



车300
che300.com



车300数据处理技术实践

车300研发部 章水鑫

2018/04/14

目录

- 车300简介
- 车300数据类型与规模
- 车300数据处理架构
- 车300数据处理监控
- 车300数据安全



车300简介

车300是独立第三方车况与车价信息查询服务提供商，为广大买车、卖车用户在车辆交易中提供全方位的车辆价格评估、车辆维修保养记录、车型识别判定和车辆保值率分析等在线信息查询服务，帮助买卖双方规避交易中的风险，买车卖车先查车300！

车300简介

2014.3
车300网站
上线

2014.7
获五岳资本
天使投资

2015.6
获源码资本
A轮投资

2015.10
获红杉资本
A+轮投资

2016.12
获产业资本
联合投资

2017.8
伽马资产管
理系统上线

2014.6
赶集网数据
合作

2014.12
车300 APP
上线

2015.10
合作伙伴突
破300家

2016.9
安吉租赁等
战略合作

2017.4
快定价服务
17万+

2018.3
服务客户超
过30亿次

车300数据类型与规模

3 亿

超过3亿条车源价格数据

200 万

200万日活产生的
行为数据

数据清洗架构

微服务集群

车型识别

车源去重

虚假数据识别

...

清洗节点

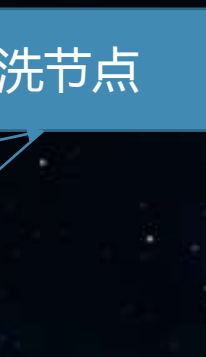
清洗节点

清洗节点

消息队列

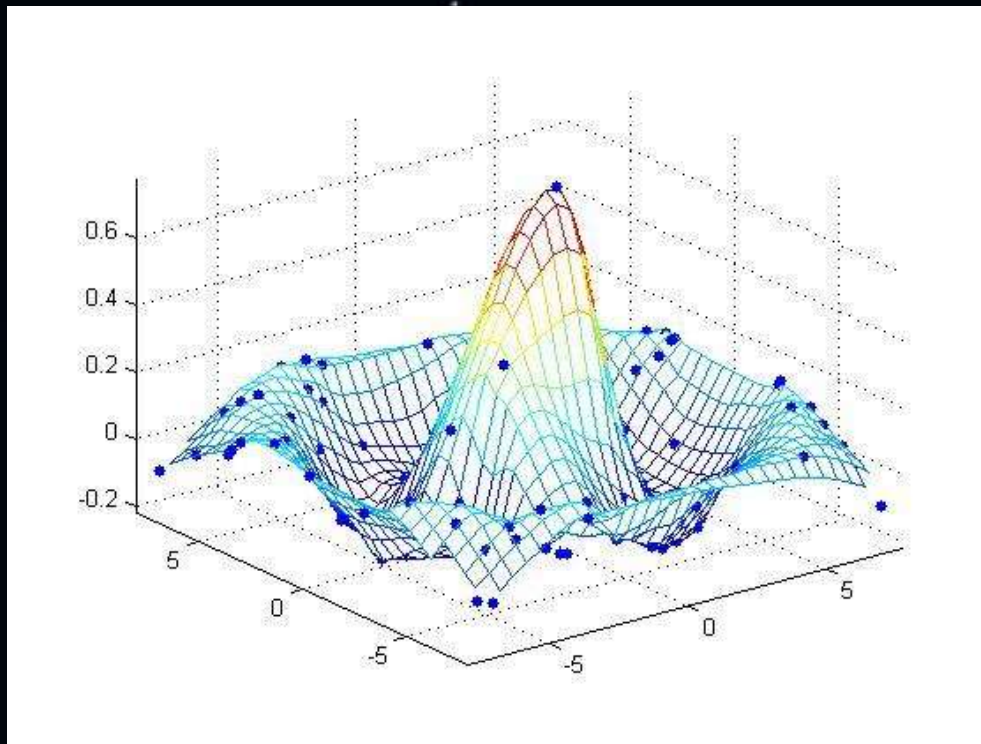
车源数据库

原始
车源数据

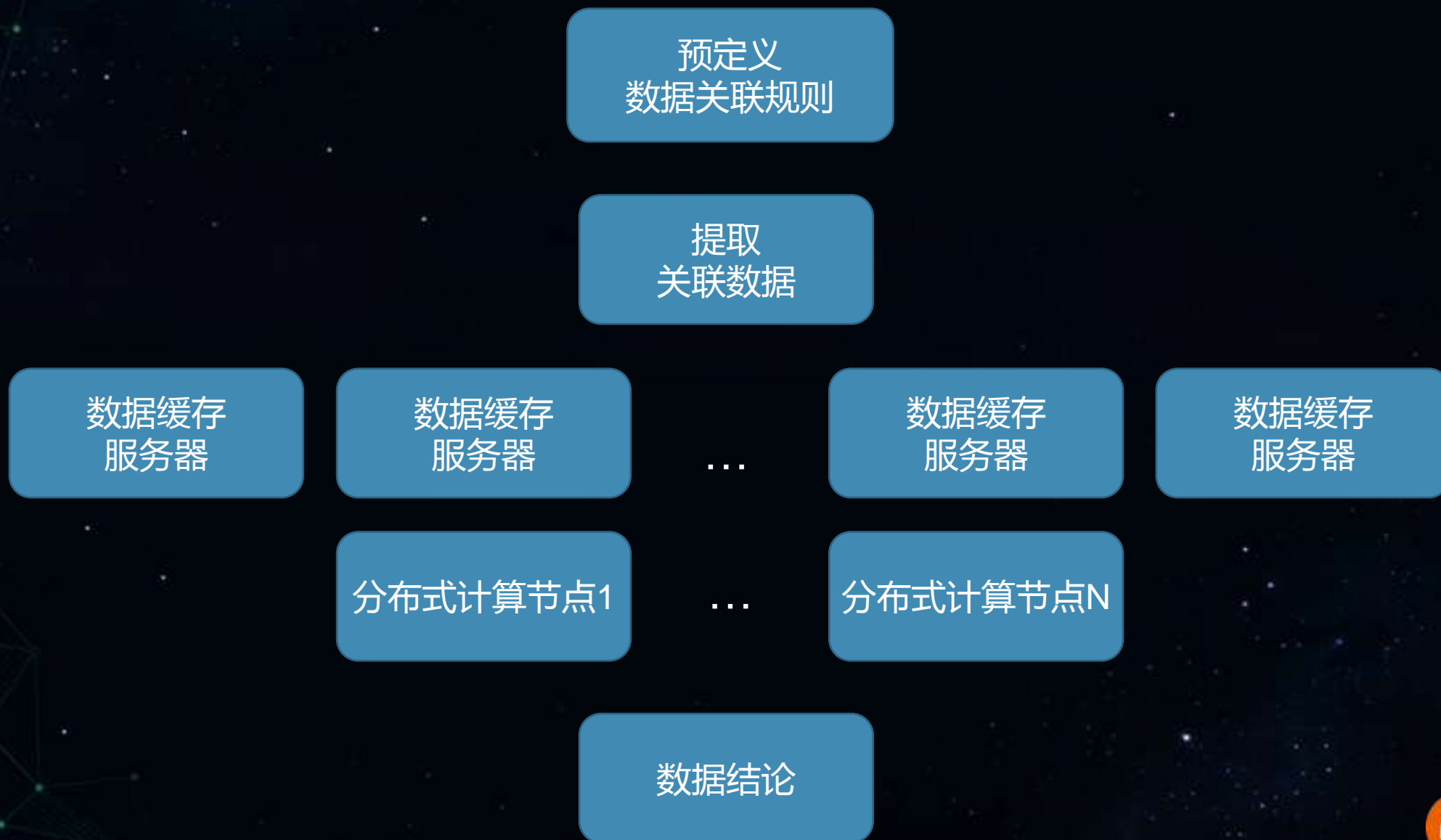


数据处理中的规模与挑战

- **车300数据模型**
 - 经纬网络模型
- **超过3亿条车源数据**
 - 多维关系网络模型
 - 每条数据都是关系网络中的一个节点
- **节点数据**
 - 可能会缺失
 - 可能存在异常数据
 - 可能存在关系错位



车源数据处理架构

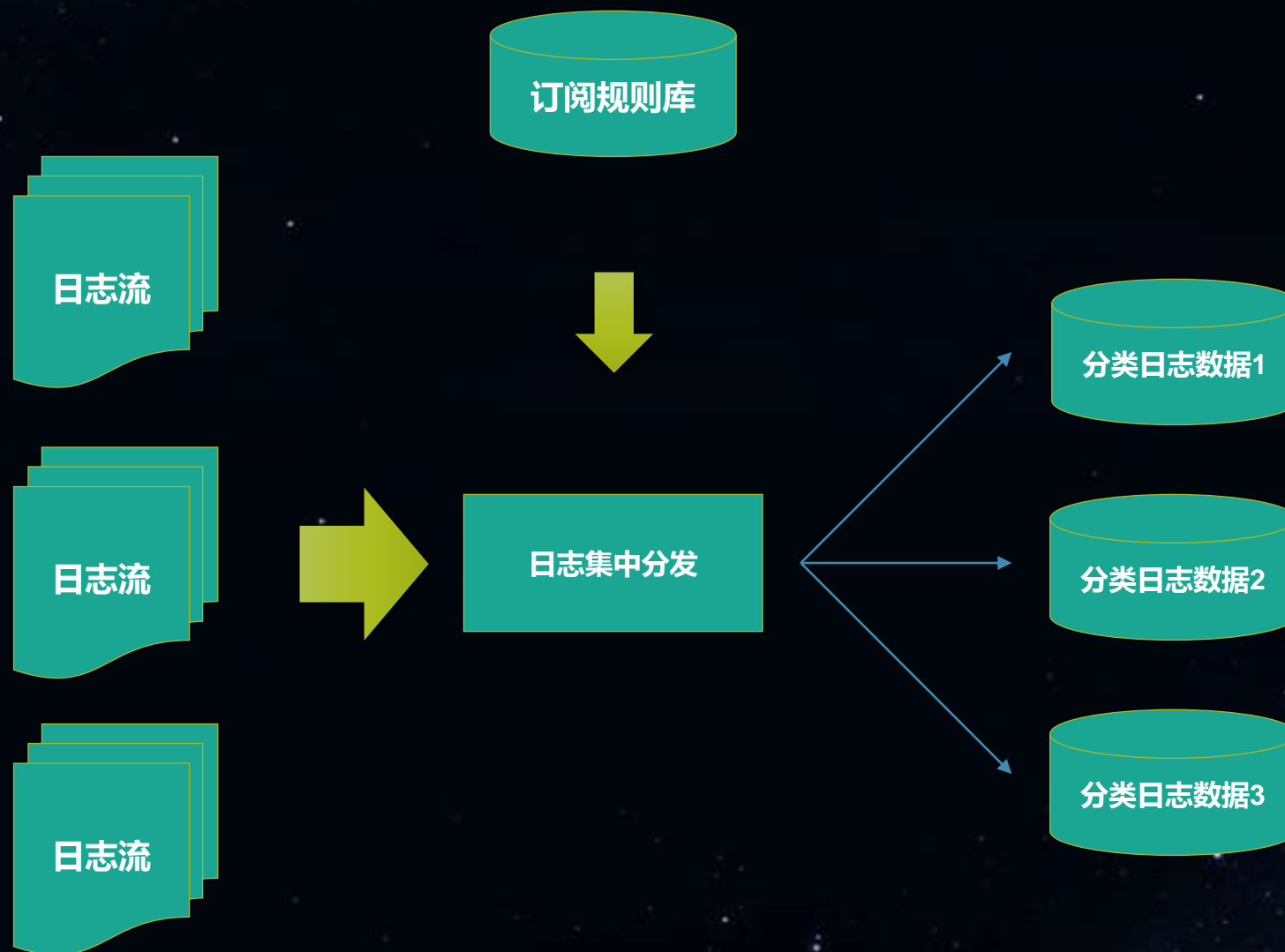


用户日志数据处理架构

海量日志数据处理逻辑

- 数据存储方式
 - 订阅模式数据处理分类，分布式存储
 - 自动智能标签索引
- 数据处理模块建立订阅规则后在指定数据节点提取数据
 - 车型/车龄/车价/位置/行为路径/行为时间/模块停留时间/输入或筛选行为
 - 外部信息标签

用户日志数据处理架构



用户身份标签



- 聚类
- K-近邻
- SVM
- 关联规则

END
T H A N K S