



开源软件供应链点亮计划—开源之夏2023

格睿科技 Greptime 开源之夏介绍





做梦都想写开源项目

目录

CONTENT

- 01 关于格睿科技 Greptime
- 02 开源项目 GreptimeDB
- 03 “开源之夏” 项目介绍
- 04 致谢

01

关于格睿科技

关于格睿科技 Greptime

我们是谁

我们是一家年轻有活力的初创公司，刚刚成立一周年，团队在**北京**和**杭州**。三位联创平均有10多年互联网大厂工作经验，团队全员皆开发，组织架构扁平，团队氛围非常和谐轻松。

我们在做什么

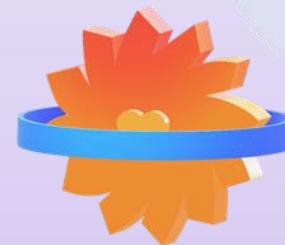
用 **Rust** 语言开发**时序数据库** GreptimeDB 和商业版 SaaS 应用 Greptime Cloud.



C GreptimeDB

| Rust 写的**云原生、分布式、融合时序和分析的开源时序数据库**

- 数据库，针对时间序列数据特别优化
- Rust 语言，注重性能和安全
- 云原生架构，算算分离，存算分离
- 融合分析，优化计算分析性能
- **GreptimeTeam/greptimedb**



Greptime 社区风采



开放、透明、尊重

2600+
star

140+
fork

20+
外部贡献者

42%
国内

58%
海外

GAIDC

2.25-2.26 团队在上海参加
2023全球人工智能开发者先
锋大会 (GAIDC)



Data
Infra

4.1 联创冯家纯做客 Data
Infra 研究社，漫谈时序数据
库的设计



分享嘉宾
冯家纯

Greptime 联合创始人，先后在阿里巴巴，蚂蚁集团工
作多年，长期从事中间件、时序数据库产品研发，曾负
责蚂蚁集团统一监控存储层，对时序存储相关技术有
丰富的实践经验

分享议题

漫谈时序数据库的设计

随着云原生可观测 (Observability) 的流行和 IoT 物
联网领域的兴起，时序数据的价值也越来越被重视起
来。时序数据库作为时序数据的存储底座，也受到了
众多的关注。此次分享将分析业界常见时序数据库的
技术选型、各自的优缺点，我们在 Rust 研发
GreptimeDB 时的设计抉择，以及最终的经验教训。

SOFA
Stack

4.19 团队参与 SOFAShark 五周年
活动



信通院
可信开源

4.21 信通院在开源生态建设论坛上为 GreptimeDB 颁布“可信开源社
区共同体”成员证书



02

开源项目 GreptimeDB

关于格睿科技 Greptime


什么是时序数据？

记录各时间点状态变化的详细过程

- 监控某个设备的温度，每隔几秒采集记录一次
- 股票每个时刻的价格
- 跟踪记录车辆的行驶轨迹
- 记录用户的点击行为，路径分析，获知用户的使用习惯、改善产品体验和挖掘商业价值等
- 记录用户的每次登录信息（时间、客户端类型、设备、地域等），获知登录频率、时长、常用设备、高峰期等使用情况，安全预判等。

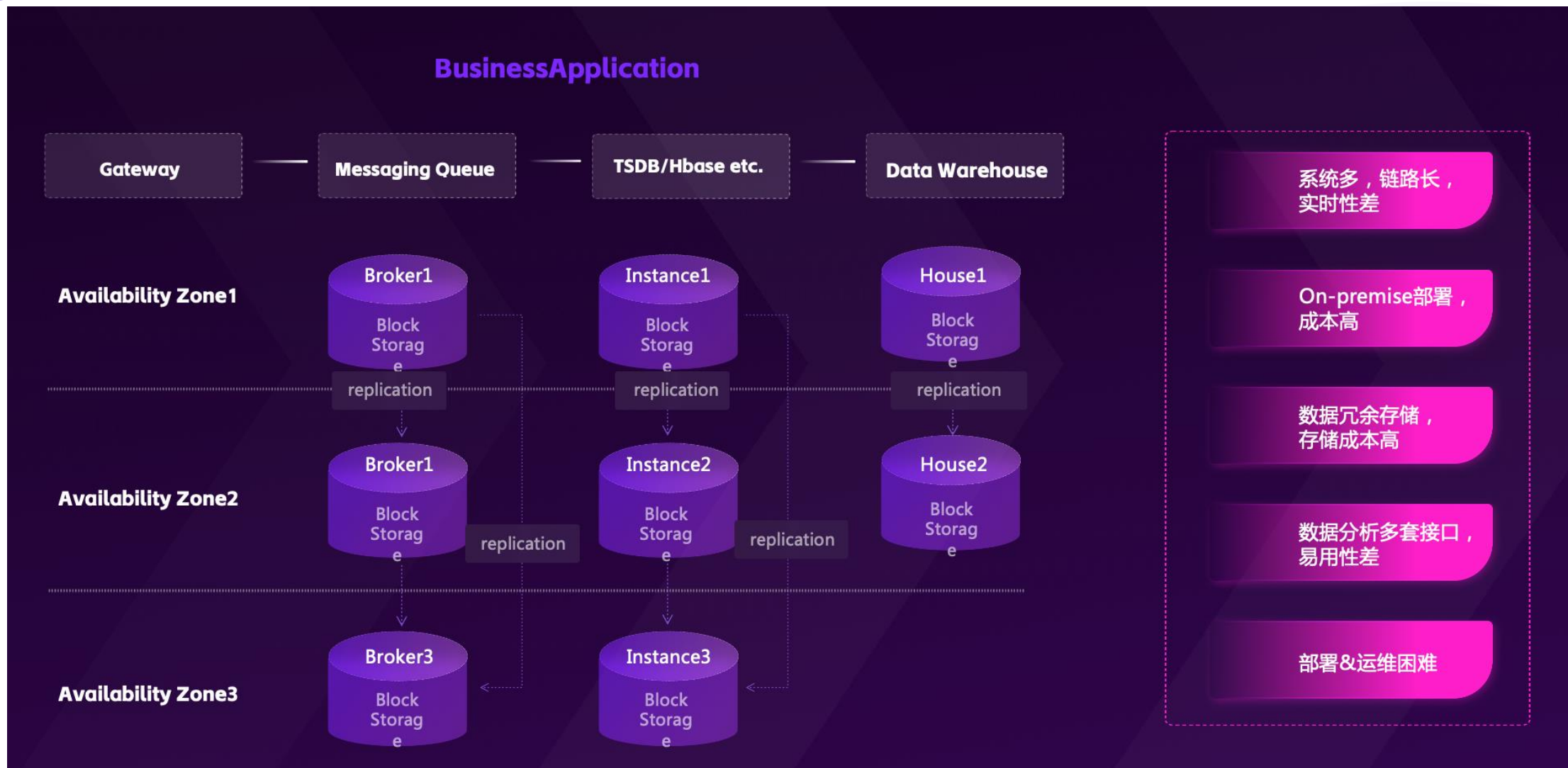
更高的数据价值：

- 实时测量
- 掌握详细过程
- 更深入地数据分析
- 预测未来

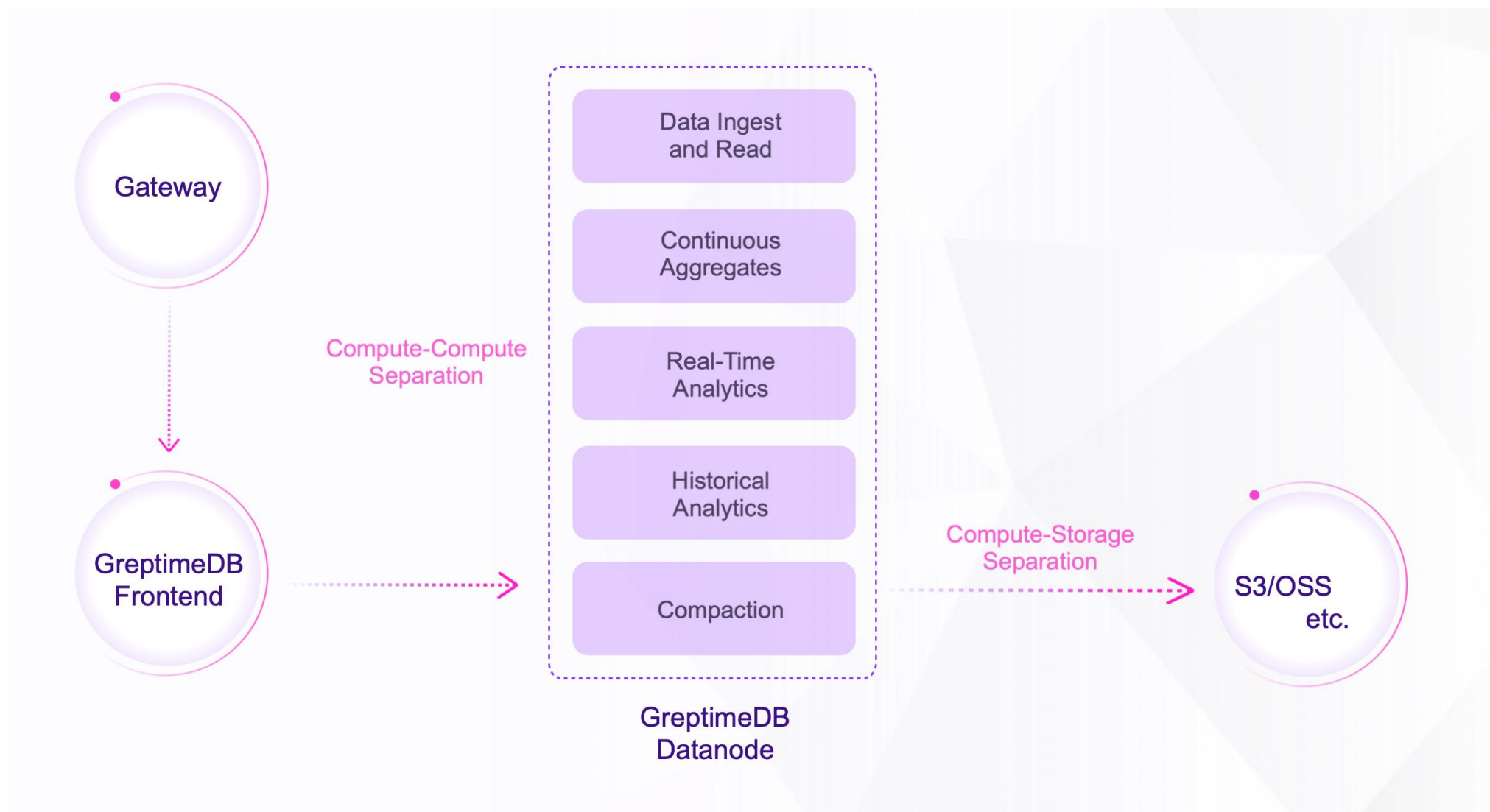


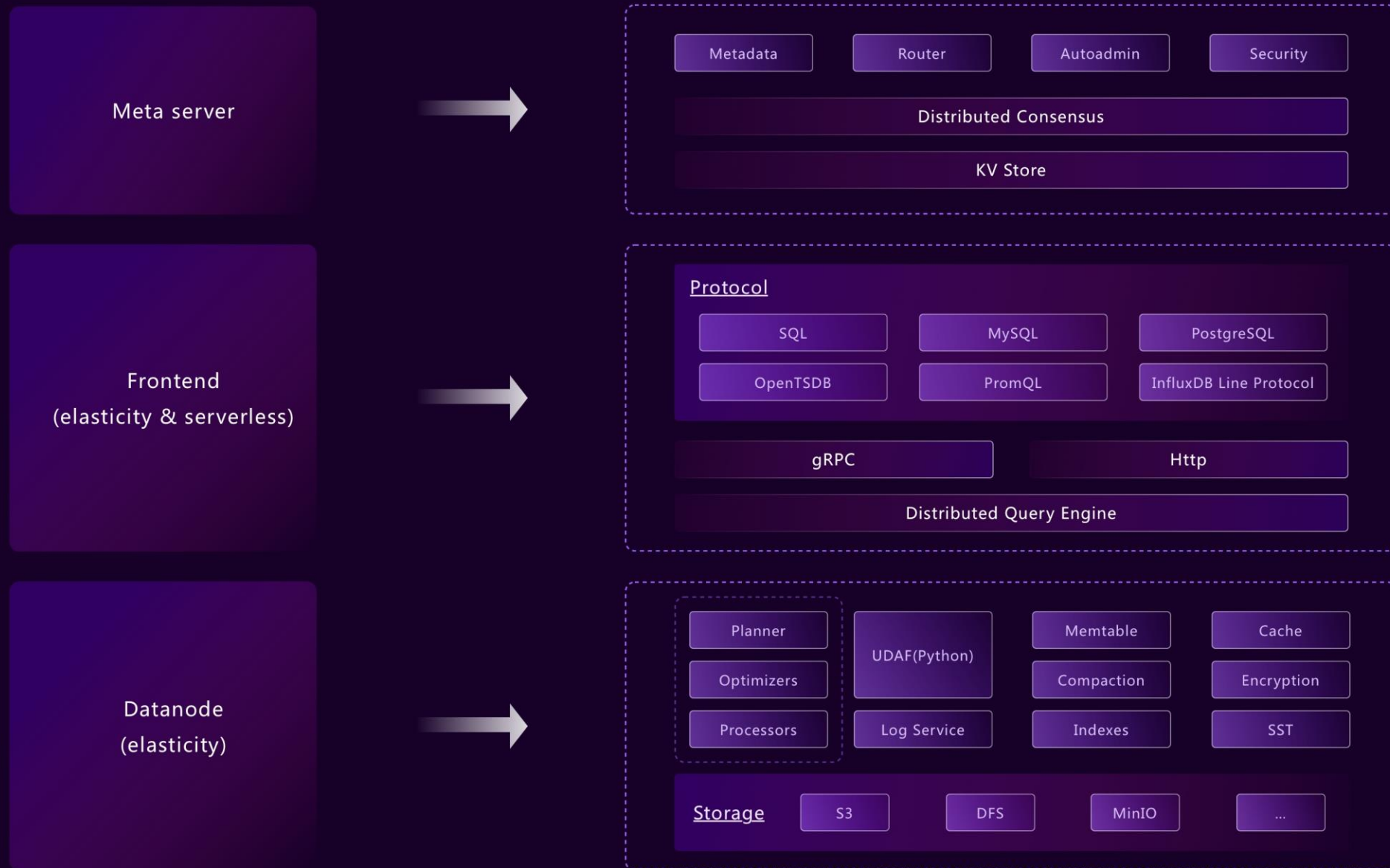
记录方式的不同，
插入，非更新

传统的架构



GreptimeDB架构





高可用 高可靠

使用透明

- 使用上与单机数据库一致，分布式对用户透明。

伸缩透明

- 自动化伸缩，增减节点无需人工干预数据，用户无感知。

灵活负载

- 写入、查询、分析、离线计算可分别路由到不同节点，分摊负载压力。

部署

- 可采用单点或集群部署。Frontend 和 Datanode 也可以部署在同一个进程内。单点部署无需Meta server。

性能

- 经验：每天万亿级别的指标点（每秒接近 1 亿数据点写入），数十万 QPS 的实时多维查询。
千万级别读写性能，写性能 1亿 数据点/秒，读性能 4000w 数据点/秒。

生态兼容

- MySQL、PostgreSQL、OpenTSDB、InfluxDB、Prometheus，用以上客户端均可写入和查询。
- 拥抱社区，易与现有生态数据栈集成

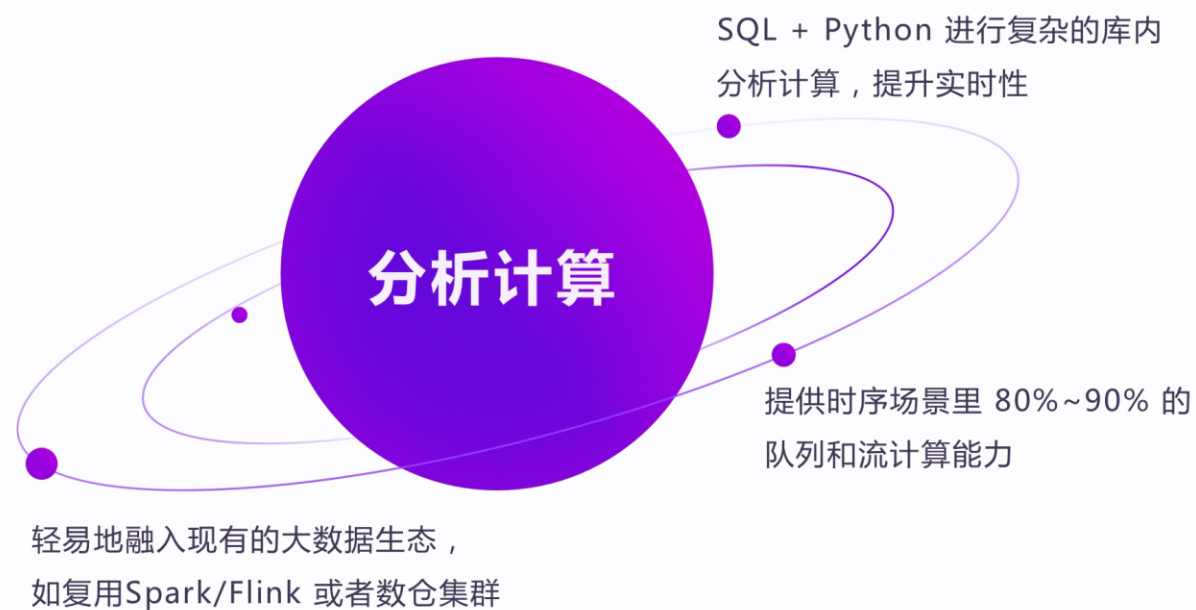
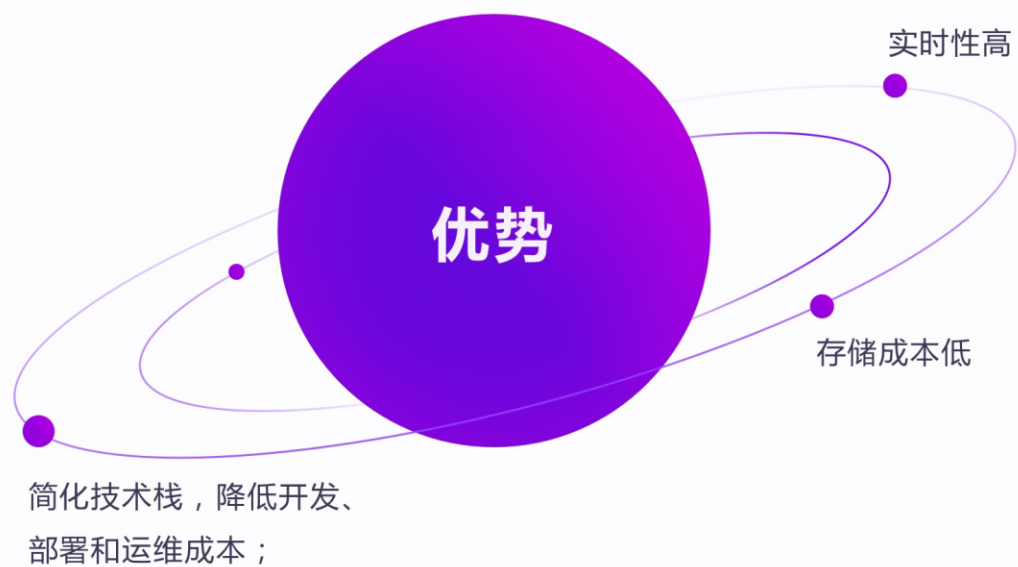
企业级能力

- 多租户、资源隔离、鉴权

超低存储成本

- 极致压缩
- 云上廉价存储
- 混合负载，一份数据支撑时序模型和分析模型查询，避免数据冗余





running Python script inside the database

01-

编写python脚本保存到文件

02-

提交

- `curl --data-binary "@system_status.py" -XPOST "http://ip:3000/v1/scripts?name=system_status"`

03-

执行

- `curl -XPOST "http://localhost:3000/v1/run-script?name=system_status"`

Python ▼

```
1 @coprocessor(args=["cpu", "mem"], returns=["overheat"])
2 def measure(cpu: vector, mem: vector):
3     overhear = (cpu > 75) & (mem > 60) # construct a new
4     # here can also contains more complicated code i.e.
5     return overhear
```


04

“开源之夏”项目介绍

Truncate Table 清空表功能的实现

项目编号：2332c0052

实现 GreptimeDB 的 TRUNCATE TABLE 功能；TRUNCATE TABLE 从表中删除所有行。它类似于没有 WHERE 子句的 DELETE 语句，但比逐行删除更快。通过该功能的实现，提升 GreptimeDB 的性能。

要求

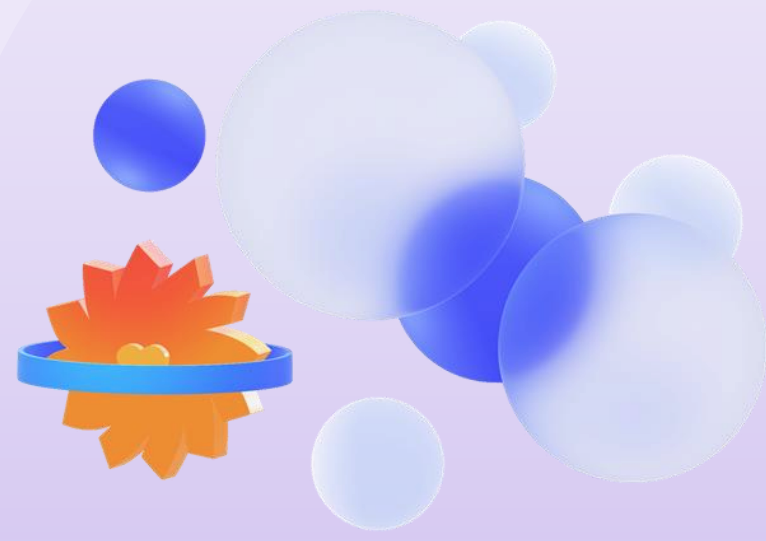
Rust 语言

SQL

导师

杨颖文 (evenyag)

- 偶像宅



实现“Duration”和“Interval”数据类型的支持

项目编号： 2332c0053

“Duration”数据类型是度量特定类型的时间，可以用秒、毫秒、微秒或纳秒表示。“Interval”数据类型通过计算两个时间点之间的天数表示时间差异。这些数据类型在表示和建模真实世界的时序场景中非常有用。

要求

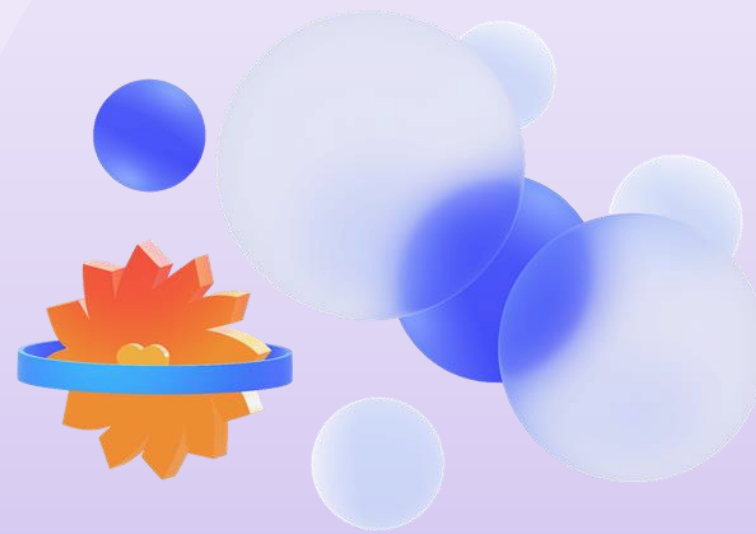
Rust 语言

Linux 系统

导师

黄磊 (v0y4g3r)

- 摄影爱好者



实现滑动采样窗口函数的语法及计算

项目编号：2332c0054

查询并聚合一个给定长度的时间范围的数据是时序数据常见的一种查询模式，例如 PromQL 中的 Range Selector，而 GreptimeDB 计划在 SQL 界面中实现类似的能力。与 WINDOW 语法不同，滑动采样窗口函数会按固定的步长及范围进行计算，例如：查询每日气温的周平均值；该查询的步长是一天（每日气温），每一次计算都使用过去一周的数据求平均。

该项目需要从拓展 SQL Parser 开始，定义并实现滑动采样窗口函数的语法，并拓展逻辑计划器来实现相关语句的计划生成，最后实现对常见聚合算子的支持。

要求

Rust 语言

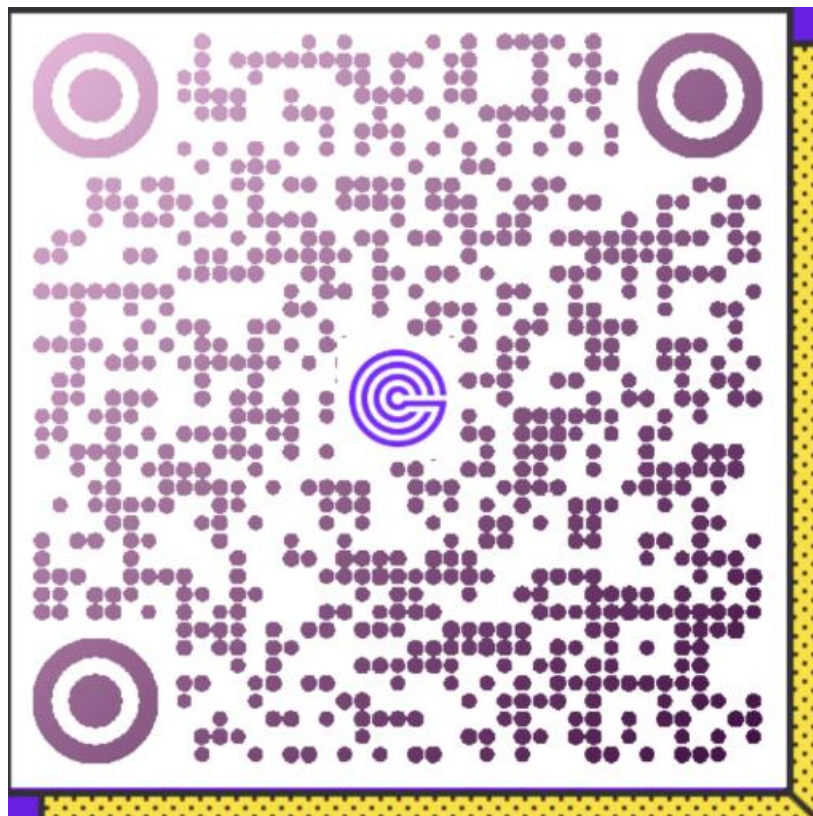
SQL

导师

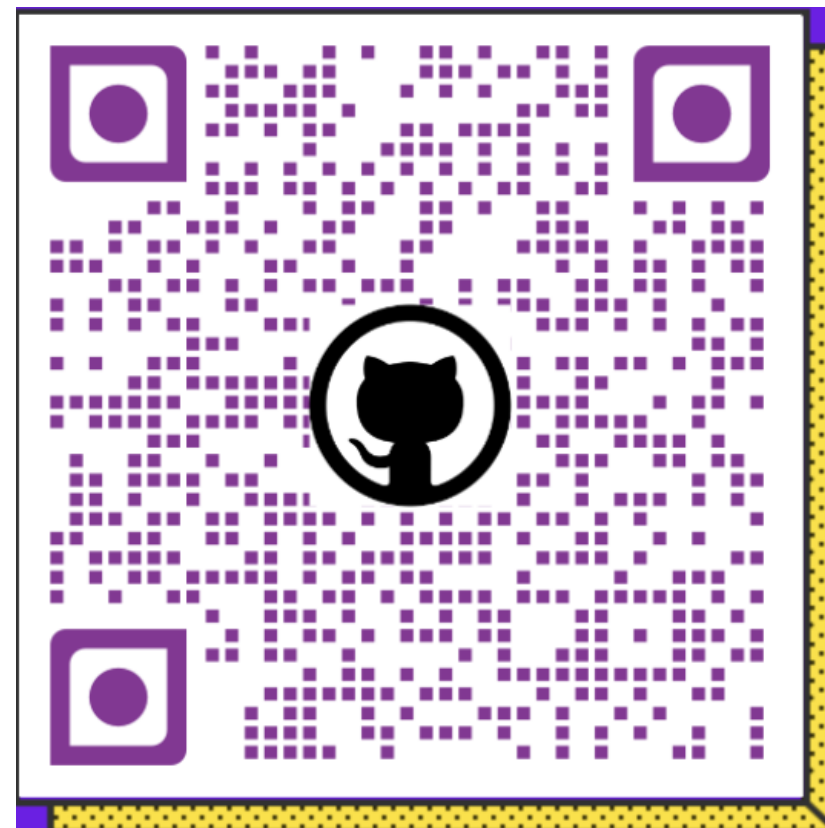
锐航 (waynexia)

- 沉迷游历海拉鲁中





Greptime 公众号



GreptimeDB GitHub



开源软件供应链点亮计划—开源之夏2023

Thanks

