# Google Bigtable 学习笔记 Bigtable 学习笔记

本或**bttps://www.ac.**yoidcn.com/)

时规心头注本站公众号,获取更多程序园信息



5%A4%A7%E6%95%B0%E6%8D%AE) 栏目 Google (http://www.voidcn.com/column/google ) ails/8435230

几个模块之上。Bigtable 使用 Google 分布式文件系统 GFS 来存储日志和数据文件。一个在同一批计算机上,Bigtable 进程与这些分布式应用的进程共享同样的计算机。Bigtable 依赖于机器故障处理以及机器状态监控。

的数据。SSTable 是一种持久的、排序的、不变的键值映射文件,其中的键值是任意的字节串。 通历制定范围的键值等。每个 SSTable 包含 64K 大小的 block 序列,并用 block 索引来定位 SSTable 文件被打开的时候 block 被装入内存。一次查询只需一次磁盘读写:首先在内存里通过 读取 block。另外,SSTable 还能完全映射到内存,这样不用读取磁盘就可以完成查询。

程序设置依赖于名为 Chubby 的高可用、分布式、持久化锁服务。一个 Chubby 服务包含 5 个活跃复本,其中一个选为 master 并响应查询。只有在多数是各有语行产品地河用互相通讯的情况下,Chubby服务才可用。Chubby 采用 paxos 算法复本之间在面对故障情况下的一致性。Chukhty操供wvople可见的名字空间。每个目录或者文件可用用作锁,对文件的读写时原子的。Chubby 客户程序库提供关于Chuhbw,这样的预数的编码。每个 Chubby 客户与服务器维持一个 session。Client 如果不能在一定时间内更新其租期,将导致 session 将过期,处于基期争争的和创展。但hubby 客户端还能注册回调函数,以便在文件目录变更或者 session 过期时得到通知。Bigtable 使用 Chubby来完成许多证券,这种保证的时候是多只有一个活跃的 master;存储 Bigtable 数据的加载地址;发现 tablet 服务器以及宣告 bablet 服务器的终止;据看。Bigtable,的模式信息(每个表的列族信息);存储访问控制列表;如果 Chubby 服务长时间不可访问,那么 Bigtable 也不能访问。

• 3. 《Google-Bigtable》读书笔记

(http://www.voidcn.com/article/p-

• 6. bigtable笔记

(http://www.voidcn.com/article/p-

我们来最高的结构tr(类似B+树)来存储 tablet 位置信息。第一层是存储在 chubby 中的一个文件,它包含 root tablet 的位置信息。而 root tablet 包含了一些用户 tablet 的表的所有 tablet 的位置信息。每个 METADATA tablet 包含了一些用户 tablet 的信息。Root tablet 其实是METADATA,表的第15页。由于12页,以外,1

hvakauhl-vm.html)

• 8. BigTable学习

METANATA 法不够估价的有限的。 由于 METANATA 是一个 METANATA METAN

(http://www.voidcn.com/article/p-

Masker 的表情的,服务器停止响应的时刻,并负责尽快重新分配该 tablet 服务器所服务的 tablet 。为了探测 tablet 服务器的问题,master 定期地向每个 tablet 服务器询问其锁的状况。如果 tablet 服务器报告锁丢失,或者 master 联系不上 tablet 服务器,master 尝试获取该服务器的独思统制为单元分配,说明 chubby 正常,而 tablet 服务器死机或者不能访问 chubby ,所以 master 删除该锁文件,这样对应证明的代据务器再也不能服务相关 tablet ,master 就可以将这些 tablet 纳入未分配 tablet 集合。如果 master 不能访问 chubby 则自动终止,master 对时间 tablet 的分配情况。

当 master 启动的时候,他需要首先发现目前 tablet 分配情况。它执行以下步骤: (1) 从 chubby 获取一个唯一的 master 锁,从而避免多个 master 并发; (2) 扫描 chubby 中的服务器目录找到所有 tablet 服务器; (3) 询问每个 tablet 服务器获知 tablet 映射情况; (4) 扫描 METADATA 表构建未映射 tablet 集合; 但是 METADATA 表的 tablet 必须先分配 master 才能扫描 METADATA 表。所以,在 (4) 之前必须先将 root tablet 分配给某个服务器,如果它还未分配的话。因为 root tablet 包括了 METADATA 表的所有tablet 信息,在扫描 root tablet 之后 master 就知道所有 METADATA tablet 了。

所有现存。Master 知道这些变化,因为其中前三个都是由 master 发起的。Master 知道这些变化,因为其中前三个都是由 master 发起的。 Tablet 的分裂是由 tablet 服务器发起的,所有特别处理:tablet 通过在 METADATA 表中登记新的 tablet 来提交分裂结果,当提交成功后它通知 本語(Latt),其一(Labet Labet 



新提交到 commit 日志以保持重做记录。在这些更新中,最近的更新存储在内存的一个叫 储在一系列 SSTable 文件中。如果要恢复一个 tablet , 服务器从 METADATA 表读它的元数据, 訂,重做点指向包含该 tablet 的 commit 日志。Tablet 服务器从 SSTable 中读取这些指标,利用 包含重做点之后的所有更新。当写请求到达时,服务器检查格式正确性及读权限。权限检查通过 合。合法的修改请求记录到 commit 日志,多次小修改可以一次提交。写提交之后,其内容插入 权限,并在由 SSTable 和 memtable 合并的视图上查询。由于他们都是排序的,效率较高。在

,增长超过一定阀值时,老 memtable 冻结并生成一个新 memtable。冻结的 memtable 转换成 个目的:减少内存的使用,减少故障恢复是需要重做的 Log 记录。次级压缩过程中读写能同时进

S加限制,读操作需要从大量 SSTable 中归并结果。通过在后台周期性地执行归并压缩可以避免 **涵条**诃靈。归并压缩将多个 SSTable 和 memtable 压缩成一个 SSTable。老的 SSTable 和 memtable 在归并压缩完成之后即可丢弃。

• 1 Google BigTable 学习笔记。 这个归并压缩过程叫做 major 压缩。次级压缩产生的 SSTable 中包含一些标志已删除的条目,major 压缩过程将丢弃这写删除条目。 (http://www.voidcn.com/article/p-

#### 优化措施zyfq-bme.html)

使用压缩Gold每个Bildestallet》(读书笔记以指定压缩方式。

(http://www.voidcn.com/article/p-使用缓存: \_ 两级缓存, SSTable 和 Tablet Server 之间的键值对缓存, 以及 GFS SSTable 块的缓存。 が vqtbvnfi-dr.html)

使用4BGoogleilBigTa为e阅读等记roup指定使用过滤器。

## <u>/(http:</u>//www.voidcn.com/article/p-

- 文章 kzidwyai-rr.html) 1. Google BigTable 学习笔记 (http://www.voidcn.com/article/p-hpwpzyfq-bme.html) 5. NoSQL 学习笔记(三)之BigTable 2. 论文学习笔记: BigTable (http://www.voidcn.com/article/p-kgmxeiiu-uk.html)
- /www.voidcn.com/article/b-Google-Bigtable》读书笔记 (http://www.voidcn.com/article/p-vqtbvnfi-dr.html)
- oi-yu.htmi) gle gigTable阅读笔记 (http://www.voidcn.com/article/p-kzjdwyai-rr.html)

- . Digitable美元 NoSQL学习笔记(三)之BigTable (http://www.voidcn.com/article/p-fwcqdaoi-yu.html) http://www.voidcn.com/article/p-. Bigtable笔记 (http://www.voidcn.com/article/p-vybwgfow-bah.html) ybwgfow-bah.html) . Google Bigtable (http://www.voidcn.com/article/p-hvakauhl-vm.html)

- 7. Google Bigtable (http://www.voiucn.com/article/p-hvakauhl-vm.htm?) (http://www.voidcn.com/article/p-hijbzpwa-hn.html) (http://www.voidcn.com/article/p-9. BigTable笔记 (http://www.voldcn.com/article/p-mzoncchg-bs.html) hvakauhl-vm.html)
- 笔记 (http://www.voidcn.com/article/p-explyamb-bku.html)

#### 相关标签/搜索hn.html)

- 9学的自由16年記/www.voidcn.com/tag/%E5%AD%A6%E4%B9%A0%E7%AC%94%E8%AE%B0) 学习笔记" (http://www.voldcn.com/tagt/%eff5%AD%A6%E4%B9%A0%E7%AC%94%E8%AE%B0%E2%80%9D) 学习笔记 (https://dom/wbsg/tdgnj).com/tag/%E5%AD%A6%E4%B9%A0%E7%AC%94%E8%AE%B0%E3%80%80) 笔记学习
- (http:// (http://www.voidcn.com/tartj/thedgren%E7%AC%94%E8%AE%B0%EF%BC%8C%E5%AD%A6%E4%B9%A0%E7%AC%94%E8%AE%B0) e菜和/a帶知識记html)

(<u>http://www.vo</u>idcn.com/tag/%E5%AD%A6%E4%B9%A0%EF%BC%8C%E5%AD%A6%E4%B9%A0%E7%AC%94%E8%AE%B0)

>>更多相关文章<< hadoop学习笔记(http://www.voidcn.com/tag/hadoop%E5%AD%A6%E4%B9%A0%E7%AC%94%E8%AE%B0)C++ 学习笔记 (http://www.voidcn.com/relative/p-(http://www.voidcn.com/tag/C%2B%2B+%E5%AD%A6%E4%B9%A0%E7%AC%94%E8%AE%B0)iOS学习笔记

kzkxvxup-zd\_html) (http://www.voidcn.com/tag/iOS%E5%AD%A6%E4%B9%A0%E7%AC%94%E8%AE%B0) PHP学习笔记

(http://www.voidcn.com/tag/PHP%E5%AD%A6%E4%B9%A0%E7%AC%94%E8%AE%B0) BigTable

(http://www.voidcn.com/cata/1245793) bigtable (http://www.voidcn.com/cata/5664057) google closure 学习笔记

(http://www.voidcn.com/cata/1254786) 学习笔记 (http://www.voidcn.com/cata/2520027) 学习笔记 (http://www.voidcn.com/cata/1760887)

学习笔记 (http://www.voidcn.com/cata/1351055) 学习笔记 (http://www.voidcn.com/cata/1411286) 学习笔记

(http://www.voidcn.com/cata/1160811) 学习笔记 (http://www.voidcn.com/cata/1917873) 学习笔记 (http://www.voidcn.com/cata/363026) Google (http://www.voidcn.com/column/google) onvif学习笔记 (http://www.voidcn.com/search/stvqcd) Flink学习笔记

(http://www.voidcn.com/search/effcbl) pcl学习笔记 (http://www.voidcn.com/search/crworm) rnn学习笔记 (http://www.voidcn.com/search/wshnyu) ros学习笔记 (http://www.voidcn.com/search/ajqesv) vagrant学习笔记 (http://www.voidcn.com/search/nvgzxq) mxnet 学习笔记 (http://www.voidcn.com/search/nvgzxq) mxnet 学习笔记 本述 (http://www.voidcn.com/search/fnmjlo) caffe 学习笔记

欢迎卷油本站公众号。获取更多程底围信息



6分享到微博

**№**分享到微信

♣ 分享到QQ

努力的曾经。

ecore查询中的破折号吗? (http://www.voidcn.com/article/p-aljjgcxb-bur.html)
Task或ServiceImpl中的线程抛出异常 (http://www.voidcn.com/article/p-ycookfyw-bur.html)
的Python GUI工具包 (http://www.voidcn.com/article/p-sjytnunc-bur.html)
(http://www.voidcn.com/article/p-emcclanf-bur.html)

/lainFormOnTaskBar属性 (http://www.voidcn.com/article/p-sesakcgl-bur.html) 定位不起作用? (http://www.voidcn.com/article/p-xvukuuze-bur.html)

년국(한왕) ascript-events -javascript internals. 如何实现事件? (http://www.voidcn.com/article/p-nvebeymv-bur.html)

- 8. c- 混合decltype和enable\_if (http://www.voidcn.com/article/p-wigvqdiz-bur.html)
- 1. Google Big Table 子グ表に
   9. Google Dig Tab

### 公众号推荐学习笔记t) BigTable

一个健康类的公众岛,c改追共进/p-

kgmxeiiu-uk.html)



 9. BigTable笔记 (http://www.voidcn.com/article/pmzoncchg-bs.html)

• 10. bigtable论文笔记

意见**依馈//www.voidcn.com/search**) 最新文章 (http://www.voidcn.com/recent) 站长统计(https://www.voidcn.com/search) 最新文章 (http://www.voidcn.com/recent) 站长统计(https://www.voidcn.com/search) 最新文章 (http://www.voidcn.com/recent) 站长统计

#### >>更多相关文章<<

(http://www.voidcn.com/relative/p-kzkxyxup-zd.html)