

OSDF 2017 开源数据库论坛(北京)  
OPEN-SOURCE DATABASE FORUM(BEIJING)

# 开源数据库正在改变世界

2017年8月24日-25日 北京-京仪大酒店





# PostgreSQL时空之门

阿里云 digoal

# 目录

## CONTENTS

背景介绍

时空业务需求与挑战

PostgreSQL时空特性

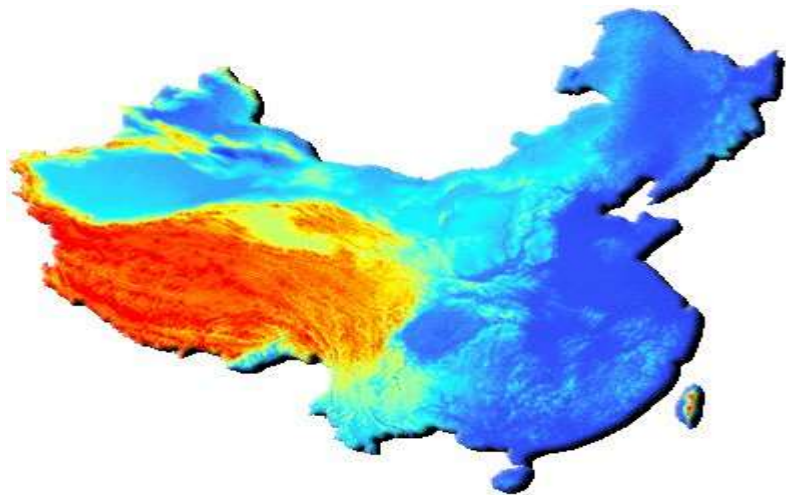
案例

展望



# 背景介绍

# 时空数据天天见

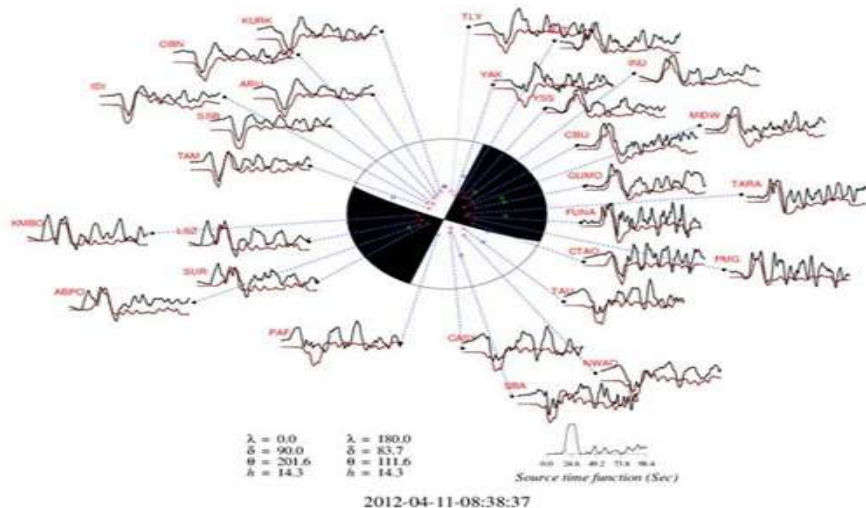


气象数据：

raster :

经纬，时间，  
温度、湿度、风速、风向、风力、  
日照、雨量、。。。。。

# 时空数据天天见

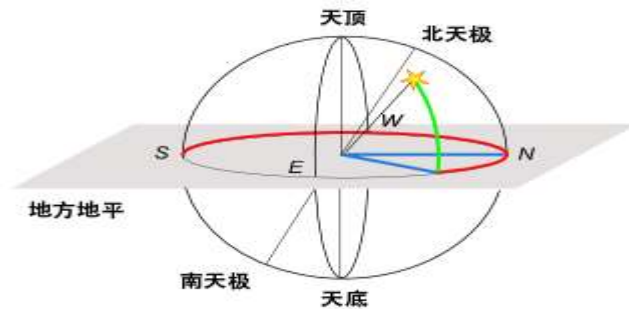


地震数据:

raster:

经纬, 时间, 高频波(50次/s)

# 时空数据天天见



天文数据：

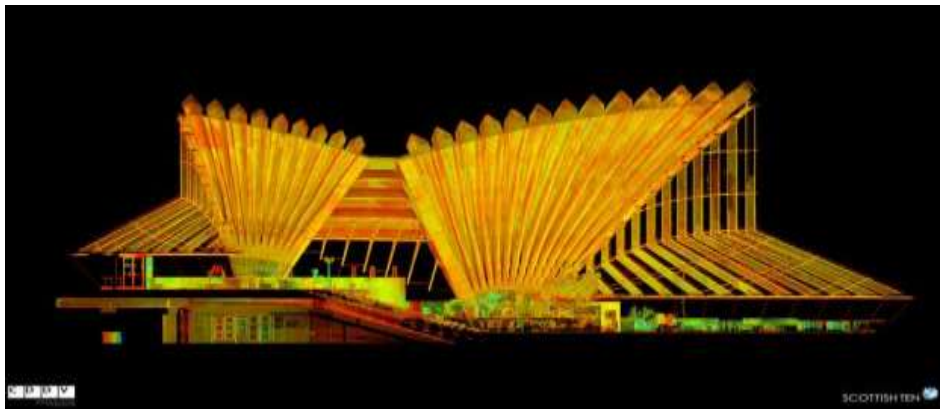
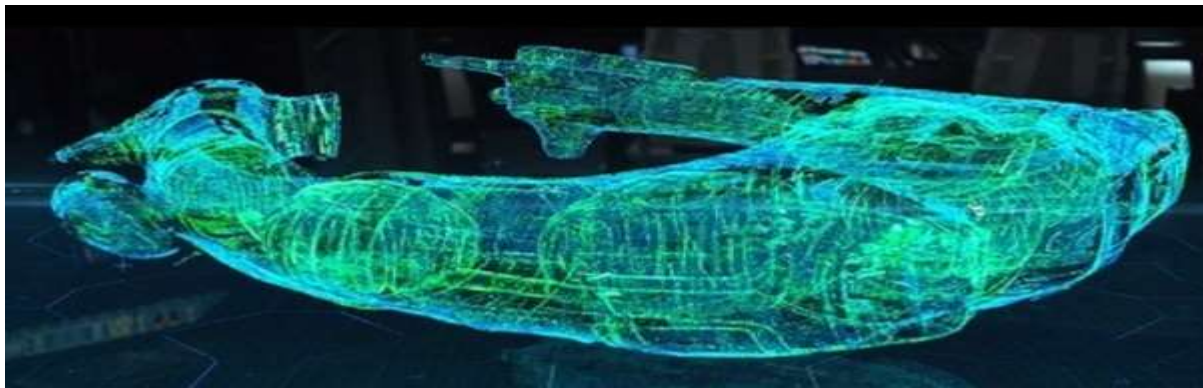
raster:

天球坐标，时间，image, 其他属性





# 时空数据天天见



点云:

x,y,z,RGB (optional k:v)



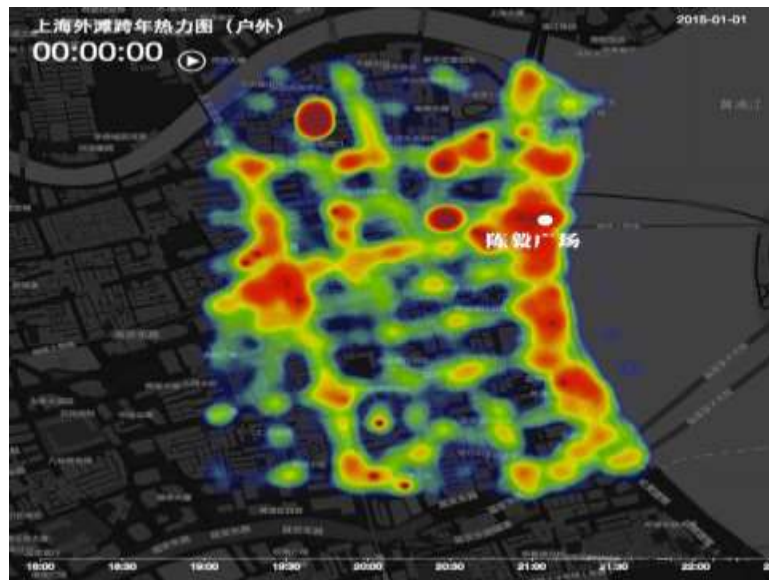
# 时空数据天天见



室内定位:

时间, 相对坐标系,  $x, y, z$ , 其他属性

# 时空数据天天见



室外定位：  
时间，经纬度



# 时空数据天天见



自动驾驶：

区域信息，线段（道路），点云  
路径规划。

# 时空数据无处不在





02

# 时空业务需求与挑战

# 业务需求

- 选址、网格运营
  - 空间数据自动聚集分析；时间+多边形圈人；驻留时间分析；舆情分析；...
- 室内定位
  - 3D坐标；相对坐标系；+以上；运营活动效果分析报表；
- 无人驾驶
  - 点云；动态路径规划；
- 空间调度（菜鸟）
  - 实时位置更新；多边形圈人；
- LBS
  - 实时位置更新；基于GIS+其他条件的推荐(探探)；附近+其他属性圈数（地图附加业务，找餐馆、加油站）
- 导航
  - 动态路径规划
- 时空透视
  - 滑动窗口；流式计算；多维分析；...
- 视觉搜索
  - 文本搜索转换为视觉搜索；(店铺id,浏览次数)，搜索浏览了某些店铺若干次到若干次或xxx的对象。



# 挑战

- 数据维度多
  - 时间维度、空间维度、对象本身属性维度
- 数据检索复杂
  - 空间检索
  - 时间范围检索
  - 对象属性检索
  - 复合检索
  - 路径检索
- 空间数据格式混乱
  - 坐标系、文本、经纬度、纠偏、POINT、raster、ESRI、KMZ、OSM ...
- 空间类型繁多
  - 平面（点、经纬、多点、线段、多边形、圆、混合对象...）、3D、4D、多维、栅格、点云、拓扑...
- **TP + AP**需求
  - 在线处理
  - 实时分析
  - 离线分析



03

# PostgreSQL时空特性

# 挑战 1

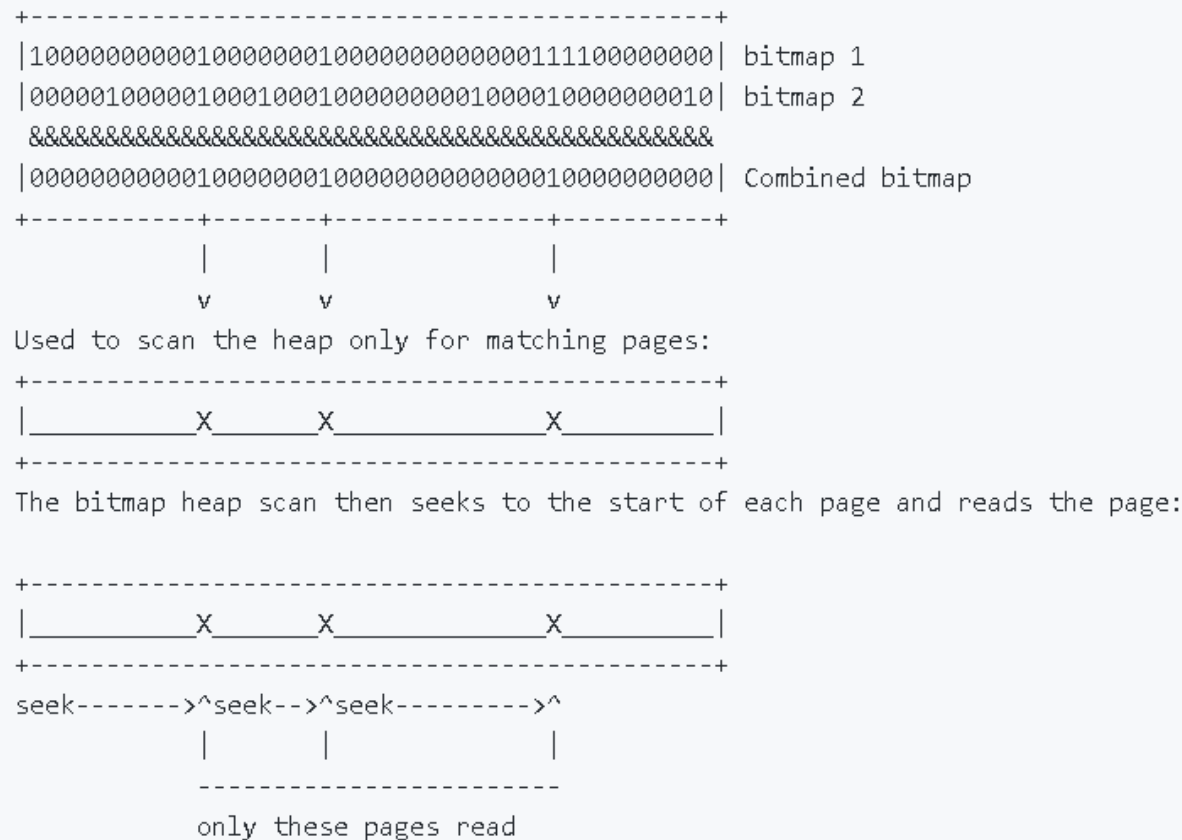
- 数据维度多
  - 时间维度、空间维度、对象本身属性维度
- 数据检索复杂
  - 空间检索
  - 时间范围检索
  - 对象属性检索
  - 复合检索
  - 路径检索

# 空间索引

- 1、平面、三维、多维对象 几何相交、不相交、相邻。
- 2、平面、三维、多维对象的方位判断（相交或严格在左边、右边、上边、下边），类似数值的大于、小于、大于等于、小于等于。
- 3、平面、三维、多维对象 包含 另一个对象
- 4、平面、三维、多维对象 等于 另一个对象
- 5、平面、三维、多维对象 与另一个对象的（边、最近、中心点的）距离。
- 6、按距离排序输出满足条件的行，输出距离在xx以内的行。
- ...

# bitmap scan

- 多索引bitmap扫描
- 支持任意维度查询



# 空间复合索引

- 一个索引包含多列（空间列+其他列）
  - 支持复合查询
  - create table test (id int, crt\_time timestamp, pos geometry);
  - create index idx\_test\_geo on test using gist (pos, crt\_time);
  - select \* from test where
    - crt\_time between x and x
    - and st\_dwithin('\$polygon'::geometry, pos);



# 空间分区索引

- create index idx1 on test using gist (geo) where 普通分区条件1;
- ...
- create index idx2 on test using gist (geo) where 普通分区条件n;
- ...
- create index idx1 on test using btree (col) where 空间分区条件1;
- ...
- create index idx2 on test using btree (col) where 空间分区条件n;
  
- 在**5000**万空间数据中，按时间、空间、对象属性进行多维检索，响应时间**0.592** 毫秒。
  
- select \*, (pos <-> \$1) as dist from tbl
- where
- crt\_time between '2017-07-22 17:59:34' and '2017-07-22 17:59:40'
- and ( c1 in (1,2,3,4,100,200,99,88,77,66,55) or c2 < 10 )
- order by pos <-> \$1

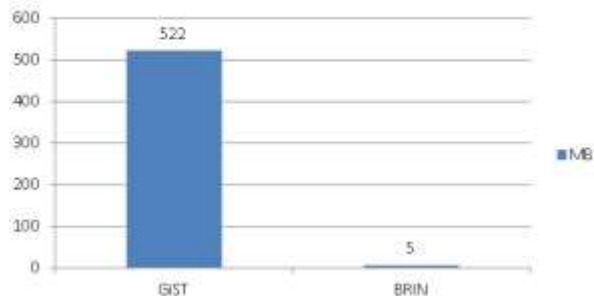
# 空间块索引

- 堆表组织形式：BLOCKs
- BRIN块级索引：每BLOCK(s)内所有空间数据的BOX BOUND。
  - 按geohash规整后，BRIN性能贼好

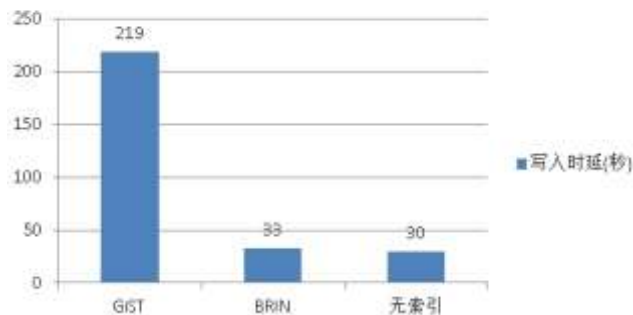
# 空间块索引

1000万空间数据

空间对比(MB)



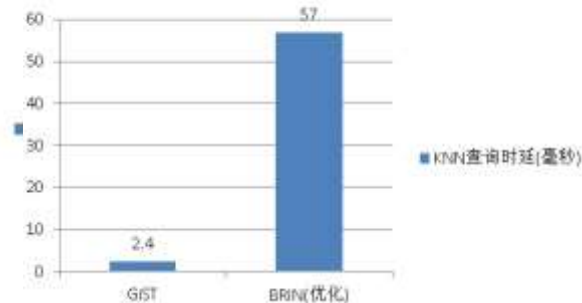
写入时延(秒)



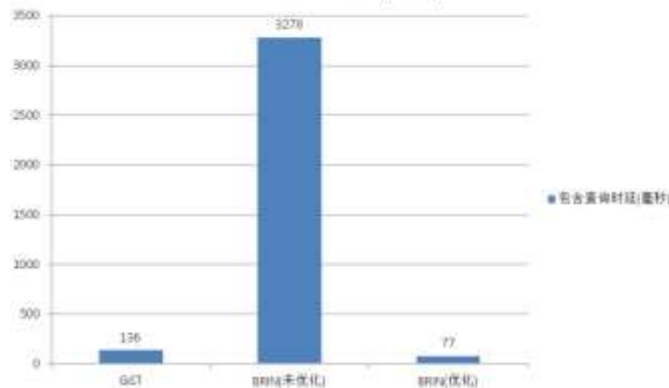
创建时延(秒)



KNN查询时延(毫秒)



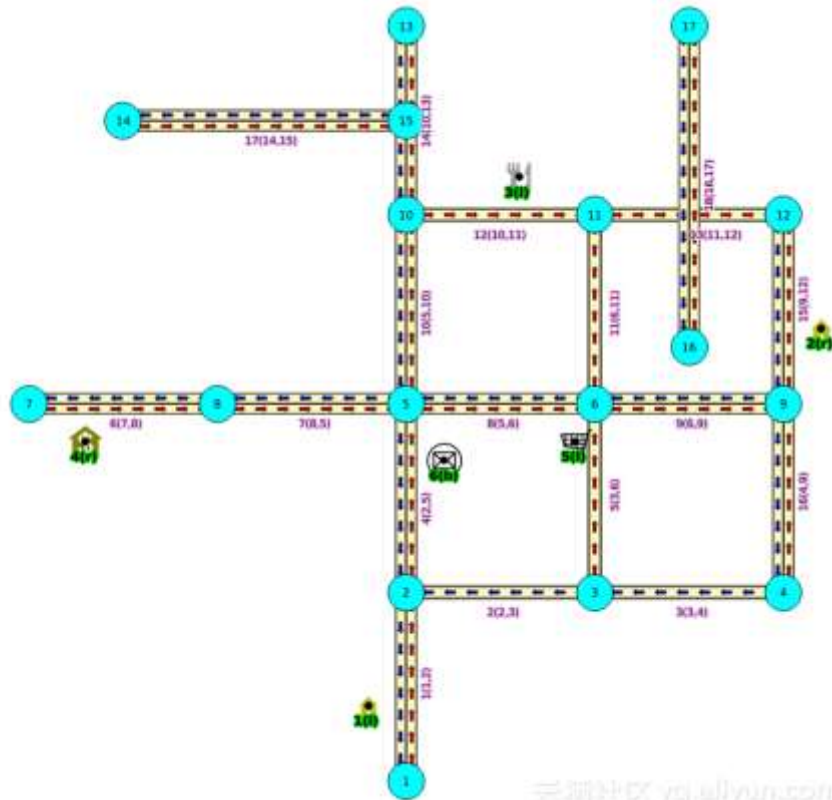
包含查询时延(毫秒)



# 理论上支持无数种索引接口

- `CREATE INDEX idx ON table USING 索引接口 (列);`
- 索引接口已支持9种
  - b-tree, hash, gin, gist, sp-gist, bloom, rum, brin, zoomdb
  - 支持扩展接口

# 路由规划 - pgrouting



# 挑战 2

- 空间数据格式混乱
  - 坐标系、文本、经纬度、纠偏、POINT、raster(PBF、sharp、...) ...
- 空间类型繁多
  - 平面（点、经纬、多点、线段、多边形、圆、混合对象...）、3D、4D、多维、栅格、点云、拓扑...

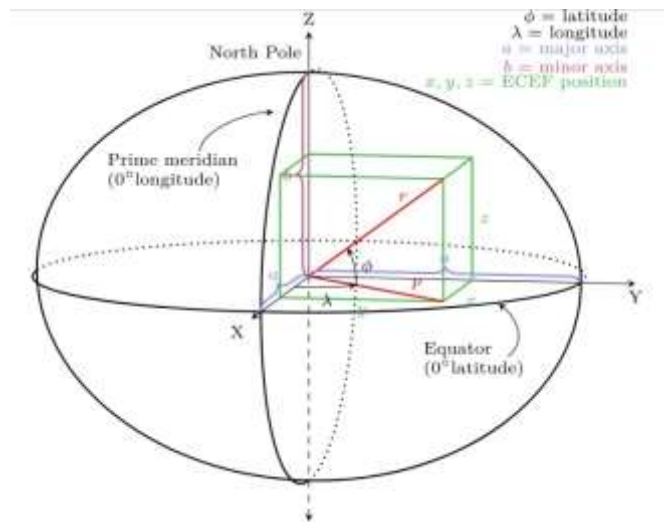


# 格式容错

- 支持UDF，格式转换。
- 规则、触发器。数据入库自动转换。
- 支持表达式索引，不转换的情况下建立空间索引。（例如在文本上创建空间索引）

# 数据类型

- 兼容OGC, 3000+ SRID 的支持
- geometry、geograph
  - 平面（点、经纬、多点、线段、多边形、圆、混合对象...）、
  - 3D、4D、nD
- 栅格
- 点云
- 拓扑
- ...

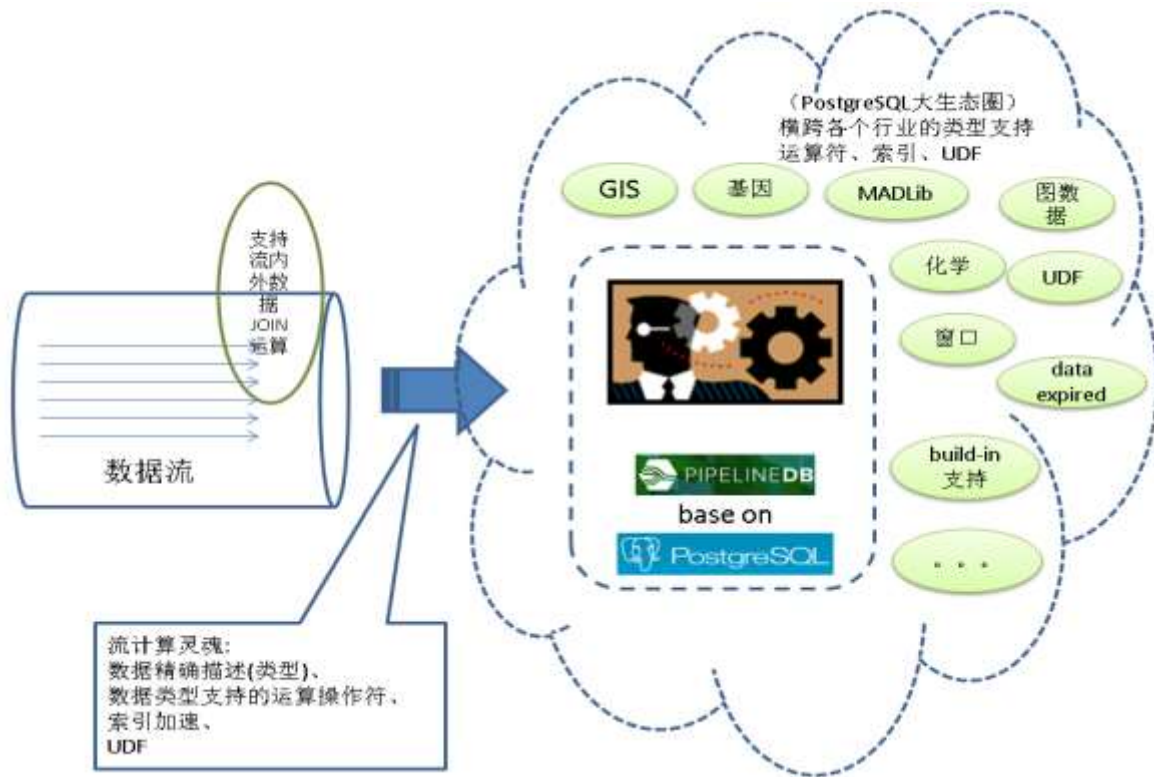


# 挑战 3

- TP + AP需求
  - 在线处理
  - 实时分析
  - 离线分析

# 空间流计算

- 流式商圈人流量统计
- 人群聚集实时舆情分析
- ...



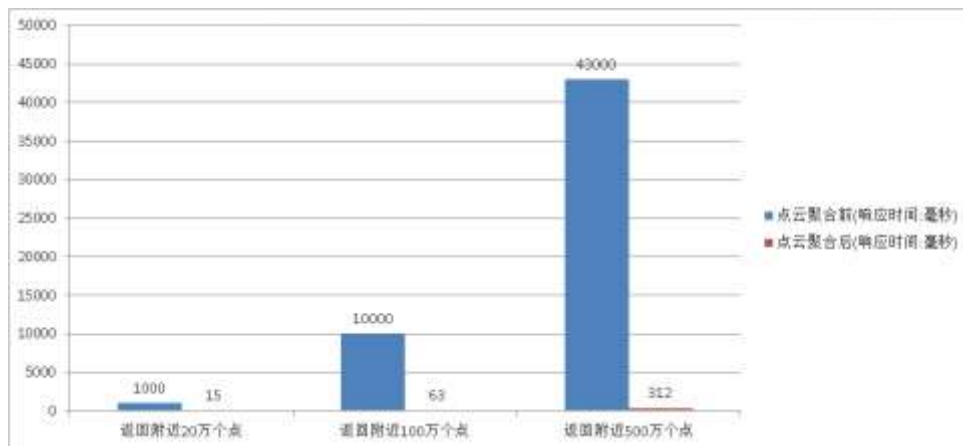
# 实时分析

- 对接R、Python
- MADlib机器学习库
- UDF
- GIS分析函数
- 窗口查询
- 多维分析语法
- 递归查询
- 多核并行
- 复杂JOIN
- GPU+FPGA插件



# 在线GIS数据处理性能

- KNN检索
  - 100亿空间数据，单机(16核)。
  - 检索任意随机点最近的点（7.4万tps，平均响应0.848毫秒）
- 点云数据吞吐

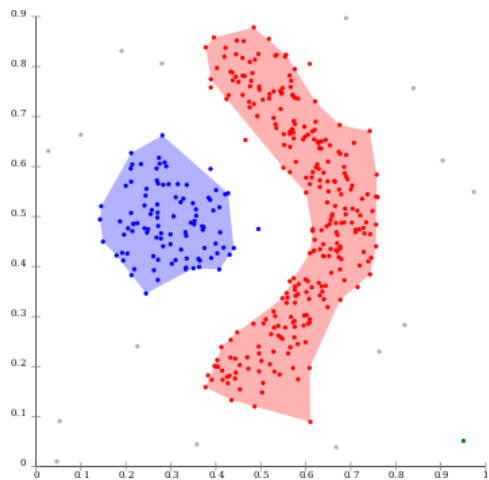


- 多维查询(时间、空间、对象多维度过滤)
  - 5000万空间数据，多维查询，0.592毫秒。

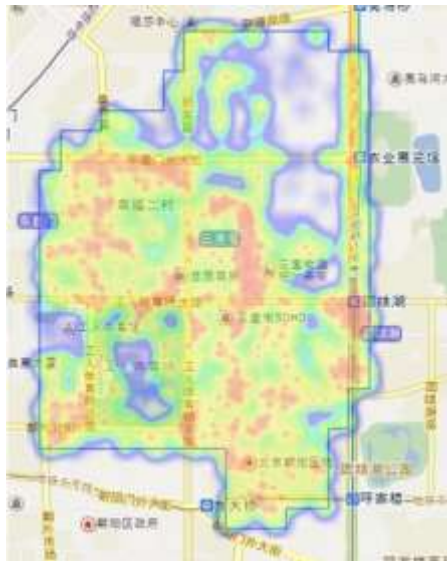


# 地上本无路、走的人多了便成了路

空间数据  
聚集分析



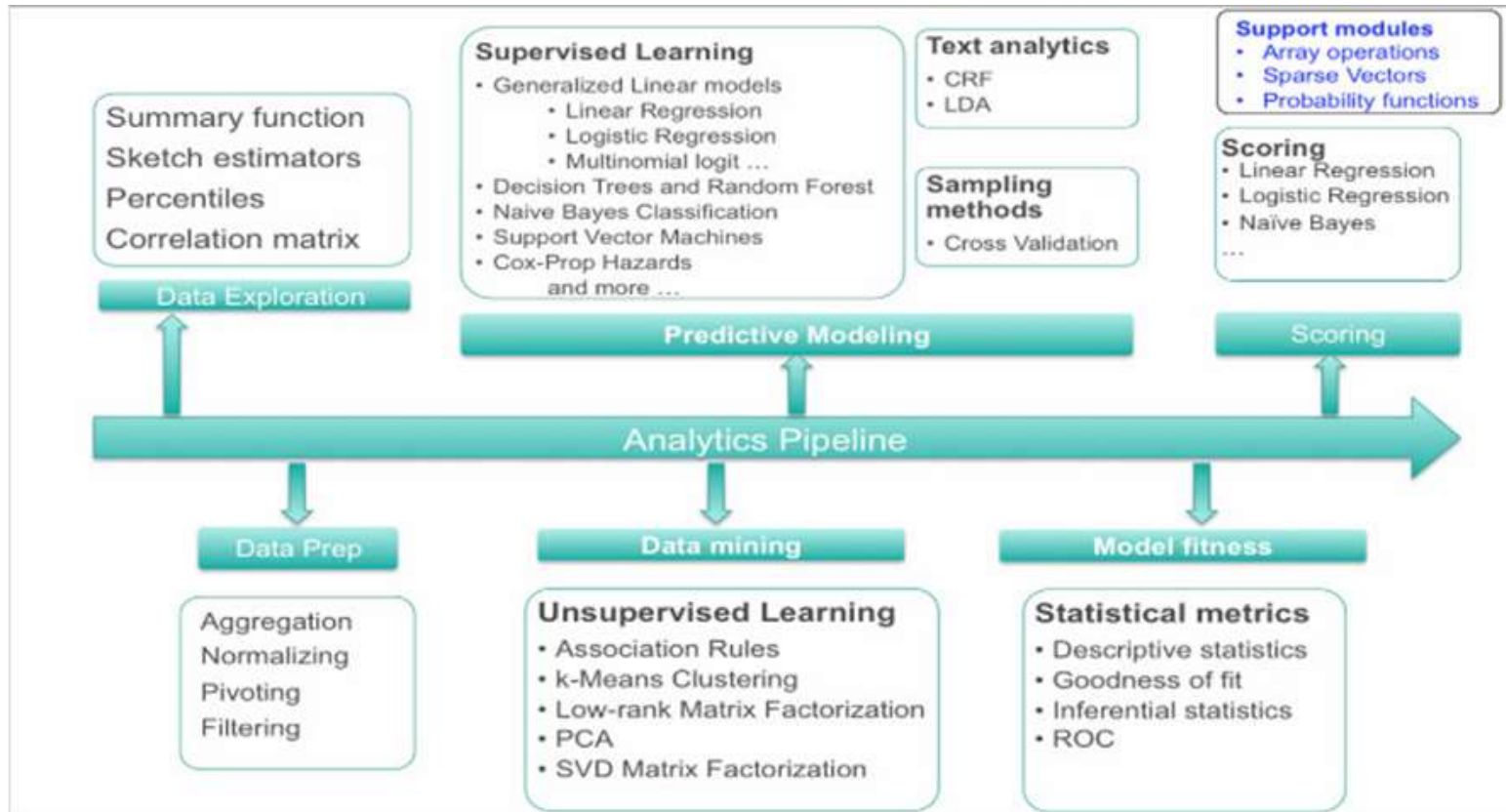
人流量大了便成了黄金旺铺



时空数据  
寻龙点穴



# MADlib



# 离线分析

- FDW接口
  - 阿里云RDS PG + OSS
  - 冷热分离、数据共享。
  - 数据高速传输通道。
  - 备份存储。
- MPP
  - 阿里云HybridDB for PG + OSS
  - 并行计算、冷热分离、数据共享。
  - 数据高速传输通道。
  - 备份存储。

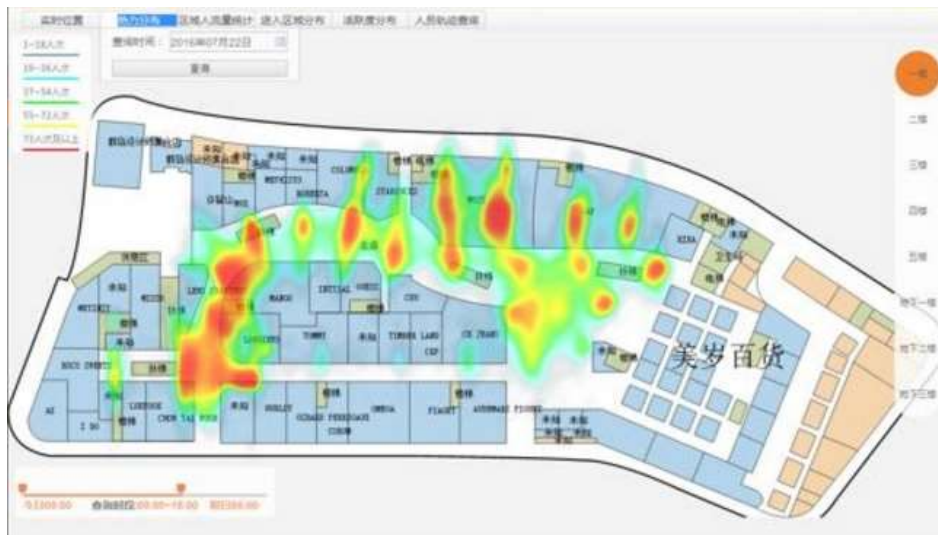


04

# 案例

# 时间、空间、对象多维数据透视

- select
- to\_char(timecol,'fmt'), -- 时间点到时间窗口
- get\_polygon(pos), -- 点到区间窗口
- obj\_pro1, obj\_pro2, ..., -- 对象属性维度
- count(\*), 其他聚合函数,...
- from table
- group by 1,2,3,4,...;



# 新零售 - 网格化运营

- 线下商铺人群位置属性+行业属性+时间属性+客户其他属性透视
- 线下商铺根据位置属性+其他条件圈选目标人群



# 大健康 - 高科技敬老院

- 1、一个人的行为轨迹，在某个范围不动，持续多久预警。
- 2、一个人在床的空间，躺了12个小时以上，预警。
- 3、一个人高度低于40公分，持续5分钟，预警
  - （周围有人，不预警）。
- 4、一个人在马桶的狭小空间，超过30分钟，预警。



# 菜鸟 - 包裹侠

- 配送实时位置跟踪
  - 86.4亿记录，17.4万/s位置更新
- 规划快递员最佳配送路径
  - pgrouting
- 实时件需求-召回快递员调度
  - 每秒处理召回请求 28万。
- 寄件调度-召回快递员调度
  - 点面判断+属性过滤
  - 1000万多边形，每秒29万处理请求，单次平均响应时间0.2毫秒。



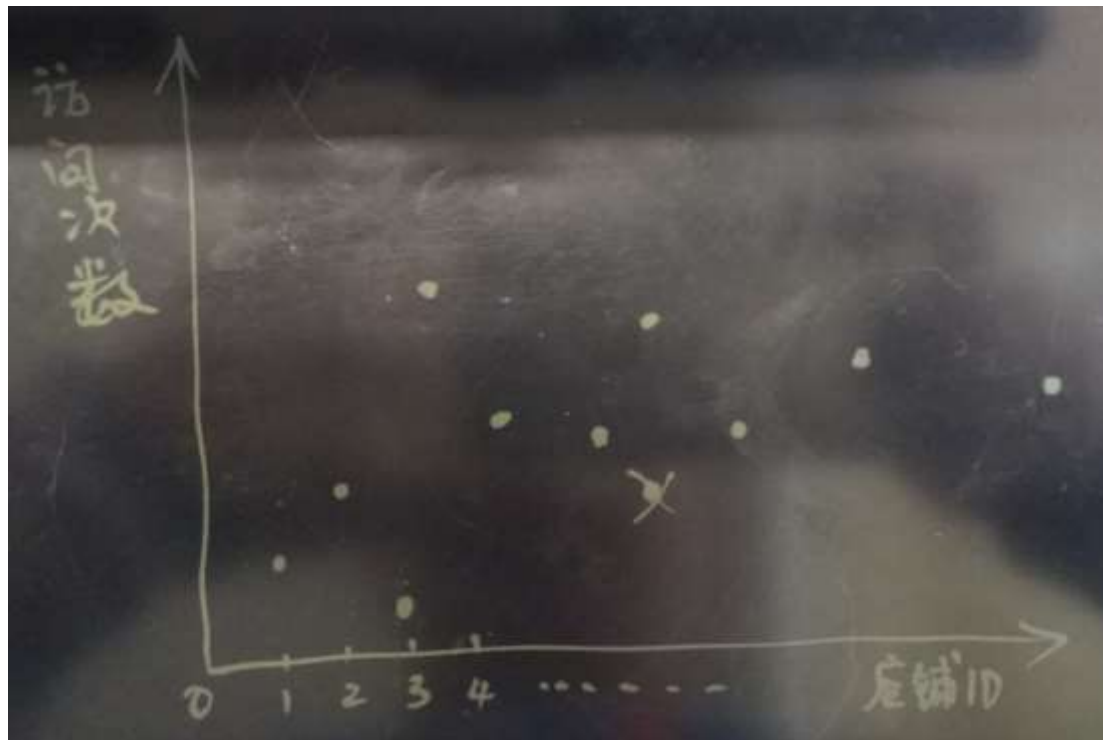
# 菜鸟 - 自动配送

- PostgreSQL点云插件pgpointcloud。

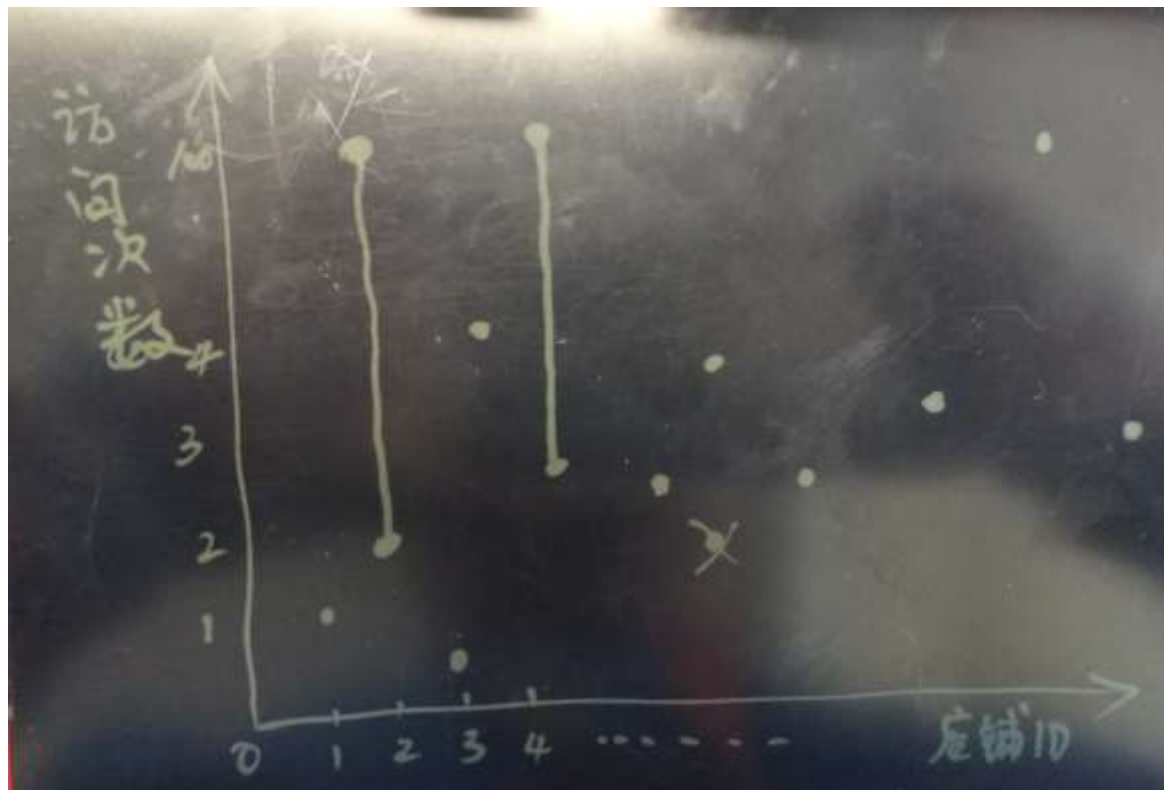
# 视觉搜索

- user, (店铺ID:访问次数;店铺ID:访问次数;....)
- 根据店铺的访问次数圈出一部分人群，比如A店铺访问超出多少次的，或者B店铺访问超过多少次的等。
- insert into test values (1, '1:1');
- insert into test values (2, '1:100');
- insert into test values (3, '2:1');
- insert into test values (4, '2:100');
- insert into test values (5, '1:100;2:100');
- GIS表达式索引，对文本建立空间索引。
- 利用GiST索引进行视觉搜索（空间"包含"、"相交"操作符。multiline 包含 multi point）

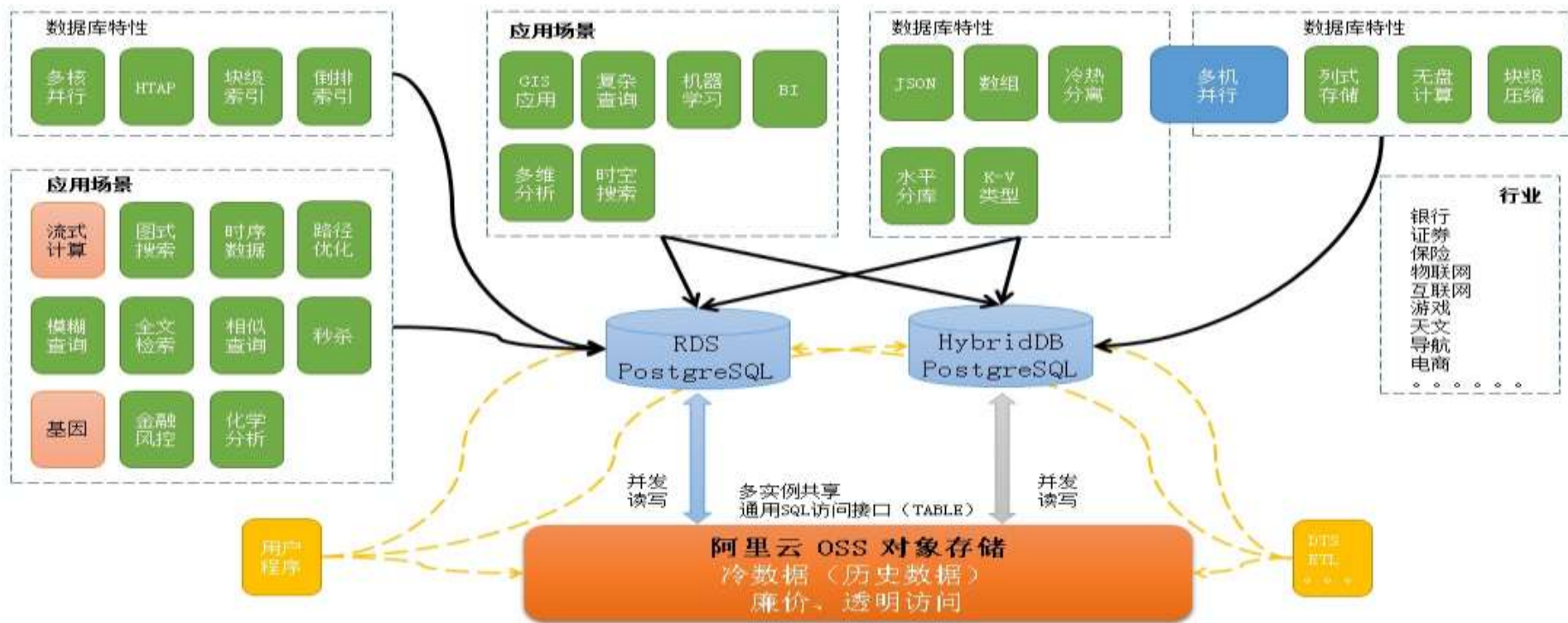
# 视觉搜索



# 视觉搜索



# 云上TP+AP经典架构



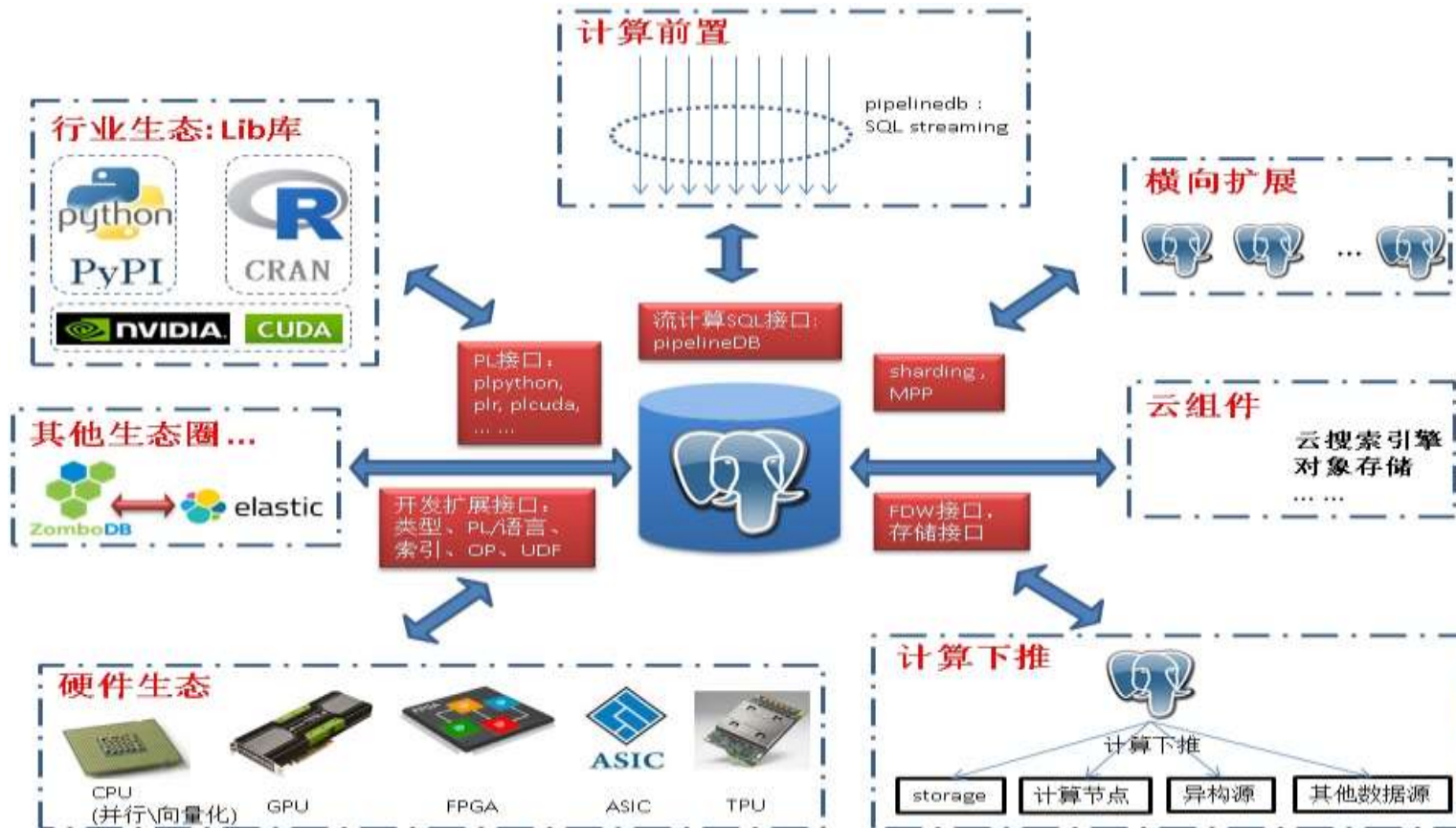
# 其他案例

- 高德
- 探探
- 中国虚拟天文台、美国宇航局、俄罗斯天文台、...
- 《基于PostgreSQL和GIS信息打造的洞察平台》
- 《PostGIS空间索引(GiST、BRIN、R-Tree)选择、优化 - 阿里云RDS PostgreSQL最佳实践》
- 《分区索引的应用和实践 - 阿里云RDS PostgreSQL最佳实践》
- 《空间复合索引加速空间搜索》
- 《多字段，任意组合条件查询(无需建模) - 毫秒级实时圈人 最佳实践》
- 《PostgreSQL 9种索引的原理和应用场景》
- 《时间、空间、对象多维属性 海量数据任意多维 高效检索 - 阿里云RDS PostgreSQL最佳实践》
- 《PostgreSQL 海量时序数据(任意滑动窗口实时统计分析) - 传感器、人群、物体等对象跟踪》
- 《PostgreSQL\GPDB 毫秒级海量时空数据透视 典型案例分享》
- 《(AR虚拟现实)红包 技术思考 - GIS与图像识别的结合》



展望

# PostgreSQL HTAP







# 谢谢



如来神掌  
二维码



# Thanks

关注开源数据库论坛