openGauss 3.0.0重点特性总览

熊小军

openGauss数据库研发工程师

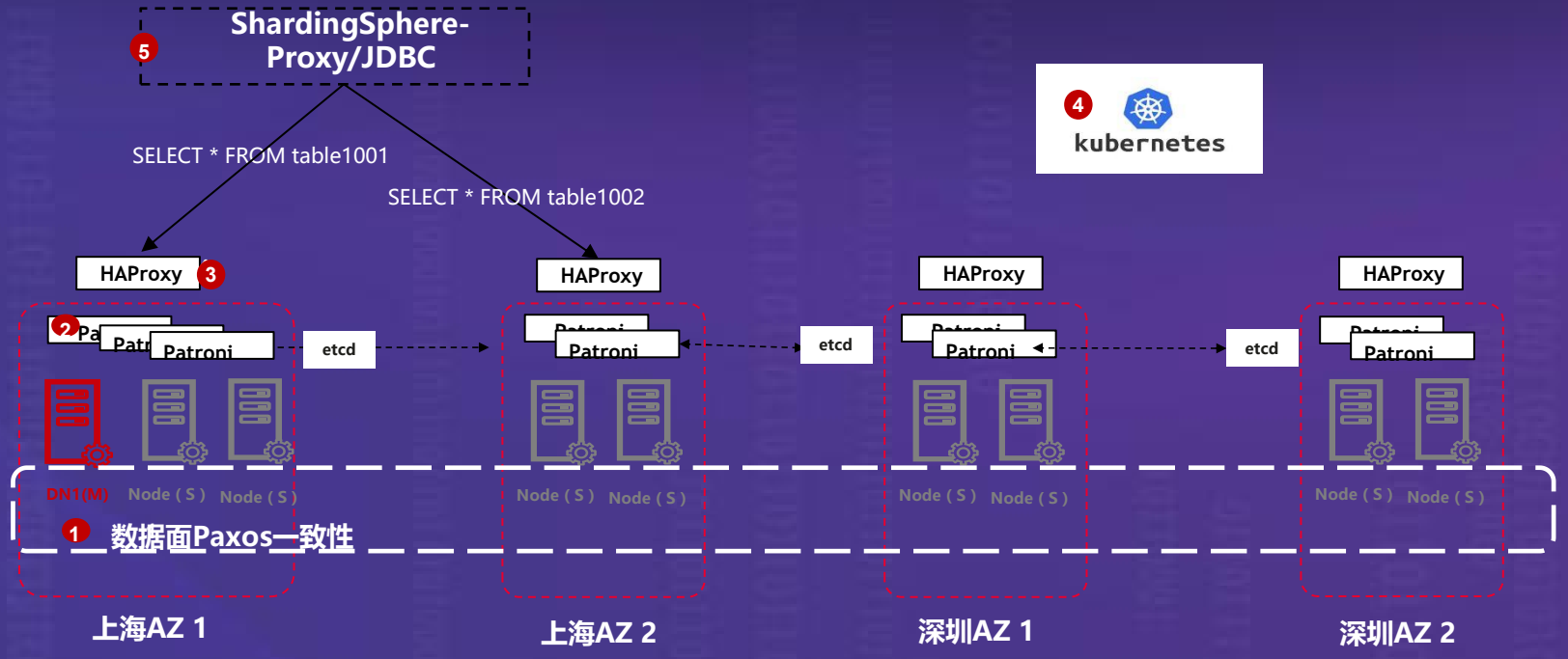
openGauss

https://opengauss.org

目录

- openGauss开源分布式解决方案
- 高性能：并行逻辑解码
- 高可用：CM（Cluster Manager）
- 高可用：Global Syscache
- AI4DB：支持Prometheus 生态，支持服务化、插件化
- 高安全：全密态数据库性能增强
- 轻量版：满足资源受限场景使用
- 工具链：Data Studio代码开源
- 其他企业级特性：发布订阅、行表压缩

openGauss开源分布式解决方案



* 数据按sharding key划分，满足大规模业务量场景
* 支持分布式查询、支持分布式事务
* 读写负载均衡
* 水平扩展，可扩展性强，在线扩缩容
* 灵活的分布式体系，无单点故障
* 吞吐量和低延时
* 易部署和运维

本方案对以上需求有良好的支持

1. 内核能力：Paxos主备副本一致性、主备自仲裁、日志并行复制
2. CM：基于etcd和Patroni的集群管理工具(将替换为openGauss-CM)
3. HAProxy：负载均衡 + 固定IP
4. K8S：支持基于Docker和K8s的集群部署，支持pg-pool的类云原生架构
5. ShardingSphere-Proxy：自动分片，读写分离，分布式事务，分布式查询，DistSQL

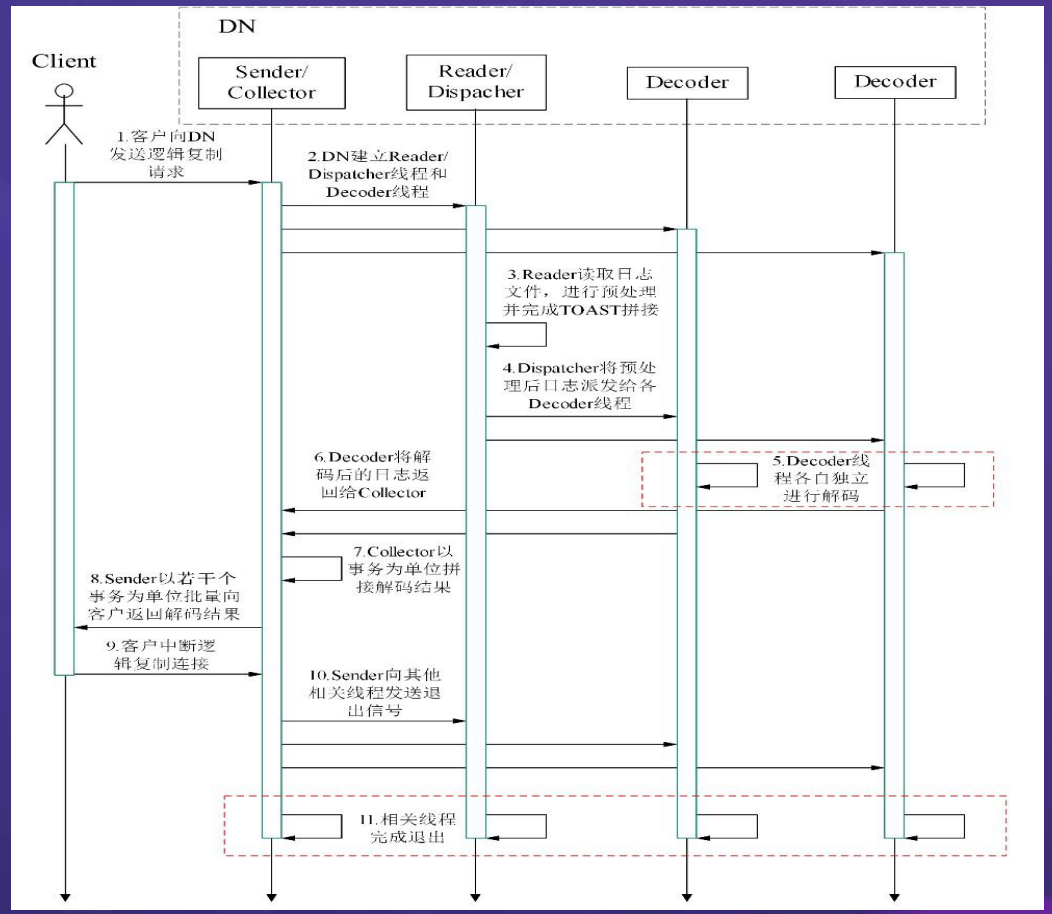
openGauss+ss的分布式解决方案使用16台服务器在超过1小时的测试中，得到超过1000万 tpmC的结果，行业同等规模下性能最好。

openGauss

https://opengauss.org

高性能：并行逻辑解码

逻辑复制串行解码：平均性能为3~5MBps，业务压力大时难以满足实时同步的需求，导致日志堆积，影响生产。其中解码流程耗时占比为70%左右，因此需通过多线程解码进行优化；
并行解码：多个线程协同并行解码从而提高解码性能，在基础场景下解码性能可达到100MBps.



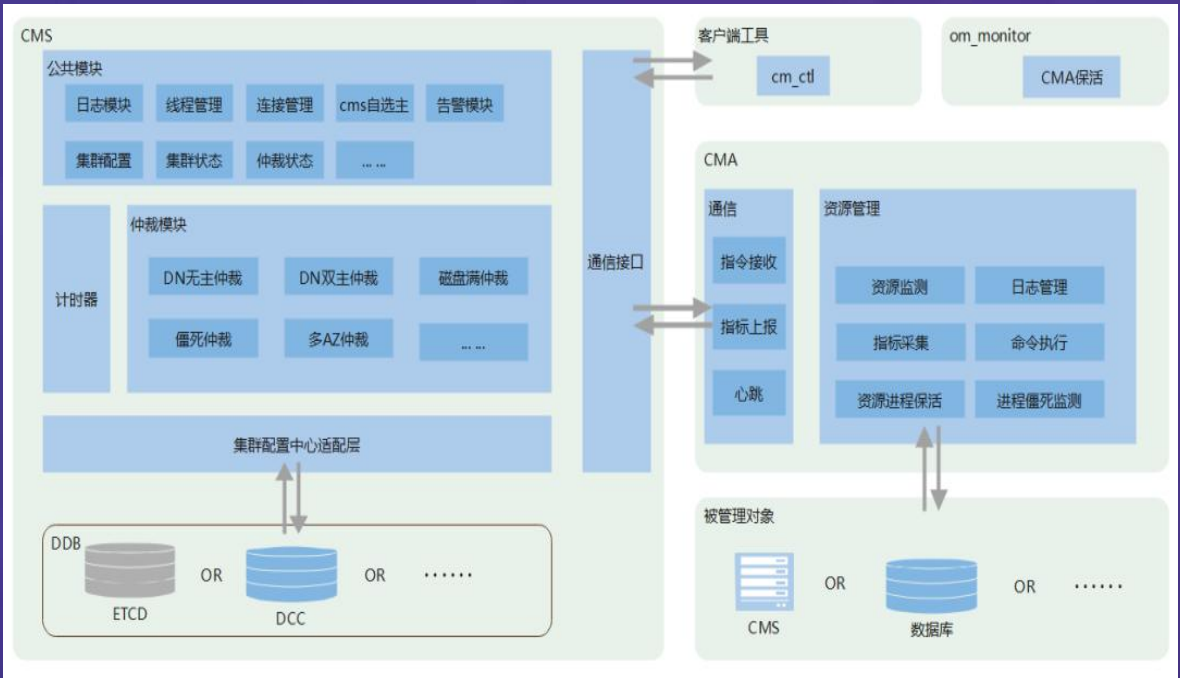
三种线程：
Sender/Collector（1个）：接收解码请求，拼接并发送解码结果
Reader/Dispatcher（1个）：读取Wal日志并分发到解码线程
Decoder（N个）：负责解码，并将解码结果发送给Sender

pg_recvlogical -d $db -p $port -o parallel-decode-num=5 -o standby-connection=true -o decode-style='t' -o white-table-list='public.t1,public.t2'
代表5并发解码、仅解码表public.t1和public.t2、启用备机连接、解码格式为text

openGauss
https://opengauss.org

高可用：CM（Cluster Manager）

CM（Cluster Manager）：集群资源管理软件。支持自定义资源监控，提供了数据库主备的状态监控、网络通信故障监控、文件系统故障监控、故障自动主备切换等能力。



cm_server：cm的服务端，负责收集cma上报的状态，并作为仲裁中心和全局配置中心，集群能否稳定运行以及在发生单点故障后，备实例能否正常切换为主来保证集群的可用性，都与CMS是否稳定相关。

cm_agent：通常集群中的每台机器都安装一个，负责管理本节点所有实例的状态检测和上报以及cms下发命令的执行。

om_monitor：通常集群中的每台机器都安装一个，负责保障本节点cm_agent进程的健康。

cm_ctl：cm的客户端工具，提供集群管理操作。

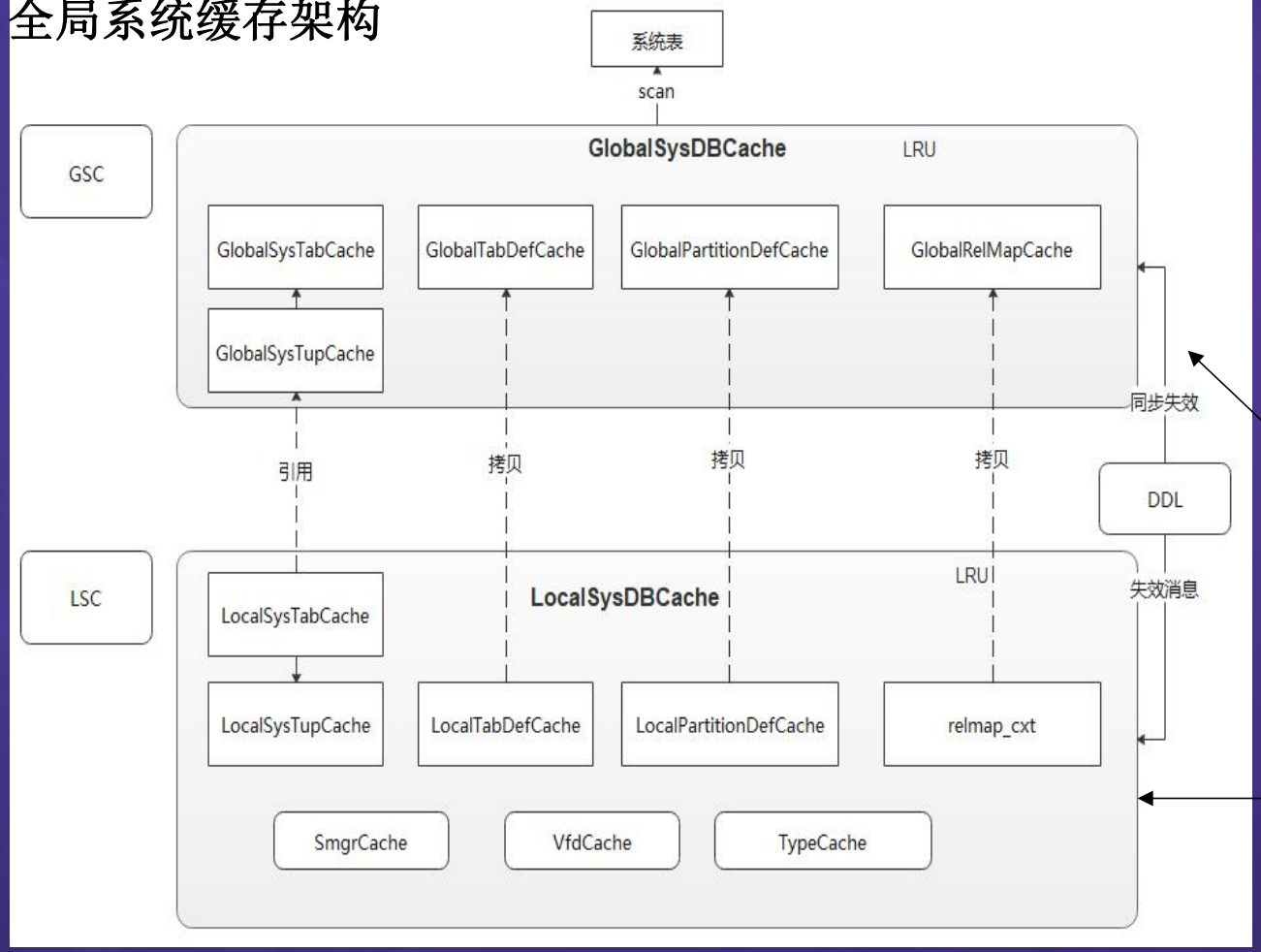
openGauss

https://opengauss.org

高可用：Global Syscache

系统缓存与线程绑定，结合全局系统缓存和线程池，降低系统缓存内存占用，提升数据库并发扩展能力

全局系统缓存架构

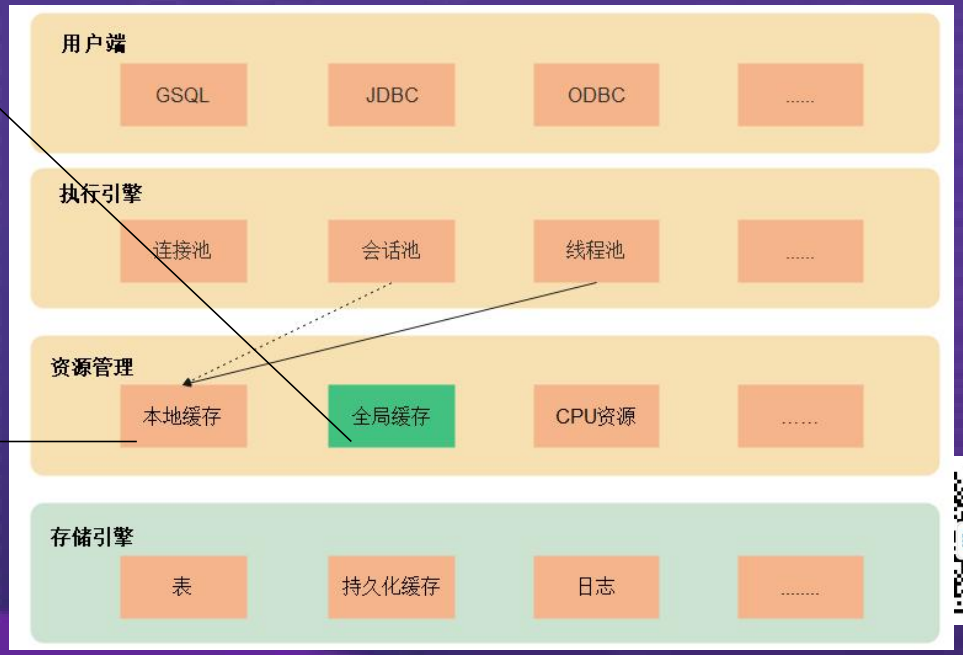


背景：

高并发场景中，需要单实例支持一万并发，在openGauss现有的实现机制下每个session自己维护一套完整私有的系统缓存，使得该并发场景下即便大量session即使没有获得CPU资源，但是却占据着大量的内存导致对系统内存挤占，严重限制了openGauss并发的扩展性。

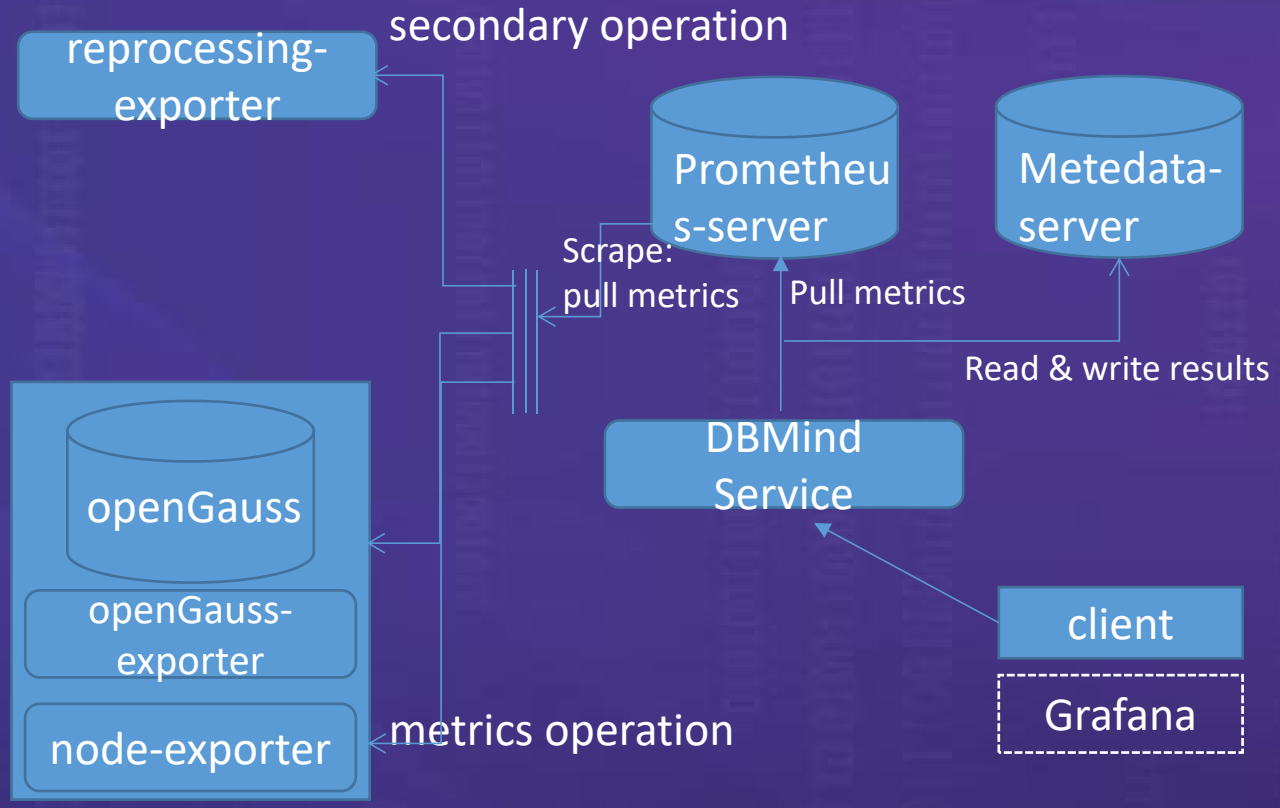
目标：

单DN存在1万以上会话链接，全局共享SysCache能够使Session使用的内存占用大小趋于恒定，不再与session并发度线性相关。DN上SessionContext中内存降低90%，性能劣化小于5%。



AI4DB：支持Prometheus 生态，支持服务化、插件化

AI4DB: 实现了后台监控服务，并在后台定期地检查数据库系统的状态，从而形成了自治数据库平台DBMind。通过离线计算的形式，将诊断结果保存，用户可以通过Grafana等软件进行可视化，从而第一时间发现问题并获知问题的根因。



openGauss-exporter 用于获取数据库系统的监控指标（metric），reprocessing-exporter用于对存储在Prometheus中的数据进行二次加工，
通过Prometheus定期采集获取exporter的数据，
DBMind系统定期从Prometheus中获取时序数据，并完成计算，计算结果存储在元数据库（meta-database）中，
用户可以从元数据库中获取诊断结果，同时可通过配置Grafana等进行可视化。

openGauss还全面整合了现有的AI能力，并重新设计了一种插件化的模式；
通过gs_dbmind 命令，可以调用所有的AI功能，通过component 子命令，可以调用具体的AI功能：
gs_dbmind component xtuner tune ...

openGauss

https://opengauss.org

高安全：全密态数据库性能增强

基于全密态数据库透明计算架构，实现新一代密态等值查询能力

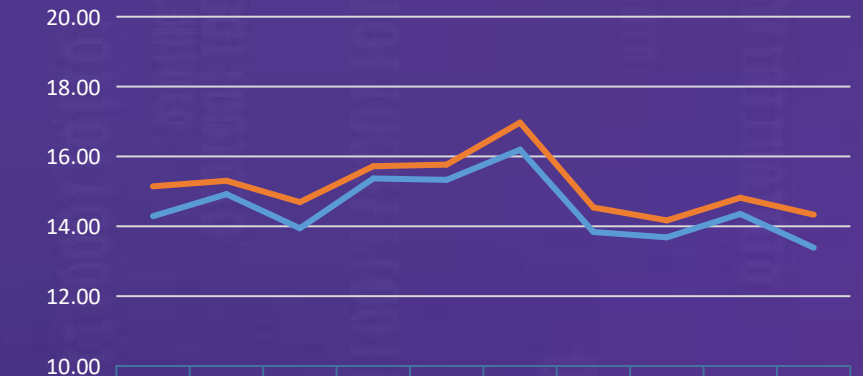
- **密态查询功能**：(1)支持存储过程和函数等值查询能力，支持通用JDBC应用开发接口；

  (2)支持国密算法

- **密态查询性能**：(1)默认开启全密态数据库能力下对系统benchmark性能影响低于10%；

  (2)在使用全密态数据库时，密文查询的性能相比密文查询的性能劣化低于10%；

```
bmsql_result_v5.py              hs_err_pid658130.log         Term-00, Running Average tpmTOTAL: 3071781.34    Current tpmTOTAL:
13:26:30,963 [Thread-37] INFO   jTPCC : Term-00,                                                  Term-00, Running Average tpm
14:45:45,359 [Thread-612] INFO   jTPCC : Term-00,
14:45:45,359 [Thread-612] INFO   jTPCC : Term-00,
14:45:45,359 [Thread-612] INFO   jTPCC : Term-00, Measured tpmC (NewOrders) = 1491493.68
14:45:45,359 [Thread-612] INFO   jTPCC : Term-00, Measured tpmTOTAL = 3314516.97
14:45:45,359 [Thread-612] INFO   jTPCC : Term-00, Session Start     = 2021-10-25 13:45:45
14:45:45,360 [Thread-612] INFO   jTPCC : Term-00, Session End       = 2021-10-25 14:45:45
14:45:45,360 [Thread-612] INFO   jTPCC : Term-00, Transaction Count = 198876597
^C
[1]+  Done                    numactl -C 32-63,64-87,96-119 ./runBenchmark.sh liu_props.run_single_node.sh.mt > test1025.log
```



	1	2	3	4	5	6	7	8	9	10
密态表	15.14	15.30	14.69	15.72	15.77	16.97	14.54	14.17	14.82	14.33
非密态表	14.29	14.92	13.94	15.37	15.33	16.19	13.84	13.68	14.35	13.39

openGauss

https://opengauss.org

轻量版：满足资源受限场景使用

当前局限：openGauss内存底噪及安装包太大，导致在资源紧张场景无法应用。
主要目标：内存底噪由当前950M下降到<250M；安装包（tar压缩包）由当前200M下降到<30M，满足资源受限场景下使用。
主要策略：GUC参数调整、三方库剥离（100+开源链接库减低到20个）、线程数（30后台线程减低到14个）、内存使用优化。

优化项	具体策略
GUC参数调优	22个GUC参数调整，涉及轻量化不使用特性，双写、Ustore、资源管控等。
调整预留线程数	16个线程默认不启动，涉及smp、asp、ustore、2pccleaner、snapshot等
系统表bucket数	系统表bucket数目优化，涉及系统表26个
三方库优化	无用三方件剥离，如OBS、ORC、Parquet、LLVM等
无用组件优化	无用组件剥离，如OM、JRE

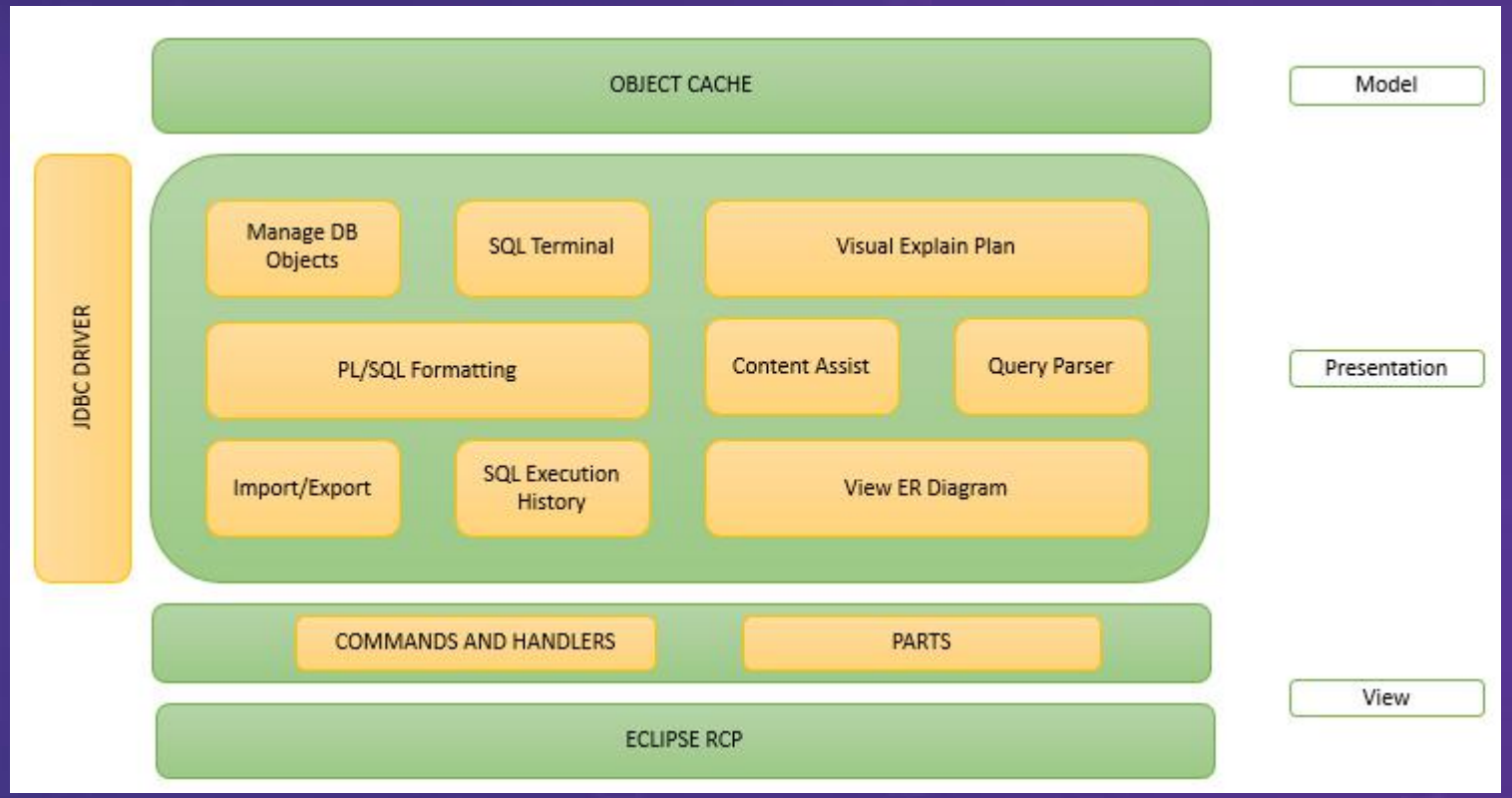
	压缩包	解压后	初始目录	空载内存
openGauss lite	<30M	117M	144M	<200M, 250M>
openGauss server	98M	349M	649M	512M

openGauss
https://opengauss.org

工具链：Data Studio代码开源

DataStudio是一个图形化的客户端工具，它通过JDBC驱动与openGauss数据库连接，采用C/S架构进行通信。

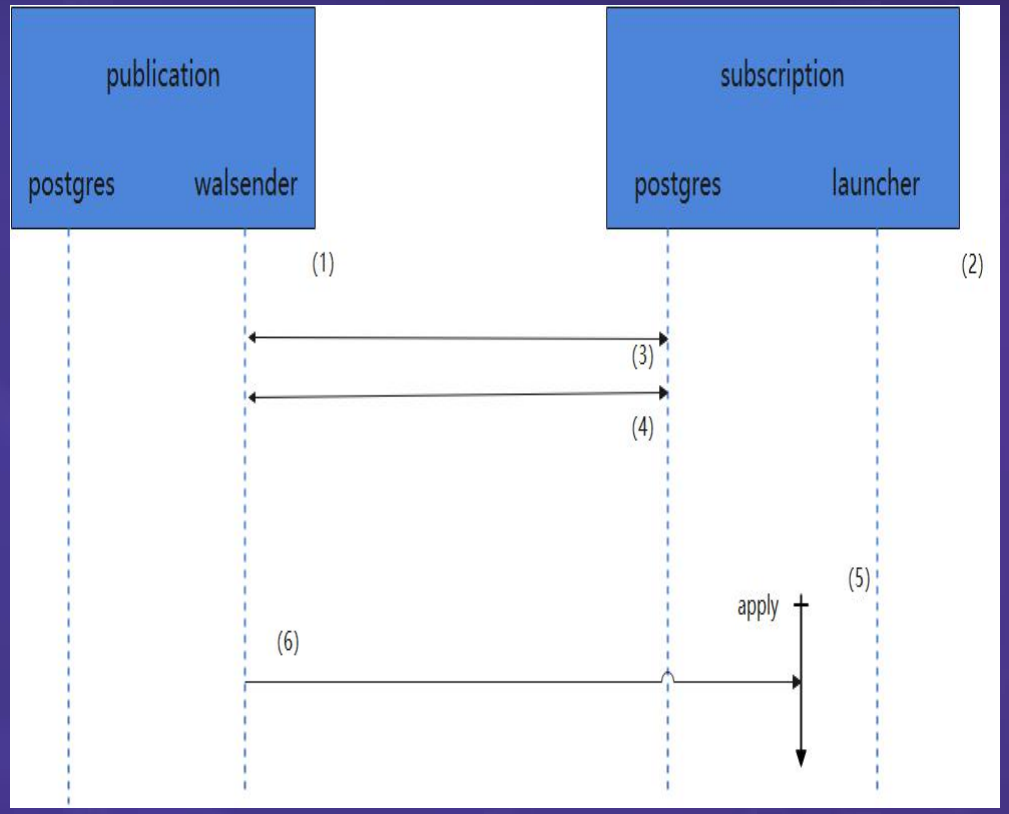
- model是用于定义将要显示的数据或者是用户所提供的待显示数据.
- view是一个被动界面用于显示数据（model），并将用户命令（events）发送到处理器(presenter)进行处理.
- presentation 作用于model 和view. 它从model读取数据并且格式化后显示在view上。

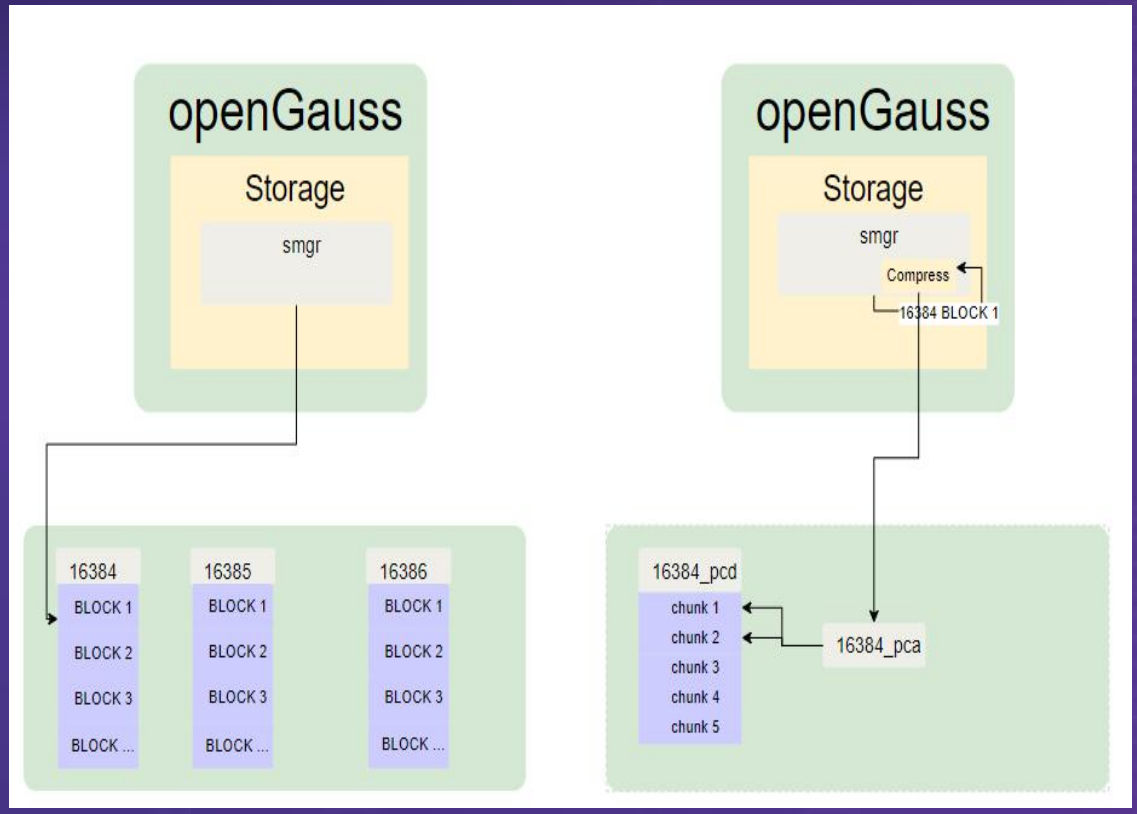


openGauss

https://opengauss.org

其他企业级特性：发布订阅、行表压缩





原理：基于逻辑复制实现，订阅者从它们所订阅的发布节点拉取数据。实现跨数据库集群的数据实时同步。
流程：创建发布->创建订阅->同步数据
应用：
1、两套集群组成互为发布、订阅的关系，用于异地双活等场景。
2、把多个数据库联合到单一数据库中，用于数据分析等场景。

通过对数据页的透明页压缩和维护页面存储位置的方式，以页（page）为粒度对数据进行压缩，将压缩后的数据存储到磁盘（pcd文件），同时用一个新文件存储数据存储的索引（pca文件）。
支持多种通用压缩算法和压缩级别设置。

微信入群助手

openGauss社区官网

openGauss

https://opengauss.org

Thank you!

openGauss

https://opengauss.org