**萧山市公安局**

**人口精准管控系统**

**建设方案**

**v1.9**

目录

[1 项目建设总概 6](#_Toc485218356)

[1.1 项目建设背景 6](#_Toc485218357)

[1.2 人口管理现状 6](#_Toc485218358)

[1.2.1 人口基数大流动性强 7](#_Toc485218359)

[1.2.2 传统管理模式单一化 7](#_Toc485218360)

[1.2.3 各部门信息采集滞后 8](#_Toc485218361)

[1.2.4 服务管理理念待提升 8](#_Toc485218362)

[1.3 总体建设目标 9](#_Toc485218363)

[2 系统建设方案 10](#_Toc485218364)

[2.1 系统构建思路 10](#_Toc485218365)

[2.2 系统总体设计框架 11](#_Toc485218366)

[2.2.1 系统架构图 11](#_Toc485218367)

[2.2.2 系统架构说明 11](#_Toc485218368)

[2.2.3 软件功能模块清单 15](#_Toc485218369)

[2.3 系统部署方案 16](#_Toc485218370)

[2.3.1 系统部署结构图 16](#_Toc485218371)

[2.3.2 系统部署说明 16](#_Toc485218372)

[2.3.3 硬件设备清单 18](#_Toc485218373)

[2.4 数据采集方案 18](#_Toc485218374)

[2.4.1 数据采集背景 18](#_Toc485218375)

[2.4.2 数据精准采集的必要性 19](#_Toc485218376)

[2.4.3 数据采集维度 19](#_Toc485218377)

[2.4.4 数据采集大纲 20](#_Toc485218378)

[2.4.5 数据清单目录 21](#_Toc485218379)

[2.4.6 数据采集方案详情 26](#_Toc485218380)

[2.5 模型积分规则及方案 32](#_Toc485218381)

[2.5.1 积分引擎 33](#_Toc485218382)

[2.5.2 积分模型 33](#_Toc485218383)

[2.5.3 积分计算 34](#_Toc485218384)

[2.5.4 单项积分计算 34](#_Toc485218385)

[2.5.5 综合积分计算 37](#_Toc485218386)

[2.5.6 积分规则模型列表 38](#_Toc485218387)

[2.5.7 规则模型使用数据清单 43](#_Toc485218388)

[2.6 绩效考核方案 48](#_Toc485218389)

[2.6.1 考核目标 48](#_Toc485218390)

[2.6.2 考核机构与主体 50](#_Toc485218391)

[2.6.3 考核标准 51](#_Toc485218392)

[2.7 数据资源管理 53](#_Toc485218412)

[2.7.1 数据采集 54](#_Toc485218413)

[2.7.2 数据归集 54](#_Toc485218414)

[2.7.3 数据清洗规则 54](#_Toc485218415)

[2.7.4 数据储存 54](#_Toc485218416)

[2.7.5 数据落地 55](#_Toc485218417)

[2.8 系统应用功能模块 56](#_Toc485218418)

[2.8.1 智能数据检索 56](#_Toc485218419)

[2.8.2 报表统计可视化 57](#_Toc485218420)

[2.8.3 历史记录追溯查询 58](#_Toc485218421)

[2.8.4 实时人员异动趋势显示 59](#_Toc485218422)

[2.8.5 可疑人员预警功能 59](#_Toc485218423)

[2.8.6 人员信息多维画像显示 60](#_Toc485218424)

[2.8.7 绩效考核模块 61](#_Toc485218425)

[3 系统核心技术 63](#_Toc485218426)

[3.1 CC-ETL数据处理软件 63](#_Toc485218427)

[3.1.1 CC-ETL基础介绍 63](#_Toc485218428)

[3.1.2 CC-ETL功能设计 64](#_Toc485218429)

[3.2 CC-Albus数据服务总线 68](#_Toc485218430)

[3.2.1 ALBus服务总线定义 68](#_Toc485218431)

[3.2.2 ALBus服务总线概念 69](#_Toc485218432)

[3.2.3 3ALBus服务总线模式 70](#_Toc485218433)

[3.2.4 ALBus服务总线作用 72](#_Toc485218434)

[3.2.5 ALBus服务总线架构特点 73](#_Toc485218435)

[3.3 CC-Model数据建模技术 79](#_Toc485218436)

[3.4 CC-Calc数据融合演算技术 81](#_Toc485218437)

[3.4.1 CC-ALCalc基础介绍 81](#_Toc485218438)

[3.4.2 CC-ALCalc软件主要功能 81](#_Toc485218439)

[4 解决方案优势介绍 82](#_Toc485218440)

[4.1 基于理论支撑 82](#_Toc485218441)

[4.2 系统前瞻性 82](#_Toc485218442)

[4.3 系统实时性 82](#_Toc485218443)

[4.4 模型精准性 83](#_Toc485218444)

文档控制页

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 版本号 | 修改日期 | 编制/修改人 | 修改说明 | 审批人 | 备 注 |
| V1.9 | 2017年6月 | 张培荷 | 新增绩效考核方案 |  |  |
|  |  | 张培荷 | 新增数据清单目录 |  |  |
|  |  | 王伟宏 | 细化积分规则 |  |  |
|  |  | 黄跃东 | 优化硬件清单、软件清单 |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

# 项目建设总概

## 项目建设背景

改革开放以后，不改变户籍而较长时期离开户籍地去外地谋生的流动人口，数量逐年增长，规模日益庞大。这一类流动人口已成为目前浙江省流动人口的主体。G20之后，随着信息化、工业化、城市化、国际化的快速发展，浙江省进入了经济转型的关键时期，也迎来了人口流动活跃时期，人口流动成为浙江省经济、社会、人口转型过程中的突出特征。

萧山区作为浙江省经济、社会建设的排头兵，多次荣获“全国农村综合实力百强县（市）”、“全国明星县（市）”、“国家卫生城市”、“浙江省品牌强县（市、区）”、“中国最令人向往的城市十强”、“浙江省科技综合实力第一名”、“全国百强县市第七名”、“中国园林绿化产业基地”等称号，是浙江省的首批小康县（市）。近几年，全区国内生产总值、工业总产值等主要经济指标成绩居浙江省县（市、区）级首位，其中2016年，萧山区（含大江东）实现生产总值1928.57亿元，按可比价格计算，比上年增长8.6%。

流动人口的涌入，为萧山经济和社会建设取得的成果做了巨大的贡献，但同时，也加大了流动人口管理的复杂性，增加了社会的不安定因素。加强和改进流动人口服务、管理质量，成为现阶段萧山社会建设的重大目标。

## 人口管理现状

人口精准化管理系统建立在人员基本信息数据库之上，通过将人员基本信息数据库中的静态数据，结合人员的动态采集信息，实现人员的动态管理。并通过建立模型算法，分析计算其安全属性，依据一定的准则，将人口进行科学合理分类，为精准化管理提供技术保障和支撑。同时将数据分类，筛选出有效数据，提高数据价值，实现公安各业务部门对信息数据的管理和数据更新状态的监控，并指导和管理人口管理人员的日常工作及信息采集内容。以解决当前人口“均同化”管理模式向“精准化”管理模式转变所遇到的人力不足、重点不清、管控不详等问题。

人口精准管控平台的积分制绩效管理：依据现有的工作机制、基础信息资源、人口信息访查平台，拓展人口信息采集项，丰富归集人口信息资源，通过科学合理的积分分类和算法，构建一套科学合理的本地和外来居住人员精准化管理系统，包括数据采集和积分制模型的构建等。

### 人口基数大流动性强

2016年末，萧山区公安部门户籍登记总人口（含大江东）127.59万人，登记在册的流动人口138.81万人。由于流动人口工作不稳当，居住地变化大等特性，造成了流动人口流动性强，分散范围大。流动人口的大量涌入，既带来了萧山区总体经济的提升，提高了萧山区经济活力，同时也为社会治安带来了很多隐患。

### 传统管理模式单一化

流动人口犯罪居高不下，管控力度未见成效的主要原因一方面在于流动人口管理机构的碎片化，政府与企业、非政府组织缺少合作，流动人口长期处于被动管理状态；另一方面，流动人口管理内容单一化，管理成本居高不下，管理手段简单滞后。上述因素造成了对流动人口的管控缺少目的性，对流动人口的服务缺少针对性。

传统对流动人口平面化的管理，已经不足以支撑现在流动人口剧增的新形势。构建新的全局性、综合性、基础性的流动人口精准管控平台，增强公安机关核心战斗力，提升“三基”工程建设水平，促进社会信用体系建设与发展，具有十分重要的意义。

### 各部门信息采集滞后

流动人口流动性强、数据庞大等原因导致数据采集工作十分艰难，采集得到的数据信息，在各部门之间数据因为各种客观因素不能共享，导致各部门之间数据隔离，且信息采集工作不够完善、全面。流动人口主动登记意识不强，也导致各种基本信息缺失，数据真实性也有待考证。

### 服务管理理念待提升

服务管理机制不够健全，服务管理理念有待提升。实际工作中并未发挥高效、协调统一的作用，尚未形成齐抓共管的协调工作机制和监督考核制度；部分职能资源难以得到有效整合；流动人口信息资源管理系统建设任务重和资源共享应用水平低的问题十分突出；一些单位和干部对流动人口存在排斥、防范的观念，具体工作中往往是歧视多于尊重、管理多于服务、防范多于保护，导致工作开展不够主动。

## 总体建设目标

为了实现对人口精准化服务管理水平，利用大数据特征标签计算技术，对人口的相关数据进行分析挖掘，通过特征模型，人员画像标签等多种数据挖掘技术，完善人口积分管理模块开发，开展人口基础数据采集归集、数据目录、数据积分配比和分级管理等工作，构建人口积分模型，支撑公安对于人口的精细化管控应用。

实现人口的积分量化管理，准确掌握人口基本情况，提升户籍制度改革积分落户的合法合理及科学性，实现对人口的精准服务和管控。

进而建设一套萧山区人口精准管控系统，可以提升萧山区人口治理主体的多元化，层次化，通过全方位的人口管理主体，提升人口管理工作的服务和管控水平；可以指导和改进警力调度，按照人口的宏观分布规律和微观活动态势，灵活部署经理，实现精准管理；可以完善城市人口治理的制度工具，丰富萧山区人口管理与服务的内容，改进人口治理工作的方式。

通过本系统的建设，主要实现：

（一）科学采集人口信息，完善人员信息库

（二）提供人口管理的精准数据支持。

（三）实现对人口科学合理化管控

（四）实现对人口精准化服务管理

# 系统建设方案

## 系统构建思路

人口精准管控系统初步建立在人员基本信息数据库之上，通过将人员基本信息数据库中的静态数据转化为动态数据，结合其他分类补充数据，筛选出有效数据，提高数据价值，实现公安各业务部门对信息数据的管理和数据更新状态的监控。初步统一的针对人口精准管控系统，为人口的服务和管理工作提供精准有效的工作支持和判断依据。

基于长期的理论研究，我们认为一个人的居住、工作、身份、社交、活动和消费行为综合全面反映了他的社会安全属性，而其中居住和工作的性质、稳定性则大致反映了其最基本生活状况和精神状态，从而直接影响到其是否存在社会安全隐患，一个居无定所、无稳定收入来源的人往往更容易发生盗窃、抢劫等非法犯罪行为。因此，全面及时掌握一个人的居住、工作情况，对一个人因为居住、工作的变化可能导致的异常行为进行预判对于有效预防某些性质的违法犯罪行为有重大的意义。基于现有的居住信息登记系统的数据梳理了反映一个人居住、工作稳定性的一些主要指标，对这些指标进行监测，通过建模来实现对人口的精准分类。

系统实现通过评价模型中量化的公共安全分值，对人口个体进行科学评判，进而确定所处的级别。对不同级别的人口采取不同力度的管理措施和方法，进而达到“精准化”管理。同时实现警员工作管理，引导人口管理员科学安排勤务，提高访查工作效率；引入KPI绩效指标对人口管理员的日常工作进行考核、管理。

## 系统总体设计框架

### 系统架构图



Figure

### 系统架构说明

如系统架构图所示，萧山公安人口精准管控系统建设中，主要包括几个部分：

#### 本地数据资源库的建设

在萧山公安，建设一套数据资源库，通过ETL系统整合、清洗、转换萧山本地各类数据资源，如政务网内各行政部门数据、医疗卫生就诊数据、水电煤数据、企业数据等。数据汇总的方式，可以采用后台数据库对接、接口服务对接，文件上传等多种方式。

在汇聚本地数据的同时，将通过接口方式，将本地的数据上传一份到市局的大数据平台，为市局层面多数据源融合模型计算提供额外数据支持。

#### 本地数据演算平台的建设

在本地数据资源库的之上，通过使用ETL数据清洗、分布式数据仓库等技术建设一套本地化数据演算平台，提供各类数据规则、模型的本地演算能力。

#### 统一演算调度管理模块的建设

本系统建设中，综合使用了萧山本地和市局信息资源服务平台的数据资源和演算平台。通过开发统一的演算调度管理模块，本地数据资源相关的模型规则在本地演算平台进行；市局数据相关的模型规则在市局演算平台进行，规则模型的演算结果，数据计算汇总结果统一存储在本地，反馈给应用系统使用方。

#### 流动人口发现规则模型

本系统中，流动人口的发现，除了采用传统的流动人口主动申报和上门访查的模式外，还通过大数据挖掘手段，利用模型规则对本地数据资源的深挖，更大范围的发现流动人口。流动人口发现模型主要包括：

* 基于企业职工数据的发现模型；
* 基于医疗就诊数据的发现模型；
* 基于社会保险数据的发现模型；
* 基于交通出行数据的发现模型；
* 基于水电煤数据的发现模型；
* 基于WIFI采集数据的发现模型；
* 基于暂住人口系统的发现模型。

#### 人口积分评判规则模型

本系统中，对于本地常口和流动人口，通过一套积分规则引擎，为每个人员形成一个社会安全积分。社会安全积分由几大类分类积分组成，如居住危险性积分、工作危险性积分、社交危险性积分等。每一个分类积分的分值由多个积分模型共同计算得到，反映了管理对象在社会生活中某一块的危险性指数。积分模型在计算分类积分时的所占的比重可以通过平台管理系统方便地进行调整，以适应外部环境的变化。

#### 人口精准管控模块

人口精准管控模块里面，包含了每个人口的基础信息，社会安全评分，管理服务状态等信息，人口在本辖区内产生的各类数据，都可以通过此平台进行展现。同时，可以对人口的缺失信息进行发现，并提醒相关民警进行落地核实和补充。

#### 异常人员触发规则模块

异常人员的触发规则，既包括对社会安高风险的人员发现，也包括对信息缺失流动人员发现，通过在后台每天运行规则，生成相关的异常人员名单。

#### 精准推送模块

精准推送模块，就是将异常人员触发规则模块发现的人员名单，按照预定义的管辖区域关系，推送给责任辖区派出所及民警。

#### 数据落地核查

数据落地核查，指一线派出所及民警接到异常人员名单后，通过电话访查、上门拜访等手段，收集完善目标对象的最新数据信息，并通过接口或者离线人工方式，将最新数据更新进人口精准管理系统。

#### 与外部数据采集系统的对接

因为萧山公安目前已经有一些相对成型的外部数据采集应用系统，所以本系统建设时可以与现有的外部数据采集系统对接，主要工作如下：

1. 常住人口分析底数采用本地化建设的人口基础库的数据，采用ETL实时同步的机制；
2. 流动人口数据采集相关的工作，可以通过完善现有外部信息采集系统的数据采集项来完成，详见数据采集方案和积分方案部分说明；
3. 通过开发推送对接模块，将本系统智能研判出来的异常人口待查名单，通过接口推送到外部采集系统中，以便一线民警落地核查；
4. 外部采集系统的异常待查人员反馈数据，应该按照双方定义的接口要求，实时推送到本系统中，以完成待核查数据项目的更新反馈工作。

### 软件功能模块清单

|  |  |  |  |
| --- | --- | --- | --- |
| **序号** | **类别** | **产品名称** | **配置说明** |
| 1 | 数据平台基础 | 本地数据集群建设 | 建设本地数据资源集群、演算集群，存储系统接入的各类数据资源，并提供演算功能 |
| ETL大数据软件 | 负责数据汇总、清洗、转换的软件，用于萧山公安本地数据的汇聚 |
| 数据分发和上传模块 | 本地数据中，需市局演算平台进行计算的数据的分发和上传管理 |
| 统一演算、建模、数据服务管理模块 | 统一协调管理本地、市局两个平台的建模、演算和数据服务调用工作 |
| 人群标签和画像模块 | 基于自定义和系统演算获得针对人员的标签，并通过标签形成对一个人员的完整画像描述 |
| 2 | 基础应用模块 | 用户和权限角色管理模块 | 用户管理，权限角色管理 |
| 系统日志和审计模块 | 系统日志管理和审计模块 |
| 3 | 分析模型挖掘 | 人口积分评价功能模块 | 基于模型规则的人口积分评价体系建设，包括： 居住危险性积分模型 工作危险性积分模型 社交危险性积分模型等 |
| 流动人口发现功能模块 | 基于模型规则的流动人口发现体系，包括： 基于企业职工数据的发现模型 基于医疗就诊数据的发现模型 基于社会保险数据的发现模型 基于交通出行数据的发现模型  基于水电煤数据的发现模型 基于WIFI采集数据的发现模型  基于暂住人口系统的发现模型等。 |
| 4 | 业务功能模块 | 人口精准管控模块 | 针对人口的基础管理模块，基于人口积分评价等，形成人口标签、画像和整体评价，并对缺失的信息或者日常的人口管理工作精心提醒并完善 |
| 异常人群触发规则模块 | 基于规则模型，对异常和搞危险人群的发现和预警模块 |
| 精准消息推送传递模块 | 按照管理辖区对应，对产生的异常人群或预警信息推送给责任单位和民警，进行有效提醒 |
| 数据落地核查模块 | 人员落地核查和信息录入模块 |
| 绩效管理模块 | 民警绩效管理模块 |
| 与采集系统对接模块 | 与现有警务基础信息采集应用对接模块 |
| 5 | 安装部署和数据服务 | 安装部署服务 | 软硬件集群部署，萧山本地数据接入、清洗服务服务 |

## 系统部署方案

### 系统部署结构图

Figure

### 系统部署说明

#### 数据采集和数据库服务器（1台）

数据采集服务器，部署ETL数据采集软件，可以通过数据库对接、数据传递接口、文件多种方式实现数据的整合。

数据库服务器，包括用户数据采集配置的存储和人口精准管控系统数据的存储。

#### 应用系统、集群管理、消息队列、演算调度服务器（3台）

应用系统，集群管理、消息队列、演算调度服务器共3台，3台服务器同时扮演多种角色，并同时对外提供服务。

#### 集群节点服务器（6台）

集群节点服务器，作为数据存储的节点和计算能力的提供者，按照3的倍数进行部署，依照一般市、区公安部门数据的规模，建议部署6台，后续可以按照实际业务量进行扩展。

#### 网络交换机（4台）

集群内部的数据通讯，通过万兆网落互联，与公安网的连接，通过千兆网络连接，考虑到网络冗余性，需要配置或者利旧2台万兆24口万兆交换机和2台48口千兆交换机

#### 网络安全边界前置服务器（2台，按实际情况配置）

考虑到本项目中，需要通过网络安全边界汇总数据，需要在网络安全边界两侧各部署一台服务器。（如果已经有这个服务器，可以考虑不配置）

### 硬件设备清单

|  |  |  |  |
| --- | --- | --- | --- |
| 机器角色 | 数量 | 参数 | 备注 |
| ETL和数据库服务器 | 1 | 2U机架服务器，配置2颗 Intel® Xeon® E5-2630 V3 CPU，128G内存，6块900GB SAS硬盘，1G以上缓存RAID卡，2个千兆网口，2个万兆网口 | 必选 |
| 数据和计算节点服务器 | 6 | 2U机架服务器，配置2颗 Intel® Xeon® E5-2630 V3 CPU，256G内存，4块800G SSD硬盘，8块4TB SATA硬盘，1G以上缓存RAID卡，2个千兆网口，2个万兆网口 | 必选 |
| 应用、消息队列、管理服务器 | 3 | 2U机架服务器，配置2颗 Intel® Xeon® E5-2630 V3 CPU，128G内存，6块600GB SAS硬盘，1G以上缓存RAID卡，2个千兆网口，2个万兆网口 | 必选 |
| 万兆交换机 | 2 | 24口万兆交换机 | 如现有设备满足接口需求，可利旧 |
| 千兆交换机 | 2 | 24口千兆交换机 | 如现有设备满足接口需求，可利旧 |
| 安全边界服务器 | 2 | 64G内存，4TB可用空间，2个千兆网口 | 可以虚拟机或者老机器利旧 |

## 数据采集方案

### 数据采集背景

2016年末，萧山区公安部门户籍登记总人口（含大江东）127.59万人，登记在册的流动人口138.81万人。庞大的流动人口基数成为落实数据精准采集的第一道障碍。流动人口流动性强、数据庞大等原因导致数据采集工作十分艰难，采集得到的数据信息，在各部门之间数据因为各种客观因素不能共享，导致各部门之间数据隔离，且信息采集工作不够完善、全面。此外流动人口主动登记意识不强，也导致各种基本信息缺失，数据真实性也有待考证。

### 数据精准采集的必要性

数据精准采集是人口管控平台建设的基础，在管控平台建设过程中，将人口各个信息系统的数据，进行采集、抽取、转换、加载，集中分类归集，形成统一、完整的数据库结构。为确保积分模型的精准性，必须从数据采集开始，保证数据源的真实性和完整性。

因此建立规范社会治安管理类信息采集标准，回访核实机制，制作采集表格，更新数据必须采集项，组织业务指导培训，审核、录入实有人口，实有房屋信息登记、出租房备案登记等一系列措施，都可以提高强对流动人口服务管理工作。

完善的数据采集机制不仅提高政府部门工作效率，避免多头采集，重复劳动，确保流动人口信息的完整、准确、鲜活，建立动态更新、数据共享、综合服务的工作机制，实现新时期全方位社会管理、人性化社会服务，构建新型社会管理服务体系。

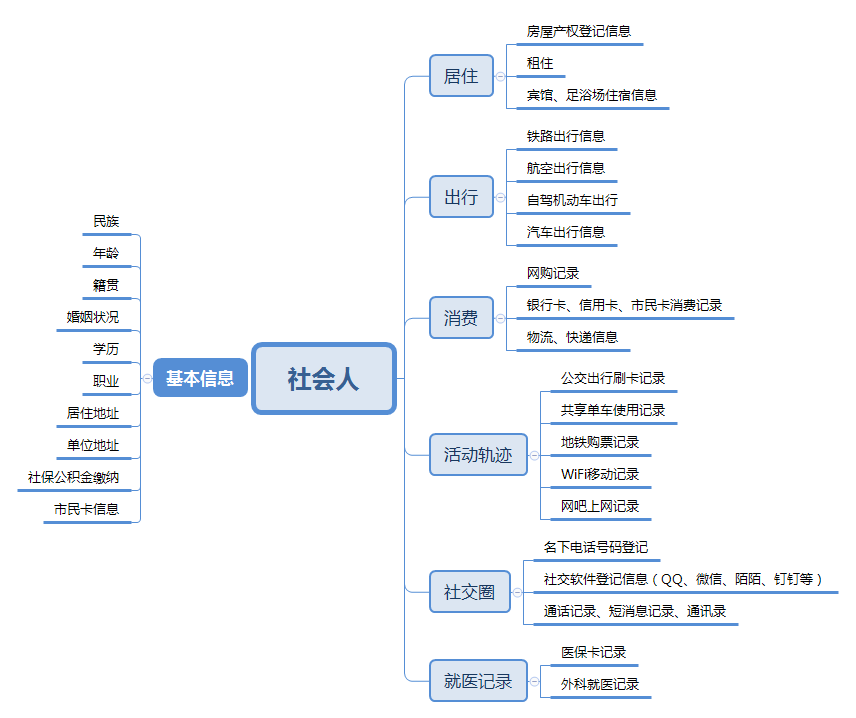
### 数据采集维度

本系统需要整合的数据主要从居住、职业、出行记录、消费记录、活动轨迹、社交圈六个维度进行采集和汇聚，积极协调各一线部门获取公安网、政务网、社会及互联网中涉及人口管理的系统资源，通过归集各系统涉及人员公共安全相关数据，数据导入系统后，将采集导入的数据按数据目录进行归集整合，进而通过分布式特征标签计算技术，将人员社会安全特性标签化。



### 数据采集大纲

依据社会人正常生活的经济需求和吃住行乐的生存需求，以及经济来源的正当性和吃住行乐的合理性的一般准则，基于人口基本信息以及劳动就业、教育、收入、社保、房产、信用、卫生计生、税务、婚姻、民族等信息，利用大数据技术等技术，通过分析其家庭状况、谋生能力及前科劣迹等基本状况，结合其居住、工作稳定性以及工作正当性和潜在危害性分析，对人口的“健康指数”的进行“诊断”。另对各类涉稳、涉恐、涉毒、在逃、重大犯罪前科、精神病人、重点上访等7类重点人员进行梳理、汇总，摸清底数，采集信息，对可能危害社会治安的重点人员落实管控措施，掌握其现实表现、行动轨迹等情况。



### 数据清单目录

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | |
| 序号 | 数据分类 | 数据项 | 数据结构 | 来源 | 数据共享方式（在线/离线） | 补充采集 |
| 1 | 基本信息 | 常住人口 | 民族、  身份证号码、  姓名、  性别、  籍贯、  年龄、  婚姻状况、 | 市云 | 在线 | 职业类别、学历、技能、婚姻等 |
| 居住人口 | 民族  身份证号码  有效标志  户籍地 | 市云 | 在线 | 职业类别、学历、技能、婚姻等 |
| 三实人口 | 民族  身份证号码  有效标志  户籍地 | 市云 | 在线 | 职业类别、学历、技能、婚姻等 |
| 公积金信息 | 单位名称  身份证号码  缴存状态  最近缴存日期 | 社保局 | 定期离线 |  |
| 民政婚姻信息 | 民族  身份证号码  有效标志  户籍地 | 市云 | 在线 |  |
| 市民卡信息 | 市民卡卡号  身份证号码  卡状态 | 市民中心 | 定期离线 | 消费流水、就医购药记录 |
|  |  | 社保信息 | 工作单位名称  身份证号码  参保日期  参保状态 | 社保局 | 定期离线 |  |
| 2 | 居住信息 | 常住人口 | 民族  身份证号码  有效标志  户籍地 | 市云 | 在线 | 出租房出租人和承租人采集登记 |
| 居住人口 | 民族  身份证号码  有效标志  户籍地 | 市云 | 在线 | 出租房出租人和承租人采集登记 |
| 出租房东登记 | 民族  身份证号码  有效标志  户籍地 | 分局 | 在线 | 出租房备案信息  （待采集） |
| 房屋产权 | 房屋地址  身份证号码  是否有房 | 分局 | 在线 | 待采集 |
| 3 | 出行信息 | 铁路 | 民族  身份证号码  有效标志  户籍地 | 市云 | 在线 |  |
| 民航 | 民族  身份证号码  有效标志  户籍地 | 市云 | 在线 |  |
| 客运 | 民族  身份证号码  有效标志  户籍地 | 市云 | 在线 | 客运数据萧山客运中心  （待采集） |
| 机动车 | 民族  身份证号码  有效标志  户籍地 | 分局 | 在线 |  |
| 4 | 消费信息 | 购物平台信息 | 身份证号码、  消费金额、  消费内容、  消费时间、  支付方式、  支付账户、  购物平台 | 购物平台 | 定期离线 | 消费金额、消费内容、消费时间  （待采集） |
| 银行卡消费 | 银行卡号、  身份证号码、  消费金额、  消费内容、  消费门店地址、  消费时间 | 各大银行 | 定期离线 | 消费金额、消费内容、消费门店地址、消费时间  （待采集） |
| 市民卡消费 | 身份证号码  市民卡号  消费金额、  消费内容、  消费门店地址、  消费时间 | 市民中心 | 定期离线 | 消费金额、消费内容、消费门店地址、消费时间  （待采集） |
| 信用卡消费 | 信用卡号、  身份证号码、  消费金额、  消费内容、  消费门店地址、  消费时间 | 信用卡中心 | 定期离线 | 消费金额、消费内容、消费门店地址、消费时间  （待采集） |
| 支付宝消费 | 支付宝账号、  身份证号码、  消费金额、  消费内容、  消费门店地址、  消费时间 | 支付宝 | 定期离线 | 消费金额、消费内容、消费门店地址、消费时间  （待采集） |
| 物流信息 | 物品名称、  物流单号、  寄件人身份证号、  寄件人联系电话、  收件人联系电话、  寄件人地址、  收件人地址、  寄件时间、  收件时间 | 各大物流公司 | 定期离线 | 物品类型、物流路径，买卖双方详细信息  （待采集） |
| 快递信息 | 寄件人身份证号、  寄件人联系电话、  收件人联系电话、  快递单号、  寄件人地址、  收件人地址、  物品名称、  寄件时间、  收件时间 | 购物平台及快递公司 | 定期离线 | 寄、收人员详细基本信息，货源地和收货地址、物品类型  （待采集） |
| 5 | 活动信息 | 公交 | 公交卡号、  刷卡时间、 | 市云 | 定期离线 | （待采集） |
| 地铁 | 身份证号码、  进站时间、  出站时间、  出发站名、  出站站名 | 市云 | 定期离线 | 进站、出站站口信息（大量单程散客数据采集归集）  （待采集） |
| 共享单车 | 使用人身份证号、  使用人手机号、  使用人银行卡号、  使用人支付宝账号、  开始时间、  开始地点、  结束时间、  租赁结束地点、  消费金额 | 共享单车平台 | 定期离线 | GPS定位信息，使用人信息，使用人使用期间轨迹  （待采集） |
| 公共自行车 | 使用人身份证号、  租车时间、  租车服务点、  还车时间、  还车服务点 | 公交集团 | 定期离线 | 使用人信息，起始结束服务店信息  （待采集） |
| 网约车 | 车辆车牌号、  网约人身份证、  网约人电话号码、  司机身份证号、  司机电话号码、  网约人上车时间、  网约人上车地点、  网约人下车时间、  网约人下车地点、  网约人消费金额 | 网约平台 | 定期离线 | 网约车归属、车辆信息、网约人信息  （待采集） |
| 家用小轿车 | 车辆归属人身份证、  驾驶人身份证、  驾驶人驾驶证、  车辆行驶证、  车牌号码、  车牌品牌、  车辆型号 | 分局卡口 | 在线 |  |
| WIFI轨迹等 | 基站信息、  经纬度、  上网记录、  终端IP、  终端mac地址 | 网警 | 在线 | 基站信息、经纬度、上网记录  （待采集） |
| 网吧记录 | IP地址，  网吧名称、  网吧地址、  上网人身份证号码、  上线时间、  下线时间 | 市云 | 定期离线 | IP地址 |
| 住宿记录 | 身份证号码、  旅馆名称、  旅馆地址、  入住时间、  离店时间、  入住房号 | 市云 | 定期离线 | 同住人员信息 |
| 6 | 社交信息 | 重点人员通讯录 | 机主手机号码、  机主身份证号、  机主通讯录信息、  机主通信话单记录、  机主短信收发记录 | 运营公司 | 定期离线 | 前科人员、敏感词汇、短信、电话记录 |
| QQ朋友圈 | 身份证号码、  QQ账号、  QQ好友信息、  QQ群信息、  QQ聊天记录、  QQ转账记录 | 腾讯 | 定期离线 | 转账记录、聊天记录、敏感词汇  （待采集） |
| 微信朋友圈 | 身份证号码  微信账号、  微信好友信息、  微信群信息、  微信聊天记录、  微信转账记录 | 腾讯 | 定期离线 | 转账记录、聊天记录、敏感词汇  （待采集） |
| 陌陌、钉钉朋友圈 | 身份证号码  账号信息、  好友信息、  群信息、  聊天记录、  转账记录 | 社交软件平台 | 定期离线 | 转账记录、聊天记录、敏感词汇  （待采集） |
| 7 | 就医信息 | 门诊记录 | 身份证号码、  外科门诊记录、  医保卡号、  就诊时间 | 医院管理系统、社保局 | 定期离线 | 外科门诊记录、医保就医明细 |

### 数据采集方案详情

#### 基本信息

**数据定义**：基本信息包括人员自身基本属性，及其他社会活动添加属性。

基本信息主要包括姓名、年龄、民族、户籍地、性别、婚姻状况、学历、工作、现居地址、单位地址、社保公积金缴纳信息、市民卡办理情况。

**数据来源：**常住人口、流动人口信息登记表、社保/公积金缴纳信息、民政婚姻登记，市民卡等信息。

**采集方式建议：**

* 直接利用市局信息资源服务平台中的常住人口、居住人口、三实人口、民政婚姻登记、社保缴纳、公积金缴纳、市民卡登记等数据
* 居住人口和实有人口信息中的部分关键信息需要通过完善采集方式进行数据标准化，如职业类别，从业单位、学历、技能、婚姻状况等模糊性信息。

**数据采集建议：**采集部门应做好辅助填写工作，确保填写的信息准确真实，建议通过升级采集端软件，特别关注职业类别、从业单位、学历、技能、婚姻状况等模糊信息，采用标准预定义选项方式采集，而不是采用自定义录入方式。

#### 居住

**数据定义：**主要针对流动人员及常住人口的居住情况，主要分为自购房屋和出租房屋给他人居住或租入房屋居住。

**数据来源：**市局信息资源服务平台中的居住人口，出租房东登记表；房产局的房屋产权登记表

**采集方式建议：**

* 直接利用市局信息资源服务平台中的常住人口，居住人口，出租房东登记表数据。
* 通过分局管辖区域内的房产局/城建局等政府部门获取管辖区域内的房屋产权登记数据，通过定期数据交换方式进入公安网，通过匡信ETL软件更新到本地计算集群。

**数据采集建议：**重点关注出租房登记信息，出租人和承租人的身份证明/档案编号/暂住详址等应详细登记。

#### 出行

**数据定义：**出行主要针对超出管辖区范围之外的出行记录，或其他跨省跨国出行。

针对出行方式可从铁路、航空、自驾、客运汽车四个方面采集出行数据。

**数据来源：**铁路部门、萧山机场、客运中心实名制订票、机动车卡口数据

**采集方式建议：**

* 直接利用市局信息资源服务平台中的客运中心实名制订票数据。
* 直接利用市局信息资源服务平台中的铁路订票、民航出行等数据（更新不够及时，如果能独立渠道获取则更好）。
* 直接利用市局信息资源服务平台中的机动车全市卡口数据。

**数据采集关注：**2016年7月1日起，在杭州客运中心站、火车东站公路汽车站、汽车西站、汽车南站、汽车北站五大客运站，实行实名制购票和实名制乘车，其他汽车站乘坐汽车大巴出行未全部实现实名制购买，数据不够完整，应完全落地实施汽车实名制购票。自驾机动车数据因活动性较大更新采集不及时，卡口信息应及时更新到系统。

#### 消费

**数据定义：**最基本的衣食住行都需要经过消费来反映，数据时代的消费方式主要通过线上购物、线下门店消费，根据支付方式不同，包括网银、支付宝、信用卡等电子支付，主要针对以上几种方式记录消费信息。便利的网购衍生快递物流信息。

**数据来源：**各类购物平台网购记录、银行卡消费记录、市名卡消费记录、信用卡消费记录、支付宝消费记录、物流公司、快递公司

**数据相关部门：**各类购物平台（京东、淘宝及阿里巴巴旗下APP）、支付宝、各大银行

**数据采集建议：**分局通过可行的渠道获取社会消费类数据，采用离线文件方式提供，通过匡信ETL软件更新到本地计算集群。

#### 活动轨迹

**数据定义：**活动轨迹主要针对在管辖区范围内的行为活动记录，主要从出行方式、WiFi移动轨迹、3G/4G基站信息、网吧上网记录、旅馆住宿方面进行数据收集记录。

**1、出行方式**

**数据来源及建议：**

|  |  |  |
| --- | --- | --- |
| 出行方式 | 数据来源 | 数据采集建议 |
| 公交 | 杭州公交集团、  市局信息资源服务平台 | 采用定期离线文件提供数据，建议采集以下数据：   1. 公交卡用户信息 2. 市民卡信息 3. 公交刷卡记录 4. 公交定位数据 |
| 地铁 | 杭州地铁集团  市局信息资源服务平台 | 需对数据进行抽取清洗规整，整合成实名制信息。 |
| 共享单车 | 各大共享单车公司（小鸣单车、hellobike、mobikie、ofo等） | 实名制使用单车，但是各大共享单车之间数据不共享，运营模式不同。  采用定期离线文件提供数据，建议采集以下数据：   1. 实名制注册信息 2. 借还车流水数据（含GPS定位数据）。 |
| 公共自行车 | 杭州公交自行车运营公司 | 实名制使用、统一服务点，需对数据进行抽取清洗规整，整合成实名制信息。  采用定期离线文件提供数据，建议采集以下数据：   1. 实名制注册信息 2. 借还车流水数据 3. 服务点信息（含GPS位置） |
| 网约车 | 滴滴、曹操等 | 网约平台通常需要实名制约车、也会绑定手机。  采用定期离线文件提供数据，建议采集以下数据：   1. 实名制注册信息 2. 约车流水数据和上线车   点位信息（含GPS位置） |
| 家用小轿车 | 卡口过车记录 | 接入分局实时卡口数据库，通过匡信ETL软件将数据实时更新到本地计算集群。 |

**2、WiFi轨迹记录、3G/4G基站信息**

**数据定义：**根据被热点WiFi采集的移动轨迹，手机电脑等通信设备的使用产生的基站定位信息，大致描绘出使用人的活动轨迹

**数据来源：**分局网警大队

**采集方式建议：**分局从本地网警大队获取数据，可采用光闸系统BCP文件交换对接方式提供，通过使用匡信ETL软件直接接入到本地计算集群。

**3、网吧上网记录**

**数据定义：**网吧上网记录主要记录上网人身份证信息、上网开始结束时间、网吧信息、使用电脑IP地址。

**数据来源：**市局信息资源服务平台中的网吧上网记录表，分局网警可选择提供的额外上网信息表

**采集方式建议：**

* 直接利用市局信息资源服务平台中的数据。
* 如分局网警大队能提供额外的上网信息记录表，则可采用光闸系统BCP文件交换对接方式提供，通过使用匡信ETL软件直接接入到本地计算集群。

**数据采集关注：**采集网吧上网记录时应特别关注重点关注人员上网信息、管辖区内的重点网吧上网记录。

**4、住宿记录**

**数据定义：**住宿记录主要针对在外住宿，住宿场所可为宾馆或足浴场所

**数据来源：**市局信息资源服务平台中的市云旅馆住宿记录表。

**采集方式建议：**直接利用市局信息资源服务平台中的数据。

**数据采集关注：**数据采集应该关注住宿人员是否为本地户籍或是否为重点关注人员，及办理入住时间，和其他异动信息、管辖区内重点旅馆。

#### 社交圈

**数据定义：**名下电话号码登记，通过手机、社交软件产生的信息记录数据，及因此展开的朋友圈。

**数据来源：**运营商、社交软件（QQ、微信、陌陌、钉钉等），市局信息资源服务平台中相关采集数据表

**采集方式建议：**

* 直接利用市局信息资源服务平台中的数据。
* 从运营商，分局网警大队获取补充数据，采用离线文件方式定期提供，通过匡信ETL软件更新到本地计算集群。

**数据采集关注：**应特别关注重点关注人员名下电话号码，及相关通话记录、短信息、通讯录。关注相关敏感词汇信息，及使用频率。

#### 就医记录

**数据定义：**就医记录指在各大医院使用医保卡就医或其他挂号就医记录。

**数据来源：**辖区内各重点医院HIS系统

**采集方式建议：**分局定期获取医院医疗就诊数据，采用离线文件方式提供，通过匡信ETL软件更新到本地计算集群。

**数据采集关注：**重点关注外科/精神科门诊记录。

## 模型积分规则及方案

人口精准管控系统的积分制绩效管理：依据现有的工作机制、基础信息资源、人口信息访查平台，拓展人口信息采集项，丰富归集人口信息资源，通过科学合理的积分分类和算法，构建一套科学合理的人口精准化管控系统，包括数据采集和积分制模型的构建。

通过积分模型：以影响大配高分、影响小配低分、重要项目一票否决的原则对分类分级后的人口数据项进行积分配比。以重要性、影响力、危害程度等指标对各数据项赋予不同分值。该分值可以随时间推移由系统自动或人工手动动态调整。最终达到对任意人口均能给出一个量化的、能自我修正的公共安全分值的最终目的。

### 积分引擎

积分引擎负责处理管理对象的社会安全积分，社会安全积分由几大类分类积分组成，如居住危险性积分、工作危险性积分、社交危险性积分。每一个分类积分的分值由多个积分模型共同计算得到，反映了管理对象在社会生活中某一块的危险性指数。积分模型在计算分类积分时的所占的比重可以通过平台管理系统方便地进行调整，以适应外部环境的变化。

积分引擎主要由积分模型、积分模型配置和积分计算两部分组成。

### 积分模型

积分模型由一组在大数据分析框架下开发实现的算法模型组成，每一个积分模型都用于分别计算管理对象在某一特定社会安全维度上的危险性指数。

积分模型配置用于配置每一个积分模型规则的可选参数。配置信息使用MySQL数据库保存，后期有新的积分模型加入或原有模型需要针对新的环境做调整时只需要增加或修改配置记录即可。 每一个积分模型规则都有自己的模型配置信息，配置信息记录了该模型在计算时的一些可选参数及满足该模型的管理对象应该获得的单项积分权重。 例如，对名为“与前科人员频繁同住旅馆”这个积分模型规则对应的模型配置信息为半年内满足1次获得危险性积分权重为1，达到10次或以上获得危险性积分权重为10。而名为“从事特殊行业”这个积分模型规则对应的模型配置信息为：两年内满足则积分权重为1。

### 积分计算

积分引擎在进行积分计算时首先调用所有积分规则进行单项积分计算，然后对每一个管理对象，将其获得的所有的单项积分按权重累加后转换为百分制的分类总积分。例如居住情况危险性积分，工作情况危险性积分。之后通过加权平均计算每个管理对象的社会危险性总积分。

安全积分依据侧重角度不同，可分为正向积分、负向积分和混和积分三类。

* 正向积分对管理对象在社会生活中的行为可能对社会安全产生下面影响的情形计算得分，得分越高代表对社会安全越有益，例如工作稳定、家庭和睦、正常作息、交际圈无危险人员等；得分越接近0则管理对象对社会安全的益处越低，需要重点关注。
* 负向积分通过对管理对象进行的社会安全可能产生负面影响的行为计算积分，得分越高危险性越高，需要重点关注，例如无稳定收入来源、家庭关系不稳定、交际圈子有很多危险人员等。
* 混合积分则同时计算正向和负向积分，很多时候能够综合反应管理对象的安全坐标，但混合积分会对某些重要的波动反应钝化，导致不能及时预警。

本次系统默认计算负向积分。

### 单项积分计算

单项积分通过分别计算所有属于该项的积分模型的模型权重分值累加得到。

首先分别计算管理对象每个模型得到的权重积分。模型权重积分代表该模型对该项的贡献度。

以居住安全性这一单项积分为例，首先得到它在数据库中的模型配置信息，即它包含了哪些模型，每个模型的类型及动作方式。属于居住安全性的积分模型在数据库中的配置有类似下表的结构：

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 单项 | 模型规则 | 模型类别 | 最低分 | 最高分 | 阈值Min | 阈值Max |
| 居住安全性（负向） | 是否租住 | BOOL | - | 10 | - | - |
| 独自租房 | BOOL | - | 10 | - | - |
| 频繁变更住处 | MIN | 5 | 10 | 3 | 5 |
| 频繁入住旅馆 | MIN | 10 | 30 | 5 | 15 |
| 与前科同住旅馆 | MIN | 10 | 40 | 2 | 5 |

* 上表中第一列为单项积分模型。
* 第二列为模型规则名。
* 第三列为规则模型的类别，分别为BOOL、MIN、MAX、RANGE。BOOL即根据管理对象是否满足条件来决定是否获得积分，条件计算结果为True则获得积分，False则不获得；MIN指模型结果大于或等于min阈值时满足条件；MAX指模型结果小于或等于Max阈值时满足条件；RANGE指模型结果同时满足Min阈值和Max阈值时满足条件。大多数情况下使用BOOL型和MIN型。
* 第四第五列为满足该模型时获得的最低和最高分。所有类型的模型在不满足模型条件时都不获得积分。BOOL型在满足模型条件时获得最高分；MIN型在结果等于Min阈值时获得最低分，等于或大于Max阈值时获得最高分，处于两者之间时按线性关系处理；MAX型在结果等于MAX阈值时获得最低分，等于或小于Min阈值时获得最高分，处于两者之间时按线性关系处理；RANGE型在结果等于Min阈值时获得最低分，等于Max阈值时获得最高分，处于两者之间时满足线性关系。
* 第六第七列模型满足条件时，获得最低和最高分时的阈值。

基于上表的配置，可以计算管理对象在每个规则上获得的积分。假设管理对象通过模型规则计算后得到的结果为R，在该模型上获得的积分为，模型配置表中的最低分、最高分别为、，Min阈值、Max阈值分别为、。

对BOOL型，有：

对MIN型，有：

对MAX型，有：

对RANGE型，有：

根据上式对管理对象A的计算结果如下表：

|  |  |  |  |
| --- | --- | --- | --- |
| 单项 | 模型规则 | 计算结果 | 获得积分 |
| 居住安全性（负向） | 是否租住 | 是 | 10 |
| 独自租房 | 否 | 0 |
| 频繁变更住处 | 1次 | 0 |
| 频繁入住旅馆 | 10次 | 20 |
| 与前科同住旅馆 | 6次 | 40 |

得到对象在每个模型上的模型积分后，可以计算在该单项上的百分制总积分。令对象获得的模型分为，该模型能获得的最高权重分为（积分模型配置表中的最高分）, 则可以通过下式计算百分制单项积分：

如果在模型配置表中最高分列总和为100，即，则上述公式可以简化为：

根据上式计算管理对象A所获得的住宿单项积分为：

即该对象获得的住宿类单项总积分（百分制）为70

### 综合积分计算

使用同样的计算方式对每一个单项计算百分制安全积分（负向），然后可以计算总积分。

同样以A为例，假设A获得的所有单项积分如下表所示：

|  |  |  |
| --- | --- | --- |
|  | 单项积分（负向） | 积分权重配置 |
| 基本信息 | 20 | 1 |
| 居住 | 70 | 3 |
| 消费 | 10 | 2 |
| 出行 | 55 | 2 |
| 活动轨迹 | 72 | 2 |
| 社交圈 | 40 | 3 |

上表中第一项为单项；第二项为管理对象在该单项的得分，由上一节计算得到；第三项为单项在计算综合积分时的权重配置信息，从配置数据库获取。令总积分为, 单项积分为, 单项积分权重为，则 由以下公式计算得出：

即

### 积分规则模型列表

以下是积分模型判断指标：

**1）、基本信息情况**

1. 来源地：

|  |  |
| --- | --- |
| 定义 | 针对外来流动人口的来源地是否为高危地区 |
| 数据来源 | 居住人口高危居住表，流动人口表 |
| 规则  （判断条件+参数） | 流动人口来源地是否为高危区 |

1. 民族：

|  |  |
| --- | --- |
| 定义 | 外来流动人口及常住人口的民族为汉族或者其他少数民族 |
| 数据来源 | 常住人口表，流动人口表 |
| 规则  （判断条件+参数） | 除去汉族之外重点关注维族、回族等少数民族人口  汉族或其他 |

3、年龄段：

|  |  |
| --- | --- |
| 定义 | 对流动人口及常住人口进行年龄段划分，重点关注青年及中年人口 |
| 数据来源 | 常住人口表，流动人口表 |
| 规则  （判断条件+参数） | 通过年龄进行划分 |

4、婚姻状况

|  |  |
| --- | --- |
| 定义 | 流动人口及常住人口婚姻状况为已婚、未婚、离异 |
| 数据来源 | 常住人口表，流动人口表 |
| 规则  （判断条件+参数） | 婚姻状态  已婚、未婚、离异 |

5、名下财产

|  |  |
| --- | --- |
| 定义 | 名下是否登记个人财产，主要财产为房产、汽车 |
| 数据来源 | 机动车信息表、房产权属登记表 |
| 规则  （判断条件+参数） | 名下登记是否房产、汽车、其他财产 |

6、名下电话登记情况

|  |  |
| --- | --- |
| 定义 | 名下登记过的通信电话 |
| 数据来源 | 常住人口表、流动人口表 |
| 规则  （判断条件+参数） | 名下登记三个或以上电话号码 |

7、是否办理居住证

|  |  |
| --- | --- |
| 定义 | 主要针对流动人口是否办理暂住登记，持有居住证件 |
| 数据来源 | 常住人口表、流动人口表 |
| 规则  （判断条件+参数） | 是否办理居住登记 |

8、是否市民卡

|  |  |
| --- | --- |
| 定义 | 是否办理杭州市民卡 |
| 数据来源 | 市民卡信息表 |
| 规则  （判断条件+参数） | 是否办理杭州市民卡 |

9、自身前科

|  |  |
| --- | --- |
| 定义 | 自身是否为前科人员 |
| 数据来源 | 打防控违法人员，全国在逃人员，全国重点人员，执法办案取保人员，执法办案相关人员，禁毒涉毒嫌疑人 |
| 参数规则  （判断条件+参数） | 是否为前科人员 |

10、强关系人员前科情况

|  |  |
| --- | --- |
| 定义 | 强关系人员包括朋友、亲戚等，是否存在有前科人员 |
| 数据来源 | 打防控违法人员，全国在逃人员，全国重点人员，执法办案取保人员，执法办案相关人员，禁毒涉毒嫌疑人 |
| 规则  （判断条件+参数） | 强关系人员是否有前科 |

**2）、居住情况**

1、住处

|  |  |
| --- | --- |
| 定义 | 住处 |
| 数据来源 | 常住人口表，流动人口表 |
| 规则  （判断条件+参数） | 通过是否有固定居处，自购还是租赁判断住所是否稳定 |

2、居住时长

|  |  |
| --- | --- |
| 定义 | 居住时长 |
| 数据来源 | 常住人口表，流动人口表 |
| 规则  （判断条件+参数） | 居住时间长短 |

3、同居情况

|  |  |
| --- | --- |
| 定义 | 同行居住情况 |
| 数据来源 | 流动人口表 |
| 规则  （判断条件+参数） | 根据居住情况判断，依据有否有随居家人或其他身份同居者 |

4、出租房登记情况

|  |  |
| --- | --- |
| 定义 | 出租房、公租房 |
| 数据来源 | 流动人口表 |
| 规则  （判断条件+参数） | 出租房登记情况 |

**3）、工作情况**

1、工作情况

|  |  |
| --- | --- |
| 定义 | 工作情况 |
| 数据来源 | 流动人口表，社保登记 |
| 规则  （判断条件+参数） | 是否有工作 |

2、社保缴纳

|  |  |
| --- | --- |
| 定义 | 社保缴纳信息 |
| 数据来源 | 公积金信息，社保登记 |
| 规则  （判断条件+参数） | 根据社保表数据信息，缴纳起始时间及时长、缴纳单位及缴纳单位变更情况 |

3、工作稳定性

|  |  |
| --- | --- |
| 定义 | 工作稳定性 |
| 数据来源 | 人员信息库，三实表、社保 |
| 规则  （判断条件+参数） | 工作是否稳定，可通过相应的社保缴纳情况侧面反映 |

4、工作类别

|  |  |
| --- | --- |
| 定义 | 工作类别 |
| 数据来源 | 娱乐场所从业人员，典当行业从业人员 |
| 规则  （判断条件+参数） | 是否属于特殊行业 |

5、就业培训参加情况

|  |  |
| --- | --- |
| 定义 | 是否参加由社保局组织的就业培训 |
| 数据来源 | 社保就业信息库 |
| 规则  （筛选条件+参数） | 是否有参加就业培训 |
| 问题 | 待采集 |

### 规则模型使用数据清单

1. 高危地区人员

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 表名 | 英文名 | 字段 | 值 | 来源 |
| 1 | 流动人口表 | gazhk\_NEWZZRK\_JB\_TAB | 户籍地 |  | 市云 |
| 身份证号码 |  |
| 注销标志 | "未注销" |
| 区县 | "萧山" |
| 2 | 高危地区表 |  | 高危地区代码 | 自动输入 |  |

1. 民族

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 表名 | 英文名 | 字段 | 值 | 来源 |
| 1 | 常住人口表 | gazhk\_CZRK | 民族 |  | 市云 |
| 身份证号码 |  |
| 有效标志 | "有效" |
| 户籍地 | "萧山" |
| 2 | 流动人口表 | gazhk\_NEWZZRK\_JB\_TAB | 民族 |  | 市云 |
| 身份证号码 |  |
| 注销标志 | "未注销" |
| 区县 | "萧山" |

1. 年龄段

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 表名 | 英文名 | 字段 | 值 | 来源 |
| 1 | 常住人口表 | gazhk\_CZRK | 出生日期 |  | 市云 |
| 身份证号码 |  |
| 有效标志 | "有效" |
| 户籍地 | "萧山" |
| 2 | 流动人口表 | gazhk\_NEWZZRK\_JB\_TAB | 出生日期 |  | 市云 |
| 身份证号码 |  |
| 注销标志 | "未注销" |
| 区县 | "萧山" |

1. 婚姻状况

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 表名 | 英文名 | 字段 | 值 | 来源 |
| 1 | 常住人口表 | gazhk\_CZRK | 婚姻状况 |  | 市云 |
| 身份证号码 |  |
| 有效标志 | "有效" |
| 户籍地 | "萧山" |
| 2 | 流动人口表 | gazhk\_NEWZZRK\_JB\_TAB | 婚姻状况 |  | 市云 |
| 身份证号码 |  |
| 注销标志 | "未注销" |
| 区县 | "萧山" |

1. 名下财产

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 表名 | 英文名 | 字段 | 值 | 来源 |
| 1 | 房屋权属登记表 | gazhk\_DZZWB\_FWQSXX | 房屋地址 |  | 待采集 |
| 身份证号码 |  |  |
| 是否有房 | "是" | 市云 |
| 2 | 机动车登记表 | gazhk\_JTXT\_HZJDCXX\_NEW | 身份证号码 |  |  |
| 机动车号牌号码 |  |  |
| 机动车状态 | 非注销 | 市云 |

名下电话登记

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 表名 | 英文名 | 字段 | 值 | 来源 |
| 1 | 常住人口表 | gazhk\_CZRK | 联系电话 |  | 市云 |
| 身份证号码 |  |
| 有效标志 | "有效" |
| 户籍地 | "萧山" |
| 2 | 流动人口表 | gazhk\_NEWZZRK\_JB\_TAB | 联系电话 |  | 市云 |
| 身份证号码 |  |
| 注销标志 | "未注销" |
| 区县 | "萧山" |

1. 市民卡办理

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 表名 | 英文名 | 字段 | 值 | 来源 |
| 1 | 市名卡信息 | gazhk\_DZZWB\_ZX\_CARD | 市民卡卡号 |  | 市云 |
| 身份证号码 |  |
| 卡状态 | "正常" |

1. 自身前科

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 表名 | 英文名 | 字段 | 值 | 来源 |
| 1 | 打防控违法人员 | gazhk\_DFK\_WFFZRY | 案件编号 |  | 市云 |
| 案件类别 |  |
| 抓获单位 |  |
| 抓获日期 |  |
| 身份证号码 |  |
| 删除标志 | "未删除" |
| 2 | 全国在逃人员 | gazhk\_ZTRY\_QG\_ZTRY\_JBXX\_II | 案件编号 |  | 市云 |
| 立案时间 |  |
| 案件类别 |  |
| 身份证号码 |  |
| 注销标志 | "未注销" |
| 3 | 全国重点人员 | gazhk\_QBPT\_ZDRYJBXX | 人员类别 |  | 市云 |
| 入库时间 |  |
| 身份证号码 |  |
| 有效性 | "有效" |
| 4 | 执法办案取保人员 | gazhk\_ELES\_SQBZJB | 案件编号 |  | 市云 |
| 案件名称 |  |
| 发生日期 |  |
| 身份证号码 |  |
| 5 | 执法办案相关人员 | gazhk\_ELES\_RYAJ | 案件编号 |  | 市云 |
| 案件名称 |  |
| 案件性质 |  |
| 人员类型 |  |
| 发案时间 |  |
| 身份证号码 |  |
| 6 | 禁毒-涉毒嫌疑人 | gazhk\_QBPT\_JD\_SDXYR | 嫌疑人类型 |  | 市云 |
| 填报单位 |  |
| 登记日期 |  |
| 身份证号码 |  |

1. 住所

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 表名 | 英文名 | 字段 | 值 | 来源 |
| 1 | 常住人口表 | gazhk\_CZRK | 登记住址 |  | 市云 |
| 身份证号码 |  |
| 有效标志 | "有效" |
| 户籍地 | "萧山" |
| 2 | 流动人口表 | gazhk\_NEWZZRK\_JB\_TAB | 暂住地址 |  | 市云 |
| 身份证号码 |  |
| 注销标志 | "未注销" |
| 区县 | "萧山" |

1. 居住时长

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1 | 常住人口表 | gazhk\_CZRK | 迁入日期 |  | 市云 |
| 身份证号码 |  |
| 有效标志 | "有效" |
| 户籍地 | "萧山" |
| 2 | 流动人口表 | gazhk\_NEWZZRK\_JB\_TAB | 登记日期/最早登记日期 |  | 市云 |
| 身份证号码 |  |
| 区县 | "萧山" |

同居情况

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 表名 | 英文名 | 字段 | 值 | 来源 |
| 2 | 流动人口表 | gazhk\_NEWZZRK\_JB\_TAB | 登记日期 |  | 市云 |
| 到期日期 |  |
| 暂住详址 |  |
| 身份证号码 |  |
| 注销标志 | "未注销" |
| 区县 | "萧山" |

1. 出租房登记情况

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 表名 | 英文名 | 字段 | 值 | 来源 |
| 2 | 流动人口表 | gazhk\_NEWZZRK\_JB\_TAB | 租赁房屋档案号 |  | 市云 |
| 身份证号码 |  |
| 注销标志 | "未注销" |
| 区县 | "萧山" |

1. 工作情况

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 表名 | 英文名 | 字段 | 值 | 来源 |
| 1 | 流动人口表 | gazhk\_NEWZZRK\_JB\_TAB | 工作单位 |  | 市云 |
| 身份证号码 |  |
| 注销标志 | "未注销" |
| 区县 | "萧山" |
| 2 | 社保登记 | gazhk\_ZFZYK\_SBXX | 工作单位名称 |  | 市云 |
| 身份证号码 |  |
| 参保日期 |  |
| 参保状态 |  |

1. 工作类别

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 表名 | 英文名 | 字段 | 值 | 来源 |
| 1 | 娱乐场所从业人员 | gazhk\_ZAZD\_YLCSCYRYXX | 单位名称 |  | 市云 |
| 身份证号码 |  |
| 2 | 典当行从业人员 | gazhk\_ZAZD\_DDHCYRY | 单位名称 |  | 市云 |
| 身份证号码 |  |
| 3 | 其它特殊行业人员登记表 |  | 单位名称 |  | 待采集 |
| 身份证号码 |  |

1. 社保缴纳

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 表名 | 英文名 | 字段 | 值 | 来源 |
| 1 | 公积金信息 | gazhk\_DZZWB\_GJJXX | 单位名称 |  | 市云 |
| 身份证号码 |  |
| 缴存状态 |  |
| 最近缴存日期 |  |
| 2 | 社保登记 | gazhk\_ZFZYK\_SBXX | 工作单位名称 |  | 市云 |
| 身份证号码 |  |
| 参保日期 |  |
| 参保状态 |  |

1. 就业培训记录

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 表名 | 英文名 | 字段 | 值 | 来源 |
| 1 | 就业培训信息表 |  | 培训时间 |  | 待采集 |
| 身份证号码 |  |
| 培训项目 |  |
| 培训机构 |  |

## 绩效考核方案

### 考核目标

公安机关的绩效考核与评估是一种部门组织用来衡量和评鉴警员某一时段的工作表现，并将结果用来激励民警、反馈民警工作效率、协助民警职业成长、了解单位现有的人力资源及未来的策略方向等功能。评估的结果可作为薪酬与职务调整的依据、决定训练的需求、改进工作和规划生涯以及协助上级领导及时准确的了解民警思想和工作现状。绩效评估对各个处室的健康发展都有正面的意义，如果能利用好绩效评估的工具及手法，正确有效地评估我国公安部门的绩效，并做出适当的反馈，那么我国公安部门的管理制度打下很好的根基，也才能促进公安机关科学健康发展，切实达到为国家和人民服务的根本目标。

与其他部门不同，公安部门需要解决各种各样的经济、政治和社会问题，所以考核目标更加灵活、弹性与多样化，这样就使得公安机关人民警察绩效的评定变成一个非常复杂的任务。

公安机关现有的考核种类繁多，几乎每个业务都有考核任务，但多头考核、多线考核由于考核部门和具体执行标准的不同，必然导致考核缺乏统一标准和不规范性，最终造成考核不能发挥应有的作用。

近几年，公安机关在信息现代化投入了大量的人力物力，并取得了一定的成效，但由于缺少相应的系统规划和具体的实践应用，尤其是信息技术手段在日常管理中的强大功能还没得到具体应用，内部管理的水平和能力提高的潜力很大。公安部门追求的目标具有多样性和弹性，目标的多样性表现为不同的公安部门追求政治、经济、社会等不同性质的目标，甚至同一部门追求多重目标且经常发生变化。目标的弹性即软目标，表述笼统、抽象、且难以量化为硬性指标，这为公安警务绩效的考核带来了一定的难度。由于以上原因，公安民警绩效的测定及评估就比较复杂和困难。本系统依据研究公安部门与辖区的特性，规划与设计一套警务机关绩效评量机制。

公安机关的工作规划与任务的评估也涉及民警的绩效考核，例如在涉及考量一项新的工作任务时，一定要参考负责实施的民警在工作任务布置前后的绩效状态，或将实施任务的单位绩效与未实施任务的单位绩效作一比较。因此，绩效考核对单位政策或任务的制定、完善是必需的。

主要是根据考核结果来作为单位与工作人员的调整依据与参考，考核的目的包括：

（一）作为任务进度的展现

（二）作为升迁调遣的标准

（三）作为教育训练的依据

（四）作为警员自我改进的了解依据

通过绩效考核的无形指挥棒，利用信息化手段，创造良好的上下级沟通环境，更高效，更科学的管理整个公安系统的运营与发展。

### 考核机构与主体

#### 考核机构

根据公安机关的领导组成来确定考核领导小组与的绩效考核办公室组成，选出考核领导小组的组长、副组长、成员与考核办公室主任（可以分别由公安行政主要领导、党委领导和相关处室负责人来担任）。

#### 考核主体

考核对象是公安机关的在职在编民警。下面例举一个县级公安机关中，其各级单位及组织功能与职责概述如下：

1.局长：负责所属公安机关全面工作。

2.政委、副局长：协助局长办理所属公安机关各警种相关业务工作。

3.第一组：负责公安机关纪检委、政工、指挥中心、通信、行政服务、出入境管理、装备保障、信访等业务。

4.第二组：负责公安机关督察、教育训练、勤务督导等业务。

5.第三组：负责公安机关所属辖区的刑事、治安犯罪侦查、处置突发性事件等业务。

6.第四组：负责公安机关国内安保业务。

7.第五组：负责公安机关所属辖区的交通维护、信息管理等业务。

8.第六组：负责公安机关所属辖区的社会安防工作、外国人管理等业务。

9.勤务中心：负责公安机关全局的指挥调度、公文收发、上传下达、会议纪录、印章管理等业务。

10.派出所：负责所属辖区的巡逻检查、值班、治安管理和案件办理、户籍业务办理等各项警察工作及为民服务等业务。

#### 考核的主要警务工作内容

（一）维护良好治安：以警民合作为基础，加强辖内情报掌握及治安要点安全维护，执行巡逻、勤区查察、守望临检勤务，并因治安特性规划项目性勤务，以维护辖内治安。

（二）防制交通事故发生：落实各项交通执法及疏导管制，确保行车安全、鼓励提供道路改善方案，降低事故发生；为提升交通事故处理质量及落实专业化理念。

（三）加强为民服务：贯彻积极为民服务政策，推行刑事案件单一窗口受理民众报案；协助处理孤寡老人、留守儿童、夜归妇女安全、加强举家外出住户、机关、学校安全维护；派出所值班台设置为民服务衔牌、服务台及民众公共休息区，于受理民众报案时，由备勤警察专责处理，随时为民服务。

### 考核标准

决定考核标准是绩效考核里的第一个步骤，在决定绩效考核标准之前，必须由绩效考核委员会或绩效考核小组，研究讨论并制定考核标准。为实现考核目的，对于不同层级的单位或民警，自应制定不同的考核标准。

考核的第二步骤必须先对所制定的考核标准的重要权重与各指标可选择的评分结果（评分值）定义其语意变量所对应的口语项，以及各个口语项所对应的模糊数，以供后续进行绩效考核。

根据公安机关绩效考核管理的可行性分析，考核标准在公安机关原有的考核标准上加以修改，首先将纵向整理各警种的业务职责和横向整理各级单位的业务职责，并且予以分类，经由多公安各警种和单位开会讨论后，并根据所提供的部门职责的相关文献，以制定出本考核研究的标准。

采用公安系统的绩效考核程序、计算方式与搜集绩效考核指标。接下来则是对于公安机关绩效考核程序、计算方式与绩效考核指标进行说明：

（一）考核程序是由部门内职位最高的主要负责人对中层负责人进行考核：部门内职位最高的主要负责人通常为“局长”或“政委”，而且每个中层部门内只有一位负责人，在考核表上对所属单位民警全年的工作表现给予评价并且也对自己进行自评，再将考核表交回政工部门。政工部门将各部门的考核结果整理后，呈报给分管负责人进行审核。

（二）考核计算方法是使用传统明确数值方法：我国公安机关是使用传统的明确数值的方法，直接在考核表上打分数再进行加总，最后依据加总后结果对单位内部民警进行排序。

（三）不同级别的民警使用不同考核表：公安机关考核表按照职责不同可分为两类：共性目标与业务目标。行政部门会以民警的等级来发放考核表，以供主要负责人评分。

* 共性目标的考核标准共有四大类别，包含党建工作、内部管理、教育训练和职业道德，第一大类“党建工作”里再分三项指标：分别为贯彻落实党的十七大和科学发展观、“三会一课”制度落实、参加党组织；第二大类“内部管理”里有四项指标：分别为严格落实“内务条令”、严格备勤和值班制度、办公环境整洁、各项规章制度落实；第三大类“教育训练”里分三项指标：分别为政治理论、业务学习、训练制度落实、领导审阅卷案制度；第四大类“职业道德”里再分三项指标：分别为日常行为规范、服务群众、社会公德，总计共有十三项考核标准。
* 业务目标的考核标准根据单位或个人所承担的业务制定，大致分为两大类别:“主体工作”和“临时工作”。

（四）考核标准因民警等级的不同而有所差异：不同的等级民警的考核标准会不尽相同，会依照该等级民警所需具备之工作能力与执行之工作项目来选取考核标准，因此“共性目标”与“业务目标”的指标会有区别，并非完全一致。

（五）依照单位内最终排序而给予不同的考核结果。考核结果将作为评判民警一年来工作情况的主要标准，与奖惩挂钩，与职务晋升挂钩。

## 数据资源管理

数据资源管理是人口精准管控系统建设的基础，在管控平台建设过程中，数据资源管理就是将人口各个信息系统的数据，进行采集、抽取、转换、加载，集中分类归集，形成统一、完整的数据库结构。

### 数据采集

数据采集，主要从市局信息资源平台及本地数据资源库两大平台进行采集。

数据通过ETL工具从其他信息系统清洗数据入库外，本地采集的离线零散数据功能可将用户自身所掌握的各类表格、文本数据导入到系统，在实现数据共享和搜索的同时，也可通过计算平台对用户上传的数据进行挖掘研判分析。

### 数据归集

本系统需要整合的数据主要从居住、职业、出行记录、消费记录、活动轨迹、社交圈六个维度进行采集和汇聚，积极协调各一线部门获取公安网、政务网、社会及互联网中涉及人口管理的系统资源，通过归集各系统涉及人员公共安全相关数据，数据导入系统后，将采集导入的数据按数据目录进行归集整合，进而通过分布式特征标签计算技术，将人员社会安全特性标签化。

### 数据清洗规则

将已采集与归集的所有人口信息数据进行目录分类，梳理提取其中可能影响人口公共安全属性的部分，确定为本积分管理模块的数据分类项。

将最终确定的人口精准管控积分计算分为基础分、扣减分、附加分等三大类和基本情况、租住情况、工作情况、关系人情况、所有物情况等五小类。

### 数据储存

流动人口精准管控系统采用分布式技术建库。通过综合使用Apache Hadoop，EDW，MongoDB，MySQL等各类数据库技术，可以在通用的PC服务器集群上面，建设集中的数据库集群系统。当一台服务器出现故障时，可以由另一台服务器承担服务任务，从而在不需要人工干预的情况下，自动保证系统能持续提供服务。分布式数据特有的横向扩展，线性性能提升、map-reduce原生支持都特点，特别适合人口数据资源库数据总类多，数据量巨大，数据访问频繁等特点。通过ETL工具，支持从ORACLE、MYSQL、SQL-SERVER、文件、接口等不同类型的数据源，实时或定时将数据清洗到数据资源库中。

### 数据落地

为保证数据安全，系统稳定性，系统提供数据服务的相关接口功能。服务接口开发工作包括系统权限及功能模块接口，各类业务数据标准地址定位数据导入接口，基本信息数据导入接口，区县零散数据导入接口，卡口数据导入接口，常口数据导入接口，人员照片库整合导入接口，人员照片服务模块SOA服务接口实现，人员主题库搜索服务SOA服务接口实现，零散数据搜索服务SOA服务接口实现，卡口数据搜索服务SOA服务接口实现，系统数据搜索服务SOA服务接口实现。

1. 各类业务数据标准地址定位数据导入接口

整理各类业务数据的地址字段，根据相应的标准地址定位接口，清洗转换业务数据地址的定位信息。如常住人口的详细地址的经纬度信息。

2. 人员信息数据导入接口

整合相关业务数据，数据导入大数据平台，并生成搜索索引。

3. 其他零散数据导入接口

开发零散数据导入接口，主要针对的是区县民警掌握的零散数据信息，充分利用大数据平台，将零散数据导入大数据平台，并生成搜索索引。

4. 卡口数据导入接口

整合区县车辆卡口数据至大数据平台，并与市区车辆卡口对接，实现数据相互推送，完成全市车辆卡口数据的整合。

5. 人员主题数据导入接口

将各类业务数据作统一主题分类，充分利用大数据平台，将现有数据主题化，要求能实时增量数据主题化。将相应的主题化数据导入大数据平台。

6. 人员照片库整合导入接口

整合人员所有照片信息，包含常住人口、在逃人员、打防控嫌疑人员、居住人口等。将照片数据分文归类，统一导入大数据平台。

通过以上接口的开发提供更多的数据导入服务，加快人口精准管控系统资源整合工作。

## 系统应用功能模块

### 智能数据检索

智能数据检索应用是综合的大数据云全文智能搜索软件。通过建立统一的大数据云平台，对基础数据、外部线索、其他非结构化生成索引，并实时同步和更新。智能搜索应用使用统一SOA标准数据访问接口，实现数据资源及访问方式的整合，屏蔽业务系统之间的数据差异，突破工种和地域的限制，实现数据全局共享和分析处理，大大推动行业创新应用设计，促进行业信息化水平进一步提高，同时为行业新一代应用架构的发展奠定新的基础。

全能搜索主要是对用户提供的搜索关键字进行全库搜索，找出匹配的关键信息，这里的全能搜索包含业务主题数据库的挖掘分析搜索和分类业务数据库的全面检索。智能搜索应用也充分利用了数据的全文索引进行数据搜索匹配，同时也对特定的数据属性进行定向检索匹配，以网购式行为完成数据精确匹配搜索。

### 报表统计可视化

平台提供自动报表统计分析功能，可以按照数据搜索返回结果的不同，自动生成报表统计，分析维度和显示效果可以实时调整，动态生成。针对人口信息、人口异动趋势、辖区内流动人口分布等信息数据的分析结果，进行可视化统计图展示，为管理部门决策及相关资源的分配调度提供依据。



### 历史记录追溯查询

使用高性能数据仓库进行特定人员社会安全指数变化历史和趋势的快速回溯和掌握。历史记录查询模块，是根据用户权限，各自查询本辖区录入的流动人口的基本信息，可以查询全部的流动人口的详细信息，查询条件可根据身份证号码、姓名等条件进行筛选数据，快速定位到每条记录，查看每条记录的详细情况。从流动人口信息，可以关联到居住信息、社保信息，从社保信息又可以关联到工作信息，完成流动人员历史追溯。

### 实时人员异动趋势显示

系统运用hadoop、Spark，分布式储存数据，实时处理更新数据库，可实现24小时不间断实时监控功能，完成人员异动趋势，行为轨迹的实时监测。特别关注重点人员及辖区内的人员异动趋势，实时计算个人及总体的健康指数积分，为人员布控及警力分布提供决策依据。

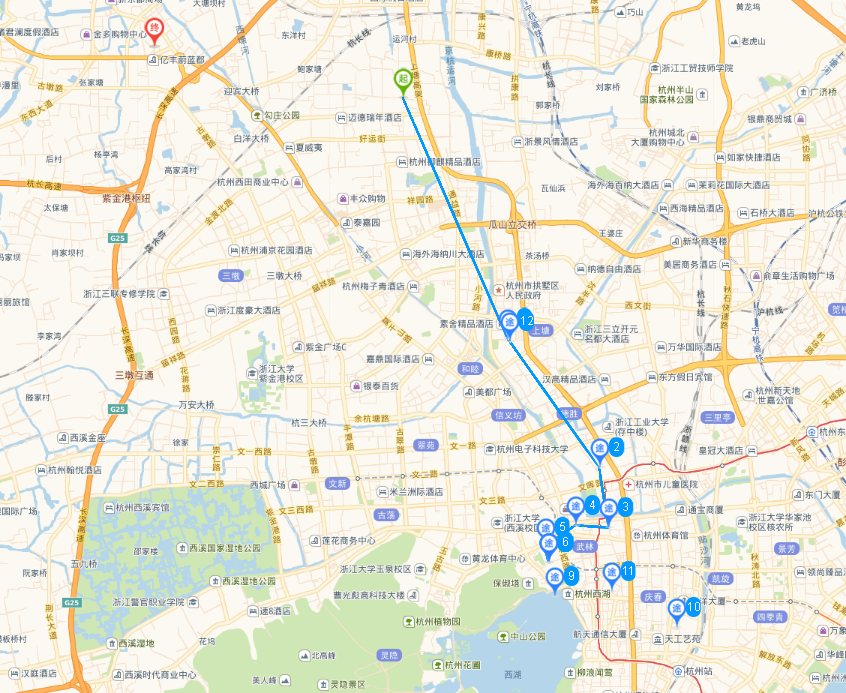


图6—4 人员行为轨迹动态图

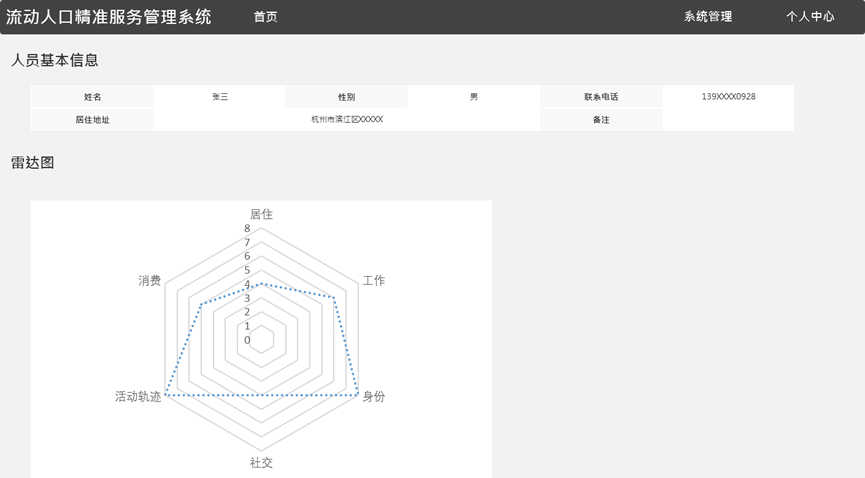
### 可疑人员预警功能

使用Spark等实时流处理引擎进行人员的实时异动检测和预警管控。可疑人员预警功能，在人口信息采集后、系统根据预设模型规则研判之后，系统会将对人员身份信息、居住信息、工作信息进行多方面、对维度比对，将可疑人员纳入可疑人员名单库的信息，系统将实时对其进行监测，关注其实时异动情况，自动发出预警提醒，决定是否采取布控行为，消除潜在危机，以消除不稳定因素的发生。



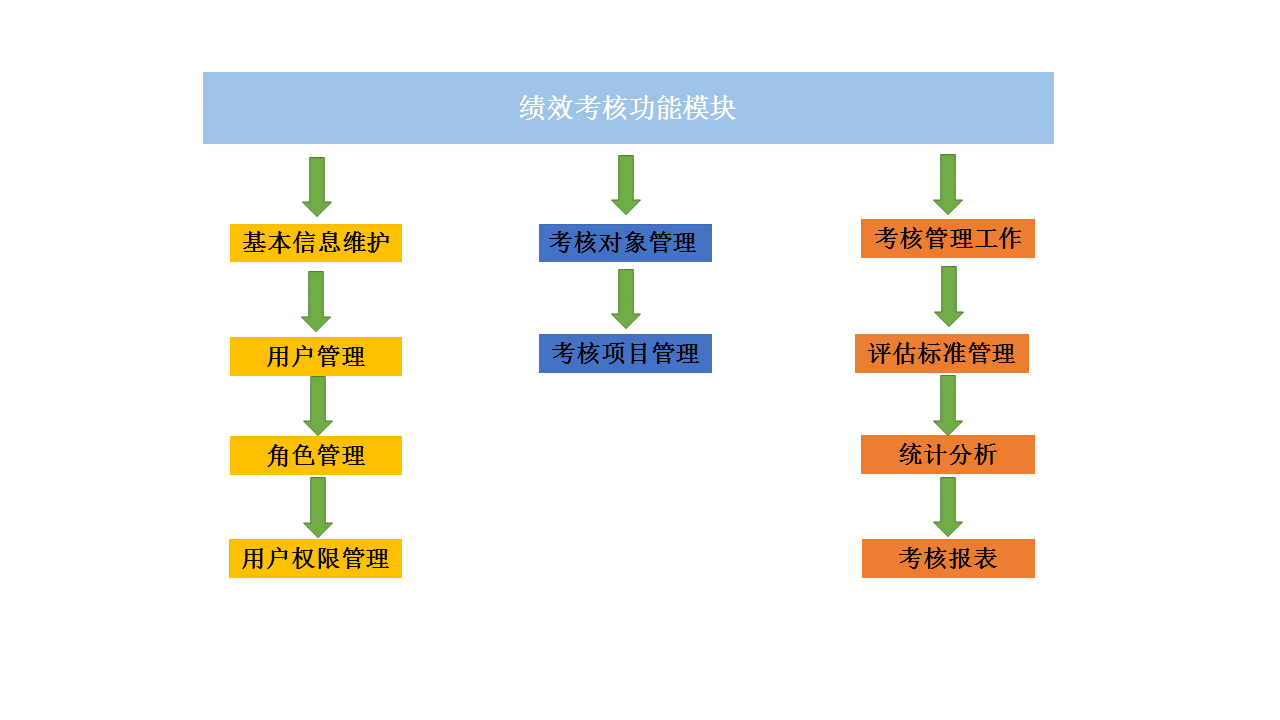
### 人员信息多维画像显示

基于人口信息基础表与人口精准积分模型相结合分析，每个人创建独立的多维社会安全特征矩阵，基于社会学理论研究成果，使用RFM、K-means、决策树、SVM、神经网络等技术对人员进行多维立体精准画像。将分析结果多维雷达图进行展示，一并展示个人详细档案。



### 绩效考核模块

#### 基础信息维护模块



**1. 角色管理**

 角色定义为具有某个职能权限的人员，如“部门领导”，“一般干警”、“管理员”等。所设计的公安绩效评估系统可以提供对不同角色的用户信息的管理。

 用户组中成员的删除、增加与修改。

**2. 权限分配**

用户管理公安绩效评估系统能够对各种类型的用户分配权限，之后才允许各个用户的访问与信息编辑等内容。

**3. 系统信息管理**

公安绩效评估系统能够及时的发布最新信息，以利于用户的查看。

**4. 数据备份**

公安绩效评估系统能够对系统的数据与信息进行热备份，以及离线备份的功能。

#### 考核管理模块

公安绩效评估系统能够实现绩效考核的评分处理。也可以允许用户在一定的范围之内调整评分结果。

* ** 评分查询**

公安绩效评估系统能够查询整体的评分结果，以及按照检索的关键词进行查分。

*  **统计分析与报表系生成**

 公安绩效评估系统能够从整体的评分结果中汇总出统计结果，并且显示与打印，生成相关的报表。

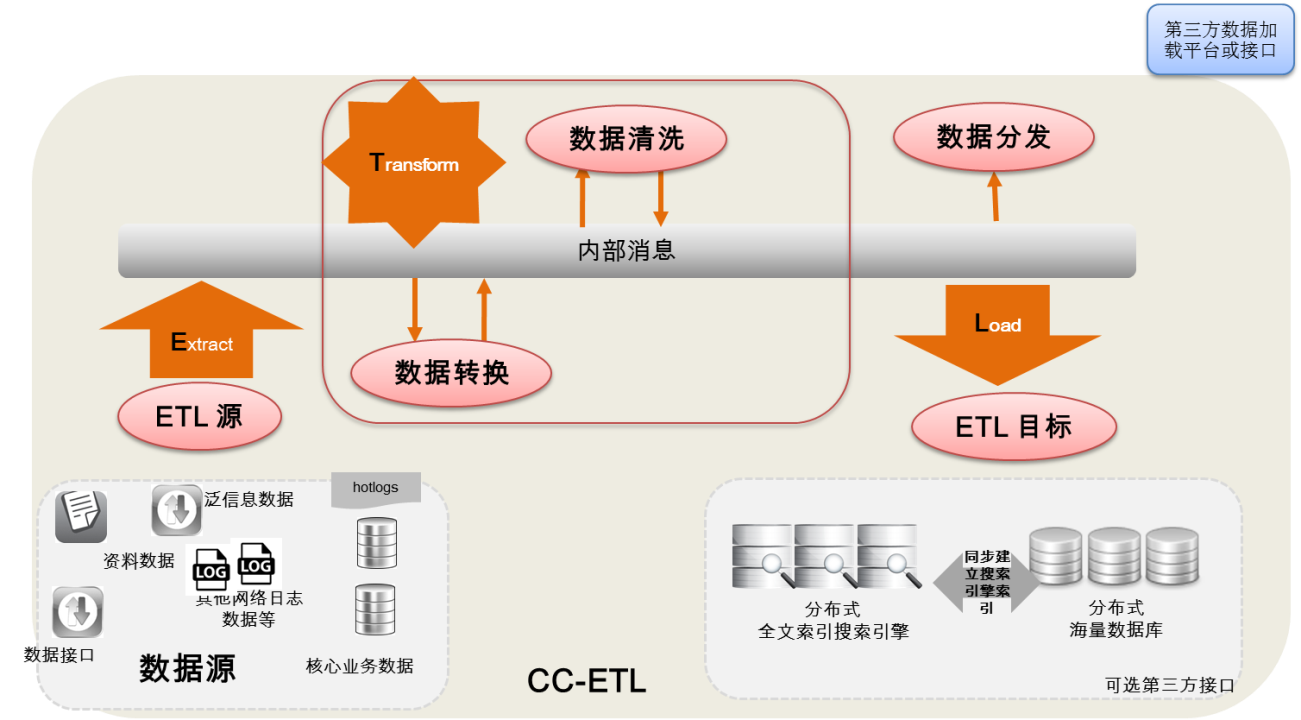
# 系统核心技术

## CC-ETL数据处理软件

### CC-ETL基础介绍

匡信大数据处理软件（以下简称CC-ETL软件），可以帮助行业用户实现多个业务系统间跨平台的原始数据整合快速实现多种数据源的数据抓取、数据转换与处理、数据质量控制及目标数据库加载，适用于将多个业务系统的海量、异构原始数据进行整合，主要作为大数据环境下数据仓库、商业智能等应用的ETL数据整合工具，也可以很好地支持目前各大行业和领域的数据大集中、主题库/基础库、运营数据存储等应用需求。

CC-ETL通过核心组件和插件结合的方式实现数据的处理。对主流数据库如Oracle、MySQL、HBASE等数据库等以插件的方式进行开发，提供数据接口（其他数据库支持插件可定制）。并在ETL中对主流数据库与数据中心数据库建立字段映射关系，因为每种数据的字段类型和命令都有所不同，以数据中心为标准进行规范统一。



CC-ETL总体架构

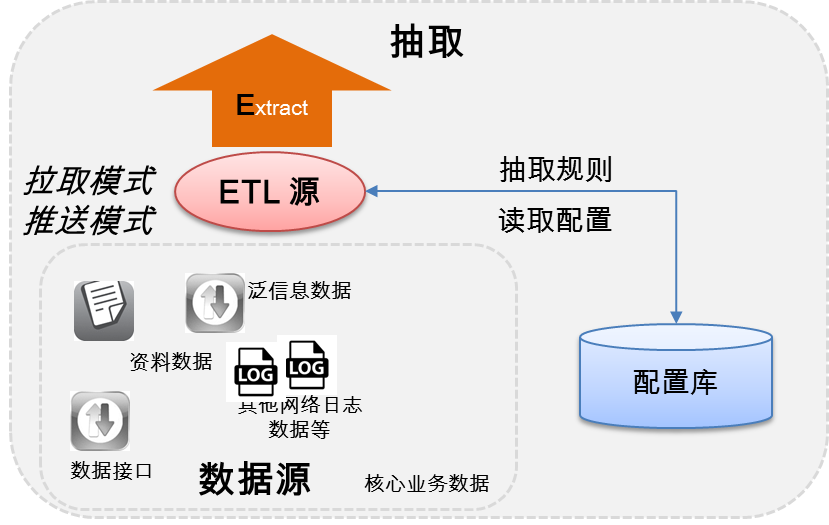
* **数据抽取：**通过各种渠道进行数据采集（存储与内存、数据库)
* **数据转换：**将各种形式的数据转换为规范格式的数据。
* **数据清洗：**将脏数据和不完整的数据过滤掉。
* **数据分发：**将订阅数据、关注数据定期推送给用户。
* **数据加载：**将所有有效数据存储起来供用户查询。

### CC-ETL功能设计

#### 数据抽取

数据抽取过程是ETL的第一个阶段。提供类似水泵的抽取动力源。需要支持众多的输入源：

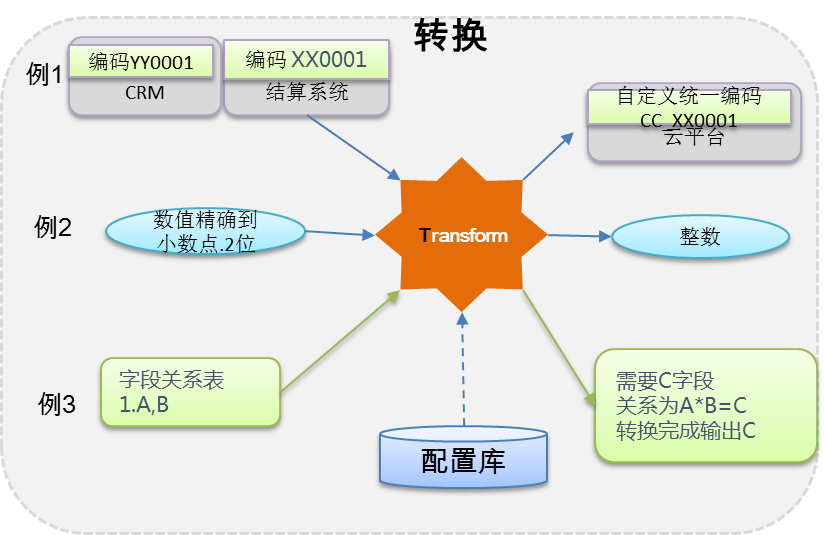
* 支持多数据库输入方式，包括Oracle/SQL Server/DB2/My SQL等
* 接口方式输入格式化数据，http/socket/webService，
* 在提供传统全量输入的同时，提供**增量输入**（输入后期增加、删除的数据，大大减少后期维护工作、节约资源、提高性能、降低成本）



#### 数据转化

数据转换的主要任务是进行不一致的数据转换、数据粒度的转换、以及一些业务规则的计算。基本任务包括：数据的选择、分离/合并、转化、汇总和丰富（变形）。

* 不一致数据转换：这个过程是一个整合的过程，将不同业务系统的相同类型的数据进行统一，比如同一个订单在结算系统的编码是XX0001，而在CRM中编码是YY0001，这样在抽取过来之后统一转换成一个编码。
* 数据粒度的转换：业务系统一般存储非常明细的数据，而数据仓库中数据是用来分析的，不需要非常明细的数据。一般情况下，会将业务系统数据按照数据仓库粒度进行聚合。
* 业务规则的计算：不同的企业有不同的业务规则、不同的数据指标，这些指标有的时候不是简单的加加减减就能完成，这个时候需要在ETL中将这些数据指标计算好了之后存储在数据仓库中，以供分析使用。



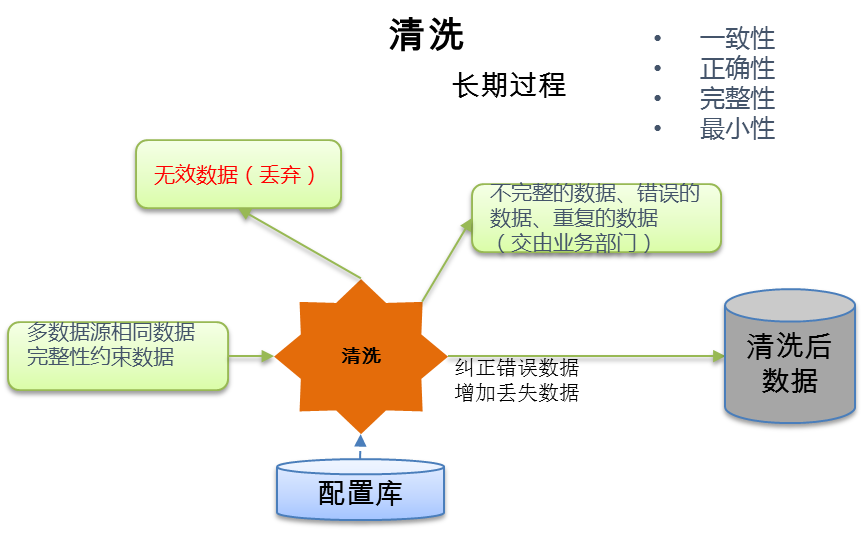
#### 数据清洗

数据清洗保证清洗后数据的数据质量，减少错误和不一致性、解决对象识别的过程

确保清洗后的数据满足一致性（consistency）、正确性(correctness)、完整性(completeness)和最小性(minimality) 4大要求。

单数据源模式层问题（如缺少完整性约束、糟糕的模式设计等）、单数据源实例层问题（如数据输入错误）、多数据源模式层问题（如异构数据模型和模式设计等）、多数据源实例层问题（如冗余、冲突、不一致的数据等）

* 对数据进行规范化，把错误数据进行纠正，缺少数据进行增加
* 对数据进行检查，把无效数据进行过滤，阻止输出
* 数据清洗的任务就是过滤掉那些不符合要求的数据，将过滤的结果交给业务部门，由业务部门确认是否修正后再进行抽取，三大类不符合要求的数据：不完整的数据、错误的数据、重复的数据。
* 数据清洗是一个反复的过程，不可能在几天内完成，只有不断的发现问题、解决问题。

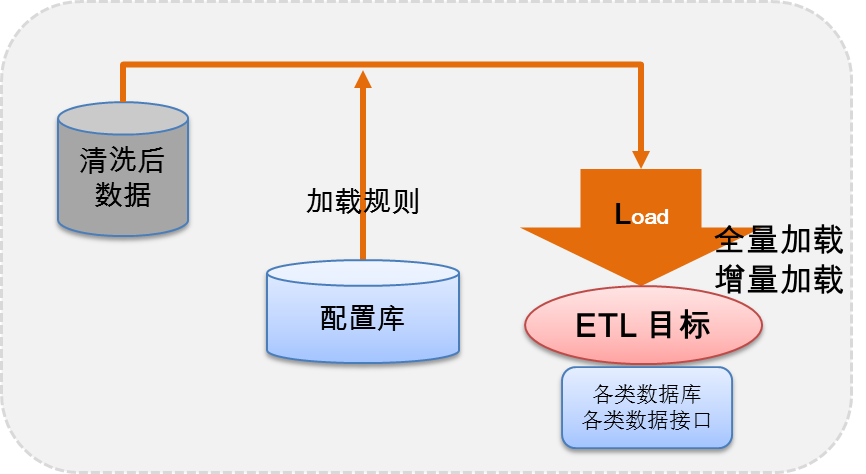


#### 数据加载

数据加载过程即水泵的水注入到水池或者另外水管的过程。

提供数据的初始装载、增量装载和完全刷新几个方面的功能

* 支持多数据库输出方式，包括Oracle/SQL Server/DB2/My SQL、kudu等
* 适配器方式提供nosql方式输出mongo/hbase
* Kafka/Mq/http/webservice等接口的输出
* 供所有有效（全量）数据的查询。



## CC-Albus数据服务总线

### ALBus服务总线定义

ALBus服务总线技术是一个基于标准化协议和框架的服务接口总线产品，属于自主开发，完全自主版权的软件产品技术。



ALBus服务总线为业务逻辑实现提供了基于多种通讯和数据协议的业务服务接口调用解决方案，拥有以下优点：

* 灵活的部署模式
* 敏捷高效的业务扩展、变更与伸缩
* 便捷丰富的前后端（业务使用者和业务提供者）开发手段
* 可配置、可插拔的系统功能扩展
* 前后端多重缓存技术支持，确保高并发下保持业务的快速响应

从而能良好的适配各种业务场景、业务量和业务模式。

### ALBus服务总线概念

1. **接口**

业务接口的缩写。在ALBus服务总线中一般“接口”和“业务”是同义词。

在ALBus服务总线中，接口被定义成一个基本业务单元，作为业务实现、发布和业务使用的基本单位。接口由接口名称（接口代码）唯一定义，接口可以拥有多个不同版本以包含相同业务的差异化实现。

在ALBus服务总线开发过程中，接口一般是ALBus服务总线服务端定义的一个方法签名，拥有确定的输入和输出类型。

在ALBus服务总线部署和运维中，接口一般包括服务端接口实现和客户端接口存根两个部分。客户端可以通过接口存根生成接口代理，实现代理模式的接口调用，也可以脱离接口存根通过接口名称和接口版本实现接口直接调用。

1. **交互**

又可以被称为业务调用、业务交互、接口调用、接口交互。

在ALBus服务总线中，业务客户端对业务接口发起一次业务调用并获得调用结果（调用成功获得输出或调用失败获得错误信息），被称为一次交互。

1. **前端**

在ALBus服务总线中，前端一般指业务接口的使用者。

1. **后端**

在ALBus服务总线中，后端一般指业务接口的实现者。

1. **总线**

在ALBus服务总线中，总线是前端与后端之间的一组（任意个）ALBus服务总线物理或逻辑节点组成的一个整体，负责接收前端输入数据，正确完成业务并返回业务输出到前端，其职责包括了接口发布和发现、路由、总线引擎、转发、业务容器管理、数据处理等部分。

### 3ALBus服务总线模式

作为一个典型的基于ALBUS的业务，主要包括以下模式

**开发模式**

1. 通过需求分析，确认业务服务接口设计，产出业务服务接口设计文档。
2. 依照ALBUS开发规范，基于前后端开发平台定义业务服务接口库，以前端为.Net，后端为J2EE平台为例，应产出接口定义一致的.Net接口定义Assembly和J2EE接口定义JAR。
3. 基于业务服务接口库，前端与后端开发团队分别实现业务逻辑开发，产出业务服务实现库。

**测试模式**

1. 使用业务服务接口库和总线服务封装模块部署业务模拟节点，作为前端黑盒测试平台，完成前端业务测试。
2. 使用业务服务接口库、业务接口实现库和总线服务封装模块部署业务服务节点，作为前后端联调测试平台
3. 按照生产环境网络拓扑部署UAT环境，其中部署完整的总线前端集群和总线后端集群。

**运行模式**

1. 前端发起业务请求
2. 总线接入节点接收业务请求，根据业务请求元数据和路由策略定位总线业务实现节点，如果需要的话转发业务请求。
3. 总线业务实现节点接收业务请求，解码业务请求数据并转发到业务单元容器。
4. 业务单元容器查询业务节点，使用业务数据完成业务请求并获取业务响应。
5. 总线获取业务响应并返回业务节点。

大致流程图如下：

ALBus服务总线组成

ALBus服务总线包括以下几个部分：

* ALBus开发框架
* ALBus测试框架
* ALBus设计规范
* ALBus开发和实现规范
* ALBus部署和运维规范

从而涵盖了复杂业务系统完整生命周期，针对模块化、规范化、敏捷化、层次化的业务系统设计、实现和运维中可能出现的问题和需求，提供了优秀、合理、完整的底层技术平台解决方案。

### ALBus服务总线作用

* 实现接口总线前端（接口使用者）与后端（接口实现者）之间稳定、高效、安全、可靠的数据和业务交互。
* 统一各类业务和接口实现之间差异，降低开发和运维成本。
* 协同接入总线的各业务子系统，实现完整业务流程，保证业务流程中处于不同角色的各子系统之间的正确交互。
* 降低业务系统中各子系统间的耦合，以接口总线为子系统间统一交互途径，从而减小系统实现复杂度和相关度，降低开发和运维成本。
* 封装节点、引擎、集群、路由等物理拓扑概念，从而解耦系统物理拓扑与业务逻辑拓扑，实现对业务需求完全透明系统需求变更（如业务量伸缩、业务规模扩展等，保证此类变更不影响现有业务和现有系统部署）。

### ALBus服务总线架构特点

#### 逻辑架构特点



* 接口封装和开放规范

ALBus服务总线为在总线上实现交互的所有接口定义了统一的接口声明和开放规范。凡遵循该规范定义的业务接口，可以被灵活的以总线支持的任意模式发布、实现、测试、部署和迁移。同时任何对总线功能的扩展均能被无缝应用到现有系统和现有业务。

* 数据处理

ALBus服务总线为接口调用和交互提供了灵活多样的数据处理手段，包括

* + 支持多种协议的序列化和反序列化适配。

当前提供了

* + - 二进制序列化/反序列化（基于Hessian协议）
    - JSON序列化/反序列化

Road Map计划中提供

* + - XML序列化/反序列化（基于WebService/SOAP协议）
    - 二进制序列化/反序列化（基于Google Ptotocol Buffer协议）
  + 智能适配各类序列化/反序列化模式的自定义数据类型处理机制
* 总线功能扩展和适配

ALBus服务总线系统所有功能均被设计为可扩展、可配置、可自定义的，且各个系统功能之间耦合度低，从而可以简单方便的实现对系统本身的调整和扩充。主要包括以下机制：

* + 可配置、可扩展的过滤器和过滤链机制，实现插拔式的接口交互中业务控制。
  + 可配置、可扩展的数据序列化/反序列化模式。
  + 可配置、可扩展的数据序列化/反序列化模式中的自定义数据类型处理器。
  + 可配置、可扩展的网络通讯协议适配。
  + 可配置、可扩展的服务节点引擎。

当前提供了

* + - 基于Embedded Jetty的内置引擎
    - 基于标准Servlet的Web应用引擎

Road Map计划中提供

* + - 基于netty的Socket引擎
    - 基于Apache的Apache Module引擎
    - 基于Microsoft IIS的IIS Filter引擎
  + 丰富的前端开发语言适配

ALBus服务总线后端（接口实现）支持Java平台开发实现。前端可以提供多种实现平台，当前包括：

* + - Java平台
    - JavaScript平台（基于AJAX技术）
    - .Net平台
    - Python平台

Road Map计划中提供

* + - PHP平台
* 安全和认证机制

通过可插拔的认证组件接口，ALBus服务总线实现了认证机制与接口平台解耦，亦即，业务系统实现的自定义认证机制可以通过扩展实现符合规范的认证适配，无缝插入接口总线，从而为所有接口使用者提供统一的认证和授权。且具体认证实现与接口总线透明无关，不影响接口总线的功能。当前实现了以下认证模式：

* + 基于用户名/密码的认证和授权机制
  + 基于密钥的认证和授权机制
* 业务功能的伸缩与扩展

ALBus服务总线通过业务单元容器解耦了业务实现和接口总线之间的依赖性，从而使被接口总线管理的业务系统实现可以灵活、无缝扩展和变更。当业务需求发生变化或扩展时，可以通过简单运维手段，在不影响接口总线配置和使用的前提下扩展已有业务子系统，也可以任意自由定义新的业务子系统插入当前接口总线中。

* 敏捷化的开发模式和规范

ALBus服务总线通过接口版本管理和智能适配，提供了相同、类似或相异业务更新和扩展的智能途径，为敏捷化的业务开发模式提供了简明合理的规范定义。遵循ALBus服务总线开发规范的同一接口多个不同版本可以被灵活的配置和使用。

#### 物理架构特点

* 网络通讯

ALBus服务总线提供了丰富、灵活的网络通讯模式，以适应不同业务场景对性能、可靠性、安全性、便利性（如防火墙通过性和开发难度）的要求。当前提供了：

* + 本地模式（无网络通讯，一般用于开发、测试和调试）
  + 基于HTTP消息（短连接）的通讯协议
  + 基于HTTP推送（长连接）的通讯协议

Road Map计划中提供

* + 基于Socket（TCP/IP）的通讯协议
  + 基于Message Queue的通讯协议
* 物理部署拓扑

ALBus服务总线提供了丰富、灵活的物理部署模式和拓扑类型，以适应不同业务场景对性能、可靠性、安全性、便利性（如防火墙通过性和开发难度）的要求。

当前ALBus服务总线提供了：

* + 本地模式（无网络通讯，前端后端在同一系统内部，一般用于开发、测试和调试）
  + 简单C/S拓扑模型



* + 基于负载均衡、节点同构、平坦拓扑的集群模型
  + 基于路由和转发的异构节点网络集群模型



* 性能

基于成熟的负载均衡技术和灵活的路由/转发策略，ALBus服务总线能够实现大规模集群部署，从而为接口总线提供了可靠的高性能处理能力。同时低耦合度的模块化设计和可插拔的运行模式，实现了总线系统单节点性能最大化。当前在PC测试环境、千兆网络环境和未做JVM调优的测试中，ALBus服务总线单个内嵌引擎可以提供简单业务400并发、2MB/s数据吞吐量。

* 稳定性

基于成熟的负载均衡和错误转移技术，ALBus服务总线集群可以简单实现极高稳定性：

* + 系统全天候不间断提供服务
  + 单次业务完整性和原子性保证
  + 业务内部错误转移
  + 完善的系统监控和异常情况通知
  + 智能自适应异常情况处理（例如：服务迁移等）
* 运维
  + 业务系统发布

ALBus服务总线以一个可嵌入的轻量级应用模式发布、部署与维护，可以简单的实现应用级系统管理运维，无需系统级管理。当前支持以下发布模式：

* + - 嵌入式发布：接口总线被嵌入到业务系统中。
    - 独立发布：接口总线包含业务系统，作为一个独立应用发布。
    - 应用容器发布：接口总线作为中间件应用，被发布到中间件应用服务器（如Tomcat、WebLogic等）中。
  + 业务监控

ALBus服务总线内核提供了完整的总线服务状态元数据监控途径，从而使业务监控可以作为一个简单的业务单元实现。在这种模式下，业务监控单元可以利用业务总线的所有特性实现诸如扩展性、灵活性、稳定性、可靠性等优点。

当前ALBus服务总线提供了基于业务服务粒度的简单监控，包括：

* + - 服务状态
    - 服务可用性
    - 服务用量
    - JVM资源监控
    - 报表输出

并可以通过扩展实现：

* + - 系统资源监控
    - 不同粒度、不同维度的计费
    - 自定义报表生成
    - 基于监控数据的智能自适应调整

## CC-Model数据建模技术

大数据主题建模又称大数据集约建模，使大数据独立于具体应用环境。

随着时间的推移，数据的各种职能会不断发生变化，但数据类型本身的变化并不经常发生，因此将数据与使用数据的职能分开是有重大意义的。数据的主题化的过程中，数据结构和存储方式都独立于应用数据的各种职能。数据按照人员、车辆、案件、通讯等业务科类，相互联系地存放在可以由众多用户共享的数据环境中。数据主题化需要进行全面的数据分析和数据模型建立工作，并及时生成数据搜索，随着该项工作的推进，可以大大加快应用项目的开发速度和加强用户与数据库的直接交互。

许多企业/政府单位已经开始进行数据主题化，但是遇到了一些问题。往往在开发一个新的应用项目时，由于某种原因，又为这个新的应用项目设计了一个新的数据库，而不是使用已经存在的或者正在建立主题数据库。通过匡信开发大数据主题建模软件，可以完全规避上述问题。通过使用简单的界面化操作即可完成数据主题模型的建立，数据主题模型的预览及其数据结构组织化。从而，数据不再分散，实现集中化管理，主题化管理。这里匡信开发的大数据主题建模软件还集成了主题数据搜索引擎的索引生成，便于用户完成信息检索。

通过开发匡信大数据主题库软件，将各类数据源字段整理统一，实现开发人员开发数据应用的高效操作。应用通过数据处理机制，将多单位不统一数据库资源，经过软件筛选聚合，整理出定义明确的大数据主题建模配置数据，实现数据按业务主题访问的统一，摒弃旧业务中需要因原数据导致多次配置操作。

* 主题数据功能
* 主题配置管理
* 主题使用表配置管理
* 主题查询参数配置管理
* 主题表查询参数配置管理
* 主题表字段配置管理
* 主题表字段内容配置管理
* 主题搜索配置管理
* 主题搜索字段配置管理
* 专题数据功能
* 专题配置管理
* 专题使用表配置管理
* 专题查询参数配置管理
* 专题表查询参数配置管理
* 专题表字段配置管理
* 专题表字段内容配置管理
* 专题搜索配置管理
* 专题搜索字段配置管理

## CC-Calc数据融合演算技术

### CC-ALCalc基础介绍

大数据处理时代，海量数据处理需要投入大量的程序员不仅要进行各类业务逻辑的开发，还需要熟悉各类计算框架，如MapReduce、Spark、Storm等，对程序员的要求较高；对于企业正常开发项目时，通常需要考虑快速迭代，通过本大数据熔合演算平台技术，能显著提升其他程序员代码产出效率，无需学习专业技术线路下的数据处理技术，可以将精力放在业务逻辑实现上。

通过一套大数据熔合演算平台，可以满足各类项目在所在计算框架中的计算需求，实现代码层面的快速切换、重用，降低大数据演算处理项目对开发人员能力要求，提升开发的效率，降低重复排错的成本。

### CC-ALCalc软件主要功能

匡信自主开发的大数据熔合演算平台，使得海量数据的离线或者实时计算可以自由地在各类大数据计算框架之间切换，并支持多数据源取数，通过拆解算法中的算子和因子，智能适配包括Spark、Spark Streaming、MapReduce、Storm等计算框架，调用大数据平台通用服务实现计算的功能。包括支持实时/批量计算、增量/存量计算、各类MapReduce计算功能。

该平台适配当前绝大部分的大数据平台，可以让开发成本显著降低，无需掌握各类大数据的所有开发技能就能实现在最短的时间内开发完成各类海量数据的业务处理实现，在最短时间内完成模型验证工作，并在将来将各类业务应用模型通过程序实现落地。

# 解决方案优势介绍

## 基于理论支撑

系统基于社会学、行为学、犯罪学三大理论支撑，理论实践个人的认知、行为与环境因素三者及其交互作用对人类行为的影响。海量模型经过迭代演算与理论相结合，筛选出更为精准的系统模型，将庞大的人口进行精准分类，精准提供服务与管理。

## 系统前瞻性

相较于现有的系统将全部重点放在重点人员及前科人员，缺乏了对潜在犯案人员的管控。系统整体具有前瞻性，提前对潜在可疑人员作出预判，达到预防作用，减少社会危害，指导警力科学合理分布。

## 系统实时性

系统运用hadoop、Spark，分布式数据库，实时数据仓库，可实现24小时不间断实时监控功能，完成人员异动趋势，行为轨迹的实时监测。

## 模型精准性

依赖大数据下的数据挖掘技术、实时数据仓库EDW、Spark等，从海量信息中发现提取影响人员社会安全的关键特征信息，在海量数据中迭代验证，结合社会学理论验算出精准可靠模型。