数据分析及架构汇总

2017-8-6 王昊尘

# 平台现有数据概述

## OEM数据统计（乘用车）

最新统计：共计27个汽车品牌，194685条OE(供货数据)

|  |  |
| --- | --- |
| **车辆品牌** | **OEM数量** |
| 一汽奥迪 | 3610 |
| 奥迪 | 19 |
| 一汽丰田 | 5345 |
| 上汽大众 | 812 |
| 上汽通用 | 417 |
| 广汽丰田 | 1486 |
| 东风日产 | 663 |
| 东风本田 | 210 |
| 广汽本田 | 353 |
| 沃尔沃 | 122 |
| 路虎 | 149 |
| 北京奔驰 | 25 |
| 梅赛德斯奔驰 | 10025 |
| 华晨宝马 | 25 |
| 宝马 | 8995 |
| 保时捷 | 6392 |
| 一汽大众 | 1368 |
| 北汽幻速 | 4831 |
| 海南马自达 | 26471 |
| 开瑞 | 20618 |
| 力帆 | 21421 |
| 郑州海马 | 6697 |
| 一汽马自达 | 1405 |
| 一汽奔腾 | 307 |
| 郑州日产 | 6 |
| 长安铃木 | 14771 |
| 奇瑞 | 58142 |
| 总计 | 194685 |

经浙江联保OE清洗和匹配系统处理后，得出下列匹配结果（由于车享配和优闪订是本周刚提交的数据，因此匹配结果未统计在下列表格中）

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **品牌** | **OEM总数** | **匹配** | **未匹配** | **匹配率** | **未匹配率** |
| 宝马 | 8984 | 8093 | 891 | 90.08% | 9.92% |
| 梅赛德斯奔驰 | 10006 | 8979 | 1027 | 89.74% | 10.26% |
| 东风本田 | 16 | 10 | 6 | 62.50% | 37.50% |
| 一汽大众 | 1326 | 789 | 537 | 59.50% | 40.50% |
| 一汽奔腾 | 307 | 174 | 133 | 56.68% | 43.32% |
| 保时捷 | 6388 | 3579 | 2809 | 56.03% | 43.97% |
| 一汽马自达 | 1405 | 624 | 781 | 44.41% | 55.59% |
| 海南马自达 | 26471 | 11344 | 15127 | 42.85% | 57.15% |
| 一汽奥迪 | 3541 | 1386 | 2155 | 39.14% | 60.86% |
| 长安铃木 | 14771 | 5466 | 9305 | 37.00% | 63.00% |
| 天津一汽 | 3370 | 1242 | 2128 | 36.85% | 63.15% |
| 广汽本田 | 52 | 19 | 33 | 36.54% | 63.46% |
| 奇瑞 | 58142 | 19804 | 38338 | 34.06% | 65.94% |
| 北汽幻速 | 4831 | 1409 | 3422 | 29.17% | 70.83% |
| 郑州海马 | 6697 | 1843 | 4854 | 27.52% | 72.48% |
| 丰田 | 12 | 3 | 9 | 25.00% | 75.00% |
| 力帆(乘用车) | 21421 | 5264 | 16157 | 24.57% | 75.43% |
| 开瑞 | 20616 | 3380 | 17236 | 16.40% | 83.60% |
| 东风日产 | 62 | 0 | 62 | 0.00% | 100.00% |
| 郑州日产 | 6 | 0 | 6 | 0.00% | 100.00% |

匹配部分包括两部分：1.清洗后供货OE与联保OE一致，2.原始供货OE与联保OE不一致，经其预处理(包括去除OE中空格，横线，和纠正配件名称)后的清洗数据与联保OE一致。不匹配部分为清洗后供货数据在联保数据库中找不到对应OE。

## 车型别名库

按品牌、车系、车型三个层级存储车辆信息。由于没有VIN解析的功能，且供应商提供的车辆信息种类比较多，平台目前尚无具体的车型定义规则。在已有的供货数据中，多以汽车具体年款或区间，辅以车辆配置来区分不同车型。例如：

|  |  |  |
| --- | --- | --- |
| **品牌** | **车系** | **车型** |
| 东风日产 | 骐达(C12) | 2011-2015 |
| 上汽大众 | 朗逸 | 朗逸15款 |
| 上汽大众 | 新速腾 | 三厢1.6手动挡 |

理赔系统中的信息比较详细，会以年款+配置的方式记录出险的车辆，例如：

|  |  |  |
| --- | --- | --- |
| **品牌** | **车系** | **车型** |
| 上海大众 | 波罗 | 2014 波罗 两厢1.4L 手动档 风尚版 |
| 东风日产 | 骐达 | 2009 骐达 两厢1.6L 手动档 时尚型 |
| 长安福特 | 福克斯 | 2015 长安福特 福克斯 两厢 1.6L 双离合变速器 舒适型 |

粗略统计，已有的供货数据共含**216**个不同的车型。因没有标准车型定义规则，故未进行精确筛选。

|  |  |
| --- | --- |
| 汽车品牌 | 车型数量 |
| 一汽丰田 | 5 |
| 一汽大众 | 80 |
| 一汽奔腾 | 1 |
| 一汽奥迪 | 5 |
| 一汽马自达 | 2 |
| 上汽大众 | 40 |
| 上汽通用 | 19 |
| 东风日产 | 8 |
| 东风本田 | 7 |
| 丰田 | 2 |
| 保时捷 | 2 |
| 北京奔驰 | 2 |
| 华晨宝马 | 2 |
| 奥迪 | 1 |
| 宝马 | 8 |
| 广汽本田 | 15 |
| 梅赛德斯奔驰 | 8 |
| 沃尔沃 | 3 |
| 路虎 | 3 |
| 郑州日产 | 2 |

# OE唯一性校验

定损单

原始上货数据

（按模板填写）

明觉

清洗系统

（车型、OE等）

车型别名库

配件别名库

清洗OE

清洗车型

筛选系统

（价格、品质）

基础数据

车型信息

VIN解析

配件信息

数据商或

其他渠道

匹配OE

导入/积累

导入/积累

配件信息

（已匹配OE及车型）

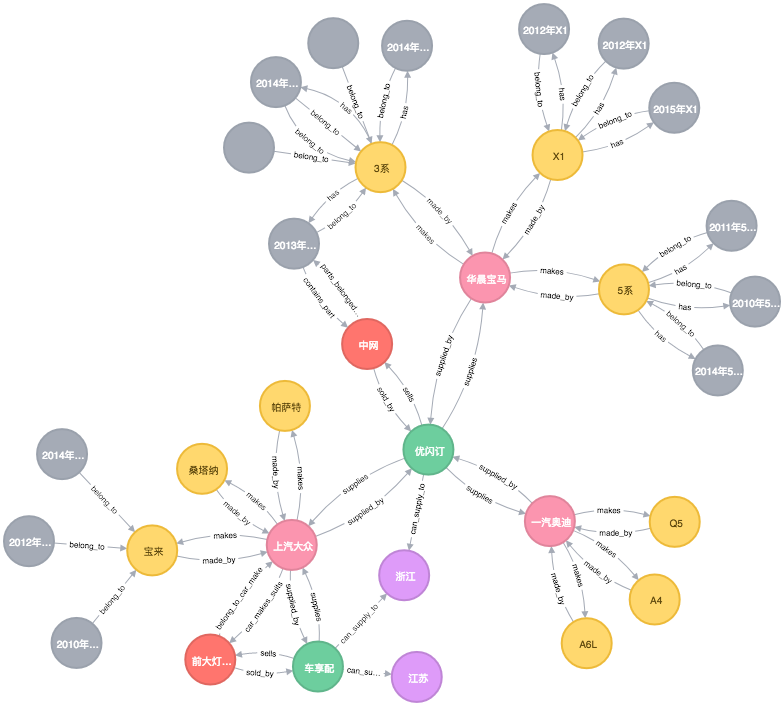
查找配件价格

返回报价

# 基础数据Relational database结构介绍

# 基础数据Graph database结构介绍

基础供货数据图建模



假设某节点（或列）有n条相关联的数据，而整体为m步（或行），查找某条数据的时间复杂度如下：

关系型数据库：O(mn)——需遍历整个表格

图数据库：采用索引图，O(mlogn)——相邻节点查找时间为logn

采用原生图，O(logn)——免索引整体遍历时间为O(1)

优势：语句简单，结构灵活，查找速度较关系型数据库有大幅度提升

缺陷：对于程序技术架构依赖性强，存在未知性