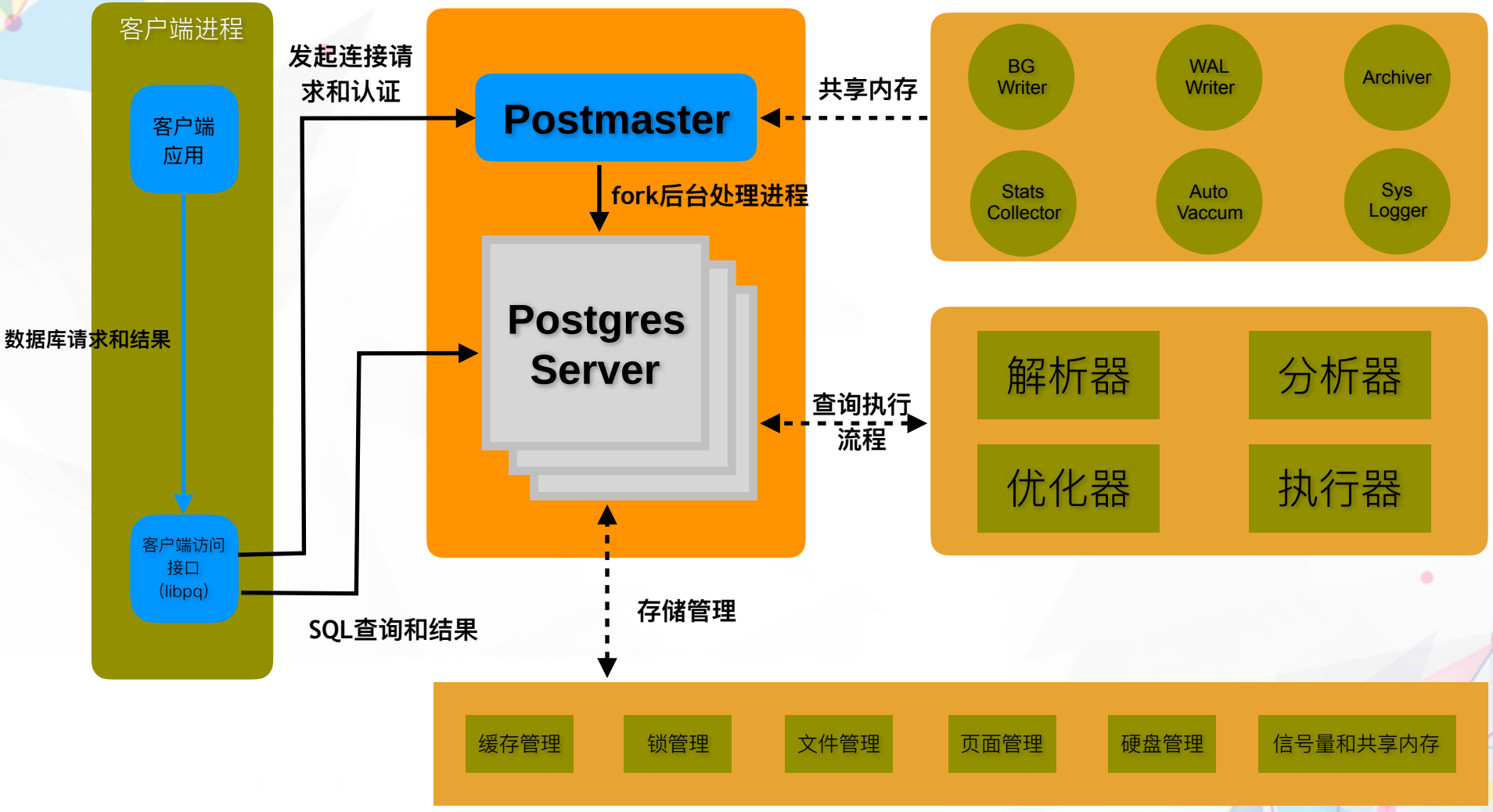
数据仓库架构的变迁

简丽荣，HashData

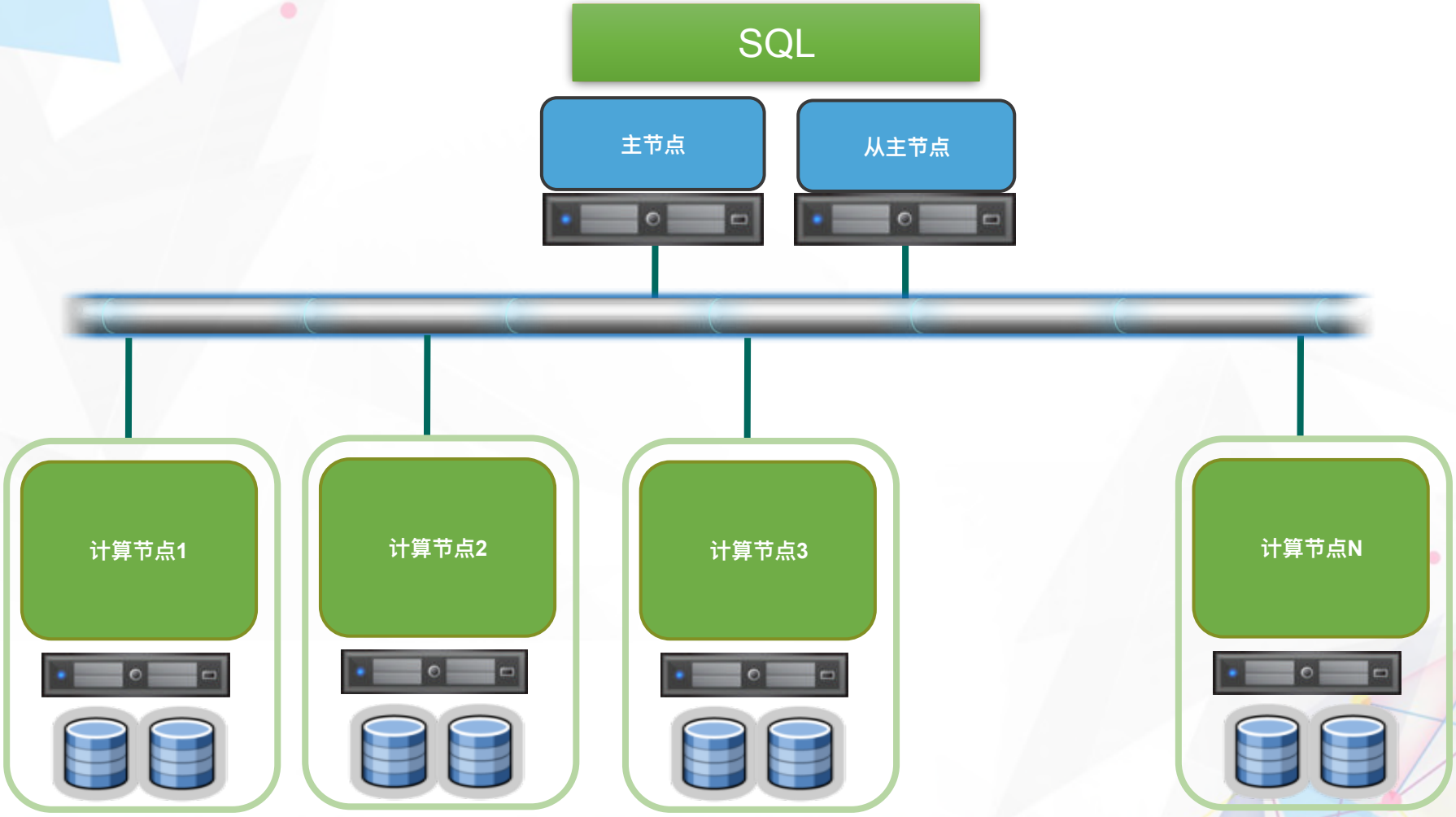
日程安排

- PostgreSQL
- Greenplum Database
- Apache HAWQ
  - Greenplum on HDFS
  - 1.x
  - 2.0
- 云端数据仓库
- 总结

PostgreSQL

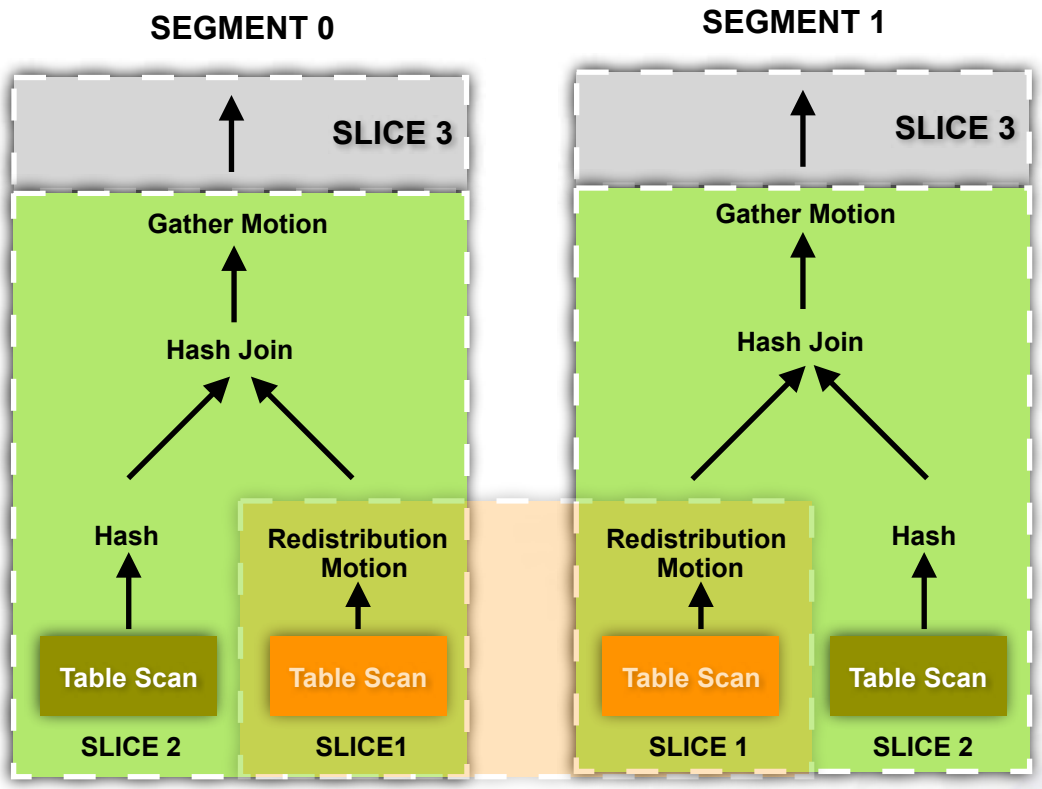

客户端进程
客户端应用
客户端访问接口(libpq)
发起连接请求和认证
数据库请求和结果
SQL查询和结果
Postmaster
fork后台处理进程
Postgres Server
共享内存
BG Writer
WAL Writer
Archiver
Stats Collector
Auto Vaccum
Sys Logger
查询执行流程
解析器
分析器
优化器
执行器
存储管理
缓存管理
锁管理
文件管理
页面管理
硬盘管理
信号量和共享内存

MPP架构的分布式数据库

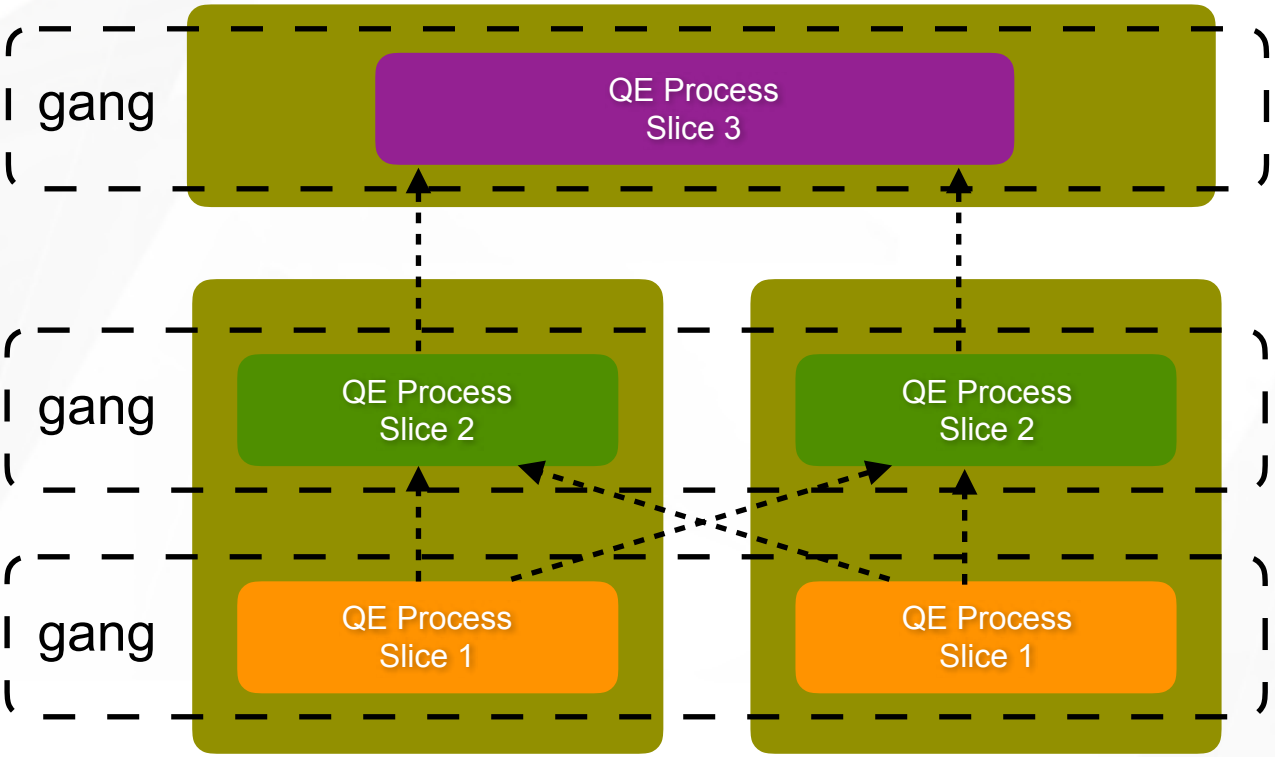


并行查询计划

```
SELECT customer,
       amount
FROM   sales
JOIN   customer
USING (cust_id)
WHERE  date=2008;
```

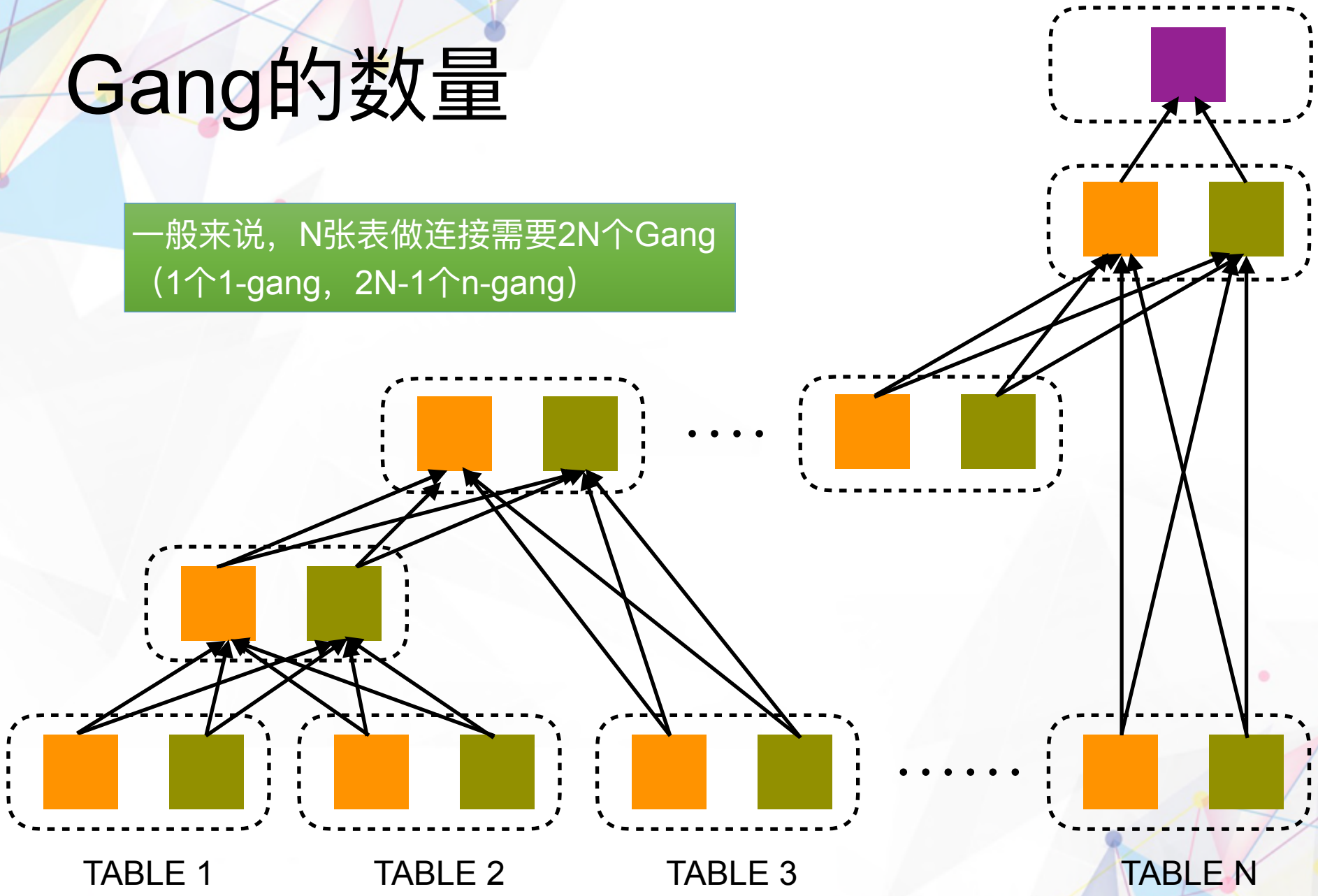


并行计划的执行


gang
QE Process
Slice 3
gang
QE Process
Slice 2
QE Process
Slice 2
gang
QE Process
Slice 1
QE Process
Slice 1

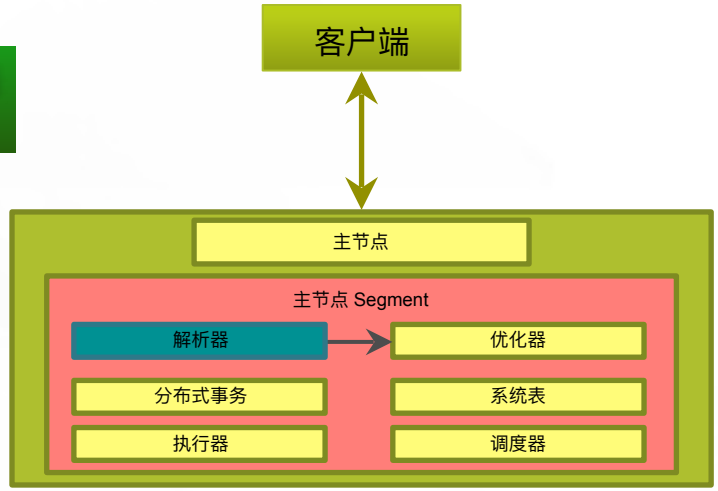
Gang的数量

一般来说，N张表做连接需要2N个Gang
（1个1-gang，2N-1个n-gang）



解析器

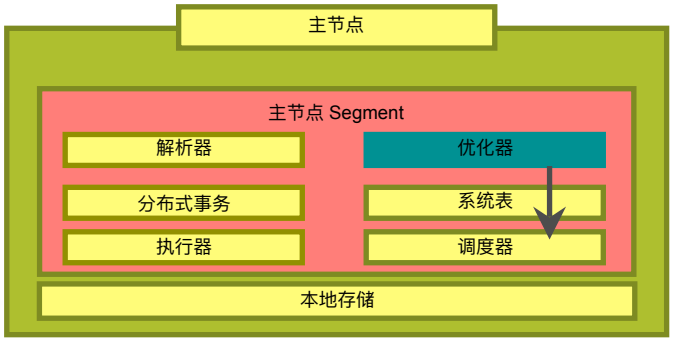
解析器执行词法分析、语法分析并生成 解析树



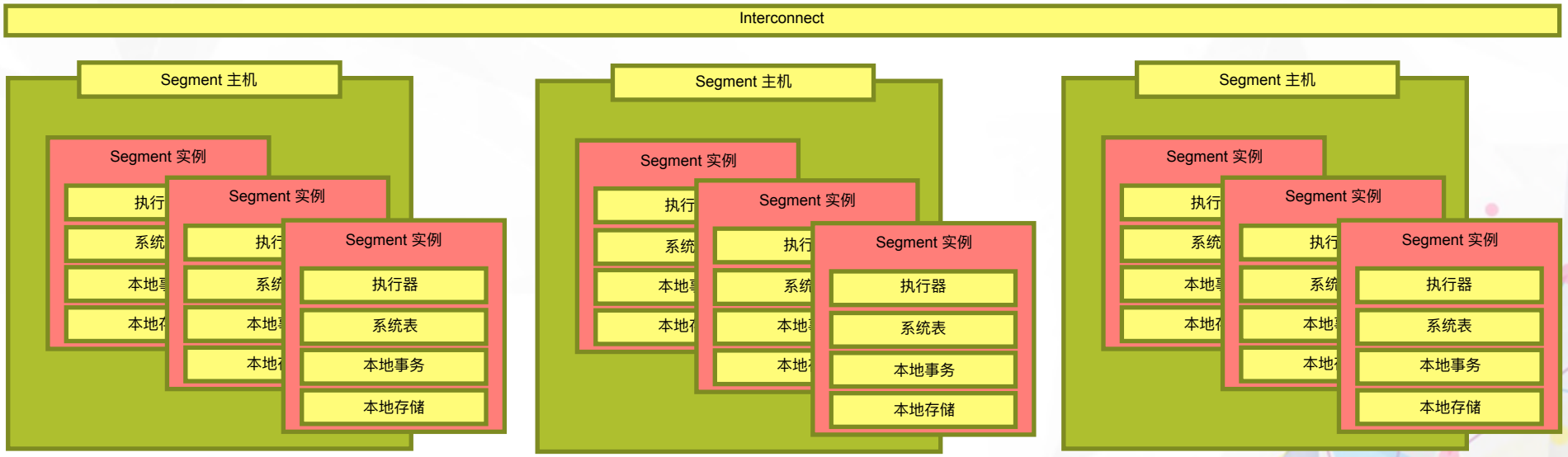
主节点接受客户连接，处理请求，执行认证

优化器

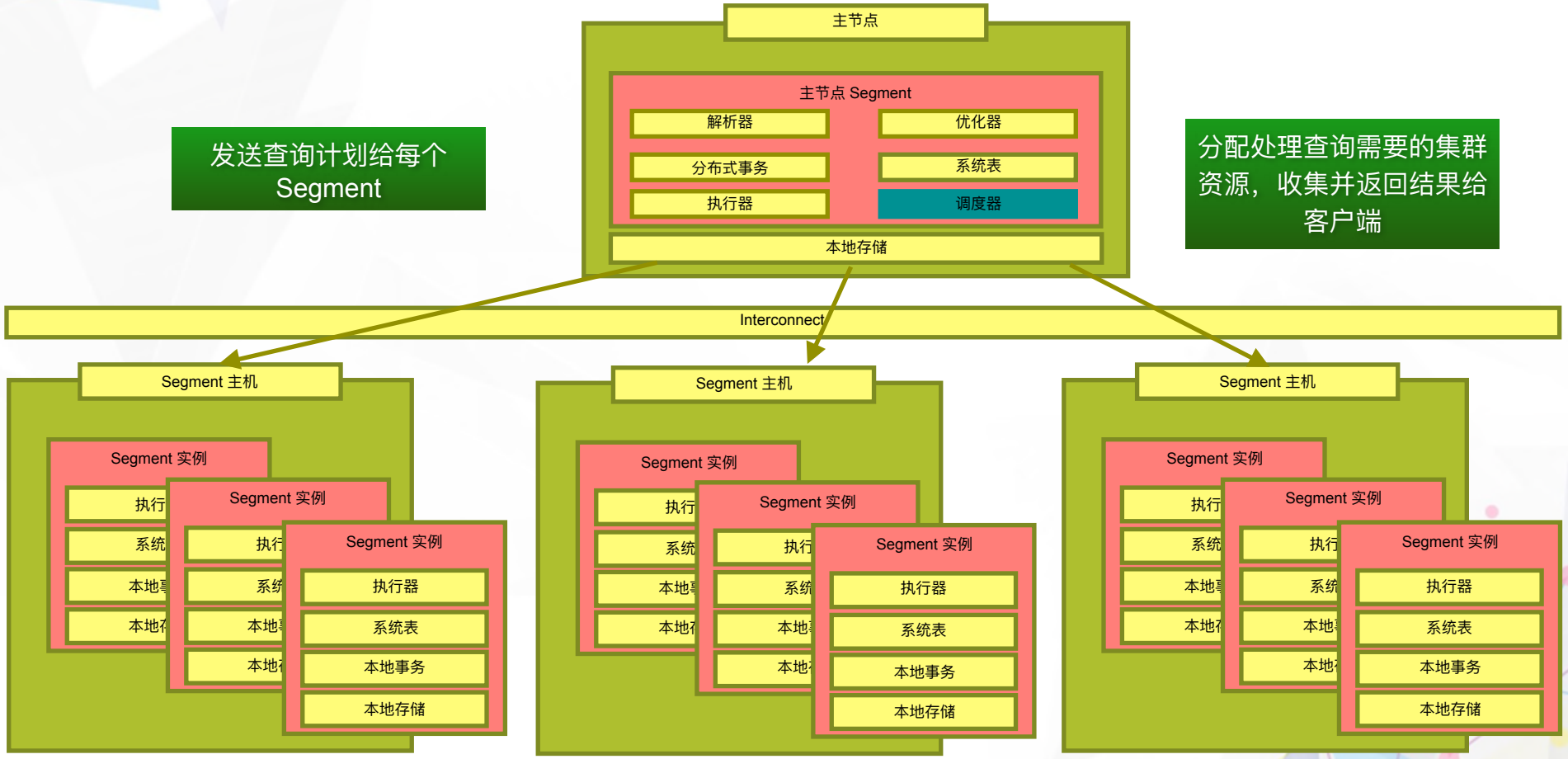
处理解析树，生成查询计划



查询计划描述了如何执行查询

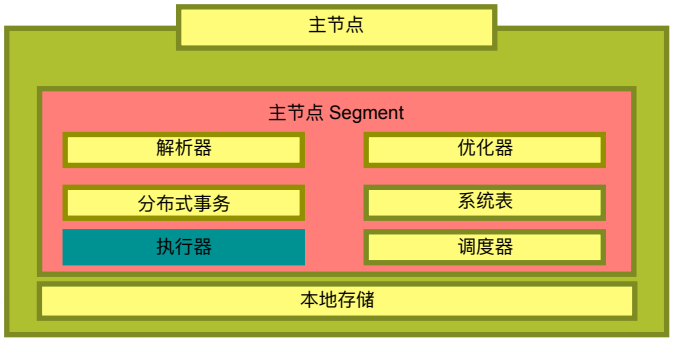


调度器

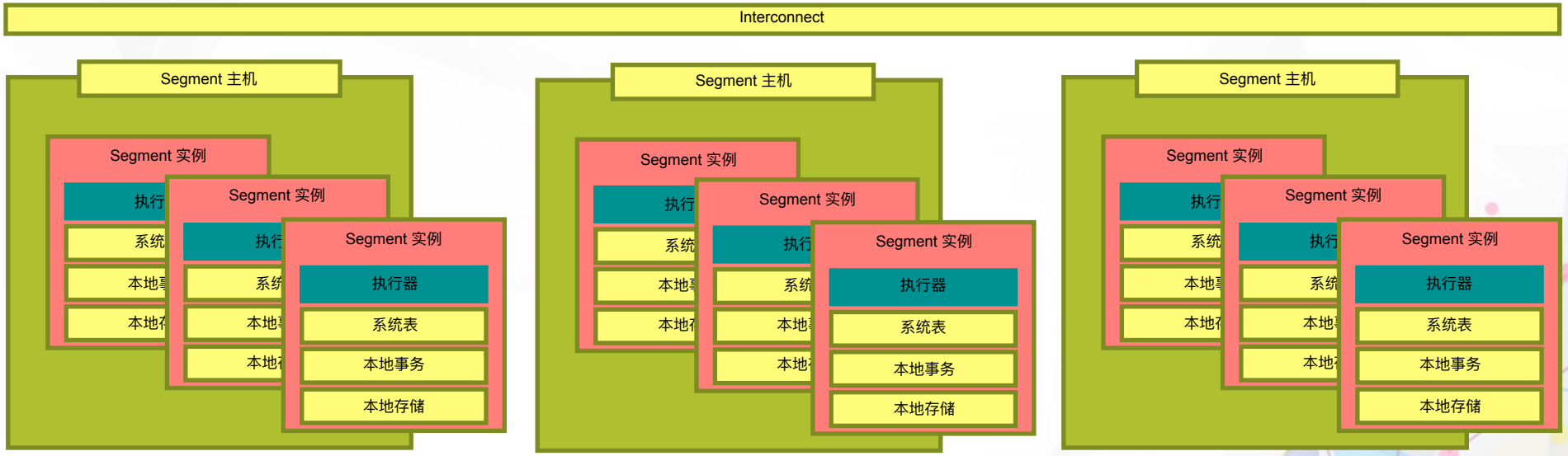

主节点
主节点 Segment
解析器
优化器
分布式事务
系统表
执行器
调度器
本地存储
发送查询计划给每个 Segment
分配处理查询需要的集群资源，收集并返回结果给客户端
Interconnect
Segment 主机
Segment 实例
执行器
系统表
本地事务
本地存储

执行器

发送查询计划给每个 Segment

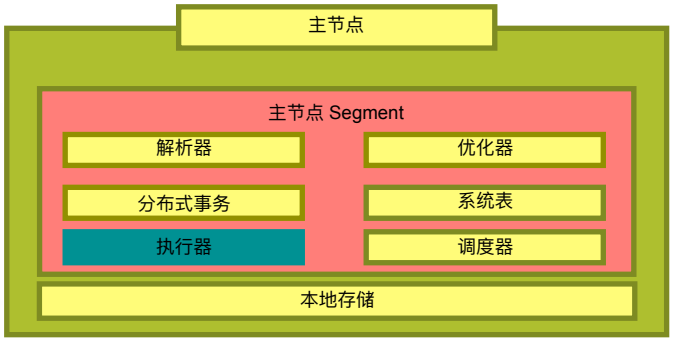


分配处理查询需要的集群资源，收集并返回结果给客户端

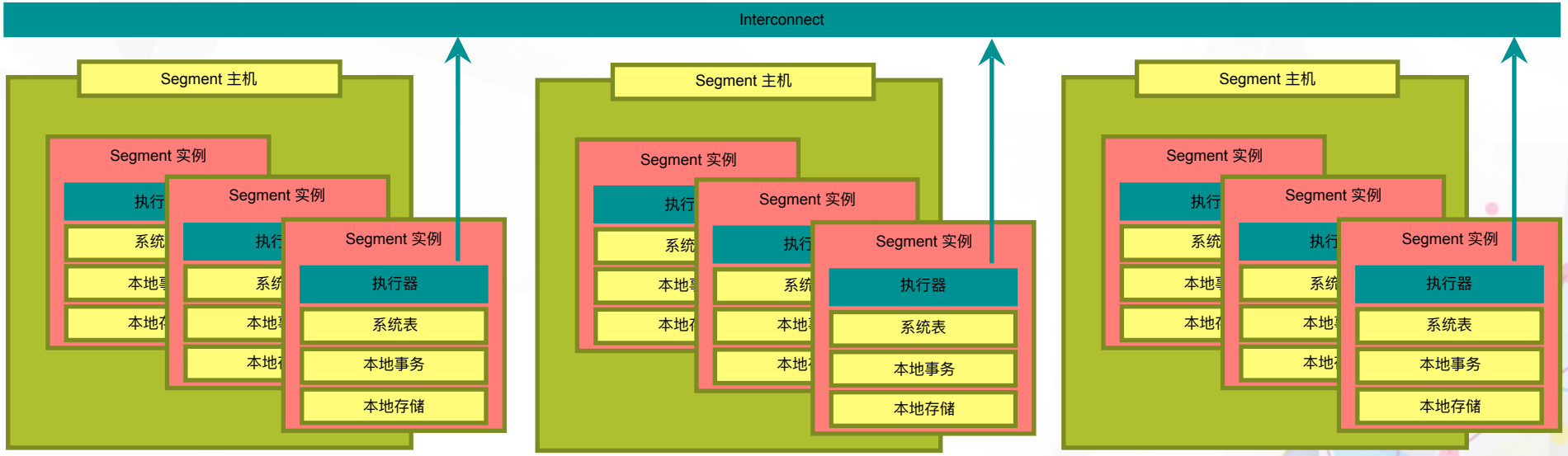


Interconnect

发送查询计划给每个 Segment

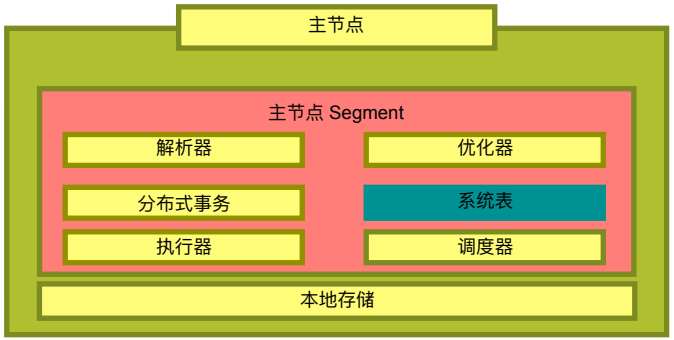


分配处理查询需要的集群资源，收集并返回结果给客户端

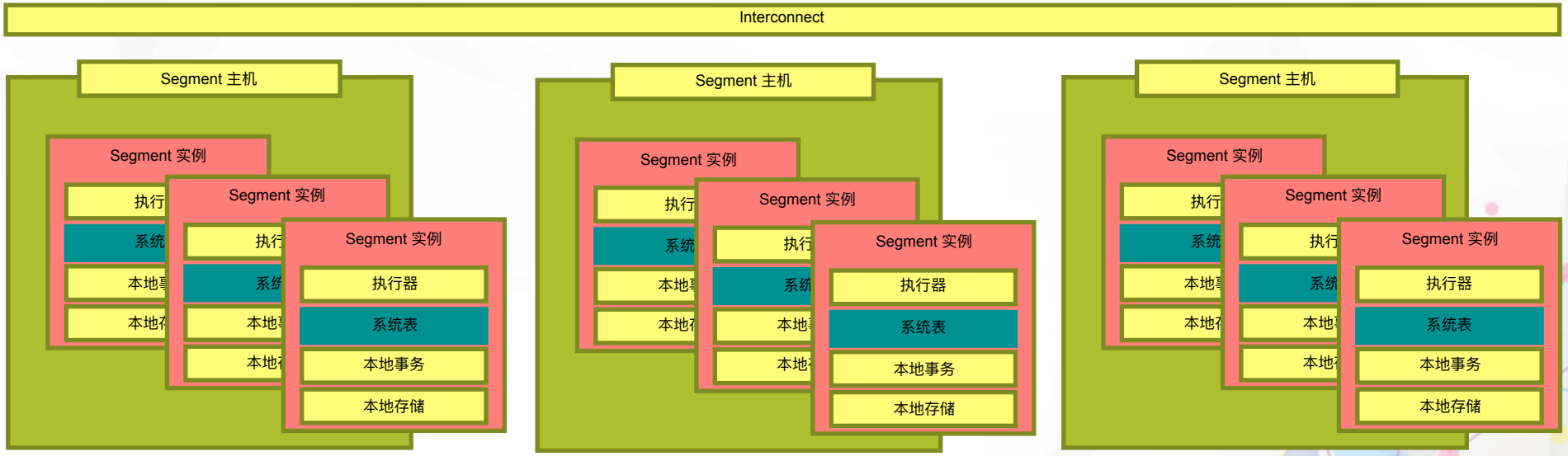


系统表

存储和管理数据库、表、字段的元数据

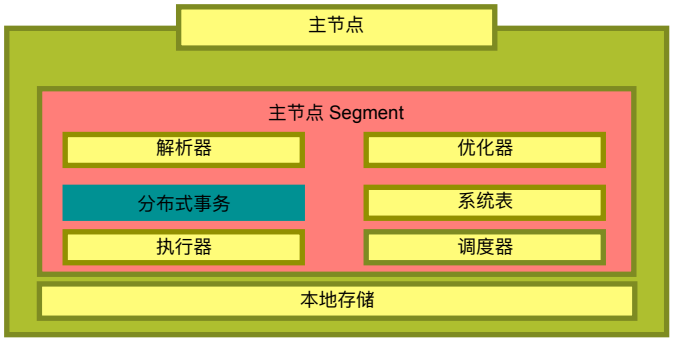


每个节点保存一个拷贝

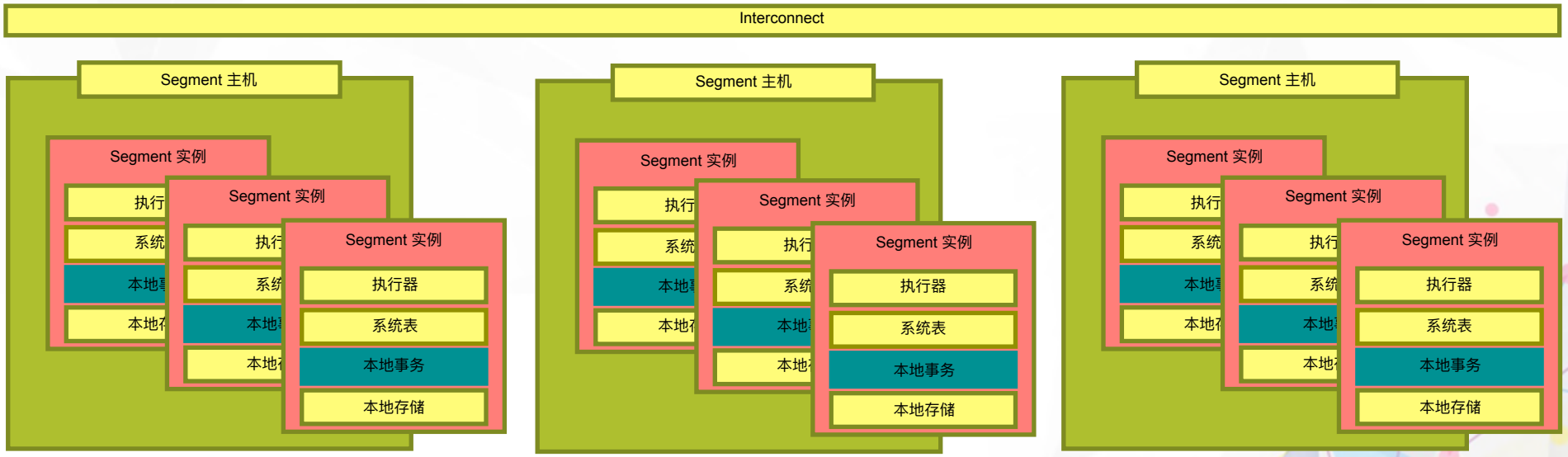


分布式事务

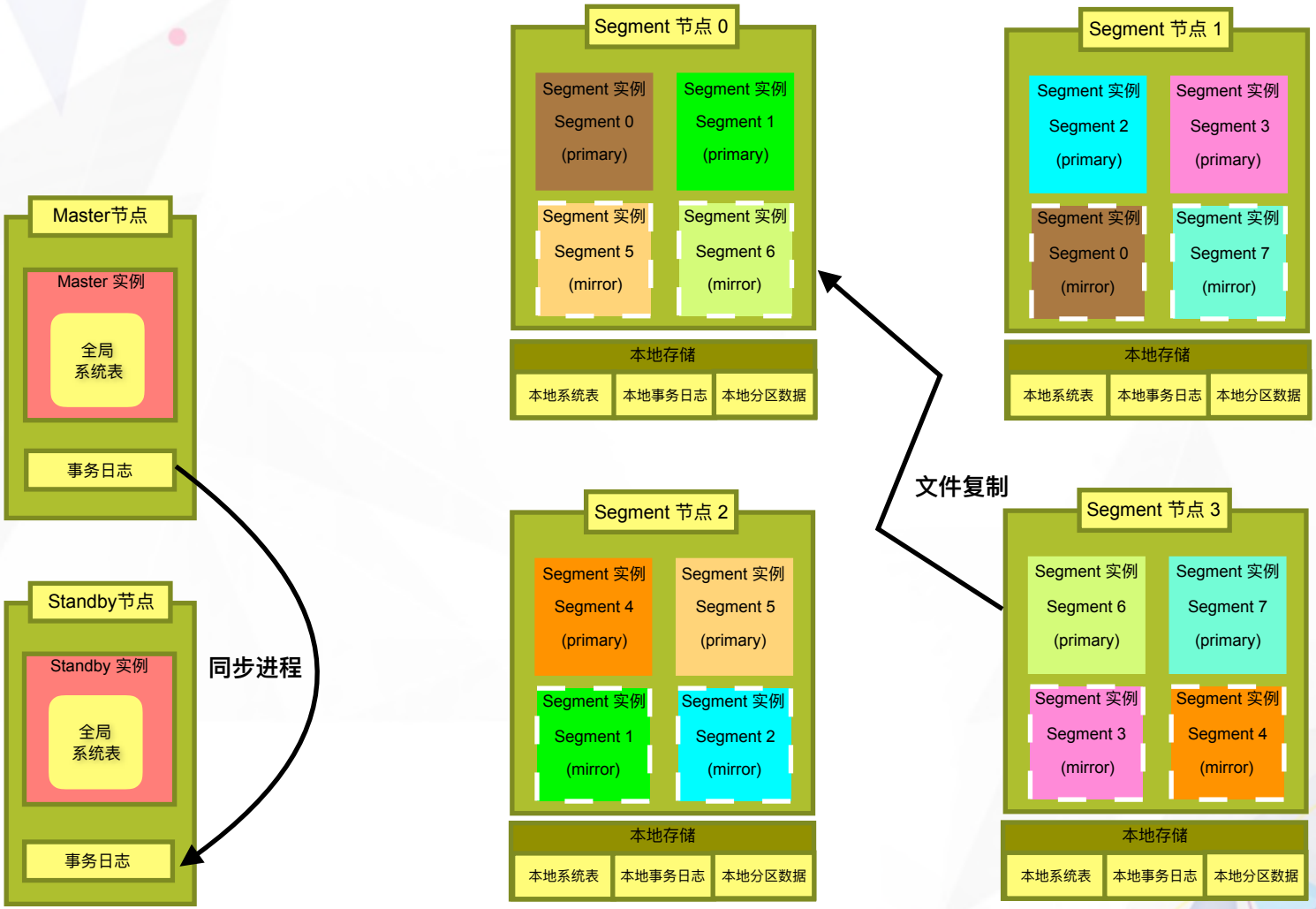
主节点上的分布式事务管理器协调Segment上的提交和回滚操作



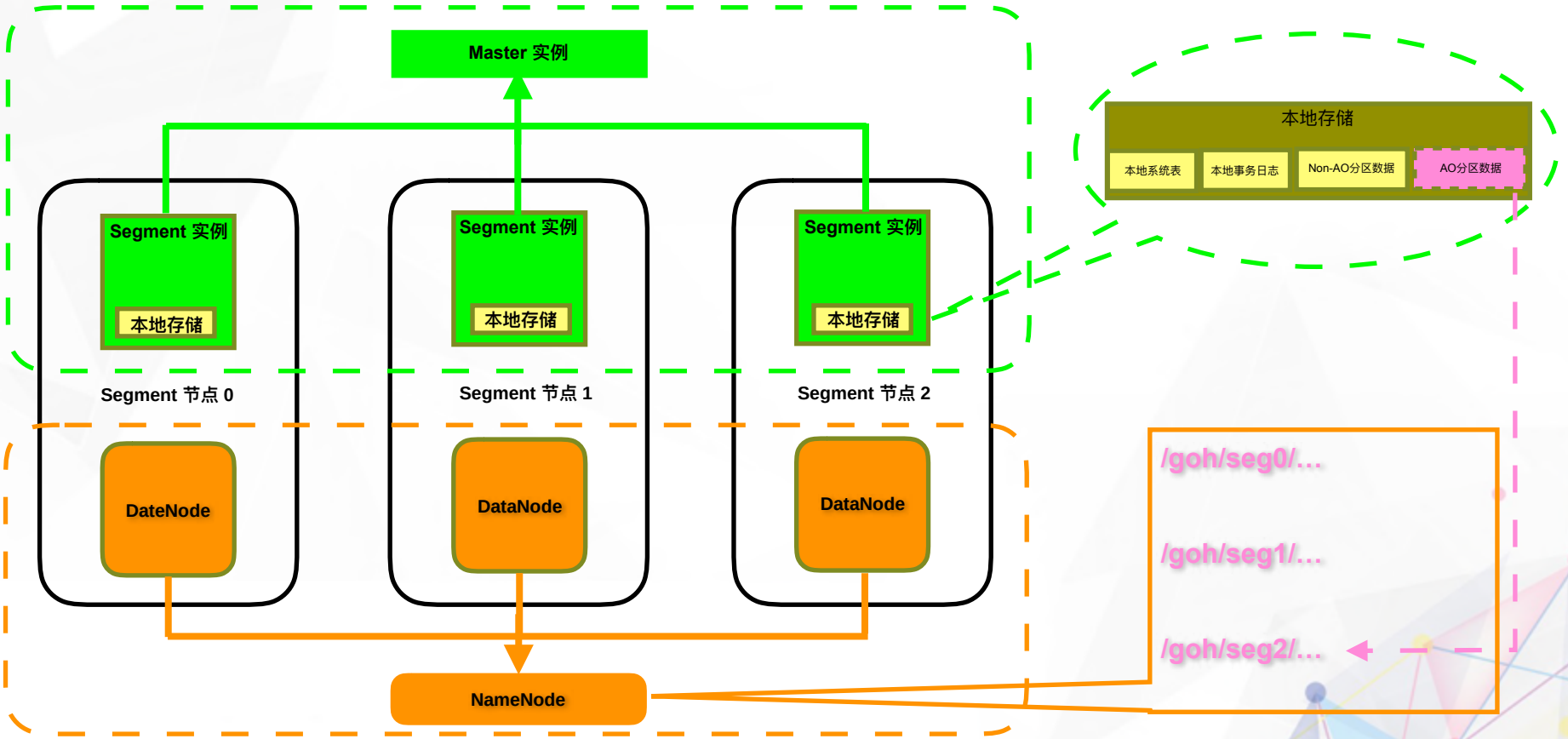
Segment有自己的事务日志，确定合适提交和回滚自己的事务



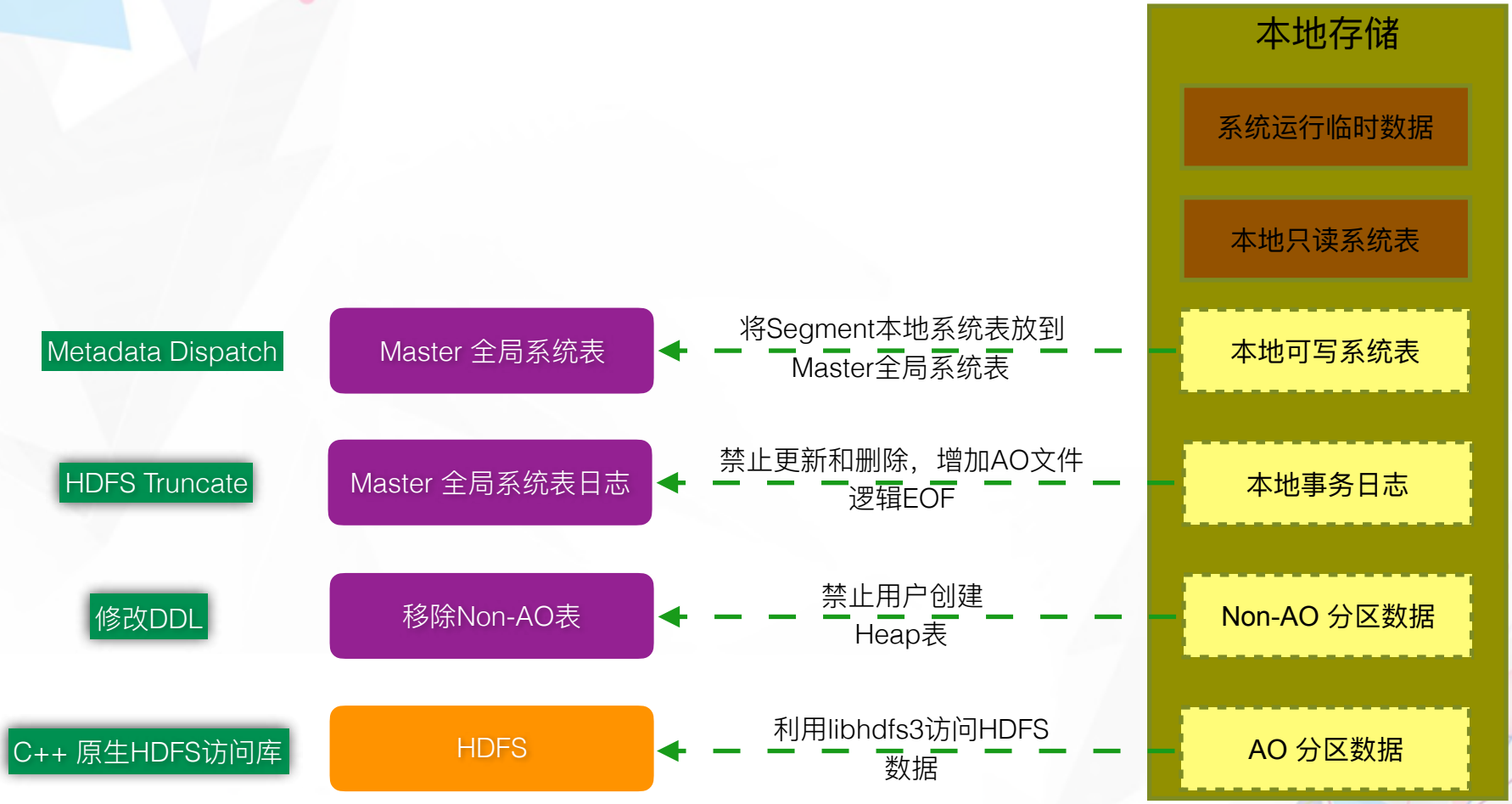
高可用


Master节点
Master 实例
全局
系统表
事务日志
Standby节点
Standby 实例
全局
系统表
事务日志
同步进程
Segment 节点 0
Segment 实例
Segment 0
(primary)
Segment 实例
Segment 1
(primary)
Segment 实例
Segment 5
(mirror)
Segment 实例
Segment 6
(mirror)
本地存储
本地系统表
本地事务日志
本地分区数据
Segment 节点 1
Segment 实例
Segment 2
(primary)
Segment 实例
Segment 3
(primary)
Segment 实例
Segment 0
(mirror)
Segment 实例
Segment 7
(mirror)
本地存储
本地系统表
本地事务日志
本地分区数据
文件复制
Segment 节点 2
Segment 实例
Segment 4
(primary)
Segment 实例
Segment 5
(primary)
Segment 实例
Segment 1
(mirror)
Segment 实例
Segment 2
(mirror)
本地存储
本地系统表
本地事务日志
本地分区数据
Segment 节点 3
Segment 实例
Segment 6
(primary)
Segment 实例
Segment 7
(primary)
Segment 实例
Segment 3
(mirror)
Segment 实例
Segment 4
(mirror)
本地存储
本地系统表
本地事务日志
本地分区数据

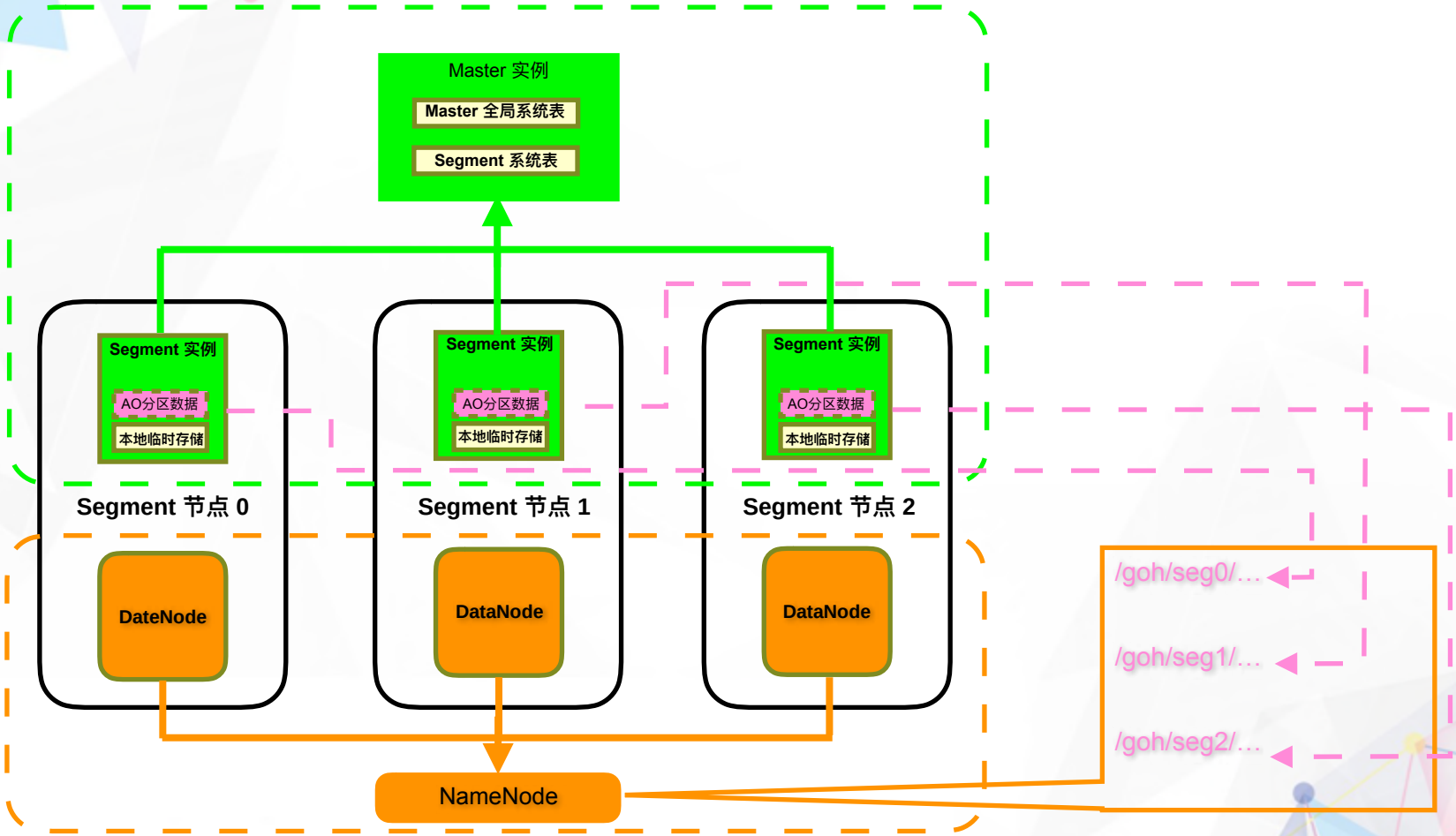
Greenplum on HDFS



HAWQ 1.x

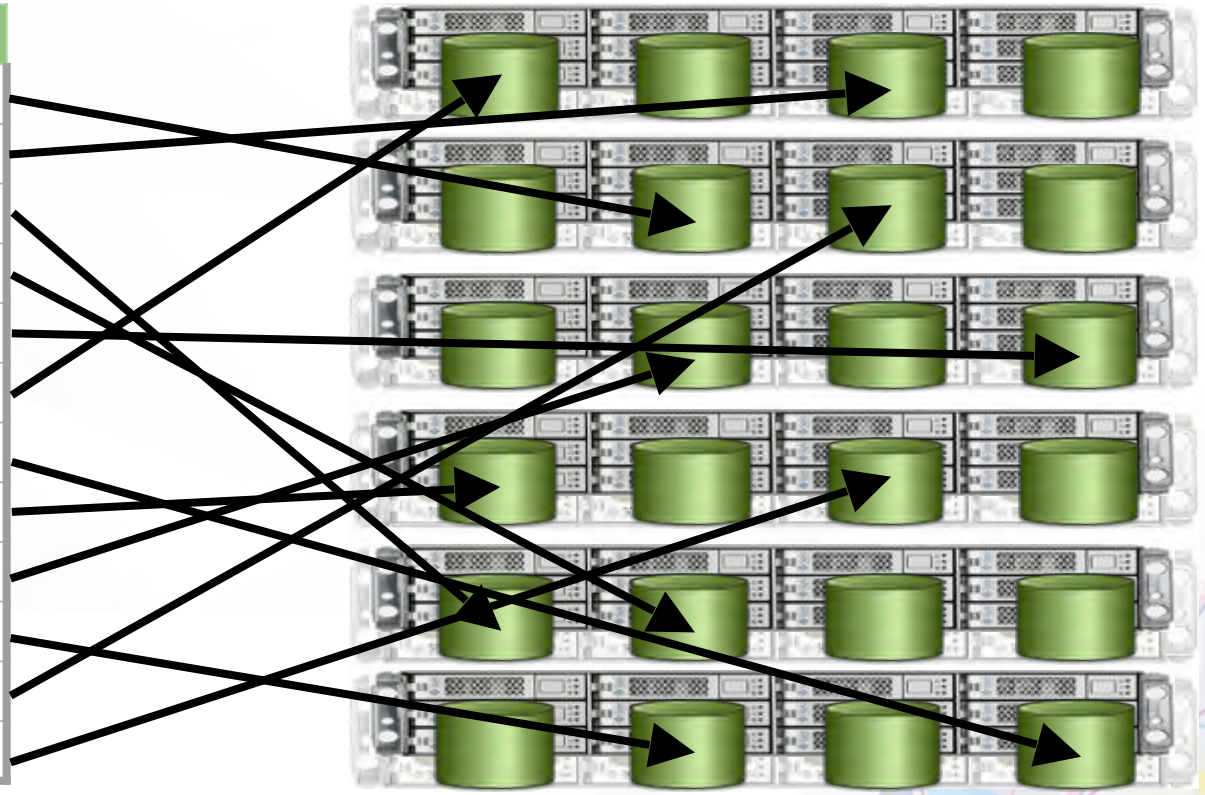

本地存储
系统运行临时数据
本地只读系统表
本地可写系统表
本地事务日志
Non-AO 分区数据
AO 分区数据
Metadata Dispatch
Master 全局系统表
将Segment本地系统表放到Master全局系统表
HDFS Truncate
Master 全局系统表日志
禁止更新和删除，增加AO文件逻辑EOF
修改DDL
移除Non-AO表
禁止用户创建Heap表
C++ 原生HDFS访问库
HDFS
利用libhdfs3访问HDFS数据

无状态Segment

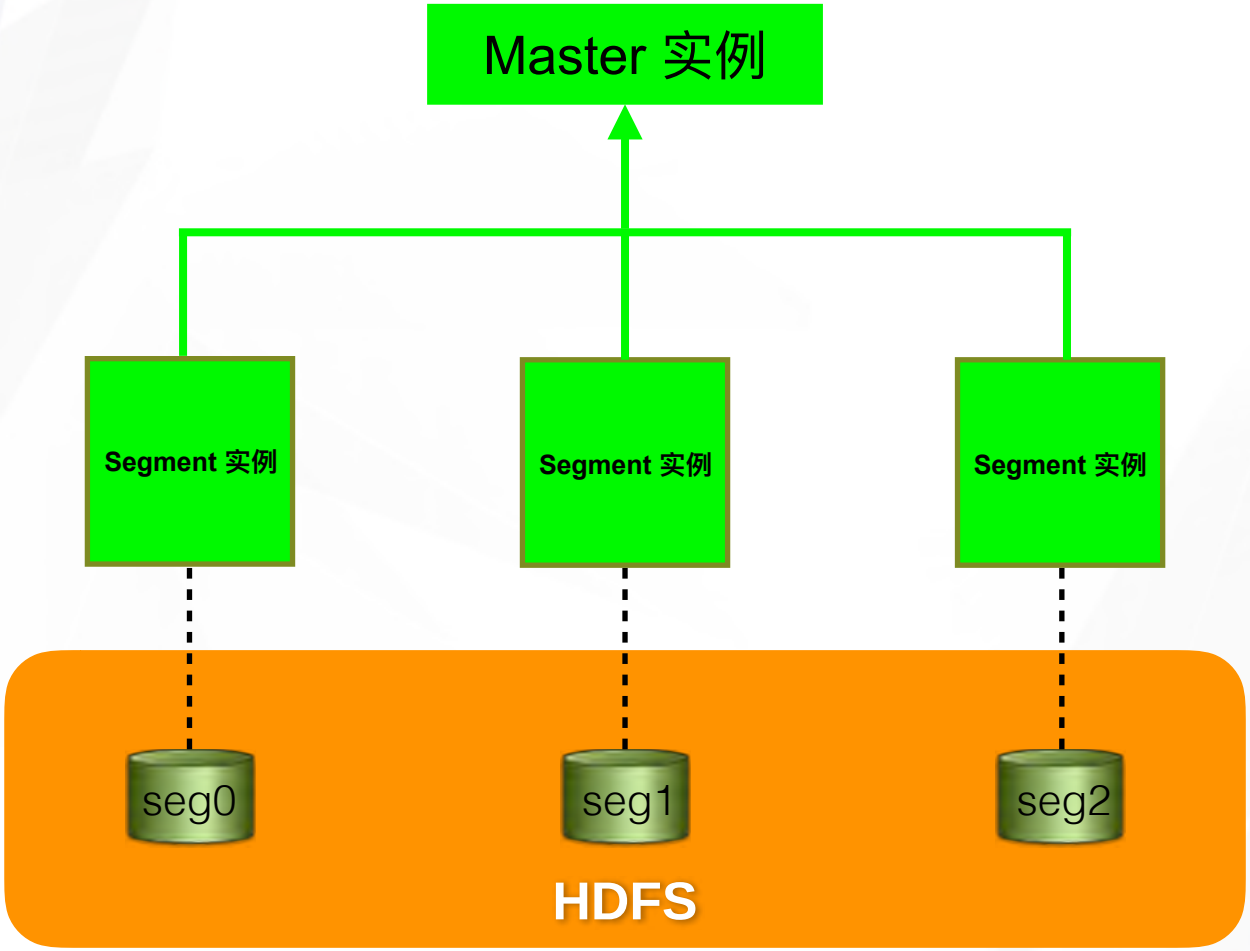

Master 实例
Master 全局系统表
Segment 系统表
Segment 实例
AO分区数据
本地临时存储
Segment 节点 0
Segment 实例
AO分区数据
本地临时存储
Segment 节点 1
Segment 实例
AO分区数据
本地临时存储
Segment 节点 2
DateNode
DataNode
DataNode
NameNode
/goh/seg0/…
/goh/seg1/…
/goh/seg2/…

数据分布：并行处理的基础

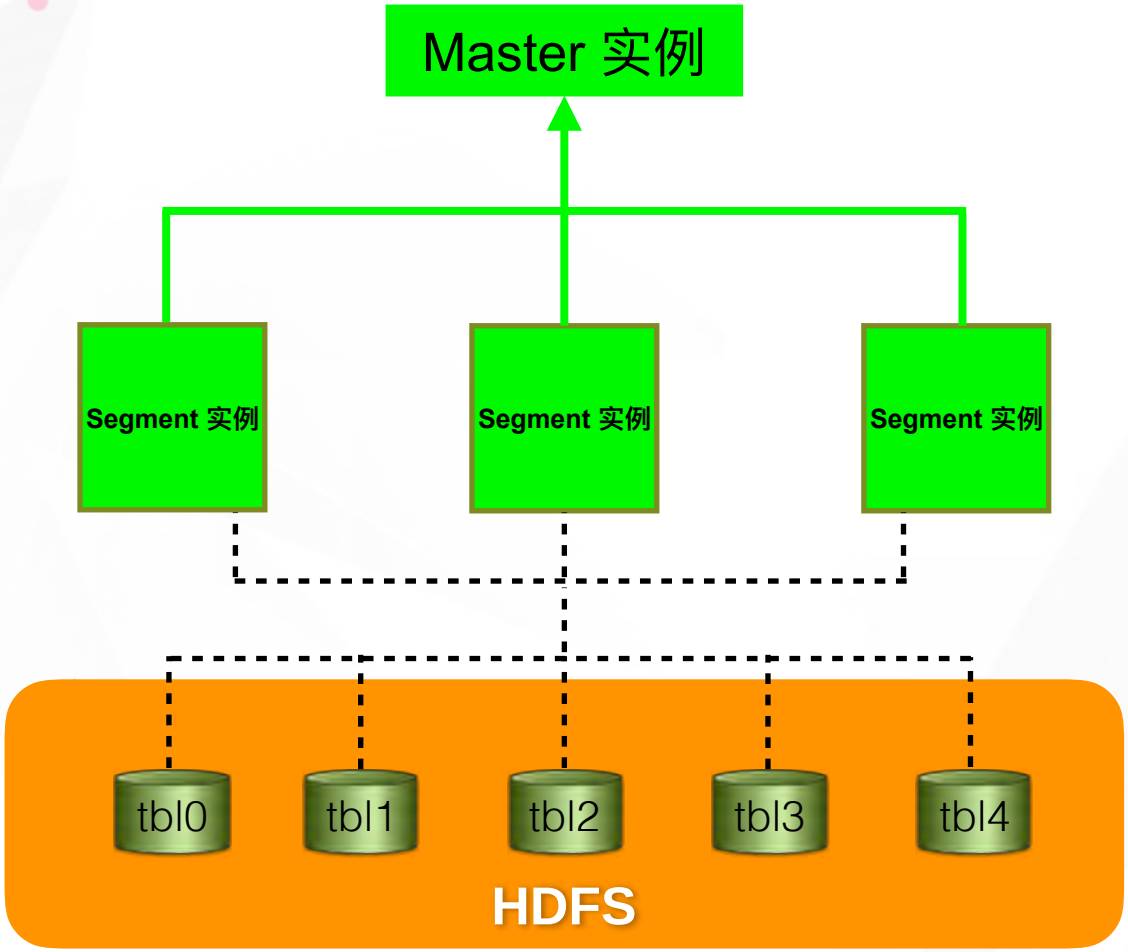
OrderID	OrderDate	CustomerID
43	Oct 20 2005	12
64	Oct 20 2005	111
45	Oct 20 2005	42
46	Oct 20 2005	64
77	Oct 20 2005	32
48	Oct 20 2005	12
50	Oct 20 2005	34
56	Oct 20 2005	213
63	Oct 20 2005	15
44	Oct 20 2005	102
53	Oct 20 2005	82
55	Oct 20 2005	55



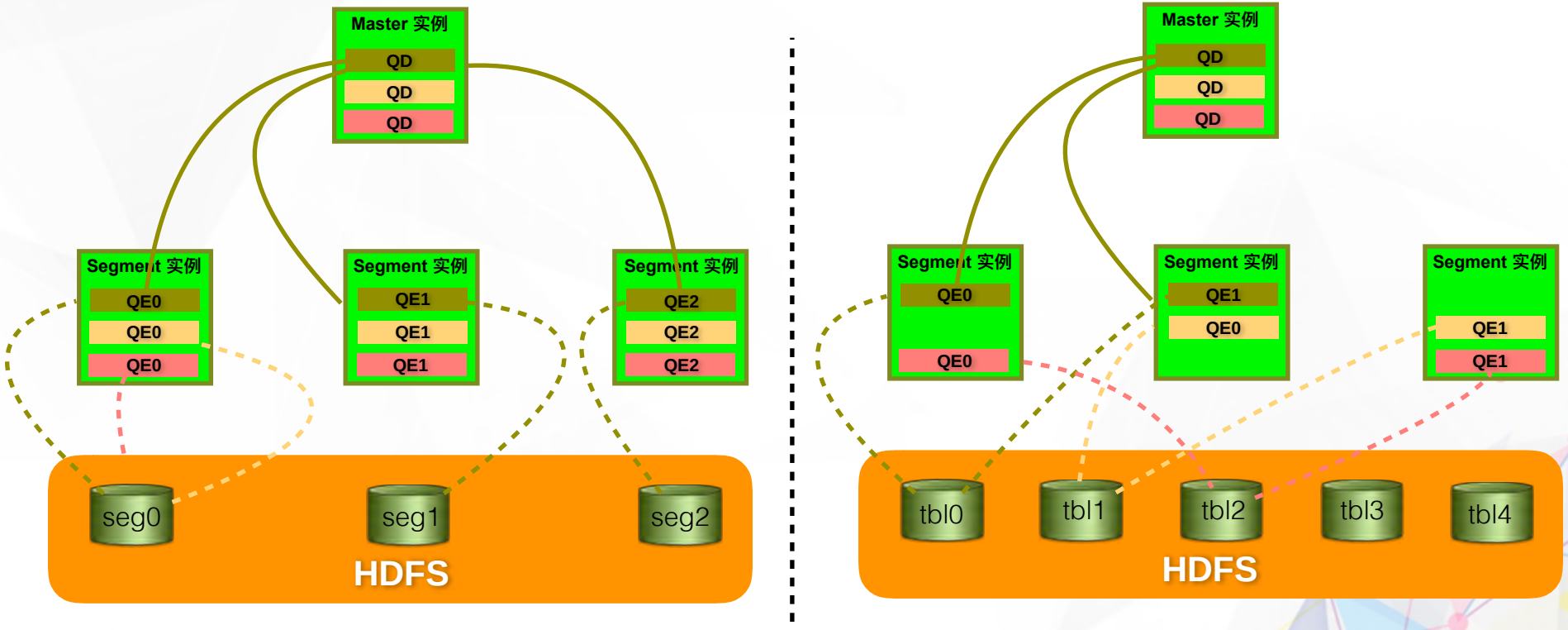
物理上分离，逻辑上集成


Master 实例
Segment 实例
Segment 实例
Segment 实例
seg0
seg1
seg2
HDFS

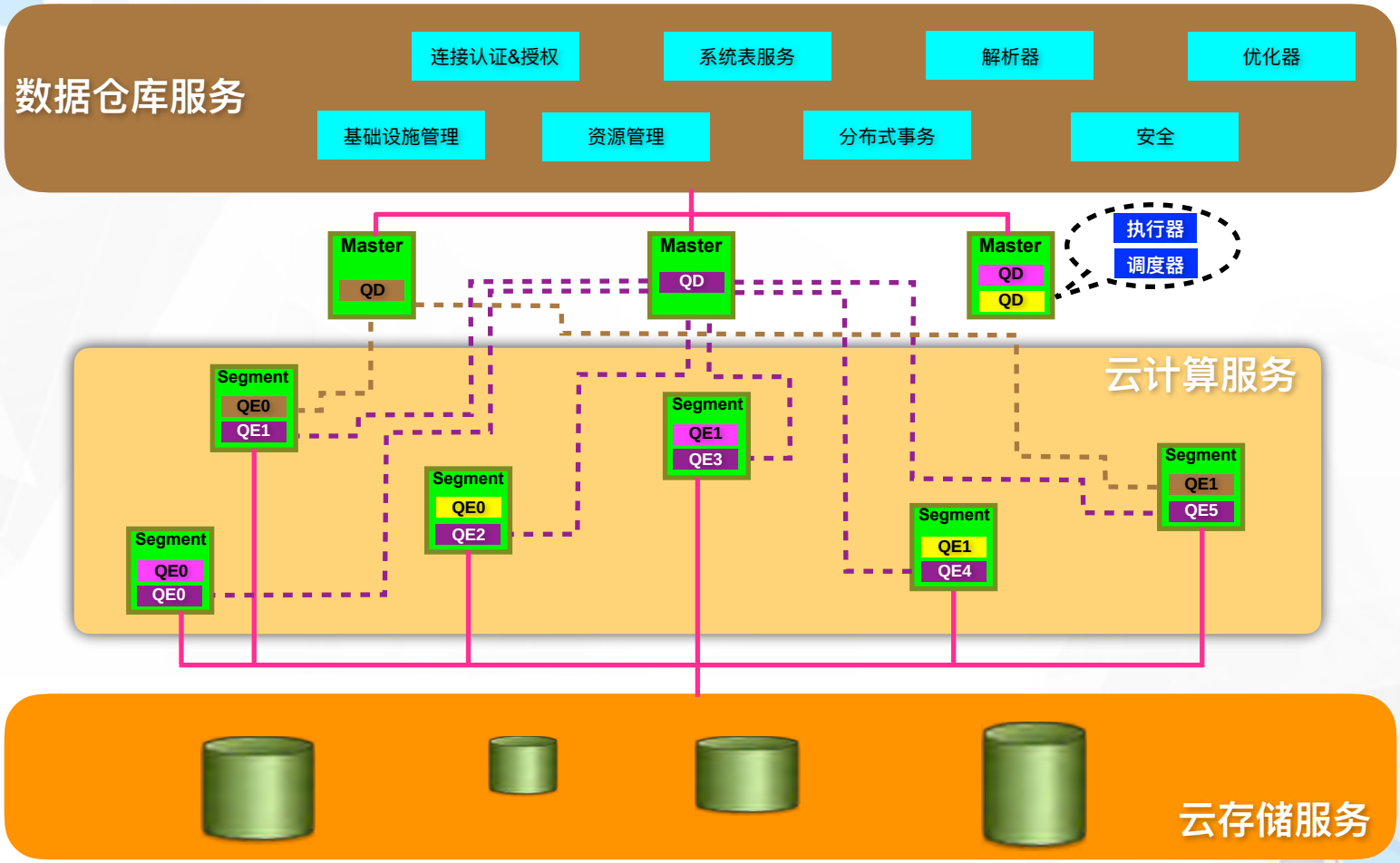
物理上分离，逻辑上也分离



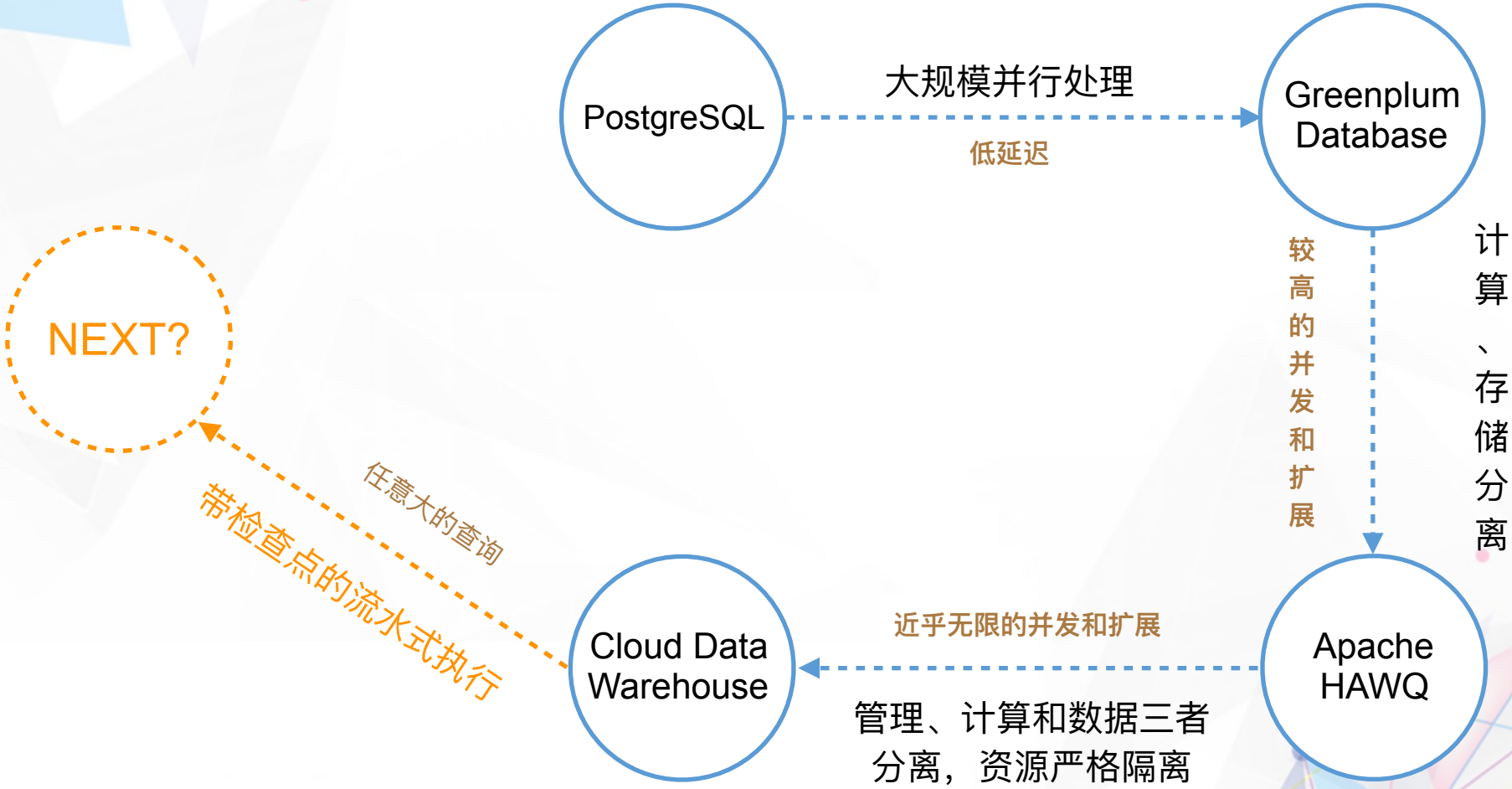
资源调度


Master 实例
QD
QD
QD
Segment 实例
QE0
QE0
QE0
Segment 实例
QE1
QE1
QE1
Segment 实例
QE2
QE2
QE2
seg0
seg1
seg2
HDFS
Master 实例
QD
QD
QD
Segment 实例
QE0
QE0
Segment 实例
QE1
QE0
Segment 实例
QE1
QE1
tbl0
tbl1
tbl2
tbl3
tbl4
HDFS

数据仓库即服务


数据仓库服务
连接认证&授权
系统表服务
解析器
优化器
基础设施管理
资源管理
分布式事务
安全
Master
QD
Master
QD
Master
QD
QD
执行器
调度器
云计算服务
Segment
QE0
QE1
Segment
QE0
QE2
Segment
QE0
QE0
Segment
QE1
QE3
Segment
QE1
QE4
Segment
QE1
QE5
云存储服务

总结


PostgreSQL
大规模并行处理
低延迟
Greenplum Database
较高的并发和扩展
计算、存储分离
Apache HAWQ
近乎无限的并发和扩展
管理、计算和数据三者分离，资源严格隔离
Cloud Data Warehouse
任意大的查询
带检查点的流水式执行
NEXT?

THANKS

SequeMedia 盛拓传媒 | IT168 专注导购16年 | ChinaUnix | ITPUB