**行为感知系统设计概述**

目录

[架构设计 2](#_Toc512523886)

[数据流导向设计（此图已失效） 2](#_Toc512523887)

[数据库设计 2](#_Toc512523888)

[基础信息库 2](#_Toc512523889)

[应用识别库 4](#_Toc512523890)

[流水数据库 6](#_Toc512523891)

[行为报警数据库 6](#_Toc512523892)

[计量数据库 7](#_Toc512523893)

[统计分析 7](#_Toc512523894)

[实时监控库 7](#_Toc512523895)

[模型数据库 8](#_Toc512523896)

[数据模式设计 9](#_Toc512523897)

[nDPI格式协议 9](#_Toc512523898)

[文件格式协议 10](#_Toc512523899)

[统一格式协议 10](#_Toc512523900)

[训练数据格式协议 10](#_Toc512523901)

## 架构设计

## 数据流导向设计（此图已失效）



## 数据库设计

Elasticsearch：索引dataflow。SQL功能

Mysql：库名 information，用户名 root，密码:admin

Redis：

### 基础信息库

#### 员工信息表

|  |  |  |  |
| --- | --- | --- | --- |
| **表名： employee\_information\_basic** | | | |
| **描述：员工信息表** | | | |
| **存储类型：MySQL** | | | |
| **字段** | **字段名** | **描述** | **类型** |
| ID | id | 主键。用户唯一编号,自增 |  |
| 工号 | number | 员工的工号 | varchar |
| 姓名 | name | 员工姓名 | varchar |
| 入职时间 | hiredate |  | datetime |
| 离职时间 | leavedate |  | datetime |
| 状态 | status | 1实习，2试用，3正式，4离职 | char(2) |
| 工龄 | seniority |  | decimal(3,1) |
| 岗龄 | post\_age |  | decimal(3,1) |
| 部门ID | department\_id | 外键，关联部门表ID。 | varchar |
| 岗位 | post | 用户所在岗位 | varchar |
| 职称 | positional\_titles |  | varchar |
| 创建时间 | create\_at |  | datetime |
| 角色ID | role\_id | 外键，关联角色表 | varchar |

#### 部门信息表

|  |  |  |  |
| --- | --- | --- | --- |
| **表名：department\_information\_basic** | | | |
| **描述：部门信息表** | | | |
| **存储类型： Mysql** | | | |
| **字段** | **字段名** | **描述** | **类型** |
| ID | department\_id | 部门ID，主键唯一编号，自增 |  |
| 部门名称 | department\_name | 部门名称 | varchar |
| 部门负责人 | department\_header |  | varchar |
| 创建时间 | create\_at |  | datetime |

#### 设备表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **表名：device\_basic** | | | | |
| **描述：设备表** | | | | |
| **存储类型： Mysql** | | | | |
| **字段** | **字段名** | | **描述** | **类型** |
| ID | id | | 主键，设备唯一编号，自增 |  |
| 设备编号 | number | |  | varchar |
| 员工ID | employee\_id | | 外键，关联员工表 | varchar |
| 设备名称 | name | | 自定义 | varchar |
| 设备种类 | type | | 用户设备类型（1:PC，2:移动设备，3:其它） | char(2) |
| 设备mac | | mac | 用户设备的mac编号 | varchar |
| 设备性质 | | property | 枚举：（1:办公，2:私有） | char(2) |
| 静态IP | | static\_ip | 浮动IP时为空 | varchar |
| 创建时间 | | create\_at |  | datetime |
| 设备状态 | | status | 枚举：使用中，未使用 | char(2) |

#### 组织结构表

|  |  |  |  |
| --- | --- | --- | --- |
| **表名：role\_basic** | | | |
| **描述：角色表** | | | |
| **存储类型： Mysql** | | | |
| **字段名** | **字段** | **描述** | **类型** |
| 角色ID | id | 主键，自增 |  |
| 角色名 | role\_name |  | varchar |
| 下级角色ID | lower\_role\_id |  | varchar |
| 上级角色ID | higher\_role\_id |  | varchar |

### 应用识别库

#### 应用识别表

|  |  |  |  |
| --- | --- | --- | --- |
| **表名： application\_identify** | | | |
| **描述：应用识别表（内置表）** | | | |
| **存储类型： Mysql** | | | |
| **字段** | **字段名** | **描述** | **类型** |
| ID | id | 主键 标识应用唯一id,自增 |  |
| 应用编号 | number |  | varchar |
| 应用名称 | name | 应用的名称  ***app.应用属性名*** | varchar |
| 应用协议 | protocol | 应用所使用协议 | varchar |
| 应用域名 | domain\_name | 应用域名 | varchar |
| 应用端口 | port | 应用服务端口 | varchar |
| 官方地址 | official\_address | 应用的官网地址 **app.官网地址** | varchar |
| 所属领域 | domain | 应用按照功能划分所属于的领域 | varchar |
| 应用类别 | category | 应用归类 ***type.关联属性*** | varchar |
| 应用标签集合 | tag\_set | 应用标签集合（内置）（关联**标签表**Id） | varchar |
| 应用描述 | description | 应用描述  **app.属性描述** | varchar |
| 版本编号 | version\_number |  | varchar |

#### 网站识别表

|  |  |  |  |
| --- | --- | --- | --- |
| **表名：web\_identify** | | | |
| **描述：网站识别表（内置表）** | | | |
| **存储类型： Mysql** | | | |
| **字段** | **字段名** | **描述** | **类型** |
| **id** | **id** | 主键. 标识网站站点id,自增 |  |
| 网站编号 | number |  | **varchar** |
| 网站名称 | name | 网站名称 **attribution.属性名** | **varchar** |
| 类别 | category | 网站的类别归类 ***type.关联属性*** | **varchar** |
| 网站协议 | protocol | 网站所使用协议 | **varchar** |
| 网站域名 | domain\_name | 网站域名 | **varchar** |
| 官方地址 | official\_address | 官方地址 **attribution.属性名** | **varchar** |
| 所属领域 | domain | 网站按照功能划分所属于的领域 | **varchar** |
| 关联应用 | association\_application | 关联 **应用识别表.id** | **varchar** |
| 站点标签集合 | site\_tag\_set | 网站标签集合（内置）（关联**标签表**Id） | **varchar** |
| 网站描述 | description | 网站介绍。**attribution.描述** | **varchar** |
| 版本编号 | version\_number |  | **varchar** |

#### 基础标签表

|  |  |  |  |
| --- | --- | --- | --- |
| **表名： tag\_basic** | | | |
| **描述：基础标签表** | | | |
| **存储类型： Mysql** | | | |
| **字段** | **字段名** | **描述** | **类型** |
| id | id | 主键。标识标签唯一id,自增 |  |
| 标签编号 | number |  | varchar |
| 标签名 | name | 标签的名字 | varchar |
| 标签描述 | description | 标签描述 | varchar |

#### 标签属性表

|  |  |  |  |
| --- | --- | --- | --- |
| **表名： tag\_property** | | | |
| **描述：标签属性表** | | | |
| **存储类型： Mysql** | | | |
| **字段** | **字段名** | **描述** | **类型** |
| 标签名 | tag\_name | 标签的名字 | varchar |
| 标签属性 | property | 0：工作不相关 1：工作相关 2：不确定 | varchar |

#### 标签扩展表

|  |  |  |  |
| --- | --- | --- | --- |
| **表名： tag\_extention** | | | |
| **描述： 标签扩展表** | | | |
| **存储类型： Mysql** | | | |
| **字段** | **字段名** | **描述** | **类型** |
| id | id | 主键。标识Id |  |
| 扩展编号 | number |  | varchar |
| 标签集 | tag\_set | 应用标签集合（内置）（关联**标签表**Id） | varchar |
| 扩展集名 | extend\_set\_name | 扩展的标签名字 | varchar |
| 描述 | description | 描述 | varchar |
| 岗位相关性 | post\_relevance | 枚举：积极性（0:消极，1:积极，2:中性） | char(2) |
| 是否启用 | is\_enabled | 枚举：（0:未启用，1:启用） | char(2) |

### 流水数据库

#### 流水数据表

|  |  |  |  |
| --- | --- | --- | --- |
| **表名： dataflow** | | | |
| **描述： 流水数据表** | | | |
| **存储类型： ElasticSearch** | | | |
| **字段名** |  | **描述** | **类型** |
| 流水Id |  | 标识记录的唯一编号 |  |
| 设备编号 | deviceID | 标识访问设备 |  |
| 源IP | sourceIP |  |  |
| 应用名称 | appName |  |  |
| 应用类别 | appType |  |  |
| 应用协议 | appProtocol |  |  |
| 目的IP | destinationIP |  |  |
| 开始时间 | startTime |  |  |
| 上传流量 | upstreamFlow |  |  |
| 下载流量 | downstreamFlow |  |  |
| 详情 | details | 未知信息 |  |
| 结束时间 | endTime |  |  |
| 域名 | host |  |  |
| 是否内网 | isIntranet | 判断是否是内网 | char |
| 应用标签集合 | appTagCollection | 应用基础标签集合（内置）（关联基础**标签表**Id） |  |

### 行为报警数据库

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **表名： behavior\_alarm** | | | | |
| **描述： 行为报警数据表** | | | | |
| **存储类型： Mysql** | | | | |
| **字段** | **字段名** | **描述** | **类型** | |
| 流水Id | id | 标识记录的唯一编号 |  | |
| 设备编号 | number | 标识访问设备 | varchar | |
| 源IP | source\_ip |  | varchar | |
| 应用名称 | application\_name |  | varchar | |
| 应用类别 | application\_ category |  | varchar | |
| 应用协议 | application\_protocol |  | varchar | |
| 目的IP | target\_ip |  | | varchar |
| 访问时间 | access\_time |  | | datetime |
| 上传流量 | upload\_flow | 本机向远端服务器发送的流量大小 | | varchar |
| 下载流量 | download\_flow | 远端服务器向本机发送的流量大小 | | varchar |
| 详情 | detail | 未知信息 | | varchar |
| 警报类型 | type | 枚举。｛0:泄密，1:病毒入侵，2:暴力请求｝ | | char(2) |
| 警报级别 | level | 枚举（缓急，紧急，火急） | | char(2) |
| 是否已处理 | is\_processed | 枚举（0:未处理，1:已处理） | | char(2) |
| 创建时间 | create\_at |  | | datetime |

### 计量数据库

### 模型数据库

*注：系统需要预制一个核心参数用于存放模型的文件路径*

|  |  |  |
| --- | --- | --- |
| **表名： 模型数据表**  **存储类型： Mysql** | | |
| **字段名** | **描述** | **类型** |
| 模型编号 |  |  |
| 模型名称 |  |  |
| 模型描述 |  |  |
| 准确率 |  |  |
| 召回率 |  |  |
| 训练样本分布 |  |  |
| 创建时间 |  |  |
| 版本号 |  |  |

### 参数配置表

|  |  |  |  |
| --- | --- | --- | --- |
| **表名： param\_config** | | | |
| **描述： 参数配置表** | | | |
| **存储类型： Mysql** | | | |
| **字段** | **字段名** | **描述** | **类型** |
| 参数名 | param\_name |  | varchar |
| 参数值 | param\_value |  | varchar |
| 参数单位 | param\_unit |  | varchar |

### 全球国家基础信息表

|  |  |  |  |
| --- | --- | --- | --- |
| **表名： country\_basic** | | | |
| **描述： 国家基础信息表（全球235个国家）** | | | |
| **存储类型： Mysql** | | | |
| **字段** | **字段名** | **描述** | **类型** |
| ID | id | 主键 | int(11) |
| 国名 | name | 国家名 | varchar(50) |
| 简称 | nick\_name | 国家（英文）简称 | varchar(20) |
| 首都位置 | capital\_geo | 首都的经纬度信息 | varchar(30) |

### 全球国家IP网段表

|  |  |  |  |
| --- | --- | --- | --- |
| **表名： global\_ip\_country\_map** | | | |
| **描述： 全球国家IP网段表（和country\_basic表有n：1关系）** | | | |
| **存储类型： Mysql** | | | |
| **字段** | **字段名** | **描述** | **类型** |
| ID | id | 主键 | int(11) |
| 起始IP | start | 网段的起始IP | varchar(15) |
| 结束IP | end | 网段的结束IP | varchar(15) |
| 国家ID | country\_id | country\_basic表的外键 | int(11) |

## 数据模式设计

### nDPI格式协议

|  |  |
| --- | --- |
| 字段名 | 字段注释 |
| ipPorto2Name | 传输层协议 |
| srcIp | 源IP |
| srcPort | 源端口 |
| dstIp | 目的IP |
| dstPort | 目的端口 |
| proto | 应用层协议 |
| srcPktNums | 源数据包个数 |
| srcBytes | 源数据流量大小 |
| dstPktNums | 目的数据包个数 |
| dstBytes | 目的数据量大小 |
| vlan | 是否是vlan |
| host | Host |
| client | SSH/SSL客户端 |
| server | SSH/SSL服务端 |
| btHash | 比特流hash值 |
| info | 其他信息 |
| timestamp | 时间戳 |
| httpURL | httpURL |
| httpMethod | httpMethod |
| srcMac | 源mac地址 |
| dstMac | 目的mac地址 |

### 文件格式协议

### 统一格式协议

|  |  |
| --- | --- |
| **格式协议： protobuf** | |
| 设备编号 | 标识访问设备 |
| 源IP |  |
| 应用名称 |  |
| 应用类别 |  |
| 应用协议 |  |
| 目的IP |  |
| 访问时间 |  |
| 上行流量 |  |
| 下行流量 |  |
| 详情 | 未知信息 |
| 应用标签集合 | 应用标签集合（内置）（关联**标签表**Id） |

### 训练数据格式协议

**草稿：**

**用户信息库（Relation Database）：**

用户ID，用户名，部门，岗位名称，终端设备id，设备种类，设备mac，设备性质（私有还是办公），固定IP

**用户上网行为流水库（NoSql）：**

设备id，源IP，应用类型（浏览器，APP），协议类型，行为类别，目的IP，目的端口，URL，访问时间，上行流量，下行流量，详情

**URL属性库：**

从行业角度： （行业相关性）正向相关，负向相关，中性

从网页自身内容： （聚类或半监督）体育，购物，媒体，新闻，IT，娱乐，社交，音乐…

从工作角度： 效率，泄密…

自定义扩展：

**APP应用分类库：**

应用属性： 购物，炒股，支付，出行，聊天…

**异常访问库：**

设备，ip，访问地址，时间，备注·

**技术选型 ：**

数据采集：nDPI

数据采集暂存队列：kafka

数据处理：spark

深度学习库：TensorFlow(内含numpy)

主要开发语言： scala/java，python

数据存储： Mysql，HBase

**URL多维度标签划分：**

**URL 与岗位的相关性判定划分： 正性，负性，中性**

**给定一个url，通过如下计算 url与岗位的相关性：**

**正性：**

**置信度：** confidence（url −>X岗位）

**对于给定的url，该url最相关的岗位**

**设在url的样本空间为R(r1,r2,r3…)，岗位的空间为P(p1,p2,p3…)，每个岗位p对应一个员工组成的员工空间G(g1,g2,g2…)。 从岗位的角度来看，**

**Max(**p1,rg **) ,** p2,rg **) ,** p3,rg **) … )**

**负性： 访问的地址与所在岗位在所有岗位中具有最小共现性。**