

vmware®

EXPLORE

2022

Greenplum 数据库集群的性能监控

付宁

Greenplum 产品经理

免责声明

本演示文稿可能包含当前正在开发的产品特性或功能。

本新技术概要介绍并不表示 VMware 承诺在任何正式推出的产品中提供这些功能特性。

产品的功能特性可能会有变更，因此不得在任何类型的合同、采购订单或销售协议中予以规定。

技术可行性和市场需求都可能影响最终提供的产品功能特性。

在本演示中讨论或展示的任何新特性/功能/技术的定价与包装都尚未确定。

Agenda

- ❖ Greenplum Database 的性能与监控
- ❖ GPCC 的架构与功能
 - ❖ 查询监控
 - ❖ 负载管理
 - ❖ 运维辅助
- ❖ Demo



Greenplum Database (GPDB)

- ❖ 基于 PostgreSQL
- ❖ MPP、分布式
- ❖ OLAP (Online analytical processing) 数据库
- ❖ 开放源代码
(<https://github.com/greenplum-db/gpdb/>)
- ❖ 将近 20 年历史，2017年发布 GPDB 5，
2019年 发布 GPDB 6



**GREENPLUM
DATABASE®**

Greenplum 数据库的性能问题

1. 查询为什么跑的慢？慢在什么地方？
2. 哪些查询占用了最多的CPU/内存/磁盘 IO？
3. 能否在一个查询消耗过多资源的时候把它终止掉？
4. 能否让一些特定的查询获得更多的资源？
5. 哪些表需要进行 VACUUM / ANALYZE？
6. 系统的内存足够？还是不够？
7. . . .



影响性能的因素

影响GPDB查询性能的主要因素

硬件性能及状态

系统资源

- CPU/内存/磁盘
- Resource Group/Queue的设置

负载 (Workload)

数据库、表的日常维护

- VACUUM
- ANALYZE

数据库的设计和优化

- 数据的分布
- 优化器和查询计划



影响性能的因素

影响GPDB查询性能的主要因素

硬件性能及状态

系统资源

- CPU/内存/磁盘
- Resource Group/Queue的设置

负载 (Workload)

数据库、表的日常维护

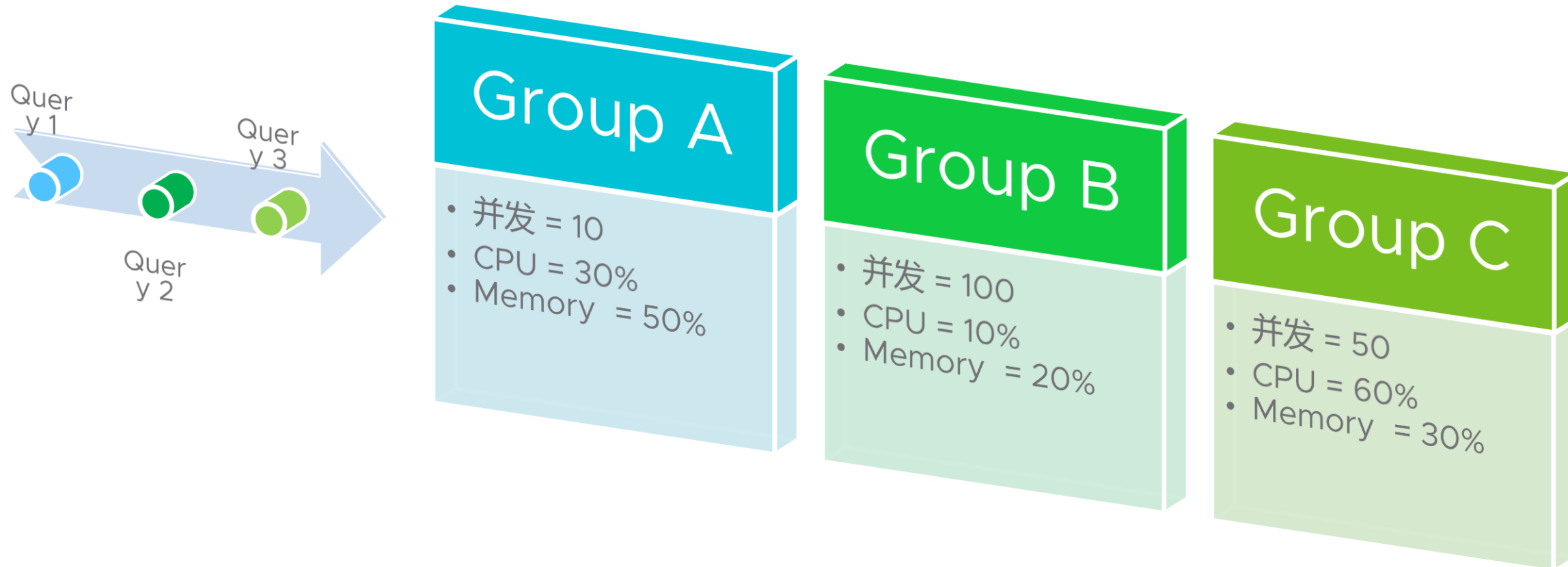
- VACUUM
- ANALYZE

数据库的设计和优化

- 数据的分布
- 优化器和查询计划



Resource Groups



影响性能的因素

影响GPDB查询性能的主要因素

硬件性能及状态

系统资源

- CPU/内存/磁盘
- Resource Group/Queue的设置

负载 (Workload)

数据库、表的日常维护

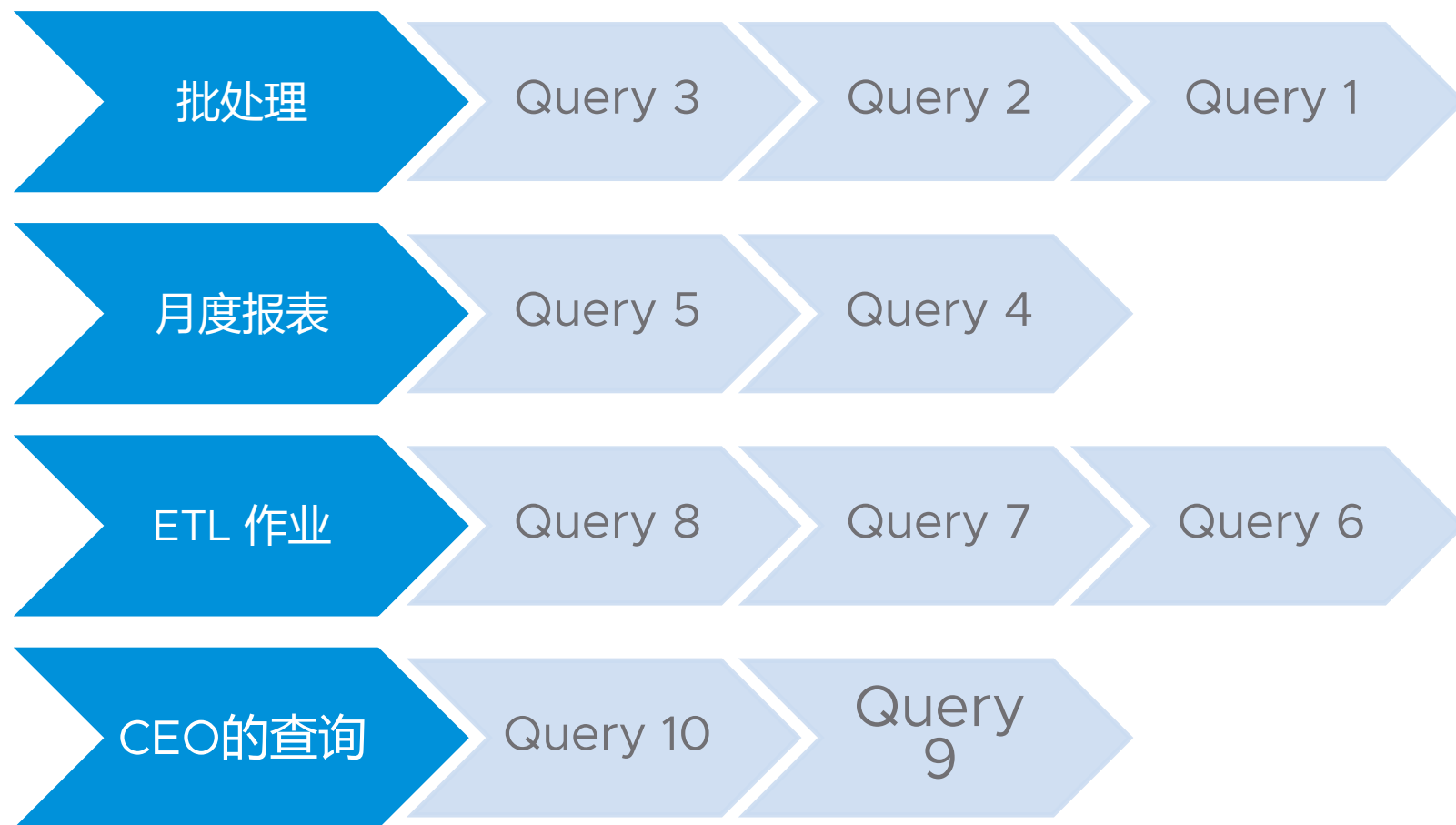
- VACUUM
- ANALYZE

数据库的设计和优化

- 数据的分布
- 优化器和查询计划



负载(Workloads)



影响性能的因素

影响GPDB查询性能的主要因素

硬件性能及状态

系统资源

- CPU/内存/磁盘
- Resource Group/Queue的设置

负载 (Workload)

数据库、表的日常维护

- VACUUM
- ANALYZE

数据库的设计和优化

- 数据的分布
- 优化器和查询计划



VACUUM & ANALYZE

❖ VACUUM/REINDEX

- ❖ 消除表/索引的膨胀(Bloat)
- ❖ Catalog 表的膨胀严重影响性能
- ❖ 需要在空闲时进行
- ❖ 尽可能经常运行 VACUUM 以避免 VACUUM FULL

❖ ANALYZE

- ❖ 收集表的统计数据
- ❖ 影响优化器对执行路径的代价(COST)的估算, 进而影响生成的查询计划
- ❖ 建议在空闲时进行

影响性能的因素

影响GPDB查询性能的主要因素

硬件性能及状态

系统资源

- CPU/内存/磁盘
- Resource Group/Queue的设置

负载 (Workload)

数据库、表的日常维护

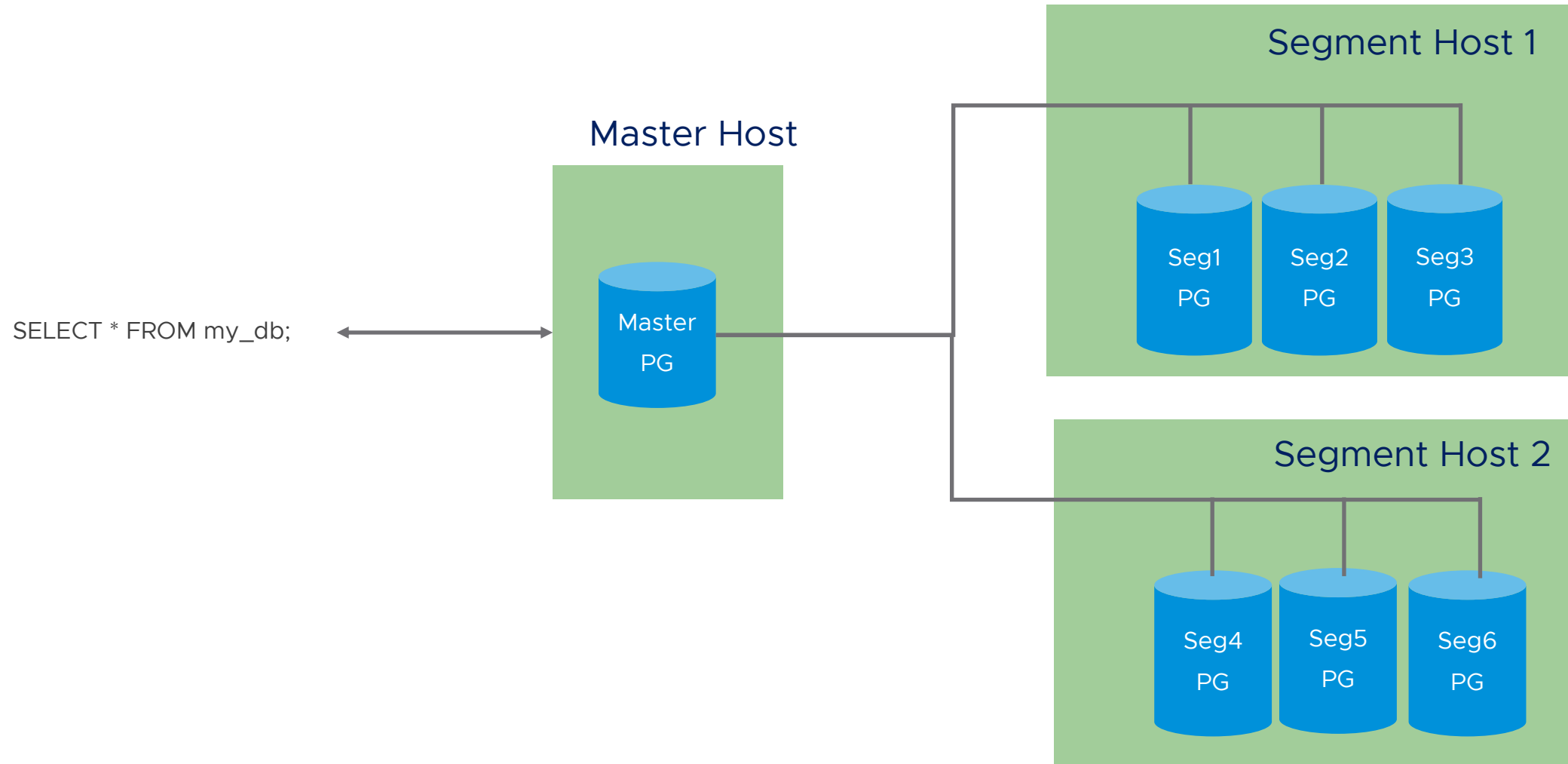
- VACUUM
- ANALYZE

数据库的设计和优化

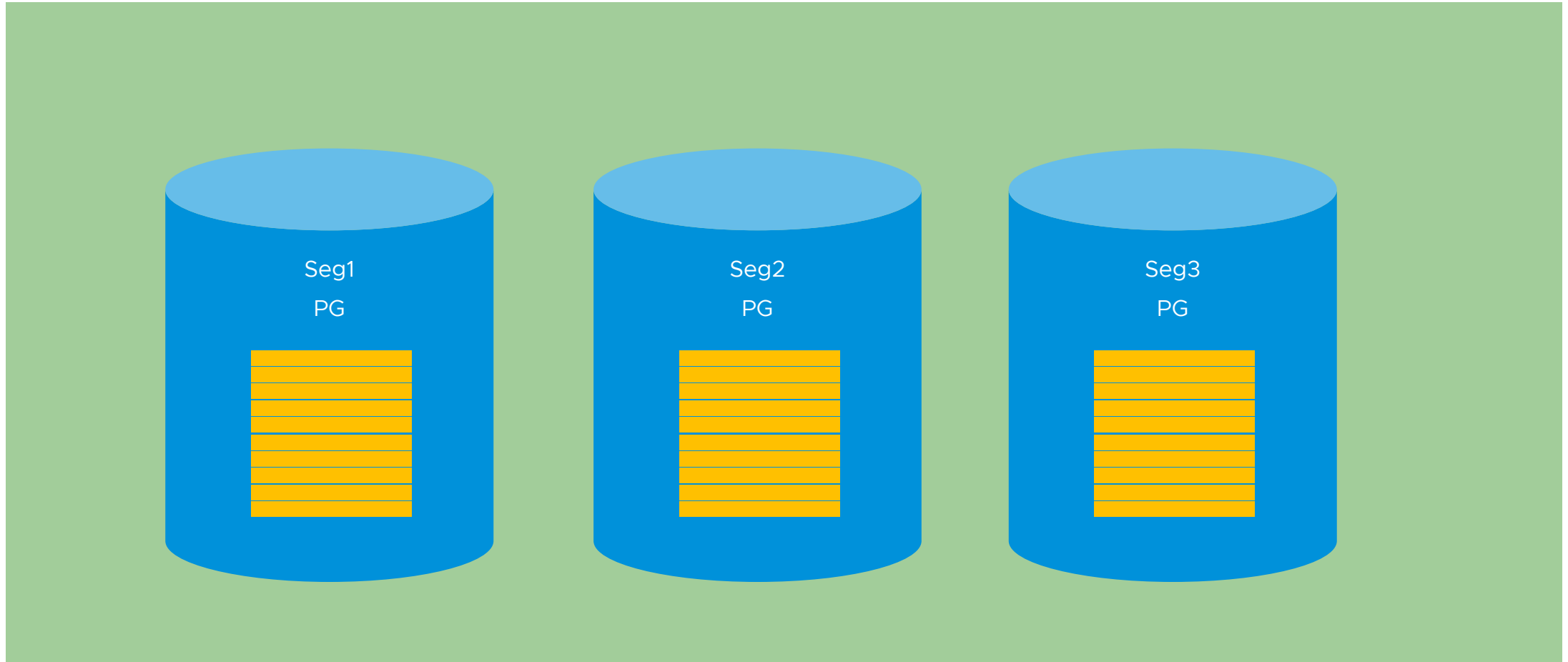
- 数据的分布
- 优化器和查询计划



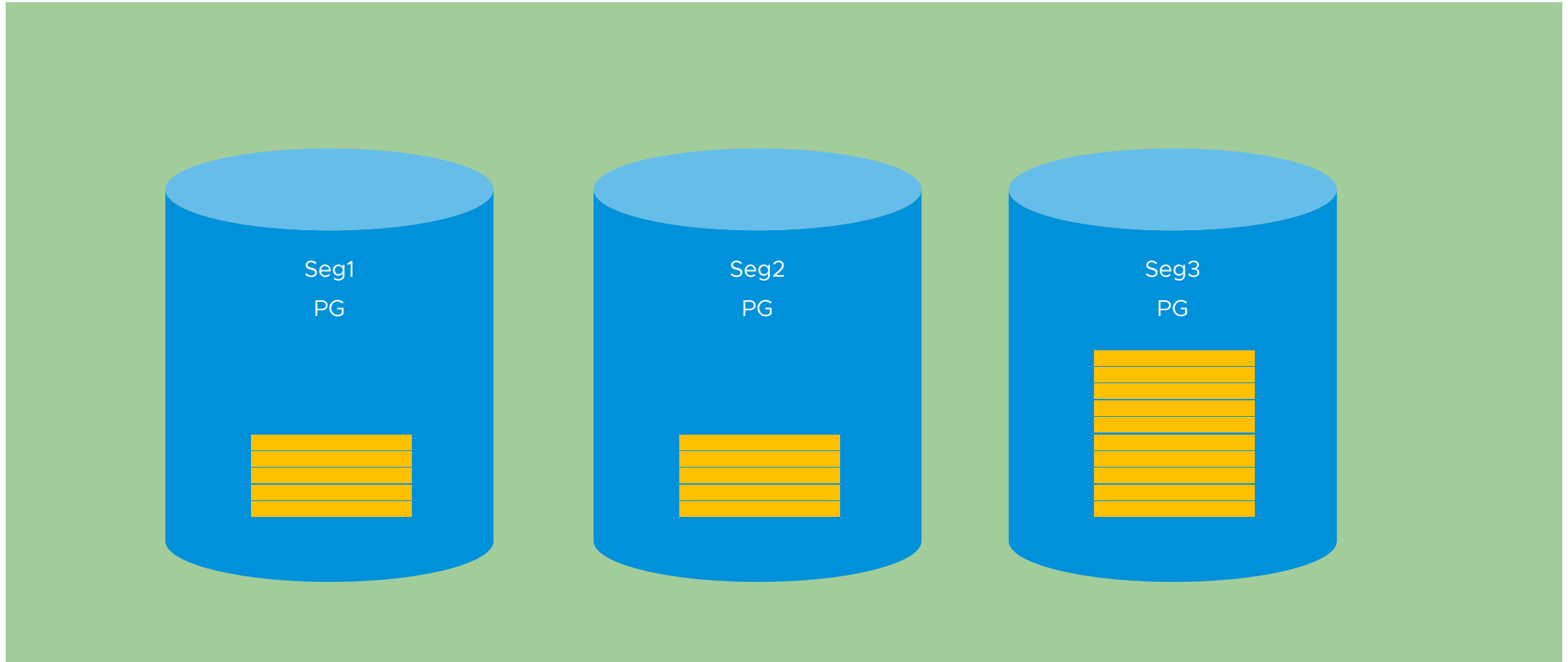
Greenplum Database



数据的分布与倾斜(Skew)



数据的分布与倾斜(Skew)



优化器和查询计划

❖ OPTIMIZER=ON (GPORCA)

```
QUERY PLAN
-----
Aggregate (cost=0.00..296.14 rows=1 width=8)
  -> Gather Motion 2:1 (slice1; segments: 2) (cost=0.00..295.10 rows=1 width=8)
        -> Aggregate (cost=0.00..294.10 rows=1 width=8)
              -> Seq Scan on part (cost=0.00..97.69 rows=100040 width=1)
Settings: optimizer=on
Optimizer status: Pivotal Optimizer (GPORCA) version 1.584
(5 rows)
```

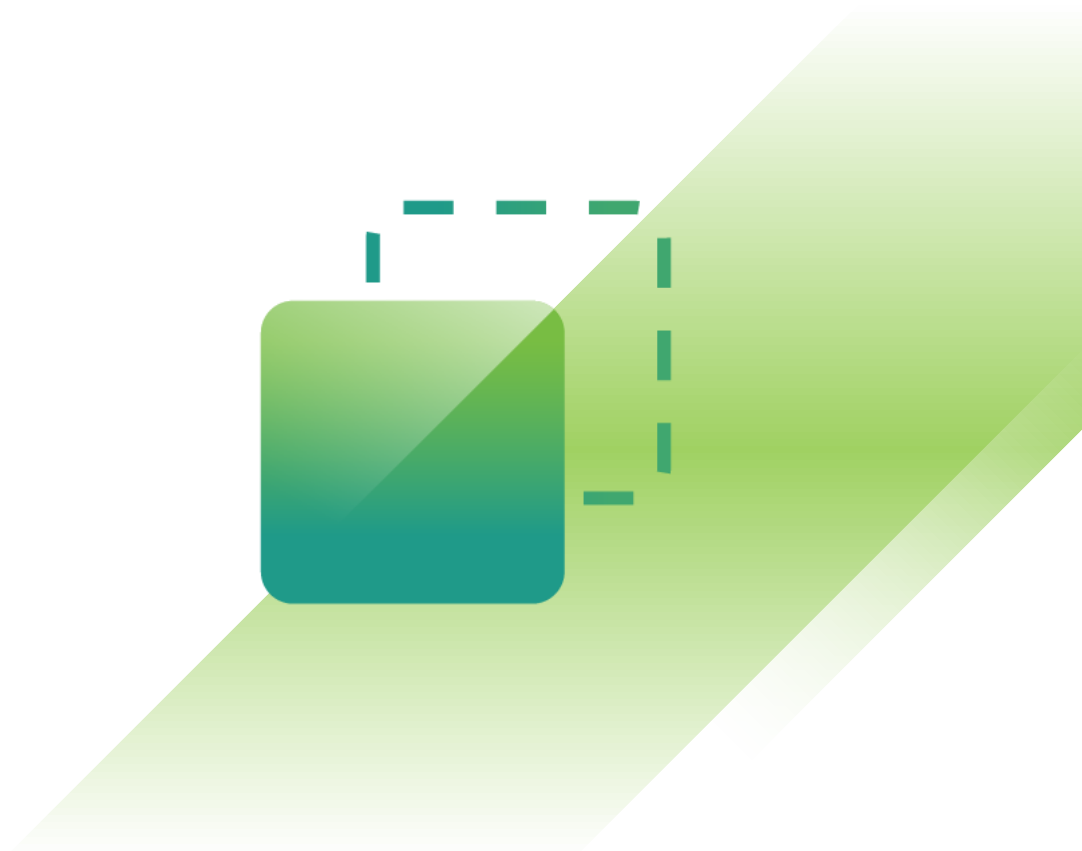
❖ OPTIMIZER=OFF (Postgres query optimizer)

```
explain select count(*) from part;
```

```
QUERY PLAN
-----
Aggregate (cost=3519.05..3519.06 rows=1 width=8)
  -> Gather Motion 2:1 (slice1; segments: 2) (cost=3518.99..3519.03 rows=1 width=8)
        -> Aggregate (cost=3518.99..3519.00 rows=1 width=8)
              -> Seq Scan on part (cost=0.00..3018.79 rows=100040 width=1)
Settings: optimizer=off
Optimizer status: Postgres query optimizer
(5 rows)
```

Greenplum Command Center

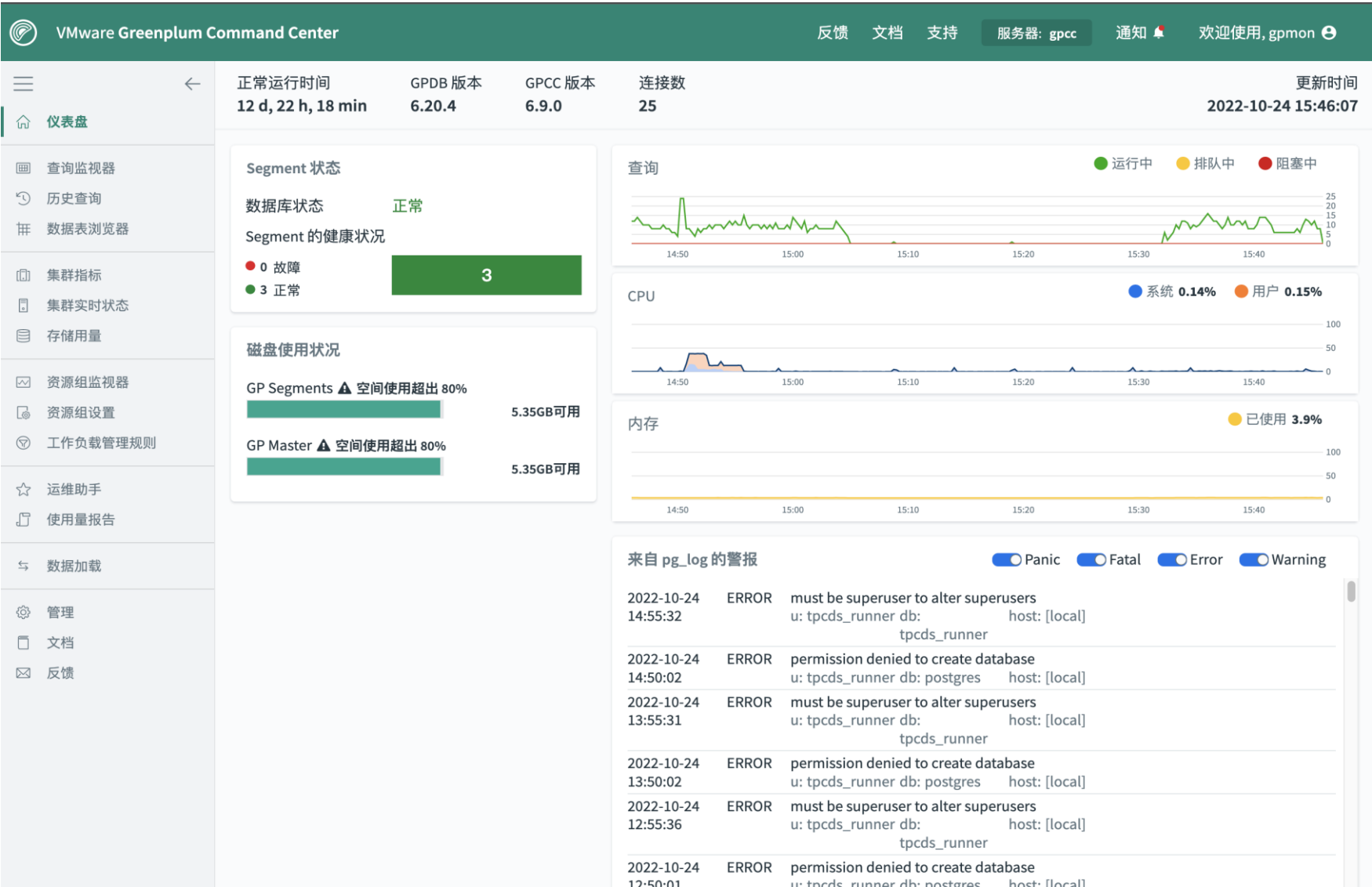
- 为 GPDB 集群提供实时的性能监控



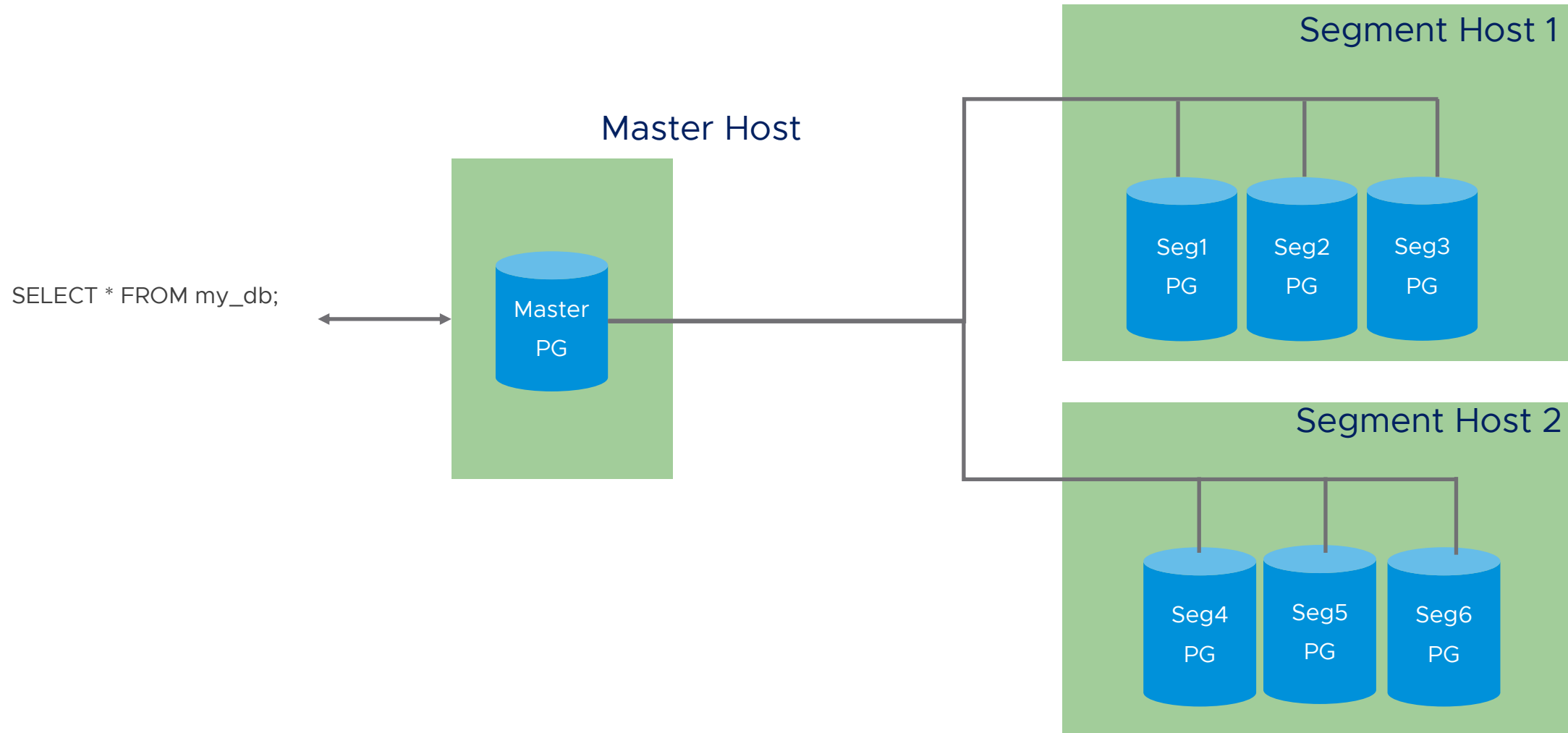
Greenplum Command Center

主要功能

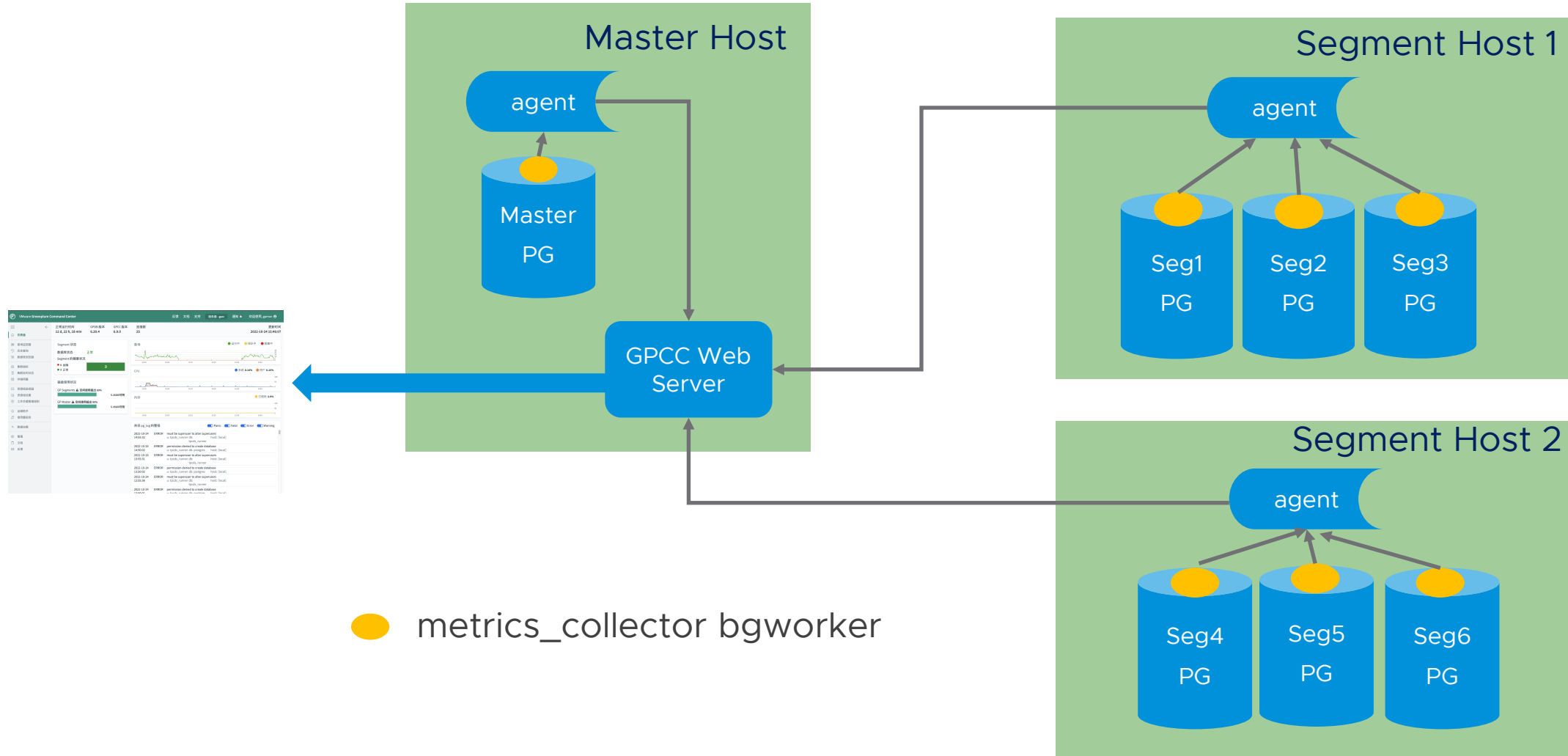
- 系统监控
- 查询监控
- 负载管理
- 运维辅助
- 异常报警
- 数据加载
- 用户管理
- . . .



Greenplum Database



Greenplum Command Center



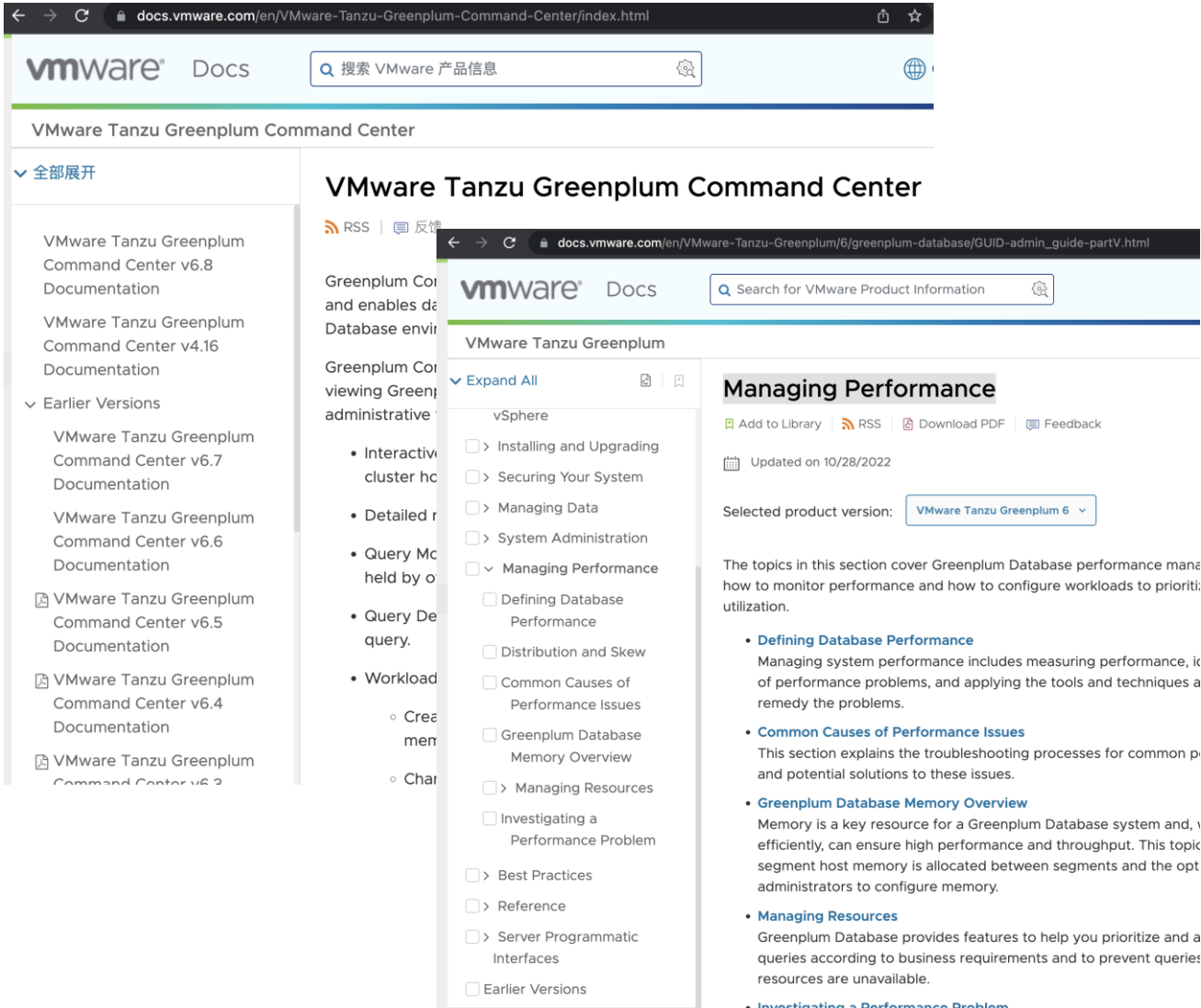
Demo

- 1. 查询监控
- 2. 负载管理
- 3. 运维助手



官方文档&社区资源

1. <https://cn.greenplum.org/>



谢谢!

vmware®
EXPLORE
2022