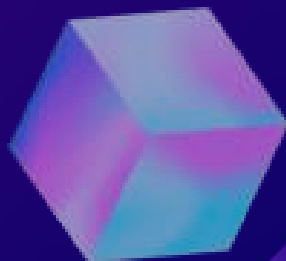




PostgreSQL中文社区

第13届PostgreSQL中国技术大会

—— 聚焦云端创新 汇聚智慧共享 ——





从Greenplum源码归档，看 国产数据库发展机遇

演讲人：熊灿灿





contents 目录

01 Greenplum与PG

02 Greenplum归档

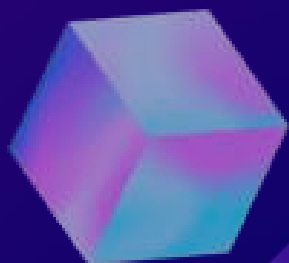
03 以简驭繁，超融合

04 Greenplum迁移



01

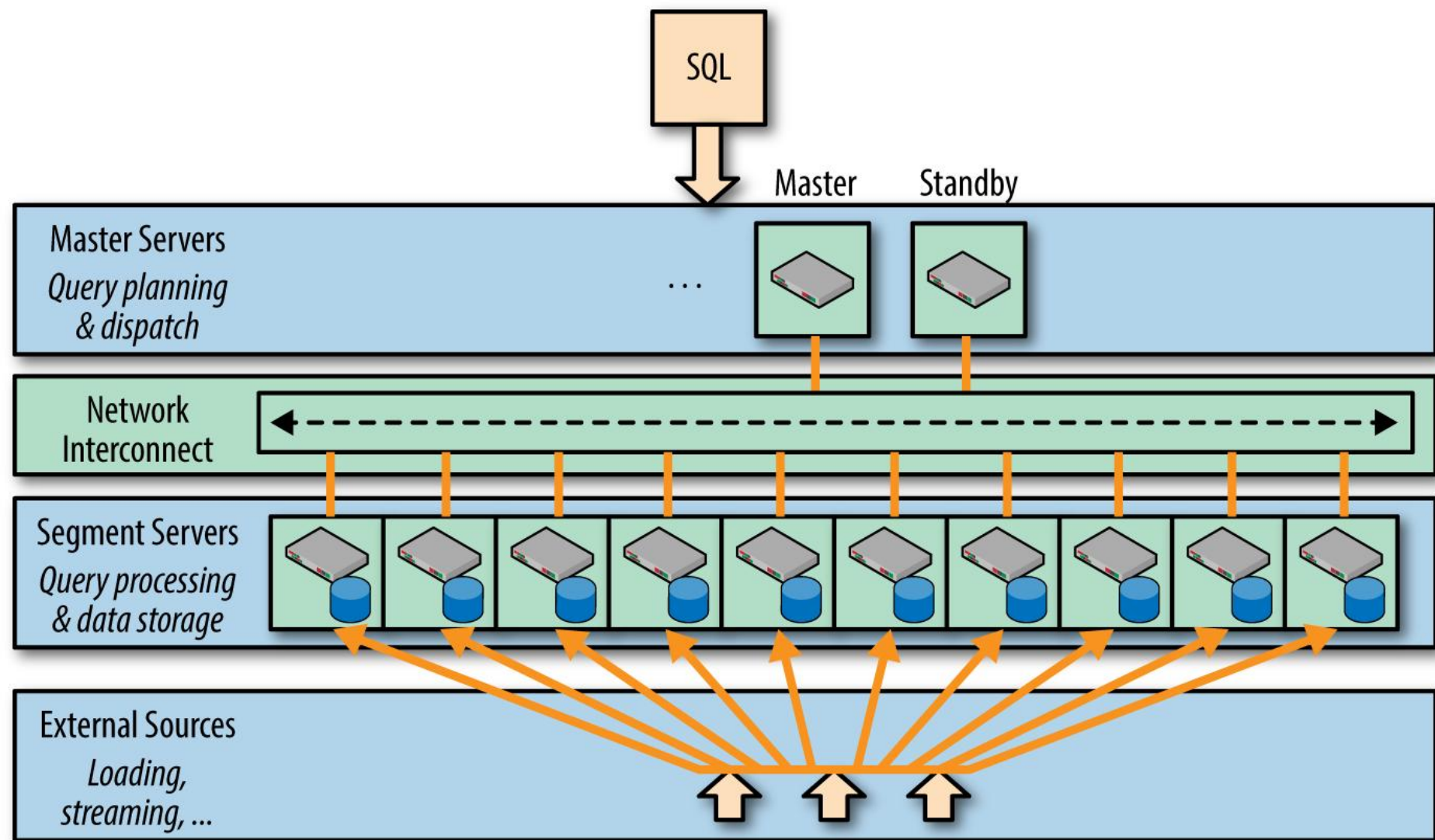
Greenplum与PostgreSQL



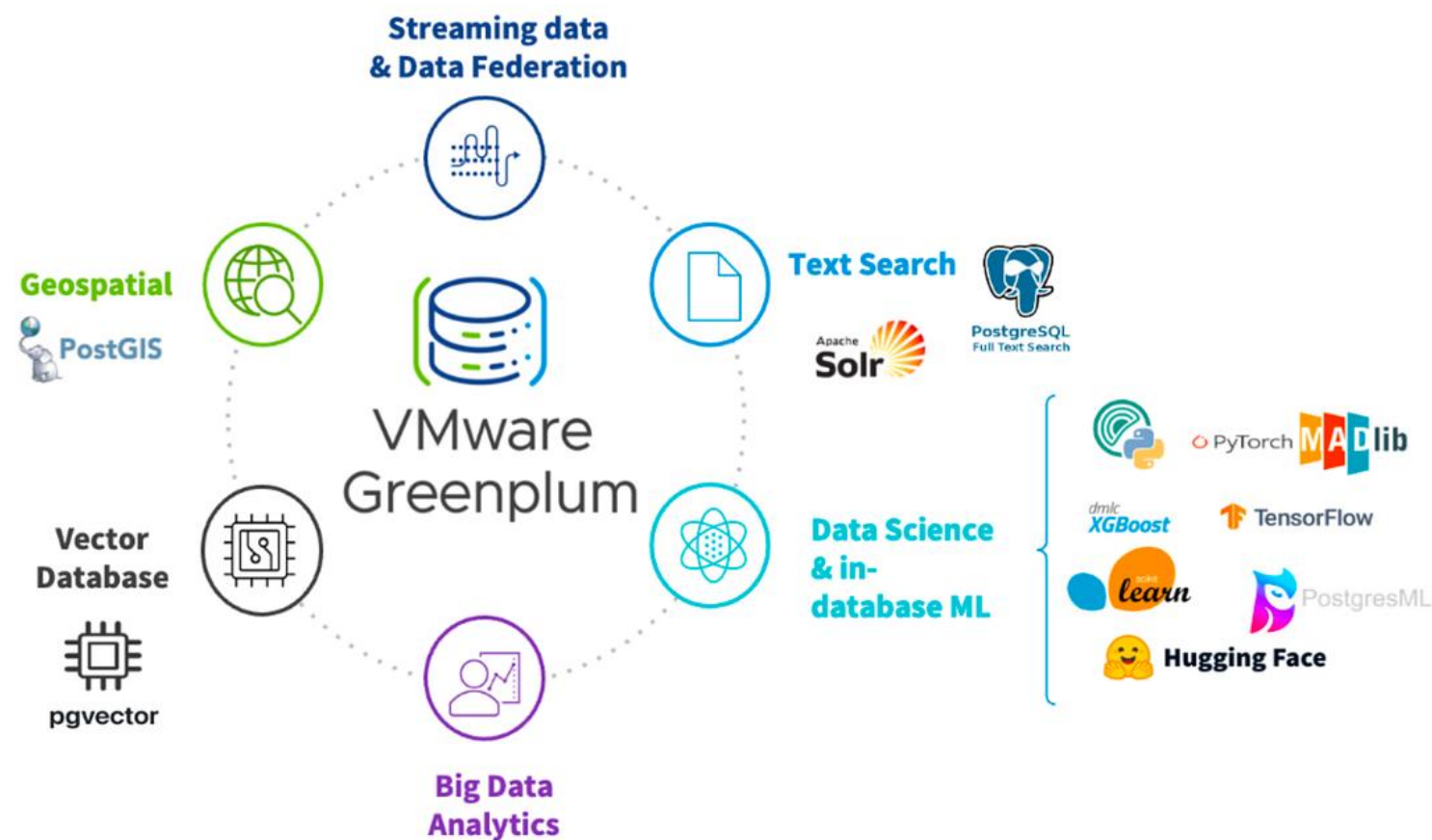
Greenplum, 常青不凋



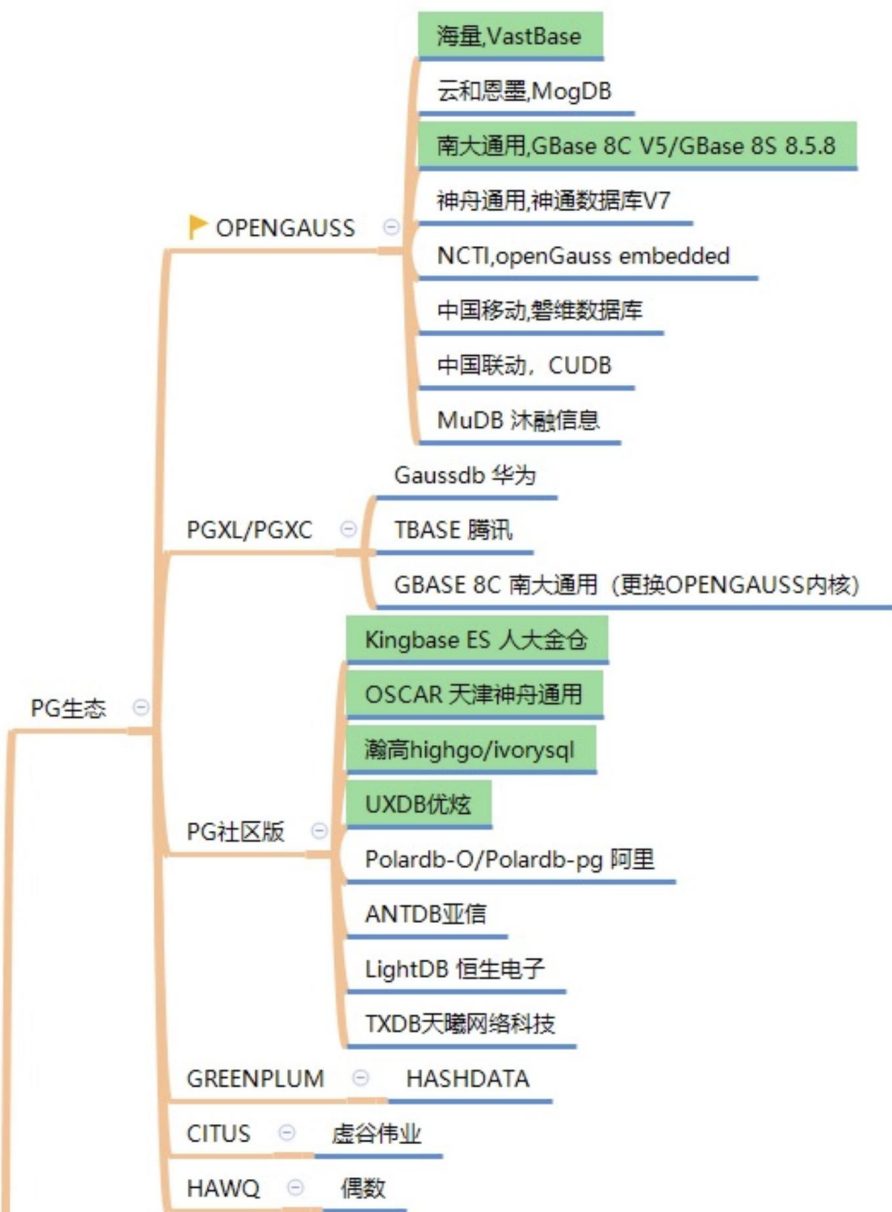
PostgreSQL中文社区



Greenplum, 全栈



Greenplum系，枝繁叶茂



北京酷克	2016	HashData	greenplum	分布式
北京人大金仓	1999	KADB	greenplum	分布式
山东瀚高	2005	HGDW	greenplum	分布式
阿里	1999	AnalyticDBforPG	greenplum	分布式
恒生电子	2021	LightDB-AP	greenplum	分布式
贵州易鲸捷	2015	QianBaseMPP	greenplum	分布式
新华三	1983	SeaSQLDWS	greenplum	分布式
杭州拓数派	2021	PieCloudDB	greenplum	分布式
百度智能云	2012	FusionDB	greenplum	分布式
中科曙光	2006	DragonDB	greenplum	分布式
东软	1991	东软思来得	greenplum	分布式
广州图灵科技	2014	羚羊数据库	greenplum	分布式
建信金科	2018	龙跃MPP	greenplum	分布式
山东符木数据科技	2017	符木FUMUDB	greenplum	分布式

Greenplum系，枝繁叶茂



PostgreSQL中文社区

 OUSHU

 HashData

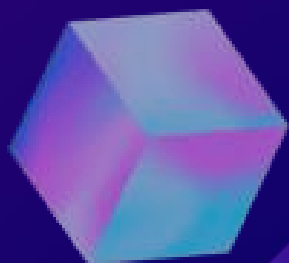
 MATRIX

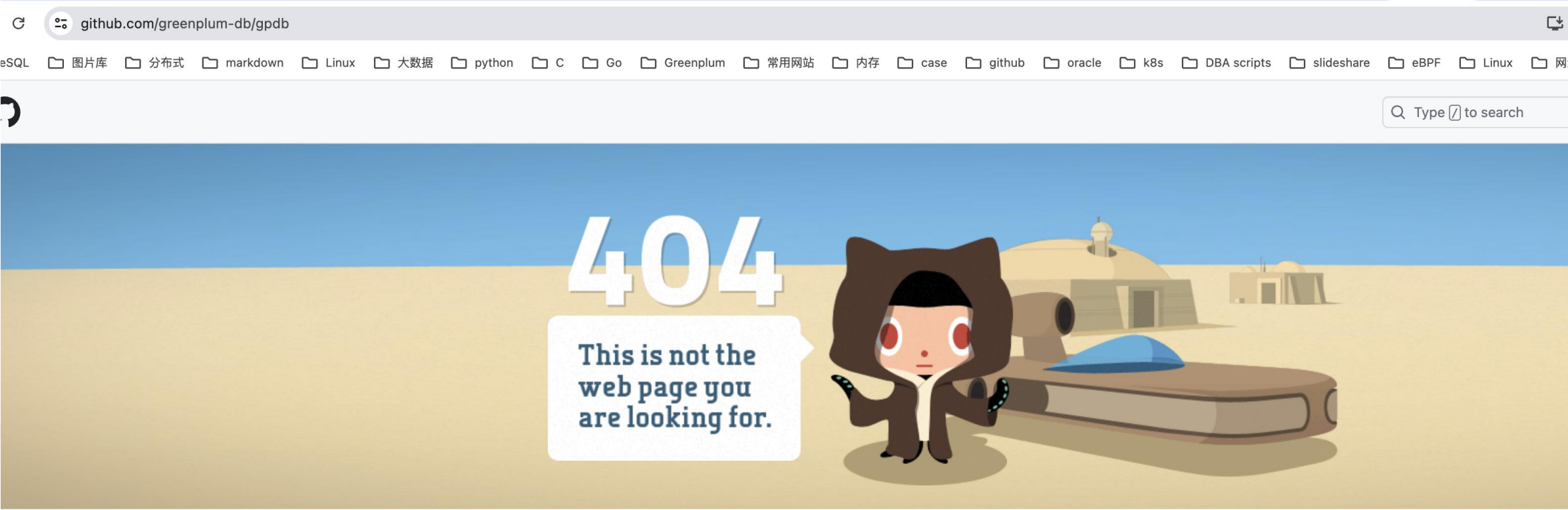
 π CloudDB



02

Greenplum归档事件





Find code, projects, and people on GitHub:

Search

greenplum-db / gpdb-archive

de Pull requests Actions Projects Security Insights

This repository has been archived by the owner on May 25, 2024. It is now read-only.

gpdb-archivePublic archive

Watch1Fork1.8kStar36

main1 Branch0 Tags

Go to fileCode

Annu149Fix compiler warnings482967c · 2 months ago69,508 Commits

.abi-check	Remove residual traces of deleted GUCs	2 months ago
.github	[main] Add support for checking ABI compatibility (#16763)	8 months ago
concourse	Add trigger times for main rocky9 pipeline	2 months ago
config	Bump minimum Perl version to 5.14	6 months ago
contrib	Optimize MPP FDW LIMIT/OFFSET push down when there...	3 months ago
doc	pg_waldump: Added option --last-valid-walname	2 months ago
gpAux	Revert "Revert "Update to python3.11 for plpython (#1713..."	4 months ago
gpMgmt	Introduce "gp init cluster --clean" command	2 months ago
gpcontrib	Add update path for gp_toolkit to version "1.6"	2 months ago
gpdb-doc	[Orca] Update doc about Fallback to planner for queries ...	2 months ago
hooks	concourse git-hook: check for pipeline generation (#4428)	6 years ago
src	Fix compiler warnings	2 months ago
.clang-tidy	[clang-tidy] Add readability braces around statements (#1...	2 years ago
.dir-locals.el	Make Emacs perl-mode indent more like perltidy.	6 years ago
.editorconfig	Merging Orca .editorconfig into gpdb file	4 years ago
.git-blame-ignore-revs	Merge with PostgreSQL 12.12	last year

About

Greenplum Database - Massively Parallel PostgreSQL for Analytics. An open-source massively parallel data platform for analytics, machine learning and AI.

[greenplum.org](#)

ReadmeApache-2.0 licenseActivityCustom properties36 stars1 watching1.8k forksReport repository


Releases

No releases published

Packages

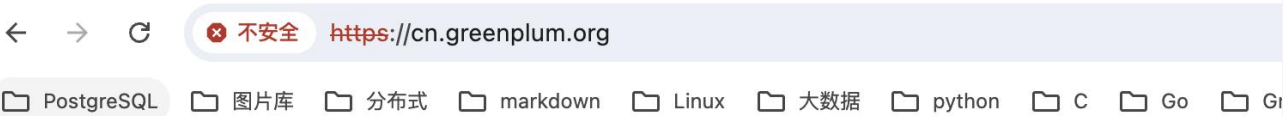
No packages published

Contributors383



Greenplum 源代码仓库的访问权限修改为了只读，同时还清空了原有的 branch、tag、pr、issue 等信息。

中文社区也人去楼空



Site is not available

The domain name in the URL is not associated with any active site on the WP Engine platform.


If this is your site, please visit your [User Portal](#) to ensure that your [domain mapping](#) and DNS records are correct.

Still have questions? [Contact Us](#).



可以通过以下途径访问原有资料：https://blog.csdn.net/gp_community、<https://space.bilibili.com/489184136/>



1. 如何更新？7.0 正式版发布不久，需要打磨
2. 如何升级？
3. 社区版用户，出了BUG怎么办？需自行解决，成本 
4. 出了问题找不到文档，中文官网 404

...

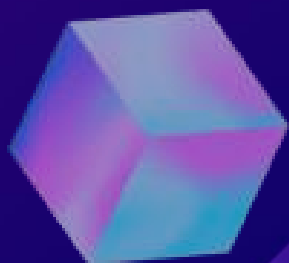
最终，

1. 比较依赖上游的国产数据库，将在后续竞争中逐渐退出
2. 拥有良好技术储备的团队，会获得更多机会
3. 新用户，会转向其他数据仓库项目，比如Doris、StarRocks等



03

以简驭繁，超融合数据库



基于多核数据库技术，一款数据库同时支持OLTP、OLAP、HTAP、时序等场景。

数据类型



结构化数据



半结构化数据
csv/text/json/xml



非结构化数据
空间/图片/音视频



流数据
IOT/传感器/车联网

MatrixGate
JDBC/ODBC
批流一体数据加载

ETL/CDC/OPC/MQTT/..

数据接入



高吞吐
低延迟



高并发连接



事务一致性

YMatrix超融合数据库

接口支持

JDBC

ODBC

HTTP

API

各种编程语言

...

可插拔优化器

Rule Based Optimizer

Cost Based Optimizer

库内海量数据计算 编程语言支持

Python



SQL



C++



R



...

多模存算引擎

执行器

存储器

microdb-oltp

执行器

存储器

microdb-olap

执行器

存储器

microdb-ts

执行器

存储器

microdb-vector

功能全面

用户认证

角色管理

安全审计

数据加密

集群监控

数据备份

集群恢复

集群高可用

客户场景



报表



商业智能 BI



机器学习/AI



数字孪生



数字应用

丰富的金融智能化场景价值体现

关键能力	描述	业务场景价值
1 多模数据支持	支持关系、时序、GIS、JSON、XML、自定义类型和相应的函数	自然表达金融各类资产、交易和对应标的物的属性、关系以及变化
2 高速写入能力	可支持每天单表百亿级别数据的毫秒级写入，支持乱序、自动合并等	交易系统 CDC、埋点和物联网收集数据都可以秒级延迟近实时入库
3 极速查询能力	标准SQL，极速执行宽表聚合、10+表复杂关联、CTE、窗口函数、数据立方体等	与业务系统、BI 和可视化工具无缝集成，即席分析体验和效率极佳
4 触发器和物化视图	分布式触发器和自动刷新物化视图，数据入库和变化时及时捕获并自动关联计算	可自动对流水等明细表进行各种维度汇聚，与产品表等关联时可以大幅降低计算量，捕获数据变化可用于上下游系统的协同
5 库内机器学习	支持 Python3 和数据科学库，单机 Python 算法可自动分布式化，计算贴近数据	数据科学家可用熟悉的 Jupyter 等工具与 Python工具库实现机器学习等算法
6 严格ACID	严格分布式事务保证，全局事务和数据节点局部事务协同	确保数据不错、不丢、不重，异常状况下有清晰的集群状态和简单的用户操作
7 高可用	内置高可用，主节点和数据节点故障自动恢复，不影响集群使用，内置告警	使用简便、运维无忧
8 大规模	经头部企业核心生产系统检验，单集群可到数百节点	随数据和业务增长，集群可线性扩展

行列混存

块内列存降低分析时文件读写开销，提升性能；整体按块组织解决纯列存超宽表写放大问题并提升点查性能。

块内数据有序

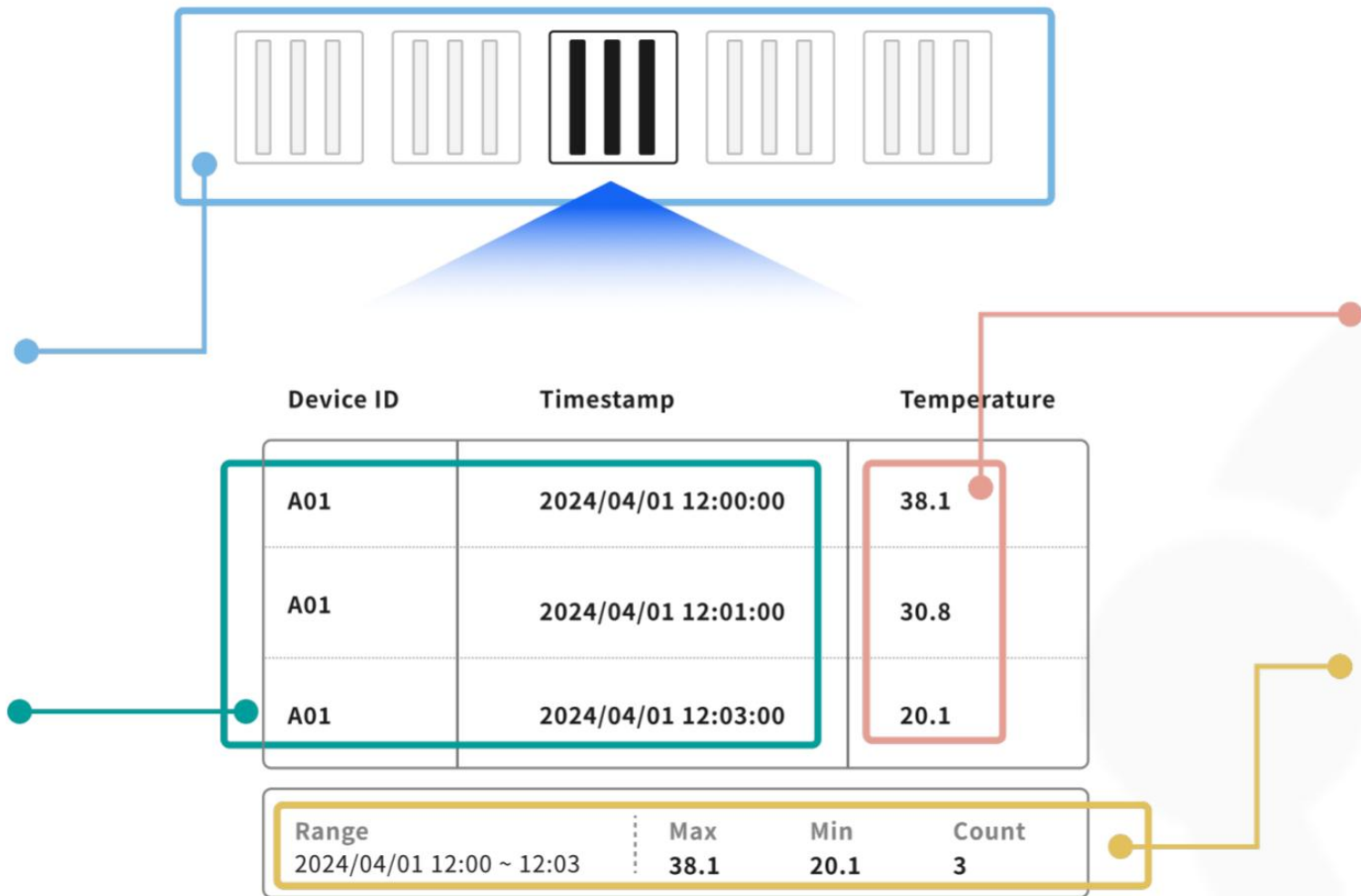
自动按指定字段联合排序，有序存储对点查、分析查询更有友好，提升性能。

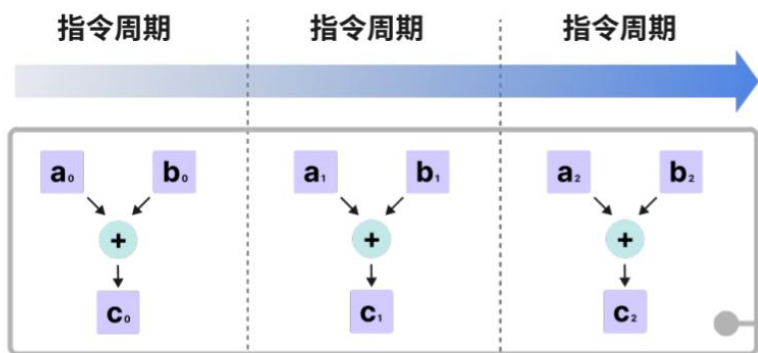
字段列式存储

显著降低大规模分析查询文件操作开销；可基于数据特征设置独立压缩算法，压缩比更高。

预聚集信息

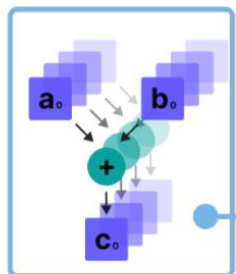
高效裁切，定位数据更智能迅捷；常用分析计算直接读取结果，极致加速。





标量计算引擎

一次计算一对数据，大数据量时需大量重复相同计算，性能较差。



向量化计算引擎

利用 SIMD 指令集一次计算一批数据，在大数据量场景下，可数量级的提升性能表现。

```
SELECT tag_id,
       COUNT("metrics"),
       AVG("metrics1"),
       SUM("metrics2"),
       MIN("metrics3"),
       MAX("metrics4")
FROM   prod_data,
WHERE  tag_id IN (950, 951, 1030, 1200, 10, 301, 900, 886)
      AND "dt" >= '2024-04-01 00:00:00'
      AND "dt" < '2024-04-01 12:00:00'
GROUP BY 1;

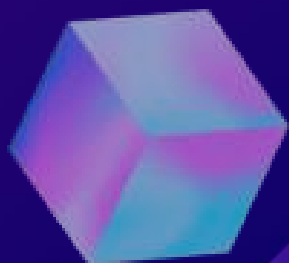
-- set matrix.enable_mxvector=off;
Time: 5088.684 ms

-- set matrix.enable_mxvector=on;
Time: 30.387 ms
```



04

Greenplum迁移替换





全量迁移

- 以数据库为单位，实现从GP 4.3.3X/5/6/YMatrix集群并行迁移数据表数据定义语言及数据到新YMatrix集群。。

条件迁移

- YMatrix支持通过WHERE条件指定数据条件和规则，用户可按需进行选择迁移，更加贴合实际场景。

增量迁移

- 以表为单位，实现从GP 5/6/YMatrix 集群并行迁移目标数据库中部分数据表的数据定义语言及数据到新的YMatrix 集群。

支持多种数据分布策略

- 复制表(REPLICATED)、哈希分布、随机分布、Master-only.....

1.设置数据源

2.设置目标库

3.设置迁移对象

4.设置迁移方式

5.一致性校验

6.校验效果确认





感谢聆听!

