

延云

千亿大数据实时解决方案

2015/11/07

延云云计算技术有限公司
大数据的未来在大索引,大索引的未来在延云

议程

第一节

- 传统常规解决方案所面临的问题。
- 大索引技术对企业大数据的影响。

第二节

- 关于我们：延云云计算。
- 公司主要产品：YDB的介绍。

第三节

- 应用场景-YDB能做什么？
- YDB与常见的几个系统的优势。

传统常规解决方案所面临的问题

超**千亿规模**的数据，数据库根本就运行不了，怎么办？

数据从产生到能够查询，要延迟一天才能看到，如何能做到**分钟级延迟**？

50台规模的hadoop集群，几亿条数据，一个MR任务要运行几小时，每天也就能进行几百次查询。

如何能让任务的执行时间缩短到**秒级响应**，每天能执行**千万次查询**。

Hbase只接受KV形式的存储，数万个维度的大宽表，如何进行**多维索引**？

Storm流计算能预计算固定的维度、粒度，但业务千变万化，突发事件很多，如何对**任意维度**的组合进行筛选、钻取、统计？

硬盘坏了，机器宕机，怎样做到**数据可靠不丢失**？

小型机太贵，我们买不起，怎么办？

大索引技术对企业大数据的影响

为什么要使用大索引？使用后会有什么好处？

1. 索引大幅度的加快数据的检索速度。
2. 索引可以显著减少查询中分组、统计和排序的时间。
3. 索引大幅度的提高系统的性能和响应时间，从而节约资源。

正因为大索引技术可以显著的降低大数据的处理成本，显著提高大数据系统的执行效率，延云自主研发了自己的分布式大索引系统YDB。

大数据的未来在大索引，大索引的未来在延云。

公司简介

- 延云云计算技术有限公司：

国内第一家“**千亿级实时多维分析**”解决方案提供商。

旨在为企业提供千亿级大数据的高可靠、低成本、高性能的实时大数据解决方案。

我们提供从咨询、设计、系统部署、软件开发、到运维管理及后续升级改造等全套解决方案和专业服务

- 我们的核心产品YDB:

是我们自主研发的一个大型分布式索引系统。旨在为总量在万亿级别、每天千亿级别数据增量的项目提供近似实时的数据导入，并提供近似实时响应的多维查询与统计服务。

我们的团队

- 延云的核心开发团队成员大多出自阿里腾讯的高级开发工程师，大多都有超10年的从业经验，也有一些留英归来的硕士。
- 团队成员中也不乏一些开源项目的核心committer，如 [JStorm](#) 与 [Mdrill](#)。
- 我们的成员曾在阿里腾讯内部设计出多个千亿规模的大数据系统，部分项目在业界也是大名鼎鼎。

11月19日阿里巴巴JStorm正式成为Apache Storm的子项目

10月7日腾讯Hermes集群单日数据规模突破3600亿，规模超过600台。

YDB的特点

千亿规模

每天千亿增量，总数据量可达几万亿

低延迟

数据从产生到能查询，根据配置的不同一般在十几秒到几分钟

查询快-高性能

常规查询毫秒级响应
常规统计秒级响应。

实时搜索

长文本字段可以进行根据关键词进行全文检索模糊匹配，并且有较高的性能

多维钻取

支持上万个维度，任意组合查询，任意维度组合过滤、分组，统计、排序。

容灾可靠

索引存储在分布式文件系统中，不因硬件的损坏或异常宕机而丢失数据。

YDB的功能

- 多维检索

支持 =,<>,>,>=,<,<=,in 以及全文检索, 模糊匹配

- 统计

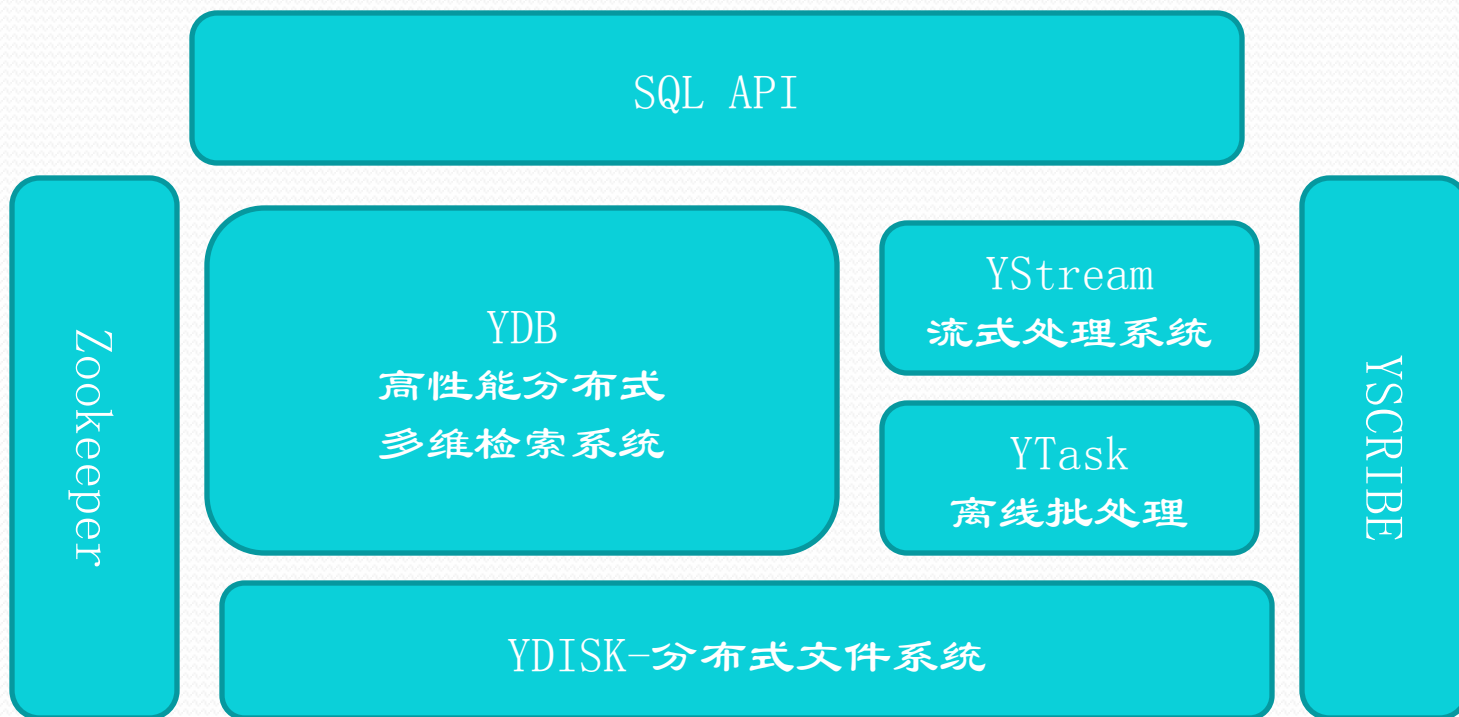
支持count,sum,max,min,avg

- 多列分组 (group by)

- 多列排序 (order by)

- 千万级别数据导出下载(新增)

YDB系统成员概览



YDB检索过程示意图

请求分发与合并

独立计算

分布式文件系统

用户查询
SQL API

查询引擎

子查询一

查询引擎一

子查询二

查询引擎二

子查询...

查询引擎...

子查询N

查询引擎N

读取索引一

索引分片一

读取索引二

索引分片二

读取索引...

索引分片...

读取索引N

索引分片N

YDB对故障的处理

指标	描述
存储节点故障	数据存储在YDISK上，多份冗余，数据不受机器硬件故障影响
查询服务节点异常	服务异常可以感知的自动重启，硬件损坏，服务可以自动迁移到其他机器上继续服务。

应用场景-YDB能做什么？

应用场景-用户画像



熟话说：知己知彼，百战不殆。用户画像就是在基于大量真实数据的基础上，从多个纬度真实反应用户特征，挖掘用户需求。从而在产品的设计、运营策略等环节上，根据目标受总用户的特征进行更符合用户需求的设计和运营方式。具体到不同行业会有不同的应用方式和应用场景。例如：广告行业会进行精准广告推送，交易类平台进行用户个性化推荐，内容类网站进行内容优化和内容推荐。总之用户的特征和需求明确后，所有的活动将具有针对性，有助于提高服务的质量和产品的ROI

棱镜门-大数据监听计划



棱镜计划（PRISM）；是一项由美国国家安全局自2007年起开始实施的绝密电子监听计划。

根据斯诺登披露的文件，棱镜"监控的主要有10类信息：电邮、即时消息、视频、照片、存储数据、语音聊天、文件传输、视频会议、登录时间和社交网络资料的细节都被政府监控。通过棱镜项目，国安局甚至可以实时监控一个人正在进行的网络搜索内容

YDB可以为这其中的海量数据提供实时的存储以及即席的搜索服务。

因YDB的数据时效性较高，并且检索速度很快，该领域未来在工信部以及公安系统上会有较大的应用前景。

应用场景-精准广告营销

● 个性化推荐

是否想过有一天当你在地铁上、公交上、电视上、马路上的大型电子显示屏或墙壁广告，他们可以感应到你，播放的广告都是为你量身定制的，都是你真正需要的。

● 富余服务能力的消化与精准投放

你是一个咖啡店主，当你的店比较清闲的时候，是否想过使用YDB搜索下周围3公里内的小资人群，告诉他们你这里有一个比较雅致的咖啡店，而且给他打5折，而当你的咖啡店人员比较满的时候就不在推送这些折扣服务。

● 阿里和腾讯已经涉水，你还在等什么？

腾讯的ADS与阿里的达摩盘升级为公司的战略产品。

YDB为此而生，立即构建属于你的DMP（大数据市场）吧。

智慧运营商

- 运营商拥有多年的数据积累，其数据资源的广度和深度是移动互联网企业难以相提并论的。每天数千亿的数据如何及时并快速的分析，是摆在运营商面前的一个难题。
YDB本身就是一个千亿级的实时即席分析系统，可以在如下几个方面帮助运营商管理数据。
- **流量精细化**
利用YDB保存详细的客户终端信息、手机上网行为轨迹，上网时长等数据。可以即席的检索出用户的上网记录，可以针对众多的客户投诉，提供更精细化的客户服务。
- **套餐精准营销**
基于客户的位置、话单，上网行为等，实时的筛选用户，进行更加精准的电话营销。
- **舆情监测-预防犯罪**
对短信，用户行为，位置进行实时的即席分析，进行多维关联，打击与预防犯罪。
- **基站优化**
实时的用户流量监控，对链接负载严重的基站进行升级扩容。
- **对外数据服务**
运营商可以利用得天独厚的大数据资产优势，利用YDB千亿级的大数据处理能力，将数据封装成服务，提供给相关行业的企业用户，为合作伙伴提供数据分析开放能力。

应用场景-图书馆与专利检索系统

- 仅仅根据书名，类目的检索太弱了，有没有想过我输入一段话，或一个人名，告诉我在哪本书的第几页里出现过。
- 一本书通常有几百万字，一个图书馆通常有数百万的藏书，这个搜索量不是普通数据库所能解决的。
- 国家专利局达几百T的专利数据的全文检索。

换成YDB吧，轻松帮你解决。

智慧城市

- 交通流量热点展示

利用延云的流式系统可以实时的对每个监控点的数据进行采集、统计与计算，从而实时的获取车辆流动信息以及交通拥堵情况。

- 交通规划

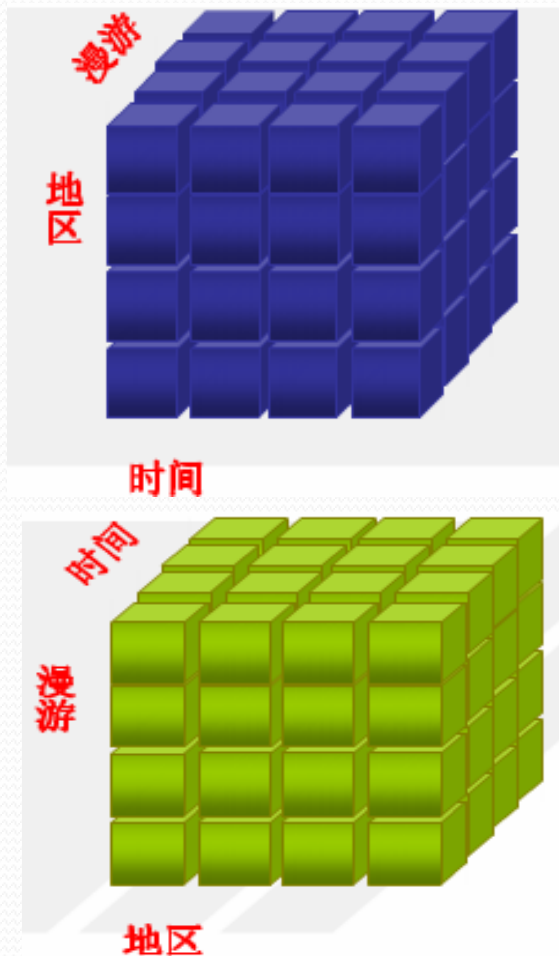
通过延云的 交通流量热点展示 大屏。

可以了解人群动向，对交通道路规划建设有参考作用。

千亿日志全文检索与多维分析

- 物流系统，网站，运营商，证券交易所，零售商每天有大量的销售，访问日志。经常需要对这些日志进行分析、过滤、筛选 从而发现系统潜在的问题。
- YDB在承接了大量千亿级的系统，对千亿级别的数据进行低延迟的导入，快速的多维分析，多维筛选。
- YDB可以做到每天接入 3000亿的系统日志，总数据量可达3万多亿。

OLAP-多维分析



- 切片与切块

在确定某些维数据的情况下对其他维进行观察，在多维结构中对二维数据切片，三维数据切块。如对城市，时间，漫游维度进行切块，可以得到城市的漫游费用情况

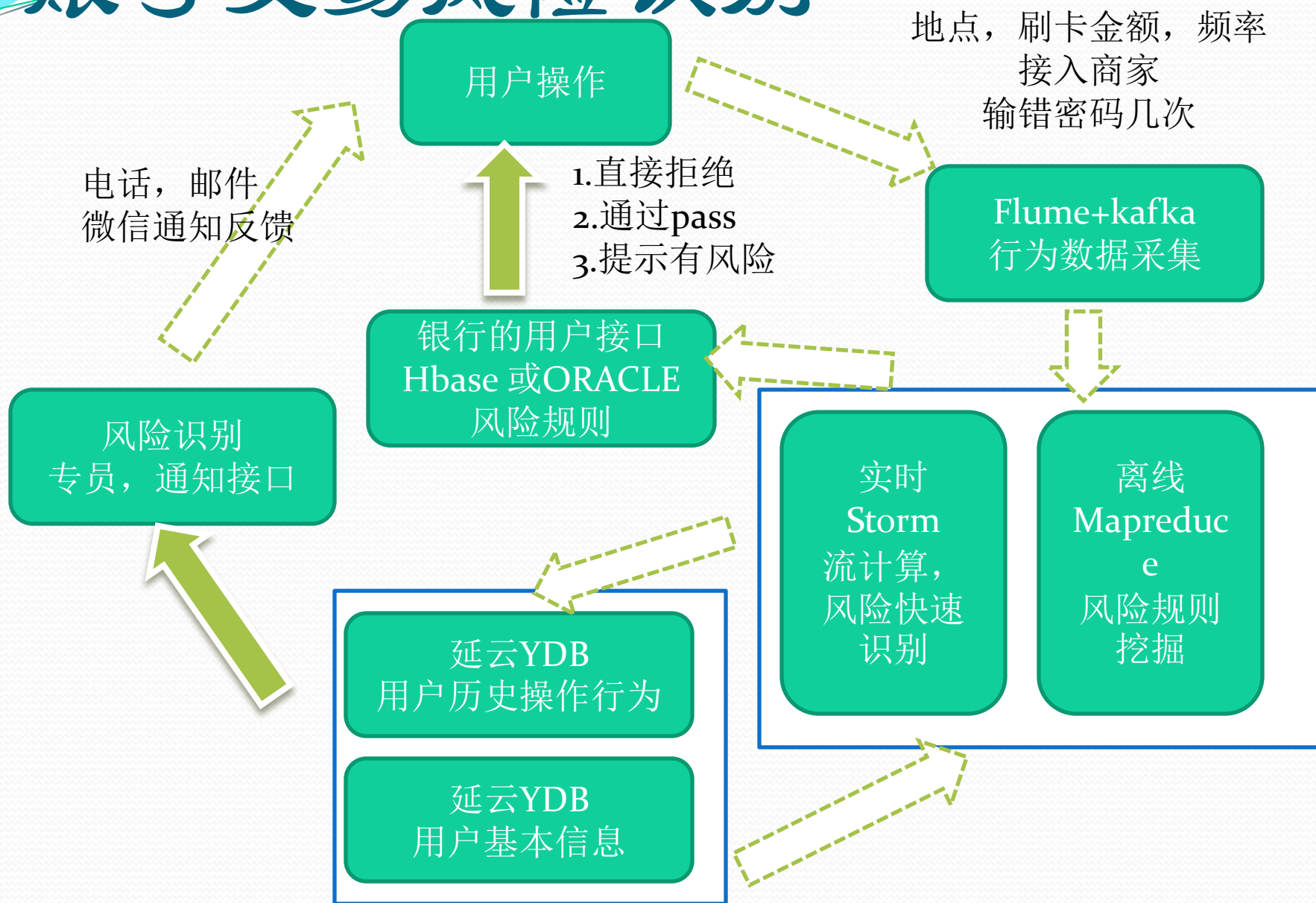
- 钻取

- 可以再一个维度从高到低或者从低到高钻取，了解不同深度的数据情况

- 旋转

对数据按照不同维度组织与考察

账号交易风险识别



YDB相对于Mdrill的优势

- 索引存储在HDFS中。硬件损坏-数据与任务自动迁移无须人工干预。
(原先需要数小时的索引迁移恢复时间)
- 数据时效性提高(由原先的几小时到现在的几分钟)。
- 每天数据增量由离线几十亿到现在的实时导入千亿。总数据量由百亿到现在的万亿规模。
- 支持的功能更多, 多列order by, 数据导入过滤, 导出下载, 拓展统计UDF。
- 简化部署与运维成本-配置文件也简洁很多。
- 更稳定-修正在某些场景下Mdrill里的死锁问题。

YDB相对于Mdrill的优势

- 动态加表，删表不需要重启，分区不在仅仅限制只能时间分区，用户任意配置分区方式。
- 过载保护、熔断机制，不因用户的误用导致服务系统宕机。
- 更少的内存消耗，更高的执行效率，全文检索模式普遍性能提升10倍以上，部分查询统计效率提升10倍以上，解决因列值数据倾斜导致的查询性能很慢。
- 来自企业技术专家的全程技术支持。
- 有过载限速逻辑，有效控制系统峰值系统负载过高的问题。
- 索引创建效率更高-IO次数明显减少。

YDB相对于SOLR、ES的优势

- 常态下solr与es的数据总量在千万以内的规模，如果数据达到千万到几亿之间通常采用离线build的方式，数据延迟较大，几乎没有超过百亿规模的索引。

YDB的常态数据规模为每天几十亿到千亿之间，总量可达万亿。

- DocValues无须在内存中保留索引与每个Segments之间的映射关系，大幅度的节省了内存。

SOLR当数据规模达到千万以后，内存与CPU的使用消耗都非常高，而且由于索引的频繁创建与合并，这种映射关系也会频繁的变化，从而导致磁盘IO也会很高，这个时候solr与ES的使用者就不得不降低数据的时效性，采用离线批量导入的方式，时效性会严重下降。

- 每条数据都要先根据uniqKey对每个索引碎片segment检索一遍，在继续添加。

这种方式如果数据量少，没有任何问题，但是ydb的系统，通常都是每天百亿的写入量，这种方式会造磁盘随机IO特别的高，吞吐量严重下降。

ydb采用append方式写入数据，没有uniqKey的限制，索引的写入采用“近似”顺序写方式写入到磁盘。

YDB相对于SOLR、ES的优势

- Solr的索引持久化的打开，即使不用，也一直占用内存资源，而YDB则采用按需加载索引的方式，不经常使用的索引，放到hdfs上，不占用内存等系统资源。
- 数据倾斜问题
倒排检索即使某个词语存在数据倾斜，因数据量比较小，solr也可以将全部的doc list都读取过来（比如说男、女），这个doc list会占用较大的内存进行Cache，当然在数据规模较小的情况下占用内存不是特别多，查询命中率很高，会提升检索速度，但是数据规模上来后，这里的内存问题越来越严重。
- Fq的Cache但数据规模变多后，cache利用率变低。
Fq采用openbitset保存匹配结果，1亿个bit约使用10m内存，缓存不了太多的key。
YDB则采用独有的缓存的方式，有效的提高缓存利用率，有效解决了opentbitset中空置过高暂用太多内存的问题。

YDB相对于SOLR、ES的优势

- 索引合并是Docvalue需要3次读取原数据，这三次还需要在内存中构建映射关系，而且读取字典有大量的随机IO。这也是为什么solr用久了，有的时候会发现突然“抽风”，CPU与磁盘瞬间IO变高的原因。
YDB拓展了这个地方，只需要1次读取原数据.IO次数明显减少。
- SOLR与Es的索引存储在本地磁盘
SOLR与ES尽管提供了hdfs的接口，但是采用直接裸读hdfs方式，性能太差。YDB的索引是存储在hdfs里的，但在读取接口上做了较多的优化，1.减少了访问namenode的频率，2. 添加了BlockBufferCache，对频繁访问的数据块进行了缓存。

只需六点教您辨识延云YDB

- **暴力扫描。**

暴力扫描的方式，不用多说，性能很差，需要狂堆机器，成本也太高。而这类系统一般的并发也不大，如果数据量在百亿级别，千台的集群规模，一天也就能进行几十万次的查询而已。

YDB则采用大索引技术，通过索引技术直接定位到相关记录，避免记录的逐条扫描，即使只有50几台的机器，百亿数据也能查询个几百万次。

- **预计算-不留原始数据,只能看特定的维度、粒度。**

如果数据经过预先汇总，原始数据没有保留，在未来某一时间如果想查看其它维度或者粒度的数据将无法实现。

- **YDB则保存每条数据的原始数据，没有经过汇总，查询的时候根据大索引技术以及独特的标签技术，在几秒的时间，返回任意维度，任意筛选条件的结果，灵活性很好，多维钻取，全文检索都是YDB的强项。**

- **只能局部计算，本质KV。**

这类系统最大的问题无法进行任意维度的筛选，因为数据时基于KV，做任何检索的前提必须要有K，虽然可以针对多个维度分别创建KV索引，但是一旦索引的维度特别多，创建索引的性能就会严重降低，而且大部分场景下我们的数据都会有数据倾斜-有数据倾斜值的列在查询的时候，因为value特别巨大，会出现内存不够的情形，也会导致索引失败。

YDB真实场景支持了长达3万个列的索引的创建，通过多级跳跃表结合查询上的巧妙处理，完美的解决了数据倾斜的问题。

只需六点教您深度辨识延云YDB

- 我们以服务的方式运行，托管你的数据。

YDB则采用交给客户独立部署的方式，但YDB提供全程的技术支持。延云公司不会接触到您的数据，您不用担心您的数据会被窃取。

- 全内存计算-海量SSD硬盘

技术实现采用将数据全部LOAD到内存中，或者使用大量内存，计算的时候通过SSD硬盘与内存进行快速的内存交换，这种时间可以保持较高的运算速度，但是运算成本太高，每天100亿，保留一个月3000亿的数据，这种严重依赖SSD盘以及内存的系统需要千台机器的成本。

YDB数据时放到HDFS上的，并没有load到内存中，但是由于采用标签标记技术，重复的值只存储一次，与磁盘交互的IO很少，如果是上述场景，YDB大约使用30台的常规机器，而且也不需要SSD盘。

只需六点教您深度辨识延云YDB

- 索引存储在本地磁盘，但可以通过副本恢复。

传统数据库的经典方案，因为数据量少，恢复也很快，所以很使用。但是并不适用于大数据的场景，真实运维经验表明，大数据场景的机器数据写入量都特别的大，频繁的磁盘读写，硬盘损坏的情况经常发生，一般一台机器有2T*10的存储空间，以目前的磁盘读写速度至少需要10多个小时，甚至几天的恢复时间，这种数据恢复速度大部分业务是不能够接受的。

通过副本技术来保障高可用的显著缺点是有多少个副本就要额外冗余出多少倍的机器，CPU内存的消耗都是副本数量的倍数。而且在CAP原则里，数据的一致性的保证很难很难。

YDB是将数据存储在海DFS中，机器宕机，服务进程会迅速的切换到其他的机器，快速的恢复服务，数据也不会丢失。

业务合作

- 电话：024-83653716
- 微信：ycloudnet
- 邮箱：ycloudnet@163.com
- QQ：1820150327
- 主页：<http://www.ycloud.net.cn>
- 地址：沈阳市浑南区新隆街万科明天广场



谢谢