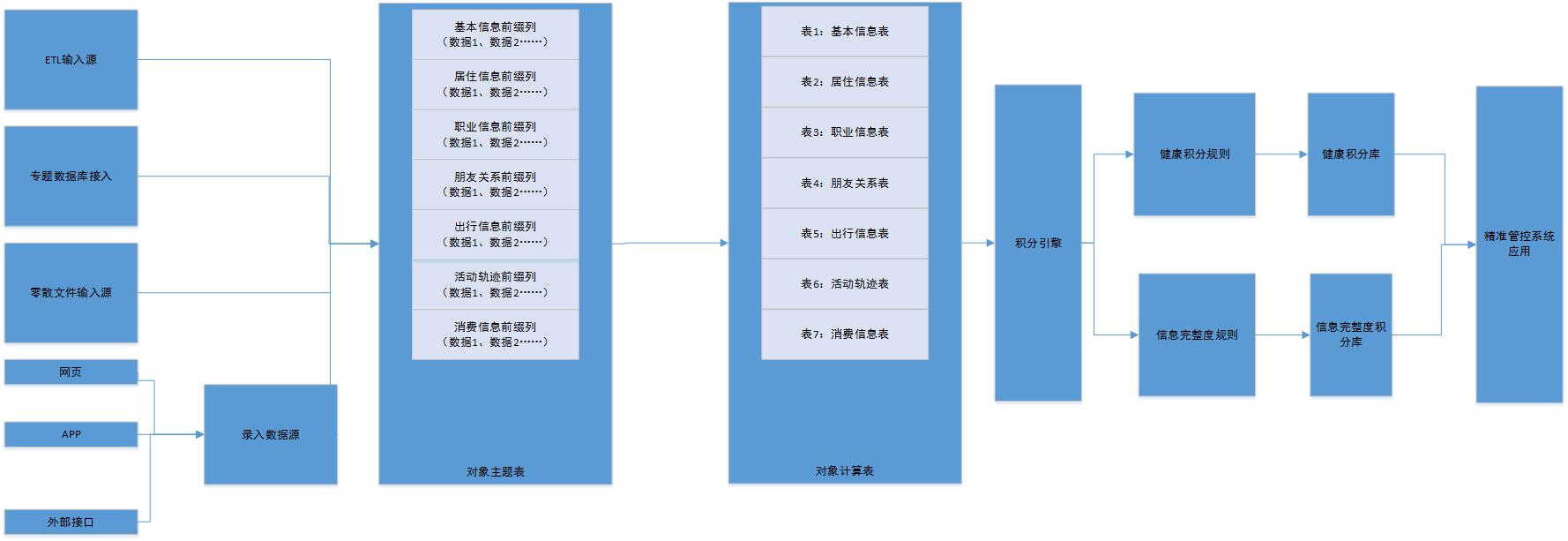
萧山人口精准管控系统整体设计方案

2017-09-12

# 文档版本

|  |  |  |  |
| --- | --- | --- | --- |
| 时间 | 人员 | 内容 | 版本 |
| 2017-09-12 | 黄跃东 | 创建此文档 | V0.5 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

# 系统逻辑图



# 数据的接入方式

## ETL数据接入

ETL数据接入方式，指的是从ETL的输出Mongo库（全量）和Kafka数据库（增量）中，接入数据资源，进入到对象主题表的某类列中。

ETL中的一类数据资源，可以加工进入主题表中的不同维度的列中。

如暂住数据，既可以加工进入主题表的居住信息维度列，也可以进入轨迹信息纬度列。

每个信息维度列中，具有的字段信息是明确的。

## 专题数据的接入

专题数据接入方式，指的是从专题数据存储的Mongo库，将数据资源，接入到对象主题表中的基本信息维度前缀列。

全量数据来自于全量mongo表，增量数据来自于mongo表中的时间字段。

## 零散数据的接入

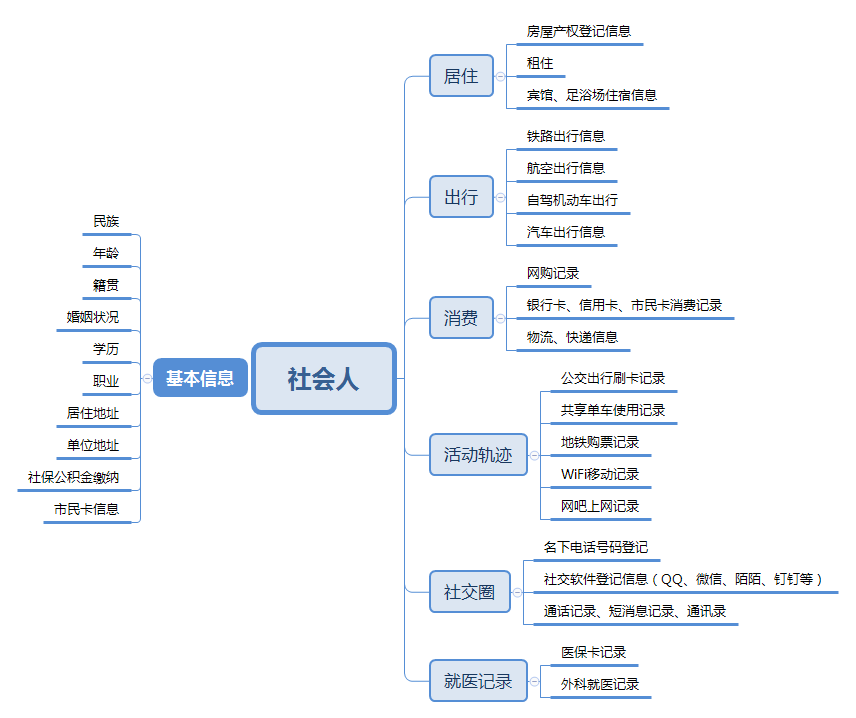
零散数据接入方式，指的是从零散文件上传的数据中，接入数据到对象主题表中指定维度前缀列中。

零散数据上传的时候，会保存到mysql（持久保存）和kafka中，通过配置，零散数据的字段分别加工存储到不同维度列中。

## 录入数据的接入

录入数据，是人工方式产生的各类数据。通过在mysql中，创建与维度名和字段信息一致的表和字段，实现录入数据与主题表中数据的同步。Mysql中表数据的支持增删改查，分别提供PC、APP、接口三种修改方式。

# 数据的主题库存储



## 数据主题库说明

数据的主题库存储，是将人、场所等客观实体对象，按照特定的安全维度，组织其相关的所有数据信息。比如，针对一个人员个体，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。

每一个社会安全属性，都有固定的字段内容。不同来源的数据，可以加工匹配进入不同的社会安全属性表中。

## 基本信息列的字段定义（后续可以继续添加）

基本信息列字段，主要有：

姓名

性别

民族

籍贯

婚姻状态

管辖单位

户籍地

文化程度

相关数据主要来源于专题库中数据。

## 居住信息列的字段定义（后续可以继续添加）

场所类型

场所ID

场所地址

居住类型

当前状态

开始居住时间

结束居住时间

管辖单位

相关数据来源与多种ETL、零散、手工输入等。

## 职业信息维度列族的字段定义（后续可以继续添加）

单位类型

单位名称

岗位类型

岗位职级

当前状态

开始时间

结束时间

管辖单位

相关数据来源与多种ETL、零散、手工输入等。

## 出行信息维度列族的字段定义（后续可以继续添加）

出行方式

出行起点

出行目的地

开始时间

结束时间

管辖单位

相关数据来源与多种ETL、零散、手工输入等。

## 轨迹信息维度列族的字段定义（后续可以继续添加）

轨迹类型

轨迹时间

轨迹地点

管辖单位

管辖单位

相关数据来源与多种ETL、零散、手工输入等。

## 消费信息维度列族的字段定义（后续可以继续添加）

消费类型

消费地点

消费时间

消费金额

管辖单位

相关数据来源与多种ETL、零散、手工输入等。

# 数据的演算库存储

## 数据演算库说明

数据演算库，是将数据主题库中，同一前缀的数据列，转换为一张表。

数据演算库，支持多表之间的关联查询，支持数据的更新。

## 基本信息表的定义

由主题表的基本信息列字段数据转换而来。

## 居住信息表的定义

由主题表的基本信息列字段数据转换而来。

## 职业信息表的定义

由主题表的职业信息列字段数据转换而来。

## 出行信息表的定义

由主题表的出行信息列字段数据转换而来。

## 轨迹信息表的定义

由主题表的轨迹信息列字段数据转换而来。

## 消费信息表的定义

由主题表的消费信息列字段数据转换而来。

# 积分引擎

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

# 人口健康积分

人口健康积分，就是基于积分引擎，配置的计算人员健康程度的一套积分，人口积分采用定期计算+特定目标对象分数计算结合的两种方式。

# 信息完整度积分

人口健康积分，就是基于积分引擎，配置的计算人员信息完整度的一套积分，信息完整度积分采用定期计算+特定目标对象分数计算结合的两种方式。