


(/apps/download?utm_source=sbc)

如何实现海量数据下有序漏斗秒查



蓝色背影 (/u/d2d4e1fd6c73)

+ 关注

2017.11.07 20:41*

字数 1721

阅读 643

评论 0

喜欢 1

(/u/d2d4e1fd6c73)

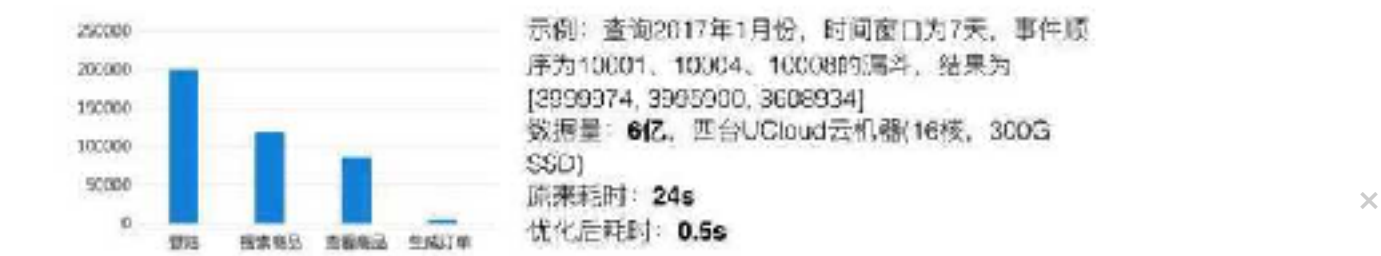
近期易观 (<https://link.jianshu.com?t=http://link.zhihu.com/?target=https%3A/www.analysys.cn/>)公司举办了一个OLAP大赛，我们队伍非常荣幸地获得了第一名，成为本次比赛最大黑马。此篇文章主要分享一下我们是如何解决有序漏斗秒查问题的

比赛地址：2017易观OLAP算法大赛 (<https://link.jianshu.com?t=http://link.zhihu.com/?target=http%3A/ds.analysys.cn/OLAP.html>)

参赛情况: <https://www.analysys.cn/media/detail/20018458> ([https://link.zhihu.com/?target=https%3A/www.analysys.cn/media/detail/20018458/](https://link.jianshu.com?t=http://link.zhihu.com/?target=https%3A/www.analysys.cn/media/detail/20018458/))

1. 题目分析：

假设我们在购买商品的过程中，需要触发的事件包括“启动”，“登陆”，“搜索商品”，“查看商品”，“生成订单”等。运营人员需要分析某段时间内，在全部用户中依次有序触发“登陆”→“搜索商品”→“查看商品”→“生成订单”事件的人群的转化流失情况，并且事件从第一个到最后一个有时间窗口限制，以完成的最长路径为准。



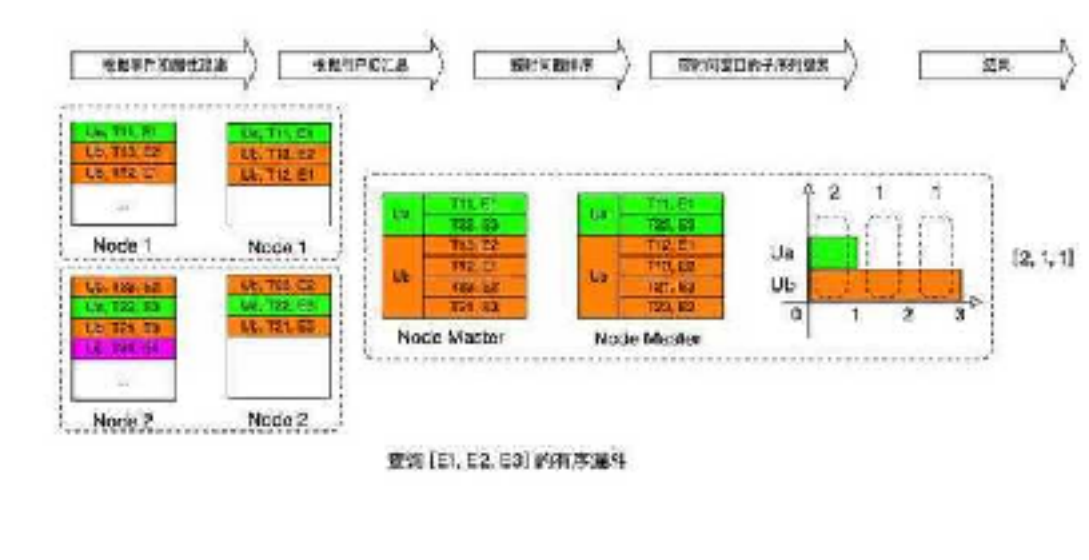
详情：2017易观OLAP算法大赛 (<http://ds.analysys.cn/OLAP.html>)

在以上示例场景下，我们在易观提供的6亿测试数据集上，在4台UCloud云主机 (16core，16G ram)机器下从24s优化到了0.5s。而在正式比赛的26亿数据集上，使用相同硬件环境，耗时1.6s。

2. 解题分析：

题目描述的有序漏斗问题可以归结为带滑动时间窗口的最左子序列问题，比如我们需要寻找，2017年7月份中，在3小时的时间窗口下，[A,B,C,D] 漏斗路径下的转化情况，单个用户只能有 NULL，[a]，[A,B]，[A,B,C]，[A,B,C,D] 五种转化结果，对应的漏斗深度我们称之为level，在[A,B,C,D]漏斗路径下，level的取值可以有[0,1,2,3,4] 四个值，题目的要求即算出所有用户的满足条件下最大level汇总结果。

理解问题之后，我们梳理了一下流程图：



我们将问题解决分为5个步骤：

filter阶段：根据时间区间和事件属性对数据进行过滤

group阶段：根据userId进行group汇总

sort阶段：按照时间进行排序

algorithm aggregate 阶段：带时间窗口的子序列搜索

合并结果

(/apps/download?
utm_source=sbc)

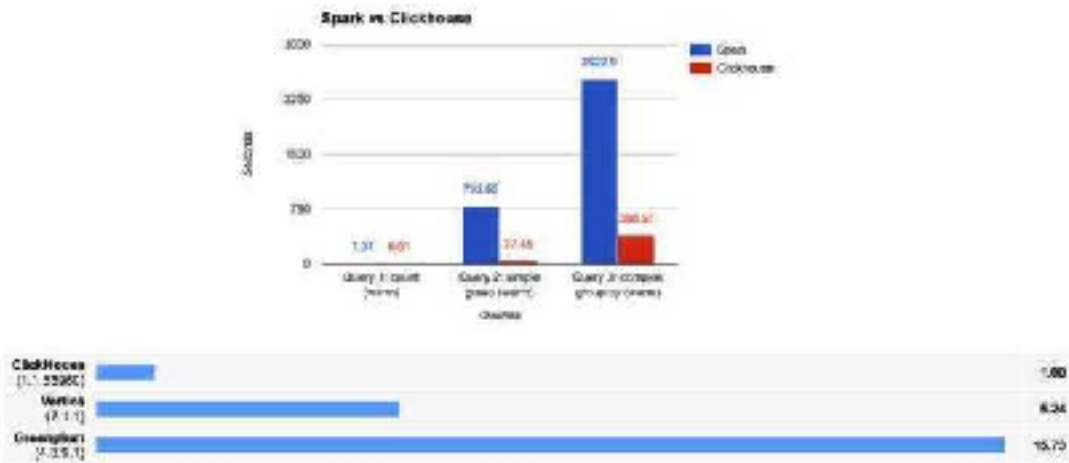
3. 数据库选型：

根据以上分析，需要filter，group，sort，aggregate等操作，数据库是必备的核心，而在OLAP领域，开源的数据库选型有很多，比如：mysql, druid, kylin hdfs + (hive,spark,presto)，imapla, kudu etc。

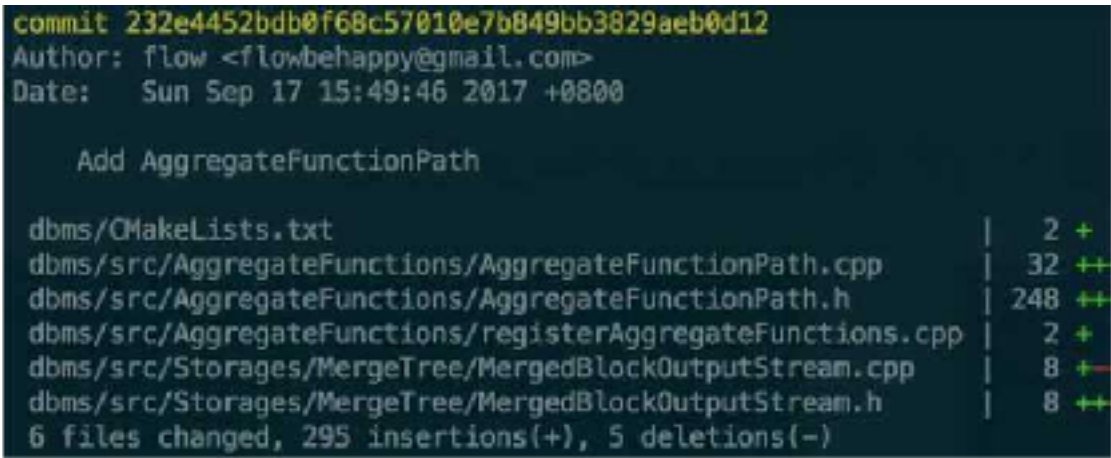
但在这个场景下，结合以往对其他数据的深入研究分析对比经验，我们几乎毫不犹豫就选择了ClickHouse (<https://link.jianshu.com?t=http://link.zhihu.com/?target=https%3A/clickhouse.yandex/>)(纵然它不支持udaf)，，我们相信ClickHouse是目前cpu领域最快的olap开源数据库，它最突出的优点就是快，如果你是第一次用，相信ClickHouse会让你感到非常惊艳。



ClickHouse 由俄罗斯Yandex开发，09年原型，12年生产可用，16年开源，目前最大的线上部署实例是 Yandex.Metrica: 472个节点，每秒处理2T数据，实时在线分析。ClickHouse 在OLAP上的查询性能非常彪悍，平均查询性能几乎是vertica的三倍。



ClickHouse不仅速度快，它系统架构灵活，性能优越，代码优雅，非常适合大数据下需要极致性能的应用场景。ClickHouse目前暂不支持UDAF，但没关系，我们可以通过修改源代码并重新编译来实现自定义AggregateFunction。

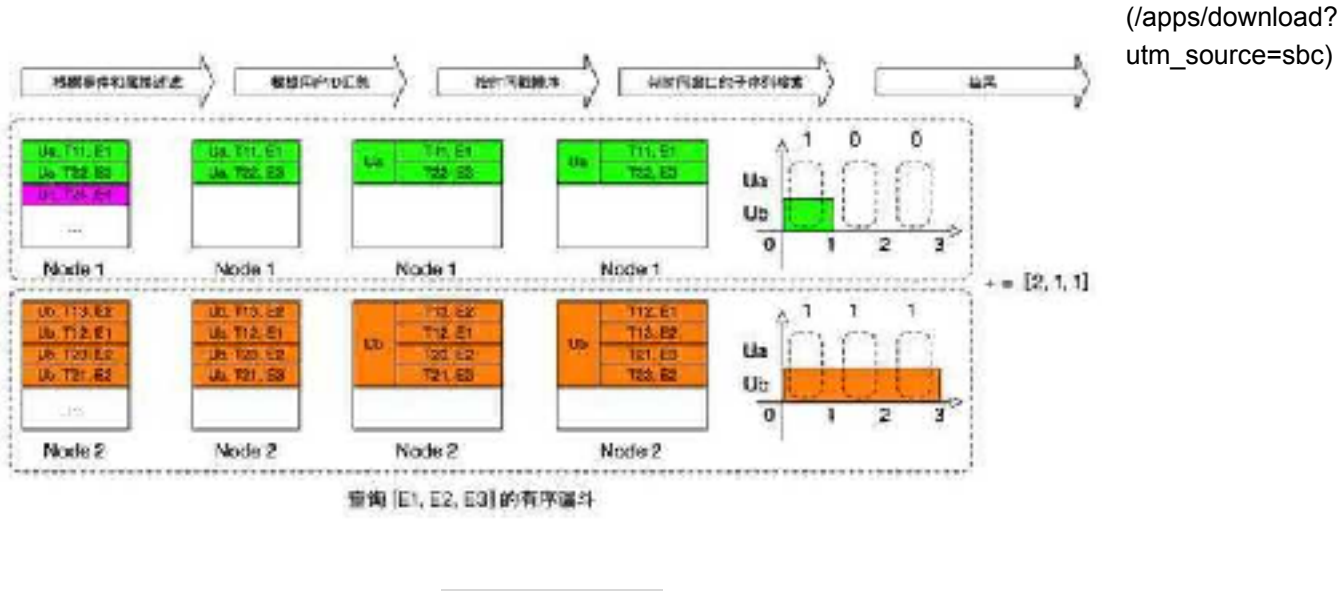


以上就是针对漏斗场景的代码修改情况，可以看出我们只用了不到300行代码就为ClickHouse加入了漏斗计算（ aggregate function path ）的功能。

对比官方的presto + hdfs 方案实现的24s速度，使用ClickHouse之后，我们在测试环境下跑的速度达到了8s。

下面开始我们的正式优化过程

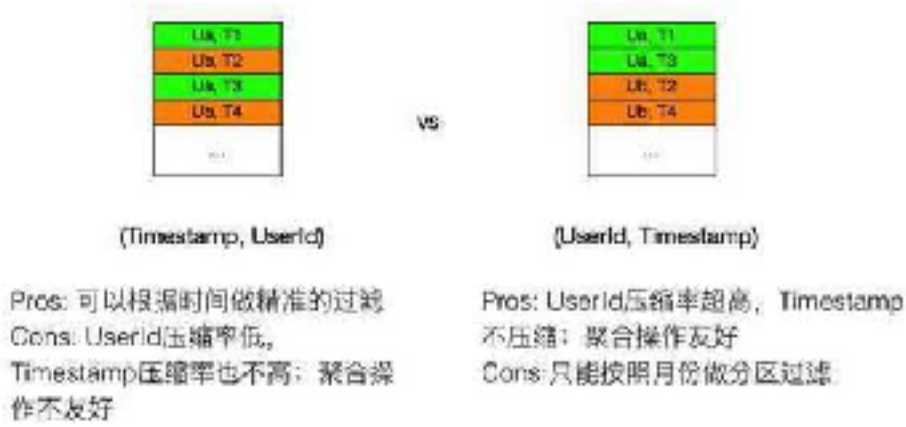
4. 按照用户ID分区：



我们注意到，漏斗的计算中，每个用户都是相互独立的，所以我们可以将数据按照uid来分区，这样就将数据分散到了四台机器上，我们可以分别向每个数据库节点发送请求，然后将数据进行汇总，得到最终结果。

通过这次优化，我们在测试环境下跑的速度达到了1.6s。

5. 以(uid,timestamp)作为primary key

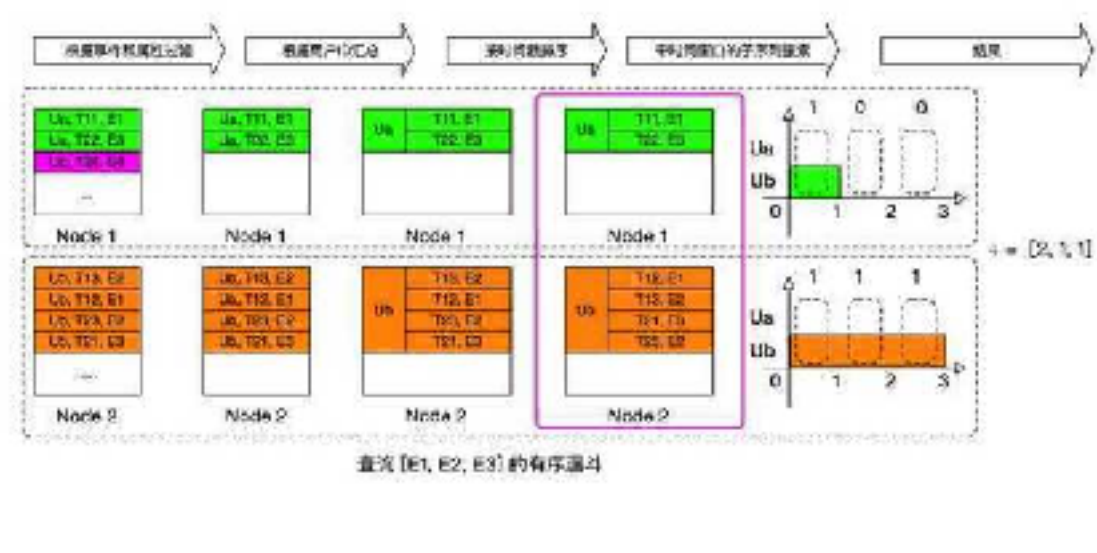


ClickHouse中primary key代表了数据的组织排序方式

左边的以（timestamp，uid）为主键，时间是有序的，方便做时间精准过滤，但uid压缩率低；右边的以（uid,timestamp）为主键，uid压缩率高，方便做uid group，但对时间过滤支持不够好。

通过测试对比，我们发现以（uid,timestamp）作为主键性能略快，查询时间达到了1.4s。

6. 分组预排序



当我们以（uid,timestamp）作为primary key后，分组内的数据其实已经有序了，我们可以去掉代码中的sort方法，来提高性能，经过这个优化，查询时间达到了0.9s。

7. 带时间窗口的子序列搜索优化



(/apps/download?
utm_source=sbc)

这里主要是用了一些 剪枝的策略，当我们从左往右去搜索 a,b,c,d 漏斗的时候，我们需要找到最大的深度，必须一直去搜索以a开头的子序列；但我们从右往左搜索时，我们只要考虑比当前结尾更大的子串即可，比如我们找到了 a,b,c，后面我们只需要考虑以d结尾的串，这样减小了搜索的复杂度，查询时间达到了0.8s。

8. 数据结构优化

事件ID到数组下标Index的映射，我们直接遍历了数组搜索，而不使用 std::unordered_map, 因为在events数据量不大的情况下，数组搜索O(1)比O(n)慢。

使用纯C++数组存储事件序列，不使用std::vector，去掉了vector的开销，灵活控制内存分配。

经此优化，查询时间达到了0.6s。

9. 部分压缩

数据库通常会对字段进行压缩，这样做节省了硬盘空间，但却浪费了cpu计算，为了提供性能，我们对字段进行了部分压缩，经此优化，查询时间达到了最终的0.5s。

uid => 压缩

timestamp => 不压缩

event_id => 不压缩

event_name => 压缩

event_tag => 压缩

date => 不压缩

总结：

我们已经将代码和PPT开源：

<https://github.com/analysys/olap> (<https://link.jianshu.com?t=http://link.zhihu.com/?target=https%3A/github.com/analysys/olap>)

0.5s当然不会是极限，当我们在这条路上渐行渐远的时候，我们发现做数据库的研究其实是一件非常有意思的事情，我们来自广州向量线科技，一家专注下一代分析技术的初创公司，期待各路人才的加入，欢迎私信勾搭。

小礼物走一走，来简书关注我

赞赏支持

📖 日记本 (/nb/2622844)

👤 举报文章 © 著作权归作者所有



蓝色背影 (/u/d2d4e1fd6c73)

写了 1721 字，被 4 人关注，获得了 1 个喜欢

(/u/d2d4e1fd6c73)

+ 关注

^

🔗

喜欢 | 1



更多分享

(http://cwb.assets.jianshu.io/notes/images/2480863/weibo/image_2 (/apps/download?utm_source=sbc)



(/apps/download?utm_source=nbc)



登录 (/sign_in?utm_source=desktop&utm_medium=not-signed-in-comment-form)

评论

智慧如你，不想发表一点想法 (/sign_in?utm_source=desktop&utm_medium=not-signed-in-nocomments-text)咩~

(/p/cdbed82a34bc?



utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommendation)

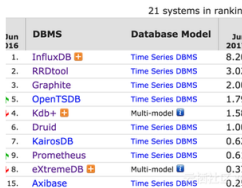
[3/4]我所经历的大数据平台发展史（三）：互联网时代•上篇 (/p/cdbed82a...

//我所经历的大数据平台发展史（三）：互联网时代•上篇http://www.infoq.com/cn/articles/the-development-history-of-big-data-platform-paet02 编者按：本文是松子（李博源）的大数据平台发展史...

葡萄喃喃呓语 (/u/2c67926c48ce?

utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendation)

(/p/9ad69141b67a?



utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommendation)

时间序列数据的存储和计算 - 开源时序数据库解析 (/p/9ad69141b67a?utm...

开源时序数据库 如图是17年6月在db-engines上时序数据库的排名，我会挑选开源的、分布式的时序数据库做详细的解析。前十的排名中，RRD是一个老牌的单机存储引擎，Graphite底层是Whisper，可以认为是一...

阿里云栖社区 (/u/12532d36e4da?

utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendation)

(/p/e789050e523a?



utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommendation)

数据库的简单学习 (/p/e789050e523a?utm_campaign=maleskine&utm_c...

需要原文的可以留下邮箱我给你发，这里的文章少了很多图，懒得网上粘啦 1数据库基础 1.1数据库定义 1) 数据库(Database)是按照数据结构来组织、存储和管理数据的建立在计算机存储设备上的仓库。简单来说...

极简纯粹_ (/u/08a03646e5f4?

utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendation)

深度访谈：华为开源数据格式CarbonData项目，实现大数据即席查询秒级...


华为宣布开源了CarbonData项目，该项目于6月3日通过Apache社区投票，成功进入Apache孵化器。 CarbonData是一种低时延查询、存储和计算分离的轻量化文件存储格式。那么相比SQL on Hadoop方案、...

曹振华 (/u/7e92f630a7d8?

utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendation)

掘金 Android 文章精选合集 (/p/5ad013eb5364?utm_campaign=maleski...

用两张图告诉你，为什么你的 App 会卡顿? - Android - 掘金Cover 有什么料？从这篇文章中你能获得这些料：知道setContentView()之后发生了什么？ ... Android 获取 View 宽高的常用正确方式，避免为零 - 掘金...

 掘金官方 (/u/5fc9b6410f4f?utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendation) (/apps/download?utm_source=sbc)


(/p/ab86db16a741?



utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommendation)

那些相机错过的镜头 || 萌宠记 (/p/ab86db16a741?utm_campaign=malesk...

图文/陆二蛋 对我我这种不识货的摄影爱好者来说，相机和手机，最大的差距大概就是像素了。所以一旦染上了单反(佳能80D)的瘾，就很抗拒手机的渣像素（ vivo x5 pro ）。 但是，单反这种东西，肯定不是时时都...

 鹿二蛋 (/u/ee32c698fd2e?utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendation)


(/p/5dcdb3e6753b?



utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommendation)

忘记悲伤 (/p/5dcdb3e6753b?utm_campaign=maleskine&utm_content=...

想哭，就痛快地哭 泪水止不住，就让它尽情地流 别再傻傻地， 抬头看天空， 指望眼泪能收回。 不是都说覆水难收， 为什么还这么拼命的， ...

 云栗 (/u/afec5a6be8a4?utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendation)


(/p/09948969ae2d?

设计梦里不见改稿的车
'方也曾经叹过苦无知
,奈杯酒入肚却一暮莫
'不知你已耗尽荏苒光

utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommendation)

甲方和设计师天生就是一对儿冤家 (/p/09948969ae2d?utm_campaign=ma...

甲方和设计师天生就是一对儿冤家，这样的爱恨情仇要多少年才能够化解？ 设计行业存在少说也几十年了，但是至今为止，甲方和设计师之间的矛盾就没断过，互相吐槽、互相埋怨。 甲方认为设计师不够专业做不...

 3D学苑 (/u/040c045fbe70?utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendation)


(/p/08df92144935?



utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommendation)


有温度的城市|合肥：这座城，这些人，这片景 (/p/08df92144935?utm_ca...

01. 两张关于孩子的图片，在合肥的天鹅湖捕捉下来。那刻，虽不能参与却能深深感受到他们世界的美好，一个专属孩子的世界。 02. 作品名：公交站的老人（之一） 拍摄者：白衡 后期：snapseed 拍摄器材:手机这...

 白衡 (/u/4bdc0b179419?utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendation)

校园|百辞令·归校 (/p/45c46f99e3c7?utm_campaign=maleskine&utm_co...

想 返航 白雪茫 叶泥护壤 黄昏人影长 轻装笑问故乡 红炉相拥梦酣香 翘首倚待春日模样 不觉以往鞭炮声声响 欲问行人处打马归学堂 云淡风清天朗抬眸望 但见草木欣欣向上 万物复苏新气象 且看春梅怒放 孤芳不自赏 ...

 东湖一枝春 (/u/5851e9fec73d?utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendation)

