

REDEFINE FUTURE

阿里云天池千里马获奖企业沟通沙龙

Zebrai
斑马数智

基于阿里云架构的车联网应用实践

架构师 洛彬



2018云栖大会·深圳峰会
THE COMPUTING CONFERENCE

阿里云 | 奥运会全球指定云服务商



1

公司介绍

2

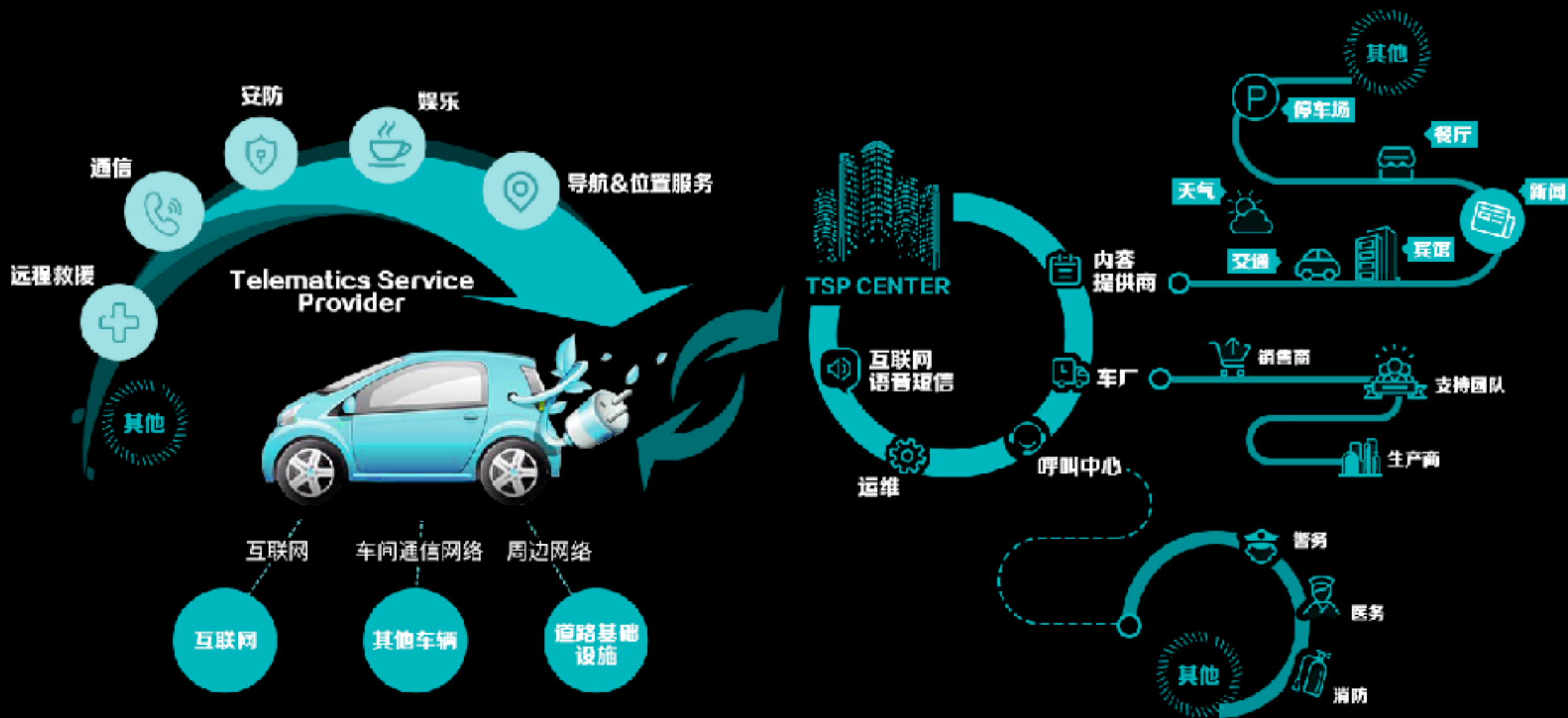
产品

3

技术实践

公司介绍

从海量车联网数据中帮助客户挖掘人车场景互联的价值





应用场景

车联网有助车厂直接获取用户用车数据，借助大数据能力提升更精准的产品营销

促进
精准营销

满足用户
个性化需求

车联网是开放平台服务体系，可以整合多种应用资源，满足用户个性化的应用场景需求，提升用户驾车体验

通过用户驾驶过程中的行为习惯做定制化的内饰和动力匹配，为新车的设计和生產提供数据支持

反向定制
柔性生产

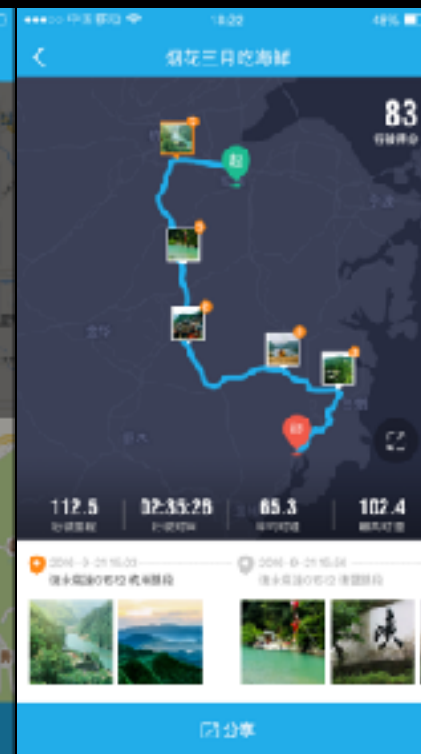
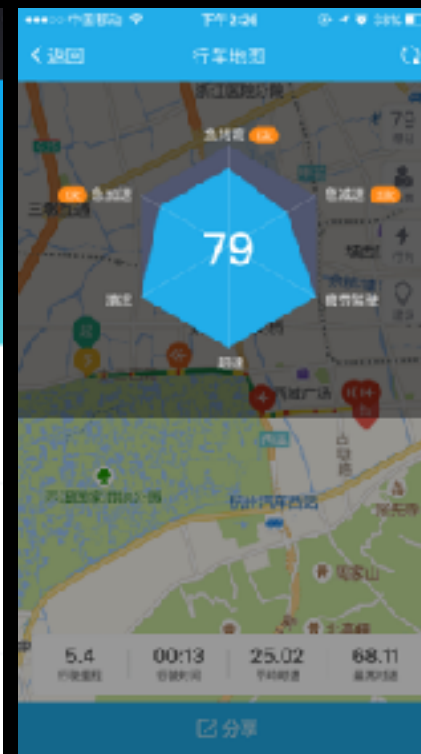
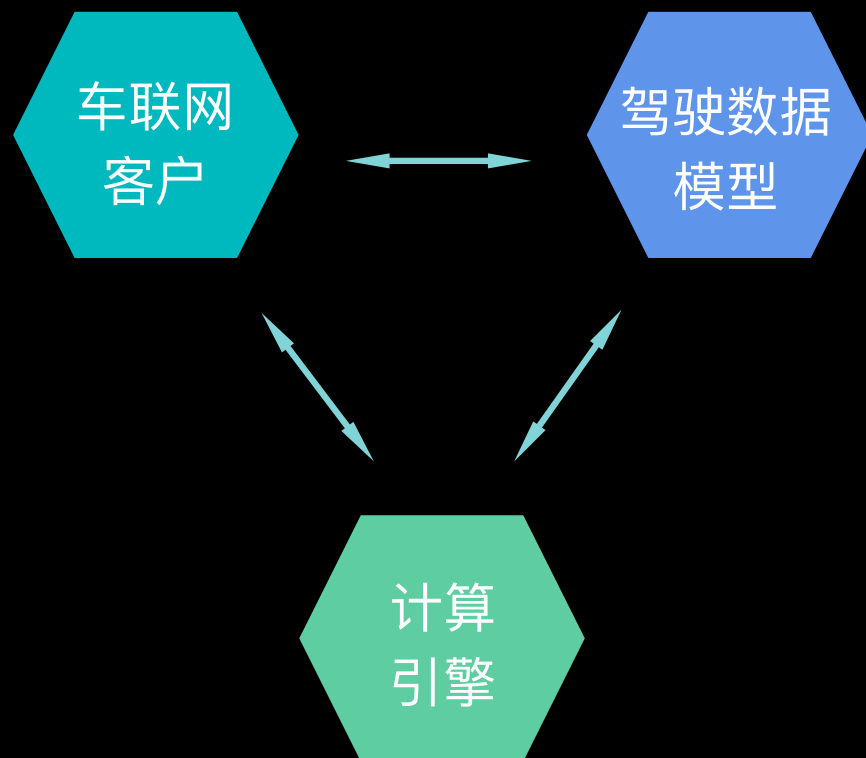
促进
客服服务

促进经销商
网络管理

车联网会增加车主、车厂、经销商直接接触，有助车厂提升对经销商网络的管理与服务创新能力

车联网产品

行车基础产品



传统车离我们整车厂很远；未来车辆实时在线、车辆中台实时数据服务、实时智能推送、场景车、场景、用户数据挖掘思考化消费体验、线上线下一体化和人车世界

车生活
听歌、FM
随我所想



AI
人工智能
通信服务



智慧化服务
智能推送
搜索信息



场景化消费
查地图
语音导航



智能化
信息服务
天气、新闻



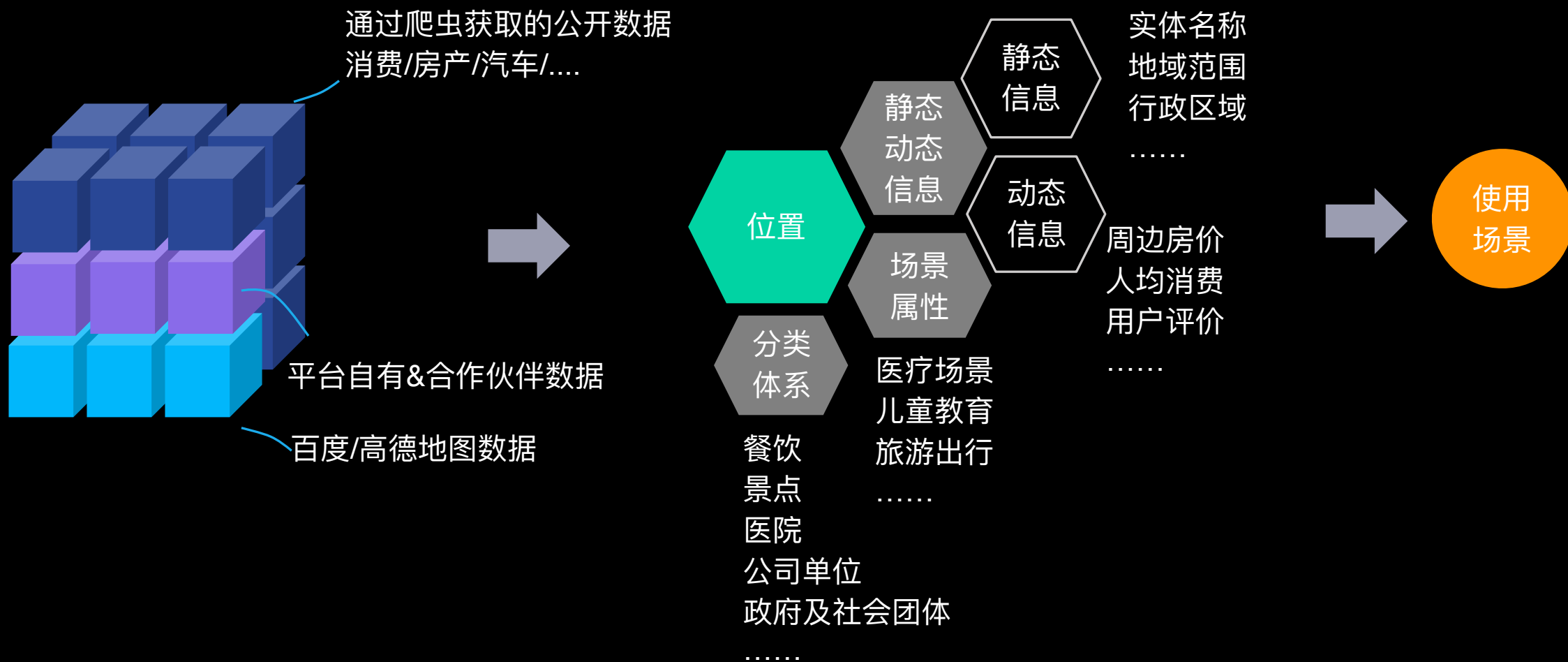
社交化
发微信
查邮件



技术驱动构建基础数据

洞察位置背后的信息，先知地而后知人

斑马“位置”领域模型

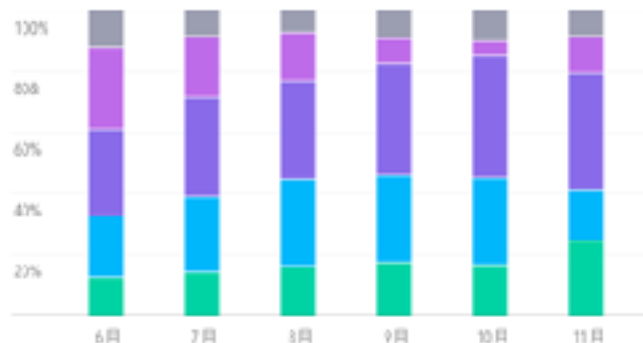


斑马综合指数人群分布



800-900	10.8%
700-799	21.6%
600-699	34.9%
500-599	26.7%
400-499	10.7%

斑马综合指数人群分布趋势



近7日榜单

榜单榜单榜单榜单

榜单榜单	榜单榜单
榜单榜单	榜单榜单
榜单榜单	榜单榜单
榜单榜单	榜单榜单
榜单榜单	榜单榜单
榜单榜单	榜单榜单
榜单榜单	榜单榜单
榜单榜单	榜单榜单
榜单榜单	榜单榜单
榜单榜单	榜单榜单

榜单榜单



用户画像热力图



人群属性标签分布

全部 800-900 700-799 600-599 500-499 400-399



属性	1.0%
81-100分	24.0%
61-80分	11.9%
41-60分	10.7%
21-40分	10.7%
0-20分	10.7%



81-100分	24.4%
61-80分	11.9%
41-60分	10.7%
21-40分	10.7%
0-20分	10.7%



81-100分	24.4%
61-80分	11.9%
41-60分	10.7%
21-40分	10.7%
0-20分	10.7%

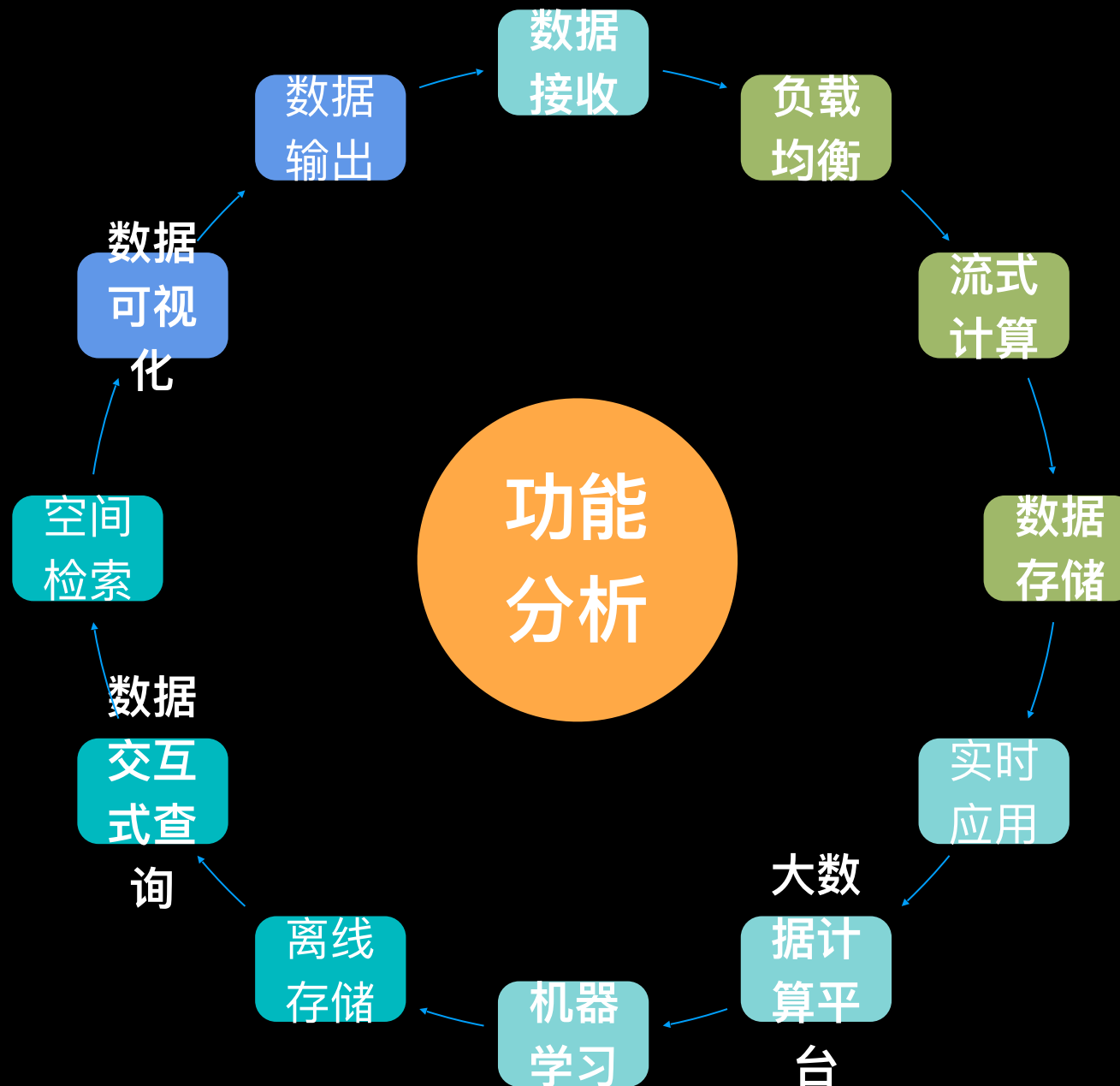


81-100分	24.4%
61-80分	11.9%
41-60分	10.7%
21-40分	10.7%
0-20分	10.7%



81-100分	24.4%
61-80分	11.9%
41-60分	10.7%
21-40分	10.7%
0-20分	10.7%

技术实践





PGSql
Impala Spark TCP
SparkStreaming
Tomcat Azkaban Flink
FTP Flume
SpringBoot Kafka Hdfs Kudu
Hbase Mysql
Netty MR Sqoop Gobblin
LVS Hue Hive Parquet ElasticSearch
HTTP Yarn Redis
Nginx MachineLearning

需要引入大框架
和中间件

运维成本极高

开发投入成本高

现阶段在业务的
快速落地和迭代

稳定性可能比较
差

人才需求极大



云产品



解决方案
Solution



APP



险公司
insurers



汽车出行
fleets



整车厂
Car manufacture



车载硬件商/
After market installments

Gateway

Input

SLB-TCP

SLB-HTTP

接收服务
器ECS

接收服务
器ECS

接收服务
器ECS

接收服务
器ECS

Output

回调服务
ECS

H5服务
ECS

API服务
ECS

CDN

分布式MQ

Kafka

Kafka

Kafka

核心服务

GIS服务

UI智
服务

行程产品
服务

位智
服务

行程报告推送
服务

标签
服务

大数据平台

DataV

QuickBI

Data
Works

MaxCompute

流计算

Spark
Streaming

Spark
Streaming

Spark
Streaming

数据存储

PgSQL

Mysql

Hbase

Redis

HybridDB

分布式离线存储

Oss对象存储

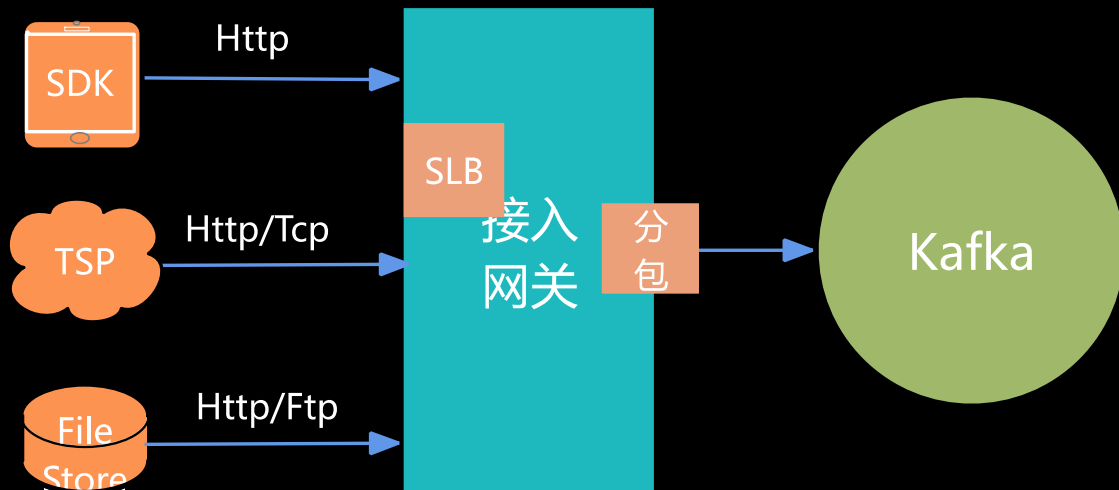
Oss归档存储



数据上传接收

多种通信协议支持,HTTP\TCP\File

高性能、高可用、高吞吐量、低延时MQ



- 基于顺序磁盘写O1级性能数据的写入
- 基于内存缓存的索引支持
- 多分区模式的支持
- 支持主备模式
- 支持各种主流流计算框架
- 低成本的扩容

数据压缩方案

列式

列式压缩可以把具有相同特征的数据聚在一起，选择最优的数据压缩和处理方式

数据Int化

数据使用Int类型的格式，可以最大化压缩数据的字节数

前缀提取

将gps数据和传感器数据中有大量相同数据前缀，比如gps经纬度数据前4位基本不变，可以大幅度压缩数据大小

混合编码

根据数据不同的特性，如波动性变化小，采用差值编码；大量数据连续，采用RLE编码；大量数据重复，采用字典编码；数据的最大值不大，采用BitPacked编码

边界值处理

对gps和sensor数据，数字都是在一定范围内波动，在采用差值编码后，存在极值像0转变，需要特殊处理

★ 100行数据，gps数据，数据大小： %20CSV； 50%Protobuf

流计算

Spark Streaming

社区的活跃度和
成熟度极高

公司的Spark和
Scala技术栈

Checkpoint机制
的支持

对滑动窗口的很
好支持

微批处理可很好
的支持车联网相
关的数据

无缝的支持
KafkaMq和各种
DB



在线应用存储

PGSQL

- 高效的R索引可以很好的支持GIS相关的应用
- GIN索引的支持，可以对基于标签的存储和大文本存储建立全文检索
- 高效的支持大数据领域的特征查询、标签存储检索
- 使用简单，单机性能卓越，成本低
- 阿里云提供一体的数据备份，可视化管理以及稳定的服务

Hbase

- 基于LSMT架构的NOSQL，可以很好的对KV进行按序的存储
- 非常适合作为多维的指标中心的存储
- 强大的分布式架构加上阿里云强大的自动化运维支持
- 支持流式计算和离线计算和应用计算的对接

HybridDB

- 基于列式存储的OLAP数据
- 基于MPP架构的高效的查询性能
- 可支持交互式的大数据可视化展示

大数据平台选型

基于Hadoop生态圈

开发投入成本
高

运维成本
高

稳定性
比较差

易用性
差

机器成本需
固定投入

计算量
规模化后成本
降低

VS

基于ODPS

只需ETL
开发

无运维成本

稳定性好

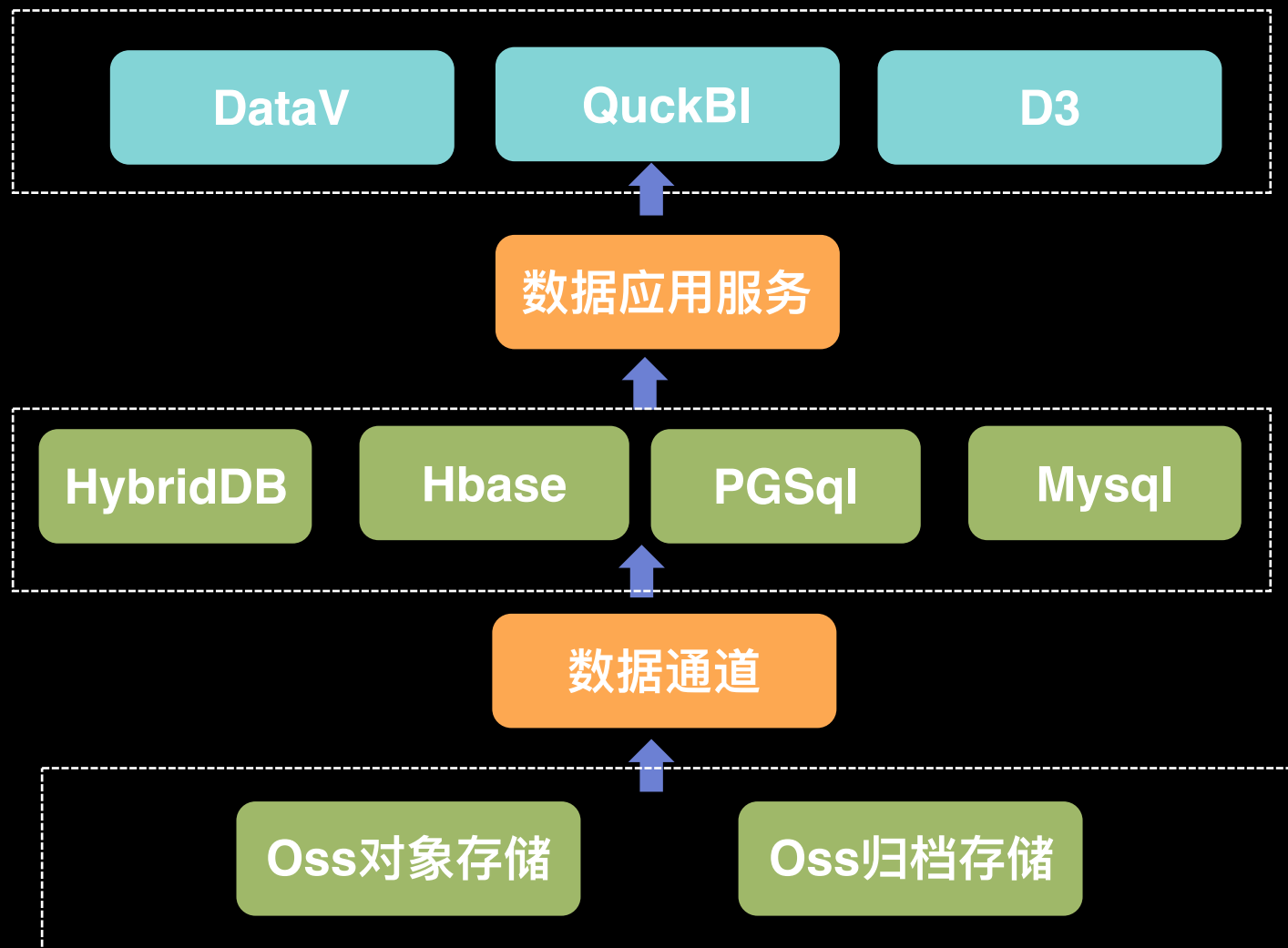
可视化
开发

按计算资源
动态收费

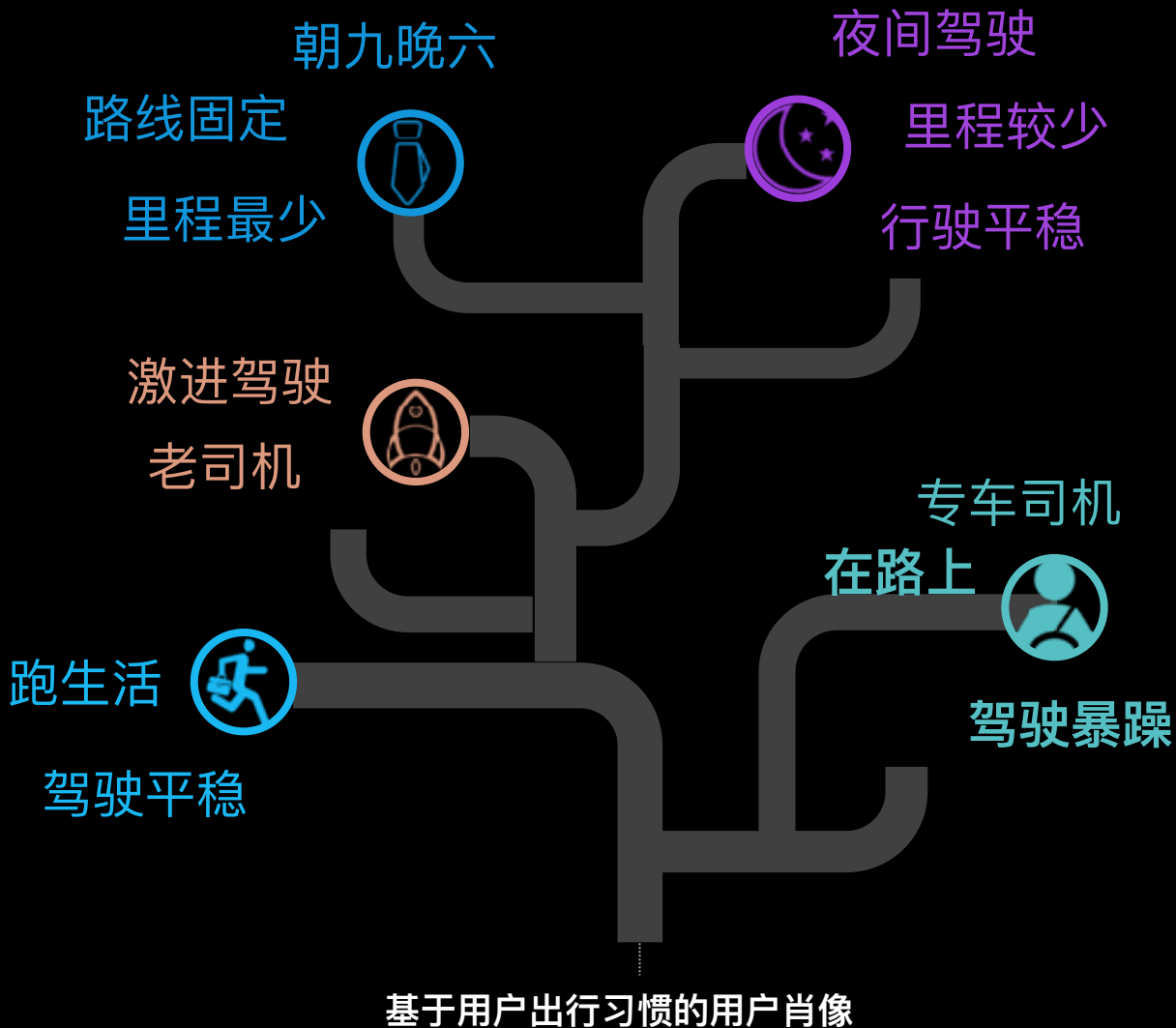
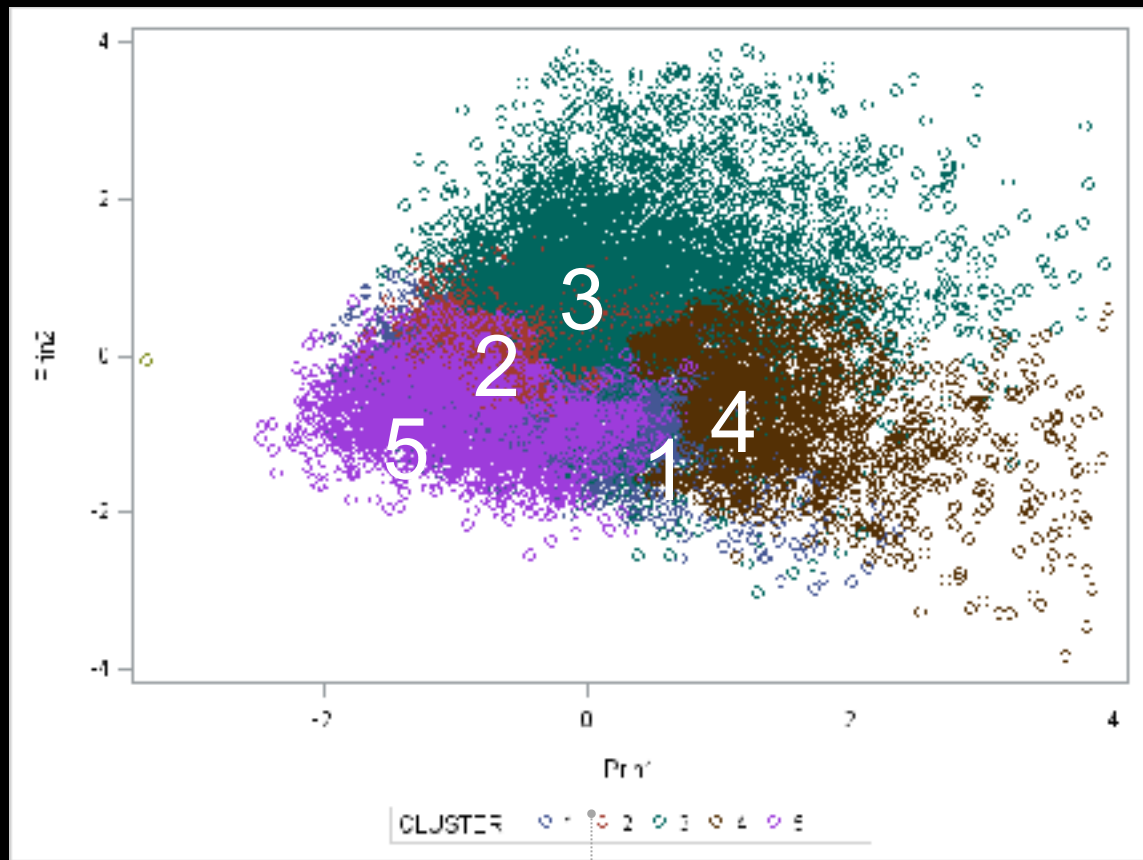
规模化
后计算成本很
高

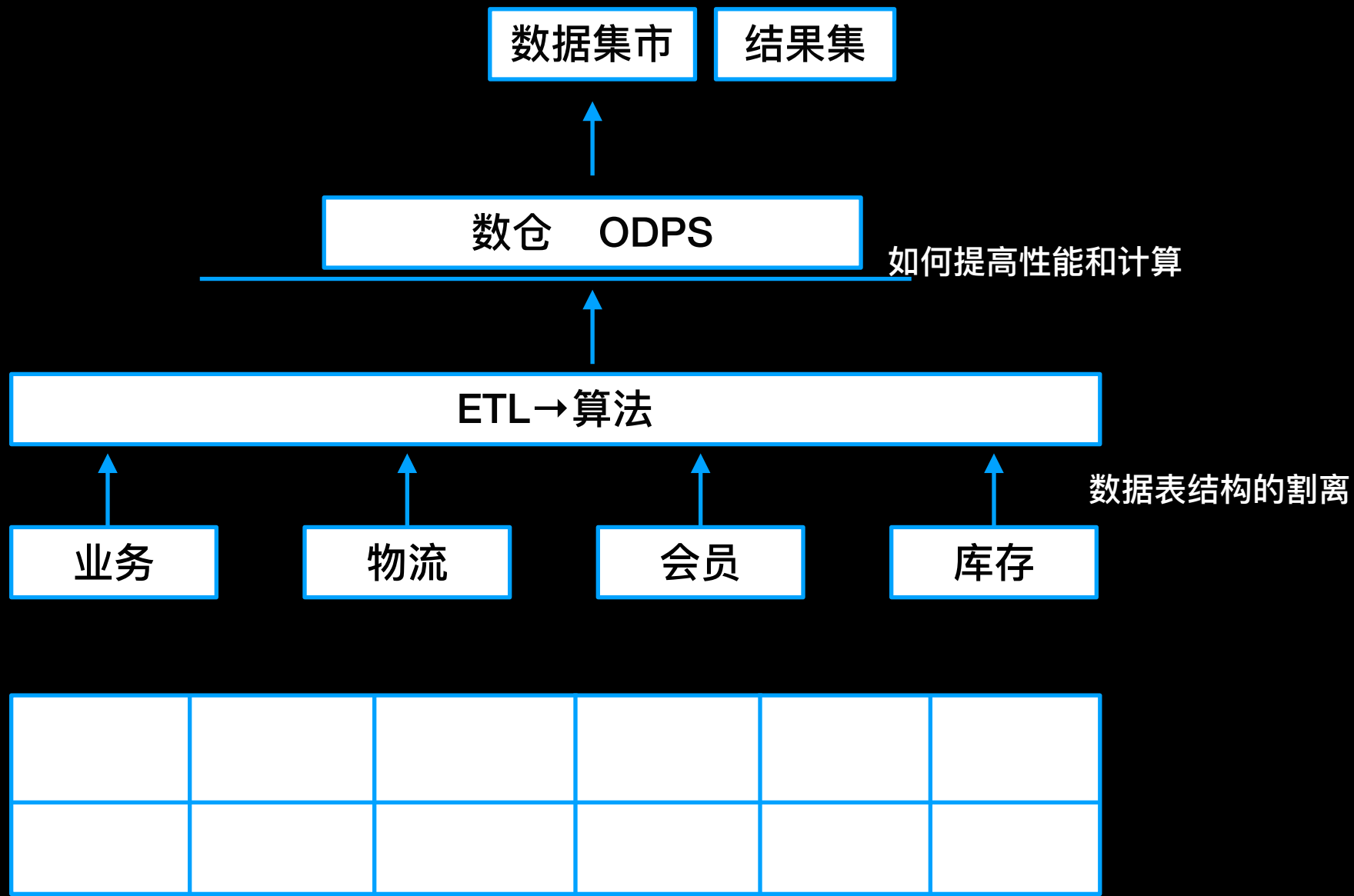
大数据存储及可视化

- 数据仓库
- 数据集市
- 原始归档数据
- GIS地理数据
- 基于Geohash地理快的统计数据
- 标签数据



基于聚类算法的人群挖掘







拿到客户企业的原数据表结构

数据库之间的关联

用户需求

业务理解想法

需求文档

分项

指标

指标定义

指标单位

数据来源

下钻

数据结构有没有？

数据结构如何处理

需求拆解：1~2个周

数据治理：6个周

表结构、数据维度 PRD（200页）

数据源→接口→计算处理



2018云栖大会·深圳峰会
THE COMPUTING CONFERENCE



2018云栖大会·深圳峰会
THE COMPUTING CONFERENCE