

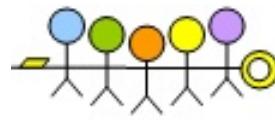
# 数据库超体



10%

- 增
- 删
- 改
- 查





20%



20%

- 一、物联网、金融、日志、运营商网关、行为轨迹类数据

数据特征

append  
only

堆存储  
线性相关

时序

量大

需求

区间查找

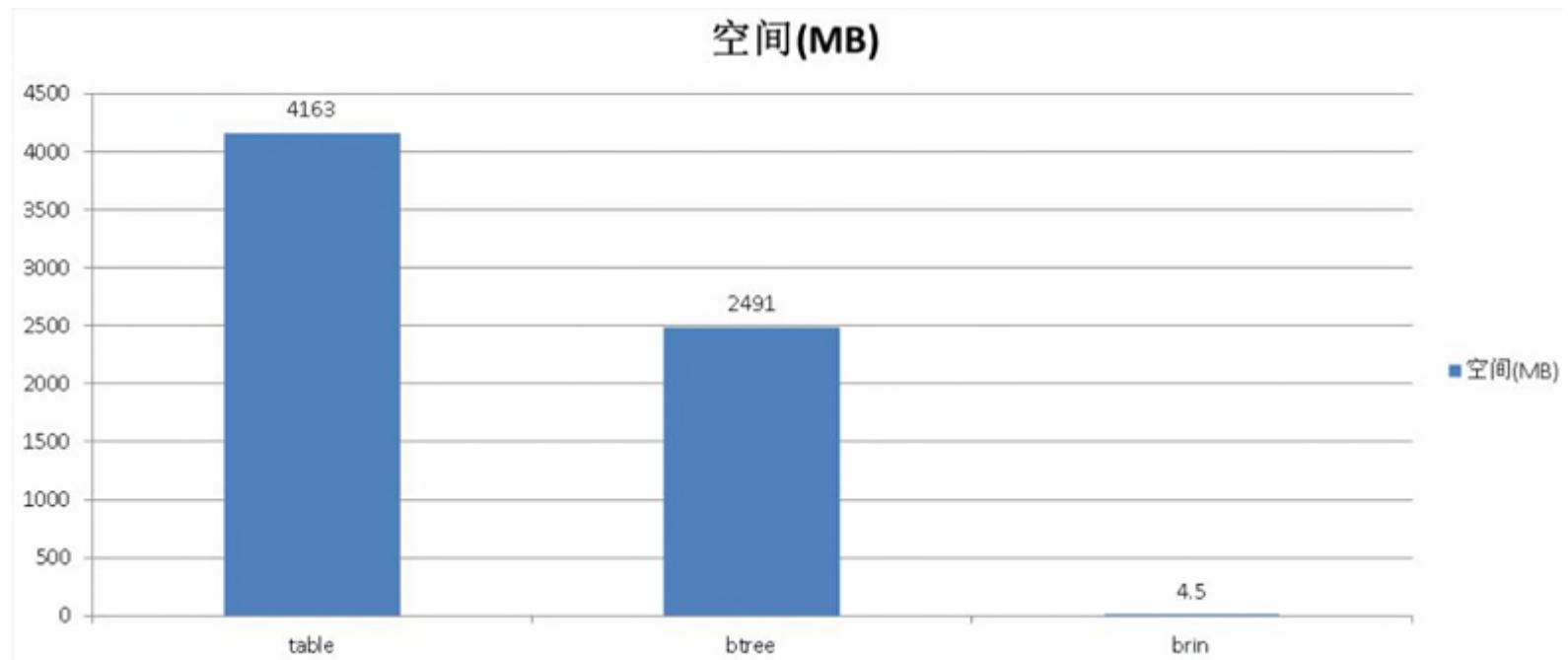
分析

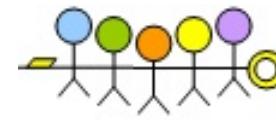
插入快

压缩  
节约空间

20%

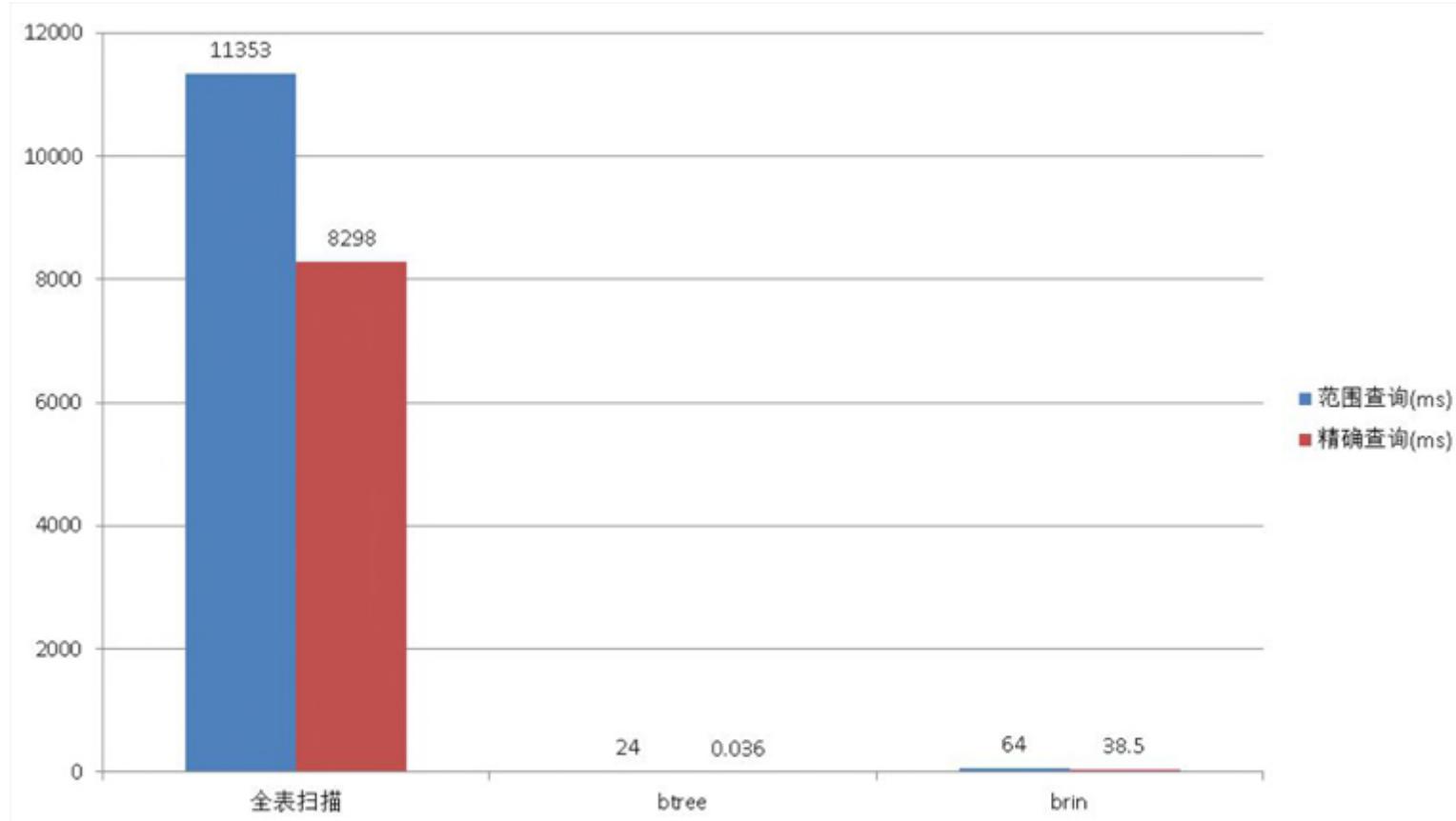
- PostgreSQL 块级别索引 - BRIN





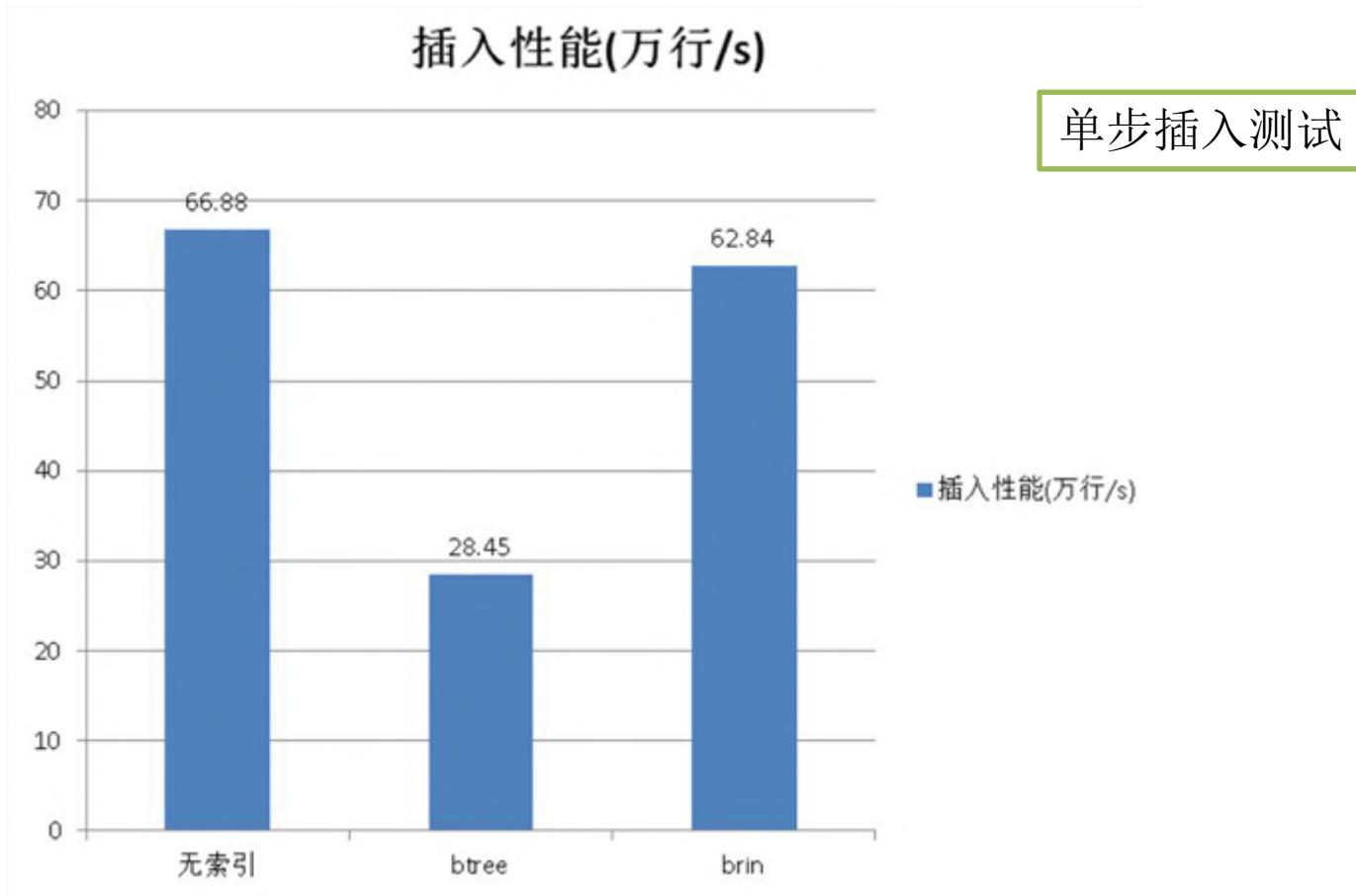
20%

- PostgreSQL 块级别索引 - BRIN



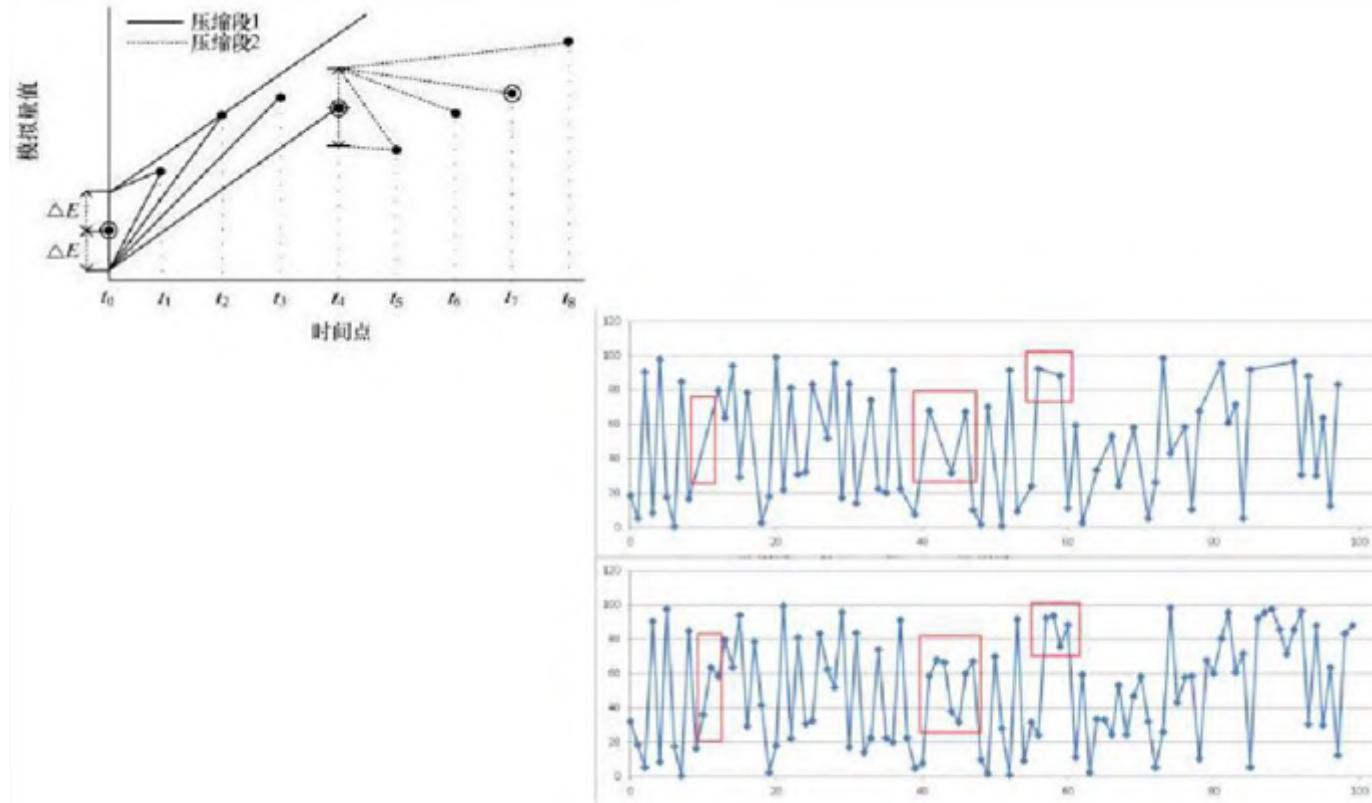
20%

- PostgreSQL 块级别索引 - BRIN



20%

- 时序类数据有损压缩 - 旋转门压缩



- 时序类数据无损压缩 - 列存储(瓦片式\内置\FDW\IMCS)

20%

## • 二、搜索类、多维度交互类场景

实例名  Select  高级搜索

实例名:

用户连接:

用户信息:

可用区:  集群名:

库类型:  库版本:

实例类型:  网络类型:

VIP类型:  TOP类型:

实例状态:  锁定模式:

服务状态:  是否非标:

SQL WALL:  业务类型:

实例角色:  网络 MODE:

过期时间:  至

创建时间:  至

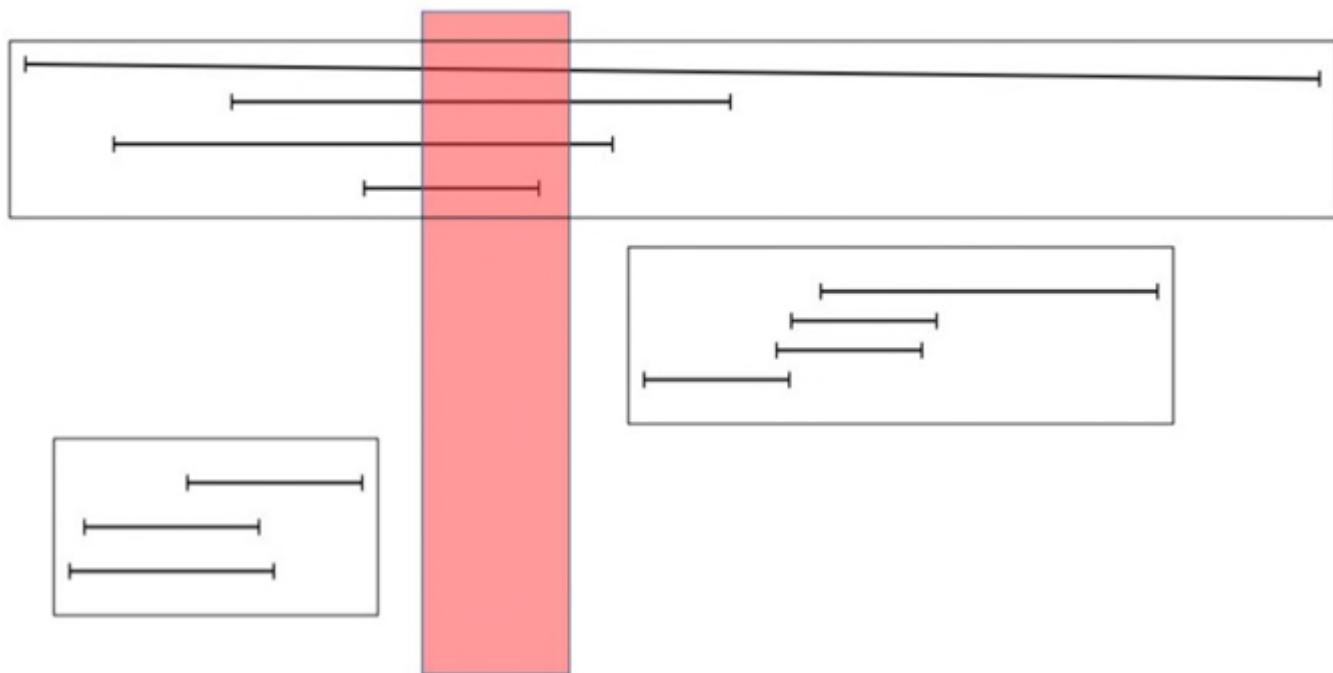


1. GIN 复合(倒排+聚集)索引
2. 分词索引
3. bloom复合索引
4. pg\_trgm, fuzzymatch

20%

- 三、高效率范围查询

Group into clusters



20%

- 三、高效率范围查询



## GIST search

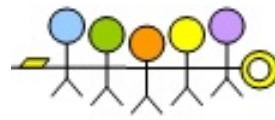
Search key: [55, 60]

[1, 200]  
[80, 150]  
[10, 45]

[1, 200]	(0, 2)
[20, 70]	(5, 1)
[30, 50]	(4, 3)
[10, 60]	(4, 2)

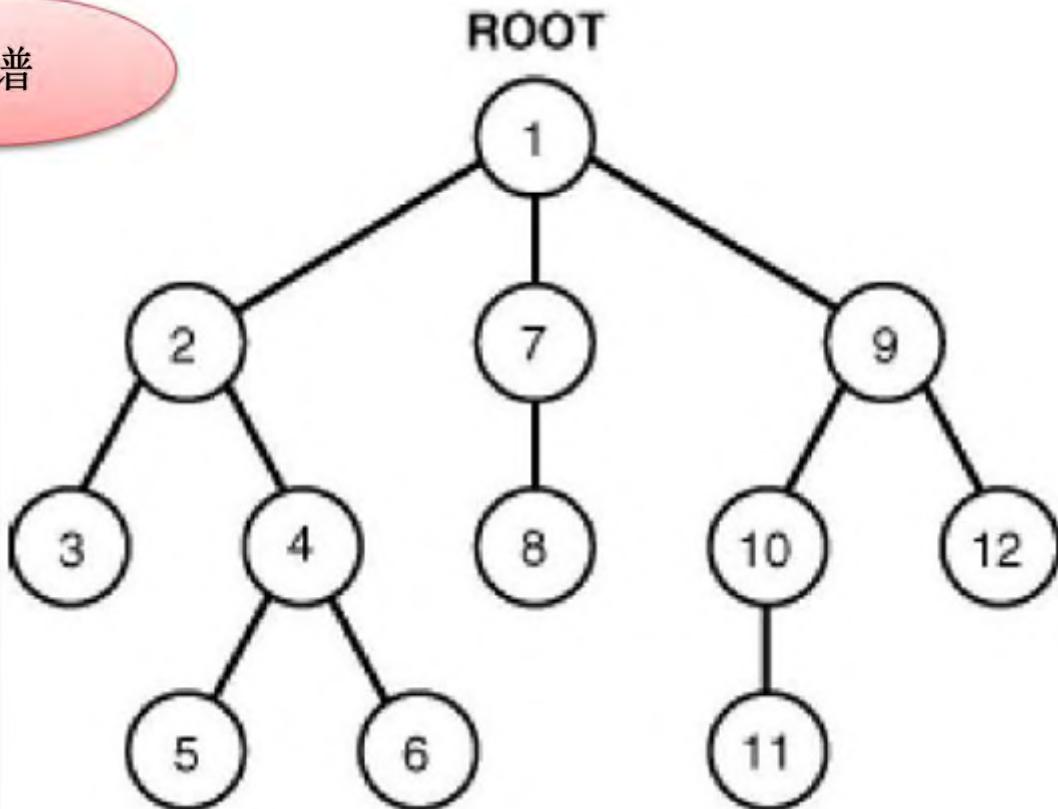
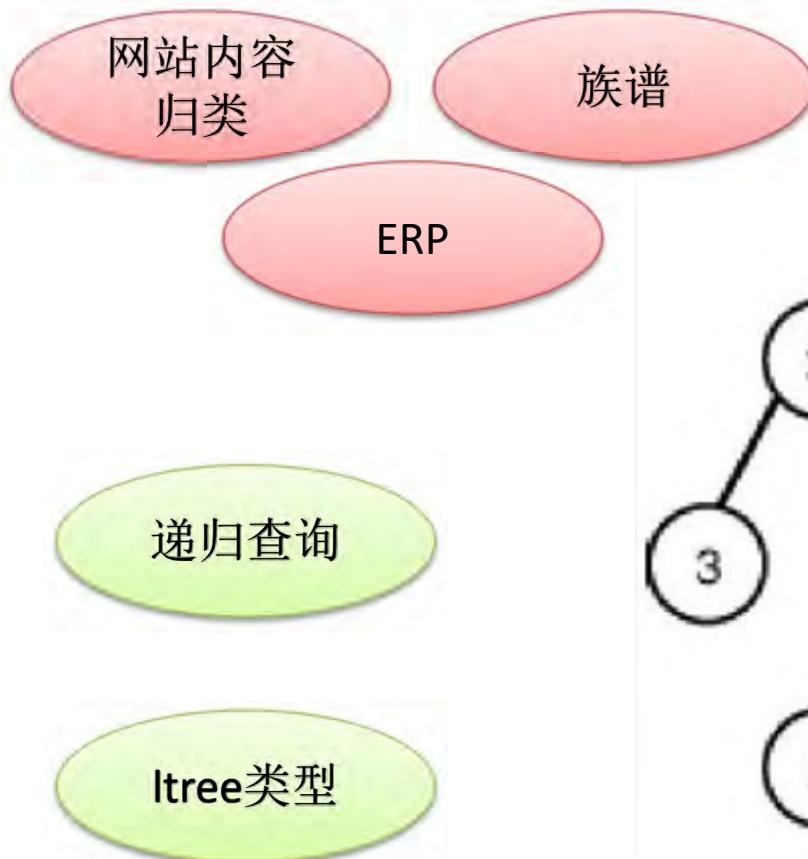
[100, 150]	(1, 10)
[110, 120]	(2, 2)
[105, 115]	(3, 4)
[80, 90]	(9, 2)

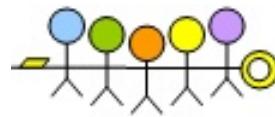
[25, 45]	(8, 1)
[15, 30]	(2, 1)
[10, 20]	(1, 7)



# 20%

- 四、悬殊数据求差、树型数据查询、收敛加速



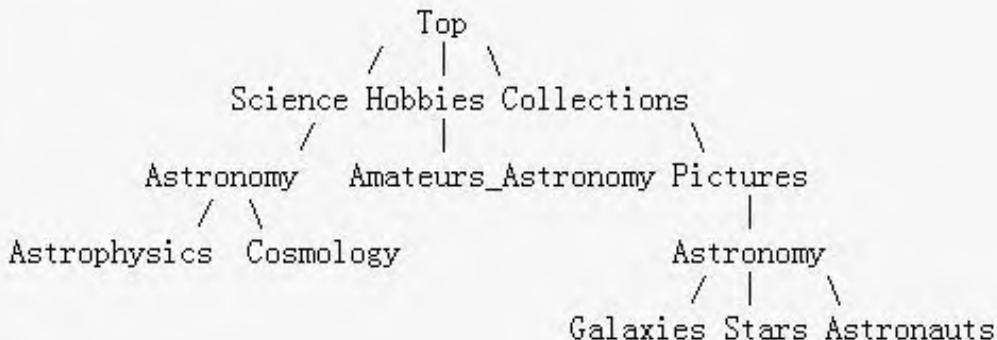


# 20%

- 四、悬殊数据求差、树型数据查询、收敛加速

递归查询

ltree类型



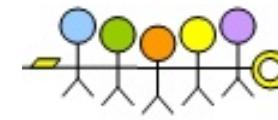
We can do inheritance:

```
lreetest=> SELECT path FROM test WHERE path <@ 'Top.Science';  
path
```

```
Top.Science  
Top.Science.Astronomy  
Top.Science.Astronomy.Astrophysics  
Top.Science.Astronomy.Cosmology  
(4 rows)
```

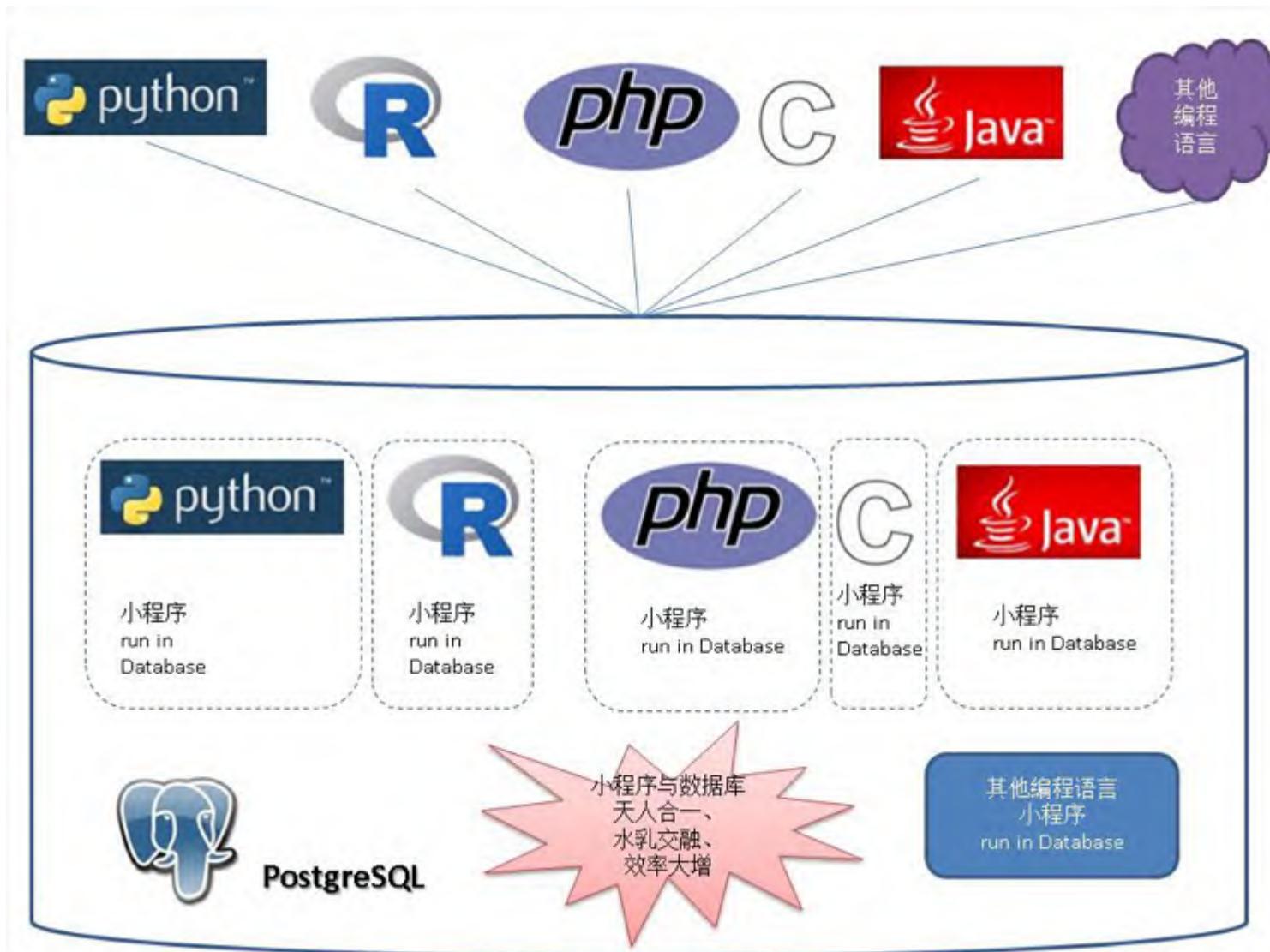
40%





40%

- 一、
- 数据库端编程
- 
- 
- 
- 
- 



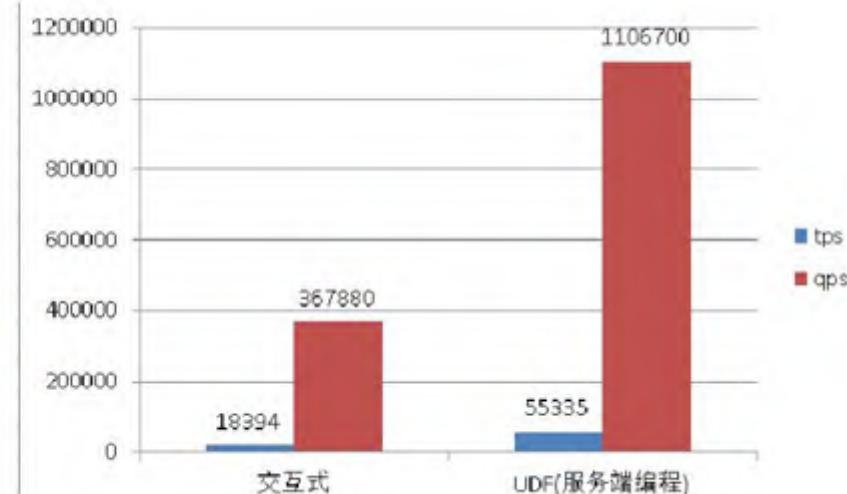
40%

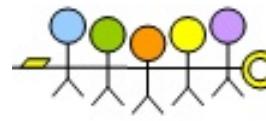
- 一、
- 数据库端编程



对比20条**QUERY(pk I,U,D,S)**的事务TPS

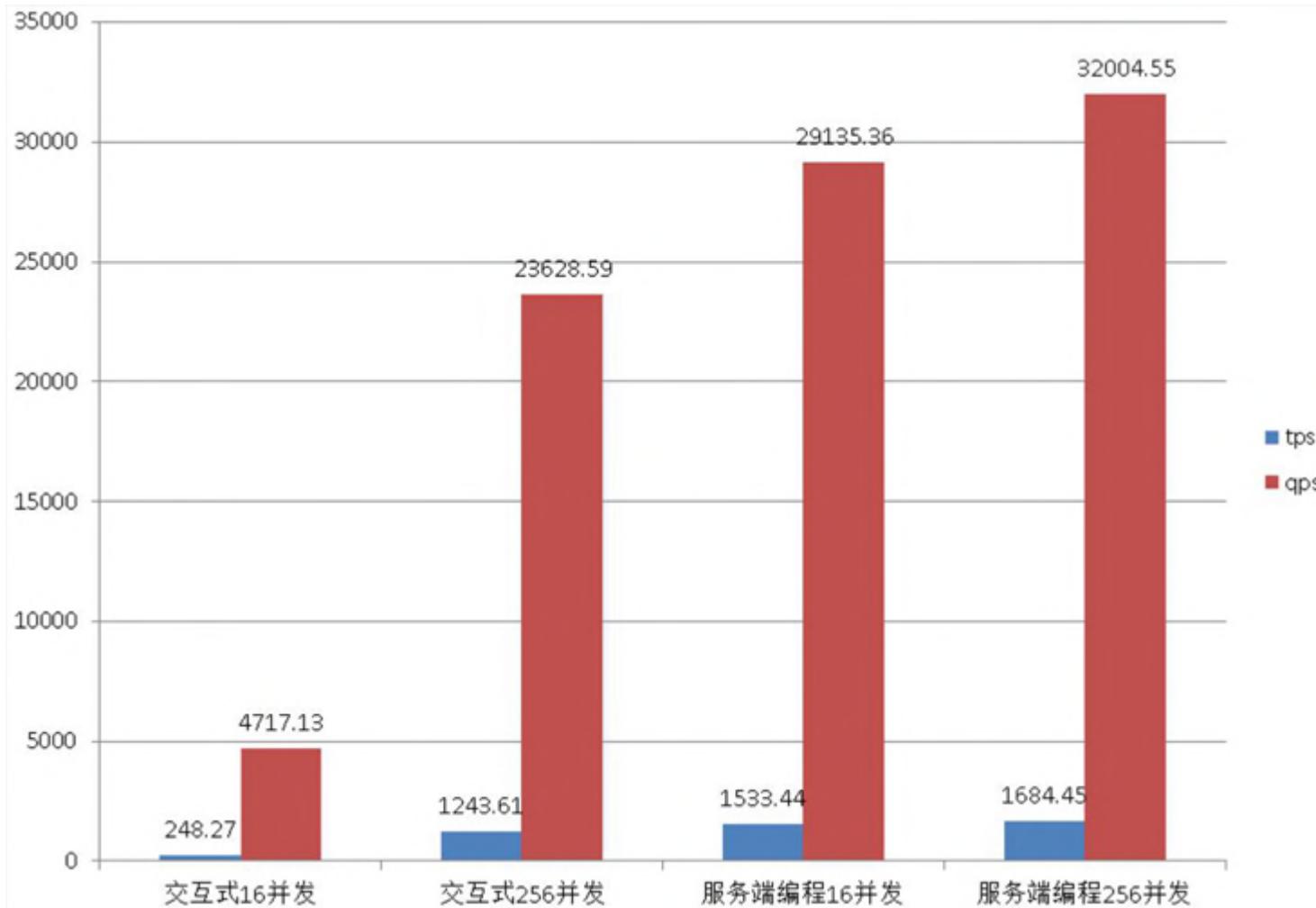
1. 交互式（业务逻辑在**client**完成）
2. 服务端编程模式（业务逻辑封装在数据库**FUNC**完成）





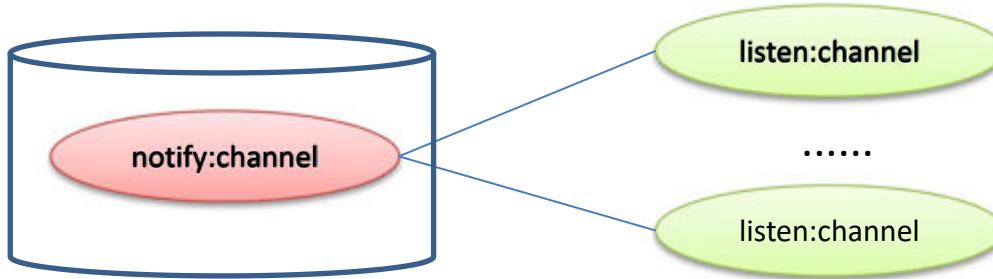
**40%**

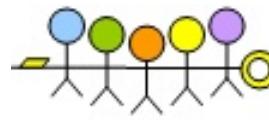
# 一、 数 据 库 端 编 程



40%

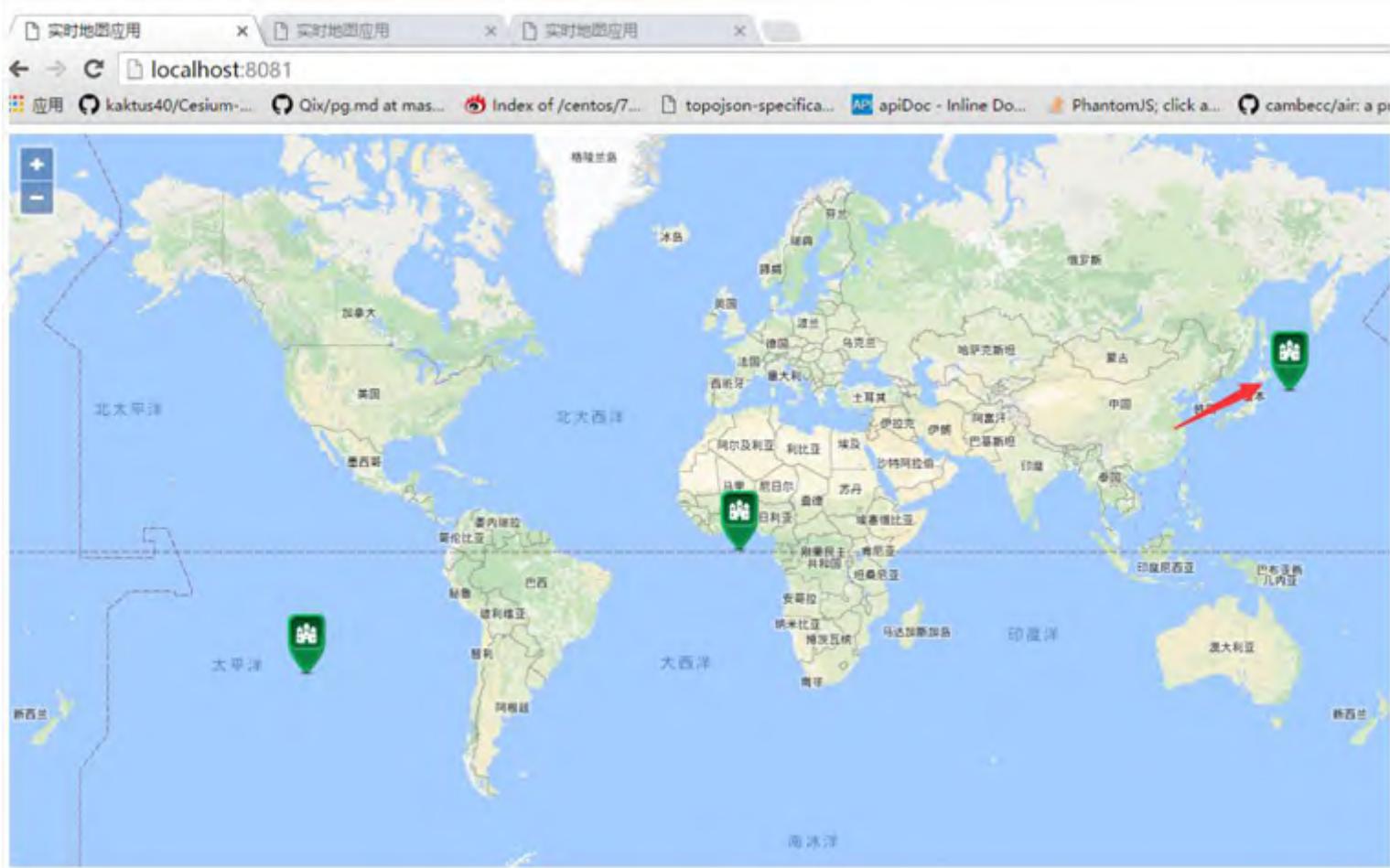
- 二、
- 异
- 步
- 消
- 息





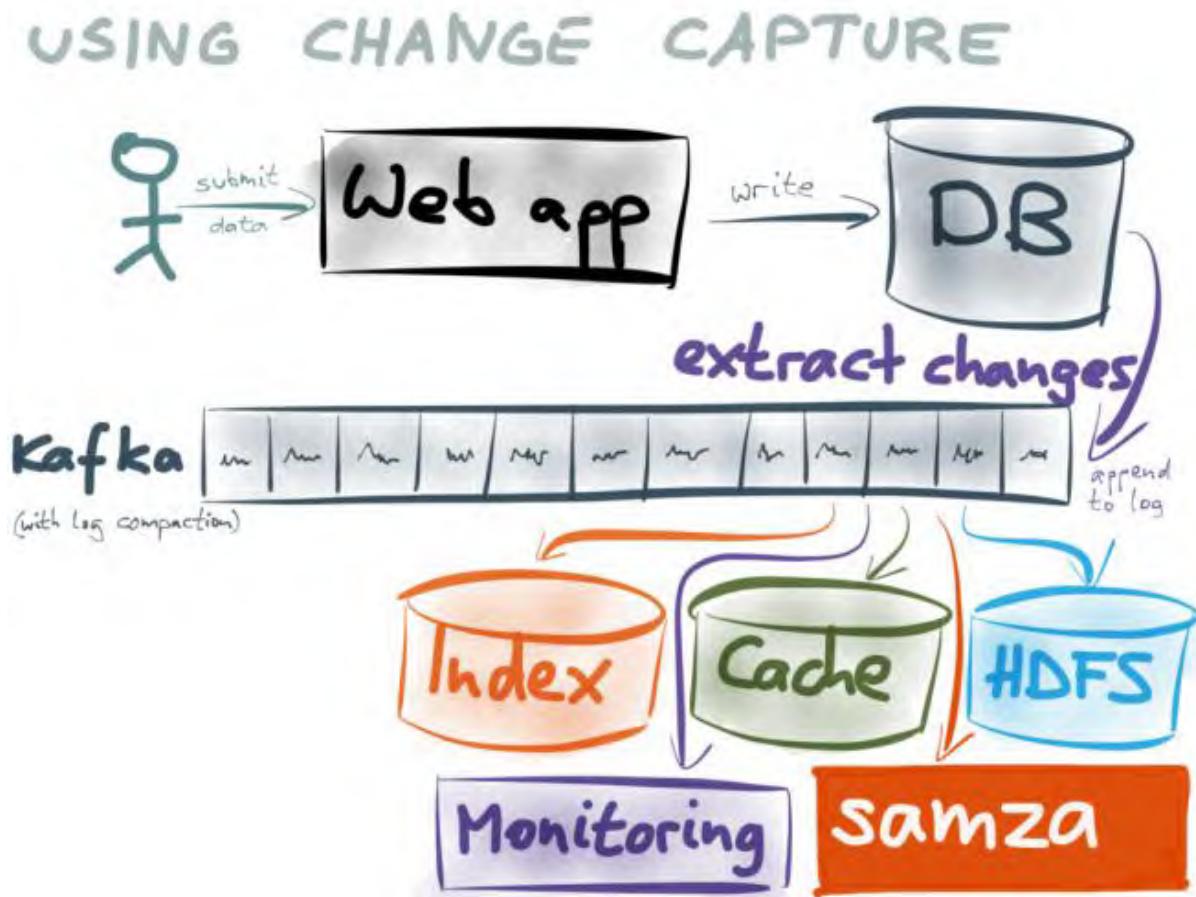
# 40%

- 二、
- 异步消息
- 
- 
- 



40%

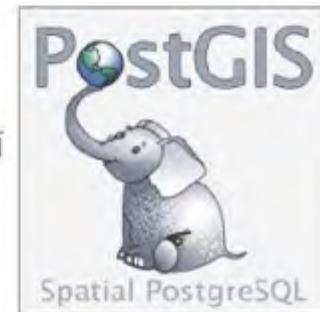
- 三、数据泵

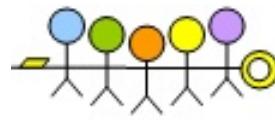


40%

- 四、天文、地理  
**PostGIS(覆盖民用、军用、科研)**

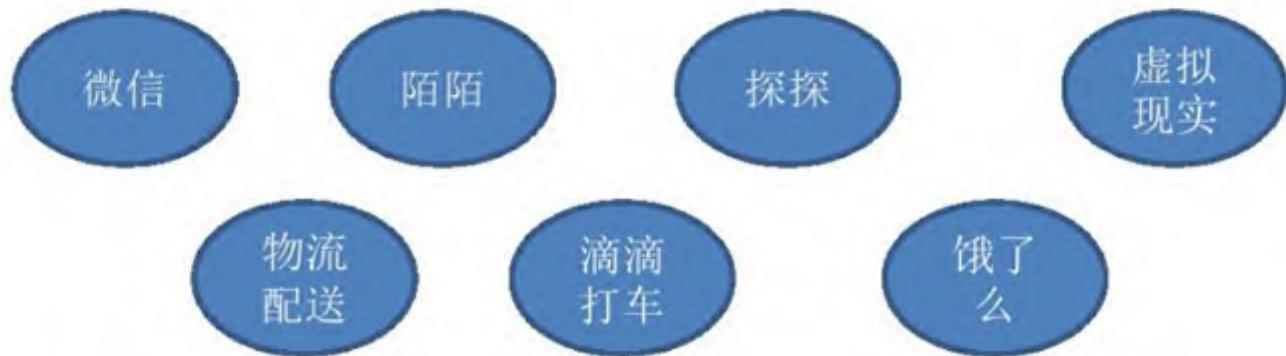
- 类型
  - 点
    - 二维、三维（经纬度、海拔）、多维、地址类型
  - 线
    - 闭合线段、开放线段、多点线段
  - 面
    - 圆、椭圆、矩形、长方形、正方形、规则多边形、不规则多边形、曲面
  - raster
    - GIS图数据, jpeg, tiffs, pngs, digital elevation models
    - 测绘、航天、天文应用
  - 拓扑
- 操作符
  - 点面判断、距离、面积、体积、叠加、相减、长度、弧度、夹角、pixel相关运算、raster相关运算
- 索引
  - 距离运算、距离排序、包含判断、相交判断、
- 函数
  - 区域内线段长度（城市道路长度）、路径成本(坡度、长度、权重)计算、最佳路径计算、区域面积（城市面积）

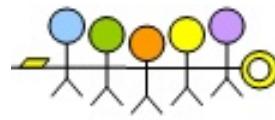




# 40%

- 近邻查询
  - 根据经纬度查询最近用户
  - 根据数值查询最相邻数值
  - 根据文本相似度查询最相似文本
- 距离排序





40%

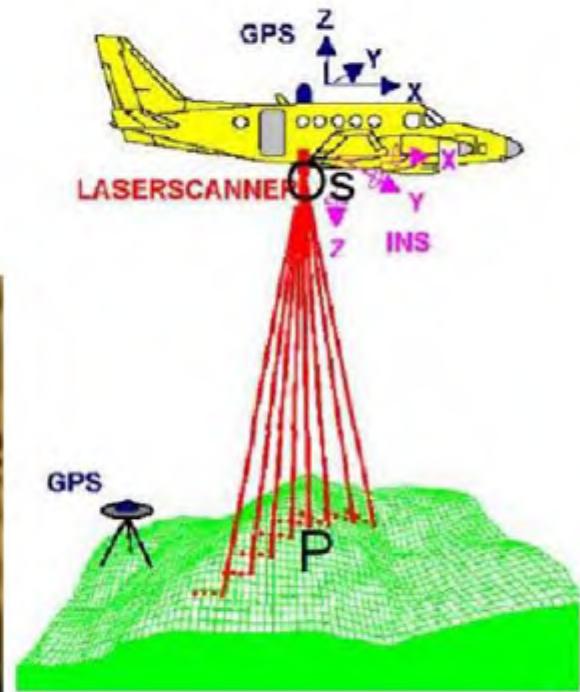
## 3D数据处理(raster,pgpointcloud)

位置、**RGB**、密度、材质、。。。位点多元属性



3D打印是近几年兴起的一种技术，除了存储物体表面的位置信息，还有颜色，密度等信息。

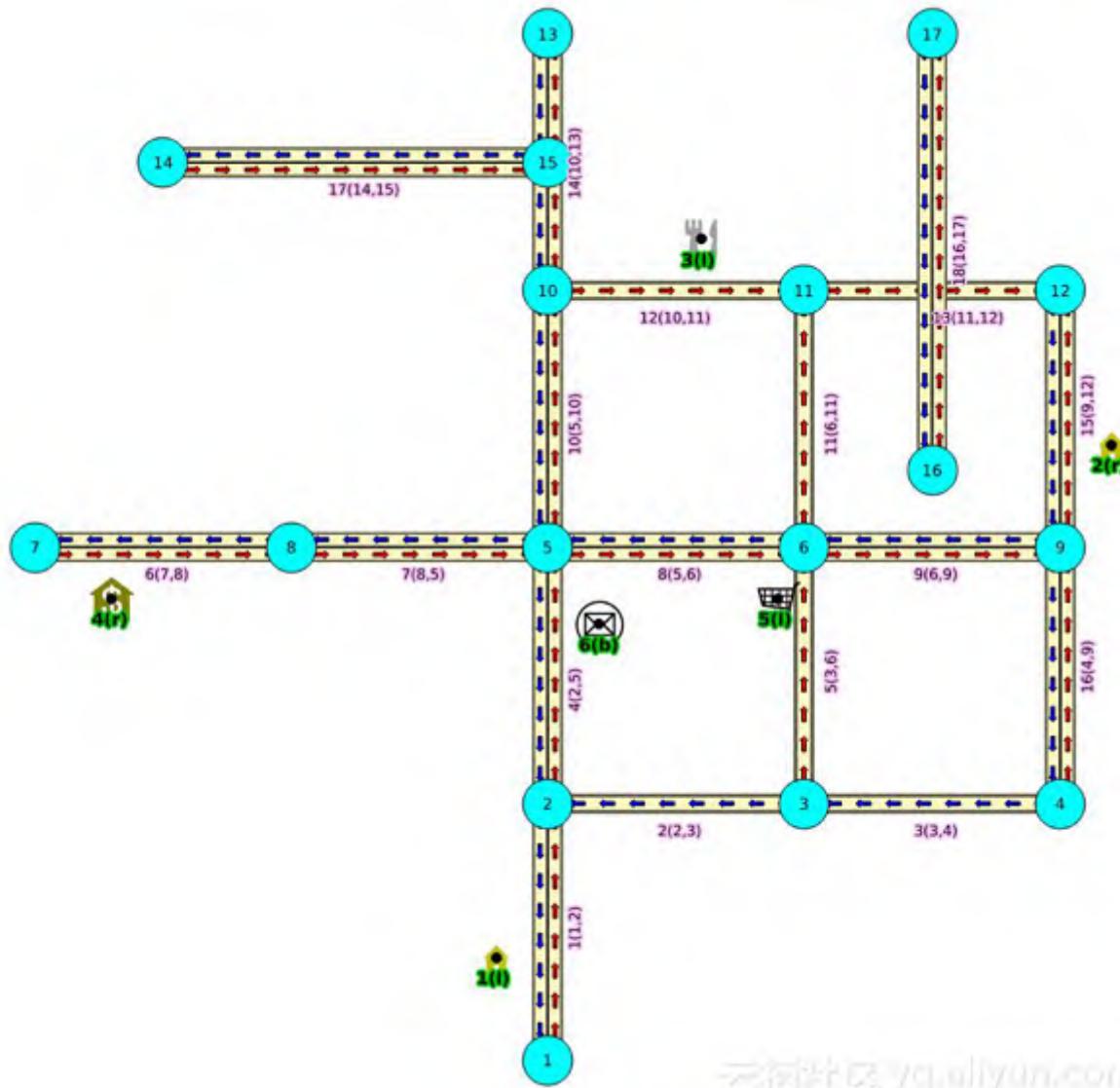
而3D扫描其实在军用领域很早以前就有了。



40%

## 动态路径规划

物流配送、  
打车软件、  
导航软件、  
出行软件、  
高速、  
高铁

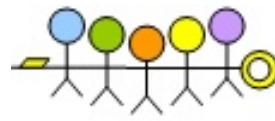


40%

- 五、C嵌入式开发(金融)
  - 与Oracle的Pro\*C功能对齐



ECPG

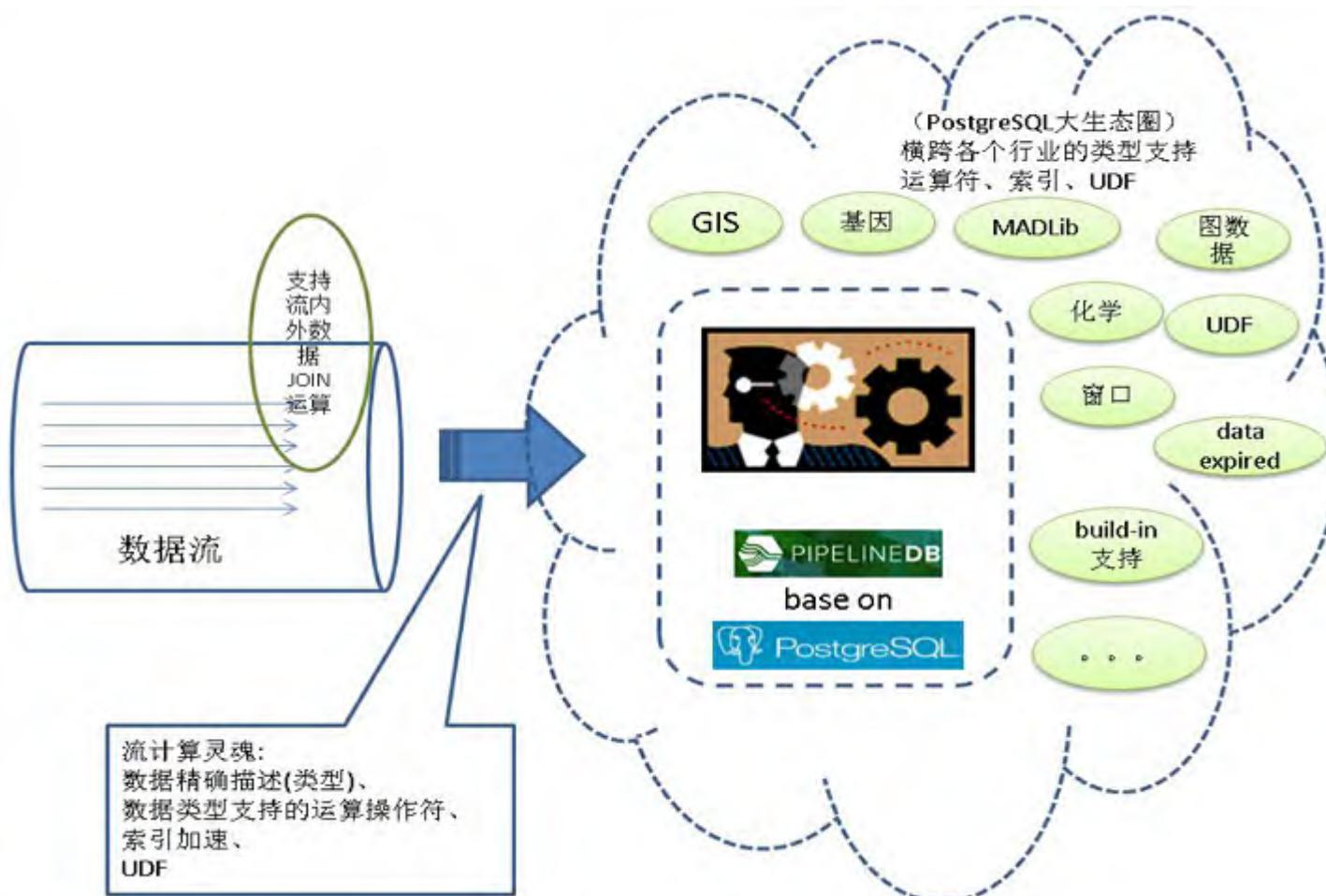


60%



60%

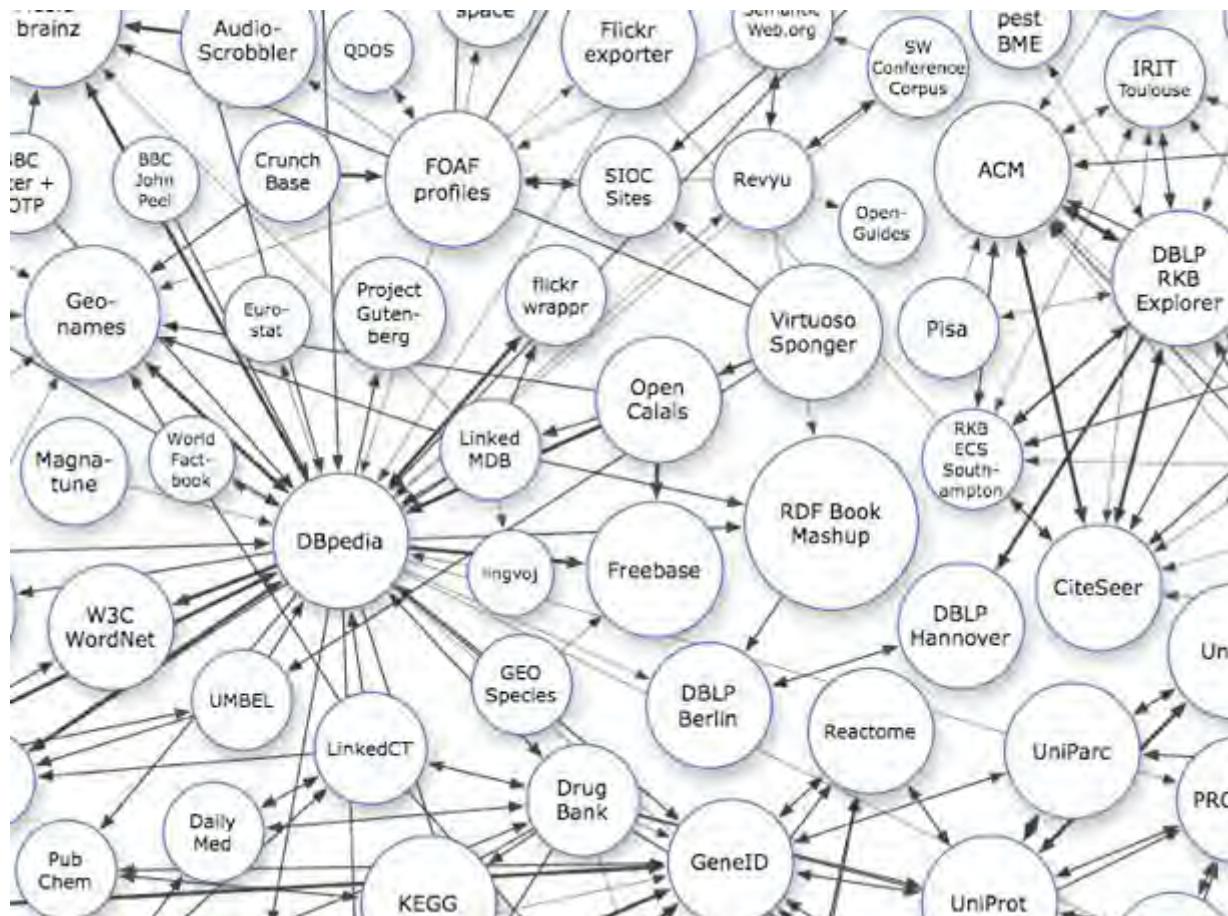
- 一、流式处理



60%

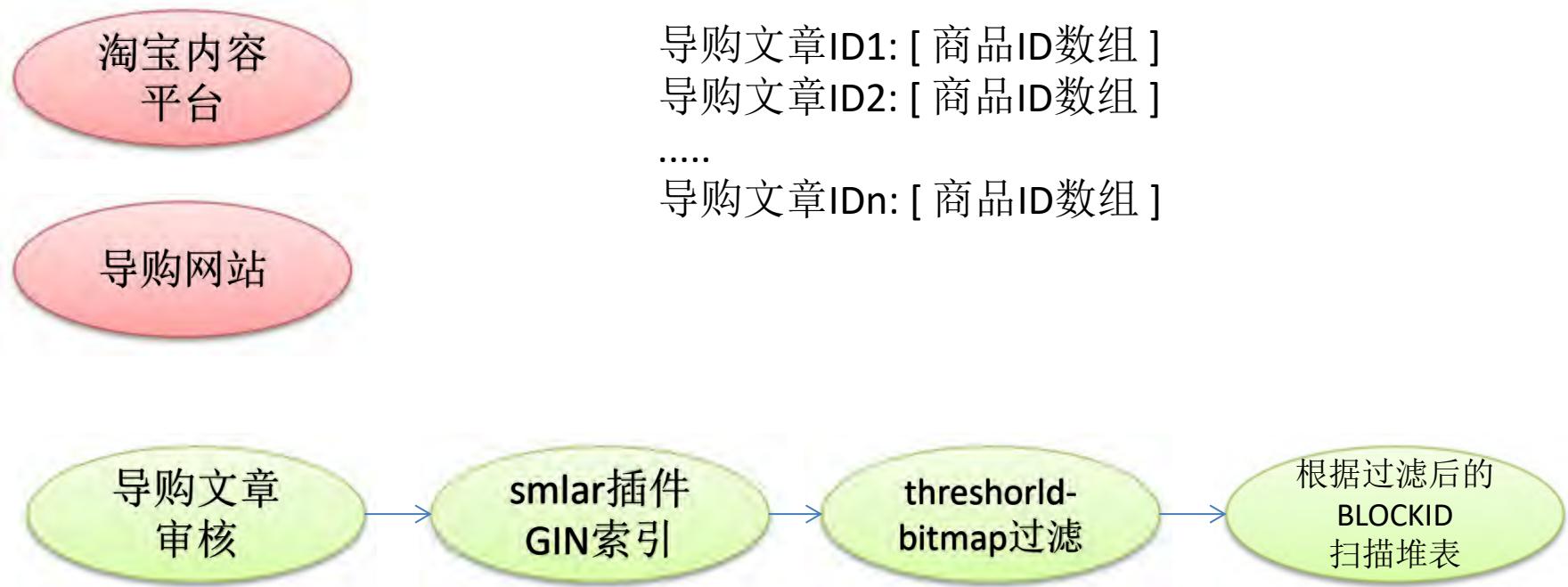
- 二、网状关系搜索应用、金融风控、公安刑侦、社会关系、人脉分析

最短路径  
N度人脉



60%

- 三、相似内容判定

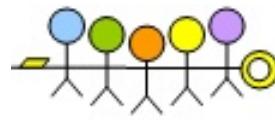


性能：  
亿级文章，亿级商品，毫秒响应

60%

- 三、相似内容判定

<del>Heap ID</del> <del>key page</del>	1	35	44	.....	101	109	1000
1	1	0	0	:	1	0	0
3	0	0	1	:	0	1	1
101	0	0	1	:	1	0	0
198	0	1	0	:	1	1	0
798	1	0	0	:	0	0	1
10009	0	0	1	:	0	0	0



60%

- 四、图像识别、AR红包

图片相似度判断

地理位置处理

Input image ( $2^N \times 2^N$ )



Transformed image ( $2^N \times 2^N$ )



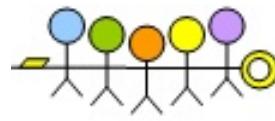
wavelet  
transform



87,55% similarity

60%





60%

- 五、real sharding（无限制sharding）

pushdown

跨节点  
JOIN

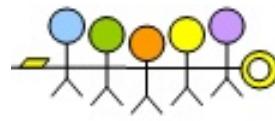
跨库事务

(agg\where\sort\JOIN\group by\  
insert\update\delete\select)

扩容、缩容

一致性备  
份、恢复

管理便捷



# 60%

- 六、资源控制

DaaS

多租户

一库多业  
务

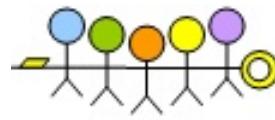
用户级

库级

内存、  
IOPS、CPU

HOOK

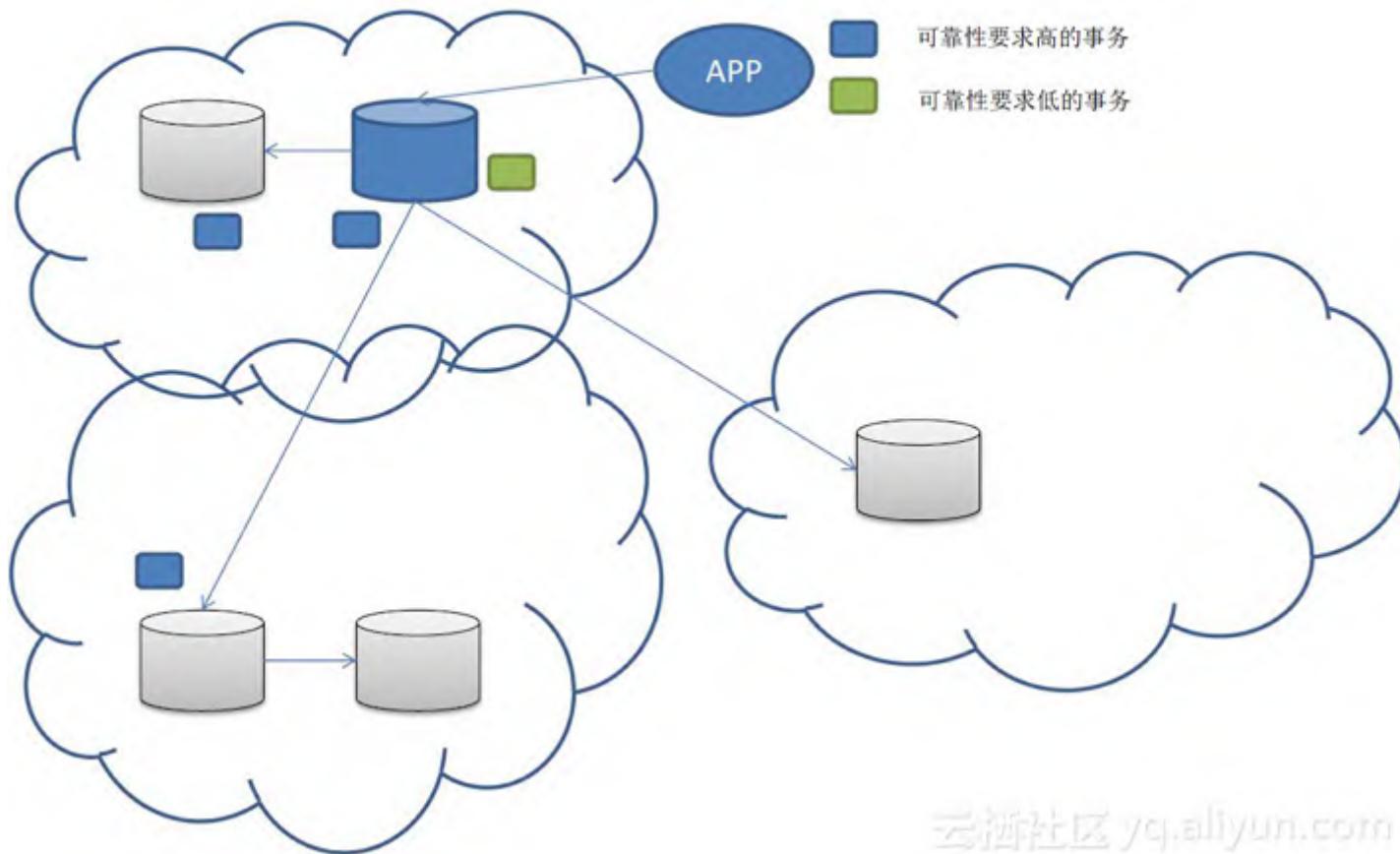
cgroup\  
backend进程

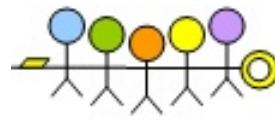


60%

- 七、任意多副本、事务级弹性选择

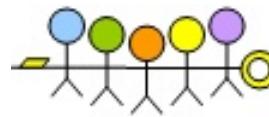
## 金融级 - 多副本可靠性





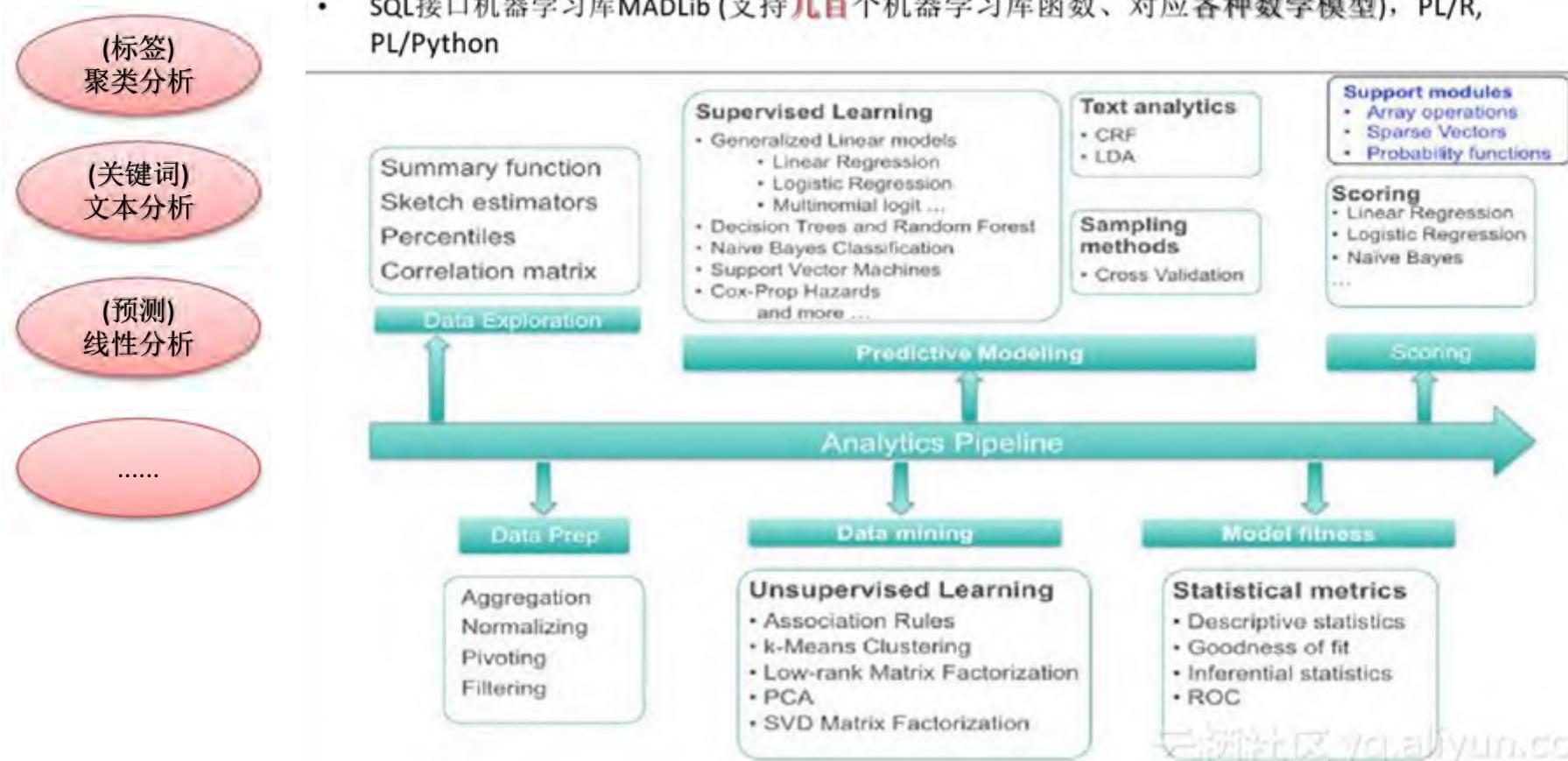
80%





80%

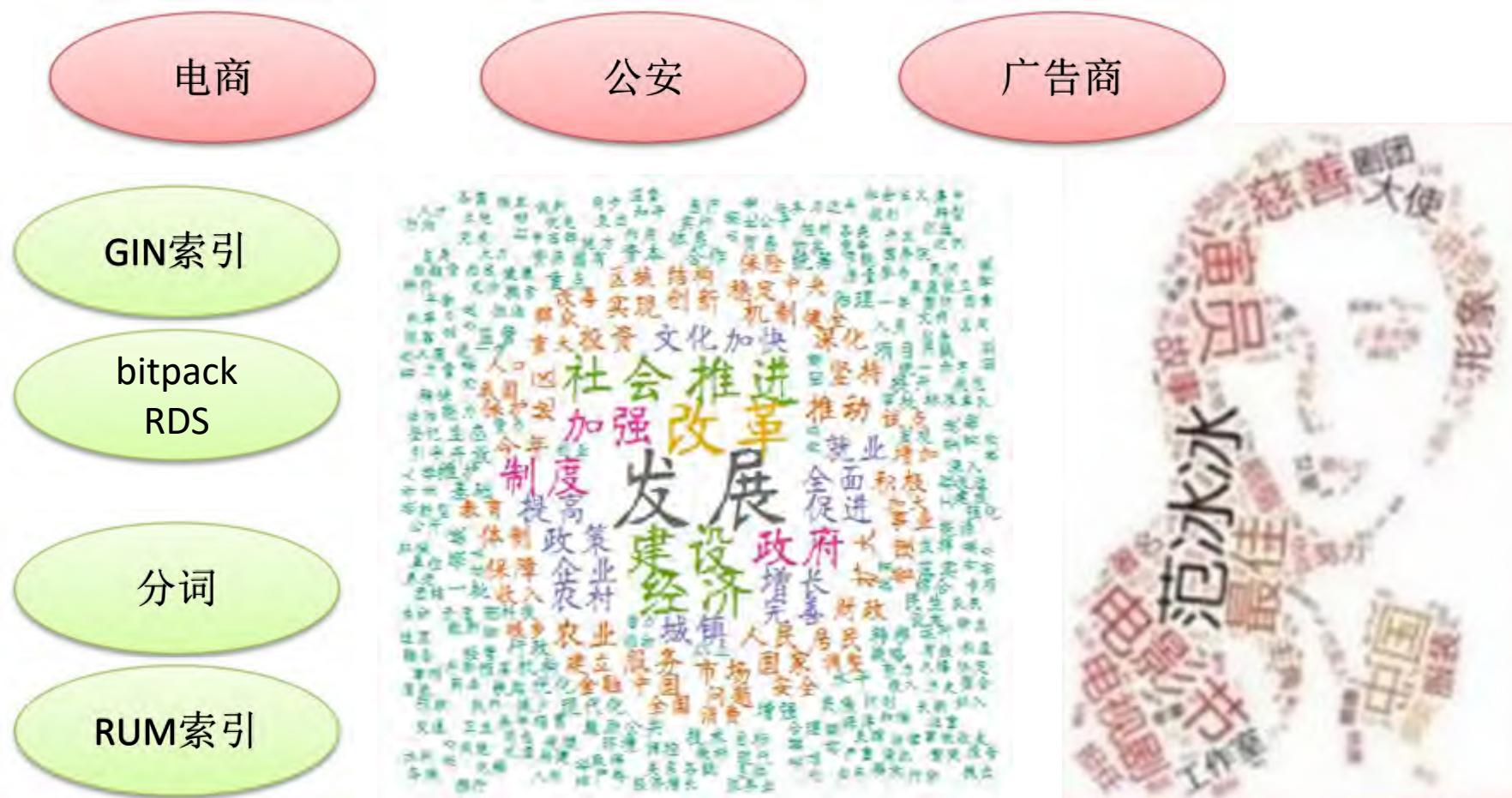
- 一、机器学习库(MADlib)





80%

- 二、实时用户画像、按标签圈人；文本关键词分析、相似文本分析

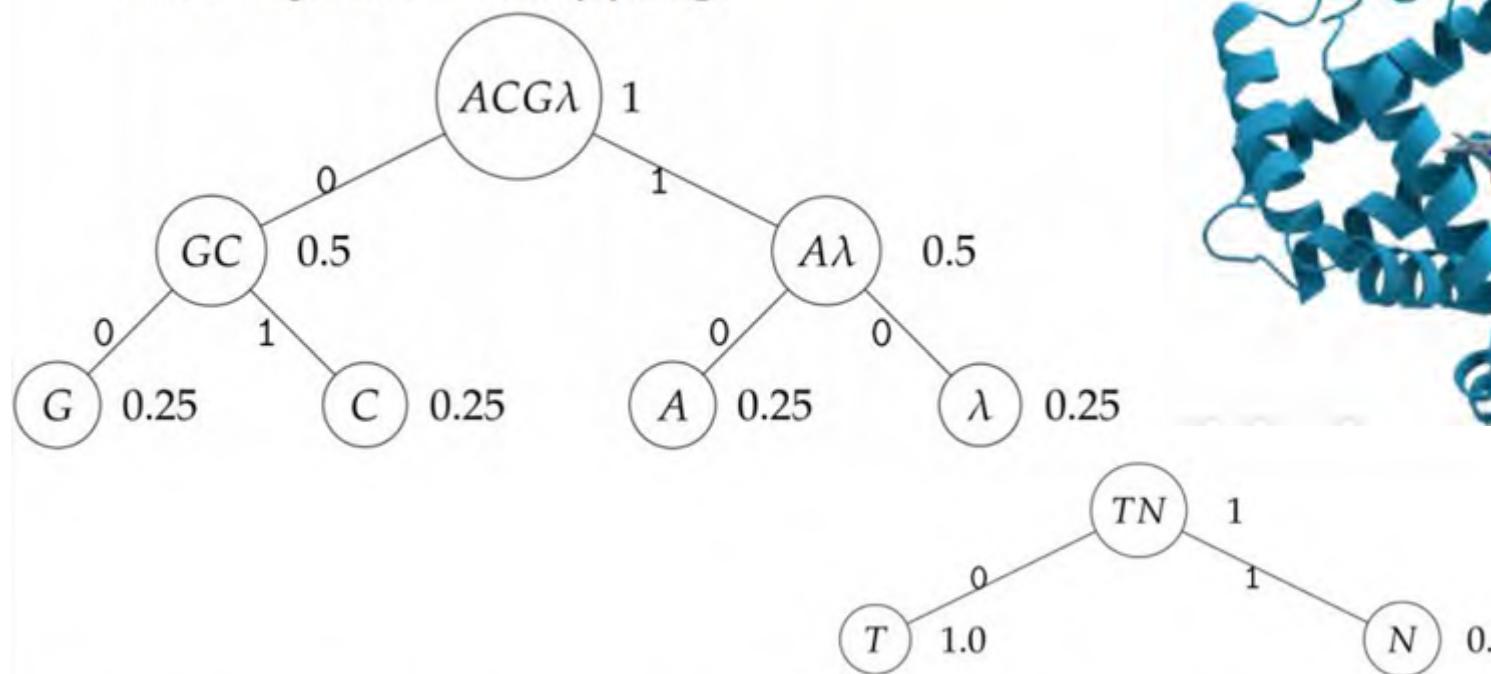


80%

- 三、基因数据处理(存、取、检索；类型、索引、操作符、函数、UDF)

## How does PostBIS compression work?

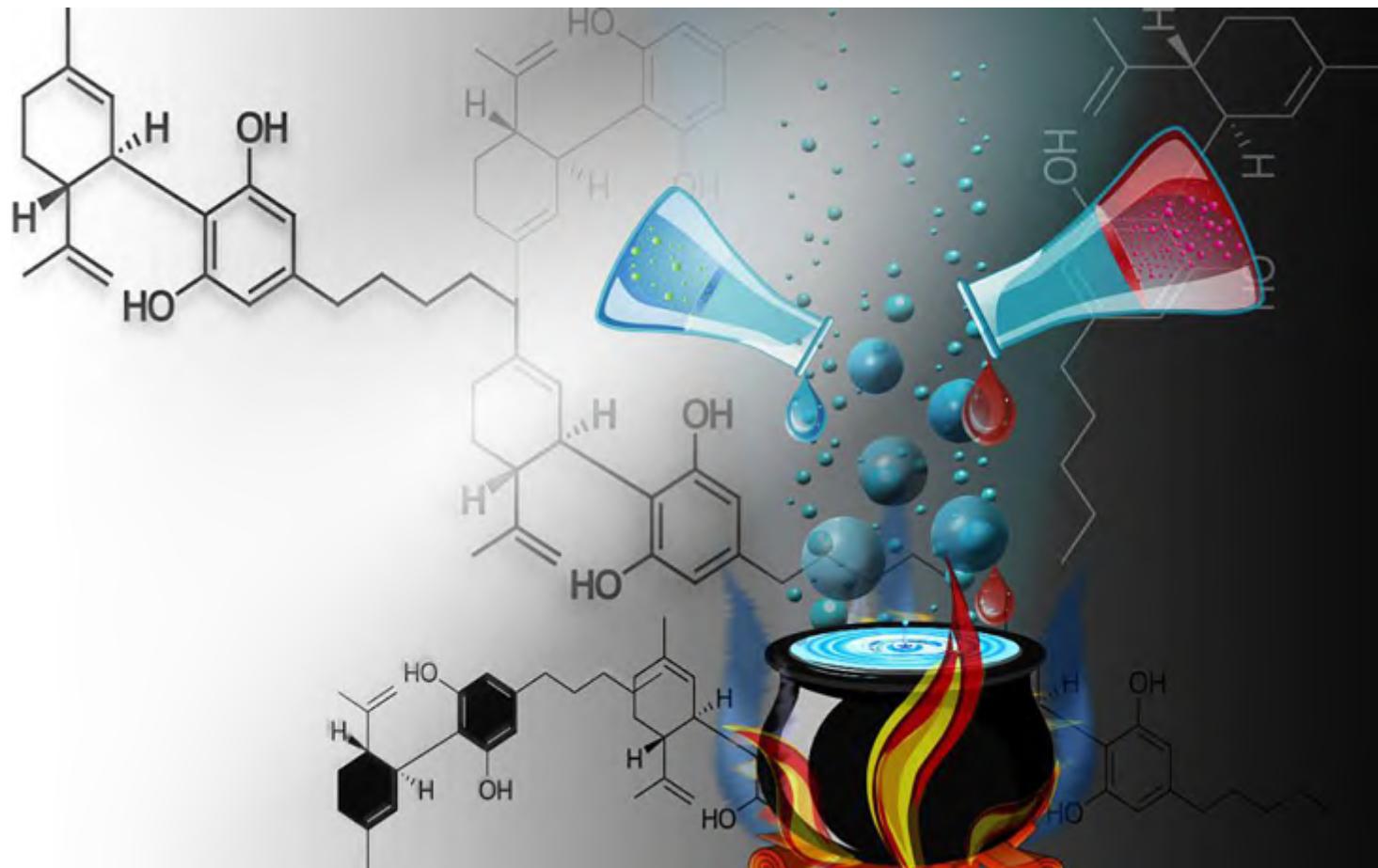
### Rare Symbol Swapping



- Lower Limit of Redundancy = 0.000003815

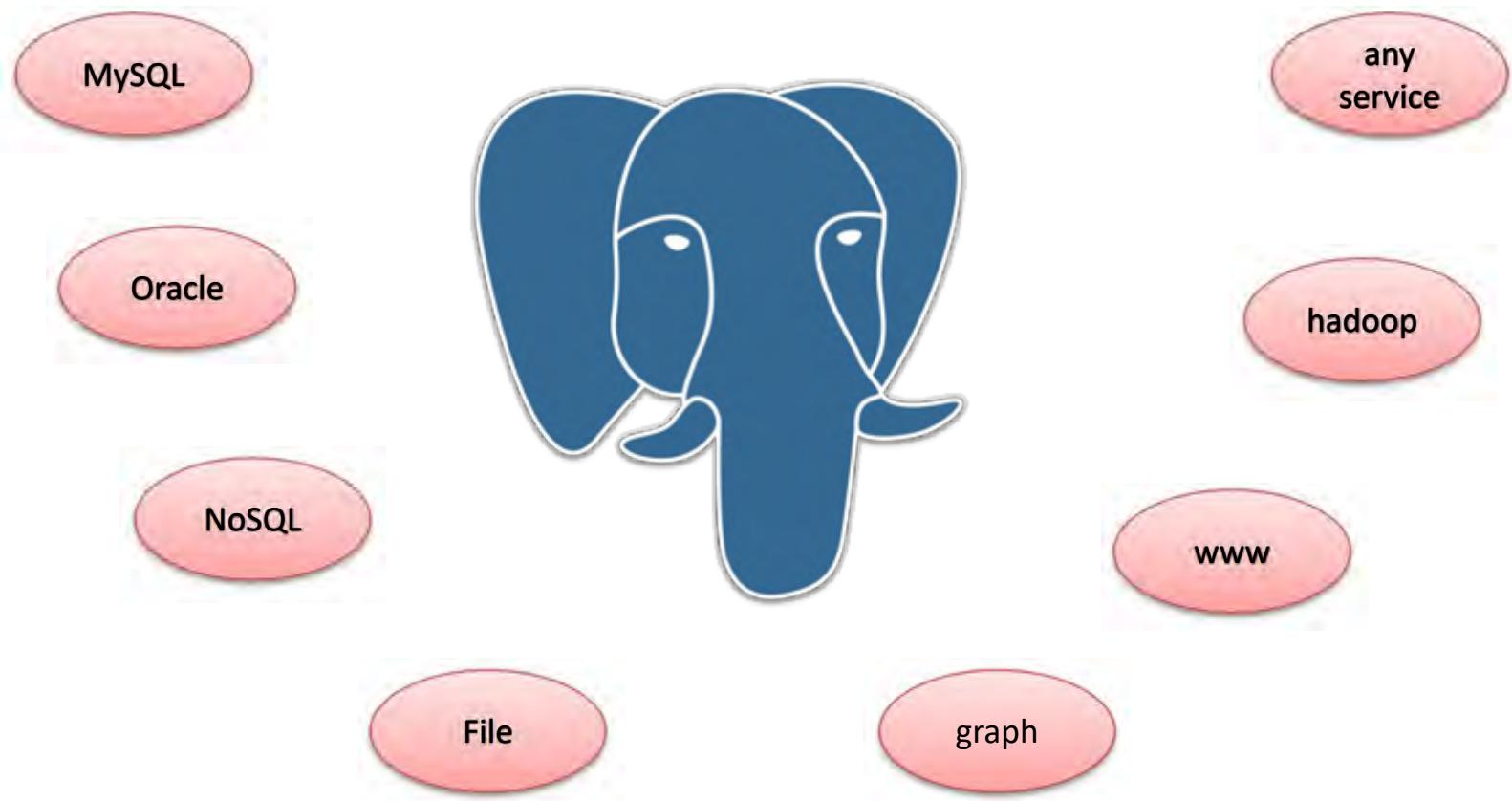
80%

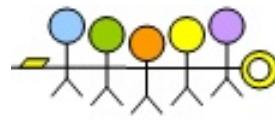
- 四、化学数据处理(存、取、检索；类型、索引、操作符、函数、UDF)



80%

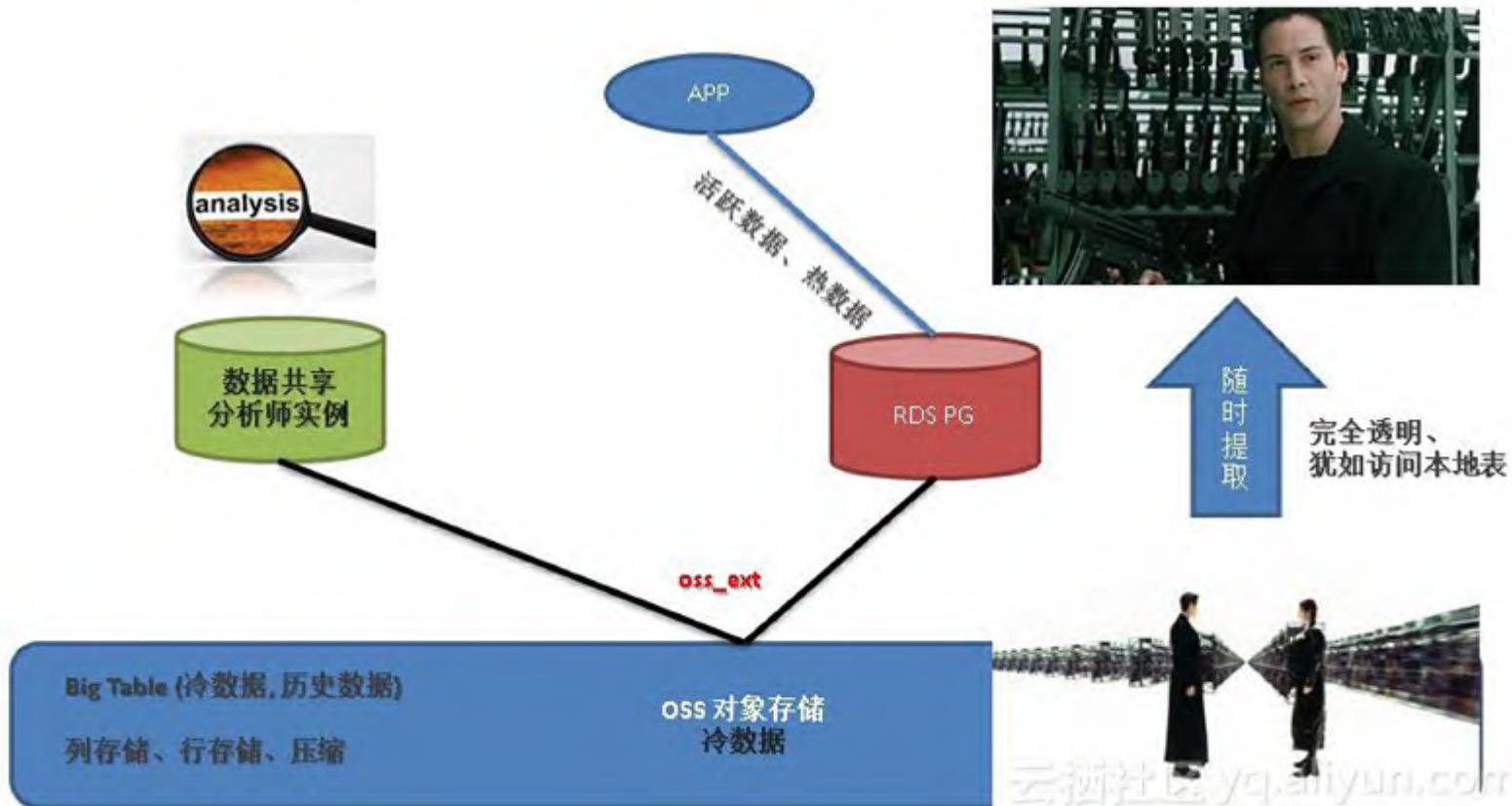
- 五、融合





# 80%

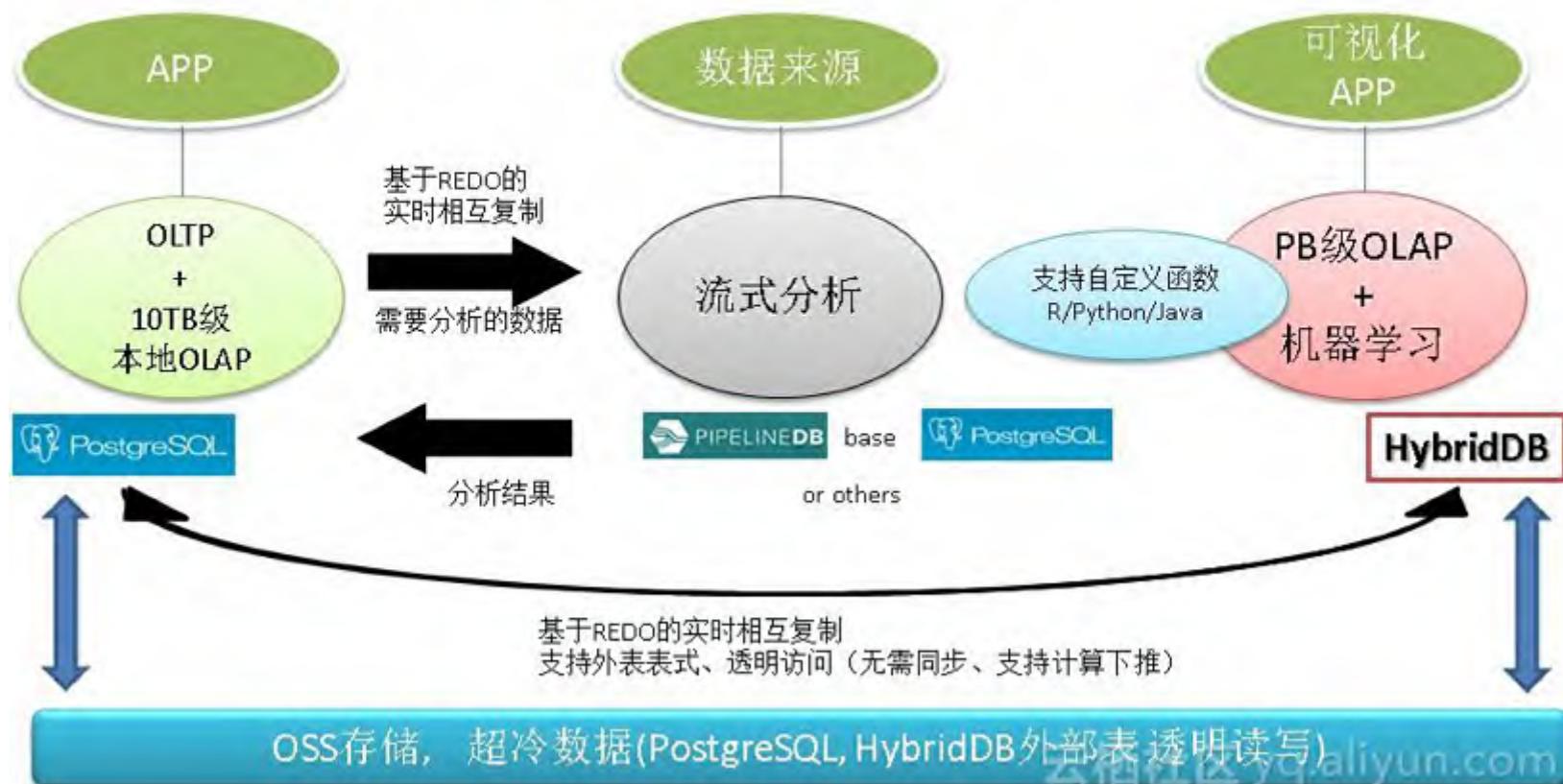
- 五、融合
  - 一份数据，分析、生产两不误



# 80%

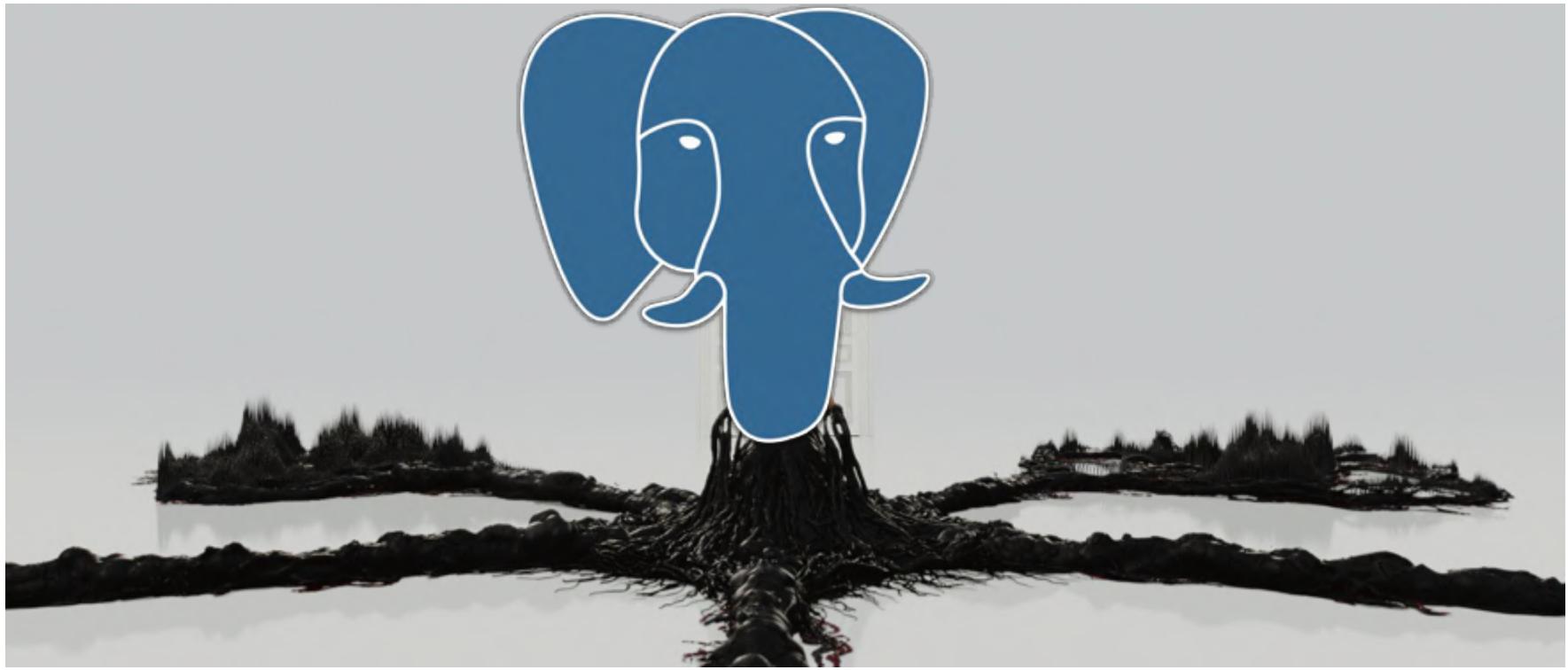
## 五、融合

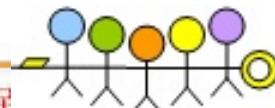
- PostgreSQL+HybridDB
- 一站式“**OLTP+流式分析+OLAP+机器学习+可视化**”解决方案



# 100%

- 一切皆可扩展





JSON

HINT

GIS数据处理

多维分析

FDW数据

异步消息

音频处理

立体几何

平面几何

基因处理

图数据处理

分词

旋转门压缩

数据走势预测

正则走索引

下棋

估值计算

探索宇宙

3D打印

GPU并行计算

流式计算

机器学习

文本挖掘

对接ES\Kafka\...

MR

LBS应用

一切皆可扩展 - 取自最终用户开源项目(**pgxn, github, pgfoundry, sourceforge, 社区.....**)



历经**43年**进化 - 相当成熟的底盘技术

开放式接口

AGG\WINDOW

HOOK language handler

FDW handler scan handler

类型扩展(IO,OP,AM,FUNC)

支持数十种流行编程语言  
编写服务端函数

SQL-2013 递归查询 MVCC 优化器 8种索引方法,20种大类

bt,hash,gin,brin,rum

ECPG 窗口查询 HASH JOIN 秒杀

bloom,gist,sp-gist

SQL兼容性

可靠性

REDO

流式复制

多副本

块级增量备份&PITR

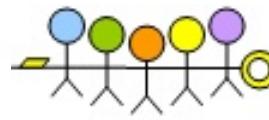
扩展性

CPU并行计算

读写分离

水平分库

多主同步



# PostgreSQL - 可编程数据库

传统数据库

能存取固定类型

能用固定模式处理固定类型

传统数据库  
眼里

这是别墅



可编程数据库

能存取固定类型

能用固定模式处理固定类型

**能新增自定义类型**

**能新增自定义数据处理和访问方法**

可编程数据库  
眼里

这是别墅  
同时允许扩展其用途  
迪斯尼  
农场  
赌场  
KTV  
银行  
超级市场  
医院  
.....

# 继Ingres之后又一个改变世界的产品

- OLAP
  - Greenplum、AsterData、matrixDB、Paraclle、Redshift、Illustra, Informix, Netezza、某些国产数据库。 . . .
- OLTP
  - EDB、国家电网xxDB、某些国产数据库
- 流式数据库
  - pipelineDB
- NewSQL
  - Postgres-XL, Postgres-XC, Postgres-XZ, . . .



[http://amturing.acm.org/stonebraker\\_1172121.pdf](http://amturing.acm.org/stonebraker_1172121.pdf)  
[http://amturing.acm.org/award\\_winners/stonebraker\\_1172121.cfm](http://amturing.acm.org/award_winners/stonebraker_1172121.cfm)  
[http://amturing.acm.org/vp/stonebraker\\_1172121.cfm](http://amturing.acm.org/vp/stonebraker_1172121.cfm)