破解程序或者绿色程序，但往往破解程序或者绿色程序会被黑客进行注码操作，比如曾经网络上流传的一个破解 PLSQL 工具，黑客对该工具的 AfterConnect.sql 脚本进行篡改，将原本应该为 0 字节的脚本文件变为 27kb 的脚本文件，脚本里增加了以下几个存储过程和触发器：

- 存储过程

1. DBMS_SUPPORT_INTERNAL
2. DBMS_STANDARD_FUN9
3. DBMS_SYSTEM_INTERNA
4. DBMS_CORE_INTERNAL

- 触发器

5. DBMS_SUPPORT_INTERNAL
6. DBMS_ SYSTEM _INTERNAL
7. DBMS_ CORE _INTERNAL

而这些存储过程和触发器的效果是一旦连接到数据库就会运行触发器不断的增加日志，导致系统使用越来越慢，最后崩溃；甚至是当检测到数据库创建日志达到 1200 天后直接清空数据表的。

为了防范该类恶意程序或者假冒程序的使用，我们将常用的工具进行 MD5 或者 HASH 值识别匹配，当再次利用工具连接数据库时，会进行攻击特征匹配，识别真实应用特征，防止人为恶意将其他的应用改成业务系统应用，防止假冒应用访问数据库，进行非法操作，从而有效的控制应用程序、应用工具的使用的非法入侵。

10、动态脱敏

为了有效的防止用户的敏感数据信息泄露，支持数据的动态脱敏功能，

对未授权的账户访问敏感数据实现动态脱敏功能，支持对字符串类型、数据类型、日期类型数据脱敏，实现随机、转换、遮盖方式实现对动态数据的脱敏效果，防止第三方人员接触重要的敏感数据信息和业务的个人隐私数据，提高运维安全。

动态脱敏内置动态脱敏屏蔽策略，用户可以根据实际使用需求选择屏蔽策略，同时支持自定义屏蔽策略，可以根据用户的实际需求来定义屏蔽策略。

11、地址仿真

目前业务、开发、运维等访问数据库，均采用数据库真实地址及端口进行连接，这种方式将数据资源库的真实地址和端口暴露在外，一旦有人恶意在政务外网中通过扫描工具扫描或者爆破就很容易获取数据资源库的相关信息比如地址、端口；具备地址和端口后黑客就能够进行进一步的攻击，因此我们采用反向代理的方式将真实数据资源库的地址和端口仿真变换后提供给业务、开发、运维等场景下使用。

（三）数据脱敏

1、静态脱敏

在大数据流转、交换、开发测试、共享开发等场景下，会存在数据从生产区域流动测试区域、终端区域、外部区域，外部两个机构间互相调用数据，数据共享交换，数据回流，远程运维等场景。数据从高安全域流动到低安全域，数据泄露的风险变高了。

对流动数据的保护，经过验证的有效的方案是在源端进行保护，比如脱敏。把脱敏后的数据共享给别人，是一种比较有效的保护方案。然后再结合审计和水印来实现数据流转过程的追溯。

数据脱敏兼容各类主流数据库，包括 Oracle、SQL Server、Mysql、DB2、Sybase、Informix、RDS、PostgreSQL、ODPS、Cache、MariaDB、TiDB 等关系型数据库以及达梦、南大通用、人大金仓、巨衫等国产主流数据库，支持 HDFS、Hive、Impala、Teradata、Greenplum、MongoDB、FusionInsight、ODPS 等大数据敏感源，支持 Kafka、Redis 等消息队列，支持 Oracle dmp、Excel、CSV、TXT、DBF、DEL、DICOM（医疗影像文件）等各种数据文件敏感源以及 RSS、XML 等内容共享或数据传输文件。

为简化脱敏工作，防止遗漏敏感项，脱敏系统预设丰富的敏感对象，包括个人隐私、金融财务、商业机密、医疗数据、相关证件等，在进行数据脱敏时，将会对设定的敏感对象按照规则进行脱敏，且支持根据实际需求自定义敏感对象，完善敏感对象库。

基于预设敏感对象自动发现，支持包括字段信息、数据信息匹配等灵活的配置方式，利用各类敏感信息规则，主动发现、分析、评估和梳理敏感信息，建设敏感性评估知识库，通过数据脱敏策略进行管理，整体构建敏感数据地图，同时支持样本数据浏览和确认、人工审核确认。

支持 20 多种脱敏算法，通过不同的算法组合进行脱敏，保证满足脱敏后数据的一致性、完整性、安全性。

支持将敏感数据变形、漂白后，得到的数据之间是保持与源数据一致，支持主外键一致、业务关联一致、有依赖字段的敏感信息脱敏一致、多次脱敏结果保持一致，整体保证脱敏前后一致性，保持数据间的逻辑关系。

脱敏后最大限度的保证数据真实性，确保交付可用、可靠的高质量数据。数据脱敏后可以保持数据原始特征，保证开发、测试、培训以及大数据利用类业务不会受到脱敏的影响，达成脱敏前后的一致性。数据脱敏后仍然保持业务规则的关联性，包括主外键关联性、关联字段的业务语义关联性等。

为了适应不同数据脱敏的应用场景，在保持数据原始特征及业务一致性的基础上，提供多种数据脱敏算法：

- 随机映射：随机生成符合数据原始特征的数据；
- 固定映射：根据用户设定的密钥，将最小数据单位根据映射算法做固定映射；
- 替换：根据用户设定的替换字符，对数据的某一段内容进行替换；
- 加减值：对数值在一定范围内做加减值；
- 范围随机：对数值在一定范围内取随机值；
- 截断：将数据根据设定长度进行截断；
- 截取：截图数据中的某一部分；
- 加密：通过 MD5、SHA1、DES、RSA 等算法对数据进行加密；
- 格式化脱敏：根据数据的格式对数据进行切分，以保证数据的原始特

征。

2、数字水印

当数据共享给第三方，即意味着丧失了对于数据安全管控的能力。数据脱敏服务通过对于数据进行水印标记，在保证数据适度安全的情况下保证第三方数据可用，并可在第三方发生数据泄露后，可以通过泄露的数据追溯到数据提供给了哪个人或者企业，从而进行数据泄露的追责。

- 安全性：嵌入在原始数据中的水印是不可除的，且能够提供完整的版权证据。数据水印不会因为数据的某种改动而导致水印信息丢失的能力，数据水印能保持完整性或仍能被准确鉴别。
- 透明性：在原始数据中嵌入水印标记信息不易被察觉，不影响原数据的可用性。
- 溯源能力：从水印数据中溯源水印标记信息的能力。
- 低错误率：误判率低（误判分两种，一是数据无水印标记时，却检测到了水印存在；二是加了水印标记信息却没有检测出水印的存在）。

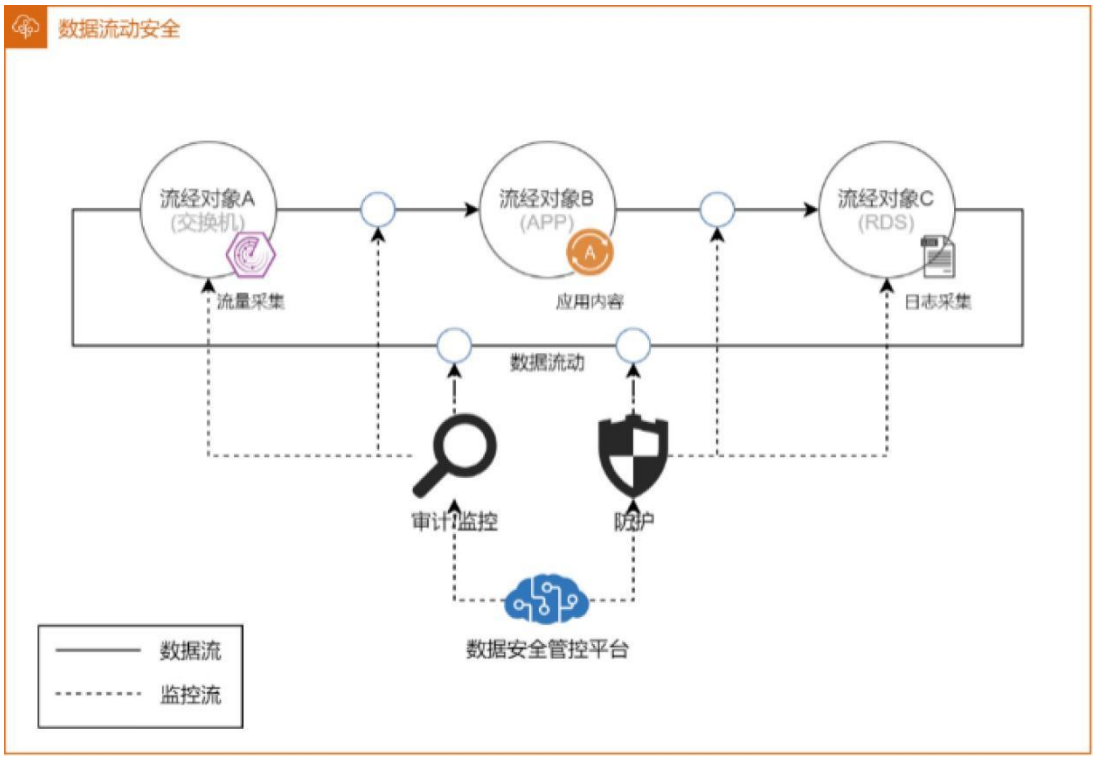
1.1.9.2.3 安全监管监测

数据安全管控平台以保障数据合规运营需要进行相应的监督管理，平台对数据安全的监管从数据层面、业务层面、用户层面和基础设施层面考虑，全面涉及到数据自身、计算流程、共享交换的参与方和数据储存安全，提供多维深度合规监管能力。

基于业务的数据安全监管监测

基于业务场景对其数据流经的节点按其属性和特点进行归类，根据数据的流向，利用数据流节点进行可视化的拓扑配置和安全监管。平台对数据监管从边界的数据流出流入开始，直到内部数据的流转均进行监控管理，保障数据的操作都可以被记录并审计，同时平台具备敏感数据的访问监管功能。

对于业务系统或场景的数据流进行梳理，对数据流经的节点结合人员账号，建立数据流程可视化图谱。基于 AI 自学习能力，可以辅助业务人员快速建立业务系统或业务场景的数据流。基于业务场景的数据安全监管监测如下图所示：



（一）边界数据安全监管监测

信息系统承担数据资源汇聚与共享的任务，涉及各单位数据接入与流出，需要在边界建设统一的数据信息安全技术保障体系，通过大数据分析、人工智能、安全监测等方式进行安全政务信息系统平台的数据进出的安全监控。平台边界数据监测模块对政务信息系统的进出流量进行实时监测，及时发现大流量数据外传等异常行为，并通过异常告警模块报警。平台边界数据安全监控通过数据库操作日志分析、接口调用日志分析、边界及关键节点流量解析等进行关联分析。并利用出口安全设备联动来实现阻断防护。

1、数据库操作日志分析

在数据库前端部署数据审计系统，监控数据库数据操作及返回数据的情况，明确有多少数据流入/流出政务信息系统。同时数据库审计系统也可以保护业务库的安全，及时发现数据库攻击、数据库后门等行为。

2、边界及关键节点流量解析

在政务信息系统边界及数据流关键节点旁路部署流量解析系统，通过解析 IP 报文字段及流特征信息，获取数据包的来源、去向、使用的协议等信息。流量解析获取的主要信息如下：

基于数据包字段解析：

源 IP 地址、目的 IP 地址、源 MAC 地址、目的 MAC 地址、源端口、目的端口、协议、报文内容及数量、开始及结束时间等。

3、接口日志分析

通过对共享平台等 API 接口日志进行分析，对 API 接口调用的名称、用户信息、IP、调用频率、数据内容、数量、和流量等进行分析。

（二）敏感数据访问监管监测

政务信息系统里汇聚了大量的数据，这些数据有不同的安全等级，数据分类分级从隐私安全与保护成本的角度出发，对数据进行分类和等级划分，进而根据不同需要对关键数据进行重点监管。通过重点节点数据访问的监控，并通过各种策略模型发现异常的访问，比如规定敏感数据只能由指定的用户从规定的 IP 地址访问，当某个用户的 IP 发生改变时，及时报警，因为有可能该用户的账号发生盗用；结合用户权限尤其是拥有特权账号的运维人员，对其访问铭感数据进行监控，一旦发现存在越权访问高敏数据等行为进行及时的告警提醒。

1、数据接口访问监测

数据接口请求监测主要是对服务节点（如 API 接口服务、数据库服务等）的流量和日志进行监测，对接口访问进行统计，结合业务规则对高频的访问行为进行及时的告警。

2、数据违规调用监测

数据违规调用监测指针对不同的节点或者人员对数据接口调用数据量进行监测，基于数据调用规则对违规的调用行为进行及时的发现和告警。

3、数据接口异常监测

数据接口异常监测对接口的健康状况进行实时监测，包括高延时、返回报错、404 异常访问进行监测和记录，一旦错误率超过设定的阈值，及时告警。

1.1.10 数据运营

数据运营是以鹿城区政务数据分平台为据点，以数据赋能、平台赋能、场景赋能、主管单位赋能为渠道，以数据资源、数据服务为介质，优化政府管理服务职能，提升社会治理服务能力现代化，同时发挥筑巢引凤的作用，带动

鹿城区数字经济发展。打造全省标杆的具有鹿城特色的数据运营体系。

1.1.10.1 数据场景研究与应用

场景是数据运营的承载体，无场景则无运营，对于省市已有的，要充分吸收转化集成，对于还没有的，要尽可能的具备本地特色；从责任主体上划分为两部分，一部分是由牵头单位主导的，另一部分是跨场景跨层级的。从用途上划分有三种：强政、惠民、兴业。目前一体化智能化公共数据平台的场景存在一种现象，惠民和强政的比较多，兴业的较少，即“产业场景”较少。基于上述现状，数据运营需要做以下工作：

1.1.10.1.1 数据场景研究

梳理省市已建场景，挑选符合鹿城需求的引进落地；针对创新场景，一种是行业强相关场景，一种是数据融合场景。行业强相关场景，需要部门为主导提出场景，紧贴业务，主管单位以配合为主；数据融合场景，主管单位定期组织头脑风暴探讨。最后我建议后续的场景应多往“产业场景”上靠，因为在当前经济形式下只有产业场景是可以持续不断地创造肉眼可见的价值，真正能够带动主管单位项目长期发展。

1.1.10.1.2 数据场景应用

主要是承接数据场景研究成果，分析场景落地的可行性，编制落地方案，形成项目。

1.1.10.2 数据场景成效校验

一体化智能化公共数据平台出现了场景的概念，经过多年的发展，不管民生领域还是管理领域都建设了许多场景，这些场景中口碑有好有坏，如何判断场景的社会效益经济效益、指导优化后续场景的设计和改造，变得至关重要。

1.1.10.2.1 数据资产效能校验

平台建有数据资产效能评价引擎，全市各单位均可以开通并使用这些能

力，数据局作为考核方，将对各单位的数据效能进行定期的盘查。对于成效差的，由主管单位牵头制定数据效能改善提升方案。

1.1.10.2.2 数据场景效能校验

建立一套测评标准，从场景的使用情况、口碑、价值等维度对场景进行评测。对于成效差的，由主管单位牵头制定改善提升方案。

1.1.10.3 数据增值服务

温州市当前处于数字经济发展的转折点，目前的大环境是：全市产业主要以传统制造业为主，数字经济产业的发展程度相较于先进市县仍有较大差距；GDP 差距较其他城市逐年增大的趋势。主要原因是产业结构失调和资源配套存在不足。随着“一体化智能化公共数据平台平台”的建设，本地数字经济基础设施正逐步完善，各种数据治理、数据交换的规则制定，已初步完成数字赋能产业的原始积累。

1.1.10.3.1 数据+产业

选取特定行业（如体育、芯片、医疗、电商），打造温州数字产业服务高地，建设独有的行业数据资源池，形成资源壁垒，筑巢引凤，吸引相关产业聚集。这里需要注意的一点是形成资源壁垒难度较大，但是可以先人一步展开筹措工作，借着某活动契机提前部署，把它当成一个功率放大器，聚焦“产业场景”。

1.1.10.3.2 数据交易

可借鉴大数据交易中心模式，依托于省政务数据开放风向标，采用“政务数据——企业数据——开放数据”模式，将政府中不宜交易的数据通过企业数据这一环进行脱敏加工处理转化形成宜交易的自由数据，在数据分平台加上一层企业数据层，在该层进行企业之间的数据交易，可以反哺“数据+产业”中所需的数据资源，形成正向循环。