

简单聊聊 C4 中的大数据库技术 ZDB2

ZDB2 具备支持未来前沿存储领域能力：

- 分布式存储：对 ZDB2 来说，整合多个数据库是大数据支持的关键技术，因此开辟了许多命名以 Z.ZDB2.Thread****.pas 为主的支持库，这些支持库是将离散的数据条目，统一起来集中管理，且没有 20 亿链表规模限制，没有增删查改的数据结构陷阱
- 没有海量数据支持上限：ZDB2 对离散数据库是没有支持上限的，用 100 万个子数据库来拼成一个大库都没问题
- 没有平台和硬件结构束缚：fpc 可以支持，delphi 可以支持，ZDB2 都可以无改支持
- 支持 hpc：由于 Z-AI 已经浸泡多年 hpc 支持体系，ZDB2 沿用的先进线程调度技术，真正将 hpc 高级计算能力提炼成了数据支持技术
- C4 的所有数据存储，一律使用 ZDB2：在 C4 的 FS（文件碎片），FS2.0（文件碎片），UserDB（用户登录信息数据库），使用了 ZDB2 很小的功能一角
- C4 的网盘支持系统：在该系统中，有一个先进 Cache 技术，它的作用是把已经上传和下载的数据，全部 Cache 下来管理，外面以无感包装，体验则是拉文件直接秒传
- Z-AI 未来对于数据库方案主要使用都会是 ZDB2：不管大估摸图片，分类器，检测器，Z-AI 未来都会使用 ZDB2 作为主要存储方案，但这种方式不会一步到位，会缓慢的进行，慢慢转变。

假定以城市档案的数据量为基础

单人头像=300k

各种证件照=2M

户口电话各种账号等等=100k

一个中等城市 1500W 人口=1500*10000*(0.3+2+0.1)，大约 36TB 数据量

ZDB2 如何保存 36TB 数据

36TB 的数据量对于 hpc 来说非常小，用 ZDB2 方案，是这样来设计和使用它的
首先，分好阵列支持体系，规划好存储路径（Linux），或则驱动器（Windows）

接下来，建立 ZDB2 分散存储，例如，36 个 ZDB2 数据库文件，每个文件预置 2TB 空间
主要支持技术来自 Z.ZDB2.Thread.pas 库，单个数据库使用 TZDB2_Th_Enginea 实例，建立 36 个，每个文件预置 2TB 空间（ZDB2 的单个数据库最大只能 136TB）

下一步，把 36 个 TZDB2_Th_Engine 实例汇集管理，开始从某种源头导入数据
主要支持技术来自 Z.ZDB2.Thread.pas 库，TZDB2_Th_Engine_Marshal 就是汇集支持实例
使用方法，TZDB2_Th_Engine_Marshal.Add_Data_To_Minimize_Size_Engine，头像，照片，各种信息，往这里推，无限的循环，这是支持 HPC 的全流水线处理机制，届时，高配 hpc 性能会被开满，整个阵列系统负载也拉满。
36TB，如果阵列按 600M/s 的物理写入计算，大约 15-17 个小时后，完成数据导入

ZDB2 如何使用 36TB 的使用

建模，分库全免，我们只需要明确，我们要怎么使用 ZDB2 数据库
例如，我们要快速根据身份证号查询 36TB

1. 先创建 36 个 TZDB2_Th_Enginea 实例，把数据给读入进来，不是全读，就是打开文件 IO
2. 在 TZDB2_Th_Engine_Marshal 把数据库给汇集了
3. 启动遍历 TZDB2_Th_Engine_Marshal.Parallel_Load，根据身份证号码，在内存 hash 指向数据位置。这一步，相当于建立查询索引，一旦启动 Parallel_load，hpc 的 cpu 和阵列均会拉满，IO 不会读整库，要比建库快上百倍
4. 在服务器里面直接给查询实现就可以了

在 C4 中，对于 ZDB2 的运用无处不在，机制简洁，代码精炼，稳稳大数据支持。

by.qq600585

2022-10-15